CLINICAL ASSOCIATIONS AND GENETIC ALTERATIONS TO PREDICT RADIOTHERAPY TREATMENT RESPONSE IN PATIENTS WITH TRIPLE NEGATIVE BREAST CANCER (TNBC)

by

Ernest E. ONUIRI

A Dissertation Submitted

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biomedical Informatics

Department of Health Informatics
Rutgers, The State University of New Jersey
School of Health Professions

August 2020

RUTGERS
School of Health Professions

Final Dissertation Defense Approval Form

CLINICAL ASSOCIATIONS AND GENETIC ALTERATIONS TO PREDICT RADIOTHERAPY TREATMENT RESPONSE IN PATIENTS WITH TRIPLE NEGATIVE BREAST CANCER (TNBC)

by

Ernest E. ONUIRI

Dissertation Committee:

Frederick Coffman, Ph.D. Committee Chair
Antonina Mitrofanova, Ph.D. Committee Member
Shankar Srinivasan, Ph.D. Committee Member

Approved by the Dissertation Committee:

_____ Date: _____

_____ Date: _____

_____ Date: _____

_____ Date: _____

TABLE OF CONTENTS

ABSTRACT

Despite the major advances in healthcare over the past century, the successful treatment of cancer has remained a significant challenge, and cancers are the second leading cause of death worldwide behind cardiovascular disease. Breast cancer is the most prevalent cancer in women, and an aggressive and difficult to treat breast cancer variant which tends to appear in younger patient populations is Triple Negative Breast Cancer (TNBC). Post-surgical adjuvant radiotherapy is frequently utilized in patients with TNBC, however little is currently known about which patient populations would significantly benefit from this procedure. In this study we take a deep look at 190 TNBC patient samples from the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) database, sourced through cBioPortal. Our goal was to identify genetic and clinical features that can be used to better understand the effects of adjuvant radiotherapy on post-surgery TNBC patients, and to build a predictive model for the identification of patient populations which would benefit from post-surgery adjuvant radiotherapy, and those which would receive minimal benefit from this procedure. Exploratory data analysis and Kaplan Meier analysis identified statistically significant genes based on Fisher's Exact Test and PAM50 (Prediction Analysis of Microarray 50) classifications. KNN machine learning algorithm was used to build the predictive models incorporating both clinical and genetic features. All analyses were carried out on RStudio 3.6.0 and cBioPortal's Onco-Query Language (OCL). The final optimized model was very efficient, with returned parameters as follows: Accuracy – 94.64%, Kappa Statistic – 88.89%, Sensitivity – 100%, Specificity – 87.50% and AUC(ROC) – 93.75%. Four genes and five clinical make up the core features of the model: the genes are AKT1, MAP3K1, MEN1 and SHANK2, while the clinical features

are survival months, survival groups, breast surgery, NPI (Nottingham Prognostic Index) and tumor size. Not only does this model have utility in making TNBC treatment decisions, but it demonstrates that a useful predictive model of cancer therapeutic responses can be constructed using a reasonable number of input parameters, even in a highly heterogeneous disease.

ACKNOWLEDGEMENTS

I give God all the praise for making it possible for me to against all the odds, embark on this journey that has brought a huge amount of reward to me as an individual and to my career. I am appreciative of all my teachers and especially my dissertation committee members on whose backs I rode to this success. To Dr. Frederick Coffman, you have been a blessing to me as regards this journey, thank you for your mentorship and continuous support. You always had my back even in difficult times and played a huge role in helping me achieve this feat. To Dr. Antonina Mitrofanova, I cannot thank you well enough. My foundation in quantitative research, necessary to thrive in the field of Biomedical Informatics was fast-tracked by your committed tutelage. I enjoyed those sessions with you because I learnt a whole lot in a short period. You kept me on my toes in a way that fostered the required growth, necessary to become what I am today. To Dr. Shankar Srinivasan, you make up the triage of the quality mentoring I enjoyed throughout my studies. You were prompt to assist every time I ran to you for one form of advice or another both administrative and academic. I earnestly look forward to future collaborations with you all as my academic mentors.

To my father, Godfrey Onuiri and my siblings Elizabeth, Jane, Victoria and Angela, thank you all for all the roles you play in different special ways. To my parents-in-law, Dr. Silvanus Chioma and Mrs. Joy Chioma, thank you for the periodic words of encouragements and prayers.

DEDICATION

To my loving wife Patricia, thank you for all you do. For the love, support, and endless encouragements. Thank you for graciously managing the ship while I got deeply engaged throughout this journey. I love you always. To my three children, Adanna, Chukwudiebube and the one yet unborn, you are the reason I strive for excellence. I love you all to the stars and back.

In loving memory of my dear mum, Catherine. I am what I have become largely because of all you showed and taught me. You will forever be in my heart.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background of the Problem

Cancer is a disease condition that results when gene mutation leads to the uncontrolled growth of cells in the body. It poses a major concern to mankind, especially since lifestyle changes associated with the current and immediate past centuries suggest that we now do all the things that ordinarily aggravates cancer risk factors.[1-3] More and more people in the last century have succumbed to one form of cancer or the other. One of such cancers is Breast Cancer, known to affect both males and females, but much more prevalent in females than it is in males. According to a 2019 World Health Organization (WHO) report, breast cancer is the most prevalent cancer ravaging the women population. It affects 2.1 million women each year and is responsible for most cancer deaths in women. The report went on to assert that in 2018 alone 627,000 women died from breast cancer which represents approximately 15% of total cancer deaths among women in that year. The reports also suggest that whereas breast cancer rates remain high in developed regions, the gap is closing because the number of cases in the developing world is on the increase. This indeed makes the breast cancer problem a global one.[4,5]

In recent years, a lot of research and funding have been channeled to breast cancer research. While it is also true that some breakthroughs have been recorded in genetics, treatment interventions and care plans, there is still more to be done to counter and arrest the devastation that breast cancer brings along with it. This research focuses on a breast cancer subtype – the Basal or Triple Negative Breast Cancer (TNBC). As the name

suggests, the TNBC tests negative for Estrogen Receptor (ER), Progesterone Receptor (PR) and Human Epidermal Growth Factor Receptor 2 (HER-2). Consequently, TNBC does not respond to receptor targeted treatments such as hormone therapy. In addition, depending on the stage of the diagnosis, TNBC can be very aggressive and more likely to relapse compared to other breast cancer subtypes. [6-9]

In view of the foregoing, whereas there is no gainsaying the fact that a lot of resources are continuously being channeled to research efforts aimed at improving treatment outcomes, there is the menace of resistance to treatment, which is usually difficult to predict.[10-12] From a precision medicine point of view, resistance to treatment is usually tied to the dynamics of both the genetic and epigenetic predispositions of the patient. This leads to questions as to why a patient survives and beats the cancer disease, while another patient with the same disease succumbs to it through death.[13-15]

1.2 Statement of the Problem

The cancer scourge is an agelong one that leaves sufferers either dead or having to "go through the valley of the shadow of death" before surviving. Recovery from cancer has never been a straight-forward path. The race for survival is often so grim, especially if detection is not early. Amid these uncertainties is also the unsavory dimension of cancer being a disease with a list of multiple causative factors that are vastly heterogenous. One of the very prevalent cancers, happen to be breast cancer, for which there are different types. Triple Negative Breast Cancer -TNBC, happens to be one of the most aggressive variants affecting about 20% of the breast cancer population and prevalent among African American and Latino women who are below the age of 40.[16-20]

With all the efforts in recent years to come up with suitable therapies capable of enhancing survival in patients with TNBC, not much has been achieved especially given that this cancer variant tests negative to hormonal receptors of interest pertaining to targeted treatment in cancer studies namely; Progesterone Receptor (PR), Estrogen Receptor (ER) and Human Epidermal (growth factor) Receptor– 2 (HER-2). Hence earning its name from this nomenclature, which also rules out hormonal therapy, thereby deeming it a non-viable treatment option. This limits therapy efforts to surgery, radiotherapy and chemotherapy in a combinational manner predicated on the genetic predisposition of the patient, while also recommending certain lifestyle changes.[16]

Furthermore, a well-documented fact in cancer studies in all its various forms is that its cause is vastly triggered by heterogeneous genetic mutations. Unfortunately, this heterogeneity varies from patient to patient, making the processes for deriving effective therapy nothing short of daunting.[21] It becomes important that this research makes effort to identify genes that play a part in the survival outcome of patients who undergo the combinational treatments of surgery and radiotherapy.

It is in view of the forgoing that the emphasis of this research work is placed on patients who undertake radiotherapy treatment after going through a surgical procedure. This research therefore seeks to investigate the chances of survival based on underlying clinical and genetic predispositions of TNBC patients.

1.3 Aim and Objectives

The main aim of this research is to ascertain the effects of germline and somatic mutations as well as Copy Number Alterations in the presence of underlying clinical

features, and how all these relate to the administration of adjuvant radiotherapy on post-surgery TNBC patients, to enable prediction of treatment outcomes.

The specific objectives are to:

1. establish the effects of adjuvant radiotherapy on post-surgery TNBC patients, both alone and in combination with other TNBC therapeutic modalities.
2. correlate genetic and transcriptomic alterations identified in these patient's tumors with the effects established in Objective 1.
3. utilize the relationships found in Objective 2 and relevant clinical features to build a predictive model for treatment response in patients with TNBC.

1.4 Hypotheses

1. Surgery is an important treatment method if TNBC patients are to survive beyond the first 5 years after diagnosis.
2. Adjuvant radiotherapy has the same effect on patients who undergo breast conserving procedure as with patients who undergo mastectomy.
3. Combination treatment approach such as surgery and radiotherapy offer a better survival outcome than the single treatment approach for TNBC patients.
4. The combination of mutation and clinical features can be used to predictively determine which TNBC patients would benefit from radiotherapy.

1.5 Significance of the Study

There is no doubt that in recent times, a lot of breakthroughs have been recorded in medicine, however, chronic diseases have continued to pose treatment challenges even to the best of healthcare methods and systems. While it is one thing to proffer solutions in the

form of potent treatments for a disease such as cancer, there is also the issue of treatment resistance leading to the undesired outcome of death. This research focuses on the resistance to treatment of TNBC patients based on their genetic and clinical predispositions. The treatment method of interest is the combination of surgery and adjuvant radiotherapy. In this study, 18 genes and 24 clinical features of interest were systematically identified. The combined 41 (multivariate) predictor features, and 1 output feature, formed the basis for the classification system derived after exploring five machine learning algorithms. To the best of my knowledge, no other research work has ever sought to build a predictive system using these collection of 42 features that this study adopts in a bid to ascertain TNBC patients that would benefit from adjuvant radiotherapy and by extension, have a good chance of survival or otherwise, based on their genetic and clinical predispositions. Research findings from this effort will be critical to improving treatment outcomes in the precision medicine domain as well.

CHAPTER 2

LITERATURE REVIEW

2.1 Preamble

Cancer is one of the leading causes of death worldwide with approximately 9.6 million deaths in 2018.[22-24] It accounts for 1 in 6 of all deaths globally, with a third of these deaths attributable to poor behavioral and dietary choices.[25-28] Cancer is defined as "a generic term for a large group of diseases that can affect any part of the body, in which abnormal cells are created rapidly and grow beyond their usual boundaries".[29-31] In addition, "Cancer is a broad term for a class of diseases characterized by the abnormal cells that grow and invade healthy cells in the body".[32-35] These abnormal cells are referred to as cancer cells, malignant cells or tumor cells.

Cancer does not only occur in humans; it also manifests in animals and other living organisms. In the body, normal cells grow, divide, and die (apoptosis). Cancer cells also grow and divide but they refuse to die. Instead of old and damaged cells to die, they survive, and grow and divide when they are not required.[36-38] In some cases, the cancerous growth spreads beyond the primary tissue or organ to secondary tissues and/or organs. The process in which cancer cells leave an area of the body to grow in another area of the body is known as metastasis.[39-42]

2.2 Types of Cancer and Prevalence

There are over 200 types of cancer. Types of cancer are named after the tissue or organ upon which the cancer cells are formed.[43] This breaks cancer into types:

a. Cancer in the Cell.

These types of cancer are described using the cells in which the cancer cells are formed. In this type of cancer, cancer cells begin in specific cells.[44] Examples include:

i. Carcinoma: Carcinoma is a type of cancer which occurs in the epithelial cells – cells that cover the internal and external surface of the body. Carcinomas are the most prevalent type of cancer.[45]

ii. Sarcoma: Sarcoma is a type of cancer that occur in the bone, soft tissues such as muscle, tendons, fat, blood vessels, lymph vessels, nerves, and tissues around the joints.[46]

iii. Leukemia: Leukemia is a type of cancer that begins in the blood-forming tissue of the bone marrow. In leukemia, no tumor is formed. However, a significant volume of deformed cells accumulate in the blood and bone marrow, thereby overpowering the normal blood cells.[47]

iv. Lymphoma: Lymphoma is a type of cancer that emanates from the lymphocytes. In lymphocytes, there is a build-up of abnormal lymphocytes in the lymph nodes, lymph vessels and other parts of the body.[48]

v. Multiple Myeloma: Multiple myeloma is a type of cancer that begins in plasma cells. In multiple myeloma, abnormal plasma cells build up in the bone marrow and form cancer cells in bones all through the body.[49]

vi. Melanoma: Melanoma is a cancer type that originates in specialized cells that produce melanin – the pigment responsible for skin coloration.[50,51]

b. Cancer in the Organ

These types of cancer are described using the organs in which the cancer cells originate.[52] For example: lung cancer begins in the lungs, cervical cancer begins in the cervix, breast cancer, begins in the breast, prostate cancer begins in the prostate gland, ovarian cancer begins in the ovaries, colon cancer begins in the colon, pancreatic cancer begins in the pancreas and so on.

c. Prevalence of Cancer

The National Cancer Institute gives prevalence of cancer across two timelines with Table 1 showing different types of cancers, the estimate of new cases and deaths for the year 2016, [39] while Table 2 captures the statistics for the year 2019 [53]:

Table 1: Cancer Types and Prevalence in 2016[39]

	Types of Cancer	**Estimated Number of New Cases**	**Estimated Number of Deaths**
1	Bladder	76,960	16,390
2	Breast (Female -- Male)	246,660 -- 2,600	40,450 – 440
3	Colon and Rectal	134,490	49,190
4	Endometrial	60,050	10,470
5	Kidney	62,700	14,240
6	Leukemia	60,140	24,400
7	Lung	224,390	158,080

8	Melanoma	76,380	10,130
9	Non-Hodgkin Lymphoma	72,580	20,150
10	Pancreatic	53,070	41,780
11	Prostate	180,890	26,120
12	Thyroid	64,300	1,980

Table 2: Cancer Types and Prevalence in 2019[53]

	Types of Cancer	**Estimated Number of New Cases**	**Estimated Number of Deaths**
1	Bladder	80,470	17,670
2	Breast (Female -- Male)	268,600 – 2,670	41,760 – 500
3	Colon and Rectal	145,600	51,020
4	Endometrial	61,880	12,160
5	Kidney (renal cell and renal pelvis)	73,820	14,770
6	Leukemia (all types)	61,780	22,840
7	Lung (including bronchus)	228,150	142,670
8	Melanoma	96,480	7,230
9	Non-Hodgkin Lymphoma	74,200	19,970
10	Pancreatic	56,770	45,750
11	Prostate	174,650	31,620

12	Thyroid	52,070	2,170
13	Liver and Intrahepatic Bile Duct	42,030	31,780

d. Symptoms of Cancer

There are different symptoms of cancer. The most common include skin changes, breast changes, lump on or under the skin, changes in bowel habits, eating problems, weight gain or loss, abdominal pain, unexplained night sweats, unusual bleeding or discharge and weakness.[54,55] Oftentimes, these symptoms are not as a result of cancer. They may be caused by benign tumors or other illnesses. However, if the listed symptoms persist for few weeks, it is important to see a doctor for early diagnosis and treatment commencement.[56]

2.3 Diagnosis of Cancer

If by meticulous observation, symptoms suggest cancer, the first step in developing a cancer treatment plan is through a thorough and accurate cancer diagnosis.[57-59] Consequently, the following are needed for accurate diagnosis:

1. A review of health history
2. Physical examination
3. Laboratory tests
4. Biopsy
5. Imaging tests
6. Nuclear medicine scans
7. Endoscopy

8. Genetic tests.[60]

2.4 Management of Cancer

There are different types of cancer treatment. Each cancer treatment depends on the type of cancer, stage at diagnosis and how its advances.[61,62] Most cancer patients use combination of treatments to manage cancer such as surgery, radiotherapy, chemotherapy, immunotherapy, targeted therapy or hormone therapy.[63-65] However, cancer patients are sometimes subjected to only one form of treatment;

i. Surgery: Surgery is a procedure used to treat cancer in which a surgeon removes tumor from the body, debulk a tumor and ease cancer symptoms.[66,67]

ii. Chemotherapy: Chemotherapy is a type of cancer treatment that works by stopping the growth of cancer cells in the body. It is used to treat cancer, lessen the likelihood of cancer recurring and helps to ease cancer symptoms. It makes use of drugs to kill cancer cells.[68,69]

iii. Radiation therapy: Radiation therapy, otherwise referred to as radiotherapy is a type of cancer treatment whereby heavy doses of radiation are used to shrink tumors and kill cancer cells. It takes days or weeks for cancer cells to die.[70,71]

iv. Immunotherapy: Immunotherapy is a biological therapy which uses substances to treat cancer. The substance is gotten from living organisms. It is a type of cancer treatment in which the immune system helps to fight cancer.[72,73]

v. Hormone therapy: Hormone therapy is a type of cancer treatment which stops or slows the growth of cancer that makes use of hormones to grow.[74,75]

vi. Targeted therapy: Targeted therapy is a cancer treatment which attacks cancer cells that help them grow, divide, and spread all around the body. It changes or blocks the effects of the cancer cells in the body.[76,77]

2.5 Breast Cancer

"Breast cancer is a malignant tumor (a collection of cancer cells) arising from the cells of the breast which occurs in both men and women". Breast cancer is caused by genetic abnormality.[30,78,79]

2.6 Types of Breast Cancer

Types of breast cancer are broken into groups of common and non-common groups. Below are the common types of breast cancer:

i. Ductal Carcinoma in Situ: The Ductal Carcinoma in Situ (DCIS) is the most common type of non-invasive breast cancer. It is a type of cancer that does not spread and usually has a very high cure rate.[80,81]

ii. Invasive Ductal Carcinoma: Invasive ductal carcinoma is a type of breast cancer that starts in a breast duct and grows into the tissue that surrounds it. It makes up approximately 80% of invasive breast cancers and is the most common type of breast cancer.[82,79]

iii. Invasive Lobular Carcinoma: Invasive lobular carcinoma is a type of breast cancer which begins in the mammary glands (milk-producing glands of the breast). It makes up about 10% of invasive breast cancers.[50,79]

Below are the much less common breast cancers:

i. Medullary Carcinoma: The medullary carcinoma is an infiltrating breast cancer that presents well-defined boundaries between the cancerous tissues and noncancerous tissue.[50,56,83]

ii. Mucinous Carcinoma: Mucinous carcinoma is a type of breast cancer which is formed from mucus-producing cancer cells.[50,56,84]

iii. Triple-Negative Breast Cancer: This type of breast cancer is invasive. The TNBC cells lack estrogen, progesterone and HER2 (excess of a specific protein) receptors respectively, on their surface. TNBC is notorious for mostly occurring in younger and African-American women, respectively.[50,56,85,86]

iv. Inflammatory Breast Cancer: Inflammatory breast cancer is a type of breast cancer which makes the skin of the breast reddish and feel warm. It makes the skin of the breast look like an infection. This occurs as a result of cancer cells blocking the lymph vessels.[50,56,87]

v. Paget's Disease of the Nipple: Paget's disease of the nipple is a type of breast cancer which begins in the breast ducts and moves to the nipple and the area that surrounds the nipple. The Paget’s disease usually presents with redness and crusting around the nipple.[50,56,88]

vi. Adenoid Cystic Carcinoma: Adenoid cystic carcinoma is a type of breast cancer which has both the glandular and cystic features. It has good prognosis and does not spread aggressively.[50,56,89]

vii. Lobular Carcinoma in Situ: Lobular carcinoma in situ is an area of abnormal cell growth found in the breast which can lead to invasive breast cancer later in life, but it is not a cancer.[50,56,90]

2.7 Symptoms of Breast Cancer

Breast cancer symptoms varies from person to person.[91] Below are the common symptoms of breast cancer:

1. Skin change
2. Redness
3. Swelling
4. Visible difference in one or both breasts
5. Increase in the size of the breast(s)
6. Change in shape of the breast(s)
7. Changes in the appearance of the nipple(s)
8. Nipple discharge
9. General pain in the breast
10. Lumps in the breast(s).[60]

2.8 Basal or Triple Negative Breast Cancer (TNBC)

Triple Negative Breast Cancer "is a type of breast cancer in which the offending tumor is estrogen receptor-negative, progesterone receptor-negative and Human epidermal growth factor receptor 2(HER-2)-negative". The negativity of these three receptors gave rise to the name "Triple Negative". TNBC is more aggressive, compared to other types of breast cancer and it has a probability of spreading beyond the breast as well as coming back within three years after treatment. TBNC does not respond to some medications that would ordinarily work for other types of cancers.[16,92]

With breast cancer responsible for a high mortality rate among women globally, records show that TNBC is responsible for 15 – 20% of new cases. More so, the six sub-types of TNBC namely; Basal-Like 1 (BL1), Basal-Like 2 (BL2), Mesenchymal (M), Mesenchymal Stem–Like (MSL), Immunomodulatory (IM), and Luminal Androgen Receptor (LAR), are said to exhibit similar patterns in their gene expressions, predicated by the fact that they are all negative to the estrogen, progesterone, and human epidermal growth factor receptors.[93-95]

Irrespective of these similarities inherent in all six subtypes, there is enough documentation suggesting that TNBC is very heterogeneous. Its heterogeneity can be classified in terms of molecular, histopathologic and clinical profiles.[96] Interestingly, a large proportion of TNBC patients are known to be quite young, typically below the age of 40 and from a racial perspective, African-Americans and Blacks form a dominant percentage of patients. In addition, TNBC patients have a comparatively shorter progression-free and overall survival rate than patients of other breast cancer types.[97-100]

Furthermore, when compared to other tumors with hormone-positive receptors, TNBC patients have a higher rate of relapsing and death which is quite congruent to the fact that the disease is also known to be more aggressive. The ability of the disease to proceed in a gradual, subtle way, but with very harmful effects is due to the high frequency of rapidly proliferating grade 3 tumors at diagnosis.[97,98,101,102] More so, between 1 to 3 years after diagnosis and treatment, TNBC patients are likely to suffer recurrence, while majority of deaths will occur within the first 5 years during which therapy is being administered.[100]

Considering the poor prognosis associated with TNBC, it is therefore imperative that efforts to discover relevant markers that can be used to predict response to treatment is of profound medical importance.

2.9 Clinical Features of TNBC

Triple-Negative Breast Cancer (TNBC) has unique features which differentiates it from others.[103] These features include:

i. The Receptor Status: The receptor status helps to identify TNBC, tests that detect estrogen-receptors, progesterone-receptors and HER-2 will not be positive, which means that the traditional treatment for breast cancer would not be effective.[104]

ii. More Aggressive: There is a great possibility that TNBC would spread and recur after treatment compared to other types of breast cancer. After the first five years of following therapy, the risk of it spreading and recurring decreases.[105]

iii. Cell type and grade: In TNBC, cancer cells are mostly 'basal-like'. They look like basal cells that line the breast ducts. These cells do not resemble the normal healthy cells.[106]

Other clinical features include:

iv. Lump in the breast

v. Pain in the breast

vi. Redness of the breast

vii. Nipple turns inward

viii. Discharge from the nipple.

2.10 Diagnosis of TNBC

When doctors find an area in the breast that is abnormal, biopsy is carried out – a little part of the tissue would be cut out in order to test the cells. Then a pathologist looks at the tissue with the use of a microscope. If the tissue has cancer cells, a test would be conducted. The cell is triple negative, if the test is negative for estrogen, progesterone, and HER-2 receptors, respectively.[16,107]

2.11 Treatment/Management of TNBC

Triple-Negative Breast Cancer is treated with a combination of therapies which includes surgery, radiotherapy, and chemotherapy. Hormonal therapy is not likely to treat TNBC. Hormone-receptor negative breast cancers respond better to chemotherapy. Therefore, TNBC would respond to chemotherapy. There is no recommendation that more treatment should be given to patients with TNBC.[108,109]

2.12 Overview of Precision Medicine

Precision medicine is a dynamic field which integrates the latest research advances into individual treatment by the most tailored diagnostic tools. Such treatment for patients with any form of disease, especially chronic diseases, such as cardiac disease, diabetes and cancer can proof to be more effective with minimal to no side effects.[110,111] Precision medicine can even be used to manage and maintain healthy people. In addition, it also entails targeted therapy against mutations found in certain cancers as well as gene

therapy.[112,113] Precision medicine drills down to know as much about the patient as possible from understanding the genomic and proteomic disposition of the patient with respect to the disease of interest. Interestingly, precision medicine can even investigate the social media patterns of a person to facilitate early detection of disease as well as provide answers as to how well a patient would respond to a treatment method.[114-116]

Precision medicine enables the ultimate delivery of personalized medicine, which means that when a patient goes to the doctor, the doctor can have access to more genetic information which the doctor can use to specifically design treatment care plans for such a patient. It is important to note also, that there will be thousands of individuals that have similar disease and genetic information with such a patient, the patient's adopted care plan can then be used to fast track treatment for a good number of other people in the same situation. In addition, with precision medicine, experts can look at individual tumors, individual cells within tumors and gain profound understanding of the molecular dynamics of the disease. It becomes hugely possible to analyze this very complex data using state of the art computer algorithms to observe the interaction of proteins in a cancer cell and consequently, to design drugs that inhibit or treat them.[117,118]

Furthermore, precision medicine is sparking an excitement that brings with it an entirely new era of medicine that makes sure new jobs that didn't exist before are now evolving with new markets and industries.[119] While it would be difficult to forecast exactly what the financial implications of precision medicine are going to be on the one hand, if drugs are tailored to very specific populations, the prices can be higher because the value to the patient is also higher on the other hand, because in normal medical practice today, many people are taking drugs that are really not doing them any good. So, we can avoid

the unnecessary costs of inappropriate medication that would be cost saving. Precision medicine is often referred to as personalized medicine.[120]

2.13 Personalized Medicine in Cancer

This largely entails taking what was learned from decoding the human genome and translating it into life changing outcomes for cancer patients using the so called "next generation sequencing" to help doctors in making personalized medicine a reality for their cancer patients.[121] The human body is made up of more than 30 trillion cells and each of these cells contains deoxyribonucleic acid (DNA), our genetic code. When cancer occurs in our cells, DNA has gone wrong. Genetic change takes more than just one mutation in a gene to result in cancer and these changes can be inherited or more commonly acquired after we are born.[122] As a matter of fact, 85% of all cancers happen as a result of changes that occur over the course of our lives by exposure to things like cigarette smoke, certain toxic chemicals or ultraviolet (UV) light from the sun. As we age our cell's DNA gets more and more of these changes which explains why it may take many years for a cancer to develop. Cancer occurs when certain combinations of these changes cause cells to lose their normal function and as a result, they grow and divide uncontrollably and together they form what is called a tumor.[123,124]

Imagine if we knew how to read the DNA sequence of the tumor, we can then see whether genetic changes are what they look like and having such information would mean doctors could provide a precise personalized management plan for patients with cancer, perhaps even targeting some of the genetic changes with specific drugs. This depth of genetic profiling provides the unique genetic information for specific cancer at incredibly high resolution and in many cases provides the ability to identify specific changes that

played a role in the development of that tumor. This is a future game changer for cancer patients with the prospects of using drugs that specifically target their tumor cells.

2.14 Machine Learning

As humans, we learn from our past experiences. In so doing, we execute future actions based on what we learnt by virtue of our successes or failures from previous experiences. It is the ability to get machines replicate this capacity that has given rise to the field of machine learning. Machine learning is a subset of Artificial Intelligence (AI), dedicated to loosely programming machines so as not to explicitly carry out tasks in a predetermined manner, but to through experience, determine how best to respond to problem solving in an optimized way.[125]

The value of machine learning is only just beginning to show itself, especially due to the explosion of data within the last two decades. This data is generated not only by people but also by computers, phones, and other devices. Traditionally, humans have the natural capacity to analyze data and adapt decision making on this basis, even when there are changes in data patterns. However, the volume of data evident today, surpasses the natural ability of humans to make sense of these intricately complex patterns without the help of automated systems that can learn from the data with profound speed and precision.[126]

Take Google Search for example, every time you use the search platform, you are using the machine learning systems at its core from understanding the text query, to adjusting the results based on your personal interests such as knowing which results to show you first using a ranking system. Today, machine learning applications are widely

used, ranging from image recognition, fraud detection, recommender systems, text, and speech systems. With these powerful capabilities, it can be applied to a wide range of fields such as precision medicine, retail marketing, weather forecasting, epidemiology, and transportation to mention but a few. As a matter of fact, machine learning is at the heart of Artificial Intelligence in general. Machine learning is simply the use of data to answer important questions. Where the questioning can be likened to using aspects of the data to create a training set with which the model will be built, whereas the answers can be likened to the ability of the model to use insights gained from the trained set to give accurate answers pertaining to the data that wasn't used in the training.[127]

Usually, machine learning goes through eight steps: data gathering, data preparation, choosing a model, training, evaluation, hyperparameter tuning, prediction, and validation.[128] In addition, the machine learning algorithms can be linear, non-linear or ensemble.[129,130]

2.15 Machine Learning as a Driver for Precision Medicine

Machine Learning plays a pivotal role in precision medicine because as an expert, you want all relevant information about the patient to enable you make a more precise diagnosis and also the best treatment decision.[131] Important questions would seek to know for example in breast cancer treatment, the kind of breast cancer, the patients age at diagnosis, family history, presence of comorbidities such as diabetes or cardiomyopathy and then you want to know what the imaging results have revealed. In addition, you want to know if an MRI was done because there was a distant metastasis and you also want to know if there was genetic testing. Harnessing all the relevant information about the patient helps to create quality insights about the disease and possible treatment plan. More so, from

the combined information, you also want to extract more information through AI, radiomics, and bringing all that together to enable the expert team to personalize the treatment suitable for the patient.[132]

CHAPTER 3

METHODS

3.1 Subjects

Data was sourced from the publicly available Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset. The METABRIC dataset was collected by the curators between 2012 and 2016.[133] In view of this research, effort was made to get as much relevant data which was primarily derived from 2509 patient samples. The data was accessed through cBioPortal for Cancer Genomics. The public cBioPortal site is hosted by the Center for Molecular Oncology at Memorial Sloan Kettering Cancer Center (MSK).[134] Table 3 shows that as a result of filtering operations to derive necessary variables key to this research in terms of the cancer subtype of interest and the availability of treatment and survival information, the number of samples was pruned from 2509 to 320 homogenous samples that consequently serves as the population used for the research analyses.

Table 3: Dataset Source and Breakdown

Dataset Source	Overall Samples	TNBC Samples only	Patients Treated with Radiation after surgery	% of TNBC Patients with surgery and Radiation Treatments	Surgery + Radiation : Living	Surgery + Radiation : Died of Disease	Surgery + Radiation : Died of Other Causes
META BRIC	2509	320	220	68.75%	107	83	30

To attain analytical efficiency, the dataset described in Table 3 was further filtered to produce the classifications captured in Table 4. The dataset sub-classes are lettered A – G. The class E dataset was predominantly used for computational analysis in this research. However, classes F and G were also used in a considerable number of instances.

Table 4: **Filtered Dataset**

Dataset Classification	**Dataset Description**	**Operation on dataset**	**Sample size**	**Percentage of Samples**
Class A	Overall Samples from METABRIC	2509	2509	100% of overall samples
Class B	TNBC Samples Only	320	320	12.75% of overall samples
Class C	Class B minus patients who did not undergo surgery	320 – 9	311	97.18% of all TNBC samples
Class D	*Class C minus patients who died from other causes*	*311 – 48*	*263*	*82.18% of all TNBC samples*
Class E	*Class D minus patients who did not undergo radiotherapy (with the inclusion of data for 28 genes)*	*263 – 73*	***190***	*59.37% of all TNBC samples*

Class F	*Patients with follow-up information of not more than 5 years (60 months)*	*190 - 120*	***70***	*21.87% of all TNBC samples*
Class G	*Patients with follow-up information of more than 5 years (60 months and above)*	*190 - 70*	***120***	*37.50% of all TNBC samples*

Table 5 is a representation of the datatypes captured from the dataset. The datatypes captured cut across phenotypic, mutation, survival/follow up, clinical and patient biodata.

Table 5: Dataset Datatypes

	DATA TYPE	METABRIC (30 Variables)
1	**Phenotypic**	Yes
2	**Mutation**	Yes (as mutation counts)
3	**Treatment**	Yes (surgery, radiotherapy, chemotherapy & hormone therapy)
4	**Survival/Follow up**	Yes
5	**Hormone Receptors**	Yes
6	**Clinical**	Yes
7	**Patient Bio**	Yes
8	**Racial/Ethnicity**	No
9	**Gene Expression**	No

3.2 Instrument

The research instrument for data collection was cBioPortal for Cancer Genomics. The cBioPortal for Cancer Genomics was originally developed at Memorial Sloan Kettering Cancer Center (MSK). The public cBioPortal site is hosted by the Center for Molecular Oncology at MSK and available at www.cbioportal.org. cBioPortal was the instrument through which both clinical and genetic data was sourced. The summarized clinical and genetic variables of interest are as follows:

i. The 24 clinical variables of interest that were captured from clinical data for all patients who underwent radiotherapy treatment are: AGE_AT_DIAGNOSIS,

AGE_GROUP, BREAST_SURGERY, CANCER_TYPE, CELLULARITY, CHEMOTHERAPY, PAM50 (based on PAM50 classification), COHORT, ER_IHC (ER status measured by Immunohistochemistry), TUMOR_GRADE, HER2_SNP6 (HER status measured by Single Nucleotide Polymorphism 6), HISTOLOGICAL_SUBTYPE, HORMONE_THERAPY, MENOPAUSE, INTCLUST, LATERALITY, MUTATION_COUNT, NPI (Nottingham Prognostic Index), SURVIVAL_MONTHS, SURVIVAL_GROUPS, THREEGENE, TUMOR_SIZE, TUMOR_STAGE, and VITAL_STATUS.

ii. The 18 genes of interest for which mutation data was sourced are as follows: AGMO, AKT1, BIRC6, CTNNA1, DNAH11, FOXO3, GATA3, LDLRAP1, MAP2K4, MAP3K1, MEN1, NCOR2, NF1, PIK3CA, SHANK2, SYNE1, TAF1 and TP53.

3.2.1. Data Transformation

The data was transformed to make it suitable for statistical analyses and machine learning operations was a binary classification problem.[135] The gene values were either 0 or 1. 0 means no mutation, while 1 means mutation occurred in the referred gene. Also, for Patient VITAL_STATUS which was the "Y" – dependent variable, the values 0 or 1 was used. 0 was used to represent "Living", while 1 was used to represent "Died of disease". All other variables (multivariate independent variables) that are non-numeric, were at some point, converted to the numeric/integer datatype. Other non-numeric variables that were assigned numeric values are as follows:

i. AGE_GROUP: Below 40 = 1; 40-52 = 2; Above 52 = 3.

ii. BREAST_SURGERY: Mastectomy = 1; Breast Conserving = 2.

iii. CANCER_TYPE: Invasive Breast Carcinoma = 1; Breast Invasive Ductal Carcinoma = 2; Breast Mixed Ductal and Lobular Carcinoma = 3; Breast Invasive Lobular Carcinoma = 4.

iv. CELLULARITY: Yes = 1; No = 2.

v. CHEMOTHERAPY: Yes = 1; No = 2.

vi. PAM50: Basal = 1; Claudin-low = 2; Her2 = 3; LumA = 4; Normal = 5.

vii. ER_IHC: Positive = 1; Negative = 2.

viii. HER2_SNP6: Gain = 1; Neutral = 2; Loss = 3.

ix. HISTOLOGICAL_SUBTYPE: DCIS = 1; IDC = 2; IDC-MED = 3; IDC-TUB = 4; IDC+ILC = 5; ILC = 6; INVASIVE TUMOR = 7; OTHER = 8.

x. HORMONE_THERAPY: Yes = 1; No = 2.

xi. MENOPAUSE: Pre = 1; Post = 2.

xii. INTCLUST: 1 = 1; 2 = 2; 3 = 3; 5 = 5; 7 = 7; 9 = 9; 10 = 10; 4ER- = 4; 4ER+ = 6.

xiii. LATERALITY: L = 1; R = 2.

xiv. THREEGENE: ER-/HER2 = 1; ER+/HER2- High Prolif = 2; ER+/HER2- Low Prolif = 3; HER2+ = 4.

xv. SURVIVAL_GROUPS: Below 5years = 1; Above 5years = 2.

3.3 Procedures

The procedures adopted in this research are discussed as methods under the following subsections:

3.3.1. Recap of Objective 1

Establish the effects of adjuvant radiotherapy on post-surgery TNBC patients, both alone and in combination with other TNBC therapeutic modalities.

3.3.2. Method 1 – Exploratory Data Analyses

Exploratory data analysis was carried out on R Studio 3.6 to have a good understanding of the interactions of the various variables in the dataset. This was done using the following packages:

i. Dplyr: this offers sufficient grammar for data manipulation. This was used to manipulate the dataset such as variable renaming and other flexible data transformation operations required during the analysis.

ii. Ggplot2: this is a popular visualization package for making plots.

iii. Mice: this stands for multivariate imputation by chained equations. It was used to check for missing values (NA's). Afterwards, the missing values were assigned with predicted values. In our analysis, mice makes use of the Predictive Mean Matching (PMM) method for integers/numeric variable values, logistic regression was used for categorical variables with two outcomes while polynomial regression was used for categorical variables with more than two outcomes.

iv. DataExplorer: was used to automate visual exploration of data and treatment.

v. Psych: used fundamentally for multivariate analysis and scale construction. To achieve this, the package makes use of functions pertaining to factor analysis, principal component analysis, cluster analysis and reliability analysis. It was used to find the correlations between predictor variables. By default, corr.test()

function in the Psych package produces pairwise correlation matrix for the entire dataset making use of the "Pearson's" correlation method.

vi. Furthermore, the cor.test () function was used to check the significance level of some correlations. This operation makes use of the base R cor() function which provides a simple way to get Pearson correlations between two variables, thus showing its significance at p-value = 0.05.

Notable base functions that were used to enhance data understanding, manipulation and visualization include boxplot and hist (for plotting histograms). Other data manipulation and exploration tasks executed using base functions included creating the Classes F and G sub-datasets respectively for targeted analysis affecting two groups of patients defined by the length of follow-up they had.

3.3.3. Method 2 – Survival Analyses

Analysis was carried out using the R function; summary(object,60). 60 represents 60 months which is equivalent to 5 years. Medical literatures and research suggest that 1 – 5 years after diagnosis is the period of significance during which there is a high possibility for treatment failure, relapse and death in TNBC patients.[136] Therefore, this research deems it imperative to use the 5-year timeline as a benchmark to test treatment effectiveness. The class E, F and G datasets with 190, 70 and 120 samples respectively were used for this analysis.

With emphasis being on the role of adjuvant radiotherapy on TNBC patients, Kaplan Meier Survival Analysis was carried out using the ggfortify

package and the survival package. The "survdiff" function from the survival package was used to carry out a comparative analysis of the P-values of the statistical significance of the survival analysis. The statistical tests used were Log-Rank (rho = 0) and Generalized Wilcoxon (rho = 1), while the Cox Proportional Hazard ratio was used to ascertain the likelihood of an event to take place while a patient is on a treatment type. The analysis was comparatively carried out on TNBC patients who undertook adjuvant radiotherapy after surgery and compared with those who did not undergo radiotherapy at all. The class E, F and G datasets with 190, 70 and 120 samples respectively were used for this analysis.

3.3.4. Recap of Objective 2

Correlate genetic and transcriptomic alterations identified in these patient's tumors with the effects established in Objective 1.

3.3.5. Method 3 – Genetic (Mutation) Analyses

To ascertain the genes that are statistically relevant, Fisher's Exact Test was used with a P-Value of 0.05.[137] The Fisher's Exact Test was used to test for associations between the categorical variable, patient's vital status and gene mutation status as well as gene amplification status. The analysis captured Copy Number Alterations (CNA), gene mutation and further analysis based on PAM50 Classification. A total of 18 statistically significant genes were derived from the gene mutation and PAM50 processes.

In addition, analysis was carried out on 190 TNBC samples (107 Living and 83 Died of disease) and excludes all patient samples that pertains to patients who

died of other causes. The samples represented in the 190-patient group were all administered radiotherapy after they had already undergone either mastectomy or breast conserving surgical procedures, respectively. Furthermore, the analysis was carried out using Onco-Query Language (OQL), on the cBioPortal web platform for Cancer Genomics. The 190 TNBC patient samples all tested negative to Estrogen, Progesterone and Human Epidermal (Growth Factor)-2 Receptors, respectively. The class E dataset with 190 samples was used for this analysis.

3.3.6. Recap of Objective 3

Utilize the relationships found in Objective 2 and relevant clinical features to build a predictive model for treatment response in patients with TNBC.

3.3.7. Method 4 – Machine Learning Analyses

This analysis was carried out on the "Class E" dataset which has a total of 190 patient samples. The dataset had 42 variables, 18 genes of interest and 24 clinical features. More so, this involved the building of three predictive models after comparatively analyzing the performance of 5 classification and ensemble learning models. The three predictive models were built using:

i. Clinical features only

ii. Genetic (mutation) features only

iii. Combination of clinical and genetic features

The learning model with the best accuracy and ROC curve statistic was adopted as the model of choice. The following algorithms were explored:

i. Linear Algorithms: Logistic Regression (LG) and Linear Discriminant Analysis (LDA).[138]

ii. Non-Linear Algorithms: K-Nearest Neighbors (KNN) and Classification and Regression Trees (CART).[139]

iii. Ensemble Method (Bagging): Random Forest (RF).[140]

Firstly, feature selection using the Boruta package is carried out to prune down the number of predictor variables by eliminating unimportant variables. Correlations operations had earlier suggested that some variables showed no significant relationship with the outcome. Feature selection ensures that only the important variables are used to build all three models. The workings of the Boruta function are outlined below:

i. Firstly, it introduces randomness into the data set by producing shuffled copies of all the features of interest (these are referred to as shadow features).

ii. Next, it trains a random forest classifier on the extended data set and applies a feature importance measure (the default is Mean Decrease Accuracy) to ascertain the relevance of each of the features. Those with higher means are classified as more important.

iii. Afterwards, at every given iteration, it checks whether a real feature has a higher importance than the best of its shadow features (in other words, it checks to see if the feature has a higher Z-score than the maximum Z-score of its shadow features) and therefore eliminates features which are termed highly irrelevant.

iv. Lastly, the Boruta algorithm stops either when all features gets confirmed or rejected or it when reaches a specified limit of random forest runs.[141]

After the feature selection operation has been finalized, the algorithms were evaluated using baseline parameters e.g. 10-fold cross validation with 3 repeats and for simplicity, the performance metrics used were "Accuracy and Cohen's Kappa". Subsequently, the algorithms were further evaluated using "transforms". This was due to the presence of skewed distributions in the dataset. However, the transform operation using Box-Cox was ignored because it failed to improve the accuracy of the models. After this phase, the algorithm with promise was further tuned for optimal performance. Subsequently, the final predictive model was built followed by a validation activity to test performance of the model. Afterwards, the confusion matrix was executed to derive statistically important parameters such as accuracy, specificity, sensitivity, and kappa. Finally, the ROC (Receiver Operating Characteristic) curve was then plotted and the AUC (Area Under the ROC Curve) was derived. The class E dataset with 190 samples was used for this analysis as shown in Table 6. 70% of the samples (134 samples) was used to train the model, while 30% of the samples (56 samples) unused for the training, was used to validate/test the model.

Lastly, a scatterplot was used to try to understand the relationship between the final predictor variables and the outcome. This was also used to demonstrate the heterogeneous nature of TNBC and cancers in general. The x-parameter was used to represent the multi-variate predictors after the pruning operation effected

using the Boruta feature selection function. On the other hand, the y-parameter was the outcome variable designated as VITAL_STATUS.

Table 6: Dataset used to Build Machine Learning Model

Datasets Source	**Class E Dataset**	**Radiation: Living**	**Radiation: Died of Disease**
METABRIC	190	107 (56.3%)	83 (43.7%)
Training Set =	**134**	**70%**	
Testing/Validation Set =	**56**	**30%**	
TOTAL	**190**	**100%**	

Other R packages that were used include

i. Mlbench: Machine Learning Benchmark Problems – is used for machine learning related operations.

ii. Caret: Classification and Regression Training – for doing classification and regression tasks in machine learning.

iii. PRROC: Precision-Recall and ROC Curves for Weighted and Unweighted data

3.4. Data Analysis

The data was collected from cBioPortal web platform. Methods 1, 2 and 4 was analyzed using R-Studio Version 3.6.0 (2019-04-26), while method 3 was analyzed using Onco-Query Language (OQL) on the cBioPortal web platform.

CHAPTER 4

RESULTS AND FINDINGS

4.1. Results from Exploratory Data Analyses

Cleaning, treating and transformation of the clinical dataset containing the 24 clinical variables described for analysis and mentioned in section 3.3.2. Among the 24 variables are two variables created to aid analyses. They are the AGE_GROUP and SURVIVAL_GROUPS variables which were created from the AGE_AT_DIAGNOSIS and SURVIVAL_MONTHS variables, respectively.

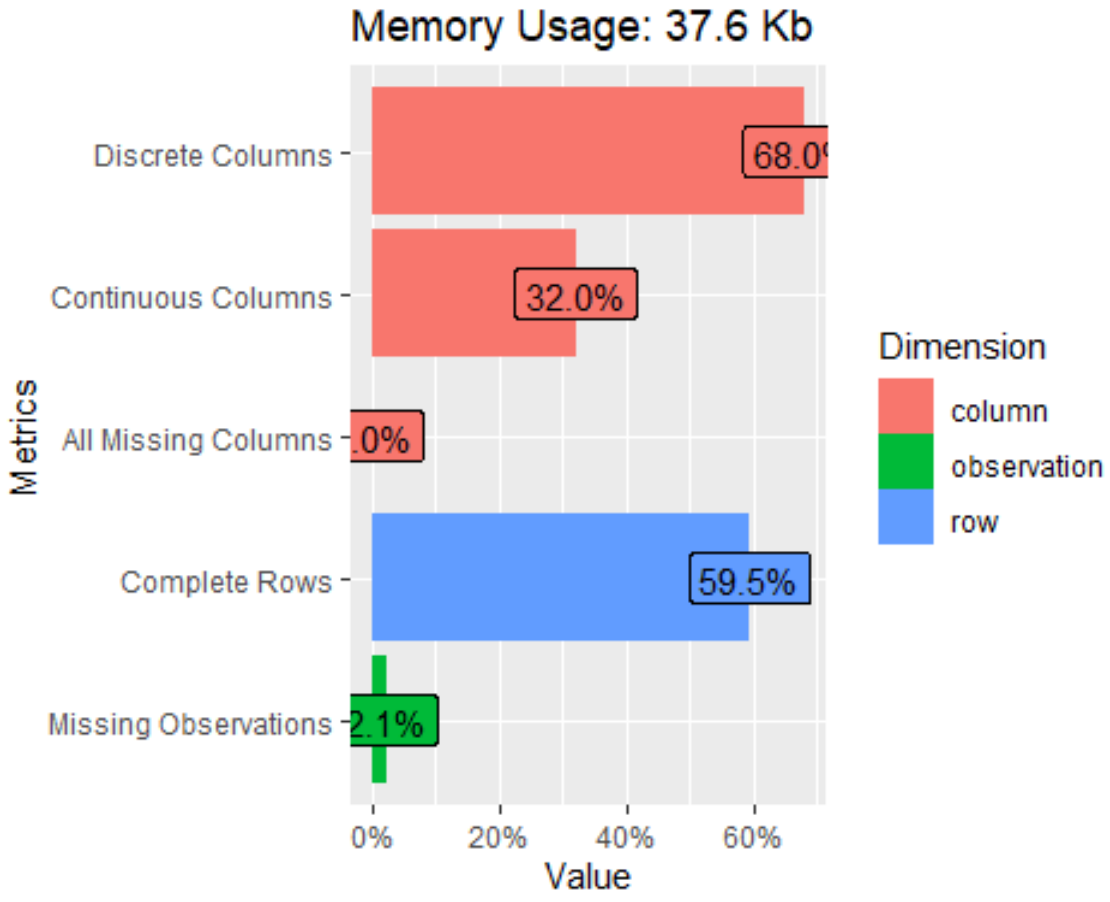


Figure 1 Plot showing status of all observations in the 190-Patient dataset

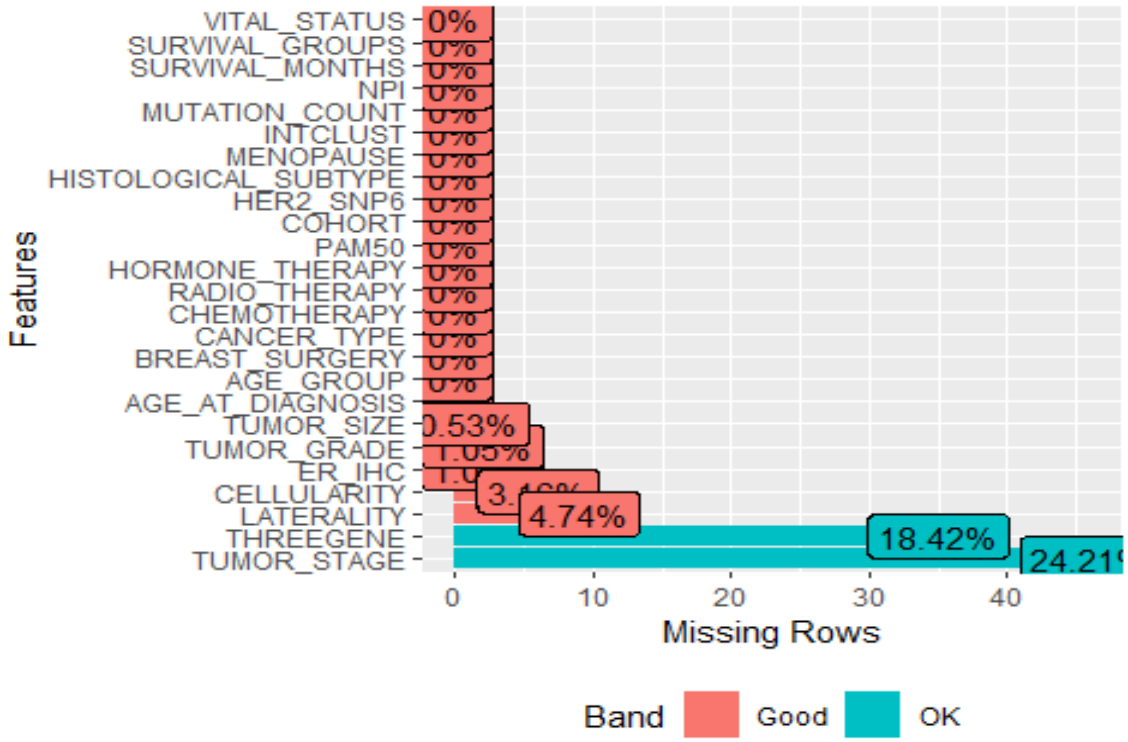


Figure 2 Plot showing dataset variables and percentage of missing values

Figure 1 shows that of the 190 observations, 113(59.5%) had complete records while 77 observations had 1 or more missing value(s). In all, 101 entries (2.1%) were missing out of 4750 entries (25columns x 190observations) across all 190 observations. Figure 2 shows that the affected variables with missing values are: CELLULARITY (6), ER_IHC (2), TUMOR_GRADE (2), LATERALITY (9), THREEGENE (35), TUMOR_SIZE (1) AND TUMOR_STAGE (46). At 46, TUMOR_STAGE had the highest number of missing values. With the use of the PMM method offered by the mice package, these missing values were computationally filled up. Predictive Mean Matching (PMM) method was applied to the following numeric variables: TUMOR_GRADE, TUMOR_SIZE, and TUMOR_STAGE. While polynomial regression was applied to the CELLULARITY and THREEGENE variables because they are categorical variables with

more than two values each. Logistic regression was applied to the ER_IHC and LATERALITY variables because they are categorical variables with two values each.

In addition, the median and mean ages of the patients at the point of diagnosis, were 52.125 years and 52.44 years, respectively. The median age for breast cancer diagnosis for white women is 62 and 59 for black women.[142] It therefore implies from this study, that there is a prevalence of TNBC occurrence in younger women relative to the specified median by approximately 10 years difference.

More so, in the 70-patient subgroup, defined as those who were tracked for exactly 5years or less, 64 patients died of the disease during the period under review which represents 91.43% mortality rate in this patient group. On the other hand, only 6 of the patients survived up till 5years, representing just 8.57% of the same patient group (see figure 3). For this group, correlation analysis indicated that there exists correlation of significant levels between VITAL_STATUS and the following 4 variables SHANK2, CANCER_TYPE, HER2_SNP6 and SURVIVAL_MONTHS respectively, tested for the 70-patient sub-group at $p = 0.05$.

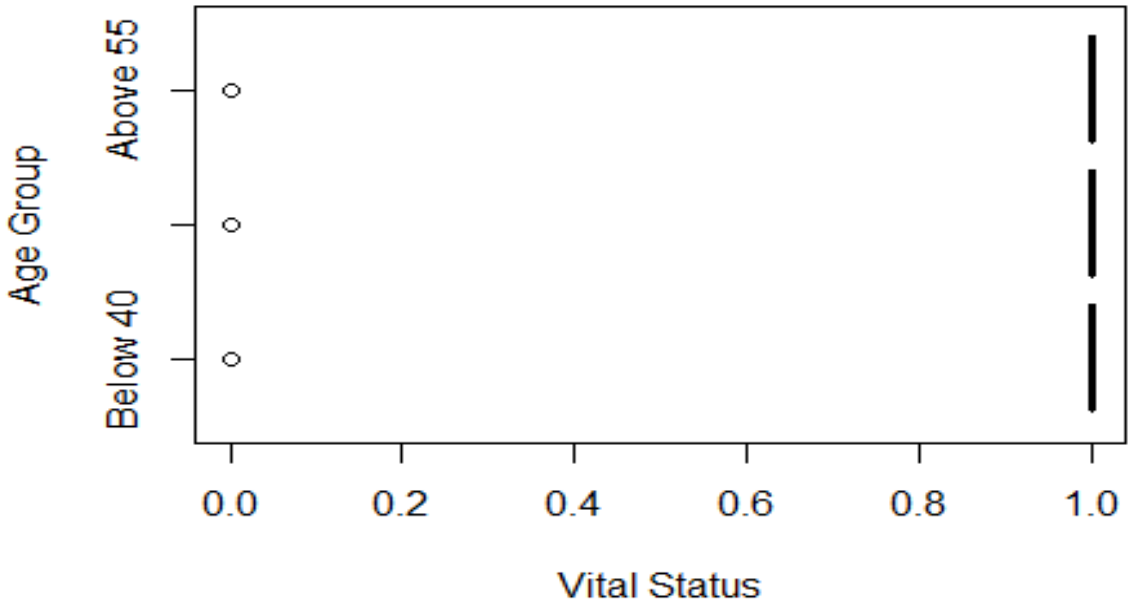


Figure 3 Boxplot showing relationship between Vital Status vs Age group for 70-Patient dataset

Conversely, in another group of 120 patients who were tracked beyond 5years, only 19 patients died of the disease during this extended period of follow-up, which represents a drastic reduction in the mortality rate to 15.83% of the patients. On the other hand, 101 of the patients in this second group survived, representing an impressive 84.17% survival rate of the patients in this group (see figure 4). For this group, correlation analysis indicated that there exists correlation of significant levels between VITAL_STATUS and the following 6 variables AKT, MAP3K1, MEN1, SURVIVAL_MONTHS, PAM50 and INTCLUST respectively, tested for the 120-patient sub-group at p = 0.05.

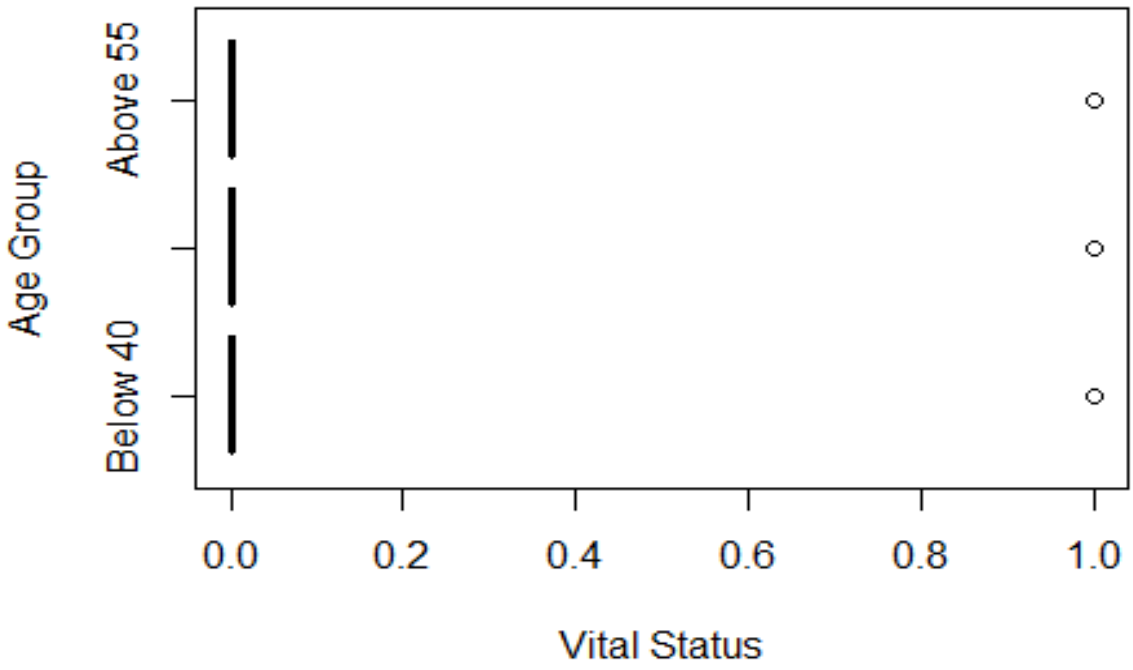


Figure 4 Boxplot showing relationship between Vital Status vs Age group for 120-Patient dataset

Furthermore, a check to ascertain the significance (p-values at 0.05) of the correlation tests carried out on the overall 190-patient group tracked beyond 5 years, showed that there exists correlation of significant levels between VITAL_STATUS and the following 10 variables AKT, MAP3K1, MEN1, SHANK2, NCOR2, SURVIVAL_MONTHS, SURVIVAL_GROUPS, NPI, BREAST_SURGERY and TUMOR_STAGE respectively, tested for the 120-patient sub-group at p = 0.05.

These correlation findings suggest that not all variables will be important as predictor features as this study intends to identify features of importance that will be used to build the predictive model that helps to ascertain TNBC patients who will benefit from post-surgery adjuvant radiotherapy.

Other findings include:

4.2. Results from Survival Analyses

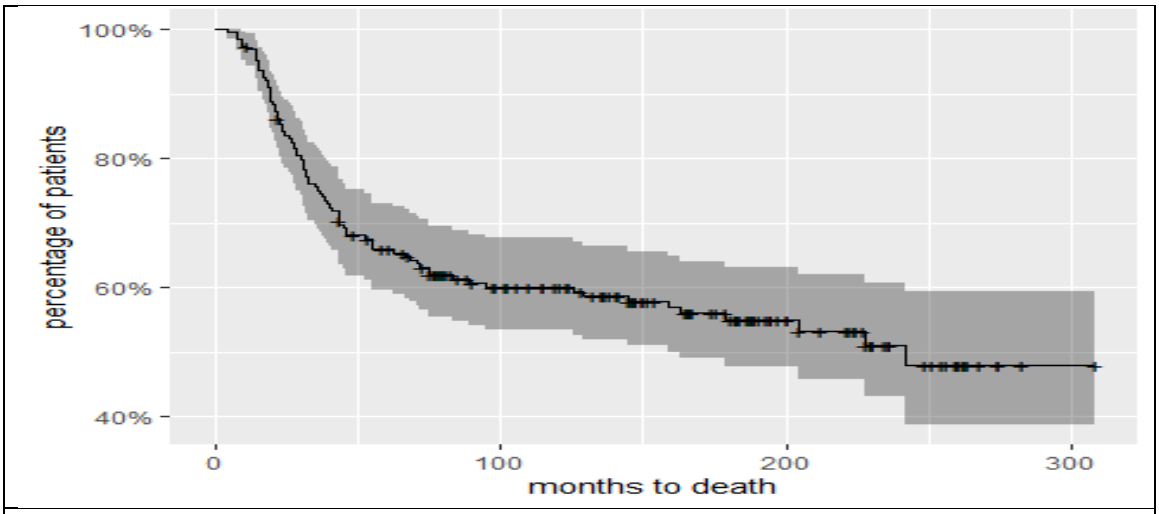
Survival Analysis with RStudio: Kaplan Meier Analysis for TNBC-METABRIC datasets of 190 samples, 70 samples and 120 samples respectively (Code segments are in the Appendices Section).

4.2.1. Survival Analysis on the Entire "Classes E, F and G" TNBC-METABRIC Datasets

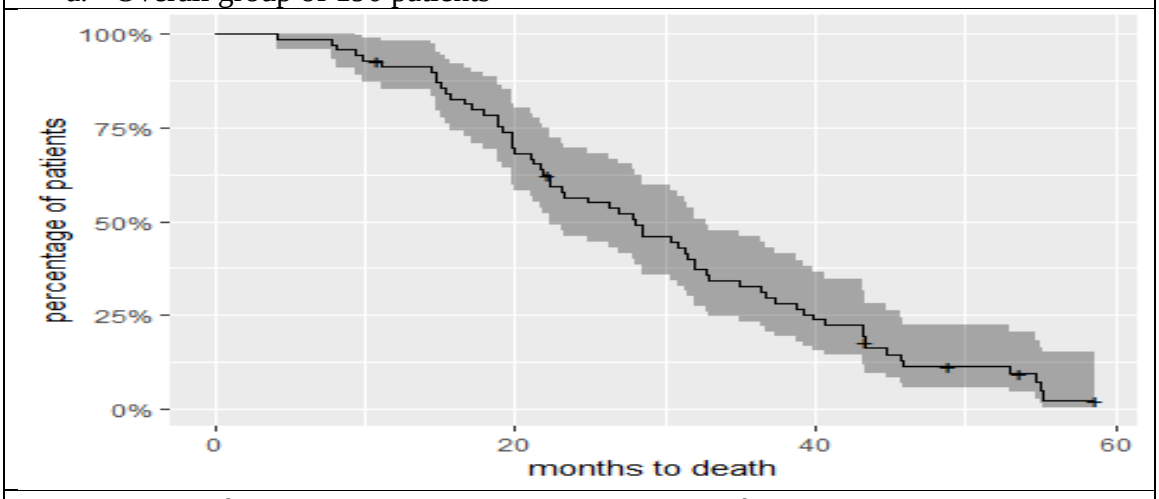
Code segment 2.8 (see Appendix) captures the survival analysis and output segment for the "Class E" TNBC-METABRIC dataset that contains 190 patient samples with treatment and follow-up information. The median survival time was 242 months which is approximately 20 years with 83 events (deaths). See Figure 5a.

Code segment 2.9 (see Appendix) captures the survival analysis and output segment for the "Class F" TNBC-METABRIC dataset that contains 70 patient samples with treatment and follow-up information of not more than 5 years. The median survival time was 28 months which is approximately 2.3 years with 64 events (deaths). Of this group, 7 patients were censored. See Figure 5b.

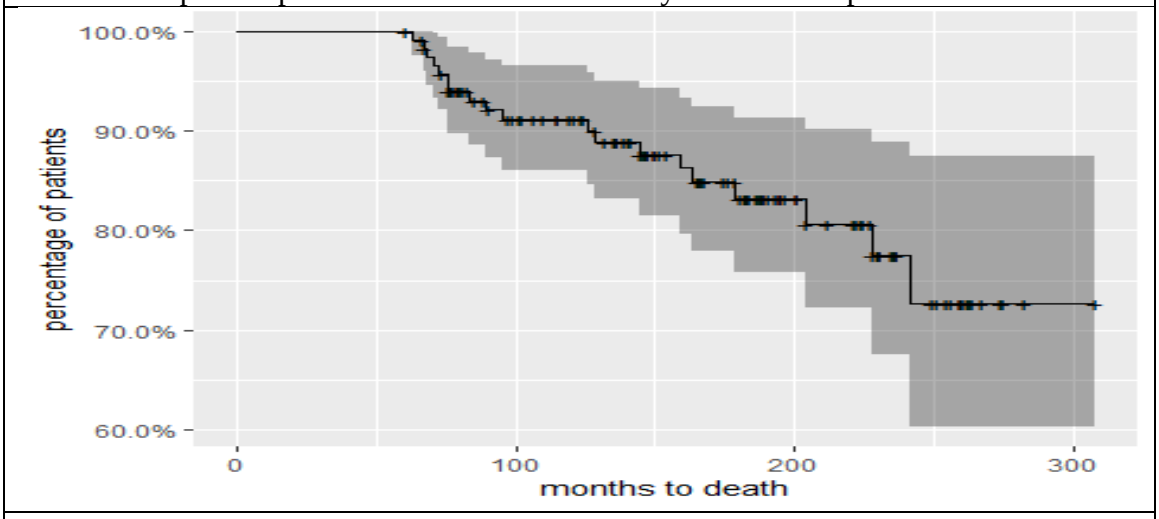
Code segment 2.10 (see Appendix) captures the survival analysis and output segment for the "Class G" TNBC-METABRIC dataset that contains 120 patient samples with treatment and follow-up information of more than 5 years. As of 241.6 months, approximately 72.6% of the patients were still alive with 19 events (deaths) recorded. See Figure 5c.



a. Overall group of 190 patients



b. Group of 70 patients with not more than 5 years follow-up



c. Group of 120 patients with more than 5 years follow-up

Figure 5 Survival Graphs for METABRIC Dataset

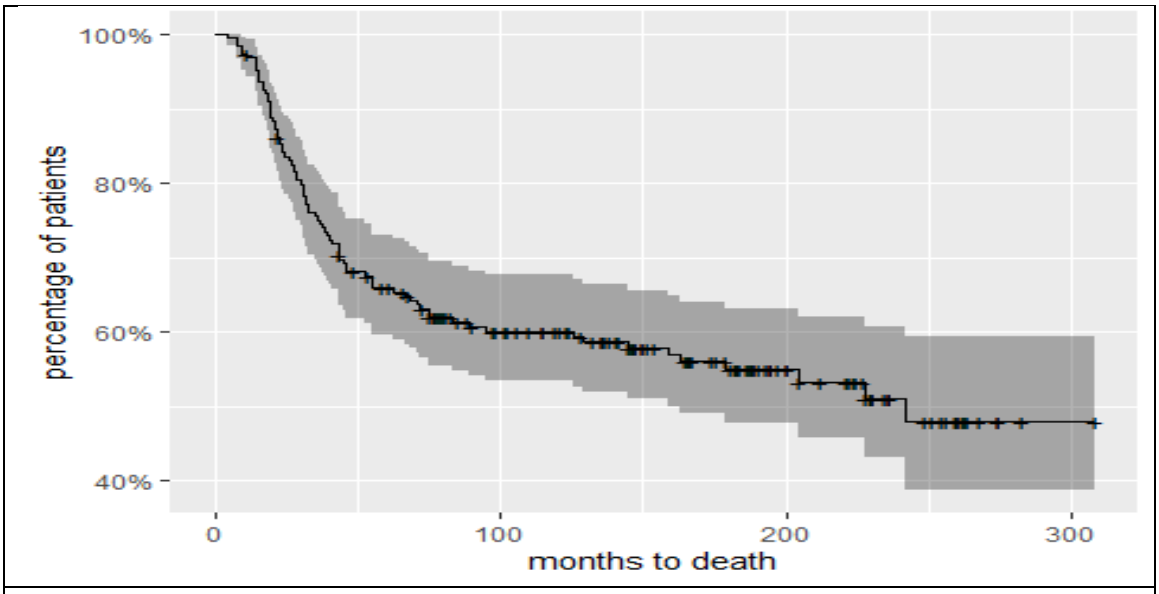
SINGLE TREATMENTS

4.2.2. Analysis for Radiation Treatment Only

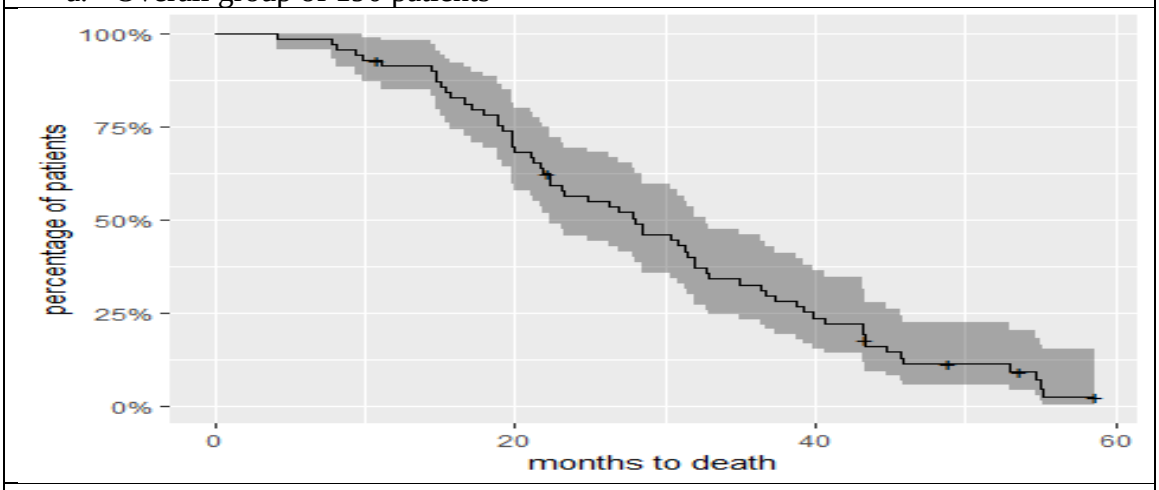
Figure 6a shows the survival graph for 190 patients who took the radiation single treatment only. It also reveals that the median survival time for those who had radiotherapy is 242 months with 83 events. (See Code Segments 2.14).

Figure 6b shows the survival graph for the group of 70 patients with follow-up of not more than 5 years, who took the radiation single treatment only. It also reveals that their median survival time is 28 months with 64 events. (See Code Segments 2.14).

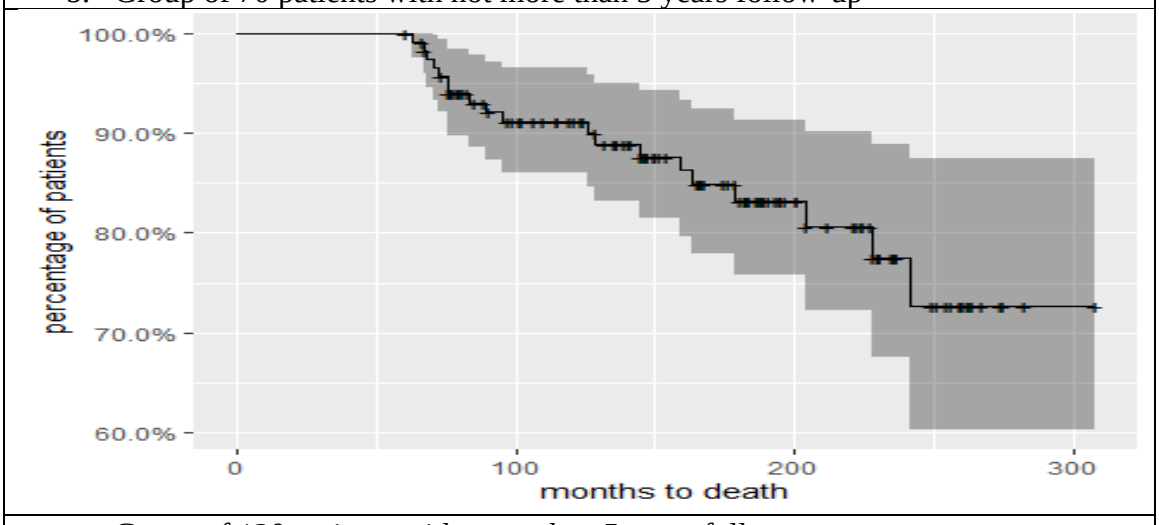
Figure 6c shows the survival graph for the group of 120 patients with follow-up of more than 5 years, who took the radiation single treatment only. It also reveals that 19 events. (See Code Segments 2.14).



a. Overall group of 190 patients



b. Group of 70 patients with not more than 5 years follow-up



c. Group of 120 patients with more than 5 years follow-up

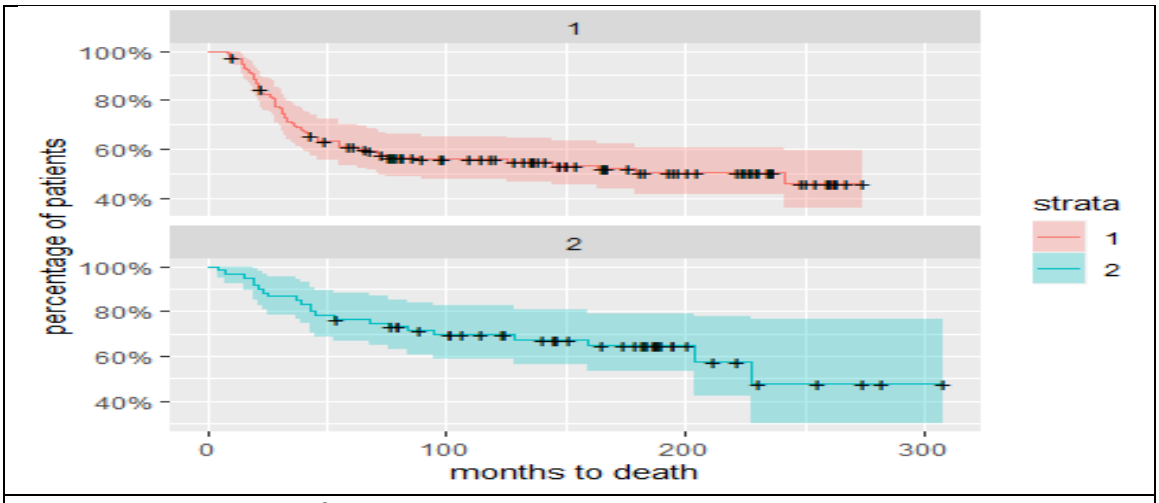
Figure 6 Survival graph for radiation treatment

4.2.3. Analysis for Chemotherapy Treatment Only

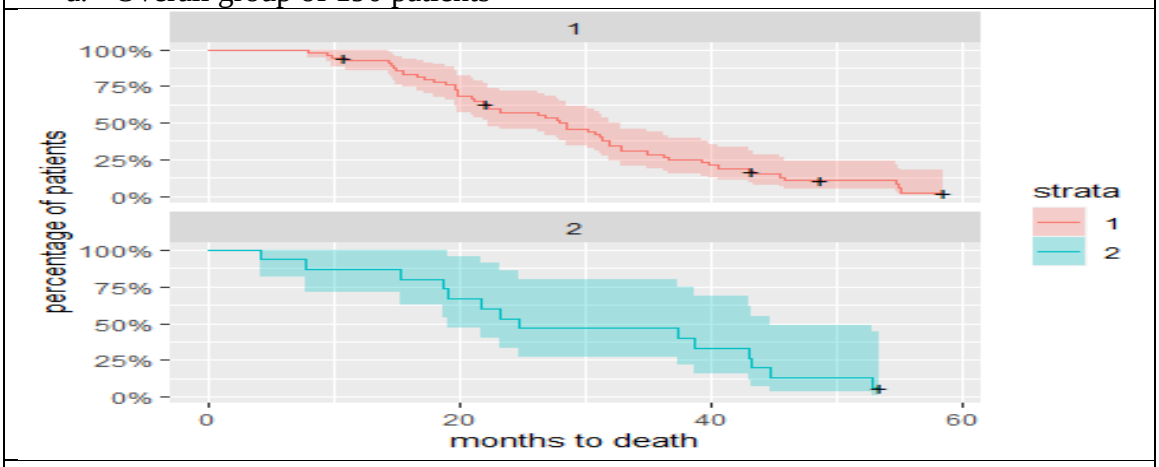
Figure 7a shows the survival graph for the 190-patient group who took the chemotherapy single treatment only. It also reveals that the median survival time for those who had chemotherapy is 242 months with 61 events. However, those in this group who did not get chemotherapy had a median survival time of 228 months with 22 events. (See Code Segments 2.15)

Figure 7b shows the survival graph for the 70-patient group who took the chemotherapy single treatment only. It also reveals that the median survival time for those who had chemotherapy is 28 months with 50 events. However, those in this group who did not get chemotherapy had a median survival time of 24.8 months with 14 events. (See Code Segments 2.15)

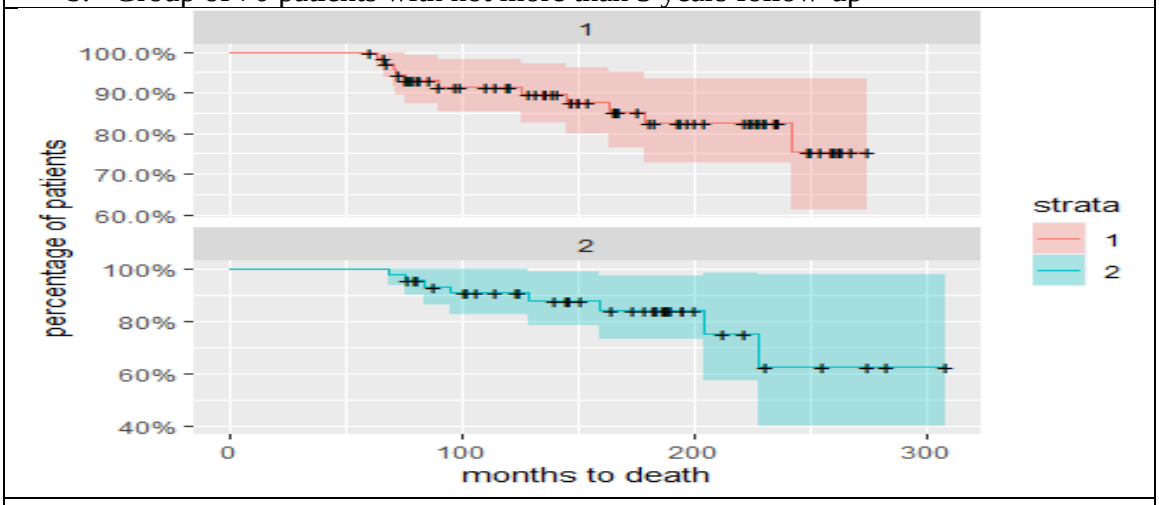
Figure 7c shows the survival graph for the 120-patient group who took the chemotherapy single treatment only. It also reveals that the patients that took chemotherapy had 11 events whereas those who did not experienced 8 events. (See Code Segments 2.15)



a. Overall group of 190 patients



b. Group of 70 patients with not more than 5 years follow-up



c. Group of 120 patients with more than 5 years follow-up

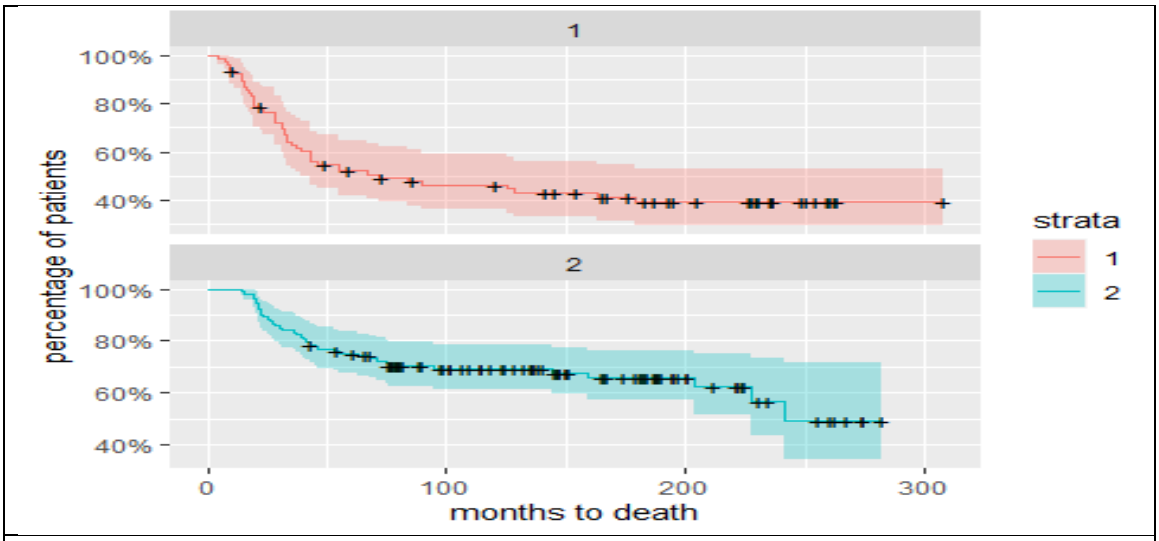
Figure 7 Survival graph for chemotherapy treatment only (Strata: 1 = Got Chemotherapy; 2 = Did not get Chemotherapy)

4.2.4. Analysis for Breast Surgery Treatment Only

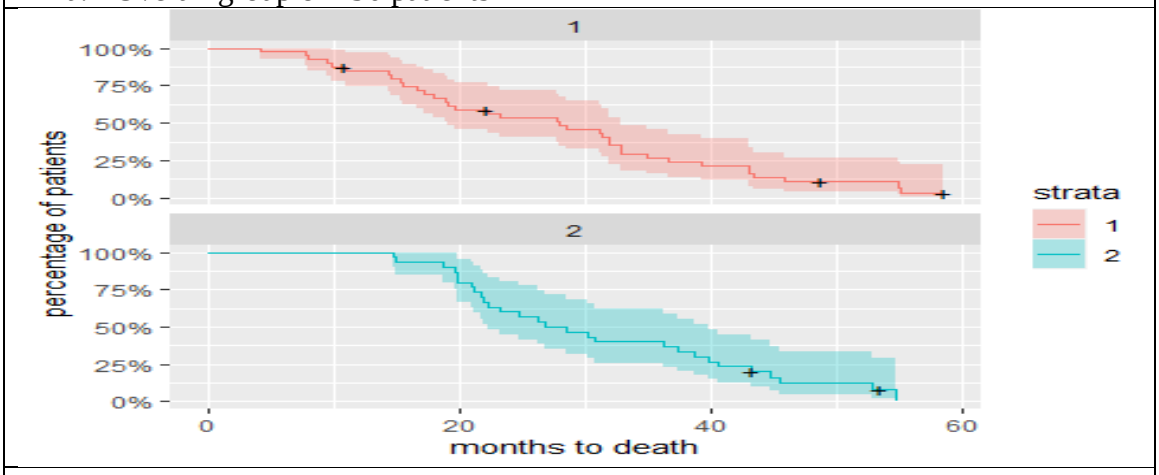
Figure 8a shows the survival graph for the 190-patient group who underwent the breast surgery single treatment only. Meaning that patients underwent either mastectomy or breast conserving. It further reveals that the median survival times for those who underwent mastectomy and breast conserving were 71.8 and 241.6 months, respectively. The number of events for mastectomy and breast conserving were 44 and 39, respectively. (See Code Segments 2.16)

Figure 8b shows the survival graph for the 70-patient group who underwent the breast surgery single treatment only. Meaning that patients underwent either mastectomy or breast conserving. It further reveals that the median survival times for those who underwent mastectomy and breast conserving were 28 and 27.7 months, respectively. The number of events for mastectomy and breast conserving were 36 and 28, respectively. (See Code Segments 2.16)

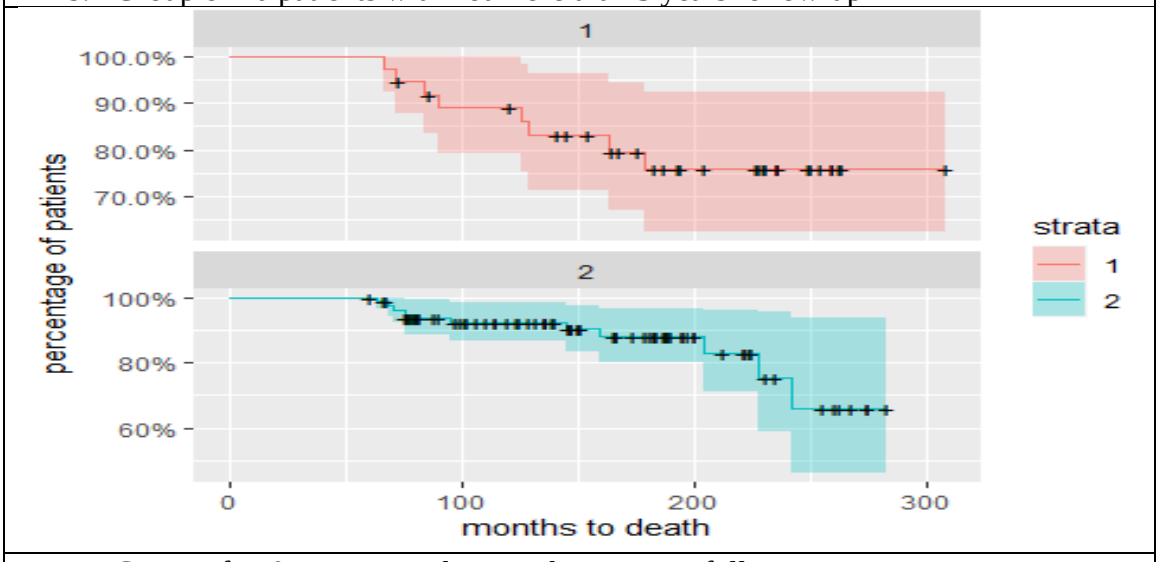
Figure 8c shows the survival graph for the 120-patient group who underwent the breast surgery single treatment only. Meaning that patients underwent either mastectomy or breast conserving. The number of events for mastectomy and breast conserving were 8 and 11, respectively. (See Code Segments 2.16)



a. Overall group of 190 patients



b. Group of 70 patients with not more than 5 years follow-up



c. Group of 120 patients with more than 5 years follow-up

Figure 8 Survival graph for breast surgery treatment only (Strata: Mastectomy = 1; Breast Conserving = 2)

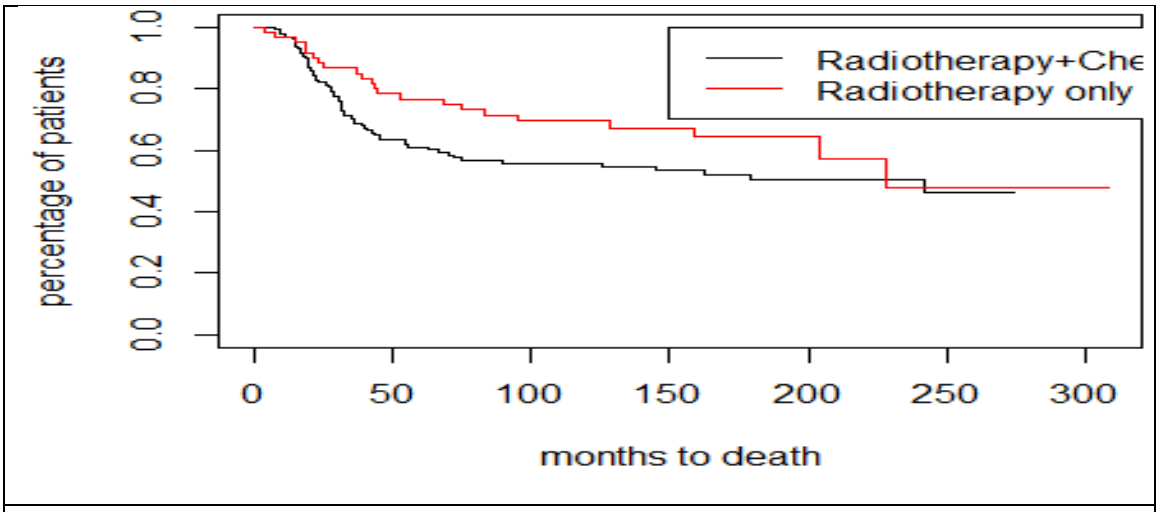
COMBINATION TREATMENTS

4.2.5. Analysis for Radiotherapy and Chemotherapy Treatments Without Surgery

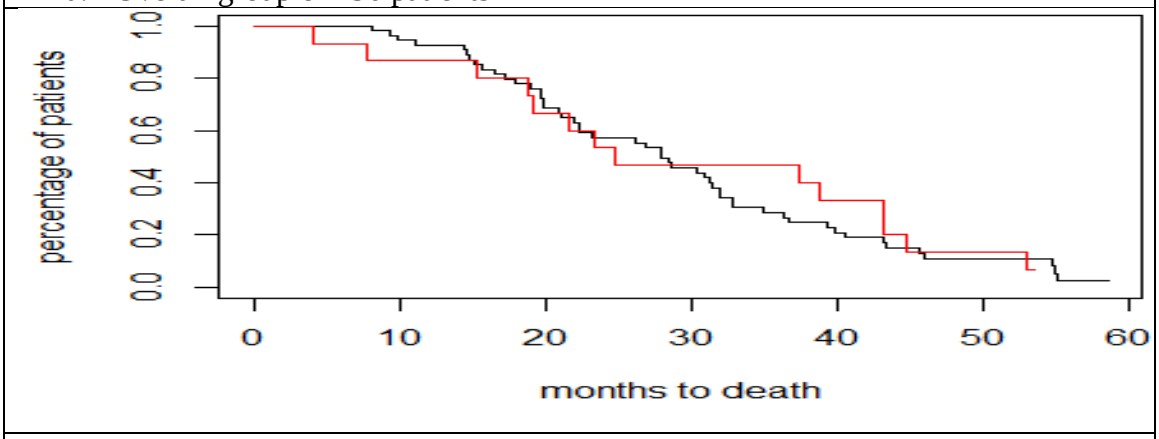
Figure 9a shows the survival graph for 190 patients who took the combination treatment of radiotherapy and chemotherapy but without breast surgery. It further reveals that those who underwent the combination treatment of chemotherapy and radiotherapy had a median survival time of 242 months with 61 events. (See Code Segments 2.17).

Figure 9b shows the survival graph for 70 patient-group who took the combination treatment of radiotherapy and chemotherapy but without breast surgery. It further reveals that those who underwent the combination treatment of chemotherapy and radiotherapy had a median survival time of 28 months with 50 events. (See Code Segments 2.17).

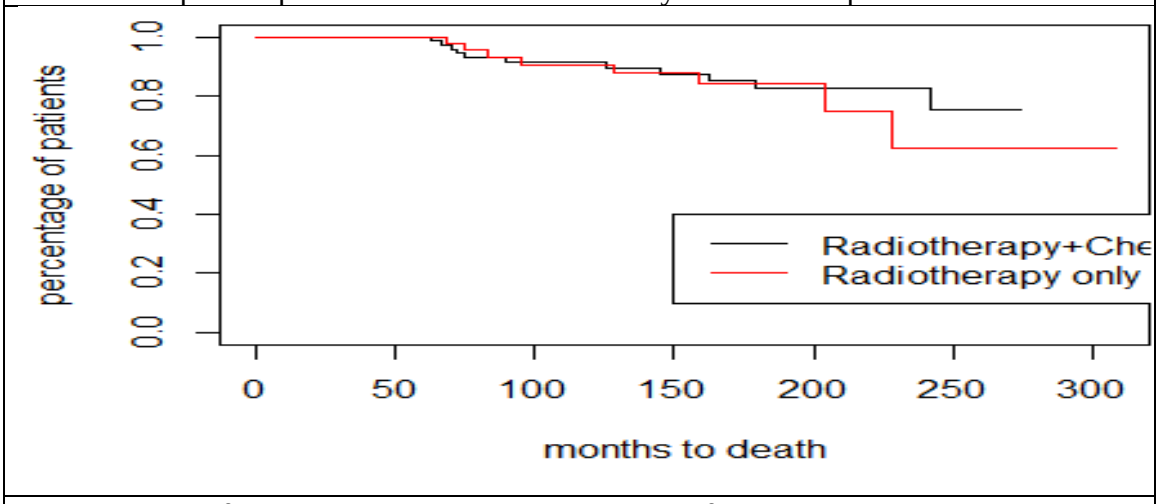
Figure 9c shows the survival graph for 120 patient-group who took the combination treatment of radiotherapy and chemotherapy but without breast surgery. It further reveals that those who underwent the combination treatment of chemotherapy and radiotherapy had 11 events. (See Code Segments 2.17).



a. Overall group of 190 patients



b. Group of 70 patients with not more than 5 years follow-up



c. Group of 120 patients with more than 5 years follow-up

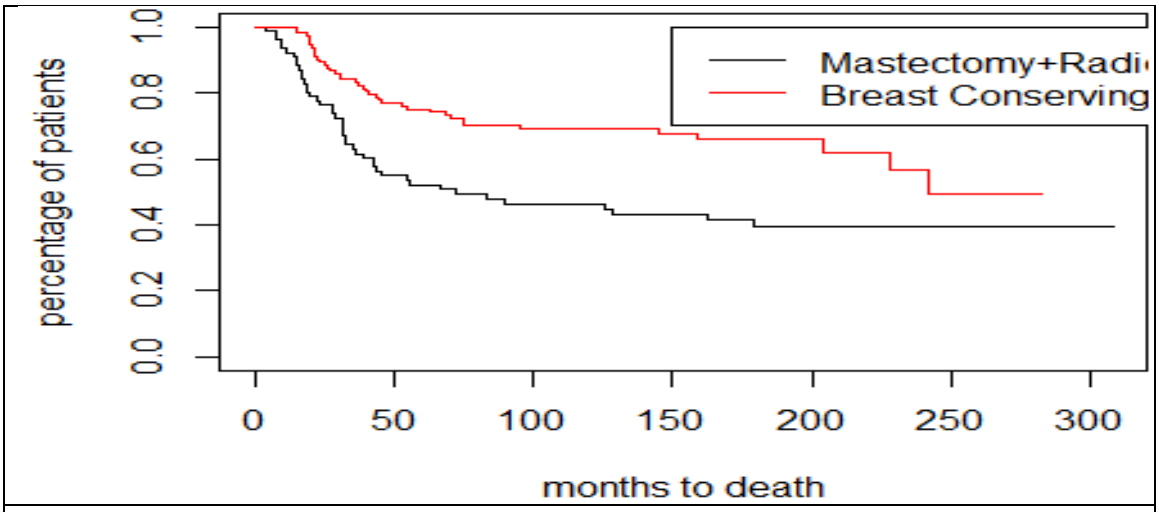
Figure 9 Survival graph for radiotherapy and chemotherapy combination treatments

4.2.6. Analysis for Breast Surgery and Radiotherapy Treatments

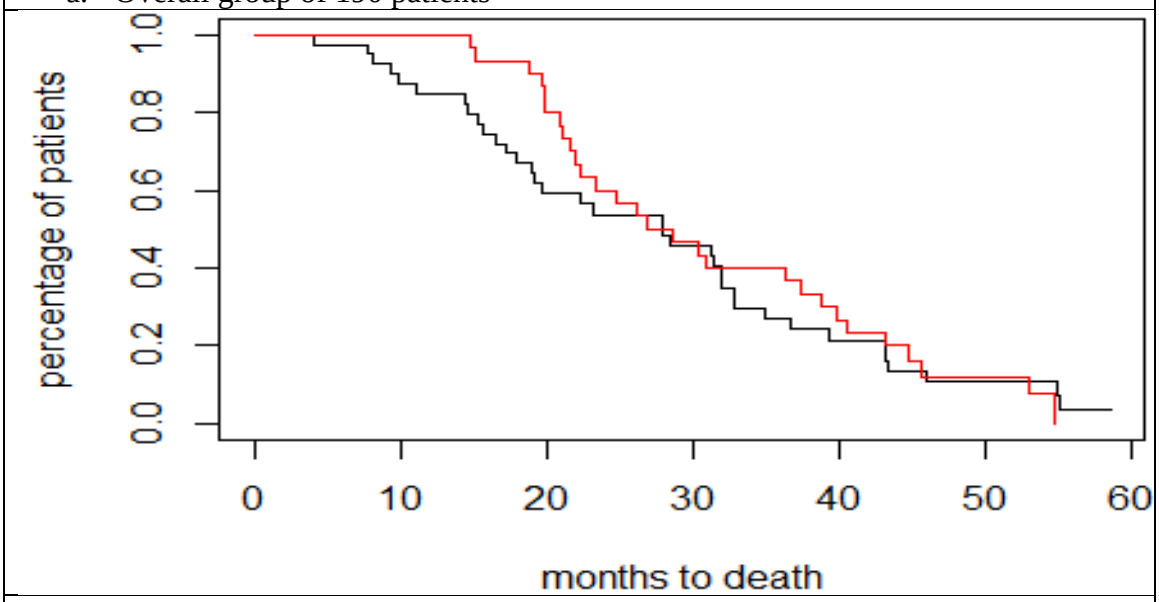
Figure 10a shows survival graph for the 190 patient-group who had a combination treatment of radiotherapy and breast surgery. It also reveals that those who had the combination treatment of breast conserving and radiotherapy had a median survival time of 241.6 months with 39 events, while those who had the combination treatment of mastectomy and radiotherapy had a median survival time of 71.8 months with 44 events (See Code Segments 2.18).

Figure 10b shows the survival graph for the 70 patient-group who had a combination treatment of radiotherapy and breast surgery. It also reveals that those who had the combination treatment of breast conserving and radiotherapy had a median survival time of 27.7 months with 28 events, while those who had the combination treatment of mastectomy and radiotherapy had a median survival time of 28 months with 36 events (See Code Segments 2.18).

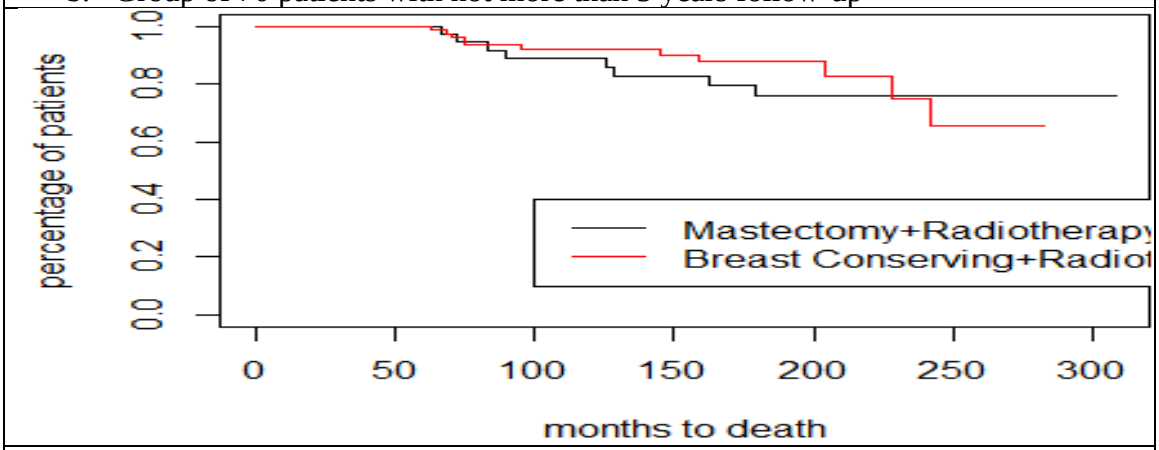
Figure 10c shows the survival graph for the 120 patient-group who had a combination treatment of radiotherapy and breast surgery. It also reveals that there were 11 events from those who had the combination treatment of breast conserving and radiotherapy, while those who had the combination treatment of mastectomy and radiotherapy experienced 8 events (See Code Segments 2.18).



a. Overall group of 190 patients



b. Group of 70 patients with not more than 5 years follow-up



c. Group of 120 patients with more than 5 years follow-up

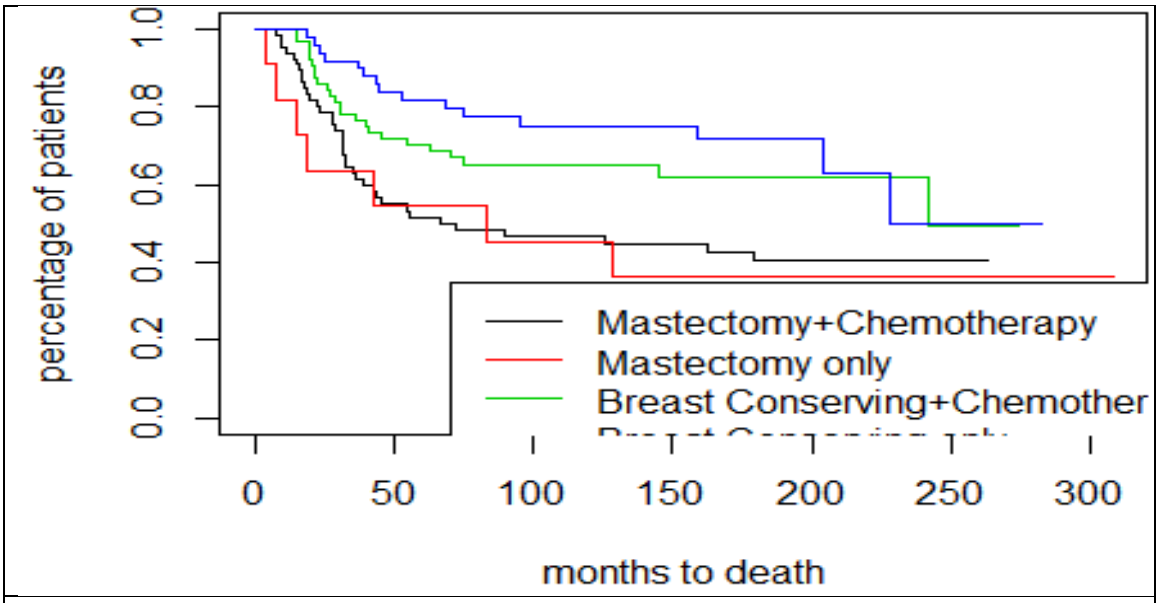
Figure 10 Survival graph for combination of surgery and radiotherapy treatments

4.2.7. Analysis for Breast Surgery and Chemotherapy Treatments

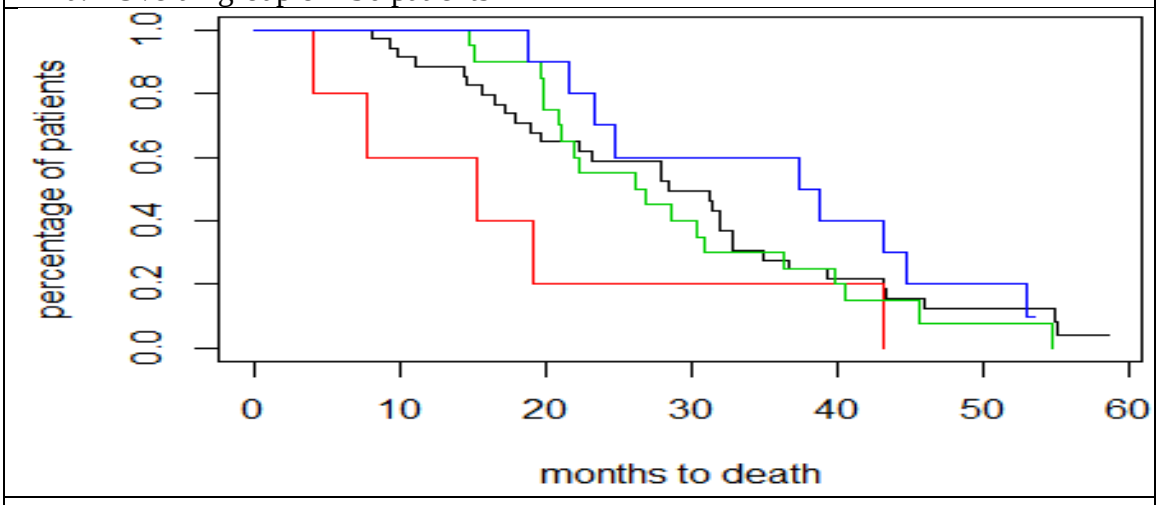
Figure 11a shows the survival graph for the 190-patient group who had the combination treatment of chemotherapy and breast surgery. It further reveals that those who had the combination treatment of mastectomy and chemotherapy had a median survival time of 71.8 months with 37 events, while those who underwent the combination treatment of breast conserving and chemotherapy had a median survival time of 241.6 months with 24 events (See Code Segments 2.19).

Figure 11b shows the survival graph for 70-patient group who had the combination treatment of chemotherapy and breast surgery. It further reveals that those who had the combination treatment of mastectomy and chemotherapy had a median survival time of 28.5 months with 31 events, while those who underwent the combination treatment of breast conserving and chemotherapy had a median survival time of 26.6 months with 19 events (See Code Segments 2.19).

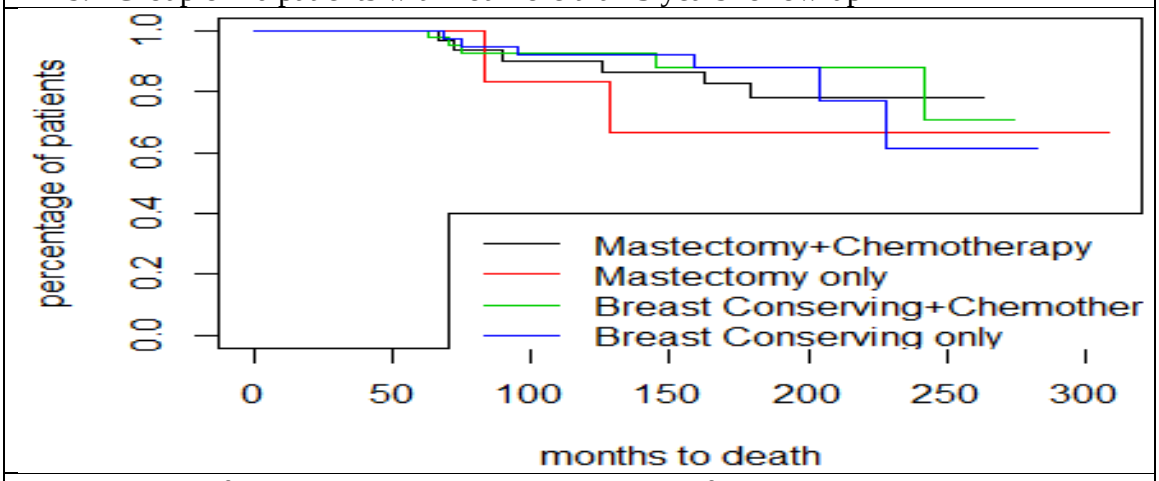
Figure 11c shows the survival graph for 120-patient group who had the combination treatment of chemotherapy and breast surgery. It further reveals that those who had the combination treatment of mastectomy and chemotherapy had total of 6 events, while those who underwent the combination treatment of breast conserving and chemotherapy had a total of 5 events (See Code Segments 2.19).



a. Overall group of 190 patients



b. Group of 70 patients with not more than 5 years follow-up

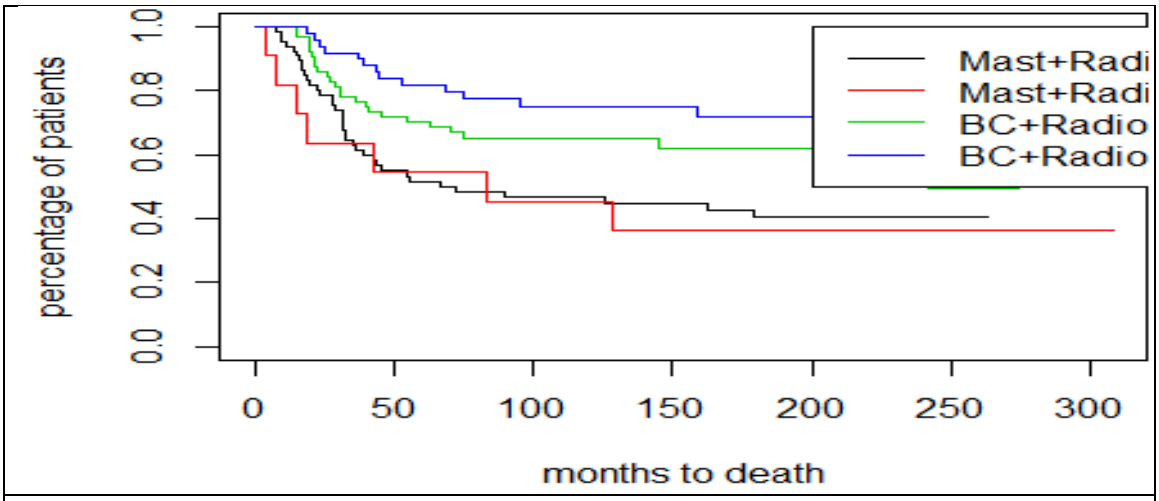


c. Group of 120 patients with more than 5 years follow-up

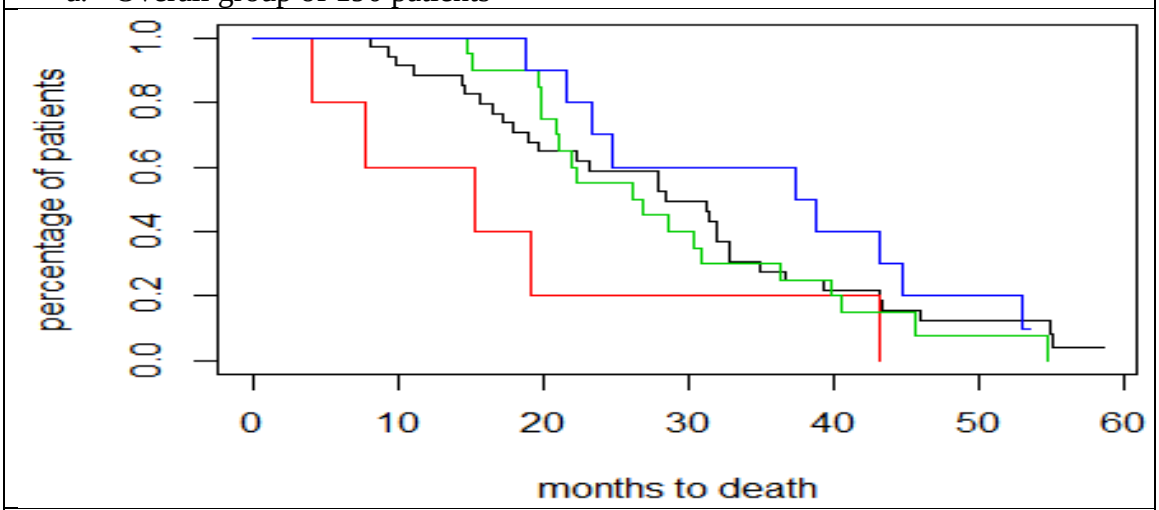
Figure 11 Survival graph for surgery and chemotherapy treatments

4.2.8. Analysis for Breast Surgery, Radiotherapy and Chemotherapy Treatments

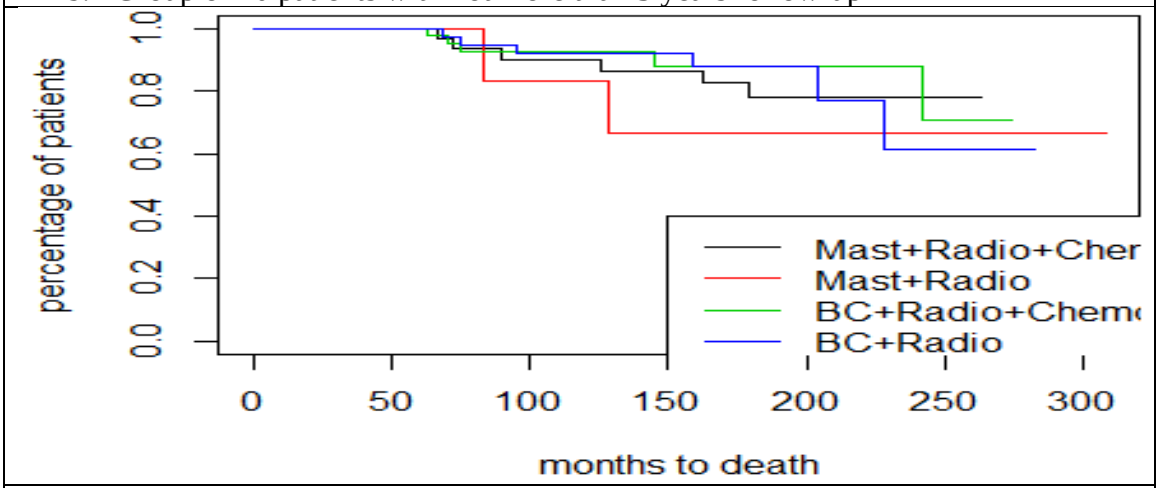
For patients who underwent the combination treatment of radiotherapy, chemotherapy and breast surgery, findings were like those reported in section 4.2.7. It further reveals that those who underwent the combination treatment of breast conserving, radiotherapy and chemotherapy had a median survival time of 241.6 months, while those who underwent the combination treatment of mastectomy, radiotherapy and chemotherapy had a median survival time of 71.8 months (See Code Segments 2.20).



a. Overall group of 190 patients



b. Group of 70 patients with not more than 5 years follow-up



c. Group of 120 patients with more than 5 years follow-up

Figure 12 Survival graph for surgery, radiotherapy and chemotherapy treatments

4.2.9. Cox Proportional Hazards Model and Tests of Significance Summary Table

Table 7 is a summary of all the p-values of the three tests used to evaluate hypotheses 1, 2 and 3 respectively. Asides the Log-Rank and Wilcoxon tests respectively, the Cox Proportional Hazards model was used to ascertain the likelihood of a patient dying while on any of the treatment type being analyzed in this study.

Table 7: Summary Table of Treatment Outcome P-Values

			P-Values			
SN	**Treatment**	**Dataset**	**LogR**	**Wil**	**CoxPH**	**Significance**
1	Radiotherapy only	E	~	~	NA	~
2		F	~	~	NA	~
3		G	~	~	NA	~
4	Chemotherapy only	E	0.1	0.08	0.112	Low
5		F	0.8	0.9	0.838	Low
6		G	0.6	0.7	0.650	Low
7	Breast Surgery only	E	0.01	5e-04	0.00141**	High
8		F	0.7	0.2	0.678	Low
9		G	0.6	0.6	0.649	Low
10	Radiotherapy and Chemotherapy	E	0.1	0.08	NA 0.112	Low
11		F	0.8	0.9	NA 0.838	Low
12		G	0.6	0.7	NA 0.65	Low
13	Breast Surgery and Radiotherapy	E	0.001	5e-04	0.00141** NA	Very High
14		F	0.7	0.2	0.678 NA	Low
15		G	0.6	0.6	0.649 NA	Low
16	Breast Surgery and Chemotherapy	E	0.008	0.003	0.00544** 0.56834	High
17		F	0.2	0.03	0.717 0.973	Low
18		G	0.8	0.7	0.501 0.502	Low
19	Breast Surgery and Radiotherapy and Chemotherapy	E	0.008	0.003	0.00544** NA 0.56834	High
20		F	0.2	0.03	0.717 NA 0.973	Low
21		G	0.8	0.7	0.501 NA 0.502	Low
Significance codes: 0 = '***' | 0.001 = '**' | 0.01 = '*'| 0.05 = '.' | 0.1 = ' ' | 1						
LogR = Log-Rank Test; Wil = Generalized Wilcoxon Test. CoxPH = Cox Proportional Hazard Ratio						
For Log-Rank and Wilcoxon analysis, see Code Segments: 2.14 – 2.20						
For Cox Proportional Hazard analysis, see Code Segments: 2.21 – 2.25						

Table 8 captures the summary of 19 treatment combinations and their survival percentages. Results show that in the absence of surgery, survival rate beyond the first 5 years after diagnosis was very poor when patients are administered with radiotherapy and/or chemotherapy as captured below:

i. percentage survival for radiotherapy & chemotherapy combination as shown in row **4d** is a paltry **4.32%.** This was the worst treatment outcome of all the treatment options.

However, when compared to ten other instances where surgery was integral to the treatment, the following outcomes were recorded:

ii. percentage survival for the first instance of breast conserving & radiotherapy combination as shown in row **5b** is **75.17%**.

iii. percentage survival for the first instance of mastectomy & radiotherapy combination as shown in row **5d** is **52.17%**.

iv. percentage survival for the first instance of breast conserving & chemotherapy combination as shown in row **6b** is **70.24%**.

v. percentage survival for the first instance of mastectomy & chemotherapy combination as shown in row **6d** is **51.69%**.

vi. percentage survival for the second instance of breast conserving & chemotherapy combination as shown in row **7b** is **33.33%**.

vii. percentage survival for the second instance of breast conserving & radiotherapy combination as shown in row **7c** is **81.63%**.

viii. percentage survival for the second instance of mastectomy & radiotherapy combination as shown in row **7g** is **54.50%.**

ix. percentage survival for breast conserving, radiotherapy and chemotherapy combination as shown in row **7d** is **70.24%.**

x. percentage survival for mastectomy, radiotherapy and chemotherapy combination as shown in row **7h** is **51.69%.**

From the summary captured in Table 8, it can be deduced that the singular instance of combination treatment that excluded surgery led to the worst percentage survival of only **4.32%**. This can be said to be a very poor outcome when compared to the ten other instances of combination treatments where surgery was featured either as breast conserving or mastectomy.

In addition, the least percentage survival outcome of **33.33%** came from the second instance of breast conserving & chemotherapy combination as shown in row **7b;** all other combinations had at least **50%** percentage survival outcome, with **81.63%** being the highest percentage survival outcome derived from the second instance of breast conserving & radiotherapy combination as shown in row **7c**.

Table 8: Survival at 60 months (5years) of treatment

		Survival at 60 months (5years) of treatment from 190 Patient Group				
	Treatment Types	**Treatment Combinations**	**Number at Risk**	**Number of Events**	**Survival**	
					Decimal	**Percentage %**
1	Radiotherapy only	a. No	~	~	~	~
		b. Yes	120	64	0.6596	65.96
2	Chemotherapy only	a. No	45	14	0.7667	76.67
		b. Yes	75	50	0.6095	60.95
3	Breast Surgery only	a. Breast Conserving	83	28	0.7517	75.17
		b. Mastectomy	37	36	0.5217	52.17
4	Radiation (Rad) and Chemotherapy (Chem)	a. Rad=No; Chem=No	~	~	~	~
		b. Rad=No; Chem=Yes	~	~	~	~
		c. Rad=Yes; Chem=No	45	14	0.0546	5.46
		d. Rad=Yes; Chem=Yes	75	50	0.0432	**4.32**
5	Breast Surgery and Radiotherapy	a. Breast Conserving & Rad=No	~	~	~	~
		b. Breast Conserving & Rad=Yes	83	28	0.7517	**75.17**
		c. Mastectomy & Rad=No	~	~	~	~
		d. Mastectomy & Rad=Yes	37	36	0.5217	**52.17**
6	Breast Surgery and Chemotherapy	a. Breast Conserving & Chem=No	39	9	0.8163	81.63
		b. Breast Conserving & Chem =Yes	44	19	0.7024	70.24
		c. Mastectomy & Chem =No	6	5	0.545	54.45
		d. Mastectomy & Chem =Yes	31	31	0.5169	51.69

7	Breast Surgery; Chemotherapy and Radiotherapy	a. Breast Conserving; Rad=No; Chem = No	7	2	0.7780	77.80
		b. Breast Conserving; Rad=No; Chem = Yes	1	2	0.3333	33.33
		c. Breast Conserving; Rad=Yes; Chem = No	39	9	0.8163	**81.63**
		d. Breast Conserving; Rad=Yes; Chem = Yes	44	19	0.7024	**70.24**
		e. Mastectomy; Rad=No; Chem = No	~	~	~	~
		f. Mastectomy; Rad=No; Chem = Yes	~	~	~	~
		g. Mastectomy; Rad=Yes; Chem = No	6	5	0.5450	**54.50**
		h. Mastectomy; Rad=Yes; Chem = Yes	31	31	0.5169	**51.69**

From Table 7, it can be inferred that the patient groups of 70 (Class F) and 120 (Class G) respectively did not get the backing of the cox proportional hazard model. It means that patients in both samples are likely to die during their respective treatments.

On the other hand, the overall group of 190 patient sample (Class E), appears to have acceptable levels of significance in terms of their cox proportional hazard model p-values. This is especially when surgery is part of the treatment plan.

In addition, a look at Table 8 suggests that this is true given the high survival rates associated with the groups that have a combination of surgery in their treatment courses.

Furthermore, the P-Values from the Log-Rank and Wilcoxon tests suggests that:

a. Breast surgery and radiotherapy – 0.001 and 5e-04 respectively, which are both far less than 0.05
b. Breast surgery, chemotherapy, and radiotherapy – 0.008 and 0.003 respectively, which are both far less than 0.05
c. Radiotherapy and chemotherapy – 0.10 | 0.08, which are both greater than 0.05
d. Breast surgery and chemotherapy – 0.008 | 0.03, which are both less than 0.05

In all these mentioned instances where surgery and radiotherapy featured, P-Values of very highly significant values were derived. More so the cox model validated the instances where surgery featured as part of the treatment combinations.

In view of the foregoing, it can be inferred that a relationship exists between the administration of adjuvant radiotherapy and enhanced survival of post-surgery TNBC patients.

It can also be inferred *that there is a significant difference in survival outcome after 5 years between patients who undergo combination treatment of chemotherapy and/or radiotherapy without surgery as compared to patients who had surgery as one of the treatment methods.*

Furthermore, analysis was carried out using the R function; **summary(object,60).** 60 represents 60 months which is equivalent to 5years.

Table 8 captures the summary of treatment combinations and their survival percentages. The breakdown relevant to this test is as follows:

i. percentage values highlighted in **blue** show a **23%** disparity between the breast conserving & radiotherapy combination (75.17% survival – see row **5b**) and the mastectomy & radiotherapy combination (52.17% survival – see row **5d**). This means that 23% more patients who underwent breast conserving exceeded the 60-month benchmark as compared to those who subscribed to mastectomy.

ii. percentage values highlighted in **green** show a **27%** disparity between the breast conserving & radiotherapy combination (81.63% survival – see row **7c**) and the mastectomy & radiotherapy combination (54.5% survival – see row **7g**). This means that 27% more patients who underwent breast conserving exceeded the 60-month benchmark as compared to those who subscribed to mastectomy.

iii. percentage values highlighted in **yellow** show an **18.55%** disparity between the breast conserving, chemotherapy & radiotherapy combination (70.24% survival – see row **7d**) and the mastectomy, chemotherapy & radiotherapy combination (51.69% survival – see row **7h**). This means that 18.55% more patients who underwent breast conserving exceeded the 60-month benchmark as compared to those who subscribed to mastectomy.

We can therefore infer that there is a difference in survival outcome after 5 years between patients who combine breast conserving procedure and radiotherapy as compared to patients who combine mastectomy and radiotherapy. The patients who adopted breast conserving as the surgical procedure of choice had a better chance of survival beyond the first 5years after diagnosis than patients who opted for mastectomy.

4.3. Results from Genetic (Mutation) Analyses

In this section, a wide range of data analyses was carried out to ascertain possible genetic markers that was used to combine already known clinical features. The combination of genetic and clinical features was subsequently used as the basis to build the predictive model to be discussed in full in section 4.4.

4.3.1. Copy Number Alterations (CNAs)

From the analysis, the genes with CN Alterations were limited to 50 with the highest significance in terms of Fisher's Exact Test (P-Value less than 0.05).

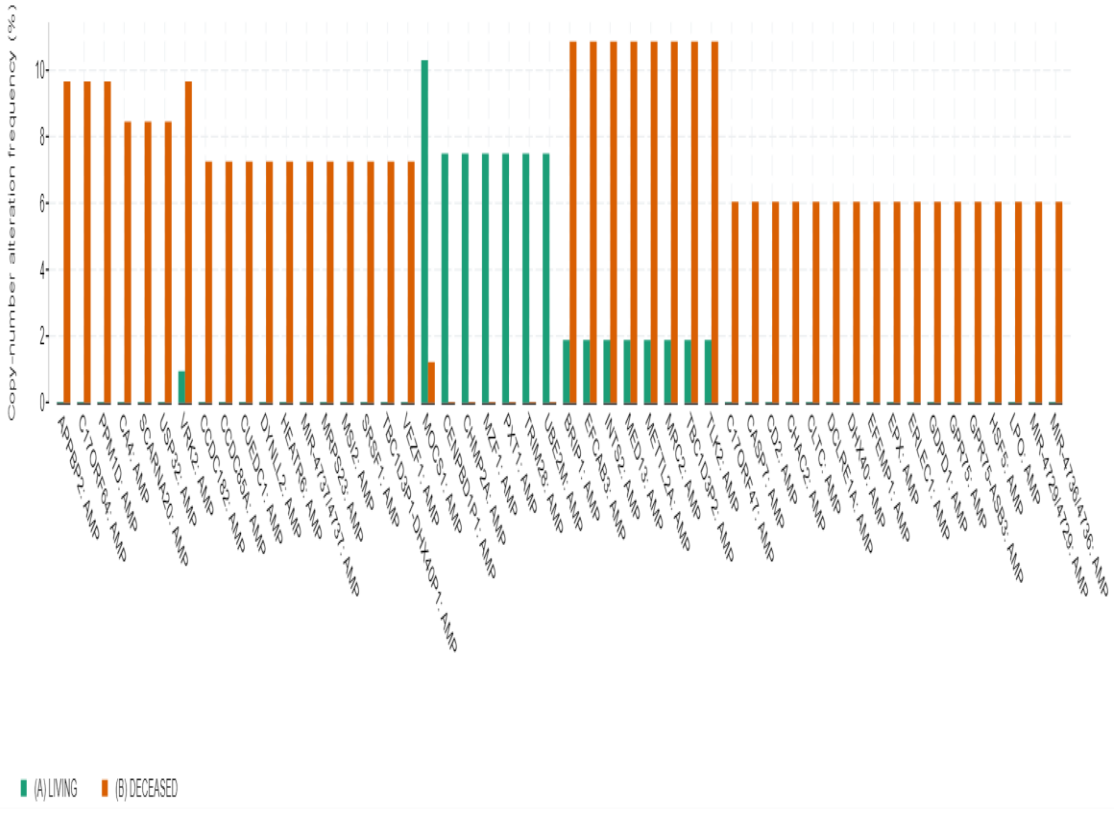


Figure 13 Plot showing 50 Genes with CNA with most significant p-values

Table 9: Summary of Genes with CN Alterations as well as the loci of occurrence and P-values

	Gene	Cytoband/ Localization	Altera tion	(A) Living	(B) Deceased	p-Value	Enriched In
1	APPBP2	17q23.2	Amp	0 (0.00%)	8 (9.64%)	1.09E-03	(B) Deceased
2	C17ORF64	17q23.2	Amp	0 (0.00%)	8 (9.64%)	1.09E-03	(B) Deceased
3	PPM1D	17q23.2	Amp	0 (0.00%)	8 (9.64%)	1.09E-03	(B) Deceased
4	CA4	17q23.1	Amp	0 (0.00%)	7 (8.43%)	2.62E-03	(B) Deceased
5	SCARNA20	17q23.2	Amp	0 (0.00%)	7 (8.43%)	2.62E-03	(B) Deceased
6	USP32	17q23.1-q23.2	Amp	0 (0.00%)	7 (8.43%)	2.62E-03	(B) Deceased
7	VRK2	2p16.1	Amp	1 (0.93%)	8 (9.64%)	6.20E-03	(B) Deceased
8	CCDC182	17q22	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
9	CCDC85A	2p16.1	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
10	CUEDC1	17q22	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
11	DYNLL2	17q22	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
12	HEATR6	17q23.1	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
13	MIR-4737/4737		Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
14	MRPS23	17q22	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
15	MSI2	17q22	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
16	SRSF1	17q22	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
17	TBC1D3P1-DHX40P1	17q23.1	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
18	VEZF1	17q22	Amp	0 (0.00%)	6 (7.23%)	6.26E-03	(B) Deceased
19	MOCS1	6p21.2	Amp	11 (10.28%)	1 (1.20%)	8.70E-03	(A) Living
20	CENPBD1P1	19q13.43	Amp	8 (7.48%)	0 (0.00%)	8.99E-03	(A) Living
21	CHMP2A	19q13.43	Amp	8 (7.48%)	0 (0.00%)	8.99E-03	(A) Living
22	MZF1	19q13.43	Amp	8 (7.48%)	0 (0.00%)	8.99E-03	(A) Living

23	PXT1	6p21.31	Amp	8 (7.48%)	0 (0.00%)	8.99E-03	(A) Living
24	TRIM28	19q13.43	Amp	8 (7.48%)	0 (0.00%)	8.99E-03	(A) Living
25	UBE2M	19q13.43	Amp	8 (7.48%)	0 (0.00%)	8.99E-03	(A) Living
26	BRIP1	17q23.2	Amp	2 (1.87%)	9 (10.84%)	9.85E-03	(B) Deceased
27	EFCAB3	17q23.2	Amp	2 (1.87%)	9 (10.84%)	9.85E-03	(B) Deceased
28	INTS2	17q23.2	Amp	2 (1.87%)	9 (10.84%)	9.85E-03	(B) Deceased
29	MED13	17q23.2	Amp	2 (1.87%)	9 (10.84%)	9.85E-03	(B) Deceased
30	METTL2A	17q23.2	Amp	2 (1.87%)	9 (10.84%)	9.85E-03	(B) Deceased
31	MRC2	17q23.2	Amp	2 (1.87%)	9 (10.84%)	9.85E-03	(B) Deceased
32	TBC1D3P2	17q23.2	Amp	2 (1.87%)	9 (10.84%)	9.85E-03	(B) Deceased
33	TLK2	17q23.2	Amp	2 (1.87%)	9 (10.84%)	9.85E-03	(B) Deceased
34	C17ORF47	17q22	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
35	CASP7	10q25.3	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
36	CD2	1p13.1	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
37	CHAC2	2p16.2	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
38	CLTC	17q23.1	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
39	DCLRE1A	10q25.3	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
40	DHX40	17q23.1	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
41	EFEMP1	2p16.1	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased

42	EPX	17q22	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
43	ERLEC1	2p16.2	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
44	GDPD1	17q22	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
45	GPR75	2p16.2	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
46	GPR75-ASB3	2p16.2	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
47	HSF5	17q22	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
48	LPO	17q22	Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
49	MIR-4729/4729		Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased
50	MIR-4736/4736		Amp	0 (0.00%)	5 (6.02%)	0.0148	(B) Deceased

From the plot captured in Figure 13, all the copy number alterations are due to amplification. Of the 50 genes profiled, MOCS1, CENPBD1P1, CHMP2A, MZF1, PXT1, TRIM28 and UBE2M are the seven instances where the enrichment is prevalent in the living patients. In the majority of 43 instances, the enrichment was found to be prevalent in the deceased patients.

From available studies, PPM1D and MSI2 are known oncogenes, BRIP1 and CLTC are known tumor suppressor genes whereas MSI2 and CLTC are fusion genes. Of the four genes, only BRIP1 is a germline gene located on chromosome 17 (17q22) while the other three are somatic. BRIP1 is known to work with BRCA1 to repair DNA. In addition, these four genes are the only well documented genes from available literature as established by the OncoKB Cancer Gene List – a

catalogue of 1039 genes, (last update: August 4, 2019) as well as the Catalogue of Somatic Mutations in Cancer (COSMIC). From observation very little is known about most of the genes identified in Figure 13. This implies that a systematic analysis of their interaction could lead to a better understanding of the implications of the administration of radiation treatment to patients of TNBC.

Table 9 shows all 50 genes with CN Alterations with their corresponding loci as well as the type of alteration which appears to be majorly amplification. This is also reflected in Figure 14.

Figure 14 Plot showing alteration profile of genes showing evidence of mostly amplification activities

4.3.2. Clinical Significance

A strong clinical significance exists between the combination treatments of surgery and radiotherapy and the overall survival of the TNBC patients with a Chi-Square derived P-value of less than 10^{-10}.

4.3.3. Mutated Genes in 190 Patient Samples

From the analysis carried out using OQL on cBioPortal, the mutated genes that carry statistical significance based on Fisher's Exact Test are MAP3K1, SHANK2, AKT1, MEN1, PIK3CA and NCOR2. All six genes are somatic mutations. Until now, the MEN1 is not known to have links to breast cancer. Other tumor suppressor genes are MAP3KI and NCOR2, while AKT1 and PIK3CA are both oncogenes. Not much is yet known about SHANK2. Except for SHANK2 and NCOR2 which were enriched in the living samples, the rest were enriched in the deceased samples as shown in Figure 15.

The six queried genes were altered in 78 (41%) of queried patients/samples.

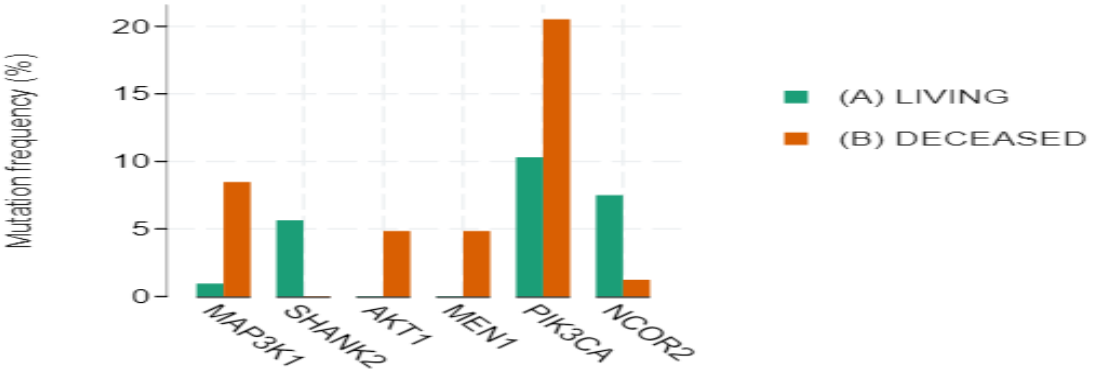


Figure 15: Plot showing 6 Mutated Genes with most significant p-values

Table 70: Summary of Mutated Genes as well as the loci of occurrence and P-values

	Gene	Cytoband/ Localization	**Role in Cancer	(A) Living	(B) Deceased	p-Value	Enriched in
1	MAP3K1	5q11.2	*TSG	1 (0.93%)	7 (8.43%)	0.0133	(B) Deceased
2	SHANK2	11q13.3-q13.4	Unknown	6 (5.61%)	0 (0.00%)	0.03	(A) Living
3	AKT1	14q32.33	Oncogene	0 (0.00%)	4 (4.82%)	0.0349	(B) Deceased
4	MEN1	11q13.1	TSG	0 (0.00%)	4 (4.82%)	0.0349	(B) Deceased
5	PIK3CA	3q26.32	Oncogene	11 (10.28%)	17 (20.48%)	0.0395	(B) Deceased
6	NCOR2	12q24.31	TSG	8 (7.48%)	1 (1.20%)	0.0418	(A) Living
*TSG – Tumor Suppressor Gene **Role in Cancer – according to OncoKB Cancer Gene List							

Table 10 reveals the six mutated genes and their respective loci as well as their P-values for which they were deemed significant since they all have values less than 0.05.

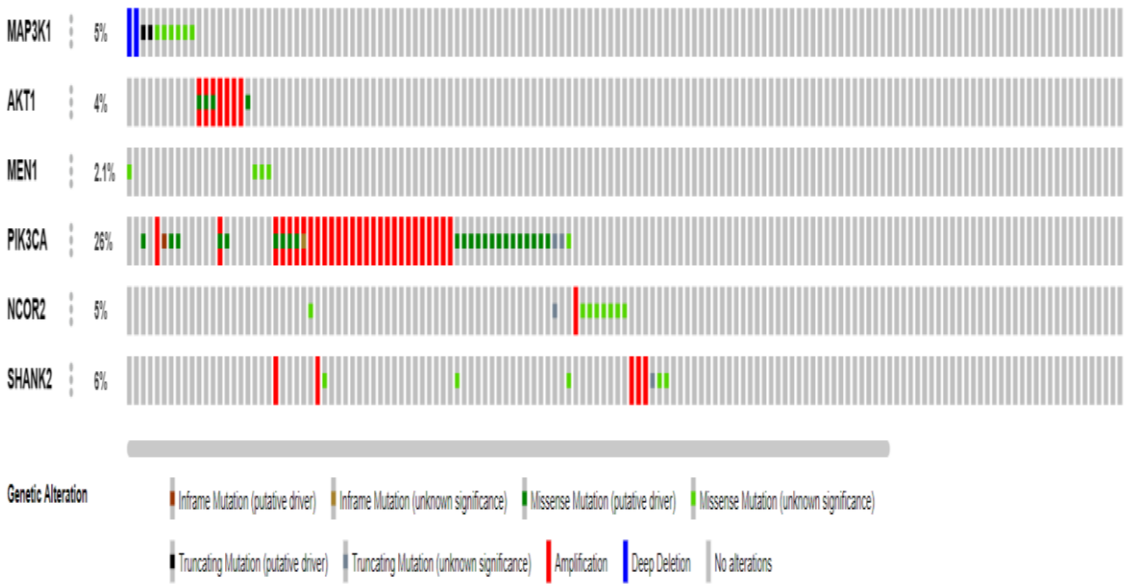


Figure 16: Plot showing the mutated profile of genes

It can be deduced from Figure 16 that the types of mutation are widespread, ranging from amplification, deep deletion, missense, truncating and in-frame mutations of varying degrees.

4.3.4. Analysis Based on PAM50 Classification

Based on the PAM 50 Classification, Figure 17, and Table 11 show 13 genes that suggests significant levels of mutations based on Fisher's Exact Test. The genes are: MAP2K4, AGMO, TAF1, GATA3, PIK3CA, BIRC6, DNAH11, NF1, TP53, SYNE1, CTNNA1 and FOXO3. Of these 13 genes only PIK3CA showed levels of significance in the prior analysis summarized in Table 10. Whereas 12 of these genes are somatic, **TAF1 is germline.** It is important to note that for TP53, a Tumor Suppressor Gene, 84% of samples in the Basal Group had mutations, while 70.77% of samples in the Claudin-Low Group had mutations. The PAM50

classification is based on early transcriptomic profiling of breast cancer that made use of microarrays to classify breast tumors. Five of these subtypes are captured in the METABRIC dataset as follows: Basal, claudin-Low, HER2-Enriched, Luminal-A (LumA), and Normal.

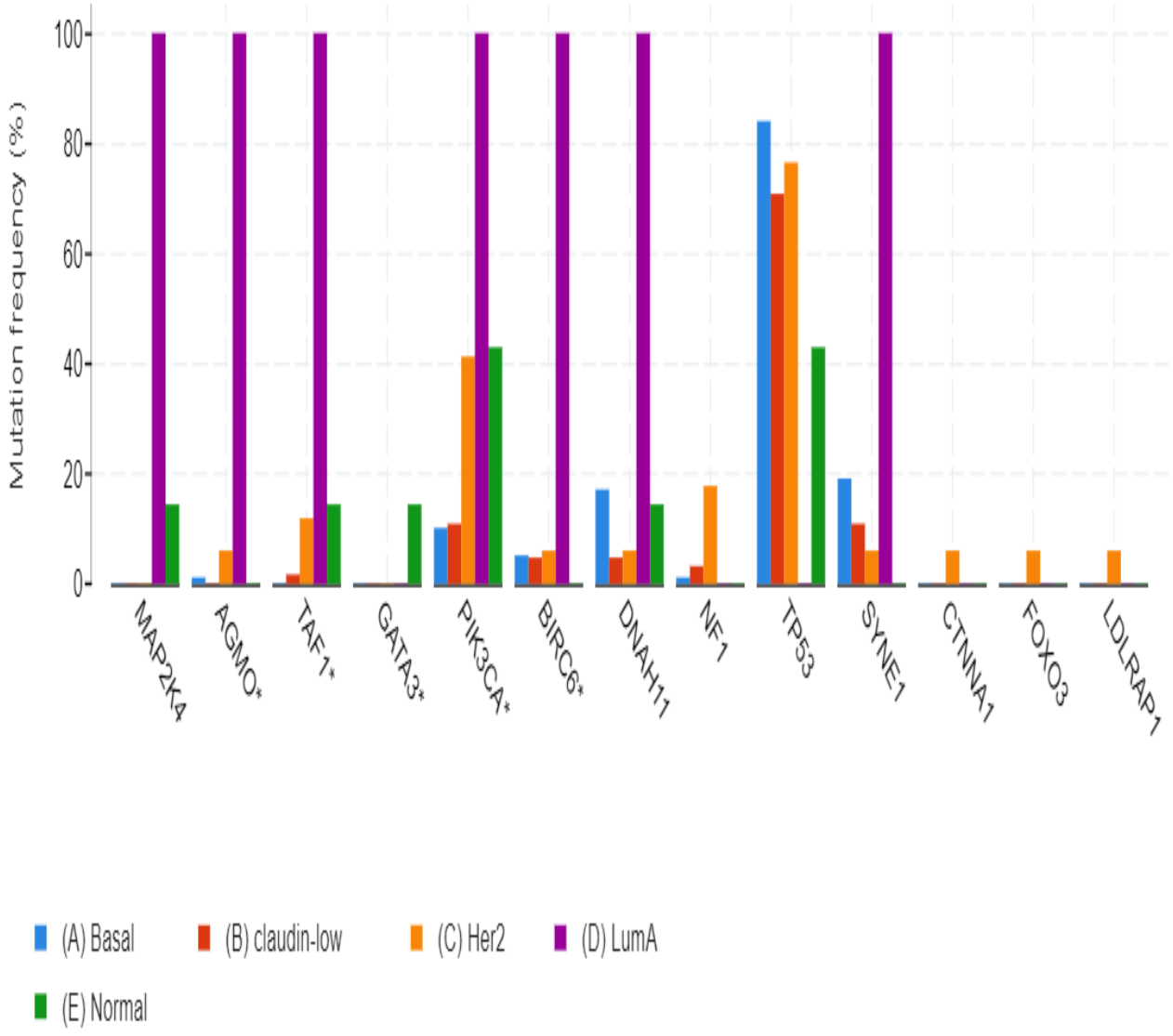


Figure 17: Plot showing the mutated profile of genes based on the PAM50 Classification

Table 8: Summary of Mutated Genes as well as the loci of occurrence and P-values based on PAM50 Classification

	Gene	Cytoband/ Localization	**Role in Cancer	(A) Basal (B)	(B) Claudin-low	(C) HER-2
1	MAP2K4	17p12	*TSG	0 (0.00%)	0 (0.00%)	0 (0.00%)
2	AGMO	7p21.2	Unknown	1 (1.00%)	0 (0.00%)	1 (5.88%)
3	TAF1	Xq13.1	Unknown	0 (0.00%)	1 (1.54%)	2 (11.76%)
4	GATA3	10p14	Oncogene/*TSG	0 (0.00%)	0 (0.00%)	0 (0.00%)
5	**PIK3CA**	3q26.32	Oncogene	10 (10.00%)	7 (10.77%)	7 (41.18%)
6	BIRC6	2p22.3	Oncogene/Fusion	5 (5.00%)	3 (4.62%)	1 (5.88%)
7	DNAH11	7p15.3	Unknown	17 (17.00%)	3 (4.62%)	1 (5.88%)
8	NF1	17q11.2	*TSG	1 (1.00%)	2 (3.08%)	3 (17.65%)
9	TP53	17p13.1	*TSG	84 (84.00%)	46 (70.77%)	13 (76.47%)
10	SYNE1	6q25.2	Unknown	19 (19.00%)	7 (10.77%)	1 (5.88%)
11	CTNNA1	5q31.2	Unknown	0 (0.00%)	0 (0.00%)	1 (5.88%)
12	FOXO3	6q21	Oncogene/*TSG /Fusion	0 (0.00%)	0 (0.00%)	1 (5.88%)
13	LDLRAP1	1p36.11	Unknown	0 (0.00%)	0 (0.00%)	1 (5.88%)

	Gene	Cytoband/ Localization	**Role in Cancer	(D) LumA	(E) Normal	p-Value	Most enriched in
1	MAP2K4	17p12	*TSG	1 (100.00%)	1 (14.29%)	0	**(D) LumA**
2	AGMO	7p21.2	Unknown	1 (100.00%)	0 (0.00%)	1.81E-13	**(D) LumA**
3	TAF1	Xq13.1	Unknown	1 (100.00%)	1 (14.29%)	5.18E-10	**(D) LumA**
4	GATA3	10p14	Oncogene/*TSG	0 (0.00%)	1 (14.29%)	2.78E-05	**(E) Normal**
5	**PIK3CA**	3q26.32	Oncogene	1 (100.00%)	3 (42.86%)	1.79E-04	**(D) LumA**
6	BIRC6	2p22.3	Oncogene/Fusion	1 (100.00%)	0 (0.00%)	9.98E-04	**(D) LumA**
7	DNAH11	7p15.3	Unknown	1 (100.00%)	1 (14.29%)	8.73E-03	**(D) LumA**
8	NF1	17q11.2	*TSG	0 (0.00%)	0 (0.00%)	9.25E-03	**(C) Her2**

9	TP53	17p13.1	*TSG	0 (0.00%)	3 (42.86%)	0.0167	**(A) Basal**
10	SYNE1	6q25.2	Unknown	1 (100.00%)	0 (0.00%)	0.0354	**(D) LumA**
11	CTNNA1	5q31.2	Unknown	0 (0.00%)	0 (0.00%)	0.0367	**(C) Her2**
12	FOXO3	6q21	Oncogene/*TSG /Fusion	0 (0.00%)	0 (0.00%)	0.0367	**(C) Her2**
13	LDLRAP1	1p36.11	Unknown	0 (0.00%)	0 (0.00%)	0.0367	**(C) Her2**

Table 9 and Figure 18 show the Overall Survival Kaplan-Meier Estimate based on the PAM 50 Classification. It shows that the median months survival for Normal-Like is quite high at 159.2 months compared to that of HER-2 which is just 43.4 months.

Table 9: Table showing Overall Survival Kaplan-Meier Estimate based on the PAM50 Classification

	Number of Cases, Total	Number of Cases, Deceased	Median Months Survival
(A) Basal	100	44	NA
(B) claudin-low	65	22	NA
(C) HER-2	17	13	43.4
(D) LumA	1	0	NA
(E) Normal	7	4	159.2333333
TOTAL	**190**	**83**	

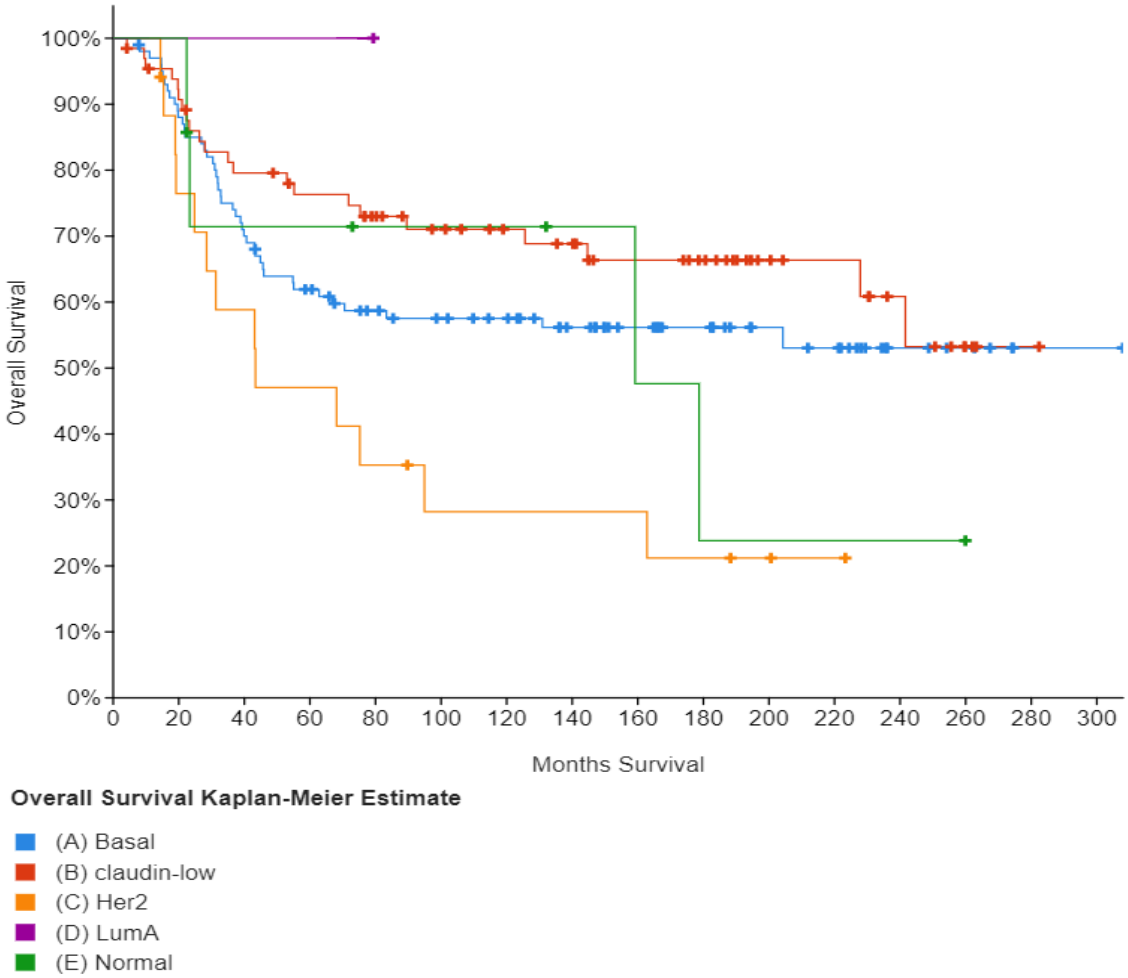


Figure 18: Plot showing Overall Survival Kaplan-Meier Estimate based on the PAM50 Classification

Table 10: Gene Classification for Predictive Model

Gene Class	Description	Genes	Sub-Total	Total
Class A	Genes derived from general mutation analysis	1. AKT1 2. MAP3K1 3. MEN1 4. NCOR2 5. PIK3CA 6. SHANK2	6	**18 Unique Genes**
Class B	Genes derived from PAM50 analysis	1. MAP2K4 2. AGMO 3. TAF1 4. GATA3 5. PIK3CA 6. BIRC6 7. DNAH11 8. NF1 9. TP53 10. SYNE1 11. CTNNA1 12. FOXO3 13. LDLRAP1	13	
***All analyses was carried out using OQL on cBioPortal**				

Table 13 formed the basis of the genetic dimension that will be integrated with known clinical features to build a predictive model capable of telling which TNBC patient will be a good candidate for the combination treatment of surgery

(preferably breast conserving) and radiotherapy.

4.4. Results from Machine Learning Analyses

In this section, three predictive models were built using various collection of features. The three models were built using clinical features only, genetic features only and lastly combination of clinical and genetic features, respectively.

4.4.1. Feature Selection

The feature selection operation is premised on the fact that earlier correlation analyses had already suggested that not all features are important, as such features were not significantly correlated to the outcome variable, which was defined as the VITAL_STATUS. The Boruta function for feature selection was therefore used to prune the predictor features, limiting the predictor set to only features that are very important. "Blue boxplots correspond to minimal, average, and maximum Z score of a shadow attribute. Red, yellow, and green boxplots represent Z scores of rejected, tentative and confirmed attributes, respectively".[141]

Model 1 was built using clinical features only. After the 23 predictor features were subjected to pruning, figure 19 shows that there were only 5 important variables, namely SURVIVAL_MONTHS, SURVIVAL_GROUPS, NPI, BREAST_SURGERY and TUMOR_SIZE.

Model 2 was built using genetic features only. After the 18 predictor features were subjected to pruning, figure 20 shows that there were only 5 variables deemed to

be important in building the model. They are AKT1, MAP3K1, MEN1, SHANK2 and NCOR2.

Model 3 was built, primarily using a combination of the 5 clinical features used to build Model 1 and the 5 genetic features used to build model 2. After feature selection operation was carried out using the 10 features, only the NCOR2 feature was dropped. Consequently, 9 were used to build model 3 as captures in figure 21. The 9 important variables used to build model 3 are AKT1, MAP3K1, MEN1, SHANK2, SURVIVAL_MONTHS, SURVIVAL_GROUPS, NPI, BREAST_SURGERY and TUMOR_SIZE.

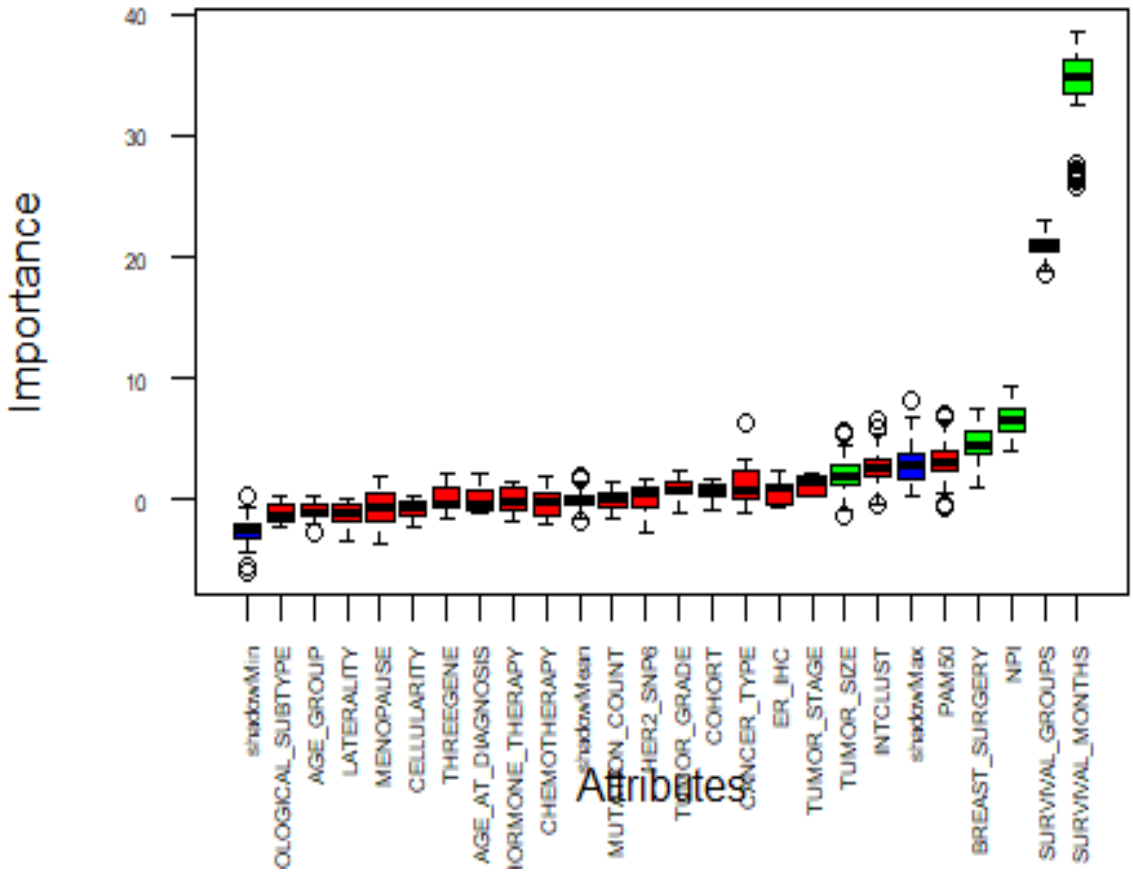


Figure 19 Feature selection plot showing confirmed (green) and rejected features (red) for the clinical features set used to build model 1.

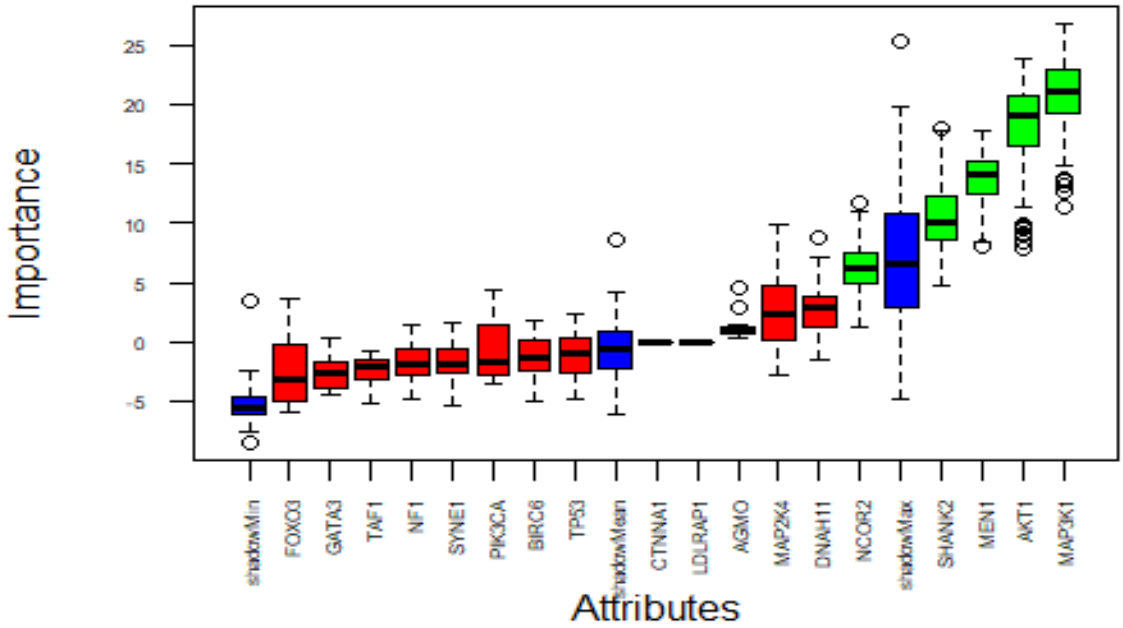


Figure 20 Feature selection plot showing confirmed (green) and rejected features (red) for the genetic features set used to build model 2.

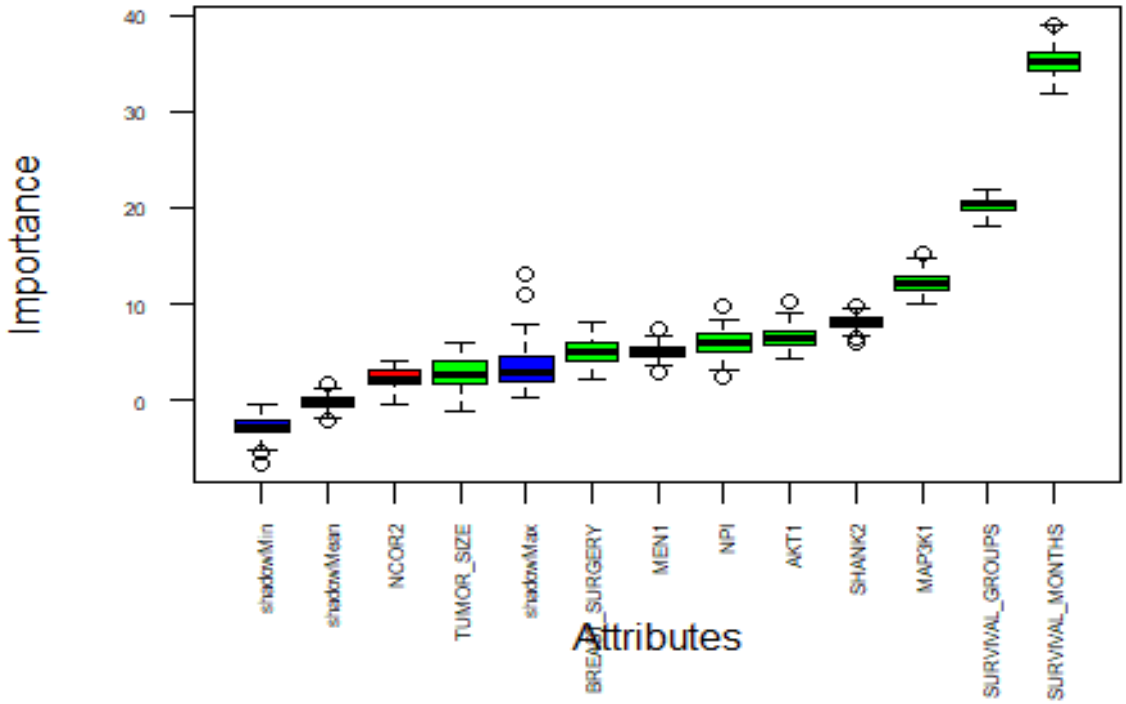


Figure 21 Feature selection plot showing confirmed (green) and rejected features (red) for the clinical + genetic features set used to build model 3.

4.4.2. Uni-Modal Visualizations

Histograms and density plots of all the predictor variables were plotted to have a feel of what the distribution looked like. The histograms are captured in figures 22 and 24, while the density plots are captured in figures 23 and 25. From figure 22 and 24, we can see that almost all the distributions have exponential or bimodal shapes. This should normally warrant the use of power transforms to flatten out the distributions.

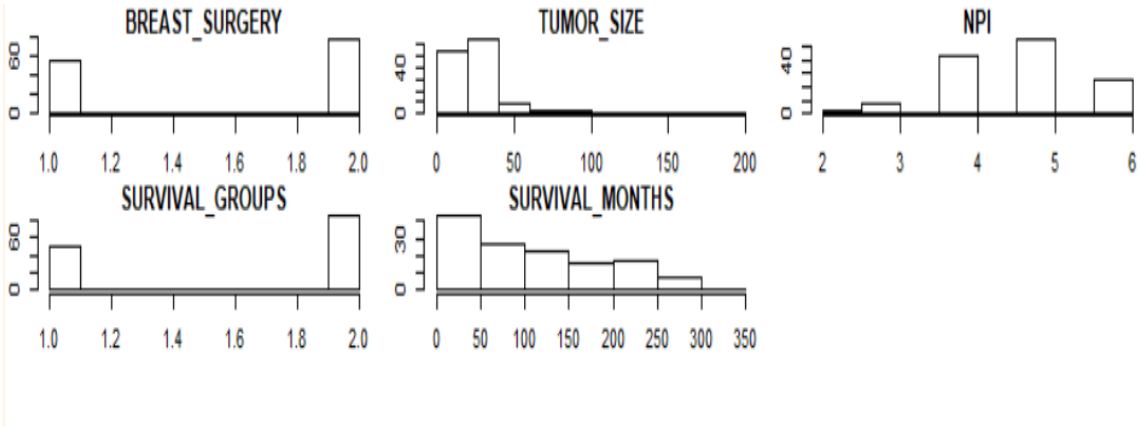


Figure 22 Plots showing histograms for all 5 clinical features confirmed after feature selection

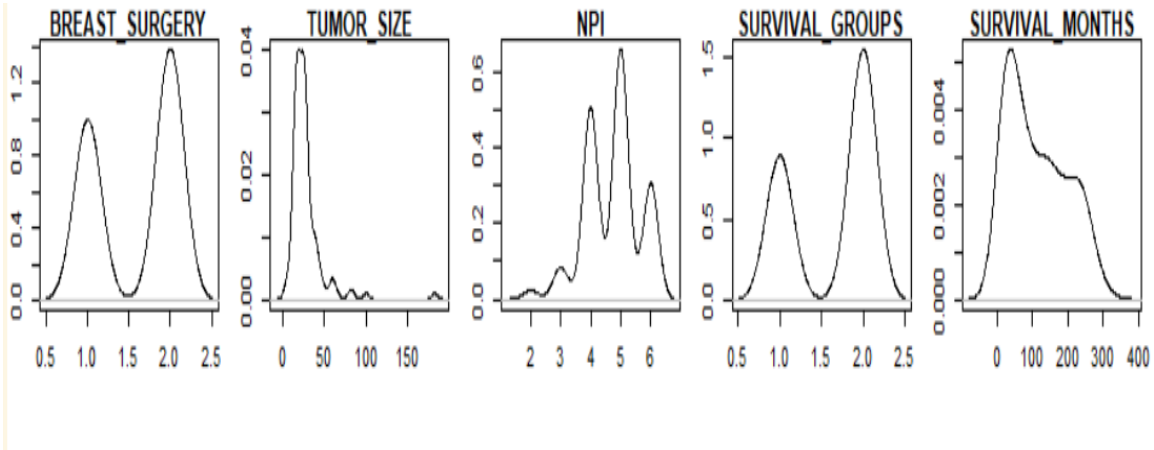


Figure 23 Plots showing the density curves for all 5 clinical features confirmed after feature selection

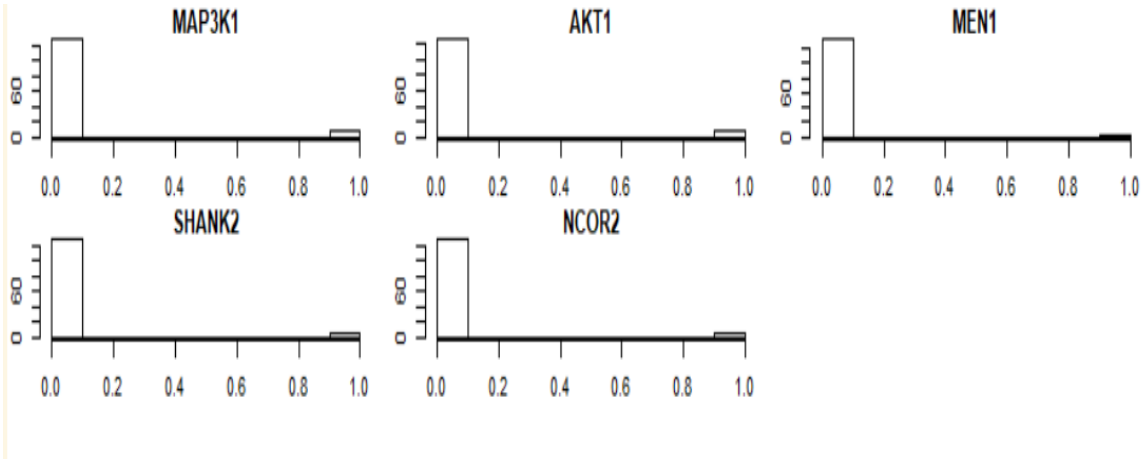


Figure 24 Plot showing the histograms for all 5 genetic features confirmed after feature selection

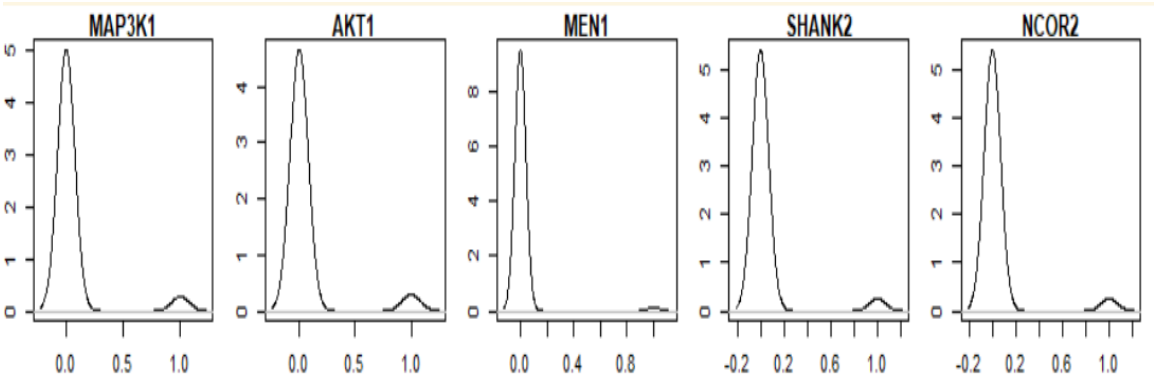


Figure 25 Plot showing the density curves for all the 5 genetic features confirmed after feature selection

4.4.3. Multi-Modal Visualizations

Further to the activity of section 4.4.1, scatterplots and bar plots of the predictor attributes were plotted to further gain more insights into the data distribution. The scatterplots are captured in figures 26, 27 and 28. The black dots

represents the living patients while the red dots represent the patients who died of the disease. Figure 28 shows many different associations simultaneously across all 9 predictor features.

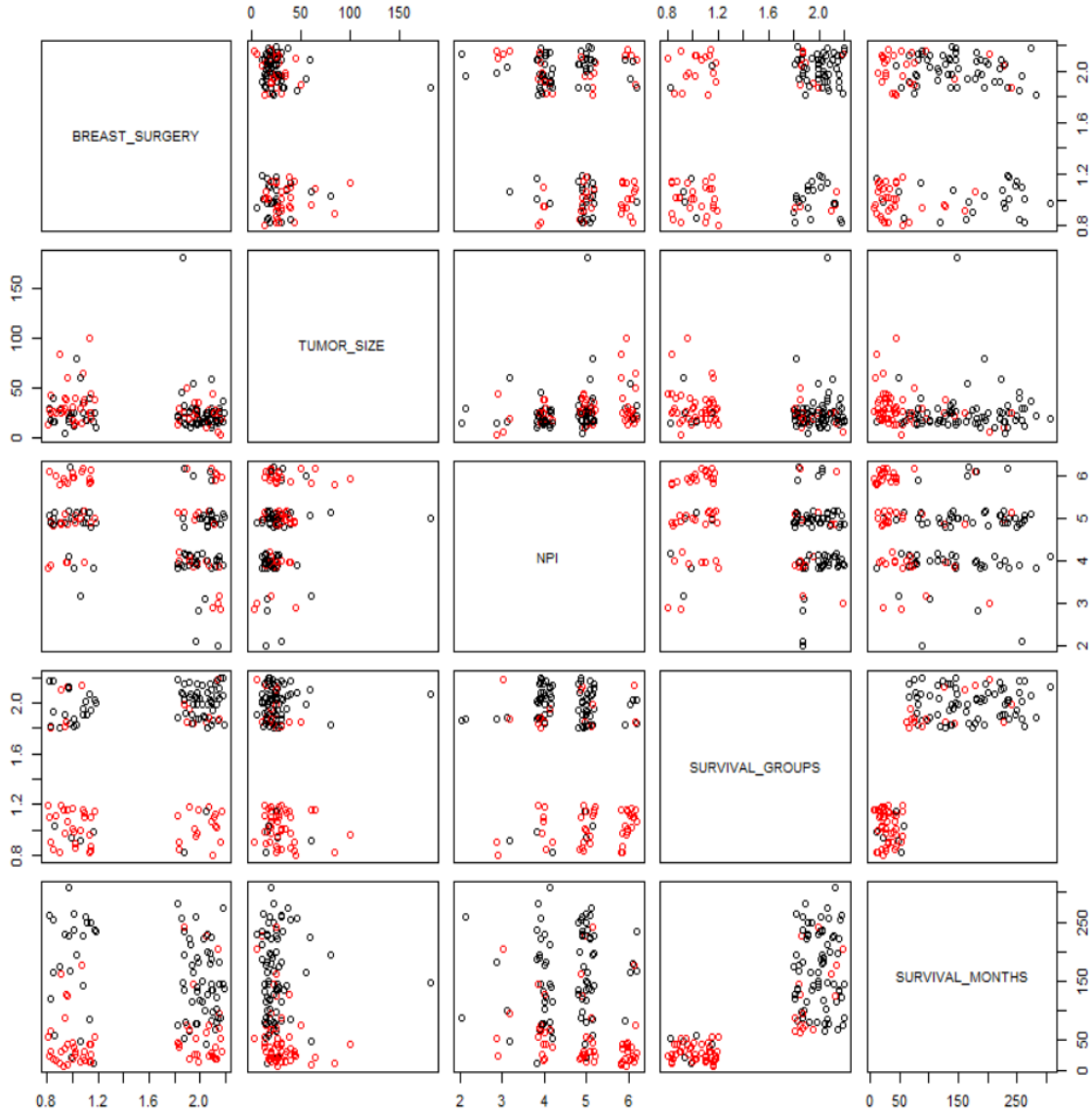


Figure 26 Scatterplots for the 5 predictor clinical features for model 1

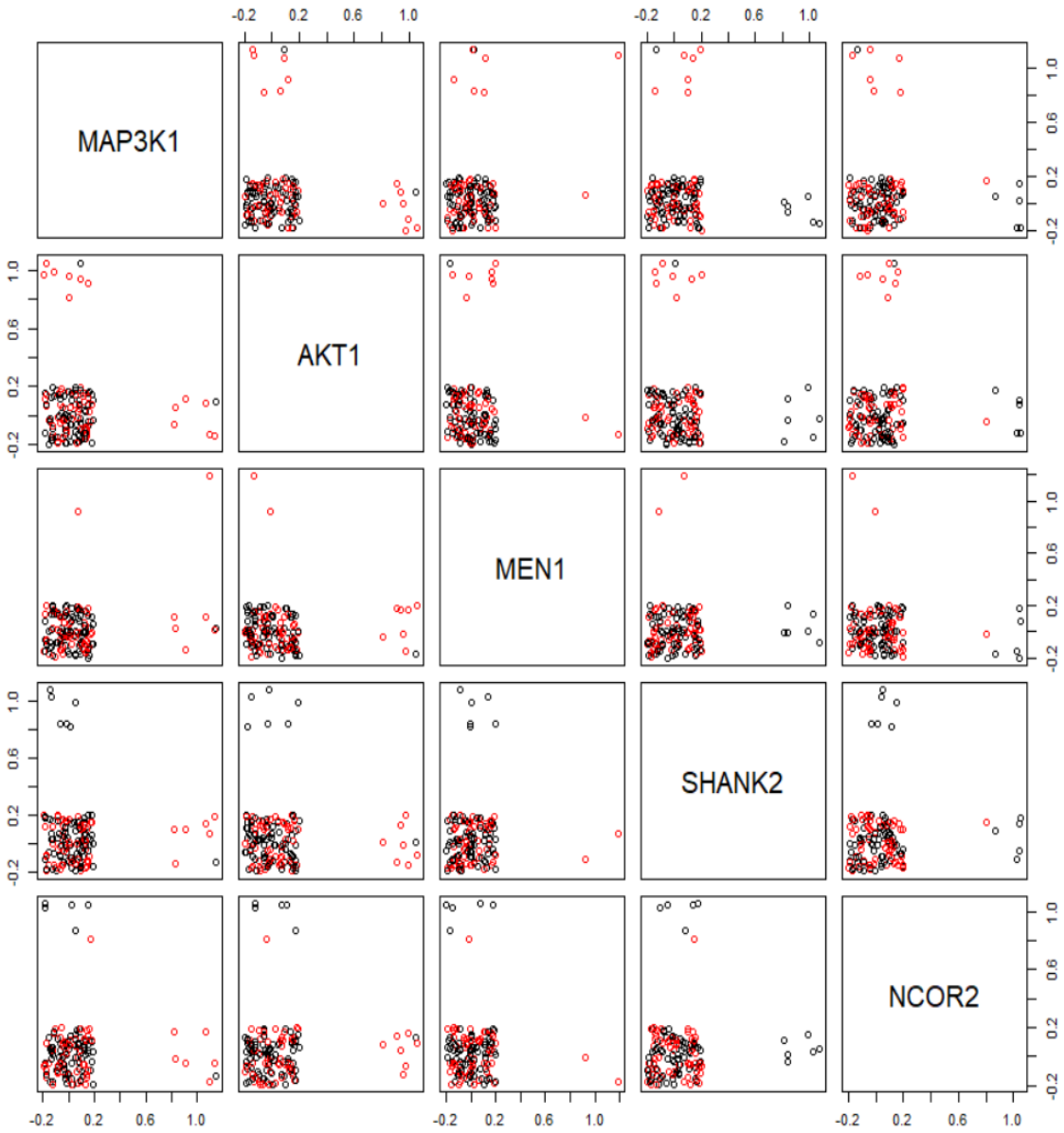


Figure 27 Scatterplots for the 5 genetic predictor features for model 2

Figure 28 Scatterplots for the 9 clinical + genetic predictor features for model 3

4.4.4. Model 1 - Evaluation of Machine Learning Algorithms Using Baseline Parameters

Model 1 makes use of 5 clinical features to make predictions. The dependent variable was the VITAL_STATUS feature, which specifies whether a patient died of the disease or is living. It is usually not so easy to determine which learning method will be ideal for a dataset from the onset. In view of this, five methods were explored cutting across linear, non-linear and ensemble methods. Figure 29 shows the output statistics for the 5 algorithms used and the corresponding graph showing accuracy and kappa plotting.

```
summary(results)

##
## Call:
## summary.resamples(object = results)
##
## Models: LG, LDA, KNN, CART, RF
## Number of resamples: 30
##
## Accuracy
##              Min.   1st Qu.    Median      Mean   3rd Qu. Max. NA's
## LG   0.6153846 0.7692308 0.8397436 0.8279304 0.8571429    1    0
## LDA  0.6428571 0.7692308 0.8461538 0.8352564 0.9065934    1    0
## KNN  0.6428571 0.7857143 0.8461538 0.8431929 0.8571429    1    0
## CART 0.7142857 0.7857143 0.8461538 0.8554249 0.9230769    1    0
## RF   0.6428571 0.7548077 0.7857143 0.7958791 0.8461538    1    0
##
## Kappa
##              Min.   1st Qu.    Median      Mean   3rd Qu. Max. NA's
## LG   0.2168675 0.5301205 0.6596452 0.6401427 0.7170833    1    0
## LDA  0.2258065 0.5301205 0.6829268 0.6541901 0.8125301    1    0
## KNN  0.2553191 0.5531915 0.6867015 0.6747001 0.7170833    1    0
## CART 0.4166667 0.5714286 0.6904762 0.7002191 0.8470588    1    0
## RF   0.2258065 0.4577546 0.5471840 0.5747909 0.6885889    1    0

dotplot(results)
```

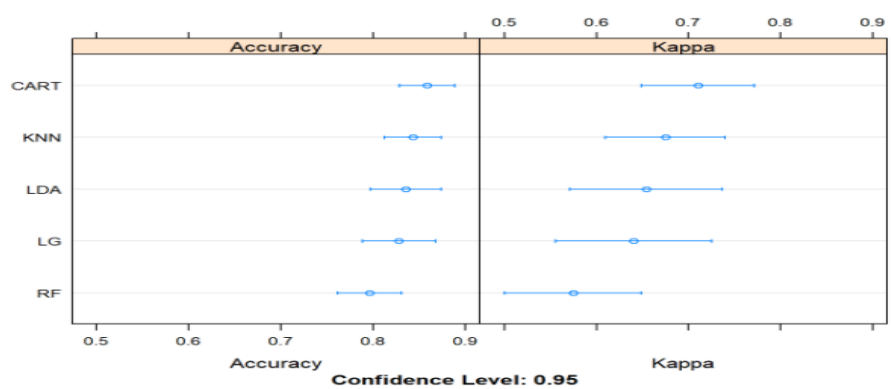


Figure 29: Baseline output analysis of 5 learning algorithms showing accuracy and kappa values

4.4.5. Fine-tuning the Best Three Machine Learning Algorithms

From section 4.4.3, the top performing algorithms using accuracy and kappa statistics are represented in table 14. It was discovered that the BoxCox Transform had no significant effect on the CART algorithm which maintained similar values for both baseline and transform operations. It was also discovered that the quality of performance for KNN was worsened by the transform operation.

Table 114: Top performing algorithms

	Algorithm	**Accuracy (%)**	**Kappa (%)**	**Operation**
1	CART	85.54	70.02	Baseline
2	KNN	84.31	67.47	Baseline

Consequently, KNN was subjected to tuning as captured in Figure 30, to see if there will be improvements. Table 15 shows its performance after tuning. Note that tuning worsened the CART algorithm.

Table 12: Performance of KNN algorithm before and after tuning

	Before Tuning		**After Tuning**	
Algorithm	**Accuracy (%)**	**Kappa (%)**	**Accuracy (%)**	**Kappa (%)**
KNN	84.31	67.47	85.56	70.23

From the analysis captured in Table 15, it can be deduced that the best algorithm is the KNN after the tuning operation.

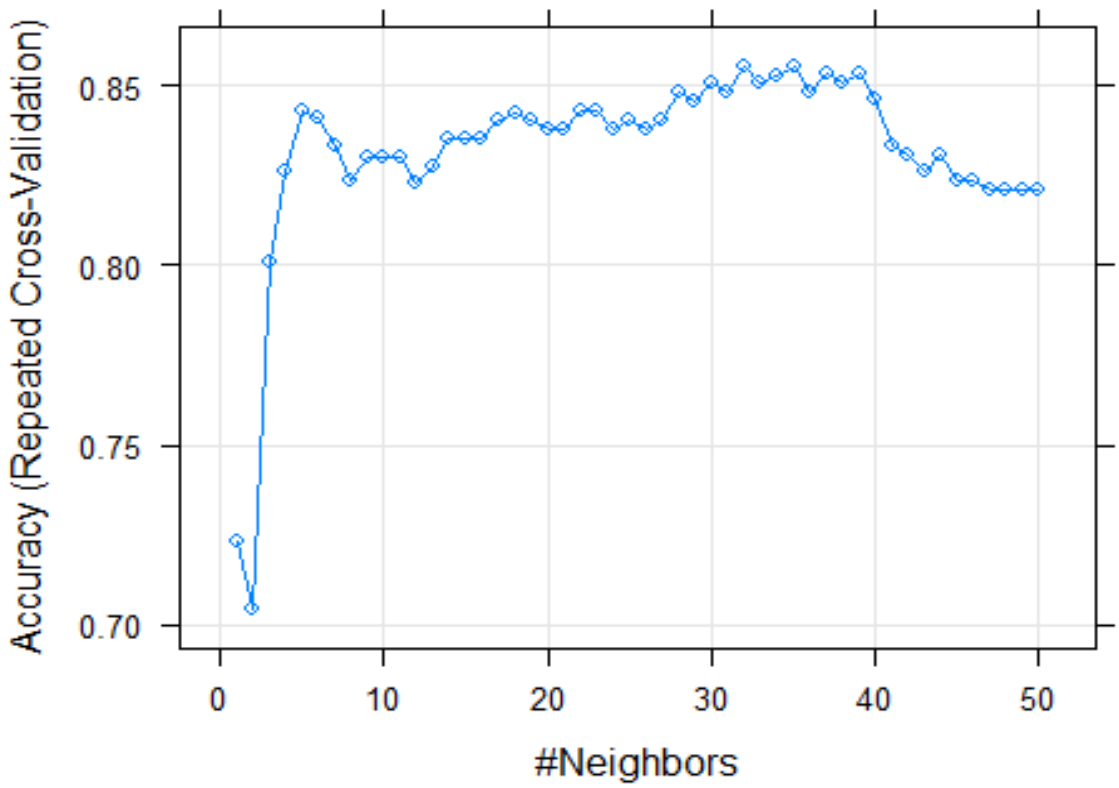


Figure 30: Graph showing tuning operation for KNN

4.4.6. Finalize and Build Model 1 Using KNN

Having established KNN as the machine learning algorithm of choice, model 1 is built with the confusion matrix, capturing relevant statistics of interest. The confusion matrix operation captures important metrics such as model accuracy, confidence interval, kappa, sensitivity, specificity, and detection prevalence. These and other statistics are captured in Figure 31. Additionally, Figure 32 presents an Area under Curve value of 93.75% which suggests that the 5 predictors used to build the model portrays a high degree of importance given the very high AUC value derived.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 32  3
##          1  0 21
##
##                Accuracy : 0.9464
##                  95% CI : (0.8513, 0.9888)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 3.093e-10
##
##                   Kappa : 0.8889
##
##  Mcnemar's Test P-Value : 0.2482
##
##             Sensitivity : 1.0000
##             Specificity : 0.8750
##          Pos Pred Value : 0.9143
##          Neg Pred Value : 1.0000
##              Prevalence : 0.5714
##          Detection Rate : 0.5714
##    Detection Prevalence : 0.6250
##       Balanced Accuracy : 0.9375
##
##        'Positive' Class : 0
##
```

Figure 31: Output statistics for the confusion matrix of Model 1

Furthermore, Table 16 shows the various guidelines by three different proponents on how to assess the kappa statistic. From Figure 31, it can be deduced that the kappa score associated with the model performance is 0.8889 (88.89%). Consequently, we can infer from Table 16 that by all standards, the performance is close to perfect.

Table 13: Guideline for assessing kappa statistic[143-145]

	Landis & Koch	**Cicchetti & Sparrow**	**Fleiss, Levin & Paik**
< 0.01 (<1%)	Poor	Poor	Poor
0.00 – 0.20 (0 – 20%)	Slight	--	--
0.21 – 0.40 (21 – 40%)	Fair	--	--
0.41 – 0.60 (41 – 60%)	Moderate	Fair	Fair to Good
0.61 – 0.75 (61 – 75%)	Substantial	Excellent	--
0.76 – 0.80 (76 – 80%)	--	--	Excellent
0.81 – 1.00 (81 – 100%)	Almost Perfect	--	--

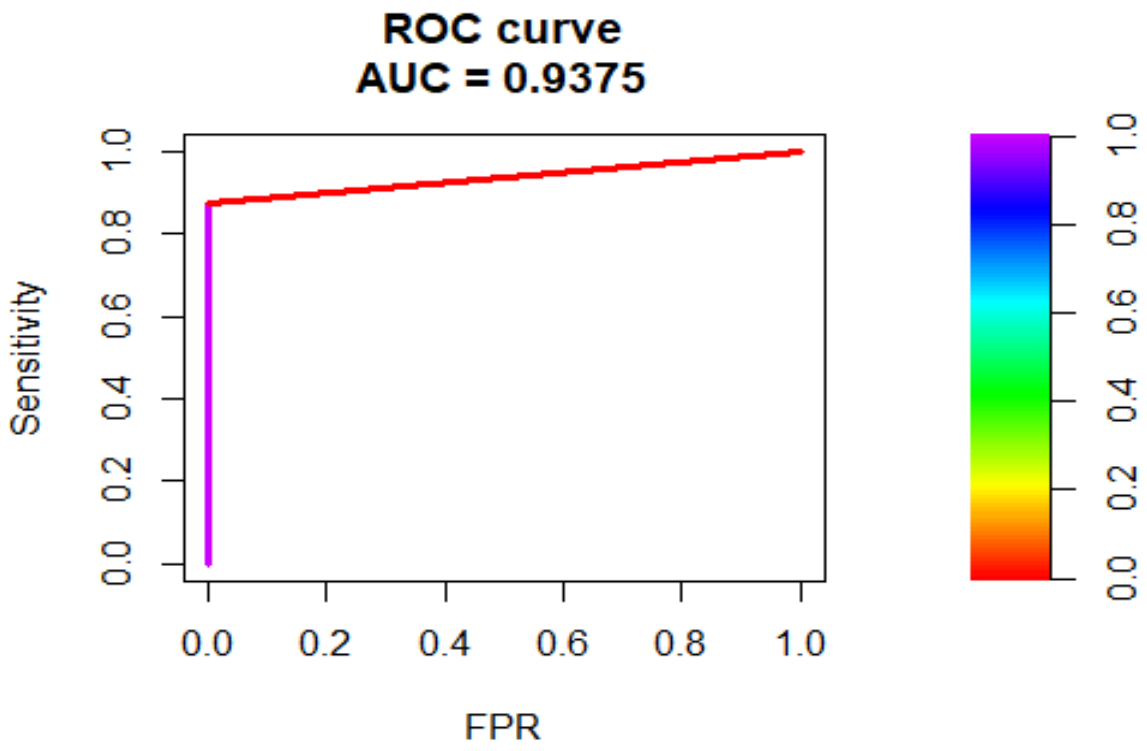


Figure 32: Graph showing the ROC Curve (Area Under Curve) for Model 1

In the subsequent sub-sections, two more models will be built and comparatively analyzed.

4.4.6 Model 2 – Predictive Model Using Genetic (Mutation) Features Only

Model 2 makes use of 5 genetic features to make predictions. The dependent variable was the VITAL_STATUS feature, which specifies whether a patient died of the disease or is living. The 5 predictor features pruned down from an initial set of 18 genes through feature selection are AKT1, MAP3K1, MEN1, SHANK2 and NCOR2 as captured in figure 20.

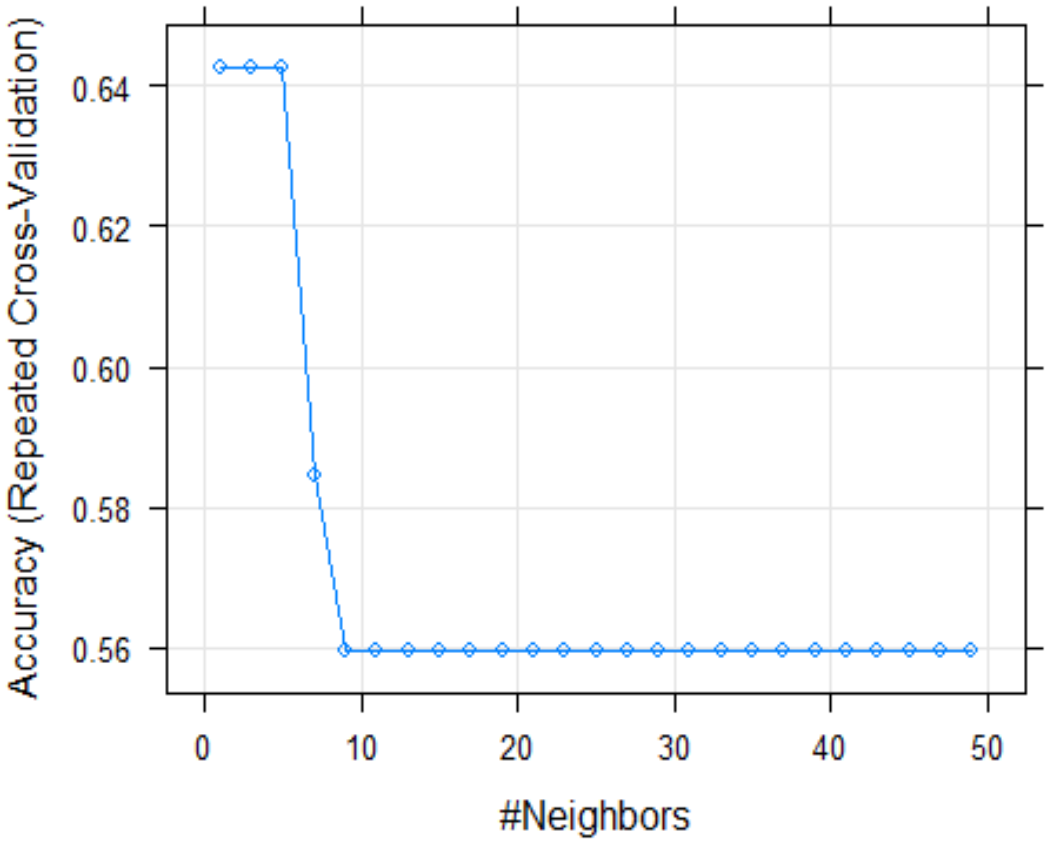


Figure 33 KNN tuning operation on model 2

Figure 33 shows that the tuning operation carried out on the KNN algorithm failed to improve the model. Figure 34 shows the confusion matrix for Model 2 which depicts a poor overall performance of the model when compared to Model 1. Model 2 displayed an accuracy of 62.5%. Figure 35 shows the ROC Curve (Area Under Curve) for Model 2 with a poor value of 0.5625 or 56.25%.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 32 21
##          1  0  3
##
##                Accuracy : 0.625
##                  95% CI : (0.4855, 0.7508)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 0.2511
##
##                   Kappa : 0.1404
##
##  Mcnemar's Test P-Value : 1.275e-05
##
##             Sensitivity : 1.0000
##             Specificity : 0.1250
##          Pos Pred Value : 0.6038
##          Neg Pred Value : 1.0000
##              Prevalence : 0.5714
##          Detection Rate : 0.5714
##    Detection Prevalence : 0.9464
##       Balanced Accuracy : 0.5625
##
##        'Positive' Class : 0
##
```

Figure 34 Output statistics for the confusion matrix of Model 2

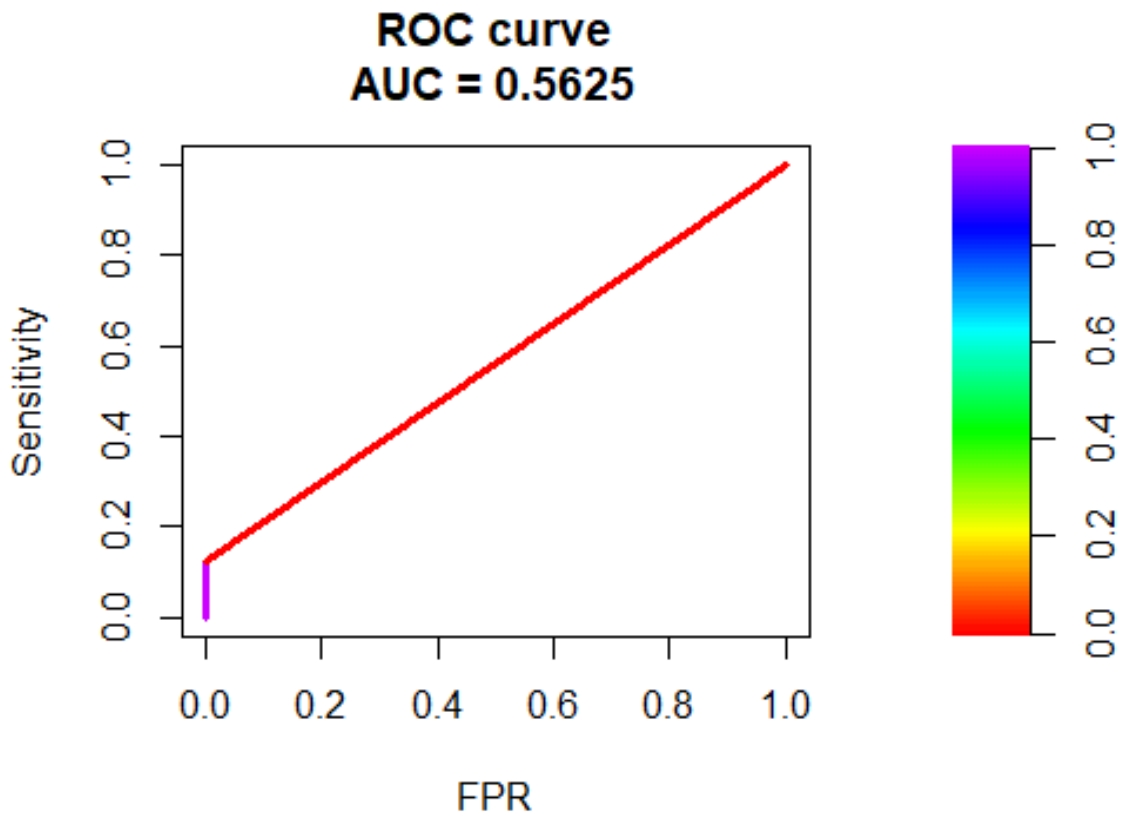


Figure 35 Graph showing the ROC Curve (Area Under Curve) for Model 2

4.4.7. Model 3 – Predictive Model Using Clinical and Genetic Features

Model 3 is based on the combination of clinical and genetic features. Initially, the 5 clinical predictors from model 1 and the 5 genetic predictors from model 2 were combined to produce 10 predictor features. These 10 predictors were subjected to feature selection. After this operation, the NCOR2 genetic feature was termed unimportant. This therefore reduced the number of predictor features for model 3 from 10 to 9 predictor features as captured in figure 21.

Figure 36 shows the confusion matrix statistics while figure 37 shows the ROC curve for model 3.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 32  3
##          1  0 21
##
##                Accuracy : 0.9464
##                  95% CI : (0.8513, 0.9888)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 3.093e-10
##
##                   Kappa : 0.8889
##
##  Mcnemar's Test P-Value : 0.2482
##
##             Sensitivity : 1.0000
##             Specificity : 0.8750
##          Pos Pred Value : 0.9143
##          Neg Pred Value : 1.0000
##              Prevalence : 0.5714
##          Detection Rate : 0.5714
##    Detection Prevalence : 0.6250
##       Balanced Accuracy : 0.9375
##
##        'Positive' Class : 0
##
```

Figure 36 Output statistics for the confusion matrix of Model 3

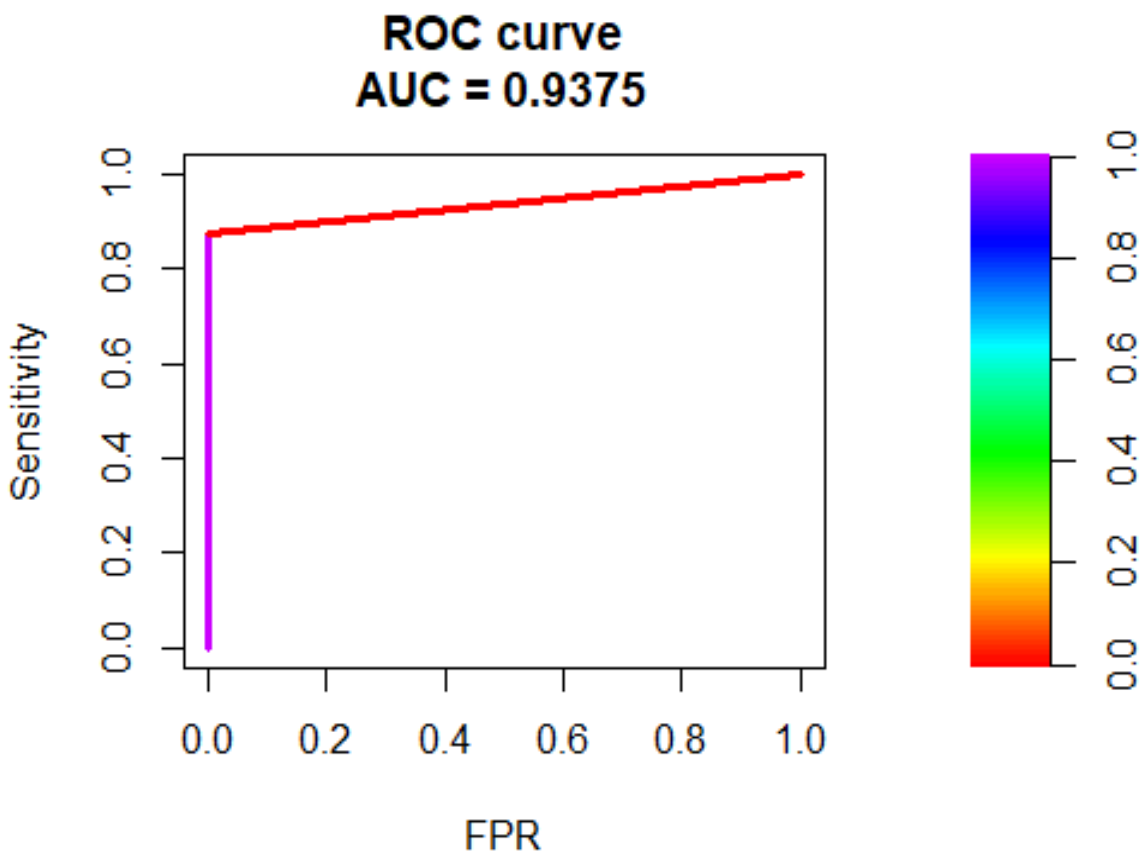


Figure 37 Graph showing the ROC Curve (Area Under Curve) for Model 3

Table 14 Performance Summary of Predictive Models

	Clinical Features	**Genetic Features**	**Clinical + Genetic Features**
Parameters	**Model 1**	**Model 2**	**Model 3**
Predictors	5	5	9
Accuracy	94.64%	62.50%	94.64%
Kappa	88.89%	14.04%	88.89%
Sensitivity	100%	100%	100%
Specificity	87.50%	12.50%	87.50%
AUC(ROC)	93.75%	56.25%	93.75%

Table 14 shows that the model efficiencies derivable from the mechanism of feature selection which has systematically reduced the predictor variables from 41 to less than 10. The three models showed varying degrees of accuracies. Model 3 validates the proposition that the combination of both mutation (genetic) and clinical features as the predictors used to build the predictive model will help determine which post-surgery TNBC patients would benefit from radiotherapy.

CHAPTER 5

DISCUSSIONS

From the analysis and results obtained, post-surgery TNBC patients have a good chance of survival if they are administered with adjuvant radiotherapy. Considering the established fact derived from numerous literatures, that the first 5 years after diagnosis is critical to the survival or demise of TNBC patients, most of the patients who lived beyond this threshold timeline underwent the combination treatment of surgery and radiotherapy. Interestingly, those who underwent breast conserving as the surgical procedure of choice had a better outcome than those who underwent mastectomy before the administration of radiotherapy.

Of the 190 patients that made up the overall sample size used mostly for analysis; 107 patients were still alive, while 83 had died of the disease. Furthermore, the 70-patient group had 64 patients that died during the 5-year period under review for that sub-group. Since the 70-patient group is a subset of the 190-patient overall sample, it can be inferred that approximately 77% of total deaths occurred in the 70-sample group. We can therefore confirm that the 5-year treatment period post-diagnosis is very critical in the life of a TNBC patient. Data suggests that those who make it past the 5-year benchmark stand a good chance of surviving in the long term.

In addition, from the 190 TNBC patient group that underwent radiotherapy treatment, there were about 355 genes of statistical significance ($p < 0.05$ – Fisher's Exact Test) with copy number alterations, mostly amplification and a few deletions. The top 50

genes captured in Table 9 were derived from this collection of 355 genes based on statistical significance.

- The prevalent loci where these alterations occurred were 17q23.1/2/3 (52 genes – enriched in only deceased patients); 19q13.43 (59 genes – enriched in only living patients); 17q22 (44 genes – enriched in only deceased patients); 16p22.2 (23 genes – enriched in only deceased patients) and 14q32.33 (20 genes – enriched in only deceased patients).

From the 50 CNA genes, the three loci – 17q23.1/2; 17q22 and 19q13.43 account for more than 80% of the altered genes.

- There is a need for future studies to investigate the gene interactions in 17q23.1/2; 17q22 in relation to the deceased patients as well as the interactions in 19q13.43 in relation to the living patients.

In addition, the best predictive model was built using 9 features that were carefully selected based on feature selection operations that pruned 41 potential predictor features to just 9 features. It comprised of 4 genetic and 5 clinical features respectively of statistical significance. The importance of this is that these collection of genes and clinical features can be used to profile TNBC patients to ascertain how well they will respond to radiotherapy as a treatment option to be administered after undergoing breast surgery, preferably the breast conserving procedure.

CHAPTER 6

SUMMARY AND CONCLUSIONS

The analysis carried out in this research establishes the importance of surgery as part of combination treatment for TNBC patients if they are to survive beyond 5 years after diagnosis. More so, it was inferred that adjuvant radiotherapy improves survival in post-surgery TNBC patients, especially where breast conserving procedure was adopted.

Given the report by the American Cancer Society that the median age for breast cancer diagnosis for white women is 62 years and 59 years for black women respectively[142], it can be inferred from this study that there is a prevalence of TNBC occurrence in women younger that the reported median age by approximately 10yrs onset difference when compared to the general population of breast cancer patients. This is because the median age at the time of diagnosis recorded in this study is 52.12 years.

In addition, genetic analysis suggests that CN alterations in chromosomes 17 and 19 might have a role to play in the drug interactions that lead to the survival status of TNBC patients. The findings pave the way for further analysis to establish any possible correlations between the CN alteration activities and the expressed genes.

Furthermore, six genes: MAP3K1, SHANK2, AKT1, MEN1, PIK3CA and NCOR2 had mutations with significant levels based on Fisher's Exact Test. All six are somatic mutations. In addition, the PAM50 Classification revealed another set of 13 genes with mutations of significant levels. They are: MAP2K4, AGMO, TAF1, GATA3, PIK3CA, BIRC6, DNAH11, NF1, TP53, SYNE1, CTNNA1, FOXO3 and LDLRAP1. Of these 13 genes only PIK3CA was part of the initial 6. Of the 13, all are somatic except TAF1 which

is a germline mutation. This brought the gene set to 18 genes. These 18 genes when queried in cBioportal were altered in 85% of 190 patients/samples predominantly used for this study. Surprisingly, neither BRCA1 nor BRCA2 played a part in the analysis because they were not statistically relevant to be factored into the workings of this research. Analyses on the 18 genes showed that only 4 of these genes showed a high level of importance. They are AKT1, MAP3K1, MEN1 and SHANK2. They were therefore combined with 5 other clinical features which were used to build the adopted model.

Interestingly, among the 70-patient subgroup which had 64 deaths, all patients who died of the disease, had no mutation in the SHANK2 gene. However, a few mutations were recorded among the 120-patient group.

Having built three models in the course of this study, model 3 exhibited an accuracy of ~95%, kappa statistic of ~89%, sensitivity of 100% and specificity of 87.5% and ROC(AUC) value of ~94%, it can be deduced that the model is a robust one with parameters that suggest reliability. Hence, model 3 is adopted as the model of choice. Given the reliability of model 3, we can therefore conclude that the following 9 features are important in ascertaining which TNBC patients will benefit from post-surgery radiotherapy treatment. The features are as follows:

Clinical Features: Breast Surgery, NPI, Tumor Size, Survival Months and Survival Groups

Mutation Features: AKT1, MAP3K1, MEN1 and SHANK2

Finally, for a very heterogenous disease (see figure 28) that proves difficult to treat and very aggressive compared to other types of cancer, the 4 biomarkers and clinical features established in the predictive model can be used as a reference point in the treatment

care plan of TNBC patients to determine those who will benefit from adjuvant radiotherapy. With this, those who will not benefit from post-surgery radiotherapy can be recommended for alternative treatment plans rather than subjecting them to a treatment course that is bound to fail.

References

1. Qu, Xinglong, Ying Tang, and Shucheng Hua. "Immunological approaches towards cancer and inflammation: a cross talk." *Frontiers in immunology* 9 (2018): 563.
2. Walsh, Michael F., et al. "Genetic Factors: Hereditary Cancer Predisposition Syndromes." *Abeloff's Clinical Oncology*. Content Repository Only!, 2020. 180-208.
3. Prabhakar, Uma, Yalia Jayalakshmi, and C. Katherine Wang. "Cancer Diagnostics and Therapeutics." *Nanotheranostics for Cancer Applications*. Springer, Cham, 2019. 33-66.
4. World Health Organization. Breast Cancer. Retrieved December 16, 2019, from World Health Organization: https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/ (2019).
5. Harris, Randall E. *Epidemiology of chronic disease: global perspectives*. Jones & Bartlett Learning, 2019.
6. Alluri, Prasanna, and Lisa A. Newman. "Basal-like and triple-negative breast cancers: searching for positives among many negatives." *Surgical Oncology Clinics* 23.3 (2014): 567-577.
7. Gerratana, L., et al. "Androgen receptor in triple negative breast cancer: a potential target for the targetless subtype." *Cancer treatment reviews* 68 (2018): 102-110.
8. Kumar, Pankaj, and Rupali Aggarwal. "An overview of triple-negative breast cancer." *Archives of gynecology and obstetrics* 293.2 (2016): 247-269.
9. Bianchini, Giampaolo, et al. "Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease." *Nature reviews Clinical oncology* 13.11 (2016): 674.
10. Nagle, Peter W., et al. "Patient-derived tumor organoids for prediction of cancer treatment response." *Seminars in cancer biology*. Vol. 53. Academic Press, 2018.

11. Lesterhuis, W. Joost, et al. "Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity." *Nature Reviews Drug Discovery* 16.4 (2017): 264.

12. Hauser, Kevin, et al. "Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations." *Communications biology* 1.1 (2018): 1-14.

13. Stulpinas, Aurimas, et al. "Recent Approaches Encompassing the Phenotypic Cell Heterogeneity for Anticancer Drug Efficacy Evaluation." *Tumor Progression and Metastasis*. IntechOpen, 2019.

14. Ginsburg, Geoffrey S., et al., eds. *Genomic and Precision Medicine: Infectious and Inflammatory Disease*. Academic Press, 2019.

15. Mughal, Tariq. *Precision Haematological Cancer Medicine*. CRC Press, 2018.

16. Triple Negative Breast Cancer Foundation. Understanding Triple Negative Breast Cancer. Retrieved January 14, 2018, from Triple Negative Breast Cancer Foundation Web site: https://tnbcfoundation.org/understanding-triple-negative-breast-cancer/(2017)

17. Christian, Caglevic, et al. "Triple-negative breast cancer: the reality in Chile and in Latin America." *ecancermedicalscience* 13 (2019).

18. Yedjou, Clement G., et al. "Assessing the racial and ethnic disparities in breast cancer mortality in the United States." *International journal of environmental research and public health* 14.5 (2017): 486.

19. Gray, Janet M., et al. "State of the evidence 2017: an update on the connection between breast cancer and the environment." *Environmental Health* 16.1 (2017): 94.

20. Gomez-Flores-Ramos, Liliana, et al. "Breast cancer genetics in young women: What do we know?." *Mutation Research/Reviews in Mutation Research* 774 (2017): 33-45.

21. Garrido-Castro, Ana C., Nancy U. Lin, and Kornelia Polyak. "Insights into molecular classifications of triple-negative breast cancer: improving patient selection for treatment." *Cancer discovery* 9.2 (2019): 176-198.

22. Bray, Freddie, et al. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." *CA: a cancer journal for clinicians* 68.6 (2018): 394-424.

23. Ferlay, J., et al. "Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018." *European Journal of Cancer* 103 (2018): 356-387.

24. Jung, Kyu-Won, et al. "Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2015." *Cancer research and treatment: official journal of Korean Cancer Association* 50.2 (2018): 303.

25. Torre, Lindsey A., et al. "Global cancer in women: burden and trends." (2017): 444-457.

26. Cao, Bochen, et al. "Effect on longevity of one-third reduction in premature mortality from non-communicable diseases by 2030: a global analysis of the Sustainable Development Goal health target." *The Lancet Global Health* 6.12 (2018): e1288-e1296.

27. López-Suárez, Alejandro. "Burden of cancer attributable to obesity, type 2 diabetes and associated risk factors." *Metabolism* 92 (2019): 136-146.

28. Springmann, Marco, et al. "Mitigation potential and global health impacts from emissions pricing of food commodities." *Nature Climate Change* 7.1 (2017): 69-74.

29. World Health Organization - Cancer. Media Centre. Retrieved March 14, 2020 from World Health Organization: https://www.who.int/news-room/fact-sheets/detail/cancer (2018).

30. Akram, Muhammad, et al. "Awareness and current knowledge of breast cancer." *Biological research* 50.1 (2017): 33.

31. Hoffbrand, A. Victor, and David P. Steensma. *Hoffbrand's essential haematology*. John Wiley & Sons, 2019.

32. National Breast Cancer Foundation. What Is Cancer? Retrieved January 14, 2018, from National Breast Cancer Foundation website: http://www.nationalbreastcancer.org/what-is-cancer (2016).

33. Haskul, M., and Emine Yaman. "Comparison of Different Machine Learning Algorithms for Breast Cancer Recurrence Classification." *Southeast Europe Journal of Soft Computing* 8.2 (2019).

34. Tanay, Mary Anne Lagmay. "Medications Used for Cancer." *Understanding Pharmacology in Nursing Practice*. Springer, Cham, 2020. 393-411.

35. Lukong, Kiven Erique, Yetunde Ogunbolude, and Jean Paul Kamdem. "Breast cancer in Africa: prevalence, treatment options, herbal medicines, and socioeconomic determinants." *Breast cancer research and treatment* 166.2 (2017): 351-365.

36. MNT Editorial Team. Cancer: What you need to know. Retrieved January 14, 2018, from Medical News Today: https://www.medicalnewstoday.com/info/cancer-oncology (2016)

37. Soteriou, Despina, and Yaron Fuchs. "A matter of life and death: stem cell survival in tissue regeneration and tumour formation." *Nature reviews cancer* 18.3 (2018): 187.

38. Singh, Rumani, Anthony Letai, and Kristopher Sarosiek. "Regulation of apoptosis in health and disease: the balancing act of BCL-2 family proteins." *Nature Reviews Molecular Cell Biology* 20.3 (2019): 175-193.

39. Davis, C. P. what is cancer. Retrieved January 14, 2018, from medicine net website: https://www.medicinenet.com/cancer/article.htm#what_is_cancer. (2016)

40. Lambert, Arthur W., Diwakar R. Pattabiraman, and Robert A. Weinberg. "Emerging biological principles of metastasis." *Cell* 168.4 (2017): 670-691.

41. Sontheimer-Phelps, Alexandra, Bryan A. Hassell, and Donald E. Ingber. "Modelling cancer in microfluidic human organs-on-chips." *Nature Reviews Cancer* 19.2 (2019): 65-81.

42. Lloyd, Mark C., Robert A. Gatenby, and Joel S. Brown. "Ecology of the Metastatic Process." *Ecology and Evolution of Cancer*. Academic Press, 2017. 153-165.

43. Er, Ekrem Emrah, et al. "Pericyte-like spreading by disseminated cancer cells activates YAP and MRTF for metastatic colonization." *Nature cell biology* 20.8 (2018): 966-978.

44. Viswanathan, Vasanthi S., et al. "Dependency of a therapy-resistant state of cancer cells on a lipid peroxidase pathway." *Nature* 547.7664 (2017): 453-457.

45. Stockfleth, Eggert, and Lutz Schmitz. "Malignant Epithelial Tumors." *Braun-Falco´ s Dermatology* (2020): 1-24.

46. Yuan, Jilong, et al. "Defect Repair After Resection of the Upper Limb Malignant Tumor." *Oncoplastic surgery*. Springer, Singapore, 2018. 433-453.

47. Letfullin, Renat R., and Thomas F. George. "Introduction to Cancer Cells and Targeting." *Computational Nanomedicine and Nanotechnology*. Springer, Cham, 2016. 63-131.

48. Adler, Elizabeth M. *Living with Lymphoma: A Patient's Guide*. JHU Press, 2016.

49. Kumar, Shaji K., et al. "Multiple myeloma, version 3.2017, NCCN clinical practice guidelines in oncology." *Journal of the National Comprehensive Cancer Network* 15.2 (2017): 230-269.

50. National Cancer Institute. Understanding cancer. Retrieved January 14, 2018, from National Cancer Institute website: https://www.cancer.gov/about-cancer/understanding/what-is-cancer (2015).

51. Swope, Viki B., and Zalfa A. Abdel-Malek. "MC1R: Front and center in the bright side of dark eumelanin and DNA repair." *International journal of molecular sciences* 19.9 (2018): 2667.

52. Hoadley, Katherine A., et al. "Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer." *Cell* 173.2 (2018): 291-304.

53. America Cancer Society: Cancer Facts and Figures 2019. Atlanta, GA: American Cancer Society, 2019. https://www.cancer.gov/types/common-cancers. Last accessed February 5, 2019.

54. Wan, Bo Angela, et al. "Not All Lumps in Male Breast Referred to the Cancer Centre are Cancerous: A Single Institution Review." *Annals of Breast Cancer and Therapy* 1.1 (2017).

55. Dawson, Rachel S., and Rosa Cervantes. "Delayed menarche with normal pubertal growth." *Contemporary Pediatrics* 34.7 (2017): 15-21.

56. National Cancer Institute. Symptoms of Cancer. Retrieved January 14, 2018, from National Cancer Institute Web site: https://www.cancer.gov/about-cancer/diagnosis-staging/symptoms (2015)

57. Glynne-Jones, R., et al. "Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." *Annals of Oncology* 28 (2017): iv22-iv40.

58. Cardoso, F., et al. "Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." *Annals of Oncology* 30.8 (2019): 1194-1220.

59. Evans, Julie, et al. "GPs' understanding and practice of safety netting for potential cancer presentations: a qualitative study in primary care." *Br J Gen Pract* 68.672 (2018): e505-e511.

60. Cancer Treatment Centers of America. Breast Cancer Symptoms. Retrieved January 14, 2018, from Cancer Treatment Centers of America Web site: https://www.cancercenter.com/breast-cancer/symptoms/ (2017).

61. Wang, J. J., K. F. Lei, and F. Han. "Tumor microenvironment: recent advances in various cancer treatments." *European review for medical and pharmacological sciences* 22.12 (2018): 3855-3864.

62. Lehmann, Vicky, and Marrit A. Tuinman. " *Body Image Care for Cancer Patients: Principles and Practice* (2018): 81.

63. Sherman, S. "Differentiated thyroid cancer refractory to standard treatment: chemotherapy." *Up to Date. Accessed March* 8 (2018).

64. Luszczak, Sabina, et al. "PIM kinase inhibition: co-targeted therapeutic approaches in prostate cancer." *Signal Transduction and Targeted Therapy* 5.1 (2020): 1-10.

65. Miller, Kimberly D., et al. "Cancer treatment and survivorship statistics, 2019." *CA: a cancer journal for clinicians* 69.5 (2019): 363-385.

66. National Cancer Institute. Surgery to Treat Cancer. Retrieved January 14, 2018, from National Cancer Institute Web site: https://www.cancer.gov/about-cancer/treatment/types/surgery (2015).

67. Gootjes, Elske C., et al. "The value of tumour debulking for patients with extensive multi-organ metastatic colorectal cancer." *European Journal of Cancer* 103 (2018): 160-164.

68. National Cancer Institute. Chemotherapy to Treat Cancer. Retrieved January 14, 2018, from National Cancer Institute Website: https://www.cancer.gov/about-cancer/treatment/types/chemotherapy (2015)

69. Saraei, Pouya, et al. "The beneficial effects of metformin on cancer prevention and therapy: a comprehensive review of recent advances." *Cancer management and research* 11 (2019): 3295.

70. National Cancer Institute. Radiation Therapy to Treat Cancer. Retrieved January 14, 2018, from National Cancer Institute Web site: https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy (2017).

71. Baumann, Michael, et al. "Radiation oncology in the era of precision medicine." *Nature Reviews Cancer* 16.4 (2016): 234.

72. National Cancer Institute. Immunotherapy to Treat Cancer. Retrieved January 14, 2018, from National Cancer Institute Web site: https://www.cancer.gov/about-cancer/treatment/types/immunotherapy (2017).

73. Mayfield, Eleanor. "The Promise of Immunotherapy." *group* (2018).

74. National Cancer Institute. Hormone Therapy to Treat Cancer. Retrieved January 14, 2018, from National Cancer Institute Web site: https://www.cancer.gov/about-cancer/treatment/types/hormone-therapy (2015).

75. Patel, Hitisha K., and Teeru Bihani. "Selective estrogen receptor modulators (SERMs) and selective estrogen receptor degraders (SERDs) in cancer treatment." *Pharmacology & therapeutics* 186 (2018): 1-24.

76. National Cancer Institute. Targeted Therapy to Treat Cancer. Retrieved January 14, 2018, from National Cancer Institute Web site: https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies (2014).

77. Senapati, Sudipta, et al. "Controlled drug delivery vehicles for cancer treatment and their performance." *Signal transduction and targeted therapy* 3.1 (2018): 1-19.

78. Balentine, J. R. Breast Cancer. Retrieved January 14, 2018, from Medicine Net Web site: https://www.medicinenet.com/breast_cancer_facts_stages/article.htm (2017).

79. Feng, Yixiao, et al. "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis." *Genes & diseases* 5.2 (2018): 77-106.

80. Gregorowitsch, M. L., et al. "Severe depression more common in patients with ductal carcinoma in situ than early-stage invasive breast cancer patients." *Breast cancer research and treatment* 167.1 (2018): 205-213.

81. Apoe, Ogheneruona, et al. "Breast CanCer: UniqUe CharaCteristiCs as targets for therapy." *Cancer: Prevention, Early Detection, Treatment and Recovery* (2019): 227-243.

82. Roberts, Sophie, and Valerie Speirs. "Advances in the development of improved animal-free models for use in breast cancer biomedical research." *Biophysical reviews* 9.4 (2017): 321-327.

83. Guo, Rongrong, et al. "Ultrasound imaging technologies for breast cancer detection and management: A review." *Ultrasound in medicine & biology* 44.1 (2018): 37-70.

84. Matsuo, Mioko, Masazumi Tuneyoshi, and Mari Mine. "Primary mucinous carcinoma with rhabdoid cells of the thyroid gland: a case report." *Diagnostic pathology* 11.1 (2016): 48.

85. Fleisher, Brett, Charlotte Clarke, and Sihem Ait-Oudhia. "Current advances in biomarkers for targeted therapy in triple-negative breast cancer." *Breast Cancer: Targets and Therapy* 8 (2016): 183.

86. Pandit, Ashish, et al. "Breast Cancer Receptors and Targeting Strategies." *Targeted Intracellular Drug Delivery by Receptor Mediated Endocytosis*. Springer, Cham, 2019. 79-108.

87. Talbert, Patricia Y., and Marquise D. Frazier. "Inflammatory Breast Cancer Disease: A Literature Review." *cancer* 2.1/4 (2019).

88. Finkelstein, E., et al. "11.1 Anatomy and Physiology of the Nipple." *Gender and Dermatology* (2018): 139.

89. Pang, Wenyang, et al. "Adenoid cystic carcinoma of the breast in a male: A case report." *Medicine* 98.32 (2019).

90. Dania, Vanessa, et al. "Associations of race and ethnicity with risk of developing invasive breast cancer after lobular carcinoma in situ." *Breast Cancer Research* 21.1 (2019): 120.

91. Vidyadhar, T. U., and A. Suresh. "Metastatic Breast Cancer: Mini Review." *Ann Oncol Radiol* 1.1 (2019): 1003.

92. Saraiva, Diana P., et al. "How many diseases is triple negative breast cancer: the protagonism of the immune microenvironment." *ESMO open* 2.4 (2017): e000208.

93. Abramson VG, Lehmann BD, Ballinger TJ, Pietenpol JA. Subtyping of triplenegative breast cancer: implications for therapy. Cancer 2015; 121:8-16.

94. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 121, 2750–2767, https://doi.org/10.1172/JCI45014 (2011).

95. Lehmann BD, Pietenpol JA. Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes. J Pathol 2014; 232:142-50.

96. Oakman C, Viale G, Di Leo A. Management of triple negative breast cancer. Breast 2010; 19:312-21.

97. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California Cancer Registry. Cancer 2007; 109:1721-8.

98. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006; 295:2492.

99. Liedtke C, Hess KR, Karn T, et al. The prognostic impact of age in patients with triple-negative breast cancer. Breast Cancer Res Treat 2013; 138:591-9.

100. Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 2007; 13:4429-34.

101. Morris GJ, Mitchell EP. Higher incidence of aggressive breast cancers in African-American women: a review. J Natl Med Assoc 2008; 100:698-702.

102. Heitz F, Harter P, Lueck HJ, et al. Triple-negative and HER2-overexpressing breast cancers exhibit an elevated risk and an earlier occurrence of cerebral metastases. Eur J Cancer 2009; 45:2792-8.

103. Azzam, Heba, et al. "The value of dynamic contrast-enhanced MRI in differentiating triple-negative breast cancer from other subtypes." *Egyptian Journal of Radiology and Nuclear Medicine* 50.1 (2019): 1-9.

104. Balayla, Jacques, et al. "Outcomes of ovarian stimulation and fertility preservation in breast cancer patients with different hormonal receptor profiles." *Journal of Assisted Reproduction and Genetics* (2020): 1-9.

105. Al-Mahmood, Sumayah, et al. "Metastatic and triple-negative breast cancer: challenges and treatment options." *Drug delivery and translational research* 8.5 (2018): 1483-1507.

106. Bierie, Brian, et al. "Integrin-β4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells." *Proceedings of the National Academy of Sciences* 114.12 (2017): E2337-E2346.

107. Nasir, Sebahadin. *Automatic Breast Cancer Detection from Biopsy fine Needle Aspiration Microscopic Images*. Diss. Addis Ababa University, 2018.

108. Martin, L. J. Triple-negative Breast Cancer. Retrieved January 14, 2018, from WebMD Web site: https://www.webmd.com/breast-cancer/triple-negative-breast-cancer#2 (2017).

109. Denkert, Carsten, et al. "Molecular alterations in triple-negative breast cancer—the road to new treatment strategies." *The Lancet* 389.10087 (2017): 2430-2442.

110. Golubnitschaja, Olga, et al. "Medicine in the early twenty-first century: paradigm and anticipation-EPMA position paper 2016." *EPMA Journal* 7.1 (2016): 23.

111. Martinez-Garcia, Miguel Angel, et al. "Precision medicine in obstructive sleep apnoea." *The Lancet Respiratory Medicine* (2019).

112. Gong, Wang, et al. "Toward the use of precision medicine for the treatment of head and neck squamous cell carcinoma." *Oncotarget* 8.2 (2017): 2141.

113. Carrasco-Ramiro, F., R. Peiro-Pastor, and B. Aguado. "Human genomics projects and precision medicine." *Gene therapy* 24.9 (2017): 551-561.

114. Rebhan, Michael. "Towards a systems approach for chronic diseases, based on health state modeling." *F1000Research* 6 (2017).

115. Özdemir, Vural, et al. "Digging deeper into precision/personalized medicine: Cracking the sugar code, the third alphabet of life, and sociomateriality of the cell." *OMICS: A Journal of Integrative Biology* 24.2 (2020): 62-80.

116. Vegter, M. W. *Big & Intimate: Selfhood in Times of Precision Medicine*. Diss. [Sl: sn], 2019.

117. Beger, Richard D., et al. "Metabolomics enables precision medicine:"a white paper, community perspective"." *Metabolomics* 12.9 (2016): 149.

118. Cryer, Alexander M., and Andrew J. Thorley. "Nanotechnology in the diagnosis and treatment of lung cancer." *Pharmacology & therapeutics* 198 (2019): 189-205.

119. Robinson, Mark Dennis. *The Market in Mind: How Financialization is Shaping Neuroscience, Translational Medicine, and Innovation in Biotechnology*. MIT Press, 2019.

120. Canonica, Giorgio W., et al. "Asthma: personalized and precision medicine." *Current opinion in allergy and clinical immunology* 18.1 (2018): 51-58.

121. Wakai, Toshifumi, et al. "Next-generation sequencing-based clinical sequencing: toward precision medicine in solid tumors." *International journal of clinical oncology* 24.2 (2019): 115-122.

122. Roy, Ananda L., and Richard S. Conroy. "Toward mapping the human body at a cellular resolution." *Molecular biology of the cell* 29.15 (2018): 1779-1785.

123. Doll, Richard. *The Prevention of Cancer: Pointers from Epidemiology*. Routledge, 2017.

124. Stewart, Bernard W. "The Mechanistic Basis of Cancer Prevention." *Translational Toxicology and Therapeutics: Windows of Developmental Susceptibility in Reproduction and Cancer* (2017): 147-170.

125. McJohn, Stephen M., and Ian McJohn. "Fair Use and Machine Learning." *Northeastern University Law Review, Forthcoming* (2019).

126. Durstewitz, Daniel, Georgia Koppe, and Andreas Meyer-Lindenberg. "Deep neural networks in psychiatry." *Molecular psychiatry* 24.11 (2019): 1583-1598.

127. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

128. Géron, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.

129. Barzegar, Rahim, et al. "Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms." *Science of the total environment* 621 (2018): 697-712.

130. Chang, Spencer, Timothy Cohen, and Bryan Ostdiek. "What is the machine learning?." *Physical Review D* 97.5 (2018): 056009.

131. Oakden-Rayner, Luke, et al. "Precision radiology: predicting longevity using feature engineering and deep learning methods in a radiomics framework." *Scientific reports* 7.1 (2017): 1-13.

132. Allen, Bibb, Robert Gish, and Keith Dreyer. "The Role of an Artificial Intelligence Ecosystem in Radiology." *Artificial Intelligence in Medical Imaging*. Springer, Cham, 2019. 291-327.

133. Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016). https://www.cbioportal.org/study/summary?id=brca_metabric

134. Pereira, Bernard, et al. "The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes." *Nature communications* 7.1 (2016): 1-16.

135. Wei, Leyi, et al. "A novel hierarchical selective ensemble classifier with bioinformatics application." *Artificial intelligence in medicine* 83 (2017): 82-90.

136. Sherry, Alexander D., et al. "Systemic Inflammation after Radiation Predicts Locoregional Recurrence, Progression, and Mortality in Stage II-III Triple-Negative Breast Cancer." *International Journal of Radiation Oncology* Biology* Physics* (2019).

137. Mi, Huaiyu, et al. "Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v. 14.0)." *Nature protocols* 14.3 (2019): 703-721.

138. Quost, Benjamin, Thierry Denœux, and Shoumei Li. "Parametric classification with soft labels using the evidential EM algorithm: linear discriminant analysis versus logistic regression." *Advances in Data Analysis and Classification* 11.4 (2017): 659-690.

139. Veena, K. M., K. Manjula Shenoy, and KB Ajitha Shenoy. "Performance comparison of machine learning classification algorithms." *International Conference on Advances in Computing and Data Sciences*. Springer, Singapore, 2018.

140. Chen, Wei, et al. "Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping." *Journal of Hydrology* 583 (2020): 124602.

141. Caraka, Rezzy Eko, et al. "Feature importance of the aortic anatomy on endovascular aneurysm repair (EVAR) using Boruta and Bayesian MCMC." Commun. Math. Biol. Neurosci. 2020 (2020): Article-ID.

142. American Cancer Society. Breast Cancer Facts & Figures 2017-2018. Atlanta: American Cancer Society, Inc. 2017 – (https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf)]

143. Landis, J. Richard, and Gary G. Koch. "The measurement of observer agreement for categorical data." *biometrics* (1977): 159-174.

144. Cicchetti, Domenic V., and Sara A. Sparrow. "Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior." *American journal of mental deficiency* (1981).

145. Fleiss, J. L., B. Levin, and M. C. Paik. "Statistical methods for rates and proportions. John Wiley & Sons." *New York* 870 (1981).

Appendices

Doctoral Dissertation - Clinical Associations and Genetic Alterations to Predict Radiotherapy Treatment Response in Patients with Triple Negative Breast Cancer [R Codes and Data Visualizations for Methods 1, 2 and 4]

Ernest Onuiri

8/13/2020

1.0. Method 1 - Exploratory Data Analysis on the METABRIC TNBC Dataset

Steps(enumerated using regular numbers)/Observations (enumerated using bullet points)

- 1.1. Confirm that the right working directory is set.
- 1.2. Load datasets containing both clinical and genetic mutation data respectively from working directory.
- 1.3. View variable (feature/column) names.
- 1.4. Filter the dataset by removing the following ten variables: PATIENT.ID, CANCER_TYPE, ER_STATUS,
- PR_STATUS, HER2_STATUS, ONCOTREE_CODE, and SAMPLE_TYPE.

```
getwd()

## [1] "C:/Users/ernes/OneDrive - Rutgers University/PhD/DISSERTATION/Thesis/New Datasets/metabric/210 Normalize
d expression data_discovery set/Experiment"

TNBC_clinical = read.csv("tnbc_clinical.csv") # clinical dataset
TNBC_mut = read.csv("tnbc_mutation.csv") # Mutation dataset
names(TNBC_clinical)

##  [1] "PATIENT.ID"                "AGE_AT_DIAGNOSIS"
##  [3] "BREAST_SURGERY"            "CANCER_TYPE"
##  [5] "CANCER_TYPE_DETAILED"      "CELLULARITY"
##  [7] "CHEMOTHERAPY"              "RADIO_THERAPY"
##  [9] "HORMONE_THERAPY"           "CLAUDIN_SUBTYPE"
## [11] "COHORT"                    "ER_IHC"
## [13] "ER_STATUS"                 "PR_STATUS"
## [15] "HER2_STATUS"               "GRADE"
## [17] "HER2_SNP6"                 "MUTATION_COUNT"
## [19] "HISTOLOGICAL_SUBTYPE"      "INFERRED_MENOPAUSAL_STATE"
## [21] "INTCLUST"                  "LATERALITY"
## [23] "NPI"                       "ONCOTREE_CODE"
## [25] "OS_MONTHS"                 "SAMPLE_TYPE"
## [27] "THREEGENE"                 "TUMOR_SIZE"
## [29] "TUMOR_STAGE"               "VITAL_STATUS"

names(TNBC_mut)

##  [1] "PATIENT.ID" "AKT1"       "MAP3K1"     "MEN1"       "NCOR2"
##  [6] "PIK3CA"     "SHANK2"     "MAP2K4"     "AGMO"       "TAF1"
## [11] "GATA3"      "BIRC6"      "DNAH11"     "NF1"        "TP53"
## [16] "SYNE1"      "CTNNA1"     "FOXO3"      "LDLRAP1"

TNBC_clinical = TNBC_clinical[,-c(1,4,13:15,24,26)]
```

- 1.5. Load dplyr and rename some variable names using the function dplyr::rename().
- 1.6. View column names, structure and Summary of sorted data for insight.
- 1.7. Display the TNBC_clinical dataset.

```
library(dplyr) # A grammar for data manipulation
```

```
## Warning: package 'dplyr' was built under R version 3.6.3

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

TNBC_clinical = TNBC_clinical %>%
  rename(
    CANCER_TYPE = CANCER_TYPE_DETAILED,
    PAM50 = CLAUDIN_SUBTYPE,
    MENOPAUSE = INFERRED_MENOPAUSAL_STATE,
    TUMOR_GRADE = GRADE,
    SURVIVAL_MONTHS = OS_MONTHS
    )
```

- 1.8. Bin the "age_at_diagnosis" into buckets
- 1.9. Segregate the age groups into 3 intervals. Summary statistics showed that median age at diagnosis is 52.12
- 1.10. Add a new variable AGE_GROUP with the bucket identifier

```
TNBC_clinical$AGE_GROUP = cut(TNBC_clinical$AGE_AT_DIAGNOSIS, breaks = c(0,40,52, 100),
                  labels = c("Below 40", "40 to 52", "Above 52"))
```

- 1.11. Bin the "SURVIVAL_MONTHS" into buckets. 60 = 60months which is the equivalent of 5 years
- 1.12. Segregate the age groups into 2 intervals
- 1.13. Add a new variable SURVIVAL_GROUPS with the bucket identifier

```
TNBC_clinical$SURVIVAL_GROUPS = cut(TNBC_clinical$SURVIVAL_MONTHS, breaks = c(0,60,320),
                      labels = c("5yrs or Less", "Above 5yrs"))
```

- 1.14. Reorder variables

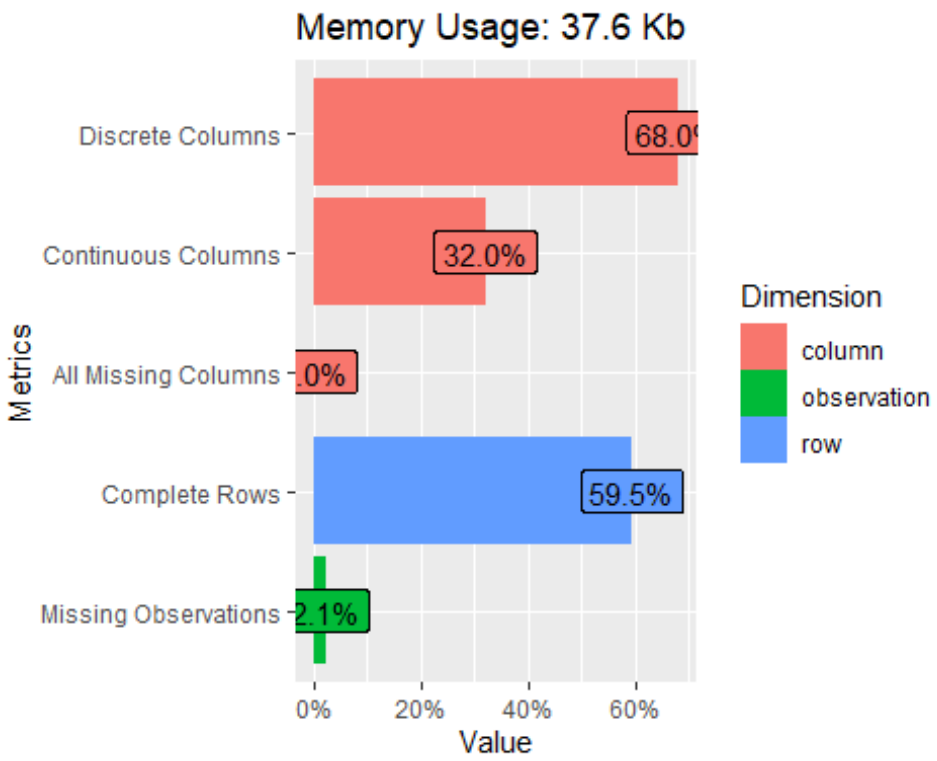
```
col_order = c("AGE_AT_DIAGNOSIS","AGE_GROUP","BREAST_SURGERY","CANCER_TYPE","CHEMOTHERAPY","RADIO_THERAPY",
              "HORMONE_THERAPY","PAM50","COHORT","ER_IHC","TUMOR_GRADE","TUMOR_SIZE","TUMOR_STAGE",
             "HER2_SNP6","HISTOLOGICAL_SUBTYPE","CELLULARITY","MENOPAUSE","INTCLUST","LATERALITY",
             "MUTATION_COUNT","NPI","SURVIVAL_MONTHS", "SURVIVAL_GROUPS","THREEGENE","VITAL_STATUS")
TNBC_clinical = TNBC_clinical[, col_order]
names(TNBC_clinical)

##  [1] "AGE_AT_DIAGNOSIS"     "AGE_GROUP"            "BREAST_SURGERY"
##  [4] "CANCER_TYPE"          "CHEMOTHERAPY"         "RADIO_THERAPY"
##  [7] "HORMONE_THERAPY"      "PAM50"                "COHORT"
## [10] "ER_IHC"               "TUMOR_GRADE"          "TUMOR_SIZE"
## [13] "TUMOR_STAGE"          "HER2_SNP6"            "HISTOLOGICAL_SUBTYPE"
## [16] "CELLULARITY"          "MENOPAUSE"            "INTCLUST"
## [19] "LATERALITY"           "MUTATION_COUNT"       "NPI"
## [22] "SURVIVAL_MONTHS"      "SURVIVAL_GROUPS"      "THREEGENE"
## [25] "VITAL_STATUS"
```

- 1.15. Check for missing values (NA's) in the records using the summary function.
- 1.16. Use DataExplorer & mice packages to further analyze missing records and fix values.
- 1.17. Using mice; m=5 is the number of times model should run, maxit: Max number of iterations.
- 1.18. Confirm the methods (algorithms) that was imposed by the mice function.
- 1.19. Input generated data for missing values (NA's) and create dataset with complete

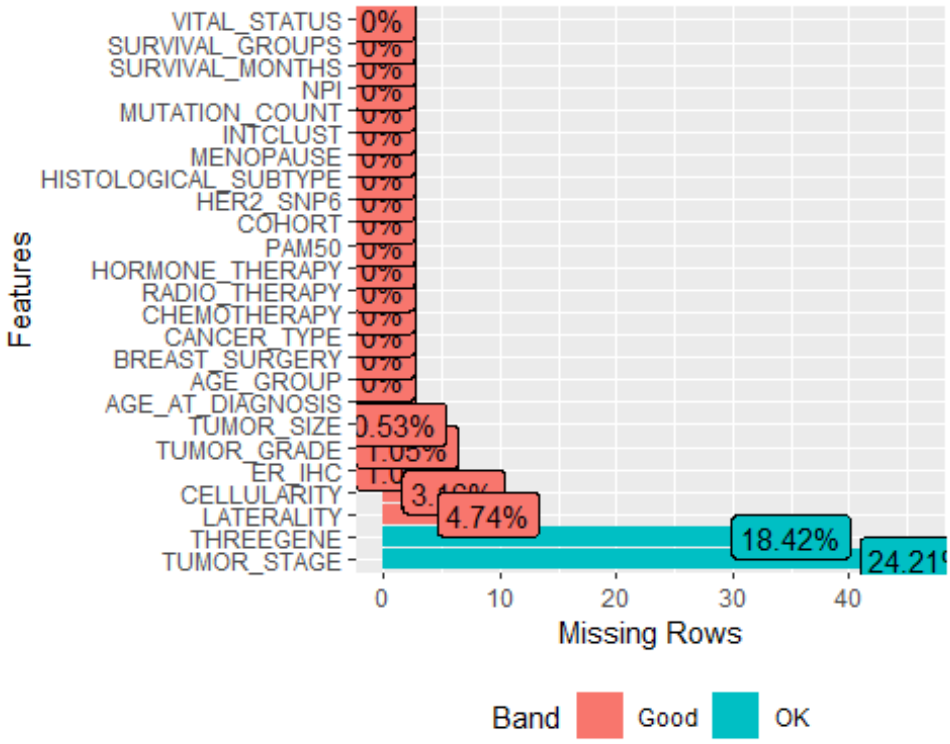
```
•       entries. In addition, view and summarize the dataset.

summary(TNBC_clinical)

##  AGE_AT_DIAGNOSIS    AGE_GROUP            BREAST_SURGERY
##  Min.   :26.72    Below 40:37   BREAST CONSERVING:113
##  1st Qu.:43.05    40 to 52:56   MASTECTOMY       : 77
##  Median :52.12    Above 52:97
##  Mean   :52.44
##  3rd Qu.:61.09
##  Max.   :96.29
##
##                                     CANCER_TYPE  CHEMOTHERAPY RADIO_THERAPY
##  Breast Invasive Ductal Carcinoma         :159   NO : 60      YES:190
```

```
##  Breast Invasive Lobular Carcinoma        :  5   YES:130
##  Breast Mixed Ductal and Lobular Carcinoma:  2
##  Invasive Breast Carcinoma                : 24
##
##
##
##  HORMONE_THERAPY         PAM50         COHORT       ER_IHC     TUMOR_GRADE
##  NO :136         Basal      :100   Min.   :1.000   neg :162   Min.   :1.000
##  YES: 54         claudin-low: 65   1st Qu.:1.000   pos : 26   1st Qu.:3.000
##                  Her2       : 17   Median :3.000   NA's:  2   Median :3.000
##                  LumA       :  1   Mean   :2.521              Mean   :2.888
##                  Normal     :  7   3rd Qu.:3.000              3rd Qu.:3.000
##                                     Max.   :5.000              Max.   :3.000
##                                                                NA's   :2
##    TUMOR_SIZE      TUMOR_STAGE    HER2_SNP6       HISTOLOGICAL_SUBTYPE
##  Min.   :  3.00   Min.   :0.000   GAIN: 19   IDC            :164
##  1st Qu.: 18.00   1st Qu.:1.750   LOSS: 12   IDC-MED        : 13
##  Median : 25.00   Median :2.000   NEUT:159   ILC            :  6
##  Mean   : 28.26   Mean   :1.889              IDC+ILC        :  2
##  3rd Qu.: 30.00   3rd Qu.:2.000              INVASIVE TUMOUR:  2
##  Max.   :182.00   Max.   :3.000              DCIS           :  1
##  NA's   :1        NA's   :46                 (Other)        :  2
##    CELLULARITY  MENOPAUSE     INTCLUST   LATERALITY MUTATION_COUNT
##  high    :108   post:108   10     :120   l   :91    Min.   : 0.000
##  low     : 18   pre : 82   4ER-   : 27   r   :90    1st Qu.: 3.000
##  moderate: 58              4ER+   : 12   NA's: 9    Median : 5.000
##  NA's    :  6              9      : 11              Mean   : 5.484
##                            1      :  8              3rd Qu.: 7.000
##                            3      :  5              Max.   :30.000
##                            (Other):  7
##       NPI        SURVIVAL_MONTHS       SURVIVAL_GROUPS
##  Min.   :1.050   Min.   :  4.167   5yrs or Less: 70
##  1st Qu.:4.042   1st Qu.: 35.375   Above 5yrs  :120
##  Median :5.034   Median : 92.350
##  Mean   :4.730   Mean   :113.128
##  3rd Qu.:5.080   3rd Qu.:182.883
##  Max.   :6.320   Max.   :307.933
##
##                  THREEGENE            VITAL_STATUS
##  ER-/HER2-            :145   Died of Disease: 83
##  ER+/HER2- High Prolif:  5   Living         :107
##  ER+/HER2- Low Prolif :  3
##  HER2+                :  2
##  NA's                 : 35
##
##
```

```
library(DataExplorer) # automate visual exploration of data and treatment

## Warning: package 'DataExplorer' was built under R version 3.6.3

plot_intro(TNBC_clinical)
```

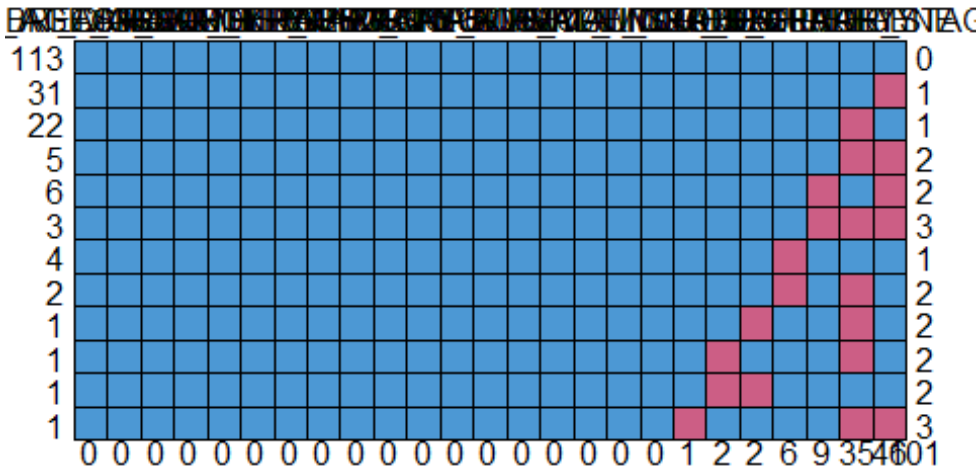



```
plot_missing(TNBC_clinical)
```




```
library(mice) # Multivariate Imputation by Chained Equations

## Warning: package 'mice' was built under R version 3.6.3
```

```
##
## Attaching package: 'mice'

## The following objects are masked from 'package:base':
##
##     cbind, rbind

md.pattern(TNBC_clinical)
```




```
##     AGE_AT_DIAGNOSIS AGE_GROUP BREAST_SURGERY CANCER_TYPE CHEMOTHERAPY
## 113                1         1              1           1            1
## 31                 1         1              1           1            1
## 22                 1         1              1           1            1
## 5                  1         1              1           1            1
## 6                  1         1              1           1            1
## 3                  1         1              1           1            1
## 4                  1         1              1           1            1
## 2                  1         1              1           1            1
## 1                  1         1              1           1            1
## 1                  1         1              1           1            1
## 1                  1         1              1           1            1
## 1                  1         1              1           1            1
##                    0         0              0           0            0
##     RADIO_THERAPY HORMONE_THERAPY PAM50 COHORT HER2_SNP6 HISTOLOGICAL_SUBTYPE
## 113             1               1     1      1         1                    1
## 31              1               1     1      1         1                    1
## 22              1               1     1      1         1                    1
## 5               1               1     1      1         1                    1
## 6               1               1     1      1         1                    1
## 3               1               1     1      1         1                    1
## 4               1               1     1      1         1                    1
## 2               1               1     1      1         1                    1
## 1               1               1     1      1         1                    1
## 1               1               1     1      1         1                    1
## 1               1               1     1      1         1                    1
## 1               1               1     1      1         1                    1
##                 0               0     0      0         0                    0
##     MENOPAUSE INTCLUST MUTATION_COUNT NPI SURVIVAL_MONTHS SURVIVAL_GROUPS
## 113         1        1              1   1               1               1
## 31          1        1              1   1               1               1
## 22          1        1              1   1               1               1
```

```
## 5           1        1              1   1               1               1
## 6           1        1              1   1               1               1
## 3           1        1              1   1               1               1
## 4           1        1              1   1               1               1
## 2           1        1              1   1               1               1
## 1           1        1              1   1               1               1
## 1           1        1              1   1               1               1
## 1           1        1              1   1               1               1
## 1           1        1              1   1               1               1
##             0        0              0   0               0               0
##     VITAL_STATUS TUMOR_SIZE ER_IHC TUMOR_GRADE CELLULARITY LATERALITY THREEGENE
## 113            1          1      1           1           1          1         1
## 31             1          1      1           1           1          1         1
## 22             1          1      1           1           1          1         0
## 5              1          1      1           1           1          1         0
## 6              1          1      1           1           1          0         1
## 3              1          1      1           1           1          0         0
## 4              1          1      1           1           0          1         1
## 2              1          1      1           1           0          1         0
## 1              1          1      1           0           1          1         0
## 1              1          1      0           1           1          1         0
## 1              1          1      0           0           1          1         1
## 1              1          0      1           1           1          1         0
##                0          1      2           2           6          9        35
##     TUMOR_STAGE
## 113           1   0
## 31            0   1
## 22            1   1
## 5             0   2
## 6             0   2
## 3             0   3
## 4             1   1
## 2             1   2
## 1             1   2
## 1             1   2
## 1             1   2
## 1             0   3
##              46 101
```

```
mice_imputes = mice(TNBC_clinical, m = 5, maxit = 5, seed = 10)
```

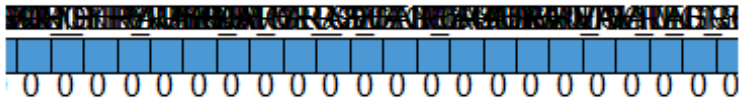
```
##
##  iter imp variable
##   1   1  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   1   2  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   1   3  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   1   4  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   1   5  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   2   1  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   2   2  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   2   3  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   2   4  ER_IHC  TUMOR_GRADE*  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   2   5  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   3   1  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   3   2  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   3   3  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   3   4  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   3   5  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   4   1  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   4   2  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   4   3  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   4   4  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   4   5  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   5   1  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   5   2  ER_IHC  TUMOR_GRADE*  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   5   3  ER_IHC  TUMOR_GRADE*  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   5   4  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##   5   5  ER_IHC  TUMOR_GRADE  TUMOR_SIZE  TUMOR_STAGE  CELLULARITY  LATERALITY  THREEGENE
##  * Please inspect the loggedEvents
```

```
## Warning: Number of logged events: 179
```

```
mice_imputes$method
```

```
##     AGE_AT_DIAGNOSIS            AGE_GROUP       BREAST_SURGERY
##                   ""                   ""                   ""
##          CANCER_TYPE         CHEMOTHERAPY        RADIO_THERAPY
```

```
##                    ""                    ""                    ""
##      HORMONE_THERAPY                PAM50                COHORT
##                    ""                    ""                    ""
##               ER_IHC           TUMOR_GRADE            TUMOR_SIZE
##             "logreg"                 "pmm"                 "pmm"
##          TUMOR_STAGE             HER2_SNP6 HISTOLOGICAL_SUBTYPE
##                 "pmm"                    ""                    ""
##          CELLULARITY             MENOPAUSE              INTCLUST
##            "polyreg"                    ""                    ""
##           LATERALITY        MUTATION_COUNT                   NPI
##             "logreg"                    ""                    ""
##      SURVIVAL_MONTHS       SURVIVAL_GROUPS             THREEGENE
##                    ""                    ""             "polyreg"
##         VITAL_STATUS
##                    ""

TNBC_clinical = complete(mice_imputes, 5) #TNBC_clinical dataset to be used for bi-variate analysis
clin = complete(mice_imputes, 5) #clin dataset to be used for uni-variate analysis

md.pattern(TNBC_clinical)

##  /\     /\
## {  `---'  }
## {  O   O  }
## ==>  V <==  No need for mice. This data set is completely observed.
##  \  \|/  /
##   `-----'
```



```
##     AGE_AT_DIAGNOSIS AGE_GROUP BREAST_SURGERY CANCER_TYPE CHEMOTHERAPY
## 190                1         1              1           1            1
##                    0         0              0           0            0
##     RADIO_THERAPY HORMONE_THERAPY PAM50 COHORT ER_IHC TUMOR_GRADE TUMOR_SIZE
## 190             1               1     1      1      1           1          1
##                 0               0     0      0      0           0          0
##     TUMOR_STAGE HER2_SNP6 HISTOLOGICAL_SUBTYPE CELLULARITY MENOPAUSE INTCLUST
## 190           1         1                    1           1         1        1
##               0         0                    0           0         0        0
##     LATERALITY MUTATION_COUNT NPI SURVIVAL_MONTHS SURVIVAL_GROUPS THREEGENE
## 190          1              1   1               1               1         1
##              0              0   0               0               0         0
##     VITAL_STATUS
```

```
## 190            1 0
##                0 0

summary(TNBC_clinical) #confirm absence of missing data and gain insigts on the data

##  AGE_AT_DIAGNOSIS    AGE_GROUP            BREAST_SURGERY
##  Min.   :26.72    Below 40:37   BREAST CONSERVING:113
##  1st Qu.:43.05    40 to 52:56   MASTECTOMY       : 77
##  Median :52.12    Above 52:97
##  Mean   :52.44
##  3rd Qu.:61.09
##  Max.   :96.29
##
##                                    CANCER_TYPE  CHEMOTHERAPY RADIO_THERAPY
##  Breast Invasive Ductal Carcinoma         :159   NO : 60      YES:190
##  Breast Invasive Lobular Carcinoma        :  5   YES:130
##  Breast Mixed Ductal and Lobular Carcinoma:  2
##  Invasive Breast Carcinoma                : 24
##
##
##
##  HORMONE_THERAPY         PAM50         COHORT      ER_IHC     TUMOR_GRADE
##  NO :136         Basal      :100   Min.   :1.000   neg:163   Min.   :1.000
##  YES: 54         claudin-low: 65   1st Qu.:1.000   pos: 27   1st Qu.:3.000
##                  Her2       : 17   Median :3.000             Median :3.000
##                  LumA       :  1   Mean   :2.521             Mean   :2.874
##                  Normal     :  7   3rd Qu.:3.000             3rd Qu.:3.000
##                                    Max.   :5.000             Max.   :3.000
##
##    TUMOR_SIZE      TUMOR_STAGE    HER2_SNP6       HISTOLOGICAL_SUBTYPE
##  Min.   :  3.00   Min.   :0.000   GAIN: 19   IDC            :164
##  1st Qu.: 17.93   1st Qu.:1.250   LOSS: 12   IDC-MED        : 13
##  Median : 25.00   Median :2.000   NEUT:159   ILC            :  6
##  Mean   : 28.19   Mean   :1.884              IDC+ILC        :  2
##  3rd Qu.: 30.00   3rd Qu.:2.000              INVASIVE TUMOUR:  2
##  Max.   :182.00   Max.   :3.000              DCIS           :  1
##                                              (Other)        :  2
##    CELLULARITY  MENOPAUSE     INTCLUST   LATERALITY MUTATION_COUNT
##  high    :111   post:108   10     :120   l:94       Min.   : 0.000
##  low     : 19   pre : 82   4ER-   : 27   r:96       1st Qu.: 3.000
##  moderate: 60              4ER+   : 12              Median : 5.000
##                            9      : 11              Mean   : 5.484
##                            1      :  8              3rd Qu.: 7.000
##                            3      :  5              Max.   :30.000
##                            (Other):  7
##       NPI        SURVIVAL_MONTHS       SURVIVAL_GROUPS
##  Min.   :1.050   Min.   :  4.167   5yrs or Less: 70
##  1st Qu.:4.042   1st Qu.: 35.375   Above 5yrs  :120
##  Median :5.034   Median : 92.350
##  Mean   :4.730   Mean   :113.128
##  3rd Qu.:5.080   3rd Qu.:182.883
##  Max.   :6.320   Max.   :307.933
##
##                  THREEGENE            VITAL_STATUS
##  ER-/HER2-            :173   Died of Disease: 83
##  ER+/HER2- High Prolif: 11   Living         :107
##  ER+/HER2- Low Prolif :  4
##  HER2+                :  2
##
##
##
```

- Median age at diagnosis is 52.12. Suspicion of outliers exist, given that maximum age is 96.
- 51% of patients were older than 52 years while 49% of patients were below the median age of 52.12.
- 59.5% of patients subscribed to Breast conserving while 40.5% went for mastectomy.
- 83.7 of the patients had the “Breast Invasive Ductal Carcinoma” Cancer type.
- 68.4 % of the patients were administered chemotherapy.
- 28.4 % of the patients were administered hormone therapy.
- More than 50% of the patients were of the Basal sub-type based on PAM50 classification.

- The Basal and Claudin-low sub-types based on PAM50 classification, both accounted for 86.8% of total patients.
- 42% of the patients were yet to attain menopause, while 56.8% had attained menopause.
- The median survival months was 92.35 months(7.7years), suspicion of outliers exist.
- 36.8% of patients had at least 5years of follow-up, while 63% had more than 5years of follow-up.
- 43.7% of the patients died from the disease while 56.3% were still living.
- Dependent Variable: VITAL_STATUS.
- All independent variables are either numeric/integer or factor data types.
- Numeric/Integer variables: Age at diagnosis, cohort, tumor grade, tumor size, tumor stage, mutation count, NPI, and survival months(8 in all).
- The 16 other variables not listed above, are of the factor data type.
- Of the 190 observations, 113(59.5%) had complete records while 77 observations had 1 or more missing value(s).
- In all 101 entries (2.1%) were missing out of 4750 entries(25columns x 190observations).
- The affected variables with missing values are: Cellularity, ER_IHC, Tumor Grade, Laterality, ThreeGene, Tumor Size and Tumor Stage.
- PMM (Predictive Mean Matching) method was applied to the following variables: TUMOR_GRADE, TUMOR_SIZE, and TUMOR_STAGE. Note that these are all numeric variables.
- Polynomial regression was applied to the CELLULARITY and THREEGENE variables.
- Logistic regression was applied to the ER_IHC and LATERALITY variables.
- 1.20. To enhance uni-variate statistical analyses, the factor type variables will be encoded using

```
•       numbers and converted to integer data type and stored in the "clin" object.
```

- 1.21. Run table function on concerned variables to correctly pick out value entries.
- 1.22. Confirm that all factor types have been converted to numeric and view “clin” object.

```
attach(clin)
"BREAST_SURGERY"

## [1] "BREAST_SURGERY"

table(BREAST_SURGERY)

## BREAST_SURGERY
## BREAST CONSERVING        MASTECTOMY
##               113                77

clin$BREAST_SURGERY = as.character(clin$BREAST_SURGERY)
clin$BREAST_SURGERY[clin$BREAST_SURGERY == "MASTECTOMY"] <- 1
clin$BREAST_SURGERY[clin$BREAST_SURGERY == "BREAST CONSERVING"] <- 2
clin$BREAST_SURGERY = as.integer(clin$BREAST_SURGERY)

"CANCER_TYPE"

## [1] "CANCER_TYPE"

table(CANCER_TYPE)

## CANCER_TYPE
##          Breast Invasive Ductal Carcinoma
##                                       159
##         Breast Invasive Lobular Carcinoma
##                                         5
## Breast Mixed Ductal and Lobular Carcinoma
##                                         2
##                 Invasive Breast Carcinoma
##                                        24
```

```
clin$CANCER_TYPE = as.character(clin$CANCER_TYPE)
clin$CANCER_TYPE[clin$CANCER_TYPE == "Invasive Breast Carcinoma"] <- 1
clin$CANCER_TYPE[clin$CANCER_TYPE == "Breast Invasive Ductal Carcinoma"] <- 2
clin$CANCER_TYPE[clin$CANCER_TYPE == "Breast Mixed Ductal and Lobular Carcinoma"] <- 3
clin$CANCER_TYPE[clin$CANCER_TYPE == "Breast Invasive Lobular Carcinoma"] <- 4
clin$CANCER_TYPE = as.integer(clin$CANCER_TYPE)

"CELLULARITY"

## [1] "CELLULARITY"

table(CELLULARITY)

## CELLULARITY
##     high      low moderate
##      111       19       60

clin$CELLULARITY = as.character(clin$CELLULARITY)
clin$CELLULARITY[clin$CELLULARITY == "low"] <- 1
clin$CELLULARITY[clin$CELLULARITY == "moderate"] <- 2
clin$CELLULARITY[clin$CELLULARITY == "high"] <- 3
clin$CELLULARITY = as.integer(clin$CELLULARITY)

"HISTOLOGICAL_SUBTYPE"

## [1] "HISTOLOGICAL_SUBTYPE"

table(HISTOLOGICAL_SUBTYPE)

## HISTOLOGICAL_SUBTYPE
##            DCIS             IDC         IDC-MED         IDC-TUB         IDC+ILC
##               1             164              13               1               2
##             ILC INVASIVE TUMOUR           OTHER
##               6               2               1

clin$HISTOLOGICAL_SUBTYPE = as.character(clin$HISTOLOGICAL_SUBTYPE)
clin$HISTOLOGICAL_SUBTYPE[clin$HISTOLOGICAL_SUBTYPE == "DCIS"] <- 1
clin$HISTOLOGICAL_SUBTYPE[clin$HISTOLOGICAL_SUBTYPE == "IDC"] <- 2
clin$HISTOLOGICAL_SUBTYPE[clin$HISTOLOGICAL_SUBTYPE == "IDC-MED"] <- 3
clin$HISTOLOGICAL_SUBTYPE[clin$HISTOLOGICAL_SUBTYPE == "IDC-TUB"] <- 4
clin$HISTOLOGICAL_SUBTYPE[clin$HISTOLOGICAL_SUBTYPE == "IDC+ILC"] <- 5
clin$HISTOLOGICAL_SUBTYPE[clin$HISTOLOGICAL_SUBTYPE == "ILC"] <- 6
clin$HISTOLOGICAL_SUBTYPE[clin$HISTOLOGICAL_SUBTYPE == "INVASIVE TUMOUR"] <- 7
clin$HISTOLOGICAL_SUBTYPE[clin$HISTOLOGICAL_SUBTYPE == "OTHER"] <- 8
clin$HISTOLOGICAL_SUBTYPE = as.integer(clin$HISTOLOGICAL_SUBTYPE)

str(clin$HISTOLOGICAL_SUBTYPE)

##  int [1:190] 6 3 2 2 2 2 2 2 2 2 ...

"CHEMOTHERAPY"

## [1] "CHEMOTHERAPY"

table(CHEMOTHERAPY)

## CHEMOTHERAPY
##  NO YES
##  60 130

clin$CHEMOTHERAPY = as.character(clin$CHEMOTHERAPY)
clin$CHEMOTHERAPY[clin$CHEMOTHERAPY == "YES"] <- 1
clin$CHEMOTHERAPY[clin$CHEMOTHERAPY == "NO"] <- 2
clin$CHEMOTHERAPY = as.integer(clin$CHEMOTHERAPY)

"RADIO_THERAPY"

## [1] "RADIO_THERAPY"

table(RADIO_THERAPY)

## RADIO_THERAPY
## YES
## 190
```

```
clin$RADIO_THERAPY = as.character(clin$RADIO_THERAPY)
clin$RADIO_THERAPY[clin$RADIO_THERAPY == "YES"] <- 1
clin$RADIO_THERAPY[clin$RADIO_THERAPY == "NO"] <- 2
clin$RADIO_THERAPY = as.integer(clin$RADIO_THERAPY)

"HORMONE_THERAPY"

## [1] "HORMONE_THERAPY"

table(HORMONE_THERAPY)

## HORMONE_THERAPY
##  NO YES
## 136  54

clin$HORMONE_THERAPY = as.character(clin$HORMONE_THERAPY)
clin$HORMONE_THERAPY[clin$HORMONE_THERAPY == "YES"] <- 1
clin$HORMONE_THERAPY[clin$HORMONE_THERAPY == "NO"] <- 2
clin$HORMONE_THERAPY = as.integer(clin$HORMONE_THERAPY)

"PAM50"

## [1] "PAM50"

table(PAM50)

## PAM50
##       Basal claudin-low        Her2        LumA      Normal
##         100          65          17           1           7

clin$PAM50 = as.character(clin$PAM50)
clin$PAM50[clin$PAM50 == "Basal"] <- 1
clin$PAM50[clin$PAM50 == "claudin-low"] <- 2
clin$PAM50[clin$PAM50 == "Her2"] <- 3
clin$PAM50[clin$PAM50 == "LumA"] <- 4
clin$PAM50[clin$PAM50 == "Normal"] <- 5
clin$PAM50 = as.integer(clin$PAM50)

"ER_IHC"

## [1] "ER_IHC"

table(ER_IHC)

## ER_IHC
## neg pos
## 163  27

clin$ER_IHC = as.character(clin$ER_IHC)
clin$ER_IHC[clin$ER_IHC == "pos"] <- 1
clin$ER_IHC[clin$ER_IHC == "neg"] <- 2
clin$ER_IHC = as.integer(clin$ER_IHC)

"HER2_SNP6"

## [1] "HER2_SNP6"

table(HER2_SNP6)

## HER2_SNP6
## GAIN LOSS NEUT
##   19   12  159

clin$HER2_SNP6 = as.character(clin$HER2_SNP6)
clin$HER2_SNP6[clin$HER2_SNP6 == "GAIN"] <- 1
clin$HER2_SNP6[clin$HER2_SNP6 == "NEUT"] <- 2
clin$HER2_SNP6[clin$HER2_SNP6 == "LOSS"] <- 3
clin$HER2_SNP6 = as.integer(clin$HER2_SNP6)

"MENOPAUSE"

## [1] "MENOPAUSE"

table(MENOPAUSE)
```

```
## MENOPAUSE
## post  pre
##  108   82

clin$MENOPAUSE = as.character(clin$MENOPAUSE)
clin$MENOPAUSE[clin$MENOPAUSE == "pre"] <- 1
clin$MENOPAUSE[clin$MENOPAUSE == "post"] <- 2
clin$MENOPAUSE = as.integer(clin$MENOPAUSE)

"INTCLUST"

## [1] "INTCLUST"

table(INTCLUST)

## INTCLUST
##    1   10    2    3 4ER- 4ER+    5    7    9
##    8  120    3    5   27   12    1    3   11

clin$INTCLUST = as.character(clin$INTCLUST)
clin$INTCLUST[clin$INTCLUST == "1"] <- 1
clin$INTCLUST[clin$INTCLUST == "2"] <- 2
clin$INTCLUST[clin$INTCLUST == "3"] <- 3
clin$INTCLUST[clin$INTCLUST == "4ER-"] <- 4
clin$INTCLUST[clin$INTCLUST == "5"] <- 5
clin$INTCLUST[clin$INTCLUST == "4ER+"] <- 6
clin$INTCLUST[clin$INTCLUST == "7"] <- 7
clin$INTCLUST[clin$INTCLUST == "9"] <- 9
clin$INTCLUST[clin$INTCLUST == "10"] <- 10
clin$INTCLUST = as.integer(clin$INTCLUST)

"LATERALITY"

## [1] "LATERALITY"

table(LATERALITY)

## LATERALITY
##  l  r
## 94 96

clin$LATERALITY = as.character(clin$LATERALITY)
clin$LATERALITY[clin$LATERALITY == "l"] <- 1
clin$LATERALITY[clin$LATERALITY == "r"] <- 2
clin$LATERALITY = as.integer(clin$LATERALITY)

"THREEGENE"

## [1] "THREEGENE"

table(THREEGENE)

## THREEGENE
##             ER-/HER2- ER+/HER2- High Prolif  ER+/HER2- Low Prolif
##                  173                    11                     4
##                 HER2+
##                     2

clin$THREEGENE = as.character(clin$THREEGENE)
clin$THREEGENE[clin$THREEGENE == "ER-/HER2-"] <- 1
clin$THREEGENE[clin$THREEGENE == "ER+/HER2- High Prolif"] <- 2
clin$THREEGENE[clin$THREEGENE == "ER+/HER2- Low Prolif"] <- 3
clin$THREEGENE[clin$THREEGENE == "HER2+"] <- 4
clin$THREEGENE = as.integer(clin$THREEGENE)

"VITAL_STATUS"

## [1] "VITAL_STATUS"

table(VITAL_STATUS)

## VITAL_STATUS
## Died of Disease          Living
##              83             107
```

```r
clin$VITAL_STATUS = as.character(clin$VITAL_STATUS)
clin$VITAL_STATUS[clin$VITAL_STATUS == "Living"] <- 0
clin$VITAL_STATUS[clin$VITAL_STATUS == "Died of Disease"] <- 1
clin$VITAL_STATUS = as.integer(clin$VITAL_STATUS)

"AGE_GROUP"

## [1] "AGE_GROUP"

table(AGE_GROUP)

## AGE_GROUP
## Below 40 40 to 52 Above 52
##       37       56       97

clin$AGE_GROUP = as.character(clin$AGE_GROUP)
clin$AGE_GROUP[clin$AGE_GROUP == "Below 40"] <- 1
clin$AGE_GROUP[clin$AGE_GROUP == "40 to 52"] <- 2
clin$AGE_GROUP[clin$AGE_GROUP == "Above 52"] <- 3
clin$AGE_GROUP = as.integer(clin$AGE_GROUP)

"SURVIVAL_GROUPS"

## [1] "SURVIVAL_GROUPS"

table(SURVIVAL_GROUPS)

## SURVIVAL_GROUPS
## 5yrs or Less   Above 5yrs
##           70          120

clin$SURVIVAL_GROUPS = as.character(clin$SURVIVAL_GROUPS)
clin$SURVIVAL_GROUPS[clin$SURVIVAL_GROUPS == "5yrs or Less"] <- 1
clin$SURVIVAL_GROUPS[clin$SURVIVAL_GROUPS == "Above 5yrs"] <- 2
clin$SURVIVAL_GROUPS = as.integer(clin$SURVIVAL_GROUPS)

str(clin)

## 'data.frame':    190 obs. of  25 variables:
##  $ AGE_AT_DIAGNOSIS    : num  70 52.1 50.4 68.7 39.8 ...
##  $ AGE_GROUP           : int  3 3 2 3 1 3 3 1 3 3 ...
##  $ BREAST_SURGERY      : int  1 1 1 1 1 2 2 1 2 1 ...
##  $ CANCER_TYPE         : int  1 1 2 2 2 2 2 2 2 2 ...
##  $ CHEMOTHERAPY        : int  1 1 1 1 1 1 2 1 1 1 ...
##  $ RADIO_THERAPY       : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ HORMONE_THERAPY     : int  2 2 2 2 2 1 2 2 2 2 ...
##  $ PAM50               : int  5 1 3 1 1 5 2 2 2 2 ...
##  $ COHORT              : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ ER_IHC              : int  2 2 2 2 2 1 2 2 2 2 ...
##  $ TUMOR_GRADE         : int  3 3 3 3 3 2 3 3 3 3 ...
##  $ TUMOR_SIZE          : num  65 17 40 39 25 13 30 16 24 34 ...
##  $ TUMOR_STAGE         : int  3 1 2 2 2 2 2 2 2 3 ...
##  $ HER2_SNP6           : int  2 2 2 2 2 2 2 2 2 1 ...
##  $ HISTOLOGICAL_SUBTYPE: int  6 3 2 2 2 2 2 2 2 2 ...
##  $ CELLULARITY         : int  3 3 3 1 2 1 2 3 1 1 ...
##  $ MENOPAUSE           : int  2 2 2 2 1 2 2 1 2 2 ...
##  $ INTCLUST            : int  4 10 10 10 10 3 4 10 4 4 ...
##  $ LATERALITY          : int  1 2 2 2 1 2 1 1 1 2 ...
##  $ MUTATION_COUNT      : int  0 3 4 1 5 5 1 4 2 0 ...
##  $ NPI                 : num  6.13 4.03 6.08 4.08 4.05 ...
##  $ SURVIVAL_MONTHS     : num  22.4 153.97 28.5 8.07 66.73 ...
##  $ SURVIVAL_GROUPS     : int  1 2 1 1 2 2 2 1 2 1 ...
##  $ THREEGENE           : int  1 1 1 1 1 3 1 1 1 1 ...
##  $ VITAL_STATUS        : int  1 0 1 1 1 0 0 0 0 1 ...
```

Univariate Analyses for clinical Features

- 43.7% of the patients died, hence the need to determine factors that drive resistance to treatment so that efforts can be made to identify important features that will help suggest measures to reduce this percentage.

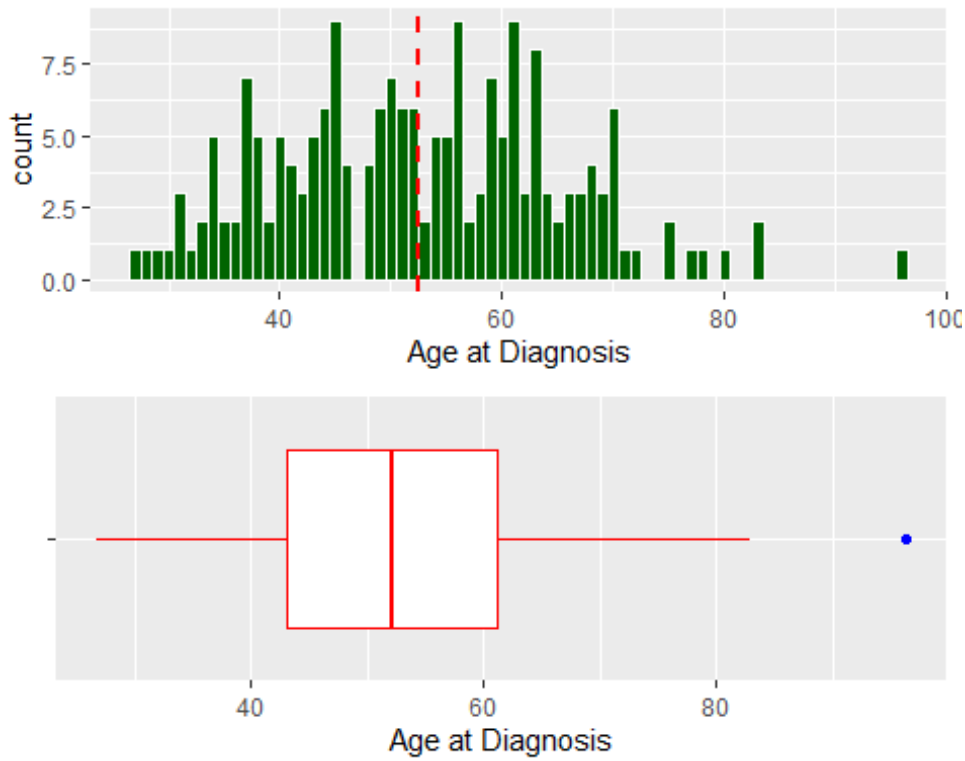
- 1.23. In section 4.0, predictive models that will help to accurately predict patients at risk of resisting successful treatment leading to death will be built.

```r
# Function to visualize histogram and boxplot of numerical variables using ggplot

library(ggplot2) # For graphs and visualizations
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3

library(gridExtra) # To plot multiple ggplot graphs in a grid

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

plot_histogram_n_boxplot = function(variable, variableNameString, binw){
  h = ggplot(data = clin, aes(x= variable))+
    labs(x = variableNameString,y ='count')+
    geom_histogram(fill = 'dark green',col = 'white',binwidth = binw)+
    geom_vline(aes(xintercept = mean(variable)),color="red", linetype="dashed", size=1)

  b = ggplot(data = clin, aes('',variable))+
    geom_boxplot(outlier.colour = 'blue',col = 'red',outlier.shape = 19)+
    labs(x = '',y = variableNameString)+ coord_flip()
  grid.arrange(h,b,ncol = 1)
}
```

- 1.24. Histogram and boxplot visualizations for the Age at diagnosis variable

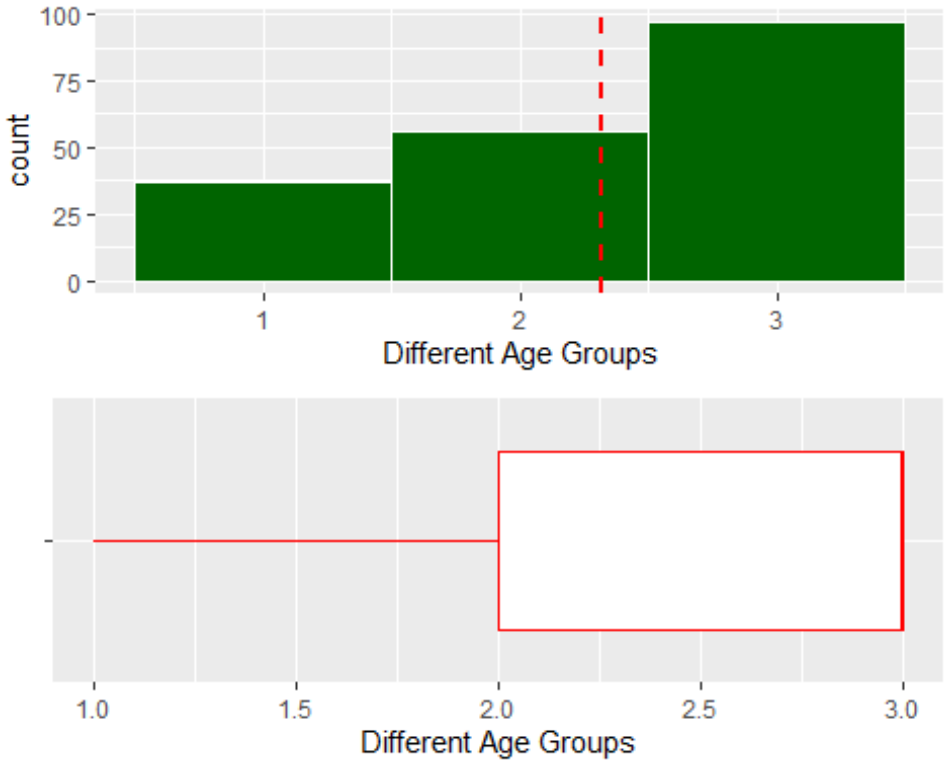
```
plot_histogram_n_boxplot(clin$AGE_AT_DIAGNOSIS, 'Age at Diagnosis', 1)
```



* Distribution looks fairly normal with one outlier.

- 1.25. Histogram and boxplot visualizations for the age groups variable

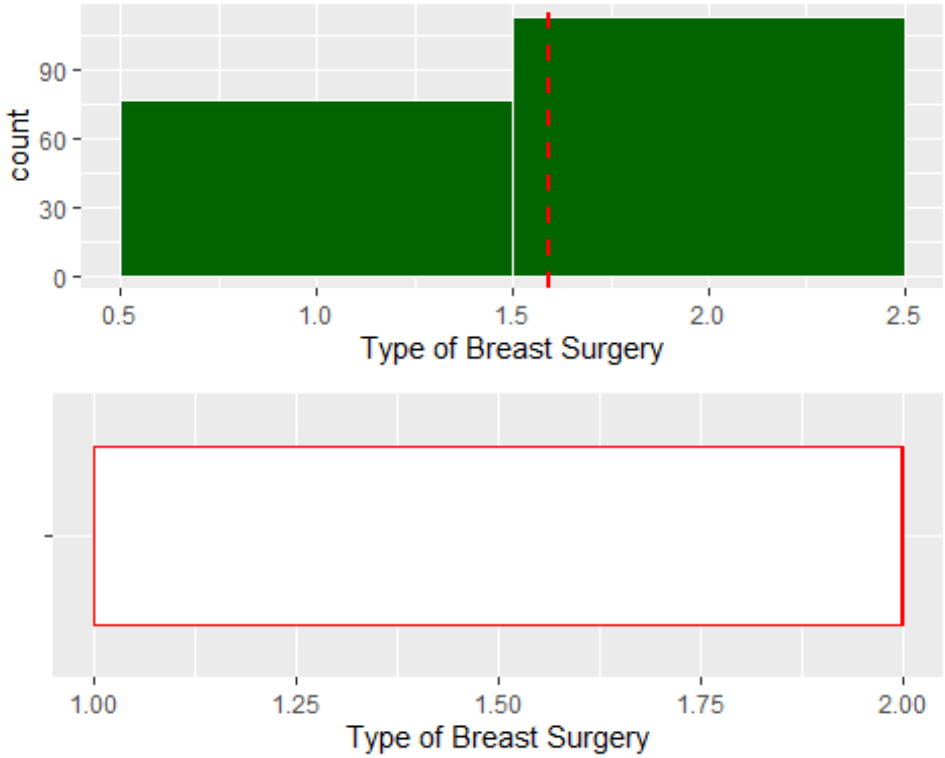
```
plot_histogram_n_boxplot(clin$AGE_GROUP, 'Different Age Groups', 1)
```



* Median age is within the group of patients who are above 52years * Distribution appears left-skewed.

- 1.26. Histogram and boxplot visualizations for the breast surgery type administered to patients

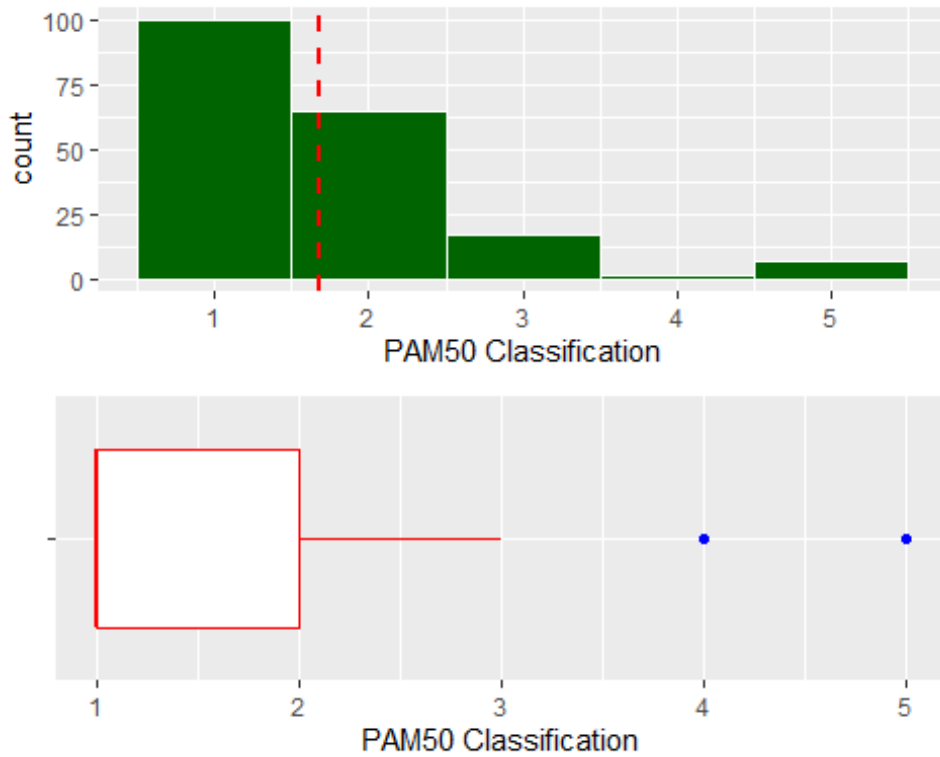
```
plot_histogram_n_boxplot(clin$BREAST_SURGERY, 'Type of Breast Surgery', 1)
```



* More patients underwent the Breast Conserving surgical procedure. * Distribution appears left-skewed.

- 1.27. Histogram and boxplot visualizations for the PAM50 classification variable

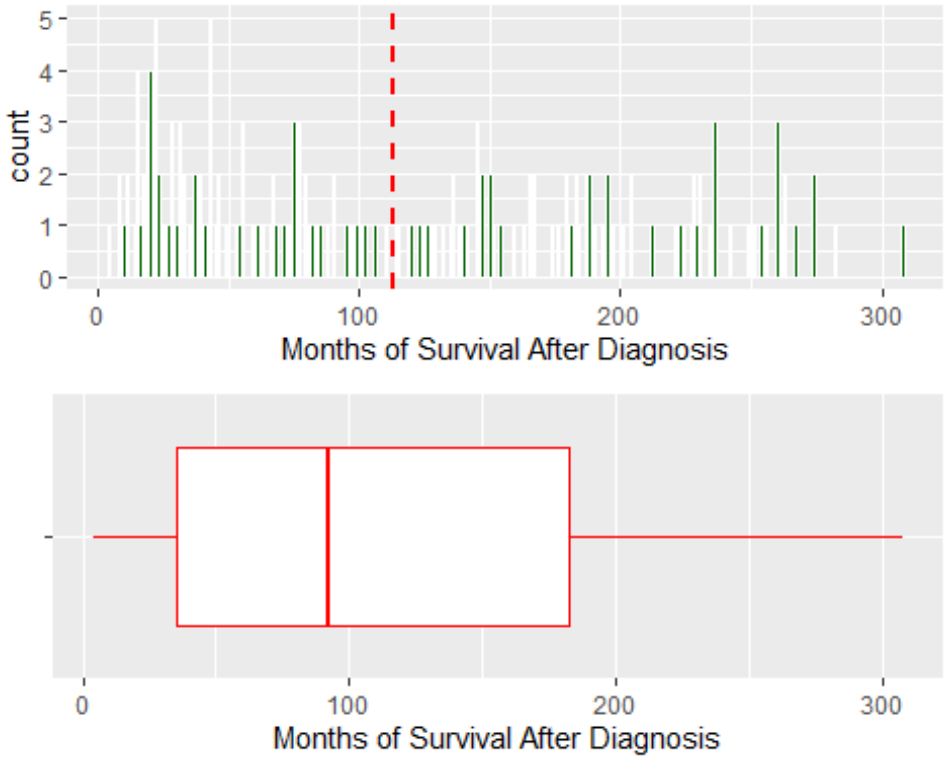
```
plot_histogram_n_boxplot(clin$PAM50, 'PAM50 Classification', 1)
```



* Distribution appears right-skewed with 2 outliers. * Most patients are of the Basal sub-type and two outliers exist in the PAM50 data points

- 1.28. Histogram and boxplot visualizations for survival months variable

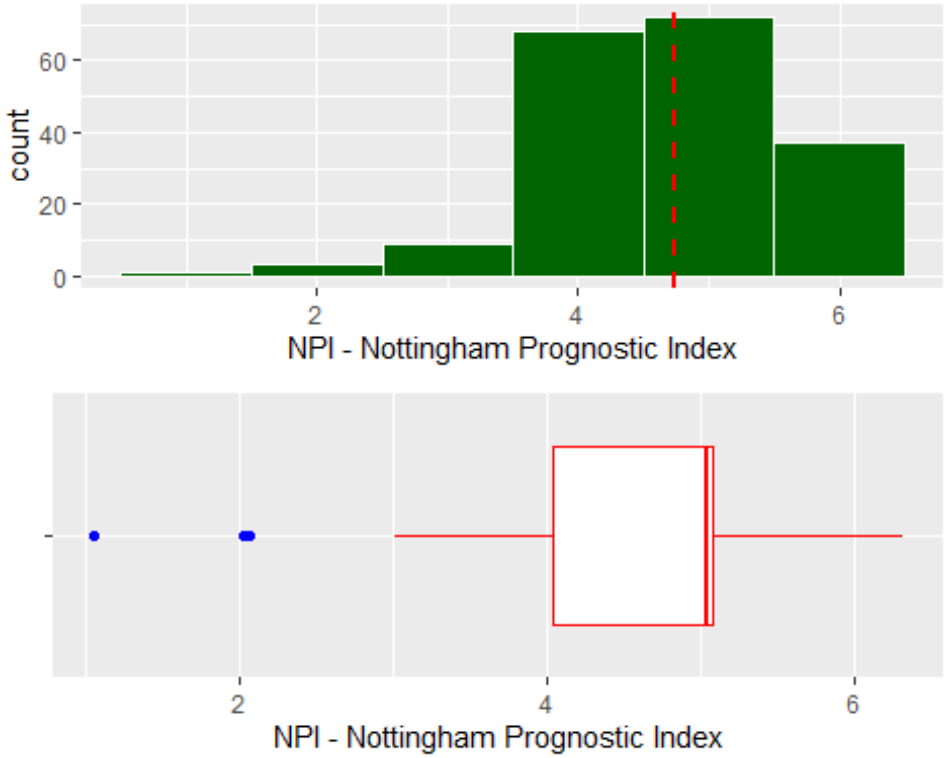
```
plot_histogram_n_boxplot(clin$SURVIVAL_MONTHS, 'Months of Survival After Diagnosis', 1)
```



* Distribution appears right-skewed.

- 1.29. Histogram and boxplot visualizations for the NPI variable

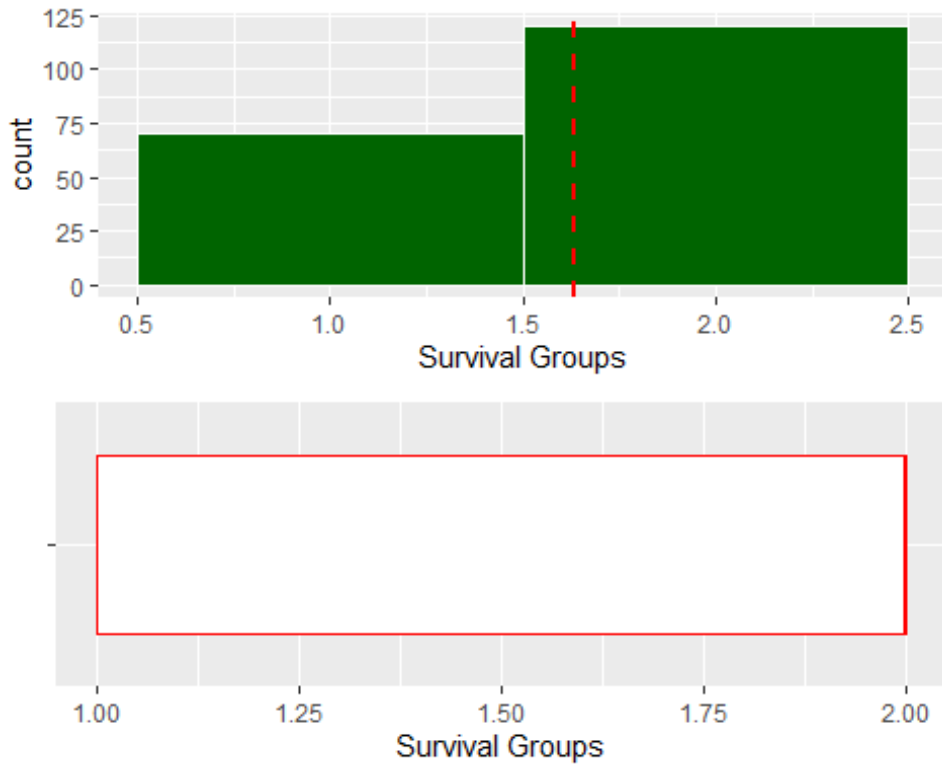
```
plot_histogram_n_boxplot(clin$NPI, 'NPI - Nottingham Prognostic Index', 1)
```



* Distribution appears left-skewed with 3 outliers.

- 1.30. Histogram and boxplot visualizations for the Survival Groups variable

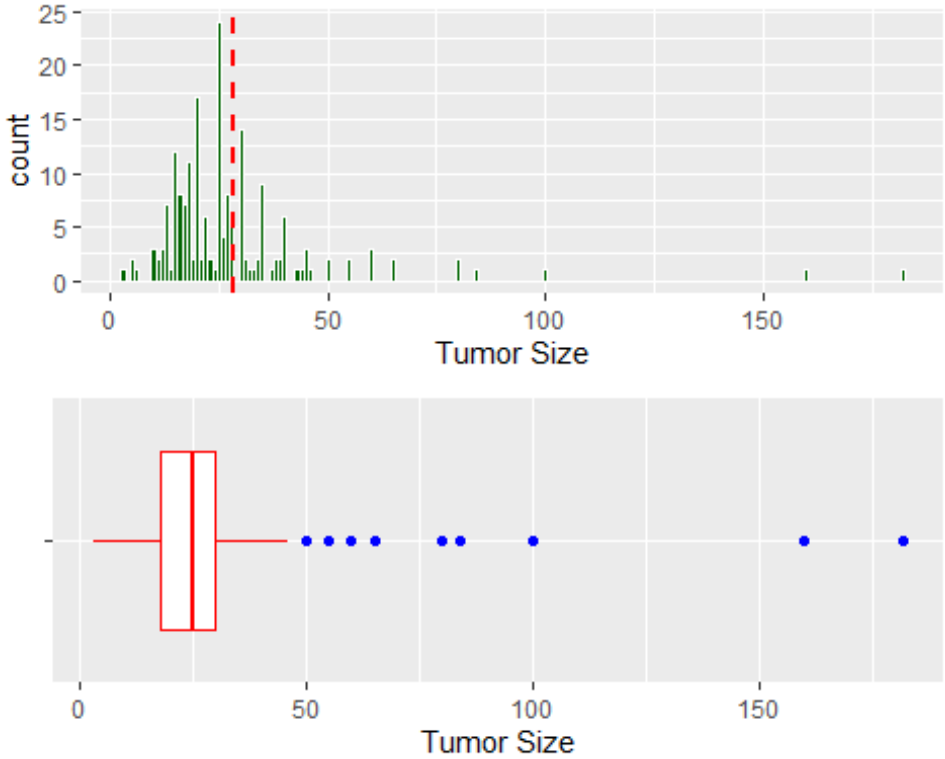
```
plot_histogram_n_boxplot(clin$SURVIVAL_GROUPS, 'Survival Groups', 1)
```



* Distribution appears left-skewed.

- 1.31. Histogram and boxplot visualizations for the Tumor Size variable

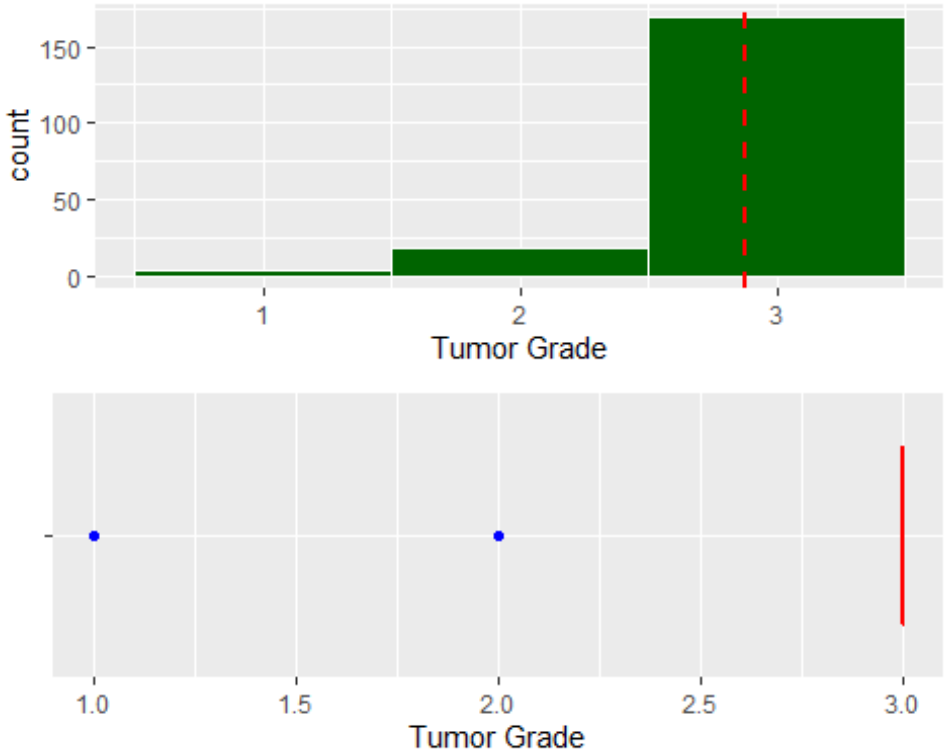
```
plot_histogram_n_boxplot(clin$TUMOR_SIZE, 'Tumor Size', 1)
```



* Distribution appears right-skewed, with 9 outliers.

- 1.32. Histogram and boxplot visualizations for the Tumor Grade variable

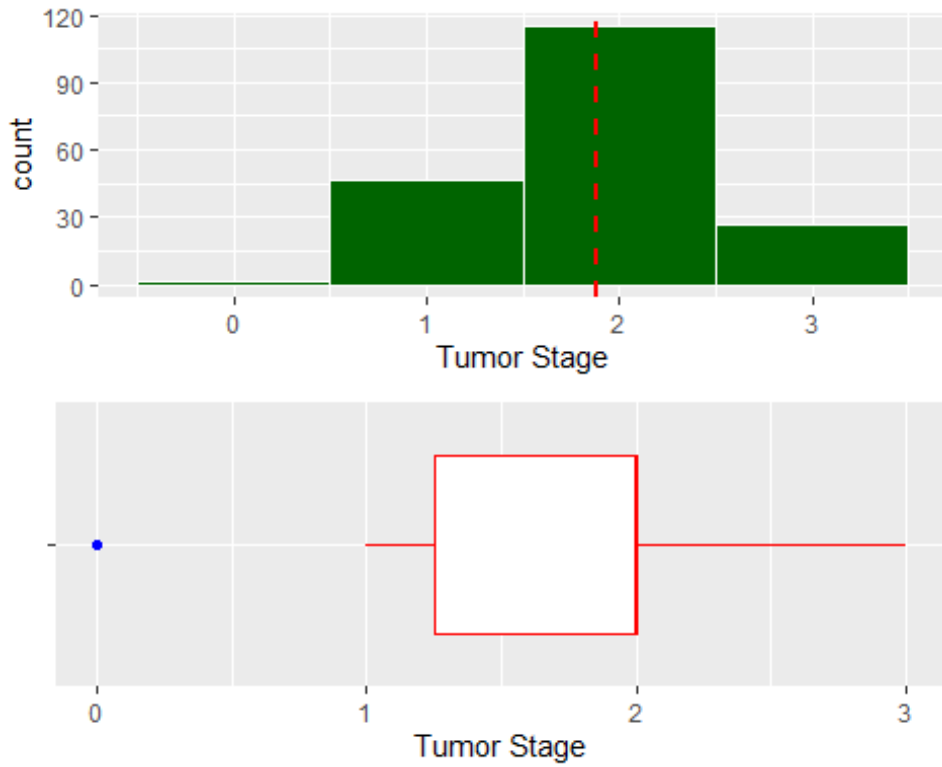
```
plot_histogram_n_boxplot(clin$TUMOR_GRADE, 'Tumor Grade', 1)
```



* Distribution appears left-skewed with 2 outliers.

- 1.33. Histogram and boxplot visualizations for the Tumor Stage variable

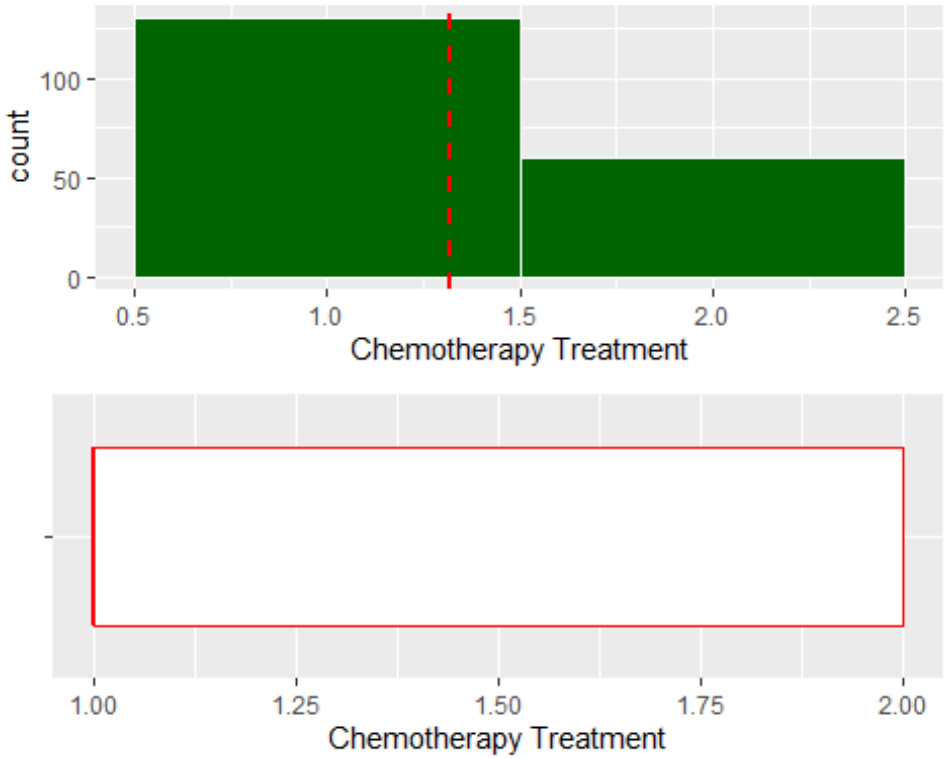
```
plot_histogram_n_boxplot(clin$TUMOR_STAGE, 'Tumor Stage', 1)
```



* Distribution appears left-skewed with 1 outlier.

- 1.34. Histogram and boxplot visualizations for the Chemotherapy variable

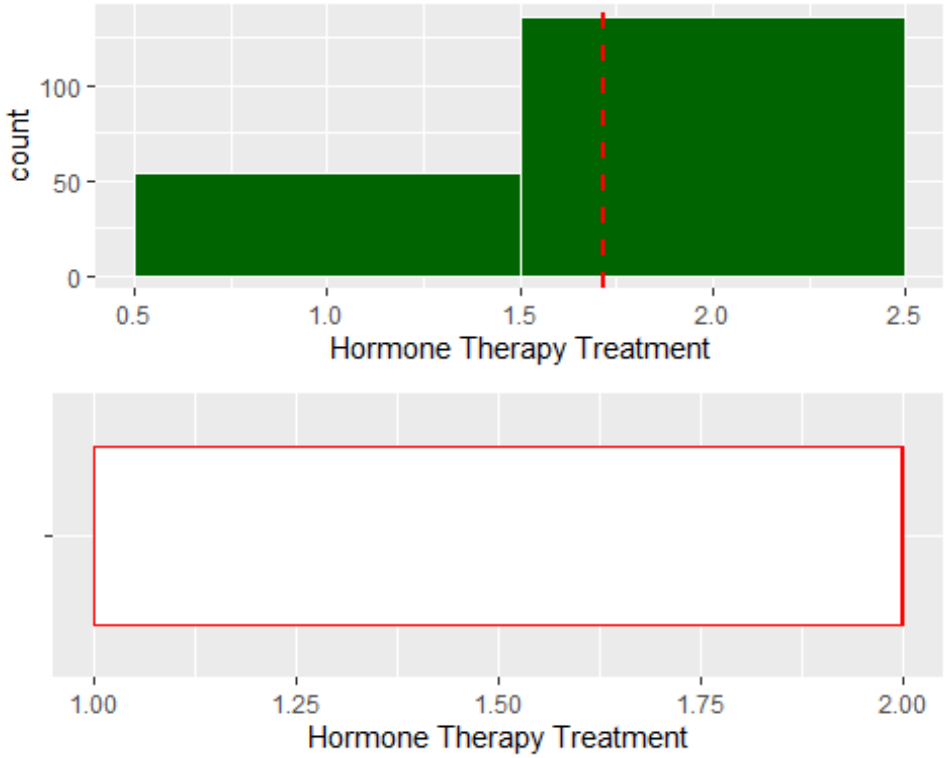
```
plot_histogram_n_boxplot(clin$CHEMOTHERAPY, 'Chemotherapy Treatment', 1)
```



* Distribution appears right-skewed.

- 1.35. Histogram and boxplot visualizations for the Hormone Therapy variable

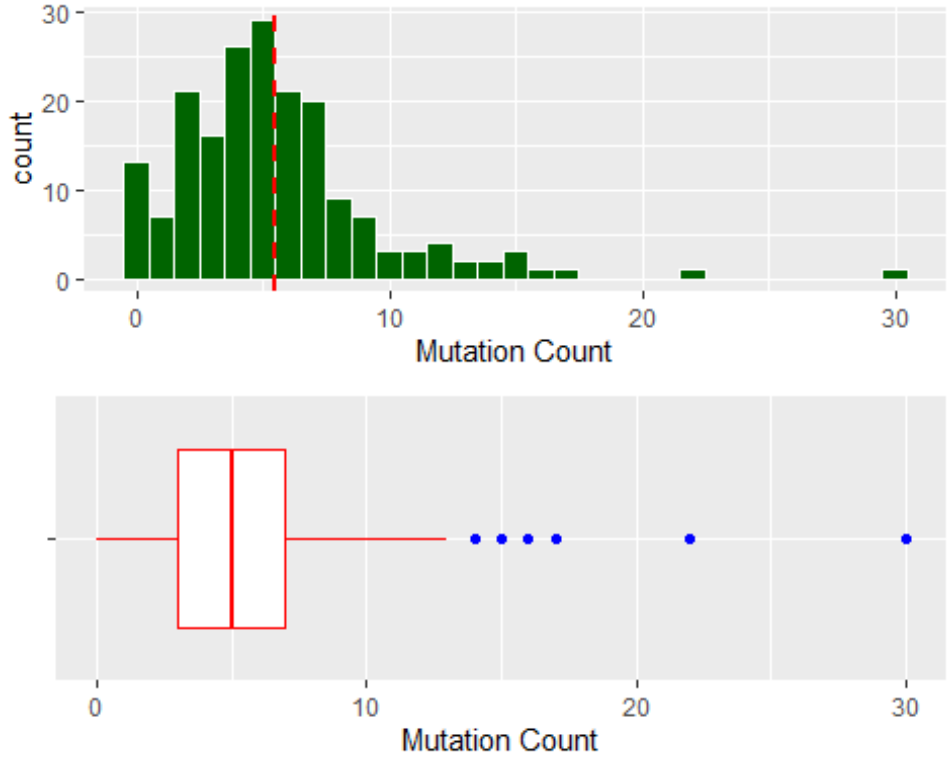
```
plot_histogram_n_boxplot(clin$HORMONE_THERAPY, 'Hormone Therapy Treatment', 1)
```



* Distribution appears left-skewed.

- 1.36. Histogram and boxplot visualizations for the Mutation Count variable

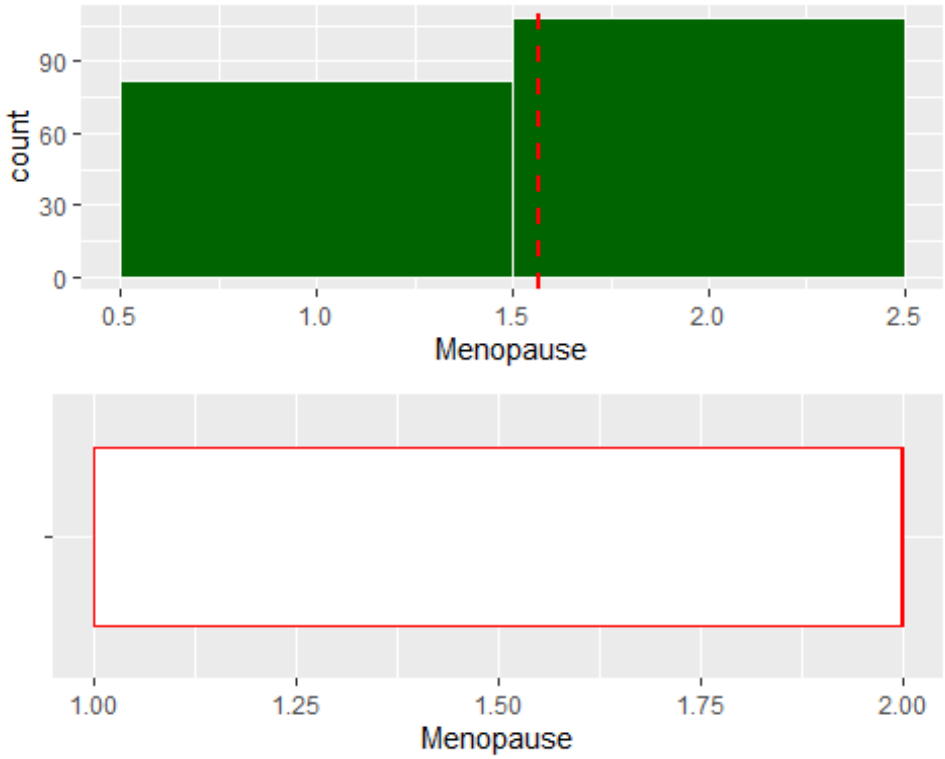
```
plot_histogram_n_boxplot(clin$MUTATION_COUNT, 'Mutation Count', 1)
```



* Distribution appears right-skewed with 6 outliers.

- 1.37. Histogram and boxplot visualizations for the Menopause variable

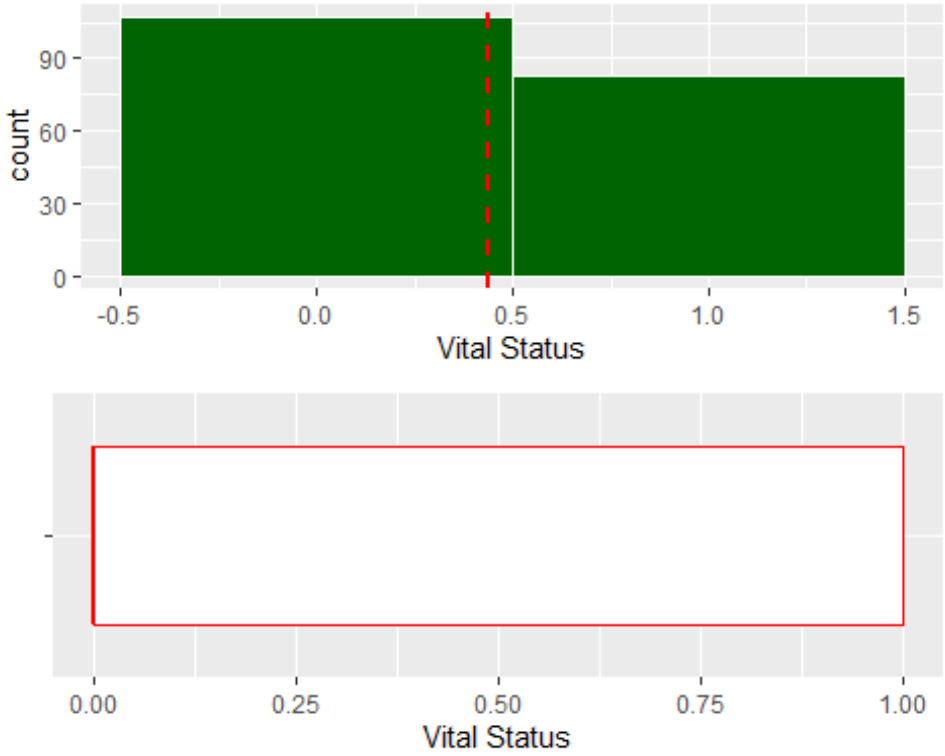
```
plot_histogram_n_boxplot(clin$MENOPAUSE, 'Menopause', 1)
```



* Distribution appears left-skewed.

- 1.38. Histogram and boxplot visualizations for the vital status variable

```
plot_histogram_n_boxplot(clin$VITAL_STATUS, 'Vital Status', 1)
```



* Distribution appears right-skewed.

- create a combined dataset of clinical and mutation features and save in the object TNBC_All

```
TNBC_All = cbind(TNBC_mut[-1],clin)
names(TNBC_All)

##  [1] "AKT1"                  "MAP3K1"                "MEN1"
##  [4] "NCOR2"                 "PIK3CA"                "SHANK2"
##  [7] "MAP2K4"                "AGMO"                  "TAF1"
## [10] "GATA3"                 "BIRC6"                 "DNAH11"
## [13] "NF1"                   "TP53"                  "SYNE1"
## [16] "CTNNA1"                "FOXO3"                 "LDLRAP1"
## [19] "AGE_AT_DIAGNOSIS"      "AGE_GROUP"             "BREAST_SURGERY"
## [22] "CANCER_TYPE"           "CHEMOTHERAPY"          "RADIO_THERAPY"
## [25] "HORMONE_THERAPY"       "PAM50"                 "COHORT"
## [28] "ER_IHC"                "TUMOR_GRADE"           "TUMOR_SIZE"
## [31] "TUMOR_STAGE"           "HER2_SNP6"             "HISTOLOGICAL_SUBTYPE"
## [34] "CELLULARITY"           "MENOPAUSE"             "INTCLUST"
## [37] "LATERALITY"            "MUTATION_COUNT"        "NPI"
## [40] "SURVIVAL_MONTHS"       "SURVIVAL_GROUPS"       "THREEGENE"
## [43] "VITAL_STATUS"
```

- 1.39. Create an object “TNBC5” for all patients who had up to five years of follow-up.
- 1.40. Create an object “TNBC_over5” of all patients who had more than five years of follow-up.

```
TNBC5 = (TNBC_All[(SURVIVAL_MONTHS<60.1),])
dim(TNBC5)

## [1] 70 43

TNBC_over5 = (TNBC_All[(SURVIVAL_MONTHS>60),])
dim(TNBC_over5)

## [1] 120  43
```

- 70 patients had follow-up data for up to five years (TNBC5)
- 120 patients had follow-up data for more than five years (TNBC_over5)

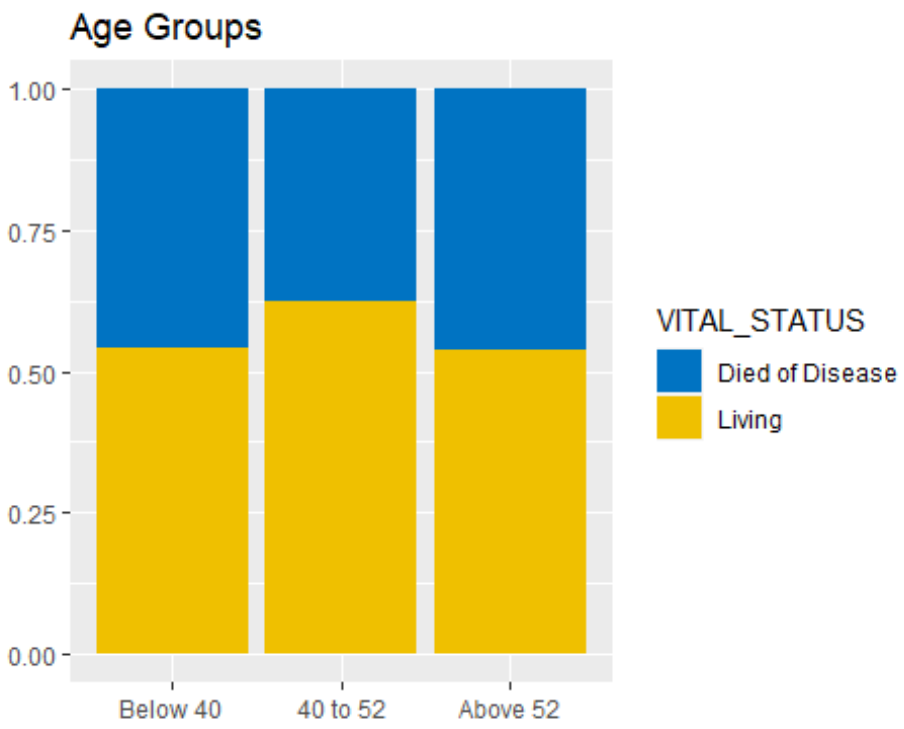
Bivariate Analysis on the 190 patient samples

- 1.41. We go on to plot percent stacked bar chart to see the effect of independent variables on the probability of vital status
- 14.1. Create function to draw percent stacked bar chart to see the effect of independent variables on the probability of vital status using ggplot

```
plot_stacked_barchart = function(variable, variableNameString){
  ggplot(TNBC_clinical, aes(fill = VITAL_STATUS, x = variable)) +
    geom_bar(position="fill")+
    labs(title = variableNameString, y = '', x = '')+
    scale_fill_manual(values=c("#0073C2FF", "#EFC000FF"))
}
```

- 1.43. Vital Status vs Age Groups

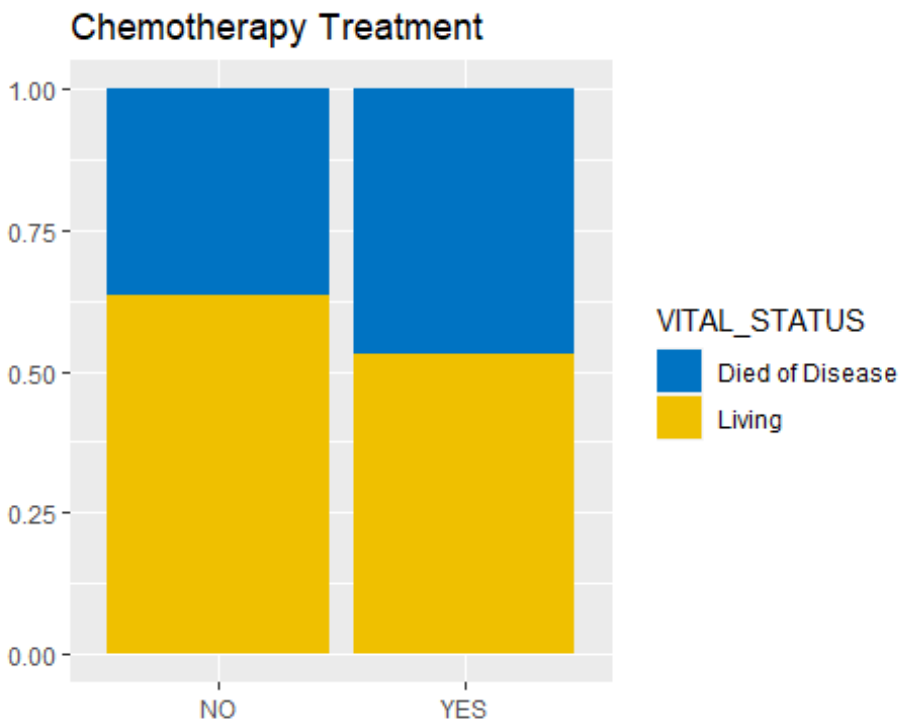
```
plot_stacked_barchart(TNBC_clinical$AGE_GROUP, 'Age Groups')
```



* Survival faired better for the 40 - 52 years age group

- 1.44. Vital Status vs Chemotherapy

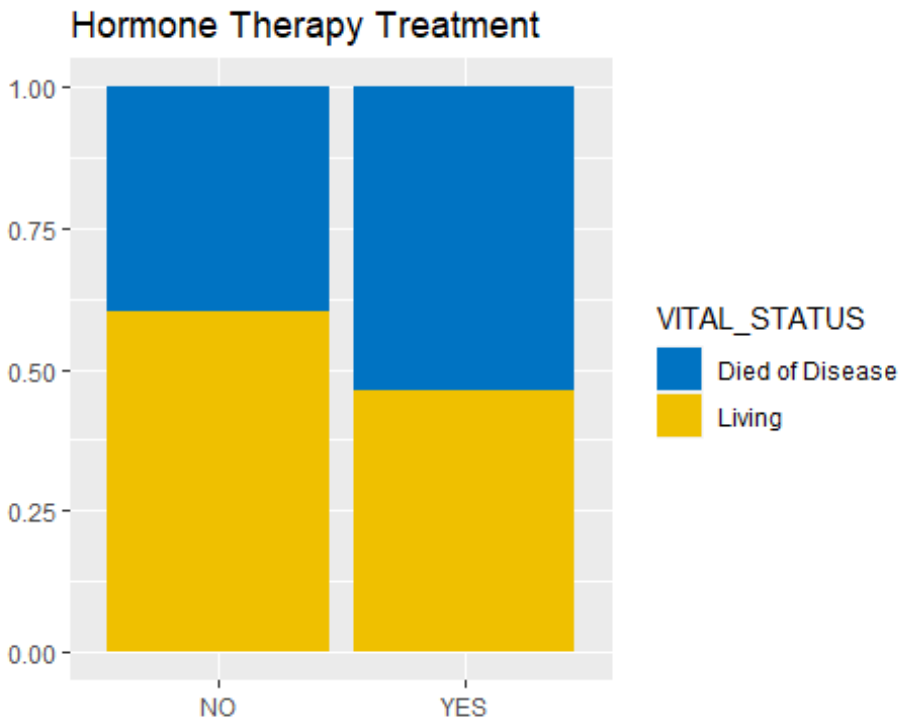
```
plot_stacked_barchart(TNBC_clinical$CHEMOTHERAPY, 'Chemotherapy Treatment')
```



* The group that did not undergo chemotherapy faired better than the group that had chemotherapy

- 1.45. Vital Status vs Hormone Therapy

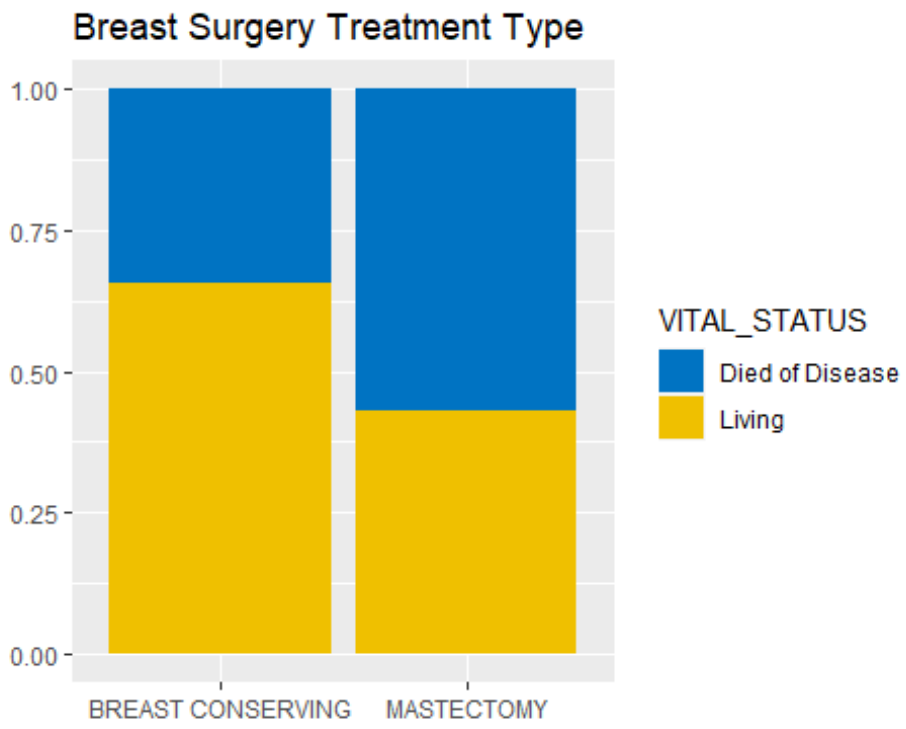
```
plot_stacked_barchart(TNBC_clinical$HORMONE_THERAPY, 'Hormone Therapy Treatment')
```



* More patients died in the group that adopted hormone therapy

- 1.46. Vital Status vs Breast Surgery

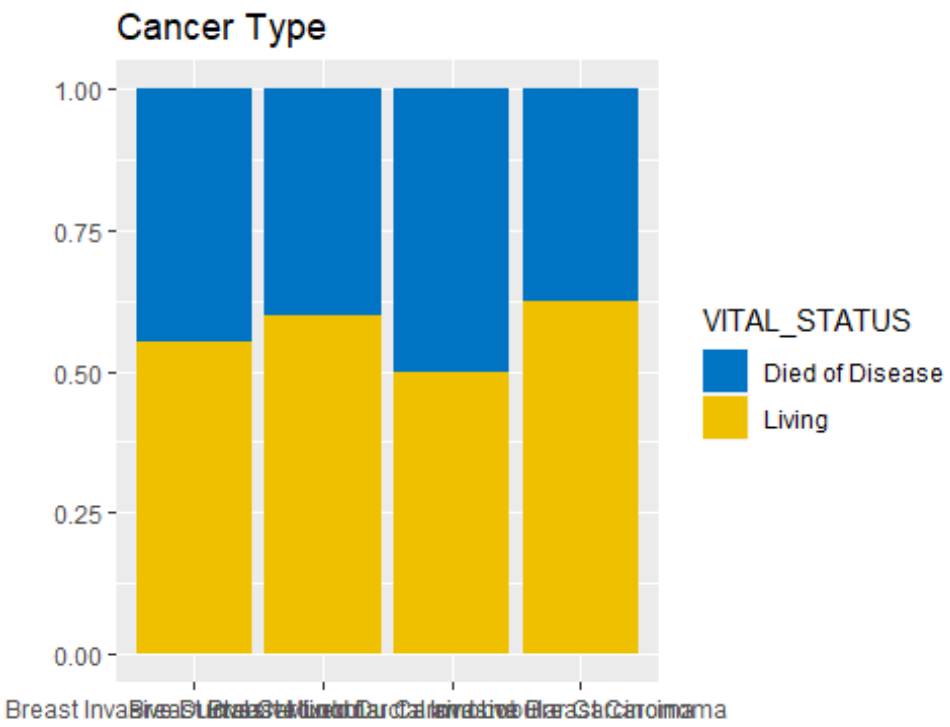
```
plot_stacked_barchart(TNBC_clinical$BREAST_SURGERY, 'Breast Surgery Treatment Type')
```



* Patients who went through breast conserving faired better than those who did mastectomy.

- 1.47. Vital Status vs Cancer Type

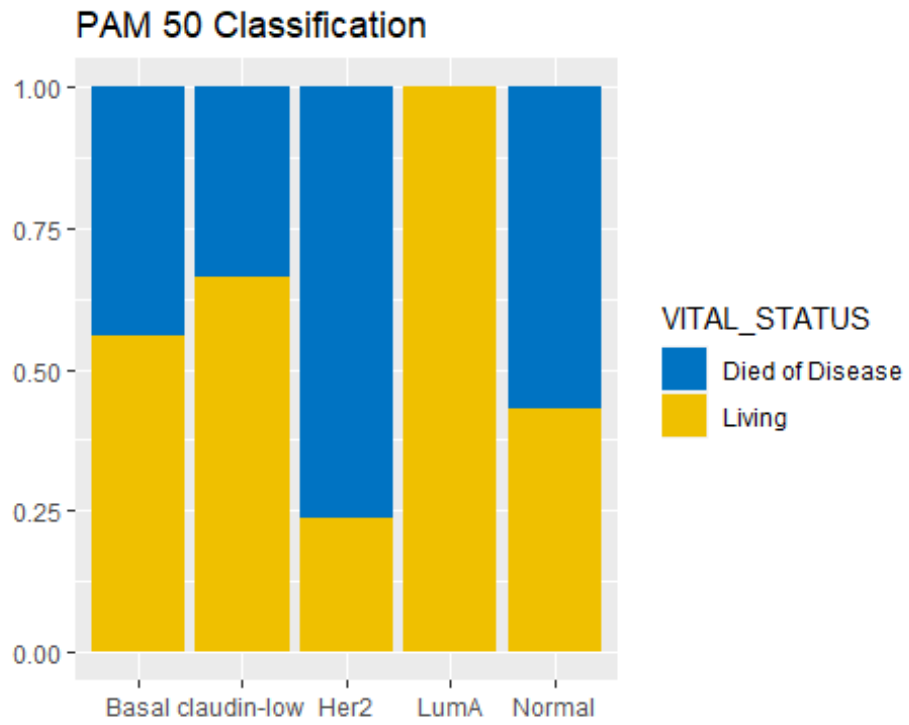
```
plot_stacked_barchart(TNBC_clinical$CANCER_TYPE, 'Cancer Type')
```



* All groups had survival of at least 50%

- 1.48. Vital Status vs PAM 50 Classification

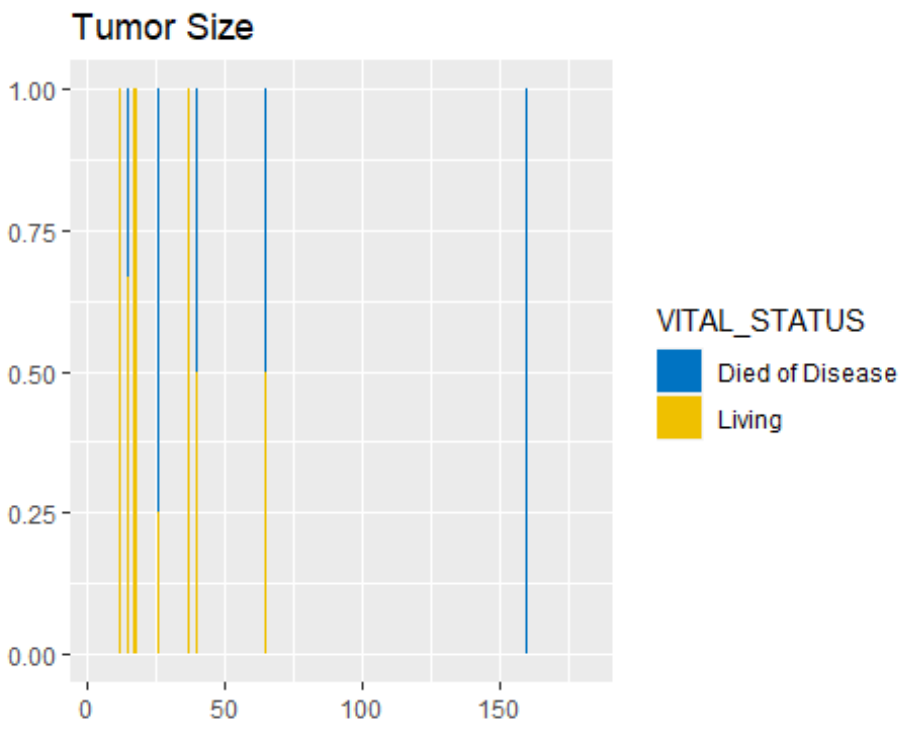
```
plot_stacked_barchart(TNBC_clinical$PAM50, 'PAM 50 Classification')
```



* All patients in the LumA group survived * The Her2 group had the worst outcome

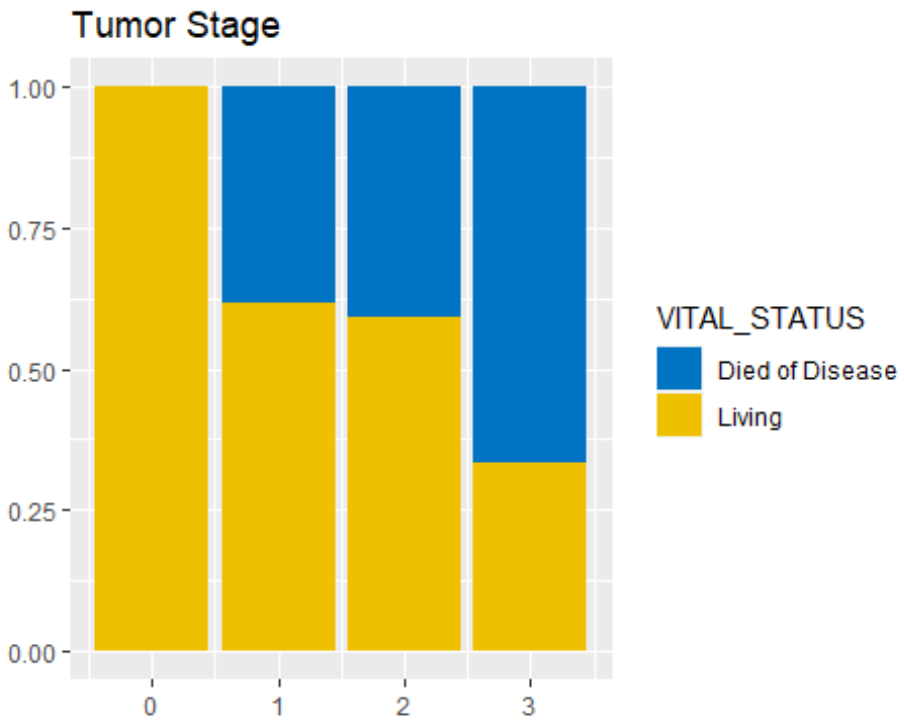
- 1.49. Vital Status vs Tumor Size

```
plot_stacked_barchart(TNBC_clinical$TUMOR_SIZE, 'Tumor Size')
```



- 1.50. Vital Status vs Tumor Stage

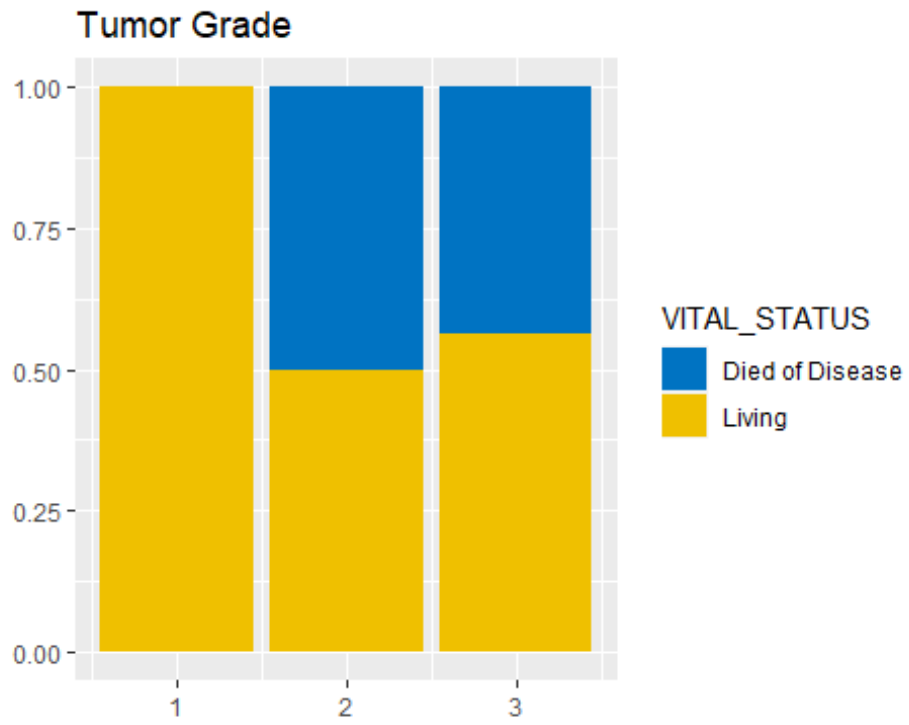
```
plot_stacked_barchart(TNBC_clinical$TUMOR_STAGE, 'Tumor Stage')
```



* Stage 0 patients were all alive * Stage 3 patients had the most deaths

- 1.51. Vital Status vs Tumor grade

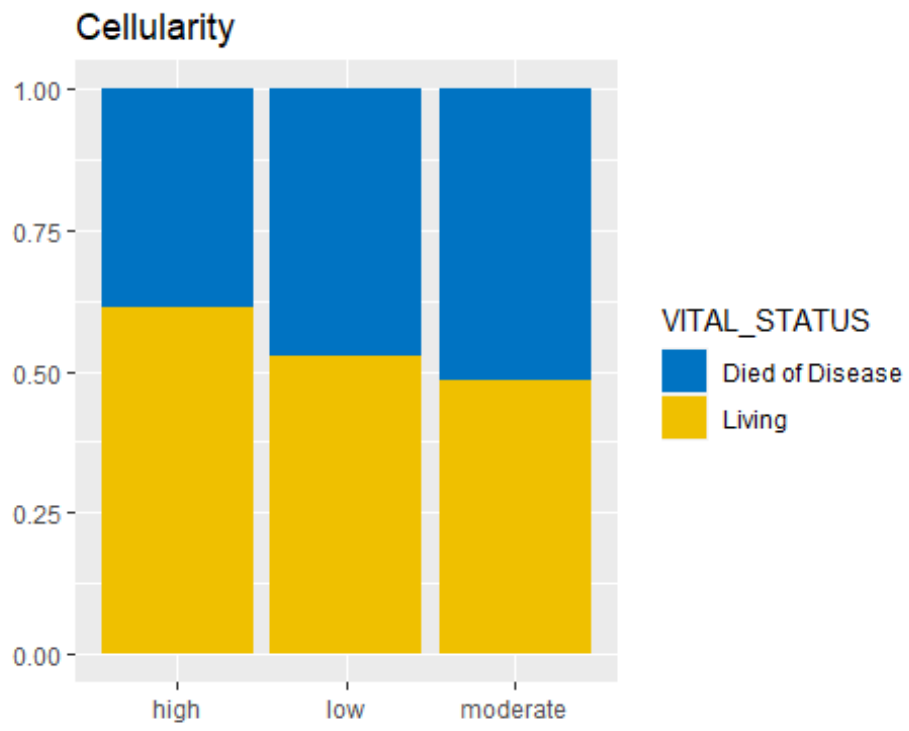
```
plot_stacked_barchart(TNBC_clinical$TUMOR_GRADE, 'Tumor Grade')
```



* Grade 1 patients were all living * Grade 2 patients had the most deaths

- 1.52. Vital Status vs Cellularity

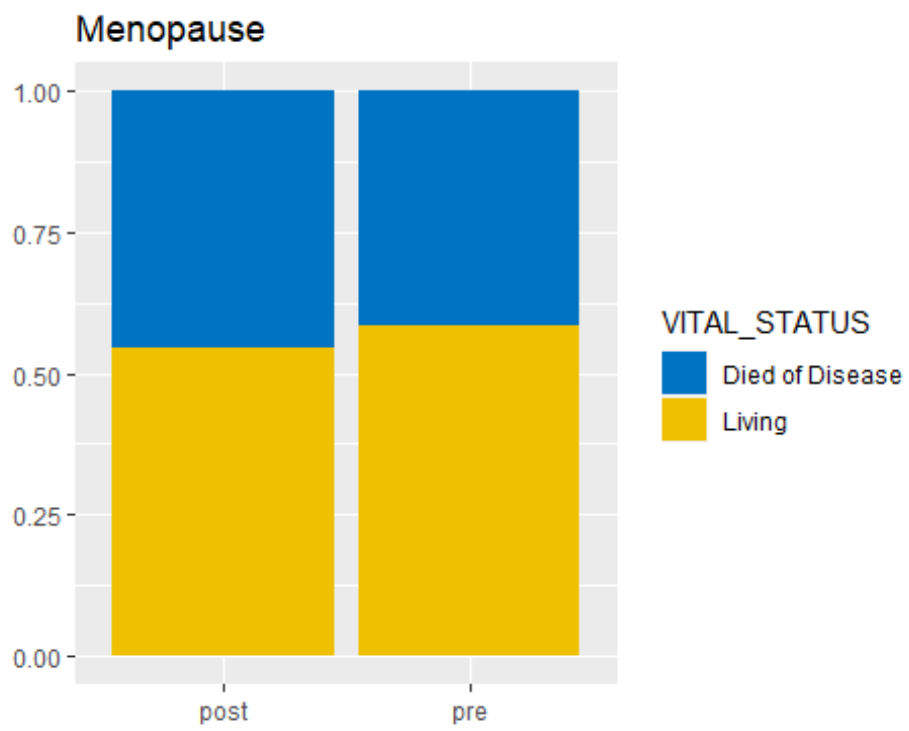
```
plot_stacked_barchart(TNBC_clinical$CELLULARITY, 'Cellularity')
```



* Group with moderate cellularity had the highest deaths

- 1.53. Vital Status vs Menopause

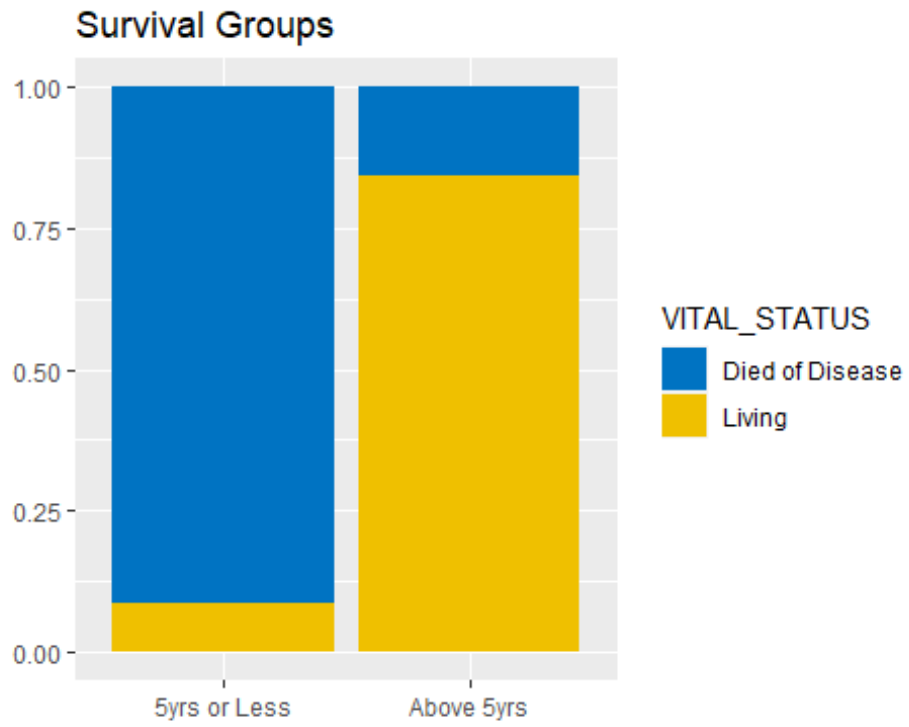
```
plot_stacked_barchart(TNBC_clinical$MENOPAUSE, 'Menopause')
```



* The post-menopausal group had more deaths

- 1.54. Vital Status vs Survival Groups

```
plot_stacked_barchart(TNBC_clinical$SURVIVAL_GROUPS, 'Survival Groups')
```



* The group with follow-up data of 5years or less had a very high rate if deaths

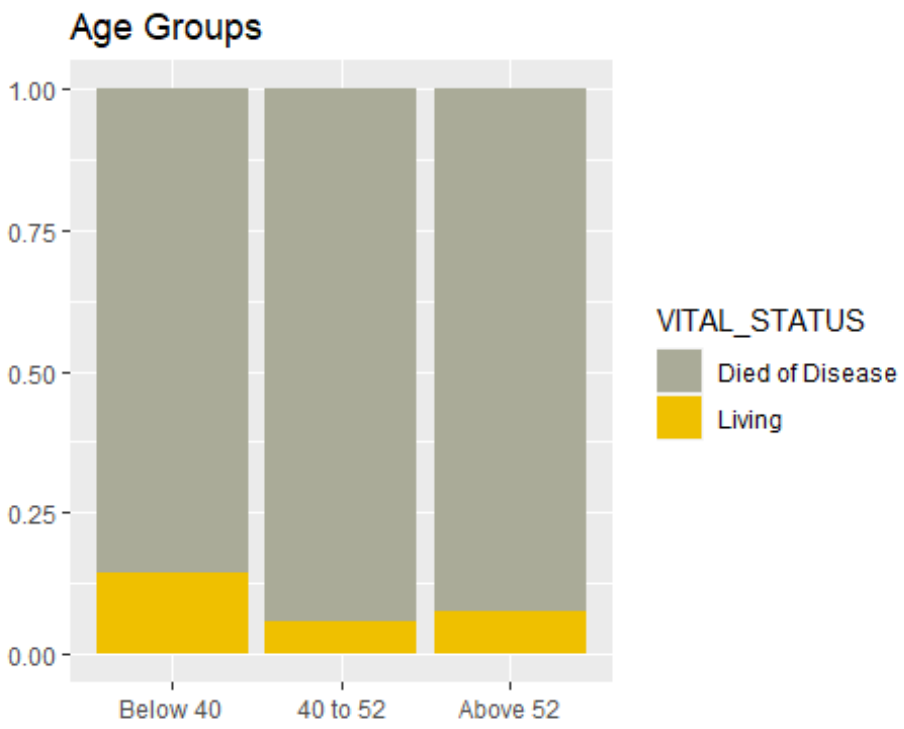
Bivariate analysis on patients with not more than 5 years follow-up

- 1.55. Create object for group with 5 years or less follow up and create function for stacked bar chart

```
clinic5 = (TNBC_clinical[(SURVIVAL_MONTHS<60.1),])

plot_stacked_barchart = function(variable, variableNameString){
  ggplot(clinic5, aes(fill = VITAL_STATUS, x = variable)) +
    geom_bar(position="fill")+
    labs(title = variableNameString, y = '', x = '')+
    scale_fill_manual(values=c("#AAAB98", "#EFC000FF"))
}
```

- 1.56. Vital Status vs Age Groups

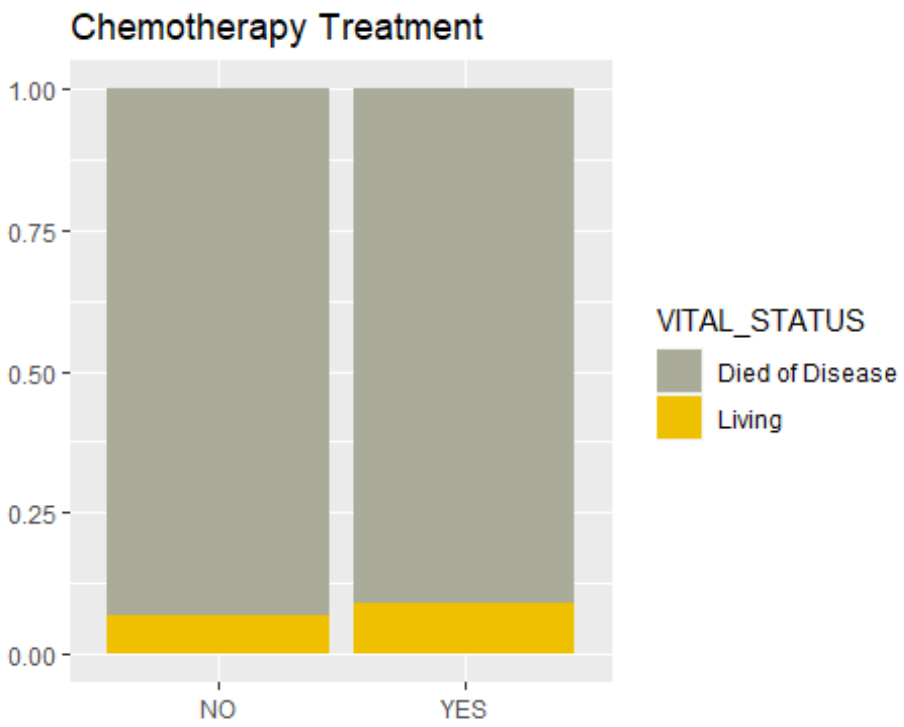
```
plot_stacked_barchart(clinic5$AGE_GROUP, 'Age Groups')
```



* The Below 40 group had the most patients alive

- 1.57. Vital Status vs Chemotherapy

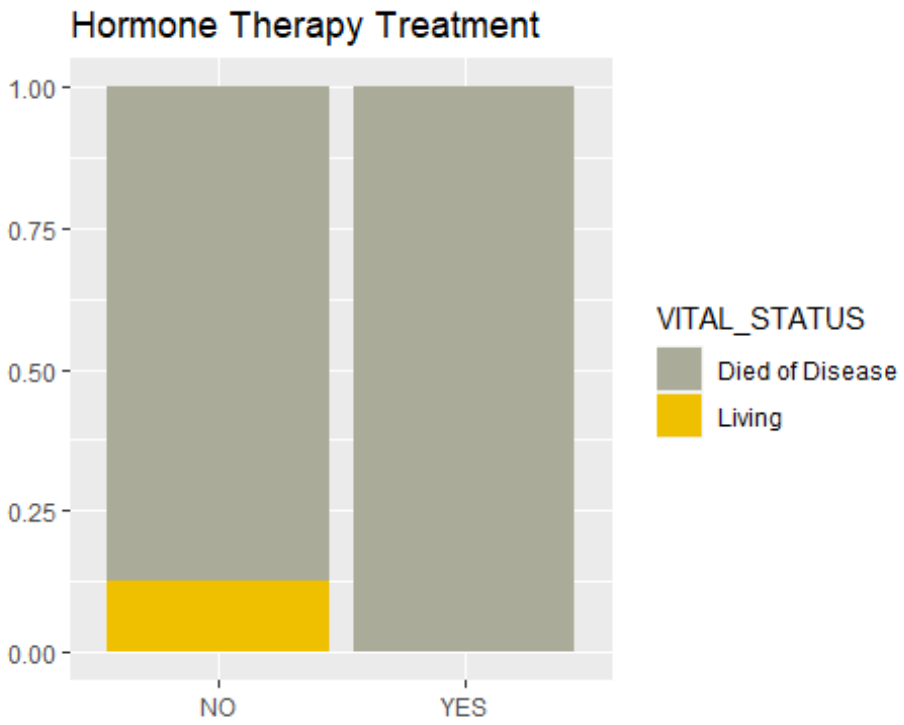
```
plot_stacked_barchart(clinic5$CHEMOTHERAPY, 'Chemotherapy Treatment')
```



* The group with no chemotherapy had more deaths

- 1.58. Vital Status vs Hormone Therapy

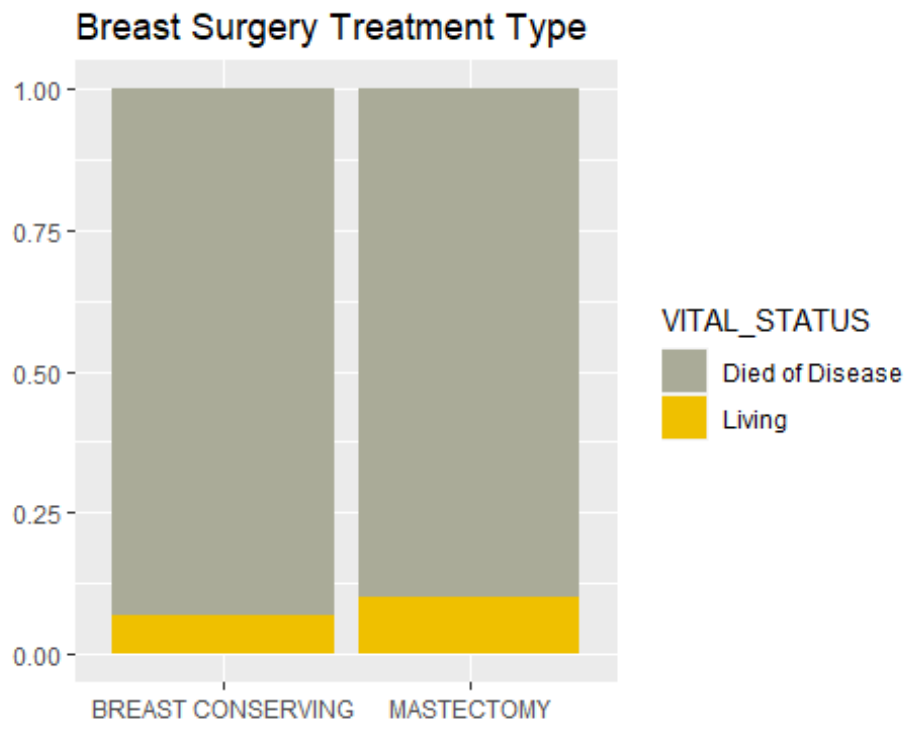
```
plot_stacked_barchart(clinic5$HORMONE_THERAPY, 'Hormone Therapy Treatment')
```



* All the patients who had hormone therapy died

- 1.59. Vital Status vs Breast Surgery

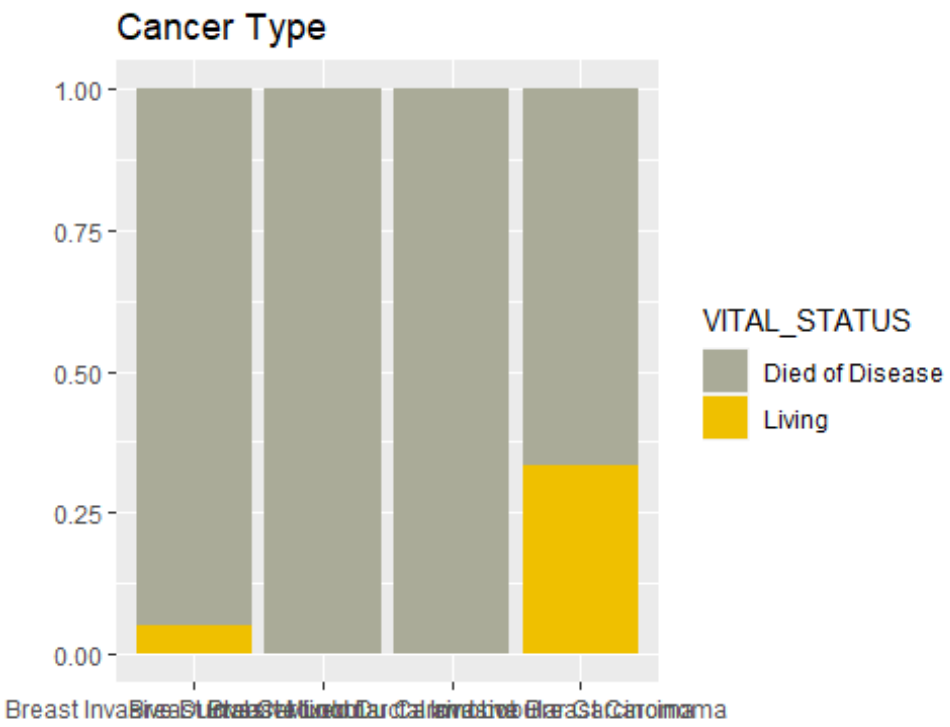
```
plot_stacked_barchart(clinic5$BREAST_SURGERY, 'Breast Surgery Treatment Type')
```



* The mastectomy group had a better outcome

- 1.60. Vital Status vs Cancer Type

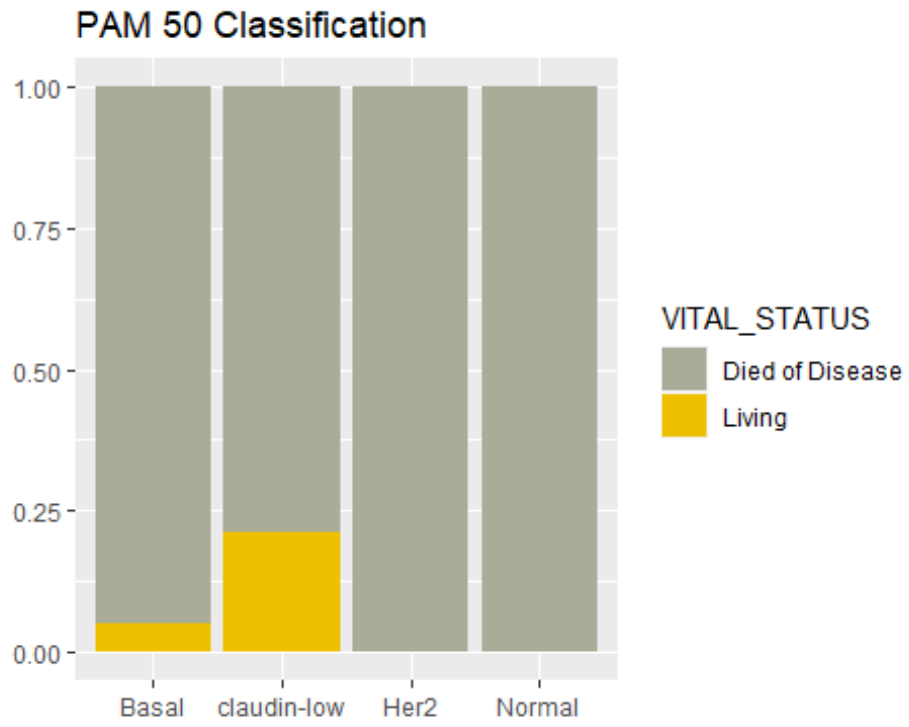
```
plot_stacked_barchart(clinic5$CANCER_TYPE, 'Cancer Type')
```



* The invasive breast carcinoma group had the best outcome

- 1.61. Vital Status vs PAM 50 Classification

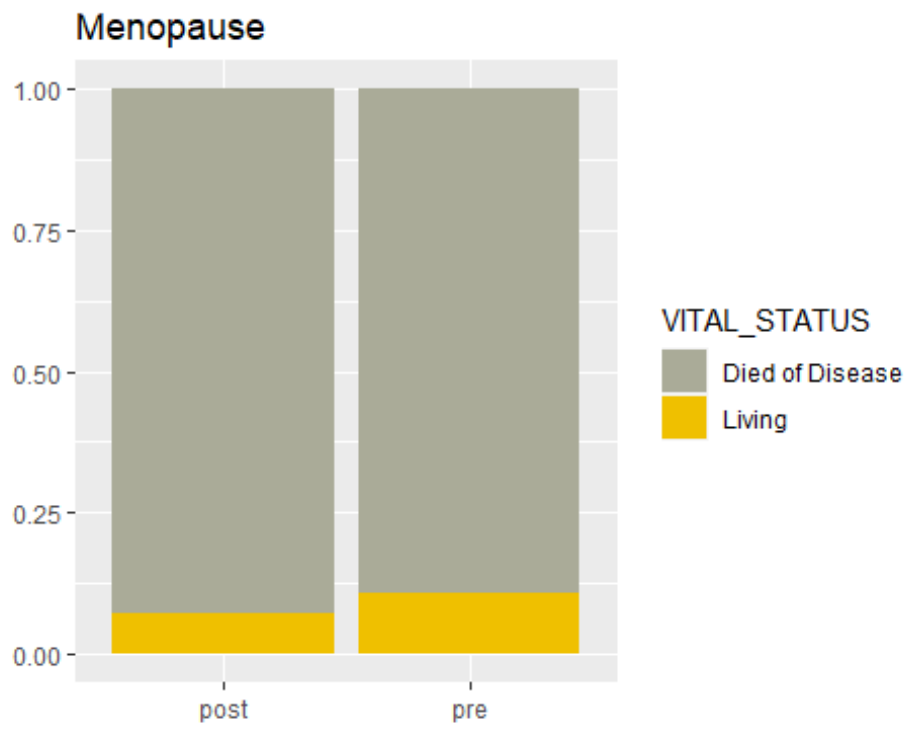
```
plot_stacked_barchart(clinic5$PAM50, 'PAM 50 Classification')
```



* The claudin-low group had the fewest deaths

- 1.62. Vital Status vs Menopause

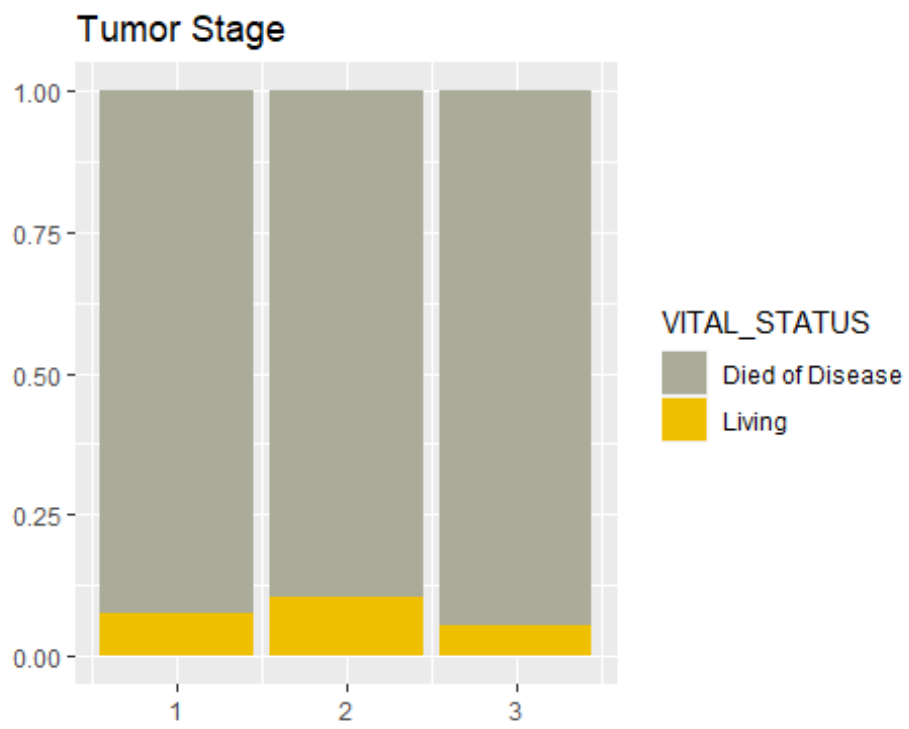
```
plot_stacked_barchart(clinic5$MENOPAUSE, 'Menopause')
```



* More deaths occured in the post-menopausal group

- 1.63. Vital Status vs Tumor Stage

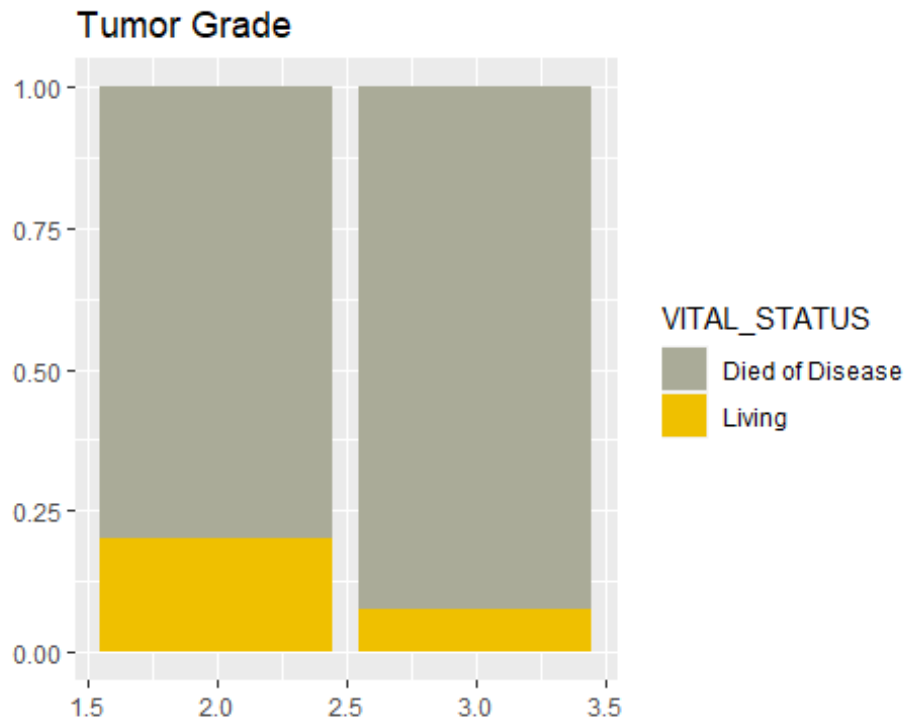
```
plot_stacked_barchart(clinic5$TUMOR_STAGE, 'Tumor Stage')
```



* Stage 2 had the best outcome

- 1.64. Vital Status vs Tumor grade

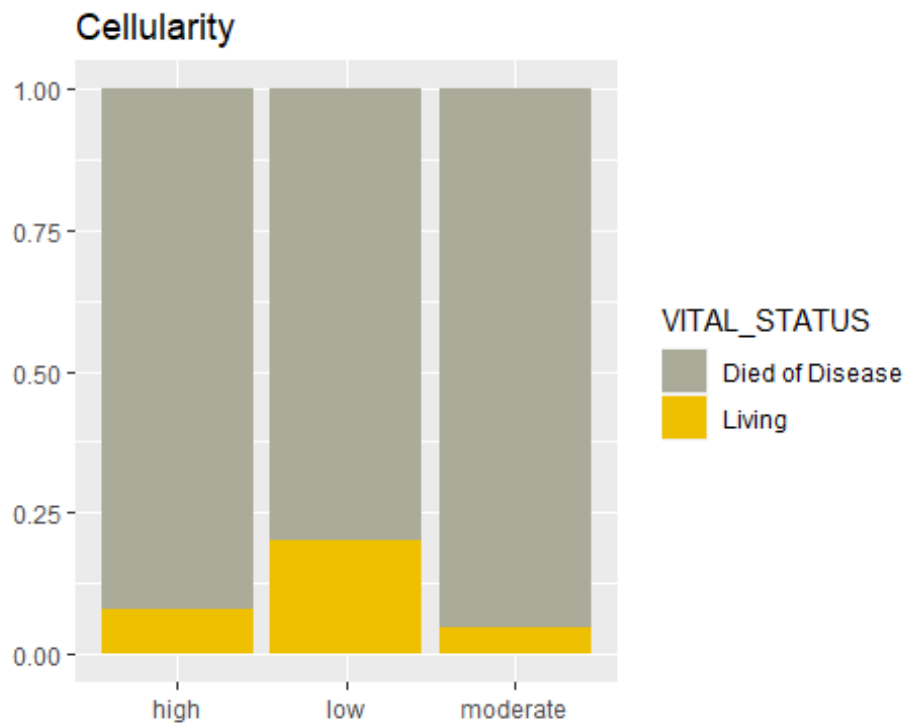
```
plot_stacked_barchart(clinic5$TUMOR_GRADE, 'Tumor Grade')
```



* Grade 2 had the best outcome

- 1.65. Vital Status vs Cellularity

```
plot_stacked_barchart(clinic5$CELLULARITY, 'Cellularity')
```



* Group with low cellularity had the best outcome

- 1.66. Use the psych package to carry out multivariate analysis
- 1.67. Carry out correlation analysis on the 190, 120 and 70-patient subgroups
- 1.68. Analysis is separately carried out on both the clinical and mutation feature sets
- By default corr.test produces pairwise "Pearson" correlation matrix for the entire dataset

```
library(psych) # multivariate analysis, FA and PCA

##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

corr.test(clin) # for all clinical variables

## Warning in cor(x, use = use, method = method): the standard deviation is zero

## Call:corr.test(x = clin)
## Correlation matrix
##                      AGE_AT_DIAGNOSIS AGE_GROUP BREAST_SURGERY CANCER_TYPE
## AGE_AT_DIAGNOSIS                 1.00      0.87          -0.09        0.07
## AGE_GROUP                        0.87      1.00          -0.09        0.04
## BREAST_SURGERY                  -0.09     -0.09           1.00        0.00
## CANCER_TYPE                      0.07      0.04           0.00        1.00
## CHEMOTHERAPY                     0.34      0.26           0.31       -0.10
## RADIO_THERAPY                      NA        NA             NA          NA
## HORMONE_THERAPY                 -0.21     -0.22          -0.02       -0.18
## PAM50                            0.21      0.18          -0.06        0.03
## COHORT                          -0.19     -0.13           0.23       -0.03
## ER_IHC                          -0.22     -0.24           0.00       -0.08
## TUMOR_GRADE                     -0.13     -0.12          -0.13       -0.19
## TUMOR_SIZE                       0.02      0.03          -0.18       -0.08
## TUMOR_STAGE                      0.00     -0.05          -0.25       -0.04
## HER2_SNP6                        0.06      0.04           0.08        0.10
```

```
## HISTOLOGICAL_SUBTYPE             0.19      0.15          -0.02        0.17
## CELLULARITY                     -0.13     -0.10           0.05       -0.05
## MENOPAUSE                        0.82      0.86          -0.11        0.08
## INTCLUST                        -0.21     -0.21           0.07       -0.09
## LATERALITY                       0.12      0.09           0.00        0.00
## MUTATION_COUNT                  -0.07     -0.08           0.14        0.25
## NPI                             -0.15     -0.15          -0.34       -0.05
## SURVIVAL_MONTHS                 -0.17     -0.11           0.11       -0.08
## SURVIVAL_GROUPS                 -0.12     -0.04           0.26       -0.01
## THREEGENE                        0.17      0.16          -0.04        0.22
## VITAL_STATUS                     0.11      0.02          -0.22        0.03
##                      CHEMOTHERAPY RADIO_THERAPY HORMONE_THERAPY PAM50 COHORT
## AGE_AT_DIAGNOSIS             0.34            NA           -0.21  0.21  -0.19
## AGE_GROUP                    0.26            NA           -0.22  0.18  -0.13
## BREAST_SURGERY               0.31            NA           -0.02 -0.06   0.23
## CANCER_TYPE                 -0.10            NA           -0.18  0.03  -0.03
## CHEMOTHERAPY                 1.00            NA           -0.12  0.10   0.03
## RADIO_THERAPY                  NA            NA              NA    NA     NA
## HORMONE_THERAPY             -0.12            NA            1.00 -0.21  -0.01
## PAM50                        0.10            NA           -0.21  1.00  -0.16
## COHORT                       0.03            NA           -0.01 -0.16   1.00
## ER_IHC                      -0.24            NA            0.38 -0.30   0.32
## TUMOR_GRADE                 -0.25            NA            0.19 -0.32   0.16
## TUMOR_SIZE                  -0.23            NA           -0.03  0.05   0.00
## TUMOR_STAGE                 -0.45            NA            0.05  0.10  -0.06
## HER2_SNP6                    0.12            NA           -0.06 -0.16  -0.11
## HISTOLOGICAL_SUBTYPE         0.06            NA           -0.05  0.29  -0.02
## CELLULARITY                 -0.19            NA            0.14 -0.24   0.16
## MENOPAUSE                    0.23            NA           -0.27  0.21  -0.16
## INTCLUST                    -0.11            NA            0.10 -0.43   0.14
## LATERALITY                   0.08            NA            0.01  0.03   0.05
## MUTATION_COUNT              -0.01            NA           -0.01 -0.12   0.31
## NPI                         -0.57            NA            0.00 -0.09   0.10
## SURVIVAL_MONTHS              0.14            NA            0.08 -0.01   0.08
## SURVIVAL_GROUPS              0.17            NA            0.05  0.03   0.03
## THREEGENE                    0.00            NA           -0.20  0.47  -0.08
## VITAL_STATUS                -0.10            NA           -0.13  0.08   0.00
##                      ER_IHC TUMOR_GRADE TUMOR_SIZE TUMOR_STAGE HER2_SNP6
## AGE_AT_DIAGNOSIS      -0.22       -0.13       0.02        0.00      0.06
## AGE_GROUP             -0.24       -0.12       0.03       -0.05      0.04
## BREAST_SURGERY         0.00       -0.13      -0.18       -0.25      0.08
## CANCER_TYPE           -0.08       -0.19      -0.08       -0.04      0.10
## CHEMOTHERAPY          -0.24       -0.25      -0.23       -0.45      0.12
## RADIO_THERAPY            NA          NA         NA          NA        NA
## HORMONE_THERAPY        0.38        0.19      -0.03        0.05     -0.06
## PAM50                 -0.30       -0.32       0.05        0.10     -0.16
## COHORT                 0.32        0.16       0.00       -0.06     -0.11
## ER_IHC                 1.00        0.42       0.00        0.12      0.00
## TUMOR_GRADE            0.42        1.00       0.11        0.09     -0.07
## TUMOR_SIZE             0.00        0.11       1.00        0.26     -0.04
## TUMOR_STAGE            0.12        0.09       0.26        1.00     -0.02
## HER2_SNP6              0.00       -0.07      -0.04       -0.02      1.00
## HISTOLOGICAL_SUBTYPE  -0.07       -0.28       0.08        0.10      0.08
## CELLULARITY            0.23        0.14       0.13       -0.07     -0.03
## MENOPAUSE             -0.20       -0.12       0.06       -0.03      0.03
## INTCLUST               0.24        0.27       0.03       -0.14      0.03
## LATERALITY             0.02       -0.02      -0.06       -0.10     -0.09
## MUTATION_COUNT         0.11        0.06       0.02        0.00      0.01
## NPI                    0.21        0.57       0.28        0.59     -0.13
## SURVIVAL_MONTHS        0.03       -0.08      -0.11       -0.11      0.03
## SURVIVAL_GROUPS       -0.06       -0.11      -0.16       -0.23      0.07
## THREEGENE             -0.34       -0.20       0.03        0.02     -0.20
## VITAL_STATUS          -0.04        0.04       0.10        0.16      0.03
##                      HISTOLOGICAL_SUBTYPE CELLULARITY MENOPAUSE INTCLUST
## AGE_AT_DIAGNOSIS                     0.19       -0.13      0.82    -0.21
## AGE_GROUP                            0.15       -0.10      0.86    -0.21
## BREAST_SURGERY                      -0.02        0.05     -0.11     0.07
## CANCER_TYPE                          0.17       -0.05      0.08    -0.09
## CHEMOTHERAPY                         0.06       -0.19      0.23    -0.11
## RADIO_THERAPY                          NA          NA        NA       NA
## HORMONE_THERAPY                     -0.05        0.14     -0.27     0.10
## PAM50                                0.29       -0.24      0.21    -0.43
## COHORT                              -0.02        0.16     -0.16     0.14
## ER_IHC                              -0.07        0.23     -0.20     0.24
## TUMOR_GRADE                         -0.28        0.14     -0.12     0.27
## TUMOR_SIZE                           0.08        0.13      0.06     0.03
## TUMOR_STAGE                          0.10       -0.07     -0.03    -0.14
## HER2_SNP6                            0.08       -0.03      0.03     0.03
## HISTOLOGICAL_SUBTYPE                 1.00       -0.03      0.14    -0.24
```

```
## CELLULARITY                        -0.03        1.00     -0.08     0.27
## MENOPAUSE                           0.14       -0.08      1.00    -0.17
## INTCLUST                           -0.24        0.27     -0.17     1.00
## LATERALITY                          0.00       -0.13      0.03     0.01
## MUTATION_COUNT                      0.04        0.14     -0.04     0.12
## NPI                                -0.10        0.06     -0.14     0.06
## SURVIVAL_MONTHS                     0.01        0.11     -0.09     0.14
## SURVIVAL_GROUPS                    -0.07        0.10     -0.05     0.09
## THREEGENE                           0.13       -0.17      0.13    -0.29
## VITAL_STATUS                        0.06       -0.10      0.04    -0.14
##                      LATERALITY MUTATION_COUNT   NPI SURVIVAL_MONTHS
## AGE_AT_DIAGNOSIS           0.12          -0.07 -0.15           -0.17
## AGE_GROUP                  0.09          -0.08 -0.15           -0.11
## BREAST_SURGERY             0.00           0.14 -0.34            0.11
## CANCER_TYPE                0.00           0.25 -0.05           -0.08
## CHEMOTHERAPY               0.08          -0.01 -0.57            0.14
## RADIO_THERAPY                NA             NA    NA              NA
## HORMONE_THERAPY            0.01          -0.01  0.00            0.08
## PAM50                      0.03          -0.12 -0.09           -0.01
## COHORT                     0.05           0.31  0.10            0.08
## ER_IHC                     0.02           0.11  0.21            0.03
## TUMOR_GRADE               -0.02           0.06  0.57           -0.08
## TUMOR_SIZE                -0.06           0.02  0.28           -0.11
## TUMOR_STAGE               -0.10           0.00  0.59           -0.11
## HER2_SNP6                 -0.09           0.01 -0.13            0.03
## HISTOLOGICAL_SUBTYPE       0.00           0.04 -0.10            0.01
## CELLULARITY               -0.13           0.14  0.06            0.11
## MENOPAUSE                  0.03          -0.04 -0.14           -0.09
## INTCLUST                   0.01           0.12  0.06            0.14
## LATERALITY                 1.00          -0.02 -0.02           -0.05
## MUTATION_COUNT            -0.02           1.00  0.09            0.05
## NPI                       -0.02           0.09  1.00           -0.17
## SURVIVAL_MONTHS           -0.05           0.05 -0.17            1.00
## SURVIVAL_GROUPS           -0.08           0.07 -0.28            0.78
## THREEGENE                  0.01           0.11 -0.06           -0.16
## VITAL_STATUS               0.09           0.05  0.25           -0.68
##                      SURVIVAL_GROUPS THREEGENE VITAL_STATUS
## AGE_AT_DIAGNOSIS               -0.12      0.17         0.11
## AGE_GROUP                      -0.04      0.16         0.02
## BREAST_SURGERY                  0.26     -0.04        -0.22
## CANCER_TYPE                    -0.01      0.22         0.03
## CHEMOTHERAPY                    0.17      0.00        -0.10
## RADIO_THERAPY                     NA        NA           NA
## HORMONE_THERAPY                 0.05     -0.20        -0.13
## PAM50                           0.03      0.47         0.08
## COHORT                          0.03     -0.08         0.00
## ER_IHC                         -0.06     -0.34        -0.04
## TUMOR_GRADE                    -0.11     -0.20         0.04
## TUMOR_SIZE                     -0.16      0.03         0.10
## TUMOR_STAGE                    -0.23      0.02         0.16
## HER2_SNP6                       0.07     -0.20         0.03
## HISTOLOGICAL_SUBTYPE           -0.07      0.13         0.06
## CELLULARITY                     0.10     -0.17        -0.10
## MENOPAUSE                      -0.05      0.13         0.04
## INTCLUST                        0.09     -0.29        -0.14
## LATERALITY                     -0.08      0.01         0.09
## MUTATION_COUNT                  0.07      0.11         0.05
## NPI                            -0.28     -0.06         0.25
## SURVIVAL_MONTHS                 0.78     -0.16        -0.68
## SURVIVAL_GROUPS                 1.00     -0.06        -0.74
## THREEGENE                      -0.06      1.00         0.05
## VITAL_STATUS                   -0.74      0.05         1.00
## Sample Size
## [1] 190
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##                      AGE_AT_DIAGNOSIS AGE_GROUP BREAST_SURGERY CANCER_TYPE
## AGE_AT_DIAGNOSIS                 0.00      0.00           1.00        1.00
## AGE_GROUP                        0.00      0.00           1.00        1.00
## BREAST_SURGERY                   0.20      0.21           0.00        1.00
## CANCER_TYPE                      0.31      0.60           0.97        0.00
## CHEMOTHERAPY                     0.00      0.00           0.00        0.18
## RADIO_THERAPY                      NA        NA             NA          NA
## HORMONE_THERAPY                  0.00      0.00           0.77        0.01
## PAM50                            0.00      0.01           0.40        0.73
## COHORT                           0.01      0.08           0.00        0.64
## ER_IHC                           0.00      0.00           0.98        0.25
## TUMOR_GRADE                      0.07      0.11           0.06        0.01
## TUMOR_SIZE                       0.82      0.73           0.01        0.27
## TUMOR_STAGE                      0.97      0.46           0.00        0.58
```

```
## HER2_SNP6                        0.41      0.61           0.25        0.19
## HISTOLOGICAL_SUBTYPE             0.01      0.04           0.81        0.02
## CELLULARITY                      0.07      0.16           0.47        0.48
## MENOPAUSE                        0.00      0.00           0.12        0.25
## INTCLUST                         0.00      0.00           0.37        0.20
## LATERALITY                       0.11      0.21           0.98        0.99
## MUTATION_COUNT                   0.32      0.25           0.05        0.00
## NPI                              0.03      0.04           0.00        0.51
## SURVIVAL_MONTHS                  0.02      0.12           0.15        0.26
## SURVIVAL_GROUPS                  0.10      0.58           0.00        0.90
## THREEGENE                        0.02      0.03           0.56        0.00
## VITAL_STATUS                     0.13      0.74           0.00        0.71
##                      CHEMOTHERAPY RADIO_THERAPY HORMONE_THERAPY PAM50 COHORT
## AGE_AT_DIAGNOSIS             0.00            NA            0.73  0.96   1.00
## AGE_GROUP                    0.06            NA            0.43  1.00   1.00
## BREAST_SURGERY               0.00            NA            1.00  1.00   0.28
## CANCER_TYPE                  1.00            NA            1.00  1.00   1.00
## CHEMOTHERAPY                 0.00            NA            1.00  1.00   1.00
## RADIO_THERAPY                  NA            NA              NA    NA     NA
## HORMONE_THERAPY              0.09            NA            0.00  0.69   1.00
## PAM50                        0.19            NA            0.00  0.00   1.00
## COHORT                       0.72            NA            0.91  0.03   0.00
## ER_IHC                       0.00            NA            0.00  0.00   0.00
## TUMOR_GRADE                  0.00            NA            0.01  0.00   0.03
## TUMOR_SIZE                   0.00            NA            0.65  0.48   0.98
## TUMOR_STAGE                  0.00            NA            0.49  0.17   0.42
## HER2_SNP6                    0.10            NA            0.43  0.03   0.12
## HISTOLOGICAL_SUBTYPE         0.44            NA            0.54  0.00   0.80
## CELLULARITY                  0.01            NA            0.05  0.00   0.03
## MENOPAUSE                    0.00            NA            0.00  0.00   0.03
## INTCLUST                     0.13            NA            0.19  0.00   0.05
## LATERALITY                   0.25            NA            0.93  0.72   0.47
## MUTATION_COUNT               0.94            NA            0.91  0.09   0.00
## NPI                          0.00            NA            0.97  0.24   0.15
## SURVIVAL_MONTHS              0.06            NA            0.30  0.86   0.29
## SURVIVAL_GROUPS              0.02            NA            0.49  0.64   0.67
## THREEGENE                    0.97            NA            0.01  0.00   0.30
## VITAL_STATUS                 0.19            NA            0.08  0.26   0.98
##                      ER_IHC TUMOR_GRADE TUMOR_SIZE TUMOR_STAGE HER2_SNP6
## AGE_AT_DIAGNOSIS       0.57        1.00       1.00        1.00      1.00
## AGE_GROUP              0.18        1.00       1.00        1.00      1.00
## BREAST_SURGERY         1.00        1.00       1.00        0.10      1.00
## CANCER_TYPE            1.00        1.00       1.00        1.00      1.00
## CHEMOTHERAPY           0.18        0.10       0.36        0.00      1.00
## RADIO_THERAPY            NA          NA         NA          NA        NA
## HORMONE_THERAPY        0.00        1.00       1.00        1.00      1.00
## PAM50                  0.01        0.00       1.00        1.00      1.00
## COHORT                 0.00        1.00       1.00        1.00      1.00
## ER_IHC                 0.00        0.00       1.00        1.00      1.00
## TUMOR_GRADE            0.00        0.00       1.00        1.00      1.00
## TUMOR_SIZE             0.98        0.12       0.00        0.07      1.00
## TUMOR_STAGE            0.11        0.20       0.00        0.00      1.00
## HER2_SNP6              1.00        0.37       0.58        0.82      0.00
## HISTOLOGICAL_SUBTYPE   0.37        0.00       0.26        0.18      0.28
## CELLULARITY            0.00        0.06       0.07        0.36      0.67
## MENOPAUSE              0.01        0.09       0.39        0.73      0.72
## INTCLUST               0.00        0.00       0.73        0.06      0.64
## LATERALITY             0.79        0.74       0.40        0.18      0.21
## MUTATION_COUNT         0.12        0.43       0.74        0.99      0.92
## NPI                    0.00        0.00       0.00        0.00      0.07
## SURVIVAL_MONTHS        0.71        0.26       0.13        0.13      0.64
## SURVIVAL_GROUPS        0.40        0.13       0.03        0.00      0.37
## THREEGENE              0.00        0.00       0.66        0.83      0.01
## VITAL_STATUS           0.62        0.57       0.18        0.03      0.70
##                      HISTOLOGICAL_SUBTYPE CELLULARITY MENOPAUSE INTCLUST
## AGE_AT_DIAGNOSIS                     1.00        1.00      0.00     0.66
## AGE_GROUP                            1.00        1.00      0.00     0.97
## BREAST_SURGERY                       1.00        1.00      1.00     1.00
## CANCER_TYPE                          1.00        1.00      1.00     1.00
## CHEMOTHERAPY                         1.00        1.00      0.39     1.00
## RADIO_THERAPY                          NA          NA        NA       NA
## HORMONE_THERAPY                      1.00        1.00      0.05     1.00
## PAM50                                0.01        0.25      0.94     0.00
## COHORT                               1.00        1.00      1.00     1.00
## ER_IHC                               1.00        0.39      1.00     0.17
## TUMOR_GRADE                          0.02        1.00      1.00     0.05
## TUMOR_SIZE                           1.00        1.00      1.00     1.00
## TUMOR_STAGE                          1.00        1.00      1.00     1.00
## HER2_SNP6                            1.00        1.00      1.00     1.00
```

```
## HISTOLOGICAL_SUBTYPE                 0.00        1.00      1.00     0.16
## CELLULARITY                          0.67        0.00      1.00     0.05
## MENOPAUSE                            0.05        0.25      0.00     1.00
## INTCLUST                             0.00        0.00      0.02     0.00
## LATERALITY                           0.96        0.07      0.68     0.88
## MUTATION_COUNT                       0.57        0.06      0.61     0.10
## NPI                                  0.18        0.37      0.05     0.40
## SURVIVAL_MONTHS                      0.94        0.12      0.23     0.06
## SURVIVAL_GROUPS                      0.31        0.19      0.50     0.21
## THREEGENE                            0.07        0.02      0.07     0.00
## VITAL_STATUS                         0.41        0.18      0.59     0.05
##                      LATERALITY MUTATION_COUNT  NPI SURVIVAL_MONTHS
## AGE_AT_DIAGNOSIS           1.00           1.00 1.00            1.00
## AGE_GROUP                  1.00           1.00 1.00            1.00
## BREAST_SURGERY             1.00           1.00 0.00            1.00
## CANCER_TYPE                1.00           0.11 1.00            1.00
## CHEMOTHERAPY               1.00           1.00 0.00            1.00
## RADIO_THERAPY                NA             NA   NA              NA
## HORMONE_THERAPY            1.00           1.00 1.00            1.00
## PAM50                      1.00           1.00 1.00            1.00
## COHORT                     1.00           0.00 1.00            1.00
## ER_IHC                     1.00           1.00 0.75            1.00
## TUMOR_GRADE                1.00           1.00 0.00            1.00
## TUMOR_SIZE                 1.00           1.00 0.02            1.00
## TUMOR_STAGE                1.00           1.00 0.00            1.00
## HER2_SNP6                  1.00           1.00 1.00            1.00
## HISTOLOGICAL_SUBTYPE       1.00           1.00 1.00            1.00
## CELLULARITY                1.00           1.00 1.00            1.00
## MENOPAUSE                  1.00           1.00 1.00            1.00
## INTCLUST                   1.00           1.00 1.00            1.00
## LATERALITY                 0.00           1.00 1.00            1.00
## MUTATION_COUNT             0.74           0.00 1.00            1.00
## NPI                        0.77           0.21 0.00            1.00
## SURVIVAL_MONTHS            0.51           0.51 0.02            0.00
## SURVIVAL_GROUPS            0.28           0.34 0.00            0.00
## THREEGENE                  0.91           0.15 0.44            0.03
## VITAL_STATUS               0.24           0.52 0.00            0.00
##                      SURVIVAL_GROUPS THREEGENE VITAL_STATUS
## AGE_AT_DIAGNOSIS                1.00      1.00         1.00
## AGE_GROUP                       1.00      1.00         1.00
## BREAST_SURGERY                  0.08      1.00         0.43
## CANCER_TYPE                     1.00      0.51         1.00
## CHEMOTHERAPY                    1.00      1.00         1.00
## RADIO_THERAPY                     NA        NA           NA
## HORMONE_THERAPY                 1.00      1.00         1.00
## PAM50                           1.00      0.00         1.00
## COHORT                          1.00      1.00         1.00
## ER_IHC                          1.00      0.00         1.00
## TUMOR_GRADE                     1.00      1.00         1.00
## TUMOR_SIZE                      1.00      1.00         1.00
## TUMOR_STAGE                     0.38      1.00         1.00
## HER2_SNP6                       1.00      1.00         1.00
## HISTOLOGICAL_SUBTYPE            1.00      1.00         1.00
## CELLULARITY                     1.00      1.00         1.00
## MENOPAUSE                       1.00      1.00         1.00
## INTCLUST                        1.00      0.01         1.00
## LATERALITY                      1.00      1.00         1.00
## MUTATION_COUNT                  1.00      1.00         1.00
## NPI                             0.02      1.00         0.10
## SURVIVAL_MONTHS                 0.00      1.00         0.00
## SURVIVAL_GROUPS                 0.00      1.00         0.00
## THREEGENE                       0.37      0.00         1.00
## VITAL_STATUS                    0.00      0.52         0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option

corr.test((clin[(SURVIVAL_MONTHS<60.1),])) # for clinical variables of 70 patients with <= 5years follow-up

## Warning in cor(x, use = use, method = method): the standard deviation is zero

## Call:corr.test(x = (clin[(SURVIVAL_MONTHS < 60.1), ]))
## Correlation matrix
##                      AGE_AT_DIAGNOSIS AGE_GROUP BREAST_SURGERY CANCER_TYPE
## AGE_AT_DIAGNOSIS                 1.00      0.85          -0.19        0.08
## AGE_GROUP                        0.85      1.00          -0.21        0.02
## BREAST_SURGERY                  -0.19     -0.21           1.00       -0.13
## CANCER_TYPE                      0.08      0.02          -0.13        1.00
## CHEMOTHERAPY                     0.33      0.25           0.25       -0.01
```

```
## RADIO_THERAPY                      NA        NA             NA          NA
## HORMONE_THERAPY                 -0.13     -0.16          -0.16       -0.08
## PAM50                            0.27      0.18          -0.20       -0.11
## COHORT                          -0.20     -0.14           0.26       -0.05
## ER_IHC                          -0.21     -0.29           0.13       -0.22
## TUMOR_GRADE                     -0.17     -0.08           0.02       -0.03
## TUMOR_SIZE                       0.22      0.25          -0.34       -0.20
## TUMOR_STAGE                      0.10     -0.02          -0.22       -0.07
## HER2_SNP6                        0.09      0.07           0.07        0.17
## HISTOLOGICAL_SUBTYPE             0.31      0.20          -0.06        0.06
## CELLULARITY                     -0.07     -0.05          -0.04       -0.02
## MENOPAUSE                        0.80      0.88          -0.18       -0.03
## INTCLUST                        -0.21     -0.21           0.24       -0.04
## LATERALITY                      -0.04     -0.07          -0.04       -0.10
## MUTATION_COUNT                   0.08      0.01           0.14        0.41
## NPI                             -0.11     -0.13          -0.28        0.05
## SURVIVAL_MONTHS                 -0.05      0.04           0.16       -0.14
## SURVIVAL_GROUPS                    NA        NA             NA          NA
## THREEGENE                        0.23      0.17          -0.12        0.26
## VITAL_STATUS                     0.12      0.07           0.06        0.27
##                       CHEMOTHERAPY RADIO_THERAPY HORMONE_THERAPY PAM50 COHORT
## AGE_AT_DIAGNOSIS              0.33            NA           -0.13  0.27  -0.20
## AGE_GROUP                     0.25            NA           -0.16  0.18  -0.14
## BREAST_SURGERY                0.25            NA           -0.16 -0.20   0.26
## CANCER_TYPE                  -0.01            NA           -0.08 -0.11  -0.05
## CHEMOTHERAPY                  1.00            NA           -0.25  0.13  -0.03
## RADIO_THERAPY                   NA            NA              NA    NA     NA
## HORMONE_THERAPY              -0.25            NA            1.00 -0.10  -0.19
## PAM50                         0.13            NA           -0.10  1.00  -0.38
## COHORT                       -0.03            NA           -0.19 -0.38   1.00
## ER_IHC                       -0.25            NA            0.43 -0.39   0.25
## TUMOR_GRADE                  -0.26            NA           -0.07 -0.41   0.20
## TUMOR_SIZE                   -0.13            NA           -0.10  0.34  -0.29
## TUMOR_STAGE                  -0.32            NA            0.31  0.26  -0.15
## HER2_SNP6                     0.08            NA            0.03 -0.16  -0.09
## HISTOLOGICAL_SUBTYPE          0.08            NA            0.13  0.48  -0.12
## CELLULARITY                  -0.34            NA            0.08 -0.13   0.14
## MENOPAUSE                     0.28            NA           -0.18  0.22  -0.19
## INTCLUST                     -0.12            NA           -0.05 -0.48   0.22
## LATERALITY                   -0.03            NA            0.08 -0.10   0.06
## MUTATION_COUNT                0.16            NA           -0.26 -0.08   0.10
## NPI                          -0.47            NA           -0.06  0.06   0.04
## SURVIVAL_MONTHS               0.04            NA           -0.05 -0.14   0.06
## SURVIVAL_GROUPS                 NA            NA              NA    NA     NA
## THREEGENE                     0.03            NA           -0.26  0.41  -0.08
## VITAL_STATUS                  0.04            NA           -0.21 -0.01   0.08
##                       ER_IHC TUMOR_GRADE TUMOR_SIZE TUMOR_STAGE HER2_SNP6
## AGE_AT_DIAGNOSIS      -0.21       -0.17       0.22        0.10      0.09
## AGE_GROUP             -0.29       -0.08       0.25       -0.02      0.07
## BREAST_SURGERY         0.13        0.02      -0.34       -0.22      0.07
## CANCER_TYPE           -0.22       -0.03      -0.20       -0.07      0.17
## CHEMOTHERAPY          -0.25       -0.26      -0.13       -0.32      0.08
## RADIO_THERAPY            NA          NA         NA          NA        NA
## HORMONE_THERAPY        0.43       -0.07      -0.10        0.31      0.03
## PAM50                 -0.39       -0.41       0.34        0.26     -0.16
## COHORT                 0.25        0.20      -0.29       -0.15     -0.09
## ER_IHC                 1.00        0.25      -0.32        0.17     -0.06
## TUMOR_GRADE            0.25        1.00      -0.07       -0.05     -0.04
## TUMOR_SIZE            -0.32       -0.07       1.00        0.12     -0.05
## TUMOR_STAGE            0.17       -0.05       0.12        1.00      0.06
## HER2_SNP6             -0.06       -0.04      -0.05        0.06      1.00
## HISTOLOGICAL_SUBTYPE  -0.17       -0.40       0.15        0.22      0.03
## CELLULARITY            0.14        0.15       0.17       -0.03     -0.17
## MENOPAUSE             -0.29       -0.11       0.25        0.00      0.06
## INTCLUST               0.21        0.40      -0.09       -0.31      0.10
## LATERALITY             0.04       -0.02       0.10       -0.12      0.05
## MUTATION_COUNT        -0.19        0.14      -0.05        0.00      0.07
## NPI                    0.04        0.49       0.21        0.48     -0.14
## SURVIVAL_MONTHS       -0.07       -0.11      -0.20       -0.21     -0.09
## SURVIVAL_GROUPS          NA          NA         NA          NA        NA
## THREEGENE             -0.50       -0.23       0.20        0.09     -0.13
## VITAL_STATUS          -0.11        0.11       0.08        0.03      0.29
##                       HISTOLOGICAL_SUBTYPE CELLULARITY MENOPAUSE INTCLUST
## AGE_AT_DIAGNOSIS                      0.31       -0.07      0.80    -0.21
## AGE_GROUP                             0.20       -0.05      0.88    -0.21
## BREAST_SURGERY                       -0.06       -0.04     -0.18     0.24
## CANCER_TYPE                           0.06       -0.02     -0.03    -0.04
## CHEMOTHERAPY                          0.08       -0.34      0.28    -0.12
## RADIO_THERAPY                           NA          NA        NA       NA
```

```
## HORMONE_THERAPY                      0.13        0.08     -0.18    -0.05
## PAM50                                0.48       -0.13      0.22    -0.48
## COHORT                              -0.12        0.14     -0.19     0.22
## ER_IHC                              -0.17        0.14     -0.29     0.21
## TUMOR_GRADE                         -0.40        0.15     -0.11     0.40
## TUMOR_SIZE                           0.15        0.17      0.25    -0.09
## TUMOR_STAGE                          0.22       -0.03      0.00    -0.31
## HER2_SNP6                            0.03       -0.17      0.06     0.10
## HISTOLOGICAL_SUBTYPE                 1.00       -0.03      0.15    -0.35
## CELLULARITY                         -0.03        1.00     -0.07     0.26
## MENOPAUSE                            0.15       -0.07      1.00    -0.19
## INTCLUST                            -0.35        0.26     -0.19     1.00
## LATERALITY                          -0.10       -0.10     -0.08     0.13
## MUTATION_COUNT                      -0.01        0.20      0.05     0.17
## NPI                                 -0.05        0.13     -0.17    -0.04
## SURVIVAL_MONTHS                     -0.12        0.03      0.05     0.04
## SURVIVAL_GROUPS                        NA          NA        NA       NA
## THREEGENE                            0.16        0.01      0.16    -0.26
## VITAL_STATUS                         0.06        0.10      0.06     0.07
##                      LATERALITY MUTATION_COUNT   NPI SURVIVAL_MONTHS
## AGE_AT_DIAGNOSIS          -0.04           0.08 -0.11           -0.05
## AGE_GROUP                 -0.07           0.01 -0.13            0.04
## BREAST_SURGERY            -0.04           0.14 -0.28            0.16
## CANCER_TYPE               -0.10           0.41  0.05           -0.14
## CHEMOTHERAPY              -0.03           0.16 -0.47            0.04
## RADIO_THERAPY                NA             NA    NA              NA
## HORMONE_THERAPY            0.08          -0.26 -0.06           -0.05
## PAM50                     -0.10          -0.08  0.06           -0.14
## COHORT                     0.06           0.10  0.04            0.06
## ER_IHC                     0.04          -0.19  0.04           -0.07
## TUMOR_GRADE               -0.02           0.14  0.49           -0.11
## TUMOR_SIZE                 0.10          -0.05  0.21           -0.20
## TUMOR_STAGE               -0.12           0.00  0.48           -0.21
## HER2_SNP6                  0.05           0.07 -0.14           -0.09
## HISTOLOGICAL_SUBTYPE      -0.10          -0.01 -0.05           -0.12
## CELLULARITY               -0.10           0.20  0.13            0.03
## MENOPAUSE                 -0.08           0.05 -0.17            0.05
## INTCLUST                   0.13           0.17 -0.04            0.04
## LATERALITY                 1.00          -0.04  0.02            0.02
## MUTATION_COUNT            -0.04           1.00  0.11           -0.07
## NPI                        0.02           0.11  1.00           -0.22
## SURVIVAL_MONTHS            0.02          -0.07 -0.22            1.00
## SURVIVAL_GROUPS              NA             NA    NA              NA
## THREEGENE                 -0.04           0.25  0.07            0.06
## VITAL_STATUS               0.04           0.21  0.23           -0.24
##                      SURVIVAL_GROUPS THREEGENE VITAL_STATUS
## AGE_AT_DIAGNOSIS                  NA      0.23         0.12
## AGE_GROUP                         NA      0.17         0.07
## BREAST_SURGERY                    NA     -0.12         0.06
## CANCER_TYPE                       NA      0.26         0.27
## CHEMOTHERAPY                      NA      0.03         0.04
## RADIO_THERAPY                     NA        NA           NA
## HORMONE_THERAPY                   NA     -0.26        -0.21
## PAM50                             NA      0.41        -0.01
## COHORT                            NA     -0.08         0.08
## ER_IHC                            NA     -0.50        -0.11
## TUMOR_GRADE                       NA     -0.23         0.11
## TUMOR_SIZE                        NA      0.20         0.08
## TUMOR_STAGE                       NA      0.09         0.03
## HER2_SNP6                         NA     -0.13         0.29
## HISTOLOGICAL_SUBTYPE              NA      0.16         0.06
## CELLULARITY                       NA      0.01         0.10
## MENOPAUSE                         NA      0.16         0.06
## INTCLUST                          NA     -0.26         0.07
## LATERALITY                        NA     -0.04         0.04
## MUTATION_COUNT                    NA      0.25         0.21
## NPI                               NA      0.07         0.23
## SURVIVAL_MONTHS                   NA      0.06        -0.24
## SURVIVAL_GROUPS                   NA        NA           NA
## THREEGENE                         NA      1.00         0.10
## VITAL_STATUS                      NA      0.10         1.00
## Sample Size
## [1] 70
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##                      AGE_AT_DIAGNOSIS AGE_GROUP BREAST_SURGERY CANCER_TYPE
## AGE_AT_DIAGNOSIS                 0.00      0.00           1.00        1.00
## AGE_GROUP                        0.00      0.00           1.00        1.00
## BREAST_SURGERY                   0.12      0.08           0.00        1.00
## CANCER_TYPE                      0.49      0.90           0.28        0.00
```

```
## CHEMOTHERAPY                    0.01      0.04           0.04        0.94
## RADIO_THERAPY                     NA        NA             NA          NA
## HORMONE_THERAPY                 0.29      0.18           0.19        0.53
## PAM50                           0.02      0.14           0.10        0.38
## COHORT                          0.10      0.23           0.03        0.69
## ER_IHC                          0.08      0.01           0.28        0.07
## TUMOR_GRADE                     0.17      0.48           0.90        0.80
## TUMOR_SIZE                      0.07      0.04           0.00        0.10
## TUMOR_STAGE                     0.41      0.86           0.06        0.55
## HER2_SNP6                       0.48      0.56           0.55        0.16
## HISTOLOGICAL_SUBTYPE            0.01      0.10           0.64        0.62
## CELLULARITY                     0.58      0.68           0.74        0.90
## MENOPAUSE                       0.00      0.00           0.14        0.78
## INTCLUST                        0.09      0.08           0.05        0.76
## LATERALITY                      0.72      0.57           0.73        0.40
## MUTATION_COUNT                  0.50      0.93           0.24        0.00
## NPI                             0.38      0.27           0.02        0.66
## SURVIVAL_MONTHS                 0.69      0.75           0.19        0.25
## SURVIVAL_GROUPS                   NA        NA             NA          NA
## THREEGENE                       0.06      0.17           0.31        0.03
## VITAL_STATUS                    0.32      0.55           0.63        0.02
##                      CHEMOTHERAPY RADIO_THERAPY HORMONE_THERAPY PAM50 COHORT
## AGE_AT_DIAGNOSIS             1.00            NA            1.00  1.00   1.00
## AGE_GROUP                    1.00            NA            1.00  1.00   1.00
## BREAST_SURGERY               1.00            NA            1.00  1.00   1.00
## CANCER_TYPE                  1.00            NA            1.00  1.00   1.00
## CHEMOTHERAPY                 0.00            NA            1.00  1.00   1.00
## RADIO_THERAPY                  NA            NA              NA    NA     NA
## HORMONE_THERAPY              0.04            NA            0.00  1.00   1.00
## PAM50                        0.29            NA            0.43  0.00   0.31
## COHORT                       0.80            NA            0.11  0.00   0.00
## ER_IHC                       0.04            NA            0.00  0.00   0.04
## TUMOR_GRADE                  0.03            NA            0.57  0.00   0.10
## TUMOR_SIZE                   0.29            NA            0.43  0.00   0.02
## TUMOR_STAGE                  0.01            NA            0.01  0.03   0.22
## HER2_SNP6                    0.50            NA            0.81  0.17   0.44
## HISTOLOGICAL_SUBTYPE         0.52            NA            0.30  0.00   0.33
## CELLULARITY                  0.00            NA            0.53  0.28   0.24
## MENOPAUSE                    0.02            NA            0.15  0.06   0.12
## INTCLUST                     0.32            NA            0.69  0.00   0.07
## LATERALITY                   0.84            NA            0.52  0.42   0.61
## MUTATION_COUNT               0.18            NA            0.03  0.51   0.39
## NPI                          0.00            NA            0.61  0.64   0.72
## SURVIVAL_MONTHS              0.72            NA            0.69  0.25   0.60
## SURVIVAL_GROUPS                NA            NA              NA    NA     NA
## THREEGENE                    0.81            NA            0.03  0.00   0.48
## VITAL_STATUS                 0.77            NA            0.09  0.95   0.53
##                      ER_IHC TUMOR_GRADE TUMOR_SIZE TUMOR_STAGE HER2_SNP6
## AGE_AT_DIAGNOSIS       1.00        1.00       1.00        1.00      1.00
## AGE_GROUP              1.00        1.00       1.00        1.00      1.00
## BREAST_SURGERY         1.00        1.00       0.99        1.00      1.00
## CANCER_TYPE            1.00        1.00       1.00        1.00      1.00
## CHEMOTHERAPY           1.00        1.00       1.00        1.00      1.00
## RADIO_THERAPY            NA          NA         NA          NA        NA
## HORMONE_THERAPY        0.04        1.00       1.00        1.00      1.00
## PAM50                  0.22        0.09       0.90        1.00      1.00
## COHORT                 1.00        1.00       1.00        1.00      1.00
## ER_IHC                 0.00        1.00       1.00        1.00      1.00
## TUMOR_GRADE            0.04        0.00       1.00        1.00      1.00
## TUMOR_SIZE             0.01        0.59       0.00        1.00      1.00
## TUMOR_STAGE            0.15        0.66       0.33        0.00      1.00
## HER2_SNP6              0.64        0.72       0.68        0.60      0.00
## HISTOLOGICAL_SUBTYPE   0.16        0.00       0.21        0.06      0.83
## CELLULARITY            0.26        0.21       0.16        0.81      0.15
## MENOPAUSE              0.01        0.35       0.04        1.00      0.60
## INTCLUST               0.08        0.00       0.48        0.01      0.43
## LATERALITY             0.73        0.84       0.43        0.32      0.68
## MUTATION_COUNT         0.11        0.24       0.70        0.98      0.56
## NPI                    0.75        0.00       0.08        0.00      0.26
## SURVIVAL_MONTHS        0.54        0.35       0.10        0.08      0.44
## SURVIVAL_GROUPS          NA          NA         NA          NA        NA
## THREEGENE              0.00        0.05       0.10        0.45      0.28
## VITAL_STATUS           0.36        0.35       0.50        0.79      0.02
##                      HISTOLOGICAL_SUBTYPE CELLULARITY MENOPAUSE INTCLUST
## AGE_AT_DIAGNOSIS                     1.00        1.00      0.00     1.00
## AGE_GROUP                            1.00        1.00      0.00     1.00
## BREAST_SURGERY                       1.00        1.00      1.00     1.00
## CANCER_TYPE                          1.00        1.00      1.00     1.00
## CHEMOTHERAPY                         1.00        1.00      1.00     1.00
```

```
## RADIO_THERAPY                          NA          NA        NA       NA
## HORMONE_THERAPY                      1.00        1.00      1.00     1.00
## PAM50                                0.01        1.00      1.00     0.01
## COHORT                               1.00        1.00      1.00     1.00
## ER_IHC                               1.00        1.00      1.00     1.00
## TUMOR_GRADE                          0.16        1.00      1.00     0.15
## TUMOR_SIZE                           1.00        1.00      1.00     1.00
## TUMOR_STAGE                          1.00        1.00      1.00     1.00
## HER2_SNP6                            1.00        1.00      1.00     1.00
## HISTOLOGICAL_SUBTYPE                 0.00        1.00      1.00     0.78
## CELLULARITY                          0.84        0.00      1.00     1.00
## MENOPAUSE                            0.20        0.55      0.00     1.00
## INTCLUST                             0.00        0.03      0.11     0.00
## LATERALITY                           0.43        0.40      0.50     0.27
## MUTATION_COUNT                       0.91        0.09      0.69     0.17
## NPI                                  0.69        0.28      0.17     0.75
## SURVIVAL_MONTHS                      0.31        0.82      0.67     0.73
## SURVIVAL_GROUPS                        NA          NA        NA       NA
## THREEGENE                            0.19        0.95      0.18     0.03
## VITAL_STATUS                         0.62        0.42      0.61     0.56
##                       LATERALITY MUTATION_COUNT  NPI SURVIVAL_MONTHS
## AGE_AT_DIAGNOSIS            1.00           1.00 1.00            1.00
## AGE_GROUP                   1.00           1.00 1.00            1.00
## BREAST_SURGERY              1.00           1.00 1.00            1.00
## CANCER_TYPE                 1.00           0.10 1.00            1.00
## CHEMOTHERAPY                1.00           1.00 0.01            1.00
## RADIO_THERAPY                 NA             NA   NA              NA
## HORMONE_THERAPY             1.00           1.00 1.00            1.00
## PAM50                       1.00           1.00 1.00            1.00
## COHORT                      1.00           1.00 1.00            1.00
## ER_IHC                      1.00           1.00 1.00            1.00
## TUMOR_GRADE                 1.00           1.00 0.00            1.00
## TUMOR_SIZE                  1.00           1.00 1.00            1.00
## TUMOR_STAGE                 1.00           1.00 0.01            1.00
## HER2_SNP6                   1.00           1.00 1.00            1.00
## HISTOLOGICAL_SUBTYPE        1.00           1.00 1.00            1.00
## CELLULARITY                 1.00           1.00 1.00            1.00
## MENOPAUSE                   1.00           1.00 1.00            1.00
## INTCLUST                    1.00           1.00 1.00            1.00
## LATERALITY                  0.00           1.00 1.00            1.00
## MUTATION_COUNT              0.73           0.00 1.00            1.00
## NPI                         0.88           0.38 0.00            1.00
## SURVIVAL_MONTHS             0.84           0.57 0.06            0.00
## SURVIVAL_GROUPS               NA             NA   NA              NA
## THREEGENE                   0.75           0.03 0.59            0.60
## VITAL_STATUS                0.77           0.08 0.05            0.04
##                       SURVIVAL_GROUPS THREEGENE VITAL_STATUS
## AGE_AT_DIAGNOSIS                   NA      1.00            1
## AGE_GROUP                          NA      1.00            1
## BREAST_SURGERY                     NA      1.00            1
## CANCER_TYPE                        NA      1.00            1
## CHEMOTHERAPY                       NA      1.00            1
## RADIO_THERAPY                      NA        NA           NA
## HORMONE_THERAPY                    NA      1.00            1
## PAM50                              NA      0.10            1
## COHORT                             NA      1.00            1
## ER_IHC                             NA      0.00            1
## TUMOR_GRADE                        NA      1.00            1
## TUMOR_SIZE                         NA      1.00            1
## TUMOR_STAGE                        NA      1.00            1
## HER2_SNP6                          NA      1.00            1
## HISTOLOGICAL_SUBTYPE               NA      1.00            1
## CELLULARITY                        NA      1.00            1
## MENOPAUSE                          NA      1.00            1
## INTCLUST                           NA      1.00            1
## LATERALITY                         NA      1.00            1
## MUTATION_COUNT                     NA      1.00            1
## NPI                                NA      1.00            1
## SURVIVAL_MONTHS                    NA      1.00            1
## SURVIVAL_GROUPS                    NA        NA           NA
## THREEGENE                          NA      0.00            1
## VITAL_STATUS                       NA      0.39            0
##
##  To see confidence intervals of the correlations, print with the short=FALSE option

corr.test((clin[(SURVIVAL_MONTHS>60),])) # for clinical variables of 120 patients with > 5years follow-up

## Warning in cor(x, use = use, method = method): the standard deviation is zero
```

```
## Call:corr.test(x = (clin[(SURVIVAL_MONTHS > 60), ]))
## Correlation matrix
##                      AGE_AT_DIAGNOSIS AGE_GROUP BREAST_SURGERY CANCER_TYPE
## AGE_AT_DIAGNOSIS                 1.00      0.90           0.03        0.07
## AGE_GROUP                        0.90      1.00           0.00        0.05
## BREAST_SURGERY                   0.03      0.00           1.00        0.10
## CANCER_TYPE                      0.07      0.05           0.10        1.00
## CHEMOTHERAPY                     0.40      0.29           0.29       -0.14
## RADIO_THERAPY                      NA        NA             NA          NA
## HORMONE_THERAPY                 -0.27     -0.26           0.05       -0.24
## PAM50                            0.17      0.19           0.00        0.10
## COHORT                          -0.17     -0.11           0.22       -0.02
## ER_IHC                          -0.24     -0.22          -0.04       -0.01
## TUMOR_GRADE                     -0.15     -0.14          -0.16       -0.25
## TUMOR_SIZE                      -0.19     -0.14          -0.01        0.00
## TUMOR_STAGE                     -0.13     -0.10          -0.20       -0.02
## HER2_SNP6                        0.05      0.02           0.07        0.04
## HISTOLOGICAL_SUBTYPE             0.05      0.11           0.05        0.26
## CELLULARITY                     -0.17     -0.13           0.08       -0.08
## MENOPAUSE                        0.84      0.85          -0.06        0.15
## INTCLUST                        -0.21     -0.20          -0.10       -0.13
## LATERALITY                       0.22      0.18           0.06        0.06
## MUTATION_COUNT                  -0.16     -0.13           0.12        0.17
## NPI                             -0.27     -0.18          -0.29       -0.12
## SURVIVAL_MONTHS                 -0.18     -0.17          -0.23       -0.14
## SURVIVAL_GROUPS                    NA        NA             NA          NA
## THREEGENE                        0.11      0.15           0.04        0.19
## VITAL_STATUS                    -0.01     -0.05          -0.11       -0.08
##                      CHEMOTHERAPY RADIO_THERAPY HORMONE_THERAPY PAM50 COHORT
## AGE_AT_DIAGNOSIS             0.40            NA           -0.27  0.17  -0.17
## AGE_GROUP                    0.29            NA           -0.26  0.19  -0.11
## BREAST_SURGERY               0.29            NA            0.05  0.00   0.22
## CANCER_TYPE                 -0.14            NA           -0.24  0.10  -0.02
## CHEMOTHERAPY                 1.00            NA           -0.08  0.08   0.05
## RADIO_THERAPY                  NA            NA              NA    NA     NA
## HORMONE_THERAPY             -0.08            NA            1.00 -0.29   0.10
## PAM50                        0.08            NA           -0.29  1.00  -0.04
## COHORT                       0.05            NA            0.10 -0.04   1.00
## ER_IHC                      -0.23            NA            0.36 -0.26   0.35
## TUMOR_GRADE                 -0.24            NA            0.31 -0.30   0.15
## TUMOR_SIZE                  -0.26            NA            0.02 -0.13   0.21
## TUMOR_STAGE                 -0.50            NA           -0.10  0.02   0.01
## HER2_SNP6                    0.13            NA           -0.13 -0.16  -0.13
## HISTOLOGICAL_SUBTYPE         0.07            NA           -0.19  0.15   0.07
## CELLULARITY                 -0.14            NA            0.18 -0.31   0.16
## MENOPAUSE                    0.22            NA           -0.32  0.20  -0.15
## INTCLUST                    -0.13            NA            0.19 -0.41   0.09
## LATERALITY                   0.16            NA           -0.03  0.10   0.05
## MUTATION_COUNT              -0.10            NA            0.12 -0.15   0.41
## NPI                         -0.60            NA            0.07 -0.16   0.16
## SURVIVAL_MONTHS              0.01            NA            0.08 -0.06   0.10
## SURVIVAL_GROUPS                NA            NA              NA    NA     NA
## THREEGENE                    0.00            NA           -0.15  0.51  -0.07
## VITAL_STATUS                 0.04            NA           -0.10  0.23   0.01
##                      ER_IHC TUMOR_GRADE TUMOR_SIZE TUMOR_STAGE HER2_SNP6
## AGE_AT_DIAGNOSIS      -0.24       -0.15      -0.19       -0.13      0.05
## AGE_GROUP             -0.22       -0.14      -0.14       -0.10      0.02
## BREAST_SURGERY        -0.04       -0.16      -0.01       -0.20      0.07
## CANCER_TYPE           -0.01       -0.25       0.00       -0.02      0.04
## CHEMOTHERAPY          -0.23       -0.24      -0.26       -0.50      0.13
## RADIO_THERAPY            NA          NA         NA          NA        NA
## HORMONE_THERAPY        0.36        0.31       0.02       -0.10     -0.13
## PAM50                 -0.26       -0.30      -0.13        0.02     -0.16
## COHORT                 0.35        0.15       0.21        0.01     -0.13
## ER_IHC                 1.00        0.48       0.17        0.07      0.04
## TUMOR_GRADE            0.48        1.00       0.17        0.12     -0.07
## TUMOR_SIZE             0.17        0.17       1.00        0.32     -0.01
## TUMOR_STAGE            0.07        0.12       0.32        1.00     -0.06
## HER2_SNP6              0.04       -0.07      -0.01       -0.06      1.00
## HISTOLOGICAL_SUBTYPE   0.00       -0.27      -0.01       -0.05      0.15
## CELLULARITY            0.29        0.16       0.14       -0.06      0.07
## MENOPAUSE             -0.16       -0.14      -0.07       -0.06      0.00
## INTCLUST               0.28        0.25       0.14        0.02     -0.03
## LATERALITY             0.00       -0.04      -0.19       -0.12     -0.19
## MUTATION_COUNT         0.26        0.04       0.09        0.03     -0.04
## NPI                    0.29        0.62       0.28        0.62     -0.10
## SURVIVAL_MONTHS        0.15        0.02       0.05        0.18     -0.02
## SURVIVAL_GROUPS          NA          NA         NA          NA        NA
## THREEGENE             -0.25       -0.22      -0.11       -0.07     -0.25
```

```
## VITAL_STATUS          -0.12       -0.11      -0.08       -0.03      0.02
##                     HISTOLOGICAL_SUBTYPE CELLULARITY MENOPAUSE INTCLUST
## AGE_AT_DIAGNOSIS                    0.05       -0.17      0.84    -0.21
## AGE_GROUP                           0.11       -0.13      0.85    -0.20
## BREAST_SURGERY                      0.05        0.08     -0.06    -0.10
## CANCER_TYPE                         0.26       -0.08      0.15    -0.13
## CHEMOTHERAPY                        0.07       -0.14      0.22    -0.13
## RADIO_THERAPY                         NA          NA        NA       NA
## HORMONE_THERAPY                    -0.19        0.18     -0.32     0.19
## PAM50                               0.15       -0.31      0.20    -0.41
## COHORT                              0.07        0.16     -0.15     0.09
## ER_IHC                              0.00        0.29     -0.16     0.28
## TUMOR_GRADE                        -0.27        0.16     -0.14     0.25
## TUMOR_SIZE                         -0.01        0.14     -0.07     0.14
## TUMOR_STAGE                        -0.05       -0.06     -0.06     0.02
## HER2_SNP6                           0.15        0.07      0.00    -0.03
## HISTOLOGICAL_SUBTYPE                1.00       -0.02      0.14    -0.14
## CELLULARITY                        -0.02        1.00     -0.08     0.26
## MENOPAUSE                           0.14       -0.08      1.00    -0.15
## INTCLUST                           -0.14        0.26     -0.15     1.00
## LATERALITY                          0.06       -0.14      0.09    -0.06
## MUTATION_COUNT                      0.10        0.09     -0.08     0.09
## NPI                                -0.19        0.07     -0.16     0.18
## SURVIVAL_MONTHS                     0.17        0.08     -0.11     0.13
## SURVIVAL_GROUPS                       NA          NA        NA       NA
## THREEGENE                           0.10       -0.30      0.11    -0.31
## VITAL_STATUS                       -0.02       -0.11     -0.02    -0.21
##                     LATERALITY MUTATION_COUNT   NPI SURVIVAL_MONTHS
## AGE_AT_DIAGNOSIS          0.22          -0.16 -0.27           -0.18
## AGE_GROUP                 0.18          -0.13 -0.18           -0.17
## BREAST_SURGERY            0.06           0.12 -0.29           -0.23
## CANCER_TYPE               0.06           0.17 -0.12           -0.14
## CHEMOTHERAPY              0.16          -0.10 -0.60            0.01
## RADIO_THERAPY               NA             NA    NA              NA
## HORMONE_THERAPY          -0.03           0.12  0.07            0.08
## PAM50                     0.10          -0.15 -0.16           -0.06
## COHORT                    0.05           0.41  0.16            0.10
## ER_IHC                    0.00           0.26  0.29            0.15
## TUMOR_GRADE              -0.04           0.04  0.62            0.02
## TUMOR_SIZE               -0.19           0.09  0.28            0.05
## TUMOR_STAGE              -0.12           0.03  0.62            0.18
## HER2_SNP6                -0.19          -0.04 -0.10           -0.02
## HISTOLOGICAL_SUBTYPE      0.06           0.10 -0.19            0.17
## CELLULARITY              -0.14           0.09  0.07            0.08
## MENOPAUSE                 0.09          -0.08 -0.16           -0.11
## INTCLUST                 -0.06           0.09  0.18            0.13
## LATERALITY                1.00          -0.01 -0.08            0.03
## MUTATION_COUNT           -0.01           1.00  0.12            0.00
## NPI                      -0.08           0.12  1.00            0.13
## SURVIVAL_MONTHS           0.03           0.00  0.13            1.00
## SURVIVAL_GROUPS             NA             NA    NA              NA
## THREEGENE                 0.03           0.03 -0.18           -0.23
## VITAL_STATUS              0.04           0.12  0.00           -0.27
##                     SURVIVAL_GROUPS THREEGENE VITAL_STATUS
## AGE_AT_DIAGNOSIS                 NA      0.11        -0.01
## AGE_GROUP                        NA      0.15        -0.05
## BREAST_SURGERY                   NA      0.04        -0.11
## CANCER_TYPE                      NA      0.19        -0.08
## CHEMOTHERAPY                     NA      0.00         0.04
## RADIO_THERAPY                    NA        NA           NA
## HORMONE_THERAPY                  NA     -0.15        -0.10
## PAM50                            NA      0.51         0.23
## COHORT                           NA     -0.07         0.01
## ER_IHC                           NA     -0.25        -0.12
## TUMOR_GRADE                      NA     -0.22        -0.11
## TUMOR_SIZE                       NA     -0.11        -0.08
## TUMOR_STAGE                      NA     -0.07        -0.03
## HER2_SNP6                        NA     -0.25         0.02
## HISTOLOGICAL_SUBTYPE             NA      0.10        -0.02
## CELLULARITY                      NA     -0.30        -0.11
## MENOPAUSE                        NA      0.11        -0.02
## INTCLUST                         NA     -0.31        -0.21
## LATERALITY                       NA      0.03         0.04
## MUTATION_COUNT                   NA      0.03         0.12
## NPI                              NA     -0.18         0.00
## SURVIVAL_MONTHS                  NA     -0.23        -0.27
## SURVIVAL_GROUPS                  NA        NA           NA
## THREEGENE                        NA      1.00        -0.05
## VITAL_STATUS                     NA     -0.05         1.00
```

```
## Sample Size
## [1] 120
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##                      AGE_AT_DIAGNOSIS AGE_GROUP BREAST_SURGERY CANCER_TYPE
## AGE_AT_DIAGNOSIS                 0.00      0.00           1.00        1.00
## AGE_GROUP                        0.00      0.00           1.00        1.00
## BREAST_SURGERY                   0.76      0.96           0.00        1.00
## CANCER_TYPE                      0.47      0.57           0.30        0.00
## CHEMOTHERAPY                     0.00      0.00           0.00        0.12
## RADIO_THERAPY                      NA        NA             NA          NA
## HORMONE_THERAPY                  0.00      0.00           0.62        0.01
## PAM50                            0.06      0.04           0.97        0.26
## COHORT                           0.06      0.22           0.02        0.80
## ER_IHC                           0.01      0.01           0.65        0.89
## TUMOR_GRADE                      0.10      0.13           0.08        0.00
## TUMOR_SIZE                       0.04      0.13           0.87        1.00
## TUMOR_STAGE                      0.16      0.30           0.03        0.80
## HER2_SNP6                        0.58      0.85           0.46        0.65
## HISTOLOGICAL_SUBTYPE             0.55      0.22           0.56        0.00
## CELLULARITY                      0.07      0.15           0.40        0.41
## MENOPAUSE                        0.00      0.00           0.52        0.09
## INTCLUST                         0.02      0.03           0.29        0.15
## LATERALITY                       0.02      0.05           0.54        0.50
## MUTATION_COUNT                   0.07      0.16           0.18        0.07
## NPI                              0.00      0.04           0.00        0.19
## SURVIVAL_MONTHS                  0.05      0.06           0.01        0.13
## SURVIVAL_GROUPS                    NA        NA             NA          NA
## THREEGENE                        0.24      0.10           0.66        0.04
## VITAL_STATUS                     0.88      0.62           0.25        0.37
##                      CHEMOTHERAPY RADIO_THERAPY HORMONE_THERAPY PAM50 COHORT
## AGE_AT_DIAGNOSIS             0.00            NA            0.62  1.00   1.00
## AGE_GROUP                    0.32            NA            0.85  1.00   1.00
## BREAST_SURGERY               0.26            NA            1.00  1.00   1.00
## CANCER_TYPE                  1.00            NA            1.00  1.00   1.00
## CHEMOTHERAPY                 0.00            NA            1.00  1.00   1.00
## RADIO_THERAPY                  NA            NA              NA    NA     NA
## HORMONE_THERAPY              0.40            NA            0.00  0.35   1.00
## PAM50                        0.41            NA            0.00  0.00   1.00
## COHORT                       0.61            NA            0.26  0.70   0.00
## ER_IHC                       0.01            NA            0.00  0.00   0.00
## TUMOR_GRADE                  0.01            NA            0.00  0.00   0.09
## TUMOR_SIZE                   0.00            NA            0.80  0.17   0.02
## TUMOR_STAGE                  0.00            NA            0.26  0.85   0.91
## HER2_SNP6                    0.16            NA            0.15  0.08   0.15
## HISTOLOGICAL_SUBTYPE         0.46            NA            0.04  0.09   0.43
## CELLULARITY                  0.14            NA            0.05  0.00   0.07
## MENOPAUSE                    0.02            NA            0.00  0.03   0.11
## INTCLUST                     0.14            NA            0.04  0.00   0.32
## LATERALITY                   0.08            NA            0.74  0.28   0.58
## MUTATION_COUNT               0.29            NA            0.18  0.11   0.00
## NPI                          0.00            NA            0.46  0.09   0.08
## SURVIVAL_MONTHS              0.93            NA            0.37  0.49   0.28
## SURVIVAL_GROUPS                NA            NA              NA    NA     NA
## THREEGENE                    0.96            NA            0.10  0.00   0.48
## VITAL_STATUS                 0.65            NA            0.28  0.01   0.91
##                      ER_IHC TUMOR_GRADE TUMOR_SIZE TUMOR_STAGE HER2_SNP6
## AGE_AT_DIAGNOSIS       1.00        1.00       1.00        1.00      1.00
## AGE_GROUP              1.00        1.00       1.00        1.00      1.00
## BREAST_SURGERY         1.00        1.00       1.00        1.00      1.00
## CANCER_TYPE            1.00        1.00       1.00        1.00      1.00
## CHEMOTHERAPY           1.00        1.00       1.00        0.00      1.00
## RADIO_THERAPY            NA          NA         NA          NA        NA
## HORMONE_THERAPY        0.01        0.16       1.00        1.00      1.00
## PAM50                  1.00        0.20       1.00        1.00      1.00
## COHORT                 0.02        1.00       1.00        1.00      1.00
## ER_IHC                 0.00        0.00       1.00        1.00      1.00
## TUMOR_GRADE            0.00        0.00       1.00        1.00      1.00
## TUMOR_SIZE             0.06        0.06       0.00        0.08      1.00
## TUMOR_STAGE            0.46        0.18       0.00        0.00      1.00
## HER2_SNP6              0.64        0.45       0.88        0.54      0.00
## HISTOLOGICAL_SUBTYPE   0.98        0.00       0.94        0.59      0.11
## CELLULARITY            0.00        0.08       0.14        0.52      0.42
## MENOPAUSE              0.08        0.13       0.44        0.50      0.96
## INTCLUST               0.00        0.01       0.13        0.80      0.76
## LATERALITY             0.99        0.68       0.04        0.19      0.04
## MUTATION_COUNT         0.00        0.64       0.35        0.77      0.65
## NPI                    0.00        0.00       0.00        0.00      0.26
## SURVIVAL_MONTHS        0.09        0.85       0.59        0.05      0.79
## SURVIVAL_GROUPS          NA          NA         NA          NA        NA
```

```
## THREEGENE              0.01        0.02       0.22        0.47      0.01
## VITAL_STATUS           0.18        0.25       0.39        0.76      0.83
##                      HISTOLOGICAL_SUBTYPE CELLULARITY MENOPAUSE INTCLUST
## AGE_AT_DIAGNOSIS                     1.00        1.00      0.00     1.00
## AGE_GROUP                            1.00        1.00      0.00     1.00
## BREAST_SURGERY                       1.00        1.00      1.00     1.00
## CANCER_TYPE                          0.78        1.00      1.00     1.00
## CHEMOTHERAPY                         1.00        1.00      1.00     1.00
## RADIO_THERAPY                          NA          NA        NA       NA
## HORMONE_THERAPY                      1.00        1.00      0.09     1.00
## PAM50                                1.00        0.12      1.00     0.00
## COHORT                               1.00        1.00      1.00     1.00
## ER_IHC                               1.00        0.26      1.00     0.50
## TUMOR_GRADE                          0.57        1.00      1.00     1.00
## TUMOR_SIZE                           1.00        1.00      1.00     1.00
## TUMOR_STAGE                          1.00        1.00      1.00     1.00
## HER2_SNP6                            1.00        1.00      1.00     1.00
## HISTOLOGICAL_SUBTYPE                 0.00        1.00      1.00     1.00
## CELLULARITY                          0.80        0.00      1.00     1.00
## MENOPAUSE                            0.14        0.36      0.00     1.00
## INTCLUST                             0.13        0.00      0.09     0.00
## LATERALITY                           0.49        0.12      0.33     0.52
## MUTATION_COUNT                       0.30        0.32      0.41     0.35
## NPI                                  0.04        0.42      0.08     0.05
## SURVIVAL_MONTHS                      0.06        0.40      0.25     0.15
## SURVIVAL_GROUPS                        NA          NA        NA       NA
## THREEGENE                            0.27        0.00      0.24     0.00
## VITAL_STATUS                         0.79        0.22      0.82     0.02
##                      LATERALITY MUTATION_COUNT  NPI SURVIVAL_MONTHS
## AGE_AT_DIAGNOSIS           1.00           1.00 0.71            1.00
## AGE_GROUP                  1.00           1.00 1.00            1.00
## BREAST_SURGERY             1.00           1.00 0.28            1.00
## CANCER_TYPE                1.00           1.00 1.00            1.00
## CHEMOTHERAPY               1.00           1.00 0.00            1.00
## RADIO_THERAPY                NA             NA   NA              NA
## HORMONE_THERAPY            1.00           1.00 1.00            1.00
## PAM50                      1.00           1.00 1.00            1.00
## COHORT                     1.00           0.00 1.00            1.00
## ER_IHC                     1.00           0.99 0.32            1.00
## TUMOR_GRADE                1.00           1.00 0.00            1.00
## TUMOR_SIZE                 1.00           1.00 0.52            1.00
## TUMOR_STAGE                1.00           1.00 0.00            1.00
## HER2_SNP6                  1.00           1.00 1.00            1.00
## HISTOLOGICAL_SUBTYPE       1.00           1.00 1.00            1.00
## CELLULARITY                1.00           1.00 1.00            1.00
## MENOPAUSE                  1.00           1.00 1.00            1.00
## INTCLUST                   1.00           1.00 1.00            1.00
## LATERALITY                 0.00           1.00 1.00            1.00
## MUTATION_COUNT             0.93           0.00 1.00            1.00
## NPI                        0.36           0.19 0.00            1.00
## SURVIVAL_MONTHS            0.77           0.96 0.15            0.00
## SURVIVAL_GROUPS              NA             NA   NA              NA
## THREEGENE                  0.74           0.77 0.05            0.01
## VITAL_STATUS               0.63           0.20 0.97            0.00
##                      SURVIVAL_GROUPS THREEGENE VITAL_STATUS
## AGE_AT_DIAGNOSIS                  NA      1.00         1.00
## AGE_GROUP                         NA      1.00         1.00
## BREAST_SURGERY                    NA      1.00         1.00
## CANCER_TYPE                       NA      1.00         1.00
## CHEMOTHERAPY                      NA      1.00         1.00
## RADIO_THERAPY                     NA        NA           NA
## HORMONE_THERAPY                   NA      1.00         1.00
## PAM50                             NA      0.00         1.00
## COHORT                            NA      1.00         1.00
## ER_IHC                            NA      1.00         1.00
## TUMOR_GRADE                       NA      1.00         1.00
## TUMOR_SIZE                        NA      1.00         1.00
## TUMOR_STAGE                       NA      1.00         1.00
## HER2_SNP6                         NA      1.00         1.00
## HISTOLOGICAL_SUBTYPE              NA      1.00         1.00
## CELLULARITY                       NA      0.24         1.00
## MENOPAUSE                         NA      1.00         1.00
## INTCLUST                          NA      0.12         1.00
## LATERALITY                        NA      1.00         1.00
## MUTATION_COUNT                    NA      1.00         1.00
## NPI                               NA      1.00         1.00
## SURVIVAL_MONTHS                   NA      1.00         0.66
## SURVIVAL_GROUPS                   NA        NA           NA
## THREEGENE                         NA      0.00         1.00
```

```
## VITAL_STATUS                      NA      0.55         0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option

t_mutate = cbind(TNBC_mut[-1],clin[c(22,25)]) #bind mutation data with SURVIVAL_MONTHS + VITAL_STATUS variables
View(t_mutate)

corr.test(t_mutate[-19]) # for all mutation variables + VITAL_STATUS

## Call:corr.test(x = t_mutate[-19])
## Correlation matrix
##               AKT1 MAP3K1  MEN1 NCOR2 PIK3CA SHANK2 MAP2K4  AGMO  TAF1 GATA3
## AKT1          1.00  -0.05 -0.03 -0.05  -0.01  -0.05  -0.03  0.30 -0.04  0.10
## MAP3K1       -0.05   1.00  0.13 -0.06   0.13  -0.06  -0.04  0.05 -0.04 -0.06
## MEN1         -0.03   0.13  1.00 -0.03  -0.09  -0.04  -0.02 -0.03 -0.03  0.11
## NCOR2        -0.05  -0.06 -0.03  1.00  -0.03  -0.06  -0.04  0.05  0.09  0.00
## PIK3CA       -0.01   0.13 -0.09 -0.03   1.00   0.11   0.05  0.07  0.10  0.07
## SHANK2       -0.05  -0.06 -0.04 -0.06   0.11   1.00  -0.04 -0.06 -0.04 -0.01
## MAP2K4       -0.03  -0.04 -0.02 -0.04   0.05  -0.04   1.00  0.11  0.35  0.09
## AGMO          0.30   0.05 -0.03  0.05   0.07  -0.06   0.11  1.00  0.09 -0.06
## TAF1         -0.04  -0.04 -0.03  0.09   0.10  -0.04   0.35  0.09  1.00 -0.01
## GATA3         0.10  -0.06  0.11  0.00   0.07  -0.01   0.09 -0.06 -0.01  1.00
## BIRC6         0.03  -0.07 -0.05 -0.07   0.11   0.16   0.06  0.17  0.15  0.08
## DNAH11        0.15   0.11 -0.06  0.04   0.01  -0.10   0.03  0.18  0.02  0.07
## NF1           0.08   0.06  0.14 -0.05   0.04  -0.06  -0.04 -0.05 -0.04 -0.05
## TP53         -0.01  -0.09  0.08  0.02   0.02   0.08  -0.07 -0.04 -0.04  0.11
## SYNE1        -0.03  -0.05  0.03  0.14   0.01   0.01   0.01  0.08  0.08  0.02
## CTNNA1       -0.02  -0.02 -0.01 -0.02   0.12  -0.02  -0.01 -0.02  0.40 -0.04
## FOXO3        -0.05  -0.06 -0.03  0.05   0.07  -0.06   0.11  0.05  0.23  0.00
## LDLRAP1      -0.02  -0.02 -0.01 -0.02   0.12  -0.02  -0.01 -0.02 -0.01  0.15
## VITAL_STATUS  0.19   0.22  0.17 -0.16   0.05  -0.22   0.05  0.13 -0.04 -0.06
##              BIRC6 DNAH11   NF1  TP53 SYNE1 CTNNA1 FOXO3 LDLRAP1 VITAL_STATUS
## AKT1          0.03   0.15  0.08 -0.01 -0.03  -0.02 -0.05   -0.02         0.19
## MAP3K1       -0.07   0.11  0.06 -0.09 -0.05  -0.02 -0.06   -0.02         0.22
## MEN1         -0.05  -0.06  0.14  0.08  0.03  -0.01 -0.03   -0.01         0.17
## NCOR2        -0.07   0.04 -0.05  0.02  0.14  -0.02  0.05   -0.02        -0.16
## PIK3CA        0.11   0.01  0.04  0.02  0.01   0.12  0.07    0.12         0.05
## SHANK2        0.16  -0.10 -0.06  0.08  0.01  -0.02 -0.06   -0.02        -0.22
## MAP2K4        0.06   0.03 -0.04 -0.07  0.01  -0.01  0.11   -0.01         0.05
## AGMO          0.17   0.18 -0.05 -0.04  0.08  -0.02  0.05   -0.02         0.13
## TAF1          0.15   0.02 -0.04 -0.04  0.08   0.40  0.23   -0.01        -0.04
## GATA3         0.08   0.07 -0.05  0.11  0.02  -0.04  0.00    0.15        -0.06
## BIRC6         1.00  -0.02 -0.07  0.08  0.10   0.23 -0.07   -0.02         0.06
## DNAH11       -0.02   1.00 -0.02 -0.04  0.14  -0.03 -0.03   -0.03        -0.07
## NF1          -0.07  -0.02  1.00 -0.05 -0.04  -0.02  0.06   -0.02         0.05
## TP53          0.08  -0.04 -0.05  1.00  0.05   0.04  0.07    0.04        -0.04
## SYNE1         0.10   0.14 -0.04  0.05  1.00  -0.03 -0.05   -0.03        -0.01
## CTNNA1        0.23  -0.03 -0.02  0.04 -0.03   1.00 -0.02   -0.01         0.08
## FOXO3        -0.07  -0.03  0.06  0.07 -0.05  -0.02  1.00   -0.02        -0.07
## LDLRAP1      -0.02  -0.03 -0.02  0.04 -0.03  -0.01 -0.02    1.00        -0.06
## VITAL_STATUS  0.06  -0.07  0.05 -0.04 -0.01   0.08 -0.07   -0.06         1.00
## Sample Size
## [1] 190
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##              AKT1 MAP3K1 MEN1 NCOR2 PIK3CA SHANK2 MAP2K4 AGMO TAF1 GATA3 BIRC6
## AKT1         0.00   1.00 1.00  1.00   1.00   1.00   1.00 0.00 1.00  1.00  1.00
## MAP3K1       0.50   0.00 1.00  1.00   1.00   1.00   1.00 1.00 1.00  1.00  1.00
## MEN1         0.67   0.07 0.00  1.00   1.00   1.00   1.00 1.00 1.00  1.00  1.00
## NCOR2        0.50   0.45 0.64  0.00   1.00   1.00   1.00 1.00 1.00  1.00  1.00
## PIK3CA       0.93   0.08 0.23  0.64   0.00   1.00   1.00 1.00 1.00  1.00  1.00
## SHANK2       0.48   0.42 0.62  0.42   0.14   0.00   1.00 1.00 1.00  1.00  1.00
## MAP2K4       0.64   0.60 0.74  0.60   0.48   0.58   0.00 1.00 0.00  1.00  1.00
## AGMO         0.00   0.49 0.64  0.49   0.32   0.42   0.14 0.00 1.00  1.00  1.00
## TAF1         0.60   0.56 0.72  0.21   0.18   0.54   0.00 0.21 0.00  1.00  1.00
## GATA3        0.19   0.44 0.12  0.97   0.35   0.91   0.24 0.44 0.86  0.00  1.00
## BIRC6        0.72   0.31 0.53  0.31   0.15   0.03   0.38 0.02 0.03  0.28  0.00
## DNAH11       0.05   0.12 0.42  0.55   0.94   0.18   0.68 0.01 0.83  0.30  0.81
## NF1          0.29   0.42 0.05  0.47   0.63   0.45   0.62 0.47 0.58  0.52  0.34
## TP53         0.90   0.20 0.27  0.81   0.82   0.26   0.37 0.60 0.55  0.12  0.25
## SYNE1        0.71   0.53 0.69  0.05   0.89   0.94   0.88 0.28 0.30  0.78  0.17
## CTNNA1       0.83   0.81 0.88  0.81   0.09   0.80   0.87 0.81 0.00  0.62  0.00
## FOXO3        0.50   0.45 0.64  0.49   0.32   0.42   0.14 0.49 0.00  0.97  0.31
## LDLRAP1      0.83   0.81 0.88  0.81   0.09   0.80   0.87 0.81 0.86  0.04  0.75
## VITAL_STATUS 0.01   0.00 0.02  0.03   0.48   0.00   0.46 0.09 0.61  0.43  0.42
##              DNAH11  NF1 TP53 SYNE1 CTNNA1 FOXO3 LDLRAP1 VITAL_STATUS
## AKT1           1.00 1.00 1.00  1.00   1.00  1.00    1.00         1.00
## MAP3K1         1.00 1.00 1.00  1.00   1.00  1.00    1.00         0.38
## MEN1           1.00 1.00 1.00  1.00   1.00  1.00    1.00         1.00
```

```
## NCOR2          1.00 1.00 1.00  1.00   1.00  1.00    1.00         1.00
## PIK3CA         1.00 1.00 1.00  1.00   1.00  1.00    1.00         1.00
## SHANK2         1.00 1.00 1.00  1.00   1.00  1.00    1.00         0.41
## MAP2K4         1.00 1.00 1.00  1.00   1.00  1.00    1.00         1.00
## AGMO           1.00 1.00 1.00  1.00   1.00  1.00    1.00         1.00
## TAF1           1.00 1.00 1.00  1.00   0.00  0.27    1.00         1.00
## GATA3          1.00 1.00 1.00  1.00   1.00  1.00    1.00         1.00
## BIRC6          1.00 1.00 1.00  1.00   0.21  1.00    1.00         1.00
## DNAH11         0.00 1.00 1.00  1.00   1.00  1.00    1.00         1.00
## NF1            0.82 0.00 1.00  1.00   1.00  1.00    1.00         1.00
## TP53           0.63 0.46 0.00  1.00   1.00  1.00    1.00         1.00
## SYNE1          0.05 0.61 0.46  0.00   1.00  1.00    1.00         1.00
## CTNNA1         0.69 0.82 0.58  0.65   0.00  1.00    1.00         1.00
## FOXO3          0.73 0.42 0.31  0.53   0.81  0.00    1.00         1.00
## LDLRAP1        0.69 0.82 0.58  0.65   0.94  0.81    0.00         1.00
## VITAL_STATUS   0.32 0.46 0.54  0.87   0.26  0.37    0.38         0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

```
corr.test((t_mutate[(SURVIVAL_MONTHS<60.1),]))# for mutation variables of 70 patients with <= 5years follow-up
```

```
## Warning in cor(x, use = use, method = method): the standard deviation is zero
```

```
## Call:corr.test(x = (t_mutate[(SURVIVAL_MONTHS < 60.1), ]))
## Correlation matrix
##                  AKT1 MAP3K1  MEN1 NCOR2 PIK3CA SHANK2 MAP2K4  AGMO  TAF1 GATA3
## AKT1             1.00  -0.08 -0.06 -0.03  -0.07  -0.03  -0.05  0.31 -0.05  0.17
## MAP3K1          -0.08   1.00  0.22 -0.03  -0.07  -0.03  -0.05  0.11 -0.05  0.02
## MEN1            -0.06   0.22  1.00 -0.03  -0.14  -0.03  -0.04 -0.06 -0.04  0.28
## NCOR2           -0.03  -0.03 -0.03  1.00  -0.08  -0.01  -0.02 -0.04 -0.02  0.26
## PIK3CA          -0.07  -0.07 -0.14 -0.08   1.00   0.18  -0.12  0.12  0.07  0.02
## SHANK2          -0.03  -0.03 -0.03 -0.01   0.18   1.00  -0.02 -0.04 -0.02 -0.05
## MAP2K4          -0.05  -0.05 -0.04 -0.02  -0.12  -0.02   1.00 -0.05 -0.03  0.15
## AGMO             0.31   0.11 -0.06 -0.04   0.12  -0.04  -0.05  1.00 -0.05  0.00
## TAF1            -0.05  -0.05 -0.04 -0.02   0.07  -0.02  -0.03 -0.05  1.00 -0.08
## GATA3            0.17   0.02  0.28  0.26   0.02  -0.05   0.15  0.00 -0.08  1.00
## BIRC6            0.07  -0.10 -0.08 -0.04   0.14   0.34  -0.06  0.21  0.21 -0.04
## DNAH11           0.41  -0.07 -0.05 -0.03  -0.03  -0.03  -0.04  0.36 -0.04  0.21
## NF1             -0.06  -0.06  0.30 -0.03   0.16  -0.03  -0.04 -0.06 -0.04 -0.10
## TP53            -0.05  -0.05  0.14  0.08   0.19   0.08   0.12 -0.01  0.12  0.14
## SYNE1           -0.13   0.02  0.09 -0.05   0.02  -0.05  -0.08  0.00 -0.08  0.09
## CTNNA1          -0.03  -0.03 -0.03 -0.01   0.18  -0.01  -0.02 -0.04  0.70 -0.05
## FOXO3           -0.06  -0.06 -0.04 -0.03   0.01  -0.03  -0.04 -0.06  0.39 -0.10
## LDLRAP1            NA     NA    NA    NA     NA     NA     NA    NA    NA    NA
## SURVIVAL_MONTHS  0.28  -0.10 -0.11  0.01   0.02   0.27   0.08  0.22 -0.12 -0.12
## VITAL_STATUS     0.08   0.08  0.06  0.04  -0.01  -0.39   0.05  0.09  0.05  0.14
##                 BIRC6 DNAH11   NF1  TP53 SYNE1 CTNNA1 FOXO3 LDLRAP1
## AKT1             0.07   0.41 -0.06 -0.05 -0.13  -0.03 -0.06      NA
## MAP3K1          -0.10  -0.07 -0.06 -0.05  0.02  -0.03 -0.06      NA
## MEN1            -0.08  -0.05  0.30  0.14  0.09  -0.03 -0.04      NA
## NCOR2           -0.04  -0.03 -0.03  0.08 -0.05  -0.01 -0.03      NA
## PIK3CA           0.14  -0.03  0.16  0.19  0.02   0.18  0.01      NA
## SHANK2           0.34  -0.03 -0.03  0.08 -0.05  -0.01 -0.03      NA
## MAP2K4          -0.06  -0.04 -0.04  0.12 -0.08  -0.02 -0.04      NA
## AGMO             0.21   0.36 -0.06 -0.01  0.00  -0.04 -0.06      NA
## TAF1             0.21  -0.04 -0.04  0.12 -0.08   0.70  0.39      NA
## GATA3           -0.04   0.21 -0.10  0.14  0.09  -0.05 -0.10      NA
## BIRC6            1.00   0.11 -0.08  0.24  0.19   0.34 -0.08      NA
## DNAH11           0.11   1.00 -0.05 -0.10 -0.11  -0.03 -0.05      NA
## NF1             -0.08  -0.05  1.00 -0.01 -0.10  -0.03 -0.04      NA
## TP53             0.24  -0.10 -0.01  1.00  0.23   0.08  0.14      NA
## SYNE1            0.19  -0.11 -0.10  0.23  1.00  -0.05 -0.10      NA
## CTNNA1           0.34  -0.03 -0.03  0.08 -0.05   1.00 -0.03      NA
## FOXO3           -0.08  -0.05 -0.04  0.14 -0.10  -0.03  1.00      NA
## LDLRAP1            NA     NA    NA    NA    NA     NA    NA      NA
## SURVIVAL_MONTHS  0.09   0.04 -0.14 -0.03  0.01  -0.04  0.06      NA
## VITAL_STATUS    -0.05   0.08  0.06  0.12  0.14   0.04  0.06      NA
##                 SURVIVAL_MONTHS VITAL_STATUS
## AKT1                       0.28         0.08
## MAP3K1                    -0.10         0.08
## MEN1                      -0.11         0.06
## NCOR2                      0.01         0.04
## PIK3CA                     0.02        -0.01
## SHANK2                     0.27        -0.39
## MAP2K4                     0.08         0.05
## AGMO                       0.22         0.09
## TAF1                      -0.12         0.05
## GATA3                     -0.12         0.14
```

```
## BIRC6                      0.09        -0.05
## DNAH11                     0.04         0.08
## NF1                       -0.14         0.06
## TP53                      -0.03         0.12
## SYNE1                      0.01         0.14
## CTNNA1                    -0.04         0.04
## FOXO3                      0.06         0.06
## LDLRAP1                      NA           NA
## SURVIVAL_MONTHS            1.00        -0.24
## VITAL_STATUS              -0.24         1.00
## Sample Size
## [1] 70
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##                 AKT1 MAP3K1 MEN1 NCOR2 PIK3CA SHANK2 MAP2K4 AGMO TAF1 GATA3
## AKT1            0.00   1.00 1.00  1.00   1.00   1.00   1.00 1.00 1.00  1.00
## MAP3K1          0.53   0.00 1.00  1.00   1.00   1.00   1.00 1.00 1.00  1.00
## MEN1            0.63   0.07 0.00  1.00   1.00   1.00   1.00 1.00 1.00  1.00
## NCOR2           0.78   0.78 0.83  0.00   1.00   1.00   1.00 1.00 1.00  1.00
## PIK3CA          0.57   0.57 0.24  0.50   0.00   1.00   1.00 1.00 1.00  1.00
## SHANK2          0.78   0.78 0.83  0.91   0.14   0.00   1.00 1.00 1.00  1.00
## MAP2K4          0.70   0.70 0.77  0.87   0.34   0.87   0.00 1.00 1.00  1.00
## AGMO            0.01   0.35 0.59  0.76   0.31   0.76   0.67 0.00 1.00  1.00
## TAF1            0.70   0.70 0.77  0.87   0.57   0.87   0.81 0.67 0.00  1.00
## GATA3           0.16   0.86 0.02  0.03   0.88   0.65   0.22 0.97 0.52  0.00
## BIRC6           0.54   0.41 0.53  0.72   0.24   0.00   0.61 0.08 0.08  0.72
## DNAH11          0.00   0.57 0.67  0.81   0.78   0.81   0.73 0.00 0.73  0.07
## NF1             0.63   0.63 0.01  0.83   0.18   0.83   0.77 0.59 0.77  0.43
## TP53            0.67   0.67 0.24  0.50   0.11   0.50   0.34 0.92 0.34  0.23
## SYNE1           0.30   0.86 0.45  0.65   0.88   0.65   0.52 0.97 0.52  0.43
## CTNNA1          0.78   0.78 0.83  0.91   0.14   0.91   0.87 0.76 0.00  0.65
## FOXO3           0.63   0.63 0.71  0.83   0.94   0.83   0.77 0.59 0.00  0.43
## LDLRAP1           NA     NA   NA    NA     NA     NA     NA   NA   NA    NA
## SURVIVAL_MONTHS 0.02   0.41 0.34  0.91   0.88   0.03   0.52 0.07 0.33  0.34
## VITAL_STATUS    0.48   0.48 0.59  0.76   0.92   0.00   0.67 0.44 0.67  0.25
##                 BIRC6 DNAH11  NF1 TP53 SYNE1 CTNNA1 FOXO3 LDLRAP1
## AKT1             1.00   0.07 1.00 1.00  1.00   1.00  1.00      NA
## MAP3K1           1.00   1.00 1.00 1.00  1.00   1.00  1.00      NA
## MEN1             1.00   1.00 1.00 1.00  1.00   1.00  1.00      NA
## NCOR2            1.00   1.00 1.00 1.00  1.00   1.00  1.00      NA
## PIK3CA           1.00   1.00 1.00 1.00  1.00   1.00  1.00      NA
## SHANK2           0.76   1.00 1.00 1.00  1.00   1.00  1.00      NA
## MAP2K4           1.00   1.00 1.00 1.00  1.00   1.00  1.00      NA
## AGMO             1.00   0.32 1.00 1.00  1.00   1.00  1.00      NA
## TAF1             1.00   1.00 1.00 1.00  1.00   0.00  0.16      NA
## GATA3            1.00   1.00 1.00 1.00  1.00   1.00  1.00      NA
## BIRC6            0.00   1.00 1.00 1.00  1.00   0.76  1.00      NA
## DNAH11           0.39   0.00 1.00 1.00  1.00   1.00  1.00      NA
## NF1              0.53   0.67 0.00 1.00  1.00   1.00  1.00      NA
## TP53             0.04   0.42 0.94 0.00  1.00   1.00  1.00      NA
## SYNE1            0.11   0.36 0.43 0.06  0.00   1.00  1.00      NA
## CTNNA1           0.00   0.81 0.83 0.50  0.65   0.00  1.00      NA
## FOXO3            0.53   0.67 0.71 0.24  0.43   0.83  0.00      NA
## LDLRAP1            NA     NA   NA   NA    NA     NA    NA      NA
## SURVIVAL_MONTHS  0.46   0.76 0.25 0.80  0.92   0.77  0.63      NA
## VITAL_STATUS     0.68   0.54 0.59 0.31  0.25   0.76  0.59      NA
##                 SURVIVAL_MONTHS VITAL_STATUS
## AKT1                       1.00         1.00
## MAP3K1                     1.00         1.00
## MEN1                       1.00         1.00
## NCOR2                      1.00         1.00
## PIK3CA                     1.00         1.00
## SHANK2                     1.00         0.13
## MAP2K4                     1.00         1.00
## AGMO                       1.00         1.00
## TAF1                       1.00         1.00
## GATA3                      1.00         1.00
## BIRC6                      1.00         1.00
## DNAH11                     1.00         1.00
## NF1                        1.00         1.00
## TP53                       1.00         1.00
## SYNE1                      1.00         1.00
## CTNNA1                     1.00         1.00
## FOXO3                      1.00         1.00
## LDLRAP1                      NA           NA
## SURVIVAL_MONTHS            0.00         1.00
## VITAL_STATUS               0.04         0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

```r
corr.test((t_mutate[(SURVIVAL_MONTHS>60),]))# for mutation variables of 120 patients with > 5years follow-up

## Warning in cor(x, use = use, method = method): the standard deviation is zero

## Call:corr.test(x = (t_mutate[(SURVIVAL_MONTHS > 60), ]))
## Correlation matrix
##                  AKT1 MAP3K1  MEN1 NCOR2 PIK3CA SHANK2 MAP2K4  AGMO  TAF1 GATA3
## AKT1             1.00  -0.03 -0.01 -0.05   0.04  -0.05  -0.03  0.27 -0.03  0.05
## MAP3K1          -0.03   1.00 -0.02 -0.06   0.28  -0.06  -0.03 -0.04 -0.04 -0.11
## MEN1            -0.01  -0.02  1.00 -0.03  -0.05  -0.03  -0.01 -0.02 -0.02 -0.05
## NCOR2           -0.05  -0.06 -0.03  1.00  -0.01  -0.09  -0.05  0.12  0.12 -0.07
## PIK3CA           0.04   0.28 -0.05 -0.01   1.00   0.12   0.16  0.01  0.12  0.11
## SHANK2          -0.05  -0.06 -0.03 -0.09   0.12   1.00  -0.05 -0.06 -0.06 -0.01
## MAP2K4          -0.03  -0.03 -0.01 -0.05   0.16  -0.05   1.00  0.27  0.56  0.05
## AGMO             0.27  -0.04 -0.02  0.12   0.01  -0.06   0.27  1.00  0.22 -0.10
## TAF1            -0.03  -0.04 -0.02  0.12   0.12  -0.06   0.56  0.22  1.00  0.02
## GATA3            0.05  -0.11 -0.05 -0.07   0.11  -0.01   0.05 -0.10  0.02  1.00
## BIRC6           -0.05  -0.06 -0.03 -0.08   0.07   0.14   0.16  0.12  0.12  0.17
## DNAH11           0.06   0.22 -0.04  0.03   0.04  -0.14   0.06  0.15  0.03  0.02
## NF1              0.21   0.14 -0.02 -0.07  -0.04  -0.07  -0.04 -0.04 -0.04 -0.02
## TP53             0.08  -0.12  0.04 -0.03  -0.10   0.06  -0.20 -0.03 -0.15  0.08
## SYNE1            0.07  -0.10 -0.04  0.20   0.01   0.02   0.07  0.16  0.16 -0.02
## CTNNA1             NA     NA    NA    NA     NA     NA     NA    NA    NA    NA
## FOXO3           -0.04  -0.05 -0.02  0.06   0.11  -0.08   0.19  0.15  0.15  0.05
## LDLRAP1         -0.01  -0.02 -0.01 -0.03   0.17  -0.03  -0.01 -0.02 -0.02  0.18
## SURVIVAL_MONTHS -0.11  -0.06 -0.13  0.19   0.03   0.07   0.00 -0.06  0.03  0.09
## VITAL_STATUS     0.22   0.37  0.21 -0.12  -0.02  -0.13   0.08  0.05 -0.08 -0.11
##                 BIRC6 DNAH11   NF1  TP53 SYNE1 CTNNA1 FOXO3 LDLRAP1
## AKT1            -0.05   0.06  0.21  0.08  0.07     NA -0.04   -0.01
## MAP3K1          -0.06   0.22  0.14 -0.12 -0.10     NA -0.05   -0.02
## MEN1            -0.03  -0.04 -0.02  0.04 -0.04     NA -0.02   -0.01
## NCOR2           -0.08   0.03 -0.07 -0.03  0.20     NA  0.06   -0.03
## PIK3CA           0.07   0.04 -0.04 -0.10  0.01     NA  0.11    0.17
## SHANK2           0.14  -0.14 -0.07  0.06  0.02     NA -0.08   -0.03
## MAP2K4           0.16   0.06 -0.04 -0.20  0.07     NA  0.19   -0.01
## AGMO             0.12   0.15 -0.04 -0.03  0.16     NA  0.15   -0.02
## TAF1             0.12   0.03 -0.04 -0.15  0.16     NA  0.15   -0.02
## GATA3            0.17   0.02 -0.02  0.08 -0.02     NA  0.05    0.18
## BIRC6            1.00  -0.05 -0.07 -0.03  0.04     NA -0.07   -0.03
## DNAH11          -0.05   1.00 -0.01 -0.05  0.24     NA -0.03   -0.04
## NF1             -0.07  -0.01  1.00 -0.09 -0.01     NA  0.11   -0.02
## TP53            -0.03  -0.05 -0.09  1.00 -0.07     NA  0.03    0.04
## SYNE1            0.04   0.24 -0.01 -0.07  1.00     NA -0.02   -0.04
## CTNNA1             NA     NA    NA    NA    NA     NA    NA      NA
## FOXO3           -0.07  -0.03  0.11  0.03 -0.02     NA  1.00   -0.02
## LDLRAP1         -0.03  -0.04 -0.02  0.04 -0.04     NA -0.02    1.00
## SURVIVAL_MONTHS -0.14   0.05 -0.12  0.17 -0.14     NA  0.08    0.09
## VITAL_STATUS     0.05   0.09  0.11  0.09 -0.08     NA -0.11   -0.04
##                 SURVIVAL_MONTHS VITAL_STATUS
## AKT1                      -0.11         0.22
## MAP3K1                    -0.06         0.37
## MEN1                      -0.13         0.21
## NCOR2                      0.19        -0.12
## PIK3CA                     0.03        -0.02
## SHANK2                     0.07        -0.13
## MAP2K4                     0.00         0.08
## AGMO                      -0.06         0.05
## TAF1                       0.03        -0.08
## GATA3                      0.09        -0.11
## BIRC6                     -0.14         0.05
## DNAH11                     0.05         0.09
## NF1                       -0.12         0.11
## TP53                       0.17         0.09
## SYNE1                     -0.14        -0.08
## CTNNA1                       NA           NA
## FOXO3                      0.08        -0.11
## LDLRAP1                    0.09        -0.04
## SURVIVAL_MONTHS            1.00        -0.27
## VITAL_STATUS              -0.27         1.00
## Sample Size
## [1] 120
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##                 AKT1 MAP3K1 MEN1 NCOR2 PIK3CA SHANK2 MAP2K4 AGMO TAF1 GATA3
## AKT1            0.00   1.00 1.00  1.00   1.00   1.00   1.00 0.52 1.00  1.00
## MAP3K1          0.72   0.00 1.00  1.00   0.34   1.00   1.00 1.00 1.00  1.00
## MEN1            0.87   0.84 0.00  1.00   1.00   1.00   1.00 1.00 1.00  1.00
## NCOR2           0.62   0.52 0.78  0.00   1.00   1.00   1.00 1.00 1.00  1.00
## PIK3CA          0.68   0.00 0.58  0.94   0.00   1.00   1.00 1.00 1.00  1.00
```

```
## SHANK2          0.60   0.50 0.76  0.35   0.20   0.00   1.00 1.00 1.00  1.00
## MAP2K4          0.78   0.72 0.87  0.62   0.07   0.60   0.00 0.52 0.00  1.00
## AGMO            0.00   0.67 0.85  0.18   0.94   0.54   0.00 0.00 1.00  1.00
## TAF1            0.75   0.67 0.85  0.18   0.20   0.54   0.00 0.01 0.00  1.00
## GATA3           0.59   0.24 0.61  0.46   0.25   0.95   0.59 0.30 0.84  0.00
## BIRC6           0.62   0.52 0.78  0.38   0.47   0.12   0.09 0.18 0.18  0.07
## DNAH11          0.50   0.01 0.64  0.76   0.63   0.12   0.50 0.10 0.73  0.81
## NF1             0.02   0.12 0.82  0.48   0.69   0.45   0.69 0.64 0.64  0.80
## TP53            0.41   0.20 0.64  0.76   0.30   0.48   0.03 0.73 0.10  0.36
## SYNE1           0.47   0.30 0.65  0.03   0.96   0.83   0.47 0.08 0.08  0.83
## CTNNA1            NA     NA   NA    NA     NA     NA     NA   NA   NA    NA
## FOXO3           0.67   0.57 0.80  0.49   0.21   0.42   0.04 0.10 0.10  0.61
## LDLRAP1         0.87   0.84 0.93  0.78   0.07   0.76   0.87 0.85 0.85  0.05
## SURVIVAL_MONTHS 0.24   0.50 0.17  0.04   0.77   0.44   0.98 0.53 0.77  0.30
## VITAL_STATUS    0.01   0.00 0.02  0.18   0.80   0.15   0.40 0.61 0.38  0.23
##                 BIRC6 DNAH11  NF1 TP53 SYNE1 CTNNA1 FOXO3 LDLRAP1
## AKT1             1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## MAP3K1           1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## MEN1             1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## NCOR2            1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## PIK3CA           1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## SHANK2           1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## MAP2K4           1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## AGMO             1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## TAF1             1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## GATA3            1.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## BIRC6            0.00   1.00 1.00 1.00  1.00     NA  1.00    1.00
## DNAH11           0.56   0.00 1.00 1.00  1.00     NA  1.00    1.00
## NF1              0.48   0.91 0.00 1.00  1.00     NA  1.00    1.00
## TP53             0.76   0.56 0.33 0.00  1.00     NA  1.00    1.00
## SYNE1            0.70   0.01 0.96 0.48  0.00     NA  1.00    1.00
## CTNNA1             NA     NA   NA   NA    NA     NA    NA      NA
## FOXO3            0.44   0.78 0.25 0.78  0.82     NA  0.00    1.00
## LDLRAP1          0.78   0.64 0.82 0.64  0.65     NA  0.80    0.00
## SURVIVAL_MONTHS  0.12   0.61 0.18 0.07  0.12     NA  0.39    0.34
## VITAL_STATUS     0.59   0.33 0.23 0.34  0.39     NA  0.24    0.67
##                 SURVIVAL_MONTHS VITAL_STATUS
## AKT1                          1         1.00
## MAP3K1                        1         0.01
## MEN1                          1         1.00
## NCOR2                         1         1.00
## PIK3CA                        1         1.00
## SHANK2                        1         1.00
## MAP2K4                        1         1.00
## AGMO                          1         1.00
## TAF1                          1         1.00
## GATA3                         1         1.00
## BIRC6                         1         1.00
## DNAH11                        1         1.00
## NF1                           1         1.00
## TP53                          1         1.00
## SYNE1                         1         1.00
## CTNNA1                       NA           NA
## FOXO3                         1         1.00
## LDLRAP1                       1         1.00
## SURVIVAL_MONTHS               0         0.50
## VITAL_STATUS                  0         0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

*Correlation test findings - All Patients (*mutation)*

- The following variables showed low correlation to vital status: AKT1*, MAP3K1*,MEN1*, NCOR2*, SHANK2*, NPI, BREAST_SURGERY, and TUMOR_STAGE
- The following variables showed high correlation to vital status: SURVIVAL_MONTHS and SURVIVAL_GROUPS
- 5 clinical variables and 5 mutation variables respectively.

Correlation test findings - 70 patients with <= 5years follow-up

- The following variables showed low correlation to vital status: CANCER_TYPE, HER2_SNP6, SURVIVAL_MONTHS, HORMONE_THERAPY, MUTATION COUNT and NPI.
- The following variables showed moderate correlation to vital status: SHANK2*
- 6 clinical variables and 5 mutation variables respectively.

Correlation test findings - 120 patients with > 5years follow-up

- The following variables showed low correlation to vital status: AKT1*, MEN1*, NCOR2*, SHANK2*, SURVIVAL_MONTHS, PAM50 and INT.CLUST.
- The following variables showed moderate correlation to vital status: MAP3K1*

- 3 clinical variables and 5 mutation variables respectively.

- 1.69. Check significance of correlations of importance to identify group that will not benefit from treatment.

```
# Check the correlation significance levels of the 10 variables (for all 190 patients)

# mutation features (5)
cor.test(TNBC_All$VITAL_STATUS,TNBC_All$AKT1)$p.value #significant

## [1] 0.01051979

cor.test(TNBC_All$VITAL_STATUS,TNBC_All$MAP3K1)$p.value #significant

## [1] 0.002278348

cor.test(TNBC_All$VITAL_STATUS,TNBC_All$MEN1)$p.value #significant

## [1] 0.02167507

cor.test(TNBC_All$VITAL_STATUS,TNBC_All$NCOR2)$p.value #significant

## [1] 0.02737616

cor.test(TNBC_All$VITAL_STATUS,TNBC_All$SHANK2)$p.value #significant

## [1] 0.002476226

# clinical features(5)
cor.test(TNBC_All$VITAL_STATUS,TNBC_All$SURVIVAL_MONTHS)$p.value #very strong significance

## [1] 3.905855e-27

cor.test(TNBC_All$VITAL_STATUS,TNBC_All$NPI)$p.value #strong significance

## [1] 0.0004132662

cor.test(TNBC_All$VITAL_STATUS,TNBC_All$BREAST_SURGERY)$p.value #strong significance

## [1] 0.001891904

cor.test(TNBC_All$VITAL_STATUS,TNBC_All$SURVIVAL_GROUPS)$p.value #very strong significance

## [1] 1.411319e-33

cor.test(TNBC_All$VITAL_STATUS,TNBC_All$TUMOR_STAGE)$p.value #significant

## [1] 0.02570364
```

- There exists correlation significance between VITAL_STATUS and each of all the 10 variables tested for the group comprising all the 190 patients at p = 0.05.

```
# Check the correlation significance levels of the 11 variables (for 70-patient group)

# mutation features (5)
cor.test(TNBC5$VITAL_STATUS,TNBC5$AKT1)$p.value

## [1] 0.4845712

cor.test(TNBC5$VITAL_STATUS,TNBC5$MAP3K1)$p.value

## [1] 0.4845712

cor.test(TNBC5$VITAL_STATUS,TNBC5$MEN1)$p.value

## [1] 0.5941214

cor.test(TNBC5$VITAL_STATUS,TNBC5$NCOR2)$p.value

## [1] 0.7619302

cor.test(TNBC5$VITAL_STATUS,TNBC5$SHANK2)$p.value #strong significance
```

```
## [1] 0.0007597998

# clinical features (6)
cor.test(TNBC5$VITAL_STATUS,TNBC5$CANCER_TYPE)$p.value # significant

## [1] 0.02427381

cor.test(TNBC5$VITAL_STATUS,TNBC5$HER2_SNP6)$p.value #significant

## [1] 0.01603582

cor.test(TNBC5$VITAL_STATUS,TNBC5$SURVIVAL_MONTHS)$p.value #significant

## [1] 0.04169608

cor.test(TNBC5$VITAL_STATUS,TNBC5$HORMONE_THERAPY)$p.value

## [1] 0.08509972

cor.test(TNBC5$VITAL_STATUS,TNBC5$MUTATION_COUNT)$p.value

## [1] 0.07613006

cor.test(TNBC5$VITAL_STATUS,TNBC5$NPI)$p.value

## [1] 0.05244881
```

- There exists correlation significance between VITAL_STATUS and the following 4 variables (SHANK2, CANCER_TYPE, HER2_SNP6 and SURVIVAL_MONTHS) tested for the 70-patient sub-group at p = 0.05.

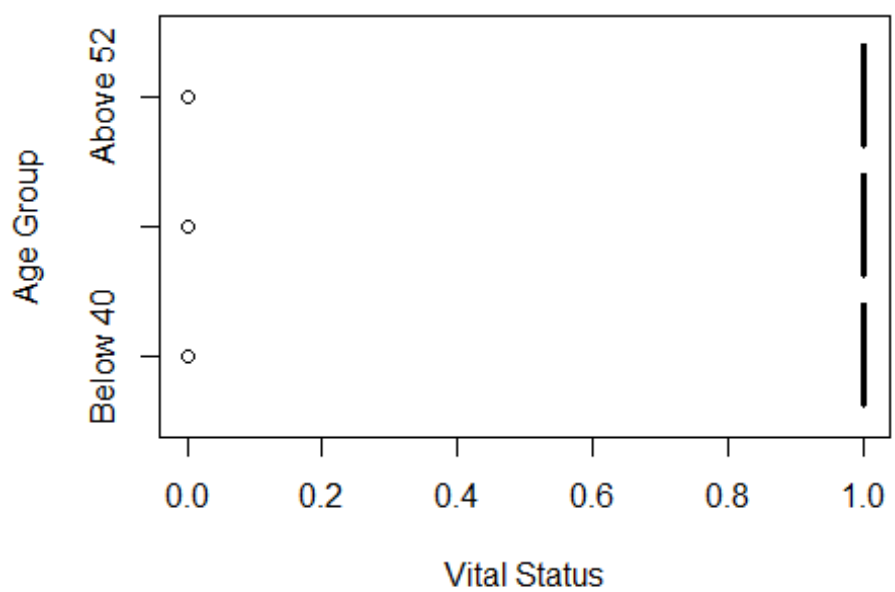
```
# Check the correlation significance levels of the 8 variables (for the 120 patient-subgroup)

# mutation features (5)
cor.test(TNBC_over5$VITAL_STATUS,TNBC_over5$AKT1)$p.value #significant

## [1] 0.01436811

cor.test(TNBC_over5$VITAL_STATUS,TNBC_over5$MAP3K1)$p.value #very strong significance

## [1] 3.836098e-05

cor.test(TNBC_over5$VITAL_STATUS,TNBC_over5$MEN1)$p.value #significant

## [1] 0.02048874

cor.test(TNBC_over5$VITAL_STATUS,TNBC_over5$NCOR2)$p.value

## [1] 0.178983

cor.test(TNBC_over5$VITAL_STATUS,TNBC_over5$SHANK2)$p.value

## [1] 0.1545393

# clinical features (3)
cor.test(TNBC_over5$VITAL_STATUS,TNBC_over5$SURVIVAL_MONTHS)$p.value #strong significance

## [1] 0.002955894

cor.test(TNBC_over5$VITAL_STATUS,TNBC_over5$PAM50)$p.value #significant

## [1] 0.01117429

cor.test(TNBC_over5$VITAL_STATUS,TNBC_over5$INTCLUST)$p.value #significant

## [1] 0.02130576
```

- There exists correlation significance between VITAL_STATUS and the following 6 variables (AKT, MAP3K1, MEN1, SURVIVAL_MONTHS, PAM50 and INTCLUST) tested for the 120-patient sub-group at p = 0.05.

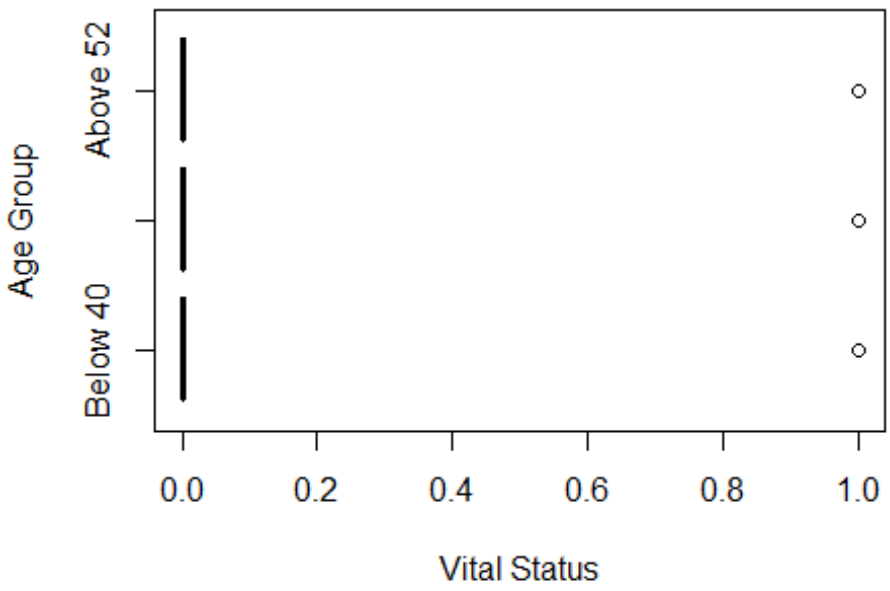
- 1.70. Analysis on the age variable to gain insight. Literature suggests that TNBC can occur in patients as young as 40years or less.

- 1.71. Perform median and mean operations on the age variable for the 70-patient and 120-patient groups respectively

```r
median(TNBC_All$AGE_AT_DIAGNOSIS)

## [1] 52.125

mean(TNBC_All$AGE_AT_DIAGNOSIS)

## [1] 52.44179

table(TNBC5$VITAL_STATUS)

##
##  0  1
##  6 64

prop.table(table(TNBC5$VITAL_STATUS))

##
##          0          1
## 0.08571429 0.91428571

boxplot(TNBC5$VITAL_STATUS~TNBC5$AGE_GROUP,horizontal = TRUE,
        xlab = "Vital Status", ylab = "Age Group", names = c("Below 40", "40-52", "Above 52"))
```




```r
table(TNBC_over5$VITAL_STATUS)

##
##   0   1
## 101  19

prop.table(table(TNBC_over5$VITAL_STATUS))

##
##         0         1
## 0.8416667 0.1583333

boxplot(TNBC_over5$VITAL_STATUS~TNBC_over5$AGE_GROUP,horizontal = TRUE,
        xlab = "Vital Status", ylab = "Age Group", names = c("Below 40", "40-52", "Above 52"))
```



- There is a negative correlation between patient vital status and survival. This suggests that as time progresses more and more patients are expected to die from the disease.
- The median and mean ages for the patients were 52.125yrs and 52.44yrs respectively.
- Note that according to the American Cancer Society, the median age for breast cancer diagnosis for white women is 62 and 59 for black women.
- It implies from this study, that there is a prevalence of TNBC occurrence in younger women by approximately 10yrs difference when compared to other cancer types.
- Of the 70 patients who were tracked for exactly 5years. 64 patients died of the disease during this period which represents 91.43% of the patients. On the other hand, only 6 of the patients survived up till this point representing just 8.57% of the patients.
- Of the 120 patients who were tracked beyond 5years, 19 patients died of the disease during this extended period which represents 15.83% of the patients. On the other hand, 101 of the patients survived, representing an impressive 84.17% of the patients.
==========

2.0. Method 2 - Survival Analysis on the METABRIC TNBC Dataset

- 2.1. Kaplan Meier Analysis [Kaplan Meier = KM for short]
- 2.2. First thing to do is to use Surv() to build the standard survival object
- 2.3. The variable 'SURVIVAL_MONTHS', records survival time - months to death;
- 2.4. "VITAL_STATUS' indicates whether the patient's death was observed
- 2.5. When death happens, VITAL_STATUS = 1, otherwise when censored, VITAL_STATUS = 0)
- 2.6. Kaplan Meier Survival Curve

```
library(survival) # Survival analysis package

## Warning: package 'survival' was built under R version 3.6.3
```

```
KM = with(TNBC_All, Surv(SURVIVAL_MONTHS,VITAL_STATUS))
KM

##   [1]  22.400000  153.966667+  28.500000    8.066667   66.733333  132.066667+
##   [7] 114.766667+  10.833333+  78.766667+  17.933333   31.300000  128.700000
##  [13] 141.166667+ 144.933333+ 146.900000+ 193.166667+ 188.333333+ 189.133333+
##  [19]  22.233333+  72.900000+ 188.133333+ 182.900000+ 164.733333+  55.200000
##  [25]  62.766667   60.666667+  26.266667   76.633333+  76.733333+   9.833333
##  [31]  43.133333   39.300000   88.233333+  36.400000   36.633333   71.800000
##  [37]  67.466667+  82.100000+ 114.466667+  65.866667+ 194.533333+   7.866667
##  [43]  48.800000+  75.400000+  24.800000   97.266667+  23.333333   18.933333
##  [49]  11.066667   16.600000  175.633333+ 136.166667+  52.966667  274.400000+
##  [55]  44.833333  235.666667+ 259.766667+  19.733333   43.200000    9.433333
##  [61] 250.666667+ 125.600000  267.400000+  31.933333   14.400000  262.133333+
##  [67] 227.833333  187.033333+ 263.233333+ 262.633333+  32.033333  254.266667+
##  [73] 146.366667+  28.566667  259.533333+ 167.433333+ 227.933333+ 145.500000+
##  [79] 248.766667+  32.933333  236.066667+  23.200000  226.733333+  32.833333
##  [85] 229.333333+ 236.033333+ 230.466667+  38.800000   75.333333  124.100000+
##  [91]  15.633333  307.933333+  68.133333   15.366667   83.366667  241.600000
##  [97] 200.600000+  39.866667  221.200000+ 149.866667+  19.733333  211.933333+
## [103]  21.700000   40.633333  274.200000+ 230.500000+  31.433333   70.600000
## [109]  19.166667   75.233333  223.300000+ 173.933333+ 178.733333   26.866667
## [115]  19.900000   30.366667   77.400000+  37.366667   43.266667+ 159.233333
## [121] 259.933333+  43.400000  144.666667   43.266667  204.200000  166.666667+
## [127]  53.566667+  79.300000+ 204.233333+ 194.300000+  45.633333   30.933333
## [133] 101.400000+ 194.566667+ 196.633333+ 162.833333  190.200000+  18.800000
## [139] 151.066667+  14.800000  183.966667+  98.566667+ 182.233333+ 186.600000+
## [145] 234.400000+ 178.633333+ 224.433333+ 180.633333+ 136.000000+  35.033333
## [151]  21.000000  200.433333+ 282.366667+ 255.600000+ 221.933333+ 165.433333+
## [157] 182.833333+  14.700000   22.400000   27.866667   89.533333   17.200000
## [163]  58.600000+  54.766667   19.833333  135.333333+  27.966667   80.233333+
## [169]  85.366667+ 120.433333+  81.033333+  94.933333  102.033333+ 106.133333+
## [175]  89.766667+   4.166667  140.233333+ 118.900000+ 109.766667+ 123.266667+
## [181] 128.366667+ 138.333333+ 149.766667+  21.933333  147.166667+ 166.033333+
## [187]  45.933333   55.000000   15.066667   21.166667
```

- 2.7. To begin analysis, we use the formula Surv(SURVIVAL_MONTHS,VITAL_STATUS) ~ 1 and the Survfit() function to produce the KM estimates of the probability of survival over time. The 'times' parameter of the summary() function gives some control over which times to print. Here it is set to print estimates for 1, 30, 60 and 90 months, and then every 90months thereafter.
- 2.8. Survival Analysis for overall 190 patient data

```
KM_fit_clinical = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~1, data = TNBC_All)
KM_fit_clinical

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ 1, data = TNBC_All)
##
##       n  events  median 0.95LCL 0.95UCL
##     190      83     242     159      NA

summary(KM_fit_clinical, times = c(10, 30, 60, 90*(1:10)))

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ 1, data = TNBC_All)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    10    185       5    0.974  0.0116        0.951        0.997
##    30    151      32    0.804  0.0289        0.749        0.863
##    60    120      27    0.660  0.0346        0.595        0.731
##    90     95       9    0.608  0.0360        0.541        0.682
##   180     51       7    0.549  0.0389        0.478        0.631
##   270      4       3    0.479  0.0521        0.387        0.593

summary(KM_fit_clinical)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ 1, data = TNBC_All)
##
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    4.17    190       1    0.995 0.00525        0.985        1.000
##    7.87    189       1    0.989 0.00740        0.975        1.000
##    8.07    188       1    0.984 0.00904        0.967        1.000
##    9.43    187       1    0.979 0.01041        0.959        1.000
##    9.83    186       1    0.974 0.01161        0.951        0.997
##   11.07    184       1    0.968 0.01270        0.944        0.994
##   14.40    183       1    0.963 0.01369        0.937        0.990
##   14.70    182       1    0.958 0.01460        0.930        0.987
##   14.80    181       1    0.953 0.01545        0.923        0.983
##   15.07    180       1    0.947 0.01624        0.916        0.980
```

```
##   15.37    179       1    0.942 0.01699        0.909        0.976
##   15.63    178       1    0.937 0.01770        0.903        0.972
##   16.60    177       1    0.931 0.01838        0.896        0.968
##   17.20    176       1    0.926 0.01902        0.890        0.964
##   17.93    175       1    0.921 0.01963        0.883        0.960
##   18.80    174       1    0.915 0.02022        0.877        0.956
##   18.93    173       1    0.910 0.02078        0.870        0.952
##   19.17    172       1    0.905 0.02133        0.864        0.948
##   19.73    171       2    0.894 0.02235        0.852        0.939
##   19.83    169       1    0.889 0.02284        0.845        0.935
##   19.90    168       1    0.884 0.02331        0.839        0.931
##   21.00    167       1    0.878 0.02376        0.833        0.926
##   21.17    166       1    0.873 0.02420        0.827        0.922
##   21.70    165       1    0.868 0.02462        0.821        0.917
##   21.93    164       1    0.863 0.02504        0.815        0.913
##   22.40    162       2    0.852 0.02583        0.803        0.904
##   23.20    160       1    0.847 0.02622        0.797        0.900
##   23.33    159       1    0.841 0.02659        0.791        0.895
##   24.80    158       1    0.836 0.02695        0.785        0.890
##   26.27    157       1    0.831 0.02729        0.779        0.886
##   26.87    156       1    0.825 0.02763        0.773        0.881
##   27.87    155       1    0.820 0.02796        0.767        0.877
##   27.97    154       1    0.815 0.02828        0.761        0.872
##   28.50    153       1    0.809 0.02860        0.755        0.867
##   28.57    152       1    0.804 0.02890        0.749        0.863
##   30.37    151       1    0.799 0.02920        0.743        0.858
##   30.93    150       1    0.793 0.02948        0.738        0.853
##   31.30    149       1    0.788 0.02976        0.732        0.849
##   31.43    148       1    0.783 0.03003        0.726        0.844
##   31.93    147       1    0.777 0.03030        0.720        0.839
##   32.03    146       1    0.772 0.03055        0.714        0.834
##   32.83    145       1    0.767 0.03080        0.709        0.830
##   32.93    144       1    0.761 0.03105        0.703        0.825
##   35.03    143       1    0.756 0.03128        0.697        0.820
##   36.40    142       1    0.751 0.03151        0.691        0.815
##   36.63    141       1    0.745 0.03173        0.686        0.810
##   37.37    140       1    0.740 0.03195        0.680        0.805
##   38.80    139       1    0.735 0.03216        0.674        0.801
##   39.30    138       1    0.729 0.03237        0.669        0.796
##   39.87    137       1    0.724 0.03257        0.663        0.791
##   40.63    136       1    0.719 0.03276        0.657        0.786
##   43.13    135       1    0.713 0.03295        0.652        0.781
##   43.20    134       1    0.708 0.03313        0.646        0.776
##   43.27    133       1    0.703 0.03330        0.640        0.771
##   43.40    131       1    0.697 0.03348        0.635        0.766
##   44.83    130       1    0.692 0.03365        0.629        0.761
##   45.63    129       1    0.687 0.03381        0.624        0.756
##   45.93    128       1    0.681 0.03397        0.618        0.751
##   52.97    126       1    0.676 0.03413        0.612        0.746
##   54.77    124       1    0.671 0.03429        0.607        0.741
##   55.00    123       1    0.665 0.03444        0.601        0.736
##   55.20    122       1    0.660 0.03458        0.595        0.731
##   62.77    119       1    0.654 0.03474        0.589        0.726
##   66.73    117       1    0.648 0.03489        0.584        0.721
##   68.13    115       1    0.643 0.03504        0.578        0.715
##   70.60    114       1    0.637 0.03518        0.572        0.710
##   71.80    113       1    0.632 0.03532        0.566        0.705
##   75.23    111       1    0.626 0.03545        0.560        0.699
##   75.33    110       1    0.620 0.03558        0.554        0.694
##   83.37    100       1    0.614 0.03577        0.548        0.688
##   89.53     97       1    0.608 0.03595        0.541        0.682
##   94.93     95       1    0.601 0.03614        0.534        0.676
##  125.60     82       1    0.594 0.03643        0.527        0.670
##  128.70     80       1    0.586 0.03673        0.519        0.663
##  144.67     72       1    0.578 0.03711        0.510        0.656
##  159.23     62       1    0.569 0.03766        0.500        0.648
##  162.83     61       1    0.560 0.03819        0.490        0.640
##  178.73     52       1    0.549 0.03894        0.478        0.631
##  204.20     33       1    0.532 0.04116        0.457        0.619
##  227.83     25       1    0.511 0.04468        0.431        0.607
##  241.60     16       1    0.479 0.05207        0.387        0.593
```

- 2.9. Survival Analysis for the 70-patient sub-group observed for 5years

```
KM_fit_clinical_5 = survfit(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~1, data = TNBC5)

KM_fit_clinical_5
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     1, data = TNBC5)
##
##       n  events  median 0.95LCL 0.95UCL
##    70.0    64.0    28.0    22.4    32.8
```

```
summary(KM_fit_clinical_5, times = c(10, 20, 40, 50, 55, 60))
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     1, data = TNBC5)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    10     65       5   0.9286  0.0308       0.8702        0.991
##    20     47      17   0.6819  0.0560       0.5805        0.801
##    40     16      30   0.2377  0.0518       0.1551        0.364
##    50      6       8   0.1134  0.0393       0.0575        0.224
##    55      3       3   0.0473  0.0300       0.0136        0.164
```

```
summary(KM_fit_clinical_5)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     1, data = TNBC5)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   4.17     70       1   0.9857  0.0142      0.95830        1.000
##   7.87     69       1   0.9714  0.0199      0.93317        1.000
##   8.07     68       1   0.9571  0.0242      0.91085        1.000
##   9.43     67       1   0.9429  0.0277      0.89002        0.999
##   9.83     66       1   0.9286  0.0308      0.87016        0.991
##  11.07     64       1   0.9141  0.0335      0.85062        0.982
##  14.40     63       1   0.8996  0.0360      0.83166        0.973
##  14.70     62       1   0.8850  0.0382      0.81317        0.963
##  14.80     61       1   0.8705  0.0403      0.79507        0.953
##  15.07     60       1   0.8560  0.0421      0.77730        0.943
##  15.37     59       1   0.8415  0.0439      0.75982        0.932
##  15.63     58       1   0.8270  0.0454      0.74259        0.921
##  16.60     57       1   0.8125  0.0469      0.72560        0.910
##  17.20     56       1   0.7980  0.0482      0.70881        0.898
##  17.93     55       1   0.7835  0.0495      0.69222        0.887
##  18.80     54       1   0.7690  0.0507      0.67581        0.875
##  18.93     53       1   0.7545  0.0517      0.65956        0.863
##  19.17     52       1   0.7400  0.0527      0.64347        0.851
##  19.73     51       2   0.7109  0.0545      0.61171        0.826
##  19.83     49       1   0.6964  0.0553      0.59604        0.814
##  19.90     48       1   0.6819  0.0560      0.58049        0.801
##  21.00     47       1   0.6674  0.0567      0.56507        0.788
##  21.17     46       1   0.6529  0.0573      0.54976        0.775
##  21.70     45       1   0.6384  0.0578      0.53457        0.762
##  21.93     44       1   0.6239  0.0583      0.51949        0.749
##  22.40     42       2   0.5942  0.0592      0.48880        0.722
##  23.20     40       1   0.5793  0.0595      0.47363        0.709
##  23.33     39       1   0.5645  0.0598      0.45858        0.695
##  24.80     38       1   0.5496  0.0601      0.44363        0.681
##  26.27     37       1   0.5348  0.0603      0.42878        0.667
##  26.87     36       1   0.5199  0.0604      0.41405        0.653
##  27.87     35       1   0.5050  0.0605      0.39942        0.639
##  27.97     34       1   0.4902  0.0605      0.38490        0.624
##  28.50     33       1   0.4753  0.0604      0.37048        0.610
##  28.57     32       1   0.4605  0.0604      0.35617        0.595
##  30.37     31       1   0.4456  0.0602      0.34196        0.581
##  30.93     30       1   0.4308  0.0600      0.32786        0.566
##  31.30     29       1   0.4159  0.0597      0.31386        0.551
##  31.43     28       1   0.4011  0.0594      0.29998        0.536
##  31.93     27       1   0.3862  0.0591      0.28620        0.521
##  32.03     26       1   0.3714  0.0586      0.27254        0.506
##  32.83     25       1   0.3565  0.0581      0.25898        0.491
##  32.93     24       1   0.3417  0.0576      0.24555        0.475
##  35.03     23       1   0.3268  0.0570      0.23223        0.460
##  36.40     22       1   0.3119  0.0563      0.21904        0.444
##  36.63     21       1   0.2971  0.0555      0.20598        0.428
##  37.37     20       1   0.2822  0.0547      0.19305        0.413
##  38.80     19       1   0.2674  0.0538      0.18025        0.397
##  39.30     18       1   0.2525  0.0528      0.16760        0.380
##  39.87     17       1   0.2377  0.0518      0.15510        0.364
##  40.63     16       1   0.2228  0.0506      0.14276        0.348
##  43.13     15       1   0.2080  0.0494      0.13059        0.331
##  43.20     14       1   0.1931  0.0480      0.11861        0.314
##  43.27     13       1   0.1783  0.0466      0.10682        0.297
```

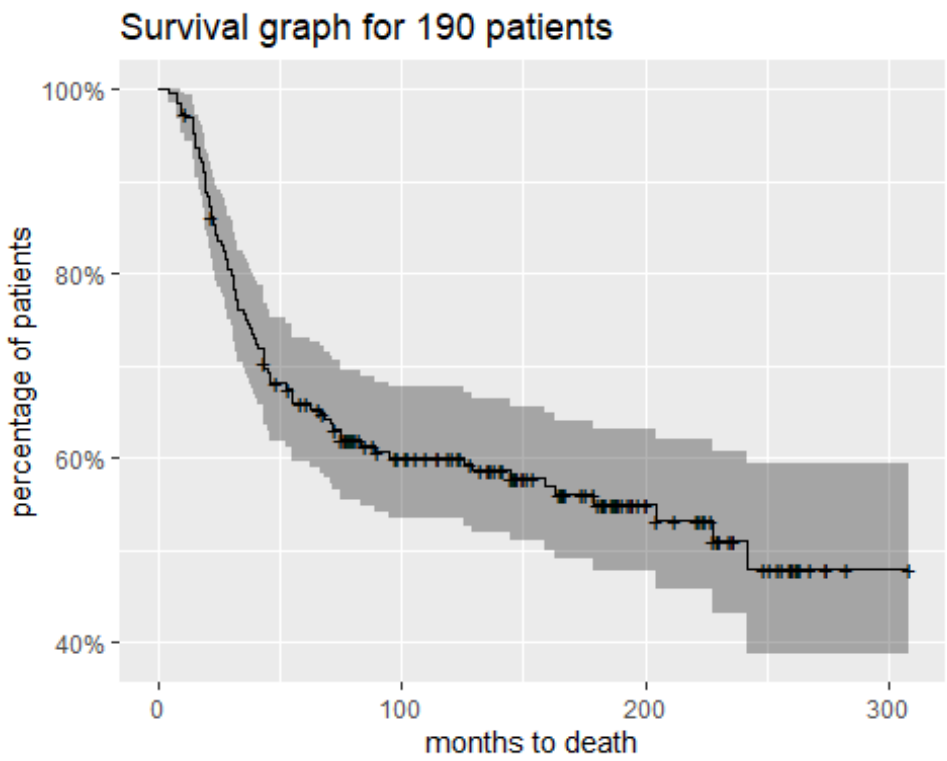
```
##  43.40     11       1   0.1620  0.0451      0.09395        0.279
##  44.83     10       1   0.1458  0.0434      0.08142        0.261
##  45.63      9       1   0.1296  0.0415      0.06925        0.243
##  45.93      8       1   0.1134  0.0393      0.05749        0.224
##  52.97      6       1   0.0945  0.0370      0.04386        0.204
##  54.77      4       1   0.0709  0.0345      0.02731        0.184
##  55.00      3       1   0.0473  0.0300      0.01361        0.164
##  55.20      2       1   0.0236  0.0225      0.00367        0.152
```

- 2.10. Survival Analysis for overall 120-patient subgroup observed for more than 5years

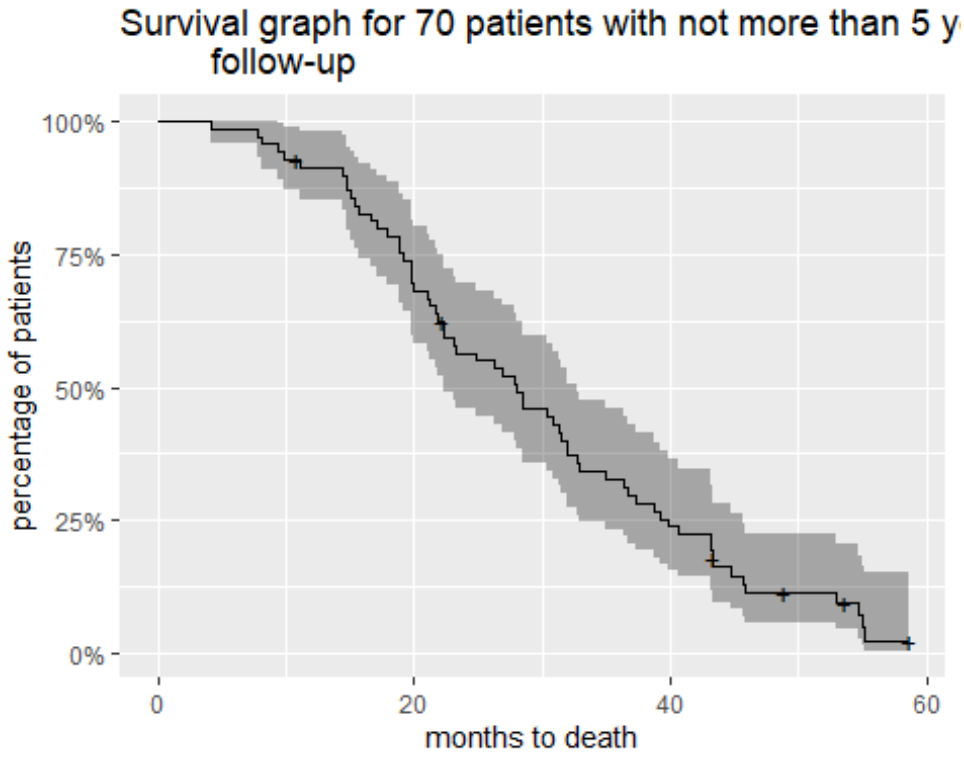
```
KM_fit_clinical_over_5 = survfit(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~1, data = TNBC_over5)
KM_fit_clinical_over_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     1, data = TNBC_over5)
##
##       n  events  median 0.95LCL 0.95UCL
##     120      19      NA      NA      NA

summary(KM_fit_clinical_over_5, times = c(10, 30, 60, 90*(1:10)))

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     1, data = TNBC_over5)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    10    120       0    1.000  0.0000        1.000        1.000
##    30    120       0    1.000  0.0000        1.000        1.000
##    60    120       0    1.000  0.0000        1.000        1.000
##    90     95       9    0.921  0.0253        0.873        0.972
##   180     51       7    0.832  0.0398        0.758        0.914
##   270      4       3    0.726  0.0691        0.603        0.875

summary(KM_fit_clinical_over_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     1, data = TNBC_over5)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   62.8    119       1    0.992 0.00837        0.975        1.000
##   66.7    117       1    0.983 0.01183        0.960        1.000
##   68.1    115       1    0.975 0.01449        0.947        1.000
##   70.6    114       1    0.966 0.01670        0.934        0.999
##   71.8    113       1    0.957 0.01861        0.922        0.995
##   75.2    111       1    0.949 0.02034        0.910        0.990
##   75.3    110       1    0.940 0.02191        0.898        0.984
##   83.4    100       1    0.931 0.02362        0.886        0.978
##   89.5     97       1    0.921 0.02525        0.873        0.972
##   94.9     95       1    0.912 0.02679        0.861        0.966
##  125.6     82       1    0.900 0.02867        0.846        0.958
##  128.7     80       1    0.889 0.03044        0.831        0.951
##  144.7     72       1    0.877 0.03243        0.815        0.943
##  159.2     62       1    0.863 0.03485        0.797        0.934
##  162.8     61       1    0.849 0.03704        0.779        0.924
##  178.7     52       1    0.832 0.03976        0.758        0.914
##  204.2     33       1    0.807 0.04586        0.722        0.902
##  227.8     25       1    0.775 0.05421        0.675        0.889
##  241.6     16       1    0.726 0.06914        0.603        0.875
```

- 2.11. Plot survival graphs using the ggfortify package

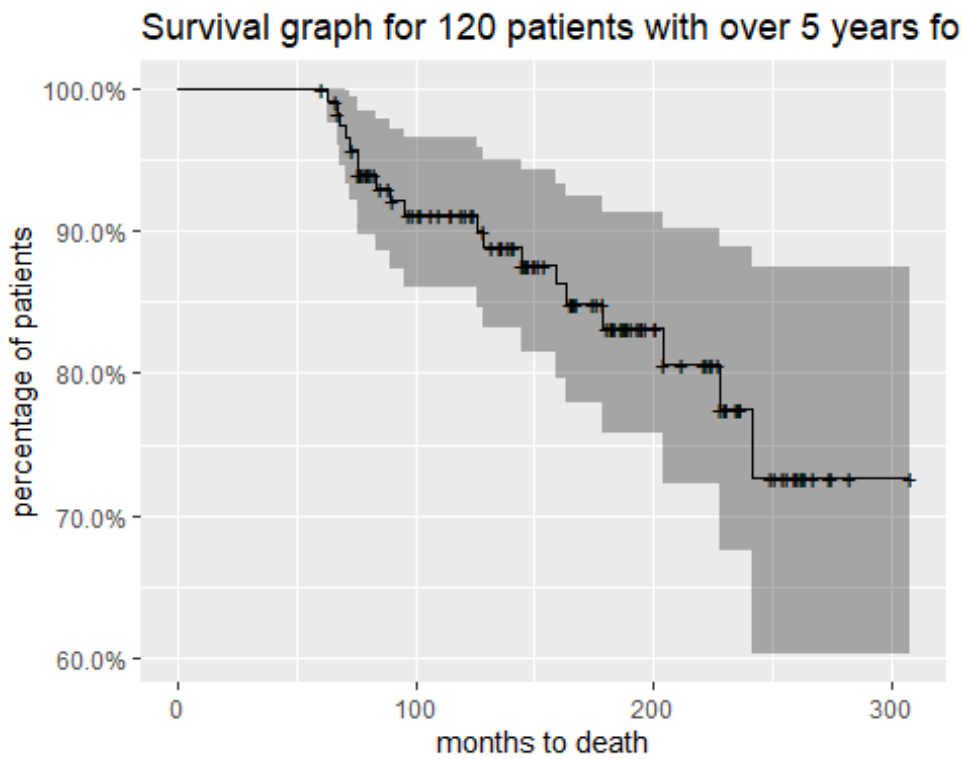
```
library(ggfortify)

## Warning: package 'ggfortify' was built under R version 3.6.3

autoplot(KM_fit_clinical,main = "Survival graph for 190 patients",
         xlab = "months to death", ylab = "percentage of patients")
```




```
autoplot(KM_fit_clinical_5,main = "Survival graph for 70 patients with not more than 5 years
         follow-up",xlab = "months to death", ylab = "percentage of patients")
```




```
autoplot(KM_fit_clinical_over_5,main = "Survival graph for 120 patients with over 5 years follow-up",
         xlab = "months to death", ylab = "percentage of patients")
```

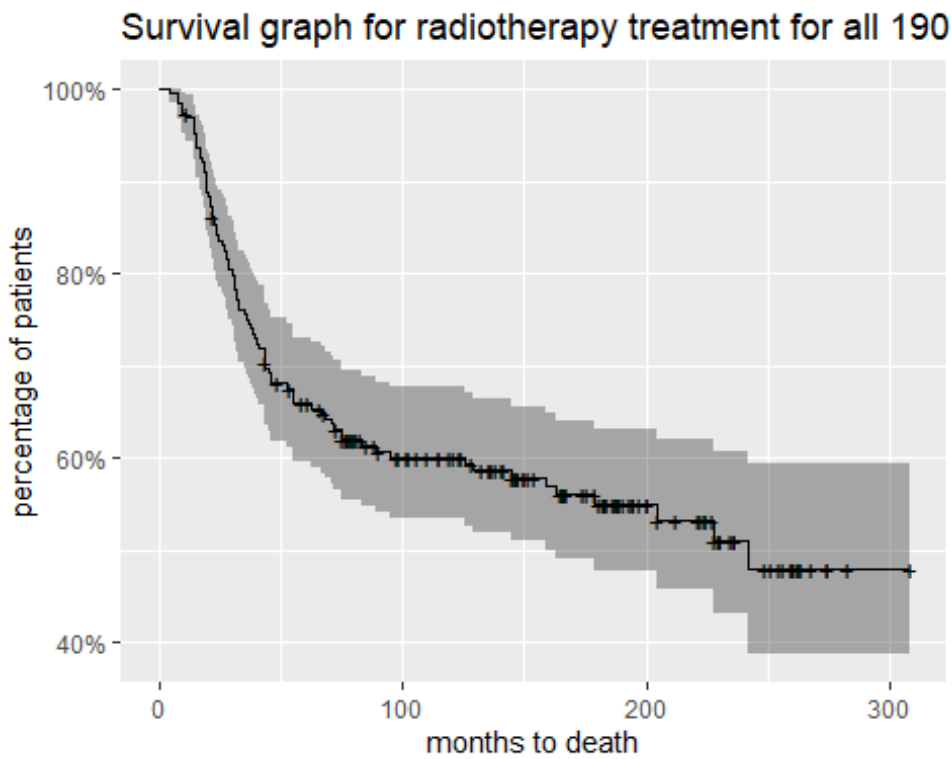


2.1.1. Single Treatments

- 2.12. Create objects for survival curve for radiotherapy treatment
- 2.13. For Log-rank test, rho = 0; whereas for Generalized Wilcoxon test, rho = 1
- 2.14 Plot survival graphs for radiotherapy treatments

```
KM_trt_fit_clinical = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~RADIO_THERAPY, data = TNBC_All)
KM_trt_fit_clinical

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC_All)
##
##       n  events  median 0.95LCL 0.95UCL
##     190      83     242     159      NA

summary(KM_trt_fit_clinical,60)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC_All)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    60    120      64     0.66  0.0346        0.595        0.731

summary(KM_trt_fit_clinical)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC_All)
##
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    4.17    190       1    0.995 0.00525        0.985        1.000
##    7.87    189       1    0.989 0.00740        0.975        1.000
##    8.07    188       1    0.984 0.00904        0.967        1.000
##    9.43    187       1    0.979 0.01041        0.959        1.000
##    9.83    186       1    0.974 0.01161        0.951        0.997
##   11.07    184       1    0.968 0.01270        0.944        0.994
##   14.40    183       1    0.963 0.01369        0.937        0.990
##   14.70    182       1    0.958 0.01460        0.930        0.987
```

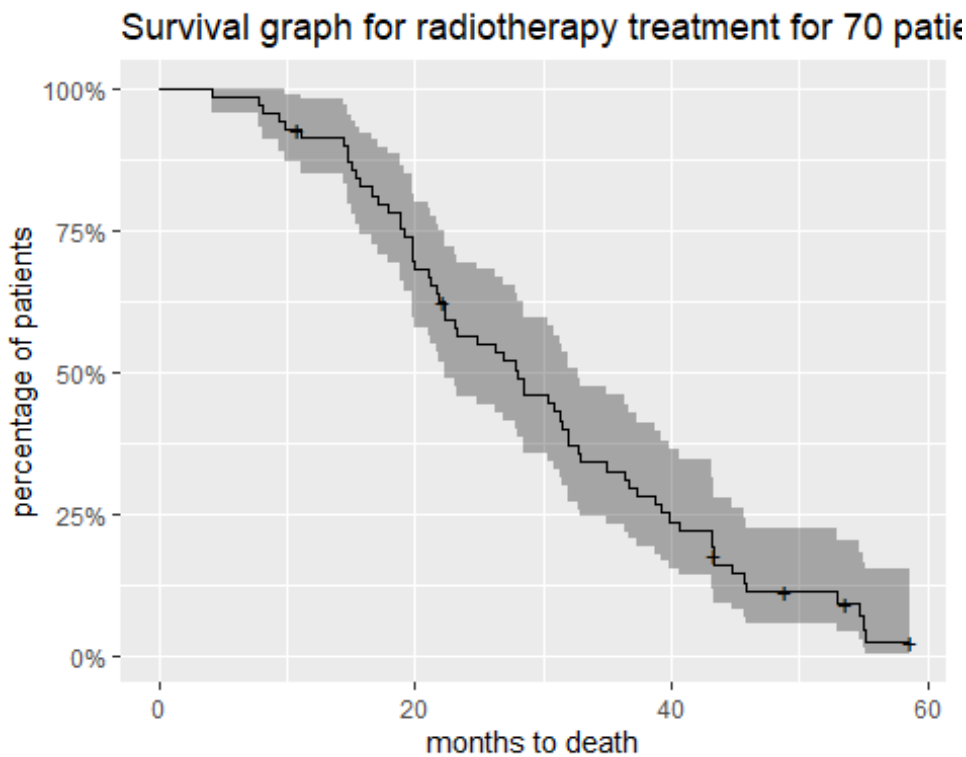
```
##   14.80    181       1    0.953 0.01545        0.923        0.983
##   15.07    180       1    0.947 0.01624        0.916        0.980
##   15.37    179       1    0.942 0.01699        0.909        0.976
##   15.63    178       1    0.937 0.01770        0.903        0.972
##   16.60    177       1    0.931 0.01838        0.896        0.968
##   17.20    176       1    0.926 0.01902        0.890        0.964
##   17.93    175       1    0.921 0.01963        0.883        0.960
##   18.80    174       1    0.915 0.02022        0.877        0.956
##   18.93    173       1    0.910 0.02078        0.870        0.952
##   19.17    172       1    0.905 0.02133        0.864        0.948
##   19.73    171       2    0.894 0.02235        0.852        0.939
##   19.83    169       1    0.889 0.02284        0.845        0.935
##   19.90    168       1    0.884 0.02331        0.839        0.931
##   21.00    167       1    0.878 0.02376        0.833        0.926
##   21.17    166       1    0.873 0.02420        0.827        0.922
##   21.70    165       1    0.868 0.02462        0.821        0.917
##   21.93    164       1    0.863 0.02504        0.815        0.913
##   22.40    162       2    0.852 0.02583        0.803        0.904
##   23.20    160       1    0.847 0.02622        0.797        0.900
##   23.33    159       1    0.841 0.02659        0.791        0.895
##   24.80    158       1    0.836 0.02695        0.785        0.890
##   26.27    157       1    0.831 0.02729        0.779        0.886
##   26.87    156       1    0.825 0.02763        0.773        0.881
##   27.87    155       1    0.820 0.02796        0.767        0.877
##   27.97    154       1    0.815 0.02828        0.761        0.872
##   28.50    153       1    0.809 0.02860        0.755        0.867
##   28.57    152       1    0.804 0.02890        0.749        0.863
##   30.37    151       1    0.799 0.02920        0.743        0.858
##   30.93    150       1    0.793 0.02948        0.738        0.853
##   31.30    149       1    0.788 0.02976        0.732        0.849
##   31.43    148       1    0.783 0.03003        0.726        0.844
##   31.93    147       1    0.777 0.03030        0.720        0.839
##   32.03    146       1    0.772 0.03055        0.714        0.834
##   32.83    145       1    0.767 0.03080        0.709        0.830
##   32.93    144       1    0.761 0.03105        0.703        0.825
##   35.03    143       1    0.756 0.03128        0.697        0.820
##   36.40    142       1    0.751 0.03151        0.691        0.815
##   36.63    141       1    0.745 0.03173        0.686        0.810
##   37.37    140       1    0.740 0.03195        0.680        0.805
##   38.80    139       1    0.735 0.03216        0.674        0.801
##   39.30    138       1    0.729 0.03237        0.669        0.796
##   39.87    137       1    0.724 0.03257        0.663        0.791
##   40.63    136       1    0.719 0.03276        0.657        0.786
##   43.13    135       1    0.713 0.03295        0.652        0.781
##   43.20    134       1    0.708 0.03313        0.646        0.776
##   43.27    133       1    0.703 0.03330        0.640        0.771
##   43.40    131       1    0.697 0.03348        0.635        0.766
##   44.83    130       1    0.692 0.03365        0.629        0.761
##   45.63    129       1    0.687 0.03381        0.624        0.756
##   45.93    128       1    0.681 0.03397        0.618        0.751
##   52.97    126       1    0.676 0.03413        0.612        0.746
##   54.77    124       1    0.671 0.03429        0.607        0.741
##   55.00    123       1    0.665 0.03444        0.601        0.736
##   55.20    122       1    0.660 0.03458        0.595        0.731
##   62.77    119       1    0.654 0.03474        0.589        0.726
##   66.73    117       1    0.648 0.03489        0.584        0.721
##   68.13    115       1    0.643 0.03504        0.578        0.715
##   70.60    114       1    0.637 0.03518        0.572        0.710
##   71.80    113       1    0.632 0.03532        0.566        0.705
##   75.23    111       1    0.626 0.03545        0.560        0.699
##   75.33    110       1    0.620 0.03558        0.554        0.694
##   83.37    100       1    0.614 0.03577        0.548        0.688
##   89.53     97       1    0.608 0.03595        0.541        0.682
##   94.93     95       1    0.601 0.03614        0.534        0.676
##  125.60     82       1    0.594 0.03643        0.527        0.670
##  128.70     80       1    0.586 0.03673        0.519        0.663
##  144.67     72       1    0.578 0.03711        0.510        0.656
##  159.23     62       1    0.569 0.03766        0.500        0.648
##  162.83     61       1    0.560 0.03819        0.490        0.640
##  178.73     52       1    0.549 0.03894        0.478        0.631
##  204.20     33       1    0.532 0.04116        0.457        0.619
##  227.83     25       1    0.511 0.04468        0.431        0.607
##  241.60     16       1    0.479 0.05207        0.387        0.593

autoplot(KM_trt_fit_clinical,main = "Survival graph for radiotherapy treatment for all 190 patients",
         xlab = "months to death", ylab = "percentage of patients", facets = TRUE, colour = TRUE)
```




```
KM_trt_fit_clinical_5 = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~RADIO_THERAPY, data = TNBC5)
KM_trt_fit_clinical_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC5)
##
##       n  events  median 0.95LCL 0.95UCL
##    70.0    64.0    28.0    22.4    32.8

summary(KM_trt_fit_clinical_5,60, extend = TRUE)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC5)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    60      0      64   0.0236  0.0225      0.00367        0.152

summary(KM_trt_fit_clinical_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC5)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   4.17     70       1   0.9857  0.0142      0.95830        1.000
##   7.87     69       1   0.9714  0.0199      0.93317        1.000
##   8.07     68       1   0.9571  0.0242      0.91085        1.000
##   9.43     67       1   0.9429  0.0277      0.89002        0.999
##   9.83     66       1   0.9286  0.0308      0.87016        0.991
##  11.07     64       1   0.9141  0.0335      0.85062        0.982
##  14.40     63       1   0.8996  0.0360      0.83166        0.973
##  14.70     62       1   0.8850  0.0382      0.81317        0.963
##  14.80     61       1   0.8705  0.0403      0.79507        0.953
##  15.07     60       1   0.8560  0.0421      0.77730        0.943
##  15.37     59       1   0.8415  0.0439      0.75982        0.932
##  15.63     58       1   0.8270  0.0454      0.74259        0.921
##  16.60     57       1   0.8125  0.0469      0.72560        0.910
##  17.20     56       1   0.7980  0.0482      0.70881        0.898
##  17.93     55       1   0.7835  0.0495      0.69222        0.887
##  18.80     54       1   0.7690  0.0507      0.67581        0.875
##  18.93     53       1   0.7545  0.0517      0.65956        0.863
##  19.17     52       1   0.7400  0.0527      0.64347        0.851
```

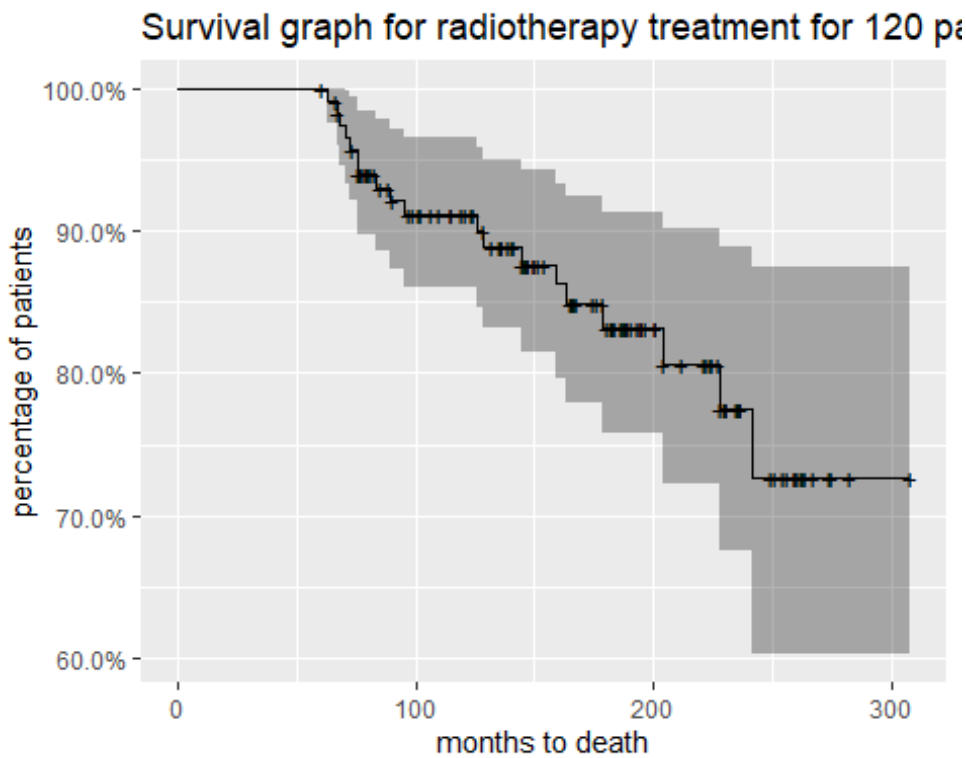
```
## 19.73     51       2   0.7109  0.0545      0.61171        0.826
## 19.83     49       1   0.6964  0.0553      0.59604        0.814
## 19.90     48       1   0.6819  0.0560      0.58049        0.801
## 21.00     47       1   0.6674  0.0567      0.56507        0.788
## 21.17     46       1   0.6529  0.0573      0.54976        0.775
## 21.70     45       1   0.6384  0.0578      0.53457        0.762
## 21.93     44       1   0.6239  0.0583      0.51949        0.749
## 22.40     42       2   0.5942  0.0592      0.48880        0.722
## 23.20     40       1   0.5793  0.0595      0.47363        0.709
## 23.33     39       1   0.5645  0.0598      0.45858        0.695
## 24.80     38       1   0.5496  0.0601      0.44363        0.681
## 26.27     37       1   0.5348  0.0603      0.42878        0.667
## 26.87     36       1   0.5199  0.0604      0.41405        0.653
## 27.87     35       1   0.5050  0.0605      0.39942        0.639
## 27.97     34       1   0.4902  0.0605      0.38490        0.624
## 28.50     33       1   0.4753  0.0604      0.37048        0.610
## 28.57     32       1   0.4605  0.0604      0.35617        0.595
## 30.37     31       1   0.4456  0.0602      0.34196        0.581
## 30.93     30       1   0.4308  0.0600      0.32786        0.566
## 31.30     29       1   0.4159  0.0597      0.31386        0.551
## 31.43     28       1   0.4011  0.0594      0.29998        0.536
## 31.93     27       1   0.3862  0.0591      0.28620        0.521
## 32.03     26       1   0.3714  0.0586      0.27254        0.506
## 32.83     25       1   0.3565  0.0581      0.25898        0.491
## 32.93     24       1   0.3417  0.0576      0.24555        0.475
## 35.03     23       1   0.3268  0.0570      0.23223        0.460
## 36.40     22       1   0.3119  0.0563      0.21904        0.444
## 36.63     21       1   0.2971  0.0555      0.20598        0.428
## 37.37     20       1   0.2822  0.0547      0.19305        0.413
## 38.80     19       1   0.2674  0.0538      0.18025        0.397
## 39.30     18       1   0.2525  0.0528      0.16760        0.380
## 39.87     17       1   0.2377  0.0518      0.15510        0.364
## 40.63     16       1   0.2228  0.0506      0.14276        0.348
## 43.13     15       1   0.2080  0.0494      0.13059        0.331
## 43.20     14       1   0.1931  0.0480      0.11861        0.314
## 43.27     13       1   0.1783  0.0466      0.10682        0.297
## 43.40     11       1   0.1620  0.0451      0.09395        0.279
## 44.83     10       1   0.1458  0.0434      0.08142        0.261
## 45.63      9       1   0.1296  0.0415      0.06925        0.243
## 45.93      8       1   0.1134  0.0393      0.05749        0.224
## 52.97      6       1   0.0945  0.0370      0.04386        0.204
## 54.77      4       1   0.0709  0.0345      0.02731        0.184
## 55.00      3       1   0.0473  0.0300      0.01361        0.164
## 55.20      2       1   0.0236  0.0225      0.00367        0.152
```

```
autoplot(KM_trt_fit_clinical_5,
         main = "Survival graph for radiotherapy treatment for 70 patients with not more than 5years follow-up",
         xlab = "months to death", ylab = "percentage of patients", facets = TRUE, colour = TRUE)
```




```
KM_trt_fit_clinical_over_5 = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~RADIO_THERAPY, data = TNBC_over5)
KM_trt_fit_clinical_over_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC_over5)
##
##       n  events  median 0.95LCL 0.95UCL
##     120      19      NA      NA      NA

summary(KM_trt_fit_clinical_over_5,60)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC_over5)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    60    120       0        1       0            1            1

summary(KM_trt_fit_clinical_over_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC_over5)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   62.8    119       1    0.992 0.00837        0.975        1.000
##   66.7    117       1    0.983 0.01183        0.960        1.000
##   68.1    115       1    0.975 0.01449        0.947        1.000
##   70.6    114       1    0.966 0.01670        0.934        0.999
##   71.8    113       1    0.957 0.01861        0.922        0.995
##   75.2    111       1    0.949 0.02034        0.910        0.990
##   75.3    110       1    0.940 0.02191        0.898        0.984
##   83.4    100       1    0.931 0.02362        0.886        0.978
##   89.5     97       1    0.921 0.02525        0.873        0.972
##   94.9     95       1    0.912 0.02679        0.861        0.966
##  125.6     82       1    0.900 0.02867        0.846        0.958
##  128.7     80       1    0.889 0.03044        0.831        0.951
##  144.7     72       1    0.877 0.03243        0.815        0.943
##  159.2     62       1    0.863 0.03485        0.797        0.934
##  162.8     61       1    0.849 0.03704        0.779        0.924
##  178.7     52       1    0.832 0.03976        0.758        0.914
##  204.2     33       1    0.807 0.04586        0.722        0.902
```

```
##  227.8     25       1    0.775 0.05421        0.675        0.889
##  241.6     16       1    0.726 0.06914        0.603        0.875

autoplot(KM_trt_fit_clinical_over_5,
         main = "Survival graph for radiotherapy treatment for 120 patients with more than 5years follow-up",
         xlab = "months to death", ylab = "percentage of patients", facets = TRUE, colour = TRUE)
```



- 2.15 Create survival objects and plot survival graphs for chemotherapy treatments

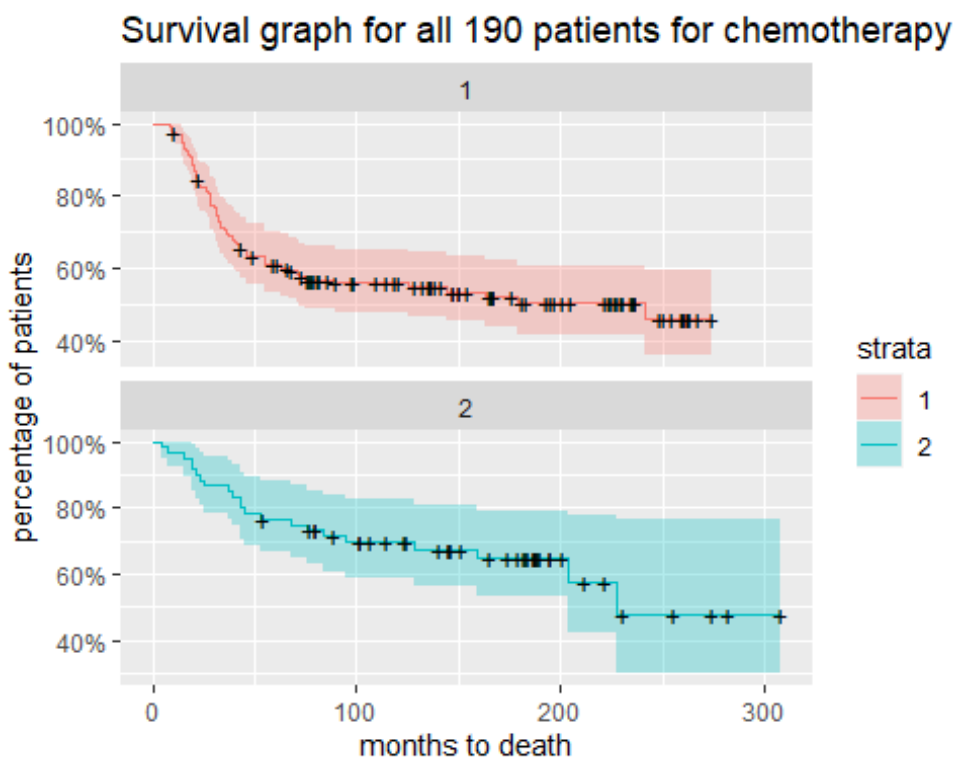
```
KM_trt_fit2_clinical = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~CHEMOTHERAPY, data = TNBC_All)
KM_trt_fit2_clinical

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ CHEMOTHERAPY,
##     data = TNBC_All)
##
##                   n events median 0.95LCL 0.95UCL
## CHEMOTHERAPY=1 130     61    242    71.8      NA
## CHEMOTHERAPY=2  60     22    228   204.2      NA

summary(KM_trt_fit2_clinical)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ CHEMOTHERAPY,
##     data = TNBC_All)
##
##                 CHEMOTHERAPY=1
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    8.07    130       1    0.992 0.00766        0.977        1.000
##    9.43    129       1    0.985 0.01079        0.964        1.000
##    9.83    128       1    0.977 0.01317        0.951        1.000
##   11.07    126       1    0.969 0.01518        0.940        0.999
##   14.40    125       1    0.961 0.01692        0.929        0.995
##   14.70    124       1    0.954 0.01847        0.918        0.991
##   14.80    123       1    0.946 0.01988        0.908        0.986
##   15.07    122       1    0.938 0.02118        0.898        0.981
##   15.63    121       1    0.930 0.02238        0.888        0.975
##   16.60    120       1    0.923 0.02350        0.878        0.970
##   17.20    119       1    0.915 0.02455        0.868        0.964
##   17.93    118       1    0.907 0.02553        0.858        0.959
##   18.93    117       1    0.899 0.02647        0.849        0.953
##   19.73    116       2    0.884 0.02819        0.830        0.941
##   19.83    114       1    0.876 0.02899        0.821        0.935
```

```
##   19.90    113       1    0.868 0.02975        0.812        0.929
##   21.00    112       1    0.861 0.03048        0.803        0.922
##   21.17    111       1    0.853 0.03117        0.794        0.916
##   21.93    110       1    0.845 0.03184        0.785        0.910
##   22.40    108       2    0.829 0.03312        0.767        0.897
##   23.20    106       1    0.822 0.03372        0.758        0.890
##   26.27    105       1    0.814 0.03429        0.749        0.884
##   26.87    104       1    0.806 0.03484        0.741        0.877
##   27.87    103       1    0.798 0.03537        0.732        0.871
##   27.97    102       1    0.790 0.03588        0.723        0.864
##   28.50    101       1    0.783 0.03637        0.714        0.857
##   28.57    100       1    0.775 0.03684        0.706        0.850
##   30.37     99       1    0.767 0.03729        0.697        0.844
##   30.93     98       1    0.759 0.03772        0.689        0.837
##   31.30     97       1    0.751 0.03813        0.680        0.830
##   31.43     96       1    0.743 0.03853        0.672        0.823
##   31.93     95       1    0.736 0.03891        0.663        0.816
##   32.03     94       1    0.728 0.03928        0.655        0.809
##   32.83     93       1    0.720 0.03963        0.646        0.802
##   32.93     92       1    0.712 0.03996        0.638        0.795
##   35.03     91       1    0.704 0.04028        0.630        0.788
##   36.40     90       1    0.696 0.04059        0.621        0.781
##   36.63     89       1    0.689 0.04088        0.613        0.774
##   39.30     88       1    0.681 0.04116        0.605        0.766
##   39.87     87       1    0.673 0.04142        0.596        0.759
##   40.63     86       1    0.665 0.04167        0.588        0.752
##   43.13     85       1    0.657 0.04191        0.580        0.745
##   43.40     83       1    0.649 0.04215        0.572        0.737
##   45.63     82       1    0.641 0.04237        0.564        0.730
##   45.93     81       1    0.634 0.04258        0.555        0.723
##   54.77     79       1    0.626 0.04279        0.547        0.715
##   55.00     78       1    0.618 0.04299        0.539        0.708
##   55.20     77       1    0.609 0.04317        0.530        0.700
##   62.77     74       1    0.601 0.04336        0.522        0.693
##   66.73     72       1    0.593 0.04356        0.513        0.685
##   70.60     70       1    0.584 0.04375        0.505        0.677
##   71.80     69       1    0.576 0.04393        0.496        0.669
##   75.23     67       1    0.567 0.04411        0.487        0.661
##   89.53     59       1    0.558 0.04440        0.477        0.652
##  125.60     51       1    0.547 0.04485        0.466        0.642
##  144.67     43       1    0.534 0.04558        0.452        0.631
##  162.83     37       1    0.520 0.04657        0.436        0.619
##  178.73     31       1    0.503 0.04799        0.417        0.606
##  241.60     12       1    0.461 0.05954        0.358        0.594
##
##                 CHEMOTHERAPY=2
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    4.17     60       1    0.983  0.0165        0.951        1.000
##    7.87     59       1    0.967  0.0232        0.922        1.000
##   15.37     58       1    0.950  0.0281        0.896        1.000
##   18.80     57       1    0.933  0.0322        0.872        0.999
##   19.17     56       1    0.917  0.0357        0.849        0.989
##   21.70     55       1    0.900  0.0387        0.827        0.979
##   23.33     54       1    0.883  0.0414        0.806        0.968
##   24.80     53       1    0.867  0.0439        0.785        0.957
##   37.37     52       1    0.850  0.0461        0.764        0.945
##   38.80     51       1    0.833  0.0481        0.744        0.933
##   43.20     50       1    0.817  0.0500        0.724        0.921
##   43.27     49       1    0.800  0.0516        0.705        0.908
##   44.83     48       1    0.783  0.0532        0.686        0.895
##   52.97     47       1    0.767  0.0546        0.667        0.882
##   68.13     45       1    0.750  0.0560        0.648        0.868
##   75.33     44       1    0.733  0.0572        0.629        0.854
##   83.37     40       1    0.714  0.0587        0.608        0.839
##   94.93     38       1    0.695  0.0601        0.587        0.824
##  128.70     31       1    0.673  0.0622        0.562        0.807
##  159.23     25       1    0.646  0.0653        0.530        0.788
##  204.20      9       1    0.574  0.0891        0.424        0.779
##  227.83      6       1    0.479  0.1147        0.299        0.766
```

```
summary(KM_trt_fit2_clinical,60)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ CHEMOTHERAPY,
##     data = TNBC_All)
##
##                 CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      75.0000      50.0000       0.6095       0.0432       0.5305
```

```
## upper 95% CI
##       0.7003
##
##                  CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      45.0000      14.0000       0.7667       0.0546       0.6668
## upper 95% CI
##       0.8815

autoplot(KM_trt_fit2_clinical,
         main = "Survival graph for all 190 patients for chemotherapy treatment(K-M estimate)",
         xlab = "months to death", ylab = "percentage of patients", facets = TRUE)
```



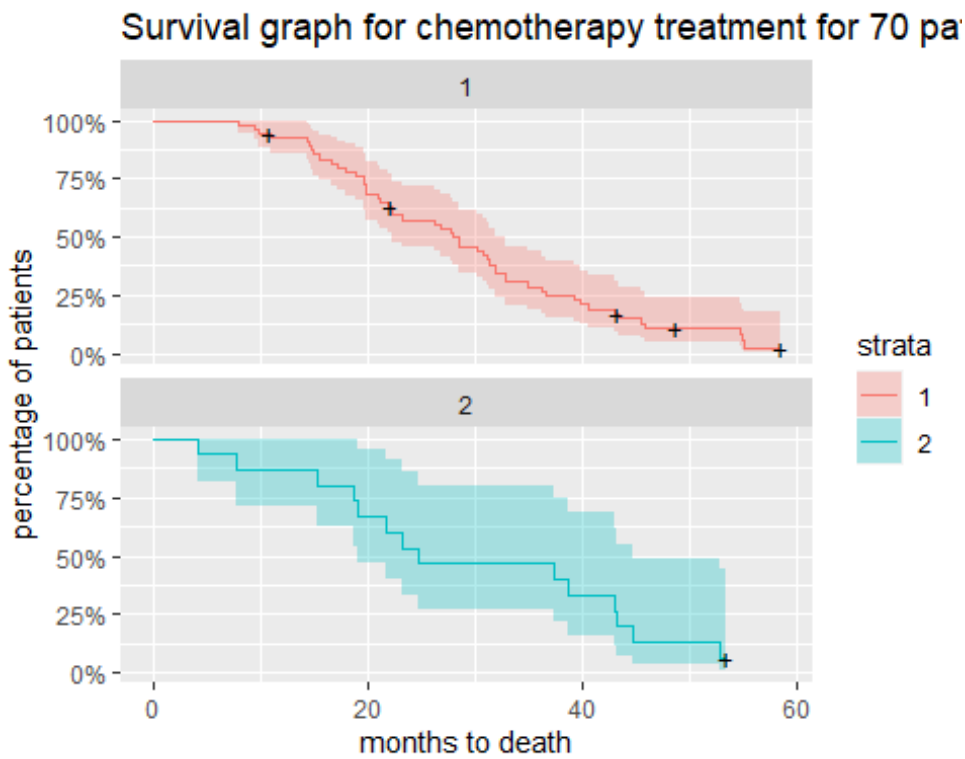

```
KM_trt_fit2_clinical_5 = survfit(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~CHEMOTHERAPY, data = TNBC5)
KM_trt_fit2_clinical_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     CHEMOTHERAPY, data = TNBC5)
##
##                  n events median 0.95LCL 0.95UCL
## CHEMOTHERAPY=1 55     50   28.0    22.4    32.0
## CHEMOTHERAPY=2 15     14   24.8    19.2    44.8

summary(KM_trt_fit2_clinical_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     CHEMOTHERAPY, data = TNBC5)
##
##                  CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   8.07     55       1   0.9818  0.0180      0.94714        1.000
##   9.43     54       1   0.9636  0.0252      0.91541        1.000
##   9.83     53       1   0.9455  0.0306      0.88730        1.000
##  11.07     51       1   0.9269  0.0352      0.86045        0.999
##  14.40     50       1   0.9084  0.0391      0.83495        0.988
##  14.70     49       1   0.8898  0.0424      0.81043        0.977
##  14.80     48       1   0.8713  0.0454      0.78667        0.965
##  15.07     47       1   0.8528  0.0481      0.76353        0.952
##  15.63     46       1   0.8342  0.0505      0.74091        0.939
##  16.60     45       1   0.8157  0.0527      0.71873        0.926
##  17.20     44       1   0.7971  0.0546      0.69695        0.912
```

```
##  17.93     43       1   0.7786  0.0564      0.67552        0.897
##  18.93     42       1   0.7601  0.0580      0.65442        0.883
##  19.73     41       2   0.7230  0.0608      0.61306        0.853
##  19.83     39       1   0.7045  0.0620      0.59277        0.837
##  19.90     38       1   0.6859  0.0631      0.57272        0.821
##  21.00     37       1   0.6674  0.0641      0.55290        0.806
##  21.17     36       1   0.6488  0.0649      0.53329        0.789
##  21.93     35       1   0.6303  0.0657      0.51389        0.773
##  22.40     33       2   0.5921  0.0670      0.47431        0.739
##  23.20     31       1   0.5730  0.0675      0.45484        0.722
##  26.27     30       1   0.5539  0.0679      0.43558        0.704
##  26.87     29       1   0.5348  0.0682      0.41652        0.687
##  27.87     28       1   0.5157  0.0684      0.39766        0.669
##  27.97     27       1   0.4966  0.0685      0.37900        0.651
##  28.50     26       1   0.4775  0.0685      0.36054        0.632
##  28.57     25       1   0.4584  0.0683      0.34227        0.614
##  30.37     24       1   0.4393  0.0681      0.32420        0.595
##  30.93     23       1   0.4202  0.0678      0.30633        0.576
##  31.30     22       1   0.4011  0.0673      0.28866        0.557
##  31.43     21       1   0.3820  0.0668      0.27120        0.538
##  31.93     20       1   0.3629  0.0661      0.25394        0.519
##  32.03     19       1   0.3438  0.0653      0.23690        0.499
##  32.83     18       1   0.3247  0.0644      0.22008        0.479
##  32.93     17       1   0.3056  0.0634      0.20349        0.459
##  35.03     16       1   0.2865  0.0623      0.18714        0.439
##  36.40     15       1   0.2674  0.0610      0.17104        0.418
##  36.63     14       1   0.2483  0.0595      0.15521        0.397
##  39.30     13       1   0.2292  0.0579      0.13966        0.376
##  39.87     12       1   0.2101  0.0562      0.12442        0.355
##  40.63     11       1   0.1910  0.0542      0.10951        0.333
##  43.13     10       1   0.1719  0.0520      0.09497        0.311
##  43.40      8       1   0.1504  0.0498      0.07863        0.288
##  45.63      7       1   0.1289  0.0471      0.06303        0.264
##  45.93      6       1   0.1074  0.0439      0.04827        0.239
##  54.77      4       1   0.0806  0.0403      0.03024        0.215
##  55.00      3       1   0.0537  0.0347      0.01516        0.190
##  55.20      2       1   0.0269  0.0257      0.00411        0.175
##
##                 CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   4.17     15       1   0.9333  0.0644       0.8153        1.000
##   7.87     14       1   0.8667  0.0878       0.7106        1.000
##  15.37     13       1   0.8000  0.1033       0.6212        1.000
##  18.80     12       1   0.7333  0.1142       0.5405        0.995
##  19.17     11       1   0.6667  0.1217       0.4661        0.953
##  21.70     10       1   0.6000  0.1265       0.3969        0.907
##  23.33      9       1   0.5333  0.1288       0.3322        0.856
##  24.80      8       1   0.4667  0.1288       0.2717        0.802
##  37.37      7       1   0.4000  0.1265       0.2152        0.743
##  38.80      6       1   0.3333  0.1217       0.1630        0.682
##  43.20      5       1   0.2667  0.1142       0.1152        0.617
##  43.27      4       1   0.2000  0.1033       0.0727        0.550
##  44.83      3       1   0.1333  0.0878       0.0367        0.484
##  52.97      2       1   0.0667  0.0644       0.0100        0.443
```

```r
summary(KM_trt_fit2_clinical_5,60, extend = TRUE)
```

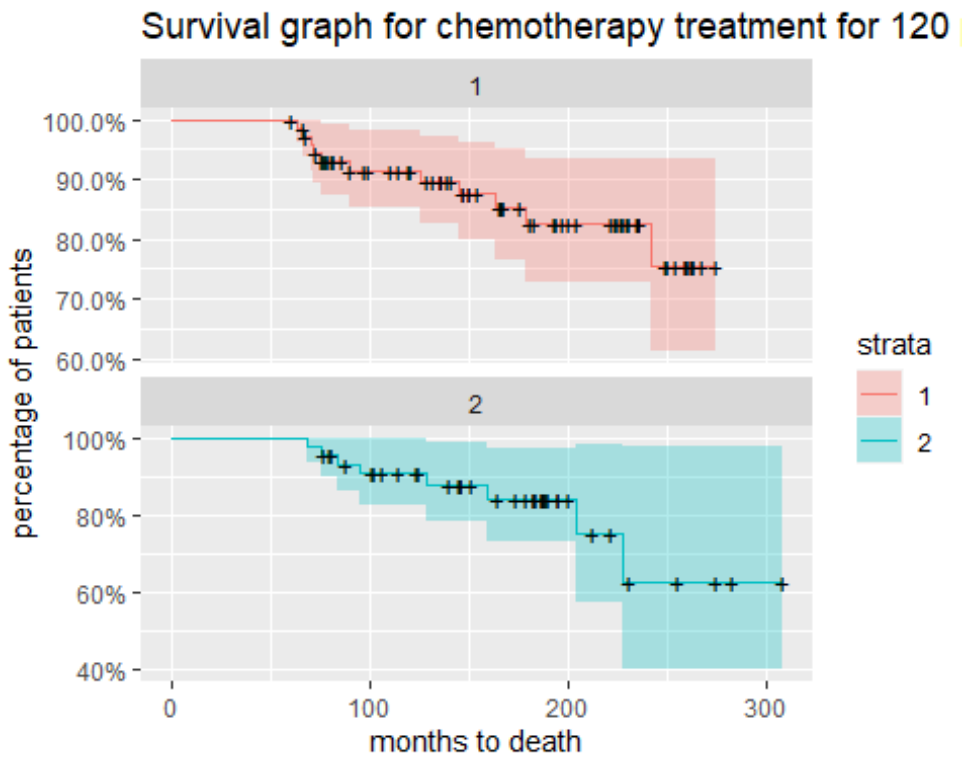
```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     CHEMOTHERAPY, data = TNBC5)
##
##                 CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##     60.00000      0.00000     50.00000      0.02686      0.02572      0.00411
## upper 95% CI
##      0.17541
##
##                 CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000       0.0000      14.0000       0.0667       0.0644       0.0100
## upper 95% CI
##       0.4428
```

```r
autoplot(KM_trt_fit2_clinical_5,
         main = "Survival graph for chemotherapy treatment for 70 patients with not more than 5years follow-up",
         xlab = "months to death", ylab = "percentage of patients", facets = TRUE)
```




```
KM_trt_fit2_clinical_over_5 = survfit(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~CHEMOTHERAPY,
                                      data = TNBC_over5)
KM_trt_fit2_clinical_over_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     CHEMOTHERAPY, data = TNBC_over5)
##
##                  n events median 0.95LCL 0.95UCL
## CHEMOTHERAPY=1 75     11     NA      NA      NA
## CHEMOTHERAPY=2 45      8     NA     228      NA

summary(KM_trt_fit2_clinical_over_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     CHEMOTHERAPY, data = TNBC_over5)
##
##                  CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   62.8     74       1    0.986  0.0134        0.961        1.000
##   66.7     72       1    0.973  0.0190        0.936        1.000
##   70.6     70       1    0.959  0.0232        0.914        1.000
##   71.8     69       1    0.945  0.0267        0.894        0.999
##   75.2     67       1    0.931  0.0298        0.874        0.991
##   89.5     59       1    0.915  0.0332        0.852        0.983
##  125.6     51       1    0.897  0.0371        0.827        0.973
##  144.7     43       1    0.876  0.0417        0.798        0.962
##  162.8     37       1    0.853  0.0468        0.766        0.950
##  178.7     31       1    0.825  0.0528        0.728        0.935
##  241.6     12       1    0.756  0.0817        0.612        0.935
##
##                  CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   68.1     45       1    0.978  0.0220        0.936        1.000
##   75.3     44       1    0.956  0.0307        0.897        1.000
##   83.4     40       1    0.932  0.0381        0.860        1.000
##   94.9     38       1    0.907  0.0443        0.824        0.998
##  128.7     31       1    0.878  0.0516        0.782        0.985
##  159.2     25       1    0.843  0.0603        0.732        0.970
##  204.2      9       1    0.749  0.1033        0.572        0.982
##  227.8      6       1    0.624  0.1428        0.399        0.978
```

```r
summary(KM_trt_fit2_clinical_over_5,60)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     CHEMOTHERAPY, data = TNBC_over5)
##
##                 CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           75            0            1            0            1
## upper 95% CI
##            1
##
##                 CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           45            0            1            0            1
## upper 95% CI
##            1

autoplot(KM_trt_fit2_clinical_over_5,
         main = "Survival graph for chemotherapy treatment for 120 patients with over 5 years follow-up",
         xlab = "months to death", ylab = "percentage of patients", facets = TRUE)
```




```r
# Using the Survdiff function to test survival curve differences
chem = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~CHEMOTHERAPY,rho = 0))
chem

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ CHEMOTHERAPY,
##     rho = 0)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## CHEMOTHERAPY=1 130       61     54.1     0.884      2.55
## CHEMOTHERAPY=2  60       22     28.9     1.654      2.55
##
##  Chisq= 2.6  on 1 degrees of freedom, p= 0.1

chem2 = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~CHEMOTHERAPY,rho = 1))
chem2

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ CHEMOTHERAPY,
##     rho = 1)
```

```
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## CHEMOTHERAPY=1 130     48.4     42.4     0.845      3.08
## CHEMOTHERAPY=2  60     16.4     22.4     1.603      3.08
##
##  Chisq= 3.1  on 1 degrees of freedom, p= 0.08
```

```
chem_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~CHEMOTHERAPY,rho = 0))
chem_5
```

```
## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     CHEMOTHERAPY, rho = 0)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## CHEMOTHERAPY=1 55       50     49.3   0.00909    0.0408
## CHEMOTHERAPY=2 15       14     14.7   0.03058    0.0408
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.8
```

```
chem2_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~CHEMOTHERAPY,rho = 1))
chem2_5
```

```
## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     CHEMOTHERAPY, rho = 1)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## CHEMOTHERAPY=1 55    27.10    26.85   0.00236    0.0161
## CHEMOTHERAPY=2 15     7.43     7.68   0.00825    0.0161
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

```
chem_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~CHEMOTHERAPY,rho = 0))
chem_ov
```

```
## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     CHEMOTHERAPY, rho = 0)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## CHEMOTHERAPY=1 75       11    11.95    0.0757     0.206
## CHEMOTHERAPY=2 45        8     7.05    0.1283     0.206
##
##  Chisq= 0.2  on 1 degrees of freedom, p= 0.6
```

```
chem2_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~CHEMOTHERAPY,rho = 1))
chem2_ov
```

```
## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     CHEMOTHERAPY, rho = 1)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## CHEMOTHERAPY=1 75    10.11    10.85    0.0509      0.15
## CHEMOTHERAPY=2 45     7.21     6.47    0.0855      0.15
##
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.7
```

- 2.16. Create survival objects and plot survival graphs for breast surgery treatments

```
KM_trt_fit3_clinical = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~BREAST_SURGERY, data = TNBC_All)
KM_trt_fit3_clinical
```

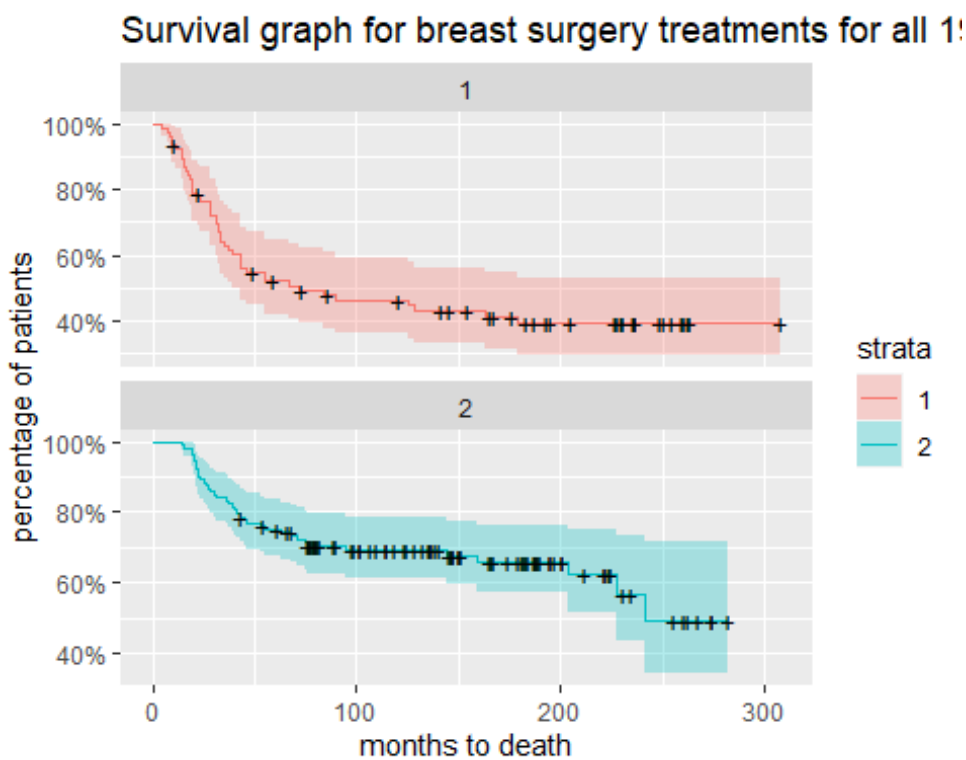
```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY,
##     data = TNBC_All)
##
##                     n events median 0.95LCL 0.95UCL
## BREAST_SURGERY=1  77     44   71.8    43.1      NA
## BREAST_SURGERY=2 113     39  241.6   227.8      NA
```

```
summary(KM_trt_fit3_clinical)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY,
##     data = TNBC_All)
##
##                  BREAST_SURGERY=1
```

```
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    4.17     77       1    0.987  0.0129        0.962        1.000
##    7.87     76       1    0.974  0.0181        0.939        1.000
##    8.07     75       1    0.961  0.0221        0.919        1.000
##    9.43     74       1    0.948  0.0253        0.900        0.999
##    9.83     73       1    0.935  0.0281        0.882        0.992
##   11.07     71       1    0.922  0.0306        0.864        0.984
##   14.40     70       1    0.909  0.0329        0.846        0.976
##   14.70     69       1    0.896  0.0350        0.830        0.967
##   15.37     68       1    0.882  0.0368        0.813        0.958
##   15.63     67       1    0.869  0.0386        0.797        0.948
##   16.60     66       1    0.856  0.0402        0.781        0.939
##   17.20     65       1    0.843  0.0417        0.765        0.929
##   17.93     64       1    0.830  0.0430        0.750        0.918
##   18.93     63       1    0.817  0.0443        0.734        0.908
##   19.17     62       1    0.803  0.0455        0.719        0.898
##   19.73     61       1    0.790  0.0466        0.704        0.887
##   22.40     59       1    0.777  0.0477        0.689        0.876
##   23.20     58       1    0.763  0.0488        0.674        0.865
##   27.87     57       1    0.750  0.0497        0.659        0.854
##   27.97     56       1    0.737  0.0506        0.644        0.843
##   28.50     55       1    0.723  0.0514        0.629        0.831
##   31.30     54       1    0.710  0.0522        0.615        0.820
##   31.43     53       1    0.696  0.0529        0.600        0.808
##   31.93     52       1    0.683  0.0535        0.586        0.796
##   32.03     51       1    0.670  0.0541        0.572        0.785
##   32.83     50       1    0.656  0.0547        0.557        0.773
##   32.93     49       1    0.643  0.0552        0.543        0.761
##   35.03     48       1    0.629  0.0556        0.529        0.749
##   36.63     47       1    0.616  0.0560        0.515        0.736
##   39.30     46       1    0.603  0.0564        0.502        0.724
##   43.13     45       1    0.589  0.0567        0.488        0.712
##   43.20     44       1    0.576  0.0570        0.474        0.699
##   43.40     43       1    0.563  0.0572        0.461        0.687
##   45.93     42       1    0.549  0.0574        0.447        0.674
##   55.00     40       1    0.535  0.0576        0.434        0.661
##   55.20     39       1    0.522  0.0577        0.420        0.648
##   66.73     37       1    0.508  0.0579        0.406        0.635
##   71.80     36       1    0.493  0.0579        0.392        0.621
##   83.37     34       1    0.479  0.0580        0.378        0.607
##   89.53     32       1    0.464  0.0581        0.363        0.593
##  125.60     30       1    0.449  0.0582        0.348        0.578
##  128.70     29       1    0.433  0.0582        0.333        0.564
##  162.83     25       1    0.416  0.0584        0.316        0.547
##  178.73     21       1    0.396  0.0589        0.296        0.530
##
##                  BREAST_SURGERY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   14.8    113       1    0.991 0.00881        0.974        1.000
##   15.1    112       1    0.982 0.01240        0.958        1.000
##   18.8    111       1    0.973 0.01512        0.944        1.000
##   19.7    110       1    0.965 0.01738        0.931        0.999
##   19.8    109       1    0.956 0.01935        0.919        0.994
##   19.9    108       1    0.947 0.02109        0.906        0.989
##   21.0    107       1    0.938 0.02268        0.895        0.984
##   21.2    106       1    0.929 0.02413        0.883        0.978
##   21.7    105       1    0.920 0.02547        0.872        0.972
##   21.9    104       1    0.912 0.02672        0.861        0.965
##   22.4    103       1    0.903 0.02789        0.850        0.959
##   23.3    102       1    0.894 0.02898        0.839        0.952
##   24.8    101       1    0.885 0.03002        0.828        0.946
##   26.3    100       1    0.876 0.03099        0.817        0.939
##   26.9     99       1    0.867 0.03192        0.807        0.932
##   28.6     98       1    0.858 0.03280        0.796        0.925
##   30.4     97       1    0.850 0.03363        0.786        0.918
##   30.9     96       1    0.841 0.03443        0.776        0.911
##   36.4     95       1    0.832 0.03518        0.766        0.904
##   37.4     94       1    0.823 0.03590        0.756        0.896
##   38.8     93       1    0.814 0.03659        0.746        0.889
##   39.9     92       1    0.805 0.03725        0.736        0.882
##   40.6     91       1    0.796 0.03788        0.726        0.874
##   43.3     90       1    0.788 0.03848        0.716        0.867
##   44.8     88       1    0.779 0.03907        0.706        0.859
##   45.6     87       1    0.770 0.03963        0.696        0.851
##   53.0     86       1    0.761 0.04017        0.686        0.844
##   54.8     84       1    0.752 0.04070        0.676        0.836
##   62.8     82       1    0.743 0.04122        0.666        0.828
##   68.1     79       1    0.733 0.04175        0.656        0.820
##   70.6     78       1    0.724 0.04226        0.645        0.812
```

```
##   75.2     77       1    0.714 0.04275        0.635        0.803
##   75.3     76       1    0.705 0.04321        0.625        0.795
##   94.9     64       1    0.694 0.04391        0.613        0.786
##  144.7     45       1    0.679 0.04556        0.595        0.774
##  159.2     37       1    0.660 0.04788        0.573        0.761
##  204.2     17       1    0.621 0.05874        0.516        0.748
##  227.8     11       1    0.565 0.07584        0.434        0.735
##  241.6      8       1    0.494 0.09363        0.341        0.716

summary(KM_trt_fit3_clinical,60)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY,
##     data = TNBC_All)
##
##                 BREAST_SURGERY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      37.0000      36.0000       0.5217       0.0577       0.4200
## upper 95% CI
##       0.6480
##
##                 BREAST_SURGERY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      83.0000      28.0000       0.7517       0.0407       0.6760
## upper 95% CI
##       0.8359

autoplot(KM_trt_fit3_clinical,
         main = "Survival graph for breast surgery treatments for all 190 patients",
         xlab = "months to death", ylab ="percentage of patients", facets = TRUE)
```



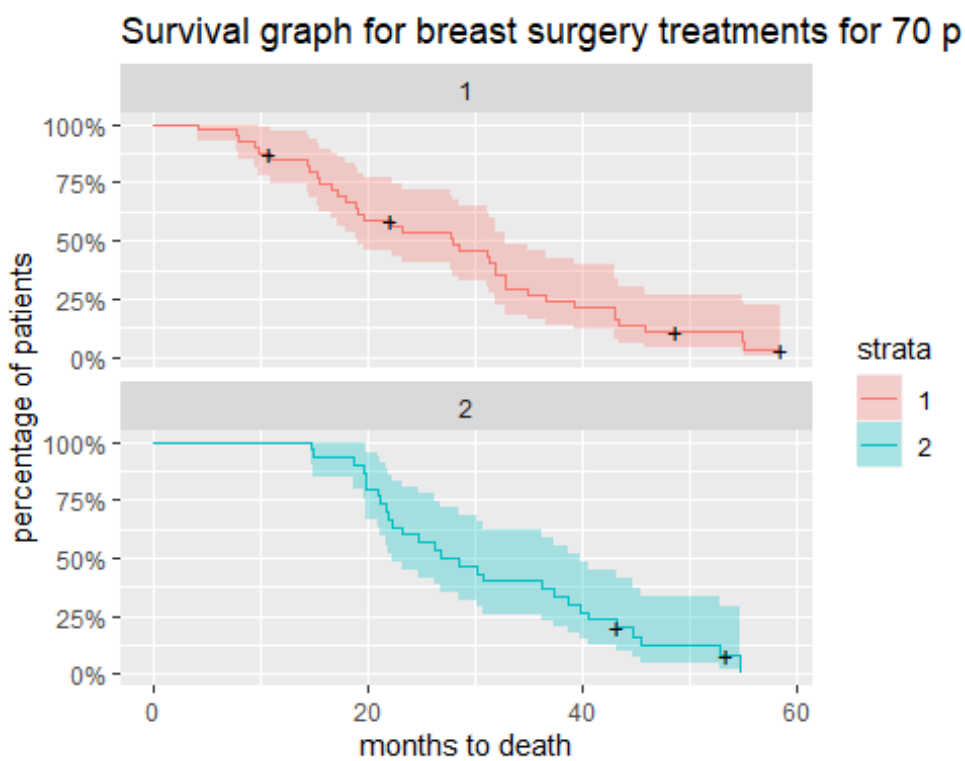

```
KM_trt_fit3_clinical_5 = survfit(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~BREAST_SURGERY, data = TNBC5)
KM_trt_fit3_clinical_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY, data = TNBC5)
##
##                    n events median 0.95LCL 0.95UCL
## BREAST_SURGERY=1 40     36   28.0    19.2    32.9
## BREAST_SURGERY=2 30     28   27.7    22.4    39.9

summary(KM_trt_fit3_clinical_5)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY, data = TNBC5)
##
##                 BREAST_SURGERY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   4.17     40       1   0.9750  0.0247      0.92780        1.000
##   7.87     39       1   0.9500  0.0345      0.88480        1.000
##   8.07     38       1   0.9250  0.0416      0.84687        1.000
##   9.43     37       1   0.9000  0.0474      0.81167        0.998
##   9.83     36       1   0.8750  0.0523      0.77829        0.984
##  11.07     34       1   0.8493  0.0567      0.74504        0.968
##  14.40     33       1   0.8235  0.0606      0.71297        0.951
##  14.70     32       1   0.7978  0.0639      0.68187        0.933
##  15.37     31       1   0.7721  0.0668      0.65158        0.915
##  15.63     30       1   0.7463  0.0694      0.62201        0.895
##  16.60     29       1   0.7206  0.0716      0.59307        0.876
##  17.20     28       1   0.6949  0.0735      0.56471        0.855
##  17.93     27       1   0.6691  0.0752      0.53688        0.834
##  18.93     26       1   0.6434  0.0766      0.50954        0.812
##  19.17     25       1   0.6176  0.0777      0.48268        0.790
##  19.73     24       1   0.5919  0.0786      0.45626        0.768
##  22.40     22       1   0.5650  0.0795      0.42882        0.744
##  23.20     21       1   0.5381  0.0801      0.40187        0.721
##  27.87     20       1   0.5112  0.0805      0.37541        0.696
##  27.97     19       1   0.4843  0.0807      0.34941        0.671
##  28.50     18       1   0.4574  0.0805      0.32389        0.646
##  31.30     17       1   0.4305  0.0802      0.29883        0.620
##  31.43     16       1   0.4036  0.0795      0.27425        0.594
##  31.93     15       1   0.3767  0.0787      0.25015        0.567
##  32.03     14       1   0.3498  0.0775      0.22654        0.540
##  32.83     13       1   0.3229  0.0761      0.20345        0.512
##  32.93     12       1   0.2960  0.0743      0.18089        0.484
##  35.03     11       1   0.2691  0.0723      0.15891        0.456
##  36.63     10       1   0.2421  0.0699      0.13754        0.426
##  39.30      9       1   0.2152  0.0671      0.11683        0.397
##  43.13      8       1   0.1883  0.0639      0.09688        0.366
##  43.20      7       1   0.1614  0.0602      0.07777        0.335
##  43.40      6       1   0.1345  0.0558      0.05965        0.303
##  45.93      5       1   0.1076  0.0507      0.04272        0.271
##  55.00      3       1   0.0717  0.0447      0.02114        0.244
##  55.20      2       1   0.0359  0.0338      0.00565        0.228
##
##                 BREAST_SURGERY=2
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  14.8     30       1    0.967  0.0328       0.9045        1.000
##  15.1     29       1    0.933  0.0455       0.8482        1.000
##  18.8     28       1    0.900  0.0548       0.7988        1.000
##  19.7     27       1    0.867  0.0621       0.7532        0.997
##  19.8     26       1    0.833  0.0680       0.7101        0.978
##  19.9     25       1    0.800  0.0730       0.6689        0.957
##  21.0     24       1    0.767  0.0772       0.6293        0.934
##  21.2     23       1    0.733  0.0807       0.5910        0.910
##  21.7     22       1    0.700  0.0837       0.5538        0.885
##  21.9     21       1    0.667  0.0861       0.5176        0.859
##  22.4     20       1    0.633  0.0880       0.4824        0.832
##  23.3     19       1    0.600  0.0894       0.4480        0.804
##  24.8     18       1    0.567  0.0905       0.4144        0.775
##  26.3     17       1    0.533  0.0911       0.3816        0.745
##  26.9     16       1    0.500  0.0913       0.3496        0.715
##  28.6     15       1    0.467  0.0911       0.3183        0.684
##  30.4     14       1    0.433  0.0905       0.2878        0.652
##  30.9     13       1    0.400  0.0894       0.2581        0.620
##  36.4     12       1    0.367  0.0880       0.2291        0.587
##  37.4     11       1    0.333  0.0861       0.2010        0.553
##  38.8     10       1    0.300  0.0837       0.1737        0.518
##  39.9      9       1    0.267  0.0807       0.1473        0.483
##  40.6      8       1    0.233  0.0772       0.1220        0.446
##  43.3      7       1    0.200  0.0730       0.0978        0.409
##  44.8      5       1    0.160  0.0685       0.0691        0.370
##  45.6      4       1    0.120  0.0620       0.0436        0.330
##  53.0      3       1    0.080  0.0527       0.0220        0.291
##  54.8      1       1    0.000     NaN           NA           NA
```

```
summary(KM_trt_fit3_clinical_5,60, extend = TRUE)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY, data = TNBC5)
##
```

```
##                  BREAST_SURGERY=1
##          time       n.risk      n.event     survival      std.err lower 95% CI
##      60.00000      0.00000     36.00000      0.03587      0.03382      0.00565
## upper 95% CI
##       0.22765
##
##                  BREAST_SURGERY=2
##          time       n.risk      n.event     survival      std.err lower 95% CI
##            60            0           28            0          NaN           NA
## upper 95% CI
##            NA

autoplot(KM_trt_fit3_clinical_5,
         main = "Survival graph for breast surgery treatments for 70 patients with not more than 5 years follow-
up",
         xlab = "months to death", ylab = "percentage of patients", facets = TRUE)
```



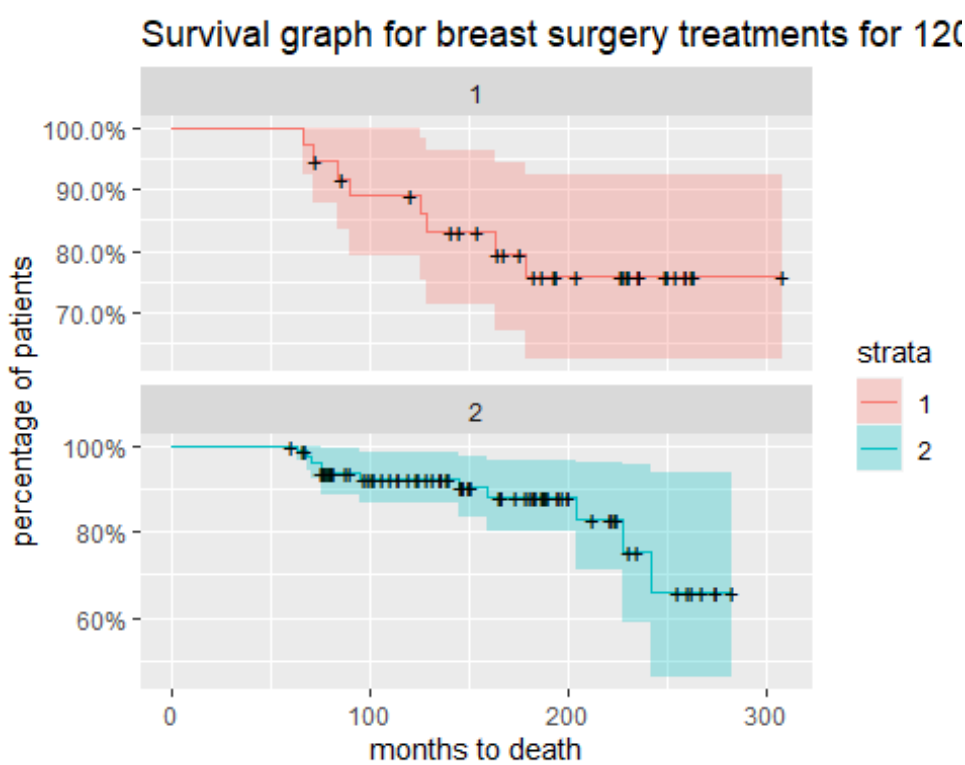

```
KM_trt_fit3_clinical_over_5 = survfit(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~BREAST_SURGERY,
                                      data = TNBC_over5)
KM_trt_fit3_clinical_over_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY, data = TNBC_over5)
##
##                     n events median 0.95LCL 0.95UCL
## BREAST_SURGERY=1 37      8     NA      NA      NA
## BREAST_SURGERY=2 83     11     NA     242      NA

summary(KM_trt_fit3_clinical_over_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY, data = TNBC_over5)
##
##                  BREAST_SURGERY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   66.7     37       1    0.973  0.0267        0.922        1.000
##   71.8     36       1    0.946  0.0372        0.876        1.000
##   83.4     34       1    0.918  0.0453        0.833        1.000
##   89.5     32       1    0.889  0.0522        0.793        0.998
##  125.6     30       1    0.860  0.0583        0.753        0.982
##  128.7     29       1    0.830  0.0634        0.715        0.964
```

```
##  162.8     25       1    0.797  0.0690        0.673        0.944
##  178.7     21       1    0.759  0.0754        0.625        0.922
##
##                 BREAST_SURGERY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   62.8     82       1    0.988  0.0121        0.964        1.000
##   68.1     79       1    0.975  0.0173        0.942        1.000
##   70.6     78       1    0.963  0.0211        0.922        1.000
##   75.2     77       1    0.950  0.0242        0.904        0.999
##   75.3     76       1    0.938  0.0269        0.886        0.992
##   94.9     64       1    0.923  0.0302        0.866        0.984
##  144.7     45       1    0.903  0.0359        0.835        0.976
##  159.2     37       1    0.878  0.0424        0.799        0.965
##  204.2     17       1    0.827  0.0641        0.710        0.962
##  227.8     11       1    0.751  0.0923        0.591        0.956
##  241.6      8       1    0.657  0.1194        0.461        0.938
```

```
summary(KM_trt_fit3_clinical_over_5,60, extend = TRUE)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY, data = TNBC_over5)
##
##                 BREAST_SURGERY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           37            0            1            0            1
## upper 95% CI
##            1
##
##                 BREAST_SURGERY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           83            0            1            0            1
## upper 95% CI
##            1
```

```
autoplot(KM_trt_fit3_clinical_over_5,
         main = "Survival graph for breast surgery treatments for 120 patients with over 5 years follow-up",
         xlab = "months to death", ylab = "percentage of patients", facets = TRUE)
```




```
# Using the Survdiff function to test survival curve differences
surg = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~BREAST_SURGERY,rho = 0))
surg
```

```
## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY,
##     rho = 0)
##
##                     N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1  77       44     29.8      6.75      10.6
## BREAST_SURGERY=2 113       39     53.2      3.78      10.6
##
##  Chisq= 10.6  on 1 degrees of freedom, p= 0.001

surg2 = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~BREAST_SURGERY,rho = 1))
surg2

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY,
##     rho = 1)
##
##                     N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1  77     35.4     23.3      6.28      12.3
## BREAST_SURGERY=2 113     29.4     41.5      3.53      12.3
##
##  Chisq= 12.3  on 1 degrees of freedom, p= 5e-04

surg_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~BREAST_SURGERY,rho = 0))
surg_5

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY, rho = 0)
##
##                    N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1 40       36     34.4    0.0770     0.173
## BREAST_SURGERY=2 30       28     29.6    0.0894     0.173
##
##  Chisq= 0.2  on 1 degrees of freedom, p= 0.7

surg2_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~BREAST_SURGERY,rho = 1))
surg2_5

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY, rho = 1)
##
##                    N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1 40     20.9     18.0     0.452       1.4
## BREAST_SURGERY=2 30     13.7     16.5     0.493       1.4
##
##  Chisq= 1.4  on 1 degrees of freedom, p= 0.2

surg_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~BREAST_SURGERY,rho = 0))
surg_ov

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY, rho = 0)
##
##                    N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1 37        8     7.05    0.1276     0.208
## BREAST_SURGERY=2 83       11    11.95    0.0753     0.208
##
##  Chisq= 0.2  on 1 degrees of freedom, p= 0.6

surg2_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~BREAST_SURGERY,rho = 1))
surg2_ov

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY, rho = 1)
##
##                    N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1 37     7.35     6.35    0.1589     0.281
## BREAST_SURGERY=2 83     9.97    10.97    0.0919     0.281
##
##  Chisq= 0.3  on 1 degrees of freedom, p= 0.6
```

2.1.2. Combination Treatments

- 2.17. Create survival objects and plot survival graphs for radiotherapy and chemotherapy combination treatments

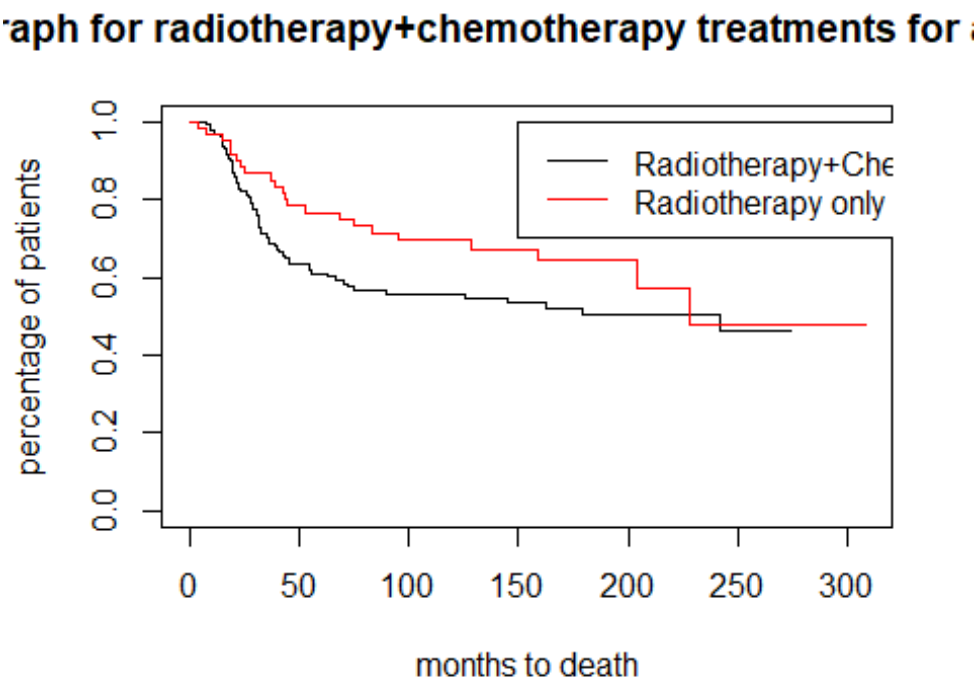
```
KM_trt_fit4_clinical = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~ RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_All
)
KM_trt_fit4_clinical

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY +
##     CHEMOTHERAPY, data = TNBC_All)
##
##                                   n events median 0.95LCL 0.95UCL
## RADIO_THERAPY=1, CHEMOTHERAPY=1 130     61    242    71.8      NA
## RADIO_THERAPY=1, CHEMOTHERAPY=2  60     22    228   204.2      NA

summary(KM_trt_fit4_clinical)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY +
##     CHEMOTHERAPY, data = TNBC_All)
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=1
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    8.07    130       1    0.992 0.00766        0.977        1.000
##    9.43    129       1    0.985 0.01079        0.964        1.000
##    9.83    128       1    0.977 0.01317        0.951        1.000
##   11.07    126       1    0.969 0.01518        0.940        0.999
##   14.40    125       1    0.961 0.01692        0.929        0.995
##   14.70    124       1    0.954 0.01847        0.918        0.991
##   14.80    123       1    0.946 0.01988        0.908        0.986
##   15.07    122       1    0.938 0.02118        0.898        0.981
##   15.63    121       1    0.930 0.02238        0.888        0.975
##   16.60    120       1    0.923 0.02350        0.878        0.970
##   17.20    119       1    0.915 0.02455        0.868        0.964
##   17.93    118       1    0.907 0.02553        0.858        0.959
##   18.93    117       1    0.899 0.02647        0.849        0.953
##   19.73    116       2    0.884 0.02819        0.830        0.941
##   19.83    114       1    0.876 0.02899        0.821        0.935
##   19.90    113       1    0.868 0.02975        0.812        0.929
##   21.00    112       1    0.861 0.03048        0.803        0.922
##   21.17    111       1    0.853 0.03117        0.794        0.916
##   21.93    110       1    0.845 0.03184        0.785        0.910
##   22.40    108       2    0.829 0.03312        0.767        0.897
##   23.20    106       1    0.822 0.03372        0.758        0.890
##   26.27    105       1    0.814 0.03429        0.749        0.884
##   26.87    104       1    0.806 0.03484        0.741        0.877
##   27.87    103       1    0.798 0.03537        0.732        0.871
##   27.97    102       1    0.790 0.03588        0.723        0.864
##   28.50    101       1    0.783 0.03637        0.714        0.857
##   28.57    100       1    0.775 0.03684        0.706        0.850
##   30.37     99       1    0.767 0.03729        0.697        0.844
##   30.93     98       1    0.759 0.03772        0.689        0.837
##   31.30     97       1    0.751 0.03813        0.680        0.830
##   31.43     96       1    0.743 0.03853        0.672        0.823
##   31.93     95       1    0.736 0.03891        0.663        0.816
##   32.03     94       1    0.728 0.03928        0.655        0.809
##   32.83     93       1    0.720 0.03963        0.646        0.802
##   32.93     92       1    0.712 0.03996        0.638        0.795
##   35.03     91       1    0.704 0.04028        0.630        0.788
##   36.40     90       1    0.696 0.04059        0.621        0.781
##   36.63     89       1    0.689 0.04088        0.613        0.774
##   39.30     88       1    0.681 0.04116        0.605        0.766
##   39.87     87       1    0.673 0.04142        0.596        0.759
##   40.63     86       1    0.665 0.04167        0.588        0.752
##   43.13     85       1    0.657 0.04191        0.580        0.745
##   43.40     83       1    0.649 0.04215        0.572        0.737
##   45.63     82       1    0.641 0.04237        0.564        0.730
##   45.93     81       1    0.634 0.04258        0.555        0.723
##   54.77     79       1    0.626 0.04279        0.547        0.715
##   55.00     78       1    0.618 0.04299        0.539        0.708
##   55.20     77       1    0.609 0.04317        0.530        0.700
##   62.77     74       1    0.601 0.04336        0.522        0.693
##   66.73     72       1    0.593 0.04356        0.513        0.685
##   70.60     70       1    0.584 0.04375        0.505        0.677
##   71.80     69       1    0.576 0.04393        0.496        0.669
##   75.23     67       1    0.567 0.04411        0.487        0.661
##   89.53     59       1    0.558 0.04440        0.477        0.652
##  125.60     51       1    0.547 0.04485        0.466        0.642
##  144.67     43       1    0.534 0.04558        0.452        0.631
##  162.83     37       1    0.520 0.04657        0.436        0.619
##  178.73     31       1    0.503 0.04799        0.417        0.606
```

```
##  241.60     12       1    0.461 0.05954        0.358        0.594
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=2
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    4.17     60       1    0.983  0.0165        0.951        1.000
##    7.87     59       1    0.967  0.0232        0.922        1.000
##   15.37     58       1    0.950  0.0281        0.896        1.000
##   18.80     57       1    0.933  0.0322        0.872        0.999
##   19.17     56       1    0.917  0.0357        0.849        0.989
##   21.70     55       1    0.900  0.0387        0.827        0.979
##   23.33     54       1    0.883  0.0414        0.806        0.968
##   24.80     53       1    0.867  0.0439        0.785        0.957
##   37.37     52       1    0.850  0.0461        0.764        0.945
##   38.80     51       1    0.833  0.0481        0.744        0.933
##   43.20     50       1    0.817  0.0500        0.724        0.921
##   43.27     49       1    0.800  0.0516        0.705        0.908
##   44.83     48       1    0.783  0.0532        0.686        0.895
##   52.97     47       1    0.767  0.0546        0.667        0.882
##   68.13     45       1    0.750  0.0560        0.648        0.868
##   75.33     44       1    0.733  0.0572        0.629        0.854
##   83.37     40       1    0.714  0.0587        0.608        0.839
##   94.93     38       1    0.695  0.0601        0.587        0.824
##  128.70     31       1    0.673  0.0622        0.562        0.807
##  159.23     25       1    0.646  0.0653        0.530        0.788
##  204.20      9       1    0.574  0.0891        0.424        0.779
##  227.83      6       1    0.479  0.1147        0.299        0.766
```

```
summary(KM_trt_fit4_clinical,60)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY +
##     CHEMOTHERAPY, data = TNBC_All)
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      75.0000      50.0000       0.6095       0.0432       0.5305
## upper 95% CI
##       0.7003
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      45.0000      14.0000       0.7667       0.0546       0.6668
## upper 95% CI
##       0.8815
```

```
plot(KM_trt_fit4_clinical,col = c(1:4),
     main = "Survival graph for radiotherapy+chemotherapy treatments for all 190 patients",
     xlab = "months to death", ylab="percentage of patients")
      legend (150,1,c("Radiotherapy+Chemotherapy", "Radiotherapy only"),col = (1:4),lwd = 0.5)
```



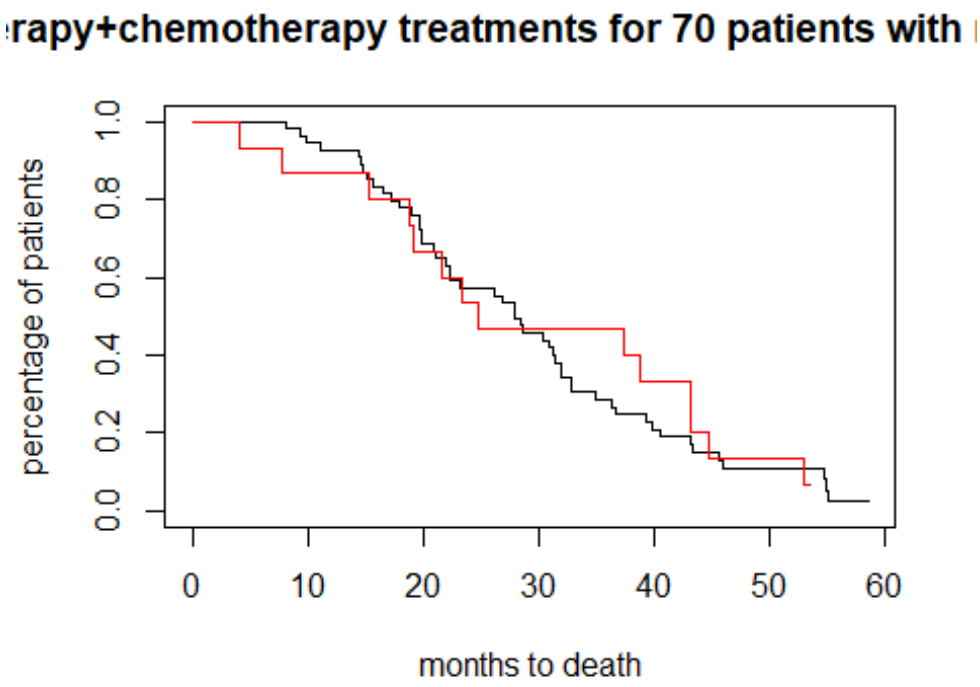

```
KM_trt_fit4_clinical_5 = survfit(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~ RADIO_THERAPY + CHEMOTHERAPY,
                                 data = TNBC5)
KM_trt_fit4_clinical_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
##
##                                   n events median 0.95LCL 0.95UCL
## RADIO_THERAPY=1, CHEMOTHERAPY=1 55     50   28.0    22.4    32.0
## RADIO_THERAPY=1, CHEMOTHERAPY=2 15     14   24.8    19.2    44.8

summary(KM_trt_fit4_clinical_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   8.07     55       1   0.9818  0.0180      0.94714        1.000
##   9.43     54       1   0.9636  0.0252      0.91541        1.000
##   9.83     53       1   0.9455  0.0306      0.88730        1.000
##  11.07     51       1   0.9269  0.0352      0.86045        0.999
##  14.40     50       1   0.9084  0.0391      0.83495        0.988
##  14.70     49       1   0.8898  0.0424      0.81043        0.977
##  14.80     48       1   0.8713  0.0454      0.78667        0.965
##  15.07     47       1   0.8528  0.0481      0.76353        0.952
##  15.63     46       1   0.8342  0.0505      0.74091        0.939
##  16.60     45       1   0.8157  0.0527      0.71873        0.926
##  17.20     44       1   0.7971  0.0546      0.69695        0.912
##  17.93     43       1   0.7786  0.0564      0.67552        0.897
##  18.93     42       1   0.7601  0.0580      0.65442        0.883
##  19.73     41       2   0.7230  0.0608      0.61306        0.853
##  19.83     39       1   0.7045  0.0620      0.59277        0.837
##  19.90     38       1   0.6859  0.0631      0.57272        0.821
##  21.00     37       1   0.6674  0.0641      0.55290        0.806
##  21.17     36       1   0.6488  0.0649      0.53329        0.789
##  21.93     35       1   0.6303  0.0657      0.51389        0.773
##  22.40     33       2   0.5921  0.0670      0.47431        0.739
##  23.20     31       1   0.5730  0.0675      0.45484        0.722
##  26.27     30       1   0.5539  0.0679      0.43558        0.704
##  26.87     29       1   0.5348  0.0682      0.41652        0.687
```

```
##  27.87     28       1   0.5157  0.0684      0.39766        0.669
##  27.97     27       1   0.4966  0.0685      0.37900        0.651
##  28.50     26       1   0.4775  0.0685      0.36054        0.632
##  28.57     25       1   0.4584  0.0683      0.34227        0.614
##  30.37     24       1   0.4393  0.0681      0.32420        0.595
##  30.93     23       1   0.4202  0.0678      0.30633        0.576
##  31.30     22       1   0.4011  0.0673      0.28866        0.557
##  31.43     21       1   0.3820  0.0668      0.27120        0.538
##  31.93     20       1   0.3629  0.0661      0.25394        0.519
##  32.03     19       1   0.3438  0.0653      0.23690        0.499
##  32.83     18       1   0.3247  0.0644      0.22008        0.479
##  32.93     17       1   0.3056  0.0634      0.20349        0.459
##  35.03     16       1   0.2865  0.0623      0.18714        0.439
##  36.40     15       1   0.2674  0.0610      0.17104        0.418
##  36.63     14       1   0.2483  0.0595      0.15521        0.397
##  39.30     13       1   0.2292  0.0579      0.13966        0.376
##  39.87     12       1   0.2101  0.0562      0.12442        0.355
##  40.63     11       1   0.1910  0.0542      0.10951        0.333
##  43.13     10       1   0.1719  0.0520      0.09497        0.311
##  43.40      8       1   0.1504  0.0498      0.07863        0.288
##  45.63      7       1   0.1289  0.0471      0.06303        0.264
##  45.93      6       1   0.1074  0.0439      0.04827        0.239
##  54.77      4       1   0.0806  0.0403      0.03024        0.215
##  55.00      3       1   0.0537  0.0347      0.01516        0.190
##  55.20      2       1   0.0269  0.0257      0.00411        0.175
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   4.17     15       1   0.9333  0.0644       0.8153        1.000
##   7.87     14       1   0.8667  0.0878       0.7106        1.000
##  15.37     13       1   0.8000  0.1033       0.6212        1.000
##  18.80     12       1   0.7333  0.1142       0.5405        0.995
##  19.17     11       1   0.6667  0.1217       0.4661        0.953
##  21.70     10       1   0.6000  0.1265       0.3969        0.907
##  23.33      9       1   0.5333  0.1288       0.3322        0.856
##  24.80      8       1   0.4667  0.1288       0.2717        0.802
##  37.37      7       1   0.4000  0.1265       0.2152        0.743
##  38.80      6       1   0.3333  0.1217       0.1630        0.682
##  43.20      5       1   0.2667  0.1142       0.1152        0.617
##  43.27      4       1   0.2000  0.1033       0.0727        0.550
##  44.83      3       1   0.1333  0.0878       0.0367        0.484
##  52.97      2       1   0.0667  0.0644       0.0100        0.443
```

```
summary(KM_trt_fit4_clinical_5,60, extend = TRUE)
```

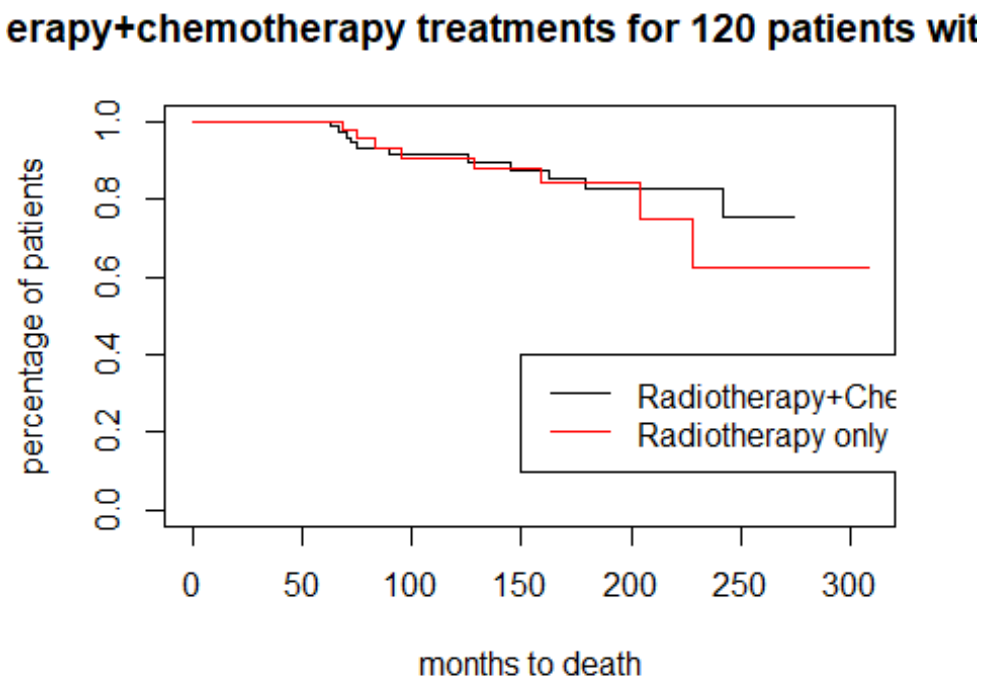
```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##     60.00000      0.00000     50.00000      0.02686      0.02572      0.00411
## upper 95% CI
##      0.17541
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000       0.0000      14.0000       0.0667       0.0644       0.0100
## upper 95% CI
##       0.4428
```

```
plot(KM_trt_fit4_clinical_5,col = c(1:4),
     main = "Survival graph for radiotherapy+chemotherapy treatments for 70 patients with not more than 5years f
ollow-up",
     xlab = "months to death", ylab = "percentage of patients")
     legend (100,1,c("Radiotherapy+Chemotherapy", "Radiotherapy only"),col = (1:4),lwd = 0.5)
```




```
KM_trt_fit4_clinical_over_5 = survfit(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~ RADIO_THERAPY + CHEMOTHE
RAPY,
                                      data = TNBC_over5)
KM_trt_fit4_clinical_over_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over5)
##
##                                    n events median 0.95LCL 0.95UCL
## RADIO_THERAPY=1, CHEMOTHERAPY=1 75     11     NA      NA      NA
## RADIO_THERAPY=1, CHEMOTHERAPY=2 45      8     NA     228      NA

summary(KM_trt_fit4_clinical_over_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over5)
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   62.8     74       1    0.986  0.0134        0.961        1.000
##   66.7     72       1    0.973  0.0190        0.936        1.000
##   70.6     70       1    0.959  0.0232        0.914        1.000
##   71.8     69       1    0.945  0.0267        0.894        0.999
##   75.2     67       1    0.931  0.0298        0.874        0.991
##   89.5     59       1    0.915  0.0332        0.852        0.983
##  125.6     51       1    0.897  0.0371        0.827        0.973
##  144.7     43       1    0.876  0.0417        0.798        0.962
##  162.8     37       1    0.853  0.0468        0.766        0.950
##  178.7     31       1    0.825  0.0528        0.728        0.935
##  241.6     12       1    0.756  0.0817        0.612        0.935
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   68.1     45       1    0.978  0.0220        0.936        1.000
##   75.3     44       1    0.956  0.0307        0.897        1.000
##   83.4     40       1    0.932  0.0381        0.860        1.000
##   94.9     38       1    0.907  0.0443        0.824        0.998
##  128.7     31       1    0.878  0.0516        0.782        0.985
##  159.2     25       1    0.843  0.0603        0.732        0.970
##  204.2      9       1    0.749  0.1033        0.572        0.982
##  227.8      6       1    0.624  0.1428        0.399        0.978
```

```
summary(KM_trt_fit4_clinical_over_5,60)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over5)
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           75            0            1            0            1
## upper 95% CI
##            1
##
##                 RADIO_THERAPY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           45            0            1            0            1
## upper 95% CI
##            1

plot(KM_trt_fit4_clinical_over_5,col = c(1:4),
     main = "Survival graph for radiotherapy+chemotherapy treatments for 120 patients with more than 5 years fol
low-up",
     xlab = "months to death", ylab = "percentage of patients")
      legend (150,0.4,c("Radiotherapy+Chemotherapy", "Radiotherapy only"),col = (1:4),lwd = 0.5)
```




```
acomb = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~RADIO_THERAPY+CHEMOTHERAPY,rho = 0))
acomb

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY +
##     CHEMOTHERAPY, rho = 0)
##
##                                   N Observed Expected (O-E)^2/E (O-E)^2/V
## RADIO_THERAPY=1, CHEMOTHERAPY=1 130       61     54.1     0.884      2.55
## RADIO_THERAPY=1, CHEMOTHERAPY=2  60       22     28.9     1.654      2.55
##
##  Chisq= 2.6  on 1 degrees of freedom, p= 0.1

acomb2 = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~RADIO_THERAPY+CHEMOTHERAPY,rho = 1))
acomb2

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY +
```

```
##     CHEMOTHERAPY, rho = 1)
##
##                                   N Observed Expected (O-E)^2/E (O-E)^2/V
## RADIO_THERAPY=1, CHEMOTHERAPY=1 130     48.4     42.4     0.845      3.08
## RADIO_THERAPY=1, CHEMOTHERAPY=2  60     16.4     22.4     1.603      3.08
##
##  Chisq= 3.1  on 1 degrees of freedom, p= 0.08
```

```
acomb_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~RADIO_THERAPY+CHEMOTHERAPY,rho = 0
))
acomb_5
```

```
## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, rho = 0)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## RADIO_THERAPY=1, CHEMOTHERAPY=1 55       50     49.3   0.00909    0.0408
## RADIO_THERAPY=1, CHEMOTHERAPY=2 15       14     14.7   0.03058    0.0408
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.8
```

```
acomb2_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~RADIO_THERAPY+CHEMOTHERAPY,rho =
1))
acomb2_5
```

```
## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, rho = 1)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## RADIO_THERAPY=1, CHEMOTHERAPY=1 55    27.10    26.85   0.00236    0.0161
## RADIO_THERAPY=1, CHEMOTHERAPY=2 15     7.43     7.68   0.00825    0.0161
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

```
acomb_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~RADIO_THERAPY+CHEMOTHERAPY
                                  ,rho = 0))
acomb_ov
```

```
## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, rho = 0)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## RADIO_THERAPY=1, CHEMOTHERAPY=1 75       11    11.95    0.0757     0.206
## RADIO_THERAPY=1, CHEMOTHERAPY=2 45        8     7.05    0.1283     0.206
##
##  Chisq= 0.2  on 1 degrees of freedom, p= 0.6
```

```
acomb2_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~RADIO_THERAPY+CHEMOTHERAP
Y,
                                   rho = 1))
acomb2_ov
```

```
## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, rho = 1)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## RADIO_THERAPY=1, CHEMOTHERAPY=1 75    10.11    10.85    0.0509      0.15
## RADIO_THERAPY=1, CHEMOTHERAPY=2 45     7.21     6.47    0.0855      0.15
##
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.7
```

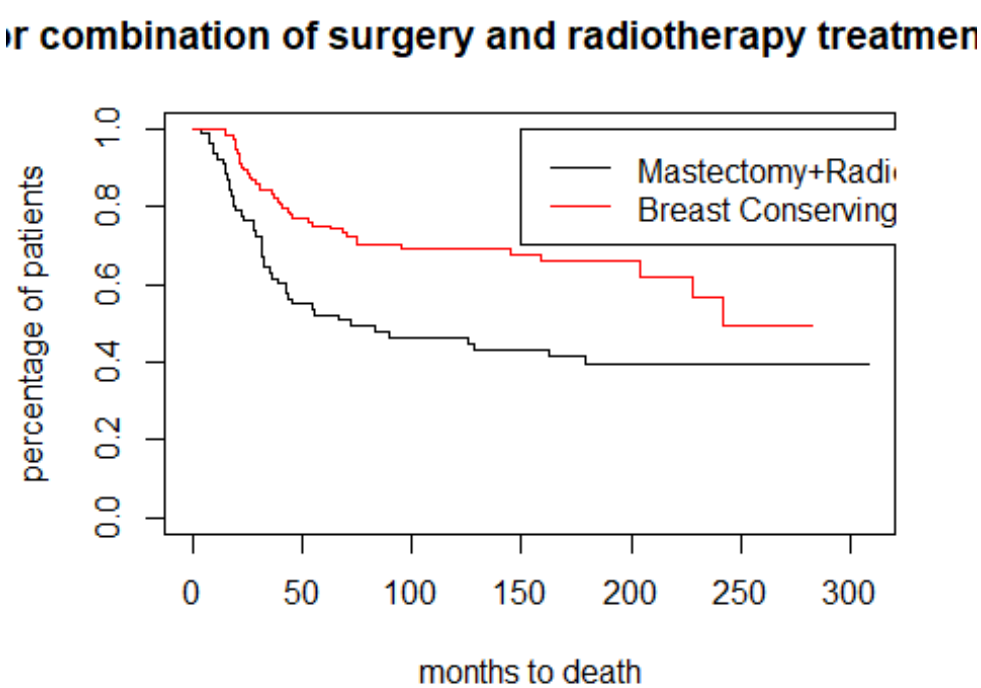
- 2.18. Create survival objects and plot survival graphs for breast surgery and radiotherapy combination treatments

```
KM_trt_fit5_clinical = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~ BREAST_SURGERY + RADIO_THERAPY, data = TNBC_A
ll)
KM_trt_fit5_clinical
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY, data = TNBC_All)
##
##                                    n events median 0.95LCL 0.95UCL
## BREAST_SURGERY=1, RADIO_THERAPY=1  77     44   71.8    43.1      NA
## BREAST_SURGERY=2, RADIO_THERAPY=1 113     39  241.6   227.8      NA
```

```r
summary(KM_trt_fit5_clinical)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY, data = TNBC_All)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    4.17     77       1    0.987  0.0129        0.962        1.000
##    7.87     76       1    0.974  0.0181        0.939        1.000
##    8.07     75       1    0.961  0.0221        0.919        1.000
##    9.43     74       1    0.948  0.0253        0.900        0.999
##    9.83     73       1    0.935  0.0281        0.882        0.992
##   11.07     71       1    0.922  0.0306        0.864        0.984
##   14.40     70       1    0.909  0.0329        0.846        0.976
##   14.70     69       1    0.896  0.0350        0.830        0.967
##   15.37     68       1    0.882  0.0368        0.813        0.958
##   15.63     67       1    0.869  0.0386        0.797        0.948
##   16.60     66       1    0.856  0.0402        0.781        0.939
##   17.20     65       1    0.843  0.0417        0.765        0.929
##   17.93     64       1    0.830  0.0430        0.750        0.918
##   18.93     63       1    0.817  0.0443        0.734        0.908
##   19.17     62       1    0.803  0.0455        0.719        0.898
##   19.73     61       1    0.790  0.0466        0.704        0.887
##   22.40     59       1    0.777  0.0477        0.689        0.876
##   23.20     58       1    0.763  0.0488        0.674        0.865
##   27.87     57       1    0.750  0.0497        0.659        0.854
##   27.97     56       1    0.737  0.0506        0.644        0.843
##   28.50     55       1    0.723  0.0514        0.629        0.831
##   31.30     54       1    0.710  0.0522        0.615        0.820
##   31.43     53       1    0.696  0.0529        0.600        0.808
##   31.93     52       1    0.683  0.0535        0.586        0.796
##   32.03     51       1    0.670  0.0541        0.572        0.785
##   32.83     50       1    0.656  0.0547        0.557        0.773
##   32.93     49       1    0.643  0.0552        0.543        0.761
##   35.03     48       1    0.629  0.0556        0.529        0.749
##   36.63     47       1    0.616  0.0560        0.515        0.736
##   39.30     46       1    0.603  0.0564        0.502        0.724
##   43.13     45       1    0.589  0.0567        0.488        0.712
##   43.20     44       1    0.576  0.0570        0.474        0.699
##   43.40     43       1    0.563  0.0572        0.461        0.687
##   45.93     42       1    0.549  0.0574        0.447        0.674
##   55.00     40       1    0.535  0.0576        0.434        0.661
##   55.20     39       1    0.522  0.0577        0.420        0.648
##   66.73     37       1    0.508  0.0579        0.406        0.635
##   71.80     36       1    0.493  0.0579        0.392        0.621
##   83.37     34       1    0.479  0.0580        0.378        0.607
##   89.53     32       1    0.464  0.0581        0.363        0.593
##  125.60     30       1    0.449  0.0582        0.348        0.578
##  128.70     29       1    0.433  0.0582        0.333        0.564
##  162.83     25       1    0.416  0.0584        0.316        0.547
##  178.73     21       1    0.396  0.0589        0.296        0.530
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   14.8    113       1    0.991 0.00881        0.974        1.000
##   15.1    112       1    0.982 0.01240        0.958        1.000
##   18.8    111       1    0.973 0.01512        0.944        1.000
##   19.7    110       1    0.965 0.01738        0.931        0.999
##   19.8    109       1    0.956 0.01935        0.919        0.994
##   19.9    108       1    0.947 0.02109        0.906        0.989
##   21.0    107       1    0.938 0.02268        0.895        0.984
##   21.2    106       1    0.929 0.02413        0.883        0.978
##   21.7    105       1    0.920 0.02547        0.872        0.972
##   21.9    104       1    0.912 0.02672        0.861        0.965
##   22.4    103       1    0.903 0.02789        0.850        0.959
##   23.3    102       1    0.894 0.02898        0.839        0.952
##   24.8    101       1    0.885 0.03002        0.828        0.946
##   26.3    100       1    0.876 0.03099        0.817        0.939
##   26.9     99       1    0.867 0.03192        0.807        0.932
##   28.6     98       1    0.858 0.03280        0.796        0.925
##   30.4     97       1    0.850 0.03363        0.786        0.918
##   30.9     96       1    0.841 0.03443        0.776        0.911
##   36.4     95       1    0.832 0.03518        0.766        0.904
##   37.4     94       1    0.823 0.03590        0.756        0.896
##   38.8     93       1    0.814 0.03659        0.746        0.889
##   39.9     92       1    0.805 0.03725        0.736        0.882
##   40.6     91       1    0.796 0.03788        0.726        0.874
##   43.3     90       1    0.788 0.03848        0.716        0.867
```

```
##   44.8     88       1    0.779 0.03907        0.706        0.859
##   45.6     87       1    0.770 0.03963        0.696        0.851
##   53.0     86       1    0.761 0.04017        0.686        0.844
##   54.8     84       1    0.752 0.04070        0.676        0.836
##   62.8     82       1    0.743 0.04122        0.666        0.828
##   68.1     79       1    0.733 0.04175        0.656        0.820
##   70.6     78       1    0.724 0.04226        0.645        0.812
##   75.2     77       1    0.714 0.04275        0.635        0.803
##   75.3     76       1    0.705 0.04321        0.625        0.795
##   94.9     64       1    0.694 0.04391        0.613        0.786
##  144.7     45       1    0.679 0.04556        0.595        0.774
##  159.2     37       1    0.660 0.04788        0.573        0.761
##  204.2     17       1    0.621 0.05874        0.516        0.748
##  227.8     11       1    0.565 0.07584        0.434        0.735
##  241.6      8       1    0.494 0.09363        0.341        0.716

summary(KM_trt_fit5_clinical,60)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY, data = TNBC_All)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      37.0000      36.0000       0.5217       0.0577       0.4200
## upper 95% CI
##       0.6480
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      83.0000      28.0000       0.7517       0.0407       0.6760
## upper 95% CI
##       0.8359

plot(KM_trt_fit5_clinical,col = c(1:4),
     main = "Survival graph for combination of surgery and radiotherapy treatments for all 190 patients",
     xlab = "months to death", ylab = "percentage of patients")
     legend (150,1,c("Mastectomy+Radiotherapy", "Breast Conserving+Radiotherapy"),col = (1:4),lwd = 0.5)
```



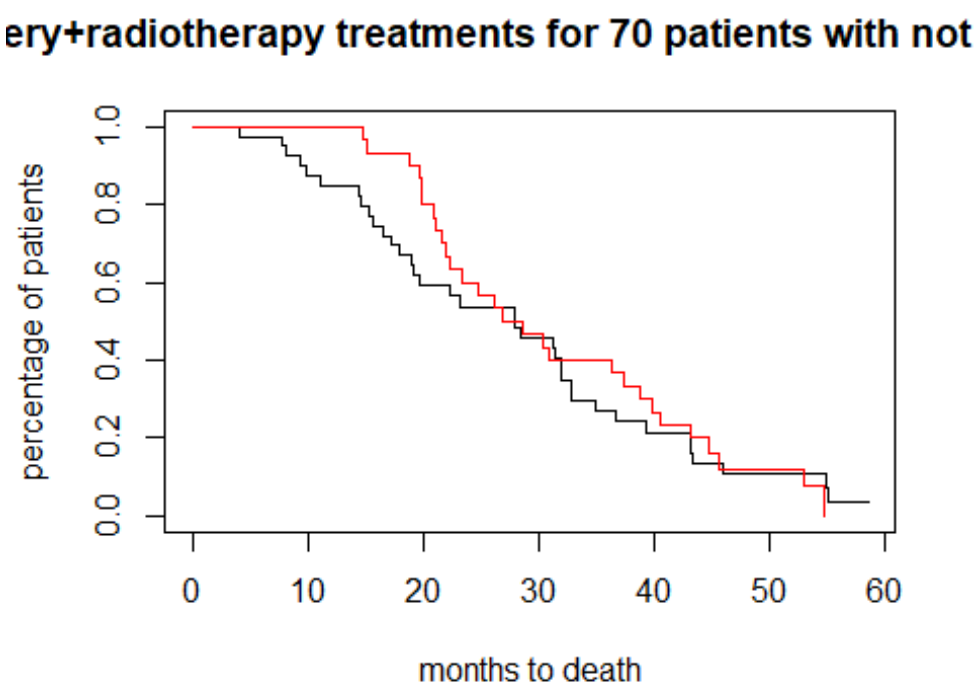

```
KM_trt_fit5_clinical_5 = survfit(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~ BREAST_SURGERY + RADIO_THERAPY,
                                 data = TNBC5)
KM_trt_fit5_clinical_5
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY, data = TNBC5)
##
##                                    n events median 0.95LCL 0.95UCL
## BREAST_SURGERY=1, RADIO_THERAPY=1 40     36   28.0    19.2    32.9
## BREAST_SURGERY=2, RADIO_THERAPY=1 30     28   27.7    22.4    39.9
```

```
summary(KM_trt_fit5_clinical_5)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY, data = TNBC5)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   4.17     40       1   0.9750  0.0247      0.92780        1.000
##   7.87     39       1   0.9500  0.0345      0.88480        1.000
##   8.07     38       1   0.9250  0.0416      0.84687        1.000
##   9.43     37       1   0.9000  0.0474      0.81167        0.998
##   9.83     36       1   0.8750  0.0523      0.77829        0.984
##  11.07     34       1   0.8493  0.0567      0.74504        0.968
##  14.40     33       1   0.8235  0.0606      0.71297        0.951
##  14.70     32       1   0.7978  0.0639      0.68187        0.933
##  15.37     31       1   0.7721  0.0668      0.65158        0.915
##  15.63     30       1   0.7463  0.0694      0.62201        0.895
##  16.60     29       1   0.7206  0.0716      0.59307        0.876
##  17.20     28       1   0.6949  0.0735      0.56471        0.855
##  17.93     27       1   0.6691  0.0752      0.53688        0.834
##  18.93     26       1   0.6434  0.0766      0.50954        0.812
##  19.17     25       1   0.6176  0.0777      0.48268        0.790
##  19.73     24       1   0.5919  0.0786      0.45626        0.768
##  22.40     22       1   0.5650  0.0795      0.42882        0.744
##  23.20     21       1   0.5381  0.0801      0.40187        0.721
##  27.87     20       1   0.5112  0.0805      0.37541        0.696
##  27.97     19       1   0.4843  0.0807      0.34941        0.671
##  28.50     18       1   0.4574  0.0805      0.32389        0.646
##  31.30     17       1   0.4305  0.0802      0.29883        0.620
##  31.43     16       1   0.4036  0.0795      0.27425        0.594
##  31.93     15       1   0.3767  0.0787      0.25015        0.567
##  32.03     14       1   0.3498  0.0775      0.22654        0.540
##  32.83     13       1   0.3229  0.0761      0.20345        0.512
##  32.93     12       1   0.2960  0.0743      0.18089        0.484
##  35.03     11       1   0.2691  0.0723      0.15891        0.456
##  36.63     10       1   0.2421  0.0699      0.13754        0.426
##  39.30      9       1   0.2152  0.0671      0.11683        0.397
##  43.13      8       1   0.1883  0.0639      0.09688        0.366
##  43.20      7       1   0.1614  0.0602      0.07777        0.335
##  43.40      6       1   0.1345  0.0558      0.05965        0.303
##  45.93      5       1   0.1076  0.0507      0.04272        0.271
##  55.00      3       1   0.0717  0.0447      0.02114        0.244
##  55.20      2       1   0.0359  0.0338      0.00565        0.228
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  14.8     30       1    0.967  0.0328       0.9045        1.000
##  15.1     29       1    0.933  0.0455       0.8482        1.000
##  18.8     28       1    0.900  0.0548       0.7988        1.000
##  19.7     27       1    0.867  0.0621       0.7532        0.997
##  19.8     26       1    0.833  0.0680       0.7101        0.978
##  19.9     25       1    0.800  0.0730       0.6689        0.957
##  21.0     24       1    0.767  0.0772       0.6293        0.934
##  21.2     23       1    0.733  0.0807       0.5910        0.910
##  21.7     22       1    0.700  0.0837       0.5538        0.885
##  21.9     21       1    0.667  0.0861       0.5176        0.859
##  22.4     20       1    0.633  0.0880       0.4824        0.832
##  23.3     19       1    0.600  0.0894       0.4480        0.804
##  24.8     18       1    0.567  0.0905       0.4144        0.775
##  26.3     17       1    0.533  0.0911       0.3816        0.745
##  26.9     16       1    0.500  0.0913       0.3496        0.715
##  28.6     15       1    0.467  0.0911       0.3183        0.684
##  30.4     14       1    0.433  0.0905       0.2878        0.652
##  30.9     13       1    0.400  0.0894       0.2581        0.620
##  36.4     12       1    0.367  0.0880       0.2291        0.587
##  37.4     11       1    0.333  0.0861       0.2010        0.553
##  38.8     10       1    0.300  0.0837       0.1737        0.518
##  39.9      9       1    0.267  0.0807       0.1473        0.483
##  40.6      8       1    0.233  0.0772       0.1220        0.446
##  43.3      7       1    0.200  0.0730       0.0978        0.409
##  44.8      5       1    0.160  0.0685       0.0691        0.370
```

```
##  45.6      4       1    0.120  0.0620       0.0436        0.330
##  53.0      3       1    0.080  0.0527       0.0220        0.291
##  54.8      1       1    0.000     NaN           NA           NA

summary(KM_trt_fit5_clinical_5,60, extend = TRUE)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY, data = TNBC5)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##     60.00000      0.00000     36.00000      0.03587      0.03382      0.00565
## upper 95% CI
##      0.22765
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60            0           28            0          NaN           NA
## upper 95% CI
##           NA

plot(KM_trt_fit5_clinical_5,col = c(1:4),
     main = "Survival graph for surgery+radiotherapy treatments for 70 patients with not more than 5years follow
-up",
     xlab = "months to death", ylab = "percentage of patients")
     legend (100,1,c("Mastectomy+Radiotherapy", "Breast Conserving+Radiotherapy"),col = (1:4),lwd = 0.5)
```



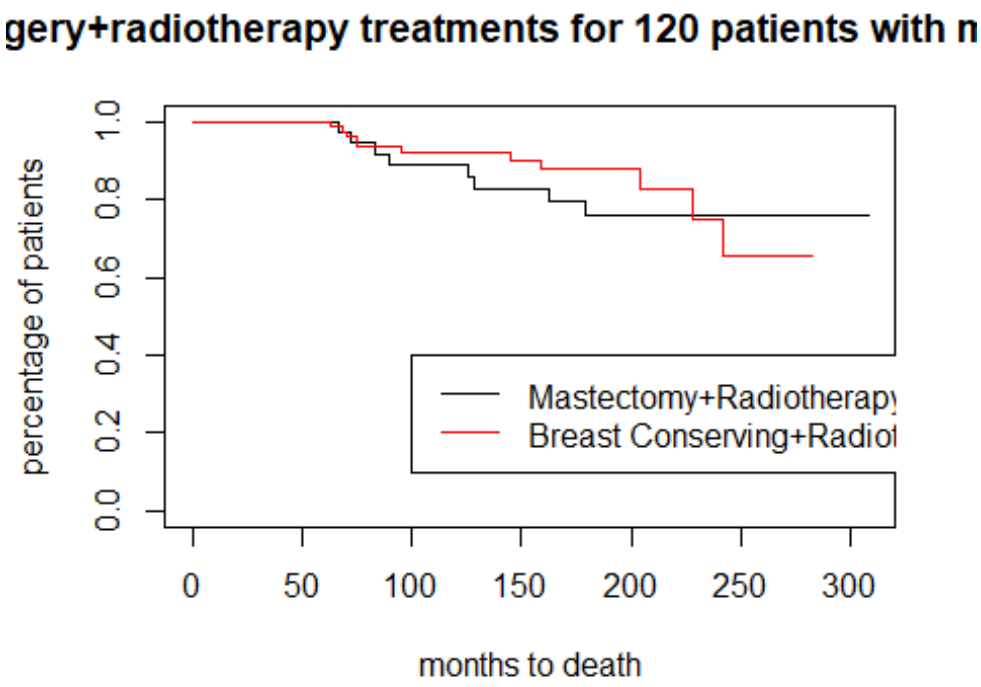

```
KM_trt_fit5_clinical_over_5 = survfit(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~ BREAST_SURGERY + RADIO_T
HERAPY,                                  data = TNBC_over5)
KM_trt_fit5_clinical_over_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY, data = TNBC_over5)
##
##                                   n events median 0.95LCL 0.95UCL
## BREAST_SURGERY=1, RADIO_THERAPY=1 37      8     NA      NA      NA
## BREAST_SURGERY=2, RADIO_THERAPY=1 83     11     NA     242      NA

summary(KM_trt_fit5_clinical_over_5)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY, data = TNBC_over5)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   66.7     37       1    0.973  0.0267        0.922        1.000
##   71.8     36       1    0.946  0.0372        0.876        1.000
##   83.4     34       1    0.918  0.0453        0.833        1.000
##   89.5     32       1    0.889  0.0522        0.793        0.998
##  125.6     30       1    0.860  0.0583        0.753        0.982
##  128.7     29       1    0.830  0.0634        0.715        0.964
##  162.8     25       1    0.797  0.0690        0.673        0.944
##  178.7     21       1    0.759  0.0754        0.625        0.922
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   62.8     82       1    0.988  0.0121        0.964        1.000
##   68.1     79       1    0.975  0.0173        0.942        1.000
##   70.6     78       1    0.963  0.0211        0.922        1.000
##   75.2     77       1    0.950  0.0242        0.904        0.999
##   75.3     76       1    0.938  0.0269        0.886        0.992
##   94.9     64       1    0.923  0.0302        0.866        0.984
##  144.7     45       1    0.903  0.0359        0.835        0.976
##  159.2     37       1    0.878  0.0424        0.799        0.965
##  204.2     17       1    0.827  0.0641        0.710        0.962
##  227.8     11       1    0.751  0.0923        0.591        0.956
##  241.6      8       1    0.657  0.1194        0.461        0.938
```

```
summary(KM_trt_fit5_clinical_over_5,60)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY, data = TNBC_over5)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           37            0            1            0            1
## upper 95% CI
##            1
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           83            0            1            0            1
## upper 95% CI
##            1
```

```
plot(KM_trt_fit5_clinical_over_5,col = c(1:4),
     main = "Survival graph for surgery+radiotherapy treatments for 120 patients with more than 5years follow-up
",
     xlab = "months to death", ylab = "percentage of patients")
     legend (100,0.4,c("Mastectomy+Radiotherapy", "Breast Conserving+Radiotherapy"),col = (1:4),lwd = 0.5)
```




```
bcomb = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~BREAST_SURGERY + RADIO_THERAPY,rho = 0))
bcomb

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY, rho = 0)
##
##                                    N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1  77       44     29.8      6.75      10.6
## BREAST_SURGERY=2, RADIO_THERAPY=1 113       39     53.2      3.78      10.6
##
##  Chisq= 10.6  on 1 degrees of freedom, p= 0.001

bcomb2 = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~BREAST_SURGERY + RADIO_THERAPY,rho = 1))
bcomb2

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY, rho = 1)
##
##                                    N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1  77     35.4     23.3      6.28      12.3
## BREAST_SURGERY=2, RADIO_THERAPY=1 113     29.4     41.5      3.53      12.3
##
##  Chisq= 12.3  on 1 degrees of freedom, p= 5e-04

bcomb_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~BREAST_SURGERY + RADIO_THERAPY,
                               rho = 0))
bcomb_5

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY, rho = 0)
##
##                                   N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1 40       36     34.4    0.0770     0.173
## BREAST_SURGERY=2, RADIO_THERAPY=1 30       28     29.6    0.0894     0.173
##
##  Chisq= 0.2  on 1 degrees of freedom, p= 0.7
```

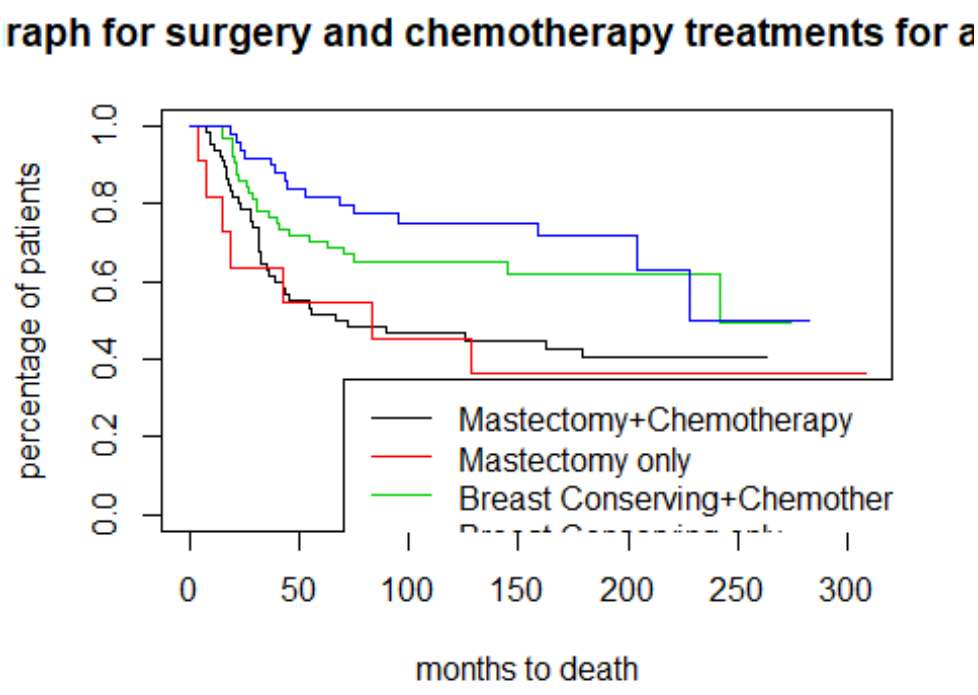
```
bcomb2_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~BREAST_SURGERY + RADIO_THERAPY,
                                rho = 1))
bcomb2_5

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY, rho = 1)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1 40     20.9     18.0     0.452       1.4
## BREAST_SURGERY=2, RADIO_THERAPY=1 30     13.7     16.5     0.493       1.4
##
##  Chisq= 1.4  on 1 degrees of freedom, p= 0.2

bcomb_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~
                                     BREAST_SURGERY + RADIO_THERAPY, rho = 0))
bcomb_ov

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY, rho = 0)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1 37        8     7.05    0.1276     0.208
## BREAST_SURGERY=2, RADIO_THERAPY=1 83       11    11.95    0.0753     0.208
##
##  Chisq= 0.2  on 1 degrees of freedom, p= 0.6

bcomb2_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~
                                      BREAST_SURGERY + RADIO_THERAPY, rho = 1))
bcomb2_ov

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY, rho = 1)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1 37     7.35     6.35    0.1589     0.281
## BREAST_SURGERY=2, RADIO_THERAPY=1 83     9.97    10.97    0.0919     0.281
##
##  Chisq= 0.3  on 1 degrees of freedom, p= 0.6
```

- 2.19. Create survival objects and plot survival graphs for breast surgery and chemotherapy combination treatments

```
KM_trt_fit6_clinical = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~ BREAST_SURGERY + CHEMOTHERAPY, data = TNBC_Al
l)
KM_trt_fit6_clinical

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     CHEMOTHERAPY, data = TNBC_All)
##
##                                 n events median 0.95LCL 0.95UCL
## BREAST_SURGERY=1, CHEMOTHERAPY=1 66     37   71.8    39.3      NA
## BREAST_SURGERY=1, CHEMOTHERAPY=2 11      7   83.4    19.2      NA
## BREAST_SURGERY=2, CHEMOTHERAPY=1 64     24  241.6   241.6      NA
## BREAST_SURGERY=2, CHEMOTHERAPY=2 49     15     NA   204.2      NA

summary(KM_trt_fit6_clinical)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     CHEMOTHERAPY, data = TNBC_All)
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=1
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    8.07     66       1    0.985  0.0150        0.956        1.000
##    9.43     65       1    0.970  0.0211        0.929        1.000
##    9.83     64       1    0.955  0.0256        0.906        1.000
##   11.07     62       1    0.939  0.0295        0.883        0.999
##   14.40     61       1    0.924  0.0328        0.862        0.990
##   14.70     60       1    0.908  0.0357        0.841        0.981
##   15.63     59       1    0.893  0.0382        0.821        0.971
##   16.60     58       1    0.878  0.0406        0.802        0.961
##   17.20     57       1    0.862  0.0427        0.782        0.950
##   17.93     56       1    0.847  0.0446        0.764        0.939
##   18.93     55       1    0.831  0.0464        0.745        0.927
##   19.73     54       1    0.816  0.0480        0.727        0.916
##   22.40     52       1    0.800  0.0496        0.709        0.904
```

```
##   23.20     51       1    0.785  0.0510        0.691        0.891
##   27.87     50       1    0.769  0.0524        0.673        0.879
##   27.97     49       1    0.753  0.0536        0.655        0.866
##   28.50     48       1    0.738  0.0547        0.638        0.853
##   31.30     47       1    0.722  0.0558        0.620        0.840
##   31.43     46       1    0.706  0.0567        0.603        0.827
##   31.93     45       1    0.690  0.0576        0.586        0.813
##   32.03     44       1    0.675  0.0584        0.570        0.799
##   32.83     43       1    0.659  0.0591        0.553        0.786
##   32.93     42       1    0.643  0.0597        0.536        0.772
##   35.03     41       1    0.628  0.0603        0.520        0.758
##   36.63     40       1    0.612  0.0608        0.504        0.744
##   39.30     39       1    0.596  0.0612        0.488        0.729
##   43.13     38       1    0.581  0.0616        0.472        0.715
##   43.40     37       1    0.565  0.0619        0.456        0.700
##   45.93     36       1    0.549  0.0621        0.440        0.686
##   55.00     34       1    0.533  0.0624        0.424        0.670
##   55.20     33       1    0.517  0.0625        0.408        0.655
##   66.73     31       1    0.500  0.0627        0.391        0.640
##   71.80     30       1    0.484  0.0628        0.375        0.624
##   89.53     27       1    0.466  0.0630        0.357        0.607
##  125.60     25       1    0.447  0.0631        0.339        0.590
##  162.83     22       1    0.427  0.0635        0.319        0.571
##  178.73     19       1    0.404  0.0640        0.296        0.551
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=2
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    4.17     11       1    0.909  0.0867        0.754        1.000
##    7.87     10       1    0.818  0.1163        0.619        1.000
##   15.37      9       1    0.727  0.1343        0.506        1.000
##   19.17      8       1    0.636  0.1450        0.407        0.995
##   43.20      7       1    0.545  0.1501        0.318        0.936
##   83.37      6       1    0.455  0.1501        0.238        0.868
##  128.70      5       1    0.364  0.1450        0.166        0.795
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   14.8     64       1    0.984  0.0155        0.954        1.000
##   15.1     63       1    0.969  0.0217        0.927        1.000
##   19.7     62       1    0.953  0.0264        0.903        1.000
##   19.8     61       1    0.938  0.0303        0.880        0.999
##   19.9     60       1    0.922  0.0335        0.858        0.990
##   21.0     59       1    0.906  0.0364        0.838        0.981
##   21.2     58       1    0.891  0.0390        0.817        0.970
##   21.9     57       1    0.875  0.0413        0.798        0.960
##   22.4     56       1    0.859  0.0435        0.778        0.949
##   26.3     55       1    0.844  0.0454        0.759        0.938
##   26.9     54       1    0.828  0.0472        0.741        0.926
##   28.6     53       1    0.812  0.0488        0.722        0.914
##   30.4     52       1    0.797  0.0503        0.704        0.902
##   30.9     51       1    0.781  0.0517        0.686        0.889
##   36.4     50       1    0.766  0.0530        0.669        0.877
##   39.9     49       1    0.750  0.0541        0.651        0.864
##   40.6     48       1    0.734  0.0552        0.634        0.851
##   45.6     46       1    0.718  0.0563        0.616        0.838
##   54.8     45       1    0.702  0.0572        0.599        0.824
##   62.8     43       1    0.686  0.0582        0.581        0.810
##   70.6     40       1    0.669  0.0592        0.562        0.796
##   75.2     39       1    0.652  0.0601        0.544        0.781
##  144.7     20       1    0.619  0.0654        0.503        0.762
##  241.6      5       1    0.495  0.1225        0.305        0.804
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   18.8     49       1    0.980  0.0202        0.941        1.000
##   21.7     48       1    0.959  0.0283        0.905        1.000
##   23.3     47       1    0.939  0.0342        0.874        1.000
##   24.8     46       1    0.918  0.0391        0.845        0.998
##   37.4     45       1    0.898  0.0432        0.817        0.987
##   38.8     44       1    0.878  0.0468        0.790        0.974
##   43.3     43       1    0.857  0.0500        0.765        0.961
##   44.8     42       1    0.837  0.0528        0.739        0.947
##   53.0     41       1    0.816  0.0553        0.715        0.932
##   68.1     39       1    0.795  0.0577        0.690        0.917
##   75.3     38       1    0.774  0.0599        0.666        0.901
##   94.9     33       1    0.751  0.0625        0.638        0.884
##  159.2     22       1    0.717  0.0683        0.595        0.864
##  204.2      8       1    0.627  0.1030        0.455        0.865
##  227.8      5       1    0.502  0.1392        0.291        0.864
```

```r
summary(KM_trt_fit6_clinical,60)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     CHEMOTHERAPY, data = TNBC_All)
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      31.0000      31.0000       0.5169       0.0625       0.4078
## upper 95% CI
##       0.6552
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##       60.000        6.000        5.000        0.545        0.150        0.318
## upper 95% CI
##        0.936
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      44.0000      19.0000       0.7024       0.0572       0.5988
## upper 95% CI
##       0.8241
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      39.0000       9.0000       0.8163       0.0553       0.7148
## upper 95% CI
##       0.9323

plot(KM_trt_fit6_clinical,col = c(1:4),
     main = "Survival graph for surgery and chemotherapy treatments for all 190 patients",
     xlab = "months to death", ylab = "percentage of patients")
     legend(70,0.35,c("Mastectomy+Chemotherapy","Mastectomy only","Breast Conserving+Chemotherapy",
               "Breast Conserving only"), col = (1:4),lwd = 0.5)
```



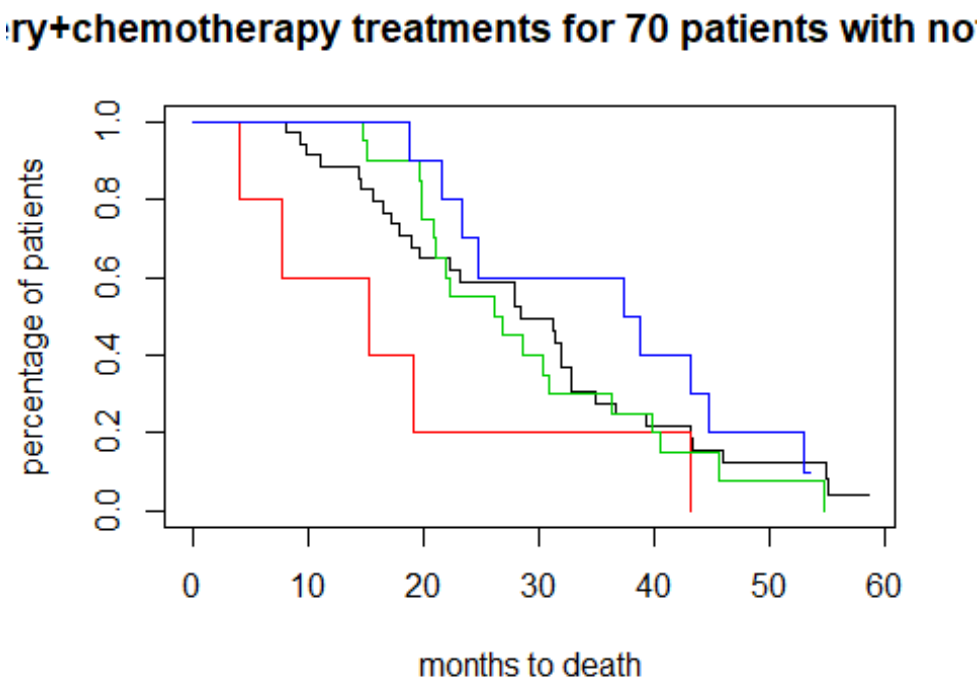

```r
KM_trt_fit6_clinical_5 = survfit(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~ BREAST_SURGERY + CHEMOTHERAPY,
                                 data = TNBC5)
KM_trt_fit6_clinical_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, data = TNBC5)
```

```
##
##                                  n events median 0.95LCL 0.95UCL
## BREAST_SURGERY=1, CHEMOTHERAPY=1 35     31   28.5   22.40    35.0
## BREAST_SURGERY=1, CHEMOTHERAPY=2  5      5   15.4    7.87      NA
## BREAST_SURGERY=2, CHEMOTHERAPY=1 20     19   26.6   21.17    39.9
## BREAST_SURGERY=2, CHEMOTHERAPY=2 10      9   38.1   23.33      NA
```

```
summary(KM_trt_fit6_clinical_5)
```

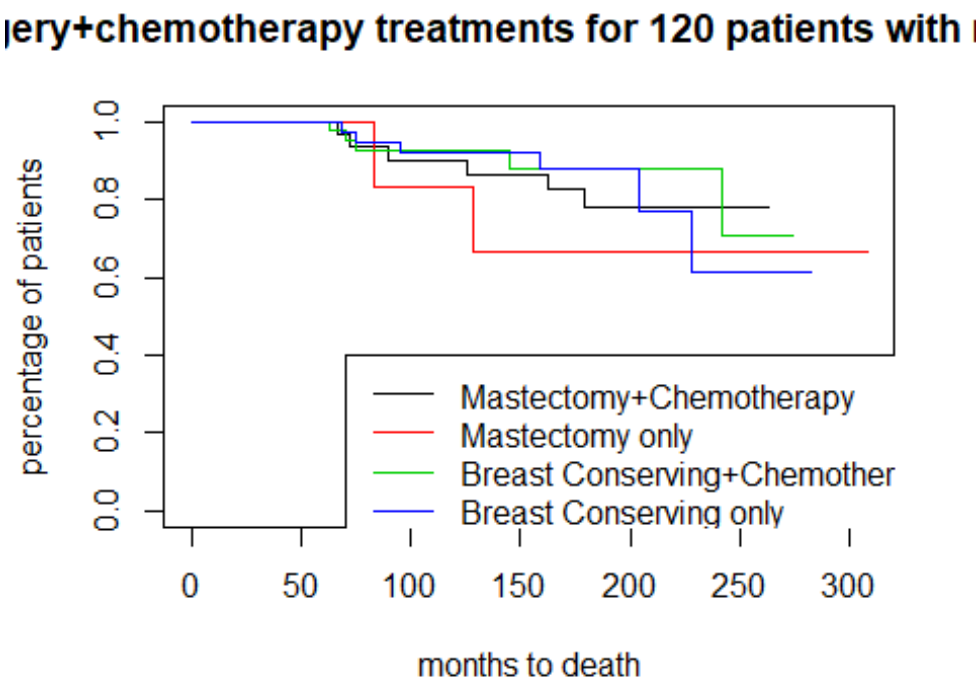
```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, data = TNBC5)
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   8.07     35       1   0.9714  0.0282      0.91777        1.000
##   9.43     34       1   0.9429  0.0392      0.86901        1.000
##   9.83     33       1   0.9143  0.0473      0.82609        1.000
##  11.07     31       1   0.8848  0.0542      0.78467        0.998
##  14.40     30       1   0.8553  0.0599      0.74561        0.981
##  14.70     29       1   0.8258  0.0647      0.70828        0.963
##  15.63     28       1   0.7963  0.0688      0.67232        0.943
##  16.60     27       1   0.7668  0.0723      0.63749        0.922
##  17.20     26       1   0.7373  0.0753      0.60363        0.901
##  17.93     25       1   0.7078  0.0778      0.57062        0.878
##  18.93     24       1   0.6783  0.0800      0.53839        0.855
##  19.73     23       1   0.6488  0.0818      0.50687        0.831
##  22.40     21       1   0.6180  0.0835      0.47418        0.805
##  23.20     20       1   0.5871  0.0848      0.44224        0.779
##  27.87     19       1   0.5562  0.0858      0.41101        0.753
##  27.97     18       1   0.5253  0.0864      0.38045        0.725
##  28.50     17       1   0.4944  0.0867      0.35056        0.697
##  31.30     16       1   0.4635  0.0866      0.32133        0.668
##  31.43     15       1   0.4326  0.0862      0.29274        0.639
##  31.93     14       1   0.4017  0.0854      0.26481        0.609
##  32.03     13       1   0.3708  0.0842      0.23756        0.579
##  32.83     12       1   0.3399  0.0827      0.21099        0.547
##  32.93     11       1   0.3090  0.0807      0.18516        0.516
##  35.03     10       1   0.2781  0.0783      0.16009        0.483
##  36.63      9       1   0.2472  0.0755      0.13585        0.450
##  39.30      8       1   0.2163  0.0721      0.11253        0.416
##  43.13      7       1   0.1854  0.0681      0.09024        0.381
##  43.40      6       1   0.1545  0.0634      0.06914        0.345
##  45.93      5       1   0.1236  0.0577      0.04947        0.309
##  55.00      3       1   0.0824  0.0511      0.02442        0.278
##  55.20      2       1   0.0412  0.0388      0.00652        0.260
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   4.17      5       1      0.8   0.179       0.5161            1
##   7.87      4       1      0.6   0.219       0.2933            1
##  15.37      3       1      0.4   0.219       0.1367            1
##  19.17      2       1      0.2   0.179       0.0346            1
##  43.20      1       1      0.0     NaN           NA           NA
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=1
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  14.8     20       1    0.950  0.0487       0.8591        1.000
##  15.1     19       1    0.900  0.0671       0.7777        1.000
##  19.7     18       1    0.850  0.0798       0.7071        1.000
##  19.8     17       1    0.800  0.0894       0.6426        0.996
##  19.9     16       1    0.750  0.0968       0.5823        0.966
##  21.0     15       1    0.700  0.1025       0.5254        0.933
##  21.2     14       1    0.650  0.1067       0.4712        0.897
##  21.9     13       1    0.600  0.1095       0.4195        0.858
##  22.4     12       1    0.550  0.1112       0.3700        0.818
##  26.3     11       1    0.500  0.1118       0.3226        0.775
##  26.9     10       1    0.450  0.1112       0.2772        0.731
##  28.6      9       1    0.400  0.1095       0.2339        0.684
##  30.4      8       1    0.350  0.1067       0.1926        0.636
##  30.9      7       1    0.300  0.1025       0.1536        0.586
##  36.4      6       1    0.250  0.0968       0.1170        0.534
##  39.9      5       1    0.200  0.0894       0.0832        0.481
##  40.6      4       1    0.150  0.0798       0.0528        0.426
##  45.6      2       1    0.075  0.0664       0.0132        0.425
##  54.8      1       1    0.000     NaN           NA           NA
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=2
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
##  18.8     10       1      0.9  0.0949       0.7320        1.000
##  21.7      9       1      0.8  0.1265       0.5868        1.000
##  23.3      8       1      0.7  0.1449       0.4665        1.000
##  24.8      7       1      0.6  0.1549       0.3617        0.995
##  37.4      6       1      0.5  0.1581       0.2690        0.929
##  38.8      5       1      0.4  0.1549       0.1872        0.855
##  43.3      4       1      0.3  0.1449       0.1164        0.773
##  44.8      3       1      0.2  0.1265       0.0579        0.691
##  53.0      2       1      0.1  0.0949       0.0156        0.642

summary(KM_trt_fit6_clinical_5,60, extend = TRUE)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, data = TNBC5)
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##     60.00000      0.00000     31.00000      0.04120      0.03875      0.00652
## upper 95% CI
##      0.26037
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60            0            5            0          NaN           NA
## upper 95% CI
##           NA
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60            0           19            0          NaN           NA
## upper 95% CI
##           NA
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000       0.0000       9.0000       0.1000       0.0949       0.0156
## upper 95% CI
##       0.6420

plot(KM_trt_fit6_clinical_5,col = c(1:4),
     main = "Survival graph for surgery+chemotherapy treatments for 70 patients with not more than 5years follow
-up",
     xlab = "months to death", ylab = "percentage of patients")
     legend(120,1,c("Mastectomy+Chemotherapy","Mastectomy only","Breast Conserving+Chemotherapy",
                "Breast Conserving only"), col = (1:4),lwd = 0.5)
```




```
KM_trt_fit6_clinical_over_5 = survfit(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~ BREAST_SURGERY + CHEMOTH
ERAPY,
                                data = TNBC_over5)
KM_trt_fit6_clinical_over_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, data = TNBC_over5)
##
##                                   n events median 0.95LCL 0.95UCL
## BREAST_SURGERY=1, CHEMOTHERAPY=1 31      6     NA      NA      NA
## BREAST_SURGERY=1, CHEMOTHERAPY=2  6      2     NA     129      NA
## BREAST_SURGERY=2, CHEMOTHERAPY=1 44      5     NA     242      NA
## BREAST_SURGERY=2, CHEMOTHERAPY=2 39      6     NA     228      NA

summary(KM_trt_fit6_clinical_over_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, data = TNBC_over5)
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   66.7     31       1    0.968  0.0317        0.908        1.000
##   71.8     30       1    0.935  0.0441        0.853        1.000
##   89.5     27       1    0.901  0.0544        0.800        1.000
##  125.6     25       1    0.865  0.0631        0.750        0.998
##  162.8     22       1    0.825  0.0714        0.697        0.978
##  178.7     19       1    0.782  0.0798        0.640        0.955
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   83.4      6       1    0.833   0.152        0.583            1
##  128.7      5       1    0.667   0.192        0.379            1
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   62.8     43       1    0.977  0.0230        0.933            1
##   70.6     40       1    0.952  0.0329        0.890            1
##   75.2     39       1    0.928  0.0401        0.853            1
##  144.7     20       1    0.882  0.0591        0.773            1
##  241.6      5       1    0.705  0.1646        0.446            1
##
```

```
##                 BREAST_SURGERY=2, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   68.1     39       1    0.974  0.0253        0.926            1
##   75.3     38       1    0.949  0.0353        0.882            1
##   94.9     33       1    0.920  0.0444        0.837            1
##  159.2     22       1    0.878  0.0589        0.770            1
##  204.2      8       1    0.768  0.1149        0.573            1
##  227.8      5       1    0.615  0.1653        0.363            1

summary(KM_trt_fit6_clinical_over_5,60)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, data = TNBC_over5)
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           31            0            1            0            1
## upper 95% CI
##            1
##
##                 BREAST_SURGERY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60            6            0            1            0            1
## upper 95% CI
##            1
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           44            0            1            0            1
## upper 95% CI
##            1
##
##                 BREAST_SURGERY=2, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           39            0            1            0            1
## upper 95% CI
##            1

plot(KM_trt_fit6_clinical_over_5,col = c(1:4),
     main = "Survival graph for surgery+chemotherapy treatments for 120 patients with more than 5years follow-up
",
     xlab = "months to death", ylab = "percentage of patients")
     legend(70,0.4,c("Mastectomy+Chemotherapy","Mastectomy only","Breast Conserving+Chemotherapy",
               "Breast Conserving only"), col = (1:4),lwd = 0.5)
```




```
ccomb = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~BREAST_SURGERY + CHEMOTHERAPY,rho = 0))
ccomb

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     CHEMOTHERAPY, rho = 0)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, CHEMOTHERAPY=1 66       37    25.82     4.836     7.079
## BREAST_SURGERY=1, CHEMOTHERAPY=2 11        7     3.99     2.276     2.393
## BREAST_SURGERY=2, CHEMOTHERAPY=1 64       24    28.26     0.642     0.978
## BREAST_SURGERY=2, CHEMOTHERAPY=2 49       15    24.93     3.954     5.678
##
##  Chisq= 11.8  on 3 degrees of freedom, p= 0.008

ccomb2 = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~BREAST_SURGERY + CHEMOTHERAPY,rho = 1))
ccomb2

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     CHEMOTHERAPY, rho = 1)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, CHEMOTHERAPY=1 66    29.59    20.19     4.384      7.96
## BREAST_SURGERY=1, CHEMOTHERAPY=2 11     5.78     3.09     2.330      3.07
## BREAST_SURGERY=2, CHEMOTHERAPY=1 64    18.78    22.20     0.527      1.00
## BREAST_SURGERY=2, CHEMOTHERAPY=2 49    10.60    19.27     3.903      6.99
##
##  Chisq= 14  on 3 degrees of freedom, p= 0.003

ccomb_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~BREAST_SURGERY + CHEMOTHERAPY,rho
= 0))
ccomb_5

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, rho = 0)
##
##                                  N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, CHEMOTHERAPY=1 35       31    32.25    0.0482     0.102
## BREAST_SURGERY=1, CHEMOTHERAPY=2  5        5     2.13    3.8852     4.065
```

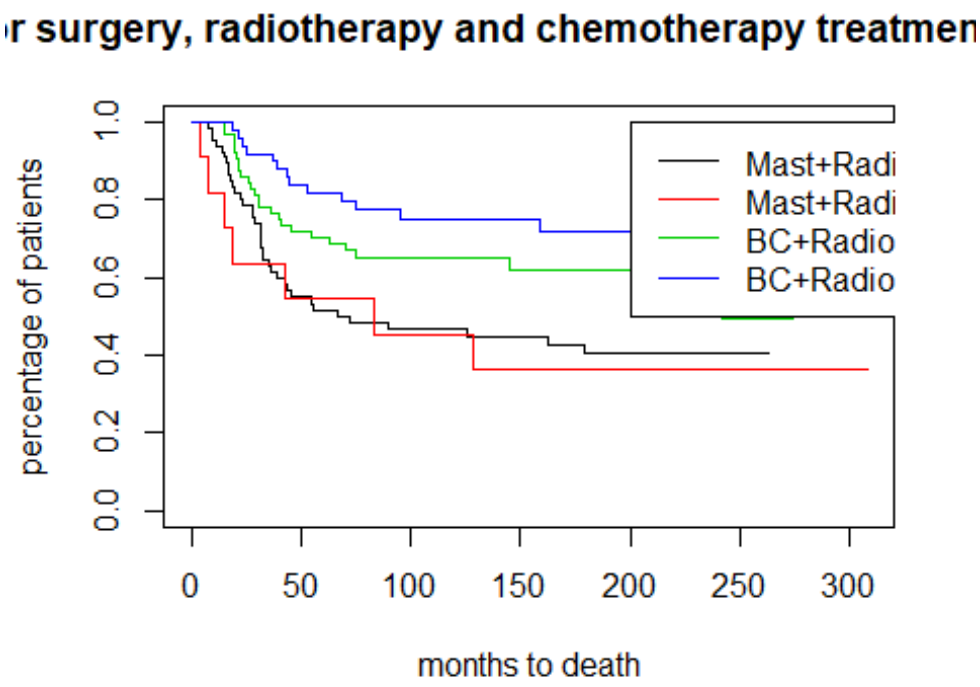
```
## BREAST_SURGERY=2, CHEMOTHERAPY=1 20       19    17.08    0.2151     0.301
## BREAST_SURGERY=2, CHEMOTHERAPY=2 10        9    12.54    1.0012     1.279
##
##  Chisq= 5.2  on 3 degrees of freedom, p= 0.2

ccomb2_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~BREAST_SURGERY + CHEMOTHERAPY,rho
= 1))
ccomb2_5

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, rho = 1)
##
##                                    N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, CHEMOTHERAPY=1 35    17.07    16.75  0.006005  0.017274
## BREAST_SURGERY=1, CHEMOTHERAPY=2  5     3.80     1.27  5.070451  7.430498
## BREAST_SURGERY=2, CHEMOTHERAPY=1 20    10.04    10.10  0.000422  0.000868
## BREAST_SURGERY=2, CHEMOTHERAPY=2 10     3.63     6.42  1.211283  2.300505
##
##  Chisq= 9.1  on 3 degrees of freedom, p= 0.03

ccomb_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~
                                       BREAST_SURGERY + CHEMOTHERAPY,rho = 0))
ccomb_ov

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, rho = 0)
##
##                                    N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, CHEMOTHERAPY=1 31        6     6.07  0.000832   0.00125
## BREAST_SURGERY=1, CHEMOTHERAPY=2  6        2     0.98  1.060855   1.12002
## BREAST_SURGERY=2, CHEMOTHERAPY=1 44        5     5.88  0.131648   0.19229
## BREAST_SURGERY=2, CHEMOTHERAPY=2 39        6     6.07  0.000782   0.00116
##
##  Chisq= 1.2  on 3 degrees of freedom, p= 0.8

ccomb2_ov = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~
                                        BREAST_SURGERY + CHEMOTHERAPY,rho = 1))
ccomb2_ov

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, rho = 1)
##
##                                    N Observed Expected (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, CHEMOTHERAPY=1 31     5.51    5.451  0.000656   0.00108
## BREAST_SURGERY=1, CHEMOTHERAPY=2  6     1.84    0.896  0.995264   1.14421
## BREAST_SURGERY=2, CHEMOTHERAPY=1 44     4.60    5.399  0.119484   0.19037
## BREAST_SURGERY=2, CHEMOTHERAPY=2 39     5.37    5.570  0.007256   0.01175
##
##  Chisq= 1.2  on 3 degrees of freedom, p= 0.7
```

- 2.20. Create survival objects and plot survival graphs for breast surgery, radiotherapy and chemotherapy combination treatments

```
KM_trt_fit7_clinical = survfit(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~ BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY
,
                               data = TNBC_All)
KM_trt_fit7_clinical

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_All)
##
##                                                   n events median 0.95LCL
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1 66     37   71.8    39.3
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2 11      7   83.4    19.2
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1 64     24  241.6   241.6
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2 49     15     NA   204.2
##                                                  0.95UCL
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1      NA
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2      NA
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1      NA
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2      NA

summary(KM_trt_fit7_clinical)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_All)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    8.07     66       1    0.985  0.0150        0.956        1.000
##    9.43     65       1    0.970  0.0211        0.929        1.000
##    9.83     64       1    0.955  0.0256        0.906        1.000
##   11.07     62       1    0.939  0.0295        0.883        0.999
##   14.40     61       1    0.924  0.0328        0.862        0.990
##   14.70     60       1    0.908  0.0357        0.841        0.981
##   15.63     59       1    0.893  0.0382        0.821        0.971
##   16.60     58       1    0.878  0.0406        0.802        0.961
##   17.20     57       1    0.862  0.0427        0.782        0.950
##   17.93     56       1    0.847  0.0446        0.764        0.939
##   18.93     55       1    0.831  0.0464        0.745        0.927
##   19.73     54       1    0.816  0.0480        0.727        0.916
##   22.40     52       1    0.800  0.0496        0.709        0.904
##   23.20     51       1    0.785  0.0510        0.691        0.891
##   27.87     50       1    0.769  0.0524        0.673        0.879
##   27.97     49       1    0.753  0.0536        0.655        0.866
##   28.50     48       1    0.738  0.0547        0.638        0.853
##   31.30     47       1    0.722  0.0558        0.620        0.840
##   31.43     46       1    0.706  0.0567        0.603        0.827
##   31.93     45       1    0.690  0.0576        0.586        0.813
##   32.03     44       1    0.675  0.0584        0.570        0.799
##   32.83     43       1    0.659  0.0591        0.553        0.786
##   32.93     42       1    0.643  0.0597        0.536        0.772
##   35.03     41       1    0.628  0.0603        0.520        0.758
##   36.63     40       1    0.612  0.0608        0.504        0.744
##   39.30     39       1    0.596  0.0612        0.488        0.729
##   43.13     38       1    0.581  0.0616        0.472        0.715
##   43.40     37       1    0.565  0.0619        0.456        0.700
##   45.93     36       1    0.549  0.0621        0.440        0.686
##   55.00     34       1    0.533  0.0624        0.424        0.670
##   55.20     33       1    0.517  0.0625        0.408        0.655
##   66.73     31       1    0.500  0.0627        0.391        0.640
##   71.80     30       1    0.484  0.0628        0.375        0.624
##   89.53     27       1    0.466  0.0630        0.357        0.607
##  125.60     25       1    0.447  0.0631        0.339        0.590
##  162.83     22       1    0.427  0.0635        0.319        0.571
##  178.73     19       1    0.404  0.0640        0.296        0.551
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2
##    time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    4.17     11       1    0.909  0.0867        0.754        1.000
##    7.87     10       1    0.818  0.1163        0.619        1.000
##   15.37      9       1    0.727  0.1343        0.506        1.000
##   19.17      8       1    0.636  0.1450        0.407        0.995
##   43.20      7       1    0.545  0.1501        0.318        0.936
##   83.37      6       1    0.455  0.1501        0.238        0.868
##  128.70      5       1    0.364  0.1450        0.166        0.795
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   14.8     64       1    0.984  0.0155        0.954        1.000
##   15.1     63       1    0.969  0.0217        0.927        1.000
##   19.7     62       1    0.953  0.0264        0.903        1.000
##   19.8     61       1    0.938  0.0303        0.880        0.999
##   19.9     60       1    0.922  0.0335        0.858        0.990
##   21.0     59       1    0.906  0.0364        0.838        0.981
##   21.2     58       1    0.891  0.0390        0.817        0.970
##   21.9     57       1    0.875  0.0413        0.798        0.960
##   22.4     56       1    0.859  0.0435        0.778        0.949
##   26.3     55       1    0.844  0.0454        0.759        0.938
##   26.9     54       1    0.828  0.0472        0.741        0.926
##   28.6     53       1    0.812  0.0488        0.722        0.914
##   30.4     52       1    0.797  0.0503        0.704        0.902
##   30.9     51       1    0.781  0.0517        0.686        0.889
##   36.4     50       1    0.766  0.0530        0.669        0.877
##   39.9     49       1    0.750  0.0541        0.651        0.864
##   40.6     48       1    0.734  0.0552        0.634        0.851
##   45.6     46       1    0.718  0.0563        0.616        0.838
##   54.8     45       1    0.702  0.0572        0.599        0.824
##   62.8     43       1    0.686  0.0582        0.581        0.810
##   70.6     40       1    0.669  0.0592        0.562        0.796
##   75.2     39       1    0.652  0.0601        0.544        0.781
##  144.7     20       1    0.619  0.0654        0.503        0.762
##  241.6      5       1    0.495  0.1225        0.305        0.804
```

```
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   18.8     49       1    0.980  0.0202        0.941        1.000
##   21.7     48       1    0.959  0.0283        0.905        1.000
##   23.3     47       1    0.939  0.0342        0.874        1.000
##   24.8     46       1    0.918  0.0391        0.845        0.998
##   37.4     45       1    0.898  0.0432        0.817        0.987
##   38.8     44       1    0.878  0.0468        0.790        0.974
##   43.3     43       1    0.857  0.0500        0.765        0.961
##   44.8     42       1    0.837  0.0528        0.739        0.947
##   53.0     41       1    0.816  0.0553        0.715        0.932
##   68.1     39       1    0.795  0.0577        0.690        0.917
##   75.3     38       1    0.774  0.0599        0.666        0.901
##   94.9     33       1    0.751  0.0625        0.638        0.884
##  159.2     22       1    0.717  0.0683        0.595        0.864
##  204.2      8       1    0.627  0.1030        0.455        0.865
##  227.8      5       1    0.502  0.1392        0.291        0.864

summary(KM_trt_fit7_clinical,60)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_All)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      31.0000      31.0000       0.5169       0.0625       0.4078
## upper 95% CI
##       0.6552
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##       60.000        6.000        5.000        0.545        0.150        0.318
## upper 95% CI
##        0.936
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      44.0000      19.0000       0.7024       0.0572       0.5988
## upper 95% CI
##       0.8241
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000      39.0000       9.0000       0.8163       0.0553       0.7148
## upper 95% CI
##       0.9323

plot(KM_trt_fit7_clinical,col = c(1:4),
     main="Survival graph for surgery, radiotherapy and chemotherapy treatments for all 190 patients",
     xlab = "months to death", ylab = "percentage of patients")
     legend(200,1,c("Mast+Radio+Chemo","Mast+Radio","BC+Radio+Chemo","BC+Radio"), col = (1:4),lwd = 0.5)
```



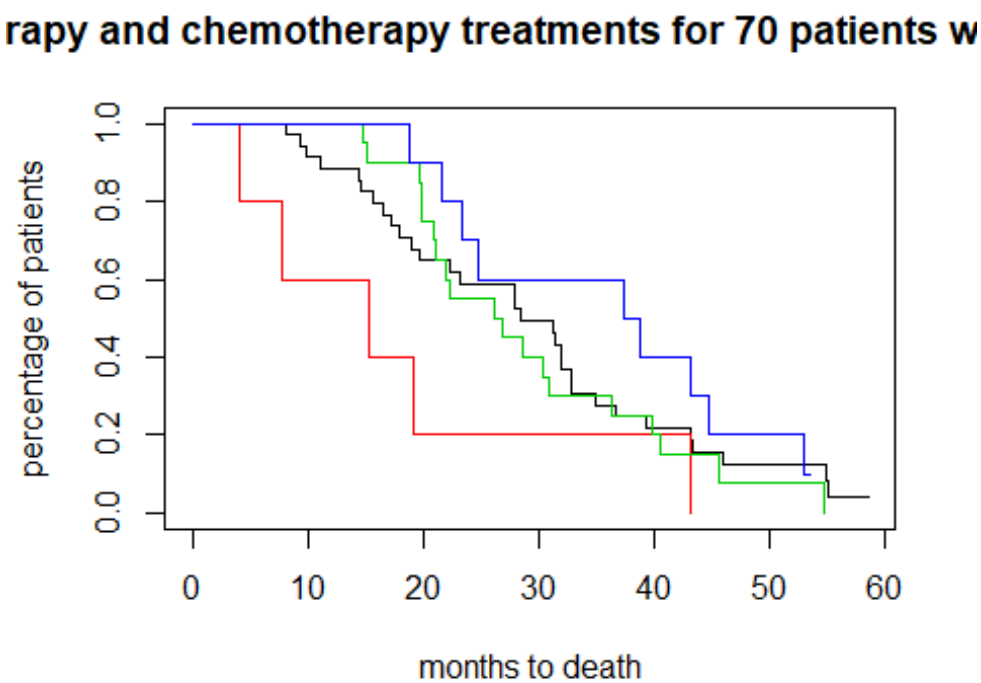

```
KM_trt_fit7_clinical_5 = survfit(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~
                                 BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
KM_trt_fit7_clinical_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
##
##                                                   n events median 0.95LCL
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1 35     31   28.5   22.40
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2  5      5   15.4    7.87
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1 20     19   26.6   21.17
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2 10      9   38.1   23.33
##                                                  0.95UCL
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1    35.0
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2      NA
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1    39.9
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2      NA

summary(KM_trt_fit7_clinical_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   8.07     35       1   0.9714  0.0282      0.91777        1.000
##   9.43     34       1   0.9429  0.0392      0.86901        1.000
##   9.83     33       1   0.9143  0.0473      0.82609        1.000
##  11.07     31       1   0.8848  0.0542      0.78467        0.998
##  14.40     30       1   0.8553  0.0599      0.74561        0.981
##  14.70     29       1   0.8258  0.0647      0.70828        0.963
##  15.63     28       1   0.7963  0.0688      0.67232        0.943
##  16.60     27       1   0.7668  0.0723      0.63749        0.922
##  17.20     26       1   0.7373  0.0753      0.60363        0.901
##  17.93     25       1   0.7078  0.0778      0.57062        0.878
##  18.93     24       1   0.6783  0.0800      0.53839        0.855
##  19.73     23       1   0.6488  0.0818      0.50687        0.831
##  22.40     21       1   0.6180  0.0835      0.47418        0.805
##  23.20     20       1   0.5871  0.0848      0.44224        0.779
##  27.87     19       1   0.5562  0.0858      0.41101        0.753
##  27.97     18       1   0.5253  0.0864      0.38045        0.725
```

```
##  28.50     17       1   0.4944  0.0867      0.35056        0.697
##  31.30     16       1   0.4635  0.0866      0.32133        0.668
##  31.43     15       1   0.4326  0.0862      0.29274        0.639
##  31.93     14       1   0.4017  0.0854      0.26481        0.609
##  32.03     13       1   0.3708  0.0842      0.23756        0.579
##  32.83     12       1   0.3399  0.0827      0.21099        0.547
##  32.93     11       1   0.3090  0.0807      0.18516        0.516
##  35.03     10       1   0.2781  0.0783      0.16009        0.483
##  36.63      9       1   0.2472  0.0755      0.13585        0.450
##  39.30      8       1   0.2163  0.0721      0.11253        0.416
##  43.13      7       1   0.1854  0.0681      0.09024        0.381
##  43.40      6       1   0.1545  0.0634      0.06914        0.345
##  45.93      5       1   0.1236  0.0577      0.04947        0.309
##  55.00      3       1   0.0824  0.0511      0.02442        0.278
##  55.20      2       1   0.0412  0.0388      0.00652        0.260
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   4.17      5       1      0.8   0.179       0.5161            1
##   7.87      4       1      0.6   0.219       0.2933            1
##  15.37      3       1      0.4   0.219       0.1367            1
##  19.17      2       1      0.2   0.179       0.0346            1
##  43.20      1       1      0.0     NaN           NA           NA
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  14.8     20       1    0.950  0.0487       0.8591        1.000
##  15.1     19       1    0.900  0.0671       0.7777        1.000
##  19.7     18       1    0.850  0.0798       0.7071        1.000
##  19.8     17       1    0.800  0.0894       0.6426        0.996
##  19.9     16       1    0.750  0.0968       0.5823        0.966
##  21.0     15       1    0.700  0.1025       0.5254        0.933
##  21.2     14       1    0.650  0.1067       0.4712        0.897
##  21.9     13       1    0.600  0.1095       0.4195        0.858
##  22.4     12       1    0.550  0.1112       0.3700        0.818
##  26.3     11       1    0.500  0.1118       0.3226        0.775
##  26.9     10       1    0.450  0.1112       0.2772        0.731
##  28.6      9       1    0.400  0.1095       0.2339        0.684
##  30.4      8       1    0.350  0.1067       0.1926        0.636
##  30.9      7       1    0.300  0.1025       0.1536        0.586
##  36.4      6       1    0.250  0.0968       0.1170        0.534
##  39.9      5       1    0.200  0.0894       0.0832        0.481
##  40.6      4       1    0.150  0.0798       0.0528        0.426
##  45.6      2       1    0.075  0.0664       0.0132        0.425
##  54.8      1       1    0.000     NaN           NA           NA
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  18.8     10       1      0.9  0.0949       0.7320        1.000
##  21.7      9       1      0.8  0.1265       0.5868        1.000
##  23.3      8       1      0.7  0.1449       0.4665        1.000
##  24.8      7       1      0.6  0.1549       0.3617        0.995
##  37.4      6       1      0.5  0.1581       0.2690        0.929
##  38.8      5       1      0.4  0.1549       0.1872        0.855
##  43.3      4       1      0.3  0.1449       0.1164        0.773
##  44.8      3       1      0.2  0.1265       0.0579        0.691
##  53.0      2       1      0.1  0.0949       0.0156        0.642
```

```
summary(KM_trt_fit7_clinical_5,60, extend = TRUE)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##     60.00000      0.00000     31.00000      0.04120      0.03875      0.00652
## upper 95% CI
##      0.26037
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60            0            5            0          NaN           NA
## upper 95% CI
##           NA
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60            0           19            0          NaN           NA
```

```
## upper 95% CI
##           NA
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##      60.0000       0.0000       9.0000       0.1000       0.0949       0.0156
## upper 95% CI
##       0.6420

plot(KM_trt_fit7_clinical_5,col = c(1:4),
     main="Survival graph for surgery, radiotherapy and chemotherapy treatments for 70 patients with not more th
an 5years            follow-up",
     xlab = "months to death", ylab = "percentage of patients")
     legend(150,1,c("Mast+Radio+Chemo","Mast+Radio","BC+Radio+Chemo","BC+Radio"), col = (1:4),lwd = 0.5)
```



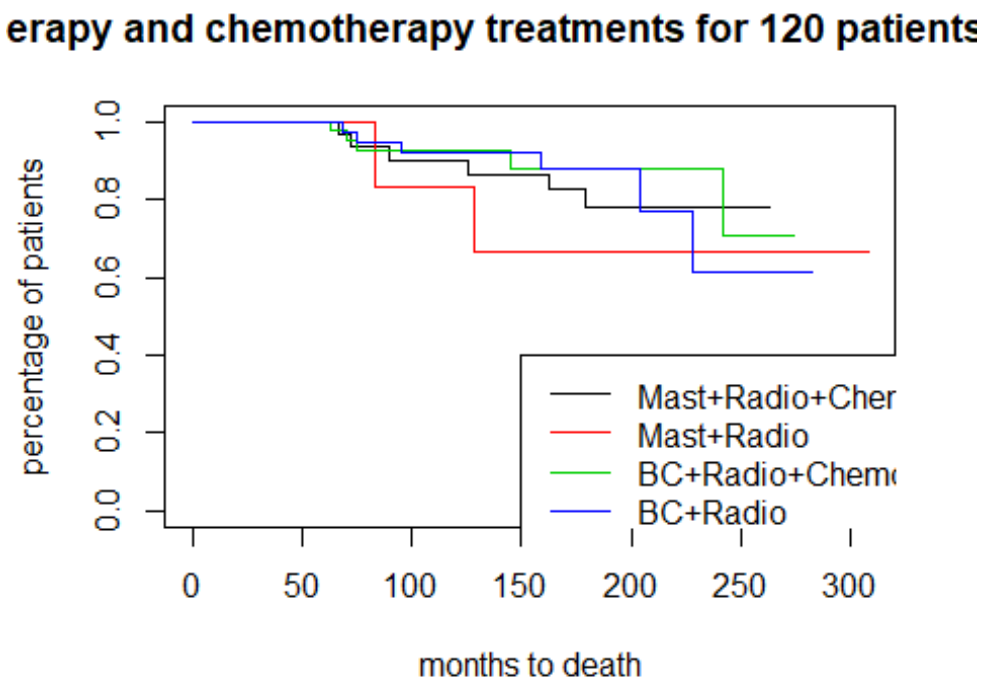

```
KM_trt_fit7_clinical_over_5 = survfit(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~
                                BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over5)
KM_trt_fit7_clinical_over_5

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over5)
##
##                                                  n events median 0.95LCL
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1 31      6     NA      NA
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2  6      2     NA     129
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1 44      5     NA     242
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2 39      6     NA     228
##                                                 0.95UCL
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1      NA
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2      NA
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1      NA
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2      NA

summary(KM_trt_fit7_clinical_over_5)

## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over5)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   66.7     31       1    0.968  0.0317        0.908        1.000
```

```
##   71.8     30       1    0.935  0.0441        0.853        1.000
##   89.5     27       1    0.901  0.0544        0.800        1.000
##  125.6     25       1    0.865  0.0631        0.750        0.998
##  162.8     22       1    0.825  0.0714        0.697        0.978
##  178.7     19       1    0.782  0.0798        0.640        0.955
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   83.4      6       1    0.833   0.152        0.583            1
##  128.7      5       1    0.667   0.192        0.379            1
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   62.8     43       1    0.977  0.0230        0.933            1
##   70.6     40       1    0.952  0.0329        0.890            1
##   75.2     39       1    0.928  0.0401        0.853            1
##  144.7     20       1    0.882  0.0591        0.773            1
##  241.6      5       1    0.705  0.1646        0.446            1
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   68.1     39       1    0.974  0.0253        0.926            1
##   75.3     38       1    0.949  0.0353        0.882            1
##   94.9     33       1    0.920  0.0444        0.837            1
##  159.2     22       1    0.878  0.0589        0.770            1
##  204.2      8       1    0.768  0.1149        0.573            1
##  227.8      5       1    0.615  0.1653        0.363            1
```

```r
summary(KM_trt_fit7_clinical_over_5,60)
```

```
## Call: survfit(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over5)
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           31            0            1            0            1
## upper 95% CI
##            1
##
##                 BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60            6            0            1            0            1
## upper 95% CI
##            1
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           44            0            1            0            1
## upper 95% CI
##            1
##
##                 BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2
##         time       n.risk      n.event     survival      std.err lower 95% CI
##           60           39            0            1            0            1
## upper 95% CI
##            1
```

```r
plot(KM_trt_fit7_clinical_over_5,col = c(1:4),
     main="Survival graph for surgery, radiotherapy and chemotherapy treatments for 120 patients with more than
5years            follow-up",
     xlab = "months to death", ylab = "percentage of patients")
     legend(150,0.4,c("Mast+Radio+Chemo","Mast+Radio","BC+Radio+Chemo","BC+Radio"), col = (1:4),lwd = 0.5)
```




```
dcomb = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~BREAST_SURGERY+RADIO_THERAPY+CHEMOTHERAPY,rho
= 0))
dcomb

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY + CHEMOTHERAPY, rho = 0)
##
##                                                N Observed Expected
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1 66       37    25.82
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2 11        7     3.99
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1 64       24    28.26
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2 49       15    24.93
##                                               (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1     4.836     7.079
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2     2.276     2.393
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1     0.642     0.978
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2     3.954     5.678
##
##  Chisq= 11.8  on 3 degrees of freedom, p= 0.008

dcomb2 = with(TNBC_All, survdiff(Surv(SURVIVAL_MONTHS,VITAL_STATUS)~BREAST_SURGERY+RADIO_THERAPY+CHEMOTHERAPY,rh
o = 1))
dcomb2

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY + CHEMOTHERAPY, rho = 1)
##
##                                                N Observed Expected
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1 66    29.59    20.19
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2 11     5.78     3.09
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1 64    18.78    22.20
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2 49    10.60    19.27
##                                               (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1     4.384      7.96
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2     2.330      3.07
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1     0.527      1.00
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2     3.903      6.99
##
##  Chisq= 14  on 3 degrees of freedom, p= 0.003
```

```
dcomb_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~
                                  BREAST_SURGERY+RADIO_THERAPY+CHEMOTHERAPY,rho = 0))
dcomb_5

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, rho = 0)
##
##                                                  N Observed Expected
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1 35       31    32.25
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2  5        5     2.13
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1 20       19    17.08
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2 10        9    12.54
##                                                 (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1    0.0482     0.102
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2    3.8852     4.065
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1    0.2151     0.301
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2    1.0012     1.279
##
##  Chisq= 5.2  on 3 degrees of freedom, p= 0.2

dcomb2_5 = with(TNBC5, survdiff(Surv(SURVIVAL_MONTHS[1:70],VITAL_STATUS[1:70])~
                                   BREAST_SURGERY+RADIO_THERAPY+CHEMOTHERAPY,rho = 1))
dcomb2_5

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, rho = 1)
##
##                                                  N Observed Expected
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1 35    17.07    16.75
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2  5     3.80     1.27
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1 20    10.04    10.10
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2 10     3.63     6.42
##                                                 (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1  0.006005  0.017274
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2  5.070451  7.430498
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1  0.000422  0.000868
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2  1.211283  2.300505
##
##  Chisq= 9.1  on 3 degrees of freedom, p= 0.03

dcomb_OV = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~
                                  BREAST_SURGERY+RADIO_THERAPY+CHEMOTHERAPY,rho = 0))
dcomb_OV

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, rho = 0)
##
##                                                  N Observed Expected
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1 31        6     6.07
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2  6        2     0.98
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1 44        5     5.88
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2 39        6     6.07
##                                                 (O-E)^2/E (O-E)^2/V
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1  0.000832   0.00125
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2  1.060855   1.12002
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1  0.131648   0.19229
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2  0.000782   0.00116
##
##  Chisq= 1.2  on 3 degrees of freedom, p= 0.8

dcomb2_OV = with(TNBC_over5, survdiff(Surv(SURVIVAL_MONTHS[1:120],VITAL_STATUS[1:120])~
                                   BREAST_SURGERY+RADIO_THERAPY+CHEMOTHERAPY,rho = 1))
dcomb2_OV

## Call:
## survdiff(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, rho = 1)
##
##                                                  N Observed Expected
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1 31     5.51    5.451
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2  6     1.84    0.896
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1 44     4.60    5.399
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2 39     5.37    5.570
##                                                 (O-E)^2/E (O-E)^2/V
```

```
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=1  0.000656   0.00108
## BREAST_SURGERY=1, RADIO_THERAPY=1, CHEMOTHERAPY=2  0.995264   1.14421
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=1  0.119484   0.19037
## BREAST_SURGERY=2, RADIO_THERAPY=1, CHEMOTHERAPY=2  0.007256   0.01175
##
##  Chisq= 1.2  on 3 degrees of freedom, p= 0.7
```

- 2.21. Create objects for Cox Proportional Hazards Model Analyses
- 2.22. Fit the cox proportional hazards model that makes use of relevant co-variates in the patient subgroups
- 2.23. Fit Cox Model for "TNBC_All" object with all 190 TNBC Patients

```
cox1 = coxph(Surv(SURVIVAL_MONTHS, VITAL_STATUS)~BREAST_SURGERY, data = TNBC_All)
summary(cox1)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY,
##     data = TNBC_All)
##
##   n= 190, number of events= 83
##
##                   coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.7045    0.4944   0.2207 -3.193  0.00141 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.4944      2.023    0.3208    0.7618
##
## Concordance= 0.6  (se = 0.028 )
## Likelihood ratio test= 10.13  on 1 df,   p=0.001
## Wald test            = 10.19  on 1 df,   p=0.001
## Score (logrank) test = 10.61  on 1 df,   p=0.001

cox2 = coxph(Surv(SURVIVAL_MONTHS, VITAL_STATUS)~CHEMOTHERAPY, data = TNBC_All)
summary(cox2)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ CHEMOTHERAPY,
##     data = TNBC_All)
##
##   n= 190, number of events= 83
##
##                 coef exp(coef) se(coef)      z Pr(>|z|)
## CHEMOTHERAPY -0.3956    0.6733   0.2492 -1.587    0.112
##
##              exp(coef) exp(-coef) lower .95 upper .95
## CHEMOTHERAPY    0.6733      1.485    0.4131     1.097
##
## Concordance= 0.552  (se = 0.025 )
## Likelihood ratio test= 2.67  on 1 df,   p=0.1
## Wald test            = 2.52  on 1 df,   p=0.1
## Score (logrank) test = 2.55  on 1 df,   p=0.1

cox3 = coxph(Surv(SURVIVAL_MONTHS, VITAL_STATUS)~RADIO_THERAPY, data = TNBC_All)
summary(cox3)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY,
##     data = TNBC_All)
##
##   n= 190, number of events= 83
##
##               coef exp(coef) se(coef)  z Pr(>|z|)
## RADIO_THERAPY   NA        NA        0 NA       NA
##
##               exp(coef) exp(-coef) lower .95 upper .95
## RADIO_THERAPY        NA         NA        NA        NA
##
## Concordance= 0.5  (se = 0 )
## Likelihood ratio test= 0  on 0 df,   p=1
## Wald test            = NA  on 0 df,   p=NA
## Score (logrank) test = 0  on 0 df,   p=1
```

```r
cox4 = coxph(Surv(SURVIVAL_MONTHS, VITAL_STATUS)~RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_All)
summary(cox4)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ RADIO_THERAPY +
##     CHEMOTHERAPY, data = TNBC_All)
##
##   n= 190, number of events= 83
##
##                  coef exp(coef) se(coef)      z Pr(>|z|)
## RADIO_THERAPY      NA        NA   0.0000     NA       NA
## CHEMOTHERAPY  -0.3956    0.6733   0.2492 -1.587    0.112
##
##               exp(coef) exp(-coef) lower .95 upper .95
## RADIO_THERAPY        NA         NA        NA        NA
## CHEMOTHERAPY     0.6733      1.485    0.4131     1.097
##
## Concordance= 0.552  (se = 0.025 )
## Likelihood ratio test= 2.67  on 1 df,   p=0.1
## Wald test            = 2.52  on 1 df,   p=0.1
## Score (logrank) test = 2.55  on 1 df,   p=0.1

cox5 = coxph(Surv(SURVIVAL_MONTHS, VITAL_STATUS)~BREAST_SURGERY + RADIO_THERAPY, data = TNBC_All)
summary(cox5)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY, data = TNBC_All)
##
##   n= 190, number of events= 83
##
##                   coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.7045    0.4944   0.2207 -3.193  0.00141 **
## RADIO_THERAPY       NA        NA   0.0000     NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.4944      2.023    0.3208    0.7618
## RADIO_THERAPY         NA         NA        NA        NA
##
## Concordance= 0.6  (se = 0.028 )
## Likelihood ratio test= 10.13  on 1 df,   p=0.001
## Wald test            = 10.19  on 1 df,   p=0.001
## Score (logrank) test = 10.61  on 1 df,   p=0.001

cox6 = coxph(Surv(SURVIVAL_MONTHS, VITAL_STATUS)~BREAST_SURGERY + CHEMOTHERAPY, data = TNBC_All)
summary(cox6)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     CHEMOTHERAPY, data = TNBC_All)
##
##   n= 190, number of events= 83
##
##                   coef exp(coef) se(coef)     z Pr(>|z|)
## BREAST_SURGERY -0.6555    0.5192   0.2358 -2.78  0.00544 **
## CHEMOTHERAPY   -0.1519    0.8590   0.2663 -0.57  0.56834
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.5192      1.926    0.3270    0.8242
## CHEMOTHERAPY      0.8590      1.164    0.5097    1.4478
##
## Concordance= 0.615  (se = 0.028 )
## Likelihood ratio test= 10.46  on 2 df,   p=0.005
## Wald test            = 10.47  on 2 df,   p=0.005
## Score (logrank) test = 10.9  on 2 df,   p=0.004

cox7 = coxph(Surv(SURVIVAL_MONTHS, VITAL_STATUS)~BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_All)
summary(cox7)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS, VITAL_STATUS) ~ BREAST_SURGERY +
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_All)
##
```

```
##   n= 190, number of events= 83
##
##                   coef exp(coef) se(coef)     z Pr(>|z|)
## BREAST_SURGERY -0.6555    0.5192   0.2358 -2.78  0.00544 **
## RADIO_THERAPY       NA        NA   0.0000    NA       NA
## CHEMOTHERAPY   -0.1519    0.8590   0.2663 -0.57  0.56834
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.5192      1.926    0.3270    0.8242
## RADIO_THERAPY         NA         NA        NA        NA
## CHEMOTHERAPY      0.8590      1.164    0.5097    1.4478
##
## Concordance= 0.615  (se = 0.028 )
## Likelihood ratio test= 10.46  on 2 df,   p=0.005
## Wald test            = 10.47  on 2 df,   p=0.005
## Score (logrank) test = 10.9  on 2 df,   p=0.004
```

- 2.24. Fit Cox Model for "TNBC5" object with the 70-Patient subgroup with not more than 5 years follow-up

```
cox1_5 = coxph(Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70])~BREAST_SURGERY, data = TNBC5)
summary(cox1_5)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY, data = TNBC5)
##
##   n= 70, number of events= 64
##
##                   coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.1064    0.8991   0.2563 -0.415    0.678
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.8991      1.112    0.5441     1.486
##
## Concordance= 0.544  (se = 0.035 )
## Likelihood ratio test= 0.17  on 1 df,   p=0.7
## Wald test            = 0.17  on 1 df,   p=0.7
## Score (logrank) test = 0.17  on 1 df,   p=0.7

cox2_5 = coxph(Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70])~CHEMOTHERAPY, data = TNBC5)
summary(cox2_5)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     CHEMOTHERAPY, data = TNBC5)
##
##   n= 70, number of events= 64
##
##                  coef exp(coef) se(coef)      z Pr(>|z|)
## CHEMOTHERAPY -0.06263   0.93929  0.30624 -0.205    0.838
##
##               exp(coef) exp(-coef) lower .95 upper .95
## CHEMOTHERAPY    0.9393      1.065    0.5154     1.712
##
## Concordance= 0.503  (se = 0.032 )
## Likelihood ratio test= 0.04  on 1 df,   p=0.8
## Wald test            = 0.04  on 1 df,   p=0.8
## Score (logrank) test = 0.04  on 1 df,   p=0.8

cox3_5 = coxph(Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70])~RADIO_THERAPY, data = TNBC5)
summary(cox3_5)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     RADIO_THERAPY, data = TNBC5)
##
##   n= 70, number of events= 64
##
##                coef exp(coef) se(coef)  z Pr(>|z|)
## RADIO_THERAPY   NA        NA        0 NA       NA
##
##               exp(coef) exp(-coef) lower .95 upper .95
## RADIO_THERAPY         NA         NA        NA        NA
##
## Concordance= 0.5  (se = 0 )
## Likelihood ratio test= 0  on 0 df,   p=1
```

```
## Wald test            = NA  on 0 df,   p=NA
## Score (logrank) test = 0  on 0 df,   p=1

cox4_5 = coxph(Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70])~RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
summary(cox4_5)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
##
##   n= 70, number of events= 64
##
##                   coef exp(coef) se(coef)      z Pr(>|z|)
## RADIO_THERAPY       NA         NA  0.00000     NA       NA
## CHEMOTHERAPY  -0.06263   0.93929  0.30624 -0.205    0.838
##
##               exp(coef) exp(-coef) lower .95 upper .95
## RADIO_THERAPY         NA         NA        NA        NA
## CHEMOTHERAPY     0.9393      1.065    0.5154     1.712
##
## Concordance= 0.503  (se = 0.032 )
## Likelihood ratio test= 0.04  on 1 df,   p=0.8
## Wald test            = 0.04  on 1 df,   p=0.8
## Score (logrank) test = 0.04  on 1 df,   p=0.8

cox5_5 = coxph(Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70])~BREAST_SURGERY + RADIO_THERAPY, data = TNBC5)
summary(cox5_5)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY, data = TNBC5)
##
##   n= 70, number of events= 64
##
##                   coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.1064    0.8991   0.2563 -0.415    0.678
## RADIO_THERAPY       NA        NA   0.0000     NA       NA
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.8991      1.112    0.5441     1.486
## RADIO_THERAPY         NA         NA        NA        NA
##
## Concordance= 0.544  (se = 0.035 )
## Likelihood ratio test= 0.17  on 1 df,   p=0.7
## Wald test            = 0.17  on 1 df,   p=0.7
## Score (logrank) test = 0.17  on 1 df,   p=0.7

cox6_5 = coxph(Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70])~BREAST_SURGERY + CHEMOTHERAPY, data = TNBC5)
summary(cox6_5)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, data = TNBC5)
##
##   n= 70, number of events= 64
##
##                    coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.10234   0.90273  0.28266 -0.362    0.717
## CHEMOTHERAPY   -0.01151   0.98856  0.33774 -0.034    0.973
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.9027      1.108    0.5187     1.571
## CHEMOTHERAPY      0.9886      1.012    0.5099     1.916
##
## Concordance= 0.539  (se = 0.037 )
## Likelihood ratio test= 0.17  on 2 df,   p=0.9
## Wald test            = 0.17  on 2 df,   p=0.9
## Score (logrank) test = 0.17  on 2 df,   p=0.9

cox7_5 = coxph(Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70])~BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY,
               data = TNBC5)
summary(cox7_5)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:70], VITAL_STATUS[1:70]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC5)
##
```

```
##   n= 70, number of events= 64
##
##                     coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.10234   0.90273  0.28266 -0.362    0.717
## RADIO_THERAPY        NA        NA  0.00000     NA       NA
## CHEMOTHERAPY   -0.01151   0.98856  0.33774 -0.034    0.973
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.9027      1.108    0.5187     1.571
## RADIO_THERAPY         NA         NA        NA        NA
## CHEMOTHERAPY      0.9886      1.012    0.5099     1.916
##
## Concordance= 0.539  (se = 0.037 )
## Likelihood ratio test= 0.17  on 2 df,   p=0.9
## Wald test            = 0.17  on 2 df,   p=0.9
## Score (logrank) test = 0.17  on 2 df,   p=0.9
```

- 2.25. Fit Cox Model for "TNBC_over5" object with 120-Patient subgroup with more than 5 years follow-up

```
cox1_ov = coxph(Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120])~BREAST_SURGERY, data = TNBC_over5)
summary(cox1_ov)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY, data = TNBC_over5)
##
##   n= 120, number of events= 19
##
##                    coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.2137    0.8076   0.4699 -0.455    0.649
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.8076      1.238    0.3215     2.028
##
## Concordance= 0.551  (se = 0.061 )
## Likelihood ratio test= 0.2  on 1 df,   p=0.7
## Wald test            = 0.21  on 1 df,   p=0.6
## Score (logrank) test = 0.21  on 1 df,   p=0.6

cox2_ov = coxph(Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120])~CHEMOTHERAPY, data = TNBC_over5)
summary(cox2_ov)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     CHEMOTHERAPY, data = TNBC_over5)
##
##   n= 120, number of events= 19
##
##                 coef exp(coef) se(coef)     z Pr(>|z|)
## CHEMOTHERAPY 0.2120    1.2361   0.4676 0.453     0.65
##
##               exp(coef) exp(-coef) lower .95 upper .95
## CHEMOTHERAPY     1.236      0.809    0.4944     3.091
##
## Concordance= 0.504  (se = 0.06 )
## Likelihood ratio test= 0.2  on 1 df,   p=0.7
## Wald test            = 0.21  on 1 df,   p=0.7
## Score (logrank) test = 0.21  on 1 df,   p=0.6

cox3_ov = coxph(Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120])~RADIO_THERAPY, data = TNBC_over5)
summary(cox3_ov)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     RADIO_THERAPY, data = TNBC_over5)
##
##   n= 120, number of events= 19
##
##                coef exp(coef) se(coef)  z Pr(>|z|)
## RADIO_THERAPY   NA        NA        0 NA       NA
##
##                exp(coef) exp(-coef) lower .95 upper .95
## RADIO_THERAPY         NA         NA        NA        NA
##
## Concordance= 0.5  (se = 0 )
## Likelihood ratio test= 0  on 0 df,   p=1
## Wald test            = NA  on 0 df,   p=NA
## Score (logrank) test = 0  on 0 df,   p=1
```

```
cox4_ov = coxph(Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120])~RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over
5)
summary(cox4_ov)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over5)
##
##   n= 120, number of events= 19
##
##                  coef exp(coef) se(coef)     z Pr(>|z|)
## RADIO_THERAPY     NA        NA   0.0000    NA       NA
## CHEMOTHERAPY  0.2120    1.2361   0.4676 0.453     0.65
##
##               exp(coef) exp(-coef) lower .95 upper .95
## RADIO_THERAPY        NA         NA        NA        NA
## CHEMOTHERAPY      1.236      0.809    0.4944     3.091
##
## Concordance= 0.504  (se = 0.06 )
## Likelihood ratio test= 0.2  on 1 df,   p=0.7
## Wald test            = 0.21  on 1 df,   p=0.7
## Score (logrank) test = 0.21  on 1 df,   p=0.6

cox5_ov = coxph(Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120])~BREAST_SURGERY + RADIO_THERAPY, data = TNBC_ov
er5)
summary(cox5_ov)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY, data = TNBC_over5)
##
##   n= 120, number of events= 19
##
##                    coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.2137    0.8076   0.4699 -0.455    0.649
## RADIO_THERAPY       NA        NA   0.0000     NA       NA
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.8076      1.238    0.3215     2.028
## RADIO_THERAPY         NA         NA        NA        NA
##
## Concordance= 0.551  (se = 0.061 )
## Likelihood ratio test= 0.2  on 1 df,   p=0.7
## Wald test            = 0.21  on 1 df,   p=0.6
## Score (logrank) test = 0.21  on 1 df,   p=0.6

cox6_ov = coxph(Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120])~BREAST_SURGERY + CHEMOTHERAPY, data = TNBC_ove
r5)
summary(cox6_ov)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + CHEMOTHERAPY, data = TNBC_over5)
##
##   n= 120, number of events= 19
##
##                    coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.3447    0.7084   0.5119 -0.673    0.501
## CHEMOTHERAPY    0.3428    1.4088   0.5100  0.672    0.502
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.7084     1.4116    0.2597     1.932
## CHEMOTHERAPY      1.4088     0.7098    0.5185     3.828
##
## Concordance= 0.558  (se = 0.072 )
## Likelihood ratio test= 0.65  on 2 df,   p=0.7
## Wald test            = 0.65  on 2 df,   p=0.7
## Score (logrank) test = 0.65  on 2 df,   p=0.7

cox7_ov = coxph(Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120])~BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY,
                data = TNBC_over5)
summary(cox7_ov)

## Call:
## coxph(formula = Surv(SURVIVAL_MONTHS[1:120], VITAL_STATUS[1:120]) ~
##     BREAST_SURGERY + RADIO_THERAPY + CHEMOTHERAPY, data = TNBC_over5)
##
```

```
##   n= 120, number of events= 19
##
##                   coef exp(coef) se(coef)      z Pr(>|z|)
## BREAST_SURGERY -0.3447    0.7084   0.5119 -0.673    0.501
## RADIO_THERAPY       NA        NA   0.0000     NA       NA
## CHEMOTHERAPY    0.3428    1.4088   0.5100  0.672    0.502
##
##                exp(coef) exp(-coef) lower .95 upper .95
## BREAST_SURGERY    0.7084     1.4116    0.2597     1.932
## RADIO_THERAPY         NA         NA        NA        NA
## CHEMOTHERAPY      1.4088     0.7098    0.5185     3.828
##
## Concordance= 0.558  (se = 0.072 )
## Likelihood ratio test= 0.65  on 2 df,   p=0.7
## Wall test            = 0.65  on 2 df,   p=0.7
## Score (logrank) test = 0.65  on 2 df,   p=0.7
```

==========
==========

4.0 Method 4: Machine Learning Analysis

Part A: This analysis captures clinical features ONLY

- 4.1. Load mlbench and caret packages respectively
- 4.2. For this analysis, remove RADIO_THERAPY variable to avoid collinearity
- 4.3. Transform and rename the object for the 190 Patient Samples from "clin" to "TNBC_1"
- 4.4. Convert integer values to numeric and ensure that dependent variable "VITAL_STATUS" is of the factor class

```
library(mlbench)

## Warning: package 'mlbench' was built under R version 3.6.3

library(caret)

## Loading required package: lattice

## Warning: package 'lattice' was built under R version 3.6.3

##
## Attaching package: 'caret'

## The following object is masked from 'package:survival':
##
##     cluster

# remove radiotherapy feature from the clin object and save in the TNBC_1 object
TNBC_1 = clin[-6]
for (i in 1:23) {
  TNBC_1[,i] = as.numeric(as.integer(TNBC_1[,i]))
}

TNBC_1$VITAL_STATUS = as.factor(as.integer(TNBC_1$VITAL_STATUS))

sapply(TNBC_1, class)

##     AGE_AT_DIAGNOSIS            AGE_GROUP       BREAST_SURGERY
##            "numeric"            "numeric"            "numeric"
##          CANCER_TYPE         CHEMOTHERAPY      HORMONE_THERAPY
##            "numeric"            "numeric"            "numeric"
##                PAM50               COHORT               ER_IHC
##            "numeric"            "numeric"            "numeric"
##          TUMOR_GRADE           TUMOR_SIZE          TUMOR_STAGE
##            "numeric"            "numeric"            "numeric"
##            HER2_SNP6 HISTOLOGICAL_SUBTYPE          CELLULARITY
##            "numeric"            "numeric"            "numeric"
##            MENOPAUSE             INTCLUST           LATERALITY
##            "numeric"            "numeric"            "numeric"
##       MUTATION_COUNT                  NPI      SURVIVAL_MONTHS
##            "numeric"            "numeric"            "numeric"
##      SURVIVAL_GROUPS            THREEGENE         VITAL_STATUS
##            "numeric"            "numeric"             "factor"
```

```r
names(TNBC_1)

##  [1] "AGE_AT_DIAGNOSIS"     "AGE_GROUP"            "BREAST_SURGERY"
##  [4] "CANCER_TYPE"          "CHEMOTHERAPY"         "HORMONE_THERAPY"
##  [7] "PAM50"                "COHORT"               "ER_IHC"
## [10] "TUMOR_GRADE"          "TUMOR_SIZE"           "TUMOR_STAGE"
## [13] "HER2_SNP6"            "HISTOLOGICAL_SUBTYPE" "CELLULARITY"
## [16] "MENOPAUSE"            "INTCLUST"             "LATERALITY"
## [19] "MUTATION_COUNT"       "NPI"                  "SURVIVAL_MONTHS"
## [22] "SURVIVAL_GROUPS"      "THREEGENE"            "VITAL_STATUS"
```

- 4.5. Perform feature selection. Correlation analyses findings in Section 1 already pointed to the fact that some features have no association with the outcome of the dependent variable.
- 4.6. Feature Selection using Boruta, mlbench and Caret packages respectively

```r
library(Boruta)

## Warning: package 'Boruta' was built under R version 3.6.3

# Feature Selection operation
set.seed(7)
boruta = Boruta(TNBC_1$VITAL_STATUS~.,data = TNBC_1, doTrace = 4, maxRuns = 100)

##  1. run of importance source...

## Assigned hit to 4 attributes out of 23 undecided.

##  2. run of importance source...

## Assigned hit to 6 attributes out of 23 undecided.

##  3. run of importance source...

## Assigned hit to 3 attributes out of 23 undecided.

##  4. run of importance source...

## Assigned hit to 7 attributes out of 23 undecided.

##  5. run of importance source...

## Assigned hit to 3 attributes out of 23 undecided.

##  6. run of importance source...

## Assigned hit to 4 attributes out of 23 undecided.

##  7. run of importance source...

## Assigned hit to 7 attributes out of 23 undecided.

##  8. run of importance source...

## Assigned hit to 3 attributes out of 23 undecided.

##  9. run of importance source...

## Assigned hit to 4 attributes out of 23 undecided.

##  10. run of importance source...

## Assigned hit to 5 attributes out of 23 undecided.

##  11. run of importance source...

## Assigned hit to 5 attributes out of 23 undecided.

##  12. run of importance source...

## Assigned hit to 3 attributes out of 23 undecided.

## After 12 iterations, +1.7 secs:
```

```
##  confirmed 3 attributes: NPI, SURVIVAL_GROUPS, SURVIVAL_MONTHS;

##  rejected 15 attributes: AGE_AT_DIAGNOSIS, AGE_GROUP, CELLULARITY, CHEMOTHERAPY, COHORT and 10 more;

##  still have 5 attributes left.

##  13. run of importance source...

## Assigned hit to 5 attributes out of 5 undecided.

##  14. run of importance source...

## Assigned hit to 3 attributes out of 5 undecided.

##  15. run of importance source...

## Assigned hit to 4 attributes out of 5 undecided.

##  16. run of importance source...

## Assigned hit to 5 attributes out of 5 undecided.

##  17. run of importance source...

## Assigned hit to 2 attributes out of 5 undecided.

##  18. run of importance source...

## Assigned hit to 1 attribute out of 5 undecided.

##  19. run of importance source...

## Assigned hit to 3 attributes out of 5 undecided.

##  20. run of importance source...

## None of undecided attributes scored a hit.

##  21. run of importance source...

## Assigned hit to 2 attributes out of 5 undecided.

##  22. run of importance source...

## None of undecided attributes scored a hit.

##  23. run of importance source...

## Assigned hit to 2 attributes out of 5 undecided.

##  24. run of importance source...

## Assigned hit to 2 attributes out of 5 undecided.

##  25. run of importance source...

## None of undecided attributes scored a hit.

##  26. run of importance source...

## Assigned hit to 4 attributes out of 5 undecided.

##  27. run of importance source...

## Assigned hit to 4 attributes out of 5 undecided.

##  28. run of importance source...

## Assigned hit to 2 attributes out of 5 undecided.

##  29. run of importance source...
```

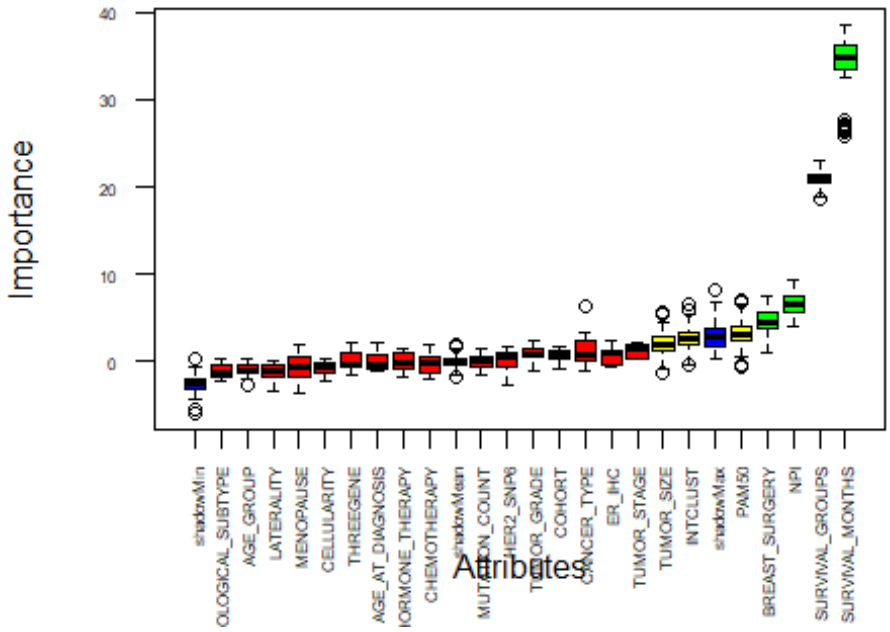
```
## None of undecided attributes scored a hit.

##  30. run of importance source...

## Assigned hit to 2 attributes out of 5 undecided.

##  31. run of importance source...

## Assigned hit to 1 attribute out of 5 undecided.

##  32. run of importance source...

## None of undecided attributes scored a hit.

## After 32 iterations, +3.1 secs:

##  rejected 1 attribute: CANCER_TYPE;

##  still have 4 attributes left.

##  33. run of importance source...

## Assigned hit to 2 attributes out of 4 undecided.

##  34. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

##  35. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

##  36. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.

##  37. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

##  38. run of importance source...

## None of undecided attributes scored a hit.

##  39. run of importance source...

## Assigned hit to 1 attribute out of 4 undecided.

##  40. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

##  41. run of importance source...

## None of undecided attributes scored a hit.

##  42. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

##  43. run of importance source...

## None of undecided attributes scored a hit.

##  44. run of importance source...

## None of undecided attributes scored a hit.

##  45. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.
```

```
##  46. run of importance source...

## Assigned hit to 1 attribute out of 4 undecided.

##  47. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

##  48. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.

##  49. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.

##  50. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

##  51. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

##  52. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

##  53. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.

##  54. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.

##  55. run of importance source...

## Assigned hit to 2 attributes out of 4 undecided.

##  56. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.

##  57. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.

##  58. run of importance source...

## Assigned hit to 3 attributes out of 4 undecided.

## After 58 iterations, +4.7 secs:

##  confirmed 1 attribute: BREAST_SURGERY;

##  still have 3 attributes left.

##  59. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  60. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  61. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  62. run of importance source...
```

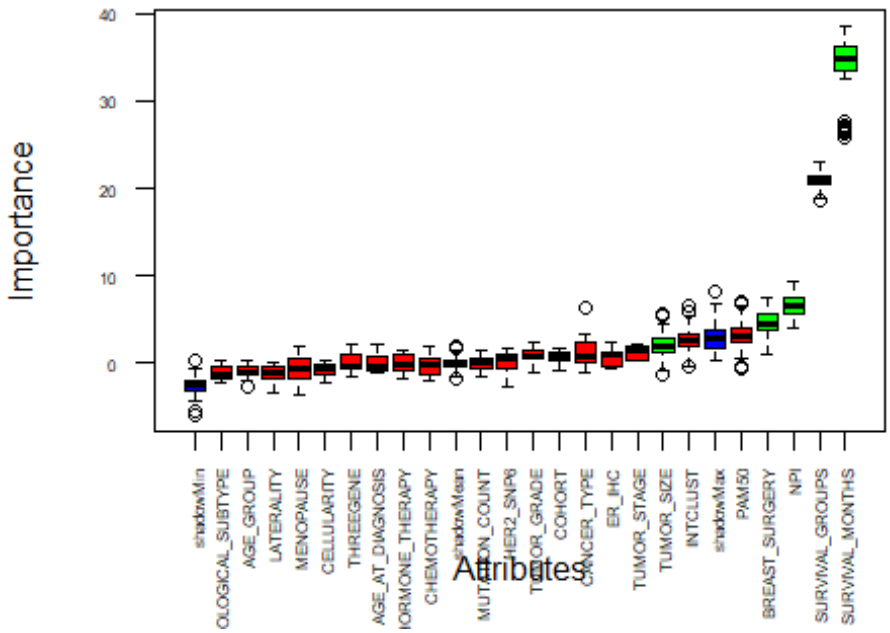
```
## Assigned hit to 1 attribute out of 3 undecided.

##  63. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

##  64. run of importance source...

## None of undecided attributes scored a hit.

##  65. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  66. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  67. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

##  68. run of importance source...

## None of undecided attributes scored a hit.

##  69. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  70. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

##  71. run of importance source...

## None of undecided attributes scored a hit.

##  72. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  73. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

##  74. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  75. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  76. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  77. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  78. run of importance source...

## None of undecided attributes scored a hit.

##  79. run of importance source...

## None of undecided attributes scored a hit.

##  80. run of importance source...
```

```
## None of undecided attributes scored a hit.

##  81. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  82. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  83. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  84. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  85. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  86. run of importance source...

## None of undecided attributes scored a hit.

##  87. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  88. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

##  89. run of importance source...

## None of undecided attributes scored a hit.

##  90. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  91. run of importance source...

## None of undecided attributes scored a hit.

##  92. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  93. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

##  94. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  95. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

##  96. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  97. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

##  98. run of importance source...
```

```
## Assigned hit to 2 attributes out of 3 undecided.

##  99. run of importance source...

## None of undecided attributes scored a hit.

plot(boruta, las = 2, cex.axis = 0.5)
```




```
# reclassify features that were designated as tentative
bor = TentativeRoughFix(boruta)
plot(bor, las = 2, cex.axis = 0.5)
```




```
attStats(boruta)

##                          meanImp  medianImp     minImp      maxImp   normHits
## AGE_AT_DIAGNOSIS      0.14507098 -0.2459541 -1.0458331  2.12717045 0.00000000
## AGE_GROUP            -1.00932434 -1.0707263 -2.7231150  0.39932954 0.00000000
## BREAST_SURGERY        4.53392288  4.5018779  0.9732793  7.55084377 0.74747475
## CANCER_TYPE           1.25016442  0.9142108 -0.9797365  6.31751628 0.06060606
## CHEMOTHERAPY         -0.20454892 -0.1470157 -1.8925093  2.02486834 0.00000000
## HORMONE_THERAPY      -0.01790394 -0.1949820 -1.7745223  1.44267414 0.00000000
## PAM50                 3.27820975  3.1426584 -0.4664953  6.99953684 0.58585859
## COHORT                0.67834767  0.7590903 -0.8591791  1.67150211 0.00000000
## ER_IHC                0.74276620  0.9281919 -0.6328097  2.35725107 0.00000000
## TUMOR_GRADE           0.94953319  0.7587302 -0.9437769  2.35719154 0.00000000
## TUMOR_SIZE            2.04582380  1.8837540 -1.2908259  5.58791348 0.37373737
## TUMOR_STAGE           1.24015970  1.4006065  0.2360888  2.20624756 0.00000000
## HER2_SNP6             0.10177859  0.4791761 -2.6915589  1.66720649 0.00000000
## HISTOLOGICAL_SUBTYPE -1.09033139 -1.2868738 -2.3154210  0.26346929 0.00000000
## CELLULARITY          -0.74809266 -0.4871961 -2.1774244  0.29271237 0.00000000
## MENOPAUSE            -0.63850713 -0.6642741 -3.4955783  1.99503517 0.00000000
## INTCLUST              2.63035806  2.6724020 -0.4025485  6.50622511 0.49494949
## LATERALITY           -1.14982277 -1.0411424 -3.3872808  0.04033673 0.00000000
## MUTATION_COUNT        0.12595963  0.1982859 -1.5241230  1.52276208 0.00000000
## NPI                   6.63742769  6.5179537  3.9709403  9.29863541 0.97979798
## SURVIVAL_MONTHS      34.12595312 34.8135810 25.7081816 38.57936616 1.00000000
## SURVIVAL_GROUPS      20.93254542 20.9208266 18.6367008 23.01524760 1.00000000
## THREEGENE             0.12283003 -0.2824960 -1.5195408  2.12336100 0.00000000
##                       decision
## AGE_AT_DIAGNOSIS      Rejected
## AGE_GROUP             Rejected
## BREAST_SURGERY       Confirmed
## CANCER_TYPE           Rejected
## CHEMOTHERAPY          Rejected
## HORMONE_THERAPY       Rejected
## PAM50                Tentative
## COHORT                Rejected
## ER_IHC                Rejected
## TUMOR_GRADE           Rejected
## TUMOR_SIZE           Tentative
## TUMOR_STAGE           Rejected
## HER2_SNP6             Rejected
## HISTOLOGICAL_SUBTYPE  Rejected
## CELLULARITY           Rejected
```

```
## MENOPAUSE             Rejected
## INTCLUST             Tentative
## LATERALITY            Rejected
## MUTATION_COUNT        Rejected
## NPI                  Confirmed
## SURVIVAL_MONTHS      Confirmed
## SURVIVAL_GROUPS      Confirmed
## THREEGENE             Rejected

print(bor)

## Boruta performed 99 iterations in 6.982829 secs.
## Tentatives roughfixed over the last 99 iterations.
##  5 attributes confirmed important: BREAST_SURGERY, NPI,
## SURVIVAL_GROUPS, SURVIVAL_MONTHS, TUMOR_SIZE;
##  18 attributes confirmed unimportant: AGE_AT_DIAGNOSIS, AGE_GROUP,
## CANCER_TYPE, CELLULARITY, CHEMOTHERAPY and 13 more;

# create object and limit features to only the 5 confirmed features from the boruta analysis + the dependent var
iable

col_order2 = c("BREAST_SURGERY","TUMOR_SIZE","NPI","SURVIVAL_GROUPS","SURVIVAL_MONTHS","VITAL_STATUS")
#View(TNBC_1)
TNBC_1 = TNBC_1[, col_order2]
str(TNBC_1)

## 'data.frame':    190 obs. of  6 variables:
##  $ BREAST_SURGERY : num  1 1 1 1 1 2 2 1 2 1 ...
##  $ TUMOR_SIZE     : num  65 17 40 39 25 13 30 16 24 34 ...
##  $ NPI            : num  6 4 6 4 4 4 4 4 4 6 ...
##  $ SURVIVAL_GROUPS: num  1 2 1 1 2 2 2 1 2 1 ...
##  $ SURVIVAL_MONTHS: num  22 153 28 8 66 132 114 10 78 17 ...
##  $ VITAL_STATUS   : Factor w/ 2 levels "0","1": 2 1 2 2 2 1 1 1 1 2 ...

levels(TNBC_1$VITAL_STATUS)

## [1] "0" "1"
```

- The feature selection operation using the Boruta package works as follows:
- Firstly, it introduces randomness into the data set by producing shuffled copies of all the features of interest (these are referred to as shadow features).
- Next, it trains a random forest classifier on the extended data set and applies a feature importance measure (the default is Mean Decrease Accuracy) to ascertain the relevance of each of the features. Those with higher means are classified as more important.
- Afterwards, at every given iteration, it checks whether a real feature has a higher importance than the best of its shadow features (in other words, it checks to see if the feature has a higher Z-score than the maximum Z-score of its shadow features) and therefore eliminates features which are termed highly irrelevant.
- Lastly, the Boruta algorithm stops either when all features gets confirmed or rejected or it when reaches a specified limit of random forest runs
- 4.7. Split the TNBC_1 clinical features object into training and validation groups (objects)
- 4.8. Split out validation dataset
- 4.9. Do this by creating a list of 70% of the rows in the original dataset to serve as training set

```
#Living
Living = TNBC_1[TNBC_1$VITAL_STATUS == 0,]

#Died of Disease
Deceased = TNBC_1[TNBC_1$VITAL_STATUS == 1,]

set.seed(7)
validationIndex = createDataPartition(TNBC_1$VITAL_STATUS, p=0.70, list = FALSE)

#select 30% of the data for validation

validation = TNBC_1[-validationIndex,]

#use the remaining 70% of data to train the model
dataset = TNBC_1[validationIndex,]
```

```r
# Check that the distribution of the dependent variable is similar in train and test sets
prop.table(table(TNBC_1$VITAL_STATUS))

##
##         0         1
## 0.5631579 0.4368421

prop.table(table(dataset$VITAL_STATUS))

##
##         0         1
## 0.5597015 0.4402985

prop.table(table(validation$VITAL_STATUS))

##
##         0         1
## 0.5714286 0.4285714
```

- The derived figures show that the ratio of the "Living":"Died of Disease" was proportionally split

- 4.10. Analyze the data using descriptive statistics

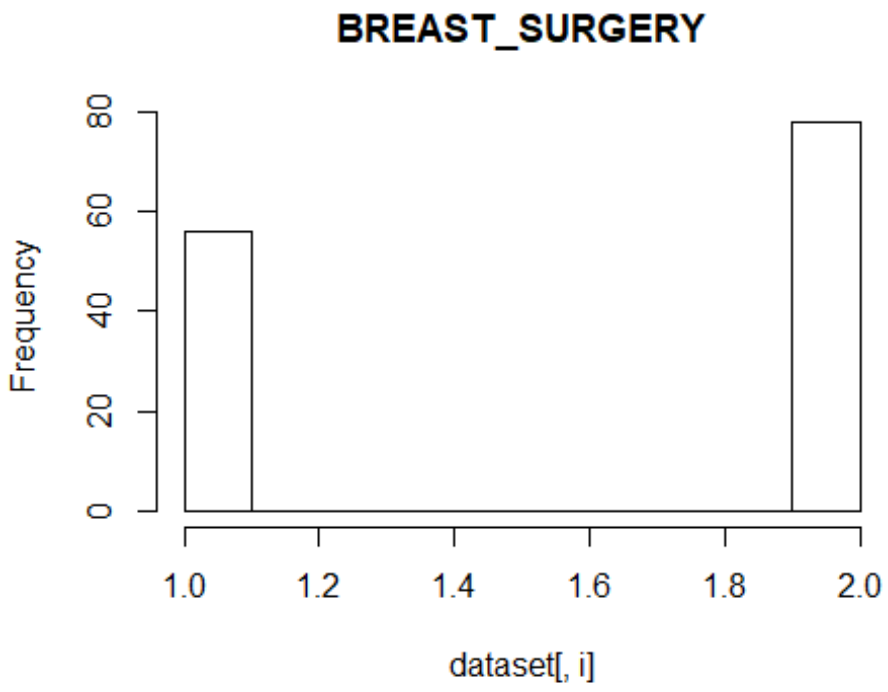
```r
# Dimension of the datasets
dim(dataset)

## [1] 134   6

dim(validation)

## [1] 56  6

#class distribution of training set
cbind(freq = table(dataset$VITAL_STATUS), percentage = prop.table(table(dataset$VITAL_STATUS))*100)

##   freq percentage
## 0   75   55.97015
## 1   59   44.02985

#summarize correlations between input variables to ascertain if there is the potential for col-linearity
cor(dataset[,1:5])

##                 BREAST_SURGERY TUMOR_SIZE        NPI SURVIVAL_GROUPS
## BREAST_SURGERY      1.00000000 -0.1401290 -0.3672388       0.2991589
## TUMOR_SIZE         -0.14012896  1.0000000  0.2220290      -0.1233120
## NPI                -0.36723879  0.2220290  1.0000000      -0.3071684
## SURVIVAL_GROUPS     0.29915885 -0.1233120 -0.3071684       1.0000000
## SURVIVAL_MONTHS     0.08953678 -0.1008098 -0.1917113       0.7706964
##                 SURVIVAL_MONTHS
## BREAST_SURGERY       0.08953678
## TUMOR_SIZE          -0.10080979
## NPI                 -0.19171129
## SURVIVAL_GROUPS      0.77069641
## SURVIVAL_MONTHS      1.00000000
```

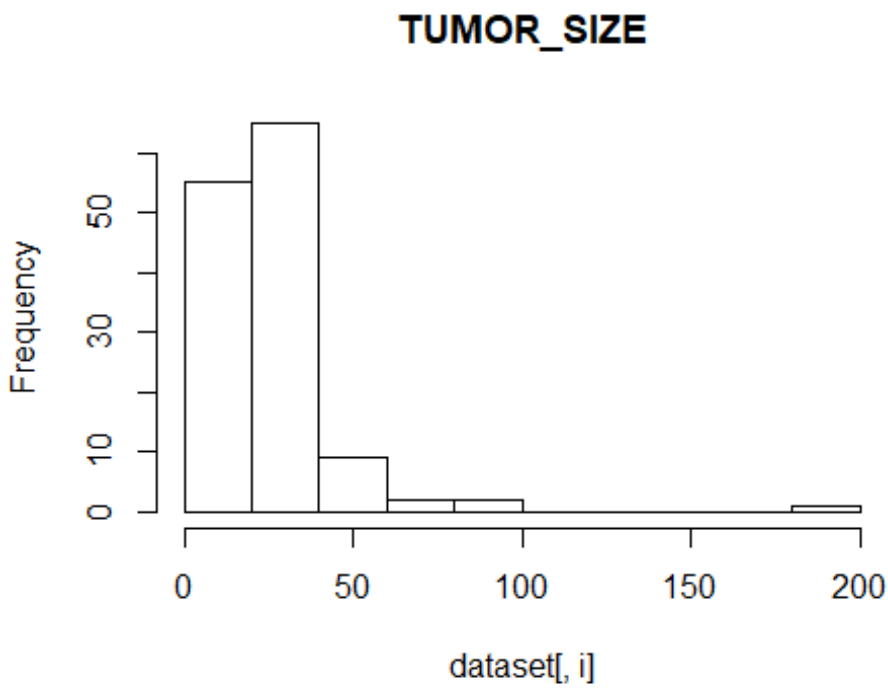
- Lowest correlation is between PAM50 and survival months

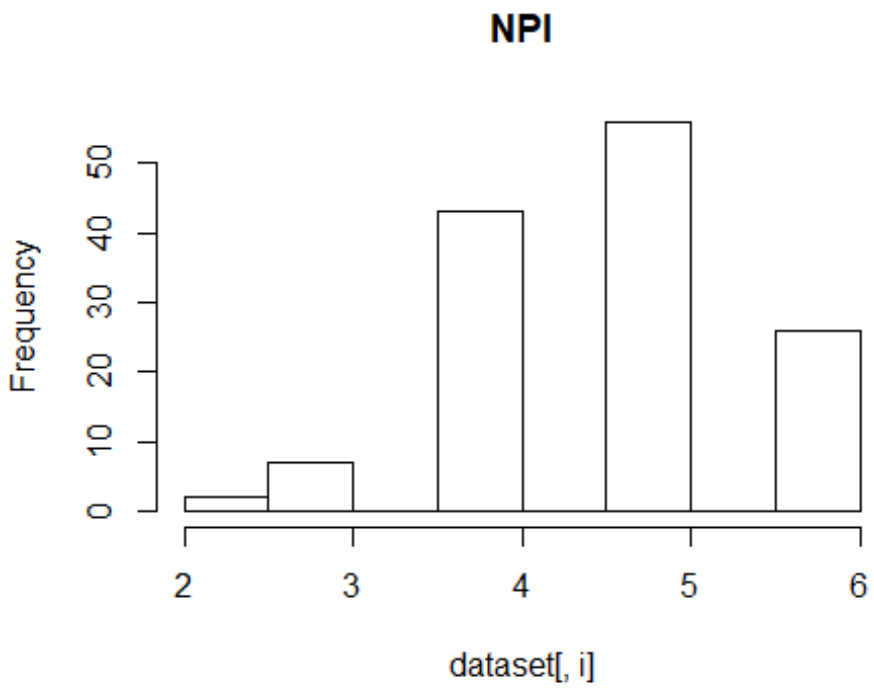
- Highest correlation is between survival groups and survival months

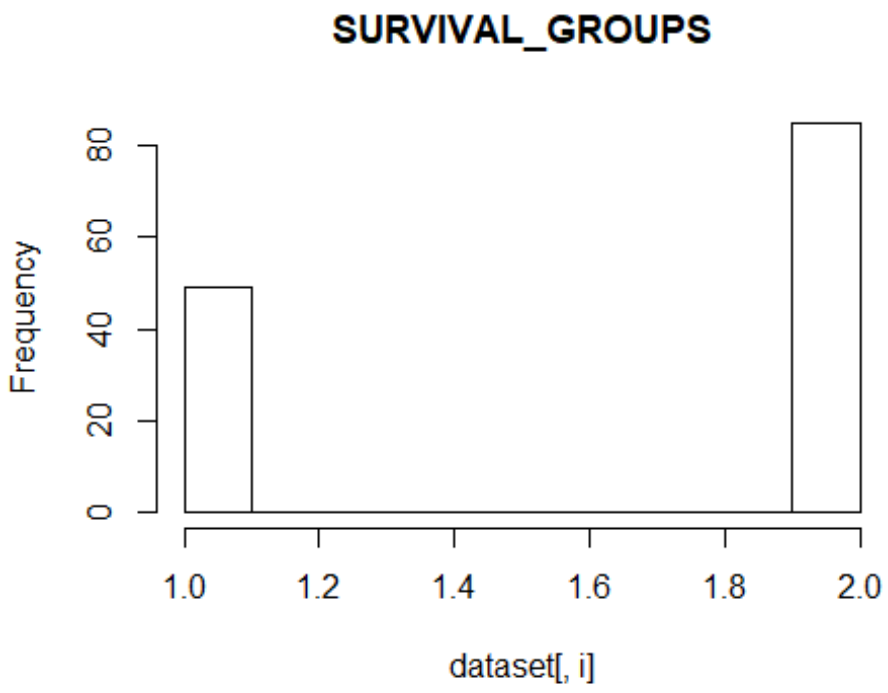
- 4.11. Uni-modal Data Visualizations of the Predictor Variables

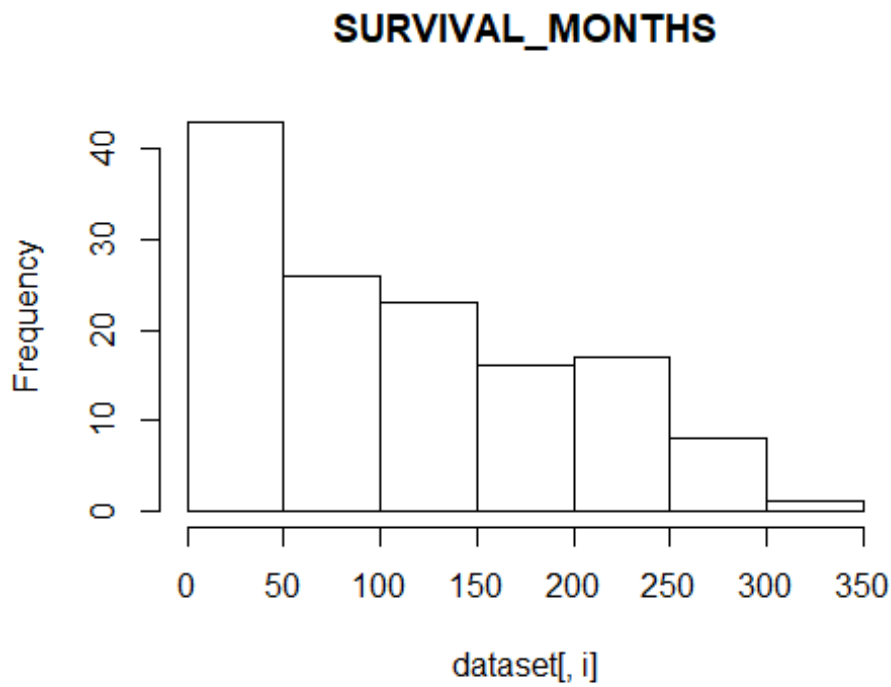
```r
#### Histograms for each attribute

#par(mfrow = c(10,3), mar=c(2,2,1.2,1.2))

for (i in 1:5) {
  hist(dataset[,i],main = names(dataset)[i])
}
```


BREAST_SURGERY
Frequency
0
20
40
60
80
1.0
1.2
1.4
1.6
1.8
2.0
dataset[, i]

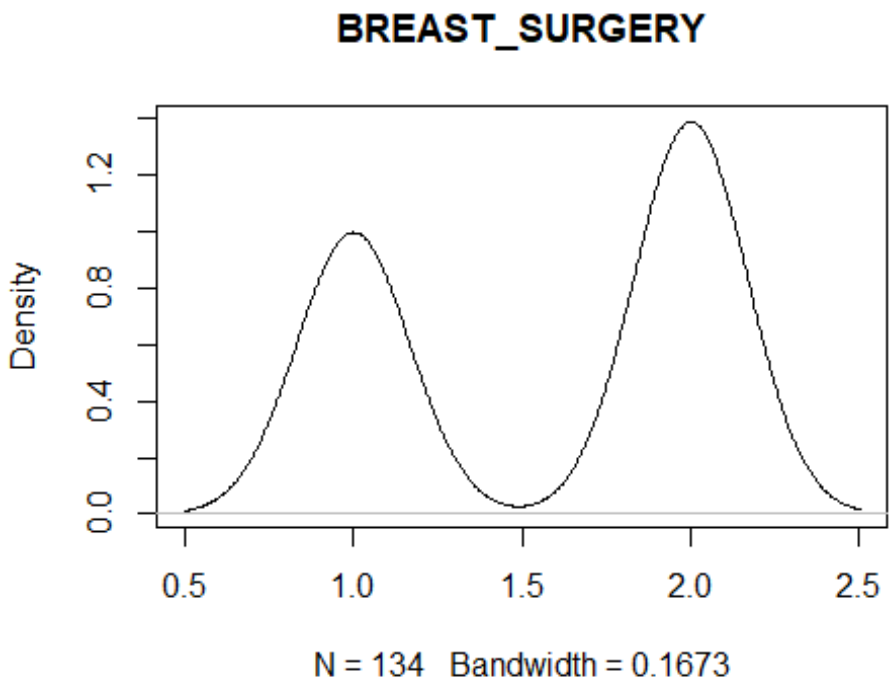

TUMOR_SIZE
Frequency
0
10
30
50
0
50
100
150
200
dataset[, i]

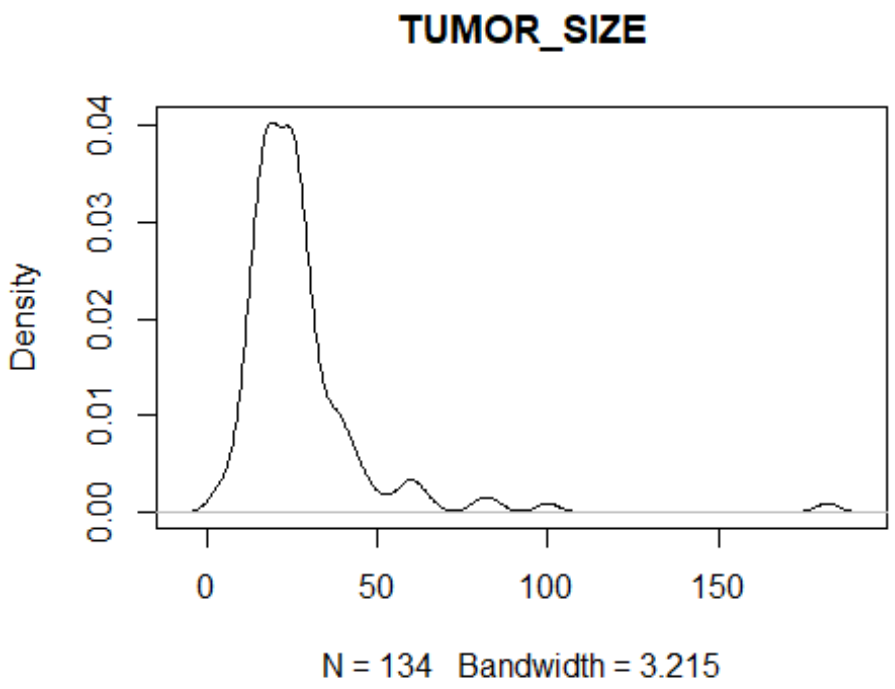

NPI
Frequency
0
10
20
30
40
50
2
3
4
5
6
dataset[, i]

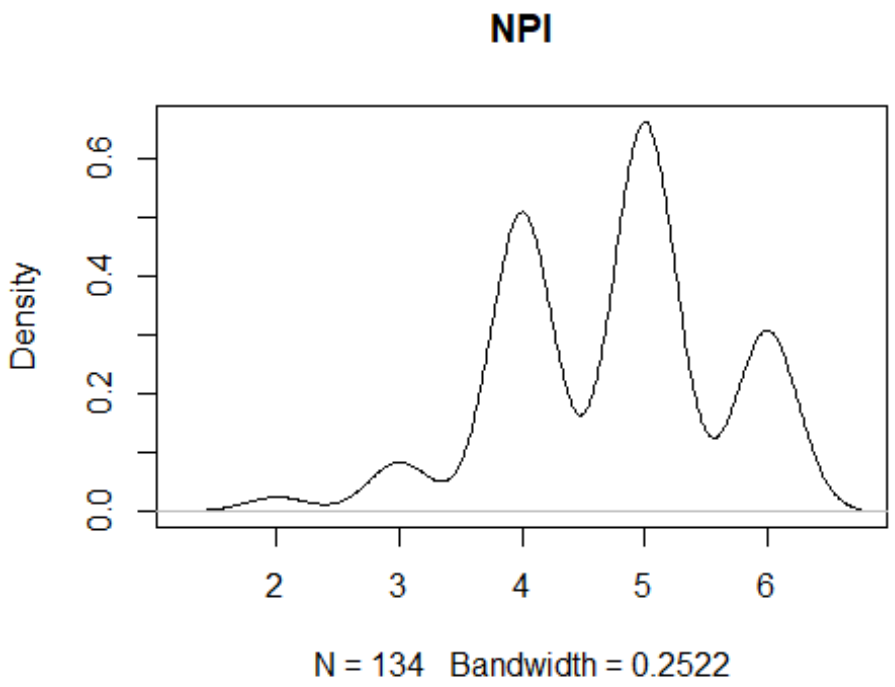

SURVIVAL_GROUPS
Frequency
0
20
40
60
80
1.0
1.2
1.4
1.6
1.8
2.0
dataset[, i]

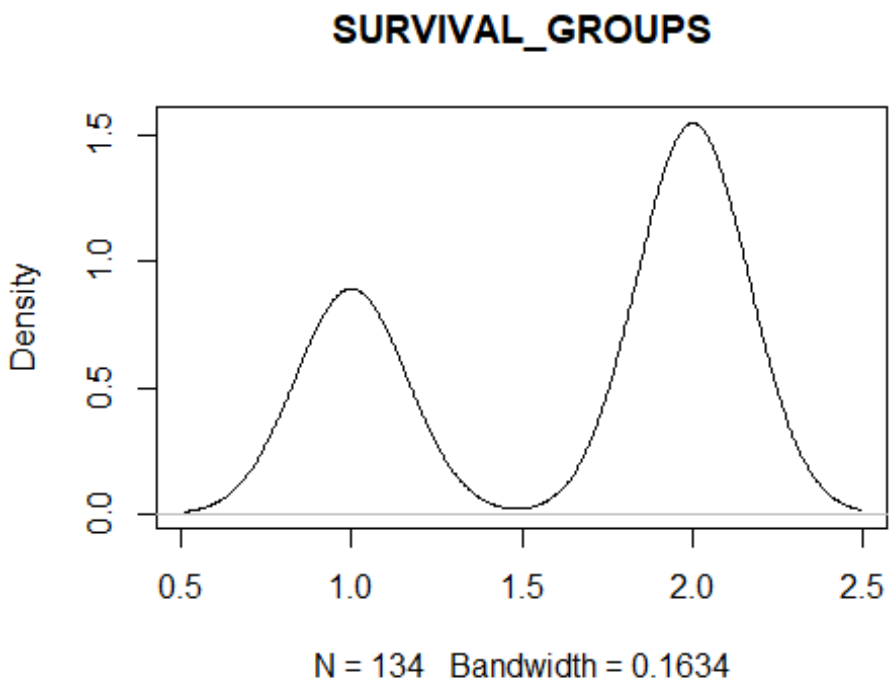


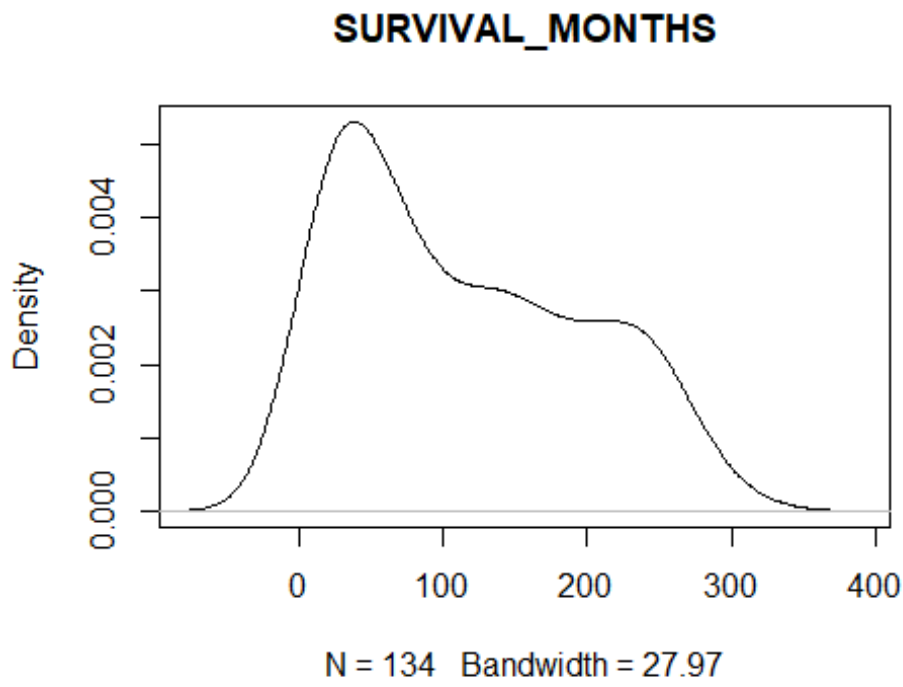

```
# density plots for each predictor attribute
#par(mfrow = c(5,5), mar=c(2,2,1.2,1.2))
for (i in 1:5) {
  plot(density(dataset[,i]),main = names(dataset)[i])
}
```


BREAST_SURGERY
Density
0.0
0.4
0.8
1.2
0.5
1.0
1.5
2.0
2.5
N = 134 Bandwidth = 0.1673


TUMOR_SIZE
Density
0.00
0.01
0.02
0.03
0.04
0
50
100
150
N = 134 Bandwidth = 3.215

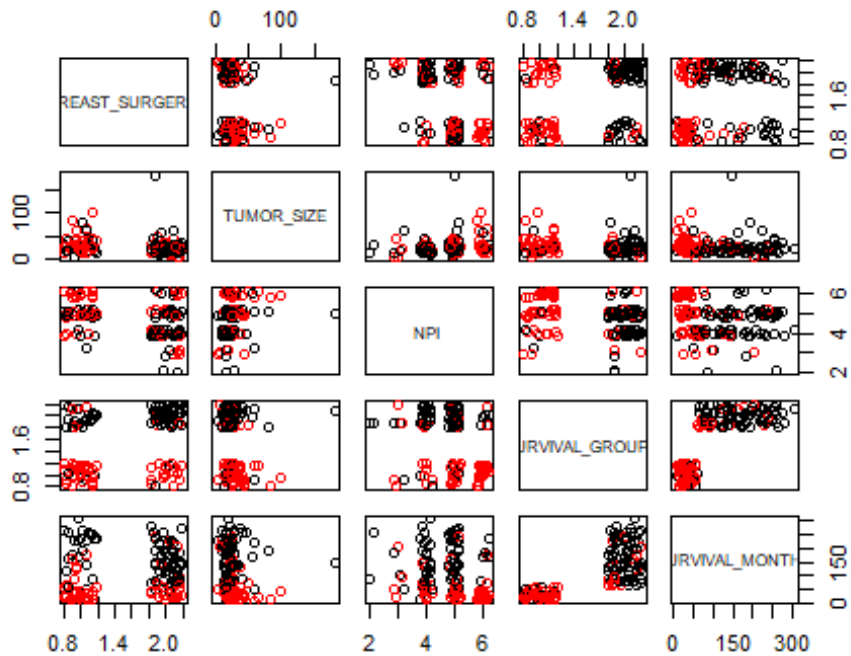

NPI
Density
0.0
0.2
0.4
0.6
2
3
4
5
6
N = 134 Bandwidth = 0.2522


SURVIVAL_GROUPS
Density
0.0
0.5
1.0
1.5
0.5
1.0
1.5
2.0
2.5
N = 134 Bandwidth = 0.1634



- 4.12. Multi-modal Data Visualizations of the Predictor Variables

```
jittered_x = sapply(dataset[,1:5], jitter)
pairs(jittered_x, names(dataset[,1:5]), col = dataset$VITAL_STATUS)
```

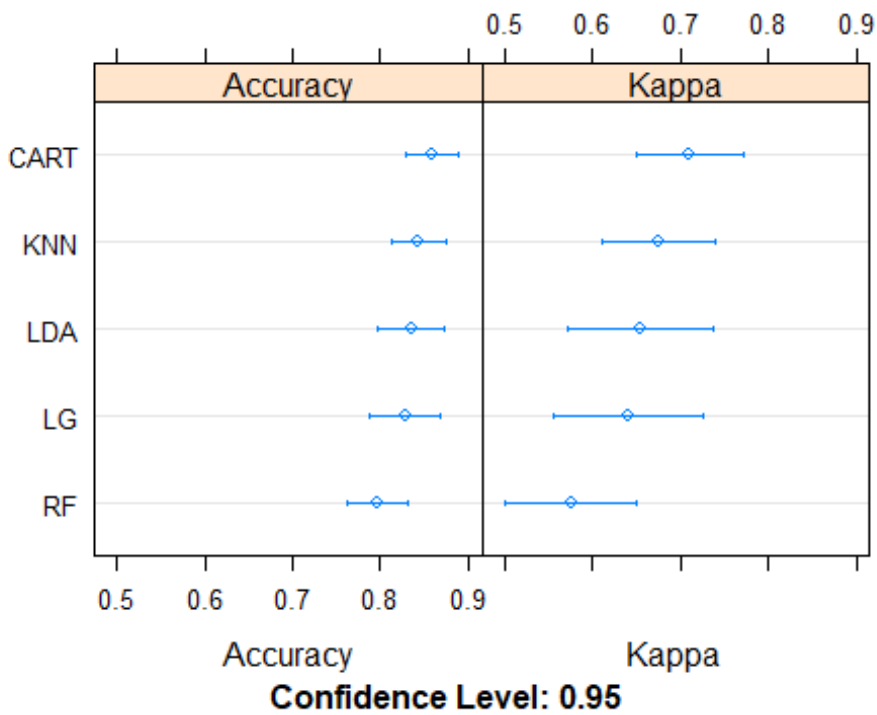


* Plot showing the relationship between the 5 clinical features and the dependent variable. * The red dots represents the patients who died from the disease while the black dots represent the patients alive

- 4.13. Evaluate 5 Machine Learning Algorithms using clinical data only
- 4.14. Linear Algorithms - Logistic Regression (LG) and Linear Discriminate Analysis (LDA)
- 4.15. Non-Linear Algorithms - K-Nearest Neighbors (KNN), Classification and Regression Trees (CART)
- 4.16. ENSEMBLE METHODS: Random Forest (RF)
- 4.17. Analysis is done using 10-fold cross validation with 3 repeats

```
trainControl = trainControl(method = "repeatedcv", number = 10, repeats = 3)
metric = "Accuracy"

#1 LG
set.seed(7)
fit.glm = train(VITAL_STATUS~., data=dataset, method="glm", metric=metric, trControl=trainControl)

#2 LDA
set.seed(7)
fit.lda = train(VITAL_STATUS~., data=dataset, method="lda", metric=metric, trControl=trainControl)

#3 KNN
set.seed(7)
fit.knn = train(VITAL_STATUS~., data=dataset, method="knn", metric=metric, trControl=trainControl)

#4 CART
set.seed(7)
fit.cart = train(VITAL_STATUS~., data=dataset, method="rpart", metric=metric, trControl=trainControl)

#5 Random Forest
set.seed(7)
fit.rf = train(VITAL_STATUS~., data=dataset, method="rf", metric=metric, trControl=trainControl)

# Compare algorithms
results = resamples(list(LG=fit.glm, LDA=fit.lda, KNN=fit.knn, CART=fit.cart, RF=fit.rf))

summary(results)

##
## Call:
## summary.resamples(object = results)
##
```

```
## Models: LG, LDA, KNN, CART, RF
## Number of resamples: 30
##
## Accuracy
##           Min.   1st Qu.    Median      Mean   3rd Qu. Max. NA's
## LG   0.6153846 0.7692308 0.8397436 0.8279304 0.8571429    1    0
## LDA  0.6428571 0.7692308 0.8461538 0.8352564 0.9065934    1    0
## KNN  0.6428571 0.7857143 0.8461538 0.8431929 0.8571429    1    0
## CART 0.7142857 0.7857143 0.8461538 0.8554249 0.9230769    1    0
## RF   0.6428571 0.7548077 0.7857143 0.7958791 0.8461538    1    0
##
## Kappa
##           Min.   1st Qu.    Median      Mean   3rd Qu. Max. NA's
## LG   0.2168675 0.5301205 0.6596452 0.6401427 0.7170833    1    0
## LDA  0.2258065 0.5301205 0.6829268 0.6541901 0.8125301    1    0
## KNN  0.2553191 0.5531915 0.6867015 0.6747001 0.7170833    1    0
## CART 0.4166667 0.5714286 0.6904762 0.7002191 0.8470588    1    0
## RF   0.2258065 0.4577546 0.5471840 0.5747909 0.6885889    1    0

dotplot(results)
```



####* Top three performing algorithms are CART, KNN and LDA, with mean accuracies of 0.858, 0.843 and 0.835 respectively * Evaluation of these 5 Machine Learning Algorithms "with transform operation" worsened their outcomes.

Tune the top three learning algorithms

- 4.18. Tuning for the Linear Discriminate Analysis (LDA) learning algorithm

```
trainControl = trainControl(method="repeatedcv", number=10, repeats=3)
metric = "Accuracy"

#LDA
set.seed(7)
grid = expand.grid(.parameter = seq(1,10, by=1))
fit.LDA_tuned = train(VITAL_STATUS~.,data = dataset, method="lda", metric=metric, tuneGrid=grid, trControl=train
Control)
fit.LDA_tuned

## Linear Discriminant Analysis
##
## 134 samples
##   5 predictor
```

```
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 120, 120, 120, 121, 120, 121, ...
## Resampling results across tuning parameters:
##
##   Accuracy   Kappa      parameter
##   0.8352564  0.6541901  1
##   0.8352564  0.6541901  1
##   0.8352564  0.6541901  1
##   0.8352564  0.6541901  1
##   0.8352564  0.6541901  1
##   0.8352564  0.6541901  1
##   0.8352564  0.6541901  1
##   0.8352564  0.6541901  1
##   0.8352564  0.6541901  1
##   0.8352564  0.6541901  1
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was parameter = 1.
```

- Tuning did not improve the LDA model

- 4.19. Tuning for the CART learning algorithm

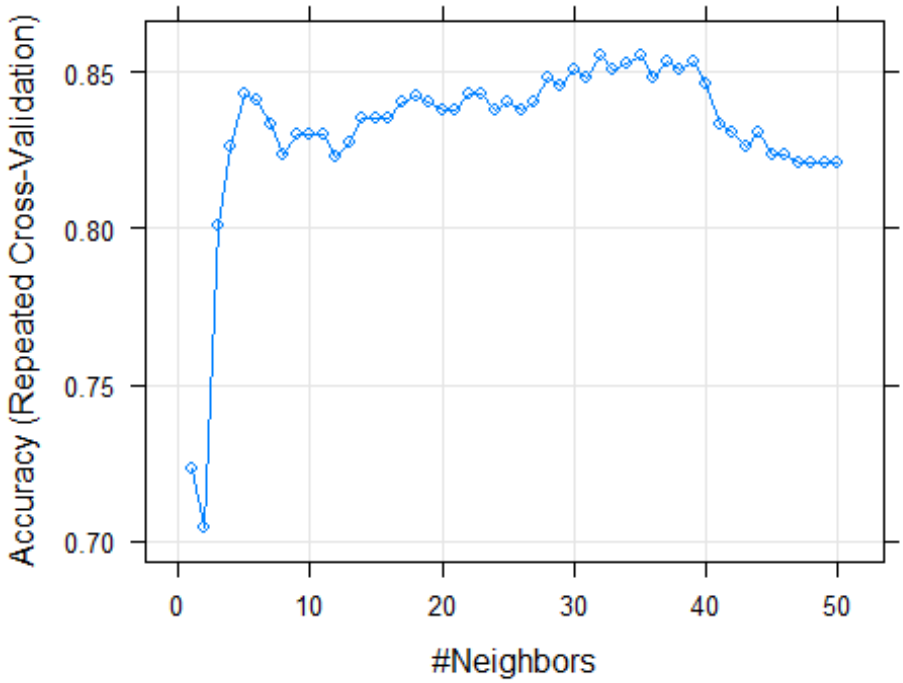
```
trainControl = trainControl(method="repeatedcv", number=10, repeats=3)
metric = "Accuracy"

#CART
set.seed(7)
grid = expand.grid(.cp = seq(1,10, by=1))
fit.CART_tuned = train(VITAL_STATUS~.,data = dataset, method="rpart", metric=metric, tuneGrid=grid, trControl=tr
ainControl)
fit.CART_tuned

## CART
##
## 134 samples
##   5 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 120, 120, 120, 121, 120, 121, ...
## Resampling results across tuning parameters:
##
##   cp  Accuracy   Kappa
##    1  0.5594322  0
##    2  0.5594322  0
##    3  0.5594322  0
##    4  0.5594322  0
##    5  0.5594322  0
##    6  0.5594322  0
##    7  0.5594322  0
##    8  0.5594322  0
##    9  0.5594322  0
##   10  0.5594322  0
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was cp = 10.
```

- Tuning did not improve the LDA model

- 4.20. Tuning for the K-Nearest Neighbor (KNN) learning algorithm

```
# KNN
set.seed(7)
grid = expand.grid(.k = seq(1,50, by=1))
fit.knn_tuned = train(VITAL_STATUS~.,data = dataset, method="knn", metric=metric, tuneGrid=grid, trControl=train
Control)
fit.knn_tuned

## k-Nearest Neighbors
##
## 134 samples
##   5 predictor
```

```
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 120, 120, 120, 121, 120, 121, ...
## Resampling results across tuning parameters:
##
##   k   Accuracy   Kappa
##    1  0.7234737  0.4315616
##    2  0.7041819  0.3959883
##    3  0.8010073  0.5880806
##    4  0.8263431  0.6405447
##    5  0.8431929  0.6747001
##    6  0.8412088  0.6717998
##    7  0.8330891  0.6532868
##    8  0.8233822  0.6326019
##    9  0.8303114  0.6461173
##   10  0.8303114  0.6454967
##   11  0.8303114  0.6454967
##   12  0.8228022  0.6293911
##   13  0.8277473  0.6402773
##   14  0.8352564  0.6556433
##   15  0.8354396  0.6556913
##   16  0.8354396  0.6546608
##   17  0.8402015  0.6652429
##   18  0.8425824  0.6699916
##   19  0.8403846  0.6659607
##   20  0.8378205  0.6608627
##   21  0.8378205  0.6611129
##   22  0.8431624  0.6730869
##   23  0.8431624  0.6730869
##   24  0.8380342  0.6626421
##   25  0.8405983  0.6690475
##   26  0.8378205  0.6622069
##   27  0.8407814  0.6690880
##   28  0.8483211  0.6862844
##   29  0.8459402  0.6820866
##   30  0.8507021  0.6926374
##   31  0.8485043  0.6872651
##   32  0.8556471  0.7023782
##   33  0.8507021  0.6928353
##   34  0.8530830  0.6979940
##   35  0.8554640  0.7031744
##   36  0.8483211  0.6888643
##   37  0.8532967  0.6995599
##   38  0.8510989  0.6947154
##   39  0.8532967  0.6995599
##   40  0.8459707  0.6854448
##   41  0.8335165  0.6624986
##   42  0.8309524  0.6570251
##   43  0.8263736  0.6486060
##   44  0.8309524  0.6570331
##   45  0.8238095  0.6435172
##   46  0.8235958  0.6430802
##   47  0.8210317  0.6379913
##   48  0.8212149  0.6384001
##   49  0.8210317  0.6379913
##   50  0.8210317  0.6379913
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 32.

plot(fit.knn_tuned)
```



* Tuning significantly improved the KNN model from an accuracy of 0.8431 to 0.8556 at k = 32

- 4.21. Finalize by building Model 1 using KNN
- 4.22. Prepare the data transform for finalizing Model 1

```
set.seed(7)
x = dataset[,1:5]
preprocessParam = preProcess(x)
x = predict(preprocessParam, x)

# prepare the validation dataset
set.seed(7)

# transform the validation dataset
validationX = predict(preprocessParam, validation[,1:5])
```

- 4.23. Make predictions and build confusion matrix for Model 1

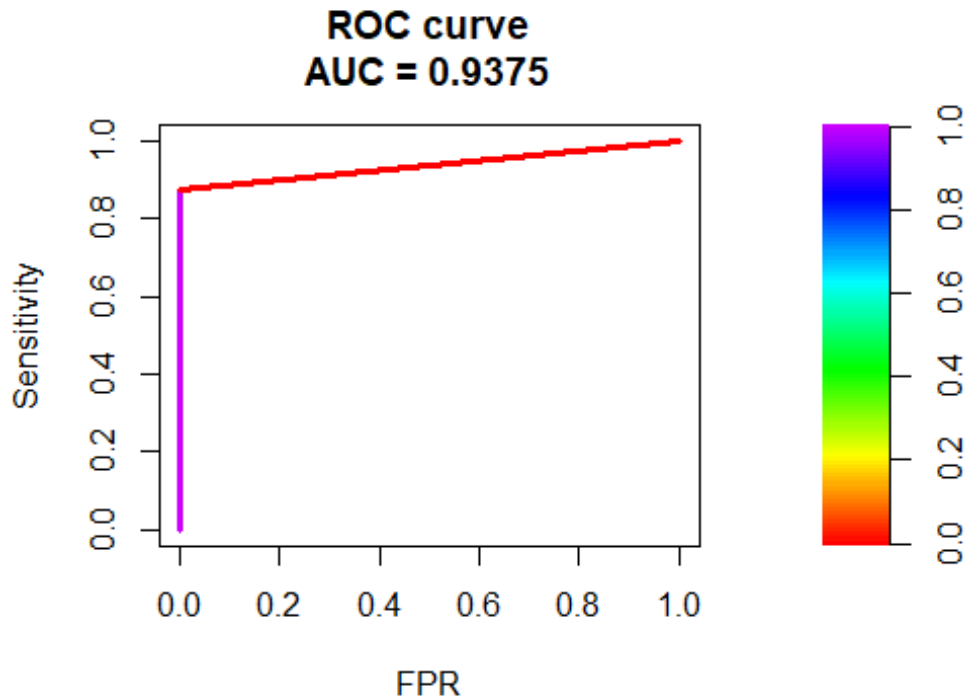
```
set.seed(7)
predictions = knn3Train(x, validationX, dataset$VITAL_STATUS, k=32, prob = FALSE)
confusionMatrix(factor(predictions), factor(validation$VITAL_STATUS))

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 32  3
##          1  0 21
##
##                Accuracy : 0.9464
##                  95% CI : (0.8513, 0.9888)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 3.093e-10
##
##                   Kappa : 0.8889
##
##  Mcnemar's Test P-Value : 0.2482
##
##             Sensitivity : 1.0000
##             Specificity : 0.8750
##          Pos Pred Value : 0.9143
##          Neg Pred Value : 1.0000
##              Prevalence : 0.5714
```

```
##          Detection Rate : 0.5714
##    Detection Prevalence : 0.6250
##       Balanced Accuracy : 0.9375
##
##        'Positive' Class : 0
##

View(TNBC_clinical)
```

- Model performed at an accuracy of 0.9463 with only 3 error calls leading to a specificity of 0.8750 but a perfect sensitivity score of 1. Kappa was very high at 0.8889.

- 4.24. Calculate AUC(ROC) for Model 1

```
library(PRROC)

## Warning: package 'PRROC' was built under R version 3.6.3

sapply(predictions, class)

##           0           1           1           1           1           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           0           1
## "character" "character" "character" "character" "character" "character"
##           0           0           0           1           1           0
## "character" "character" "character" "character" "character" "character"
##           1           0           1           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           1           0           0           0           1
## "character" "character" "character" "character" "character" "character"
##           1           0           0           0           1           0
## "character" "character" "character" "character" "character" "character"
##           0           0           1           0           1           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           1           1
## "character" "character" "character" "character" "character" "character"
##           0           1           0           1           0           1
## "character" "character" "character" "character" "character" "character"
##           0           1
## "character" "character"

sapply(validation, class)

##  BREAST_SURGERY      TUMOR_SIZE             NPI SURVIVAL_GROUPS SURVIVAL_MONTHS
##       "numeric"       "numeric"       "numeric"       "numeric"       "numeric"
##    VITAL_STATUS
##        "factor"

pred = as.numeric(as.character(predictions))
valid = as.character(as.factor(validation$VITAL_STATUS))
valid = as.numeric(as.character(validation$VITAL_STATUS))
prroc_obj = roc.curve(scores.class0 = pred, weights.class0 = valid, curve = TRUE)
plot(prroc_obj)
```



Part B: This analysis captures genetic features ONLY

- 4.25. For this analysis, remove SURVIVAL_MONTHS variable from the t_mutate object and save in TNBC_2 object
- 4.26. Convert integer values to numeric and ensure that dependent variable "VITAL_STATUS" is of the factor class

```
# remove SURVIVAL_MONTHS feature from the t_mutate object and save in the TNBC_2 object
TNBC_2 = t_mutate[-19]

for (i in 1:18) {
  TNBC_2[,i] = as.numeric(as.integer(TNBC_2[,i]))
}

TNBC_2$VITAL_STATUS = as.factor(as.integer(TNBC_2$VITAL_STATUS))

sapply(TNBC_2, class)

##         AKT1       MAP3K1         MEN1        NCOR2       PIK3CA       SHANK2
##    "numeric"    "numeric"    "numeric"    "numeric"    "numeric"    "numeric"
##       MAP2K4         AGMO         TAF1        GATA3        BIRC6       DNAH11
##    "numeric"    "numeric"    "numeric"    "numeric"    "numeric"    "numeric"
##          NF1         TP53        SYNE1       CTNNA1        FOXO3      LDLRAP1
##    "numeric"    "numeric"    "numeric"    "numeric"    "numeric"    "numeric"
## VITAL_STATUS
##     "factor"

names(TNBC_2)

##  [1] "AKT1"          "MAP3K1"        "MEN1"          "NCOR2"         "PIK3CA"
##  [6] "SHANK2"        "MAP2K4"        "AGMO"          "TAF1"          "GATA3"
## [11] "BIRC6"         "DNAH11"        "NF1"           "TP53"          "SYNE1"
## [16] "CTNNA1"        "FOXO3"         "LDLRAP1"       "VITAL_STATUS"
```

- 4.27. Perform feature selection. Correlation analyses findings in Section 1 already pointed to the fact that most of the mutation features have no association with the outcome of the dependent variable.
- 4.28. Feature Selection using once again using Boruta, mlbench and Caret packages respectively

```
# Feature Selection operation using the boruta function
set.seed(7)
boruta2 = Boruta(TNBC_2$VITAL_STATUS~.,data = TNBC_2, doTrace = 4, maxRuns = 100)
```

```
##  1. run of importance source...

## Assigned hit to 3 attributes out of 18 undecided.

##  2. run of importance source...

## Assigned hit to 5 attributes out of 18 undecided.

##  3. run of importance source...

## Assigned hit to 6 attributes out of 18 undecided.

##  4. run of importance source...

## Assigned hit to 2 attributes out of 18 undecided.

##  5. run of importance source...

## Assigned hit to 4 attributes out of 18 undecided.

##  6. run of importance source...

## Assigned hit to 5 attributes out of 18 undecided.

##  7. run of importance source...

## Assigned hit to 7 attributes out of 18 undecided.

##  8. run of importance source...

## Assigned hit to 5 attributes out of 18 undecided.

##  9. run of importance source...

## Assigned hit to 6 attributes out of 18 undecided.

##  10. run of importance source...

## Assigned hit to 4 attributes out of 18 undecided.

##  11. run of importance source...

## Assigned hit to 4 attributes out of 18 undecided.

## After 11 iterations, +1.5 secs:

##  confirmed 2 attributes: MAP3K1, SHANK2;

##  rejected 11 attributes: AGMO, BIRC6, CTNNA1, FOXO3, GATA3 and 6 more;

##  still have 5 attributes left.

##  12. run of importance source...

## Assigned hit to 3 attributes out of 5 undecided.

##  13. run of importance source...

## Assigned hit to 5 attributes out of 5 undecided.

##  14. run of importance source...

## Assigned hit to 3 attributes out of 5 undecided.

##  15. run of importance source...

## Assigned hit to 1 attribute out of 5 undecided.

##  16. run of importance source...

## Assigned hit to 1 attribute out of 5 undecided.
```

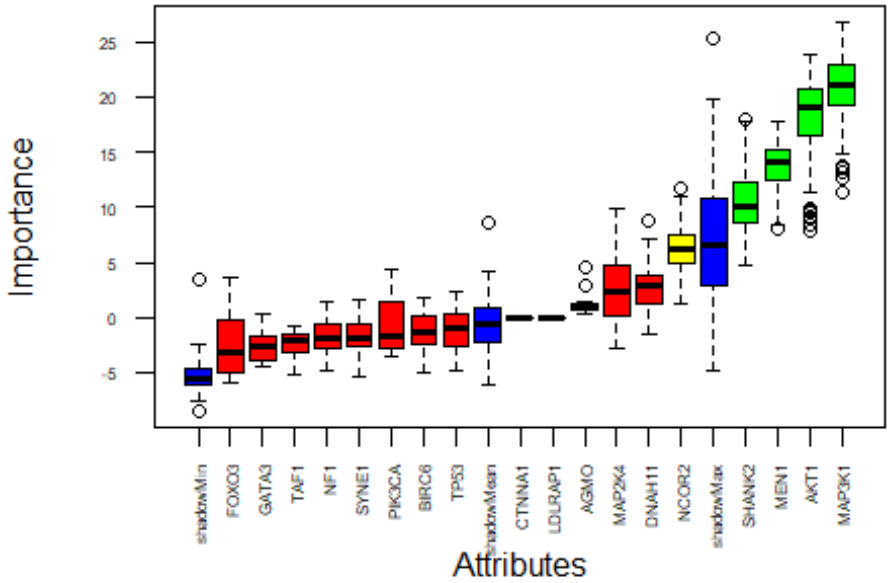
```
##  17. run of importance source...

## Assigned hit to 4 attributes out of 5 undecided.

##  18. run of importance source...

## Assigned hit to 4 attributes out of 5 undecided.

##  19. run of importance source...

## Assigned hit to 3 attributes out of 5 undecided.

## After 19 iterations, +2 secs:

##  confirmed 1 attribute: AKT1;

##  still have 4 attributes left.

##  20. run of importance source...

## Assigned hit to 1 attribute out of 4 undecided.

##  21. run of importance source...

## Assigned hit to 2 attributes out of 4 undecided.

##  22. run of importance source...

## None of undecided attributes scored a hit.

##  23. run of importance source...

## Assigned hit to 1 attribute out of 4 undecided.

##  24. run of importance source...

## Assigned hit to 2 attributes out of 4 undecided.

##  25. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.

##  26. run of importance source...

## Assigned hit to 2 attributes out of 4 undecided.

##  27. run of importance source...

## None of undecided attributes scored a hit.

##  28. run of importance source...

## Assigned hit to 1 attribute out of 4 undecided.

##  29. run of importance source...

## Assigned hit to 1 attribute out of 4 undecided.

##  30. run of importance source...

## Assigned hit to 1 attribute out of 4 undecided.

##  31. run of importance source...

## Assigned hit to 2 attributes out of 4 undecided.

## After 31 iterations, +2.8 secs:

##  confirmed 1 attribute: MEN1;

##  rejected 2 attributes: DNAH11, MAP2K4;
```

```
##  still have 1 attribute left.

##  32. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  33. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  34. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  35. run of importance source...

## None of undecided attributes scored a hit.

##  36. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  37. run of importance source...

## None of undecided attributes scored a hit.

##  38. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  39. run of importance source...

## None of undecided attributes scored a hit.

##  40. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  41. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  42. run of importance source...

## None of undecided attributes scored a hit.

##  43. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  44. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  45. run of importance source...

## None of undecided attributes scored a hit.

##  46. run of importance source...

## None of undecided attributes scored a hit.

##  47. run of importance source...

## None of undecided attributes scored a hit.

##  48. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  49. run of importance source...
```

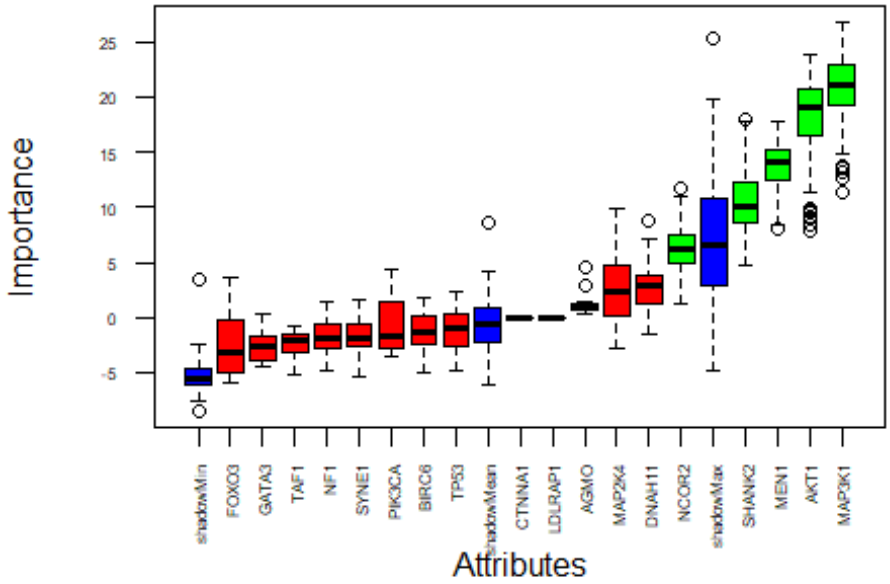
```
## None of undecided attributes scored a hit.

##  50. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  51. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  52. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  53. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  54. run of importance source...

## None of undecided attributes scored a hit.

##  55. run of importance source...

## None of undecided attributes scored a hit.

##  56. run of importance source...

## None of undecided attributes scored a hit.

##  57. run of importance source...

## None of undecided attributes scored a hit.

##  58. run of importance source...

## None of undecided attributes scored a hit.

##  59. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  60. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  61. run of importance source...

## None of undecided attributes scored a hit.

##  62. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  63. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  64. run of importance source...

## None of undecided attributes scored a hit.

##  65. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  66. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  67. run of importance source...
```

```
## None of undecided attributes scored a hit.

##  68. run of importance source...

## None of undecided attributes scored a hit.

##  69. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  70. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  71. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  72. run of importance source...

## None of undecided attributes scored a hit.

##  73. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  74. run of importance source...

## None of undecided attributes scored a hit.

##  75. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  76. run of importance source...

## None of undecided attributes scored a hit.

##  77. run of importance source...

## None of undecided attributes scored a hit.

##  78. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  79. run of importance source...

## None of undecided attributes scored a hit.

##  80. run of importance source...

## None of undecided attributes scored a hit.

##  81. run of importance source...

## None of undecided attributes scored a hit.

##  82. run of importance source...

## None of undecided attributes scored a hit.

##  83. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  84. run of importance source...

## None of undecided attributes scored a hit.

##  85. run of importance source...
```

```
## None of undecided attributes scored a hit.

##  86. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  87. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  88. run of importance source...

## None of undecided attributes scored a hit.

##  89. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  90. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  91. run of importance source...

## None of undecided attributes scored a hit.

##  92. run of importance source...

## None of undecided attributes scored a hit.

##  93. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  94. run of importance source...

## None of undecided attributes scored a hit.

##  95. run of importance source...

## None of undecided attributes scored a hit.

##  96. run of importance source...

## None of undecided attributes scored a hit.

##  97. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  98. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  99. run of importance source...

## None of undecided attributes scored a hit.

plot(boruta2, las = 2, cex.axis = 0.5)
```




```
# reclassify features that were designated as tentative
bor2 = TentativeRoughFix(boruta2)
plot(bor2, las = 2, cex.axis = 0.5)
```




```
attStats(boruta2)
```

```
##            meanImp  medianImp     minImp     maxImp   normHits  decision
## AKT1    17.9188615 18.9636993  7.8801865 23.7668864 0.96969697 Confirmed
## MAP3K1  20.8355911 21.0551587 11.3498043 26.7741660 0.98989899 Confirmed
## MEN1    13.7539941 14.1581585  8.0313042 17.8382329 0.84848485 Confirmed
## NCOR2    6.3168929  6.2069631  1.3251337 11.7613976 0.50505051 Tentative
## PIK3CA  -0.6191905 -1.6353639 -3.5499042  4.3707642 0.00000000  Rejected
## SHANK2  10.6660232 10.0700849  4.8142449 17.9537847 0.71717172 Confirmed
## MAP2K4   2.6983386  2.3383209 -2.7249150  9.9540623 0.06060606  Rejected
## AGMO     1.3255964  0.9796938  0.2687342  4.5208907 0.00000000  Rejected
## TAF1    -2.4437595 -2.0843589 -5.1875873 -0.7780414 0.00000000  Rejected
## GATA3   -2.5780892 -2.5583918 -4.4554781  0.4093262 0.00000000  Rejected
## BIRC6   -1.2008749 -1.4101927 -5.0893223  1.7584991 0.00000000  Rejected
## DNAH11   2.9478000  2.8022284 -1.5427801  8.7575446 0.06060606  Rejected
## NF1     -1.7335133 -1.9513663 -4.7685990  1.3588660 0.00000000  Rejected
## TP53    -1.0925369 -0.9755629 -4.8244629  2.3227284 0.00000000  Rejected
## SYNE1   -1.6834784 -1.9151533 -5.3652120  1.6691511 0.00000000  Rejected
## CTNNA1   0.0000000  0.0000000  0.0000000  0.0000000 0.00000000  Rejected
## FOXO3   -2.3626614 -3.1702467 -5.9497583  3.5692188 0.00000000  Rejected
## LDLRAP1  0.0000000  0.0000000  0.0000000  0.0000000 0.00000000  Rejected
```

```
print(bor2)
```

```
## Boruta performed 99 iterations in 5.705732 secs.
## Tentatives roughfixed over the last 99 iterations.
##  5 attributes confirmed important: AKT1, MAP3K1, MEN1, NCOR2, SHANK2;
##  13 attributes confirmed unimportant: AGMO, BIRC6, CTNNA1, DNAH11,
## FOXO3 and 8 more;
```

```
# create object and limit features to only the 5 confirmed features from the boruta analysis and the dependent v
ariable

col_order3 = c("MAP3K1","AKT1","MEN1","SHANK2","NCOR2","VITAL_STATUS")
View(TNBC_2)
TNBC_2 = TNBC_2[, col_order3]
names(TNBC_2)
```

```
## [1] "MAP3K1"       "AKT1"         "MEN1"         "SHANK2"       "NCOR2"
## [6] "VITAL_STATUS"
```

```
str(TNBC_2)
```

```
## 'data.frame':    190 obs. of  6 variables:
##  $ MAP3K1      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ AKT1        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ MEN1        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ SHANK2      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ NCOR2       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ VITAL_STATUS: Factor w/ 2 levels "0","1": 2 1 2 2 2 1 1 1 1 2 ...
```

- 4.29. Split the TNBC_2 genetic features object into training and validation groups (objects)
- 4.30. Split out validation dataset
- 4.31. Do this by creating a list of 70% of the rows in the original dataset to serve as training set

```
#Living
Living2 = TNBC_2[TNBC_2$VITAL_STATUS == 0,]

#Died of Disease
Deceased2 = TNBC_2[TNBC_2$VITAL_STATUS == 1,]

set.seed(7)
validationIndex2 = createDataPartition(TNBC_2$VITAL_STATUS, p=0.70, list = FALSE)

#select 30% of the data for validation

validation2 = TNBC_2[-validationIndex2,]

#use the remaining 70% of data to train the model
dataset2 = TNBC_2[validationIndex2,]

# Check that the distribution of the dependent variable is similar in train and test sets
prop.table(table(TNBC_2$VITAL_STATUS))
```

```
##
##         0         1
## 0.5631579 0.4368421
```

```r
prop.table(table(dataset2$VITAL_STATUS))

##
##         0         1
## 0.5597015 0.4402985

prop.table(table(validation2$VITAL_STATUS))

##
##         0         1
## 0.5714286 0.4285714
```

- The figures show that the ratio of the “Living”:“Died of Disease” was proportionally split

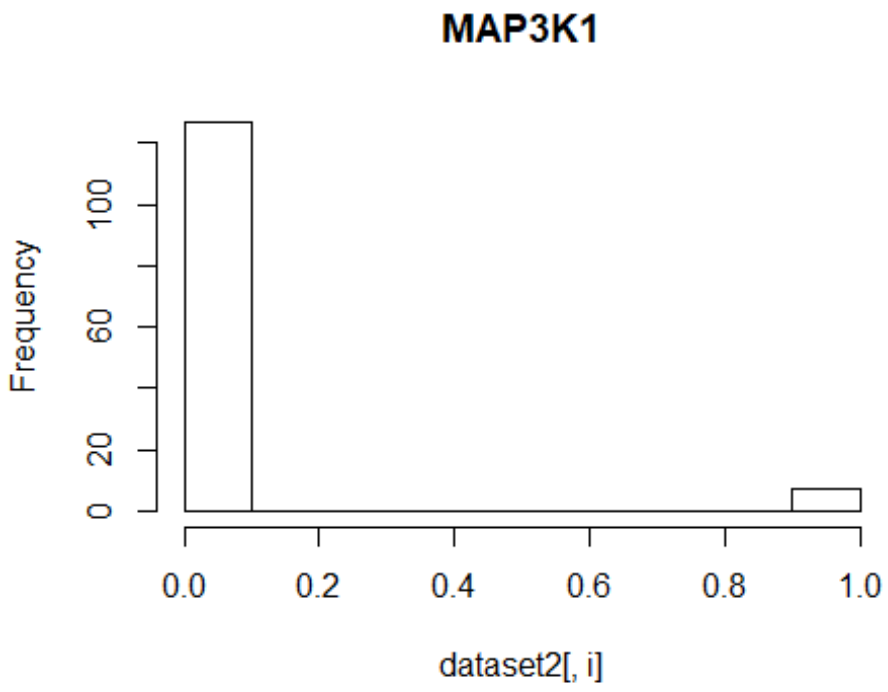
- 4.32. Analyze the data using descriptive statistics

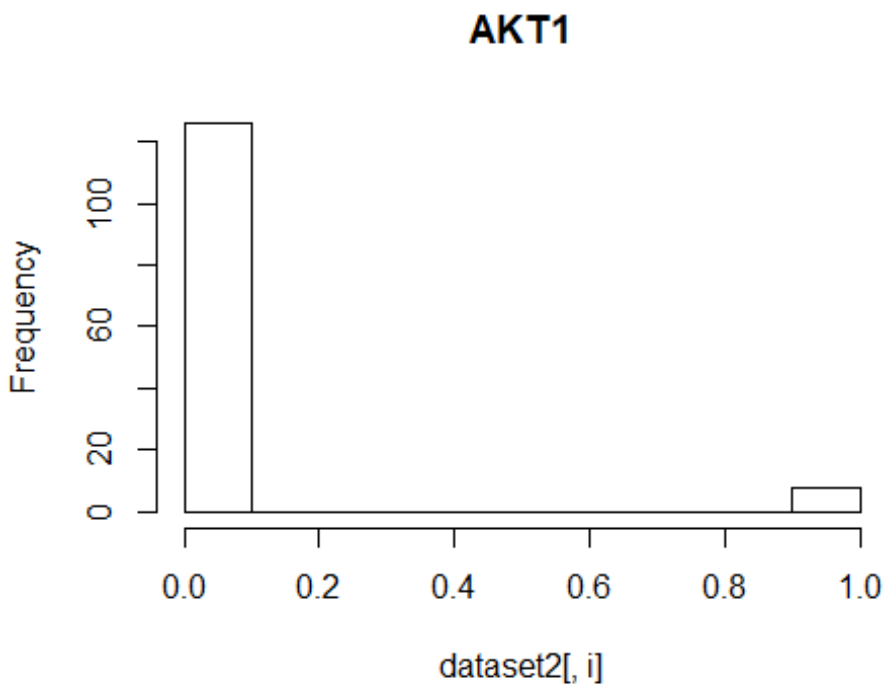
```r
# Dimension of the datasets
dim(dataset2)

## [1] 134   6

dim(validation2)

## [1] 56  6

#class distribution of training set
cbind(freq = table(dataset2$VITAL_STATUS), percentage = prop.table(table(dataset2$VITAL_STATUS))*100)

##   freq percentage
## 0   75   55.97015
## 1   59   44.02985

#summarize correlations between input variables to ascertain if there is the potential for col-linearity
cor(dataset2[,1:5])

##             MAP3K1        AKT1        MEN1      SHANK2       NCOR2
## MAP3K1  1.00000000 -0.05915710  0.24770137 -0.05082973 -0.05082973
## AKT1   -0.05915710  1.00000000 -0.03101614 -0.05455447 -0.05455447
## MEN1    0.24770137 -0.03101614  1.00000000 -0.02665009 -0.02665009
## SHANK2 -0.05082973 -0.05455447 -0.02665009  1.00000000 -0.04687500
## NCOR2  -0.05082973 -0.05455447 -0.02665009 -0.04687500  1.00000000
```

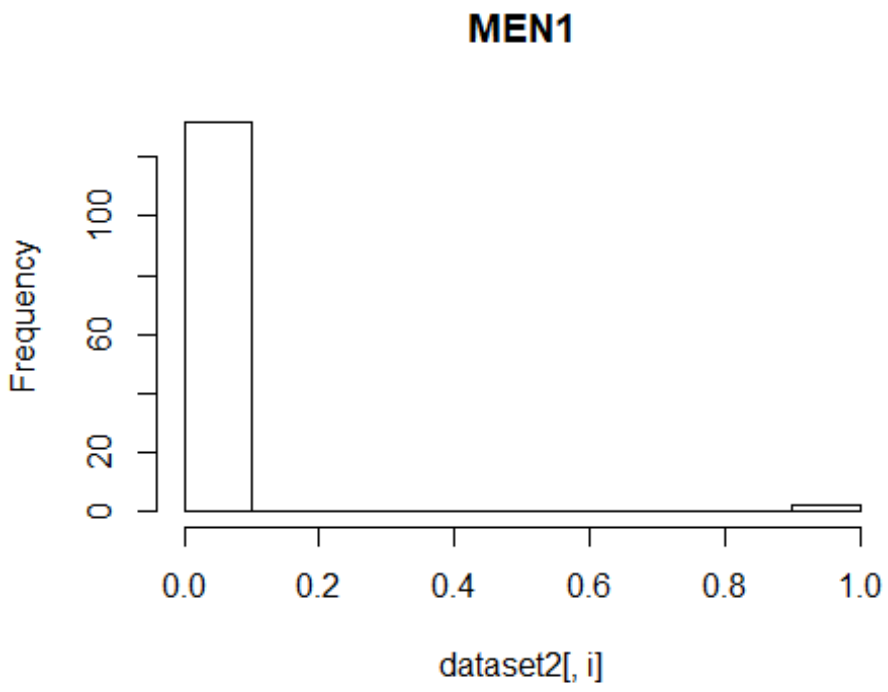
- Highest correlation is between MAP3K1 and MEN1 genes

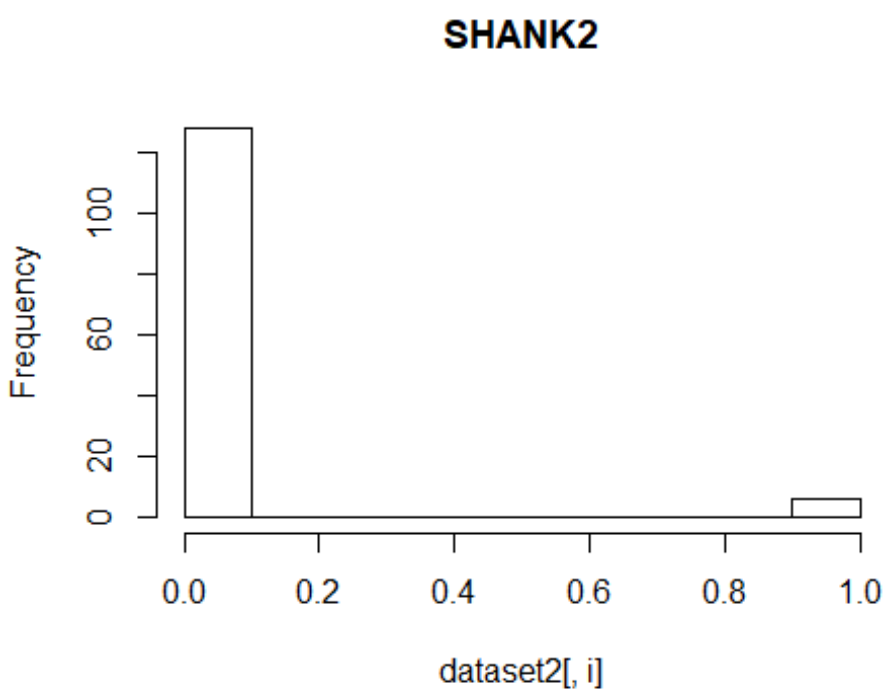
- 4.33. Uni-modal Data Visualizations of the Predictor Variables

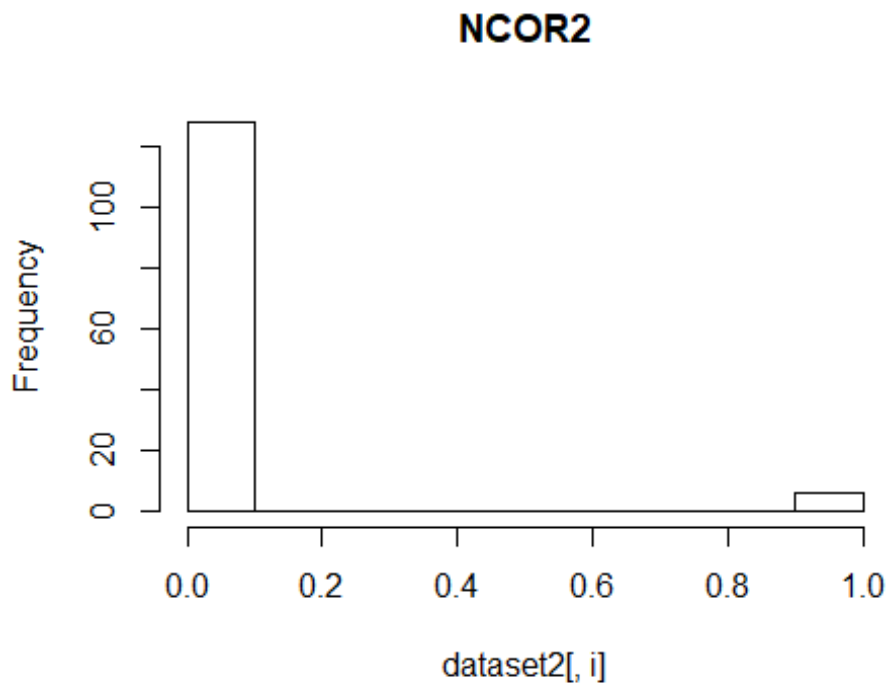
```r
#### Histograms for each attribute
#par(mfrow = c(10,3), mar=c(2,2,1.2,1.2))

for (i in 1:5) {
  hist(dataset2[,i],main = names(dataset2)[i])
}
```


MAP3K1
Frequency
0
20
60
100
0.0
0.2
0.4
0.6
0.8
1.0
dataset2[, i]

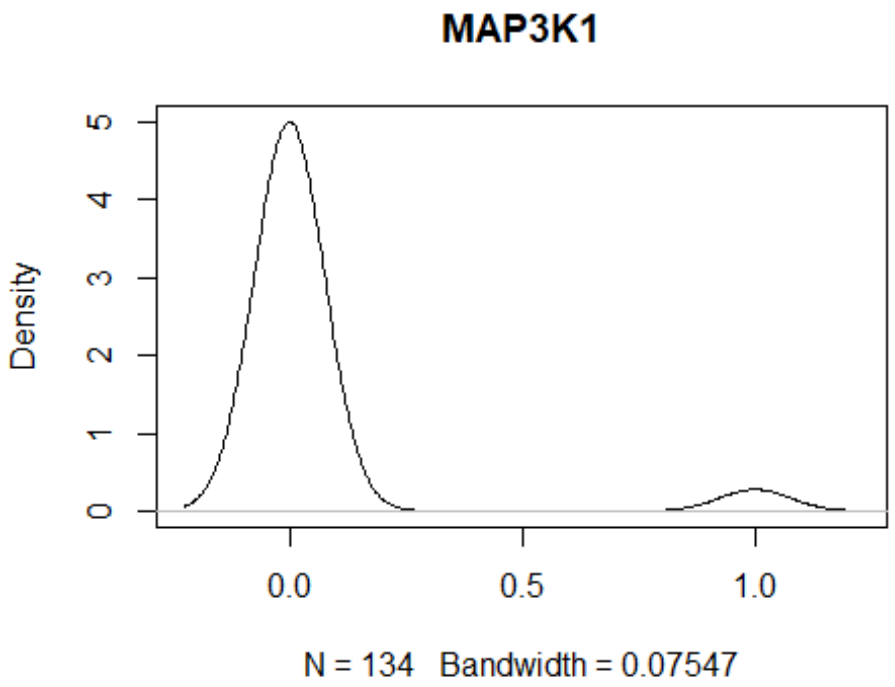

AKT1
Frequency
0
20
60
100
0.0
0.2
0.4
0.6
0.8
1.0
dataset2[, i]

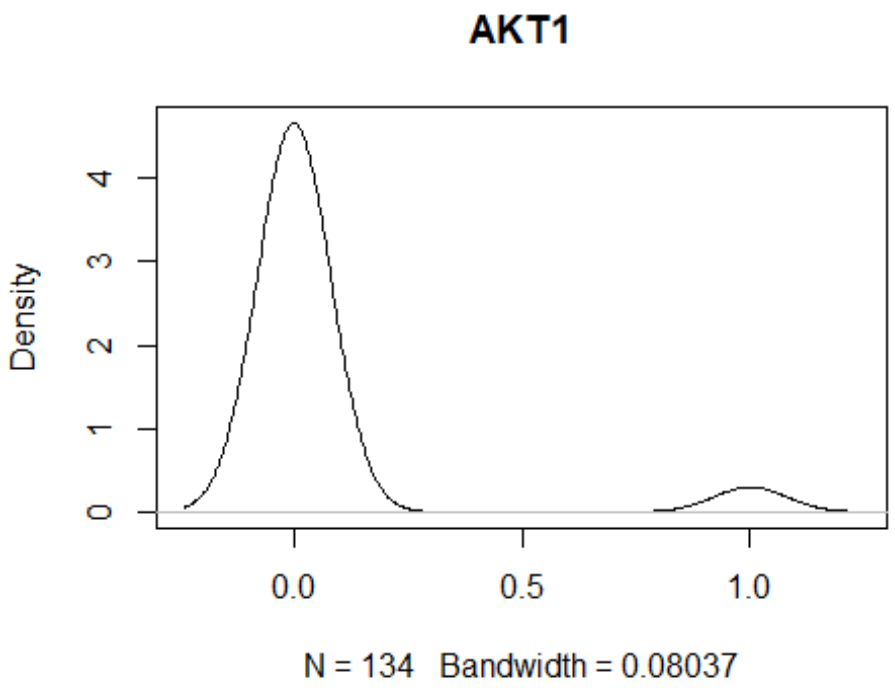

MEN1
Frequency
0
20
60
100
0.0
0.2
0.4
0.6
0.8
1.0
dataset2[, i]

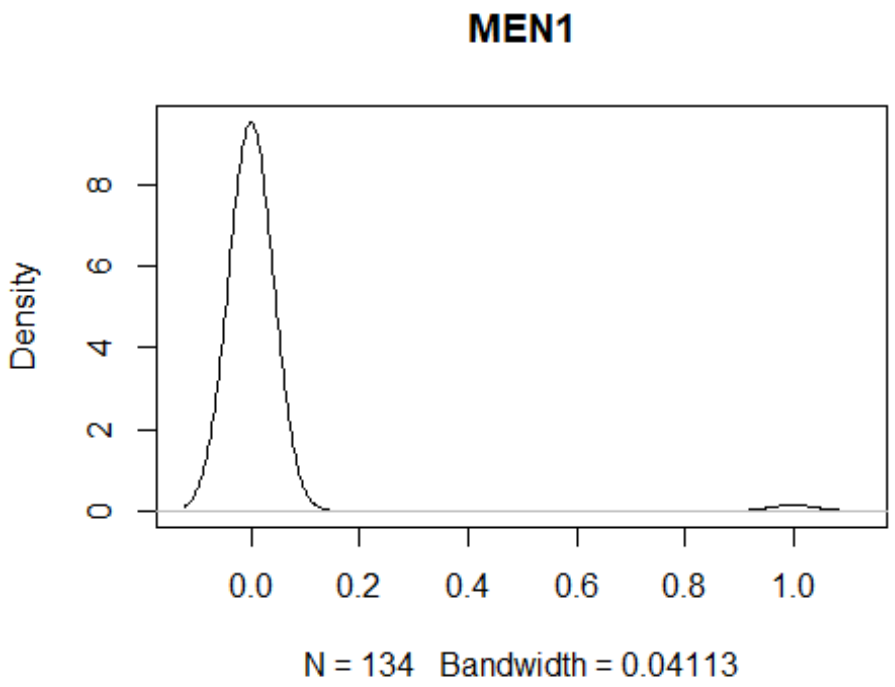

SHANK2
Frequency
0
20
60
100
0.0
0.2
0.4
0.6
0.8
1.0
dataset2[, i]

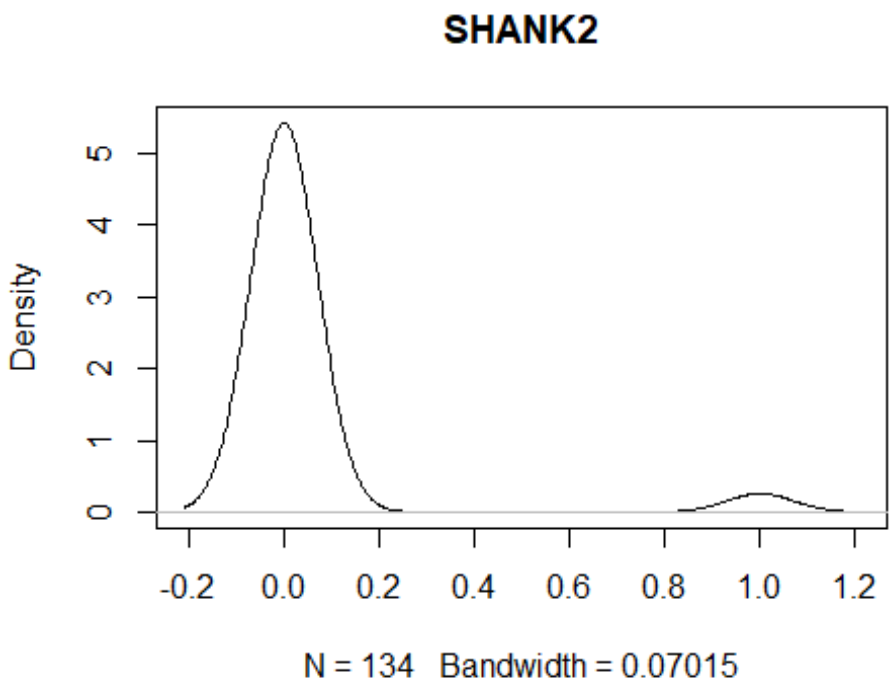


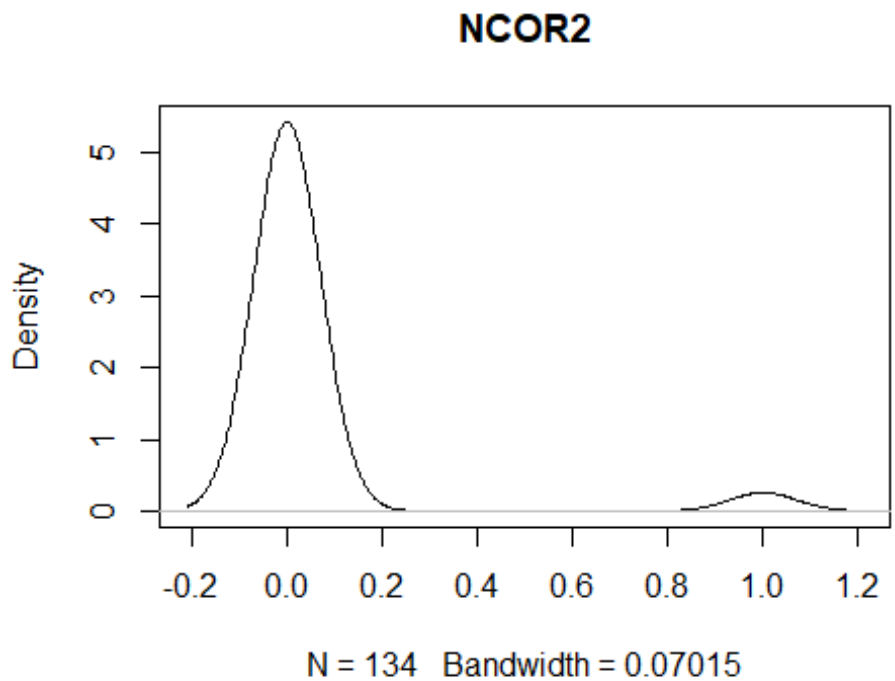

```
# density plots for each predictor attribute
#par(mfrow = c(5,5), mar=c(2,2,1.2,1.2))
for (i in 1:5) {
  plot(density(dataset2[,i]),main = names(dataset2)[i])
}
```


MAP3K1
Density
0
1
2
3
4
5
0.0
0.5
1.0
N = 134 Bandwidth = 0.07547


AKT1
Density
0
1
2
3
4
0.0
0.5
1.0
N = 134 Bandwidth = 0.08037

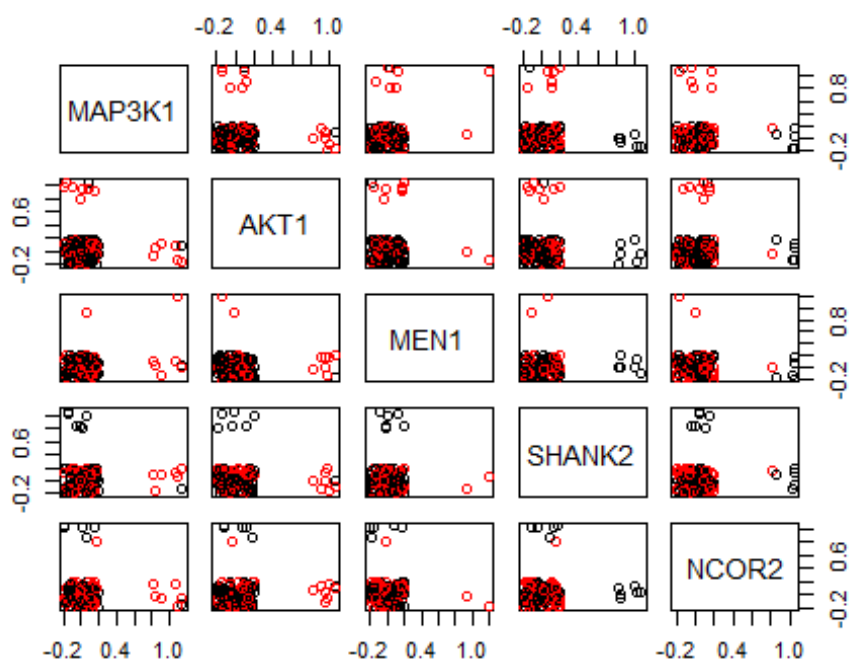

MEN1
Density
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
N = 134 Bandwidth = 0.04113


SHANK2
Density
0
1
2
3
4
5
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
N = 134 Bandwidth = 0.07015



- 4.34. Multi-modal Data Visualizations of the Predictor Variables

```
jittered_x = sapply(dataset2[,1:5], jitter)
pairs(jittered_x, names(dataset2[,1:5]), col = dataset2$VITAL_STATUS)
```



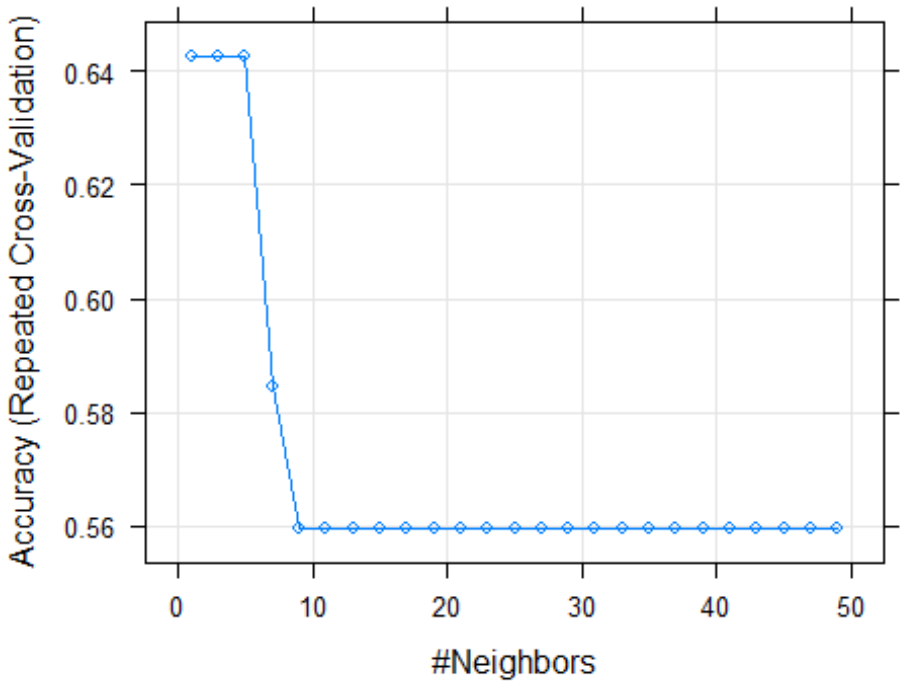
- 4.35. Apply the KNN Learning Algorithms on the mutation data
- 4.36. Analysis is done using 10-fold cross validation with 3 repeats

```
trainControl2 = trainControl(method = "repeatedcv", number = 10, repeats = 3)
metric2 = "Accuracy"

#KNN
set.seed(7)
fit.knn2 = train(VITAL_STATUS~., data=dataset2, method="knn", metric=metric2, trControl=trainControl2)

fit.knn2

## k-Nearest Neighbors
##
## 134 samples
##   5 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 120, 120, 120, 121, 120, 121, ...
## Resampling results across tuning parameters:
##
##   k  Accuracy   Kappa
##   5  0.6427045  0.20701582
##   7  0.5844017  0.06604211
##   9  0.5594322  0.00000000
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 5.
```

- 4.37. Tuning for the K-Nearest Neighbor (KNN) learning algorithm

```
# KNN
set.seed(7)
grid2 = expand.grid(.k = seq(1,50, by=2))
fit.knn_tuned2 = train(VITAL_STATUS~.,data = dataset2, method="knn", metric=metric2, tuneGrid=grid2,
                       trControl=trainControl2)
fit.knn_tuned2

## k-Nearest Neighbors
##
## 134 samples
##   5 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 120, 120, 120, 121, 120, 121, ...
## Resampling results across tuning parameters:
##
##   k   Accuracy   Kappa
##    1  0.6427045  0.20864108
##    3  0.6427045  0.20864108
##    5  0.6427045  0.20701582
##    7  0.5844017  0.06604211
##    9  0.5594322  0.00000000
##   11  0.5594322  0.00000000
##   13  0.5594322  0.00000000
##   15  0.5594322  0.00000000
##   17  0.5594322  0.00000000
##   19  0.5594322  0.00000000
##   21  0.5594322  0.00000000
##   23  0.5594322  0.00000000
##   25  0.5594322  0.00000000
##   27  0.5594322  0.00000000
##   29  0.5594322  0.00000000
##   31  0.5594322  0.00000000
##   33  0.5594322  0.00000000
##   35  0.5594322  0.00000000
##   37  0.5594322  0.00000000
##   39  0.5594322  0.00000000
##   41  0.5594322  0.00000000
##   43  0.5594322  0.00000000
##   45  0.5594322  0.00000000
##   47  0.5594322  0.00000000
##   49  0.5594322  0.00000000
##
```

```
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 5.

plot(fit.knn_tuned2)
```



* Accuracy after tuning for KNN showed no improvement

- 4.38. Finalize by building Model 2 using KNN
- 4.39. Prepare the data transform for finalizing Model 2

```
set.seed(7)
y = dataset2[,1:5]
preprocessParam2 = preProcess(y)
y = predict(preprocessParam2, y)

# prepare the validation dataset
set.seed(7)

# transform the validation dataset
validationy = predict(preprocessParam2, validation2[,1:5])
```

- 4.40. Make predictions and build confusion matrix for Model 2

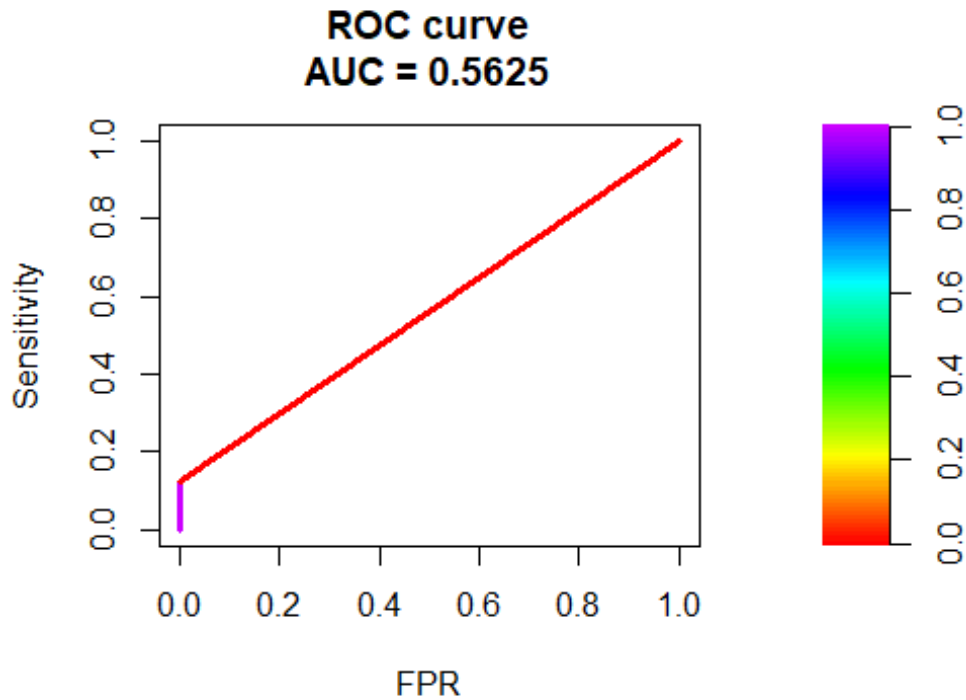
```
set.seed(7)
predictions2 = knn3Train(y, validationy, dataset2$VITAL_STATUS, k=5, prob = FALSE)
confusionMatrix(factor(predictions2), factor(validation2$VITAL_STATUS))

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 32 21
##          1  0  3
##
##                Accuracy : 0.625
##                  95% CI : (0.4855, 0.7508)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 0.2511
##
##                   Kappa : 0.1404
##
##  Mcnemar's Test P-Value : 1.275e-05
```

```
##
##             Sensitivity : 1.0000
##             Specificity : 0.1250
##          Pos Pred Value : 0.6038
##          Neg Pred Value : 1.0000
##              Prevalence : 0.5714
##          Detection Rate : 0.5714
##    Detection Prevalence : 0.9464
##       Balanced Accuracy : 0.5625
##
##        'Positive' Class : 0
##
```

- Model performed at an accuracy of 0.625 with 21 error calls leading to a very poor specificity of 0.1250 but a perfect sensitivity score of 1. Kappa was very poor at 0.1404

- 4.41. Calculate AUC(ROC) for Model 2

```
library(PRROC)
sapply(predictions2, class)

##           0           0           0           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           0           0
## "character" "character" "character" "character" "character" "character"
##           1           0           0           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           1           1           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           0
## "character" "character"

sapply(validation2, class)

##       MAP3K1         AKT1         MEN1       SHANK2        NCOR2 VITAL_STATUS
##    "numeric"    "numeric"    "numeric"    "numeric"    "numeric"     "factor"

pred2 = as.numeric(as.character(predictions2))
valid2 = as.character(as.factor(validation2$VITAL_STATUS))
valid2 = as.numeric(as.character(validation2$VITAL_STATUS))
prroc_obj2 = roc.curve(scores.class0 = pred2, weights.class0 = valid2, curve = TRUE)
plot(prroc_obj2)
```



Part C: This analysis captures a combination of clinical and genetic features

- 4.42. For this analysis, remove RADIO_THERAPY variable from the TNBC_All object and save in TNBC_3
- 4.43. Convert integer values to numeric and ensure that dependent variable "VITAL_STATUS" is of the factor class

```
# remove RADIO_THERAPY feature from the TNBC_All object and save in the TNBC_3 object

TNBC_3 = cbind(TNBC_1,TNBC_2)
TNBC_3 = TNBC_3[-6]

str(TNBC_3)

## 'data.frame':    190 obs. of  11 variables:
##  $ BREAST_SURGERY : num  1 1 1 1 1 2 2 1 2 1 ...
##  $ TUMOR_SIZE     : num  65 17 40 39 25 13 30 16 24 34 ...
##  $ NPI            : num  6 4 6 4 4 4 4 4 4 6 ...
##  $ SURVIVAL_GROUPS: num  1 2 1 1 2 2 2 1 2 1 ...
##  $ SURVIVAL_MONTHS: num  22 153 28 8 66 132 114 10 78 17 ...
##  $ MAP3K1         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ AKT1           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ MEN1           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ SHANK2         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ NCOR2          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ VITAL_STATUS   : Factor w/ 2 levels "0","1": 2 1 2 2 2 1 1 1 1 2 ...
```

- 4.44. Perform feature selection.
- 4.45. Feature Selection using Boruta, mlbench and Caret packages respectively

```
# Feature Selection operation using the boruta function
set.seed(7)
boruta3 = Boruta(TNBC_3$VITAL_STATUS~.,data = TNBC_3, doTrace = 4, maxRuns = 100)

##  1. run of importance source...

## Assigned hit to 7 attributes out of 10 undecided.

##  2. run of importance source...

## Assigned hit to 9 attributes out of 10 undecided.
```

```
##  3. run of importance source...

## Assigned hit to 10 attributes out of 10 undecided.

##  4. run of importance source...

## Assigned hit to 10 attributes out of 10 undecided.

##  5. run of importance source...

## Assigned hit to 8 attributes out of 10 undecided.

##  6. run of importance source...

## Assigned hit to 8 attributes out of 10 undecided.

##  7. run of importance source...

## Assigned hit to 10 attributes out of 10 undecided.

##  8. run of importance source...

## Assigned hit to 7 attributes out of 10 undecided.

##  9. run of importance source...

## Assigned hit to 10 attributes out of 10 undecided.

##  10. run of importance source...

## Assigned hit to 5 attributes out of 10 undecided.

## After 10 iterations, +0.68 secs:

##  confirmed 5 attributes: AKT1, MAP3K1, SHANK2, SURVIVAL_GROUPS, SURVIVAL_MONTHS;

##  still have 5 attributes left.

##  11. run of importance source...

## Assigned hit to 1 attribute out of 5 undecided.

##  12. run of importance source...

## Assigned hit to 4 attributes out of 5 undecided.

##  13. run of importance source...

## Assigned hit to 3 attributes out of 5 undecided.

##  14. run of importance source...

## Assigned hit to 4 attributes out of 5 undecided.

## After 14 iterations, +1.1 secs:

##  confirmed 1 attribute: NPI;

##  still have 4 attributes left.

##  15. run of importance source...

## Assigned hit to 2 attributes out of 4 undecided.

##  16. run of importance source...

## Assigned hit to 4 attributes out of 4 undecided.

##  17. run of importance source...

## Assigned hit to 2 attributes out of 4 undecided.
```

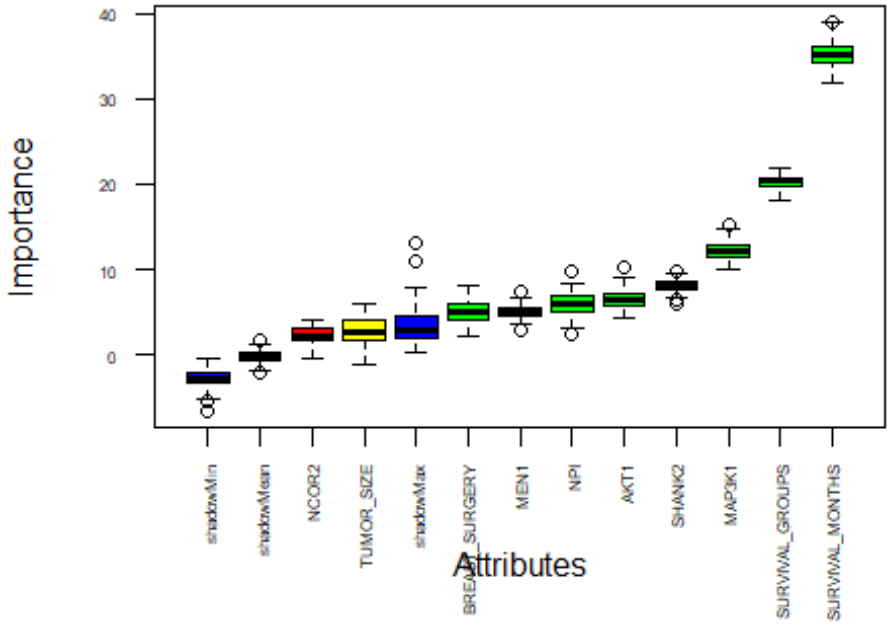
```
##  18. run of importance source...

## Assigned hit to 2 attributes out of 4 undecided.

## After 18 iterations, +1.5 secs:

##  confirmed 1 attribute: MEN1;

##  still have 3 attributes left.

##  19. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  20. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  21. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  22. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  23. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

##  24. run of importance source...

## None of undecided attributes scored a hit.

##  25. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  26. run of importance source...

## Assigned hit to 3 attributes out of 3 undecided.

##  27. run of importance source...

## None of undecided attributes scored a hit.

##  28. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  29. run of importance source...

## Assigned hit to 2 attributes out of 3 undecided.

##  30. run of importance source...

## Assigned hit to 1 attribute out of 3 undecided.

## After 30 iterations, +2.6 secs:

##  confirmed 1 attribute: BREAST_SURGERY;

##  still have 2 attributes left.

##  31. run of importance source...

## None of undecided attributes scored a hit.

##  32. run of importance source...

## None of undecided attributes scored a hit.
```

```
##  33. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  34. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  35. run of importance source...

## None of undecided attributes scored a hit.

##  36. run of importance source...

## None of undecided attributes scored a hit.

##  37. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  38. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  39. run of importance source...

## None of undecided attributes scored a hit.

##  40. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  41. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  42. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  43. run of importance source...

## None of undecided attributes scored a hit.

##  44. run of importance source...

## None of undecided attributes scored a hit.

##  45. run of importance source...

## None of undecided attributes scored a hit.

##  46. run of importance source...

## None of undecided attributes scored a hit.

##  47. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  48. run of importance source...

## None of undecided attributes scored a hit.

##  49. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  50. run of importance source...

## None of undecided attributes scored a hit.
```

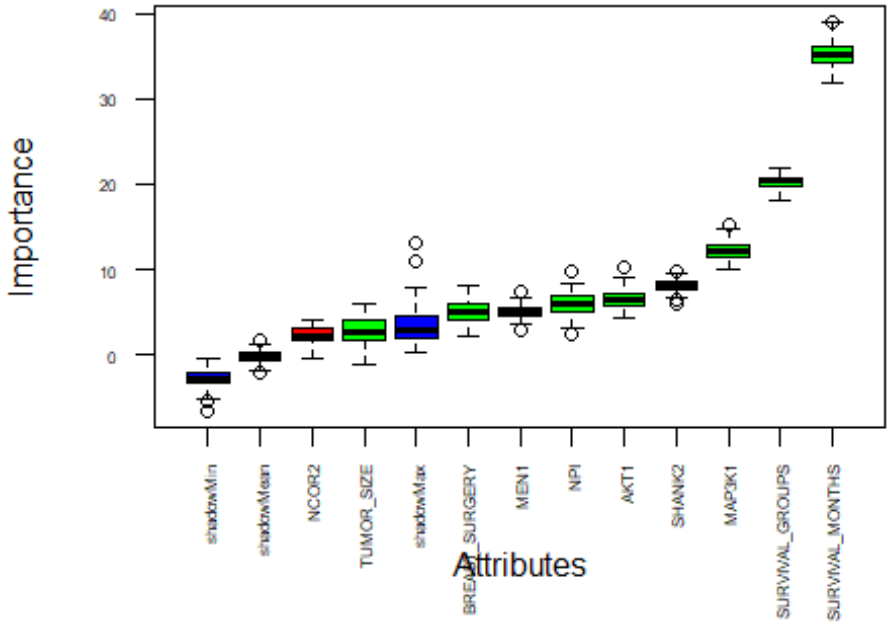
```
##  51. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  52. run of importance source...

## None of undecided attributes scored a hit.

##  53. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  54. run of importance source...

## None of undecided attributes scored a hit.

##  55. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  56. run of importance source...

## None of undecided attributes scored a hit.

##  57. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  58. run of importance source...

## None of undecided attributes scored a hit.

##  59. run of importance source...

## None of undecided attributes scored a hit.

##  60. run of importance source...

## None of undecided attributes scored a hit.

##  61. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  62. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  63. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  64. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  65. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  66. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  67. run of importance source...

## None of undecided attributes scored a hit.

##  68. run of importance source...

## None of undecided attributes scored a hit.
```

```
##  69. run of importance source...

## None of undecided attributes scored a hit.

##  70. run of importance source...

## None of undecided attributes scored a hit.

##  71. run of importance source...

## None of undecided attributes scored a hit.

##  72. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  73. run of importance source...

## None of undecided attributes scored a hit.

##  74. run of importance source...

## Assigned hit to 2 attributes out of 2 undecided.

##  75. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  76. run of importance source...

## None of undecided attributes scored a hit.

##  77. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  78. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

##  79. run of importance source...

## None of undecided attributes scored a hit.

##  80. run of importance source...

## None of undecided attributes scored a hit.

##  81. run of importance source...

## Assigned hit to 1 attribute out of 2 undecided.

## After 81 iterations, +6.7 secs:

##  rejected 1 attribute: NCOR2;

##  still have 1 attribute left.

##  82. run of importance source...

## None of undecided attributes scored a hit.

##  83. run of importance source...

## None of undecided attributes scored a hit.

##  84. run of importance source...

## None of undecided attributes scored a hit.

##  85. run of importance source...
```

```
## Assigned hit to 1 attribute out of 1 undecided.

##  86. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  87. run of importance source...

## None of undecided attributes scored a hit.

##  88. run of importance source...

## None of undecided attributes scored a hit.

##  89. run of importance source...

## None of undecided attributes scored a hit.

##  90. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  91. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  92. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  93. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  94. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  95. run of importance source...

## None of undecided attributes scored a hit.

##  96. run of importance source...

## Assigned hit to 1 attribute out of 1 undecided.

##  97. run of importance source...

## None of undecided attributes scored a hit.

##  98. run of importance source...

## None of undecided attributes scored a hit.

##  99. run of importance source...

## None of undecided attributes scored a hit.

plot(boruta3, las = 2, cex.axis = 0.5)
```




```
# reclassify features that were designated as tentative
bor3 = TentativeRoughFix(boruta3)
plot(bor3, las = 2, cex.axis = 0.5)
```




```
attStats(boruta3)
```

```
##                   meanImp medianImp     minImp    maxImp  normHits  decision
## BREAST_SURGERY   5.111934  5.120375  2.1115642  8.232683 0.7979798 Confirmed
## TUMOR_SIZE       2.789070  2.751431 -1.2100698  6.043410 0.4444444 Tentative
## NPI              5.985522  5.999452  2.5347494  9.839659 0.8787879 Confirmed
## SURVIVAL_GROUPS 20.229976 20.357407 18.1170256 21.984505 1.0000000 Confirmed
## SURVIVAL_MONTHS 35.317004 35.094013 31.8683125 39.026247 1.0000000 Confirmed
## MAP3K1          12.267995 12.214001 10.0146685 15.208270 0.9898990 Confirmed
## AKT1             6.444094  6.488535  4.2142956 10.207389 0.9191919 Confirmed
## MEN1             5.126237  5.129232  2.8158367  7.426156 0.8383838 Confirmed
## SHANK2           8.083977  8.115455  5.9705706  9.822149 0.9797980 Confirmed
## NCOR2            2.334204  2.183119 -0.4298524  4.168576 0.2626263  Rejected

print(bor3)

## Boruta performed 99 iterations in 7.845032 secs.
## Tentatives roughfixed over the last 99 iterations.
##  9 attributes confirmed important: AKT1, BREAST_SURGERY, MAP3K1, MEN1,
## NPI and 4 more;
##  1 attributes confirmed unimportant: NCOR2;

# create object and limit predictor features to the 9 confirmed features from the boruta analysis and dependent
variable

col_order4 = c("SURVIVAL_MONTHS","SURVIVAL_GROUPS","NPI","BREAST_SURGERY","TUMOR_SIZE",
               "MAP3K1","SHANK2","AKT1","MEN1","VITAL_STATUS")
TNBC_3 = TNBC_3[, col_order4]
```

- 4.46. Split the TNBC_3 clinical + genetic features object into training and validation groups (objects)
- 4.47. Split out validation dataset
- 4.48. Do this by creating a list of 70% of the rows in the original dataset to serve as training set

```
#Living
Living3 = TNBC_3[TNBC_3$VITAL_STATUS == 0,]

#Died of Disease
Deceased3 = TNBC_3[TNBC_3$VITAL_STATUS == 1,]

set.seed(7)
validationIndex3 = createDataPartition(TNBC_3$VITAL_STATUS, p=0.70, list = FALSE)

#select 30% of the data for validation

validation3 = TNBC_3[-validationIndex3,]

#use the remaining 70% of data to train the model
dataset3 = TNBC_3[validationIndex3,]

# Check that the distribution of the dependent variable is similar in train and test sets
prop.table(table(TNBC_3$VITAL_STATUS))

##
##         0         1
## 0.5631579 0.4368421

prop.table(table(dataset3$VITAL_STATUS))

##
##         0         1
## 0.5597015 0.4402985

prop.table(table(validation3$VITAL_STATUS))

##
##         0         1
## 0.5714286 0.4285714
```

- The figures show that the ratio of the “Living”:“Died of Disease” was proportionally split

- 4.49. Analyze the data using descriptive statistics

```
# Dimension of the datasets
dim(dataset3)

## [1] 134  10
```

```r
dim(validation3)

## [1] 56 10

#class distribution of training set
cbind(freq = table(dataset3$VITAL_STATUS), percentage = prop.table(table(dataset3$VITAL_STATUS))*100)

##   freq percentage
## 0   75   55.97015
## 1   59   44.02985

#summarize correlations between predictor variables to ascertain if there is the potential for col-linearity
cor(dataset3[,1:9])

##                  SURVIVAL_MONTHS SURVIVAL_GROUPS         NPI BREAST_SURGERY
## SURVIVAL_MONTHS      1.00000000      0.77069641 -0.19171129     0.08953678
## SURVIVAL_GROUPS      0.77069641      1.00000000 -0.30716837     0.29915885
## NPI                 -0.19171129     -0.30716837  1.00000000    -0.36723879
## BREAST_SURGERY       0.08953678      0.29915885 -0.36723879     1.00000000
## TUMOR_SIZE          -0.10080979     -0.12331195  0.22202900    -0.14012896
## MAP3K1              -0.02902820     -0.03066152 -0.04047493    -0.00507467
## SHANK2               0.20458413      0.08946061 -0.01400807    -0.03603479
## AKT1                -0.13188091     -0.13567725  0.00744267    -0.04193833
## MEN1                -0.09315831     -0.03433152  0.03843534     0.10429770
##                   TUMOR_SIZ      MAP3K1      SHANK2        AKT1        MEN1
## SURVIVAL_MONTHS -0.10080979 -0.02902820  0.20458413 -0.13188091 -0.09315831
## SURVIVAL_GROUPS -0.12331195 -0.03066152  0.08946061 -0.13567725 -0.03433152
## NPI              0.22202900 -0.04047493 -0.01400807  0.00744267  0.03843534
## BREAST_SURGERY  -0.14012896 -0.00507467 -0.03603479 -0.04193833  0.10429770
## TUMOR_SIZE       1.00000000 -0.09031983 -0.06712603 -0.04572776 -0.06203120
## MAP3K1          -0.09031983  1.00000000 -0.05082973 -0.05915710  0.24770137
## SHANK2          -0.06712603 -0.05082973  1.00000000 -0.05455447 -0.02665009
## AKT1            -0.04572776 -0.05915710 -0.05455447  1.00000000 -0.03101614
## MEN1            -0.06203120  0.24770137 -0.02665009 -0.03101614  1.00000000
```

- Highest correlation is between SSURVIVAL_MONTHS and SURVIVAL_GROUPS genes

- 4.50. Apply the KNN Learning Algorithm on the clinical + mutation data

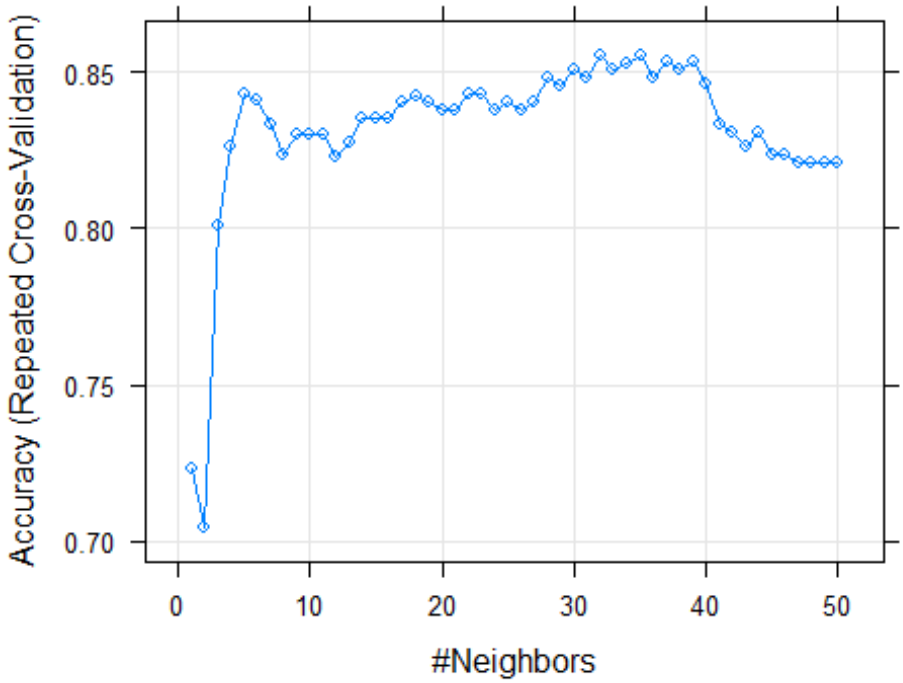
- 4.51. Analysis is done using 10-fold cross validation with 3 repeats

```r
trainControl3 = trainControl(method = "repeatedcv", number = 10, repeats = 3)
metric3 = "Accuracy"

set.seed(7)
fit.knn3 = train(VITAL_STATUS~., data=dataset3, method="knn", metric=metric3, trControl=trainControl3)

fit.knn3

## k-Nearest Neighbors
##
## 134 samples
##   9 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 120, 120, 120, 121, 120, 121, ...
## Resampling results across tuning parameters:
##
##   k  Accuracy   Kappa
##   5  0.8431929  0.6747001
##   7  0.8330891  0.6532868
##   9  0.8303114  0.6461173
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 5.
```

- 4.52. Tuning for the K-Nearest Neighbor (KNN) learning algorithm

```r
# KNN
set.seed(7)
grid3 = expand.grid(.k = seq(1,50, by=1))
fit.knn_tuned3 = train(VITAL_STATUS~.,data = dataset3, method="knn", metric=metric3, tuneGrid=grid3,
                       trControl=trainControl3)
fit.knn_tuned3
```

```
## k-Nearest Neighbors
##
## 134 samples
##   9 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 120, 120, 120, 121, 120, 121, ...
## Resampling results across tuning parameters:
##
##   k   Accuracy   Kappa
##    1  0.7234737  0.4315616
##    2  0.7041819  0.3959883
##    3  0.8010073  0.5880806
##    4  0.8263431  0.6405447
##    5  0.8431929  0.6747001
##    6  0.8412088  0.6717998
##    7  0.8330891  0.6532868
##    8  0.8233822  0.6326019
##    9  0.8303114  0.6461173
##   10  0.8303114  0.6454967
##   11  0.8303114  0.6454967
##   12  0.8228022  0.6293911
##   13  0.8277473  0.6402773
##   14  0.8352564  0.6556433
##   15  0.8354396  0.6556913
##   16  0.8354396  0.6546608
##   17  0.8402015  0.6652429
##   18  0.8425824  0.6699916
##   19  0.8403846  0.6659607
##   20  0.8378205  0.6608627
##   21  0.8378205  0.6611129
##   22  0.8431624  0.6730869
##   23  0.8431624  0.6730869
##   24  0.8380342  0.6626421
##   25  0.8405983  0.6690475
##   26  0.8378205  0.6622069
##   27  0.8407814  0.6690880
##   28  0.8483211  0.6862844
##   29  0.8459402  0.6820866
##   30  0.8507021  0.6926374
##   31  0.8485043  0.6872651
##   32  0.8556471  0.7023782
##   33  0.8507021  0.6928353
##   34  0.8530830  0.6979940
##   35  0.8554640  0.7031744
##   36  0.8483211  0.6888643
##   37  0.8532967  0.6995599
##   38  0.8510989  0.6947154
##   39  0.8532967  0.6995599
##   40  0.8459707  0.6854448
##   41  0.8335165  0.6624986
##   42  0.8309524  0.6570251
##   43  0.8263736  0.6486060
##   44  0.8309524  0.6570331
##   45  0.8238095  0.6435172
##   46  0.8235958  0.6430802
##   47  0.8210317  0.6379913
##   48  0.8212149  0.6384001
##   49  0.8210317  0.6379913
##   50  0.8210317  0.6379913
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 32.

plot(fit.knn_tuned3)
```



* Tuning improved

KNN model from an accuracy level of 0.8431 to 0.8556 at k = 32

- 4.53. Finalize by building Model 3 using KNN
- 4.54. Prepare the data transform for finalizing Model 3

```
set.seed(7)
z = dataset3[,1:9]
preprocessParam3 = preProcess(z)
z = predict(preprocessParam3, z)

# prepare the validation dataset
set.seed(7)

# transform the validation dataset
validationz = predict(preprocessParam3, validation3[,1:9])
```

- 4.55. Make predictions and build confusion matrix for Model 3

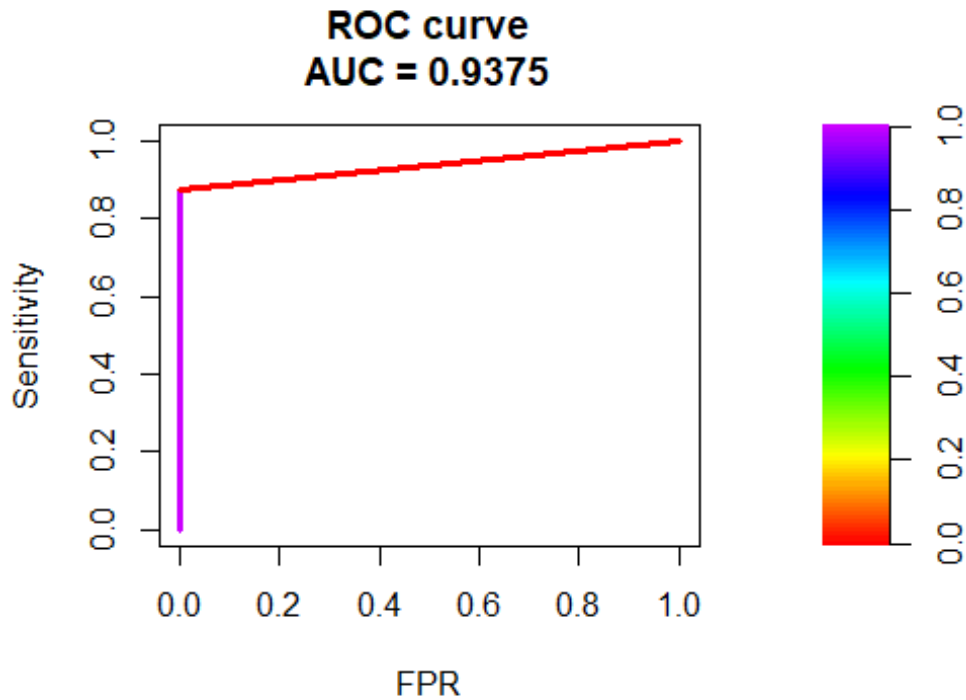
```
set.seed(7)
predictions3 = knn3Train(z, validationz, dataset3$VITAL_STATUS, k=32, prob = FALSE)
confusionMatrix(factor(predictions3), factor(validation3$VITAL_STATUS))

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 32  3
##          1  0 21
##
##                Accuracy : 0.9464
##                  95% CI : (0.8513, 0.9888)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 3.093e-10
##
##                   Kappa : 0.8889
##
##  Mcnemar's Test P-Value : 0.2482
##
##             Sensitivity : 1.0000
##             Specificity : 0.8750
##          Pos Pred Value : 0.9143
##          Neg Pred Value : 1.0000
##              Prevalence : 0.5714
```

```
##          Detection Rate : 0.5714
##    Detection Prevalence : 0.6250
##       Balanced Accuracy : 0.9375
##
##        'Positive' Class : 0
##
```

- Model performed at an accuracy of 0.9464 with only 3 errors calls leading to a specificity of 0.8750 but a perfect sensitivity score of 1. Kappa was very high at 0.8889.

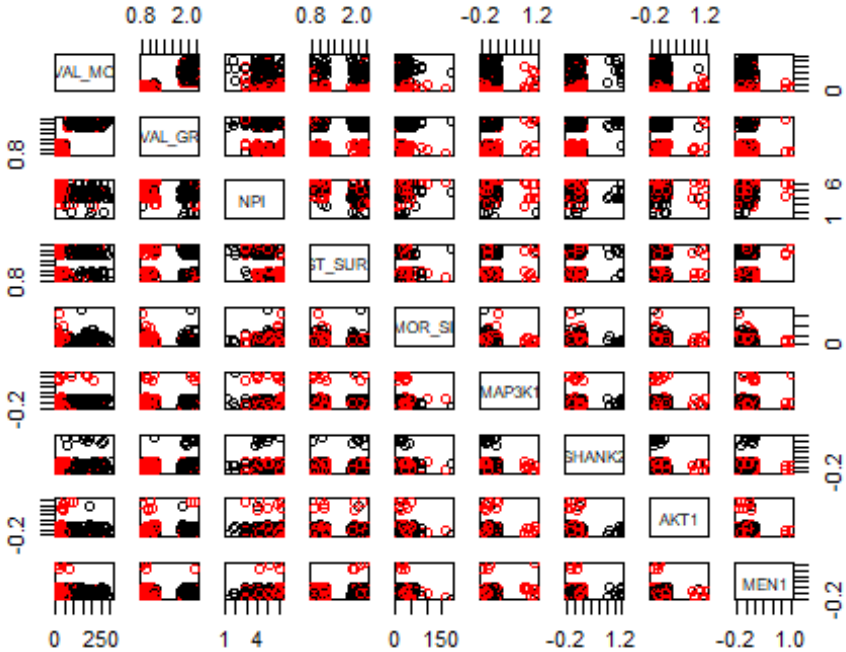
- 4.56. Calculate AUC(ROC) for Model 3

```
library(PRROC)
sapply(predictions3, class)

##           0           1           1           1           1           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           0           1
## "character" "character" "character" "character" "character" "character"
##           0           0           0           1           1           0
## "character" "character" "character" "character" "character" "character"
##           1           0           1           0           0           0
## "character" "character" "character" "character" "character" "character"
##           0           1           0           0           0           1
## "character" "character" "character" "character" "character" "character"
##           1           0           0           0           1           0
## "character" "character" "character" "character" "character" "character"
##           0           0           1           0           1           0
## "character" "character" "character" "character" "character" "character"
##           0           0           0           0           1           1
## "character" "character" "character" "character" "character" "character"
##           0           1           0           1           0           1
## "character" "character" "character" "character" "character" "character"
##           0           1
## "character" "character"

sapply(validation3, class)

## SURVIVAL_MONTHS SURVIVAL_GROUPS             NPI  BREAST_SURGERY      TUMOR_SIZE
##       "numeric"       "numeric"       "numeric"       "numeric"       "numeric"
##          MAP3K1          SHANK2            AKT1            MEN1    VITAL_STATUS
##       "numeric"       "numeric"       "numeric"       "numeric"        "factor"

pred3 = as.numeric(as.character(predictions3))
valid3 = as.character(as.factor(validation3$VITAL_STATUS))
valid3 = as.numeric(as.character(validation3$VITAL_STATUS))
prroc_obj3 = roc.curve(scores.class0 = pred3, weights.class0 = valid3, curve = TRUE)
plot(prroc_obj3)
```



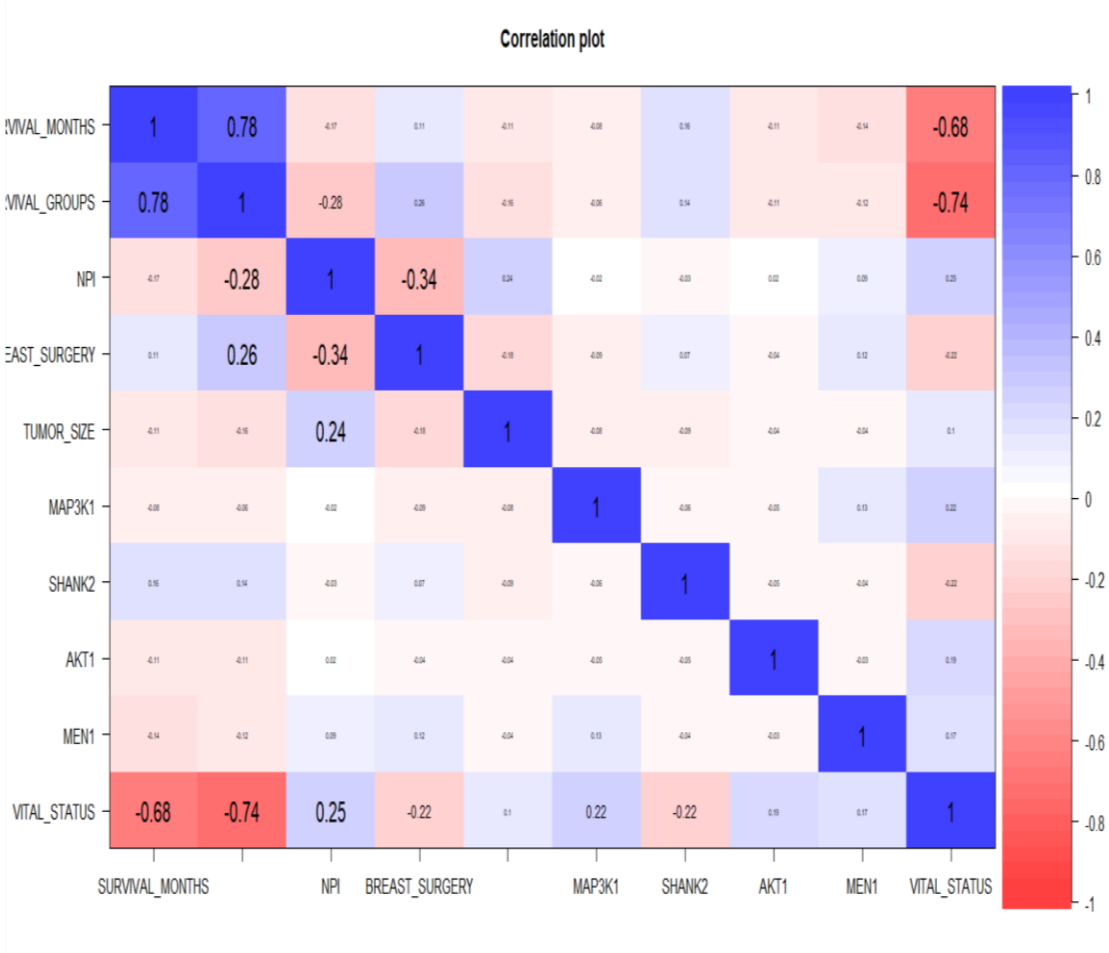
- 4.57. Multi-modal Data Visualizations of the 9 Predictor variables and the dependent Variable

```
jittered_a = sapply(TNBC_3[,1:9], jitter)
pairs(jittered_a, names(TNBC_3[,1:9]), col = TNBC_3$VITAL_STATUS)
```



- 4.58. Correlation plot of Models Predictor Variables and the dependent variable

```
TNBC_1B = TNBC_3

TNBC_1B$VITAL_STATUS = as.character(TNBC_1B$VITAL_STATUS)
TNBC_1B$VITAL_STATUS[TNBC_1B$VITAL_STATUS == "Living"] <- 0
TNBC_1B$VITAL_STATUS[TNBC_1B$VITAL_STATUS == "Died of Disease"] <- 1
TNBC_1B$VITAL_STATUS = as.numeric(TNBC_1B$VITAL_STATUS)


cor.plot(TNBC_1B)
```



The End