THE APPLICATION OF DATA VISUALIZATION IN AUDITING

By

ABDULLAH ALAWADHI

A dissertation submitted to the

Graduate School- Newark

Rutgers, The State University of New Jersey

in partial fulfillment of requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

Written under the direction of

Professor Miklos A. Vasarhelyi

and approved by

____

Professor Miklos A. Vasarhelyi

____

Professor Michael Alles

____

Professor Helen Brown

____

Professor Asli Basoglu

Newark, New Jersey

May, 2015

Abstract of the Dissertation

The Application of Data Visualization in Auditing

By Abdullah Alawadhi

Dissertation Chairman: Prof. Miklos A. Vasarhelyi

Albert Einstein once wrote, "My particular ability does not lie in mathematical calculation, but rather in visualizing effects, possibilities, and consequences." Today, the web, cloud, virtual reality, enterprise resource planning (ERP) systems and other evolving technologies are increasing the stream of data in all areas, thus the need for "visualizing effects, possibilities, and consequences", is necessary to reduce the volume to a manageable size and focus on those crucial data points.

With the advent of such "Big Data", businesses are now so overwhelmed by these data cascading into and through their business operations. They are faced with major challenges in processing and analyzing data. Furthermore, auditors will need to rely on more advance techniques throughout the audit cycle since traditional techniques are becoming less and less effective and efficient in the Big Data environment.

Data visualization is one solution that presents information in ways that engages the use of the human's cognitive and visual abilities. It is the process of converting raw data into some visual form. It helps by shifting the cognitive load coupled with the understanding of Big Data, to the human perceptual system through graphics and visuals. Data visualization helps auditors gain better insights, draw better conclusions and ultimately improve the audit process.

This dissertation will contribute to the existing literature by exploring the application of various data visualization techniques in auditing. Despite being widely used in other fields such as medicine and engineering, the auditing profession has been behind in their usage of data visualization, and this dissertation will attempt to bridge this gap. Specifically, it will look into and analyze the existing literature and their usage of data visualization techniques to shed light and understand its current evolution and its future trend. It will also contribute by providing two illustrations of how exploratory and explanatory data visualization can be applied in the audit profession. The first will demonstrate how it can be applied in the audit planning stage, where little is known about the data. The second will illustrate how an expert visual dashboard can be applied in the area of procurement cards for the purpose of audit monitoring and control.

ACKNOWLEDGMENTS

It is my great pleasure to thank the many people who have made this dissertation possible. It is difficult to overstate my gratitude to my mentor and Ph.D. advisor, Dr. Miklos Vasarhelyi. His encouragement and support during my Ph.D. study at Rutgers University helped me complete this dissertation. His enthusiasm, inspiration, and critiques helped me become a better researcher.

I would also like to gratefully acknowledge the contributions and guidance of my dissertation committee members, Dr. Michael Alles for his insightful feedback and open door policy, Dr. Helen Brown-Liburd for her supportive words and advice, and Dr Asli Basoglu for her appreciative comments and guidance.

I am indebted to my many colleagues and friends for providing helpful comments and a stimulating environment in which to learn and grow. Particularly, Deniz Applebaum and Paul Byrnes for their participation and feedback during data collection and analysis, and Dr. Hussain Issa for his passionate support and help during my research. I would also like to thank Lisa Koors for giving her time and effort during the knowledge elicitation phase.

I also wish to thank my entire family for providing love, support, and encouragement throughout my academic years, specifically my mother and father who raised me, supported me, taught me, and loved me.

Lastly, and most importantly I would like to give great acknowledgement, appreciation, and thanks to my lovely wife Moneera Altawari, who stood beside me and

supported me throughout my studies. I will never forget the sacrifices she made on behalf of my education and career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: INTRODUCTION

This dissertation presents three main essays on data visualization and its application in auditing. Chapter one of this dissertation introduces the motivation behind the topic and provides a background and general overview. Chapters two, three and four present the remaining three essays. The last chapter concludes the dissertation by discussing the limitations and future research areas.

1.1 INTRODUCTION

In the year of 2002, researchers from the University of Berkeley have estimated that we generate about 1 Exabyte (1 Million Terabyte) of data every year (Sagiroglu and Sinanc, 2013). The ease of gathering such data has been facilitated by several technologies, such as Radio Frequency Identifiers (RFID). Many other sources of data, such as social media networks, video and audio streams, cell phone Global Positioning System (GPS) data, and ERP systems have also contributed to the large structure of today's "Big Data". In Big Data, the amount and size of data is beyond the limit of what information systems can store, process, and/or analyze (Vasarhelyi et al., 2015, Warren et al., 2015). The accumulation and evaluation of such Big Data are becoming key elements for a firm's competitive advantage (Bughin et al., 2011). Furthermore, it has been found that firms who leverage on Big Data by utilizing data analytics effectively achieve five to six percent gains in productivity (Brynjolfsson et al., 2011). Auditors on the other hand are also likely to benefit from Big Data as it provides them with more access to information. This can lead to effective identification of anomalies in the data that may or may not suggest related risks (Brown-Liburd et al., 2015).

Nevertheless, researchers have recognized the importance of Big Data and the benefits, obstacles and changes it brings to the auditing and accounting profession (Alles. 2015, Warren et al.,2015, Vasarhelyi et al.,2015, Yoon et al.,2015, Brown-Liburd et al.,2015, Zhang et al.,2015). However, despite being an important resource, it is generally difficult to extract relevant information from Big Data. Especially when many parameters are recorded, resulting in multi-dimensional data with a high dimensionality. Such complexity has lead companies to face major challenges in processing and analyzing their data (Moffitt & Vasarhelyi 2013). Furthermore, the problem of efficiently analyzing such multi-dimensional data is becoming increasingly challenging (Cuzzocrea & Mansmann, 2009).

1.2. INFORMATION OVERLOAD AND HUMAN INFORMATION PROCESSING

With the availability and access of such Big Data, the possibility of information overload is likely. Prior research in psychology has demonstrated that decision makers have limited ability to process large amounts of information required for complex decision making (Kleinmuntz 1990; Iselin 1988). Therefore, the decision-maker must reduce the volume to a manageable size and focus on those data points which are crucial to the task at hand (Chorba & New 1980). One solution could be to present information in ways that engage the use of the human's cognitive and visual abilities (Sloman 1996). Chervany and

Dickson (1974) suggested a way to deal with large volumes of data by simply summarizing them. They state that summarized data presentations, such as tables and graphs, are preferred to large amounts of data since summarization leads to lower cost decisions.

The human visual cortex dominates our perception and is highly specialized for processing information. The average number of neurons in the entire human visual cortex has been estimated at around 280 million (Leuba & Kraftsik, 1994). Moreover, more than 30 million neurons are activated in the visual cortex when we see a single-object image (Levy et al., 2004). Therefore, humans have great visual and spatial abilities, and are able to detect outliers, variations in color, shape, and motion. We are able to recognize patterns and retrieve information using various visual cues (Kosslyn & Cunningham 1994). Therefore, visualization is a sound and efficient way of presenting data and analyzing it effectively. Many tools have appeared that promise to help decision makers reduce large data sets to simple visuals (Lurie & Mason, 2007). These visual tools range from common bar charts to sophisticated virtual environments. Data visualization helps by shifting the cognitive load coupled with the understanding of Big Data, to the human perceptual system through graphics and visuals (Lohse 1997; Zhang & Whinston 1995).

Decision-making is a cognitive process that leads to the selection of a course of action among alternatives that produces a decision outcome (Libby, 1981; Cloyd, 1995). It involves three central stages: input; processing and output (Libby & Lewis, 1977). Since the decision-making process involves reliance on the nature and content of information being entered (Big Data versus non-Big Data), and how the information is being presented (visually versus non-visually), there is a possibility that decision makers would not be able to fully process the information due to human information processing limitations. (Cloyd, 1995; Roberts, 2002).

The source of deficiency in decision making resulting from the limitations in human information processing can be caused by two main sources: the decision maker or the task. When the deficiency is caused by the decision makers, it generally relates to limited skills and knowledge the decision maker has, and with proper training, the deficiency maybe alleviated (Roberts, 2002). However, if the deficiency is caused by the task, then a primary reason can be due to size and volume of information at hand. In such instances, the decision maker may potentially face information overload, which is the notion of receiving too much information (Eppler & Mengis, 2004). Due to this information overload, overall decision making tends to be less efficient and effective (Casey 1980; Malhotra 1982; Jacoby 1984; Herbig & Kramer 1994)

Once these limits of human information processing were established, researchers began exploring ways to overcome them, and today, there are many ways available to assist decision-making (Libby, 1981; Libby & Lewis, 1982; Libby et al., 2002). However, these methods may not be effective in today's Big Data environment. As such more advanced methods to address information processing in the Big Data environment is necessary. One potential solution that improves decision-making and addresses the limitations of human information processing discussed above, is the use of graphics and visuals (Ghani et al., 2009). Numerous studies have suggested the use of various presentation formats to minimize human information processing limitations (Chervany & Dickson, 1974; DeSanctis & Jarvenpaa, 1989; Iselin, 1988; Hard & Vanecek, 1991; Stone & Schkade, 1991; Frownfelter-Lohrke, 1998; Stocks & Tuttle, 1998; Dull et al., 2003). These studies suggest that presentation format, such as the use of various visualization techniques, could overcome the effect of increased information (Roberts, 2002) and also improve ways of thinking (Schick et al., 1990).

1.3. DATA VISUALIZAION

Prior studies have used different terms to refer to the presentation of information in visual form, such as "information visualization" (Card et al., 1999), "data visualization" (Green 1998), or "scientific visualization" (DeFanti et al., 1989). Despite the variation in naming, most have a common theme, and in this paper, I will use the term "data visualization" and define it as "*The selection, transformation, and presentation of various forms of data in visual form that helps facilitate exploration and understanding*"

In general, there are two categories of data visualization, each serving different purposes: explanation and exploration (Steele & Iliinsky, 2011). Explanatory data visualization is appropriate when we know what the data is and has to say. Specifically, explanatory data visualization is part of a presentation phase, where we want to convey certain information in a visual form. With Big Data, companies need better ways, to not only explore data, but to synthesize meaning from it. Producing visuals that provide explanation and understanding can have significant effects in guiding users toward a conclusion, persuading them to take different actions, or inviting them to ask entirely new questions. Nevertheless, creating such visuals requires preplanning, setting clear objectives, and obtaining the right visual elements

In contrast, visual data exploration is appropriate when little is known about the data and the exploration goals are vague. Translating large data sets into a visual medium can help in identifying interesting trends and/or outliers. Exploratory data visualization facilitates the user exploring the data, helping them unearth their own insights. Depending on the user's context, it is a discovery process that could or could not potentially lead to the finding of many different insights. Ultimately though, it could help users obtain interesting information, and build hypothesis from large amount of data (Steele & Iliinsky, 2011).

Users who utilize exploratory data visualization generally do not know what the data will show, and would usually analyze and look at the data from a couple of different angles, searching for relationships, connections, and insights that might be concealed in the data. In contrast, users who use explanatory data visualization can typically be called presenters as they are already experts in their own data. They have already explored and analyzed the data and highlighted the data points that support the core ideas they want to communicate (Fisher, 2010).

The effectiveness of a visualization techniques depends on several factors. These include perception, cognition, and the users' specific tasks and goals. Some visuals may convey meaning to a user in a certain way, while conveying it to others differently. How a user perceives a visualization depends on factors such as visual awareness, lighting conditions, color scales, and previous experience (Ware, 2012).

1.4. MOTIVATION AND CONTRIBUTION

The motivation behind this dissertation is that by presenting information in ways that would engage the human's cognitive and visual abilities, the effects of information overload associated with processing and analyzing Big Data is potentially reduced. Additionally, by utilizing both the explanatory and exploratory aspects of data visualization, auditors may obtain better insights during exploration, and communicate results more effectively during explanation.

The auditing and accounting literature have lagged in showcasing the application of data visualization. Therefore this dissertation attempts to contribute to the literature by presenting a full understanding of data visualization and showcasing its applications in different areas. It will also explore what role visualization can play in the audit process. Given that traditional accounting and auditing data is highly structured, one hypothesis is that in this setting: *"Visualization adds greatest value as an explanation role within the audit opposed to an exploration role as it is in medical and other sciences."* The exploration role is generally best when one cannot predefine the nature of relationships between the data, basically when the exploration goals are vague. Explanation on the other hand is best when aimed to highlight significant predefined/predetermined data points in a relationship.

The dissertation is organized as follows. The first essay will discuss data visualization and the different visual cues available. Not all visuals and graphs are effective for certain task and data types, hence this essay will attempt to shed light on the issue. Additionally, it will discuss the evolution of data visualization from the early 1500s till today, presenting the major milestones throughout history. Finally, it will present an in-depth visual analysis of the auditing literature and their research on data visualization. This section specifically contributes to the literature by using exploratory data visualization as means to explore and analyze the literature.

The second essay will talk about data visualization for knowledge discovery, which is an effective form of hypothesis generation. Specifically, this essay will showcase the application of explanatory and exploratory data visualization in Medicare health insurance industry for the purpose of knowledge discovery, to assist auditors in their planning stage. This essay will contribute to the literature by presenting different methodologies by which data visualization can be applied throughout the audit planning phase in a Big Data environment, whether the need is to explore and analyze the data, or to explain and understand it.

The final essay will present a paper that illustrates how the implementation of a visual dashboard can provide flexibility and improved performance in an audit setting. Specifically, this essay discusses the implementation of a prototype visual expert system in a firm by eliciting the knowledge of an expert in the procurement card (p-card) area. The contribution of this study is to bridge the gap in the literature by utilizing real world procurement data and applying a multi-dimensional approach for the detection of p-card misuse. Additionally, it will illustrate how visualization can be used as means to provide meaning and understanding throughout the process. Finally, the fifth and final chapter will summarize the conclusions and discuss limitations and future work.

CHAPTER 2: DATA VISUALIZATION, ITS EVOLUTION AND USAGE AND RESEARCH IN AUDITING: A LITERATURE ANALYSIS

2.1 INTRODUCTION

The main goal of data visualization is to help users gain better insights, draw better conclusions and eventually generate hypotheses. This is achieved by integrating the user's perceptual abilities to the data analysis process, and applying their flexibility, creativity, and general knowledge to the large data sets available in today's systems. Data visualization involves several main advantages. It deals more easily with highly noisy data. It usually allows for faster data exploration in large data sets. Finally it is intuitive and requires no understanding of complex mathematical or statistical algorithms (Keim, 2001).

McCandless (2012) presents four elements of a good data visualization shown in Figure 1. The first element is interestingness, which is concerned in making sure the audience cares about what you want to say, in that the data should be meaningful, relevant and new. Next is integrity, which deals with the question of whether you are saying the truth or not, and whether you are being consistent, honest and accurate. Form deals with the way you are trying to convey your message. In other words, the visuals need to be structured and pleasant to the human eyes. Finally we have function which is concerned with whether you are able to make your message easily understood. In order for a visual to be effective, all four elements need to be present, but if for example only integrity, function and form, are available without interestingness, the visualization may end up being slightly boring.

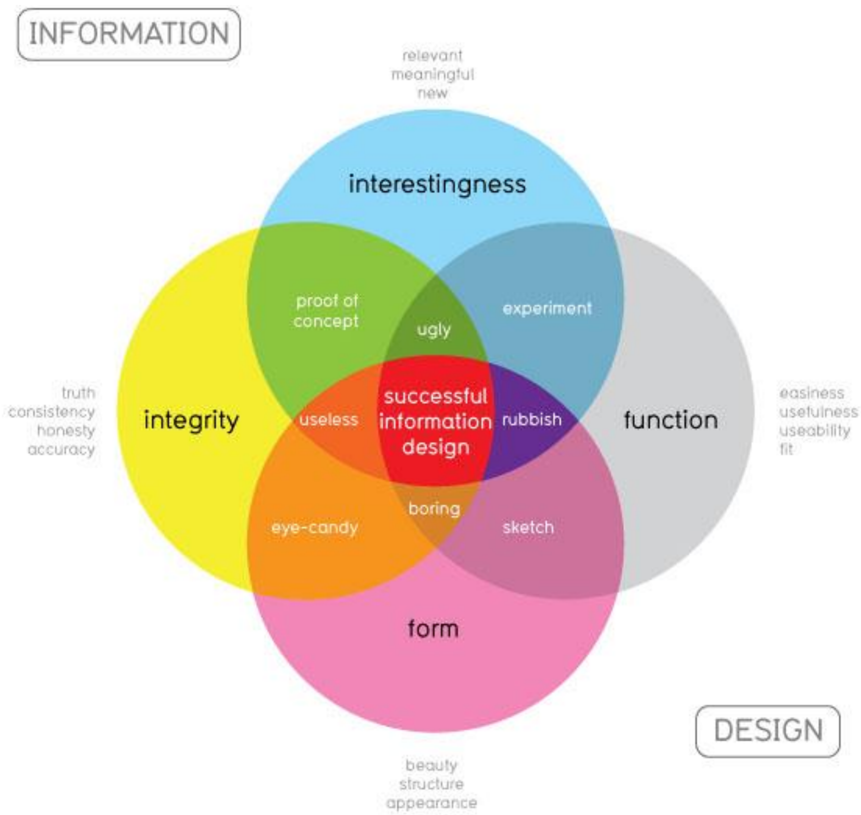


Figure 1: Four main elements of a successful data visualization

In general there are five main stages to data visualization: The collection and storage of data, the preprocessing of data, the hardware used for display, the algorithms used to visualize the data, and finally, and most importantly, the human perceptual and cognitive system (the process of thinking). A final stage can also be added where we can provide interactability with the visualization, giving the users the ability to manipulate the data or control what features are visible (such as those applied in visualization dashboards). Data visualization has several benefits: 1) it can easily deal with highly diverse and noisy data. 2) It is intuitive and requires no understanding of complex mathematical or statistical algorithms. 3) It usually allows for faster data exploration and often provides better results.

4) And finally data visualization may be useful in quality control, meaning that with an appropriate visualization, errors and artifacts in the data often become visible.

Figure 2 illustrates how visualization leverages the abilities of our eyes and brains and helps in detecting patterns or pattern violations fast and efficiently. By looking at the Figure, deviations from the norm can be observed. Simple changes in skew, position, color, size, blur or even shape of different objects in the visualization can be detected by our powerful visual and cognitive capabilities.

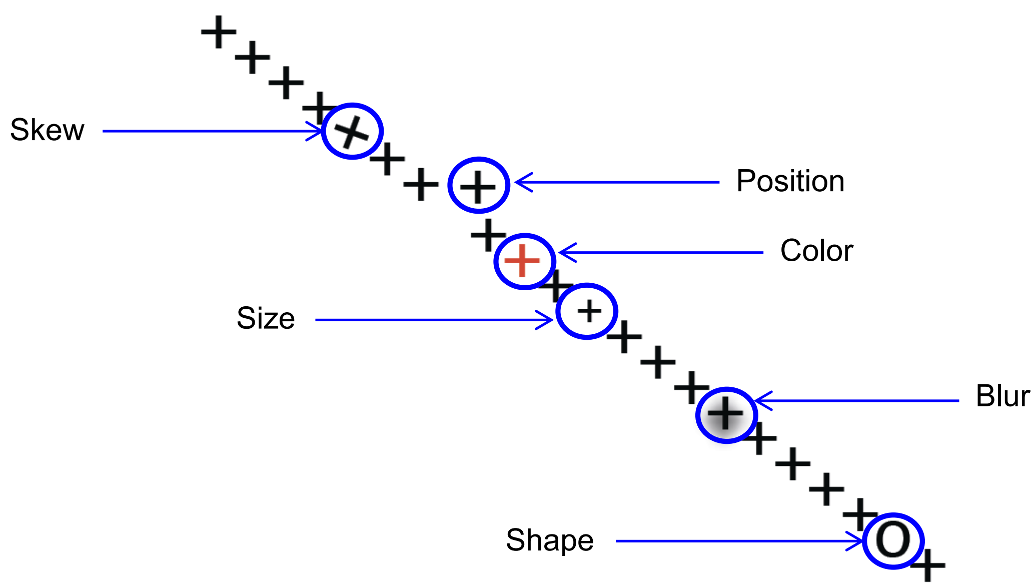


Figure 2: Visualization and Outlier Detection

This chapter will attempt to contribute to the literature by providing a full understanding of data visualization. It will also contribute by illustrating how exploratory data visualization can be used as means to study and analyze the literature. The first section will discuss the different types of data visualization and the associated visual ques. Not all visuals or graphs work for certain tasks, thus this section will present the necessary information to help map visuals to the data, making sure they are not only effective but relevant as well. The second section will discuss the evolution of data visualization from the early 10th century to our time today. The final section will conduct a visual literature analysis of the usage and research of data visualization in auditing, and attempt to present new methodologies for analyzing and reviewing the literature.

2.2. VISUAL CUES AND TYPES OF DATA VISUALIZATION

Data visualization in its most basic form is simply mapping data to geometry and color. However, an important aspect of visuals is to be able to map to the data, making sure its essence is intact, otherwise the visualizations will appear as nothing but meaningless shapes. Being able to choose the right visual cue is crucial. These cues generally change based on the task at hand, which depends on how varied shapes, sizes and shades are perceived (Keller & Keller, 1993; Yau, 2013). Figure 3 presents the 10 common visual cues (Yau, 2013). The following paragraphs will discuss these different cues and corresponding types of data visualization.

Position

The first visual cue is position, where values placed on a given space or coordinate system are compared. It helps in spotting clusters, trends, and outliers by plotting all the data at once. One example here is a scatterplot as seen in Figure 4. In scatterplots, data points, represented as dots, are judged based on their X and Y coordinates and where they are relative to others. Scatterplots can be useful when the data is large in size, since it draws all the data within the X- and Y-plane, taking less space.

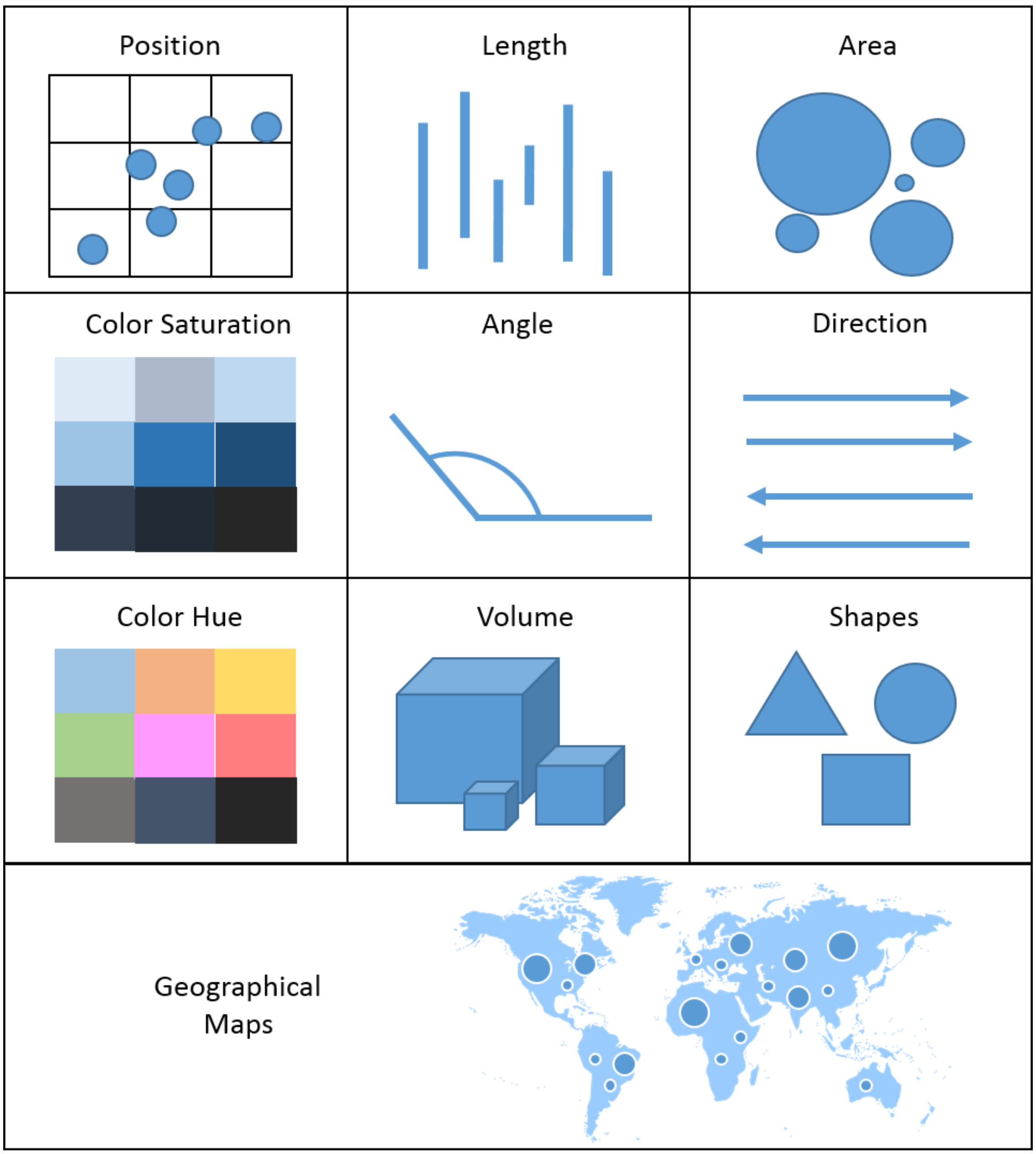


Figure 3: 10 Common Visual Cues

Clustering is another example of using position as the visual cue. It is a technique that deals with grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. It is a very useful technique for detecting outliers. Network graphs (Figure 5) is a more recent data visualization that is becoming more popular with social media being widely used. They specialize in modeling pairwise relations between data points, and add an additional feature of visualizing these relationships between the data points. A hybrid approach would be the combination of clustering and network graphs, as shown in Figure 6, where a set of similar objects are grouped together and a pairwise relations between them is modeled.

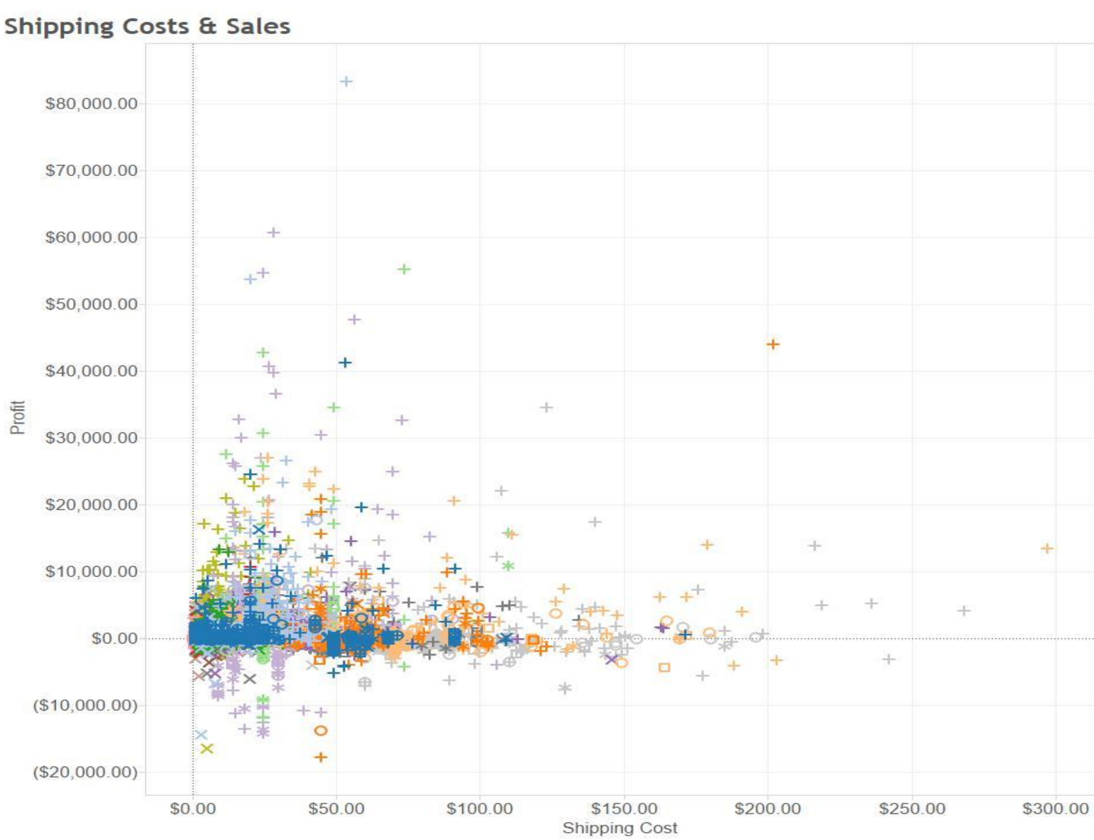


Figure 4: Scatterplot Visual

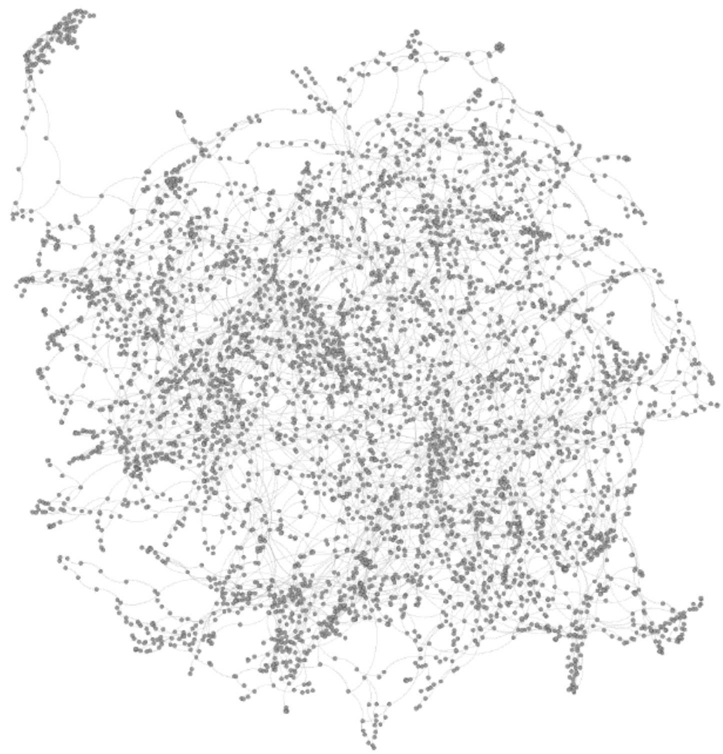


Figure 5: Network Graphs

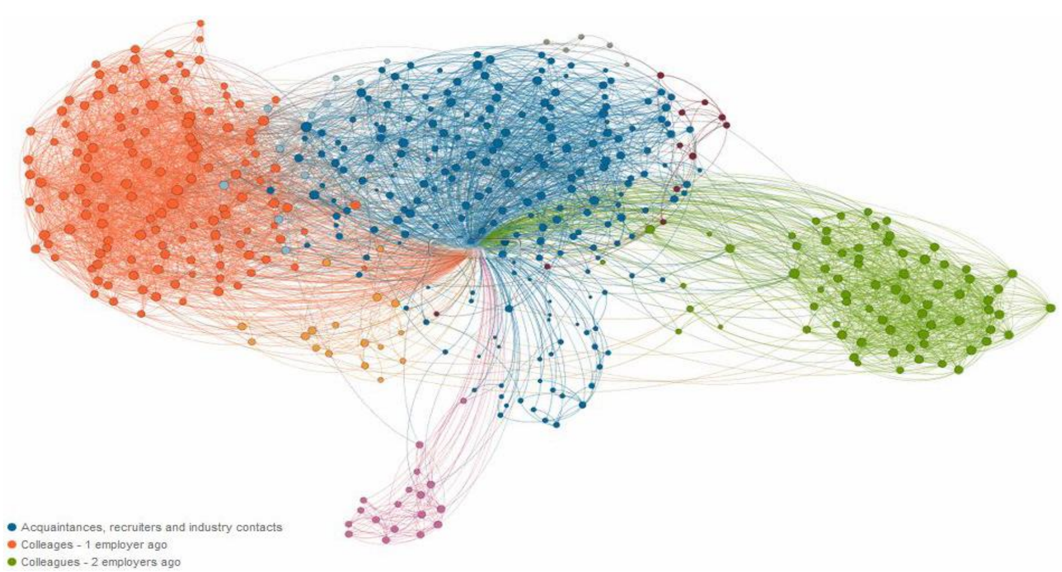


Figure 6: Network Cluster Graph

Length

Length on the other hand is most commonly used in the context of bar charts. Bar charts visualize data by displaying rectangular bars (horizontal or vertical) with lengths proportional to the values that they represent (Figure 7), the longer a bar is, the greater the absolute value. Bar charts are used to compare different values based on their bar lengths.

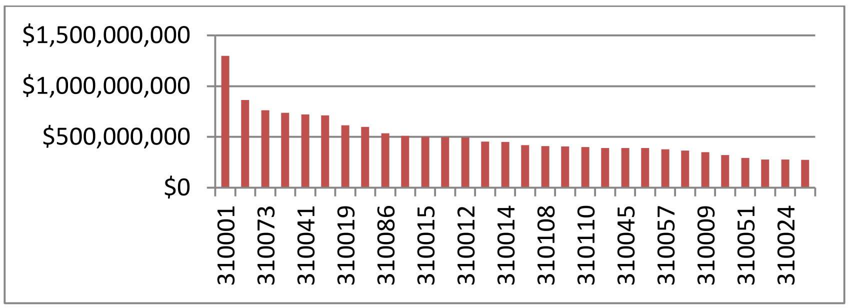


Figure 7: Bar chart

Angle and Direction

Angles are generally represented on a circle with range from zero to 360 degrees. Each angle between the zero and 360 degrees, has an implied opposite angle that completes the circle. The visual cue, angle, is commonly used to represent pie charts. A pie chart is a simple circle divided into sectors with areas proportional to the quantity it represents. It is usually used to present the frequency distributions of qualitative variables. Similar to angle is direction, but instead it relies on a single vector's orientation. It helps to determine the increases, decreases, fluctuations and slope of the data. Direction as a visual cue can be seen in time series plots, as shown in Figure 8. Time series graphs are useful for visualizing a sequence of data points, measured typically at successive points in time, with vectors fluctuating across time.

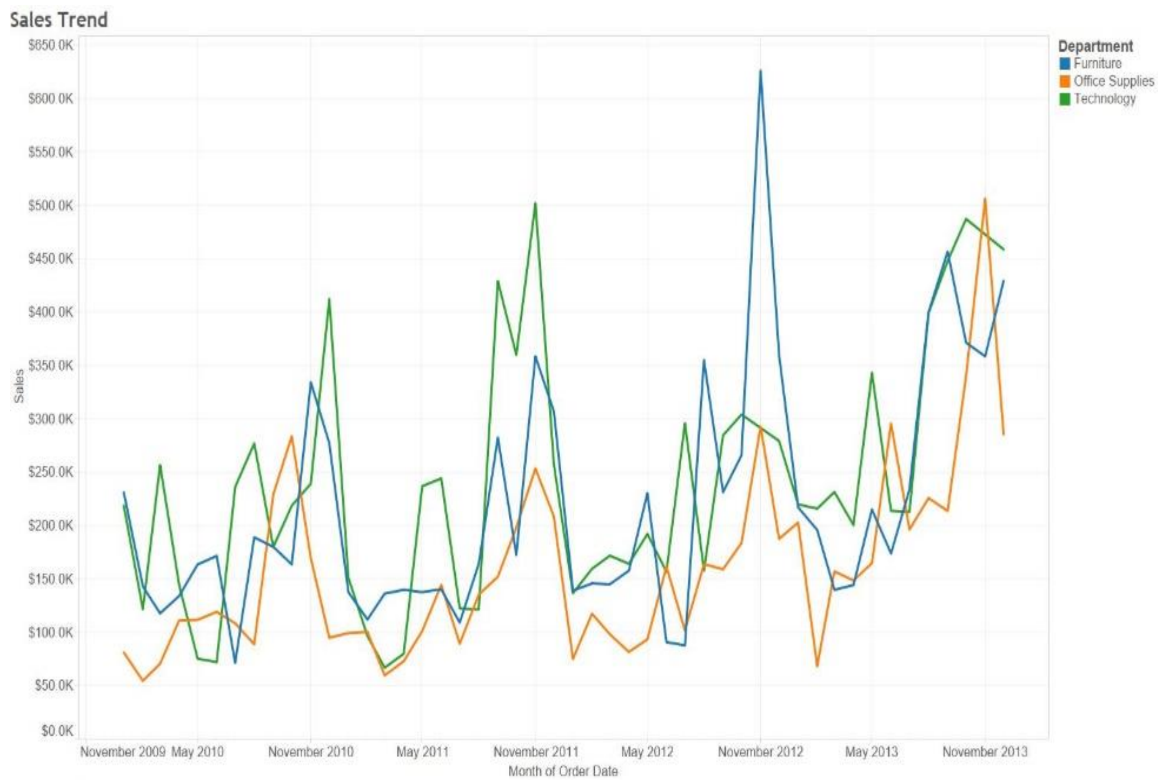


Figure 8: Time Series Visual

Shapes, Area and Volume

Shapes are commonly used to differentiate between categories. Compared to single dots on a scatterplot, the usage of different shapes can provide context, and help categorize the data. When dealing with area and volume as the visual cues, generally what that means is the greater the value is, the bigger or larger the object will be. Similar to length, area and volume can be used to represent data with size. For example, circles and rectangles can be used to represent 2-dimensions, while cubes and spheres can represent 3-dimensions. Bubble charts are one example of area and volume visual cues.

Color Hue and Color Saturation

Color can be spilt into two categories: hue and saturation. Color hue is simply color (Blue, red, green…etc.). Similar to shapes, when color is used it generally indicates categorical data, where each color represents a group. Saturation on the other hand is the amount of hue in a color (bright green to dark green). Combining both hue and saturation, visuals can have multiple hues that represent different categories, while each category can have varying scales in saturation. Heat maps are a good example of graphs that combine the previous 3 visual cues, namely area, hue and saturation. As shown in Figure 9, heat map values are presented in a matrix and represented by hue and saturation. Size or area of the individual boxes also represent an additional attribute. Heat Maps can help to quickly identify outliers in the data.

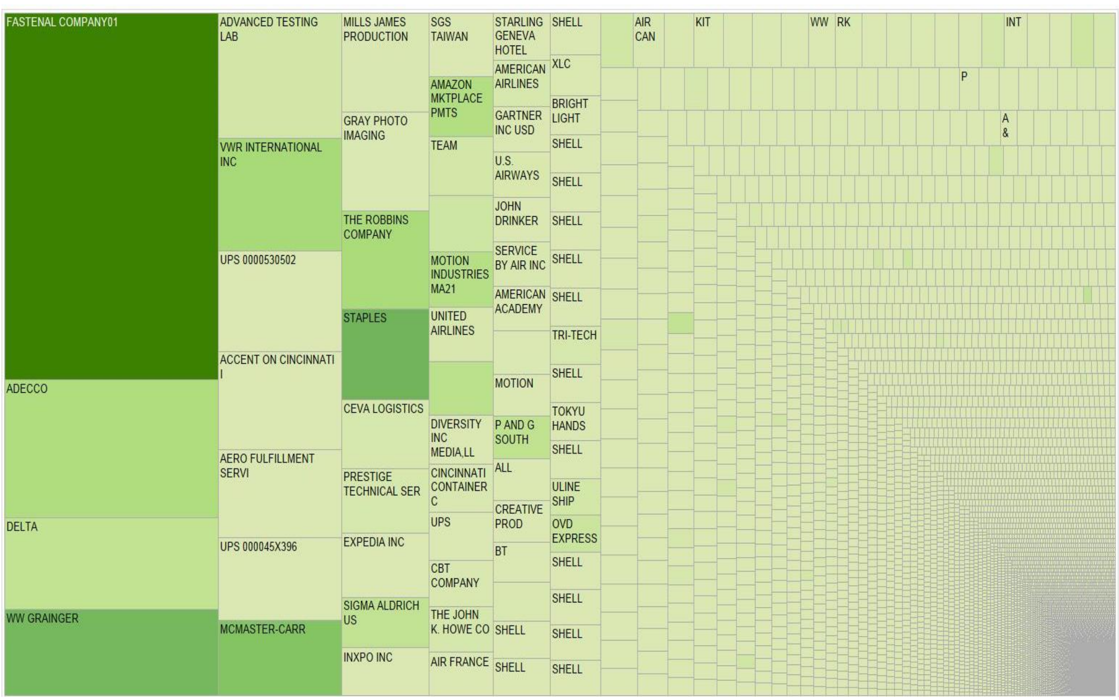


Figure 9: Heat Map Visual

Geographical Maps

In some instances, visual data points that stem from visual cues need to be placed somewhere. Generally what is used are charts or simple X-Y axis. However, when location data points need to be represented, a geographic map is required. Geographic visuals are very helpful when dealing with addresses, zip codes, or even cities. They can be combined with different visual cues, such as bubble charts and/or hue/saturation. For example in Figure 10, we see a geographical map showing a company's international sale revenues represented by bubble size and color intensity.

Figure 10: Geographical Visualization

Cleveland and Mcgrill (1986) conducted a study to determine how accurately people read the visual cues discussed above (excluding shapes, color and geographical maps). Their results ranked the visual cues from most accurate to least accurate, as shown in Figure 11. However, this ranking doesn't mean that scatter plots are always better than bubble charts. It may serve as an initial guide, but it will still depend on the task at hand and what data type is available.

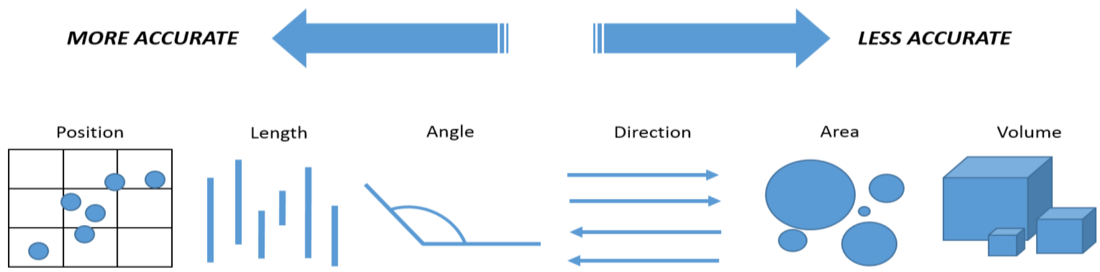


Figure 11: Ranking of Visual Cues

There are several data types, such as quantitative, ordinal, categorical, and relational. However, knowing which visual cue is suited best for which data type is a challenge. One solution to help identify which visual cue is suited for a specific task is to see whether their property is naturally ordered, and how many distinct values the reader will be able to perceive and differentiate (Steele & Iliinsky, 2011). For example, the visual cue position has a natural ordering, same with area and volume, however, shape doesn't. Hue is not naturally ordered as well, however saturation is. As for the distinct values, users need to be able to easily differentiate between the different features of visual cues. For example, hue has a limited capacity, at least in terms of our visual capabilities, the more colors are introduced the harder it is for our eyes to tell them apart. However, shapes can be more easily differentiated, as well as positions on an axis.

The variety of visual tools that are available can potentially make it difficult to determine the best tool to use. Steele and Iliinsky (2011) present a guide as to how to select certain visual properties based on the kind of data type available. As shown in Figure 12, some visual properties are limited to few data types, while other visual properties can be used to visualize multiple data types. For example, the visual properties of position, placement, and text, can be used to visualize any type of data.

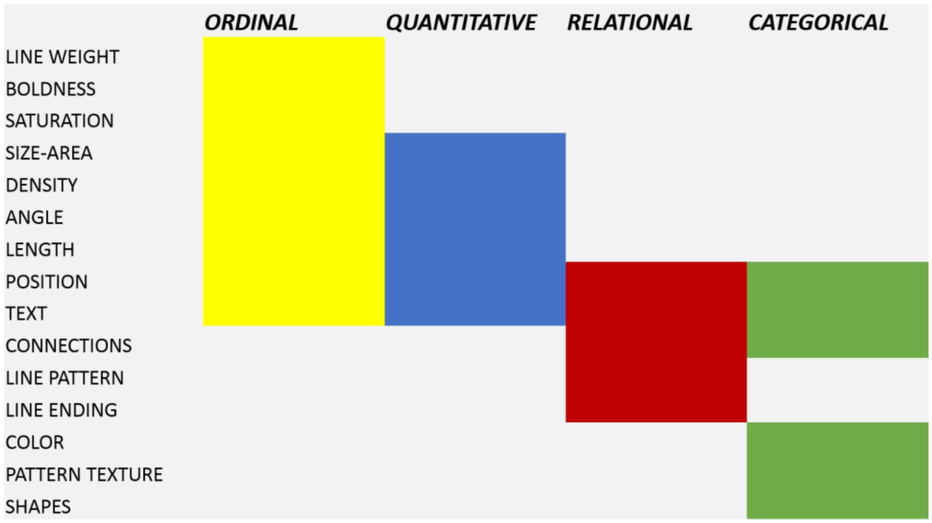


Figure 12: Visual Properties vs Data Types

2.3. THE EVOLUTION OF DATA VISUALIZATION

Data visualization has deep roots in history from the earliest map-making to the later fields of statistics, medicine, and engineering. Throughout history, a wide variety of advancements have contributed to the widespread use of data visualization today. One of the first known graphics is shown in Figure 13 which is a chart of planetary movements introduced in the 10th century. Despite being about one thousand years old, it is intuitive and easy to understand. It shows the position of several celestial bodies over time, possibly making it the earliest version of what is called a multiple times series chart (Beniger & Robyn, 1978).

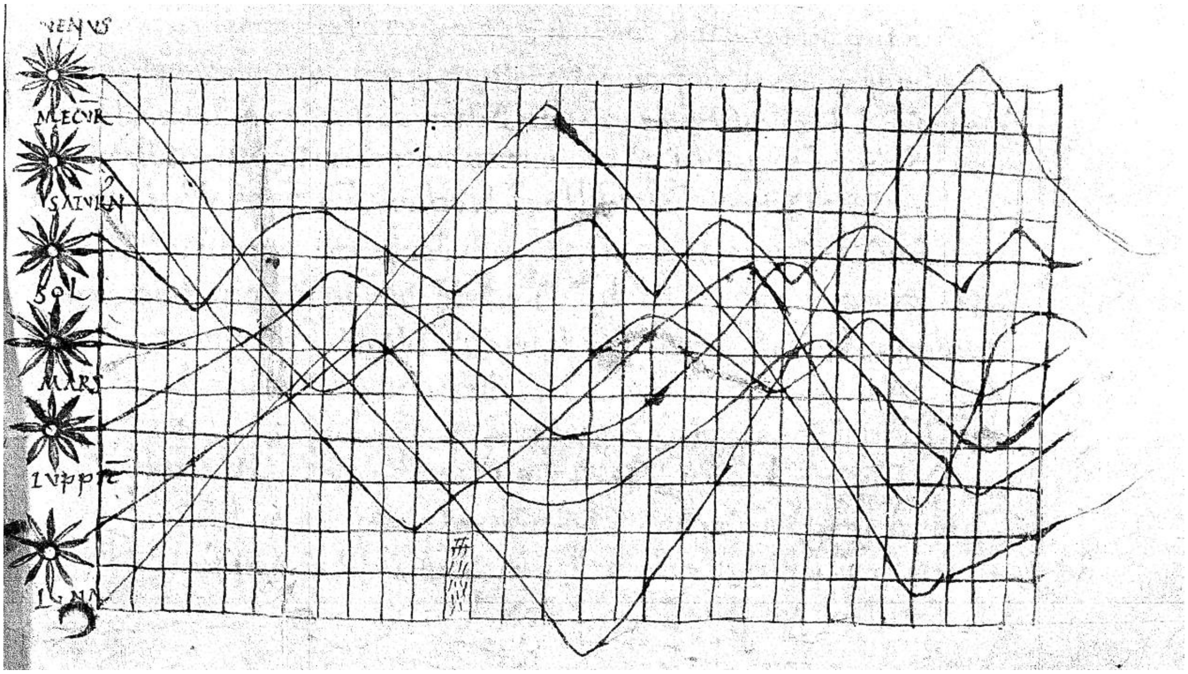


Figure 13: One of the earliest forms of data visualization

Friendly (2008), in Figure 14, showcases the development of data visualization throughout the ages. The following paragraphs will discuss this development in more details. In the 16th century, the earliest forms of data visualization arose in geometric diagrams, graphical positioning of stars, and in the making of maps to aid in navigation and exploration. These early steps comprise the beginnings of data visualization. By the 17th century, physical measurements of time, distance and space for fields such as astronomy, map making, navigation, and territorial expansion began to appear (Friendly, 2005). Figure 15 shows an example by Michael Florent van Langren, a Flemish astronomer to the court of Spain, who in 1644, believed to be the first to present a visual representation of statistical data, graphing the determinations of distance, in longitude, from Toledo to Rome (Tufte, 1997).

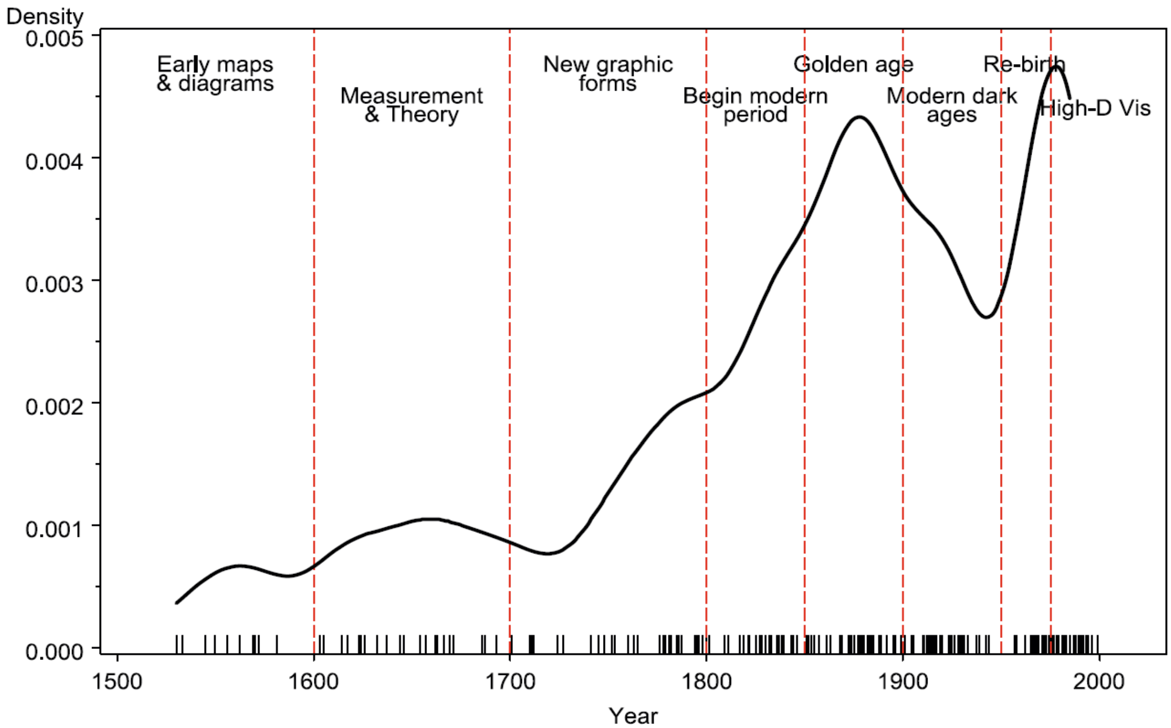


Figure 14: Time distribution of events considered milestones in the history of data visualization

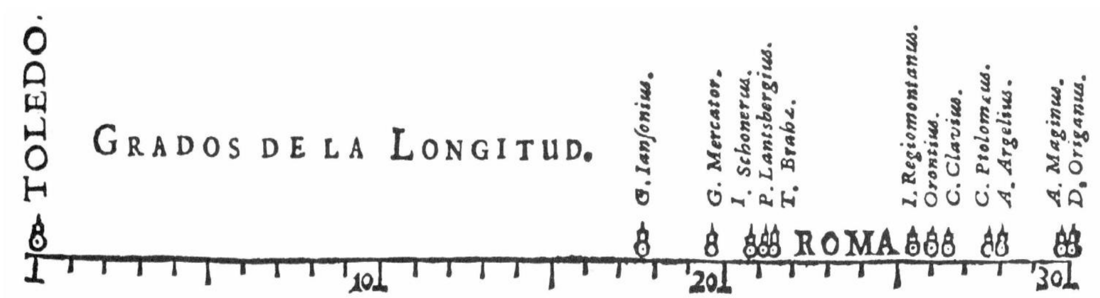


Figure 15: Langren's 1644 graph of determinations of the distance, in longitude, from Toledo to Rome

Along with statistical theory in the 18th century, abstract graphs, and graphs of functions became more widespread. For example, in cartography, the study and practice of making maps, new data representations, such as isolines and contours, were invented, as well as introduction of the concept of thematic mapping (Friendly, 2005). In that century, some of the most widely used graphics today were invented. One notable Figure was William Playfair, a Scottish engineer and political economist. In 1786 playfair created what is considered by many to be the first bar chart as shown in Figure 16.

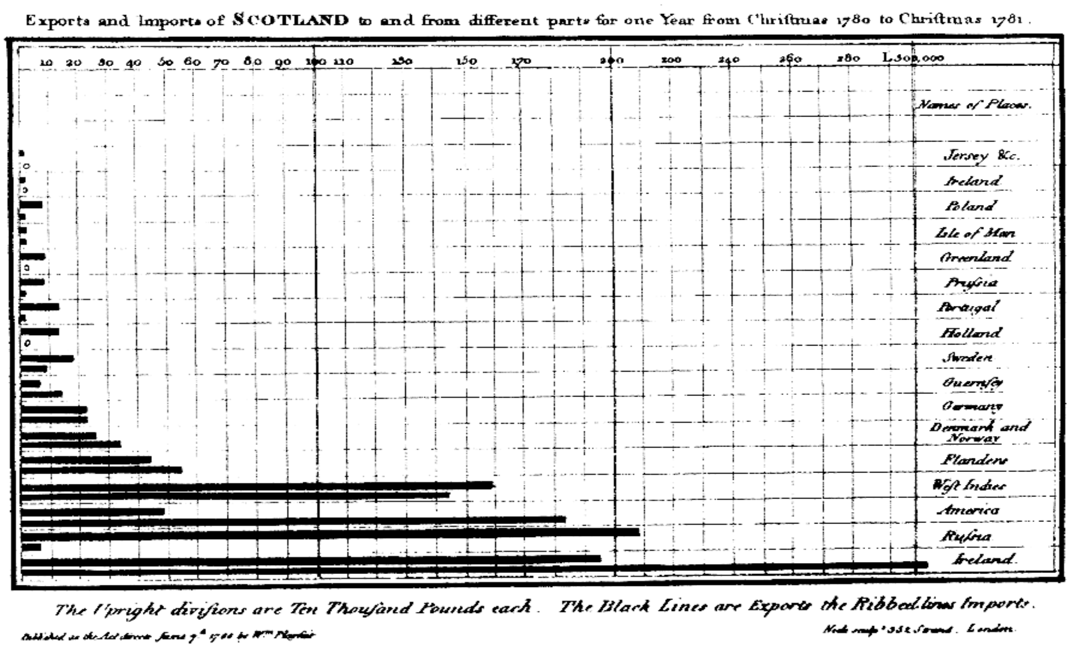


Figure 16: Bar chart created by Playfair in 1786

Ironically, Playfair only created this chart out of necessity, since he did not have time dimensional data to construct a line graph. The graph shows a series of 34 plates about the imports and exports of 17 different countries over the years. However, since Playfair lacked the time series data for Scotland, he graphed its trade data as a series of 34 bars, one for each of 17 trading partners for a single year (Playfair, 1786; Playfair, 2005; Spence, 2006).

Later in 1801, Playfair introduced the pie chart and circle graph (Playfair, 1801). Figure 17 presents a more sophisticated graph by Playfair, whereby he used three parallel time series to show the price of wheat, weekly wages, and reigning monarch from the year 1565 to 1820. Based on his graphical representation, Playfair argued that workers had become better off in the most recent years (Playfair, 1821).

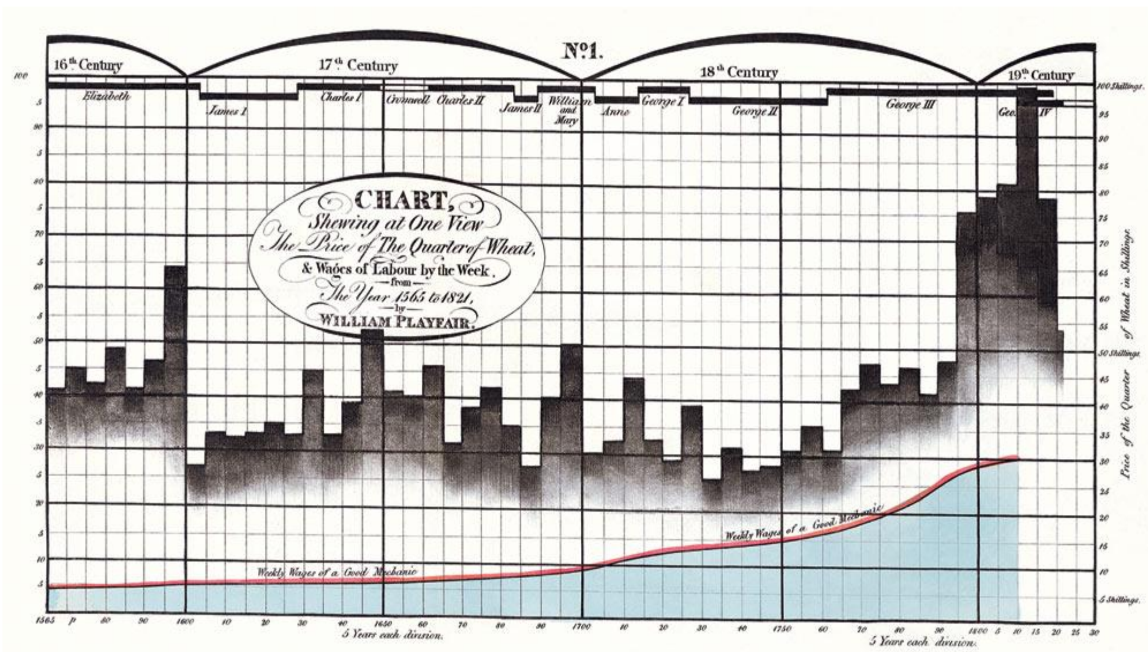


Figure 17: A time-series graph of wages, prices, and reigning ruler over a 250 year period.

By the 19th century, an explosive growth in statistical graphics and thematic mapping emerged. Many of today's modern forms of data visualizations were developed, such as bar and pie charts, histograms, line graphs, time-series plots, scatterplots and so forth (Friendly, 2005). Around the 1820s, Baron Charles Dupin, a French mathematician, introduced the use of continuous shadings to show the distribution and degree of illiteracy in France, and is considered the first known instance of what is called a choropleth map today (Dupin, 1826).

By the mid-1800s, all the conditions for the rapid growth of visualization had been established, and what was known as the "Golden Age" of statistical graphics began. Figure 18 presents a good example of how data visualization was at the time of the golden ages. The graph uses polar time-series diagrams on a map of France to show the changes in population by department at the 5-year intervals of the national census from 1801 to 1881 (Friendly, 2008).

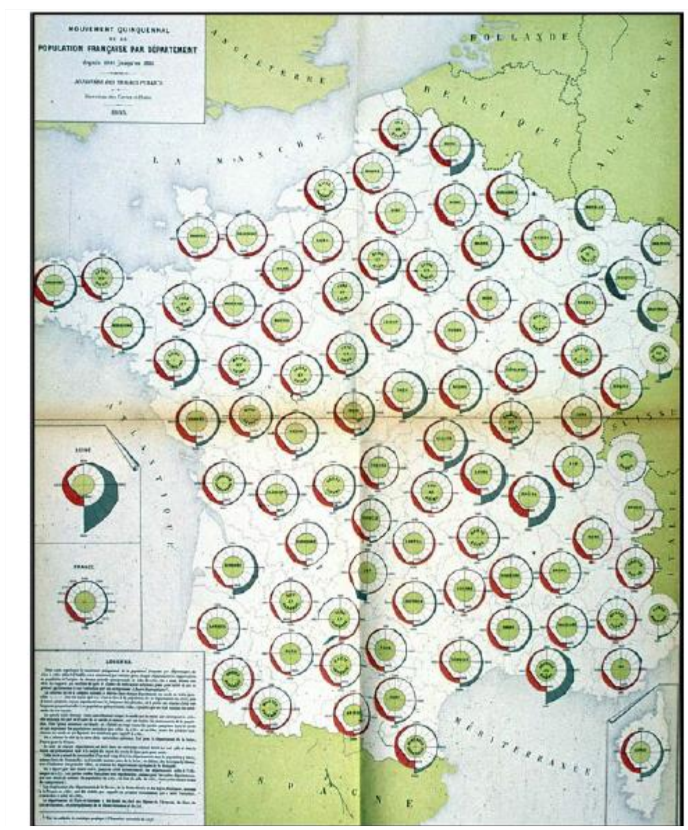


Figure 18: A polar time-series diagram on map of France

Opposed to the “Golden Age” of statistical graphics in the 1800s, the early 1900s could be called the “modern dark ages” of visualization (Friendly & Denis, 2000). The enthusiasm for visualization had largely been replaced by the rise of quantification and statistical models in the social sciences. Many statisticians believed that graphs and pictures were suggestive, and often incapable of stating a fact to three or more decimals. Numbers and parameter estimates, on the other hand, were precise (Friendly, 2008).

However, in the mid-1960s, data visualization began to spur and rise from dormancy largely due to three significant developments. These developments began in 1962, when John Tukey, an American statistician, issued a call for the recognition of data analysis as a legitimate branch of statistics, and by 1977, he introduced exploratory data analysis (EDA), were he proposed a wide variety of new, simple and effective graphic methods. (Tukey, 1962, 1977). In France, Jacques Bertin, a French cartographer and theorist, published S´emiologie Graphique (Bertin, 1967, 1983), were he showed how to organize the visual and perceptual elements of graphics according to the features and relations in data. Finally, by the 1980s, computer science research, along with the developments in data analysis (e.g. EDA) and the introduction of display and input technologies (graphic terminals, digitizer tablets, etc.), would provide new paradigms for implementing data visualizations and lead to an explosive growth in new methods and techniques. Specifically, with the introduction of high-dimensional data, new techniques such as scatterplot matrix (Tukey & Tukey, 1988), and parallel coordinates plot (Inselberg, 1985; Wegman, 1990) were developed.

2.4. DATA VISUALIZATION IN AUDITING

Graphs and visuals have been used extensively in business to help detect changes in trends not easily identified by standard statistical models; examples include those of the Dow-Jones Index, money supply, unemployment rates, and accounting numbers like income, rate of return, and sales (Moriarity, 1979).

In the accounting and auditing literature, prior research has examined the importance of presentation format and its linkages to decision-making performance. They put focus on the comparison of different visual techniques and their impact on decision making. Furthermore, the growing number of studies examining presentation format provide an indication of its importance on decision-making (Vessey & Galletta, 1991; Ramarapu et al., 1997; Frownfelter-Lohrke, 1998; Dull & Tegarden, 1999; Hodge, 2001; Dull et al., 2003; Hodge et al., 2004; Hodge & Pronk, 2006).

Some of these studies compared multi-dimensional graphics with tabular presentations, and by using schematic faces concluded that the multi-dimensional graphic presentations were more efficient in communicating financial information than tabular presentations (Moriarity 1979). Some also concluded that human judgment accuracy can be influenced by the accounting report format used (Stock & Watson, 1984). Other accounting research took the idea of presentation format from a different perspective. They examined the relationship between presentation format and decision quality. Wright (1995) found that auditors benefited from graphical formats in their financial judgments, while Anderson and Reckers (1992) found that visuals and graphs aided the subject's judgments in assessing correlations in analytical procedures. However, not all prior studies found similar results. Nibbelin et al. (1992) did not find much difference over the use of graphics over tabular formats for a bond-rating change decision. Bricker and Nehmer (1995) found that graphics did not have an impact on accuracy in evaluating financial ratios and Blocher et al. (1986) found that visual formats did not increase accuracy for more complex tasks.

Numerous studies have also examined the manner in which data visualization and presentation format affects the judgment development process with respect to analytical procedures (Kaplan 1988, Anderson & Reckers 1992, Anderson & Kaplan 1992, Schulz & Booth 1995). Analytical procedures provide important evidence for the initial planning, substantive tests, and overall review of an audit. However, according to the environmental complexity model (Schroder et al., 1967), dealing with large amounts of information may impair performance in decision making. Therefore, when applying analytical procedures to today's Big Data, auditors may face complex scenarios, and may wish to present the data in some organized visual format that may facilitate the decision making process.

However, the studies mentioned above examining the effect of presentation format on analytical procedures, also found varying results. Kaplan (1988) used trend analysis as the experimental task and found that presentation format had no effect on the accuracy of assessing expected sales dollars. Anderson and Reckers (1992) extended Kaplan's study by examining more complex tasks. In their study, participants with graphs performed significantly better than participants with tables. A more recent study by Anderson and Mueller (2011) found that participants using graphs perform significantly better both in assessing correlations and in making predictions during analytical review of the sales account. This performance advantage was found across experience levels for students and auditors.

Frownfelter-Lohrke (1998) suggests that one of the primary reasons for such conflicting results in prior studies is the lack of a theoretical basis in explaining the relationship between presentation format and the task at hand. Vessey (1991) attempted to resolve this by directly relating task type to presentation format. She came up with a theory that suggests when the problem solving aids, including visual graphics, support the task strategies required for that task, then we have what is said to be a “cognitive fit”, which will ultimately lead to effective and efficient problem-solving performance. This theory was named the theory of Cognitive Fit (Vessey, 1991). Frownfelter-Lohrke (1998) study failed to support the theory of cognitive fit, and suggested that these inconsistent results may be a function of task complexity.

2.4.1 Visual Literature Analysis of Data Visualization Research in Auditing

This section attempts to present a methodology by which data visualization can be used to analyze the auditing literature in terms of their usage and research of data visualization. Data visualization has been researched extensively in the past. The auditing and accounting literature has allotted a great amount of resources in understanding the effects of different data visualizations techniques in decision making and analytical procedures. However, as technology is evolving, and the size and volume of data is constantly increasing, new ways to present information are emerging, and as this continues to develop, it is important for accounting and auditing research to examine newer data visualization techniques. Therefore, understanding the types of research conducted in the past relative to data visualization is crucial in bridging a path towards more relevant research in the future.

The data used to analyze the accounting and auditing literature consists of 35 publications in the related research area. Table 1 provides a description of the data collected for the analysis.

Table 1: Literature analysis data set

Journals	Description
Accounting and Finance *Accounting Education* *Accounting Horizons* *Accounting, Organizations and Society* *Advances in Accounting* *Auditing: A journal of practice and theory* *Communications of the ACM* *Contemporary Accounting Research* *Information Systems Research* *Journal of Accounting Research* *Journal of Information Systems* *Journal of Management* *Journal of Management Information Systems* *Management Science*	A total of 35 Publications From 15 Different Journals Ranging from year 1979 to year 2011

The research publications were searched from the Business Source Premier - EBSCO Publishing website. A combination of keywords were used to obtain the papers. The primary keywords were: Data visualization and auditing/auditors, visualization and auditing/auditors, presentation format and auditing/auditors, tables versus graphs and auditing/auditors. The papers that were included in the data set are those that relate to data visualization in 3 ways: 1) they investigate the effects of data visualization and/or presentation format on decision making. 2) They investigate the effects of data visualization and/or presentation format on auditor judgments, and finally 3) they investigate the effects of data visualization and/or presentation format on outcomes of analytical procedures. Those publications that do not fit the criteria were not included. The final research papers were read thoroughly, and the relevant information was extracted. Specifically, the author name, the journal name, the date or year of publication, the major findings, and whether the author(s) find results that's support the use of data visualization, don't find support, or find varying results. Appendix 1 provides a table with a summary of the data collected.

The analysis of the auditing literature is presented using two types of visualizations. Figure 19 presents the first visual which shows a trend analysis of data visualization publications over time. The graph shows the number of data visualization studies from year to year, starting from 1979, with Moriarity's paper "*Communicating financial information through multidimensional graphics*". In his paper, he found that his subjects made more accurate predictions with multidimensional graphics than with the direct use of financial ratios or other financial information.

Despite a slow start of research, the number of publications generally maintained a consistent pace, as seen by the dotted trend line. However, around 1991, data visualization research reached a peak. One can speculate that this phenomena can relate to the introduction of the theory of Cognitive Fit. Vessey (1991) researched the relationship between task type and presentation format and came up with the theory of Cognitive Fit. It can be assumed that as she developed her theory, researchers began to heavily study the effects of when visual graphics support the task strategies required for that task, and hence the increased amount of publications.

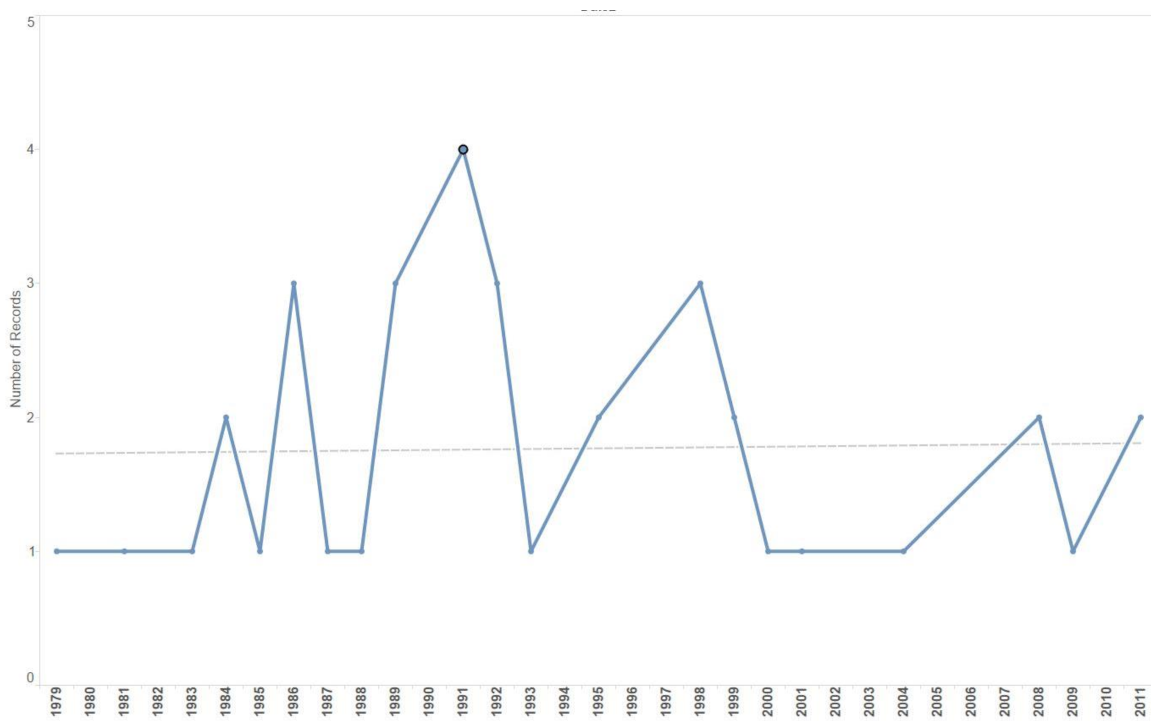


Figure 19: Trend analysis of publications over time

Figure 20 presents the second visual analysis showing a scatter plot of the papers across time. The plots also categorizes each paper in terms of whether the main findings support the use of graphics or not, or whether they are inconclusive. Inconclusive means that the results of the study could have gone either way, in that for certain tasks, data visualization was useful, and in others it was not. The color coding used is: Green for supporting graphics, red for not supporting, and yellow for inconclusive results.

Starting from 1979, Moriarity's paper that presented the usefulness of data visualization is shown. However, after this paper, and up until the early 1990's, researchers have found little support for the usage of data visualization compared to other methods, such as tables and ratios. This again can relate to the fact that during this time, little was known about the effects of the judgement task and how the theory of Cognitive Fit may apply. However, looking around the center, and specifically around the year 1991, there are a large number of research publications. Again, it can be speculated that due to the introduction of the theory of cognitive fit, researchers started realizing how data visualization can be useful for certain tasks. For example, Hard and Vanecek (1991) found that financial predictions were more accurate for graphical presentations during an estimation task but more accurate for tabular presentations during an accumulation task. Similarly, Amer (1991) found that tabular presentation resulted in more accurate assessments of debt covenant violations in a selective task but had no effect on predictions of bond ratings during an integrative task.

Figure 20: Scatter plot of major findings in Data visualization literature across time

Evidently, as time progressed and researchers started to further understand the effects and usefulness of data visualization, most publications supported the use of graphics. This can be attributed to several factors. First was due to the theory of cognitive fit. Researchers now have an understanding of how task environment can affect the outcome when using graphics. Furthermore, compared to before, we see a new breed of applications for data visualization, applications that have flexibility and provide multiple functionality. Finally, the vast amount of data that has been generated and gathered over the last several years has influenced researchers in acknowledging the benefits of using data visualization.

2.5. DISCUSSION AND CONCLUSION

Prior research in auditing and accounting has long recognized the importance of presentation format to decision making (O'Donnell & David, 2000). However, despite the rise of ERP systems, the advances in the accumulation of accounting data, and the appearance of Big Data, the presentation format of accounting information has not kept up with the advancements (Dull & Tegarden, 1999), and such increasing complexity can simply overwhelm the decision-maker.

Nevertheless, data visualization has been successfully deployed in medicine, genetics, biology, engineering, and many other scientific fields, and despite the availability and use of visualization tools in these fields (Kraemer & Ferrin 1998; Montgomery et al.,2004; Sinha et al.,2002; Trelease, 2002), business applications are said to lag the visual sciences by as much as ten years (West, 1995). Auditing on particular has been behind in the usage of data visualization. As mentioned before, the auditing and accounting literature has put more focus on the comparison of different visual techniques and their impact on decision making, particularly limited to the traditional “graph versus table” studies. Not only that, but most of the visual techniques studied in the literature are primitive compared to what is available today.

However, there are only a handful of studies in the accounting and auditing literature that look at more recent forms of visualization techniques. For example, Dilla et al., (2012) studied the impact of graphical displays of pro forma earnings information on professional and nonprofessional investors' earnings judgments, while Elliott et al. (2011) studied the influences of online video to announce a restatement on investment decisions and the mediating role of trust. Nevertheless, what is really lacking in the literature are examples of how certain data visualization techniques can be helpful in the field of auditing. How they can aid auditors in the audit cycle, from the planning stage, to fieldwork, and finally to reporting the results to management.

Hence, it may be acknowledged that the research is lagging behind in innovation, and therefore this dissertation attempts to contribute to that literature by illustrating different methodologies by which data visualization can be applied. This essay specifically, contributed by presenting a methodology by which data visualization can be used to analyze the auditing literature, and provide understanding of their research of data visualization, in attempt to bridge a path to more relevant research in the future.

CHAPTER 3: THE APPLICATION OF DATA VISUALIZATION IN MEDICARE HEALTH INSURANCE FOR THE PURPOSE OF KNOWLEDGE DISCOVERY AND COMMUNCATION

3.1. INTRODUCTION

3.1.1. The Medicare Healthcare Program

King (2010) views Medicare as a high-risk program, at least partially because its size and complexity create opportunities for abuse, waste, and fraud. Daly and King (2011) indicate that estimated improper Medicare fee-for-service payments were $35.4 billion in 2009. More importantly, the estimate for 2010 ballooned to $48 billion. This suggests that problems are significant and expanding within the Medicare system, and that solutions are desperately needed to both detect previous errors and/or irregularities as well as prevent improper payments in the future. Moreover, due to the nature of these programs, researchers have found that the fundamental reason why such systems are targeted for fraud is that the industry´s standard control and detection systems are not aimed at criminal fraud (Sparrow, 2008).

There are many examples of healthcare fraud that can be hidden in many different forms. According to a 2012 report by the Centers for Medicaid Services (CMS) fraudulent activities include billing multiple times for single procedures, billing for procedures not carried out, billing for more than what the procedures actually cost, running excessive expensive tests, implementing ineffective treatments, filing false cost reports, providing kickbacks for referrals, and substituting drugs for cheaper generics. In the U.S., healthcare fraud is investigated by the Centers for Medicare and Medicaid Services (CMS), the Federal Bureau of Investigation (FBI), and the U.S. Attorney General, State District

Attorneys, as well as by other private sector organizations. Fighting healthcare fraud more effectively would help reduce unnecessary costs and free up resources to promote public health and provide better patient care. Healthcare auditors can leverage data analytics to detect and prevent fraud and abuse across the healthcare industry, including health plans, hospitals, and pharmacies. However, before auditors can identify fraud and abuse, it is necessary to effectively manage the healthcare data and understand its format and organization (Wiedemann, 2014).

The vast amount of data generated by these healthcare insurance providers today is far too complex and voluminous to be processed and analyzed via traditional methods. Consequently, auditors must rely on advanced techniques to find and track offenders (Koh & Tan, 2011). Data visualization is one solution that can help find useful and valid information in large volumes of data. It consists of a collection of descriptive and graphical statistical tools for discovering patterns in data, and facilitating hypothesis development and refinement (Tukey, 1977). One advantage of this method involves user interactivity, whereby a human is actively engaged in the data exploration process, leveraging his/her abilities to perceive patterns and structures in visual representations and ultimately interpret and explain what is seen. Data visualization is especially useful when little is known about the data, and exploration goals are vague (Gahegan et al., 2001; Keim, 2002).

3.1.2. Data Visualization for Knowledge Discovery in Auditing

Analytical auditing procedures play an important role in assisting the auditor in determining the timing, nature, and extent of his or her substantive testing. According to

AU Section 329, analytical procedures should focus on two main factors: 1) enhancing the auditor's understanding of the client's business in terms of the transactions and events that have occurred, and 2) identifying areas of risk that might be relevant to the audit (AICPA, 2001).

There are many types of techniques that can be used to conduct analytical procedures. Blocher and Patterson (1996) identified three general types of analytical auditing techniques, including ratio analysis, trend analysis and model-based procedures. However, with the advent of Big Data, such as that generated from Medicare, the audit scope has substantively expanded. Big Data now allows for a more practical exploration of data to develop testable assertions (Liu, 2013). It has also given rise to new audit techniques, such as pattern recognition, data visualization, and natural-language processing, and created many new forms of audit evidence, such as alarms/alerts, text mining, continuity equations, and exceptional exceptions (Vasarhelyi & Halper, 1991, Kogan et al, 2011, Issa, 2013).

Auditors must also be able to interpret multiple data points and relationships within Big Data, and this can be challenging and time consuming. Additionally, when dealing with Big Data, auditors will not only need to extract and analyze large amounts of data, but will still need to interpret, understand, and explain and communicate findings, and this may be problematic because of the overwhelming amount of data that will still be generated (Steele & Iliinsky, 2011). Further, traditional audit methods and tools may not always be suitable to effectively and efficiently analyze Big Data. For example, when looking at fraud risk assessment, traditional audit methods, such as fraud risk checklists and standard planning programs, have not always been very effective in identifying fraud risks (Eining et al., 1997; Asare & Wright 2004).

One solution to address these challenges is data visualization, which can be useful in analyzing and evaluating Big Data. Data visualization can help auditors identify relationships between different attributes in a multidimensional dataset, potentially producing better fraud risk assessments (Humpherys et al., 2011). It is a technique that accommodates large datasets, and can further assist in the process of knowledge discovery. Data visualization can also help pinpoint areas where further inquiries are needed to gain a better understanding of the underlying transactions. For example, certain outliers in the data may be visible and/or questionable items may need further investigation. Therefore, this ensures that the auditors are not only asking the required questions but also asking the right ones. Data visualization has the potential to enhance the efficiency and effectiveness of audit procedures. By applying different data visualization techniques, auditors can identify unknown patterns or relationships between the data, and communicate findings to stakeholders.

Auditors are likely to benefit from data visualization throughout the audit cycle. By using data visualization, auditors can understand the entity and its environment, identify and assess business risks and test for internal control weaknesses and/or fraud with regard to those assessed risks. Figure 21 illustrates how and where data visualization can be applied by auditors throughout the audit cycle, including the planning, fieldwork and reporting stages. In the planning stages, auditors start by gaining a general understanding of the overall processes. They also conduct risk assessments to identify any potential material weaknesses, and check overall data integrity and validity. In this stage, data visualization can be used as a tool for knowledge discovery, producing descriptive visuals that would help auditors further understand the data and business, and identify any potential risks in order to create an effective audit plan.

When conducting fieldwork, auditors may utilize data visualization as a supplement to their analytical procedures and other forms of evidence collection. Visualization techniques such as clustering, network analysis, box plots, scatter plots, and geo graphical analysis can help auditors analyze the entire population of data. Such techniques allow for faster data exploration in highly diverse and noisy data, and are extremely useful in detecting pattern violations and potential outliers in the data

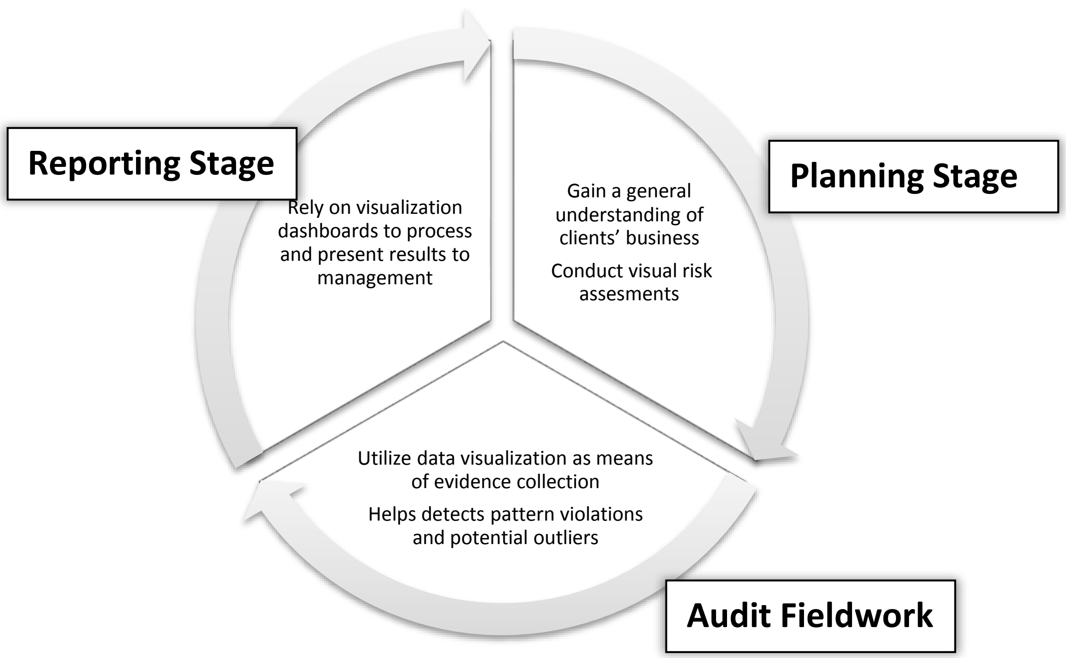


Figure 21: Application of Data Visualization throughout the Audit Cycle

In the final, reporting, stage of the audit cycle, auditors may rely on explanatory data visualization, such as dashboards, to process and present results to management. Dashboards allow users to switch between alternative presentation formats, filter down to the necessary information required, and obtain a general summarization of the results in one view.

In this study, the focus will be on the use of data visualization to assist auditors in the planning stage of the audit cycle. Prior studies have utilized this technique in variety of areas as means of knowledge construction (Keim & Kriegel, 1994; Card et al., 1998; Ribarsky et al., 1999; MacDougall, 1992; Lee & Ong, 1996; Keim & Kriegel, 1996). However, limited research has been carried out concerning the uses of data visualization as an enabling technology for comprehensive knowledge construction in the audit planning stage. Hence, the aim and contribution of this study is to help bridge this gap by relying on data visualization for knowledge discovery. Specifically, this study will illustrate how data visualization techniques can be used to assist auditors in exploring and explaining large datasets, finding relations, detecting patterns, and communicating results during the planning stage of their audit.

3.2. METHODOLOGY

Visualization leverages the capabilities and bandwidth of the human visual system. This enables the transfer of large amounts of information into our brain. Visualization can help in deeper exploration, identify sub-problems, give rise to new questions and identify trends and outliers within the data (Steele & Iliinsky, 2011). However, for data visualization to be successful, it must be novel, informative, and efficient. An important aspect is novelty; data visualizations should offer a fresh look at the data and present results in a new level of understanding. Additionally, visuals should be able to convey information so that the decision maker may gain knowledge. Finally, data visualizations need to be efficient. They should have a clear goal or message without sacrificing necessary information or increasing complexity (Steele & Iliinsky, 2010).

As discuss in the first chapter, there are two general categories of data visualization: explanation and exploration. In general, the exploratory part of data visualization is used as part of the initial analysis. Focused on finding insights, patterns, and drawing conclusions. On the other hand, the explanatory part of data visualization is typically used to present the resulting insights and conclusions discovered during the analysis to clarify or illuminate any relationships in the data.

The knowledge development process is comprised of five stages: data selection, data pre-processing, data transformation, data mining, and interpretation/evaluation (Fayyad et al., 1996) and visual presentations. Figure 22 depicts the six stages of the visual knowledge construction procedure. The objectives of portraying data in this visual form generally relate to stimulating pattern recognition and hypothesis generation rather than simply presenting results.

The first stage is data collection and selection. Collection refers to physically obtaining the data while selection refers to selecting the tables and/or data points required. The second and third stages are data preprocessing and data transformation. In these stages the data is preprocessed and any missing values or errors in the data are addressed. Then the data is transformed based on the needed requirements. Qualitative data can be transformed to quantitative and vice versa. Data analysis, visual presentation and data interpretation are generally interrelated. These final 3 stages are the core of data visualization.

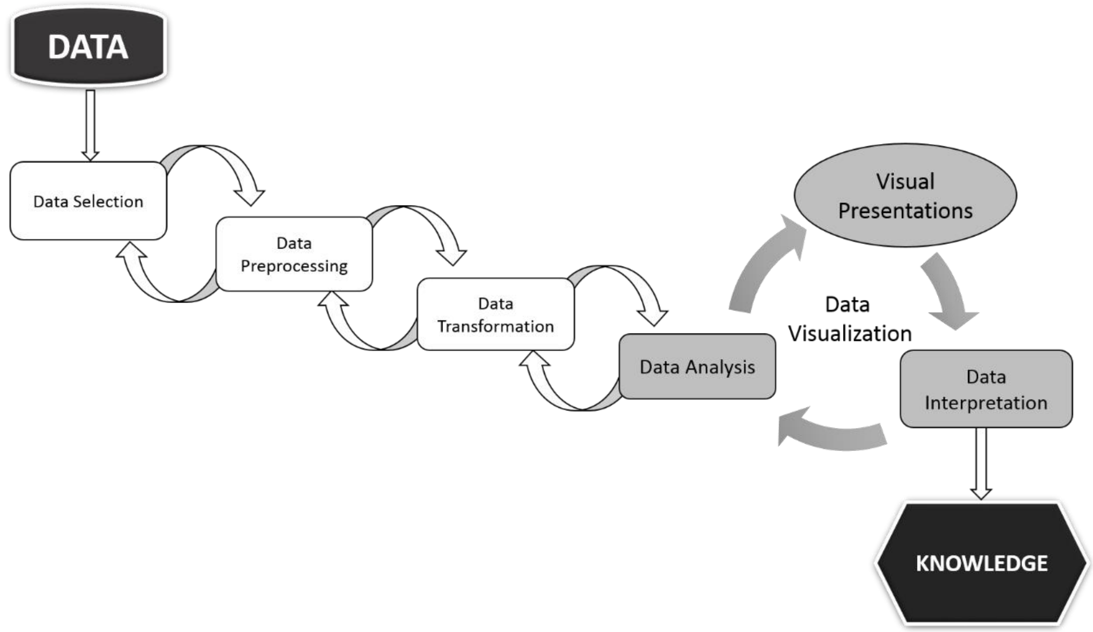


Figure 22: Stages of Visual Knowledge Construction Process

Some visualization packages are constrained to generating grids of visual presentations of uniform granularity and limited dimensionality. Other researchers suggest that for better visualizations, one should consider those enriched by extending basic graphic parameters or by employing novel and less-known visualization techniques (Tegarden, 1999). Nevertheless, the exploratory and explanatory visual analysis of Medicare data begins with tabular statistics and basic graphs, followed by descriptive visual dashboards. Lastly, the main analysis includes multiple advanced visuals and in-depth exploration.

Before conducting the in-depth visual analysis, understanding the different levels of data visualization is necessary. Data visualization can generally be considered at two main levels: syntax and semantics. Syntax concerns the actual visual marks, which include identifying marks such as labels, framing marks such as gridlines and axes, and data-representative marks such as points, lines and bars. Semantics analysis on the other hand focuses on the meaning of the visuals and the underlying data values and relations that the marks represent (Kosslyn & Cunningham 1994). Both levels are needed to effectively model data visualizations. At the syntax level, marks need to be identified. As for the semantic level, it involves the representation of data points by mapping the visual marks to the related visual cues or properties, such as shape, color, and position (Mackinlay, 1986). Tegarden (1999) suggests the use of several advanced visual metaphors for multivariate data, such as scatter-grams, and map-based bar-charts. Nevertheless, the in-depth analysis for this study will include geographical visuals, scatter plots, box plots, and trend analysis.

Furthermore, healthcare fraud is expected to be hidden in the relationships between providers and beneficiaries when making insurance claims (Chandola et al., 2013). In general, a patient will see a limited number of providers during a specific year. Additionally, a practitioner and beneficiary will normally be limited to a specific geographic area. Therefore, it is necessary to further understand such relationships and detect any unusual behavior patterns exhibited by a group of providers frequently sharing large numbers of patients. Hence, network analysis of providers and their beneficiaries will also be conducted as means to further explore this relationship.

Finally, data animation will be used. Data animation is one promising technique that facilitates perception related to changes in and between transitioning data graphics. It has been applied within statistical data graphics, examples include morphing and translation of shapes, and the movement of data points to convey change over time. Unlike static data visualization, visual motion is highly effective at attracting attention, and is easily perceived in our peripheral vision (Palmer 1999), suggesting that animation may be effectively applied to direct attention to outliers in the data. Prior research has found that animated transitions increase levels of engagement (Tversky et al., 2002), facilitate decision-making (Gonzalez, 1996) and learning (Bederson & Boltman, 1999), and may help keep viewers oriented (Robertson et al., 1991, Tversky et al., 2002). Prior literature has also studied animation visualization in terms of transitions between one view to another (Bederson & Boltman,1999), transitions of data from one state to another (Robertson et al.,1991; Robertson et al.,2002; Heer & Robertson 2007), and illustration of trends (Rosling, 2009; Robertson et al.,2008). Gapminder, developed by Rosling (2009), is one example that relates to the illustration of trends and presents major global developments using animated statistics and colorful graphics.

3.3. DATA

Medicare is a health insurance program administered by the United States federal government. The program provides healthcare access and assistance for individuals 65 or older, under 65 with certain disabilities, and with End-Stage Renal Disease. The Medicare program typically maintains three types of data for their operations. The claim information that captures information such as nature and cost of the service. Patient data that captures demographic and eligibility information about the patients. Finally, provider data that includes information about physicians, hospitals, and other healthcare providers (CMS, 2012).

The Medicare program follows a prospective payment system (PPS) for hospital care, meaning that providers are paid for services at predetermined rates. Therefore, if the costs of the actual service is more than the allowed cost, the provider will be responsible for the difference. However, if it happens to be the other way around, the provider will end up keeping the difference. Such system forces some providers to charge for unnecessary services, for example, they may make an extreme diagnosis to safeguard against any potential losses (Chandola et al., 2013).

The data obtained for the purpose of this study consists of a primary data set and a supplementary data set. The primary data set was from 4 main sources. The first relates to Medicare patient claims and provider details for the 2010 fiscal year. The second and third relates to the National Provider Identifier (NPI) details and ICD-9-CM Diagnosis and Procedure Codes from the Centers for Medicare and Medicaid Services. The final source is the New Jersey census data for the year 2010. The 4 sources were then merged into one comprehensive data set. Figure 23 provides the relationships between the different tables. The NJ census table was linked to the Medicare patient claims and provider detail table using the “county” attribute. The ICD-9-CM Diagnosis and Procedure Code table was linked to the Medicare patient claims and provider detail table using the “diagnosis code”. Finally, the National Provider Identifier table was linked using the relevant “NPI” codes.

A supplementary data set was also constructed focusing on the number of major hospitals in different counties in NJ and the number of staffed beds for those major hospitals. The supplementary data set was obtained from two different sources. The first is the Department of Health for the State of New Jersey[1], which deals with data related to the number of major hospitals per county. The second relates to the American Hospital Directory (AHD[2]), which provides data and statistics about more than 6,000 hospitals

[1] http://web.doh.state.nj.us/apps2/hpr/county.aspx?by=county
[2] http://www.ahd.com/states/hospital_NJ.html

nationwide, and includes information such as the number of staffed beds per hospital and the total number of patient days. This supplementary data set will be used in the exploration processes to provide further understanding. Appendix B provides a table compiling the primary and secondary data sources and their attributes.

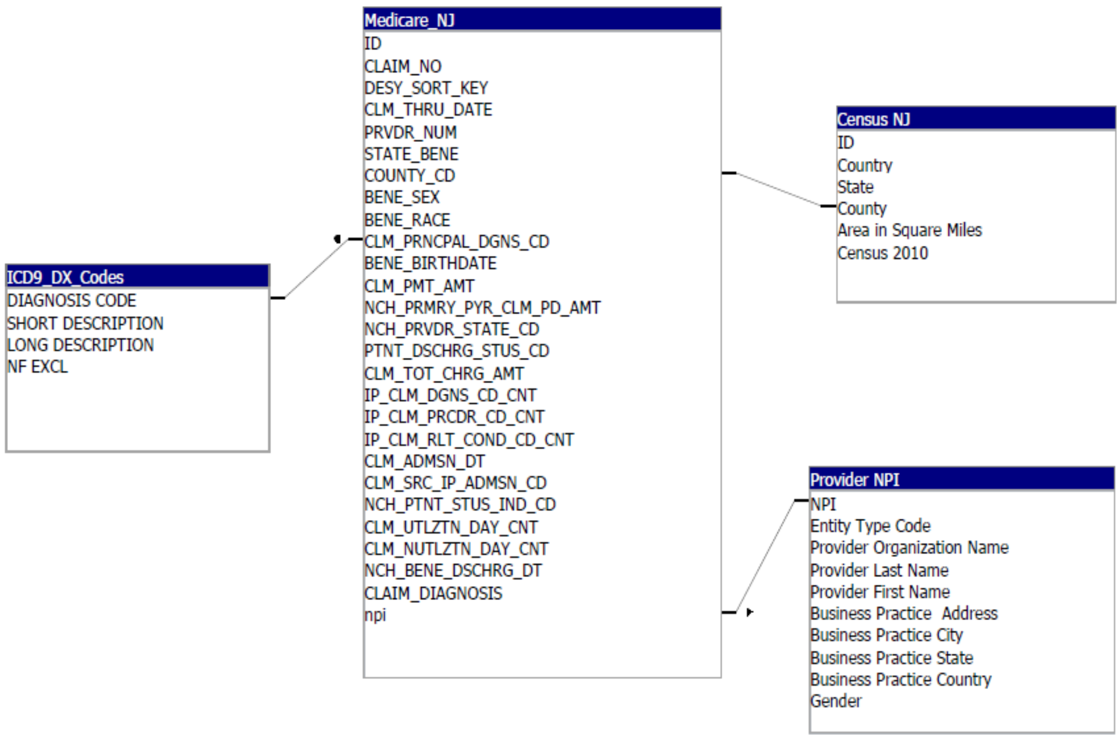


Figure 23: Join operations on the 4 different data sources

Several pre-processing tasks were conducted on the primary data set. The first relates to missing data fields. Any records that have missing information is discarded. Secondly, for NPI's that were not matched during the table join were also removed. Finally, geo-locational data, such as addresses and zip codes, were converted to longitudes and latitudes. In the end, the final primary data set consisted of 383,547 records and 36 attributes relative to the state of New Jersey.

3.4. ANALYSIS AND DISCUSSION

The initial analysis starts with basic tabular statistics and basic graphs followed by descriptive visual dashboards. Finally, an in-depth analysis providing more advanced visualizations and insights will be conducted.

3.4.1. Tabular Statistics and Basic Graphs

The analysis begins in the tabular form, with information on claim and dollar amounts. In order to obtain a per day representation, each pertinent value was further divided by the number of days. This per day rate was based on three separate scenarios that this paper proposes: liberal, base, and conservative. The rationale for having these separate scenarios is that providers generally differ in nature. They may range from individuals, to small clinics, and finally large institutions. Nevertheless, the liberal view assumes a four day work week and 45 work weeks per year, for a total of 180 working days. The base case is computed from a five day work week and 48 work weeks during the year, for a total of 240 working days. The conservative view is not adjusted, and, consequently, simply based upon 365 working days.

Table 2 below shows the provider descriptive statistics. Given that providers are of an institutional nature (e.g. hospitals, clinics), it may be reasonable to believe that many would operate 240 to 365 days per year, depending on whether they are small or large hospitals, respectively. Hence, the base and conservative scenarios might offer the best representation of daily provider activity in many cases. From this viewpoint, in 2010, the top five providers filed an average of 31 to 42 claims per day for the conservative view, and 48 to 63 claims per day for the base view. Furthermore, for that same period, these providers averaged between $1.7 and $4.2, and $2.6 and $6.3 million, respectively in claims per day for the conservative and base views. In examining payments actually issued to providers during 2010, they were found to be substantially lower, having mean values between $330,000 and $614,000 per day for the conservative view and between $383,000 and $932,000 for the base view.

Table 2: "Top 5 Providers" Descriptive Statistics

Provider. #	Total Claim Count	Claims Per day		
		Liberal (180 day year)	Base (240 day year)	Conservative (365 day year)
310041	15,221	84.6	63.4	41.7
310001	13,929	77.4	58.0	38.2
310012	12,699	70.6	52.9	34.8
310022	12,106	67.3	50.4	33.2
310015	11,467	63.7	47.8	31.4
Provider. #	**Total Claim Dollars**	**Dollars Per day**		
		Liberal (180 day year)	**Base (240 day year)**	**Conservative (365 day year)**
310001	1,521,053,291	8,450,296	6,337,722	4,167,269
310041	833,012,463	4,627,847	3,470,885	2,282,226
310022	805,830,992	4,476,839	3,357,629	2,207,756
310015	655,385,631	3,641,031	2,730,774	1,795,577
310012	616,884,547	3,427,136	2,570,352	1,690,095
Provider. #	**Total Payment Dollars**	**Dollars Per day**		
		Liberal (180 day year)	**Base (240 day year)**	**Conservative (365 day year)**
310001	223,831,957	1,243,511	932,633	613,238
310041	111,873,435	621,519	466,139	306,503
310022	92,122,802	511,793	383,845	252,391
310015	153,346,460	851,925	638,944	420,127
310012	121,205,061	673,361	505,021	332,069

While it is true that a timing difference would exist pertaining to a given claim and the associated payment, it can be assumed that claim activity does not fluctuate substantially from year to year. As such, claims and payments should not experience significant variations in a given period if all claim filings are complete and legitimate. Given this, the noted differences between claims and payments warrant further exploration.

This explanation is displayed in Figure 24 that depicts the most noteworthy differences by using a basic form of visualization, a histogram/bar chart. This histogram shows the highest 30 providers in terms of differences between claim amounts and payments. From this initial graph, it can be seen that provider 310001 stands out from the others by having a total difference of almost $1.3 billion. This could have several implications. For example, perhaps the provider experienced difficulty with respect to proper completion and submission of legitimate claims documentation. This would likely be indicative of a staff-level problem.

As another example, it may be that this provider generated claims that were inappropriate. This is a much more egregious issue, and would suggest that Medicare abuse and/or fraud is present. Whatever the case, this provider should be reviewed more closely in an effort to unearth the cause(s) of the excessive discrepancy between claims and payments during 2010. In addition, other outliers in terms of claim-payment differences should be similarly addressed. After the discussion of providers, attending physicians are examined.

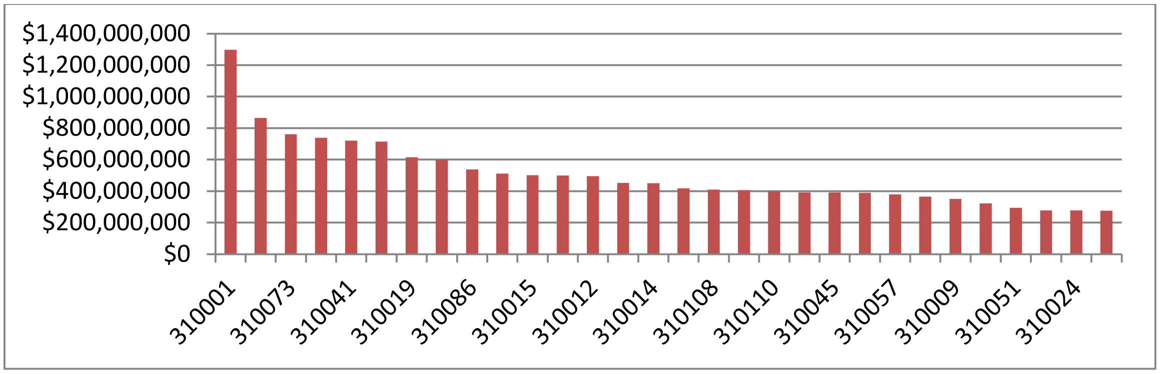


Figure 24: Providers - Claim Dollars in Excess of Payments

VISUAL THOUGHT (1)

Visualization helps explore the data, but to be able to visualize it in a certain way, users generally need to decide before hand on the metrics, attributes and parameters to choose from. Once the criteria is chosen users may have the option of presenting the data in either tabular or graphical form. It depends on what the users would want to look at, as well as how they want to look at it. In certain cases, it depends on how the user prefers to look at the data during their exploratory analysis. Some may prefer bubble charts or heat maps, while others may simply decide to use ranked tables. People might argue on whether which method produces more obvious results, but in reality, it depends. So for example, Figure 24 can also be represented in a table or even added to Table 2. Whether the information here is graphed or tabled, both produce similar outcomes.

As before, a table with descriptive statistics about claims, claim dollars, and payment dollars is presented first. Table 3 shows this information for the five most significant attending physicians. Attending physicians are generally physicians who work at an institutional provider. Therefore, in this analysis, it would be unreasonable to assume that attending physicians work 365 days a year. As such, the liberal and/or base scenario(s) would be most appropriate for physician-level evaluation. In examining these two views, it can be seen that total claims per day for attending physician number 4515525515 is 6.8

and 9.0 for the base and liberal cases, respectively. This substantial daily value may give rise to several important questions regarding the integrity of such claims.

Table 3: "Top 5 Attending Physicians" Descriptive Statistics

Attending Phy. #	Total Claim Count	Claims Per day		
		Liberal (180 day year)	Base (240 day year)	Conservative (365 day year)
4515525515	1,624	9.0	6.8	4.4
4230461102	1,363	7.6	5.7	3.7
4673852456	1,179	6.6	4.9	3.2
4693822477	931	5.2	3.9	2.6
4389568347	873	4.9	3.6	2.4
Attending Phy. #	**Total Claim Dollars**	**Dollars Per day**		
		Liberal (180 day year)	**Base (240 day year)**	**Conservative (365 day year)**
4515525515	93,611,441	520,064	390,048	256,470
4673852456	88,937,819	494,099	370,574	243,665
4230461102	86,044,769	478,026	358,520	235,739
4761662774	70,331,491	390,731	293,048	192,689
4848261167	69,416,981	385,650	289,237	190,184
Attending Phy. #	**Total Payment Dollars**	**Dollars Per day**		
		Liberal (180 day year)	**Base (240 day year)**	**Conservative (365 day year)**
4515525515	19,334,212	107,412	80,559	52,970
4791176537	10,341,818	57,455	43,091	28,334
4848261167	10,068,616	55,937	41,953	27,585
4673852456	10,043,765	55,799	41,849	27,517
4427701086	9,675,799	53,754	40,316	26,509

For example, one question can be asked on whether it is reasonable that a given attending physician would have seven to nine claims per business day during 2010? To connect the above noted claim counts with dollar values, the same attending physician filed claims amounting to daily averages of $390,048 and $520,064 for the base and liberal scenarios, respectively. To emphasize, these values are on a per day basis and pertain to a single physician. Once again, such high amounts may create genuine concerns about whether insurance abuse and/or other problems are occurring in this situation. Moving forward, payments received by attending physicians are examined. Once again, significant differences are observed between claims and payments relative to certain doctors. For example, in 2010 physician 4515525515 had $93.6 million in claims, but only received $19.3 million in payments.

To obtain an enhanced understanding of differences, the top 30 attending physicians regarding the claims-payment disparity are visualized in Figure 25 below. While many potential problems exist, it is instructive to note that six attending physicians had claims-payment differences exceeding $50 million in 2010. Furthermore, physician number 4673852456 had a claims-payment difference of nearly $80 million. Based upon this visualization, it would be useful to carefully review several of the attending physicians appearing as outliers in the data set.

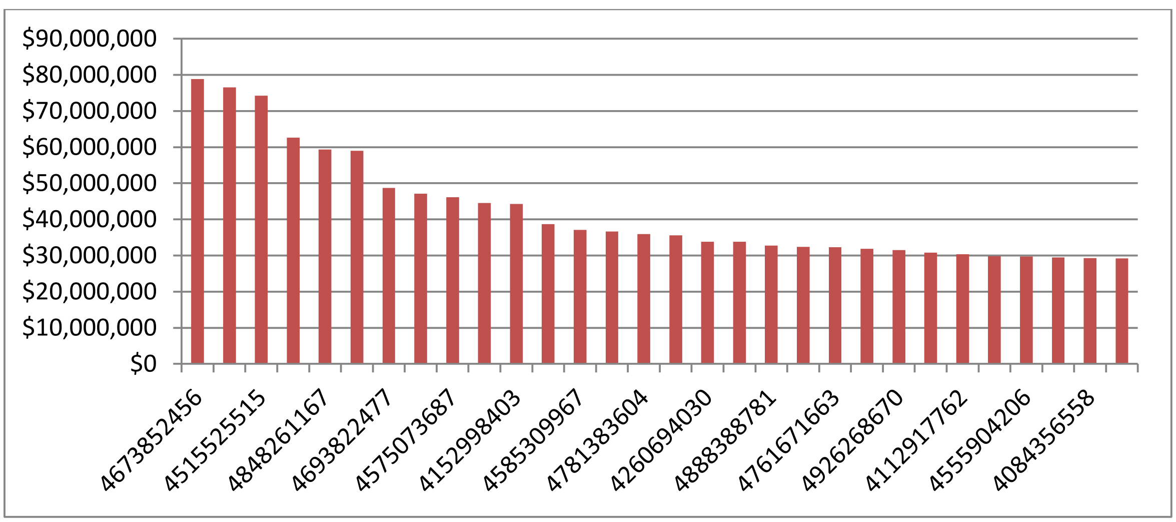


Figure 25: Attending Physicians - Claims in Excess of Payments

The third analysis relates to the top five operating physicians in terms of claims, claim dollars, and payment dollars. Similar to before, Table 4 below provides the descriptive statistics. To simplify the evaluation, it might be plausible to believe that operating physicians would generally work similar hours to attending physicians but have more hours per patient, as they deal with operating procedures. If implementing this viewpoint, then the liberal scenario might be most applicable to this physician type. Consequently, claim counts range from an average of 2.8 to 5 per day. Which can be considered acceptable as the average number of surgical and non-surgical cases for operating rooms per day in 2010 was 3.1 and 4.6 respectively (VMG Health, 2011). In terms of daily claim dollars, they fall between $254,027 and $536,083 on average. Interestingly, payment amounts are substantially lower than claim amounts. For example, whereas physician number 4750190404 has $96.5 million in claims during 2010, he/she only received $14.8 million in payments.

Table 4: "Top 5 Operating Physicians" Descriptive Statistics

Operating Phy. #	Total Claim Count	Claims Per day		
		Liberal (180 day year)	Base (240 day year)	Conservative (365 day year)
4868676523	869	4.8	3.6	2.4
4585505710	636	3.5	2.7	1.7
4838978677	588	3.3	2.5	1.6
4878133455	510	2.8	2.1	1.4
4230461102	493	2.7	2.1	1.4
Operating Phy. #	**Total Claim Dollars**	**Dollars Per day**		
		Liberal (180 day year)	**Base (240 day year)**	**Conservative (365 day year)**
4751090404	96,494,851	536,083	402,062	264,370
4868676523	65,934,941	366,305	274,729	180,644
4505710852	53,208,981	295,605	221,704	145,778
4240238180	51,848,667	288,048	216,036	142,051
4122701486	45,724,954	254,027	190,521	125,274
Operating Phy. #	**Total Payment Dollars**	**Dollars Per day**		
		Liberal (180 day year)	**Base (240 day year)**	**Conservative (365 day year)**
4751090404	14,758,625	81,992	61,494	40,435
4240238180	10,820,927	60,116	45,087	29,646
4791176537	9,999,546	55,553	41,665	27,396
4828732881	9,521,023	52,895	39,671	26,085
4505710852	9,178,432	50,991	38,244	25,146

Given the noteworthy differences between claims and payments, the top 30 operating physicians relative to this dimension were captured. Once again, the results are displayed in a histogram format in Figure 26. In representing the data in this manner, it clearly shows that one operating physician stands out from the others. Specifically, physician number 4751090404 has claims that are more than $80 million greater than payments. This individual clearly deserves closer attention and/or scrutiny. Also, in an overall sense, each of the entries in Figure 26 has claim-payment differences in excess of $20 million. At a minimum, it would seem prudent to conduct more thorough reviews of all these physicians to determine the legitimacy of 2010 claims activity.

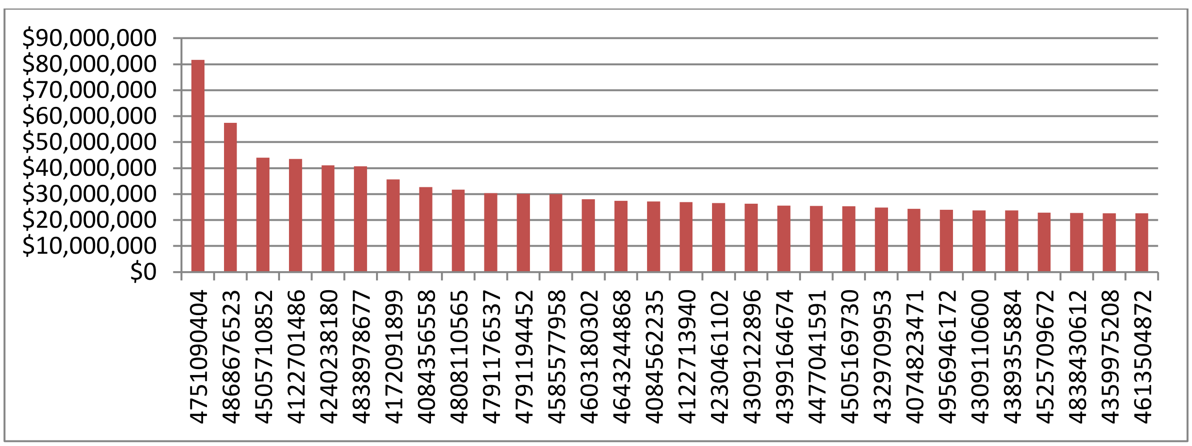


Figure 26: Operating Physicians – Claims in Excess of Payments

The final analysis in this section pertains to other physicians, and Table 5 summarizes the descriptive statistics for the five most noteworthy individuals of this type. Given that these doctors should more closely resemble attending physicians than operating physicians, it seems reasonable to perceive that the base and/or liberal scenarios would be most useful for this analysis. In examining the data in this manner, it can be seen that total claims per day for other physicians range from 1.9 to 3.0 for the base case, and 2.5 to 4.0 for the liberal view. In addition, the daily claim dollars range from an average of $122,718 to $212,483 for the base scenario, and $163,624 to $283,310 for the liberal perspective. Once again, because of the magnitude of daily claim dollars, it seems that additional investigations would be warranted. Finally, it can be noted that payment dollars are considerably less than claim amounts for the top 5 physicians in this category. Specifically, mean daily differences range from $31,901 to $43,091 in the base case, and $42,534 to $57,455 in the liberal scenario.

Table 5: "Top 5 Other Physicians" Descriptive Statistics

Other Phy. #	Total Claim Count	Claims Per day		
		Liberal (180 day year)	Base (240 day year)	Conservative (365 day year)
4575073687	715	4.0	3.0	2.0
4906160955	593	3.3	2.5	1.6
4389568347	535	3.0	2.2	1.5
4064660997	454	2.5	1.9	1.2
4034451005	453	2.5	1.9	1.2
Other Phy. #	**Total Claim Dollars**	**Dollars Per day**		
		Liberal (180 day year)	**Base (240 day year)**	**Conservative (365 day year)**
4575073687	50,995,847	283,310	212,483	139,715
4791176537	42,172,249	234,290	175,718	115,540
4084562235	30,833,574	171,298	128,473	84,476
4359964135	29,932,423	166,291	124,718	82,007
4653012895	29,452,409	163,624	122,718	80,692
Other Phy. #	**Total Payment Dollars**	**Dollars Per day**		
		Liberal (180 day year)	**Base (240 day year)**	**Conservative (365 day year)**
4791176537	10,341,818	57,455	43,091	28,334
4906160955	9,371,403	52,063	39,048	25,675
4064660997	8,242,325	45,791	34,343	22,582
4926728436	7,704,934	42,805	32,104	21,109
4034451005	7,656,197	42,534	31,901	20,976

As before, claim-payment differences were separately visualized. Figure 27 presents this information for the top 30 other physicians. Again, it is noticed that one individual stands out. In particular, physician number 4575073687 has a total difference of about $45 million, and this warrants additional review.

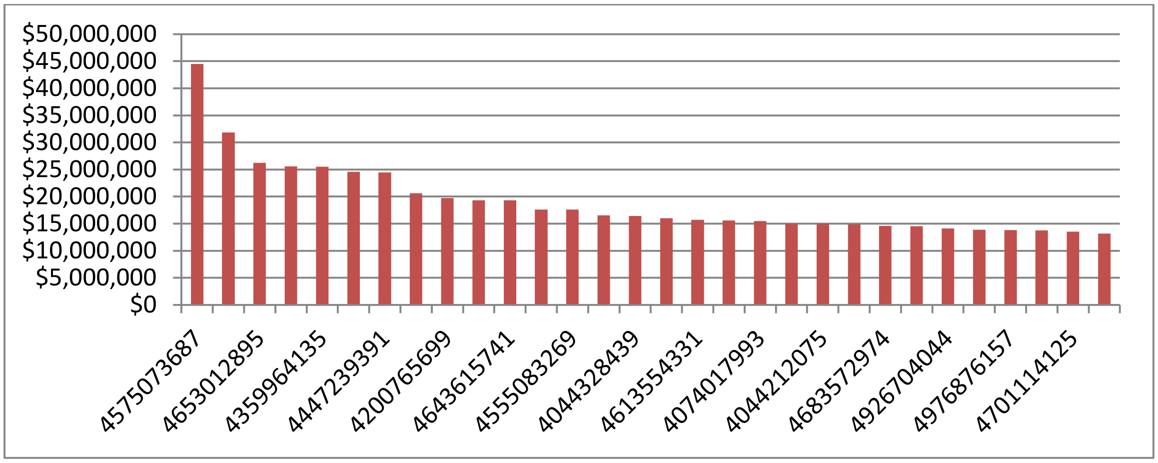


Figure 27: Other Physicians – Claims in Excess of Payments

Finally, Figure 28 presents a basic graph of age and gender distributions for Medicare claims. It can be seen by viewing the graph, that the percentage of female beneficiaries in the Medicare program is higher than male claimants for a majority of the age groups. This can be justified due to the fact that percentage of female residents in the state of New Jersey are 5% more than those of Male residents (U.S. Census, 2010). Furthermore, most of the claims cluster around the ">84 yrs" category, suggesting that beneficiaries above the age of 84 are more likely to use Medicare. In particular, more than 23% of all claims arose within the eldest age grouping. In addition, within this category nearly two-thirds of claimants were female.

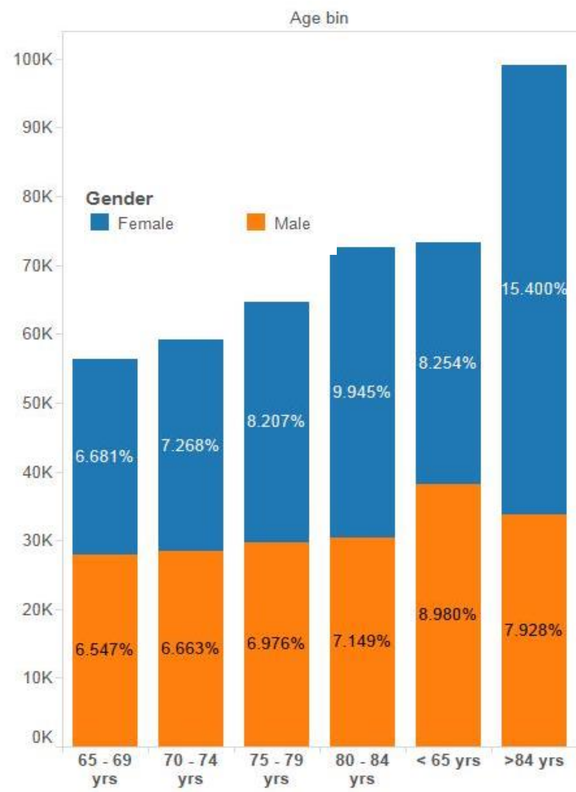


Figure 28: Age - Gender Population

VISUAL THOUGHT (2)

Figure 28 provides an example of how data visualization can be used to explain and present data in ways that would enlighten the decision maker. The same results can be presented in a table as seen below. Probably, with more time to analyze the table, the same insights can be gained. So whether or not tables or graphs are used depends, not on what the user wants to portray, but how and to whom they are communicating it to.

Gender	Age bin	% Count	Number of Records
Female	65 - 69 yrs	6.68%	28,428
	70 - 74 yrs	7.27%	30,888
	75 - 79 yrs	8.21%	34,903
	80 - 84 yrs	9.95%	42,249
	< 65 yrs	8.25%	35,079
	>84 yrs	15.40%	65,391
Male	65 - 69 yrs	6.55%	27,855
	70 - 74 yrs	6.66%	28,365
	75 - 79 yrs	6.98%	29,687
	80 - 84 yrs	7.15%	30,429
	< 65 yrs	8.98%	38,171
	>84 yrs	7.93%	33,695

3.4.2. Descriptive Visual Dashboards

The next analysis is based upon a series of descriptive dashboards, each of which is comprised of four different visual representations, including relevant histograms, heat maps, and packed bubble charts. A separate dashboard is constructed for providers, attending physicians, operating physicians, and other physicians. To conclude, a final dashboard is created that depicts claim information whereby a given physician performed two or more roles.

In Figure 29, pertinent provider information is presented, and the most noteworthy elements are labeled. In examining the dashboard beginning with the histogram visualization in the upper left (view 1), it can readily be seen that provider number 310041 has the largest number of Medicare claims at just over 14,000. In the remaining three views, shape size is positively related to relative volume of activity. Given this, it is clear that provider number 310001 has the largest claims (top right, view 2) and payment activities (bottom left, view 3) in terms of dollars, and experienced the greatest differential between claims made and payments received (bottom right, view 4). In returning to the histogram, it is noted that this provider is the second largest regarding claim count, having about 14,000 of them during 2010.

By glancing at the visualization dashboard, questions immediately surface. For example, is it reasonable for a given provider to have over 14,000 legitimate Medicare claims during a fiscal year? Is it reasonable for the top provider in terms of claim dollars to submit nearly twice the quantity of the second leading provider? Why does significant disparity exist between claim dollars and payments for certain providers? Whatever the case, by observing the data in this dashboard format, the outliers become glaringly apparent such that investigatory resources might be better allocated for resolution purposes.

<u>VISUAL THOUGHT (3)</u>

The descriptive visual dashboards in this section are considered explanatory data visualizations. Metrics and attributes have already been decided on, and the analysis has already been conducted. These dashboards are only a mean to present the data in ways that would summarize multiple information in one simple view. Since in these examples, simple ranking is used, then one can argue again that tables or graphs will do the job similarly.

The table below shows proof of how such a dashboard can be presented in table form while still maintaining its integrity. However, one benefit of using visualization is it makes it easier and faster to pin-point any patterns that may arise by means of graphs and colors. Though a benefit of tables remains in how they show the actual data as numbers and/or text.

PRVDR NUM	Clms Count	CLM PMT AMT	Clm Pmt Diff	CLM CHRG AMT
310001	14,792	223,831,957.36	1,297,221,333.99	1,521,053,291.35
310012	12,711	103,373,614.75	495,679,485.78	616,884,547.05
310015	11,483	153,346,460.31	502,039,170.94	655,385,631.25
310019	10,061	124,966,590.72	615,140,874.39	740,107,465.11
310022	12,106	92,122,802.36	713,708,189.67	805,830,992.03
310038	10,318	159,198,547.08	864,690,029.18	1,023,888,576.26
310041	15,221	111,873,435.23	721,139,028.21	833,012,463.44
310045	6,793	76,186,210.35	392,232,855.93	468,419,066.28
310048	6,861	55,914,071.68	510,874,122.97	566,788,194.65
310051	8,397	85,496,249.20	294,628,353.91	380,124,603.11

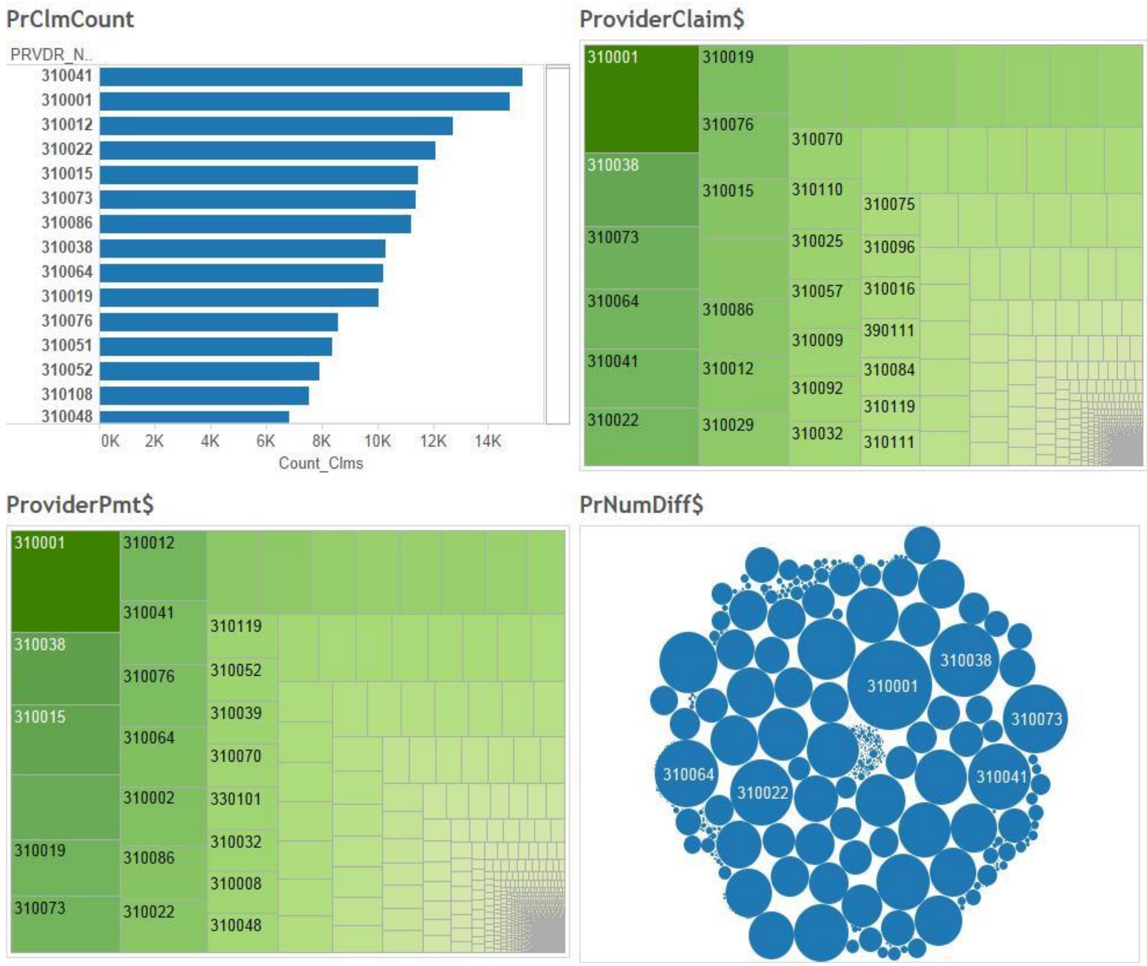


Figure 29: Provider Visual Dashboard

In Figure 30, it is once again the case that a single item appears as an exception. Specifically, attending physician number 4515525515 is the largest in terms of claim count, claim dollars, and payment dollars. Furthermore, he/she is third highest relative to excess of claims over payments. It is also notable to mention that this physician has about 300 more claims than the second largest competitor in this physician grouping. Interestingly, while physician numbers 4673852456 and 4230461102 have the second and third highest claim dollars, they are the first and second highest respectively in terms of excess of claims over payments. Once again, a substantial amount of potentially actionable information is depicted in this dashboard, and it should facilitate efficiency and cost effectiveness in coordinating appropriate investigation efforts.

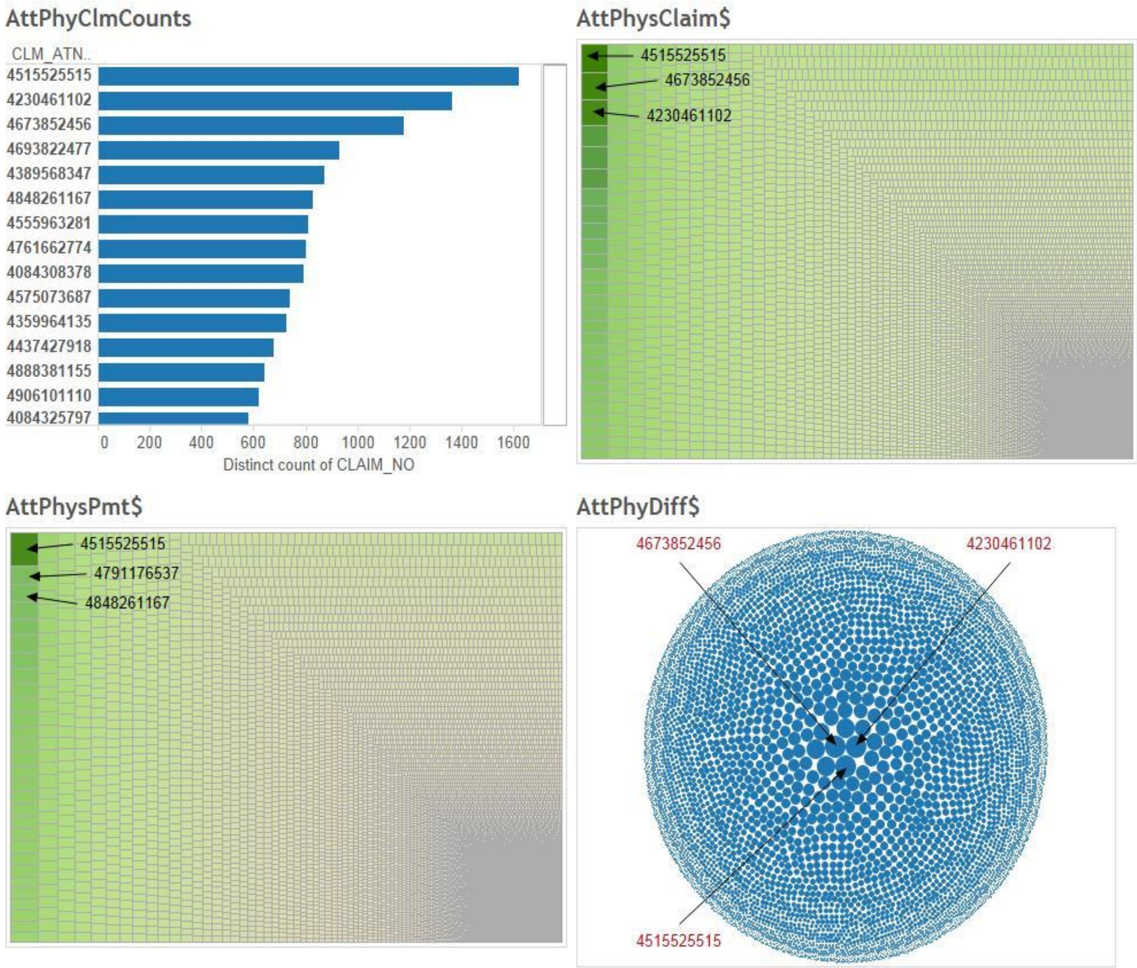


Figure 30: Attending Physician Visual Dashboard

It is likely the case in Figure 31 that at least two physicians may be considered as outliers of interest. First, operating physician number 4868676523 has the largest claim count during 2010, and was affiliated with nearly 50 percent more claims than the second highest individual. Additionally, this doctor is second largest in terms of both dollar value of claims as well as excess of claims over payments. Next, while operating physician number 4751090404 did not appear among the top 15 physicians in terms of claim count, he/she had the highest claim dollars, payments received, and excess of claims over payments. This might suggest that this physician is associated with high-cost procedures and claims, and has a relatively substantial amount of unpaid claim activity.

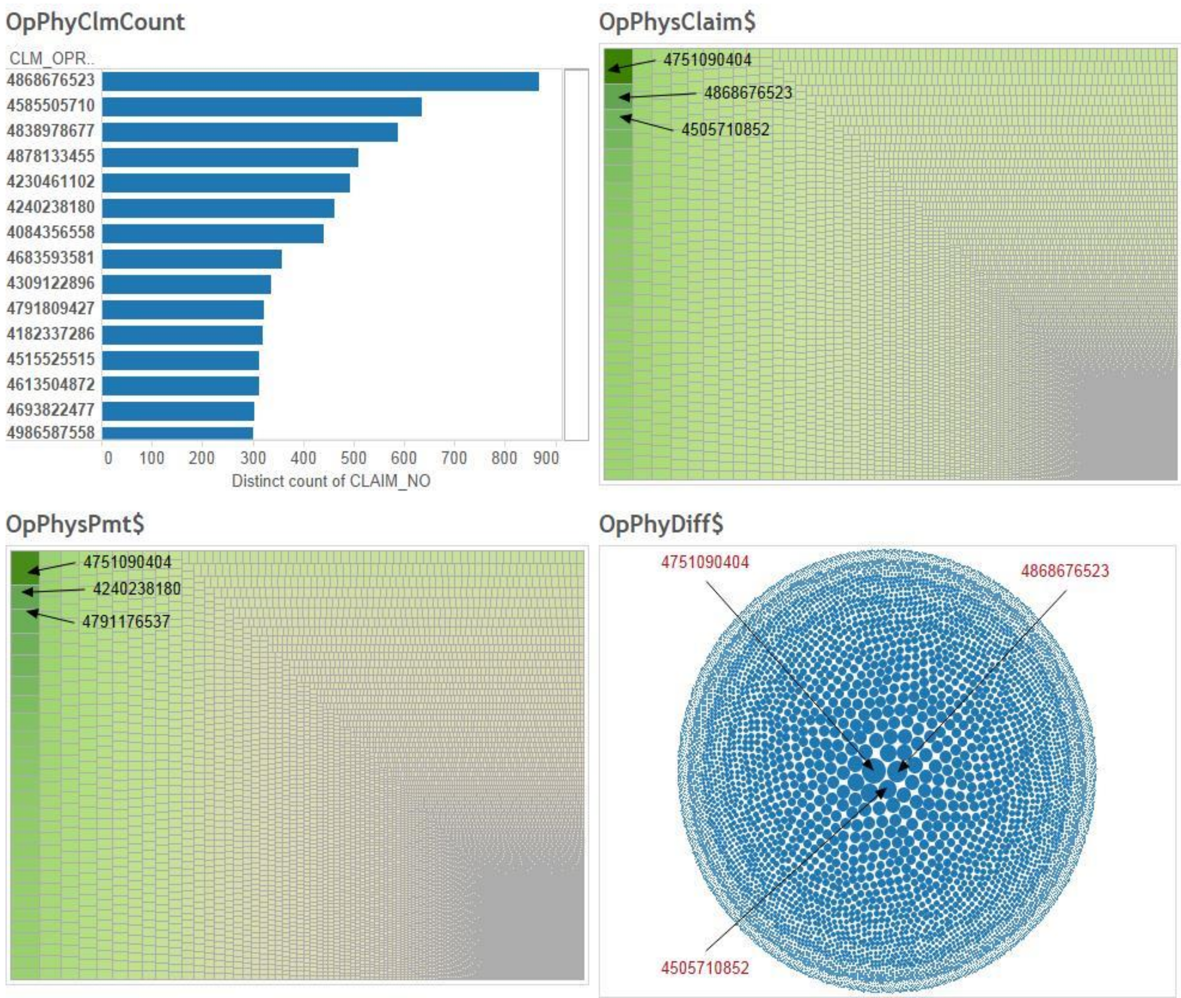


Figure 31: Operating Physician Visual Dashboard

In addition to the above physicians, one might also tend to question operating physician number 4505710852. In particular, this individual has the third highest ranking relative to excess of claims over payments, and this might generate some skepticism concerning claims activity. At any rate, based upon the visualization, several questions are unearthed that deserve to be answered. For example, is it normal for an operating physician to be involved with nearly 900 claims in a given year, particularly when the next highest claimant has only about 600? Is it reasonable that the operating physician with the most activity in terms of total claim amount submits about 50 percent more claim dollars than that of the second largest competitor? Through knowledge discovery via visualization these and other relevant questions are able to be identified.

In Figure 32, other physician number 4575073687 appears as the most potentially problematic entity. In particular, this individual has the greatest number of claims, dollar amount of claims, and excess of claims over payment dollars. As can be seen, while this doctor far surpasses the second largest claimant in terms of both count and dollar amount, he/she is not represented in the top three relative to payments received. Consequently, there is substantial disparity between claims and payment history, and this might warrant closer investigation to determine the legitimacy of Medicare claims affiliated with this physician. On the other hand, while physician number 4906164955 has the second largest claim count and payments received amount, he/she is not represented in the top three relative to claim dollars. As such, the claim-payment difference for this individual does not appear as an extreme concern. This might suggest that this doctor does not represent a particular problem, although the sheer volume of claim activity might substantiate further review procedures.

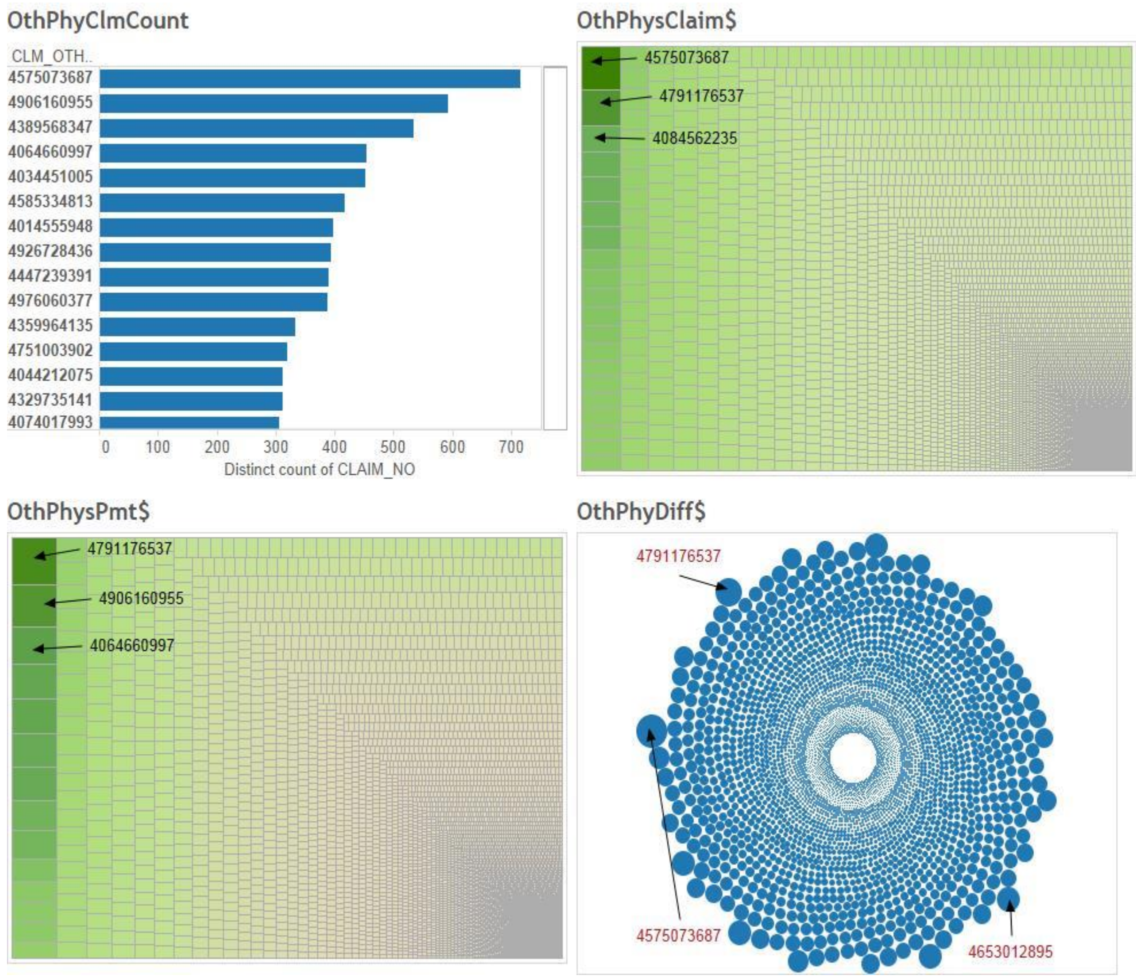


Figure 32: Other Physician Visual Dashboard

The visual dashboard in Figure 33 is comprised of all claims whereby a given physician performed two or more roles (i.e. attending, operating, and/or other) in the same claim. At this point, an immediate question concerns whether or not such circumstances are considered normal. If not, then substantial review would be warranted based upon the accumulated information.

In the upper left heat map, claim counts of physicians who acted as both the operating and attending physician in each event are depicted. Based upon the size of the rectangles, physician number 4230461102 has substantially more claim activity in which he/she performed multiple duties. In the upper right heat map, claim counts for doctors who were engaged as both operating and other physician in each claim are shown. While the differences in this view do not appear as dramatic as the previous image, one physician (i.e. 4791176537) nevertheless stands out in comparison with the others. In the lower left visualization, claim counts of physicians who acted as both the attending and other physician in each claim are presented.

Similar to the first view in this dashboard, the dual-role physician associated with the largest number of claims (i.e. 4575073687) appears to far surpass all of the competition. Interestingly, there are also cases where certain doctors apparently perform the attending, operating, and other physician duties. The bubble chart in the lower right of Figure 28 depicts this information. On an individual basis, physician number 4791176537 has the greatest volume of claims whereby he/she served all three physician roles. In addition, there are other doctors who, in a relative sense, have frequently performed in this capacity.

Taken collectively, it may be that several physicians should be considered for additional review based upon their existence as multiple-role medical providers. However, this would obviously be contingent upon whether it is reasonable for a physician to perform two or more roles in a particular Medicare claim. Whatever the case, a primary candidate for investigation would be physician number 4791176537. Moving forward, other outliers from the lower right bubble chart should be considered for closer review.

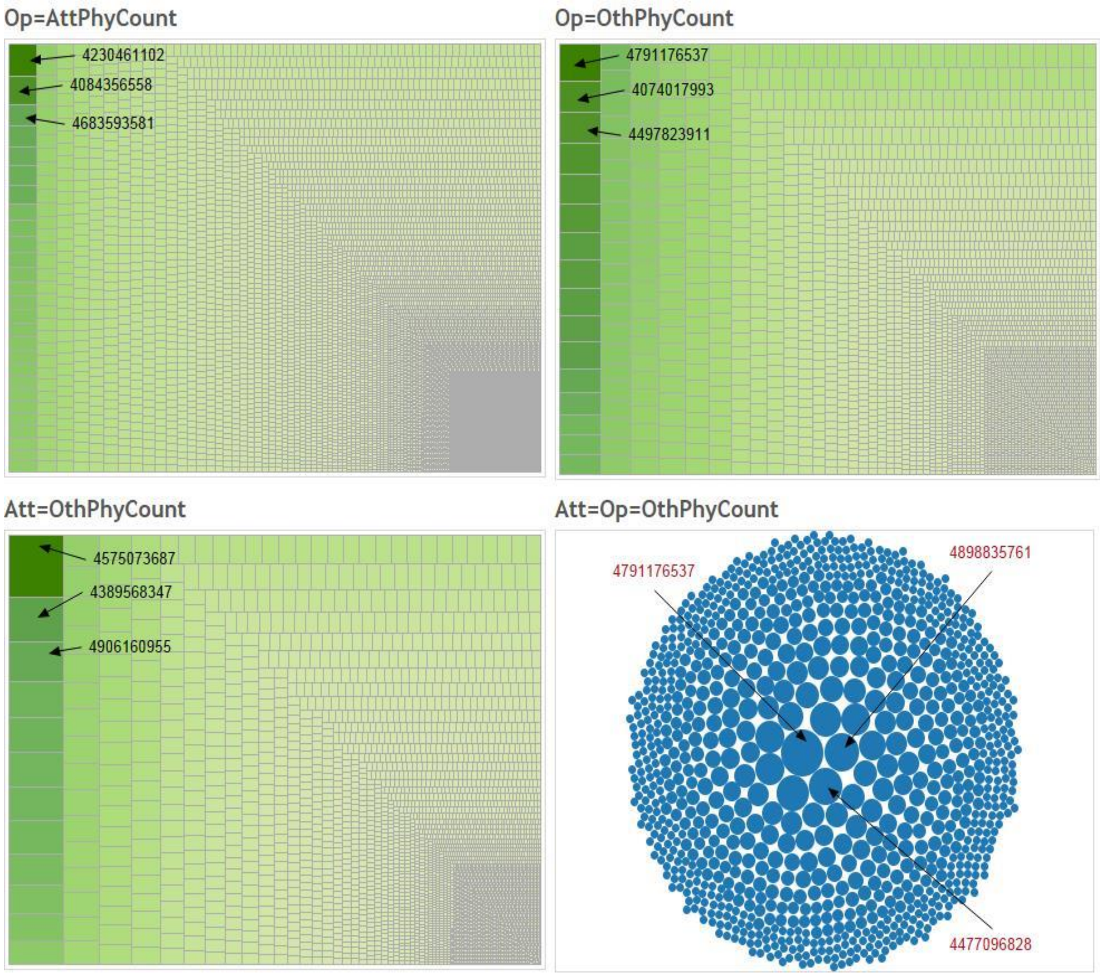


Figure 33: Visual Dashboard of Claims with Same Physician Performing Multiple Roles

3.4.3. In-Depth Analysis and Advanced Visuals

The following sections will provide in depth analysis focusing primarily on providers and beneficiaries of Medicare by utilizing advanced visual techniques. The analysis starts with a geographic visualization of claim percentages and New Jersey population percentages. Figure 34 presents two visuals that depicts this information. From the graph on the right (Green), it can be seen that two counties, Ocean and Bergen, generate the greatest number of claims within the state of New Jersey based on the color intensity of the heat map. However, according to U.S. Census for the state of New Jersey in 2010 as shown in left graph (Brown), the top two counties in terms of population are Bergen and Middlesex, with 10.29% and 9.2% respectively. Also, despite Bergen County being smaller than Ocean and several others in terms of square miles, it generates the largest percentage of total Medicare claim dollars. However, reviewing the population graph provides at least a partial rationale for this situation. Specifically, even though Bergen occupies a relatively small area, its population is over 50 percent larger than Ocean County. In fact, of the nine counties evaluated, Bergen is the largest in terms of residency level.

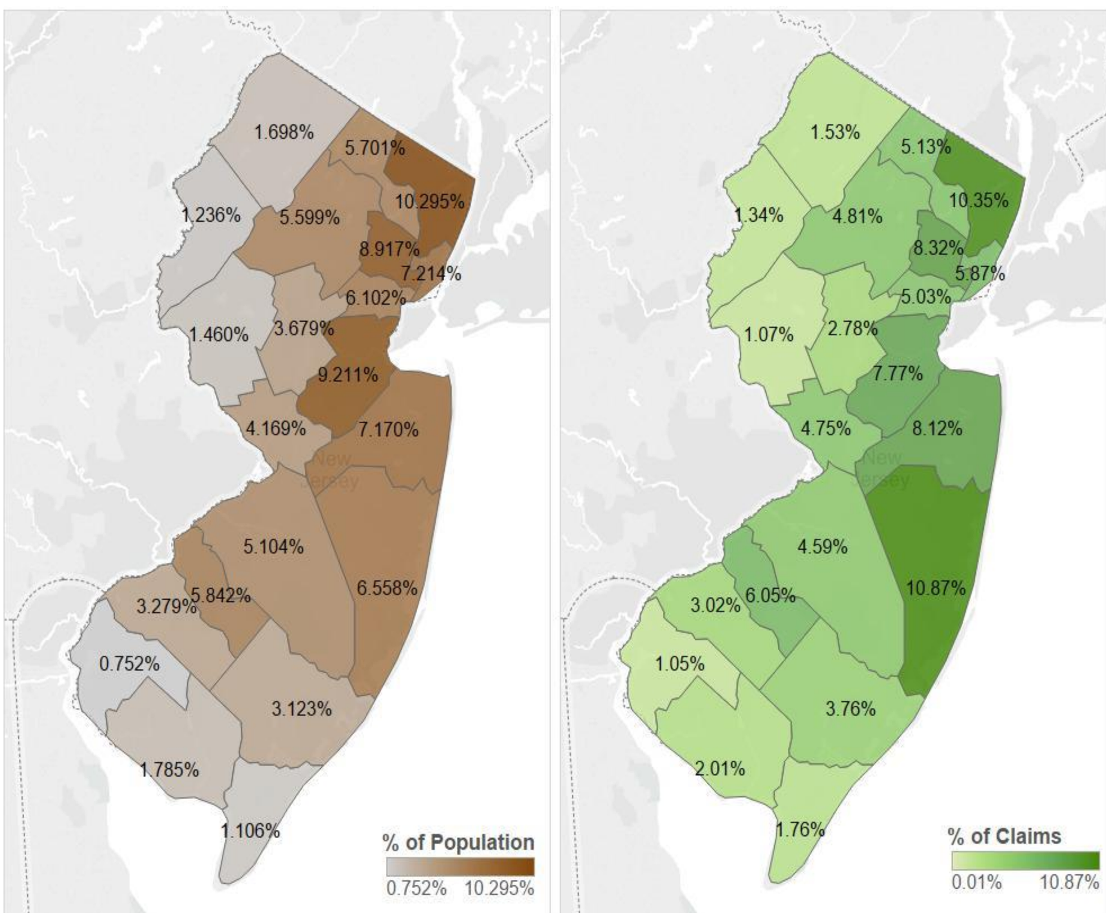


Figure 34: Geographic Visualization of Claims and County Population

VISUAL THOUGHT (4)

Geographical visualizations may be considered both exploratory and explanatory data visualization. Locational data is generally best explored using geographical visualizations, they help picture the data points across the globe and discover patterns and outliers within the data. They are also useful in explanatory data visualization when they are used to present and communicate results and findings to decision makes. Even though Figure 34 can be presented in a table, as shown below, one can argue that when it comes to visualizing a large list of locations, viewing and comparing multiple lines of data in a table is cumbersome, and hence visualizing them would be more pleasing and easier on the eyes. Nevertheless, as the graphs become more complex, as it will be seen in the next couple of pages, it takes more and more tables to present data non-visually.

County	% of Total Count of Claim No
Ocean	10.87%
Bergen	10.35%
Essex	8.32%
Monmouth	8.12%
Middlesex	7.77%
Camden	6.05%
Hudson	5.87%
Passaic	5.13%
Union	5.03%
Morris	4.81%
Mercer	4.75%
Burlington	4.59%
Atlantic	3.76%
Gloucester	3.02%
Somerset	2.78%
Cumberland	2.01%
Cape May	1.76%
Sussex	1.53%
Warren	1.34%
Hunterdon	1.07%
Salem	1.05%

County	% of Total Census 2010
Bergen	10.29%
Middlesex	9.21%
Essex	8.92%
Hudson	7.21%
Monmouth	7.17%
Ocean	6.56%
Union	6.10%
Camden	5.84%
Passaic	5.70%
Morris	5.60%
Burlington	5.10%
Mercer	4.17%
Somerset	3.68%
Gloucester	3.28%
Atlantic	3.12%
Cumberland	1.78%
Sussex	1.70%
Hunterdon	1.46%
Warren	1.24%
Cape May	1.11%
Salem	0.75%

In looking for potential issues within certain regions, it is instructive to compare the claim percentages with the associated population percentages in the geographic heat map of Figure 34. Other things equal, one would generally expect the two rates to be comparable for a given county, and any significant differences could be indicative of problems. By approaching the analysis in this manner, it is in fact noted that most of the counties exhibit comparability between claim and population percentages. However, six negative outliers exist. More particularly, Atlantic, Cape May, Hunterdon, Mercer, Morris, and Ocean Counties have claim dollar rates that are in the neighborhood of 30 to 45 percent higher than the associated population rates. For example, while Ocean County accounts for 6.56 percent of the New Jersey population, the region issued 10.87 percent of all Medicare claim dollars during the same year.

In examining counties using the techniques outlined above, it is also advisable to take into consideration whether notable retirement communities exist in any of the potentially problematic areas. If so, then this might provide at least a partial justification for the observed differences between population and claim percentages, and perhaps even claim count and claim dollar percentages. Whatever the case, in optimizing the use of investigatory resources, it could be quite advantageous to disaggregate to the county level when deciding how to conduct Medicare claims reviews.

Additional exploration can be done relative to the number of hospitals per county and the number of staffed beds per hospital. Figure 35 presents a visualization that explores this. The graph on the right (orange) provides information on the number of major hospitals per county, while the graph on the left provides information on number of hospital beds per city. Investigating the graph on the right, Bergan and Ocean counties have 7.4% (5) and 5.8% (4) of total number of NJ major hospitals respectively. Despite these two counties having the largest number of Medicare claims, they are behind Essex County which has

10.3% (7) of total number of NJ major hospitals. Nevertheless, Essex County is still third in terms of Medicare claims having 8.3% of the total count.

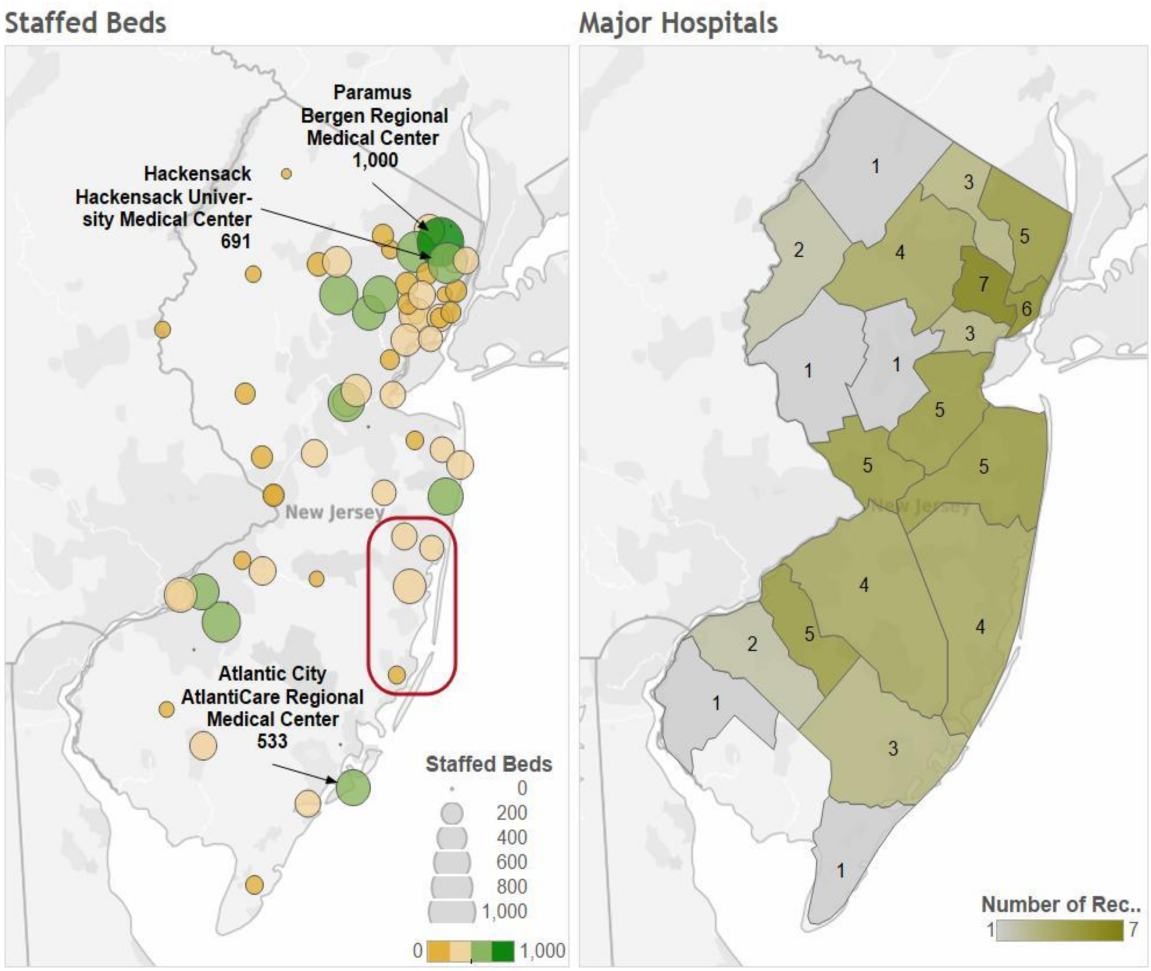


Figure 35: Geographic Visualization of Number of Hospitals and Hospital beds

Moving to the graph on the left, the number of total beds per hospital is presented. This graph provides a good rational for Bergen County observing the highest claim count. Two major hospitals in Bergen County, Bergen Regional Medical Center and Hackensack University Medical Center, have hospital beds in excess of 500, with the former observing the highest number of hospital beds in all of NJ, at a total of 1000 beds. As for Essex County, despite having the largest number of major hospitals, it only has hospitals providing beds of less than 500 each. However, due to this condensation of small hospital beds, Essex County produced the third largest count of Medicare claims as seen in Figure 34.

Two potential investigatory points here. The first relates to AtlantiCare Regional Medical center. Despite this hospital having beds in excess of 500, Atlantic County only has 3.8% of all Medicare claims. The other point relates to Ocean County. As seen in Figure 34, Ocean County produces the largest number of Medicare claims, however, in observing its hospital's bed count, none have beds in excess of 500, as highlighted in red. Auditors may potentially want to investigate these hospitals further to understand the discrepancy between the small bed count and high claims produced.

It can be assumed, that for any provider under the Medicare system, any amounts charged for the provided services should be comparable to the amounts paid back by Medicare. Figure 36 provides a visualization that potentially investigates this matter. The graph shows a scatter plot of providers having the X-axis represent the total payments made from Medicare, and the Y-axis represent the total amounts charge by providers for services rendered. The scatter plot is also color coded based on the type of diagnosis. From the graph, several outliers exist. From the upper left quadrant, "Shore Memorial Hospital", Acutecare Health Systems LLC", and "The Cooper Health System". From the lower right quadrant is "AHS Hospital Corp." and "Centerstate Medical Center Inc.". Such situations may be of high risk and auditors may consider further review of these providers.

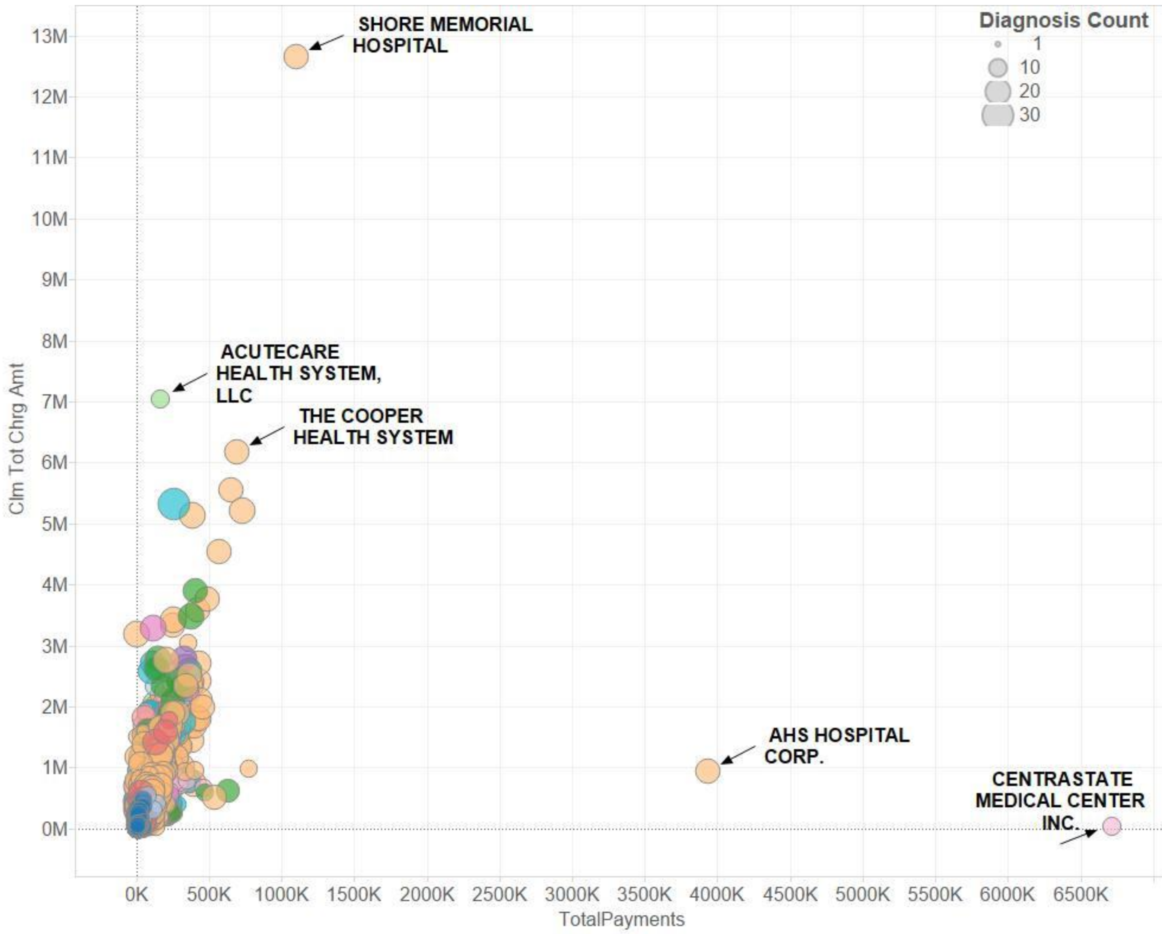


Figure 36: Amount charged for services vs amounts paid back by Medicare[3]

Figure 37 presents a visualization of the payments made by Medicare to providers per diagnosis made. Box plots are used to visualize the variation in each type of diagnosis (color coded separately) performed by providers. Spacing between the different parts of the box indicate the degree of dispersion, skewness, and outliers for each. It would be assumed here that diagnoses for certain illnesses or diseases, despite the size and/or popularity of the provider, should be comparable. However, several outliers are present

[3] Color coding for Figure 36 and 37 found in Appendix C

and a few have been marked on the graph. Again, whether such circumstance are ordinary or not, warrant further investigation.

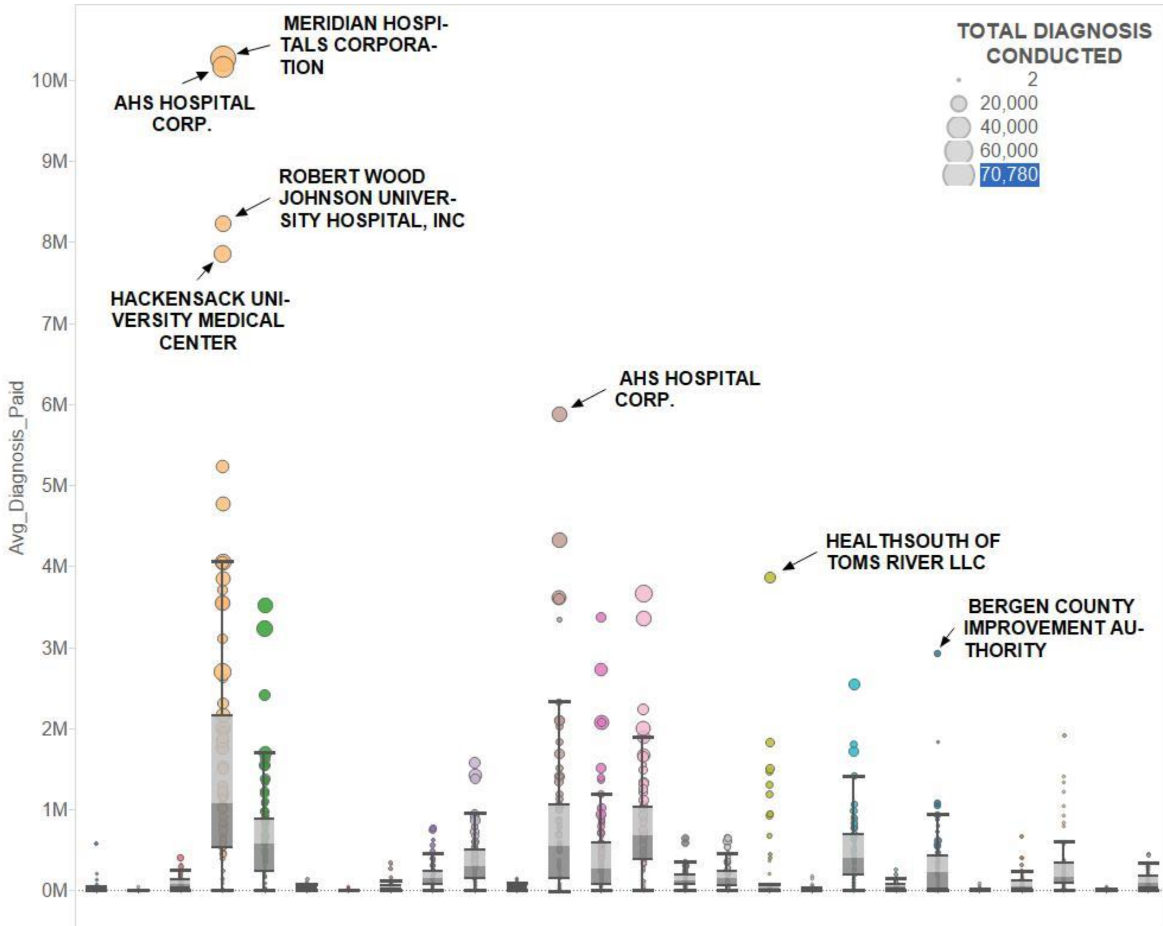


Figure 37: Payments made by Medicare to providers per diagnosis made

Further analysis was conducted specifically to see whether providers were still getting paid for claim related to dead patients. The visualization in Figure 38 shows a monthly timeline of payments made by Medicare for deceased patients. Two questionable items appear: Patient IDs 496572183 and 499832903, relating to providers, "Trinitas Regional Medical center" and "East Orange General Hospital" respectively. These patients appear to have been billed twice, despite having died the first time. For example, patient 496572183 died on April of 2010 while Trinitas Regional Medical center received their payment from Medicare for the services provided amounting to $38,679. However, on May 2010, Trinitas Regional Medical center received additional payments amounting to $14,705 from Medicare for the same deceased patient. Similar instances occurred for patient 499832903. This patient was registered as being dead on March of 2010 by the hospital after receiving $9,336, however the same patient was used on a claim for Medicare payments again on October of 2010 amounting to $4,230. Such examples are of high risk, and auditors would want to investigate the reasons behind such behaviors.

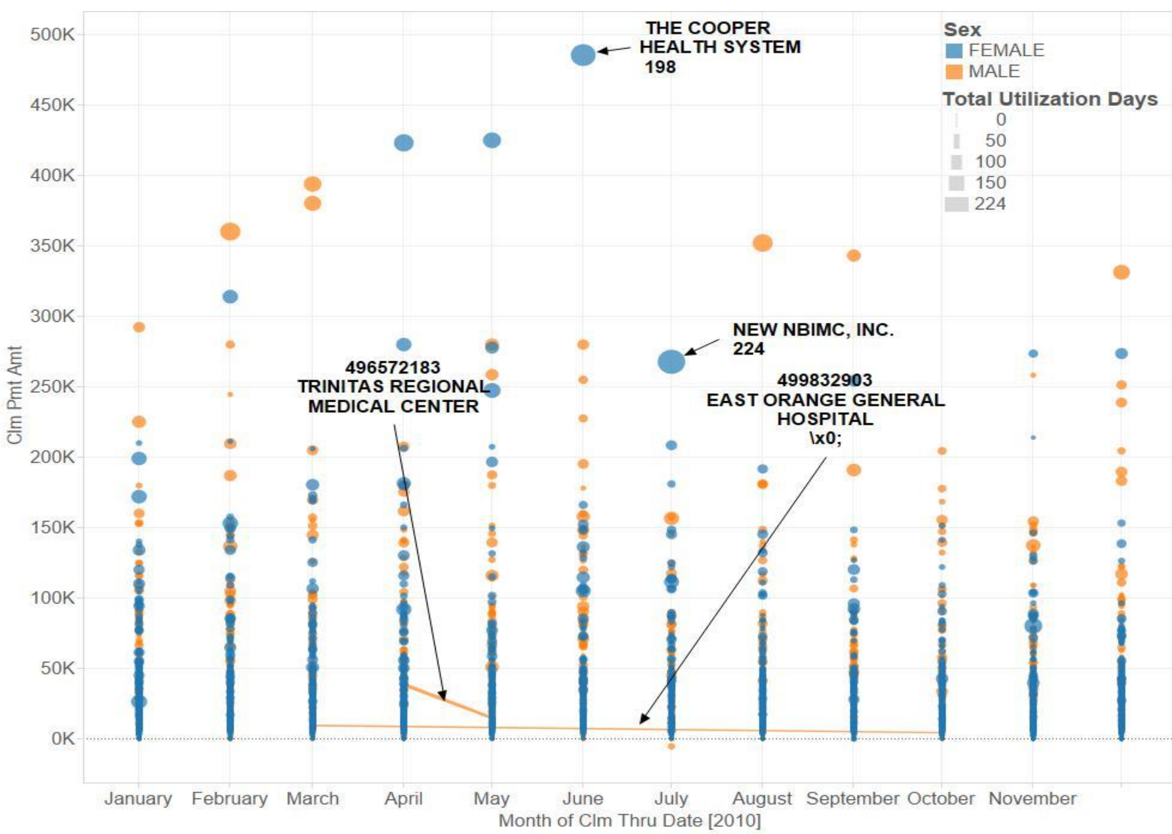


Figure 38: Monthly timeline of payments made by Medicare for deceased patients

VISUAL THOUGHT (5)

Figure 38, presents an example of where data was explored first using exploratory data visualizations, and then presented in a way to communicate certain insights using explanatory data visualization. Nevertheless, graphs usually works best in these circumstances, especially when dealing with multi-dimensional data, and incorporating multiple sources of Big Data. Visuals here outperform tables as the patterns and insights become more easily obtained from the graphics. Depending on the number of dimensions, for example when using ranking as in the highest and lowest values, then tables or graphs will do the job similarly as seen before. However, the more dimensions added (highest, lowest, mean, category, percent of...etc.), the more useful visualization becomes. Below is a table that represents a sample of the information visualized in Figure 38. It can be observed from this comparative example, that in these cases, it is hard to read and analyze the data, and this visuals outperform tables.

Bene Sex	Desy Sort Key	Measure Names	Month of Clm Thru Date	Provider Organization Name	Measure Values
FEMALE	100034390	Clm Pmt Amt	Jan-10	ROBERT WOOD JOHNSON UNIVERSITY HOSPITAL, INC	6,850.54
FEMALE	100034390	Clm Utlztn Day Cnt	Jan-10	ROBERT WOOD JOHNSON UNIVERSITY HOSPITAL, INC	18
FEMALE	100035113	Clm Pmt Amt	Feb-10	SAINT PETER'S UNIVERSITY HOSPITAL	14,141.63
FEMALE	100035113	Clm Utlztn Day Cnt	Feb-10	SAINT PETER'S UNIVERSITY HOSPITAL	2
FEMALE	100051101	Clm Pmt Amt	Jul-10	SOUTHERN OCEAN COUNTY HOSPITAL	9,221.44
FEMALE	100051101	Clm Utlztn Day Cnt	Jul-10	SOUTHERN OCEAN COUNTY HOSPITAL	2

Another insight that can be made from the same graph relates to the total payments made from Medicare to the providers for these departed patients. One feature of this graph is that it visualizes the amount of utilization days by circle size. So in utilizing this feature, two providers can be compared, namely “The Cooper Health System” and “New NBIMC Inc.”. The first provider received payments from Medicare exceeding $480,000 for a patient who died in June for a total of 198 utilization days (Days spent as in-patient). As for New NBIMC Inc., in July, they received an amount of around $260,000 for a departed patient who spent a total of 224 utilization days. Why would more days spent in a hospital be charged less than spending fewer days? Are these differences normal between different hospitals? Are the medications and medical equipment used the same? Nevertheless, despite this being a high level view, such discrepancies are worth investigating.

The next visual focuses on the number of diagnoses made and dollar amounts received for discharged patients. Figure 39 provides a graph that captures such information. Based on this visual, it can be seen that most providers are below the $500,000 mark despite the number of diagnoses performed. Generally, it can be assumed that as the number of diagnoses is increased, the payments will increase, which is evidently displayed by the trend line.

One could also get an insight from the color coding (male vs female), in that males were generally charged higher compared to females. However, one significant item here is the outliers as highlighted in the graph. Both “AHS Hospital Corp” and “Centerstate Medical Center Inc.” have received payments of around $3.9 and $6.7 million respectively. These providers appear to be significantly over the current trend compared to the others, and thus require further investigation on the cause of this issue.

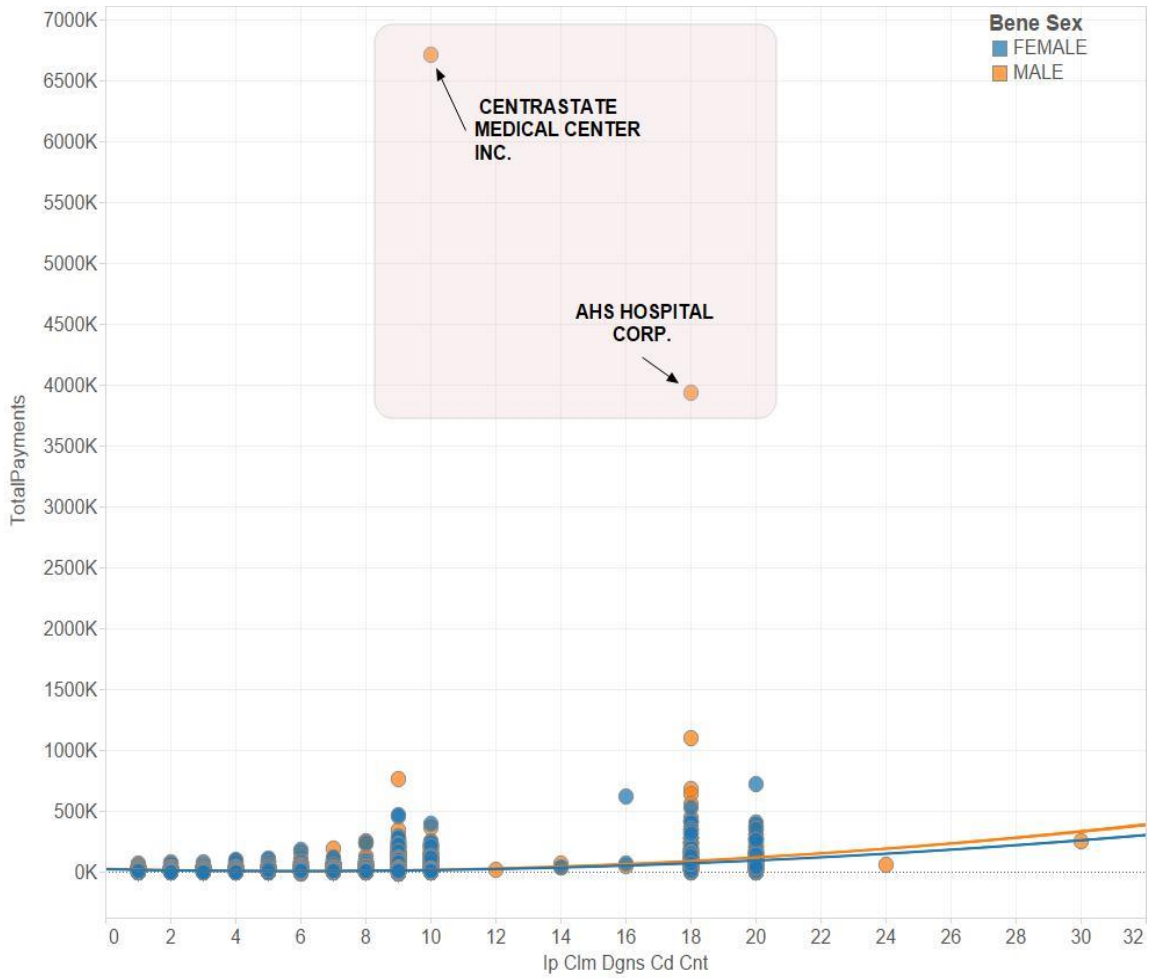


Figure 39: The number of diagnoses and dollar amount of discharged patients

Moving to Figure 40, this visual presents payment trends for each provider throughout the year. There are several ups and downs from month to month for different providers, which can be assumed as normal behavior, however, one significant outlier exists, as highlighted in red. This provider had a steady trend of Medicare payments from February till December averaging around $3.9 million, however in January at the beginning of the year, Centrastate Medical Center Inc. received in excess of $10 million. Whether or not this is normal, when compared to others this provider had the only significant change and thus is worth examining further.

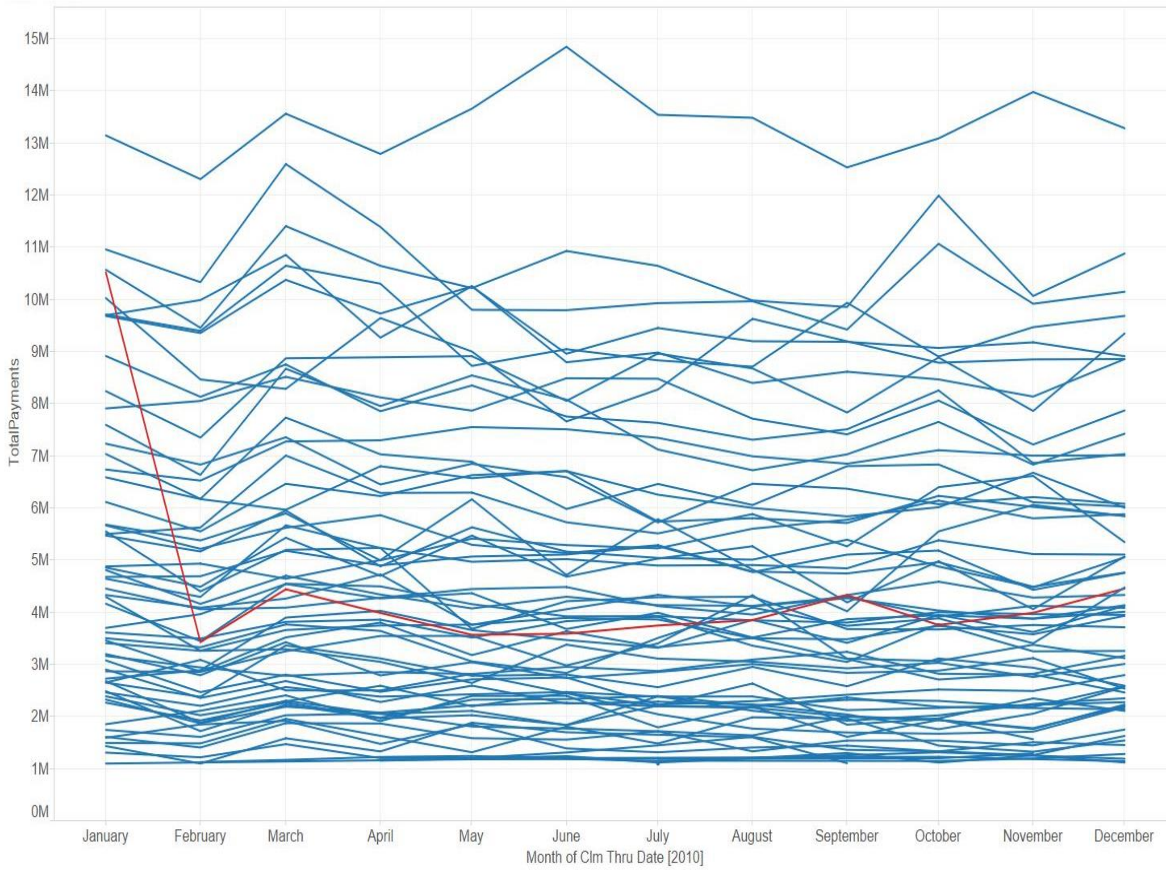


Figure 40: Annual payment trends for each provider

Network analysis of Medicare providers and their beneficiaries is performed next. ForceAtlas2 algorithm is used to visualize the network. It integrates different techniques to provide a continuous graph layout network visualization. Its very essence is to turn structural proximities into visual proximities, facilitating the analysis. It applies a linear attraction linear repulsion model with few approximations. Nodes repulse each other (like magnets) while edges attract the nodes they connect (like springs). These forces create a movement that converges to a balanced state. This final configuration is expected to help the interpretation of the data (Jacomy et al., 2011). Additionally, a modularity measure is used to further structure the network graph. Modularity is often used in optimization methods for detecting community structure in networks (Newman & Girvan, 2004).

Figure 41 shows the network analysis of providers and their beneficiaries, both represented by nodes, while edges represent the relationship among them. Looking at the graph, several clusters of nodes can be seen, this indicates a close-knit community of providers who share common beneficiaries. The size of the nodes indicates the number of patients handled by that specific provider. The visual graph shows several community clusters colored similarly. These cluster groups represent providers and their beneficiaries within a specific county. Clear relationships can be observed between providers and their patients. Looking back at Figure 27, we can further examine two providers that were of question, namely provider 310001 and 310041. The orange colored node and surrounding clusters represent provider number 310001.

By examining the relationships between this provider and its beneficiaries, we see their patients focused around the center node (main provider). Compared to other nodes, beneficiaries tend to disperse less. Furthermore, edges (or relationships) from other neighboring nodes (other counties) are linked to this provider. Due to these findings, we can assume that the provider may be located in a high populated county, and or the provider is a centralized hospital, and less mobile compared to others. Additionally, we can assume that based on the large amount of relationships with beneficiaries within the same county and outside, this provider may have large dollar amounts in terms of claims, which evidently has been shown in Figure 29.

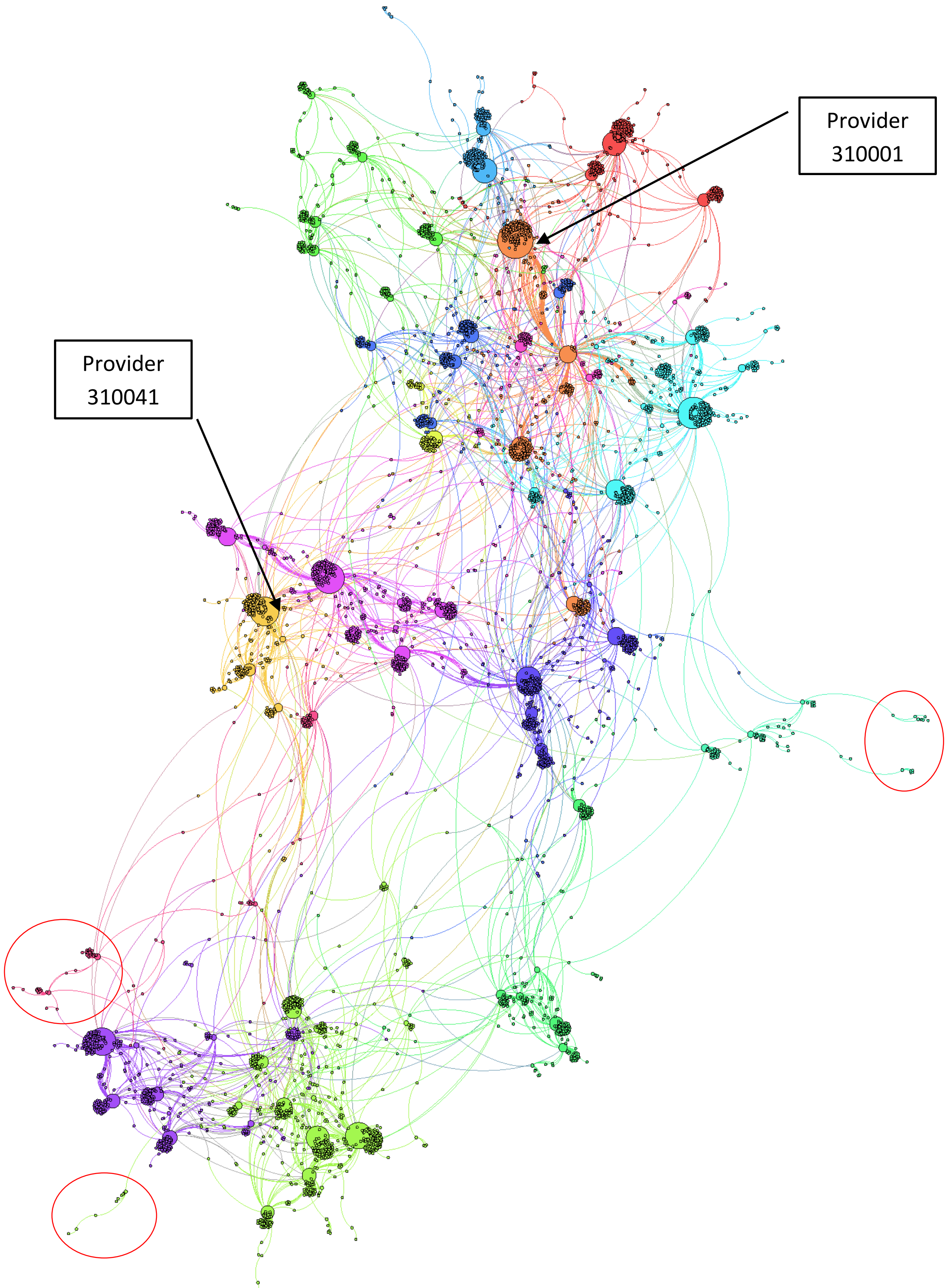


Figure 41: Network cluster analysis of providers and their beneficiaries

As for provider 310041, who had the largest number of Medicare claims at just over 14,000, their relationship with beneficiaries seems to be slightly more dispersed. Despite being less in size (node size), this provider has more beneficiaries generating within the same county. Due to this it can be assumed that this provider is located within a small-medium sized county and/or is a small hospital or practice.

VISUAL THOUGHT (6)

Figure 41 illustrates an optimal form of exploratory data visualization. In this example, no prior imagination of how the data would be portrayed was available. The visualization package simply produced a network analysis of providers and beneficiaries for the user to explore. However, a more ideal form of exploratory analysis would be to have the visualization package itself present a multiple of visualizations, without any user intervention made, and no metrics, attributes, or parameters decided on. By simply taking the data and creating different visuals randomly, users would purely be exploring the data without any bias in terms of limiting the visuals to what the user's imagination has already predetermined. In such cases, if the insights were achieved as a result of the visual portrayal and no by tables and statistics alone, then it can be truly said that the findings were due to the visual exploration of data.

Furthermore, in investigating the graph further, it can be noted that several providers seem to be isolated or secluded from others, as those circled in red. Three possible insights can be gained from this observation. The first is that these provider(s) give services to only a handful of recurring beneficiaries, and they are considered locals and/or regulars. The second, is that throughout the year, these provider(s) had non-recurring visits from only one or two beneficiaries only. The final insight may relate to specialized clinics/hospitals. These isolated nodes may relate to highly specialized clinics and/or hospitals that patients only visit a couple of times per year. Nevertheless, whatever the case, such scenarios require further review and investigation by auditors.

Finally, data animation is conducted. The animated visualization of claims was produced with the following criteria: 1) Patients were discharged the same day, 2) More than 8 diagnoses were conducted, and 3) the providers responsible for the claim received payments from Medicare in excess of $100,000. Timestamp transition of claim dates was applied to show temporal changes to data points. The animated visual was also based on an underlying layer graphing NJ county population rates as a heat map. Finally, the data points were categorized based on male patients (color coded in green) and female patients (color coded in red). Screenshots of the animated visualization can be seen in Appendix D[4].

Based on the animated visualization, the data points observed throughout the year require further investigation. Generally, patients that are discharged the same day usually indicate less severe cases than those required to be admitted. Moreover, these patients received more than 8 diagnoses. It can be assumed that the more diagnoses performed, the more likely the severity of the illness. However, these patients were discharged regardless. Furthermore, the payments received from Medicare by the providers exceed $100,000. Is it necessary to conduct more than 8 diagnoses for patients and discharging them the same day? Whether or not is this general practice, these providers received a large sum per patient and such cases are worth investigating.

[4] Full video can be found at: http://raw.rutgers.edu/docs/videos/Medicare.mp4

3.5. DISCUSSION OF RESULTS

Discussion is split into two parts: 1) Discussion on the Methodology used and 2) discussion on the results and findings.

3.5.1. Discussion on Results

Table 6 compiled the list of providers that were considered to be questionable during the in-depth visual analysis, as well as the number of times they were mentioned. It can be noted that based on the list of questionable providers, "AHS Hospital Corp", "Centerstate Medical Center Inc.", and "The Cooper Health System" were questioned the most. Due to exploratory nature of this study, auditors can utilize various forms of data visualization to pinpoint and focus on such questionable high risk providers and spend more time during the fieldwork stage investigating the causes and concerns. Furthermore, regardless of whether these providers are committing Medicare fraud or not, exploring the data in such ways may assist audits in conducting their fraud risk assessments.

Table 6: List of Providers in Question

	Providers in Question	**# Item was in Questions**
1	AHS Hospital Corp	3
2	Centerstate Medical Center Inc.	3
3	The Cooper Health System	2
4	Meridian Hospitals Corporation	1
5	Robert Wood Johnson University Hospital, Inc.	1
6	Hackensack University Medical Center	1
7	Health South of Toms River LLC	1
8	Acutecare Health Systems LLC	1
9	Trinitas Regional Medical center	1
10	East Orange General Hospital	1
11	Shore Memorial Hospital	1

In order to provide some form of validation for the results, the providers listed on Table 6 were researched using the United States Department of Justice website, and a few findings are shown in Figure 42[5]. Based on the research, 3 providers were found to be allegedly committing Medicare fraud, as highlighted in Figure 42. There were "AHS Hospital Corp", "The Cooper Health System", and "Trinitas Regional Medical Center".

Department of Justice

Office of Public Affairs

FOR IMMEDIATE RELEASE Thursday, June 21, 2012

New Jersey Hospital Pays U.s. $8,999,999 to Settle False Claims Act Allegations

AHS Hospital Corp., Atlantic Health System Inc., and Overlook Hospital, located in New Jersey, have agreed to pay the United States $8,999,999 to settle allegations that they violated the False Claims Act, the Justice Department announced today.

New Jersey Hospital to Pay $3 Million to Resolve Allegations of Medicare Fraud

WASHINGTON – The United States has entered into a settlement with a New Jersey hospital and filed a motion to intervene in a lawsuit against a New York hospital involving allegations that the hospitals defrauded Medicare, the Justice Department announced today.

The United States settled for $3.02 million, plus interest, with Trinitas Regional Medical Center in Elizabeth, N.J. The United States is also seeking to intervene in a lawsuit brought against Brookhaven Memorial Hospital in East Patchogue, N.Y.

NEWARK, N.J. – The Cooper Health System has agreed with the U.S. Attorney's Office for the District of New Jersey and the State of New Jersey to pay $12.6 million to settle allegations that it violated the federal False Claims Act and New Jersey False Claims Act by making improper payments to physicians under so-called "consulting" and "compensation" agreements as it sought to build its cardiology program.

Figure 42: Results on questionable providers

To recall, AHS Hospital Corp was flagged due to 3 reasons: Amounts charged for the provided services were not comparable to the amounts paid back by Medicare, the number of diagnoses and dollar amount of discharged patients differ significantly from the

[5] Results were obtained from the United States department of justice website http://www.justice.gov/

current trend, and had significantly higher payments from Medicare per diagnosis made. As for The Cooper Health System, it was flagged due to 2 reasons: Similar to AHS Hospital Corp, amounts charged for the provided services were not comparable to the amounts paid back by Medicare, and the total payments made from Medicare for deceased patients were considered anomalies compared to others. Finally, Trinitas Regional Medical center was flagged due to it receiving payments from Medicare for deceased patients. Nevertheless, the hope is that, through these efforts, direction may be established in deciding subsequent courses of action relative to activities such as information gathering, investigation, and review

3.5.1. Discussion on Methodology

The analysis conducted above illustrates the use of both exploratory and explanatory data visualization. The first section, tabular statistics and basic graphs, illustrates how explanatory data visualization can be used to communicate and present findings of predefined tests. The analysis had already been conducted and the primary purpose here was to present and discuss results. In these scenarios, explanatory data visualization is used. Results are graphed in ways that convey meaning and understanding to decision makers. However, it can be argued that tables can be used in a similar way to present such results. A counter argument can be made that data visualization has an advantage in its ability to summarize entire populations, and deal with multiple attributes and sources of data in one picture. However, users may still choose to use tables in their explanatory visualization, however, in doing so comes at a cost. Tables lack the ability to present Big Data. As attributes increase, and multiple datasets are used, presenting findings in tabular form becomes more difficult and time consuming. Hence, visualizing and explaining results using graphs is more effective and efficient in these scenarios.

The second section presented descriptive visual dashboards whose primary purpose again were to explain and communicate results of analysis in visual form. “Visual thought (3)” presented a table showing proof of how some dashboards can be presented in tabular form while still maintaining its integrity. Ultimately, it depends on the user’s perceptual preference, whether they prefer tables or graphs, and generally it depends not on what they wants to portray, but how and to whom they want to communicate it to. Nevertheless, one benefit of using visualization is it makes it easier and faster to pin-point any patterns that may arise by means of size and colors.

The final section introduced the in-depth analysis and advanced visuals. Starting with the geographical analysis, and despite the analysis being restricted to the state of New Jersey, when dealing with locational data, it is generally best explored using graphs that help picture the data points across the globe and discover patterns and outliers. When the data is large, and specific addresses are used, exploring and explaining results in a table is cumbersome, and hence visualizing them would be more pleasing and easier on the eyes.

Additionally, for most of the remaining visuals in this section, they illustrate how data was explored first using exploratory data visualizations, and then presented in a way to communicate certain insights using explanatory data visualization. Furthermore, as mentioned before, to be able to visualize data in a certain way, users generally need to decide before hand on the metrics, attributes and parameters to choose from. Users may choose to present the data in either tabular or graphical form, whether they are utilizing exploratory or explanatory data visualization. However, whether or not tables perform a similar job as data visualization depends on several aspects, including whether or not Big Data is in the picture, as well as the user's preference. Nevertheless, the power of data visualization in showing outliers and patterns within the data is a tremendous advantage.

3.6. CONCLUSION

In this exploratory study, a sample of Medicare data was analyzed for the purposes of developing questions, making observations, and facilitating knowledge discovery by means of exploratory data visualization. The information gathered and processed will be useful in ultimately conducting initiatives to address Medicare abuse, waste, and fraud.

The aim and contribution of this study is to showcase the use of data visualization throughout the planning stage of the audit cycle. Data visualization is valuable in many ways, especially with the advent of Big Data and potential information overload. It can help in exploring the data, gaining more knowledge and understanding, and conducting effective risk assessment tasks, all qualities that are useful when preparing the audit plan.

Furthermore, the unique insights gained from data visualization can help auditors assess and assign risk to certain items. Most of these insights would have taken more time and/or would have been difficult to gain if the auditors were to use just tables and/or other methods, specifically when considering the size and complexity of Medicare Data. Simple statistics may help, but not much is gained overall. In conclusion, the hope is through such efforts, auditors can utilize data visualization throughout their audit engagements to perform more effectively and efficiently.

CHAPTER 4: EXPERT KNOWLEDGE ELICITATIONS IN A PROCUREMENT CARD CONTEXT: A VISUAL DASHBOARD

4.1. INTRODUCTION

In the traditional audit process, auditors perform their audits by manually testing samples of data on an annual or quarterly basis which can become labor and time intensive. Moreover, with the advent of Big Data, several legitimate or illegitimate errors may go undetected. However, in order to provide continuous assurance in this era of data deluge, auditors will need to rely on more advanced analytical tools throughout the audit process in order to reduce such inefficiencies in time and labor, and further provide better detection of fraud and misuse.

With the adoption of Procurement card (p-card), as a function of purchasing in either a banking/financial firm or manufacturer, several critical internal controls have been found to be lacking. Specifically, separation of duties is not maintained (Gillett, 1997); employees no longer need authorization to purchase certain items; the receiving department does not receive the merchandise for verification; and the purchasing department no longer has the ability to shop around for the best deal. The same individual, the employee, is now making the purchasing decision, executing the purchase, receiving the merchandise directly, and submitting the verification paperwork (Gillett, 1997). With this substantial reduction in internal controls, the likelihood of employee fraud is drastically increased. Additionally, these transactions are frequent, and with this higher transaction volume, the temptation is greater to rationalize “sneaking in” a small personal use item within a large shopping basket of legitimate purchases.

Due to such lack of internal controls, Palmer and Gupta (2003) found an average procurement fraud abuse of $270 for each $1 million spent. Therefore, any firm planning to administer a p-card program should strengthen its internal controls for that area as well as its internal audit functions (NAPCP, 2004). Hence for a manufacturing firm that employs p-cards and that has many diverse functions and products, detecting p-card misuse will require a more advanced approach compared to a traditional one. Therefore, the main contribution of this paper is to present a methodology by which multiple techniques can be applied and combined to effectively and efficiently detect misuse. Additionally, this essay will illustrate how explanatory data visualization, in the form of a visual dashboard, can be used as means to display and present results of analysis.

4.1.1. Computer Assisted Audit Techniques (CAATS)

Basic analytics and general commercial computer assisted audit techniques (CAATS) serve as a basis for misuse or outlier detection. CAATS help analyze data, and turn facts and Figures to vital information. They allow auditors to perform previous manual intensive tasks quickly and efficiently thus improving audit efficiency (Zhao et al., 2004). The flexibility and power of CAATS help bring effectiveness and efficiency when dealing with huge amounts of data. CAATS also provide auditors and fraud examiners the ability to quickly and efficiently extract information from enterprise resource planning (ERP) systems. They offer a host of new opportunities to analyze data and find anomalies to detect potential fraud (Coderre 2009).

There are many CAATS and data analytics available (Coderre, 2009; Mueller, 2011; IDEA, 2013). These CAATs are more powerful than simple audit tests. They provide the possibility of mining and analyzing 100% of the data, turning facts and Figures to vital information, and potentially detect any violations of internal controls. Researchers and practitioners argue that CAATs will improve audit efficiency and effectiveness (Winogard et al., 2000; Manson et al., 2001; Bell et al., 2002; Braun and Davis 2003). Therefore, the flexibility and power of CAATs help bring effectiveness and efficiency when dealing with huge amounts of data. They also provide auditors and fraud examiners the ability to quickly and efficiently extract information and detect anomalies.

SAS No. 99 encourages auditors to use CAATs in evaluating fraud risks, evaluating inventory existence and completeness, and identifying journal entries. Specifically, the new risk standards (SAS Nos. 104-111) suggest that auditors use CAATs to select sample transactions to audit from key electronic files, test entire populations, sort transactions with specific characteristics, and obtain evidence about control effectiveness (AICPA 2002, 2006).

Companies should insure that internal controls for p-card usage are in place for both the authorization process and payment process. Nevertheless, control weakness can still occur and go undetected throughout the p-card process. For example, in the payment process, companies may have daily limits in place, where in such cases, employees may misuse the card by spending large amounts purposely split into many transactions that may not necessary exceed their per-transactional limit. Another example of control weakness in the authorization process is when managers do not review employee purchases to the full extent, and may authorize certain purchases that include items bought for personal use. Additionally, incomplete or missing information related to p-card purchases could be considered a control weakness. CAATS are effective and efficient in performing basic audit functions that would help resolve some of these issues (Coderre, 2009). Nevertheless, fraud and misuse may still occur and likely go undetected. Therefore, a more advanced and/or tailored approach is necessary.

However, CAATs are generally not tailored for company specific anomaly detection. One solution would be to utilize a rule-based expert system, an intelligent computer system that emulates the decision making ability of a human expert. Rule-based systems can capture and preserve human experiences and develop a more consistent system. These systems collect small fragments of human knowhow into a knowledge base used to reason through a specific problem.

4.1.2. Expert Systems

The audit environment is a highly complex decision making environment (O'Leary & Watkins, 1989), partially due to the increased amount of data generated by ERP systems in companies. Holstrom et al. (1987) identified numerous trends in information technology that are likely to have significant impact on audit evidence and the audit process in the next 10 to 15 years. O'Leary and Watkins (1989) believe that there is an indication of increased use of expert systems in the future. Nevertheless, expert systems have been successfully applied to several fields such as geology, chemistry, medicine, and engineering (Waterman, 1986; Harmon and King, 1985)

In the early 1960's, Edward Feigenbaum, reoriented the work in Artificial Intelligent (AI) by identifying and formalizing human expertise (Feigenbaum, 1961), and ultimately led to the creation of Expert Systems (ES). ES are learning algorithms that classify transactions by means of expert knowledge to solve real-world problems that normally would require human intelligence. These systems can capture and preserve human experiences. They collect small fragments of human knowhow into a knowledge base used to reason through a specific problem. This knowledge base is constructed through the formulation of conditional statements by using if (certain condition), and then (a consequent condition) rules (Flegel et.al, 2010; Abraham, 2005). Combinations of these rules based algorithms can be used to detect fraud (Chan, et.al, 1999). Figure 43 displays the main elements that encompass expert system architectures (Vasarhelyi, 1998).

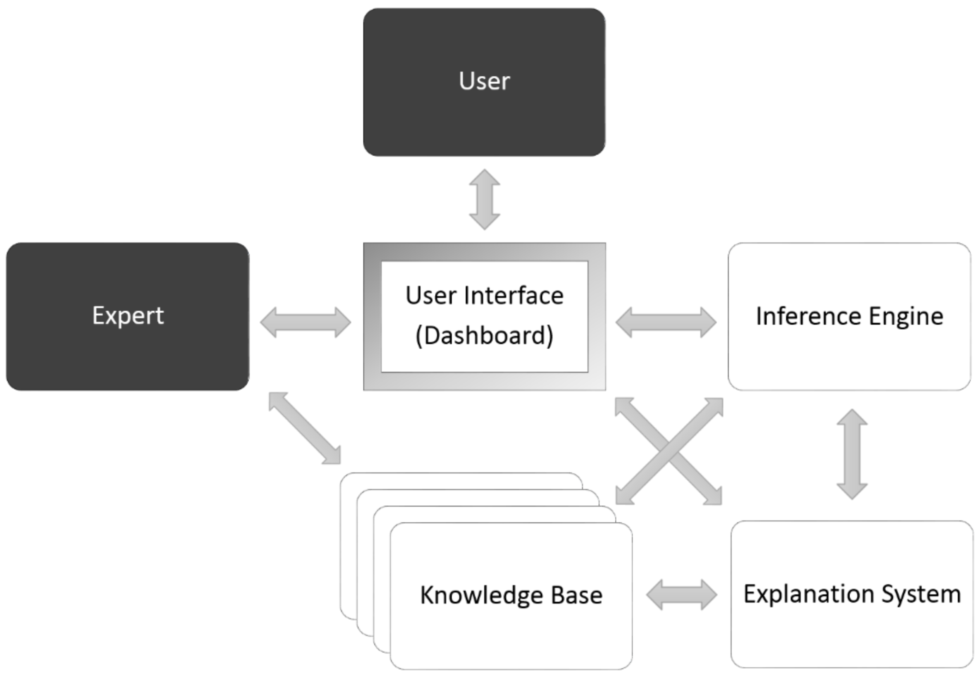


Figure 43: Typical expert system architecture

With respect to the accounting domain, ES were only introduced by 1977, when McCarthy (1977) developed TAXMAN, the first ES for tax applications. Following McCarthy, numerous examples of accounting-related ES emerged, including those used to reduce the risk of commercial bank loan default (Sangster, 1995), improve the productivity of credit authorizers (Leonard-Barton & Sviokla, 1988), improve problem solving in financial risk analysis (Fedorowicz et al., 1992), and increase accuracy of accounting reports (Back, 1993).

As for the auditing domain, prior research has found support for the use of ES. Eining and Dorr (1991) investigated the impact of an ES decision aid on novice auditors' experiential Learning and found that groups using expert systems significantly outperformed non-users on both accuracy and efficiency. Trewin (1996) investigated the impact of a tax accrual and planning ES on three offices of a national accounting firm and concluded that use of the expert systems improved productivity in many aspects. Finally, Eining et al. (1997), in an experimental study, compared different levels of decision aids, including ES, checklists, and statistical models, to determine which influences auditors most in assessing management fraud risk. Their results demonstrate that auditors make more accurate assessments with the use of ES than any other method. Other studies include those that found ES ensure consistency in audit planning (Baldwin-Morgan, 1994), enhance decision making (Gal & Steinbart 1987; Borthick & West 1987), and proved to aid auditors in developing more appropriate risk assessments (Lombardi, 2012).

There are also numerous examples of actual ES applications in auditing. AUDITPLANNER was a system that investigates how qualitative and quantitative information affects the materiality judgments carried out by auditors during their planning stage (Steinbart, 1987). Another example is RISK ADVISOR, an ES that evaluates audit risks and client's economic performance (Graham et al., 1991). RISC (Auditor Response to Identified Systems Controls) is another example of an expert systems that mimics the auditors' decision process in internal control evaluation (Meservy et al., 1986). Gal (1985) introduced an expert systems that specializes in assisting auditors in assessing the accounting internal controls in the income cycle.

In general, ES are reserved for accounting tasks that require expert judgement and heuristics in order to reach fast response. So, in structured accounting tasks that can be expressed algorithmically, such as the preparation of financial statement, sampling and ratio analysis, it is possible to use traditional computing software and word processing (Tomas, 1998). Nevertheless, expert systems are effective in responding to questions from a wide domain of knowledge as well as handling repetitive tasks in fuzzy domain of knowledge. Furthermore, ES have become means to achieve strategic competitive advantage for several firms and has become an important part of their overall strategy. Additionally, accounting firms have also recognized the importance of ES as a competitive tool in the accounting and auditing profession (Yang & Vasarhelyi, 1995).

ES allow for improved judgment and decision quality, and lower cognitive barriers when making decisions in a complex environment (Rose 2002) such as that of the audit. ES have also shown to have an overall positive effect on company performance and decision making (Mauldin 2003; Elmer and Borowski 1988). Moreover, prior studies have suggested that expert systems have a high level of accuracy (Bell et al., 1993). O'Keefe et al. (1993) suggests that expert systems have real-world impacts such as improving work quality, reducing time on tasks, and providing faster decision-making.

According to Hayes-Roth, (1984), ES generally have several common characteristics. They can generally solve difficult problems better than or as good as human experts. They can contemplate multiple hypotheses simultaneously and use heuristic programming and rules that handle uncertainty. Additionally, rules in ES can be easily added or deleted, the system is easily flexible to adapt any changes. Furthermore, because of their simple logic, they are relatively easy to implement in practice. Finally, ES can explain why they are asking a question and justify their conclusion. In order to utilize such systems to analyze p-cards, one needs to derive specific if-then rules and construct a knowledge base by reviewing internal procedures, audit programs, and interviewing auditors. Once a certain business process is mapped, rules can be identified.

4.1.3.. The Visual Dashboard

Explanatory data visualization is one type of data visualization that is part of a presentation phase, where we want to convey certain information in a visual form. With Big Data, companies need better ways, to not only explore data, but to synthesize meaning from it. Despite the use of analytic techniques by firms, an overwhelming amount of exceptions are still generated (Alles et al., 2006, 2008; Debreceny et al., 2003), causing the overall efficiency to decrease due to the limitations of human processing. Alles et al.,(2006, 2008) and Debreceny et al.,(2003) discussed this issue and pointed out that these exceptions are generally generated and sent to auditors without prior processing or sub-filtering. These scenarios raise the question of how users can organize and make sense of such voluminous data. Issa (2013) attempted to resolve the issue of information overload by proposing methodologies that would prioritize exceptions. Such attempts can help auditors focus on the more suspicious cases and make further investigation be more efficient.

Individuals have limited working memory, and alongside the challenges of processing large amounts of data, some valuable information may be disregarded during decision making. Visual dashboards may reduce this effect by optimizing information load and enabling users to focus on the relevant information at hand. By amplifying cognition and capitalizing on human perceptual capabilities, visual dashboards are expected to improve decision making. Hence, interest in visual dashboards has been well received and is growing, which is also evident from the proliferation of dashboard solution providers in the market (Yigitbasioglu & Velcu, 2012). Negash and Gray (2008) also regard them as one of the most useful analysis tools in business intelligence (BI).

Visual dashboards help managers identify trends, patterns and anomalies about their business. They can provide multiple purposes such as planning, monitoring, consistency, and communication (Pauwels et al., 2009). Since data analytics such as those of fraud detection tools, produce voluminous, complex data, they raise the question of how decision makers can organize and make sense of the large amount of data. An interactive data visualization dashboard that allows users to select and filter the information they wish to view, and make sense of the data produced, is an important tool for dealing with such issues (Dilla et al.,2010).

Dashboard terminology originates from the vehicle dashboard, which presents metrics for the driver, such as a speedometer, tachometer, odometer and fuel gauge. However, there is no clear definition of visual dashboards. Dashboard vendors generally define dashboards from the perspective of characteristics that their products have, while researchers talk about the different types of applications and different stages in dashboard development (Pauwels et al., 2009). Yigitbasioglu & Velcu (2012) provide a generic description of dashboards in that they are “a graphical user interface that contains measures of business performance to enable managerial decision making.”

In general, visual dashboards are graphical presentation of the current data, in a single, one page view. It incorporates various concepts and applications such as geographical maps, scorecards, and BI into one manageable solution. They provide users with the ability to visualize trends, key indicators, monitor activities, and evaluate performance. Their main purpose is to enable instantaneous and informed decisions effectively and efficiently.

Yigitbasioglu and Velcu (2012) distinguish between two types of design features for a visual dashboard: functional features and visual features. Functional features are the features that relate indirectly to visualization but describe what the dashboard can do. In this study, the functional feature will be the underlying expert system in place. The visual features, on the other hand, refer to how efficiently and effectively the information is visualized. Figure 44 illustrate how the two dashboard design features fit into the overall design architecture of a visual dashboard.

Technically, building a dashboard that presents several indicators and provides summarization is easy, however the main challenges is how to build one that is conceptually sound. One of the challenges here is to build a dashboard that despite being broad, remains simple, useful and easy to use. Therefore, if the constructs and features are not carefully crafted and kept to a minimum, the dashboard will be too complex for auditors to utilize and comprehend. Additionally, a challenge here relates to the way different users can interact with the dashboard and minimize the time spent to accesses the information. Therefore, defining user interaction and navigation features is essential. Finally, from an audit perspective, a dashboard needs to be in line with the criteria and approach that auditors have for control and compliance (Silveira et al., 2010).

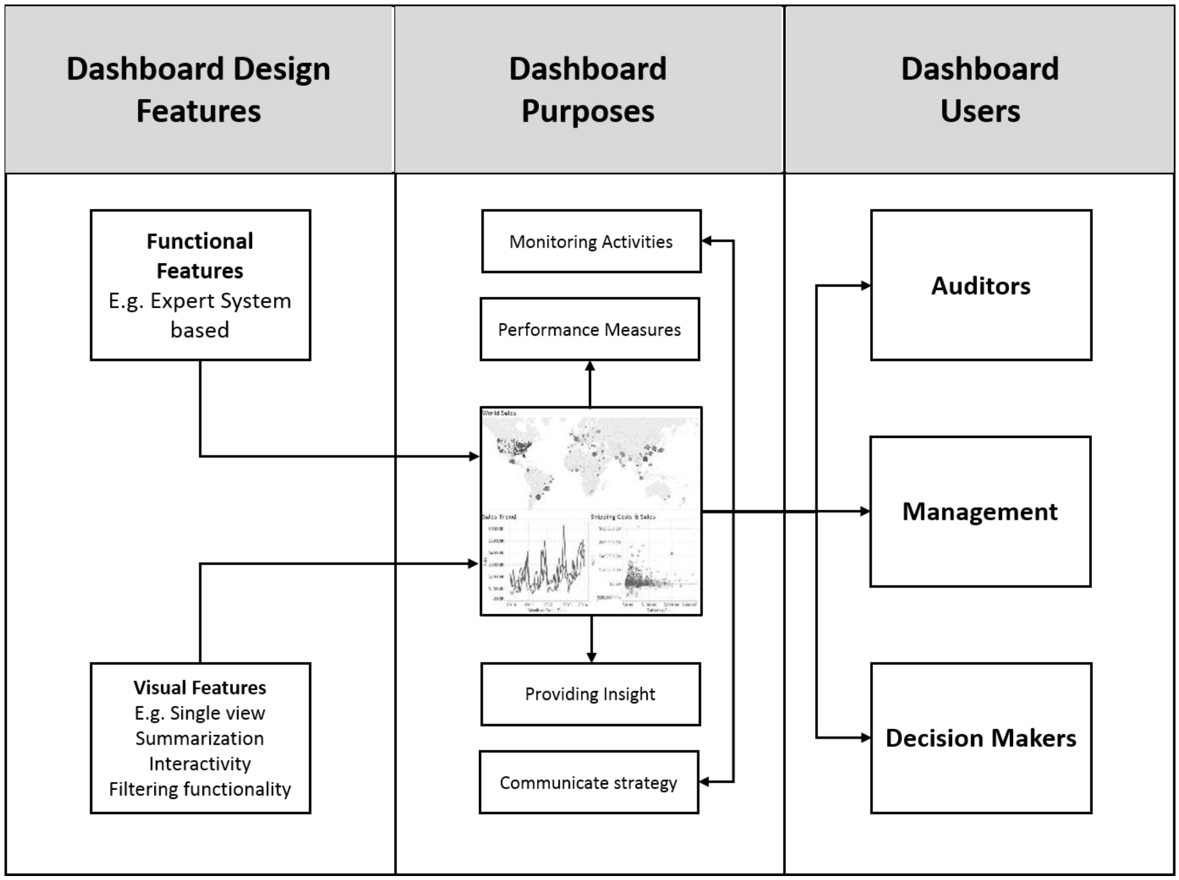


Figure 44: Visual Dashboard Design Architecture

Several studies have attempted to address the challenges relative to designing effective visual dashboards. Bellamy et al. (2007) studies the problem of designing dashboards for risk and compliance management, while Chowdhary et al. (2006), provide a model-driven framework that jointly helps in designing a business performance dashboard. Nevertheless, prior studies in visual dashboard implementations have found support for their use. Schulte (2006) found IBM's Business Objects Dashboard Manager usage at Edward Hospital improved its cash flow through better account receivables management. Also, Miller and Cioffi (2004) found that Unisys' marketing dashboard led to improved accountability, budget allocation, and performance management. However, only a handful of studies in the literature can be found providing guidance for practitioners (Pauwels et al., 2009) and researchers (Yigitbasioglu & Velcu, 2012).

Nevertheless, utilizing explanatory data visualization, in the form of visual dashboards, can further help facilitate the detection process and assist auditors in decision-making. They can have significant effects in guiding users toward a conclusion or inviting them to ask entirely new questions Nevertheless, creating such visuals requires preplanning, setting clear objectives, and obtaining the right visual elements.

Fraud detection is an area that requires many different approaches (Coderre, 2009; Breiman, 2002). Despite the abundance of studies related to general fraud and credit card fraud (Quah & Sriganesh, 2007; Brause et al., 1999), little could be found evaluating p-card misuse within a corporate setting. This lack of research attention is primarily due to the scarcity of feasible real-world data (Amat, 2002). There exist only a few studies that were undertaken with hypothetical, manufactured data (Phua et.al, 2010). Therefore, the contribution of this study is to bridge this gap by utilizing real world procurement data and applying a multi-dimensional approach for p-card misuse detection. Additionally, this paper contributes to the professional and research literature by introducing a methodology whereby an expert visual dashboard prototype is developed and implemented in a multinational corporation to detect p-card misuse.

Visualization in this essay plays a dual role: First in explaining how the expert system is constructed, and secondly, in displaying the results of that system by means of dashboard. The motivation behind the usage of the visual dashboard is that it may reduce the effect of limited working memory by optimizing information load and enabling users to focus on the relevant information at hand. Therefore, this essay will also contribute by illustrating how explanatory visualization can be used to provide meaning throughout the process via diagrams and graphs, as well as to communicate findings in a more effective and efficient way.

The paper is organized as follows. The first section will discuss the data used and discuss the overall p-card and monitoring processes. The next section will present the methodology used. It will then present the analysis and results. The final section concludes the paper and discusses relevant areas for future research.

4.2. THE DATA

4.2.1. Procurement Cards

Employee p-cards have been commended by public and private companies for their success in reducing purchasing department costs and increasing individual department purchasing decisions (Daly and Buehner, 2003). These procurement cards, which are company credit cards given to employees, were intended originally to replace petty cash. These cards typically enforce employee spending limits of under $1000 (Gillett, 1997). They have been utilized by many major U.S. corporations since 1991, such as PepsiCo (Garrison, 1997), and Merck & Company (Murphy, 1998).

P-cards generally operate like traditional charge cards where the user makes a purchase directly from a merchant, the merchant is paid by the bank that issued the card, and then finally the user’s company pays the issuing bank. P-cards system growth has been substantial over the past decade. Basically, the use of p-cards by firms reflects their desire to leverage their process efficiencies. Their use reduces costly delays from processing documents and purchase order requisitions for petty cash items and MRO (Maintenance, Repair, and Operating) supplies. Also, users may purchase items on a “just in time” basis, which helps departments run more smoothly. Furthermore, not only do p-card systems result in favorable overall and per transaction cost saving, but they also offer end-users greater flexibility to purchase critical resources (Daly and Buehner, 2003). However, procurement card users have a high volume of everyday purchase transactions, in contrast to travel cards that have purchasing typically linked to a specific event.

4.2.2. The Data

The data obtained for the purpose of this case study was from a multi-national firm that details every transaction from the preceding month of employee p-card use, and averages about 50,000 transactions with 55 attributes. The firm is headquartered in the U.S. with manufacturing, warehousing, and distribution centers located worldwide. The firm utilizes p-cards, however, there was not a comprehensive automated audit system in place to serve as an internal control for those p-card transactions. The firm had tried to integrate various CAATS and other tools as suggested by the credit card vendor, but these standardized applications were inadequate. The dataset was extensive and informative and included such attributes as dates, purchases limits, purchase amounts, items purchased, merchant/supplier details, authorization levels and department information. Table 7 provides descriptive statistics for the overall dataset obtained. Appendix E provides a summary list of the attributes used in the analysis.

However, several of the data fields had missing values. For example, vendors may choose the level of information that they want to provide, and some opt out of supplying purchase item description information due to data privacy concerns. Particularly troublesome is that these vendors are large retailers that carry a large variety of merchandise. So employees could be buying computer accessories legitimately or fraudulently from these stores, and there would be no way to verify except to pull the physical purchase records.

Table 7: Descriptive Statistics for P-Card Data

TRANSACTION DATE	EMPLOYEES	TOTAL AMOUNT	TOTAL MERCHANTS	ITEMS BOUGHT
MAR-13	2,921	17,894,293.99	11,229	20,688
APR-13	2,944	15,060,392.77	10,994	22,160
MAY-13	3,018	16,507,915.42	11,634	21,123
JUN-13	2,978	20,586,405.31	11,726	21,482
JUL-13	2,915	13,985,134.46	11,410	20,080
AUG-13	2,938	15,465,536.00	10,883	19,789
TOTAL	*17,714*	*99,499,678*	*67,876*	*125,322*

4.2.3. The Procurement Card Process

The firm’s procurement card process begins with assignment of p-cards to employees based on their authorization groups. Once p-cards are assigned, the cardholder can make purchases. A purchase is copied from the bank’s credit system and is then posted to the firm’s ERP system. The bank is then paid. The procurement card team then downloads all transactions and then uploads them for division management review. Meanwhile, the cardholder would file a basic expense report without receipts to his/her division manager. The manager either marks the transactions as reviewed and approved or needing clarification. Figure 45 summarizes the firm's overall procurement Process following the assignment of p-cards to employees.

As for the p-card monitoring process by the auditor, it first begins when they receive the monthly list of p-card transactions. Then the auditor would manually review each p-card transaction line-by-line to detect any misuse. Any transaction that violate the firm's p-card policies would then be flagged and sent to human resources (HR) for further investigation. After HR investigation, feedback is sent to the auditors regarding those flagged transactions. Figure 46 shows the overall p-card monitoring process for the auditor.

Figure 45: Firm's Procurement Process

Figure 46: Auditor's Monitoring Process

4.3. THE METHODOLOGY

The methodology undertaken in this case study is a multi-dimensional approach. First is that of basic CAAT tests to understand the data and construct an initial knowledge base. Then is the construction of the expert system by conducting interviews. Finally the visual dashboard is built having the expert system as the underlying facilitator.

The project began by establishing a basic framework which would apply to the creation of an expert system, an "electronic auditor", that requires intimate knowledge of the data. The initial dataset started in 2013 with the month of March and is has been updated monthly. The project would be conducted in tandem with the internal audit department's own audit procedures as the control. The first and initial processes would be of basic analytic CAAT tests, which provided the firm with immediate feedback. Like most businesses, management wanted to see results quickly and in a simple format (Kohavi et.al 2004). Following this, the second process would begin, which involves that of duplicating the department's manual tests by creating specific rules and relationship based codes.

The main objective of this project requires the elicitation of an expert's knowledge. Basically developing an expert system from a domain expert's knowledge. The first step in developing such a system is characterizing the knowledge of the expert (Hoffman, 1987). This exploratory stage can present a significant bottleneck in the system development process. Much time is consumed with the elicitation, identification and encoding of the expert's knowledge (Duda & Shortliffe, 1983; Bramer, 1982; Denning, 1986).

Depending on the size of the data files and complexity of the domain expert's knowledge, building a system can require from several months up to several years.

Furthermore, Hu (1987) suggests that expert systems will generally be successful if the task at hand is well understood and the human experts needed are available for knowledge extraction. Furthermore, certain conditions, if presented, will make the development of an expert system worthwhile. Namely a shortage of human experts in the domain and a high cost or critical requirement of expert advice (Hu, 1987).

Nevertheless, expert knowledge can be elicited with the following methods (Hoffman, 1987), as shown below in Table 8. The method of familiar tasks denotes observations of the expert at work with procedures that are typical (Hoffman, 1987). Although it is difficult to grasp at this stage exactly how an expert arrives at decisions (Mintzer & Messmore, 1984), a general idea of the knowledge and skills required for the project can be attained.

Table 8: Types of Methods That Can Extract the Knowledge of an Expert

METHOD CATEGORY	DESCRIPTION
METHOD OF “FAMILIAR” TASKS	Analysis of the tasks that the expert usually performs
STRUCTURED AND UNSTRUCTURED INTERVIEWS	The expert is queried with regard to knowledge of facts and procedures
LIMITED INFORMATION TASKS	A familiar task is performed, but the expert is not given certain information that is typically available
COSTRAINED PROCESS TASKS	A familiar task is performed, but the expert must do so under time or other constraints
METHOD OF “TOUGH CASES”	Analysis of a familiar task that is conducted for a set of data that presents a “tough case” for the expert

(Hoffman, 1995)

The structured and unstructured interviews is one of the most popular methods of attaining expert knowledge (Weiss & Kulikowski, 1984). However, this can be a particularly lengthy process, particularly for a data base as large as that provided for this project. These interviews are ideal for building initial knowledge (Hoffman, 1995). For the purpose of this study, interviews were conducted with the primary expert auditor responsible for detecting p-card misuse. Weekly calls were conducted as the auditor reviewed p-card transactions.

The rules were designed and tested during three iterative training cycles. After each training cycle the expert provided knowledge refinement to the rules, based on a review of the prior month results. After three training months, the first pass was conducted with the data from the month of June. By the second pass run, additional knowledge was attained due to further unstructured interviews, structured interviews, limited information tasks, constrained processing tasks, and methods of tough cases. The entire project structure as discussed before is shown in Figure 50.

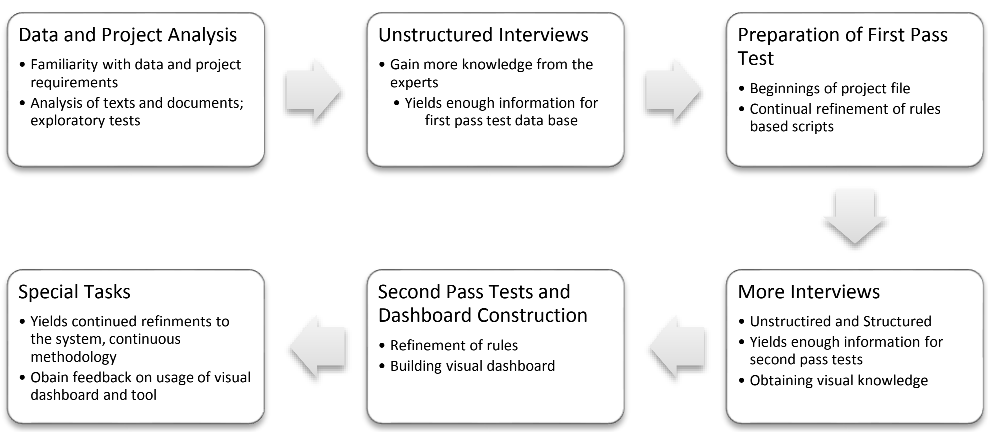


Figure 47: Steps for Extracting and Characterizing the Knowledge of the Procurement Card Experts

Finally the visual dashboard would serve as the system interface. It will provide the auditor the ability to monitor and control, and will consist of three main elements, Interaction, selection, and representation. Interaction involves the exploration of data by the decision maker. Selection refers to the navigation of complex data sets and choosing the optimal subsets. Finally, representation, which involves the mapping of data to a visual format (Spence, 2007). The main purpose of this visualization dashboard is to provide the experts with the ability to filter and select the data points they wish to focus on and ultimately have more flexibility in the analysis and review.

The main objective of this dashboard is to visualize the resulting data generated from the expert system. The first step was to locate the data and import it into the dashboard. The second step was to visualize this data. The third and final step is to provide interaction. The criteria for choosing the visuals were based on what the expert auditor found to be relevant. The third and final step was to provide filtering capabilities for the auditor in order for them to have flexibility and interactability. The visual dashboard is dynamic and can be changed based on the level of exceptions chosen.

4.4. ANALYSIS AND RESULTS

4.4.1. Initial CAAT Tests

Before constructing the expert system and visual dashboard, first preliminary CAAT tests were conducted. These tests include those of limits, as shown in Table 9. These tests were conducted with the initial two year dataset and not the monthly training sets.

Examining table 8, some transactions that have suspicious behaviors can be observed. For example, ID1929 has 574 transactions per day, which accounts for about 71 transactions per hour (assuming an 8 hour work schedule) and 1.2 transactions per minute.

Table 9: Results of Limit Tests

ID	Purchase Date	Total Dollar Amount Spent/day	Monthly Credit Limit	Single Transaction Limit	Transaction Per Day	Difference – Single Limit	Difference – Monthly Limit
ID2974	11/29/2012	267,087.6	75,000	2,500	141	264,587.6	192,087.6
ID1929	9/10/2012	136,551.8	60,000	10,000	574	126,551.8	76,551.8
ID5209	5/17/2012	99,599.0	75,000	2,500	3	97,099.0	24,599.0
ID1967	12/19/2012	96,250.9	75,000	2,500	3	93,750.9	21,250.9
ID1929	11/12/2012	99,821.1	60,000	10,000	193	89,821.1	39,821.1
ID3723	5/15/2012	89,625.3	75,000	10,000	421	79,625.3	14,625.3

Due to this initial investigation, there were transactions that exceeded monthly credit limits and/or single transaction limits. The firm established that some of the limits that were in these data fields reflected newly changed limits and not the historic limits that corresponded to the transactions at that time. Nevertheless, it was a concern that no daily limits were established. Either way, internal audit needs to review such cases to see if such behavior is normal or not and conduct further investigation.

During these tests, an internal control weakness became apparent, namely missing key information. Table 10 provides an example of such scenario where a key data field, such as purchased item description, is missing. In this example, one major vendor did not provide any item description information. This vendor accounts for 1% of the entire dataset; and all purchases with missing item descriptions count for about 25% of the entire population, amounting to around $33 million. Internal audit and management need to direct more attention to such cases where the opportunity to commit fraud is more apparent.

Table 10: Tests of Items with Missing Information

ID	City	Original Currency Amount	Merchant Name	Item Description	Product Code	Purchase Date
ID0484	ORLANDO	2,367.7	WM SUPERCENTER			
ID2934	CINCINNATI	2,472.9	WM SUPERCENTER			
ID0918	CINCINNATI	2,231.7	WM SUPERCENTER			
ID0918	CINCINNATI	2,450.2	WM SUPERCENTER		?	
ID0918	CINCINNATI	2,454.9	WM SUPERCENTER			
ID0918	CINCINNATI	2,499.4	WM SUPERCENTER			
ID0918	CINCINNATI	2,320.7	WM SUPERCENTER			
ID4347	JACKSON	2,459.8	WALMART.COM			
ID4347	JACKSON	2,384.5	WALMART.COM			

After gaining an understanding of the data and business process, and completing the initial CAAT tests, textual analytics were then conducted as the first stage of building the expert tool. Text analytics look for structure that is inherent in unstructured textual data and either applies semantic and/or statistical techniques to extract knowledge and patterns. Typical steps in text analytics include, retrieving the data for analysis, analyzing the data to identify, tag, and extract concepts, relationships, and information within different data sets, and finally classifying and organizing the extracted information according to a specified or generated taxonomy.

Text or keyword tagging, forms an important component of text analytics, and consists of identifying the names of entities in unstructured text (Prasad & Ramakrishna, 2010). Therefore a code was written to find specific predefined keyword tags and key phrases in a string of text, such as Gift Cards, Beer, Poker, and Alcohol. More than 300 keywords were obtained from the expert and some from other researched sources.

Table 11 provides an initial run test. It can be observed that results of the textual tagging show that all transactions were flagged as questionable, with one case (highlighted in red) identified immediately as fraudulent by the company. It is worthy to point out that after follow-up, the other cases were determined to be for legitimate use, however this does not mean that the system produced false positives as these cases are considered legitimate exceptions and were correctly flagged. However, with further rule refinements, the hope is to reduce any such instances.

Table 11: Results of Textual Analysis

ID	Purchase Date	Original Currency Amount	Extended Item Amount	Merchant Name	Item Description
ID1637	2/17/2011	0.0	50.0	STAPES 00101907	$50 APPLES ITUNES
ID1917	2/22/2012	0.0	7.6	STAPLES 00014472	POKER CHIPS 11.5G GAME ESSEN
ID0925	3/25/2011	84.9	75.0	AMAZON MKTPLACE PMTS	ITUNES GIFT CARD
ID4720	7/22/2011	0.0	10.0	BOLDEN INSTRUMENT	FUEL CHARGE $10
ID2503	10/6/2011	31.9	31.9	AMAZON MKTPLACE PMTS	PROACTIV SOLUTION ORIGINAL
ID0305	10/11/2011	16.3	12.9	AMAZON.COM	CONAIR TOUCH & TONE MASSAGER
ID2315	10/11/2012	49.7	41.7	STAPLES	STRESS BUSTER MASSAGE FOOT
ID5477	11/14/2012	24.5	22.0	AMAZON MKTPLACE PMTS	BRIDAL WEDDING JEWERLY HAIR

4.4.2. Knowledge Base Construction

Throughout the initial tests, weekly calls were taking place with the expert, and data relationships and associations were obtained to build a more refined rule based model. During these calls, the auditor would present p-card exceptions flagged and reasons for flagging would be discussed. Any specific reasoning the auditor had for flagging transactions was recorded and later formulated into specific rules. For example, under the

Merchant Category Analysis, one of the primary associations is between the merchant category code and credit card type; another is between merchant category code and department cost center. Table 12 provides a sample of a few of the merchant category transaction rules related to the keyword “car wash”. These rules a very specific and are tailored to specific cases. Most “car wash” transactions would fail the rule and be flagged as suspicious. This is based on the concept that most of the keywords recorded are considered inappropriate, and only a few cases of these keywords, if tagged, would pass.

Table 12: Rules by Merchant Category Analysis

MCH CODE 7542 (CAR WASHES):

IF (MCH_MCC_Description = “Car Washes”)
AND (Department_Cost_Center CONTAINS “Facilities Management” OR “Executive” OR “Buildings and Grounds”)
THEN → PASS.
 	
IF (MCH_MCC_Description = “Car Washes”)
AND (ACC_Master_Accounting_Code = EQUAL “GAS” OR “INCIDGAS”)
THEN → PASS.
 	
IF (MCH_Merchant_Name = “MR CLEAN CAR WASH”)
AND (Department_Cost_Center CONTAINS “PANELS”)
AND (FIN_Original_Currency_Amount > $50)
THEN → PASS.

Moreover, merchant categories that contain thousands of purchase items and transaction origins are complex and require frequent interviews with the auditor expert in order to replicate their procedures and judgments. The issue here is that purchases may be legitimate for one department but not for another. There are hundreds of associations that had and will need to be scripted for large vendors. For example, an employee working in the firm’s bakery division may purchase competitor products for research purchases but the employee working in the microchip division has no business buying that same bakery item. By the end of the project, more than 300 complex association and IF-THEN rules were collected, and, each rule branching into various sub-rules and relationships. Appendix F provides a sample of rules coded into the system.

Rules were categorized based on different levels of analysis, from those generating high false positives to those generating high false negatives. This will give the auditor the ability to decide whether they would want to look at the data from a higher level (levels 1, 2 and 3), generating more exceptions (probability of higher false positives), or a narrowed down level (level 4) to get a more focused view, generating only exceptions that adhere to very specific rules (possibility of higher false negatives).

Looking at the concept of employees getting a “car wash” for example. Figure 47 illustrates how the basic first 3 levels of analysis operate. Level 1 conducts a textual analysis on the items purchased and flags any transactions that have the keyword “car wash” tagged. For example, if an employee would drive their car to a local garage and receive a car wash by paying via their p-card, then the item description, “car wash”, is recorded by the bank after receiving the card information from the merchant.

As for Level 2, it works the same way in that it searches for the keyword “car wash” but in merchant names. So for the previous example, the keyword “car wash” would be recorded as the merchant’s name. However, in some instance the merchant name would not indicate “car wash” and might simply have the name “car garage”. Hence the availability of Level 1. Level 3 on the other hand does not conduct textual analysis, but conducts basic filtering on predefined rules specific to merchant categories.

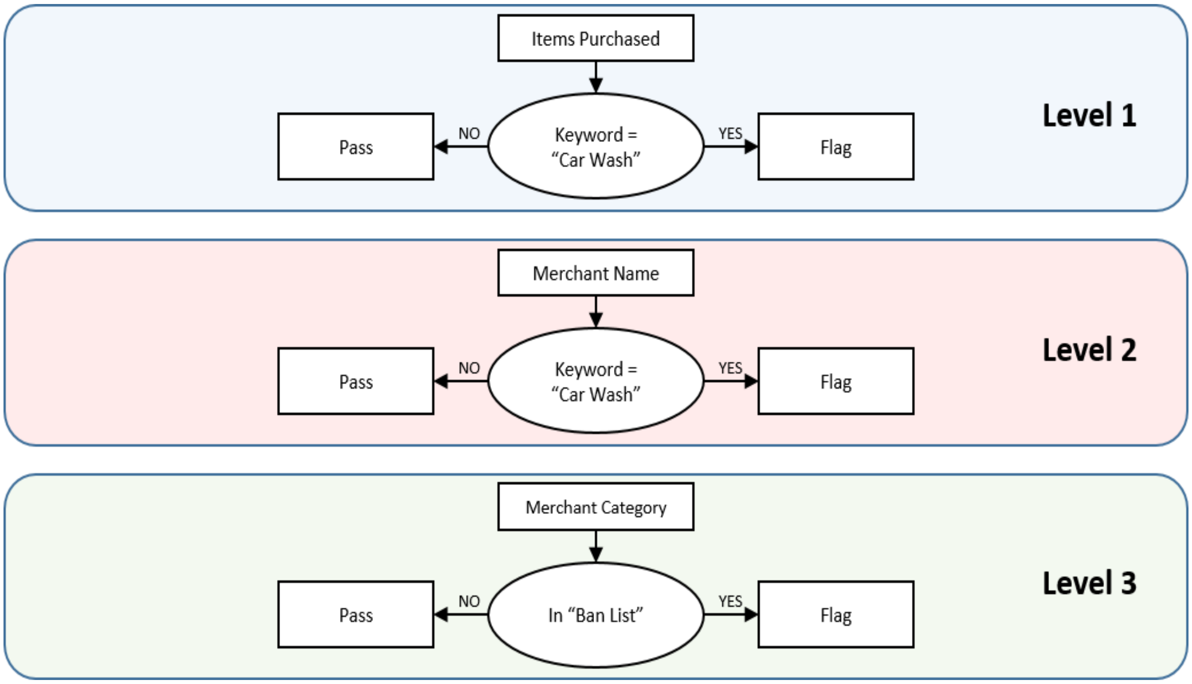


Figure 48: "Car Wash" Example of Levels 1, 2, and 3 of the Expert System

Figure 48 looks at how the more specific rules apply in level 4 of the system. This level contains most of the complex association and IF-THEN rules defined. It further breaks down textual analytics by applying specific rules relative to certain department policies or authorization codes. For example in this scenario, if an employee takes their car for a car wash, and the keyword is tagged by the system, a secondary stage of ruling is applied. One case might be to look at the department cost center the employee works at. If the cost enter contains the keyword "facility", then it means that the employee is allowed to receive car washes by paying with his company p-card. However, if the keyword "facility" is not present, then an additional stage of rulings is applied, and so on. This occurs until transaction are finally flagged as exceptions, or when all criteria for receiving a car wash are there and zero transactions are flagged.

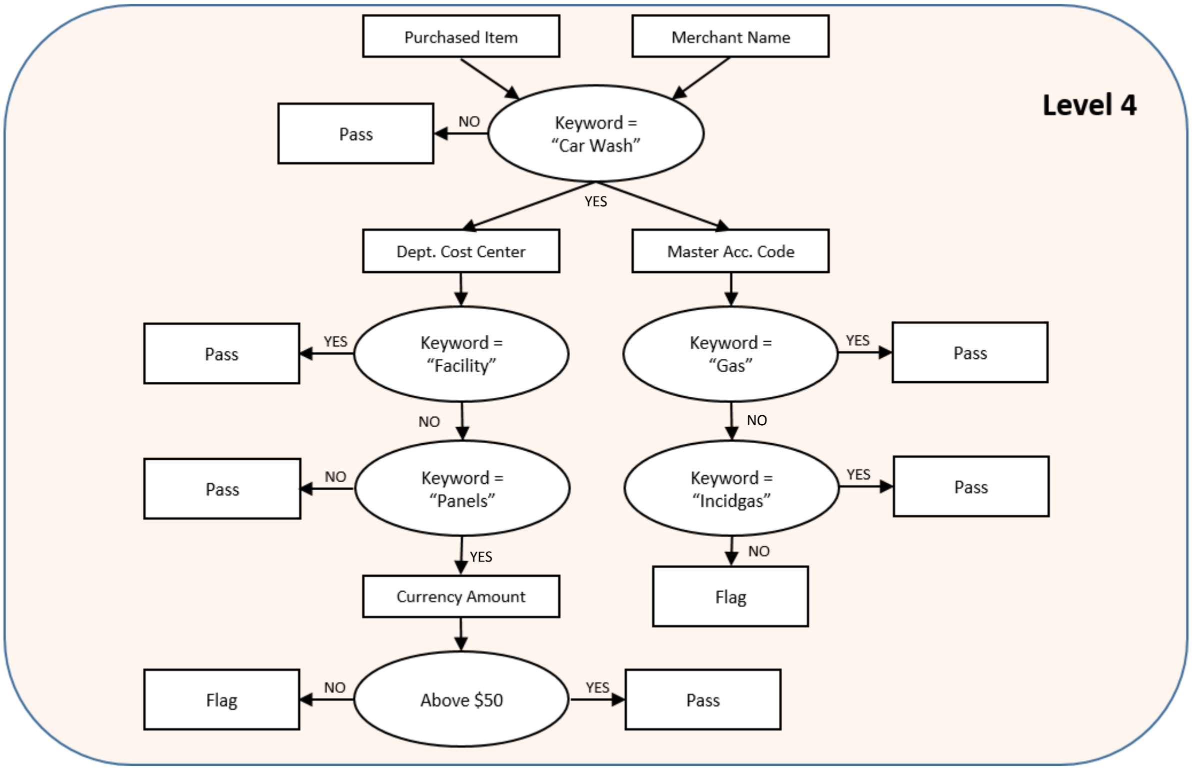


Figure 49: "Car Wash" Example of Level 4 of the Expert System

The rules of the expert system were coded in Visual Basic (VB) in EXCEL. Rules were coded based on a looping structure, were a set of filters will be applied until a stand point is reached. Figure 49 illustrates the coding concept. The code begins by identifying the level of analysis requested. For all levels, the code will run the keyword search based on the pre-defined list obtained from the expert. Once results are produced, the code will tag them and export them to a separate table. As for level 4 however, the code will continue the loop based on the specific rules and relationships defined. After the code obtains the keyword results for level 4, it will run additional set of rules defined. For every set of rules, the code will either delete the results (if transaction does not defy the rule), or continue to another set of rules (if the transaction defies the rule). After reaching the lowest set of rules, or after reaching zero results, the code will stop, and the output is extracted to a separate table.

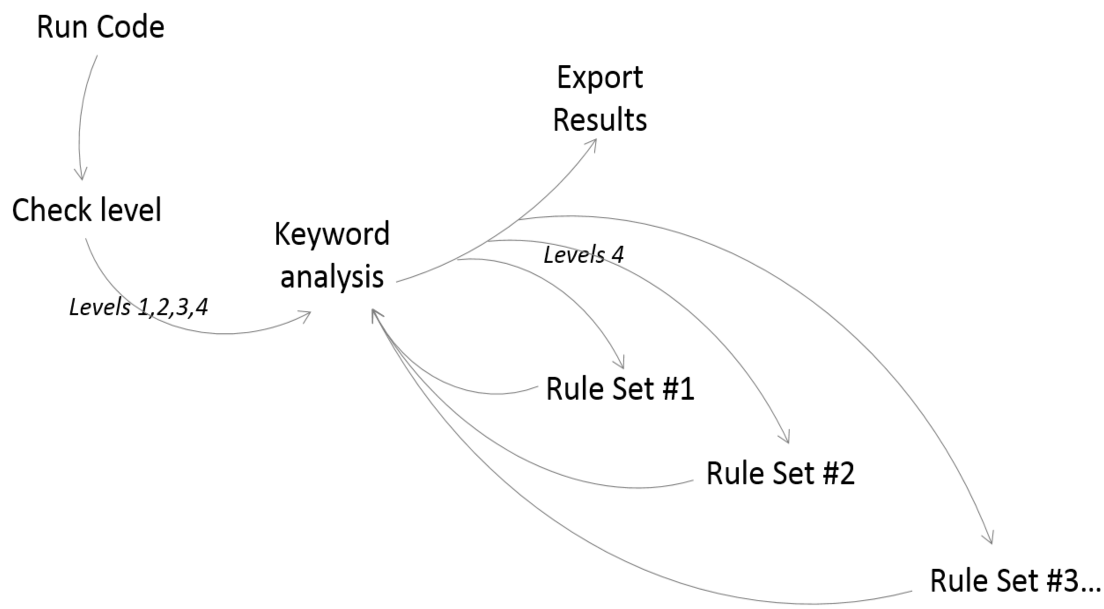


Figure 50: Looping Structure of Programmed Code

VISUAL THOUGHT (7)

Figures 47, 48 and 49 demonstrate how explanatory data visualization can be used to present and explain textual data, i.e. the rule system. Explanatory data visualization is not only restricted to numerical data, but these examples show how textual data, in the form of rules and associations can be represented in visual form, such as diagrams and graphs, to provide meaning and understanding to the readers. Another example of the understanding and explaining role of data visualization.

Additional knowledge was also extracted from the expert in terms of what are the optimal visuals that would help analyze and review the exceptions generated. Knowledge on type of filters most commonly used was also obtained from the expert during the weekly interviews. Based on the expert's perceptual view, the dashboard was constructed with 4 main visuals: Total dollar amount per user or employee, total dollar amount per merchants or seller, dollar amounts for user/merchant combination, and the distribution of tagged keywords. Six filtering functions were also implemented based on what the auditor looks for. These include: filtering by employee account name, merchant name, posting date, organization name and department cost center, and finally keywords flagged.

Rules obtained were validated by the experts during separate test runs. As the expert was interviewed and gave a list of rules and relationships, they would be tested and sent back for validation. The expert would then be interviewed a second time to discuss these results and they would either confirm the rules or provide additional feedback and refinement on them. This validation process was conducted multiple times until rules were fine-tuned or fully confirmed by the expert.

4.4.3. The Visual Dashboard

Explanatory data visualization can help pinpoint areas where further inquiries are needed to gain a better understanding of the underlying transactions. For example, certain outliers in the data may be visible and/or questionable items may need further investigation. Data visualization has the potential to enhance the efficiency and effectiveness of audit procedures. By applying different data visualization techniques, auditor can identify unknown patterns or relationships between the data, and communicate findings to stakeholders.

As discussed earlier in this chapter, the purpose of the visualization dashboard is to provide the experts with the ability to decide which level to focus on and have more flexibility in the analysis and review. It will be considered the main interface of the expert system. Dashboards take the center stage for reporting and performance management. Users generally would like to consolidate views of multiple sources and types of information into their dashboard workspaces, and discover hidden data relationships within a range of internal and external sources. Figure 51 provides a screenshot of the visual expert dashboard prototype. Appendix G presents detailed screen shots of each view shown in Figure 51.

Figure 51: P-Card Visual Expert System User Interface

4.4.4. Results

Connell (1987) presents 3 general criteria that can be applied to expert systems in order to measure their success: 1) their output accuracy compared to the human expert. 2) The extent to which their use saves time and money. 3) The interaction with the users in presenting and explaining the output. This first run measured the success of the visual dashboard in replicating the knowledge of the experts in detecting misuse. At that stage, the tool flagged 68% of the transactions that would have been pulled as suspect by the real auditor. After several hundred additional rules were added, the second pass test was conducted with the same dataset. The goal was to design an anomaly detection system that picks out as accurately as possible those outlier transactions that would be pulled by the expert and be subsequently subject to an internal audit.

According to feedback obtained by the firm and calculation made by the team, the tool replicated the expert's decisions with a 95% accuracy rate as seen in table 13. Additionally, based on the feedback obtained from firm, the tool managed to help the auditors be more efficient in analyzing the p-card transactions. They now can dedicate more time in other high risk areas instead of wasting time manually running through these transactions manually. Finally, the visual dashboard provided an effective and efficient portal for explaining the results and offering auditors insights on the resulting data. Based on the weekly calls conducted after the implementation of the prototype, the auditor mentioned their use of the visual dashboard as the primary interface for analyzing the exceptions.

Table 13: Results of first and second pass runs

	Red Flags Produced by Prototype	Red Flags Confirmed by Expert	Effectiveness
First Pass	1408	957	***68%***
Second Pass	1300	1235	***95%***

VISUAL THOUGHT (8)

With respect to audit tests and procedures, visualizing and communicating data in a certain way would be more useful than looking at the entire list of exceptions generated. Nevertheless, when auditors do find results or findings and would like to communicate these results to managers, they may be faced with the question of how those results would be communicated and explained best. Therefore, data visualization may play an important role as a learning tool in audit teams. Since one role of extracting knowledge from experts is to help and assist novice auditors in improving their experiential Learning, visual knowledge extracted from experienced auditors, can also provide guidance in how data is best visualized and looked at. It can also provide guidance on how to create flexible user interfaces, navigation tools, and search methods appropriate for each type of user, application, and task. Nevertheless, in future research, more focus can be put on extracting visualization knowledge of experts, and not only rules on procedures, policies and practices, and see if novice auditors would learn and benefit from the visualization knowledge extracted from those experts..

4.5. DISCUSSION & CONCLUSION

Throughout the study, there have been 172 cases of confirmed misuse procurement card transactions. Although a measure of confirmed fraudulent transactions from previous years could not be provided by the firm, its management did confide that the detection rate had increased by a large magnitude. The company is considering implementing the tool in their system in North America, followed by a modification to be adopted internationally.

Management is also discussing a transition from its current batch processing format of the procurement card transactions to one of a real time nature.

Detecting p-card misuse will require a more advanced approach compared to a traditional one, especially considering the voluminous amount of data associated with Big Data. This paper contributes to the auditing literature by introducing a multi-layer approach for which p-card misuse can be detected. Utilizing CAATs as a first layer of attack, followed by the construction of an expert system with a visual dashboard as the main interface, is an effort to enhance auditor performance in their fraud risk assessment.

In the future, the expectation is that this multi-pronged simultaneous attack of using CAAT tests, generating automated rules and relationships, and utilizing data visualization in the form of visual dashboards to detect unusual p-card activity will assist enormously in the area of occupational fraud detection. Furthermore, the tool can be applied on a continual basis, contributing to the continual journey of expert knowledge elicitation in a continuous auditing and monitoring environment.

Further research can also be conducted with respect to resolving the issue of missing information. Vital information was unavailable due to certain merchants not providing them. Such information may be the difference between legitimate transactions and illegitimate ones. Utilizing different techniques such as profiling, pattern recognition, and visual exploration may serve as starting points to dealing with this issue. Additionally, future research can be conduct in generalizing the visual expert dashboard. Currently rules in place only pertain to company specific policies. The hope is, in the future, such systems can be modified to operate in certain industries and not just one specific company.

CHAPTER 5: DISCUSSION AND CONCLUSION

5.1. GENERAL DISCUSSION

With the advent of Big Data, companies are facing major challenges in terms of processing and analyzing it. Such Big Data has been facilitated by many technologies, such as RFIDs, social media networks, video and audio streams, cell phone GPS data, and ERP systems. Nevertheless, companies who are able to leverage on Big Data have shown to have gains in productivity. Despite that, one obstacle companies still face is making sense of such voluminous amounts. Auditors too can benefit from Big Data as it provides them with more access to information, leading to effective identification of anomalies and risk assessment tests.

Data visualization is one technique that helps auditors summarize large volumes of data to a manageable size and focus on those data points which are critical to the task at hand. It aims to integrate individuals in the data exploration process, applying their perceptual abilities, flexibility, creativity, and general knowledge to the large data sets available in today's systems. As well as facilitating the explanation and understanding of the results generated from the exploration phase, to effectively communicate these findings to the decision makers. Data visualization is less complex compared to other analytical tools and is easily interpreted, it helps auditors gain insights, draw conclusions, come up with hypotheses, and present and convey meaning and understanding to the stakeholders.

In this era of Big Data, traditional audit procedures are not sufficient to provide high-level assurance. Several forms of risks can be missed throughout the audit process. One reason for this is that traditional audit procedures do not allow auditors to effectively utilize big data to allow for a more complete and through risk assessment process. Data visualization can provide auditors with some assistance in fulfilling their tasks such as that of fraud detection and risk assessment. In addition, it can be used to supplement their analytical procedures in order to discover previously unknown risks and ultimately improve audit quality. By integrating data visualization into the audit process, it allows auditors to generate more effective risk-oriented audit objectives, providing deeper insights into the data, ensure high quality services.

The primary goals of data visualization are to explore and/or explain the data. This dissertation has presented different essays by which both techniques can be applied. However, when it comes to deciding which is optimal, it depends primarily on the task and setting at hand. Accounting data, for example, is simple as compared to, let’s say, medical data, and therefore visualization techniques would differ. Additionally, financial and/or transactional data may limit the imagination of users to produce unique visuals due to their simplicity. Hence, one would argue that in a simple auditing setting, explanatory data visualizations is more relevant.

The first essay presented a good example of utilizing explanatory data visualization in conducting the in-depth literature analysis. Communicating findings in such visual form help users gain insights from the data and come up with unique conclusions. Another example of a successful explanatory data visualization comes from Hans Rosling at his talk on February 2006 at the Technology, Entertainment, and Design (TED) conference. Rosling is a professor of Global Health from Sweden, he collected public health statistics for different countries for a given year, plotting them on a scatterplot, and then tracing their progress through time, providing the audience with a highly effective explanatory visual of the data

Accountants and auditors are generally limited in their visual imagination primarily because they only deal with the same data periodically. Possibly, when auditors start implementing and utilizing Big Data in their audits, then the scope of their imagination would likely to broaden. Therefore, one could argue, that for a simple accounting or auditing setting (non-big data), the understanding/explanation aspect data visualization becomes more relevant, and possibly the sole driver. On the other hand, when Big Data comes into the picture, then visual exploration would lead, followed by visual explanation. The second essay illustrated how this can be applied when dealing with Big Data. Both exploratory and explanatory data visualization was used throughout the analysis. It also discussed the argument of whether or not tables may perhaps do a similar (or better) job as graphs. Generally, this goes back to the user's preference, however, when it comes to Big Data, graphs will outperform tables.

Anscombe (1973) presents a phenomena, named Anscombe's quartet, which demonstrates how important it is to visually explore and explain data, rather than relying on tabular statistics alone. The example is show in Table 14 which provides 4 datasets that have the same mean, variance, correlation coefficient, and line of best fit. One could assume that these datasets having the same statistics would be very similar when visualized. However that is not the case, as presented in Figure 52, where we see when each dataset is visualized they have different features. Effects of curvature and outliers have significantly thrown off the statistics, and by exploring the data this way, patterns and/or deviations in the data become clear (Anscombe, 1973; Tufte & Graves-Morris, 1983). Anscombe's example shows proof of how powerful data visualization is when it comes to the detection of patterns and/or outliers.

Table 14: Anscombe Sample Data Set

Dataset 1		Dataset 2		Dataset 3		Dataset 4	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

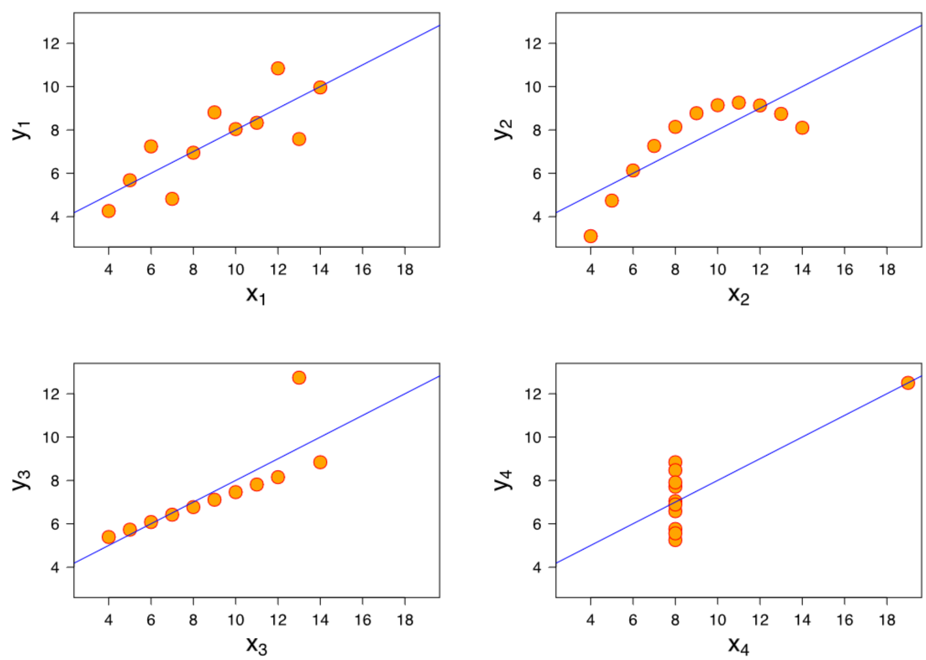


Figure 52: Datasets from Table 15 graphed and visualized

Data visualization also plays an important role in teaching and conveying understanding through graphics and diagrams. As demonstrated in the third essay, visualization is not only restricted to numerical data, but the examples shown in this chapter illustrate how textual data, in the form of rules and associations can be represented in visual form, providing meaning and understanding. There is no point having a visual expert system if you can't convince decision makers of its effectiveness. Graphical displays, such as those presented in this essay, provide effective ways of communicating complex content, such as that of the coding structure. This can be related back to the visual argument hypothesis, which concentrates on the perceptual and interpretation processes that take place when users learn from graphical representations. It claims that diagrams and visual graphs are more effective than text for communicating complex content because they are less demanding in terms of processing (Waller, 1981; Vekiri, 2002)

5.2. CONCLUSION, LIMITATIONS AND FUTURE RESEARCH

Current data visualization research in auditing focuses on the comparison between different presentation formats and their effects on auditors. Little to none can be found on the application of data visualization techniques through the audit process. Therefore, this dissertation attempts to contribute to the auditing literature by exploring the concept of data visualization and its applications in the audit process. It introduces a new approach for conducting audit tasks. Data visualization can enhance existing audit procedures by exploring and testing predefined audit objectives derived from fixed management assertions. The dissertation is comprised of three main essays, each addressing different aspects and contributing in unique ways. The following summarizes the 3 essays and their findings.

The first essay provides a literature analysis of data visualization. It begins by discussing the major milestones in history of data visualization by showing the time distribution of events. From the 16th and 17th century, were the earliest forms of data visualization were introduced in terms geometric diagrams, astronomy, map making, and navigation, to the mid-1800s, which was known as the “Golden Age” of statistical graphics, where data visualization was at its highest. Nevertheless, data visualization has been around for many years, and throughout history, a wide variety of advancements have contributed to the widespread use of it today.

This essay also investigates the usage and research of data visualization in the auditing literature. Prior research has examined the importance of visualization and presentation format and its linkages to decision-making performance as well as how it affects the judgment development process with respect to analytical procedures. However, many of these studies found varying results. So understanding the past literature relative to data visualization is crucial in bridging a path towards better research in the future. Hence this essay conducts an in-depth visual literature analysis of 35 research publications in the area of data visualization, focusing on the audit and decision making literature.

Findings suggest that data visualization started slow, and as time progressed, more and more research was conducted, but with non-supporting results. However, by the year 1991, when the theory of Cognitive fit was introduced, many researchers experimented with the concept and found support for the usage of data visualization in certain tasks.

Nevertheless, as time got closer to our century, and we see a new breed of applications that have flexibility and provide multiple functionality, researchers started acknowledging the benefits and usage of data visualization.

The second essay focuses on the application of data visualization during the audit planning stage. Specifically, this essay will showcase the application of exploratory and explanatory data visualization in Medicare health insurance for the purpose of knowledge discovery. One benefit of Data visualization is that it accommodates large datasets, and can further assist in the process of knowledge discovery by leveraging the auditor's abilities to perceive patterns and structures in visual representations. In doing so, it can help auditors gain a general understanding of the overall processes, raise questions and conduct risk assessment tests to identify any potential weaknesses, and check overall data integrity and validity.

The analysis focuses on three main methodologies, tabular statistics and basic graphs, descriptive visual dashboards, and most importantly in-depth analysis and advanced visuals. The insights gained from viewing the visuals produced in this essay can help auditors assess and assign risk to certain items. Furthermore, most of these insights would have taken more time and/or would have been difficult to gain if the auditors were to use just tables and other methodologies, specifically when considering the size and complexity of Medicare Data.

The third and final essay deals with the construction of the expert visual dashboard. It utilizes real world procurement card data and applies a multi-dimensional approach for p-card misuse detection. P-cards have several benefits, but despite that they are still prone for misuse, and the current state with which the company this essay deals with analyzes procurement card data for misuse, is manually. Therefore the main objective of this essay is to construct a visual expert dashboard, and in order to do so, it requires the elicitation of an expert's knowledge.

Expert systems are learning algorithms that classify transactions by means of expert knowledge to solve real-world problems. As for Visual dashboards, they are graphical presentation of current data, in a single, one page view. They incorporates various concepts and applications such as geographical maps, scorecards, and BI tools into one manageable solution. They provide users with the ability to visualize trends, key indicators, monitor activities, and evaluate performance. In this study, the visual dashboard will serve as the primary user interface for the expert system.

The data obtained consists of 50,000 records with 51 attributes for each month. Several training months were conducted and by the end, more than 300 complex association and IF-THEN rules were collected, each branching into various sub-rules and relationships. Additionally, perceptual views of the expert were extracted, in terms of different visuals and filtering methods the expert would likely to use to explore and explain results. The rules were then coded in and the visual dashboard was constructed as the user interface by which the auditor would review, filter, and focus on specific exceptions. The final tool managed to replicate the expert's decisions with a 95% accuracy rate. Additionally, the tool managed to help the auditors be more efficient in analyzing the p-card transactions. They now can dedicate more time in other high risk areas instead of wasting time manually running through these transactions.

5.2.1. Limitations

This dissertation has several limitations. Data visualization has a limitation on the output method used. Computers and display hardware make it easy to produce pictures, and if they are used to display the visuals then generally there is no issue. However, when visuals are displayed on paper, like in an academic journal for example, then the graph will be limited in terms of different aspects. For one, depending on what kind of printer is used, the colors of the visuals may not be available. Another limitation relates to interaction. If visuals are printed on paper, then there is zero interactability. Users will not be able to zoom, filter, and select data points within the visualization. A final example of the limitations of data visualization is on the dynamic change of data, or animated data. When data is animated it is difficult to present it in a meaningful way on paper. Without a display interface, animated visuals will simple lose their functionality.

Another limitation relates to visualization packages. These packages generally have a limited capability of displaying unique or custom visuals. If a user would want to present data in a unique way, they may be limited by types of visuals available in the software package. These users would have to custom code visuals in order to portray what their imagination entails.

Specific to the first essay, one limitation relates to the small number of publications used as the dataset in analyzing the literature. Moreover, the methodology used to analyze the literature is considered basic. In the second essay, due to computational processing power and certain time restrictions, it only managed to analyze Medicare data for the state of New Jersey. Furthermore, there was no proper validation of results. Aside from the results researched from the United States department of justice website, no official feedback was obtained from Medicare. Finally, the third essay deals with a small expert panel, namely one expert auditor in the area of p-cards. This limitation was due to the fact that the company only hired one auditor to analyze the p-card transactions. Moreover, due to the large amount of rules and relationships implemented in the tool, it has become specifically tailored to that company. Thus any attempts to generalize this tool to other companies might produce less effective results.

Nevertheless, despite these limitations, this dissertation fills a gap in the auditing literature by showcasing the application of data visualization in different areas. Data visualization has been successfully deployed in medicine, genetics, biology, engineering, and many other scientific fields. However, auditing on particular has been behind in the usage of data visualization. The auditing and accounting literature has put more focus on the comparison of different visual techniques and their impact on decision making, particularly limited to the traditional “graph versus table” studies. Moreover, what is really lacking in the literature are examples of how certain data visualization techniques can be helpful in auditing. For example how they can aid auditors in the audit cycle, from the planning stage, to fieldwork, and finally to reporting the results to management

5.2.1. Future Work

Further research can be conducted to not only explore and/or explain data using data visualization, but to tell stories too. The human brain has evolved to remember and understand stories more easily than other presentations of information. Prior research has shown that when people talk about facts, only two areas of the brain are activated (language processing and comprehension), however, when a story is heard, multiple areas of the brain are activated (Gershon & Page, 2001).

Potential future work related to the first essay can relate to the usage a wider range of publications, in terms of depth (number of articles) and breadth (number of journals). Additionally, more advanced visual methods can be applied to further understand the related literature. Future work in the second essay can start by focusing on the entire United States, establishing benchmarks, providing comparative visuals of different states, and ultimately delivering more insights. Furthermore, working directly with the Centers for Medicare and Medicaid Services, feedback can be obtained regarding the visual analysis, and in doing so, it brings validation to the results.

As for the final essay, in an attempt to resolve issues related to the small expert panel these, more experts from different companies can be interviewed and a more extensive, but generalized, knowledge base can be developed. Extracting the visual knowledge from the expert in terms of how they look at the data and what filters they would apply, may not only serve means to explain the data but can also serve as a learning tool to help novice auditors. So as future research, it would be good to see whether novice auditors would benefit from those experience auditors who suggest visualizing data in certain ways may help in the review process of exceptions generated relative to p-cards analysis.

Daniel Kohn, a Brooklyn-based painter and conceptual artist, has been collaborating with Albert Einstein College of Medicine in New York to continually reimagine the ways in which data can be visualized. His main role is to help scientists with

Big Data overload. This new venture would jolt scientists out of their comfort zone of seeing data in certain ways. By the help of graphic designers, visual artists, and/or conceptual painters, such as Daniel Kohn, Data Scientist are pushed into thinking about representations in new and different ways. Having designers and artists chose how graphs are displayed and what to look for beforehand, may help scientist look at data in ways they did not before.

Perceptual psychologists and graphic designers have long understood many of the benefits and challenges associated with visualization. However, due to the relative isolation of various scientific disciplines, very little of their knowledge appears to have reached other fields (Reuter et al., 1990), especially those currently dealing with large amounts of data and require means to make sense of them all. The aspects of good graphic designs and human visual perceptions must be taken into consideration, and users who wish to design effective visualizations must collaborate with these fields in order to make visuals convey the insights that they are meant to convey (Reuter et al., 1990).

The hope in the near future is for visualization software developers, graphic designers, artists, and perceptual psychologists to collaborate in order to understand how humans interact with and perceive information, and ultimately design effective visualizations that convey the right meaning the right way.

REFERENCES

1. Abraham, A., (2005). Rule-based Expert Systems. Handbook of Measuring System Design, John Wiley & Sons, Ltd, 909-919(8)

2. Alles, M., Brennan, G., Kogan, A., & Vasarhelyi, M. (2006). Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at Siemens. International Journal of Accounting Information Systems, 7(2), 137–161

3. Alles, M., Kogan, A., & Vasarhelyi, M. (2008). Putting continuous auditing theory into practice: Lessons from two pilot implementations. Journal of Information, 22(2), 195–214

4. Alles, M. (2015) Drivers of the Use and Facilitators and Obstacles of the Evolution of Big Data by the Audit Profession. Accounting Horizons In-Press

5. Altman, E. I. (1983). Multidimensional graphics and bankruptcy prediction: a comment. Journal of Accounting Research, 297-299.

6. Amat, J.L. (2002) Using reporting and data mining techniques to improve knowledge of subscribers; applications to customer profiling and fraud management. Journal of Telecommunications and Information Technology, 2/2002.

7. Amer, T. (1991). An experimental investigation of multi-cue financial information display and decision making. Journal of Information Systems, 5(2), 18-34.

8. American Institute of Certified Public Accountants (AICPA), (2001). Analytical procedures. American Institute of Certified Public Accountants

9. American Institute of Certified Public Accountants (AICPA). (2002). Consideration of Fraud in Financial Statement Audit. Statement of Auditing Standards No. 99. New York NY: AICPA

10. American Institute of Certified Public Accountants (AICPA). (2006). Audit Risk Exposure Standards. Statements of Auditing Standards New York NY: AICPA

11. Anderson, J. C, & P. M. J. Reckers. (1992). An empirical investigation of the effects of presentation format and personality on auditor's judgment in applying analytical procedures. Advances in Accounting 10:19-43

12. Anderson, J. C., & Mueller, J. M. (2011). The effects of experience and data presentation format on an auditing judgment. Journal of Applied Business Research (JABR), 21(1).

13. Anderson, J.C. & S. E. Kaplan. (1992). "An Investigation of the Effect of Presentation Format on Auditors" Noninvestigation Region Judgments" Advances in Accounting Information Systems. (1): 71-88.

14. Anscombe, F. J. (1973). Graphs in statistical analysis. The American Statistician, 27(1), 17-21.

15. Asare, S. K., & A. M. Wright. (2004). The effectiveness of alternate risk assessment and program planning tools in a fraud setting. Contemporary Accounting Research 21 (2): 325-352

16. Association of Certified Fraud Examiners (ACFE), (2010). Report to the Nations on Occupational Fraud and Abuse. ACFE 2010 Global Fraud Study

17. Back, B. (1993). Validating an expert system for financial statement planning. Journal of Management Information Systems, 157-177.

18. Baldwin-Morgan, A. A. (1994). The impact of an expert system for audit planning: Evidence from a case study. International journal of applied expert systems, 2(3), 159-174.

19. Beattie, V. A., & Jones, M. J. (1993). Effect of graphical presentations on insights into a company's financial position-an innovative educational approach to communicating financial information in financial reporting: a comment. Accounting Education, 2(4), 303-309.

20. Bell, T. B., S. Szykowny, and J. J. Willingham. (1993). Assessing the Likelihood of Fraudulent Financial Reporting: A Cascaded Logit Approach. Working paper, KPMG Peat Marwick, Montvale, NJ.

21. Bell, T. B., Bedard, J. C., Johnstone, K. M., & Smith, E. F. (2002). KRiskSM: A computerized decision aid for client acceptance and continuance risk assessments. Auditing: A Journal of Practice & Theory, 21(2), 97-113.

22. Bellamy, R. K., Erickson, T., Fuller, B., Kellogg, W. A., Rosenbaum, R., Thomas, J. C., & Vetting Wolf, T. (2007). Seeing is believing: Designing visualizations for managing risk and compliance. IBM Systems Journal, 46(2), 205-218.

23. Benbasat, I., & Dexter, A. S. (1985). An experimental evaluation of graphical and color-enhanced information presentation. Management science, 31(11), 1348-1364.

24. Beniger, J. R., & Robyn, D. L. (1978). Quantitative graphics in statistics: A brief history. The American Statistician, 32(1), 1-11

25. Bertin, J. (1967). Semiologie Graphique: Les Diagrammes. Les Reseaux, Les Cartes.

26. Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps.

27. Blocher, E., R. Moffie, & R. Zmud. (1986). Report format and task complexity: interaction in task judgments. Accounting. Organizations and Society 11 (6): 457-470.

28. Blocher, E., & Patterson Jr, G. F. (1996). The use of analytical procedures. Journal of Accountancy, 181(2), 53

29. Bolton, R.J. and Hand, D.J. (2001). Peer-Group Analysis - Local Anomaly Detection in Longitudinal Data. Technical Report 2001, Department of Mathematics, Imperial College, London, UK

30. Boritz, J. E. (1985). The effect of information presentation structures on audit planning and review judgments*. Contemporary Accounting Research, 1(2), 193-218.

31. Borthick, A. F., and O. D. West. (1987). Expert Systems--A New Tool for the Professional. Accounting Horizons 1 (1):9-16.

32. Bramer, M.A. (1982). A Survey and Critical Review of Expert Systems Research. In Introductory Readings in Expert Systems, ed. D. Mitchie, 3-29. London: Gordan and Breach

33. Braun, R. L., & Davis, H. E. (2003). Computer-assisted audit tools and techniques: Analysis and perspectives. Managerial Auditing Journal, 18(9), 725-731.

34. Brause, R., Langsdorf, T., & M. Hepp (1999). Neural Data Mining for Credit Card Fraud Detection. Working paper, J.W. Goethe-Universitat, Germany 1999.

35. Breiman, L. (2002). Comment, Statistical Fraud Detection: A Review. Statistcial Science, 2002, Volume 17, No. 3, 325-255

36. Bricker, R., & Nehmer, R. (1995). Information presentation format, degree of information processing and decision quality. Advances in Accounting Information Systems, 3, 3-29.

37. Brown-Liburd, H., Issa, H. & Lombardi, D.R. (2015) Behavioral Implications of Big Data's Impact on Audit Judgment and Decision Making and Future Research Directions. Accounting Horizons In-Press.

38. Brynjolfsson, E., J. Hammerbacher, & B. Stevens. (2011). Competing Through Data: Three Experts Offer Their Game Plans. McKinsey Quarterly (October): 36-47

39. Bughin, J., J. Livingston, & S. Marwaha. (2011). Seizing the Potential of 'Big Data'. McKinsey Quarterly (October): 103-109

40. Card, S., Mackinlay, J. & Shneiderman, B. (1998). Information visualization. Readings in Information Visualization, San Francisco: Morgan-Kaufmann, pp. 1-34.

41. Card, Stuart K., Jock D. Mackinlay, & Ben Shneiderman (1999), Readings in Information Visualization: Using Vision to Think. San Francisco: Morgan Kaufmann.

42. Cardinaels, E. (2008). The interplay between cost accounting knowledge and presentation formats in cost-based decision-making. Accounting, Organizations and Society, 33(6), 582-602.

43. Carey, B. (2015). Learning to See Data. The New York Times. Http://www.nytimes.com/2015/03/29/sunday-review/learning-to-see-data.html?_r=0

44. Casey, C. J. (1980). Variation in accounting information load: The effect on loan officers' predictions of bankruptcy. The Accounting Review, 55(1), 36-49.

45. Carey, J. M., & White, E. M. (1991). The effects of graphical versus numerical response on the accuracy of graph-based forecasts. Journal of Management, 17(1), 77-96.

46. Center for Medicare and Medicaid services (CMS), (2012), http://www.cms.gov

47. Chan, P.K., Fan, W., Prodromidis, A.L., & Stolfo, S.J. (1999). Distributed Data Mining in Credit Card Fraud Detection. IEEE Intelligent Systems, 14(6) 67-74

48. Chan, S. Y. (2001). The use of graphs as decision aids in relation to information overload and managerial decision quality. Journal of information science, 27(6), 417-425.

49. Chandola, V., Sukumar, S. R., & Schryver, J. C. (2013). Knowledge discovery from massive healthcare claims data. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1312-1320). ACM.

50. Chervany, N.L. & Dickson, G.W. (1974). Experimental evaluation of information overload in a production environment, Management Science, 20(10), pp 1335-1344.

51. Chorba, R. W. & New, J. L. (1980). Information Support for Decision-Maker Learning in a Competitive Environment: An Experimental Study, Decision Sci., 11, 4 (1980), 603-615.

52. Chowdhary, P., Palpanas, T., Pinel, F., Chen, S. K., & Wu, F. Y. (2006). Model-driven dashboards for business performance reporting. In Enterprise Distributed Object Computing Conference, 2006. EDOC'06. 10th IEEE International (pp. 374-386). IEEE.

53. Cleveland, W. S., & McGill, R. (1986). An experiment in graphical perception. International Journal of Man-Machine Studies, 25(5), 491-500

54. Cloyd, B.C. (1995). Prior knowledge, information search behaviour and performance in tax research tasks, The Journal of American Taxation Association, 17, Supplement, pp 82-107.

55. Coderre, G.D. (2009). Computer Aided Fraud Prevention and Detection. John Wiley & Sons, Inc. 2009, Hoboken, New Jersey.

56. Connell, N. A. D. (1987). Expert systems in accountancy: A review of some recent applications. Accounting and Business Research, 17(67), 221-233

57. Cuzzocrea, A., & Mansmann, S. (2009). Olap visualization: models, issues, and techniques. Encyclopedia of Data Warehousing and Mining,, 1439-1446.

58. Daly, J. L. & Buehner, M. (2003). P-Card Utilization in Municipal Government: Advantages and concerns. Journal of Public Procurement, Vol. 3, Issue 1, 75-94.

59. Daly, K. L., & King, K. M. (2011). Improper Payments: Reported Medicare Estimates and Key Remediation Strategies, United States Government Accountability Office, GAO-11-842T, July, 2011.

60. Davis, L. R. (1989). Report format and the decision maker's task: An experimental investigation. Accounting, Organizations and Society, 14(5), 495-508.

61. Debreceny, R., Gray, G. L., Tham, W., Goh, K., & Tang, P. (2003). The Development of Embedded Audit Modules to Support Continuous Monitoring in the Electronic Commerce Environment, 185(February), 169–185

62. DeFanti, T. A., Brown, M. D., & McCormick, B. H. (1989). Visualization: expanding scientific and engineering research opportunities. Computer, 22(8), 12-25.

63. Denning, P.J. (1986). Towards a Science of Expert Systems. IEEE Expert 1(2): 80-83.

64. Dennis, A. R., & Carte, T. A. (1998). Using geographical information systems for decision making: Extending cognitive fit theory to map-based presentations. Information Systems Research, 9(2), 194-203.

65. DeSanctis, G. & Jarvenpaa, S.L. (1989). Graphical presentation of accounting data for financial forecasting: An experimental investigation, Accounting, Organisations and Society, 14(6), pp 509-525.

66. Dickson, G. W., DeSanctis, G., & McBride, D. J. (1986). Understanding the effectiveness of computer graphics for decision support: a cumulative experimental approach. Communications of the ACM, 29(1), 40-47.

67. Dilla, W. N., Janvrin, D. J., & Jeffrey, C. (2012). The impact of graphical displays of pro forma earnings information on professional and nonprofessional investors' earnings judgments. Behavioral Research in Accounting, 25(1), 37-60

68. Duda, R. O., & Shortliffe, E. H. (1983). Expert Systems Research. Science 220:261-276.

69. Dull, R. B., & Tegarden, D. P. (1999). A comparison of three visual representations of complex multidimensional accounting information. Journal of Information Systems, 13(2), 117-131.

70. Dull, R.B; Graham, A.W & Baldwin, A.A. (2003). Web-based Financial Statements: Hyperlinks to Footnotes and their Effect on Decisions, International Journal of Accounting Information Systems, 4, pp 185-203.

71. Dupin, C. (1826). figurative Map of popular education of France

72. Eining, M., & Dorr, P. (1991). The impact of expert system usage on experiential learning in an auditing setting. Journal of Information Systems, 5(1), 1-16.

73. Eining, M. M., Jones, D. R., & Loebbecke, J. K. (1997). Reliance on decision aids: An examination of auditors' assessment of management fraud. Auditing: A Journal of Practice & Theory, 16(2)

74. Elmer, P. J., and D. M. Borowski. (1988). An Expert System Approach to Financial Analysis: The Case of S&L Bankruptcy. FM: The Journal of the Financial Management Association 17 (3):66-76.

75. Elliott, W. B., Hodge, F. D., & Sedor, L. M. (2011). Using online video to announce a restatement: Influences on investment decisions and the mediating role of trust. The Accounting Review, 87(2), 513-535.

76. Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. The information society, 20(5), 325-344.

77. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, Vol. 39, No. 11, pp. 27-34.

78. Fedorowicz, J., Oz, E., & Berger, P. D. (1992). A Learning Curve Analysis of Expert System Use*. Decision Sciences, 23(4), 797-818.

79. Feigenbaum, E. A. (1961). The simulation of verbal learning behavior. InPapers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference (pp. 121-132). ACM.

80. Fisher, D. (2010). Animation for visualization: opportunities and drawbacks. Ch,19, 329-352

81. Flegel, U., Vayssiere, J., & Bitz, G. (2010) A State of the Art Survey of Fraud Detection Technology. Insider Treats to Cyber Security, C.W. Probst et al.

82. Friendly, M. (2005). Milestones in the history of data visualization: A case study in statistical historiography. In Classification—the Ubiquitous Challenge (pp. 34-52). Springer Berlin Heidelberg.

83. Friendly, M. (2008). The golden age of statistical graphics. Statistical Science, 502-535.

84. Friendly, M. & Denis, D. (2000). The roots and branches of statistical graphics. Journal de la Soci´et´e Franc¸aise de Statistique, 141(4), 51–60.

85. Frownfelter-Lohrke, C. (1998). The effects of differing information presentations of general purpose financial statements on users' decisions. Journal of Information Systems, 12(2), 99-107.

86. Gahegan, M., Wachowicz, M., Harrower, M., & Rhyne, T. M. (2001). The integration of geographic visualization with knowledge discovery in databases and geocomputation. Cartography and Geographic Information Science, 28(1), 29-44.

87. Gal, G., "Using Auditor Knowledge to Formulate Data Model Constraints: Expert Systems for Internal Control Evaluation", Ph.D. dissertation, Michigan State University, 1985.

88. Gal, G., & P. Steinbart. (1987). Artificial Intelligence and Research in Accounting Information Systems: Opportunities and Issues. Journal of Information Systems 2(1):54.

89. Garrison, R.D., Wiesel, J.A., & Hart, A.J. (1997). Cut and Deal New Savings with Procurement Cards. Management Accounting, 79, 47-51

90. Gershon, N., & Page, W. (2001). What storytelling can do for information visualization. Communications of the ACM, 44(8), 31-37.

91. Ghani, E., Laswad, F., Tooley, S., & Jusoff, K. (2009). The role of presentation format on decision-makers' behaviour in accounting. International Business Research, 2(1), 183-195

92. Gillett, J.W., Fink, R.L. & Johnson, L.T. (1997). Accounting Control in a Procurement Environment. Management Accounting, 79, 47-51

93. Goswami, S., Chan, H. C., & Kim, H. W. (2008). The role of visualization tools in spreadsheet error correction from a cognitive fit perspective. Journal of the Association for Information Systems, 9(6), 16.

94. Graham, L. E., Damens, J. & Van Ness, G., (1991). Developing Risk Advisor: an Expert System for Risk Identification, Auditing: a Journal of Practice 6 Theory, pp. 69

95. Green, M. (1998), Toward a Perceptual Science of Multidimensional Data Visualization: Bertin and Beyond, ERGO/ GERO Human Factors Science.

96. Hard, N.J. & Vanecek, M.T. (1991). The implications of tasks and format on the use of financial information, Journal of Information Systems (Fall), pp 35-49.

97. Hayes-Roth, F. (1984). Knowledge-based expert systems—the state of the art in the US. The Knowledge Engineering Review, 1(02), 18-27.

98. Herbig, P. A., & Kramer, H. (1994). The effect of information overload on the innovation choice process: Innovation overload. Journal of Consumer Marketing, 11(2), 45-54.

99. Hodge, F.D. (2001). Hyperlinking un-audited information to audited financial statements: Effects on investor judgments, The Accounting Review 76(10), pp 675-691.

100. Hodge, F.D. & Pronk, M. (2006). The impact of professionalize and investment familiarity on investors' use of online financial reporting information, Journal of Accounting, Auditing and Finance, pp 267-292.

101. Hodge, F.D.; Kennedy, J.J & Maines, L.A. (2004). Does search facilitating technology improve the transparency of financial reporting?, The Accounting Review 79 (3), pp 687-703.

102. Hoffman, R. (1985). The Problem of Extracting the Knowledge of Experts from the Perspective of Experimental Psychology. AI Magazine Volume 8 Number 2 (1987)

103. Hoffman, R., Shadbolt, N., Burton, M.A., & Klein, G (1995). Eliciting Knowledge from Experts: A Methodological Analysis. Organizational Behaviorand Human Decision Processes, Volume 62, No. 2, May, pp. 129-158, 1995.

104. Holstrum, G. L., Mock, T. J., & West, R. N. (1986). The impact of technological events and trends on audit evidence in the year 2000: Phase I, Auditing Symposium VIII. In Proceedings of the 1986 Touche Ross--University of Kansas Symposium on Auditing Problems, Lawrence, KS (pp. 125-146).

105. Häubl, G., & Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. Marketing science, 19(1), 4-21.

106. Hu, S. D. (1987). Expert Systems for Software Engineers and Managers (Chapman and Hall: New York).

107. Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. Decision Support Systems, 50(3), 585-594

108. IDEA, (2013). IDEA Version Nine Workbook. CaseWare IDEA workbook, 2013.

109. Inselberg, A. (1985). The plane with parallel coordinates. The Visual Computer, 1(2), 69-91.

110. Iselin, E.R. (1988). The effects of information load and information diversity on decision quality in a structured decision task, Accounting, Organizations and Society 13(2), pp 147-164.

111. Issa, H. (2013). Exceptional exceptions (Doctoral dissertation, Rutgers University-Graduate School-Newark).

112. Jacoby, J. (1984). Perspectives on information overload. Journal of consumer research, 432-435.

113. Jacomy, M., Heymann, S., Venturini, T., & Bastian, M. (2011). Forceatlas2, a continuous graph layout algorithm for handy network visualization. Medialab center of research, 560

114. Kaplan, S.E. (1988). An Examination of the Effect of Presentation Format on Auditor's Expected Value Judgments. Accounting Horizons. (September): 90-95.

115. Keller, P. R., & Keller, M. M. (1993). Visual cues: practical data visualization (Vol. 2). Los Alamitos, CA: IEEE Computer Society Press.

116. Keim, D. A. (2001). Visual exploration of large data sets. Communications of the ACM, 44(8), 38-44.

117. Keim, D. A. (2002). Information visualization and visual data mining. Visualization and Computer Graphics, IEEE Transactions on, 8(1), 1-8.

118. Keim, D. A & Kriegel, H. P. (1994). VisDB: Database exploration using multidimensional visualization. Computer Graphics and Applications, September 1994, pp. 44-49.

119. Keim, D. A., & Kriegel, H. P. (1996). Visualization techniques for mining large databases: A comparison. Knowledge and Data Engineering, IEEE Transactions on, 8(6), 923-938.

120. King, K. M. (2010). Medicare Fraud, Waste, and Abuse: Challenges and Strategies for Preventing Improper Payments, United States Government Accountability Office, GAO-10-844T, June, 2010.

121. Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. Psychological Bulletin; Psychological Bulletin, 107(3), 296–310

122. Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. Journal of Healthcare Information Management—Vol, 19(2), 65.

123. Kohavi, R., Mason, L., Parekh, R., & Zheng, Z. (2004) Lessons and Challenges from Mining E-Commerce Data. Machine Learning, 57 (1-2)(2004), 83-113.

124. Kogan, A., Alles, M.G., Vasarhelyi, M.A. & Wu, J (2011). Analytical Procedures for Continuous Data Level Auditing: Continuity Equations. Working paper, Rutgers Accounting Research Center.

125. Kosslyn, S. M., & Cunningham, D. H. (1994). Elements of graph design. New York: WH Freeman.

126. Kraemer, Eileen T. & Thomas E. Ferrin (1998), "Molecules to Maps: Tools for Visualization and Interaction in Support of Computational Biology," Bioinformatics Review, 14 (2), 764–71.

127. Kumar, N., & Benbasat, I. (2004). The effect of relationship encoding, task type, and complexity on information representation: An empirical evaluation of 2D and 3D line graphs. Mis Quarterly, 255-281.

128. Lee, H. Y. & Ong, H. L. (1996). Visualization support for data mining. IEEE Expert Intelligent Systems and their Applications, Vol. 11, No. 5, pp. 69-75

129. Leonard-Barton, D., & Sviokla, J. J. (1988). Putting expert systems to work.Harvard Business Review, 66(2), 91-98.

130. Leuba, G., & Kraftsik, R. (1994). Changes in volume, surface estimate, three-dimensional shape and total number of neurons of the human primary visual cortex from midgestation until old age. Anatomy and embryology, 190(4), 351-366.

131. Levy, I., Hasson, U., & Malach, R. (2004). One picture is worth at least a million neurons. Current Biology, 14(11), 996-1001.

132. Libby, R. (1981). Accounting and human information processing theory: Theory and applications, Englewood Cliffs, NJ: Prentice Hall.

133. Libby, R. & Lewis, B.L (1977). Human information processing research in accounting: The start of the art, Accounting, Organizations and Society 2(3), pp 245-268.

134. Libby, R. & Lewis, B.L. (1982). Human information processing research in accounting: The start of the art in 1982, Accounting, Organizations and Society 7(3), pp 231-285.

135. Libby, R.; Bloomfield, R & Nelson, M.W. (2002). Experimental research in financial accounting, Accounting, Organizations and Society (27), pp 775-810.

136. Liberatore, M. J., Titus, G. J., & Dixon, P. W. (1988). The effects of display formats on information systems design. Journal of Management Information Systems, 85-99.

137. Liu, Q. (2013). The Application of exploratory Data Analysis in Auditing, Dissertation Proposal, Rutgers Business School, Continuous Audit and Reporting Laboratory.

138. Lohse, Gerald L. (1997), The Role of Working Memory on Graphical Information Processing, Behaviour and Information Technology, 16 (6), 297–308.

139. Lombardi, D. R. (2012). Using an expert system to debias auditor judgment: An experimental study, PhD Dissertation, Rutgers Business School

140. Lucas Jr, H. C. (1981). An experimental investigation of the use of computer-based graphics in decision making. Management Science, 27(7), 757-768.

141. Lurie, N. H., & Mason, C. H. (2007). Visual representation: Implications for decision making. Journal of Marketing, 71(1), 160-177.

142. MacDougall, E. B. (1992). Exploratory analysis, dynamic statistical visualization and geographic information systems. Cartography and Geographical Information Systems, Vol. 19, No. 4, pp. 237-246

143. Malhotra, Naresh K. (1982). Information Load and Consumer Decision Making, Journal of Consumer Research, 8 (March), 419-430.

144. Manson, S., McCartney, S., & Sherer, M. (2001). Audit automation as control within audit firms. Accounting, Auditing & Accountability Journal, 14(1), 109-130.

145. Mauldin, E. G. (2003). An Experimental Examination of Information Technology and Compensation Structure Complementarities in an Expert System Context. Journal of Information Systems 17 (1):19-41

146. McCandless, D. (2012). Information is beautiful. Collins.

147. McCarty, L. T. (1977). Reflections on" Taxman": An Experiment in Artificial Intelligence and Legal Reasoning. Harvard Law Review, 837-893.

148. Meservy, R. D., Bailey, A. D., & Johnson, P. E. (1986). Internal control evaluation: A computational model of the review process. Division of Research, College of Administrative Science, Ohio State University.

149. Miller A, Cioffi J. (2004). Measuring marketing effectiveness and value: the unisys marketing dashboard. J Advert Res 44:237–43

150. Mintzer, O., & Messmore, J. (1984). Terrain Analysis Procedural Guide for Surface Configuration, Technical Report, ETL-0352, Engineer Topographical Laboratories.

151. Moffitt, K. C., & Vasarhelyi, M. A. (2013). AIS in an Age of Big Data. Journal of Information Systems, 27(2), 1-19.

152. Montgomery, S. B., Astakhova, T., Bilenky, M., Birney, E., Fu, T., Hassel, M., & Jones, S. J. (2004). Sockeye: a 3D environment for comparative genomics. Genome Research, 14(5), 956-962.

153. Moriarity, S. (1979). Communicating financial information through multidimensional graphics. Journal of Accounting Research, 205-224.

154. Mueller, J.P. (2011) Mastering IDEAScript, The Definitive Guide. 2011, John Wiley & Sons, Hoboken New Jersey.

155. Murphy, E.E. (1998). Cards are a Key Part of Reengineering. Purchasing, 125(6), 75-76.

156. National Association of Purchasing Card Professionals (NAPCP), (2004). Fraud Prevention and Detection: Establishing and Maintaining a Purchasing Card Program with Adequate Management Controls to Prevent Fraud, Misuse, and Abuse. NAPCP Best Practices Task Force, January 2004.

157. Negash S, Gray P. (2008). Business intelligence, international handbook on information systems. Handbook on decision support systems. Berlin Heidelberg: Springer. p. 175—193.

158. Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical review E, 69(2), 026113

159. Nibbelin, M. C, C. D. Bailey, & R. W. Zmud. (1992). The effects of mode of information presentation and cognitive style on bond rating change decisions. Advances in Accounting 10: 159-174

160. O'Donnell, E., & J. S. David. (2000). How information systems influence user decisions: A research framework and literature review. International Journal of Accounting Information Systems 1: 178–203.

161. O'Keefe, R. M., O'Leary, D., Rebne, D., & Chung, Q. B. (1993). The impact of expert systems in accounting: System characteristics, productivity and work unit effects. Intelligent Systems in Accounting, Finance and Management, 2(3), 177-189.

162. O'Leary, D. E., & Watkins, P. R. (1989). Review of expert systems in auditing.EXPERT SYST. REV., 2(1), 3-22

163. Palmer,R. & Gupta, M. (2003). Purchasing Card Benchmark Survey. RPGM Research 2003, (http://www.napcp.org/palmergupta2003)

164. Paul, H., & Davis, K. (1985). Expert systems: artificial intelligence in business.John Willey.

165. Pauwels K, Ambler T, Bruce HC, LaPointe P, Reibstein D, Skiera B, et al.,(2009). Dashboards as a service: why, what, how, and what research is needed? J Serv Res 12:175–89

166. Phua, C., Lee, V., Smith, K., & Gaylor, R. (2010) A Comprehensive Survey of Data Mining-based Fraud Detection Research. Cornell University Library, Sep 30, 2010

167. Playfair, W. (1786). Commercial and political atlas: Representing, by copper-plate charts, the progress of the commerce, revenues, expenditure, and debts of England, during the whole of the eighteenth century. Corry, London. Republished in The Commercial and Political Atlas and Statistical Breviary.

168. Playfair, W. (1801). The commercial and political atlas: representing, by means of stained copper-plate charts, the progress of the commerce, revenues, expenditure and debts of England during the whole of the eighteenth century. T. Burton.

169. Playfair, W. (1821). A Letter on Our Agricultural Distresses, Their Causes and Remedies: Accompanied with Tables and Copper-plate Charts, Shewing and

Comparing the Prices of Wheat, Bread, and Labour, from 1565 to 1821. Addressed to the Lords and Commons. W. Sams.

170. Playfair, W. (2005). Playfair's commercial and political atlas and statistical breviary. Cambridge University Press.

171. Prasad, K., & Ramakrishna, S. (2010). Text Analytics to Data Warehousing. International Journal on Computer Science & Engineering, 2(6), 2201-2207.

172. Quah, J.T., & Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. Expert Systems with Applications 35 (2008) 1721-1732

173. Ramarapu, N. K., Frolick, M. N., Wilkes, R. B., & Wetherbe, J. C. (1997). The Emergence of Hypertext and Problem Solving: An Experimental Investigation of Accessing and Using Information from Linear versus Nonlinear Systems*. Decision Sciences, 28(4), 825-849.

174. Remus, W. (1984). An empirical investigation of the impact of graphical and tabular data presentations on decision making. Management Science, 30(5), 533-542.

175. Remus, W. (1987). A study of graphical and tabular displays and their interaction with environmental complexity. Management Science, 33(9), 1200-1204.

176. Reuter, L. H., Tukey, P., Maloney, L. T., Pani, J. R., & Smith, S. (1990). Human perception and visualization. In Proceedings of the 1st conference on Visualization'90 (pp. 401-406). IEEE Computer Society Press.

177. Ribarsky, W., Katz, J. & Holland, A. (1999). Discovery visualization using fast clustering. IEEE Computer Graphics and Applications, September/October 1999, pp. 32-39.

178. Roberts, F.D. (2002). The effects of decision aid recommendation on users' cognitive processes, memory and judgments, Unpublished thesis, University of Tennessee

179. Roscoe, P., & Howorth, C. (2009). Identification through technical analysis: A study of charting and UK non-professional investors. Accounting, Organizations and Society, 34(2), 206-221.

180. Rose, J. M. (2002). Behavioral Decision Aid Research: Decision Aid Use and Effects. In: Arnold V, Sutton SG, editors. Researching Accounting as an

Information Systems Discipline. Sarasota, FL: American Accounting Association: 111-33

181. Sagiroglu, S., & Sinanc, D. (2013). Big Data: A review. In Collaboration Technologies and Systems (CTS), 2013 International Conference on (pp. 42-47). IEEE.

182. Sangster, A. (1995). The Bank of Scotland's lending adviser expert system, COMPASS. In Artificial Intelligence for Applications, 1995. Proceedings., 11th Conference on (pp. 24-30). IEEE.

183. Schick, A.G., Gordon, L.A & Haka, S. (1990). “Information overload: A temporal approach”, Accounting, Organisations and Society, 15 (3), pp 199-220.

184. Schroder, H. M., Driver, M. J., & Streufert, S. (1967). Human information processing: Individuals and groups functioning in complex social situations. New York: Holt, Rinehart and Winston.

185. Schulte MF. (2006). Business objects dashboard manager. DM Rev 16(2):49

186. Schulz, A. & P. Booth. (1995). The Effects of Presentation Format on the Effectiveness and Efficiency of Auditors’ Analytical Review Judgments. Accounting and Finance. (May): 107-131.

187. Silveira, P., Rodríguez, C., Casati, F., Daniel, F., D’Andrea, V., Worledge, C., & Taheri, Z. (2010). On the design of compliance governance dashboards for effective compliance and audit management. In Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops (pp. 208-217). Springer Berlin Heidelberg.

188. Sinha, U., Bui, A., Taira, R., Dionisio, J., Morioka, C., Johnson, D., & Kangarloo, H. (2002). A review of medical imaging informatics. Annals of the New York Academy of Sciences, 980(1), 168-197.

189. Sloman, S. A. (1996). The empirical case for two systems of reasoning. Psychological bulletin, 119(1), 3.

190. Sparrow, M. K. (2008). Fraud in the U.S. Healthcare System: Exposing the Vulnerabilities of Automated Payment Systems. Social Research, 75(4), 1151-1180.

191. Spence, I. (2006). William Playfair and the psychology of graphs. 2006 JSM proceedings, American Statistical Association, Alexandria, 2426-2436.

192. Steele, J., & Iliinsky, N. (2010). Beautiful visualization: looking at data through the eyes of experts. " O'Reilly Media, Inc.".

193. Steele, J., & Iliinsky, N. (2011). Designing Data Visualizations: Representing Informational Relationships. " O'Reilly Media, Inc.".

194. Steinbart, P. J. (1987). The construction of a rule-based expert system as a method for studying materiality judgments. Accounting Review, 97-116.

195. Stock, D., & C. J. Watson. (1984). Human judgment accuracy, multidimensional graphics, and humans vs. models. Journal of Accounting Research (Spring): 192-206.

196. Stocks, M.H. & Tuttle, B. (1998). An examination of information presentation effects on financial distress predictions, Advances in Accounting Information Systems 6, pp 107-128.

197. Stone, N.D. & Schkade, D.A. (1991). Numeric and linguistic information representation in multi-attribute choice, Organizations Behaviour and Human Decision Processes 49, pp 42-59.

198. Strong, J. M., & Portz, K. (2011). A Further Investigation Of Tables Versus Graphs For Decision-Making: Does Accounting Knowledge Make A Difference?. Review of Business Information Systems (RBIS), 7(2), 39-48.

199. Tegarden, D.P. (1999). Business Information Visualization. Communications of the AIS, 1(1), Article 4.

200. Tomas, A. S. (1998). Expert Systems Applications In Accounting. Artificial Intelligence in Accounting and Auditing: Towards New Paradigms, Volume 4, 11.

201. Tractinsky, N., & Meyer, J. (1999). Chartjunk or goldgraph? Effects of presentation objectives and content desirability on information presentation. MIS Quarterly, 397-420.

202. Trelease, R. B. (2002). Anatomical informatics: Millennial perspectives on a newer frontier. The Anatomical Record, 269(5), 224-235.

203. Trewin, J. (1996). The impact of the introduction of an expert system on a public accounting organization. Intelligent Systems in Accounting, Finance and Management, 5(3), 185-197.

204. Tufte, E. R., & Graves-Morris, P. R. (1983). The visual display of quantitative information (Vol. 2). Cheshire, CT: Graphics press.

205. Tufte, E. R., & Robins, D. (1997). Visual explanations (Vol. 25). Cheshire, CT: Graphics Press.

206. Tukey, J. W. (1962). The future of data analysis. The Annals of Mathematical Statistics, 1-67.

207. Tukey J. W. (1977). Exploratory data analysis. Reading, Addison-Wesley

208. Tukey, P., & Tukey, J. (1988). Graphic display of data sets in 3 or more dimensions. The collected works of John Tukey, 5, 189-288.

209. Tuttle, B., & Kershaw, R. (1998). Information presentation and judgment strategy from a cognitive fit perspective. Journal of Information Systems, 12(1), 1-17.

210. U.S Census. 2010. www.census.gov

211. Vasarhelyi, M. A. & Halper. F. B. (1991). The continuous audit of online systems. Auditing: A Journal of Practice and Theory. 1991. 10 (1): 110-125

212. Vasarhelyi, M. A. (1998). Toward an intelligent audit: Online reporting, online audit, and other assurance services. Advances in Accounting Information Systems, 6, 207-222

213. Vasarhelyi, M. A., Kogan, A. & Tuttle, B. (2015) Big Data in Accounting: An Overview. Accounting Horizons In-Press

214. Vekiri, I. (2002). What is the value of graphical displays in learning? Educational Psychology Review, 14(3), 261-312

215. Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graph vs. tables literature. Decision Sciences 22: 219-240.

216. Vessey, I. & Galletta, D (1991). Cognitive fit: An empirical study of information acquisition, Information Systems Research, 2(1), pp 63-84.

217. VMG Health (2011). Intellimarker. Ambulatory Surgical Centers Financial & Operational Benchmarking Study. Sixth Edition, November 2011

218. Waller, R. (1981). Understanding network diagrams. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, April 1981.

219. Ware, C. (2012). Information visualization: perception for design. Elsevier.

220. Warren, J. D. Jr., Moffitt, K .C., & Byrnes, P. (2015) How Big Data Will Change Accounting. Accounting Horizons In-Press

221. Waterman, D. A. (1985). A guide to expert systems.

222. Wegman, E. J. (1990). Hyperdimensional data analysis using parallel coordinates. Journal of the American Statistical Association, 85(411), 664-675.

223. Weiss, S. & Kulikowski, C. (1984). A Practical Guide to Designing Expert Systems. Totowa, N.J.: Rowman and Allanheld.

224. West, T. G. (1995). Forward into the past: A revival of old visual talents with computer visualization. ACM SIGGRAPH Computer Graphics, 29(4), 14-19.

225. Wiedemann, L.A. (2014). Leveraging Data and Analytics to Detect Fraud and Abuse in Healthcare. Journal of AHIMA

226. Winograd, B. N., Gerson, J. S., & Berlin, B. L. (2000). Audit practices of PricewaterhouseCoopers. Auditing: A Journal of Practice & Theory, 19(2), 176-182.

227. Wright, W. F. (1995). Superior loan collectability judgments given graphical displays. Auditing: A Journal of Practice & Theory (Fall): 144-154.

228. Yang, D. C., & Vasarhelyi, M. A. (1995). The Application of Expert Systems In Accounting. Unpublished working paper.

229. Yau, N. (2013). Data Points: Visualization That Means Something: Visualization That Means Something. John Wiley & Sons.

230. Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. International Journal of Accounting Information Systems, 13(1), 41-59.

231. Yoon, K., Hoogduin, L.A. & Zhang, L. (2015) Big Data as Complementary Audit Evidence. Accounting Horizons In-Press.

232. Zhang, P. & Andrew B. W. (1995), Business Information Visualization for Decision-Making Support: A Research Strategy, in Proceedings of the First Americas Conference on Information Systems, Manju K. Ahuja, Dennis F. Galletta, and Hugh J. Watson, eds. Pittsburgh: Association for Information Systems.

233. Zhang, J., Yang, X., & Appelbaum, D. (2015). Toward Effective Big Data Analysis in Continuous Auditing. Accounting Horizons In-Press.

APPENDIX A

	Authors	Date	Paper Title	Journal Name	Major Finding	Y/N/V*
1	Altman, Edward	1983	*Multidimensional Graphics and Bankruptcy Prediction: A Comment*	Journal of Accounting Research	Found that subjects made more accurate predictions with the direct use of financial ratios or other financial information than with multidimensional graphics	N
2	Amer	1991	*An experimental investigation of multi-cue financial information display and decision making*	Journal of Information Systems	Found that tabular presentation resulted in more accurate assessments of debt covenant violations in a selective task but had no effect on predictions of bond ratings during an integrative task	V
3	Anderson and Kaplan	1992	*The Effects of Output Interference on Analytical Procedures Judgments*	Auditing: A journal of practice and theory	Found that graphs were less effective for non-investigation tasks but effective for investigative tasks	V
4	Anderson and Mueller	2011	*The effects of experience and data presentation format on an auditing judgment. Journal of Applied Business Research*	Journal of applied business research	While both students and audit practitioners were aided with graphs in assessing correlations, students were aided relatively more than were practitioners	Y

5	Anderson and Reckers	1992	*An empirical investigation of the effects of presentation format and personality on auditor's judgment in applying analytical procedures*	Advances in Accounting	Graphs outperformed tables for tasks that require comparing data and understanding relationships	Y
6	Beattie and Jones	1993	*Effect of graphical presentations on insights into a company's financial position-an innovative educational approach to communicating financial information in financial reporting*	Accounting Education	The experimental studies suggested that graphs were more effective than numerical presentation	Y
7	Benbasat and Dexter	1986	*An experimental evaluation of graphical and color-enhanced information presentation*	Management Science	Found that tables generally were better than graphs for simple tasks that require precise estimates or specific data values	N
8	Blocher et al	1986	*Report format and task complexity: interaction in task judgments. Accounting*	Accounting, Organizations and Society	Found that decision making is improved when participants use tabular formats on high-information load tasks.	N
9	Boritz, Jefim	1985	*The effect of information presentation structures on audit planning and review judgments*	Contemporary Accounting Research	The use of graphical presentation format on internal control information appeared to have	N

					created more difficulty for users	
10	Cardinaels, Eddy	2008	*The interplay between cost accounting knowledge and presentation formats in cost-based decision-making*	Accounting, Organizations and Society	Found that decision makers with a low level of cost accounting knowledge attain higher profits when they use a graphical format in comparison to a tabular format.	Y
11	Carey and White	1991	*The effects of graphical versus numerical response on the accuracy of graph-based forecasts*	Journal of Management	Found that graph-based forecasts were significantly more accurate when reported graphically than when made numerically	Y
12	Chan, Siu	2001	*The use of graphs as decision aids in relation to information overload and managerial decision quality*	Journal of information Science	Found that type of presentation format (graphs or tables) alone did not have any significant effect on decision quality.	V
13	Davis	1989	*Report format and the decision maker's task: An experimental investigation.*	Accounting, Organizations and Society	Results show that tabular format were superior as the task became more complex	N
14	Denniz and Carte	1998	*Using Geographical Information Systems for Decision Making: Extending Cognitive Fit Theory to Map-Based Presentations*	Information Systems Research	The experiment found that using a map-based presentation helped decision makers make faster and more accurate decisions	Y

15	DeSanctis and Jarvenpaa	1989	*Graphical presentation of accounting data for financial forecasting: An experimental investigation*	Accounting, Organizations and Society	Found that the use of graphs improves decision accuracy only after participants have practice using the graphs.	V
16	Dickson et al.	1986	*Understanding the effectiveness of computer graphics for decision support: a cumulative experimental approach*	Communications of the ACM	The experiment Conducted found that tabular reports were rated as “easier to read and understand” than graphical reports.	N
17	Dull and Tegarden	1999	*A comparison of three visual representations of complex multidimensional accounting information*	Journal of Information Systems	Results suggest that decision accuracy improved with the use of multidimensional graphs	Y
18	Frownfelter-Lohrke	1998	*The effects of differing information presentations of general purpose financial statements on users' decisions*	Journal of Information Systems	The presentation format predicted to support each task did not significantly affect accuracy.	V
19	Goswami et al.	2008	*The Role of Visualization Tools in Spreadsheet Error Correction from a Cognitive Fit Perspective*	Journal of the association of information systems	Results indicate that subjects using a visualization tool were significantly faster in identifying and correcting link errors	Y
20	Hard and Vanecek	1991	*The implications of tasks and format on the use of financial information.*	Journal of Information Systems	Found that financial predictions were more accurate for graphical presentations during an estimation task but more	V

					accurate for tabular presentations during an accumulation task	
21	Huabl and Trifts	2000	*Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids*	Marketing science	Findings suggest that the interactive visual tool had strong favorable effects on both the quality and the efficiency of consumers' purchase	Y
22	Kaplan	1988	*An Examination of the Effect of Presentation Format on Auditors" Expected Value Judgments."*	Accounting Horizons	Graphs and tables appear to be equally effective for expected value judgments of auditors.	V
23	Kumar and Benbasat	2004	*The effect of relationship encoding, task type, and complexity on information representation: An empirical evaluation of 2D and 3D line graphs*	Management Information Systems Quarterly	Results suggest that 3-D graphs often outperform 2-D graphs for both elementary and advanced tasks	Y
24	Liberatore et al.	1989	*The effects of display formats on information systems design*	Journal of Management Information Systems	Found that tabular format was superior as the questions to be answered with decision cues became more complex	N
25	Lucas JR., Henry	1981	*An experimental investigation of the use of computer-based graphics in decision making*	Management Science	The results of the experiment found limited support for the use of graphics presentation in an information system	N

26	Moriarity	1979	*Communicating financial information through multidimensional graphics*	Journal of Accounting Research	Found that subjects made more accurate predictions with multidimensional graphics than with the direct use of financial ratios or other financial information	Y
27	Nibbelin et al.	1992	*The effects of mode of information presentation and cognitive style on bond rating change decisions*	Advances in Accounting	Found that presentation format in terms of faces used did not result in differences in decision accuracy	V
28	Remus, William	1984	*An empirical investigation of the impact of graphical and tabular data presentations on decision making.*	Management Science	Found that the tabular aids outperform the graphical aids when the erratic components of the decisions are reduced	N
29	Remus, William	1987	*A study of graphical and tabular displays and their interaction with environmental complexity.*	Management Science	Results suggest that tables are better for tasks that are low in environmental complexity and graphs are better for tasks with intermediate environmental complexity	V
30	Roscoe and Horwoth	2009	*Identification through technical analysis: A study of charting and UK non-professional investors*	Accounting, Organizations and Society	The results suggested that charting techniques may help decision techniques to organize data	Y
31	Schulz and Booth	1995	*The Effects of Presentation Format on the Effectiveness and*	Accounting and Finance	Found that graphs produced more accurate correlation	V

			Efficiency of Auditors" Analytical Review Judgments."		estimates and decreased time on task but did not influence decision confidence	
32	Strong and Portz	2011	*A Further Investigation Of Tables Versus Graphs For Decision-Making: Does Accounting Knowledge Make A Difference?*	Review of Business Information Systems	Found that individuals do realize the benefits of graphical information presentation and perform significantly better using graphs	Y
33	Tractinsky and Meyer	1999	*Chartjunk or goldgraph? Effects of presentation objectives and content desirability on information presentation*	Management Information Systems Quarterly	Result suggest that individuals prefer 2-D graphs over 3-D graphs for effective decision making	Y
34	Tuttle and Kershaw	1998	*Information presentation and judgment strategy from a cognitive fit perspective.*	Journal of Information Systems	Results found that graphs produced better performance evaluations under conditions that require holistic decision strategies but did not influence decision performance when analytical strategies were mandated	V
35	Vessey and Galletta	1991	*"Cognitive fit: An empirical study of information acquisition"*	Information Systems Research	Found that participants preferred to use tables rather than graphs to solve symbolic problems	N
36	Wright	1995	*Superior loan collectability*		Found that auditors who	Y

			judgments given graphical displays		used graphical presentations made more accurate loan collectability predictions	

* Support the Use of Data Viz, Y/N/V means: Y=yes, N=no, V=varying results

APPENDIX B

FIELD NAME	DATA SOURCE	FIELD DESCRIPTION
Primary Data Source		
CLAIM_NO	Medicare patient claims and provider details for the 2010 fiscal year	Claim number
DESY_SORT_KEY		Patient ID number
CLM_THRU_DATE		Claim through Date
PRVDR_NUM		Provider Number
STATE_BENE		Patient State
COUNTY_CD		Patient County
BENE_SEX		Patient Sex
BENE_RACE		Patient Race
CLM_PRNCPAL_DGNS_CD		Claim Principle Diagnosis
BENE_BIRTHDATE		Patient Birthday
CLM_PMT_AMT		Claim payment amount
NCH_PRMRY_PYR_CLM_PD_AMT		Primary payer payment amount
NCH_PRVDR_STATE_CD		Provider State
PTNT_DSCHRG_STUS_CD		Patient Discharge Status
CLM_TOT_CHRG_AMT		Claim total charged amount
IP_CLM_DGNS_CD_CNT		Claim diagnosis counts
IP_CLM_PRCDR_CD_CNT		Claim proceder counts
IP_CLM_RLT_COND_CD_CNT		Claim condition counts
CLM_ADMSN_DT		Claim admission date
CLM_SRC_IP_ADMSN_CD		Claim inpatient admission code
NCH_PTNT_STUS_IND_CD		Patient status indicator
CLM_UTLZTN_DAY_CNT		Claim utlization day count
CLM_NUTLZTN_DAY_CNT		Claim non-utlization dat count
NCH_BENE_DSCHRG_DT		Patient discharge date
CLAIM_DIAGNOSIS		Claim diagnosis
NPI_NUM		National provider number
NPI	National Provider Identifier (NPI) details	National provider number
Entity Type Code		Entity type code
Provider Organization Name		Provider Organization Name
Provider Last Name		Provider Last Name
Provider First Name		Provider First Name
Business Practice Address		Business Practice Address
Business Practice City		Business Practice City
Business Practice State		Business Practice State
Business Practice Country		Business Practice Country
Gender		Provider Gender

DIAGNOSIS CODE	ICD-9-CM Diagnosis and Procedure Codes	DIAGNOSIS CODE
SHORT DESCRIPTION		DIAGNOSIS SHORT DESCRIPTION
LONG DESCRIPTION		DIAGNOSIS LONG DESCRIPTION
Country	NJ Census 2010	Country
State		State
County		County
Census 2010		Census population for 2010
Secondary Data Source		
Hospital Name	American Hospital Directory	Hospital Name
City		City
State		State
Staffed Beds		Total number of Staffed Beds
Total Discharges		Total annual Discharges
Patient Days		Total Patient Days
Hospital Name	New Jersey Department of Health	Hospital Name
County		County
State		State

APPENDIX C

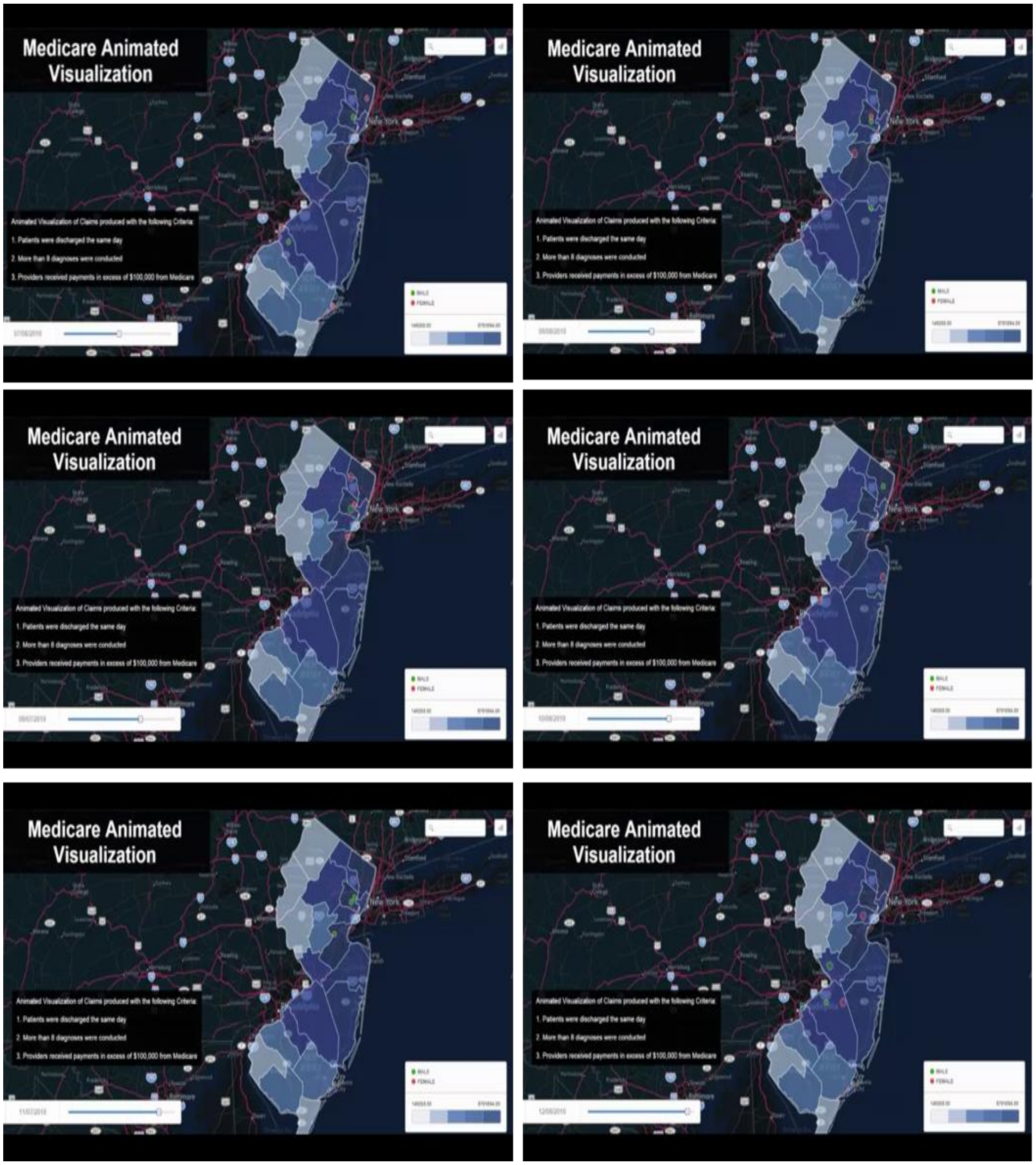
Claim Diagnosis Color Coding:

Claim Diagnosis

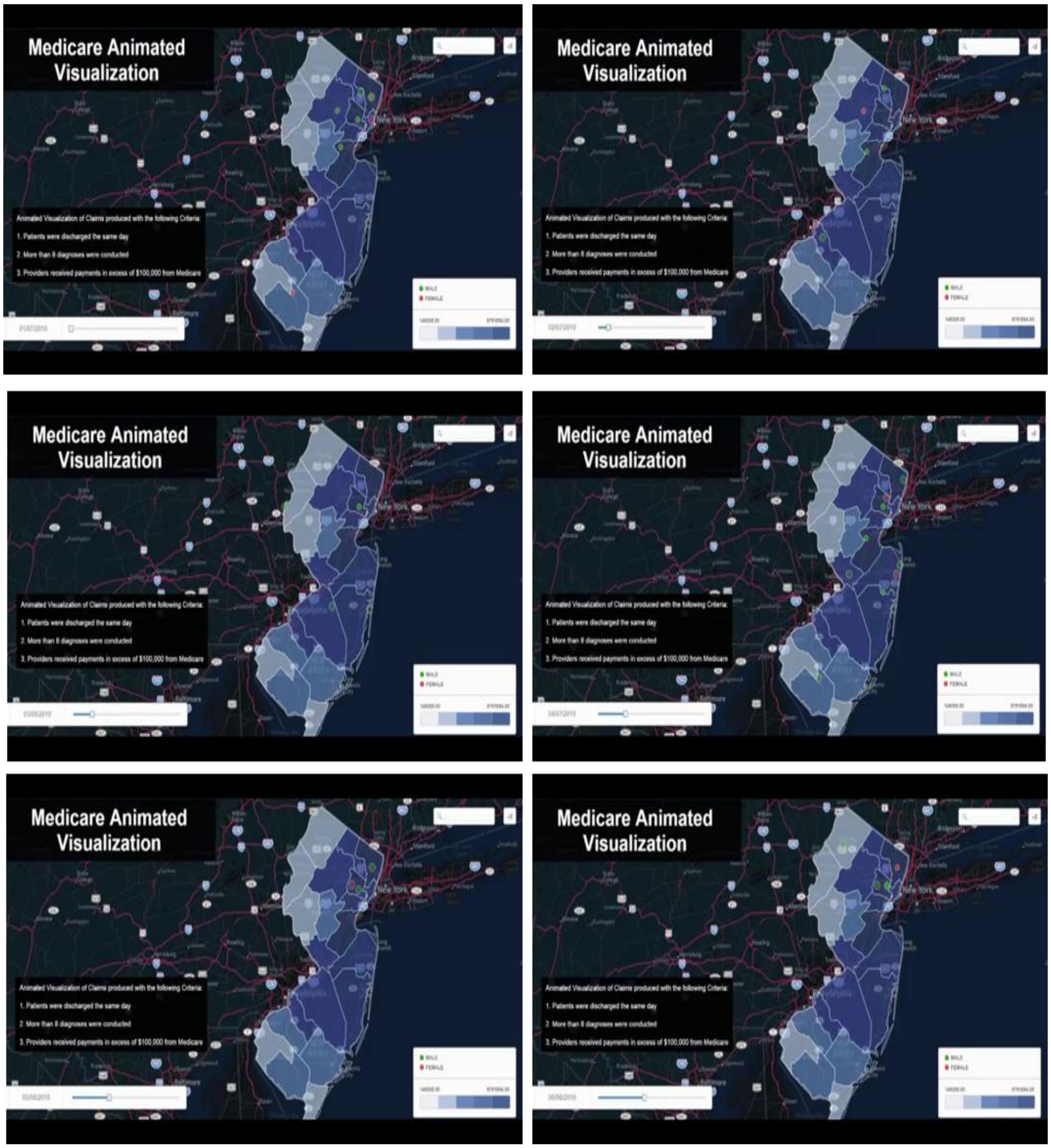
- ALCOHOL/DRUG USE & ALCOHOL/DRUG INDUCED ORGANIC MENTAL DISORDERS
- BURNS
- DISEASES & DISORDERS OF BLOOD, BLOOD FORMING ORGANS, IMMUNOLOG DISORD
- DISEASES & DISORDERS OF THE CIRCULATORY SYSTEM
- DISEASES & DISORDERS OF THE DIGESTIVE SYSTEM
- DISEASES & DISORDERS OF THE EAR, NOSE, MOUTH & THROAT
- DISEASES & DISORDERS OF THE EYE
- DISEASES & DISORDERS OF THE FEMALE REPRODUCTIVE SYSTEM
- DISEASES & DISORDERS OF THE HEPATOBILIARY SYSTEM & PANCREAS
- DISEASES & DISORDERS OF THE KIDNEY & URINARY TRACT
- DISEASES & DISORDERS OF THE MALE REPRODUCTIVE SYSTEM
- DISEASES & DISORDERS OF THE MUSCULOSKELETAL SYSTEM & CONN TISSUE
- DISEASES & DISORDERS OF THE NERVOUS SYSTEM
- DISEASES & DISORDERS OF THE RESPIRATORY SYSTEM
- DISEASES & DISORDERS OF THE SKIN, SUBCUTANEOUS TISSUE & BREAST
- ENDOCRINE, NUTRITIONAL & METABOLIC DISEASES & DISORDERS
- FACTORS INFLUENCING HLTH STAT & OTHR CONTACTS WITH HLTH SERVCS
- HUMAN IMMUNODEFICIENCY VIRUS INFECTIONS
- INFECTIOUS & PARASITIC DISEASES, SYSTEMIC OR UNSPECIFIED SITES
- INJURIES, POISONINGS & TOXIC EFFECTS OF DRUGS
- MENTAL DISEASES & DISORDERS
- MULTIPLE SIGNIFICANT TRAUMA
- MYELOPROLIFERATIVE DISEASES & DISORDERS, POORLY DIFFERENTIATED NEOPLASM
- PRE-MDC
- PREGNANCY, CHILDBIRTH & THE PUERPERIUM
- UNRELATED OPERATING ROOM PROCEDURES

APPENDIX D

Animated Data Visualization of Medicare Claims (January to June):



Animated Data Visualization of Medicare Claims (July to December):



APPENDIX E

FIELD NAME	FIELD DESCRIPTION
ACC.Account Name	Account holder name
ACC.Employee ID	Employee ID number
ACC.Master Accounting Code	Master Accounting Code
PUR.Purchase Date	Purchase Date
FIN.Posting Date	Posting Date
FIN.Original Currency Amount	Original Currency Amount
Acc.Single Transaction Dollar Limit	Single Transaction Dollar Limit
PUR.Item Description	Description of items purchased
MCH.Merchant Name	Merchant Name
MCH.MCC Description	Merchant category code Description
MCH.Merchant Category Code (MCC)	Merchant Category Code
FIN.Cardholder Transaction Type	Cardholder Transaction Type
PUR.Extended Item Amount	Extended Item Dollar Amount
PUR.Item Quantity	Item Quantity Purchased
PUR.Line Item Total Amount	Total Dollar Amount per line item
PUR.Product Code	Product Code
PUR.Line Item Total Sign	Line item sign – Debit or Credit
PUR.Customer Code	Customer Code
FIN.Update User ID	ID of user who made an update
FIN.Transaction Date	Transaction Date
PUR.Extended Item Sign	Extended Item Sign – Debit or Credit
ACC.Credit Limit	Credit Limit for Account holder
Acc.Acount Number	Account Number
Fin.Original Currency Code	Country’s Original Currency Code
Fin.Posted Currency Code	Posted Currency Code
FIN.Transaction Reference Number	Transaction Reference Number
Company	Company Name
Account Type	Account Type
Primary Location - Building/Office	Primary Location - Building/Office
Primary Location - Site/Country	Primary Location - Site/Country
Org Type	Type of the Organization
Org Name	Name of the Organization
Department Name	Department Name
Department Cost Center	Department Cost Center

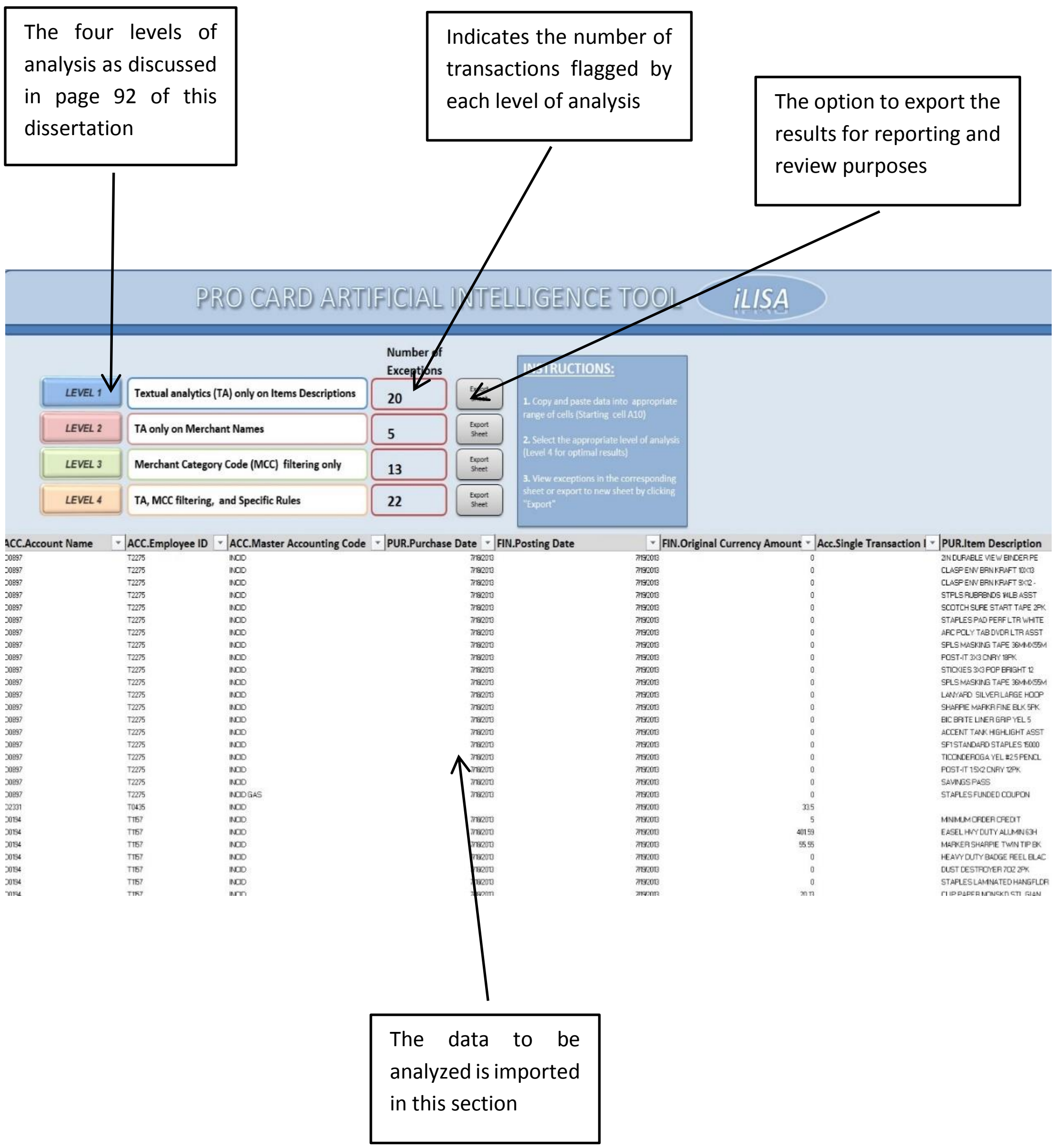
APPENDIX F

Sample of Coded Rule:

```
ActiveSheet.ShowAllData
Columns("A:A").Select
Selection.Delete Shift:=xlToLeft
Range("A2").Select
LastRow = ActiveSheet.Cells(Rows.Count, 1).End(xlUp).Row
ActiveSheet.Range("$A$1:$AR$" & LastRow).AutoFilter Field:=1, Criteria1:= _
   "=*gift*", Operator:=xlAnd
LastRow = ActiveSheet.Cells(Rows.Count, 1).End(xlUp).Row
ActiveSheet.Range("$A$1:$AR$" & LastRow).AutoFilter Field:=9, Criteria1:= _
   "=*10 greatest*", Operator:=xlAnd
If Application.WorksheetFunction.Subtotal(3, ActiveSheet.Columns(1)) > 1 Then
ActiveSheet.UsedRange.Offset(1, 0).Resize(ActiveSheet.UsedRange.Rows.Count - 1).Rows.Delete
Else
End If

ActiveSheet.ShowAllData
 LastRow = ActiveSheet.Cells(Rows.Count, 1).End(xlUp).Row
ActiveSheet.Range("$A$1:$AR$" & LastRow).AutoFilter Field:=1, Criteria1:= _
   "=*gift*", Operator:=xlAnd
LastRow = ActiveSheet.Cells(Rows.Count, 1).End(xlUp).Row
ActiveSheet.Range("$A$1:$AR$" & LastRow).AutoFilter Field:=32, Criteria1:= _
   "=*prestige*", Operator:=xlAnd
If Application.WorksheetFunction.Subtotal(3, ActiveSheet.Columns(1)) > 1 Then
ActiveSheet.UsedRange.Offset(1, 0).Resize(ActiveSheet.UsedRange.Rows.Count - 1).Rows.Delete
Else
End If

ActiveSheet.ShowAllData
LastRow = ActiveSheet.Cells(Rows.Count, 1).End(xlUp).Row
ActiveSheet.Range("$A$1:$AR$" & LastRow).AutoFilter Field:=1, Criteria1:= _
   "=*Cruise*", Operator:=xlAnd
LastRow = ActiveSheet.Cells(Rows.Count, 1).End(xlUp).Row
ActiveSheet.Range("$A$1:$AR$" & LastRow).AutoFilter Field:=35, Criteria1:= _
   "=*baby*", Operator:=xlAnd
If Application.WorksheetFunction.Subtotal(3, ActiveSheet.Columns(1)) > 1 Then
ActiveSheet.UsedRange.Offset(1, 0).Resize(ActiveSheet.UsedRange.Rows.Count - 1).Rows.Delete
Else
End If

ActiveSheet.ShowAllData
LastRow = ActiveSheet.Cells(Rows.Count, 1).End(xlUp).Row
ActiveSheet.Range("$A$1:$AR$" & LastRow).AutoFilter Field:=1, Criteria1:= _
   "=*apple*", Operator:=xlAnd
LastRow = ActiveSheet.Cells(Rows.Count, 1).End(xlUp).Row
ActiveSheet.Range("$A$1:$AR$" & LastRow).AutoFilter Field:=12, Criteria1:=Array( _
   "5111", "5732", "5942"), Operator:=xlFilterValues
If Application.WorksheetFunction.Subtotal(3, ActiveSheet.Columns(1)) > 1 Then
ActiveSheet.UsedRange.Offset(1, 0).Resize(ActiveSheet.UsedRange.Rows.Count - 1).Rows.Delete
Else
End If
```

APPENDIX G

Expert Visual Dashboard Main View:



Expert Visual Dashboard Keyword View:

Keyword used in the rule engine. Each column is color coded based on the level of analysis. Users will have ability to add/delete keywords for each level based on audit requirements

Keywords for Items in LEVEL 1	Keywords for Merchants Name in LEVEL 2	Keywords for Items in LEVEL 4	Keywords for Merchants Name in LEVEL 4	Merchant CC Search	Merchant Desc to Delete
leotard	Towing	leotard	Towing	5813	ADVERTISING SERVICES
Massage	Massage	Massage	Massage	5921	AGRICULTURAL COOPERATIVES
Neon	casino	Neon	casino	7273	AIR CONDITIONING AND REFRIGERATION REPAIR SHOPS
Pringles	parking	Pringles	parking	7297	AIR CONDITIONING, HEATING, PLUMBING CONTRACTORS
Coke 20oz	Animal	Coke 20oz	Animal	7841	ALTERATIONS, MENDING, SEAMSTRESSES, TAILORS
beer	Antiques	beer	Antiques	7995	ASSOCIATIONS CIVIC, SOCIAL, AND FRATERNAL
casino	Bar	casino	Bar	8351	ATTORNEYS, LEGAL SERVICES
parking	Rental	parking	Rental	9211	BANDS, ORCHESTRAS, & MISC ENTRTNRS-NOT ELSWHR CLAS
Alcohol	Cruise	Alcohol	Cruise	9222	BUILDING MATERIALS, LUMBER STORES
Animal Crackers	Dating	Animal Crackers	Dating	9223	BUSINESS SERVICES-NOT ELSEWHERE CLASSIFIED
Antiques	Club	Antiques	Club	8675	CARPENTRY CONTRACTORS
towing	Entertainment	towing	Entertainment	8368	CATERERS
Bar	Escort	Bar	Escort	5814	CHEMICALS/ALLIED PRODUCTS NOT ELSEWHERE CLASSIFIED
Cash	Furniture	Cash	Furniture	5944	CLEANING AND MAINTENANCE, JANITORIAL SERVICES
Rental	Fuel	Rental	Fuel	8211	COMMERCIAL EQUIPMENT NOT ELSEWHERE CLASSIFIED
Cigar	Gambling	Cigar	Gambling	7829	COMMERCIAL FOOTWEAR
Cruise	Motel	Cruise	Motel	7832	COMP PROGRAMING,DATA PRCSNG,INTGRTD SYS DSGN SRVS
Dating	iTunes	Dating	iTunes	7841	COMPUTER MAIN./REPAIR/SERVICES NOT ELSEWHERE CLASS
Club	Jewelry	Club	Jewelry	7911	COMPUTER NETWORK/INFORMATION SERVICES
Drug	Lease	Drug	Lease	7922	COMPUTERS, COMPUTER PERIPHERAL EQUIPMENT, SOFTWARE
Entertainment	Medical	Entertainment	Medical	7929	CONCRETE WORK CONTRACTORS
Escort	Movie	Escort	Movie	7932	CONSTRUCTION MATERIALS NOT ELSEWHERE CLASSIFIED
Furniture	Pharmacy	Furniture	Pharmacy	7933	CONSULTING, MANAGEMENT, AND PUBLIC RELATIONS
Fuel	Resort	Fuel	Resort	7941	CONTRACTORS, SPECIAL TRADE NOT ELSEWHERE CLASS
Gambling	gas	Gambling	gas	7991	COURIER SERVICES AIR & GROUND, FREIGHT FORWARDERS
Gift	Spa	Gift	Spa	7992	DENTAL AND MEDICAL LABORATORIES
Motel	motel	Motel	motel	7993	DENTAL/LAB/MED/OPHTHALMIC HOSP EQUIP & SUPPLIES
iTunes	porn	iTunes	porn	7994	DETECTVE/PROTECTVE AGNCY,SECURITY SRVS, ARMOR CARS
Jewelry	Tickets	Jewelry	Tickets	7995	DIRECT MARKETING - CATALOG MERCHANTS
Lease	Kids	Lease	Kids	7996	DIRECT MARKETING - COMBINATION CATALOG AND RETAIL
Medical	Puzzle	Medical	Puzzle	7997	DIRECT MARKETING - INBOUND TELESERVICES MERCHANTS
Meal	Game	Meal	Game	7998	DIRECT MARKETING CONTINUITY/SUBSCRIPTION MERCHANTS
Movie	University of	Movie	University of	7999	DIRECT MARKETING-OTHER DIRECT MARKETERS/NOT ELSEW.
Pharmacy	Baseball	Pharmacy	Baseball	8011	DURABLE GOODS NOT ELSEWHERE CLASSIFIED
Poker	Basketball	Poker	Basketball	8021	ELECTRICAL CONTRACTORS
Resort	Football	Resort	Football	8042	ELECTRICAL PARTS AND EQUIPMENT
gas	BEDDING	gas	BEDDING	4900	ELECTRONIC REPAIR SHOPS
Spa	CRAYOLA	Spa	CRAYOLA	5812	EMPLOYMENT AGENCIES, TEMPORARY HELP SERVICES
hotel	HERSHEY	hotel	HERSHEY	7631	EQUIPMENT RENTAL&LEASING SVS, FURNTURE/TOOL RENTAL
porn	BARBIE	porn	BARBIE	8031	EXTERMINATING AND DISINFECTING SERVICES
Tickets	SANDISK	Tickets	SANDISK	8041	FREIGHT CARRIER,TRUCKING-LCL/LNG DIST, MVG/STORAGE

Users will have ability to modify the keyword data base as needed without affecting the system

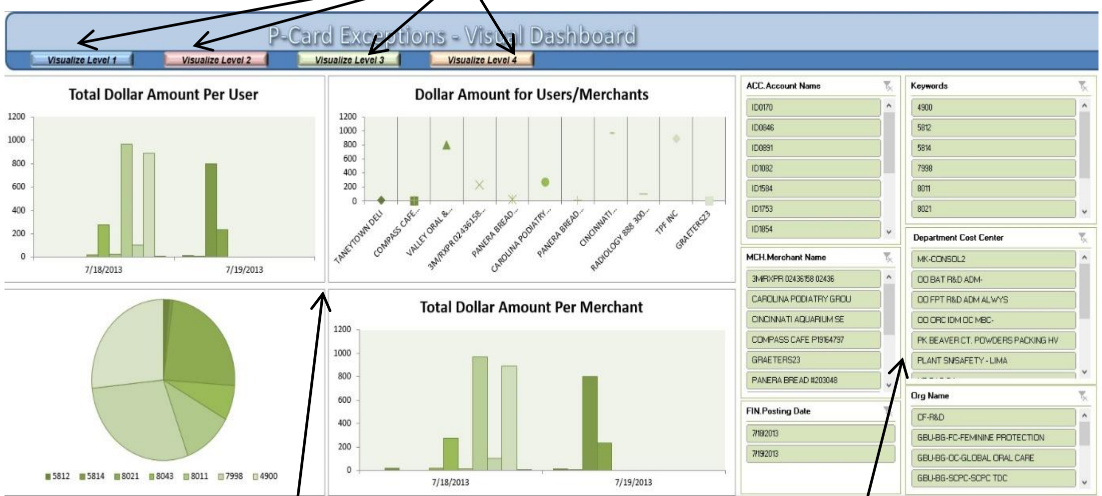
Expert Visual Dashboard Table Results View:

The results of the analysis is shown in the first column. This example relates to level 4, and results are color coded to reflect each category of analysis

Results	ACC.Account Name	ACC.Employee ID	ACC.Master Accounting Code	PUR.Purchase Date	FIN.Posting Date	FIN.Original Currency Amc
Furniture	ID0838	T2613	P&GSOURCED	7/18/2013	7/19/2013	
Medical	ID3238	T1022	INCID	7/18/2013	7/19/2013	
HORMEL	ID1977	T2220	INCID	7/18/2013	7/19/2013	
SAUSAGE	ID1977	T2220	INCID	7/18/2013	7/19/2013	
CHEX	ID2038	T2666	PARCEL	7/17/2013	7/18/2013	
casino	ID2241	T3253	INCID		7/18/2013	
casino, hotel	ID3211	T1054	INCID GAS		7/18/2013	
5812	ID1753	T3005	INCID		7/19/2013	
5814	ID1082	T1059	INCID		7/19/2013	
5812	ID3058	T1251	INCID		7/18/2013	
8011	ID2221	T1439	INCID		7/18/2013	
5814	ID0170	T0054	INCID		7/18/2013	
5812	ID0170	T0054	INCID		7/18/2013	
8011	ID3236	T1094	INCID		7/18/2013	
5814	ID2261	T2973	INCID		7/18/2013	
EXT-AMT > ORG-AMT	ID3175	T0457	P&GR&D	7/18/2013	7/19/2013	
EXT-AMT > ORG-AMT	ID2430	T1299	INCID	7/18/2013	7/19/2013	
EXT-AMT > ORG-AMT	ID2433	T0152	INCID	7/18/2013	7/19/2013	
EXT-AMT > ORG-AMT	ID3166	T2672	INCID	7/18/2013	7/19/2013	
EXT-AMT > ORG-AMT	ID3121	T2846	INCID	7/18/2013	7/19/2013	
EXT-AMT > ORG-AMT	ID3295	T1018	PARCEL	7/17/2013	7/18/2013	
EXT-AMT > ORG-AMT	ID0097	T0397	INCID	7/17/2013	7/18/2013	

Expert Visual Dashboard Visual View:

Ability to visualize each level of analysis. Each color coded to represent its level.



Main visual view. Presents 4 main visualizations that summarize the exceptions generated as discussed in page 99 of the dissertation. The Dashboard is interactive and users can hoover over certain visual outliers for further investigation.

Interaction and navigation for the visual dashboard can be achieved via this side panel. Users will have the ability to filter and focus on certain data points.