UNDERSTANDING WATER USE IN PHOENIX, AZ:

A SPATIAL STATISTICS APPROACH

By

JOSHUA NICHOLAS RANDALL

A thesis submitted to the

Graduate School – New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Geography

Written under the direction of

Laura C. Schneider

and approved by

____

____

____

New Brunswick, New Jersey

May, 2014

ABSTRACT OF THE THESIS

Understanding Water Use in Phoenix, AZ: A Spatial Statistics Approach

by JOSHUA NICHOLAS RANDALL

Thesis Director:

Laura C. Schneider

Water managers in Arizona are facing difficulties due to population and urban infrastructure growth. An understanding of water use patterns is critical to the management of urban water resources. This study focuses on estimating spatial and socioeconomic patterns of water users in the Phoenix Metro area through a series of statistical analysis. Using total water use for 2010 as a dependent variable and 36 socioeconomic characteristics as explanatory variables, four statistical methods were used to analyze the relationships: 1) Individual regression analyses, 2) A multivariate regression analysis 3) a principal components analysis (PCA) and 4) a principal components regression. Results show that water users between ages of 55 to 69 by census tract correlated strongest with total water use in 2010. Results of the multivariate regression of seven socioeconomic variables were able to explain 77% of the variability of water use across the study area. PCA analysis identified three components of socioeconomic variables that in combination explained 73% of water use. From the components four specific socioeconomic groups were identified: high income retiree populations, large Hispanic families, high income families, and low to middle income populations. To analyze the spatial clustering of water use and socio-economic data, local index of spatial autocorrelation (LISA) mapping was used. The identified socioeconomic clusters were found to overlay political boundaries. Recommendations presented include possible water use patterns for each identified socioeconomic group and some suggested programs that may be beneficial to water management. LISA results also suggest that addressing intra-city water management to account for the spatial variability of water use their users across political boundaries is important. The analysis presented here may be used as tool to identify broad spatial and statistical water use patterns, but it has limitations to understanding patterns at the level of households.

Acknowledgements

The completion of this project was not possible without the help and support of many people. I would like to first and foremost Laura Schneider not only for her guidance throughout this thesis (and the many, many drafts it entailed), but for her mentorship and support throughout the entirety of my time Rutgers. I would also like to thank my committee members Asa Rennermalm and David Tulloch for their support on this project. Thank you to the Rutgers Graduate School and the Rutgers Geography Department for providing me an opportunity and academic space to pursue this project. Thank you to the City of Glendale, City of Phoenix, City of Goodyear and the Decision Center for Desert Cities at Arizona State University for providing me with the data necessary to complete this project. Finally thank you to my family and friends who supported me throughout.

Table of Contents

List of Tables

List of Figures

Understanding Water Use in Phoenix, AZ: A Spatial Statistics Approach

Joshua Randall

I. Introduction

Water use across space and scales and between individuals and populations is quite variable. Many factors have been shown to have a possible influence on how much water a population may use (Domene and Saurí 2006; House-Peters and Chang 2011; Wentz and Gober 2007). These factors include physical environment, demographics such as age and income, regional factors, and water pricing. While the large scale management of natural resources is often compounded by the comprehension of these factors, the management of water is difficult due to predictions of water loss and drought, particularly in arid regions (Gober 2010). Water use management is a number one priority in many local and state governments in the southwest US. The ability to make relevant decisions on water use is pressurized by rapid increase in water demand and outdated water laws (White, Corley, and White 2008; Keys, Wentz, and Redman 2007). Understanding the spatial distribution of population and water use provides a nuanced view of this relationship. Despite attention paid to understanding water use factors, little work has been done to address specific spatial variations in conjunction large-scale socioeconomic analysis, particularly in Phoenix and surrounding areas.

The objective of this study is to conduct a spatial socioeconomic analysis of communities in Phoenix, Arizona and their relation to water use. The results are intended to be used as decision making tools for water managers in the area. Through statistical and spatial analyses of socioeconomic variables, this investigation explores patterns of water use in the region. Characterization of the socioeconomic and spatial variability among water users in the Phoenix Metro Area is an integral part of this study. To provide this characterization, three distinct analyses will be used. First, individual socioeconomic variables are analyzed independently and through multivariate regression to determine importance of variables in explaining water use. Second, a principal components analysis and principal components regression will be used to develop the descriptive power of certain socioeconomic variables that affect water use. Lastly, a brief analysis of the autocorrelation trends developed by Chang, Parandvash, and Shandas (2010), of the socioeconomic population variables, and components from PCA analysis will provide an explicit spatial clustering characterization of the population.

Understanding Water Use in Arizona

Between 1990 and 2000 the Phoenix Metropolitan Area grew by 3.5 million people. Residential land use in the area grew by about 250%, attributed mostly to local politics, development regulations, social and cultural choices, and economic factors (Keys, Wentz, and Redman 2007). There were about 4.2 million people in 26 different towns and municipalities in the area in 2010 (Shrestha et al. 2012). Because of this large land use and population growth, present and future availability of water resources are perhaps the biggest issues facing the Phoenix Metro Area. Overall, any sort of intensive land use, be it urbanization or agricultural production, consumes large amounts of water (Stonestrom, Scanlon, and Zhang 2009). In Phoenix, the rate of water use has reached about 2 million acre-feet (651.7 billion gallons) per year. 57% goes to the industrial and municipal sector, 33% to agricultural use, and the remaining 10% to Native American communities (Wentz and Gober 2007). The water inflows of this system include three major sources: The Colorado River, the Salt-Verde Rivers, and groundwater aquifers (Gober et al. 2011). Water sources are managed through various policies and laws at many government levels. The Colorado River was allocated by the federal government to various states through treaties and Supreme Court rulings in the 1930's, and Arizona's allocation of Colorado River water is managed through the Central Arizona Project (CAP). The Salt-Verde River surface water is managed by the Salt River Project (SRP). Groundwater is managed by individual water managers for each town. The Arizona Department of Water Resources is the central state water authority. In 1980 the Groundwater Management Act (GMA) was passed in order to secure water sources for new housing developments with minimal use of groundwater. It also provided conservation provisions for municipal, industrial and agricultural sectors (White, Corley, and White 2008). Attempts at conservation have been implemented in and throughout the Phoenix Metro area, including mandatory retrofitting of low-flow fixtures on old houses and set standards for new ones, public education, price schemes, mailings and so forth (Robert C. Balling and Gober 2007; Larson, Ibes, and Wentz 2013). These programs have been met with varied success, with economic incentives and personal communication having the greatest benefit (Larson, Ibes, and Wentz 2013).

There are many different methods proposed by researchers for policy and decision makers in the realm of urban water use and management. Modeling methods often provide extrapolation of data for future prediction of water use based on infrastructure factors, or a combination of infrastructure and socioeconomic factors (House-Peters and Chang 2011). The variables used are often structural, such as age or size of house, lawn size, room numbers, garden size and so on (Chang, Parandvash, and Shandas 2010;

Runfola et al. 2013; House-Peters and Chang 2011). Many studies focus on socioeconomic variables, but not exclusively, and almost always either using age, population in a household, income, or education (House-Peters and Chang 2011). Attempts to combine socioeconomic variables at multiple layers into one study *and* situate them in a spatial context are rare (Shandas and Parandvash 2010; Chang, Parandvash, and Shandas 2010; Wentz and Gober 2007; Wentz et al. 2013; Franczyk and Chang 2008). Other modes of analysis focus on the political implications of natural resource management, both from a critical perspective (Bakker 2010; Swyngedouw 2005) and a political science perspective (Campbell, Johnson, and Larson 2004; Stoutenborough and Vedlitz 2014; Browne, Medd, and Anderson 2012; Randolph and Troy 2008). Finally, some studies show the viability of mapping the spatial patterns of water use and other variables through an analysis of autocorrelation errors and patterns (Chang, Parandvash, and Shandas 2010; Franczyk and Chang 2008).

II. Methodology

Study Region

The study site is located in the Phoenix Metro area and consists of two cities, the City of Phoenix and to the west - northwest the City of Glendale (figure 1). Combined they consisted of a population around 1.6 million people in 2010 (US Census Bureau 2010). For this study 411 census tracts were used, as delineated by the 2010 Census. All census tracts with only one water customer were removed previous to data acquisition for privacy protection.

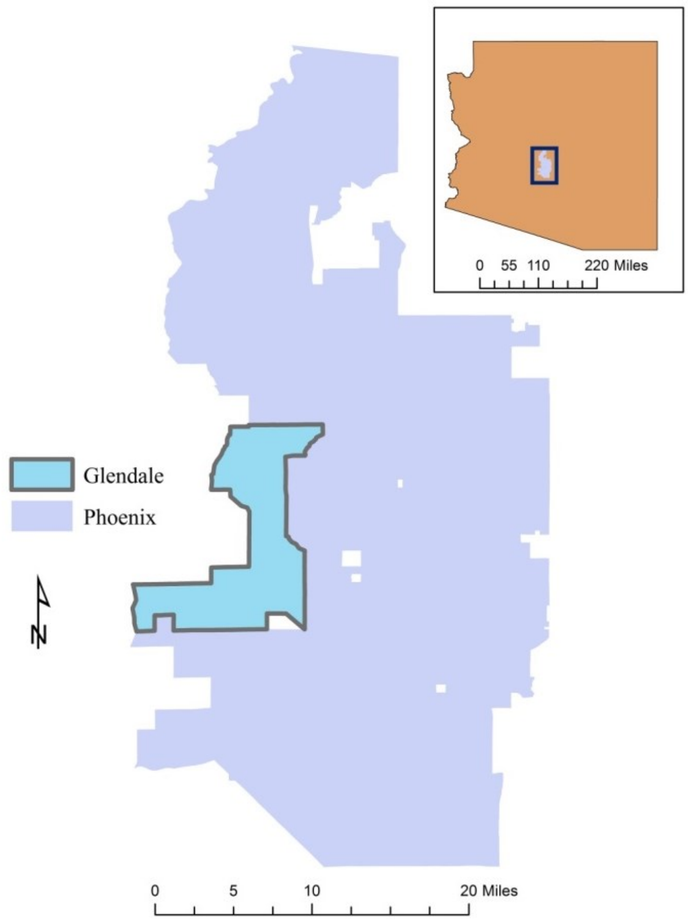

Figure 1: Study Area, Glendale and Phoenix

Water Use Data and Socioeconomic Data

Water use data were acquired from both the City of Phoenix and the City of Glendale (City of Phoenix 2010; City of Glendale 2010). The data from the City of Phoenix came as monthly water use by census block. The data were aggregated spatially by census tract, and summed to total water use in 2010. The City of Glendale data also came by census block, for total water use in the year 2010 and was also aggregated by census tract. Both were converted into total gallons per census tract. Population size and socioeconomic variables data were acquired from the American Communities Survey and Census Survey from the year 2010, through American Fact Finder (U.S. Census Bureau 2010). Thirty-six variables were selected as common measures of socioeconomic status or description (table 1) and many had been used in previous studies in relation of socioeconomic descriptors and water use (House-Peters and Chang 2011).

Table 1: List of Socioeconomic Variables

Independent Variables	
Average Income	Two or More Races
Income $0 - $15,999	Hispanic
Income $15,000 - $34,999	1 person household
Income $35,000 - $74,999	2 person household
Income $75,000 - $150,999	3 person household
Income $150,000+	4 person household
Median Age	5 person household
Age 0 – 19	6 person household
Age 20 – 34	7+ person household
Age 35 – 54	Average Household
Age 55 – 69	Family Household
Age 70+	Non Family Household
White	Average Family Size
Black	Less Than HS
American Indian	HS Grad
Asian	Some College
Hawaiian	Bachelors and above
Other (other races)	Education Average

Spatial Data

To map socioeconomic variables and results of the analysis, the Census Tiger File shapefiles (TIGER 2010) were used as the political boundaries. The shapefiles contain census tract identifiers that match identifiers on water use and population data. Data were imported into the shapefile data using ArcGIS.

III. Statistical and Spatial Analysis

The following methods were used to investigate the relation between water use and socioeconomic data: regression analyses, a multivariate regression, a principal components analysis and Principal Components Analysis (PCA) regression. The use of these methods was expected to provide a clearer understanding of the relationships between socioeconomic characteristics and water use. The final analysis consists of an autocorrelation clustering to investigate the spatial trends of the population characteristics and water-use.

a. Regression Analysis

A Pearson's correlation coefficient was calculated for each variable to examine the individual relationship with water use. The correlation measures provide an indication of how water use is associated with population. The r-squared, or coefficient of determination was also calculated to provide a goodness of fit for each variable (ordinary least-squares regression). Statistical significance of r-squared and log-likelihood power of explanation was also estimated. Log-likelihood power of explanation allows for a comparable measure of model power or explanation between variables. Each variable was also examined for autocorrelation, all of which displayed spatial autocorrelation, thus a spatial error regression was run independently in order to correct for autocorrelation. Spatial error (SE) regressions use a co-efficient term in addition to the standard correlation equation that takes into account the surrounding observations of the variables in conjunction with the identified correlation. The program GeoDa was used to perform the spatial error regressions (Anselin, Syabri, and Kho 2006).

b. Multivariate Regression

A multivariate regression was used in order to study the linear relationships between socioeconomic data and water use. A multivariate correlation coefficient measures the variables' joint degree of association with the dependent variable, and r-squared value is the joint coefficient of determination. The joint coefficient of determination measures how well the variables together explain the variability in the dependent variables (Burt, Barber, and Rigby 2009, 504). Variables significant in either an individual OLS regression or spatial error regression were placed in a simple linear multivariate model. The model was highly multicollinear, indicated by the multicollinearity condition number in GeoDa (generally any condition number above 30 requires review). Variables that were non-significant at a 0.05 level were removed. This process was then repeated twice more and a final equation was obtained. Deemed sufficiently non-collinear based on the multicollinearity condition number, the model was then tested for autocorrelation, and the variables were found to be significantly spatially autocorrelated. The final model developed was corrected for autocorrelation based on only significant variables

c. Autocorrelation

Auto-correlation measures whether adjacent regions exhibit similar or dissimilar patterns in terms of a single variable. This may affect assumptions of traditional statistical tests. Traditional statistical models will not take into account the distribution of variable observations over space, simply looking at the correlation between observations of independent and dependent variables. However, there are inherent issues when dealing with spatial data in an aspatial analysis. People of similar actions or types tend to located together or, the nearer they are in space the more similar they are. In the context of this study, the issue manifests itself in the form of a spatial error. If either the dependent or independent variable is auto-correlated, then observations will be skewed because the variables will be more dependent on itself in the surrounding census tracts (hence "autocorrelation") than on an observation of the other variable. This is corrected through the addition of a spatial error regression. Analysis of spatial autocorrelation can be done through the process of creating "LISA" maps, or maps of the Local Index of Spatial Auto-correlation (LISA). Moran's Index (or Moran's I) is an indicator of spatial autocorrelation, with -1 being perfect distribution of variables in the global space, and +1 being perfect autocorrelation where all similar variables are located next to each other. This is a global statistic, providing one value of spatial autocorrelation for the entire study area. A LISA analysis gives a *local* z-statistic of spatial autocorrelation for each individual census tract, and then maps how similar they are too each other in space (Anselin, Syabri, and Kho 2006).

d. Principal Components Analysis (PCA)

One of the most common data reduction techniques when dealing with many independent variables is a Principal Components Analysis. This technique transforms the data into a new set of linear weighted combinations of the variables, or components. The components provide groupings of variables that, in this case, combine separate socioeconomic variables into groups of weighted variables (McGarigal, Cushman, and Stafford 2000, 23). The 36 variables (table 1) were placed in a matrix linked by census tract identifiers in the statistical program *R*. Using the *princomp* function in R (R Core Team 2013), the data were transformed using a PCA analysis into components. The outputs of the analysis were the weights of each variable in the components. This also includes the variable scores for each census tract. With the variable weights, the components were analyzed for explanatory power based on their variable makeup. Each component had the variables ranked by power of their influence on the component. Because each variable has an influence on the component, the included variables in the component for analysis were cutoff at a loading of 0.10. With the variable z-scores, the components were mapped and analyzed spatially.

e. PCA Regression

Because the components are linear combinations, the subsequent component scores can be used in basic statistical analysis (McGarigal, Cushman, and Stafford 2000, 23). The components were placed into a multivariate regression, and ranked in order of total variance explained. The top three components in explanatory power were then placed in a linear regression model in GeoDa and analyzed for autocorrelation effects. A spatial error regression was then run correcting spatial autocorrelation and to improve the explanatory power of the model.

IV. **Results**

General descriptive statistics on single family residential water use, including mean and standard deviation, can be found in table 2. Average water use in 2010 was around 128 million gallons a year per census tract. The highest use was around 737 million gallons, and the lowest 30 thousand gallons. Spatially the two highest areas of water use are in east-central Phoenix, southeast Phoenix and northwest Phoenix (figure 2). Descriptive statistical summaries of individual independent socioeconomic variables used can be found in table 3. This includes units of measure for each variable.

Water Use

Table 2: Single Family Residential Water Use, Phoenix and Glendale (units in gallons of water, total 2010) (411 observations)

Mean	Median	SD	Min	Max
128,570,908	117,307,344	93377,537	30,668	737,048,532

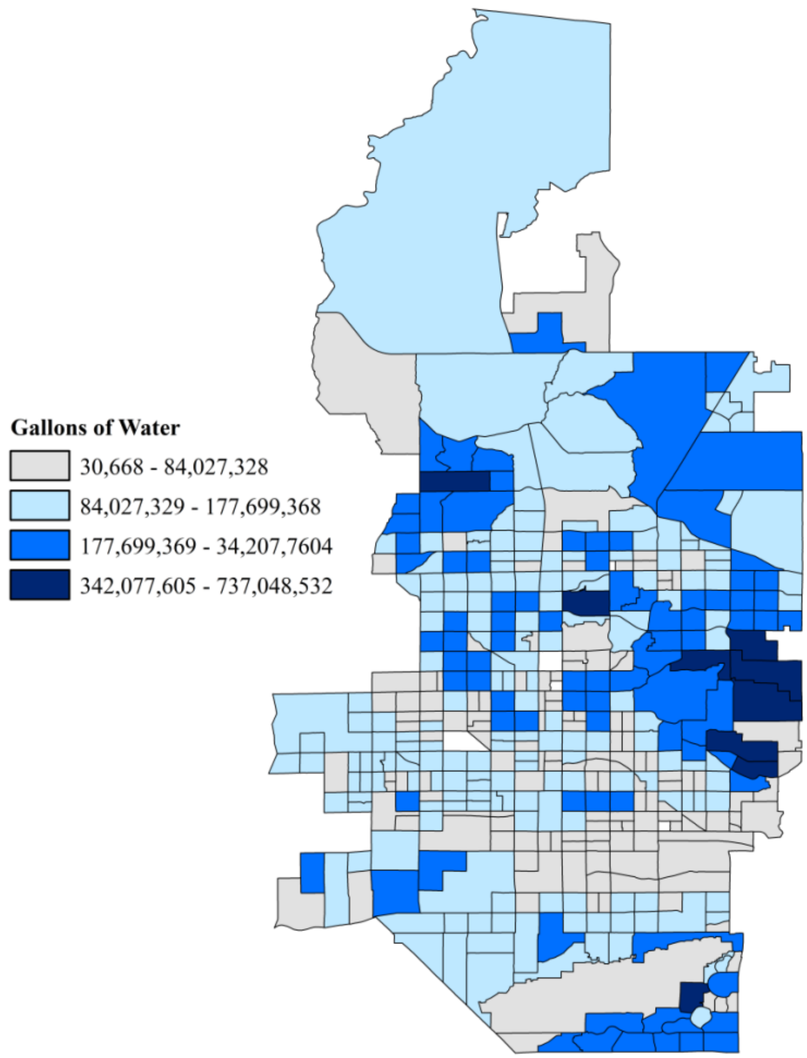

Figure 2: Single Family Residential Water Use, Total 2010

Individual Variables

Table 3: Independent Variable Summary Statistics (411 observations)

Variable	Min	Median	SD	Mean	Max	Unit
Average Income	12009	60595.5	37188.70	68479.72	317222	Dollar
Income $0 - $14,999	0	156	150.54	188.61	1015	Households
Income $15,000 - $34,999	0	301	193.80	324.97	970	Households
Income $35,000 - $74,999	0	489	249.51	494.92	1366	Households
Income $75,000 - $149,999	0	321	229.53	351.11	1231	Households
Income $150,000+	0	55	155.72	117.30	892	Households
Median Age	18.9	32.8	7.46	33.85	60.6	Age
Age 0 - 19	2	1221	635.62	1290.74	3625	Total Population
Age 20 - 34	6	880	465.84	923.34	4532	Total Population
Age 35 - 54	3	1110	437.82	1154.97	2650	Total Population
Age 55 - 69	4	486	262.38	536.73	1544	Total Population
Age 70+	2	199	172.43	238.23	1330	Total Population
White	4	2627	1179.50	2766.32	6591	Total Population
Black	8	173	258.34	258.50	1903	Total Population
American Indian	1	73	69.70	87.03	487	Total Population
Asian	0	93	129.79	133.26	1113	Total Population
Hawaiian	0	6	7.31	7.27	62	Total Population
Other Race	1	518	725.98	741.67	4007	Total Population
Two or More Races	0	143	70.00	149.96	375	Total Population
Hispanic	7	1221	1428.35	1636.07	7344	Total Population
1 person household	1	318	264.98	388.23	1430	Households
2 person household	1	398	218.11	430.01	1115	Households
3 person household	0	219	97.88	227.16	530	Households
4 person household	0	188	91.44	198.52	597	Households
5 person household	0	109	63.61	116.57	376	Households
6 person household	0	47	41.25	58.35	215	Households
7+ person household	0	36	51.62	53.35	272	Households
Average Household	1.19	2.75	0.65	2.87	4.61	Household Population
Family Household	2	930	376.48	961.90	2038	Households
Non Family Household	2	443	328.06	510.29	1770	Households
Average Family Size	2.09	3.25	0.53	3.41	4.78	Family Population
Less Than HS	0	186	219.29	243.49	1208	Homeowner Population
HS Grad	0	316	177.92	332.85	856	Homeowner Population
Some College	0	456	260.53	490.24	1385	Homeowner Population
Bachelors and above	0	337	339.62	410.33	1722	Homeowner Population
Education Average	0	2.72	0.57	2.66	3.72	*

** (1*Less than HS+2*HS Grad+3*Some College+ 4 * Bachelors and above) / (Less than HS+HS Grad + Some College + Bachelors and above)*

Correlation Analysis

The highest correlated variables are shown in table 4, which contains both an ordinary least squares (OLS) regression and a spatial error (SE) regression for the top correlated variables. Regression coefficients explain how much water would be added on average if the unit of observation increased by one in a census tract (table 4). Individual regressions show explanatory power across the entire study area. Because each individual regression was flagged as being spatially autocorrelated, each was run in a spatial regression model. This improved nearly all of the explanatory values of the variable due to a reduction in effect of autocorrelation. Traditional models assume error values of the population are random, allowing for estimates of variability. A spatial error correlation specifically assumes that the errors of the model are spatially autocorrelated, and corrects for the assumption that the error values are random, providing greater calculation of variance in the data set (Anselin, Syabri, and Kho 2006; Chang, Parandvash, and Shandas 2010). It also changed the explanatory power of the top variables. Age and income correlations remained high, while family households correlations improved, as well as white population correlations. This is an indication that the variables regressed to water use are spatially autocorrelated, and when they are mapped should be clustered. The high r-squared value of ages 55-69 in significantly high alone, indicating this variable would be adept in explaining the variability of water use in the study area, and should warrant further review.

Table 4: Individual Regression Results

(dependent variable: water use in gallons per year)

	OLS Regression		SE Regression	
Variable	**R^2**	**Co-ef**	**R^2**	**Co-ef**
Age 55 - 69	0.54	261,686.8	0.63	244,682.9
Income $150,000+	0.50	426,037.7	0.60	456,087.5
Income $75k-$150k	0.44	269,696.7	0.56	259,417.1
Family Household	0.36	149,078.1	0.56	132,862.0
White	0.40	50,081.8	0.56	45,723.9
Bachelors and above	0.40	174,955.1	0.53	183,892.2
Average Income	0.38	1,556.8	0.44	1,475.3

(co-ef in gallons of water) (all variables significant at .05 level)

Multivariate Regression Analysis

In order to provide a greater model description of water users across the valley, a multivariate regression was used. The resulting formula had seven significant variables all at a .05 level (table 5), including income, age and household size and education variables. The model had an R-squared of 0.75, indicating a moderately high explanatory power from a simple multivariate regression. In order to correct for autocorrelation, a spatial error model was used (table 6), as with the individual variables. The improvement was significant as indicated by a log likelihood score increase, and the r-squared improved from 0.75 to 0.77. This indicated that out of the 36 variables that were initially used, these six, after correction for spatial errors caused by autocorrelation, accounted for 77 percent of the variation of water use in the area. This model could be considered a collection of the most statistically important explanatory variables of water use in the study area. They were all significant and provided variability of selection across socioeconomic categories.

Table 5: Individual Ordinary Least Squares (OLS) Multivariate Regression

Variable	Coeff.	Std. Error	t
Constant	*-14,883,350*	*10,246,170*	*-1.45*
Average Income	326.98	118.72	2.75
Income $35,000 - $74,999	56,021.37	14,817.8	3.78
Income $150,000+	229,430.6	30,967.78	7.41
Age 55 – 69	177,795.5	15,162.47	11.73
1 Person HH	-93,768.95	13,579.24	-6.91
Less than HS Education	-51,892.26	21,574.73	-2.41
7 Person Household	376,549.5	94,545.18	3.98
R^2	0.757	**Log-lh**	-7834.89

(co-ef in gallons of water) (all variables significant at .05 level)

Table 6: Spatial Error Multivariate Regression

Variable	Coeff.	Std. Error	z-value
Constant	*-9,617,475*	*107,350,80*	*-0.896*
Average Income	280.36	119.99	2.34
Income $35,000 - $74,999	59,635.46	15,315.75	3.89
Income $150,000+	242,870.7	31,262.10	7.77
Age 55 – 69	174,489.1	15,982.8	10.92
1 Pe^rson HH	-102,283.6	14,309.5	-7.15
Less than HS Education	-55,721.49	21,542.87	-2.59
7 Person Household	382,310.5	98,226.15	3.89
Lambda	*0.34*	*.07*	*4.66*
R^2	0.772	**Log-lh**	-7826.17
(co-ef in gallons of water) (all variables significant at .05 level)			

Principal Components Analysis

The first three components from the PCA were retained for analysis, as they were the only components individually above 5% variance explanation. The total population variance explained for the first three components combined was 78%. Variable loadings used for analysis were limited to above 0.10 or below -0.10 in order to simplify interpretation of the components makeup. Component 2 will be discussed last, as it was an outlier population component, but provided a large portion of explanation in terms of population variability.

Component 1: High-income Retirees/Large Hispanic Families

The first component (table 7) explained 34% of the population variance. Component 1 is split, into two distinct socioeconomic groups. In the case of this study, the divergent groups seem to be driven by a combination of all the categories of socioeconomic variables. This component contains high loadings in median age, high education, and high-income. Also included are small households, indicating the component represent highly educated, and high-income retirees and near retirees. The negative loadings indicate low-income, large, poorly educated Hispanic households (reinforced by the age 0 – 19 group).

Table 7: Results of PCA: Component 1 Variable Loadings

Variable	Loading	Variable	Loading
Median Age	0.23	Hispanic	-0.27
Education Avg.	0.23	Other Race	-0.27
Income Average	0.18	7 Person Household	-0.26
Bachelors and Above	0.18	6 Person Household	-0.26
Income $150,000+	0.17	Less than HS (Edu.)	-0.25
Income $75,000 - $149,999	0.11	Family Average	-0.24
2 Person Household	0.11	Age 0 - 19	-0.23
		5 Person Household	-0.23
		Average Household	-0.21
		Two or More Races	-0.21
		Age 20 - 34	-0.18
		American Indian	-0.16
		Black	-0.15
		Income $15,000 - $34,999	-0.15
		HS Grad	-0.14
		$0 - $14,999	-0.12

Component 3: High-income Families/Lower to Middle Class

The third component (table 8) contained about 12% of the variance of the population. Although this component is similar to the first in income, it also loads heavily in mid-household size range and a young to middle age range, indicating high-income family households. The negatively loaded variables indicate these families live in areas with homogenous income, shown by the variables of non-family households, one person households, and the entire 0 to 75k bracket. The negatively loading variables form a socioeconomic group of their own, non-family low to middle income households.

Table 8: Results of PCA: Component 3 Variable Loadings

Variable	Loading	Variable	Loading
Average Household	0.26	1 Person Household	-0.34
Income Average	0.25	Non-Family Household	-0.34
4 Person Household	0.25	$0 - $14,999	-0.31
Income $150,000+	0.21	Income $15,000 - $34,999	-0.28
5 Person Household	0.18	American Indian	-0.24
Family Household	0.17	HS Grad	-0.18
Income $75,000 - $149,999	0.16	Some College Edu.	-0.11
Age 0 - 19	0.15	Income $35,000 - $74,999	-0.11
Family Average	0.14		
Age 35 - 54	0.14		
6 Person Household	0.13		
Asian	0.11		

Component 2: (Outlier/High-Low Population)

The second component explained 32% of the variance (table 9). In component 2, all the variables were positively loading except one. This is likely an outlier component as there is no socioeconomic group distinguishable. Due to the structural patterns of this component, it is most likely areas of low and high populations and/or households. A spatial error correlation was run between the total populations of census tracts and component 2 which resulted in an r-squared of .90, supporting the claim that component 2 accounts for total population.

Table 9: Results of PCA: Component 2 Variable Loadings

Variable	Loading
3 Person Household	0.27
White	0.27
Family Household	0.27
Age 35 - 54	0.26
2 Person Household	0.26
Some College Edu.	0.25
Age 55 - 69	0.24
Income $35,000 - $74,999	0.23
Income $75,000 - $149,999	0.22
4 Person Household	0.22
Bachelors and Above	0.19
Non-Family Household	0.19
1 Person Household	0.18
HS Grad	0.18
Age 70+	0.17
Age 20 - 34	0.16
Two or More Races	0.16
Asian	0.15
Age 0 - 19	0.14
5 Person Household	0.12
Income $150,000+	0.12
Income $15,000 - $34,999	0.12
Education Avg.	0.12
Family Average	-0.10

PCA Regression

Each of the three components was placed in an individual OLS model and evaluated for autocorrelation. Each was flagged for spatial error, and was subsequently run as a spatial error regression. They give an indication of the general interaction between water use and the components. The regression was limited to the first 3 components, as the rest of the components did not explain as much variance in the population. Basic correlations for water use were measured, as components with small population variance explanation may have a large correlation (Jolliffe 1982). No component was found to have a large correlation. Each of the three components retained showed moderately high autocorrelation, as can be shown by the vast improvement between the OLS regression and SE regression of each component (table 10). However, component 1 was non-significant in the SE correlation analysis. The three components were also placed in an SE regression, creating a multivariate principal component analysis (table 11). Based on the z-values, component 2 and component 3 had the most influence on the model. In this model, four separate socioeconomic groups (two in component 1, two in component 3) along with the low-high population component 2 explain over 73% of the water use variability in Phoenix and Glendale in 2010.

Table 10: Individual Component Correlations

Component	OLS R^2	SE R^2
1 (High-income Retirees/Large Families, Hispanic)	0.080	0.337*
2 (Outlier/Low-High Population)	0.301	0.490
3 (High-income Families/Lower to Middle Income)	0.251	0.508
*(*non-significant) (all other variables significant at .05 level)*		

Table 11: Spatial Error Multivariate Component Regression

Variable	Coeff.	Std. Error	z-value
Constant	*127,185,800*	*6,440,468*	*19.75*
1 (High-income Retirees/Large Families, Hispanic)	7,879,336	1,156,097	6.82
2 (Outlier/Low-High Population)	16,147,460	856,383.1	18.86
3 (High-Income Families/Lower to Middle Class)	28,620,960	1,580,117	18.11
Lambda	0.63	0.05	12.09
R^2	0.734	**Log-lh**	-7870.46
(all variables significant at .05 level) (gallons of water)			

Mapping Autocorrelation (Local Index of Spatial Autocorrelation)

The following three variables: ages 50 – 69, income $150,000+, Family Households were mapped using the LISA technique (figure 3, figure 4, figure 5) as well as the three components (figure 6, figure 7, figure 8). For figures 3-8, red indicates areas of high instances of the variable or components surrounded by other areas of high counts of the same socioeconomic characteristic. Blue is a low counts surrounded by other low counts, light red is a high count surrounded by low counts, and light blue is a low count surrounded by high count. For example for component 1 (figure 6) the red areas would be areas of high counts of the component, and the blue would be areas of low counts of the component. Because there is a positive loading and a negatively loading group of variables associated with this component, it means that red is considered an indication of where the positively loaded variables are, and blue is where the negatively loading variables are. The colored areas then indicate areas of patterns in a spatial context, indicating clusters of both the negatively and positively loading variables. This clustering is also true of water use and component 3. Clustering indicates spatial autocorrelation. In this case component 2 shows fairly moderate spatial autocorrelation with a Moran's I of 0.26, with the rest showing moderate to high spatial autocorrelation. This indicates that through all of the variables there are underlying spatial patterns, with similar rates of water use and similar types of population likely to occur in the same area. As evidenced by the cluster maps, variables often overlap with each other, with water use, or both. By using a PCA to reduce the variables and regressions to relate to water use, analysis of neighborhoods of populations and water use can be more efficient.

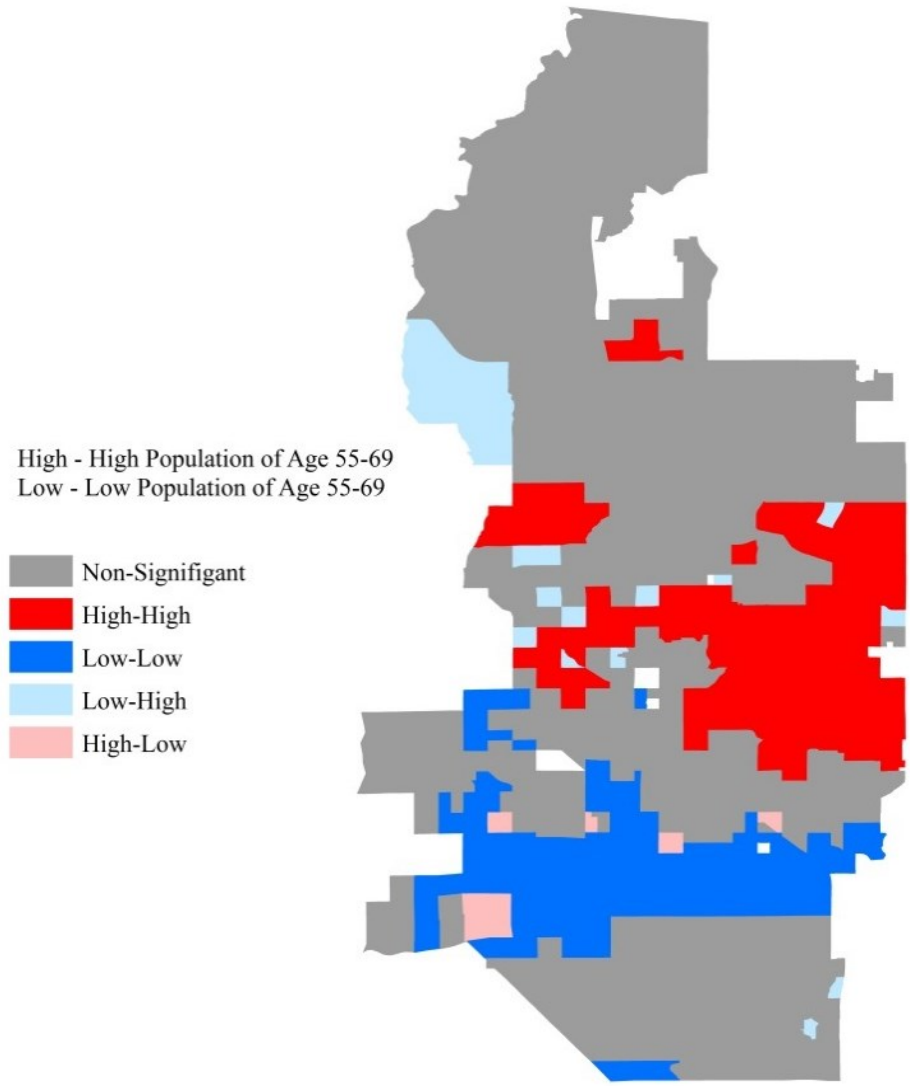

Figure 3: Age 55-69 LISA Cluster Map

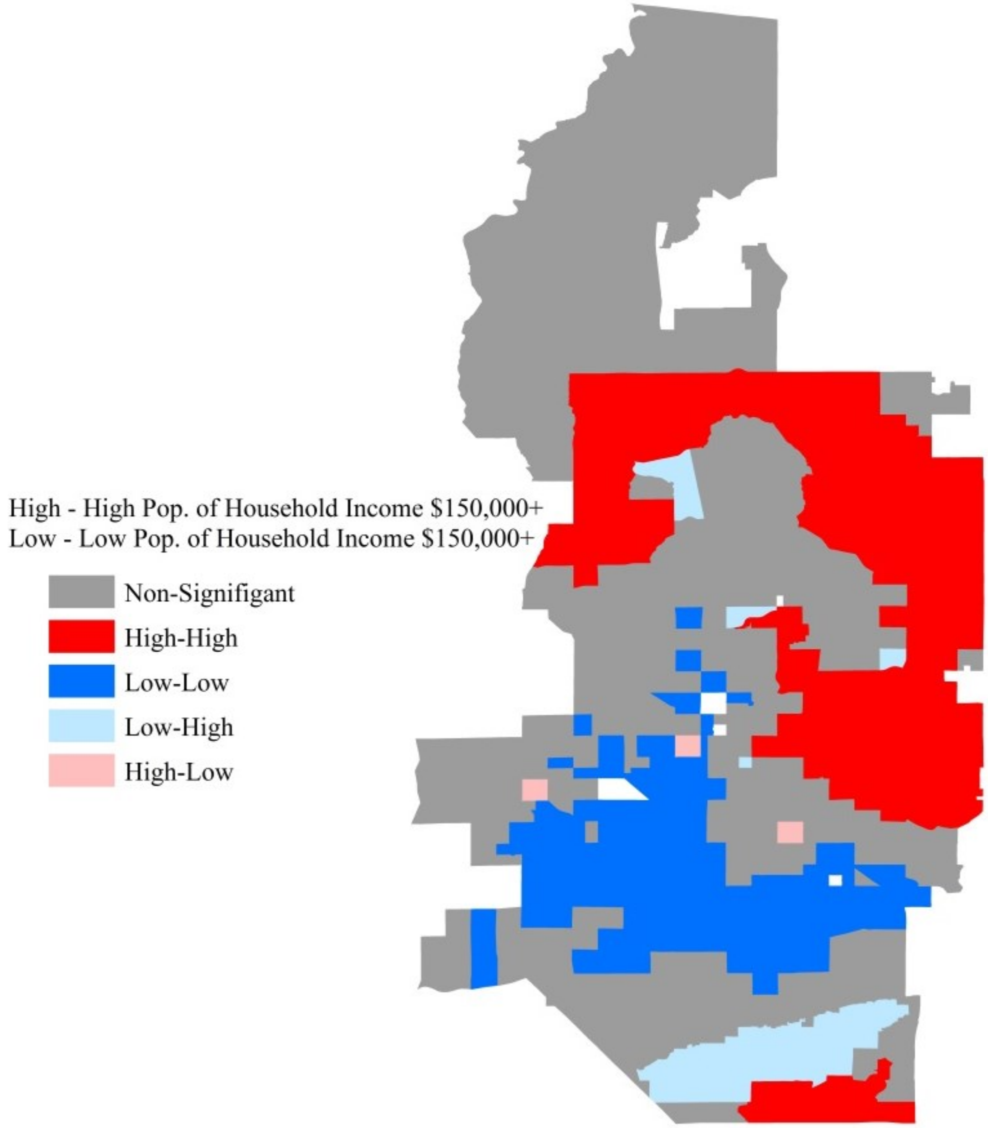

Figure 4: Income $150,000+ LISA Cluster Map

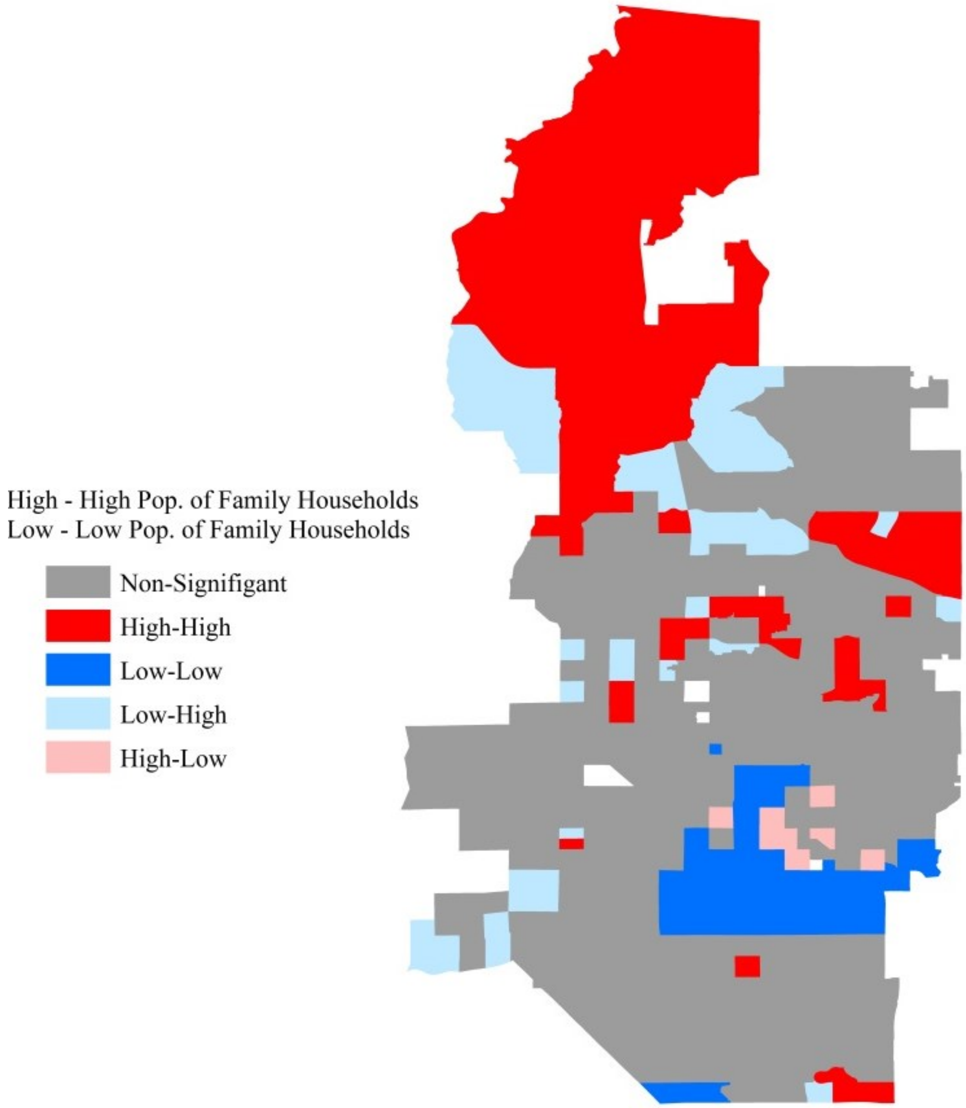

Figure 5: Family Households LISA Cluster Map

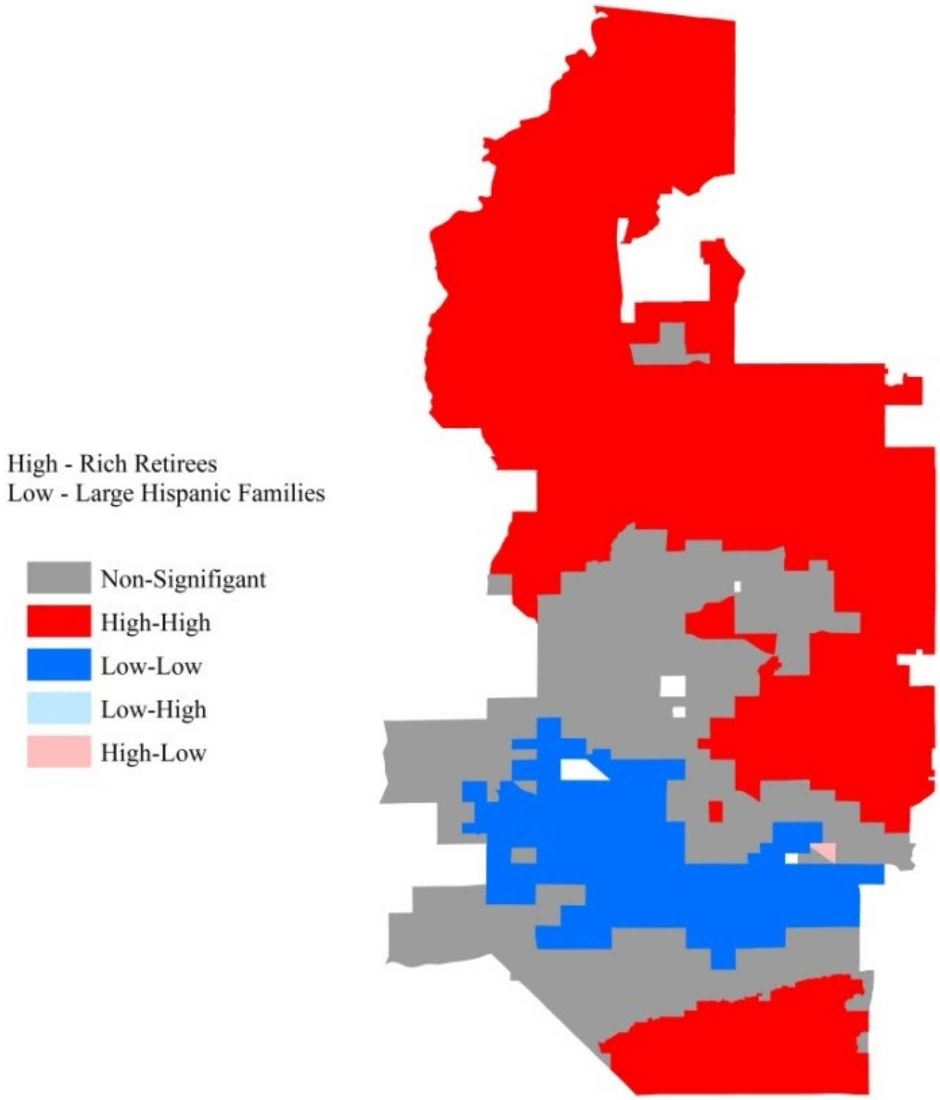

Figure 6: Component 1 LISA Cluster map

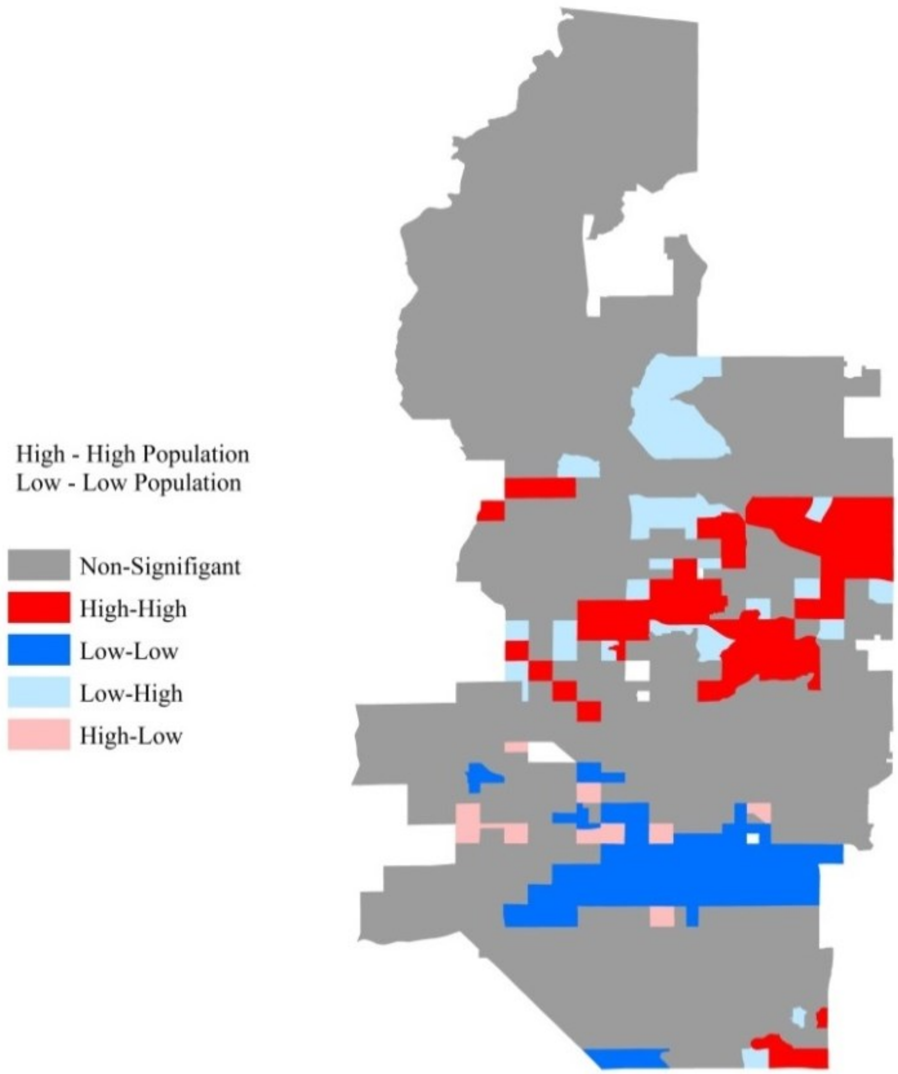

Figure 7: Component 2 LISA Cluster Map

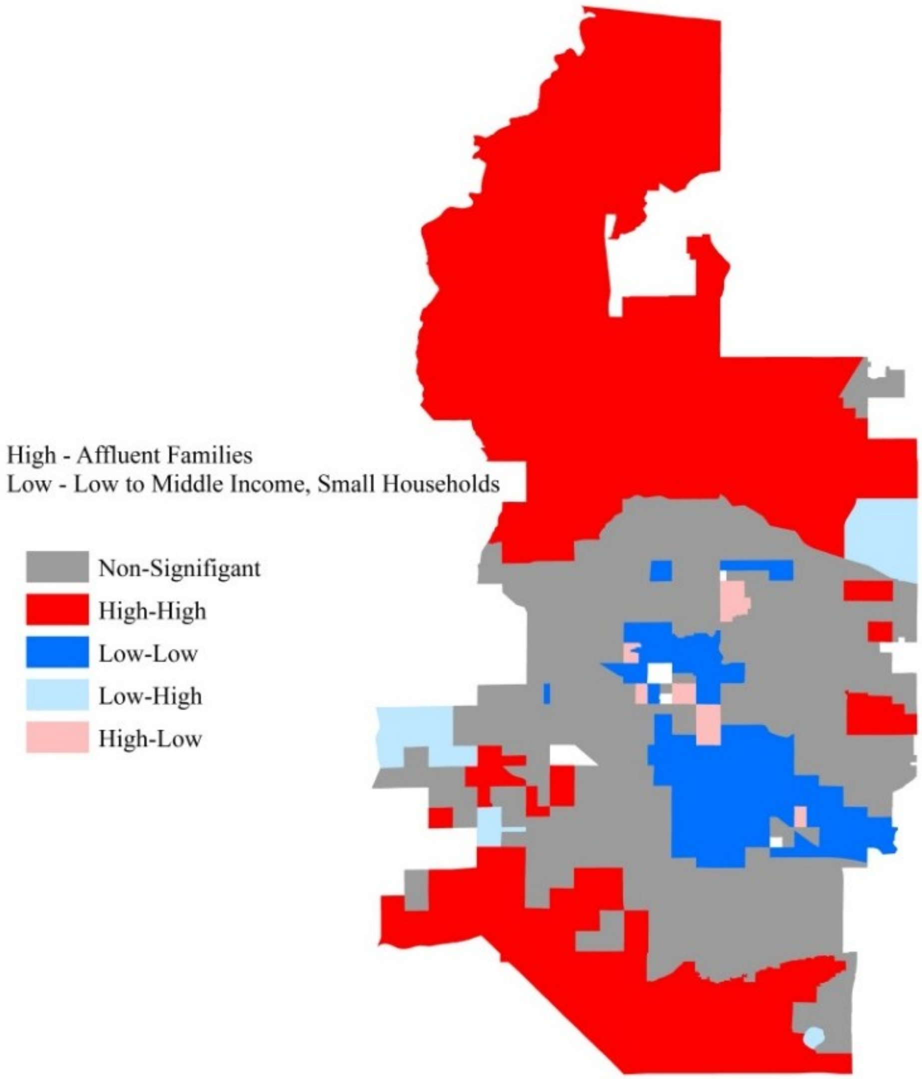

Figure 8: Component 3 LISA Cluster Map

V. **Discussion**

Socioeconomic Data

One of the goals of this study is to define water users based on their socioeconomic characteristics. The resulting groups could provide an understanding on decision making by communities in Phoenix, including planning and public service education on water conservation and use. The use of socioeconomic variables allows for decision makers and researchers to look at the variables as descriptors of water use. It is not intended to provide a future water model in terms of water use totals. Presence of high-income, family households, and older age explained most of the variability of water use across the study area. $150,000+ households use much more water than $75,000 - $150,000 households. The addition of one $150,000+ household leads to an increase of about 456,088 gallons of water a year, while an addition of one $75,000 - $150,000 household leads to an increase of about 259,417 gallons of water per year. Data provide estimates on the importance of socioeconomic variables when making decisions, or rough estimates on how much water may be used in certain areas based on the socioeconomic makeup of a census tract.

Multivariate Regression

This model attributed explanation to education, age, income, and household size, with the most important variable being the age group 55-69. The negative coefficients include one person households and lower income. These variables drive down water use in census tracts, while the remaining variables indicate higher use of water in the census tract. The multivariate regression provides a model the strongest explanatory variables, but removes any that may be used in describing the socioeconomic patterns of the study area. A multivariate analysis includes the interaction between independent variables as well as their joint relationship with water use. This is especially important as it includes variables that did not have high explanatory power individually, but did in combination with other variables. For example, both 1 person and 7 person households were positively correlated with water use, but neither is strong individually in explaining variance in water use.

However within the multivariate analysis, both explained enough unique variance in combination to be included in a multivariate analysis. Further, 1 person households had a negative coefficient. This means that independent of other socioeconomic variables, the presence of 1 person households means an increase in water use. But when included in the multivariate regression, the presence of 1 person households reduces water use, in combination with the other variables in the same location.

PCA

Indications of multicollinearity and lack of variables found in the multivariate equation showed that a data reduction method would be appropriate, as many of the variables explained similar variability in terms of water use. Interpretation of these components can provide an explanation on how socioeconomic data overlap with each other providing a much more defined set of groups. The PCA resulted in five distinct groups: one outlier group, as well as two divergent groups in the first and third component. The three components and five groups account for over 78% of the variability in the population only, without respect for water use. The second component warrants further analysis, but seems to be pulling the outlier group of low and high population single family residential areas in the valley. This is indicated in the high levels of nearly every socioeconomic population variable in those tracts. The first component was a divergent component, with two discernable socioeconomic groups: large, Hispanic households and high-income small, older households. Household size splits this component, with large households also being linked, low-education and the very old and very young. This indicates multi-generational households and households with children. The positively loading variables include higher age, very high end income, small households, all of which indicate retiree or near-retiree couples. The divergence means broadly that the divergent groups do not live in the same areas, and are spatially distinct. This separation is important in terms of conservation education and planning. Small households with large disposable incomes will have different water priorities compared to those of large families with multiple children. Of the top three components this has the smallest correlation to water use. This is most likely due to the fact that the negatively socioeconomic group was contained large households. As the negative loading increases, the amount of large households also increases, and with it water use. The relationship is then less linear, creating a less robust model. The third component was the highest in terms of correlation to water use and contained most of the highest correlated individual variables from the individual correlation analysis. This component contains high-income, medium sized to large families/household sizes. Higher incomes and larger households often leads to the ability to perform other water related activities with such as pools, lawns, gardens (Balling, Gober, and Jones 2008). It also reinforces the fact that family households with tend to be next to other households of similar makeup. The second half of the component is identified as lower middle class, due to the fact the lowest 3 income levels were loaded in combination with small non-family households. As before, there is a trend of high levels of the negative variables (or extreme low end of the component) seeing an increase in water use. This is due to the fact that an increase on the extreme ends of each component means an increase in population. For both component 1 and 3 the overall trend is still positive to the positive side of the component, also meaning that those socioeconomic groups, at a comparable population level, will use more water. The four identified major socioeconomic groups in Phoenix are high-income retirees, high-income families, non-single family residential/low population areas, large Hispanic families, and lower to middle class population.

PCA Regression

PCA regression was estimated to gain a comparison of how three PCA components explain water use. Results were compared to the previous multivariate regression (spatial error). The r-squared for the SE regression of the original variables is 0.75 and the SE r-squared of the PCA regression was 0.73. The difference in interpretation is important to note. The component regression is using five distinct groups and three combinations of variables to examine water use. While this may have (slightly) less explanatory power, it does provide more information in regards to populations and their water use. This is due to the increased amount of information from the ability to combine variables into components. The individual variable regression may provide more variables to make decisions on the surface, and can be used as such. But the increase of information from the components allows for an in-depth view of correlated socioeconomic variables in one analysis. One example is income above $150,000. This bracket is prevalent in two major socioeconomic groups, high-income retirees, and high income families. The two distinct groups may have different political and social implications in terms of water use (Campbell, Johnson, and Larson 2004) which will be discussed in the next section. An important management aspect of this equation is that water variability overall is explained to nearly the same extent as with the multivariate analysis of individual variables. This allows for the components to replace individual variables, as not much explanatory power is lost between the two in explaining water variability.

Spatial Autocorrelation

The concept of autocorrelation as something that needs to be corrected can be held true when applying statistical analysis to a spatial data set, but it is also something that can be used to accurately describe the spatiality of variable clustering. If groups of variables exhibit areas of similarity when they are variables of human activity or measures of human profile, they can accurately be described as water use “neighborhoods”, as termed by Chang, Parandvash, and Shandas (2011). This distinction is important in this project for a few reasons. “Neighborhoods” can be determined as specific groups of water users, and if they overlap city boundaries there may be discrepancies in how water managers may deal with specific groups of water users. The presence of bureaucratic splits can be confusing and/or counter-productive to conservation and educational efforts. The ability to map clusters allows variables and components to be used in spatial analysis. The 3 variables mapped, income $150,000+, age 55-69, and family households (figure 3, figure 4, figure 5) do overlay water use indicating why they were the highest correlated variables with water use. Spatial analysis shows how the variables like high-income (income $150,000+) was split between components 1 and 3 and water use (figure 6, figure 8). Family households appear clustered in high similarity in the north and northwest portions of Phoenix, while 50-69 appears to be clustered in East Phoenix. These two areas have high water use, and high-income. The PCA split the two populations as high-income and high amounts of family households (Component 3), and high-income and high age (Component 1). Both correlated highly with water use but are much different in makeup. Further, each of the top variables had cold spots in the low population area that is indicated in Component 2 (Southeast Phoenix). High autocorrelation was expected for the last two components in particular because of the distinct groupings that resulted from it. By default, the components with highly positive and negative correlations contain groups that are likely not to occur in the same areas. This is supported by the high Moran's I value (Component 1 = 0.69, Component 3 = 0.52) for the last two components.

VI. Water Use and Management in Phoenix: Discussion of Results

The previous section identified the relationships between water use socioeconomic characteristics in the Phoenix Metro area through both an analysis of the spatial clustering with autocorrelation and an analysis of the content of the clusters with principal component analysis (PCA). This section will explore social implications that the results may have.

There are three points from the analysis that warrant further discussion. The first is a focus on socioeconomic differences in how water use is related within the two components and four major sub-populations identifies, specifically on conservation education and the patterns of water use in different types of households. The second point is a discussion on city organization and how water managers may have to deal with neighborhoods of similar water users across the physical boundaries of a city. This section will also focus briefly on the hierarchy of the water political environment (state, local, federal, tribal), as well as the management of sources of urban water in a political context. The third point will focus on the differences of socioeconomic data used in analysis of urban water use, specifically on the $150,000+ income users and the differences both spatially and contextually within the Phoenix Metro area.

1. Water "Neighborhoods"

This section discusses water use and conservation techniques which are considered suitable for the four major groups identified by the principal component analysis.

The first specific socioeconomic group identified contains four variables including high age, highly educated, high-income, and two person households. Based on the socioeconomic characteristics, it was labeled "high-income retirees". In Phoenix, retirees may often be non-native to Arizona. Mobile retirees are often married, newly retired and relatively well off financially (Monk 1994). This population may be considered "amenity migrants" (Monk 1994) as they are often well off financially and socially. According to the Journal of the American Society on Aging: "Phoenix, long a haven for retiree migrants from the Midwest, also shows sharp disparities between the number of whites in its senior population compared with the number of minorities in its youth population" (Longino Jr. 1994). The Phoenix Metro area as a whole had the largest age gap between minorities under 18 (56.2%) and above 65 (14.8%) in the country, at a difference of 41.5%. This is only expected to increase from 2010 to 2030, as Arizona's 65+ age population may increase 157% (Longino, 1994). In 2010, the 55-69 age bracket correlated the highest with water use. If the population is expected to increase drastically in that age bracket, all else equal, water use alone will increase. As people aged they tend to use more water (Schleich and Hillenbrand 2009; Kenney et al. 2008). An explanation may be that near-retirees work less hours and retirees spend more time at home. Retirement age populations may use more water do to simply being home more often.

Another key variable explaining increase in water use is income. Income generally leads to an increase in water use, often due to affluence (Larson, Ibes, and

Wentz 2013), particularly in Arizona. High-income households generally have larger lots, larger homes, are more likely to have swimming pools and are generally more likely to use more water because. Water prices are not an issue, as prices are low enough in the area to eliminate this as a decision making factor (Wentz and Gober 2007).

It is common to address the inability to reduce water use through economic pricing through conservation education. High-income areas often do not respond well to monetary incentives, and Larson (2013) suggests targeting stewardship programs and mandates on specific practices, such as pool covers to reduce water use. Retiree (often from out of state) communities may be defined by extreme geographic clustering, and this may be used to water managers' advantage. By creating a sense of place and environmental stewardship in a desert, retirees may feel more in touch with the area and more willing to limit unnecessary water use. Campbell, Johnson, and Larson (2004) showed that those areas with a higher population originally from Phoenix and originally from the Southwestern US generally used less water, indicating residents from the southwest may indeed have a sense of water limitations in the region.

The second group identified from component three was also high-income. This group was associated with high-income families as well as households with children and likely characterized by the presence of two generations (parents living with children). This group most likely follows the trends of affluent households, in particular use of water for residential activities such as lawns and pools. Family households and 4 person households are relatively good indicators of high water use. Generally, children use less water than adults, but in combination with affluence and increased household populations that may results in high water use. It would be important then to focus on high-income family households regarding conservation efforts. Studies have shown that an increase of households with children in an area will lead to a decrease in water use per capita, all other factors equal (Campbell, Johnson, and Larson 2004; Wentz and Gober 2007). This is likely due to the fact that children use less water than adults. However, there are expectations that children are more receptive to water conservation messages, particularly at school and in the media. This effort is supported by the importance of youth water education in a family setting. It may affect a family's water use, as the parents may be more conscious of water use due to ethical and role model expectations (Campbell, Johnson, and Larson 2004). Previously it was discussed that those who grow up in the southwest, are typically water-conscious and use less water. Part of this may be attributed to water use education as children. As a water manager it would be beneficial to target this group of high-income families particular in the context of a family water ethic, or through youth education. These techniques may not help reduce water in the present in typically affluent households, but may have lasting impacts on future residents in Phoenix. Though the commonality between the two groups mentioned above is high-income, it may be effective to approach water conservation for these groups through age and community relevant conservation techniques and education. In Phoenix, there are promising results of conservation education from targeting efforts at children and assisting elderly (Larson, Ibes, and Wentz 2013).

The third water-use "neighborhood" that was identified was the negatively loading variable group from component 1, or large, Hispanic, multi-generational (grandparents and children living together) households. This group showed generally lower water use in comparison to high-income retirees. Large Hispanic households do not occupy the same locations as high income retirees and near-retirees. Within the Hispanic communities, the size of the household itself may not have much effect on residential water demand (Wentz and Gober 2007). It does however provide an indication of the types of water users in these areas. For water managers the concepts of youth and family education still apply due to the presence of younger generations. Further analysis may be needed to determine if these programs would be beneficial being conducted in Spanish as well as English. Race/ethnicity was identified by Campbell, Johnson, and Larson (2004) as a possible barrier to conservation practices in Phoenix, particularly in the Hispanic communities, possibly due to political alienation, cultural differences, and language. They found that there was indeed an increase compared to all other minority water users in Phoenix, although to find the target of the conservation efforts (cultural or language) would require further analysis. It should however remain a focus, as the Hispanic population expects to increase drastically. This study specifically showed that the presence of large Hispanic households tended to lead to an increase of water use.

The fourth group was the negatively loading variables within component 3, which here is identified as the low to middle income range. The most prominent issue is tailoring a specific water management program to low-to-middle income households. In households on the low range of this income, there can often be an increase in proportional water-use due to failing infrastructure and the lack of disposable income to replace and fix high water use appliances (Campbell, Johnson, and Larson 2004; Chang, Parandvash, and Shandas 2010). This group combined with the high-income families had the strongest correlation of any component of water use, and was the most significant variable in the PCA regression in terms of explanatory power. High-income households as discussed previously, generally result in increased water use through affluence as opposed to directly through the ability to purchase more water. Therefore areas of low-income may be influenced directly by the *lack* of affluence. Water is cheap in Phoenix, generally enough to where it does not directly interfere with decisions making about water (Wentz and Gober 2007). However it was found that an increase in water use from low-income area all else equal was present, most likely due to lack of proper infrastructure (Campbell, Johnson, and Larson 2004). Overuse of water can be found in both low and high-income households, the high to affluence and the low to failing infrastructure. Because of the structural implications of water use, it may be more appropriate to focus efforts in areas where infrastructure failures may be an issue, for example programs specially dealing with replacement of faucets and appliances. Some programs, including rebate systems and retrofit training have been attempted with success (Larson, Ibes, and Wentz 2013). Lower income often means less outdoor activities, particularly with water use (pools, lawns) (Domene and Saurí 2006; Wentz and Gober 2007) which not only indicate different ideas of water in terms of income but that usage patterns are not totally dependent on income but definitively a byproduct of it.

2. Spatial and Hierarchical Water Politics

Each city has its own water managers and water laws, and this section will provide evidence that similar populations occur across borders, at least in this instance, and water management in urban areas may need to be re-evaluated. Understanding the spatial scales at which water use operates is important, as water demand patterns and management may change at different scales. There has been an increase on the use of spatial data for urban water studies as spatial data have increased in collection, accuracy and resolution (House-

Peters and Chang 2011). It includes the use of spatial statistics with increasing accuracy and inclusion of different types of data. Two cities were used for this study, Glendale and Phoenix. Glendale contained 53 of the 411 census tracts and is west to northwest of Phoenix. The proximity of the cities and amount of integration (or lack of distinction of a border) is indicative of the political boundaries of many urban areas, including Phoenix. Both Phoenix and Glendale have individual water management sectors, with individual providers and laws (among the other cities in the Phoenix area). In this study, the identified socioeconomic groups and water use clusters also overlay these boundaries (figure 9, figure 10, figure 11). There is a pattern for northern Glendale consistently splitting both high-income retirees, and affluent families. Management issues may be heightened across the borders, in anything from prices to water access to educational management, especially with high-income families and retirees loading highly in conjunction with high water use. However, the cluster of similar neighborhoods also allows for collaboration in management processes, between Phoenix and Glendale. Particular focusing may be placed on these water use neighborhoods. Many towns in the area have focused collaborated in ad-hoc groups discussing management and water issues across the valley. The results show that a similar plan for water conservation education and management may be needed. The locations of clusters and associations with water used need attention, specifically the differences in spatial configurations (figures 9, figure10, figure 11), and do not provide the range of actual populations in the census tracts. Many clusters do overlap as well, providing cases where is may be appropriate to address multiple techniques of conservation education and management. For example, the previous analysis on income and education that high-income areas see an increase in water use, while high education areas see a decrease in water use. However, there are some areas in Phoenix where low education areas are spatially similar to high-income large family areas. This would mean that these areas may see extremely high water use, and this is in fact supported by a brief look at the LISA maps clustering from both components 1 and 3 and water use (figure 9). In fact the southwest corner of Phoenix there are some census tracts that are in the top quantile of water use in the study area, that load high in component 3 (affluent families) and low in component 1 (large Hispanic families) indicating a mixing of components can lead to high water use. This does not necessarily mean both components describe the same population (although it might) but that they are at least present in the same census tracts.

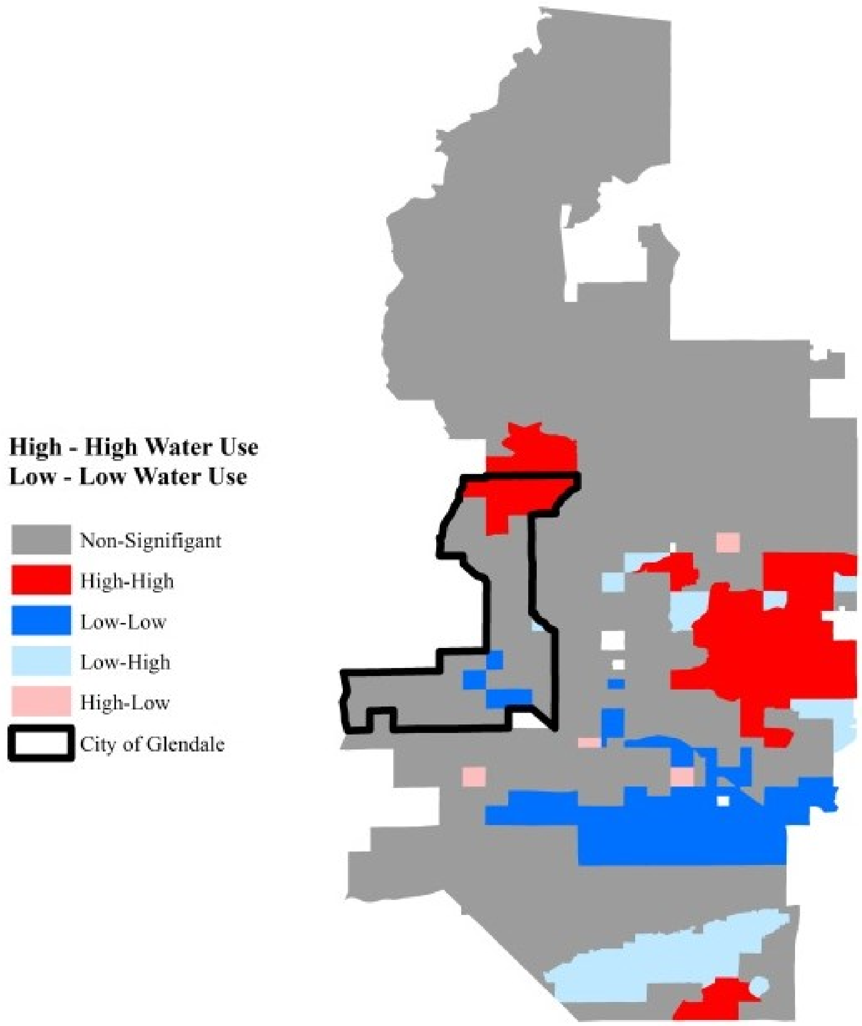

Figure 9: LISA Water Use w/Glendale

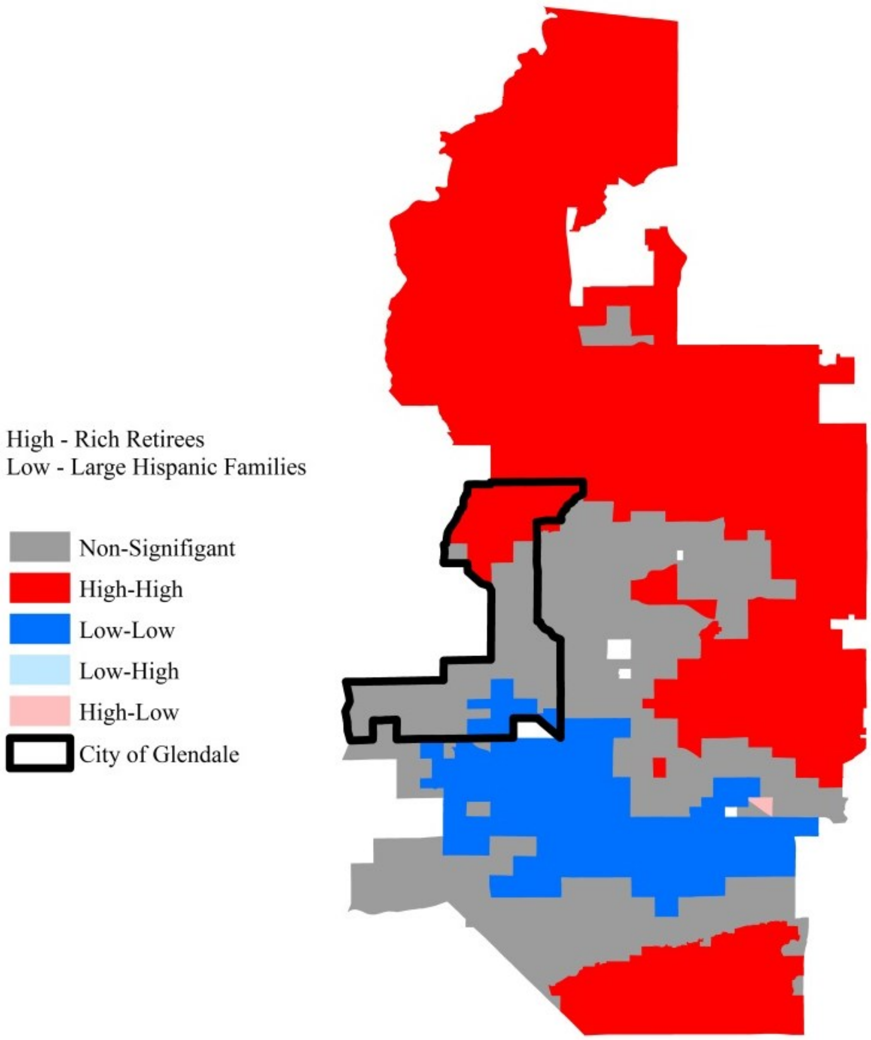

Figure 10: LISA Component 1 w/ Glendale

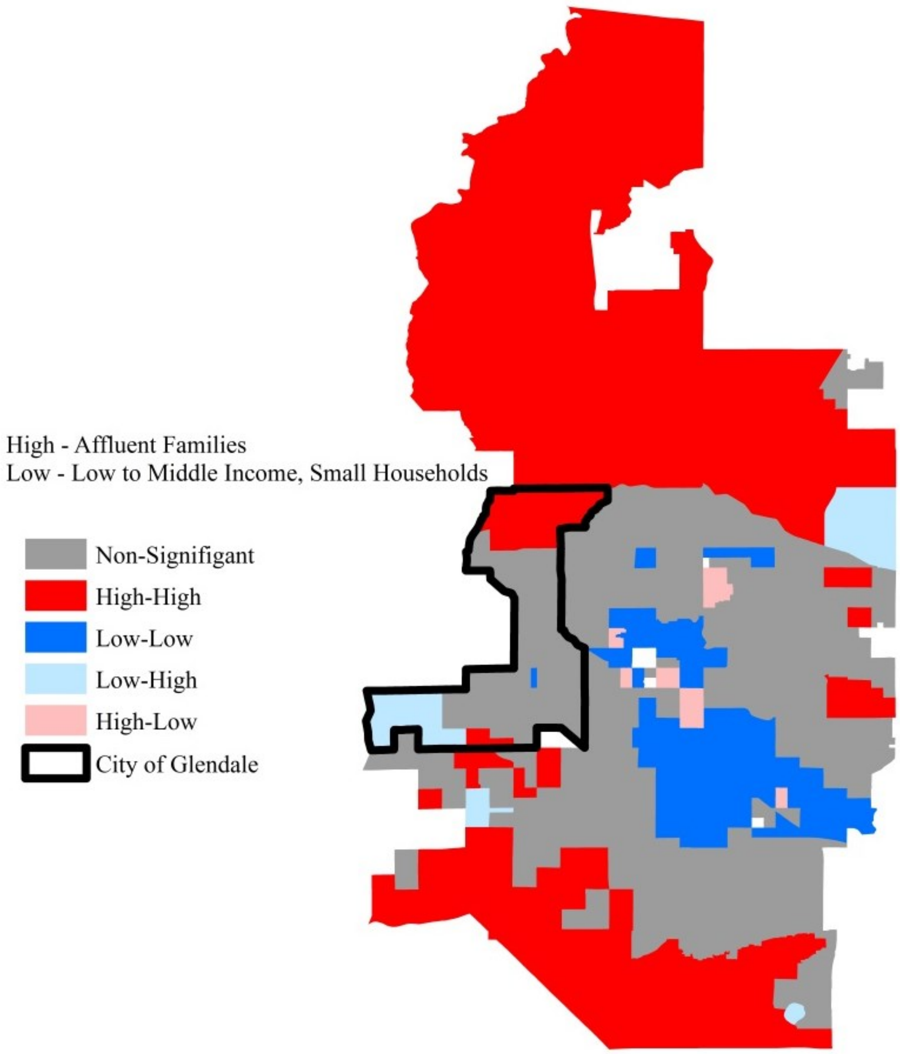

Figure 11: LISA Component 3 w/ Glendale

Urban water management is practiced at multiple government levels, which has implications in water planning. According to House-Peters and Chang (2011), this process is permeating through the study of urban water: "Governance structures exist at multiple scales, from the neighborhood to the city to the region, and can influence water consumption decisions" (House-Peters and Chang 2011). In the Phoenix Metro area there are approximately 33 incorporated communities, and 9 of the communities have over 100,000 inhabitants (US Census Bureau 2010) including both Phoenix and Glendale. While some of this is to be given in an urban area, the large growth of the Phoenix area in a short period of time is what makes this area somewhat unique. Many of these municipalities were incorporated between 1890 and the 1950's. The growth of over 3 million between 1960 and 2010 meant the once separate towns became a conglomerate. One of the biggest logistical concerns of this area is how to provide water to entire population of individuals over a large area in an increasingly intricate infrastructure. To further complicate issues very few water sources are available. Increasing uncertainties over climate change and drought also exist (White, Corley, and White 2008). Historically, water appropriation in Arizona was based on a first come first served basis. This allowed for the incorporation of many different cities and agricultural land in the early 1900's (Keys, Wentz, and Redman 2007). Now, there are over 4 million people and 33 incorporated communities in the Phoenix Metro area (US Census Bureau 2010). This is a complex scenario for water managers.

In terms of water sources, currently the Phoenix Metro area has four main sources: 1) The Colorado River, which supplies water to Arizona through the Central Arizona (CAP) project, 2) surface water from the Salt River, 3) surface water from the Verde River, and 4) groundwater from aquifers. The main problem in efficiency is that all of these sources are managed by different institutions. Colorado River water alone is allocated by many interstate compacts, federal laws, Supreme Court decisions and international treaties. Due to these factors government interplay issues have increased, the consequences of which will be discussed later. Water managers in Arizona must traverse these laws in order to provide water to residents. The Salt River Project (SRP) manages both the Verde and the Salt River water. This is a private company and is the largest single manager of water in the state. Groundwater is managed by the Groundwater Management Act of 1980. It is intended to manage sustainable groundwater access by regulating wells and groundwater use (White, Corley, and White 2008). "Phoenix water managers must consider multiple decision inputs, balance the needs of current and future citizens, protect environmental values, and negotiate complex scientific, technical, and political issues" (White, Corley, and White 2008). There have been unofficial organizations that have been created to help deal with this, but the issue remains. To further complicate the issue, many Native American tribes around the area that are not under federal jurisdiction add to the demand of water. Uncertainties surround their claims, especially to the Colorado River (White, Corley, and White 2008). The goal of water managers is to provide water to their constituents. Every water manger from every town in the area must collaborate with other towns in the area to align their interest in order to provide a more efficient governance system. Along with the collaboration among bordering towns, there is a vertical interplay that these towns must to deal with, including the federal government and federal water laws as well as a state policy board for management, and water managers. The policies and effectiveness of water management may be inefficient when government interplay is high, and the fact that there is a lack of clear governmental structure is a road-block to efficient water use management. Ad-hoc coalitions are often formed for certain goals, but with high levels of regulation from the federal government this could result in even more confusion. Looking at the State of Arizona as a water manager, the same issues arise. Arizona must deal with California, Nevada and a host of other states, the federal government and a variety of local governments, including Native American tribes. Government structures are imbedded within the water use in many ways and the basic understanding of both the spatial and political complexities of how water gets to users may be important.

<u>*3. Separation of Individual Variables in Decision Making*</u>

Those with high-income are often lumped into one analysis but this section suggest that these water users may populate more than one urban water use neighborhood and may in fact warrant a more specific approach based on identified populations. There are a variety of previous studies that have focused on each of the variables in this study, and a variety of others, including many structural variables and specific economic variables to predict water use for future planning or for description of water users (House-Peters and Chang 2011). House-Peters and Chang provide a review of most of the studies on urban water use the variables and techniques used. There have been a variety of techniques that have been used to study different aspects of water use, including time-series analysis, econometric modeling including multivariate analysis. Many previous studies focus on the multiplicity of spatial and temporal scales in studies providing broad pictures and regional based urban water use information. Multiple regression analysis of urban water users started in the 1980's, greatly improving demand forecasting and allowing multiple variables on a large content scale to be used in a singular study. The studies were, however, aspatial. Multivariate analysis continued to be used in terms of pricing structure and continue to be used as such in many studies, particular in terms of temporal pricing and climate models (House-Peters and Chang 2011). In many studies multivariate analysis is used in determining multiple factors' influence (or description) on water use. There is a distinction that needs to be made in terms of using singular variables as broad descriptors of water users. They are helpful determine broad factors associated with water use, and have been shown to be strong in future prediction of water use. This is particularly true when dealing with structural variables (or climate variables), as the presence of many variables can be used to help improve water conservation. But when dealing with socioeconomic variables, especially based on the results of this study, there should be particular attention paid to the variability of these specific variables. Building off of the previous sections, in Phoenix in particular there is distinct spatial pattern for those with income over $150,000. This variable is also associated with two distinct socioeconomic groups: 1) high-income retirees and near retirees, and 2) affluent families. Both socioeconomic groups are similar in many aspects but can be vastly different in others. In dealing with conservation education, the fact that there may be children in the house changes how these neighborhoods view water, make decisions about water, and how they should be approached in terms of water conservation education. Spatially, these neighborhoods do overlap in some areas, but can be defined by their own neighborhoods. Water managers can take this into account, making decisions on a broad level on what how to integrate management and conservation education. Therefore spatially and in the composition of the socioeconomic groups through the PCA there is a distinct divide in this $150,000+ income group. By looking at a specific socioeconomic variable over an entire study area, special attention needs to be paid to the spatial and compositional relations to other variables the study variable has. But the differences between the socioeconomic groups may unique enough to warrant a deeper look at how to approach education and management.

VII. Further Research

Methods that include rates of water use per population and per PCA component would be useful, across man temporal spans. Improvements can also be made in the data reduction methods, particular in the use of geographically weighted regression and geographically-weighted principal components analysis (Demšar et al. 2013). Concerning populations, it would be beneficial to study each of these communities independent of each other. Both component 1 and 3 contained two distinct socioeconomic groups, but they were tied to each other within the bounds of a principal components analysis. Attempting to analyze these groups independent of each other may provide more insight into water use patterns. Analysis should also be undertaken concerning climate variability, traditional high-risk populations, and the populations identified in this study. This is especially important concerning the dwindling water resources and the prolonged droughts that are becoming commonplace in the area. Multi-family residences are an important part of understanding urban water use (Wentz et al. 2013). However, due to availability of data from Glendale and Phoenix, this study did not use water data from multi-family residences. The original goal of this study was to focus on the spatial variations of the individual variables across political boundaries in order to show that water use management should focus on the water use neighborhoods as opposed to focusing solely on arbitrary political spaces. This process may be automated fairly simply either through python or R and streamlining the techniques in order to identify clusters of water users can be useful to water managers in urban areas. Studies should include an understanding of township side-by-side management and the imbedded politics of a top down management from federal to state to local within household water use. The spatial and aspatial implications are broad, but identifying qualitative aspects of this political structure would be helpful for any study area.

Limitations

There were some limitations within the statistical analysis of this study that warrant mention. The data used for the socioeconomic population variables were acquired from the ACS and the 2010 Census. Both are sampled data, leading to populations that may be over or under represented within the independent variables of this study. For example, minorities living in areas that are not highly populated with those of similar race are often underrepresented due to sampling techniques (Hirschman, Alba, and Farley 2000). Water use data from Phoenix and Glendale was collected independently of each other, and the collection of each cannot be guaranteed to be consistent, however each was labeled as 2010 single-family residential water use totals and is assumed to be. Data are single family residential water use, while the socioeconomic population variable totals for each census tract are not limited to single family residence. Most of these are household populations, but there may be differences in population representation in the socioeconomic data and in water use data. Interpretation of coefficients (amount of water use) should be limited as there is no way to directly identify water use populations to total water data. Coefficients cannot be assumed to be accurate in terms of actual water use, because the water use populations represented by the water use data are independent of the population data. Lastly, the Census and ACS are the most efficient ways to acquire large amounts of socioeconomic data. However this data is only collected every 10 years in the case of the Census. The water use data are also only for 2010 total. Any long term trends or short term-anomalies in population counts or in water use will not be reflected in the data.

VIII. <u>Conclusions</u>

Results of statistical analysis show how individual socioeconomic variables are important in explaining water use in the Phoenix Metro area. Principal components analysis allowed for a spatial and statistical analysis of the prominent water use "neighborhoods" in the study area. The variables of highest explanatory power individually included the two highest income brackets, family households, and the retirement age population. In order to include depth into the socioeconomic description of water use, a PCA reduced these variables into four distinct socioeconomic groups. The groups include high-income retirees and near retirees, large Hispanic households, high-in affluent families, and small, low to middle income households.

There is a distinct difference of certain individual socioeconomic variables, most notably in the high-income ($150,000+) range at explaining water use. This variable is quite important to water use, yet it is in two distinct groups, high-income families, and high-income retirees. The separation of the variable into two distinct socioeconomic groups may not have been recognized if a principal components analysis had not been pursued. Second, the principal component regression and socioeconomic population variable regressions results are similar, both providing over 70% variance explanations. With the use of principal components a more socioeconomic descriptors of water users may be presented with no loss in explanatory power. Individual variables can be used to generate rough estimates of which socioeconomic variables may be important in explaining urban water use, but are often more suited for simplistic predictions of water use as opposed to more developed descriptors of water users. By the same process, a multivariate regression provides a more nuanced view of socioeconomic variables related to water use in combination, but still takes into account only the most statistically important variables. High-income loaded highly in component 1 and component 3, showing that there are different types of high-income populations. A principal component analysis allows for a broad and developed understanding of the unique populations of the area. Performing a regression using the components from PCA analysis with water use allows a population to be associated with water use. Methods used in the study (save the multivariate regressions) can then be mapped, traditionally or with the LISA method in order to provide a clear spatial pattern. Differences in socioeconomic components or neighborhoods allow in-depth approach to water planning and conservation education. Components/neighborhoods move across city lines, or overlap one another. This suggests that planning and conservation education needs to be focused on a more micro or macro scale than at city level, perhaps both scales in different areas. The availability of spatial could be of importance for the managers of urban water and the scientists who study it. However, the increasing number of water managers (private and public) and different sources and formats of data will continue to hinder the ability to quickly analyze large amounts of population and water use data efficiently. Finally, spatial statistical analysis cannot be taken without qualitative context. The discussion on the individual socioeconomic neighborhoods showed that there are major differences between how different socioeconomic characteristics can affect conservation practices and water use patterns. However, broad definitions of groups of water users should always be taken in context. Considering the limitations of the study the data are meant to be a tool to water users, and should not replace the knowledge of the population and the area held by water managers. Both the use of independent variables and the PCA can provide information on the spatial distribution of water use, water use populations and the ability for water managers to combine both population analysis and water use analysis into decisions making processes. Spatial statistical methods may be used as exploratory tools for this purpose. An approach to similar studies that combines both the quantitative water use metrics with the qualitative contextual approach of understanding of the way water is used in neighborhoods and in the home will provide the most complete analysis for urban water use in Phoenix.

References

Anselin, Luc, Ibnu Syabri, and Youngihn Kho. 2006. "GeoDa: An Introduction to Spatial Data Analysis." *Geographical Analysis* 38 (1) (January): 5–22. doi:10.1111/j.0016-7363.2005.00671.x.

Bakker, KJ. 2010. *Privatizing Water: Governance Failure and the World's Urban Water Crisis*. *World Watch*.

Balling, R. C., P. Gober, and N. Jones. 2008. "Sensitivity of Residential Water Consumption to Variations in Climate: An Intraurban Analysis of Phoenix, Arizona." *Water Resources Research* 44 (10) (October 2): n/a–n/a. doi:10.1029/2007WR006722.

Balling, Robert C., and Patricia Gober. 2007. "Climate Variability and Residential Water Use in the City of Phoenix, Arizona." *Journal of Applied Meteorology and Climatology* 46 (7) (July): 1130–1137. doi:10.1175/JAM2518.1.

Browne, Alison L, Will Medd, and Ben Anderson. 2012. "Developing Novel Approaches to Tracking Domestic Water Demand Under Uncertainty—A Reflection on the 'Up Scaling' of Social Science Approaches in the United Kingdom." *Water Resources Management* 27 (4) (August 4): 1013–1035. doi:10.1007/s11269-012-0117-y.

Burt, James E., Gerald M. Barber, and David L. Rigby. 2009. *Elementary Statistics for Geographers*. New York: Guilford Press.

Campbell, Heather E., Ryan M. Johnson, and Elizabeth Hunt Larson. 2004. "Prices, Devices, People, or Rules: The Relative Effectiveness of Policy Instruments in Water Conservation1." *Review of Policy Research* 21 (5) (September): 637–662. doi:10.1111/j.1541-1338.2004.00099.x.

Chang, Heejun, G. Hossein Parandvash, and Vivek Shandas. 2010. "Spatial Variations of Single-Family Residential Water Consumption in Portland, Oregon." *Urban Geography* 31 (7) (October 1): 953–972. doi:10.2747/0272-3638.31.7.953.

City of Glendale. 2010. "Monthly Single Family Residential Water Use." (accessed October 2013) [computer file]

City of Phoenix. 2010. "Monthly Single Family Residential Water Use." (accessed October 2013) [CD-ROM]

Demšar, Urška, Paul Harris, Chris Brunsdon, A. Stewart Fotheringham, and Sean McLoone. 2013. "Principal Component Analysis on Spatial Data: An Overview." *Annals of the Association of American Geographers* 103 (1) (January): 106–128. doi:10.1080/00045608.2012.689236.

Domene, Elena, and David Saurí. 2006. "Urbanisation and Water Consumption: Influencing Factors in the Metropolitan Region of Barcelona." *Urban Studies* 43 (9) (August 1): 1605–1623. doi:10.1080/00420980600749969.

Franczyk, Jon, and Heejun Chang. 2008. "Spatial Analysis of Water Use in Oregon, USA, 1985–2005." *Water Resources Management* 23 (4) (July 17): 755–774. doi:10.1007/s11269-008-9298-9.

Gober, Patricia. 2010. "Desert Urbanization and the Challenges of Water Sustainability." *Current Opinion in Environmental Sustainability* 2 (3) (August): 144–150. doi:10.1016/j.cosust.2010.06.006.

Hirschman, Charles, Richard Alba, and Reynolds Farley. 2000. "The Meaning and Measurement of Race in the U.S. Census: Glimpses into the Future." *Demography* 37 (3) (August): 381. doi:10.2307/2648049.

House-Peters, Lily a., and Heejun Chang. 2011. "Urban Water Demand Modeling: Review of Concepts, Methods, and Organizing Principles." *Water Resources Research* 47 (5) (May 28): W05401. doi:10.1029/2010WR009624.

Jolliffe, Ian T. 1982. "A Note on the Use of Principal Components in Regression." *Applied Statistics* 31 (3): 300. doi:10.2307/2348005.

Kenney, Douglas S., Christopher Goemans, Roberta Klein, Jessica Lowrey, and Kevin Reidy. 2008. "Residential Water Demand Management: Lessons from Aurora, Colorado 1." *JAWRA Journal of the American Water Resources Association* 44 (1) (February): 192–207. doi:10.1111/j.1752-1688.2007.00147.x.

Keys, E, EA Wentz, and CL Redman. 2007. "The Spatial Structure of Land Use from 1970–2000 in the Phoenix, Arizona, Metropolitan Area." *The Professional Geographer*.

Keys, Eric, EA Wentz, and CL Redman. 2007. "The Spatial Structure of Land Use from 1970–2000 in the Phoenix, Arizona, Metropolitan Area." *The Professional Geographer* (March 2013): 37–41.

Larson, KL, DC Ibes, and EA Wentz. 2013. "Identifying the Water Conservation Potential of Neighborhoods in Phoenix, Arizona: An Integrated Socio-Spatial Approach." *Geospatial Tools for Urban Water ...*.

Longino Jr., CF. 1994. "Where Retirees Prefer to Live: The Geographical Distribution and Migration Patterns of Retirees." In *Columbia Retirement Handbook*, 405–416.

McGarigal, Kevin, Sam Cushman, and Susan Stafford. 2000. *Multivariate Statistics for Wildlife and Ecology Research*. New York: Springer.

Monk, Abraham. 1994. *The Columbia Retirement Handbook*. New York: Columbia University Press.

Randolph, Bill, and Patrick Troy. 2008. “Attitudes to Conservation and Water Consumption.” *Environmental Science & Policy* 11 (5) (August): 441–455. doi:10.1016/j.envsci.2008.03.003.

Runfola, Daniel Miller, Colin Polsky, Craig Nicolson, Nicholas M. Giner, Robert Gilmore Pontius, Joseph Krahe, and Albert Decatur. 2013. “A Growing Concern? Examining the Influence of Lawn Size on Residential Water Use in Suburban Boston, MA, USA.” *Landscape and Urban Planning* 119 (November): 113–123. doi:10.1016/j.landurbplan.2013.07.006.

Schleich, Joachim, and Thomas Hillenbrand. 2009. “Determinants of Residential Water Demand in Germany.” *Ecological Economics* 68 (6) (April): 1756–1769. doi:10.1016/j.ecolecon.2008.11.012.

Shandas, Vivek, and G. Hossein Parandvash. 2010. “Integrating Urban Form and Demographics in Water-Demand Management: An Empirical Case Study of Portland, Oregon.” *Environment and Planning B: Planning and Design* 37 (1): 112–128. doi:10.1068/b35036.

Shrestha, Milan K., Abigail M. York, Christopher G. Boone, and Sainan Zhang. 2012. “Land Fragmentation due to Rapid Urbanization in the Phoenix Metropolitan Area: Analyzing the Spatiotemporal Patterns and Drivers.” *Applied Geography* 32 (2) (March): 522–531. doi:10.1016/j.apgeog.2011.04.004.

Stonestrom, David A., Bridget R. Scanlon, and Lu Zhang. 2009. “Introduction to Special Section on Impacts of Land Use Change on Water Resources.” *Water Resources Research* 45 (June 17): W00A00. doi:10.1029/2009WR007937.

Stoutenborough, James W., and Arnold Vedlitz. 2014. “Public Attitudes Toward Water Management and Drought in the United States.” *Water Resources Management* 28 (3) (January 10): 697–714. doi:10.1007/s11269-013-0509-7.

Swyngedouw, Erik. 2005. “Dispossessing H2O: The Contested Terrain of Water Privatization.” *Capitalism Nature Socialism (after Jan 1, 2004)* 16 (1) (March 1): 81–98. doi:10.1080/1045575052000335384.

US Census Bureau. 2010. *State and Metropolitan Area Data Book : 2010*.

Wentz, Elizabeth a., and Patricia Gober. 2007. “Determinants of Small-Area Water Consumption for the City of Phoenix, Arizona.” *Water Resources Management* 21 (11) (February 2): 1849–1863. doi:10.1007/s11269-006-9133-0.

Wentz, Elizabeth A., Angela J. Wills, Won Kyung Kim, Soe W. Myint, Patricia Gober, and Robert C. Balling. 2013. “Factors Influencing Water Consumption in Multifamily Housing in Tempe, Arizona.” *The Professional Geographer* (June 26): 130626081911000. doi:10.1080/00330124.2013.805627.

White, Dave D., Elizabeth a. Corley, and Margaret S. White. 2008. “Water Managers’ Perceptions of the Science–Policy Interface in Phoenix, Arizona: Implications for an Emerging Boundary Organization.” *Society & Natural Resources* 21 (3) (February 15): 230–243. doi:10.1080/08941920701329678.