Obesity and Healthy Food Accessibility: Case Study of Minnesota, USA

by

Elliott Wayne Ingram Jr.

A Thesis Presented to the
Faculty of the USC Graduate School
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
Master of Science
(Geographic Information Science and Technology)

December 2019

To my queen, Shoreé, two princesses, Erynn and Sekai, my father, my mother, and my siblings. You all have sacrificed much and guided me through this journey. Thank you and I love you all!

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like give praise to the one above all. He has never left or forsaken me. He gave me the resolve to persevere through this journey. I will forever be grateful for my family, friends, and peers, as they stabilized my foundation and kept my backbone aligned. It was an honor to have Dr. Jennifer Bernstein as my faculty advisor. She was genuine, compassionate, empathetic, sympathetic and understood the haze of the parenting, working, and continuing education gauntlet. I am also appreciative of the constructive criticism received from my juries and committee staff of Dr. Su Jin Lee and Dr. Katsuhiko Oda. The feedback steered me in the right direction to complete this thesis.

List of Abbreviations

ACS	American Community Survey
ADJR$_2$	Adjusted R-Squared (R$_2$)
AICc	Alkaike's Information Criterion
CDC	Centers for Disease Control and Prevention
GIS	Geographic Information System
GISci	Geographic Information Science
JB	Jarque-Bera p-value
K(BP)	Koenker (BP) Statistic (p-value)
MDA	Minnesota Department of Agriculture
MDHFS	Minnesota Department of Health and Family Support
MGC	Minnesota Geospatial Commons
MDH	Minnesota Department of Health
PA	Physical Activity
SA	Global Moran's I p-value
SHIP	Statewide Health Improvement Partnership
SSI	Spatial Sciences Institute
USC	University of Southern California
USCB	United States Census Bureau
VIF	Variance Inflation Factor

Abstract

Since the 1980s, obesity has been categorized as a national and global phenomenon. Although obesity rates in Minnesota have been consistently lower than the nation's and neighboring states' median, the rates have been gradually increasing. The disproportionateness of obesity rates between Minnesota, Minnesota's neighboring states, and the United States suggest that aspects of the Minnesota environment are different. Potential explanatory variables included are linked to economic opportunity, demographics, healthy food availability, and health policies. Utilizing the methodology employed by Shresta et. al (2013), this study expanded it by incorporating more explanatory variables with the intention of building the best model to showcase the impact these variables have on obesity levels and disparities within the study area. Ordinary Least (OLS) and Exploratory Regression analyses were used to assess the spatial relationship between explanatory variables (socio-economics, socio-demographics, and healthy food accessibility) and the dependent variable (obesity levels) over space in Minnesota. The results suggested that the rate of obesity correlates weakly with diabetes, median family income, age, education, and healthy food availability at the county level. The analysis yielded an AICc = 402.068415 and AdjR2 = 0.231832 compared to hypothesis values of AICc = 410.857562 and AdjR2 = 0.162779. The explanatory variables included in the model did not have a strong relationship with the dependent variables in space. Given the relatively low correlations between the predicted relationships, the findings indicate that additional social, cultural, and behavioral factors are required to better explain the prevalence of obesity within Minnesota.

Chapter 1 Introduction

Obesity is an ongoing social and health issue worldwide. The prevalence of the phenomenon is influenced by many social, cultural, and behavioral variables. This study spatially analyzed and modeled the relationship between socio-economics, socio-demographics, accessibility to healthy food options, and their correlations with obesity levels in Minnesota. Socio-economic and demographic explanatory variables included physical inactivity, population size, education attainment, income, employment, race and ethnicity, language spoken, poverty, and access to healthy food. Chapter 1 introduces the problem, presents the study area, and discusses the motivation behind the research. Chapter 2 highlights previous work wherein authors incorporated socio-economic and socio-demographic variables and food accessibility to spatially display the relationships between the factors and how they contribute to obesity. It also articulates gaps in previous research and discusses how these gaps are addressed by this project. Chapter 3 explains the methodology behind the research to be conducted in the thesis. Chapter 4 presents the results of the study. Chapter 5 provides an in-depth summary of the study, discusses its strengths and weaknesses, and provides direction for future research on this topic.

1.1 What is Obesity?

Obesity is an abnormal or excessive fat accrual that threatens an individual's health.

It defines individuals with a body mass index (BMI) above 30 kg/m^2 (Shrestha et al. 2013). Consumption of foods saturated with high levels of sodium, added sugars, and sugar-sweetened beverages contribute to obesity. Individual dietary behaviors such as low intakes of vegetables, fibers, and milk in children, adolescents and adults also contribute to obesity. In the United States, 35.7% of adults and 16.9% of children are considered obese (Chi et. al., 2013). In the United States, approximately 365,000 deaths per year are related to obesity, only second to tobacco (Shrestha et al. 2013). Chronic health conditions like high blood pressure, high cholesterol, diabetes, coronary heart diseases, strokes, cancer, and poor sexual health are directly related to obesity. Obesity has negative economic effects, totaling $117 billion dollars in health care costs in the United States (Shrestha et al. 2013).

1.1.1 Obesity in Minnesota

Minnesotans spend an estimated $2.8 billion each year on obesity related health care costs alone (MDH, 2017). However, Minnesota's obesity rates have been consistently lower than the U.S. median, with exceptions in 2001 and 2002. As of 2017, Minnesota ranks 35th in adult and youth ages 10-17 obesity rates in the nation. 28.4% of adult Minnesotans are obese, up from 16.4% in 2000 and from 10.3% in 1990 (MDH, 2017). Between 2000-2007, the obesity rate in Minnesota increased from 17.4% to 26%. From 2007-2017, the obesity trend slowed from 26% to 28.4% (MDH, 2017). Overall, the state of Minnesota has a lower obesity rate than the U.S. as a whole (See Figure 1).

Minnesota's obesity rate followed the U.S. median from 2001-2007, but the rate was significantly lower compared to neighboring states (Iowa, North Dakota, South Dakota, and Wisconsin) in 2009 and from 2011-2017 (MDH, 2017). In 2008, the Minnesota obesity rate diverged from the U.S. median, as did the obesity rates in Minnesota's neighboring states. These statistics were affected by sample size and demographic compositions of reported surveys (MDH, 2017). Economic opportunity, differences in population demographics, and the availability of healthy food options all are variables contributing to obesity which are different between Minnesota and neighboring states (MDH, 2017).

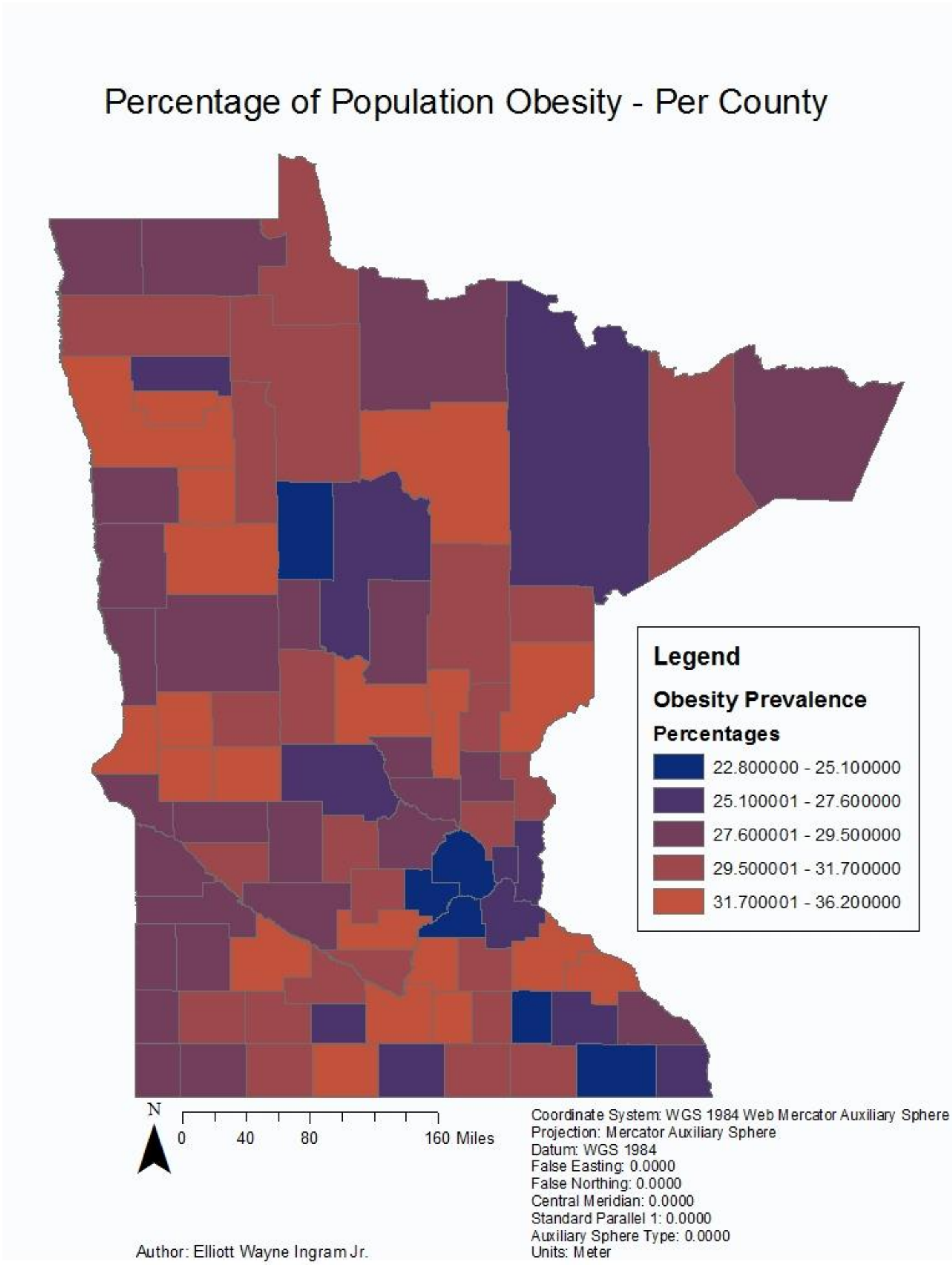


Figure 1 Obesity Prevalence in Minnesota: Percentage of Population Per County

1.2 Study Area

The state of Minnesota (MN), USA is the focus of this research. Of the 48 contiguous of the United States, Minnesota is the northernmost state in the country. Located in the upper Midwest, it lies north central in the United States. Minnesota borders Canada, Iowa, Wisconsin, North Dakota, and South Dakota. Geographically, it is over 400 miles in length and 200-350 miles in wide. Minnesota was ranked as the 12th largest state in the United States (ACS, 2017). Minnesota has 87 counties, with being a major component of scope of this research. Minnesota experienced an incremental population growth of 0.92% in 2018 and accounts for 1.72% of the United States total population (ACS, 2017).

1.3 Socioeconomics and Sociodemographics in Minnesota

Socioeconomics is the social science that studies how economic activity affects and is shaped by social processes (Hellmich, 2015). It analyzes how societies progress, stagnate, or regress because of their local or regional economy, or the global economy (Hellmich, 2015). Sociodemographics are characteristics of a population. Sociodemographic factors including age, race, ethnicity, as well as language and socio-economic variables of income and education all influence health outcomes. It is easy to assume that poverty stricken and low-income communities, for example, are more susceptible to obesity. However, there exist disparities within each study and what constitute contributing variables to obesity (CDC, 2019). Obesity also varies geographically (CDC, 2019).

The total population of Minnesota is 5,303,925 (USCB, 2018). Of the total population, 83.75% are White, 5.95% Black or African American, 4.66% Asian, and the remaining are other races (ACS, 2017). 67% of all Minnesotans are employed, with an unemployment rate of 4%. The median household income of all Minnesotans is $68,400 (ACS, 2017). Minnesota's overall poverty rate was 10.8% in 2017, a slight increase from 10.2% in 2015. However, over 500,000 Minnesotans live below the poverty threshold (ACS, 2017). At 48%, Minnesota ranks 2nd nationally with the percentage of the population age 25-64 earning an associate degree or higher. However, there are major disparities in degree attainment among racial and ethnic population groups over age 25, with only Asian (50%) and white (44%) Minnesotans exceeding the state average. In 2012, 70% of Minnesota adults had at least some college or higher (ACS, 2017).

1.4 Healthy Food Accessibility in Minnesota

1.6 million Minnesotans have low levels of access to healthy food sources (Mattessich, 2016). 235,000 Minnesotans live more than 10 miles from a large grocery store or supermarket (Mattessich, 2016). 49% of Minnesotans report that not having a store nearby that sells healthy food directly impacts what they eat (Mattessich, 2016). Price and distance create barriers to healthy food options. Around 341,000 Minnesotans encounter this barrier (Mattessich, 2016). Approximately 16% of Minnesota's census tracts are considered food deserts, defined as areas with a high proportion of residents who live far from a full-service grocery store and a high proportion of residents who are low-to-moderate income (Mattessich, 2016). Counties in rural Minnesota have a disproportionate number of food deserts relative to their population and geographic area (Mattessich, 2016). It is stated that rural residents, low-income residents, senior residents, and residents of color have relatively low access to healthy food in their communities (Mattessich, 2016). This indicates that thousands of Minnesotans don't have access to healthy food whether it be because of distance, income, or both. These trends continue to contribute to rising obesity rates in Minnesota.

Minnesota ranks seventh-worst in the nation for the share of residents, about one-third of its population, with no grocery options close to their homes (Minnesota Department of Health and Family Support 2012). The saturation of fast food restaurants and lack of farmers' markets, supermarkets, co-ops, and other stores deemed as providing healthy foods are highly noticeable in Minneapolis communities. It is vital for people to have access to places providing healthy foods to help prevent the effects of obesity, including high cholesterol, heart disease, high blood pressure, and additional risks associated with the phenomenon. Proximity to healthy food has the potential to mitigate the obesity epidemic.

1.5 Obesity Disparities Associated with Socio-economics and Socio-demographics in Minnesota

Minnesota is considered one of the healthiest states in the US in terms of obesity trends. However, obesity disproportionately affects many population groups and communities including older adults and seniors, areas of low-income, poverty, low education, US-born Blacks, Hispanics/Latinos, older residents with disabilities, residents with mental illnesses, and female LGBT's (Survey, 2010). Per the 2010 census, 38.5% of US-born adult Blacks and 29.5% adult Hispanic/Latino were obese. 31.4% of high school adults with a high school education, 26.4% of adults with less than a high school education, 25.5% of adults with some college education, and 15.9% of adults with a college education or higher were obese (United States Census Bureau 2010).

The research project looked at the relationship between accessibility to healthy foods, socio-economics, and socio-demographics in the hopes of identifying trends. The successes and failures of this study can provide guidance for researchers wanting to study similar trends within their communities. The project hopes to ultimately help address how all populations can access healthy food to mitigate obesity.

Chapter 2 Background and Literature Review

Areas with greater access to healthy foods tend to have lower obesity rates. However, research conducted to analyze relationships between obesity and healthy food accessibility is complex. Larson et al. (2009) researched and analyzed the presence, nature, and implications of neighborhood differences in access to food using a snowball sampling strategy. They found that national as well as local studies in the United States indicate disparities in socio-economics and demographics and accessibility to healthy resources. The authors emphasized that there are neighborhood disparities in access to food. Larson et al. (2009) suggest that additional research is required to address limitations of current studies promote better healthy food accessibility.

Morland et al. (2002) examined the distribution of food stores and food service locations, sorting each by neighborhood wealth and segregation. The names and addresses of places to buy food in Mississippi, North Carolina, Maryland, and Minnesota were obtained from their respective state Departments of Health and Agriculture. The addresses were then geocoded to census tracts. Median home values were used to estimate neighborhood wealth, while the proportion of black residents was used to measure neighborhood racial segregation. Their study showed that there are four times more supermarkets established in predominantly white communities compared to predominantly black communities. Without access to supermarkets and healthier food options, which offer a wide variety of foods at lower prices, impoverished and minority communities may not have equal access to the array of healthy food choices available to predominantly white and/or wealthy communities.

Boone-Heinonen et al. (2011) conducted a study where they modeled fast food consumption, diet quality, and adherence to fruit and vegetable recommendations as a function of fast food chain, supermarket, or grocery store availability over fixed distances. Their models took into consideration gender, individual sociodemographic characteristics, and community poverty, and tested for interaction by individual-level income. The authors concluded that fast food consumption was directly proportional to fast food availability among low income individuals. However, greater supermarket accessibility was not related to diet quality and fruit and vegetable consumption. Correlations between grocery store availability and individual diets showed mixed results.

Bressie (2016) wrote a thesis analyzing spatial patterns of food accessibility in Lane County, Oregon. The goal was to quantify food retail dispersion in the study area of Lane County, Oregon, in the context of proximity, affordability, diversity (types of food venues), perception, food supply (availability), and socio-economics. The authors' methodology was composed of four steps: (1) food store classification, (2) measurement calculations, (3) aggregation of areal units, and (4) statistical analysis (Bressie 2016). Using Esri's Network Analyst to measure residential proximity to five different food store types over a road network, the study showed that deprived and minority-dense communities in Lane County, Oregon had better access to healthy food sources (Bressie, 2016). The results of this study eliminate stereotypical assumptions of urban and rural food environments and that evaluations of these areas' food environments should be conducted separately.

2.1 GIS-based Analysis of Obesity

One way to analyze obesity trends is to use GIS to better understand the role of social and economic factors. Specifically, "spatially-varying coefficient models such as OLS and GWR have become statistical methods for identifying local variations in relationships between outcome and explanatory variables" (Wen et al. 2010, 263). These complex models benefit researchers who seek to identify spatial variations in relationships.

In Pennsylvania USA, it was found that obesity rates were impacted by many factors, including physical activity, diabetes, and average distance to the nearest healthy food source. Shrestha et. al (2013) conducted a study using Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) to spatially analyze the relationship between socio-economic and physical health in the region. The researchers' goal was to better understand and regulate obesity trends, incorporating exploratory variables including diabetes, physical inactivity, and average distance to healthy food stores. The results of OLS and GWR analyses were compared to determine which method produced the best model to analyze the relationship between socio-economics and obesity rates. It was concluded that GWR generated the best results. Because of only three explanatory variables were used in their analysis, the results had low levels of variance, indicating that additional factors were needed to better explain the distribution of obesity in Pennsylvania. This analysis influenced the approach of the study conducted here, suggesting more explanatory variables should be incorporated to better understand obesity trends in the study area.

Wen et al. (2010) performed a study analyzing 29,273 working adults aged 21-65 years of age, in which they used GWR to inspect geographical variations in the relationship between poverty and obesity. The study revealed geographical inequalities in poverty and that poverty was a key contributor of obesity in Taiwan. Results from the study concluded that poverty and obesity were prominent in less developed areas and that poverty and obesity were locally variable. The impact of poverty on obesity was shown to be locally specific. Variables such as community low income and deprivation increased the prevalence of obesity.

Obesity and socio-environmental variables have been shown to be correlated. Between the three primary socio-economic factors of employment, education, and income, Chalkais et al.

(2013) revealed that education was the most significant indicator of increased obesity rates in Athens, Greece, among 18,296 children 8-9 years if age. Using GWR, Chalkais et al. (2013) concluded that low educational level, high population density, low family income, and green space availability constituted an “obesogenic” environment Chalkais et al. (2013). Although findings by Chalkais et al. (2013) displayed a significant relationship between childhood obesity and socio-economic heterogeneity, further research was needed to understand how socio-economics and environmental factors interact with one another to better understand the obesity epidemic. Chalkais et al. (2013) ultimately called for preventative tactics to combat childhood obesity including changes in diet and physical inactivity.

In a similar study, Drewnowski et al. (2014) linked low socio-economic status to high obesity rates in both Paris and Seattle, despite differences in urban form, food environments, and health care systems. The objective of the study was to compare the relationship between the food environment at the individual level, socio-economic status, and obesity in Paris and Seattle. The researchers collected sociodemographic data, geocoded home addresses and food source locations, and calculated the distance between home and supermarkets. A Modified Poisson regression model was used to test the association between socio-economic status, food environmental variables, and obesity. Results of the study concluded that distance to supermarkets did not have a direct link to obesity; however, low income and education, coupled with low property values and shopping at lower cost stores were directly correlated with high obesity rates.

Obesity and other chronic conditions linked with low levels of physical activity (PA) are associated with deprivation of accessibility to recreational physical activities (Ferguson et. al 2013). Ferguson et al. (2013) used GIS car and bus networks in Scotland to determine the number of PA facilities accessible within travel times of 10, 20, and 30 minutes. The accessibility by car to recreational physical activity facilities greatly exceeded that by bus (Ferguson et. al 2013). Low income communities were deprived of access to facilities that offer recreational activities (Ferguson et. al 2013). It was found that access to physical activity facilities by car was much more significant for the most affluent quintiles of area-based income deprivation than for most affluent quintiles in small towns and rural areas. Facilities were much less accessible compared to bus travel for the most affluent quintile than for other quintiles in urban areas and small towns. The most disadvantaged groups were those without access to a car in rural areas (Ferguson et. al 2013).

Low accessibility to healthy foods and greater access to unhealthy foods are variables in dietary habits leading to obesity. Cubbin et al. (2012) found that neighborhoods that have experienced long-term poverty have the greatest access to both healthy and unhealthy food sources compared to more economically advanced neighborhoods in Alameda County, California. This is counter to stereotypical assumptions that minorities and urban areas have less access to healthy food sources. Blacks and Latino neighborhoods had the greatest access to healthy food sources. The results of their study suggested that spatial relationships between sociodemographic characteristics and healthy food accessibility at the community level depends on place and level of urbanization (Cubbin et al., 2013).

The suburbanization of food retailers in North America and United Kingdom have contributed to urban food deserts (Larsen and Gilliland, 2008). Larsen and Gilliland (2008) used GIS and multiple network analyses were implemented to assess supermarket accessibility in relation to location, socio-economic characteristics, and access to public transit. They found that residents in urban communities with low economic status have the lowest levels of access to supermarkets and that spatial inequality have increased.

Obesity continues to rise and will grow with the increase in of obesity among younger people (Daniel et al. 2009). This trend is due to the consumption of high-dense energy food, reduced energy expenditure, and failure to meet daily fruit and vegetable intake. Daniel et al. (2009) looked at the density of fast food outlets and stores selling fruits and vegetables. Socio-demographic predictors including income, household structure, language, education, and urban form measures (road and highway densities) were assigned. A regression analysis showed that socio-demographic and urban form measures accounted for 60% and 73% of the variance densities of fast food outlets and stores selling fruits and vegetables, respectively (Daniel et al. 2009). Fast food outlets were more prevalent in areas with full-time students and households without fluent speakers of French or English. Stores selling fruits and vegetables were more prevalent in communities with high proportions of single-status residents and university-educated residents.

As this literature review shows, obesity is a complex phenomenon that is influenced by multiple factors. Previous studies provided the blueprint on the best approach to analyzing the prevalence of obesity, including socio-economics, socio-demographics, and accessibility to healthy food outlets. Based on past studies, this study anticipated that physical inactivity, median family income, poverty prevalence, unemployment, and healthy food source density would be the greatest contributors obesity in Minnesota. 14 explanatory variables, including the five variables listed above were included in regression analyses to test which factors most contributed to obesity and constituted the best regression model. This builds on past research by determining which variables, of which there are many, are the strongest predictors of obesity. Ideally, the results of the study will be used to mitigate obesity in the future.

Chapter 3 Methodology

This chapter explains the process of data acquisition, data preparation, and the regression analyses used in this study. The approach to this analysis followed the study conducted by Shrestha et al. (2013). First, socio-economic and socio-demographic variables were acquired and aggregated from non-spatial data in census and county databases. This data was inserted into an Excel spreadsheet for analyses. The non-spatial data was later formatted and aggregated to the state level as polygons, with each polygon representing a single county in ArcMap. In ArcMap, a projected coordinate system was established to best display the data across the study area (NAD 1983 (2011) StatePlane Minnesota Central FIPS 2202 (US feet)). Data on businesses serving healthy food and grocery stores were plotted as vector points in ArcGIS software for reference. In the study, healthy food accessibility was generated by dividing the number of healthy food sources by the area of each county per square mile. This gave the healthy food density per county. Population density was calculated to identify possible correlations between population per square mile and the prevalence of obesity. After identifying the explanatory variables, OLS analyses were conducted to test the correlation of hypothesized explanatory variables against obesity and a list of 14 explanatory variables and their correlation with obesity using exploratory regression. The analyses sought to test the statistical significance of each explanatory variable on obesity in Minnesota. Figure 2 shows a workflow of the methodology.

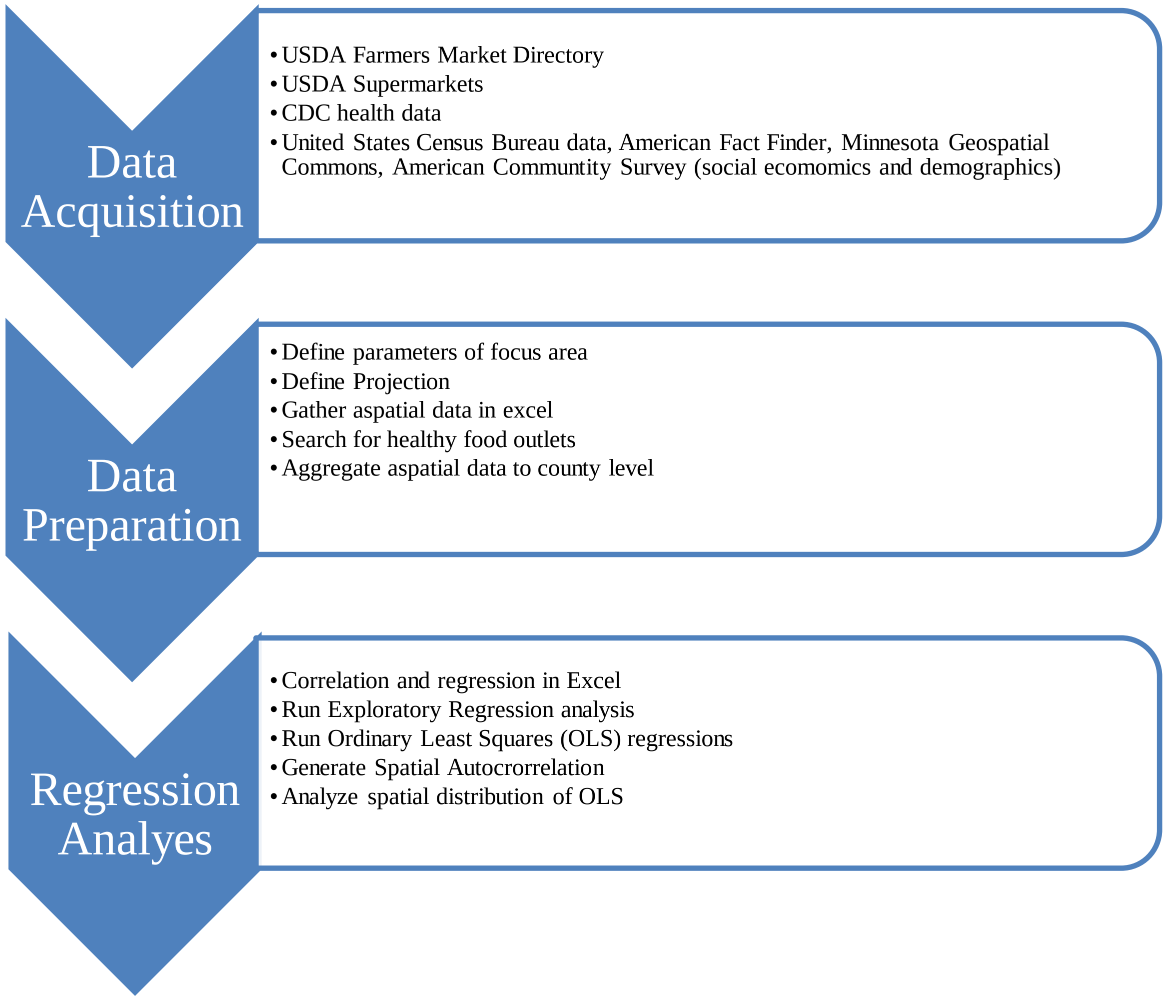


Figure 2 Summary of Workflow

3.1 Data Acquisition

Data was acquired from the US Census Bureau, Minnesota Department of Health databases, Minnesota GIS databases, and the CDC. Data was used to map how which variables negatively or positively impacted the prevalence of obesity the most. Socio-economic, socio-demographic variables, and healthy food accessibility was needed to be thoroughly investigated to help mitigate the obesity epidemic in Minnesota.

Table 1 Data Types and Sources

Category	Factors	Data Source	Geographic Scale	Data Type
Administrative boundaries	Census Tract, County, Municipality, State	Tiger Line Data, Minnesota Geospatial Commons, Explore Minnesota	State/County	Vector polygons
Access to healthy food	Healthy food businesses/stores	Exploring Food Environments (ESRI), ArcGIS Online, MetroGIS DataFinder, Minnesota Department of Agriculture	State/County	Vector points and polygons
Health	Obesity, diabetes, physical inactivity	Minnesota Public Health Data, Center for Disease Control and Prevention (CDC)	State/County	Vector polygons
Socio-economics and Socio-demographics	Poverty, income, population, age, gender, ethnicity and race, employment education attainment, language spoken,	United States Census Bureau, American Fact Finder, American Community Survey, Minnesota Geospatial Commons, Minnesota Geographic Data Clearinghouse Data, MetroGIS DataFinder	State/County	Vector polygons

3.2 Data Preparation

3.2.1 Data Aggregation

Health, socio-economic, and sociodemographic variables were aggregated in an Excel data sheet in tabular format and appended as aspatial data into county polygons (see Appendix A). Most of the aspatial data was aggregated to the county scale and by percentage of the total population per county. Population density and healthy food source density were calculated by taking the quotient of total population by the area of the county and quotient of the number of healthy food sources by the area of the county, respectively.

3.2.2 Healthy Food Sources

Healthy food sources are defined as businesses that offer foods that provide nutrients needed to sustain health and provide energy (Richardson, 2010). These outlets sell health foods, organic foods, local produce, and nutritional supplements. Generally, supermarkets, grocery stores, food co-ops are grouped as one entity and farmer's markets fall into the category of healthy food sources (Richardson, 2010). They offer an array of nutritious foods on the food pyramid that are beneficial to healthy living. Table 2 lists the number of healthy food sources in Minnesota. Supermarket, grocery store, and food co-op data was attained from the Supermarket Access Map in ArcGIS Online (Richardson, 2010). Data pertaining to the number of farmer's markets in Minnesota was gathered from the Minnesota Department of Agriculture Directory (2019).

Table 2 Healthy Food Sources in Minnesota

Store Types	**Original Counts**
Supermarkets, Grocery Stores, Food Co-ops	1332
Farmer's Markets	196
Totals	**1528**

3.2.3 Correlation of the Dependent Variable and Explanatory Variables

Correlation was used to test the relationship between the potential explanatory variables and the dependent variable. In Microsoft Excel, the CORREL function was used to find the correlation between two variables. A correlation coefficient of +1 indicates a perfect positive correlation, which means that as x increases, variable y increases and while variable x decreases, variable y decreases. In contrast, a correlation of -1 indicates a perfect negative correlation, as variable x increases, variable z decreases and as variable x decreases, variable z increases. When visualized, the x-axis represents the explanatory variables and the y-axis represents the dependent variable (in this case obesity prevalence).

3.3 Regression Analysis ArcGIS and ArcMap

3.3.1 Ordinary Least Squares

The relationship between the dependent variable and explanatory variables was examined on a county-wide basis with a cross-sectional analysis by using Ordinary Least Squares (OLS). Multicollinearity refers to the state of very high inter-correlations or inter-associations among independent variables (Shrestha et al. 2013). The OLS model assigns an equation to all the features being analyzed and predicted. OLS's purpose is to test the significance of explanatory variables and potential multicollinearity amongst the variables. Using Variation Inflation Factor (VIF) and Variable Significance (VS) values, multicollinearity addressed by removing variables with a VIF over 7.5. OLS was run again to mitigate multicollinearity.

Five potential explanatory variables of physical inactivity, median family income, poverty prevalence, unemployment, and heathy food source density were selected for regression analysis. The purpose of OLS was provide a global model of the dependent variable, obesity prevalence, and try to predict the phenomenon by creating a regression equation to represent the process. In ArcMap, the OLS tool prompted a pop-up screen, in which an Input Feature Class with Unique ID Field was specified. Obesity prevalence was acknowledged as the Dependent Variables and the five explanatory variables was listed in the Explanatory Variables section. The OLS analysis was run and generated an output feature class.

3.3.2 Exploratory Regression Analysis

Exploratory Regression Analysis was used to evaluate all possible combination of the input variables, searching for OLS models that best explained the dependent variable within guidelines of criteria specified. The Exploratory Regression tool mined data for all possible combinations of explanatory variables to see which models passed all the OLS diagnostics. The minimum and maximum number of explanatory variables in each model was set at 1 and 5 respectively, with default threshold criteria for Adjusted R_2, coefficient p-values, VIF values, Jarque-Bera values, and spatial autocorrelation p-values. The Exploratory Regression analysis ran OLS on every possible combination of explanatory variable listed in Table 3, with at least the minimum number of explanatory variables and no more than the maximum number of explanatory variables specified. The dependent variable was obesity prevalence. Each model was assessed against the default threshold criteria. If the model exceeded the specified Adjusted R_2 threshold, had coefficient p-values for all explanatory variables less than the threshold, had coefficient VIF values for all explanatory variables less than the threshold, and returned a Jarque-Bera p-value larger than anticipated, the Spatial Autocorrelation tool was run on the model's residuals. If the spatial autocorrelation p-value was larger than the specification in the search criteria, the model was deemed to have passed. A properly specified OLS model is validated with statistically significant explanatory variables, with small VIF values indicating non-redundancy. The coefficients reflect the strength of the relationship between the explanatory variables and the dependent variable. Normally distributed residuals indicated a non-biased model, namely a Jarque-Bera value that is not statistically significant. A properly specified OLS model also has a random distribution of over and under predictions.

Table 3 Explanatory Variables (Units)

Explanatory Variables (Per County)
Obesity Prevalence (%)
Physical Inactivity (%)
Total Population (#)
Median Family Income ($)
Poverty Prevalence (%)
Language Other Than English in Household (%)
Foreign Born (%)
Unemployment (%)
Population 25 and Over with Associates Degree (%)
Population 25 and Over with Bachelor's Degree (%)
Population 25 and Over with Master's or Professional Degree (%)
Source Count (Accumulation of Supermarkets and Farmer's Markets) (#)
Population Density (#)
Source Density (#)

3.3.3 Spatial Autocorrelation

Spatial autocorrelation measures the correlation between variable in space. Spatial autocorrelation, also known as the clustering of residuals, is a symptom of misspecification. This occurs when key explanatory variables are missing. Moran's I was utilized to test for spatial autocorrelation and verify that systematic patterns and biases did exist in the model.

After running an exploratory regression analysis of the 14 explanatory variables in relation to the dependent variable, a spatial autocorrelation analysis was conducted. Spatial autocorrelation indicates whether there was clustering or dispersion in the correlation between the explanatory variables and dependent variable. Spatial autocorrelation confirmed if there's a significant statistical pattern in the data. A positive Moran's I indicates that the data was clustered. In contrast, a negative Moran's I implies that the data was dispersed. The Spatial Autocorrelation tool in the Spatial Statistics toolbox of ArcMap used Global Moran's I function to compute the z-score value of the correlation between the explanatory variable and the dependent variable. The z-score value helped determine whether Moran's I should be classified as positive, negative, or no spatial autocorrelation. Spatial autocorrelation displayed how the explanatory variables and the dependent variable spatial relationship geographically.

Chapter 4 Results

The results of the analysis are highlighted in this chapter. To test the correlations between the explanatory variables and the dependent variable, explanatory variables that were thought to contribute most to obesity prevalence were examined. Five explanatory variables were analyzed in OLS to correlate with obesity prevalence. 14 explanatory variables were analyzed in an exploratory regression analysis to find the best model that showcased the spatial relationship between the explanatory variables and the dependent variable. First, each explanatory variable was correlated with obesity prevalence in Excel to visually show the statistical significance between the explanatory variables and the dependent variable before running the regression analysis.

4.1 Excel Correlations of Explanatory Variables and Dependent Variable

Overall, the relationships between the explanatory variables and the dependent variables yielded low R-Squared values and poor regression line fits, though some relationships of course were stronger than others.

In Figure 3, the R-Squared value equaling 0.0825 of Physical Inactivity vs. Obesity Prevalence indicated a very poor regression line fit. Eight percent of the variation in obesity prevalence is explained by the independent variable physical inactivity.

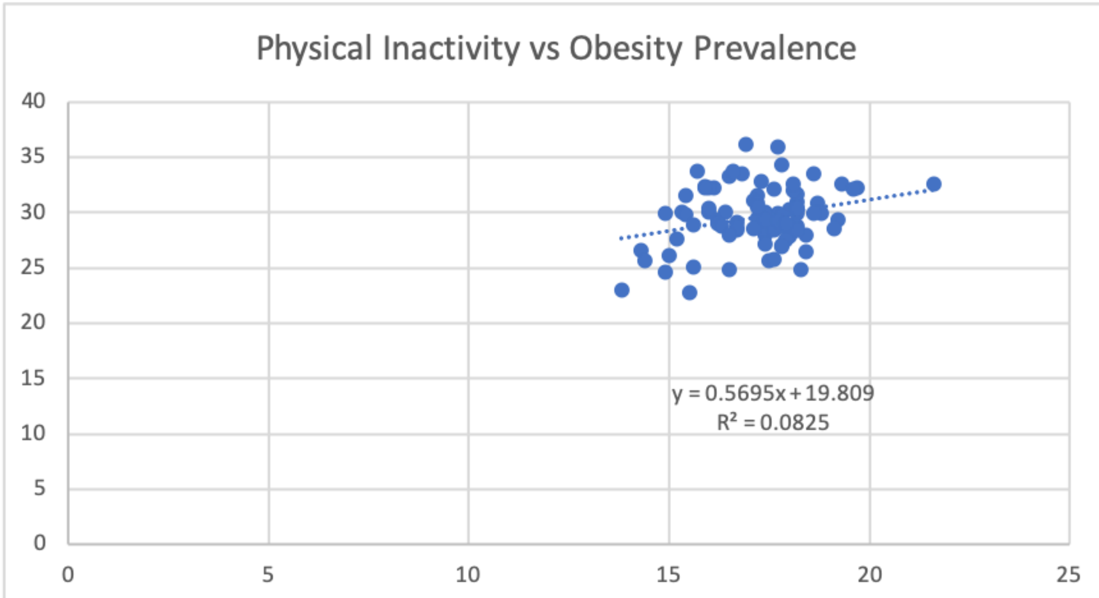


Figure 3 Physical Inactivity vs. Obesity Prevalence Correlation

In Figure 4, the R-Squared value equaling 0.1123 of Diabetes Prevalence vs. Obesity Prevalence indicated a very poor regression line fit. 11% of the variation in obesity prevalence is explained by the independent variable diabetes prevalence.

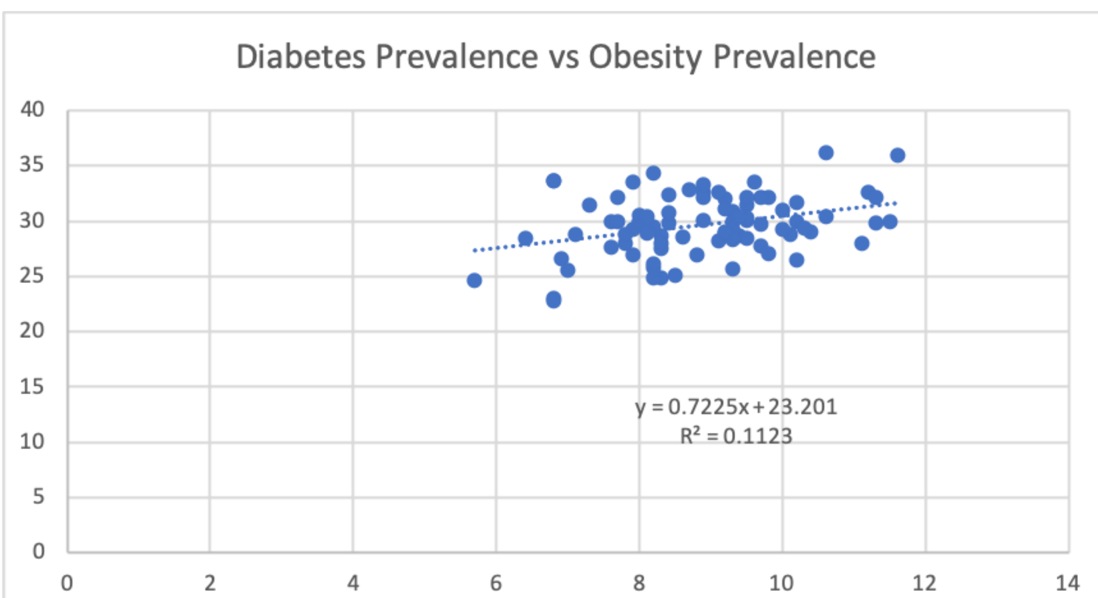


Figure 4 Diabetes Prevalence vs. Obesity Prevalence correlation

In Figure 5, the R-Squared value equaling 0.1541 of Total Population vs. Obesity prevalence indicated a very poor regression line fit. 15% of the variation in obesity prevalence is explained by the independent variable total population.

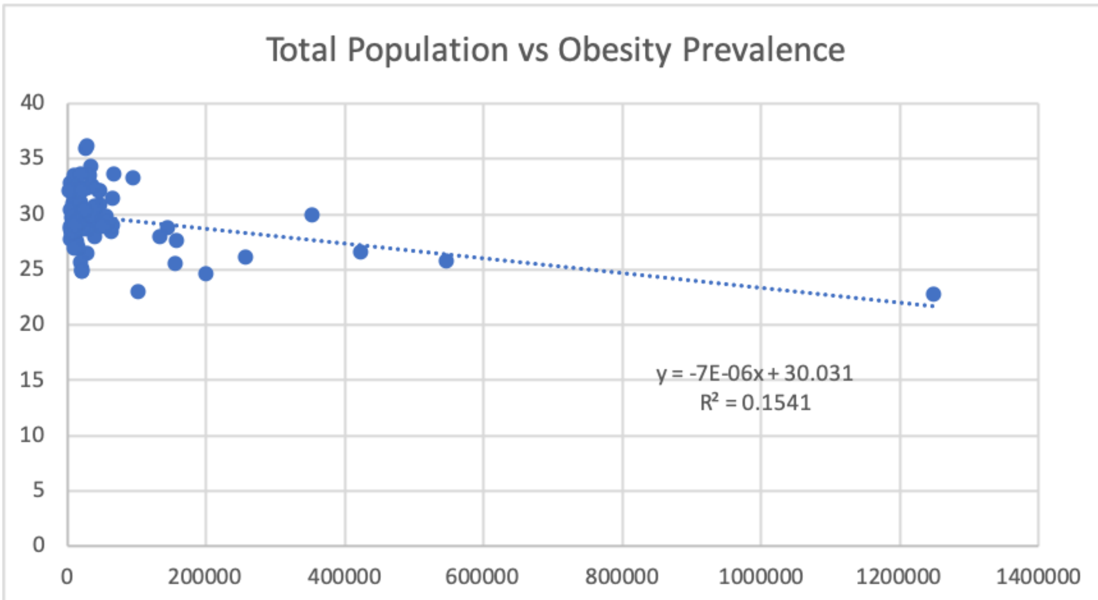


Figure 5 Total Population vs. Obesity Prevalence Correlation

In Figure 6, the R-Squared value equaling 0.0182 of Median Family Income vs. Obesity Prevalence indicated a very poor regression line fit. 2% of the variation in obesity prevalence is explained by the independent variable median family income.

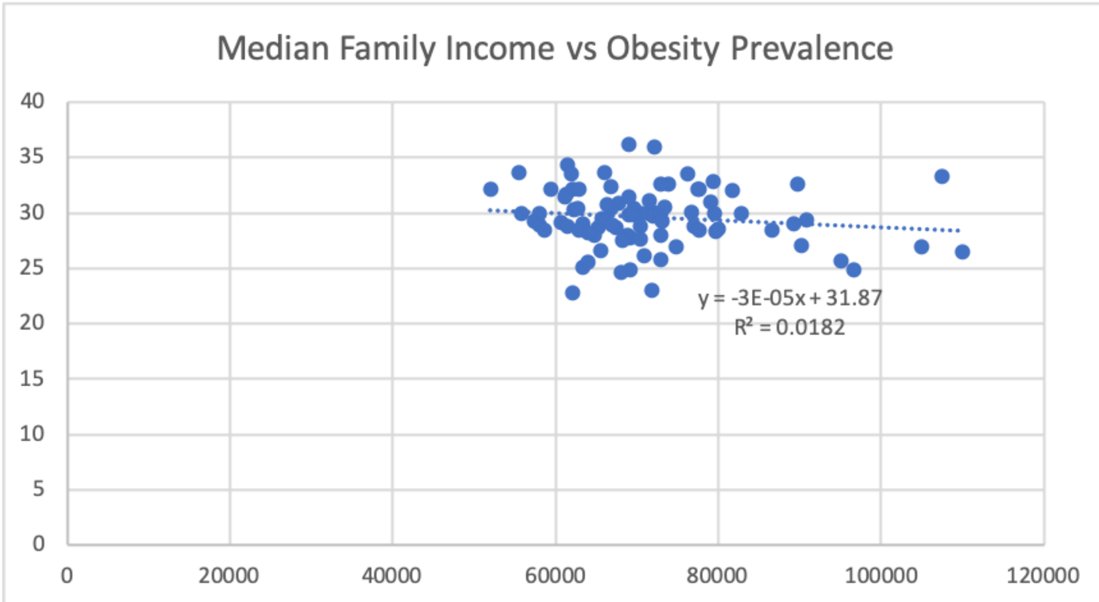


Figure 6 Median Family Income vs. Obesity Prevalence Correlation

In Figure 7, the R-Squared value equaling 0.0133 of Poverty Prevalence vs. Obesity Prevalence indicated a very poor regression line fit. 1% of the variation in obesity prevalence is explained by the independent variable poverty prevalence.

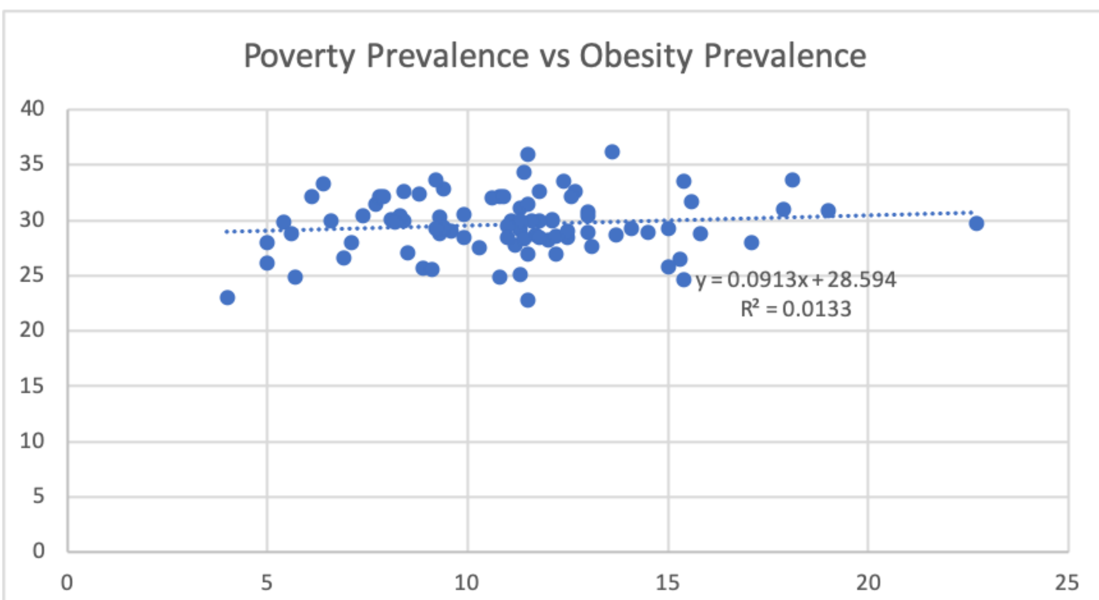


Figure 7 Poverty Prevalence vs. Obesity Prevalence Correlation

In Figure 8, the R-Squared value equaling 0.0676 of Language Other Than English vs. Obesity Prevalence indicated a very poor regression line fit. 7% of the variation in obesity prevalence is explained by the independent variable language other than English.

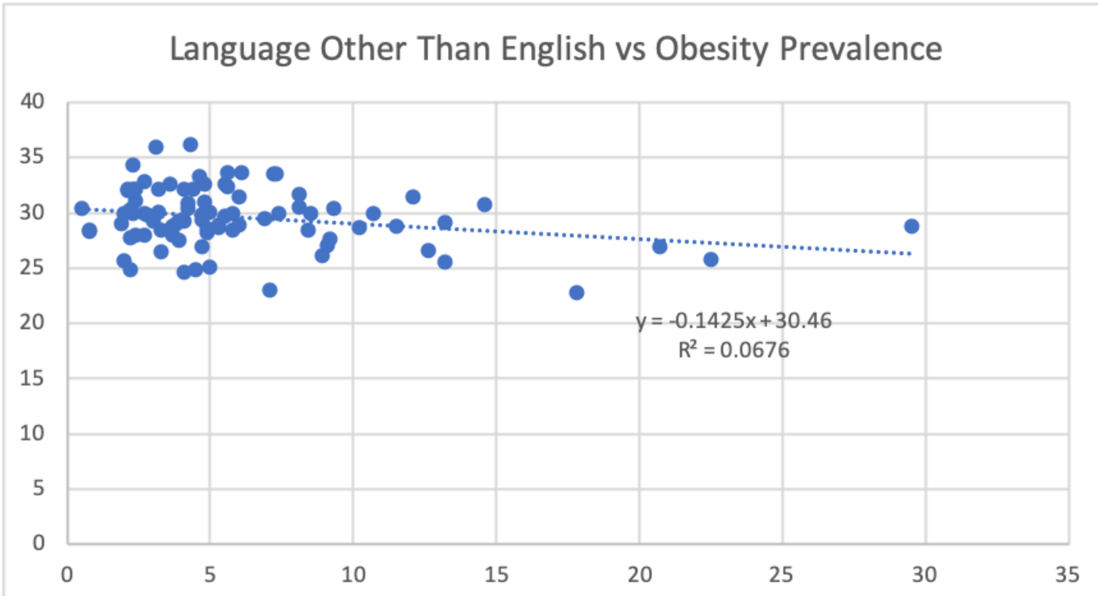


Figure 8 Language Other Than English vs. Obesity Prevalence Correlation

In Figure 9, the R-Squared value equaling 0.0987 of foreign-born vs obesity prevalence indicated a very poor regression line fit. 10% of the variation in obesity prevalence is explained by the independent variable foreign born.

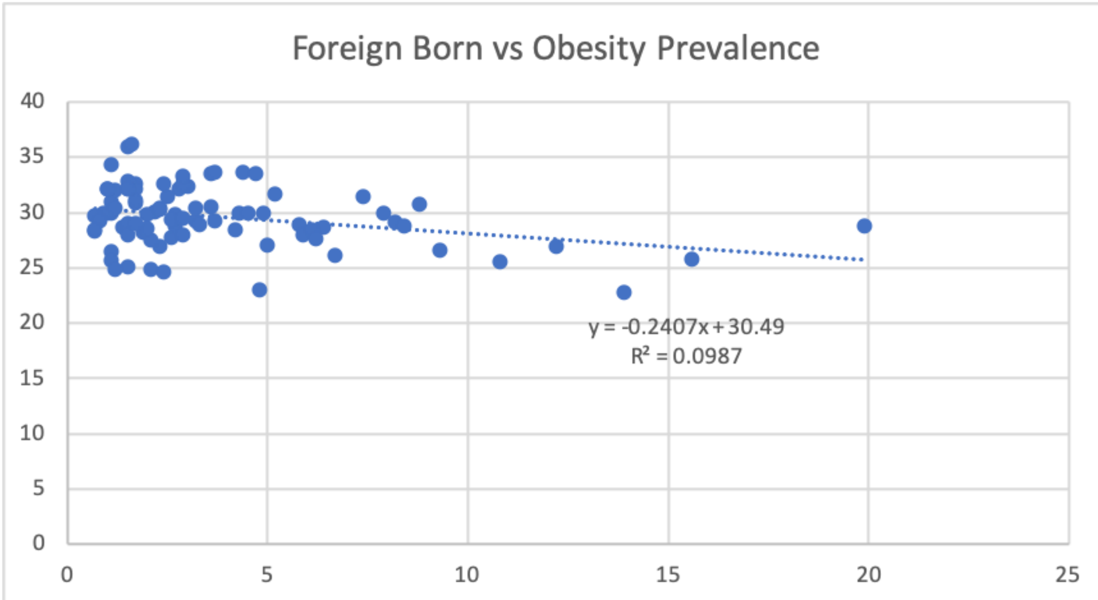


Figure 9 Foreign Born vs. Obesity Prevalence Correlation

In Figure 10, the R-Squared value equaling 0.0543 of Unemployment vs. Obesity Prevalence indicated a very poor regression line fit. 5% of the variation in obesity prevalence is explained by the independent variable unemployment.

Figure 10 Unemployment vs. Obesity Prevalence Correlation

In Figure 11, the R-Squared value equaling 0.1785 of Bachelor's Degree vs. Obesity Prevalence indicated a very poor regression line fit. 17% of the variation in obesity prevalence is explained by the independent variable bachelor's degree.

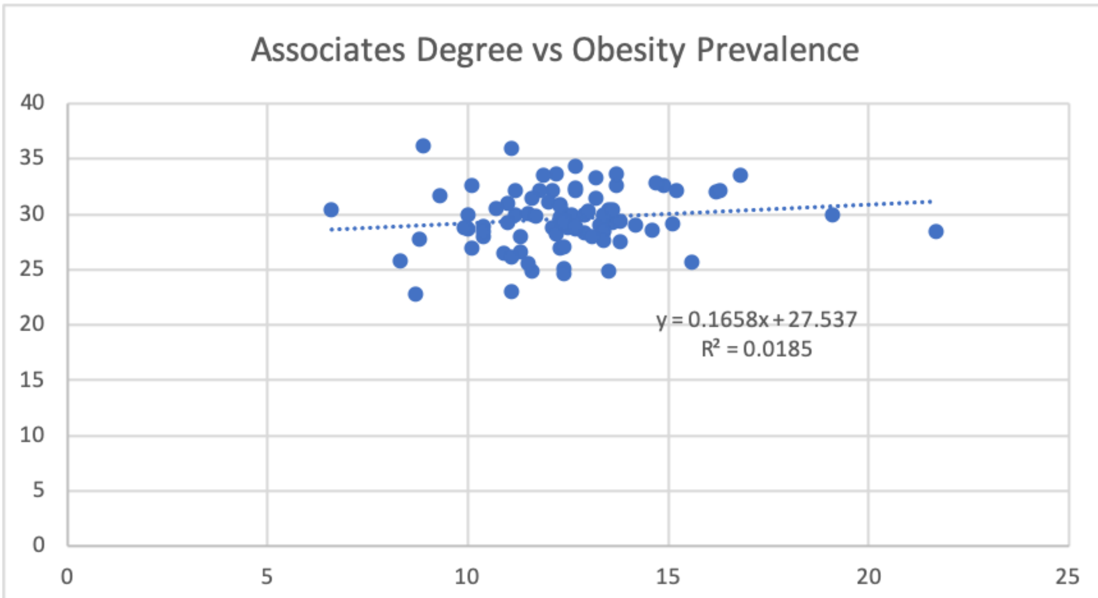


Figure 11 Associates Degree vs. Obesity Prevalence Correlation

In Figure 12, the R-Squared value equaling 0.0185 of Associates Degree vs. Obesity Prevalence indicated a very poor regression line fit. 2% of the variation in obesity prevalence is explained by the independent variable associate degree.

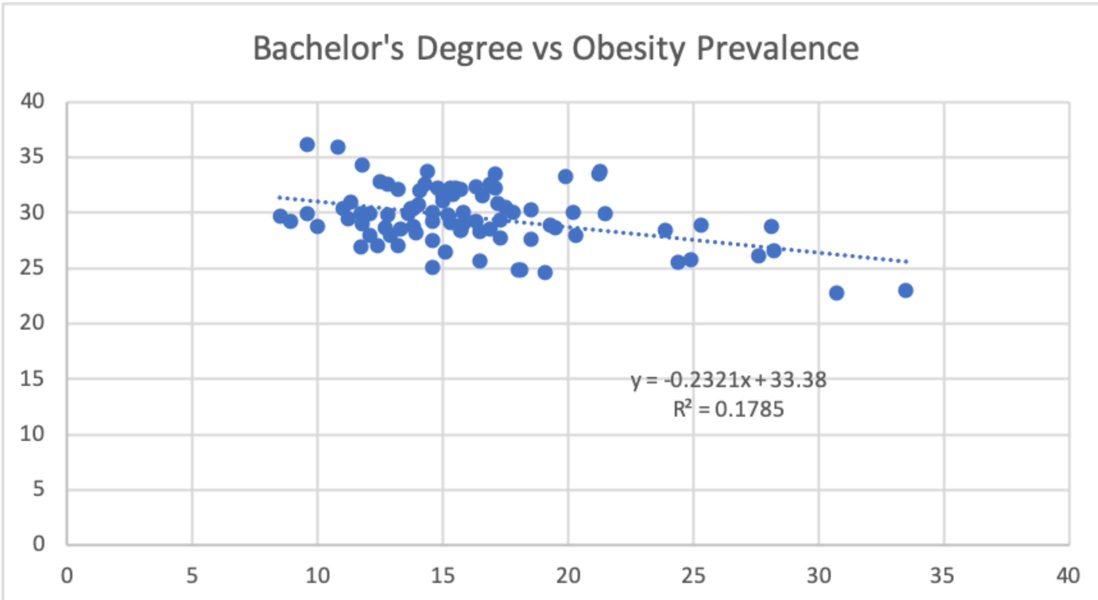


Figure 12 Bachelor's Degree vs. Obesity Prevalence Correlation

In Figure 13, the R-Squared value equaling 0.1461 of Professional/Master's Degree vs. Obesity Prevalence indicated a very poor regression line fit. 15% of the variation in obesity prevalence is explained by the independent variable professional/master's degree.

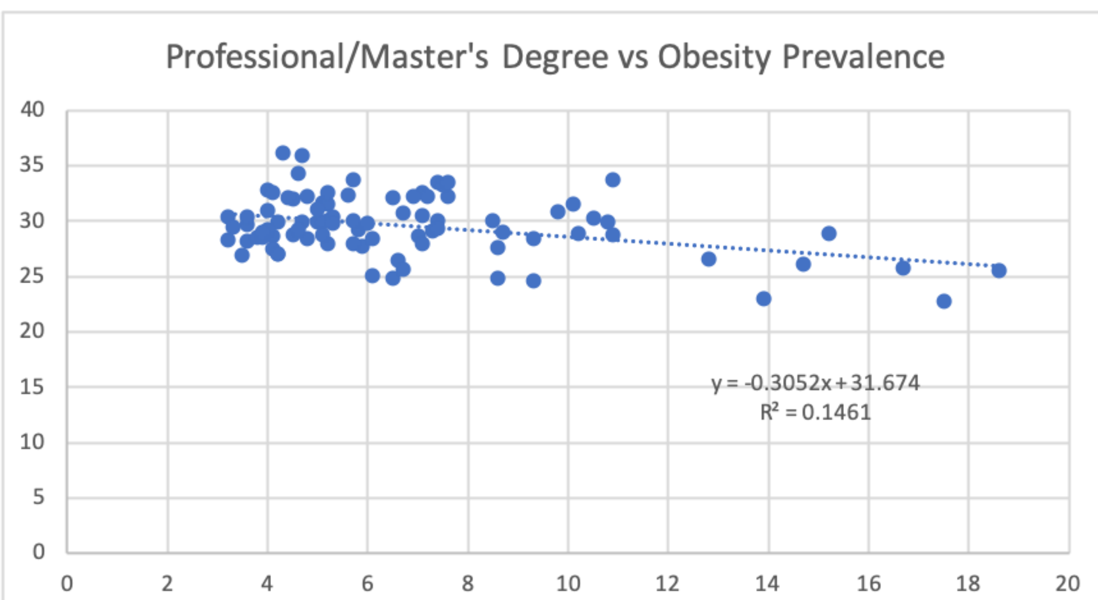


Figure 13 Professional/Master's Degree vs. Obesity Prevalence Correlation

In Figure 14, the R-Squared value equaling 0.1513 of Source Count vs. Obesity Prevalence indicated a very poor regression line fit. 15% of the variation in obesity prevalence is explained by the independent variable source count.

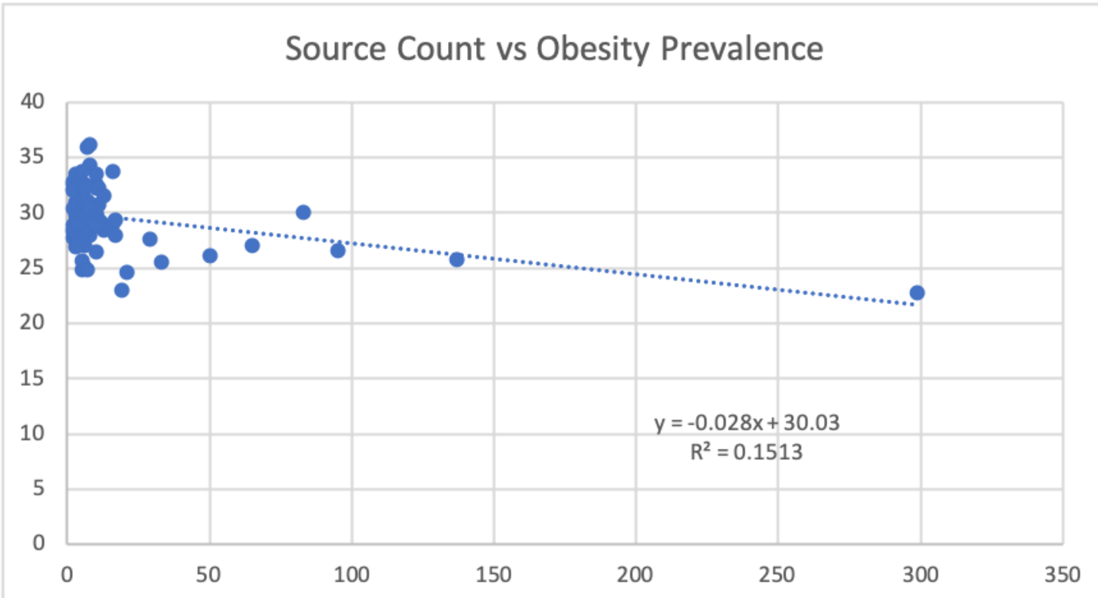


Figure 14 Source Count vs. Obesity Prevalence Correlation

In Figure 15, the R-Squared value equaling 0.1272 of Population Density vs. Obesity Prevalence indicated a very poor regression line fit. 13% of the variation in obesity prevalence is explained by the independent variable population density.

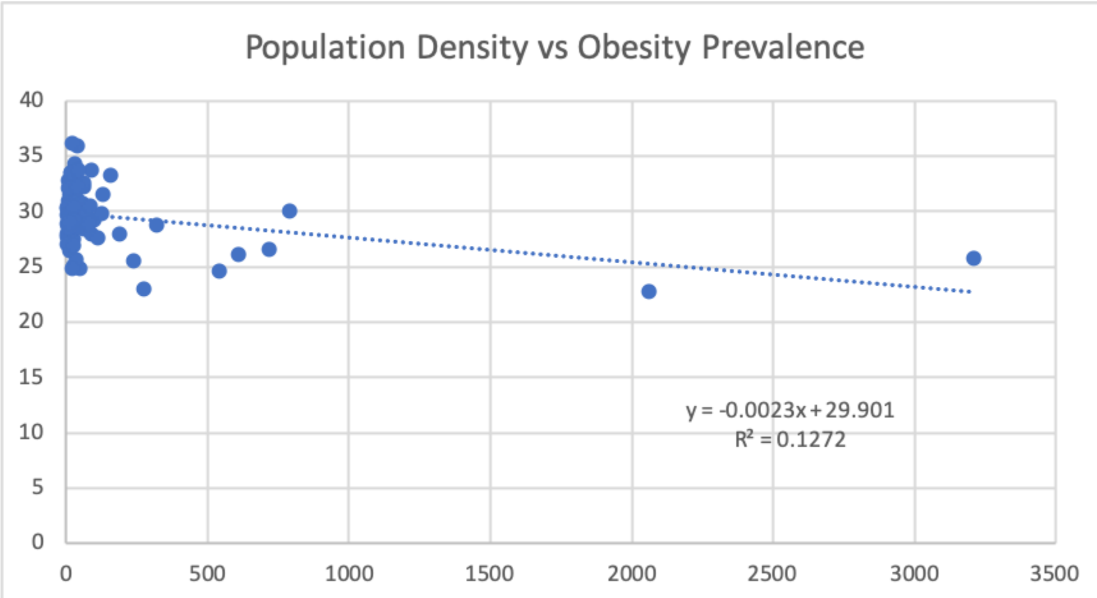


Figure 15 Population Density vs. Obesity Prevalence Correlation

In Figure 16, the R-Squared value equaling 0.1109 of Source Density vs. Obesity Prevalence indicated a very poor regression line fit. 11% of the variation in obesity prevalence is explained by the independent variable source density.

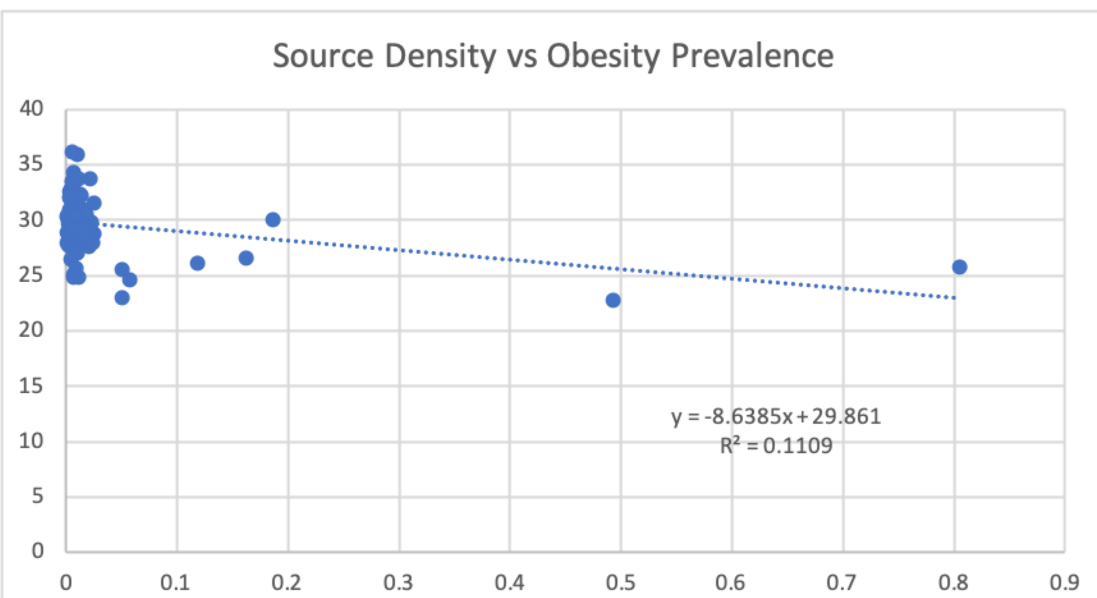


Figure 16 Source Density vs. Obesity Prevalence Correlation

As displayed in Figures 3-16, all 14 explanatory variables have poor regression line fits, though some explanatory variables performed better than others. Explanatory variables 25 and over with a bachelor's degree, total population, and source count had the highest R-Squared values, indicating that those three had the strongest relationships with obesity prevalence. Explanatory variable 25 and over with a bachelor's degree was positively correlated with obesity prevalence. In contrast, explanatory variables total population and source count were negatively correlated with the dependent variable. Although the explanatory variables are not statistically significant, an objective of the analysis was to correlate the explanatory variables against obesity prevalence in Minnesota prior to conducting regression analyses. It was necessary to include all these explanatory variables for comparison between the hypothesized and actual results. A thorough analysis of the statistics is shown in Table 4.

Table 4 Correlations of Explanatory Variables to Obesity Prevalence Chart

Explanatory Variable	**R-Squared**	**R-Value**	**Correlation**
25 and Over with Bachelor's Degree	0.1785	0.422492603	-0.422449991
Total Population	0.1541	0.392555729	-0.392602856
Source Count	0.1513	0.388973007	-0.388988064
25 and Over with Professional/Master's Degree	0.1461	0.382230297	-0.382206529
Population Density	0.1272	0.35665109	-0.356698191
Diabetes Prevalence	0.1123	0.335111922	0.335083855

Source Density	0.1109	0.333016516	-0.333033709
Foreign Born	0.0987	0.314165561	-0.314123159
Physical Inactivity	0.0825	0.287228132	0.287206316
Language Other Than English	0.0676	0.26	-0.259958827
Unemployment	0.0543	0.233023604	0.232968462
25 and Over with Associate's Degree	0.0185	0.136014705	0.136115005
Median Family Income	0.0182	0.134907376	-0.134798968
Poverty Prevalence	0.0133	0.115325626	0.11539687

4.2 Regression Modeling

4.2.1 Ordinary Least Squares Results

The second step in the analysis involved running an OLS Regression analysis. The results of the hypothesized OLS in Table 5 produced a linear model with the 5 explanatory variables that were thought to have produced the best model to analyze obesity. Physical inactivity, poverty prevalence, and unemployment all showed positive relationships with the rate of obesity. Explanatory variables of median family income and source density had indirect relationships with obesity rates. The t-statistic and probability values for the explanatory variables physical inactivity, median family income, and source density suggest that those variables are statistically significant to the model at the 95%, 95%, and 99% confidence level, respectively. All VIF values for the explanatory variables in the model are indicative of the removal of multicollinearity insofar as they all have VIF values < 7.5. Although explanatory variables of poverty prevalence and unemployment were not statistically significant, an objective of the analysis was to compare influence of space on obesity prevalence in Minnesota. It was necessary to include these explanatory variables for comparison between the hypothesized and actual results.

Table 5 OLS Model of Hypothesized Contributing Variables to Obesity

Variable	Coefficient	t-Statistic	Probability	VIF
Intercept	24.641593	6.662621	0.000000*	----
Physical Inactivity	0.437778	2.009759	0.047785*	1.239712
Median Family Income	-0.000048	-2.029751	0.045664*	1.034531
Poverty Prevalence	0.035709	0.423837	0.672814	1.166013
Unemployment	0.168059	0.706591	0.481845	1.372532
Source Density	-7.504010	-2.794633	0.006484*	1.100765

In Table 6, the values for multiple r-squared and adjusted r-squared were 0.211 and 0.162779 respectively, resulting in an AICc value of 410.857562. These values indicate a less than optimum model fit. The Koenker statistic, p-value, yielded a value of 8.804676, and also wasn't statistically significant. The explanatory variables in the analysis do not have a consistent relationship with the dependent variable of obesity prevalence in both geographic and data space. This indicates that the model represents stationarity of the variables. However, the model is unbiased. The standardized residuals of the model follow a normal distribution.

Table 6 OLS Diagnostics of Hypothesized Contributing Variables to Obesity

Number of Observations	87
Multiple R-Squared [d]	0.211
Joint F-Statistic [e]	4.344152
Joint Wald Statistic [e]	23.006088
Koenker (BP) Statistic [f]	8.804676
Jarque-Bera Statistic [g]	1.586573
Akaike's Information Criterion (AICc) [d]	410.857562
Adjusted R-Squared [d]	0.162779
Prob(>F), (5.81) degrees of freedom	0.001492*
Prob(>chi-squared), (5) degrees of freedom	0.000337*
Prob(>chi-squared), (5) degrees of freedom	0.117113
Prob(>chi-squared), (2) degrees of freedom	0.452356

4.2.2 Exploratory Regression Analysis Results

An Exploratory Regression analysis was run after the OLS Regression analysis. The results of the exploratory regression analysis produced a linear model using the four explanatory variables that produced the best model to analyze obesity (see Table 7). These explanatory variables accounted for most of the variance observed for obesity prevalence. Diabetes prevalence was the lone variable to have positive relationship with the rate of obesity. Explanatory variables of median family income, adults 25 and over with a bachelor's degree, and source density all showed indirect relationships with obesity rates. The t-statistic and probability values for the explanatory variable source count suggest that it was statistically significant at the 95%

confidence level. All VIF values for the explanatory variables in the model were indicative of the removal of multicollinearity. All have VIF values < 7.5. Although explanatory variables of diabetes prevalence, median family income, and adults 25 and over with a bachelor's degree are not statistically significant, as a reminder, the objective of the analysis was to compare influence of space on the obesity prevalence in Minnesota. It was necessary to include these explanatory variables for comparison between the hypothesized and actual results.

Table 7 Exploratory Regression Analysis

Variable	Coefficient	t-Statistic	Probability	VIF
Intercept	32.103069	9.911752	0.000000*	----
Diabetes Prevalence	0.328595	1.346891	0.181730	1.433286
Median Family Income	-0.000044	-1.950838	0.054492	1.012067
Bachelor's Degree	-0.124009	-1.800979	0.075386	1.758920
Source Count	-0.017784	-2.245523	0.027421*	1.355902

The diagnostics of the regression analysis is shown in Table 8. The values for multiple r-squared and adjusted r-squared were 0.267561 and 0.231832 respectively, resulting in an AICc value of 402.068415. These values are conclusive of a less than ideal model fit. The Koenker Statistic, p-value, yielded a value of 5.376321, and isn't statistically significant. The explanatory variables from the results of the exploratory analysis do not have a consistent relationship with the dependent variable of obesity prevalence in both geographic and data space. This indicates that the model represents stationarity of the variables. However, the model, like the hypothesized model, is unbiased. The standardized residuals of the model follow a normal distribution. Table 9 shows a simplified synopsis of the statistics associated with Exploratory Regression analysis.

Table 8 Diagnostics of Regression Analysis

Number of Observations	87
Multiple R-Squared [d]	0.267561
Joint F-Statistic [e]	7.488662
Joint Wald Statistic [e]	176.735692
Koenker (BP) Statistic [f]	5.376321
Jarque-Bera Statistic [g]	2.139391
Akaike's Information Criterion (AICc) [d]	402.068415
Adjusted R-Squared [d]	0.231832
Prob(>F), (4,82) degrees of freedom	0.000034*
Prob(>chi-squared), (4) degrees of freedom	0.000000*
Prob(>chi-squared), (4) degrees of freedom	0.250817
Prob(>chi-squared), (2) degrees of freedom	0.343113

Table 9 Exploratory Analysis: Highest Adjusted R-Squared Results

AdjR2	**AICc**	**JB**	**K(BP)**	**VIF**	**SA**	**Model**
0.23	402.07	0.34	0.25	1.76	0.89	Diabetes Prevalence, Median Family Income*, Bachelor's Degree*, Source Count***

4.3 Spatial Autocorrelation Results

A graphical summary of the Spatial Autocorrelation Report was generated as an HTML file after running the tool for both OLS and Exploratory Regression analyses in ArcMap as shown in Figure 17. Given a set of explanatory variables and a dependent variable for each analysis. The Spatial Autocorrelation tool evaluated whether the pattern expressed was clustered, dispersed, or random. The tool also calculated the z-score and p-value to verify the significance of the contributing explanatory variables to the dependent variable. Figure 17 was used as reference.

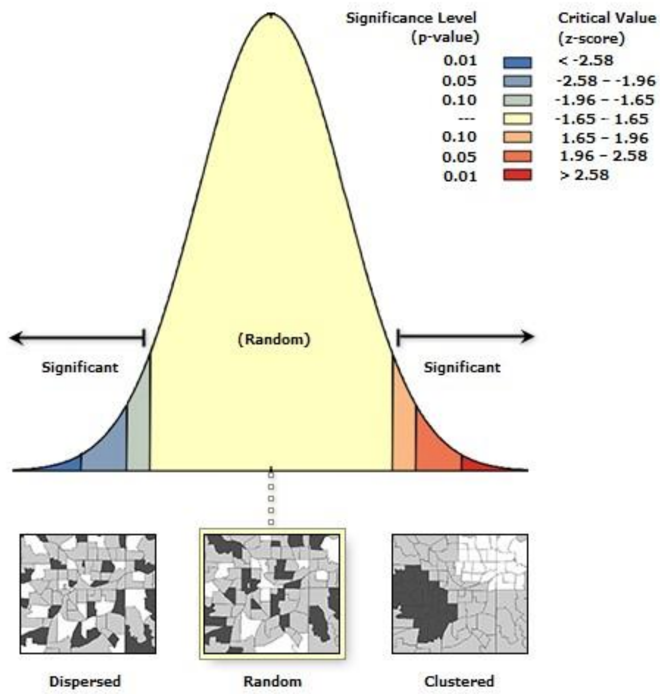


Figure 17 Graphical Summary of Spatial Autocorrelation for Reference

The statistical output of Moran's I for the hypothesis trial is shown in Table 10. Yielding a z-score of 0.491087, the pattern does not appear to be significantly different than random, with reference to Figure 17. The illustration in Figure 18. shows the standardized distribution of residuals across Minnesota's counties. The figure does not display spatial patterns, hence the results agreeing Moran's I. The model performed decently in 4 of the 87 counties, with those counties correlating with a salmon color, being under predicted (having a standard deviation of residual between 1.5 – 2.5).

Table 10 Global Moran's I Summary of Hypothesized Variables

Global Moran's I Summary – Hypothesis	
Moran's Index	0.014294
Expected Index	-0.011628
Variance	0.002786
z-score	0.491087
p-value	0.623365

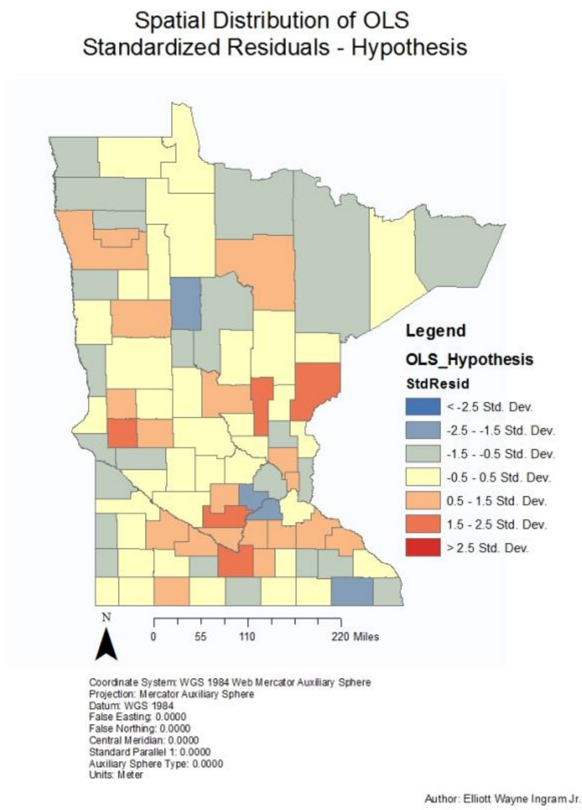


Figure 18 Spatial Distribution of OLS Standardized Residuals of Hypothesized Variables

The statistical output of Moran's I for the Exploratory Regression analysis of explanatory variables that calculated the best model is shown in Table 11. Yielding a z-score of 0.599203, the pattern does not appear to be significantly different than random, with reference to Figure 17. The illustration in Figure 19 shows the standardized distribution of residuals across Minnesota's counties. The figure does not display spatial patterns, hence (similar to the results in the hypothesized analysis) concurs with the findings from Moran's I. The model performed decently in 8 of the 87 counties in Minnesota, with those counties being under predicted (having a standard deviation of residual between 1.5 – 2.5) (See Figure 19).

Table 11 Global Moran's I Summary of Exploratory Regression Variables

Global Moran's I Summary – Exploratory Regression	
Moran's Index	0.020037
Expected Index	-0.011628
Variance	0.002793
z-score	0.599203
p-value	0.549037

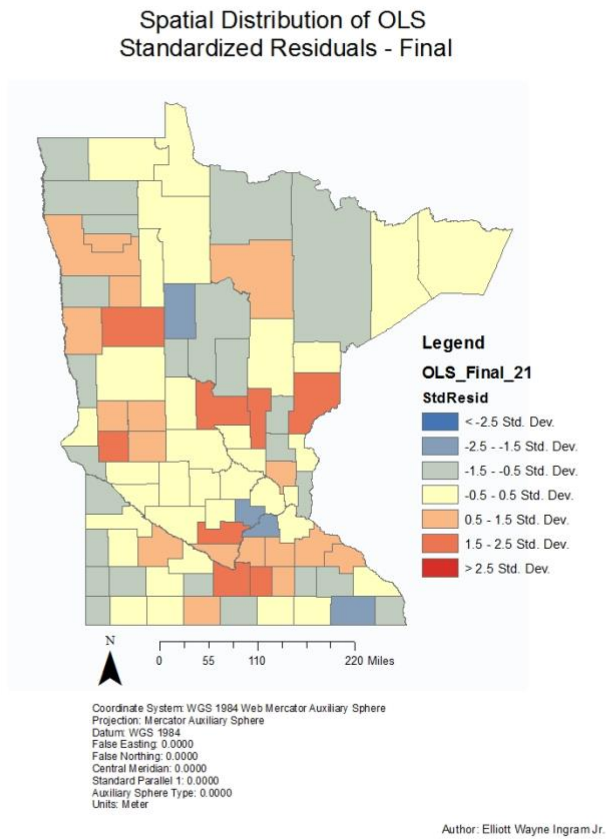


Figure 19 Spatial Distribution of Exploratory Regression Variables Contributing to Obesity

4.4 Geographically Weighted Regression

The Koenker test for both analyses were statistically insignificant, implying non-stationarity of the relationship between the explanatory variables and dependent variable. Therefore, Geographically Weighted Regression (GWR) was not necessary.

Chapter 5 Discussion and Conclusion

5.1 Summary and Significance of Findings

Obesity is a significant public health issue and approaching the topic spatially may provide stakeholders with direction as to how address the phenomenon. It is also a complicated topic that has defied many researchers attempts to understand it. This study showed that obesity rates in Minnesota are impacted by various factors according to the OLS model. The AICc and r-squared values for the hypothesized model were 410.857562 and 0.211 respectively. The AICc and r-squared values for the model after running an exploratory regression of all explanatory variables were 402.068415 and 0.23 respectively. These values suggest that under a quarter of the variance in obesity rates can be explained using this model. Consequently, these models are poor, and reflect the challenges in modeling the spatial relationship between obesity and other demographic and economic factors.

One explanation of the low variances in these models could be that additional factors are required to explain the distribution of obesity rates across Minnesota. This could be because only 14 variables were included in this study, and many other social, cultural, and behavioral variables were left out. The scale of the analysis, the county level, may have generalized the data. Looking at the problem using a different scale of analysis, such as the census tract, may yield more conclusive results. Using disaggregated data could make future analysis more statistically robust. In addition, studies have shown that obesity is directly influenced seasonal eating habits, a temporal scale that this study did not take into account. This could be incorporated into future analyses. Other factors such as healthy food affordability, purchasing decisions, transportation, walkability, technical and regulatory protocol may also help explain the spatial distribution and prevalence of obesity, but were not included in this study.

5.2 Study Limitations and Future Research

5.2.1 Study Limitations

Although this study examined the spatial distribution of obesity at the county level, studies that examine obesity at the community scale have been recommended for prevention and intervention purposes. To do so, a cross-sectional spatial data analysis is ideal. Results would yield geographical variations in obesity between rural and urban communities. Understanding community scale obesity trends would better highlight associated behavioral determinants like diet and physical activity, as well as built environments, socio-economics, and how each determinant contributes to obesity.

Of the 5.611 million residents of Minnesota, 3.6 million live in Hennepin and Ramsey Counties, accounting for ~65% of the total population of the state (ACS, 2017). Yet the state of Minnesota consists of 87 counties. The aggregated data of the explanatory variables could be subject to the Modifiable Areal Unit Problem (MAUP), where the choice of analytical entities may influence the spatial patterning and variability of the data and any ensuing interpretations (Sharkey et al. 2009). Hence, the presence or absence of healthy food outlets would vary greatly across the study area as more healthy food sources are directly proportional to population. In this study, proportional comparisons of socioeconomics and sociodemographics revealed disparities on the county level.

The results of this study suggest the complexity of anticipating phenomenon like obesity, and that methodological and analytical approaches must be carefully chosen. The study only analyzed accessibility to healthy food sources and did not include accessibility to unhealthy food sources, which provides a less comprehensive analysis. The study only measured access to healthy food sources based on count instead of incorporating roadways and routes. The study also only uses statistical methods such as correlation and regression to examine the relationship between healthy food sources, socio-economics, and socio-demographics, and their correlations with obesity. This suggests that statistical data assumes observations are independent or the statistical relationships remain unchanged across the study area.

5.2.2 Future Analyses

Given the relatively weak findings in this study, it is suggested that future analyses implement average nearest distances to fast food outlets and/or healthy food sources to examine their prevalence on obesity rates. Shresta et. al (2013) results showed that obesity rates in Pennsylvania correlated with diabetes, physical inactivity, and average nearest distance to the nearest healthy food store after running an OLS Regression analysis. The AICc and R-Squared values were 299.87 and 0.34 respectively, inferring that only 34% of the variance in obesity rates were explained using OLS. Shresta et. al (2013) also ran a GWR analysis, which yielded AICc and R-Squared values of 261.59 and 0.45, respectively. Their model suggested that additional explanatory variables were required to account for the variance in obesity rates in Pennsylvania. Shresta et. al (2013) approach was different in comparison to this analysis. For example, Shresta et. al (2013) used the Network Analyst tool in ArcMap to successfully determine food accessibility and the average nearest health food facilities to the centroid of each census tract in Pennsylvania. In a network analysis, accessibility can be measured in terms of travel time, distance, or other criteria. Evaluating accessibility can help answer basic questions such as how many people live within a 10-minute drive from healthy food outlet? How many people live within a half-kilometer walking distance from a grocery store? A network analysis was not used in this study due to the large extent of the study area's scale, data restrictions, and the overall time to upload and download massive amounts of data from the ArcMap file. Implementing a network system may have contributed to better results in this analysis. Examining accessibility helps determine whether a community has access to healthy food outlets. Results would yield geographical variations in obesity between rural and urban communities. It better highlights behavioral determinants like diet and physical activity, as well as built environments, socio-economics, and how each determinant contributes to obesity.

In conclusion, this study attempted to understand the relationship between a variety of predictive factors and obesity rates within Minnesota. The existing literature was carefully reviewed and, despite using Shrestha et al. (2013) as a template for the methodology, the results were suggestive rather than conclusive. That said, diabetes rates, education, and age were all found to have some relationship with obesity, despite not being statistically significant. The relationship between obesity, the decisions individuals make about their dietary health, and geographic proximity is a complex phenomenon. This study, by pointing out which variables were *not* explanatory at the county level, can prompt future researchers to choose a wider variety of variables, a different statistical technique, and/or a different scale of analysis.

References

Bressie, Shanna. "Spatial Patterns of Food Accessibility Across Lane County, Oregon in 2015-2016." Master's Thesis, University of Southern California, 2016.

Boone-Heinonen, Janne, Kiefe, Catarina I., Gordon-Larsen, Penny. "Fast Food Restaurants and Food Stores Longitudinal Associations with Diet in Young to Middle-Aged Adults: The CARDIA Study." *Arch Intern Med* 171, (2011): 1162-1170. doi:10.1001/archinternmed.2011.283.

Centers for Disease Control and Prevention. "Adult Obesity Prevalence Maps." Center for Disease Control and Prevention, last modified 25 March 2019, accessed 18 February 2019, 2019, https://www.cdc.gov/obesity/data/prevalence-maps.html

Centers for Disease Control and Prevention. "Physical Activity." Center for Disease Control and Prevention, last modified 13 May 2019, accessed 18 February 2019, https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html

Centers for Disease Control and Prevention. "Diabetes." Center for Disease Control and Prevention last modified 30 May 2019, 2019, accessed 18 February 2019, https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html.

Chalkias, Christos, Apostolos G. Papadopoulos, Kleomenis Kalogeropoulos, Kostas Tambalis, Glykeria Psarra, and Labros Sidossis. "Geographic heterogeneity of the relationship between childhood obesity and socio-environmental status: Empirical evidence from Athens, Greece." *Applied Geography* 37, (2013): 34-43. doi: http://dx.doi.org/10.1016/j.apgeog.2012.10.007

Chen, Xiang. "Take the edge off: A hybrid geographic food access measure." *Applied Geography* 87, (2017): 149-159. doi: http://dx.doi.org/10.1016/j.apgeog.2017.07.013

Chi, Sang-Hyun, Diana S. Grigsby-Toussaint, Natalie Bradford, and Jinmu Choi. "Can Geographically Weighted Regression improve out contextual understanding of obesity in the US? Findings from the USDA Food Atlas." *Applied Geography* 44, (2013): 134-142. doi: http://dx.doi.org/10.1016/j.apgeog.2013.07.017

Cubbin, Catherine, Jina Jun, Claire Margerison-Zilko, Nicolas Welch, James Sherman, Talia McCray, and Barbara Parmenter. "Social Inequalities in neighborhood conditions: spatial relationships between sociodemographic and food environments in Alameda, California." *Journal of Maps* 8, no.4 (2012): 344-348. doi: 10.1080/17445647.2012.747992

Cruz, Hildemar. "A Geospatial Analysis of Income Level, Food Deserts and Urban Agriculture Hot Spots." Master's Thesis, University of Southern California, 2016

Daniel PhD, Mark, Yan Kestens, PhD and Catherine Paquet, PhD. "Demographic and Urban Form Correlates of Healthful and Unhealthful Food Availability in Montréal, Canada." *Canadian Journal of Public Health* 100, no.3 (2009): 189-193. url: https://www-jstor-org.libproxy1.usc.edu/stable/41995243

Deitz, Shiloh L. "A Spatial Analysis of the Relationship Between Obesity and the Built Environment in Southern Illinois." Master's Thesis, Southern Illinois University Carbondale, 2014

Drewnowski, A., AV Moudon, J. Jiao, A. Aggarwal, H. Charreire, and B. Chaix. "Food environment and socioeconomic status influence obesity rates in Seattle and in Paris." *International Journal of Obesity*, no.38 (2014): 308-314. doi: 10.1038/ijo.2013.97

ESRI 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute

Ferguson, Neil S., Karen E. Lamb, Yang Wang, David Ogilvie, and Anne Ellaway. "Access to Recreational Physical Inactivity by Car and Bus: An Assessment of Socio-Spatial Inequalities in Mainland Scotland." *PLoS ONE* 8(2) (2013): e55638. doi: 10.1371/journal.pone.0055638.

Fleischhacker, S.E., K.R. Evenson, D.A. Rodriguez, and A.S. Ammerman. "A systematic review of fast food access studies." *Obesity Reviews* 12, (2011): e460-e471. doi: 10.1111/j.1467-789X.2010.00715.x

Gard, Julienne. "A Spatial Accessibility Analysis of the Los Angeles Foodscape." Doctorate's Dissertation, University of Southern California, 2015

Helbich, Marco, Bjorn Schadenberg, Julian Hagenauer, and Maartje Poelman. "Food deserts? Healthy food access in Amsterdam." *Applied Geography* 83, (2017): 1-12. doi: http://dx.doi.org/10.1016/j.apgeog.2017.02.015

Hellmich, Simon N. "What is Socioeconomics? An Overview of Theories, Methods, and Themes in the Field." *Forum for Social Economics* 46, (2015): 3-25. doi:10.1080/07360932.2014.999696

Hilmers, Angela, MD, MS, David C. Hilmers, MD, MPH, and Jayna Dave, PhD. "Neighborhood Disparities in Access to Healthy Foods and Their Effects on Environmental Justice." *American Journal of Public Health* 102, no. 9 (2012): 1644-1654. doi: 10.2105/AJPH.2012.300865

Horner, Mark W., and Brittany S. Wood. "Capturing individuals 'food environments using flexible space-time accessibility measures." *Applied Geography* 51, (2014): 99-107. doi: http://dx.doi.org/10.1016/j,apgeog.2014.03.007

Howard, Philip H., Margaret Fitzpatrick, and Brian Fulfrost. “Proximity of food retailers to schools and rates of overweight ninth grade students: an ecological study in California.” *BMC Public Health* 11, no. 68 (2011): 1-8. doi: 10.1186/1471-2458-11-68

Jeffery, Robert W., Judy Baxter, Maureen McGuire, and Jennifer Linde. “Are fast food restaurants an environmental risk factor for obesity?” *International Journal of Behavioral Nutrition and Physical Activity* 3, no. 2 (2006). doi: 10.1186/1479-5868-3-2

Larsen, Kristian and Jason Gilliland. “Mapping the evolution of ‘food deserts’ in a Canadian city: Supermarket accessibility in London, Ontario, 1961-2005.” *International Journal of Health Geographics* 7, no. 16 (2008): 1-16. doi: 10.10.1186/1476-072X-7-16

Larson, Nicole I., PhD, MPH, RD, Mary T. Story, PhD, RD, and Melissa C. Nelson, PhD, RD. “Neighboring Environments: Disparities in Access to Healthy Foods in the U.S.” *American Journal of Preventive Medicine* 36, no. 1 (2009): 74-81.

Malina, Robert M. “Ethnic variation in the prevalence of obesity in North American children and youth.” *Critical Reviews in Food Science and Nutrition* 33, (1993): 389-396. doi: http://dx.doi.org/10.1080/10408399309527637

Mattessich, Paul W., Rausch, Ela J. “Healthy Food Access: A View of the Landscape in Minnesota and Lessons Learned from Healthy Food Financing Initiatives.” *Wilder Research*: (2016). 91-54. url: https://www.wilder.org/WilderResearch/Publications/Studies/Healthy.pdf.

Minnesota Compass. Wilder Research (2017). “Near-North Neighborhood,” www.mncompass.org/profiles/neighborhoods/minneapolis/near-north#notes.

Minnesota Department of Agriculture. “Minnesota Grown Directory.” https://minnesotagrown.com/search-directory/farmers-markets/?gclid=EAIaIQobChMI-aPZ2pHI5AIVF6SzCh25UQuQEAAYASAAEgIut_D_BwE

Minnesota Department of Health. 2017. “Adult Obesity in Minnesota 2017: Data Brief.” https://www.health.state.mn.us/people/obesity/docs/obesitybrief.pdf.

Minneapolis Department of Health and Family Support. 2012. “Minneapolis Healthy Corner Store Program: Making produce more visible, affordable and attractive,” www.health.state.mn.us/divs/oshii/docs/Mpls_Healthy_Corner.pdf.

“Minneapolis, Minnesota (MN) Poverty Rate Data,” 2017. Avameg Inc. www.city-data.com/poverty/poverty-Minneapolis-Minnesota.html.

Morland, Kimberly, PhD, Steve Wing, PhD, Ana Diez Roux, MD, PhD, and Charles Poole, ScD. “Neighborhood Characteristics Associated with the Location of Food Stores and Food Service Places.” *American Journal of Preventive Medicine* 22, no.1 (2001): 23-29.

Olson, Jeremy. 2016. “Minnesota among 10 worst states for access to fresh food: Leaders look for Legislature for state funding solutions.” *Star Tribune*. http://www.startribune.com/minnesota-among-10-worst-states-for-food-deserts/375573111/.

Penney, T. L., D. G. C. Rainham, T. J. B. Dummer, and S. F. L. Kirk. “A spatial analysis of community level overweight and obesity.” *Journal of Human Nutrition and Dietetics* 27, no. 2 (2013): 65-74. doi: 10.1111/jhn.12055

Richardson, Karen. “Exploring Food Environments: Assessing access to nutritious food.” *ArcUser Online* (2010): 50-52. accessed 1 March 2019 https://www.esri.com/news/arcuser/1010/files/foodataset.pdf.

Sharkey, Joseph R., Cassandra M. Johnson, Wesley R. Dean, and Scott A. Horel. “Focusing on fast food restaurants alone underestimates the relationship between neighborhood deprivation and exposure to fast food in a large rural area.” *Nutrition Journal* 10, no. 10 (2011): 1-14. doi: http://www.nutritionj.com.content/10/1/10

Shrestha, Ranjay, Ron Mahabir, and Liping Di. *Healthy Food Accessibility and Obesity: Case Study of Pennsylvania, USA.* 2013.

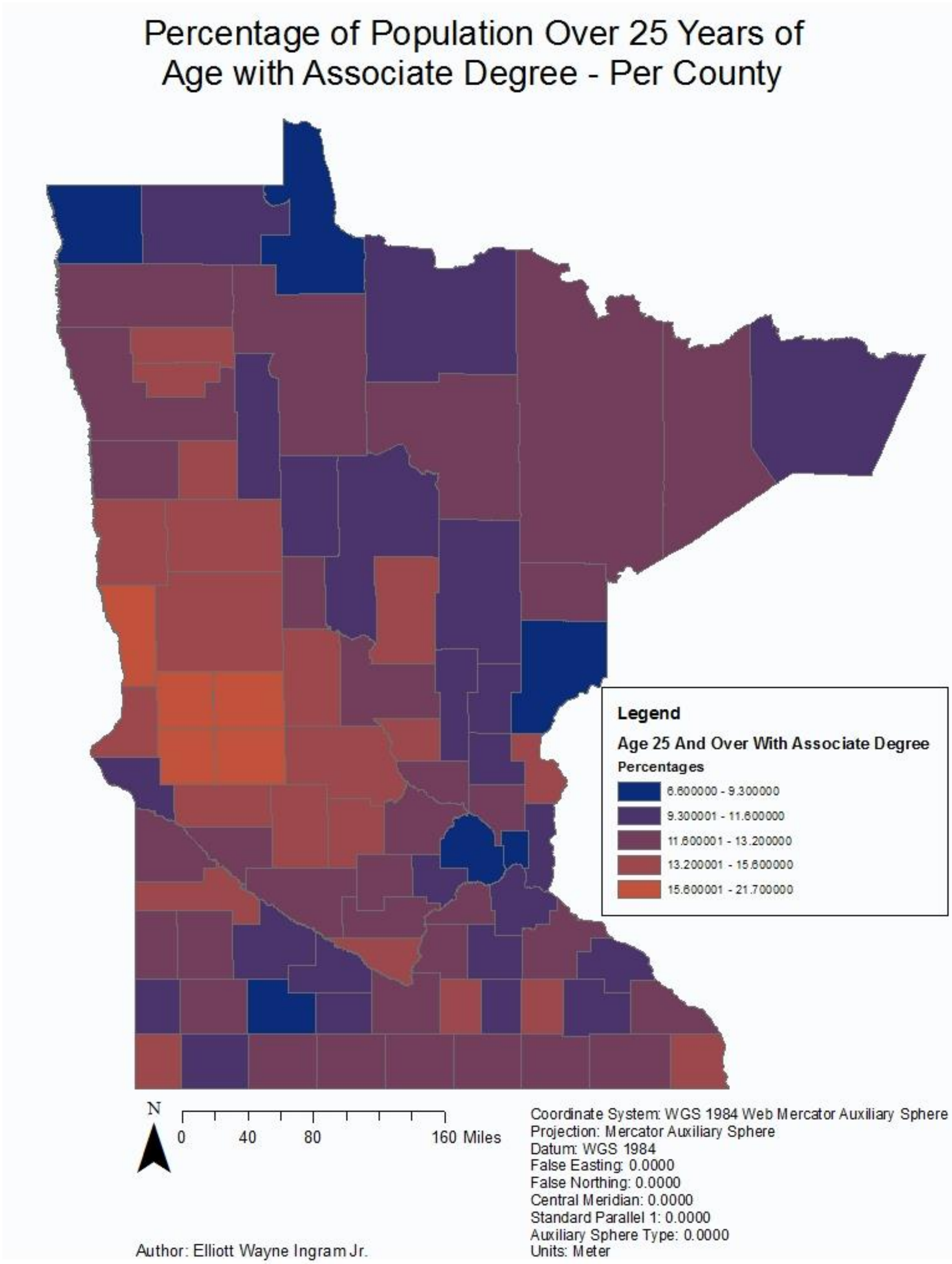
Stein, Dana Beth. “‘Food Deserts’ and ‘Food Swamps’ in Hillsborough County, Florida: Unequal Access to Supermarkets and Fast-food Restaurants.” Master’s Thesis, University of South Florida, 2011.

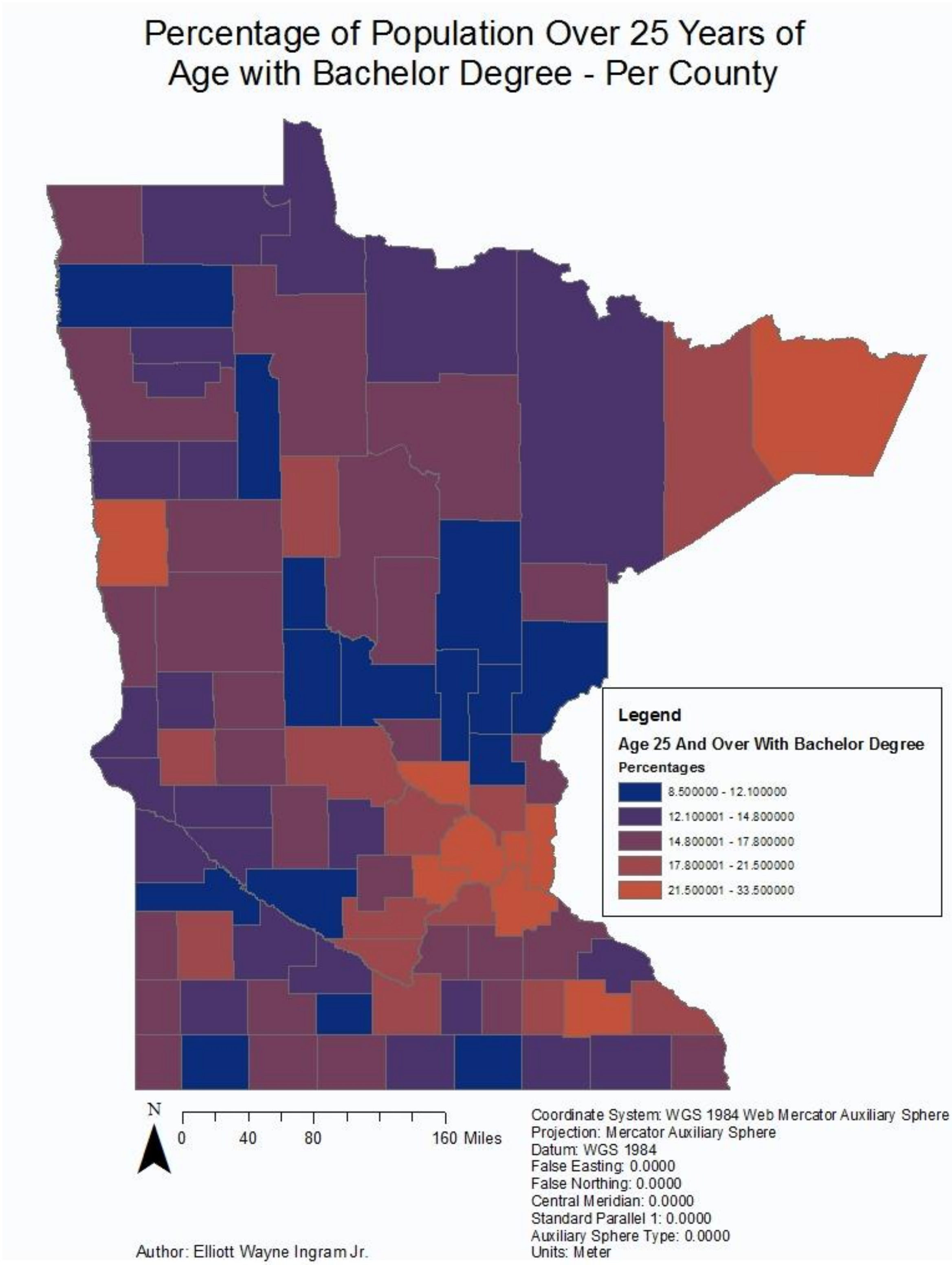
Todd, Michael, Mark A. Adams, Jonathan Kurka, Terry L. Conway, Kelli L. Cain, Matthew P. Bunman, Lawrence D. Frank, James F. Sallis, and Abby C. King. “GIS-measured walkability transit, and recreation environments in relation to older Adults’ physical inactivity: A latent profile analysis.” *Preventive Medicine* 93, (2016): 57-63. doi: http://dx.doi.org/10.1016/j.ypmed.2016.09.019

Xu, Yanqing, Wang, Lei. “GIS-based analysis of obesity and the built environment in the US.” *Cartography and Geographic Information Science* 42, no.1 (2015): 9-21. doi: 0.1080/15230406

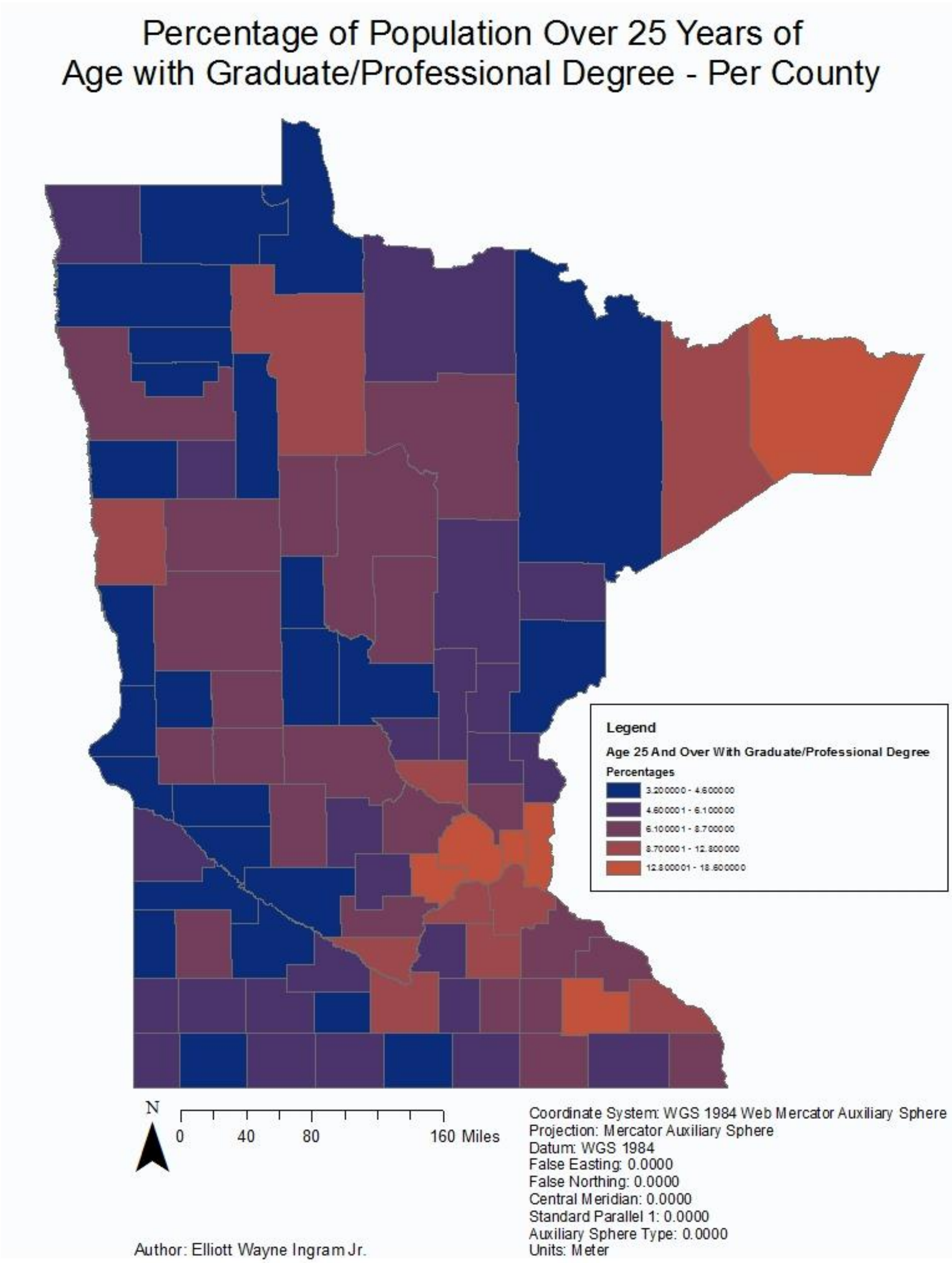
Wen, Tzai-Hung, Duan-Rung Chen, and Meng-ju Tsai. “Identifying geographical variations in poverty-obesity relationships: empirical evidence from Taiwan.” *Geospatial Health* 4, no. 2 (2010): 257-265. doi: https://doi.org.10.4081/gh.2010.205

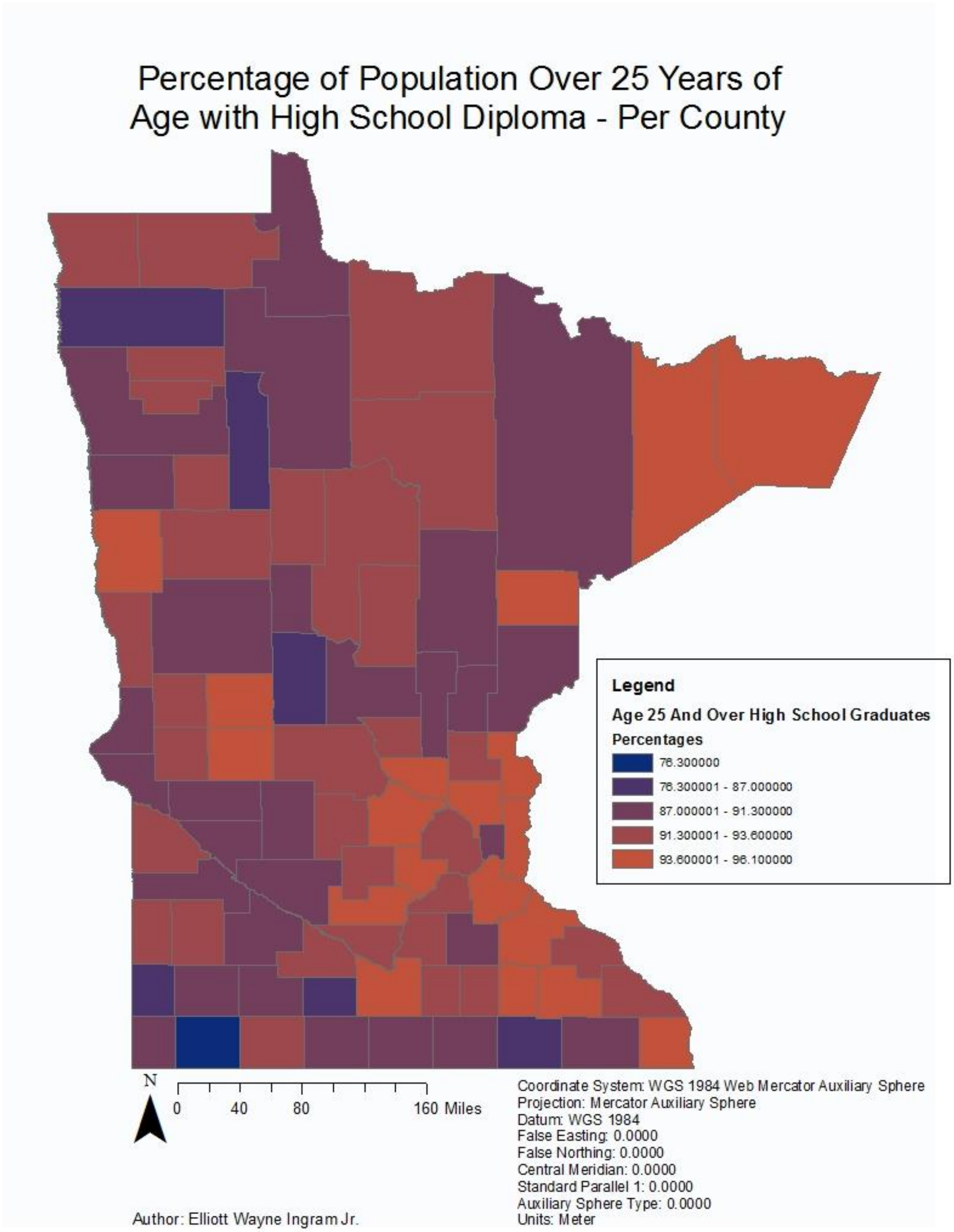
Appendix A Maps of Aggregated Data Used in Analysis

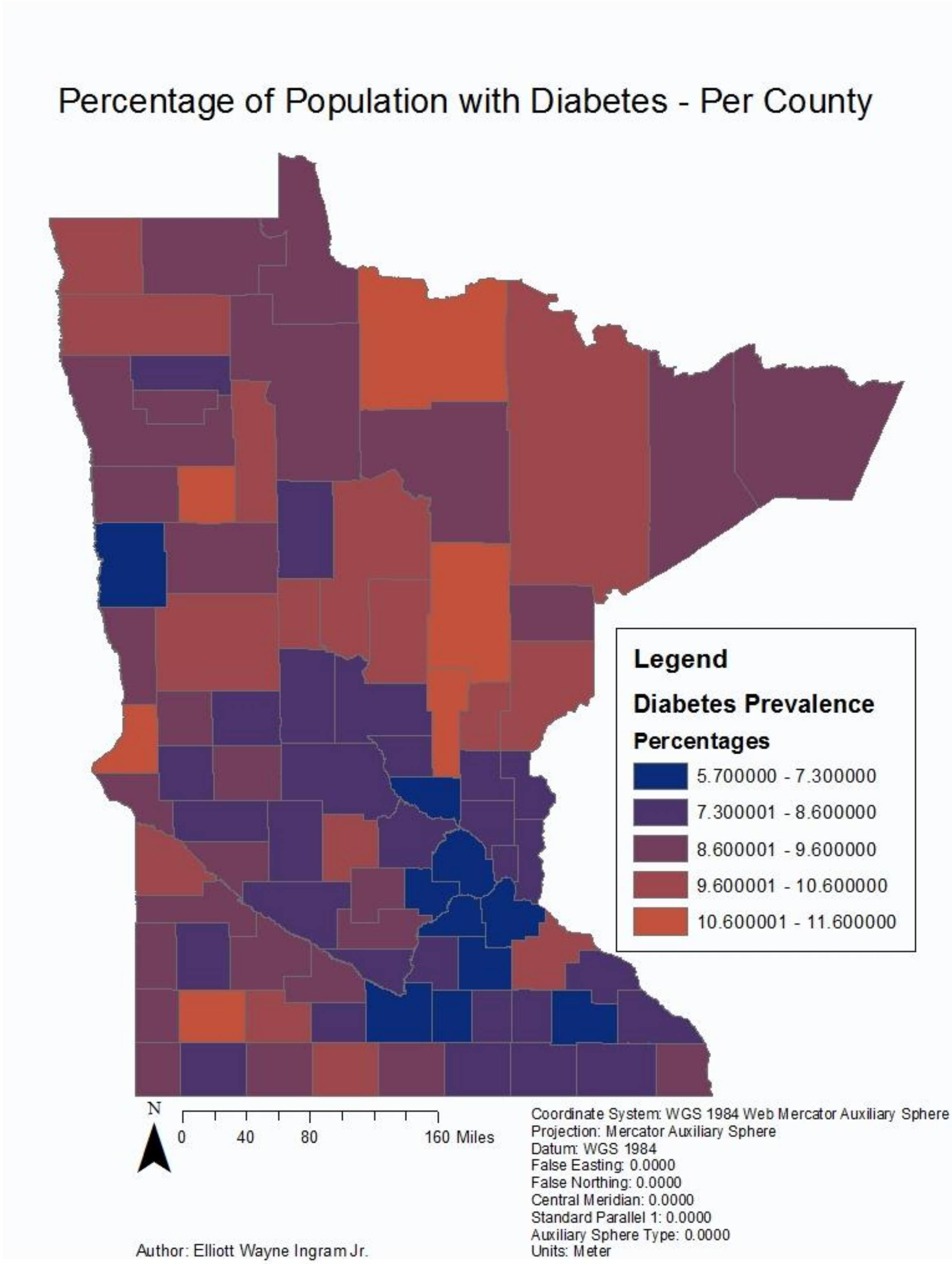


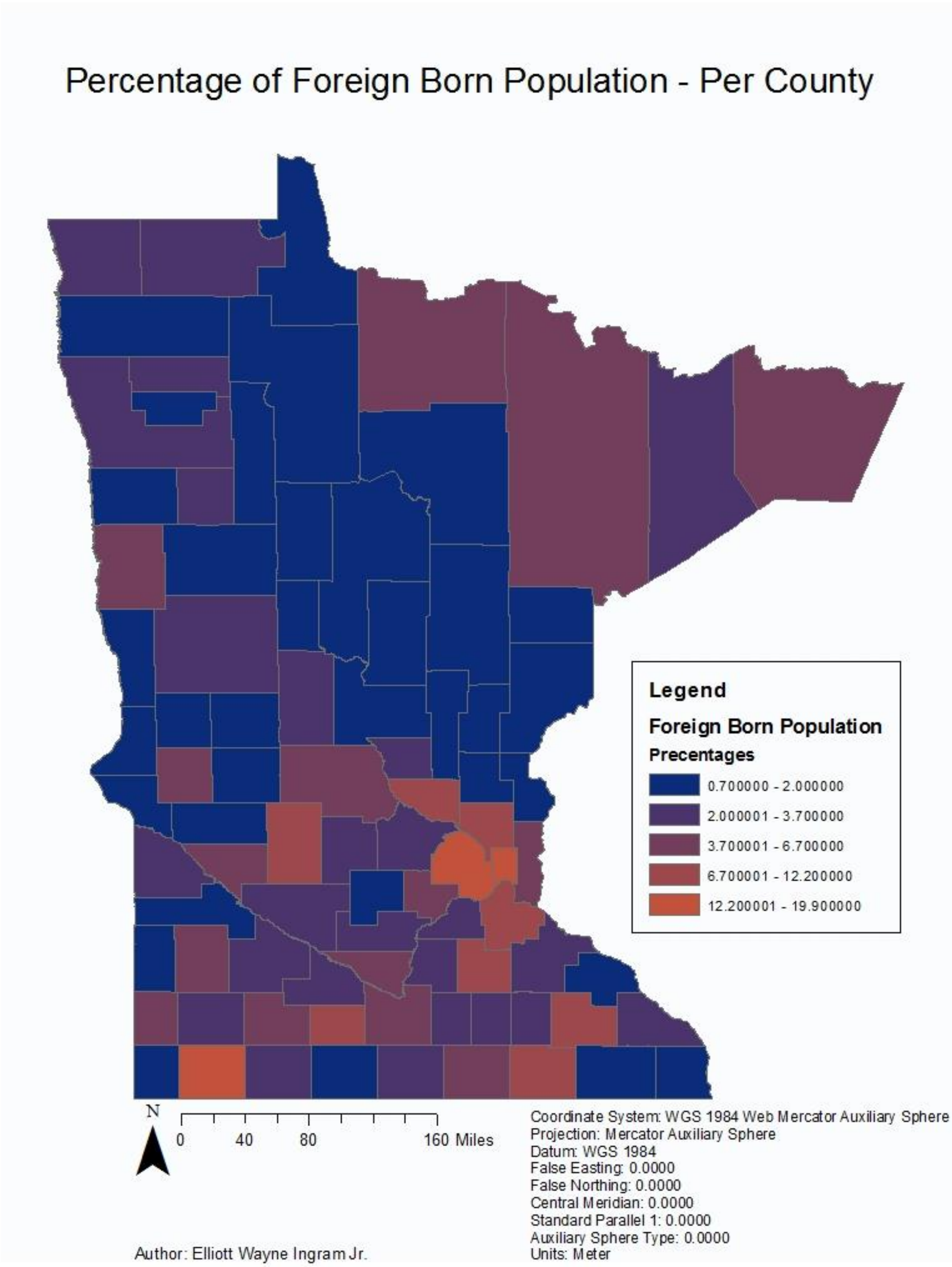


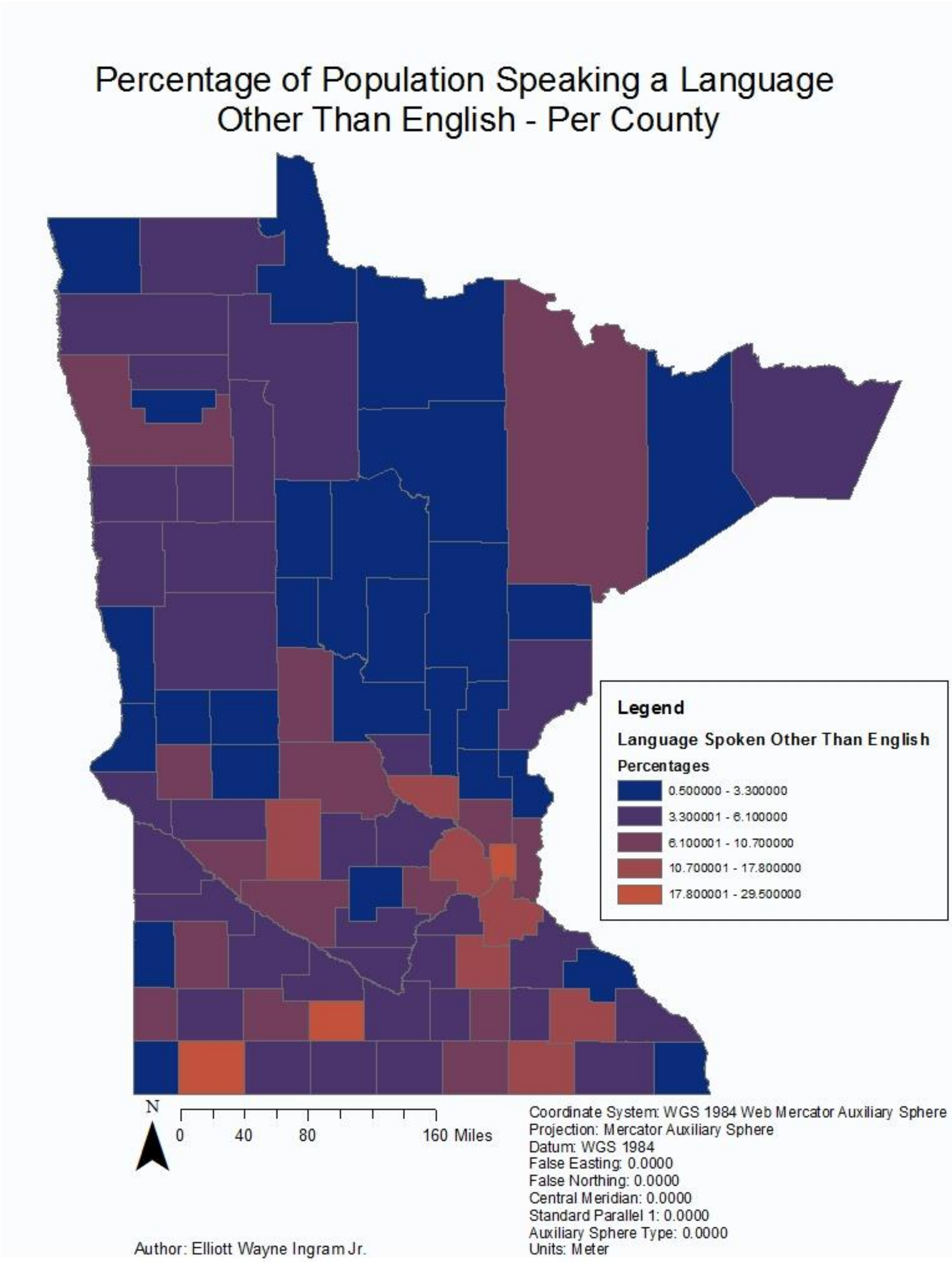
Percentage of Population Over 25 Years of Age with Bachelor Degree - Per County
Legend
Age 25 And Over With Bachelor Degree
Percentages
8.500000 - 12.100000
12.100001 - 14.800000
14.800001 - 17.800000
17.800001 - 21.500000
21.500001 - 33.500000
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.

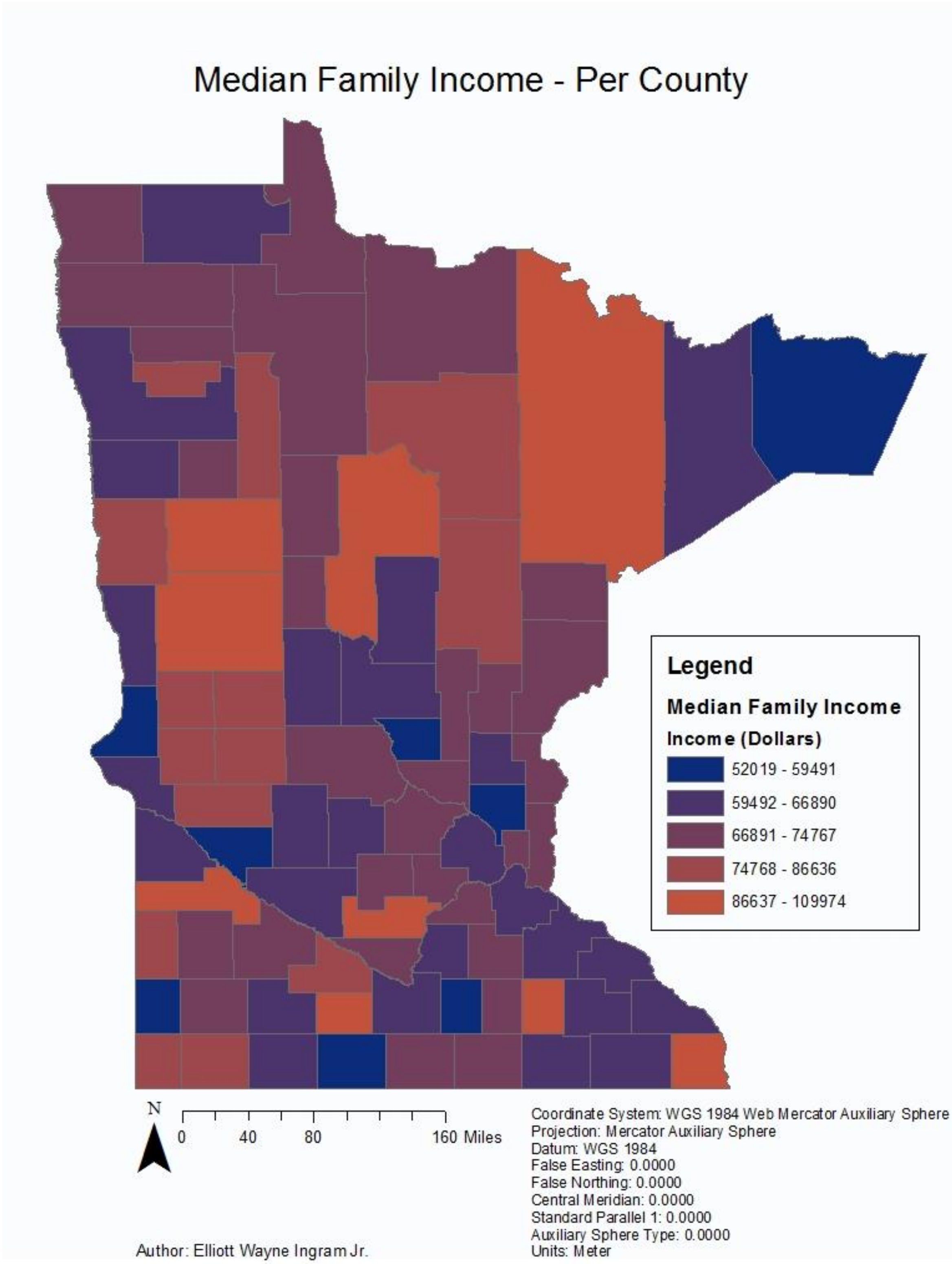

Percentage of Population Over 25 Years of Age with Graduate/Professional Degree - Per County
Legend
Age 25 And Over With Graduate/Professional Degree
Percentages
3.200000 - 4.600000
4.600001 - 6.100000
6.100001 - 8.700000
8.700001 - 12.800000
12.800001 - 18.600000
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.

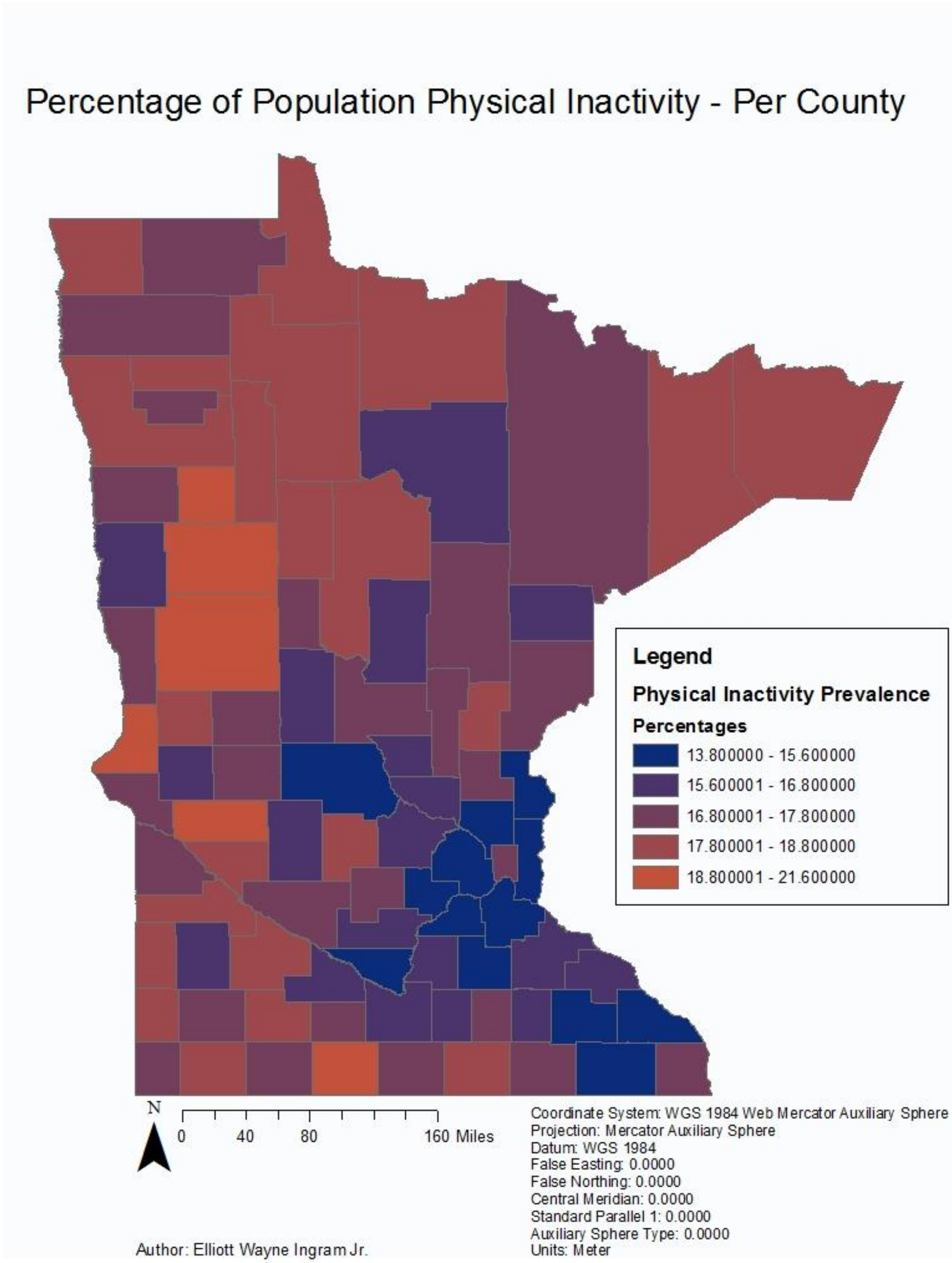

Percentage of Population Over 25 Years of Age with High School Diploma - Per County
Legend
Age 25 And Over High School Graduates
Percentages
76.300000
76.300001 - 87.000000
87.000001 - 91.300000
91.300001 - 93.600000
93.600001 - 96.100000
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.

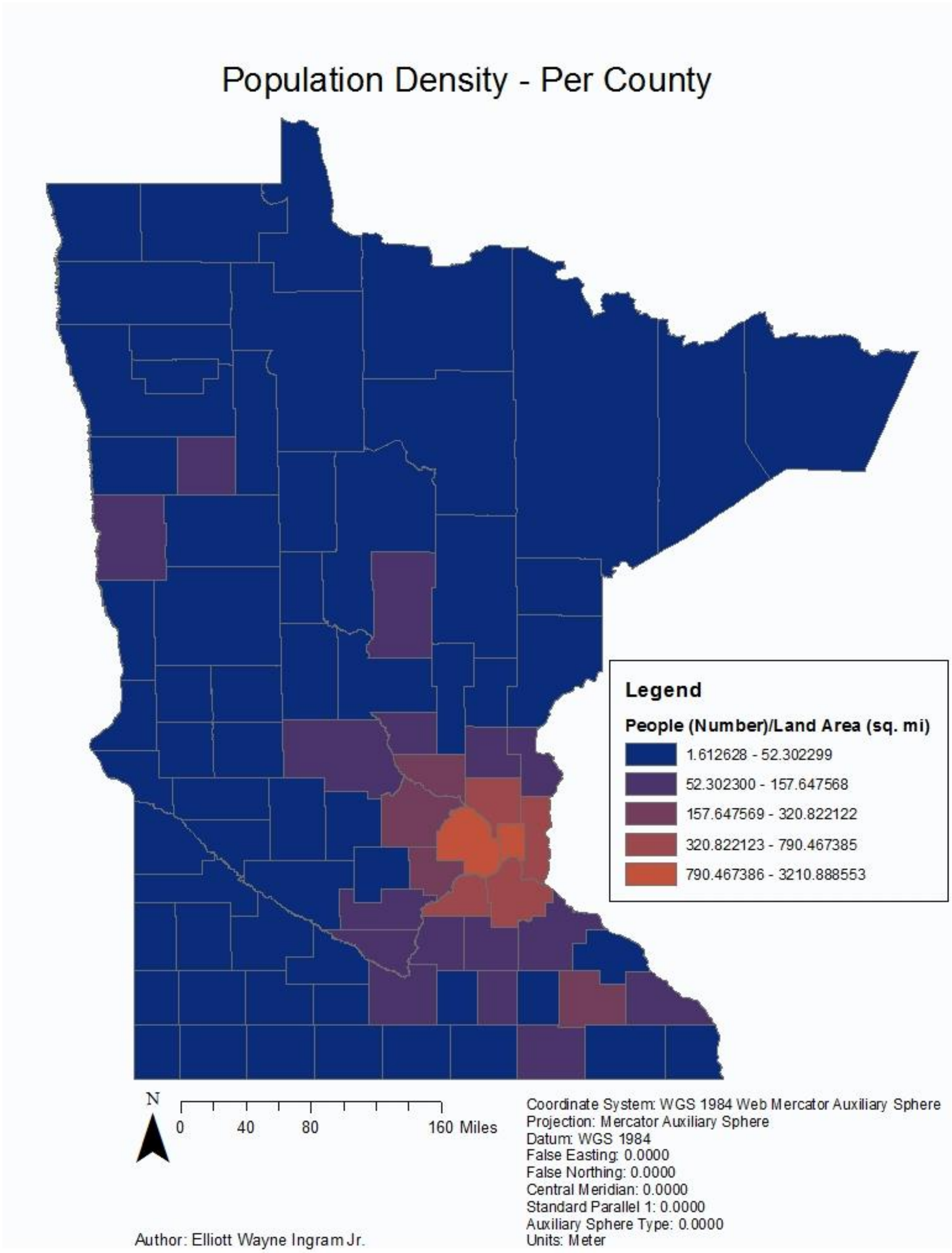

Percentage of Population with Diabetes - Per County
Legend
Diabetes Prevalence
Percentages
5.700000 - 7.300000
7.300001 - 8.600000
8.600001 - 9.600000
9.600001 - 10.600000
10.600001 - 11.600000
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.

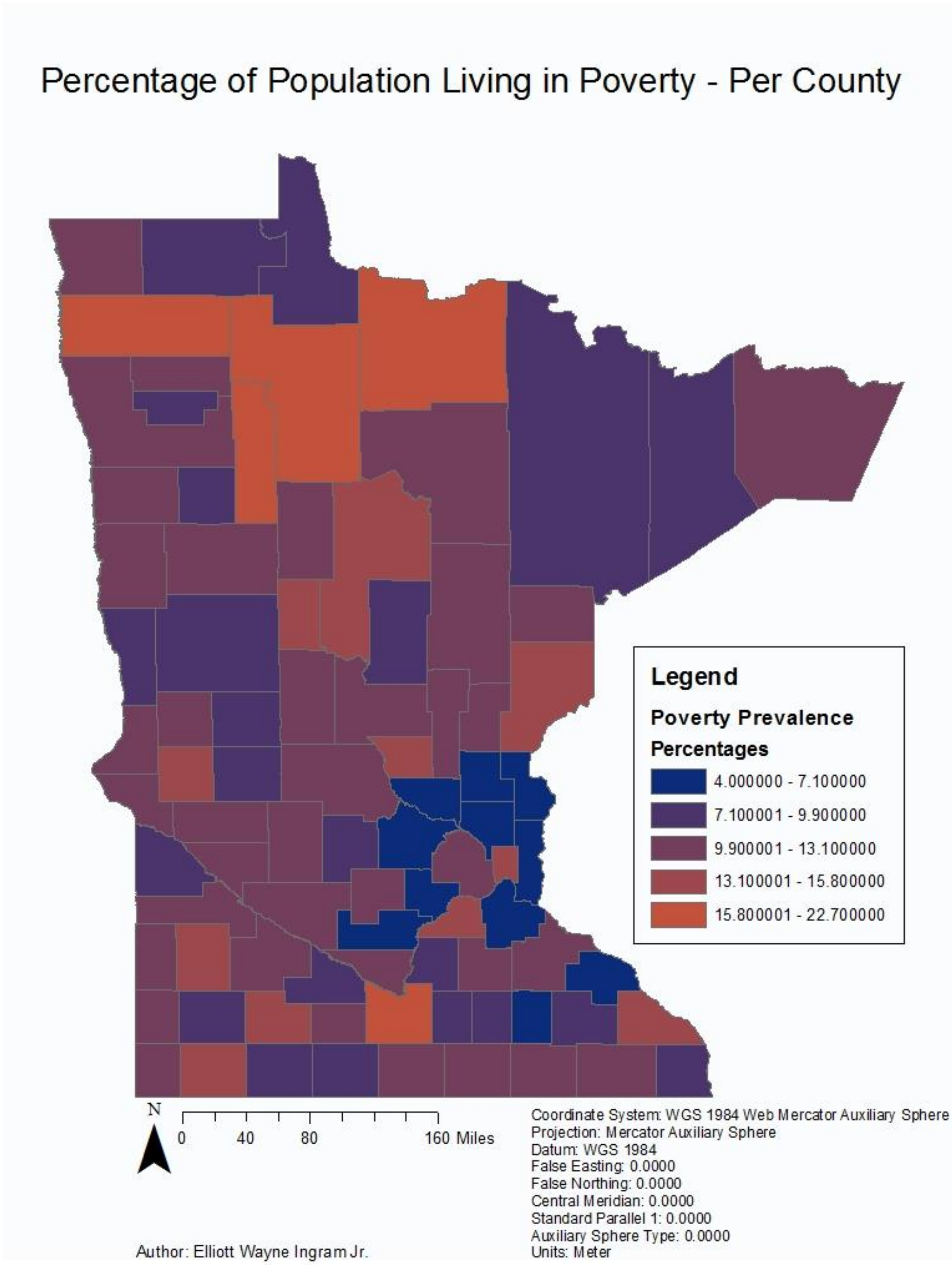

Percentage of Foreign Born Population - Per County
Legend
Foreign Born Population
Precentages
0.700000 - 2.000000
2.000001 - 3.700000
3.700001 - 6.700000
6.700001 - 12.200000
12.200001 - 19.900000
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.

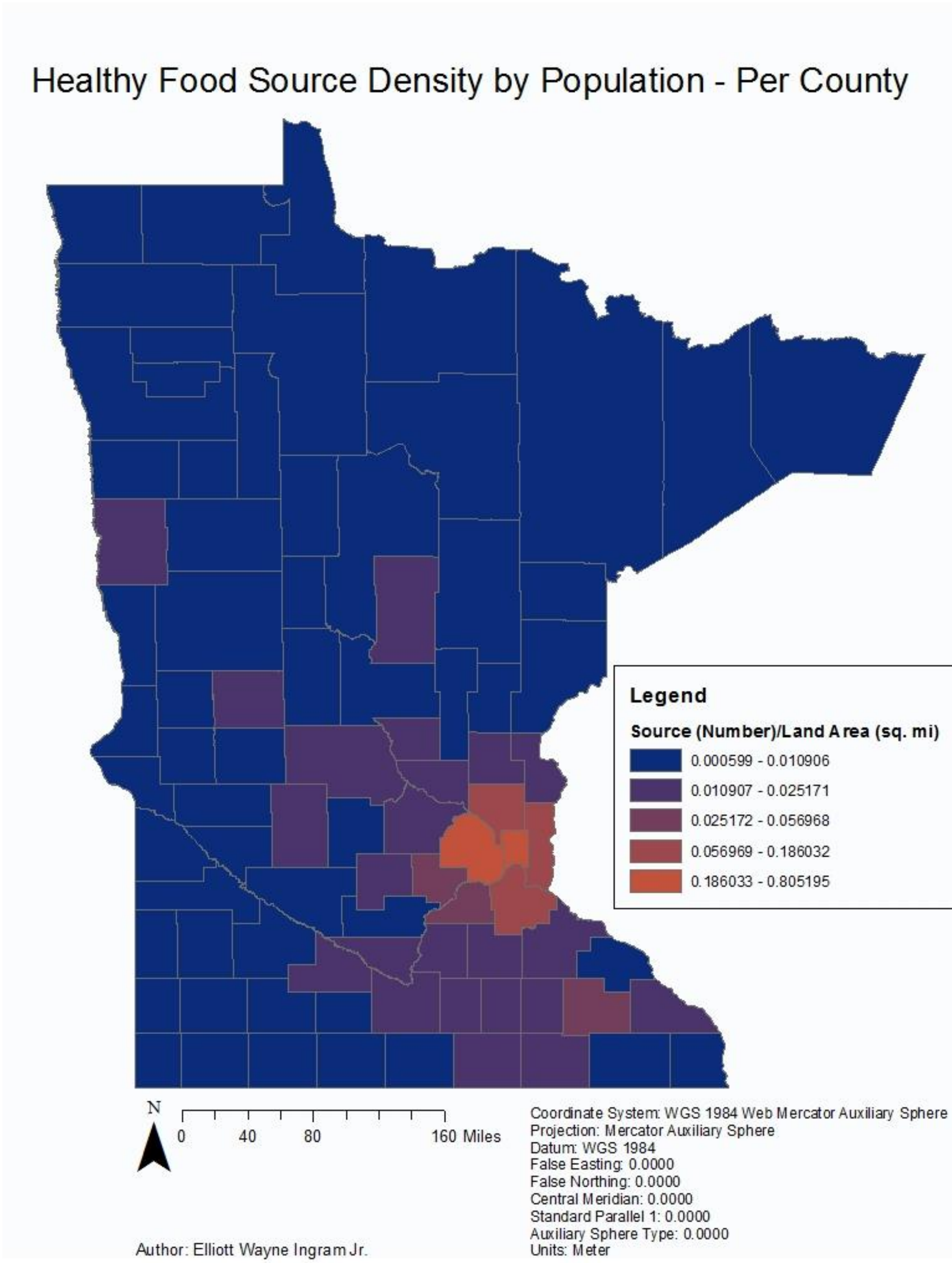

Percentage of Population Speaking a Language Other Than English - Per County
Legend
Language Spoken Other Than English
Percentages
0.500000 - 3.300000
3.300001 - 6.100000
6.100001 - 10.700000
10.700001 - 17.800000
17.800001 - 29.500000
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.

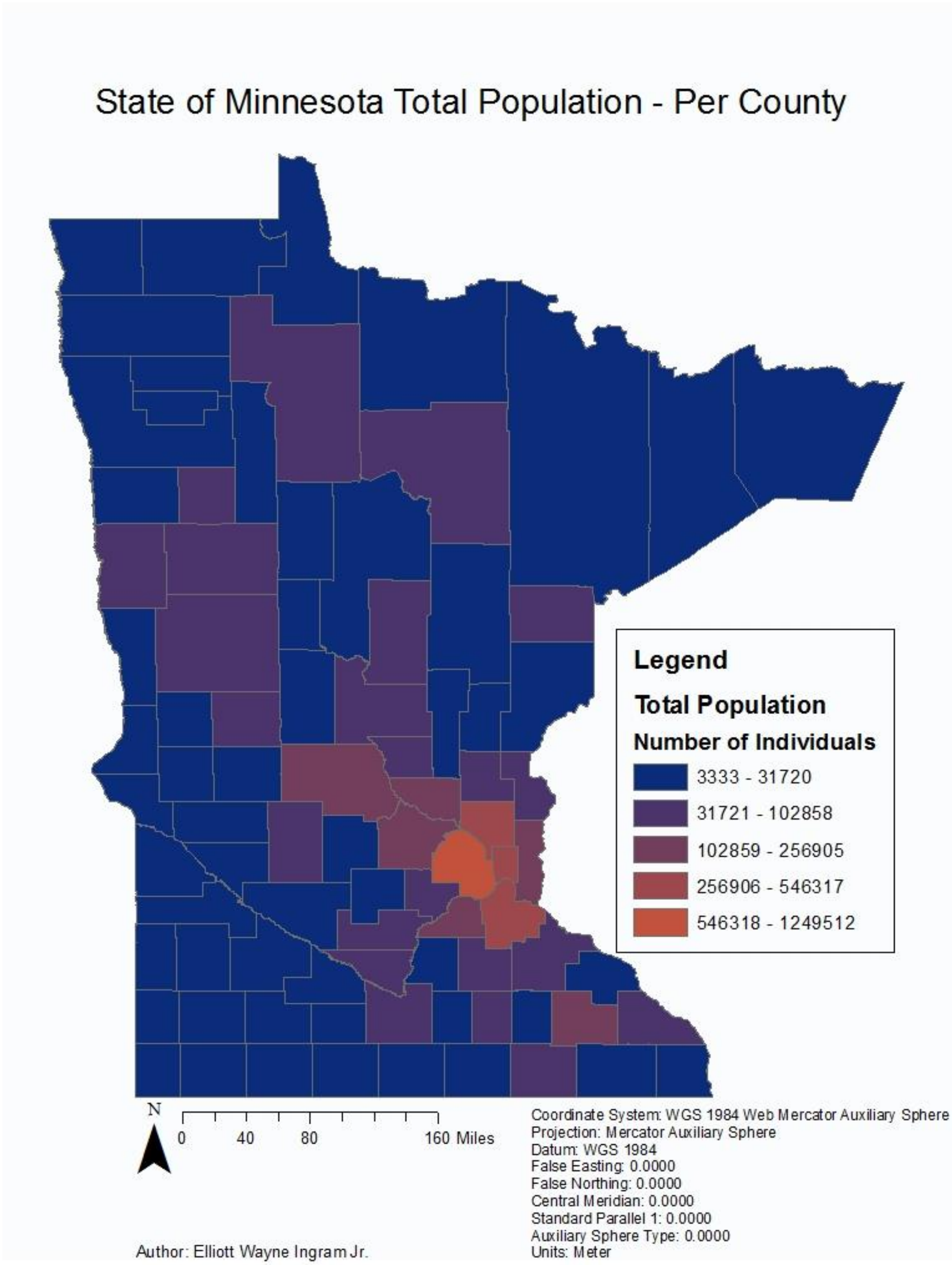

Median Family Income - Per County
Legend
Median Family Income
Income (Dollars)
52019 - 59491
59492 - 66890
66891 - 74767
74768 - 86636
86637 - 109974
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.

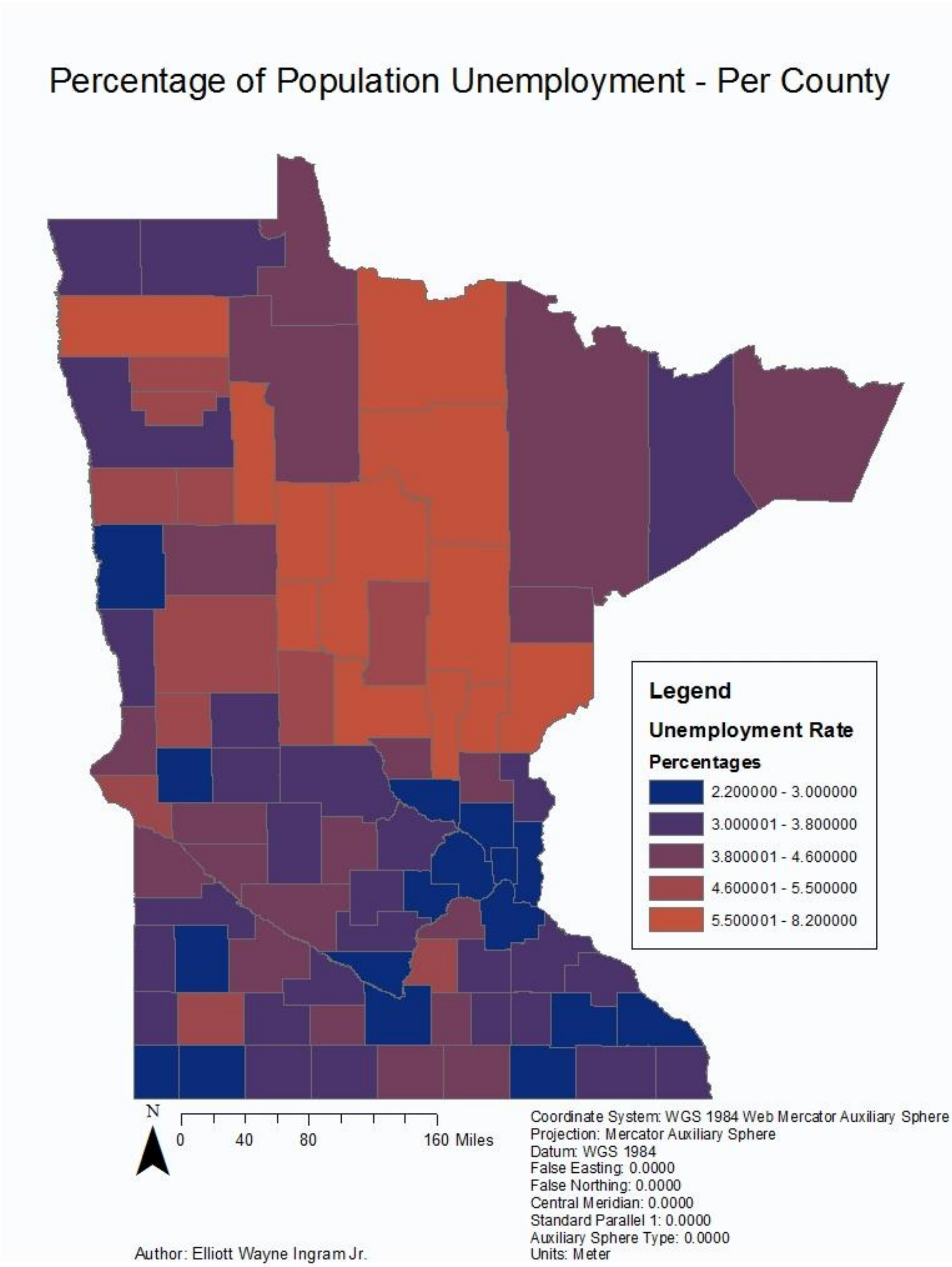

Percentage of Population Physical Inactivity - Per County
Legend
Physical Inactivity Prevalence
Percentages
13.800000 - 15.600000
15.600001 - 16.800000
16.800001 - 17.800000
17.800001 - 18.800000
18.800001 - 21.600000
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.


Population Density - Per County
Legend
People (Number)/Land Area (sq. mi)
1.612628 - 52.302299
52.302300 - 157.647568
157.647569 - 320.822122
320.822123 - 790.467385
790.467386 - 3210.888553
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.


Percentage of Population Living in Poverty - Per County
Legend
Poverty Prevalence
Percentages
4.000000 - 7.100000
7.100001 - 9.900000
9.900001 - 13.100000
13.100001 - 15.800000
15.800001 - 22.700000
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.


Healthy Food Source Density by Population - Per County
Legend
Source (Number)/Land Area (sq. mi)
0.000599 - 0.010906
0.010907 - 0.025171
0.025172 - 0.056968
0.056969 - 0.186032
0.186033 - 0.805195
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.


State of Minnesota Total Population - Per County
Legend
Total Population
Number of Individuals
3333 - 31720
31721 - 102858
102859 - 256905
256906 - 546317
546318 - 1249512
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.


Percentage of Population Unemployment - Per County
Legend
Unemployment Rate
Percentages
2.200000 - 3.000000
3.000001 - 3.800000
3.800001 - 4.600000
4.600001 - 5.500000
5.500001 - 8.200000
N
0
40
80
160 Miles
Coordinate System: WGS 1984 Web Mercator Auxiliary Sphere
Projection: Mercator Auxiliary Sphere
Datum: WGS 1984
False Easting: 0.0000
False Northing: 0.0000
Central Meridian: 0.0000
Standard Parallel 1: 0.0000
Auxiliary Sphere Type: 0.0000
Units: Meter
Author: Elliott Wayne Ingram Jr.

Appendix B Exploratory Regression Model – Raw Results

****
Choose 5 of 15 Summary
Highest Adjusted R-Squared Results
AdjR2 AICc JB K(BP) VIF SA Model
0.23 403.34 0.31 0.50 1.76 0.76 +DIABETES_PREVALENCE -MEDIAN_FAMILY_INCOME** +POPULATION_25_AND_OVER_ASSOCIATES_DEGREE -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -SOURCE_COUNT*
0.23 403.59 0.48 0.19 3.10 0.76 +DIABETES_PREVALENCE +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -MEDIAN_FAMILY_INCOME* -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -SOURCE_COUNT*
0.23 403.59 0.46 0.19 2.57 0.68 +PHYSICAL_INACTIVITY_PREVALENCE +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -MEDIAN_FAMILY_INCOME** -POPULATION_25_AND_OVER_BACHELORS_DEGREE** -SOURCE_COUNT*
Passing Models
AdjR2 AICc JB K(BP) VIF SA Model

****
Choose 6 of 15 Summary
Highest Adjusted R-Squared Results
AdjR2 AICc JB K(BP) VIF SA Model
0.23 405.18 0.35 0.47 24.91 0.75 +DIABETES_PREVALENCE +TOTAL_POPULATION -MEDIAN_FAMILY_INCOME* +POPULATION_25_AND_OVER_ASSOCIATES_DEGREE -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -SOURCE_COUNT
0.23 405.42 0.40 0.26 3.23 0.70 +DIABETES_PREVALENCE +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -MEDIAN_FAMILY_INCOME* +POPULATION_25_AND_OVER_ASSOCIATES_DEGREE -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -SOURCE_COUNT*
0.22 405.53 0.53 0.18 24.87 0.75 +DIABETES_PREVALENCE +TOTAL_POPULATION +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -MEDIAN_FAMILY_INCOME* -POPULATION_25_AND_OVER_BACHELORS_DEGREE** -SOURCE_COUNT
Passing Models
AdjR2 AICc JB K(BP) VIF SA Model

****
Choose 7 of 15 Summary
Highest Adjusted R-Squared Results
AdjR2 AICc JB K(BP) VIF SA Model
0.22 407.27 0.45 0.23 20.39 0.69 +DIABETES_PREVALENCE +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -MEDIAN_FAMILY_INCOME** +LANGUAGE_OTHER_THAN_ENGLISH -

FOREIGN_BORN -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -
SOURCE_COUNT*
0.22 407.35 0.45 0.25 24.94 0.69 +DIABETES_PREVALENCE +TOTAL_POPULATION
+POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -
MEDIAN_FAMILY_INCOME* +POPULATION_25_AND_OVER_ASSOCIATES_DEGREE
-POPULATION_25_AND_OVER_BACHELORS_DEGREE* -SOURCE_COUNT
0.22 407.40 0.37 0.23 19.97 0.62 +PHYSICAL_INACTIVITY_PREVALENCE
+POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -
MEDIAN_FAMILY_INCOME** +LANGUAGE_OTHER_THAN_ENGLISH -
FOREIGN_BORN -POPULATION_25_AND_OVER_BACHELORS_DEGREE** -
SOURCE_COUNT*
Passing Models
AdjR2 AICc JB K(BP) VIF SA Model

****
Choose 8 of 15 Summary
Highest Adjusted R-Squared Results
AdjR2 AICc JB K(BP) VIF SA Model
0.22 409.01 0.33 0.77 177.22 0.90 +DIABETES_PREVALENCE +TOTAL_POPULATION -
MEDIAN_FAMILY_INCOME* +POPULATION_25_AND_OVER_ASSOCIATES_DEGREE
-POPULATION_25_AND_OVER_BACHELORS_DEGREE* -SOURCE_COUNT -
POPULATION_DENSITY +SOURCE_DENSITY
0.22 409.03 0.50 0.50 177.86 0.84 +DIABETES_PREVALENCE +TOTAL_POPULATION
+POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -
MEDIAN_FAMILY_INCOME -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -
SOURCE_COUNT -POPULATION_DENSITY +SOURCE_DENSITY
0.22 409.10 0.47 0.46 178.13 0.79 +PHYSICAL_INACTIVITY_PREVALENCE
+TOTAL_POPULATION +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE
-MEDIAN_FAMILY_INCOME* -
POPULATION_25_AND_OVER_BACHELORS_DEGREE** -SOURCE_COUNT -
POPULATION_DENSITY +SOURCE_DENSITY
Passing Models
AdjR2 AICc JB K(BP) VIF SA Model

****
Choose 9 of 15 Summary
Highest Adjusted R-Squared Results
AdjR2 AICc JB K(BP) VIF SA Model
0.22 410.93 0.44 0.59 179.29 0.84 +DIABETES_PREVALENCE +TOTAL_POPULATION
+POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -
MEDIAN_FAMILY_INCOME +LANGUAGE_OTHER_THAN_ENGLISH -
POPULATION_25_AND_OVER_BACHELORS_DEGREE** -SOURCE_COUNT* -
POPULATION_DENSITY +SOURCE_DENSITY
0.21 411.04 0.38 0.23 25.15 0.70 +PHYSICAL_INACTIVITY_PREVALENCE
+DIABETES_PREVALENCE +TOTAL_POPULATION
+POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -

MEDIAN_FAMILY_INCOME** +LANGUAGE_OTHER_THAN_ENGLISH -
FOREIGN_BORN -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -
SOURCE_COUNT
0.21 411.08 0.40 0.51 179.63 0.79 +PHYSICAL_INACTIVITY_PREVALENCE
+TOTAL_POPULATION +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE
-MEDIAN_FAMILY_INCOME* +LANGUAGE_OTHER_THAN_ENGLISH -
POPULATION_25_AND_OVER_BACHELORS_DEGREE** -SOURCE_COUNT -
POPULATION_DENSITY +SOURCE_DENSITY
Passing Models
AdjR2 AICc JB K(BP) VIF SA Model

****
Choose 10 of 15 Summary
Highest Adjusted R-Squared Results
AdjR2 AICc JB K(BP) VIF SA Model
0.22 412.49 0.48 0.55 179.86 0.79 +DIABETES_PREVALENCE +TOTAL_POPULATION
+POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -
MEDIAN_FAMILY_INCOME* +LANGUAGE_OTHER_THAN_ENGLISH -
FOREIGN_BORN -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -
SOURCE_COUNT* -POPULATION_DENSITY +SOURCE_DENSITY
0.21 412.74 0.37 0.50 180.25 0.78 +PHYSICAL_INACTIVITY_PREVALENCE
+TOTAL_POPULATION +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE
-MEDIAN_FAMILY_INCOME* +LANGUAGE_OTHER_THAN_ENGLISH -
FOREIGN_BORN -POPULATION_25_AND_OVER_BACHELORS_DEGREE** -
SOURCE_COUNT -POPULATION_DENSITY +SOURCE_DENSITY
0.21 413.13 0.38 0.58 180.84 0.87 +PHYSICAL_INACTIVITY_PREVALENCE
+DIABETES_PREVALENCE +TOTAL_POPULATION
+POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -
MEDIAN_FAMILY_INCOME* +LANGUAGE_OTHER_THAN_ENGLISH -
POPULATION_25_AND_OVER_BACHELORS_DEGREE** -SOURCE_COUNT -
POPULATION_DENSITY +SOURCE_DENSITY
Passing Models
AdjR2 AICc JB K(BP) VIF SA Model

****
********* Exploratory Regression Global Summary (OBESITY_PREVALENCE) *********

Percentage of Search Criteria Passed
Search Criterion Cutoff Trials # Passed % Passed
Min Adjusted R-Squared > 0.50 28886 0 0.00
Max Coefficient p-value < 0.05 28886 0 0.00
Max VIF Value < 7.50 28886 12064 41.76
Min Jarque-Bera p-value > 0.10 28886 28886 100.00
Min Spatial Autocorrelation p-value > 0.10 21 21 100.00

Summary of Variable Significance

Variable	% Significant	% Negative	% Positive
MEDIAN_FAMILY_INCOME	30.45	100.00	0.00
POPULATION_25_AND_OVER_BACHELORS_DEGREE	21.98	100.00	0.00
DIABETES_PREVALENCE	3.02	0.00	100.00
SOURCE_COUNT	2.80	100.00	0.00
TOTAL_POPULATION	1.60	54.19	45.81
POPULATION_25_AND_OVER_GRADUATE_OR_PROFESSIONAL	1.14	56.88	43.12
POPULATION_DENSITY	1.09	99.47	0.53
SOURCE_DENSITY	0.80	44.59	55.41
PHYSICAL_INACTIVITY_PREVALENCE	0.49	0.07	99.93
FOREIGN_BORN	0.42	93.01	6.99
LANGUAGE_OTHER_THAN_ENGLISH	0.19	38.72	61.28
POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE	0.10	28.82	71.18
POPULATION_25_AND_OVER_ASSOCIATES_DEGREE	0.02	0.00	100.00
UNEMPLOYMENT_RATE	0.01	61.84	38.16
POVERTY_PREVALENCE	0.00	0.17	99.83

Summary of Multicollinearity

Variable	VIF	Violations	Covariates
PHYSICAL_INACTIVITY_PREVALENCE	2.11	0	--------
DIABETES_PREVALENCE	2.38	0	--------
TOTAL_POPULATION	43.85	7007	SOURCE_COUNT (100.00), SOURCE_DENSITY (21.19), LANGUAGE_OTHER_THAN_ENGLISH (21.19), POPULATION_DENSITY (21.19), FOREIGN_BORN (21.19)
POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE	7.42	0	--------
MEDIAN_FAMILY_INCOME	1.13	0	--------
POVERTY_PREVALENCE	1.57	0	--------
LANGUAGE_OTHER_THAN_ENGLISH	22.40	7007	FOREIGN_BORN (100.00), TOTAL_POPULATION (21.19), SOURCE_DENSITY (21.19), POPULATION_DENSITY (21.19), SOURCE_COUNT (21.19)
FOREIGN_BORN	18.01	7007	LANGUAGE_OTHER_THAN_ENGLISH (100.00), TOTAL_POPULATION (21.19), SOURCE_DENSITY (21.19), POPULATION_DENSITY (21.19), SOURCE_COUNT (21.19)
UNEMPLOYMENT_RATE	2.47	0	--------
POPULATION_25_AND_OVER_ASSOCIATES_DEGREE	1.43	0	--------
POPULATION_25_AND_OVER_BACHELORS_DEGREE	7.47	0	--------
POPULATION_25_AND_OVER_GRADUATE_OR_PROFESSIONAL	6.62	0	--------
SOURCE_COUNT	36.83	7007	TOTAL_POPULATION (100.00), SOURCE_DENSITY (21.19), POPULATION_DENSITY (21.19), LANGUAGE_OTHER_THAN_ENGLISH (21.19), FOREIGN_BORN (21.19)

POPULATION_DENSITY 184.07 7007 SOURCE_DENSITY (100.00), TOTAL_POPULATION (21.19), SOURCE_COUNT (21.19), LANGUAGE_OTHER_THAN_ENGLISH (21.19), FOREIGN_BORN (21.19)
SOURCE_DENSITY 169.61 7007 POPULATION_DENSITY (100.00), TOTAL_POPULATION (21.19), LANGUAGE_OTHER_THAN_ENGLISH (21.19), SOURCE_COUNT (21.19), FOREIGN_BORN (21.19)

Summary of Residual Normality (JB)

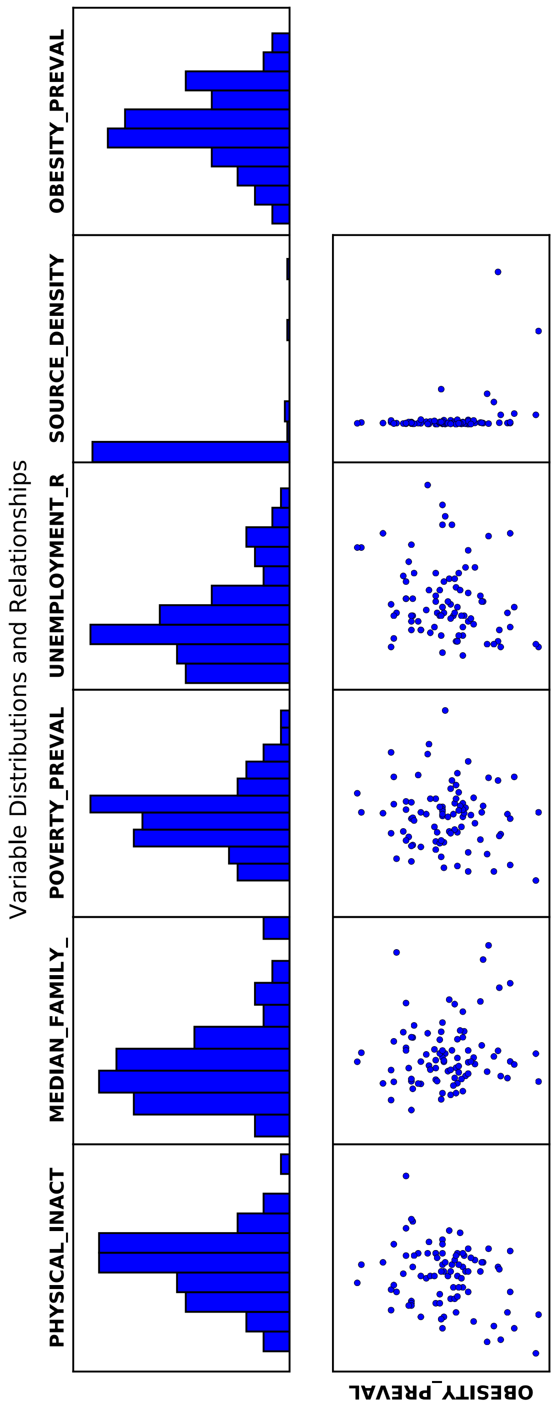
JB AdjR2 AICc K(BP) VIF SA Model
0.999679 0.113295 418.598264 0.559382 21.346142 0.615356 -TOTAL_POPULATION +POVERTY_PREVALENCE -FOREIGN_BORN +UNEMPLOYMENT_RATE +POPULATION_25_AND_OVER_ASSOCIATES_DEGREE -SOURCE_COUNT -POPULATION_DENSITY
0.999625 0.118684 416.670770 0.688669 3.164811 0.615985 +POVERTY_PREVALENCE -LANGUAGE_OTHER_THAN_ENGLISH +UNEMPLOYMENT_RATE +POPULATION_25_AND_OVER_ASSOCIATES_DEGREE -SOURCE_COUNT* -SOURCE_DENSITY
0.999616 0.114588 418.471318 0.639007 14.395781 0.594059 -TOTAL_POPULATION +POVERTY_PREVALENCE +LANGUAGE_OTHER_THAN_ENGLISH -FOREIGN_BORN +UNEMPLOYMENT_RATE +POPULATION_25_AND_OVER_ASSOCIATES_DEGREE -POPULATION_DENSITY

Summary of Residual Spatial Autocorrelation (SA)

SA AdjR2 AICc JB K(BP) VIF Model
0.895426 0.218912 409.013337 0.330985 0.770122 177.218120 +DIABETES_PREVALENCE +TOTAL_POPULATION -MEDIAN_FAMILY_INCOME* +POPULATION_25_AND_OVER_ASSOCIATES_DEGREE -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -SOURCE_COUNT -POPULATION_DENSITY +SOURCE_DENSITY
0.866338 0.209368 413.131510 0.376597 0.578981 180.842188 +PHYSICAL_INACTIVITY_PREVALENCE +DIABETES_PREVALENCE +TOTAL_POPULATION +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -MEDIAN_FAMILY_INCOME* +LANGUAGE_OTHER_THAN_ENGLISH -POPULATION_25_AND_OVER_BACHELORS_DEGREE** -SOURCE_COUNT -POPULATION_DENSITY +SOURCE_DENSITY
0.838562 0.218796 409.026201 0.497854 0.499425 177.860337 +DIABETES_PREVALENCE +TOTAL_POPULATION +POPULATION_25_AND_OVER_HIGH_SCHOOL_GRADUATE -MEDIAN_FAMILY_INCOME -POPULATION_25_AND_OVER_BACHELORS_DEGREE* -SOURCE_COUNT -POPULATION_DENSITY +SOURCE_DENSITY

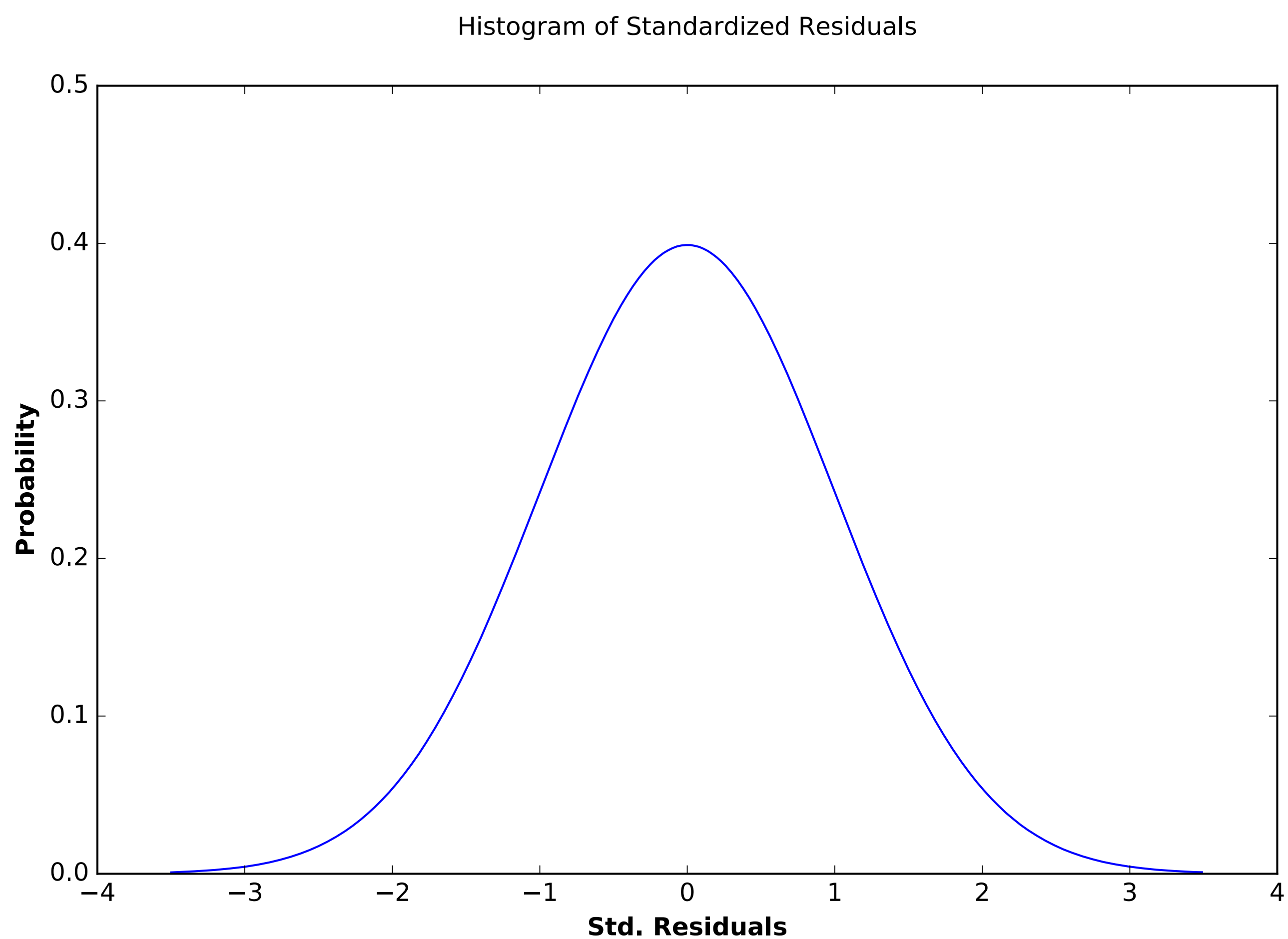
Table Abbreviations
AdjR2 Adjusted R-Squared
AICc Akaike's Information Criterion
JB Jarque-Bera p-value
K(BP) Koenker (BP) Statistic p-value
VIF Max Variance Inflation Factor
SA Global Moran's I p-value
Model Variable sign (+/-)
Model Variable significance (* = 0.10; ** = 0.05; *** = 0.01)

Appendix C Ordinary Least Squares (OLS) Results – Hypothesis

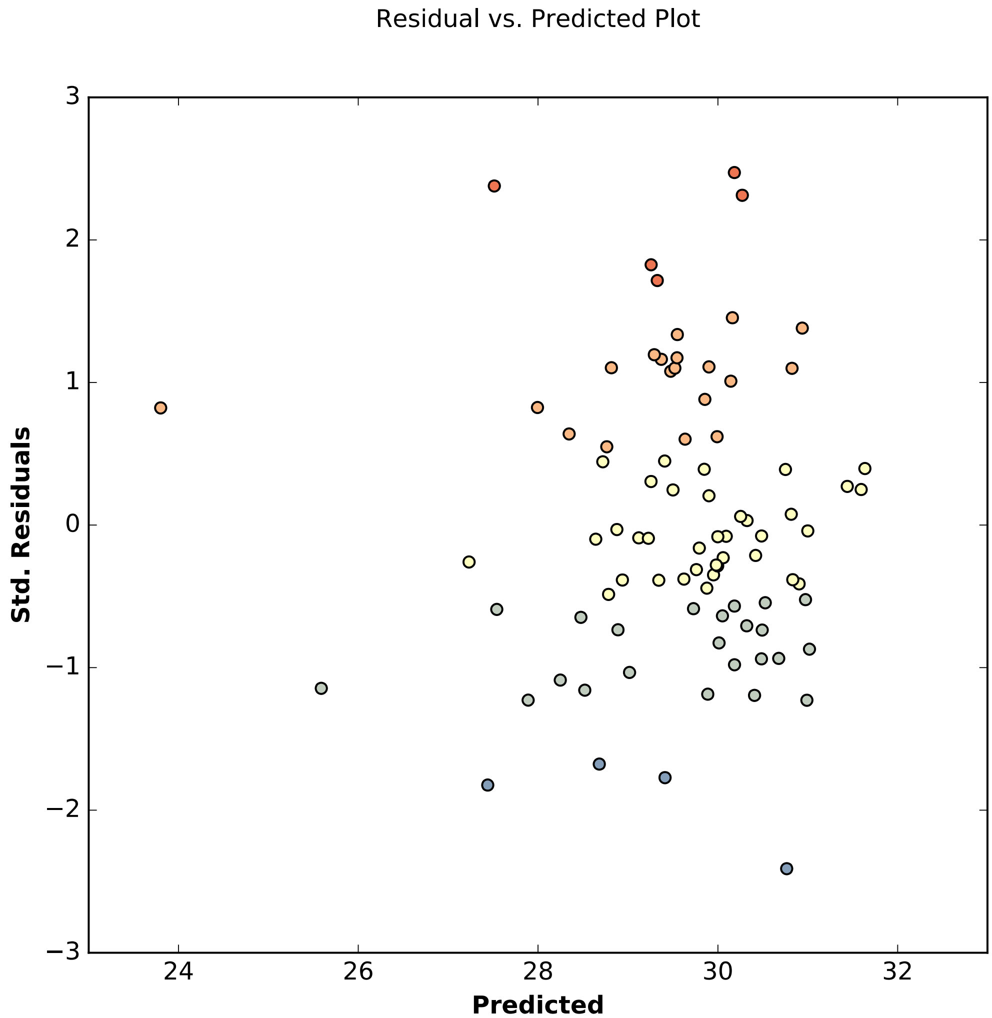


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

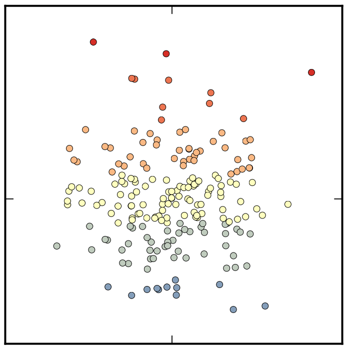
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).

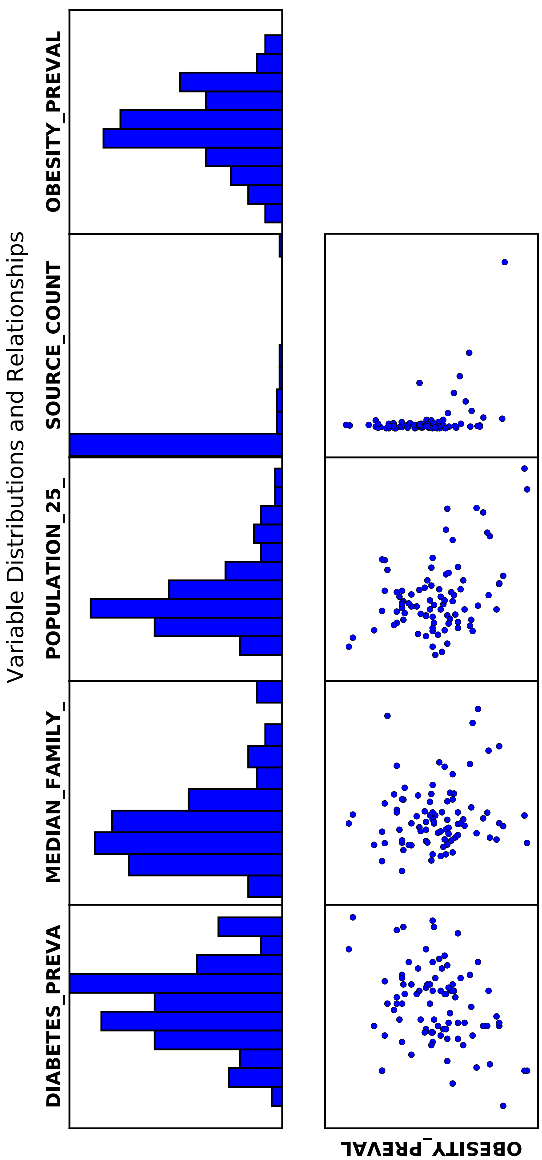


This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



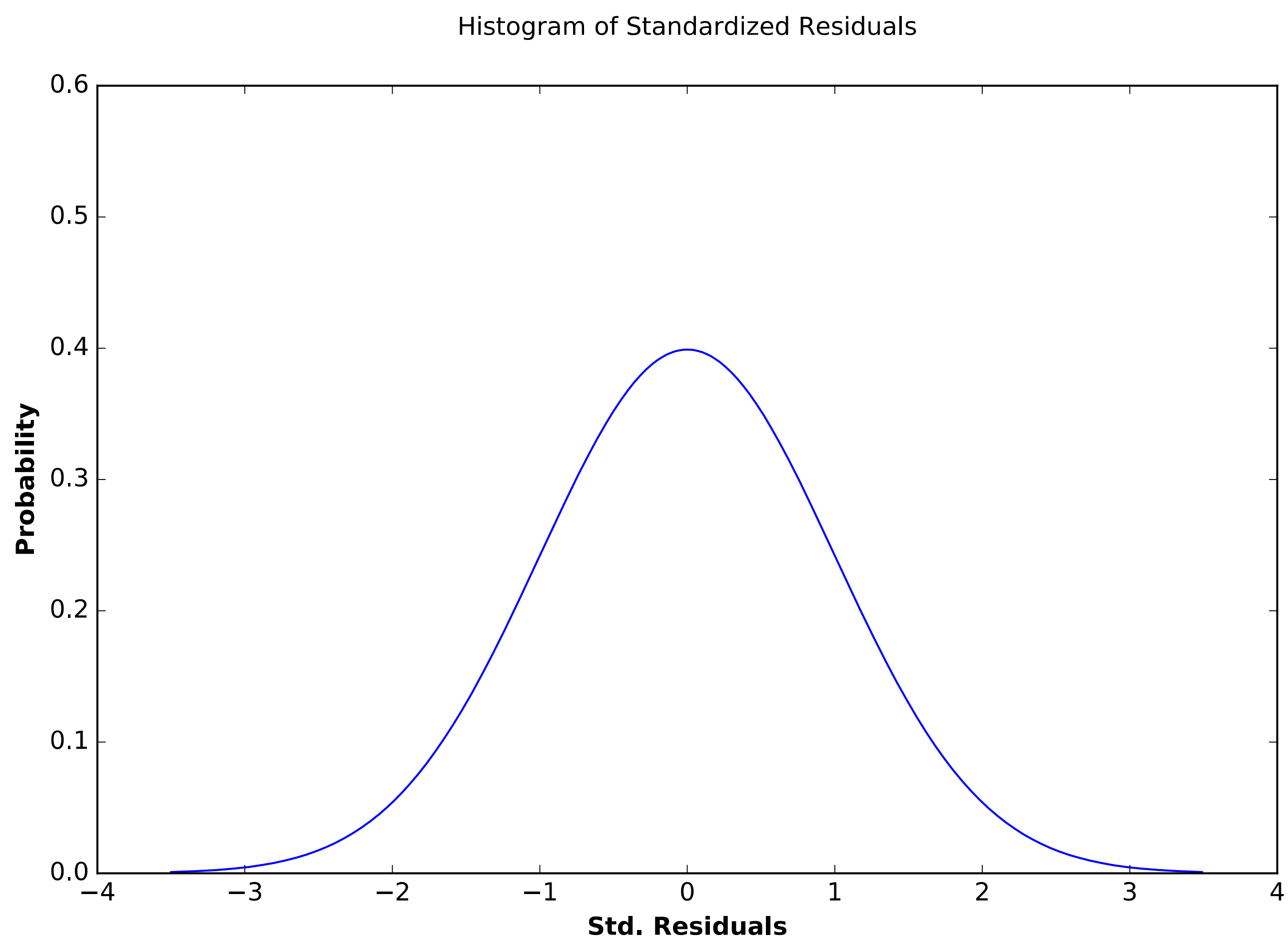
Random Residuals

Appendix D Ordinary Least Squares (OLS) Results – Exploratory Regression



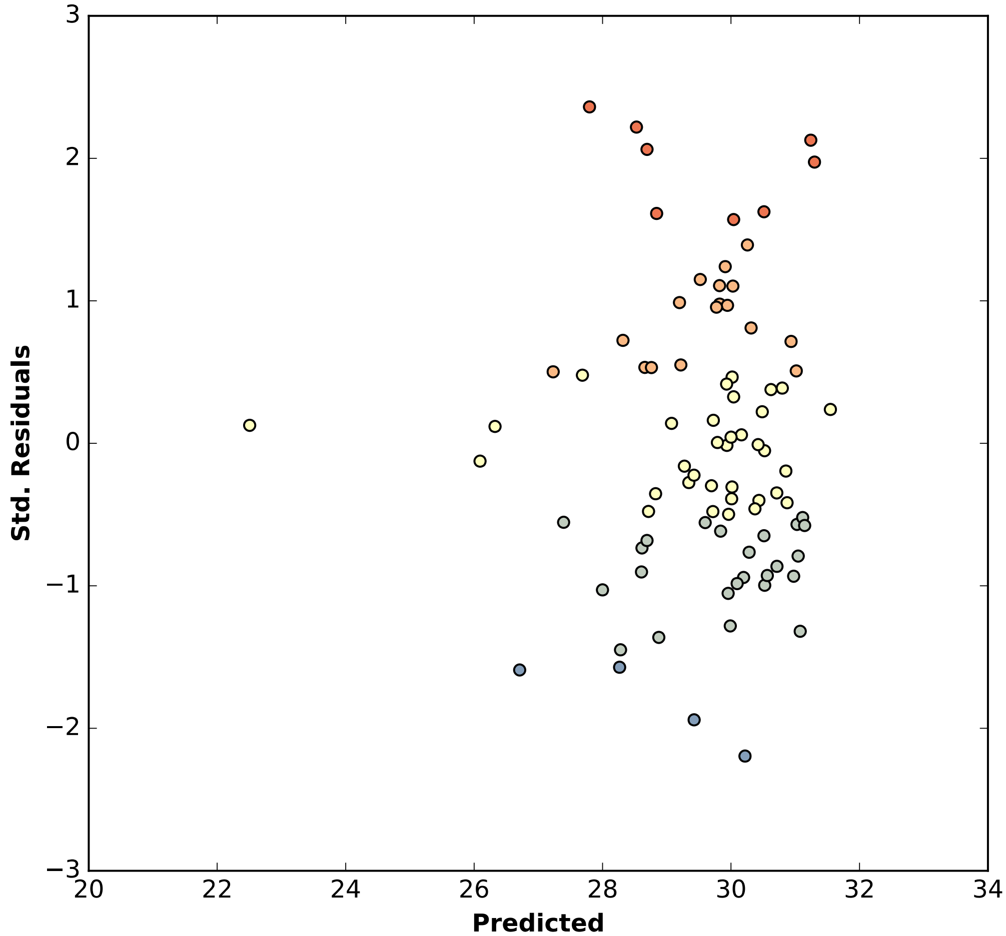
The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.

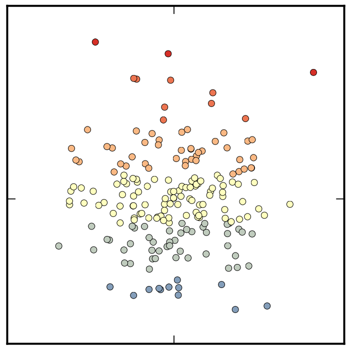


Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).

Residual vs. Predicted Plot



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals