5-2015

Spatial Incidence of Adult Obesity in the State of Alabama, US

Mary G. Odhiambo
University of North Alabama

Spatial Incidence of Adult Obesity in the State of Alabama, US

Mary G. Odhiambo

MS Thesis Committee

Dr. Francis Koti –Chair

Dr. Lisa Keys-Mathews

Dr. Mario Mighty

In partial fulfillment for the degree of

Master of Science in Geospatial Science

Department of Geography

University of North Alabama

May 2015

Table of Contents

List of abbreviations

BMI – Body Mass Index

CDC – Centers for Disease Control and Prevention

GIS – Geographic Information Systems

GWR – Geographically Weighted Regression

OLS – Ordinary Least Square

OBM – Office of Budget and Management

SPSS – Statistical Package for the Social Sciences

AICc – Akaike Information Criterion

AdjR2 – Adjusted R- Squared

List of tables

List of figures

Acknowledgement

I would never have been able to finish my thesis without the guidance of my adviser, committee members, help from friends, and support from my family.

I am using this opportunity to express my gratitude to everyone who supported me throughout the course of my graduate studies at the University of North Alabama. I express my warm thanks to all the faculty and staff in the Department of Geography for their aspiring guidance and support. My deepest appreciation to my adviser Dr. Francis Koti for his invaluably constructive criticism and friendly advice during my thesis research, and to my committee members Dr. Lisa Keys-Mathews and Dr. Mario Mighty for their useful comments, remarks, and engagement through the learning process of this master thesis.

I would also like to thank the Freddie Wood Research Center at University of North Alabama, Department of Geography for providing funding to present my research at the South Eastern Division of the Association of American Geographers (SEDAAG). This was a great experience that will never be forgotten.

A special thanks to my bosses at Benchmark Electronics Inc. Miss Rhonda Bennett, Mr. David Allen and Mrs. Darlene Mangrum. Thank you for being understanding and providing a flexible work schedule so that I could continue with my education.

Most importantly, I would like to thank my family and friends. Words cannot express how grateful I am for your prayer that sustained me this far. And to my fellow graduate students at the University of North Alabama, thank you for your support in writing my thesis and your encouragements to strive towards my goal. You will be missed.

Abstract

In the past few decades obesity has been among the most studied health issues globally. In the United States, studies have indicated that obesity rates are rising in most states with growing evidence that obesity in the US is largely related to economic factors (Chou et al., 2004; Chang et al., 2005; Rosin, 2008). This paper provides an overview and spatial analysis of adult obesity in the state of Alabama. Although research has linked obesity prevalence to different economic factors, other variables are often excluded; hence this study will incorporate factors that are often omitted such as lack of health insurance, physical inactivity, access to recreational facilities, and limited access to healthy food. Demographic, economic, health and environmental data were collected from the US Census bureau 2010 datasets, health and medical data from United States Centers for Disease Control and Prevention, Small Area Health Insurance Estimates (SAHIE),County Business Patterns, and USDA Food Environment Atlas. These data were analyzed using cluster analysis (Getis-Ord GI), Spatial Autocorrelation (Global Moran's I) and Wilcoxon-Mann-Whitney test to assess the role of location in health analysis. Multiple Regression, Global Ordinary Least Square (OLS) and Geographically-Weighted Regression (GWR) were used to determine spatial relationships between variables and location. Analysis indicated that obesity rates are higher in rural than urban counties and also confirmed that there is spatial relationship between socio-economic, demographic, health, and built environment variables although the relationship varies with specific factors and by location.

Chapter 1: Introduction

There is increasing evidence that obesity and overweight in the United States (US) poses a greater health concern than society realizes. The data is are alarming: obesity rates in the US have doubled in the last four decades; more than two-thirds of Americans were classified as overweight in 2010 and during the same year, obesity prevalence in all states had exceeded twenty percent (Hojjat, 2013). The data continue to show that 36 states had a prevalence of 25% or more with 12 of them having prevalence rates over 30% (Alabama, Arkansas, Kentucky, Louisiana, Michigan, Mississippi, Missouri, Oklahoma, South Carolina, Tennessee, Texas, and West Virginia) (Hojjat, 2013).

Despite efforts in obesity research and education programs, there is still limited focus on place-based approaches to the obesity phenomena. Studies that focus on causes of obesity have tended to link obesity disparities with socio-economic status such as inequality in household income, education attainment, and unemployment among other factors (Nayga, 2001; Pickett, Brunner & Wilkinson, 2005; Rosin, 2008). Others seem to indicate that obesity rates tend to differ significantly between race, gender, and ethnic backgrounds, as well as geographic regions (Peytremann, Faeh, & Santos, 2007; Chen & Truong, 2012; D'Agostino, Gennarelli, Lyons, & Goodman, 2013; Le et al., 2014).

This study focuses on the importance of location in understanding obesity. The overarching goal is to offer a place-based approach to understanding obesity – especially in rural Alabama.

1.1 Problem Statement and Research Objectives

Contemporary obesity research is multidisciplinary in nature, drawing largely from but not limited to: biology, psychology, epidemiology, geography, sociology – to name but a few. Most of these studies focus on the causes, effects, and preventive measures of obesity. While most of these studies identify the role of economic factors in the prevalence of obesity, there is still a notable bias in urban America and a limited focus in rural communities. This study examines adult obesity in the state of Alabama with a focus on the importance of place in understanding obesity.

The study objectives are to:

- Establish the relationship between socio-economic and demographic factors and obesity across Alabama counties.
- Assess the role of place in understanding obesity.

1.2 Study Area

This study focused on the state of Alabama in the American South. With a total population of 4,779,736 in 2010 (Census 2010), Alabama is divided into 67 counties, 29 of which are categorized as urban and 38 as rural. Fifty nine percent of the state population is considered urban while 41 percent live in rural areas. More than half of the rural counties are within the Alabama Black Belt Region (Figure 1) in the south-central part of the State. Historically, the Black Belt Region has been known for its fertile black clay soil and large share of African American population (Jeffries, 2009). The area is also characterized by various aspects of socio-economic depression such as poor education, income below poverty, and high rates of unemployment (Geronimus, Bound, Waidmann, Hillemeier, & Burns 1996; Carter, Vivian, Dawkins & Howard, 2010).

The term Black Belt has been perceived by many scholars to have two meanings, one referring to the trans-south band rich black topsoil and the other referring to the concentration of African American population still living in those former cotton-growing counties (Bliss, Howze & Teeter, 1993). Black belt counties consistently rank last in the state and nation's per capita income with their economies revolving around low wage cotton production (Jeffries, 2009). Gibbs, 2003:257) states that "very low level of human capital are the underlying limiting factors in the region's growth and development" , indicating that adult academic attainment is highly racially uneven with whites' academic completion rates and participation in the labor force higher than the blacks.

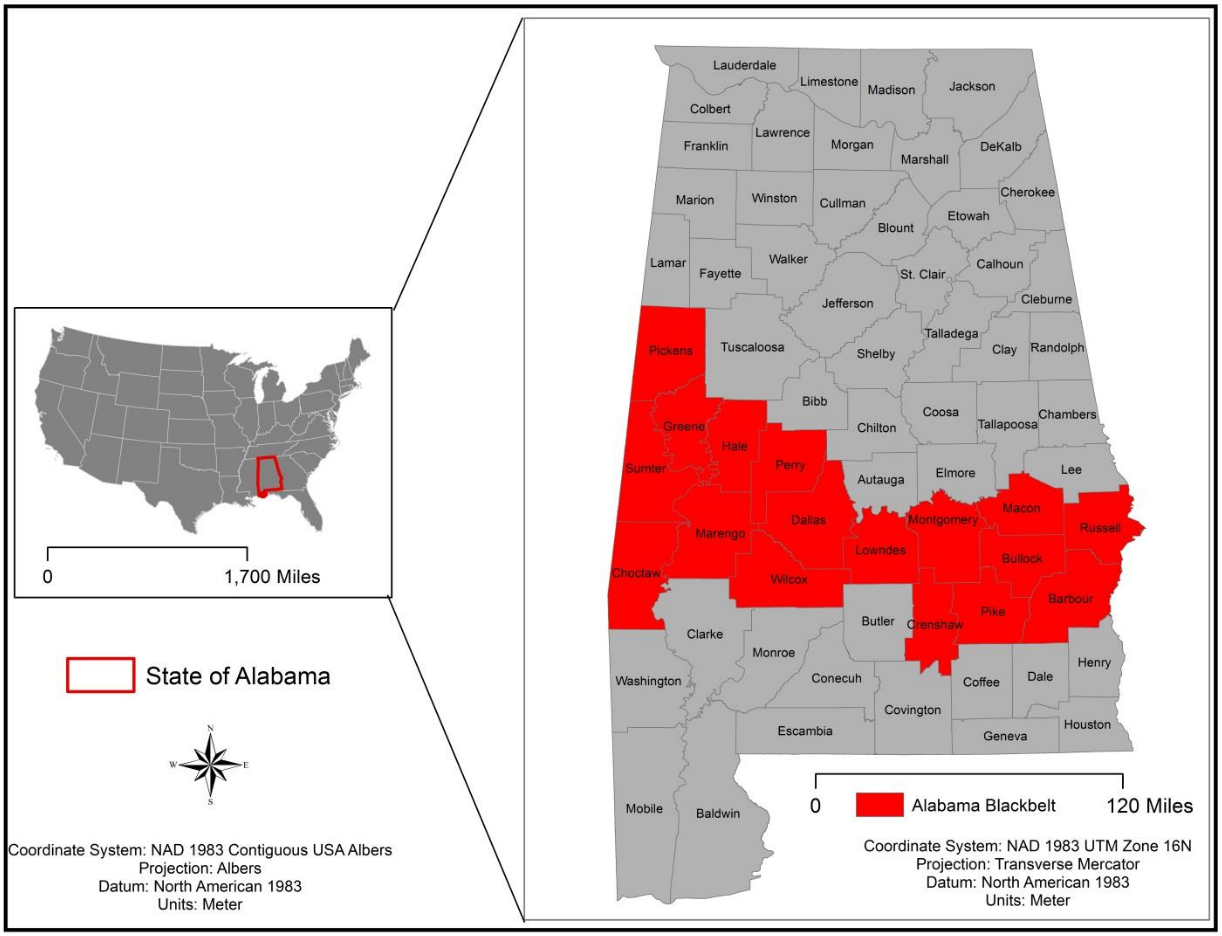


Figure 1: Study Area, the state of Alabama.

The State of Alabama is selected for this study for various reasons including:

1. Disparities in levels of economic development between the urban and rural counties.
2. History of regional disparities especially in the Black Belt area in south-central part of the Sate.
3. High prevalence of adult obesity rates reported by the Centers for Disease Control and Prevention (CDC) which indicates that that Alabama is among the states with high prevalence (more than 30%) of the adults being obese / overweight (CDC 2010).

4. Living in the area affords first-hand knowledge and also facilitates accessibility of data and resources.
5. Limited attention in the literature and public debates.

1.3 Significance of Study

The effects of obesity are diverse, including the fact that obese individuals are susceptible to weight-related illnesses, and other life-threatening diseases. Figures show that, obesity is related to 300,000 premature deaths per year in the United States - which is higher than deaths related to alcohol and illegal drug use. This makes obesity the second leading preventable cause of death in the United States (Hojjat, 2013). Direct medical costs associated with obesity are argued to be as much as 100 percent higher than for healthy weight adults, and nationwide medical spending may amount to as much as $147 billion annually for adults and $ 14.3 billion annually for children (Hojjat, 2013).

Unfavorable socioeconomic status and/ or environmental conditions often translate into regional disparities in obesity rates (Peytremann et al., 2007). In the United States for example, some studies suggest that obesity prevalence may be higher in rural than urban areas. These studies cite economic disparities, differential access to opportunities for physical activity, healthy nutrition, and health care; lower standards of living, and few opportunities of employment among other factors. This gap presents an opportunity for a place-based study and approach that addresses variations in obesity incidence across space. Such an approach may help to tailor public health interventions directed to the management and prevention of obesity (Peytremann et al., 2007; Jokela et al., 2009). Investigating local and regional disparities in obesity rates – especially at the county level is important for four main reasons. First, disparities in level of socio-economic development have resulted in the societal cost of obesity being higher for certain areas than others (Peytremann et al., 2007). Secondly, policies and strategies to reduce health disparities are often implemented at the national and state levels – far removed from the local scale where individual health outcomes are realized (Peytremann et al., 2007). Third, studies such as (Peytremann et al., 2007; D’Agostino et al., 2013) have recommended a shift in focus to the local, especially county-level, indicating that analyzing health issues at this level may accelerate progress in reducing health disparities. Finally, large-scale data such as those at the census tract, or block group level would be more useful because health disparities are always experienced at such large scale data. Unfortunately, there is limited large-scale data sufficient for effective evaluation of public health policy, programs, and interventions that occur at the local level (D’Agostino et al., 2013).

1.3: Definitions and Terminologies:

Geographically-Weighted Regression - Geographically Weighted Regression is a statistical techniques used to examine the spatial variability of regression results across a region to inform on the presence of spatial nonstationarity (Brunsdon, 1998).

Geographic Information Systems - A Geographical Information System (GIS) constitutes a system of hardware and software used for storage, management, retrieval, manipulation, analysis, modeling, and mapping of geographical data (Aimone, Perumal & Cole, 2013).

Medical geography – Medical geography is a sub-discipline that studies public health using the concepts, theories, methodologies, and perspectives of the discipline of geography to analyze spatial patterns of disease, their relationships to the natural and social environment, and their expression in the health of people in places (Meade, 2010).

Metropolitan – Metropolitan statistical area (MSA) is defined as an entity that contains a core urban area of 50,000 or more population (US Census Bureau).

Micropolitan – Micropolitan statistical area is defined as an entity with an urban core of at least 10,000, but less than 50,000 populations people (US Census Bureau).

Multiple regression – Multiple regression is a statistical analysis method used to model relationships among data variables associated with geographic features, allows for data examination, exploration and better understanding of key factors influencing the variable being modelled. Regression also verifies that relationships exist and measures strengths of those relationships (Crawley, 2005).

Overweight and obesity – Overweight and Obesity are term used for ranges of weight that are greater than what is generally considered healthy for a given height, also used to identify ranges of weight that has been shown to increase the likelihood of certain diseases and other health problems (CDC).

Rural counties – Rural counties are counties that are not designated as parts of metropolitan areas are considered rural (Office of Budget and Management)

Spatial analysis – Spatial analysis is the ability to manipulate spatial data into different forms and extract additional meaning as a result. In medical geography, spatial analysis involves quantitative study of disease distribution and the pattern of health care and service availability (Miller & Wentz, 2003)

Spatial epidemiology – Spatial epidemiology is the description and analysis of geographically-indexed health data with respect to demographic, environmental, behavioral, socioeconomic, genetics, and infectious risk factors (Elliott & Wartenberg, 2004).

Urban counties – Urban counties are counties that are designated as parts of metropolitan statistical area (OBM).

Chapter 2: Review of the Literature

This study draws from and is also informed by three bodies of literature: Obesity in society, Geographic information Systems (GIS) and Spatial Analysis in Medical Geography, multiple regression and Geographically Weighted regression (GWR). These are discussed below.

2.1 Obesity in Society: Causes and Effects

According to the (CDC) Overweight and obesity are health conditions that are defined by ratio of weigh to height. The two measures are used to calculate body mass index (BMI) that is used to determine if a given weight is healthy for a given height. Overweight and obesity also identify ranges of weight that have been shown to increase the likelihood of certain diseases and other health problems. For adults BMI between 25 and 29.9 is considered overweight and BMI of 30 or higher is considered obese (ibid).

The history of obesity dates back to more than 30,000 years ago but it was rarely recognized and rarely studied (Haslam, 2007). It was in the early 1600's when it first gained medical attention and from late 1600 that it was linked with other diseases. Contrary to today, obesity at that time was viewed as a sign of high status and wealth in various cultures (Must & Strauss, 1999).

In the past few decades, obesity and obesity-related illnesses have been one of the most studied health issues and its rising trends have raised concerns for close monitoring. Wang, Beydoun, Liang, Caballero & Kumanyika, 2008) for example estimated the progression and cost of the US obesity epidemic and their results indicated that obesity and overweight in adults is increasing faster than in children, and in women than in men. They (ibid) continue to argue that if these trends continue, by 2030, 86.3 percent of adults will be overweight or obese, and by 2048, all American adults would become overweight or obese. Their study also demonstrated that the total health-care costs attributable to obesity/overweight would double every decade to 860.7–956.9 billion US dollars. Hojjat (2013) has similarly indicated that obesity is related to 300,000 premature deaths per year in the United States which is higher than deaths related to alcohol and illegal drug use – making it the second leading preventable cause of death in the country.

The health, economic, environmental, and social implication of obesity in society have also been well documented (Nayga, 2001; Linne et al., 2004; Chou et al., 2004; Pickett 2005; Drewnowski et al., 2005; Gordon et al., 2006; Liese et al., 2007; Papas, 2007; Rosin, 2008; Whiteman et al., 2008; Healy et al., 2008; Salome et al., 2010). Some studies have shown that the obesity epidemic is strongly related to a wide range of behavioral and lifestyle factors (Drewnowski et al., 2005; Healy et al., 2008). Others link obesity to social and environmental factors citing: high consumption of fast food and foods prepared away from home, increase to hereditary pursuits such as television viewing, the use of computer and other forms of electronic entertainment, reduction in walking and cycling as a means of transportation, increase in availability and marketing of food, and reduction in physical education in schools (Chou et al., 2004; Gordon et al., 2006; Liese et al., 2007; Papas, 2007; Healy et al., 2008). While (Boutin et al., 2001; Speakman et al., 2004) link obesity to genetic factors, (Chou et al., 2004; Chang et al., 2005; Rosin, 2008) associate obesity to economic factors.

There is an increasing body of literature that has also focused on the relationship between obesity and different health conditions including cardiovascular disease, coronary artery heart disease, diabetes, cancer, increased morbidity, and mortality and with children; for example, obesity is known to cause hypertension, dyslipidemia, chronic inflammation, increase in blood clotting and hyperinsulinaemia (Nayga, 2001; Linne et al., 2004; Salome et al., 2004; Whiteman et al., 2008; Salome et al., 2010). These studies continue to show a variety of other medical complications that can be directly linked to obesity. Stereotyping obese individuals is also said to have negative social impacts especially on children. For instance, obese children stereotyped as unhealthy, academically unsuccessful, socially inept, unhygienic, and lazy, and also uniformly ranked by other children as the least desired friends. Obese girls were observed to have obsessive concern with body image as well as expectation of rejection and progressive withdrawal (Must, 1999; Carr, 2005).

In the workforce, evidence shows that obese employees are considered to have greater rates of absence with statistics indicating that in the United states the cost of obesity among obese employees amounts to $73.1 billion per year with 18 % due to sick days, 41% due to lack of productivity from health issues and 41% due to medical expenses, (Finkelstein et al., 2011). Scores of other scholars have also examined obesity mitigation and preventive measures (Coleman, 2007; Weiss et al., 2010; Cassel et al., 2010; Gillman et al., 2013). To help manage obesity rates Bogart, (2013) illustrates that

> At the moment we have a set of norms buttressing stigma towards and discrimination against fat people. Somehow we need to shift to norms that encourage nutritious eating and drinking, active lifestyles, and a fundamental acceptance of bodies of many shapes and sizes. For that transformation to occur there will need to be great societal change. If we get things more or less right, law can have a role: a complicated and limited one. Perhaps law can even do something for that child in Atlanta. As a start, let’s ask: which is the bigger problem — her chubbiness or the way society treats her? (Bogart, 2013:38)

2.2 GIS and Spatial Analysis in Medical Geography

Geographic information systems (GIS) are “automated systems for the capture, storage, retrieval, analysis, and display of spatially-referenced data (Miller & Wentz, 2003:575). Spatial analysis (SA) on the other hand refers to the “…ability to manipulate spatial data into different forms and extract additional meaning as a result.” (Miller & Wentz, 2003:575). In medical geography, spatial analysis involves the quantitative study of disease distribution and including patterns of health care and service availability (Clarke, McLafferty & Tempalski, 1996; Miller & Wentz, 2003).The primary methodological approach for both GIS and SA is quantitative analysis and both share geographic location as a central organizing principle, with the goal of enhancing understanding of geographic phenomena and solving geographic problems (Clarke, McLafferty & Tempalski, 1996; Miller & Wentz, 2003). Elliott & Wartenberg, 2004: 998 notes that “Advances in GIS and statistical methodologies together with the availability of high-resolution, geographically-referenced health databases present unprecedented new opportunities to investigate the environmental, social, and behavioral factors underlying geographic variations in disease rate, improving on the traditional reporting of diseases at national or regional scale.”
As a field of study, spatial analysis of diseases dates back to the 1800s when different maps showing spread, causes and outbreaks of diseases begun to emerge from different countries (Elliott & Wartenberg, 2004). The first mapping in epidemiology was in 1854, when Dr. John Snow identified the broad street pump as the source of an intense cholera outbreak by plotting the location of cholera deaths on a dot-map. Since then, many scholars have incorporated Geographic Information Systems (GIS) and spatial analysis techniques in their studies, a “concept that has been missing in health literature recently” (Chen et al., 2012). Advantages of

using GIS to investigate global health issues have been captured in (Aimone et al., 2013:2) when the authors indicate that GIS:

> 1) allows the exploration of the role of geographical or environmental factors in the prevalence or incidence of a health outcome of interest; 2) the combination of cartography and multivariate analysis allows investigation of complex spatial relationships (e.g. linking people and health outcomes to space and time); 3) GIS software enables the presentation of research findings in a visual manner that can be easily interpreted across disciplines; and 4) the technique can be applied to a range of analysis units, which may provide insight into relationships between health outcomes and other social, demographic, or economic variables at various jurisdictional level. (Aimone et al., 2013:2)

Some scholars such as Jacquez et al., (2000) have indicated that application of GIS in studying public health has not been successful as expected due to “lack of spatial knowledge to effectively demonstrate unique and substantial contributions of GIS in epidemiology, and failure of commercial off-site-self GIS to provide appropriate tools for spatial epidemiology” (Jacquez et al., 2000:92).

Spatial analysis provides researchers with different methods of studying phenomena. These include visualization, exploratory, and quantitative modelling methods. These capabilities of spatial analysis have been identified to allow examination and display of health data effectively. Visualization assists in indicating change in disease distribution and pattern over a given period of time. Exploratory analysis enables extraction of meaningful information from the data to help formulate hypotheses for future research. Modelling on the other hand includes procedures for hypotheses testing on causes of diseases, their nature and disease transmission (Clarke, McLafferty & Tempalski 1996; Gatrell & Bailey 1996; Bhatt & Joshi, 2012).

Among scholars that have used spatial analysis to study health related issues, Comber, Brunsdon & Radburn, (2011) used spatial analysis to analyze variations in health access using regression analysis. They concluded that difficulty in accessing different health facilities was found to be significantly related to health status (Ibid: 9). Aimone, Perumal & Cole, (2013) reviewed the application and utility of geographical information systems in exploring disease relationships and indicated that " investigation of geographic relationships with specific health outcomes has extended beyond simply mapping and describing spatial distribution patterns, to more complex analyses and predictive modelling to incorporate the effect of other environmental and spatial factors, such as regional variations in climate and distributions in population density" (Ibid: 11) . Koch & Denike, (2001) analyzed GIS approaches to the problem of disease clusters. In their study they emphasized on the importance of using cartographic solution in medical cartography in general, and GIS-based mapping in particular. Their study indicated that there are significant advantages in using fundamental cartographic approach to the problem of disease clusters. Sui, (2007) presented a review on interaction between GIS and medical geography. This study discussed the need for a better synergy between the two fields and it indicated that GIS applications are important and have contributed to the rapid growth of medical geography in recent years. The study also showed that advances in medical geography can also have significant implications on the future development of GIScience (Sui, 2007: 573).

Elliott & Wartenberg, (2004) have discussed some challenges of spatial analysis in epidemiology. Their research indicated that these challenges include data availability and quality, data protection and confidentiality, exposure assessment, exposure mapping and study design issues. To overcome the challenges they suggested studies to be guided by well stated questions, excellent statistical methodologies, and sound epidemiology principles including taking proper account of problem of data quality and the potential for bias and confounding.

2.3 Multiple Regression and Geographically Weighted Regression (GWR)

Spatial analyses are commonly used to study relationships between place-level disadvantages and health outcomes. Scores of scholars have turned to multiple regression analysis as a preferred statistical method for modelling cause and effect relationships. Allison, (1999) explains multiple regression as a statistical method for studying the relationship between single dependent variable and one or multiple independent variable for causal and prediction analysis. Regression analysis is sometimes referred to as ordinary least square multiple linear regressions and is expressed as

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots. + \beta_n X_n + \varepsilon}$$

Where y = dependent variable

X_1 to X_n = independent variable

b_0 to b_n = regression coefficients

$_{e\,=}$ a random error.

Least square is the method used to estimate the regression equation, multiple is a term that indicates the use of more than one independent variable and linear describes the equation used by multiple regression method. Multiple regression is widely used because it is user-friendly compared to other statistical methods such as logistic regression, Poisson regression and structural equation models. It enables combination of many variables to produce most favorable predictions of the dependent variable. It also separates the effects of independent variables on the dependent variable for easy examination of every variable contribution (Paul, 1999; Charlton, 2009).

Errors in regression analysis are related to measurement errors, sampling error and uncontrolled variation (Berry, 1993). Regression assumptions and model selection criterions are used to minimize errors in the analysis. These assumptions include test for linearity and additivity of the relationship between dependent and independent variables, statistical independence of the errors, homoscedasticity of the errors, and normality of the error distribution (Berry, 1993)

In this study two types of regression models were used. A global regression model (Ordinary Least Square Regression) was used to examine the relationship between socio-economic and demographic factors and obesity across Alabama Counties and a local regression model (Geographically Weighted Regression) was used to analyze spatial variation of the relationship. The difference between the two models is that, OLS model predicts the response coefficient from a linear predictor generated from the independent terms assuming that variables are stationary over geographic space, while GWR allows test for variables to vary over geographic space (Comber et al., 2011). GWR is described by the equation

$$\text{Eq (1)} \qquad Y_i = \mathbf{X}_i^t \beta(u_i, v_i) + \varepsilon_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} X_{ik} \beta_k(u_i, v_i) + \varepsilon_i$$

Where

β(ui,vi) - Indicates the vector of the location-specific parameter estimates

(ui,vi) - Represents the geographic coordinates of location i in space, and is the error term with mean zero and common variance σ2.

Excluding the geographic coordinates, (ui,vi), will make the GWR equation a multiple regression Other spatial analysis technique used in this study includes cluster and hot spot analysis which are used to analyze spatial patterns.

In a study by Chen & Truong (2012) a multilevel modeling and geographically weighted regression was used to identify spatial variations in the relationship between place-level disadvantages and obesity. Pal & Bhattacharya (2013) adopted cluster analysis and multiple regression analysis in a case study on the financial health of the main steel Producing segment in India and (Kirby et al, 2012) used series of linear regression models to determine complex relationships among community racial/ethnic composition, individual race/ethnicity, and obesity in the United States.

The bodies of literature surveyed here provide a snapshot of the state of knowledge in obesity studies in general. The review reveals the multidisciplinary nature of obesity research most of which tend to focus heavily on causes, effects and preventive measures. A notable bias is also noted on economic, social, genetic, and environmental arguments regarding obesity prevalence.

A common thread in the literature is the limited engagement of place-based factors in understanding the obesity phenomena. For example, the literature on obesity studies in rural communities is limited. This research hopes to fill in these gaps.

Chapter 3: Research Questions, Methods, and Data

3.1 Research Questions

This study seeks to answer two main research questions:

1) **Is there a significant relationship between geographic location and the incidence of obesity at the county level in Alabama?**

The focus of this question is to determine if adult obesity rates differ between urban and rural counties. Geographic location has been identified to play a major role in health outcomes and many studies such as (Brown, Young & Byles, 1999; Strong et al., 2001; Andrews, Henderson & Hall, 2001; Duncan et al., 2009). These same studies have indicated that individuals living in rural areas experience variety of health disparities than those in urban areas. Given the established relationships between geographic location and health disparities, it is useful to consider the role geographic location plays in adult obesity in the state of Alabama. To answer these questions three methods were: Hot Spot Analysis (Getis-Ord GI), Spatial Autocorrelation (Global Moran's I) and Wilcoxon-Mann-Whitney test.

2) **Is there a statistical relationship between socio-economic, demographic, health and environmental variables and obesity at the county level in Alabama?**

This question seeks to identify spatial relationships and patterns in the data set. By understanding the spatial relationship between socio-economic, demographic, health, and environmental variables and obesity at the county level, it is possible to identify factors that best explain adult obesity prevalence in the state of Alabama. Knowledge of the spatial distribution across different scales may help policy makers in identifying intervention measures for specific geographic locations. This question will be answered using multiple linear regressions: Global Ordinary Least Square (OLS) and Geographically Weighted Regression (GWR) methods.

3.2 Methods

This study employs quantitative methods to examine the relationship between geographic location and obesity incidence at county level (detailed discussion of multiple regression and regression analysis in a previous section). The dataset included 29 independent variables that are directly related to the prevalence of adult obesity - the dependent variable in the study.

These datasets were classified into variables for preprocessing. All 29 variables were standardized into Z-scores in SPSS statistical software to insure comparability of variables. All the data were aggregated in the form of percentages and rates to aid in the minimization of errors. Data were assembled in ArcGIS and R programming environment where the following analyses were run: Cluster analysis, hotspot analysis, exploratory regression, multiple linear regression analysis, and Wilcoxon-Mann-Whitney test.

3.3 Data

County-level geodemographic data, health, socio-economic and built environment information were obtained from various sources. County level line shapefiles for the state of Alabama were obtained from US Census Bureau TIGER products that contain spatial data for use in GIS. The counties were categorized into urban and rural based on criteria used by the Office of Management and Budget (OBM) definition (Figure 2). OBM defines as “rural” all counties that are not designated as parts of a metropolitan area. A metropolitan area is defined as an entity that contains a core urban area of 50,000 or more population, while a micropolitan as an entity with an urban core of at least 10,000, but less than 50,000 people (US Census Bureau). This definition is also one of the two methods often used by the Office of Rural Health Policy to determine geographic eligibility for its grant programs (Health Resources and Services Administration).

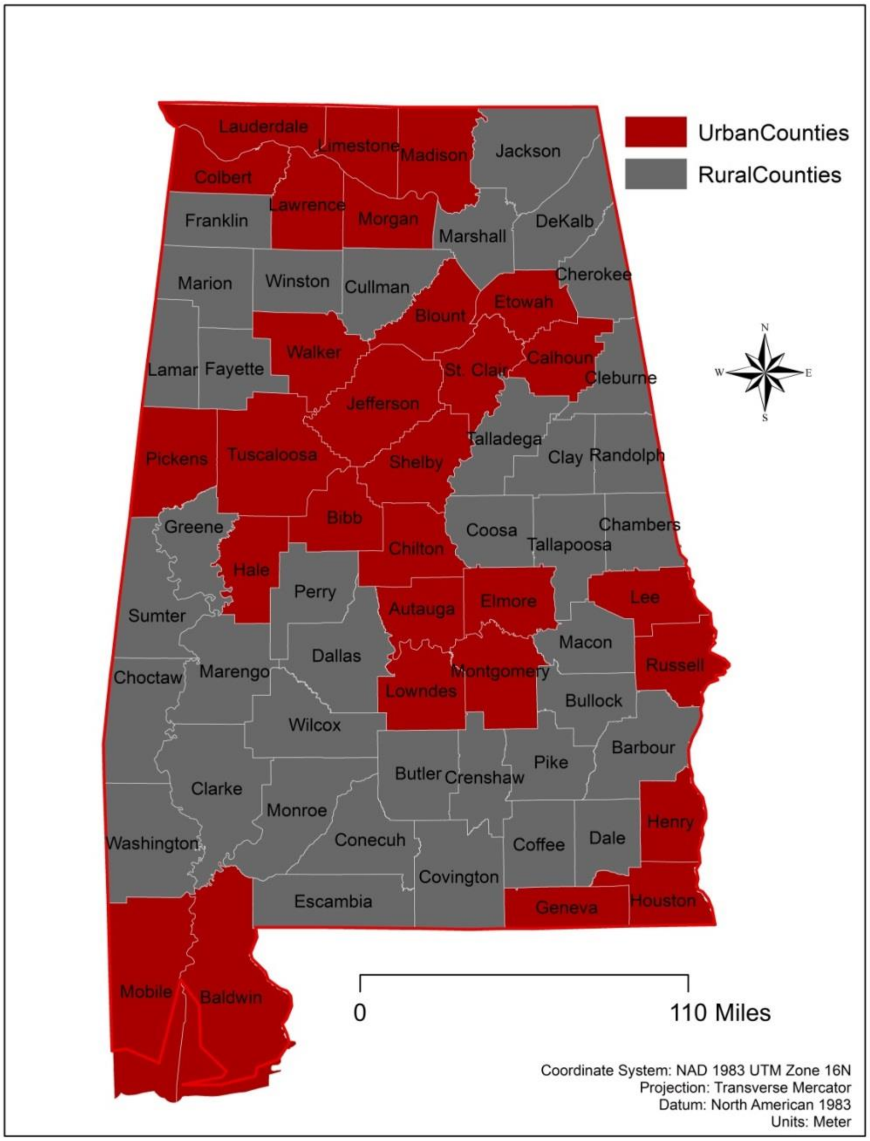


Figure 2: Classification of Alabama counties into urban and rural based on OBM definition of counties.

As the Figure 2 shows, of the 67 counties in the Alabama, 38 (56.7%) are classified as rural while 29 counties are classified as urban. Interestingly, most of the rural counties lie within the Alabama black belt region (Figure 1).

Health data were obtained from two different databases: Behavioral Risk Factor Surveillance System and Small Area Health Insurance Estimates (SAHIE). Data include: Adult obesity rates, female obesity rates, male obesity rates, number of adults with no insurance and physical activity rates (Table 1). Built environment data were obtained from County Business Patterns and the United States Department of Agriculture (USDA) Food Environment Atlas databases and includes variables such as access to recreational facilities, access to healthy foods and 2009 fast food rates (Table 3). Other datasets used include demographic, socioeconomic and age groups data from the United States Census Bureau (Table 2, 4 and 5),

3.3.1 Variable in the Study

Data	Sources
Adult obesity rates	Behavioral Risk Factor Surveillance System
Female obesity rates	Behavioral Risk Factor Surveillance System
Male obesity rates	Behavioral Risk Factor Surveillance System
Female 18-65 uninsured	Small Area Health Insurance Estimates (SAHIE)
Male 18-65 uninsured	Small Area Health Insurance Estimates (SAHIE)
Female Physically inactive	Small Area Health Insurance Estimates (SAHIE)
Male physically inactive	Small Area Health Insurance Estimates (SAHIE)

Table 1: Health Data

Data	Sources
Male population	United States Census Bureau
Female population	United States Census Bureau
White population	United States Census Bureau
Black population	United States Census Bureau
Hispanic population	United States Census Bureau
Asian population	United States Census Bureau
Native Hawaiian and other Pacific Islanders	United States Census Bureau

Table 2: Population Data

Data	Sources
Access to recreational facility	County Business Patterns
Limited access to healthy foods	USDA Food Environment Atlas
2009 Fast food rates	County Business Patterns

Table 3: Built Environment Data

Data	Sources
Agesunder 5	United States Census Bureau
Ages 5 to 85	United States Census Bureau
Ages 85 and above	United States Census Bureau

Table 4: Age Groups Data

Data	Sources
Education high school and above	United States Census Bureau
2010 poverty rates	United States Census Bureau

Table 5: Socioeconomic Data

Chapter 4: Analysis

4.1 Exploratory Regression

Exploratory regression was used to determine the most important variables; this method systematically quantifies the relative importance of variables. Exploratory analysis helps to improve accuracy by minimizing human error as even perception of expert can be misleading (Braun & Oswald, 2011). In ArcGIS, exploratory regression was applied to variables Z-scores. Out of the 29 independent variables, 24 variables were identified as important predictors (Table 2); four variables were eliminated due to multicollinearity (a state of very high intercorrelations or inter-associations among the independent variables) and four additional variables were removed because their coefficients were not significant. Multicollinearity was determined using Maximum Variance Inflation Factor that "reports how much the variance of the estimated coefficients increase is due to collinear independent variables" (Craney & Surles, 2002:392).

Explanatory Variable (Z-scores)	**VIF**	**VIF violations**	**Significance**
Male obese	1.75	0	100.00
Poverty rates	1.85	0	100.00
Hispanics	2.17	0	89.90
Limited access to healthy foods	3.10	0	85.78
Ages 35_44	4.09	0	75.84
High school education and higher Ed	1.94	0	55.80
Ages 65_74	3.08	0	55.33
Ages 5_14	2.30	0	46.32
Ages 15_24	6.00	0	46.21
Asian	4.30	0	45.34
Ages 25_34	3.00	0	45.18
Ages 45_54	2.44	0	32.46
Access to recreational facilities	7.13	0	16.331
Ages under 5	3.32	0	14089
Ages 75_84	1.46	0	10.46
Ages 85 and over	3.19	0	6.49

Table 6: Selected Variables from Exploratory Regression.

The table above contains twenty four variables that were identified as important predictors. The variables were selected based on their Max Variance Inflation Factor scores from exploratory analysis.

Explanatory Variable (Z-scores)	**VIF**	**VIF violations**	**Significance**
Female physically inactive	19.55	1794*	80.12
Male physically inactive	18.16	1794*	63.46
Female obese	7.76	3*	100
Ages 55_64	8.20	5*	14.18
Female 18-65 uninsured			0.04**
Male 18 – 65 uninsured			0.00**
Fast food restaurants			0.00**
Native Hawaiian			0.46**
White population	***	***	***
Black population	***	***	***
Male population	***	***	***
Female population	***	***	***
Unemployment	***	***	***

Table 7: Eliminated Variables

Table 7 above contains variables that were eliminated from the exploratory analysis. Reasons for their elimination are indicated below.

* = VIF violation

**= insignificant

***= never selected

4.2 Wilcoxon-Mann-Whitney Test

The Wilcoxon Mann-Whitney test is a nonparametric test for comparing two populations (Crawley, 2012). Wilcoxon Mann-Whitney, tests the null hypothesis that two populations have identical distribution functions against the alternative hypothesis that the two distribution functions differ only with respect to location (ibid). This test was used to determine whether adult obesity rates between rural and urban counties differ. The test was run within R the programming environment. Hypotheses are.

- **Null hypothesis:** There is no significant difference between adult obesity rates in urban and rural counties
- **Alternative hypothesis:** There is a significant difference in adult obesity rates between rural and urban counties

After running the codes a p-value of 0.04621 was computed. This p-value was less than alpha level of 0.05 which led to the rejection of the null hypothesis that there is no difference between obesity rates in the urban and rural counties. Rejection of the null hypothesis leads to the supporting of the alternative hypothesis. The results indicated that there is a difference in adult obesity rates between rural and urban counties. However, the difference may not be considered as significant since 0.04621 is close to the alpha level 0.05.

4.3 Hot Spot and Spatial Autocorrelation (Global Moran's I)

Hot Spot and spatial autocorrelation (Global Moran's I) analysis are spatial statistics tools within Arc GIS used to map and analyze clusters and patterns of features under study. These methods were used to analyze patterns of adult obesity within the state of Alabama. The Hot Spot Analysis Tool (Getis-Ord Gi) was used to compare the distribution of adult obesity rates in the state while spatial autocorrelation (Moran's I) tool was used to analyze the overall patterns and trend of the data to evaluate whether features are clustered, dispersed, or random. The difference between the two methods is that Hot Spot Analysis identifies spatial concentration of values and distinguishes between hot spots (areas of high values) and cold spots (areas of low values), while spatial autocorrelation analysis only indicates clustering (areas where similar features are grouped together) and cannot tell if these are hot spots (high values), cold spots (low values), or both. These two methods are based on hypothesis testing. For each method two hypotheses was formulated as indicated below.

4.3.1 Hot Spot Analysis

For the Hot Spot Analysis the hypotheses were:

Null hypothesis: values are randomly distributed

Alternative hypothesis: Values are clustered

The output for Hot Spot Analysis was a map indicating where high and low values of adult obesity are clustered (Figure 3). The results indicate that four counties were identified as hot spots 99% confidence, four counties hot spots 95% confidence, and one county hot spot 90% confidence (Table 8). In the cold spot, three counties were identified as cold spots 99% confidence; two counties cold spots 95% confidence and two counties cold spots 90% confidence (Tables 9). The Presence of high and low clusters indicated that adult obesity in the state of Alabama is not randomly distributed. Considering the results above, the null hypothesis was rejected- indicating that adult obesity within the state of Alabama is clustered with most of the counties in the hot spot zones being identified as rural counties. These counties also fall within the Alabama Black Belt Region. This is the most economically depressed and socially disenfranchised region of the state of Alabama – generally characterized by low socioeconomic levels, low levels of educational attainment, unemployment, high rates of poverty and poor access to health care service among other things.

4.3.2 Spatial Autocorrelation Analysis

For the Spatial Autocorrelation Analysis the hypotheses were:

Null Hypothesis: there is no spatial autocorrelation in the incidence of obesity in the State of Alabama.

Alternative Hypothesis: spatial autocorrelation exists in the incidence of obesity in the State of Alabama.

Spatial autocorrelation is the measure of how much close objects are in comparison with other close objects (William, 1993), and the results can be classified as positive, negative or no spatial auto-correlation. Positive spatial autocorrelation is when similar values cluster together on a map. Negative spatial autocorrelation is when dissimilar values cluster together on a map. This tool calculates the Moran's I Index value, a z-score and p-value to evaluate the significance of the Index. For this test the results indicated a Moran's I Index of 0.257, a Z-score of 5.083 and a p-value of 0.000 (Figure 4). According to Spatial (Moran's I) analysis, if the Z test statistic is > 1.96 (or < -1.96) the null hypothesis is usually rejected. According to the results, the z-score was greater than 1.96 so the null hypothesis was rejected. This indicated that adult obesity rates are NOT random within the state of Alabama.

Hot Spots 99% confidence	**Hot Spots 95% confidence**	**Hot Spots 90% confidence**
Sumter	Pickens	Butler
Greene	Perry	
Hale	Wilcox	
Dallas	Lowndes	

Table 8: Counties in Hot Spot Zones

Cold Spots 99% confidence	**Cold Spots 95% confidence**	**Cold Spots 90% confidence**
Madison	Mobile	Blount
Marshal	Baldwin	Etowah
Dekalb		

Table 9: Counties in Cold Spot Zones

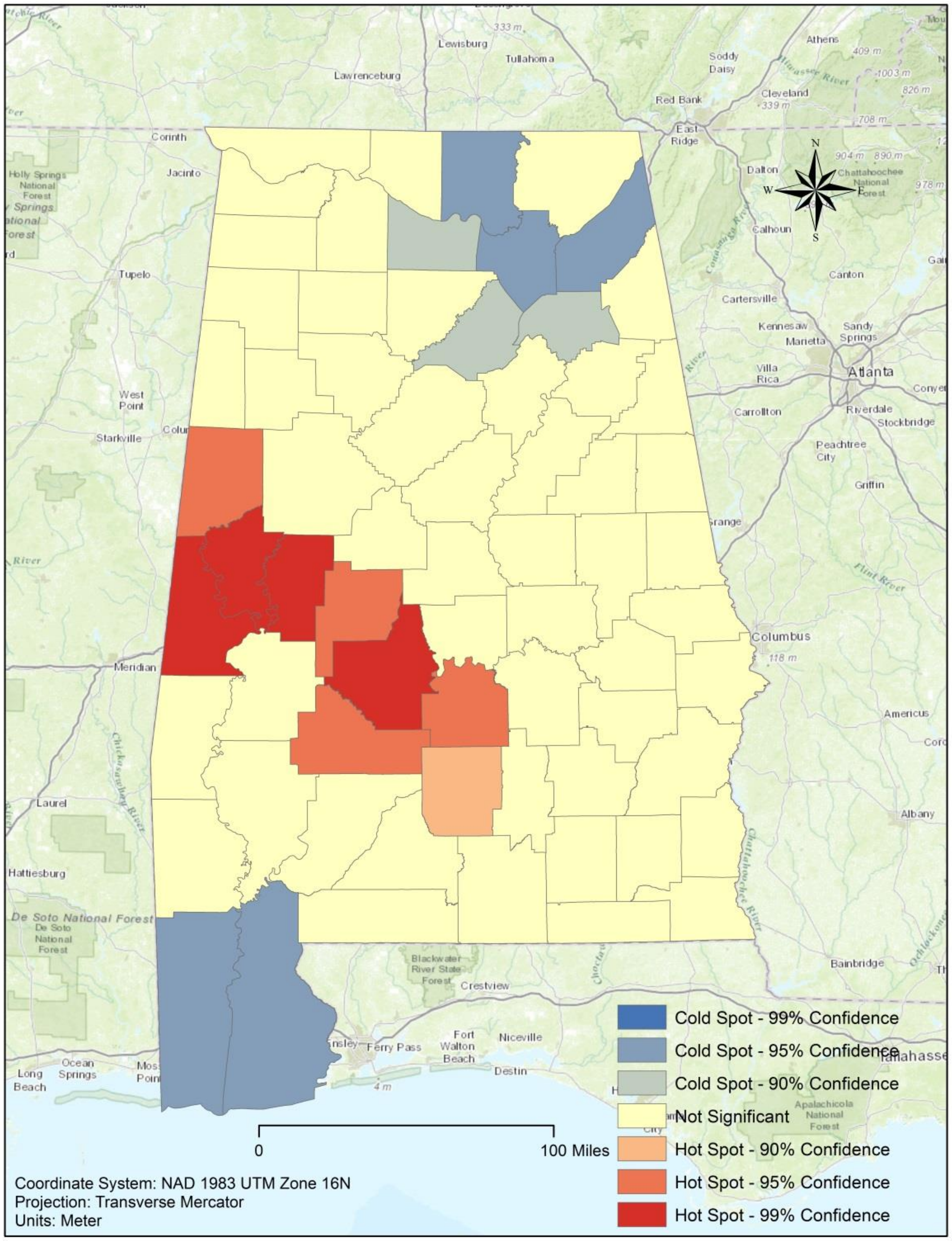


Figure 3: Adult obesity Hot Spot Analysis Map.

Hot Spot analysis map above indicates where high values and low values of adult obesity are within the state of Alabama.

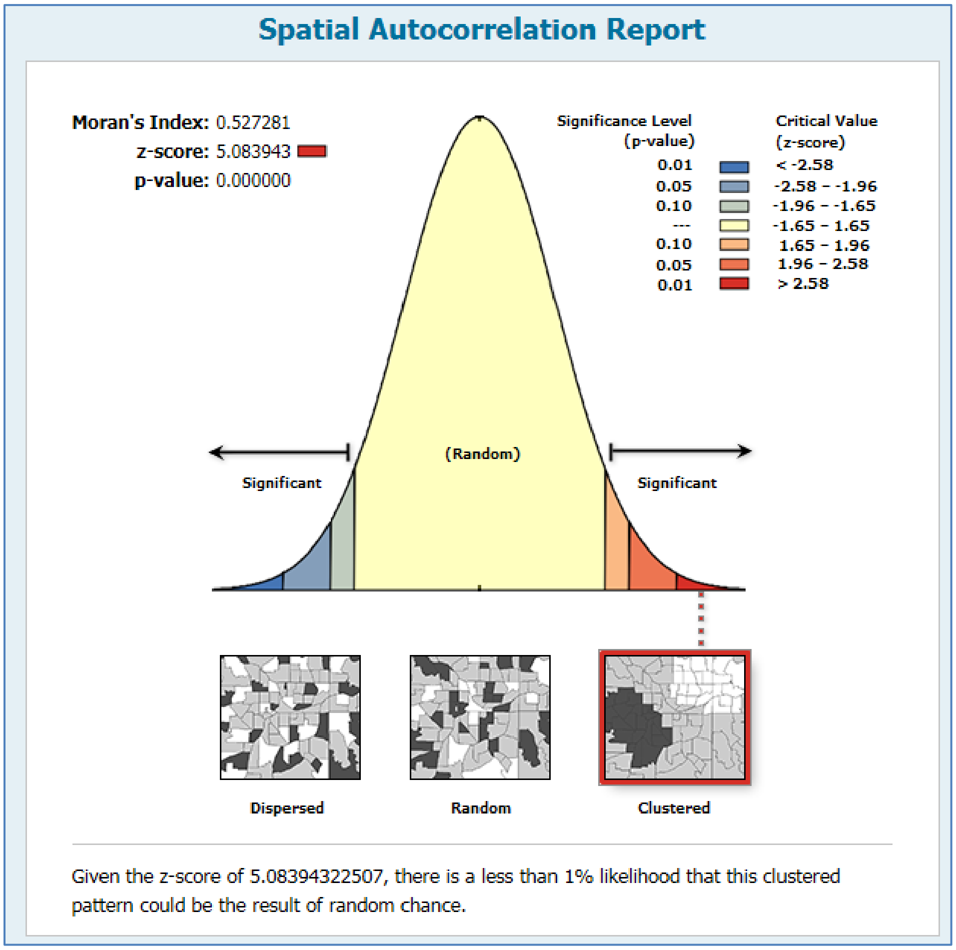


Figure 4: Spatial Autocorrelation Analysis of Adult Obesity.

Spatial autocorrelation result above shows the statistics computed for Spatial Autocorrelation (Moran's I). The statistics indicate that adult obesity rates in the state of Alabama are not out of a random chance.

4.4 Multiple Regression Analysis

Spatial relationships between demographic, socioeconomic, health, and built environment factors and adult obesity were modeled using a stepwise method for variable selection in linear regression. Regression analysis models relationships among data variables, allowing data examination, exploration and better understanding of key factors influencing the variable being modelled. Regression also verifies that relationships exist and measures strengths of those relationships (Crawley, 2005).

Regression analysis was carried out in steps.

1. First, a global (OLS) model was run with all the variables from exploratory regression analysis. To obtain a reliable model, different models were run with combination of different variables and the best model selected using the procedure illustrated by Yamashita et al, (2007).
2. Secondly, five model sets: Health, Age, Population, Built environment and Economic models were also run selectively using the same procedure as the global OLS model. Within each model set, different models were run and the best model was selected following the linear model performance criterion by Montgomery et al., (2012).
3. These six models were finally categorized and ranked to assist in determining factors that influence adult obesity in the state of Alabama.

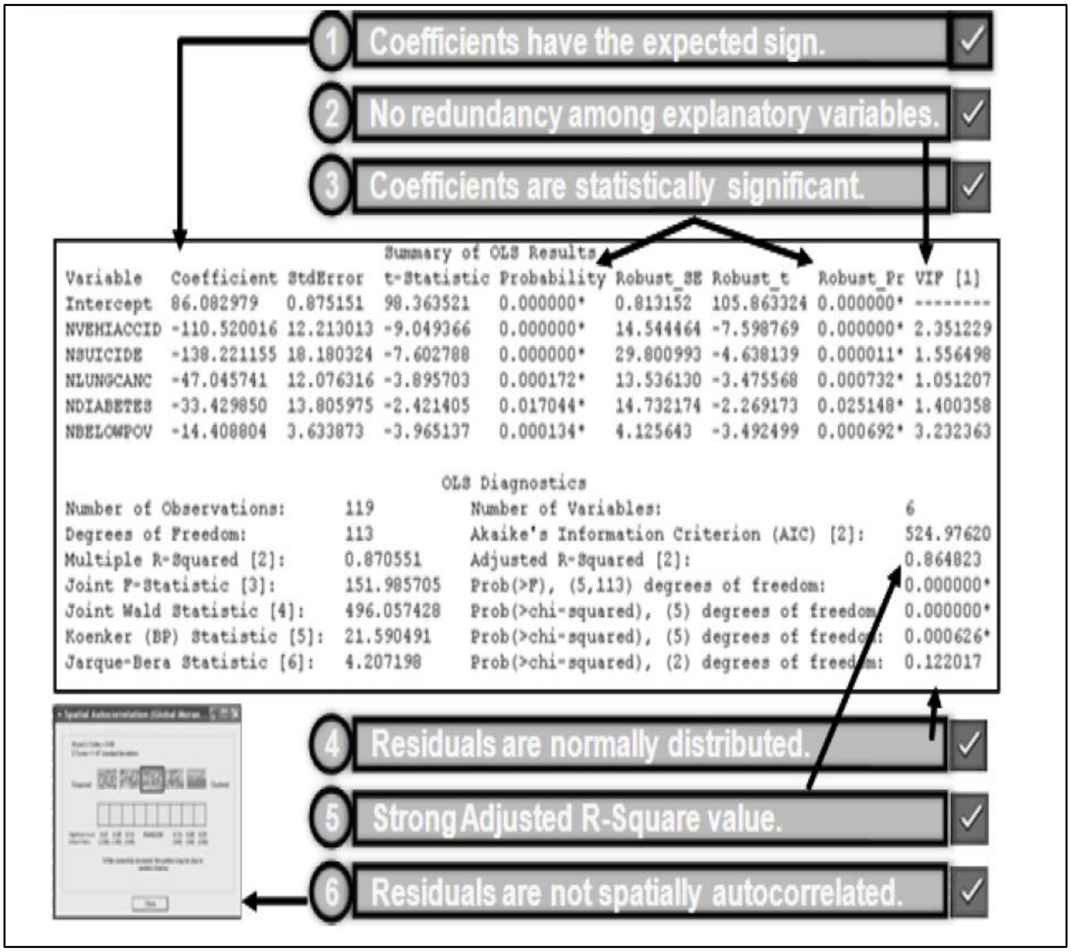


Figure 5: Steps for Examining the Model Summary Reports

The above six steps were used in examining models summary reports to ensure the very best OLS model is identified.

To rank the models, AICc- Akaike's Information Criterion was used. AICc "measure allows one to compare and rank multiple competing models and to estimate which of them best approximates the "true" process underlying the biological phenomenon under study" (Symonds & Moussalli, 2011: 13). Wagenmakers & Farrell (2004) have indicated that the use of AICc to evaluate models is one of the most popular methods of comparing multiple models; as it takes into account both descriptive accuracy and parsimony.

After categorizing and ranking the models, the global OLS model was selected as the best model because it has the lowest AIC value of the 6 candidate models, however considering only the models set; Health model seemed to be the best model with AIC of 61, followed by Economic, Age, Built Environment and lastly Population model as indicated in (Table 10).

To assess if a pattern exists in the spatial distribution of the variable, spatial autocorrelation (Global Moran's I) was run on the general OLS model residuals to analyze the overall patterns and trend of the data. A Moran's I of 0.044, z-score of -0.55 and p-value of 0.577 was computed. Based on spatial autocorrelation Moran's I statistics and hypotheses testing, the z-score (0.55) computed led to the supporting of the Null hypothesis (Figure 5 & 6). There was no statically significant Spatial Autocorrelation.

Since OLS model assumes that variables are stationary over geographic space, GWR was used to test for variability of data over geographic space. The best model (global OLS model) from the six models was selected to be run using GWR. The GWR results were adjusted R2 value of 94.6 % and AIC of 9.97. The values of this model were compared to the values of the global OLS which was an adjusted R2 of 94.6 and AIC of 9.95%. Using AIC measure for model performance, if the AICc values for two models differ by more than 3, the model with the lower AICc is held to be best model. For this analysis the difference in the models was 0.02 indicating that both models were good models.

The stepwise regression method used was adopted from Yamashita,Yamashita, & Kamimura, (2007). Their method has explicitly explained in the most simplest and understandable way how stepwise regression works. This method has been illustrated below with the variables and steps by step application explained.

Variables identification:

(a) xi, xii are variables to select, and xj, xjj are variables to delete.

(b) p_m − 1_, p_m_: respectively, number of input variables at the end of step m − 1, and number of input variables considered in step m and at the end of step m.

(c) x_m − 1_, x_m_: respectively, vector of input variables at the end of step m − 1, and vector of input variables considered in step m and at the end of step m

How to run the model:

Start with no input variables in model (we may re-start with a subset of variables, and in this case, if there are more than or equal to three variables in the model go to Step 3), otherwise go to Step 2.

(1) Select (add) one significant variable xi. Compare the criterion value of all models with one variable xi. Select xi if the model with xi gives the best criterion value. If there is no selection, go to Step 4.

(2) Select one more significant variable xii. “Selection method: compare criterion value of all models that include the first xi (or xi’s) and one additional xii. Select xii if the model with the additional xii gives the best criterion value when the first xi (or xi’s) are already in the model.” If there is no selection go to Step 4. If there are more than or equal to three variables in the model go to Step 3, otherwise go to Step 2.

(3) Delete (remove) one insignificant variable xj. “Deletion method: compare the criterion value of all models that include the xi’s without one xj. Delete xj if the model removing xj gives the best criterion value.” If there is no deletion or no more deletion, go to Step 2, otherwise go again to Step 3.

(4) Stop the stepwise method for variable selection.

4.5 Models

This section contains detailed information on the models that were run. For each model variables used have been identified and the results captured in histograms and scatter plots for visual analysis. 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. All the models were run using the same criteria as illustrated in the OLS model below.

4.5.1 Global OLS Model

For this model, all the 16 variables selected for analysis from the explanatory regression were inputted into the model. Using a stepwise model selection criterion different models were run and 7 variables were eliminated for not passing linear model assumptions, and 9 variables were selected as the key variables that best explain adult obesity rates in the state of Alabama (Figure 10). The OLS results show how significant the model parameters are. With 9 variables the model explained 95 % of adult obesity prevalence in the state of Alabama indicating that some other unknown factors were responsible for the 5% of obesity unexplained. Among all the seven models this was the best model with an AIC of 9 (Figure 9 &10).

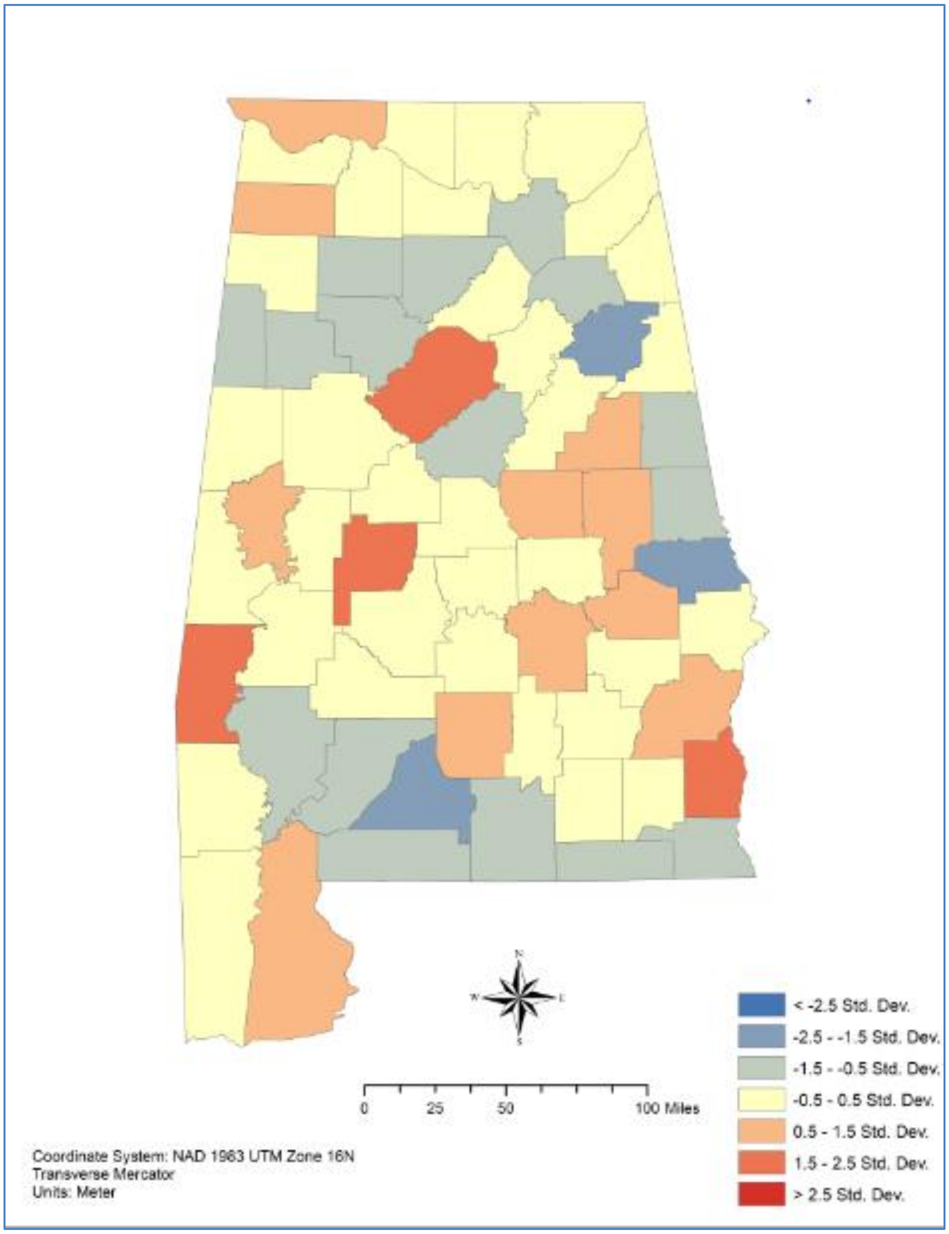


Figure 6: Global OLS model Map.

Figure 6 above show the OLS model residuals. It indicates how randomly the over and under predictions are distributed.

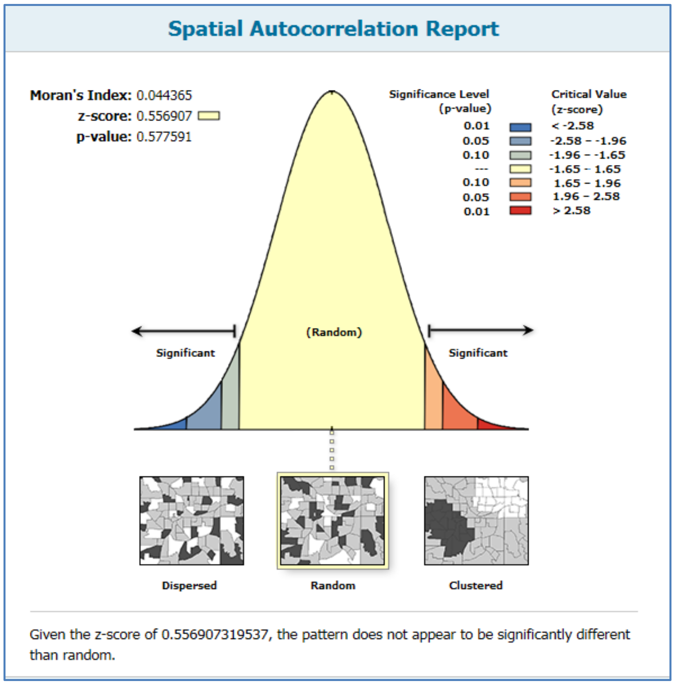


Figure 7: Global OLS Spatial Autocorrelation

Figure 7 above indicates that the model residuals are randomly distributed. This supports the Null hypothesis of complete spatial randomness.

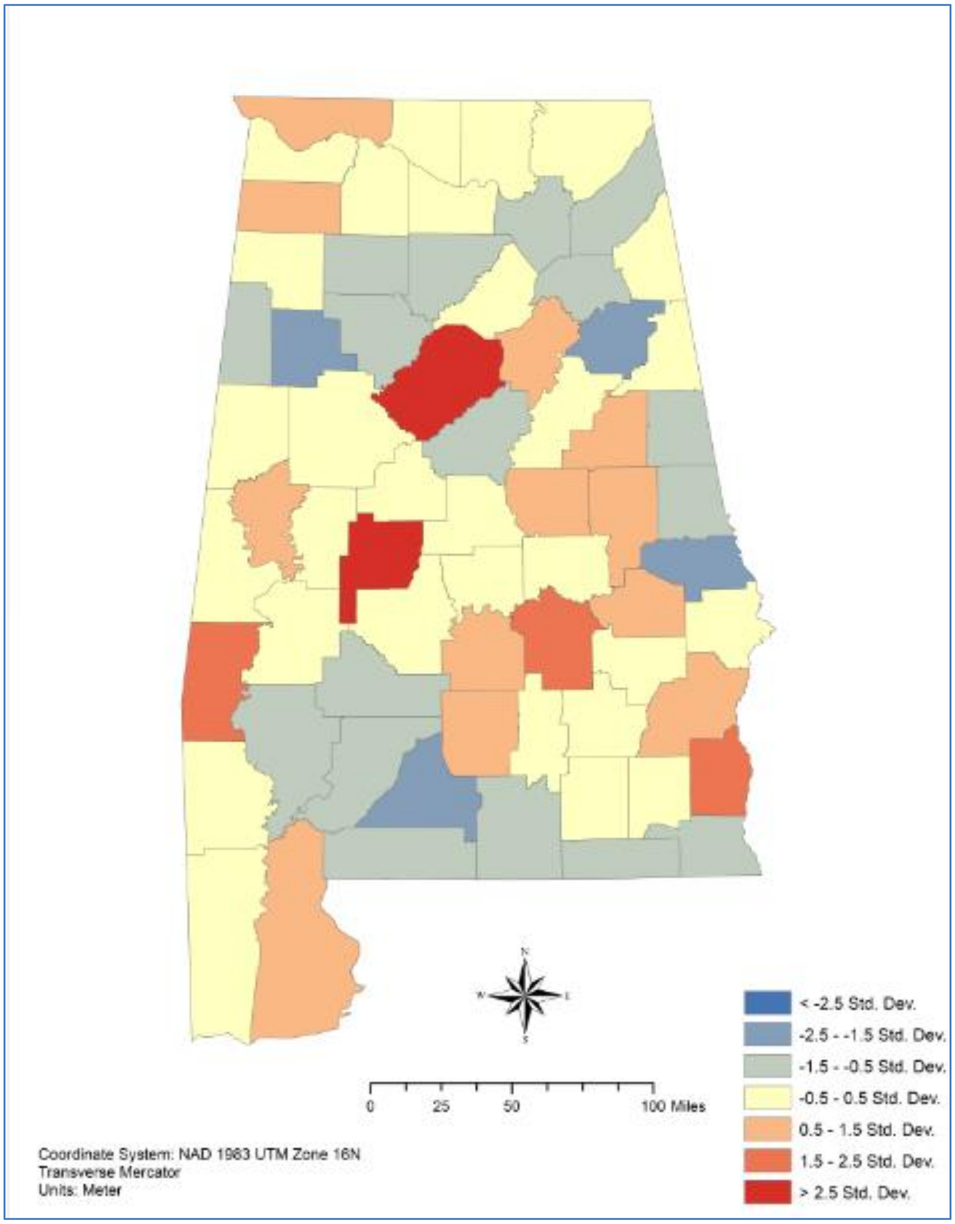


Figure 8: GWR Map.

Figure 8 shows how randomly the over and under predictions are randomly distributed.

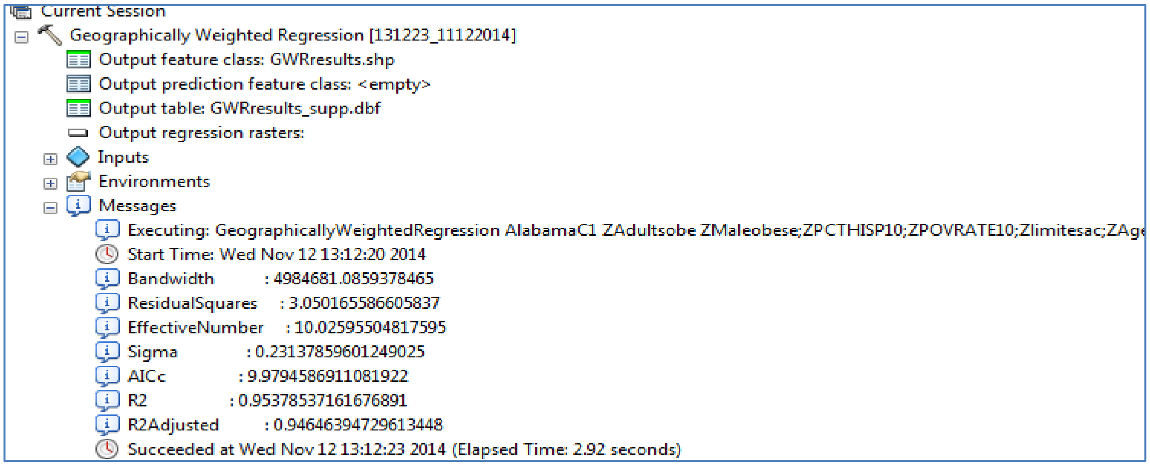


Figure 9: GWR Results

Figure 9 above shows the statistical summary report indicating the GWR model performance

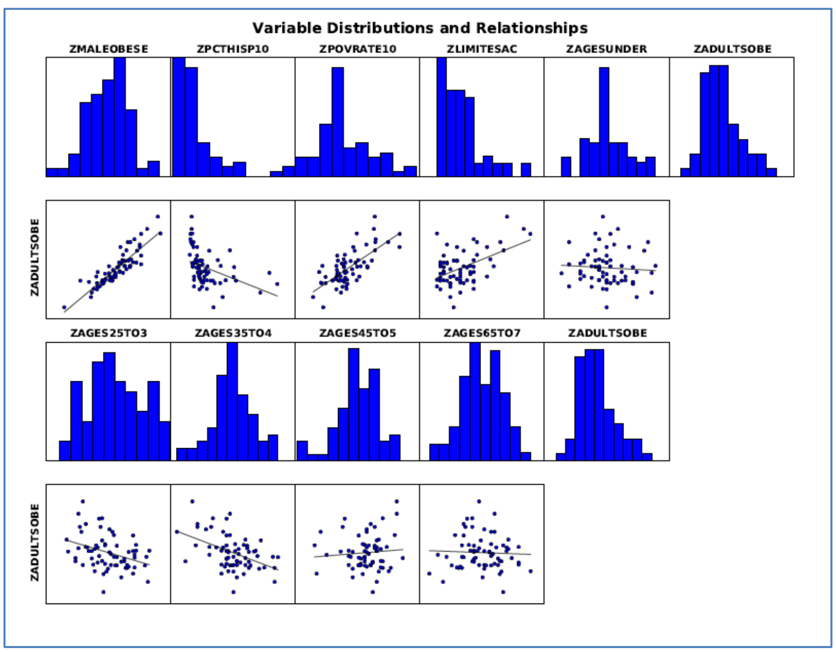


Figure 10: Histogram and Scatterplots for Global OLS Model.

OLS Diagnostics			
Input Features:	AlabamaC1	Dependent Variable:	ZADULTSOBE
Number of Observations:	67	Akaike's Information Criterion (AICc) [d]:	9.955677
Multiple R-Squared [d]:	0.953776	Adjusted R-Squared [d]:	0.946477
Joint F-Statistic [e]:	130.679380	Prob(>F), (9,57) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	2193.980599	Prob(>chi-squared), (9) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	15.214683	Prob(>chi-squared), (9) degrees of freedom:	0.085205
Jarque-Bera Statistic [g]:	1.356709	Prob(>chi-squared), (2) degrees of freedom:	0.507451

Figure 11: Global OLS Model Diagnostics.

4.5.2 Health *OLS* model

This model was created using only health variables which included female obesity rates, male obesity rates, female 18-65 uninsured, male 18-65 uninsured, female physically inactive, and male physically inactive. After running several models only three of the variables; male obesity rates, female-physically inactive, and male physically inactive met the criteria for the best model as indicated below (Figure 11 & 12). With only health factors the model explained 87 % of adult obesity rates in the state of Alabama.

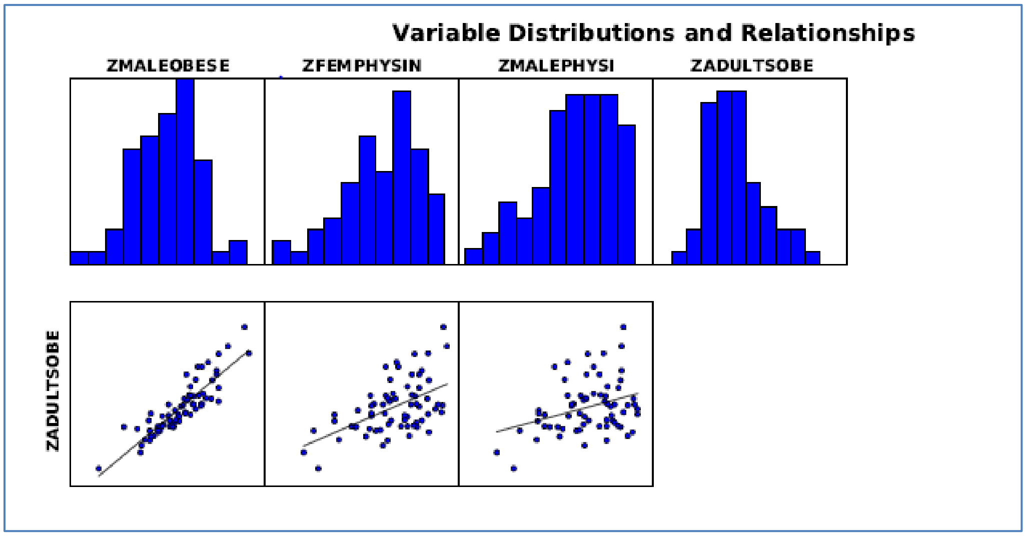


Figure 12: Histogram and Scatterplots for Health Model.

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	0.000000	0.044593	0.000000	1.000000	0.043242	0.000000	1.000000	--------
ZMALEOBESE	0.763915	0.056093	13.618826	0.000000*	0.056339	13.559174	0.000000*	1.558622
ZFEMPHYSIN	0.963917	0.142110	6.782909	0.000000*	0.136564	7.058359	0.000000*	10.004069
ZMALEPHYSI	-0.927836	0.132214	-7.017660	0.000000*	0.122486	-7.575013	0.000000*	8.659391

Figure 13: Health Model Diagnostics.

4.5.3 Age Model

Age model consisted of nine age groups. After running several models the best model had four age groups (ages 5- 14, ages 25 -34, ages 35 – 44, and ages 65 -74) that were considered significant as shown in the scatter plots and OLS results below. With the above age categories the model explained 39% of adult obesity in the state of Alabama (Figure 13 & 14).

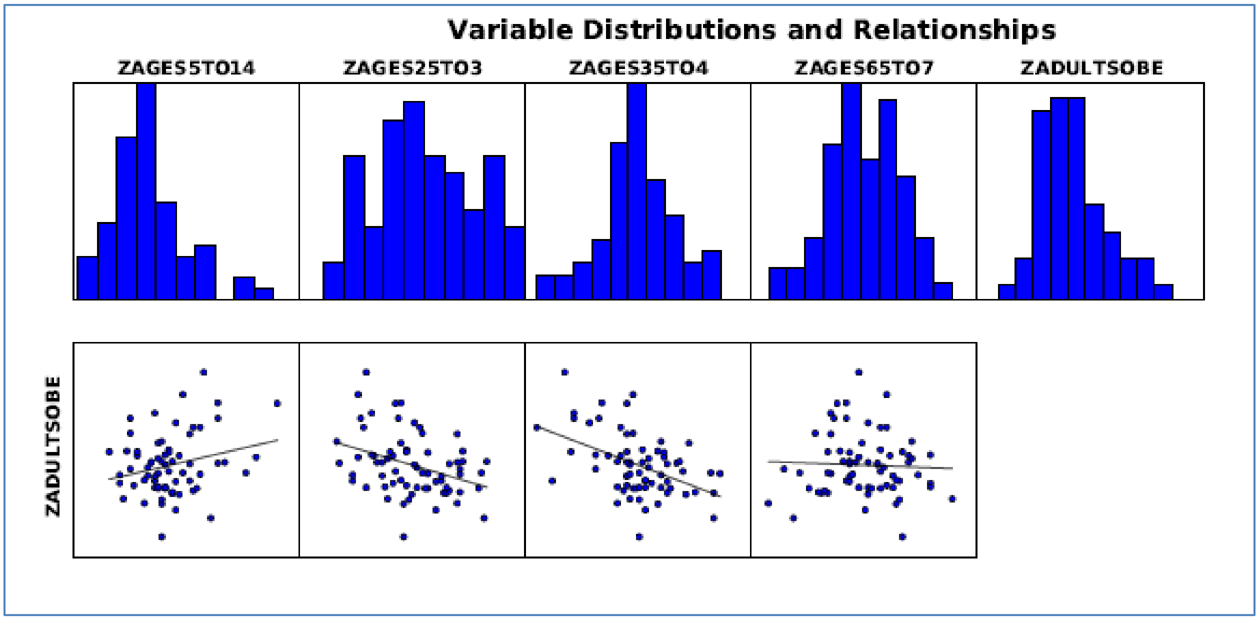
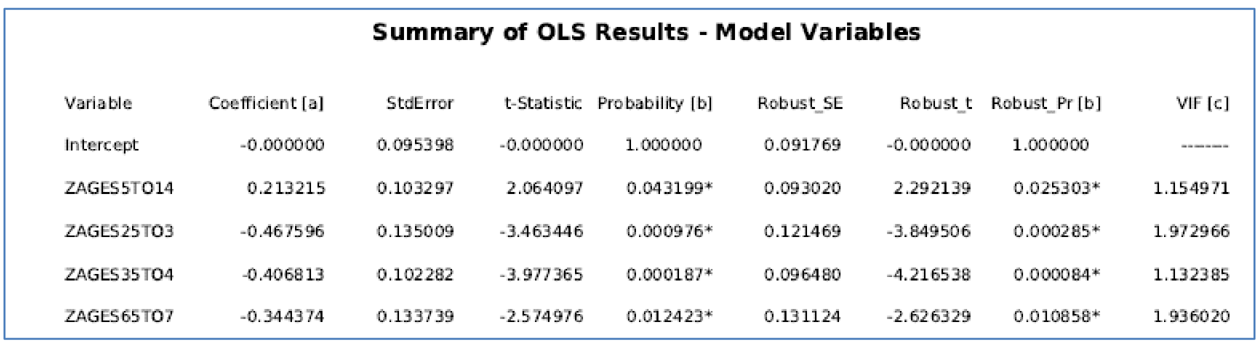


Figure 14: Histogram and Scatterplots for Age Model.



Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-0.000000	0.095398	-0.000000	1.000000	0.091769	-0.000000	1.000000	--------
ZAGES5TO14	0.213215	0.103297	2.064097	0.043199*	0.093020	2.292139	0.025303*	1.154971
ZAGES25TO3	-0.467596	0.135009	-3.463446	0.000976*	0.121469	-3.849506	0.000285*	1.972966
ZAGES35TO4	-0.406813	0.102282	-3.977365	0.000187*	0.096480	-4.216538	0.000084*	1.132385
ZAGES65TO7	-0.344374	0.133739	-2.574976	0.012423*	0.131124	-2.626329	0.010858*	1.936020

Figure 15: Age Model Diagnostics.

4.5.4 Population Model

Variable for this model included sex and race data. For race only White, Black, Hispanic, Asian, Native Hawaiian and other Pacific Islanders population were included in the study. Of which only Hispanic and Asian population were significant as indicated below in the scatterplot and

OLS results. This model explained model 23% of adult obesity in the state of Alabama (Figure 15 & 16).

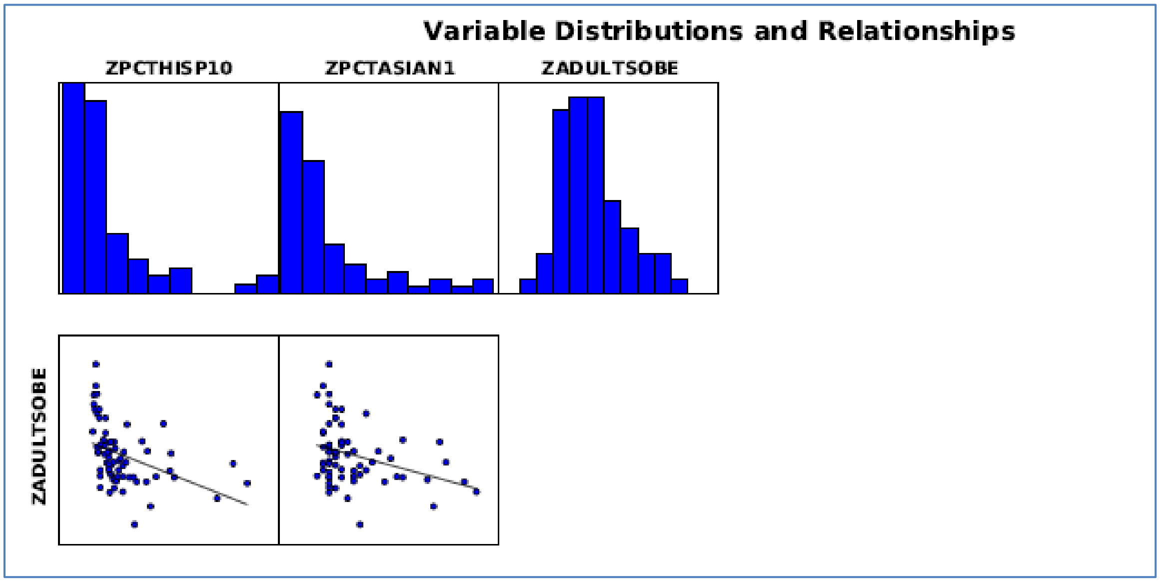


Figure 16: Histogram and Scatterplots for Population Model.

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-0.000000	0.107155	-0.000000	1.000000	0.104729	-0.000000	1.000000	--------
ZPCTHISP10	-0.381045	0.108756	-3.503661	0.000846*	0.090044	-4.231765	0.000077*	1.014733
ZPCTASIAN1	-0.287129	0.108756	-2.640112	0.010398*	0.079920	-3.592680	0.000639*	1.014733

Figure 17: Population Model Diagnostics.

4.5.5. Built Environment model

Variables used in the model were access to recreational facility, limited access to healthy foods and 2009 fast food rates. Fast food rate was eliminated from the model leaving the best model with two variables as shown in the scatter plot and summary of model results below. The model explained 32% of adult obesity in the state of Alabama (Figure 17 &18) .

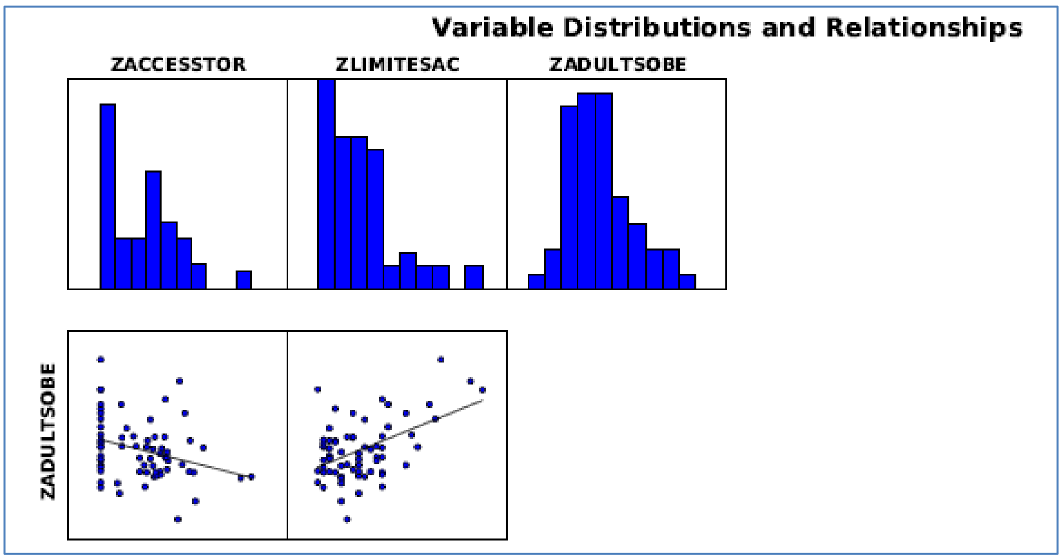


Figure 18: Histogram and Scatterplots for Environment Model.

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-0.000000	0.107155	-0.000000	1.000000	0.104729	-0.000000	1.000000	--------
ZPCTHISP10	-0.381045	0.108756	-3.503661	0.000846*	0.090044	-4.231765	0.000077*	1.014733
ZPCTASIAN1	-0.287129	0.108756	-2.640112	0.010398*	0.079920	-3.592680	0.000639*	1.014733

Figure 19: Environment Model Diagnostics.

4.5.6 Economic Model

Variables in the model were 2010 poverty rate and education rates. Education rates were eliminated from the variable for being insignificant. With only one variable the model explained 50% of the obesity rates (Figure 19 & 20).

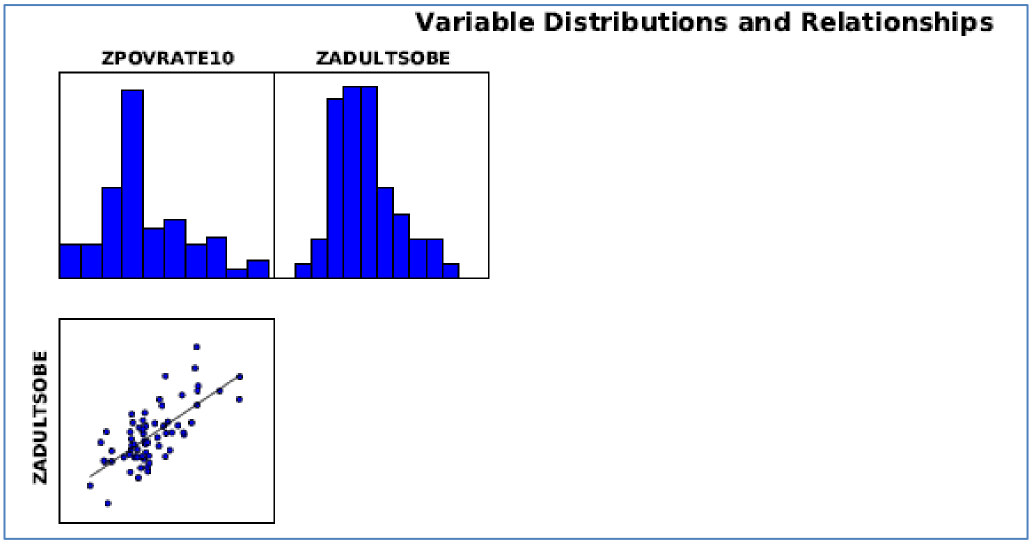


Figure 20: Histogram and Scatterplots for Economic Model.

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	0.000000	0.086456	0.000000	1.000000	0.085156	0.000000	1.000000
ZPOVRATE10	0.711892	0.087108	8.172496	0.000000*	0.087700	8.117327	0.000000*

Figure 21: Economic Model Diagnostics.

Models	Variables	#variables	AIC	∆AIC	AdjR2	Ranks
Global OLS	ZAdultobesity = ZMaleobese + ZHispanic + ZPovertyrate + ZLimitedaccesstohealthy food + ZAgesunder5 + ZAges25-34 + ZAges35-44 + Ages45-55 + ZAges65-74	9	9	0.00	0.946	1
Health	ZAdultobesity = ZMaleobses + ZFemalephyicallyinactive + ZMalephysicallyinactive	3	61		0.866	2
AGE	ZAdultobesity = ZAges5-14 + ZAges25-34 + ZAges35-44 + ZAges65-74	4	165		0.390	4
Population	ZAdultobesity = ZHispanics + ZAsians	2	178		0.230	6
Built Environment	ZAdultobsity= ZAccesstorecreationalfacilities + ZLimittedaccesstohealthy	2	169		0.323	5
Economic	ZAdultobesity = ZPovertyrate	1	148		0.499	3

Table 10: Summary of Models Results.

Table 10 above shows the six models created and the statistical measures used to evaluate the models.

Chapter 5: Summary of Results

The results of this study indicate that there is a significant relationship between geographic location and the incidence of obesity at county level. As the results show, rural counties exhibit higher rates of adult obesity than urban counties. Most of the counties in the Hot Spot zones were identified as rural counties while the cold spot areas were predominantly urban. The results also indicate that statistical relationship exists between socio-economic, demographic, health and environmental variables and obesity at county level. Stronger relationships were observed between health, economic factors, age groups, built environment and population factors. These results show that the variables selected were able to explain more of the variations in the model.

Chapter 6: Findings and Conclusion

The main focus of this study was to demonstrate the role of place in understanding obesity patterns in Alabama. The main assumption was that unlike popular beliefs, incidence of obesity varies by place and across different scales. The overarching goal was to contribute to the existing body of knowledge on obesity and to help provide a place-based approach applicable in formulating policies and usable in tailoring public health interventions directed towards management and prevention of the obesity epidemic.

This study focus was to answer the following questions.

1. Is there a significant relationship between geographic location and the incidence of obesity at county level In Alabama?

2. Is there a statistical relationship between socio-economic, demographic, health and environmental variables and obesity at the county level in Alabama?

According to the analysis, obesity rates were identified to be higher in rural counties than in urban counties with p-value of 0.04621 at α 0.05 significance level, but the difference is small. The difference between rural and urban adult obesity rates was determined using hot spot analysis that indicated that 66.7% of the counties in the hot spot zone were rural counties. Seven counties were in the cold spot zone (5 urban; 2 rural) and fifty one counties were insignificant. Most of these rural counties lie within the Black Belt Region of Alabama.

Multiple regression analysis confirmed that there is relationship between socio-economic, demographic, health, and built environment variables. However, the relationship varies with specific factors as indicated in Figure 3 and by location. Marked variation in obesity incidence across counties can be linked to among others things socio-economic, demographic, health and built environment factors.

Among all the factors analyzed, adult obesity in the state of Alabama was more related to (1) Gender - male being obese than women (2) Race – higher among Hispanic population (3) Poverty rate (4) Limited access to healthy food (5) Ages under 5years (6) Ages 25 to 34 (7) Ages 35 to 44 (8) Ages 45 to 54 and (9) Ages 65 to 74. When variables were grouped into factors and categorized into different models, health factors were more related to adult obesity - with higher rates of obesity in males than females even when study shows both as physically inactive. The second best model was economic factors model although the model had only one variable, poverty rates. Age model was the third best model with ages 5 to 14 more related to adult obesity and ages 65 to 74 least related to obesity. In the built environment model obesity was more related to access to recreational activities than limited access to healthy foods. The least significant model was population / race model which indicated that obesity was more prevalent among Hispanic population.

Contemporary obesity studies are multidisciplinary in nature with the main focus on causes, effects and preventive measures (Linne et al., 2004; Drewnowski et al., 2005; Colman, 2007; Healy et al., 2008; Whiteman et al., 2008; Weiss et al., 2008; Salome et al., 2010; Cassel et al., 2010; Gillman et al., 2013) with limited focus on place based factors in understanding the obesity phenomena.

From the analysis 100% of all counties in the hot spot zone fall within Alabama's Black Belt Region in the south central part of the State. None of the cold spots fall within this region. The

Black belt region, which is predominantly African-American, is characterized by high levels of poverty, illiteracy, unemployment and histories of social and spatial exclusion.

5.1 Study Limitations

The research results established a relationship between socio-economic and demographic factors and obesity across Alabama Counties and assessing the role of place in understanding obesity. Like many other previous studies, the study experienced some limitations which included lack of sub-county level data. Data used in the study was at county level which was readily available. Data at the zip code, census block or census tract level would have been more informative for the analysis. The study also employed aggregated datasets from sources such as the Behavioral Risk Factor Surveillance System (BRFSS). Using aggregated data obscures internal differences. Lastly, obesity rate data were obtained from the BRFS system. These rates are from self-reported information on individual height and weight which can sometimes be misreported or inaccurate. While it was possible to answer the research questions, it is recommended that a place-based analysis of such a topic as obesity should consider sub-county level data for analysis.

Bibliography

Aimone, A. M., Perumal, N., & Cole, D. C. (2013). A systematic review of the application and utility of geographical information systems for exploring disease-disease relationships in paediatric global health research: the case of anaemia and malaria. Int J Health Geogr, 12(13), 10-1186.

Akil, L., & Ahmad, H. A. (2011). Effects of socioeconomic factors on obesity rates in four southern states and Colorado. Ethnicity & disease, 21(1), 58.

Australia. Australian Institute of Health and Welfare, Canberra

Bhatt, B. M., & Joshi, J. P. (2012). GIS IN EPIDEMIOLOGY: APPLICATIONS AND SERVICES. National Journal of Community Medicine, 3(2).

Bogart, W. A. (2013). Law as a Tool in "The War on Obesity": Useful Interventions, Maybe, But, First, What's the Problem?. The Journal of Law, Medicine & Ethics, 41(1), 28-41.

Braun, M. T., & Oswald, F. L. (2011). Exploratory regression analysis: A tool for selecting models and determining predictor importance. Behavior research methods, 43(2), 331-339.

Brown, W.J., Young, A.F., Byles, J.E., 1999. Tyranny of distance? The health of mid- age women living in five geographical areas of Australia. Australian Journal of Rural Health 7, 148–154.

Carr, D., & Friedman, M. A. (2005). Is obesity stigmatizing? Body weight, perceived discrimination, and psychological well-being in the United States. Journal of Health and Social Behavior, 46(3), 244-259.

Census Bureau (2014). Alabama Demographics [Community Facts]. Retrieved from http://factfinder2.census.gov/faces/nav/jsf/pages/community_facts.xhtml

Center for Disease Control and Prevention. (2014) Overweight and Obesity. Retrieved from http://www.cdc.gov/obesity/index.html

Charlton, M., Fotheringham, S., & Brunsdon, C. (2009). Geographically weighted regression. White paper. National Centre for Geocomputation. National University of Ireland Maynooth.

Chen, D. R., & Truong, K. (2012). Using multilevel modeling and geographically weighted regression to identify spatial variations in the relationship between place-level disadvantages and obesity in Taiwan. Applied Geography, 32(2), 737-745.

Chou, S. Y., Grossman, M., & Saffer, H. (2004). An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. Journal of health economics, 23(3), 565-587.

Clarke, K. C., McLafferty, S. L., & Tempalski, B. J. (1996). On epidemiology and geographic information systems: a review and discussion of future directions. Emerging infectious diseases, 2(2), 85.

Clarke, K. C., McLafferty, S. L., & Tempalski, B. J. (1996). On epidemiology and geographic information systems: a review and discussion of future directions. *Emerging infectious diseases*, *2*(2), 85.

Comber, A. J., Brunsdon, C., & Radburn, R. (2011). A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International journal of health geographics*, *10*(1), 44-55.

Craney, T. A., & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. Quality Engineering, 14(3), 391-403.

Crawley, M. J. (2005). Statistics: an introduction using R.

Crawley, M. J. (2012). The R book. John Wiley & Sons

D'Agostino-McGowan, L., Gennarelli, R. L., Lyons, S. A., & Goodman, M. S. (2013). Using Small-Area Analysis to Estimate County-Level Racial Disparities in Obesity Demonstrating the Necessity of Targeted Interventions. International journal of environmental research and public health, 11(1), 418-428.

D'Agostino-McGowan, L., Gennarelli, R. L., Lyons, S. A., & Goodman, M. S. (2013). Using small-area analysis to estimate county-level racial disparities in obesity demonstrating the necessity of targeted interventions. *International journal of environmental research and public health*, *11*(1), 418-428.

Drewnowski, A., & Darmon, N. (2005). Food choices and diet costs: an economic analysis. The Journal of nutrition, 135(4), 900-904.

Duncan, M. J., Mummery, W. K., Steele, R. M., Caperchione, C., & Schofield, G. (2009). Geographic location, physical activity and perceptions of the environment in Queensland adults. Health & place, 15(1), 204-209.

Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: current approaches and future challenges. Environmental health perspectives, 998-1006.

Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: current approaches and future challenges. Environmental health perspectives, 998-1006.

Finkelstein, E. A., Khavjou, O. A., Thompson, H., Trogdon, J. G., Pan, L., Sherry, B., & Dietz, W. (2012). Obesity and severe obesity forecasts through 2030. American journal of preventive medicine, 42(6), 563-570.

Gatrell, A. C., & Bailey, T. C. (1996). Interactive spatial data analysis in medical geography. Social Science & Medicine, 42(6), 843-855.

Gibbs, R. M. (2003). Reconsidering the southern black belt. The Review of Regional Studies, 33(3), 254-263.

Gordon-Larsen, P., Nelson, M. C., Page, P., & Popkin, B. M. (2006). Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*, *117*(2), 417-424.

Gordon-Larsen, P., Nelson, M. C., Page, P., & Popkin, B. M. (2006). Inequality in the built environment underlies key health disparities in physical activity and obesity. Pediatrics, 117(2), 417-424.

Griffith, D. A. (1992). What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. Espace géographique, 21(3), 265-280.

Haslam, D. (2007). Obesity: a medical history. Obesity reviews, 8(s1), 31-36.

Healy, G. N., Wijndaele, K., Dunstan, D. W., Shaw, J. E., Salmon, J., Zimmet, P. Z., & Owen, N. (2008). Objectively measured sedentary time, physical activity, and metabolic risk the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). Diabetes care, 31(2), 369-371.

Hojjat, T. A. THE ECONOMIC ANALYSIS OF OBESITY. Review of Business & Finance Studies, 81.

Jackson, J. E., Doescher, M. P., Jerant, A. F., & Hart, L. G. (2005). A national study of obesity prevalence and trends by type of rural county. *The Journal of Rural Health*, *21*(2), 140-148.

Jacquez, G. M. (2000). Spatial analysis in epidemiology: Nascent science or a failure of GIS?. Journal of Geographical Systems, 2(1), 91-97.

Jeffries, H. K. (2009). Bloody Lowndes: Civil Rights and Black Power in Alabama's Black Belt. NYU Press.

Jokela, M., Kivimäki, M., Elovainio, M., Viikari, J., Raitakari, O. T., & Keltikangas-Järvinen, L. (2009). Urban/rural differences in body weight: evidence for social selection and causation hypotheses in Finland. Social Science & Medicine, 68(5), 867-875.

Kirby, J. B., Liang, L., Chen, H. J., & Wang, Y. (2012). Race, place, and obesity: The complex relationships among community racial/ethnic composition, individual race/ethnicity, and obesity in the United States. American journal of public health, 102(8), 1572-1578.

Koch, T., & Denike, K. (2001). GIS approaches to the problem of disease clusters: a brief commentary. Social science & medicine, 52(11), 1751-1754.

Le, A., Judd, S. E., Allison, D. B., Oza-Frank, R., Affuso, O., Safford, M. M., ... & Howard, G. (2014). The geographic distribution of obesity in the US and the potential regional differences in misreporting of obesity. Obesity, 22(1), 300-306.

Liese, A. D., Weis, K. E., Pluto, D., Smith, E., & Lawson, A. (2007). Food store types, availability, and cost of foods in a rural environment. Journal of the American Dietetic Association, 107(11), 1916-1923.

Linne, Y. (2004). Effects of obesity on women's reproduction and complications during pregnancy. Obesity reviews, 5(3), 137-143.

Miller, H. J., & Wentz, E. A. (2003). Representation and spatial analysis in geographic information systems. *Annals of the Association of American Geographers*, *93*(3), 574-594.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). John Wiley & Sons.

Mota, J., Ribeiro, J. C., & Santos, M. P. (2009). Obese girls differences in neighbourhood perceptions, screen time and socioeconomic status according to level of physical activity. Health education research, 24(1), 98-104.

Must, A., & Strauss, R. S. (1999). Risks and consequences of childhood and adolescent obesity. International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity, 23, S2-11.

Nakaya, T. (2001). Local spatial interaction modelling based on the geographically weighted regression approach. GeoJournal, 53(4), 347-358.

Nayga Jr, R. M. (2001). Effect of schooling on obesity: is health knowledge a moderating factor?. Education Economics, 9(2), 129-137.

Pal, S., & Bhattacharya, M. (2013). An empirical study on the financial health of the main steel producing segment in India: application of factor analysis and multiple regression analysis. DECISION, 40(1-2), 47-55.

Papas, M. A., Alberg, A. J., Ewing, R., Helzlsouer, K. J., Gary, T. L., & Klassen, A. C. (2007). The built environment and obesity. Epidemiologic reviews, 29(1), 129-143.

Paul D. Allison. (1999). Multiple regression: A primer. Pine Forge Press.

Peytremann-Bridevaux, I., Faeh, D., & Santos-Eggimann, B. (2007). Prevalence of overweight and obesity in rural and urban settings of 10 European countries. Preventive medicine, 44(5), 442-446.

Pickett, K. E., Kelly, S., Brunner, E., Lobstein, T., & Wilkinson, R. G. (2005). Wider income gaps, wider waistbands? An ecological study of obesity and income inequality. Journal of Epidemiology and Community Health, 59(8), 670-674.

Rosin, O. (2008). The economic causes of obesity: a survey. Journal of Economic Surveys, 22(4), 617-647.

Sallis, J. F., & Glanz, K. (2009). Physical activity and food environments: solutions to the obesity epidemic. *Milbank Quarterly*, *87*(1), 123-154.

Salome, C. M., King, G. G., & Berend, N. (2010). Physiology of obesity and effects on lung function. Journal of Applied Physiology, 108(1), 206-211.

Strong, K., Trickett, P., Titulaer, I., Bhatia, K., 2001. Health in rural and Remote

Sui, D. Z. (2007). Geographic information systems and medical geography: Toward a new synergy. *Geography compass*, *1*(3), 556-582.

Symonds, M. R., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behavioral Ecology and Sociobiology, 65(1), 13-21.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. Psychonomic bulletin & review, 11(1), 192-196.

Wang, Y., Beydoun, M. A., Liang, L., Caballero, B., & Kumanyika, S. K. (2008). Will all Americans become overweight or obese? Estimating the progression and cost of the US obesity epidemic. Obesity, 16(10), 2323-2330.

Whiteman, D. C., Sadeghi, S., Pandeya, N., Smithers, B. M., Gotley, D. C., Bain, C. J., ... & Green, A. C. (2008). Combined effects of obesity, acid reflux and smoking on the risk of adenocarcinomas of the oesophagus. Gut, 57(2), 173-180.

William D. Berry. (1993). Understanding regression assumptions (Vol. 92). Sage.

Yamashita, T., Yamashita, K., & Kamimura, R. (2007). A stepwise AIC method for variable selection in linear regression. Communications in Statistics—Theory and Methods, 36(13), 2395-2403.

Yamashita, T., Yamashita, K., & Kamimura, R. (2007). A stepwise AIC method for variable selection in linear regression. Communications in Statistics—Theory and Methods, 36(13), 2395-2403.