2015

A Spatial Analysis of Obesity and Its Associations with the Built and Natural Environment, Physical Inactivity, and Socioeconomic and Demographic Conditions in the United States of America

Mustafa Erdem
Louisiana State University and Agricultural and Mechanical College

A SPATIAL ANALYSIS OF OBESITY AND ITS ASSOCIATIONS WITH THE BUILT AND NATURAL ENVIRONMENT, PHYSICAL INACTIVITY, AND SOCIOECONOMIC AND DEMOGRAPHIC CONDITIONS IN THE UNITED STATES OF AMERICA

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Geography and Anthropology

by
Mustafa Erdem
B.A., Yildiz Technical University, 2001
M.L.A., Louisiana State University, 2006
May 2016

To express my sincere appreciation for so much faithful support, I would like to dedicate this dissertation to my lovely wife, Dr. S. Sibel Erdem, my precious son, Sansalp Doruk Erdem, my dear father, Yusuf Ziya Erdem, and my dear mother, Emine Erdem.

ACKNOWLEDGEMENTS

During the course of my research at Louisiana State University, I am indebted to a number of people. Initially, I want to express my sincere gratitude to my advisor Dr. Michael Leitner for his support and useful discussions that we had. Without his willingness this would not be possible. I would like to thank my committee members, Dr. Fahui Wang, Dr. Robert V. Rohli, and Dr. Michael Barton for agreeing to serve on my advisory committee and taking time from their busy schedules to read this dissertation. I would like to thank Dr. Xingyou Zhang for generously sharing parks and open spaces data with me. I am also thankful to Dr. Kent Mathewson for his encouragement and belief in me. I would like to thank the Graduate Secretary, Ms. Dana C. Sanders, for her assistance whenever I needed it most.

Special thanks go to my spouse, Dr. S. Sibel Erdem, for her endless support, love, encouragement, and belief in me. Without her and her endless support on my side, none of this would have been possible. I also would like to thank to my adorable son Sansalp Doruk Erdem for his unconditional love and being a great motivation for me. He is the most precious person in our life. In closing, I want to thank my parents, parents-in-law, brothers, sister, brother-in-law, and extended family for their endless support from thousands of miles away.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Obesity has strong genetic determinants but the genetic composition of the population does not change rapidly. Thus in this study, the major changes in non-genetic factors such as the development of obesogenic environments and shifting socioeconomic status and lifestyle of individuals are hypothesized to increase the risk of obesity. As the prevalence of obesity continue to increase worldwide with substantial attention in the US, a clearer understanding of how spatial associations between obesity and confounding factors are interrelated is crucial to better tackle the issue of obesity.

This study employs the 'global' and 'local' Exploratory Spatial Data Analysis (ESDA) methods including the Ordinary Least Squares (OLS) regression and Geographically Weighted Regression (GWR) to investigate obesity and its spatial associations with environmental, behavioral, socioeconomic, sociodemographic, and population based dynamics at the county level. The results from this study have generated empirically-based and useful insights for the 3,105 counties and county-equivalents across the 48 contiguous states, also known as the continental US.

A major contribution of this study is exploring obesity and its confounding associations with various factors not only spatially but also temporally for the first time, revealing the temporal changes from 2004 to 2007 and to 2010. By utilizing the ESDA methods, a consistent answer obtained significantly indicates that positive spatial associations exist between obesity and physical inactivity (PIA), poverty, and population-weighted distance (PWD) to parks. Conversely, negative spatial associations exist between obesity and ratio of jobs to employed residents (JER) and population density.

Another major contribution of this study is examining and revealing geographic variability in the association between obesity rates and various explanatory variables both nationwide and regionally at the county level for the entire US. By utilizing the GWR, a significant spatial nonstationarity is identified. This finding suggests that the strength of associations between obesity and each of the explanatory variables vary depending on the spatial location. It is also revealed that the confounding variables PIA, high educational attainment, African-American population, and poverty are identified as the top four variables by having relatively stronger effects in explaining obesity rates at the county level both nationwide and regionally.

CHAPTER 1. INTRODUCTION

Obesity kills. According to the most widely as well as the most recently accepted estimates, the number of deaths that may be attributable to obesity each year makes obesity the second leading cause of preventable death in the US after smoking and it is likely to become the first unless this chronic issue is addressed successfully (Danaei et al., 2009; Flegal et al., 2005; Mokdad et al., 2004). Obesity has already reached epidemic proportions across the nation and, most recently in 2013, it has been officially classified as a disease by the American Medical Association (AMA), the nation's largest physician organization (AMA, 2013). In addition to the population-wide obesity trends, it is known that the burden of obesity affects some groups, such as low-income and minority populations, more than others (Hill et al., 2014; Flegal et al., 2012). Similarly, disparities in obesity are also evident by geographic location, such as existence of higher rates of obesity in the southern states than elsewhere in the US (Hill et al., 2014). Compared to smoking, obesity is more prevalent, yet similar to smoking, obesity is also a risk factor for many chronic health conditions and is also highly associated with overall poor physical health (Sturm, 2001). Therefore, obesity is seen as an important determinant of overall heath.

In many different cultures around the world, the importance of being healthy is emphasized by the saying "the greatest wealth is health" yet people only give health value when they lose it. Powerful and rather complex relationships exist between health and the factors affecting health. The range of biological, demographic, behavioral, social, economic, and physical environmental factors that influence an individual's or population's health are known as determinants of health. The examples of demographic, behavioral, social, economic, and physical environmental determinants of health, respectively include: race and ethnicity, age and gender; diet, physical activity, and alcohol and cigarette use; availability of resources to meet daily needs such as educational/job opportunities, living wages, or healthy foods and exposure to crime and violence; poverty, income, and transportation options; geographic location, built environmental features, such as buildings, recreational settings, and public transportation, and natural environmental features, such as open spaces and weather or climate change (Healthy People 2020, 2014). In the literature, the influence of such factors is also documented as notable determinants of obesity.

When then US Secretary of Health and Human Services Margaret Heckler's (1985) landmark report on minority health was released (Heckler, 1985), obesity was among the modifiable risk factors associated with leading causes of death. With the release of this report, an overall message, pertaining to all risk factors including obesity, was framed in a way that "bad behavior" was the key issue and therefore changes in individuals' behaviors were the ultimate solution. In other words, if individuals change their behaviors, they would be able to modify risk factors and that would help disparities, which were experienced mostly by African-Americans and other minority Americans, to disappear. However, when especially the last two decades are considered from the obesity point of view, the pervasiveness of disparities affecting diverse populations proved otherwise and showed that it was not simply a behavior issue but there were also structural factors involved (Kumanyika, 2005). Individuals' behaviors and biology may govern individual risk, yet even the amalgamation of the two does not fully explain the dramatic increase in obesity prevalence over the past decades in the US (Huang and Glass, 2008).

Health disparities among individuals of various demographics and socioeconomic status and across different geographic locations have been a topic attracting recent interest not only in academia but also in the public realm due to the increasing recognition of strong links between social, economic, and environmental factors and health and the widening disparities in such factors (WHO, 2015). The National Institutes of Health (NIH) developed the first official definition of health disparities in September 1999 in response to a White House initiative: "Health disparities are differences in the incidence, prevalence, mortality, and burden of diseases and other adverse health conditions that exist among specific population groups in the United States." In 2000, United States Public Law 106-525 (106th Congress, 2000) provided a legal definition of health disparities: "A population is a health disparity population if there is a significant disparity in the overall rate of disease incidence, prevalence, morbidity, mortality, or survival rates in the population as compared to the health status of the general population." In addition, Healthy People 2020 defined health disparity as "a particular type of health difference that is closely linked with social, economic, and/or environmental disadvantage (Healthy People 2020, 2014). Health disparities adversely affect groups of people who have systematically experienced greater obstacles to health based on their racial or ethnic group; religion; socioeconomic status; gender; age; mental health; cognitive, sensory, or physical disability; sexual orientation or gender identity; geographic location; or other characteristics historically linked to discrimination or exclusion."

These three definitions have several things in common and they define health disparities based on: 1) The differences in the overall rate of disease incidence, prevalence, morbidity, mortality or survival rates among different population groups; 2) The factors that contribute to health disparities. Even though the term "disparities" is often interpreted to imply racial or ethnic disparities, many dimensions of health disparities exist in the US. In other words, racial or ethnic background, sex, age, disability, socioeconomic status, physical environment, or geographic location all contribute to an individual's ability to achieve good health and the effects of such differences are applicable to obesity as well.

Disparities in health outcomes among communities across the nation are a central concern in public health and epidemiology. Due to the spatial aspects of health disparities issue, public health professionals and medical scholars have been increasingly seeking assistance from geographers. Geographers utilize concepts and techniques to investigate the spatial patterns of health outcomes and the independent and alternate processes that lead to such spatial patterns. Detecting where events occur often yields valuable clues about why these events occur. In other words, the identification of spatial patterns (or clusters) leads to creating hypotheses to better understand the underlying processes in effect. Hence, the application of geographic methods and techniques in health-related research has been gaining popularity to tackle intricate health issues, and obesity is definitely as one such health issue. Methods of exploratory spatial data analysis and geovisualization are widely used to provide insight into the patterns of health outcomes and to establish possible linkages to the built environmental, socioeconomic, and sociodemographic factors, which individually or in combination may play a role in the resulting health outcomes (Cromley and Cromley, 2009).

Geographic Information Systems (GIS) are computer-based sets of procedures that capture, store, manipulate, analyze, model, and visualize data with spatial characteristics, therefore providing an excellent platform to study health-related spatial phenomena. Although GIS was initially developed in the 1960s, the more modern version of GIS used today emerged in the 1980s (Foresman et al., 1998). Over the last two decades, with the widespread use of GIS and increased demand for more sophisticated applications of GIS, computer-based geographic technologies and methods including remote sensing (RS), global positioning systems (GPS), mathematical models, and quantitative methods showed a tremendous growth (Brown et al., 2005; Legates et al., 2005). Especially in the 1990s some of the most remarkable developments and improvements have taken place on hardware and software technologies (Legates et al., 2005; Goodchild, 1992). As a result, GIS and spatial analysis and modeling have greatly benefited from these innovations (e.g., Fischer and Getis, 1997; Longley et al., 1997; Cho, 1996; Fotheringham and Rogerson, 1994; and Fischer and Nijkamp, 1993). These progresses not only lowered the cost of data gathering, storing, and processing but also increased the speed and efficiency in both data processing and analyzing. Furthermore, the establishment of academic GIS centers, organizations, conferences, and new journals has further accelerated the pace of spatial analysis research (Legates et al., 2005, 443).

Even though GIS and spatial analysis initially started as two separate fields of research and application, over time they have become closer fields that support each other and add value to the other (Goodchild and Haining, 2004). The advances in GIS and spatial analysis and modeling have not only transformed the field of geography but also brought geography into many other fields of research where geography did not exist before. This has increased the appreciation and popularity of geography and its tools (e.g., GIS and spatial analysis methods and techniques) in solving spatial problems. Health-related research has been one of the many academic fields of research that benefited greatly from such advances and collaborations (Gatrell, 2002).

One of the biggest advantages of using GIS is its capability to analyze both temporally and spatially complex data sets. The research proposed in this dissertation involves the use of GIS and Exploratory Spatial Data Analysis (ESDA) techniques to investigate the spatiotemporal distribution of obesity at the county level in the US and its associations with built environmental, socioeconomic, and sociodemographic factors. A spatial database, a GIS mapping software, and a standalone spatial analysis platform designed to implement methods and techniques for ESDA were used to capture, store, manipulate, edit, analyze, and visualize the outcomes and relations produced from the analyses conducted in this study.

1.1 Problem Statement

The Centers for Disease Control and Prevention (CDC) has reported substantial increases in obesity rates across all ages, genders, and racial and ethnic groups in every state in the US especially since 1991 (CDC, 2014; Flegal et al., 2010; Ogden and Carroll, 2010a, 2010b; Mokdad et al., 2001, 2000, 1999). On the other hand, based on previous research, it is also evident that obesity has affected some groups and sub-populations more than others (Hill et al., 2014; Flegal et al., 2012). Hill and Trowbridge (1998) emphasized that "despite obesity having strong genetic determinants, the genetic composition of the population does not change rapidly. Therefore, the large increase in obesity must reflect major changes in non-genetic factors."

The literature on obesity has been growing especially since 2002. Although there are many publications examining the relationship between obesity and diverse key factors using various spatial and statistical methods, only a few studies have been detected at the county level for the entire US in the literature (e.g., Xu and Wang, 2015a; Xu and Wang, 2015b; and Chi et al., 2013). There are obesity studies covering the entire US at the state level or the entire US at the county level but only for those counties included in the Metropolitan Statistical Area (MSA) regions (therefore not contiguous) or a whole state at the county/zip code/Traffic Analysis Zone (TAZ) levels or parts of a county at the census tract/neighborhood levels (Durand et al., 2011; Feng et al., 2010; Galvez et al., 2010; Li et al., 2009; Ewing et al., 2008; Mujahid et al., 2008; Drewnowski et al., 2007; Holt and Sui, 2007; Rundle et al., 2007; Schlundt et al., 2006). Even though the contributions of such publications are essential to see the obesity phenomenon and its interactions at different scales, studying and documenting it at the county level is as valuable if not more. In addition, despite the numerous studies, the great heterogeneity across studies in terms of the key variables, statistical methods, and spatial extent used and the number and range of places compared has produced different results that are somewhat incomparable (Feng et al., 2010).

Therefore, this dissertation aims to fill that gap and contribute to the literature on spatial analysis of obesity by documenting spatio-temporal interactions, patterns, and trends based on the obesity and the selected key variables using ESDA methods including the Geographically Weighted Regression (GWR) technique at the county level covering the entire US. The main research questions include: How has obesity increased? Where and what type of populations are most affected? How does obesity relate to the built environment? Considering the above overview, a spatial analysis of obesity based on the behavioral, built environmental, socioeconomic, and sociodemographic conditions at the county level is expected to provide a better understanding of the obesity epidemic and its previously overlooked or unknown confounding associations with these dynamics.

1.2 Research Objectives

The goal of this research is to investigate the spatial associations between obesity and built environmental, socioeconomic, and sociodemographic dynamics in the US and to determine if such dynamics can explain disparities in obesity prevalence. Addressing disparities has much to contribute to health research in general, and to obesity research, in particular. Especially in recent years, inequities in the social and built environments have increasingly been the focus of research that aim to explain weight-related disparities, such as obesity (Kershaw et al., 2013; Powell et al., 2012; Wen and Kowaleski-Jones, 2012; Lovasi et al., 2009). Therefore, the specific objectives of this dissertation research are:

1) To explore and identify spatio-temporal patterns and trends of obesity and its associations with respect to behavioral, environmental, socioeconomic, and sociodemographic conditions in the US using county level data.
2) To examine whether these associations are based on clusters of occurrences or are randomly dispersed across space by applying a set of spatial and non-spatial analytical methods. If statistically significant clusters exist, then additional research questions are: a) Where are these clusters located? b) Do they spatially converge?
3) To identify predictors of obesity prevalence at the county level and to reveal how the effects of these predictors vary spatially.

1.3 Research Hypotheses

The hypotheses of this research are:

1) There is significant spatial clustering of high (or low) values of obesity or the selected environmental, socioeconomic, and behavioral variables.
2) There are significant spatial clusters of obesity in areas with lower population density, poorer access to parks, and less mixed land use.
3) There are significant spatial clusters of obesity in areas where poverty levels are high.
4) There are significant spatial clusters of obesity in areas where the physical inactivity levels of the underlying population are high.
5) Predictors of obesity prevalence and effects and variations of such factors across space at the county level can be identified and revealed by the GWR technique.

1.4 Expected Significance

The expected significance of this research is threefold:

1) This dissertation will outline geographic areas where disparities in obesity occur to identify specific populations at risk at a county level for the entire US. More particularly, this study identifies geographic clusters (hot spots) of obesity and therefore certain groups and/or subpopulations that require special attention. Spatial clusters can reveal information about the underlying geographical process that generates the spatial pattern which can further assist in the comprehension of the underlying geographical process and its relationship with the phenomenon under investigation. In order to develop successful public health interventions to reduce obesity disparities, one must thoroughly define spatial patterns and understand the forces that lead to such clustering.
2) Part of the problem is that the causal factors behind obesity disparities or health disparities, more broadly, are complex and not fully understood. Nonetheless, what may overall affect obesity risks and obesity outcomes lies beyond the medical model which focuses primarily on medical treatment options. Therefore, in the literature regarding health disparities, an emerging focus has been the spatial context (Richardson et al., 2013). In other words, what role do individuals' living environments play in explaining poorer health outcomes among certain groups and/or subpopulations? This dissertation examines whether disparities in obesity can be linked to the contextual factors that refer to the characteristics of an area such as the features of the physical environment or the level of poverty in those particular geographic areas as well as to the population composition variables that refer to the characteristics of the underlying population such as income, educational level, race, or age. Hence, the variables scrutinized in this study

were chosen carefully to depict the different aspects of such influences. More specifically, this study will attempt to specify area-based built environmental, socioeconomic, and sociodemographic variables to identify predictors of obesity and how the effects of these predictors vary spatially across the US counties. Understanding the possible linkages between obesity and its predictors can help reduce obesity prevalence throughout the nation. Otherwise, the cumulative impact of these forces will leave a permanent mark in the obesity disparities that exist today.

3) This study is among the first ones to use ESDA methods, including the GWR technique, to study obesity at the county level in the United States. The results from this study will generate useful insights that will help generate new hypotheses and define study areas for further detailed scale studies.

1.5 Chapter Organization

The current chapter provided an overview of this dissertation research including its objectives, hypotheses, and expected significance. Chapter 2 documents the general trends in obesity and physical activity in the US to date, associations between various factors and obesity, built environmental characteristics including design standards and their effects on obesity and physical activity, and last but not least various public policies on food and urban planning and their effects on obesity. Chapter 3 describes the study area, the data sources, and data used in this research along with the spatial and non-spatial analytical methods applied to this research. Chapter 4 analyzes the spatial distribution of obesity and the associated built environmental, socioeconomic, and behavioral factors to identify spatio-temporal patterns and trends at the county level in the US by employing ESDA techniques available in GeoDa 1.6.7 software (univariate and bivariate Moran's I and LISA statistics). Chapter 5 examines the obesity's associations with the built environmental, socioeconomic, and sociodemographic dynamics to identify predictors of obesity at the county level in the US by employing the GWR technique utilizing the GWR 4.0 software (https://geodacenter.asu.edu/gwr).

CHAPTER 2. LITERATURE REVIEW ON OBESITY, DIET, PHYSICAL ACTIVITY, AND BUILT AND SOCIAL ENVIRONMENT

2.1 Obesity

Obesity is defined by the Body Mass Index (BMI) which is a measure of fat content in the human body based on both height and weight. According to the World Health Organization (WHO), National Institute of Health (NIH), and the Centers for Disease Control and Prevention (CDC) guidelines, individuals having a BMI of 30kg/m^2 or greater are considered obese, and those having a BMI between 25 kg/m^2 and 29.9 kg/m^2 are considered overweight. The CDC calculates the BMI using the formula of (weight [kg] / height [m]^2) and the self-reported height and weight information. Table 2.1 presents more detailed definitions of the BMI according to WHO guidelines.

Table 2.1 Definitions of the Body Mass Index (BMI) according to the World Health Organization (WHO) guidelines

Categories	Body Mass Index (BMI) in kg/m^2
Underweight	<18.5
Normal weight	18.5-24.9
Overweight	≥ 25.0
Pre-obese	25-29.9
Obesity	≥ 30.0
Obese class I	30-34.9
Obese class II	35-39.9
Obese class III	≥ 40.0

There is an alternative and more accurate method to calculate the BMI which uses waist circumference to measure abdominal body fat (CDC, 2015; Hu, 2007; Janssen et al., 2002). This method also takes gender and age of individuals into account. However, use of this method is not widespread since it is difficult to collect waist circumference data (CDC, 2015; Hu, 2007; Janssen et al., 2002). Therefore, the CDC uses the above mentioned BMI formula. With these definitions given, the majority of the population, both men and women, in the US may be categorized as either overweight or obese (Ogden et al., 2014).

The burden of obesity is substantial both in the US and globally, and it is predicted that a rise in obesity prevalence is likely to continue to 2030 (Wang et al., 2011; Kelly et al., 2008). Rather high rates of obesity are observed in the US when compared to the rest of the world (https://www.cia.gov/library/publications/the-world-factbook/rankorder/2228rank.html). Although the obesity issue has captured the public's attention especially in recent years due to the increasing prevalence of obesity reported, the obesity problem in the US is older than one might think. National nutrition-monitoring data from 1976–1980 showed a gradual increase in the prevalence of adult obesity (Kuczmarski et al., 1994). Over the last decades the Behavioral Risk Factor Surveillance System's (BRFSS) survey data have revealed more recent obesity trends in the US. The CDC has been conducting the BRFSS survey annually since 1984 (http://www.cdc.gov/brfss/annual_data/all_years/states_data.htm). There are a total of 31 surveys to date. The most recent survey was conducted in 2014. The CDC has reported substantial increases in obesity rates in all ages, genders, races, and ethnic groups in every state since 1991 (CDC, 2014; Flegal et al., 2010; Ogden and Carroll, 2010a, 2010b; Mokdad et al., 2001, 2000, 1999).

Figure 2.1 presents an overview of obesity prevalence among adults in 1990, 2000, and 2010 in the US. In 1990, among the 45 states participating in the BRFSS's survey, ten states had a prevalence of obesity less than 10% and no state had a prevalence equal to or greater than 15%.

Based on the surveys conducted, there has been an obvious increase in obesity prevalence in between the 1991-1998 periods (Mokdad et al., 2001, 2000, 1999). Among respondents aged 18 years and older, the obesity prevalence increased by around 50% (from 12% to 17.9%) between 1991 and 1998 (Mokdad et al., 2001, 2000, 1999). In addition, by 1999 13% of children aged 6 to 11 and 14% of adolescents aged 12 to 19 were reported as overweight (Office of the Surgeon General, 2014). The increase in obesity prevalence continued through 1999-2009 periods as well (CDC, 2010).

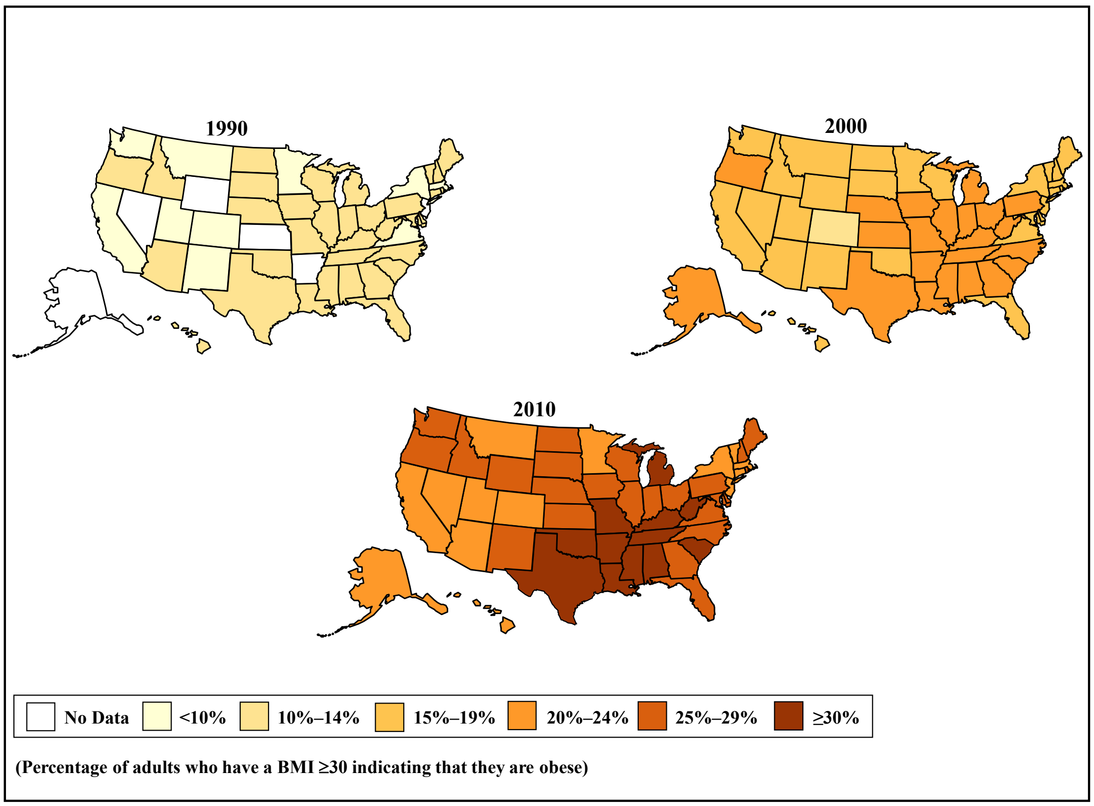

Figure 2.1 Obesity prevalence rates at the state level in the US for 1990, 2000, and 2010

By 2000, no state had a prevalence of obesity less than 10%, twenty-three states had a prevalence between 20%-24%, and no state had a prevalence equal to or greater than 25% (Mokdad et al., 2001). In 2010, no state had a prevalence of obesity less than 20%. Thirty-six states had a prevalence equal to or greater than 25%. Twelve of these thirty-six states (Alabama, Arkansas, Kentucky, Louisiana, Michigan, Mississippi, Missouri, Oklahoma, South Carolina, Tennessee, Texas, and West Virginia) had a prevalence equal to or greater than 30% (CDC, 2010).

Figure 2.2 presents an overview of obesity prevalence among adults in 2011, 2012, 2013, and 2014 in the US. Based on the surveys conducted in the 2011-2012 periods, 34.9% of US adults aged 18 years and older and 17% of US children and adolescents aged 2-19 years were obese (CDC, 2014; Ogden et al., 2014). By 2013, for the first time ever in the US history, two states (Mississippi (35.1%) and West Virginia (35.1%)) had a prevalence of obesity above 35%.

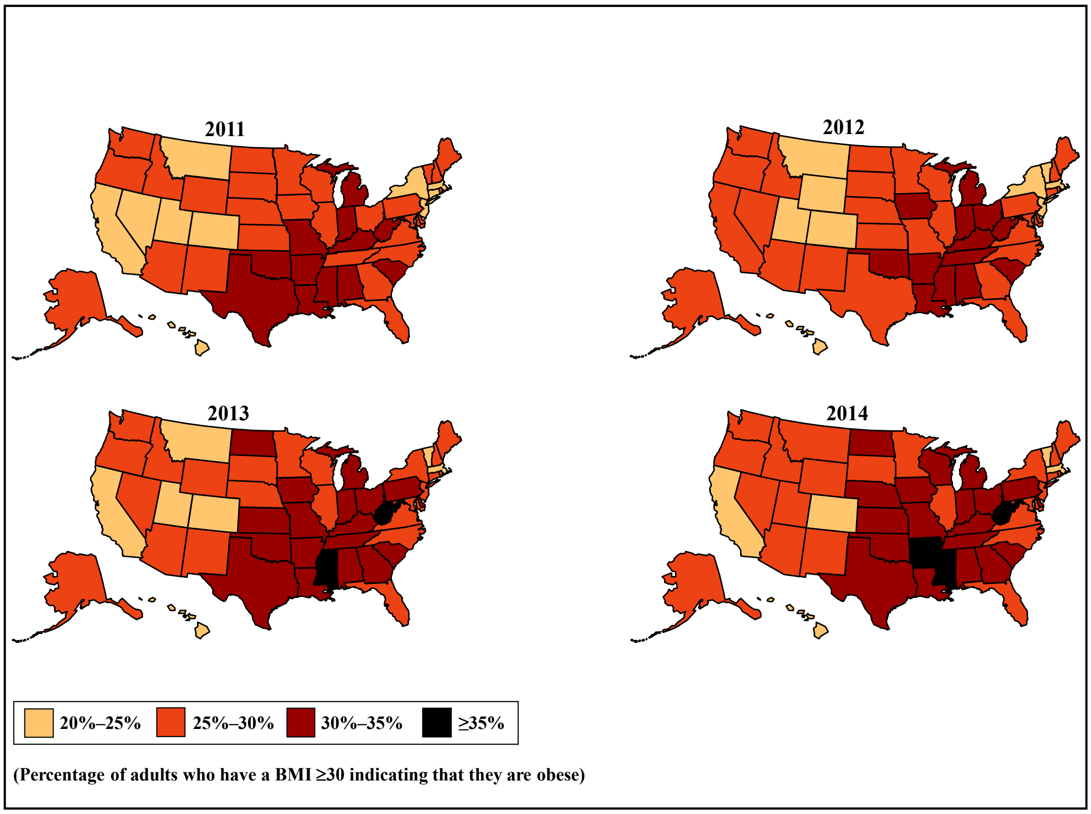

Figure 2.2 Obesity prevalence rates at the state level in the US for 2011, 2012, 2013, and 2014

In addition, based on the most recently conducted BRFSS survey in 2014, in no state was the prevalence of obesity less than 20%. The prevalence of obesity ranged from 21.3% in Colorado to 35.9% in Arkansas. The national median obesity prevalence was 29.5%. These trends show that overall, the prevalence of obesity among the individuals aged 18 years or older is more than 1 in 4 in the US. What is worse, more than 1 in 3 individuals aged 18 years or older are obese in Alabama, Louisiana, Mississippi, West Virginia, and Arkansas (CDC, 2015).

Although one of the goals of the US government's Healthy People 2000 initiative was to reduce the prevalence of obesity, it was evident by January 1994 that this goal was not being reached even when population surveys showed that more than half of the US population actively attempted weight loss or weight control during that period (Serdula et al., 1999; Russel et al., 1995). Following this, yet again obesity was listed as among the leading health indicators in Healthy People 2010 and 2020 initiatives (Healthy People 2020, 2014). Unfortunately, in spite of the national efforts, the progress toward the goal of reducing the proportion of adults who are obese has been minimal during the Healthy People 2010 initiative yet the Healthy People 2020 initiative still aims to reduce the proportion of obese adults by 10% (Heron et al., 2009).

Despite the debates over the actual risks attributable to obesity and the factors leading to high obesity prevalence, the epidemic itself remains a major public health concern. Being overweight or obese are considered risk factors for a number of health problems including chronic cardiovascular diseases, stroke, hypertension, dyslipidemia, type 2 diabetes, various cancer types, liver and gallbladder disease, osteoarthritis, lower-back and chronic neck pain, sleep apnea, and other respiratory problems, infertility, as well as poor general health and some cancers (CDC, 2014; Ahima and Lazar, 2013). Also obesity is known to be associated with psychological disorders and mental health since there are social standards of appearance and attractiveness linked with body size and shape across cultures (Sutin et al., 2011). Obesity's distinct impact is mostly on morbidity and disability rather than on mortality. However, studies still confirm that obesity is a risk factor for all-cause mortality (Flegal et al., 2013; Whitlock et al., 2009).

As mentioned in the introduction section of this study, some 200,000 deaths may be attributable to obesity each year making obesity the second-leading cause of preventable death after smoking in the US. It is likely to become the first unless this chronic issue is addressed successfully (Wang et al., 2011; Danaei et al., 2009; Flegal et al., 2005; Mokdad et al., 2004). In addition, the total annual economic costs associated with obesity in the US were estimated to be in excess of $215 billion, with $147 billion in direct costs such as medical care expenditures and the rest in indirect costs including lost wages, productivity loss (due to absenteeism from the job), disability, premature deaths, health insurance, or transportation costs (Hammond and Levine, 2010; Finkelstein et al., 2009).

The above review frames obesity in terms of the definition of obesity, obesity trends in the US, and obesity's health and economic burden along with projections for the future. Nonetheless, obesity is a complex issue and thus, a further understanding is needed especially in terms of the obesity's confounding associations. In industrialized societies, almost all individuals are exposed to an array of behavioral and environmental influences that provoke obesity. There has been a great debate about the effects as well as the magnitude of such effects of the confounding factors' on obesity such as individual behaviors, including diet and physical inactivity, and obesogenic environments, both within the built and social environmental context. Therefore, the main concepts relating to such confounding factors are essential to review and will be elaborated further.

2.2 Diet and Physical Activity

The causes of obesity are intricate. However, it is known that poor diet and physical inactivity are major lifestyle contributors to obesity. In fact, there is a rather simple equation. Weight gain is the outcome of more calories consumed than burned. In other words, an individual's excessive daily food intake gets converted to body fat and stored unless it is used through a recreational or utilitarian physical activity. It is known that there is a global shift in diet toward increased consumption of energy dense foods that are low in vitamins and minerals and high in fat and sugars (Economos et al., 2015; WHO, 2015). The relative costs of energy-dense foods, composed of processed grains, sugar, and fat, have decreased while prices of fruits and vegetables have increased progressively in the more recent years (Popkin, 2011; Christian and Rashad, 2009). Although a range of individuals from various income levels are affected due to this price structure, individuals with low income profile are affected more substantially (Hawkes et al., 2015; Maillot et al., 2010). What is worse, those energy-dense foods and beverages have become easily accessible anywhere including vending machines, school cafeterias, pharmacies, book stores, and even hardware stores (Gordon-Larsen, 2014). Especially since the early 1980s, while energy intake has gradually increased, energy expenditure has decreased in the same rate if not more. Also, the trend toward increasing portion sizes has started in the same years (early 1980s) and has gradually increased ever since (Barbara et al., 2014). The increase in portion sizes is most evident and well documented in the US not only for eating out but also eating home among the adults, adolescents, and children alike (Barbara et al., 2014). These trends led to growing energy imbalance which strongly reflects the high obesity rates in the US (Economos et al., 2015).

In addition, a recent report by US Department of Agriculture and US Department of Health and Human Services claims that people are generally unaware of how many calories they need each day (http://health.gov/dietaryguidelines/dga2010/dietaryguidelines2010.pdf). Also, many people are either unaware of or underestimate how many calories their snacks, beverages, and meals have. Therefore, national improvements in dietary intake, in terms of food quality (i.e., healthy and nutritious foods), energy density, and portion size, have been identified as the main direction to prevent undesired weight gain and associated inverse health outcomes in the US (Karl and Roberts, 2014; Austin et al., 2011; Briefel and Johnson, 2004). However, there is still a great debate on how to implement improvements in dietary intake in practice (Karl and Roberts, mediation 2014). For instance, as documented by Hawkes et al. (2015) and Finkelstein et al. (2014), although the effects are modest, research suggests that food taxes and subsidies to encourage consumption of healthier foods and beverages may play an important role in reducing obesity especially when implemented along with other initiatives such as nutrition education and nutrition labeling on foods and beverages.

On the other hand, it is also important to emphasize that the type of ingredients (e.g., sugar types and many artificially produced additives and flavors) vary considerably across different foods and beverages (Pereira, 2014). It is believed that the type of ingredients plays a central role for palatability and thus greatly influences individuals' food selection (Johnson and Wardle, 2014). However, the effect of food ingredients and the extent of such an effect on eating behavior, obesity, or overall health are still not fully understood (Pereira, 2014). Therefore, food availability and convenient access to healthy (or unhealthy) food choices have received great attention recently in the obesity and nutrition literature (Gordon-Larsen, 2014). Neighborhood environments that provide access to healthy and nutritious foods are hypothesized to improve diet as well as weight outcomes of individuals living in such neighborhoods (Gordon-Larsen, 2014; Walker et al., 2010; Sturm and Cohen, 2009). Conversely, neighborhoods that limit access to healthy foods or food deserts have particularly received equal attention, if not more, in the literature.

Similar trends that strongly effects energy imbalance experienced by individuals also exist in terms of physical activity. There is a trend toward decreased physical activity due to the modern and relatively recent conveniences provided with technological advances. Immense improvements in science and technology made life much easier than it was a decade or two ago and people have become more inactive and automobile dependent. In other words, the modern conveniences have brought many sedentary forms of working, schooling, traveling, cooking, and even eating, to name a few. Therefore, some focus must be placed on different environments (e.g., social, physical, and cultural) that have an effect on individuals' choices for food intake and physical activity.

Since 1965, leisure time has increased more than 4 hours per week; time spent at productive activities such as working, cooking, cleaning, repairing things, and childcare has diminished; transportation time has increased to some extent; and time used for personal care including taking showers, getting dressed, and eating (roughly 2 hours a day) remained the same (Sturm and An, 2014; Finkelstein and Zuckerman, 2008). Thus, there has not been much change except the increase in leisure time. Nevertheless, according to various surveys, the majority of individuals tend to use much of their leisure time for sedentary activities rather than physical activities. For example, according to the data from the American Time Use Survey (ATUS), most recently collected in 2014, the US individuals only spend 17 minutes of their daily leisure time in average to participate in sports, exercise, and recreation. The total time for leisure and sports (in average per day) is calculated to be 5.1 hours based on the 2014 survey. The collected data include all days of the week and are annual averages for all persons 15 years and older (http://www.bls.gov/tus/charts/leisure.htm). Considering the fact that individuals spend only 17 minutes (out of 5.1 hours) to be physically active, it is quite insufficient in view of the expert recommendations.

Experts currently recommend a minimum of 30 minutes of moderate physical activity for five days or more per week, or a minimum of 20 minutes of vigorous physical activity for three days or more per week (Goodman and Fuller, 2015). Moderate physical activities cause small increases in breathing or heart rate and includes such activities as brisk walking, bicycling, vacuuming, and gardening. Vigorous physical activities cause large increases in breathing or heart rate and include running, aerobics, and heavy yard work (CDC, 2015). However, only less than half of the US adult population (48%) achieved those suggested levels of physical activity in between the 2001-2013 periods (CDC, 2015; Goodman and Fuller, 2015; NCHS, 2013).

As evident by the above review, lifestyle is one of the major contributors to the obesity epidemic. Therefore, the social and more specifically, physical environments that greatly affect individuals' lifestyle choices have become the focus to better understand the linkages that lead to obesity (Durand et al., 2011; Feng et al., 2010; Papas, 2007). In this case, it is crucial to review the literature focusing on built and social environments and thus they are discussed further.

2.3 Built and Social Environment

There is substantial and growing research linking characteristics of the built environment to health behaviors and outcomes such as diet, physical activity, and obesity. The same research also associates characteristics of the built environment to the disproportionate burden of these health risks among certain groups and/or subpopulations (Ferdinand et al., 2012; Casagrande et al., 2011; Ruel et al., 2010; Franco et al., 2009; Harrington and Elliott, 2009; Larson et al., 2009; Lovasi et al., 2009; Sallis and Glanz, 2009; Story et al., 2008; Frank et al., 2007; Handy and Clifton, 2007; Papas et al., 2007; Yancey and Kumanyika, 2007; Lopez and Hynes, 2006; Nelson et al., 2006; Frumkin et al., 2004; Dannenberg et al., 2003; Ewing et al., 2003; Hill et al., 2003; Srinivasan et al., 2003; Hill and Peters, 1998).

A thorough and recent literature search reveals that the studies of Ewing et al. (2003), Saelens et al. (2003) and Handy et al. (2002) are fundamental and pioneer studies linking aspects of the built environment with obesity and/or physical activity. It is also noticed that these studies have still been referred to by many recent studies. Ewing et al. (2003) has discovered significant associations between obesity, physical activity and the county level sprawl index that they have created. The study concludes that individuals living in densely populated counties weighted less, with lower BMI, than individuals residing in sprawling counties. Saelens et al. (2003) has emphasized that more walkable environments are usually associated with higher physical activity and lower obesity rates (Kent and Thompson, 2014; Ferdinand et al., 2012; Durand et al., 2011; Brownson et al., 2009; Papas et al., 2007). Furthermore, Ferdinand et al. (2012) have suggested that characteristics of the built environment are more likely to be associated with self reported than objectively measured physical activity. Handy et al. (2002) has stated that urban plans adopting land use and transportation concepts that promote walking and/or bicycling would help create more active, healthier, and livable communities (Kent and Thompson, 2014; Durand et al., 2011; Kaczynski et al., 2008). In that regard, there are two core concepts: land use, which consists of density and land use mix; and transportation, which consists of automobile dependence and connectivity (Jordan et al., 2015; Kent and Thompson, 2014; Frank et al., 2006; Frumkin et al., 2004).

Land use patterns are the main determinants of the proximity between different urban functional areas and therefore such patterns could be the key factors affecting individuals' physical activity decisions and hence energy expenditure. Density by definition is a measurement of the quantity of individuals, households, or businesses per unit of area such as an acre, a square mile, or a square kilometer. In urban planning, higher density of urban development is generally associated with more frequent and shorter trips, reduced automobile ownership, increased transportation options, and increased walkability and walking compared to lower density development (Frank et al., 2006; Frank et al., 2005; Saelens et al., 2003). The land use mix refers to the types of uses, such as housing, offices, retail, services, recreation, and so forth over an area. High level of land use mix is usually associated with reduced automobile dependence, reduced travel needs to distant places, and increased walkability (Frumkin et al., 2004; Saelens et al., 2003).

Similarly, transportation characteristics can also be decisive factors shaping individuals' physical activity decisions and hence alter energy expenditure. Automobile dependence is generally determined by multiple factors such as proximity between destinations, availability of different modes of transportation, and quality of connectivity. Transportation connectivity refers to how different destinations are linked through means of transportation. A poor connectivity between destinations, such as poor street networks, can make even nearby destinations inconvenient to reach (Frumkin et al., 2004). In planning, the number of intersections in an area is usually viewed as a good measure of connectivity (Saelens et al., 2003). Higher connectivity is generally associated with shorter automobile trips, decreased automobile dependence, and increased walkability and walking (Saelens et al., 2003).

Consequently, researchers in planning and transportation fields have identified land use mix, density, and connectivity as the key aspects to create walkability indices (Jordan et al., 2015; Kent and Thompson, 2014; Frank et al., 2006; Frank et al., 2005; Leslie et al., 2005) while controlling potential confounding factors such as socioeconomic and sociodemographic attributes. However, compared to adults, positive associations between density and walking are stronger and more consistent findings in youth while positive associations between land use mix and walking are stronger and more consistent findings in adolescents (Jordan et al., 2015). On the other hand, positive associations between street connectivity and walking are weaker and less inconsistent findings (Ding et al., 2011; Kerr et al., 2007; Frank et al., 2007). A number of studies have also highlighted other determinants of physical activity in the built environment in addition to those mentioned above, such as overall neighborhood design features, the presence and quality of nearby sidewalks and footpaths, the existence of enjoyable scenery, safety including lighting at night, and the presence of other individuals who are physically active (Durand et al., 2011; Ferdinand et al., 2012; Kaczynski et al., 2008; Frumkin et al., 2004).

All these built environmental attributes of a place play an important role on individuals' decisions about whether to engage in physical activity or not. However, the research investigating the links between the built and social environments and health outcomes has been unable to confirm a causal relationship. Because, individuals choose where to live and those who are conscious about certain health behaviors may be more likely to choose neighborhoods that provide better opportunities for physical activity and healthy diet. Nonetheless, even after controlling for possible effects of individuals' neighborhood choices, significant associations between characteristics of the built environment and health outcomes have been identified (Frank et al., 2007; Plantinga and Bernell, 2007). Hence, these findings indicated that the features of the built environment have an impact on individual behaviors and therefore on health outcomes (Frank et al., 2007; Plantinga and Bernell, 2007).

As illustrated in Figure 2.3, adverse characteristics of the social environment such as poverty, lack of education, unemployment, and crime may act synergistically to produce a cumulative impact of various stressors in individuals that directly increase their risk of disease. On the other hand, adverse characteristics of the physical environment such as availability of parks, grocery stores, restaurants, and public transportation create the context in which individual behavioral choices are made concerning physical activity, nutrition, tobacco and alcohol use, and other health-related behaviors (Iton et al., 2010). As a result poor health outcomes and thus health disparities becomes inevitable. It is clear that health disparities are often produced by multiple and complex adverse effects such as racial or ethnic background, sex, age, disability, socioeconomic status, physical environment, or geographic location.

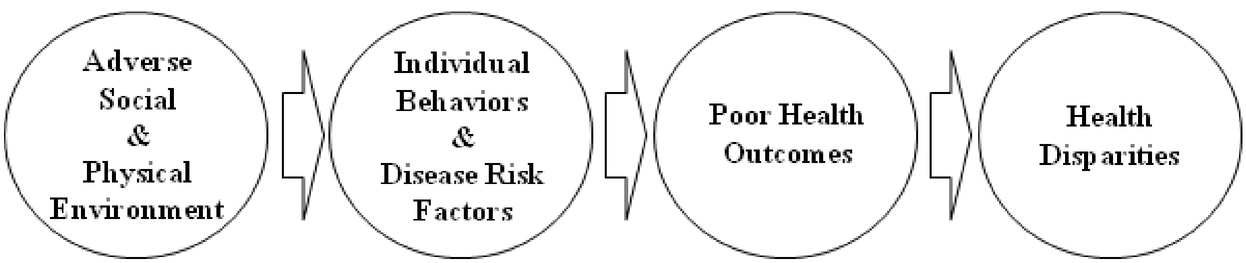

Figure 2.3 Possible pathways through which social and physical environment effects may operate

The cumulative effect of these built environmental, socioeconomic, and sociodemographic factors reflects the complex ways in which systematic inequalities and disadvantages translate into the places individuals live and the types and degrees of access to various resources and opportunities they have (Rossen and Pollack, 2012; Gordon-Larsen et al., 2006; Schulz and Northridge, 2004). For instance, low-income and some racial and ethnic subpopulations usually inhabit locations with greater access to fast-food restaurants, corner stores, and unsafe or undesignated urban open spaces. As a consequence, these subpopulations usually have limited access to preferable outlets such as healthy restaurants and supermarkets, gyms, or safe parks and playgrounds (Abercrombie et al., 2008; Franco et al., 2008; Gordon-Larsen, et al., 2006; Zenk et al., 2005; Block et al., 2004).

In addition, previous research also reported that individuals facing demographic, social, economic, environmental, and/or geographic (spatial) disadvantages are more likely to be vulnerable to risk factors and stressors (Lovasi et al., 2009; Wilson, 2009; Wilson et al., 2008). In return, being more vulnerable to risk factors in the social and built environments shapes an individual's propensity toward health and therefore contributes to numerous health disparities including disparities in obesity (Wilson, 2009; Wilson et al., 2008; Schulz and Northridge, 2004). Consequently, a number of studies (e.g., Taveras et al., 2013; Wen and Kowaleski-Jones, 2012; Wilson, 2009; Wilson et al., 2008; Drewnowski et al., 2007; and Swinburn et al., 1999) have emphasized that inequalities and disadvantages in the social and built environments underlie disparities in obesity risks and outcomes.

Based on the literature review above that primarily focused on the US, it is clear that the large increase in obesity prevalence in the US may reflect variations in built environmental, behavioral, socioeconomic, and sociodemographic factors. However, the extent to which these factors are causally related to the obesity phenomenon, in general and disparities in obesity, in particular is still poorly understood. This study will investigate the spatial associations between obesity and built environmental, behavioral, socioeconomic, and sociodemographic dynamics at the county level in the US and will attempt to determine whether such dynamics can explain the current obesity epidemic and disparities in obesity. Therefore, it is important to identify key variables to establish the possible spatial associations with obesity to reveal the pathways through which the underlying effects operate. The key variables used and the spatial and non-spatial methods applied to this research are discussed in Chapter 3 in detail.

CHAPTER 3. DATA AND METHODS

3.1 Study Area

The datasets used in this study come from various resources. The original datasets were at the county level for the entire US. However, due to the spatial nature of this study, only the contiguous counties and county equivalents (i.e., parishes in Louisiana, the District of Columbia, and the independent cities in Virginia, Maryland, Missouri, and Nevada) within the continental US are included in the analyses. The study area consists of the 3,105 counties and county-equivalents across the 48 contiguous states, also known as the continental US (all except Alaska and Hawaii) (Figure 3.1).

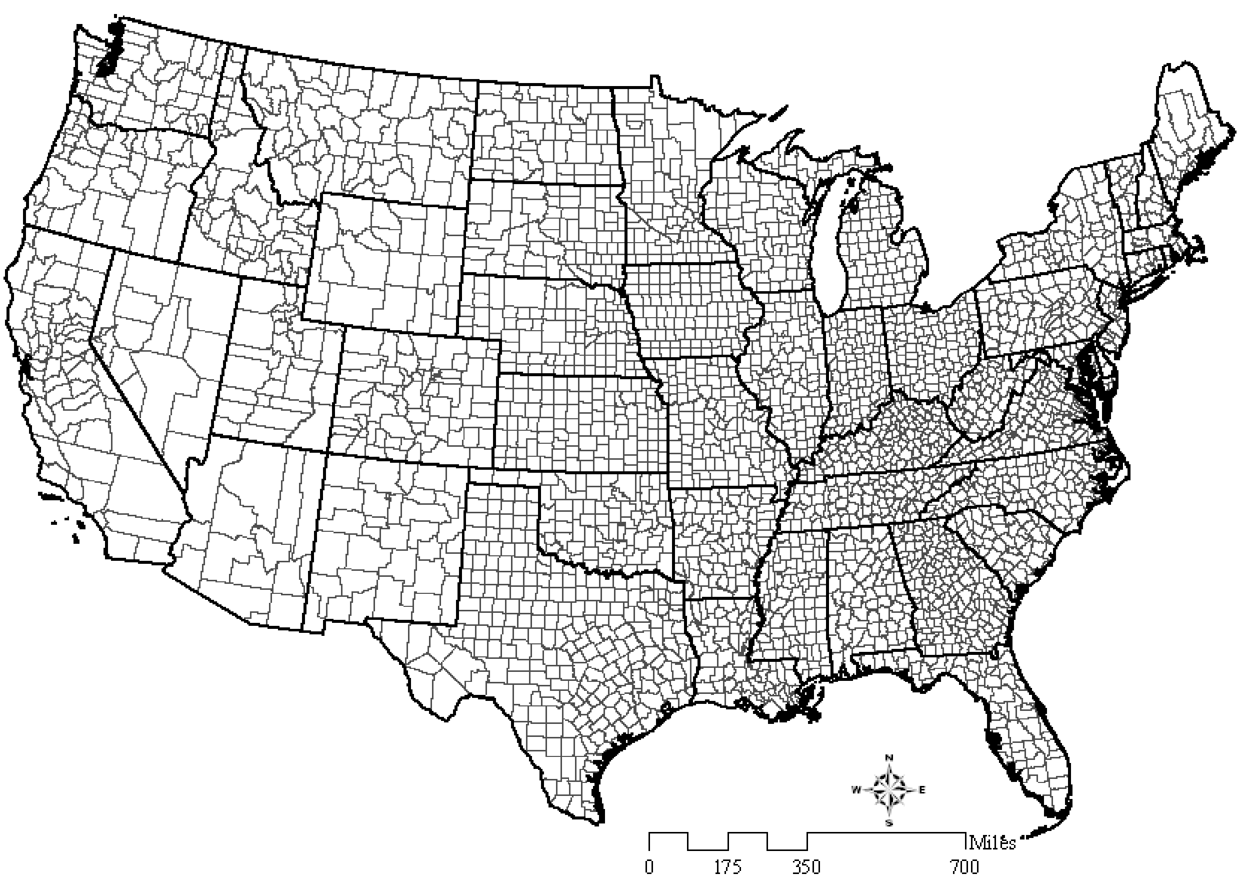

Figure 3.1 Study area showing the boundaries for counties, county-equivalents, and states for the continental US

3.2 Data

Thirteen datasets are used, as shown in Table 3.1. The first set of data includes: 1) obesity; 2) physical inactivity (PIA); 3) poverty; 4) ratio of jobs to employed residents (JER); 5) population-weighted distance (PWD) to parks; and 6) population density (Table 3.1). The secondary set of data includes: 1) obesity; 2) physical inactivity (PIA); 3) commute to work by car; 4) natural amenities index; 5) ratio of jobs to employed residents (JER); 6) fast-food restaurants per capita; 7) poverty; 8) African-American population; 9) high educational attainment; 10) rural population; and 11) native residents (Table 3.1).

The first set of data is mainly used to conduct the global and local Exploratory Spatial Data Analyses (ESDA) in Chapter 4. The variables included in the first set of data are selected to explore especially the built environmental, as well as the behavioral (PIA) and socioeconomic (poverty), associations with obesity. Based on the literature review as well as the earlier data exploration processes performed, these variables are indentified to represent such associations rationally at the county level. On the other hand, the secondary set of data is used to perform the global Ordinary Least Squares (OLS) regression and the local Geographically Weighted Regression (GWR) in Chapter 5. The variables included in the secondary set of data consist of only some of the same variables from the first set of data (i.e., obesity, PIA, poverty, and JER) and additional variables to capture sociodemographic and population associations with obesity as well. Although it is more logical and common practice to include all variables from a primary analysis when conducting a secondary analysis (for this study, the global OLS regression and the local GWR), in this study some of the variables have either been excluded or replaced with new ones due to issues of multicollinearity and/or low explanation power. Furthermore, a number of other variables (e.g., Gini index, per capita income, public transportation use, crime, female

Table 3.1 Description of the data sources

Variable	Link to the Data	Data Produced by	Year
Obesity	http://www.cdc.gov/diabetes/atlas/countydata/County_ListofIndicato	Behavioral Risk Factor Surveillance System (BRFSS)	2004, 2007, and 2010
Physical Inactivity	http://www.cdc.gov/diabetes/atlas/countydata/County_ListofIndicato	Behavioral Risk Factor Surveillance System (BRFSS)	2004, 2007, and 2010
Poverty	http://www.census.gov/did/www/saipe/data/statecounty/data/index.ht	US Census Small Area Income and Poverty Estimates (SAIPE) program	2004, 2007, and 2010
Ratio of Jobs to Employed Residents	http://www.fhwa.dot.gov/planning/census_issues/ctpp/data_products/	US Census Transportation Planning Products (CTPP)	2010
Population-weighted Distance to Parks	http://www.esri.com/data/data-maps/data-and-maps-dvd	Environmental Systems Research Institute (ESRI), Redlands, CA	2008
Population Density	http://www.census.gov/popest/data/index.html	US Census Population Estimates Program (PEP)	2004, 2007, and 2010
Commute to Work by Car	http://factfinder.census.gov/faces/tableservices/jsf/pages/productview	US Census 2010 American Community Survey (ACS) 5-Year Est.	2010
Natural Amenities Index	http://www.ers.usda.gov/data-products/natural-amenities-	US Department of Agriculture (USDA) Economic Research Service (ERS)	1999
Fast-food Restaurants Per Capita	http://ers.usda.gov/data-products/food-environment-	US Department of Agriculture (USDA) Food Environment Atlas	2010
African American Population	http://factfinder.census.gov/faces/tableservices/jsf/pages/productview	US Census 2010 Census Summary File 1 (SF 1)	2010
High Educational Attainment	http://factfinder.census.gov/faces/tableservices/jsf/pages/productview	US Census 2010 American Community Survey (ACS) 5-Year Est.	2010
Rural Population	http://factfinder.census.gov/faces/tableservices/jsf/pages/productview	US Census 2010 Census Summary File 1 (SF 1)	2010
Native Residents	http://factfinder.census.gov/faces/tableservices/jsf/pages/productview	US Census 2010 American Community Survey (ACS) 5-Year Est.	2010
State, County, and Parish Boundaries	http://www.census.gov/geo/maps-data/data/tiger-line.html	US Census TIGER products	2010

employment, unemployment, occupation type, gender, median age, etc.) are also explored in the earlier phases of this study, however, not all are used due to various reasons including multicollinearity and/or low correlation with obesity.

The obesity and PIA data used in this study are basically the county level prevalence rates obtained from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) (http://www.cdc.gov/diabetes/atlas/countydata/County_ListofIndicators.html). The BRFSS has been established and administered by the Behavioral Surveillance Branch of the Centers for Disease Control and Prevention (CDC). The CDC collaborates with states and other partners to provide health promotion, prevention, and preparedness and maintain national health statistics (CDC, 2014). Established in 1984 by the CDC, the BRFSS is an ongoing state-based system of health surveillance tracking health conditions and risk behaviors of the adult population in the US (CDC, 2014). The BRFSS collects the surveillance data through monthly telephone interviews from a random sample of adult population who are 18 years of age or older. The BRFSS questionnaire used in the interviews consists of three parts: 1) the core component; 2) optional module; and 3) state-added questions (BRFSS, 2014). The core component is a standard set of questions asked by all states and varies for each specific year. The core variables that are common to each year within these periods include obesity (BMI), exercise (leisure time physical activity), general health status, heath care access, immunization, diabetes, asthma, HIV/AIDS, tobacco use, alcohol use, and demographic variables such as race, age, education level, and income. The moderate physical activity and vigorous physical activity, two of the core variables, exist only in the odd years. In addition, core variables such as women’s health, prostate cancer, colorectal cancer, nutrition, hypertension, disability, and so on are not common for each year yet exist in some of the years. On the other hand, the optional module is a set of questions submitted by the CDC for each specific year and the states elect which questions to be used in their questionnaires (BRFSS, 2014). The optional module variables include oral health, heart attack and stroke, anxiety and depression, arthritis, family planning, home environment, and more. In addition to the core component and optional module, the states can also develop their own questions for use in their questionnaire. However, the CDC does not evaluate the state-added questions (BRFSS, 2014).

Consequently, in this study the county level obesity and PIA prevalence rates depict the percentage of adult individuals who are obese and the percentage of adult individuals who are physically inactive, respectively. The obesity and PIA rates do not include children and/or adolescents. These prevalence rates are derived based on the responses to the questions regarding height, weight, and exercise in the above mentioned core component. The respondents are considered obese if their body mass index (BMI) is 30 or greater. The respondents are considered to be physically inactive if they answered "no" to the question, "During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?" (BRFSS, 2014). The obesity prevalence rates are calculated by dividing the total number of adult individuals who are obese (having a BMI value greater than or equal to 30) in the county by the total county population (18 years and over) and then multiplying it by 100. Similarly, the PIA prevalence rates are calculated by dividing the total number of adult individuals who are physically inactive (answered "no" to the physical activity question) in the county by the total county population (18 years and over) and then multiplying it by 100.

The poverty data used in this study are the county level prevalence rates acquired from the US Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program (http://www.census.gov/did/www/saipe/data/statecounty/data/index.html). The official poverty thresholds do not vary geographically. However, the thresholds are updated for inflation using Consumer Price Index (CPI-U) for each estimation period to classify impoverished individuals more rationally (https://www.census.gov/hhes/www/poverty/about/overview/measure.html). Otherwise, it is difficult to know that an increasing poverty rate really means that more people are impoverished. The SAIPE program provides annual estimates of income and poverty statistics for all school districts, counties, and states in the United States. The SAIPE program produces county estimates for the total number of people in poverty, number of related children ages 5 to 17 in families in poverty, number of children under age 18 in poverty, and median household income (US Census Bureau/SAIPE, 2014). Therefore, in this study, the county level poverty prevalence rates represent the percentage of individuals living in poverty. The poverty prevalence rates are calculated by dividing the total number of individuals who live in poverty in the county by the total county population and then multiplying it by 100.

The JER data used in this study are the county level rates obtained from the US Census Transportation Planning Products (CTPP). The CTPP is a set of special tabulations and uses the American Community Survey (ACS) sample to create the special tabulations (http://www.fhwa.dot.gov/planning/census_issues/ctpp/data_products/acsdataprod.cfm). The CTPP data mainly consists of three parts: 1) residence-based tabulations summarizing worker and household characteristics; 2) workplace-based tabulations summarizing worker characteristics; and 3) worker flows between home and work, including travel mode. The CTPP provides data at various geographic levels ranging from state to Transportation Analysis Zone (TAZ) levels (CTPP, 2014). Hence, for this study the county level JERs are calculated based on the information on the number of jobs and employed residents available in the CTPP data. The county level JERs are calculated by dividing the total number of jobs in the county by the total number of employed residents in the county.

In this study, the JER variable is used as a proxy to capture the degree of land use mix due to limitations in data availability to capture built environmental aspects at the county level. Although the JER variable may have its own limitations to use as a built environmental variable to capture the degree of mixed land use, it is hypothesized that higher JERs are associated with lower obesity rates especially considering use of the JER variable in various studies. Higher mixed land uses decrease the average commute distances and per capita vehicle travel in general since with higher mixed land uses, proximity to various places to do daily activities such as commuting to work or school, shopping, and recreation reduce (Litman, 2008). In that regard, Litman (2008) has used JER in an area to measure land use mix. Furthermore, in urban growth and planning literature, JER is used as an important determinant for commuting times (Sultana, 2002; Cervero, 1996). Higher JER is found to be associated with a shorter residence commuting time and a longer workplace commuting time. For instance, jobs are better supplied in the downtowns and that indicates a higher JER (Peng, 1997; Cervero, 1989). In other words, a JER higher than 1 in an area indicates that it is a job rich area with commercial uses (i.e., high land use mix). On the other hand, a JER lower than 1 in an area indicates that it is a job poor area with mostly residential land use (i.e., low land use mix). Therefore, the JER variable is used as a proxy to capture the degree of land use mix.

The PWD to parks data used in this study are the county level distances in miles constructed using the US park and US Census population data. The US park data is obtained from the park GIS layer in the ArcGIS 9.3 Data DVD (Environmental Systems Research Institute (ESRI), Redlands, CA). The particular data DVD, which was used to create the data used in this research, was created in 2008. An updated version of the data DVD exists with the release of ArcGIS 10.1 recently in 2010 (http://www.esri.com/data/data-maps/data-and-maps-dvd). The data DVD included 35,436 public park and/or forest units at various geographic levels, ranging from state to local, in the 50 states and DC. Very small parks that are less than 0.1 acres are not available in the dataset and therefore are not included in this study. Thus, in this study the county level PWDs to parks provide a measure in terms of park accessibility. Six sequential steps are involved in calculating and creating PWDs to parks at the county level. As stated first by Zhang et al. (2011) and then by Wen et al. (2013), these steps include: 1) A Euclidean straight line distance between a census block centroid and a park centroid is calculated; 2) The access potential from a census block to a park is calculated as the ratio of park size and the squared distance between them; 3) A sum of a census block's access potentials to its nearest seven parks is calculated as its spatial park access index (based on cognitive research on choice set formation, seven is the most likely size for a spatial destination choice set); 4) The access probability from a census block to a park is calculated as the ratio of the access potential between them and the census block's park access index (the sum of all seven access potentials); 5) PWD to parks for a census block is calculated as the sum of census block population multiplied by access probability and distance for all its seven nearest parks; 6) A county's PWD to parks is calculated as a sum of block PWD to parks multiplied by block population divided by total county population.

The population density data used in this study are the county level density measures obtained from the US Census Bureau's Population Estimates Program (PEP). Each year, the PEP uses current data to produce series of estimates of population, demographic components of change, and housing units (http://www.census.gov/popest/data/index.html). These official estimates of the population are at the national, state, county, and city and town levels, while the estimates of housing units are only at the state and county levels (US Census Bureau/PEP, 2014). The population densities are derived using the estimates of the population and corresponding land areas (in square miles) available in the PEP data. Therefore, in this study the county level population densities show the number of individuals living per square mile. The county level population densities are calculated by dividing the total number of individuals in the county by the total county land area (in square miles).

The commute to work by car data used are the county level percentages obtained from the US Census Bureau's 2010 American Community Survey (ACS) 5-Year Estimates (http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_5YR_S0802&prodType=table). In this study, the county level commute to work by car percentages represent the percentage of individuals who commute to work by car, truck, or van, and drive alone. The commute to work by car percentages are calculated by dividing the total number of individuals who commute to work by car in the county by the total county population (16 years and over) and then multiplying it by 100. Furthermore, in US cities car is the main mode of transportation for all purposes but especially to commute to work. Commute to work by car may be considered as a passive commuting type when compared to the act of walking or biking and even when compared to the use of public transportation to get to work (Zwald et al., 2014). Public transportation is considered as an active commuting type mainly because it is multimodal often involving walking or biking. Thus, it is an important variable to assess both the behavioral and built environmental characteristics.

The natural amenities index data used in this study are the county level measure of the physical characteristics of a county area obtained from the US Department of Agriculture's

(USDA) Economic Research Service (ERS) (http://www.ers.usda.gov/data-products/natural-amenities-scale.aspx). The USDA has assumed that the index is relatively constant since it is not expected to change much over time. Thus, the USDA has not updated its data beyond the initial 1999 construction. The natural amenities index is mainly created as "a measure of the physical characteristics of a county area that enhance the location as a place to live" (USDA/ERS, 2015). The index is constructed by combining six measures of climate (i.e., warm winter, winter sun, temperate summer, and low summer humidity), topography (i.e., topographic variation), and water area (i.e., access to a body of water such as lakes, ponds, or oceanfront) that reflect environmental qualities that most people prefer, as USDA suggests. The index ranges from 1 to 6, where 1 is the lowest amenity score and therefore the least preferable while 6 is the highest score and thus the most preferable. Consequently, in this study the county level natural amenities index represents the aspects of the natural environment. The index for a specific county is calculated based on the standard deviation from the mean for all counties. A county having a large negative value gets a much lower index score than a county with a large positive value which indicates the existence of natural environment superiority. Furthermore, the natural amenities index variable is included in this study with the notion that it is an essential one for such study. For example, Michimi and Wimberly (2012) found that recreational opportunities and therefore propensity for physical activity were positively correlated with natural amenities index. It is also widely accepted that the aspects of environments that encourage walking and other physical activities have a great potential to reduce obesity. While much of the literature has focused on urban form as well as planning characteristics, additional aspects of urban environments such as natural amenities should also be considered (Lovasi et al., 2012).

The fast-food restaurants per capita data used in this study are the county level rates acquired from the US Department of Agriculture's (USDA) Food Environment Atlas (http://ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads.aspx). The USDA uses the US Census Bureau's County Business Patterns (CBP) series to create the fast-food restaurants data. CBP is an annual series that provides economic data by industry at the subnational scale. This series includes data on the number of establishments, employment, first quarter payroll, and annual payroll (US Census Bureau/CBP, 2015). Fast-food restaurants include establishments which primarily provide food services (except snack and nonalcoholic beverage bars) where customers generally order or select items and pay before eating. Food and drink may be consumed on premises, taken out, or delivered to the customer's address (USDA/Fast Food Atlas, 2015). Therefore, in this study the county level fast-food restaurants per capita rates show the number of fast-food restaurants in the county per 1,000 county residents. The fast-food restaurants per capita rates are calculated by dividing the total number of fast-food restaurants in the county by the total county population and then multiplying it by 1,000.

The African-American population data used in this study are the county level percentages obtained from the US Census Bureau's 2010 Census Summary File 1 (SF 1) (http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_SF1_P3&prodType=table). SF 1 contains the data compiled from the questions asked of all people and about every housing unit. The data collected include detailed information about age, sex, race, Hispanic/Latino origin, households, families, housing units, owner/renter status, and urban/rural identification (US Census Bureau/2010 Census SF 1, 2015). Therefore, in this study the county level African-American population percentages represent the percentage of mono racial Black or African-American individuals. The African-American population percentages are calculated by dividing the total number of Black or African-American alone individuals in the county by the total county population and then multiplying it by 100.

The high educational attainment data used are the county level percentages obtained from the US Census Bureau's 2010 American Community Survey (ACS) 5-Year Estimates (http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_5YR_DP02&prodType=table). The ACS is a nationwide survey that produces a detailed statistical portrait of the US social, economic, housing, and demographic characteristics through 1-year, 3-year, and 5-year ACS products (US Census Bureau/2010 ACS 5-Year Estimates, 2015). The 5-year estimates have the largest sample size and include data for all counties. Therefore, in this study the county level high educational attainment percentages represent the percentage of individuals who completed at least a Bachelor's degree or higher. The high educational attainment percentages are calculated by dividing the total number of individuals who completed a Bachelor's degree or higher in the county by the total county population (18 years and over) and then multiplying it by 100.

The rural population data used in this study are the county level percentages acquired from the US Census Bureau's 2010 Census Summary File 1 (SF 1) (http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_SF1_P2&prodType=table). SF 1 contains information about urban/rural status (US Census Bureau/2010 Census SF 1, 2015). Therefore, in this study the county level rural population percentages show the percentage of individuals living in rural areas. The rural population percentages are calculated by dividing the total number of individuals living in rural areas of the county by the total county population and then multiplying it by 100.

The native residents data used in this study are the county level percentages acquired from the US Census Bureau's 2010 American Community Survey (ACS) 5-Year Estimates (http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_5YR_DP02&prodType=table). In this study, the county level native residents percentages show the percentage of individuals who were born in the state in which they reside. The native residents percentages are calculated by dividing the total number of county residents who were born in the state in which they reside by the total county population and then multiplying it by 100.

Many times, residents who reside in the state in which they were born address to themselves by expressions such as "homegrown" or "born and raised" to indicate that they have strong ties with the place in which they reside. In this study, the main drive in using native residents as an explanatory variable stems from the idea that communities or places, to where individuals are belong, shape some lifestyle habits (e.g., eating, having fun, living, and working) through local culture. Therefore, place of birth may be used to establish linkages between obesity outcomes and domestic immobility. On the other hand, "native population" also includes return migrants. Therefore, places with high percentages of return migrants (or boomerang migrants) may have more contamination of the characteristics captured with the "native residents" variable and thus, caution needed when interpreting analysis results.

The state and county level boundary data used in this study are downloaded from the US Census Bureau's TIGER (Topographically Integrated Geographic Encoding and Referencing System) products (http://www.census.gov/geo/maps-data/data/tiger-line.html). The TIGER products include various geodatabases and shapefiles for the entire US for mapping applications. Therefore, in this study these boundary files are used as a means to display the county level data (US Census Bureau/TIGER, 2014).

3.3 Methods

3.3.1 Main Concepts in Spatial Analysis

Spatial data and spatial analysis concepts arise when dealing with problems of space. Spatial data are those that have a spatial component and combine locational information with attribute information recorded at a specific location. Spatial analysis represent a step beyond simple cartography and mapping by comparing the spatial distribution of a set of features to a hypothetically-based random spatial distribution and by determining the relationship between the spatial distribution of variables over space (Mitchell, 2005). Such spatial distributions or patterns are of crucial interest to many areas of geographic research mainly because they can help identify and quantify patterns of features in space so that the underlying processes that lead to the particular distribution can be determined (Fotheringham et al., 2002).

One of the fundamental rationales behind using spatial analysis in geographic research, in general and in health-related research, in particular is the concept of "spatial autocorrelation." The first law of geography, also known as Tobler's first law, states that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). Spatial autocorrelation, measured by various formal statistics, provides coefficients that quantify what this law states. When locations with similar attribute values (high or low) are closely distributed in space, the spatial pattern is considered to exhibit positive spatial autocorrelation. Conversely, if closely associated locations have dissimilar attribute values, the spatial pattern is considered to show negative spatial autocorrelation. If no spatial relationship exists between attribute values and their locations, the spatial pattern is considered to exhibit zero spatial autocorrelation. According to Anselin and Bera (1998), of the three types (i.e., positive, negative, and neither) of spatial autocorrelation, positive autocorrelation is by far the most intuitive.

Spatial analysis can basically be classified into three major categories consisting of exploratory spatial data analysis (ESDA), visualization, and spatial modeling (Vinnakota, 2006). ESDA is one of the most prevalent methods of spatial analysis (Mitchell, 2005). ESDA consists of a set of statistical and analytical methods and techniques designed to describe spatial distributions, identify spatial patterns and/or variations such as unusual locations or outliers, and discover spatial clusters (Munch et al., 2003). However, ESDA methods and techniques attempt neither to explain the occurrence of patterns nor to make any causal assumptions. These methods mainly attempt to assess and depict non-random and interesting spatial patterns (Gatrell, 2002). Therefore, ESDA is traditionally considered as a fundamental step in the spatial analysis of health-related phenomena.

Visualization of spatial data is most commonly accomplished using a geographic information system (GIS). GIS simply consists of computer based sets of procedures that capture, store, manipulate, analyze, model, and visualize data with spatial characteristics. Although GIS and spatial analysis initially started as two separate fields of research and application, they have become closer fields over time that each support and add value to the other (Goodchild and Haining, 2004). Thus, GIS's capabilities to analyze and visualize both temporally and spatially complex data sets make it an excellent platform to study health-related spatial phenomena.

Spatial modeling used in the analysis of health-related patterns can be undertaken using a variety of spatial statistical techniques such as spatial clustering, spatial dispersion, spatial auto-correlation, and Local Indicators of Spatial Association (LISA) (Vinnakota, 2006; Anselin, 1995). Spatial clustering techniques are designed primarily to detect spatial clusters of the phenomena under investigation across space. In health-related research, the detection of clusters through spatial analysis can offer crucial insights to tackle both causation and identification of potential risk factors. A number of methods have been developed specifically to detect spatial clusters. The main concern of the various cluster detection methods is to identify locations where there exists an unusually higher or lower number of events than expected (widely also known as hotspots and coldspots) (Lloyd, 2007). Cluster detection methods may be divided into two groups: The ones formulated for identifying clusters and the ones formulated for statistical significance testing of clustering (Besag and Newell, 1991). Futhermore, "tests may be either focused or general in nature," depending on whether we have prior considerations for any links between locations and occurrence of events (Besag and Newell, 1991). In the literature, cluster detection methods such as the geographical analysis machine (Openshaw et al., 1987), Besag-Newell clustering test (Besag and Newell, 1991), spatial scan statistic (Kulldorff and Nagarwalla, 1995), and local indicators of spatial association (LISA) (Anselin, 1995) are fundamental in terms of detecting and testing clusters.

3.3.2 Spatial Analysis Methods Applied to This Research

Global and local ESDA techniques include the Moran's I and local indicators of spatial association (LISA) as the primary spatial analysis methods and the Geographically Weighted Regression (GWR) as the secondary spatial analysis methods. Both sets of techniques are applied to this research and will be described in the following section of this dissertation.

3.3.2.1 Global Moran's I and Local Indicators of Spatial Association (LISA)

ESDA consists of a set of techniques that are used to describe and visualize spatial distributions, identify atypical locations or spatial outliers, and discover patterns of spatial associations and clusters. The process of ESDA starts with simple mapping and geovisualization, moves on to exploration, and concludes with spatial autocorrelation analysis (Anselin, 2005).

Several formal statistics are designed to measure spatial autocorrelation and these statistics include simple indices such as the Moran's I, Geary's C, and Getis-Ord G (Levine, 2015). However, of all three statistics, the Moran's I is considered as a more powerful test than the other two simply because of its ability to correctly reject a false null hypothesis or in other words, to avoid a Type II error. For instance, a data set for which the Moran's I is barely statistically significant for positive spatial autocorrelation might simply fail to be statistically significant with the Geary's C or the Getis-Ord G due to the fact that the Geary's C and the Getis-Ord indices are not as powerful as the Moran's I index (Levine, 2015). On the other hand, only the Getis-Ord G can differentiate between high positive and low positive spatial autocorrelation. High positive indicates that there are more zones with high values located near to other zones also with high values whereas low positive indicates the opposite case where low values are near to other low values. The Moran's I and Geary's C statistics do not distinguish between these conditions. Furthermore, of these three statistics, only the Moran's I and Geary's C can identify negative spatial autocorrelation (Levine, 2015). In other words, the Getis-Ord G does not indicate a dispersion process, even if it is present in the data.

When the advantages and disadvantages of all three formal statistics are weighted along with the diversity of the datasets under examination and the scope of this study, the Moran's I appear to be a better test to use. More specifically, the use of the Getis-Ord G is eliminated mainly because it cannot detect negative spatial autocorrelation. Areas with negative spatial autocorrelation may reveal interesting associations that need further exploration and therefore such areas needs to be detected. On the other hand, the Geary's C is inversely related to Moran´s

I and thus the Geary's ratio is similar to Moran's I. However, the Moran's I is more of a global indicator whereas the Geary coefficient is more sensitive to differences in small neighborhoods. In other words, the Geary's C emphasizes differences in values between pairs of observations, rather than the covariation between the pairs. Therefore, use of the Moran's I over the Geary's C suits better to detect spatial autocorrelation at the county level for the entire US. Last but not the least, based on the literature, it is also evident that the Moran's I test (Moran, 1950) is one of the oldest and most commonly used indicators of spatial autocorrelation (Holt, 2007).

The global Moran's I statistic is a single measure that describes the general extent of spatial clustering of observed values across a region. It indicates the relationship between a vector of observed values, y, and a weighted average of values that are contiguous to y, conditional on the specific neighborhood structure imbedded in the chosen weights matrix (Levine, 2015; Bailey and Gatrell, 1995; Ebdon, 1985; Moran, 1950). The weighted average of values are generally referred to as the "spatial lag of y" and expressed as W_y. The resulting value of Moran's I is basically the slope of a linear regression of W_y on y (Anselin and Bera, 1998; Anselin, 1988).

Associations among locations are defined using a spatial weights matrix. The weights matrix defines the neighbors for each location in the study area using either a distance matrix that contains distances between all possible pairs of locations or a contiguity matrix that evaluates and specifies common boundaries (e.g., rook contiguity, queen contiguity) (Martinez et al., 2014).

Formally, the global Moran's I is defined as:

$$I = \frac{n}{\sum_i^n \sum_j^n w_{ij}} \frac{\sum_i^n \sum_j^n w_{ij}\,(y_i - \bar{y})(y_j - \bar{y})}{\sum_i^n (y_i - \bar{y})^2} \qquad \text{Formula 3.1}$$

where n is the total number of observations, y_i is the value of a variable at a particular location, the *ith* location, y_j is the value of the same variable at another location, the *jth* location, (where $i \neq j$), $\bar{y}$ is the mean of the variable, and w_{ij} is a weight applied to the comparison between location i and location j. The weight variable, w_{ij}, is a contiguity matrix. If locations i and j are adjacent w_{ij} =1 and zero otherwise (w_{ii} =0, a region is not adjacent to itself). This type of contiguity is referred to as a "first-order contiguity." Therefore, the resulting I is used to test hypotheses concerning similarity.

The global Moran's I values range from -1.0 indicating the highest value for dissimilarity among neighbors (perfect dispersion or negative spatial autocorrelation) to +1.0, indicating the highest value for similarity (perfect clustering or positive spatial autocorrelation). A value slightly below 0 (negative), but very close to 0 indicates a random spatial pattern (no spatial relationship). Since the Moran's I is considered as a special type of a slope coefficient of a correlation in that it compares the sum of the cross-products of values at different locations, the magnitude of the Moran's I value itself is also important (Levine, 2015). In general, when nearby areas have similar values, their cross product is high and conversely, when nearby areas have dissimilar values, their cross product is low. Therefore, a Moran's I statistic that is high is a sign of more spatial autocorrelation than a Moran's I statistic that is low (Levine, 2015). Tests for significance use z-scores.

However, the "global" measure of spatial autocorrelation cannot identify the locations of spatial clusters or spatial outliers that are exhibited as a result of the spatial autocorrelation process. The statistic only indicates the presence of spatial autocorrelation globally for the entire study area. To generate statistics and map the location of spatial clusters, the Local Indicators of Spatial Association (LISA), the local counterpart of the Moran's I, has been developed (Anselin, 1995). It is an important tool to gain an understanding of the localized extent and nature of spatial clustering in a dataset (Voss et al., 2006).

The LISA provides a measure of spatial autocorrelation for every areal unit and it is formally defined as:

$$I_i = (y_i - \bar{y}) \sum_j^n w_{ij} (y_j - \bar{y}) \qquad \text{Formula 3.2}$$

where n is the total number of observations, y_i is the value of a variable at a particular location, y_j is the value of the same variable at another location, $\bar{y}$ is the mean of the variable, and w_{ij} is a weight applied to the comparison between location i and location j.

With the utilization of the LISA statistic, cluster maps and corresponding significance maps are generated. A cluster map is basically a special type of a choropleth map illustrating locations based on the significant local Moran's I statistics. These locations on the cluster map are classified by the type of spatial autocorrelation. The high-high and low-low locations indicate clustering of similar values while the high-low and low-high show spatial outliers (dissimilar values that are close to each other). Likewise, the significance map is a special type of choropleth map. It shows the very same locations included in the cluster map but reclassifies them based on the significance levels. Depending on the spatial distribution of the data and the number of permutation runs, the significance levels can be insignificant, $p<0.05$, $p<0.01$, $p<0.001$, and $p<0.0001$ (Anselin, 2003).

The global Moran's I and LISA tests can be either in the univariate or the bivariate form. Even though an univariate measure is used in most studies, it is also possible to get statistics based on bivariate measures. When an univariate form is used, the Moran's I and/or LISA statistics compare the value of a variable at any one location with the value of the same variable at all other locations. On the other hand, the same two statistics compare the value of a variable at any one location with the value of a second variable at all other locations when bivariate measures are used (Martinez et al., 2014; Levine, 2015). In other words, in the univariate form of the Moran's I and/or the LISA, the observed values, y, and the spatial lag of y are based on the same variable, whereas in the bivariate form, the observed values of a variable, y, are regressed on the spatial lag of the observed values of another variable, x.

According to Anselin (1995), it is a good practice to supplement the assessment of global spatial autocorrelation by local measures of spatial dependence such as LISA, especially with larger data sets for thorough understanding of spatial associations and processes.

In this study, the global Moran's I and LISA algorithms both in the univariate and bivariate forms were employed using GeoDa 1.6.7 to compute global and local Moran's I indices for each county and county equivalent in the continental US. GeoDa 1.6.7 is a stand-alone program developed by Luc Anselin and his team. Both the most recent (GeoDa 1.6.7) and older (GeoDa 0.9.5-i that runs on Windows XP only) versions of the software can be downloaded from the GeoDa center at Arizona State University (ASU) https://geodacenter.asu.edu/software/downloads to do exploratory and confirmatory spatial data analyses. It uses ESRI shapefiles as input and output coverage (Anselin, 2005). It includes functionality for data manipulation, mapping, and spatial regression analysis, as well. More specifically, GeoDa includes applications to perform both global Moran's I and LISA along with other supporting analytical applications. GeoDa is particularly useful for the analysis of area patterns, since the statistical applications included are mostly limited to performing polygon pattern analysis. GeoDa can do analysis of lattice (polygon) data, but is very limited in the analysis of point data.

The GeoDa outputs for the global Moran's I and LISA statistics also help to test hypotheses for clustering. In hypothesis testing, the initial assumption is that the attribute distribution is one of complete spatial randomness (CSR). In other words, the CSR assumption is set as the null hypothesis (H_0: The spatial distribution of polygons and their attributes are spatially independent; H_A: The spatial distribution of polygons and their attributes are not spatially independent) and the attribution distribution is compared against a set significance level to either accept or reject the null hypothesis (Eck et al., 2005). If the z statistic lies beyond a critical value, then the null hypothesis is rejected. In a case where the z-score is positive (positive spatial autocorrelation), the spatial distribution of locations with similar attribute values (high or low) is more spatially clustered than would be expected if underlying spatial processes were random. Conversely, in a case where the z-score is negative (negative spatial autocorrelation), the spatial distribution of locations with high or low attribute values is more spatially dispersed than would be expected if underlying spatial processes were random.

GeoDa uses permutations to calculate significance levels for both global and local spatial autocorrelation statistics. Since the p-values are dependent on the number of permutation runs, they are pseudo p-values (pseudo significance levels). In other words, permutations are used to determine how likely it would be to observe the existing spatial distribution of the data at hand, if the existing values are reshuffled over space under conditions of spatial randomness. The number of permutations is specified by the user (Anselin, 2003).

The pseudo significance level is derived from:

$$(M+1)/(R+1) \qquad \text{Formula 3.3}$$

where M is the number of instances where a calculated statistic from the permutations equals or exceeds the observed value for positive Moran's I or equal to or less than the observed value for negative Moran's I and R is the number of permutation runs.

As an example, if the observed Moran's I exceeds any of the randomly generated Moran's I values, meaning that M is equal to zero, the pseudo p-value would be 1/1000= 0.001 for 999 permutations (R being equal to 999). Similarly, the pseudo p-value would be 1/100= 0.01 under the same conditions except with 99 permutations (Anselin, 2003).

All aspatial and spatial exploratory data analyses, particularly the tests for spatial autocorrelation, assist to answer various questions about a dataset including: 1) How are the features distributed? 2) What is the pattern created by the features? 3) Where are the clusters? 4) How do patterns and clusters of different variables compare to one another? When patterns are explored both statistically and spatially as in global Moran's I and LISA tests, this leads to a better understanding of geographic phenomena especially in terms of patterns and clusters.

3.3.2.2 Geographically Weighted Regression (GWR)

There are various approaches to the analysis of spatial data. However, logical and suggested steps in spatially analyzing event distributions include creating rate or event distribution maps, using spatial statistics to determine whether or not the events or rates are spatially autocorrelated, and detecting and identifying the locations of clusters (e.g., hot-spots or cold-spots) and spatial outliers (Wilson and Fotheringham, 2007). These are considered the primary steps in Exploratory Spatial Data Analysis (ESDA). In order to assess why clusters and outliers exist where they do as well as to assess the underlying processes that created such patterns through the relationships between the phenomenon and selected variables under investigation, secondary approaches, such as regression analysis or spatial data process models are needed. Examining "causality" through multiple regression analysis, specifically using the Geographically Weighted Regression (GWR), which is a local spatial statistical technique used to test for spatial nonstationarity, is one of the most commonly used techniques. Spatial nonstationarity is a condition in which the relationships between influences and outcomes of interests are heterogeneous (not uniform) over the entire area under consideration (Goodchild and Janelle, 2010). In other words, the relationships vary across the geographic area under investigation reflecting the spatial structure embedded within the covariates and the outcome variable. Therefore the "global" models cannot explain such a geographically varying relationship. The main concepts of the GWR will be introduced briefly.

In a global model, the underlying assumption is that the relationships between the covariates and the outcome variable are homogeneous (or stationary) across space. In other words, the global model assumes that the same stimulus provokes the same response across the entire study area. However, the geographic phenomena and underlying processes and therefore the relationships between variables are often intrinsically spatial (space dependent) and vary geographically in 'real-world' situations (Fotheringham et al., 2002).

Traditional global modeling techniques such as the ordinary least squares (OLS) linear regression or spatial regression methods cannot detect nonstationarity and therefore cannot adequately capture spatially varying relationships between the covariates and the outcome variable. In this case, the parameters calculated by the OLS linear regression will likely be biased and generate low estimates of the confidence intervals and real variance (Ward and Gleditsch, 2008). Therefore, in contrast to traditional linear regression models, the GWR technique was created and proposed in the geography literature to allow relationships in a regression model to vary over space and be measured within a single modeling framework (Fotheringham et al., 2002).

The GWR is a locally weighted, linear, and nonparametric estimation method and, therefore spatial effects from a local point of view can be examined utilizing the GWR (Brunsdon et al., 1996). Using the GWR approach, the spatial variation of the regression coefficients can be captured for each event or area based on the attribute values of the neighboring observations. This method uses a kernel function to determine the size of the window, which generates a subset of the data around a specific point. Furthermore, the GWR method allows estimating local parameters rather than global ones through the allocation of weights.

In summary, the GWR extends the capabilities of the OLS linear regression models by accounting for the spatial structure of the data. It estimates a separate model as well as local parameter estimates for each geographic location based on a local subset of the data using a differential weighting scheme (Matthews and Yang, 2012).

In the literature, the GWR technique is applied to research in a wide range of academic fields. Examples of research topics include but not limited to the analysis of health and disease (Chen et al., 2010; Yang et al., 2009; Goovaerts, 2005; Nakaya et al., 2005), environmental equity (Mennis and Jordan, 2005), housing markets (Fotheringham et al., 2002), population density and housing (Mennis, 2006), poverty (Partridge and Rickman, 2005; Longley and Tobon, 2004), traffic models (Zhao and Park, 2004), as well as environmental conditions (Foody, 2003).

Formally, the GWR model is defined as:

$$y_i = \beta_0 (u_i, v_i) + \sum_{j=1}^{k} \beta_j(u_i, v_i) x_{ij} + \varepsilon_i \qquad \text{Formula 3.4}$$

where y_i is the value of the outcome variable at a particular location, the *ith* location; where (u_i, v_i) denote the coordinates, β_o and β_j represent the local estimated intercept and effect of variable *j* for location *i*, respectively. The locations near *i* have a stronger influence in the estimation of β_j (u_i, v_i) than locations farther from *i*. Therefore, using this model, localized parameter estimates can be generated for any location *i*, which allows users to create a map showing a continuous surface of parameter values and to examine the spatial variability of these parameters (Matthews and Yang, 2012).

As an exploratory technique, GWR provides outcomes that are certainly powerful and offers significant advantages over simple linear regression procedures. It provides users a set of location specific parameter estimates not only to map and visualize but also to interpret regarding the relationships between predictors and the outcome variable (Bitter et al., 2007; Wheeler and Tiefelsdorf, 2005).

The global and local ESDA (Moran's I and LISA) and GWR techniques described above in detail are applied to this study. The additional methodological steps taken and the results of the spatial analyses are discussed in a greater detail in Chapter 4 and Chapter 5 of this dissertation. In this dissertation, the explanations of the analysis results are written based on the US Census regions and division definitions as shown in Figure 3.2. The West (combining the Pacific and Mountain divisions) includes Arizona, California, Colorado, Idaho, Montana, Nevada, Oregon, New Mexico, Utah, Washington, and Wyoming; the Midwest (combining the West North Central and East North Central divisions), Kansas, Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin; the Northeast (combining the New England and Middle Atlantic regions), Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and

Vermont; and the South (combining the West South Central, East South Central, and South Atlantic divisions), Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia.

Figure 3.2 Boundaries of the US regions and divisions based on the US Census

CHAPTER 4. A SPATIO-TEMPORAL ANALYSIS OF OBESITY AND CORRELATES OF PHYSICAL INACTIVITY, POVERTY, AND BUILT ENVIRONMENT IN THE US: A COUNTY-LEVEL ANALYSIS

4.1. Introduction

The obesity issue has captured the public's attention especially in recent years due to the increasing prevalence of obesity reported worldwide yet it is particularly substantial in the US. A study published very recently shows that the obesity rates vary worldwide and it is estimated that there are 671 million obese individuals around the world as of 2013 (The GBD 2013 Obesity Collaboration et al., 2014). More than half of these obese individuals live in the US, China, India, Russia, Brazil, Mexico, Egypt, Germany, Pakistan, and Indonesia. Among these ten countries, the US alone accounts for more than 13% of the obese individuals while China and India account for 15% combined (The GBD 2013 Obesity Collaboration et al., 2014). Despite the debates over the actual risks attributable to obesity and the factors leading to high obesity prevalence, the epidemic itself remains as a major public health concern in the US (CDC, 2014).

Obesity has strong genetic determinants but the genetic composition of the population does not change rapidly. Thus, the major changes in non-genetic factors such as the development of obesogenic environments and shifting socioeconomic status and lifestyles of individuals are hypothesized to increase the risk of obesity (Hill and Trowbridge, 1998). The obesogenicity of an environment has been defined as "the sum of influences that the surroundings, opportunities, or conditions of life have on promoting obesity in individuals or populations" (Lake and Townsend, 2006).

The etiology of obesity is multi-causal in nature and the causes leading to high obesity prevalence are mainly related to demographic factors, socioeconomic status, and lifestyle choices as well as "spatial" factors that determine the physical and social environments in which people live. The physical environment includes all individuals' surroundings either designed or natural. The built environment is a part of the physical environment and it consists of places built or designed by humans including buildings, grounds around buildings, layout of communities, transportation infrastructure, and park and trails (Sallis et al., 2012).

Consequently, the variables scrutinized in this study are chosen carefully to depict the different aspects of such influences. The poverty and physical inactivity (PIA) variables were chosen to represent the socioeconomic status and lifestyle choices while the ratio of jobs to employed residents (JER), population-weighted distance (PWD) to parks, and population density variables were selected to characterize the built environment factors in relation to obesity for all counties and county equivalents in the continental US.

The main objectives of this study are: 1) To explore and identify spatio-temporal patterns and trends of obesity and their associations with respect to the aforementioned variables; 2) To examine whether these associations are based on clusters of occurrences or they are dispersed randomly across space. Furthermore, the main hypothesis aims to test whether a spatial component is involved in these temporal observations such as a significant spatial convergence of high obesity/PIA, poverty, JER, PWD to parks, and population density rates and/or low obesity/PIA, poverty, JER, PWD to parks, and population density rates.

The relationships between obesity and each of the aforementioned variables were explored separately in both univariate and bivariate forms in three waves (2004, 2007, and 2010) to quantify and to gain a better understanding of their spatio-temporal interactions, patterns, and trends. Exploratory spatial data analysis (ESDA) techniques available in GeoDa, in particular, univariate and bivariate Moran's I and Local Indicators of Spatial Association (LISA) statistics were employed. The results of univariate Moran's I are used to determine whether statistically significant clusters of obesity exist. If obesity clusters exist, where they are spatially observed is mapped and documented using univariate LISA test results. Likewise, the results of univariate spatial analyses for the other variables (e.g., PIA, poverty etc.) are also documented and mapped in the same way. To explore the clustering patterns between obesity and each of the other variables further, two at a time, bivariate spatial analyses are conducted (e.g., between obesity and lagged poverty or between obesity and lagged PIA, and so on). The results of bivariate Moran's I are used to determine whether a general clustering of high or low attribute values exists for the pairs of variables analyzed. If clustering exists, where and how they spatially converge are also mapped and documented by utilizing bivariate LISA test results.

Therefore, the outcomes and insights of this study will be useful to identify and select the locations to further study and explicate the underlying processes involved in forming such spatial patterns (clusters and/or spatial outliers).

4.2. Data and Methods

4.2.1 Data

The datasets used in this study come from various sources. The original datasets are at the county level for the entire US mainly for three years, including 2004, 2007, and 2010. However, due to the spatial nature of this study, only the contiguous counties and county equivalents within the continental US are included in the analyses. In other words, the study area consists of the 3,105 counties and county-equivalents across the 48 contiguous states, also known as the continental US (all except Alaska and Hawaii).

As discussed in detail in Chapter 3 of this dissertation, there are mainly six types of datasets used in this study. These datasets include: 1) obesity; 2) PIA; 3) poverty; 4) JER; 5) PWD to parks; and 6) population density.

4.2.2 Methods

ESDA consists of a set of techniques that are used to describe and visualize spatial distributions, identify atypical locations or spatial outliers, and discover patterns of spatial associations and clusters. The process of ESDA starts with simple mapping and geovisualization, moves on to exploration, and ends with spatial autocorrelation analysis (Anselin, 2005).

In this study, as an initial part of an ESDA process, choropleth maps were constructed, each using a specific type of data classification algorithm, to gain a better understanding of the data by describing the overall distribution and visually interpreting spatial patterns.

Choropleth mapping is a crucial step to visualize patterns and obtain a general picture of the spatial distribution of the phenomenon under examination across the defined geographic boundaries. However, when interpreting choropleth maps caution is needed, since attention is often focused on either the relatively larger areas that tend to dominate maps or the areas that misleadingly stand out due to the improper use of classification methods. Without statistical support, the interpretations can be subjective and/or simply inaccurate.

An alternative and complementary technique for exploring the spatial structure of a dataset is to utilize a statistical measure that accounts for the spatial locations of each data observation in conjunction with the observed data value at each location. One family of such measures was developed to assess and quantify spatial autocorrelation (Holt, 2007). The concept of spatial autocorrelation is based on the idea that observations that are located near each other are influenced by each other and thus not distributed in space by random chance alone.

The notion of spatial autocorrelation can be defined as the coincidence of value similarity with locational similarity (Anselin and Bera, 1998). In other words, if locations with similar attribute values (high or low) are closely distributed in space, the spatial pattern is considered to exhibit positive spatial autocorrelation. Conversely, if closely associated locations have dissimilar attribute values, the spatial pattern is considered to show negative spatial autocorrelation. If no spatial relationship exists between attribute values and their locations, the spatial pattern is considered to exhibit zero spatial autocorrelation. According to Anselin and Bera (1998), of the two types of spatial autocorrelation, positive autocorrelation is by far the more intuitive.

The analyses for this study were performed in four steps: 1) Acquisition and manipulation of the data to be used in ArcGIS 10.3 and GeoDa 1.6.7; 2) Generation of descriptive statistics including histograms to see overall distribution of the data; 3) Creation of choropleth maps in ArcGIS 10.3 to visually interpret the spatial distribution of the data; and 4) Implementation of ESDA techniques (global Moran's I and LISA) in GeoDa to determine if there was significant spatial clustering of variables at the county level.

First of all, the datasets for all the variables used in this study were downloaded from various sources as explained above in the data section and imported into ArcGIS 10.3, where the data were joined to a geographic boundary file, also known as a shapefile, for 3,105 counties and county equivalents across the 48 contiguous states (all except Alaska and Hawaii), also known as the continental Unites States. The data were joined using the common field of counties' five-digit Federal Information Processing Standards (FIPS) codes. The final merged shapefile was projected using the Albers equal-area conic projection.

The created shapefile was then imported into GeoDa to perform an initial exploration of the overall data distribution through the descriptive statistics, boxplots, and histograms. The class intervals are calculated by dividing the range (i.e., highest value-lowest value) by the number of classes (determined to be 25 classes for n=3,105) for each variable in each data period. Even though GeoDa includes mapping functionality, it is somewhat outdated in terms of graphic representation and user friendliness when compared to ArcGIS 10.3. Therefore, ArcGIS 10.3 was utilized for the creation of the choropleth maps for all variables. The county-level data for the variables were initially tested and mapped using various data classification methods to determine data cut points. The data classification based on the standard deviation and natural breaks (Jenks' optimal algorithm) methods gave better results in terms of the spatial representation of the data. However, the finalized versions of the choropleth maps were derived using the standard deviation method to emphasize the distribution of the data in reference to the means of the datasets.

In the final step, GeoDa was used to perform the advanced spatial analyses and illustrate the results in the form of Moran scatter plots and cluster maps. Initially, a spatial weights matrix was generated based on contiguity using the US county boundaries in the joined shapefile. Contiguity-based spatial weights are chosen because the main interest lies in understanding the spatial interdependence at the areal level. The spatial weights matrix was created using the first order "rook's contiguity" which uses only common boundaries to define neighbors. The global Moran's I statistics were calculated in a traditional univariate form as well as in a bivariate form for all the variables for the years 2004, 2007, and 2010. Following that, the LISA tests were conducted in univariate and bivariate forms for each year to see exactly where the spatial clusters and/or outliers are located and how pairs of variables are spatially associated in the bivariate form. The number of permutations chosen to construct the reference distribution was 999 for both the global Moran's I and LISA tests. The corresponding Moran scatter plots based on the global Moran' I statistics and the cluster/significance maps based on the LISA statistics are presented in the results section. The presented cluster maps are significant at the $p < 0.05$ level.

All aspatial and spatial exploratory data analyses, particularly the tests for spatial autocorrelation, address various questions about a dataset including: 1) How are the features distributed? 2) How does the pattern created by the features appear? 3) Where are the clusters? 4) How do patterns and clusters of different variables compare to one another? When patterns are explored both statistically and spatially as in global Moran's I and LISA tests, this leads to a better understanding of the distribution of geographic phenomena, especially in terms of patterns and clusters.

4.3. Results

4.3.1 Descriptive Statistics and Basic Mapping of Variables

The following findings document the overall data distribution and spatial patterns through descriptive statistics, histograms, and choropleth maps for all county-level variables for the three waves (2004, 2007, and 2010) explored in this research. These basic maps visually support the descriptive statistics and mainly depict how the county specific rates vary from one county to another. The spatial patterns of the county-level variables for the entire US start to emerge visually by examining these maps.

4.3.1.1 Obesity

Among the counties, the obesity rates in 2004, 2007, and 2010 range from 12.30% (Boulder, Colorado) to 38.00% (Jefferson, Mississippi), from 12.50% (Boulder, Colorado) to 43.50% (Greene, Alabama), and from 13.10% (Teton, Wyoming) to 47.90% (Greene, Alabama), respectively. As expected, in all years, the highest obesity rates are observed in the South. Greene County in Alabama remains unchanged, having the highest obesity rate in both 2007 and

2010, while Boulder County in Colorado remains unchanged having the lowest obesity rate in both 2004 and 2007. The mean obesity rates and the standard deviations are 25.29% (3.24), 28.29% (3.61), and 30.58% (4.23) in 2004, 2007, and 2010, respectively, indicating that the mean obesity rate has increased continuously from 25.29% (2004) to 30.58% (2010). The median obesity rates are 25.40, 28.50, and 30.70 over this time period (Table 4.1).

Table 4.1 Descriptive statistics for county level obesity prevalence rates in the US for 2004, 2007, and 2010

	2004	2007	2010
n	3,105	3,105	3,105
Mean	25.29	28.29	30.58
Std. Deviation	3.24	3.61	4.23
Median	25.40	28.50	30.70
Minimum	12.30	12.50	13.10
Maximum	38.00	43.50	47.90

The histograms (Figures 4.1, 4.2, and 4.3) below show the distribution of the county-level obesity prevalence rates for the years 2004, 2007, and 2010. In all years, the histograms show approximately a symmetric (normal) distribution of the obesity rates (the skewness values are -0.37, -0.47, and -0.33 respectively with a standard error of 0.04).

In 2004, 56.65% of all counties have an obesity prevalence of 25% or more and 0.68% of all counties have an obesity prevalence of 35% or more (Figure 4.1). In 2007, the percentages of the counties that have obesity prevalence above 25% and above 35% are 87.15% and 3%, respectively (Figure 4.2). In 2010, 90.85% of all counties have an obesity prevalence exceeding 25%, 12.59% have an obesity prevalence exceeding 35%, and 1.52% have an obesity prevalence exceeding 40% (Figure 4.3). By 2010, it is already observed that 2 counties have obesity prevalence higher than 45%. As presented by these findings, it is clear that the magnitude of obesity prevalence rates has gradually increased at the county level for the period 2004, 2007, and 2010.

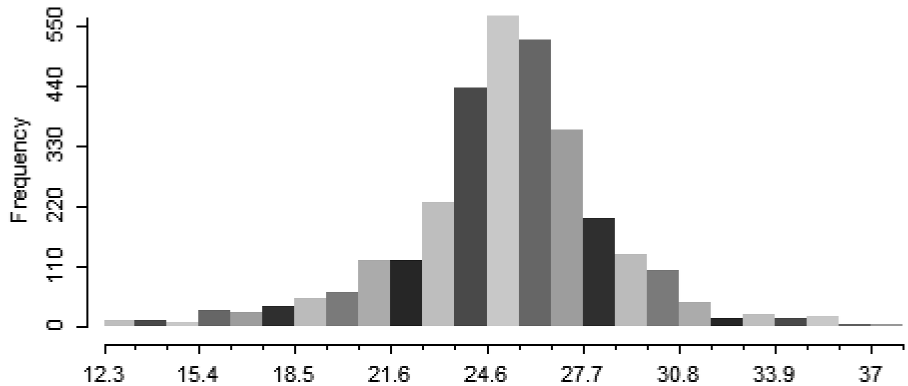

Figure 4.1 Histogram of county level obesity prevalence rates in the US in 2004

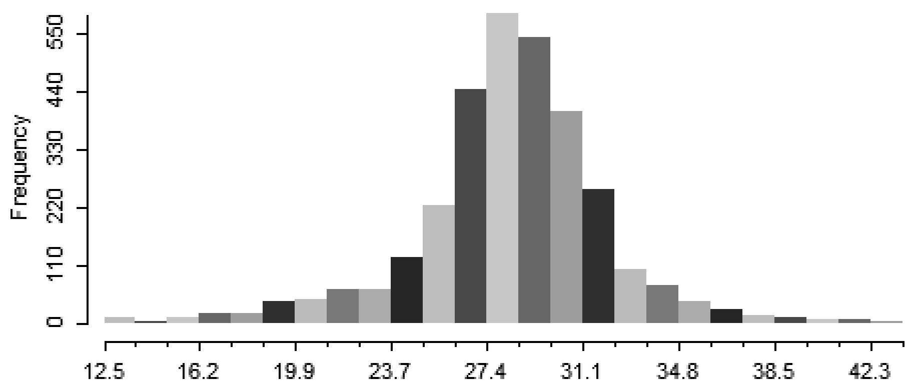

Figure 4.2 Histogram of county level obesity prevalence rates in the US in 2007

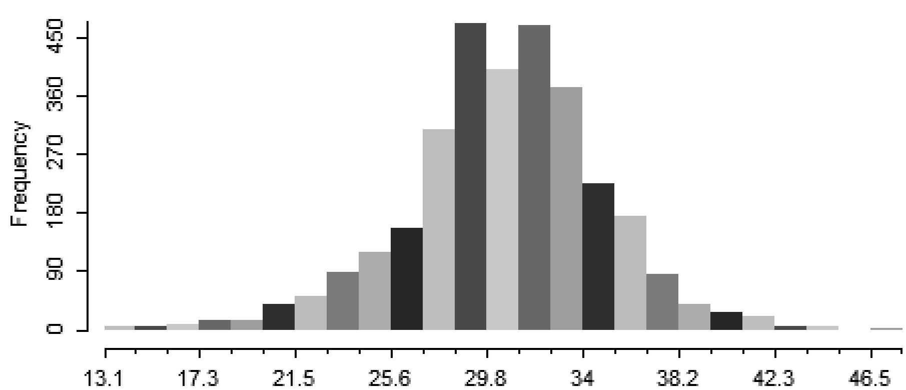

Figure 4.3 Histogram of county level obesity prevalence rates in the US in 2010

Figures 4.4, 4.5, and 4.6 illustrate the county-level obesity prevalence rates for the years 2004, 2007, and 2010 using the standard deviation classification method. Overall, high obesity rates are seen in the Southern US while a large concentration of the low obesity rates are observed in the Western and the Northeastern US.

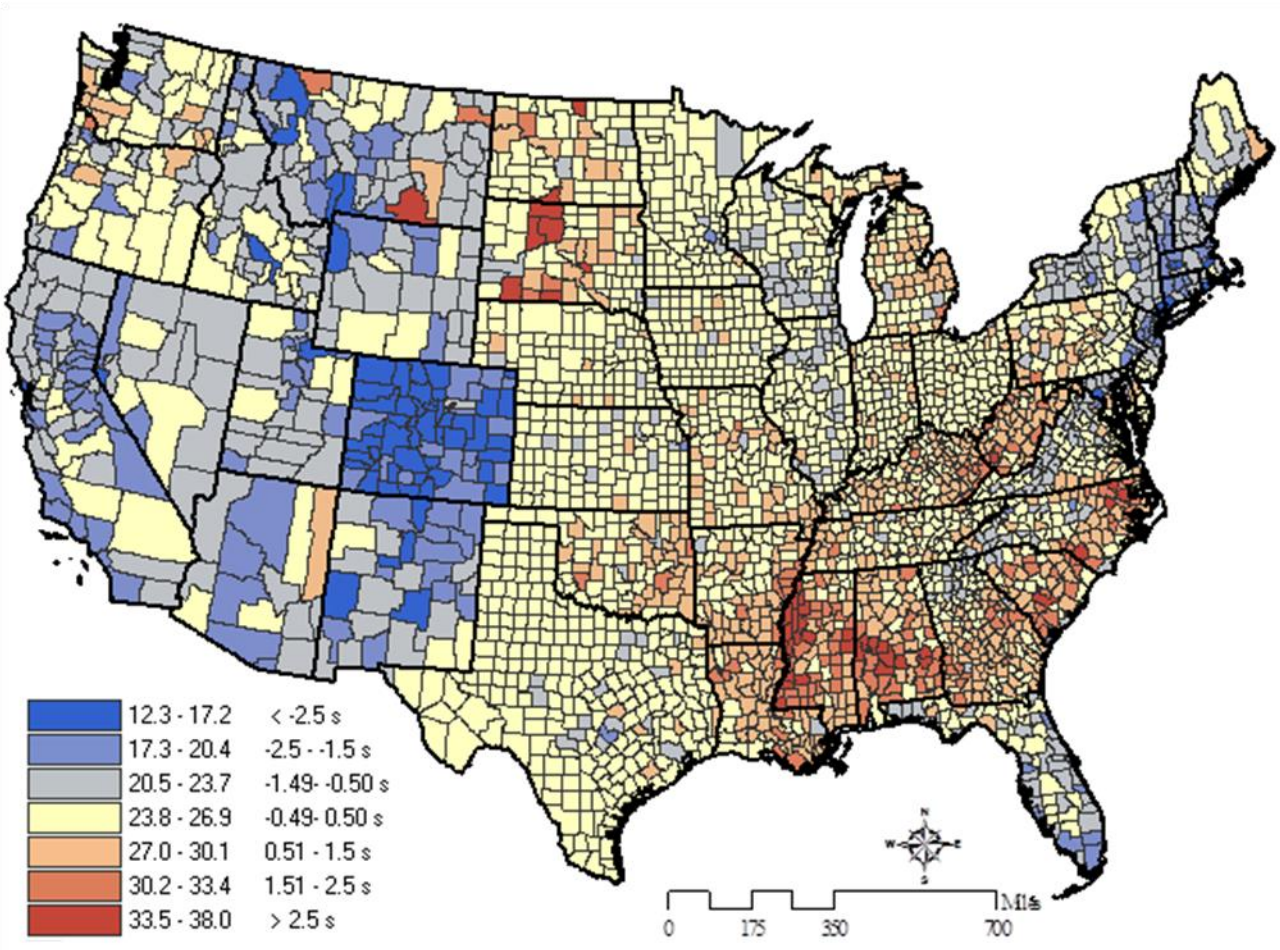

Figure 4.4 Geographic distribution of obesity prevalence rates at the county level in the US for 2004

In 2004, of all counties, 4.22% have obesity rates below -2 standard deviations (SD)s of the mean and 12.98% have obesity rates below -1 SD of the mean, while 11.63% have obesity rates above +1 SD of the mean and 2.13% have obesity rates above +2 SDs of the mean (Figure

4.4). In 2007, of all counties, 4.38% have obesity rates below -2 SDs of the mean and 12.11% have obesity rates below -1 SD of the mean, whereas 10.85% have obesity rates above +1 SD of the mean and 2.35% have obesity rates above +2 SDs of the mean (Figure 4.5). In 2010, of all counties, 3.57% have obesity rates below -2 SDs of the mean and 13.59% have obesity rates below -1 SD of the mean while 13.27% have obesity rates above +1 SD of the mean and 2.03% have obesity rates above +2 SDs of the mean (Figure 4.6).

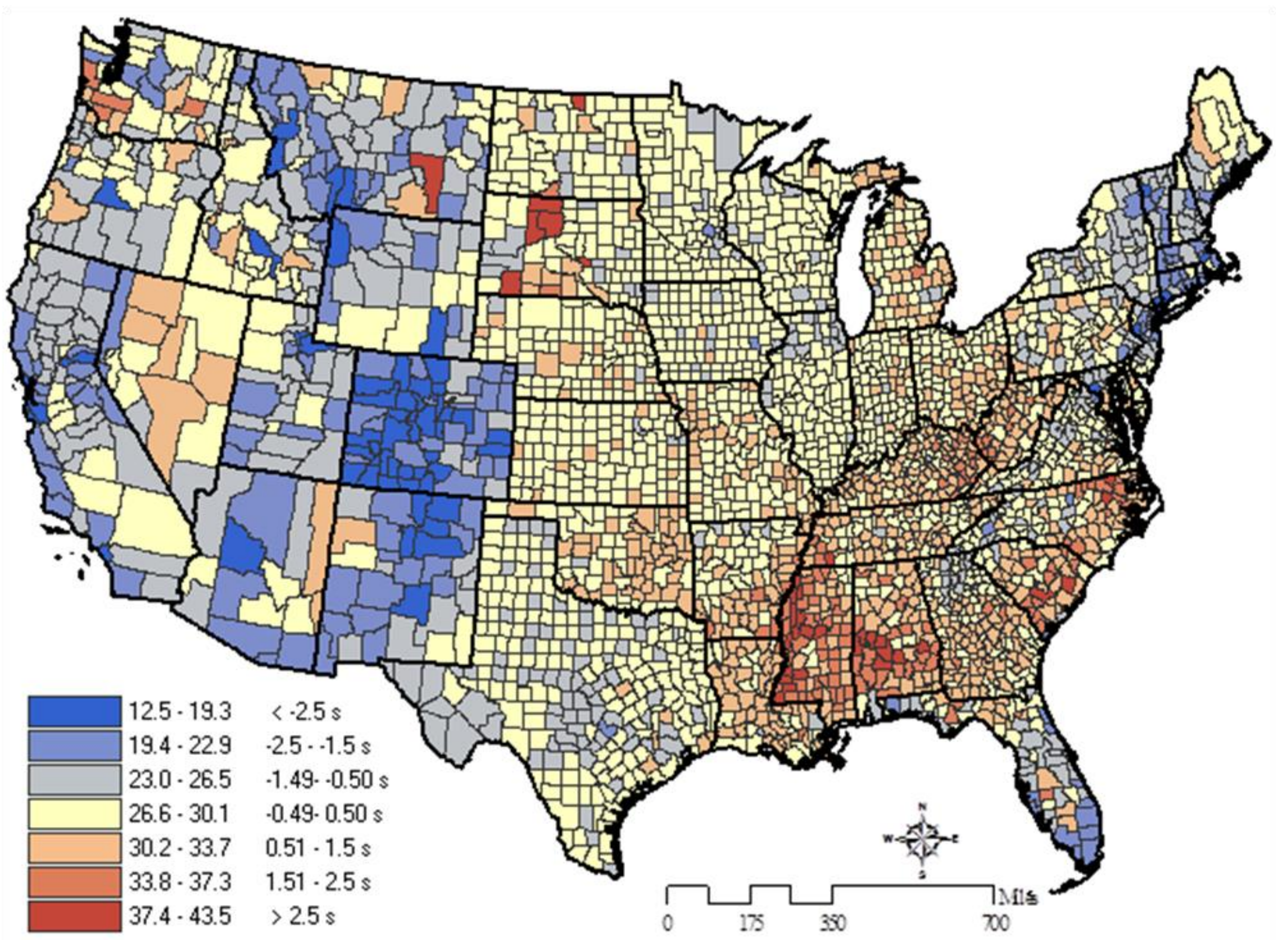

Figure 4.5 Geographic distribution of obesity prevalence rates at the county level in the US for 2007

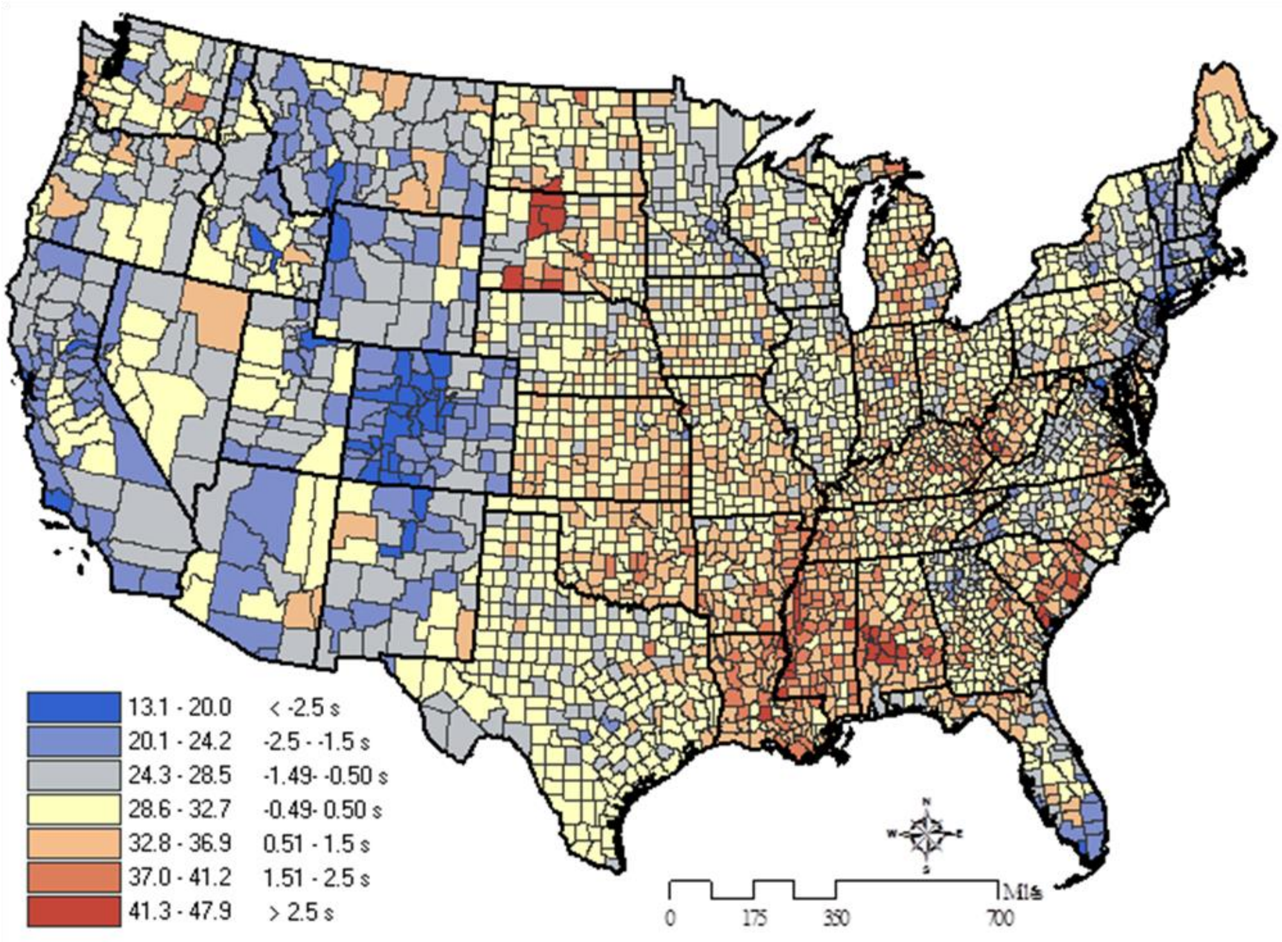

Figure 4.6 Geographic distribution of obesity prevalence rates at the county level in the US for 2010

4.3.1.2 Physical Inactivity (PIA)

The physical inactivity (PIA) rates for the 3,105 US counties and county-equivalents range from 9.2% (Boulder, Colorado) to 42.4% (Mingo, West Virginia) in 2004, from 9.8% (Boulder, Colorado) to 44.4% (Pike, Kentucky) in 2007, and from 10.4% (Routt, Colorado) to 44.9% (Wyoming, West Virginia) in 2010. In all years, the highest PIA rates are seen in the South, while the lowest rates are observed in the West. The state of West Virginia has the highest PIA rates in both 2004 and 2010 and the state of Colorado has the lowest PIA rates in all three years. In addition, Boulder County in Colorado stands out by having the lowest PIA rate both in 2004 and 2007. The mean PIA rates and the standard deviations are 25.98% (5.23) in

2004, 26.74% (5.09) in 2007, and 27.90% (5.28) in 2010. The mean PIA rate has increased gradually from 25.98% in 2004 to 26.74% in 2007 and to 27.90% in 2010. The medians are 25.80, 26.80, and 28.10 in 2004, 2007, and 2010, respectively (Table 4.2).

Table 4.2 Descriptive statistics for county level physical inactivity (PIA) prevalence rates in the US for 2004, 2007, and 2010

	2004	2007	2010
n	3,105	3,105	3,105
Mean	25.98	26.74	27.90
Std. Deviation	5.23	5.09	5.28
Median	25.80	26.80	28.10
Minimum	9.20	9.80	10.40
Maximum	42.40	44.40	44.90

The histograms below show approximately a symmetric (normal) distribution of the county-level PIA prevalence rates in all years (the skewness values are 0.11, -0.07, and -0.23, respectively with a standard error of 0.04) (Figures 4.7, 4.8, and 4.9).

The observed trends indicate a gradual increase in PIA prevalence rates throughout the years 2004, 2007, and 2010. In 2004, 56.65% of all counties have a PIA prevalence of 25% or more and 0.68% of all counties have a PIA prevalence of 35% or more (Figure 4.7). In 2007, the percentages of the counties that have a PIA prevalence of 25% or more and 35% or more are 87.15% and 3%, respectively (Figure 4.8). In 2010, 90.85% of all counties have a PIA prevalence above 25% and 12.59% of all counties have a PIA prevalence above 35% (Figure 4.9).

The following maps (Figures 4.10, 4.11, and 4.12) show the county-level PIA rates for the years 2004, 2007, and 2010 using the standard deviation classification method. Overall, high PIA rates are observed in the Southern US, while a large concentration of low PIA rates are seen in the Western and the Northeastern US. These PIA rate patterns show a high similarity to those observed in the obesity rate maps.

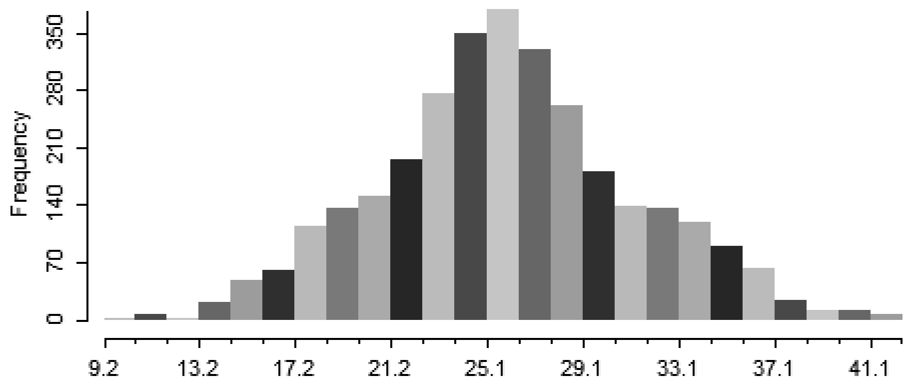

Figure 4.7 Histogram of county level physical inactivity (PIA) prevalence rates in the US in 2004

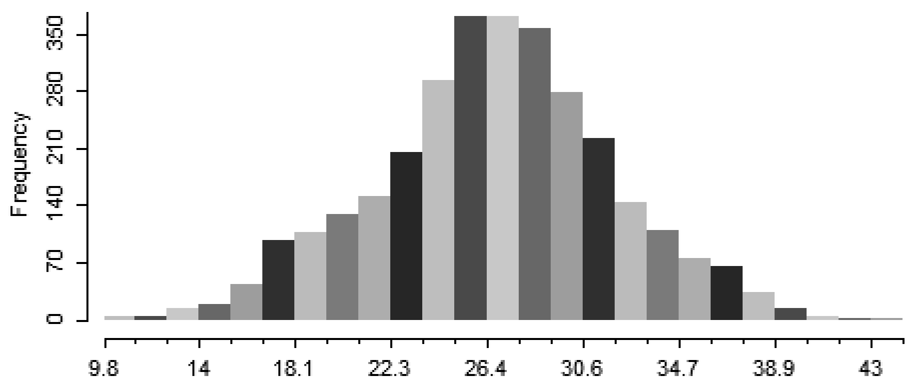

Figure 4.8 Histogram of county level physical inactivity (PIA) prevalence rates in the US in 2007

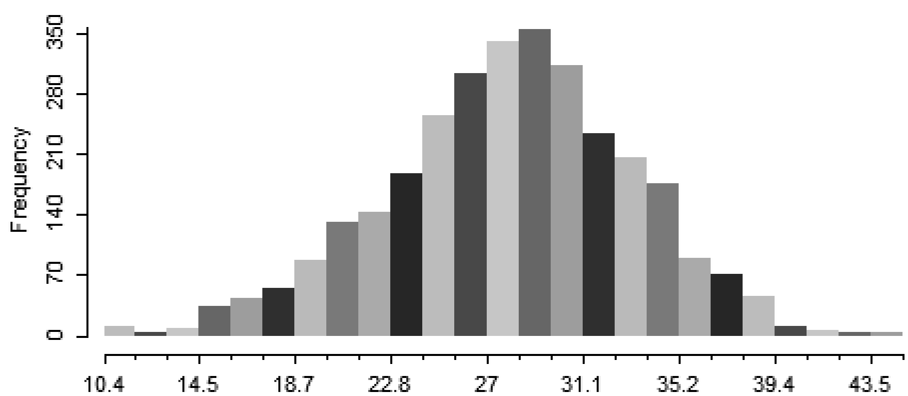

Figure 4.9 Histogram of county level physical inactivity (PIA) prevalence rates in the US in 2010

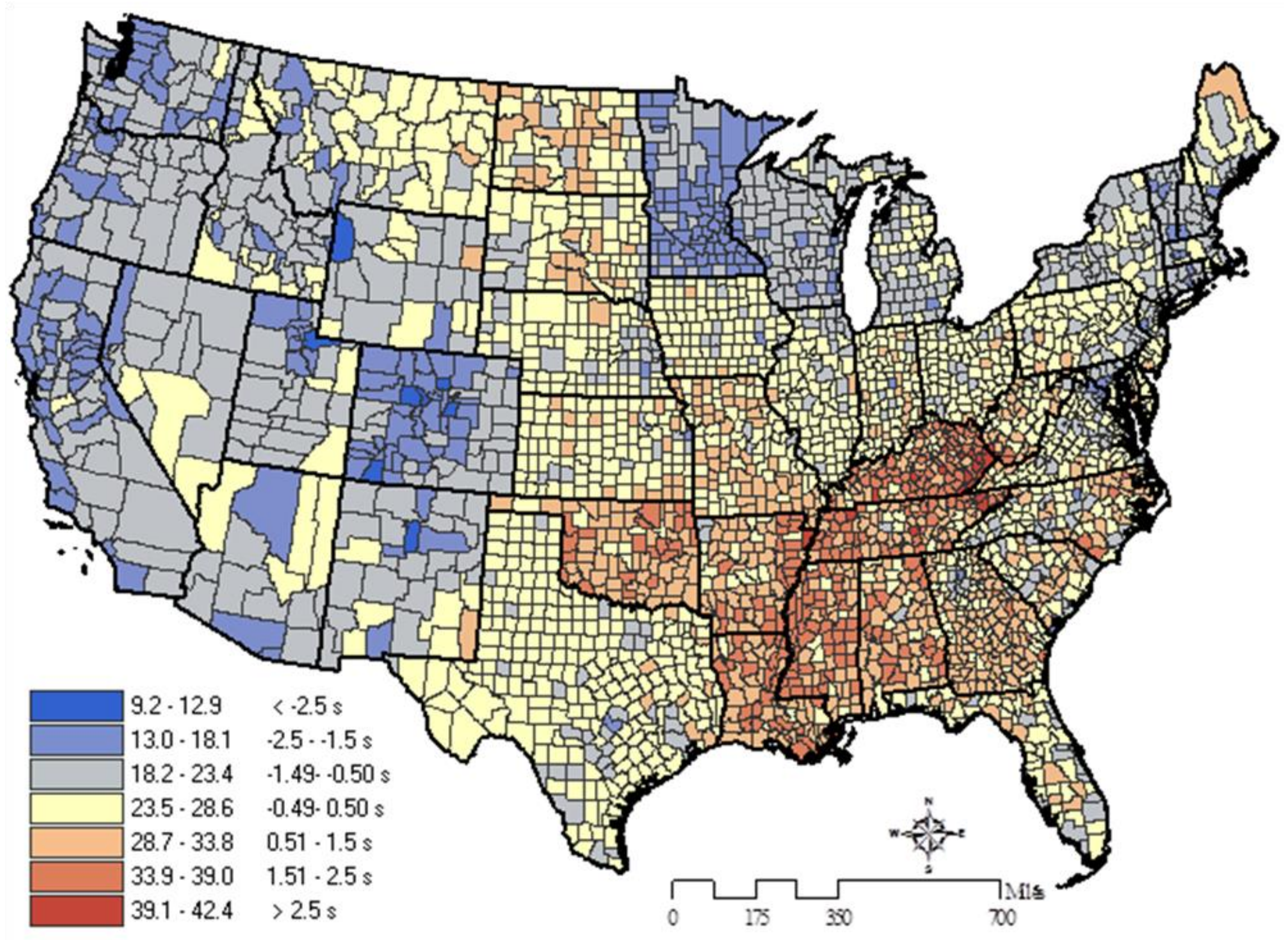

Figure 4.10 Geographic distribution of physical inactivity (PIA) prevalence rates at the county level in the US for 2004

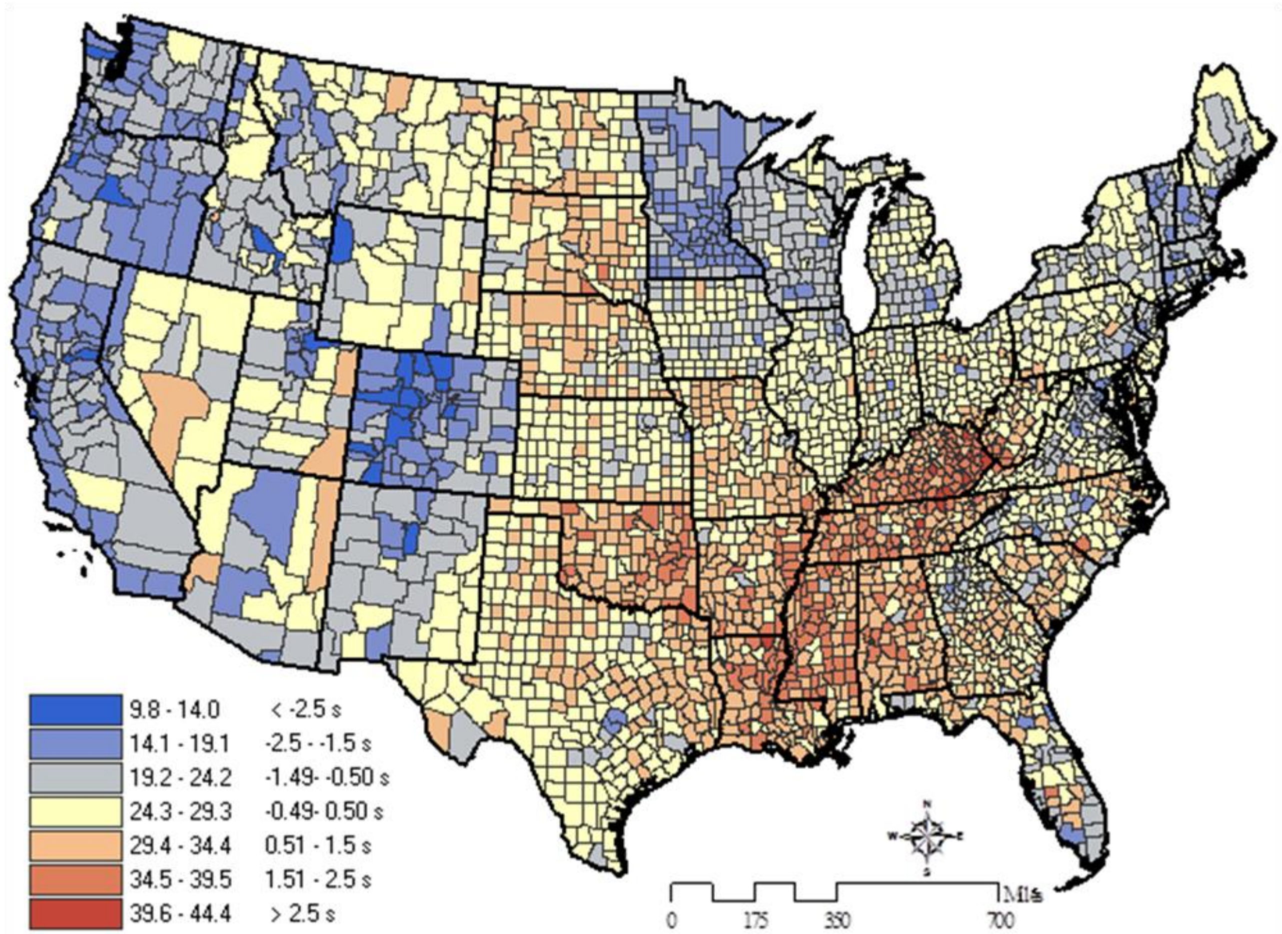

Figure 4.11 Geographic distribution of physical inactivity (PIA) prevalence rates at the county level in the US for 2007

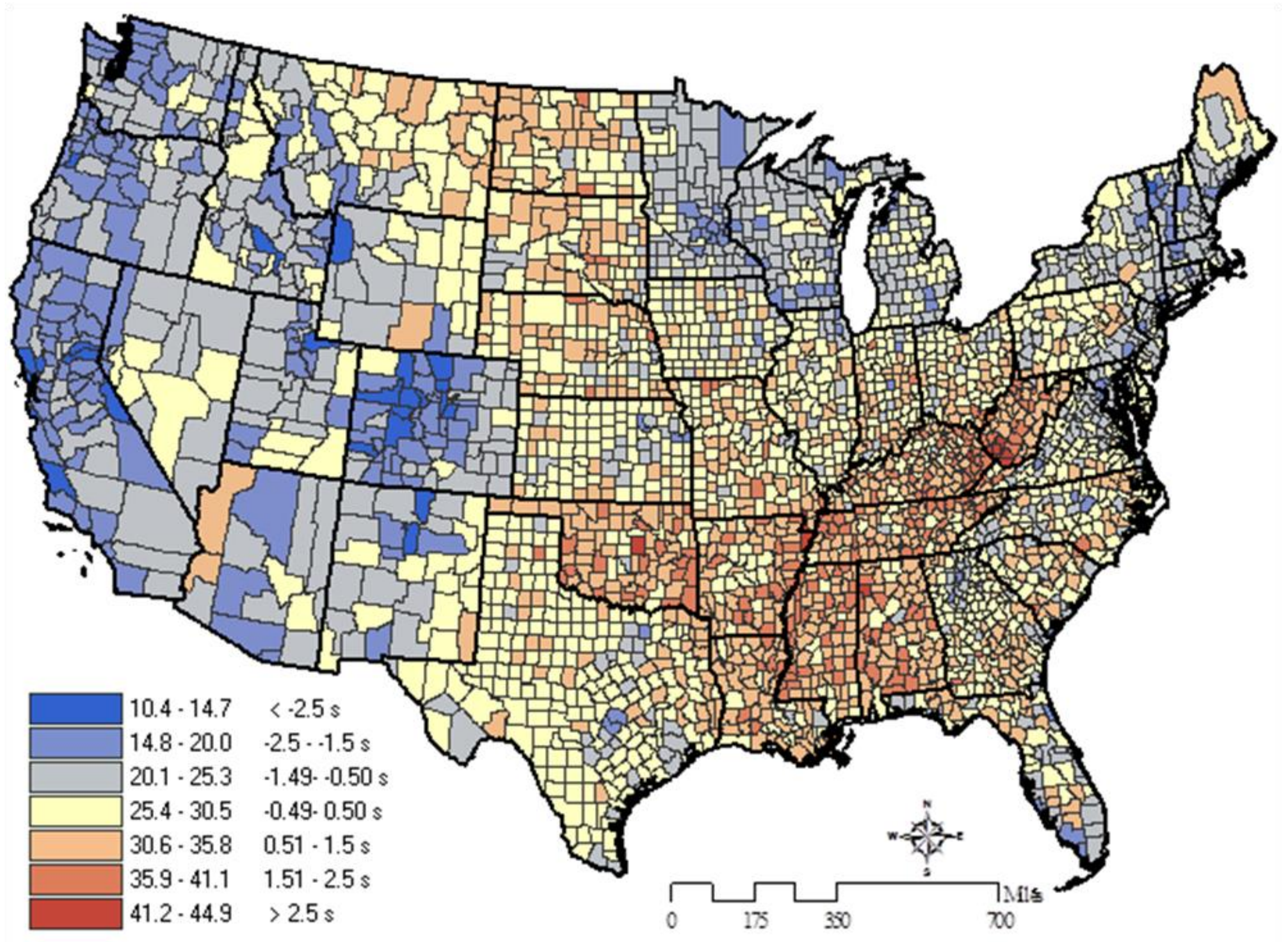

Figure 4.12 Geographic distribution of physical inactivity (PIA) prevalence rates at the county level in the US for 2010

In 2004, of all counties, 2.09% have PIA rates below -2 SDs of the mean and 16.20% have PIA rates below -1 SD of the mean, while 16.30% have PIA rates +1 SD and higher than the mean and 2.35% have PIA rates +2 SDs and higher than the mean (Figure 4.10). In 2007, of all counties, 2.54% have PIA rates below -2 SDs of the mean and 16.23% have PIA rates below -1 SD of the mean, whereas 14.91% have PIA rates +1 SD and higher than the mean and 2.38% have PIA rates +2 SDs and higher than the mean (Figure 4.11). In 2010, of all counties, 3.22% have PIA rates below -2 SDs of the mean and 15.88% have PIA rates below -1 SD of the mean, while 15.39% have PIA rates +1 SD and higher than the mean and 1.29% have PIA rates +2 SDs and higher than the mean (Figure 4.12).

4.3.1.3 Poverty

Among the counties for the years 2004, 2007, and 2010, the poverty rates range from 2.6% (Falls Church, Virginia) to 39.4% (Ziebach, South Dakota) in 2004, from 2.4% (Douglas, Colorado) to 55.9% (Ziebach, South Dakota) in 2007, and from 3.1% (Falls Church, Virginia) to 50.1% (Ziebach, South Dakota) in 2010 (Table 4.3). In all years, the highest poverty rates are observed in Ziebach County in South Dakota. Falls Church County in Virginia stands out having the lowest poverty rate both in 2004 and 2010. The mean poverty rates and the standard deviations are 13.78% (5.14) in 2004, 15.13% (6.24) in 2007, and 16.80% (6.23) in 2010. The medians are 13.00, 14.10, and 16.00 in 2004, 2007, and 2010, respectively.

Table 4.3 Descriptive statistics for county level poverty prevalence rates in the US for 2004, 2007, and 2010

	2004	2007	2010
n	3,105	3,105	3,105
Mean	13.78	15.13	16.80
Std. Deviation	5.14	6.24	6.23
Median	13.00	14.10	16.00
Minimum	2.60	2.40	3.10
Maximum	39.40	55.90	50.10

The histograms below show a positively skewed distribution of the county-level poverty rates in all years (the skewness values are 0.97, 1.22, and 0.98 respectively with a standard error of 0.04) (Figures 4.13, 4.14, and 4.15).

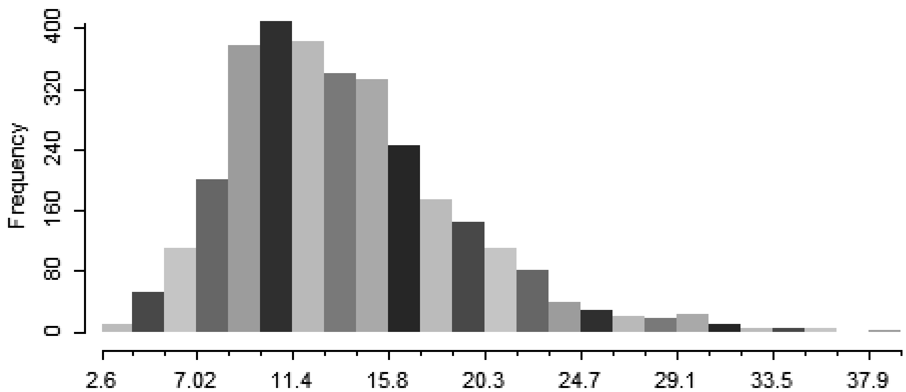

Figure 4.13 Histogram of county level poverty prevalence rates in the US in 2004

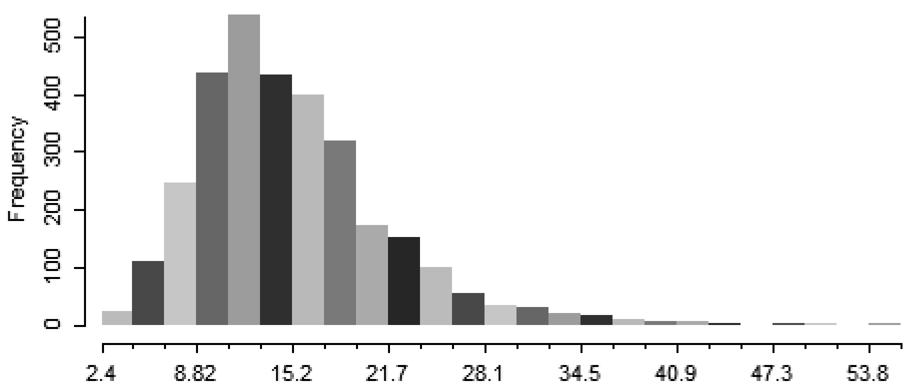

Figure 4.14 Histogram of county level poverty prevalence rates in the US in 2007

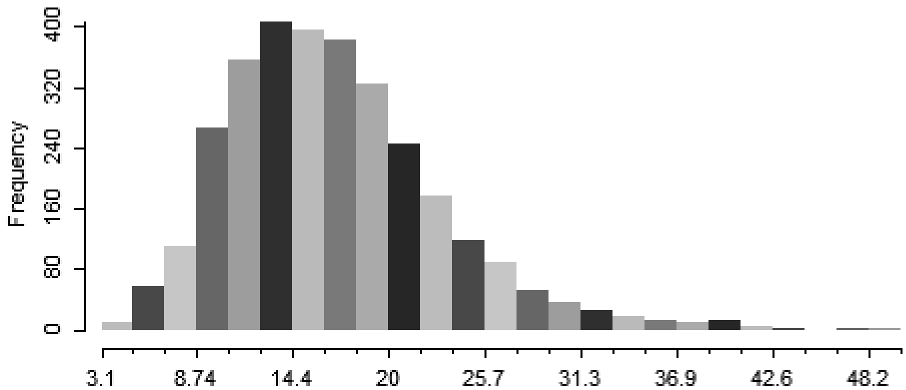

Figure 4.15 Histogram of county level poverty prevalence rates in the US in 2010

The trends throughout the years 2004, 2007, and 2010 indicate a slight increase in poverty prevalence rates. The percentages of the counties that have a poverty prevalence of 25% or more are 3.22% in 2004, 6.73% in 2007, and 9.47% in 2010 (Figures 4.13, 4.14, and 4.15). Overall, the percentages of the counties that have a poverty prevalence of 35% or more are less than 1% in 2004 and slightly more than 1% in 2007 and 2010.

Figures 4.16, 4.17, and 4.18 illustrate the county-level poverty rates for the years 2004, 2007, and 2010 using the standard deviation classification method. As seen in the maps there is a distinctive separation between the northern and southern counties and county equivalents. It is almost as if an imaginary horizontal line divides the US in the middle and this only applies to the eastern half of the US. Overall, the counties in the southern half of this divide are dominated by higher poverty rates, while the counties in the northern half are dominated by lower poverty rates. In 2004, of all counties, 0.13% have poverty rates below -2 SDs of the mean and 12.98% have poverty rates below -1 SD of the mean, whereas 14.56% have poverty rates +1 SD and higher than the mean and 3.74% have poverty rates +2 SDs and higher than the mean (Figure 4.16). In 2007, of all counties, 0.04% have poverty rates below -2 SDs of the mean and 12.79% have poverty rates below -1 SD of the mean, while 14.14% have poverty rates +1 SD and higher than the mean and 4.06% have poverty rates +2 SDs and higher than the mean (Figure 4.17). In 2010, of all counties, 0.16% have poverty rates below -2 SDs of the mean and 14.27% have poverty rates below -1 SD of the mean, whereas 14.17% have poverty rates +1 SD and higher than the mean and 3.96% have poverty rates +2 SDs and higher than the mean (Figure 4.18).

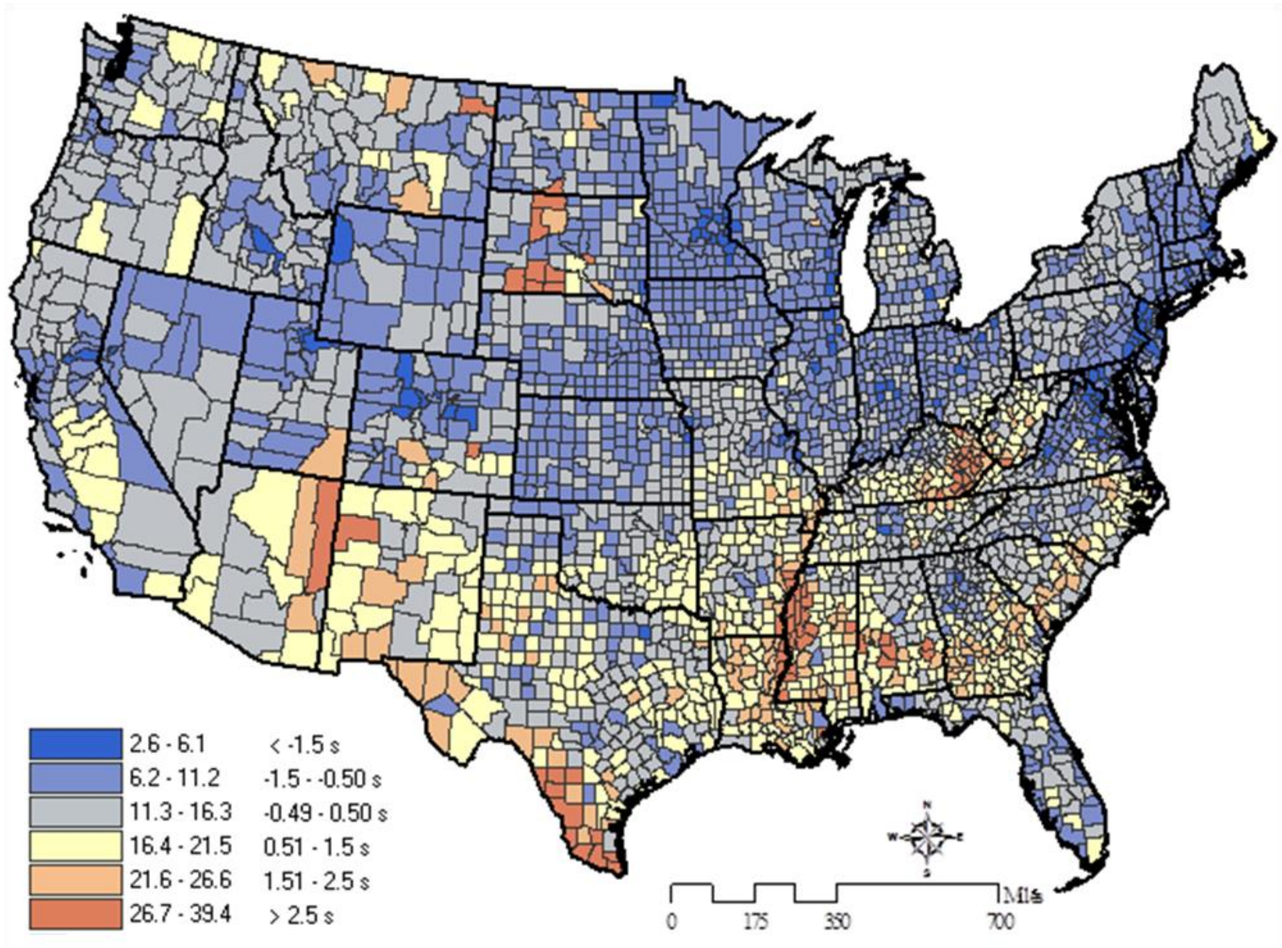

Figure 4.16 Geographic distribution of poverty prevalence rates at the county level in the US for 2004

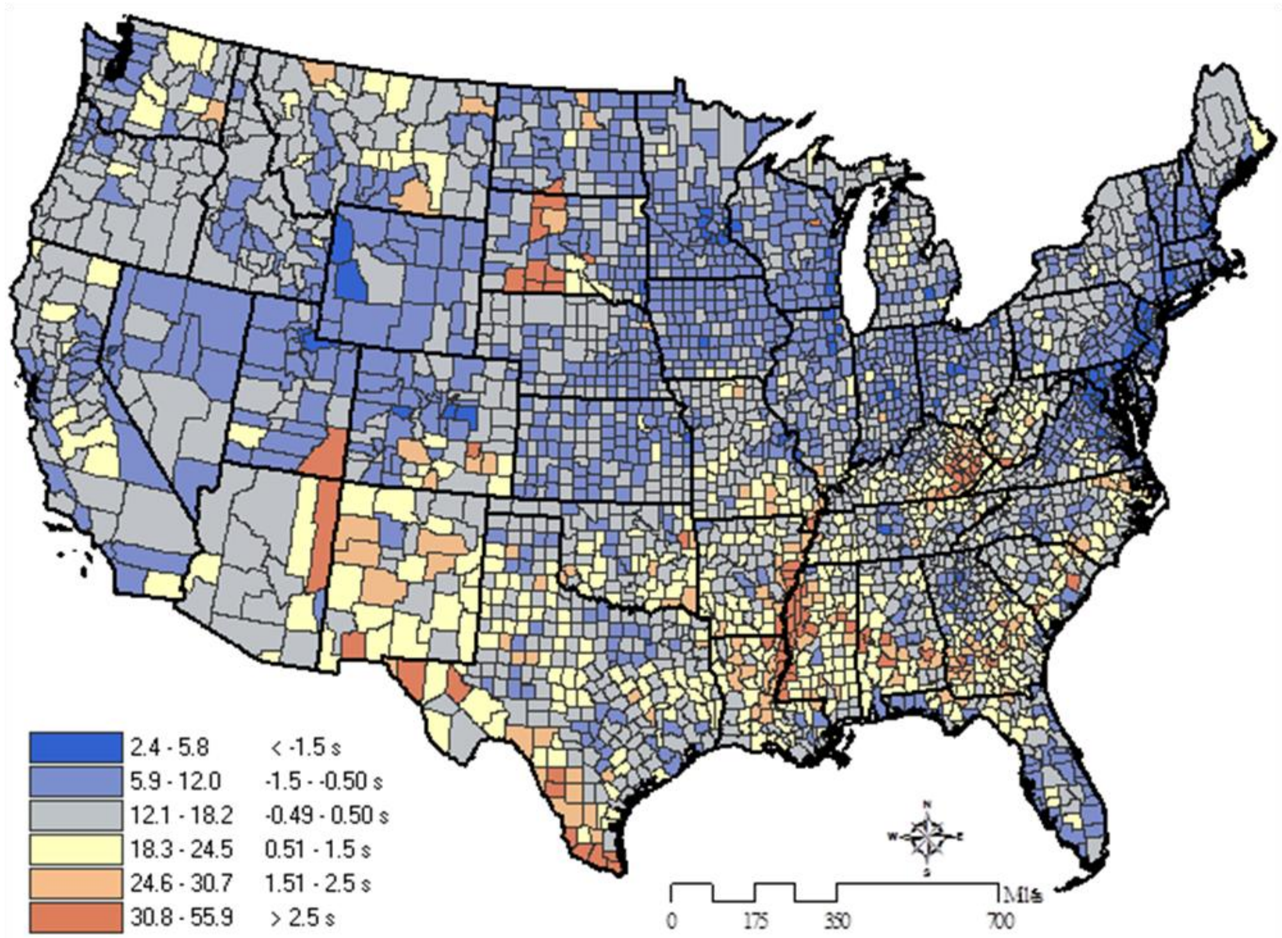

Figure 4.17 Geographic distribution of poverty prevalence rates at the county level in the US for 2007

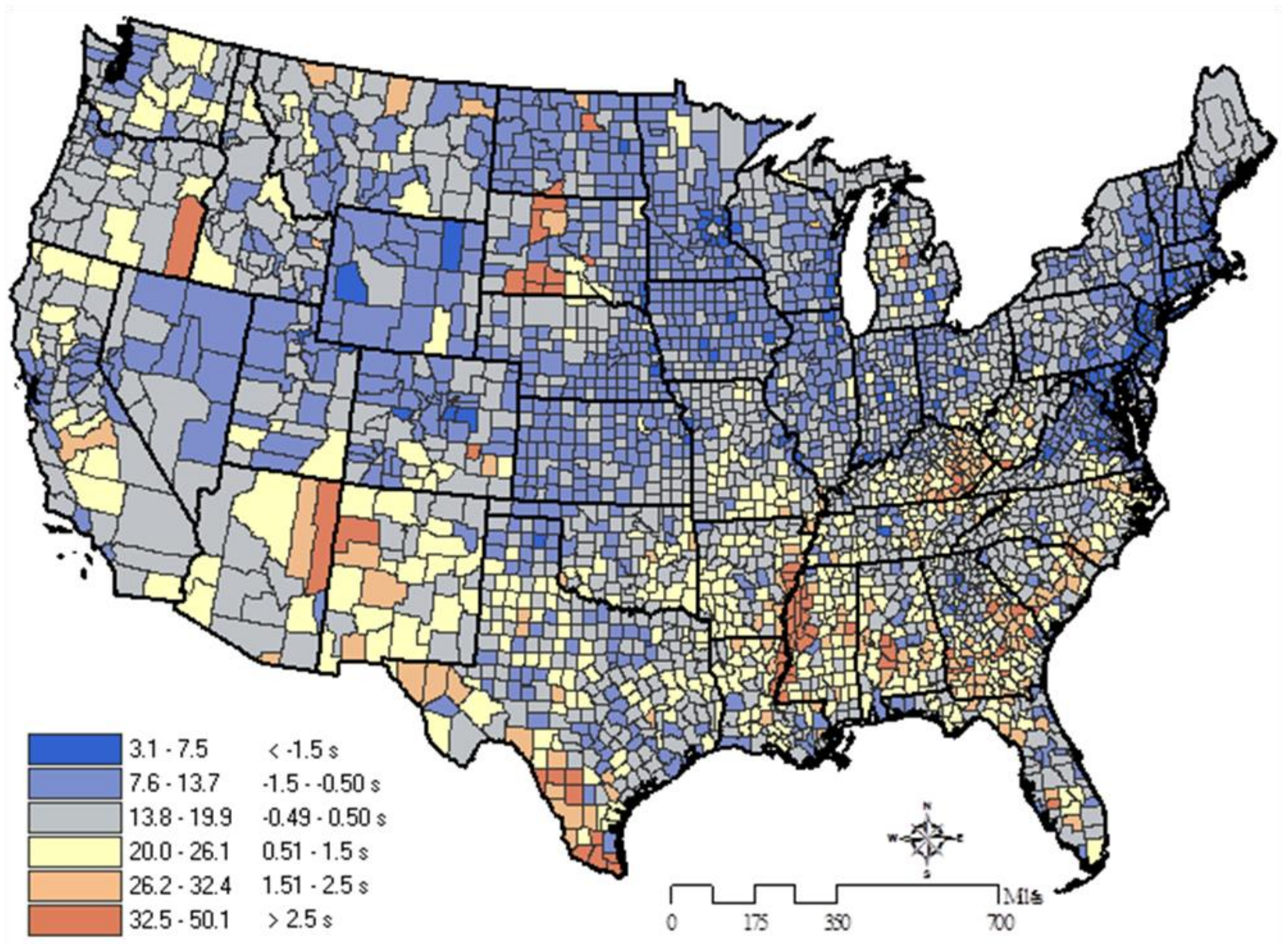

Figure 4.18 Geographic distribution of poverty prevalence rates at the county level in the US for 2010

4.3.1.4 Ratio of Jobs to Employed Residents (JER)

In 2010, among the 3,105 counties in the US, the ratio of jobs to employed residents (JERs) range from 0.20 (Long, Georgia) to 3.92 (Tunica, Mississippi). The mean JER is 0.89 with a standard deviation of 0.24. The median JER is 0.90 (Table 4.4).

Table 4.4 Descriptive statistics for county level ratio of jobs to employed residents (JERs) in the US for 2010

	2010
n	3,105
Mean	0.89
Std. Deviation	0.24
Median	0.90
Minimum	0.20
Maximum	3.92

The histogram below shows a positively skewed distribution of the county-level JERs in 2010 (the skewness value is 2.40 with a standard error of 0.04) (Figure 4.19). In 2010, more than half of the counties have JERs between 0.80 and 1.09. In addition, only 27.15% of all counties have JERs equal to or greater than 1.00 (Figure 4.19).

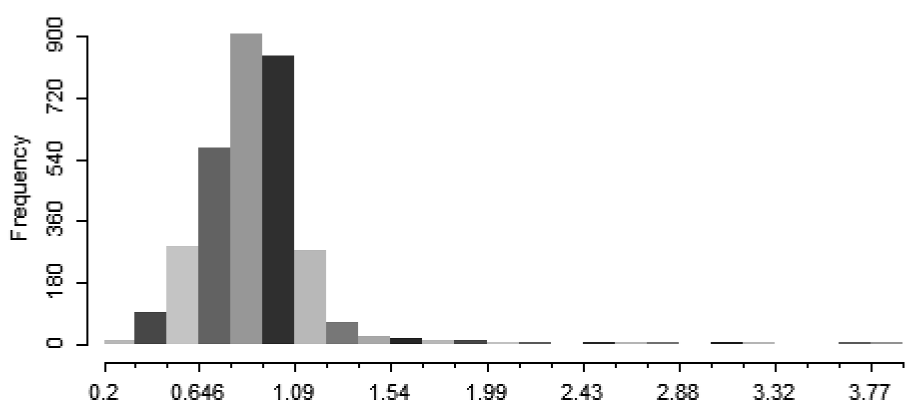

Figure 4.19 Histogram of county level ratio of jobs to employed residents (JER) in the US in 2010

Figure 4.20 shows the county-level JERs in 2010 using the standard deviation classification method. This map shows how county specific JERs vary among the counties throughout the US. As seen in Figure 4.20, the western half of the US has higher JERs compared to the eastern half. In addition, the lower JERs are observed mostly in the South and the south of the Midwest region.

In 2010, of the 3,105 counties, 1.06% have JERs below -2 SDs of the mean and 13.49% have JERs below -1 SD of the mean, whereas 8.82% have JERs +1 SD and higher than the mean and 2.09% have JERs +2 SDs and higher than the mean (Figure 4.20).

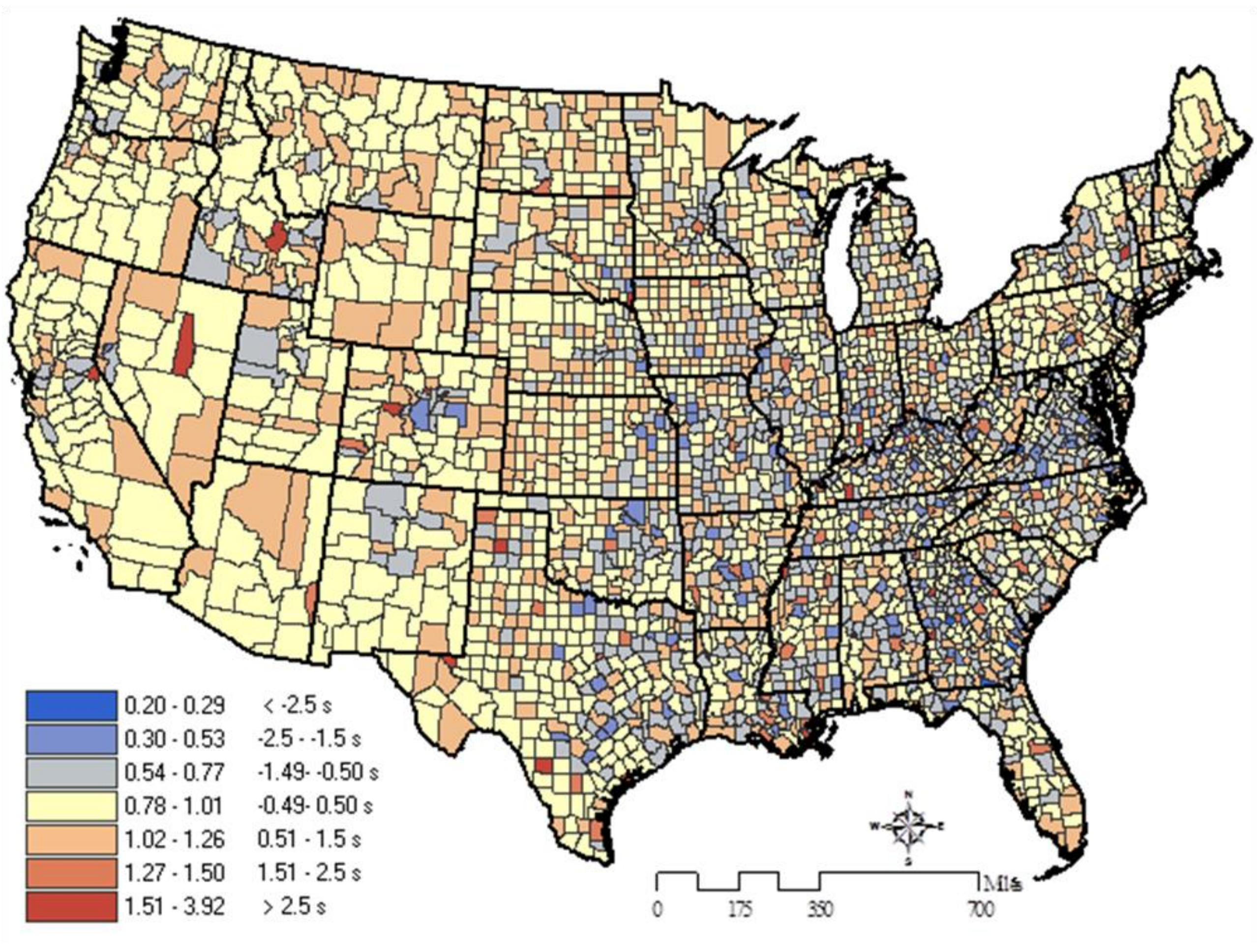

Figure 4.20 Geographic distribution of ratio of jobs to employed residents (JERs) at the county level in the US for 2010

4.3.1.5 Population-Weighted Distance (PWD) to Parks

In 2008, among the 3,105 counties, the population-weighted distances (PWDs) to parks range from 0.50 miles (Suffolk, Massachusetts) to 172.80 miles (Petroleum, Montana). The mean PWD to parks is 18.86 with a standard deviation of 17.08. The median PWD to parks is 14.54 (Table 4.5).

Table 4.5 Descriptive statistics for county level population-weighted distances (PWDs) to parks in the US for 2008

	2008
n	3,105
Mean	18.86
Std. Deviation	17.08
Median	14.54
Minimum	0.50
Maximum	172.80

The histogram shows a positively skewed distribution of the county-level PWDs to parks (the skewness value is 2.65 with a standard error of 0.04) (Figure 4.21).

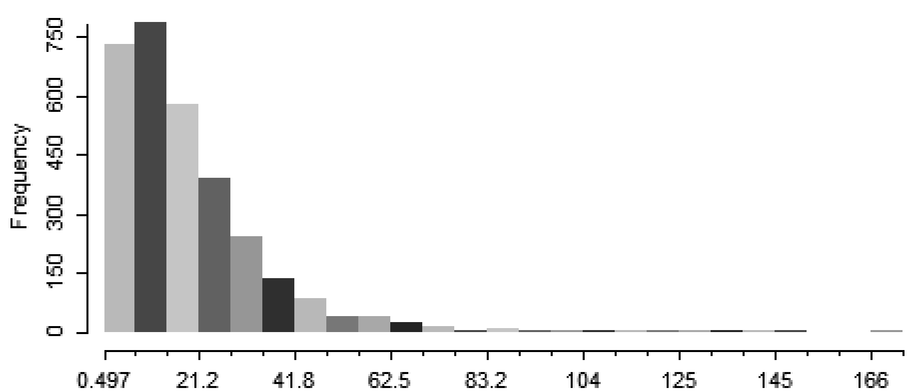

Figure 4.21 Histogram of county level population-weighted distance (PWD) to parks in the US in 2008

Figure 4.22 shows the county-level PWDs to parks in 2008 using the standard deviation classification method. This map shows how county specific PWDs to parks differ among the counties throughout the United States. As shown in Figure 4.22, the higher PWDs to parks are especially seen in the western half of the US and in the southern states. In addition, the state of Maine stands out by having higher PWDs to parks than other states in the Northeast.

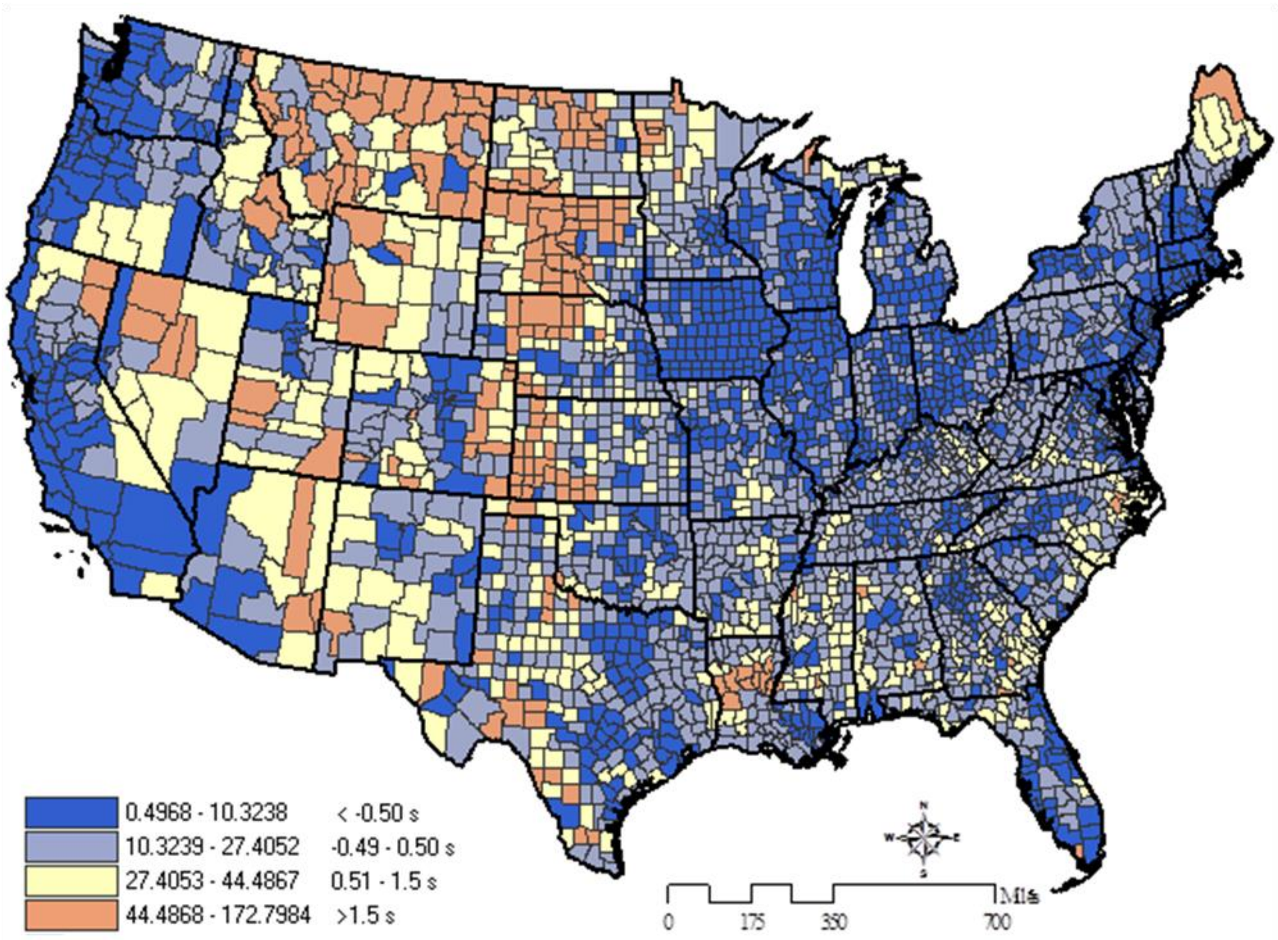

Figure 4.22 Geographic distribution of population-weighted distances (PWDs) to parks at the county level in the US for 2008

In 2008, of the 3,105 counties, not a single county has PWD to parks below -2 SDs of the mean and 4.99% have PWDs to parks below -1 SD of the mean while 10.89% have PWDs to parks +1 SD and higher than the mean and 3.90% have PWDs to parks +2 SDs and higher than the mean (Figure 4.22).

4.3.1.6 Population Density

Among the counties for the years 2004, 2007, 2010, the population densities range from 0.08 (Loving, Texas) to 68,084.38 (New York, New York) in 2004, from 0.08 (Loving, Texas) to 71,146.03 (New York, New York) in 2007, and from 0.12 (Loving, Texas) to 69,538.93 (New York, New York) in 2010. In all years, the highest population densities are observed in New York County in New York in the Northeast while the lowest population densities are seen in Loving County in Texas in the South. The mean population densities and the standard deviations are 248.35 (1689.27) in 2004, 254.70 (1742.84) in 2007, and 259.56 (1731.13) in 2010. The medians are 44.41, 44.65, and 45.59 in 2004, 2007, and 2010, respectively (Table 4.6).

Table 4.6 Descriptive statistics for county level population densities in the US for 2004, 2007, and 2010

	2004	2007	2010
n	3,105	3,105	3,105
Mean	248.35	254.70	259.56
Std. Deviation	1689.27	1742.84	1731.13
Median	44.41	44.65	45.59
Minimum	0.08	0.08	0.12
Maximum	68084.38	71146.03	69538.93

The histograms below show the distribution of the population densities in all years (Figures 4.23, 4.24, and 4.25). The original data is transformed using the logarithmic transformation in order to better reveal the patterns in all three years.

Figures 4.26, 4.27, and 4.28 illustrate the county-level population densities for the years 2004, 2007, and 2010. These maps show how county specific population densities vary throughout the US. As seen in the maps, the highest population densities are observed in the metropolitan areas throughout the US, as expected. In particular, the Megalopolis region that stretches from Virginia in to New Hampshire emerges as the most distinctive areas having high population densities over 1,000.

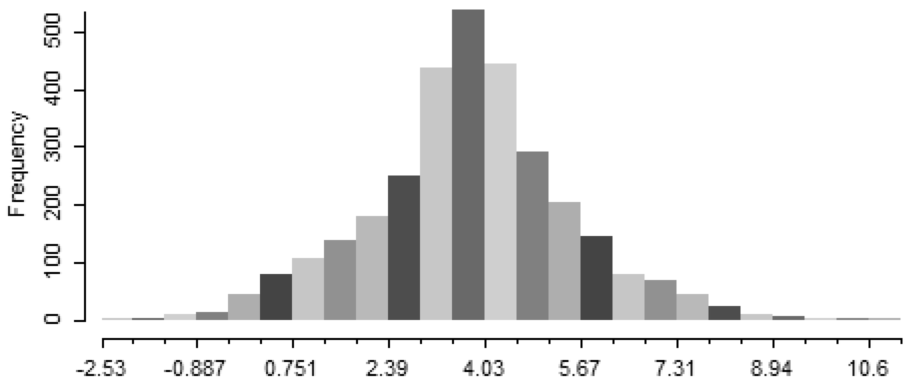

Figure 4.23 Histogram of county level population densities in the US in 2004

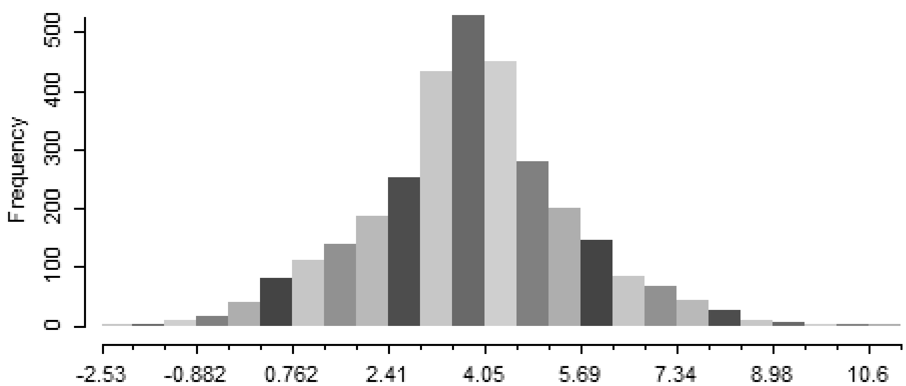

Figure 4.24 Histogram of county level population densities in the US in 2007

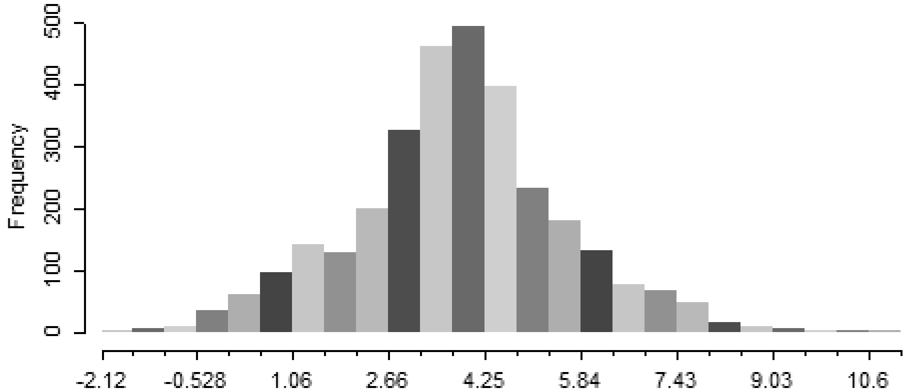

Figure 4.25 Histogram of county level population densities in the US in 2010

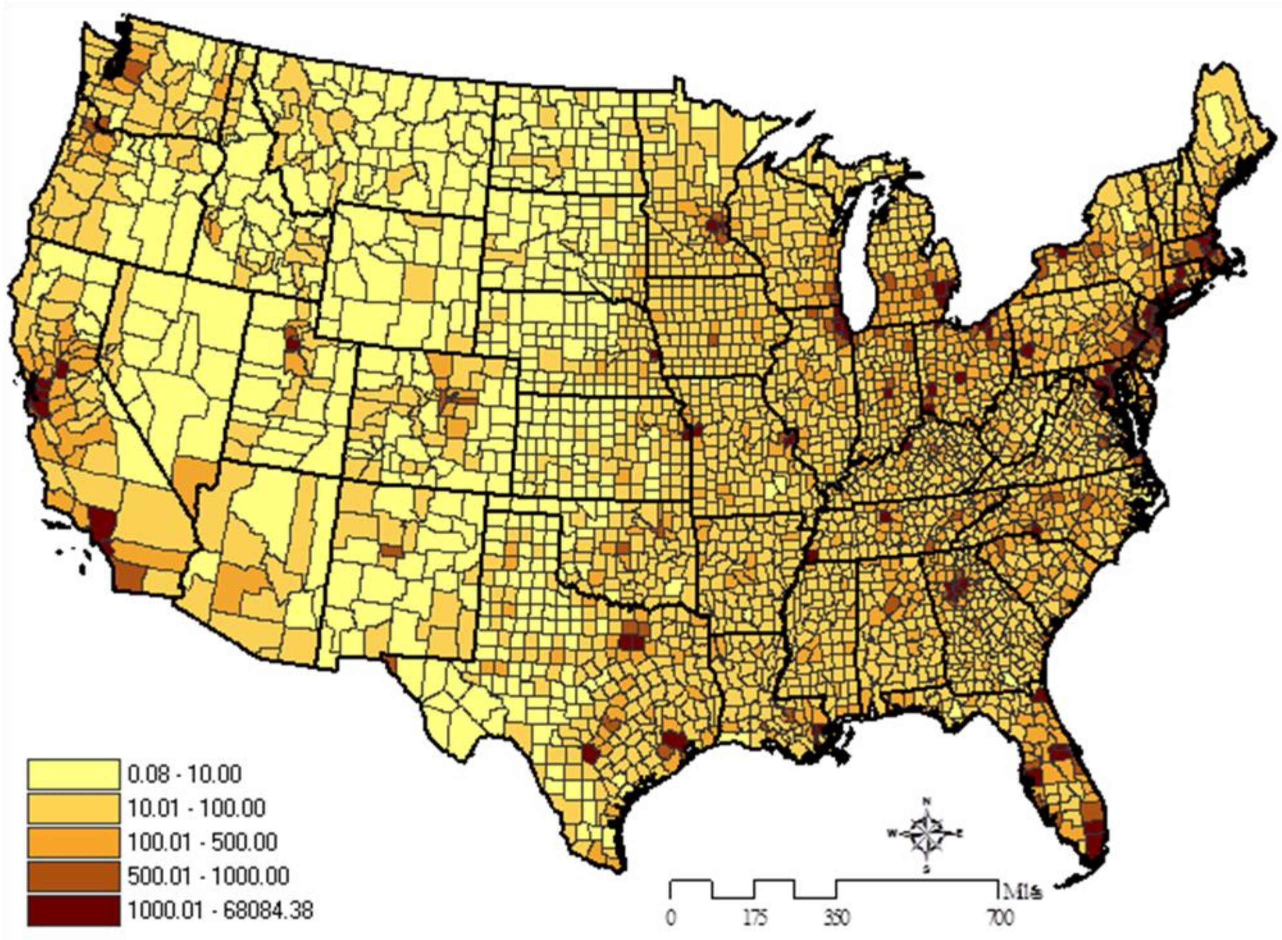

Figure 4.26 Geographic distribution of population densities at the county level in the US for 2004

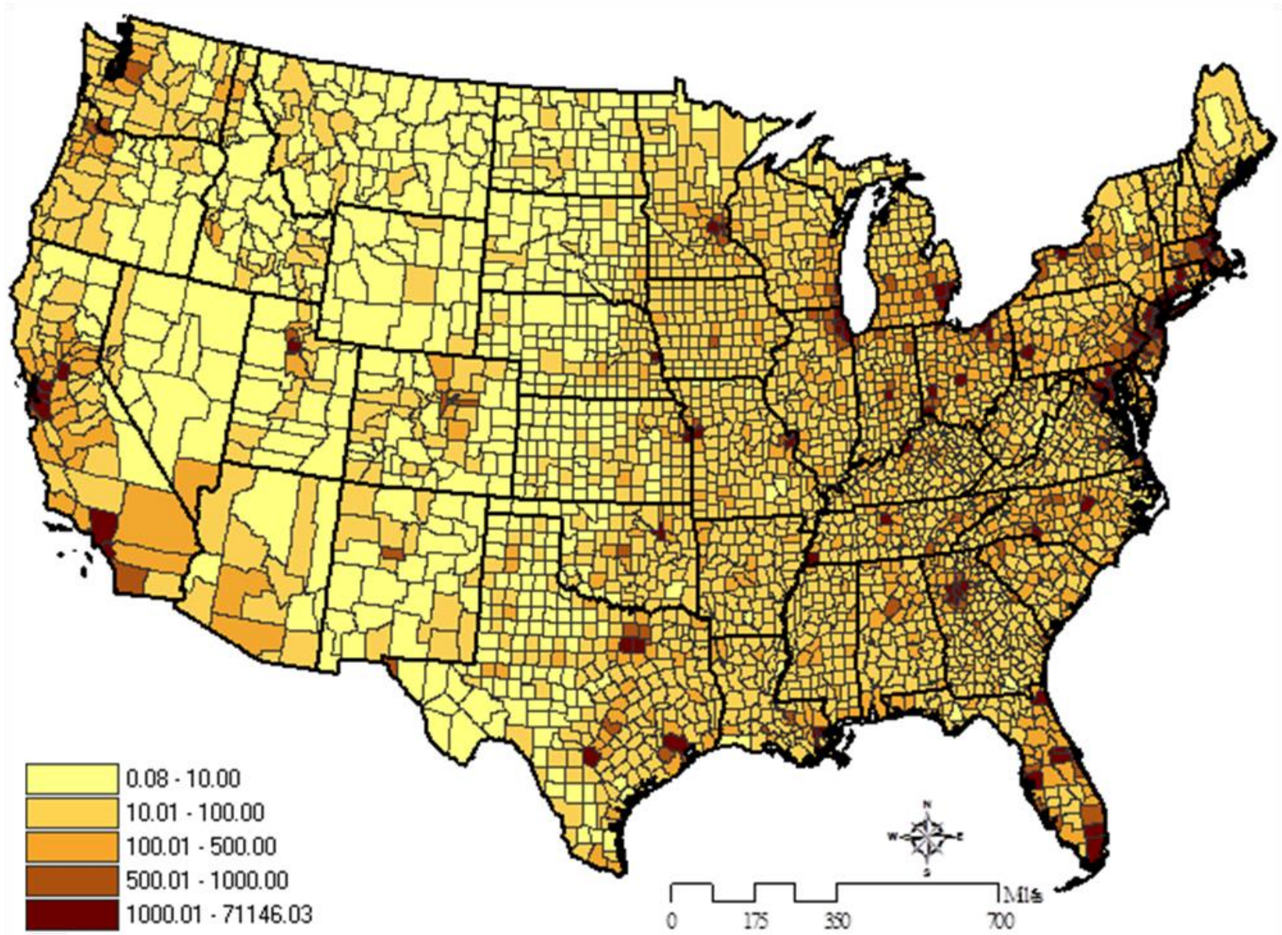

Figure 4.27 Geographic distribution of population densities at the county level in the US for 2007

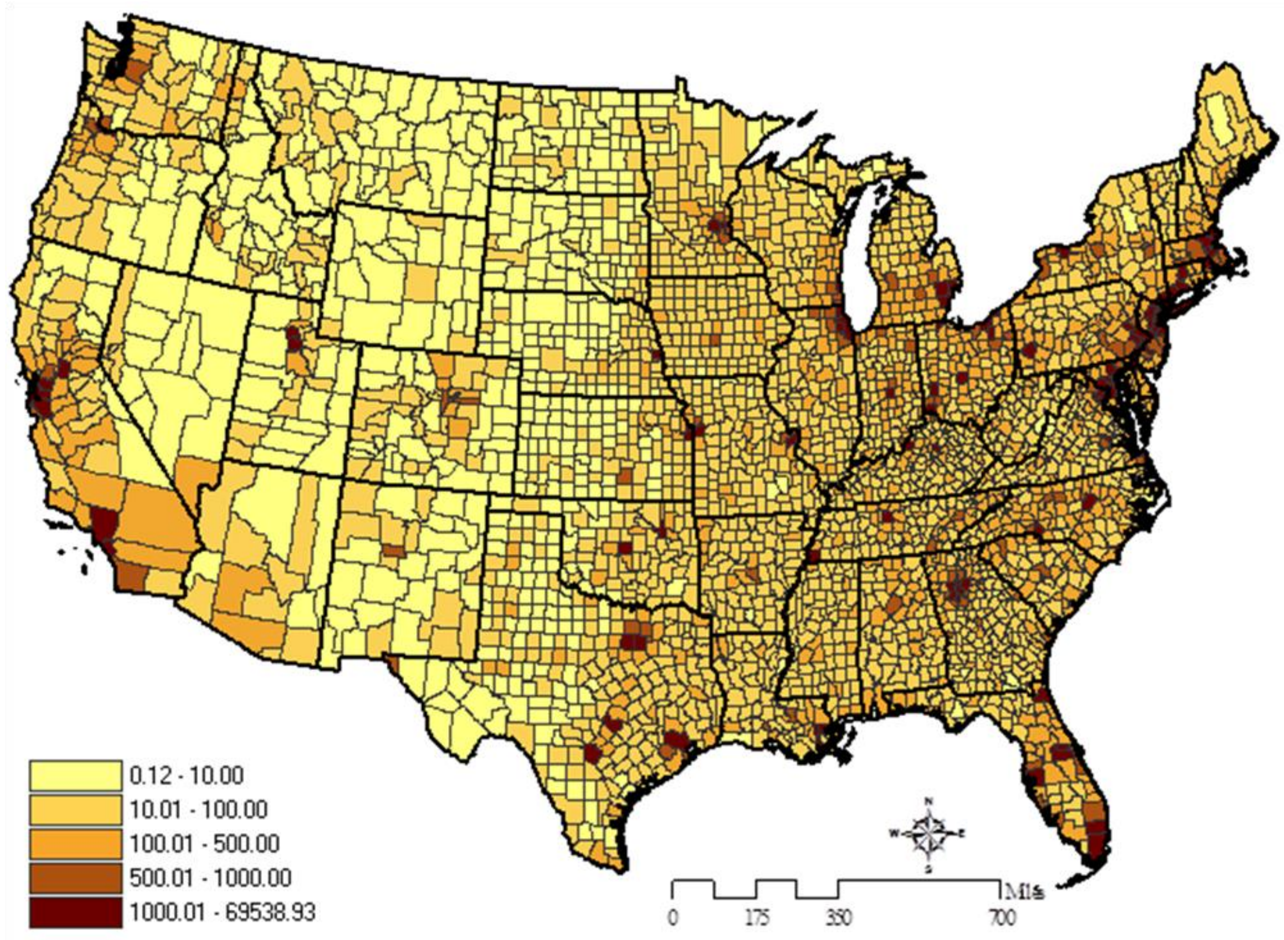

Figure 4.28 Geographic distribution of population densities at the county level in the US for 2010

In all years, of all counties, not a single county has a population density below -2 SDs and/or below -1 SD of the mean. In 2004, of all counties, 1.96% have population densities of +1 SD higher than the mean and 0.81% have population density of +2 SDs higher than the mean (Figure 4.26). In 2007, of all counties, 1.90% have population densities of +1 SD higher than the mean and 0.84% have population density of +2 SDs higher than the mean (Figure 4.27). In 2010, of all counties, 1.96% have population densities of +1 SD higher than the mean and 0.84% have population density of +2 SDs higher than the mean (Figure 4.28).

4.3.2 Univariate Moran's I and LISA Results

The following findings are based on advanced spatial analyses performed in GeoDa for all county-level variables for the years 2004, 2007, and 2010. The univariate global Moran's I and LISA statistics were utilized and the results, in the forms of Moran scatter plots and cluster maps, are presented here. While the global Moran's I statistics describe the general extent of spatial clustering of observed values across the study area, the LISA statistics identify the locations of clusters or spatial outliers that are exhibited as a result of the spatial autocorrelation process. In univariate forms of the Moran's I and/or LISA, the tests compare the value of a variable at any one location with the value of the same variable at all other locations. In other words, in the univariate forms of Moran's I and/or LISA, the observed values, y, and the spatial lag of y were based on the same variable.

The Moran scatter plots not only quantify the degree of clustering (positive spatial autocorrelation) in each year, but also show how the magnitude of clustering (positive spatial autocorrelation) has changed among the years studied. The results of the univariate global Moran's I are supported by the visual indications of clustering and/or dispersion at a localized extent in the LISA cluster maps. With the utilization of the univariate LISA tests, cluster maps were generated and the locations were classified and shown by the type of spatial correlation. Four resulting classifications are represented. The high-high (areas in red) and low-low (areas in blue) locations indicate clustering of similar values (i.e., hot spots and cold spots, respectively). On the other hand, the high-low (areas in lighter red) and low-high (areas in lighter blue) locations illustrate spatial outliers (dissimilar values). More specifically, the high-low locations are indicative of that high values are surrounded by low values whereas the low-high locations show where low values are surrounded by high values. Only the statistically significant patterns are visualized on the maps ($p<0.05$) while insignificant spatial associations are shown in light grey. In addition, the corresponding significance maps are presented in Appendix A.

4.3.2.1 Obesity

In Figure 4.29, the results of the univariate Moran's I statistics for the county-level obesity rates in the US for the years 2004, 2007, and 2010 are presented. In all years, the univariate Moran's I statistics indicate positive spatial autocorrelation. The results of the Moran's I statistics are 0.68 (2004), 0.66 (2007), and 0.59 (2010) and based on the 999 permutations run, they are significant ($p<0.001$). These results reveal that there is continued spatial clustering of the high and/or low obesity rates at the county level in the US. However, the statistics also indicate that the magnitude of spatial autocorrelation has decreased slightly during the seven-year observation period. This decline suggests that the obesity rates between some counties and their neighbors have become less concentrated over the years. This is applicable to both hot spot and/or cold spot cluster locations. This indicates that obesity has become more prevalent over the years and thus it has become less concentrated.

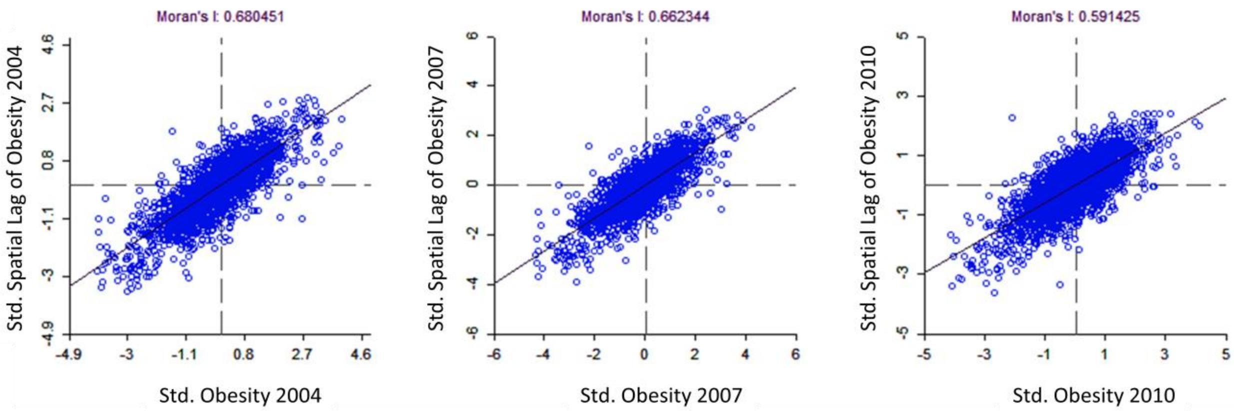

Figure 4.29 Univariate Moran's I scatter plots for county level obesity prevalence rates in the US for 2004, 2007, and 2010

As shown in Figures 4.30, 4.31, and 4.32, the obesity rates are mapped and categorized based on the results of the univariate LISA statistics ($p<0.05$) to illustrate the localized extent and the nature of the spatial clustering in the obesity data for the years 2004, 2007, and 2010. In all three years, the number of counties categorized as hot spot obesity clusters exceeds the number of counties categorized as cold spot obesity clusters.

As seen in Figures 4.30, 4.31, and 4.32, in all years, overall clear high obesity clusters are mainly seen in the East South Central and South East (Atlantic) regions (excluding Texas and Florida) in the South. Particularly, there are two distinguished and belt-like high obesity clusters, in the direction of southwest to southeast, that run almost parallel to each other. The first one extends from Louisiana to Virginia covering Louisiana, Mississippi, Alabama, Georgia, South Carolina, North Carolina, and Virginia. The second one runs through Oklahoma, Arkansas, Tennessee, Kentucky and West Virginia. These two parallel and belt-like clusters run mainly along the southern and northern boundaries of the South US. The states included in these two belts largely represent cultural and geographical subregion of the American South and therefore they stand out as two separate but associated belts within the context of obesity.

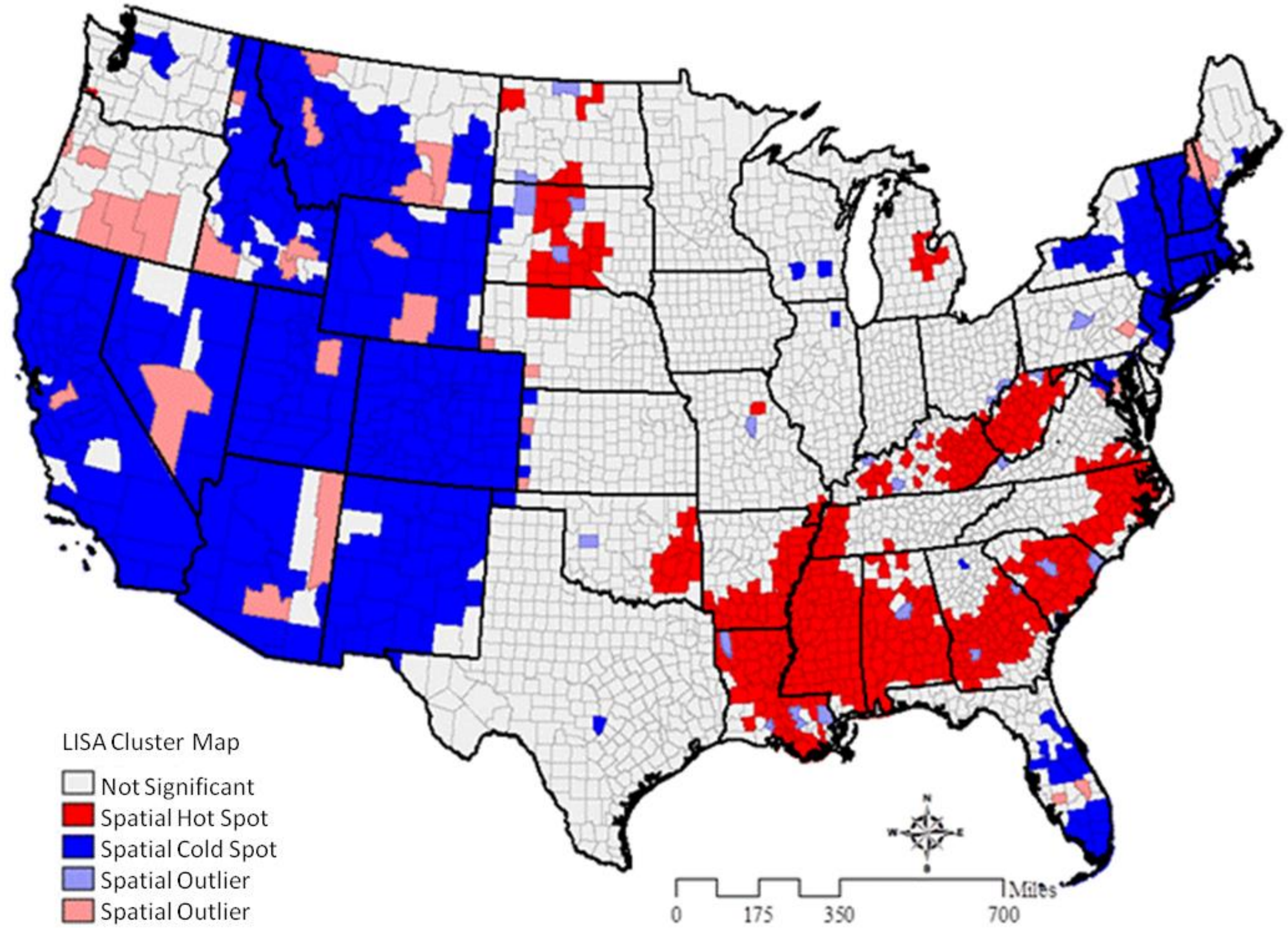

Figure 4.30 Univariate LISA cluster map for county level obesity prevalence rates in the US for 2004

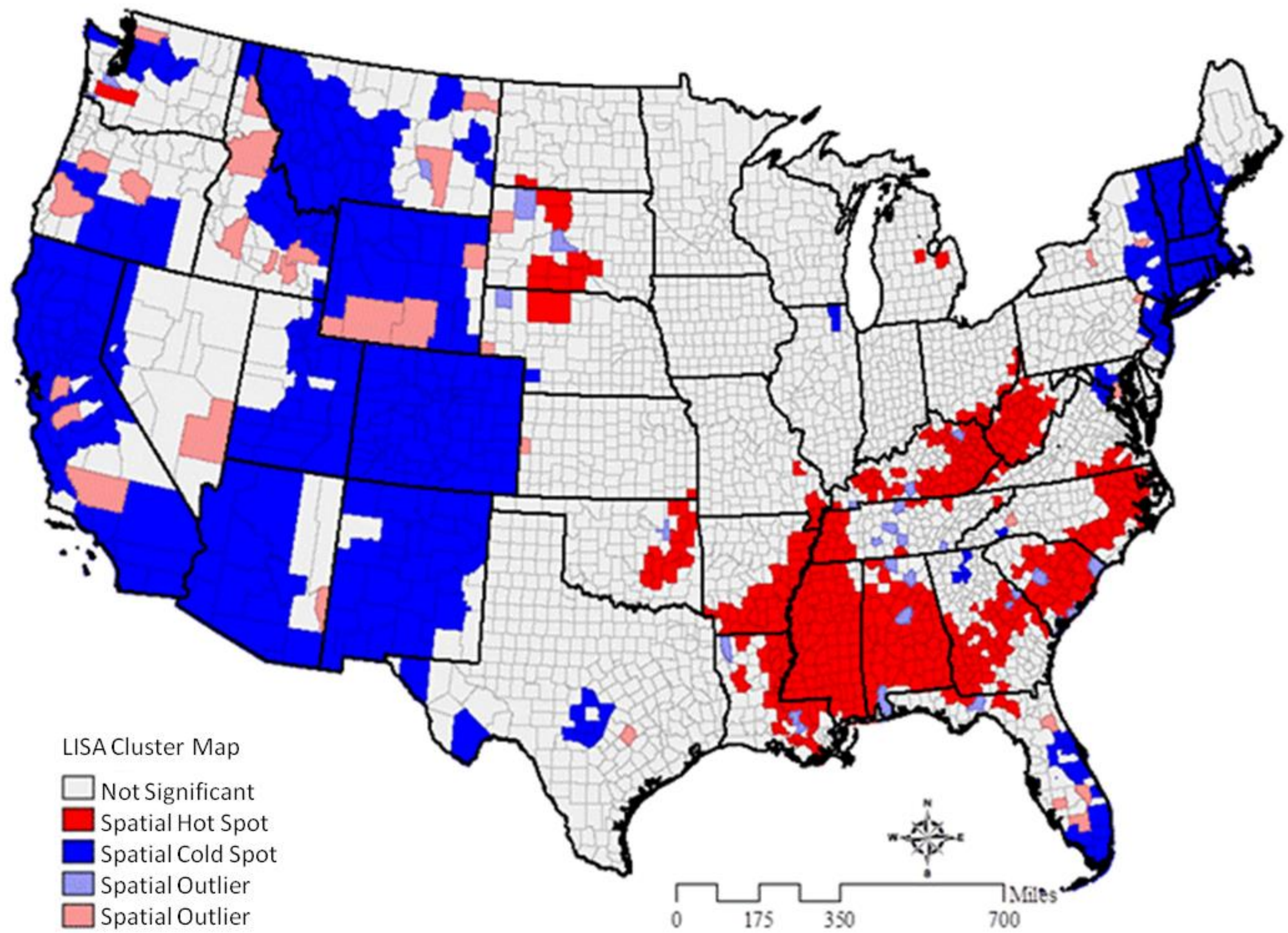

Figure 4.31 Univariate LISA cluster map for county level obesity prevalence rates in the US for 2007

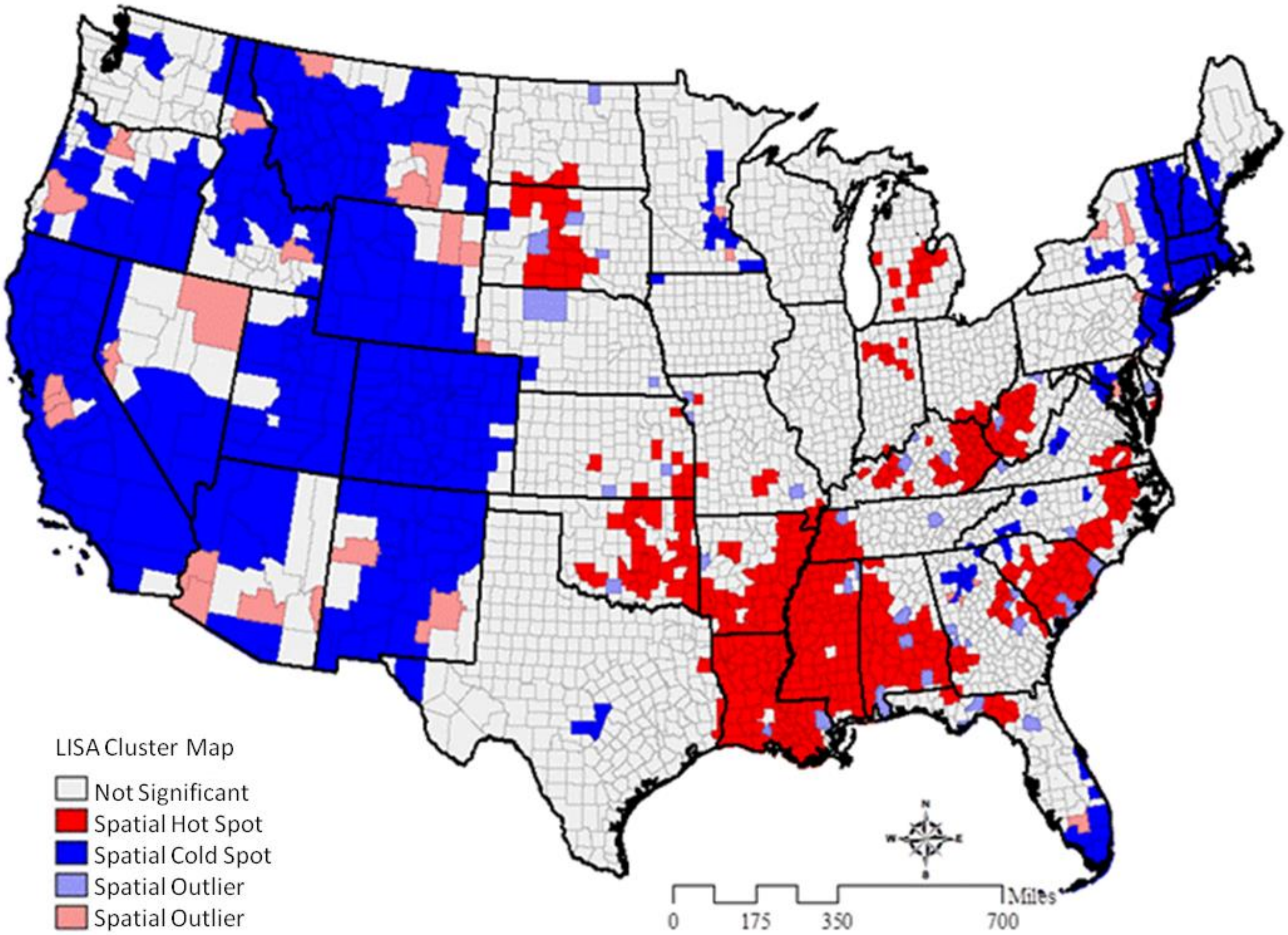

Figure 4.32 Univariate LISA cluster map for county level obesity prevalence rates in the US for 2010

There are also distinct high obesity clusters in the Dakotas region which are observed more predominantly in the state of South Dakota. Over this time period, a growing cluster of high obesity is also observed in Oklahoma in the South and in Michigan and Indiana in the East North Central region in the Midwest. On the other hand, slightly diminishing high obesity clusters are detected in the states of Georgia, North Carolina, and Kentucky in the South.

In all three years, overall clear low obesity clusters are mainly seen in the West and Northeast. In the South, Texas and Florida are the only states that showed a distinct low obesity clustering patterns throughout this time period. Even though the Southern US is dominated by high obesity clusters, growing clusters of low obesity are also observed in Georgia, North

Carolina, and Virginia by 2010. A similar growth pattern in the low obesity clusters is also seen in Minnesota in the West North Central region. Further examination of the map reveals that these low obesity cluster growth areas correspond to the counties or county equivalents where metropolitan areas exist.

4.3.2.2 Physical Inactivity (PIA)

In Figure 4.33, the results of the univariate Moran's I statistics for the county-level PIA rates in the US for the years 2004, 2007, and 2010 are presented. In all years, the univariate Moran's I statistics indicate high positive spatial autocorrelation. The results of the Moran's I statistics are 0.76 (2004), 0.73 (2007), and 0.67 (2010) and based on the 999 permutations run, they are significant ($p<0.001$). Similar to the Moran's I results for the obesity rates, these results reveal that there is continued spatial clustering of the high and/or low PIA rates at the county level in the US. However, the statistics also indicate that the magnitude of spatial autocorrelation has slightly decreased during the seven-year observation period. This decline suggests that the spatial autocorrelation of the PIA rates between some counties and their neighbors have become less concentrated over the years. This is applicable to both hot spot and/or cold spot cluster locations.

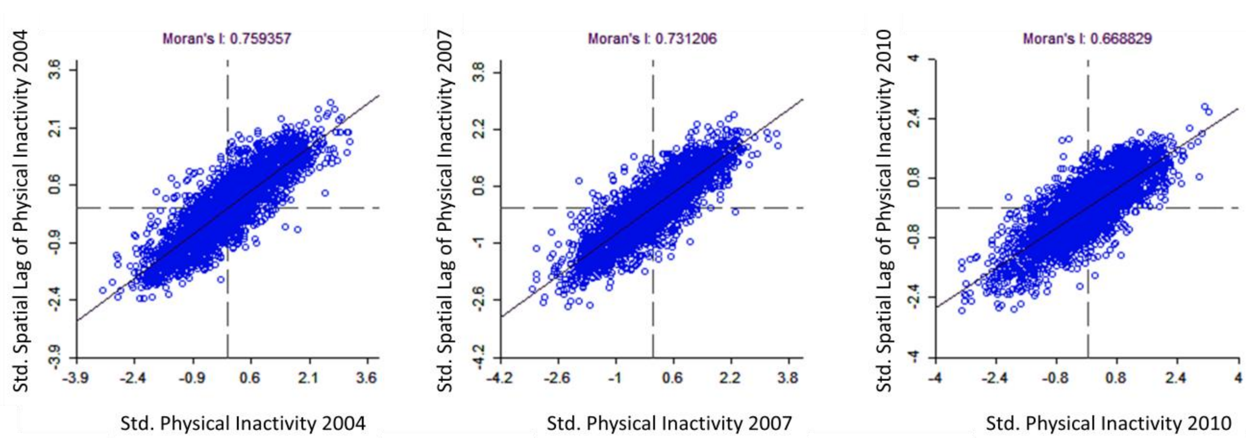

Figure 4.33 Univariate Moran's I scatter plots for county level physical inactivity (PIA) prevalence rates in the US for 2004, 2007, and 2010

Figures 4.34, 4.35, and 4.36 illustrate the mapped and categorized PIA rates based on the results of the univariate LISA statistics ($p<0.05$) for the period 2004, 2007, and 2010. In all three years, the number of counties categorized as hot spot PIA clusters exceeds the number of counties categorized as cold spot PIA clusters.

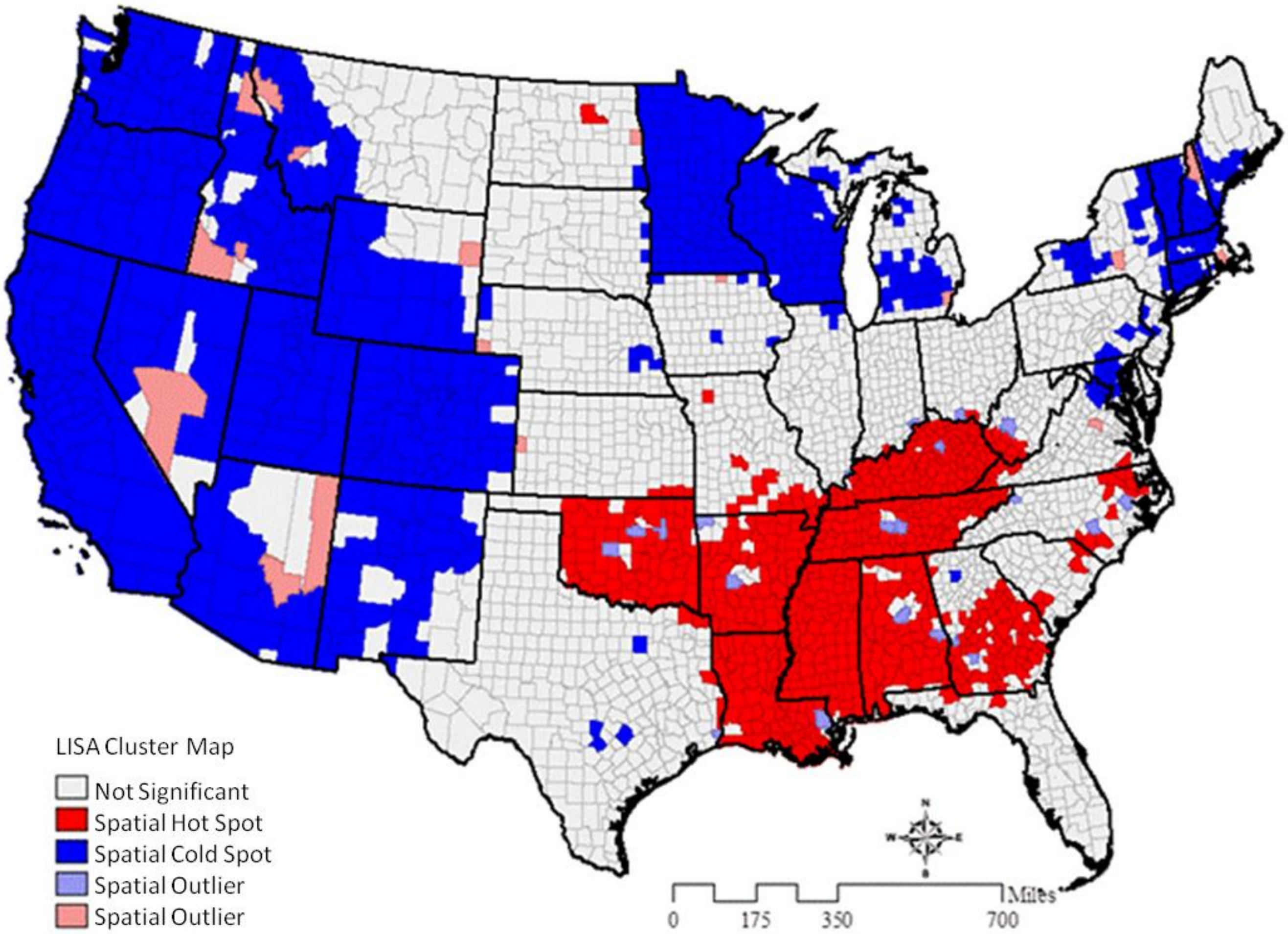

Figure 4.34 Univariate LISA cluster map for county level physical inactivity (PIA) prevalence rates in the US for 2004

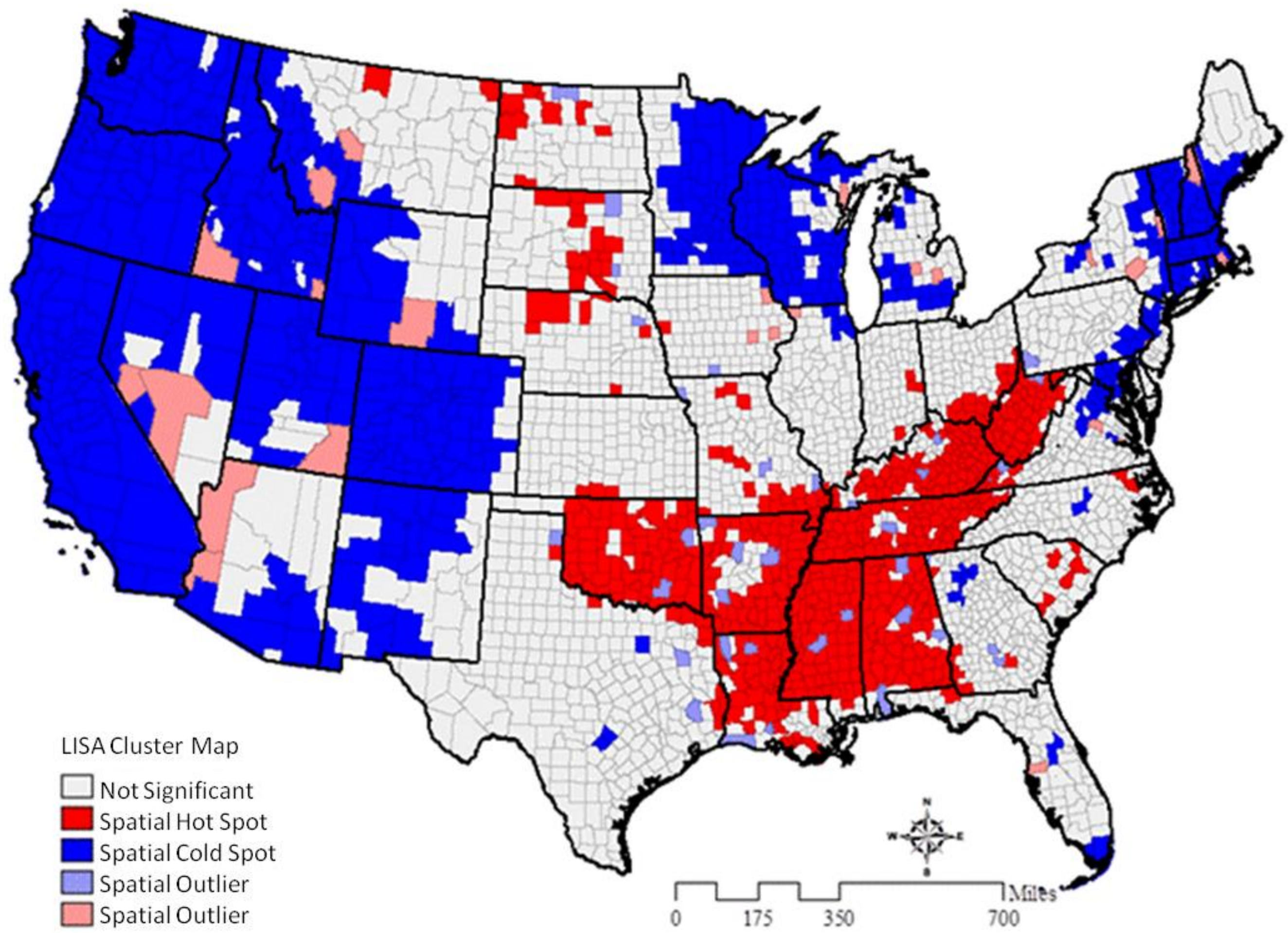

Figure 4.35 Univariate LISA cluster map for county level physical inactivity (PIA) prevalence rates in the US for 2007

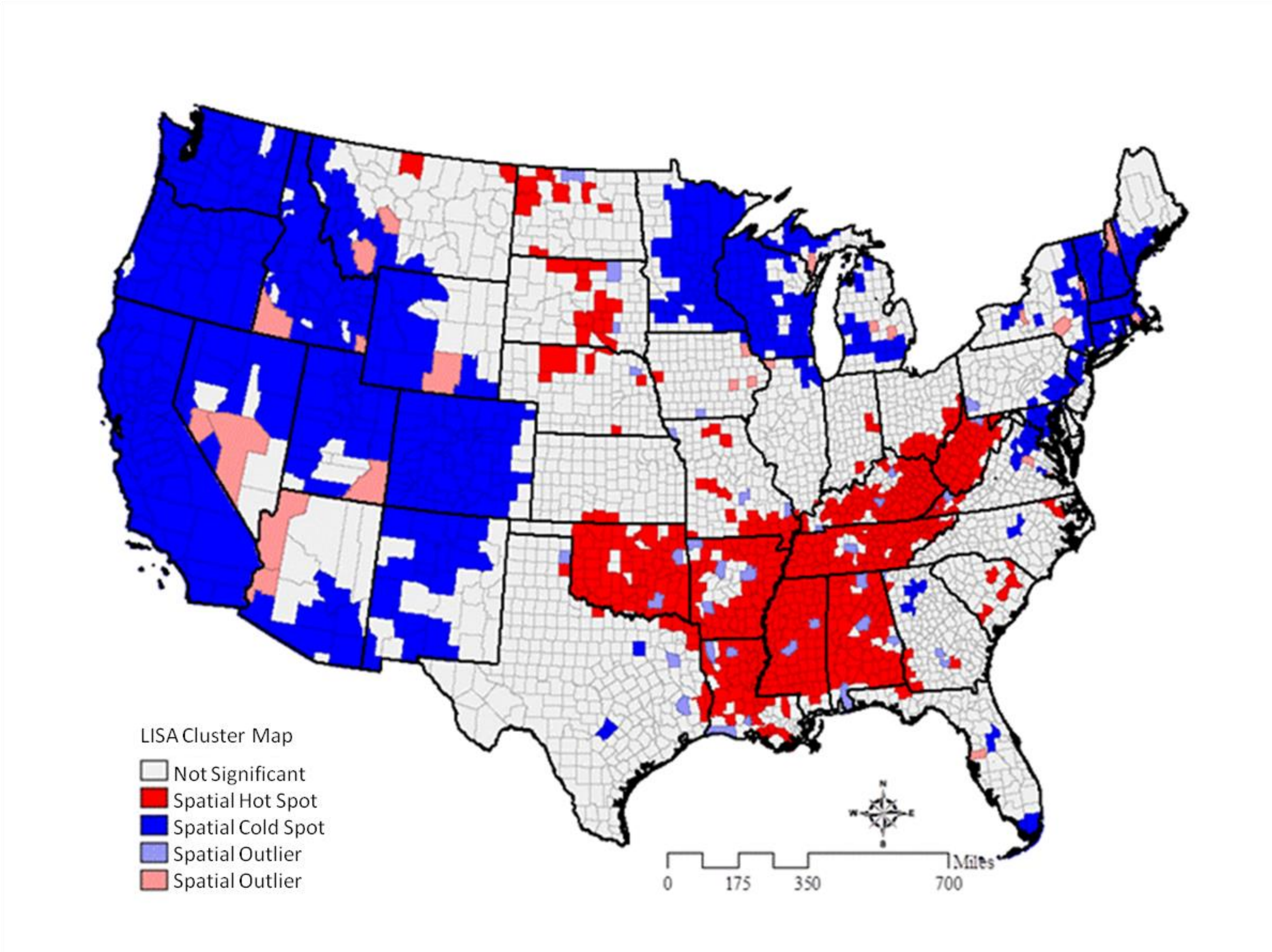

Figure 4.36 Univariate LISA cluster map for county level physical inactivity (PIA) prevalence rates in the US for 2010

4.3.2.3 Poverty

Figure 4.37 presents the results of the univariate Moran's I statistics for the county-level poverty rates in the US for the years 2004, 2007, and 2010. In all years, the univariate Moran's I statistics indicate positive spatial autocorrelation. The results of the Moran's I statistics are 0.65 in 2004, 0.59 in 2007, and 0.56 in 2010 and based on the 999 permutations run, they are significant ($p<0.001$). These results reveal that there is continued spatial clustering of the high and/or low poverty rates at the county level in the US. However, the statistics also indicate that the magnitude of spatial autocorrelation has decreased during the seven year observation period. This is a small decline, yet it still implies that the spatial correlation of the poverty rates between some counties and their neighbors have become less concentrated over the years for the hot spot and/or cold spot cluster locations.

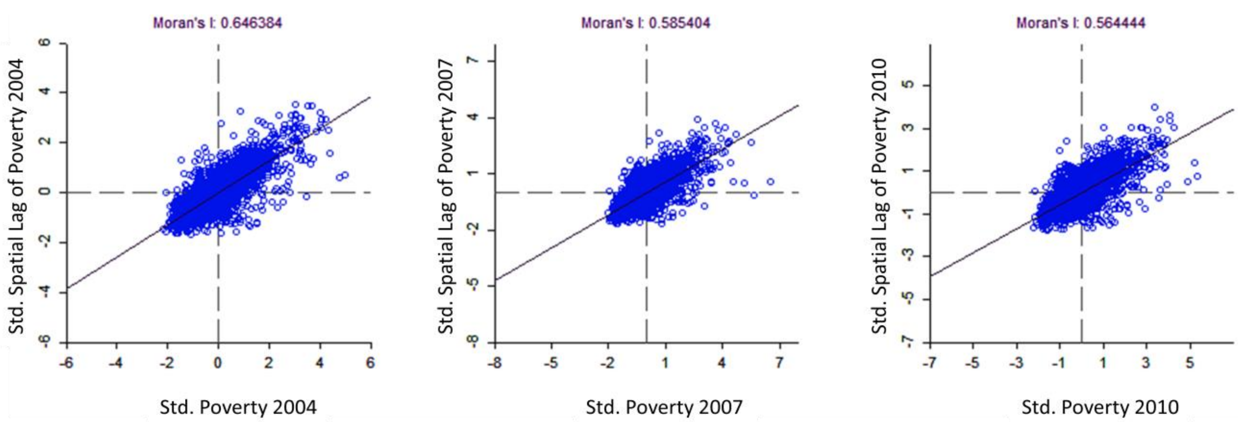

Figure 4.37 Univariate Moran's I scatter plots for county level poverty prevalence rates in the US for 2004, 2007, and 2010

Figures 4.38, 4.39, and 4.40 demonstrate the mapped and categorized poverty rates based on the results of the univariate LISA statistics ($p< 0.05$). In all three years, the number of counties categorized as hot spot poverty clusters is lower than the number of counties categorized as cold spot poverty clusters as clearly shown in the maps.

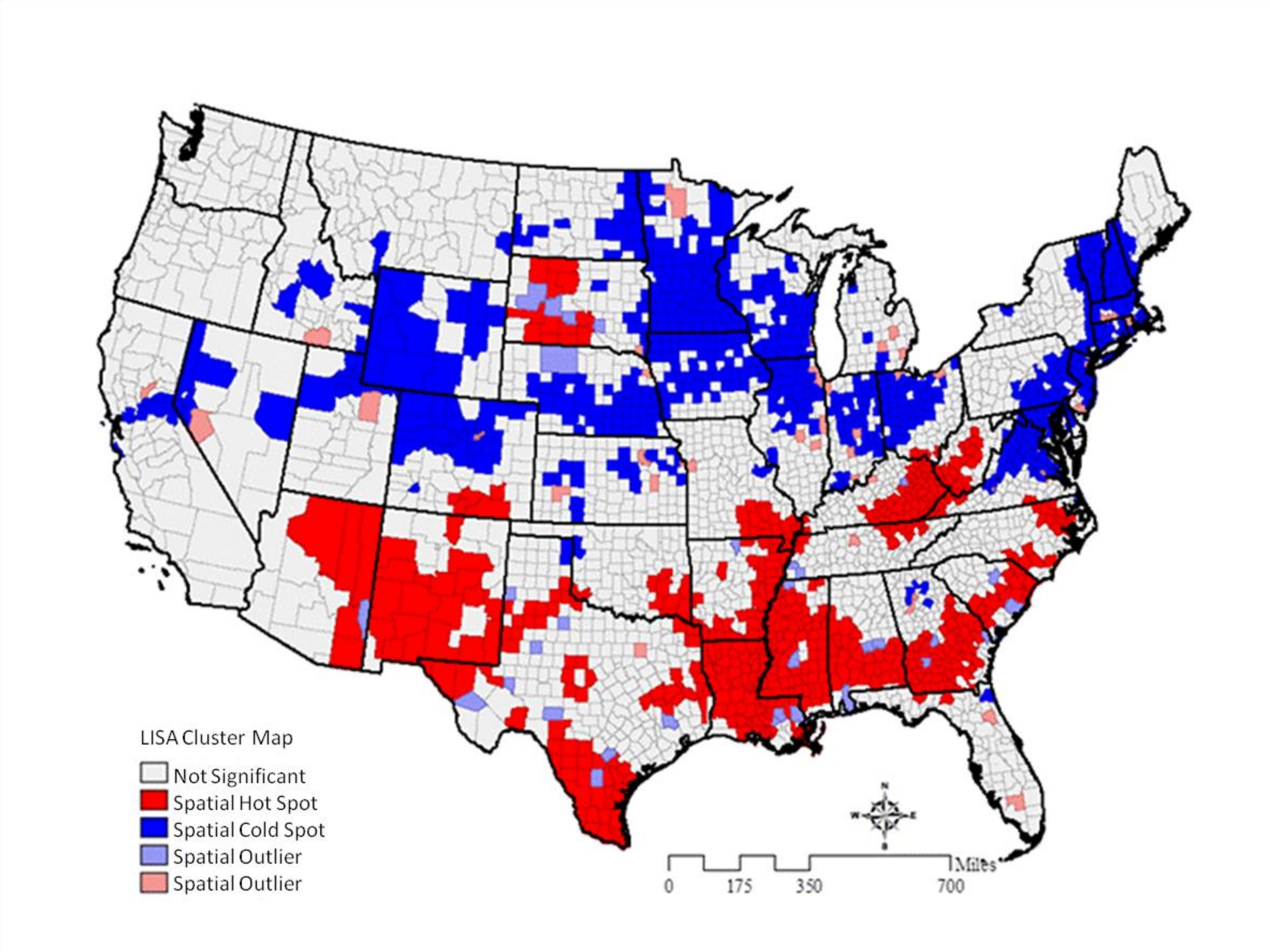

Figure 4.38 Univariate LISA cluster map for county level poverty prevalence rates in the US for 2004

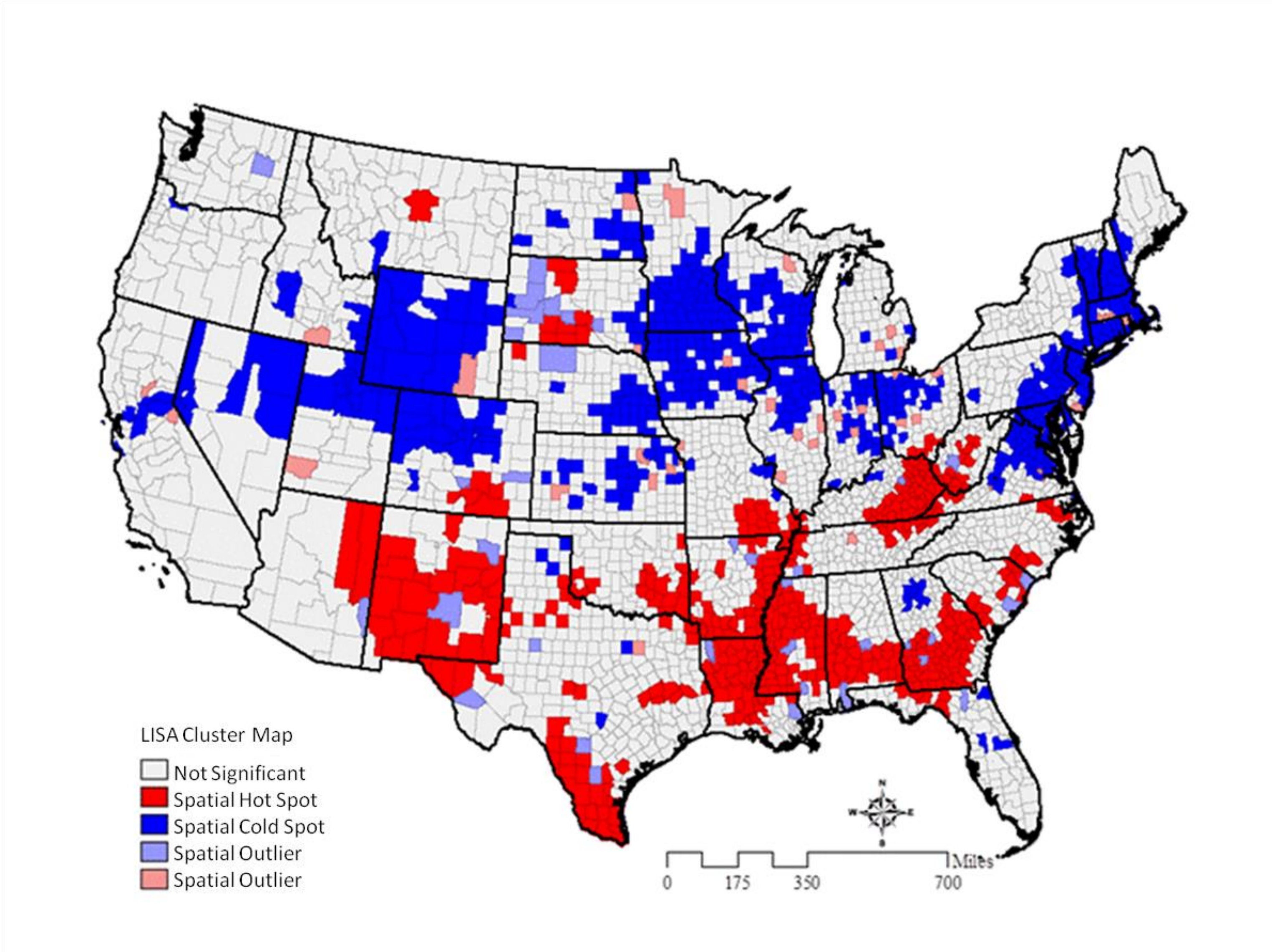

Figure 4.39 Univariate LISA cluster map for county level poverty prevalence rates in the US for 2007

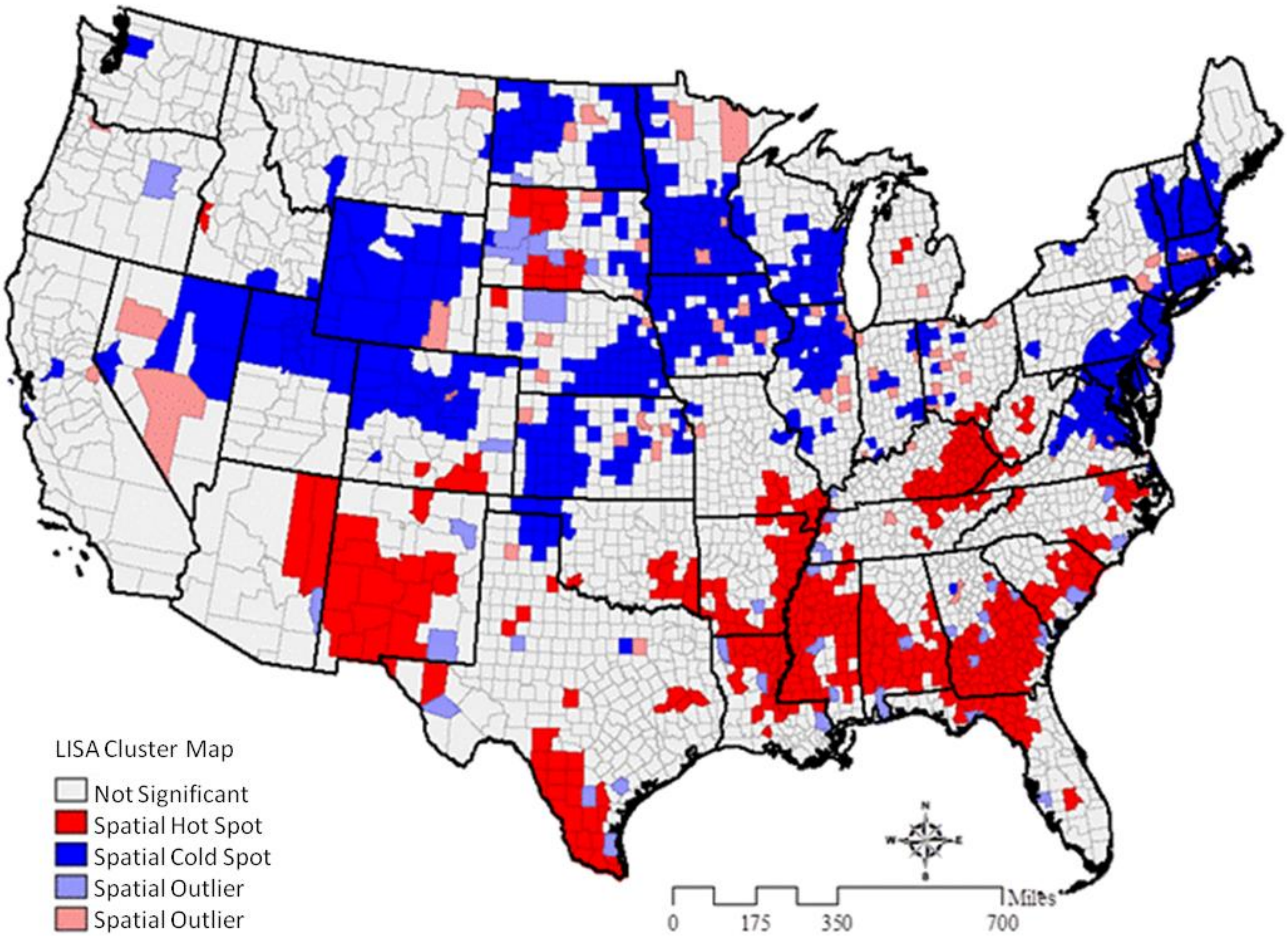

Figure 4.40 Univariate LISA cluster map for county level poverty prevalence rates in the US for 2010

4.3.2.4 Ratio of Jobs to Employed Residents (JER)

Figure 4.41 presents the results of the univariate Moran's I statistic for the county-level JERs in the US for the year 2010. The univariate Moran's I statistic indicates a negative spatial autocorrelation. The result of the Moran's I statistic is -0.02 in 2010 and based on the 999 permutations run, it is significant ($p<0.001$). This result shows that the high and/or low JERs are spatially dispersed at the county level in the US. Figure 4.42 illustrates the mapped and categorized JERs based on the results of the univariate LISA statistics ($p<0.05$). Even though the magnitude of the univariate Moran's I statistic is negative, suggesting a dispersed pattern, some clustering is still observed at the local level.

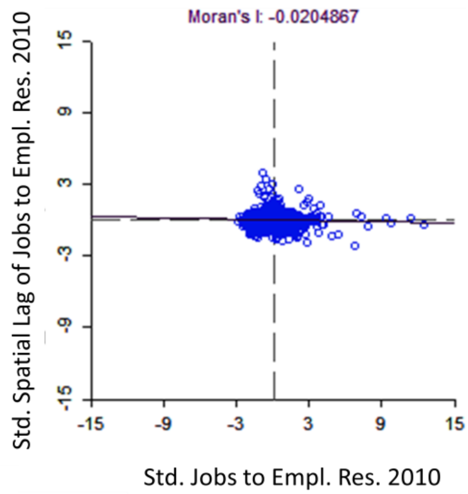

Figure 4.41 Univariate Moran's I scatter plots for county level ratio of jobs to employed residents (JER) in the US for 2010

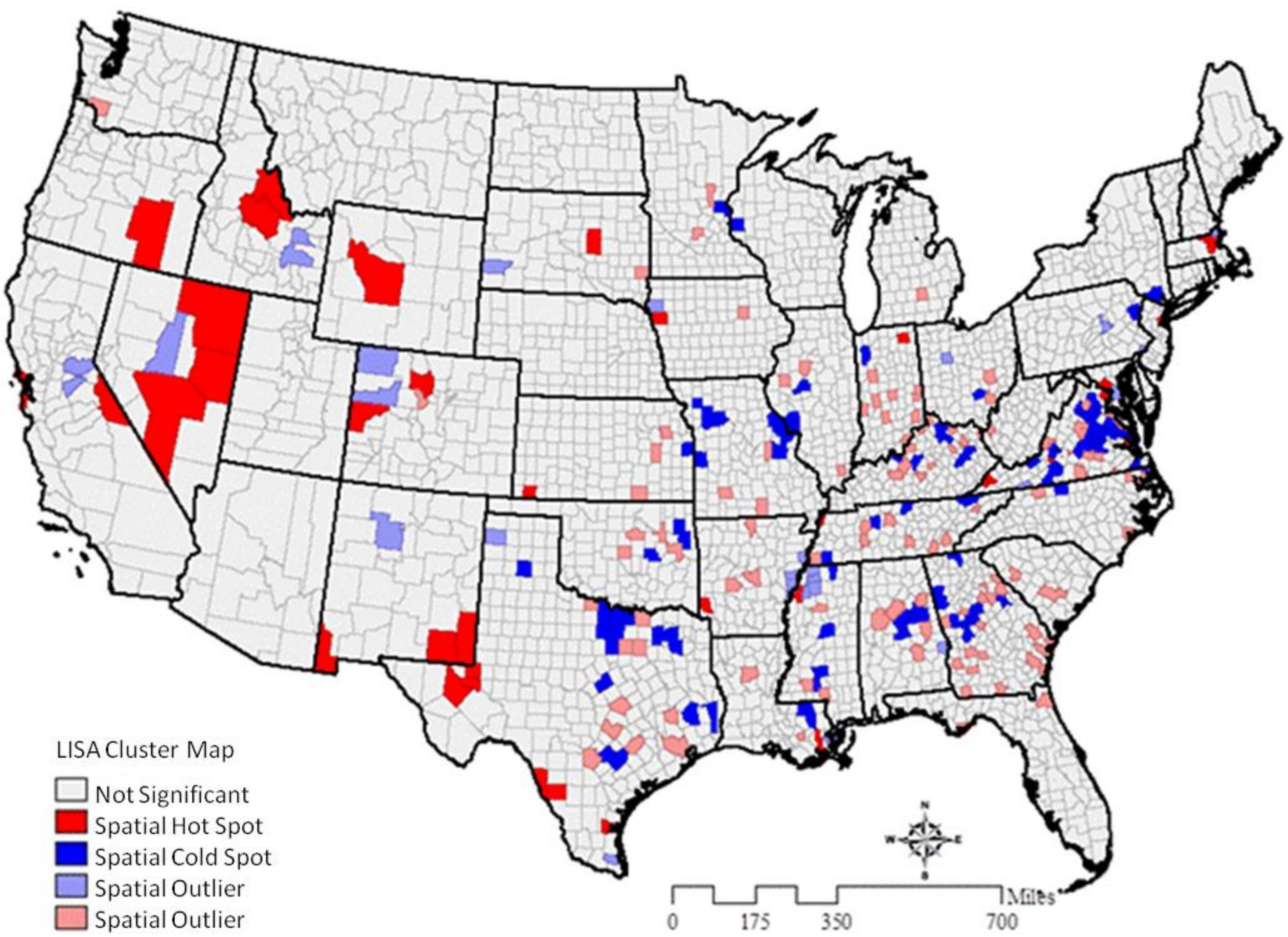

Figure 4.42 Univariate LISA cluster map for county level ratio of jobs to employed residents (JERs) in the US for 2010

4.3.2.5 Population-Weighted Distance (PWD) to Parks

Figure 4.43 presents the results of the univariate Moran's I statistic for the county-level PWD to parks in the US for the year 2008. The univariate Moran's I statistic indicates a positive spatial autocorrelation. The result of the Moran's I statistic is 0.64 in 2008 and based on the 999 permutations run, it is significant ($p<0.001$). This result shows that the high and/or low PWDs to parks are highly clustered at the county level in the US.

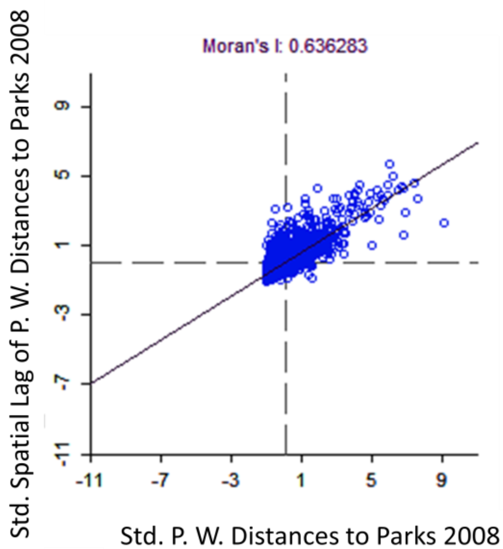

Figure 4.43 Univariate Moran's I scatter plots for county level population-weighted distance (PWD) to parks in the US for 2008

Figure 4.44 illustrates the mapped and categorized PWD to parks based on the results of the univariate LISA statistics ($p< 0.05$).

4.3.2.6 Population Density

Figure 4.45 presents the results of the univariate Moran's I statistics for the county-level population densities in the US for the years 2004, 2007, and 2010. In all years, the univariate Moran's I statistics indicate positive spatial autocorrelation. The results of the Moran's I statistics are 0.66 in 2004 and 0.65 in 2007 and in 2010 and based on the 999 permutations run, they are significant ($p<0.001$). These results reveal that there is continued spatial clustering of

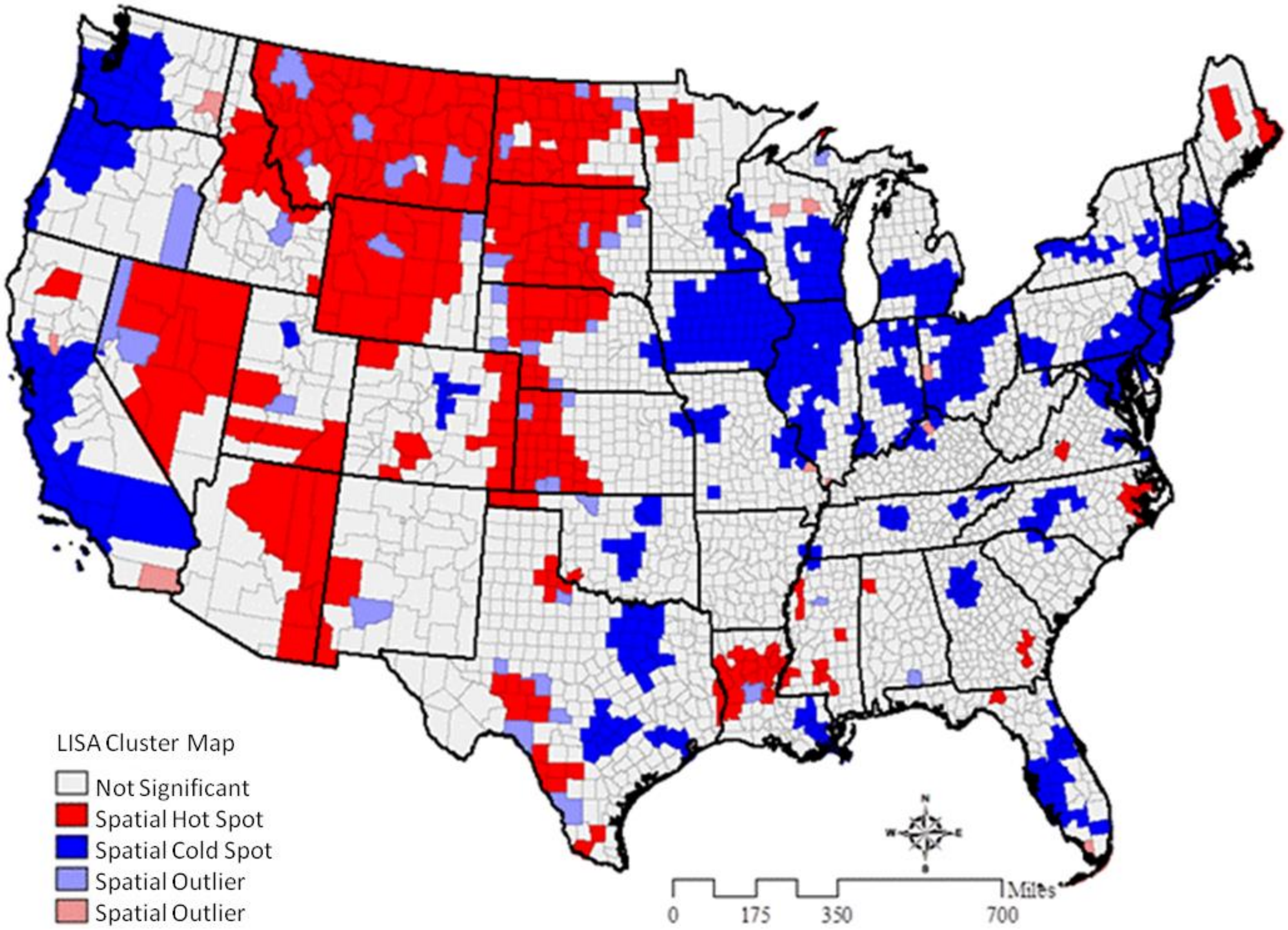

Figure 4.44 Univariate LISA cluster map for county level population-weighted distances (PWDs) to parks in the US for 2008

the high and/or low population densities at the county level in the US. The statistics also indicate that the magnitude of spatial autocorrelation has decreased very slightly during the seven-year observation period (Figure 4.45).

Figures 4.46, 4.47, and 4.48 demonstrate the mapped and categorized population densities based on the results of the univariate LISA statistics ($p<0.05$). In all three years, the number of counties categorized as hot spot population density clusters is lower than the number of counties categorized as cold spot population density clusters.

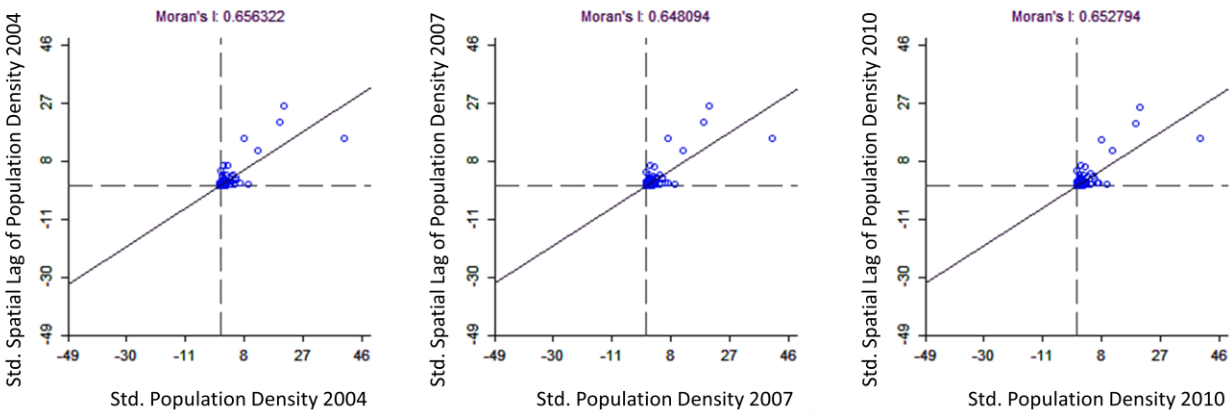

Figure 4.45 Univariate Moran's I scatter plots for county level population densities in the US for 2004, 2007, and 2010

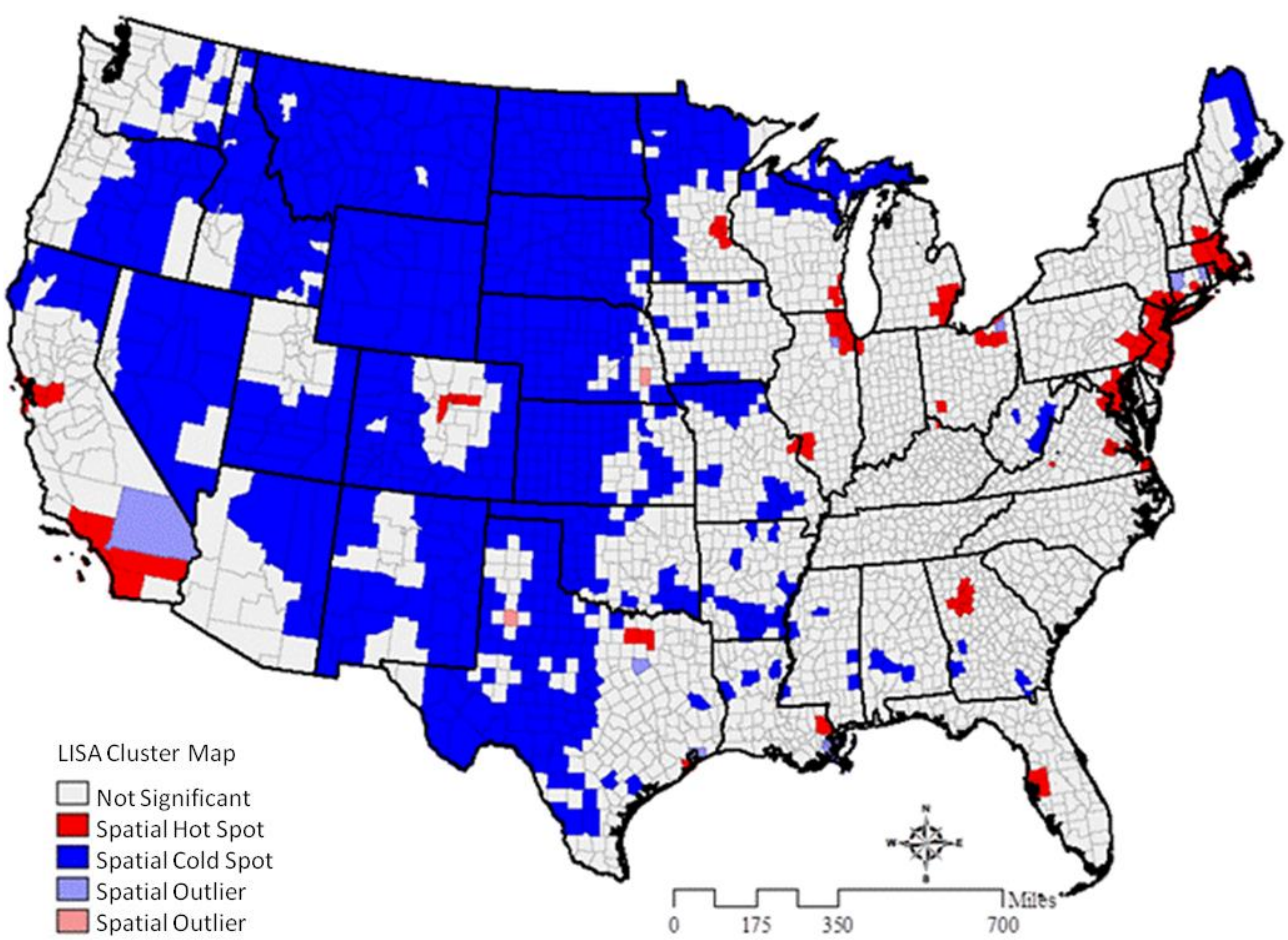

Figure 4.46 Univariate LISA cluster map for county level population densities in the US for 2004

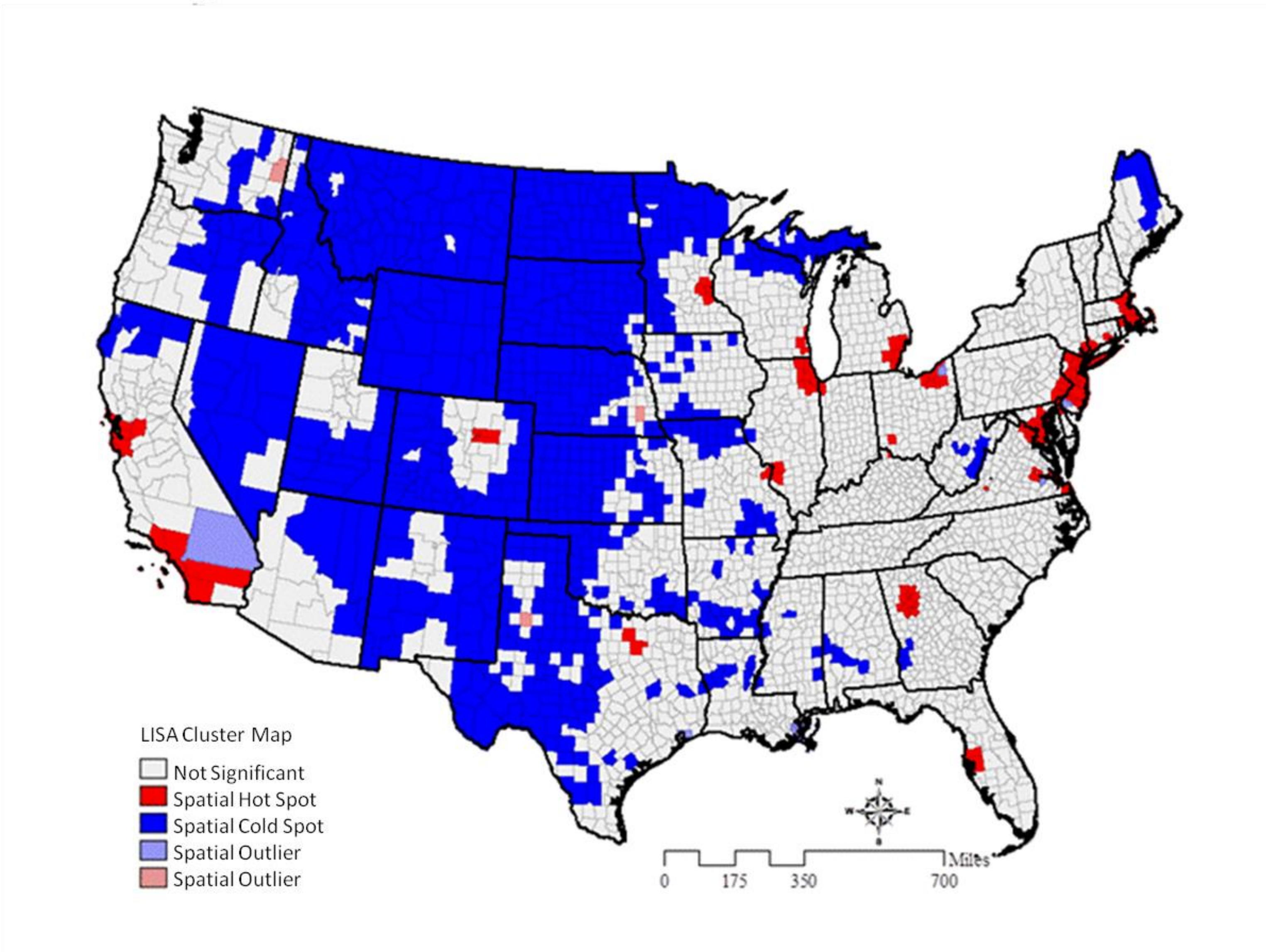

Figure 4.47 Univariate LISA cluster map for county level population densities in the US for 2007

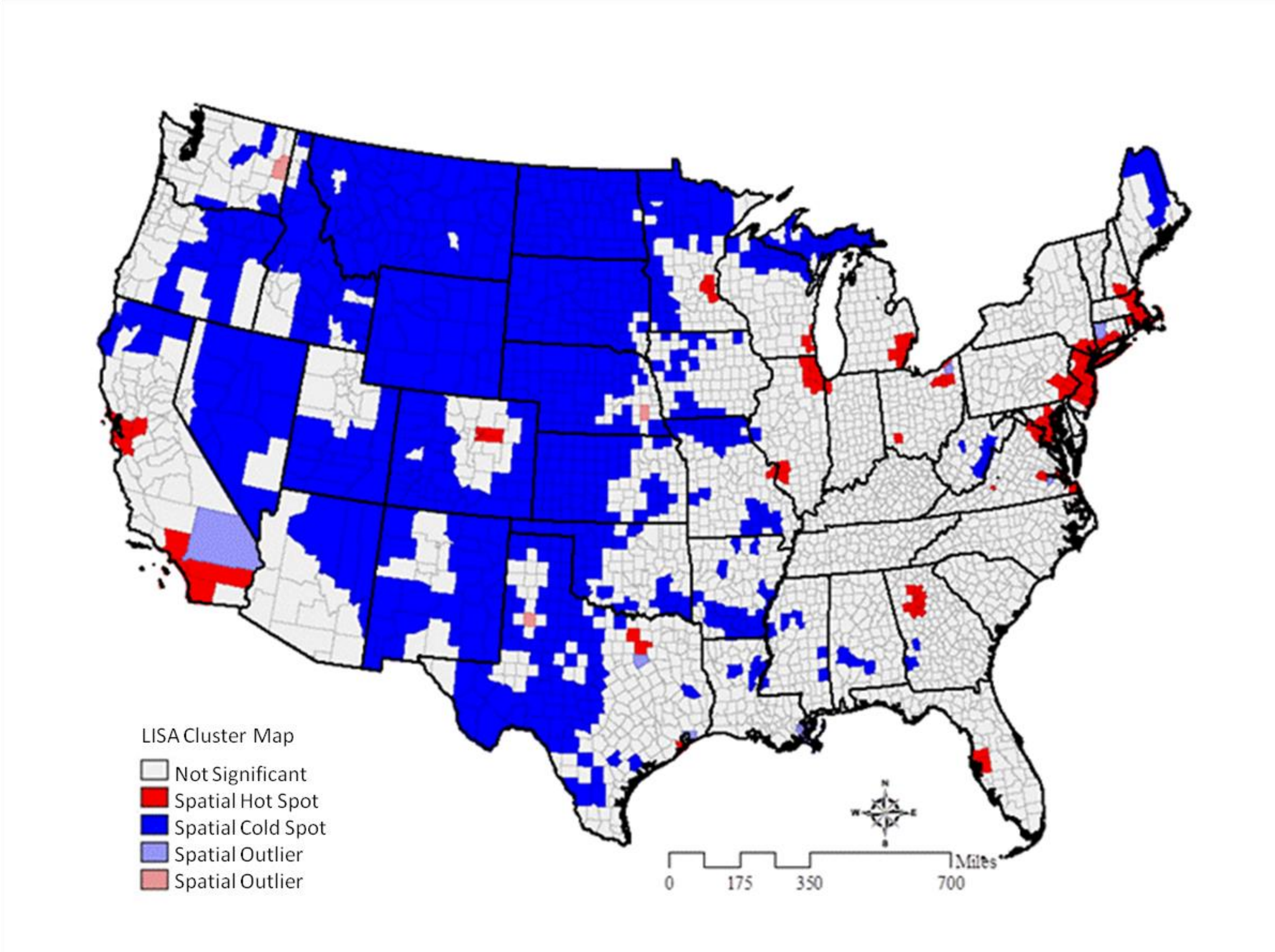

Figure 4.48 Univariate LISA cluster map for county level population densities in the US for 2010

4.3.3 Bivariate Moran's I and LISA Results

To further test the hypothesis that the clustering of the obesity prevalence rates is spatially related to the clustering of the rates of the other (independent) variables, the bivariate global Moran's I and LISA statistics are calculated in GeoDa. Therefore, the following findings are based on the bivariate global Moran's I and LISA statistics. In the bivariate form of the Moran's I and/or LISA, the tests compare the value of a variable at any one location with the value of a second variable at all other locations. In other words, unlike in the univariate form, the observed values of a variable, y, are regressed with the spatial lag of the observed values of another variable, x, in the bivariate form.

Similar to the univariate tests, the bivariate test results are shown in the forms of Moran scatter plots and cluster maps. In the cluster maps, the hot spot (areas in red) and cold spot (areas in blue) locations indicate clustering of similar values while the two types of spatial outlier (areas in lighter red and in lighter blue) locations illustrate dissimilar values that are near each other. Only the statistically significant patterns are visualized on the cluster maps with these colors ($p<0.05$), while insignificant spatial associations are shown in light grey. In addition, the corresponding significance maps are presented in Appendix B.

Initially, scatter plots are created to show the correlation between the obesity rates and each independent variable. Scatter plots show both the type and magnitude of the relationship that exist between the two variables as shown in Table 4.7. In all three years, obesity has a positive association with PIA, poverty, and PWD to parks while it has a negative association with JER and population density. Obesity has clearly stronger relationships with PIA and poverty based on the correlation coefficients observed.

Table 4.7 Correlation coefficients between county level obesity prevalence rates and each independent variable in the US for 2004, 2007, and 2010

	2004	2007	2010
Obesity-Physical Inactivity (PIA)	0.68	0.69	0.70
Obesity-Poverty	0.50	0.47	0.45
Obesity- Ratio of Jobs to Employed Residents (JER)	-0.11	-0.13	-0.13
Obesity- Population-Weighted Distance (PWD) to Parks	0.09	0.08	0.10
Obesity-Population Density	-0.13	-0.13	-0.15

4.3.3.1 Spatial Association between Obesity and Physical Inactivity (PIA)

In Figure 4.49, the results of the bivariate Moran's I statistics for the county-level obesity rates and PIA rates in the US for the years 2004, 2007, and 2010 are presented. Similar to the univariate statistics, these results not only quantify the degree of spatial clustering (spatial autocorrelation) in each year, but also show how the magnitude of spatial clustering (spatial autocorrelation) varied among the years studied. In all three years, the bivariate Moran's I statistics indicate high positive spatial autocorrelation. The results of the Moran's I statistics are 0.55 (2004), 0.54 (2007), and 0.52 (2010) and based on the 999 permutations run, they are significant ($p<0.001$). These results reveal that there is continued spatial clustering of high and/or low rates at the county level in the US. However, the statistics also indicate that the magnitude of spatial autocorrelation has slightly decreased during the seven-year observation period. This decline suggests that the spatial autocorrelation between the obesity rates and PIA rates in some counties and their neighbors have become less spatially concentrated over the years. This is applicable to both hot spot and/or cold spot cluster locations.

Figures 4.50, 4.51, and 4.52 illustrate the mapped and categorized rates (i.e., local associations between obesity rates and lagged PIA rates) based on the results of the bivariate LISA statistics ($p<0.05$) for the period 2004, 2007, and 2010. In all years, the number of counties categorized as hot spot clusters exceeds the number of counties categorized as cold spot clusters.

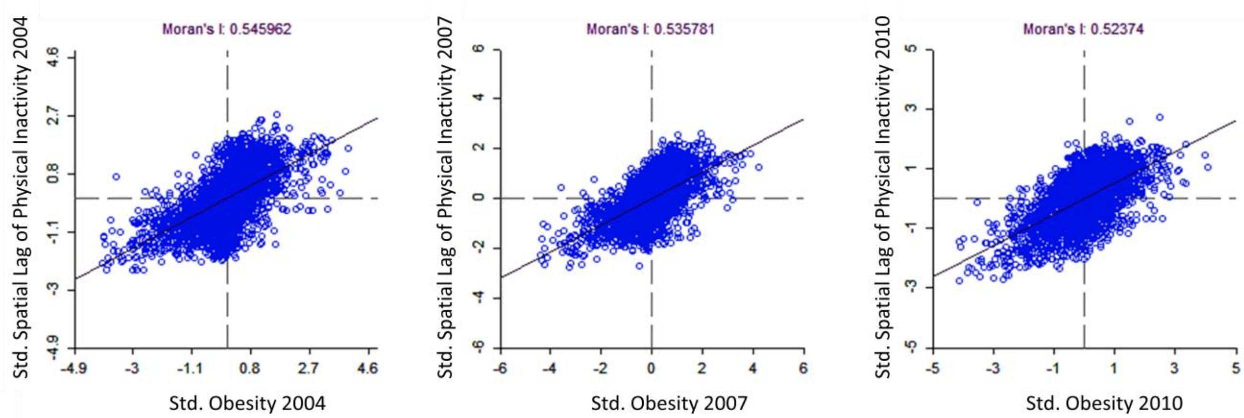

Figure 4.49 Bivariate Moran's I scatter plots for county level obesity prevalence rates and physical inactivity (PIA) prevalence rates in the US for 2004, 2007, and 2010

In all three years, a distinct cluster of high obesity with high physical inactivity (PIA) is observed in the Southern US, especially the South Central region. The high obesity and PIA clusters especially exist in Oklahoma, Arkansas, Louisiana, Mississippi, Alabama, Tennessee, Kentucky, and West Virginia. In addition, Georgia, South Carolina, and North Carolina in the South East (Atlantic) region exhibit a cluster of high obesity and PIA in these three years. There is not a distinct clustering of high obesity and PIA in Texas, except in 2007. Furthermore, over this time period, a growing cluster of high obesity and PIA in Missouri, South Dakota, and North Dakota in the West North Central region in the Midwest are observed. On the other hand, there is a diminishing high obesity and PIA cluster in the South East (Atlantic) region of the South, particularly in Georgia from 2004 to 2010 (Figures 4.50, 4.51, and 4.52).

In all years, a constant and distinct cluster of low obesity and PIA is observed in Washington, Oregon, and California in the Pacific region in the West. Another distinct cluster of low obesity and PIA is apparent in the Mountain region in the West despite the fact that there is a slight decrease in the number of clusters in Arizona and New Mexico from 2004 to 2010 period.

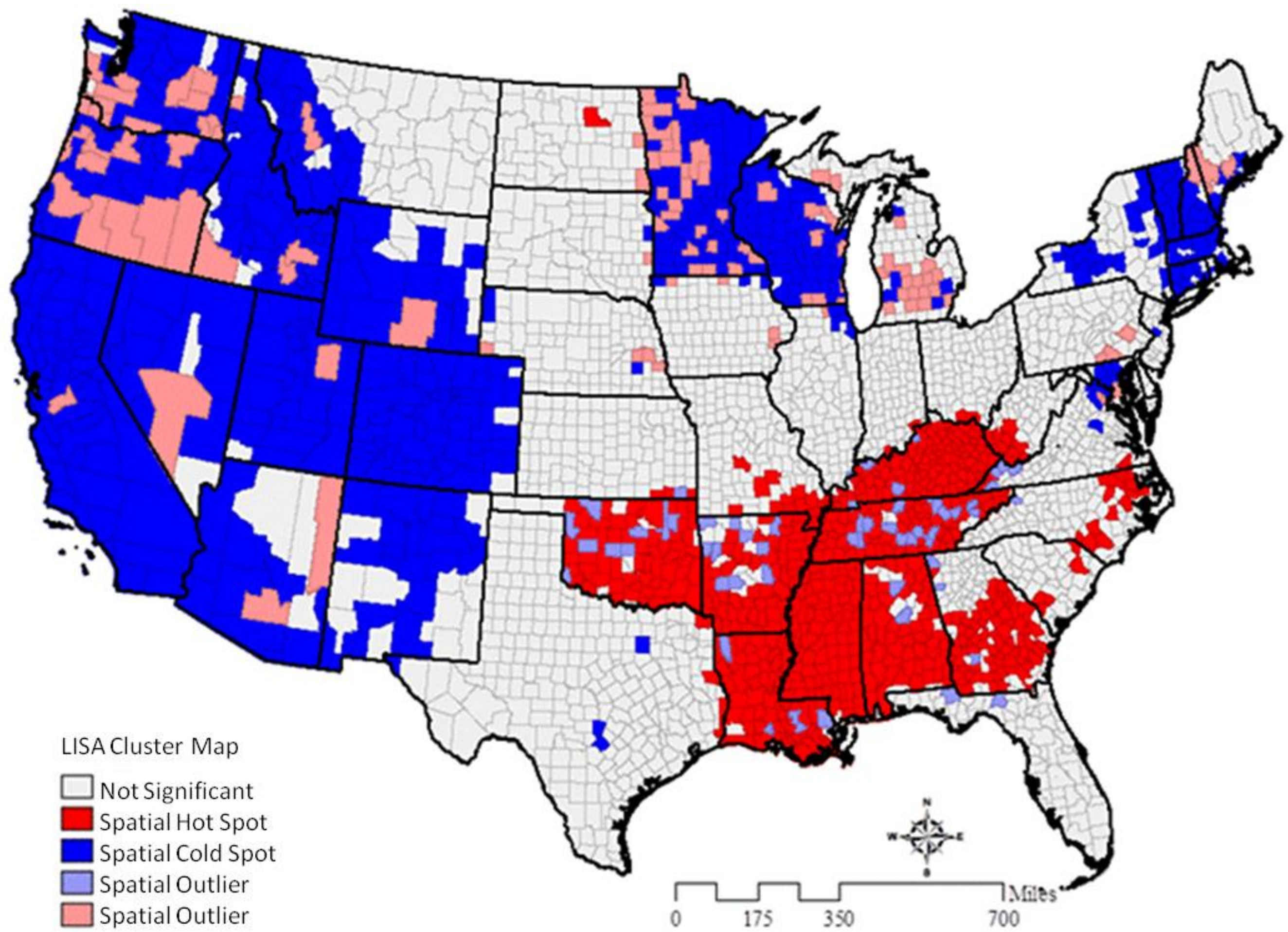

Figure 4.50 Bivariate LISA cluster map for county level obesity prevalence rates and physical inactivity (PIA) prevalence rates in the US for 2004

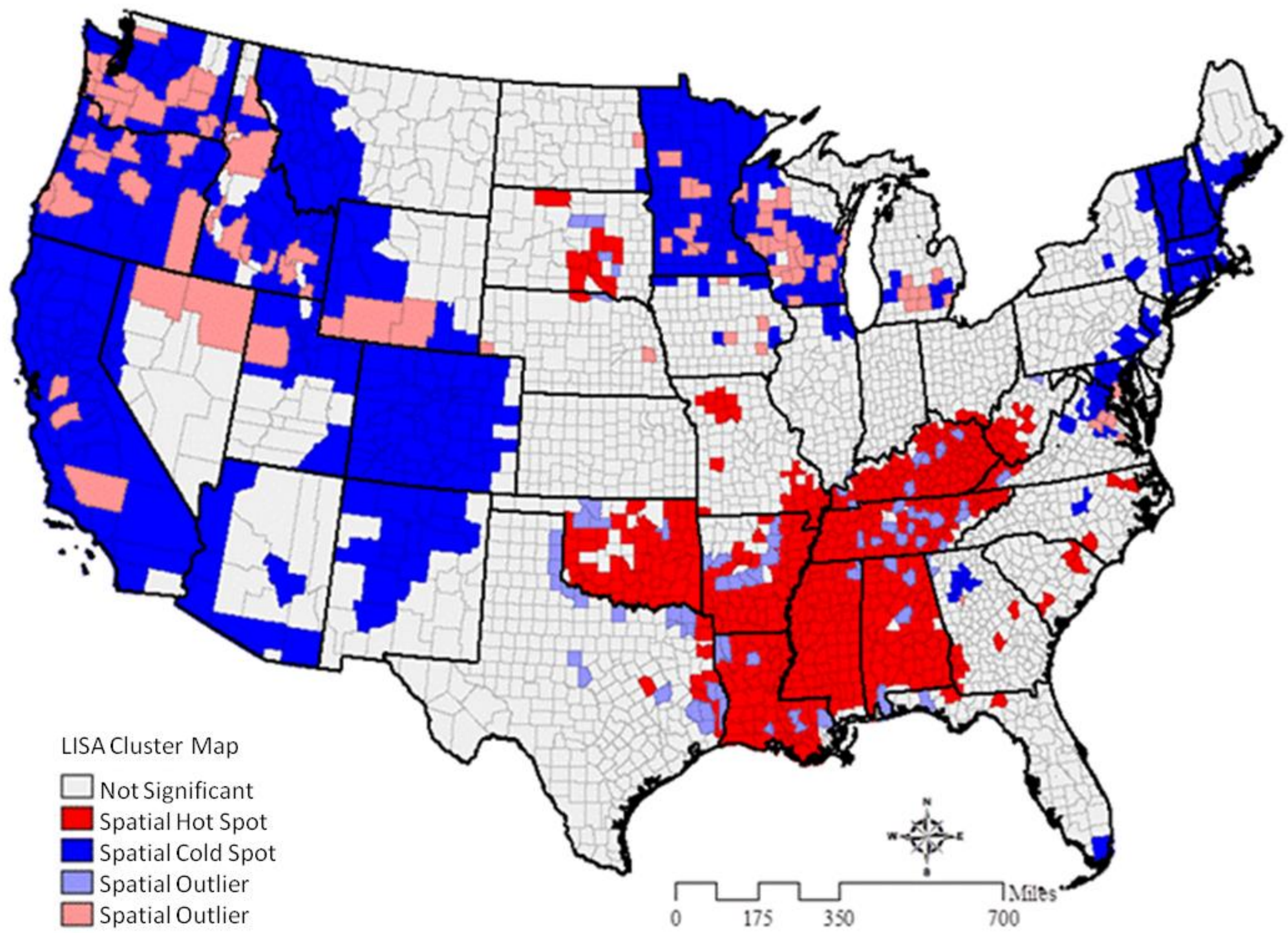

Figure 4.51 Bivariate LISA cluster map for county level obesity prevalence rates and physical inactivity (PIA) prevalence rates in the US for 2007

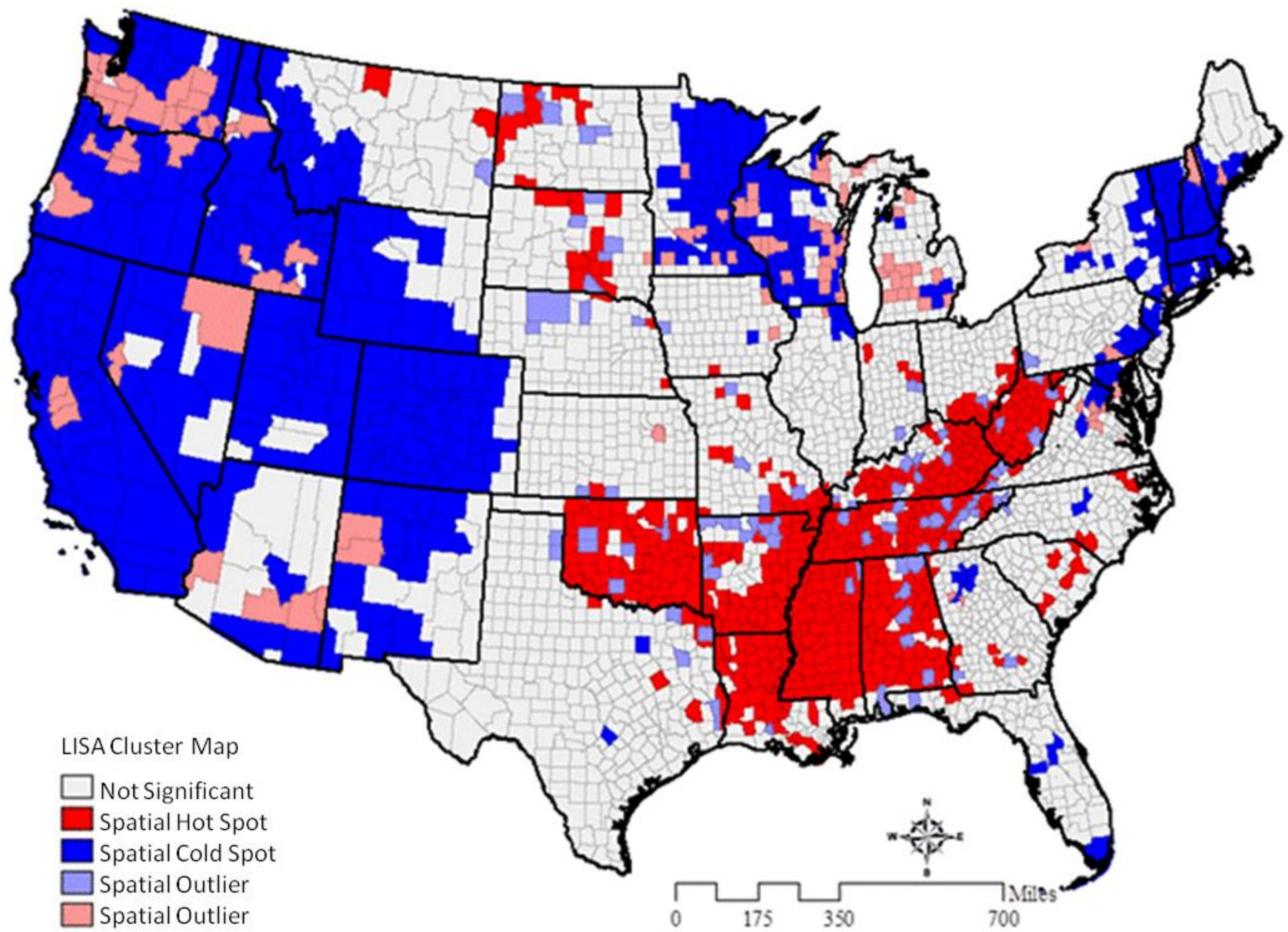

Figure 4.52 Bivariate LISA cluster map for county level obesity prevalence rates and physical inactivity (PIA) prevalence rates in the US for 2010

Over the same time period, a distinct cluster of low obesity and PIA is also observed in Minnesota, Wisconsin, and Michigan in the Midwest. In addition, there is another distinct low obesity and PIA clustering shown in the Northeast, especially in the New England region. Moreover, from 2004 to 2010, a fairly distinct cluster of low obesity and PIA is apparent in Virginia and Maryland in the South. While the South US is dominated by clusters of high obesity and PIA areas, growing clusters of low obesity and PIA areas are also observed in Virginia, Georgia, and Florida in the South East (Atlantic) region of the South from 2004 to 2010 (Figures 4.50, 4.51, and 4.52).

On the other hand, in all three years, high obesity with low PIA and low obesity with high PIA spatial outliers exist throughout the US. The high obesity and low PIA spatial outliers are mostly found in the West and Midwest. Washington, Oregon, and California in the West and the states of Minnesota, Wisconsin, and Michigan stand out having the most outliers in all three years. The low obesity and high PIA spatial outliers are mainly observed in Texas, Oklahoma, Arkansas, Alabama, Tennessee, and Kentucky in the South in all three years and in Nebraska, South Dakota, and North Dakota in the Midwest, particularly in 2010 (Figures 4.50, 4.51, and 4.52).

4.3.3.2 Spatial Association between Obesity and Poverty

In Figure 4.53, the results of the bivariate Moran's I statistics for the county-level obesity rates and poverty rates in the US for the years 2004, 2007, and 2010 are presented. In all three years, the bivariate Moran's I statistics indicate positive spatial autocorrelation. The results of the Moran's I statistics are 0.36 (2004), 0.33 (2007), and 0.31 (2010) and based on the 999 permutations run, they are significant ($p<0.001$). These results reveal that there is continued spatial clustering of high and/or low rates at the county level in the US. However, the statistics also indicate that the magnitude of spatial autocorrelation has decreased slightly during the seven-year observation period. This decline suggests that the spatial autocorrelation between the obesity rates and poverty rates in some counties and their neighbors have become less concentrated over the years. This is applicable to both hot spot and/or cold spot cluster locations.

Figure 4.53 Bivariate Moran's I scatter plots for county level obesity prevalence rates and poverty prevalence rates in the US for 2004, 2007, and 2010

Figures 4.54, 4.55, and 4.56 illustrate the mapped and categorized rates (i.e., local associations between obesity rates and lagged poverty rates) based on the results of the bivariate LISA statistics ($p<0.05$) for the period 2004, 2007, and 2010. In all years, the number of counties categorized as hot spot clusters is lower than the number of counties categorized as cold spot clusters.

In all three years, there is a distinctive north–south divide across the US in which the Southern US is dominated by high obesity with high poverty clusters, while the Northern US is dominated by low obesity with low poverty clusters. More specifically, the high obesity and poverty clusters that are existent within the South are mostly observed in Texas, Louisiana, Mississippi, Alabama, Georgia, Florida, South Carolina, North Carolina, Virginia, West Virginia, Kentucky, Tennessee, Arkansas, and Oklahoma in the South and in New Mexico and Arizona in the West and in Missouri in the Midwest. South Dakota, located in the West North

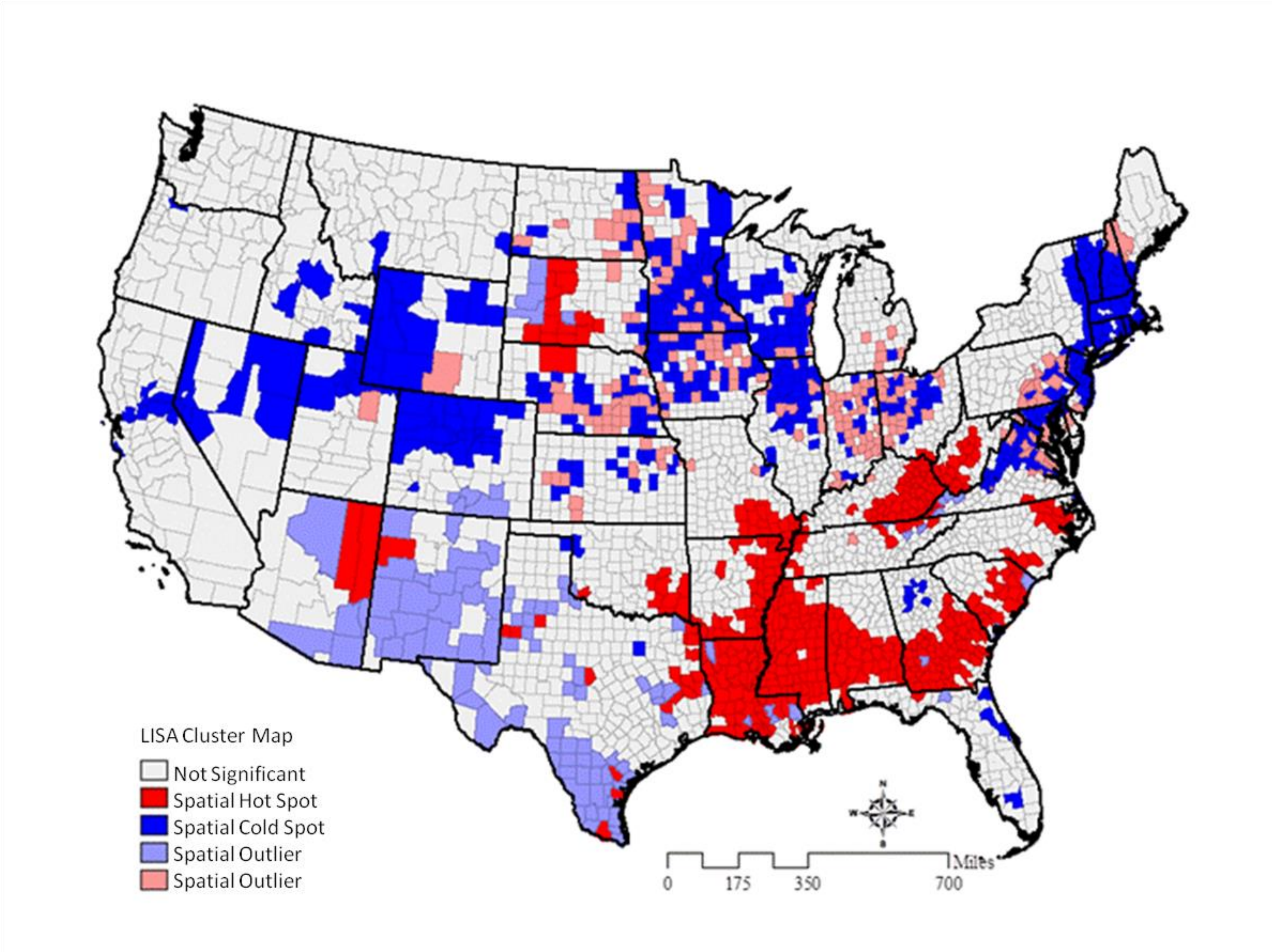

Figure 4.54 Bivariate LISA cluster map for county level obesity prevalence rates and poverty prevalence rates in the US for 2004

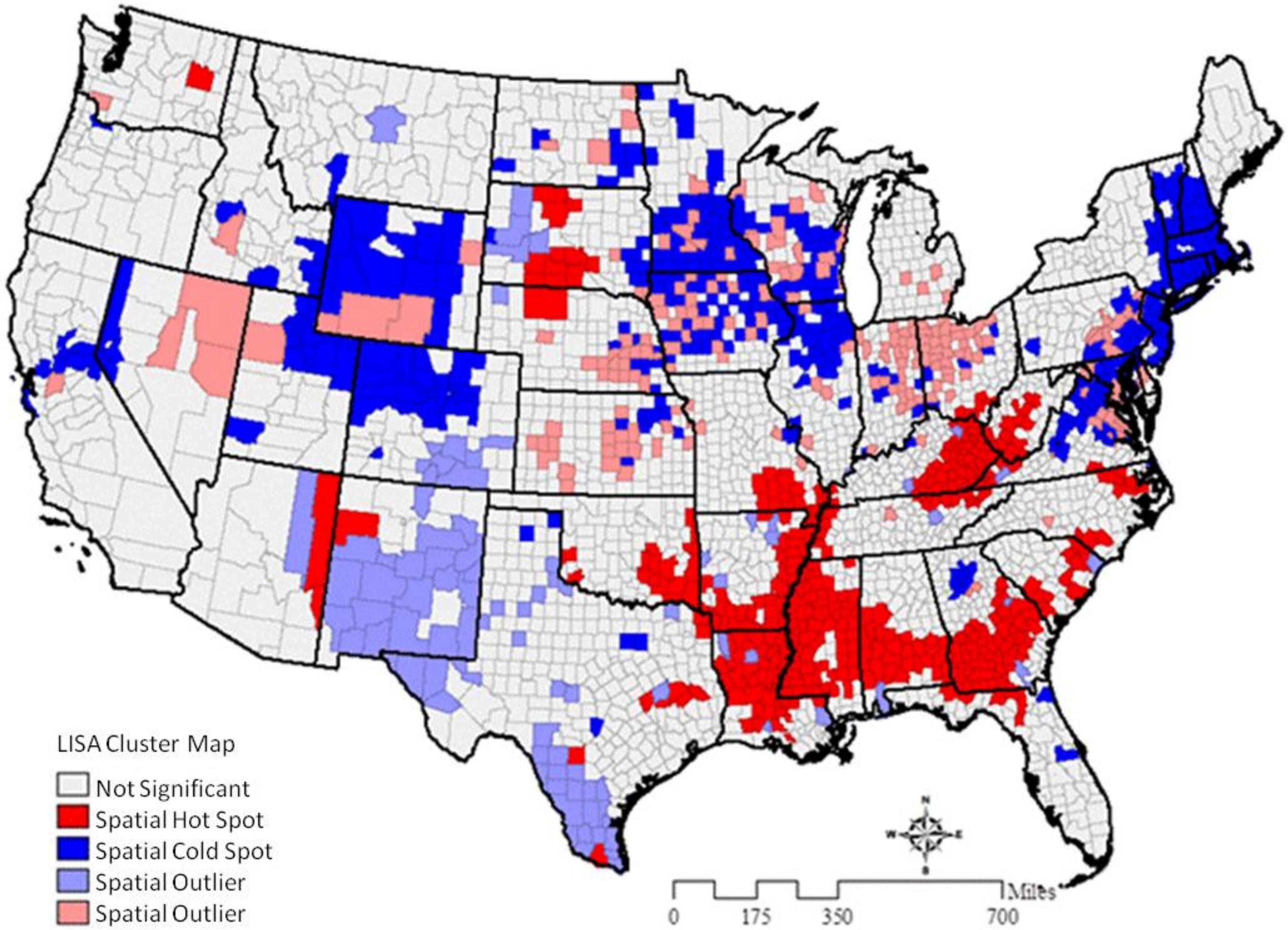

Figure 4.55 Bivariate LISA cluster map for county level obesity prevalence rates and poverty prevalence rates in the US for 2007

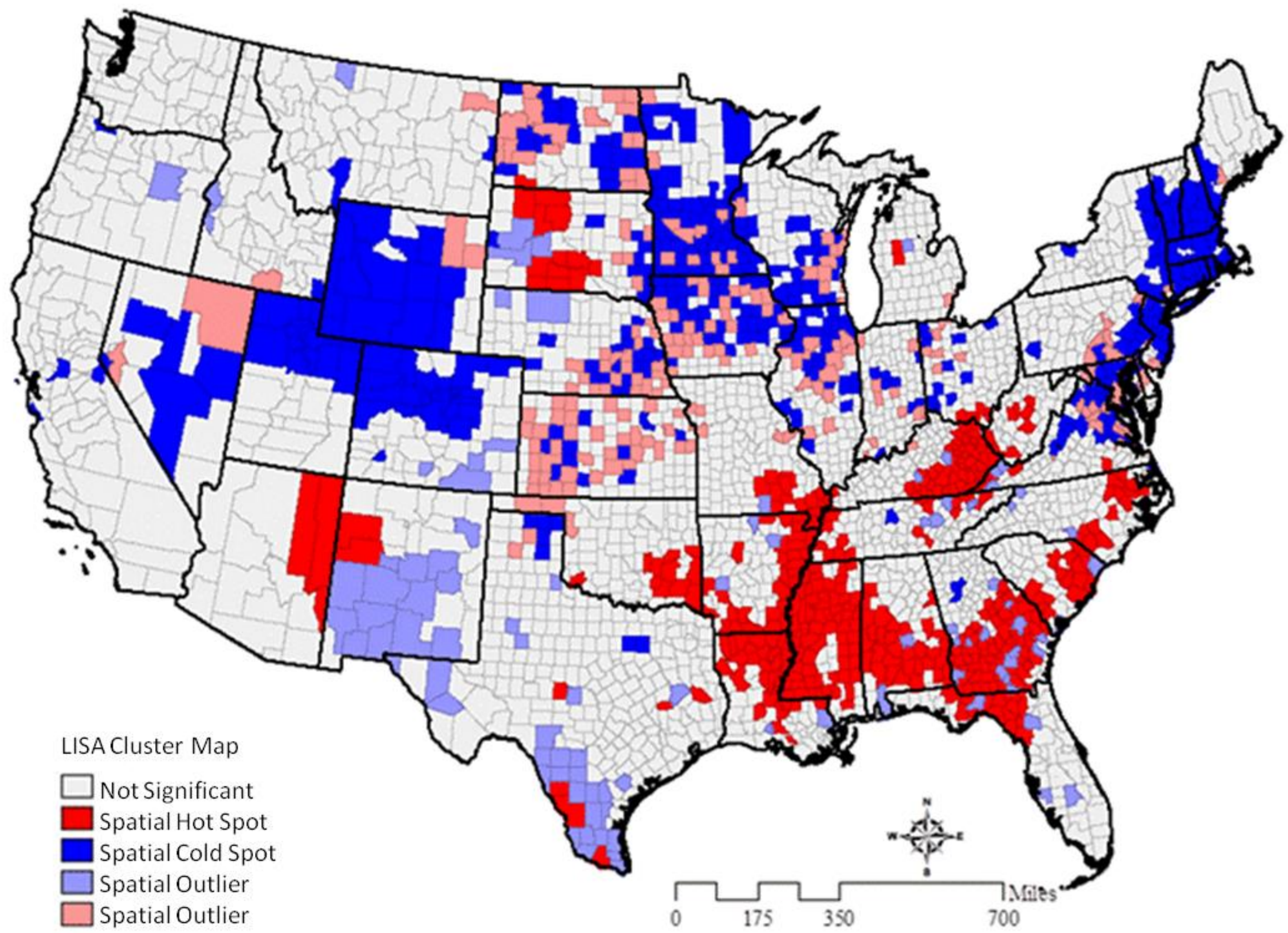

Figure 4.56 Bivariate LISA cluster map for county level obesity prevalence rates and poverty prevalence rates in the US for 2010

Central region has the only distinctive high obesity and poverty clusters within the abovementioned divide. In addition, there is a growing cluster of high obesity and poverty in Florida in these periods, whereas the high obesity and poverty clusters are slightly diminished in Texas from 2004 to 2010. Nonetheless, the majority of the high obesity and poverty clusters are confined to the South Central and South East (Atlantic) regions in the Southern US from 2004 to 2010 (Figures 4.54, 4.55, and 4.56).

Over this time period, distinct clusters of low obesity and poverty are observed in the abovementioned divide across the US. These clusters extend from Maine in the Northeastern US to California in the far Western US, with low obesity and poverty clusters in every state except in Oregon along this path. Based on Figures 4.54, 4.55, and 4.56, Georgia, Virginia, Maryland, and Delaware have the only distinctive low obesity and poverty clusters in the Southern US over this time period.

On the other hand, the high obesity and low poverty spatial outliers are mostly confined to the Midwest yet existent to a certain level in the Mountain region of the West. Pennsylvania in the Northeast and Delaware, Maryland, and Virginia in the South have also distinct high obesity and low poverty spatial outliers. Distinctive low obesity and high poverty spatial outliers are mainly observed in Texas, New Mexico, Arizona, Colorado, and South Dakota (Figures 4.54, 4.55, and 4.56).

4.3.3.3 Spatial Association between Obesity and Ratio of Jobs to Employed Residents (JER)

Figure 4.57 presents the results of the bivariate Moran's I statistics for the county-level obesity rates and JERs in the US for the years 2004, 2007, and 2010. In all three years, the bivariate Moran's I statistics indicate negative spatial autocorrelation. The results of the Moran's I statistics are -0.07 (2004), -0.07 (2007), and -0.06 (2010) and based on the 999 permutations run, they are significant (p<0.001). These results reveal that there is not spatial clustering of high and/or low values at the county level in the US. The statistics also indicate that the magnitude of negative spatial autocorrelation remained nearly constant during the seven-year observation period.

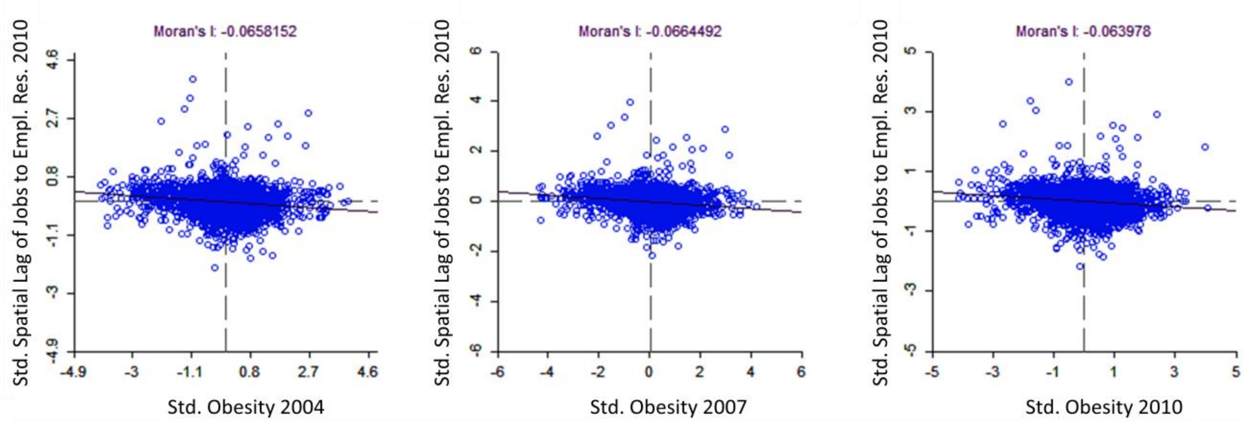

Figure 4.57 Bivariate Moran's I scatter plots for county level obesity prevalence rates and ratio of jobs to employed residents (JERs) in the US for 2004, 2007, and 2010

Figures 4.58, 4.59, and 4.60 illustrate the mapped and categorized rates (i.e., local associations between obesity rates and lagged JERs) based on the results of the bivariate LISA statistics (p<0.05) for 2004, 2007, and 2010. In all years, the number of counties categorized as hot spot clusters is lower than the number of counties categorized as cold spot clusters. In addition, the number of counties categorized as spatial outliers exceeds the number of counties categorized spatial clusters.

Overall, the spatial association between obesity and JER is dispersed at the county level in all three years. However, as observed in the LISA cluster maps, spatial clustering of high obesity with high JER and low obesity with low JER still exists to a certain degree. In all three years, the clusters of high obesity and JER are distinctively observed in Louisiana and Mississippi in the South, in Pennsylvania in the Northeast, and in Nevada in the West. Texas

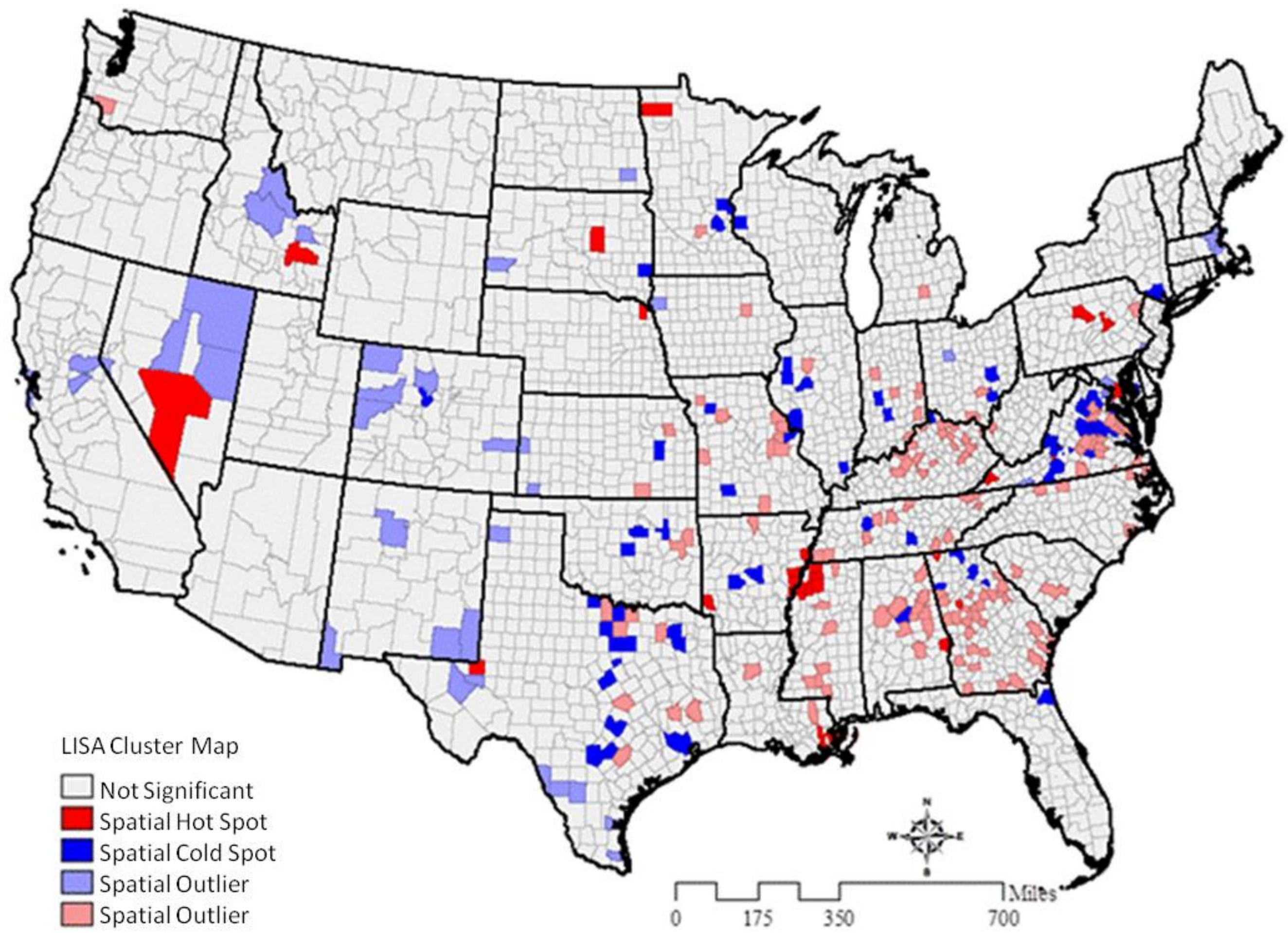

Figure 4.58 Bivariate LISA cluster map for county level obesity prevalence rates and ratio of jobs to employed residents (JERs) in the US for 2004

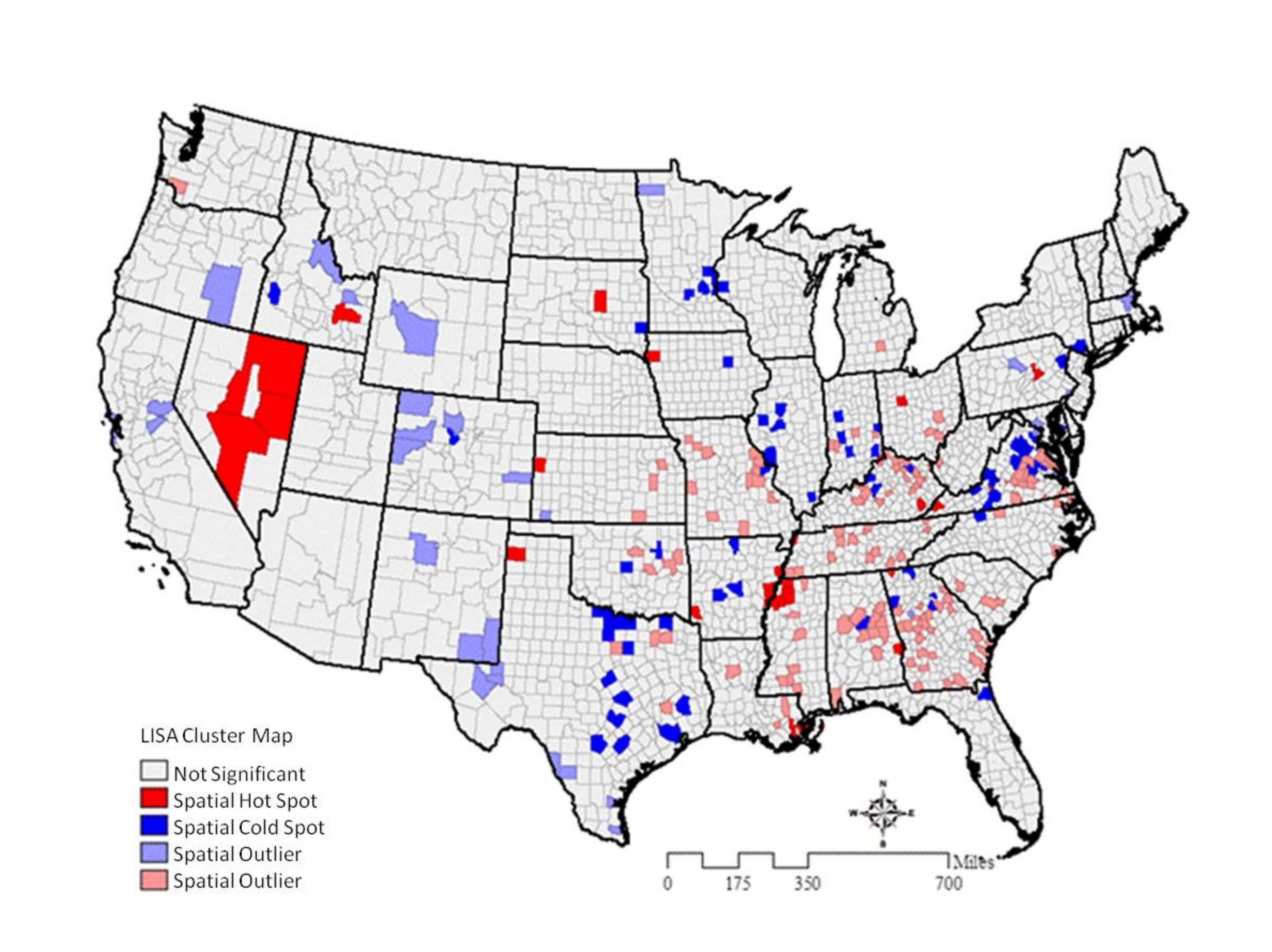

Figure 4.59 Bivariate LISA cluster map for county level obesity prevalence rates and ratio of jobs to employed residents (JERs) in the US for 2007

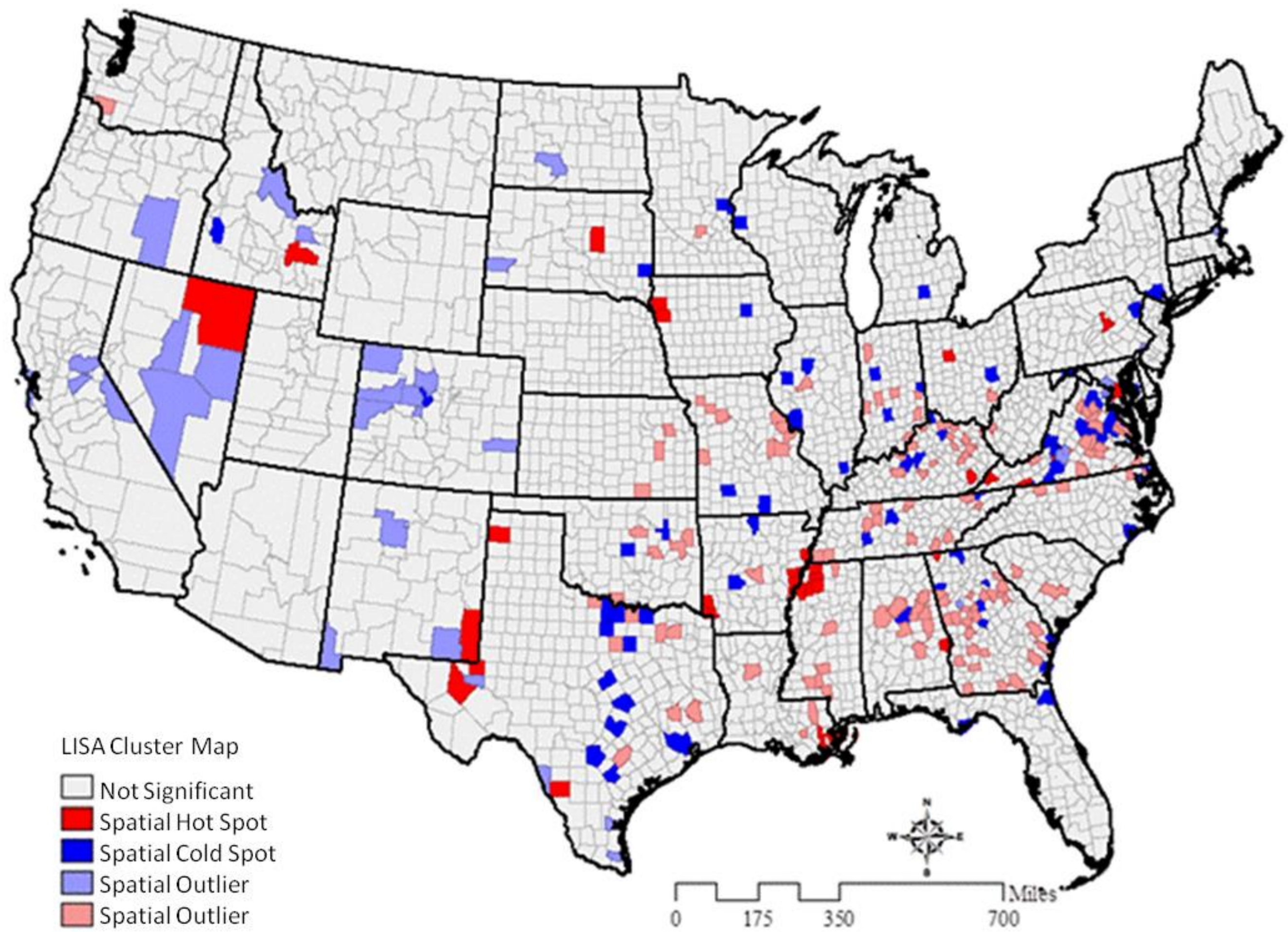

Figure 4.60 Bivariate LISA cluster map for county level obesity prevalence rates and ratio of jobs to employed residents (JERs) in the US for 2010

and Virginia in the South and Illinois, Indiana, and Minnesota in the Midwest stand out having the most of the low obesity and JER clusters (Figures 4.58, 4.59, and 4.60).

On the other hand, the high obesity and low JER spatial outliers are mainly observed in the Southern US while low obesity and high JER spatial outliers are mainly observed in the Western US in all three years. More specifically, the high obesity and low JER spatial outliers are seen in Texas, Louisiana, Mississippi, Alabama, Georgia, South Carolina, North Carolina, Virginia, Kentucky, Tennessee, and Oklahoma in the South while the low obesity and high JER spatial outliers are observed in New Mexico, Colorado, Idaho, Nevada, and California in the West and in Texas in the South US. In addition, high obesity and low JER spatial outliers are also detected in Kansas, Missouri, Indiana, and Ohio in the Midwest. Overall for both the high obesity and low JER and low obesity and high JER spatial outliers, the observed spatial outliers in California, Nevada, Colorado, New Mexico in the West, in Texas, Mississippi, Alabama, Georgia, Virginia, Kentucky, and Tennessee in the South, and in Missouri in the Midwest stand out in terms of visual detection on the map (Figures 4.58, 4.59, and 4.60).

4.3.3.4 Spatial Association between Obesity and Population-Weighted Distance (PWD) to Parks

Figure 4.61 presents the results of the bivariate Moran's I statistics for the county-level obesity rates and PWDs to parks in the US for the years 2004, 2007, and 2010. In all three years, the bivariate Moran's I statistics indicate positive spatial autocorrelation. The results of the Moran's I statistics are 0.07 (2004), 0.05 (2007), and 0.07 (2010) and based on the 999 permutations run, they are significant ($p<0.001$). These results reveal that there is continued spatial clustering of high and/or low values at the county level in the US. However, the statistics also indicate that the magnitude of spatial autocorrelation has slightly decreased in between 2004 and 2007 and increased back by 2010. Overall, this decline followed by an increase suggests that the spatial correlation between the obesity rates and PWDs to parks in some counties and their neighbors have become less concentrated by 2007 and then become more concentrated by 2010. This is applicable to both hot spot and/or cold spot cluster locations.

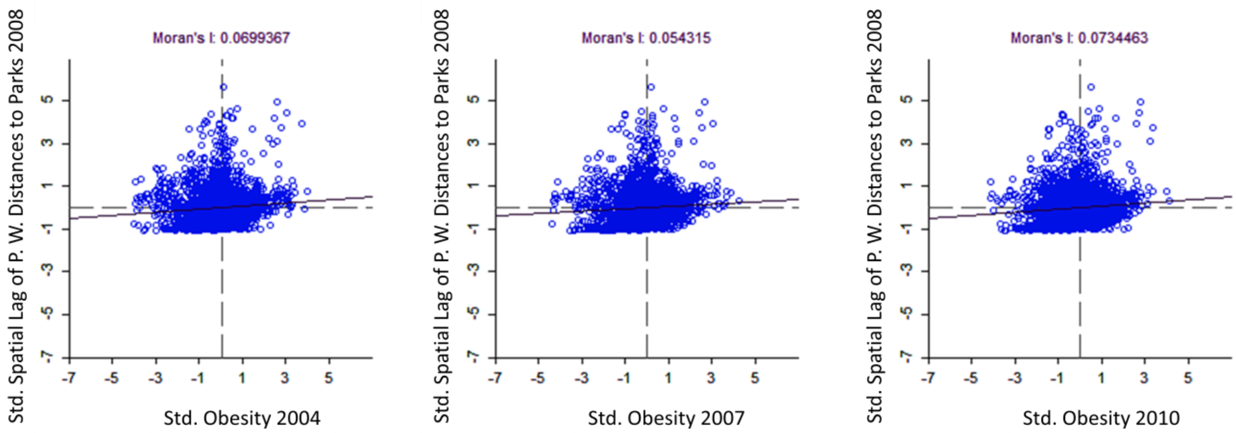

Figure 4.61 Bivariate Moran's I scatter plots for county level obesity prevalence rates and population-weighted distances (PWDs) to parks in the US for 2004, 2007, and 2010

Figures 4.62, 4.63, and 4.64 illustrate the mapped and categorized rates (i.e., local spatial association between obesity rates and lagged PWDs to parks) based on the results of the bivariate LISA statistics ($p < 0.05$) for 2004, 2007, and 2010. In all years, the number of counties categorized as hot spot clusters is lower than the number of counties categorized as cold spot clusters.

Overall, in all three years, a distinct cluster of high obesity with high PWD to parks is mainly observed in the Mountain region of the West and in the West North Central region of the Midwest. More specifically, the area consisting of Nevada, Montana, and Wyoming in the West and North Dakota, South Dakota, and Nebraska in the Midwest stands out having a very distinctive cluster of high obesity and high PWD to parks. The clusters of high obesity and high PWD to parks are also seen in the South, mainly in Texas, Louisiana, Mississippi, Alabama, Georgia, and North Carolina. Maine is the only state with a cluster of high obesity and high PWD to parks in the Northeast (Figures 4.62, 4.63, and 4.64).

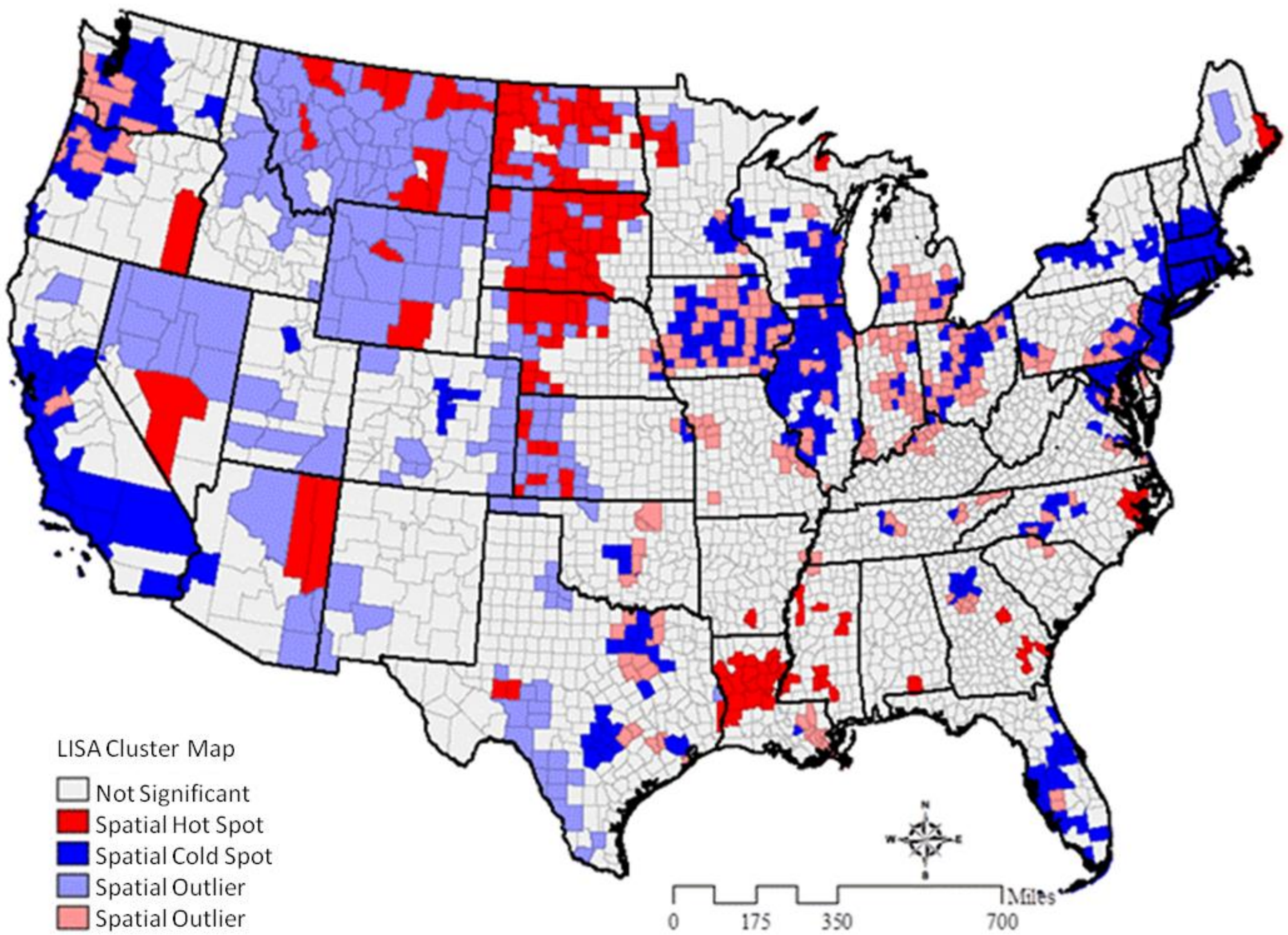

Figure 4.62 Bivariate LISA cluster map for county level obesity prevalence rates and population-weighted distances (PWDs) to parks in the US for 2004

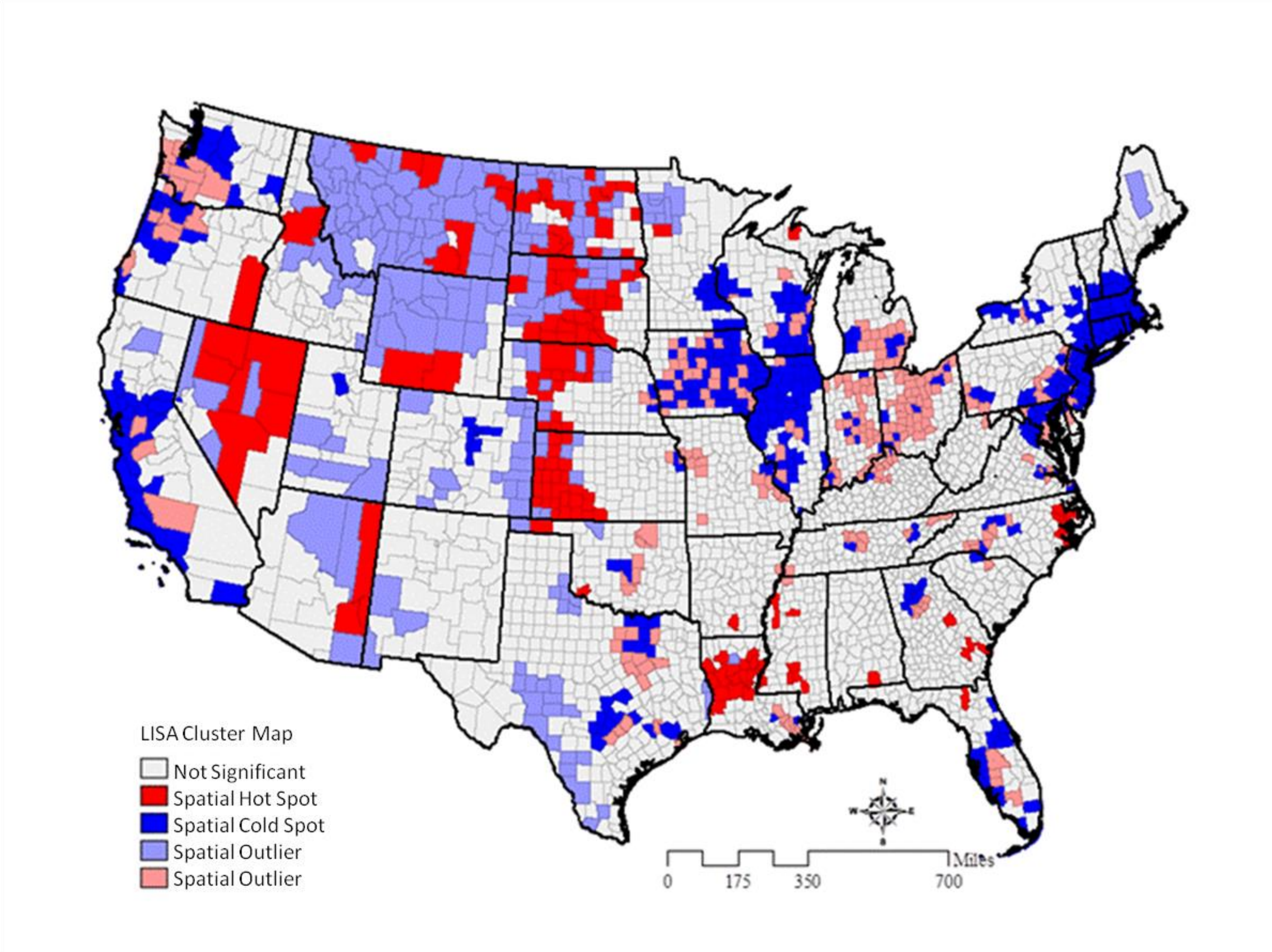

Figure 4.63 Bivariate LISA cluster map for county level obesity prevalence rates and population-weighted distances (PWDs) to parks in the US for 2007

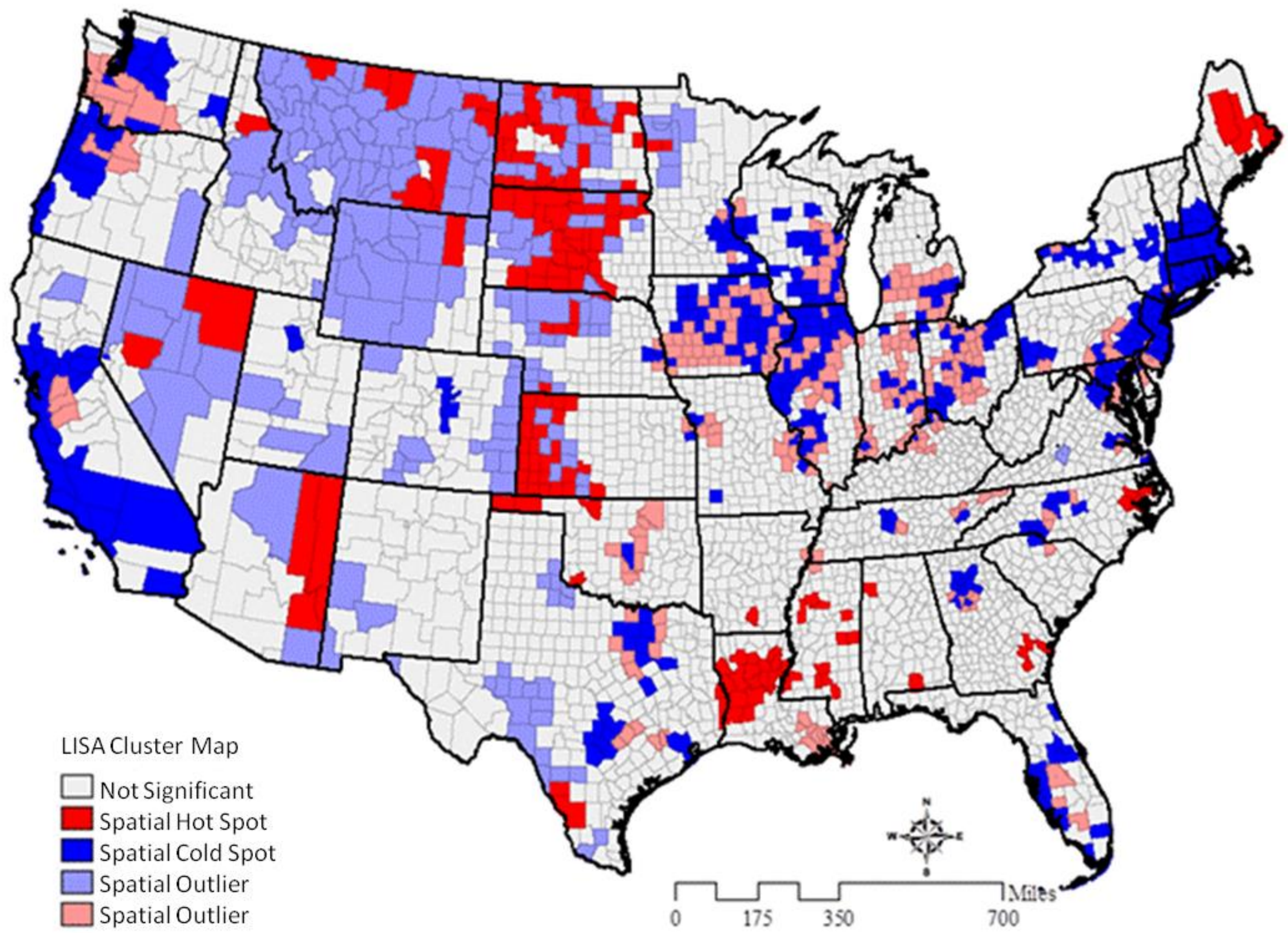

Figure 4.64 Bivariate LISA cluster map for county level obesity prevalence rates and population-weighted distances (PWDs) to parks in the US for 2010

Conversely, in all three years, a fairly distinct cluster of low obesity and low PWD to parks is observed predominantly in the Northeast Megalopolis region that covers the District of Columbia and the states of Virginia, Maryland, Delaware, Pennsylvania, New Jersey, New York, Connecticut, Rhode Island, Massachusetts, Vermont, and New Hampshire. In the Northeast, Maine is the only state that does not have clusters of low obesity and low PWD to parks. Another distinctive cluster of low obesity and low PWD to parks is seen in Minnesota, Iowa, Wisconsin, and Illinois in the East North Central region of the Midwest and in Washington, Oregon, and California in the Pacific region of the West. Texas, Georgia, Florida, North Carolina, Tennessee, and Oklahoma in the South also have fairly distinct cluster of low obesity and low PWD to parks. Overall, the observed patterns reveal that the clusters of high obesity with high PWD to parks are more prevalent in the rural counties and county equivalents while the clusters of low obesity with low PWD to parks are mainly seen in the urban counties and county equivalents (Figures 4.62, 4.63, and 4.64).

On the other hand, the high obesity and low PWD to parks spatial outliers are most distinctively observed in the Midwest, particularly in Wisconsin, Iowa, Missouri, Illinois, Michigan, Indiana, and Ohio. In addition, the high obesity and low PWD to parks spatial outliers are also seen in Pennsylvania and New Jersey in the Northeast, in Texas, Oklahoma, and Louisiana in the South, and in Washington and Oregon in the West. On the contrary, the entire Mountain region of the West, North Dakota, South Dakota, Nebraska, and Kansas in the Midwest, and Texas in the South are covered by the low obesity high PWD to parks spatial outliers (Figures 4.62, 4.63, and 4.64).

4.3.3.5 Spatial Association between Obesity and Population Density

In Figure 4.65, the results of the bivariate Moran's I statistics for the county-level obesity rates and population densities in the US for the years 2004, 2007, and 2010 are presented. In all three years, the bivariate Moran's I statistics indicate negative spatial autocorrelation. The results of the Moran's I statistics are -0.11 (2004), -0.10 (2007), and -0.12 (2010) and based on the 999 permutations run, they are significant ($p<0.001$). These results reveal that the high and/or low values are dispersed at the county level in the US. The statistics also indicate that the magnitude of negative spatial autocorrelation has slightly increased during the seven-year observation period.

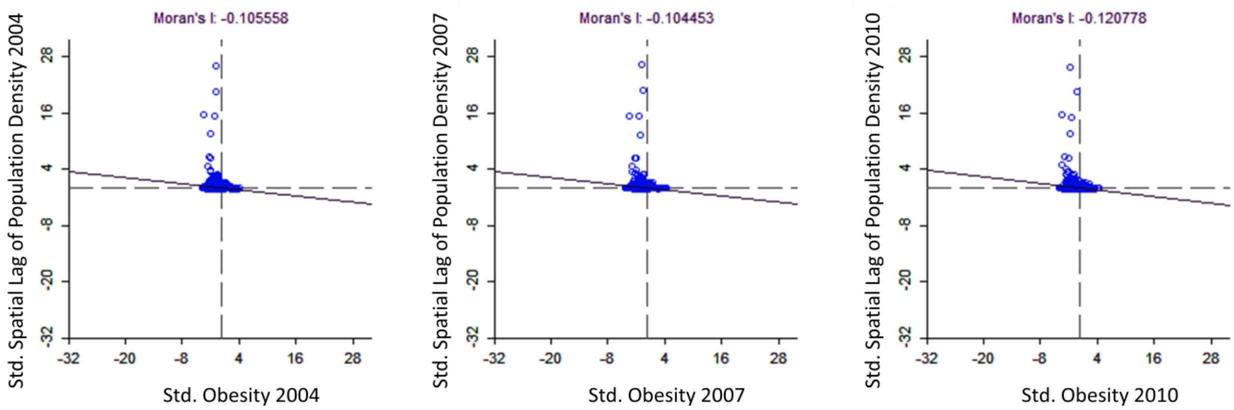

Figure 4.65 Bivariate Moran's I scatter plots for county level obesity prevalence rates and population densities in the US for 2004, 2007, and 2010

Figures 4.66, 4.67, and 4.68 illustrate the mapped and categorized rates (i.e., local associations between obesity rates and lagged population densities) based on the results of the bivariate LISA statistics ($p< 0.05$) for the period 2004, 2007, and 2010. In all years, the number of counties categorized as hot spot clusters is lower than the number of counties categorized as cold spot clusters.

Similar to the results of bivariate global Moran's I for the obesity and JER, the spatial association between obesity and population density is dispersed at the county level in all three

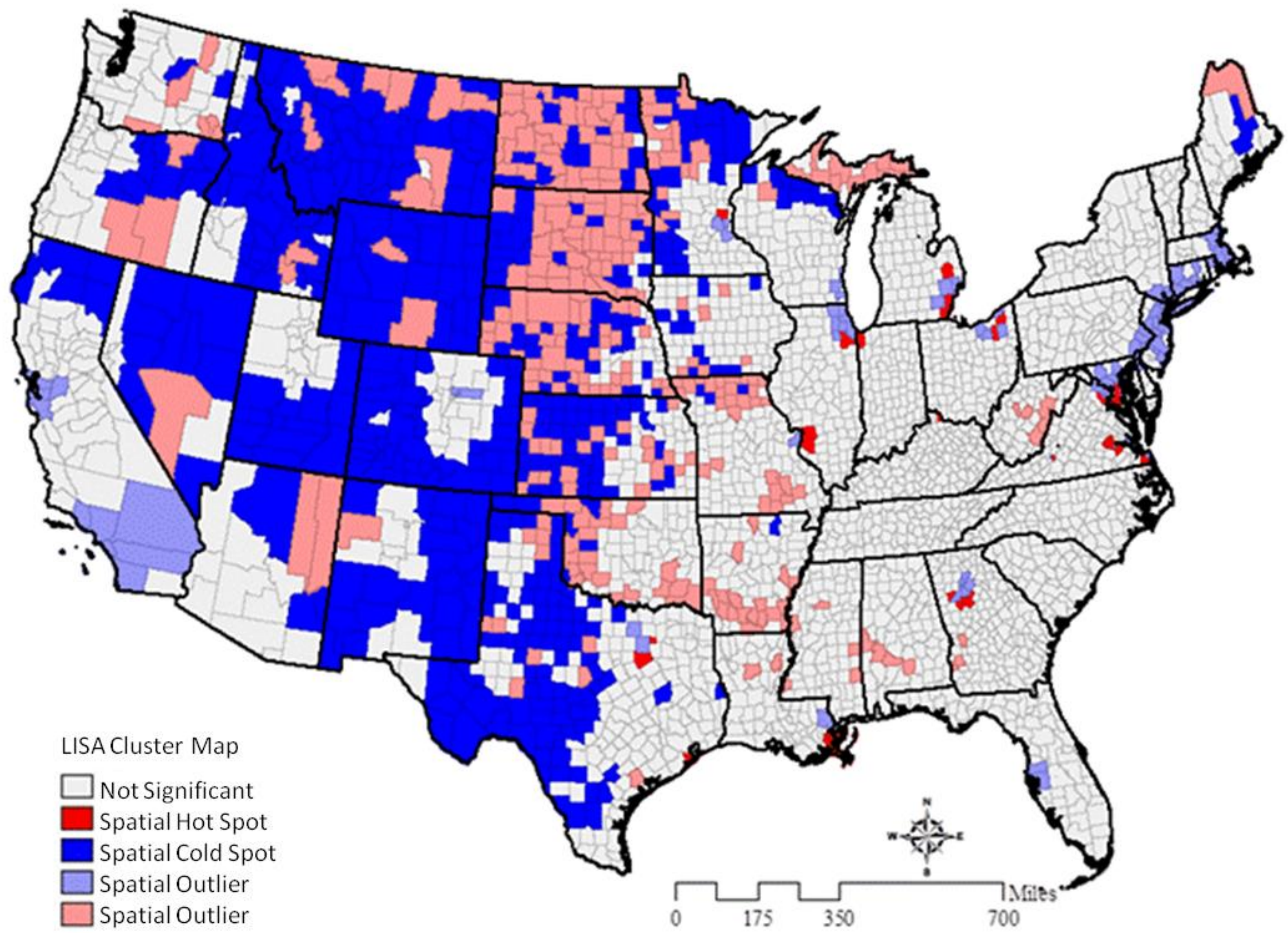

Figure 4.66 Bivariate LISA cluster map for county level obesity prevalence rates and population densities in the US for 2004

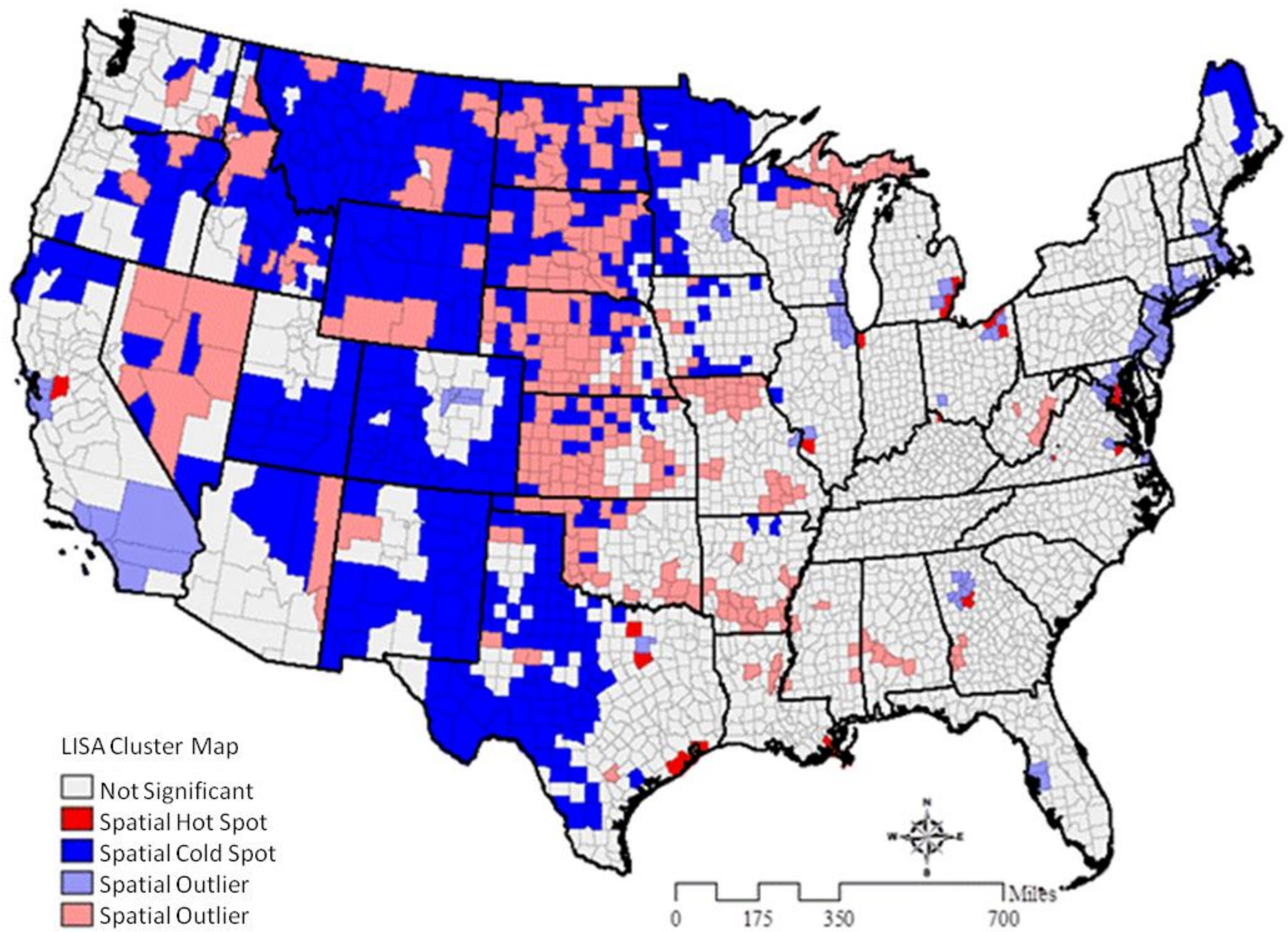

Figure 4.67 Bivariate LISA cluster map for county level obesity prevalence rates and population densities in the US for 2007

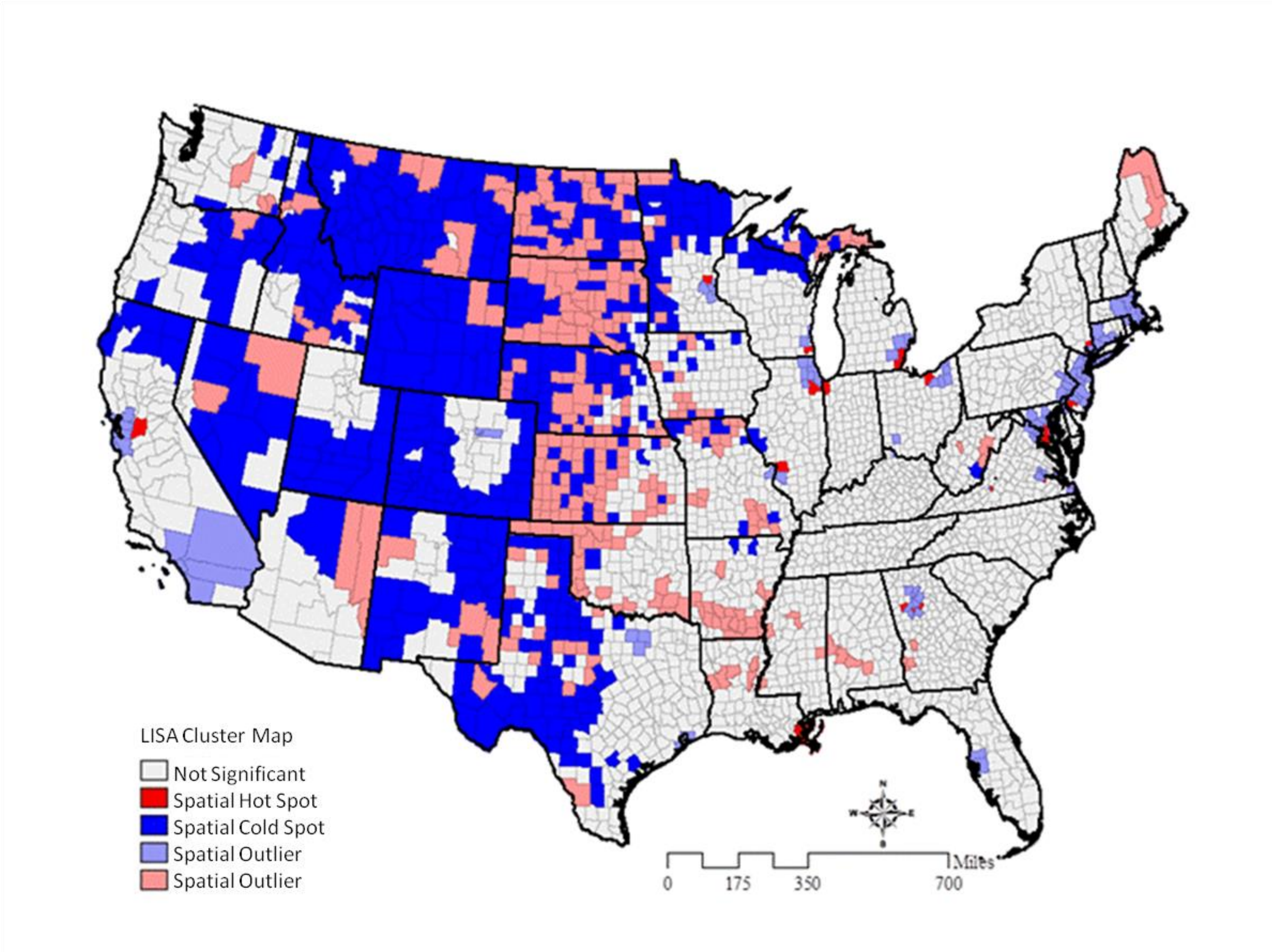

Figure 4.68 Bivariate LISA cluster map for county level obesity prevalence rates and population densities in the US for 2010

years. However, as observed in the LISA cluster maps, spatial clustering of high obesity with high population density and low obesity with low population density still exists to a certain degree. The clusters of high obesity and high population density are seen in Texas, Louisiana, Georgia, and Virginia in the South and in Illinois, Indiana, Michigan, and Ohio in the Midwest. However, a distinct cluster of low obesity and low population density is predominantly observed in the western half of the US. More specifically, these clusters are mainly seen in the entire Western US, the West North Central region in the Midwest, and the West South Central region (Texas and Oklahoma) in the Southern US. In addition, the clusters of low obesity and low population density exist in Wisconsin and Illinois in the East North Central region in the Midwest (Figures 4.66, 4.67, and 4.68).

On the other hand, in all three years, the high obesity and low population density spatial outliers are observed in the entire West North Central region of the Midwest and in the entire West South Central region of the South. The high obesity and low population density spatial outliers also foremost exist in the Mountain region in the West. The other notable high obesity and low population density spatial outliers exist in Mississippi, Alabama, Georgia, and West Virginia in the South, in Maine in the Northeast, and in Wisconsin and Michigan in the East North Central region in the Midwest (Figures 4.66, 4.67, and 4.68).

Over this time period (i.e., 2004, 2007, and 2010), the low obesity and high population density spatial outliers are predominantly seen in Massachusetts, Connecticut, New York, New Jersey, and Pennsylvania in the Northeast, in Maryland, Virginia, Florida, Georgia, and Texas in the South, in Ohio, Michigan, Illinois, Wisconsin, and Minnesota in the North Central region in the Midwest, and in Colorado and California in the West. Of all these low obesity and high population density spatial outlier areas, the Northeast Megalopolis region that stretches from

Virginia in the South to New Hampshire in the Northeast and California in the West emerge as the most distinctive areas (Figures 4.66, 4.67, and 4.68).

4.4. Discussion

As a major contribution, this study examined the obesity issue simultaneously accounting for space and time (i.e., the temporal changes from 2004 to 2007 and to 2010) while exploring the complex interplay of physical inactivity (PIA), poverty, ratio of jobs to employed residents (JER), population-weighted distance (PWD) to parks, and population density with the obesity phenomenon at the county level in the US. The main objectives of this study were to explore and identify spatio-temporal patterns and trends of obesity and their associations with the selected key variables, to examine whether these associations are based on clusters of occurrences or whether they are dispersed randomly across space, and to identify where these patterns have been manifested.

The main hypothesis aimed to test whether there was a spatial component involved in these temporal observations such as a significant spatial convergence of high obesity/PIA, poverty, JER, PWD to parks, and population density rates and/or low obesity/PIA, poverty, JER, PWD to parks, and population density rates. The more specific hypotheses included were: 1) There is a general clustering of high (or low) values of obesity and of the other tested variables; 2) Areas with high rates of PIA are more likely to have high rates of obesity; 3) Areas with high rates of poverty are more likely to have high rates of obesity; 4) Areas with high JERs are more likely to have low rates of obesity; 5) Areas with high PWDs to parks are more likely to have high rates of obesity; and 6) Areas with high population densities are more likely to have low rates of obesity.

As presented in the results section, a consistent answer obtained in this study significantly indicated ($p<0.001$) that the high (or low) values of the obesity, PIA, poverty, PWD to parks, and population density at the county level in the US were spatially clustered while the high (or low) values of the JER were spatially dispersed. These findings were yielded through univariate global Moran's I tests and confirmed through LISA tests at the local level. The term 'local' here refers to a scale for which a new set of results are obtained by utilizing the localized counterpart of the global Moran's I, LISA. For all variables, p-values were statistically significant and z-scores were positive except for JER, which had negative z-scores. Therefore, the null hypothesis that "the spatial distribution of polygons and their attributes are spatially independent" was rejected. In other words, for the obesity, PIA, poverty, PWD to parks, and population density variables, the spatial distribution of high (or low) values in the dataset was more spatially clustered than would be expected if underlying spatial processes were random, while the spatial distribution of high (or low) values was more spatially dispersed for the JER.

Further tests were conducted in a bivariate form to explore whether any spatially significant convergence of high (or low) obesity or the other tested variables exist. The following findings were yielded through bivariate global Moran's I tests and confirmed through LISA tests at the local level. For all variables, the p-values were statistically significant and the z-scores were positive for the obesity/PIA, obesity/poverty, and obesity/PWD to parks combinations, whereas the z-scores were negative for the obesity/JER and obesity/population density combinations. Therefore, the null hypothesis of spatial independence was rejected for all tested combinations. These significant associations indicated that high (or low) (i.e., similar) rates of obesity, PIA, poverty, and PWD to parks were spatially associated. In other words, areas with high (or low) rates of PIA or poverty or PWD to parks are more likely to have high (or low) rates of obesity. Conversely, it was also indicated that high (or low) rates of obesity, JER, and population density were spatially dispersed (i.e., inversely associated). In other words, areas with high rates of JER or population density are more likely to have low rates of obesity and vice versa. Furthermore, although all bivariate findings were statistically significant ($p<0.001$), the magnitudes of the spatial associations were higher for the obesity/PIA and obesity/poverty combinations than for the obesity/JER, obesity/PWD to parks, and obesity/population density combinations. This was also evident in the LISA cluster maps by visual detection of spatial clusters and spatial outliers both in terms of quantity and location. Consequently, all findings indicated that the actual spatial patterns of obesity and the selected key variables were strongly associated spatially and temporally with each other.

Moreover, it should also be noted that in general the focus is given to the areas, which are the so-called spatial clusters as shown on the LISA cluster maps. A cluster is classified as such when a value at a location, either high or low, is more similar to its neighbors, as calculated by averaging the neighboring values (i.e., the spatial lag), than it would be under the case of Completer Spatial Randomness (CSR). However, an equally high attention should be given to the areas which are the so-called outlier locations, where a value at a location is more dissimilar to its neighbors. This is especially important in the bivariate form. Such spatial outlier locations often reflect some type of competitive process and thus warrant further investigation to see if other factors are at work. Therefore, in this study, spatial outlier locations on the LISA cluster maps were also identified and discussed.

There are several key contributions of this study to the existing body of knowledge in obesity research. First of all, a key contribution was that the global and local findings helped to quantify and map the spatial distribution of obesity and the selected variables and to identify clusters (hot spots or cold spots) for all variables. With the utilization of both univariate and bivariate forms of spatial analyses, the identification of clusters for each unique variable, as well as between variable pairs, was accomplished. For instance, cluster locations for obesity (high or low obesity locations surrounded by neighboring locations with high or low obesity) in the univariate form and very similar cluster locations for obesity and PIA (high or low obesity locations surrounded by neighboring locations with high or low PIA) in the bivariate form were identified. Second, another key finding of the study was the identification of locations that presented inconsistent and/or unexpected spatial relationships (i.e., negative for the obesity/PIA, obesity/poverty, and obesity/PWD to parks combinations and positive for the obesity/JER and obesity/population density combinations) between obesity and the other variables (e.g., the locations that exhibited high obesity but low PIA). Last but not least, one of the gaps observed in the obesity literature was the lack of studies at the county level for the entire US. Therefore, this study fills that gap by exploring obesity and its confounding associations at the county level for the entire US. This study is a novel work especially by being the first study to explore obesity not only spatially but also temporally revealing the temporal changes from 2004 to 2007 and to 2010.

On the other hand, this study was subject to some limitations, as well. First of all, the heights and weights that the BRFSS uses to calculate BMI values were based on telephone self-reported information. It is known that both men and women may underreport weight and men may over report height in telephone surveys (Merrill and Richardson, 2009). Therefore, the prevalence of obesity rates used in this study should be considered somewhat conservative. Nonetheless, the very same BRFSS data are the most widely used obesity data and also assumed to present the best picture of the obesity epidemic in the US (BRFSS, 2014). Second, the spatial patterns of obesity identified may change depending on the spatial scale used, an issue also known as the Modifiable Areal Unit Problem (MAUP). Similar to most geographic research, the results of this study were scale-dependent, as well. Therefore, the findings were significant and important at the scale of counties and county equivalents within the US as a whole.

In conclusion, this study verified that the application of the univariate and bivariate Moran's I and LISA tests, as advanced ESDA methods, was effective in detecting clusters of obesity, as well as to identify obesity's complex spatial interplay with the selected independent variables. The identification of spatial clusters provided valuable information on the local geographical variation of obesity in the US and these findings can be very crucial to identify areas of elevated risk to tackle the problem of obesity. In addition, the outcomes and insights from this study will be useful to further study and explicate the underlying processes involved in forming such spatial patterns (spatial clusters and/or spatial outliers) and the determinants of obesity on the spatially clustered areas.

Therefore, further research will be conducted utilizing geographically weighted regression (GWR) with the addition of other key variables. So far, it was found that obesity and the selected variables were spatially autocorrelated. In such cases, using analytical methods that rely upon assumptions of the independence of observations, such as ordinary least squares (OLS) regression, can create biased findings. The data used in this study clearly violated that assumption and thus a spatial alternative to the traditional OLS method, such as the GWR, should be utilized. The GWR method has been specifically developed to deal with such situations (Fotheringham et al., 2002).

CHAPTER 5. A SPATIAL ANALYSIS OF COUNTY-LEVEL OBESITY PREVALENCE AND ITS ASSOCIATIONS WITH THE ENVIRONMENTAL, BEHAVIORAL, SOCIOECONOMIC, AND SOCIODEMOGRAPHIC FACTORS IN THE US

5.1. Introduction

Obesity has already reached epidemic proportions across the US and recently in 2013 it has been officially classified as a disease by the American Medical Association (AMA), the nation's largest physician organization (AMA, 2014). As previously mentioned, approximately 200,000 deaths that may be attributable to obesity each year makes obesity the second-leading cause of preventable death after smoking in the US and it is likely to become the first, unless this chronic issue is addressed successfully (Danaei et al., 2009; Flegal et al., 2005; Mokdad et al., 2004). In addition to this very serious issue, the total annual economic costs associated with obesity in the US were estimated to be in excess of $215 billion as of 2010 (Hammond and Levine, 2010).

Based on the literature review in Chapters 1 and 2, it was clear that the large increase in obesity prevalence in the US, especially since 1991, may reflect variations in behavioral, built environmental, socioeconomic, and sociodemographic factors. However, the extent to which these factors are related to the obesity phenomenon is still poorly understood. While this study may not identify all causal factors of the obesity phenomenon, it will investigate the spatial associations between obesity and behavioral, natural and built environmental, socioeconomic, sociodemographic, and population based dynamics at the county level in the US.

Behavioral factors play an important role as a major determinant of obesity by being associated with obesity risk throughout an individual's life course. More specifically, sedentary behaviors are often seen as causal factors and physical inactivity is by far the most commonly investigated factor in that regard. In the literature, studies exist from neighborhood to national levels and the common finding suggests that physical inactivity is positively associated with obesity (Schuurman et al., 2012; Durand et al., 2011; Sallis and Glanz, 2009; Ewing et al., 2008; Kim et al., 2006; Lopez and Hynes, 2006; Sturm, 2004; Saelens et al., 2003; Pate et al., 1995). Similarly, automobile dependency, which is considered as another form of sedentary behavior, has been found to act to discourage activity and therefore increases the risk of obesity (Lathey et al., 2009; Frank et al., 2007; Lopez and Hynes, 2006; Berrigan and Troiano, 2002).

Environmental factors associated with obesity primarily include various aspects of the natural and built environments. The natural and built environments consist of features such as green spaces, houses, businesses, and street networks that are quite stable and therefore have a great potential to affect health-related behaviors and thus health (Lovasi, 2012; Das, 2010). Previous research has highlighted that both the natural and built environmental factors such as availability of parks, recreational facilities, and dense mixed-use, therefore walkable, neighborhoods promote physical activity and thus they have been associated with lower rates of obesity (Jilcott Pitts et al., 2013; Lamichhane et al., 2012; Durand et al., 2011; McCormack and Shiell, 2011; Feng et al., 2010; Mujahid et al., 2008; Diez Roux et al., 2007). However, there have also been other studies that concluded that the effect of the built environment on obesity is not as clear (Durand et al., 2011; Feng et al., 2010). More specifically, features of the built environment can only encourage physical activity within the limitations and/or opportunities presented by the natural environment. For instance, sidewalks may be used less in areas where the summer is very hot or the winter is very cold and a bicycle lane or trail may get little use in areas where the landscape is steeper than normal (von Hippel and Benson, 2014). In other words, the characteristics of the natural environment may have power over the advantages of the built environment. In addition, especially in the more recent literature, food environments have been included as a part of the environmental determinants of obesity. Studies concerning food-related behaviors have concluded that disparities in the availability of restaurants and supermarkets are related to differences in caloric intake and therefore obesity rates across populations (Chi et al., 2013; Morland et al., 2012; Bodor et al., 2010; Cummins and Macintyre, 2006). On the other hand, there are mixed findings regarding the relationship between fast-food restaurant density and obesity. For instance, significant positive associations between the density of fast-food restaurants and obesity are shown both at the state and neighborhood levels (Li et al., 2009; Maddock, 2004). Similarly, in another study, a significant positive association is observed across 26 advanced economies (De Vogli et al., 2011). Conversely though, Ahern et al. (2011) found significant negative associations between the density of fast-food restaurants and obesity at the county level.

Socioeconomic indicators including but not limited to education, income, and poverty have been linked to both obesity and physical activity at individual and area scales, yet in the literature, trends in these effects are still subject to debate (Cohen et al., 2013; Cheng, 2012; Bennett et al., 2011; Levine, 2011; Wen et al., 2010; McLaren, 2007). Sociodemographic factors and population characteristics are also considered as important determinants of obesity in the literature. Previous studies have shown differences in obesity prevalence rates among minority populations as well as urban and rural populations (Jackson et al., 2015; Michimi and Wimberly, 2010; Ramsey and Glenn, 2002). Moreover, native residency is used as one of the population characteristics in this study. Interestingly, the use of this factor has not been detected in any obesity related studies in the literature.

Based on the above review, it is evident that physical and social environments may be characterized as obesogenic environments by encouraging or discouraging individuals' health related behaviors and therefore they have been documented as the determinants of the obesity issue in the US. For that reason, understanding which of these underlying factors contribute the most to the obesity issue in the US is crucial especially for determining possible prevention targets. However, one of the biggest challenges in understanding the associations between these underlying factors and obesity is that all these components vary across the US. Thus, the Geographically Weighted Regression (GWR) is an appropriate tool to further investigate the issue of obesity and its confounding relationships.

In the light of the literature review, the variables scrutinized in this part of the study were chosen to depict the different aspects of the possible determinants of the obesity issue in the United States at the county level. The physical inactivity (PIA) and commute to work by car variables were chosen to represent the behavioral factors, while the natural amenities index, ratio of jobs to employed residents (JER), and fast-food restaurants per capita variables were selected to characterize the natural and built environmental factors. On the other hand, poverty, African-American population, and high educational attainment were chosen to represent the socioeconomic and demographic factors, whereas rural population and native residents were selected to depict population characteristics.

In conclusion, the main objective of this part of the study is to identify predictors of obesity prevalence at the county level and to reveal how the effects of these predictors vary spatially. The main hypothesis aims to test whether predictors of obesity prevalence and effects and variations of such factors across space can be identified and revealed by the GWR technique at the county level. Consequently, the results from this study will generate useful insights that will help national as well as local obesity programs and interventions to better target areas and/or populations and factors associated with high obesity prevalence across the US.

5.2. Data and Methods

5.2.1 Data

The datasets used in this part of the study come from various sources and they are integrated to perform the GWR analysis at the county level. The original datasets are at the county level mainly for 2010. As discussed in detail in Chapter 3 of this dissertation, there are mainly eleven types of datasets. These datasets include: 1) obesity; 2) physical inactivity (PIA); 3) commute to work by car; 4) natural amenities index; 5) ratio of jobs to employed residents (JER); 6) fast-food restaurants per capita; 7) poverty; 8) African-American population; 9) high educational attainment; 10) rural population; and 11) native residents.

5.2.2 Methods

In this part of the study, the GWR is employed as a tool to further investigate the issue of obesity and its confounding relationship with various factors in the US. As discussed in depth in Chapter 3, there are various approaches to the analysis of spatial data including the GWR. Chapter 4 mainly focused on ESDA to detect and identify locations of obesity clusters. In order to assess why clusters and spatial outliers exist where they do, as well as to assess the underlying processes that created such patterns through the relationships between obesity and selected independent variables under investigation, secondary (further) approaches, such as regression analysis or spatial data process models are needed. These secondary approaches are equally important as the primary ESDA steps, if not more. Hence, the GWR, which is one of the most increasingly used local spatial statistical techniques to test for spatial nonstationarity, is employed as the secondary approach in this study (Hipp and Chalise, 2015; Fraser et al., 2012; Maroko et al., 2009).

The geographic phenomena and underlying processes and therefore the relationships between variables are often intrinsically spatial (space dependent) and vary geographically in 'real-world' situations (Fotheringham et al., 2002). As an exploratory technique, the GWR provides outcomes that are certainly powerful and offers significant advantages over simple linear regression procedures. It provides users a set of location specific parameter estimates not only to map and visualize, but also to interpret regarding the relationships between predictors and the outcome variable (Bitter et al., 2007; Wheeler and Tiefelsdorf, 2005).

The GWR model building process for this study were performed in mainly four steps: 1) Acquisition and manipulation of the data to be used in the global OLS regression and the local GWR; 2) Creation of descriptive statistics in the SPSS 23 software (IBM Corporation, 2015) and choropleth maps in ArcGIS 10.3 to initially explore data and visually interpret the spatial distribution of the data; 3) Specification of a valid global OLS regression model in the SPSS 23 software in order to identify variables to use in the GWR analysis; and 4) Implementation of the local GWR analysis in the GWR 4.0 software (https://geodacenter.asu.edu/gwr) to determine whether significant spatial heterogeneity is in effect at the county level and if so, to reveal spatial patterns by mapping parameter estimates in ArcGIS 10.3.

First of all, the datasets for all variables used in this study were downloaded from various sources as explained in the data section in Chapter 3. These data were imported into ArcGIS 10.3, where the data were joined to a geographic boundary file, also known as a shapefile, for 3,105 counties and county equivalents across the 48 contiguous states (all except Alaska and Hawaii) also known as the continental US. The data were joined using the common field of the counties' five-digit Federal Information Processing Standards (FIPS) codes. The finally merged shapefile was projected using the Albers equal-area conic projection.

Data were then imported into the SPSS 23 software to perform an initial exploration of the overall data distribution. To further explore the spatial distribution of the data, ArcGIS 10.3 was utilized for the creation of choropleth maps for all variables. The finalized versions of the choropleth maps were derived using the standard deviation method to emphasize the distribution of the data in reference to the means of the datasets. At the end of these steps, all candidate variables to be used in the global OLS regression were defined.

The specification of a valid global OLS regression model was accomplished in the SPSS 23 software using the forward stepwise selection technique. At each step, the variable that provides the greatest additional improvement to the fit is added to the model. These variables have also been further verified based on the review of the relevant literature. Moreover, possible multicollinearity issues among the independent variables were examined through the Variance Inflation Factor (VIF). The VIF is a measure that is commonly used to determine whether independent variables are multicollinear, in other words, if they are strongly intercorrelated and redundant from an information content point of view. A VIF value above 7.5 exceeds the common threshold and is considered problematical. Consequently, using the chosen variables, a valid and efficient Ordinary Least Squares (OLS) regression model has been fit to the data.

The OLS regression is a common statistical method for testing relationships between variables. Before modeling a spatial relationship it is always a good practice to start with an OLS regression, since OLS provides important diagnostic tests that inform the analyst whether a good model is achieved (Fotheringham et al., 2002).

Formally, an OLS regression model with multiple independent variables is defined as:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_n X_n + \varepsilon \qquad \text{Formula 5.1}$$

where y is the dependent variable, X_is are the independent variables, β_is are the parameters (coefficients) to be estimated, and ε is the random error term (residuals). The assumption is that the values of β_is are constant across the entire study area.

The OLS model is considered to be global, since one parameter is estimated for each variable included in the model. In such a model, the relationship between the dependent and the independent variables is assumed to be stationary at all locations in the study area. However, spatial data often violate the assumptions of OLS regression and thus further analysis is necessary to investigate whether significant spatial heterogeneity is in effect. Under such circumstances, a local model has more relevance to the real-world problem in hand compared to an OLS model. In other words, standard global modeling techniques, such as the OLS regression method, cannot identify nonstationarity and therefore utilizing such a method may obscure local variations in the relationships between predictors and the outcome variable (Matthews and Yang, 2012). Therefore, to deal with local variations in OLS regression models, methods that incorporate local variations into the regression model, such as the GWR, are essential. The GWR provides a well-designed and easily comprehended means of modeling such relationships (Fotheringham et al., 2002).

In the final step, the GWR analysis was implemented using the same set of variables as determined in the OLS regression by utilizing the GWR 4.0 software to determine whether significant spatial heterogeneity is in effect. Similar to the global OLS regression, the GWR 4.0 software calculates local parameter estimates for the independent variables as well as local R^2 statistics. The adaptive kernel, which is produced using the bi-square weighting function, was used when conducting the GWR analysis. The adaptive kernel approach uses varying spatial areas but with a fixed number of observations (although with different weights) for each estimation. This method is most appropriate when there is a large variation in the geographical density of the observed data as in the case of US counties (Fotheringham et al., 2002). Furthermore, the small sample bias corrected Akaike Information Criteria (AICc) was used to determine the optimal bandwidth size (Fotheringham et al., 2002). Finally, all parameter estimates as well as local R^2 statistics obtained from the GWR analysis were exported and mapped in ArcGIS 10.3 to reveal spatial patterns at the county level across the US.

When using the GWR to model a phenomenon (the dependent variable), the general interest usually involves either predicting values or understanding the dynamics that contribute to dependent variable outcomes. Another interest is examining how spatially consistent relationships between the dependent and each of the independent variable are across the entire study area. Examining the coefficient distribution as a surface reveals where and how much variation is present. In this study, the GWR analysis can help answer important questions including: 1) Is the general assumption made in the global modeling of spatial data, often without question, that the processes being modeled are stationary over space valid or not for obesity? 2) What is the relationship between obesity and the key factors investigated (behavioral, natural and built environmental, socioeconomic and demographic, and population characteristics) across the study area? 3) Are those relationships consistent across the study area? 4) Where are the specific areas within the study area in which ‘interesting’ relationships occur and therefore where further investigation is necessary?

5.3. Results

5.3.1 Descriptive Statistics and Basic Mapping of Variables

Details of the summary statistics of the obesity and all selected variables investigated for the counties across the continental US are shown in Table 5.1. The mean obesity rate among US adults and the standard deviation are 30.58% and 4.23, respectively (Table 5.1).

Table 5.1 Descriptive statistics for dependent and independent variables

	Minimum	Maximum	Mean	Std. Deviation	n
Obesity	13.10	47.90	30.58	4.23	3,105
PIA	10.40	44.90	27.90	5.28	3,105
Commute by Car	6.90	91.70	77.94	6.88	3,105
Natural Amenities Index	-6.40	11.17	0.05	2.28	3,105
JER	0.20	3.92	0.89	0.24	3,105
Fast-food Restaurants Per Capita	0.00	7.21	0.57	0.31	3,105
Poverty	3.10	50.10	16.80	6.23	3,105
African American Population	0.00	85.70	8.97	14.56	3,105
High Educational Attainment	3.70	71.00	18.98	8.65	3,105
Rural Population	0.00	100.00	58.55	31.42	3,105
Native Residents	8.50	97.50	68.00	14.54	3,105

The highest rate of obesity is observed in Greene, Alabama (47.90%) while the lowest rate of obesity is observed in Teton, Wyoming (13.10%) (Table 5.1). As previously noted, although the rates of obesity vary considerably among the counties across the entire US, the highest obesity rates are observed in the South while a large concentration of substantially lower obesity rates is observed in the West and Northeast (Figure 5.1).

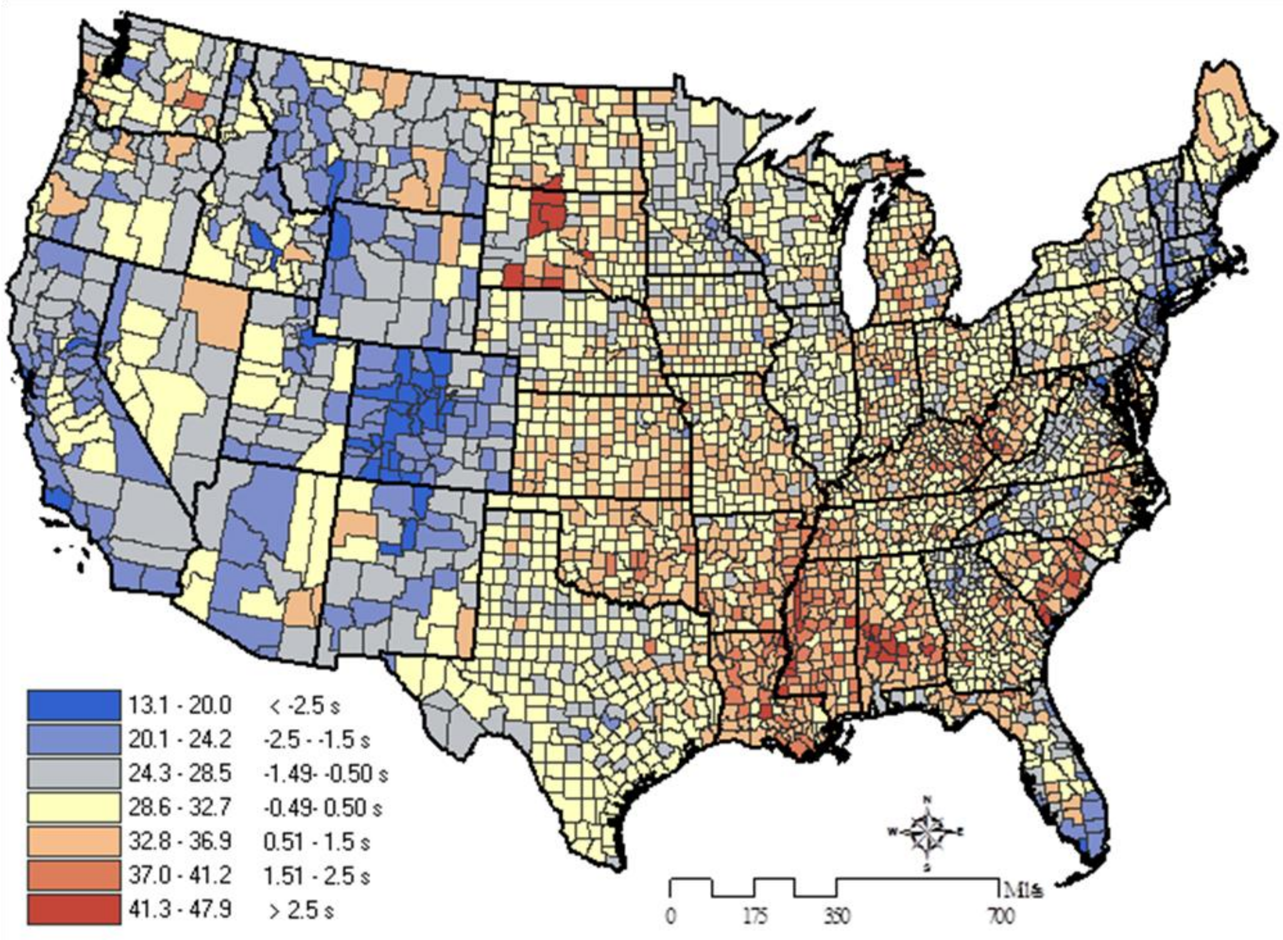

Figure 5.1 Geographic distribution of obesity rates at the county level in the US for 2010

In regard to the behavioral variables at the county level, the physical inactivity (PIA) rates among US adults range from 10.4% (Routt, Colorado) to 44.9% (Wyoming, West Virginia). The mean PIA rate is 27.90% (Table 5.1). Overall, higher PIA rates are seen in the South while lower rates are observed in the West (Figure 5.2). Similarly, on average 77.94% of the total population chose to commute to work by car, which is another inactive behavior type (Table 5.1). Higher rates of commute by car are mainly observed in the South and the East North Central Region in the Midwest (Figure 5.3). On the other hand, lower rates are mostly found in the West and partly in the West North Central Region in the Midwest (Figure 5.3).

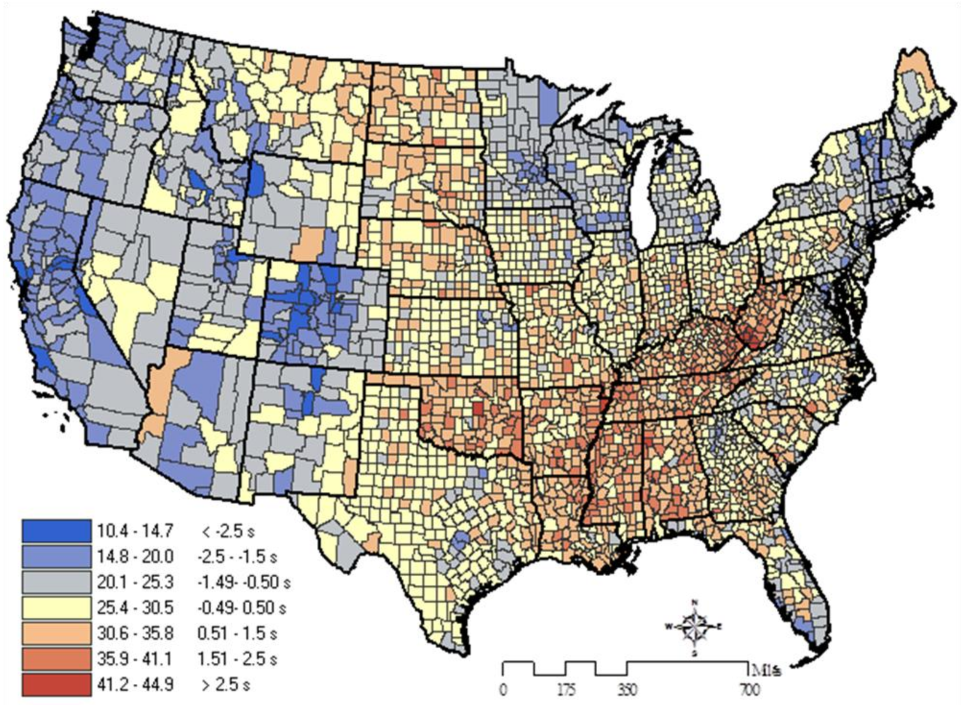

Figure 5.2 Geographic distribution of physical inactivity (PIA) rates at the county level in the US for 2010

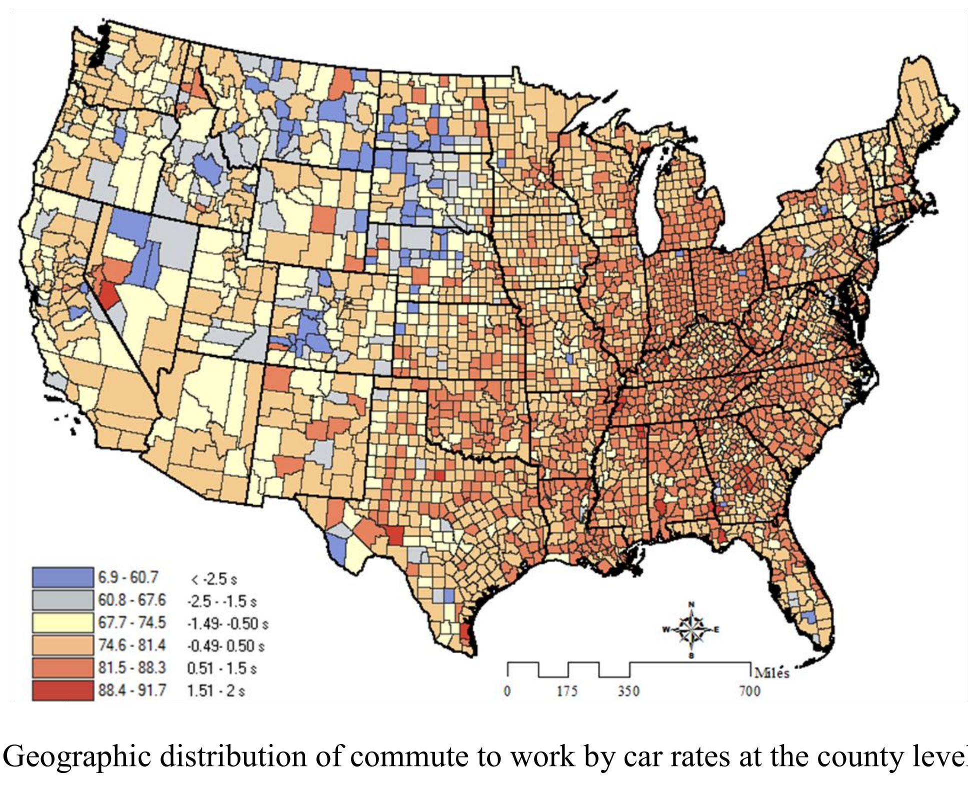

Figure 5.3 Geographic distribution of commute to work by car rates at the county level in the US for 2010

The natural and built environmental variable patterns are shown in Figures 5.4, 5.5, and 5.6. Based on the USDA's measured scores, the mean natural amenities index is 0.05 (Table 5.1). Overall, higher index values are observed in the West as well as in the State of Florida while lower values are mainly seen in the Midwest (Figure 5.4). The mean JER is 0.89 (Table 5.1). The JER composition of US counties varies considerably. However, the lower JERs are mainly observed in the South as well as in the Midwest (Figure 5.5). The average fast-food restaurant per capita was 0.57 (Table 5.1). Although there is not a clear distribution pattern, relatively higher values are observed in the West (Figure 5.6).

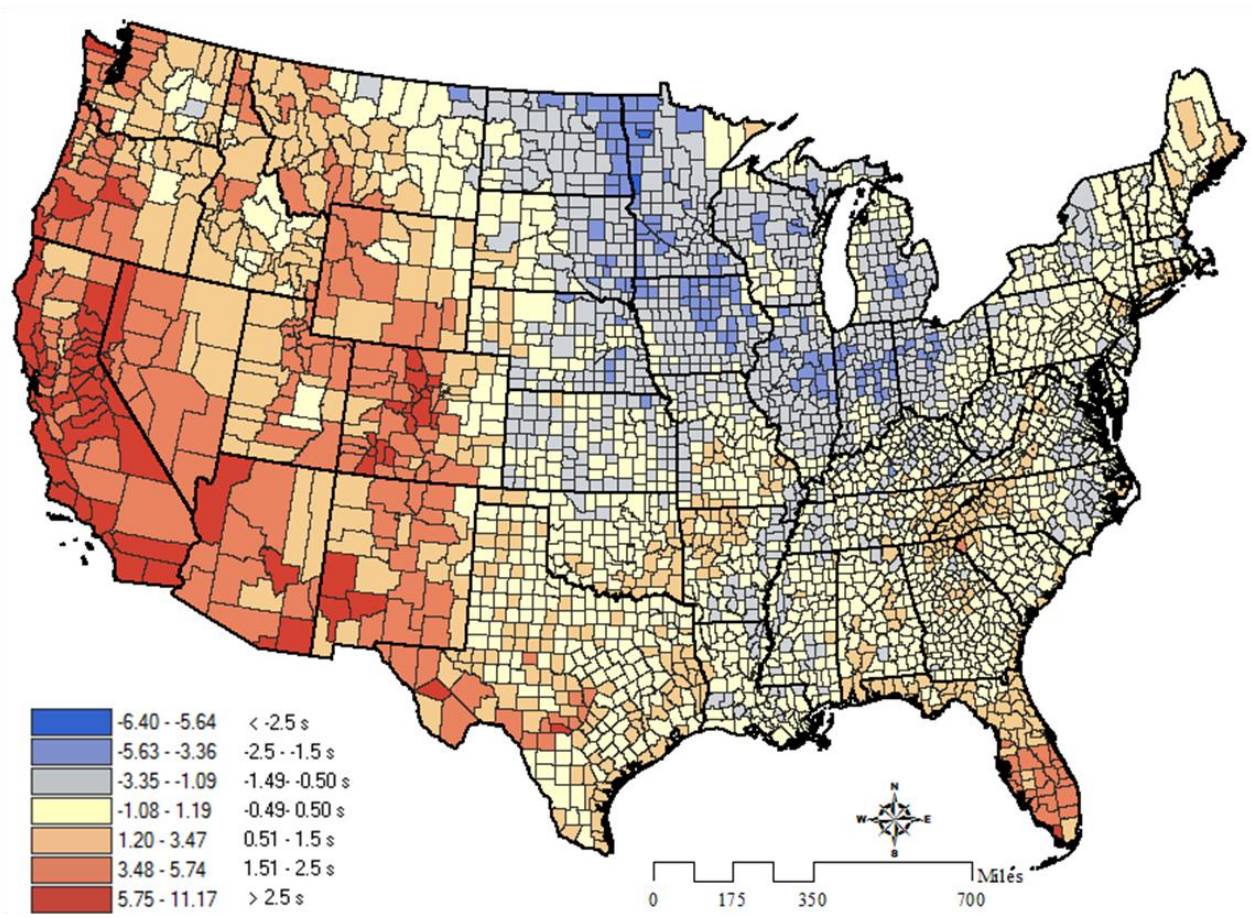

Figure 5.4 Geographic distribution of natural amenities index at the county level in the US for 1999

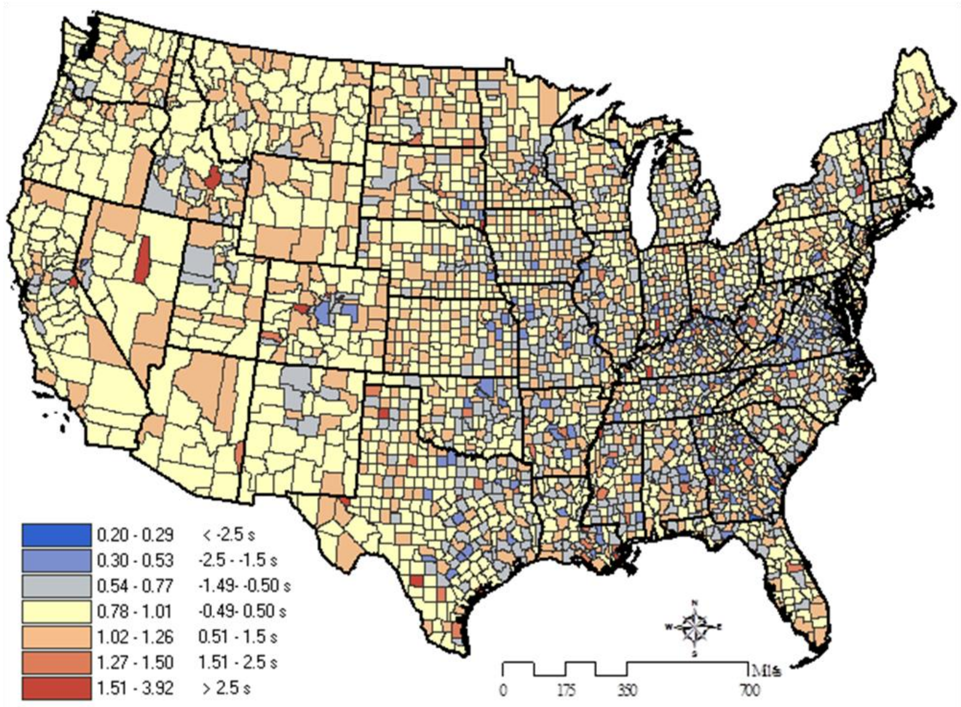

Figure 5.5 Geographic distribution of ratio of jobs to employed residents (JERs) at the county level in the US for 2010

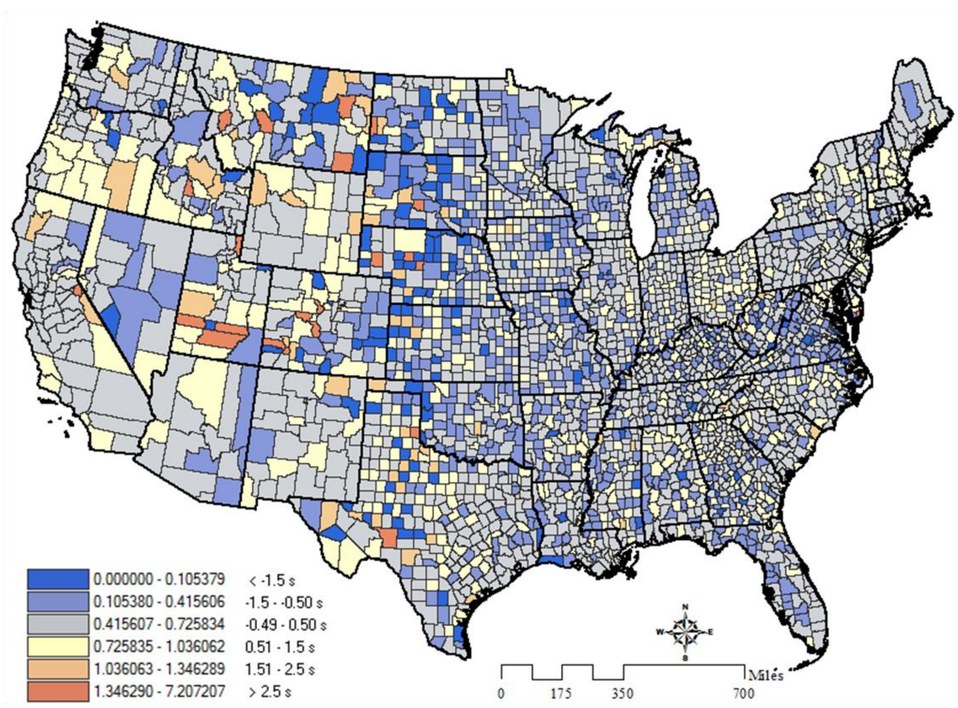

Figure 5.6 Geographic distribution of fast-food restaurants per capita at the county level in the US for 2010

Figures 5.7, 5.8, and 5.9 illustrate distribution patterns of the socioeconomic and demographic variables. The average poverty rate is 16.80% at the county level (Table 5.1). As seen in Figure 5.7, there is almost a horizontal line that divides the US in the middle and the counties in the southern half of this divide are dominated by higher poverty rates, while the counties in the northern half are dominated by lower poverty rates. In addition, the counties in the Northeast emerge as areas with lower poverty rates (Figure 5.7).

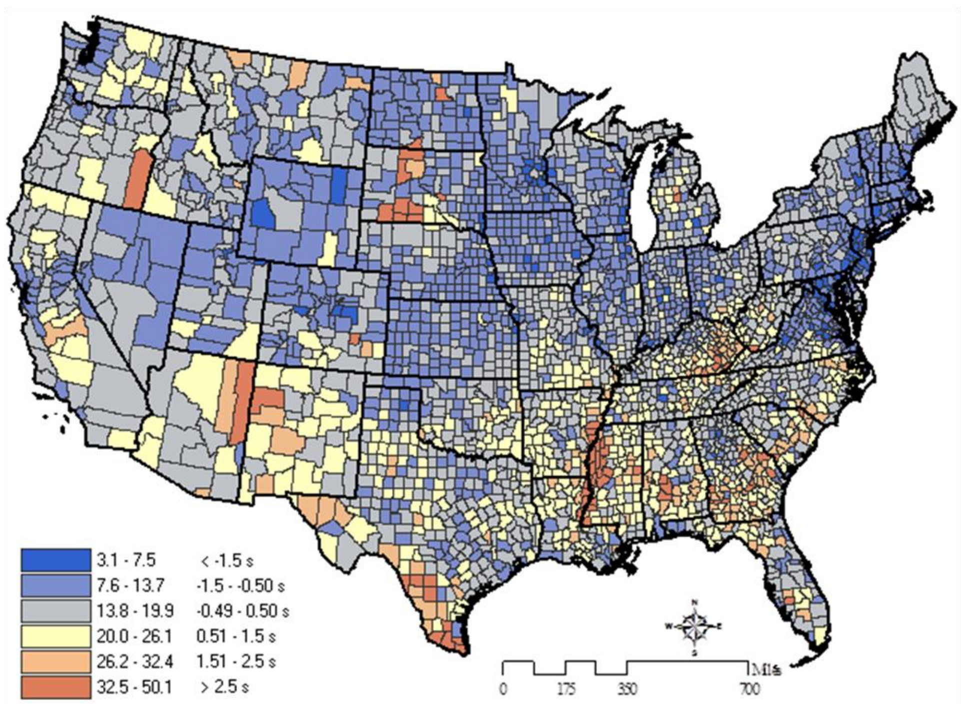

Figure 5.7 Geographic distribution of poverty at the county level in the US for 2010

On average about 9% of the total population is African-American (Table 5.1). Jefferson County in Mississippi (85.7%) has the highest percentage of African-American population. According to Figure 5.8, the entire Southern US is heavily populated by African-American individuals at the county level. The high educational attainment rate is 18.98% on average (Table 5.1). Higher overall high educational attainment rates are clustered mostly in the

Northeast as well as in the West (Figure 5.9). Conversely, lower rates of high educational attainment rates are clustered predominantly in the South and partly in the Midwest (Figure 5.9).

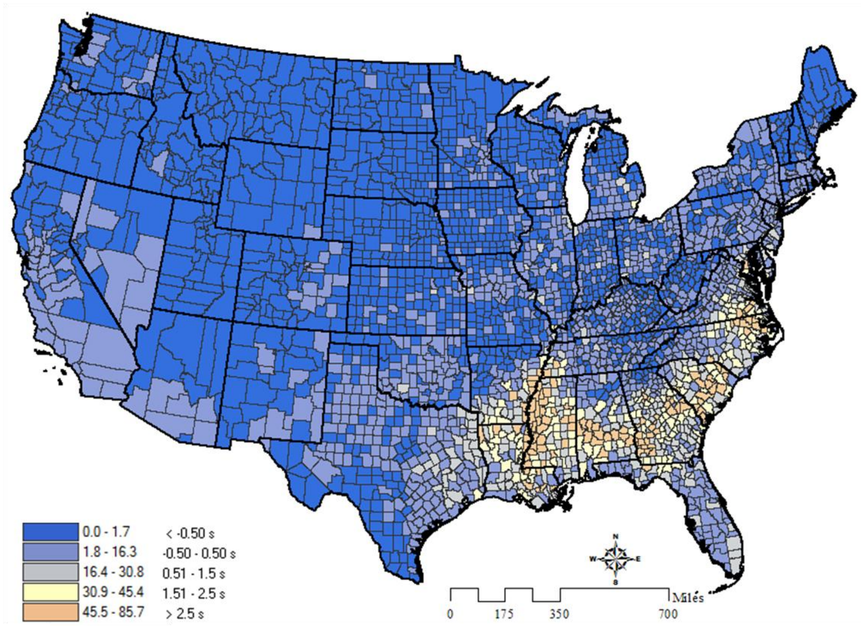

Figure 5.8 Geographic distribution of African American population rates at the county level in the US for 2010

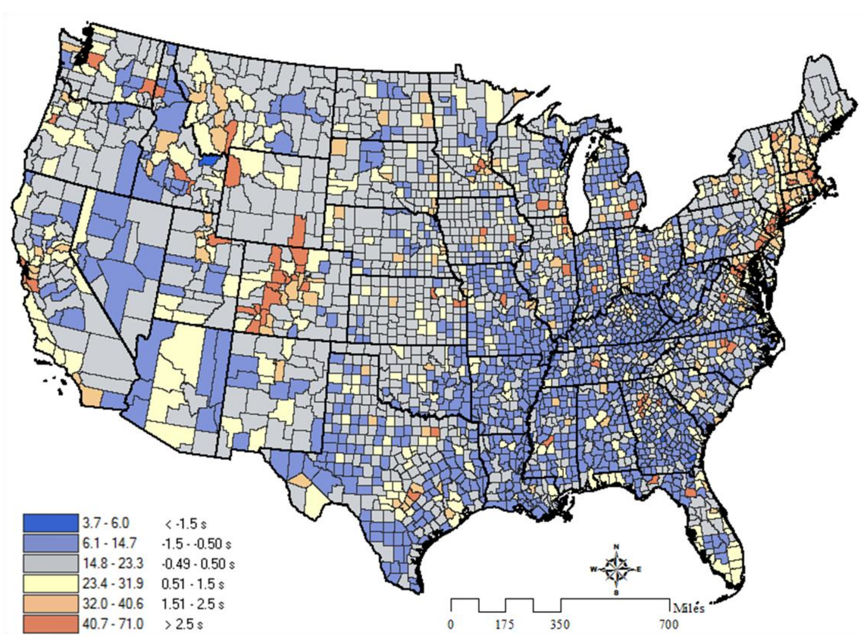

Figure 5.9 Geographic distribution of high educational attainment rates at the county level in the US for 2010

In regard to population characteristics, on average 58.55% of the US population resides in rural areas (Table 5.1). Overall though, the percentage of individuals residing in rural areas varies widely across the US without showing a clear clustering pattern (Figure 5.10). In addition, Figure 5.11 shows the spatial distribution of the percentage of the native residents. The average native resident rate is about 68% (Table 5.1). Higher percentages of native residents are observed mostly in the South (especially in Louisiana and Mississippi), except in Florida, and in the Midwest extending through Pennsylvania and New York (Figure 5.11). Lower rates of native residents are especially found in the entire West (especially in Nevada, Arizona, Colorado, and Wyoming), the New England Region of the Northeast, and in Florida in the South, suggesting that more people migrate to these areas (Figure 5.11).

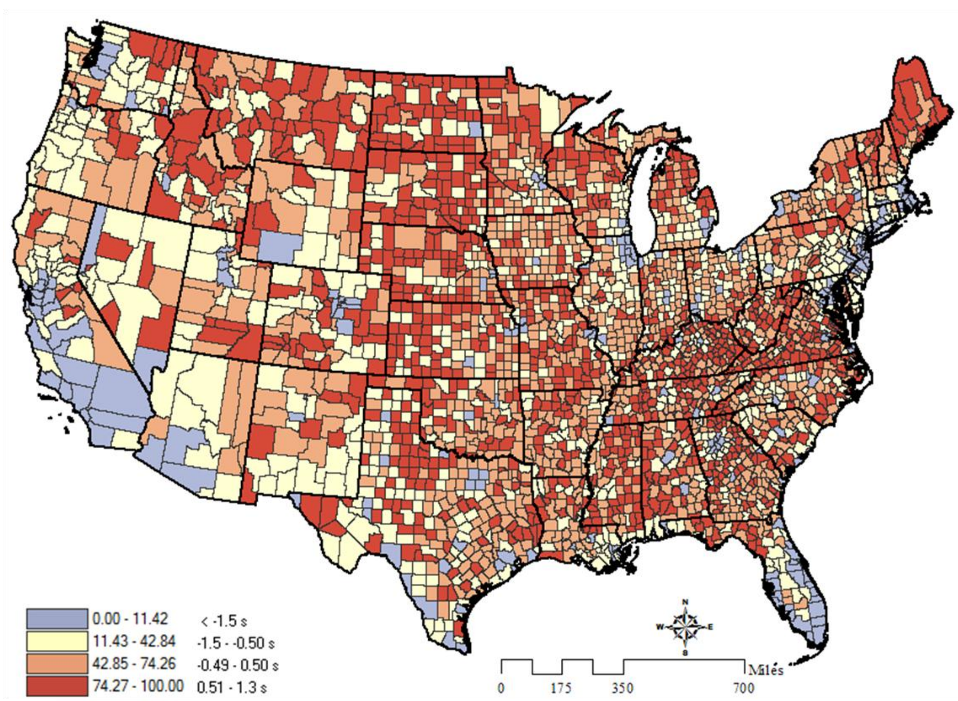

Figure 5.10 Geographic distribution of rural population rates at the county level in the US for 2010

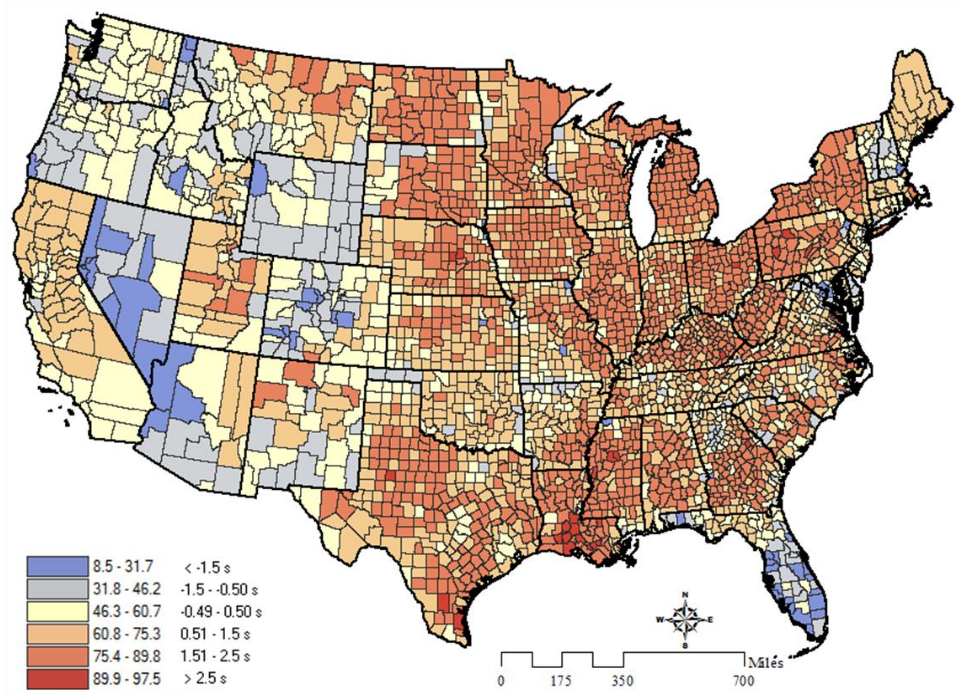

Figure 5.11 Geographic distribution of native residents rates at the county level in the US for 2010

5.3.2 Ordinary Least Squares (OLS) Regression

5.3.2.1 Nationwide OLS Results

For this part of the study, the global estimates for the entire study area are shown in Table 5.2. The global OLS regression results show that the OLS model performs quite well and that the parameter signs are as expected. All parameter estimates are significant based on the p-values as shown in Table 5.2.

As to the behavioral variables, the PIA variable has a positive coefficient (0.287), signifying that the relationship with obesity is positive. More specifically, it indicates that obesity rates are higher in areas where higher PIA rates exist. As the rate of PIA increases by one unit, the obesity rate increases by 0.287%. Similarly, the commute by car variable has a positive coefficient (0.059), indicating that areas with residents who choose to commute by car are more likely to have higher rates of obesity. In regard to the behavioral variable estimates at the county level, it can be concluded that overall, more inactive behavior types trigger higher obesity rates.

Table 5.2 Results of global ordinary least squares (OLS) regression for the US

	Coefficient	SE	*p*-Value	VIF
Intercept	18.036	0.817	0.000	
PIA	0.287	0.013	0.000	2.146
Commute by Car	0.059	0.007	0.000	1.289
Natural Amenities Index	-0.369	0.025	0.000	1.543
JER	-0.420	0.191	0.028	1.396
Fast-food Restaurants Per Capita	-0.629	0.169	0.000	1.354
Poverty	0.068	0.010	0.000	1.974
African American Population	0.064	0.004	0.000	1.443
High Educational Attainment	-0.117	0.008	0.000	2.457
Rural Population	0.005	0.002	0.010	1.899
Native Residents	0.012	0.004	0.002	1.638
Adjusted R-square	0.648			
AIC	14,546.713			

In regard to the natural and built environmental variables, the negative sign of the natural amenities index variable suggests that areas with higher natural amenities index are more likely to have lower obesity rates. More specifically, with every one unit increase in the natural amenities index, the obesity rate decreases by about 0.37%. Similarly, the JER variable has a negative coefficient (-0.420), indicating that the relationship with obesity is negative. It confirms that higher JER values are associated with lower obesity rates. The coefficient for fast-food restaurant per capita is negative (-0.629) suggesting that areas with higher fast-food restaurant per capita rates possess lower obesity rates.

As to the socioeconomic and demographic variables, the poverty variable has a positive coefficient (0.068), indicating that the relationship is positive. More specifically, it indicates that obesity rates are higher in areas where high poverty rates exist. With every one unit increase in the poverty rate, the obesity rate increases by roughly 0.07%. Likewise, the positive sign of the African-American population variable suggests that obesity rates are higher in areas where higher percentage of African-American individuals reside. As the percentage of African-Americans increases by one unit, the obesity rate increases by about 0.06%. Conversely, the high educational attainment variable has a negative coefficient (-0.117), signifying that the relationship with obesity is negative. This basically indicates that areas populated by highly educated residents are more likely to have lower obesity rates.

In regard to the population characteristics, the positive sign of the rural population variable suggests that areas with higher rural populations are more likely to have higher obesity rates. Similarly, the native residents variable has a positive coefficient (0.012), indicating that obesity rates are higher where native resident rates are higher. Among all the variables, a change in the rate of the rural population or native residents' variables by one unit generates the least change in obesity rates based on the 'global' OLS results.

The VIF values for each variable are also shown in Table 5.2 and they do not suggest any multicollinearity among the independent variables. The coefficient of determination, R^2, is a measure of model performance. However, it is a common practice to report adjusted R^2**.** The adjusted R^2 is a modified version of R^2 that has been adjusted based on the number of independent variables included in the model. Values vary from 0.0 to 1.0, with higher values being preferable. Thus, overall the 'global' OLS model is robust having an adjusted R^2 at a satisfactory level of 0.648 and therefore explaining a significant amount of variance (almost

65%) in the obesity rates at the county level ($p<0.001$) (Table 5.2). The Akaike Information Criterion (AIC) (Akaike, 1974) is a measure that is commonly used to verify the relative quality of statistical models (Matthews and Yang, 2012). The AIC comparisons among different models reveal whether the overall model fits are improved or not (Fotheringham et al., 2002). The AIC for the global OLS model is 14,546.713 and this specific AIC value is compared with the AIC value obtained from the local GWR to compare the relative quality of the two models (i.e., 'global' vs. 'local'). The lower the AIC value, the better the model fit is.

5.3.2.2 Regional OLS Results

A series of additional global OLS regression analyses is also conducted in order to elaborate and explore further whether geographic variability in the association between obesity rates and explanatory variables exist. For that reason, the data for the entire study area is disaggregated into regional areas based on the US Census Bureau's region definitions and used. The global OLS estimates for the Northeast, Midwest, South, and West are presented in Tables 5.3, 5.4, 5.5, and 5.6, respectively. As evident in these tables, the global OLS results for the regions indicate that geographic variability in the association between obesity rates and explanatory variables exists. It is also observed that only some of the parameter estimates are significant in the global OLS for the US regions while all parameter estimates are significant in the global OLS for the entire US as presented previously in Table 5.2. The details of the regional OLS results are interpreted further as follows.

In the Northeast, the significant parameter estimates include PIA (0.182), commute by car (0.090), and high educational attainment (-0.229) while all others are insignificant based on the p-values (Table 5.3). Among these significant variables, a change in the rate of the high educational attainment variable by one unit generates the most change in obesity rates. In other words, high educational attainment is the most important predictor of obesity in the Northeast. On the other hand, commute by car is the least important predictor of obesity in the Northeast. On the other hand, in the Midwest only the parameter estimates for the PIA (0.187), commute by car (0.036), fast-food restaurants per capita (-0.687), poverty (0.170), and high educational attainment (-0.090) variables are significant based on the p-values (Table 5.4). Among these significant variables, fast-food restaurants per capita is the most important predictor of obesity while commute by car is the least important predictor (i.e., with one unit change, generates the least change in obesity rates) in the Midwest.

Table 5.3 Results of global ordinary least squares (OLS) regression for the Northeast

	Coefficient	SE	*p*-Value	VIF
Intercept	20.644	3.470	0.000	
PIA	0.182	0.062	0.004	2.861
Commute by Car	0.090	0.020	0.000	2.243
Natural Amenities Index	-0.119	0.142	0.401	1.181
JER	1.221	0.700	0.083	1.622
Fast-food Restaurants Per Capita	-1.064	0.985	0.281	1.786
Poverty	0.058	0.053	0.274	2.468
African American Population	0.038	0.031	0.216	2.545
High Educational Attainment	-0.229	0.033	0.000	3.948
Rural Population	0.005	0.008	0.500	3.158
Native Residents	0.000	0.012	0.991	1.859
Adjusted R-square	0.730			

Table 5.4 Results of global ordinary least squares (OLS) regression for the Midwest

	Coefficient	SE	*p*-Value	VIF
Intercept	22.107	1.754	0.000	
PIA	0.187	0.022	0.000	1.559
Commute by Car	0.036	0.013	0.007	1.544
Natural Amenities Index	-0.005	0.056	0.927	1.160
JER	0.510	0.433	0.240	1.615
Fast-food Restaurants Per Capita	-0.687	0.341	0.044	1.488
Poverty	0.170	0.018	0.000	1.484
African American Population	-0.004	0.019	0.825	1.508
High Educational Attainment	-0.090	0.015	0.000	2.095
Rural Population	-0.002	0.004	0.577	2.859
Native Residents	0.005	0.007	0.524	1.146
Adjusted R-square	0.332			

In the South, the significant parameter estimates include PIA (0.375), natural amenities index (-0.314), fast-food restaurants per capita (-0.661), African American population (0.080), and high educational attainment (-0.081) while all others are insignificant based on the p-values (Table 5.5). Among these significant variables, fast-food restaurants per capita is the most important predictor of obesity in the South. On the other hand, African American population and high educational attainment variables are, almost equally, the least important predictors of obesity in the South. On the other hand, in the West only the parameter estimates for the PIA (0.240), natural amenities index (-0.239), high educational attainment (-0.224), and rural population (0.036) variables are significant (Table 5.6). Among these significant variables, PIA is the most important predictor of obesity (i.e., with one unit change, generates the most change in obesity rates) while rural population is the least important predictor in the West.

Table 5.5 Results of global ordinary least squares (OLS) regression for the South

	Coefficient	SE	*p*-Value	VIF
Intercept	18.373	1.295	0.000	
PIA	0.375	0.019	0.000	2.007
Commute by Car	0.025	0.013	0.054	1.189
Natural Amenities Index	-0.314	0.050	0.000	1.258
JER	-0.490	0.304	0.107	1.975
Fast-food Restaurants Per Capita	-0.661	0.280	0.018	1.839
Poverty	0.015	0.014	0.284	2.041
African American Population	0.080	0.004	0.000	1.444
High Educational Attainment	-0.081	0.012	0.000	2.636
Rural Population	0.004	0.003	0.188	2.124
Native Residents	0.007	0.006	0.217	1.760
Adjusted R-square	0.637			

Table 5.6 Results of global ordinary least squares (OLS) regression for the West

	Coefficient	SE	*p*-Value	VIF
Intercept	24.811	2.837	0.000	
PIA	0.240	0.045	0.000	2.381
Commute by Car	0.030	0.024	0.220	1.521
Natural Amenities Index	-0.239	0.071	0.001	1.453
JER	0.484	0.360	0.179	1.035
Fast-food Restaurants Per Capita	-0.073	0.322	0.820	1.132
Poverty	0.027	0.032	0.407	1.338
African American Population	-0.097	0.080	0.229	1.290
High Educational Attainment	-0.224	0.023	0.000	2.412
Rural Population	0.036	0.006	0.000	1.869
Native Residents	0.012	0.012	0.318	1.177
Adjusted R-square	0.558			

Although some of the variables are referred to as the most and least important predictors of obesity in the above interpretations, the designation of the "most" and "least" only refers to the magnitude of the effect. Otherwise, it should be noted that all significant variables are important in explaining obesity rates in their regional context. Furthermore, overall, the parameter estimates for the PIA and high educational attainment are significant in all US regions. On the other hand, the parameter estimates for the poverty, African American, and rural population are only significant in the Midwest, South, and West, respectively.

The VIF values for each variable in each region are also shown in Tables 5.3, 5.4, 5.5, and 5.6 and they do not suggest any multicollinearity among the independent variables. Overall, the global OLS models for the Northeast, South, and West are robust having adjusted R^2 values of 0.730, 0.637, and 0.558, respectively while the model for the Midwest is relatively quite weak having an adjusted R^2 of 0.332 ($p<0.001$). In other words, the model for the Midwest explains an insignificant amount of variance (only 33%) in the obesity rates at the county level when compared to the models for rest of the regions explored. In the end, the above results show that geographic variability in the association between obesity rates and explanatory variables can be detected to a certain extent by conducting the OLS for the US regions instead of the entire US.

5.3.3 Geographically Weighted Regression (GWR)

As stated previously, spatial data often violate the assumptions of OLS regression and therefore further analysis is crucial to investigate whether significant spatial heterogeneity is in effect. In that case, using methods such as the GWR that incorporate local variation into the regression model is essential. The above global OLS results confirm that the global OLS model performs well and thus the variables included in the global model can be carried on to run the GWR analysis. It is expected that the GWR analysis will not only improve the overall model fit but also reveal spatial patterns such as local variations that are not visible otherwise. Details of the GWR analysis will be examined in the following section.

The GWR model summary is presented in Table 5.7 and it confirms that there is a significant improvement over the OLS model. The 'local' GWR model is thus more robust. The GWR returns an overall adjusted R^2 of 0.778, indicating that the model explains about 78% of the variance in the obesity rates at the county level ($p<0.001$) (Table 5.7). In addition, the AIC in the GWR model is 13,830.341. Since the AIC value from the GWR is smaller than the AIC value obtained from the OLS (14,546.713), it implies that the GWR model has been improved by allowing relationships to vary across the study area. The values of both the adjusted R^2 and AIC confirm that the 'local' GWR model is much better than the 'global' OLS model.

Table 5.7 Results of local geographically weighted regression (GWR)

	Min	25% Quartile	50% Quartile	75% Quartile	Max
Intercept	2.739	15.811	20.544	25.363	38.824
PIA	0.031	0.171	0.246	0.316	0.609
Commute by Car	-0.118	0.014	0.042	0.080	0.206
Natural Amenities Index	-0.866	-0.342	-0.207	-0.001	0.452
JER	-20.208	-2.718	-0.746	0.927	11.850
Fast-food Restaurants Per Capita	-3.479	-0.868	-0.258	0.325	3.623
Poverty	-0.485	0.082	0.198	0.329	0.832
African American Population	-0.691	0.001	0.063	0.100	1.272
High Educational Attainment	-0.248	-0.148	-0.102	-0.076	0.059
Rural Population	-0.039	-0.014	-0.004	0.005	0.033
Native Residents	-0.094	-0.019	0.004	0.030	0.092
Adjusted R-square	0.778				
AIC	13,830.34102				

Furthermore, all regression coefficients obtained from the local GWR model show evidence of non-stationarity. As shown in Table 5.8, this is confirmed by the fact that the interquartile ranges of the local GWR regression coefficients were larger than twice the standard errors of the regression coefficients of the global model (Ailes et al., 2012; Weisent et al., 2012). This indicates that the regression coefficients of each of the independent variables included in the model change across all counties in the study area. In other words, this implies that the strength of the associations between obesity and each of the independent variables varies depending on the spatial location. Therefore, assuming a causal relationship, the effects of the determinants of obesity are not constant across the US, but are mainly dependent on the geographical location.

Table 5.8 Assessment of the stationarity based on the local GWR model coefficients

	Global SE	Global SEx2	Local GWR Interquartile R.	Is Coefficient Non-Stationary?
PIA	0.013	0.025	0.145	Yes
Commute by Car	0.007	0.015	0.065	Yes
Natural Amenities Index	0.025	0.049	0.341	Yes
JER	0.191	0.381	1.236	Yes
Fast-food Restaurants Per Capita	0.169	0.338	1.193	Yes
Poverty	0.010	0.020	0.107	Yes
African American Population	0.004	0.007	0.099	Yes
High Educational Attainment	0.008	0.016	0.072	Yes
Rural Population	0.002	0.004	0.019	Yes
Native Residents	0.004	0.008	0.049	Yes

Figure 5.12 illustrates the spatial distribution of the geographically weighted local R^2 values ($p<0.05$). Local R^2 values range from 0.11 to 0.87. Some counties have quite high R^2 values, where the model performs well, while others have very low values of R^2, where the model fit is less precise (Figure 5.12). Overall, the GWR model's highest performance is mainly observed in Florida and parts of Alabama, Georgia, South Carolina, and North Carolina in the South, and with a second area mainly in Colorado and parts of New Mexico, Utah, Arizona, and Wyoming in the Mountain Region in the West (Figure 5.12). Also, most of the Northeast and counties around the DC area have relatively high local R^2 values, as well. On the other hand, poorer model fits are seen among the states in the Midwest and partly in the South including Texas, Oklahoma, Tennessee, and Kentucky (Figure 5.12).

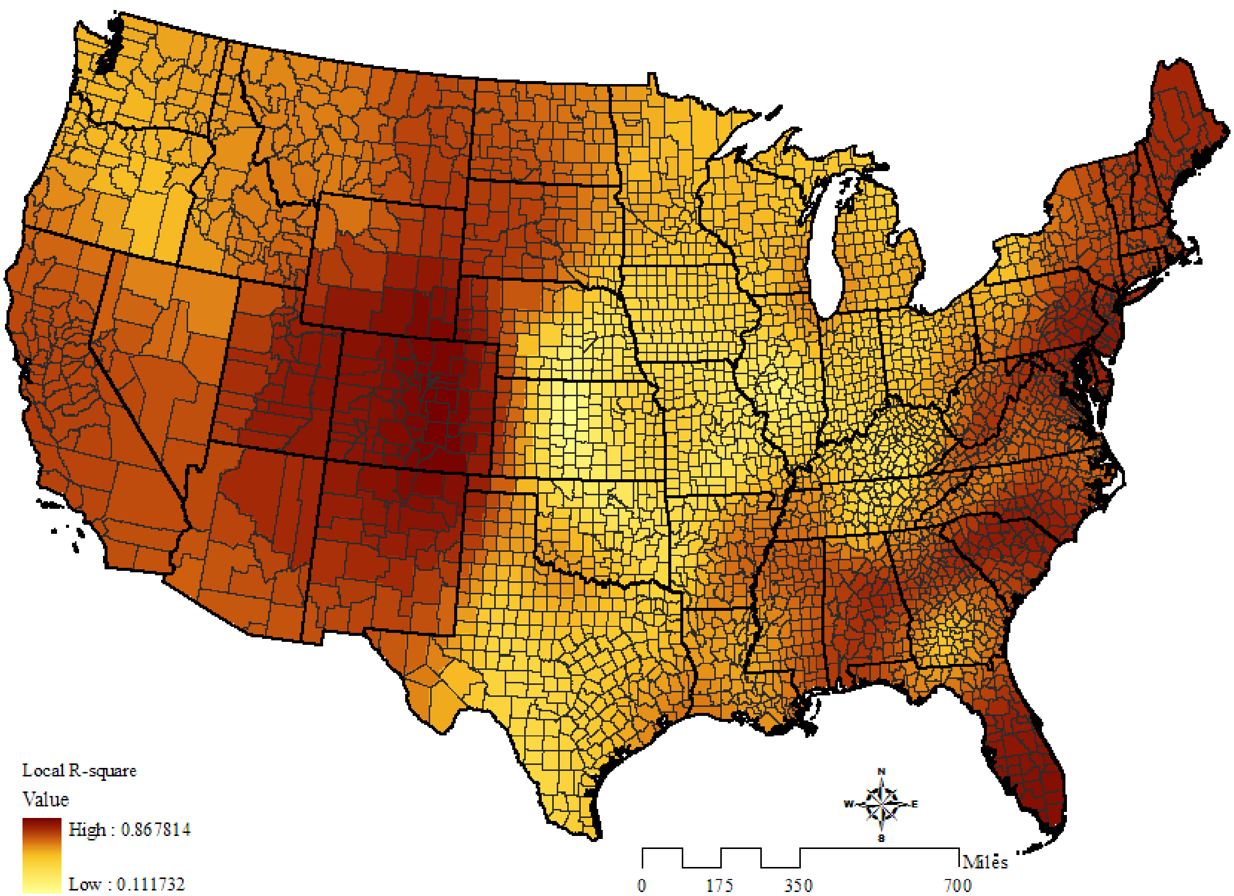

Figure 5.12 Local coefficient of determination (R^2) map based on the local GWR model

Although Table 5.7 above presents the intercept and coefficient ranges across the study area as obtained from the GWR model, it is crucial to map the local parameter estimates for the independent variables used to observe the spatial patterns, to reveal where significant nonstationarity is in effect and where it is not. Therefore, local GWR parameter estimates are mapped and compared to obtain further and useful insight concerning spatial variations in the relationships. Primarily the GWR method itself and in particular the maps generated from the GWR outputs are exploratory rather than explanatory, showing mainly where particular covariates contribute strongly and where they do not. The spatial patterns of the local GWR coefficients of the independent variables are illustrated in Figures 5.13 through 5.22 ($p<0.05$ shown on the maps). The white areas show an insignificant relationship with obesity.

Figure 5.13 shows significant positive coefficients of PIA for most of the US. Although the magnitude of the effect of PIA on obesity varies across the US, the PIA effect is particularly pronounced in Alabama and Florida in the South, in the four corners area (Colorado, Utah, Arizona, and New Mexico) in the Mountain Region of the West, and in Oregon and California in the Pacific Region of the West. The positive correlation pattern observed throughout the US proves that the PIA is a major determinant of obesity rates in the US. This finding suggests that targeting the PIA to lower high obesity rates will be effective in all areas across the US, yet the effectiveness of such an approach may be more substantial in the counties located in the South and West due to the fact that the relationship is more pronounced in these areas.

Similarly, the relationship between commute by car rate and obesity is positive in most counties except in Arkansas and Mississippi (Figure 5.14). But for most counties, the relationship is insignificant. The positive relationship is most pronounced in Iowa, Wisconsin, and Illinois in the Midwest and in the entire Northeast. This suggests that in order to reduce

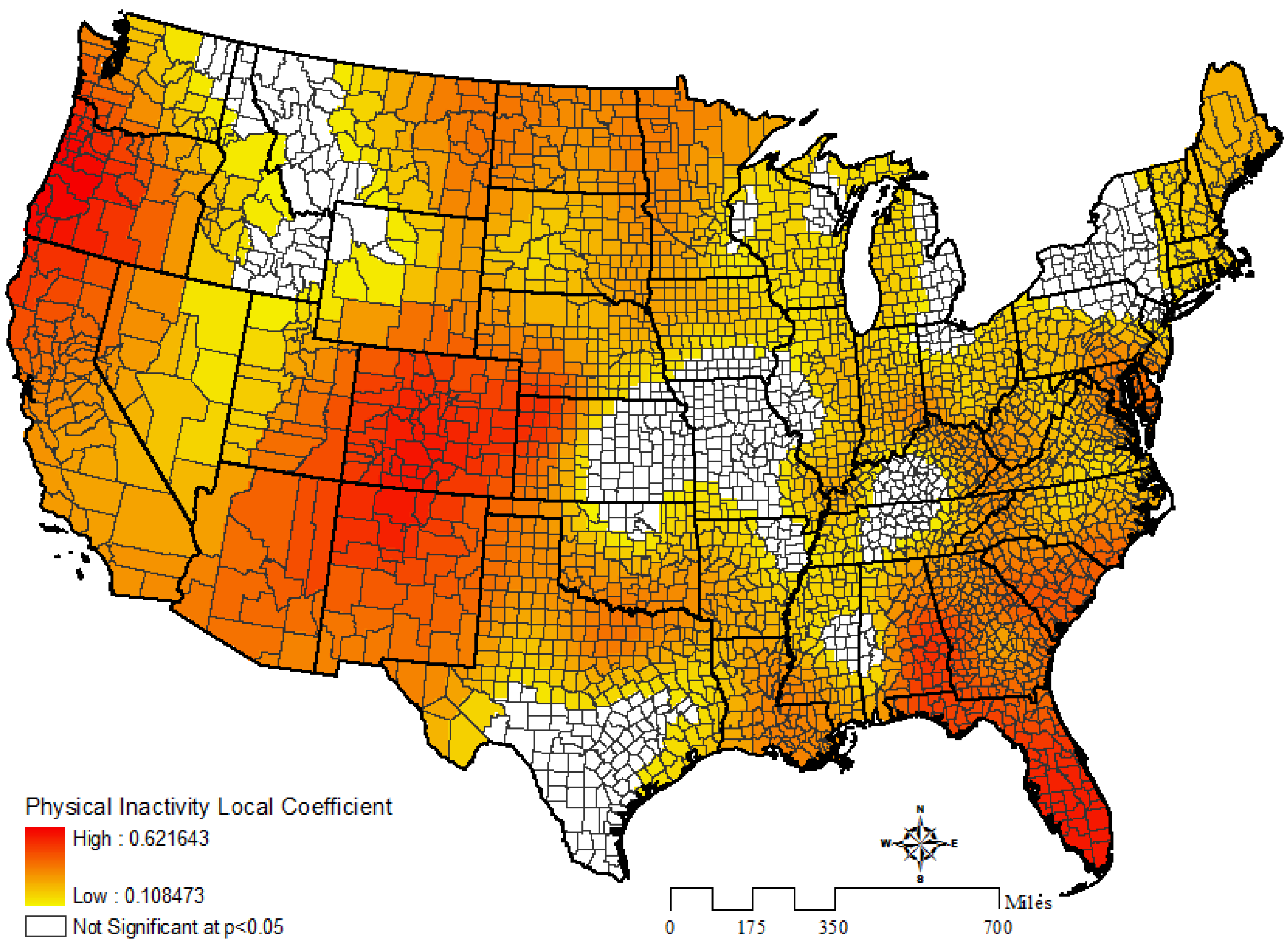

Figure 5.13 Significant and insignificant local GWR coefficient parameter estimates for physical inactivity (PIA) rates as a determinant of obesity prevalence at the county level in the US

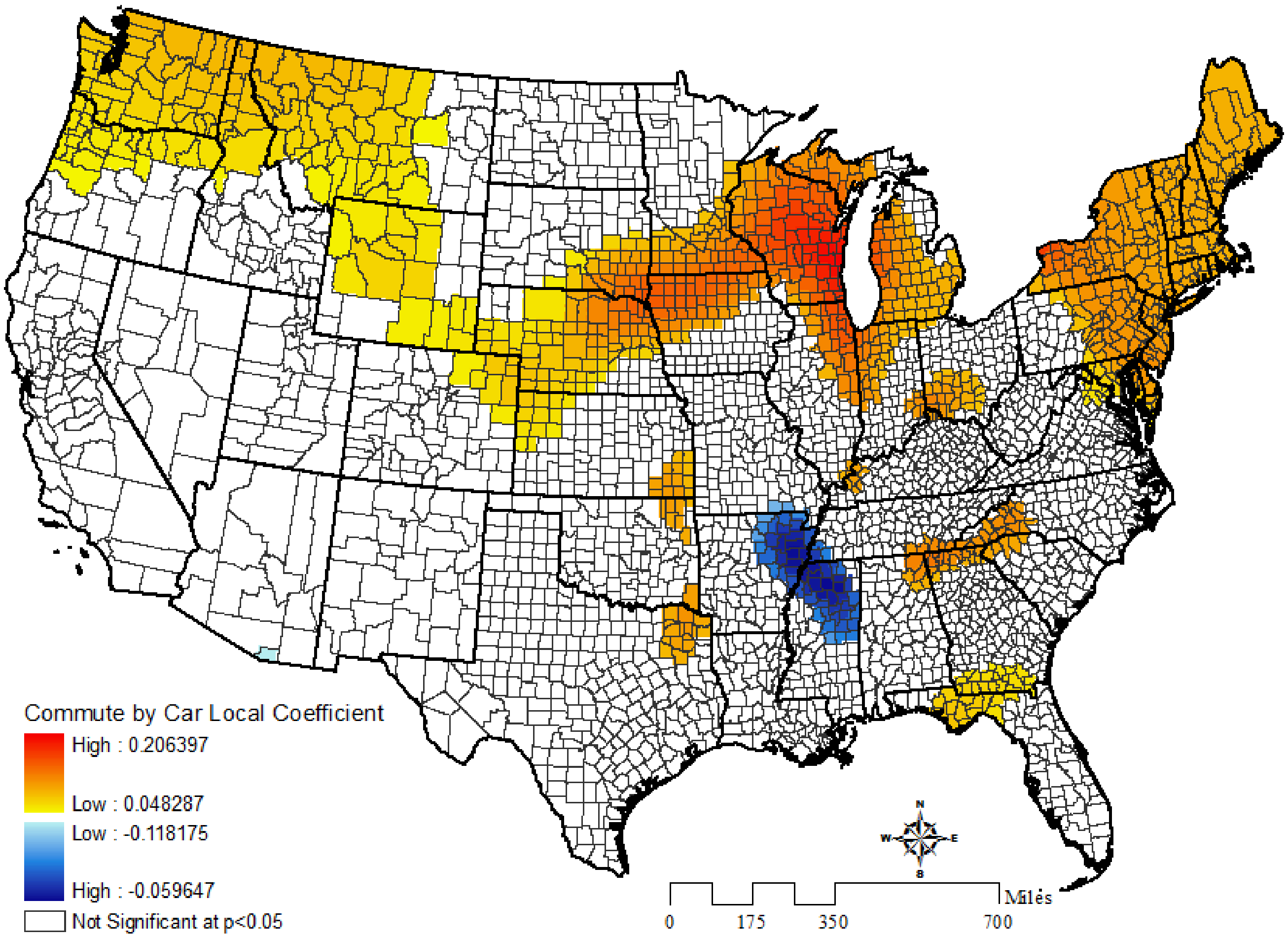

Figure 5.14 Significant and insignificant local GWR coefficient parameter estimates for commute to work by car rates as a determinant of obesity prevalence at the county level in the US

obesity rates, initiatives aiming to reduce commute by car rates may very well work in these areas. In other words, initiatives introducing and/or encouraging other modes of transportation (e.g., public transportation, bicycling, and/or walking) and therefore decreasing commute by car rates may be very practical to reduce obesity rates in these areas.

The natural amenities index is strongly and extensively associated with obesity in the West and partially in the South (Figure 5.15). There are both positive and negative coefficients. As seen in Figure 5.15, negative coefficients are predominantly higher than positive coefficients. The negative effect of the natural amenities index is most noticeable in the Texas Panhandle Region and its bordering counties in Oklahoma, Kansas, and Colorado, in Pennsylvania and New Jersey in the Northeast, in North Carolina and Virginia in the South, in Utah and Wyoming in the West, and in Montana and North Dakota in the Midwest. This suggests that in these areas, an increase in natural amenities may reduce obesity. Public health interventions should definitely consider the opportunities and limitations offered by the natural environment to promote physical activity and therefore reduce obesity rates especially in these areas, where there is an inverse association between natural amenities index and obesity. On the other hand, the positive effect is observed in Kentucky, Indiana, Illinois, and Missouri as well as Michigan and South Dakota in the Midwest. In these positively related areas, the relationship is in the unexpected direction.

The JER variable is negatively correlated with obesity most prominently in Florida in the South and in California in the West (Figure 5.16). The negative relationship is also observed in Maine, Massachusetts, New Hampshire, Vermont, New York, Pennsylvania, and New Jersey in the Northeast, in West Virginia, South Carolina, Kentucky, Alabama, Mississippi and in the DC area in the South, in Ohio, Wisconsin, Illinois, Michigan, Minnesota, and Iowa in the Midwest, and in Wyoming, Oregon, and Washington. Conversely, the JER variable is strongly and

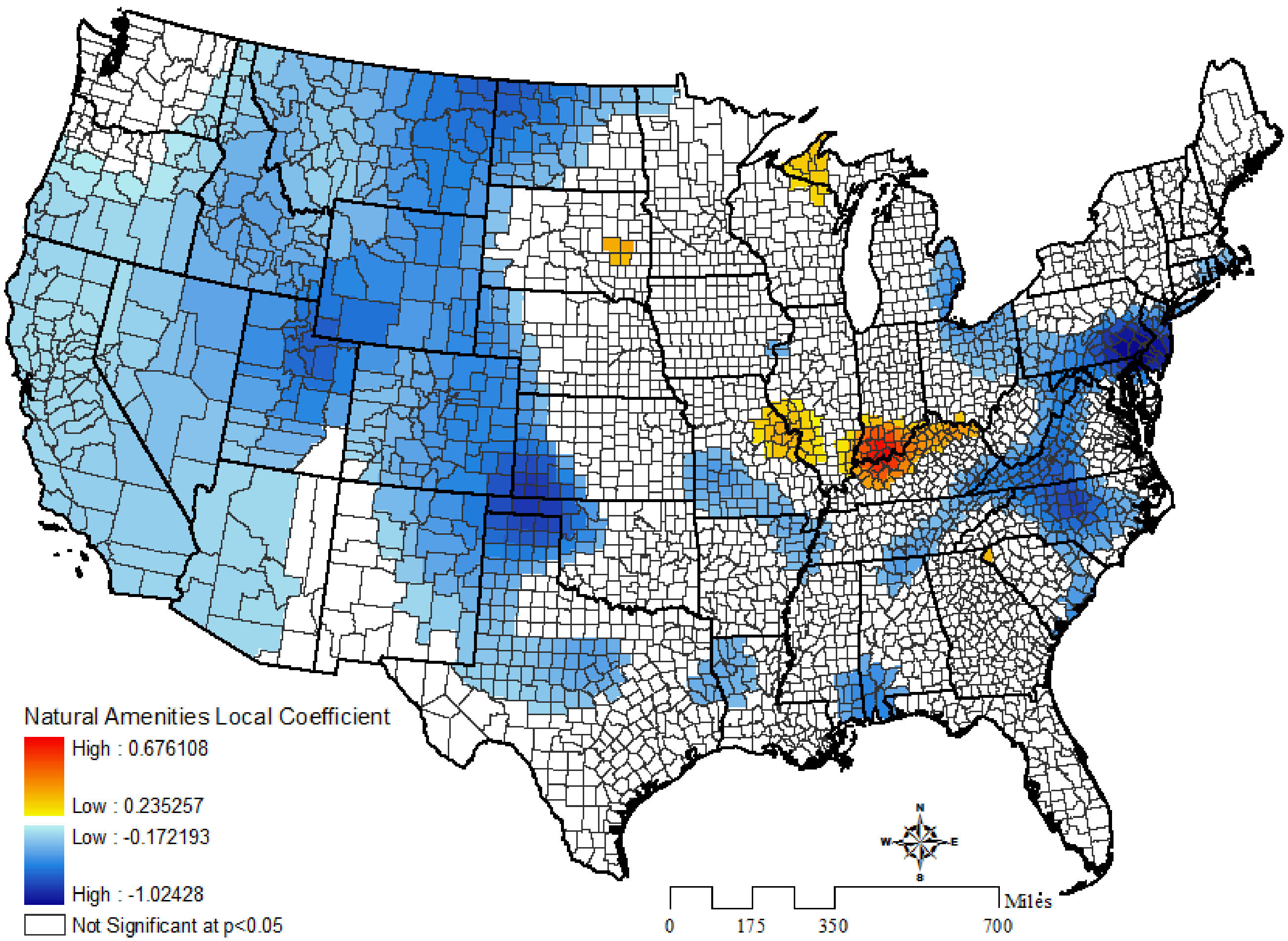

Figure 5.15 Significant and insignificant local GWR coefficient parameter estimates for natural amenities index as a determinant of obesity prevalence at the county level in the US

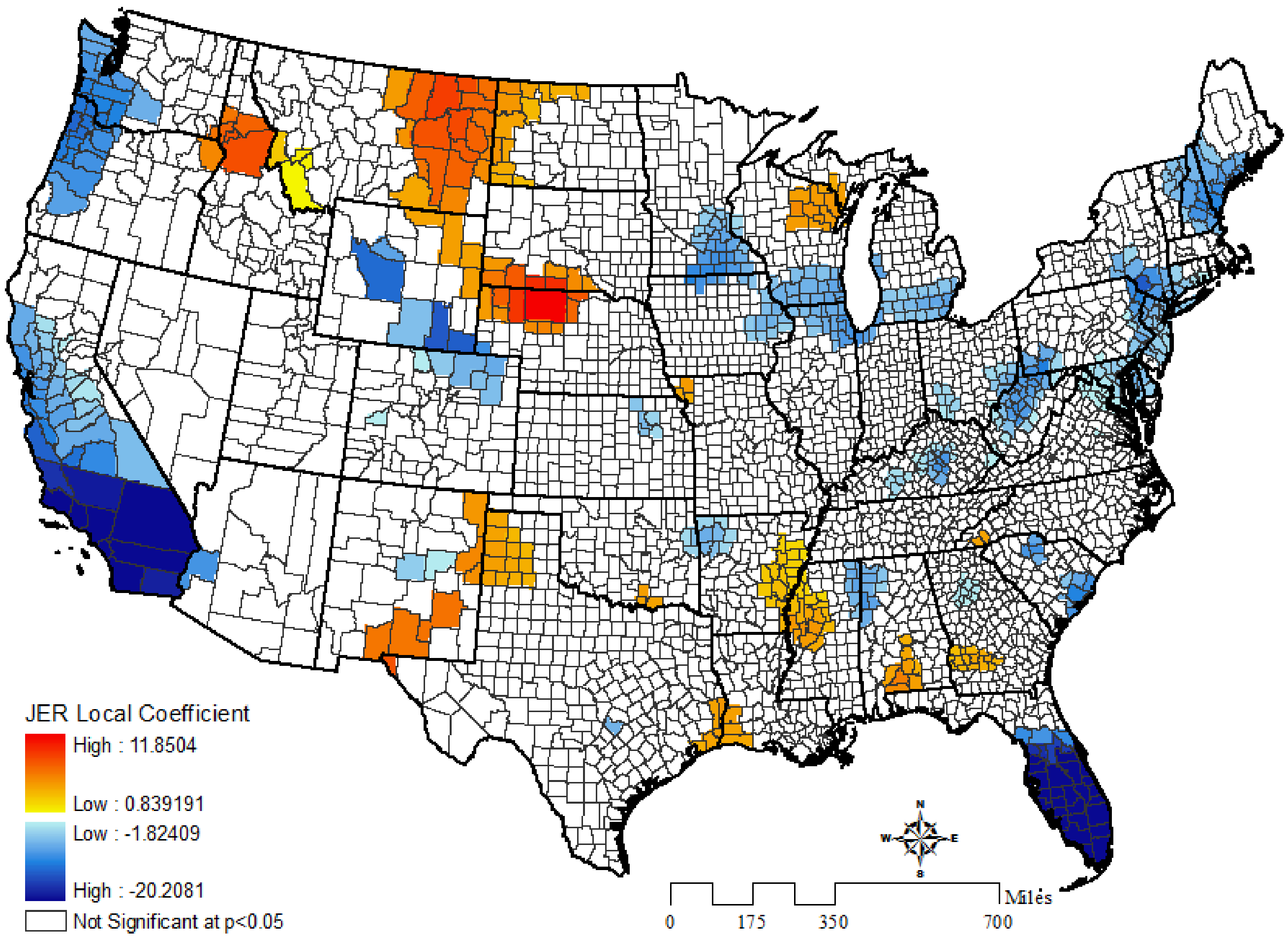

Figure 5.16 Significant and insignificant local GWR coefficient parameter estimates for ratio of jobs to employed residents (JERs) as a determinant of obesity prevalence at the county level in the US

positively correlated with obesity most prominently in Montana and Idaho in the West and in Nebraska in the Midwest (Figure 5.16). The positive relationship also exists most clearly in New Mexico in the West, in Wisconsin in the Midwest, and in Texas, Mississippi, Alabama, and Georgia in the South. It is notable that the negative relationship is primarily observed in the areas where large metropolitan areas and major cities exist while the positive relationship is mainly seen in rural areas. In this study, the JER variable is used as a proxy to capture the degree of mixed land use as discussed in detail in Chapter 3. Mixed land use is positively associated with street connectivity and negatively associated with automobile dependence. Therefore, these results indicate that in the negatively correlated areas, the higher the mixed land use (i.e., higher street connectivity and lower automobile dependence) is, the lower the obesity rates are. Such a finding also indicates that increasing JER may be used as a tool to lower obesity rates effectively in areas, where the relationship with obesity is negative.

The relationship between fast-food restaurants per capita and obesity is mainly negative. The negative relationship is observed in all US regions except the Northeast. The most noticeable negative relationship is observed in Wisconsin, Texas, and Kentucky, followed by Florida, Kansas, and Missouri (Figure 5.17). These areas are mostly rural areas. On the other hand, the positive relationship is observed in Illinois and Indiana in the Midwest, in Pennsylvania and New Jersey in the Northeast, and in South Carolina in the South (Figure 5.17). The positive effect is much more pronounced in Illinois and Indiana followed by South Carolina. Conversely, these positively related areas are mostly urban areas. The observed negative relationship between fast-food restaurants per capita and obesity is in the unexpected direction. As previously pointed out by Ahern et al. (2011), this negative association may be due to existence of a higher density of fast-food restaurants along interstates and other major highways in rural

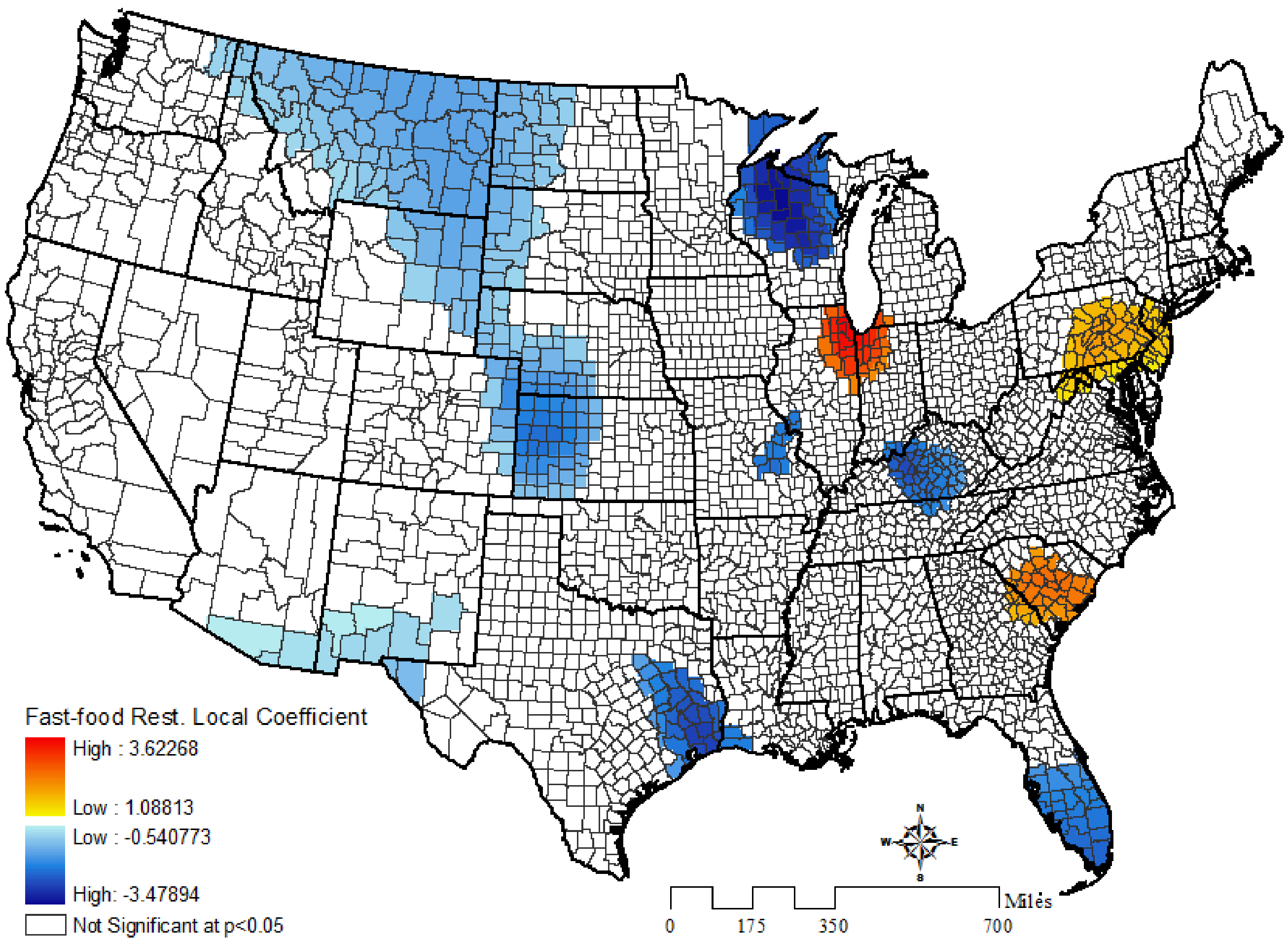

Figure 5.17 Significant and insignificant local GWR coefficient parameter estimates for fast-food restaurants per capita as a determinant of obesity prevalence at the county level in the US

areas particularly where the population density is low yet the obesity rates of the population is also low. In that regard, the patterns observed in Minnesota in the Midwest and in Texas in the South can be good examples. In other words, the supply of these fast-food outlets may be serving more travelers and/or non-residents, rather than the residents of the county.

The effect of poverty on obesity is mainly positive and observed across the entire US, except in most counties in the state of Texas and to some degree in some counties in the Midwest. This positive association is most pronounced in the West, South (except Texas), and Northeast. In particular, the magnitude of the poverty's effect on obesity is most visible in Maine in the Northeast, in Florida, South Carolina, and North Carolina in the South, as well as in California in the West (Figure 5.18). In contrast, the effect of poverty is most prominently negative in the peripheral counties of Oklahoma, Kansas, Colorado, and New Mexico and in some peripheral counties of New York and Pennsylvania (Figure 5.18). The positive sign is the expected direction of this relationship. Such consistent and strong positive correlation indicates that in these positively related areas, reducing poverty may help reduce obesity rates, yet such an approach will be more effective in some areas than others based on the strength of the relationship. On the other hand, counties showing a negative relationship with obesity may not benefit from poverty related (poverty reducing) policy changes. Furthermore, the global OLS model used in this study, as well as previously conducted studies from the literature, have confirmed a positive relationship between poverty and obesity. However, the findings from the local GWR model showed that there are also local patterns in the opposite direction (negative) and such finding is quite interesting. A detailed exploration of areas with negative coefficients may be interesting for future studies.

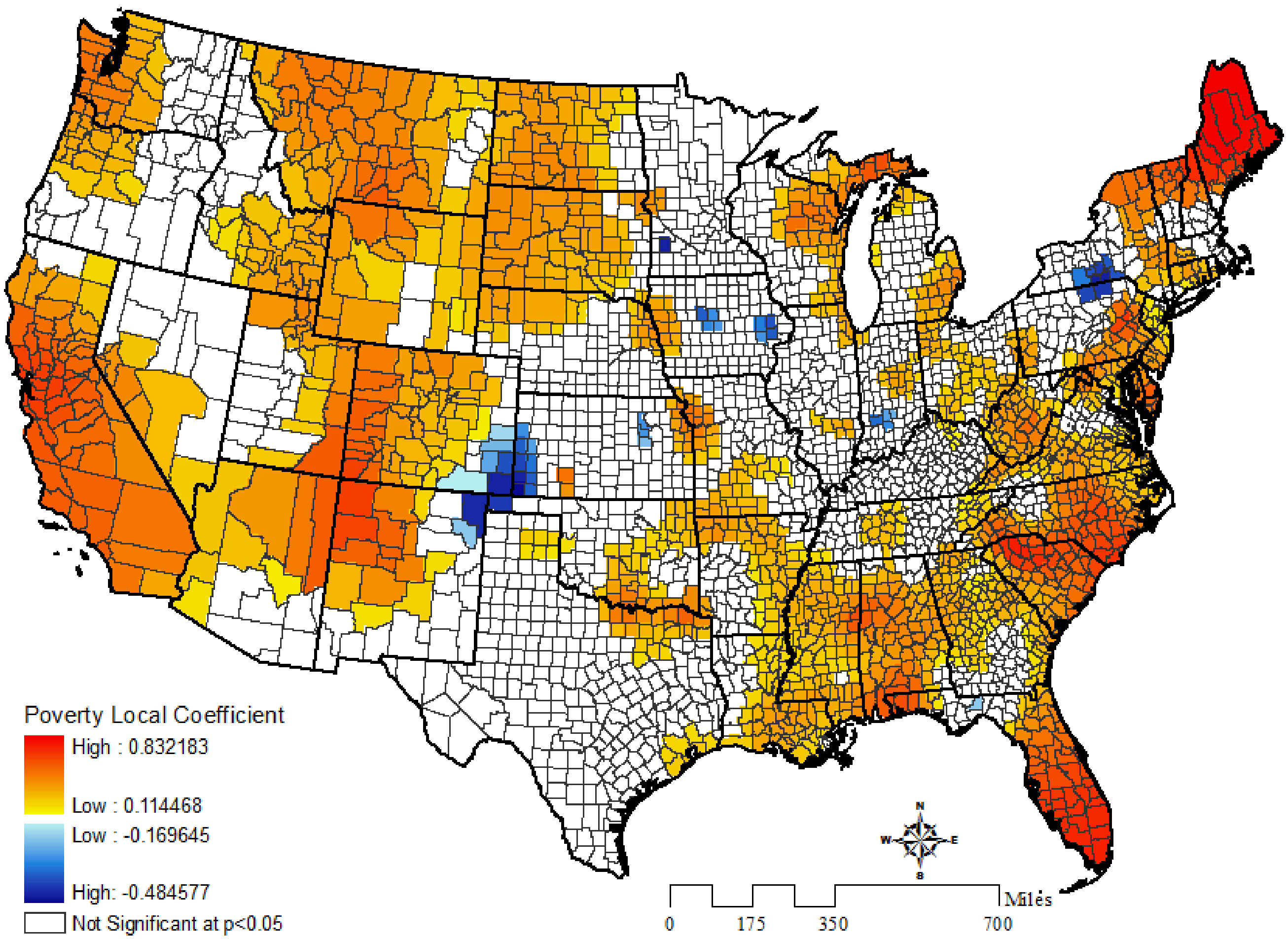

Figure 5.18 Significant and insignificant local GWR coefficient parameter estimates for poverty rates as a determinant of obesity prevalence at the county level in the US

The African-American population rate is strongly and positively correlated with obesity in most counties across the US (Figure 5.19). The positive effect on obesity is quite significantly consistent across the entire South. In addition, the positive relationship is observed in Michigan, Indiana, and Missouri in the Midwest and in Arizona, Wyoming, Montana, Idaho, Washington, and Oregon in the West. The magnitude of the positive effect is most pronounced in Montana and its surrounding counties in the West. On the other hand, the negative relationship is observed most distinctly in the Midwest, especially in North Dakota and South Dakota (Figure 5.19). In addition, the negative relationship is also seen in the West (Wyoming) and Northeast (Maine). In conclusion, the consistent pattern of the positive African-American population coefficients, especially in the South, proves that racial disparities are one of the important factors behind the obesity issue in the US.

The relationship between high educational attainment rate and obesity is negative across the entire US, except in some counties in North Dakota in the Midwest (Figure 5.20). The magnitude of the effect is most pronounced in the entire Northeast followed by most of the states in the Mountain Region in the West. This consistent and strong negative correlation shows that education may be a great tool to combat obesity in these areas. It is obvious that such an approach will be more effective in some regions than others due to the varying coefficient values. For instance, the success rates may be much higher in the Northeast compared to the other regions, since the magnitude of the effect is most pronounced in the Northeast.

Figure 5.21 shows significantly negative coefficients of the rural population rate in the West as well as in some inland areas in the Midwest and the South. Based on the magnitude of the negative effect, the counties in Washington, Oregon, California, Missouri, Nebraska, and Kansas are distinct by their strong negative effect. On the other hand, there are significantly

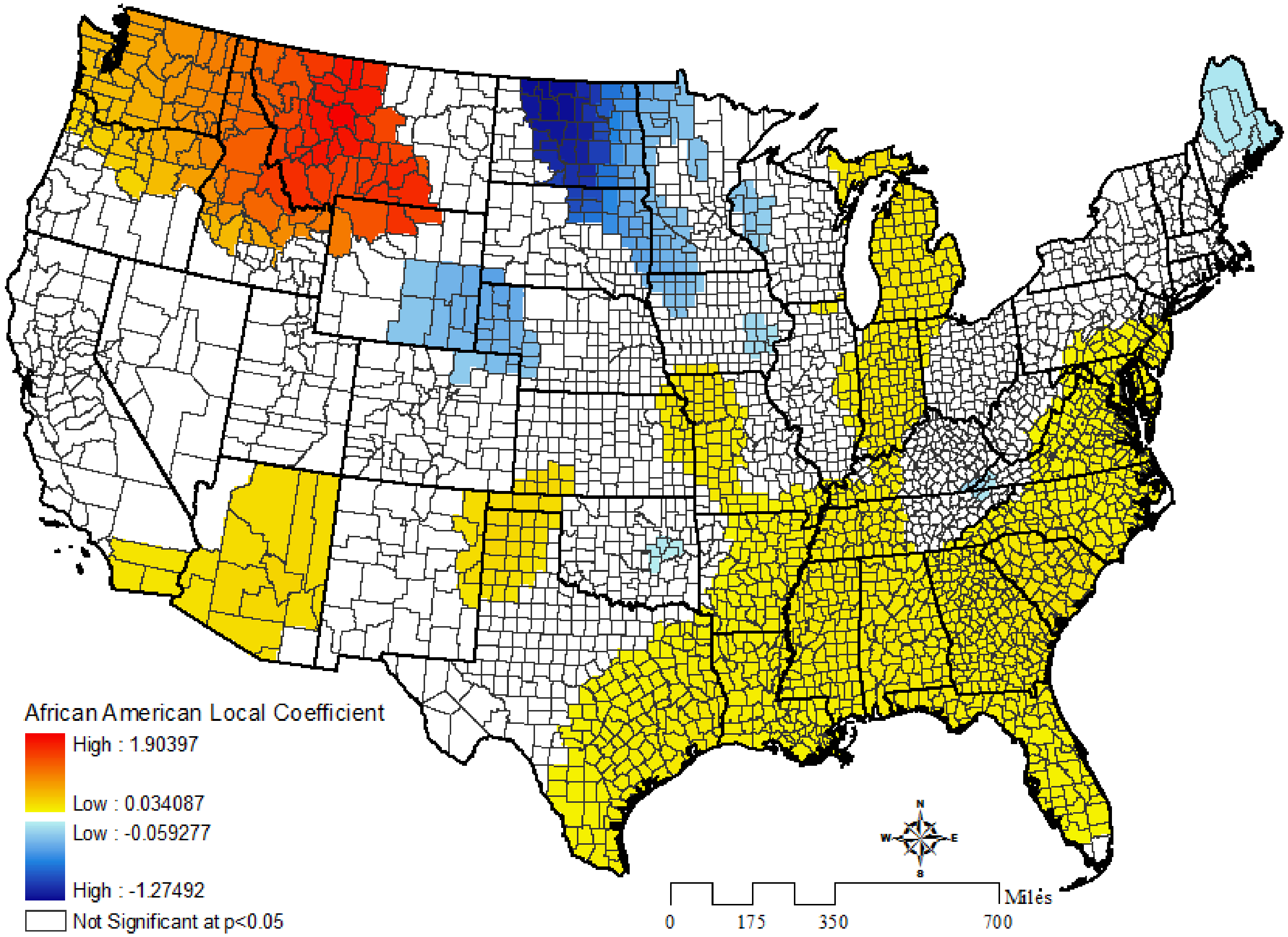

Figure 5.19 Significant and insignificant local GWR coefficient parameter estimates for African American population rates as a determinant of obesity prevalence at the county level in the US

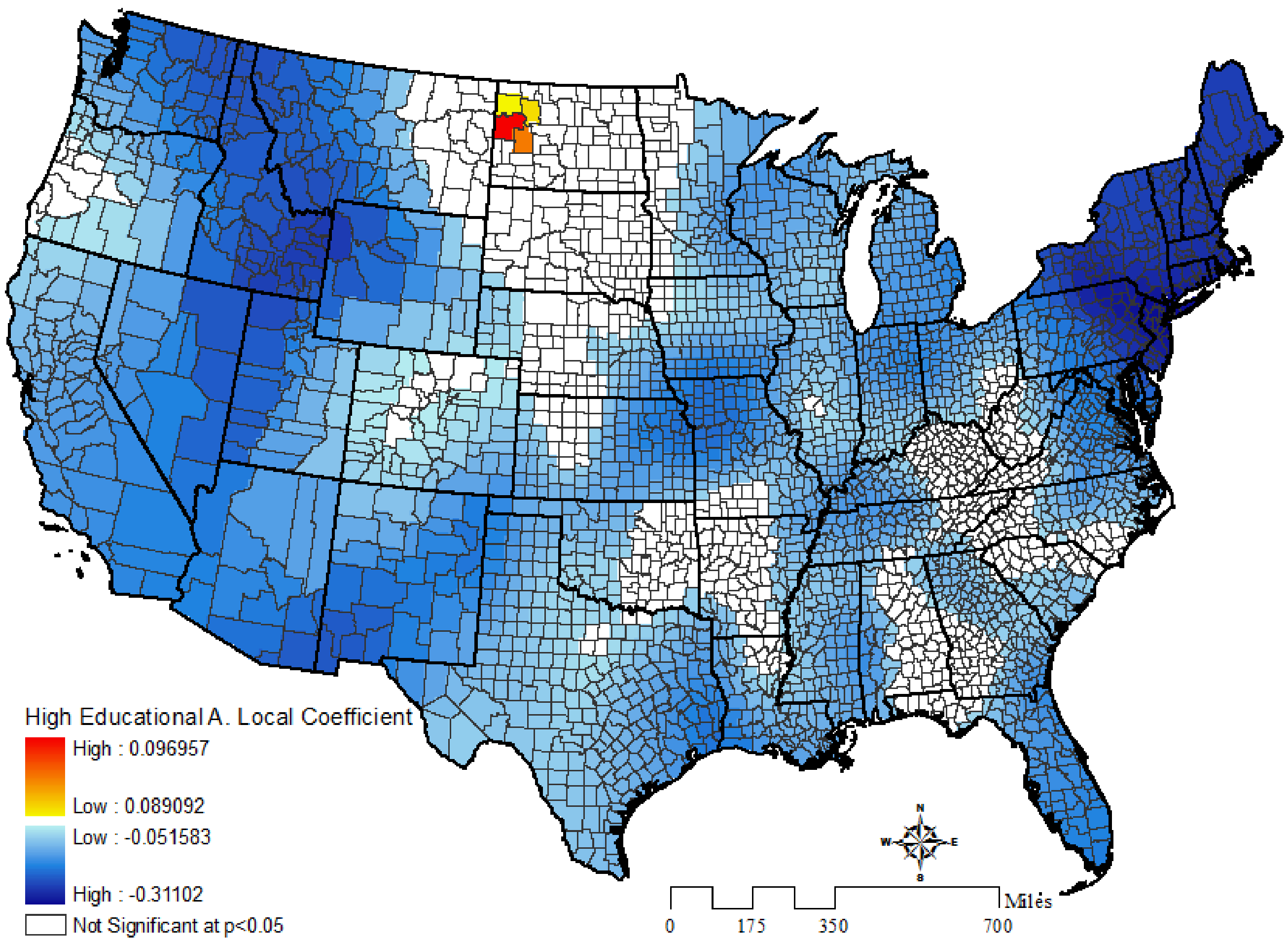

Figure 5.20 Significant and insignificant local GWR coefficient parameter estimates for high educational attainment rates as a determinant of obesity prevalence at the county level in the US

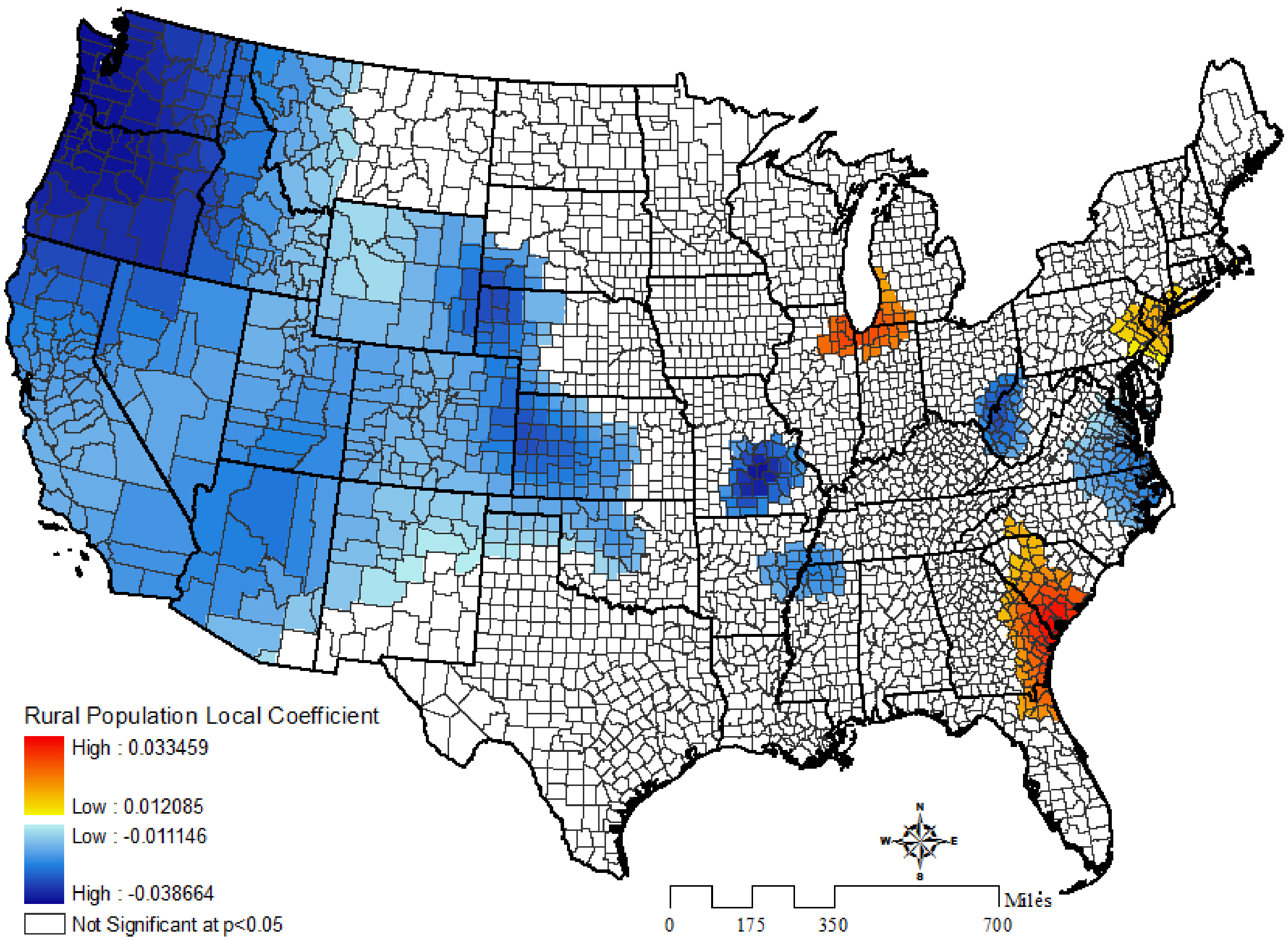

Figure 5.21 Significant and insignificant local GWR coefficient parameter estimates for rural population rates as a determinant of obesity prevalence at the county level in the US

positive coefficients of the rural population rate in the South (Georgia and South Carolina), Midwest (Illinois and Indiana), and Northeast (Pennsylvania and New Jersey). According to the government definition, all areas that are not urban, suburban, or metropolitan are classified as rural, by default. Traditionally, rural areas are often agricultural (HRSA, 2015). The varying relationship (negative vs. positive) at the local level might very well be due to the lifestyle of individuals living in these areas classified as rural areas. For instance, in correctly classified rural areas, agriculture or labor intense jobs are the primary industries. In these rural areas, it is also more typical that people live or work on farms or ranches. In other words, people generally work in heavy labor industries, which naturally increase the energy expenditure. Such factors may contribute to the negative relationship with obesity. On the other hand, some rural areas, consisting mostly of suburban areas, are, in general, incorrectly classified as rural areas (HRSA, 2015) and the lifestyle of the suburban population in such areas is totally different (e.g., more car dependent and less active). Therefore, such a factor may be attributed to the positive relationship with obesity in such incorrectly classified areas.

The relationship between native resident rate and obesity is positive in most counties (Figure 5.22). The positive relationship is mostly observed in the West, more specifically in the Mountain Region, Midwest, and South, while the negative relationship is observed in the South and Midwest (the West North Central Region) (Figure 5.22). There is no significant relationship, either positive or negative, in the Northeast. The strength of the positive relationship is especially noticeable in Kentucky and Tennessee in the South followed by Oklahoma, Wisconsin, Michigan, South Dakota, and Colorado. Conversely, the magnitude of the negative effect is more pronounced in Texas, Oklahoma, and Virginia in the South.

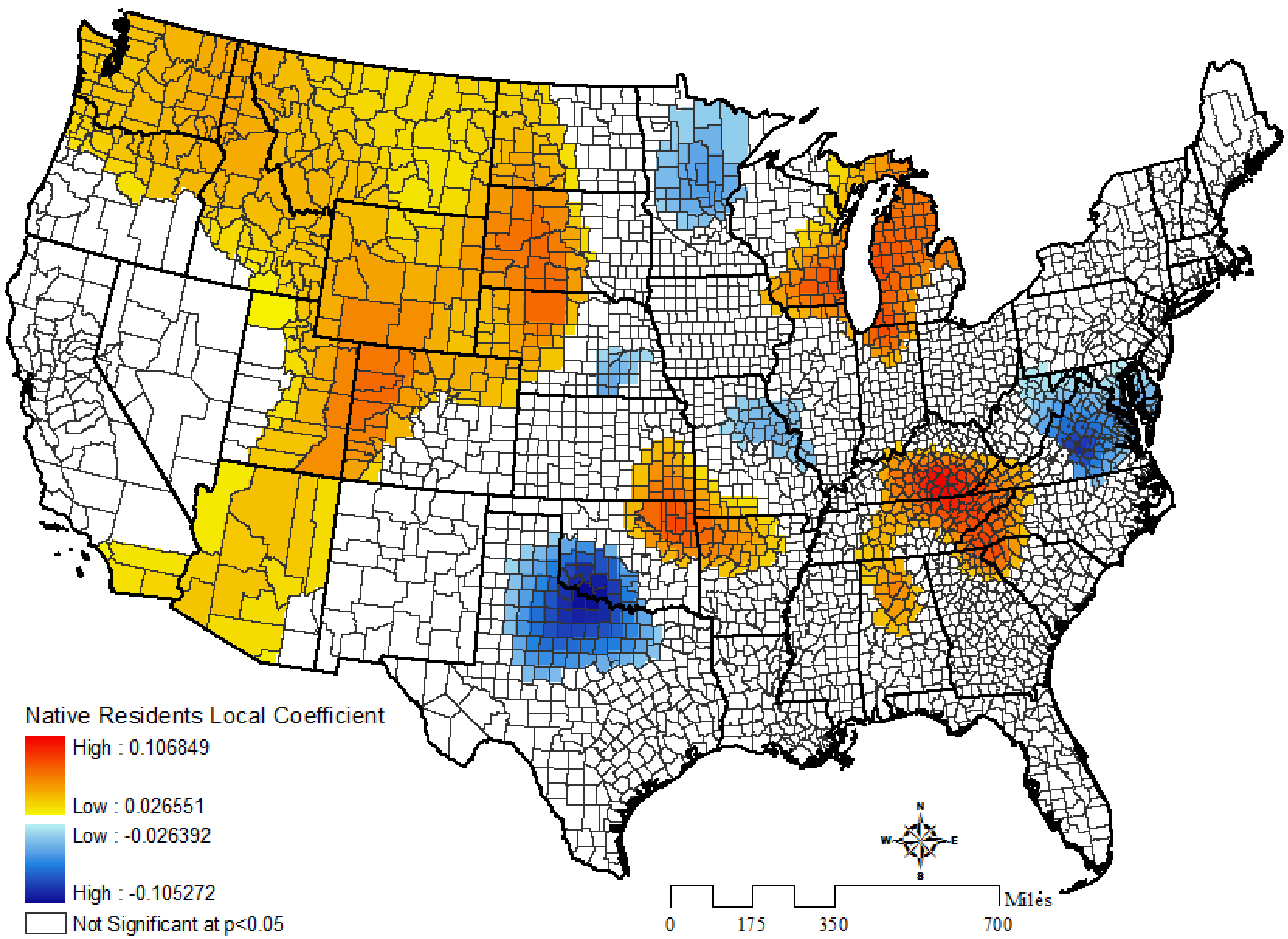

Figure 5.22 Significant and insignificant local GWR coefficient parameter estimates for native residents rates as a determinant of obesity prevalence at the county level in the US

In addition to the above results, Figure 5.23 shows the overall distribution of significant and insignificant parameter estimates. It indicates that overall, 43% of all parameter estimates (cumulatively added for all variables) are significant at the county level. Furthermore, Figure 5.24 illustrates the percentages of significant and insignificant parameter estimates separately for each variable used in this study. It points out that with 82% of all parameter estimates being significant, the PIA's effect on obesity is the strongest among all independent variables tested, while fast-food restaurants per capita is weakest, with only 16% of all parameter estimates being significant.

A further exploration of the results of the GWR analysis by US regions is also possible regarding the significance of parameter estimates for all variables used in this study. In the Northeast, high educational attainment is by far the most important predictor of obesity, while percentage of native residents is by far the least important predictor (Figure 5.25). The effects of the three variables (commute to work by car, PIA, and poverty) are also significant in most counties in the Northeast (Figure 5.25). Similarly, high educational attainment and PIA are the most important predictors of obesity in the Midwest, whereas rural population and JER are the least important predictors (Figure 5.25). In the Midwest, the effects of the three variables (commute by car, African-American population, and poverty) are also significant in most counties (Figure 5.25).

In the South, the estimate for the local regression coefficients of the PIA is the most important predictor of obesity, while the estimate of fast-food restaurants per capita is the least important predictor (Figure 5.25). The effects of the African-American population, high educational attainment, and poverty are also significant in most counties in the South (Figure 5.25). In the West, high educational attainment, rural population, PIA, and natural amenities

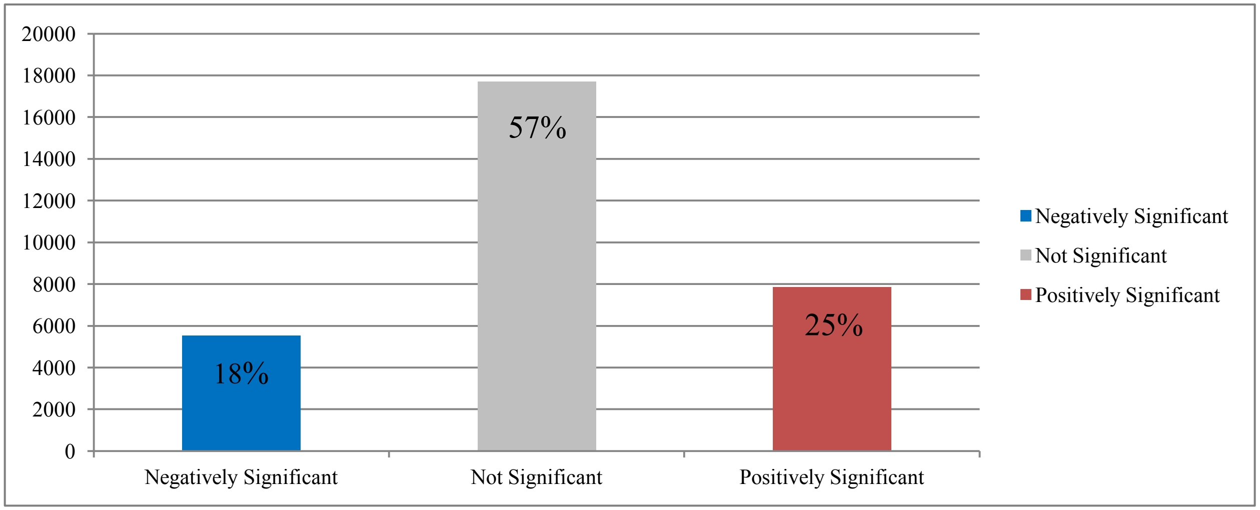

Figure 5.23 Summary of statistical significance of local GWR coefficient estimates (%) across all US counties

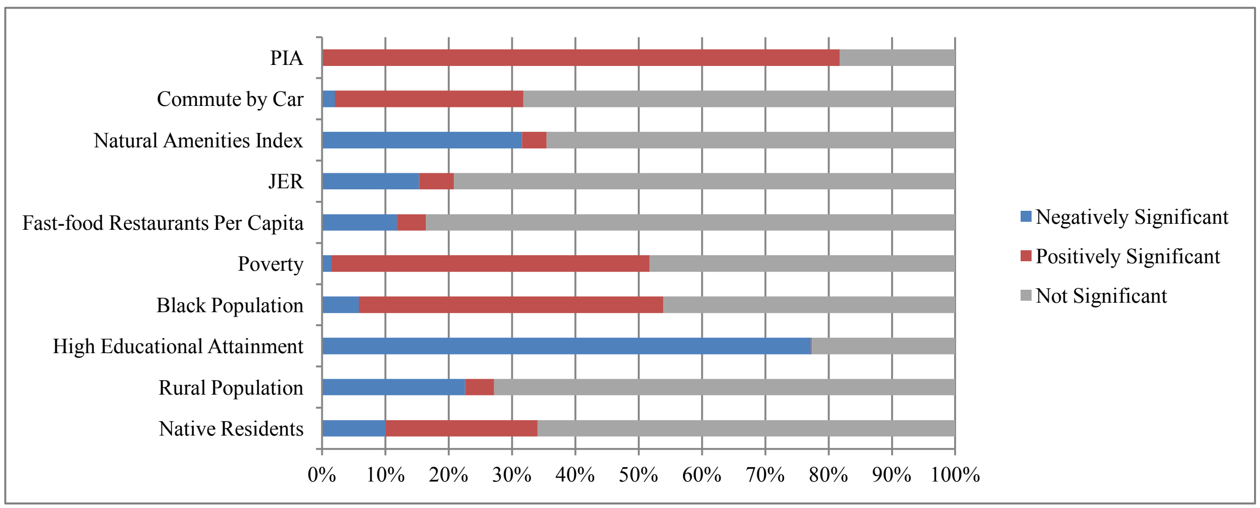

Figure 5.24 Summary of statistical significance of local GWR coefficient estimates by variables (%)

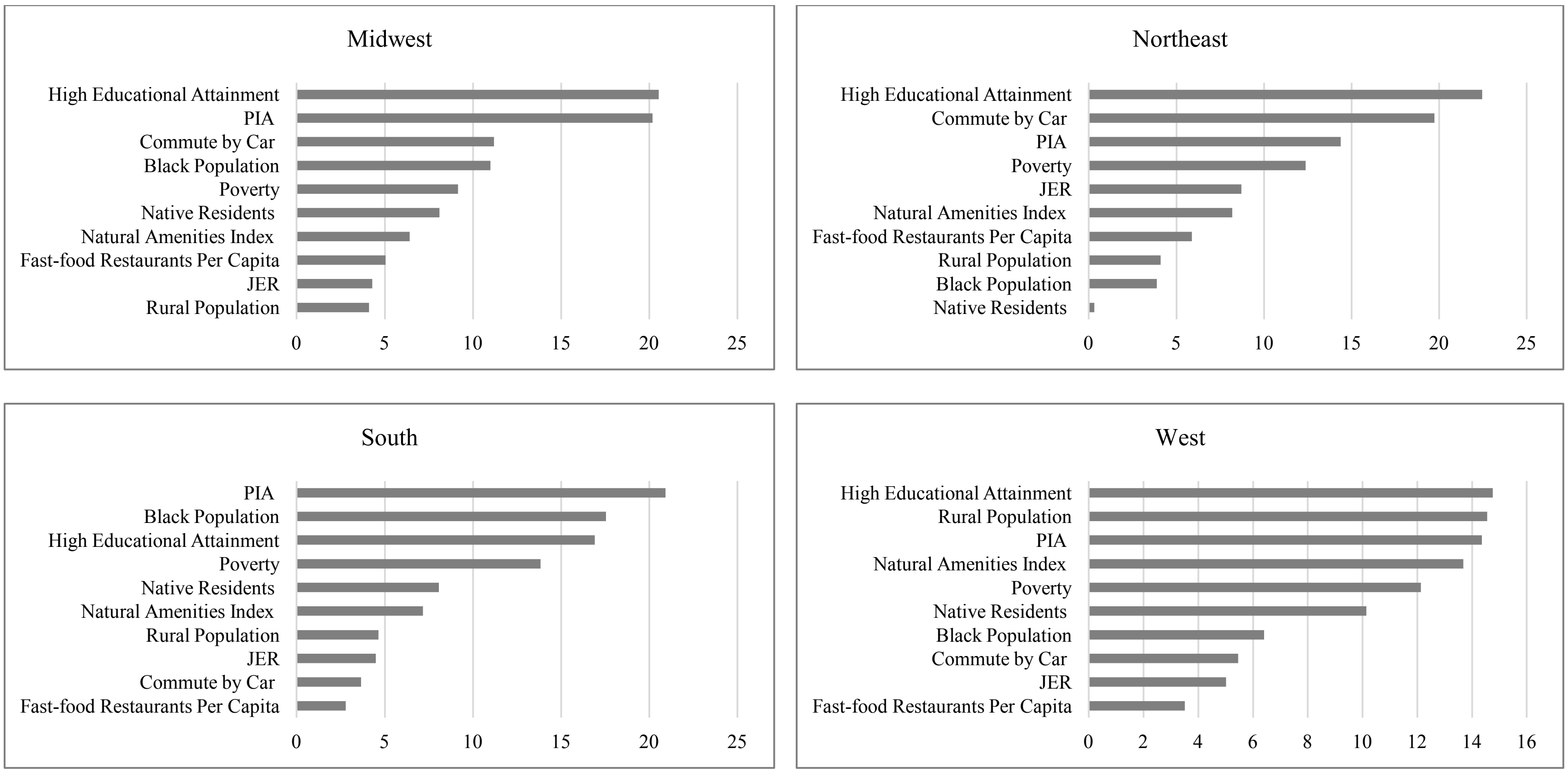

Figure 5.25 Summary of statistical significance of local GWR coefficient estimates by variables and regions (%)

index are the most important predictors of obesity, whereas fast-food restaurants per capita is the least important predictor (Figure 5.25). In the West, the effect of poverty and percentage of native residents are also significant in most counties (Figure 5. 25).

Consequently, as far as global models are concerned, the local GWR analysis has demonstrated that a spatial regression model adds significantly to the understanding and interpretation of spatially varying obesity rates in the US. The use of a GWR model allowed for an assessment of spatial heterogeneity, when exploring the relationships between the independent variables and obesity rates at the county level. Geographically weighted estimations provided a better fit to the data. Independent variables as well as obesity rates showed strong local variation, indicating that policymakers may best address problems at the local level. Further insights that can be deduced from above results are included in the discussion section, which follows next.

5.4 Discussion

A major contribution of this study is examining and revealing geographic variability in the association between obesity rates and various explanatory variables both nationwide and regionally at the county level for the entire US. The main strength of this study is the use of the GWR in the analysis of the spatial distribution and associations of the obesity prevalence at the county level in the US. As a modeling approach the 'local' GWR is a novel model by providing powerful tools to spatially disaggregated 'global' model statistics. Although a local GWR model may offer valuable insights concerning the risk factors and their confounding relationships with the obesity issue, only a few studies have utilized this approach (e.g., Xu and Wang, 2015; Lee et al., 2014; Chi et al., 2013; Wen et al., 2010). Most of these studies typically focused on one type of association (e.g., built environmental, food environment, etc.). However, this current study investigated a variety of factors, including behavioral, natural and built environmental, socioeconomic and demographic, and population characteristics to get a more comprehensive picture of the obesity issue and its spatial associations in the US both nationwide and regionally. Particularly, the use of variables such as commute to work by car, ratio of jobs to employed residents (JER), and percentage of native residents have not yet been included in the obesity literature.

This study verified the superiority of the 'local' GWR over the 'global' OLS model in spatial analysis of the relationships between obesity and numerous confounding variables. Such superiority was mainly confirmed by the spatial variation of relationships across the study area. The global OLS regression model for the entire US in this study showed that 64.8% of county level obesity prevalence was explained by the confounding variables used. Similarly, the global OLS models for the US regions showed that in the Northeast 73%, in the South 63.7%, in the West 55.8%, and in the Midwest 33.2% of county level obesity prevalence was explained by the same set of variables. An improvement in the global OLS model was valid only for the Northeast. A geographic variability in the association between obesity rates and explanatory variables were detected to a certain extent by conducting the OLS for the US regions. On the other hand, an overall percentage of 77.8 of county level obesity prevalence was explained by the local GWR model with the same set of variables. More specifically, at the individual county level, the explained obesity prevalence ranged from 11% to 87% in the GWR model. These local R^2 measures, as well as the regression coefficient estimates from the local GWR model, were mapped as surfaces and examined in detail in the results section (Figures 5.12-5.22). Examining the spatial distribution of local R^2 measures and coefficient estimates as a surface revealed where and how much variation was present. Understanding of such variations provided insights that may be used to inform local policymakers of how to best address the obesity issue at the county level.

All regression coefficients obtained from the local GWR model showed evidence of nonstationarity (Table 5.8). This suggested that the strength of associations between obesity and each of the independent variables varied depending on the spatial location. Accordingly, the observed patterns in the maps, directions of the relationships (positive vs. negative and/or expected vs. unexpected), and possible implications deduced from those patterns were discussed in detail in the results section of this study.

The confounding variables PIA, high educational attainment, African-American population, and poverty were identified as the top four variables by having relatively stronger effects in explaining obesity rates at the county level. By looking at the PIA's relationship with obesity it can be concluded that PIA rates and, consequently, individuals' caloric expenditures are the most significant determinants of obesity rates. On the other hand, natural and built environments partly influences physical activity (by encouraging or discouraging and/or creating activity spaces) and thus caloric expenditure, while the fast-food environment controls dietary habits and consequently caloric intake to a certain extent. In this study, the natural amenities index and the JER are used as measures of the physical characteristics of a county, while fast-food restaurants per capita is used as a measure of the food environment of a county. When the natural amenities index, JER, and fast-food restaurants per capita's relationships with obesity are scrutinized independently, it is concluded that the effect of the natural amenities index and/or JER on obesity rates is more prominent than fast-food restaurants per capita. In other words, natural amenities and JER, by influencing activity spaces and/or potentials for physical activity and therefore caloric expenditure, are more important determinants of obesity rates than fast-food restaurants that shape food outlets and thus opportunities for caloric intake. These three variables define opportunities and/or accessibility to places (for physical activity and/or fast-food); however, accessibility alone does not explain much about obesity rates. Accessibility becomes meaningless unless awareness is present. For instance, although physical activity can be largely affected by an individual's natural and/or built environment, it is also influenced by an individual's awareness and desire to be active. A similar approach is true for dietary habits. Therefore, awareness is a very critical factor and it often increases as the educational attainment increases. The significant and widespread local GWR coefficient patterns concerning the high educational attainment rate's correlation with obesity across the US support and strengthen this idea as can also be seen in Figure 5.20. To support a similar idea, the African-American population rate's strong correlation with obesity across the US can be further scrutinized. For instance, African-American populations in general have lower income and lower education and usually inhabit locations with greater access to fast-food restaurants, corner stores, and unsafe or undesignated urban open spaces, while they usually have limited access to preferable outlets such as healthy restaurants and supermarkets, gyms, or safe parks and playgrounds (Rossen and Pollack, 2012; Abercrombie et al., 2008; Franco et al., 2008; Gordon-Larsen et al., 2006; Zenk et al., 2005; Block et al., 2004). All these socioeconomic and environmental disadvantages translate into limited access to various resources and opportunities, as well as awareness about how or how not to use such resources and opportunities. In that regard, as shown previously in Figure 5.18 (poverty relationship with obesity) and Figure 5.19 (African-American population relationship with obesity), the significant positive associations observed are convincing representations, especially in the US South.

Further elaboration of the local GWR results revealed several particularly important points. First, an examination of the frequency of the significant parameter estimates at the local level revealed that overall, 43% of all parameter estimates are significant, while 57% of all parameter estimates are insignificant (Figure 5.23). Of those significant parameter estimates, 18% are negatively significant, while 25% are positively significant (Figure 5.23).

Second, with the exception of the PIA (82% significant), high educational attainment (77% significant), African-American population (54% significant), and poverty (52% significant), the majority of the regression coefficient estimates for all other independent variables are insignificant (Figure 5.24). Among all variables, fast-food restaurants per capita (16% significant) has the least significant coefficient estimates. These results indicate that among all variables, the effect of the PIA is by far most strongly associated with obesity, whereas the effect of fast-food restaurants per capita is least strongly associated. It is also quite clear from Figures 5.13 through 5.22 that positive or negative global effects of independent variables on obesity do not hold across all counties. Therefore, it confirms that it is quite crucial to examine variation at the local scale through GWR, rather than exploring this variation at the global scale only, by utilizing OLS or similar global models.

Third, the effects of the remaining variables, which are not mentioned above, are rather low based on the significant parameter estimate frequencies. However, the variables natural amenities index (35% significant), percentage of native residents (34% significant), and commute by car (32% significant) are especially worth mentioning, since their effect on obesity is still quite noteworthy (Figure 5.24).

Fourth, global parameter estimates obtained from OLS would have masked much of the variation at the local level, if the local GWR model were not utilized. Even conducting the global OLS for the US regions helped reveal only some of the variation. For instance, nationwide, the global parameter estimate for the PIA is 0.287, although parameter estimates at the local level range from 0.031 to 0.609. Similarly, the global parameter estimate for high educational attainment is -0.117, while the local parameter estimates range from -0.248 to 0.059 or as in the case of the African-American population, the global estimate is 0.064, whereas local estimates range from -0.691 to 1.272.

Last but not least, overall by regions and variable groups, the effect of socioeconomic and demographic variables on obesity is the most important group for all regions (highest in the South), except in the Northeast (Figure 5.26). In the Northeast, the effect of behavioral variables is the most pronounced among all groups of variables. On the other hand, the effect of natural and built environmental variables is most important in the Northeast followed by the West, while its effect is least important in the South (Figure 5.26). Furthermore, in the West, the effect of different variable groups, except the socioeconomic and demographic variables, is almost equally important in predicting obesity, while such variable mix cannot be found in any other region. These varying associations and patterns across the US regions suggest that targeting a certain factor (variable) to lower high obesity rates will not be equally effective in all areas across the US simply due to the fact that one variable or group of variables may be a more or less important predictor of obesity in some regions, states, and/or counties, than in others. In other words, the global model "one fits all" approach will not be the most effective approach to tackle the issue of obesity in all US regions. Also to further elaborate, similar facts and patterns can be obtained by dissecting states independently to distinguish what variables or group of variables are more or less important in predicting obesity in different states.

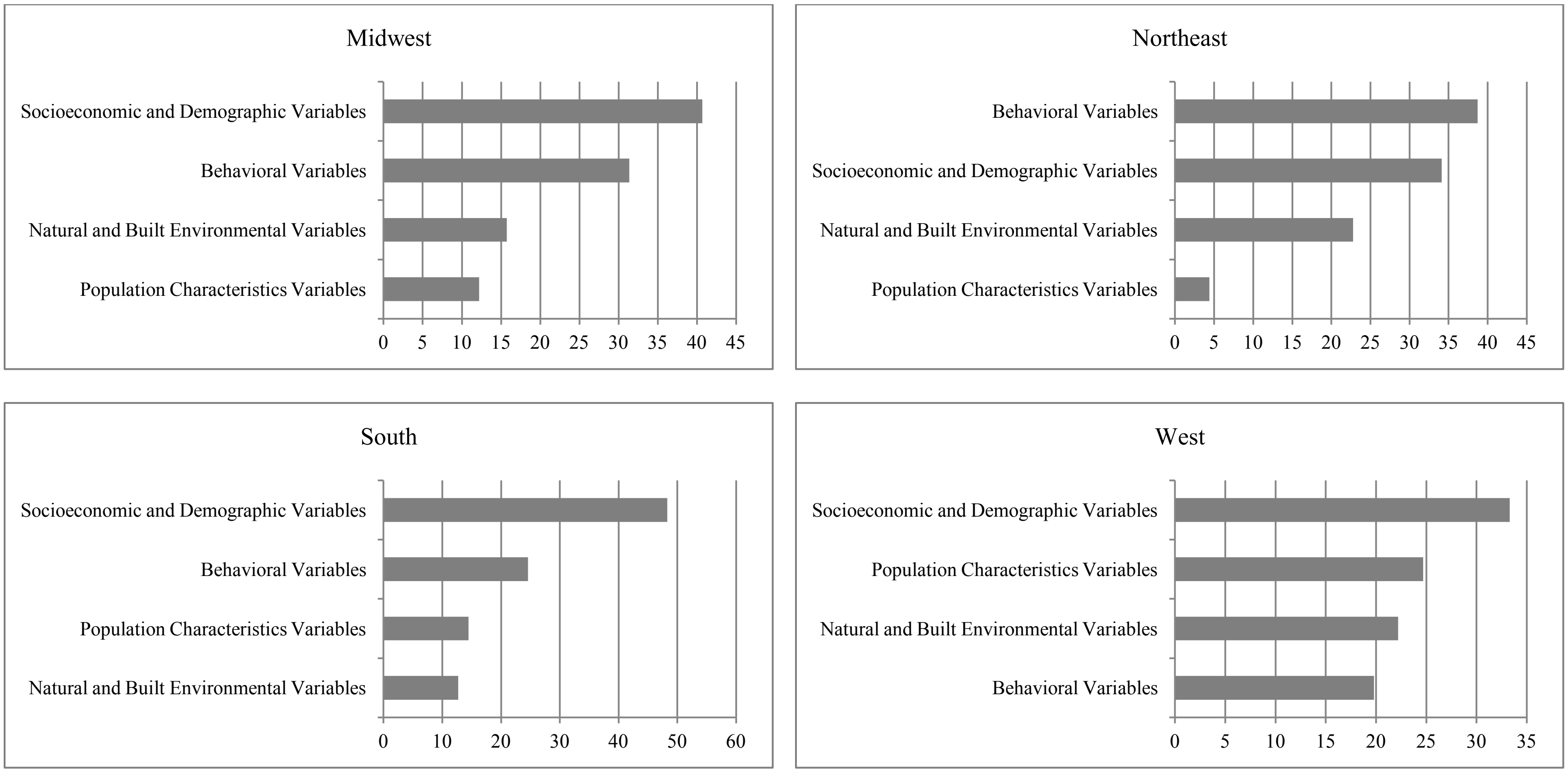

Figure 5.26 Summary of statistical significance of local GWR coefficient estimates by variable groups and regions (%)

On the other hand, this study was subject to some limitations, as well. First of all, heights and weights that the BRFSS use to calculate BMI values were based on telephone self-reported information. It is known that both men and women may underreport weight and men may over-report height in telephone surveys (Merrill and Richardson, 2009). Therefore, the prevalence of obesity rates used in this study should be considered somewhat conservative. Nonetheless, the very same BRFSS data are the most widely used obesity data and also assumed to present the best picture of the obesity epidemic in the US (BRFSS, 2014). Second, the spatial patterns of obesity identified may change depending on the spatial scale used, an issue also known as the Modifiable Areal Unit Problem (MAUP). Similar to most geographic research, the results of this study were scale dependent, as well. Therefore, the findings were significant and important at the scale of counties and county equivalents within the US as a whole.

There were also limitations to the GWR, as well as the findings of this study. All types of spatial analysis that generate localized and cartographic outputs are subject to edge effects and the GWR is no exception. Edge effects occur when counties are located at the outer boundaries of the US (i.e., US borders with Canada and Mexico). These counties do not have the influence of neighboring counties all around them, compared to the counties not located along the border areas. In regard to limitations to the findings, the local R^2 accounted for 11% to 87% of county level obesity prevalence by using the 10 independent variables included in the GWR model. This means that in the counties where the local R^2 values are low, there must have been missing important factors that are associated with obesity in the model tested.

Although a number of variables were not included but were explored in the earlier phases of this study (e.g., Gini index, per capita income, public transportation use, crime, female employment, unemployment, occupation type, gender, median age, etc.). These variables were discarded for various reasons, such as multicollinearity or low correlation with obesity. However, future studies should definitely reconsider some of those variables to use them especially for counties where the local R^2 values are low. In those areas, the low local R^2 values may be indicating that the current model is missing some important factors that are associated with obesity in those specific areas. In addition, future studies should also consider exploring the areas (counties) where the sign of the correlation is not as expected. Such areas are often very interesting and warrant further exploration in order to reveal what might be causing such unexpected patterns.

In conclusion, the main objective of this study was to identify predictors of obesity prevalence at the county level and to reveal how the effects of these predictors vary spatially. The main hypothesis aimed to test whether the predictors of obesity prevalence and effects and variations of such factors across space can be identified and revealed by the GWR technique at the county level. Significant local parameter estimates were found for the independent variables, confirming spatial heterogeneity in the effects of these variables on obesity and providing insights into the spatial scale at which processes may be operating. In addition, scrutinizing local parameter estimates based on the effect of these variables in different regions revealed further insights. Consequently, the results from this study generated empirically-based and useful insights that will help national but, most importantly, local obesity programs and interventions to better target areas/populations and specific factors associated with obesity prevalence across the US. The robust results obtained through the utilization of a GWR model in this study also proved that the GWR as a spatial analysis approach is valuable and should definitely be used in future obesity, as well as public health, studies.

REFERENCES

106th Congress. 2000. *Public Law 106-525: Minority Health and Health Disparities Research and Education Act of 2000*.

Abercrombie, L. C., Sallis, J. F., Conway, T. L., Frank, L. D., Saelens, B. E., and Chapman, J. E. 2008. Income and racial disparities in access to public parks and private recreation facilities. *American Journal of Preventive Medicine*, 34: 9-15.

Ahern, M., Brown, C., and Dukas, S. 2011. A national study of the association between food environments and county-level health outcomes. *The Journal of Rural Health*, 27: 367-379.

Ahima, R. S. and Lazar, M. A. 2013. The health risk of obesity-better metrics imperative. *Science*, 341: 856-858.

Ailes, E., Scallan, E., Berkelman, R. L., Kleinbaum, D. G., Tauxe, R. V., and Moe, C. L. 2012. Do differences in risk factors, medical care seeking, or medical practices explain the geographic variation in campylobacteriosis in foodborne diseases active surveillance network (foodnet) sites? *Clinical Infectious Diseases*, 54(Suppl. 5): 464-471.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6): 716-723.

American Medical Association. Website http://www.amaassn.org/ama/pub/news/news/2013/2013-06-18-new-ama-policies-annual-meeting.page. Last accesed September 26, 2014.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic.

Anselin, L. 1995. Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2), 93-115.

Anselin, L. 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2), 153-166.

Anselin, L. 2005. *Exploring Spatial Data with GeoDa™. A workbook*. IL, USA: Spatial Analysis Laboratory, Department of Geography, University of Illinois and the Centre for Spatially Integrated Social Science (CSISS).

Anselin, L. and Bera, A. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. Editors: Ullah, A. and Giles, D. E. *Handbook of Applied Economic Statistics*, 237-289. New York: Marcel Dekker.

Austin, G. L., Ogden, L. G., and Hill, J. O. 2011. Trends in carbohydrate, fat, and protein intakes and association with energy intake in normal-weight, overweight, and obese individuals: 1971-2006. *American Journal of Clinical Nutrition*, 93: 836-843.

Bailey, T. C. and Gatrell, A. C. 1995. *Interactive Spatial Data Analysis*. Harlow, England: Longman Publishers.

Barbara, M., Livingstone, E., and Pourshahidi, L. K. 2014. Portion size and obesity. *Advances in Nutrition*, 5: 829-834.

Behavioral Risk Factor Surveillance System. Website http://www.cdc.gov/brfss. Last accessed July 30, 2014.

Bennett, K., Probst, J., and Pumkam, C. 2011. Obesity among working age adults: The role of county-level persistent poverty in rural disparities. *Health and Place*, 17(5): 1174-1181.

Berrigan, D. and Troiano, R. 2002. The association between urban form and physical activity in U.S. adults. *American Journal of Preventive Medicine*, 23(2S): 74-79.

Besag, J. and Newell, J. 1991. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society*, A154: 143-155.

Bitter, C., Mulligan, G. F., and Dall'erba, S. 2007. Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical System*, 9: 7-27.

Block, J. P., Scribner, R. A., and DeSalvo, K. B. 2004. Fast food, race/ethnicity, and income: A geographic analysis. *American Journal of Preventive Medicine*, 27: 211-217.

Bodor, J., Rice, J., Farley, T., Swalm, C., and Rose, D. 2010. The association between obesity and urban food environments. *Journal of Urban Health,* 87(5): 771-781.

Briefel, R. R. and Johnson, C. L. 2004. Secular trends in dietary intake in the United States. *Annual* Review of *Nutrition*, 24: 401-431.

Brown, D. G., Elmes, G., Kemp, K. K., Macey, S., and Mark, D. 2005. Geographic information systems. Editors: Gaile, G. L. and Willmott, C. J. *Geography in America at the Dawn of the 21st Century*, 353-375. New York, USA: Oxford University Press Inc.

Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A., and Sallis, J. F. 2009. Measuring the built environment for physical activity: State of the science. *American Journal of Preventive Medicine*, 36(4 Suppl 1): 99-123.

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28(4): 281-298.

Casagrande, S.S., Franco, M., Gittelsohn, J., Zonderman, A. B., Evans, M.K., Kuczmarski, M.F., and Gary-Webb, T. L. 2011. Healthy food availability and the association with BMI in Baltimore, Maryland. *Public Health Nutrition,* 14: 1001-1007.

Census Transportation Planning Products. Website http://www.fhwa.dot.gov/planning/census_issues/ctpp. Last accessed July 30, 2014.

Centers for Disease Control and Prevention (CDC). 2010. Vital signs: State-specific prevalence of obesity among adults-United States, 2009. *Morbidity and Mortality Weekly Report,* 59: 1-5.

Centers for Disease Control and Prevention (CDC). Website http://www.cdc.gov. Last accessed July 30, 2014.

Centers for Disease Control and Prevention (CDC). Website http://www.cdc.gov. Last accessed October 31, 2015.

Cervero, R. 1989. Jobs-housing balancing and regional mobility. *Journal of American Planning Association*, 1: 136-150.

Cervero, R. 1996. Job-housing balance revisited: Trends and impacts in the San Francisco Bay Area. *Journal of American Planning Association*, 62: 492-511.

Chen, V. Y. J., Wu, P. C., Yang, T. C., and Su, H. J. 2010. Examining non-stationary effects of social determinants on cardiovascular mortality after cold surges in Taiwan. *Science of the Total Environment* 408(9): 2042-2049.

Cheng, J. K. 2012. Confronting the social determinants of health-obesity, neglect, and inequity. *New England Journal of Medicine*, 367(21): 1976-1977.

Chi, S., Grigsby-Toussaint, D. S., Bradford, N., and Choi, J. 2013. Can geographically weighted regression improve our contextual understanding of obesity in the US? Findings from the USDA Food Atlas. *Applied Geography*, 44: 134-142.

Cho, Y. H. 1996. *Exploring Spatial Analysis in GIS*. Santa Fe, NM, USA: On Word Press.

Christian, T. and Rashad, I. 2009. Trends in U.S. food prices, 1950-2007. *Economics and Human Biology*, 7: 113-120.

Cohen, A. K., Rehkopf, D. H., Deardorff, J., Abrams, B. 2013. Education and obesity at age 40 among American adults. *Social Science and Medicine*, 78: 34-41.

Cromley, R. G. and Cromley, E. K. 2009. Chloropleth map legend design for visualizing community health disparities. *International Journal of Health Geographics*, 8(52): 1-11.

Cummins, S. and Macintyre, S. 2006. Food environments and obesity: Neighborhood or nation? *International Journal of Epidemiol*ogy, 35(1): 100-104.

Danaei, G., Ding, E. L., Mozaffarian, D., Taylor, B., Rehm, J., Murray, C. J. L., Ezzati, M. 2009. The preventable causes of death in the United States: Comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Medicine,* 6: 4.

Dannenberg, A. L., Jackson, R. J., Frumkin, H., Schieber, R. A., Pratt, M., Kochtitzky, C., and Tilson H. H. 2003. The impact of community design and land-use choices on public health: A scientific research agenda. *American Journal of Public Health,* 93: 1500-1508.

Das, U. N. 2010. Obesity: Genes, brain, gut, and environment. *Nutrition*, 26(5): 459-473.

De Vogli, R., Kouvonen, A., and Gimeno, D. 2011. 'Globesization': Ecological evidence on the relationship between fast food outlets and obesity among 26 advanced economies. *Critical Public Health*, 21: 395-402.

Diez Roux, A. V., Evenson, K. R., and McGinn, A. P. 2007. Availability of recreational resources and physical activity in adults. *American Journal of Public Health*, 97(3): 493-499.

Ding, D., Sallis, J. F., Kerr, J., Lee, S., and Rosenberg, D. 2011. Neighborhood environment and physical activity among youth: A review. *American Journal of Preventive Medicine*, 41(4): 442-455.

Drewnowski, A., Rehm, C. D., and Solet, D. 2007. Disparities in obesity rates: Analysis by ZIP code area. *Social Science and Medicine*, 65(12): 2458-2463.

Durand, C. P., Andalib, M., Dunton, G. F., Wolch J., and Pentz, M.A. 2011. A systematic review of built environment factors related to physical activity and obesity risk: Implications for smart growth urban planning. *Obesity Reviews,* 12(5): 173-182.

Ebdon, D. 1985. *Statistics in Geography*. Oxford, England: Blackwell Publishers.

Eck, J., Chainey, S., Cameron, J., Leitner, M., and Wilson. R. E. 2005. *Mapping Crime: Understanding Hot Spots*. Washington, DC: U.S. Department of Justice Office of Justice Programs.

Economos, C. D., Hatfield, D. P., King, A. C., Ayala, G. X., and Pentz, M. A. 2015. Food and physical activity environments an energy balance approach for research and practice. *American Journal of Preventive Medicine*, 48(5): 620-629.

Ewing, R., Schmid, T., Killingsworth, R., and Raudenbush, S. 2003. Relationship between urban sprawl and physical activity, obesity, and morbidity. *American Journal of Health Promotion*, 18(1): 47-57.

Ewing, R., Schmid, T., Killingsworth, R., Zlot, A., and Raudenbush, S. 2008. Relationship between urban sprawl and physical activity, obesity, and morbidity. *Urban Ecology*, 567-582.

Feng, J., Glass, T. A., Curriero, F. C., Stewart, W. F., and Schwartz, B. S. 2010. The built environment and obesity: A systematic review of the epidemiologic evidence. *Health and Place*, 16: 175-190.

Ferdinand, A. O., Sen, B., Rahurkar, S., Engler, S., and Menachemi, N. 2012. The relationship between built environments and physical activity: A systematic review. *American Journal of Public Health*, 102(10): 7-13.

Finkelstein, E. A. and Zuckerman, L. 2008. *The Fattening of America, How the Economy Makes us Fat, If It Matters, and What to Do about It*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Finkelstein, E. A., Strombotne, K. L., Zhen, C., and Epstein, L. H. 2014. Food prices and obesity: A review. *Advances in Nutrition*, 5: 818-821.

Finkelstein, EA, Trogdon, JG, Cohen, JW, and Dietz, W. 2009. Annual medical spending attributable to obesity: Payer and service specific estimates. *Health Affairs,* 28(5): w822-w831.

Fischer, M. M. and Getis, A. 1997. *Recent Developments in Spatial Analysis*. Berlin, Germany: Springer Verlag.

Fischer, M. M. and Nijkamp, P. 1993. *Geographic Information Systems, Spatial Modeling and Policy Evaluation*. Berlin, Germany: Springer Verlag.

Flegal, K. M., Carroll, M. D., Kit, B. K., and Ogden, C. L. 2012. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *The Journal of the American Medical Association*, 307(5): 491-497.

Flegal, K. M., Carroll, M. D., Ogden, C. L., and Curtin, L. R. 2010. Prevalence and trends in obesity among U.S. adults, 1999-2008. *The Journal of the American Medical Association,* 303: 235-241.

Flegal, K. M., Graubard, B. I., Williamson, D. F., and Gail, M. H. 2005. Obesity. *The Journal of the American Medical Association*, 293(15): 1861-1867.

Flegal, K. M., Kit, B. K., Orpana, H., and Graubard, B. I. 2013. Association of all-cause mortality with overweight and obesity using standard body mass index categories: A systematic review and meta-analysis. *The Journal of the American Medical Association*, 309(1): 71-82.

Foody, G.M. 2003. Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI-rainfall relationship. *Remote Sensing of the Environment* 88(3): 283-293.

Foresman, T. 1998. *The History of Geographic Information Systems: Perspectives from the Pioneers*. Upper Saddle River, NJ, USA: Prentice Hall.

Fotheringham, A. S. and Rogerson, P. 1994. *Spatial Analysis and GIS*. London, England: Taylor & Francis.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, UK: John Wiley & Sons.

Franco, M., Diez Roux, A. V., Glass, T. A., Caballero, B., and Brancati, F. L. 2008. Neighborhood characteristics and availability of healthy foods in Baltimore. *American Journal of Preventive Medicine,* 35: 561-567.

Franco, M., Diez-Roux, A. V., and Nettleton, J. A. 2009. Availability of healthy foods and dietary patterns: The multi-ethnic study of atherosclerosis. *American Journal of Clinical Nutrition* 89: 897-904.

Frank, L. D., Saelens, B. E., Powell, K. E., and Chapman, J. E. 2007. Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Social Science and Medicine*, 65: 1898-1914.

Frank, L. D., Sallis, J. F., Conway, T. L., and Chapman, J. E. 2006. Many pathways from land use to health. *Journal of the American Planning Association*, 72(1): 75-89.

Frank, L. D., Schmid, T. L., Sallis, J. F., Chapman, J., and Saelens, B. E. 2005. Linking objectively measured physical activity with objectively measured urban form: Findings from SMARTRAQ. *American Journal of Preventative Medicine*, 28(Suppl 2): 117-25.

Fraser, L. K., Clarke, G. P., Cade, J .E., and Edwards, K. L. 2012. Fast food and obesity: A spatial analysis in a large United Kingdom population of children aged 13-15. *American Journal of Preventive Medicine*, 42(5): 77-85.

Frumkin, H., Frank, L., and Jackson, R. 2004. *Urban Sprawl and Public Health: Designing, Planning, and Building for Healthy Communities*. Washington, DC, USA: Island Press.

Galvez, M. P., Pearl, M., and Yen, I. H. 2010. Childhood obesity and the built environment. *Current Opinion in Pediatrics,* 22: 202-207.

Gatrell, A. C. 2002. *Geographies of Health: An Introduction*. Oxford: Blackwell Publishers.

Goodchild, M. F. 1992. Geographical Information Science. *International Journal of Geographical Information Systems*, 6(1): 31-45.

Goodchild, M. F. and Haining, R. P. 2004. GIS and spatial data analysis: Converging perspectives. *Papers in Regional Science,* 83(1): 363-385.

Goodchild, M. F. and Janelle, D. G. 2010. Toward critical spatial thinking in the social sciences and humanities. *GeoJournal*, 75(1): 3-13.

Goodman, C. C. and Fuller, K. S. 2015. *Pathology: Implications for the Physical Therapist*. St. Louis, MO: Elsevier-Saunders.

Goovaerts, P. 2005. Analysis and detection of health disparities using geostatistics and a space-time information system: The case of prostate cancer mortality in the United States, 1970-1994. *Proceedings of GIS Planet* May 30-June 2, 2005, Estoril, Portugal.

Gordon-Larsen, P. 2014. Food availability/convenience and obesity. *Advances in Nutrition*, 5: 809-817.

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

Hammond, R. A. and Levine, R. 2010. The economic impact of obesity in the United States. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 3: 285-295.

Handy, S. and Clifton, K. 2007. Planning and the built environment: Implications for obesity prevention. Editors: Kumanyika, S. and Brownson, R. C. *Handbook of Obesity Prevention,* 171-192. Springer Science and Business Media, LLC.

Handy, S. L., Boarnet, M. G., Ewing, R., and Killingsworth, R. E. 2002. How the built environment affects physical activity: Views from urban planning. *American Journal of Preventive Medicine*, 23(2): 64-73.

Harrington, D. W. and Elliott, S. J. 2009. Weighing the importance of neighborhood: A multilevel exploration of the determinants of overweight and obesity. *Social Science and Medicine*, 68: 593-600.

Hawkes, C., Smith, T. G., Jewell, J., Wardle, J., Hammond, R. A., Friel, S., Thow, A. M., and Kain, J. 2015. Smart food policies for obesity prevention. *Lancet*, 385: 2410-2421.

Healthy People 2020. Website http://www.healthypeople.gov/2020/about/DOHAbout.aspx. Last accessed July 30, 2014.

Heckler, M. M. 1985. *Report of the Task Force on Blacks and Minority Health*. Washington, DC: U.S. Department of Health and Human Services.

Heron, M., Hoyert, D. L., Murphy, S. L., Xu, J., Kochanek, K. D., and Tehada-Vera, B. 2009. Deaths: Final data for 2006. *National Vital Statistics Report*, 57: 1-134.

Hill, J. L., You, W., and Zoellner, J. M. 2014. Disparities in obesity among rural and urban residents in a health disparate region. *BMC Public Health*, 14: 1051.

Hill, J. O. and Peters, J. C. 1998. Environmental contributions to obesity. *Science*, 282: 1371-1373.

Hill, J. O. and Trowbridge, F. L. 1998. Childhood obesity: Future directions and research priorities. *Pediatrics*, Supplement 571(101): 570-574.

Hill, J. O., Wyatt, H. R., Reed, G. W., and Peters, J. C. 2003. Obesity and the environment: Where do we go from here? *Science*, 7: 853-855.

Hipp, J. A. and Chalise, N. 2015. Spatial analysis and correlates of county-level diabetes prevalence, 2009-2010. *Preventing Chronic Disease*, 12: 140404.

Holt, J. B. 2007. The topography of poverty in the United States: A spatial analysis using county-level data from the community health status indicators project. *Preventing Chronic Disease*, 4(4): 1-9.

Holt, J. B. and Sui, D. Z. 2007. *A Spatio-temporal Analysis of Obesity and Poverty in the United States*. U.S. Centers for Disease Control and Prevention.

Hu, F. B. 2007. Obesity and mortality: Watch your waist, not just your weight. *Archives of Internal Medicine*, 167: 875-876.

Huang, T. T. K. and Glass, T. A. 2008. Transforming research strategies for understanding and preventing obesity. *The Journal of the American Medical Association*, 300: 1811-1813.

Iton, A. B., Witt, S., Desautels, A., Schaff, K., Luluquisen, M., Maker, L., Horsley, K., and Beyers, M. 2010. Tackling the root causes of health disparities through community capacity building. Editors: Hofrichter, R. and Bhatia, R. *Tackling Health Inequities through Public Health Practice: Theory to Action,* 380. New York, NY: Oxford University Press.

Jackson, J. E, Doescher, M. D., Jerant, A. F., and Hart, L. G. 2015. A national study of obesity prevalence and trends by type of rural county. *Journal of Rural Health*, 21(2): 140-148.

Janssen, I., Katzmarzyk, P. T., and Ross, R. 2002. Body mass index, waist circumference, and health risk: Evidence in support of current National Institutes of Health guidelines. *Archives of Internal Medicine*, 162(18): 2074-2079.

Jilcott Pitts, S. B., Edwards, M., Moore, J. B., Shores, K. A., Dubose, K. D., and McGranahan, D. 2013. Obesity is inversely associated with natural amenities and recreation facilities Per capita. *Journal of Physical Activity and Health*, 10(7): 1032-1038.

Johnson, F. and Wardle, J. 2014. Variety, palatability, and obesity. *Advances in Nutrition*, 5: 851-859.

Jordan, A. C., Saelens, B. E., Kerr, J., Schipperijn, J., Conway, T. L., Frank, L. D., Chapman, J. E., Glanz, K., Cain, K. L., and Sallis, J. F. 2015. Association between neighborhood walkability and GPS-measured walking, bicycling and vehicle time in adolescents. *Health and Place*, 32: 1-7.

Kaczynski, A. T., Potwarka, L. R., and Saelens, B. E. 2008. Association of park size, distance, and features with physical activity in neighborhood parks. *American Journal of Public Health*, 98(8): 1451-1456.

Karl, J. P. and Roberts, S. B. 2014. Energy density, energy intake, and body weight regulation in adults. *Advances in Nutrition*, 5: 835-850.

Kelly, T., Yang, W., Chen, C. S., Reynolds, K, and He, J. 2008. Global burden of obesity in 2005 and projections to 2030. *International Journal of Obesity*, 32(9): 1431-1437.

Kent, J. L. and Thompson, S. 2014. The three domains of urban planning for health and well-being. *Journal of Planning Literature*, 29(3): 239-256.

Kerr, J., Frank, L., Sallis, J. F., and Chapman, J. 2007. Urban form correlates of pedestrian travel in youth: Differences by gender, race-ethnicity and household attributes. *Transportation Research Part D: Transport and Environment*, 12: 177-182.

Kershaw, K. N., Albrecht, S. S., and Carnethon, M. R. 2013. Racial and ethnic residential segregation, the neighborhood socioeconomic environment, and obesity among Blacks and Mexican Americans. *American Journal of Epidemiology*, 177(4): 299-309.

Kim, D., Subramanian, S. V., Gortmaker, S. L., and Kawachi, I. 2006. US state and county-level social capital in relation to obesity and physical inactivity: A multilevel, multivariable analysis. *Social Science and Medicine*, 63(4): 1045-1059.

Kuczmarski, R. J., Flegal, K. M., Campbell, S. M., and Johnson, C. L. 1994. Increasing prevalence of overweight among US adults. The National Health and Nutrition Examination Surveys, 1960-1991. *The Journal of the American Medical Association*, 272: 205-211.

Kulldorff, M and Nagarwalla, N. 1995. Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14(8): 799-811.

Kumanyika, S. 2005. Obesity, health disparities, and prevention paradigms: Hard questions and hard choices. *Preventing Chronic Disease*, 2(4): A02.

Lake, A. and Townshend, T. 2006 Obesogenic environments: Exploring the built and food environments. *The Journal of the Royal Society for the Promotion of Health,* 126(6): 262-267.

Lamichhane, A. P., Puett, R., Porter, D. E., Bottai, M., Mayer-Davis, E. J., and Liese, A. D. 2012. Associations of built food environment with body mass index and waist circumference among youth with diabetes. *International Journal of Behavioral Nutrition and Physical Activity*, 9: 81.

Larson, N., Story, M., and Nelson, M. 2009. Neighborhood environments: Disparities in access to health foods in the U.S. *American Journal of Preventive Medicine*, 36: 74-81.

Lathey, V., Guhathakurta, S., and Aggarwal, R. M. 2009. The impact of subregional variations in urban sprawl on the prevalence of obesity and related morbidity. *Journal of Planning Education and Research,* 29(2): 127-141.

Lee, J., Alnasrallah, M., Wong, D., Beaird, H., and Logue, E. 2014. Impacts of scale on geographic analysis of health data: An example of obesity prevalence. *ISPRS International Journal of Geo-Information*, 3: 1198-1210.

Legates, D. R., Gopal, S., and Rogerson, P. 2005. Mathematical models and quantitative methods. Editors: Gaile, G. L. and Willmott, C. J. *Geography in America at the Dawn of the 21st Century*, 442-457. New York, USA: Oxford University Press Inc.

Leslie, E., Saelens, B., Frank, L., Owen, N., Bauman, A., Coffee, N., and Hugo, G. 2005. Residents' perceptions of walkability attributes in objectively different neighborhoods: A pilot study. *Health and Place*, 11: 227-236.

Levine, J. 2011. Poverty and obesity in the U.S. *Diabetes*, 60(11): 2667-2668.

Levine, N. 2015. Chapter 5: Spatial autocorrelation statistics. *CrimeStat IV Manual*. Houston, TX.

Li, F., Harmer, P., Cardinal, B. J., Bosworth, M., and Johnson-Shelton, D. 2009. Obesity and the built environment: Does the density of neighborhood fast-food outlets matter? *American Journal of Health Promotion*, 23(3): 137-143.

Litman, T. 2008. Evaluating Accessibility for Transportation Planning, Victoria Transport Policy Institute. Available at www.vtpi.org/access.pdf.

Lloyd, C. D. 2007. *Local Models for Spatial Analysis*. Boca Raton, FL, USA: CRC Press by Taylor & Francis Group, LLC.

Longley, P., Goodchild, M., Maguire, D., and Rhind, D. 1997. *Geographical Information Systems: Techniques, Management and Applications*. Cambridge, England: Geoinformation International.

Longley, P.A. and Tobon, C. 2004. Spatial dependence and heterogeneity in patterns of hardship: An intra-urban analysis. *Annals Association of American Geographers* 94(3): 503-519.

Lopez, R. and Hynes, P. 2006. Obesity, physical activity, and the urban environment: Public health research needs. *Environmental Health*, 18: 5-25.

Lovasi, G. S. 2012. Built environment and health. Editor: Rippe, J. M. *Encyclopedia of Lifestyle Medicine and Health*. Thousand Oaks, CA: Sage Publications.

Lovasi, G. S., Grady, S., and Rundle, A. 2012. Steps forward: Review and recommendations for research on walkability, physical activity and cardiovascular health. *Public Health Review*, 33(4): 484-506.

Lovasi, G. S., Hutson, M. A., Guerra, M., and Neckerman, K .M. 2009. Built environments and obesity in disadvantaged populations. *Epidemiologic Reviews*, 31: 7-20.

Maddock, J. 2004. The relationship between obesity and the prevalence of fast food restaurants: State-level analysis. *American Journal of Health Promotion*, 19: 137-143.

Maillot, M., Darmon, N., and Drewnowski, A. 2010. Are the lowest-cost healthful food plans culturally and socially acceptable? *Public Health Nutrition*, 13: 1178-1185.

Maroko, A. R., Maantay, J. A., Sohler, N. L., Grady, K. L., and Arno P. S. 2009. The complexities of measuring access to parks and physical activity sites in New York City: A quantitative and qualitative approach. *International Journal of Health Geographics*, 8(1): 34.

Martinez, A. N., Mobley, L.R., Lorvick, J., Novak, S. P., Lopez, A. M., and Kral, A. H. 2014. Spatial analysis of HIV positive injection drug users in San Francisco, 1987 to 2005. *International Journal of Environmental Research and Public Health*, 11: 3937-3955.

Matthews, S. A. and Yang, T.C. 2012. Mapping the results of local statistics: Using geographically weighted regression *Demographic Research*, 26(6): 151-166.

McCormack, G. R. and Shiell, A. 2011. In search of causality: A systematic review of the relationship between the built environment and physical activity among adults. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1): 125.

McLaren, L. 2007. Socioeconomic status and obesity. *Epidemiologic Reviews,* 29(1): 29-48.

Mennis, J. L. 2006. Mapping the results of geographically weighted regression. *The Cartographic Journal* 43(2): 171-179.

Mennis, J. L. and Jordan, L. M. 2005. The distribution of environmental equity: Exploring spatial nonstationarity in multivariate models of air toxic releases. *Annals of the Association of American Geographers* 95(2): 249-268.

Merrill, R. M. and Richardson, J. S. 2009. Validity of self-reported height, weight, and body mass index: Findings from the National Health and Nutrition Examination Survey, 2001-2006. *Preventing Chronic Disease*, 6(4): A121.

Michimi, A. and Wimberly, M. C. 2010. Spatial patterns of obesity and associated risk factors in the conterminous U.S. *American Journal of Preventive Medicine,* 39: 2.

Michimi, A. and Wimberly, M. C. 2012. Natural environments, obesity, and physical activity in nonmetropolitan areas of the United States. *The Journal of Rural Health*, 28(4): 398-407.

Mitchell, A. 2005. *The ESRI Guide to GIS Analysis: Spatial Measurements and Statistics*, Volume 2. ESRI Press.

Mokdad, A. H., Bowman, B. A., Ford, E. S., Vinicor, F., Marks J. S., and Koplan, J. P. 2001. The continuing epidemics of obesity and diabetes in the United States. *The Journal of the American Medical Association,* 286(10): 1519-1522.

Mokdad, A. H., Marks, J. S., Stroup, D. F., and Gerberding, J. L. 2004. Actual causes of death in the United States, 2000. *The Journal of the American Medical Association*, 291: 1238-1245.

Mokdad, A. H., Serdula, M. K., Dietz, W. H., Bowman, B. A., Marks, J. S., and Koplan, J. P. 1999. The spread of the obesity epidemic in the United States, 1991-1998. *The Journal of the American Medical Association*, 282(16): 1519-1522.

Mokdad, A. H., Serdula, M. K., Dietz, W. H., Bowman, B. A., Marks, J. S., and Koplan, J. P. 2000. The continuing epidemic of obesity in the United States. *The Journal of the American Medical Association*, 284: 1650-1651.

Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37: 17-23.

Morland, K., Wing, S., and Diez Roux, A. 2012. The contextual effect of the local food environment on resident's diets: The atherosclerosis risk in communities study. *American Journal of Public Health*, 92(11): 1761-1768.

Mujahid, M. S., Diez Roux, A. V., Shen, M., Gowda, D., Sa´nchez, B., Shea, S., Jacobs, D. R., and Jackson, S. A. 2008. Relation between neighborhood environments and obesity in the multi-ethnic study of atherosclerosis. *American Journal of Epidemiology*, 167(11): 1349-1357.

Munch, Z., Van Lill, S. W. P., Booysen, C. N., Zietsman, H. L., Enarson, D. A., and Beyers, N. 2003. Tuberculosis transmission patterns in a high-incidence area: A spatial analysis. *International Journal of Tuberculosis and Lung Disease, 7*(3): 271-277.

Nakaya, T., Fotheringham, A.S., Brunsdon, C., and Charlton, M. 2005. Geographically weighted poisson regression for disease association mapping. *Statistics in Medicine* 24(17): 2695-2717.

National Center for Health Statistics (NCHS). 2013. *Health, United States, 2012: With Special Feature on Emergency Care*. Hyattsville, MD.

Nelson, M. C., Gordon-Larsen, P., Song, Y., and Popkin, B. M. 2006. Built and social environments. *American Journal of Preventive Medicine*, 31: 109-117.

Office of the Surgeon General. Website http://www.surgeongeneral.gov. Last accessed July 30, 2014.

Ogden, C. and Carroll, M. 2010a. *Prevalence of Overweight, Obesity, and Extreme Obesity among Adults: United States, Trends 1960-1962 through 2007-2008*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Health Statistics.

Ogden, C. and Carroll, M. 2010b. *Prevalence of Obesity among Children and Adolescents: United States, Trends 1963-1965 through 2007-2008*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Health Statistics.

Ogden, C. L., Carroll, M. D., Kit, B. K., and Flegal, K. M. 2014. Prevalence of childhood and adult obesity in the United States, 2011-2012. *The Journal of the American Medical Association*, 311(8): 806-814.

Openshaw, S., Charlton, M., Wymer, C., and Craft, A. 1987. A mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems*, 1: 335-358.

Papas, M. A., Alberg, A. J., Ewing, R., Helzlsouer, K. J., Gary, T. L., and Klassen, A. C. 2007. The built environment and obesity. *Epidemiologic Reviews*, 29: 129-143.

Partridge, M.D. and Rickman, D.S. 2005. *Persistent Pockets of Extreme American Poverty: People or Place Based?* Columbia, MO: Rural Poverty Research Center (RPRC).

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., Buchner, D., Ettinger, W., Heath, G. W., King, A. C., Kriska, A., Leon, A. S., Marcus, B. H., Morris, J., Paffenbarger, R. S., Patrick, K., Pollock, M. L., Rippe, J. M., Sallis, J., and Wilmore, J. H. 1995. Physical activity and public health: A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *The Journal of the American Medical Association*, 273: 402-407.

Peng, Z.R. 1997. Jobs-housing balance and urban commuting. *Urban Studies*, 34(8): 1215-1235.

Pereira, M. A. 2014. Sugar-sweetened and artificially-sweetened beverages in relation to obesity risk. *Advances in Nutrition*, 5: 797-808.

Plantinga, A. J. and Bernell, S. 2007. The association between urban sprawl and obesity: Is it a two-way street? *Journal of Regional Science*, 47(5): 857-879.

Popkin, B. M. 2011. Agricultural policies, food and public health. *EMBO Reports*, 12: 11-18.

Powell, L. M., Wada, R., Krauss, R. C., and Wang, Y. 2012. Ethnic disparities in adolescent body mass index in the United States: The role of parental socioeconomic status and economic contextual factors. *Social Science and Medicine*, 75(3): 469-476.

Ramsey, P. W. and Glenn, L. L. 2002. Obesity and health status in rural, urban and suburban southern women. *The Southern Medical Journal*, 95(7): 666-671.

Richardson, D. B., Volkow, N. D., Kwan, M. P., Kaplan, R. M., Goodchild, M. F., and Croyle, R. T. 2013. Spatial turn in health research. *Science*, 339(6126): 1390-1392.

Rossen, L. M. and Pollack, K. M. 2012. Making the connection between zoning and health disparities. *Environmental Justice*, 5: 3.

Ruel, E., Reither, E. N., Robert, S. A., and Lantz, P. M. 2010. Neighborhood effects on BMI trends: Examining BMI trajectories for black and white women. *Health and Place,* 16: 191-198.

Rundle, A., Roux, A. V., Free, L. M., Miller, D., Neckerman, K. M., and Weiss, C. C. 2007. The urban built environment and obesity in New York City: A multilevel analysis. *American Journal of Health Promotion,* 21: 326-334.

Russel, C. M., Williamson, D. F., and Byers, T. 1995. Can the year 2000 objective for reducing overweight in the United States be reached? A simulation study of the required changes in body weight. *International Journal of Obesity*, 19: 149-153.

Saelens, B., Sallis, J. F., Black, J. B., and Chen, D. 2003. Neighborhood-based differences in physical activity: An environment scale evaluation. *American Journal of Public Health*, 93(9): 1552-1558.

Sallis, J. F., Floyd, M. F., Rodríguez, D. A., and Saelens, B. E. 2012. Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation,* 125: 729-737.

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

Schlundt, D. G.; Hargreaves, M. K., and McClellan, L. 2006. Geographic clustering of obesity, diabetes, and hypertension in Nashville, Tennessee. *Journal of Ambulatory Care Management*, 29(2): 125-152.

Schulz, A. and Northridge, M. E. 2004. Social determinants of health: Implications for environmental health promotion. *Health Education and Behavior*, 31: 455-471.

Schuurman, N., Peters, P. A., and Oliver, L. N. 2012. Are obesity and physical activity clustered? A spatial analysis linked to residential density. *Obesity*, 17(12): 2202-2209.

Serdula, M. K., Mokdad, A. H., Williamson, D. F., Galuska, D. A., Mendlein, J.M., and Heath, G. W. 1999. Prevalence of attempting weight loss and strategies for controlling weight. *The Journal of the American Medical Association*, 13(14): 1353-1358.

SPSS 23 software, IBM Corporation. 2015. Website http://www-01.ibm.com/support/docview.wss?uid=swg24038592. Last accessed November, 10, 2015.

Srinivasan, S., O'Fallon, L. R., and Dearry, A. 2003. Creating healthy communities, healthy homes, healthy people: Initiating a research agenda on the built environment and public health. *American Journal of Public Health*, 93: 1446-1450.

Story, M., Kaphingst, K. M., Robinson-O'Brien, R., and Glanz, K. 2008. Creating healthy food and eating environments: Policy and environmental approaches. *Annual Review of Public Health*, 29: 253-72.

Sturm, R. 2001. Does obesity contribute as much to morbidity as poverty or smoking? *Public Health*, 115(3): 229-235.

Sturm, R. 2004. The economics of physical activity: Societal trends and rationales for intervention. *American Journal of Preventive Medicine*, 27(3S): 126-135.

Sturm, R. and An, R. 2014. Obesity and economic environments. *CA: A Cancer Journal for Clinicians*, 64(5): 337-350.

Sturm, R. and Cohen, D. A. 2009. Zoning for health? The year-old ban on new fast-food restaurants in South LA. *Health Affairs*, 28: 1088-1097.

Sultana, S. 2002. Job/housing imbalance and commuting time in the Atlanta Metropolitan Area: Exploration of causes of longer commuting time. *Urban Geography,* 23: 728-49.

Sutin, A. R., Ferrucci, L., Zonderman, A. B., and Terracciano, A. 2011. Personality and obesity across the adult life span. *Journal of Personality and Social Psychology*, 101(3): 579-592.

Swinburn, B., Egger, G., and Razza, F. 1999. Dissecting obesogenic environments: The development and application of a framework for identifying and prioritizing environmental interventions for obesity. *Preventive Medicine*, 29(6): 563-570.

Taveras, E. M., Gillman, M. W., Kleinman, K. P., Rich-Edwards, J. W., and Rifas-Shiman, S. L. 2013. Reducing racial/ethnic disparities in childhood obesity: The role of early life risk factors. *JAMA Pediatrics*, 167(8): 10.

The GBD 2013 Obesity Collaboration, Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., and Gakidou, E. 2014. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the global burden of disease study 2013. *The Lancet*, 384(9945): 766-781.

Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2): 234-240.

US Census Bureau, 2010 American Community Survey (ACS) 5-Year Estimates. Website https://www.census.gov/programs-surveys/acs/guidance/estimates.html. Last accessed November 10, 2015.

US Census Bureau, 2010 Census Summary File 1 (SF 1). Website http://www.census.gov/prod/cen2010/doc/sf1.pdf. Last accessed November 10, 2015.

US Census Bureau, County Business Patterns (CBP). Website http://www.census.gov/econ/cbp/index.html. Last accessed November 10, 2015.

US Census Bureau Population Estimates Program. Website http://www.census.gov/popest/about/index.html. Last accessed July 30, 2014.

US Census Bureau Small Area Income and Poverty Estimates. Website http://www.census.gov/did/www/saipe/about/index.html. Last accessed July 30, 2014.

US Census Bureau Topologically Integrated Geographic Encoding and Referencing (TIGER). Website http://www.census.gov/geo/maps-data/data/tiger.html. Last accessed July 30, 2014.

US Department of Agriculture (USDA), Economic Research Service (ERS). Website http://www.ers.usda.gov/data-products/natural-amenities-scale.aspx. Last accessed November 10, 2015.

US Department of Agriculture (USDA), Food Environment Atlas. Website http://ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads.aspx. Last accessed November 10, 2015.

US Department of Health and Human Services: Health Resources and Services Administration - Rural Health (HRSA) Website http://www.hrsa.gov/ruralhealth/policy/definition_of_rural.html Last accessed August 29, 2015.

Vinnakota, S. K. 2006. Socioeconomic characteristics of cancer mortality in the United States of America: A spatial data mining approach. *Doctoral Dissertation*. Louisiana State University.

von Hippel, P. and Benson, R. 2014. Obesity and the natural environment across US counties. *American Journal of Public Health*, 4(104): 7.

Voss, P. R.; White, K., Curtis J., and Hammer, R. B. 2006. Explorations in spatial demography. Editors: Kandel, W. A. and Brown, D. L. *Population Change and Rural Society,* 407-429. Dordrecht, Netherlands: Springer.

Walker, R. E., Keane, C. R., and Burke, J. G. 2010. Disparities and access to healthy food in the United States: A review of food deserts literature. *Health and Place*, 16: 876-884.

Wang, Y. C., McPherson, K., Marsh, T., Gortmaker, S. L., and Brown, M. 2011. Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet*, 378: 815-825.

Ward, M. D. and Gleditsch, K. 2008. *Spatial Regression Models*. London: Sage.

Weisent, J., Rohrbach, B., Dunn, J. R., and Odoi, A. 2012. Socioeconomic determinants of geographic disparities in campylobacteriosis risk: A comparison of global and local modeling approaches. *International Journal of Health Geographics*, 11: 45.

Wen, M. and Kowaleski-Jones, L. 2012. The built environment and risk of obesity in the United States: Racial-ethnic disparities. *Health Place*, 18(6): 1314-1322.

Wen, M., Zhang, X., Harris, C. D., Holt, J. B., and Croft, J. B. 2013. Spatial access to local public parks and neighborhood socio-demographic characteristics in the United States: A nationwide ecological analysis. *Annals of Behavioral Medicine*, 45: 18-27.

Wen, T. H., Chen, D. R., and Tsai, M. J. 2010. Identifying geographical variations in poverty-obesity relationships: Empirical evidence from Taiwan. *Geospatial Health*, 4(2): 257-265.

Wheeler, D. and Tiefelsdorf, M. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical System*, 7: 161-187.

Whitlock, G., Lewington, S., Sherliker, P., Clarke, R., Emberson, J., Halsey, J., Qizilbash, N., Collins, R., and Peto, R. 2009. Body-mass index and cause-specific mortality in 900 000 adults: Collaborative analyses of 57 prospective studies. *Lancet*, 373: 1083-1096.

Wilson, J. and Fotheringham, A. S. 2007. *The Handbook of Geographic Information Science*. Blackwell Publishing.

Wilson, S., Hutson, M., and Mujahid, M. 2008. How planning and zoning contribute to inequitable development, neighborhood health, and environmental injustice. *Environmental Justice,* 1: 211-216.

Wilson, S. M. 2009. An ecologic framework to study and address environmental justice and community health issues. *Environmental Justice,* 2: 15-24.

World Health Organization (WHO). Website http://www.who.int/dietphysicalactivity/childhood_why/en/. Last accessed October 31, 2015.

Xu, Y. and Wang, F. 2015a. Built environment and obesity by urbanicity in the U.S. *Health and Place*, 34: 19-29.

Xu, Y. and Wang, L. 2015b. GIS-based analysis of obesity and the built environment in the US. *Cartography and Geographic Information Science*, 42: 1.

Yancey, A. K. and Kumanyika, S. K. 2007. Bridging the gap: Understanding the structure of social inequities in childhood obesity. *American Journal of Preventive Medicine*, 33: S172-174.

Yang, T. C., Wu, P. C., Chen V. Y. J., and Su H. J. 2009. Cold surge: A sudden and spatially varying threat to health? *Science of the Total Environment* 407(10): 3421-3424.

Zenk, S. N., Schulz, A. J., and Israel, B. A. 2005. Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. *American Journal of Public Health*, 95: 660-667.

Zhang, X, Lu, H., and Holt, J. B. 2011. Modeling spatial accessibility to parks: A national study. *International Journal of Health Geographics*, 10: 31.

Zhao, F. and Park, N. 2004. Using geographically weighted regression models to estimate annual average daily traffic. *Transportation Research Record 1879*, 99-107. Washington, DC: Transportation Research Board, National Research Council.

Zwald, M. L., Hipp, J. A., Corseuil, M. W., and Dodson, E. A. 2014. Correlates of walking for transportation and use of public transportation among adults in St Louis, Missouri, 2012. *Prevalence of Chronic Disease*, 11: 112.

APPENDIX A
UNIVARIATE LISA SIGNIFICANCE MAPS

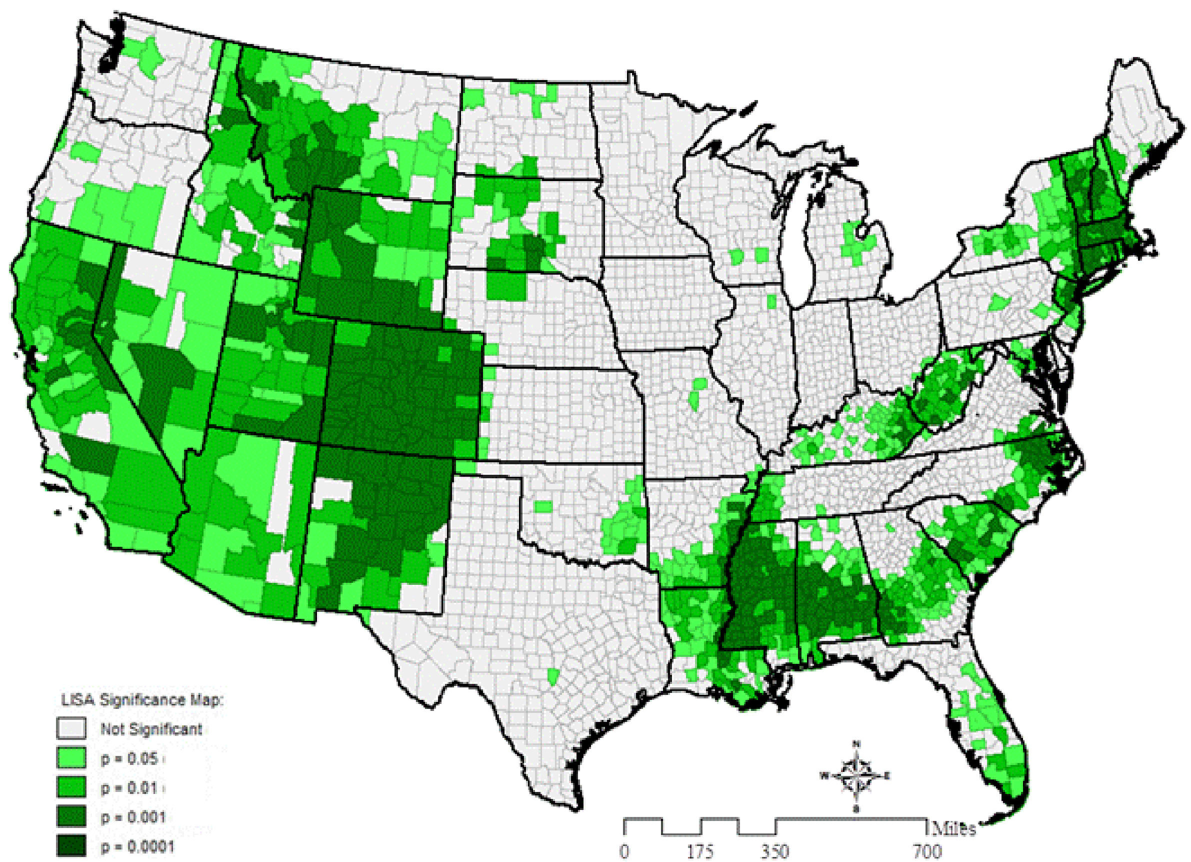

Figure A.1 Univariate LISA significance map for county level obesity prevalence rates in the US for 2004

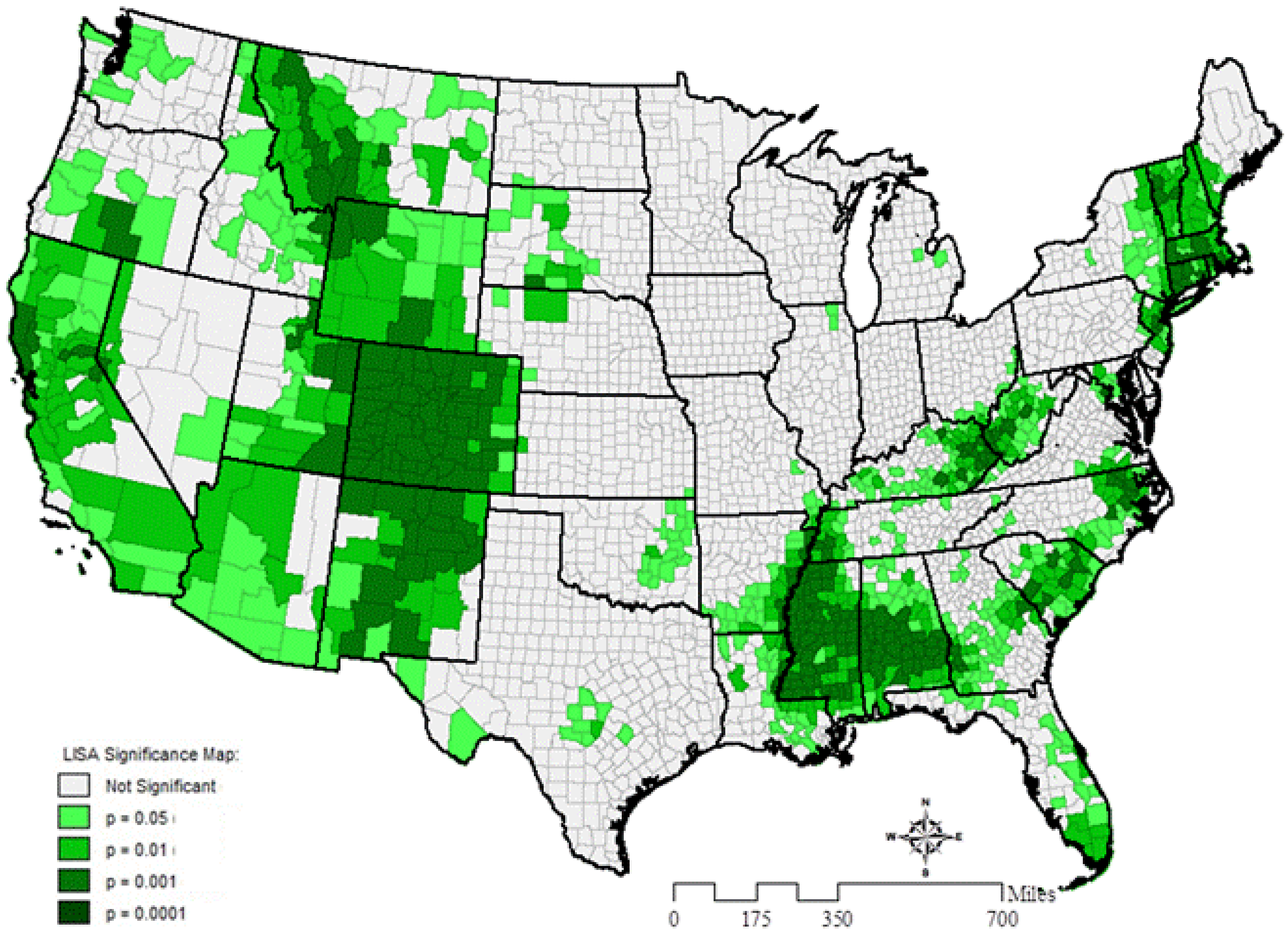

Figure A.2 Univariate LISA significance map for county level obesity prevalence rates in the US for 2007

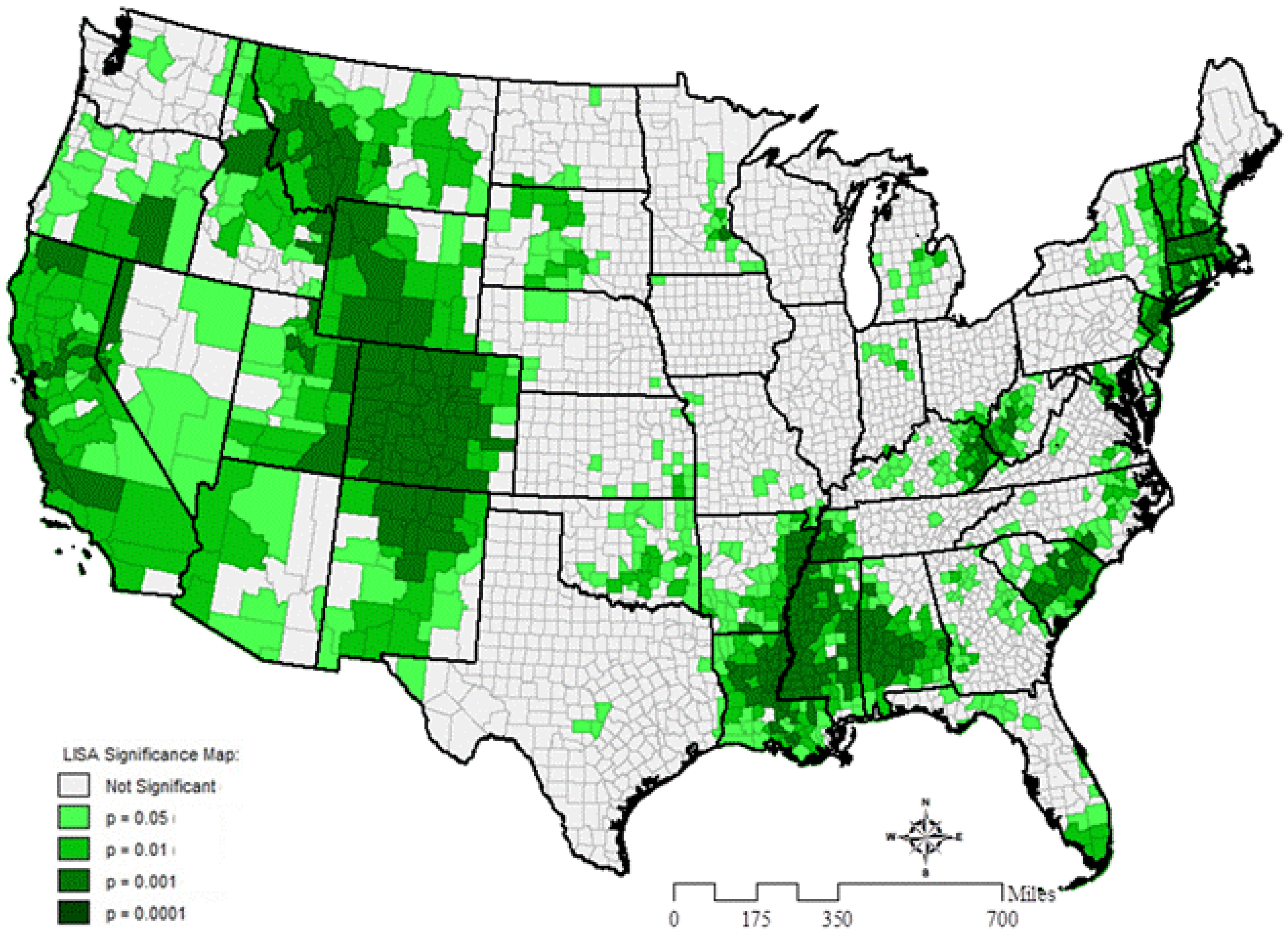

Figure A.3 Univariate LISA significance map for county level obesity prevalence rates in the US for 2010

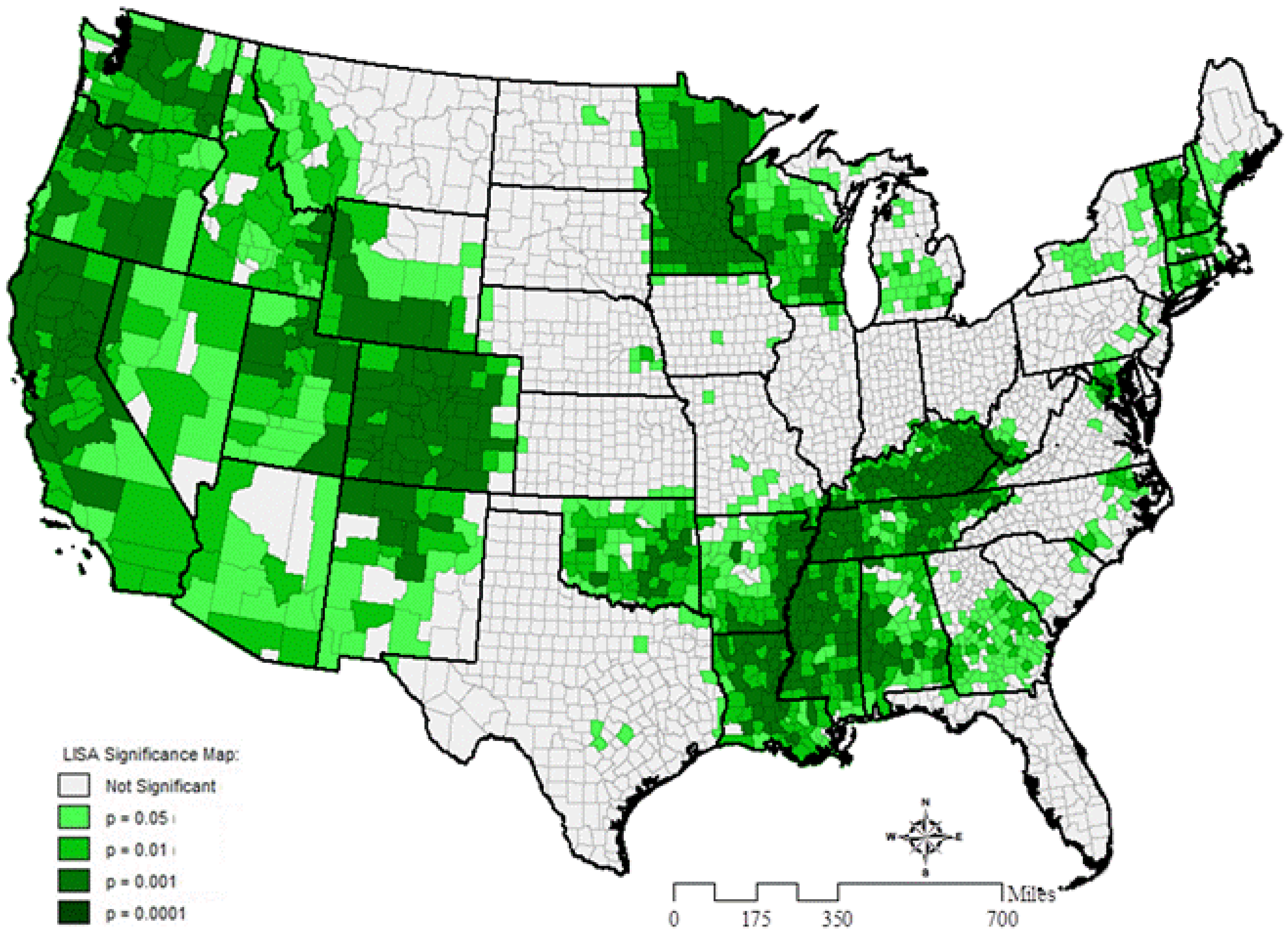

Figure A.4 Univariate LISA significance map for county level physical inactivity (PIA) prevalence rates in the US for 2004

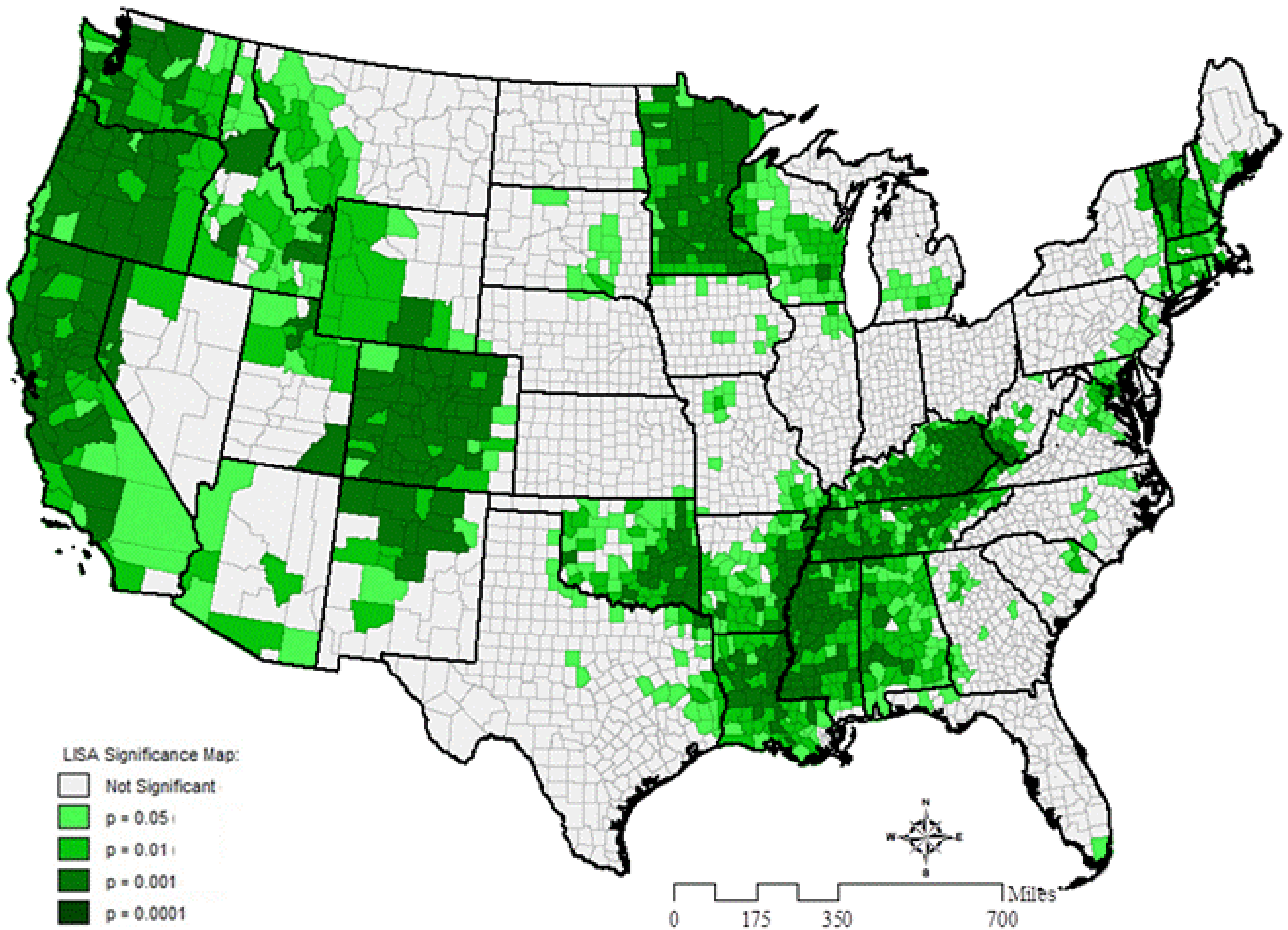

Figure A.5 Univariate LISA significance map for county level physical inactivity (PIA) prevalence rates in the US for 2007

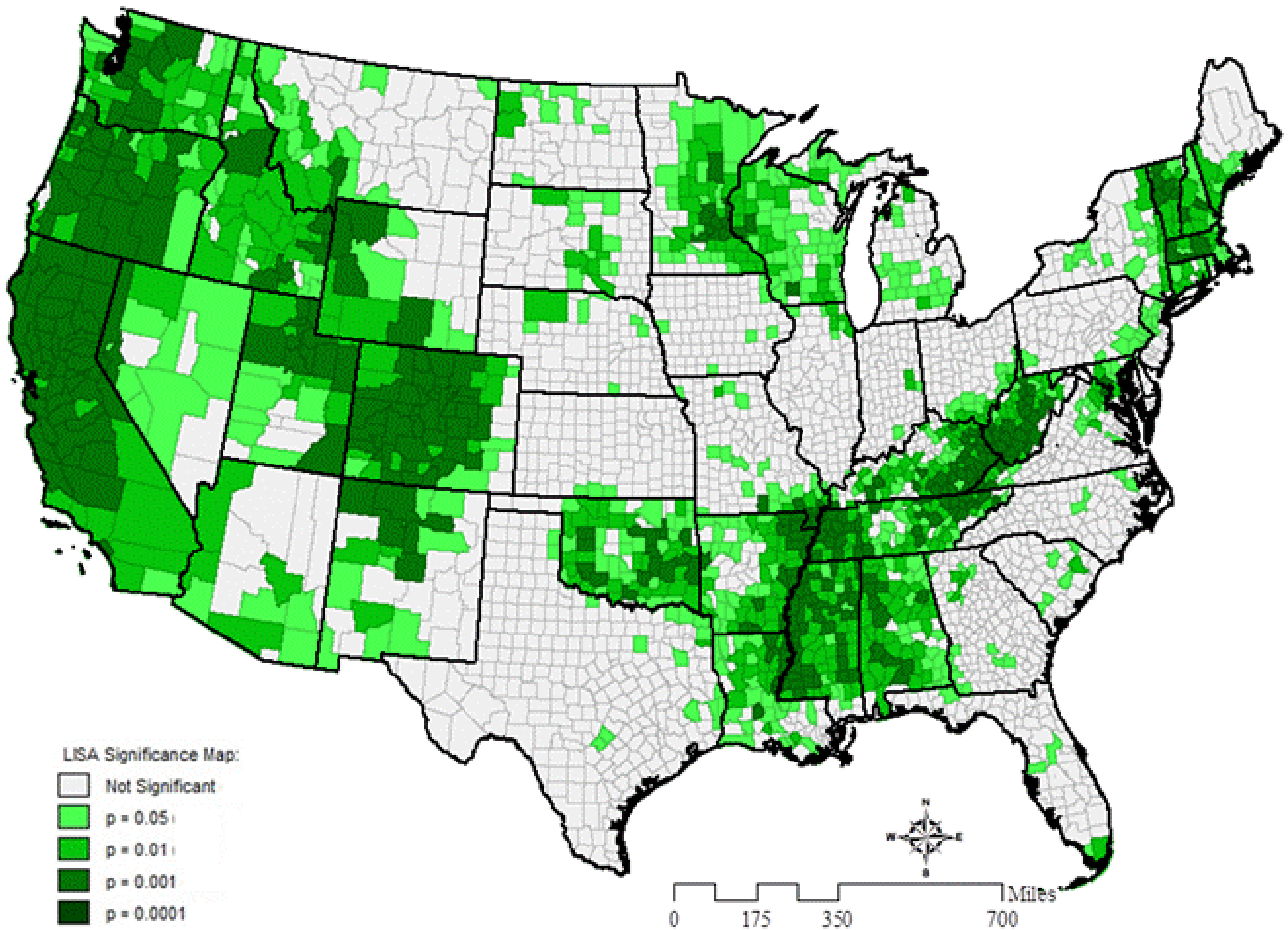

Figure A.6 Univariate LISA significance map for county level physical inactivity (PIA) prevalence rates in the US for 2010

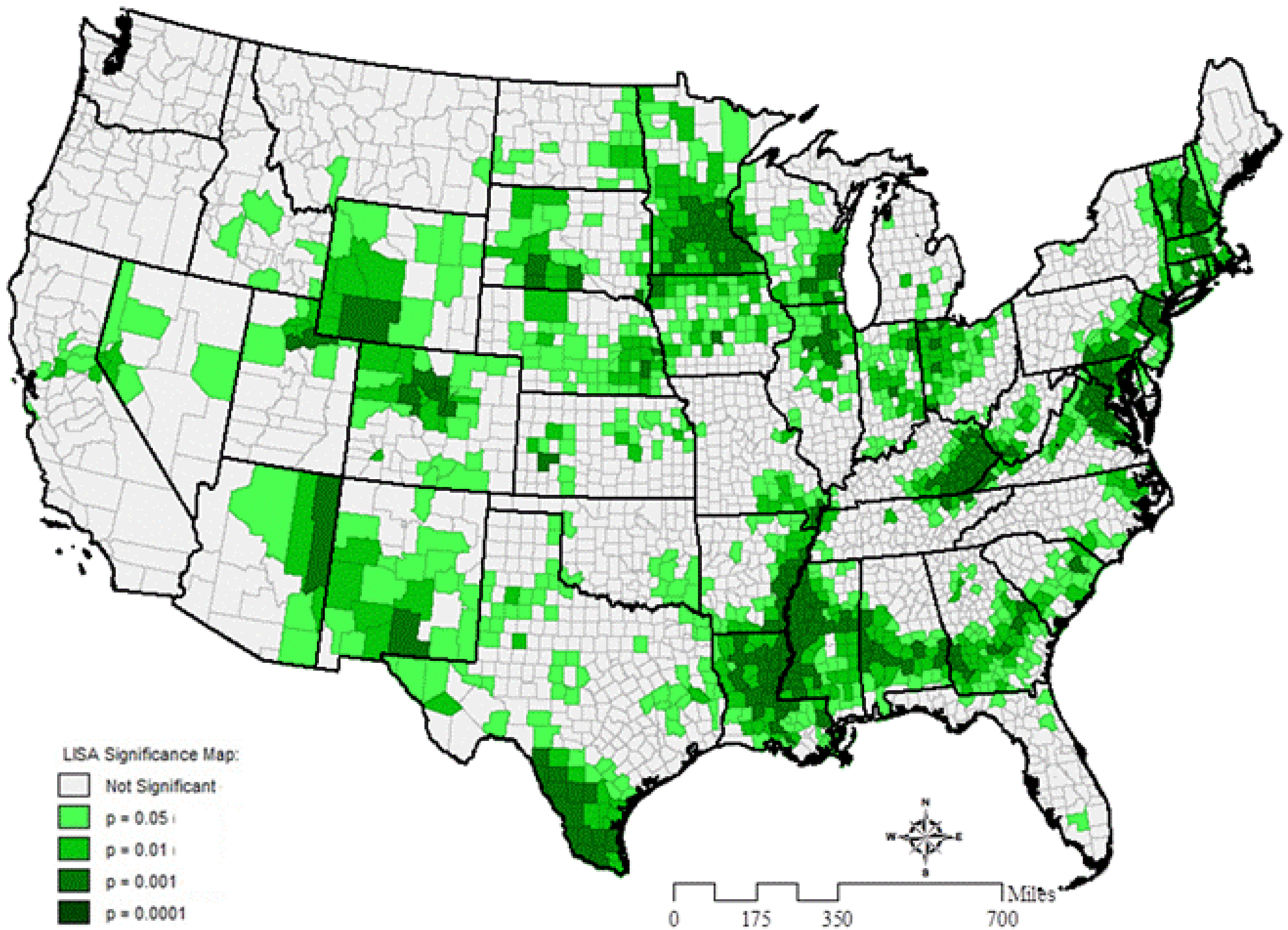

Figure A.7 Univariate LISA significance map for county level poverty prevalence rates in the US for 2004

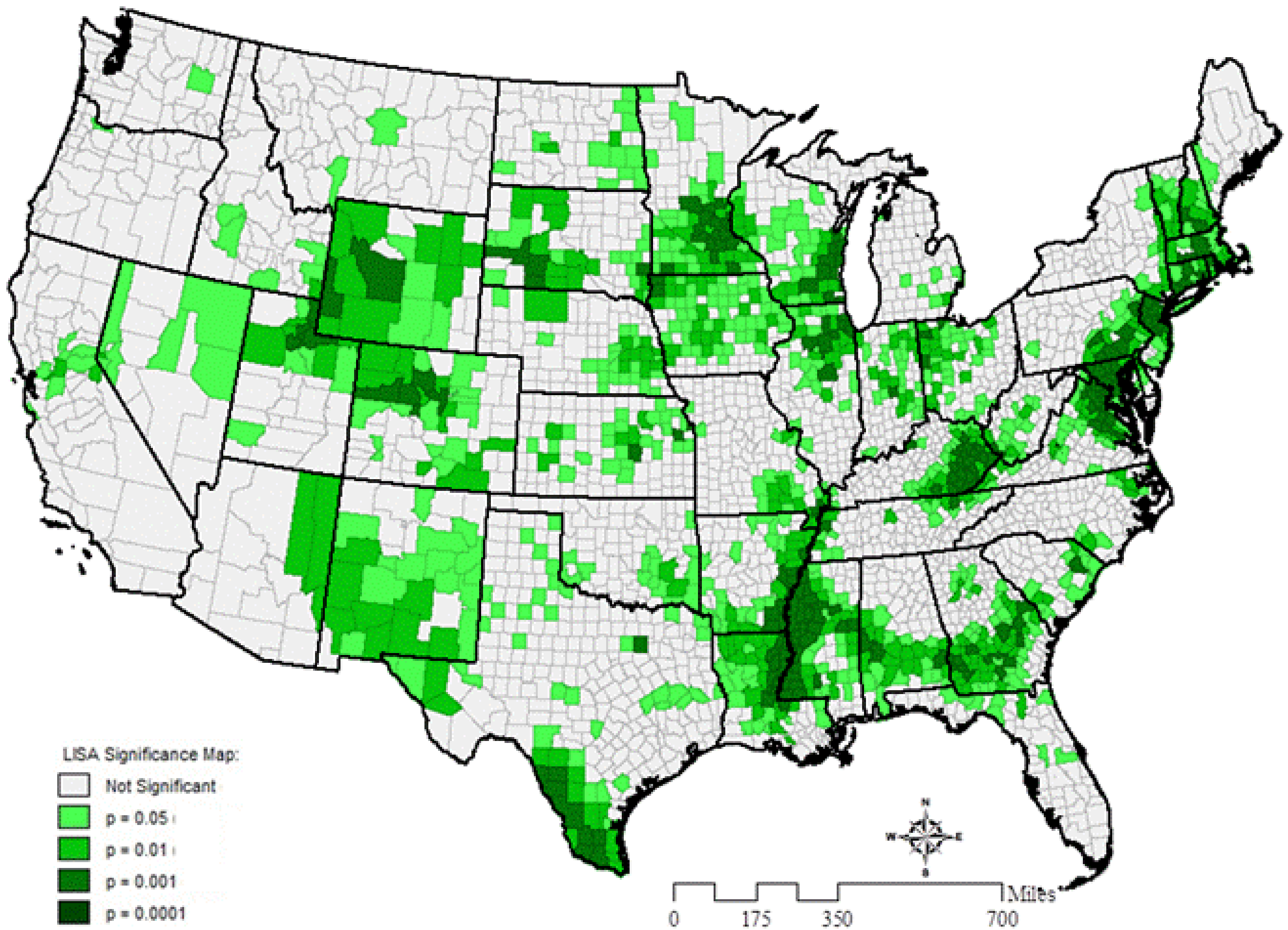

Figure A.8 Univariate LISA significance map for county level poverty prevalence rates in the US for 2007

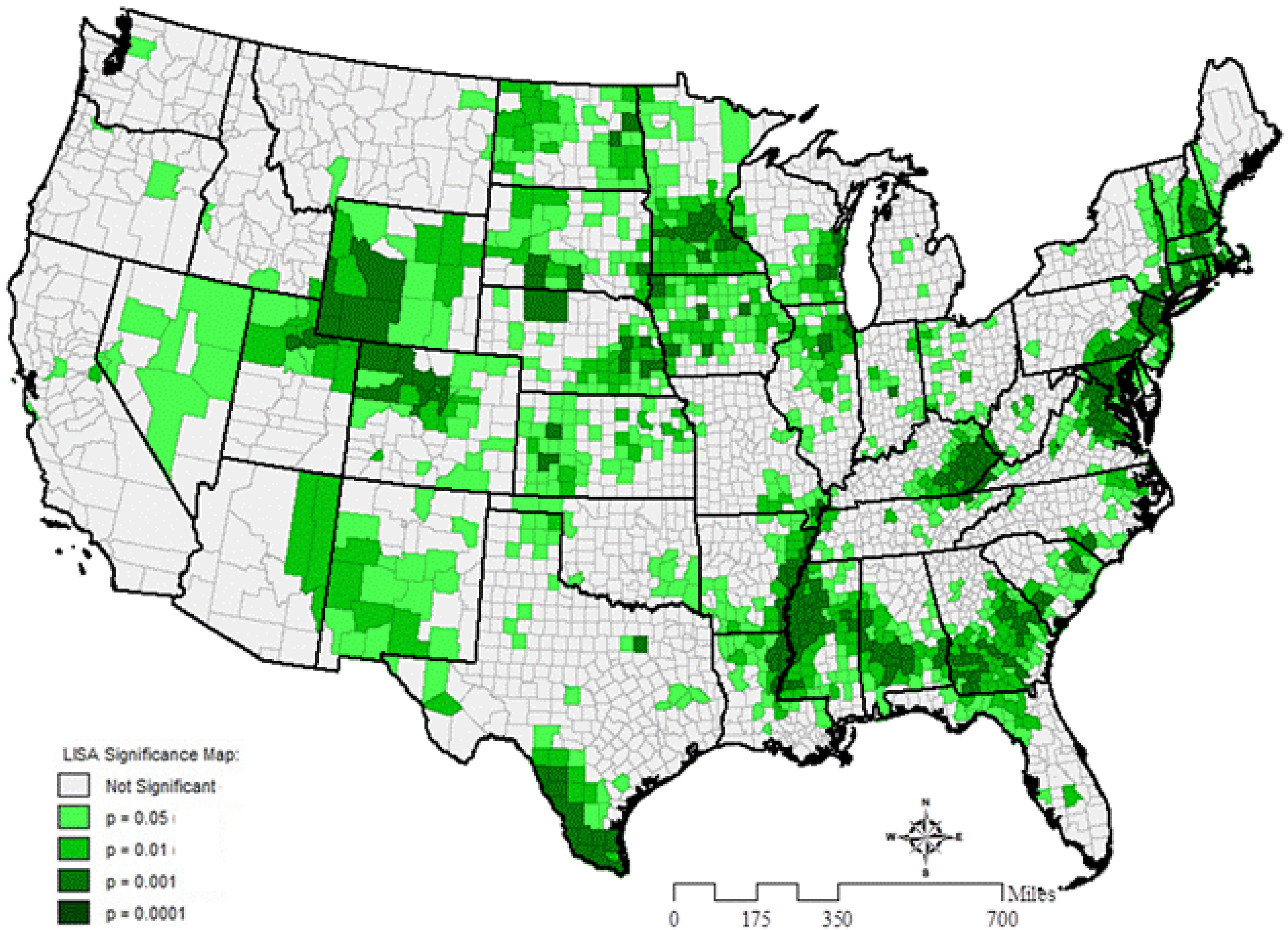

Figure A.9 Univariate LISA significance map for county level poverty prevalence rates in the US for 2010

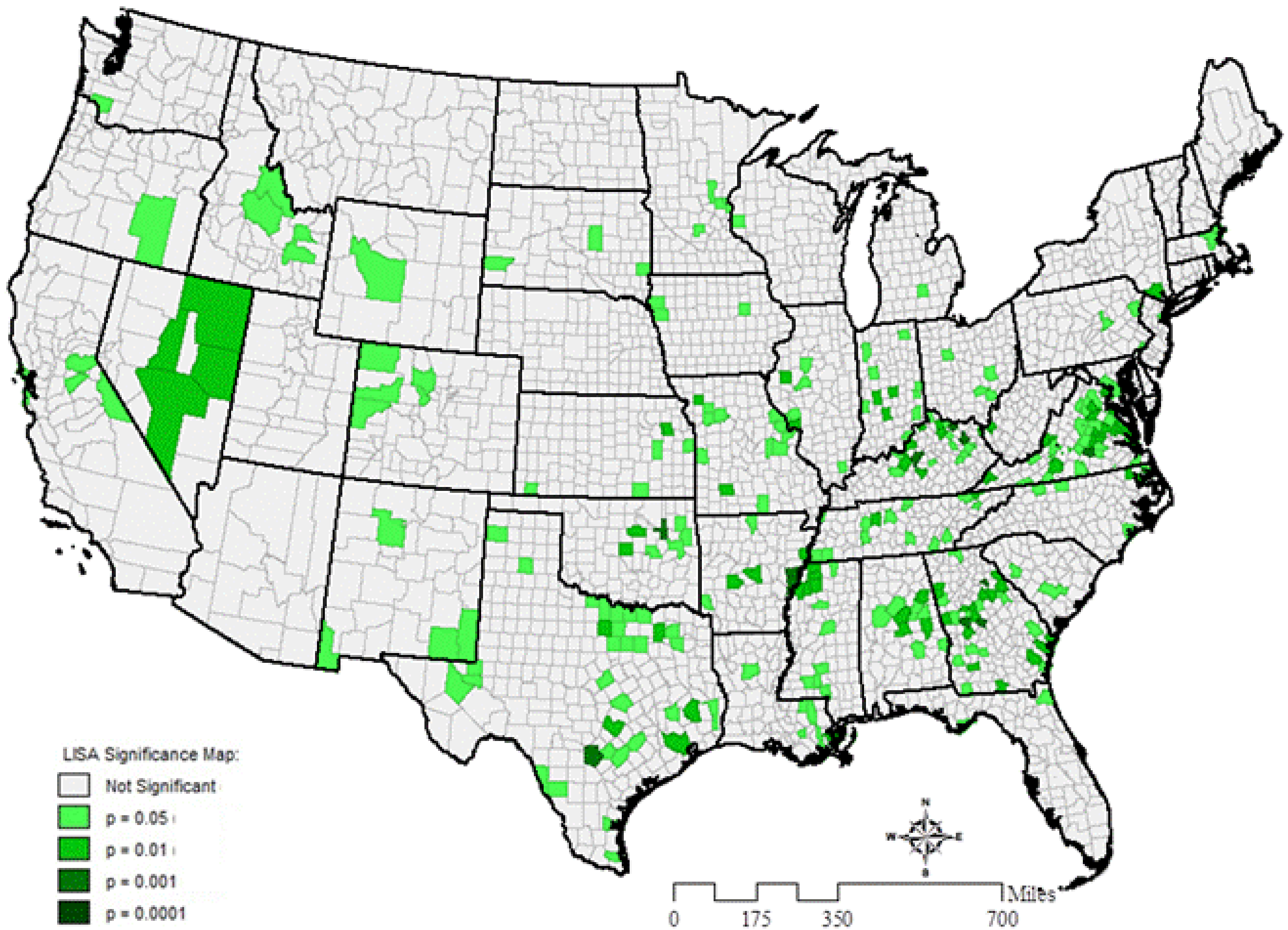

Figure A.10 Univariate LISA significance map for county level ratio of jobs to employed residents (JERs) in the US for 2010

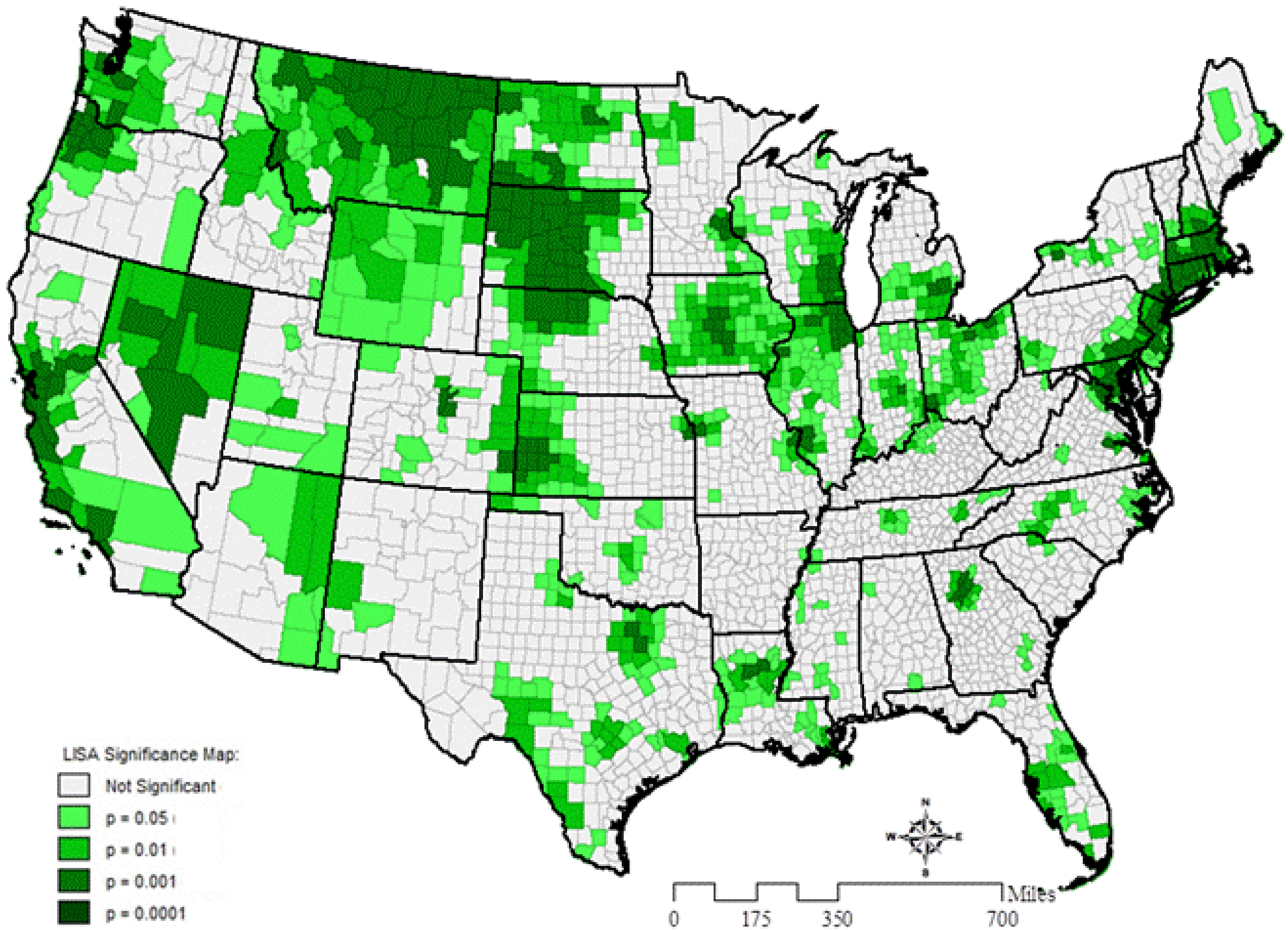

Figure A.11 Univariate LISA significance map for county level population-weighted distances (PWDs) to parks in the US for 2008

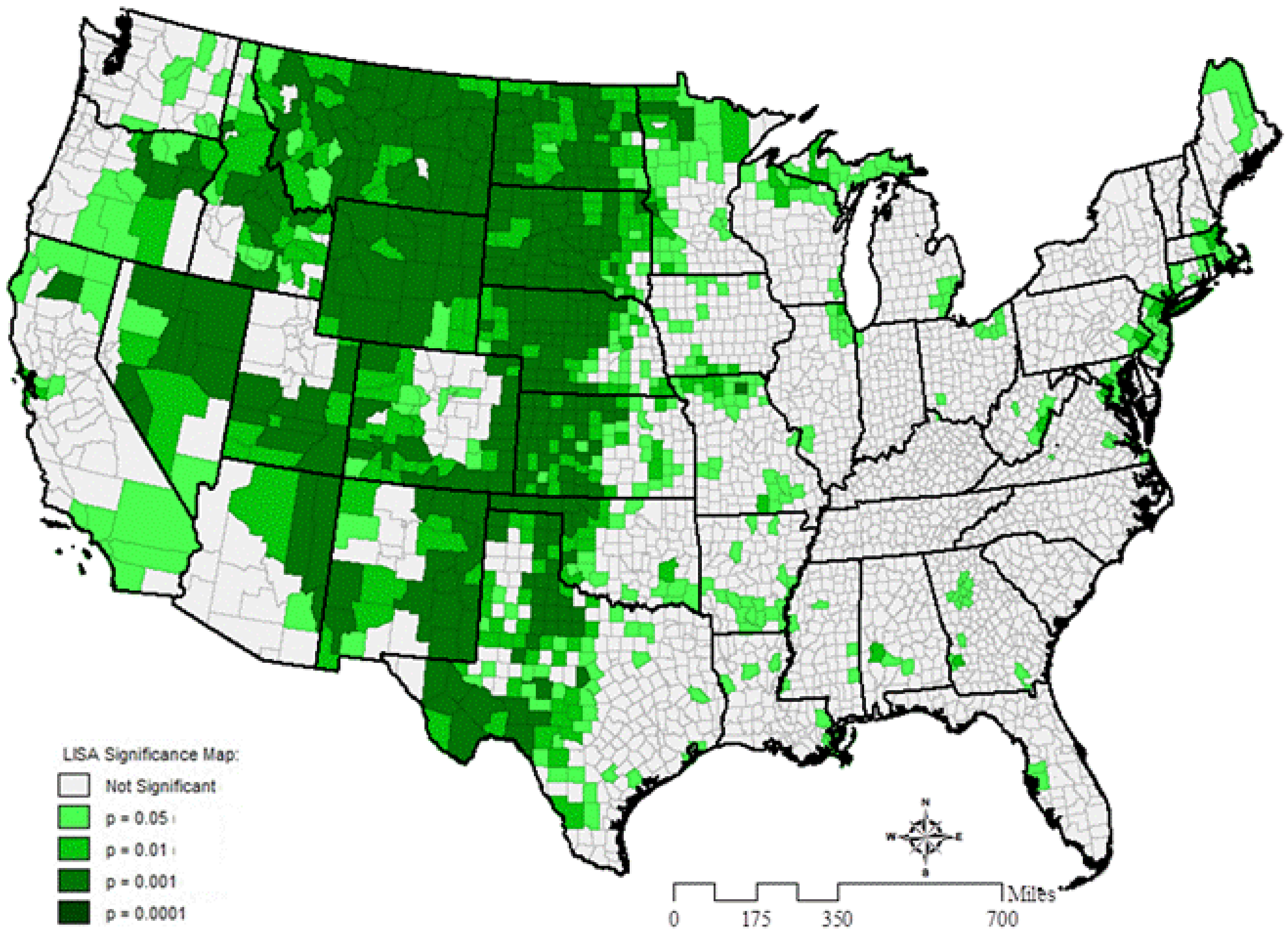

Figure A.12 Univariate LISA significance map for county level population densities in the US for 2004

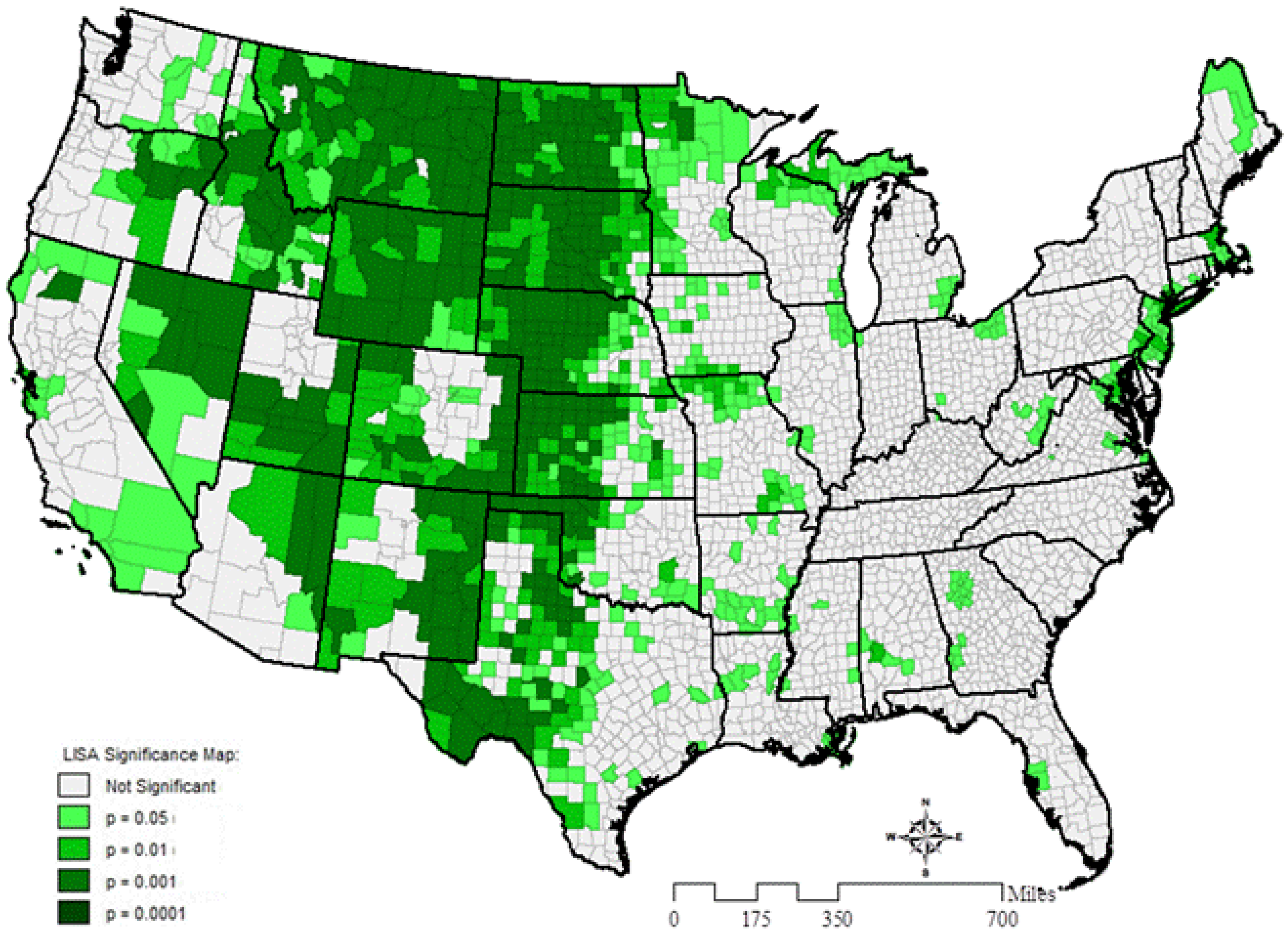

Figure A.13 Univariate LISA significance map for county level population densities in the US for 2007

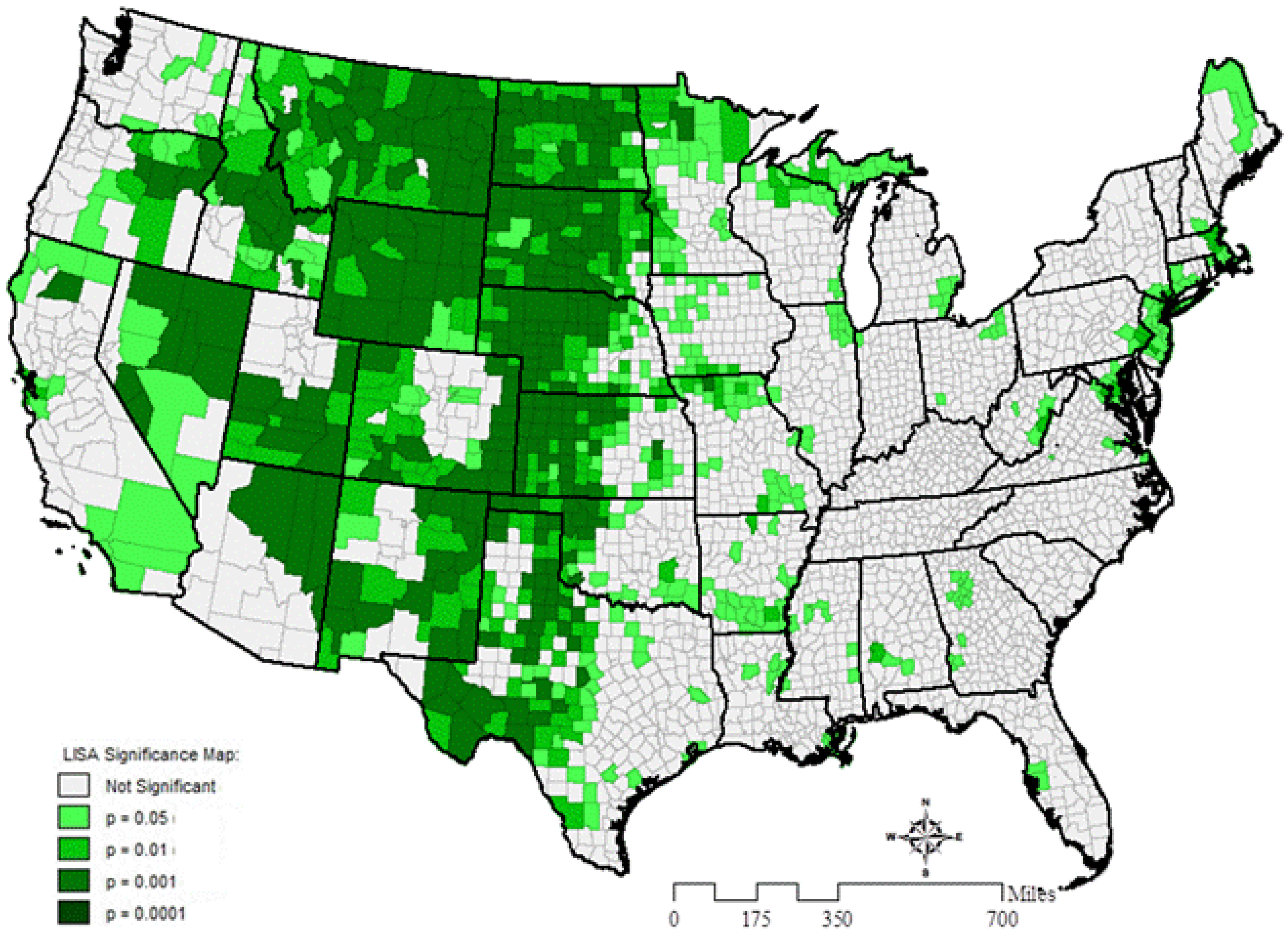

Figure A.14 Univariate LISA significance map for county level population densities in the US for 2010

APPENDIX B
BIVARIATE LISA SIGNIFICANCE MAPS

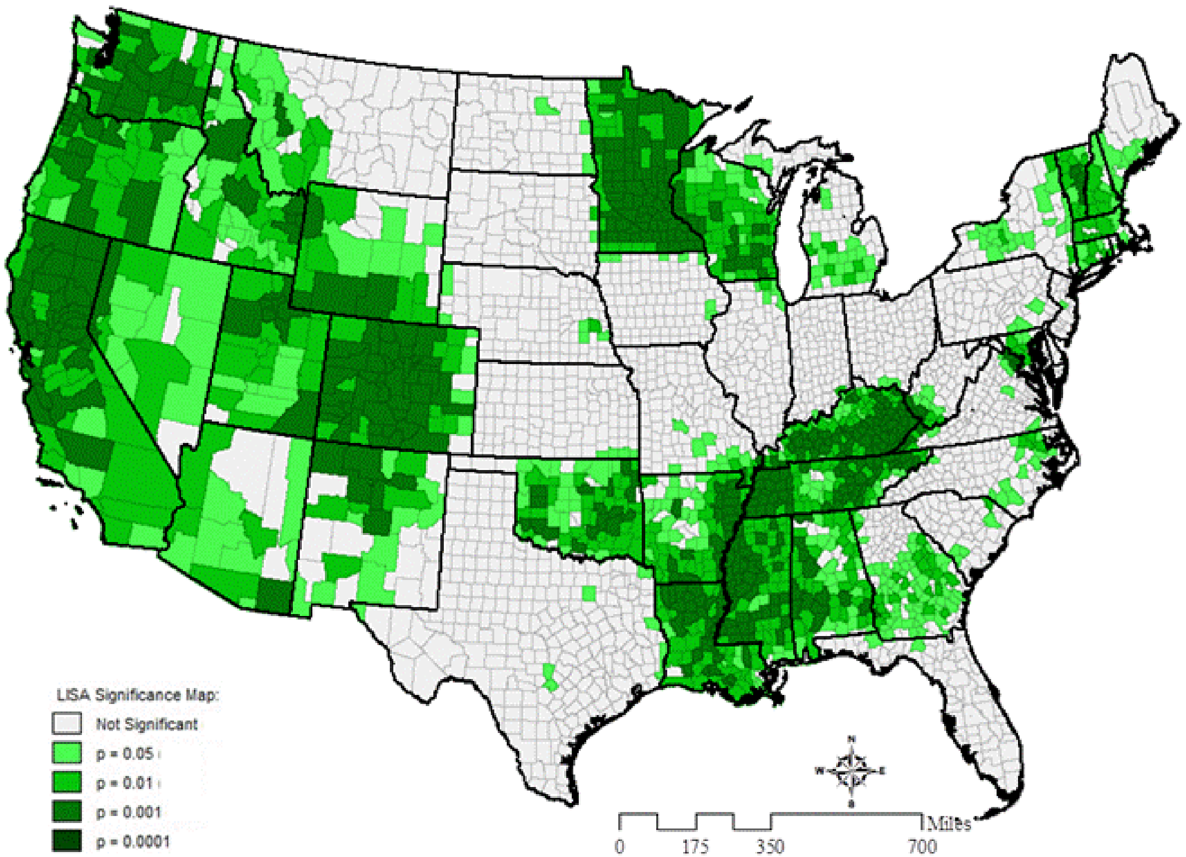

Figure B.1 Bivariate LISA significance map for county level obesity prevalence rates and physical inactivity (PIA) prevalence rates in the US for 2004

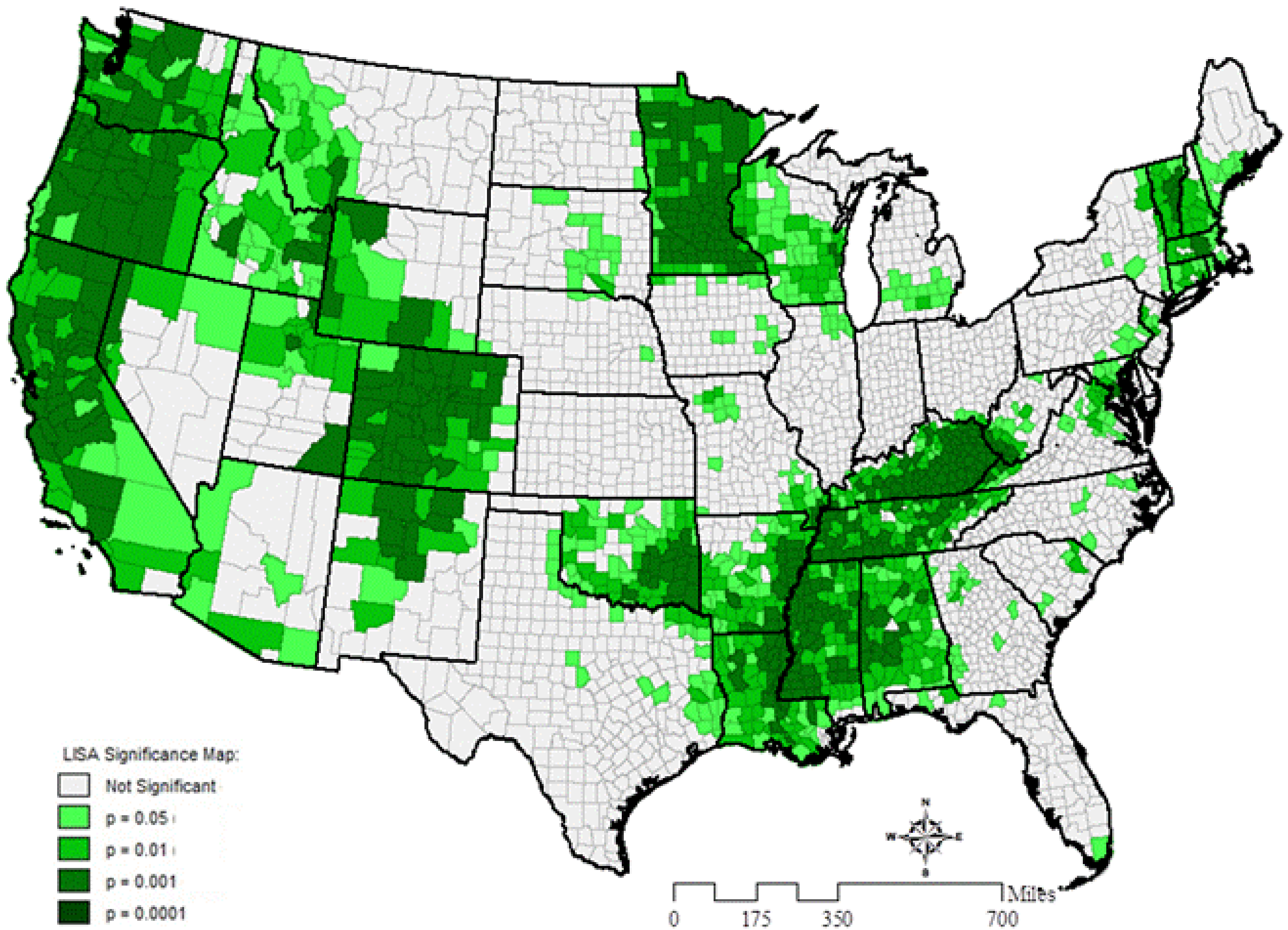

Figure B.2 Bivariate LISA significance map for county level obesity prevalence rates and physical inactivity (PIA) prevalence rates in the US for 2007

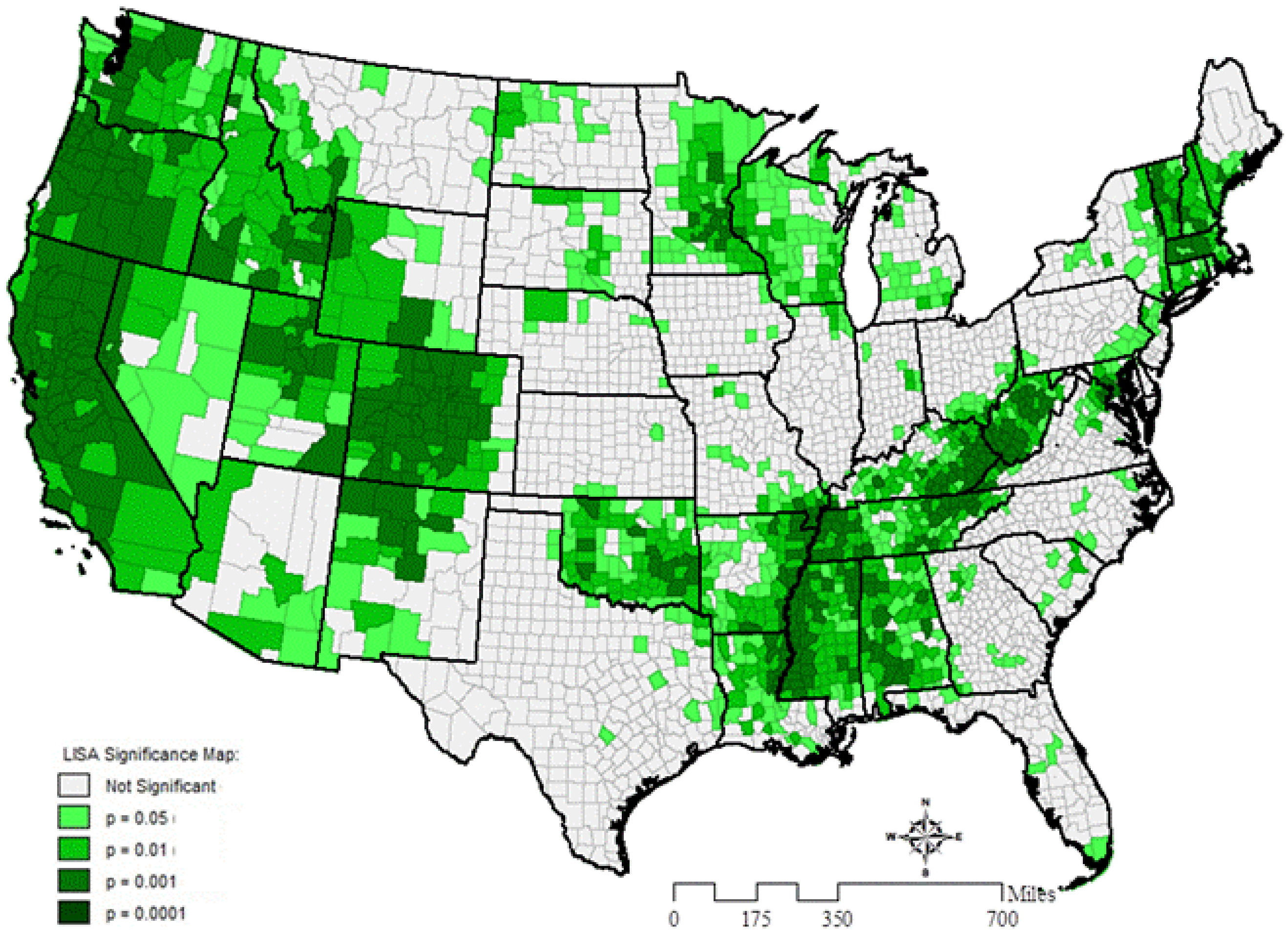

Figure B.3 Bivariate LISA significance map for county level obesity prevalence rates and physical inactivity (PIA) prevalence rates in the US for 2010

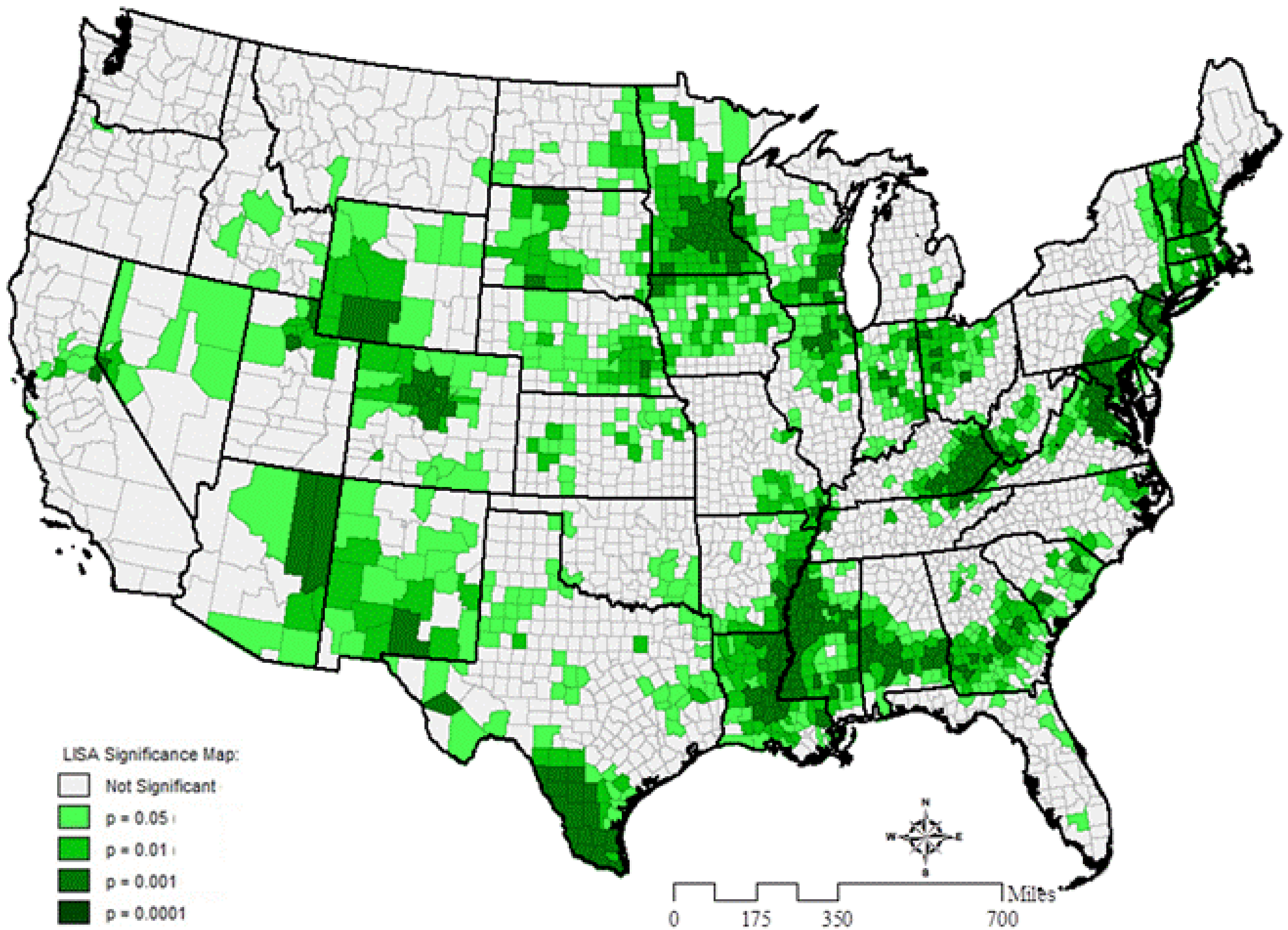

Figure B.4 Bivariate LISA significance map for county level obesity prevalence rates and poverty prevalence rates in the US for 2004

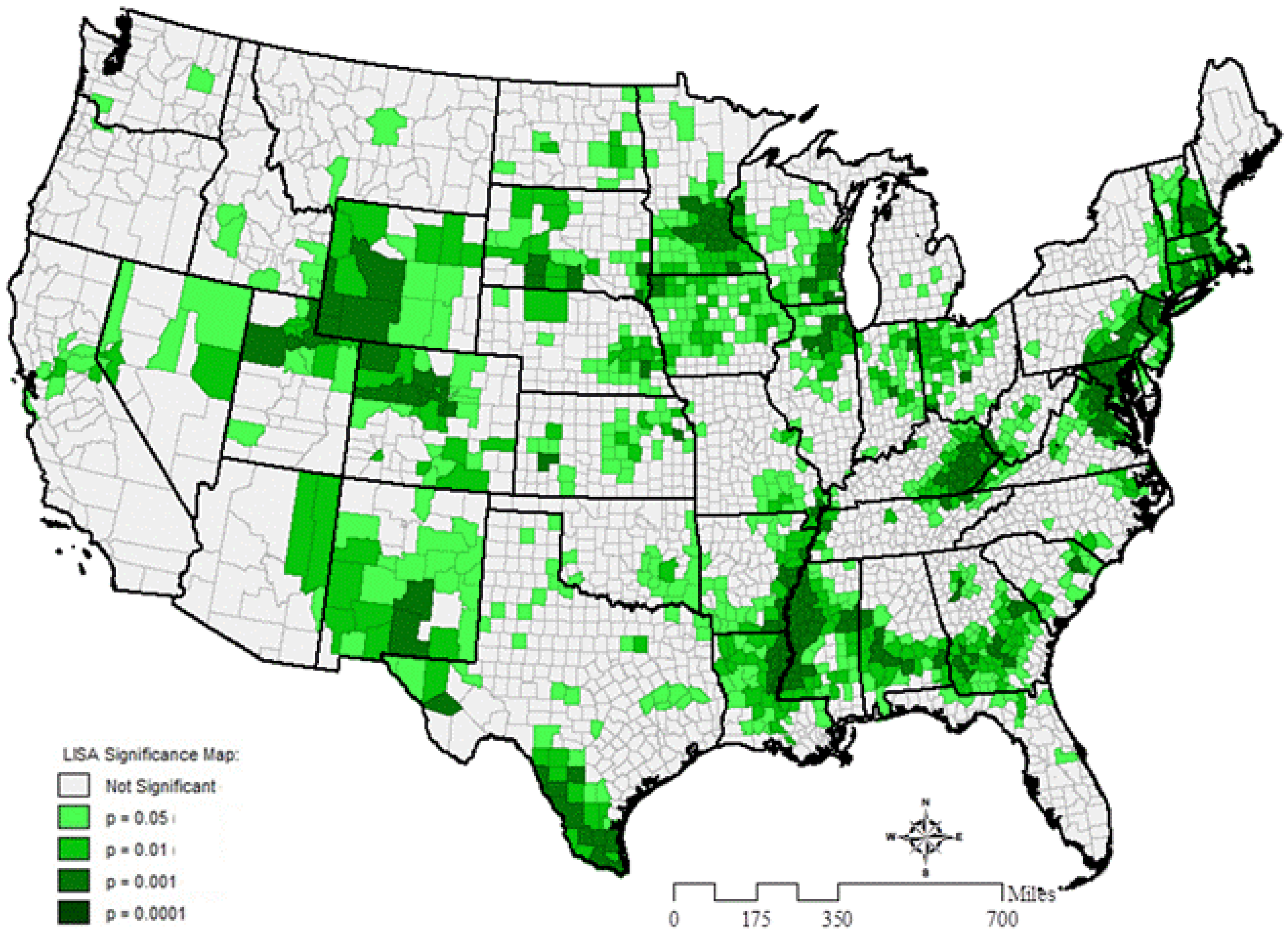

Figure B.5 Bivariate LISA significance map for county level obesity prevalence rates and poverty prevalence rates in the US for 2007

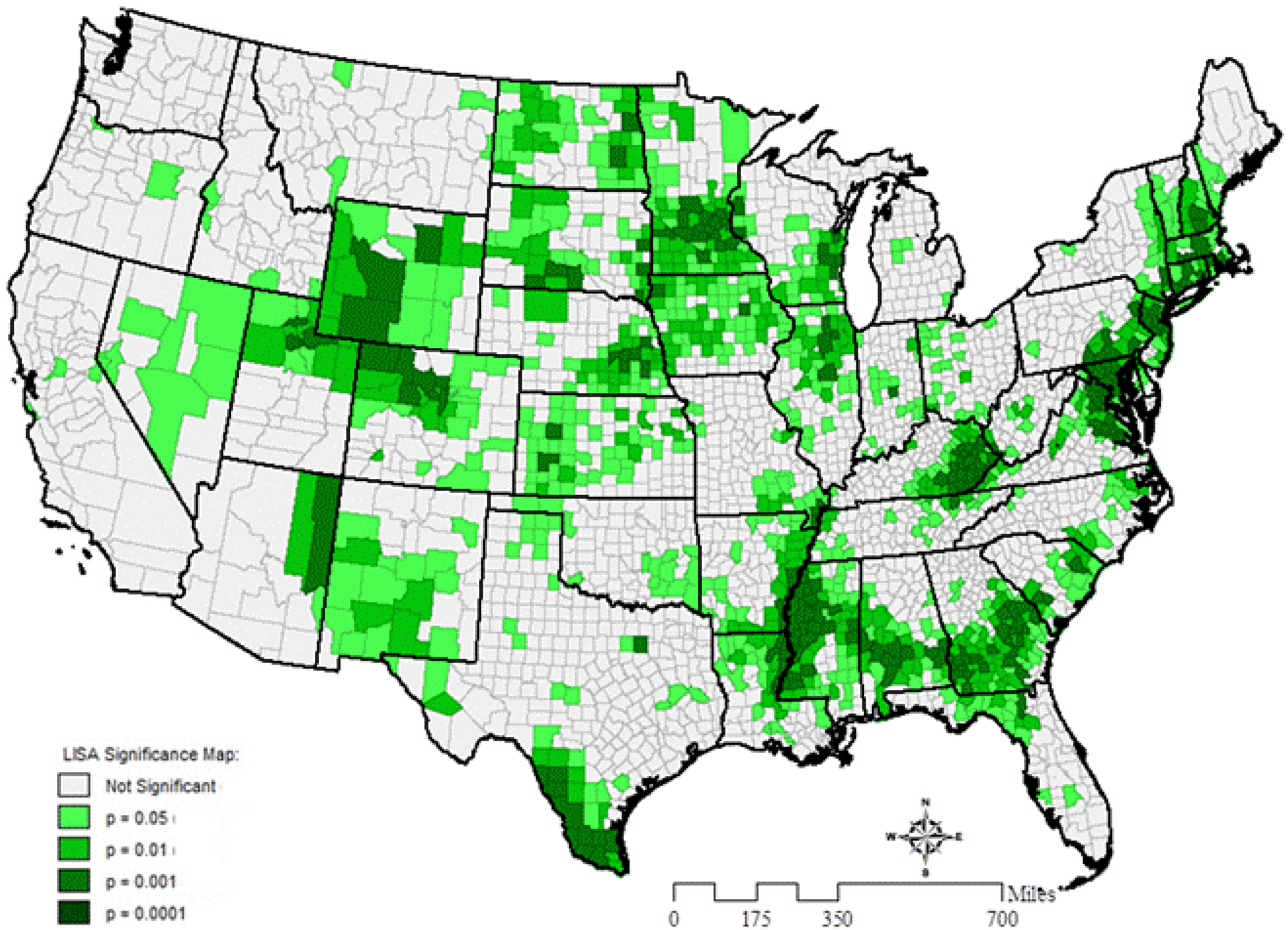

Figure B.6 Bivariate LISA significance map for county level obesity prevalence rates and poverty prevalence rates in the US for 2010

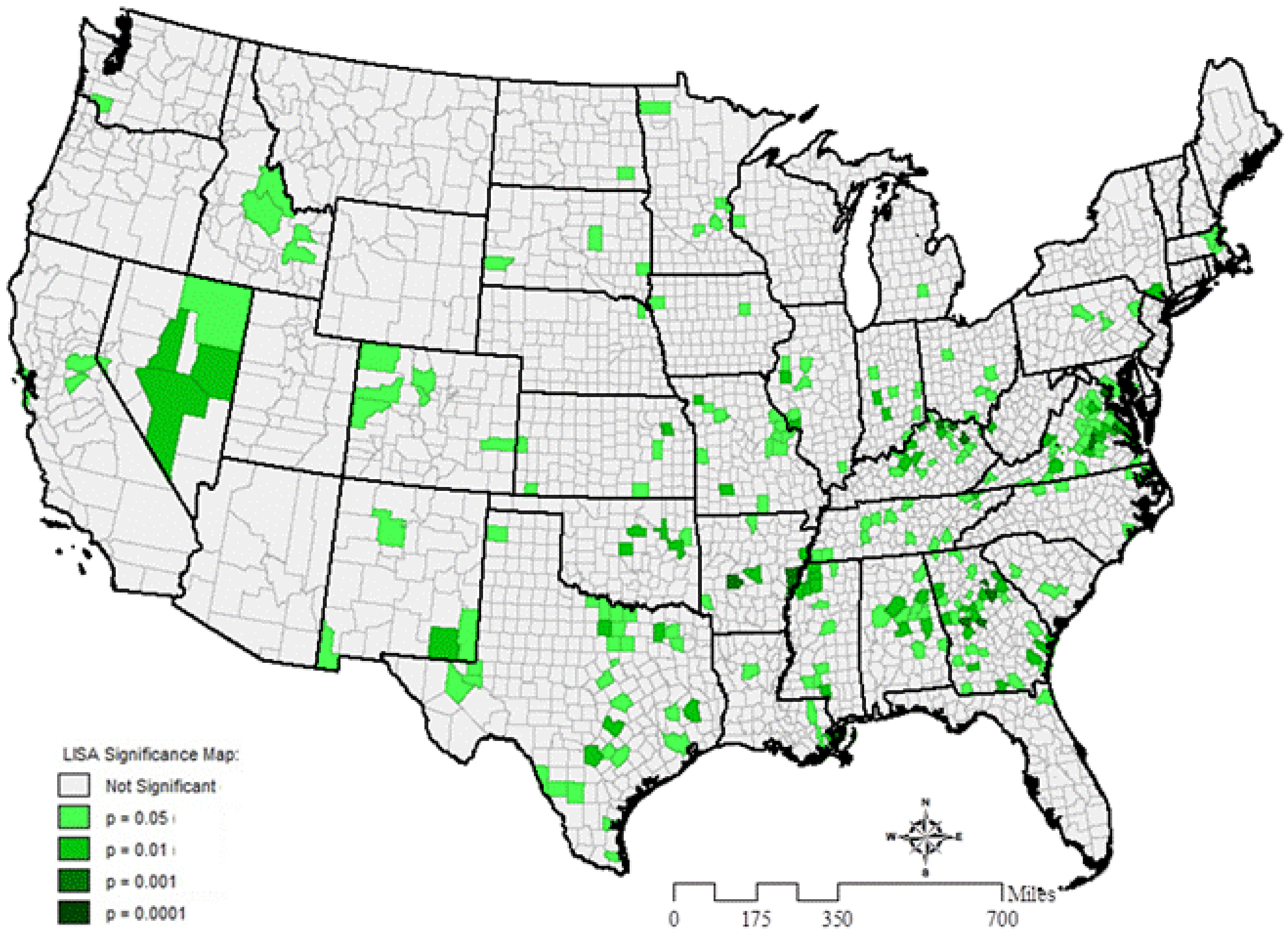

Figure B.7 Bivariate LISA significance map for county level obesity prevalence rates and ratio of jobs to employed residents (JERs) in the US for 2004

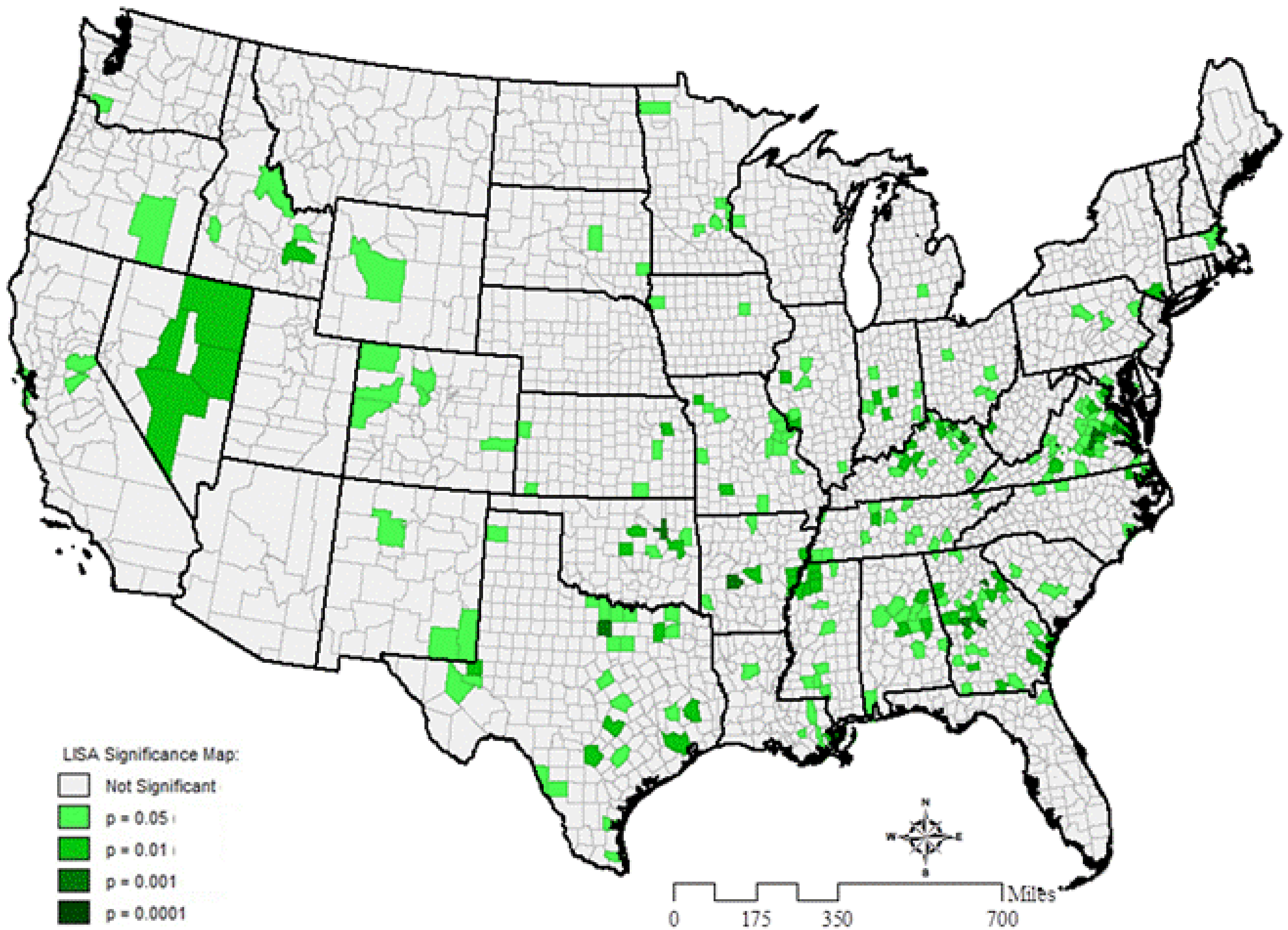

Figure B.8 Bivariate LISA significance map for county level obesity prevalence rates and ratio of jobs to employed residents (JERs) in the US for 2007

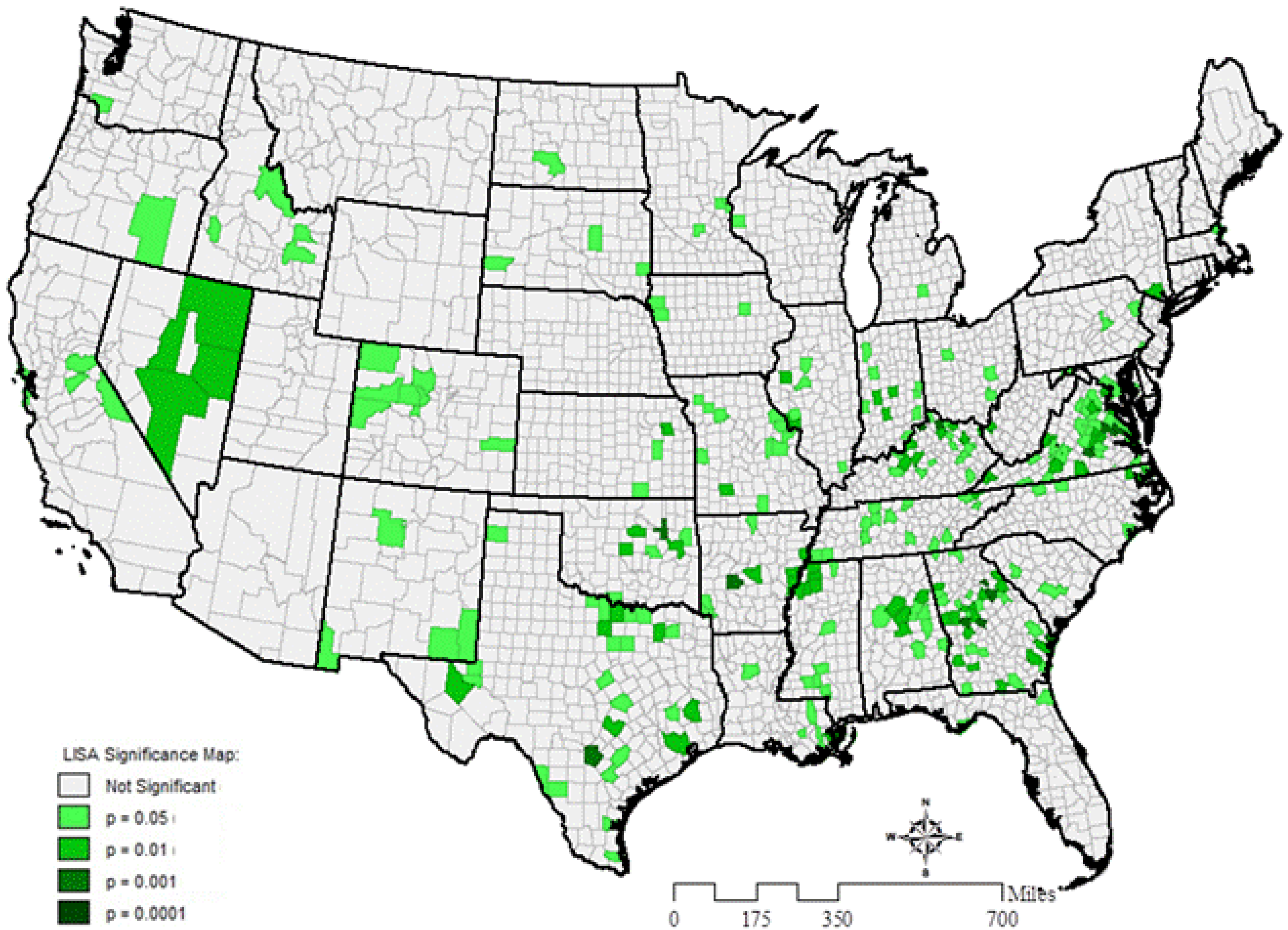

Figure B.9 Bivariate LISA significance map for county level obesity prevalence rates and ratio of jobs to employed residents (JERs) in the US for 2010

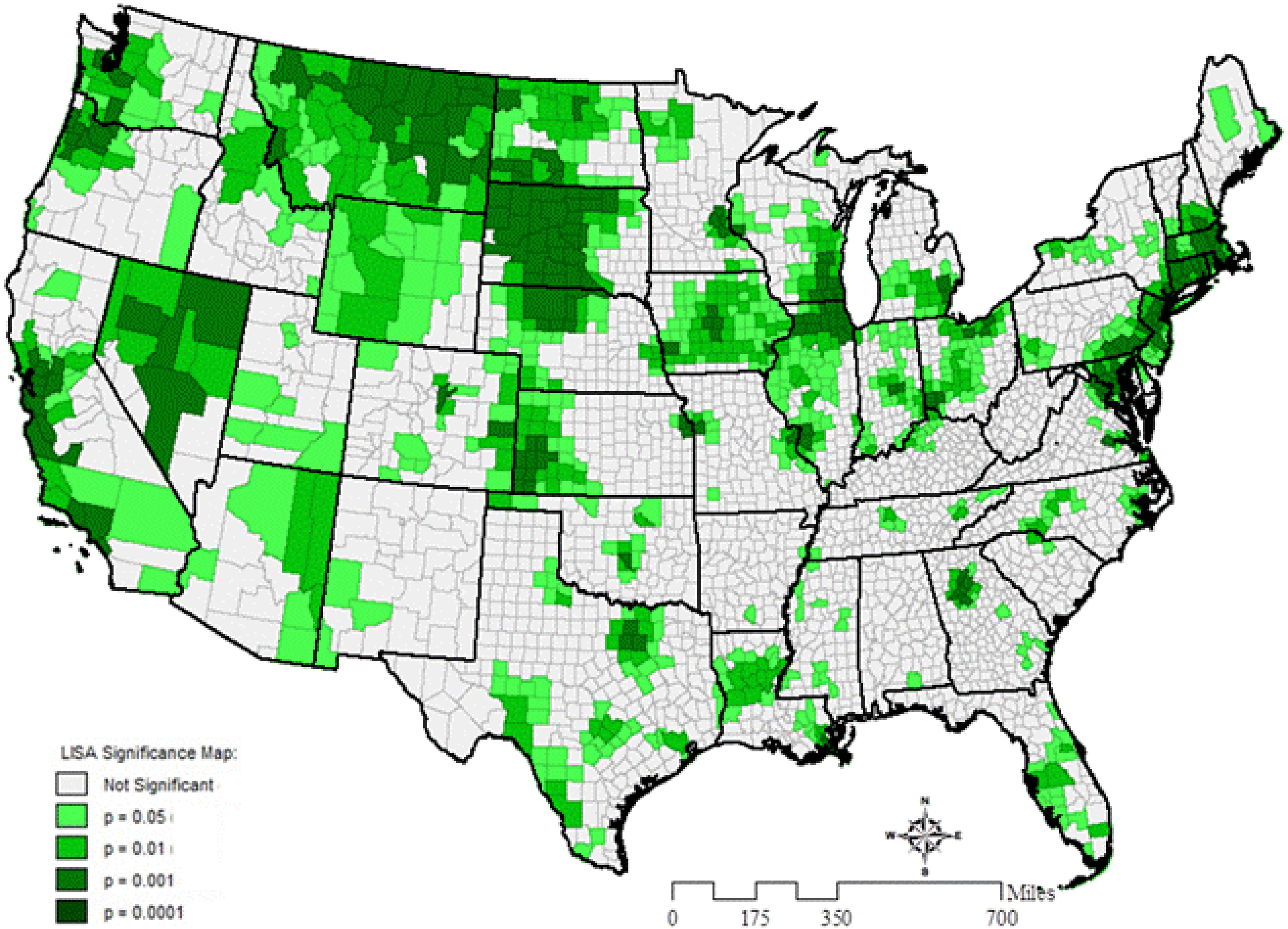

Figure B.10 Bivariate LISA significance map for county level obesity prevalence rates and population-weighted distances (PWDs) to parks in the US for 2004

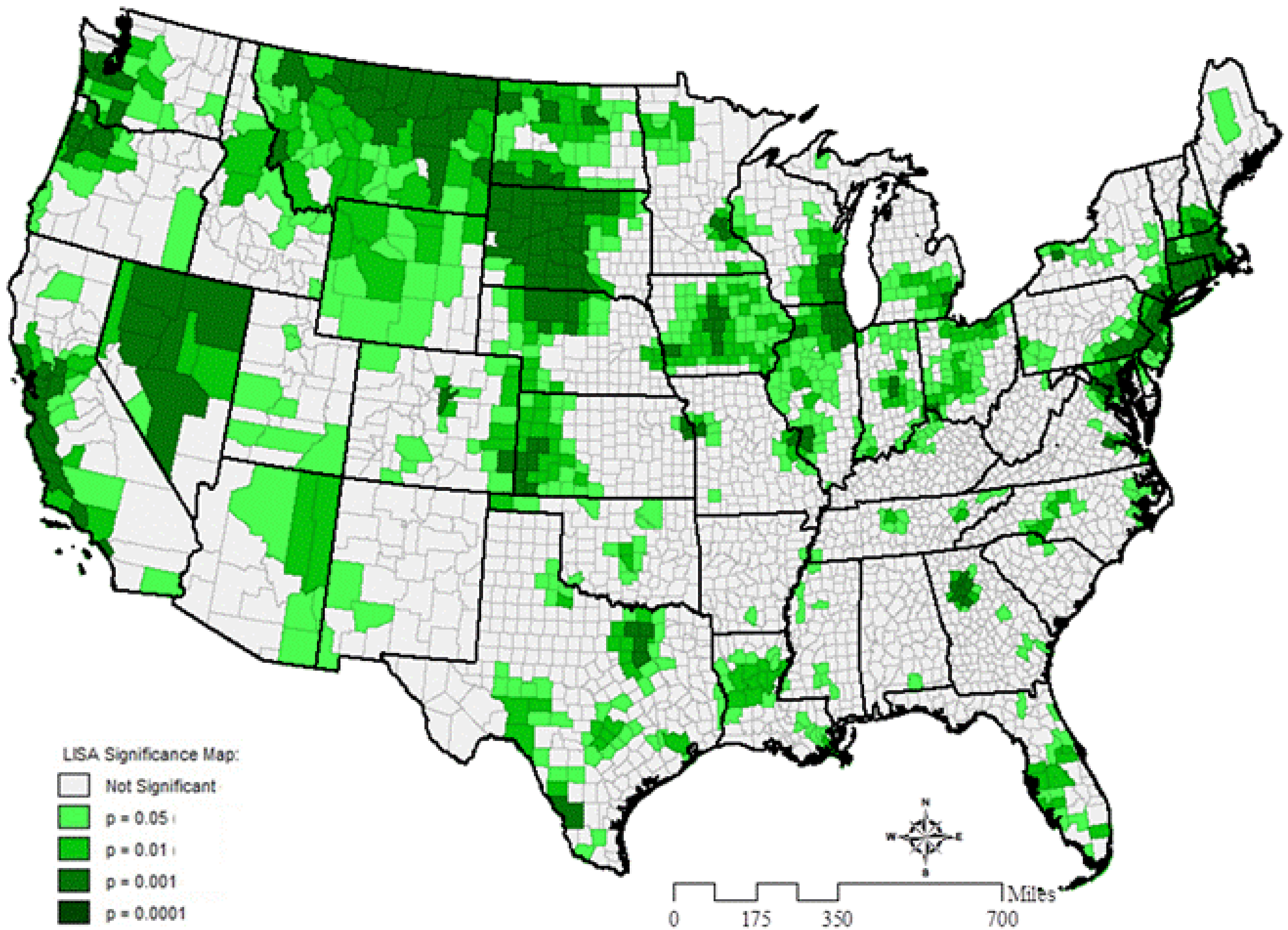

Figure B.11 Bivariate LISA significance map for county level obesity prevalence rates and population-weighted distances (PWDs) to parks in the US for 2007

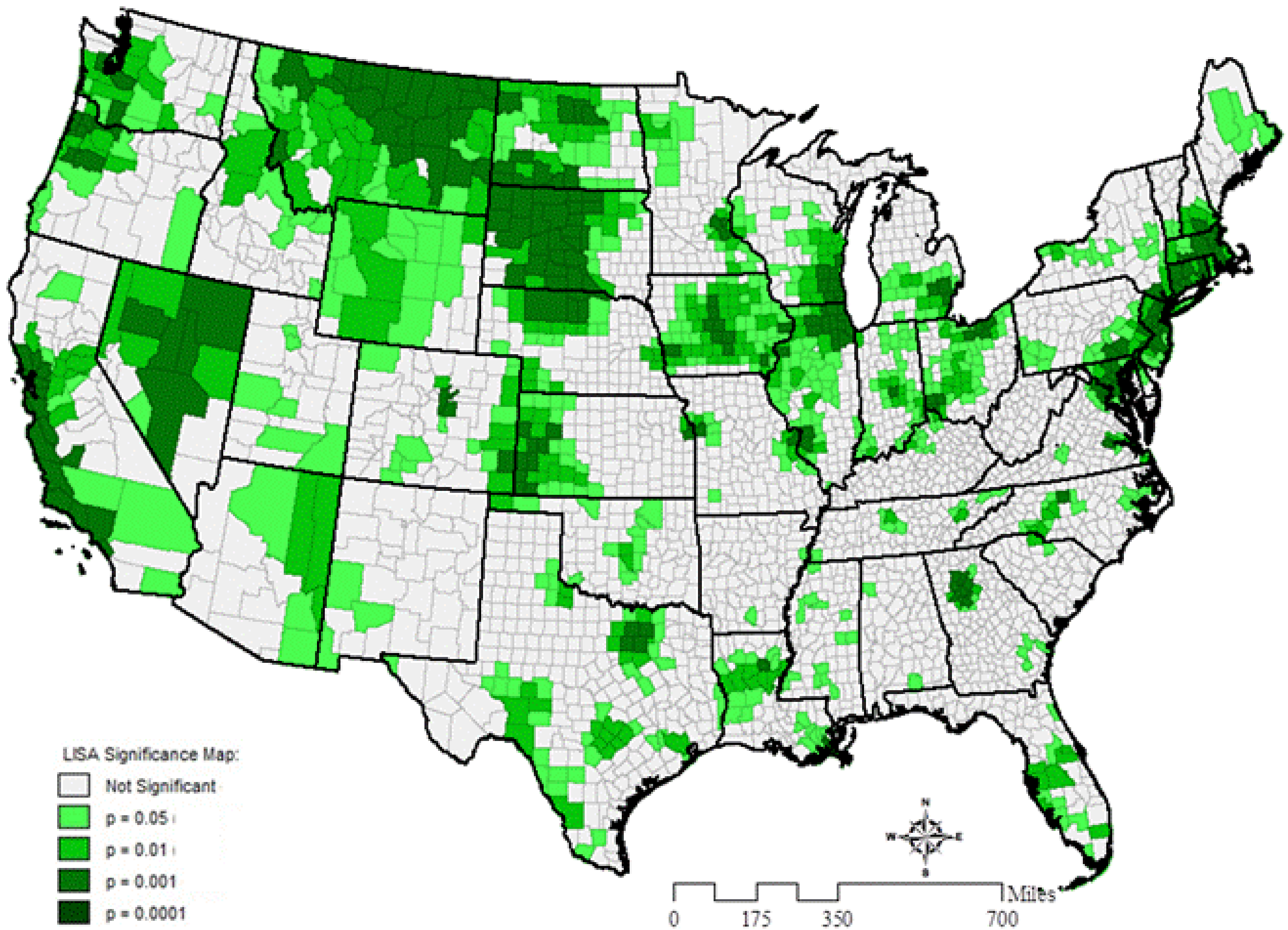

Figure B.12 Bivariate LISA significance map for county level obesity prevalence rates and population-weighted distances (PWDs) to parks in the US for 2010

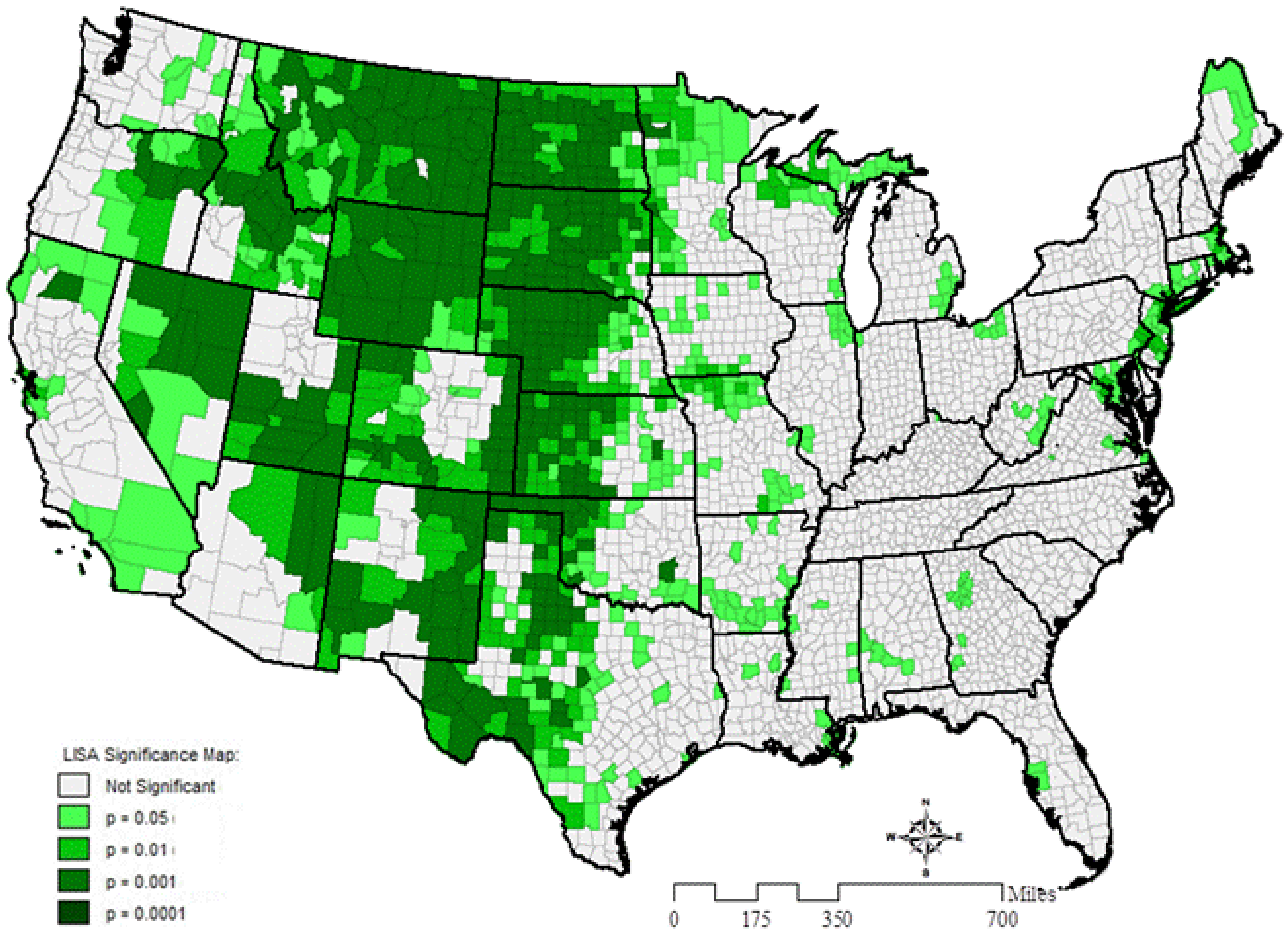

Figure B.13 Bivariate LISA significance map for county level obesity prevalence rates and population densities in the US for 2004

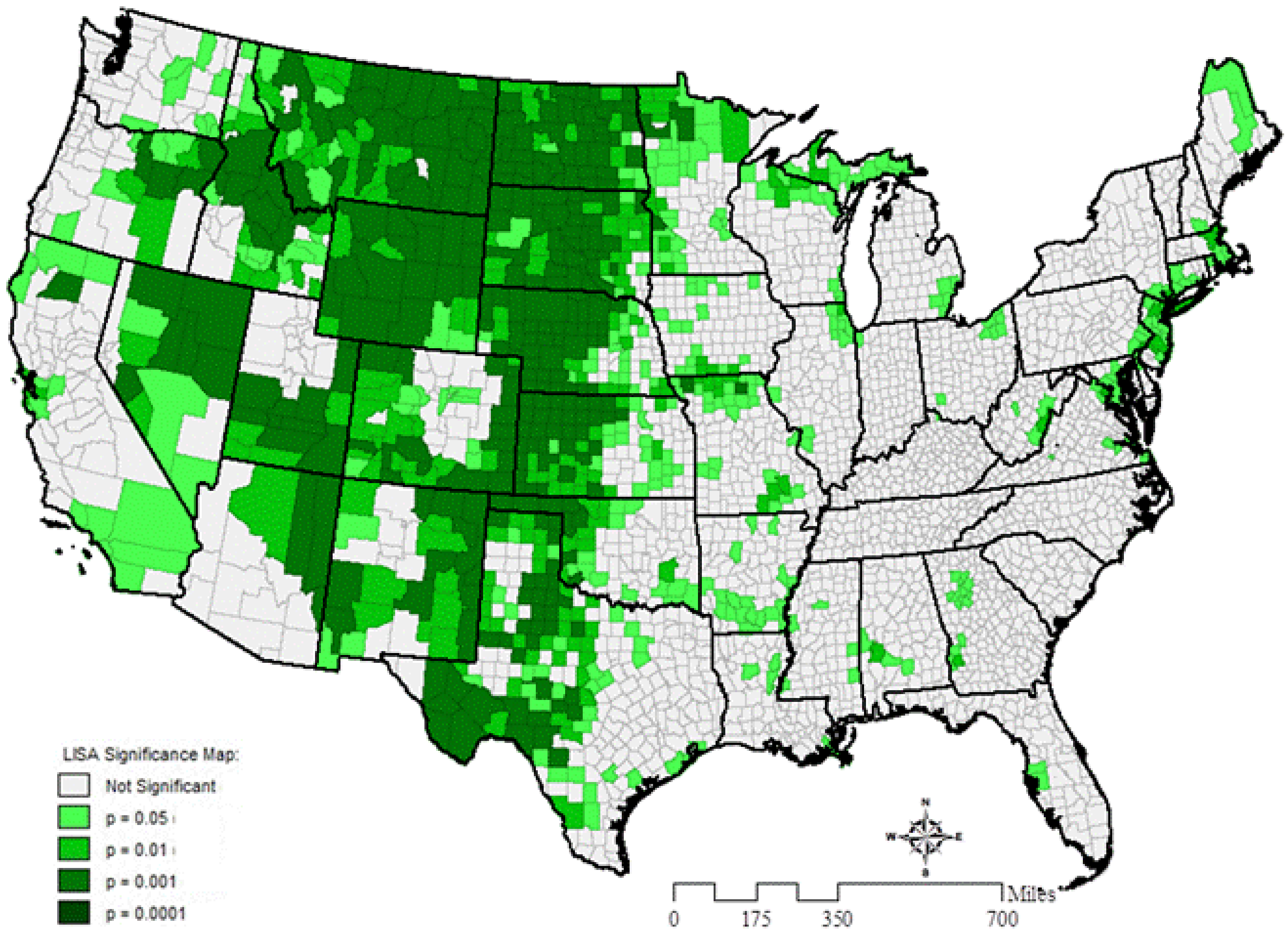

Figure B.14 Bivariate LISA significance map for county level obesity prevalence rates and population densities in the US for 2007

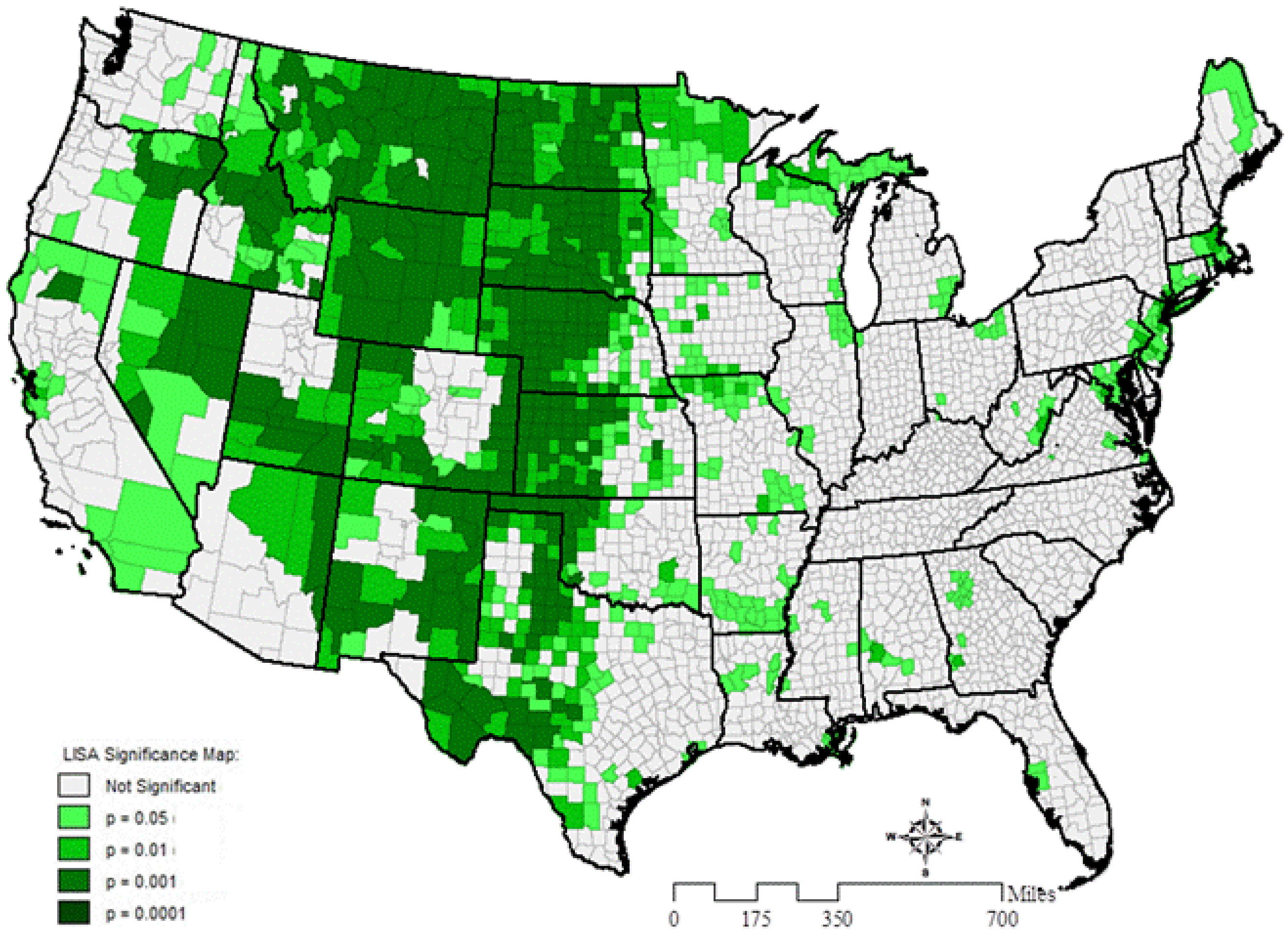

Figure B.15 Bivariate LISA significance map for county level obesity prevalence rates and population densities in the US for 2010

VITA

Mustafa Erdem was born in Isparta, Turkey in 1977. He graduated from Yildiz Technical University, Istanbul, Turkey with a Bachelor of Arts degree in City and Regional Planning. During his undergraduate study, he received a number of scholarships from the Turkish Government for his successful undergraduate studies. Immediately after his graduation, he moved to the United States of America and enrolled at Louisiana State University to pursue his Master's degree in Landscape Architecture. His exceptional contributions to a project for reconstruction of New Orleans after the Hurricane Katrina received acknowledgement from the Center for Academic Success at Louisiana State University. He was also the recipient of Helen Rich Memorial Scholarship for his outstanding graduate studies in the Robert Reich School of Landscape Architecture at Louisiana State University. He continued his academic career in the Department of Geography and Anthropology to pursue his Ph.D. degree under the academic guidance of Professor Michael Leitner. During his Ph.D. studies he worked in the Louisiana Department of Transportation and Development and Highway Safety Research Group at Louisiana State University as a research assistant. His research interests include public health, obesity, spatial and statistical analysis, GIS, health disparities in urban areas, and traffic crash analysis. He is a candidate to receive his Doctor of Philosophy degree in Geography on May 12th, 2016.