SMART GROWTH AND WALKABILITY

AFFECT ON VEHICLE USE AND OWNERSHIP

By

Derek Richard Newland

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

May 2015

DEDICATION

I dedicate this document to my parents Douglas and Susan Newland, my girlfriend Anna Simpson and my employer Jim Minnick for their constant support and understanding, and to my committee chair Robert Vos, who has helped me through this project in understanding the concepts and theories behind the research and for his support in completing this project.

ACKNOWLEDGEMENTS

I will be forever grateful to my mentor, Professor Robert Vos. Thank you for your patience and understanding. Tenacity and persistence are qualities I have come to appreciate through this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study tests the effects of the built environment on vehicle miles traveled (VMTs) and automobile ownership, with specific reference to aspects of neighborhood walkability studies and research design at nested spatial scales of metropolitan regions and neighborhoods. This adds to existing smart growth studies as they tend to focus on Census data and non-walkability land use variables such as rail transit infrastructure. This study looks at 75 census block group samples within 5 metropolitan statistical areas (MSAs). The variables measured for the study include, bus stops per square mile, jobs within 45 minute transit ride, gross activity density, temperature, distance to retail, distance to transit, and slope among others. This study also looks at including regionally measured variables such as people per transit station, MSA density, and transit expenditure in conjunction with neighborhood scaled variables in order to test if there are any interactions between neighborhood and regional variables. The variables are entered into a multivariate regression model to find the best-fit model in order to explain the relationships between the dependent and independent variables. The study finds that the new walkability variables added to the research add significantly to the explanatory value of regression models beyond studies that use just smart growth land use variables. The implications for this study are that there is ground work laid for a new type of smart growth and walkability joint study at a multiple region level.

CHAPTER 1: INTRODUCTION

The purpose of this study is to examine how smart growth strategies and walkability of the built environment impacts vehicle behavior and automobile ownership. Vehicle behavior will be measured by vehicle miles traveled (VMTs). While this is a dual study of smart growth and walkability the primary emphasis will be on the walkability variables. The smart growth portion of the study, mainly the variables and a similar analysis method are built in part from the initial work in this area conducted by Eisman (2012).

Earlier smart growth studies, like Eisman (2012), measure the built environment by population density, job access, connections to transit networks, and street connectivity. A high level of street connectivity, measured as intersection nodes, is thought to be a proxy for walkability of neighborhoods. Other studies look at rail variables such as passenger miles per capita and mass transit infrastructure density (Cervero and Murakami 2009). Also some studies look at jobs-housing balance along with other land use variables (Cervero and Duncan 2006). Although the underlying sampling framework and initial data sets are built from Eisman (2012), this study will both develop some new independent variable datasets and add to the existing set of independent variables by utilizing the basic smart growth variables found with the literature to measure the mix of economic uses and distance to jobs.

The walkability aspect of this study will take a different approach than most walkability studies. This study is a multiple region study utilizing random samples of census block groups from five Metropolitan Statistical Areas (MSA). Because of the scale of the study area, a typical walkability study is not suitable due to the type of variables used in such a study. The issue at hand is scale. Most walkability studies utilize variables found at medium to small scales: county, city, and most often neighborhood or street. Datasets are independently created for various regions. Because of this the data involved in a walkability study can be very detailed. Variables for a walkability study often include, tree canopy, side walk quality, American with Disability Act codes such as curb ramps, public safety data, street illumination and street width (Spoon 2005). These variables are difficult and time consuming to collect at neighborhood scale samples within a multiple region study. This is because a region not only contains many neighborhoods but also many different city and county jurisdictions which may or may not maintain such data or have GIS/mapping programs at all. However, there are walkability variables that can be constructed from national datasets to be utilized in a multiple region walkability study. Some of these variables are neighborhood slope (Villanueva et al, 2013), mass transit access, climate, and mixed land use (Spoon, 2005).

This study will conduct a multiple region smart growth study with an emphasis on regionally favorable walkability aspects. This is something rarely seen in the literature with multiple region level study areas, especially in so far as studies are able to utilize detailed variables at the street level (see Appendix A for MSAs, their locations and sample locations within the MSAs).

1.1 Motivation

The method of how we travel in the future is already becoming an issue with fossil fuel usage ever increasing and long term threats to fuel supplies. The impact of fossil fuel consumption from automobile transport on global warming and its effect on our lives is also a growing concern. Large cities have always been in the news for traffic jams and general everyday congestion. Highway construction can even lead to near Hollywood style traffic jams such as the 2011 “Carmageddon” in Los Angeles, CA (Reuters, 2011) or the 9 day 60 mile traffic jam near

Beijing in 2010 (Reuters, 2010). These concerns and events show that there is a need for effectively evaluating how we design urban areas and even entire regions.

As well as global warming from vehicle emissions, there is also the issue of pollution. Health issues from smog can be severe. Smog (lower atmosphere ozone) can cause several respiratory issues, such as aggravating asthma, reduced lung function, and irritate the respiratory system (EPA 1999). Along with regulation on vehicle emissions, reducing people's need to use their vehicle by smart urban design that promotes walking or the use of public transportation can reduce the emission of chemicals that form smog and reduce the health hazard attributed to vehicle as well as industrial causes of smog.

While alternative fuel for transportation holds much promise from hydrogen to biofuels (Gajendra Babu and Subramanian, 2013), there also needs to be emphasis and effort from all levels of government in designing urban and suburban environments that contribute to better walkable options and efficient suitable transit networks to encourage lower usage of personal vehicles. In 2008 California enacted California Senate Bill 375, which creates a formal process for local planning officials to develop land use plans that reduce vehicle miles travelled (VMTs and thus reduce greenhouse gas emissions from automobile travel (CEPA 2014). Unfortunately there do not appear to be any penalties for not adopting CA SB 375 and it is merely an incentive plan to encourage local governments and developers to adopt its policies to meet California's greenhouse gas emission goals (CEPA 2014).

1.2 Study Overview

This study seeks to aid in the evaluation of land use policies by identifying how the built environment effects automobile transit and apply that to new development or re-development of city centers through statistical analysis of human and geographic variables. The study employs multivariate linear regression to accomplish this. The intent is to use a multiple region study to identify how the different smart growth and walkability variables affect VMT and automobile ownership so that comparisons can be made for different regions and land use policies.

One of the study regions for this project is the Portland, Oregon metropolitan statistical area (MSA). This region is particularly known for its smart growth policies. Jun (2008) studied whether their policies were related to reduced automobile dependence and found mixed results. However, areas with higher accessibility to light rail and bus services and more mixed land used were associated with a higher probability of alternative modes of transportation to personal vehicles or driving alone. However, areas with higher employment and residential densities alone did not see a reduction in driving. The effect of accessibility of light rail and bus routes with high mixed land use appears to show the positive effects of smart growth and walkability polices. The inclusion of the Portland MSA then, should help within the study when utilizing other MSAs, several with greater populations, from different regions of the country.

The assumption that various changes in land use policies and land use itself can raise or lower VMTs deserves robust study. The analysis and method used within this study adds to the existing research by implementing a multiple region smart growth and walkability evaluation. It is important to consider variables both at the neighborhood and regional scales, as well as potential interaction between variables at these two scales. For example, is a densely built neighborhood in a region with high mass transit network density more likely to have lower VMT's or automobile ownership than a similar neighborhood situated in a region with low mass transit network density? This is a detailed version of the overall research question that motivates this study: What elements of urban built environment at the neighborhood and regional scales are associated with reductions in VMTs or automobile ownership?

Following from this research question, this study proposes and tests a set of 14 hypotheses. Of these, 6 of the hypotheses rely on newly created independent variables. The study makes extensive use of geospatial datasets and ArcGIS to develop these variables. The hypotheses, variable definitions and construction are carefully described in the Methods Chapter.

In the analysis, many of these hypotheses are initially rejected as bivariate correlations and some are further rejected in considering the total regression model. In its Results Chapter, the study identifies key differences between the regions and tries to explain these differences by testing for interactions of regional scale variables with neighborhood scale variables. Two complete regression models are presented and need for future work is identified.

CHAPTER 2: BACKGROUND

Interest in research on smart growth approaches to urban development has increased along with concerns over rising fuel cost, decreasing oil reserves, and climate change. Smart growth aims in large part to locate and design neighborhoods in ways that decrease vehicle miles traveled and potentially also decrease automobile ownership (Cervero and Murakami 2009). Interest in walkability of neighborhoods has also grown in parallel to health concerns of obesity, diabetes, and other conditions thought to be due in part to a lack of exercise (Miranda et al. 2012). Walkability research aims to understand the built environment or neighborhood factors that may encourage people to exercise.

The research done in this area is generally separated between studies focused on smart growth versus walkability. The two types of studies are rarely integrated in a comprehensive way, especially in terms of explanatory variables. Smart growth studies tend to look at transit, population density, employment, road networks and housing density (Eisman, 2012, Cervero and Murakami, 2009). Alternatively, walkability studies use a wide range of variables, some similar or the same as smart growth but many are very different such as, tree cover, sidewalk conditions, climate, terrain (slope), and street width to name a few. Walkability studies may also include variables related to individual human behavior such as habits or family circumstances (Spoon 2005).

Aside from contrasting research variables, the theme of scale is also an important point of difference between the smart growth and walkability literatures. Smart growth studies tend to have small to medium sized study areas consisting of cities or urban areas, with neighborhoods or census block groups as the focus point for data. Few studies address sample or study areas at the regional level. Walkability studies typically use even more focused study areas than smart growth research. Most if not all walkability studies deal with study areas at very local scales such as streets, neighborhoods and cities, there was no literature found during research for this study to indicate studies at larger scales.

Within this study, smart growth and walkability variables effect on VMT and automobile ownership are analyzed together. Walkability variables will be analyzed along with smart growth variables at a large regional scale utilizing Metropolitan Statistical Areas which in some cases may consist of over 20 counties. Five different MSAs across the country will be included with samples consisting of neighborhood scale samples measured as census block groups. The variables chosen for measurements of smart growth consist of nearly all variables typically used within the literature while measurements of walkability will be chosen based on aspects that are conducive to the size of the sample (n=75) along with scale of the study areas. Variables that require extensive fieldwork to gather data or hand digitization of image files, such as side walk conditions and tree canopy cover are not conducive to the scale of this study.

This chapter reviews the literature on how smart growth affects VMTs and automobile ownership, placing an emphasis on the scales used for each. Next, the chapter reviews walkability studies focusing on walkability measures of communities, cities, and counties. In general walkability studies do not appear to deal with walkability's effect on VMTs or automobile ownership. One key idea tested in this study is that neighborhoods that are more walkable might also result in fewer VMTs and/or reduced ownership of automobiles.

2.1 Influence of Smart Growth non VMTs and Automobile Ownership

The majority of VMT and automobile ownership studies tend to focus on small scale study areas such as neighborhoods, cities, counties, block groups or Census tracts. Few if any of these studies look at very large study areas such as an MSA (Metropolitan Statistical Area), include multiple metropolitan regions, or utilize walkability variables along with smart growth variables to determine the effects that the combination of these variables may have on VMTs and automobile ownership. That being said, there are two studies that address multiple regions consisting of block groups within an MSA. However, neither includes a combined look at the influence of smart growth and walkability effect on VMTs or automobile ownership

In his research on smart growth, Daniel Eisman (2012) looked at how smart growth within the built urban environment affected automobile ownership and vehicle miles traveled (VMTs). His method of analysis involved examining several land use and statistical variables at the neighborhood level. The study involved 75 samples that consisted of census block groups as neighborhoods and were located within 5 Metropolitan Statistical Areas (MSAs) covering 5 separate regions of the United States. The study evaluated distances to jobs, retail and transit as well as population density, transit networks income and several other population statistics. Eisman found based on his series of best fit regression models that there was a statistically significant association of VMTs and automobile ownership with the built urban environment. Eisman's study was the starting point for the research performed in this study (Eisman 2012).

Robert Cervero and Jin Murakami (2009) have looked at effects on VMTs from the built environment by evaluating 370 urbanized areas around the US. Urbanized areas within the study consisted of individual and connected cities (such as the Los Angeles, California area). These urbanized areas vary in size and the variables for the study aggregate metrics for the entire study area (i.e., region). The variables used within the study were VMTs, rail variables such as passenger miles per capita and infrastructure density, population and employment densities, income, and areal size variables. The variables were put into a path dependent model (structural equation model (SEM)) to test their effects on VMTs and it was found that population density has the strongest association with VMTs at a direct coefficient of -0.604, indirect coefficient of 0.233 and a total coefficient of -0.381. The total coefficient results from the overall model, allowing for other variables in the path. This is the result that the authors report as their major finding. Some of their other results included automobile commute shares at a total coefficient of .602, and road density at 0.415. Employment, size of urbanized area, and rail-transit were found to have less effect on VMTs (Cervero and Murakami 2009). This study was one of the few articles in the literature that looks across multiple regions. However, unlike this study, the authors use only regional data or regionally defined variables, computed as averages of neighborhood scale analysis. Additionally, unlike this study they do not test for the interaction between the neighborhood and regional scales.

In a separate study Cervero and Duncan (2006) explore whether retail-housing mixing (mixed land use) or jobs-housing balance have the greatest effect on VMTs and VHTs (vehicle hours traveled). Jobs-housing balance is a form of land use planning that attempts to bring jobs and residents into closer proximity in any given community. For example, one policy sometimes used is to offer grants to workers who purchase residences close to their job. Utilizing regression models for their variables they found that jobs-housing balance had a greater effect on VMTs than mixed land use (Cervero and Duncan 2006). However for jobs-housing balance to be effective, it requires incentives to employees and meaningful efforts by local jurisdictions to support it by rezoning and finding funding for these types incentives.

While most studies look at land use effect on VMTs and ownership, not all studies use land use as the variable of interest. Kim and Brownstone (2010) studied the impact of residential density on VMTs as well as fuel consumption. The scale of the study is at a neighborhood and midsized city or urban scale in that the data was sampled and measured utilizing census tracts and block groups but the regions to study were chosen at the regional scale utilizing MSAs. The housing density variables were taken from census data and covered most available population and household variables available. These variables were population per square mile by block group and tract level and housing units per square mile. It is not stated where the tract level variables are considered regional variables but typically variables at the tract level are not regional. The variables were imputed into a simultaneous equation model in order to calculate the effects on VMTs and fuel usage.

The results of the research were that there was a statistically significant effect of land use and population density on VMTs and fuel usage. The main result was that more densely populated areas had few VMTs and less fuel usage. Results for the study were broken up by variable as well based on the types of variables. Socio-demographic variables results were broken down into individual results. A number of drivers had a strong influence on household vehicle behavior. The number or workers had a nonlinear effect on annual mileage and fuel usage. Income, which is a variable within this study as well, was found to be statistically significant and fuel usage increased linearly with income. Number of children was overall statistically insignificant. “Life cycle effects” which consisted of three variables, retired households, single person households, and non-single person households, had different statistical significance for each variable. Retired households had a negative direct effect on mileage traveled, while single person households were statistically significant for household density, fuel consumption and annual mileage (Kim and Brownstone 2010).

2.2 Walkability studies and vehicle behavior

This study's main focus is aspects of walkability effect on VMTs and automobile ownership in combination with smart growth and transit variables. There is a large literature on measurement and indexing of walkability. However, most walkability studies focus on county, city, neighborhood, or street study areas within a specific region. Also, the literature does not combine walkability metrics with the more commonly investigated transit and smart growth variables. The goal here, as identified above, is to use nested neighborhood and regional scales of multiple regional (MSA) study areas to better understand the role of walkability in influencing vehicle behavior.

Stephanie Chen's (2012) study on bus route walkability along two Orange County bus routes is scaled at the level of a single county, focusing on the neighborhoods along the lengths of existing bus routes. The variables Chen chose for her study were population density, street connectivity, steepness, and tree canopy, along with 3 different buffer types: half-mile radii, route-adjacent, and stop-and-route-adjacent. The buffers were used to calculate a walkability score and then compared to the scores for stops along each of two bus routes. She found that in general, walking paths to bus stops routed along grid street neighborhoods were more walkable than cul-de-sac neighborhoods (Chen 2012).

Another type of multiple location walkability assessment, was performed by Robert Stevens who looked at determining the effectiveness of a walkability assessment type to assess the walkability of four local area parks in Springfield, Oregon (Stevens 2005). The study area scale for the study is the neighborhood scale; however the scale of measurement was at the street level specifically street segment. The data collected for use in the study was TIGER file street data from the US Census and then field data collected by Stevens himself.

The appeal of this study, in part due to its study scale, is that the data can be collected locally and by the researcher themselves. The data was collected using ArcPad software on a PDA and utilizing a data collection method call the Pedestrian Environment Data Scan (PEDS) (Stevens 2005). The PEDS assessment tool contained 77 walkability indicators but Stevens chose 20 of those indicators that were found to be the most important in assessing walkability. Being a street level study, these indicators/identifiers can be very specific such as: attractiveness for walking, safety, traffic volume, sidewalk condition, land uses, number of traffic lanes, presence of building setbacks, crossing aids, tree count along streets, and if there are parking lots that can be walked through. Each street segment being assessed was given a score based on presence of these indicators. The analysis of the data was done using ArcGIS with the sampled street segments being aggregated to census blocks to create easy to read polygons. Next, catchments, defined as the ½ mile area where streets are available for pedestrians to use were created utilizing multiple centroids in the polygons to indicate park entrances (Stevens, 2005). The hostile streets were then removed as not being suitable for walking. The results of this data acquisition and creation were then compared against TIGER and LCOG street classifications (Stevens, 2005).

The findings from Stevens (2005) determined which parks had better walkability amongst the four tested. The purpose of the study was to test the method of walkability assessment to aid in future development of Springfield, Oregon as well as determine the parks with the best walkability. It is different than this study which seeks assess aspects of walkability of a region. Still, Stevens' research gives good insight into the types of variables used in determining walkability and the study scales required to utilize each type of variable. There is a tension between the scale of the study area and level of detail that can be gathered in terms of walkability variables.

Anupama Mantri (2008) studied a GIS approach to measuring walkability within a neighborhood and applied the variables within his GIS model to parcel level data. The measures of walkability used by Mantri were connectivity (road network), proximity (access/proximity to activities), density (residential density), land use mix, and safety measures (Mantri 2008). His approach and scale of his study area allowed for in depth analysis of the neighborhood including location information for many different activities and destinations specific to the neighborhood such as theatres, restaurants, pharmacies, churches etc. This could be considered an ideal method of studying walkability due to the amount of data available to a small study area such as a neighborhood rather than analyzing larger study areas such counties or regions.

While researching walkability studies, it became apparent that most walkability studies deal with the cities, neighborhoods, or smaller landscapes. However there are studies that deal with these small study areas but also link several of them in a region or nationally. Horacek et al. (2012) studied the walkability and bikeability of US postsecondary educational institutions, which could be considered neighborhood study areas depending on the size of the campus. The variables used for studying walkability and bikeability and which were scored on a set of standards, were sets of criteria for “safety, path quality and path temperature comfort” (Horacek et al. 2012, 10). Safety criterion were variables such as crosswalk quality, night time safety such as lighting and side walk existence and quality. Path quality was made up of variables such as path size, buffer zone from road ways, and terrain (elevation change, slope etc.) among others. Path temperature constituted whether there was shade or not (Horacek et al. 2012). These types of variables are difficult to measure and obtain for larger study areas making it difficult to determine the walkability of multiple neighborhoods in multiple counties and in multiple regions.

There does not seem to be a single group of walkability variables though many seem constant in several studies. In order to determine what makes up walkability measurements and what variables may be used in measuring walkability, Steven Spoon (2005) researched the literature to create a relatively comprehensive list of what is considered important in defining and measuring walkability as well as the variables that are most prevalent amongst the different studies. The variables that are found in several different literature reviews and are to be looked at in this paper are density, access to transit, and mixed land use. These are the variables that make the most sense and are most accessible for the scale and diversity of the study areas of this study.

The variables that were chosen for this study were taken from the list created from Spoon's research of walkability studies. These variables were chosen based predominately on the study regions and scale of the samples along with the availability of data due to the scales. The first variable was elevation on the prediction that flatter land is easier to walk on. Temperature and precipitation was chosen next on the prediction that milder climates would encourage more walking. The next variable, bus stops per square mile, was chosen as a prediction that many bus stops make it easier to walk to access public transit networks. The next chapter will discuss the reasons and methods of acquisition of the data for these variables.

While there are many studies and assessments of walkability within the literature, there are also websites that allow the general public to find the walking score for their streets or neighborhoods by simply entering an address. Walkscore.com is a website that will calculate walk score utilizing a web application designed to calculate walk score based on measurements of distance from a chosen location to surrounding amenities such as super markets, restaurants, entertainment venues, and parks as well distance to transit stops. There are other data types used in the calculations such as road network and route directness. Since the data is calculated on the fly through a web app the data used in the application would only be applicable as individual data types and thus the walk score would not be incorporable into the data sheet for the regression models used in this study (Walk Score 2014).

After reviewing the literature for smart growth and walkability there are several consistencies within each topic. Smart growth research uses study areas at many different scales from regional to neighborhood and has a large number of census based variables such as population density, income, employment, etc. Geographic variables within the studies are all similar as well, utilizing land use and transit variables in most cases. Walkability studies use smaller scales with the largest being at the city or county scale. The variables within these studies cover a very wide range detail and scales. The most detailed study areas appear to be small scaled studies at the street level that in some cases can only be collected in the field by the researcher themselves. This study looks to complete a multiple region smart growth study with regional and neighborhood scaled variables at neighborhood scaled samples. This study will add to the literature by adding walkability variables that can be measured at the neighborhood scale but can be acquired from multiple regions with some ease as variables such as sidewalk condition are neither feasible nor cost effective for a study with region scaled study areas.

CHAPTER 3: METHODOLOGY

The purpose of this study is to look at how smart growth, along with aspects of walkability, may impact vehicle miles traveled (VMTs) and automobile ownership at the neighborhood scale. The locations for this study were five (5) metropolitan statistical areas chosen to represent regions of the United States with different development histories and resulting urban forms. The samples are neighborhood level and represented as census block groups as the method for applying census data and acquiring and processing non census data.

The sampling framework and initial model of Eisman (2012) is a foundation for this study. But this study develops new variables related to walkability, applies variables from EPA's Smart Location Database (EPA, 2013), and deepens analysis of regional effects. As stated previously, the majority of walkability studies tend to only consider variables measurable at county, city, neighborhood, or street scales and rarely compare or analyze multiple regions or locations. Utilizing variables used within the literature and obtainable by geoprocessing for multiple regions and jurisdictions, it is hypothesized that the predictive power of existing regression models for vehicle miles traveled (VMTs) and automobile ownership at the neighborhood scale may be improved.

3.1: Sampling Framework

This study adopts the sample of 75 Census block groups developed for Eisman's (2012) study. Based on Eisman's logic, block groups make sense because they are essentially predetermined neighborhoods consisting of clusters of several city blocks with the exception of more rural areas where a block group can be many square miles. It is important to note that a few of these larger block groups are included in the random sampling in some of the five Metropolitan Statistical

Areas (MSAs) that make up the study regions when the samples include block groups in the outskirts of the urban area.

Census block groups were selected by Eisman because they are the best scale at which to get detailed information for key variables due to their small size (Eisman 2012) and they are the closest in scale to what could be considered a neighborhood. The new independent variables developed for this study, however, do not use census data, and while spatially matched to the block groups, they are not directly reliant on the block group as the spatial unit of measurement. Still, the unit of analysis remains the census block group because several variables gathered by Eisman (2012) and tracked in EPA's Smart Location Database are only available at the block group level. Using a different scale for this study could negate the attempt to apply both smart growth and walkability into a single examination of transportation behavior by possible eliminating the smart growth variables.

The MSAs chosen for the study were Chicago, IL; Miami, FL; Portland, OR; San Diego, CA; and Washington DC. These areas were chosen because they represent different regions around the United States: the Midwest, South East, North West Coast, South West Coast and the East Coast. MSAs were chosen as the regional unit of measure because they are defined in part by transportation needs including commuting across the area. MSAs consist of 50,000 people or more and contain the counties that consist of the core urban area and adjacent counties that have a high degree of social and economic integration with the urban core. This integration is measured by the commute to work between adjacent areas and the urban core (US Census, 2014)

Eisman (2012) selected 15 sample block groups from each of the five MSA's. In general, the samples were selected using a process known as stratified random sampling (Eisman, 2012). Within the selected MSAs, a "random" field within an Excel spreadsheet was assigned to all the block groups. Within the added “random” field, a random number generator was used to place random values between 0 and 1 in the column. The spreadsheet was then ordered based on the “random” field and the lowest thirteen random numbers were chosen. The fourteenth and fifteenth sample block groups of each sample region was determined by taking the most and least dense block groups within each of a given MSA. Thus, it is important to note that the sample is not a pure stratified random sample, but includes two outliers for each region with regards to neighborhood population density.

3.2: Hypothesis and Independent Variables

Eisman hypothesized that the built environment would influence the number of automobiles owned per household as well as the vehicle miles traveled (VMTs) and that automobile ownership will be greater in sprawl neighborhoods (Eisman 2012). The same holds true with this study with the added hypothesis that a built environment conducive to walkability will affect the number of automobiles owned as well as VMTs. It is predicted that areas with greater walkability will have reduced automobile ownership and vehicle miles traveled. See Table 1 for research variables and their hypotheses.

The variables used in this study to determine walkability are; neighborhood and street slope, number of accessible bus stops per square mile, climate, and land use. There have been various studies on what defines walkability and the variables used to measure it. The variables for this study were decided upon based in part on Spoon (2005) in which several pieces of literature were analyzed and their variables assembled. Many of the variables used to measure walkability are available at only small analytical scales. The included variables such as sidewalk quality, tree cover, or walkway availability. These types of variables are not conducive to a regional study such as this which compares several study areas across several different MSAs.

While MSAs such as San Diego CA, consist of only one county, the others can consist of nearly twenty counties which make local data acquisition difficult, if not impossible.

For this reason, variables following from Spoon (2005) that can be determined from national datasets or a single provider such as Google or the USGS are ideal as they provide consistent or as close to consistent data across the United States. Such variables include bus stops per square mile, temperature and precipitation raster datasets and Data Elevation Model (DEM) raster datasets. These variables can be acquired for the entire nation and processed down to the neighborhood scale for analysis.

Table 1 Hypotheses of Variables

Dependent Variables	Hypothesis
Automobile Ownership	Land use and walkability aspects will influence automobile ownership.
Vehicle Miles Traveled	VMTs will be reduced by land use and walkability aspects.
Independent Variables	**Hypothesis**
Neighborhood Slope	Will have a direct relationship with VMTs and Automobile ownership.
Bus Stops per sq/mile	Will have an inverse relationship with VMTs and Automobile ownership.
Minimum Temperature	Will have an inverse relationship with VMTs and Automobile ownership.
Maximum Temperature	Will have an inverse relationship with VMTs and Automobile ownership.
Minimum Temperature	Will have an inverse relationship with VMTs and Automobile ownership.
Precipitation	Will have an inverse relationship with VMTs and Automobile ownership.
Jobs Within 45 min Transit Ride	Will have an inverse relationship with VMTs and Automobile ownership.
Jobs Per Household	Will have an inverse relationship with VMTs and Automobile ownership.
Land Use Diversity (Mixed land Use)	Will have an inverse relationship with VMTs and Automobile ownership.
Density	Will have an inverse relationship with VMTs and Automobile ownership.
Distance to Job Centers	Will have a direct relationship with VMTs and Automobile ownership.
Distance to Retail Centers	Will have a direct relationship with VMTs and Automobile ownership.
Distance to Transit	Will have a direct relationship with VMTs and Automobile ownership.
Transit Expenditure	Will have an inverse relationship with VMTs and Automobile ownership.
People Per Transit Station	Will have an inverse relationship with VMTs and Automobile ownership.
Regional Density (MSA Density)	Will have a direct relationship with VMTs and Automobile ownership.
Confounding Variables	**Hypothesis**
Income	Will have a direct relationship with VMTs and Automobile ownership.

3.3: Neighborhood Slope and Walkability

Slope was chosen and developed as a variable in this study because it is hypothesized that areas that are flatter will be more appealing and conducive to walking as a mode of transportation and cost saving method. If it takes less physical effort for someone to walk back and forth from work to home or to the shops it may be more likely for people to do so and perhaps to drive less as a result. While flatter land maybe best for general walking commutes it could be suggested that for recreation and health, steeper slopes may be appealing as well (Villanueva et al. 2013). However, this study is focusing on walkability as a means of everyday transportation and not primarily as a method of exercise.

This study measures slope both for the overall neighborhood and for the street network. Both slopes were measured because: a) sidewalks are built along streets and thus should have the same slope and streets may be graded for more gradual inclines than the surrounding areas, and b) walkable areas do not occur only along sidewalks. There may be parks and walking paths between buildings that can be used as shortcuts through city blocks that allow for more efficient travel by walking. By measuring both slopes, differences in slope of possible walking space can be measured.

Neighborhood and street slope were derived from USGS Digital Elevation Model (DEM) 10 meter raster files and were chosen as variables for walkability based on the prediction that neighborhoods with a greater slope percentage would not be as walkable as a flatter area. Neighborhood slope (slope without water) is the percent slope of the entire block group with a quarter mile buffer in order to ensure measurement of rasters immediately adjacent to the boundaries of the neighborhood. This buffer also helps to measure the percent slope of the roads which in many cases do not fall inside the block group for the reasons stated above. Street slope was calculated by extracting raster values from the neighborhood slope raster using a street network data. Street slope was chosen along with neighborhood slope in order to cover areas that maybe more level due to the leveling and or raising of road networks during construction.

Figure 1 San Diego, CA MSA DEM Raster Data by Sample Block Groups

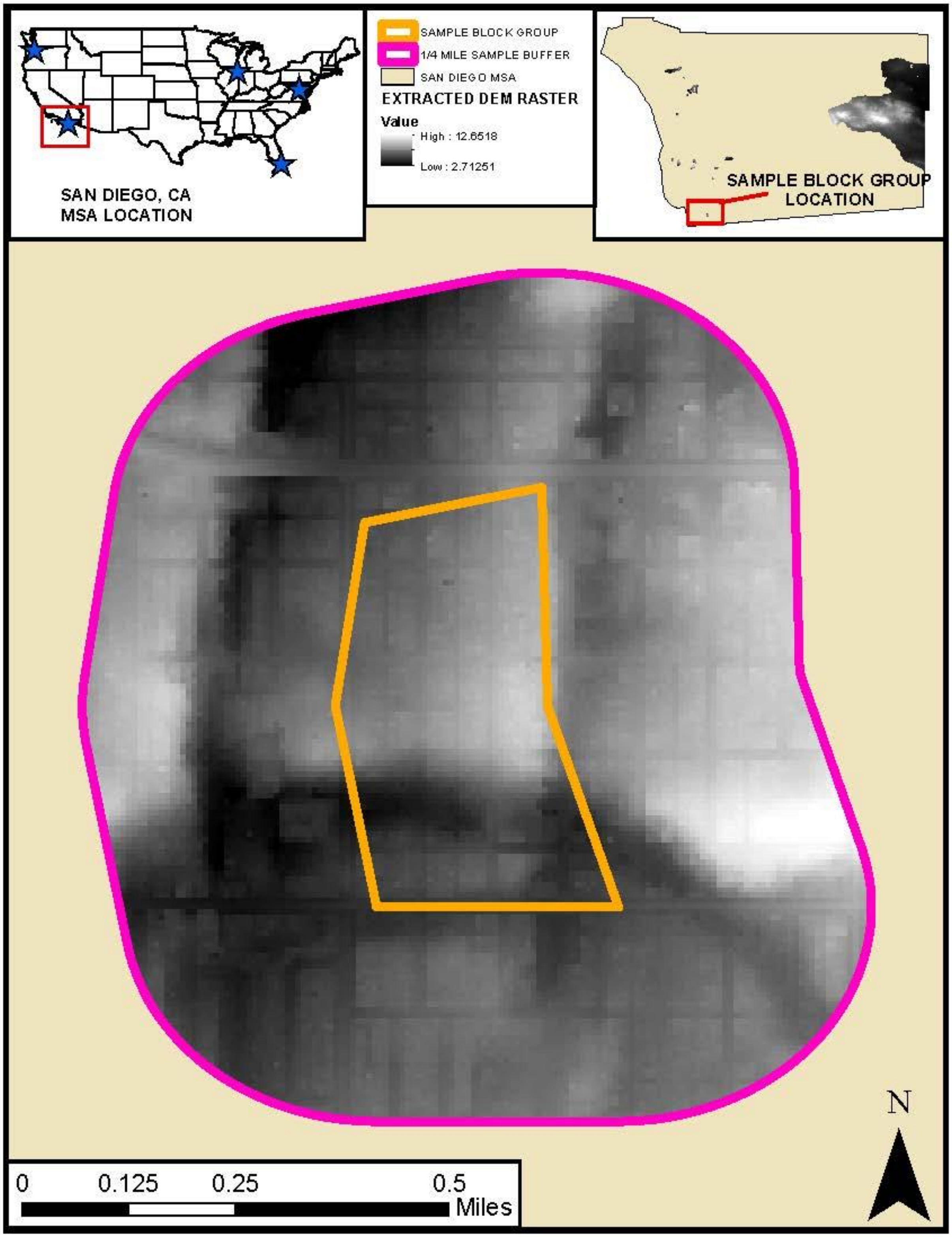
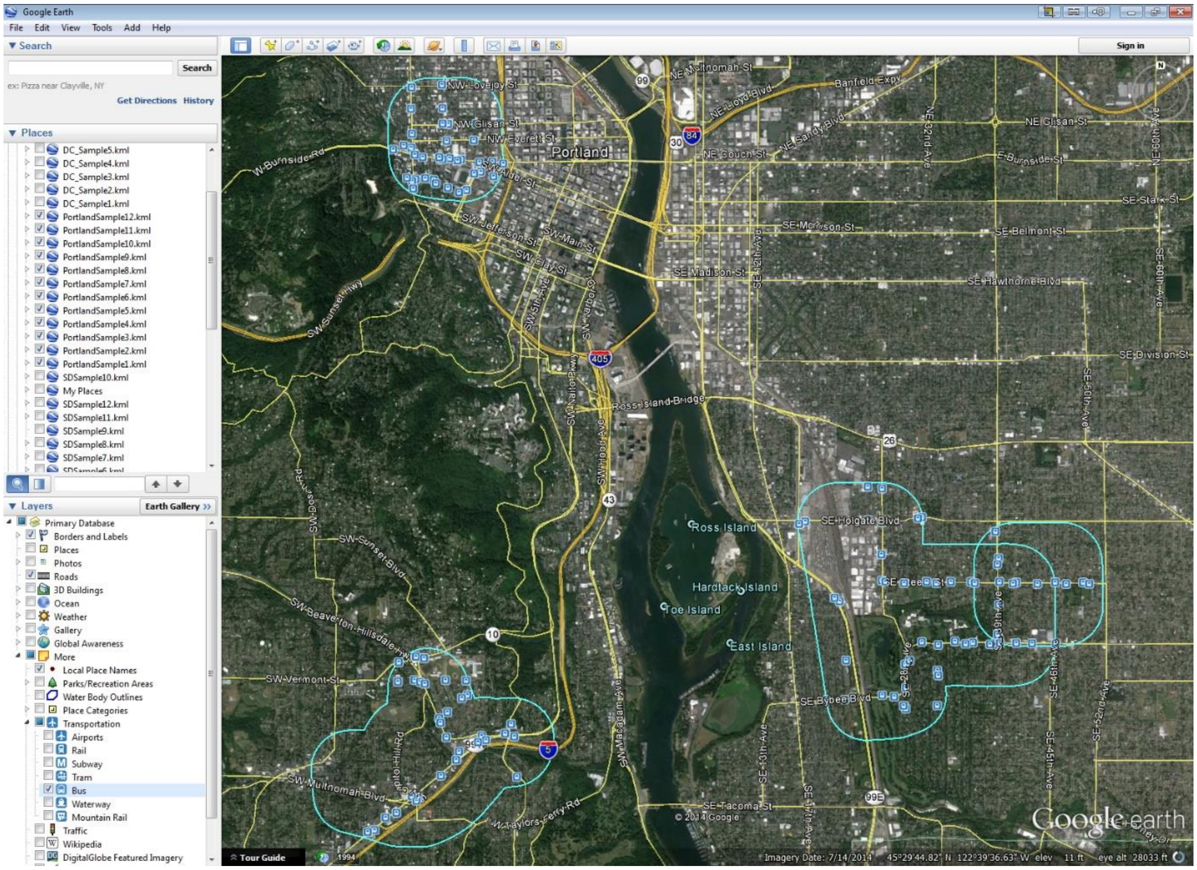


Figure 2 San Diego, CA MSA Slope Without Water by Sample Block Groups

3.4: Proximity of Bus Stops

Bus stops per square mile was the next variable looked at to define walkability. The hypothesis is that if there are a large number of available bus stops in a neighborhood, VMTs and automobile ownership will be low because people can easily walk to a bus stop. Previous studies have focused on rail access because bus stop data have been hard to develop across multiple regions. However, this study hypothesizes that bus stops may be particularly important because in many U.S. cities bus transportations is the only mass transit network covering most of the region. Also, even in cities with highly developed rail networks, bus networks are critical for access on the first and last mile of trips using the rail network. For this study, bus stop data was acquired through Google Earth data as it was the best source of bus stop data available nationwide and was verifiable through the use of their “street view” function to see if stops are actually present (Google Earth, 2013). The number of stops was counted and normalized by the areal extent of the block group.

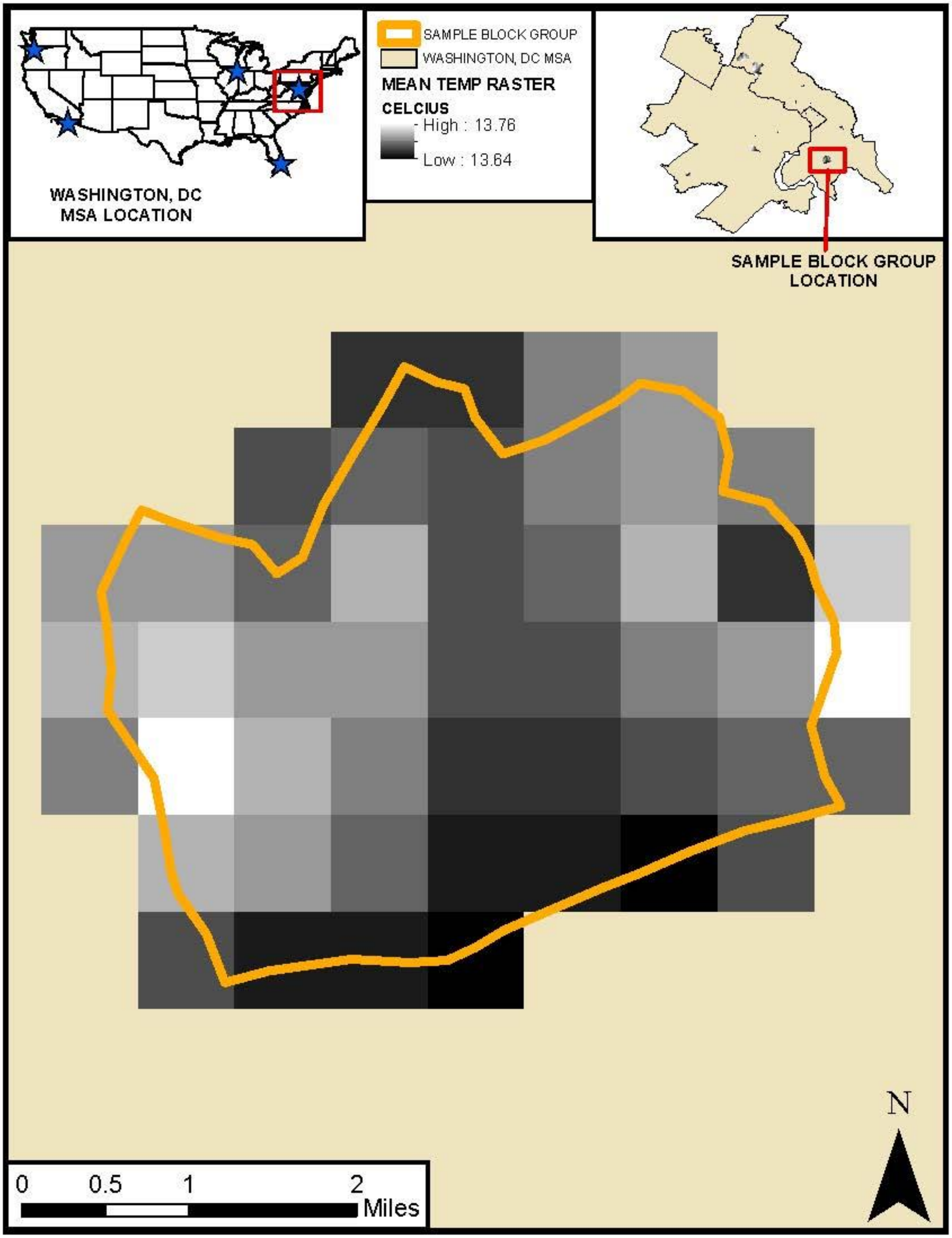
Figure 3 Example of KML Acquisition Within Google Earth for Four Sample Block Groups in the Portland, OR MSA



3.5: Climate and Walkability

The third variable to determine walkability at a regional scale is climate. The prediction is that moderate climates encourage walking. Areas with a cooler summers as well as warmer winters are presumed to be nicer to walk in, thus encouraging walking and reducing use of automobiles. The data used to determine temperature were thirty year normal maximum temperature, thirty year normal minimum temperature and thirty year mean temperature. The thirty year mean temperature was chosen as a variable along with minimum and maximum temperature to cover for the potential of abnormal minimum or maximum temperatures that could affect the analysis, such as 1 year out of 30 where an area may experience an abnormally cold winter. The variables were based on the averages for their respective measurements over 30 years. The reason for the different variables for climate is to simply cover the range of possible temperature variables. Areas may have extreme cold or heat for only a month or two, but the mean temperature may be mild. Thus by modeling all three temperature variables there is a better chance of determining if climate has any factor. The data came in the form of raster data from the Prism Climate Group and the cells were extracted based on the sample block group shape file (Prism, 2014). Aside from temperature, precipitation was also looked into for determining climate. Areas with substantial yearly rainfall may not be as walkable, or as nice to walk around, if it rains most of the year. The data from precipitation came from a thirty year normal precipitation raster file. The precipitation data also came from the Prism Climate Group.

Figure 4 Mean Temperature by Sample Block Group for Washington, DC MSA



3.6: Land Use and Walkability

The final variable category used to define walkability was based on mix land use. This variable was made up of three variables obtained through the Smart Location Database from the EPA (EPA, 2013). These variables consisted of jobs within a forty-five minute transit ride, jobs per household and land use diversity. Land use was chosen as a walkability variable because it is based on the idea that areas with a diverse mixture of land use will encourage walking because people may more easily be able to walk from home to work, school, or shopping if these locations are within their census block group. Thus, land use diversity could lead to lower vehicle ownership as families might forgo a car with less need to drive, or it might lead to lower VMTs as people find work and shopping closer to home.

The variable for jobs within a forty-five minute transit ride reflects accessibility to jobs from homes and therefore may show no need for automobile ownership or at least fewer VMTs. This variable was chosen to reflect land use and walkability on the hypothesis that if people have jobs close to their home they may be more likely to take transit to their job instead of using a personal vehicle.

Jobs per household is used as an indicator of walkability on the hypothesis that areas with a better balance of jobs to residences may promote walking to work or shorter drives to work. Thus neighborhoods with a higher ratio of jobs per household would be predicted to have fewer VMTs and lower rates of automobile ownership.

3.7: Independent Smart Growth Variables

As mentioned at the outset, this study seeks to add variables on smart growth and walkability to Eisman's (2012) initial study. The objective is to understand some of his original variables and models in the context of new and additional independent variables. Thus, this study also incorporates the majority of smart growth variables used within his study. The variables used to determine and measure smart growth effect on VMTs and automobile ownership were density, distance to jobs, distance to retail centers, and transit access.

Densification of neighborhood regional urban development is a leading smart growth strategy. Dense neighborhoods often, if not always, mean multi-family residences such as multi-story apartment complexes. These types of environments often make parking difficult and expensive for both residents and people visiting nearby business or the families in those apartments. Because of this walking or public transit such as light rail or bus lines, make more sense for some people in how they travel in their daily lives. The hypothesis for density is that the higher the density there will be less vehicles owned and less VMTs. Density was measured as the number of households in a block group divided the acreage of the block group.

Distance to jobs is the measure of the distance from the block group to an "employment center." Eisman (2012) defined an employment center as Census tract in the top ten percent of an MSA with total number of jobs. He obtained the job data from the 2000 Census Transportation Planning Package (CTPP). The distance between sample block groups and Census tracts was from the centroids of each polygon. The hypothesis for the variable is that the closer somebody lives to their job, the more likely they are to take public transit and potentially walk if they are close enough. In turn, the farther a person lives from their place of employment the more likely they are to use their own vehicle.

On top of going to work, people need to shop for food, clothing, amenities, etc. Distance to retail is the measure of the sample block groups to the Census tracts in the top ten percent of the MSA in terms of aggregate retail jobs. Like distance to jobs this data was also obtained from the CTPP. The hypothesis is that the closer the people live to retail centers the less they will need to drive and may choose other modes of transportation. However, it should be noted that some forms of shopping or long shopping trips may require the need for the use of a personal vehicle to transport purchases to a person's residence.

The last smart growth variable utilized from Eisman's study is distance to transit stations. This variable refers to rail transit only making the walkability variable of bus stops per square mile viable and a nice addition to the walkability variables. Rail transit can be a great mode of transportation for some people. Light rail services like trams or trolleys can be a comfortable and cost effective way for people to get to and from work but only if the stations are accessible. Distance to transit is the measure of the sample block groups to a transit station. The hypothesis is that better access to transit stations will encourage lower VMTs and automobile ownership

3.8: Independent Regional Variables

Originally these independent regional variables were utilized in Eisman's study as confounding variables. However, this study spends more time and delves deeper into their possible effects on VMTs and automobile ownership through scatterplot analysis and analysis of potential interactions with block group scale variables. The independent regional variables were individual transit expenditure, regional (MSA) density, and people per transit station (rail).

Individual transit expenditure is a variable that measures the annual consumer expenditure on public transportation per person for a given region. This variable is measured at the MSA level and the data came from the Bureau of Labor Statistics' 2000-2001 Consumer Expenditure Survey. It is expected that higher expenditures will have a negative effect on VMTs and automobile ownership and should have a correlation with the land use and number of bus stop variables.

People per transit station is the measure of the total MSA population divided by the total number of transit stations in the MSA. It is expected that the fewer people there are per transit station the lower VMTs and automobile ownership will be, because fewer people per transit station means that there more of an abundance of transit stations for people to have access to.

The last variable, regional density is a measure of the MSA population per square mile of area. It is expected that more dense regions will have lower VMTs and automobile ownership as a more dense region is likely to have larger more densely populated urban areas (Cervero and Murakami 2009). People living in a highly populated area may be more likely to use bus transit, rail transit, or live in areas where walking is a quicker and more economical mode of transportation. By looking at the regional density as well as neighborhood density of the samples a better picture of the population within the entirety of the study area.

3.9: Dependent variables

This study explores whether neighborhood walkability affects vehicle miles traveled and automobile ownership. Automobile ownership was calculated by taking 2000 census data and using a variable for total number of vehicles available in a block group. This number is then divided by the total number of households in the corresponding block group. This calculation gives a ratio measure of the average number of automobiles owned within a block group by household.

Vehicle miles traveled is a measurement used to determine the number of miles traveled by people going to and from work, shopping, school, etc. It was derived from data taken from the 2001 National Household Travel Survey (NHTS) conducted by the Federal Highway Administration. Utilizing a model by Hu et al. (2007), VMTs at the Census tract level were measured using household size, household income, and employment rate. In order to get the VMTs for Census block groups, they were assigned the VMTs from their corresponding Census Tract. This method was used by Eisman to calculate the VMTs as NHTS data is not available at the block group level. The estimates for the VMTs are per household on an average weekday, and the tract level estimates were given based on vehicles available to a household and the size of the household (Eisman 2012).

3.10: Confounding Variables

Other variables in this study used from Eisman's work are the confounding variables of household income and age of neighborhood population. Confounding variables are variables that may affect the outcome of the study and are not based on the primary tested variables. However, there may be relationships within the built environment that may exist between these the variables and the rest and there needs to be control for these unrelated variables in order to see these relationships properly.

The income variable was measured as the average income of a block group. Differences in income may affect vehicle ownership based on whether people can afford a single vehicle or multiple vehicles. For example, lower income households may have no vehicle or a single parent household might have lower income and also use only a single vehicle. In contrast, higher income households could have multiple vehicles, for example, in cases where two parents are working or there are households with multiple drivers.

3.11: Geoprocessing of variables

The data for this study was processed using Esri ArcGIS 10.1 desktop software. Where data from Eisman's work was used in this study, no additional processing was done with his data. On the new independent variables developed for this study, the analysis was performed on data acquired from several different sources: Google Earth, the USGS, the PRISM Climate Group and the EPA. The data were geoprocessed to develop the measures described above before the values were taken in and added to a spreadsheet for analysis.

The first step in the process was acquiring the bus stop data. To acquire bus stop data, the original method searched for data from each county in which the sample block groups were contained. This proved to be time consuming as well as fruitless, as many counties that may have bus routes did not have the GIS data for them. To solve the problem, the Google Earth software and its extensive spatial database was utilized. As Google has already aggregated nationwide bus data displaying bus stops and routes as well as schedules, it was a great source for this hard to acquire data. Additionally due to the street view feature it was possible to check through samples to see if bus stops did in fact exist at the specified locations. A ¼ mile buffer file was created around each sampled block group file and imported as a KML file into Google Earth. Each sample was then checked to see if any bus stations exist within each buffer and those that were, were selected and individually exported as KML files. These files were then exported into shape files and merged with each other in their corresponding regions. Bus stops were calculated by measuring total bus stops divided by the area of the sample and its ¼ buffer and recorded.

Figure 5 Google Earth KML Data for Bus Stops in ArcMap for Conversion to Shapefile

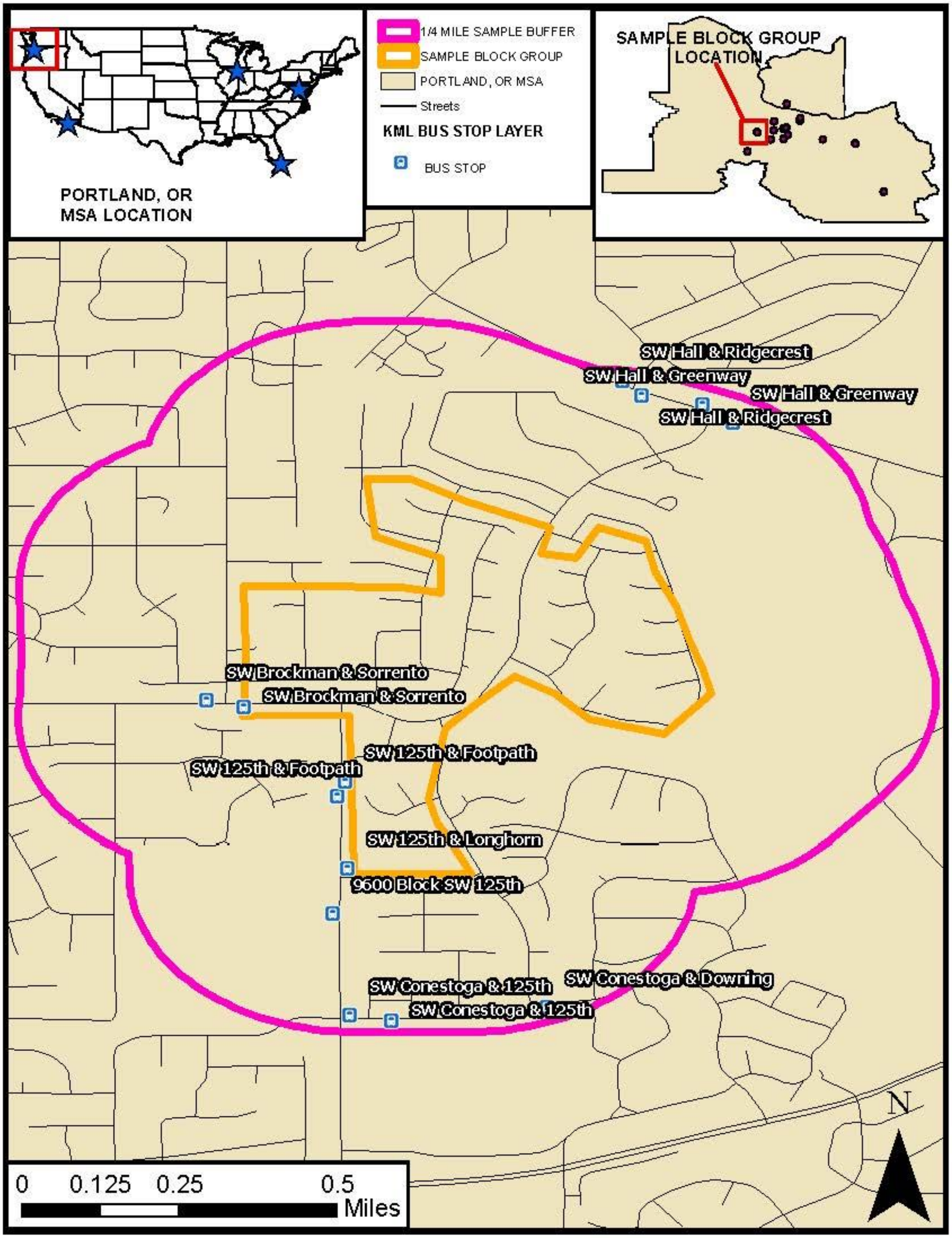
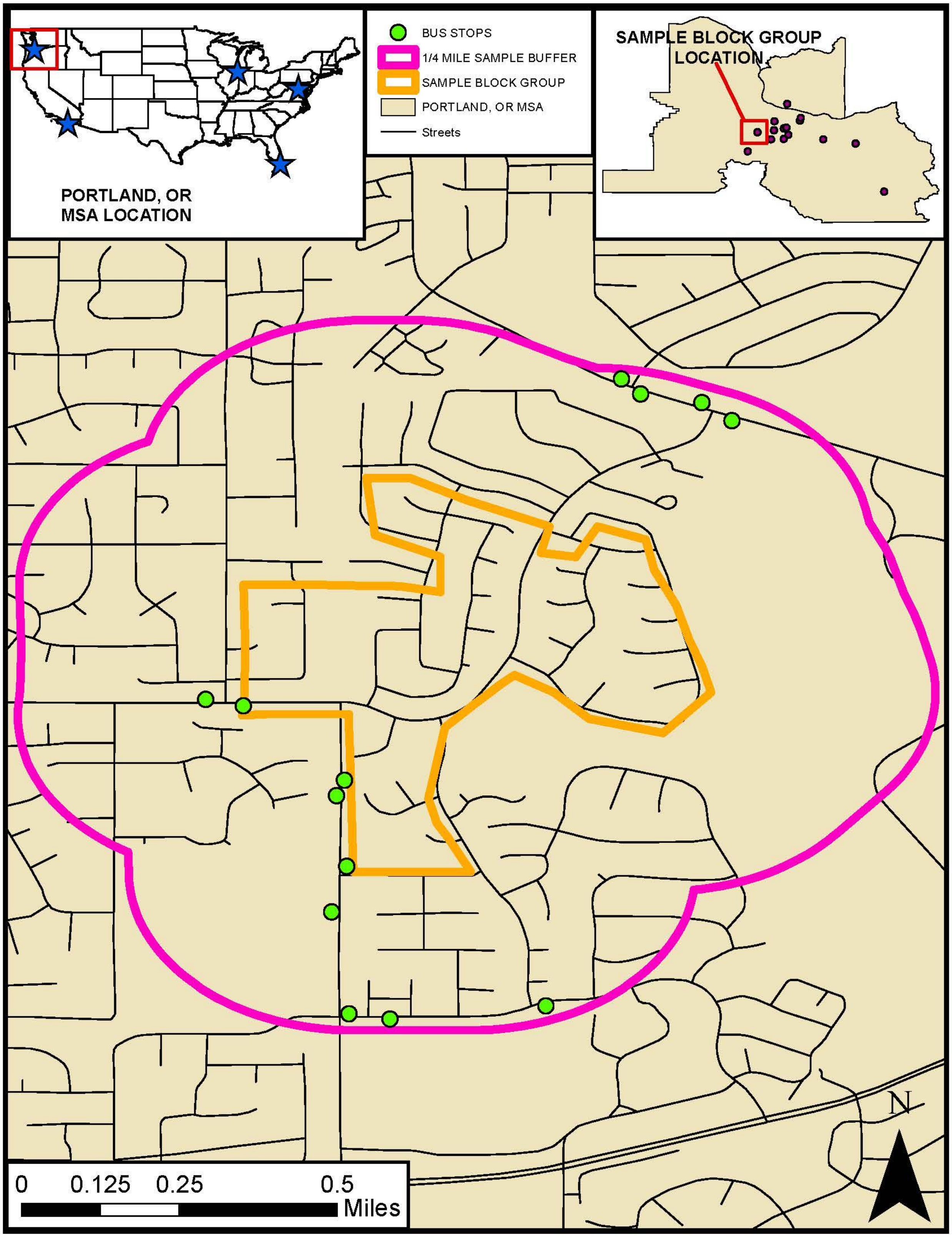


Figure 6 Bus Stop Shapefile created from Google Earth KML Files



The elevation variable was acquired utilizing United States Geological Survey, Digital Elevation Model (DEM) raster data at 1/3 arc (10 meter) resolution (USGS, 2014). To extract the elevation values for each sample, a modified ¼ mile buffer was created. Since the data is a walkability variable, water was deemed "not walkable" and needed to be removed from the elevation raster. This was done by the use of a modified ¼ mile buffer. The buffer which was used to locate bus stations was clipped with a North American Water Polygons layer found on ArcGIS online and produced by Esri. The water layer was clipped with the ¼ mile buffer to create a ¼ mile buffer with water removed. After the new ¼ mile buffer was created, the "extract by mask tool" from the "Spatial Analyst Toolbox" within ArcMap was used to create new a DEM raster for each sample, the result was 75 DEMs (i.e., one for each study area) with water removed.

The digital elevation models are used in this study to measure slope. As mentioned above, two different slope measurements were created for each census block group. One measurement describes how hilly the entire block group land is (omitting only water) and the other measure focuses just on the hilliness of the road network. In order to calculate slope for the entire block group, the "slope" tool also within in the "Spatial Analyst Toolbox" was used with the "percent rise" measurement. This calculation was performed on each raster cell within each sample DEM for each sample area. Next, to calculate slope for just the road network, the road network was described using shapefile data from US Census Tiger files. The road network polylines were used with the "extract by mask" tool to create slope data for only roads. In both cases the mean slope of the raster data was recorded for analysis.

The climate variables included 30 year normals for maximum temperature, minimum temperature, mean temperature and precipitation and was acquired from the Prims Climate

Group (NACSE 2014). To process these 4 raster datasets, the normal ¼ mile buffer was used to extract the raster data for each sample area from each dataset. Next the mean temperature or precipitation for all the raster grids within each study area was recorded in the data table.

The last variable analyzed was for mixed land use. The data to use for these variables was difficult to determine as to what would work best for the large scale of the study. Acquiring individual land use data for counties or regions proved difficult and time consuming as many of the study areas do not have their own GIS data available for this type of data.

Fortunately, there is a pre-existing data set that estimates the mix of land use at the level of the census block group for the entire United States: the Smart Location Database (SLD) from the United States Environmental Protection Agency (EPA 2013). This database was created using data from the 2010 Census, American Community Survey, Longitudinal Employer-Household Dynamics, InfoUSA, NAVTEQ, PAD-US, TOD Database, and GTFS (General Transit Feed Specification). The specific variables within the Smart Location Database that were chosen to best determine mixed land use were activity density, jobs per household, and jobs within a 45 minute transit ride.

Activity density is measured by employment plus housing units and jobs per household is measured by total employment divided by households (TotEmp/HH). Employment is measured in the SLD by using the Longitudinal Employer-Household Dynamics (LEHD) which consists of US Census LEHD Origin-Destination Employment Statistics (LODES) tables that summarizes employment at the census block level for all 50 states and territories. Additionally within the LODES there are Work Area Characteristics (WAC) tables which are used for employment tabulations. Household units are calculated by block group utilizing population, housing or employment within a block group and fall under a density category within the SLD. This data is calculated using the US Census data. Jobs within a 45 minute transit ride was calculated based on walk network travel time and GTFS schedules (Ramsey & Bell 2014).

One issue with using the Smart Location Database in this study is that all of the data are based on the 2010 census and 2010 census tiger files while this study is based on 2000 census data. This creates an important limitation in that the measured state of the study areas does not match precisely in time. The assumption is that the land use mix did not change drastically in a ten-year period in any of the study areas.

Another aspect of the mismatch of the land use data in time is that the block group data does not overlap to the precise spatial boundaries of the polygons found in the 2000 census data used in this study. When the SLD data was clipped with the sample block groups, it was found that block groups from the samples overlaid across multiple block groups from the SLD. In order to deal with this when determining values, the 2010 polygon sharing the largest area with the sample polygon was selected. This is an acknowledged limitation of the data but in most sample areas there are not be significant change in structures and there is fairly close overlap between block group polygons as defined for the 2000 and for 2010 Census.

3.12: Statistical analysis

Statistical analysis for this study was performed utilizing SPSS for linear regression modeling. In order to create linear regression models the data needed to be assembled into a table. This was done by taking the values from the data in ArcMap and importing them into the Excel sheet with one row for each sample. Once the data was in the Excel sheet, it was loaded into SPSS v 21 where analysis of the variables was done. Histograms and descriptive statistics were drawn for each variable. The variables, excluding regional variables, were all non-normally distributed. To make the variables fit better into the regression models based on "normal" distributions, the natural log was taken for each of these variables. Next in order to determine if the variables had a correlation between each other and especially the dependent variables, a bivariate correlation table was created in SPSS.

Once the data was set up and analyzed, the multivariate linear regression models were developed within SPSS. Variables were added and removed within the models to achieve the best fit based on the significance of the relationship within the regression model. To determine the strength of the relationships between the independent variables and the dependent variable in the model, the adjusted R squared was used.

CHAPTER 4: RESULTS

This chapter reports the results from the data analysis, starting with bivariate correlations of each independent variable with the two dependent variables. Next, the chapter reports the results of exploration of the possible regional effects. Last, the chapter reports the results of linear regression models for both vehicle miles travelled (VMTs) and average household automobile ownership.

4.1 Bivariate Correlations

The first part of the analysis process involved creating a bivariate correlation table in order to determine if there were any significant correlations between the dependent variables (VMTs and automobile ownership) and the independent and confounding variables. The following table (Table 4.1) shows the variables with significant correlations with the two dependent variables. A full matrix of the bivariate correlation of each variable in the study was created. It was inspected both for correlations with the dependent variable and for possible correlations among the independent variables. High correlation among the independent variables is important to consider in building regression models because it can make it difficult to compare and understand the explanatory power of any individual predictor in the overall regression model, i.e., the problem of "multicolinearity" (Allison 1999).

Table 2 Table of Significant Correlations

		Vehicle Miles Traveled	Automobile Ownership
Density	Pearson Correlation	-.596**	-.638**
	Sig. (2-tailed)	.000	.000
Distance to Job Center	Pearson Correlation	.501**	.499**
	Sig. (2-tailed)	.000	.000
Distance to Retail Center	Pearson Correlation	.349**	.353**
	Sig. (2-tailed)	.002	.002
Distance to Transit	Pearson Correlation	.489**	.547**
	Sig. (2-tailed)	.000	.000
Bus Stops per Square Mile	Pearson Correlation	-.699**	-.681**
	Sig. (2-tailed)	.000	.000
Mean Slope	Pearson Correlation	.275**	.280*
	Sig. (2-tailed)	.017	.015
Minimum Temperature	Pearson Correlation	-.315**	No Correlation
	Sig. (2-tailed)	.006	
Mean Temperature	Pearson Correlation	-.261*	No Correlation
	Sig. (2-tailed)	.024	
Gross Activity Density	Pearson Correlation	-.702**	-.739**
	Sig. (2-tailed)	.000	.000
Jobs Within a 45 Min Transit Ride	Pearson Correlation	-.554**	-.633**
	Sig. (2-tailed)	.000	.000
MSA Density	Pearson Correlation	No Correlation	-.234*
	Sig. (2-tailed)		.043
Transit Spending	Pearson Correlation	.248*	No Correlation
	Sig. (2-tailed)	.032	
People Per Transit Station	Pearson Correlation	-.278*	No Correlation
	Sig. (2-tailed)	.016	

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2.tailed).

For ease of presentation, the following (Table 2) shows the variables with significant correlations with the two dependent variables. From the correlation tables it was found that the dependent variable for the average household VMTs in each census block group has many significant correlations with both smart growth and walkability variables. Slightly more of the independent variables have significant bivariate correlations with the VMTs than automobile ownership. There are thirteen variables where VMTs show a correlation with most being significant at the .01 level of significance. Automobile ownership had significant correlations with ten of the testing variables. Most variables had a significant correlation with both dependent variables, and as expected the two dependent variables are themselves correlated with a Pearson's R of (-.875* at the .000 level (i.e., households with more vehicles also drive more).

It is an interesting note that of the few variables with correlations with only one or more dependent variables, most are measured at the regional level. For example, MSA density, as measured at the regional level does not correlate with VMTs but does correlate with automobile ownership. The four variables significant with only VMTs are transit spending, people per transit station, minimum temperature, and mean temperature. Transit spending and people per transit station are explicitly measured at the regional level. Also, the variations in climate among the study regions mean the temperature variables are implicitly regional, though each census block group has a slightly different measure of temperature due to microclimates within regions. In looking at bivariate correlations with regional variables, an important caveat to bear in mind is that there are really just 5 independent measures of each variable, even though the total number of cases is 75.

VMTs had the strongest significant correlation with gross activity density (Pearson's R= -.702**), bus stops per square mile (Pearson's R = -.699**), Density (Pearson's R = -.596**) and jobs within a 45 min transit ride (Pearson's = -.554**). Gross activity density is the measurement for mixed land use or land use diversity (shops mixed with residences etc.). The significant correlation shows that as the land use becomes more mixed (i.e., economically diverse) the number of VMTs goes down. This supports the original hypothesis for the variable. The significant correlation results for the other three variables show the same inverse relationship. As bus stops increase, VMTs go down and the same goes for jobs within a 45 minute transit ride. These results also support the hypothesis for the use of the variables within the study and are promising for use within the linear regressing model.

Automobile Ownership had the strongest significant correlation with gross activity density (Pearson's R = -.739**), bus stops per square mile (Pearson's R = -.681**), Density (Pearson's R = -.638**), and jobs within a 45 min transit ride (Pearson's R = -.633**). Like VMTs, these variables show an inverse relationship. It should be noted that the top four strongest significant correlations for both dependent variables are all the same variables in the same order. The reason for this is unclear but shows that these four variables are strong indicators both for VMTs and automobile ownership, again pointing to significant overlap in what these two variables measure.

4.2 Regional Variables

In addition to the variables at the individual census block group level, there are three regional variables. Regional variables only consist of five values, one value per region, instead of the usual fifteen independent values per region for each sample (i.e., one value per census block group). Because of the nature of the regional variables there was further evaluation of the variables to explore their significance. The mean values by region for each dependent variable were calculated and difference of means testing was performed on them. Table 3 shows the means grouped by their statistically significant differences. Scatterplots were also drawn of the regional variables with each of the dependent variables to inspect for discernible trends that might point to relationships.

Table 3 VMT and Automobile Ownership Mean Values

MSA	Mean VMTs
Washington DC	59.62219998
San Diego	52.3365337
Portland	50.83582958
Chicago	49.58853866
Miami	36.88243465

MSA	Mean Automobile Ownership
San Diego	1.851924968
Washington DC	1.801894624
Portland	1.744232531
Chicago	1.507537922
Miami	1.49282606

Mean VMTs differ greatest between Washington DC and Miami with DC having the highest average vehicle miles traveled in its neighborhoods at 59.6 VMTs and Miami having the least at 36.8 VMTs. Automobile ownership has little difference between San Diego, Washington DC and Portland who all are at or close to 1.8 vehicles per household, while Chicago and Miami have the least with both at or near 1.5 vehicles per household.

A scatter plot was drawn for each of the three regional variables with each dependent variable. The regional variables were transit spending, MSA density, and people per transit stations. The purpose of creating scatter plots for these variables is to perform visual inspection to identify potential relationships between the regional and dependent variables. To read a scatterplot, a line is drawn through the clusters of points on the individual columns. Depending on the angle of the line, a relationship between the two variables can be determined. For example in Figure 7, VMTs tend to drop as MSA density increases. This shows there is a possible relationship between VMTs and MSA density even though San Diego is fairly spread along its column there is still a decreasing trend. Additionally this scatterplot shows Chicago and Portland having similar VMTs, this can be easily explained in that Portland has an active smart growth development plan and promotes mixed land use and walkability (Jun, 2008). Automobile ownership and transit expenditure shows no relationship (Figure 8), the clusters along the columns are relatively close and have a minimal rise and drop as transit expenditure goes up. This rise and fall shows no constant trend and thus indicates there may be no relationship between the variables. After creating the scatterplots for the 3 regional variables, it was concluded that VMTs had potential relationship with MSA density and people per transit station. VMTs had no relationship with transit expenditure. Automobile ownership had a potential relationship with MSA density and no relationship with people per transit station or transit expenditure.

Figure 7 Scatterplot of VMTs and MSA Density

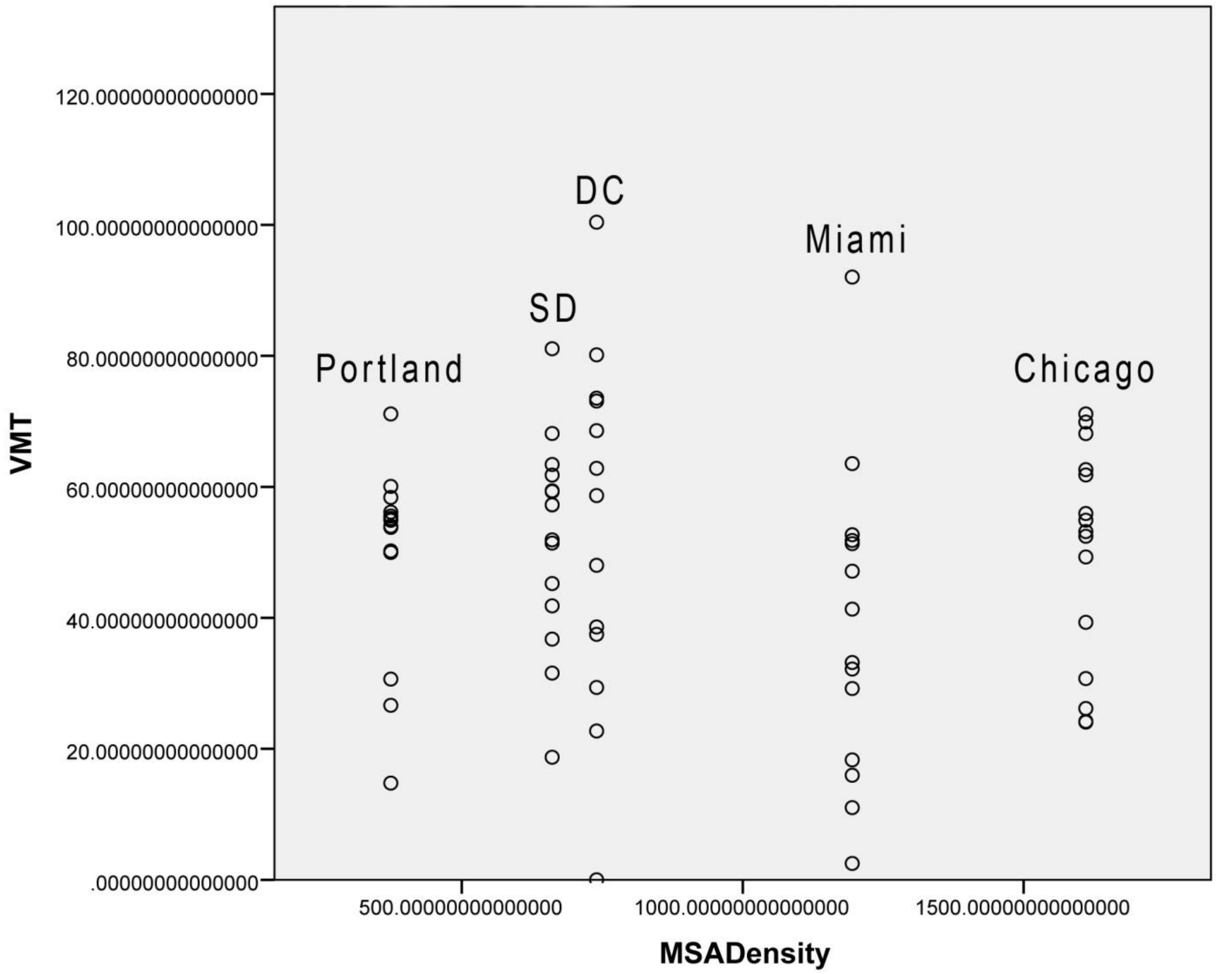
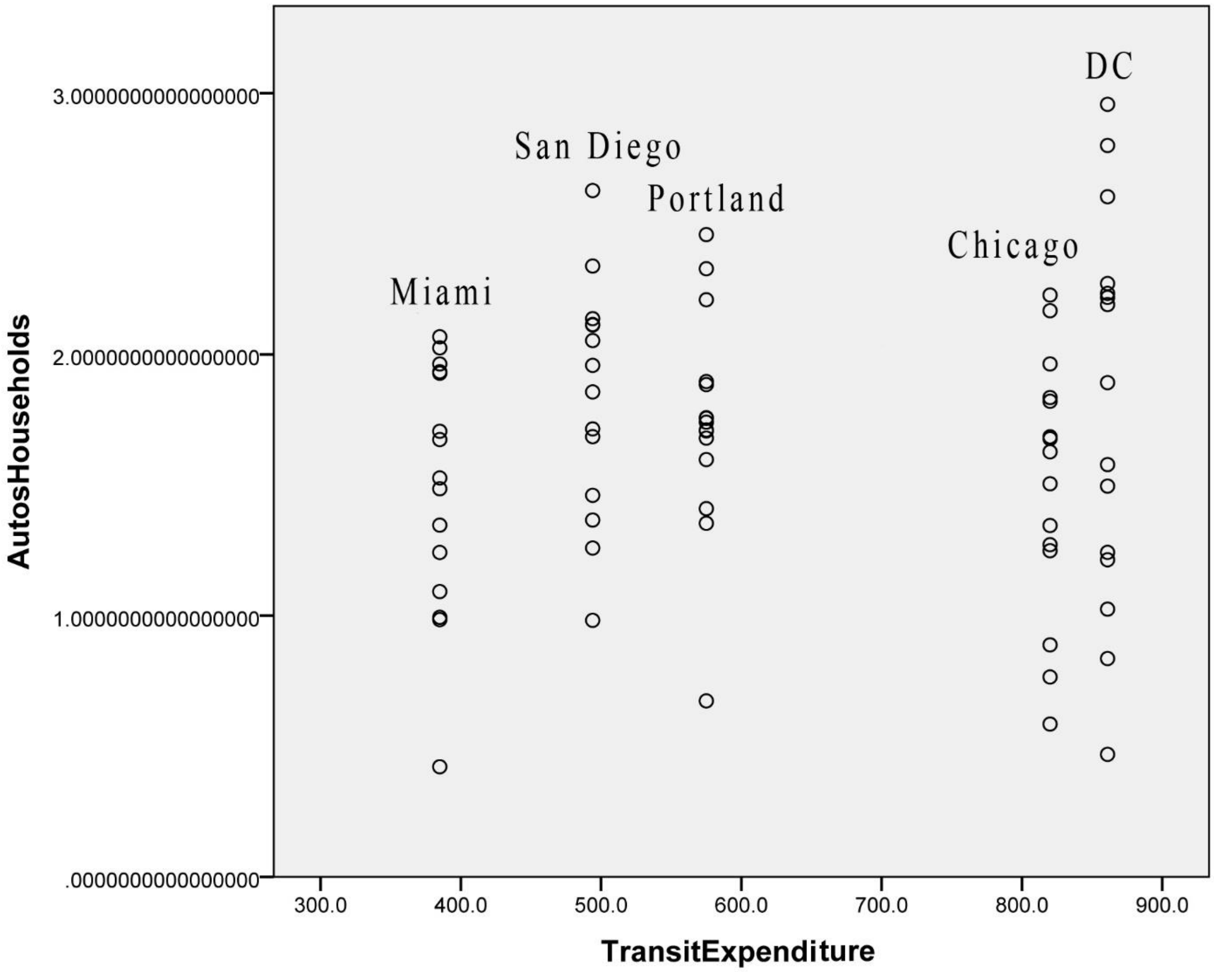


Figure 8 Scatterplot of Automobile Ownership and Transit Expenditure



4.3 Interaction Testing of Regional Variables

In order to test the regional model's effect within the regression model, interactions were performed between the regional variable MSA density and the independent variables found in the best fit regression model. An interaction is performed by multiplying the regional variable with the neighborhood variable, this gives a new variable to be included in the regression model along with the variables that were used to create the interaction. By doing this, the study attempts to find strong relationships with the dependent variable.

MSA density was chosen for the interaction due to its possible relationship with both dependent variables. The independent variables chosen for the interaction were bus stops per square mile and gross activity density (mixed land use) for the VMT regression model and jobs within a 45 min transit ride and gross activity density for the automobile ownership regression model. These variables were chosen for the interactions because they were most likely to be affected by MSA density.

The interaction between MSA density and the independent variables showed no relationship and did not add to the regression model. When the interactions were included within the linear regression model (one at a time) the interaction with bus stops per square mile had a significance of .455 which shows no significance in the model, and gross activity density had a significance of .396, also showing no significance. The interactions within the automobile ownership best fit regression model were more interesting, as the interaction between jobs within 45 minute transit ride was just significant at .052 but within the model made the jobs within 45 minute transit ride variable not significant by raising it to a significance of .588. This makes the interaction void as the variable is eliminated from the model. Gross activity density was also not significant at .215.

In addition to performing interactions with the MSA density regional variable, a dummy variable was created for regions based on the means tests for VMTs and automobile ownership. The variable was created by assigning a single number to all the block groups in an MSA based on the MSA's ranking from highest to lowest mean for VMTs and then for automobile ownership. The dummy variable was then entered into the best-fit models to test for regional variation relationship with the independent or dependent variables.

Both the interactions with MSA density and the regional dummy variable showed no impact or significant relationship. Because of this, the null hypothesis for regional variation as a predictor within the models was accepted.

4.4 Linear Regression Modeling

Best-fit regression models were created for both VMTs and automobile ownership taking into account the new independent variables developed for this study. The adjusted R-squared was used to compare the predictive power of the different models as this is a common method of determining the strength of relationships in linear regression models with relatively small numbers of observations (Allison 1999).

4.5 Vehicle Miles Traveled Linear Regression Model

The best-fit linear regression model for VMTs utilized bus stops per square mile, distance to retail centers, gross activity density and income (table 4.6). With the exception of the confounding variable of income all other variables used the natural log in the model. The adjusted R-squared for the model was .598 indicating a near 60% explanation for variation in VMTs from the four variables. Within this model, income is an important control variable: wealthier neighborhoods drive more. However, even when controlling for income the other variables are all significant predictors for VMTs. In other words, on average households in wealthy neighborhoods with more bus stops, longer distance to major retail centers, and mixed land use drive less than wealthy households in neighborhoods without these features.

Table 4: Vehicle Miles Traveled Best-fit Regression Model

Vehicle Miles Traveled		
Adjusted R-Squared	0.598	
Model	**Intercept**	**Significance**
Constant	11.842	.000
Bus Stops Per Square Mile	-2.684	.009
Distance to Retail Center	-2.08	.041
Gross Activity Density	-4.603	.000
Income	2.579	.012

Eisman's (2012) best-fit model had an adjusted R-squared of .404 with population density by census block group and transit network (distance to rail stops) being the predictive variables. It was found when running models with his original best-fit as the starting point that, when using the new variables developed for this study, the distance to rail and population density variables dropped out and another of the variables Eisman (2012) developed (distance to retail center) became significant in the model. Bus stops per square mile superseded distance to rail and in fact on its own, bus stops per square mile had an adjusted R-squared of .482 making it a very strong predictor of VMTs. Population density which is one of the top four in the bivariate correlation is superseded within the regression model by bus stops per square mile, gross activity density, and distance to retail. It is interesting to see that distance to retail is significant with these new variables from this study. The intercept for distance to retail has also changed from the bivariate correlation meaning that even as distance to retail goes down, VMTs still go up.

This could be from the nature of shopping trips to retail centers, people may just be buying more or larger items than they can walk with or carry on public transit.

An important caveat to mention here is that three independent variables had a strong correlation with each other. Table 5 shows this relationship between the three variables. These levels of correlation between the predictor variables may indicate issues with multicolinearity, making it difficult to specify the precise strength of explanation from each variable. Income had no bivariate correlation with either these variables or the dependent variables.

Table 5 Correlation between the 3 independent variables in the best-fit regression model

		Bus Stops per Square Mile	Gross Activity Density	Distance to Retail
Bus Stops per Square Mile	Pearson Correlation	1	.770*	-.525**
	Sig. (2 tailed		.000	.000
Gross Activity Density	Pearson Correlation	.770*	1	-.675**
	Sig. (2 tailed	.000		.000
Distance to Retail	Pearson Correlation	-.525**	-.675**	1
	Sig. (2 tailed	.000	.000	

There were other variables which showed promise when creating the best-fit models. The top four of the bivariate correlation table were each significant within the model but were superseded by other variables. The variable for jobs within a 45 min transit ride was significant within the model but was superseded by distance to retail centers. This is interesting as it plays a role in the model for automobile ownership. Again, the caveat with regards to multicolinearity applies because both of these variables have a strong correlation between the dependent variables as well as with each other (Pearson's R= -.304** at the .008 level) and with other variables within the best-fit model.

4.6 Automobile Ownership Linear Regression Model

The best-fit linear regression model for automobile ownership utilized jobs within a 45 min transit ride, gross activity density, and income. The adjusted R-squared was .641 indicating a 64% explanation of automobile ownership. Like the best-fit for VMTs, income is again a control variable.

Table 6: Automobile Ownership Best-fit Linear Regression Model

Automobile Ownership		
Adjusted R-Squared	0.641	
Model	**Intercept**	**Significance**
Constant	23.592	.000
Jobs Within a 45 min Transit Ride	-3.887	.000
Gross Activity Density	-0.612	.000
Income	3.102	.003

Eisman's best-fit model had an adjusted R-squared of .445 with density, distance to transit, and income being the predictive variables. Again when running the linear regression models his model was a starting point. Density which shows up in both of this models was superseded by gross activity density (mixed land use) and jobs within a 45 min transit ride superseded distance to transit. Interestingly income remained as a control variable within the model.

Most of the variables had little significance on automobile ownership including bus stops per square mile which was so strong for VMTs. With the exception of income which had no significant correlation with either the dependent or independent variables, again, like the VMTs best-fit model, the independent variables had a significant relation to each other.

4.7 Conclusion

This analysis has looked at both smart growth variables from a previous study and new variables intended to measure aspects of walkability. These new variables show great promise with determining a smart growth and walkability design plan for urban and suburban development. While not all variables show promise there are some stand out predictor variables: bus stops per square mile, gross activity density, and jobs within a 45 min transit ride.

Bus stops per square mile is a strong indicator of a walkable environment and the linear regression model appears to confirm the hypothesis that a greater number of bus stops, and one might thus hypothesize a larger bus network overall, promotes less driving and allows people to conduct their business without driving or spending money on gas. Gross activity density is not only a walkability indicator but also a smart growth indicator. The models show that a more diverse land use promotes fewer VMTs as well as reduces the need to own and vehicle. Jobs within a 45 min transit ride appears to support the hypothesis that nearby transit can reduce the need for owning a vehicle or at least multiple vehicles, though it does not add to the explanation in the best-fitting multivariate regression model for predicting VMTs.

CHAPTER 5: DISCUSSION AND CONCLUSION

The evidence from the analysis of the variables in this study supports the hypothesis that smart growth and walkability elements within the built environment promote lower VMTs and lower automobile ownership.

The findings of this study support the findings from the other smart growth studies, though in some cases in different ways as some of the studies used different variables and methods to conduct their research. Cervero and Murakami (2009) found that population density was a factor in reducing VMTs, while other variables such as transportation did not have as strong of an effect on VMTs as population density. This study found that the number of bus stops and land use have a greater effect on VMTs than population density. This study also tested interaction between neighborhood and regional scales more thoroughly than Cervero and Murakami (2009).

Cervero and Duncan (2006) used a different method and variable types for their study. Instead of testing individual land use and census variables, they tested two different urban development types/concepts: jobs-housing balance and retail-housing mixing (mixed land use). They found that jobs-housing balance had a greater effect than mixed land use. That is, living closer to the place of employment through assistance of grant programs is more effective at reducing VMTs than mixed land use of housing and retail. This is a different outcome from this study which found that mixed land use (gross activity density) did have a significant effect on VMTs. I would argue however, that the very strong significant association of bus stops per square mile with reduced VMTs supports the findings with jobs-housing balance. If a person lives close to their job and there are nearby bus systems, this would allow people to commute without using a personal vehicle.

Kim and Brownstone found that land use and population density had statistically significant effects on VMTs with the emphasis of their study being on population and socio-demographic variables. This study found that population density while significant, is less significant than other variables, bus stops and land use being the most significant.
This is a similar result found by Eisman (2012). Eisman's study found that VMTs were significantly associated with the neighborhood's population density and extent of region's transit (rail) network. As density and transit network increased, VMTs decreased. Arguably the association is significant and helps prove the hypothesis, however, regional transit and density do not give a clear picture of the built environment as regional transit is a regional variable that does not study the neighborhood level and thus does not give very thorough explanation of the built environment, additionally rail is not the only mode of mass transit and runs to fixed locations that are generally less flexible and available than bus networks.

This study uses Eisman's (2012) original work and makes an effort to expand on his idea while utilizing walkability aspects to expand the regional smart growth study. The results from this study supports Eisman's findings while giving new explanations for reduced VMTs and automobile ownership. The best-fit regression model for VMTs from this study showed a significant association between the built environment and VMTs. Specifically, bus stops per square mile, distance to retail centers, gross activity density (mixed land use), and income. While two of the indicators (distance to retail centers and income) are from Eisman's original study, the other two important variables are new and come from the determined aspects of walkability variable group. In addition to having a greater R-squared than Eisman's best-fit model, the variables are more descriptive in their assessment of the built environment and its effect on VMTs. The analysis shows that as the number of bus stops increases and land uses become more mixed, VMTs decrease which supports the hypothesis. Curiously the model also shows that as distance to retail centers increase, VMTs decrease, that is counter to the intial hypothesis. It is assumed that the closer to a retail center a person lives, the less they would drive. Income, while not a land use measurement it is a strong control variable that can also help explain people's vehicle usage. As income goes down, VMTs go down and when income goes up VMTs go up. Essentially VMTs go up for people who have the money to buy vehicles and afford to use them. This of course isn't the only explanation for reduced VMTs, but helps explain them in conjunction with the other variables.

Within the best-fit linear regression model for automobile ownership, Eisman found that there was a significant association with neighborhood population density, distance to transit, and income. His findings support the hypothesis that smart growth has an effect on lower automobile ownership. The same holds true for this study as well, and it is interesting in that the variables within Eisman's best-fit regression model are similar to those within the best-fit regression model for automobile ownership within this study. While density did not play a role in the model for this study, income and jobs within a 45 minute transit ride did. While both distance to transit (rail) and jobs within 45 min transit (all modes) ride do not have exactly the same measure, they are both transit variables that measure similar things. The fact that both variables show up within the two different models suggests that availability of public transit is an important factor in determining how many cars, if any, are owned by individual households. With many two income households, this finding makes sense because with availability of transit, it is perhaps possible for at least one person to get to work without a car.

Income is again an important control variable in this study and in Eisman's. Income directly effects whether people can afford to own and operate a vehicle. Gross activity density (mixed land use) is another variable that has a statistically significant association within both best-fit regression models. This means that areas with a more diverse land use and suitable access to transit either have less of a need to own a vehicle or more than one.

Of the new variables added to this study, a few stood out as highly significant: bus stops per square mile, jobs within 45 minute transit ride, and gross activity density were all highly significant during the development of the models. Bus stops per square mile was associated significantly with VMTs and had influence on all other variables within the models. This suggests that access to bus stops and in turn bus routes is important when people make decisions about driving.

Jobs within 45 minute transit ride was another significant variable that in early models for VMTs was significant. However when put into a model with bus stops it lost its significance. It was however stronger in models regarding automobile ownership and held its significance in most models it was incorporated into.

Lastly, gross activity density was the third variable that was very significant within the models. The mixed land use variable had a significant association with both dependent variables and ended up in the best-fit regression models for both. While bus stops per square mile was extremely significant within the VMT regression model, the importance of mixed land use cannot be ignored. The idea of mixed land use is one of the aspects of smart growth. To have it be a significant part of both best-fit models supports the idea of smart growth as well as supporting the hypothesis within this study.

5.1 Assumptions and Limitations

Within this study there were some data issues and limitations. As stated earlier, the data from the Smart Location Database were based on the 2010 census because the SLD does not have 2000 Census data available. This is an acknowledged limitation and could skew the data. However, major development generally has not occurred within the selected study are areas in the intervening decade. This was checked utilizing historical aerial imagery through Google Earth, showing that significant new structures or road network were non-existent or rare in the study areas. Additionally, where 2010 Census block groups have changed since the 2000 Census data, this was corrected for, by taking the values from the 2010 block group that was the majority within the 2000 Census sample block group.

The second limitation and assumption on data was the bus stops. While it can be confirmed that bus stops exist in 2014, there was no way of confirming their existence in 2000. Attempts were made but the sheer number of bus lines and operators as well as jurisdictions involved made confirming installation periods difficult if not impossible to do in any reasonable amount of time. In future studies of smart growth, the availability of bus stop data will be critical because this variable is very strong and an important aspect of a walkability and smart growth study. In future studies, data on the dependent variables may also be updated, eliminating the concern with projecting bus stops back to the year 2000.

Sorting out data between 2000 and 2010 data proved to be a difficult. This makes conducting a longitudinal study with this array of variables difficult and extremely time consuming. This is in part because of the number of different variables, the difference in variables, the number of jurisdictions involved in a multiple region study and that many counties and cities within these regions don’t maintain their own GIS data or have GIS systems. Bus stops, for example, were extremely difficult to find until the Google Earth data was adopted. In order to perform a longitudinal study at these study scales, national datasets need to be used in order to obtain all the data. Unfortunately those types of datasets are hard to come by, especially a dataset with both historical and current values.

The last of the limitations within this study, and I would argue Eisman's as well, is the nature of the regional variables, mainly the extent of region transit network and people per transit stations. These measurements may not be as comprehensive as they should be and may need to be looked at in the future. It may be that there are important regional effects or important interactions between regional and neighborhood variables. Indeed, the difference of means results for the regions suggest this. However, the regional variables explored here do not offer good explanations, it may be that the regional variables need more robust measurement than was developed for this study.

5.2 Future Research

This method of comparing smart growth and walkability at the neighborhood level amongst regions is a good way to identify the importance of smart growth as well as identifying regions that implement it. However, there is some variable identification and measurement that need to be refined as well.

One of the variables that was originally assumed to be important in determining walkability was climate. For this study, the 30-year normal measurements for minimum, maximum and mean temperature were used assuming that this would be a suitable method of measurement. This data had no real affect within the regression models. There may be however, a better method for measuring climate within multiple regions. While the climate data raster files were clipped down to the neighborhood level for this study, it may make more sense to turn it into a regional variable. Whether or not that would have any significant impact within this study is unknown and something to be tested in the future. There may also be more robust ways of indexing the climate variables to describe the real effects of climate on pedestrians.

As stated within the limitations section, regional variables need a revamping in how they are measured. There may need to be considerable thought on the matter to determine the best ways to measure regional data and which variables make the most sense to be measured regionally. Transit network data may need further development to test interaction between regional and neighborhood scales. The area of network analysis may provide possible, more complex methods at measuring the impact of transit.

Another aspect that should be looked at is the counter to the hypothesis that distance to retail centers has with VMTs. That is, as VMTs decrease, distance to retail centers increase. There could be some simple explanations to this such as that people generally use their vehicle when shopping in order to haul their purchases. Large retail centers with big box stores tend to promote large or bulk purposes which would require the use of a vehicle to haul these purchases back to a person's home. If this is the case, then that would show that there is no really way that urban design around large retail centers would have any effect on reducing VMTs.

Lastly bus stops per square mile may be an even stronger determinant of walkability and smart growth if other bus data types are acquired. Bus route length with and number of stops per line may give a better description about how useful a bus network really is in the built environment. There was an attempt in the study to locate this data but no reliable source could be found to consolidate bus network data.

REFERENCES

Allison, Paul D. 1999. *Multiple Regression: A Primer*. Thousand Oaks, California: Pine Forge Press.

Chen, Stephanie. 2012. "Investigating Bus Route Walkability: Comparative Case Study in Orange County, California." Master's thesis, University of Southern California. Accessed March 3, 2013 http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll3/id/98075

CEPA (California Environmental Protection Agency). 2014. "Sustainable Communities." Accessed November 19, 2014. http://www.arb.ca.gov/cc/sb375/sb375.htm

Cervero, Robert and Michael Duncan. 2006. "Which reduces vehicle travel more: Jobs-housing balance or retail-housing mixing?" *Journal of the American Planning Association*. 72: 475–490.

Cervero, Robert and Jin Murakami. 2010. "Effects of built environments on vehicle miles traveled: Evidence from 370 urbanized areas." *Environment and Planning*. 42: 400-418.

Eisman, Daniel Currie. August 2012. "Spatial Analysis of Urban Built Environments and Vehicle Transit Behavior". Master's thesis, University of Southern California. Accessed February 1, 2013 http://spatial.usc.edu/wp-content/uploads/2014/04/EismanDanielThesis.pdf

EPA (Environmental Protection Agency). 1999. "Smog-Who Does it Hurt? What You Need to Know About Ozone and Your Health" Accessed February 17, 2014. http://www.epa.gov/airnow/health/smog.pdf

———, 2013. "Smart Location Database". Accessed January 15, 2014 http://www.epa.gov/smartgrowth/smartlocationdatabase.htm

Google Earth. Bus stop KML data. Google Inc. Accessed January 10, 2014.

Horeck, Tanya M. et al. 2012. "Sneakers and Spokes: An Assessment of the Walkability and Bikeability of U.S. Postsecondary Institutions." *Journal of Environmental Health.* 74: 8-15.

Jun, Myung-Jin. 2008. "Are Portland's Smart Growth Policies Related to Reduced Automobile Dependence?" *Journal of Planning Education and Research*. 28: 100-107.

Kim, Jinwon and David Brownstone. 2010. "The impact of residential density on vehicle usage and fuel consumption." University of California: Department of Economics. http://escholarship.org/uc/item/31m0w2x3.

Mantri, Anupama. 2008. "A GIS Based Approach to Measure Walkability of a Neighborhood." Master Thesis, University of Cincinnati. Accessed May 20, February 25, 2013 https://etd.ohiolink.edu/ap/10?0::NO:10:P10_ACCESSION_NUM:ucin1205012111

M.K. Gajendra Babu and K.A. Subramanian. 2013. *Alternative Transportation Fuels: Utilization in Combustion Engines.* Boca Raton, Florida: CRC Press

Prism Climate Group. 2014. Oregon State University. Climate raster data. Accessed October 18, 2014 http://www.prism.oregonstate.edu/normals/

Ramsey, Kevin and Alexander Bell. 2014. Smart Location Database: Version 20. User Guide. Accessed January 15, 2014 http://www.epa.gov/smartgrowth/pdf/sld_userguide.pdf

Reuters. 2010. "Chinese Divers sit it out in 9-day traffic jam." Accessed October 12, 2014. http://in.reuters.com/article/2010/08/23/idINIndia-51010420100823

Reuters. 2011. "Los Angeles braces for weekend of 'Carmageddon.'" Accessed October 12, 2014. http://www.reuters.com/article/2011/07/14/us-carmageddon-losangeles-idUSTRE76D2D720110714

Spoon, Steven Chadwick. 2005. "What Defines Walkability: Walking Behavior Correlates." Masters Project, University of North Carolina at Chapel Hill. Accessed November 15, 2013 https://cdr.lib.unc.edu/indexablecontent/uuid:384e8784-d3aa-4d8e-b15c-0f01083f5c16

Stevens, Robert David. 2005. "Walkability Around Neighborhood Parks: An Assessment of Four Parks in Springfield, Oregon." Master's Thesis, Graduate School of the University of Oregon. Accessed September 3, 2014 https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/1288/stevens_thesis.pdf?sequence=1

U.S. Bureau of the Census. 2014. *Metropolitan and Micropolitan Statistical Areas Main.* Accessed November 20, 2014 http://www.census.gov/population/metro/

USGS (United States Geologic Survey). 2014. National Elevation Dataset. Accessed March 8, 2014. http://ned.usgs.gov/

Villanueva, Karen et al. 2013. "People living in hilly residential areas in metropolitan Perth have less diabetes: spurious association or important environmental determinant?" *International Journal of Health Geographics* 12: 59.

Walk Score. 2014. Live Where You Love. Accessed November 23, 2014. www.walkscore.com

Appendix A

Figure 9 Map of MSA Locations

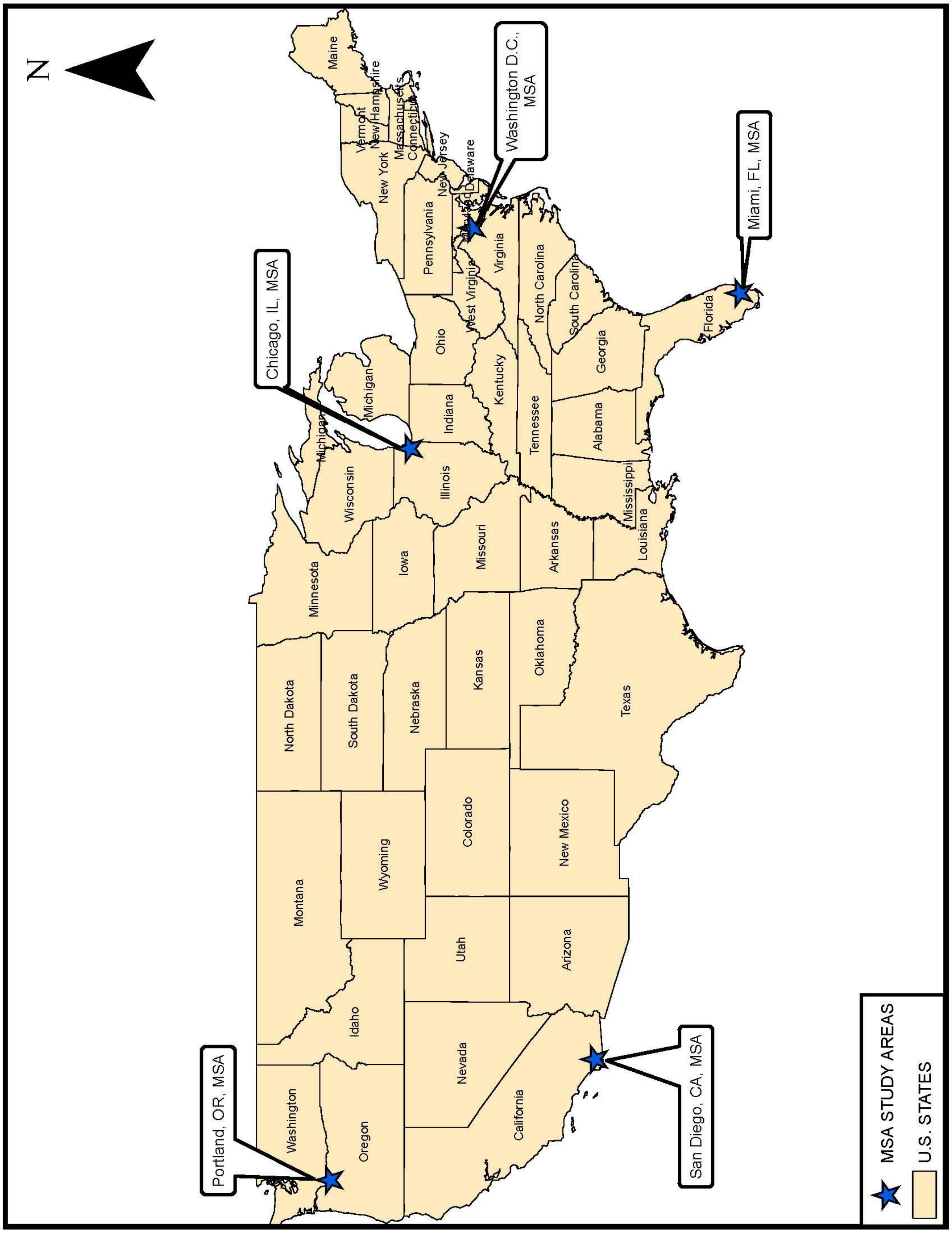


Figure 10 Chicago MSA and Samples

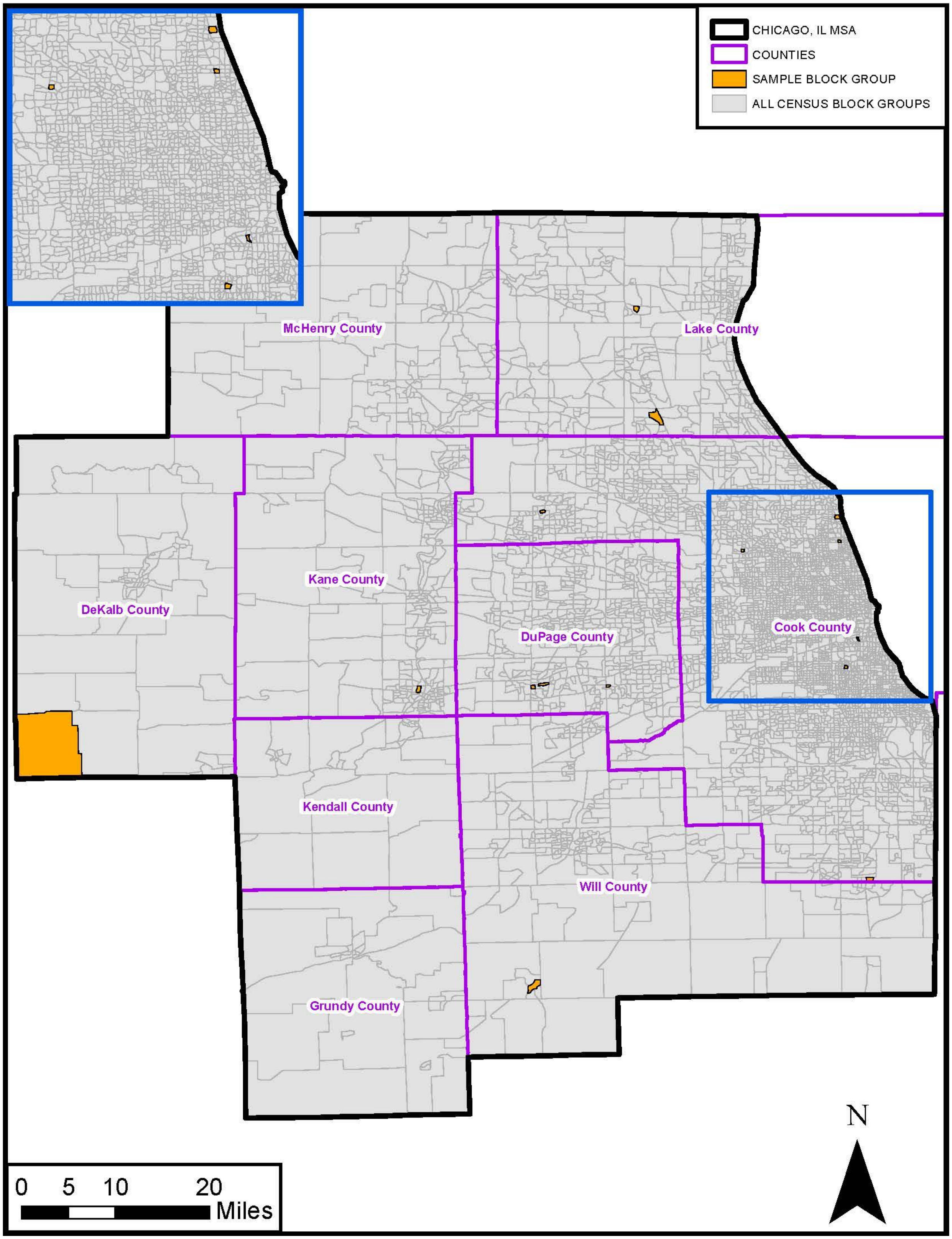


Figure 11 Miami MSA and Samples

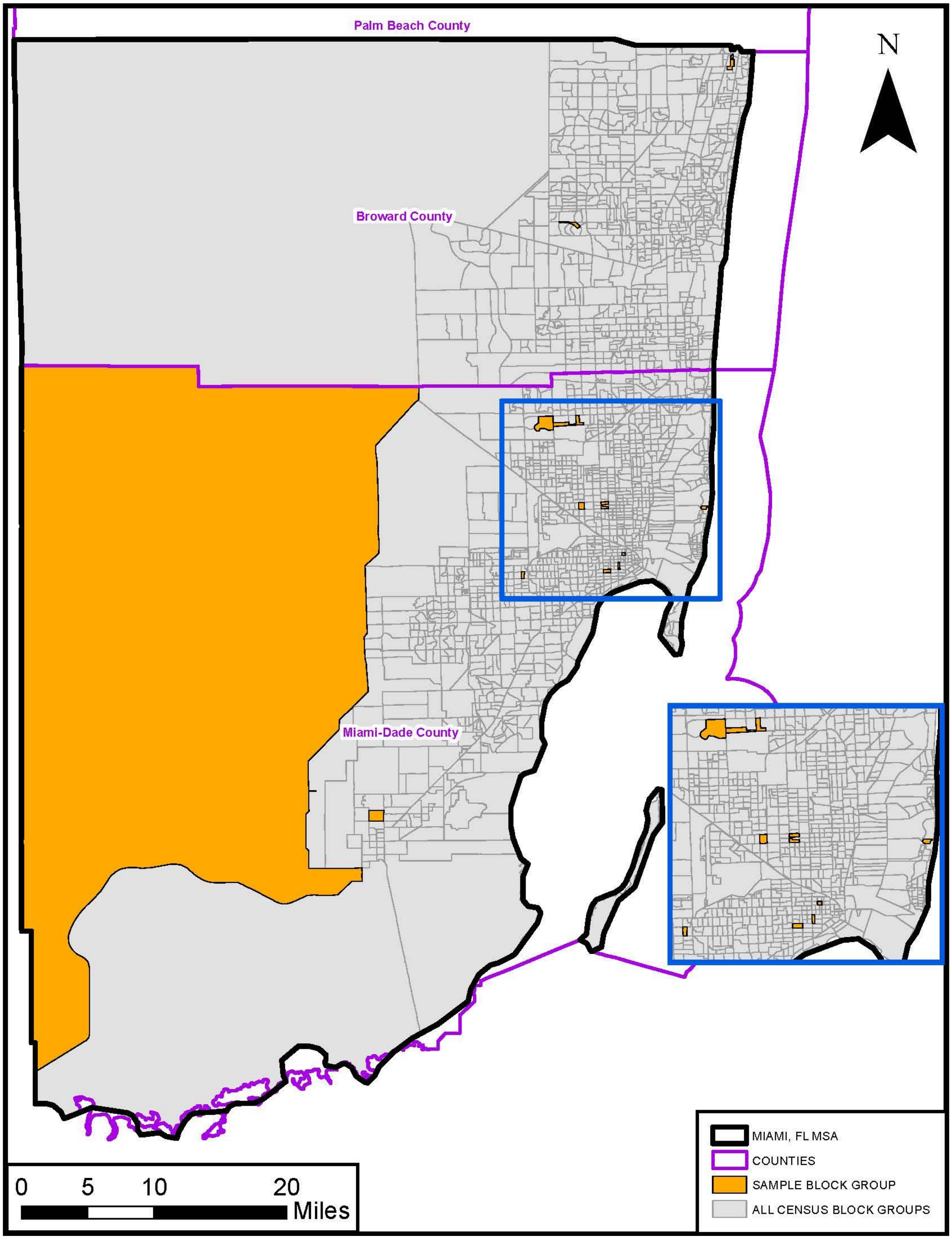


Figure 12 Portland MSA and Samples

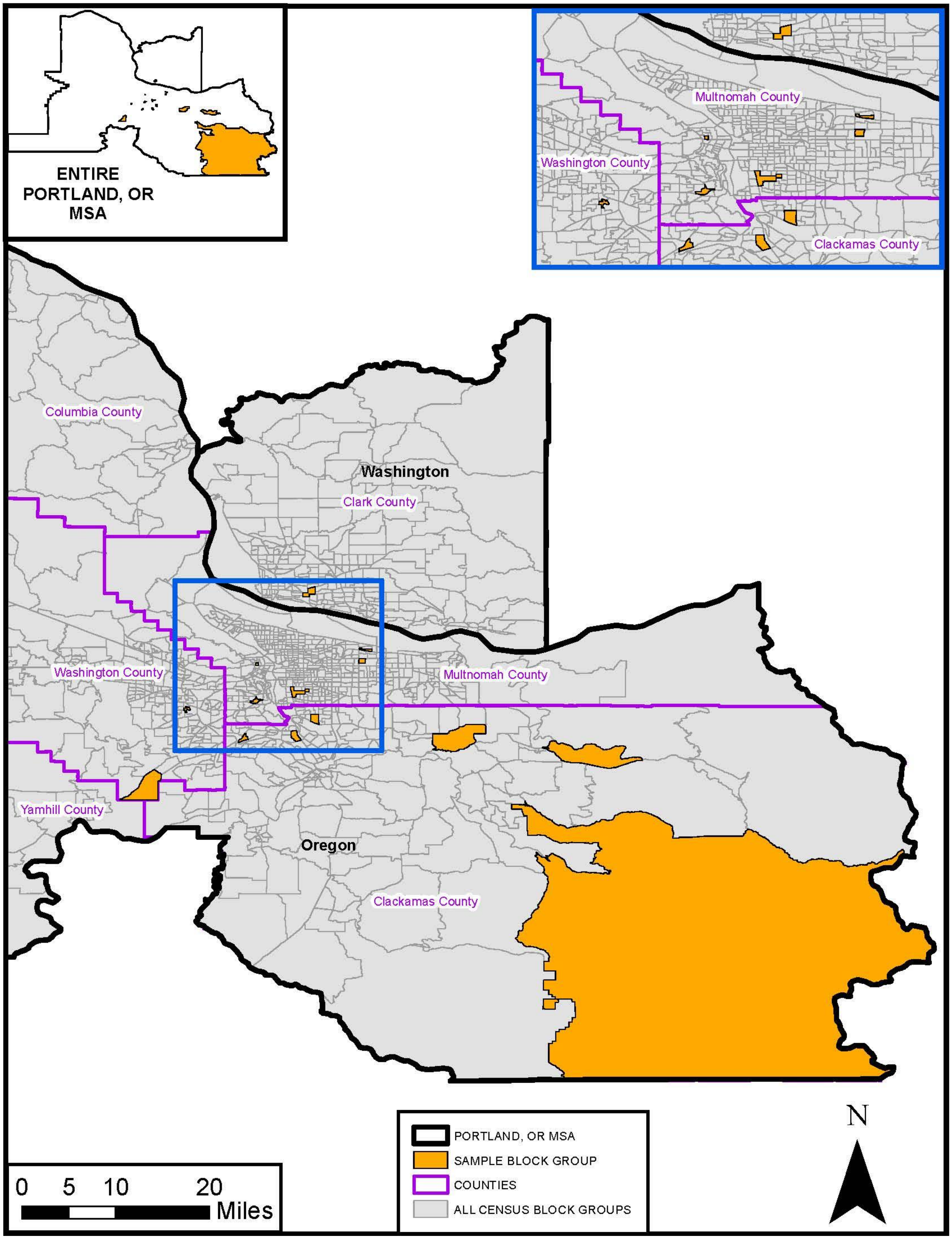


Figure 13 San Diego MSA and Samples

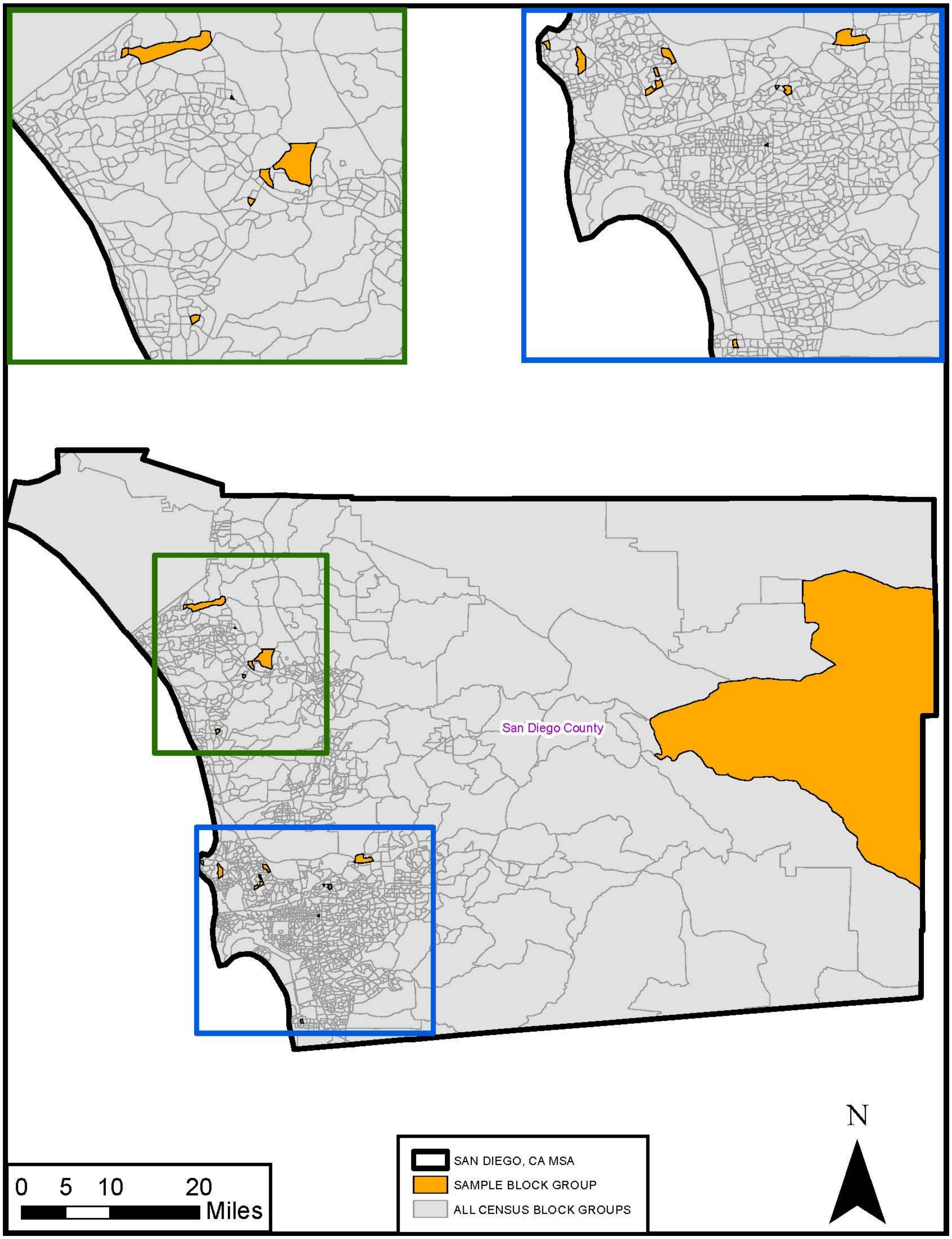


Figure 14 Washington D.C. MSA and Samples