Food Access, Race, and Poverty: Associations of Neighborhood Characteristics and Grocery Access in Portland, Oregon

---

A Thesis
Presented to
The Established Interdisciplinary Committee for Mathematics and Economics
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Joan J. Wang

May 2014

Approved for the Committee
(Mathematics and Economics)

Jeffrey A. Parker

Albert Y. Kim

Acknowledgements

Jeff, thank you for the weekly dose of grammar jokes and econometrics.

Noelwah, thank you for introducing me to the side of economics I believe in.

Albert, thank you for reminding me that correlation does not imply causation.

Rebecca, thank you for agreeing to be my outside reader.

Kristin, thank you for making me answer my own questions.

Joe, thank you for putting up with my requests for good data and restaurants.

Tyler, David, and Steve, thank you for reminding me how awesome Portland is.

Pierre-Louis, thank you for your USB, and being French.

Ruben, thank you for making this thesis so hard to finish.

Taylor and Phoebe, thank you for all those breakfast days.

Jenn, thank you for moping while making art with me.

Shruti, thank you for maintaining our friendship beyond the thesis desk.

Sunny, thank you for never judging me, and eating all those prunes with me.

Sean, thank you for never giving up on me.

Mom and dad, thank you for loving me despite my spasmodic skype calls.

Finally, I want to thank everyone I have had a chance to interact with during this past year, as well as those who have stuck with me throughout my time at Reed.

Contents

List of Tables

List of Figures

Abstract

This thesis uses GIS and network analysis to construct various spatial measures of access to grocery stores in Portland, Oregon, in 2000 and 2010. Then, it investigates the associations of a census tract's social, economic, and spatial characteristics with its access to grocery stores by multivariate regression analysis. Results show that the interaction between race and poverty is significant in characterizing the associations with access to groceries. In addition, a higher proportion of African American residents is associated with lower access to large groceries while more poverty is associated with higher access to large groceries.

Dedication

For Mom, Dad, and Sean.

Introduction

This thesis examines the effects of neighborhood characteristics on food access in Portland, Oregon, in 2000 and 2010. Within the growing literature on food security, many studies have relied on neighborhood characteristics combined with physical characteristics of the local food environment to identify areas of insufficient food access (Apparicio, Cloutier, and Shearmur 2007; Eckert and Shetty 2011; Larsen and Gilliland 2008; Leete, Bania, and Sparks-Ibanga 2012; Smoyer-Tomic, Spence, and Amrhein 2006; Sparks 2008). Many have taken one step further to conduct econometric models to parse out the magnitudes of the neighborhood factors that impact food access (Black et al. 2011; Dai and Wang 2011; Franco et al. 2008; Hall 1983; Krukowski et al. 2010; Moore and Diez Roux 2006; Morland et al. 2002; Paez et al. 2010; Powell et al. 2007; Raja, Ma, and Yadav 2008; Sharkey et al. 2009; Small and McDermott 2006; Smoyer-Tomic et al. 2008; Svastisalee et al. 2011; Wrigley, Warm, and Margetts 2003; Zenk et al. 2005). However, there has not been a study of food access done for Portland, Oregon that employs regression analysis despite the growing concerns of food access issues in the city. Effective food access policies necessitate careful examination of how different demographic and socioeconomic attributes play a role in the formation of local food environments. Using Portland, Oregon as a case study, this thesis constructs measures of food access at the census tract level using ArcGIS software. Then it analyzes how these measures are affected by changes in different census tract characteristics with regression models performed in Stata.

Concerns with food security have led to examinations of food deserts from a diverse range of fields. The term "food desert" originated from the United Kingdom in the mid 1990s to describe dense urban areas with insufficient access to healthful and affordable food (Wrigley, Warm, and Margetts 2003). Since then, many studies have attempted to define food deserts in various regional contexts: Apparicio, Cloutier, and Shearmur (2007) in Montreal, Leete, Bania, and Sparks-Ibanga (2012) in Portland, etc. Although there is no one consistent standard method of identification, most food deserts are found to be areas of inadequate access to healthy food, correlated with concentrated socioeconomic vulnerability.

In the United States, the growing focus on food deserts stems from a national obesity problem, with policy solutions from the United States Department of Agriculture (USDA), the President's budget, and the 2012 Farm Bill. The most concrete example is Michelle Obama's National Healthy Food Financing Initiative, part of which involves helping "national and regional chains to open or expand more than 1500 retail stores in an effort to bring healthier food to underserved areas" (Donald 2013, 232). Parallel to the national political attention given to food deserts, academic research has examined food security from a public health angle among many other social science approaches, using food access as a determinant of health conditions (Black and Macinko 2008; Laraia et al. 2004; Wrigley, Warm, and Margetts 2003). One segment of the literature seeks to explain how neighborhood characteristics influence the local food environment, which affects the health conditions of residents. Particularly, food access is used to connect neighborhood characteristics to health conditions (Powell et al. 2007).

Food access is inherently an economic issue as well. Basic economics informs us that the interaction of supply and demand determines the availability of products, the market where the products exist, and at what price they are available (Bitler and Haider 2011). In the context of food access, several economic aspects dictate how food access is characterized. First, the "product," healthy and affordable food, can take on many different definitions. Secondly, demand for such product is essentially determined by income, prices, and preferences, with preferences being the hardest to quantify and analyze (Bitler and Haider 2011). On the other hand, supply depends on input costs including labor, land, capital infrastructure, transportation, etc., which are largely determined by the demographic, economic, and environmental characteristics at the local level.

Finally, as McClintock (2011, 91) points out, "[few] studies move beyond a geospatial or statistical inventory of food deserts to unearth [how the] historical processes" of transforming the urban landscape and demography relate to food access. Indeed, the consideration of historical and/or current trends such as racial segregation and the urban-to-suburban migration over time can deepen the understanding of neighborhood food access. This thesis draws on sociological and economic theories to shed light on the mechanisms through which neighborhood conditions, specifically race and poverty, lead to changes in the spatial access to food over time.

This thesis is motivated by all four aforementioned segments of discussions on food access: identification of food deserts, food access as an indicator of health outcomes, food access characterized by supply and demand, and food access through structural changes of neighborhoods over time. By examining the effects of neighborhood traits on the geographic patterns of stores at a local level between different time periods, this thesis characterizes the demand and supply factors that shape the market for healthy and affordable food. This contributes to the health literature that seeks to draw a clear connection between neighborhood traits and health conditions via food access. Last but not least, this study helps understand whether the identification of food deserts should be contingent upon neighborhood socioeconomic vulnerability by showing how much these neighborhood traits are associated with food access.

This thesis is organized as follows. Chapter One consists of a literature review on measuring food access, factors that affect food access, and regression analyses on food access. Chapter Two provides background information on food access in Portland. Chapter Three explains the data and variables used in the analysis. Chapter Four explores different econometric models and their results. The implications of this study are discussed in the Conclusion.

Chapter 1

Literature Review

This study draws from two bodies of literature: how food access is measured and how food access is explained. Following a brief introduction of some basic concepts of food access, I first explore different ways of measuring food access by comparing their assumptions, advantages, and limitations. I then propose a set of hypotheses of how neighborhood conditions characterize food access by discussing several sociological and economic theories. Finally, I examine how a few past studies have employed regression analysis to study food access.

1.1 Food Access: Background and Assumptions

I use a legislative definition of "food desert" to launch the discussion on the background and assumptions of food access. In Title VI, Sec. 7527 of the 2008 Farm Bill, a food desert in the United States is defined as an area "with limited access to affordable and nutritious food, particularly such an area composed of predominantly lower income neighborhoods and communities" (United States, Department of Agriculture, and Economic Research Service 2009). As mentioned in this definition, the issue of food access extends beyond the physical proximity of grocery stores; other important factors are at play. Food access is the ability to obtain healthful affordable food. It should be evaluated based on proximity, selection, affordability, and awareness. Proximity concerns people's physical ability to reach grocery stores, which is contingent upon geographical proximity to a grocery store and availability of transport options. Selection is the requirement that people are presented with healthy food choices that necessarily include fresh produce and meat at grocery stores. Affordability identifies a key barrier to food access as the lack of expenditure allocated for groceries. The awareness aspect of food access is concerned with the effects of misinformation about healthful food as well as time and skill constraints to prepare healthy food that inhibit access. In this section, I discuss each of these evaluators of food access, and summarize with a list of crucial assumptions made in this thesis.

1.1.1 Proximity

Research on food access has focused on identifying inadequate food access (often called food deserts) in a geographical unit such as a census tract, a block group, or some other pre-determined small neighborhood. (For the definitions of census tracts and census block groups, please refer to Section 1.2.) The fundamental assumption is that people shop for food that is physically close to them, based on the fact that every analysis looks at food access in a given small-scale geographical boundary. This assumption is derived from Christaller's central place theory that food is a good with a low range, "the furthest distance that consumers are willing to travel to buy the good" (Church and Murray 2008, 6). Paired with the notion of threshold, "the distance from a [store] at which the demand for the good is large enough to satisfy the requirements for a vender to remain in business," a store location is deemed profitable if its range exceeds its threshold (6). When stores locate, they enter the market and compete until the economic profit is zero, reaching long-run equilibrium. A direct result is that grocery stores, selling mostly low-ordered goods, should be abundant and widely distributed. Thus examining food access via grocery access from a small geographical scale is justified.

On the other hand, there may be cases that show that low-income residents with close access to full-service grocery stores travel far for affordable quality food (Coalition for a Livable Future 2007). To these residents, after juggling between travel costs and food costs, food access is better from far away than what is physically near. This could contradict Christaller's assumption that "each customer would travel to the closest [grocery store] for the good of interest," depending on how "the good of interest" is defined (Church and Murray 2008, 6). If, given budget constraints, these customers identify their good of interest to be affordable food, then traveling further has not violated Christaller's assumptions, since the good of interest is necessarily further away in this case. Studies that take into account travel time to get to a grocery store may help parse out the elasticity between travel costs and food costs.

In this thesis, I take Christaller's theory as given, and assume that the proximity aspect of food access, measured by the spatial proximity to grocery stores, can be evaluated on a small geographical scale such as census tracts.

1.1.2 Other Evaluators

When identifying food deserts in neighborhoods, some only look at the characteristics of grocery stores in the geographical unit of analysis, while most others rely on a combination of grocery store traits as well as demographic information such as socioeconomic status. With the latter approach, the crucial assumption is that demographic characteristics provide additional explanatory power on top of grocery-store characteristics to derive meaningful evaluation of food access.

Awareness

One evaluator of food access other than proximity is awareness. This aspect consists of the effects of misinformation about healthful food as well as time and skill constraints to prepare healthful food that inhibit access (*Portland Plan Background Reports* 2014). Awareness partly explains the demand for healthful food. Explicitly observing awareness is very difficult due to data collection limitations, partly thanks to its individualized nature.

However, awareness could be correlated with certain demographic traits that are more easily observable. For example, the level of education is a good proxy for the level of knowledge on healthful food and nutrition, which is key to having adequate food access. In this thesis, since food access is quantified solely by the proximity aspect of grocery access, the potential inclusion of education as an explanatory variable assumes that some of the awareness aspect of food access is associated with the proximity aspect of food access.

Selection and Affordability

When one examines data on non-spatial store-specific characteristics, there exist two basic types of information: product availability and prices. The information on product availability determines selection. And while the prices of products paint half of the picture for affordability, some information on income of the households in the neighborhood is necessary to deduce whether residents can afford food offered near them. A neighborhood of only high-income residents can afford grocery stores with higher prices, thus affordability is not an issue even with the high prices.

Due to data constraints, this thesis does not examine the product availability or prices of grocery stores. And as it does with awareness, this thesis assumes that income as an explanatory variable implies that the affordability aspect of food access is associated with the proximity measure of food access.

1.1.3 Assumptions

To understand how demographic characteristics play a role in determining food access, whether through awareness or affordability or some other channel, one needs to look at how each demographic trait affects food access while controlling for all other factors. Here the dependent variables are measures of spatial access to grocery stores that are directly observable (as proxies for food access): the density of stores, the presence/absence of stores, some rating of the stores present, distance to the nearest store, etc. The independent variables are demographic characteristics that are hypothesized to affect food access, such as income, age, education, race, population density as a measure of urbanization, public/private transport options, and crime. The effects of these variables on store access are explored in this thesis. I list a few main assumptions taken:

1. Grocery stores are the primary places people use to obtain food, as opposed to alternative means such as farmer's markets and urban gardens.
2. The proximity aspect of food access, illustrated by the spatial access to grocery access, can be evaluated on a small geographical scale such as census tracts.
3. With the first two assumptions, the spatial access to grocery stores on a small geographical scale can be a proxy for food access.
4. The level of education is a proxy for the level of knowledge on healthful food and nutrition, and is also a potential explanatory variable for spatial access to stores.
5. Household income is a proxy for affordability, and can also act as an explanatory variable for spatial access to stores.

1.2 Food Access: Measurements

1.2.1 Store Types and Categories

Before discussing the different possibilities for the dependent variable, the categorization of different types of grocery stores is first required. Store types are generally defined by conventional standards or industry codes. For example, food markets versus supermarkets are different types of grocery stores. On the other hand, the categorization of stores can take many forms. Grocery stores can be categorized between chain and non-chain, by employee size, by sales volume, etc. The types and categories of grocery stores are important in contextualizing the environment within which food access is measured. For instance, the nature of a large chain supermarket is fundamentally different from a small independent mom-and-pop grocery store. A large supercenter could be the only grocery store option in a region, with adequate selection and affordability but lacks the close physical proximity to residents. On the other hand, a dispersion of small non-chain grocery stores provides the neighborhood convenience but perhaps sacrifices selection and affordability for close proximity. Past researchers have employed different ways to include variability in grocery store types in their analysis. Some rely on past findings to rank supermarkets higher than smaller stores and chains higher than non-chain supermarkets on the likelihood that they will provide healthful and affordable foods (Powell et al. 2007). Others choose to only look at a subset of grocery stores that are most likely to have adequate selection, and in some cases affordable prices are taken into consideration as well. For example, one paper used supermarkets as the grocery stores studied, which are defined as either supercenters or "full-line grocery stores ... associated with a national or regional chain" (Zenk et al. 2005, 661). Most studies related to grocery access only use full-service grocery stores as the requirement, each adopting a different definition of "full service." This method limits how food access is determined, and it also leaves a large uncertainty as to what full service entails. To sum up, there exist a variety of types and categories of grocery stores that have been adopted by past research. I determine the most adequate method based on findings from past researchers given data available in Portland, Oregon.

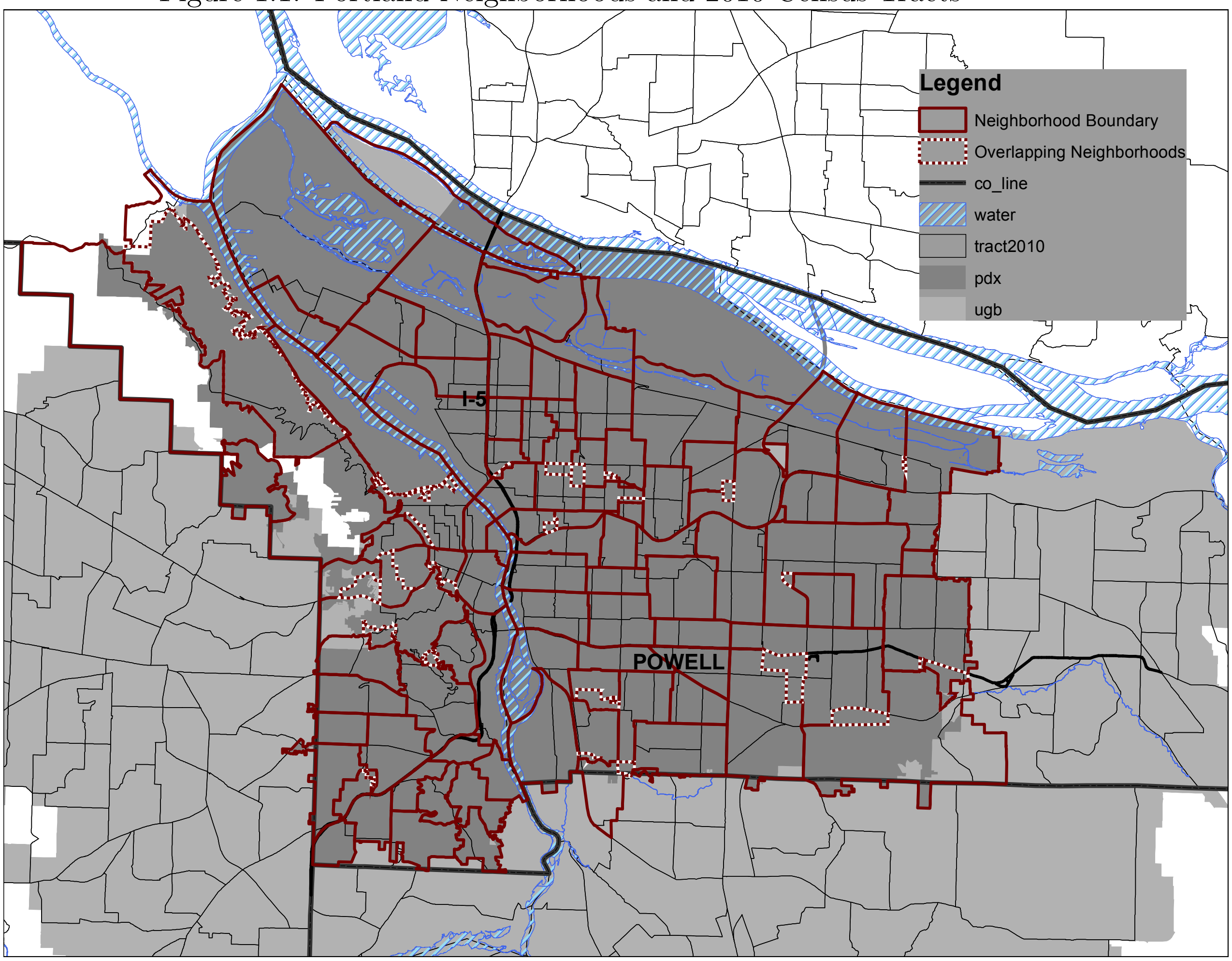
1.2.2 Geographic Units

As mentioned in Section 1.1.1, analyses of food access are based on small geographic units of analysis. The most common geographic units in research done in the United States are census tracts, block groups, and blocks, all defined by the U.S. Census Bureau. "Census tracts are small, relatively permanent statistical subdivisions of a county ... [that] generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people [, and] usually covers a contiguous area" (U.S. Census Bureau 2014b). "Block Groups are statistical divisions of census tracts, [and] are generally defined to contain between 600 and 3,000 people" whereas blocks are the smallest geographic unit that constitue block groups (U.S. Census Bureau 2014a).

The boundaries for these three geographic units may adjust for every decennial census, depending on population changes. Since the boundaries of census tracts are defined by population size, the areas of census tracts can vary drastically. High density leads to census tracts with very small areas. The same concept applies to block groups and blocks.

Figure 1.1 is a map comparing Portland Neighborhoods and census tracts defined by the U.S. Census Bureau in 2010. The thicker lines demarcate the Neighborhoods while the thinner lines represent the boundaries of census tracts. Neighborhoods and census tracts have roughly the same geographic precision, thus studying access to stores from census tracts can lead to substantial implications for food access at the neighborhood level.

Figure 1.1: Portland Neighborhoods and 2010 Census Tracts



1.2.3 Sparks, Bania, and Leete (2011)

Sparks, Bania, and Leete (2011) conducted background research on the differences of various food-access measures. Note that their research is only conducted on grocery stores and supermarkets that belong to chains. This study compiled an extensive list of methods of measuring food access from those who seek to identify food deserts explicitly as well as those who use food-access measures to explain local food environments. Then, it sampled and tested four different measures (M), three spatial aggregation levels (A), and two distance constructs (D), using Portland as its case study. I introduce their results in three tables, each accompanied by a brief explanation and summary.

Four Measures of Food Access (M)

Table 1.1: Comparison of four different measures of food access (M)

Name	Description	Type of Measure	Purpose
M1	Number of supermarkets within 1 km	Coverage; discrete count	Variety and competition
M2	Distance to nearest supermarket (meters)	Distance; continuous	Proximity
M3	Average distance to three closest supermarkets, different parent companies (meters)	Distance; continuous	Variety and competition
M4	Average distance to three closest supermarkets (meters)	Distance; continuous	Variety and competition

Table 1.1 shows the four measures of food access that are compared by Sparks, Bania, and Leete (2011). Every measure is calculated using the centroid of a census block as the starting point. M1, the only discrete variable, counts the number of supermarkets within a circle of a certain distance from the centroid. Under the assumption that identical supermarkets would not locate close to one another, the number captured by M1 can be an estimate of the variety of supermarkets of each census block. M2 to M4 are all continuous distance measures. M3 and M4 are modifications of M2, which estimates each census block's proximity to a supermarket.

I summarize the important findings from Sparks, Bania, and Leete (2011) in the bullet points below:

- All four measures are highly correlated with one another by Pearson and Spearman correlation coefficients: M1 is less so; M3 and M4 are particularly correlated, which leads to the second bullet point.

- M3 and M4 are redundant: "[accounting] for competition within or between parent companies has little effect on the competitive environment in terms of access to three different stores" (Sparks, Bania, and Leete 2011, 1723).

- Moderate and statistically significant spatial correlation of all four measures by Moran's Index of global spatial autocorrelation (Moran's I): census tracts that cluster together are likely to have similar levels of access. (Moran's I is a statistic of spatial autocorrelation that takes the value -1 if there exists perfect dispersion, and 1 if there exists perfect clustering. If there exists no spatial autocorrelation as in a random arrangement, Moran's I would take the value of 0.)

Three Aggregation Methods (A)

Table 1.2 compares three different aggregation methods of geographical units. All three methods produce measures of food access for a given census tract. However, the difference between them resides in the level of geographical detail with the measurement. A1 and A2 measure food access for a given block and block group, respectively. Then, they are aggregated to the census tract level using a population-weighting formula. Note that none of the measures of food access is restricted by the scale of the geographical unit of measurement. For example, the number of supermarkets within one kilometer of a block group (M1) should not be confused with the number of supermarkets within the area of a block group. Thus, M1 measured at the block group level, and then population-weighted to the tract level (A1), is directly comparable to M1 measured at the tract level (A3).

Table 1.2: Comparison of 3 different aggregation methods (A)

Name	**Distance from supermarket to:**	**Aggregation level**
A1	Block centroids	Tract (population weighted)
A2	Block group centroids	Tract (population weighted)
A3	Tract centroids	Tract

I summarize the important findings from Sparks, Bania, and Leete (2011) in the bullet points below:

- M1 decreases as the aggregation method moves from A1 to A3. Accordingly, the distance measures (M2, M3, and M4) increase moving from A1 to A3. Note that all aggregation methods produce measures at the census tract level. A possible explanation for the increase in distance measures from A1 to A3 is that A1 takes the weighted average of multiple distances (from the centroids of blocks) to characterize each census tract whereas A3 only has one measurement for each census tract. So for A1, it is more likely that some of the distances from the blocks are shorter than others, pulling the weighted-average distance down. The same logic could be applied to M1.

- "Less geographical detail is also associated with higher variability of access measurement" as evident in the increase in standard deviation (1726). As explored in the possible explanation for the preceding result, A1 has a smoothing effect compared to A3 because it is an average. Thus it naturally makes sense that the variability of measures is lower for A1 versus A3.

- M1, as an absolute measure of food access, is more sensitive to differences in aggregation method than M2, M3, and M4: aggregation method used for M1 might impact the food-access patterns observed.

- For the relative measures of food access (M2, M3, and M4), A1 to A3 all yield similar relative patterns of food access.

Two Methods of Distance Construct (D)

Table 1.3 compares two common distance construct methods used to create the food-access measures. The Euclidean method calculates the straight line distance between points a and b using the formula $d_{ab} = \sqrt{(a_i - b_i)^2 + (a_j - b_j)^2}$, where i and j are the x and y coordinates that describe points on the earth's surface. The network method connects points a and b using the lines in a given network, then sums up the distances of those lines. An example of networks is a street network, which constitutes all of the streets in a given region. The network method using a street network more accurately models the distance it takes to travel, which is usually significantly longer than the Euclidean distance, the distance calculated by the Euclidean method. Nevertheless, the Euclidean method is computationally simpler and might be a good way to compare distances relatively.

I summarize the important findings in the bullet points below:

- As expected, the distance measures (M2, M3, M4) increase by a large amount (about 35%) from when distance construct changes from D1 to D2; and M1 decreases accordingly.

Table 1.3: Comparison of 2 different methods of distance construct (D)

Name	**Distance from supermarket to:**	**Aggregation level**	**Distance construct**
D1	Tract centroids	Tract	Euclidean
D2	Tract centroids	Tract	Network

- For M2, M3, and M4, there exist a high rank correlation between measures computed with D1 and D2, which demonstrates that measures using D1 and D2 are highly correlated and exhibit similar relative patterns.

- M1 is more sensitive to changes between D1 and D2, which might be due to the lack of variation of M1.

1.2.4 Leete, Bania, and Sparks-Ibanga (2012)

Leete, Bania, and Sparks-Ibanga (2012) builds on findings from Sparks, Bania, and Leete (2011) by applying different food-access measures to analyze various methods of food desert identification adopted by past researchers. This paper applies three different methods, including both distance and coverage measures, for defining food deserts in the Portland Urban Growth Boundary (UGB). It concludes that the resulting food deserts are comparable, although none of the methods is sufficient for food desert identification on its own.

There is a variety of methods to measure food access used by past researchers aiming specifically to identify food deserts. They generally construct food-access measures, then produce food desert identification by combining food-access measures and demographic characteristics of a geographic unit of analysis. Through exploring the different ways to identify food deserts, Leete, Bania, and Sparks-Ibanga (2012) have summarized three common methods to measure food access. The authors calculate a measure of proximity by distance to nearest supermarket, a measure of competition and variety by number of stores in a circle, and another measure of competition and variety by mean distance to nearest three stores of different chains.

I present two tables that summarize methods of measuring food access from those who seek to explicitly identify food deserts as well as those who use food-access measures to explain local food environments. The tables present studies done in the United States (Table 1.4) and elsewhere (Table 1.5), respectively.

Table 1.4: Summary of studies measuring grocery access in the US

Reference	Study Area	Store definition	Access measures	Unit of analysis	Distance construct
Alwitt and Donley (1997)	Chicago, Illinois	All retail outlets (including grocery, restaurants, apparel)	Number of retail outlets (by type and size) per zip code	Zip codes	Euclidean
Zenk et al. (2005)	metropolitan Detroit, MI	Supercenters and full-line grocery stores of a chain with 11 or more stores	Distance to the nearest store from the centroid	Census tract	Manhattan block
Block and Kouba (2006)	Austin and Oak Park, Illinois	Retail outlets that carry food (including supermarkets, convenience stores, drug stores, liquor stores) from InfoUSA	Percentage of population within one-quarter mile of store	Not reported	Euclidean
Moore et al. (2006)	selected samples from counties in NC, MD, and NY from a large multiethnic study of atherosclerosis	additional categorization based on SIC codes	Ratio of store per 10000 population	Census tract	N/A
Powell et al. (2006)	nation-wide	categorization using SIC codes: chain supermarkets, non-chain supermarkets, grocery stores, and convenience stores	Number of stores	Zip codes	N/A
Franco et al. (2008)	Baltimore City and Baltimore County, MD	additional categorization based on SIC codes (taken from Moore et al. (2006)	an ordinal variable HFAI that assesses the availability of eight food groups for each store, then the mean of all stores in a census tract is calculated	Census tract	N/A
Raja et al. (2008)	Erie County, Pennsylvania	All food destinations including specialty shops and restaurants by SIC	Number of food destinations per 10,000 people; per block group; within 5-minute trip of block group centroid by walking, bike, or driving	Census block groups	Street network
Fan et al. (2009)	Salt Lake County, Utah	Grocery stores carrying fresh produce and/or meats	Presence of stores within 1000 metres	Census block group	Euclidean
Neckerman et al. (2009)	New York City	Supermarkets, fruit and vegetable markets, and farmer's markets	Kernel density measure of food outlets within one-half mile of population-weighted census tract centroid	Census tract	Euclidean
Krukowski et al. (2011)	selected samples within 45 minutes of Little Rock, AR and Burlington, VT	stores with NAICS codes divided between supermarkets as large, corporate-owned chain stores and grocery stores as smaller, local, not corporate-owned food stores	store size by the number of cash registers or NEMS-S, a standardized observational measure of availability, price, and quality of foods	Census tract	N/A

Table 1.5: Summary of studies measuring grocery access outside the US

Reference	Study Area	Store definition	Access measures	Unit of analysis	Distance construct
Clarke et al. (2002)	Leeds/Bradford and Cardiff, UK	Municipal surveys of grocery retailers	Presence of stores within 500 metres	Postal sector	Euclidean
Smoyer-Tomic et al. (2006)	Edmonton, Alberta, Canada	Grocery stores carrying a full range of items	Number of stores within 1000 metres; distance to nearest supermarket from postal code centroids	Population-weighted neighborhood	Street network
Apparicio et al. (2007)	Island of Montreal, Quebec, Canada	supermarkets associated with one of seven major chains in Quebec	Distance to the closest supermarket; number of supermarkets within a walkable distance of 1000 metres; mean distance to three supermarkets of different chains	population-weighted average of census block measures calculated for analysis at the census tract level	Street network
Larsen and Gilliland (2008)	London, Ontario, Canada	locally identified grocery stores without further description	Number of stores within 1000 metres; distance to nearest grocery store; percentage of population with grocery stores within 1000 metres walking distance; percentage of population with bus access to grocery store (10-minute direct bus ride plus maximum 500-metre walk at either end of trip)	population-weighted average of census block measures calculated for analysis at the census tract level	Street network
Ball et al. (2009)	randomly selected neighborhoods, Melbourne, Australia	Fruit and vegetable grocery stores and major chain supermarkets	Number of stores within 2000 metres; distance from resident's home to store; mean number of stores in neighborhood per 10,000 residents	Neighborhood	Street network

1.3 Effects of Neighborhood Traits: Hypotheses

With an established correlation between food access and health conditions of a neighborhood (Black and Macinko 2008; Laraia et al. 2004; Wrigley, Warm, and Margetts 2003), a hypothesis is that neighborhood conditions affect health through access to healthful foods. Studies have generally found a significant correlation between lack of physical proximity to grocery stores and socioeconomic vulnerability, characterized by both low-income and minority communities (Baker et al. 2006; Franco et al. 2008; Moore and Diez Roux 2006; Morland et al. 2002; Powell et al. 2007).

Although this thesis attempts to examine all neighborhood traits that are associated with food access, it follows the aforementioned studies in concentrating on answering this question: How do race and poverty correlate with access to groceries?

In this section, I present a brief overview of the conclusions of past studies regarding this question. Then, I rely on Small and McDermott (2006) to discuss the views of Park, Burgess, and McKenzie (1984), Shaw (1969), and most primarily Wilson (1987) in shaping some hypotheses that help answer this question.

Theoretical considerations of other neighborhood traits that are associated with food access are laid out in Section 3.2, where the independent variables used in this thesis are introduced.

1.3.1 Race and Poverty: Conflicting Views

As mentioned before, there is a sizable number of studies that relate high poverty and minority communities to insufficient food access. However, some found opposing effects, notably that low-income neighborhoods are associated with better access (Black et al. 2011; Dai and Wang 2011; Paez et al. 2010; Small and McDermott 2006; Pearce, Day, and Witten 2008). A few studies argue that poverty insufficient food access are not correlated (Eckert and Shetty 2011; Svastisalee et al. 2011).

In terms of the African American composition, some break down the results by asserting that only in poor neighborhoods does African American composition negatively associate with food access (Zenk et al. 2005). And others suggest that the negative association is only with access to supermarkets, and that African American composition is positively correlated with access to small groceries (Moore and Diez Roux 2006; Raja, Ma, and Yadav 2008). And still a few assert that the African American composition is not associated with food access (Eckert and Shetty 2011; Krukowski et al. 2010).

These conflicting results suggest that race and poverty need to be examined in relation with each other in the issue of food access. They also suggest that changes in the relationship between these two factors can be associated with different changes in food access, depending on other considerations such as how food access is measured.

In the next section, I summarize how Small and McDermott (2006) draw from theories of prominent sociologists that analyze ecological and economic factors in urban societies. Then, I list their hypotheses on how trends in race and poverty across time and space relate to access to groceries. The regression analyses that this thesis conducts will provide concrete results that correspond to these hypotheses.

1.3.2 Race and Poverty: Theories

Research by Small and McDermott (2006) is one of a few in this body of literature that provides a theoretical framework while attempting to quantify the correlation between poverty and race and access to groceries, as it focuses on understanding the relationships between neighborhood conditions and neighborhood resources in general.

The study begins by introducing Park, Burgess, and McKenzie (1984) of the Chicago urban school and their view that "cities [are] fundamentally market-driven entities in which competition among businesses and groups resulted in natural areas – specifically, a set of increasingly affluent, concentric circles that radiated outwardly from a central business district" (Small and McDermott 2006). Later, the social disorganization theory proposed by Shaw (1969) posited that "[poor], ethnically heterogeneous, residentially unstable neighborhoods were unable to sustain businesses and organizations because they lacked economic stability and social organization" (Small and McDermott 2006). Finally, Small and McDermott (2006) brings in the argument of Wilson (1987) that "the de-institutionalization of neighborhoods is the product of middle-class flight and the ensuing concentration of poverty" because the middle class is pivotal in sustaining the economic viability of neighborhoods (Small and McDermott 2006). According to Wilson's argument, as "black working-class families ... leave the inner cities in large numbers, [inner-city ghettos]" have become poorer. And as inner-city ghettos have historically high concentrations of African American residents, "concentration of poverty ... has increased sharply" Greenstein (1987).

To apply the three theories to grocery access, the notion of urban-suburban migration must be addressed. The migration of residents from urban to suburban parts of a city is a central mechanism through which neighborhood demographics change, and these changes are part of the cause of insufficient access to neighborhood resources such as groceries. More specifically, these three theories convey the following:

1. The migration of economically able residents from urban to suburban is a natural occurrence in cities.
2. Those that are not able to migrate due to lower socioeconomic status are left in neighborhoods with insufficient resources.
3. Given the first two points, racial segregation helps explain the emergence of neighborhoods that are both racially heterogeneous and economically unstable, which is correlated with the lack of neighborhood resources such as groceries.
4. The interaction between race and poverty is strong and negative in their correlation with neighborhood resources such as groceries.

Based on the above arguments, Small and McDermott (2006) lays out several hypotheses on how race and poverty correlate with food access: (I have made some minor wording changes, and I have substituted "access to groceries" for "the number of establishments".)

1. Poverty — Access to groceries in the neighborhood will decrease as the poverty level increases, other factors held constant.
2. Race — Other factors held constant, as the proportion of residents who are African American increases, access to groceries in the neighborhood will decrease.
3. Interaction of poverty and race — Other factors held constant, as the proportion of residents who are African American increases, the negative association between neighborhood poverty and access to groceries will increase.

1.4 Effects of Neighborhood Traits: Models

The mixed results of effects of neighborhood traits on food access (as discussed in Section 1.3) signify the importance of accounting for differences in neighborhood characteristics "across time and place" (Zenk et al. 2005, 660). Indeed, only investigating the impact of one neighborhood characteristic on food access does not account for other factors at play, and thus provides inaccurate results. The univariate analysis method cannot provide meaningful insight due to many econometric inconsistencies such as omitted variable bias. Acknowledging that many studies have used regression analysis as one of the tools to examine food access, this section explores a few studies that primarily focus on conducting multivariate regression models to analyze the effects of neighborhood characteristics on food access.

Beyond the review of past studies, an important caveat to address is the difficulty of establishing causality in these regression models. Given the observational nature of how the data for this kind of study are typically obtained, the regression models cannot easily establish causal relationships between the independent and dependent variables. Therefore, this caveat should be taken into account when discussing the effects of the demographic factors on access to groceries, in this thesis as well as in other studies.

1.4.1 Zenk et al. (2005)

Zenk et al. (2005, 660) parses out the different effects of race composition and poverty on supermarket accessibility, given that both "racial residential segregation and economic restructuring [played a role] in concentrating poverty in African American neighborhoods" in many cities. Essentially, the paper asks whether "racial disparities in supermarket accessibility occur only in higher-poverty contexts" or not. The paper finds that having a significant negative effect of high proportion of African Americans on supermarket access is contingent upon the neighborhood being impoverished.

The paper used the Detroit metropolitan area as the scope of analysis due to its extreme racial divide: African Americans overwhelmingly reside in the city while non-Hispanic White live outside the city in the Detroit metro area. The paper only focused on supermarket accessibility, based on the assumption that they "[provide] generally better availability and selection, higher quality, and lower cost of foods compared to smaller food stores" (660). With an extensive verification of data, the paper defines supermarkets as supercenters and full-service grocers that are chains "with 11 or more retail stores" (661). In addition, supermarkets within a 5-mile buffer around the sample neighborhoods are included to ensure the accurate capture of supermarket environment for each sample neighborhood. The geographic unit of analysis is neighborhood, with unspecified definitions.

The dependent variable measures the spatial accessibility of these supermarkets, which is estimated using the Manhattan block distance defined by Zenk et al. (2005, 661) as $d_{ab} = |a_i - b_i| + |a_j - b_j|$. (I use different notations from Zenk in order to match the other notations in this thesis.) The centroid of each neighborhood is represented by a while b is the location of the nearest supermarket. The latitude and longitude coordinates are denoted by i and j, respectively. The paper argues that residents are more likely to travel "on an angular route rather than in a straight line (a Euclidean distance)" (660). Nevertheless, the Manhattan distance also assumes that the streets are on a perfect North-South East-West grid. The explanatory variables for race are two dummy variables defined by the medium and high tertile of percentage of African Americans, with the reference group being the low tertile. The same method was applied to create the two dummy variables for the poverty rate.

The paper first used ordinary least squares (OLS) to regress the Manhattan block distance to the nearest supermarket on the race and poverty dummies, finding that both high proportion of African Americans and high poverty rate have a significant positive effect on distance to nearest supermarket. To observe the interaction effect between race and poverty, the paper added four possible interaction variables from Medium African American x Medium Poverty to High African American x High Poverty.

Additionally, to address spatial autocorrelation in the residuals detected by Moran's I statistic, the paper employed the moving average spatial regression. The method for the moving average spatial regression was not specified in the paper. Based on the final best-fit model of spatial regression with interaction variables, the paper finds that only neighborhoods with high poverty have significant differences in supermarket access based on racial disparity. Accordingly, a low-poverty neighborhood has roughly the same supermarket access regardless of its proportion of African Americans.

The paper begins with three different measures of the dependent variable but found comparable results in terms of the overall characterization of supermarket access given racial composition and poverty. The first is the Manhattan distance, and then the number of supermarkets within a 3-mile radius (considered reachable by car). The last one is not as commonly used, and perhaps could be explored: "potential supermarket accessibility (sum of the inverse Euclidean distances between the neighborhood and all supermarkets)" (663). The paper points out the importance of panel data to parse out how changes over time of a particular neighborhood affect supermarket accessibility. It hypothesizes that economic development might be the best way to improve African American neighborhoods if the racial composition remains similar over time. The paper also recommends including more variables such as store opening hours, crime, access to car, income, and time constraints in future analysis.

1.4.2 Powell et al. (2007)

Powell et al. (2007) improves upon the study done by Zenk et al. (2005) by incorporating different types of food stores into its analysis as well as accounting for more factors that affect neighborhood food access. Given evidence that chain supermarkets are more likely to satisfy the selection and affordability aspect of food access than non-chain grocery stores, Powell et al. (2007) ran regressions individually by each grocery type: chain supermarket, non-chain supermarket, grocery stores, and convenience, in the presumed order of selection from best to worst. These store types are defined by an industry code called the Standard Industrial Classification (SIC), which is explained in detail in Section 3.1.3. To produce more accurate results, this study not only relied on income, race, and ethnicity as its explanatory variables, but it also controlled for population, urbanization, and regions in all of its regressions.

In terms of measuring food access, the paper uses the number of food stores in each zip code, which is its geographic unit of analysis. The geographic scope covers 28,050 zip codes across the United States, which is a significantly larger sample than the sample in Zenk et al. (2005). Using zip codes as the unit of analysis is questionable because zip codes characterize efficient carrier routes and do not take into account geographical boundaries nor population size. And compared to other distance-based measures, this discrete count dependent variable does not control for population size nor density. However, the paper's inclusion of population and urbanization (created based on density) as explanatory variables effectively takes care of that omitted variable bias. Nevertheless, there probably exist more spatial issues related to zip codes that could potentially impair the paper's analysis. Furthermore, not accounting for spatial autocorrelation could worsen the accuracy of the analysis.

The paper employed the Poisson model since its dependent variable is a discrete count. Whenever there exists over-dispersion of data, as demonstrated when the dependent variable's variance is greater than its mean, the paper switches to the negative binomial model. The reason is that the expected value and the variance need to be equal in the Poisson model, whereas the negative binomial model relaxes this constraint. This occurred for the grocery stores and convenience stores, the lower-quality store types.

The paper's findings generally concurred with previous studies showing fewer available chain supermarkets and more non-chain supermarkets and grocery stores in low-income neighborhoods. The paper uses this result to support the claim that low-income households "face higher food prices" due to the inadequate access to chain supermarkets, especially considering higher mobility constraints for low-income communities (Powell et al. 2007, 193). Last but not least, by controlling for income in its multivariate regressions, the paper demonstrated that zipcodes with higher proportions of African Americans still have significantly fewer chain supermarkets regardless of income. Based on this result, the paper supports the initial findings of Zenk et al. (2005), but does not explore further the interaction effects between race and income.

Chapter 2

Food Access in Portland

As part of the City of Portland's planning guide for the next 20 years (the Comprehensive Plan), the Food System's Existing Conditions Report gives an unprecedented effort to characterize Portland's urban food system (City of Portland Bureau of Planning and Sustainability 2014b). Among different aspects of the urban food system, the report gives significant attention to food access: its issues, local conditions, conclusions, and policy alternatives. The city explains its growing attention to food access by summarizing the health issues involved as well as its lack of policies regarding food access. More concretely, Portland is experiencing rising rates of obesity and Type 2 diabetes, hence it has the pressure to reverse this trend. At the same time, the report acknowledges that one must consider the economic implications at every stage of the food-access dialogue. The two most immediate and traditional providers of healthful, affordable food are grocery stores and convenience stores, on which this thesis focuses. Nevertheless, alternative access to food such as farmer's markets are gaining increasing awareness, thus should not be ignored in conversations about food access.

Portland's supermarket hierarchy is diverse: within a 5-mile radius of Reed, there is everything from Whole Foods to Trader Joe's to Fred Meyer, and of course, Wal-Mart on 82nd. Portland seems to be increasingly invested in investigating its food deserts. The media has raised ample awareness of this issue. East of 82nd, food deserts developed after several grocery centers shut down due to lack of demand (*Addressing Portland's Food Deserts* 2013). Up to this point, there exist four important studies on grocery access in Portland: the Regional Equity Atlas by the Coalition for a Livable Future published in 2007, a study done by Metroscape magazine in 2007, a graduate thesis by Andrea Sparks at the University of Oregon in 2008, and several collaborative research papers published by Sparks and Laura Leete in 2011 and 2012. I summarize the results of these studies in this section, providing the background context for the scope of study in my analysis.

At the end of this section, I provide a map (Figure 2.1) of the neighborhoods indicated by the studies as having insufficient food access in Portland. In addition, the Portland Food Policy Council and Active Living by Design grant team conducted a food assessment and market basket survey in the ethically diverse and low-income neighborhood of Lents in 2004. The qualitative results show that half of Lents residents "reported a 15-minute or longer trip to a grocery store" (Coalition for a Livable Future 2007). As this qualitative survey serves as an introduction of studies of food access in Portland, Lents is also highlighted in Figure 2.1.

2.1 Coalition for a Livable Future (2007)

The Regional Equity Atlas was conducted by the Coalition for a Livable Future in 2007. Coalition for a Livable Future is a non-profit organization that uses research, policy and education to promote healthy, equitable and sustainable communities in the greater metro region. In its examination of supermarkets, it acknowledges that most residents drive to do their grocery-shopping. But in order to paint a picture of livability and completeness of the urban food system, it uses walkable distance to search for food deserts in Portland. Coalition for a Livable Future (2007) uses a scoring system as a combination of distance on the street network to the nearest supermarket and "a general classification of population size relative to the number of stores in the vicinity" to determine the food access for a particular neighborhood. The areas of poor food access identified by the Coalition for a Livable Future, quoted from the summary from City of Portland Bureau of Planning and Sustainability (2014b), include the following:

- The Wilkes neighborhood in outer Northeast Portland
- Along I-5 in North and Northeast Portland, including the Boise neighborhood
- South of Powell in outer East Portland
- In the area south of downtown, including the Homestead and South Portland neighborhoods

The few identifiable neighborhoods with insufficient food access (Wilkes, Boise, Homestead, and South Portland) are highlighted in Figure 2.1.

2.2 Metroscape (2007)

Metroscape is the biannual publication of the Institute of Portland Metropolitan Studies, which is a research center at Portland State University that delves into community issues in the metro area. In addition to some areas already identified by the Coalition for a Livable Future, Metroscape found the following areas of low food access, summarized in City of Portland Bureau of Planning and Sustainability (2014b): St. Johns and the Portsmouth neighborhoods, the southern border of east Portland, and some neighborhoods by the border of west Portland. St. Johns and Portsmouth are the two identifiable neighborhoods added to the neighborhoods of insufficient food access in Figure 2.1.

In addition, Metroscape also found that areas having poor food access are generally correlated with low population densities, high rates of poverty, but did not have concentration of households without a car.

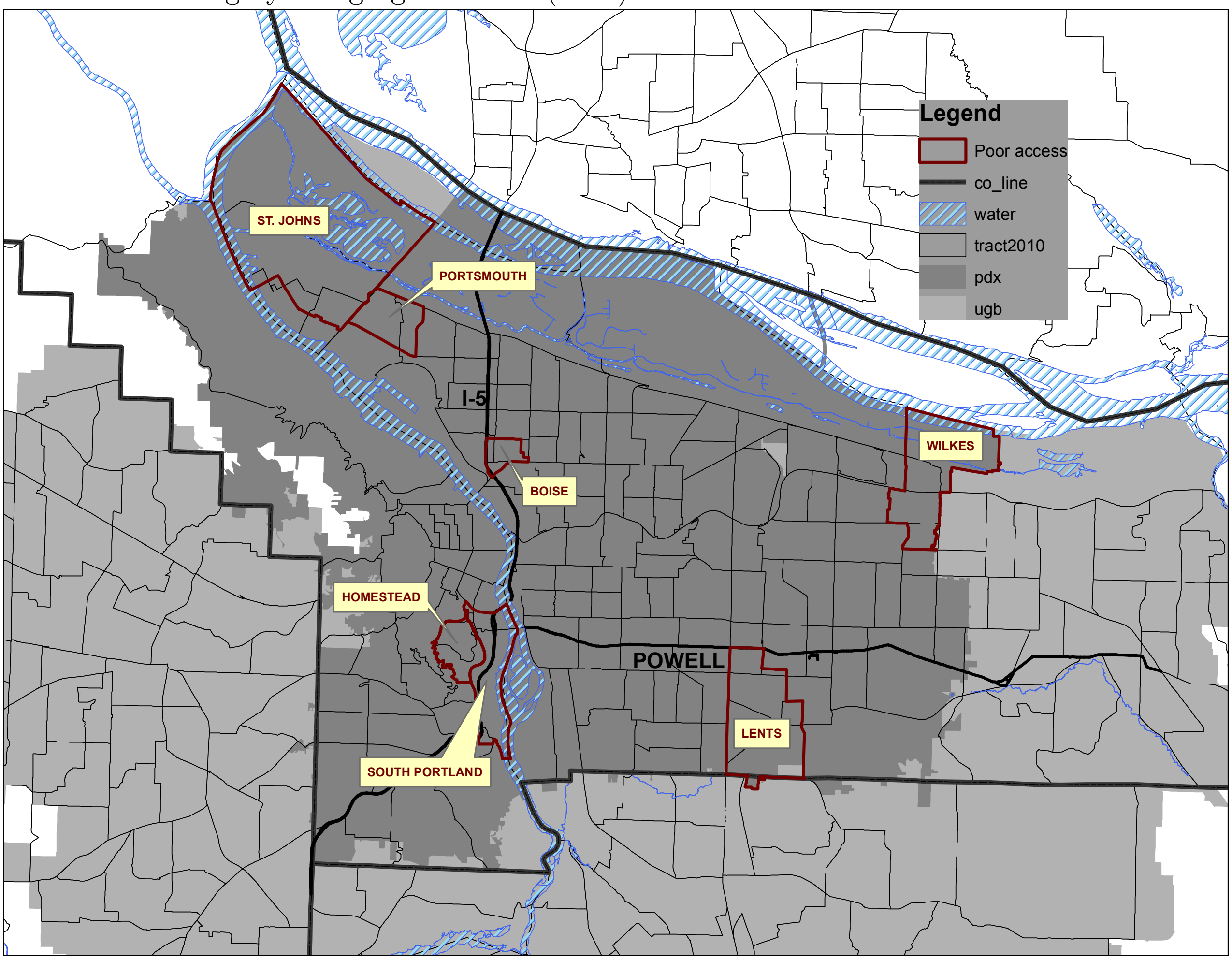
2.3 Sparks (2008) and Leete, Bania, and Sparks-Ibanga (2012)

Sparks (2008) found patterns that did not necessarily coincide with general trends. First, there exists a significant but weak correlation between higher-poverty rates and shorter average distances to the nearest supermarkets. Along the same lines, a higher-poverty rate is positively correlated with the number of supermarkets within walking distance. However, these results do not imply that poorer neighborhoods lead to better access to supermarkets.

After Sparks' thesis, she conducted two more studies on food access in Portland with Laura Leete from the University of Oregon. The latest one brought attention to a complexity to the food-access issue, food hinterlands. The paper points out that past studies have missed a significant number of areas where food access is inadequate but, "they lack concentrated socioeconomic vulnerability" to be considered food deserts (Leete, Bania, and Sparks-Ibanga 2012). The paper calls these regions "food hinterlands," and proceeds to identify them using the various food-access measures. These food hinterlands are areas where food access is really low, and even though their demographic composition is representative of the entire region, together they account for a much higher number of the poor/elderly/immobile. For example, these areas can be less dense than urban centers, but lack of mobility can overwhelmingly put residents in food-insecure environments. When considered, they tend to account for a nontrivial proportion of socioeconomically disadvantaged residents at an aggregate level.

2.4 Neighborhoods with Insufficient Food Access

Figure 2.1: Identifiable neighborhoods with insufficient food access from Coalition for a Livable Future (2007), Metroscape (2007), and the Portland Food Policy Council and Active Living by Design grant team (2004)



Chapter 3

Data and Variables

3.1 Sources of Data and Sampling Methodology

The data for this thesis were obtained from multiple sources. There are four types of data: census data, store data, Portland metro's Regional Land Information System (RLIS) data, and the City of Portland data. I first describe the census data obtained from the U.S. Census Bureau to generate most of the independent variables. The majority of the store data was sampled and purchased from referenceUSA, a private database of North American businesses. To complete the store data, some information from the current referenceUSA online database and a list of USDA SNAP retailers were also used. Finally, the data from RLIS and the City of Portland provided the spatial information for the construction of numerous variables and the maps presented in this thesis. All work related to spatial information was conducted using ArcGIS Desktop 10.1. All other data-related and statistical work was conducted in Stata. I summarize the data types and their sources in Table 3.1. More information about each data type is presented in the subsequent sections.

Table 3.1: Summary of data types and their sources

Data Type	**Data Source**
Census	U.S. Census Bureau
Store	referenceUSA; purchased
	referenceUSA; online
	USDA SNAP Retailer Locator
RLIS	Metro RLIS database
City of Portland	City of Portland Bureau of Planning and Sustainability

3.1.1 Scope of Analysis

Temporal Scope

The temporal scope of analysis is determined by the following factors. This project wishes to examine how changes in neighborhood demographic composition over time affect changes in food access. Thus, the time interval needs to be long enough to allow changes to occur in both the demographic composition and the food environment of a neighborhood without losing precision. The referenceUSA database where store data were primarily obtained only carries historical data back to 1997. In terms of census data, the most accurate and comprehensive database is either the decennial census or the American Community Survey (ACS)'s 5-year estimate. Using the 2000 Census as a starting point of comparison, the next best year where equivalent census data are available is 2010. Therefore, the constraints set by the store data and the census data determine the temporal scope of analysis to be a comparison between two years: 2000 and 2010.

Geographic Scope

In terms of the geographic scope of analysis, two different boundaries are proposed in this project: the Portland metro's Urban Growth Boundary (UGB) and the City of Portland. Both boundaries are in the urban setting so that this study could relate to the large literature on the urban food environment. The UGB effectively "separates urban land from rural land" in the Portland metropolitan region (*Metro: Urban growth boundary* 2014). The UGB and the metro boundary are very similar, but the UGB was ultimately chosen to conform with the geographic scope of analysis used by past studies analyzing food access in Portland. The second geographic scope of analysis is limited to the city boundary of Portland to incorporate more factors that are hypothesized to have a significant influence on the urban food environment. These additional factors tend to be city-specific and not comparable metro-wide, such as certain local economic development projects.

Geographic Unit

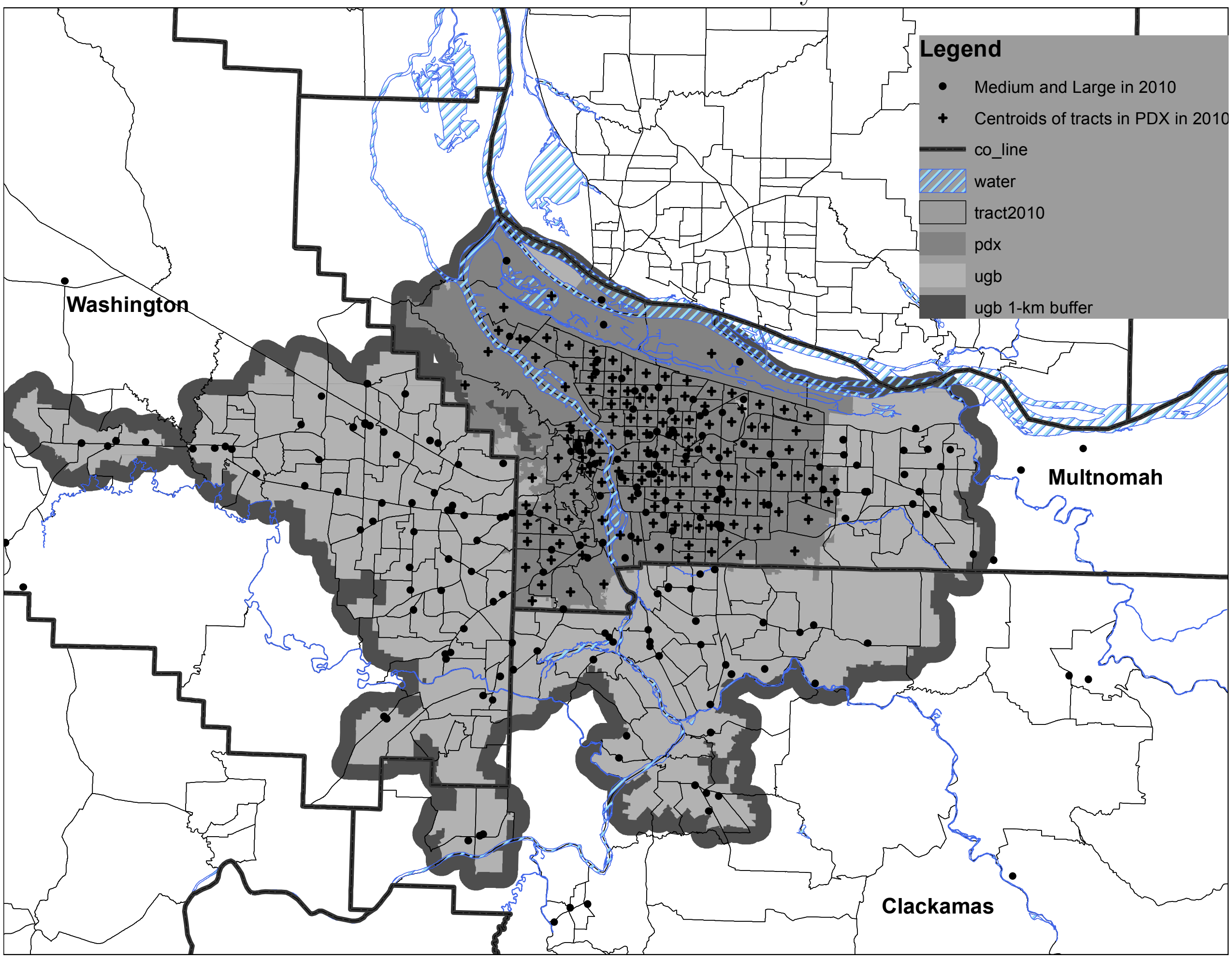
Regarding the spatial precision of the analysis, data were aggregated at the census tract level. This is the smallest geographic level at which it was possible to collect all of the information needed. For example, most of the collected census data were only available at the census tract level. Therefore, census tracts are the geographic unit of analysis in all regression models. The centroid of each census tract represents the geographic location of the census tract when constructing spatial variables. Even though some studies like Raja, Ma, and Yadav (2008) assert that census tracts are not demarcated small enough to characterize neighborhood food access, Figure 1.1 shows that most of the census tracts are about the same size as the Portland neighborhoods, and they are smaller for denser areas where people might rely more on walking proximity to groceries. In addition, results from Sparks, Bania, and Leete (2011) show that the measures of access to stores are comparable between analysis at the census tract level and analysis at the more detailed census block level.

The Edge-Effect Issue

An important issue with any spatial analysis is the edge effect. For the dependent variables, this thesis measures access to stores from the centroids of census tracts. If a census tract borders one of the geographic scopes determined above, then its residents are likely to visit stores that situate outside the geographic scope. Hence, an edge effect might appear if the stores outside the geographic scope are not included in the data, causing the access to stores for this census tract to be inaccurately measured. Thankfully, the store data acquired for this study cover at least the three Oregon counties that together contain both geographic scopes for this study. The counties are Clackamas, Multnomah, and Washington (hereafter together referred to as Tri-County). For the most part, the borders for both geographic scopes rest at least one kilometer away from the boundary of the three counties combined. The only exception is the northern border of Multnomah county that is a part of both the UGB and the City of Portland boundary. As a result, available stores for the census tracts that situate along the border are restricted by the store data since they did not include stores on the other side of the border. Nevertheless, I assume that the Columbia River acts as a "natural" restriction on the available stores for these census tracts because I do not expect residents to cross the river for stores, especially since there is a sales tax across the river.

Figure 3.1 provides the map for this discussion. Note that one kilometer is an adequate buffer distance for the UGB since it is the maximum travel distance to stores that this study adopts. One kilometer was also used to create all buffer areas for the coverage measures of access to stores. In Figure 3.1, "ugb 1-km buffer" is the one-kilometer buffer of the UGB that lies under the UGB layer, and "Medium and Large in 2010" represents grocery stores with at least 5 employees. I used this category of stores to best illustrate that there are stores that are not convenience stores or really small groceries outside the UGB buffer that the residents who live along the border can access.

Figure 3.1: The absence of edge effect due to the natural barrier placed by the Columbia River at the northern border of Multnomah county



A Caveat with Panel Data

In the following description of each data type, data were collected and manipulated separately for each year, and there were ultimately two master datasets of regression variables. In this case, regression analysis can be performed separately for each year, or in an un-balanced pooled setting. However, if any panel-data analysis is performed, the census tracts need to be identical in both years. Since there exist some minor boundary changes between the census tracts in 2000 and those in 2010, a separate sampling methodology was used to compile the master dataset that includes identical census tracts from both years. This panel master dataset is described in more detail in Section 3.4.

3.1.2 Census Data

Demographic factors constitute most of the independent variables in this study. I obtained all demographic information from the U.S. Census Bureau using its online American FactFinder tool (U.S. Census Bureau 2014c). Some basic categories of demographic information such as average household size, median household income, and unemployment were searched. These categories produced the independent variables that had been regarded as relevant in past literature and empirical studies. In addition, I also searched for certain non-basic categories such as bike commuters and vehicle access.

I searched for the demographic data from the decennial census data from 2000 and 2010. For some data found in the 2000 Census, equivalent data in 2010 were found in the 2010 ACS instead of the 2010 Census. This was because a portion of the decennial census was migrated to the ACS after the ACS was officially implemented in 2005. Summarized in Table 3.2 are the different censuses from which the data for this study were obtained, compared between 2000 and 2010.

Table 3.2: Summary of censuses comparable for 2000 and 2010

Name of Census in 2000	**Name of Census in 2010**
2000 Census Summary File 1 (SF1)	2010 Census Summary File 1 (SF1)
2010 Census Summary File 3 (SF3)	2010 American Community Survey (ACS) 5-Year Estimates

Data tables in .csv format were downloaded for the three counties of Tri-County to cover the geographic scopes of analysis identified in the previous section. The tables were extensively manipulated in Stata to generate the independent variables that were then merged into one master dataset for each year. The generated variables are summarized in the next section while the generation process is detailed in the dofile indep.do in Appendix E. Table 3.3 details the data tables obtained and provides examples of the independent variables produced from these tables.

Table 3.3: Description of data tables from the U.S. Census Bureau

Data Table	**Example of Independent Variable Produced**	**Table # for 2000**	**Table # for 2010**
Age	Percentage of total population 65 years and over	QTP1	QTP1
Average Household Size	Average household size	H018	B25010
Bike Commuters	Percentage of workers 16 and over commuting to work via bicycles	P030	B08301
Education	Percentage of adult population (25 and up) without high school diploma or equivalent	P037	B15002
Foreign Born	Percentage of foreign–born population	P021	B05002
Hispanic	Percentage of Hispanic or Latino in total population	P007	B03002
Household Size	Percentage of 2–person households	P014	B11016
Median Household Income	Median household income in the past 12 months (in thousands of 2010 inflation–adjusted dollar)	P053	B19013
Median House Value	Median value for all owner–occupied housing units in tens of thousands of dollars	H085	B25077
Poverty	Percentage of households with income in the past 12 months below poverty level	P092	B17017
Race	Percentage of total population identified as Black or African American alone or in combination	QTP5	QTP5
Single Parent	Percentage of single–parent households	P015	B11003
Total Population	Population density (Total pop in thousands/sqkm.)	P001	B01003
Unemployment	Percentage of population 16 and over unemployed	P043	B23001
Vehicle Access	Percentage of households with at least one vehicle	H044	B25044

3.1.3 Store Data

The store data in this analysis primarily consist of data purchased from referenceUSA. The referenceUSA database provides a directory of business information that compiles data from yellow pages, postal records, etc. An account executive at Infogroup, Clint Meyers, facilitated the purchase of the data. Many past studies have used referenceUSA to obtain store-level information on the grocery industry. Most relevantly, Sparks, Bania, and Leete (2011) and Leete, Bania, and Sparks-Ibanga (2012) both use referenceUSA to obtain data on grocery stores in Portland.

Data from referenceUSA were given in two Excel documents: one for 2000, and the other for 2010. Each document was sampled with the geographic scope specified as the Portland metropolitan statistical area (MSA). The Portland MSA covers the UGB, and thus transitively covers the City of Portland. Since the UGB is only bounded by Tri-County, the stores outside of the UGB were dropped if they did not belong in Tri-County.

Store Types

The store types of interest in this thesis are broadly defined as convenience stores and stores that sell fresh groceries. There are two common industry classifications used by referenceUSA: Standard Industrial Classification (SIC) and North American Industry Classification System (NAICS). SIC is older and was last updated in 1987. However, SIC has a more detailed breakdown of industry types. NAICS is the newer classification system that more or less replaced SIC. It classifies businesses based on economic factors, where "economic units that use like processes to produce goods or services are grouped together" (*What are NAICS and SIC industry codes and how do I use them?* 2013). In Table 3.4, I present a comparison of SIC and NAICS codes that could potentially be included.

Table 3.4: Potential SIC and NAICS codes

Type	**SIC**	**SIC Description**	**NAICS**	**NAICS Description**
Grocery	541101	food markets	445110	supermarkets & other grocery (No convenience)
	541104	food products – retail		
	541105	grocers – retail		
Convenience	541103	convenience stores	445120	convenience stores

The two purchased referenceUSA datasets were sampled by Clint Meyers using NAICS codes 445110 and 445120. After a thorough examination of the datasets, I concluded that every store coded NAICS–445110 was coded SIC–541105, and every store coded NAICS–445120 was coded SIC–541103. None of the stores were coded SIC–541101 or SIC–541104, hence "food markets" and "food products–retail" defined by SIC were excluded from this study. I relied on NAICS to differentiate between grocery stores and other stores that sell food, also known as stores coded SIC–541101 or SIC–541104. For a detailed record of store type examination process, please refer to the dofile stores.do in Appendix E.

Additional Store Data

During a spot check using the list of supermarket chains in Portland identified by Sparks, Bania, and Leete (2011), I discovered that the purchased referenceUSA data did not include the Fred Meyer stores because they were categorized as department stores by NAICS. Moreover, Walmart and Target were excluded for the same reason despite that some of their branches began carrying groceries around the 2000s, and have been specifically regarded as important stores for groceries (Sharkey et al. 2009). Thus, I used the online referenceUSA database to obtain the records of currently existing stores for Fred Meyer, Target and Walmart (all coded as NAICS–452111, and from now on abbreviated as FTW). I also downloaded a .csv dataset of all SNAP retailers in Oregon, which includes the FTW stores (*SNAP Retailer Locator — Food and Nutrition Service* 2014). SNAP requires its retailers to either carry fresh groceries or have more than half of the sales volume attributed to staple foods (*Retail Store Eligibility USDA Supplemental Nutrition Assistance Program — Food and Nutrition Service* 2014). Therefore, I used SNAP to broadly identify the stores that could be considered grocery stores. Then, I merged the online referenceUSA dataset with the SNAP retailers, in order to exclude the Target and Walmart branches that definitely do not fit the grocery type. Lastly, I telephoned each FTW store branch to confirm its fresh grocery availability, verify its address, and inquire whether the store existed in 2000 and/or 2010. Information from the phone verification was used to create the final FTW dataset for each year, which was subsequently appended to the purchased referenceUSA dataset of the corresponding year.

Store Categories

In the final stage of data preparation, records with duplicate addresses were dropped, keeping the unique copies with the most recent information before 2000 for the 2000 dataset, and similarly for the 2010 dataset. For records with missing or P.O. Box addresses, I updated their addresses from online research. When no credible information was found online, the records were dropped. Ultimately, 14 records in 2000 and 8 records in 2010 were dropped due to address issues. In the end, there were 258 convenience stores and 433 grocery stores in 2000, and 304 convenience stores and 425 grocery stores in 2010.

There were two ordinal variables that classify the employee size and sales volume for each store, which I utilized to categorize the grocery stores. Since there was no price-related information given about each store, employee size and sales volume provided the most detailed and effective breakdown of stores for this study, which is more-or-less a store-size categorization. Figure 3.2 is a tabulation of all grocery stores (all stores excluding convenience stores) by the two ordinal variables in the 2010 dataset. Note that the number of stores in the tabulation is 423 instead of the total of 425 because two stores did not have a value for the sales volume variable.

Figure 3.2: Tabulation of employee size and sales volume for all grocery stores in 2010

Location Employee Size Range	Location Sales Volume Range									
	Less than	$500,000	$ 1 Milli	$ 2.5 Mil	$ 5 Milli	$ 10 Mill	$ 20 Mill	$ 50 Mill	$ 100 Mil	Total
1 - 4	139	72	0	0	0	0	0	0	0	211
5 - 9	0	0	34	0	0	0	0	0	0	34
10 - 19	0	0	11	17	0	0	0	0	0	28
20 - 49	0	0	0	3	10	5	0	0	0	18
50 - 99	0	0	0	0	0	30	25	0	0	55
100 - 249	0	0	0	0	0	1	66	1	0	68
250 - 499	0	0	0	0	0	0	0	6	0	6
500 - 999	0	0	0	0	0	0	0	0	3	3
Total	139	72	45	20	10	36	91	7	3	423

Given the high correlation between employee size and sales volume (0.9809), I could categorize the grocery stores using only one of the two variables without losing pertinent information. Employee size was chosen over sales volume because every store has a value for employee size whereas two stores are missing the value for sales volume. At least one other study uses employee size to category stores (Raja, Ma, and Yadav 2008; Moore and Diez Roux 2006). Table 3.5 is the final categorization of grocery stores used to construct the measures of access to stores.

Table 3.5: Categorization of grocery stores

Store Category	Employee Size Range	Number of Stores in 2010 (425)	Number of Stores in 2000 (433)
Small groceries	1 – 4	211	232
Medium groceries	5 – 49	80	91
Large groceries	50 – 999	134	110

3.1.4 Spatial Data from RLIS and the City of Portland

Spatial Data from RLIS

Once store data were obtained and properly cleaned, they were brought to ArcGIS for three purposes: geo-location, visualization, and construction of measures of access to stores. Here I discuss the spatial data downloaded from the online RLIS database, which provided all spatial data needed to achieve the purposes stated above.

All spatial data from RLIS were contained in a spatial data format called shapefiles (.shp). The only exception was the RLIS Address Locator, which has its own data format. The RLIS database comprises numerous ArcGIS shapefiles that provide spatial information about the Portland metro. It is managed by Metro, and thus regarded as an organized, comprehensive, and trusted data source. Acquiring all spatial data from RLIS also ensures a seamless integration of data. There would be minimal discrepancies between the methodologies of the generation and maintenance of the shapefiles. For instance, the combination of the RLIS Streets shapefile and the RLIS Address Locator would most accurately match and display an address to its geographic location since the naming convention of streets is the same in both files.

The spatial data from RLIS are Analysis Centers, County Lines, City Limits, Streets, Taxlots, UGB, 2000 Census Tracts, 2010 Census Tracts, and the RLIS Address Locator. The RLIS shapefiles are mostly background geographic information that sets up the geographic scopes and spatial precision for this study. The only shapefile from RLIS that is not background information is Analysis Centers, which I used to construct an independent variable. Metro (2014) defines Centers as "compact, mixed-use areas of high-density housing, employment, and retail that are pedestrian-oriented and well served by public transportation and roads." Centers are a part of the 2040 Growth Concept Map from Metro, and they are also reflected in the City of Portland Comprehensive Plan. Analysis Centers are the actual geographic boundaries of Centers, which are normally presented conceptually in the shape of a circle.

I summarize all of the data collected in Table 3.6. For the detailed process of how these spatial data were used to construct the variables, see Appendix A.

Spatial Data from City of Portland

The spatial data from the City of Portland were obtained in a map package called "PDX Comprehensive Plan Background Layers." This map package contains all the background spatial information shown in the Comprehensive Plan MapApp created by the City's Bureau of Planning and Sustainability. The MapApp is an online open source spatial visualization of the City's Comprehensive Plan.

In a map document, the presentation of a shapefile's spatial data is called a layer. The individual data layers of interest from the map package were Neighborhoods, Transit Access, Urban Renewal Areas (URA), and Water. These four data layers were extracted from the map package, saved as shapefiles, and integrated with the RLIS shapefiles.

Neighborhoods are the official neighborhoods the the City of Portland has adopted. Transit Access represents "places within $\frac{1}{4}$ mile of the frequent transit," which consists of the "MAX [(Metropolitan Area Express) Light Rail], Portland Streetcar and frequent TriMet bus service" (City of Portland Bureau of Planning and Sustainability 2014a). URA represents the areas under the Portland Development Commission (PDC)'s "state-authorized, redevelopment and finance program designed to help communities improve and redevelop areas that are physically deteriorated, suffering economic stagnation, unsafe or poorly planned" (Commission 2014a). Major Waterbodies consist of all bodies of water in the Portland metro, some of which are important "natural" boundaries that may influence access to stores (such as the Columbia River).

Table 3.6 summarizes the City of Portland shapefiles along with the RLIS shapefiles, and provides a brief description of how each shapefile was utilized.

3.2 Independent Variables

The independent variables for this thesis were generated using census data and a few selected spatial shapefiles. I briefly describe the independent variables that each data type produced. Note that only the independent variables used in regression analyses are presented here. For example, the following variables were generated from the same census data table: percentage of households with no vehicle and percentage of

Table 3.6: Spatial Data from RLIS and the City of Portland

Filename	**Description or Usage**	**Data Source**
Analysis Centers	To construct the independent variable "Distance to Centers"	RLIS
Transit Access	To construct the independent indicator variable "Census Tract Centroid within Transit Access Area"	City of Portland
URA	To construct the independent indicator variable "Census Tract Centroid within an URA"	City of Portland
Taxlots	To visualize the space of the store lots	RLIS
Neighborhoods	To compare to census tracts	City of Portland
Streets	To transform into a network dataset	RLIS
Major Waterbodies	To visualize the natural boundaries from the bodies of water in the region	City of Portland
City Limits	To set the geographic scope to the City of Portland	RLIS
UGB	To set the geographic scope to the UGB	RLIS
County Lines	Clackamas, Multnomah, and Washington counties	RLIS
2000 Census Tracts	To demarcate the units of analysis in this study	RLIS
2010 Census Tracts	To demarcate the units of analysis in this study	RLIS
RLIS Address Locator	To geocode addresses of stores	RLIS

households with at least one vehicle. Since the two are perfectly collinear, only one of them is presented here. For a list of all generated independent variables, please refer to Appendix C.1.

3.2.1 Census Independent Variables

The census independent variables were produced from the following data tables: Age, Average Household Size, Bike Commuters, Education, Foreign Born, Hispanic, Household Size, Median Household Income, Median House Value, Poverty, Race, Single Parent, Total Population, Unemployment, and Vehicle Access. Most of these data tables produced independent variables that past researchers had used to associate with food access. The independent variables from Bike Commuters are the only variables that were not included in the past studies reviewed by this thesis.

Race and Poverty

As examined in the literature review chapter, researchers have explored the effect of race and poverty rate on food access extensively. Zenk et al. (2005) focuses exclusively on the poverty rate and the percentage of African Americans in its regression analysis; and the same approach of pairing race and poverty rate is used in numerous studies including Baker et al. (2006) and Franco et al. (2008). Most of these studies follow Zenk et al. (2005) in constructing indicator variables for different intervals of the racial composition and the poverty rate. For example, if a census tract has a poverty rate in the lowest tertile, then the "low-poverty" indicator variable would be 1 for that census tract. Having indicator variables leads to a simpler interpretation of results. For instance, instead of analyzing the effect of a one percent decrease in the poverty rate on the distance to the nearest grocery store, the studies analyze the effect of being classified as a "low-poverty" neighborhood on the distance to the nearest grocery store. On the other hand, there is a loss of information associated with having indicator variables, precisely because they are designed to generalize the census tract characteristics. This thesis attempts to construct variables that retain as much information as possible, so most of the census independent variables are percentages. In addition to the percentage of African Americans, this thesis also includes more race categories such as the percentage of Asians and the percentage of Hispanic.

Education and Income

The City of Portland Bureau of Planning and Sustainability has interviewed storeowners for their criteria to locate in specific neighborhoods (Cohen 2013). Even though the set of criteria differs depending on the type of stores and its consumer base, it generally includes education and income. Thus, percentages of a few education categories are constructed, as well as the median household income. The education variables include the percentage of the adult population without a high school diploma and the percentage of the adult population without a bachelor's degree. These two variables should embody the effect of education levels of residents on access to stores. Note that median household income is naturally strongly correlated with the poverty rate, which means that collinearity issues may occur if they are both included in a regression.

Age and Household Size

Age is added to the list of demographic information because age is often an indicator of the types of food retail demanded in an area. Percentage of the elderly population is often used as an indicator of the neighborhood's lack of mobility, as argued in Leete, Bania, and Sparks-Ibanga (2012). I divided age into five categories, from under 18, to over 65. As a result, five percentage variables for the composition of each age category were constructed.

Household size is a key indicator of demand for groceries as well as food consumption patterns (MacDonald and Nelson 1991). It could also act as a proxy for residential areas, since household size is generally bigger in residential areas. This proxy is important to distinguish the composition of stores in residential areas versus industrial areas. I have two groups of household size variables. The first group consists of seven percentage variables that each describes the composition of n-person households in the tract. The next group consists of one variable: the average household size. These two groups are considered substitutes. If the effect of an increase in the percentage of 1-person households is the same as the effect of an increase in the percentage of 2-person households less the effect of 1-person households, and so forth, then average household size could be adopted without a major loss of information.

Unemployment and House Value

The unemployment rate characterizes the general economic condition of a neighborhood. It is employed in at least three different studies, often in addition to a measure of income or the poverty rate (Apparicio, Cloutier, and Shearmur 2007; Larsen and Gilliland 2008; Leete, Bania, and Sparks-Ibanga 2012). Since the unemployment rate does not account for retired people, this factor should be accompanied by the percentage of population over the age of 65, which is done in Leete, Bania, and Sparks-Ibanga (2012). I include the unemployment rate in the list of independent variables.

In addition, house value is also argued to be a good indicator of neighborhood economic condition by Morland et al. (2002). I include it to act as a potential substitute for median household income. However, the collected data for median house value only characterize owner-occupied housing units, whereas median household income applies to all households regardless of tenure status.

Density and Vehicle Access

More control variables include density and vehicle access. Density is a factor to be accounted for in any analysis related to spatial retail. Indeed, density has perhaps the most direct effect on access to stores since high density implies a profitable market area for any stores to locate in. The only exception is perhaps stores that prefer to be on the outskirts of a city, such as certain large supermarkets. These stores may have a large threshold, so they do not need to locate in the densest areas. In addition, they have a high land requirement, which is easier to satisfy in areas with a lower density due to fewer land-use restrictions. This is based on the assumption that land-use restrictions tend to increase land prices. (For a definition of threshold as introduced by Church and Murray (2008, 6), please refer to Section 1.1.1.) Density is also included to account for the zoning restrictions that may dictate where stores can locate by implicitly separating what areas are primarily residential versus industrial (Small and McDermott 2006). In this thesis, density is calculated as the total population per square kilometers. This variable requires the area of each census tract, which is calculated using ArcGIS.

Studies have also explored the link between the availability of vehicles and access to food, demonstrating that car ownership significantly increases access to healthful, affordable food. Studies that correlated vehicle ownership with access to stores include Leete, Bania, and Sparks-Ibanga (2012) and MacDonald and Nelson (1991). In this thesis, vehicle access is characterized by the percentage of all housing units with access to at least one vehicle.

Bike Commuters

Commuting by bike is a significant means of commuting to work in the City of Portland. It would be interesting to explore whether a neighborhood with more bike commuters affects access to the stores located nearby. This is a factor that has not appeared in the past research reviewed by this thesis. Note that there may be a strong correlation between bike commuters and how close they live from downtown, which weakens the direct link between bike commuters and access to stores. I include the percentage of workers who bike to work as the variable for the Bike Commuters factor.

3.2.2 Spatial Independent Variables

The spatial independent variables constructed for this thesis are derived from three shapefiles collected from RLIS and the City of Portland: Analysis Centers, Transit Access, and URA. The variables created from Analysis Centers from RLIS are distance variables while the variables constructed from the other two shapefiles from the City of Portland are indicator variables. The construction of these variables is described in detail in Appendix A.

Distance to Central Business District (CBD) and Distance to Nearest Center

As mentioned in the discussion of bike commuters, the distance from a neighborhood to the CBD is a significant factor that may potentially dominate over some other neighborhood characteristics. Distance to downtown Portland measures the proximity to the CBD of Portland, where there are more stores (but not necessarily groceries) due to the concentration of economic and employment opportunities. Although being closer to the CBD does not guarantee being closer to a grocery or having more groceries within a distance of a neighborhood, it is a good indicator of a more sufficient general level of goods and services available to a neighborhood. The distance to downtown variable is constructed by measuring the distance on a street network between the tract centroid and the centroid of downtown (defined as the Portland Central City) in the Analysis Centers shapefile.

An alternative measure is distance to the nearest Center. Centers not only include the CBD, but also other areas that have concentrated economic and employment opportunities, with high traffic and diverse transportation options. Other than the Central City (also known as the CBD), Centers can be categorized as Regional Centers

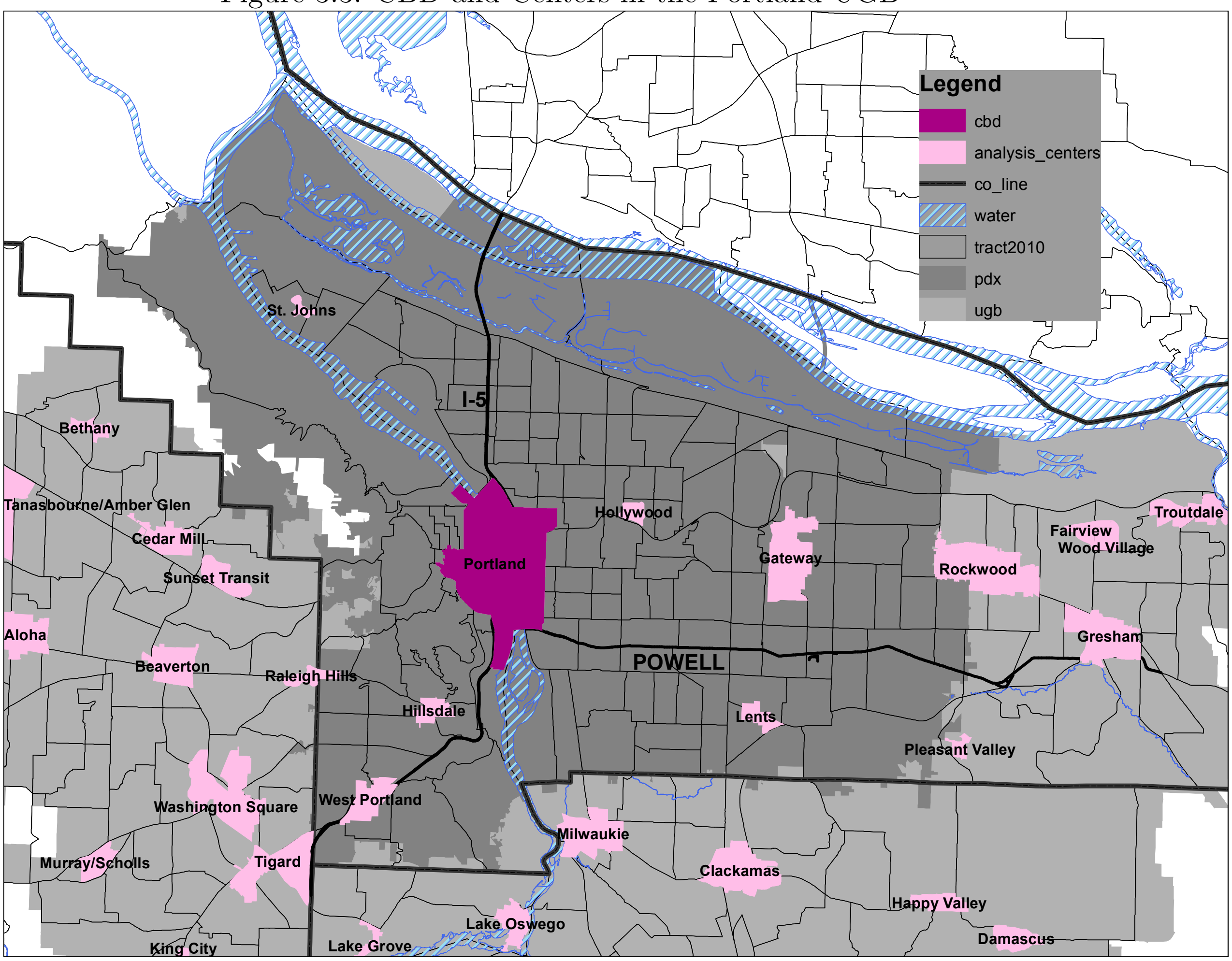
Table 3.7: Description for selected census independent variables

Variable Name	Variable Description	Data Table
AGE18	Percentage of total population under 18 years	Age
AGE1824	Percentage of total population 18 to 24 years	Age
AGE4564	Percentage of total population 45 to 64 years	Age
AGE65	Percentage of total population 65 years and over	
HIGHSCHNO	Percentage of adult population (25 and up) without high school diploma or equivalent	Education
BACHELORNO	Percentage of the adult population (25 and up) without a bachelor's degree	Education
HHSIZE2	Percentage of 2–person households	Household Size
HHSIZE3	Percentage of 3–person households	Household Size
HHSIZE4	Percentage of 4–person households	Household Size
HHSIZE5	Percentage of 5–person households	Household Size
HHSIZE6	Percentage of 6–person households	Household Size
HHSIZE7UP	Percentage of 7–or–more–person households	Household Size
AVHHSIZE	Average household size	Average Household Size
BLACK	Percentage of total population identified as Black or African American alone or in combination	Race
ASIAN	Percentage of total population identified as Asian alone or in combination	Race
HISP	Percentage of Hispanic or Latino in total population	Hispanic
POV	Percentage of households with income in the past 12 months below poverty level	Poverty
MEDHHINC	Median household income in the past 12 months (in tens of thousands of 2010 inflation–adjusted dollars)	Median Household Income
UNEMP	Percentage of population 16 and over unemployed	Unemployment
MEDHV	Median value for all owner–occupied housing units in tens of thousands of dollars	Median House Value
DENS	Population density (Total pop in tens of thousands/sqkm.)	Total Population
VEH	Percentage of households with at least one vehicle	Vehicle Access
SING	Percentage of single–parent households	Single Parent
BIKE	Percentage of workers 16 and over commuting to work via bicycles	Bike Commuters
FOREIGN	Percentage of foreign–born population	Foreign Born

such as Gateway, and Town Centers such as Hollywood. The street network distance from the tract centroid to the nearest Center is measured. Note that for certain census tracts in Portland, their nearest Center could be outside of the city boundary, just as their nearest grocery could lie outside of the city boundary as well. Another caveat is that the some of the Centers have not been adopted by Metro, but are still taken into consideration for the distance variable because they are more likely to possess features similar to the adopted Centers.

Because the data for these variables are static, these variables cannot be included in a panel dataset with fixed effects. Figure 3.3 displays the CBD named "Portland," and the Centers within and around the City of Portland.

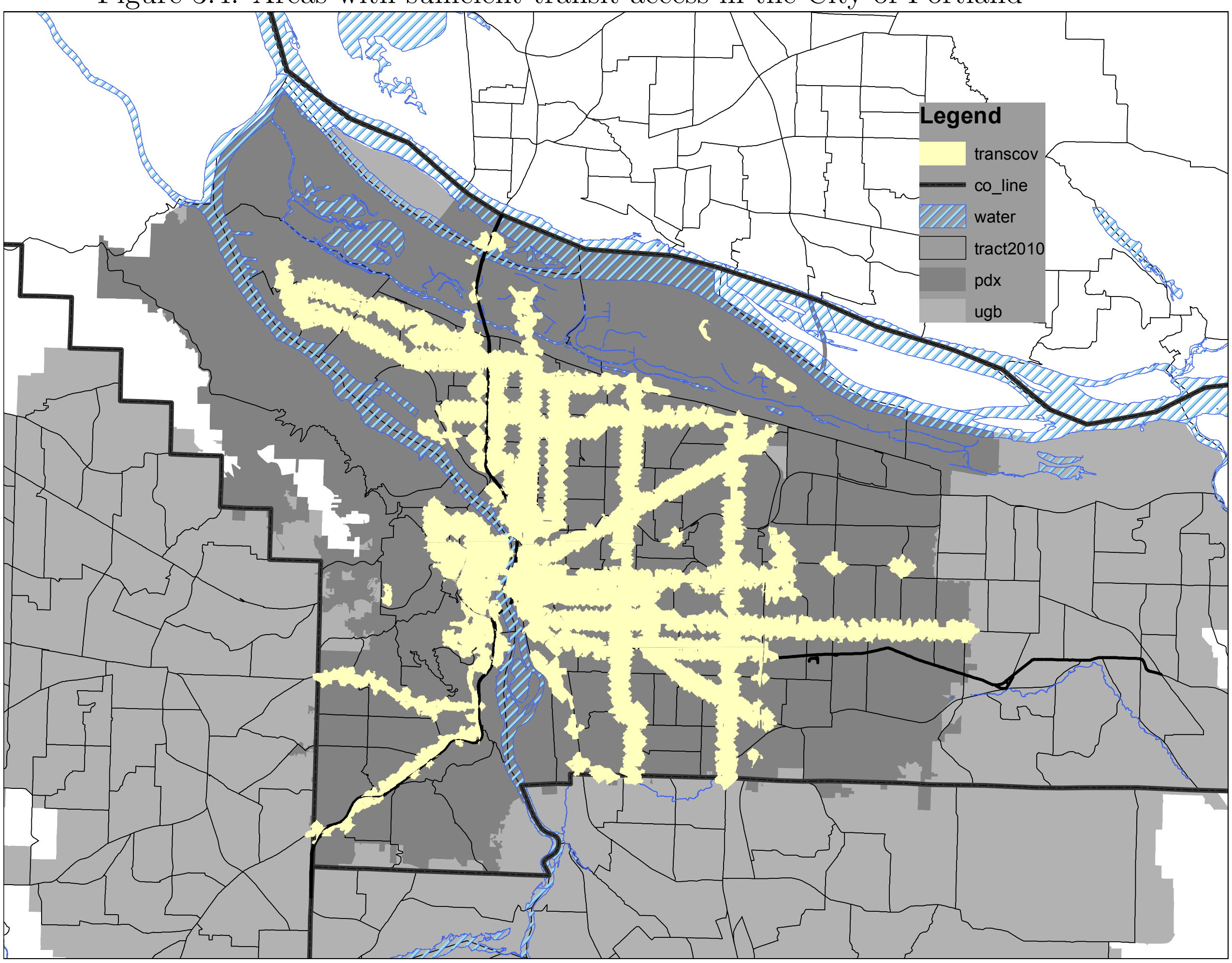
Figure 3.3: CBD and Centers in the Portland UGB



Transit Access

Transit Access is a shapefile that the City of Portland created, which presents areas in the city with sufficient transit access. The Bureau of Planning and Sustainability states that "47 percent of Portland households are located within a convenient ($\frac{1}{4}$–mile) walk to the frequent transit network" (Portland Bureau of Planning and Sustainability 2014). Therefore, a census tract is considered to have sufficient transit access if its centroid is within the Transit Access boundaries, resulting in an indicator variable. Note that this variable reflects the current state of transit options, which means it cannot be compared between years. Figure 3.4 displays the Transit Access shapefile.

Figure 3.4: Areas with sufficient transit access in the City of Portland



URA and URA Lag

The Portland Development Commission (PDC) is "Portland's urban renewal and economic development agency" (Commission 2014b). Its Urban Renewal Area (URA) program is one of the largest neighborhood economic development programs in Portland. Through this program, many projects that improve the neighborhood's economic, social, and environmental outlook can be capitalized. The PDC also has other development projects and a program called the Neighborhood Prosperity Initiative (NPI). Most of PDC's other investment project areas are covered by the URAs, thus they are not included here. In terms of the NPI, most projects of the program are in 2011, past the temporal scope of analysis for this study. All in all, if the centroid of a neighborhood is within an URA, then it is considered to be in the process of improving its economic and planning conditions. Being in an URA may improve access to stores, especially when this indicator variable is compared between years.

In order to ensure that the efforts of the URA program are implemented in time for an effect on access to stores to be realized, a lag indicator variable is constructed. The centroid of the census tract must be within an URA that was created at least two years before 2000 for the 2000 dataset, and similarly for the 2010 dataset. Figure 3.5 displays the URA in the City of Portland.

Table 3.8: Description for selected spatial independent variables

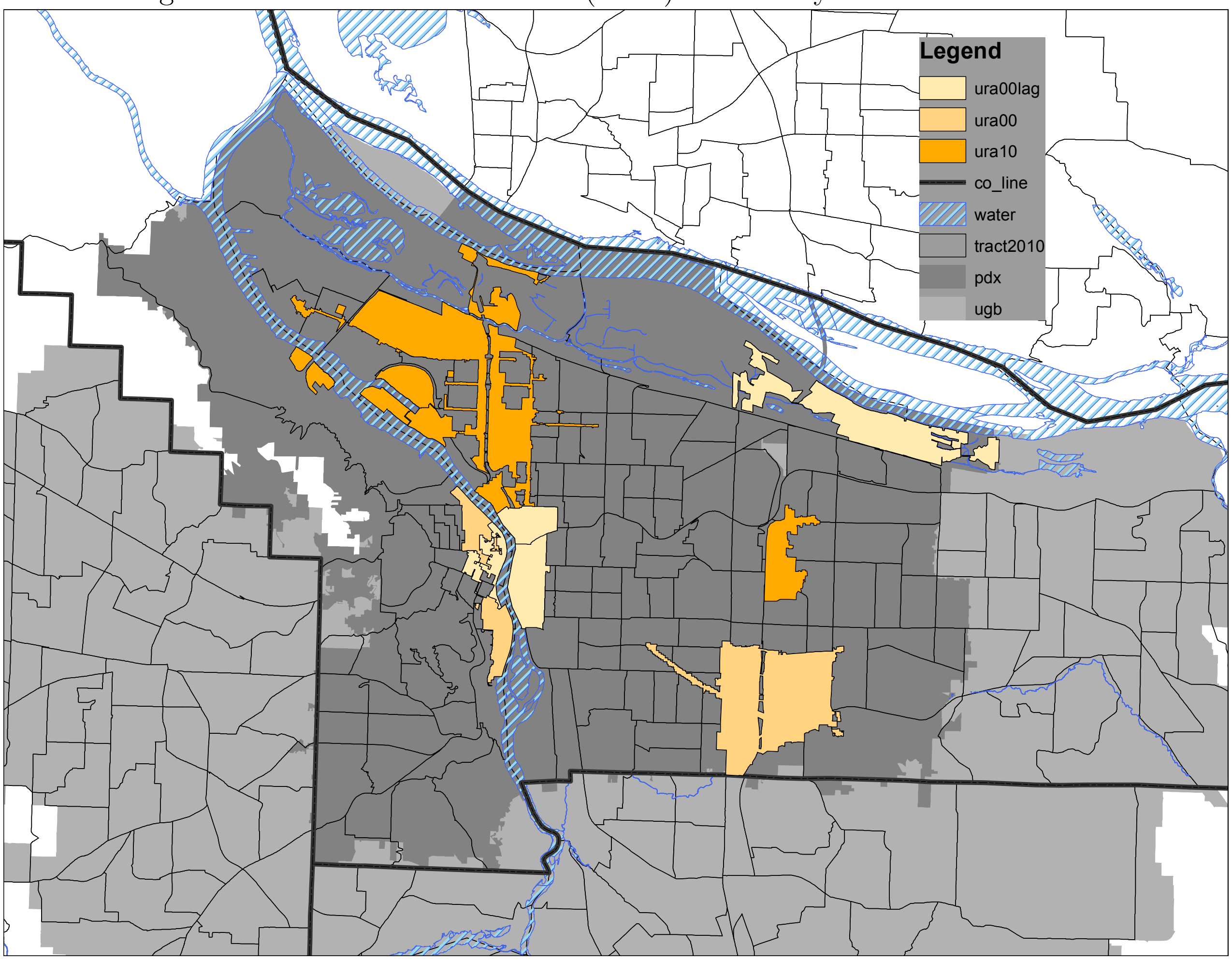
Variable Name	Variable Description	Filename of Shapefiles Used
CBD	Distance (km) to the Portland Central City on street network	cent10_pdx.shp, cbd_cent.shp
CENTER	Distance (km) to the nearest Center on street network	cent_pdx.shp, center.shp
TRANSCOV	Centroid is within places within $\frac{1}{4}$ mile of the frequent transit	transcov.shp
URA	Centroid is within an URA	ura10.shp
URALAG	Centroid is within an URA at least two years before current year	ura10lag.shp

3.2.3 Summary Statistics of Independent Variables

I discuss a few interesting observations regarding the summary statistics of the independent variables used in this thesis, presented in Table 3.9.

- The age compositions are roughly the same between two years.

Figure 3.5: Urban Renewal Areas (URA) in the City of Portland



- Level of education has improved based on a higher proportion of residents having high school or college education.
- The proportion of residents without a bachelor's degree varies more than the proportion of residents without a high school diploma.
- Generally, there is over a third of 2-person households out of all households, with average household size being just above 2 people in both years.
- Generally, the African American composition has decreased, the Asian and Hispanic compositions increased.
- Most noticeably, the maximum proportion of African Americans in a census tract decreased from almost 60% to about 30%.
- At the average, the proportion of households in poverty and median household income have stayed roughly the same between the two years.
- Unemployment has risen while median household value has gone up even more.
- Density and proportion of households with a vehicle have stayed the same between 2000 and 2010. The density measure may reflect zoning and growth control practices in the city.
- Note that distance to downtown and distance to the nearest Centers vary slightly, but they are identical in the analysis since I switch to using the same census tract centroids to calculate them.
- The proportion of single-parent households and the proportion of foreign-born have remained consistent between two years, at about 15% and 12% respectively.
- The proportion of bike commuters has risen from two percent to five percent.
- As with distance to downtown and distance to the nearest Center, the minor differences between 2000 and 2010 disappear once the census tract boundary is uniform between years.
- There's generally more probability of being in an Urban Renewal Area in 2010, as there has been an increase in the number of URAs between 2000 and 2010.

3.3 Dependent Variables

3.3.1 Construction of Dependent Variables

The dependent variables of this study are spatial measures of access to stores in four different variants. The general process of constructing these measures using ArcGIS is described below. This procedure was repeated for each year.

1. Geocode all store addresses using the RLIS Address Locator

Table 3.9: Summary Statistics of selected independent variables

Variable Name	**Mean in '00/'10**	**Std. Dev. in '00/'10**	**Min. in '00/'10**	**Max. in '00/'10**
AGE18	.200 / .183	.072 / .066	.016 / .015	.328 / .32
AGE1824	.103 / .094	.049 / .061	.038 / .038	.367 / .484
AGE4564	.227 / .256	.046 / .053	.135 / .117	.360 / .399
AGE65	.114 / .104	.045 / .040	.023 / .036	.314 / .307
HIGHSCHNO	.141 / .100	.084 / .081	.010 / 0	.348 / .341
BACHELORNO	.663 / .572	.195 / .197	.231 / .208	.967 / .938
HHSIZE2	.331 / .352	.059 / .069	.075 / .129	.571 / .567
HHSIZE3	.141 / .139	.045 / .052	0 / 0	.225 / .294
HHSIZE4	.104 / .101	.043 / .053	0 / 0	.199 / .273
HHSIZE5	.046 / .035	.023 / .026	0 / 0	.095 / .123
HHSIZE6	.019 / .013	.016 / .016	0 / 0	.069 / .080
HHSIZE7UP	.014 / .009	.014 / .014	0 / 0	.074 / .079
AVHHSIZE	2.30 / 2.27	.415 / .391	1.09 / 1.13	2.93 / 3.13
BLACK	.086 / .076	.120 / .071	.006 / .004	.580 / .311
ASIAN	.069 / .083	.036 / .042	.017 / .026	.185 / .284
HISP	.065 / .081	.045 / .060	0 / 0	.279 / .270
POV	.122 / .145	.084 / .081	0 / 0	.584 / .429
MEDHHINC	5.41 / 5.47	2.05 / 2.28	1.03 / 1.32	14.0 / 14.1
UNEMP	.068 / .086	.053 / .042	0 / .012	.517 / .387
MEDHV	17.6 / 33.2	7.02 / 12.0	0 / 17.7	47.3 / 81.0
DENS	.244 / .258	.136 / .137	.002 / .003	.916 / .928
VEH	.857 / .862	.137 / .123	.096 / .161	.995 / 1
CBD	7.48 / 7.63	3.77 / 3.73	.215 / .558	16.6 / 16.6
CENTER	3.20 / 3.19	1.51 / 1.50	.215 / .287	7.81 / 7.80
BIKE	.020 / .053	.026 / .052	0 / 0	.221 / .226
SING	.153 / .153	.079 / .080	0 / 0	.545 / .422
FOREIGN	.123 / .126	.058 / .072	0 / .022	.300 / .403
TRANSCOV	.284 / .276	.452 / .448	0 / 0	1 / 1
URA	.104 / .156	.306 / .364	0 / 0	1 / 1
URALAG	.048 / .156	.215 / .364	0 / 0	1 / 1

2. Generate shapefiles of stores categorized by Table 3.5.

3. Construct the four different variants of measures for each category of stores

4. Export the datasets of measures to .txt files in order to merge them into the master dataset of regression variables in Stata

Each store record contained the name and address that are pertinent to identification and geocoding using ArcGIS. "Geocoding is the process of transforming a description of a location — such as ... an address ... — to a location on the earth's surface ... for mapping or spatial analysis" (*ArcGIS Help 10.1 - What is geocoding?* 2014). Using the RLIS Address Locator, all but six store addresses were successfully geo-coded. When the RLIS Address Locator could not find matching addresses for certain stores, I used Google Maps to find the correct addresses. The six addresses (four from 2000 and two from 2010) without any online verification were ultimately dropped. Moreover, if the RLIS Address Locator found tied addresses for a store, I looked up the visual location of the referenceUSA store address on Google Maps, then I used the Taxlots shapefile to pinpoint the RLIS address that was visually the closest match. This happened mostly for stores that are in a shopping area with multiple store lots. In the end, there existed 727 stores in 2010 and 687 stores in 2000 (including convenience stores).

Using the store categories defined in Table 3.5, I generated several shapefiles that contain the potential groupings of stores. Other than the shapefiles that each contain one category of stores, I also generated a shapefile of convenience stores. Table 3.10 is the list of store shapefiles generated.

Table 3.10: Store Shapefiles

Store Shapefile	**Filename of Shapefile**
All convenience stores	con.shp
Small groceries	empcd1.shp
Medium groceries	empcd2to4.shp
Large groceries	empcd5to.shp

The dependent variable for this study is either a coverage measure: the number of stores present within one kilometer of the census tract centroid (M1), or a distance measure: distance to the nearest store from the census tract centroid in kilometers

(M2). Each measure was constructed with two different distance constructs: Euclidean (D1) and street network (D2). All in all, there were four variants of measures constructed for each store category: M1D1, M1D2, M2D1, and M2D2.

According to Sparks, Bania, and Leete (2011), M1 for supermarkets — otherwise defined as the large groceries in this study — has little variation between census tracts, and thus might be an insufficient dependent variable to use. On the other hand, the higher dispersion of convenience stores and small groceries makes M1 a better candidate for the dependent variable for these store categories. As for M2, I hypothesize that it is a more effective measure of access for large groceries since it is more robust to aggregation errors thanks to the smaller dispersion of large groceries. In terms of the two distance constructs, Sparks, Bania, and Leete (2011) concluded that D1 produces comparable results to D2, and is thus an acceptable distance construct for measuring food access, especially when the goal is to compare measures between neighborhoods. However, since this thesis attempts to produce more accurate and "absolute" measures of access to stores, D2 is still preferred over D1 due to its relatively more realistic depiction of travel distance. I used both D1 and D2 to construct the measures, and I will compare the results they produce. Table 3.11 summarizes the four variants of measures constructed for each store type.

Table 3.11: Variants of measures of access to stores

Measure Type	**Distance Construct**	**Variant Name**
Number of stores present within 1km of census tract centroid	Euclidean Distance	M1D1
	Street Network	M1D2
Distance to the nearest store from census tract centroid	Euclidean Distance	M2D1
	Street Network	M2D2

The ArcGIS tools and procedures used to construct these variants are detailed in Appendix A. For each store shapefile in Table 3.10, the four variants of measures were constructed. Each variable name consisted of the shapefile name and the variant name. Table 3.12 describes all of the dependent variables constructed.

3.3.2 Summary statistics of All Dependent Variables

Table 3.13 present the summary statistics of the measures of access to stores in 2000 and 2010. I give an overview of the statistics here:

- As expected, there are more convenience stores or small groceries than medium or large groceries.

Table 3.12: Description of all dependent variables

Variable Name	Variable Description
CONM1D1	Number of convenience stores within 1km Euclidean buffer
CONM1D2	Number of convenience stores within 1km buffer on street network
CONM2D1	Distance (km) to nearest convenience store in Euclidean distance
CONM2D2	Distance (km) to nearest convenience store on street network
SMALLM1D1	Number of small groceries within 1km Euclidean buffer
SMALLM1D2	Number of small groceries within 1km buffer on street network
SMALLM2D1	Distance (km) to the nearest small grocery in Euclidean distance
SMALLM2D2	Distance (km) to the nearest small grocery on street network
MEDM1D1	Number of medium groceries within 1km Euclidean buffer
MEDM1D2	Number of medium groceries within 1km buffer on street network
MEDM2D1	Distance (km) to the nearest medium grocery in Euclidean distance
MEDM2D2	Distance (km) to the nearest medium grocery on street network
LARGEM1D1	Number of large groceries within 1km Euclidean buffer
LARGEM1D2	Number of large groceries within 1km buffer on street network
LARGEM2D1	Distance (km) to the nearest large grocery in Euclidean distance
LARGEM2D2	Distance (km) to the nearest large grocery on street network

- Using the network distance over the Euclidean distance reduces the average number of convenience stores or small groceries while raising the average number of medium or large groceries.
- Using the network distance over the Euclidean distance decreases the the distance to the nearest convenience or small grocery, but increases the distance to the nearest medium or large grocery store.
- It is harder to compare and contrast measures of store access between years, as they are all roughly similar.

3.4 Panel Data Sampling Methodology

As previously stated, a panel dataset with census tracts in 2000 and 2010 requires the tracts to be identical in both years. However, certain tracts in 2000 either split into multiple tracts, merged into one tract, or had minor boundary adjustments in 2010. For example, there were 28 tracts within the UGB in 2000 that were split into at least two tracts in 2010. These boundary changes not only give rise to a different number of observations for each year, more importantly they also imply that some observations represent completely different geographic areas. There are two steps to mend this problem: determine a common boundary definition applied to both years, and transform the variables to reflect the tracts determined by the common boundary

Table 3.13: Summary statistics of all dependent variables in 2000/2010

Store Category	**Variable Name**	**Mean in '00/'10**	**Std. Dev. in '00/'10**	**Min. in '00/'10**	**Max. in '00/'10**
Convenience Stores	CONM1D1	2 / 2.340	1.809 / 2.086	0 / 0	9 / 12
Convenience Stores	CONM1D2	1.236 / 1.347	1.44 / 1.549	0 / 0	7 / 9
Convenience Stores	CONM2D1	.7811 / .7473	.5792 / .4888	.111 / .1088	4.309 / 2.937
Convenience Stores	CONM2D2	1.126 / 1.092	1.024 / .9218	.1324 / .1119	6.285 / 5.723
Small groceries	SMALLM1D1	2.375 / 1.865	2.341 / 1.741	0 / 0	12 / 8
Small groceries	SMALLM1D2	1.513 / 1.12	1.881 / 1.308	0 / 0	12 / 6
Small groceries	SMALLM2D1	.7245 / .8134	.5746 / .5813	.0047 / .0045	4.252 / 4.199
Small groceries	SMALLM2D2	1.069 / 1.146	1.011 / .980	.0036 / .004	6.195 / 5.991
Medium groceries	MEDM1D1	.6041 / .6099	.8212 / .8516	0 / 0	4 / 4
Medium groceries	MEDM1D2	.3194 / .3191	.5751 / .6014	0 / 0	3 / 3
Medium groceries	MEDM2D1	1.327 / 1.416	.7701 / .8689	.0848 / .0725	4.065 / 4.552
Medium groceries	MEDM2D2	1.791 / 1.902	1.155 / 1.2	.2045 / .0716	6.875 / 7.078
Large groceries	LARGEM1D1	.5138 / .7943	.8192 / 1.065	0 / 0	4 / 5
Large groceries	LARGEM1D2	.3333 / .4326	.6477 / .7955	0 / 0	3 / 5
Large groceries	LARGEM2D1	1.341 / 1.195	.7355 / .7396	.084 / .1241	4.100 / 3.832
Large groceries	LARGEM2D2	1.838 / 1.65	1.086 / 1.173	.0535 / .1930	6.158 / 7.158

definition.

I brainstormed three possible common boundary definitions: the 2000 boundary (the RLIS shapefile, 2000 Census Tracts), the 2010 boundary (the RLIS shapefile, 2010 Census Tracts), and a condensed boundary where the smaller tracts are aggregated to the larger boundaries. If boundary changes occur, the condensed boundary would take the more aggregated tracts in either year. As a result, the condensed boundary would lead to the smallest number of observations. As for adopting the original boundary in either year, some variables would need to be disaggregated since some census tracts would be divided in the adopted boundary definition. Both aggregation and disaggregation of variables require the application of weights; either population weights, or area weights, or both.

The 2010 boundary was adopted as the common boundary definition for several reasons. First, using one of the two original boundaries only requires the transformation of data from one year to the other year, whereas the condensed boundary requires that data from both years be transformed. Furthermore, the 2010 boundary can better reflect the results of neighborhood population growth between 2000 and 2010, since boundary changes occur precisely to adjust for neighborhood population growth. However, some may argue that the condensed boundary is more reliable because it relaxes the assumption that the two tracts in 2010 that split from one tract in 2000 had grown at the same rate over time. So using the 2010 boundary leads to the splitting of data for some 2000 tracts, which may result in an undeserved added degrees of freedom from the increase in the number of tracts. The condensed method should be an alternative approach to use in future research.

To manipulate non-spatial data from tracts in 2000 to fit the 2010 boundary definition, I relied on the methodology created by US2010, a research program from Brown University. US2010 provided all necessary tools to complete the data manipulation: a "crosswalk" dataset and a code program, both in Stata format. All non-spatial variables could be transformed with US2010's methodology. More information on US2010's methodology is presented in Appendix B.2. On the other hand, since the spatial variables were constructed using the centroids of the tracts, they needed to be re-constructed using the tracts in the common boundary definition. The re-construction of the spatial variables followed the same set of procedures as the construction of spatial variables for each year's data, detailed in Appendix A.

3.5 Spatial Distribution of Stores

This section presents a few maps visualizing the locations of groceries in the City of Portland. The groceries in 2000 are represented by diamonds and the groceries in 2010 are represented by circles. The bigger the shape of the store, the larger its employee size classification, defined by Table 3.5

Figure 3.6 and Figure 3.7 present the locations of groceries in 2000 and 2010, respectively. Figure 3.8 through Figure 3.10 compare the existence of groceries by each category between 2000 and 2010. The stores represented by squares in all three maps are stores that existed both in 2000 and 2010.

Figure 3.6: Groceries in 2000

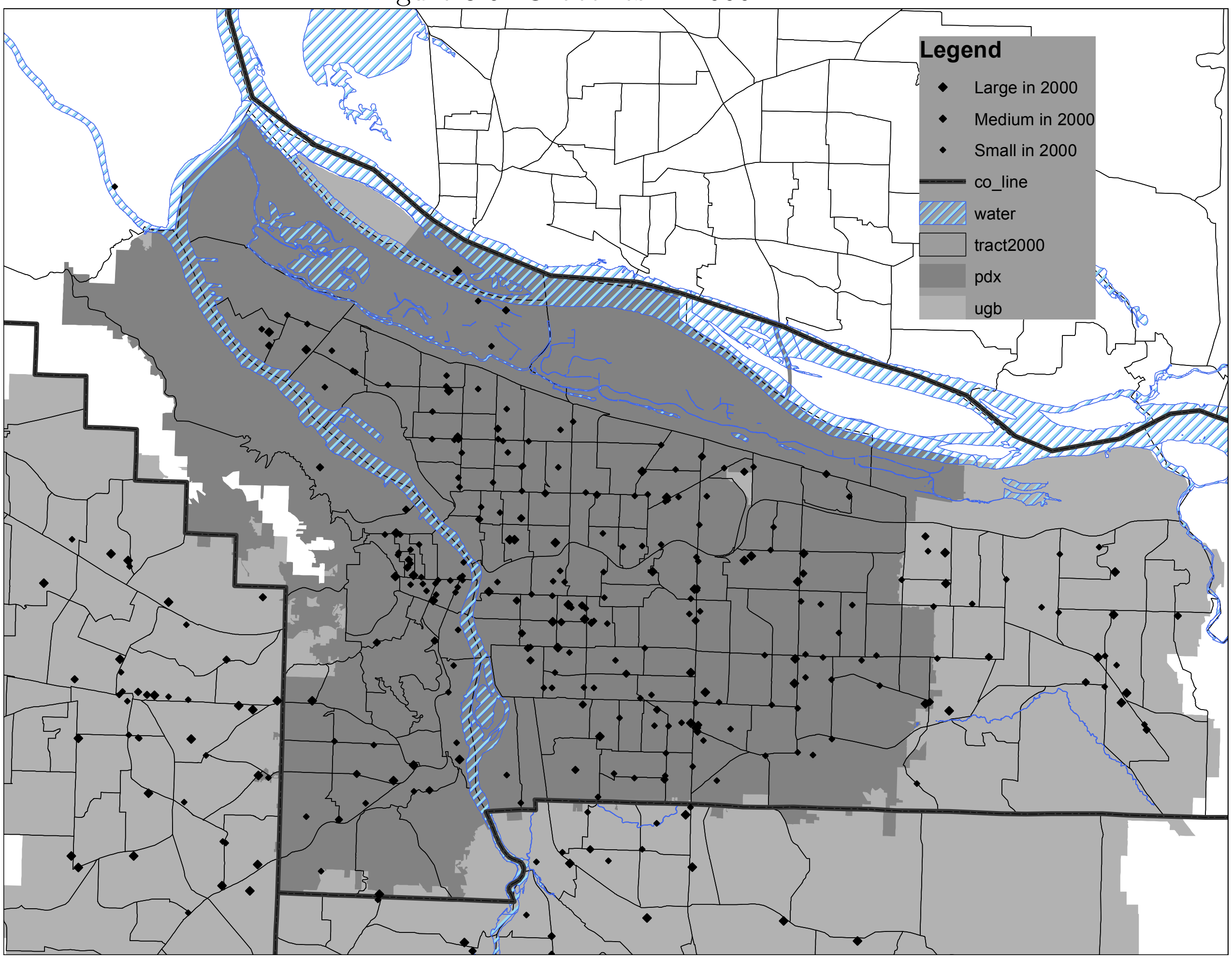


Figure 3.7: Groceries in 2010

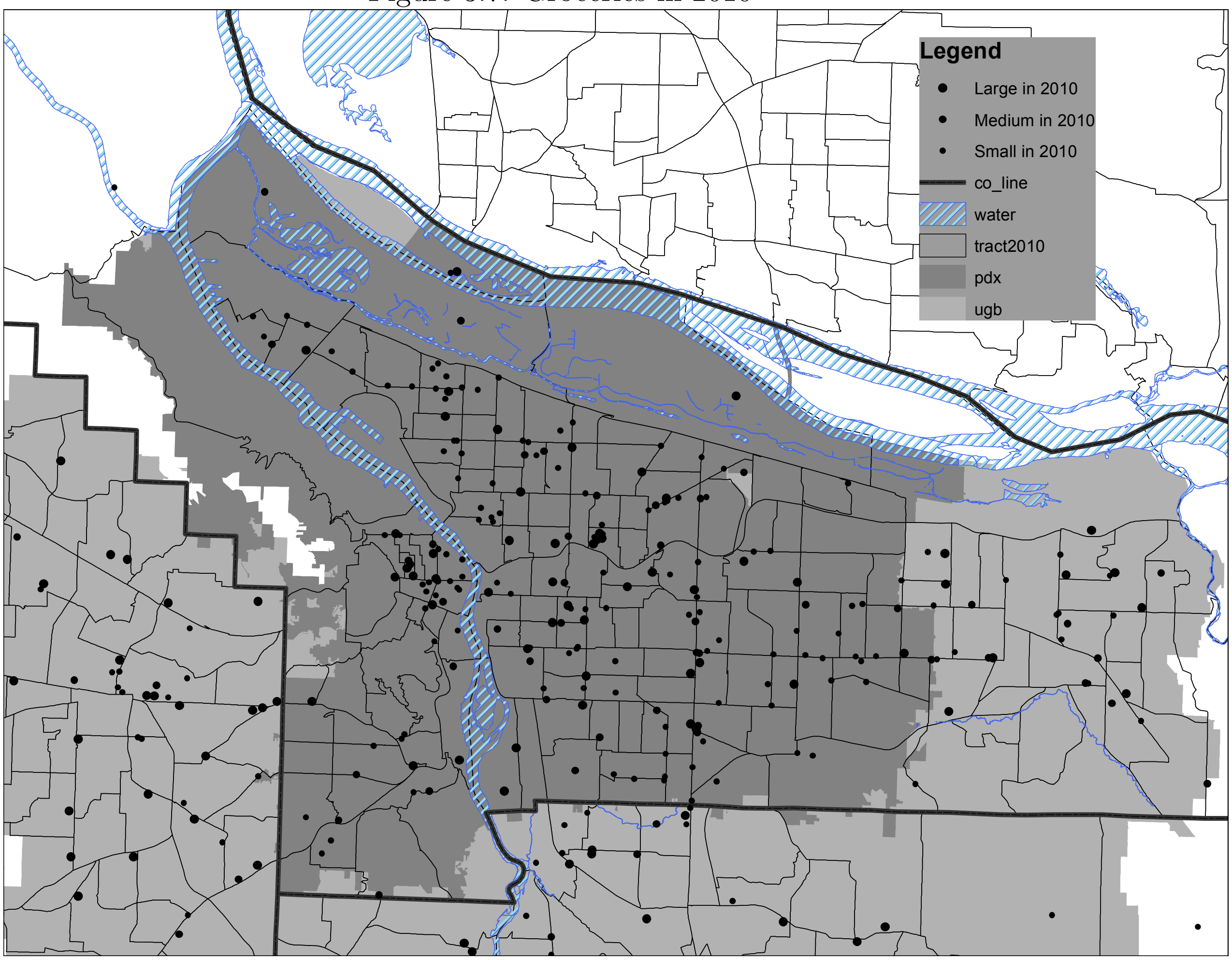


Figure 3.8: Small groceries in 2000 and 2010

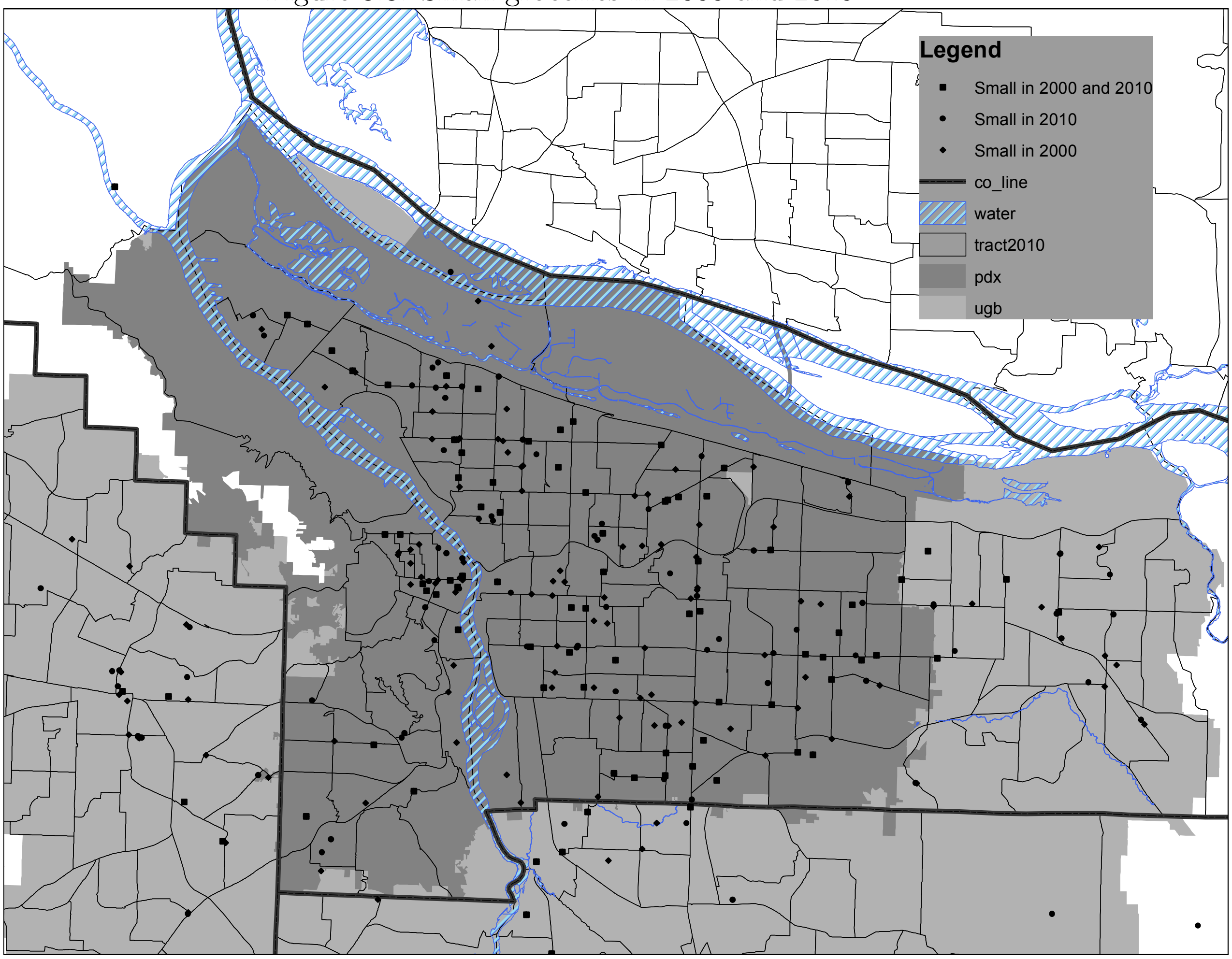


Figure 3.9: Medium groceries in 2000 and 2010

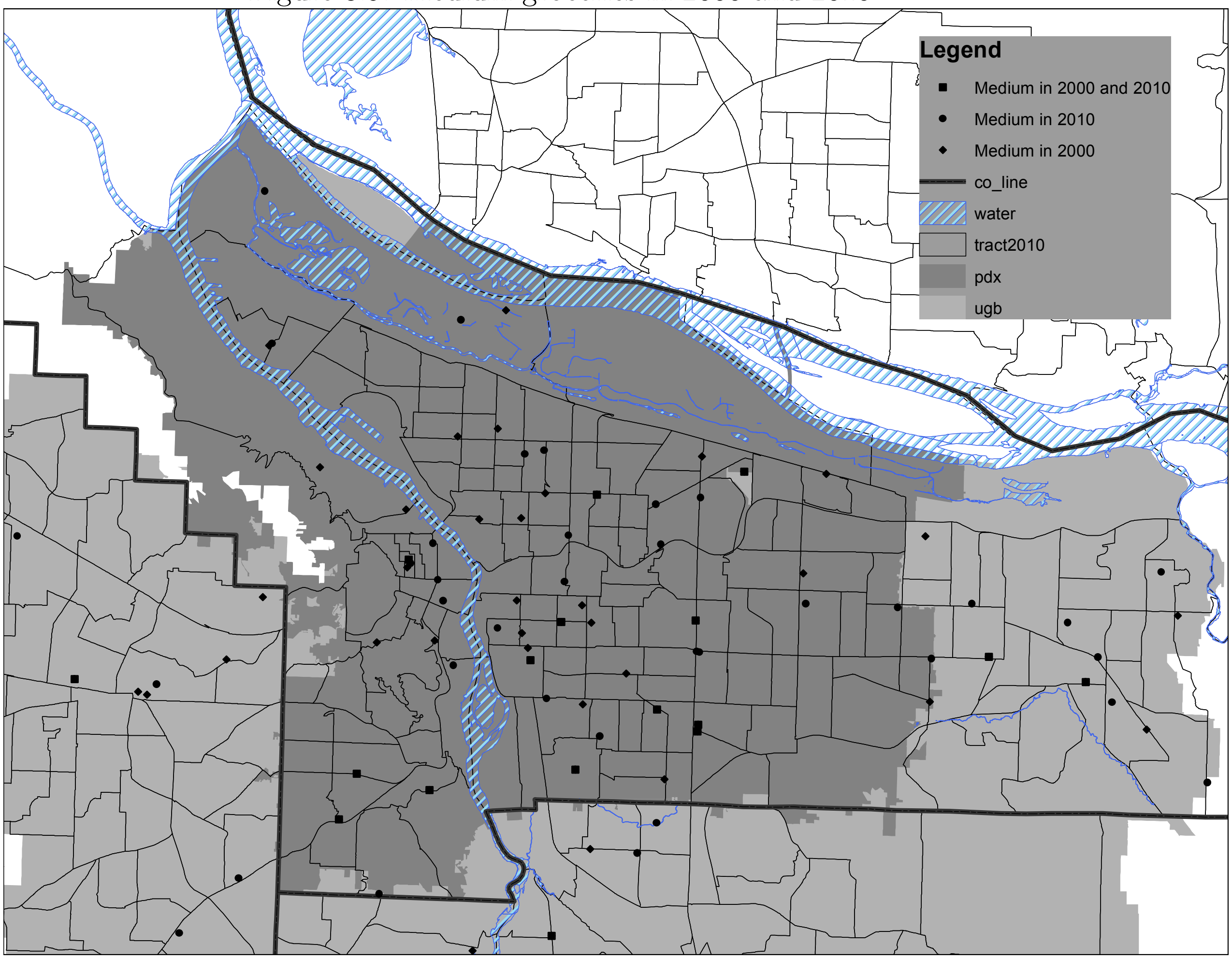
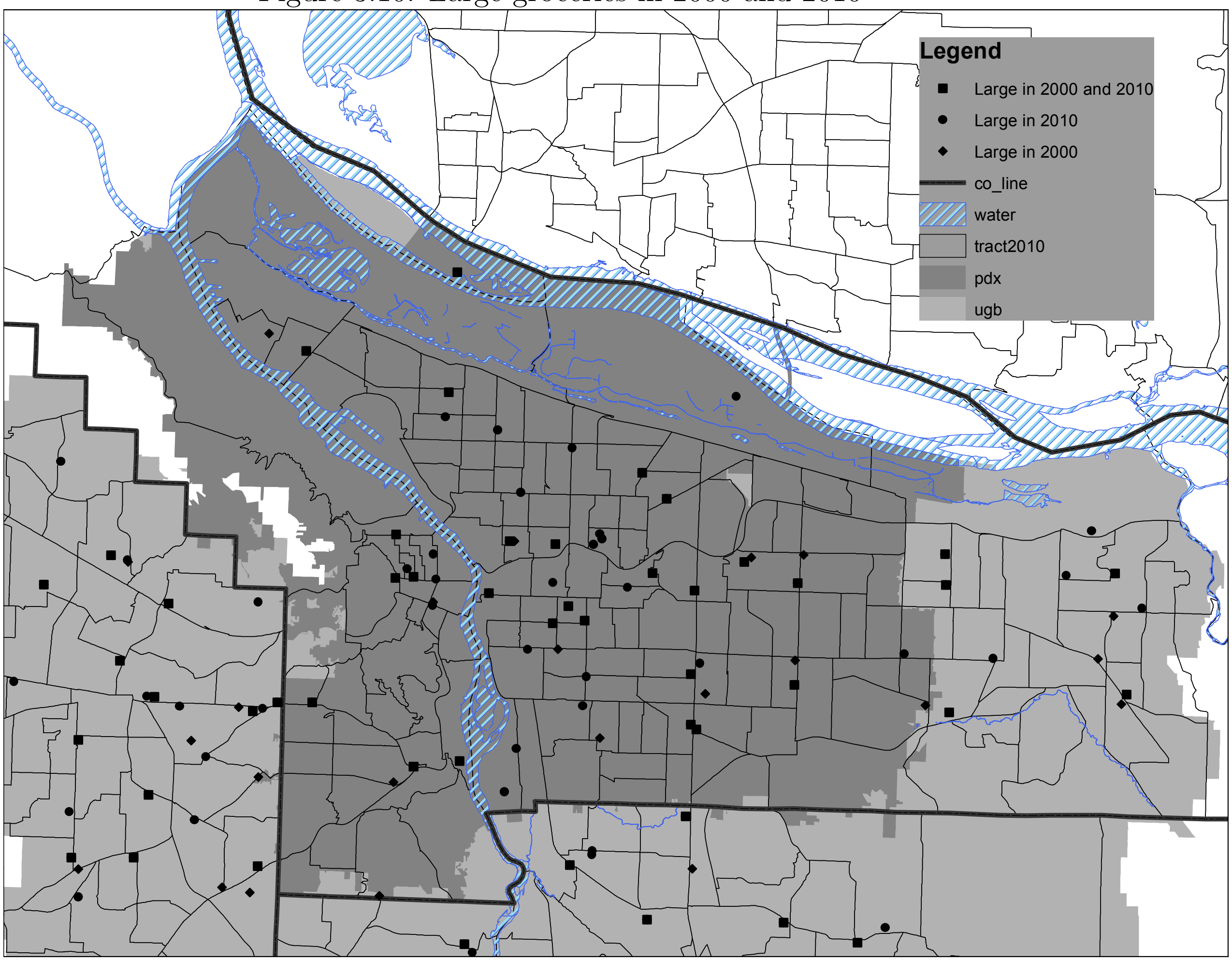


Figure 3.10: Large groceries in 2000 and 2010



Chapter 4

Models and Results

Section 4.1 presents the core model, which is used in Section 4.2 to compare the different variants of dependent variables measuring store access. Finally, Section 4.3 and Secion 4.4 analyze how neighborhood demographics affect access to stores of different sizes and between two years.

4.1 Development of Core Model

The core model is developed using data from 2010 only. The observations are the census tracts defined by the 2010 boundary (n = 141). The dependent variable is the distance to the nearest large grocery on the street network ($LARGEM2D2$). This dependent variable allows the core model to be OLS because it is a continuous variable, as opposed a discrete count variable, which would require a Poisson regression. The adoption of a street network versus a Euclidean distance more realistically depicts the travel distance to stores. Finally, only stores with over 50 employees are chosen in order to narrow the focus to large supermarkets. This would potentially allow more variation in the dependent variable given the relatively smaller number of such stores, each covering a much bigger service area. Note that although $LARGEM2D2$ is positively skewed like most continuous distance variables, taking its logarithmic form leads to a worse overall fit of the model and some variables become insignificant. However, the degrees of freedom is large enough such that the least square estimators are approximately normally distributed even if the dependent variable is not. Therefore, the dependent variable $LARGEM2D2$ remains untransformed in the core model.

Section 4.1.1 through Section 4.1.3 explain the inclusion of three broad groups of variables in the core model, which is developed upon examining the statistical significance of each selected independent variable as well as the fit of the model as a whole based on several statistical indicators. Section 4.1.4 analyzes the results of the developed core model, expressed in the equation below:

$$\begin{aligned} LARGEM2D2 = {} & \beta_0 + \beta_1 DENS + \beta_2 VEH + \beta_3 DENS * VEH + \beta_4 CBD + \beta_5 CBD * CBD + \beta_6 CENTER \\ & + \beta_7 ASIAN + \beta_8 HISP + \beta_9 BLACK + \beta_{10} MEDHHINC + \beta_{11} BLACK * MEDHHINC \\ & + \beta_{12} HIGHSCHNO + \epsilon \end{aligned} \tag{4.1}$$

4.1.1 Spatial Control Variables

The base of the core model is a few control variables that mostly pertain to the spatial characteristics of the census tracts: density, the distance to the Central Business District (CBD), and the distance to the nearest Center. I also examine vehicle access here because it acts as a control variable for the ability to access different parts of the city. Table 4.1 describes the spatial control variables included in the core model.

Table 4.1: Spatial control variables in the core model

Variable Name	**Variable Description**
DENS	Population density (Total pop in tens of thousands/sqkm.)
VEH	Percentage of households with at least one vehicle
DENS*VEH	Interaction of dens10000 and veh1
CBD	Distance (km) to downtown Portland on street network
CBD*CBD	Square of the distance (km) to downtown Portland on street network
CENTER	Distance (km) to the nearest Center on street network

The distance to the nearest Center has a positive and significant effect on the distance to the nearest large grocery. This confirms the hypothesis that the further away one lives from a Center, the longer they need to travel to access a large grocery. Similarly, the distance to CBD is positive and significant, as expected. The scatterplot of the distance to CBD and the dependent variable shows that the distance to a supermarket increases at an increasing rate as the distance to CBD increases. At least one study uses quadratic polynomials to characterize the relationship between distance to CBD and property values in a hedonic price model (Netusil 2013). This prompts an investigation in the non-linearity of the distance to the CBD and the distance to the nearest Center. The model with the distance to the CBD and its square has a slightly higher R^2, but the square of the distance to the nearest Center does not improve the model. As a result, the distance to the CBD is now in a quadratic polynomial (a linear term and a square term) whereas the distance to the nearest Center is kept as a linear term.

Vehicle access, as the only regressor, has a positive effect on the distance to the nearest large grocery. When the model accounts for density, vehicle access is negative and insignificant. This implies that perhaps vehicle access affects the distance to nearest large grocery through its interaction with density; hence the interaction of vehicle access is included. As a result, the model becomes a much better fit based on the increase in adjusted R^2 and F-stat as well as the big drop in both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Therefore, despite the high variance inflation factor (VIF) of density and interaction term, the model with the interaction term is adopted. The VIF is measures the proportion of variation of an independent variable that is explained by variation of other independent variables. A high VIF of a variable indicates a higher proportion explained, which implies that there is more collinearity associated with this variable.

Including $DENS * VEH$ actually makes density have a positive effect on distance to the nearest large grocery, and the interaction has a negative effect. Vehicle access returns to its expected positive sign. According to the partial derivative $\frac{\partial(LARGEM2D2)}{\partial VEH} = \beta_2 + \beta_3 DENS$ derived from Equation 4.1, the effect of vehicle access on distance to store now has two components. Vehicle access itself enables distance to store to be longer, but this effect decreases as the neighborhood becomes denser. In other words, the effect of vehicle access on distance to store is canceled out as density increases. Intuitively, a very dense neighborhood is a much more appealing service area to stores, which makes companies drop the consideration that they do not need to locate in such a neighborhood if it has a high vehicle access rate. In parallel, the effect of density on distance is also broken down in two parts: $\frac{\partial(LARGEM2D2)}{\partial DENS} = \beta_1 + \beta_3 VEH$. Firstly, density itself actually causes a longer distance perhaps due to the aforementioned consideration. Perhaps these big supermarkets prefer to locate not right in the middle of a dense neighborhood, but in a less dense area surrounded by the dense neighborhoods. Perhaps travel time in a really dense neighborhood is longer due to the externality of traffic and congestion created by high density, thus driving away supermarkets (O'Sullivan 2008, 228). Secondly, as the neighborhood's vehicle access improves, density then causes the distance to store to decrease. This is the expected density effect coming through, which is found here to be heavily dependent on vehicle access.

4.1.2 Race and Economic Status

The racial compositions, poverty rate, and median household income are explored here, as they are the basic variables of interest in the two past studies discussed in Section 1.4.

The racial and ethnic composition variables are percentage of African American population, percentage of Asian population, and percentage of Hispanic population. All three percentages are jointly significant. The percentage of Hispanic population ends up not being significant, but is kept in the model for a more comprehensive depiction of the racial and ethnic composition of a neighborhood.

For the indicator of a neighborhood's economic status, Zenk et al. (2005) used the poverty rate while Powell et al. (2007) used the median household income. Since they are highly negatively correlated (correlation coefficient of about -0.7), I explore them as substitutes in the model. The rate of households below poverty is not significant but its close substitute, medium household income, is very significant. As a result, the economic status of a census tract is characterized by the median household income in the core model. Although the logarithmic transformation for the medium household income is often used in econometric models due to its positive skewness, it does not improve the core model in any way. Thus, the logarithmic transformation is not adopted in order to simplify the interpretation of the effects of median household income on store access.

Finally, Zenk et al. (2005) discusses the importance of the interaction between the African American composition and the poverty rate to determine whether these two variables depend on each other to influence the store access of a neighborhood. This prompts the addition of an interaction variable between median household income and the percentage of African American population in the core model. Results show that the increase in distance to nearest large grocery due to an increase in percentage of African American population is mitigated as median household increases. And the greater the African American population, the shorter the distance to nearest large grocery due to an increase in median household income. Table 4.2 lists the racial and ethnic variables as well as economic status variables included in the core model.

4.1.3 More Independent Variables

More independent variables are explored in the core model: age, education, household size, unemployment rate, and median house value. They are either commonly used by previous researchers or are suggested by the City of Portland as being the criteria

Table 4.2: Race and economic status variables in the core model

Variable Name	**Variable Description**
ASIAN	Percentage of total population identified as Asian alone or in combination
BLACK	Percentage of total population identified as Black or African American alone or in combination
HISP	Percentage of Hispanic or Latino in total population
MEDHHINC	Median household income in the past 12 months (in tens of thousands of 2010 inflation–adjusted dollars)
BLACK*MEDHHINC	Interaction of BLACK and MEDHHINC

that companies use to locate their stores. In the end, the only variable included in the core model explored in this section is the percentage of adult population (25 and up) without high school diploma or equivalent ($HIGHSCHNO$).

In terms of the age variables, the percentage of population under 18 and the percentage of population over 65 are not significant when regressed separately or together. Including all four age compositions (leaving one age group out to avoid perfect collinearity) results in the insignificance of each age composition as well as joint insignificance. Therefore, the age variables are not included in the core model.

Both education variables (the percentage of adult population without a high school diploma and the percentage of adult population without a bachelor's degree) have a negative effect on the distance to the nearest large supermarket. The percentage of adult population without a high school diploma is chosen over the other because it possesses more statistical significance and renders a better core model based on the standard model specification tests.

For the household size variables, either the average household size is used or a group of percentages of households of different household sizes (2-person to 7-person and up). Neither the group of percentages of household sizes nor the average household size is jointly significant even at the 10% significance level. Since average household size is hypothesized to be a proxy for residential areas, it is perhaps also a proxy for the distance to the CBD, and is only indirectly affecting access to large groceries. Thus, it makes sense that the household size variables are not significant as the distance to downtown variable is in the model. Therefore, the household size variables are not included in the core model.

The unemployment rate is not included due to its large insignificance. Using median house value to substitute for median household income only results in a minuscule improvement in the VIF. Note that median household value has been argued to characterize the wealth of a neighborhood, as opposed to median household income that characterizes the wealth of the individuals living in the household. Therefore, they are not exactly perfect substitutes in these models. When regressed alongside the median household income, the median house value is insignificant while the median household income remains significant. Since the median house value is not a theoretically critical variable and it is highly correlated with median household income (correlation coefficient of 0.62), it is not included in the core model.

4.1.4 Analysis of the Core Model

According to the final version of the Core Model, the factors that are significantly associated with access to large groceries are density, vehicle access, distance to downtown (CBD), distance to the nearest Center, the Asian and African American compositions, median household income, and education. The only variable with no significant effect in the regression is the Hispanic composition, which has a p-value of 0.289. This variable becomes insignificant with the inclusion of the percentage of adult population without a high school diploma. This indicates that the Hispanic composition is mainly a proxy for the education variable due to a high correlation (correlation coefficient of 0.7271), and does not possess a direct effect on store access by itself. All other variables are significant at a 1% significance level, with the exception of vehicle access which is significant at a 5% significance level. Note that the linear and square terms of distance to CBD are jointly significant at a 1% significant level ($F(2, 128) = 15.67$). Table 4.3 displays the regression results.

The positive sign of the density measure indicates that large groceries tend not to locate within the most dense areas, given the vehicle access of the households in the neighborhoods. However, if over 50% of households in a given census tract have access to a vehicle, then an increase in the population density would result in a decrease in distance to the nearest large grocery (note the interaction between vehicle access and density). For a census tract with a median percentage of households with access to a vehicle (about 90 percent), an increase of ten thousand people in a square-kilometer would cause the distance to the nearest grocery store to decrease by over a third of a kilometer. The magnitude of the density effect increases as vehicle access improves for the census tract. In parallel, a neighborhood's improved vehicle access enables large groceries to locate further away, but this effect decreases as the neighborhood becomes denser. With a median density of roughly 2500 residents per square-kilometer, a one percent increase in the percentage of households with vehicle access on the distance

Table 4.3: Final Core Model: Regression Results

	LARGEM2D2
DENS	4.569***
	(1.425)
VEH	2.429*
	(1.369)
DENS*VEH	-9.370***
	(1.985)
CBD	-0.012
	(0.105)
CBD*CBD	0.011*
	(0.006)
CENTER	0.107**
	(0.044)
ASIAN	4.466**
	(1.871)
HISP	-1.524
	(1.431)
BLACK	12.092***
	(2.632)
MEDHHINC	0.174***
	(0.052)
BLACK*MEDHHINC	-2.013***
	(0.513)
HIGHSCHNO	-3.952***
	(1.385)
_cons	-1.633*
	(0.868)
R^2	0.68
N	141

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

to the nearest grocery is small and insignificant.

The distance to downtown Portland and the nearest Center are both key indicators of access to the nearest large grocery. Both measures have a positive and significant association with distance to the nearest large grocery. For a census tract that is about seven kilometers away from downtown (the median), as it moves one more kilometer away from downtown, the distance to the nearest large grocery increase by about 0.14 kilometers, which is more than ten percent of the median distance to the nearest large grocery. And in general, being one kilometer further away from the nearest Center leads to an increase of roughly 0.11 kilometers in the distance to the nearest Center.

Consider the associations of race and economic status on access to a large grocery. An increase of one standard deviation of the percentage of Asian population (about four percent) means the census tract is further away from the nearest large grocery by almost 0.2 kilometers. For a median census tract with the median household income of about $48,880, an increase of one standard deviation of the percentage of African American population (about seven percent) leads to an increase in the distance to the nearest large grocery of about 0.16 kilometers. These results support past studies that find a negative association between African American composition and food access by showing that the African American composition of a neighborhood still significantly reduces its access to large groceries in neighborhoods where median household income is lower than roughly $60,000. The positive effect of median household income suggests that large groceries do not locate near richer residents, given that the population of African Americans in the census tract is lower than about nine percent, which is the case for over 70% of the census tracts in the sample.

Lastly, a ten percent increase in the population without a high school education is associated with a decrease of approximately 0.4 kilometers in the distance to the nearest large grocery store.

4.2 Comparison of Dependent Variables using Groceries in 2010

I evaluate the four different variants of access measures by substituting them in iterations of the core model developed in Section 4.1. I compare the measures constructed by two different distance constructs, as well as two different types of measures. In the end, one access measure is chosen to be used in Section 4.3 and Section 4.4, where the results and analyses of the main regression models of this thesis are discussed.

4.2.1 Distance Constructs: Euclidean or Street Network?

In this section, access measures from two different distance constructs, Euclidean distance (D1) and street network distance (D2), are compared. I use the distance to the nearest grocery (M2) as the type of access measure that is constructed by the two different distance constructs. Distance to nearest store (M2) allows the use of OLS models, whose goodness of fit can be easily assessed using their R^2 measures. The other type of access measure is the number of stores within a one-kilometer buffer (M1), which requires a model for count data, and thus does not have a R^2 measure. The next section (4.2.2) shows that the number of stores in a buffer (M1) and distance to nearest store (M2) provide analogous results. Thus, the conclusions regarding the distance constructs using M2 can be applied to M1 as well.

Theory suggests that distance measures using the street network distance (D2) are more realistic than those using the Euclidean distance (D1) since people are more likely to travel using constraints set by the street network and not in a straight line between two locations. Sparks, Bania, and Leete (2011) finds that both distance constructs lead to similar identification of food deserts; but the distances measured with street network distance (D2) are longer than those with Euclidean distance (D1).

I compare the measures of distance to the nearest store (M2) built separately from the two distance constructs using all three categories of groceries in 2010. Table 4.4 displays the regression results of the different measures regressed on the same set of independent variables. In all three store categories, most of the independent variables exhibit the same significance in estimating the two different access measures. There are only two exceptions: $HIGHSCHNO$ is significant in estimating the measure of access to small groceries with the street network distance (D2) but not the measure with the Euclidean distance (D1), and the same goes for $CENTER$ in access to large groceries. Only one independent variable has opposite signs when estimating the two measures: $HIGHSCHNO$ in access to medium groceries. But its insignificance in either regression makes the contradictory signs not surprising nor alarming. To conclude, the effects of the independent variables on access measure with the two distance constructs are comparable. Therefore, an analysis of the R^2 can identify the distance construct whose access measure can be estimated more accurately by the same set of independent variables.

In all three store categories, the R^2 is higher in estimating the measure with the street network distance (D2) by at least 0.06. This suggests that the variation in the explanatory variables can explain the proportion of variation in the measures with street network distance (D2) better than explaining the proportion of variation in the measures with Euclidean distance (D1). As a result, the measures with street network distance (D2) are deemed more appropriate for an analysis of access for two reasons: it more realistically depicts travel distance, and it can be better explained using the independent variables constructed in this thesis.

Table 4.4: Comparison of distance constructs (D) using groceries in 2010

Store Category/D1 or D2	SMALL/D1	SMALL/D2	MED/D1	MED/D2	LARGE/D1	LARGE/D2
DENS	2.404**	4.220***	0.020	0.255	2.604***	4.569***
	(1.000)	(1.579)	(1.437)	(1.917)	(0.980)	(1.425)
VEH	0.883	1.551	-1.231	-2.454	2.163**	2.429*
	(0.851)	(1.213)	(1.323)	(1.850)	(0.946)	(1.369)
DENS*VEH	-4.740***	-8.801***	-2.354	-4.047	-5.506***	-9.370***
	(1.401)	(2.194)	(2.047)	(2.714)	(1.430)	(1.985)
CBD	0.029	0.014	0.038	0.108	-0.023	-0.012
	(0.050)	(0.084)	(0.097)	(0.153)	(0.063)	(0.105)
CBD*CBD	0.000	0.004	0.002	0.001	0.007*	0.011*
	(0.003)	(0.005)	(0.006)	(0.009)	(0.004)	(0.006)
CENTER	0.114***	0.226***	0.229***	0.330***	0.034	0.107**
	(0.036)	(0.055)	(0.043)	(0.060)	(0.031)	(0.044)
ASIAN	0.228	0.001	0.042	0.196	3.014**	4.466**
	(1.444)	(1.927)	(1.639)	(2.113)	(1.389)	(1.871)
HISP	-0.326	-1.866	0.129	0.838	-0.129	-1.524
	(0.969)	(1.400)	(1.475)	(1.923)	(1.037)	(1.431)
BLACK	4.559***	6.557***	1.650	2.686	8.018***	12.092***
	(1.563)	(2.428)	(2.713)	(3.957)	(1.983)	(2.632)
MEDHHINC	0.096**	0.118*	0.118***	0.213***	0.077**	0.174***
	(0.047)	(0.061)	(0.043)	(0.057)	(0.031)	(0.052)
BLACK*MEDHHINC	-1.164***	-1.742***	-0.166	-0.505	-1.263***	-2.013***
	(0.260)	(0.424)	(0.520)	(0.732)	(0.372)	(0.513)
HIGHSCHNO	-0.882	-2.351*	0.536	-0.801	-2.547**	-3.952***
	(0.971)	(1.311)	(1.240)	(1.598)	(1.046)	(1.385)
_cons	-0.517	-0.620	1.043	1.650	-1.170*	-1.633*
	(0.467)	(0.741)	(0.948)	(1.324)	(0.625)	(0.868)
R^2	0.47	0.57	0.45	0.51	0.62	0.68
N	141	141	141	141	141	141

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

4.2.2 Access Measures: Number of Stores or Distance to Nearest Store?

Using the more suitable distance construct, street network distance (D2), according to Section 4.2.1, I compare the two types of measures constructed in this thesis: number of stores within a one-kilometer buffer (M1) and distance to the nearest store (M2). A Poisson regression model is adopted to estimate the number of stores (M1) because it is a count measure, whereas the distance measure (M2) is regressed using OLS. I compare the two types of measures using all three categories of groceries in 2010. Table 4.5 displays the regression results of the different measures regressed on the same set of independent variables.

As pointed out in the study done by Powell et al. (2007), the Poisson regression is the appropriate model for count dependent variables, unless there is a significant difference between the mean and the variance of the count measure, the number of stores in a buffer (M1). An alternative model that relaxes the restriction of equal mean and variance is the negative binomial model. This model provides a parameter of over-dispersion, α, that informs whether the variance differs too much from the mean to follow a Poisson model. I first run the M1 regressions in Table 4.5 using the negative binomial model, and I test whether α significantly differs from zero using Stata's built-in likelihood ratio test. For the number of stores (M1) for all three store categories, the hypothesis that α is zero cannot be rejected even at a 25% significance level, so it is safe to assume that there is not an over-dispersion problem that delegitimizes the use of the Poisson model for the measures of number of stores (M1).

Turning back to the Poisson regressions of the number of stores (M1) (displayed in Table 4.5), I conduct some specific goodness-of-fit tests that inform us if the data fit well with the Poisson distribution. For the medium and large groceries, the hypothesis that the data follow a Poisson distribution cannot be rejected even at a 90% significance level based on two statistics for the Poisson model (the deviance statistic and the Pearson statistic). The deviance statistic is described as the R^2 measure for generalized linear models (versus OLS), of which the Poisson regression is one. And the Pearson statistic is a measure of how the observed distribution of the number of stores differs from the theoretical Poisson distribution. The aforementioned hypothesis result shows that the Poisson model is appropriate in estimating the number of stores (M1) for the medium and large groceries. Interestingly, the same hypothesis for the small groceries is rejected at the 5% significance level with the deviance statistic, and rejected at the 10% level with the Pearson statistic. This suggests that the Poission model is not a good fit for the measures of number of stores (M1) for the small groceries.

Theoretically, the estimators of both types of measures should have opposing signs because a longer distance to the nearest store should correspond to a smaller number of stores within a buffer. Results show that this is the case for most of the estimators. And when the estimators have matching signs, they are mostly not significant in the models. Therefore, the independent variables seem to estimate the two types of measures in a consistent fashion. In addition, most of the independent variables exhibit similar levels of significance in estimating the two different access measures. Unfortunately, the Poisson regressions for the number of stores (M1) do not produce a R^2 measure, thus I cannot easily compare the goodness of fit between the M1 and M2 models.

In terms of interpreting the results, the marginal effects of the independent variables on the number of stores (M1) in Poisson are dependent on the values of the independent variables. Therefore, because the effect of each independent variable on the number of stores (M1) is generally comparable to that on the distance to nearest store (M2) in meaning, I adopt the distance to nearest store (M2) in my analysis for a simpler interpretation of results.

Table 4.5: Comparison of access measures (M) using groceries in 2010

Store Category/M1 or M2	SMALL/M1	SMALL/M2	MED/M1	MED/M2	LARGE/M1	LARGE/M2
DENS	-2.838	4.220***	-5.654*	0.255	-4.331	4.569***
	(1.911)	(1.579)	(3.008)	(1.917)	(3.157)	(1.425)
VEH	-2.151	1.551	-3.815	-2.454	-5.810***	2.429*
	(1.531)	(1.213)	(2.820)	(1.850)	(2.226)	(1.369)
DENS*VEH	6.149**	-8.801***	11.153**	-4.047	9.563**	-9.370***
	(2.981)	(2.194)	(4.432)	(2.714)	(4.867)	(1.985)
CBD	-0.119	0.014	0.079	0.108	0.618**	-0.012
	(0.139)	(0.084)	(0.279)	(0.153)	(0.290)	(0.105)
CBD*CBD	0.000	0.004	-0.014	0.001	-0.063***	0.011*
	(0.008)	(0.005)	(0.018)	(0.009)	(0.023)	(0.006)
CENTER	-0.040	0.226***	-0.388***	0.330***	-0.205	0.107**
	(0.086)	(0.055)	(0.137)	(0.060)	(0.141)	(0.044)
ASIAN	1.983	0.001	3.959	0.196	0.070	4.466**
	(2.297)	(1.927)	(4.380)	(2.113)	(5.137)	(1.871)
HISP	-1.966	-1.866	1.941	0.838	4.929	-1.524
	(2.225)	(1.400)	(4.574)	(1.923)	(4.177)	(1.431)
BLACK	-5.598*	6.557***	-4.550	2.686	-20.861***	12.092***
	(2.982)	(2.428)	(10.268)	(3.957)	(6.941)	(2.632)
MEDHHINC	-0.100	0.118*	-0.072	0.213***	-0.164	0.174***
	(0.077)	(0.061)	(0.116)	(0.057)	(0.114)	(0.052)
BLACK*MEDHHINC	1.612***	-1.742***	0.280	-0.505	3.402***	-2.013***
	(0.549)	(0.424)	(1.895)	(0.732)	(1.178)	(0.513)
HIGHSCHNO	4.318**	-2.351*	-4.282	-0.801	2.749	-3.952***
	(2.075)	(1.311)	(3.425)	(1.598)	(4.161)	(1.385)
_cons	2.163**	-0.620	2.819	1.650	3.239**	-1.633*
	(0.932)	(0.741)	(1.724)	(1.324)	(1.375)	(0.868)
N	141	141	141	141	141	141
R^2		0.57		0.51		0.68

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

4.3 Comparison of Stores Categories

In this section, the effects of neighborhood characteristics on store access are examined across stores categories of different employee sizes. According to Small and McDermott (2006), "large, established stores may steer clear of high poverty areas." Additionally, the same authors argue that an increase in poverty may lead to more access to small stores. Their results suggest that the directions of the effects of poverty on stores are different depending on the store size. Therefore, it is imperative to test their theory, as the results would have contrasting policy implications. More broadly, a comparison of the effects of other characteristics on stores in different categories can paint a more comprehensive picture of store access.

4.3.1 Access by Store Categories in 2010

Table 4.6 presents the regressions of the core model for access to stores of different categories in 2010. Firstly, note that the results for store access to medium groceries have a lower fit relative to the other regressions. The R^2 measure of goodness-of-fit is lower, the F-test shows a smaller significance of the model overall, and fewer estimators are significant. And the insignificant estimators render their signs unreliable as well. This may be caused by a smaller number of medium groceries in the sample with which access measures were constructed. There were 80 medium groceries in 2010, compared to 211 small groceries and 134 large groceries. From now on, I focus my analysis on the comparison between large and small groceries.

Comparing the small groceries to the large groceries, the signs of the effects of neighborhood conditions are the same. (For a detailed analysis of the regression for large groceries, refer to the analysis of the core model in Section 4.1.4.) Yet, there are certain pronounced differences in the magnitudes of these effects. For example, a one-kilometer increase in the distance to the nearest Center is associated with the increase in distance to the nearest small grocery by 0.226 kilometers, which is more than twice the effect for large groceries. The most drastic difference is the Asian composition. Whereas a larger Asian composition leads to a large and significant increase in distance to the nearest large grocery (a one standard deviation increase is associated with an increase in distance of 0.16 kilometers), it has no effect on the distance to the nearest small grocery.

To directly compare the results of the variables with interaction terms, I calculate their marginal effects based on their median values (Table 4.7). These variables are CBD (squared), $DENS$, VEH, $BLACK$, and $MEDHHINC$. Out of these four

Table 4.6: Access by store categories in 2010

	SMALLM2D2	MEDM2D2	LARGEM2D2
DENS	4.220***	0.255	4.569***
	(1.579)	(1.917)	(1.425)
VEH	1.551	-2.454	2.429*
	(1.213)	(1.850)	(1.369)
DENS*VEH	-8.801***	-4.047	-9.370***
	(2.194)	(2.714)	(1.985)
CBD	0.014	0.108	-0.012
	(0.084)	(0.153)	(0.105)
CBD*CBD	0.004	0.001	0.011*
	(0.005)	(0.009)	(0.006)
CENTER	0.226***	0.330***	0.107**
	(0.055)	(0.060)	(0.044)
ASIAN	0.001	0.196	4.466**
	(1.927)	(2.113)	(1.871)
HISP	-1.866	0.838	-1.524
	(1.400)	(1.923)	(1.431)
BLACK	6.557***	2.686	12.092***
	(2.428)	(3.957)	(2.632)
MEDHHINC	0.118*	0.213***	0.174***
	(0.061)	(0.057)	(0.052)
BLACK*MEDHHINC	-1.742***	-0.505	-2.013***
	(0.424)	(0.732)	(0.513)
HIGHSCHNO	-2.351*	-0.801	-3.952***
	(1.311)	(1.598)	(1.385)
_cons	-0.620	1.650	-1.633*
	(0.741)	(1.324)	(0.868)
R^2	0.57	0.51	0.68
N	141	141	141

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

variables, only an increase in density has similar magnitudes to which it draws closer large groceries and small groceries, for a census tract with around 90% of households with vehicle access. With a one-kilometer increase in distance to downtown, the increase in distance to the nearest large grocery (0.145 kilometers) is more than twice as much as that to the nearest small grocery (0.068 kilometers). On the other hand, the African American composition leads to opposing effects on distance to the nearest small versus large groceries. As noted in Section 4.1.4 with the discussion of the core model, one standard deviation of the African American composition (about seven percent) is associated with the increase in the distance to the nearest large grocery of 0.16 kilometers, given a median household income of $48,880. In contrast, the same increase in the African American composition is correlated with a decrease in the distance to the nearest small grocery by almost 0.14 kilometers. This suggests that small groceries are more accessible in neighborhoods with more African Americans, and large groceries are less accessible. Lastly, the marginal effects of vehicle access and median household income are insignificant for both small and large groceries.

Table 4.7: Marginal effects of variables with interaction terms for store categories in 2010

Variable x	**Median**	$\frac{\partial(\mathbf{SMALLM2D2})}{\partial \mathbf{x}}$	$\frac{\partial(\mathbf{LARGEM2D2})}{\partial \mathbf{x}}$
CBD	7.335258	.0684675***	.1445305***
DENS	.2556868	-3.62763***	-3.785902***
VEH	.8916894	-.6988846	.032894
BLACK	.054	-1.956519*	2.254289**
MEDHHINC	4.888	.0235449	.0653451

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

4.3.2 Access by Store Categories in 2000

As with the medium groceries in 2010, the estimators for the medium groceries in 2000 are mostly insignificant. The variation in the estimators can explain the variation in the access measures for medium stores more than 10% less than the variation in the access measure of other store categories. Hence the medium groceries are omitted from the analysis. For the variables without an interaction term, they exhibit mostly similar effect sizes and levels of significance. The Asian and Hispanic compositions are insignificant in all regressions. A one-kilometer increase in distance to the nearest Center causes both the distance to the nearest large grocery and small grocery to increase by about 0.17 kilometers. The percentage of population without a high school education only reduces the distance to the nearest small grocery (by about 0.29 kilometers with a 10% increase), and not the large grocery. Therefore, I focus on the analysis of the marginal effects of the variables with interaction terms, displayed in Table 4.8.

As is the case for the 2010 stores, the extent to which shorter distances to downtown is comparable between small and large groceries. With a median distance to downtown of roughly seven kilometers, as distance to downtown increases by one standard deviation (3.77 kilometers), distance to the nearest small or large store increases by about 0.24 kilometers. Moreover, the extent to which an increase of ten thousand population per square-kilometer (density) improves store access is also comparable between small and large groceries (over a third of a kilometer), given the median percentage of households with vehicle access (89%).

For a neighborhood with a median density, better vehicle access is associated with a decrease in the distance to the nearest small grocery, but has no significant effect on the large grocery. More precisely, a ten percent increase in the population of households with vehicle access draws closer the nearest small grocery by about 0.22 kilometers.

On the other hand, the African American composition in a census tract with a median household income does not seem to have an effect on access to small groceries, whereas a one standard deviation increase (about 12%) dramatically increases the distance to the nearest large grocery by about 0.39 kilometers. This suggests that in 2000, the locations of large groceries were more segregated than small groceries.

Lastly, whereas income did not have significant marginal effects on access to either small or large groceries for a census tract of median African American composition in 2010, it has a strong correlation with both store categories in 2000. For small groceries, a one standard deviation increase in median household income (about $22868) leads to an increase in the distance to the nearest small grocery by close to half a kilometer. In other words, the poorer the neighborhood, the more access to small groceries. A similar but smaller effect is found for the large groceries: about 0.31 kilometers. Thus, poorer neighborhoods actually draw closer groceries regardless of store size.

Table 4.8: Marginal effects of variables with interaction terms for store categories in 2000

Variable x	Median	$\frac{\partial(\textbf{SMALLM2D2})}{\partial \mathbf{x}}$	$\frac{\partial(\textbf{LARGEM2D2})}{\partial \mathbf{x}}$
CBD	7.19416	.0618026*	.0650448**
DENS	.2437021	-3.804344***	-3.358817***
VEH	.8906413	-2.214914**	.7669369
BLACK	.0375	-.676885	3.20442***
MEDHHINC	5.035965	.2168168***	.1341868**

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

4.4 Access to Stores between 2000 and 2010

When comparing the correlations of neighborhood traits and store access across different store categories, I have hinted at the similarities and differences in trends between 2000 and 2010. In this section, I proceed to examine the access to stores between these two years in detail, utilizing a fixed-effects model on top of OLS. I analyze access to large groceries using the distance to the nearest large grocery on a street network ($LARGEM2D2$) as my dependent variable. Then, I analyze the same model for the small groceries.

For each of the two store categories, I run the OLS regression models for separate years, one OLS model with the observations from both years pooled together, and one fixed-effects model (a total of four regressions). Note that all regressions run in this section use the panel dataset, which is built based on only the 2010 Census Tract Boundary for a unified identification of census tracts. Therefore, the summary statistics of the variables in 2000 would vary slightly from the dataset used in the previous sections. For more information on the panel dataset, please refer to Section 3.4 and Appendix B.2.

The pooled OLS disregards the panel nature of the dataset, although the use of cluster-robust standard errors allows for individual errors between years to be correlated. The fixed-effects model takes it further by capturing and controlling for individual-specific and time-invariant characteristics in the intercepts. More specifically, the fixed effects model creates estimators that are only based on the variation of variables within individual census tracts between the two years.

The same core model is used to run model iterations in this section, except for the removal of distance to downtown (CBD) and distance to nearest Center ($CENTER$) and the addition of proportion of bike commuters ($BIKE$). The CBD

and $CENTER$ terms are dropped because they cannot be included in the fixed effects model due to their time-invariant nature. $BIKE$ is added to be a proxy for CBD, since there is a strong correlation between the two variables (correlation coefficient of -0.6 in 2010). This is a reasonable substitution since we expect neighborhoods closer to downtown to have more bike commuters. The core model with $BIKE$ produces fewer significant estimators. However, the significant effects are consistent in signs with the corresponding results in the regressions using the previous core model.

4.4.1 Access to Large Groceries between 2000 and 2010

The OLS models for separate years and the pooled model have almost the same set of significant estimators, shown in Table 4.9. Looking at the marginal effects of median characteristics of census tracts in Table 4.10, only a higher density (with a median proportion of households with a vehicle), a lower proportion of households with a vehicle (with a median density), and a lower proportion of African Americans (given a median of the median household income) are consistently correlated with better access to large groceries with significance, demonstrated by a decrease in distance to the nearest large grocery. Interestingly, higher vehicle access in this version of the core model is significantly correlated with longer distance to store, highlighting the hypothesis that large groceries do not need to locate near neighborhoods where residents are able to drive to shop, whereas the marginal effects of this variable in Section 4.3 were not significant.

Considering the fixed-effects model, the joint F-test of the hypothesis that all errors of variation between census tracts are zero is rejected at the 1% significance level, which means that the variation of individual-specific time-invariant traits of census tracts are significant for at least some census tracts, and the use of the fixed-effects model is apt. This also implies that there are differences between the pooled OLS and the fixed-effects model.

To understand the impact of including fixed effects on the effect sizes of estimators, I compare the results of the pooled OLS and the fixed-effect model. The effect sizes of all significant estimators are similar to those from the pooled OLS, with the exception of the African American composition. One peculiar difference is that the coefficient for the interaction between African American composition and median household income has a positive sign, which is discussed in further detail at the end of this section. The effect size of the same increase in the proportion of African American residents in the fixed effects is more than twice as much as in the pooled OLS, given the median of

Table 4.9: Access to Large Groceries between 2000 and 2010

Year/Type	2000/OLS	2010/OLS	Both/Pooled	Both/Fixed
DENS	10.898***	7.460***	8.592***	-0.311
	(2.456)	(2.008)	(1.970)	(2.987)
VEH	7.122***	5.654***	6.101***	3.450**
	(1.591)	(1.295)	(1.204)	(1.716)
DENS*VEH	-17.932***	-14.716***	-15.594***	-4.852
	(3.355)	(2.913)	(2.799)	(3.845)
BIKE	0.533	-2.215	-2.906**	-2.193*
	(3.383)	(1.474)	(1.148)	(1.109)
ASIAN	-0.055	5.922*	3.191	0.763
	(2.478)	(3.145)	(2.242)	(1.710)
HISP	-0.742	0.251	-0.687	-1.058
	(2.105)	(1.638)	(1.153)	(1.471)
BLACK	3.920**	9.617***	6.631***	1.047
	(1.837)	(3.176)	(1.725)	(2.116)
MEDHHINC	0.094	0.133*	0.110	0.060
	(0.085)	(0.080)	(0.075)	(0.055)
BLACK*MEDHHINC	-0.109	-1.461**	-0.675*	1.522**
	(0.412)	(0.663)	(0.389)	(0.602)
HIGHSCHNO	1.587	-0.968	0.354	-0.271
	(1.596)	(1.449)	(1.093)	(1.057)
_cons	-4.284***	-3.231***	-3.419***	-0.984
	(1.301)	(0.975)	(0.929)	(1.528)
R^2	0.49	0.55	0.51	0.41
N	141	141	282	282

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4.10: Marginal effects of variables on distance to the nearest large grocery in 2000 and 2010

Variable x	**Median in '00/'10**	$\frac{\partial(\text{LARGEM2D2})}{\partial x}$ **in 2000**	$\frac{\partial(\text{LARGEM2D2})}{\partial x}$ **in 2010**	**Median in Panel**	$\frac{\partial(\text{LARGEM2D2})}{\partial x}$ **in Pooled**	$\frac{\partial(\text{LARGEM2D2})}{\partial x}$ **in Fixed**
DENS	.243 / .256	-5.133***	-5.662***	.249	-5.328***	-4.642**
VEH	.894 / .892	2.768**	1.892**	.893	2.224***	2.243*
BIKE	.011 / .041	-	-	.0197	-2.906**	-2.193**
ASIAN	.065 / .071	-	5.922*	.068	-	-
HISP	.058 / .066	-	-	.0598	-	-
BLACK	.037 / .054	3.370***	2.476**	.043	3.273***	8.619***
MEDHHINC	5.06 / 4.89	-	-	4.98	-	.125**
HIGHSCHNO	.128 / .081	-	-	.104	-	-

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

the median household income. This indicates that merely the change in the African American composition in a census tract between 2000 and 2010 is able to characterize the negative correlation between African American composition and access to large groceries (a longer distance to the nearest large grocery). In the pooled model, as the proportion of African Americans of a census tract increases by one standard deviation (about nine percent), the census tract is more than 0.3 kilometers further away from a large grocery. Whereas with fixed effects, as the proportion of African Americans of a census tract increases by the same amount, the census tract is more than 0.8 kilometers further away from a large grocery.

Considering the correlation with poverty, whereas the marginal effect of median household income is not significant in the pooled model, it is significant and positive at a 5% significant level in fixed effects. In particular, as the median household income of a census tract increases by one standard deviation (about $21400), the census tract is about 0.27 kilometers further away from a large grocery.

These results have the following implications about access to large groceries:

- When individual-specific characteristics of census tracts that do not change over time are controlled, the changes of African American composition and poverty (with median household income as proxy) within each census tract over time are more pronounced.
- Results from both OLS and fixed effects suggest that in contrast with the poverty hypothesis from Small and McDermott (2006) in Section 1.3, more poverty (a smaller median household income) is correlated with better access to large groceries (shorter distance to the nearest large grocery).
- Results from both OLS and fixed effects suggest in accordance with the race hypothesis from Small and McDermott (2006), more proportion of African American residents is correlated with more insufficient access to large groceries, as illustrated by the increase in distance the nearest large grocery.
- There is a switch of signs between the statistically significant interactions of the race and poverty, which conveys different conclusions presented below.

In OLS

  - The estimator of the interaction is negative, suggesting the following:
  - As African American composition increases, the degree of association of more poverty with better access to large groceries decreases.
  - As poverty increases, the degree of association of more African American composition with worse access to large groceries increases.
  - The results go along with the interaction hypothesis from Small and Mc-

Dermott (2006) by saying that increasing African American composition worsens the association between poverty and access.

In Fixed Effects

- The estimator of the interaction is positive, suggesting the following:
- As African American composition increases, the degree of association of more poverty with better access to large groceries increases.
- As poverty increases, the degree of association of more African American composition with worse access to large groceries decreases.
- The results do not match with the interaction hypothesis from Small and McDermott (2006) by saying that increasing African American composition actually improves the association between poverty and access.

4.4.2 Access to Small Groceries between 2000 and 2010

The results of the same set of regression models for the small groceries are displayed in Appendix D.0.1. The insignificance of the African American compositions and median household income terms in the fixed-effects model renders it hard to provide significant implications for the relationship between the race and poverty factors and access to small groceries.

Conclusion

A thorough econometric examination of access to groceries in the City of Portland in 2000 and 2010 was conducted in this thesis. Prior to the regression analysis, a spatial mapping of stores to census tracts was conducted using ArcGIS software. Although this thesis only analyzed store access in depth for the small and large groceries in the City of Portland, it has generated all the data and preliminary examination of access to stores for at least one other store category (convenience stores).

The results of the econometric model generally confirm the hypotheses developed in the literature review. In terms of measures of spatial access to stores, the thesis finds that the distance to the nearest store on the street network distance renders a best fit for the models used. In examining neighborhoods traits that are neither race nor poverty, the results show that the distance to a CBD and/or areas with high employment and retail opportunities is a key indicator of access to the nearest large grocery. The further a census tract is from downtown and/or a Center, the further it is from the nearest large grocery. Furthermore, vehicle access only has significant marginal effects on access to large groceries when the distance to downtown is dropped. More specifically, the neighborhoods with more vehicle access have longer distances to the nearest large grocery. A higher density is consistently associated with a shorter distance to the nearest grocery, regardless of store size. In terms of level of education, more proportion of residents without a high school diploma is associated with a shorter distance to the nearest large grocery.

In general, the African American composition is negatively correlated with access to stores while poverty is positively correlated with access to stores, for large groceries. The same correlation with poverty is reflected for small groceries as well. However, the African American composition is positively correlated with access to small groceries in both 2000 and 2010. This suggests that small groceries are more accessible in neighborhoods with a higher proportion of American Americans, and large groceries are less accessible. On the other hand, median household income increases the distance to the nearest small and large grocery if the population of African

American in the census tract is low. This suggests that poverty may be correlated with better access to groceries regardless of store size.

For future research, an examination of the functional form and non-linearity is necessary, as certain variables may not follow a normal distribution. For example, median household income has a tendency to be skewed to the right, so it should be put in log form to approximate a normal distribution. Next, variables such as percentage of bike commuters, percentage of single-parent households, and percentage of foreign-born population should be further explored. The interest in bike commuters is explained in Section 3.2.1, whereas percentage of single-parent households and percentage of foreign-born could potentially affect the kinds of stores located in a neighborhood. For example, if a neighborhood has relatively more mom-and-pop grocery stores versus big supermarkets (just in terms of employee size), the mom-and-pop stores may mostly be stores that sell different ethnic food products to serve the surrounding ethnic enclaves. To be further informed of the grocery access in the City of Portland, the URA and Transit Access indicators should be included in the model and examined. The City of Portland has also created a measure of gentrification risk, which could be a potential factor to include, especially in a panel setting with more time periods. In terms of measures of food access, incorporating evaluators of price and product availability of the groceries as well as alternative food sources such as farmer's markets would help arrive at a deeper understanding of food access.

Another serious problem that this thesis has not accounted for is endogeneity. Endogeneity very likely exists in the models developed by the thesis. A fundamental assumption this thesis rests on is that neighborhood socio-economic and demographic factors influence the presence and absence of stores within the neighborhoods, but not vice versa. However, it is highly plausible that the stores that locate in a neighborhood also influence the characteristics of such neighborhood, leading to a correlation between the error term of the dependent variable of store access and the independent variable representing the neighborhood characteristic. In future research, a Hausman test should be performed, and a model with instruments should be adopted in the presence of endogeneity. Finally, to develop a sound reliable model is to account for spatial autocorrelation using spatial autoregressive models. Accounting for spatial autocorrelation should lead to a significantly more robust model, as the model would provide insight about the kinds of factors that would always affect a neighborhood's access to stores, regardless of the characteristics of the surrounding neighborhoods. In terms of external validity, the models conducted in this study may not apply to other cities. Accounting for more city-specific traits should be taken seriously in any application of the results of this thesis to other cities. Finally, as acknowledge before, the data of this study were clearly observational, which renders it difficult to establish causal relationships.

Appendix A

Construction of Variables with ArcGIS

When working with spatial data, I used two applications of ArcGIS Desktop 10.1: ArcCatalog and ArcMap. ArcCatalog was used for the general management of my geodatabase. Most of the work was performed in map documents (.mxd format) in ArcMap.

Each map document created has a corresponding Stata master dataset. The compilation of the master datasets is detailed in Appendix B. Each map document has a unique combination of tract boundary definition (2000 boundary or 2010 boundary), geographic scope (UGB or City of Portland), and temporal scope (2000 or 2010). For example, the spatial data based on the 2000 boundary (as defined in the RLIS shapefile, 2000 Census Tracts) regarding information within the City of Portland in 2000 were compiled in one map document, tct00_pdx_00.mxd.

For the panel dataset, a new map document, tctcw_pdx_00.mxd, was created using the 2010 boundary (as defined in the RLIS shapefile, 2010 Census Tracts), but constructed access to stores in 2000 and other spatial characteristics in 2000. The spatial variables generated in tctcw_pdx_00.mxd were eventually appended in Stata to the 2010 dataset (2010 boundary, City of Portland, 2010 data) to create the panel dataset, tctcw_pdx_0010.

The RLIS data use the State Plane North projection (NAD1983 HARN Stateplane OR North FIPS 3601 Feet Intl) and a coordinate system built on the North American Datum of 1983 (geographic coordinate system = GCS North American 1983 HARN). I used these same projection/coordinate system settings for all of my spatial data.

A.1 Manual Generation of Shapefiles

In tct00_pdx_00.mxd, I used the shapefiles from the RLIS and the City of Portland to generate some new shapefiles that were essential to constructing the independent and dependent spatial variables. A feature dataset (another way of collecting spatial data) was also generated for building the street network. The same procedure was repeated for tct10_pdx_10.mxd. On the other hand, tctcw_pdx_00.mxd adopted the shapefiles that were generated from the first two map documents. The generation for each shapefile is explained in this section. Refer to Table A.1 to for a summary.

Table A.1: Manually Generated Shapefiles

Filename of Shapefile	**Description or Usage of Shapefile**	**Filename of Shapefiles Used**
streets_ND (feature dataset)	To create all spatial data that rely on the street network	Streets
pdx.shp	City of Portland polygon	City Limits
cent00.shp	Centroids of 2000 Census Tracts	2000 Census Tracts
cent00_pdx.shp	Centroids of 2000 Census Tracts in Portland	cent00.shp, pdx.shp
cent00_pdx_buf1km.shp	Euclidean 1km-buffers of cen00_pdx.shp	cent00_pdx.shp
cent00_pdx_ser1km.shp	Street network 1km-buffers of cent00_pdx.shp	cent00_pdx.shp
center.shp	Centroids of Centers	Analysis Centers
cbd.shp	The Portland CBD polygon	Analysis Centers
cbd_cent.shp	Centroid of the Portland CBD polygon	cbd.shp
ura00.shp	All URA that existed in 2000	URA
ura00lag.shp	All URA that existed at least two years before 2000	ura00.shp

Street Network Dataset [street_ND]

To generate all spatial data that rely on the street network, a network dataset called "streets_ND" was created using the RLIS Streets shapefile.

1. In ArcCatalog, right-click to create a New File Geodatabase named "arcgis.gdb" for storing the all information for the network datasets.
2. In ArcCatalog, create a feature dataset called "streets."
   (a) Right-click on arcgis.gdb to create a new Feature Dataset named "streets."

   (b) In the immediate step in creating the feature dataset, edit the coordinate system settings to be the same as the one used in the RLIS shapefiles.
3. In ArcMap, create a feature class called “streets.”
   (a) In tct00_pdx_00.mxd, right-click on the RLIS Streets shapefile to Data, and choose Export Data.
   (b) Save the exported data to a Feature Class called “streets” inside the “streets” feature dataset.
4. In ArcCatalog, add the Network Analyst extension.
5. Right-click on the feature dataset “streets” to create a new Network Dataset.
   (a) Save the network dataset as streets_ND.
   (b) Use the “streets” feature class.
   (c) “Yes” to Global Turns, enabling the presence of a turn at every transition between two edges in the network.
   (d) “Any Vertex” for Connectivity Policy, as opposed to “End Point,” which would only allow streets to be connected to other streets at their end points.
   (e) “No” to Elevation because the RLIS Streets shapefile does not contain information about elevation.
   (f) “No” Historical Travel Data because the RLIS shapefile does not contain information regarding historical travel data such as travel time.
   (g) “Length” for Attribute of Evaluators, as distances are measured using this street network.
   (h) In Directions, use “feet” as Units.

City of Portland polygon [pdx.shp]

1. Select By Attribute the polygons from the RLIS City Limits shapefile.
2. Select the polygon with “CITYNAME = Portland.”
3. Save the selection as pdx.shp.

Centroids of 2000 Census Tracts [cent00.shp]

1. Add two fields to the table for the RLIS 2000 Census Tracts shapefile named x and y in “Double” format.
2. Right-click on x to Calculate Geometry as the “X coordinate of centroid.”
3. Right-click on y to Calculate Geometry as the “Y coordinate of centroid.”
4. Export the table to the map document, right-click on the table to Display XY Data.

5. Export the displayed data as cent00.shp.

Centroids of 2000 Census Tracts in Portland [cent00_pdx.shp]

1. Select By Location all centroids from cent00.shp that are within pdx.shp.
2. Export the selection as cent00_pdx.shp.

Euclidean 1km-buffers of cent00_pdx.shp [cent00_pdx_buf1km.shp]

1. Click Geoprocessing on the toolbar for Buffer.
2. Load cent00_pdx.shp as the Input Features.
3. Save the Output Features as cent00_pdx_buf1km.shp.
4. Set the Distance to "3280.84 ft," which is equal to one kilometer.

Street network 1km-buffers of cent00_pdx.shp [cent00_pdx_ser1km.shp]

1. Click Network Analyst on the toolbar and select "New Service Area."
2. Load cent00_pdx.shp as the Facilities.
3. In Layer Properties, set Impedance to "Length" and Default Breaks to "3280.84 ft."
4. Export the newly created "Polygons" layer as cent00_pdx_ser1km.shp.

Centroids of Centers [center.shp]

1. Add two fields to the table for the Analysis Centers shapefile named x and y in "Double" format.
2. Right-click on x to Calculate Geometry as the "X coordinate of centroid."
3. Right-click on y to Calculate Geometry as the "Y coordinate of centroid."
4. Export the table to the map document, right-click on the table to Display XY Data.
5. Export the displayed data as a shapefile called "center.shp."

The Portland CBD polygon [cbd.shp]

1. Select By Attribute the Centers from Analysis Centers.
2. Select the Center with "Name = Portland."
3. Save the selection as cbd.shp.

Centroid of the Portland CBD polygon [cbd_cent.shp]

1. Add two fields to the table for cbd.shp, and name them "x" and "y" in "Double" format.
2. Right-click on x to Calculate Geometry as the "X coordinate of centroid."
3. Right-click on y to Calculate Geometry as the "Y coordinate of centroid."
4. Export the table to the map document, righedt-click on the table to Display XY Data.
5. Export the displayed data as cbd_cent.shp.

All URA that existed in 2000 [ura00.shp]

1. Select By Attribute with URAs from the RLIS shapefile URA.
2. Select the URAs if they existed before 2000; the URAs dropped were:
   (a) Gateway Regional Center (created in 2001)
   (b) Interstate Corridor (created in August 2000)
   (c) Willamette Industrial (created in 2004)
3. Save the selection as ura00.shp.

All URA that existed at least two years before 2000 [ura00lag.shp]

1. Select By Attribute with URAs from ura00.shp
2. Select the URAs if they existed at least two years before 2000; the URAs dropped were:
   (a) Lents Town Center (created in 1998)
   (b) North Macadam (created in 1999)
   (c) River District (created in 1998)
3. Save the selection as ura00lag.shp

A.2 Geocoding Addresses

The procedure below describes the process for the 2000 dataset, which is repeated for 2010.

1. In Stata, save all stores records in 2000 to a new dataset, 2000_all.csv.
2. In ArcMap, add 2000_all.csv to tct00_pdx_00.mxd.
3. Right-click on the table to Geocode Addresses, and specify the address locator to be the RLIS Address Locator.
4. Save the geocoded addresses as 2000_all_geocode.shp.

5. In the Interactive Window, rematch both "Tied Addresses" and "Unmatched Addresses" by looking up both the store name and address on Google Maps.
   (a) For "Tied Addresses," visually pinpoint the exact store lot in Google Maps, and rematch the address that zooms to the closest store lot in ArcMap
      i. Look up the address in Google Maps
      ii. Visually locate the exact store lot in both Google Maps
      iii. The Interactive Window in ArcMap presents a few store lots with the same tied address, but occupy different parts of a block.
      iv. In the Interactive Window, visually pick the store lot that most closely matches the store location in Google Maps, and select the address of that store lot to be geocoded.
   (b) For "Unmatched Addresses," fix minor errors in the address and rematch, or drop the addresses with no possible online verification. Please refer to Table A.2 for examples of the minor errors in addresses that were fixed for a rematch.
6. From 2000_all_geocode.shp, I generate all shapefiles of stores specified in Table 3.10 using the NAICS code (naicsprim) and the employee size code (emp_cd).
   (a) For All Convenience Stores, Select By Attribute the store records with "naicsprim = 44512001" and save as 00_con.shp.
   (b) For Small Groceries, Select By Attribute the store records with "em_cd = 1" and save as 00_empcd1.shp.
   (c) The rest of the store shapefiles were generated in the same fashion.

Table A.2 provides information on every "Unmatched Address" in both the 2000 and 2010 store datasets. If a store has "R" as its Status, then its address was successfully rematched. On the other hand, if a store has "U" as its status, the store could not be geocoded and was dropped.

A.3 Construction of Dependent Variables

The procedures for constructing the four variants of measures are laid out below. These procedures are described using small groceries, 2000 data and 2000 boundary. The same procedures were applied to the other store shapefiles, 2010 data, as well as data generated using different boundary definitions. All data tables of measures were exported in .txt format from their respective shapefiles.

Table A.2: Information on "Unmatched" Addresses

Year	Name	Original Address Information	Updated Address Information	Status
2000	PLAID PANTRY MARKETS	10177 SE SUNNYSIDE RD	10178 SE SUNNYSIDE RD	R
2000	JACKPOT FOOD MART	PORTLAND 97266	HAPPY VALLEY 97086	R
2000	SELF SERVICE MARKET	PORTLAND 97266	HAPPY VALLEY 97086	R
2000	MURPHY'S MARKET	148 SE POWELL BLVD	14902 SE POWELL BLVD	R
2000	PLAID PANTRY MARKETS	161 NE SANDY BLVD	16152 NE SANDY BLVD	R
2000	PLAID PANTRY MARKETS	210 SE HIGHWAY 212	10218 SE HIGHWAY 212	R
2000	STOP N GO	19889 SE HIGHWAY 212 BORING 97009	19889 SE SUNNYSIDE RD DAMASCUS 97089	R
2000	MARIA'S MARKET	24180 SE BORGES RD GRESHAM	n/a	U
2000	GET & GO	346 SE HIGHWAY 211 CLACKAMAS 97015	346 SE 5TH AVE ESTACADA 97023	R
2000	BILL'S KWIK MART	3510 NE UNION AVE PORTLAND	n/a	U
2000	ORIENT ROAD FOOD MARKET	1220 SE ORIENT DR	1350 SE ORIENT DR	R
2000	CARVER STORE	CLACKAMAS 97015	DAMASCUS 97089	R
2000	MINIT MART FOOD STORE	TROUTDALE 97060	FAIRVIEW 97024	R
2000	SAFEWAY	250 TIGARD PLZ TIGARD	n/a	U
2000	LIBERAL COUNTRY STORE	28565 S HIGHWAY 213 MULINO	n/a	U
2000	MC PIER GROCERY & DELI	600 NW FRONT AVE	600 NW NAITO PKWY	R
2000	SOUTH END GROCERY	1033 S SOUTH END RD	1033 SOUTH END RD	R
2000	CUB FOODS	PORTLAND 97266	HAPPY VALLEY 97086	R
2000	BLACKMAN'S 4 WAY GROCERY	12700 OREGON 211	12700 S HWY 211	R
2000	SAFEWAY	LAKE OSWEGO	TUALATIN	R
2000	BIG BEAR'S COUNTRY MARKET	31815 E HISTORIC COLUMBIA RVR	31815 E HISTORIC COLUMBIA RIVER HWY	R
2000	SAFEWAY	BORING 97009	DAMASCUS 97089	R
2000	HOODLAND THRIFTWAY	68280 U.S. 26	68280 E HIGHWAY 26	R
2000	JIM's THRIFTWAY	660 S MAIN ST	12660 NW MAIN ST	R
2000	MT HOOD FOODS	73265 U.S. 26	73265 E HIGHWAY 26	R
2010	LIBERAL COUNTRY STORE	28565 S HIGHWAY 213 MULINO	n/a	U
2010	VILLAGE MARKET	46760 NW SUNSET HWY Manning	n/a	U
2010	SAFEWAY	LAKE OSWEGO	TUALATIN	R
2010	HAGGEN FOOD & PHARMACY	BEAVERTON 97006	HILLSBORO 97124	R
2010	GASTON MARKET	2222 FRONT ST	222 FRONT ST	R
2010	TIENDA CHIHUAHUA	23345 NE HALSEY ST # 1	1560 SW CLARA ST	R
2010	BIG BEAR'S COUNTRY MARKET	31815 E HISTORIC COLUMBIA RVR	31815 E HISTORIC COLUMBIA RIVER HWY	R
2010	JIM's THRIFTWAY	660 S MAIN ST	12660 NW MAIN ST	R
2010	HOODLAND THRIFTWAY	68280 U.S. 26	68280 E HIGHWAY 26	R
2010	MT HOOD FOODS	73265 U.S. 26	73265 E HIGHWAY 26	R

M1D1 [empcd1_m1d1.shp]

A Spatial Join of the specified category of stores to the Euclidean buffers of centroids was used to construct m1d1. This Spatial Join tool joins the stores (as points) to the buffers (as polygons) if the stores are located within the buffers. And it automatically creates a field that contains the number of stores that are joined to each buffer, which is exactly m1d1.

1. Spatial Join the stores to centroid buffers
   (a) Right-click on cent00_pdx_buf1km.shp, the shapefile of Euclidean 1km-buffers of centroids, to Join.
   (b) Choose "Join data from another layer based on spatial location."
   (c) Choose 00_empcd1.shp as the store shapefile to be joined to the buffer shapefile.
2. Export the joined shapefile named "empcd1_m1d1.shp" to the folder "tct00_pdx_00."

M1D2 [empcd1_m1d2.shp]

Another Spatial Join was performed for m1d2, now joining the specified stores to the street network buffers of centroids.

1. Spatial Join the stores to centroid buffers.
   (a) Right-click on cent00_pdx_ser1km.shp, the shapefile of street network 1km-buffers of centroids, to Join.
   (b) Choose "Join data from another layer based on spatial location."
   (c) Choose 00_empcd1.shp as the store shapefile to be joined to the buffer shapefile.
2. Export the joined shapefile named "empcd1_m1d2.shp" to the folder "tct00_pdx_00."

M2D1 [empcd1_m2d1]

The Near Table tool takes a list of input centroids and input stores, and outputs a table that contains the Euclidean distance to the nearest store. This Euclidean distance is m2d1. Note that the Near Table tool produces an Info Table, not a shapefile.

1. Generate an Info table with Near Table.
   (a) In Near Table, set Input Features to cent00_pdx.shp and Near Features to 00_empcd1.shp.
   (b) Save the Info table as empcd1_m2d1 to the folder "tct00_pdx_00."
2. Merge with the cent00_pdx table to obtain the field "fips," which is essential as a merging variable when compiling master datasets.
   (a) Right-click on the Info table to Join.
   (b) Choose "Join attributes from a table."
   (c) Choose field to join to be "IN_FID," choose cent00_pdx as the table to join, and "FID" as the join variable.

M2D2 [empcd1_m2d2.shp]

I used the Network Analyst Origin Destination (OD) cost matrix tool to calculate the distance between the centroid and the nearest store on the street network. The OD matrix outputs several layers, one of which contains the street network distance between each input centroid and each input store, from which I could extract the smallest distance for each input centroid to get m2d2.

1. Generate the OD matrix using Network Analyst.
   (a) Click Network Analyst on toolbar and select "New OD Cost Matrix."
   (b) Right-click "Orgins" to load cent00_pdx.shp as the origins, set Field Name to "fips."
   (c) Right-click "Destinations" to load 00_empcd1.shp as the destinations, set

Field Name to "name."

(d) Set Impedance to "Length (feet)", and click the Solve button.

2. Select the smallest distance to store for each centroid.
   (a) Select By Attribute from the "Lines" layer in the OD matrix.
   (b) Select the records if "DestinationRank = 1."
   (c) Export selection as empcd1_m2d2.shp to the folder "tct00_pdx_00."

A.4 Construction of Independent Variables

URA [ura00_tct00.shp]

The indicator variable is 1 for a census tract if the centroid of that census tract is within an URA that existed in 2000. I first selected the census tracts whose centroids are within an URA using Select By Location. Then, I merged the list of selected census tracts to the list of all census tracts, and assigned 1 for the indicator variable to the merged census tracts.

1. Select By Location the centroids of census tracts from cent00_pdx.shp.
2. Select the centroids that are within ura00.shp.
3. Save the selection as ura00_tct00.shp and export the shapefile as ura00_tct00.txt.
4. In Stata, merge ura00_tct00.txt to tct00_pdx_00.dta by fips.
   (a) If the census tract, identified by its fips, is succesfully merged, then the variable "ura" is 1.
   (b) If the census tract is not in ura00_tct00.txt, then the variable "ura" is 0.

URA Lag [ura00lag_tct00.shp]

The indicator variable is 1 for a census tract if the centroid of that census tract is within an URA that existed at least two years before 2000. I first selected the census tracts whose centroids are within a lag URA using Select By Location. Then, I merged the list of selected census tracts to the list of all census tracts, and assigned 1 for the indicator variable to the merged census tracts.

1. Select By Location the centroids of census tracts from cent00_pdx.shp.
2. Select the centroids that are within ura00lag.shp.
3. Save the selection as ura00lag_tct00.shp and export the shapefile as ura00lag_tct00.txt.
4. In Stata, merge ura00lag_tct00.txt to tct00_pdx_00.dta by fips.
   (a) If the census tract, identified by its fips, is succesfully merged, then the variable "uralag" is 1.

(b) If the census tract is not in ura00lag_tct00.txt, then the variable "uralag" is 0.

Transit Access [transcov00.shp]

The indicator variable is 1 for a census tract if the centroid of that census tract is within a polygon in the RLIS Transit Access shapefile. I first selected the census tracts using Select By Location. Then, I merged the list of selected census tracts to the list of all census tracts, and assigned 1 for the indicator variable to the merged census tracts.

1. Select By Location the centroids of census tracts from cent00_pdx.shp.
2. Select the centroids that are within the RLIS Transit Access shapefile.
3. Save the selection as transcov00.shp and export the shapefile as transcov00.txt.
4. In Stata, merge transcov00.txt to tct00_pdx_00.dta by fips.
   (a) If the census tract, is succesfully merged, then the variable "transcov" is 1.
   (b) If the census tract is not in transcov00.txt, then the variable "transcov" is 0.

Distance to Central Business District (CBD) [cbd_pdx_d2_00.shp]

The street network distance from the tract centroid to the CBD is measured using the OD matrix.

1. Generate the OD matrix using Network Analyst.
   (a) Click Network Analyst on toolbar and select "New OD Cost Matrix."
   (b) Right-click "Orgins" to load cent00_pdx.shp as the origins, set Field Name to "fips."
   (c) Right-click "Destinations" to load cbd_cent.shp as the destinations, set Field Name to "name."
   (d) Set Impedance to "Length (feet)", and click the Solve button.
2. Select the smallest distance to store for each centroid.
   (a) Select By Attribute from the "Lines" layer in the OD matrix.
   (b) Select the records if "DestinationRank = 1."
   (c) Export selection as cbd_pdx_d2_00.shp.

Distance to the Nearest Center [center_pdx_d2_00.shp]

The street network distance from the tract centroid to the nearest Center is measured using the OD matrix.

1. Generate the OD matrix using Network Analyst.
   (a) Click Network Analyst on toolbar and select "New OD Cost Matrix."
   (b) Right-click "Orgins" to load cent00_pdx.shp as the origins, set Field Name to "fips."
   (c) Right-click "Destinations" to load center.shp as the destinations, set Field Name to "name."
   (d) Set Impedance to "Length (feet)", and click the Solve button.
2. Select the smallest distance to store for each centroid.
   (a) Select By Attribute from the "Lines" layer in the OD matrix.
   (b) Select the records if "DestinationRank = 1."
   (c) Export selection as center_pdx_d2_00.shp.

Appendix B

Compilation of Master Datasets

As described in Appendix A, all manually created shapefiles and the data of constructed spatial variables for the same year, geographic scope, and census tract boundary definition are kept in the same map document. The name of the map document matches that of the Stata dataset, named in the format tctBD_GEO_YR. BD denotes the census tract boundary adopted: either the 2000 boundary ("00"), the 2010 boundary ("10"), or the panel dataset boundary ("cw"). GEO denotes the geographic scope, which is either the City of Portland ("pdx"), or the UGB ("ugb"). YR denotes the year of the data, which is either 2000 ("00"), or 2010 ("10"), or panel ("0010").

In this Appendix, I describe the process of compilation for all master datasets. The two master datasets where each year's data are aggregated by the same year's boundary definition (the general master datasets) are compiled first, followed by the compilation of the panel master dataset.

B.1 General Master Dataset

In the two general master datasets, the BD and YR are identical. I describe the process of compiling tct00_pdx_00.dta in Stata, which is repeated for tct10_pdx_10.dta. The Stata dofile used in each step of the compilation is documented as well. All dofiles are attached at the end of the appendices.

1. Construct the census independent variables (indep.do)
2. In ArcMap, export the data tables for all constructed variables in .txt format (Appendix A)
3. Merge the data in .txt for the dependent variables and save the dataset as tct00_pdx_00.dta (datasets.do)
4. Construct the spatial independent variables using data in .txt from Arcmap, as

well as some census data (indep.do)

5. Merge the census independent variables and the spatial independent variables to tct00_pdx_00.dta (indep.do)

B.2 Panel Master Dataset

Data from US2010

US2010 is a research program that compiles demographic data from the U.S. Census Bureau across different years and surveys, and conducts research on American demographics. It is a credible program "supported by the Russell Sage Foundation and Brown University" (*US2010: History* 2014). One of its projects, the Longitudinal Tract Data Base (LTDB), provides the necessary tools to transform census tract data from prior years to the 2010 boundary.

For the master panel dataset, I am transforming census tract data from 2000 to the 2010 boundary, which requires one crosswalk Stata dataset and one Stata dofile for the actual transformation (*US2010: LTDB4* 2014). The crosswalk dataset named "crosswalk_2000_2010" contains rows of all tracts in 2000 and their corresponding tracts in 2010, and vice versa. Each row of a 2000–2010 tract pair also contains a weight variable produced by US2010 that combines area and population weights. For example, if two tracts in 2000 are combined in 2010, then the weight for both 2000 tracts would equal 1. This means that the data values for each of the two 2000 tracts entirely contribute to the data values for the tract in 2010. In another basic case, if one tract in 2000 is split evenly to two tracts in 2010 by both area and population, then the 2000 tract would appear in two records: one for each tract in 2010. And the weight for both records would equal 0.5. This means that the data values for each of the tracts in 2010 have precisely half of the data values for the 2000 tract.

Preliminary Setup for Variable Transformation

The US2010 dofile called "interpolate_to_2010" merges the crosswalk file and an input 2000 dataset, and outputs the same information (dis)aggregated to the 2010 tracts. The input dataset is "tct00_pdx_00_cw," which includes all non-spatial variables in "tct00_pdx_00" plus the median weight for each non-spatial median variable. The merge variable is the fips code of the 2000 tracts. The merged dataset contains all of the census tracts in 2010 that correspond to all or a part of the input 2000 census tracts. The output dataset saved as "tctcw_pdx_00" contains the transformed variables, as well as the median weights.

I specify the non-spatial variables to be transformed, which US2010 views as either counts or medians. For counts, US2010 only relies on its weight variable to make the calculations. For other variables such as averages, medians, and rates, US2010 views them as medians, and requires that the respective median weights to be specified. For example, if the variable is the percentage of 1–person households, then its median weight would be the total number of households. If the variable is median household income, then its median weight is also the total number of households. Most median weights are found in the data tables used to construct the medians, and they can be directly merged to the input dataset.

For some variables whose data tables do not provide the median weights, I use median weights from other data tables as proxies. For average household size (avhhsize), I use the total number of households in the data table, Household Size. For median household income, I also use the total number of households in Household Size. For median house value, I use the total number of owner-occupied housing units from the data table, Vehicle Access, because median house values are only sampled from owner-occupied housing units.

US2010 Methodology for Variable Transformation

In US2010's formula for variable transformation, *count* is a count variable, *median* is a median variable, *median_weight* is the weight applied to the corresponding median variable, and *weight* is US2010's own area and population interpolation weight.

The formula for count variables:

$$\sum count * weight \tag{B.1}$$

The formula for median variables:

$$\frac{\sum median * median\_weight * weight}{\sum median\_weight * weight} \tag{B.2}$$

Example of Variable Transformation: Two Tracts in 2000 Merged into One Tract in 2010

Two tracts in 2000, 23.01 and 23.02, were merged into one tract in 2010, 23.03. I demonstrate how the average household size (avhhsize) variable is transformed. In this example, a stands for average household size, awt is the total number of households, and the subscripts denote the tract numbers.

$a_{23.03} = \frac{a_{23.01}*awt_{23.01}*weight_{23.01to23.03}+a_{23.02}*awt_{23.02}*weight_{23.02to23.03}}{awt_{23.01}*weight_{23.01to23.03}+awt_{23.02}*weight_{23.02to23.03}} = \frac{1.47*126*1+1.66*702*1}{126*1+702*1}$
≈ 1.631

Example of Variable Transformation: One Tract in 2000 Split into Two Tracts in 2010

The tract numbered 64.01 in 2000 was split into two tracts in 2010, 64.03 and 64.04. I demonstrate how the total population (popnum) variable is transformed. In this example, p is the total population, there is no median weight, and the subscripts denote the tract numbers. A caveat is that tract 64.04 in 2010 also comprises a very small amount of area from two other surrounding tracts, so this calculation is not entirely accurate. The actual total population from US2010 is approximately 3378.

$p_{64.03} = p_{64.01} * weight_{64.01to64.03} = 7140 * 0.52549022 \approx 3752$
$p_{64.04} \approx p_{64.01} * weight_{64.01to64.04} = 7140 * 0.47450981 \approx 3388$

Procedure in Compiling the Panel Master Dataset

1. Merge in median weights for all non-spatial variables to be transformed to tct00_pdx_00_cw (indep.do)
2. Set the inputs to the US2010 dofile (interpolate_to_2010.do)
3. Append the output dataset tctcw_pdx_00 to tct10_pdx_10 to generate the panel dataset tctcw_pdx_0010 (datasets.do)
4. In ArcMap, re-construct the spatial variables from 2000 using 2010 boundary (Appendix A)
5. Merge the re-constructed spatial variables into tctcw_pdx_0010

Appendix C

Complete List of Variables

Table C.1: Description for a complete list of independent variables

Variable Name	**Variable Description**	**Data Table**
area	Area of census tract (sqkm.)	
age_18under	Percentage of total population under 18 years	
age_18to24	Percentage of total population 18 to 24 years	
age_25to44	Percentage of total population 25 to 44 years	
age_45to64	Percentage of total population 45 to 64 years	
age_65over	Percentage of total population 65 years and over	
highsch_atleast	Percentage of adult population (25 and up) with at least high school diploma or equivalent	
highsch_below	Percentage of adult population (25 and up) without high school diploma or equivalent	
hhsize_1	Percentage of 1–person households	
hhsize_2	Percentage of 2–person households	
hhsize_3	Percentage of 3–person households	
hhsize_4	Percentage of 4–person households	
hhsize_5	Percentage of 5–person households	
hhsize_6	Percentage of 6–person households	
hhsize_7up	Percentage of 7–or–more–person households	
hisp	Percentage of Hispanic or Latino in total population	
medhhinc	Median household income in the past 12 months (in 2010 inflation–adjusted dollar)	
popnum	Total Population	
white_ormore	Percentage of total population identified as White alone or in combination	
black_ormore	Percentage of total population identified as Black or African American alone or in combination	
asian_ormore	Percentage of total population identified as Asian alone or in combination	
veh_0	Percentage of households with no vehicle	
veh_1	Percentage of households with at least one vehicle	
pov_below	Percentage of households with income in the past 12 months below poverty level	
pov_above	Percentage of households with income in the past 12 months at or above poverty	
medhhinc_1000	Median household income in thousands	
popnum_1000	Total population in thousands	
dens	Population density (Total pop/sqkm.)	
dens_1000	Population density (Total pop in thousands/sqkm.)	
avhhsize	Average household size	
avhhsize_own	Average household size in owner–occupied units	
avhhsize_rent	Average household size in renter–occupied units	
hhsize_3less	Percentage of households with less than three people	
hhsize_3to	Percentage of households with three people or more	
medhv	Median value for all owner–occupied housing units	
medhv_10000	Median value for all owner–occupied housing units in tens of thousands of dollars	
lmedhv	Logarithm of medhv	
ura	Centroid is within an Urban Renewal Area	
sing	Percentage of single–parent households	
bike	Percentage of workers 16 and over commuting to work via bicycles	
unemp	Percentage of population 16 and over unemployed	
foreign	Percentage of foreign–born population	
bachelor	Percentage of the adult population (25 and up) with at least a bachelor's degree	
bachelorno	Percentage of the adult population (25 and up) without a bachelor's degree	
dummy10	Year is 2010 versus 2000	
uralag	Centroid is within an Urban Renewal Area at least two years before current year	

Appendix D

Complete Regression Results

D.0.1 Access to Small Groceries between 2000 and 2010

Table D.1: Access to Small Groceries between 2000 and 2010

Year/Type	2000/OLS	2010/OLS	Both/Pooled	Both/Fixed
DENS	9.368***	7.666***	8.346***	0.094
	(2.750)	(2.280)	(2.246)	(2.766)
VEH	3.813**	4.654***	4.265***	1.711
	(1.578)	(1.386)	(1.302)	(1.325)
c.DENS#c.VEH	-16.277***	-14.484***	-15.342***	-2.622
	(4.029)	(3.382)	(3.392)	(3.844)
BIKE	3.713	1.555	2.095*	2.651***
	(3.230)	(1.513)	(1.263)	(0.996)
ASIAN	1.396	1.628	1.750	0.280
	(1.969)	(2.597)	(1.798)	(1.376)
HISP	1.431	-0.993	0.291	0.969
	(2.155)	(1.604)	(1.148)	(0.919)
BLACK	1.145	6.516**	3.212**	-0.968
	(1.324)	(2.788)	(1.396)	(1.500)
MEDHHINC	0.227***	0.108	0.148*	-0.027
	(0.085)	(0.081)	(0.076)	(0.056)
c.BLACK#c.MEDHHINC	-0.286	-1.393***	-0.713**	0.343
	(0.307)	(0.513)	(0.314)	(0.308)
HIGHSCHNO	0.308	-0.388	-0.343	-0.401
	(1.576)	(1.470)	(1.103)	(0.698)
_cons	-2.726**	-2.433**	-2.452***	0.107
	(1.092)	(0.935)	(0.867)	(1.108)
R^2	0.56	0.45	0.49	0.10
N	141	141	282	282

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Appendix E

Dofiles

Please see attached CD for dofiles.

References

Apparicio, Philippe, Marie-Soleil Cloutier, and Richard Shearmur. 2007. “The case of Montreal’s missing food deserts: Evaluation of accessibility to food supermarkets.” *International Journal of Health Geographics* 6, no. 1 (February): 4.

ArcGIS Help 10.1 - What is geocoding? 2014, March.

Baker, Elizabeth A, Mario Schootman, Ellen Barnidge, and Cheryl Kelly. 2006. “The role of race and poverty in access to foods that enable individuals to adhere to dietary guidelines.” *Preventing chronic disease* 3, no. 3 (July): A76.

Bitler, Marianne, and Steven J. Haider. 2011. “An Economic View of Food Deserts in the United States.” *Journal of Policy Analysis and Management* 30 (1): 153–176.

Black, Jennifer L., Richard M. Carpiano, Stuart Fleming, and Nathanael Lauster. 2011. “Exploring the distribution of food stores in British Columbia: Associations with neighbourhood socio-demographic factors and urban form.” *Health & Place* 17, no. 4 (July): 961–970.

Black, Jennifer L, and James Macinko. 2008. “Neighborhoods and obesity.” *Nutrition Reviews* 66 (1): 2–20.

What are NAICS and SIC industry codes and how do I use them? 2013, December.

Church, Richard L., and Alan T. Murray. 2008. “Introduction.” In *Business Site Selection, Location Analysis and GIS,* 1–17. John Wiley & Sons, Inc. ISBN: 9780470432761.

City of Portland Bureau of Planning and Sustainability. 2014a. *City of Portland Growth Scenarios Report Released,* April.

———. 2014b. *The Portland Plan Backgroun Reports,* April.

Coalition for a Livable Future. 2007. *The regional equity atlas: metropolitan Portland’s geography of opportunity.* Portland, Oregon: Coalition for a Livable Future.

Cohen, Steve. 2013. *Reed student seeking advice,* November.

Commission, Portland Development. 2014a. *About the Portland Development Commission,* April.

———. 2014b. *Urban Renewal,* March.

Dai, Dajun, and Fahui Wang. 2011. "Geographic disparities in accessibility to food stores in southwest Mississippi." *Environment and Planning B-Planning & Design* 38, no. 4 (July): 659–677.

Donald, Betsy. 2013. "Food retail and access after the crash: rethinking the food desert problem." *Journal of Economic Geography* 13, no. 2 (March): 231–237.

Eckert, Jeanette, and Sujata Shetty. 2011. "Food systems, planning and quantifying access: Using GIS to plan for food retail." *Applied Geography* 31, no. 4 (October): 1216–1223.

Franco, Manuel, Ana V Diez Roux, Thomas A Glass, Benjamín Caballero, and Frederick L Brancati. 2008. "Neighborhood characteristics and availability of healthy foods in Baltimore." *American journal of preventive medicine* 35, no. 6 (December): 561–567.

Greenstein, Robert. 1987. *Prisoners Of The Economy,* October.

Hall, Bruce F. 1983. "Neighborhood Differences in Retail Food Stores: Income Versus Race and Age of Population." *Economic Geography* 59, no. 3 (July): 282–295.

Krukowski, Rebecca A., Delia Smith West, Jean Harvey-Berino, and T. Elaine Prewitt. 2010. "Neighborhood Impact on Healthy Food Availability and Pricing in Food Stores." *Journal of Community Health* 35, no. 3 (June): 315–320.

Laraia, Barbara A., Anna Maria Siega-Riz, Jay S. Kaufman, and Sonya J. Jones. 2004. "Proximity of supermarkets is positively associated with diet quality index for pregnancy." *Preventive Medicine* 39 (5): 869–875.

Larsen, Kristian, and Jason Gilliland. 2008. "Mapping the evolution of 'food deserts' in a Canadian city: Supermarket accessibility in London, Ontario, 1961–2005." *International Journal of Health Geographics* 7, no. 1 (April): 16.

Leete, Laura, Neil Bania, and Andrea Sparks-Ibanga. 2012. "Congruence and Coverage: Alternative Approaches to Identifying Urban Food Deserts and Food Hinterlands." *Journal of Planning Education and Research* 32, no. 2 (June): 204–218.

MacDonald, James M., and Paul E. Nelson. 1991. "Do the poor still pay more? Food price variations in large metropolitan areas." *Journal of Urban Economics* 30, no. 3 (November): 344–359.

McClintock, Nathan. 2011. "From Industrial Garden to Food Desert Demarcated Devaluation in the Flatlands of Oakland, California." Edited by A. H. Alkon and J. Agyeman. *Cultivating Food Justice: Race, Class, and Sustainability:*89–120.

Metro. 2014. *RLIS Discovery: Analysis centers,* March.

Metro: Urban growth boundary. 2014, March.

Moore, Latetia V., and Ana V. Diez Roux. 2006. "Associations of Neighborhood Characteristics With the Location and Type of Food Stores." *American Journal of Public Health* 96, no. 2 (February): 325–331.

Morland, Kimberly, Steve Wing, Ana Diez Roux, and Charles Poole. 2002. "Neighborhood characteristics associated with the location of food stores and food service places." *American journal of preventive medicine* 22, no. 1 (January): 23–29.

Addressing Portland's Food Deserts. 2013, September.

Netusil, Noelwah R. 2013. "Urban environmental amenities and property values: Does ownership matter?" *Land Use Policy* 31 (March): 371–377.

O'Sullivan, A. 2008. *Urban Economics.* McGraw-Hill Series in Urban Economics. McGraw-Hill Education.

Paez, Antonio, Ruben Gertes Mercado, Steven Farber, Catherine Morency, and Matthew Roorda. 2010. "Relative Accessibility Deprivation Indicators for Urban Settings: Definitions and Application to Food Deserts in Montreal." *Urban Studies* 47, no. 7 (June): 1415–1438.

Park, R.E., E.W. Burgess, and R.D. McKenzie. 1984. *The City.* Heritage of Sociology Series. University of Chicago Press.

Pearce, Jamie, Peter Day, and Karen Witten. 2008. "Neighbourhood Provision of Food and Alcohol Retailing and Social Deprivation in Urban New Zealand." *Urban Policy and Research* 26 (2): 213–227.

Portland Bureau of Planning, City of, and Sustainability. 2014. *Comp Plan Map App,* March.

Portland Plan Background Reports. 2014, March.

Powell, Lisa M., Sandy Slater, Donka Mirtcheva, Yanjun Bao, and Frank J. Chaloupka. 2007. "Food store availability and neighborhood characteristics in the United States." *Preventive Medicine* 44, no. 3 (March): 189–195.

Raja, Samina, Changxing Ma, and Pavan Yadav. 2008. "Beyond Food Deserts Measuring and Mapping Racial Disparities in Neighborhood Food Environments." *Journal of Planning Education and Research* 27, no. 4 (June): 469–482.

Retail Store Eligibility USDA Supplemental Nutrition Assistance Program — Food and Nutrition Service. 2014, March.

Sharkey, Joseph R., Scott Horel, Daikwon Han, and John C. Huber Jr. 2009. "Association between neighborhood need and spatial access to food stores and fast food restaurants in neighborhoods of Colonias." *International Journal of Health Geographics* 8, no. 1 (December): 1–17.

Shaw, H.D.M.K., Clifford. 1969. *Juvenile Delinquency and Urban Areas.*

Small, M. L., and M. McDermott. 2006. "The presence of organizational resources in poor urban neighborhoods: An analysis of average and contextual effects." *Social Forces* 84, no. 3 (March): 1697–1724.

Smoyer-Tomic, Karen E., John C. Spence, and Carl Amrhein. 2006. "Food Deserts in the Prairies? Supermarket Accessibility and Neighborhood Need in Edmonton, Canada." *Professional Geographer* 58, no. 3 (August): 307–326.

Smoyer-Tomic, Karen E., John C. Spence, Kim D. Raine, Carl Amrhein, Nairne Cameron, Vladimir Yasenovskiy, Nicoleta Cutumisu, Eric Hemphill, and Julia Healy. 2008. "The association between neighborhood socioeconomic status and exposure to supermarkets and fast food outlets." *Health & Place* 14, no. 4 (December): 740–754.

SNAP Retailer Locator — Food and Nutrition Service. 2014, March.

Sparks, Andrea. 2008. *Measuring Food Deserts: A Comparison of Models Measuring the Spatial Accessibility of Supermarkets in Portland, Oregon.*

Sparks, Andrea L., Neil Bania, and Laura Leete. 2011. “Comparative Approaches to Measuring Food Access in Urban Areas The Case of Portland, Oregon.” *Urban Studies* 48, no. 8 (June): 1715–1737.

Svastisalee, Chalida M., Helene Nordah, Charlotte Glumer, Bjorn E. Holstein, Lisa M. Powell, and Pernille Due. 2011. “Supermarket and fast-food outlet exposure in Copenhagen: associations with socio-economic and demographic characteristics.” *Public Health Nutrition* 14, no. 9 (September): 1618–1626.

United States, Department of Agriculture, and Economic Research Service. 2009. *Access to affordable and nutritious food measuring and understanding food deserts and their consequences: report to Congress.* [Washington, D.C.]: U.S. Dept. of Agriculture, Economic Research Service.

U.S. Census Bureau. 2014a. *2010 Geographic Terms and Concepts - Block Groups,* April.

———. 2014b. *2010 Geographic Terms and Concepts - Census Tract,* April.

———. 2014c. *American FactFinder,* March.

US2010: History. 2014, March.

US2010: LTDB4. 2014, March.

Wilson, W.J. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.* Sociology, urban studies, black studies. University Press.

Wrigley, Neil, Daniel Warm, and Barrie Margetts. 2003. “Deprivation, diet, and food-retail access: findings from the Leeds ’food deserts’ study.” *Environment and Planning A* 35 (1): 151 – 188.

Zenk, Shannon N., Amy J. Schulz, Barbara A. Israel, Sherman A. James, Shuming Bao, and Mark L. Wilson. 2005. “Neighborhood Racial Composition, Neighborhood Poverty, and the Spatial Accessibility of Supermarkets in Metropolitan Detroit.” *American Journal of Public Health* 95, no. 4 (April): 660–667.