

FOOD ENVIRONMENT, GROCERY STORE CHOICE, AND BODY WEIGHT:  
GEOGRAPHICAL ANALYSIS AMONG SNAP-ED PARTICIPANTS IN GEORGIA

by

YANGJIAXIN WEI

(Under the Direction of JERRY SHANNON)

ABSTRACT

Under the prevalence of obesity among the United States, there has been growing attention on food environment and health. Past research used various methods to measure individuals' spatial access to healthy food. However, less is known about how proximity matters to participants' grocery store choices and how they navigate the choices available to them. This study examines the grocery store choice of SNAP-Ed participants in Clayton, DeKalb and Fulton counties in Georgia, how their proximity to preferred grocery stores relates to individual characteristics, household food environment and neighborhood socio-economic characteristics. The results indicate that proximity didn't seem to be a significant factor influencing participants' grocery store choices, especially on bigger chains. Two measures of proximity (network distance and travel time by car) are correlated and have slight differences in modeling the relationships between individual and neighborhood socio-economic characteristics. Also, BMI does not vary by store choice.

INDEX WORDS: SNAP-Ed, Food Environment, Geographical Information Systems (GIS), Grocery Store Choice, Clustered Standard Error

FOOD ENVIRONMENT, GROCERY STORE CHOICE, AND BODY WEIGHT:  
GEOGRAPHICAL ANALYSIS AMONG SNAP-ED PARTICIPANTS IN GEORGIA

by

YANGJIAXIN WEI

BE, Wuhan University, China, 2016

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment  
of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2018

© 2018

Yangjiaxin Wei

All Rights Reserved

FOOD ENVIRONMENT, GROCERY STORE CHOICE, AND BODY WEIGHT:  
GEOGRAPHICAL ANALYSIS AMONG SNAP-ED PARTICIPANTS IN GEORGIA

by

YANGJIAXIN WEI

Major Professor: Jerry Shannon  
Committee: Lan Mu  
Jung Sun Lee

Electronic Version Approved:

Suzanne Barbour  
Dean of the Graduate School  
The University of Georgia  
August 2018

## DEDICATION

To my parents, Shengjun Wei and Hanhong Yang, who always love me and support every decision of my life.

## ACKNOWLEDGEMENTS

I would like to express my sincere thanks to everyone helped me during these two years in my masters.

First of all, I wish to thank my advisor, Dr. Jerry Shannon, for leading me in the academic world of Geography, guiding me from the beginning of the thesis process, providing me unique and invaluable ideas and comments and giving me confidence and encouragement. I would also like to express my appreciations to my committee members: Dr. Lan MU, for answering my questions on the technical and statistical side and giving me helpful suggestions; Dr. Jung Sun Lee, for providing valuable data on this study and providing me knowledge and insights from the nutrition and health side.

Special thanks to Joanna Lynn Akin who helped me through the data obtaining and cleaning process of my study. Also, thanks to everyone in the UGA SNAP-Ed team who helped me better understanding the SNAP-Ed program. I also owe grateful acknowledgment to those SNAP-Ed participants who involved in the nutrition-related survey.

Finally, I would like to thank my friends Xuan Zhang, Qinjin Fan, Dorris Scott and Doori Oh for giving me both academic assistance and bringing happiness into my life.

## TABLE OF CONTENTS

|  | Page |
|--|------|
| ACKNOWLEDGEMENTS .....                                       | v    |
| LIST OF TABLES .....   | viii |
| LIST OF FIGURES .....  | ix   |
| CHAPTER  |      |
| 1 INTRODUCTION .....   | 1    |
| 2 LITERATURE REVIEW AND RESEARCH QUESTIONS.....              | 3    |
| 2.1 Food Environment and Health .....                        | 3    |
| 2.2 Existing Food Environment Research and Limitations ..... | 5    |
| 2.3 Research Questions.....                                  | 7    |
| 3 DATA AND METHOD.....                                       | 9    |
| 3.1 Human Subjects .....                                     | 9    |
| 3.2 SNAP and SNAP-Ed Program.....                            | 9    |
| 3.3 Study Area .....   | 12   |
| 3.4 Store Survey Data .....                                  | 13   |
| 3.5 Food Retailers Data.....                                 | 15   |
| 3.6 Road Network and Census Data .....                       | 18   |
| 3.7 Network Analysis.....                                    | 18   |
| 3.8 Statistical Analysis.....                                | 20   |
| 4 RESULTS .....  | 24   |

|                                   |    |
|-----------------------------------|----|
| 5 DISCUSSION AND CONCLUSION ..... | 36 |
| REFERENCES .....                  | 39 |

LIST OF TABLES

|   | Page |
|---|------|
| Table 1: Demographic Characteristics of UGA SNAP-Ed Outreach Counties .....                             | 11   |
| Table 2: Independent Variables for Proximity Regression Model .....                                     | 21   |
| Table 3: Summary Statistics of Participants' Proximity to Each Nearest Store Chains.....                | 24   |
| Table 4: Wilcoxon Test of Participants' Proximity to each Nearest Store Chains .....                    | 28   |
| Table 5: Summary Statistics of Participants' Minimum Network Distance (mi) to Preferred<br>Stores ..... | 29   |
| Table 6: Network Distance and Travel Time Regression Models .....                                       | 33   |
| Table 7: BMI Regression Model.....  | 35   |

## LIST OF FIGURES

|  | Page |
|--|------|
| Figure 1: UGA SNAP-Ed Programming Counties .....   | 11   |
| Figure 2: Study Area .....   | 13   |
| Figure 3: Participants' home addresses .....   | 15   |
| Figure 4: SNAP-Authorized Grocery Retailers .....  | 17   |
| Figure 5: Boxplots of Participants' Minimum Network Distances to each Grocery Store Chains | 26   |
| Figure 6: Boxplots of Participants' Travel Time to each Grocery Store Chains.....          | 27   |
| Figure 7: Scatter Plot of Travel Time and Distance .....                                   | 30   |

## CHAPTER 1

### INTRODUCTION

Over the decades, there has been increasing attention on food environment and health. Lots of researchers have found that diet and environment play an important role in individuals' health outcomes (Black & Macinko, 2008; Bleich, Jones-smith, Wolfson, & Zhu, 2015; Cummins & Macintyre, 2006; Willett, 1994). As a consequence of the growing prevalence of obesity in the US, researchers have increasingly focused on disparities and access to healthy food. To measure and describe individuals' exposure to food, and take environment context into consideration, the idea of "food environment" has been brought up. The Center for Disease Control and Prevention (CDC) defined it as "Physical presence of foods that affects a person's diet, a person's proximity to food store locations, the distribution of food stores, food service and physical entity by which food may be obtained, or a connected system that access to food." (Centers for Disease Control and Prevention, 2014). Geographical information system (GIS) technology is widely used in measuring food environment. (Ball, Timperio, & Crawford, 2009; Lytle & Sokol, 2017; Smoyer-Tomic et al., 2008). Another issue with food environment research is the disparities of people's access to healthy food, areas with low income and low access are generally recognized as food deserts (Cummins & Macintyre, 2006; Gallagher, 2011), Socioeconomic status and racial composition are also commonly studied in research on food desert. Individuals' mobility, transportation mode and time are also important factors in food environment research (Charleux, 2017; Chen & Clark, 2016).

There has been much research on the food environment and food deserts, but most of them only took demographic characteristics and geographical characteristics into consideration and overlooked factors that influence individuals' own grocery store choice and what drives them. It is of high importance to examine individual's perception when looking into food environment and how it relates to public health concerns.

Many studies have specifically looked into food environment of the low-income population. Supplemental Nutrition Assistance Program Education (SNAP-Ed) is a national program supported by U.S. Department of Agriculture (USDA) serving individuals at risk of food insecurity. The University of Georgia (UGA) SNAP-Ed is one of the four implementing agencies in Georgia. Though there are many studies on the residential food environment, fewer have studied the food choices of individuals.

In this research, we focused on the relationship between food environment and grocery store choice. We assessed the local grocery store environment (focusing on main chains) of SNAP-Ed participants in Georgia, examined participants' proximity to their preferred grocery store and how it relates with their individuals' characteristics, neighborhood socio-economic characteristics, local food environment variables. We also looked into how health characteristics, specifically body mass index (BMI) relates with participants' grocery store choice.

## CHAPTER 2

### LITERATURE REVIEW AND RESEARCH QUESTIONS

#### 2.1 Food Environment and Health

People's dietary intake can have significant effects on their health outcome. Different food intake patterns can lead to opposite consequences towards one's health (Bleich et al., 2015). In both the prevalence and the prevention of many diseases including heart disease, cancer, and obesity, diet plays an important role. Evidence has shown that vegetables and fruits can protect against those diseases (Willett, 1994). Researchers acknowledged that lack of healthy food could result in bad health outcomes such as obesity, diabetes and other common chronic diseases (Black & Macinko, 2008; Cummins & Macintyre, 2006; National Research Council, 1989).

Over the last three decades, Americans had a prevalence of being overweight, and the rate keeps increasing. As a result of growing attention to the prevalence of obesity in the relationship between diet and health, there has been more research emphasis on disparities and access to healthy food. In order to address the obesity problem, it's critical to take the environment setting into consideration. Some researchers drawing on the concept of the social-ecological model have emphasized the significance of neighborhood context influence such as food promotion and the interaction between individuals sensitivity and recognition of the surrounding environment in food and health research, (Chen & Kwan, 2015; Shannon, 2014).

To fully describe people's access and exposures to food, CDC defined the concept of food environment (Centers for Disease Control and Prevention, 2014). Various methods have been used to measure food environment, and the most common methodology is using GIS

technology (Lytle & Sokol, 2017). One way to measure spatial access to grocery stores is by counting the number of stores within a certain Euclidean distance or network distance from a unit (Ball et al., 2009; Smoyer-Tomic et al., 2008). Another way to measure food access is by measuring the distance to the nearest or second nearest grocery store (Laraia, Siega-Riz, Kaufman, & Jones, 2004; Zenk et al., 2005; Zhang et al., 2016). Research on food environment revealed large inequities of people's availability to healthy food and exposure to unhealthy food, common in research on food deserts. The term "food deserts" was first introduced in the 1990s by a resident of a public sector housing scheme in Scotland, then it was further defined by Cummins and Macintyre as poor urban areas where residents cannot buy affordable healthy food (Cummins, Macintyre, & Glasgow, 2002). While food deserts exist not only in urban areas but also in rural areas, the United States Department of Agriculture (USDA) defined it as parts of the country void of fresh fruit, vegetables, and other healthful whole foods, usually found in impoverished areas, which is largely due to a lack of grocery stores, farmer's markets, and healthy food providers (Gallagher, 2011). Past research has applied different methods to identify food deserts and found out that neighborhoods with low socio-economic status (SES) have less access to fresh food. Some research pointed out that low-income neighborhoods generally have less access to supermarkets or healthy food providers (Lamichhane et al., 2013; Powell, Slater, Mirtcheva, Bao, & Chaloupka, 2007). Neighborhood access to healthy food also vary differently by race. Healthy food providers are absent in many urban neighborhoods of color (Larson, Story, & Nelson, 2009; Lewis, Sloane, Nascimento, Diamant, & Guinyard, 2005; Moore & Roux, 2006; Zenk et al., 2005). Some other researchers found associations between neighborhood deprivations and healthy food access differs by environment settings, rural area healthy food accessibility is not associated with racial composition or poverty while the pattern

of healthy food access in urban areas are more similar (Bleich et al., 2015; Bower, Jr, Rohde, & Gaskin, 2014; Ghosh-Dastidar et al., 2017; Sharkey & Horel, 2008; Smith et al., 2009). Most of the food access studies focused on supermarkets access, but some have also included small ethnic markets, which play a positive role in providing healthy food especially in immigrant communities, might be neglected by researchers (Joassart-Marcelli, Rossiter, & Bosco, 2017).

## 2.2 Existing Food Environment Research and Limitations

Food environment research has focused a lot on demographics of residential neighborhoods such as socioeconomic status and racial segregation, and used a diversity of measurements including network, distance, buffer analysis, travel time and spatial-temporal approach. However, less is known about how people navigate the choices available to them. Cummins et al. mentioned a key issue in food environment study is the correct specification of the environments to which people are exposed or the uncertain geographic context problem (Cummins, Clary, & Shareck, 2017; Kwan, 2012). Some recent research incorporated the global positioning system (GPS) data or travel survey data on tracking people's daily travel pattern to understand their exposure to food or grocery stores (Charleux, 2017; Sadler & Gilliland, 2015; Shannon, 2016). Transportation mode was also taken into consideration by many studies. People's movements among the space are complicated, as their activities are different at different time and locations with different transportation mode. Only considering spatial access from the residential place can be misleading, as people's choices of grocery shopping vary dynamically by individual mobility pattern and transportation environment (Charleux, 2017; Coveney & O'Dwyer, 2009; Widener, 2017). While proximity to grocery stores is the most obvious way to describe food access, temporal access is another important constraint. A spatial-temporal

approach is another growing way to analyze the food environment (Chen & Clark, 2016; Widener et al., 2017; Widener, Metcalf, & Bar-Yam, 2011).

Although the approach above considers mobility pattern on measuring food environment and avoid predefined geographic boundaries, it still fails to capture other factors that influence an individual's grocery store choice regardless of spatial proximity. Although mobility pattern and geographical location are important components in characterizing food environment, other factors such as accessibility to certain transportation mode, the price of the food and the quality of the food are also of great importance in shaping individual's grocery shopping choices. So, it's of high importance to examine individuals perceptions and avoid misrepresentation of the geographic context to which people actually exposed to when we do food environment research (Mazidi & Speakman, 2017). While people perceive the food environment differently from what researchers have observed objectively (Chen & Kwan, 2015), it's necessary to know individuals' choices of grocery stores and how the proximities to their store choices relate to other individual, neighborhood, and local food environment characteristics. This could provide us with some insights on evaluating the food environments more correctly.

A majority of food environment studies have characterized the food environment of the national population or the low-income neighborhoods, but less is known about the grocery store choices of vulnerable individuals. By knowing more about where participants in the SNAP-Ed program, who are economically vulnerable, regularly shop for food, go grocery shopping, we can have a better understanding on how they actually use stores in their neighborhoods and target healthy food advertisements more efficiently to those population. Evaluating SNAP-Ed participants perceived food environments and its relationship with health outcome and dietary

quality would also provide valuable information for policymakers to target vulnerable population and outreach nutrition educations more efficiently.

### 2.3 Research Questions

In this study, we will contribute to the research gaps and limitations mentioned above by answering the following questions:

1) Is proximity a factor that is associated with grocery choices of SNAP-Ed participants?

Many past studies have emphasized that proximity to grocery stores especially supermarkets affect the healthfulness of food choices (Cummins et al., 2017; Gustafson, Christian, Lewis, Moore, & Jilcott, 2013; Mazidi & Speakman, 2017). The assumption under the above conclusion is that proximity does matter on people's food choices. Most research assumes that a person would go to a store close to them, from this research question, we would like to find out if proximity is an important factor influencing individuals' food choices.

2) What's the relationship of SNAP-Ed participants' proximities to their preferred grocery stores with individual, neighborhood and local food environment characteristics? And how these relationships vary by different measurements of proximity (network distance and travel time)? Previous research has explored the roles of socio-economic status on healthy food access and food environment, little of them used individual proximity data and characteristics. Also, researchers used different proximity measures such as network distance, travel time by car or travel time by public transit, but not much research compared these measures of proximity: are they different and how they are different?

3) Does BMI of SNAP-Ed participants associate with participants' store choice? Other articles studied the relationship between people's health outcome with food environment. Instead

of examining this relationship from food environments that researchers objectively observed. I'd like to explore it from the perspective of people's own grocery store choice.

## CHAPTER 3

### DATA AND METHOD

#### 3.1 Human Subjects

To ensure and protect the human subjects in this research, UGA SNAP-Ed got the approval from the Institutional Review Board (IRB). The latest UGA SNAP-Ed IRB ID is MOD00006052 (STUDY00000941).

#### 3.2 SNAP and SNAP-Ed Program

The Supplemental Nutrition Assistance Program (SNAP) offers nutrition and economic assistance to low-income individuals and households and is the largest program in the domestic hunger safety net (United States Department of Agriculture, 2017). SNAP-Ed is funded by the USDA's Food and Nutrition Service and offered nationwide. The SNAP-Ed program aims to improve the nutrition and healthy lifestyle knowledge of individuals who are SNAP participants and low-income individuals eligible to receive SNAP benefits or other federal assistance and will teach them how to make their benefits stretch further. Agencies in each state contract with USDA to provide classes in nutritional education and sponsor initiatives to encourage healthy food choices.

UGA SNAP-Ed program aims to help the low-income population in Georgia establish healthy eating habits and a physically active lifestyle through evidence-based nutrition education and obesity interventions based on the collective capacity of UGA and the augmented infrastructure and network of the UGA Cooperative Extension.

The UGA SNAP -Ed program has four types of services currently: 1) Food Talk, a direct, face-to-face nutrition education via paraprofessional peer educator model in 8 urban and rural counties (i.e., Clarke, Fulton, DeKalb, Bartow, Clayton, Gilmer, Coffee, and Lowndes County), 2) Food eTalk, an innovative smartphone-friendly online eLearning nutrition education provided to all 159 counties in Georgia, 3) Social Marketing Food Talk intervention through newsletters, social media, and segmented marketing in all 159 counties in Georgia, 4) Healthy Child Care Georgia (HCCG), formative study combining both Policy, Systems, and Environmental change approach and direct nutrition education in two selected urban counties. Figure 1 is a map of UGA SNAP-Ed current eight programming counties. Table 1 described the population, racial composition and income characteristics of these eight Food Talk programming counties based on 2012 – 2016 American Community Survey 5-Year Estimates.

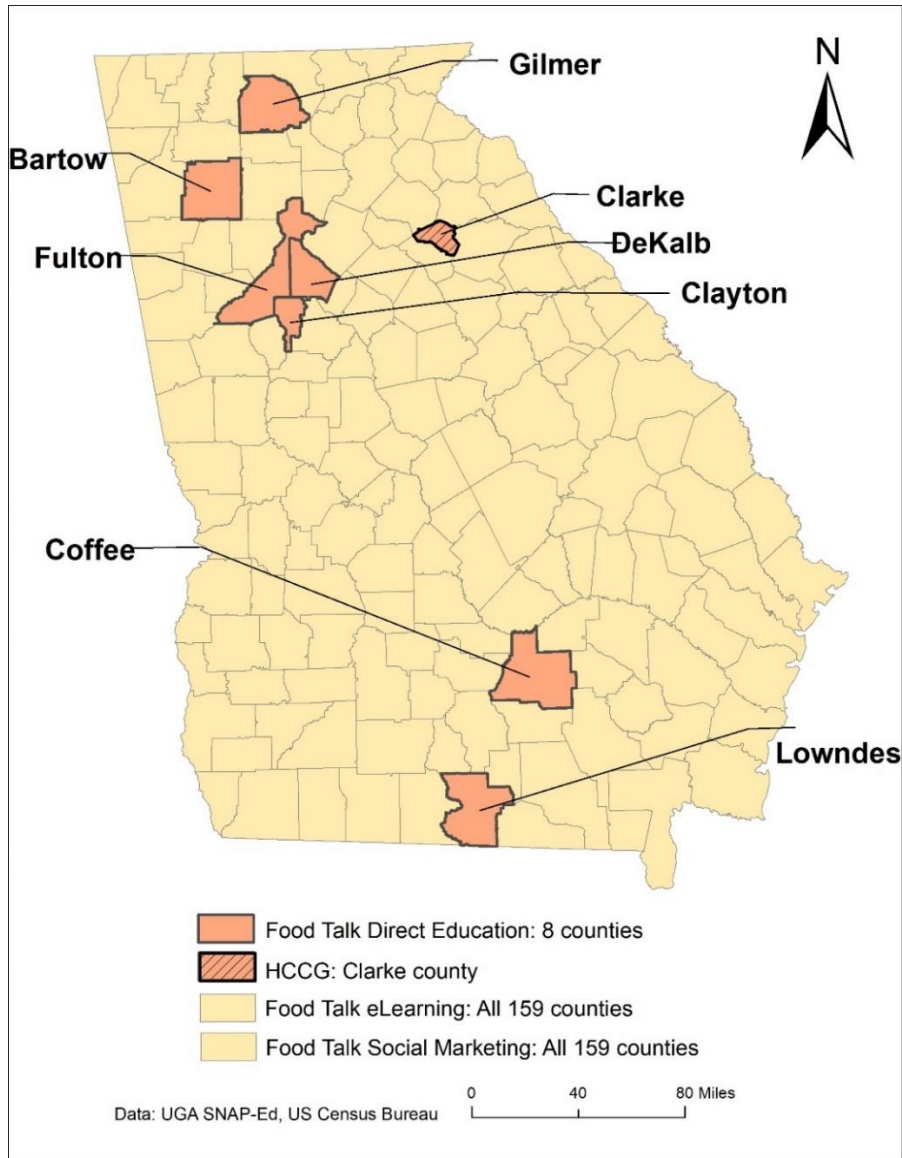


Figure 1 UGA SNAP-Ed Programming Counties

Table 1 Demographic Characteristics of UGA SNAP-Ed Outreach Counties

|                                     | Bartow | Clarke | Clayton | Coffee | DeKalb | Fulton | Gilmer | Lowndes |
|-------------------------------------|--------|--------|---------|--------|--------|--------|--------|---------|
| Total Population                    | 101942 | 122037 | 270269  | 43042  | 723902 | 996757 | 28956  | 109233  |
| % White alone                       | 83.40% | 63.90% | 19.60%  | 67.30% | 34.80% | 45.20% | 90.40% | 58.00%  |
| % Black or African American alone   | 11.00% | 27.20% | 68.10%  | 28.10% | 54.30% | 44.10% | 0.90%  | 36.40%  |
| % Asian alone                       | 0.80%  | 4.40%  | 5.20%   | 0.80%  | 5.90%  | 6.40%  | 0.70%  | 1.90%   |
| % American Indian and Alaska Native | 0.50%  | 0.20%  | 0.20%   | 0.10%  | 0.50%  | 0.40%  | 2.50%  | 0.20%   |
| % Two or More Races                 | 4.30%  | 4.30%  | 6.90%   | 3.70%  | 4.50%  | 3.90%  | 5.50%  | 3.50%   |

To be eligible for SNAP participants, household gross income should be equal to or less than 130% of the federal poverty line. SNAP-Ed are delivered to target population at qualifying locations that meet the general low-income standard, for example, more than 50% of the population have gross income at or below 185% of federal poverty line. In UGA SNAP-Ed program, they specifically targeted on low-income adults in Georgia who are caregivers of children and preschool children, parents of preschool children and Early Care and Education teachers.

### 3.3 Study Area

In this study, we mainly focused on Clayton, DeKalb and Fulton county since most of the survey participants lived in these counties. These three counties are also the three main counties in the Atlanta Metropolitan statistical area which is the most populous metro area in the state of Georgia.

Below is a reference map of the study area:

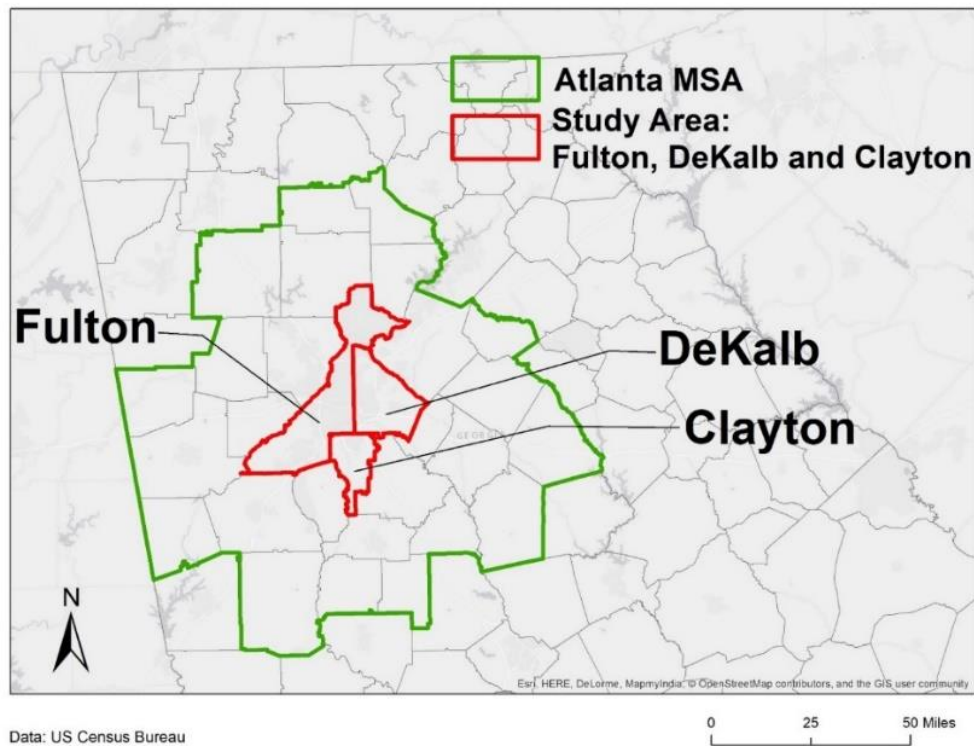


Figure 2 Study Area

### 3.4 Store Survey Data

As part of the SNAP-Education direct nutrition education, Food Talk program participants in the eight programming counties completed a survey on various nutrition-related topics, and one of these is the food environment. In the year 2015, there were 910 Food Talk participants and 746 (82%) of them provided consent; in the year 2016, there were 1567 participants and 1368 (87.3%) of them provided consent; in the year 2017, there were 1285 participants and 1135 (88.3%) of them provided consent.

We obtained data of participant's grocery store choices from participants' responses to one of the survey questions - "The store I most use for grocery shopping is:" and we also obtained background information including home address, age, gender, race and whether

received SNAP or not from other survey responses. Participants also provided data on self-reported weight and height.

We identified grocery stores chains based on participants survey response and coded store chains dummy variables for each participant. Participants grocery store choices include Kroger, Piggly Wiggly, Publix, Ingles Markets, IGA, Whole Foods, Food Depot, Wayfield, Costco, Walmart, Big Bear, Save-a-lot, Aldi and Sam's Club. The value is 1 for the dummy variable if the participant preferred that chain, and the value is 0 if not.

To get the location information of the participants, we geocoded participants' home address using World Geocode Service, Google Map Application Programming Interface (API), and Bing Map API. Most of the address were geocoded by World Geocode Service provided by Esri company using ArcGIS. For those records that couldn't be recognized by World Geocode Service, they were geocoded by Google Map API using "ggmap" package in R (David Kahle, Hadley Wickham, 2013) or they were manually geocoded using Bing Maps. Figure 3 is the maps showing the distribution of participants home address. From the map, most of the participants are clustered in the south metro of the study area.

There were 615 valid participants responses in my study area based on surveys carried out in 2015, 2016 and 2017. Most of the participants were female which accounted for 72% of the participants. The average age of the participants was 53 years old. For those participants that reported their race ethnicity, 86.8% were identified as African American, and 12% were identified as White.

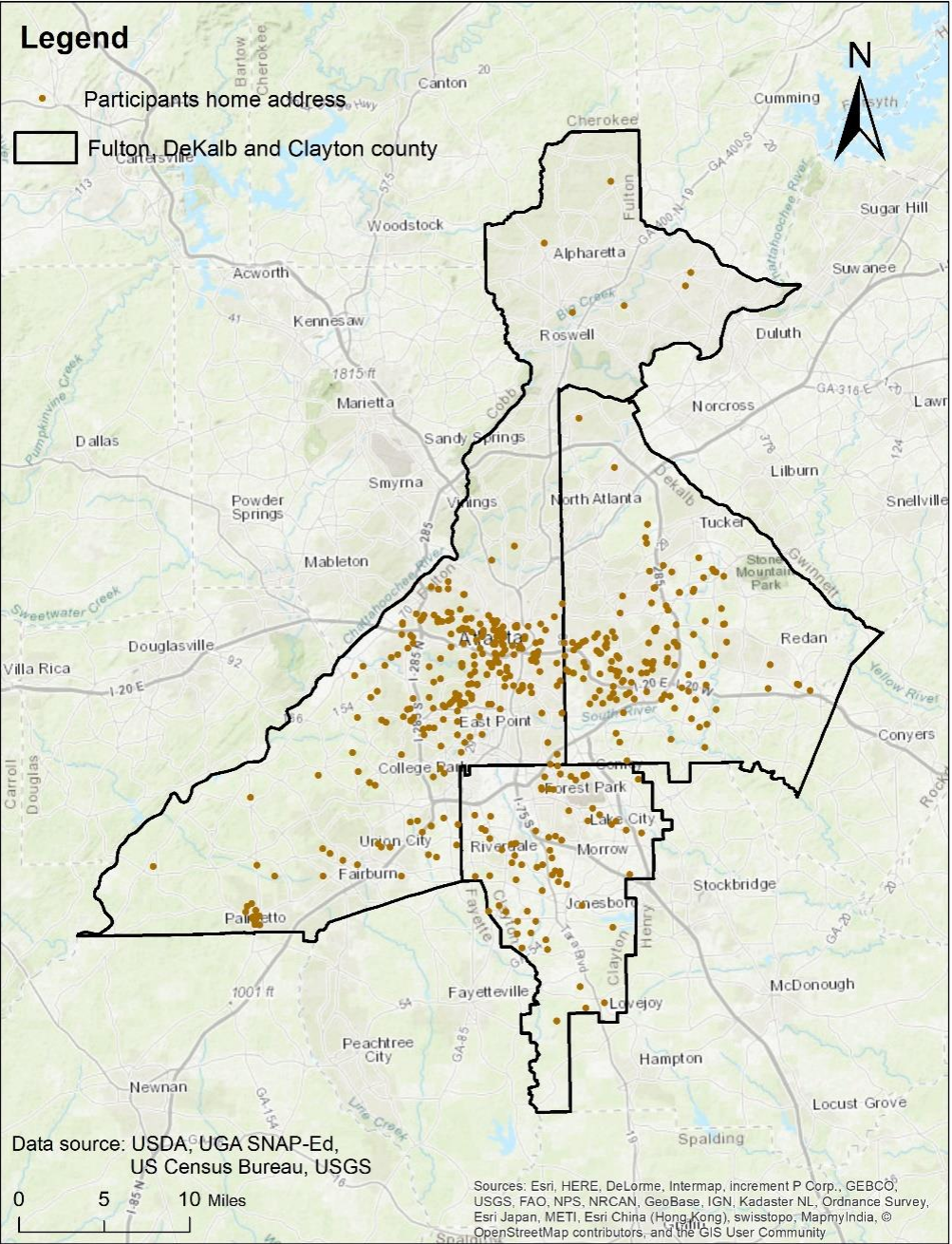


Figure 3 Participants’ Home Addresses

3.5 Food Retailers Data

We obtained food retailers data including supermarkets, convenience stores, gasoline stations and dollar stores from a commercial business and consumer database – InfoUSA

(<https://www.infousa.com/>). In this study, supermarkets are defined as retailers have more than five employees and under the category of the North American Industry Classification System (NAICS) code 44511. The NAICS is the standard used by Federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. In this case, general convenience under the NAICS code 445120, gasoline stations under the NAICS code 447190, and major chain dollar stores under the NAICS code 452990 are considered as convenience store retailers.

We also obtained grocery store data of authorized SNAP retailers from USDA. This dataset includes general information for each retailer such as name, address, and store size. To match up the USDA dataset and the InfoUSA dataset, we harmonized them based on store name and address, then assigned a unique ID to each authorized SNAP retailer. We also developed a coding scheme of store size based on the USDA classification. A list of 254 SNAP authorized retailers under 14 chains identified by participants were subsetted from USDA SNAP authorized data.

We geocoded the location of both general food retailers and SNAP authorized retailers using World Geocode Service and Google Maps API. Figure 4 is the map showing the distribution of SNAP authorized retailers in the study areas. According to the map, bigger chains including Kroger, Walmart, and Publix are more evenly distributed while smaller and local chains are clustered in the center of the study area.

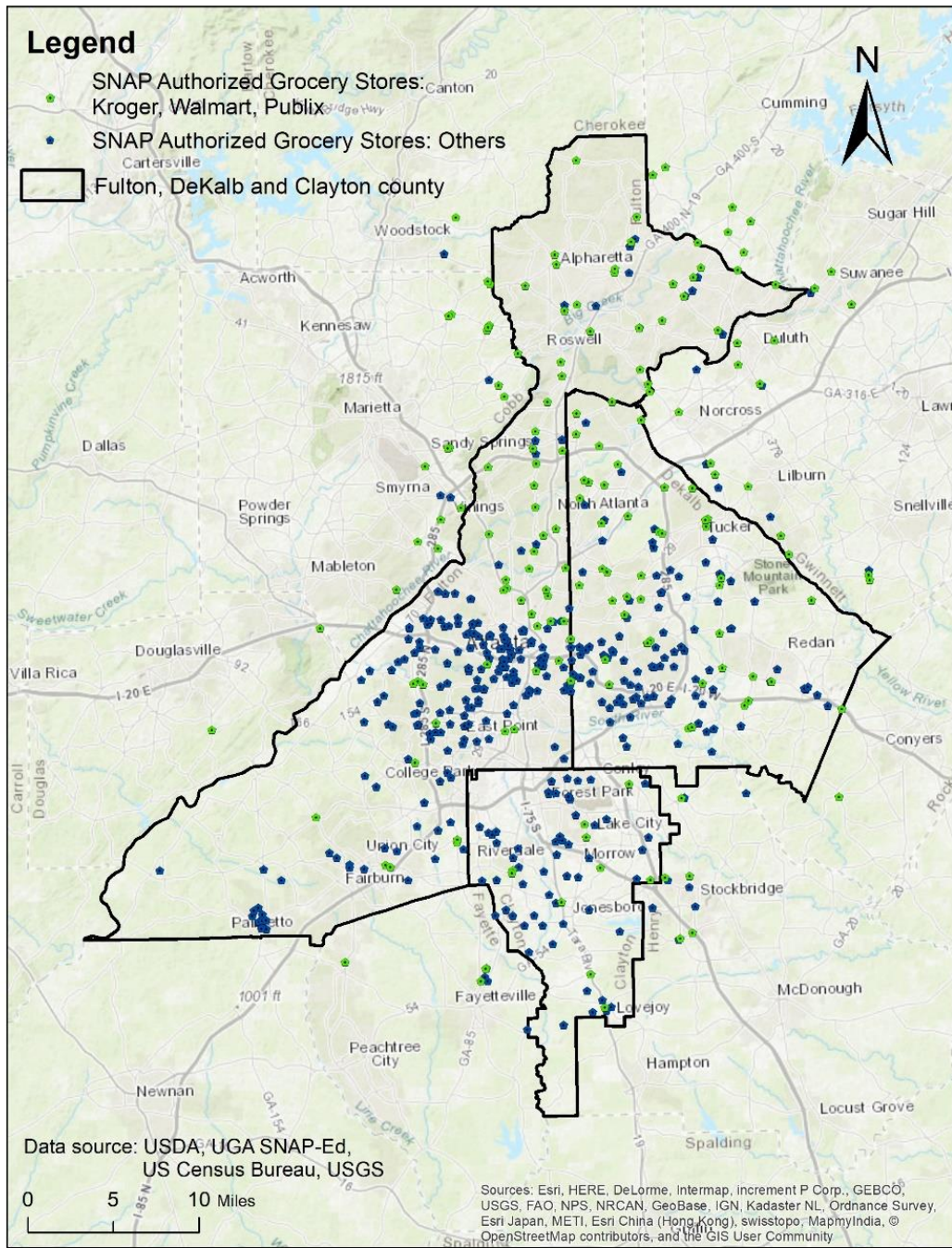


Figure 4 SNAP-Authorized Grocery Retailers

### 3.6 Road Network and Census Data

Road network data were obtained from the U.S. Geological Survey, National Geospatial Technical Operations Center (<https://ngtoc.usgs.gov/>). In this study, we used census tract as neighborhood unit. Census tracts are small, relatively permanent statistical subdivisions of a county with average about 4000 inhabitants of each (“Geographic Terms and Concepts - Tribal Census Tract,” n.d.). Census tract level socio-economic variables including age, population, race ethnicity, education status, gender, household characteristics, housing characteristics and poverty status were obtained from 2015 American community survey five year estimate data in National Historical Geographic Information System (NHGIS)(Manson, Schroeder, Van Riper, & Ruggles, 2017).

### 3.7 Network Analysis

To evaluate the participants’ grocery store environment and how it influences participants’ proximity to preferred stores, we calculated the minimum network distance from participants’ home addresses to the closet location of each participant’s preferred chain. The reason network analysis was used instead of Euclidean distance is that network distance can better represent real-world road information and built environment, which is more accurate than the Euclidean distance to measure food access(Wang, Tao, Qiu, & Lu, 2016). While the county boundaries are not physical boundaries but for administrative purpose, if we conduct spatial analysis within a specific unit, we will probably have bias results along the edge (James E, Gerald M, & L. Rigby, 2009). In this research, participants who live near the county boundaries may also visit grocery stores outside of the boundaries. Neglecting facilities around the boundary of the study area could result in inaccurate accessibility measure. Therefore, a three miles buffer

was added to the boundary of my study area to minimize edge effect considering most of the participants lived in the center part of my study area. Road networks and grocery stores within the buffer were taken into consideration when calculating the network distance. In addition to the distance measure, travel time is another common measurement in lots of food accessibility research. Both network distance and travel time to each nearest grocery store chains of each participant by car were calculated using Google Distance Matrix API (Google, n.d.). Using JavaScript and HTML code, we calculated network distance and travel time with given origin and destination coordinates. The departure time parameter was set default to the current time which was the time we run the script. However, taking computational efficiency into account, for each participant, we selected nearest five stores via Euclidean distance for each chain. To ensure that five was an adequate number, we did a pilot study on Kroger. Distance is not equal to travel time, but the two are related. In this pilot study, we used distance as a proxy for travel time. we identified 5 closet Kroger stores via Euclidean distance for each participant and we also identified the closet store via network distance for each participant using ArcGIS 10.2. Comparing these two results, all the store location in network distance are included in the five nearest pairs in Euclidean distance. Given the result, five is an adequate cut-off number which ensures both accuracy and computational efficiency. After that, we calculated network distance and car travel time of these selected participant-store pairs. Finally, we selected the shortest network distance and travel time from participants' homes to each chain among the five closet stores.

To measure households exposures to food retailers, we measured both supermarkets and convenience stores exposures. Past research has suggested that supermarkets often offer the best access to healthy food options while convenience stores are not good healthy food provides

(Moore & Roux, 2006; Morland, Wing, Diez Roux, & Poole, 2002). There are two common GIS methods for measuring food accessibility: One way is measuring spatial access to stores by counting the number of stores within a certain Euclidean distance or network distance from a unit (Ball et al., 2009; Smoyer-Tomic et al., 2008); another way is by measuring the distance to the nearest or second nearest grocery store (Laraia et al., 2004; Zenk et al., 2005; Zhang et al., 2016). However, the proximity measure doesn't take individual's choices into account, people may not shop at their nearest store, while the network buffer measure gives us a better sense of the overall neighborhood food environment (Chen, 2017). In this case, to measure participants exposure to supermarkets and convenience stores, we used Network Analyst extensions in ArcGIS 10.2 to calculate the number of supermarkets and number of convenience stores within certain network distances from participants' residence places. A total of 210 supermarkets and 1041 convenience stores are identified in my study area. To define the proper buffer distances, we used a buffer distance of 6.87 miles, which is the average distance of the summary statistics of participants minimum network distance to their preferred stores.

### 3.8 Statistical Analysis

First, we calculated descriptive statistics such as mean, median and standard deviations to examine the overall pattern of SNAP-Ed participants' proximities to their preferred food store. For each chain, participants were divided into two groups: one group preferred to that chain for grocery shopping and one not. To examine whether proximity is a factor associated with grocery store choice of SNAP-Ed participant, we used the Wilcoxon test to determine if there is a significant difference in the distances to the chains between these two groups.

To answer the second research question, we used an ordinary least square (OLS) regression model to examine the relationships between SNAP-Ed participants' proximities to their preferred grocery stores and individual, neighborhood and local food environment characteristics. In statistics, with continuous scalar dependent variable and one or more explanatory variables, people usually use OLS to examine their linear relationship. It's also a typical and the most popular linear techniques used to choose the best regression line. (Rocconi, 2013). The traditional OLS has been widely and commonly used in food environment literature to examine the relationship between socio-economic status and food accessibility (Black, Carpiano, Fleming, & Lauster, 2011; Yamashita & Kunkel, 2012).

In my study, my independent variables are organized in two different level as shown in Table 2. The individual and household level characteristics include gender, age, and race which are obtained from survey responses. Also, grocery stores accessibility and convenience stores accessibility are in this level. The neighborhoods census tract level socio-economic variables include age, population, race ethnicity, education status, unemployment, gender, household characteristics, and poverty status.

Table 2 Independent Variables for Proximity Regression Model

| Individual and household level | Neighborhood (tract) level                        |
|--------------------------------|---|
| Age                            | Population Density                                |
| Gender                         | Percent Senior Population                         |
| Race                           | Percent African American                          |
| Count of supermarkets          | Percent Hispanic                                  |
| Count of convenience store     | Percent Female Headed Households                  |
|                                | Percent Population Below 185% of the poverty line |
|                                | Percent Unemployment                              |
|                                | Percent Without High School Diploma               |
|                                | Median Housing Value                              |

However, studies that consider individual variables alone fail to take neighborhood context into consideration may cause atomistic fallacy (ALKER & Jr, 1969), while studies use only aggregated data to infer individual behaviors may cause ecological fallacy (Robinson, 1950). In this case, adjusting for clustered standard errors (CSEs) will provide a solution to this situation. CSEs is usually used in cases when some observations in the data are related to each other. For example, in our data, the correlation happens while participants in the same clusters (tracts, in this case) have identical neighborhood socio-economic variables. In the traditional regression model, the assumption for error term is that errors are independently and identically distributed. But, if observations are grouped into clusters, the model errors may correlate with each other within clusters while model errors are assumed to be uncorrelated. Accurate standard errors are important for an accurate statistical inference, Failing to take error consideration with clusters into consideration can lead to misleading results such as small standard errors, narrow confidence intervals, large t-statistics and small p-value (Cameron & Miller, 2015; Stephanie, 2017). Thus, by adjusting for CSEs, we account for different covariance structures within our data.

In the individual level variables, gender and race variables are categorical variables. Because 72% of the participants are female, to represent the two genders: female and male, we use one dummy variable “gender”: (1) female: gender = 0; (2) male: gender=1. In that way, the coefficient of “gender” in the regression model represents the impact of gender of male on the dependent variable (proximity to the preferred grocery store) compared to the gender being female. In the same way, for race attributes, 86.8% of the participants are black, we use two dummy variables: (1) Black: Race=0; (2) White: Race=1.

While network distance is one way to measure proximity, travel time is a more realistic indicator of proximity. According to the 2012-2016 American Community Survey 5-Year Estimates, more than 93% of households in Atlanta metro statistical area have one or more than one vehicle available. Since travel modes of participants are unknown to us, to examine the sensitivity of this analysis to different measures of proximity, we compared models with two different dependent variables: network distance and car driving time.

To test whether health characteristics of SNAP-Ed participants, specifically Body Mass Index (BMI), associates with grocery store choices. we also used OLS regression to examine the relationship between BMI and grocery store choice. The dependent variable is BMI, the independent variables include age, gender and 14 chain dummy variables. For each chain dummy variable, 0 indicates the participant didn't prefer that chain while 1 indicates the participant prefers that chain.

## CHAPTER 4

### RESULTS

As we can see from Table 3, the most popular three chains are Kroger, Walmart, and Publix. The “count” column indicated how many participants chose this chain as their preferred stores, Kroger was chosen by 388 participants, Walmart was chosen by 140 participants and Publix was chosen by 91 participants. These three chains were also the closest chains both in network distance measure and travel time measure according to the mean and median distance from participants home address to the closest location for each chain. Kroger was the closest chains of these chains with a mean distance of 5.49 miles, followed by Publix with a mean distance of 6.49 miles and Walmart with a mean distance of 6.80 miles. Regarding travel time by car, Kroger was still the closest chain with an average travel time of 12.95 minutes, followed by Publix with a travel time of 14.22 minutes and Walmart with a travel time of 14.5 minutes.

Table 3 Summary Statistics of Participants' Proximity to Each Nearest Store Chains

| Store         | Count | Distance (mi) |        |        |        | Travel Time (min) |        |       |       |
|---------------|-------|---------------|--------|--------|--------|-------------------|--------|-------|-------|
|               |       | Min           | Median | Mean   | Max    | Min               | Median | Mean  | Max   |
| Kroger        | 388   | 0.53          | 4.72   | 5.49   | 18.34  | 2.25              | 12.65  | 12.95 | 27.07 |
| Walmart       | 140   | 0.54          | 6.19   | 6.80   | 21.42  | 2.63              | 14.4   | 14.5  | 31.38 |
| Publix        | 91    | 0.18          | 5.74   | 6.49   | 26.58  | 0.57              | 12.95  | 14.22 | 32.67 |
| Aldi          | 30    | 1.21          | 12.65  | 12.80  | 32.88  | 3.92              | 20.15  | 21.03 | 50.8  |
| Big Bear      | 28    | 0.42          | 11.47  | 12.95  | 40.71  | 2.78              | 17.18  | 18.13 | 47.7  |
| Wayfield      | 26    | 1.57          | 11.90  | 11.87  | 47.17  | 3.13              | 18.2   | 18.75 | 53.85 |
| Food Depot    | 20    | 3.15          | 14.20  | 14.28  | 30.17  | 8.47              | 21.02  | 21.78 | 40.87 |
| Save A Lot    | 13    | 0.96          | 8.60   | 10.26  | 35.76  | 2.92              | 16.17  | 17.32 | 45.38 |
| Sam's Club    | 8     | 1.25          | 22.25  | 22.98  | 48.25  | 4.95              | 26.37  | 27.3  | 54.5  |
| Piggly Wiggly | 7     | 0.72          | 10.00  | 10.86  | 40.07  | 2.78              | 19.05  | 19.28 | 51.08 |
| Costco        | 5     | 2.42          | 20.18  | 22.24  | 47.51  | 9.28              | 25.42  | 27.33 | 52.18 |
| Ingles Market | 3     | 5.92          | 29.10  | 28.22  | 52.35  | 10.17             | 346.65 | 34.22 | 56.4  |
| IGA           | 2     | 0.92          | 15.17  | 14.79  | 45.36  | 3.02              | 22.03  | 22.18 | 54.45 |
| Wholefoods    | 1     | 3.101         | 13.896 | 16.087 | 38.498 | 9.72              | 24.52  | 24.68 | 43.52 |

For each chain, the participants were divided into two groups: based on whether they listed the chain as a preferred store. According to the boxplots of network distances of figure 5, most of the chains, the distance from home to that chain in the group preferred is smaller than the group not preferred. However, if we zoom into Kroger, Walmart, and Publix, the three most popular and closest grocery stores among participants, there was not much difference in the network distance between the group who prefer and do not prefer each chain. As for the driving time, from Figure 6, the overall pattern was similar to network distance except for the Costco. In network distance measure, the median distance of from home to Costco for the preferred group was smaller than the do not prefer one. However, this pattern was reversed for the driving time measure. Possible explanations for this situation could be the small sample size. From Table 3, we can see only 5 participants preferred Costco.

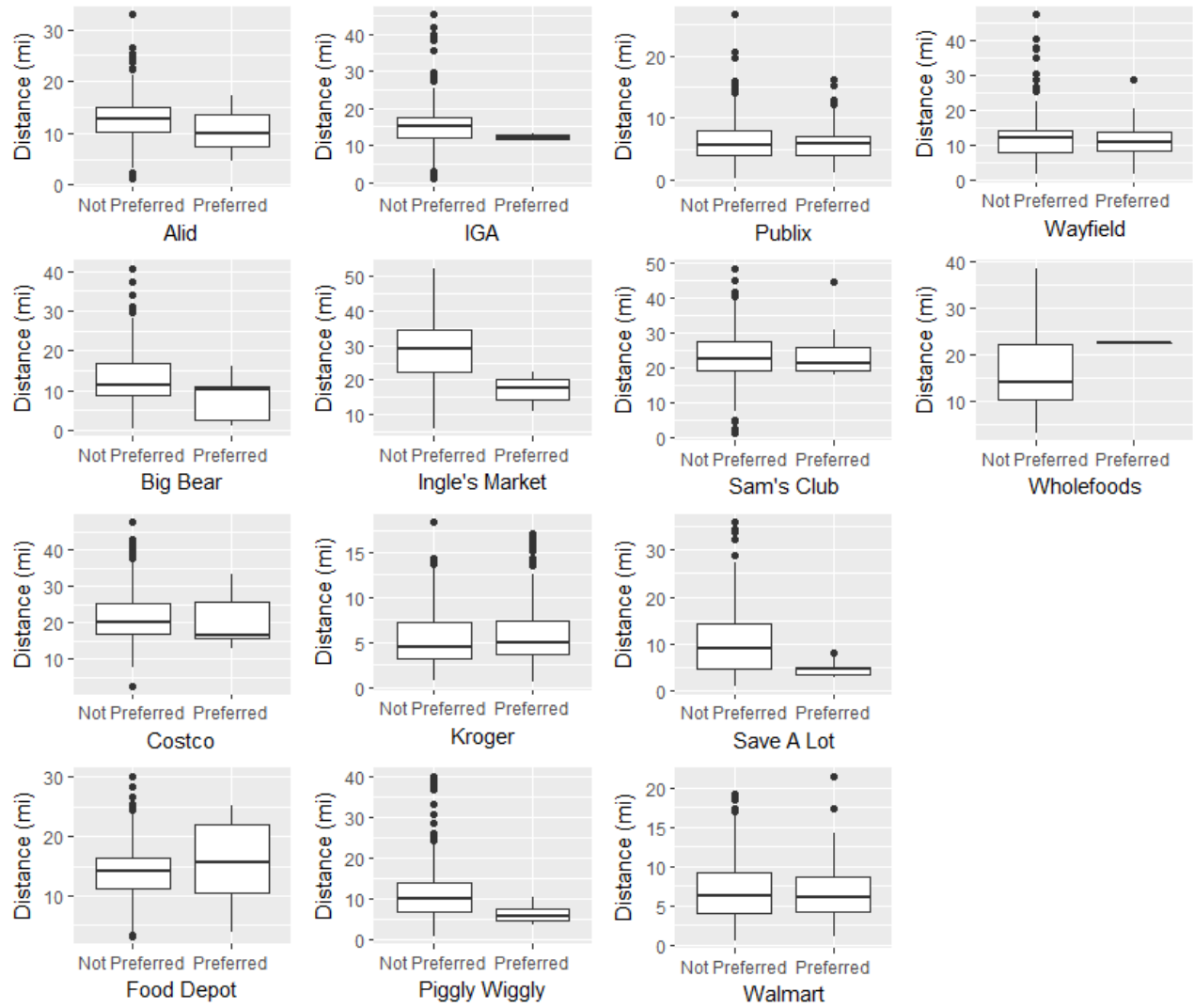


Figure 5 Boxplots of Participants' Minimum Network Distances to each Grocery Store Chains

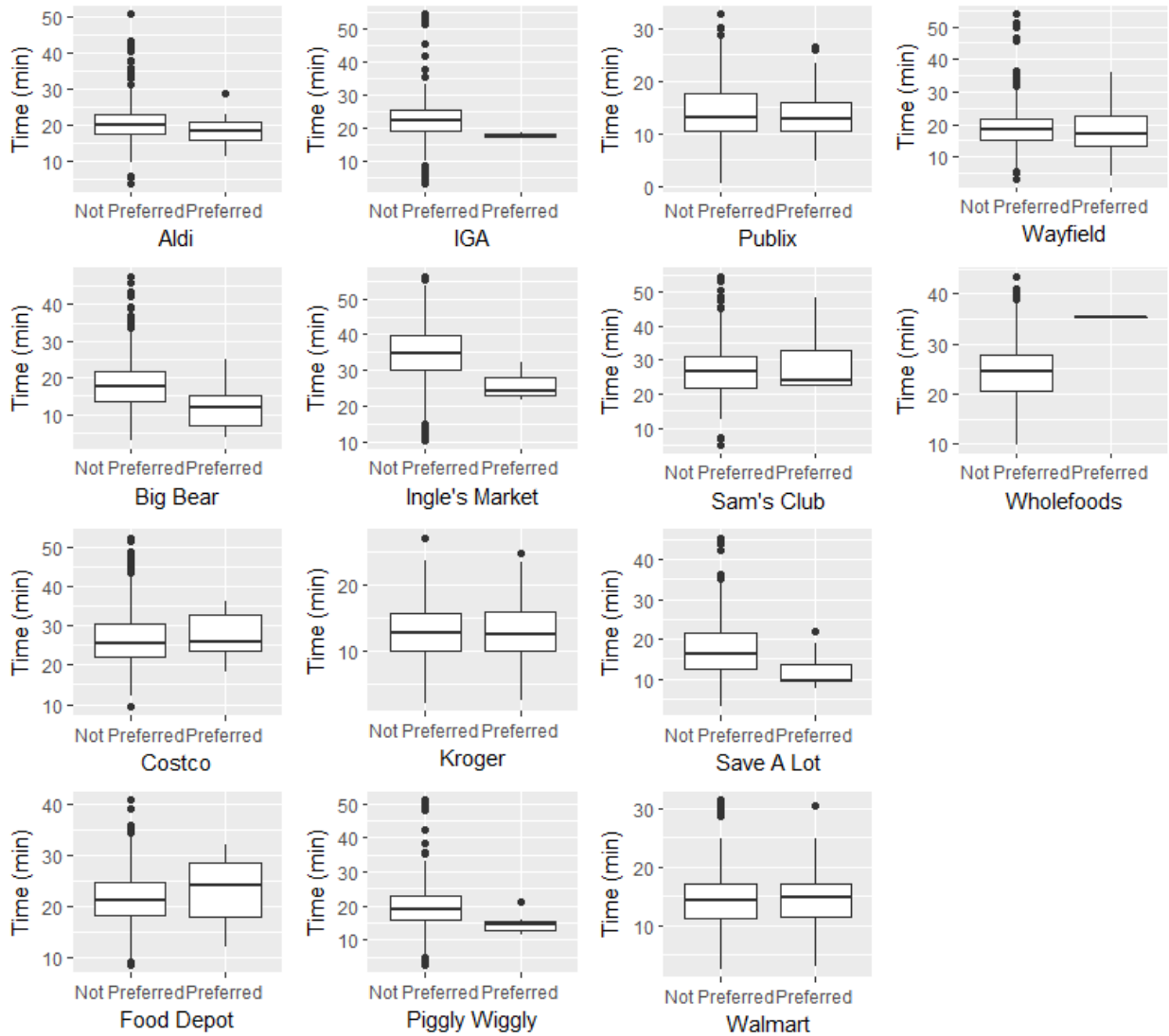


Figure 6 Boxplots of Participants' Minimum Travel Time to each Grocery Store Chains

To test the normality of network distance and driving time, we conducted Shapiro-Wilk normality test. Results indicated that both of these two measures are not normally distributed. Then we conducted Wilcoxon Test, a non-parametric alternative for two-sample t-test, to quantify the relationship of the group preferred and the group not preferred for each chain as reported in Table 4, significant values are highlighted in star signs. The null hypothesis for the Wilcoxon Test is that there is no significant difference between the two groups. The overall

pattern aligned with the box-plot result. Under the network distance measures, there was a significant difference in proximity among the group preferred and the group not preferred for Aldi, Big Bear, Save A Lot, Piggly Wiggly and Ingles market. As for car travel time measure, the pattern was consistent, except for Ingles market, those significant chains remained the same.

While many but not all small chains had significant differences in proximity, bigger chains such as Kroger, Walmart and Publix didn't have significant differences in proximity. We might conclude that proximity seemed to be a less important factor influencing participants' store choices for bigger chains such as Kroger, Walmart and Publix, and there may be some other factors influencing participant's choice such as price, quality, and the varieties of food provided within those chains. Also, the distribution of those chains is also worth considering.

Table 4 Wilcoxon Test of Participants' Proximity to each Nearest Store Chains

| Store         | P-Value    |              |
|---------------|------------|--------------|
|               | Distance   | Driving Time |
| Kroger        | 0.1046     | 0.9182       |
| Walmart       | 0.9942     | 0.1715       |
| Publix        | 0.4120     | 0.1835       |
| Aldi          | 0.0028**   | 0.0311*      |
| Big Bear      | <0.0010*** | <0.0010***   |
| Wayfield      | 0.7732     | 0.6405       |
| Food Depot    | 0.2165     | 0.2461       |
| Save A Lot    | <0.0010*** | 0.00130**    |
| Sam's Club    | 0.9641     | 0.8000       |
| Piggly Wiggly | 0.0136*    | 0.0170*      |
| Costco        | 0.4752     | 0.8865       |
| Ingles Market | 0.0364*    | 0.0699       |
| IGA           | 0.2610     | 0.0840       |

(\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ )

Table 5 is the summary statistics of minimum network distance of each participant's home address to the closest grocery stores they preferred for grocery shopping. For participants

with multiple store choices, average minimum distances were calculated. As we can tell from Table 5, the average minimum distance from participants residential place to their preferred stores is 6.87 miles, the median minimum distance is 5.90 mile. For travel time, the average time is 14.28 minutes and the median travel time is 13.72 minutes. The third quartile for network distance is 8.52 miles.

Table 5. Summary Statistics of Participants' Minimum Network Distances (mi) to Preferred stores

| Proximity             | Min  | Q1    | Median | Mean  | Q3    | Max   |
|-----------------------|------|-------|--------|-------|-------|-------|
| Network Distance (mi) | 0.53 | 3.94  | 5.90   | 6.87  | 8.52  | 33.22 |
| Driving Time (min)    | 2.5  | 11.05 | 13.72  | 14.28 | 16.74 | 38.73 |

Figure 7 is a scatter plots of distance measure and travel time measure. The correlation coefficient of these two measures is 0.85, which indicated there was strong correlation between these two measures.

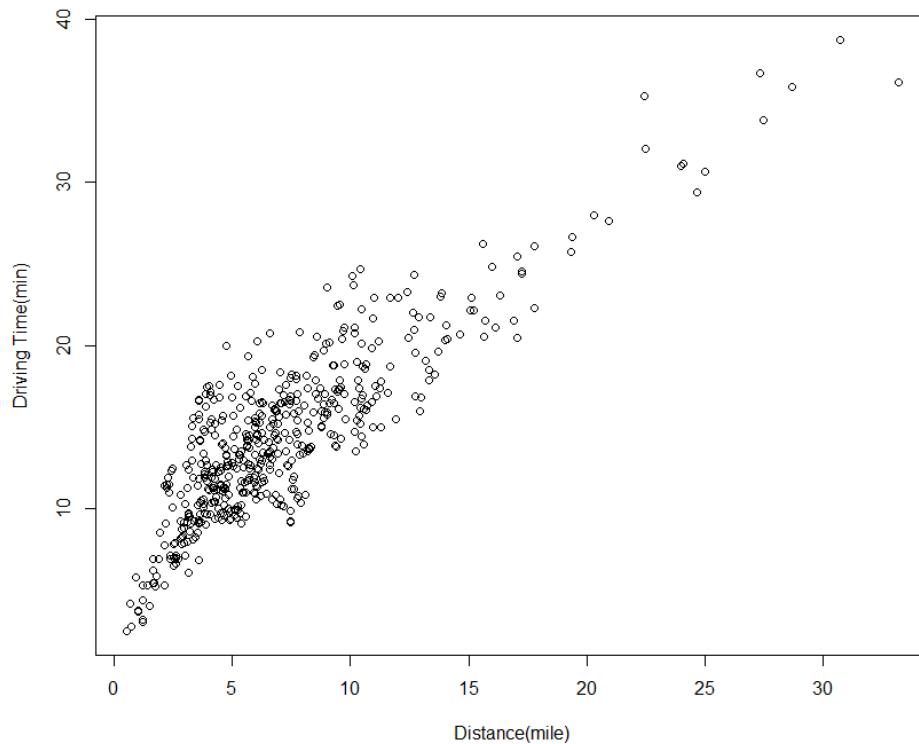


Figure 7 Scatter Plot of Travel Time and Distance

Table 6 showed the results of different regression models. First, we used R-Square ( $R^2$ ) to compare one-level regression model and the clustered standard error model with neighborhood socio-economic variables.  $R^2$ , the percent of the dependent variable variation that is explained by the model, is a common statistical measure to estimate how close the data are to the fitted regression model.  $R^2$  always ranges from 0 to 1, the higher the value, the better the model fit. For distance models, the  $R^2$  of clustered standard error model with neighborhood socio-economic variables (model 2-1) 0.222 was higher than the corresponding one level model (model 1-1) 0.1966. For travel time models, the  $R^2$  of clustered standard error model with neighborhood socio-economic variables (model 2-2) 0.1362 was slightly higher than the corresponding one

level model (model 1-2) 0.1038. The results above indicated that the added socio-economic variables helped explained more variation of the dependent variables for distance models.

Focusing on individual level variables, most of the signs were consistent across distance models (Model 1-1) and travel time models (Model 1-2). In the individual variables distance model (Model 1-1), number of convenience stores within buffer was significantly and negatively related to network distance. That is to say, for each additional convenience store within the buffer area, the distance decrease by 36.53 meters which indicates better accessibility to preferred grocery stores. The number of convenience stores within buffer variable ranged from 0 to 260, the median is 191, the first quartile is 136 and the third quartile is 241. Given the wide range of this variable, distance can decrease about 1 kilometer with additional 27 convenience stores. differences in number of convenience stores can have significant effect on proximity to preferred store. So, the magnitude of this variable is relatively big. In travel time model, number of convenience stores variable was not significant but had the same direction with distance model. Large number of convenience stores are usually associated with high density of commercial area which indicated higher accessibility to different retailers including grocery stores.

For clustered standard error models, the signs were consistent between distance (Model 2-1) and travel time model (Model 2-2). As for the clustered standard error distance model (Model 2-1), number of convenience stores within buffer, was found to have negative relationships with network distance, and was significant at 0.001 level. This relationship between convenience stores numbers and distance were consistent through one level model and the clustered standard error model. The percent population without high school diploma variable was significant and positively associated with distance to grocery stores. Neighborhoods with high

low-educated population generally have low socio-economic status (SES). The outcomes of this variable aligned with previous literature that most low SES communities have less access to healthy food providers. However, other important SES variables such as senior population, minority population, poverty status and unemployment were not significant in this model.

In the clustered standard error travel time model (Model 2-2), the overall model fit is low, with a  $R^2$  of only 0.1362. Given the above results, model 2-1 turns out to be a better model in terms of  $R^2$ . Consistent with the network distance model, number of convenience stores within buffer area us negatively and significantly related to travel time. As for neighborhood level variables, percent population below 185% of the poverty line variable was negatively associated with travel time and significant at 5% level. Participants in neighborhoods with a high percentage of population in poverty appeared to travel less to grocery stores. This outcome was somewhat surprising given the fact that high percentages of poverty population usually associated with low SES.

Overall, regression models of the two different proximity measures are very similar regarding the signs of variables. But they still have slight differences such as model fit and significant variable.

Table 6 Network Distance and Travel Time Regression Models

|  | One Level Model          |                           | Clustered Standard Error Model |                           |
|--|--------------------------|---------------------------|--------------------------------|---------------------------|
|  | Distance<br>(meter, 1-1) | Travel Time<br>(sec, 1-2) | Distance<br>(meter, 2-1)       | Travel Time<br>(sec, 2-2) |
| Intercept                                  | 18477.0***<br>(1707.4)   | 1041.9***<br>(75.37)      | 13293.5*<br>(3.43)             | 1053.8***<br>(5.32)       |
| Individual Level                           |                          |                           |                                |                           |
| Age  | 6.452<br>(19.73)         | 0.487<br>(0.871)          | 10.52<br>(16.46)               | 0.795<br>(0.885)          |
| Gender: Male                               | -2.199<br>(-834.2)       | 42.44<br>(36.83)          | 535.3<br>(851.2)               | 59.65<br>(42.66)          |
| Race Ethnicity: White                      | 350.8<br>(1102.8)        | 29.45<br>(48.68)          | 1047.0<br>(1527.7)             | 30.05<br>(57.11)          |
| Number of Supermarkets within Buffer       | -34.62<br>(52.25)        | -1.765<br>(2.307)         | 71.80<br>(63.47)               | -2.024<br>(3.393)         |
| Number of Convenience Stores within Buffer | -36.53**<br>(13.39)      | -0.945<br>(0.591)         | -52.94***<br>(13.86)           | -1.749*<br>(0.711)        |
| Neighborhood Socio-economic Level          |                          |                           |                                |                           |
| Population Density                         |                          |                           | 0.00576<br>(0.614)             | 0.0413<br>(0.0316)        |
| Percent Senior Population                  |                          |                           | -22.23<br>(36.70)              | -0.0263<br>(2.009)        |
| Percent African American                   |                          |                           | 45.01<br>(26.04)               | 0.972<br>(1.725)          |
| Percent Hispanic                           |                          |                           | 105.3<br>(77.05)               | 4.246<br>(4.675)          |
| Percent Female Headed Households           |                          |                           | 102.4<br>(58.87)               | 0.395<br>(2.860)          |
| Percent Population Below 185% poverty line |                          |                           | -74.45<br>(42.83)              | -5.768*<br>(2.453)        |
| Percent Unemployment                       |                          |                           | -46.14<br>(112.8)              | 6.900<br>(5.540)          |
| Percent without High School Diploma        |                          |                           | 138.0*<br>(62.93)              | 6.414<br>(4.456)          |
| Median Housing Value                       |                          |                           | -0.000694<br>(0.00599)         | -0.000558<br>(0.000376)   |
| R-Square                                   | 0.1966                   | 0.1038                    | 0.2222                         | 0.1362                    |

(Standard errors in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ )

Body Mass Index(BMI) is a commonly used proxy of thinness and fatness, defined as weight in kilograms divided by the square of height in meters (World Health Organization, 2000; World Health Organization & Organization, 2004). Because of data limitations, we only got self-reported BMI data, 480 out of 615 participants have valid self-reported BMI, age and gender information. For those participants, we conducted OLS regression to examine the relationship between BMI and participants' age, gender and grocery store choices. Table 7 showed the results of the regression model, as we can see from the model, participants gender was significantly associated with BMI. Compared to females, male participants tend to have higher BMI compared to female participants. As for grocery store choices, none of the grocery choices dummy variables was found significant.

Table 7 BMI Regression Model

|               | Coefficients        |
|---------------|---------------------|
| Intercept     | 33.78***<br>(1.951) |
| Age           | -0.0212<br>(0.0256) |
| Gender: Male  | -2.442*<br>(0.984)  |
| Ingles Market | 1.994<br>(9.346)    |
| Piggly Wiggly | 1.576<br>(4.008)    |
| Kroger        | 0.352<br>(1.187)    |
| Publix        | -1.019<br>(1.273)   |
| Walmart       | 1.370<br>(1.224)    |
| Aldi          | 0.805<br>(2.073)    |
| Food Depot    | 1.432<br>(2.603)    |
| Save A Lot    | 0.941<br>(3.083)    |
| Sam's Club    | 1.295<br>(3.758)    |
| Wholefoods    | 0.488<br>(9.348)    |
| Wayfield      | 4.273<br>(2.294)    |
| Big Bear      | -0.757<br>(2.090)   |
| Costco        | -0.282<br>(4.478)   |
| IGA           | 0.0569<br>(6.656)   |

(standard errors in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ )

## CHAPTER 5

### DISCUSSION AND CONCLUSION

The first research question in this study is to answer whether proximity is a factor that is associated with grocery choice of SNAP-Ed participants. According to the results, for bigger chains such as Kroger, Walmart and Publix, there is not much difference in the proximity between the group preferred and the group not preferred. Proximity seems to be a less important factor influencing participants' store choices for those chains. The patterns are a little bit vague for other chains due to relatively small sample sizes. Grocery store choice is very complicated and can be affected by numerous factors not just proximity but others including price, food quality, food varieties, etc. Other studies have indicated that price is an important motivation influencing people's food purchase behavior (French, 2003; Mabli & Worthington, 2015).

The second research question in this study is to find out the relationship between SNAP-Ed participants' proximity to their preferred grocery stores and individual, neighborhood and local food environment characteristics. And how these relationships vary by different measurements of proximity (network distance and travel time)? From the result, considering only individual characteristics of participants alone is insufficient to explain their proximity to grocery stores, adding neighborhood level socio-economic data helps to explain more variations in distance model. Household level characteristics such as number of convenience stores within buffer distance, and neighborhood level characteristics such as percent population in poverty and percent population without high school diploma are helpful in explaining participants' proximities to their preferred stores. Also, even the two different measures of proximity (network

distance and car travel time) are correlated, there are still differences in model results between these two measures. While measuring proximity, many factors need to be considered: what are participants' mobility patterns, what transportation mode do they take, what are the physical transportation infrastructures around them? Lack of controlling for the above factors might be an explanation to the differences in distance measures and travel time measure.

The third research question in the study is to find out do BMI of SNAP-Ed participants associate grocery store choices. The result indicates that BMI of participants don't seem to vary by their store choices. However, because the limit of the data, we couldn't make any confident conclusion on this.

Overall, from this research, proximity doesn't seem to be a significant factor influencing UGA SNAP-Ed participants' grocery store choices, especially on bigger chains. Various measures of proximity don't vary much, and only have slight differences in modeling the relationships between individual and neighborhood socio-economic variables. In another word, in this research, the two measures basically tell the same story. Also, given the data we have, participants' BMI do not vary by store choices among bigger chains.

This study also has some limitations: 1) The sample size of participants' survey data is not big enough and the data could be biased. In this study, participants are low-income population, most of them are female and most of them are African American. 2) we don't know the actual store location that participants go, the proximities are calculated based on the closet chains they prefer. Also, the departure time we used to calculate network distance and travel time using Google Distance Matrix was not specified and varied across the data. 3) The overall  $R^2$  of our regression models are not good, we still have space to improve the models by selecting proper independent variables. 4) We didn't take urban and rural settings into account. 5) The

accuracy of self-reported BMI is doubted, according to literature, people tend to report their BMI lower than the actual value. For example, only 60% of participants in 2015 accurately perceived their weight status (Bailey & Lee, 2017).

For future research, it would be helpful to obtain more SNAP-Ed participants survey data and also get participants' in-store shopping records data, which could help us find other significant factors affecting grocery store choices. For example, we can ask participants to keep the receipts of their grocery shopping. From the receipt, we would be able to know which store they go for shopping and what they buy in the stores. Also, it's necessary to involve participants' mobility pattern and transportation modes in measuring spatial proximity to grocery stores. Wearable GPS tracking devices would be able to help us track participants' mobility pattern and we can add a question about transportation method for grocery shopping to the survey to obtain transportation modes information.

Understanding how proximity and other factors affecting SNAP-Ed participants' food choices, examining the relationships between socio-economic status and food environment could help policymakers make better decision on targeting nutrition intervention and take neighborhood effects into account.

## REFERENCES

- ALKER, & Jr, H. R. (1969). A typology of ecological fallacies. *Quantitative Ecological Analysis in the Spatial Science*, 69–86. Retrieved from <http://ci.nii.ac.jp/naid/10003011526/en/>
- Bailey, C., & Lee, J. S. (2017). Overweight and Obesity, Weight Perception, and Weight Management Practices Among Supplemental Nutrition Assistance Program–Education (SNAP-Ed) Participants in Georgia: A Needs Assessment. *Journal of Nutrition Education and Behavior*, 49(5), 422–426.e1. <https://doi.org/10.1016/j.jneb.2017.02.007>
- Ball, K., Timperio, A., & Crawford, D. (2009). Neighbourhood socioeconomic inequalities in food access and affordability. *Health and Place*, 15(2), 578–585. <https://doi.org/10.1016/j.healthplace.2008.09.010>
- Black, J. L., Carpiano, R. M., Fleming, S., & Lauster, N. (2011). Exploring the distribution of food stores in British Columbia: Associations with neighbourhood socio-demographic factors and urban form. *Health and Place*, 17(4), 961–970. <https://doi.org/10.1016/j.healthplace.2011.04.002>
- Black, J. L., & Macinko, J. (2008). Neighborhoods and obesity. *Nutrition Reviews*, 66(1), 2–20. <https://doi.org/10.1111/j.1753-4887.2007.00001.x>
- Bleich, S. N., Jones-smith, J., Wolfson, J. A., & Zhu, X. (2015). The Complex Relationship Between Diet And Health. *Bleich, S. N., Jones-Smith, J., Wolfson, J. A., & Zhu, X. (2015). The Complex Relationship Between Diet And Health.* *Htts://Doi.Org/10.1377/Hlthaff.2015.0606*, 34(11), 1813–1820. [Htts://Doi.Org/10.1377/Hlthaff.2015.0606](https://doi.org/10.1377/hlthaff.2015.0606)
- Bower, K. M., Jr, R. J. T., Rohde, C., & Gaskin, D. J. (2014). The intersection of neighborhood racial segregation , poverty , and urbanicity and its impact on food store availability in the

- United States. *Preventive Medicine*, 58(1), 33–39.  
<https://doi.org/10.1016/j.ypmed.2013.10.010>
- Cameron, C. A., & Miller, M. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Centers for Disease Control and Prevention. (2014). General Food Environment Resources. Retrieved from <https://www.cdc.gov/healthyplaces/healthtopics/healthyfood/general.htm>
- Charleux, L. (2017). Deriving Mobility Archetypes from Household Travel Survey Data. *The Professional Geographer*, 0124(August), 1–12.  
<https://doi.org/10.1080/00330124.2017.1338588>
- Chen, X. (2017). Take the edge off: A hybrid geographic food access measure. *Applied Geography*, 87, 149–159. <https://doi.org/10.1016/j.apgeog.2017.07.013>
- Chen, X., & Clark, J. (2016). Measuring Space–Time Access to Food Retailers: A Case of Temporal Access Disparity in Franklin County, Ohio. *The Professional Geographer*, 68(2), 175–188. <https://doi.org/10.1080/00330124.2015.1032876>
- Chen, X., & Kwan, M. P. (2015). Contextual uncertainties, human mobility, and perceived food environment: The uncertain geographic context problem in food access research. *American Journal of Public Health*, 105(9), 1734–1737. <https://doi.org/10.2105/AJPH.2015.302792>
- Coveney, J., & O’Dwyer, L. A. (2009). Effects of mobility and location on food access. *Health and Place*, 15(1), 45–55. <https://doi.org/10.1016/j.healthplace.2008.01.010>
- Cummins, S., Clary, C., & Shareck, M. (2017). Enduring challenges in estimating the effect of the food environment on obesity. *American Journal of Clinical Nutrition*, 106(2), 445–446. <https://doi.org/10.3945/ajcn.117161547>
- Cummins, S., & Macintyre, S. (2006). Food environments and obesity — neighbourhood or

- nation? *International Journal of Epidemiology*, 35(1), 100–104.  
<https://doi.org/10.1093/ije/dyi276>
- Cummins, S., Macintyre, S., & Glasgow, G. (2002). “ Food deserts ”— evidence and assumption in health policy making. *Education and Debate*, 325(August), 436–438.
- David Kahle, Hadley Wickham. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, 5(1), 144–161. Retrieved from <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>
- French, S. A. (2003). Pricing effects on food choices. *The Journal of Nutrition*, 133(12), 841S–843S.
- Gallagher, M. (2011). USDA defines food deserts. Retrieved from <http://americannutritionassociation.org/newsletter/usda-defines-food-deserts>
- Geographic Terms and Concepts - Tribal Census Tract. (n.d.). Retrieved from [https://www.census.gov/geo/reference/gtc/gtc\\_tr\\_ct.html](https://www.census.gov/geo/reference/gtc/gtc_tr_ct.html)
- Ghosh-Dastidar, M., Hunter, G., Collins, R. L., Zenk, S. N., Cummins, S., Beckman, R., ... Dubowitz, T. (2017). Does opening a supermarket in a food desert change the food environment? *Health and Place*, 46(June), 249–256.  
<https://doi.org/10.1016/j.healthplace.2017.06.002>
- Google. (n.d.). Maps JavaScript API. Retrieved from <https://developers.google.com/maps/documentation/javascript/tutorial>
- Gustafson, A., Christian, J. W., Lewis, S., Moore, K., & Jilcott, S. (2013). Food venue choice, consumer food environment, but not food venue availability within daily travel patterns are associated with dietary intake among adults, Lexington Kentucky 2011. *Nutrition Journal*, 12(1), 17. <https://doi.org/10.1186/1475-2891-12-17>

- James E. B., Gerald M. B., & L. Rigby, D. (2009). *Elementary Statistics for Geographers*. Guilford Press.
- Joassart-Marcelli, P., Rossiter, J. S., & Bosco, F. J. (2017). Ethnic markets and community food security in an urban “food desert.” *Environment and Planning A*, 49(7), 1642–1663.  
<https://doi.org/10.1177/0308518X17700394>
- Kwan, M.-P. (2012). The Uncertain Geographic Context Problem Mei-Po. *Annals of GIS*, 18(4), 245–255. <https://doi.org/10.1080/19475683.2012.727867>
- Lamichhane, A. P., Warren, J., Puett, R., Porter, D. E., Bottai, M., Mayer-davis, E. J., & Liese, A. D. (2013). Spatial patterning of supermarkets and fast food outlets with respect to neighborhood characteristics. *Health & Place*, 23(1), 157–164.  
<https://doi.org/10.1016/j.healthplace.2013.07.002>
- Laraia, B. A., Siega-Riz, A. M., Kaufman, J. S., & Jones, S. J. (2004). Proximity of supermarkets is positively associated with diet quality index for pregnancy. *Preventive Medicine*, 39(5), 869–875. <https://doi.org/10.1016/j.ypmed.2004.03.018>
- Larson, N. I., Story, M. T., & Nelson, M. C. (2009). Neighborhood Environment: Disparities in Access to Healthy Foods in the U . S . *American Journal of Preventive Medicine*, 36(1), 74–81. <https://doi.org/10.1016/j.amepre.2008.09.025>
- Lewis, L. B., Sloane, D. C., Nascimento, L. M., Diamant, A. L., & Guinyard, J. J. (2005). African Americans ’ Access to Healthy Food Options in South Los Angeles Restaurants. *American Journal of Public Health*, 95(4), 668–673.  
<https://doi.org/10.2105/AJPH.2004.050260>
- Lytle, L. A., & Sokol, R. L. (2017). Measures of the food environment: A systematic review of the field, 2007–2015. *Health and Place*, 44(July 2016), 18–34.

<https://doi.org/10.1016/j.healthplace.2016.12.007>

Mabli, J., & Worthington, J. (2015). The Food Access Environment and Food Purchase Behavior of SNAP Households. *Journal of Hunger & Environmental Nutrition*, 10(1), 132–149.

<https://doi.org/10.1080/19320248.2015.1004221>

Manson, S., Schroeder, J., Van Riper, D., & Ruggles, S. (2017). IPUMS National Historical Geographic Information System: Version 12.0 [Database]. Retrieved from

<http://doi.org/10.18128/D050.V12.0>

Mazidi, M., & Speakman, J. R. (2017). Higher densities of fast-food and full-service restaurants are not associated with obesity prevalence. *American Journal of Clinical Nutrition*, 106(2), 603–613. <https://doi.org/10.3945/ajcn116.151407>

Moore, L. V., & Roux, A. V. D. (2006). Associations of Neighborhood Characteristics With the Location and Type of Food Stores. *American Journal of Public Health*, 96(2), 325–331.

<https://doi.org/10.2105/AJPH.2004.058040>

Morland, K., Wing, S., Diez Roux, A., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine*, 22(1), 23–29. [https://doi.org/10.1016/S0749-3797\(01\)00403-2](https://doi.org/10.1016/S0749-3797(01)00403-2)

National Research Council. (1989). *Diet and health: implications for reducing chronic disease risk*. National Academies Press.

Powell, L. M., Slater, S., Mirtcheva, D., Bao, Y., & Chaloupka, F. J. (2007). Food store availability and neighborhood characteristics in the United States. *Preventive Medicine*, 44(3), 189–195. <https://doi.org/10.1016/j.ypmed.2006.08.008>

Robinson, W. S. (1950). Ecological correlations and the behaviors of individuals. *American Sociological Review*, 15(January 2009), 351–357. <https://doi.org/10.1093/ije/dyn357>

- Rocconi, L. M. (2013). Analyzing multilevel data: Comparing findings from hierarchical linear modeling and ordinary least squares regression. *Higher Education*, 66(4), 439–461.  
<https://doi.org/10.1007/s10734-013-9615-y>
- Sadler, R. C., & Gilliland, J. A. (2015). Comparing children's GPS tracks with geospatial proxies for exposure to junk food. *Spatial and Spatio-Temporal Epidemiology*, 14–15, 55–61.  
<https://doi.org/10.1016/j.sste.2015.09.001>
- Shannon, J. (2014). Food deserts: Governing obesity in the neoliberal city. *Progress in Human Geography*, 38(2), 248–266. <https://doi.org/10.1177/0309132513484378>
- Shannon, J. (2016). Beyond the Supermarket Solution: Linking Food Deserts, Neighborhood Context, and Everyday Mobility. *Annals of the American Association of Geographers*, 106(1), 186–202. <https://doi.org/10.1080/00045608.2015.1095059>
- Sharkey, J. R., & Horel, S. (2008). Neighborhood Socioeconomic Deprivation and Minority Composition Are Associated with Better Potential Spatial Access to the Ground-Truthed Food Environment in a Large Rural Area 1 , 2. *The Journal of Nutrition*, 138(3), 620–627.
- Smith, D. M., Cummins, S., Taylor, M., Dawson, J., Marshall, D., Sparks, L., & Anderson, A. S. (2009). Neighbourhood food environment and area deprivation : spatial accessibility to grocery stores selling fresh fruit and vegetables in urban and rural settings. *International Journal of Epidemiology*, 39(1), 277–284. <https://doi.org/10.1093/ije/dyp221>
- Smoyer-Tomic, K. E., Spence, J. C., Raine, K. D., Amrhein, C., Cameron, N., Yassenovskiy, V., ... Healy, J. (2008). The association between neighborhood socioeconomic status and exposure to supermarkets and fast food outlets. *Health and Place*, 14(4), 740–754.  
<https://doi.org/10.1016/j.healthplace.2007.12.001>
- Stephanie. (2017). Statistics How To. Retrieved from <http://www.statisticshowto.com/clustered->

standard-errors/

United States Department of Agriculture. (2017). Supplemental Nutrition Assistance Program (SNAP). Retrieved from <https://www.fns.usda.gov/snap/supplemental-nutrition-assistance-program-snap>

Wang, H., Tao, L., Qiu, F., & Lu, W. (2016). The role of socio-economic status and spatial effects on fresh food access: Two case studies in Canada. *Applied Geography*, 67, 27–38. <https://doi.org/10.1016/j.apgeog.2015.12.002>

Widener, M. J. (2017). Comparing Measures of Accessibility to Urban Supermarkets for Transit and Auto Users. *The Professional Geographer*, 69(3), 362–371. <https://doi.org/10.1080/00330124.2016.1237293>

Widener, M. J., Metcalf, S. S., & Bar-Yam, Y. (2011). Dynamic Urban food environments: A temporal analysis of access to healthy foods. *American Journal of Preventive Medicine*, 41(4), 439–441. <https://doi.org/10.1016/j.amepre.2011.06.034>

Widener, M. J., Minaker, L., Farber, S., Allen, J., Vitali, B., Coleman, P. C., & Cook, B. (2017). How do changes in the daily food and transportation environments affect grocery store accessibility? *Applied Geography*, 83, 46–62. <https://doi.org/10.1016/j.apgeog.2017.03.018>

Willett, W. C. (1994). Diet and Health : What Should We Eat ? *Science(Washington)*, 264(5158), 532–537.

World Health Organization. (2000). *Obesity: preventing and managing the global epidemic*. (No. 894). World Health Organization.

World Health Organization, & Organization, W. H. (2004). Public health Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Public Health*, 363(9403), 157–163. [https://doi.org/10.1016/S0140-6736\(03\)15268-3](https://doi.org/10.1016/S0140-6736(03)15268-3)

- Yamashita, T., & Kunkel, S. R. (2012). Geographic access to healthy and unhealthy foods for the older population in a U.S. Metropolitan area. *Journal of Applied Gerontology*, *31*(3), 287–313. <https://doi.org/10.1177/0733464810386368>
- Zenk, S. N., Schulz, A. J., Israel, B. A., James, S. A., Bao, S., & Wilson, M. L. (2005). Neighborhood Racial Composition , Neighborhood Poverty , and the Spatial Accessibility of Supermarkets in Metropolitan Detroit. *American Journal of Public Health*, *95*(4), 41–43. <https://doi.org/10.2105/AJPH.2004.042150>
- Zhang, Y. T., Laraia, B. A., Mujahid, M. S., Blanchard, S. D., Warton, E. M., Moffet, H. H., & Karter, A. J. (2016). Is a reduction in distance to nearest supermarket associated with BMI change among type 2 diabetes patients? *Health and Place*, *40*, 15–20. <https://doi.org/10.1016/j.healthplace.2016.04.008>