Gravity models and location-allocation techniques are fundamental to analyzing current retail location’s market areas as well as to conducting retail site suitability analysis. Several current research directions are using fundamental geographic location analysis principles and incorporating some of the fundamental principles from other disciplines such as marketing, real estate and urban planning. Examples of such studies establish models of retail site selection to maximize profits or to minimize costs. An important cost factor is the distance from the customer to the retail site. While most prior studies use Euclidean distance for this type of modeling efforts, most researchers agree that Euclidean distance is not an accurate measure. One objective of this research is to implement a methodology and conduct a case study for constructing a retail analysis along a road network rather than using the more common and less precise Euclidean distance. This study looks at whether the difference between the two methods is significant. The results of this inquiry are mixed in that although there does not exist a significant difference between the two approaches in many of the cases, significant differences
were observed in some of the cases. Differences between approaches are most apparent in study areas with a less complete road network. Another goal of this study was to develop a methodology to include aspects of population mobility, such as rates of disability, levels of car ownership and the use of alternate transportation means into the retail market area analysis. This is an exploratory investigation into one of many other factors that are typically not taken into consideration from other studies with the goal of creating retail market areas. The purpose of this is to see how the inclusion of other factors of population mobility affects the space and distance that each retail location can realistically draw its customers from. Spatial regression analysis was run in Geoda to explore the relationship between the amount of sales for an individual grocery store location and the surrounding socioeconomic variables. The dependent variable was sales and the independent variables were: white population, black population, Asian population, Hispanic population, vacancies, owner occupied housing, renter occupied housing and median household income. We typically found that there was a significant relationship between sales for the stores and median household income, black population and owner occupied housing among others results subsequently discussed.

Index Words: Network, Market Areas, Retail, Transportation
Retail Market Area Analysis Using a Transportation Network with Consideration of Population Mobility: A Case Study of Grocery Store Locations in Downtown Atlanta, Georgia

by

Kevin Russell Krivacsy

B.S., Kennesaw State University, 2006

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

2009
Retail Market Area Analysis Using a Transportation Network with Consideration of Population Mobility: A Case Study of Grocery Store Locations in Downtown Atlanta, Georgia

by

Kevin Russell Krivacsy

Major Professor: Xiaobai Yao
Committee: Marguerite Madden
           Lan Mu

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2009
ACKNOWLEDGEMENTS

I would like to acknowledge and thank my family and friends for supporting me through this process even though none of them were really sure what a thesis consists of. I would also like to thank my fellow students in the Geography building who were always ready to delay working for a moment to relax. I would also like to acknowledge and thank all of the people working in the geography department office for helping me out with getting all of my paperwork completed and always helping me out with whatever I needed. Hwahwan Kim also needs to be acknowledged for allowing me to use some of his business related point data, as I would not have been able to complete this thesis or developed suitable research questions without it. I would also like to acknowledge Dr. Okabe and his research team at the Center for Spatial Information Science at the University of Tokyo for allowing me to incorporate the use of his SANET program into my research and for all of the helpful advice presented on his website. Dr. Okabe is affiliated with both the University of Tokyo as well as Aoyama Gakuin University. Acknowledgement and thanks is also given to Dr. Yao who has helped me out tremendously both with this thesis and my coursework as well as in general guidance. Dr. Madden and Dr. Mu Lan are also thanked for helping me out immensely both in terms of thesis research as well as my coursework.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ....................................................................................................................... iv

LIST OF TABLES ........................................................................................................................................ vi

LIST OF FIGURES ...................................................................................................................................... vii

Introduction .................................................................................................................................................. 1

Research Questions and Objectives ........................................................................................................ 11

Significance of Study .................................................................................................................................. 14

Literature Review ....................................................................................................................................... 16

  Overview of Location Models and GIS .................................................................................................... 16

  Gravity Modeling ...................................................................................................................................... 21

  The Voronoi Diagram ............................................................................................................................... 23

  Market Areas ........................................................................................................................................... 26

  Mobility of Population .............................................................................................................................. 28

  Voronoi Diagrams Along a Transportation Network .............................................................................. 28

Data and Software ....................................................................................................................................... 31

Methodology .............................................................................................................................................. 33

Results .......................................................................................................................................................... 44

Discussion and Conclusion ....................................................................................................................... 81

Scope of Study/Possibilities for Future Research ................................................................................... 84

REFERENCES ............................................................................................................................................. 88
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1: Results from T Test for 1st Research Question</td>
<td>64</td>
</tr>
<tr>
<td>Table 2: Regression Output for Independent Grocery Stores</td>
<td>67</td>
</tr>
<tr>
<td>Table 3: Regression Output for Specialty and International Grocery Stores</td>
<td>70</td>
</tr>
<tr>
<td>Table 4: Regression Output for Chain Grocery Stores</td>
<td>72</td>
</tr>
<tr>
<td>Table 5: Regression Output for Convenience Stores</td>
<td>73</td>
</tr>
<tr>
<td>Table 6: Results from Mobility T Test</td>
<td>76</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1: Flow Chart for comparison of Network market areas to Euclidean market areas .....39
Figure 2: Overview map of the study areas .......................................................................43
Figure 3: Downtown Atlanta Roads ................................................................................46
Figure 4: Colquitt County Roads ......................................................................................47
Figure 5: Independent Grocery Stores in Downtown Atlanta ...........................................47
Figure 6: Chain Grocery in Downtown Atlanta.................................................................48
Figure 7: Specialty and International Grocery Stores.......................................................48
Figure 8: Convenience Stores..........................................................................................49
Figure 9: Market Areas weighted with sales data for specialty and international grocery store locations using Euclidean distance .........................................................50
Figure 10: Market Areas Weighted with Sales data for convenience store locations using Euclidean distance ..................................................................................................50
Figure 11: Market Areas Weighted with Sales data for independent grocery store locations using Euclidean distance ...................................................................................51
Figure 12: Market Areas Weighted with Sales data for chain grocery store locations using Euclidean distance .................................................................................................51
Figure 13: Unweighted Voronoi diagram for Chain grocery stores in Atlanta..................52
Figure 14: Unweighted Voronoi diagram for specialty and international grocery stores in Atlanta ................................................................................................................................53
Figure 15: Unweighted Voronoi diagram for Convenience stores in Atlanta.........................53
Figure 16: Unweighted Voronoi diagram for Independent Grocery Stores in Atlanta...........54
Figure 17: Convenience Stores Weighted Voronoi Diagram – Network..........................55
Figure 18: Chain Stores Weighted Voronoi Diagram – Network.................................55
Figure 19: Specialty and International Grocery Weighted Voronoi Diagram – Network........56
Figure 20: Independent Grocery Stores Weighted Voronoi Diagram – Network..............56
Figure 21: Chain Grocery Stores Convex Hull..........................................................57
Figure 22: Convenience Store Convex Hull..............................................................58
Figure 23: Independent Grocery Stores Convex Hull..................................................58
Figure 24: Specialty and International Grocery Stores Convex Hull............................59
Figure 25: Colquitt County area Grocery Stores Convex Hull......................................59
Figure 26: Independent Grocery Store Unweighted Voronoi Diagram – Network..........60
Figure 27: Specialty and International Grocery Store Unweighted Voronoi Diagram – Network
Diagram – Network ..................................................................................................61
Figure 28: Chain Grocery Stores Unweighted Voronoi Diagram – Network.................61
Figure 29: Convenience Stores Unweighted Voronoi Diagram – Network....................62
Figure 30: Weighted Market Areas for Chain Grocery Stores with Factors of Mobility
Included .......................................................................................................................78
Figure 31: Weighted Market Areas for Convenience Stores with Factors of Mobility
Included .......................................................................................................................79
Figure 32: Weighted Market Areas for Independent Grocery Stores with Factors of Mobility
Included .......................................................................................................................79
Figure 33: Weighted Market Areas for Specialty and International Grocery Stores with Factors
of Mobility Included..................................................................................................80
Introduction

The study of locational optimization and the creation of retail market areas have a long history. These issues have been in existence since the development of classic urban geographical principles such as central place theory. However, these principles of urban geography were not necessarily incorporated into the retail scenario until it became necessary as the consumer retail industry started to become more complex in these urban environments. It dates back to 1909 when Weber studied the locational optimization of a firm in a region, called the Weberian problem. Since then, various kinds of locational optimization problems have been studied in operations research, geography and spatial economics, and have been reviewed from many viewpoints (Okabe and Suzuki, 1997:445). These include everything from location analysis techniques for obtaining the optimal location for a new retail chain to locating emergency services to creating areas of influence for cities. Location analysis has a main goal of optimization according to certain criteria. When looking at retail analysis this goal may be the optimization of sales or perhaps locating to serve an untapped market. In the case of emergency services however, the goal would likely be to minimize the distance or time between clients and facilities. Also, one may desire to create a threshold distance where clients must be able to reach a facility in a certain amount of time and facilities then would need to be located as such concerning facilities such as emergency services. In either scenario, location plays an important role in the identification of areas where the clients have the ability to access the facilities that
they need to have access to. The inclusion of population mobility characteristics is one of many things that a researcher could incorporate when looking at retail geography. Most studies have established market areas based on store attributes and distance. Few have taken into consideration the details about the population (potential customers) themselves.

Most agree that one of the most important decisions a retailer can make is where to locate a retail outlet. Because convenience is so important to today’s consumers, a retail store can prosper or fail solely based on its location (Mendes and Themido, 2004:1). “Store location sometimes contributes as much to a retailer’s brand value as some forms of marketing” (Nwogugu, 2006). Others argue that while store location may be important for a successful retail operation, it is becoming less so. I would agree that there is some truth to this concern, with the internet and other retailing avenues decreasing the importance of location, however in certain situations the location aspects of a retail location are just as important now as they ever have been. For example, location still plays a very important role when either establishing or perpetuating one’s brand image. Locating on Rodeo Drive in Southern California or Peachtree Road in Atlanta, GA can serve to heighten the public’s impression of the brand name even if it is not necessarily optimally located to serve its customers. Also, for some sectors, location remains vital such as grocery stores or restaurants as the internet and other means of receiving these goods and services are not able to handle service of these sectors satisfactorily as of this time for most consumers. This study looks at market areas for grocery store locations because grocery stores remain a sector in the retail industry that has largely not been affected by internet retailing in most areas. These remain places that people must travel to on a regular basis and location continues to play an important role. Defining market areas for other types of retail locations may not be as pertinent as they may not require regular trips or their sales may be partially
cannibalized by online sales. Grocery stores were selected for this study because of the different categories available for analysis as well as the fact that location still plays a vital role in this retail sector, while the importance of location may have diminished in other sectors.

Central place theory is a theoretical concept that concerns things such as core and periphery, attractiveness to a center etc. Originally developed by Christaller, this concept at its most basic level says that cities are central places that provide services to the surrounding area. This is one of the major theoretical constructs in the discipline of geography and its general concepts can be applied to look at a variety of problems. In this study we are essentially incorporating portions of this theory to look at attractiveness to a certain retail location and how a large number of core sites (retail locations in this case) compete with one another for market share across space. The idea of regions of influence, which we are using for retail market area analysis here, can be applied to a variety of problems and across a variety of different scales. Of course areas of influence can be developed in other areas such as areas of influence on a global scale. For example, using financial institution holdings, number of global corporations or military power to look at the areas of influence at a global scale can be accomplished. Another example would be using population of cities to determine their relative influence throughout space. Essentially, the process is the same regardless of the specific type of analysis you are doing and looking at these issues is a fundamental goal of geographers, regardless of the scale. While we are looking at retail locations and are doing this along a shortest path network distance, we are using fundamental geographic ideas related to the classic central place theory to accomplish these goals. Looking at areas of influence spatially is one of the central goals of many geographers, both human geographers and physical geographers.
While extensive past research has been conducted concerning retail location analysis, most studies have conducted this analysis using Euclidean distance instead of network distance. This is troublesome for accurate retail analysis because consumers in many parts of the world do not reach their desired location by traveling in a direct linear way from their home to the retail location. Instead, most consumers must travel using the local road network which could alter the retail location that is closest to them and the location that provides the greatest attractiveness compared to using the Euclidean distance. Perhaps extensive research has not been done on this topic because of the assumption that Euclidean distance and shortest path network distances are close enough to each other to accurately describe a retail location’s market area in most cases. This may very well be the case for most analyses. Yet, based on my literature review, the only research that has been conducted to see if there is a significant difference between the two approaches comes from Dr. Okabe and his research team at the Center for Spatial Information Science at the University of Tokyo. They found that there is indeed a significant difference between defining retail market areas using weighted Voronoi diagrams along a shortest path transportation network compared to using Euclidean distance (Okabe et al., 2008). Granted, this study was conducted in a highly urbanized and densely populated city in Japan with many directed links (one-way streets). The study area was also very small with the retail locations close to one another in this case. It is unlikely that one would see as much of a difference when conducting a similar study in a North American city such as Atlanta, Georgia with fewer directed links, but differences between the two approaches are still very possible even in this environment. While our study area has a relatively complete road network, and the difference may be significant, it is unlikely that the results will be similar to those in Japanese cities. This
is because according to the article, the large number of directed links is one reason for the significant difference between the two approaches.

Perhaps one of the reasons that issues such as the one presented here in this research have not been investigated fully by other researchers is the fact that available software to conduct these analyses was either unavailable or very limited. New developments are changing this situation and allowing for more complicated retail geographic studies to be undertaken. The recent progress in geographic information systems is changing the character of retail marketing and retail geographical analysis. Rich spatial data and easy spatial operations provided by GIS are increasing the capability of analyzing detailed spatial factors. The increasing availability of data as well as new software and increased computer capacity allows users the ability to conduct analysis in different, more computer intensive ways such as along a shortest path network distance. The development and subsequent diffusion of geodemographic databases is seen as being expressly relevant for retailers(O’Malley et al., 1997:188). Over the past 20 years, UK retailers have made increasing use of Geographical Information Systems (GIS). These systems were initially used to help with site selection decisions but have developed as Decision Support Systems (DSS) to help in many areas of marketing decisions(Birks et al., 2003:73-74). These systems allow for a great level of customization keeping in mind that every organization will have different goals and therefore different parameters in which to conduct their studies.

While we have discussed the reasons for measuring distance differently than in most market area analysis studies, distance is only one element in determining the reach of a retail location’s market area. In marketing, one of the most important tasks is the estimation of demand. Many models have been developed since Reilly (1931) that serve to accomplish this goal. Reilly devised a theory that the larger the city, that it would draw from a larger hinterland
surrounding the city. While he applied this theory to cities, it later became an important geographical theory to study a wider range of topics in all subfields of geography. Among some of these models developed since Reilly is the Huff (1963) model, which is one of the most frequently used models in practice. Since 1963, the original Huff model has been extended in numerous ways, but it is still formulated under the assumption that a market area is represented by a continuous plane with Euclidean distance (Okabe and Okunuki, 2001:210). While in this study we are essentially using the Huff model, this has rarely been used along a shortest path network distance in the literature. The Huff model or some variation of it is the preferred method of creating market areas based on some gravity or measure of attractiveness for each location even today. The fact that this model is still being used 45 years after its development proves that it remains one of the most effective means to conduct studies regarding spatial gravity models.

One way to model retail market areas is the use of Voronoi diagrams. While many techniques for location analysis, such as weighted Voronoi diagrams are being used, there are however detractors from the point of view that location analysis techniques using weighted Voronoi diagrams to model retail trade areas are an accurate approach. They feel as though this geometric technique along with many of the other existing store–location models don’t incorporate the distance element properly and that while location may be important for successful retail operations, the current approaches to modeling are ineffective and are no better at analyzing a retail strategy approach than experienced professionals are by just “gut feel” and general observation. For example, criticism is that these models erroneously assume that distance from each community in the trade area to the store is constant and remains constant for each customer, for all trips to the store and for all time periods. While this is a valid concern, it
remains difficult to model this change in travel time along the network throughout the day, although it is possible. While the methods proposed in this research may not produce a completely accurate description of reality, it does address the concern that distance is not incorporated properly. While not perfect, we hope to improve on the current methods by bringing in shortest path network distance into the methodology to improve overall accuracy of the analysis. Future research should include animated maps showing market area changes based upon traffic, network flows, time of day, weekday vs. weekend etc. as well as many other factors including possibly multi-purpose trips made by consumers and how these types of trips can affect and change the market areas.

These models also assume that all residents of each community travel the full distance between their community and the store location whenever they shop – some location models use Newton’s gravity models for location analysis which has also drawn criticism (Nwogugu, 2006). One of the major criticisms that I agree with is the fact that these models don’t consider the effects of site-specific operating costs on the location decision (insurance, taxes etc.), the economics of store operations, the retailer’s cost structure and leverage etc. (Nwogugu, 2006). This study will address some of these concerns by incorporating some of the economics of the store operations such as sales data into this analysis as a measure of attractiveness. Yet, this criticism is a valid one and an analyst can obviously not make a retail strategic decision by simply analyzing locational attributes alone. Of course finance, land values, and a host of other factors will need to be taken into consideration when conducting an all encompassing retail analysis although that level of investigation is beyond the scope of this particular, purely geographical study.
The problem to be examined in this research is the development of a methodology and the conducting of a case study to more accurately model retail market areas and the competitive retail situation in the study area compared to the less accurate and less complete techniques that are currently being used to accomplish this goal. The research problem involves defining market areas for each retail location based on a combination of factors including variables of the store itself as well as the shortest path network distance from the client. This method allows market areas not to simply be defined by the distance and assuming that customers will simply travel to the closest location, but it takes into account attributes of the store itself that will make it more relatively attractive than other locations such as sales or square footage. Instead of just assuming that customers travel to the nearest location, this method recognizes that certain attributes of the stores may give the store more relative clout in the region than other locations, thus overcoming at least a portion of the distance constraints. Factors that may give one retail location more clout than another retail location could include total sales, number of employees or gross leasable area, among other possibilities. These are the most common measures that are found in the literature to use as weights.

When we include attributes of the store location such as total sales, number of employees and GLA(Gross Leasable Area) or a combination of such factors, we are defining the relative attractiveness of one location over another. For example, if we are using only total sales data for each retail location we would say that store locations with a greater total sales number have a greater gravitational pull relative to the other locations in the area. Of course distance still plays a role in this as even if a store location has very large total sales values, its market area may not be that big depending on the number of competitors in close proximity there are. While a retail location with a small total sales value may have a larger market area if there are no competitors
close by, and a location with a large number of total sales may have a small market area if there are many competitors nearby. Gravitational models are derived from the laws of Newtonian physics, based on the balance between the store attractiveness and the distance to the potential customers. In the initial work of Reilly (1931), the law of retail gravitation related the share of customers that an outlet attracts as being inversely proportional to distance they must travel and directly related to the store dimension. A similar formulation was pioneered by Huff (1963) to calculate a probability that a customer patronizes a facility (Mendes and Themido, 2004: 5). This is relevant because of the fact that the Huff model was able to expand on the study by Reilly to include things such as attributes of the store itself which lent itself to a more realistic description of relative attractiveness of a store because they were no longer being treated as equals.

One of the main difficulties encountered in the application of gravitational or other dependence models is the definition of a store retail trade area, which is essential for turnover evaluation. A common geometrical diagram, called the Voronoi diagram, allows the combination of location information and other store attributes, with consumer behavior, to generate influence areas. Voronoi diagrams are often attributed to Thiessen, Voronoi or Dirichlet (Mendes and Themido, 2004: 11). To go into more algorithmic details on how these Voronoi diagrams work, Voronoi diagrams are a geometrical method for determining the area surrounding a certain point that is closer to that point than any other point in the study area. Essentially, Voronoi diagrams divide an area in smaller areas containing just one point and all of the area within these smaller areas is closer to that specific point than any other point in the study area. Using weights on the creation of the Voronoi diagrams adds to the distance element attributes of the location itself to determine how big its portion of the overall Voronoi diagram should be relative to the other locations. When using a weighted Voronoi diagram, the
assumption is that customers choose the store considering a trade-off between proximity and store attractiveness which is based on one attribute of the store or some sort of combination of several store attributes. Attributes that are commonly used in other studies that seek to define trade areas for retail establishments are gross leasable area (GLA), total sales for the location and the total number of employees at the particular location.

Yet, studies demonstrate that consumers can be less optimal in their shopping trip behavior than could be expected from a purely travel-cost minimizing perspective. Also, the desire to combine shopping purposes into one trip varies widely depending on the type of retail store we are looking at. Of course all of this would also depend on the study area (Benedict et al., 1998). Therefore, these approaches are not a perfect representation as no study can effectively model consumer behavior 100% of the time. It is impossible to accurately create a retail market area without having survey data from individual customers because it is difficult to model personal choices without this type of data. A plethora of future research opportunities will be outlined in a later section that may have the ability to address some of these issues.

Although formal techniques of location analysis have been available for over 50 years, most retailers traditionally make little use of them, relying instead on intuition guided by experience and ‘common sense.’ However, new circumstances such as the changing retail environment, concentration of market power, the increasing number of demanding customers, retail trends towards multi-outlet smaller quality shops and evidence of a growing use of GIS make models and quantitative techniques especially relevant (Mendes and Themido, 2004:15). The importance of conducting this analysis along a transportation network is evident by the fact that this is how consumers reach their destinations, not by a direct linear fashion.
Research Questions and Objectives

1. When defining market areas for retail locations using weighted Voronoi diagrams, is there a significant difference between the result from creating these Voronoi diagrams using Euclidean distance and that using network-based shortest-path distance in the study area?

2. Is there any relationship between store attributes and the surrounding socioeconomic attributes of the population?

3. How can we define retail market areas incorporating mobility aspects of the population such as levels of disability, use of alternate transportation and levels of car ownership?

Our stated problem for this study is how to develop a methodology for accurately conducting a retail analysis and delineating market areas using weighted Voronoi diagrams based upon a more realistic notion of shortest path network distance rather than Euclidean distance. This overall problem can be divided into sub-problems and research questions that, when combined, will allow us to produce a result for the overall research problem. When defining market areas for retail locations using weighted Voronoi diagrams, is there a significant difference between the result from creating these Voronoi diagrams using Euclidean distance and that using network-based shortest-path distance in the study area? This research question will look at whether or not it is worth it for researchers to take into account shortest path network distance when conducting retail trade area analysis or are the results from the more often used
Euclidean distance methods sufficient for this goal. Market areas have been created for both
types of distances and then were compared. Is there any relationship between store attributes
and the surrounding socioeconomic attributes of the population? This is a simple investigation
into how the certain socioeconomic variables present in our established market areas relate to
certain attributes of the store location in that area, such as the total sales. The investigation is to
look at any relationship that may be present between these two variables. The dependent
variable is store sales and the independent variables are: white population, black population,
Asian population, Hispanic population, vacancies, owner occupied housing, renter occupied
housing and median household income.

How can we define retail market areas incorporating mobility aspects of the population
such as levels of disability, the use of alternate means of transportation and car ownership as a
weight for defining market areas? This research question seeks to investigate how the mobility
of the population will affect the retail location’s market area. For example, just using sales as a
weight to create market areas will give us a theoretical market area for that location based on
attractiveness and distance. Yet, what if the population in that defined market area has a high
level of disability and low levels of car ownership? These factors may lead the market area for
that particular retail location to shrink in comparison to other locations that are surrounded by a
very mobile population. This begins to address the issue of looking at the potential customers
themselves rather than simply store attributes when creating market areas.

Most studies however, when looking at the relationship between facilities and clients
assume that the clients all have the same ability to reach certain facilities. These studies may
incorporate distance or travel time to determine what clients are allocated to certain facilities, yet
few take into consideration any aspects of the people themselves and their overall level of
mobility and ability to reach the facility sites. This research attempts to incorporate some of these aspects of population mobility to discover any adjustments that retail planners may need to make when looking at their potential market areas.
Significance of Study

The incorporation of a network when conducting retail analysis studies is rarely used. Compared with the weighted planar voronoi diagram, the weighted network voronoi diagram is rarely found in the related literature (Okabe et al., 2008). The value of this study is great because the anticipated results will be arguably applicable in a broad context, with applicable content in a variety of different fields including marketing, real estate and urban planning in addition to geography. The use of Euclidean distance is very restrictive when retailers consider their locational strategies in detail. Consumers typically access stores through streets, and so distance along streets is different from its corresponding Euclidean distance, particularly in downtown areas with a large number of one-way streets (directed network). When conducting retail analysis in general, I would argue that using Euclidean distance for these studies is not appropriate as consumers are unable to travel to their locations in a direct, linear fashion. This research can also serve as a basis for future research studies such as incorporating flows and travel time rather than distance to look at how the market areas of these retail locations change throughout the day, year etc. Approaches such as this one would also serve to assist marketers to target their potential customers much more effectively. Research such as this could also serve to determine if conducting a wide variety of analysis type functions along a network is worth doing compared to using Euclidean distance. This knowledge could serve to encourage the development of more user-friendly tools and techniques to assist in these types of analyses.
Vandell and Carter (1993) give an overview of the location analysis and market analysis methods being used as well as synthesize the research from all the relevant disciplines including geography, urban planning, marketing, and real estate. Spatial retail analysis studies can be found in a wide variety of disciplines with each discipline approaching these issues from a different theoretical angle. This allows scholars conducting this type of research to explore these types of issues from a variety of different perspectives. It is also important to incorporate mobility aspects of the population into retail market area analysis as it assists planners in determining whether their potential customers have the ability to reach the desired retail location. This allows for a more accurate market area.
Literature Review

Overview of Location Models and GIS

Vlachopoulou et al. (2001) present a method about the site selection process concerning a new set of warehouses to be constructed. By combining a GIS as well as a decision support system (DSS), their study devises a system for the warehouse site selection process, enabling managers to use quantitative and qualitative criteria in order to classify alternative warehouses or visualize the best one (Vlachopoulou, 2001). The methods behind this particular study have been used in the past and this study does not really present anything entirely new to us. It does however give us an example of how these location-allocation and gravity modeling techniques can be customized to incorporate particular decision making elements unique to an organization as opposed to using across the board assumptions. This is important to include in any potential software that may be developed in the future for location analysis as the customizing aspects must be present to accommodate the differences in strategies for different organizations.

Bateman et al. have conducted research using a GIS-based individual travel cost model to examine the impact of certain common simplifying assumptions regarding journey outset origin, routing, and consequent travel cost measures (Bateman et al., 1999). This is a good area to conduct future research as these market areas that we are attempting to develop in this study can be developed by using a wide variety of different measures of distance. Perhaps things such as time, traffic flows etc. could expand upon our knowledge of market areas and areas of influence.
beyond using shortest path network distance to be an even more accurate description of reality and how these areas of influence change throughout the day. This would definitely be an area for future research as one could use animated mapping techniques to show the change in a retail location’s market areas based upon the time of the day and current traffic congestion.

Benoit and Clarke give an overview of how GIS can be used in retail location planning research areas and contrasts the significance of using basic GIS functions such as buffer and overlay with more sophisticated functions such as the incorporation of spatial interaction models. While basic overlay functions allow demographics and the relationship between retail store locations and socio-economic factors, most feel as though it is not sufficient to conduct retail location analysis and more sophisticated tools are not only useful, but necessary for an accurate analysis (Benoit and Clarke, 1997). It remains important however to use these functions such as buffering to gather the demographic attributes of the population as some retail organizations conduct their marketing to target a population that lives a certain number of miles or less from their location. For example, Publix Supermarkets targets their customers three miles out from their stores. By conducting a three mile buffer from a potential store, the demographics of this newly defined region can be obtained to assist in the marketing goals of the organization.

Facility location models deal, for the most part, with the location of plants, warehouses, distribution centers and other industrial facilities. These location models do not account for competition or for differences across facilities and therefore allocate customers to facilities by proximity. In reality, retail facilities operate in a competitive environment with an objective of profit or market share maximization. These facilities are also different from each other in their overall attractiveness to consumers. Facilities differ in the total ‘bundle of benefits’ they offer customers. The basic problem is the optimal location of one or more new facilities in a market
where competition already exists. Assuming that profit increases when market share increases, maximizing profit is equivalent to maximizing market share. It follows that the location objective is to locate the retail outlet at the location that maximizes its market share. Such models are essential for location decisions in the retail industry (Drezner and Drezner, 2004:193). The basis of this research suggests that the larger the market share of a retail location, the better because of the assumption that the larger the market share, the larger the profits. This of course is not always the case as this approach does not take into account the demographics in the market area. One may rather have a very small market area containing extremely wealthy residents than have a larger market area containing poor residents. Depending on the scale of the study and the nature of the study though, one could possibly assume that the larger the market area the better especially if access to any other information of the area was limited.

Clarke (1998) presents an overview of the historical methods that have been used in store location research and retail planning in the UK. The argument here is that many retailers have been slow to adopt new technologies in order to assist in their location analysis goals. When retail organizations do adopt some of these tools and techniques, they often use them at their most basic level and do not incorporate the full range of applications available to them.

One of the central aspects of location-allocation discussions is the p-median problem. The p-median problem is defined as follows: locate a fixed number of facilities, p, in such a manner as to minimize the total weighted distance of serving all customers, when each customer is served by its closest facility (Church and Murray, 2008:282). Suzuki and Hodgson (2005) present a method for incorporating multi-purpose trip making and are critical of classic location-allocation techniques such as the p-median model which they feel erroneously assumes that
patrons seeks only one type of service at a time, patrons obtain this service at the nearest facility, and that the patron is best served by facilities being as close to them as possible. Their research fails to provide new approaches to dealing with these criticisms, although I would agree that the current location-allocation techniques are not perfect. Some of this criticism is also dependent upon what type of facility we are looking at. For example, the criticism of a patron perhaps not being served best by the nearest facility is really dependent on the type of facility being studied and this criticism could not be applied to every location-allocation study. Rahman and Smith (2000) provide an example of how hierarchical location-allocation techniques can be used in facility planning. These techniques would be important when conducting retail planning for a company that utilizes different levels of service such as banks etc. as well as of course many non-retail services such as healthcare.

GIS as well as location-allocation models have been extensively used in facilities planning. There are in fact three main GIS methods in facilities planning. The first method is the buffer zones method which draws buffers around existing facilities proportional to the latter’s size and capacity. It finds holes in the urban areas which cannot be served by the existing facilities. This method does not take into consideration the distribution of population, nor does it consider whether the land identified is suitable for the facility or not. The second method is the allocation method in which population in a network is allocated to the closest known or planned facility. This is similar to the buffer zone method but it takes population distribution into consideration. However, it requires data to be available in a network which is often not available in a GIS database. The third method is land suitability analysis which is one of the most commonly used spatial analysis functions in GIS(Gar-On Yeh et al., 1996:340-341). A true analysis would incorporate all three of these methods. Church (2002) presents a historical
summary and analysis of the developments of GIS and location science and the merging of the two. This research tracks how many of the developments and advances in location science could not have taken place without the increased sophistication of GIS. Many of the location analysis techniques simply would not have been possible using the old analysis techniques. Van Wee et al. (2001) discuss a way to measure accessibility along a network. This research could be expanded upon to potentially define retail market areas based not solely on network distance and attractiveness but also incorporating accessibility of populations into the analysis. This would allow for perhaps an even more realistic depiction of actual retail market areas as some populations that live relatively close to the retail establishment may not have the means of transportation to access that retail establishment.

In service system design there is a distinct difference between a system operating one facility and those operating many facilities. Within a system that operates one facility, all customers or demand are served by that single facility. However, when a system has more than one facility, then the demand must be divided up between the facilities so as to achieve service provision efficiencies. Allocation is the process of determining who is served by which facility (Church and Murray, 2008:259). One also has the opportunity to put constraints on this allocation process. For example, if one facility can only contain a certain number of people or has different operating hours, it would have a capacity constraint where the clients would then need to be allocated elsewhere where capacity remains. Current methods and software allow for the analyzer to set these capacity constraints when running the location-allocation model.

Location models are defined in the literature based on spatial domain (e.g., network, continuous space, hybrid, etc.), type of facility (e.g., point, line, area, etc.), metrics (e.g., covering, median, cost, etc.), number of sited facilities (e.g., one or multiple) and special
structure or constraints (e.g., hierarchical, capacitated, etc.). These basic properties can be used to classify specific types of problems. In the 1970s, popular classifications were public-and private-sector distinctions, as well as continuous space and network-based models. Although these categories remain valid today, models are often classified in terms of the underlying intent. For example, obnoxious location models deal with siting dangerous or unwanted facilities and median models address issues of access in minimizing total weighted distance in locating facilities (Church and Murray, 2008:281-282). This change in classification of location models is likely a good thing as the purpose of the facility location process would drive the decision making process in terms of what type of model to incorporate.

ReVelle et al. (1970) were the first to classify location models into two broad classes of problems: continuous space and discrete network-based models. Location model development has encouraged concentration on the latter of these two areas, yet not enough research has been done. Some of this modeling bias is because a network can capture a number of nuances of spatial variability that are often assumed away in an unbounded, continuous, planar representation. A network model may accurately represent travel distances in an urban area, whereas an unbounded planar model might not when numerous barriers exist (Church, 2002:552).

**Gravity Modeling**

The gravity model is based on Newton’s hypothesis that interaction between the two objects is directly proportional to the mass of the objects and inversely proportional to the
distance between the two objects (Thrall, 2002:89). Hotelling suggested that each customer patronizes the closest facility. These assumptions lead to a large body of literature that culminated in location-allocation models. Huff proposed a model for the estimation of market share captured at a given location. His model is based on the gravity model yet, in his model Huff suggested that the probability that a customer patronizes a facility is proportional to some attractiveness measure (such as square footage of the retail facility) and inversely proportional to some power of distance. This expanded upon the original ideas that customers automatically patronize the closest location and added another element of attractiveness into the equation. Huff’s model is used mainly by marketers as well as urban planners and geographers. Drezner and Drezner conducted a study that applied the Huff gravity model to competitive facility location models (2004). This is important as the Huff model uses both distance as well as a measure of the relative attractiveness of the facility itself to estimate the market share of a facility.

The well-known gravity model of retail trade links the three factors of expenditure, supply and accessibility. This is also one of the oldest retail models, generally attributed to Huff (1963). The model derives its name from the analogue to the physical model of gravitational attraction, which states that the force between two bodies (e.g. planets: analogue – a retailer and a group of consumers) is proportional to the product of the mass of the two bodies (analogue – retail spending and retail floorspace) and inversely proportional to the distance between them (Birkin et al., 2002:145).

The gravity model can be used to derive the primary market area of a real estate development. The gravity model is a procedure whereby the market share of population resident at one location is calculated to interact at another location. Trade area delineation with a gravity
model largely depends on characteristics of the real estate development, as opposed to characteristics of the population(Thrall, 2002:88-89). This is one of the negatives of using these techniques in that we don’t know for sure the actual habits of the customers. Perhaps the customers don’t mind traveling further to another facility based on habit or a host of other reasons that prove difficult to measure. We are unable to gather data relating to personal preferences in this case, although surveys may be an avenue available for future, more accurate research results. This would allow for the creation of retail market areas based on where actual customers are located rather than models.

The Voronoi Diagram

When used in appropriate situations, Voronoi diagrams provide reasonable approximations of real trade areas, quickly and inexpensively, and without the requirement of extensive retail expertise on the part of the analyst(Boots and South, 1997:520). There are several different types of Voronoi diagrams that can be used when modeling retail trade areas. The ordinary Voronoi diagram considers only the locations of the facilities and assumes that customers patronize the nearest facility. The multiplicatively weighted Voronoi diagram considers both locational and non-locational attributes of facilities (size, sales, # of employees etc. as a relative measure of attractiveness) and assumes that the customers select stores on the basis of a trade-off between distance and attractiveness. This is clearly the more complete representation of reality and what will be used in this study. These two types of Voronoi diagrams assign customers to only one facility, therefore the trade areas are spatial monopolies.
This can be seen as a negative because in reality retail market areas will likely overlap with one another. This is a difficult thing to measure though and should be a priority of future research.

The Voronoi diagram is one of the most fundamental data structures in computational geometry. The history of Voronoi diagrams can be traced back to the middle of the nineteenth century. Although the spectrum of scientific disciplines that include interest in Voronoi diagrams is broad, three aspects have been emphasized: (1) their use in modeling natural phenomena, (2) the investigation of their mathematical, in particular, geometrical, combinatorial, and stochastic properties, (3) their computer construction and representation. Accordingly, Voronoi diagrams are useful in three respects: as a structure per se that makes explicit natural processes, as an auxiliary structure for investigating and calculating related mathematical objects, and as a data structure for algorithmic problems that are inherently geometric. In all three applications, efficient and practical algorithms for computing Voronoi diagrams are required (Aurenhammer, 1991).

Okabe and Suzuki review how Voronoi diagrams can solve optimal location problems. The Voronoi diagram is a very simple diagram. Given a set of two or more, but a finite number of distinct points in the Euclidean plane, we associate all locations in that space with the closest members of the point set with respect to the Euclidean distance. The result is a tessellation of the plane into a set of the regions associated with the members of the point set. We call this tessellation the ordinary Voronoi diagram generated by the point set, and the regions constituting the Voronoi diagram, ordinary Voronoi polygons (Okabe and Suzuki, 1997:445).

Boots and South produced a study critical of multiplicatively weighted Voronoi diagrams to model retail trade areas because the results were the production of mutually exclusive trade
areas since the assumption is that customers patronize only one facility, based on a measure of travel cost and the attractiveness of the site. This method proposed allows for the production of overlapping trade areas (Boots and South, 1997). In theory, I think this is an appropriate method and perhaps a better one than the one proposed for this study. This allows some overlapping of market areas along the border which to me seems like a more accurate analysis of reality.

Okabe et al. (2000) goes into great detail concerning the various types of Voronoi diagrams as well as their applications. There are three main types of weighted Voronoi diagrams: the multiplicatively weighted Voronoi diagram, the additively weighted Voronoi diagram, and the compoundly weighted Voronoi diagram. The multiplicatively weighted Voronoi diagram is simply the process of assigning variables to the generator point that serve as a weight relative to the other points. For example, this is commonly used in retail market area analysis where each retail location will have a different weight (ex. Sales figures etc.) that will determine how large its area of influence will be in the structure of the Voronoi diagram. The additively weighted Voronoi diagram is used in scenarios where it is essential to model growth of the areas over time. This allows visualization of how different growth rates for each location will affect the overall market areas. This could also be used using a retail market area analysis as an example. For instance, establishing sales as the attractiveness measure and then assigning each retail location a sales percentage growth goal for each year and see how that would affect the relationship of market areas over time for each location. The compoundly weighted Voronoi diagram is simply a combination of the two other approaches (Okabe et al., 2000).

The general form of Voronoi models is that they combine information on store locations and attributes with assumptions about consumer behavior to generate trade areas. They are most useful in those situations where detailed consumer patronage data is either unavailable or
deemed too costly or time consuming to acquire. Another positive feature, not found in some other models, is that they can be used in either a descriptive or a predictive way. For example, when applied to the outlets of a single retail chain, Voronoi diagrams provide a visual representation of the chain’s location strategy, including its evolution over time. Voronoi diagrams can also be used to identify potential sites for new facilities, as well as indicating the impact of these and other changes on the existing set of facilities (Boots and South, 1997:519-520).

**Market Areas**

The purpose of establishing market areas is to generate areas that surround a particular location that is based on the probability of an individual shopping at that particular location. In a study containing several locations, this is calculated in the way that typically uses gravity models to determine the attractiveness of one location in relation to another location as well as distance. A variety of factors can be used to measure this attractiveness level of one location over another. Location attributes such as the age of the store, total sales, parking spaces available, operating times can all be factored into the gravity model to measure attractiveness. While all these attributes in theory can be used within the gravity model to determine trade areas between retail locations, the most common attribute that the literature points to being used is the square footage of the retail location and total sales of the location, in that larger stores and stores with a higher amount of sales will have a stronger gravitational pull compared to stores with smaller square footage numbers and less total sales. Perhaps the most common gravity model that is used in determining retail trade area analysis is the Huff Model developed decades ago, yet still the most
commonly used model. The Huff Model basically defines a retail location’s trade area by demonstrating that the probability of an individual \(i\) will select alternative \(j\) when given the utility of \(j\) relative to the sum of the utilities of all other choices of which there are \(n\) considered by the individual \(i\). The majority of users of this model incorporate an aspect of accessibility such as travel time or road distance along with the variable that speaks to the attractiveness of a certain site compared to another site such as the square footage of the location. One of these measures can be used for attractiveness or a combination of these factors can be used as the measure of attractiveness as noted by (Suzuki and Hodgson, 2007; O’Kelly, 1999; Boots and South, 1997).

O’Kelly (1999) presents a study where trade areas are defined using data on existing, known customers to the particular facilities from collected data from shoppers. This is a difficult process to conduct as most retail organizations are not willing to share information concerning the location of their customers openly even if they are already collecting this type of information themselves. Although this remains one of the goals for future research, it is difficult to access this type of data to look at the accuracy of our predicted market areas.

“A trade area is a geographic region from which a real estate project draws most of its customers. The population within the trade area has the greatest probability of engaging in exchange with the project being evaluated. Outside a trade area, the population is expected to have a lower level of interaction and exchange with the project. Within the trade area, customers of the real estate project are generally in greater geographic proximity to one another. The converse is expected for customers outside the trade area. Because analysis of a trade area is inseparable from analysis of phenomena and information on a map, application of geographic analysis and geographic technology are required (Thrall, 2002:79).” GIS supports the
identification of catchment or trade areas through creation of Thiessen polygons, also referred to as a Voronoi diagram. A Catchment area is the area serviced by a facility (Church and Murray, 2008:44-45).

**Mobility of Population**

There are additional considerations to be taken when using gravitational model principles to model trade areas such as the impact of differential mobility levels (Mendes and Themido, 2004). Measures of consumer’s mobility will vary, but some studies have pointed to using only rates of car ownership as the measure (Murad, 2007). This is an important aspect to include when modeling retail market areas. Envision, for instance that you create a market area for a retail location based on previously discussed factors and the end result is a large market area. Yet, if the population within that defined market area has a low level of car ownership or a high level of physical disability, then the market area will likely be smaller for this retail location than originally thought. Overall, there was a lack of literature that was discussing this issue. Therefore, it proved difficult to decide which population characteristics are good measures of mobility in this study.

**Voronoi Diagrams Along a Transportation Network**

Most of the major work being done concerning Voronoi diagrams along a defined transportation network comes out of the Center for Spatial Information Science at the University of Tokyo. This may be because of the higher density involved in Japan cities where the network
distance and the Euclidean distances may be vastly different, especially when one is concerning the directionality of networks such as one-way streets in urban areas. Empirical findings have shown that Euclidean distance is significantly different from the shortest path network distance in an urbanized area if the distance is less than 500m (Okabe et al., 2008).

Okabe and Okunuki describe a computational method for estimating the demand of retail stores on a street network using GIS. The “network Huff model” is formulated on a network with the shortest-path distance as an extension of the ordinary Huff model (which assumes a continuous plane with Euclidean distance). This research also demonstrates a method of implementing a computational method in a GIS environment.

Okabe et al. give a general overview of the use of Voronoi diagrams along a network and the significance of them. An empirical finding is shown that demonstrates that Euclidean distance is significantly different from the shortest path distance in an urbanized area if the distance is less than 500m. This means that service areas in urbanized areas cannot be well represented by Voronoi diagrams defined on a plane with Euclidean distance, hence the need for defining trade areas based upon shortest-path network distance or perhaps network time values rather than Euclidean distance (Okabe et al., 2008). Also, a toolbox is proposed to deal with some of these issues. Some of the possible reasons that Euclidean distance is more frequently used to approximate service areas for retail facilities even though service is most commonly achieved through a network is simply that many believe that the shortest-path distance can be approximated by the corresponding Euclidean distance, which may or not be true depending on a variety of factors including the study area involved. This has been proven to be incorrect in highly urbanized areas, especially those containing directed network, i.e. networks containing one-way streets. Also, many feel as though managing spatial data is more difficult on a network
than on a plane. While this may have been true in the past, new tools such as the SANET toolbox are making network-based spatial analysis much easier (Okabe et al., 2008).

An earlier article by Okunuki and Okabe (2002) focuses on the development of a computational method by incorporating the Huff gravity model along a transportation network. Won Bae and Chwa (2004) present a way to compute the Voronoi diagram along a transportation network using shortest path distance.

There are also higher-order Voronoi diagrams which would obviously be applicable in the study of any type of facilities that provided different levels of service at different locations such as healthcare (Okabe et al., 2000). Perhaps Okabe’s main contribution is however the adoption of the Voronoi diagram techniques along a transportation network rather than using the traditional method of Euclidian distance.
Data and Software

The data that I have used in this study include Census demographic data aggregated at the block group level, road network data obtained from the Georgia GIS Data Clearinghouse as well as Business Points Data from MapInfo. MapInfo Business Points is a comprehensive business database containing geographic points of business location in the United States, and corresponding information about those businesses. MapInfo obtained the data from GeoResults, Inc., a database-marketing firm. The data includes both the location of the businesses as well as a variety of attributes corresponding to these business locations such as number of employees, sales volumes, ownership structure and more. The SANET program we are using for spatial analysis along a network was received and downloaded from Dr. Okabe’s research center. The version we are using is SANET 3 and is a toolbox that can be added to ArcGIS. Any data that needed to be obtained concerning mobility of populations, including levels of car ownership, level of disability and alternate transportation means was obtained from Census Transportation Planning Package sources. This data will be at the TAZ level rather than the block group level but this discrepancy is fine for the type of analysis we are doing. This is because we simply had to gather the CTPP data that fell within each market area, so the data being at the TAZ level was not a problem.

The software that was used for this study include ArcGIS 9.3 and 9.2, Geoda, XtoolsPro, ETGeowizards, SANET. Geoda was used for the regression analysis pertaining to the second
research question. SANET was used to create the market areas along the network, while ArcGIS and the toolboxes were used for all other needed functions.
Methodology

The data that we currently have is of the entire state of Georgia, so we obviously needed to select the downtown Atlanta area and create a new layer from that, both the base data as well as the retail point’s data. The retail locations that we have used in this study are only the ones in downtown Atlanta and only those that are classified as “Grocery Stores (except convenience stores and grocery stores with substantial general merchandise)” by the North American Industry Classification Code. The code for this is 445110 and all points with this code have been extracted from the retail point’s layer to form a new layer. The data also contains SIC-8 codes which are the Standard Industrial Classification codes developed by the Department of Commerce. Then, using the SIC-8 codes, these points were divided further into four categories of retail locations: chain grocery stores, independently owned grocery stores, specialty and international grocery stores and convenience stores. This separation of the 445110 retail locations was done in order to reduce the data stress on the various software programs involved in this study and because of the immense amount of variety presented in the 445110 points in terms of the location’s purpose and likely marketing goals. For example, there are both large, chain grocery stores in the 445110 point set as well as small convenience type stores. It makes little sense to analyze the comparative situation between convenience stores and large chain grocery stores as they are less likely to be competing with each other for market space. The decision was also made that it is better to not use all grocery store locations in one analysis, but it
was decided to separate this category into four separate different categories and conduct a separate analysis for each category. It makes more sense to analyze just specialty or international grocery stores separately as they normally don’t compete against large chain grocery stores. The stores were separated into four different categories: chain grocery stores, convenience stores, independent grocery stores and specialty/international groceries based on their SIC-8 codes available to us in the dataset which give specific information concerning the nature of the business. These SIC-8 codes are more specific regarding the purpose of the business than the NAIC codes are. All weights throughout this study have come from the sales data for each retail location that is provided in this dataset. We have total sales data for each point in our dataset as well as the total sales value for each company. Therefore, with each analysis type conducted and discussed in this study, there will be four different categories of results; one for each of the various types of grocery stores.

While it was decided to have four separate grocery store categories for analysis in this study, it should be noted that the specialty and international grocery stores in this case perhaps would have to be studied differently for greater accuracy. When creating market areas for specialty and international grocery stores, it is likely that the rules of market area creation for the other categories do not apply in this situation. This is because of the assumption that these types of grocery stores may have a more specific customer base based on ethnicity than the other grocery store categories. Therefore, the market areas will probably be much more complicated and based more on ethnic populations that these grocery stores are targeting rather than just distance and the store attributes. For example, as discussed earlier, it makes little sense to analyze the market areas for convenience stores and chain grocery stores together as these types of stores will not typically compete with each other for market space. It is difficult to analyze
the specialty and international grocery stores as a whole because there are different categories within this group that will probably not compete with each other. For example, there may be Brazilian grocery stores and Chinese grocery stores both in this category and it is safe to say that these types of grocery stores may not be in direct competition with each other. While a trade off between the weight of sales and distance is an acceptable method for creating market areas for most retail categories, it is likely to be less accurate when looking at specialty and international grocery stores. For these types of grocery stores, it would be better to create market areas based on where the specific ethnicity that it is targeting is living rather than assuming all people in the area are interested in visiting the retail location as we did with the other categories.

When looking at the factors of population mobility, it was brought up that perhaps space should be left between market areas for this analysis if one shrinks. For example, when a market area shrinks because of low levels of mobility, should its neighbors grow as well or should there be empty space in between these market areas to account for this discrepancy? The purpose of using Voronoi diagrams to create market areas is to look at relative market areas so I decided against using this approach. However, market areas with a consideration of population mobility could be conducted in two different ways. One was to gather the demographic information from the original market areas and assigning them to their respective retail point. We could also do this by getting the mobility information using a certain distance buffer around each retail location. For example, simply disregard the original market areas and just create 0.5 mile buffers around each retail location and gather the mobility information from there using a spatial overlay. However, this second method produced no noticeably different results and was not pursued further. In some cases, where retailers are aware of the distance that they are targeting their customers, this may be appropriate.
In order to create Voronoi diagrams as market areas using Euclidean distance and using shortest path network distance, we will have to incorporate a variety of software programs. The weighted Voronoi diagrams can be created for each retail location using Euclidean distance using Pinliang Dongs’s Arcscript in ArcGIS. To create the weighted Voronoi diagrams along a shortest path transportation network, we have used the SANET toolbox to accomplish this. Testing to see if these two approaches are significantly different from one another is a difficult process but should be able to be accomplished using GeoDA or by using the spatial statistics toolbox in ArcGIS as well as the overlay analysis functions that are available in ArcGIS. For this study, the decision was made to conduct an overlay analysis and then compare the areas for any significant difference that may be present. There are other ways that this comparison could have been accomplished however.

Okabe et al. have developed a toolbox called SANET to deal with some analysis techniques along shortest path transportation network distance rather than Euclidean distance. This is what was used to create the retail market areas along shortest path network distance and all of the data preparation required. The toolbox can be added to ArcMap to conduct a wide variety of analysis functions as well as for the preparation of suitable data (Okabe et al., 2006). This toolbox was used in order to create weighted Voronoi diagrams along the prepared transportation network. The SANET toolbox analysis functions require a considerable amount of preprocessing to be conducted before the data is suitable for analysis within this toolbox. The first thing that is needed is to use the “clean polyline shapes” tool included with SANET which cuts a polyline into individual line segments. The “continuous graph” function allows for any possible resulting isolated polylines to be eliminated from the network. SANET also provides a tool for creating a network index file from a polyline shapefile. This must be done in order to
convert a polyline shapefile (in this case, our roads shapefile) into a data set that contains three files to identify and assign link-node connectivity: polyline point shapefile, adjacent node table, and the network index file. Next, we will have to attach points to the network, such as the retail locations as supply points and nodes as demand points. The “network Voronoi diagram” and “huff model” analysis tools were the main analysis functions used in this study. Other analysis and various editing functions available in the SANET toolbox were used as necessary. Using the retail locations as the supply points weighted with their respective sales data figures, market areas were created along the network for each location. Weighted Voronoi diagrams were also created for each retail point using Euclidean based distances using ArcGIS. In order to compare the results of this process with the resulting market areas using Euclidean distance, some conversions had to take place before the data was ready for comparison.

The results of the SANET analysis gives us market areas along the network and it is displayed as polylines. The output also however gave us a large collection of point outputs (the demand points) that are within that section of the Voronoi diagram which correspond to the supply point (the retail location) in this analysis. Using these types of outputs allowed us to convert this into irregular shaped polygons which are needed to compare to the polygons created as market areas using the weighted Voronoi diagram tool and Euclidean distance. Although since nearly every statistical program will simply use the centroid of these polygons to compare them to one another, it is also possible to not create these polygons but rather simply find the centroid of the network market area and use that for comparison. In this case, it was decided to compare the two sets of polygons with an overlay analysis. The Euclidean distance weighted Voronoi diagram was created using the downloaded weighted Voronoi diagram toolbox.
Since the polyline SANET output had to be converted into polygons, the decision was made to convert the polyline output into polygons using the convex hull tool. Polygons were created based on the polylines output using the “features to convex polygons” tool in the ET GeoWizards toolbox. This tool created convex polygons for the polylines that corresponded (were assigned) to each retail location. The convex hull process essentially creates the smallest polygons possible that contain all of the links and nodes that were assigned to each retail point. This had to be done because there is no good, accurate way to compare the extent of polygons (our Euclidean based market areas) and polylines (our network based market areas). Once this process was accomplished, we now had two sets of polygons to compare with one another. This made using the overlay analysis functions in ArcGIS possible in order to calculate the area overlap for each pair of polygons based on the business ID. Once we had completed this process and had the area that polygons overlapped based on the business ID, we could then compare this to the area of the Euclidean distance based market areas. Simply joining these two layers together and then calculating a percentage overlap for each retail location allowed us to compare this percentage with what was our anticipated results which is 1, meaning a perfect overlap. So, we had a column containing percentage overlap for each retail location and then a column of 1 to test it against. The data was then brought into Excel where a two-tailed T test was done in order to see if there was a significant difference between these two sets of data. The null hypothesis for the two-tailed T test was 0, which would mean that there is no difference between the data column containing percentage overlap and the data column containing 100% overlap. After interpretation of the t score, the t critical value and the p value, we were able to determine if there was a significant difference in area overlap between our results and 1. Therefore, after interpreting these statistics, we can discuss whether there is a significant difference in market
areas between creating them using Euclidean distance or creating them using shortest path network distance. This process that was undertaken to compare Euclidean based retail market areas to network-based market areas is outlined below in Figure 1.

This process was also undertaken using non-weighted Voronoi diagrams as market areas. ArcMap as well as Geoda allows for the creation of Voronoi diagrams using the Thiessen polygon creation tool. SANET as well as the Network Analyst Extension for ArcGIS can also be used to create non-weighted market areas along the transportation network. The process of

Figure 1: Flow Chart for comparison of Network market areas to Euclidean market areas
comparing the two results and the creation of the market areas themselves is essentially the same here. It was done simply to test to see if the two types of measuring distance alone were significantly different when weights were involved versus non-weighted situations. The overlay process and t test discussed earlier was also conducted on these non-weighted market areas in order to look for any difference in results that may have come.

Once again, our study area for this was downtown Atlanta and this is an area that has a very complete road network. Knowing this and anticipating any results that we may have seen, it seemed like perhaps there may not be a significant difference between the two approaches because the road network was so complete. In order to test this, the same process discussed above was completed for the Colquitt County, Georgia area, which has a much less complete road network than Atlanta, Georgia does. We examined the results to determine if this makes a difference in any of our results.

The research question concerning the relationship between store attributes and surrounding socioeconomic characteristics is fairly simple. We simply conducted a regression analysis using GeoDA. This was simply exploratory as I looked for any relationship between the sales of one of the retail locations and the surrounding socioeconomic characteristics of the population, specifically in terms of average income of the surrounding population. The regression process in Geoda provided us with a statistical output for interpretation to see what the spatial relationship is between the sales figures for each retail location and the socioeconomic characteristics of the surrounding area.

In order to incorporate aspects of the population relating to their level of mobility, we have utilized the Census Transportation Planning Package data, available to us for free. This data provides information to us on several possible measures of consumer mobility at the census
tract level or TAZ level. These include disability status, use of non-automobile transportation and the number of vehicles available per household. After the market areas for each retail location are established using the sales figures as a weight, we combined the disability, use of non-automobile transportation and car ownership values in this area to the sales data to create a new measure of attractiveness for each retail location. This will be done by creating an equation that combines sales, disability in the predefined study area, use of non-automobile transportation and levels or car ownership in the predefined study area to create a new measure of attractiveness. The CTPP data for each of these categories was joined with the retail points based on the original market areas that were created.

In order to accomplish this, the formula that was created assumed several things. One, as stated previously, that the total sales for a retail location is an accurate measure of attractiveness. Also, a high number of disabled people in an area will lesson that retail location’s market area or clout relative to other areas, not necessarily absolutely. Third, if an area has low number of vehicles available per household, a high level of non-automobile transportation use can make up for a portion of this deficiency. Since both the number of automobiles available and the use of public transportation or non-personal automobile transportation can be viewed as beneficial for increasing market areas and high levels of disability can be seen as a factor that may decrease a location’s market area, a simple formula of adding up sales, automobiles, alternate transportation means and subtracting disability was created as a new weight for analysis. That being said, the formula that was created is certainly not one that is a purely accurate depiction of how taking into consideration aspects of the population itself will affect the retail market areas, it nevertheless gives us insight into how it will or may change.
Once these market areas weighted with mobility were created, an overlay with the sales weighted market areas was conducted. After obtaining the area percentage overlap, we compared it with 1 as in the previous examples regarding the first research question. The test to determine if the difference is significant was also conducted in the same way discussed earlier for the first research question. Figure 2 below gives an overview of the study area that we are dealing with. Downtown Atlanta was used for all three research questions, while Colquitt County was only used for our first research question (When defining market areas for retail locations using weighted Voronoi diagrams, is there a significant difference between the result from creating these Voronoi diagrams using Euclidean distance and that using network-based shortest-path distance?).
Figure 2: Overview map of the study areas
Results

The anticipated results of this study were that there will be a slight significant difference or no significant difference at all between defining retail market areas based on shortest path network distance compared to using Euclidean distance in the downtown Atlanta area. This is important because it is required that we have some sort of hypothesis in order to correctly run the T test with the percentage overlap for our first research question. We have two sets of data that we are comparing with the T test, one containing percentage overlap and the other data set containing 1. Our null hypothesis is 0 for the two-tailed T test because the expectation was that there would be a perfect overlap between the two approaches described in our first research question. Therefore, the null hypothesis is 0, meaning no difference between these two sets of data. The reason that this hypothesis was developed is because there are simply not enough directed links and the store location density is too low to make a difference between the two approaches as we saw in Japanese cities with the study conducted by Okabe and his team. Also, the road network in the chosen study area is very complete which may translate into results of no significant difference. If we had gone ahead and used a limited network (aka, just major roads), there is a better chance that we would see a significant difference between this approach and Euclidean distance. This method was tested by looking at a more rural study area with a less complete road network, Colquitt County. It seems as though the relationship between store attributes such as sales and the surrounding demographic characteristics was quite apparent and obvious in this situation. The results that follow in this section demonstrate that median incomes
have a significant relationship to sales figures for the retail locations in all four of our grocery store categories. There are also many other relationships that were present in the data that can be found in the regression outputs section. Several other of our independent variables in this analysis showed a strong relationship to store sales such as owner occupied housing. As with the first research question, in order to conduct a t test for the third research question concerning the market areas weighted with population mobility, we had to establish a hypothesis. The hypothesis for this research question would be no difference between the two approaches as in the first research question, meaning a perfect overlap between sales weighted market areas and mobility weighted market areas. The reason that the hypothesis in this case was that there would be no difference between approaches is the thought that the sales weighted market areas in this study area would have contained a strong relationship to population mobility and therefore that much of the variation that one would expect when using the mobility weights had already taken place when we used the sales weighted market areas. Therefore, I think that some of the factors of population mobility that we are using are related to the sales of retail stores already so I was not confident that using these factors would change the market areas dramatically. Certainly, areas with high levels of mobility may see their market areas slightly increase relative to other areas with low levels of mobility. That being said, I anticipated that the market areas simply weighted with sales will anticipate this as I believe sales will have a relationship to population mobility. Overall, I didn’t think there would be a change in the market areas when factoring in population mobility or when creating them along shortest path network distance and so our hypothesis for both the first and the third research questions was that there will be no difference between approaches.
The six figures shown below (figures 3-8) demonstrate an overview of the study area as well as the location of our retail locations. As we can see, there is an obvious difference in the spatial distribution between the four different categories of grocery stores in this study. Also, there is a wide variety in terms of the number of locations in each grocery store category which may be causing some of the discrepancy in results discussed later. Figure 3 below shows the full extent of our downtown Atlanta study area. While figure 4 below shows our other study area, Colquitt County where you can see that the road network in this area is obviously less complete than the road network in downtown Atlanta.

Figure 3: Downtown Atlanta Roads
Figure 4: Colquitt County Roads

Figure 5: Independent Grocery Stores in Downtown Atlanta
Figure 6: Chain Grocery in Downtown Atlanta

Figure 7: Specialty and International Grocery Stores
The next four figures (figures 9-12) shown below demonstrate market areas for each retail location based on a weight of store sales. These market areas are using Euclidean distance in this case and have been compared to the sales weighted market areas created along the network.
Figure 9: Market Areas weighted with sales data for specialty and international grocery store locations using Euclidean distance.

Figure 10: Market Areas Weighted with Sales data for convenience store locations using Euclidean distance.
Figure 11: Market Areas Weighted with Sales data for independent grocery store locations using Euclidean distance

Figure 12: Market Areas Weighted with Sales data for chain grocery store locations using Euclidean distance
The following four figures (figures 13-16) are market areas that were created with no weight using the “create thiessen polygons” tool in ArcGIS. Therefore, the following market areas simply used distance alone to create market areas. These are again market areas created using Euclidean distance. This was done in order to determine if there was any difference in results between comparing the non-weighted Euclidean and network market areas and the sales weighted Euclidean and network market areas.

Figure 13: Unweighted Voronoi diagram for Chain grocery stores in Atlanta
Figure 14: Unweighted Voronoi diagram for specialty and international grocery stores in Atlanta

Figure 15: Unweighted Voronoi diagram for Convenience stores in Atlanta
Figures 17-20 show the market areas weighted with sales along shortest path network distance created using the SANET program. Each color on the links corresponds to their respective supply point (retail location) that the link was assigned to in the market area creation process. These results were compared with the market areas in figures 9-12 to determine if they were significantly different.
Figure 17: Convenience Stores Weighted Voronoi Diagram - Network

Figure 18: Chain Stores Weighted Voronoi Diagram - Network
Figure 19: Specialty and International Grocery
Weighted Voronoi Diagram - Network

Figure 20: Independent Grocery Stores
Weighted Voronoi Diagram - Network
Figures 21-24 show the convex hull output that was required to convert our polyline output from SANET into polygons in order to conduct the overlay analysis with our original sales weighted market areas. There was no effective way to accomplish this without having two sets of polygons. Figure 25 demonstrates this same process for the Colquitt County area where we conducted the same analysis in an area with a less complete network to look for and differences in results from our downtown Atlanta results.

Figure 21: Chain Grocery Stores Convex Hull
Figure 22: Convenience Store Convex Hull

Figure 23: Independent Grocery Stores Convex Hull
Figure 24: Specialty and International Grocery Stores Convex Hull

Figure 25: Colquitt County area Grocery Stores Convex Hull
The following four figures (figures 26-29) show the non-weighted market areas for each grocery store category along a network. These market areas with no weight were created using the network analyst extension in ArcGIS. As discussed earlier, this was done purely for comparison purposes and these results were compared with those in figures 13-16.

Figure 26: Independent Grocery Store Unweighted Voronoi Diagram - Network
Figure 27: Specialty and International Grocery Store Unweighted Voronoi Diagram - Network

Figure 28: Chain Grocery Stores Unweighted Voronoi Diagram - Network
When looking at the results of our first research question (When defining market areas for retail locations using weighted Voronoi diagrams, is there a significant difference between the result from creating these Voronoi diagrams using Euclidean distance and that using network-based shortest-path distance in the study area?), we have found that there does appear to be a significant difference in the results of the these two different approaches in most cases with a few exceptions resulting in no significant difference. Essentially, the results of the first research question are somewhat mixed with some of the tests resulting in a significant difference between the two approaches and others showing no significant difference between the two approaches. While it is clear that the two approaches are producing different results, when comparing the overall results across space with each other, the difference appears to be
significant in many cases but not all cases. There didn’t seem to be any difference when comparing weighted sales market areas and comparing non-weighted market areas; both seemed to have statistically significant differences as well as non-significant differences. It seems as though the results were just as dependent on the store category that we were looking at as they were the weighted or non-weighted aspects. This may have something to do with the wide differences in the number of locations between each retail category, which is likely causing some greater differences in the categories with the larger number of locations. For example, the chain grocery stores only have a total of 15 locations in our study area while the specialty and international grocery stores have a total of 58 locations.

The way this was compared was to take the Voronoi diagrams from each of the retail location categories and conduct an overlay analysis. An overlay analysis was conducted for the unweighted market areas by overlaying the unweighted network based polygons with the unweighted Euclidean based polygons and gathering the area that both of those polygons overlapped for each retail location. This area was then compared with the polygon area from the Euclidean non-weighted market areas to establish a percentage overlap for each retail location. Using a T Test in Excel, these values were compared with 1 (our expected value since we are hypothesizing in this case that there will be no difference between the approaches, aka 1 or 100% would represent a perfect overlap). The null hypothesis that has to put in for the T test was 0, meaning no difference between the two sets of data. This process was also completed for the weighted market areas for each retail category. However, in this process the weighted market areas that were created using the SANET program along a transportation network gave us an output of polylines, so initially an overlay with the weighted market areas polygons was not possible. The polylines needed to be converted into a polygon in order to conduct the overlay
analysis. In order to accomplish this, the Convex Hull tool was used in the ET GeoWizards toolbox (here called the “features to convex polygons” tool). This created a polygon for each retail point that was the smallest possible polygon that encompassed all of the nodes and links that had been assigned to that retail location. While this approach is perhaps not the best way to calculate overlap between the two market areas, it was the best approach that we could come up with. We did not look at each polygon (market area) individually, but rather used a t test to look at the difference in all of the outputs (overlapping areas) against the assumed value of 1. Comparing retail market areas individually would have likely given us a substantially different result. However, the purpose of this study is to look at the overall change in the retail environment in our study area. Therefore, looking at individual market areas is a process that was not undertaken.

Table 1: Results from T Test for 1st research question

<table>
<thead>
<tr>
<th>Category</th>
<th># of Locations</th>
<th>T Score</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Grocery Stores – Non-weighted</td>
<td>20</td>
<td>2.45</td>
<td>0.024</td>
</tr>
<tr>
<td>Independent Grocery Stores - Weighted</td>
<td>20</td>
<td>1.21</td>
<td>0.23</td>
</tr>
<tr>
<td>Convenience Stores – Non-weighted</td>
<td>27</td>
<td>1.56</td>
<td>0.129</td>
</tr>
<tr>
<td>Convenience Stores - Weighted</td>
<td>27</td>
<td>2.99</td>
<td>0.006</td>
</tr>
<tr>
<td>Chain Grocery Stores – Non-weighted</td>
<td>15</td>
<td>-0.133</td>
<td>0.895</td>
</tr>
<tr>
<td>Chain Grocery Stores - Weighted</td>
<td>15</td>
<td>1.29</td>
<td>0.217</td>
</tr>
<tr>
<td>Specialty and International Grocery Stores – Non-weighted</td>
<td>58</td>
<td>1.012</td>
<td>0.315</td>
</tr>
<tr>
<td>Specialty and International Grocery Stores - Weighted</td>
<td>58</td>
<td>5.43</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Colquitt County Grocery Stores – Non-weighted</td>
<td>25</td>
<td>4.082</td>
<td>0.0004</td>
</tr>
<tr>
<td>Colquitt County Grocery Stores - Weighted</td>
<td>25</td>
<td>2.726</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Bold P Value = Significant Difference
When conducting the overlay analysis as well as the t test for the independent grocery stores in our study area, we have a total of 20 locations. For our unweighted market areas, the T score for the unweighted market areas was 2.45 which falls outside of the critical value of 2.09. Also, the p value was 0.024 which is below 0.05 so using these indicators we can say that for unweighted market areas for independent grocery stores in downtown Atlanta, there is a significant difference between these market areas created along a shortest path network distance and those market areas created using Euclidean distance. Therefore, when looking at independent grocery stores in the downtown Atlanta area, we see a significant difference between the market areas created using Euclidean distance and those created using the network when we are not using any weights.

The results from the T test for weighted independent grocery stores is a T score of 1.21 and a P value of 0.23. Therefore, the two approaches were not significantly different in this case. Elaborating on my earlier point, this case has 20 locations, which is much less than the specialty and international grocery store locations. While the purpose of this study is not to compare the results and test it against the number of retail locations in the area, it would certainly make for an interesting future research possibility.

Concerning the results for the convenience stores, there was a significant difference between Euclidean based market areas and network based market areas for the sales weighted market areas. Yet, no significant difference between approaches was present when looking at the non-weighted market areas for convenience stores. As the table shows, there was no significant difference between Euclidean based and network based market areas for the chain grocery stores, either weighted or non-weighted.
The specialty and international grocery store category has a total number of 58 locations, which is much higher than any of our other grocery store categories. When looking at the market areas weighted with sales, there was a significant difference between the two types of market areas. As we can see from the table, not only is the difference between weighted Euclidean distance market areas and weighted network market areas significant in this case, but with a T score of 5.43 the difference is very significant, more so than many of the other analyses. This may have something to do with the number of locations being much larger than the other categories. This could be causing the difference between approaches to be greater than the other categories in general.

Once again, the original anticipation of this study was that a significant difference between the two approaches would not appear. Therefore, we decided to look at a more rural area in Georgia with a less complete road network such as Colquitt County, Georgia to see if there are any differences between results. In this case, the results demonstrate that there is a significant difference between the market areas created using Euclidean distance and market areas created using network distance. The significant results are apparent in both the weighted as well as the non-weighted market areas. There appears to be some sort of relationship between this and the fact that the network in this area is not nearly as complete as the one in downtown Atlanta. This makes sense as one may expect a bigger difference between the two approaches with a less complete network.

Overall, regarding our first research question, we can’t say too much because there is a relatively similar number of results showing a significant difference and those showing no significant difference. There also does not appear to be any relationship between the nonweighted tests and the weighted tests as both show a similar number being significant and not
significant. What does seem clear however is that there appears to be in general a substantial significant difference between the two approaches when dealing with a less than complete road network.

When looking at the results of the second research question (Is there any spatial relationship between store attributes and the surrounding socioeconomic characteristics?), there were several discoveries that appeared to us. One of those results is the relationship between median household income and the sales figures for our retail locations. We have conducted a spatial regression analysis in Geoda for all four sets of retail points and have used the same variables each time. The dependent variable is store sales and the independent variables are white population, black population, Asian population, Hispanic population, vacancies, owner occupied housing, renter occupied housing and median household income (Houseold_18). The results of the regression analysis for independent grocery stores in the study area are as follows:

Table 2: Regression Output for Independent Grocery Stores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-Statistic</th>
<th>Probability</th>
</tr>
</thead>
</table>

When looking at the results of the second research question (Is there any spatial relationship between store attributes and the surrounding socioeconomic characteristics?), there were several discoveries that appeared to us. One of those results is the relationship between median household income and the sales figures for our retail locations. We have conducted a spatial regression analysis in Geoda for all four sets of retail points and have used the same variables each time. The dependent variable is store sales and the independent variables are white population, black population, Asian population, Hispanic population, vacancies, owner occupied housing, renter occupied housing and median household income (Houseold_18). The results of the regression analysis for independent grocery stores in the study area are as follows:

Table 2: Regression Output for Independent Grocery Stores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-Statistic</th>
<th>Probability</th>
</tr>
</thead>
</table>
There are several things that we can discuss here relating to the regression output. The main values to focus on are the probability values, the coefficients and the adjusted R-squared value. While the t-statistics can tell us some things about the relationship between variables, the software itself compares the t-statistic with values in the student’s t distribution to determine the p-value, which is the number that we should really be looking at here. “The t statistic is the coefficient divided by its standard error. The standard error is an estimate of the standard deviation of the coefficient, the amount it varies across cases. It can be thought of as a measure
of the precision with which the regression coefficient is measured. If a coefficient is large compared to its standard error, then it is probably different from 0 (Princeton University, 2007).”

The interpretation of the T statistic tells us the same thing as the interpretation of the P value.

When looking at this regression output for independent grocery stores in the study area, we find that approximately 49% of the spatial variability that is present in the total sales for each of the retail locations can be explained by the spatial variability present in the white population, black population, Asian population, Hispanic population, number of vacancies, owner occupied housing, renter occupied housing and the median income of the residents. By examining the t-statistic as well as the probability, we can see which of our independent variables in this case are statistically significant at the commonly used 95% confidence level. “The 95% confidence interval for your coefficients shown by many regression packages gives you the same information. You can be 95% confident that the real, underlying value of the coefficient that you are estimating falls somewhere in that 95% confidence interval, so if the interval does not contain 0, your P value will be .05 or less(Princeton University, 2007).” The confidence level can be adjusted (typically to either 90% or 99%), but a 95% confidence interval is the most common and for the purpose of this research, there was no reason to alter that. Here we see that the independent variables of black population, Asian population and median household income are statistically significant. “In simple or multiple linear regression, the size of the coefficient for each independent variable gives you the size of the effect that variable is having on your dependent variable, and the sign on the coefficient (positive or negative) gives you the direction of the effect(Princeton University, 2007).” The coefficient figures presented here are also useful as they are able to tell us that for example, for every unit increase in the black population, we can expect a 0.53 decrease in sales is predicted when holding all other variables constant. Another
finding in this case is that for every unit increase in owner occupied households, we can expect a 1.3 increase in total sales when holding all other variables constant.

The regression analysis for specialty and international grocery store locations output can be found here:

Table 3: Regression Output for Specialty and International Grocery Stores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-Statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>1940.046</td>
<td>81.13942</td>
<td>23.91003</td>
<td>0.0000000</td>
</tr>
<tr>
<td>WHITE</td>
<td>0.2760683</td>
<td>0.2332991</td>
<td>1.183323</td>
<td>0.2383132</td>
</tr>
<tr>
<td>BLACK</td>
<td>-0.7442133</td>
<td>0.1233801</td>
<td>-6.031874</td>
<td>0.0000000</td>
</tr>
<tr>
<td>ASIAN</td>
<td>-2.387136</td>
<td>0.8357699</td>
<td>-2.856212</td>
<td>0.0048158</td>
</tr>
<tr>
<td>HISPANIC</td>
<td>0.3036744</td>
<td>0.8219039</td>
<td>0.3694768</td>
<td>0.7122288</td>
</tr>
<tr>
<td>VACANT</td>
<td>-1.588556</td>
<td>1.341933</td>
<td>-1.183781</td>
<td>0.2381325</td>
</tr>
<tr>
<td>OWNER_OCC</td>
<td>0.9748023</td>
<td>0.7802699</td>
<td>1.249314</td>
<td>0.2132476</td>
</tr>
<tr>
<td>RENTER_OCC</td>
<td>-0.7480911</td>
<td>0.3655552</td>
<td>-2.046452</td>
<td>0.0422341</td>
</tr>
<tr>
<td>HOUSEHO_18</td>
<td>-0.03689431</td>
<td>0.0051304</td>
<td>-7.191312</td>
<td>0.0000000</td>
</tr>
</tbody>
</table>

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   7.897816
TEST ON NORMALITY OF ERRORS
TEST       DF       VALUE       PROB
Jarque-Bera 2       28.19575  0.0000008
There are several items to note in this output for specialty and international grocery stores, such as the adjusted R-squared value of 65%. This means that 65% of the spatial variability that is present in the total sales for each of the retail locations can be explained by the spatial variability present in our independent variables. The independent variables of black population, Asian population, renter occupied housing and median household income are statistically significant in this case. The size of the coefficient for each of the independent variables here gives the size of the effect that that variable is having on the dependent variable (sales). The sign of the coefficient (positive or negative) gives the direction of that effect. In this example, for every unit increase in the Asian population, we can expect a 2.38 decrease in sales is predicted when holding all other variables constant. For every unit increase in vacancies in the area, we can expect a 1.58 decrease in total sales when holding all other variables constant. Also, for every unit increase in owner occupied housing in the area, we can expect a 0.974 increase in sales.
The results of the regression analysis for chain grocery store locations can be found here:

Table 4: Regression Output for Chain Grocery Stores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>3684.615</td>
<td>645.245</td>
<td>5.710412</td>
<td>0.0000001</td>
</tr>
<tr>
<td>WHITE</td>
<td>0.7344725</td>
<td>1.125839</td>
<td>0.6523782</td>
<td>0.5153376</td>
</tr>
<tr>
<td>BLACK</td>
<td>-1.361823</td>
<td>0.6206602</td>
<td>-2.194152</td>
<td>0.0300437</td>
</tr>
<tr>
<td>ASIAN</td>
<td>-5.178146</td>
<td>4.091046</td>
<td>-1.265727</td>
<td>0.2079276</td>
</tr>
<tr>
<td>HISPANIC</td>
<td>1.079384</td>
<td>3.970842</td>
<td>0.2718275</td>
<td>0.7861909</td>
</tr>
<tr>
<td>VACANT</td>
<td>-3.744361</td>
<td>6.597362</td>
<td>-0.5675542</td>
<td>0.5713414</td>
</tr>
<tr>
<td>OWNER_OCC</td>
<td>1.759534</td>
<td>3.597552</td>
<td>0.489092</td>
<td>0.6256220</td>
</tr>
<tr>
<td>RENTER_OCC</td>
<td>-1.43384</td>
<td>1.784477</td>
<td>-0.8035075</td>
<td>0.4231834</td>
</tr>
<tr>
<td>HOUSEHO_18</td>
<td>-0.07110456</td>
<td>0.02829531</td>
<td>-2.512945</td>
<td>0.0132248</td>
</tr>
</tbody>
</table>

REgression Diagnostics

MULTICOLLINEARITY CONDITION NUMBER 8.527882

TEST ON NORMALITY OF ERRORS

<table>
<thead>
<tr>
<th>Test</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarque-Bera</td>
<td>2</td>
<td>8694.265</td>
<td>0.0000000</td>
</tr>
</tbody>
</table>

DIAGNOSTICS FOR HETEROSEDASTICITY

RANDOM COEFFICIENTS

<table>
<thead>
<tr>
<th>Test</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breusch-Pagan test</td>
<td>8</td>
<td>70.06207</td>
<td>0.0000000</td>
</tr>
<tr>
<td>Koenker-Bassett test</td>
<td>8</td>
<td>3.54748</td>
<td>0.8954737</td>
</tr>
</tbody>
</table>

SPECIFICATION ROBUST TEST

<table>
<thead>
<tr>
<th>Test</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>44</td>
<td>20.46827</td>
<td>0.9990635</td>
</tr>
</tbody>
</table>

DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX: chainswht.GWT (row-standardized weights)

<table>
<thead>
<tr>
<th>Test</th>
<th>MI/DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moran's I (error)</td>
<td>0.18264</td>
<td>0.6513831</td>
<td>0.5147991</td>
</tr>
<tr>
<td>Lagrange Multiplier (lag)</td>
<td>1</td>
<td>0.0052838</td>
<td>0.9420531</td>
</tr>
<tr>
<td>Robust LM (lag)</td>
<td>1</td>
<td>0.6381091</td>
<td>0.4243964</td>
</tr>
</tbody>
</table>
The adjusted R-squared value for the chain grocery stores regression output is 12%. This is not a good R-squared value because it means that the vast majority of the spatial variability present in our dependent variable (sales) can not be explained by the variability present in the independent variables and is also a much lower value than what we find in our other categories. This could be for a variety of reasons including the low number of locations in this retail category. Therefore, the spatial regression process is likely looking at the demographic information further away from each retail location compared to some of the other categories. For the chain grocery stores, there were only two independent variables that were statistically significant: black population and median household income. For every unit increase in the white population, we can expect a 0.73 increase in sales is predicted when holding all other variables constant. Another finding in this case is that for every unit increase in vacancies in the area, we can expect a 3.74 decrease in total sales when holding all other variables constant.

The result from the regression analysis that was run for convenience stores is as follows:

Table 5: Regression Output for Convenience Stores

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Number of Observations</th>
<th>Number of Variables</th>
<th>Degrees of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>SALES</td>
<td>119</td>
<td>9</td>
<td>110</td>
</tr>
</tbody>
</table>

| R-squared            | 0.596051               | F-statistic        | 20.2889           |
| Adjusted R-squared   | 0.566673               | Prob(F-statistic)  | 1.51013e-018      |
| Sum squared residual | 3.18701e+007           | Log likelihood     | -912.488          |
The adjusted R-squared value was 57% for the convenience stores output, meaning that 57% of the spatial variability present in sales for convenience stores can be explained by the spatial variability of white population, black population, Asian population, Hispanic population, vacancies, owner occupied housing, renter occupied housing and median household income.
The independent variables of black population, Asian population, renter occupied, and median household income are statistically significant. For every unit increase in the white population, we can expect a 0.46 increase in sales is predicted when holding all other variables constant. Also, for every unit increase in owner occupied housing in the area, we can expect a 1.5 increase in total sales when holding all other variables constant.

When looking at the results of the third research question (How can we define retail market areas incorporating mobility aspects of the population such as levels of disability, levels of car ownership and the use of public transportation/alternate travel means?), we found that these factors are useful for investigation and demonstrated dramatic changes to the various grocery store market areas. Not surprisingly, incorporating these factors in the weight for the creation of market areas produced a significant difference in the same way that some of the market areas along a shortest path transportation network distance did. After creating the market areas with mobility aspects incorporated as defined in the methodology, another approach was also attempted. This involved instead of simply using the data that was contained by the original market areas, creating half mile buffers around each point and conducting a spatial join to gather the demographic information from these areas. This process produced no noticeable advantage compared to the approach discussed in the methodology so it was not pursued any further. The creation of the mobility weighted Voronoi diagrams and market areas produced results that showed similar differences to the original sales weighted market areas a more significant way than the network based market areas did. It was decided that the conducting of an overlay analysis and subsequent T test was suitable in this situation as well. So, in the same way that we determined if the results of the original retail market areas were significantly different from the
network based retail market areas, we have also compared the mobility weighted Voronoi diagrams to the sales weighted Voronoi diagrams in the same way.

We can see from our maps that incorporating mobility aspects into the weight and creating retail market areas results in a difference in retail location’s comparative influence across space. While the difference between the mobility influenced market areas and the original market areas is a significant difference as perhaps unexpected, it has demonstrated a difference in the same way that the creation of market areas along shortest path network distance did. It must also be noted that the creation of a mobility weight in this instance was not any sort of accurate depiction of how the population may or may not be able to reach a certain retail location. For instance, we had three factors that could be seen as positives and only one factor that could be seen as a negative (disability). In reality, there should be greater research done in this area to determine the factors that truly determine whether a population has the mobility necessary to reach retail locations. This is difficult to determine simply by looking at the factors chosen in this study. In order to get a more accurate result for how population mobility can affect market areas, survey data would likely need to be collected.

Table 6: Results from Mobility T Test

<table>
<thead>
<tr>
<th>Category</th>
<th># of Locations</th>
<th>T Score</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convenience Stores</td>
<td>27</td>
<td>3.987</td>
<td><strong>0.0006</strong></td>
</tr>
<tr>
<td>Specialty and International Grocery Stores</td>
<td>58</td>
<td>-11.872</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Independent Grocery Stores</td>
<td>20</td>
<td>-3.701</td>
<td><strong>0.0015</strong></td>
</tr>
<tr>
<td>Chain Grocery Stores</td>
<td>15</td>
<td>1.36</td>
<td>0.194</td>
</tr>
</tbody>
</table>

**Bold P Value = Significant Difference**
When conducting the overlay analysis as well as the t test for the convenience stores in the study area, we have a total of 27 locations. For our mobility weighted market areas, the T score for the mobility market areas was 3.987 which falls outside of the critical value of 2.059. Also, the p value was 0.0006 which is below 0.05 so using these indicators we can say that for mobility market areas for the convenience store locations in the downtown Atlanta area, there does appear to be a significant difference between these market areas created using a variety of population mobility factors compared to the market areas weighted with only store sales values.

For the specialty and international grocery store locations in the study area, the T score was -11.872 and the P value was well below 0.05. A significant difference between sales weighted and mobility weighted market areas was apparent. As we can see in this case, once again the difference between methods is highly significant in the case of specialty and international grocery stores. This may be because this category has by far the largest number of locations. With a T score of -3.701 and a P value of 0.0015, the results for the independent grocery stores were also significant. The only category in this part of the study that did not see a result of significant difference between market areas is the chain grocery stores. These also contain the least number of locations out of all the categories.

In conclusion, when looking at the difference between the original weighted market areas and the population mobility weighted market areas we see a significant difference between the two approaches in all but one case. The category that did not see a significant difference was that of chain grocery store locations, which also has the least number of point locations. This could be because the weights for this part of the study may not be a completely accurate depiction of population mobility in the study area. Nevertheless, these results show that the
incorporation of population characteristics such as mobility alter the sales weighted market areas in a significant way and is a factor that should be included in market area analyses.

Figures 30-33 below show the market areas created using mobility as a weight for each of the four retail categories in this study. These were compared with the original sales weighted market areas based on the business ID to test for any difference.

Figure 30: Weighted Market Areas for Chain Grocery Stores with Factors of Mobility Included
Figure 31: Weighted Market Areas for Convenience Stores with Factors of Mobility Included

Figure 32: Weighted Market Areas for Independent Grocery Stores with Factors of Mobility Included
Figure 33: Weighted Market Areas for Specialty and International Grocery Stores with Factors of Mobility Included
Discussion and Conclusion

This study demonstrated the creation of weighted market areas for grocery stores in the downtown Atlanta, Georgia area. By using shortest path network distance in creating some of our market areas, we were able to compare the results of both approaches. The regression analysis that was conducted gave us insight into the relationship between sales of a grocery store location and the surrounding socioeconomic variables. The creation of a new weight for the creation of market areas that factored in several attributes to give us insight into the mobility of the population was conducted.

The goal of this study was to develop a methodology and conduct a case study for retail analysis and the creation of retail market areas more accurately than current approaches. This goal was achieved as results concerning the creation of market areas using network distance and incorporating mobility aspects of the population were achieved. The first of our research questions (When defining market areas for retail locations using weighted Voronoi diagrams, is there a significant difference between the results from creating these Voronoi diagrams using Euclidean distance and that using network-based shortest-path distance in the study area?) produced mixed results as the some of the tests resulted in a significant difference between approaches while others did not. When conducting this analysis for sales weighted market areas, it was found that there was a significant difference between approaches for convenience stores and specialty and international grocery stores in our study area. There was also a significant
difference between approaches when looking at sales weighted market areas in Colquitt County. For the non-weighted market areas, there was a significant difference between approaches for independent grocery stores in downtown Atlanta as well as for grocery stores in Colquitt County. Colquitt County grocery stores are the only category that showed a significant difference between approaches for both sales weighted and non-weighted market areas. Therefore, it seems as though the areas with a less complete network are more likely to see a significant difference between Euclidean based market areas and network based market areas. This is an important aspect as future studies could be done to determine how complete a network needs to be to make the differences between approaches no longer significant. This study showed that there is a significant difference when using an incomplete road network, yet it would be interesting to see how complete the road network would have to be in Colquitt County in order for there to be no significant difference between approaches.

The third research question in this study (How can we define retail market areas incorporating mobility aspects of the population such as levels of disability, use of alternate transportation and levels of car ownership) sought to develop a methodology for the inclusion of population mobility aspects into the creation of retail market areas. The goal was to see how incorporating mobility factors would affect the market areas in our study area. The results being that there was a significant difference between sales weighted market areas and mobility weighted market areas for convenience stores, independent grocery stores and specialty and international grocery stores in the study area, with no significant difference present for chain grocery stores. More research should be done in this area, but this study showed that the inclusion of population mobility produces a significantly different market area compared to just using a trade-off between store attributes and distance.
Overall, the results of this study show that using shortest path network distance for retail market area analysis is more necessary in areas with a less than complete road network. While there were significant differences between approaches in our downtown Atlanta study area, the differences were more apparent when looking at Colquitt County with a less complete road network. For most analyses however, it appears it is the most necessary in situations with either a large number of retail locations or a less than complete road network. In situations with a low number of locations and a complete network, we were less likely to see any significant differences between these two approaches. Also, the regression analysis proved a strong relationship between store sales and some of the surrounding demographic attributes. For example, there is a strong relationship between store sales and the surrounding median household income figures. In fact, for all four of our grocery store categories in this study (chain grocery stores, independent grocery stores, convenience stores and specialty and international grocery stores), there was a significant relationship between sales for a retail location and median household income. Black population, Asian population and renter occupied housing had a significant relationship to store sales in several cases as well. The inclusion of the mobility weights for the creation of market areas also produced a significantly different result in most cases. While our methods for this may not be perfect, it at least begins to look at the population mobility itself when creating market areas, rather than assuming that the population has the same mobility levels.
Scope of Study/Possibilities for Future Research

Although objective measures of retail stores, such as opening hours, floor space and parking space, are available, these indices do not necessarily measure store attractiveness very accurately. For example, the quality and price of goods or atmosphere of stores are occasionally crucial to understand store choices by consumers, though such information is difficult to obtain (Nakaya et al., 2007:350). The use of shortest path distances could also potentially be replaced with travel time, costs or other methods of determining distance. One could also model retail market areas throughout the day along a network by obtaining information on network flows at different times of the day etc. Future research should include animated maps showing market area changes based upon traffic, network flows, time of day, weekday vs. weekend etc. as well as many other factors including the possibility of multi-purpose trips made by consumers.

Several other research questions were proposed for this study, but deemed insufficient to undertake as part of this study: how can we apply realistic growth models to each retail store based on past sales growth to demonstrate how their trade area would grow relative to others given a certain percentage increase in sales figures? This part of the research was not conducted because of simplicity and less than ideal data available to use. All that would have needed to be done here was to get industry wide grocery store sales projections data and adjust the sales for each of our stores using the growth percentage. For example, if we obtained data saying that sales were projected to grow by 2%, we would simply adjust our sales value upward by 2% and
run the weighted Voronoi diagram function again. This was deemed too simple an undertaking to be worth it and to say anything substantial about the retail situation in our study area so it was not completed. Now, if data was available for each company instead of being industry wide data, this may produce more meaningful results. How does the detail of the network, aka a complete network versus a hierarchical network affect the results of the differences between Euclidean and network based weighted Voronoi diagrams? This research question was also deemed to not be sufficiently relevant to study as the dividing of the network into a non-complete network would have been a biased undertaking as there was no good way to determine what roads should be selected out of the overall data set. While it likely would have produced different results, it is practically unrealistic to assume that perhaps customers only use certain roads and do not utilize the entire network available to them. Although we did look at Colquitt County which has a less complete network than downtown Atlanta.

One of the main limits to accurately defining market areas based on demographic attributes and distance are that we are not using data of the people who actually visit each particular location. Consumer behavior cannot be accurately explained by aggregated data such as census data, and the information supplied by customers has become more important for retailers. As the use of a credit card is an information-giving exercise by the customer, retailers can obtain information about their customers, the goods and services that they purchase and how often they visit etc. An interesting study conducted by Hiroyuki Kohsaka of Nihon University in Japan was able to use GIS to map market areas for each store in a particular shopping center in suburban Tokyo by using this method of actual customer locations. This is accomplished by imputing the buying information obtained from point’s card into a GIS(Kohsaka, 1997). While this is obviously perhaps the most ideal and accurate method to describe retail trade areas, it
remains impractical in the United States despite the recent emergence of data collection by retailers here as well. While the collection of customer’s zip codes may be possible, the collection of their actual addresses may be more difficult. In some cases, this method and similar methods may be necessary to impose capacity constraints on the allocation (Liao and Guo, 2008). Along with increasing the accuracy of distance by using time, flow etc., and one would also want to consider the mobility of the population itself. A retail location’s market area may be quite small if the population surrounding the site has no personal means of transportation and no access to public transportation services. While we looked at this in this study, future work should be done that improves the accuracy of these estimates. Another possibility for future research is to look at changing urban land uses and relate this to the changing retail environment. Also, developing a methodology that allows for the creation of overlapping retail market areas would be beneficial and more accurately represent reality.

In reality, there is likely to be no strict line between market areas, but rather a changing and fluctuating boundary. Creating market areas using strength of probability for customers visiting a certain location is also a possibility for future research. For example, a customer next to the retail location may have strength of 90% chance that they will visit that location. Meanwhile, as you get further away and towards the boundaries of market areas, those customers may only have a 50% chance of visiting that location.

Of course, this research is not a completely accurate depiction of reality and it will not demonstrate exactly where each retail location’s customers are coming from. With the data available to us, it is simply too difficult to model consumer behavior and be entirely accurate. Without conducting surveys, we will never know why some consumers travel to a store that is not the one in their respective market area. Many factors are likely to be responsible here and
many of them may be individual circumstances, making it difficult to create a one hundred percent accurate representation of the real situation. Also, the process of using overlay analysis to determine the difference between methods is a difficult one. While the convex hull method seemed to produce accurate polygons from the polyline layer, there nevertheless has to be at least some discrepancy in total size when making a conversion such as this.

Since we simply extracted a North American Industry Classification code from the dataset (445110, grocery stores), we are not looking at retail locations where an increasing number of consumers obtain their food products. Stores such as Target and Wal-Mart are not included in this study because they contain a substantial amount of general merchandise and therefore are classified differently by the North American Industry Classification system. Yet, many people do their grocery shopping at stores like this so while we extracted all retail locations with the 445110 code from the dataset, this does not mean that we are looking at every location where the population will conduct their grocery shopping.

This study has shown that the use of the transportation network is important for retail geography studies, especially in areas with a less than complete road network, as that is how consumers reach their destination. The discovery of relationships between store sales and the surrounding demographic attributes can help to predict store sales changes into the future. The use of population mobility factors produced significantly different results than the original market areas and should be expanded upon to increase accuracy in future research.
References


Clarke, Graham. 1998. “Changing Methods of Location Planning for Retail Companies.”


