Accessing Grocery Stores Without Driving

Connor Flynn, Trevor Rosberg, Sandra Iradukunda, Aboubacar Daou

DTSC 4302- Data Science for Social Good B

4/18/2024

Introduction

Access to fresh and affordable food is a fundamental aspect of urban life. Yet, for many residents across the country, this essential need remains unmet due to inadequate infrastructure and transportation options. The USDA reports that 2.3 million people in the United States lack access to a grocery store within a mile of their homes and without access to a personal vehicle. While urban environments often prioritize access to grocery stores through driving, this perspective neglects the sizable portion of the population that rely on alternative modes of transportation such as walking or public transit.

This research examines the intricate challenges faced by pedestrians in accessing grocery stores without a car. These challenges are particularly prevalent in communities with lower median incomes and marginalized populations, where limited access to transportation infrastructure worsens food insecurity rates.

Our research focuses on addressing the insufficient access to fresh and affordable food within Mecklenburg County and Charlotte's urban areas, reflecting broader social equity concerns. This issue stems from Charlotte's disparities in infrastructure, including the lengths and accessibility of sidewalks and the accessibility of public transit systems. By examining the relationship of transportation infrastructure, neighborhood design, and socio-economic factors, we seek to understand how Charlotte can promote fair access to grocery stores for all residents, regardless of their socio-economic status or mode of transportation.

Charlotte, one of the largest and fastest-growing cities nationwide, underlines the urgency of addressing this challenge. Despite a rapid growth in population, disparities in grocery store access persists, affecting a substantial part of Charlotte's population. Our research seeks to identify areas of Charlotte where food access is at its lowest using ArcGIS and spatial analysis techniques.

ArcGIS is an online geographic information system software, used to create maps and analyze data, providing a visualization route for spatial data. We have utilized ArcGIS to create necessary maps for our research, adding layers to existing maps, and creating our first dataset using ArcGIS.

By integrating food access considerations into urban planning initiatives, we can foster healthier, more resilient communities. Improving food access not only enhances public health but also promotes sustainable transportation options, reduces the reliance on personal vehicles, and strengthens local economies. Our research is particularly relevant given Charlotte's commitment to fostering fair access to essential amenities, goods, and services by 2040, as outlined in the "Charlotte Future 2040 Comprehensive Plan."

Through our research, we aim to shed light on the complexities of food access in urban environments and offer actionable insights to inform policy interventions. This includes enhancing pedestrian infrastructure and public transportation accessibility in underserved communities, contributing to broader discussions of social justice and equity in Charlotte's urban environments.

Background

The accessibility of grocery stores in urban areas is a diverse issue that intersects with community health, socio-economic equity, and urban planning. In recent years, Bastian and Napieralski 2016 and Chisuta 2018. Scholars have increasingly turned their attention to understanding the spatial distribution of grocery stores and its implications for residents' access to healthy food options. This literature review focuses on synthesizing existing research findings on grocery store accessibility in Charlotte, North Carolina. Charlotte, like many other cities experiencing rapid population growth and socio-economic diversity, presents a compelling case study for investigating food disparities and the need for policy interventions to promote fair access to healthy foods.

Several studies have delved into the intricate relationship between the local food environment, neighborhood characteristics, and transportation infrastructure. For instance, Bastian and Napieralski 2016 conducted a comprehensive assessment of food availability and walkability in inner-ring suburbs of metropolitan Detroit using the Nutrition Environment Measures Survey for Stores (NEMS–S) and Geographic Information Systems (GIS) tools. Their findings revealed stark disparities in access to fresh produce, particularly within minority communities. Similarly, Chisuta 2018 underscored the significance of evaluating grocery store accessibility in urban areas using GIS tools while considering pedestrian and biking modes of transportation. Through the analysis of commute time and distance, Chisuta demonstrated how the spatial distribution of grocery stores significantly influences accessibility outcomes.

Charlotte, as one of the fastest-growing cities in the United States, presents a dynamic urban landscape where issues of food access are of paramount importance. Despite its economic prosperity, Charlotte grapples with food deserts—areas where residents have limited access to affordable and nutritious food. This phenomenon is often worsened by socio-economic factors, including income inequality and racial segregation. Moreover, the built environment and transportation infrastructure play pivotal roles in shaping residents' ability to access grocery stores and fresh produce.

While existing studies offer valuable insights into grocery store accessibility in urban areas, there are notable gaps in the literature, particularly concerning Charlotte, North Carolina. Firstly, there is a need for more localized research that specifically examines the spatial distribution of grocery stores and its relationship with socio-economic factors in Charlotte's diverse neighborhoods. Additionally, the intersectionality of race, income, and transportation access requires further exploration to understand how these factors collectively influence food access disparities. Furthermore, the role of policy interventions and community-based initiatives in addressing food disparities in Charlotte calls for deeper investigation.

In conclusion, the accessibility of grocery stores in urban areas such as Charlotte, North Carolina, is a complex issue influenced by multiple factors, including socio-economic disparities, transportation infrastructure, and urban planning policies. While existing research provides valuable insights, there is a clear need for more localized studies and comprehensive policy interventions to address food access disparities effectively. By understanding the unique challenges faced by diverse communities in accessing healthy foods, stakeholders can work towards building more inclusive and resilient food systems that promote equitable outcomes for all residents.

Hypotheses

Identified benefits of our research include its relevance to urban planning, potential for policy impact, and addressing social justice issues. We focus on addressing disparities in access to grocery stores and transportation infrastructure, critical concerns for urban planners and policymakers looking to create a more fair and sustainable

Charlotte. Findings from our research can aid policy interventions aimed at improving pedestrian infrastructure and public transportation accessibility in underserved communities, contributing to broader discussions of social justice and equity in Charlotte's urban environments.

Identified limitations include the complexity of our factors, limitations of data, and generalizability. The relationship between sidewalk availability, public transportation access, and grocery store proximity is influenced by multiple interconnected factors, including socioeconomic status, land usage patterns, and historical investment decisions. While GIS data offers valuable insights into spatial patterns of infrastructure and resource distribution, it may not capture the full range of factors shaping individuals' access to grocery stores. Additionally, the findings of our research will be focused on the Charlotte area and may not fully generalize to other urban contexts with different geographic, demographic, and socioeconomic factors.

Our research identifies and operationalizes three main variables: sidewalk availability, public transportation accessibility, and grocery store accessibility. Sidewalk availability refers to the presence or absence of sidewalks in Charlotte's urban areas, measured as the proportion of street segments with sidewalks within a specified radius of residential areas. Public transportation accessibility is defined by the availability and connectivity of the public transportation system in Charlotte, measured as the frequency of transit stops within a specific distance of residential areas. Grocery store accessibility refers to the proximity and availability of grocery stores to the public, measured as the distance from residential areas to the nearest grocery store and the density of grocery stores within a certain radius.

Based on our theory, we propose the following hypotheses for investigation in our research:

1. Greater availability of sidewalks along main streets in the Charlotte area will be positively associated with increased accessibility to grocery stores. We predict that areas with extensive sidewalk infrastructure will ease pedestrian travel, therefore enhancing residents' ability to reach nearby grocery stores without relying on public transportation or personal vehicles.
2. We predict that neighborhoods that have an all-inclusive public transit system, including bus routes and/or light rail networks, will prove greater access to grocery stores, reducing the impact of limited sidewalk infrastructure by offering alternative modes of transportation.
3. The combination of a complete sidewalk infrastructure and efficient public transportation networks will synergistically contribute to greater accessibility to grocery stores in the Charlotte area. Communities with abundant sidewalks and comprehensive public transit systems are expected to have the highest accessibility to grocery stores due to residents having multiple transportation options to reach essential local food retailers.

Methodology

Our dataset contains 249 entries corresponding to each census tract in Charlotte-Mecklenburg area. It has 113 columns representing various variables related to food access, transportation, and infrastructure from 2020 datasets, as well as demographic information from the 2010 census. The dataset contains multiple set of variables that capture various aspects of food access, including proximity to grocery store, public transportation availability pedestrian infrastructure, and socioeconomic factors such as education levels, poverty rates and income. The methodology employed combines spatial analysis with machine learning modeling to address the multifaceted issue of food access in urban environments.

To begin with implementing these methods, we began by obtaining layer and shape mappings of locations for grocery stores, farmers markets, bus stops, train stops, all sidewalks, all streets, and a census tract layer map for the Charlotte-Mecklenburg area. Given the small quantity of farmers markets in Mecklenburg County with just over 20 compared to the 300 grocery stores, we had decided to converge them together into one equal term from here on just as grocery stores. After obtaining these layer maps, they were combined into one using ArcGIS, so that each variable can cross-reference their respective census tract area boundaries. Once all were viewable together, they needed to be looked through and those that were lacking sidewalk data, everything outside of Mecklenburg, needed to be culled to preserve the homogeneity of the layer map data. There was also an outlier such as the airport tract which needed to be culled given the fact it’s in the Mecklenburg area but it’s 90% airport. From here, ArcGIS was used to calculate the census tracts’ areas and create radius buffers around the grocery stores, farmers’ markets, train stations, and bus stops. The grocery stores and trains have 1-mile radius buffers, and the bus stops have 0.5-mile radius. These buffer assessments have shown to be very beneficial by completing an accessibility study for walkers, increasing the potential of living where a person may not own a vehicle.

Once these new buffers were created, we were able to go ahead and begin to join together the sidewalk and buffers into a new layer map which allowed the ability to begin calculations. Along with these, we took the buffers and created a new map for when they were erased from the census tract maps to calculate the remaining area locations that do not have access to these resources. From here, we obtained the following values from these different layer maps. Before performing any ARCGIS work, we had started by having the total census tract area value in square meters, the total sidewalk feet in each census tract, the total street length in feet for each census tract, as well as the census tract numbers that each value was correlated to. After creating the buffers and manipulating the layer maps of the program we obtained the total census tract areas without grocery stores, bus stops, and train stations, as well as the sidewalk values within these grocery store, bus stop, and train station buffers. After extracting all of these values using ARCGIS and exporting them out of the program, we were able to derive percentages and ratios to all of these values in order to homogenize the census tracts given their size discrepancies.

From here we were able to begin gathering and attaching all of the census tract data as well as the food access to the geospatial data obtained before. Important demographic data such as average age, gender distribution, racial makeup, education level, poverty rate, and median income were included. Because the demographic's data was complex, some cells were relabeled to ensure that the data was standardized and straightforward.

A crucial problem that arose at this point that halted progress, was that we had all of the census data with census tract numbers and all our ARCGIS data with census tract numbers. However, even though these locations on the map were the same when we tried pairing the datasets together, we noticed 42% of the data did not align and it was classifying as “missing”. When looking into this we noticed that even though the maps lined up, what had happened was that the census tracts split into multiple new census tracts. For example, what was known as census tract 3 back in 2010, was now in 2020 referred to as 3.01 and 3.02. These two new census tracts we had zero data for and if combined encompassed the exact location that was once census tract 3, and so the concern was what do we do to go forward. The approach that we took was that we took the population values and demographics from these larger old census tracts and divide them amongst these new split census tracts. We recognize that in doing so we run the risk of creating errors inside of the data and it is not exactly accurate amongst the data from these two separate census tracts but given when this problem arose it was the quickest solution to proceed forward. In an ideal world the best approach would be to either gather all the exact same data all over again from the 2020 census, or to use a ratio of the areas of the new census tracts and divide amongst those ratios and multiply by the average population change. However, this second approach can only at best predict values for quantities of populations and not percentages of things such as education attainment, poverty rates, and race percentages. Regardless, we understand the risk that is done at this stage in the cleaning and gathering stage so late on and the potential bias that might ensue.

Analysis & Results

Data cleansing has been a critical stage in research to ensure data integrity and dependability. The resolution of challenging scenarios, such as duplicate regions or not covering all of the space in spatial data, was addressed, and only essential variables for analysis were included.

Our target variable is 'LILA', a binary classification variable where we are trying to predict if an area has high or low food access based on the predictor variables. The predictor variables are specified in the 'numeric_columns' list. This list contains ['Percentage_Grocery_Area', 'Percentage_Bus_Area', 'Percentage_Train_Area', 'Percent_SW_Bus_Area', 'Percent_SW_Train_Area', 'Percent_SW_Grocery_Area', 'Ratio_SW_To_Street', 'Low Vehicle Access', 'Total Population', '% Male', 'Median age', '18+%', 'White %', 'African American %', 'American Indian %', 'Asian %', 'Pacific Islander %', 'Other %', 'Hispanic %', 'No HS', 'HS', 'Associate or less', 'Bachelor or more', 'Total Poverty 25+ %', 'Median Earnings'].

Machine learning methods such as Classification and Regression Trees (CART), Random Forest, and Naive Bayes were chosen. Such models can handle curved and complicated interactions. These models are pertinent given the complexities of the urban environment and its impact on food availability. The Naive Bayes algorithm makes predictions based on the feature's independence or neutrality priors. On the one hand, a fundamental model serves as a basis for this complicated line of models. The models' efficacy was assessed using the cross-validation phase known as K-Fold Cross-Validation, and their generalizability was examined to ensure that the model did not overfit the unique dataset.

A baseline accuracy of 0% was established, indicating a potential for improvement. The accuracy of our baseline could present five problems: random guessing, data imbalance, poor feature representation, model complexity, and/or data quality. The CART model achieved an accuracy of 0.92, Which is higher than the baseline. The naive bayes achieved an accuracy of 0.72 on the training set and 0.78 on the test set. The random forest model had an accuracy of 0.92. K-Fold cross validation(k=10) was then employed to assess the generalization performance of CART, C5.0 and random forest models. The average cross validation scores across all folds were:

- CART: 0.87
- C5.0: 0.87
- Random forest: 0.9

We can conclude that the random forest model had the best generalization performance.

After testing out multiple machine learning models to predict areas in Charlotte that have low access to grocery stores, we conclude that the Random Forest model works the best out of the models used. Naive Bayes had the lowest score, with an accuracy of 0.78 on the test set. This makes sense because Naive Bayes assumes all the predictor variables are independent and variables like income levels and education are related.

Overall, the Random Forest model achieved our goal of identifying neighborhoods that lack easy access to grocery stores based on transportation options, income levels, sidewalk availability and other factors. The model can be useful for the city to figure out which areas need improvements like more public transit, better sidewalks, or for investors places to build new grocery stores.

However, there are some limitations to our research. The models used did not include how well, safe, and/or affordable the sidewalks and food selling points are, and other transportation methods such as bikes or by water.

Conclusion

This research has examined the complex issue of insufficient access to fresh and affordable food within Charlotte's urban areas, with a focus on identifying neighborhoods and communities that face limited access to grocery stores. By analyzing spatial patterns and assessing infrastructure disparities, we have identified key areas where targeted policy interventions are necessary to improve food access without relying on personal vehicles.

Our findings underscore the critical role of urban planning in addressing food access challenges and promoting social equity. Creating walkable neighborhoods with convenient access to grocery stores not only enhances food access but also fosters sustainable transportation options and improves overall urban livability. By integrating food access considerations into urban planning initiatives, we can work towards creating healthier, more resilient communities.

However, it is crucial to acknowledge the limitations and potential biases present in our work. While our integrated dataset provided a rich source of information, there may be additional nuances or factors that were not fully captured such as the quality of sidewalks, the overall safety of the area, the frequency and reliability of public transportation, or the affordability of grocery options.

Moving forward, it is essential to extend this research by exploring techniques for model interpretation and feature importance analysis. This could provide deeper insight into the relative contributions of different predictor variables and potential interactions among them. This can guide more targeted interventions by policymakers to increase food accessibility.

In conclusion, our research contributes to a growing body of literature on urban food access and highlights the importance of interdisciplinary approaches to addressing complex social challenges. By working together to prioritize fair access to fresh and affordable food for all residents, we can create more just and resilient cities for generations to come.

References

Bastian, E., & Napieralski, J. (2015, December 9). Suburban Food Security: Walkability and Nutritional Access in Metropolitan Detroit. *The Professional Geographer*, *68*(3), 462–474. https://doi.org/10.1080/00330124.2015.1099447

Bradley, T. (n.d.). *Open mapping*. Open Mapping - Data. https://maps.mecknc.gov/openmapping/data.html

GIS, M. C. (2024, January 31). *Farmers markets*. ArcGIS StoryMaps. https://storymaps.arcgis.com/stories/2499a44e1d9b4f65879eeacfb406024c

Go to the atlas. USDA ERS - Go to the Atlas. (2023, July 6). https://www.ers.usda.gov/data-products/food-access-research-atlas/go-to-the-atlas/

Hill, K. (2022, January 14). *Speaking up for Grocery Access - Monroe Road advocates*. MoRA. https://moraclt.org/speaking-up-for-grocery-access/

Kate Hosford, Jeneva Beairsto, Meghan Winters, Is the 15-minute city within reach? Evaluating walking and cycling accessibility to grocery stores in Vancouver, Transportation Research Interdisciplinary Perspectives, Volume 14, 2022, 100602, ISSN 2590-1982, https://doi.org/10.1016/j.trip.2022.100602.(https://www.sciencedirect.com/science/article/pii/S2590198222000641)

Lee-Gardner, A. (2022, June). *Measuring accessibility for pedestrians, bicyclists, and transit riders to grocery stores in the Excelsior/Outer Mission neighborhoods of San Francisco*. DigitalCommons@CalPoly. https://digitalcommons.calpoly.edu/theses/2577/

Largen, C. C. & I. (2021, August 5). *Charlotte's food environment*. ArcGIS StoryMaps. https://storymaps.arcgis.com/stories/4c6789d9316f415999024aec4d8a572a

Office of Management and Budget: Office of Management and Budget. Office of Management and Budget | Office of Management and Budget. (n.d.). https://budget.mecknc.gov/?utm_source=CF&utm_medium=DirectSP13&utm_campaign=Redirect

Plan Policy. Goal 1: 10-Minute Neighborhoods | Charlotte Future 2040 Comprehensive Plan. (n.d.). https://www.cltfuture2040plan.com/content/goal-1-10-minute-neighborhoods-0

United States Department of Agriculture. (2009, June). *Access to affordable and nutritious food: Measuring and ...* Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences. https://www.ers.usda.gov/webdocs/publications/42711/12716_ap036_1_.pdf