Analyzing the Impact of Economic and Demographic Factors on Grocery Store Availability in the U.S.

SI 671 Fall 2024

Trace Grain (tgrain), Lucy Wang (lucywang), Ashwin Kumar (ashwinkm)

Equal Contributions

Abstract

This study investigates the relationships between economic and demographic factors and the availability of grocery stores across the U.S. over time. Using a combination of datasets from the U.S. Census Bureau, we explore changes in grocery store density from 2019 to 2022 at multiple geographic levels. Our findings highlight regions at risk of food deserts, providing actionable insights for policymakers to address food insecurity and improve community well-being.

1. Introduction

Research Question

To understand the complexities of food deserts, it is important to examine how economic and demographic factors influence the availability of grocery stores in different regions of the United States. Our research seeks to answer the following question:

How do economic and demographic factors influence the availability of grocery stores in various regions of the U.S. over time, and what implications does this have for identifying areas at risk of food deserts?

This question forms the basis of our investigation into the socio-economic factors that influence food accessibility in urban and rural areas. By analyzing how income, population density, education, and employment impact the availability of grocery stores, we aim to identify the key forces driving the emergence of food deserts. The findings will highlight geographic and socio-economic patterns, providing insights for targeted interventions to improve food access in at-risk communities.

Understanding how food insecurity varies across regions and the role grocery store availability plays in this is crucial for identifying communities in need. This research will serve as a guide for future third parties to identify areas most affected and help in the development of policies and strategies to improve food access and reduce the impact of food deserts.

Significance

Food deserts are a major challenge, particularly where economic hardship and limited transportation overlap. These areas, with scarce access to affordable, nutritious food, disproportionately impact vulnerable populations such as low-income families and communities of color. The lack of healthy food options contributes to rising health issues, including obesity and diabetes, while also affecting overall well-being.

Understanding the economic and demographic factors influencing grocery store distribution is key to addressing these disparities. The following are key points of significance:

- **Health Impacts**: Limited access to healthy food increases chronic health conditions like obesity, diabetes, and heart disease, especially in low-income communities.
- **Economic Development**: Addressing food deserts can boost local economies by attracting grocers, creating jobs, and promoting entrepreneurship in underserved areas.
- **Social Equity**: Reducing food deserts ensures equal access to essential resources, improving quality of life and empowering communities by reducing food insecurity.
- **Policy Impact**: Understanding the factors behind food deserts can help design effective policies, such as subsidies or incentives for grocery stores in underserved areas.

This research is essential for improving public health, fostering social equity, and guiding policies that promote sustainable community development.

Motivation

In light of the current socio-economic climate in the United States, our team is compelled to explore the relationships between economic factors and demographic characteristics across various regions. The nation is facing a pivotal moment in its history, characterized by heightened concerns about the health of the economy, the growing inaccessibility of the American Dream for younger generations, and the rising costs across labor, food, and essential goods. As these pressures intensify, it becomes crucial to examine the far-reaching implications of these trends on the most basic of human needs—access to food.

In this context, the issue of food availability and affordability has become a central point of discussion. While political debates often focus on the rising costs of food, we believe that a more fundamental question deserves attention: how accessible are food sources across different regions, particularly in underserved or economically disadvantaged areas? We recognize that the affordability of food is only one piece of the puzzle. For many, the real challenge lies in the lack of convenient access to grocery stores, farmers' markets, and other essential food outlets, which exacerbates food insecurity, especially in rural or low-income urban areas.

Our research seeks to shed light on the groups most impacted by these disparities, with a focus on the intersection of economic and demographic factors that contribute to food deserts across the United States. This investigation goes beyond the immediate costs of groceries to address the broader, more systemic issue of food accessibility. We believe it is crucial for both policymakers and private-sector leaders to understand how these regional gaps in food availability affect different communities.

2. Data and Methodology

Datasets

1. **Grocery Store Data (2019 - 2022):** U.S. Census Bureau data detailing grocery store counts by county.
    a. Loaded from multiple CSV files for different years (e.g., `merged_data_2019_EMPSIZE_001.csv`).
    b. Important columns include `EMPSIZES`, `ESTAB`, `PAYANN`
    c. Processed to include state and county identifiers, and merged across multiple years.
2. **Population Data:** Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2022 from U.S. Census Bureau, Population Division.
    a. Loaded from a CSV file (`cc-est2023-alldata.csv`).
    b. Contains columns like `STATE`, `COUNTY`, `STNAME`, `CTYNAME`, `YEAR`, `AGEGRP`, `TOT_POP`, `TOT_MALE`, `TOT_FEMALE`, and various demographic breakdowns by race and gender.
    c. Processed to include age ranges and renamed columns for clarity.
3. **Geographic Data:** County and Partial Economic Area shapefiles from the U.S. Census.
    a. Loaded from a DBF file (`tl_2021_us_county.dbf`).
    b. Contains columns like `STATE`, `COUNTY`, `ALAND` (area in square meters).

Methods

1. **Time-Series Analysis:** Tracked changes in grocery store availability from 2019 to 2022.
2. **Correlation and Regression Analysis:** Assessed relationships between grocery store availability, gender, income, and demographic factors.
3. **Comparative Analysis:** Evaluated disparities across geographic boundaries to identify at-risk regions using metrics such as ANOVA.

4. **Network Analysis**: Calculate PageRank scores for U.S. counties based on directed graphs that incorporate the establishment density data to identify counties that are most influential in terms of specific economic metric.

3. Related Work

The phenomenon of food deserts has been extensively studied, particularly in relation to its socioeconomic and health implications. Karpyn et al. (2019) in The Changing Landscape of Food Deserts provided a foundational understanding of how food deserts emerge and evolve over time. The authors explored how geographic, economic, and policy factors influence the distribution of food retail establishments, emphasizing the role of urbanization and economic inequality in shaping food access. They highlighted that food deserts are not static but dynamic, with their prevalence and characteristics shifting due to changes in infrastructure, population demographics, and market trends.

Karpyn et al. also noted that rural and urban food deserts present distinct challenges. Rural areas often face issues of physical distance to grocery stores, while urban areas experience economic barriers despite geographic proximity. This study informs our analysis by underscoring the importance of both geographic and economic variables in understanding food access disparities.

Building on Karpyn et al. 's findings, our research expands the scope to investigate how specific economic indicators, such as number of employment and payroll income, along with demographic variables, affect grocery store availability. While their work focused on general patterns and policy implications, our study contributes new insights by examining temporal changes (2018–2021) and performing detailed geographic analyses across multiple levels, including counties, states, and Partial Economic Areas (PEA).

Additionally, Karpyn et al. emphasized the need for actionable data to guide interventions. Our work addresses this by using regression and risk assessment analyses to identify specific regions at risk of food deserts. By integrating their broader conclusions with a data-driven approach, our research complements and extends the understanding of food accessibility challenges in the United States.

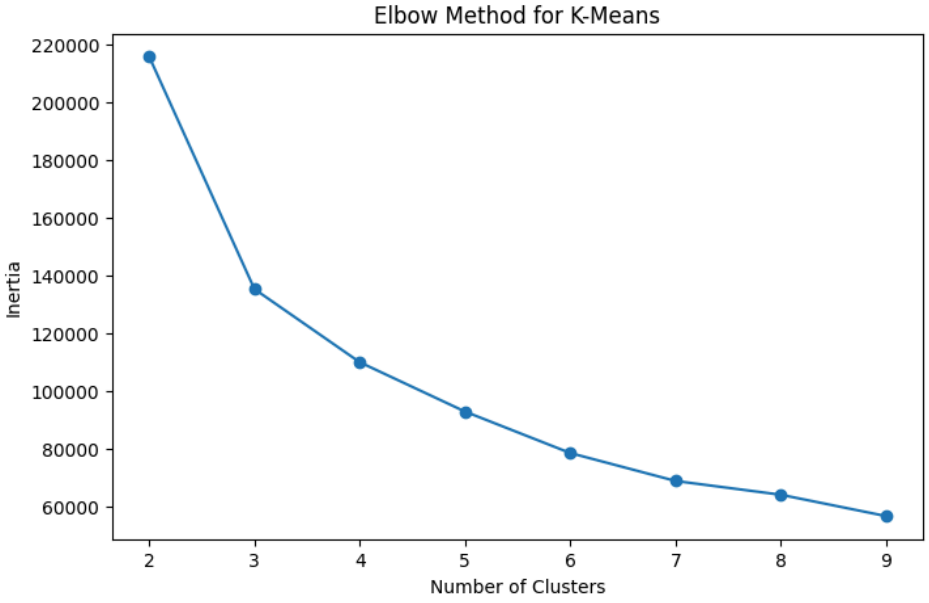
4. Analysis

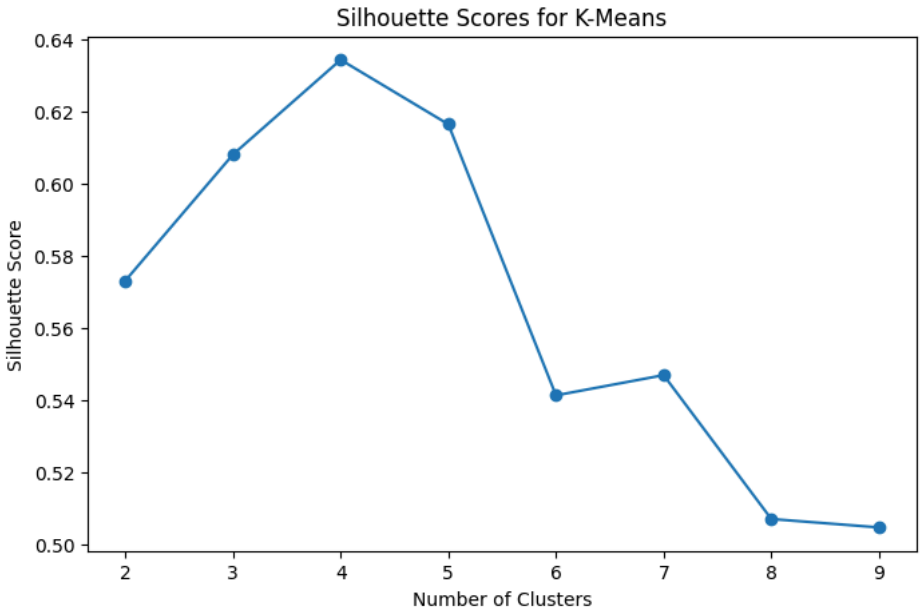
In this section, we present the results of our empirical analyses, examining county-level grocery store distributions, demographic compositions, and their evolving relationships between 2019 and 2022. We employed clustering, correlation assessments, network-based measures, and statistical tests to uncover patterns and differences that may inform our understanding of food deserts and the economic and demographic factors that influence them.

4.1 County-Level Similarities and Clustering

Our initial step was to characterize U.S. counties using a combination of demographic and economic variables. After selecting factors such as total population, gender distribution, demographic percentages (e.g., racial/ethnic breakdowns), and metrics of grocery store availability (e.g., establishments per area), we standardized these features and computed cosine similarities to gauge county-level resemblance. Using these similarity measures as input, K-means clustering was performed to identify groups of counties with comparable profiles.

Elbow and silhouette analyses suggested that a four-cluster solution provided a balanced representation of these groups. Counties within the same cluster tended to share analogous demographic patterns and comparable trajectories of grocery store availability over the study period. For example, one cluster might encompass more populous, demographically diverse counties with relatively stable or improving grocery store availability, whereas another might represent more rural or less diverse regions experiencing limited or declining grocery access. Although these cluster assignments are not absolute, they offer a meaningful lens to view the national landscape, revealing distinct "types" of counties and their varying risk levels for food deserts.



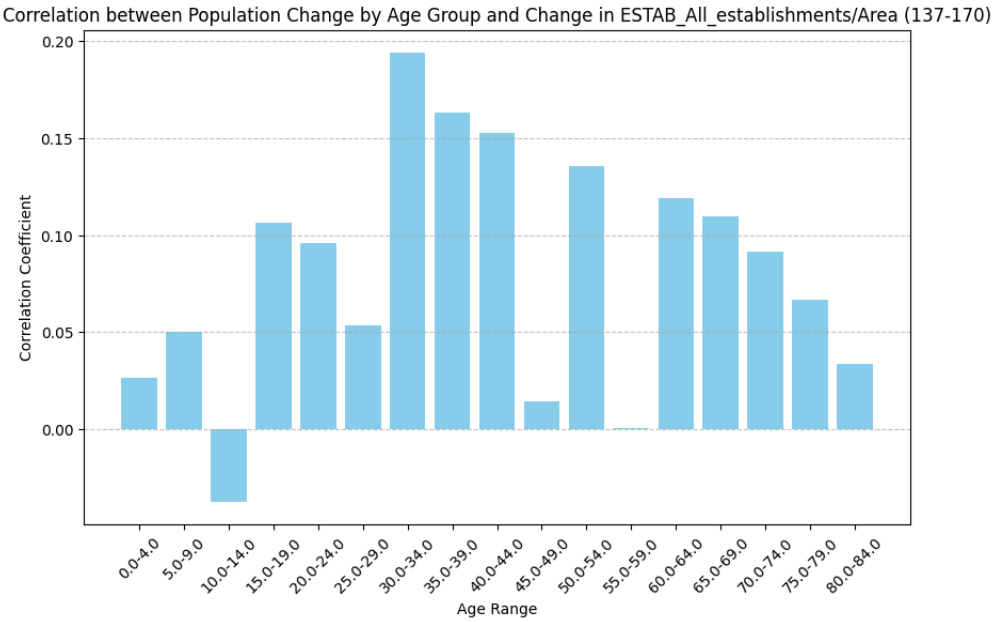


4.2 Correlating Demographic Changes with Grocery Store Availability

We further explored how shifts in demographic indicators correlated with changes in grocery store establishment density from 2019 to 2022.

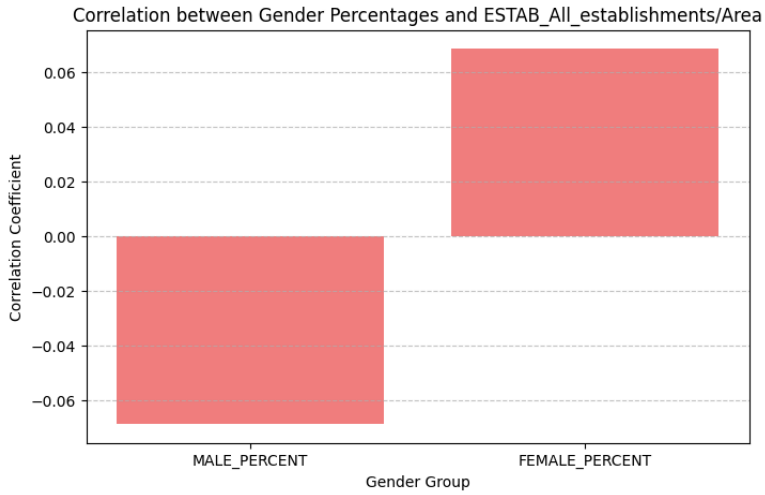
Age Dynamics:
Examining correlations by age cohort revealed that certain age groups were more closely associated with changes in grocery store availability. Specifically, the 30–34 age group exhibited a moderate positive correlation (approximately 0.2) with increases in establishment density,

suggesting that growth in this working-age demographic may attract more retailers or support existing ones. Other groups, such as 35–39 and 60–64, also displayed positive but weaker associations. By contrast, age cohorts like 10–14 and 40–44 demonstrated negligible relationships. These patterns suggest that the presence of economically active adult populations could be a factor influencing grocery store distribution decisions.



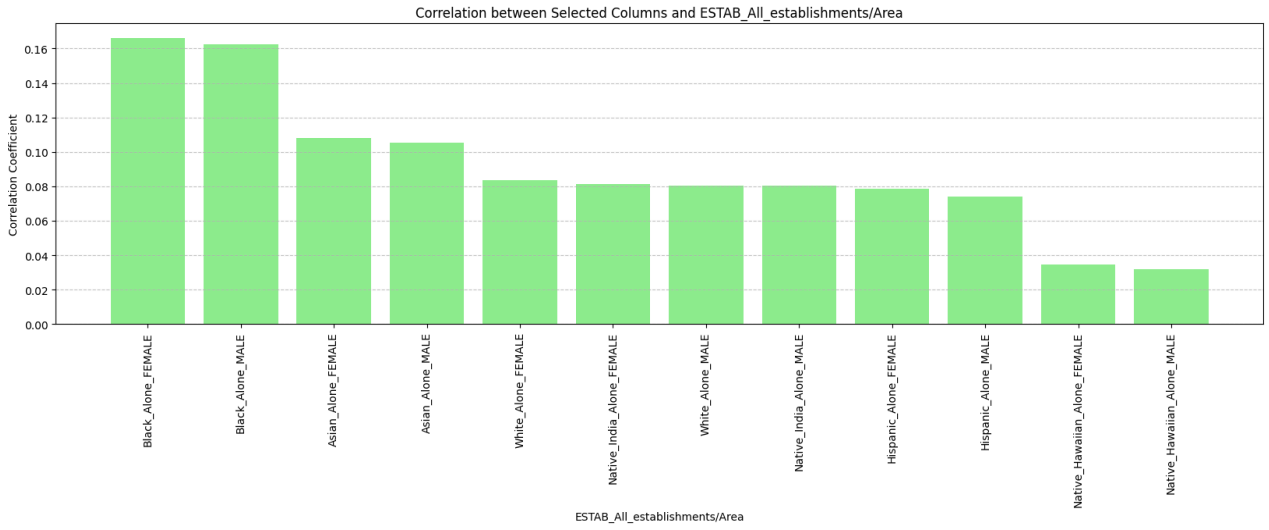
Gender Composition:

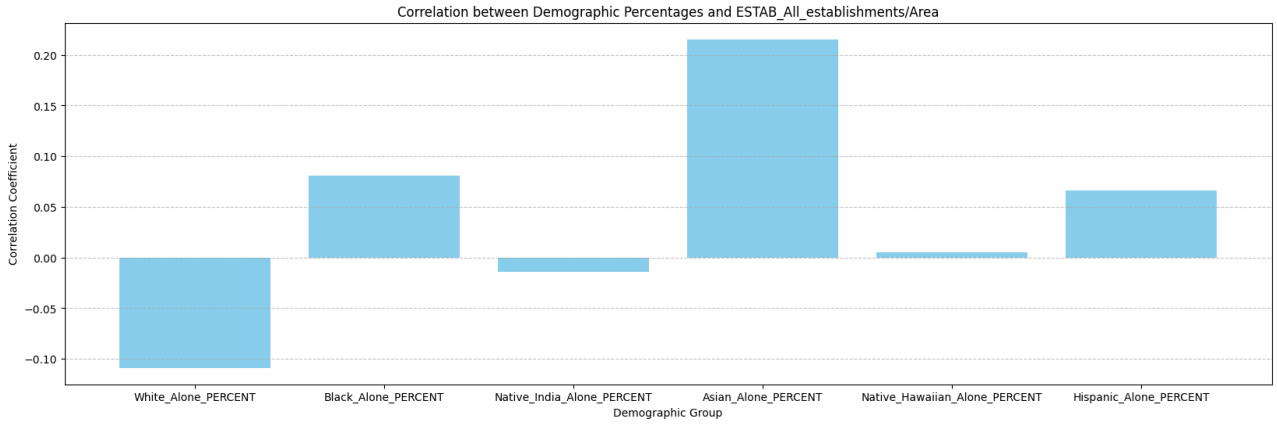
Counties with a higher proportion of females showed a modest positive correlation (0.07) with increased grocery store availability, while those with a higher proportion of males had a slightly negative correlation (-0.07). Although these effects are subtle, the results align with research suggesting that consumption patterns and household roles in shopping behavior may influence the local retail ecosystem. Still, the correlations are relatively weak, and gender differences alone do not appear to strongly shape grocery store density.



Racial and Ethnic Composition:

The racial and ethnic makeup of counties also exhibited distinct patterns. Counties with a higher percentage of the Asian Alone population had a markedly positive association (around +0.4) with grocery establishment density, indicating that these areas may be more attractive markets or have structural advantages that support retail proliferation. Hispanic Alone and Black Alone populations also correlated positively, though to a lesser degree, suggesting that demographic diversity may encourage a more robust grocery landscape. Conversely, areas with larger shares of White Alone or Native (American Indian/Alaska Native) Alone residents tended to have weaker or even negative correlations with grocery store density, hinting at long-standing disparities in investment or access. The near-zero correlation for Native Hawaiian Alone percentages suggests little direct relationship with grocery availability.





4.3 Network-Based Analysis of Geographic Connectivity

To capture the spatial interdependencies between counties, we constructed a geographic adjacency network, where each county is represented as a node and edges indicate direct geographic neighbors (i.e., counties sharing a boundary). This network structure allows us to explore the influence of spatial connectivity on economic indicators related to food access. In this approach, we not only considered which counties are connected but also integrated an economic measure - annual payroll per area (PAYANN/Area) in this example - to weight the edges.

Network Construction:
Using U.S. Census Bureau shapefiles, we identified counties and their neighboring counterparts by checking for shared borders. Each county was represented as a node, and directed edges were established from a county to its neighbors. To incorporate economic information, we assigned weights to these edges based on the neighbor's PAYANN/Area value. This transforms the adjacency network into a weighted directed graph, where high weights represent neighbors with stronger economic signals.

PageRank Application:
In this analysis, PageRank serves as a network-based metric that quantifies the relative importance of each county within the interconnected system. A county with many well-connected, high-value neighbors will receive a higher PageRank score, reflecting its position within a robust economic "neighborhood." By interpreting PageRank in this spatial-economic context, counties that appear consistently connected to neighbors exhibiting stronger economic characteristics (e.g., higher PAYANN/Area) can be considered more "central" or "influential" in the economic landscape.

Conceptually, if a county sits in a region where surrounding counties have higher economic vitality, that county's PageRank will increase. Conversely, counties that are isolated or surrounded by economically weaker neighbors will have lower scores. PageRank thus provides a

lens to identify counties that, by virtue of their location and connections, may be better situated to sustain or attract grocery stores and other critical services.

Top 10 Counties		
Rank	PageRank	County Name
1	0.005143	San Francisco County, CA
2	0.004005	Polk County, IA
3	0.003941	Douglas County, NE
4	0.003882	Davidson County, TN
5	0.003769	Cook County, IL
6	0.003741	Jefferson County, KY
7	0.00362	Sedgwick County, KS
8	0.003556	Tulsa County, OK
9	0.003556	Ramsey County, MN
10	0.003538	St. Louis city, MO

Bottom 10 Counties		
Rank	PageRank	County Name
1	0.000048	Henderson County, IL
2	0.000048	Scott County, IL
3	0.000048	Hardin County, IL
4	0.000048	Jefferson County, OK
5	0.000048	Blaine County, OK
6	0.000048	Collingsworth County, TX
7	0.000048	Johnson County, IL
8	0.000048	Perry County, AL
9	0.000048	Hale County, AL
10	0.000048	Rusk County, WI

Findings and Interpretation:

The PageRank results highlight a set of counties that stand out due to their economic connectivity. The top-ranking counties tend to be those embedded in clusters of high-payroll counties, suggesting that spatial context and economic synergy may play a role in reinforcing robust retail and employment ecosystems. On the other hand, counties with low PageRank scores are often at the periphery—either geographically isolated or surrounded by counties facing economic challenges. Such areas may struggle to attract and maintain grocery outlets, and thus may be at a higher risk of food insecurity.

By examining PageRank in conjunction with other demographic and economic analyses, policymakers and stakeholders can better understand how a county's position in a regional network influences its prospects for sustaining robust grocery access. This network-centric perspective can inform targeted interventions—such as improving transportation routes, encouraging collaborative economic development efforts between neighboring counties, or directing resource investments—to improve food access and reduce the incidence of food deserts.

4.4 Statistical Tests on At-Risk vs. Safe Counties

To further assess the disparities underlying food deserts, we classified counties into "at-risk" and "safe" categories based on their grocery store availability and related economic indicators. We then conducted inferential statistical tests to determine whether significant differences exist between these groups.

T-Test Analysis (At-Risk vs. Safe Counties):
Using a two-sample t-test, we compared the mean annual pay per area (PAYANN/Area) between at-risk and safe counties. The results were striking:

- T-Statistic: -60.55
- P-Value: 0.0 (effectively zero)

This highly significant result indicates that at-risk counties have significantly lower annual pay per area compared to safe counties. These findings lend credence to the notion that economic hardship is a key driver of grocery store scarcity, reinforcing the link between lower income and the heightened risk of residing in a food desert.

ANOVA Tests Across Multiple Geographic Units:
We conducted one-way ANOVA tests to compare differences in key metrics—such as establishments per area (ESTAB_All_establishments/Area), employees per area (EMP/Area), and annual pay per area (PAYANN/Area)—across multiple county groups. Each test returned a highly significant P-value, indicating that at least one group's mean differs substantially from the others:

- ESTAB_All_establishments/Area: $F = 1.73$, $P \approx 2.9e\text{-}14$
  Although the F-statistic is relatively modest, the extremely low P-value suggests statistically significant, if subtle, differences in establishment density among groups.
- EMP/Area: $F = 4.15$, $P \approx 3.1e\text{-}115$
  These results reveal stronger distinctions in employee densities across groups,

highlighting the uneven distribution of employment opportunities and labor markets that may influence or reflect retail development patterns.

- PAYANN/Area: F = 4.50, P ≈ 4.0e-132
  The largest F-statistic and an extremely small P-value indicate that differences in annual pay per area are the most pronounced. This metric appears highly sensitive to regional disparities, mirroring the t-test results and suggesting that economic vitality—or the lack thereof—is a crucial element in shaping the retail environment.

4.5 Summary of Analytical Findings

The analyses collectively underscore the multifaceted nature of grocery store availability and food desert formation. Clustering identified distinct types of counties, while correlation analyses pointed to demographic and socio-economic factors that influence or co-occur with grocery store presence. Spatial network considerations highlight the geographic context within which these patterns unfold. Most importantly, the inferential tests show stark economic differences between at-risk and safe counties, emphasizing that areas with lower income levels are disproportionately vulnerable to food access challenges.

While demographic factors—particularly certain age groups and racial/ethnic compositions—play a role, economic conditions are fundamentally intertwined with food access. This information can guide policymakers, local planners, and community organizations as they develop targeted interventions aimed at improving food access in economically strained regions and addressing the systemic inequalities that underlie food deserts.

5. Future Work

Future studies could expand upon our findings by examining the role of alternative food sources, such as farmers' markets and community-supported agriculture (CSA) programs, in mitigating the impact of food deserts. Additionally, longitudinal studies that incorporate more recent data beyond 2022 could capture the long-term effects of economic recovery or decline on grocery store availability. Exploring the impact of state-level policies, such as tax incentives for grocery retailers or urban planning initiatives, could provide deeper insights into effective interventions. Further research could also investigate how advancements in technology, like online grocery delivery services, influence food accessibility in both rural and urban areas. Finally, incorporating qualitative approaches, such as interviews with community members and local stakeholders, could offer a richer understanding of the lived experiences of those in food-insecure regions.

6. Conclusion/Summary

This study highlights the critical relationships between economic and demographic factors and the availability of grocery stores across the United States from 2019 to 2022. By employing a combination of clustering, regression, network-based analysis, and statistical tests, we identified significant regional disparities in grocery store access, illuminating the multifaceted nature of food deserts.

Our findings underscore that economic vitality, measured through metrics such as payroll and employment density, plays a pivotal role in determining grocery store availability. Counties with higher income levels and stronger economic connectivity exhibit greater access to grocery stores, while economically disadvantaged regions face elevated risks of becoming food deserts. Demographic factors, including age distribution and racial and ethnic composition, also influence food access, but their impact is secondary to economic variables.

The spatial network analysis revealed how geographic and economic interconnectivity shapes food access across counties, highlighting the importance of regional collaboration in addressing food insecurity. Furthermore, statistical tests confirmed stark economic disparities between at-risk and safe counties, emphasizing the need for targeted interventions in vulnerable areas.

These insights provide a valuable framework for policymakers and community leaders seeking to address food deserts. Potential solutions include incentivizing grocery store development in underserved areas, enhancing regional transportation networks, and fostering economic development initiatives to boost local purchasing power and attract retail investments.

By addressing the systemic factors that underlie food deserts, this research contributes to ongoing efforts to promote equitable food access, improve public health, and foster sustainable community development across the United States.

Project Video Link

Link:
https://drive.google.com/file/d/1-vdDLPLn9gcqZ6tA_QovS0-NKlngQ3d_/view?usp=drive_link

Citations:

Karpyn, Allison E et al. “The changing landscape of food deserts.” UNSCN nutrition vol. 44 (2019): 46-53.

Code:

https://drive.google.com/drive/folders/1doJJTI7qtpqLQqCT8_PQZRdeMMokSFa_?usp=sharing