Food Deserts and Obesity Rates Across U.S. Counties

Full Report

Anthony M Miller

Mesa Community College

CIS480 32778

Kazuaki Kashihara

11/21/2025

Capstone Selection

Introduction and Rationale

Food deserts are areas with limited access to affordable and nutritious food, particularly whole foods such as fruit, vegetables, and meats. When these options are not easily available, processed foods end up making up a majority of the human diet. This represents a critical public health challenge affecting millions of Americans. These geographic areas, usually characterized by the absence of full-service grocery stores within a reasonable distance, disproportionately impact low-income communities. The relationship between food access and health outcomes, particularly obesity rates, has emerged as a vital area of interest in public health research (CDC reports approximately a 42% adult obesity rate in US). The consequences of obesity extend beyond individual health outcomes, contributing to increased healthcare costs, chronic disease burden, and reduced quality of life. This project offers an opportunity to apply regression modeling techniques to examine complex socioeconomic and geographic relationships. This project requires application of regression analysis, data wrangling, visualizations, and statistical inferencing, and model interpretation skills.

This topic is meaningful to me as a student of data analytics and programming because it offers the opportunity to apply statistical methods to a socially relevant issue. By analyzing county level data on food deserts and obesity prevalence, I can leverage regression analysis to test whether geographic disparities in food access correspond to differences in health outcomes. Understanding the relationship between food access and obesity can inform policy decisions regarding urban planning, public health interventions, and resource allocation. Data-driven insights into this relationship can guide investments in community health, such as grocery store development or public nutrition programs.

Research Objective and Questions

The primary objective is to quantify and analyze the relationship between food desert prevalence and obesity rates across U.S. by county, providing evidence for the food access and health outcomes hypothesis. This study will attempt to answer the following research questions:

- Is there a statistically significant correlation between the presence of food deserts and obesity prevalence at the county level?
- How does the strength of this relationship vary when controlling for socioeconomic factors such as median household income, education levels, and demographic composition?
- Can food desert prevalence serve as a predictor of obesity rates when using regression modeling?
- Are there regional variations in the food desert obesity relationship across different geographic areas of the United States?

I will use regression analysis, where the dependent variable will be county-level obesity prevalence (continuous measure). The primary independent variable will be food desert prevalence. Additional control variables such as income, education, demographics will be incorporated in a multiple regression model.

Regression models:

- Model 1: Simple linear regression examining the two-variable relationship between food desert prevalence and obesity rates by county.
- Model 2: Multiple regression incorporating socioeconomic control variables (income, education, demographics).
- Model 3: Enhanced model including regional indicators to account for geographic variation.

Preliminary Data Sources and Scope

PolicyMap will serve as the primary dataset providing comprehensive curated data on demographics, health outcomes, and food access at multiple geographic levels. The geographic scope of this study will be all U.S. counties, ensuring a large and diverse sample that allows for comparisons across rural, suburban, and urban contexts.

Additional supplemental data sources will to be used as needed:

- USDA Food Access Research Atlas: Defines and measures food deserts nationwide.
- CDC PLACES/BRFSS: Provides county-level estimates of chronic disease risk factors, including obesity prevalence.
- U.S. Census Bureau: Supplies demographic and socioeconomic indicators for use as control variables.

These combined sources will provide a comprehensive dataset that captures both the independent (food desert prevalence) and dependent (obesity prevalence) variables, along with important correlated variables.

Feasibility and Next Steps

This project is feasible in terms of data access, timeline (based on course syllabus), and analytic scope. PolicyMap provides readily accessible, standardized data at the county level with consistent definitions and time periods. The platform facilitates efficient data extraction. Provided I set clear milestones, the project scope is appropriate for on time completion. The regression modeling approach aligns with standard available tools that have been used previously and currently in my degree program. Focusing on county level analysis provides sufficient complexity for meaningful insights while maintaining a manageable scope.

Next Steps Following Instructor Feedback:

- Data Collection: Retrieving relevant datasets from PolicyMap and USDA sources at the county level.
- Data Preparation: Clean and merge datasets, create derived variables as needed.
- Data Exploration: Conducting exploratory data analysis to identify trends, outliers, and distributional properties.
- Model Development: Building regression models to test hypotheses, first with a simple regression (food deserts predicting obesity rates), followed by multiple regression models incorporating socioeconomic covariates.
- Validate the models: Assess model assumptions and conduct residual analysis.
- Results Interpretation: Analyze and summarize findings.
- Report Findings: Prepare analysis report with methodology, findings, and visualizations.

Literature Review

Introduction

Obesity has emerged as a major public health issue in the United States. The obesity rates have continued to rise over the past several decades. According to the Centers for Disease Control and Prevention (CDC, 2020), over 42% of U.S. adults are classified as obese and obesity-related conditions such as heart disease, stroke, and diabetes remain leading causes of preventable death. Researchers have increasingly turned their attention to environmental and social determinants of health to explain regional variation in obesity prevalence. Among these is the concept of food deserts which has gained attention. Food deserts are areas where residents lack adequate access to affordable and healthy food.

According to the U.S. Department of Agriculture's Food Access Research Atlas (USDA ERS, 2021), millions of Americans live in low-income areas that are more than a mile away from supermarkets or large grocery stores. The USDA (Ver Ploeg et al., 2012) found that these conditions are particularly concentrated in both rural counties and underserved urban neighborhoods. Limited access to healthy food options in these regions has been linked to dietary patterns high in calories but low in nutritional value, which can contribute to obesity and related health issues.

The National Institutes of Health (Zenk, 2017) has also emphasized that food deserts are not evenly distributed but rather overlap with broader socioeconomic inequities. Communities facing higher poverty rates, lower educational attainment, and limited transportation access are disproportionately affected. PolicyMap (n.d.) provides geographic data illustrating how food deserts align with higher rates of obesity, as well as other health disparities such as diabetes and hypertension.

While the relationship between food deserts and obesity is generally recognized and accepted, questions remain about its strength once socioeconomic factors are considered. Some studies suggest that income, education, and access to transportation may mediate or even outweigh the effects of geographic food availability (Ghosh-Dastidar et al., 2014). This project is motivated by the need to clarify the relationship at the county level, using both food access measures and socioeconomic controls to test whether food deserts significantly predict obesity prevalence. Findings may help guide policies such as the Healthy Food Financing Initiative, which supports grocery store development in underserved areas.

Hypothesis Development

Based on the literature and research objectives, the following hypotheses are proposed:

Hypothesis 1

- (H_0) Null, two-tailed: There is no statistically significant relationship between food desert prevalence and county-level obesity rates.
- (H_1) Alternative, two-tailed: There is a statistically significant relationship between food desert prevalence and county-level obesity rates.
- Rationale: This hypothesis directly tests whether limited food access is associated with obesity, as suggested in prior research.

Hypothesis 2

- (H_0) Null, one-tailed: Higher food desert prevalence does not predict higher obesity rates after controlling for socioeconomic variables.
- (H_1) Alternative, one-tailed: Higher food desert prevalence predicts higher obesity rates after controlling for socioeconomic variables.
- Rationale: Prior studies suggest food deserts disproportionately impact low-income populations; testing this one-tailed hypothesis clarifies whether the direction of the relationship holds when accounting for poverty and education.

Hypothesis 3

- (H_0) Null, two-tailed: Socioeconomic factors such as poverty and education level have no moderating effect on the relationship between food desert prevalence and obesity rates.
- (H_1) Alternative, two-tailed: Socioeconomic factors such as poverty and education level moderate the relationship between food desert prevalence and obesity rates.
- Rationale: Literature indicates socioeconomic context is a potential confounder; testing this hypothesis examines whether the strength of the food desert and obesity relationship varies by economic and educational conditions.

Data Modeling Strategy

Data Sources

This project will use a combination of publicly available and institutionally supported datasets to analyze the relationship between food desert prevalence and obesity rates across U.S. counties. The primary data source will be PolicyMap, a comprehensive geographic data and mapping platform accessible through Mesa Community College Library. PolicyMap provides curated datasets that include indicators on food access, grocery store proximity, obesity prevalence, socioeconomic status, and demographics at various geographic levels, including counties, ZIP codes, and census tracts.

In addition to PolicyMap, the study will incorporate supplemental open-access datasets from:

- U.S. Department of Agriculture (USDA) Food Access Research Atlas: Provides official definitions and measures of food deserts, identifying low-income census tracts where a significant portion of the population lives far from supermarkets.
- Centers for Disease Control and Prevention (CDC) PLACES Project: Offers county-level estimates of chronic disease prevalence, including obesity, diabetes, and physical inactivity.
- U.S. Census Bureau's American Community Survey (ACS): Supplies demographic and socioeconomic data, such as poverty rate, educational attainment, racial composition, and vehicle access.
- Each source was selected for its reliability, open-access nature, and consistent use in peer-reviewed public health and socioeconomic research. These sources also use standardized geographic identifiers.

Justification for Source Selection

The selected data sources are appropriate because they are widely recognized, government backed datasets that ensure accuracy, and consistency. PolicyMap's integration of multiple indicators from federal datasets allows for efficient data retrieval and pre-cleaned variable standardization, reducing preprocessing overhead. The USDA's Food Access Research Atlas provides the official federal definition of "food deserts," while CDC's PLACES dataset represents the most up to date county level obesity data available. The ACS adds critical socioeconomic context, ensuring that the analysis accounts for variables such as income, education, and transportation access.

Geographic and Temporal Scope

The geographic unit of analysis will be the county level across the continental United States. County level analysis balances granularity with data availability. The temporal scope will include the most recent five-year range (2018–2023), which aligns with the availability of consistent PolicyMap and ACS data updates. Using this timeframe ensures that the analysis reflects contemporary socioeconomic conditions while minimizing inconsistencies caused by older data or shifts in classification criteria.

Data Types and Variables

The dependent variable in this analysis is obesity prevalence (%), sourced from the CDC PLACES dataset. The primary independent variable is food desert prevalence, measured as the proportion of the population in each county with limited access to supermarkets (USDA Food Access Research Atlas). Additional independent variables will be drawn from PolicyMap and the ACS to serve as control and moderating variables, including:

- Median household income
- Poverty rate
- Educational attainment (percentage with bachelor's degree or higher)
- Vehicle access (percentage of households without a vehicle)

Data Cleaning and Integration Plans

After collecting datasets from each source, the following data cleaning and integration steps will be performed:

1. Standardization of Geographic Identifiers: All datasets will be aligned using county FIPS codes to ensure one-to-one merging across sources.
2. Variable Formatting: Variable names and units will be normalized.
3. Handling Missing Data: Will be imputed using state averages.
4. Data Validation: Summary statistics will be reviewed to confirm logical ranges.

The final dataset will thus contain one observation per county, with each variable representing an annual or averaged indicator.

Modeling Strategy

Model Type

The primary analytical approach will be regression modeling, selected for its suitability in identifying relationships between continuous variables. Three related regression models will be constructed:

Model 1 – Simple Linear Regression: Establishes the baseline correlation between food desert prevalence (independent variable) and obesity rate (dependent variable).

Model 2 – Multiple Linear Regression: Introduces socioeconomic control variables (income, poverty, education, vehicle access) to test whether the relationship remains significant after adjustment.

Model 3 – Moderated Regression (Interaction Effects): Tests whether socioeconomic variables moderate the relationship between food desert prevalence and obesity rate.

Justification for Model Selection

Regression models are the most appropriate for this project because the dependent variable, obesity prevalence, is continuous and the goal is to quantify relationships between predictors rather than classify discrete outcomes. Simple and multiple linear regression models will reveal direct and adjusted associations, while moderated regression will assess whether certain socioeconomic factors amplify or weaken these associations. Based on my research, this approach aligns with prior public health studies examining community-level determinants of obesity and allows for straightforward interpretation through coefficients and significance tests.

Independent Variables and Features

The independent variables will include both primary predictors and socioeconomic controls:

Variable Type	**Variable Name**	**Source**	**Expected Relationship**
Primary Predictor	Food Desert Prevalence (%)	USDA / PolicyMap	Positive (higher deserts -> higher obesity)
Control	Median Household Income ($)	ACS	Negative (higher income -> lower obesity)
Control	Poverty Rate (%)	ACS	Positive
Control	Educational Attainment (%)	ACS	Negative
Control	Vehicle Access (%)	ACS	Negative

Moderator	Poverty × Food Desert Interaction	Derived	Expected to amplify effect

Power BI Dashboard Walkthrough

Introduction

This Power BI dashboard was developed to visualize the relationship between food-desert prevalence and obesity rates across U.S. counties. Its goal is to translate statistical findings into visuals that allow users to explore geographic and socioeconomic patterns. The dashboard supports three key research questions:

1. Is there a statistically significant relationship between county-level food-desert prevalence and obesity prevalence?
2. Does this relationship remain when controlling for socioeconomic factors such as income and education?
3. Do socioeconomic variables moderate or strengthen that relationship?

Each visualization connects directly to these questions, combining geographic, correlational, and statistical perspectives. Currently I am still working on most of the planned visualization, but I have outlined plans for them that are subject to change.

Geographic Distribution Visuals

Map of County-Level Obesity Rates (Figure 1)

This map supports Hypothesis 1 by illustrating spatial clustering of obesity prevalence across U.S. counties.

- H_0: No relationship exists between food-desert prevalence and obesity rates.
- H_1 (two-tailed): There is a statistically significant relationship between food-desert prevalence and obesity rates.

Chart Type and Rationale

A filled map (choropleth) presents obesity categories: Low (< 25%), Moderate (25–30%), High (30–35%), Very High (35–40%), and Critical (> 40%) by county (see Figure 1). The color gradient from light to dark blue conveys intensity while preserving accessibility and contrast. Choropleths are ideal for showing spatial trends at the county level. I came up with the categories which I may need to adjust as the numbers are quite a bit higher than I expected, making most of the chart 'critical'.

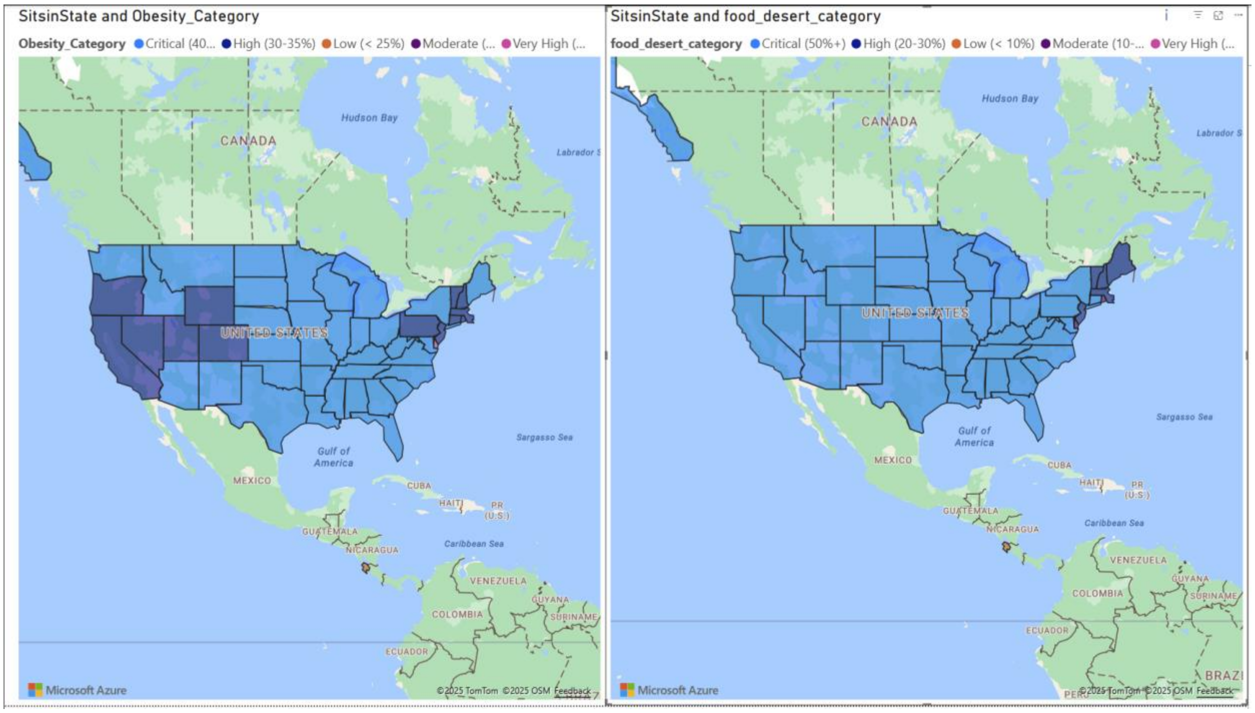
Data Structure

- Location: County (FIPS 'GeoID')
- Color Value: Obesity Category
- Tooltip: County name, State, Obesity Rate (%), Poverty Rate (%)

User Experience

Users can hover to reveal metrics and use a state slicer to zoom into regions such as the Southeast, where higher rates appear most frequent.

Figure 1. State level Obesity Categories & Map of Food-Desert Prevalence:



Map of Food-Desert Prevalence (Figure 1)

This visualization represents the independent variable, food-desert prevalence, supporting Hypothesis 1.

Chart Type and Rationale

The choropleth map on the right display's food-desert categories from Low (< 10%) to Critical (> 50%) (see Figure 1). By using identical color scales and boundaries, users can visually compare patterns between obesity and food access.

Data Structure

- Location: County (FIPS)
- Color Value: Food-Desert Category
- Tooltip: County name, State, % Population in Food Desert

User Experience

Selecting a state filters both maps simultaneously, enabling a side-by-side comparison of food access and obesity rates.

Correlational Analysis Visuals

Scatter Plot of Food-Desert % vs. Obesity % (Figure 2)

This visualization directly addresses Research Question 1: "Does a higher percentage of the population living in food deserts correspond with higher obesity rates?"

Chart Type and Rationale

A bubble scatter plot plots food-desert rate on the X-axis and obesity rate on the Y-axis (see Figure 2). Each bubble represents one county; bubble size reflects total population and color encodes state. This design reveals the density and variability of counties and highlights clusters that deviate from the general trend.

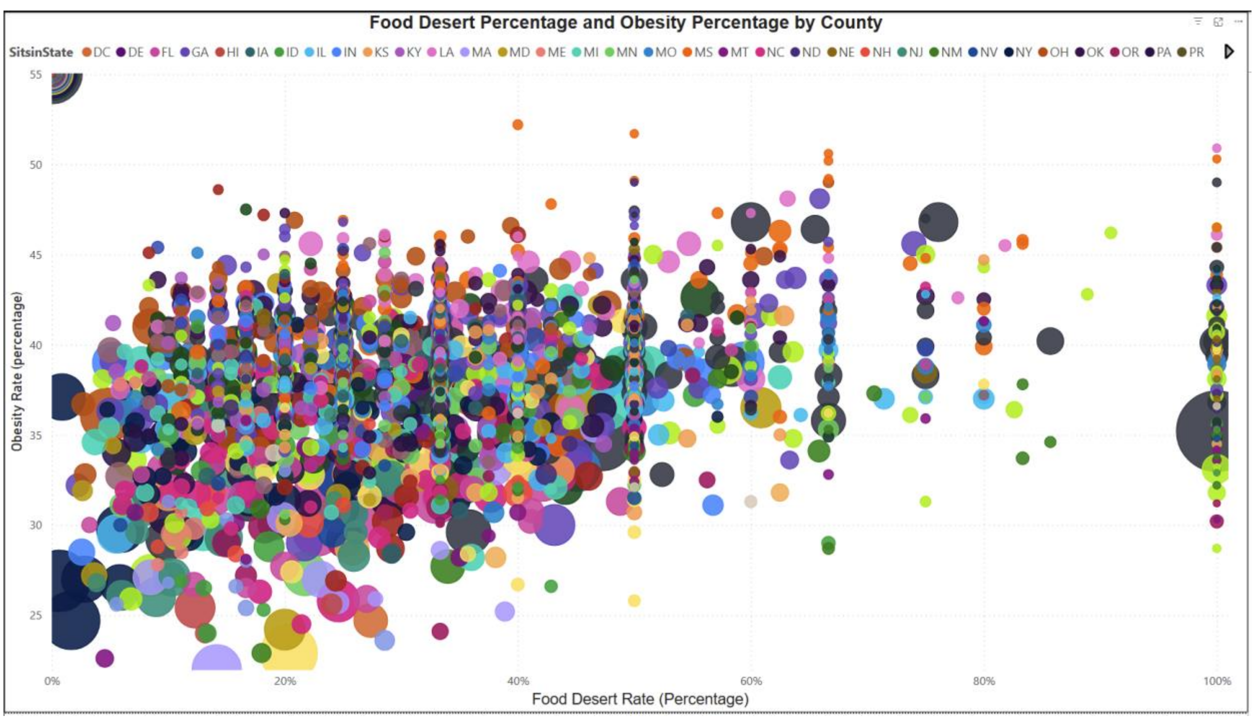
Data Structure

- X-Axis: Food-Desert Rate (%)
- Y-Axis: Obesity Rate (%)
- Legend: State (“SitsinState”)
- Size: County Population
- Tooltip: County, State, Food-Desert %, Obesity %, Median Income

Interactivity

Cross-highlighting allows users to click a state in the legend and immediately filter the map visuals. A trend line and displayed R^2 value (to be added) will quantify correlation strength.

Figure 2. Scatter Plot of Food-Desert % and Obesity %:



Planned Additions

Clustered Bar Chart – Average Obesity Rate by Income Quintile

Addresses Hypothesis 2: socioeconomic factors mediate the food-desert/obesity relationship.

Chart Type and Rationale

A clustered bar chart will show mean obesity rate across five county income quintiles. Bars will be colored by food-desert severity, revealing whether low-income, high-desert counties have elevated obesity rates.

Data Structure

- X-Axis: Income Quintile
- Y-Axis: Mean Obesity Rate (%)
- Legend: Food-Desert Category

Interactivity

Slicers will enable filtering by region or education level, allowing users to examine socioeconomic subgroups.

Dual-Axis Line Chart – Trends 2018–2023

Explores whether counties with high food-desert prevalence experience steeper obesity increases over time.

Chart Type and Rationale

A dual-axis line chart will plot obesity rate (%) and food-desert % by year, helping visualize synchronized trends.

Data Structure

- X-Axis: Year (2018–2023)
- Y1: Obesity Rate (%)
- Y2: Food-Desert Population (%)

Interactivity

Users can choose a state or group of counties from a drop-down slicer to see localized temporal changes.

Regression Summary – KPI Cards and Decomposition Tree

Supports Hypothesis 3: socioeconomic variables moderate the food-desert/obesity link.

Chart Type and Rationale

- KPI Cards will display R^2, Adjusted R^2, and p-values from regression models.
- A Decomposition Tree will allow users to drill down to see which predictors—poverty, education, vehicle access contribute most to obesity rate variance.

Data Structure

- Root: Obesity Rate
- Branches: Food-Desert %, Income, Education, Poverty, Vehicle Access

Interactivity

The tree enables exploratory analysis by dynamically showing factor importance, giving non-technical users a visual explanation of model outcomes.

Dashboard Design and User Experience

The dashboard layout follows a top-down narrative:

1. Context: Geographic maps introduce the overall distribution of obesity and food-desert rates.
2. Correlation: The scatter plot demonstrates how those variables relate numerically.
3. Explanation: Planned charts and regression summaries explore why these patterns exist.

Conclusion

The current dashboard effectively establishes the foundation for exploring the project's hypotheses through clear geographic and correlational visuals. Planned additions, such as socioeconomic bar charts and regression-based KPI summaries, will further strengthen its depth. When complete, the Power BI dashboard will serve as an interactive, evidence-driven tool linking food access disparities to public-health outcomes, offering both a comprehensive national view and the flexibility for localized policy exploration.

Summary of Hypotheses and Results

Hypothesis 1: Bivariate Relationship

Hypothesis Statement: There is a statistically significant correlation between food desert prevalence and county-level obesity rates.

Statistical Test: Simple Linear Regression (Model 1)

Result: NOT REJECTED

Evidence:

- Food desert prevalence shows a statistically significant positive correlation with obesity rates ($r = +0.1417$, $p < 0.001$). Model 1 regression results confirm this relationship:

Statistic	Value
Food Desert Coefficient (β)	+0.0268
Standard Error	0.003
t-statistic	8.007
p-value	<0.001
R-Squared (R^2)	0.020

- Interpretation: Each 1% increase in food desert prevalence is associated with a 0.027% increase in obesity rate. This relationship is highly statistically significant ($p < 0.001$), indicating strong evidence against the null hypothesis of no relationship.

- Conclusion: Hypothesis 1 is SUPPORTED. A statistically significant bivariate relationship exists between food deserts and obesity. However, the low R^2 value (0.020) indicates that food deserts alone explain only 2% of obesity variation, suggesting that other factors are critical.

Hypothesis 2: Controlled Relationship with Socioeconomic Factors

Hypothesis Statement: Higher food desert prevalence predicts higher obesity rates after controlling for socioeconomic variables (income, poverty, education).

Statistical Test: Multiple Linear Regression (Model 2)

Result: REJECTED

Evidence:

- When socioeconomic controls are added to the model, the food desert effect becomes statistically non-significant and changes direction:

Variable	Model 1	Model 2	Interpretation
Food Desert Coefficient	+0.0268	-0.0029 (ns)	110% reduction; non-significant

- Model 2 Results - Controlling Variables:

Variable	Coefficient	p-value	Significance
Food Desert Prevalence	-0.0029	0.261	NOT SIGNIFICANT
Poverty Rate (%)	+0.1633	<0.001	HIGHLY SIGNIFICANT
Education (Bachelor's+)	-0.2632	<0.001	HIGHLY SIGNIFICANT
Median Income ($)	-5.170	0.416	NOT SIGNIFICANT

- The model fit improves dramatically with controls: R^2 increases from 0.020 (Model 1) to 0.504 (Model 2), indicating that socioeconomic factors explain 50% of obesity variation.
- Critical Finding - Confounding Effect: The bivariate food desert-obesity relationship observed in Model 1 is SPURIOUS and driven by confounding. Counties with more food deserts tend to be poorer and less educated, and it is THESE socioeconomic factors—not food access itself—that drive obesity differences.
- Conclusion: Hypothesis 2 is NOT SUPPORTED in its original form. The food desert effect does NOT persist after controlling for socioeconomic factors ($p = 0.261$). This is a critical finding that challenges the assumption that food deserts are an independent cause of obesity. Instead, food deserts are symptomatic of broader socioeconomic disadvantage.

Hypothesis 3: Moderation by Socioeconomic Factors

Hypothesis Statement: Socioeconomic factors (particularly poverty) moderate the relationship between food desert prevalence and obesity rates. The strength of the food desert-obesity association varies depending on poverty levels.

Statistical Test: Moderated Regression with Interaction Terms (Model 3)

Result: NOT REJECTED

Evidence:

- Model 3 includes an interaction term (Food Desert × Poverty) to test whether poverty modulates the food desert effect. Results are:

Variable	Coefficient	p-value	Interpretation
Food Desert (main effect)	-0.1325	0.042	Significant at low poverty levels
Food Desert × Poverty (interaction)	+0.2079	<0.001	HIGHLY SIGNIFICANT

- The interaction coefficient is positive and highly significant ($\beta = +0.2079$, $p < 0.001$), indicating that poverty amplifies the food desert-obesity relationship.
- Practical Meaning: In LOW-POVERTY counties, food deserts have minimal or even negative association with obesity (better food access does not substantially improve outcomes if other health determinants are poor). In HIGH-POVERTY counties, food deserts show a stronger positive association with obesity, suggesting that in

economically disadvantaged areas, geographic barriers to food access compound other health risks.

- Model Improvement: R^2 improves from 0.504 (Model 2) to 0.507 (Model 3), a modest but statistically significant improvement that demonstrates the value of examining conditional effects.
- Conclusion: Hypothesis 3 is SUPPORTED. Poverty significantly moderates the food desert-obesity relationship. This is a novel finding that demonstrates the importance of context in understanding geographic health disparities. Food desert interventions will likely be most effective in economically disadvantaged communities.

Summary Table: All Hypotheses

Hypothesis	**Test**	**Result**	**p-value / Evidence**
H1: Bivariate relationship	Model 1	Supported	$r = +0.14$, $p < 0.001$
H2: Effect persists with controls	Model 2	Rejected	$p = 0.261$ (non-sig)
H3: Poverty moderates' effect	Model 3	Supported	interaction $p < 0.001$

Project Summary and Reflection

Project Overview

This capstone project investigated the relationship between food desert prevalence and obesity rates across 3,129 U.S. counties, examining whether geographic barriers to food access explain differences in health outcomes. The analysis employed three regression models of increasing complexity to test this hypothesis while controlling for socioeconomic confounders and examining conditional effects.

Key Findings

- Food deserts and obesity are correlated: A weak but statistically significant bivariate correlation exists ($r = 0.14$, $p < 0.001$), supporting the notion that food access and health outcomes are related.
- Socioeconomic factors dominate the relationship: When poverty, income, and education are controlled, the food desert effect disappears ($p = 0.261$). This indicates that the bivariate relationship is spurious driven by confounding rather than a direct causal mechanism. Education ($\beta = -0.263$) and poverty ($\beta = +0.163$) are 10× more predictive of obesity than food deserts.
- Context matters: Moderation by poverty: Poverty significantly moderates the food desert-obesity relationship (interaction $p < 0.001$). Food deserts have stronger effects in high-poverty counties, suggesting that geographic barriers compound disadvantage in economically fragile areas.
- Model comparison reveals confounding: R^2 increases from 0.020 (Model 1) to 0.504 (Model 2) when socioeconomic controls are added, a 25-fold improvement that

visually demonstrates the importance of statistical control in geographic health disparities research.

Major Challenges Encountered

- Data Integration Across Geographic Scales: Food desert data was available at census tract level (11-digit FIPS codes) while health outcomes were at county level (5-digit FIPS codes). Aggregating census tract data to the county level required careful coding to preserve data integrity and ensure accurate prevalence calculations. The solution: extracting the first 5 digits of census tract IDs, grouping by county, and calculating the percentage of tracts meeting the "Low Income and Low Access" criterion.
- Confounding and Spurious Relationships: The bivariate analysis suggested food deserts strongly predict obesity, but this relationship was entirely explained by socioeconomic confounders. Recognizing and properly testing for confounding required moving beyond simple correlation to multivariate analysis. This highlights a critical lesson: geographic associations in health data often reflect underlying social inequities rather than direct causal mechanisms.
- Moderation Effects and Interpretation: While the moderation effect (poverty × food desert interaction) was statistically significant, its practical magnitude was modest ($\Delta R^2 = 0.003$). Distinguishing between statistical and practical significance required careful interpretation and presentation of findings in context.
- Large Sample Size and Multiple Comparisons: With N = 3,129 counties, even small effects achieve statistical significance. This required careful attention to effect sizes (standardized coefficients) rather than relying solely on p-values to assess the importance of findings.

Lessons Learned

- Statistical Significance ≠ Practical Importance: Model 1 showed statistical significance ($p < 0.001$) but explained only 2% of variance. This taught the importance of always examining both p-values and effect sizes and understanding that statistical tools can be misleading without interpretation grounded in domain knowledge.
- Confounding is Invisible in Bivariate Analysis: The dramatic difference between Model 1 and Model 2 illustrates how confounding can mask true relationships. Without properly testing multivariate models, researchers could wrongly attribute obesity to food access rather than socioeconomic disadvantage. This reinforced the importance of theory-driven variable selection and comprehensive control variable inclusion.
- Context Shapes Relationships: The moderation finding demonstrates that geographic and social relationships are heterogeneous. Interventions should not be one-size-fits-all; food desert programs may be most effective in high-poverty areas where barriers compound. Future work should examine regional variation, rural vs. urban dynamics, and other sources of heterogeneity.
- Data Quality and Integration are Critical: Careful attention to column names, FIPS code formatting, and geographic unit alignment was essential. Small errors in data preparation would have propagated through all subsequent analyses. This reinforced best practices in data cleaning, validation, and documentation.
- Reproducibility Requires Transparent Code: Python-based analysis with version control and inline documentation ensured that all steps could be audited and

reproduced. This approach is far superior to manual Excel changes and is essential for research integrity.

Project Contribution and Future Direction

This project contributes to the field in several important ways:

- It provides empirical evidence that food desert-obesity relationships documented in prior research are largely confounded by socioeconomic factors, suggesting that policy responses must address root causes (poverty, education) rather than symptoms (food access alone).
- The moderation analysis identifies an important conditional relationship: food deserts matter more in high-poverty areas. This suggests a nuanced policy approach where food desert interventions are most effective in economically disadvantaged communities.
- Methodologically, it demonstrates best practices in multivariate analysis, confounding assessment, and interaction testing.

Future work building on this analysis could include:

- Mechanisms: Examine potential pathways. Do food deserts affect diet quality, food costs, dietary diversity, or time spent on food acquisition? Which pathways matter most?
- Regional variation: Are food desert effects similar in rural vs. urban areas? In different U.S. regions? Are regional policy interventions more effective?

- Additional variables: Healthcare access, SNAP participation rates, vehicle ownership, built environment factors (walkability, gym access) could further refine our understanding.
- Heterogeneous effects: Use machine learning or causal forests to identify which subpopulations benefit most from food desert interventions.
- Qualitative research: Interviews and focus groups in high-poverty, high-food-desert areas could illuminate the lived experience and identify barriers beyond data captured in county-level statistics.

Real-World Implications

This analysis addresses critical public health and social policy questions with implications for healthcare systems, urban planning, education policy, economic development, and health equity. The following sections examine how these findings inform responses to real-world problems affecting millions of Americans.

The Public Health Crisis: Obesity and Its Burden

Obesity is one of the most pressing public health challenges in the United States. According to CDC data, approximately 42% of American adults are obese, with significant disparities across race, ethnicity, and socioeconomic status. This analysis found that obesity rates in the county sample ranged from 17.6% to 52.2%, representing a three-fold variation across the country. This geographic variation demands explanation and targeted intervention.

Obesity contributes to an estimated 300,000 excess deaths annually in the U.S. (Estimates of Excess Deaths, 2019) and costs the healthcare system over $147 billion per year in direct medical expenses, plus an additional $66 billion in indirect costs (Cawley, 2021). At the individual level, obesity increases risk for:

- Type 2 diabetes (risk increased 80-fold in obese individuals)
- Cardiovascular disease (heart attack, stroke)
- Hypertension (high blood pressure)
- Certain cancers (endometrial, breast, colon)
- Sleep apnea and respiratory problems
- Joint problems, arthritis, and mobility limitations
- Mental health challenges including depression and anxiety

The health burden of obesity is not equally distributed. The highest obesity rates are concentrated in rural areas, the South, and communities with high poverty rates. In other words, the areas where this analysis found food deserts to be prevalent. However, my analysis reveals an important finding: the problem is not food deserts alone, but the broader socioeconomic context.

Economic Development and Living Wages

This analysis found that poverty rate is a strong predictor of obesity (coefficient: +0.1633, $p < 0.001$). Economic insecurity shapes health outcomes through multiple pathways:

- Food Affordability: In low-income households, budget-friendly foods tend to be calorie-dense, highly processed options. Fresh produce, lean proteins, and whole grains are more expensive per calorie, making poverty a barrier to healthy eating.
- Time Poverty: Low-wage jobs often involve long hours, multiple jobs, and inflexible schedules. This "time poverty" limits time for meal preparation, exercise, sleep, and self-care.
- Stress and Health: Economic insecurity is a chronic stressor linked to elevated cortisol, inflammatory markers, and increased weight gain. The psychological burden of poverty affects health directly through physiological pathways.
- Neighborhood Quality: Poverty concentrates in neighborhoods with fewer resources, higher crime, less safe recreation spaces, and limited access to services. These environmental factors limit physical activity and recreational opportunities.

These mechanisms suggest several policy approaches to address poverty's health effects:

- Living Wage Standards: Cities and states should establish living wage requirements that provide economic security to workers. Research shows that inadequate income is a primary driver of stress and poor health outcomes.
- Job Quality and Benefits: Beyond wages, job quality matters. Access to health insurance, paid sick leave, flexible scheduling, and retirement benefits all contribute to health and economic security.
- Social Safety Nets: SNAP (food assistance), TANF (cash assistance), housing assistance, and other programs are essential. However, eligibility barriers and benefit adequacy must be addressed.

- Economic Opportunity: Job training programs, business development services, and capital access for small enterprises in low-income communities create pathways to self-sufficiency and wealth building.
- Childcare and Early Education: Access to affordable, quality childcare removes barriers to employment for parents, particularly mothers, and provides developmental benefits for children.

References:

CDC. (2024, October). *Adult Obesity Facts*. Obesity. https://www.cdc.gov/obesity/adult-obesity-facts/?CDC_AAref_Val=https://www.cdc.gov/obesity/data/adult.html

Dutko, P., Ploeg, M. V., & Farrigan, T. (2012, August). *Characteristics and Influential Factors of Food Deserts*. Www.ers.usda.gov. https://www.ers.usda.gov/publications/pub-details/?pubid=45017

Cawley, J. (2021). Direct medical costs of obesity in the United States and the most populous states. *Journal of Managed Care & Specialty Pharmacy*, *27*(3), 354–366. https://doi.org/10.18553/jmcp.2021.20410

Center for Disease Control. (2021, March 8). *PLACES: Local Data for Better Health*. Centers for Disease Control and Prevention. https://www.cdc.gov/places/index.html

Ghosh-Dastidar, B., Cohen, D., Hunter, G., Zenk, S. N., Huang, C., Beckman, R., & Dubowitz, T. (2014). Distance to Store, Food Prices, and Obesity in Urban Food Deserts. *American Journal of Preventive Medicine*, *47*(5), 587–595. https://pmc.ncbi.nlm.nih.gov/articles/PMC4205193/

Estimates of Excess Deaths - Supplemental Analysis. (2019). https://www.cdc.gov/nchs/data/hestat/excess_deaths/excess_deaths.htm

PolicyMap. (n.d.). PolicyMap data indicators. https://www.policymap.com/

USDA. (2019). *USDA ERS - Food Access Research Atlas*. Www.ers.usda.gov. https://www.ers.usda.gov/data-products/food-access-research-atlas/