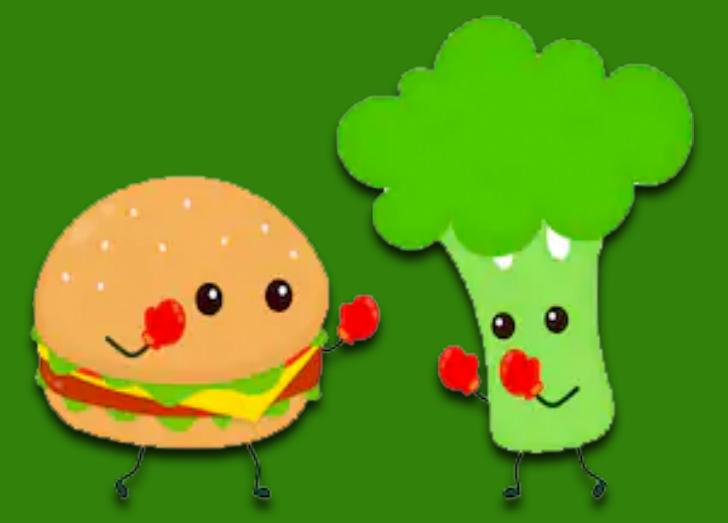# Food For Thought
## *Understanding how food access influences obesity & income*

Megha Malpani, Maddie Becker, Devin Howard, Will Richardson, and Shehryar Hasan

Introduction

Problem

- Our project aims to explore the intersection between food access, health, and economic disparity in Los Angeles county. As the following two graphs show, ***there is a clear link between household income and obesity.*** Using these graphs as a starting point, we explore how food plays into this equation.
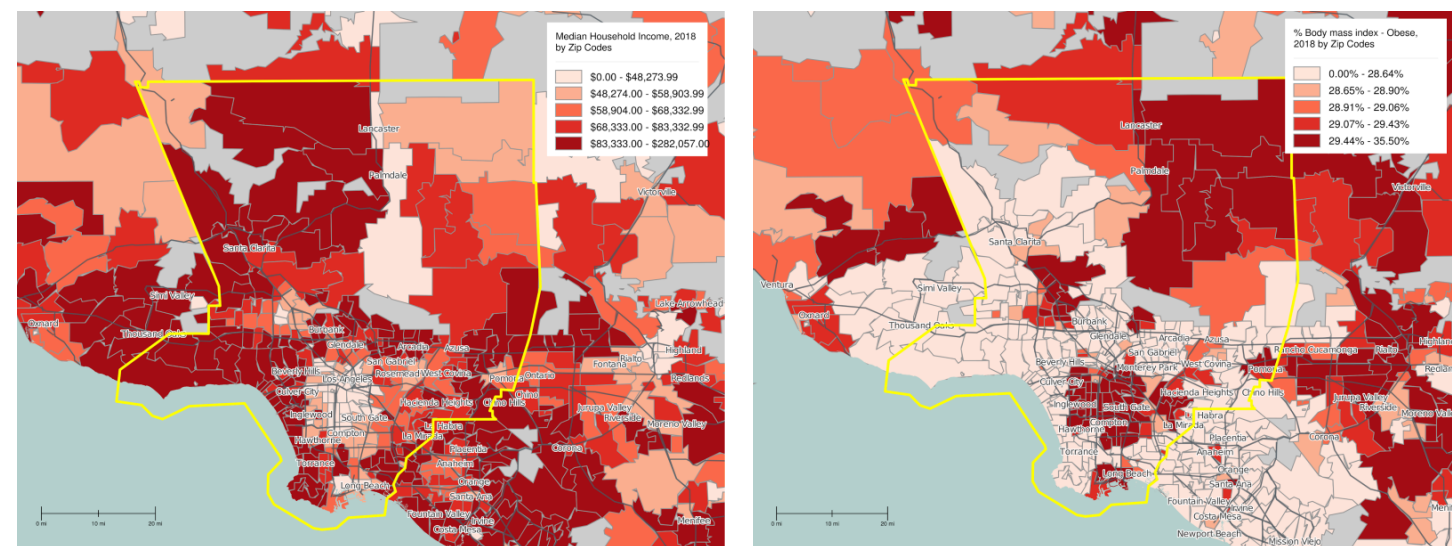
Median household income in LA County (left). Body mass index in LA County (right).

Purpose

- Our goal is to better understand the extent to which access to healthy food plays a role in income and health in LA. Existing studies argue that food access is a crucial determinant for neighborhood health, evidenced by high rates of obesity and diabetes in areas with minimal healthy food access. Our focus is on Los Angeles, as it has the highest number of fast food restaurants per capita in the United States, and because there is ample data available on its neighborhoods.

Hypothesis

- We hypothesize that supermarket and fast food access for a given neighborhood are inversely related (i.e., the more supermarkets, the less fast food restaurants) and that the number of supermarkets is directly correlated to income. We believe that this should explain above average rates of obesity in low income neighborhoods that lack access to healthy food options.
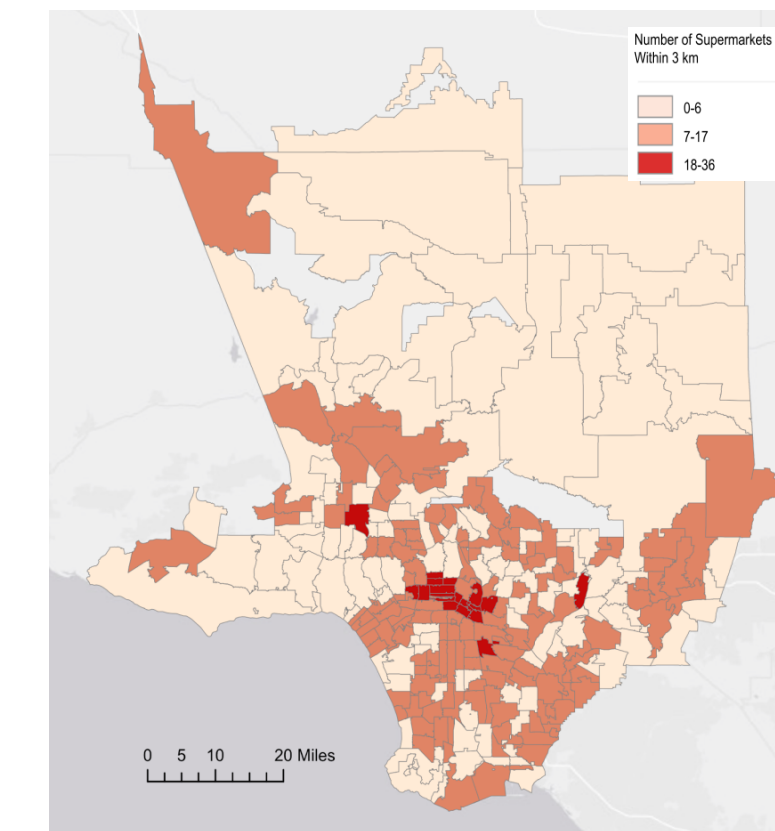
Results

Number of Fast Food Restaurants
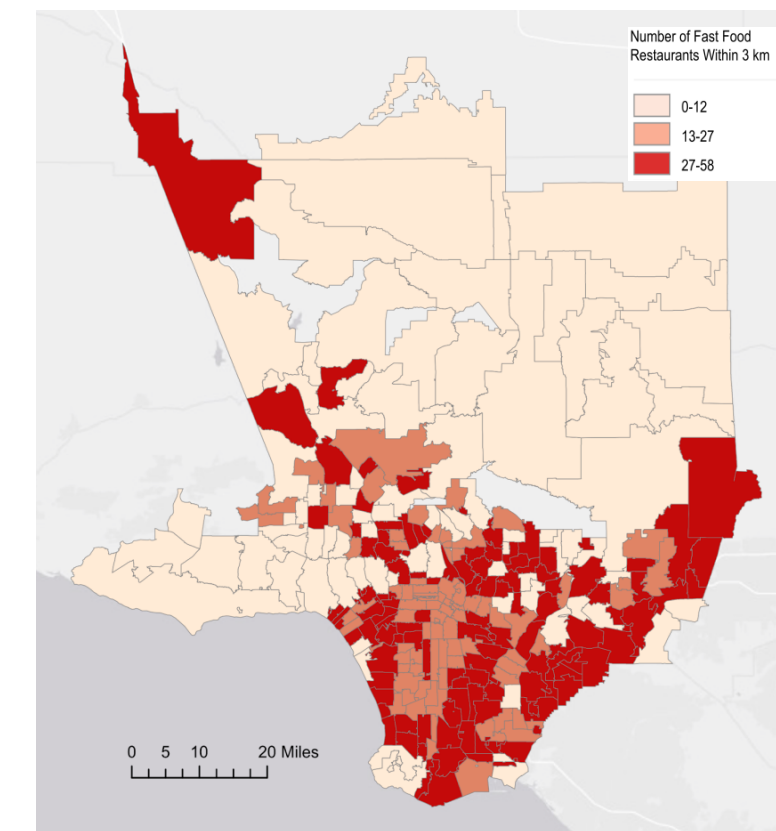
- Higher densities of fast food restaurants are visible in the center of LA
- This is **the same area where we see low income neighborhoods with high percentages of obesity**
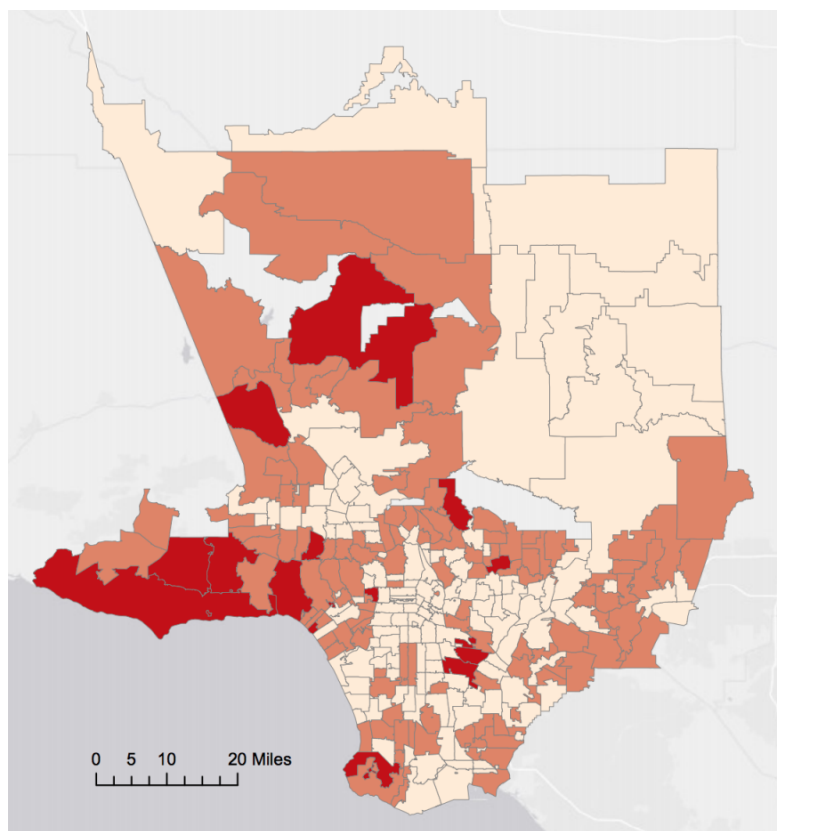
Number of Supermarkets

- Supermarket density reveals a similar, but less obvious pattern: areas near the center of LA contain more supermarkets than areas outside
- Supermarkets do **not** appear to be in different areas than fast food restaurants

Clustering of Economic and Health Data
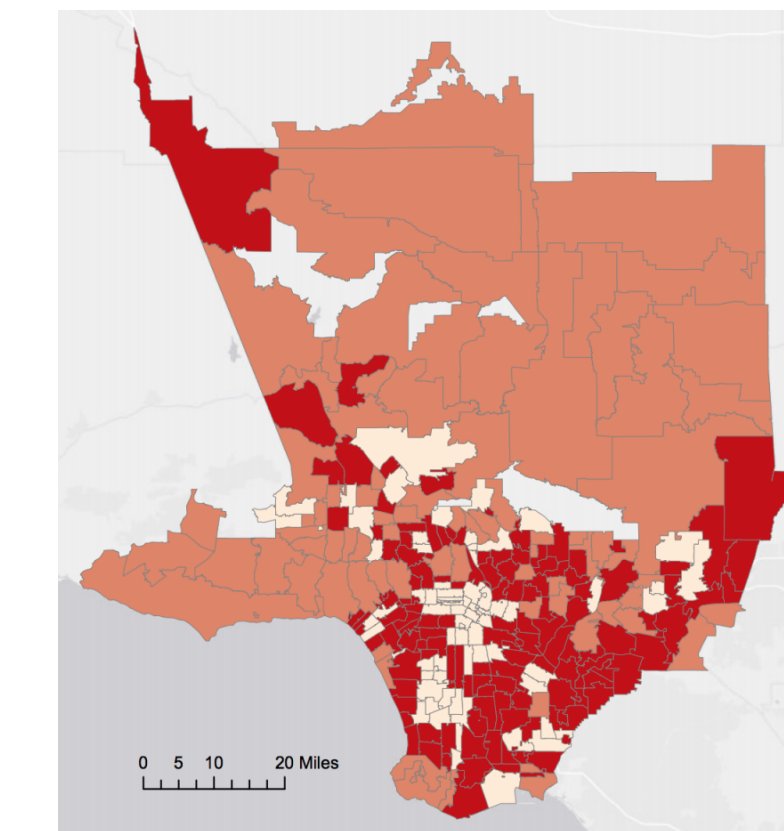
- Zip codes clustered using KMeans with k=3 on income, unemployment, obesity, and diabetes data
- A "center" of LA can be found, but is less clearly defined when adding unemployment and diabetes
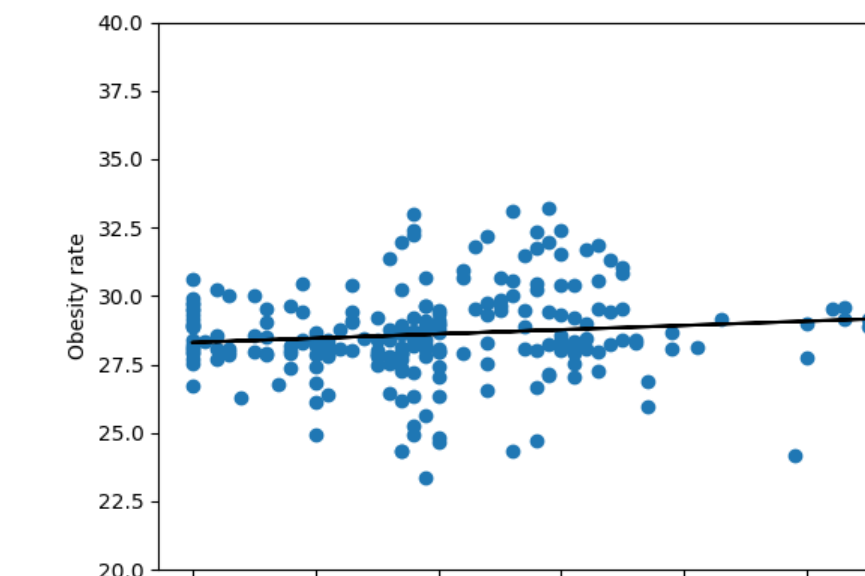
Clustering of Food Access

- Zip codes clustered using KMeans with k=3 on supermarket and fast food data
- Clear "centers" of LA can also be found

Obesity vs Supermarkets
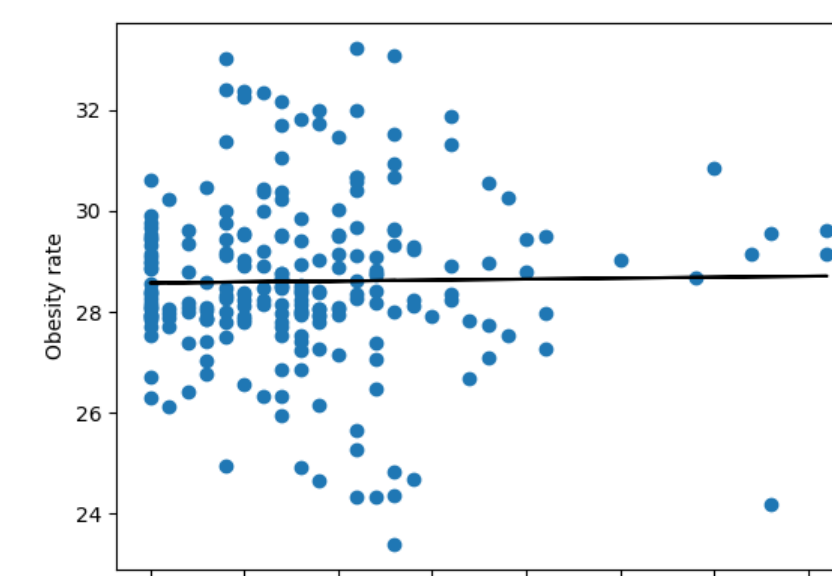
- P-value = 0.817
- As the number of supermarkets in a 3km radius increases, the neighborhood's obesity rate seems to go up
- However, this conclusion is not statistically significant

Obesity vs Fast Food

- P - value = 0.087691
- As the number of fast food restaurants within a 3km radius increases, the neighborhood's obesity rate increases.
- This is a logical conclusion given the unhealthy reputation of fast food

Supermarkets vs Income
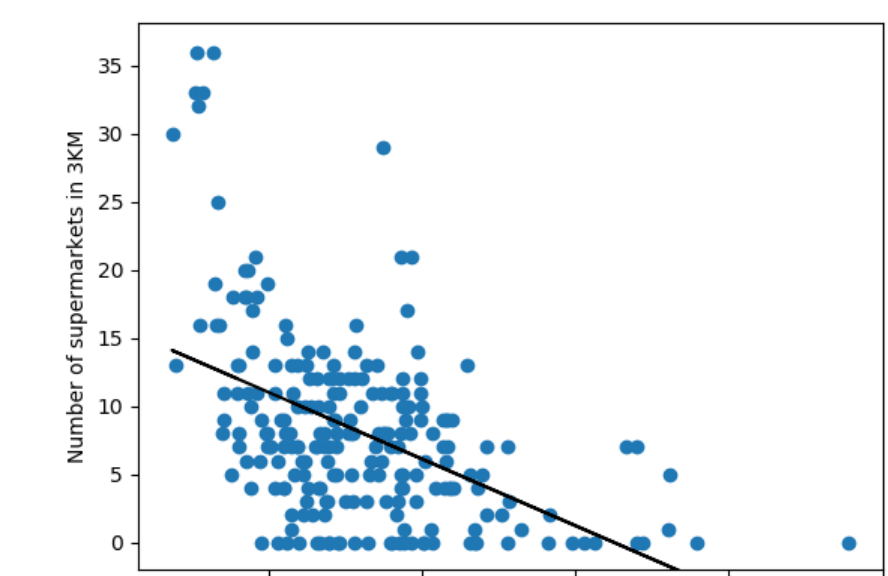
- P-value = 1.820766e-14
- As the average income of a neighborhood increases, the number of supermarkets in the neighborhood falls
- This is an unexpected result, but may be due to higher real estate prices or the spread out nature of wealthier neighborhoods

Fast Foods vs Income
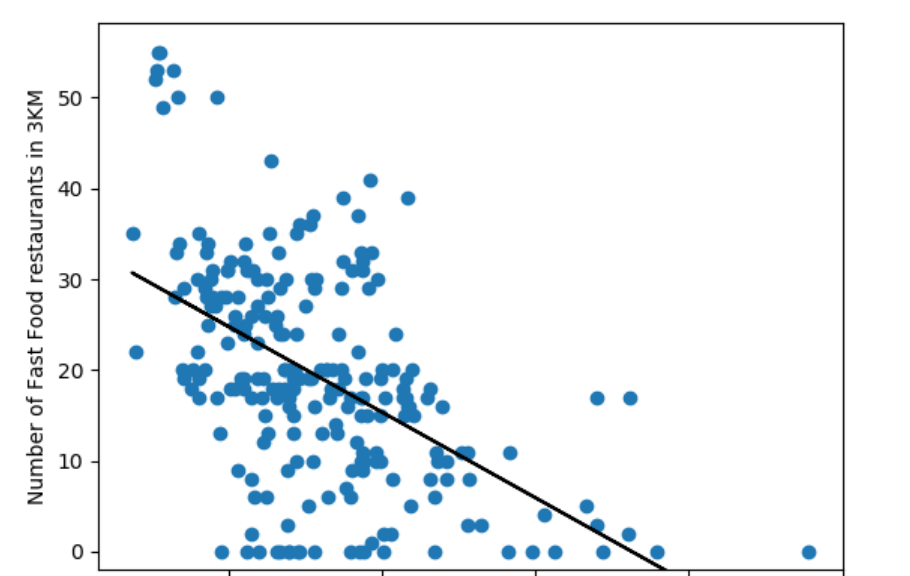
- P value = 0.0
- As the average income of a neighborhood increases, the number of fast food restaurants in the neighborhood falls.
- This is an indicator that poor neighborhoods have high access to unhealthy foods

Data

Simply Analytics

- To construct our dataset on health and wealth in LA, we initially gathered data from different websites, joining them with zip code as the primary key. However, we later discovered SimplyAnalytics, which allowed us to generate a CSV file with all our desired fields: employment, income, obesity and diabetes data, grouped by zip code.

Google Maps API

- As SimplyAnalytics.com did not offer data on fast food restaurants and supermarkets in specific geographical location, we generated a second dataset using the Google Maps Places API. For this, we scraped coordinates approximating the centers of each of the zip codes in Los Angeles county to retrieve the numbers of supermarkets and fast food restaurants within a 1km and 3km radius. A table with this data and a table with the data above comprised our SQL Database.

Challenges

- Data collection served as the most challenging aspect of this project. We found it hard to find a consistent concept of "neighborhoods," as different datasets were grouped by different location indicators (i.e geoIDs, census tracts, county districts, and zip codes). We ultimately chose to represent neighborhoods by zip codes and scraped corresponding latitude and longitude data. However, we recognize that this solution may result in overlapping information between zip codes and/or data that is not representative of all communities living in one zip code.

Main Takeaways

- **The number of fast food restaurants directly relates to both income and obesity, while unemployment, diabetes, and number of supermarkets require further study.** In other words, more fast food restaurants in lower income neighborhoods often mean higher rates of obesity.

- What is most unexpected is the finding that the prevalence of supermarkets is **not** a meaningful indicator of either income or health. This implies that simply having access to supermarkets does not lead to a healthier lifestyle. There may be many other factors at play, such as prices, food preference, and convenience.

Moving Forward

- We'd like to understand the price ranges of supermarkets in each zip code, as well as the revenues generated by both supermarkets and fast food restaurants. Though we found a high number of supermarkets in low income neighborhoods, we don't have a concept of who frequents them and how much money they spend. It's possible that they are visited less often than fast food restaurants or lack healthy options (like an average corner store), which could help explain the relationship between income and obesity that we know exists. In other words, we don't believe our Google Maps API data had a strict enough definition of a supermarket, so future exploration could look into this.