Abby Kim, Griffin Homan, Jiahua Chen, Michelle Zhang

Professor Jeff Rzeszotarski

INFO 3300 / CS 3300

11 October 2024

Project One Final Report

For our final project visualization, our team attempts to visualize the relationship between fast-food restaurant density and obesity rates across U.S. states. With fast-food chains being a prominent part of American culture, we wanted to investigate whether there was a correlation between states with a high concentration of these restaurants and higher obesity rates. In addition, we hoped to explore how the density of individual fast-food chains vary across states to see if certain chains are more common in areas with higher levels of obesity. Through these visualizations, we hope to see a clearer picture as to how fast-food availability may influence obesity rates across the United States.

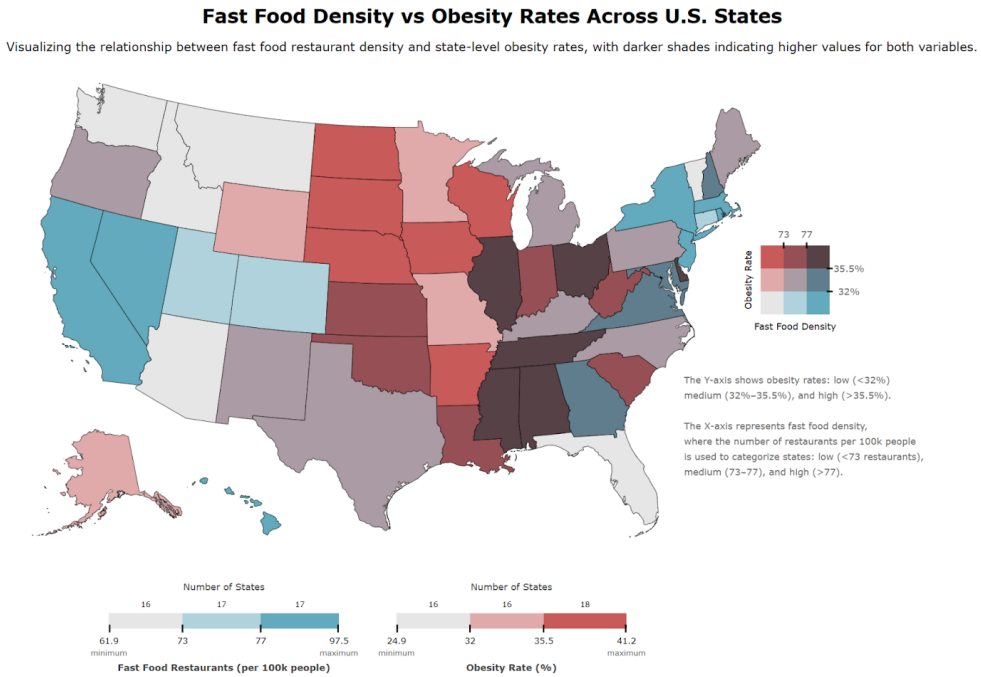
Data Used

The data used in this project comes from multiple sources, Fast food restaurant data, including the number of fast food restaurants per 100k people was sourced from Nicerx.com, which aggregates data from the U.S Census Bureau and population estimates. The dataset comprises the 10 most popular fast food brands (Dunkin' Donuts, Subway, Starbucks, McDonald's, Dairy Queen, Wendy's, Domino's, Taco Bell, Burger King, and KFC). The obesity data was gathered from Statista, providing the percentage of adults with obesity in the United States as of 2023, broken down by states. The obesity percentages for each state were extracted manually to incorporate into the dataset. The datasets from both sites were then manually combined into a JSON file (state_obesity_fastfood_data.json), which includes key variables for each state: obesity percentage, total fast food restaurants, full-service restaurants, and counts for individual fast food chains. Additionally, geographic data was sourced from Eric Clst's US GeoJSON dataset, where the 20m-resolution US States GeoJSON was downloaded and used to create the map of the United States. For the second visualization, additional data processing was per

Data processing for the second visualization:

> Data for states with the maximum, minimum, median, lower quartile, and upper quartile obesity rates were selected after ordering the dataset in ascending order. Only the obesity percentages and columns for the restaurants of interest were retained. This preprocessing was conducted in Python using Pandas, and the resulting target data was produced in dictionary format and copied over to index.html.

Visualization #1 (Bivariate Choropleth Map)



A bivariate choropleth map was chosen for this visualization because it allows for the simultaneous representation of two variables (obesity and fast food density) using color. Our approach was able to eliminate the need for an extensive amount of charts or graphs while conveying the same information.

Colors

The color scheme used in the map is a 9-color matrix that represents three obesity levels and three fast food density levels. In order to guarantee distinct visual contrasts between each bin, the colors were carefully selected. Increasing values for both variables are represented by shades that go from lighter to darker.

Binning

The obesity bins are as follows: low obesity (0%-32%), medium obesity (32%-35.5), and high obesity (>35.5%). Additionally, the fast food density bins are low (0-73 restaurants per 100,000 people), medium (73-77), and high (>77). This approach was determined after we reviewed our data, ensuring that the cutoffs were suitable and the map shows sufficient color variation.

Marks and Visual Channels

Each state is represented by an SVG <path> element. These paths serve as the marks, defining the geographic boundaries of each state. The AlbersUSA projection fixes each state's geographic location and guarantees that the spatial relationship between states are accurately depicted. Color

intensity and hue serves as the primary channels to represent the binned categories of obesity rates and fast food density.

Legends

A 3x3 grid helps users interpret color combinations, with each cell representing a specific mix of obesity and fast food density bins. Fast food density (low to high, left to right) is represented by the x-axis, while obesity rates (low to high, bottom to top) are represented by the y-axis. We believe this layout can ensure easy understanding of the color coding system used across the map. We also thought it was wise to implement two gradient legends to assist in clarifying how colors relate to specific ranges for each variable. We specifically positioned these two legends at the bottom to help avoid clutter and help maintain focus on the map as the central visual element.

Trade-offs

While using a bivariate map adds value by visualizing two variables at once, it introduces challenges in finding distinct colors that don't overwhelm or confuse viewers. To keep each color combination distinct and simple to understand, careful balancing was needed to prevent clutter. Another trade-off encountered was the attempt to add state abbreviations directly onto the map. While labeling the states would enhance identification, it made the map overwhelming and interfered with the ability to clearly interpret the colors, so it was ultimately excluded to maintain a clean and readable presentation.

Visualization #2 (Lollipop chart)

Our second visualization compares four fast food restaurants—McDonald's, Starbucks, Subway, and Taco Bell—across five U.S. states with varying obesity rates: Colorado, Nevada, Virginia, South Carolina, and West Virginia. The states are organized from the lowest to the highest obesity rates, with Colorado having the lowest and West Virginia the highest. The specific restaurants were selected to provide a diverse range of fast food chains, allowing for a more comprehensive comparison between states with different obesity rates.

Marks and Visual Channels

For this visualization, we used the logos of the restaurants as marks, along with lines to connect them. Initially, we debated representing our data with a bar chart, using different colors for each restaurant and a legend on the side. However, we ended up choosing a lollipop chart where the restaurant logos are the “lollipop heads.” This decision was made to help readers easily identify the trends, as these logos are well-known and recognizable. By using logos directly on the chart, the users wouldn’t have to constantly look back at the legend every time to see which restaurant is which.

The main visual channels we used were the position on a common scale (y-axis) and the length of the lollipop sticks. Positioning the logos on the y-axis according to the number of restaurants per 100K people provided clarity in visualizing trends. Additionally, varying the length of the sticks highlighted differences between restaurant counts across states.

Choice of Visualization:

We chose the lollipop chart to group the four fast food restaurants for each state, making it easier to compare them visually. By using the logos of the different restaurants, we made it so that we can help the readers quickly and easily understand the trends of each fast food chain without relying on additional legends or color schemes.

Colors:

For this visualization, we decided to use a black and gray color scheme. This choice allowed the focus to remain on the trends in restaurant counts, shown by the position of the logos. Using vibrant colors could have drawn attention away from the position of the logos and the differences between the states, which are the main point we are trying to convey.

Choice of the Restaurants:

We selected four different fast food restaurants to avoid overwhelming readers with too much information. These particular restaurants were chosen because they had complete data (no missing values) across all five states, and they are well-known chains that would be interesting to compare.

Choice of the States:

We thought it would be interesting to compare the states based on their obesity rates, organizing them from lowest to highest. So through analyzing the data through code, we used a statistical 5-point summary, finding that Colorado had the lowest obesity rate, Nevada ranked 13th lowest, Virginia was at the median (25th lowest), South Carolina was 35th lowest, and West Virginia had the highest obesity rate. This order allowed for a clearer visualization of the trends in restaurant across states with different obesity profiles.

Additionally, we included text beneath each state of this information to provide more context for why these states were chosen and to clarify the ordering of the states by obesity rate.

Position of the Logos:

We aligned the center of each restaurant logo's radius with the corresponding number of restaurants per 100K people on the y-axis, making it easier for readers to understand the data. Furthermore, we included the exact values above each logo to provide clarity and avoid confusion.

D) The story. What does your visualization tell us? What was surprising about it? What insights do you want to convey to the viewer of your visualization?

Our visualizations explore the relationship between fast food density and obesity rates across different states in the U.S. We were curious to see if a higher concentration of fast food restaurants correlates with higher obesity rates, which is often assumed due to the association between fast food consumption and unhealthy eating habits.

Our state map visualization shows that states in the West and Northeast regions tend to have a high density of fast food restaurants while maintaining relatively lower obesity rates, shown by a blue hue. This was unexpected, as we anticipated a more direct relationship between fast food availability and obesity. Conversely, in the Midwest and Southeast regions, obesity rates are high without a consistent increase in fast food density, indicated by the red hues. Toward the middle of the region, many states have high obesity rates despite lower fast food density. However, moving east, there is a stronger connection between high fast food density and high obesity rates. Five states were identified that are low in both fast food density and obesity rates, six states high in both, and eight states with moderate values for both. Overall, fast food density appears to be more related to geographic and economic factors than obesity rates alone.

In the restaurant visualization, we were surprised to find that **Starbucks** locations decrease almost linearly as obesity rates increase. Starbucks is often linked with leisure and higher socioeconomic status, suggesting that its distribution may be influenced more by income levels than obesity rates. Other fast food chains that offer quick and convenient meals fluctuate in number across states but do not show clear patterns like Starbucks. A notable exception is Subway in West Virginia, state with the highest obesity rate. Subway's presence spikes significantly in West Virginia compared to other states, which could be interpreted in various ways by different readers.

Overall, these visualizations suggest that while fast food density might align with demand in certain areas, it is more strongly influenced by geographic or economic factors than by obesity rates. In future studies, it would be insightful to explore how **socioeconomic status** correlates with obesity across different states. Understanding the relationship between income levels, access to healthier food options, and fast food availability could provide deeper understanding into the root causes of obesity and help shape more effective public health strategies.

Map:

- West and north east: high fast food density while generally low obese rate (blue hue)
- Middle West and southeast: high obesity rate without high fast food density (towards the middle of states) high obesity and high density towards the east (signified by the red hue)
- 5 states that are low in both values
- 6 states that are high in both values
- 8 states moderate in both values

- Fast food density might by in sync with obesity rate to support demand but fast food density seems to be more related to geographic location (economic development)

Lollipop:

- Surprising: Starbukcs decreases almost linearly as obesity rate is lower
    - Products at starbucks - leisure
    - Other restaurants (quick food choices) fluctuates but don't have trends as obvious
    - Except subway number increases drastically in west virginia (most obese) compared to other chains
- Obesity is often related to socioeconomic status
- Campaigns to reduce obesity should focus on healthy food access

Future Study

- It would be interesting to compare the different socioeconomic status of each state compared to obesity rate.

**NOTE:
For our Visualization 2, we wanted to group the restaurants by each state but we didn't learn in class on how to group using d3. So, one of our teammates went to office hours and the TAs instructed them to use d3.group and scaleBand to group the restaurants by the states. The TA also provided us with documentation and an example of how the d3.group and the scaleBand could be used so we followed along this example: https://observablehq.com/@d3/grouped-bar-chart/2

Team contributions
an outline of team contributions to the project. Identify how work was broken down in the group and explain each group member's contributions to the project. Give a rough breakdown of how much time you spent developing and which parts of the project took the most time.