Project 2 Report
By: Zhan Wu, Yingshi Zhu, Minwei Jiang, Qi Su
Due: Friday, Nov. 19, 2021

Topic:

Our group visualizes the number of US fast food restaurants by state on a choropleth map. Each state will have a click on event to show more information on the right side of the map. Those information including cancer rate, obesity rate, heart disease rate, and races, can be shown as bar graphs, line charts, pie charts, etc. We can use this interactive map to see the relationship between the number of fast food restaurants and different types of disease rate.

1. **Data Description:**
    a. **Map:**
        i. **The United States map data:**
            This dataset is provided by the professor during class. We use this dataset to construct the bottom layer of the map which contains the area and the border of each state.

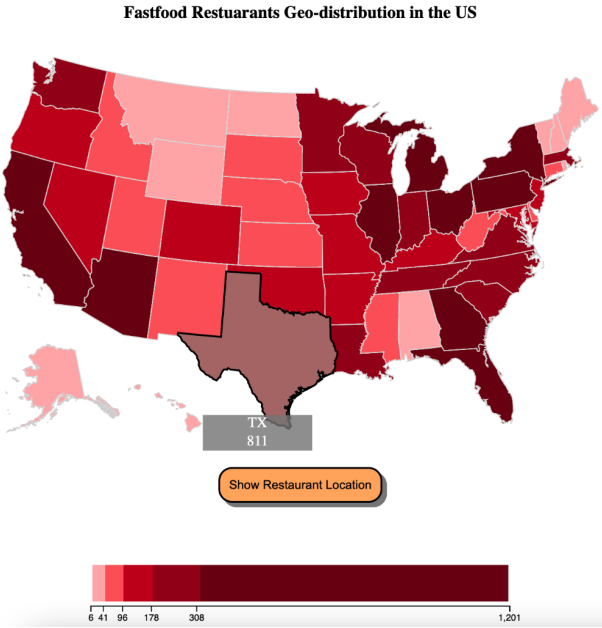
        ii. **Fast food restaurants location in US:**
            https://www.kaggle.com/khushishahh/fast-food-restaurants-across-us
            This dataset contains 10 columns of information of 10000 fast food restaurants in the United States. Each row in the dataset represents the detailed information of a single fast food restaurant. There are 10 attributes that describe a fast food restaurant, which are address, categories, city, country, latitude, longitude, name, postalCode, province, and websites.
            After evaluating the topic and our visualization ideas, we used Python to pre-process the dataset so that we can remove those unrelated columns that we will not use in the visualization, then we grouped rows by states so that we can compute the number of fast food restaurants in each state.
            We also need the latitude and longitude information of each fast food restaurant to indicate their geographic location on the map, so we kept a copy of the original dataset and used it to draw circles on the map, where each circle represents one fast food restaurant. In addition, we add state code to the new dataset so that we can assign value to each state by using state code later.

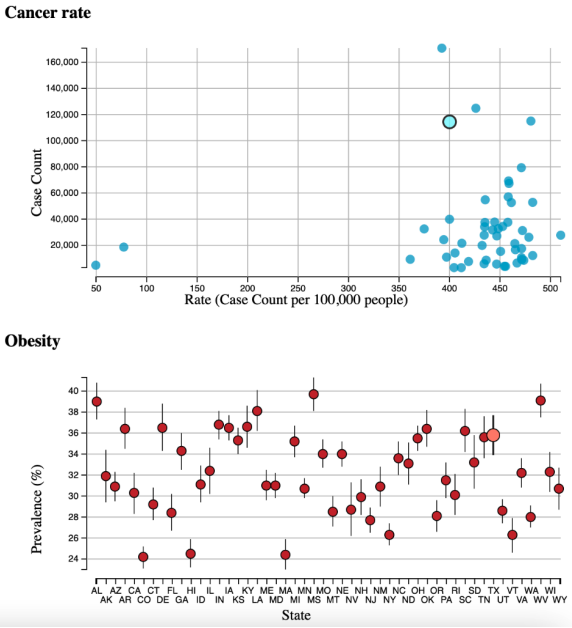
    b. **Plots:**
        i. Obesity rate by state:
            https://www.cdc.gov/obesity/data/prevalence-maps.html

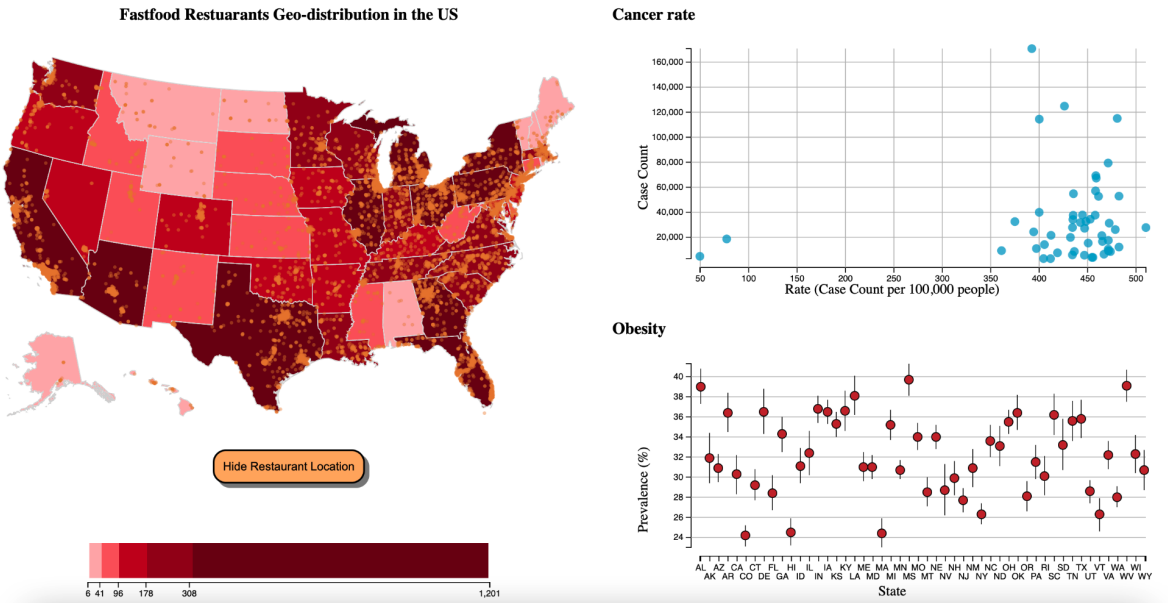
The dataset contains three columns of data of the latest adult obesity prevalence in the United State from CDC. Each row in the dataset represents the detailed information of a state. Each state has two attributes which are prevalence mean and its CI (Confidential Interval) 95%.

ii. Cancer rates by state:
https://gis.cdc.gov/Cancer/USCS/#/AtAGlance/
The dataset contains three columns of data of the new cancer rate in the United State from CDC. Each row in the dataset represents the detailed information of a state. Each row has two attributes which are cancer case count and cancer rate (case count per 100,000 people).

2. Overview of Visual Design Rationale







a. Map

We choose to use a choropleth map to visualize the fast food restaurants geographic distribution in the US. We use polygon, path, and circle as marks, and varying saturation and hue of color as channel. We choose to fill each state's polygon with colors approximate to red and fill circles of each restaurant location with a color of yellow. This color combination matches the color in the McDonald's Logo which helps to remind users this visualization is about fast food restaurants.

A state's polygon will have a deeper, more saturated red color if the number of fast food restaurants in that state is high, and a lighter red color if the number of fast food restaurants is low. This design follows the intuition of people that if a distribution is dense and concentrated it tends to have a deeper color. We used white color in the border of each contingent state, this color can provide a distinct visual separation between states. We also draw a color scale legend below the map to indicate the range of numbers represented by each color.

Yellow circles with half opacity will be shown after clicking the Show Restaurant Location button. Using not a full opacity of circles can make the location with a high fast food restaurants concentration more stand out from the graph.

b. Plots

We choose to use a scatter plot to visualize the data for cancer rate and a lollipop chart to visualize the data for obesity prevalence. In the scatter plot for cancer rate, the x axis is the cancer rate (case count) and the y axis is the case count. The circles representing each state are scattered according to the cancer rate and case count of each state in this plot . We fill each circle with a blue color, which is distinguished from the map. Besides, we add jitters to each circle because some circles overlap with each other.

In the lollipop chart for obesity, the x axis is the state and the y axis is prevalence. The circles representing each state are located in the plot according to the prevalence mean in each state. The length of the line across each circle is the 95% confidence interval. The percentage interval that the line represents is the interval associated with 95% confidence interval. We fill each circle with a red color, which is in contrast to the circles in the upper plot.

3. **Overview of Interactive Elements Design Rationale**
    a. **Map**

We design several interactive elements on the map.

First, we add a hover event on each state in the map. A tooltip will show on the middle bottom of the state polygon when the mouse hovers on. Users can conveniently know the state code and the total number of fast food restaurants in that state from the tooltip. The background color of tooltip has a half opacity, so that the user can still see the color of state underneath the tooltip. The opacity of the filling color as well as the border color of the state's polygon change when the mouse hover over. This visual effect helps users quickly identify which state the user is currently looking at.

Second, we add a Show Restaurant Location button to show the detailed geographic location of every restaurant. Circles will show on the map after users click the button, and the hover event is disable when the yellow circles are shown. The text content in the button will change to Hide Restaurant Location. This feature can help users dive deeper into the detailed distribution of restaurants in each state, and by hiding these circles at the beginning can provide a cleaner graph as a first impression. Click the button again, the visibility of yellow circles will be set to none, and the hover event will resume. We also design a custom style for the button, the opacity of background color in the button changes every time users hover and click on the button. This design makes the button more distinguishable and interactive.

    b. **Plots**

For the first plots on the right side, we add a hover event on each circle in the scatter plot. When the mouse hovers over a circle, the circle will become larger and lighter, and a black stroke will be added. Some circles still overlap with each other after adding jitters so we highlight the circle by changing these three attributes to make the selected circle stand out. Additionally, a tooltip will show beside the circle. Users can know the cancer case count and cancer rate of each state from the tooltip.

For the second plot on the right side, we add a hover event on each circle in the lollipop chart. When the mouse hovers over a circle, the circle will become larger and lighter. This lets the users easily notice the selected circle. Additionally, a tooltip will show beside the circle. Users can know the prevalence mean and 95% confidence interval of prevalence from the tooltip.

    c. **Map and plots**

We add some interactive elements between the map and the two plots. First, the two plots will show if any state is clicked on the map. And clicking any state can also close the two plots.

Second, when the mouse hovers over a state in the map, the lightness and size change of the corresponding circle in the scatter plot and the lillpop chart also occur.

4. **Story**
   a. **Map**

      By looking at the map, we found out that west coast, midwest, and east coast states tend to have more fast food restaurants than rocky mountain and gulf coast states. In other words, southern and eastern states tend to have more fast food restaurants than northern and western states. This trend might also be caused by the limited number of fast food restaurants in our dataset or the differences in population among these states. We also noticed that the area of each state is not strongly correlated with the number of fast food restaurants in that state, for example, Montana state is much larger than Ohio state, but it only has 38 fast food restaurants, compared to 522 fast food restaurants in Ohio state.

   b. **Plots**

      We want to find out the relationship between the number of fast food restaurants and the cancer rate and obesity prevalence in each state. By hovering over some states with lighter color to the state with darker color, we can see that the state with more fast food restaurants can have lower cancer rate and obesity prevalence. The number of fast food restaurants does not have a high positive correlation with the cancer rate and obesity prevalence.

5. **Outline of Team Contributions**

Member	Key Contributions
Zhan Wu	Data Pre-processing, Map Designing, Map Implementation, Report
Yingshi Zhu	Map Designing, Implement interactive elements on Map, Report
Minwei Jiang	Plots design and implementation, Report
Qi Su	Interaction between map and plots, Report

It took about 4 hours for our team to discuss the visualization ideas and make the final decision on the topic. And we have spent about 3 hours to complete the report book.

Zhan and Yingshi are mainly in charge of designing and implementing the map, and it took about 10 hours in total, which contains data pre-processing, basic map implementation, applying different colors to each state, drawing circles on the map, adding interactive elements, adding legends, and debugging.

Qi added some interactive elements between the map and the two plots and wrote the content about plots in the report.