Data Visualization Project Proposal - Group_A_Nutrition

Final Project -- Project Idea

Team Members:
Allison Jiang
Qinyao Xia
Cherie Yu

Project Title: Obesity and Nutrition in the United States

Abstract:

In this project, we aim to present a static display of obesity rates in adults across the different states in the U.S., and more pertinently, analyze how differences in the level of nutritional intake across the different states correlate with obesity rates. Since it is widely known that one's diet plays a key role in affecting one's weight and health, we believe that states with a widespread number of fast food outlets see greater occurrence of obesity as the presence of such unhealthy food options fuel adults living in these to adopt unhealthy diets, thereby neglecting their weight and health. In particular, our analysis consist of three major components:

Obesity:
1. How do obesity rates differ across states? We plan to provide a more detailed analysis by providing the demographic information, such as age group, gender, and income level of the people who are obese for each state.
2. How have obesity rates for each state change over time? Which are the states that experience the greatest increase in obesity rates over time, and which are the states that saw a decline in obesity rates over time?

Adult Diet and Nutrition:
1. Across the states, what is the percent of adults who consume fruits and vegetables less than one time daily?
2. Does there exist a negative relationship between obesity rates and the percent of adults who adopt a healthy diet?

Fast Food Prevalence:
1. How prevalent is fast food across states?
2. For states with high fast food prevalence, are sentiments towards fast food more positive as compared to states with lower fast food prevalence? We plan to analyze positive and negative sentiments using Twitter data.

Method:
We intend to utilize ggplot2, spatial data, as well as text mining techniques in our visualization project .

<u>**Dataset selection:**</u>

The visualization of our project will mainly employ two datasets: The Nutrition, Physical Activity, and Obesity data set from CDC and the a sample of 10000 fast food dataset provided by Datafiniti.

CDC Nutrition, Physical Activity, and Obesity data:
This dataset contains 58,408 observations of 33 variables regarding the obesity in the U.S. from 2011 to 2017. We are interested in variables such as the location of the survey, data for specific questions such as percent of adults aged 18 and older who have obesity problem, education, age range, gender, income, race, and geolocation. We are also interested in the diets of the sample, as surveyed through questions including the percent of adults who consumer fruits and vegetables as least once a day.

Fast Food data:
This dataset contains 10,000 observations of 15 variables regarding fast food restaurants across the U.S. We are interested in looking at the top states where all the fast food restaurants are located and see if states with higher fast food prevalence (greater number of fast food restaurant) is associated with higher obesity rates in those states.

Twitter API data:
We would like to extract text data on individuals' thoughts and opinions on obesity and using text cloud to visualize what are the key determinants people believe that could be correlated with the obesity problem in the U.S. In addition, we would like to analyze the positive and negative sentiments towards fast food.

<u>**Visualization Brainstorming**</u>:

Top states with obesity problems
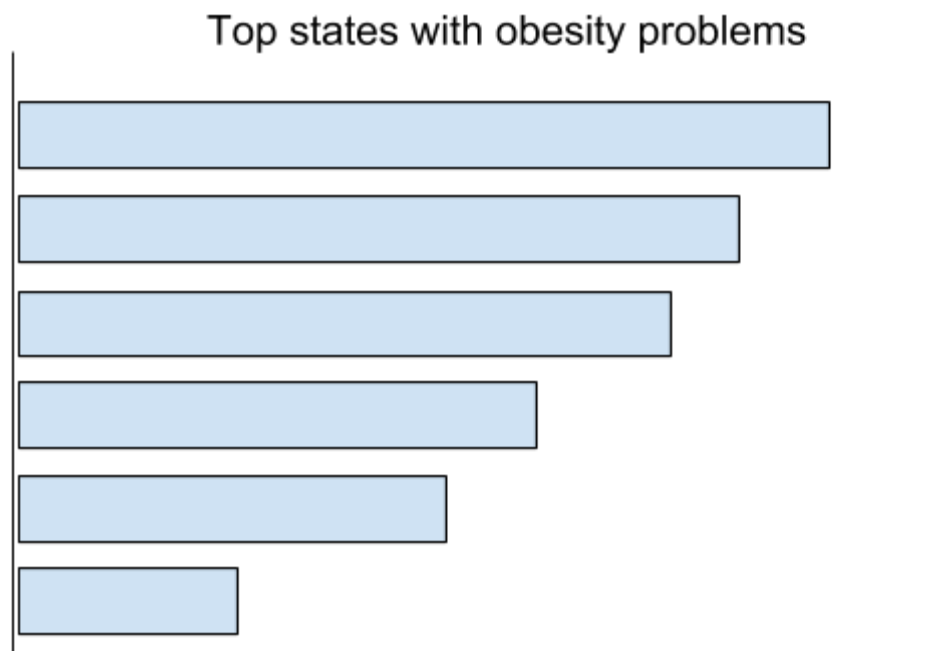
obesity
Diet
Fried
Chicken
FOOD
leisure-time
Health Concern

Three mapping exercise to show nutrition/obesity throughout U.S.

50%	50%
25%	25%
0%	0%
Obesity %	Veggie intake %
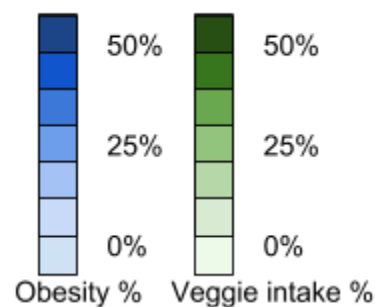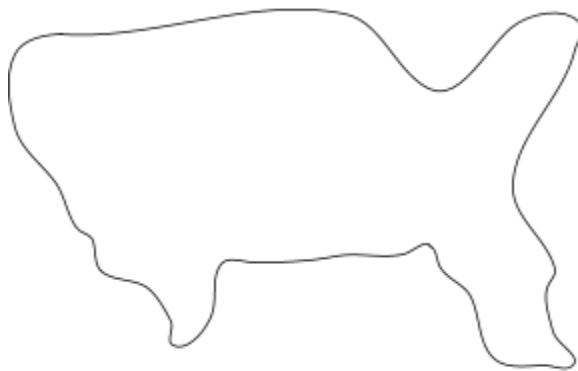
1. Obesity problem intensity throughout the U.S.
2. fruit/veggie intake throughout the U.S.
3. Fast food restaurant geolocations in the U.S.