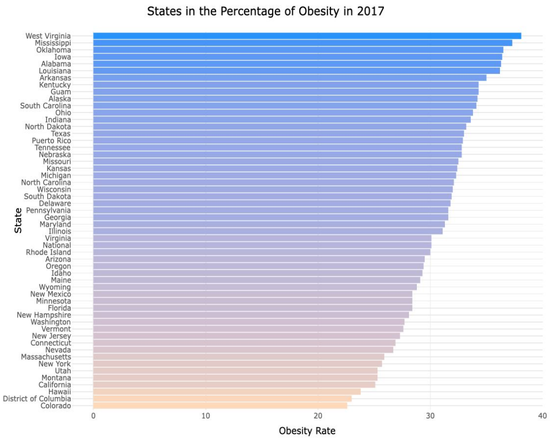
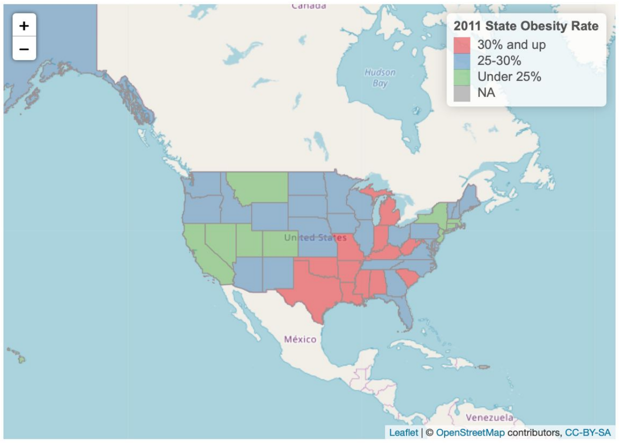
Obesity and Nutrition in the US

Process Notebook

By Qinyao Xia, Qinyue Yu, Allison Jiang





fat
kill
risk
death
loss
hard
cancer
blame
lack
wrong
flaw
shit
sever
addict
object
overweight
outcast
lost
disgust
murder
perish
backwood
weird
odd
depress
stress
limit
rampant
assault
outlaw
rant
reject
sick
scandal
ruin
bad
cheat
lose
drain
burn
critic
bitch
deceit
fuck
error
joke
worst
alarm
mess
strict
cold
hang
broken
worsen
lemon
dumb
hate
fail
vomit
fume
foul
shock
pain
die
stupid
rage
insult
threat
wild
chronic
devil
complain
morbid
poor
dark
grief
freak
slow
stigma
awkward
fear
twist
complex
disdain
poison

Introduction

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 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 analyze positive and negative sentiments using Twitter data.

Initial Project Plan: Data

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:

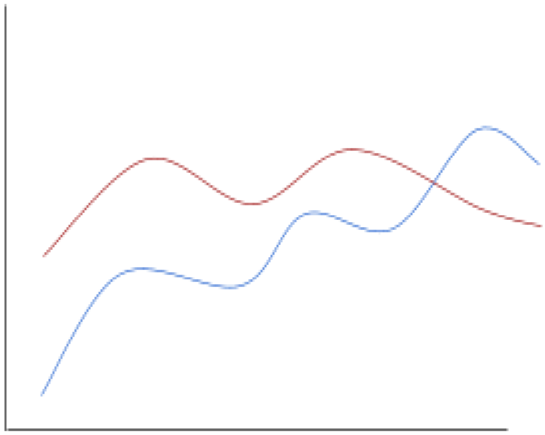
We extract text data on individuals' thoughts and opinions on fast food, obesity and diet and using text cloud to visualize what are the key determinants people believe that could be correlated with the obesity problem in the U.S. In addition, we analyze the positive and negative sentiments towards fast food. The key purpose of the twitter analysis is to see the degree of which tweets that people are posting reflect official figures on the propensity to improve nutritional intake as well as exercise as predictors of obesity rates across states.

Initial Project Plan: Plots

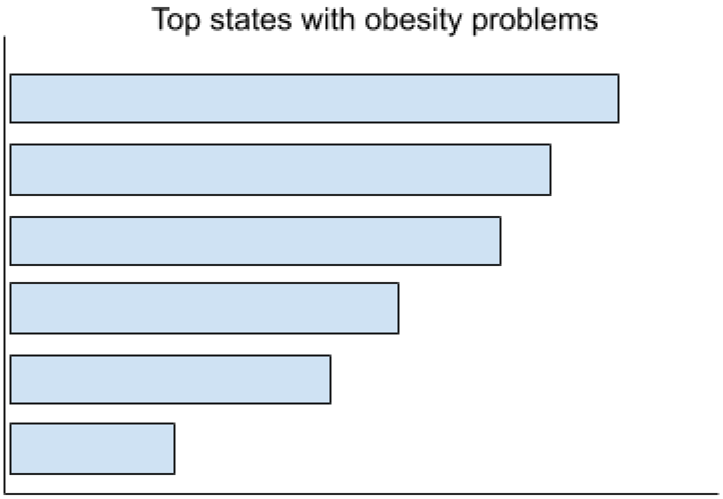
We initially decide to apply:

Basic Data Exploration

1. **Line graphs of obesity rate over time by gender, age, education and income.** This is to explore the basic demographics and characteristics of people who are obese. This will allow us to see if there is a systematic pattern surrounding people with a higher propensity to be obese (hypothesize that they are low income and low-educated Americans).



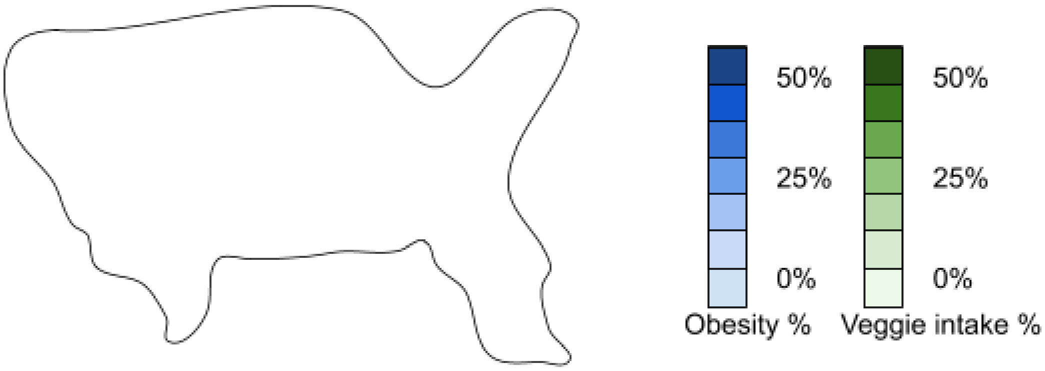
2. **Barplot of top states with obesity problems.** This plot shows the states with the highest occurrence of obesity.



Maps of Obesity & Fast food

3. Maps of obesity/nutrition/fast foods in the US

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



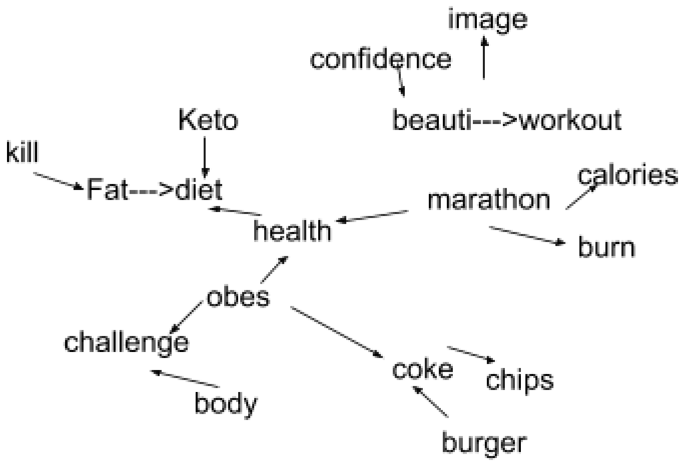
1. Obesity problem intensity throughout the U.S.
2. fruit/veggie intake throughout the U.S.
3. Fast food restaurant geolocations in the U.S.

Text Analysis

4. Word clouds of tweets. The word cloud presents a general picture of the highest occuring words in the corpus of text. We would that that these are words related to obesity and diet.

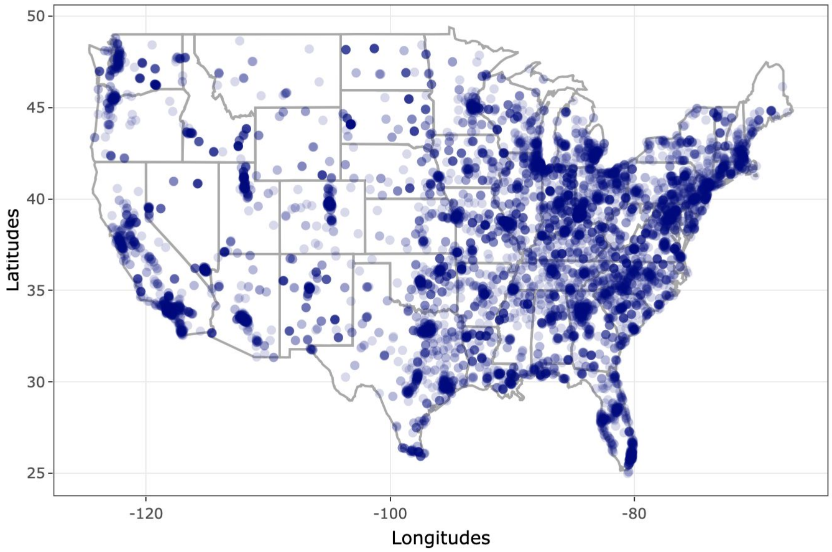
obesity
Diet
Fried
Chicken
leisure-time
FOOD
Health
Concern

5. Network graph representing network of bigrams in obesity related tweets. The network graph of bigrams display the relationship of words. Take for instance the phrase “childhood obesity”. The phrase loses its meaning when the words are individually tokenized into “childhood” and “obesity”. Therefore, it is necessary that we explore compound words with bigram analysis.



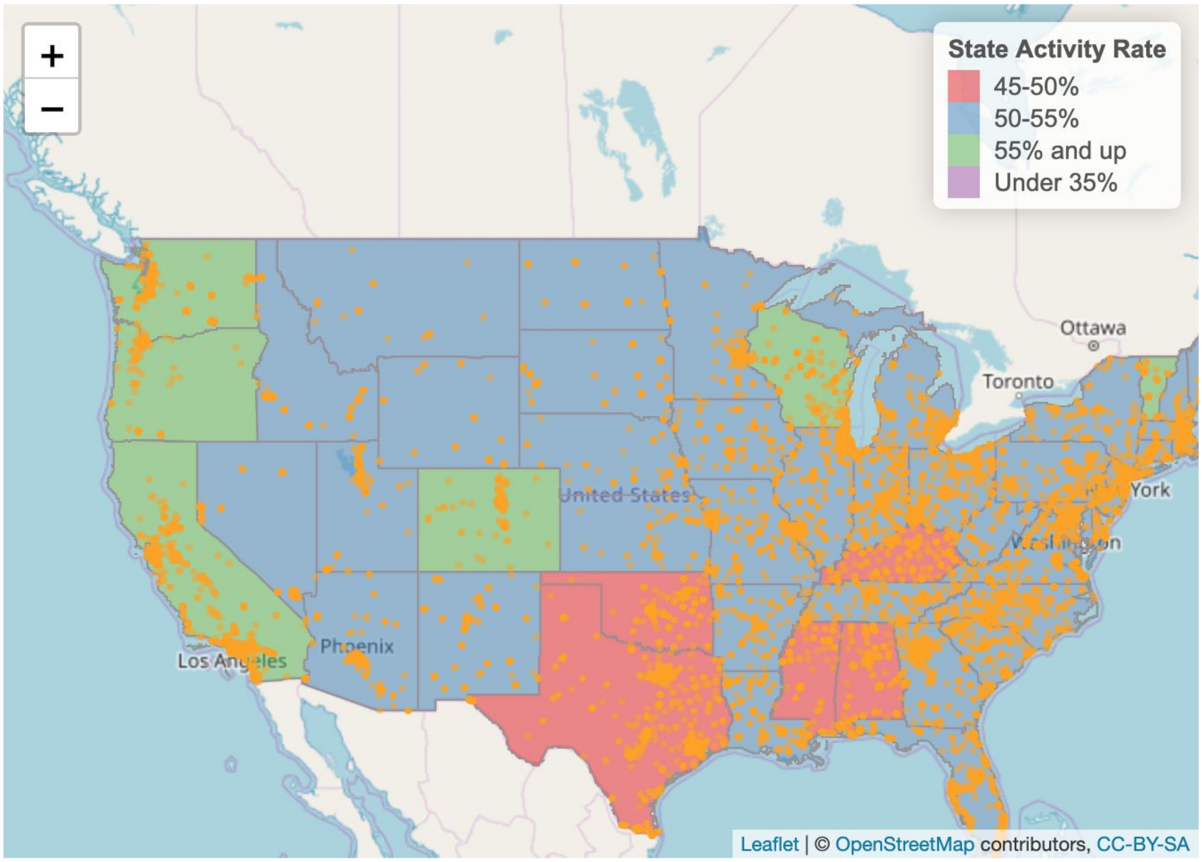
Choosing between leaflet and ggplot

ggplot graph



Using ggplot can generate beautiful graph that are clear and much easier for configuration with website publishing tools. However, there is a lack of interactivity for the visualization. To show the at the level of detail that one can observe the fast food distribution on U.S. contiguous states, we have to keep Alaska, Hawaii and Puerto Rico outside the mapping exercise.

Leaflet, on the other hand, provides interactivity for one to zoom in and out to check each state's stat on fast food, obesity and others. The effect looks like below:



We choose to keep both. For the graphic representation of fast food restaurant, we use ggplot so that the picture zooms in more and shows the states. For the graph with both state level info on obesity, activity or nutrition intake and geolocation for the fast food restaurant, we use leaflet to incorporate more complex info and allow user to zoom in and out to check various states.

Choosing Rpubs for presentation

Since we are using leaflet for mapping activity, we choose to publish our website on Rpubs, which provides better compatibility with leaflet than shiny.

Project link

Our project site is up at:

http://rpubs.com/QMSSnutrition/494313

Final data sources used

- Fast food Data https://www.kaggle.com/datafiniti/fast-food-restaurants
- CDC data https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html
- Twitter API

Conclusions

Our data visualization allow us to conclude as the following:

1. Obesity rate in the US increases from 2011 to 2017, and it still has trend to go up.
2. Male obesity rate is higher than female; obesity rate has negative relationship with education and income; middle-age groups from 35 to 64 have the highest obesity rate, while the young adults with age from 18 to 24 have the lowest obesity rate.
3. States with high physical activities participation rate tend to have low obesity rate.
4. The distribution of the fast food Restaurant is disproportionately denser in the East and West coast than Mid America, which is understandable as coastal area has higher population density.
5. Throughout the past 6 years, the number of U.S. states with less than 25% obesity shrink from 9 states to only Colorado left. Colorado is also the state with the highest exercise rate.
6. Puerto Rico does not have info regarding the fast food distribution. However, it has high obesity rate, low exercise rate, and have the highest portion of its residents eating vegetable less than once daily.

7. The text analysis primarily shows that people from states with low obesity tend to be more cautious of their lifestyle and diet, as seen through the higher occurrence of words like “diet” as opposed to people from states with high obesity.
8. Most of the words related to obesity seem to be about the adverse health complications of obesity, and words that reflect rising obesity in the U.S. However, there seems to be a lack of conversation on how to reduce obesity.