

# DSO 545: Statistical Computing and Data Visualization Final Project

Obesity & the Fast Food Restaurant Industry



Dani Cheng

Mario Guerrero

Jade Stevens

# **DSO: 545 Statistical Computing and Data Visualization**

Team 8: Jade Stevens, Mario Guerrero, Dani Cheng

Spring 2018

## **Executive Summary**

For years obesity has been a growing concern in the United States. Obesity rates have grown considerably and is now being treated as a nationwide epidemic. There are many causes that contribute to this nationwide problem, including higher poverty rates, dual income households (both parents work full-time), low income per capita, technology replacing exercise in children, and the increase of social media availability

For our project, we wanted to explore the relationship between fast food restaurants and associated rates of obesity in that area. Many of us have instinctively known that fast food is inherently unhealthy, but this critical examination will enable us to support and confirm our suspicions with concrete data.

We believe that there exists a positive correlation between fast food restaurant density and average BMI in the associated area. We would like create a heat map for the purpose of observing whether obesity occurs within areas that are saturated with a high number of fast food restaurants.

In our report, we lay out our findings and what it could potentially mean for the future. Based on our analysis we determined that there is somewhat of a correlation between the location of the fast food restaurants and the high percentage of obesity in each state.

## **Introduction**

Obesity in the United States has been increasingly cited as a major health issue in recent decades, resulting in diseases such as coronary heart disease that lead to mortality. While many industrialized countries have experienced similar increases, obesity rates in the United States are the highest in the world.

Obesity has continued to grow within the United States. Two of every three American men are considered to be overweight or obese, but the rates for women are far higher. The United States contains one of the highest percentage of obese people in the world. An obese person in America incurs an average of \$1,429 more in medical expenses annually. Approximately \$147 billion is spent in added medical expenses per year within the United States. This number is suspected to increase approximately \$1.24 billion per year until the year 2030.

Obesity has been cited as a contributing factor to approximately 100,000–400,000 deaths in the United States per year and has increased health care use and expenditures, costing society an estimated \$117 billion in direct (preventive, diagnostic, and treatment services related to weight) and indirect (absenteeism, loss of future earnings due to premature death) costs. This exceeds health care costs associated with smoking and accounts for 6% to 12% of national health care expenditures in the United States.

Fast food has been cited as a contributing factor to obesity in the United States by various sources over the years. Around one third of children aged 4 to 19 eat fast food every day in the U.S.

Many popular American foods, including hamburgers, french fries, bacon, and doughnuts, are high in fat and carbohydrate content. These food items are also relatively inexpensive and available at fast food chains across the country (for ex. value menu). These convenience foods are frequently fried and are high in calories. Consumption of foods exceedingly high in fat calories can lead to obesity. Many menu items at fast food establishments do not fulfill basic nutritional requirements of a healthy meal.

Fast food chains and restaurants have experienced improved sales with larger portion sizes. Research cited by the Center for Disease Control estimates restaurant meal sizes to be four times larger than they were in the 1950s.

There are four basic reasons why fast food is linked to obesity:

- First, unhealthy ingredients. Most of the fast food contains a large amount of sugar, fats and carbs and less minerals and vitamins. This means that you are taking in large amount of unhealthy calories in the shape of fast food which leads to weight gain and ultimately obesity. Most of the fast foods have exceeding levels of sugar and fats which are directly associated with increasing weight.
- Second, larger portions. The unhealthy ingredients of the fast food are further aggravated with increased portion sizes which have grown parallel with the average body weight of a person from the 70s to date. While the portions have become large, the person will still eat the complete meal regardless of feeling full or not. This means that people are eating bigger portions of junk food leading to abnormal weight gain.
- Third, lower cost. One of the main reasons people cannot stop eating fast food is the low costs it is available in. As per a research conducted by the University of Washington, a diet containing 2000 calories of fast food costs much less than a diet with 2000 calories of healthy food. This makes it more affordable than going for healthy food and is a leading cause of obesity in lower-income classes.
- Fourth, convenience. Fast food restaurants offer convenience. You can always find one in close proximity of your home and can also get food easily delivered to your

house. This makes it a convenient option than making food at home using healthier ingredients.

According to studies, the consumption of fast food has increased over the past four decades at an alarming rate. This increased calorie intake is a leading cause of obesity in the USA. If it is not controlled, health issues are going to rise in the coming years as more people will become obese eating unhealthy fast food.

Based on the available information, we will explore available datasets to determine the relationship if any between the location of fast food restaurants and the associated rates of obesity in the given area.

## Data Description

To determine what the best approach would be we wanted to find public data that could show us where all fast food restaurants are located in United States.

### BMI Variables

Variable	Description
Year	Year
Location	State Location
BMI	Body Mass Index (BMI)
Lon	Longitude Coordinates
Lat	Latitude Coordinates

### Fast Food Restaurant Variables

Variable	Description
ID	Observer ID
Address	Address of restaurant
City	City of restaurant
Abbrv	Abbreviation of State
State	State
Name	Name of Restaurant

Lon	Longitude Coordinate
Lat	Latitude Coordinates
Postal Code	Zip Code

## Data Cleaning

Our team analyzed data from two separate data sources to draw correlations and validate our hypothesis.

Our first data set, supplied by Kaggle, is a random sampling of 10,000 fast food restaurants across the United States which provided insights into the distribution of restaurants by zip code and state. Our second data source examined BMI by state and year across a wide array of socioeconomic, demographic, and psychographic metrics, including age, gender, income, and frequency/quality of exercise. Both of these data sources were first cleaned in Excel before being imported into R.

Relatively little cleanup was needed in the Kaggle fast food data set. No rows were removed, but some column headers were removed due to redundancy or blank fields to make the final data set more presentable when loaded into R.

For our obesity dataset, we did a bit more work. While there was a wealth of information about BMI across different age groups, fitness and income levels, and gender, we didn't want to make any heavy-handed assumptions about the types of people that frequent fast food restaurants, so we took the most generalized comparison for the most recent year. This reduced our obesity dataset to 50 rows that compared overall BMI by state in 2016. We removed extraneous dimensions and columns that didn't benefit in our data analysis, and also excluded rows that analyzed BMI on more granular metrics, such as age range, gender, or fitness level.

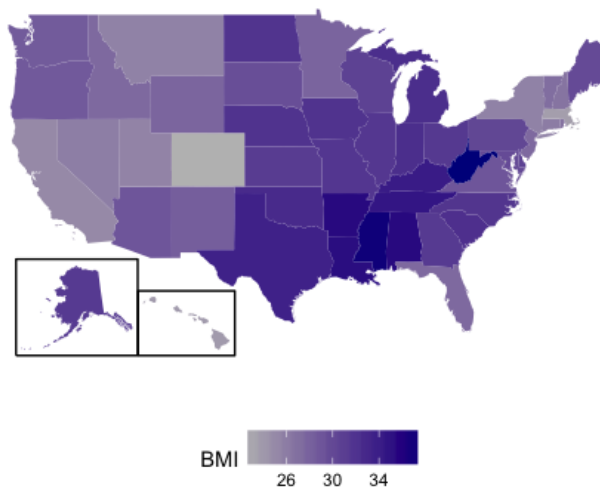
To join our two data sets, we rolled up zip codes in the fast food dataset to the state level. We then used state as our primary key to join the two data sets. This allowed us to directly compare number of fast food chains to BMI in each state. We represented fast food locations by zip code with dots to show their distribution across the US, and colored the dots along a gradient, with lighter dots representing higher BMI.

## Exploratory Data Analysis:

We first wanted to see how the data looked for each data set as a stand alone.

## Average BMI Across the State

```
new_obese = obese_map %>%  
  separate(col = Locator,  
    into = c("lon", "lat"),  
    sep=",",  
    remove = T)  
  
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 1 rows [11].  
  
new_obese$LocationDesc = tolower(new_obese$LocationDesc)
```



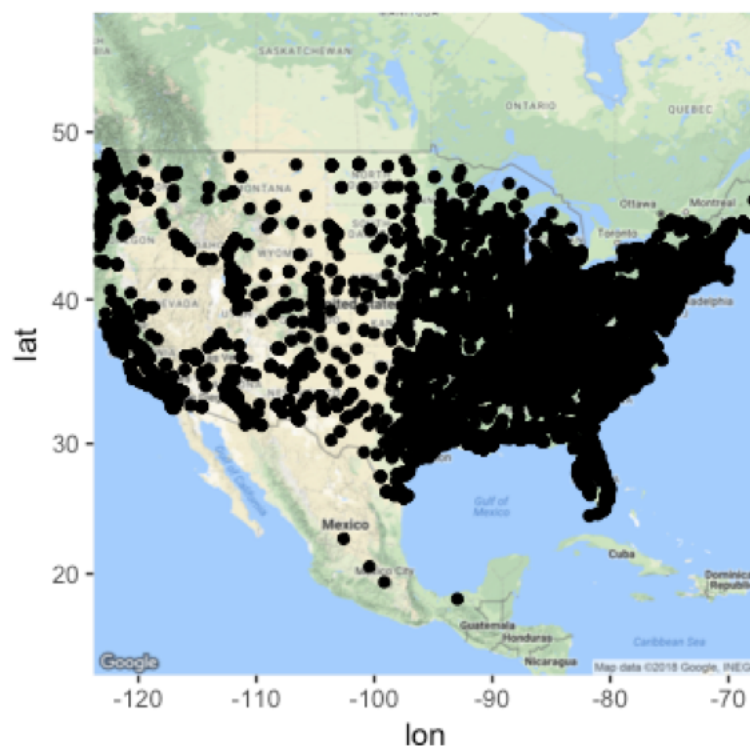
Analysis of the obesity dataset shows us the distribution of overweight persons by BMI in the United States. Darker states represent a higher average BMI, which suggests that the Midwestern states have a higher concentration of obesity than the western states or the Northeastern states, especially New York and Massachusetts.

```
fastfood_map = read.csv("fastfoodtest.csv")  
obese_map = read.csv("scrub_obese_test.csv")  
obese_map$LocationDesc = tolower(obese_map$LocationDesc)  
fastfood_map$State = tolower(fastfood_map$State)  
  
test_map = full_join(fastfood_map, obese_map,  
  by = c("state" = "LocationDesc"))  
usa_center = as.numeric(geocode("United States"))  
  
## Information from URL :  
http://maps.googleapis.com/maps/api/geocode/json?address=United%20States&sensor=false  
  
USAMap = ggmap(get_googlemap(center=usa_center,  
  scale=2,  
  zoom=4))
```

```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=37.09024,-95.712891&zoom=4&size=640x640&scale=2&maptype=terrain&sensor=false
```

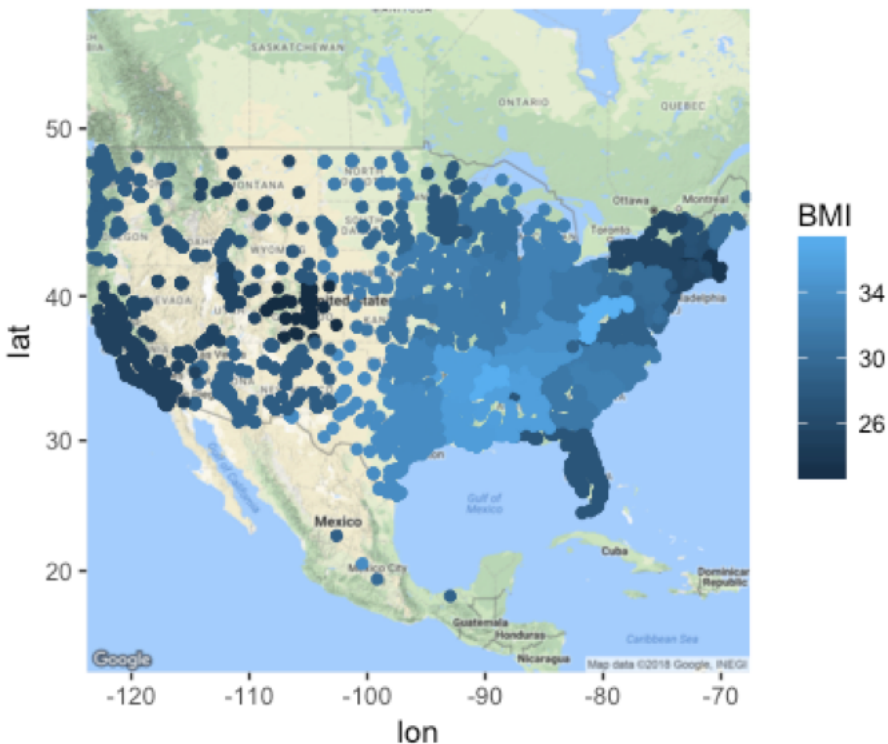
```
USAMap +  
  geom_point(aes(x=lon, y=lat),  
    data=test_map) +  
  scale_fill_gradient(low = "white",  
    high = "darkblue") +  
  ggtitle("NEEDS TITLE")
```

```
## Warning: Removed 123 rows containing missing values (geom_point).
```



The fast food dataset shows us the distribution of fast food restaurants by zip code. Based on the distribution, we can see that most of the fast food chains are located along the eastern United States, and become more sparse heading west until reaching the West Coast. Given our hypothesis, we would expect a positive correlation between fast food chain density and average BMI.

## Dashboard



When we combine the two datasets, we can see that our hypothesis is mostly confirmed. The higher concentration of fast food chains in the Midwest heading eastward correlates with the higher average BMI. This suggests that the obesity epidemic is somewhat spurred on by the influx of fast food restaurants. A couple anomalies exist, however - North Dakota also has high rates of obesity, but a relatively scattered distribution of fast food restaurants. The opposite holds true for the west coast, where there are high concentrations of fast food chains but lower average BMI. This could potentially be explained by health trends in California, Oregon, and Washington, where the typical consumer tends to be more inclined toward organic and wholesome foods rather than fast food.

## Conclusion

Based on our analysis of the data provided, we can see that there is indeed somewhat of a correlation between the number of fast food restaurants in a community and percentage of individuals that are considered obese in that area. Proximity and low-cost food items make fast food restaurants a convenient choice for adults, when the alternative of a grocery store may not be as close. It would be interesting to also map the location of grocery stores in each community to see if that may also be a contributing factor to BMI.

It should of course be noted that there other factors that go into an individual's BMI. Considering external factors only, we know that income, education, transportation are



just a few of the factors that play a part in influencing an individual's behavior towards health and food choices.

Furthermore, we should not be fully reliant on just the number of fast food restaurants as being the sole contributors to an individual's behavior. We should also consider the type of restaurant and their menu items.

Lastly, as mentioned earlier this dataset does not fully represent all fast food restaurants in the country. The map is clearly lacking fast food restaurants located from the Midwest to the West Coast.

In order for this analysis to be fully conclusive, we would need all of these concerns addressed so that we can give the best representative. Looking towards the future, we look forward to gathering additional data points and doing an even more deep analysis of the information to determine if there is indeed a true relationship between the number of restaurants and a community's average BMI.