

Process Book

Martin Reindl - Andre Souffrant

Background and Motivation

Our project aims to map obesity rates in the United States to the prevalence of major US fast food chains (McDonalds, Burger King, Dairy Queen, Starbucks etc.). Obesity and weight issues are some of the biggest health problems currently facing the United States. Roughly 30% of the US population is overweight ($25 < \text{BMI} < 29.9$) and another 35% of the US population is obese ($\text{BMI} > 30$) according to the Center for Disease Control. These conditions significantly increase the likelihood of major medical conditions, such as stroke, cancer and diabetes, which cost the United States about \$190 billion in 2014 according to [a recent Cornell study](#).

Fast food consumption has often been linked to obesity, and is commonly accepted as unhealthy. The United States has a very high per capita intake of fast foods, amounting to [11.3% of daily caloric intake](#). A visualization that allows users to explore a possible positive correlation between the geographic locations of fast food restaurants and measures of poor health will allow those users to have a deeper understanding of the potential negative effects of fast food on health.

Although none of us have worked with population data on health measures and fast food consumption before, we all feel highly prepared for this project given our lifelong personal consumption of fast food. Additionally, we chose to focus on this topic for the following reasons:

- It is easy to understand and highly accessible to most users
- It is interesting and relates to a topic with which we are all well acquainted
- Easily available data exists through the Kaiser Family Foundation and the Real World Data Series (see section on Data for more information)
- Visualizing the impact of fast food consumption on health can help policy makers make smart decisions.

Project Objectives and Questions

The primary questions we hope to answer with our project are:

- Does the quantity of fast food restaurants in a state correlate with measures of poor health in that state?
- Are there geographical trends in the measures of poor health (e.g. North vs South)?
- Are there geographical trends in the quantity of certain fast food chains?

We hope to learn the answers to these questions by means of an interactive, fun, and clear visualization. Ideally this visualization would accomplish the following:

- Get people thinking about the impact of fast food on their lives and the United States as a whole.
- Make a fun visualization with which people enjoy interacting

- Allow policy makers to easily visualize the prevalence of fast food consumption and poor health in the United States.
- Learn some D3 and visualization in the process!

Data and Data Processing

We specifically conceived our project in a way that requires relatively little data collection, manipulating and cleaning. Data on health and obesity is collected regularly by the Kaiser Family Foundation. The data is readily available in many formats, and we have found [csv files](#) for the year 2013, which we plan on using. We were able to get the percentage of overweight and obese people, the percentage of people with cardiovascular disease, the percentage of people with diabetes, and the percentage of people diagnosed with mental health problems for each state from this dataset.

The United States Census Bureau regularly collects [population estimates](#). We will be using their population data to get per capita fast food restaurants for each state (a more accurate reflection of fast food consumption than absolute quantities of franchises). This data will require a small amount of cleaning and filtering.

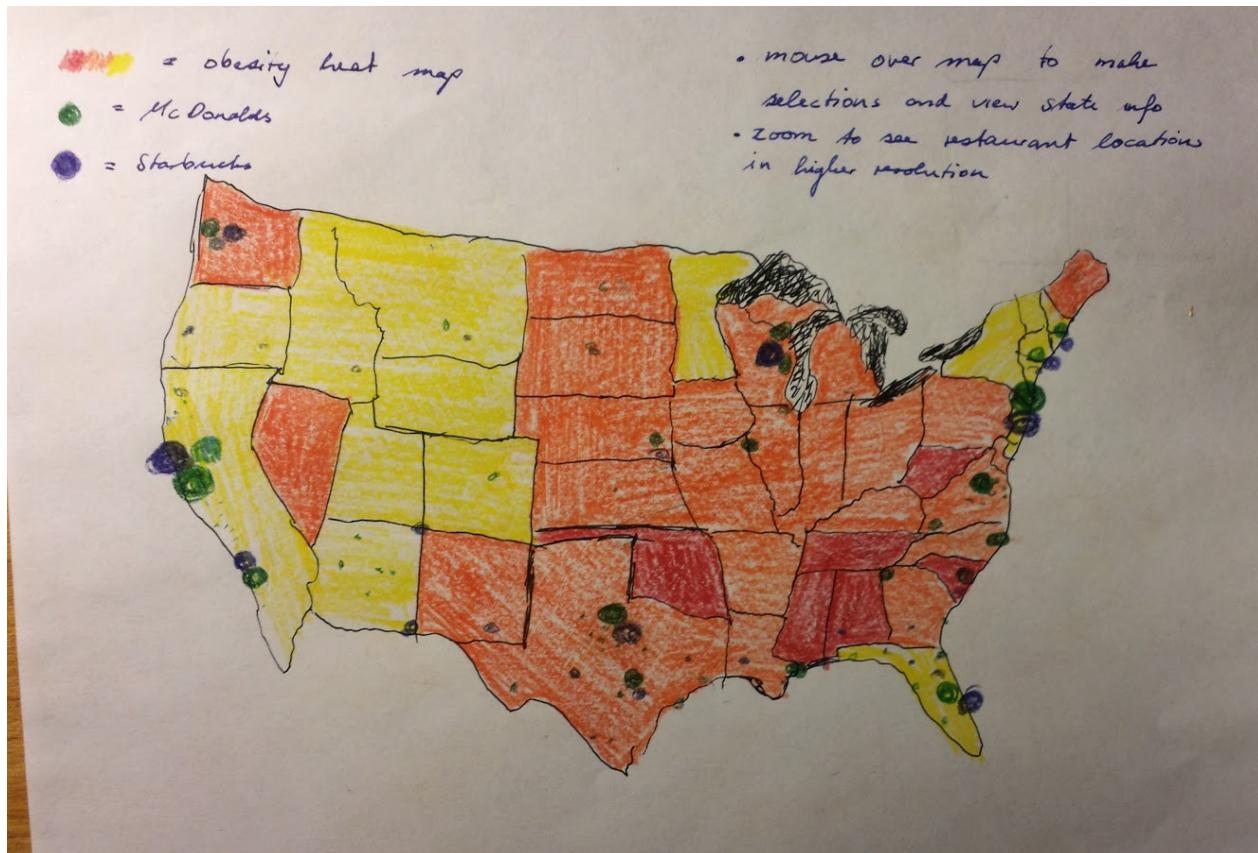
The [Real World Data Series](#) has comprehensive lists of all franchise locations for multiple major American fast food chains and is updated regularly. These csv files consist of latitude/longitude values and an address for each franchise restaurant. We parsed the state from this address using python.

Using the states parsed from each restaurants address, we were able to calculate the total number of fast food restaurants in each state. We combined this total number with health measure information in a single data structure that stores all relevant information about each state. We have another data structure that stores location information for each restaurant.

The last data structure we are working with is the data structure that corresponds to the map visualization. The implementation of the map was created by Mike Bostock, and the [source file](#) can be found here.

Visualization

At heart our visualization will consist of a map of all fifty states. We will allow users to toggle multiple views of this map and will generate smooth, non-distracting transitions between views. The map will have several features. We will show health information by heat mapping. In addition to the health information, we also want to add information about the location of fast food restaurants in the United States. We will thus treat the map as a scatter plot and add dots for every location of a fast food restaurant of one of the four major chains (Starbucks, McDonalds, Burger King, Dairy Queen) onto the map. Ideally this will allow users to view and correlate the number of fast food restaurants with health data. An initial sketch of the central map is shown below.

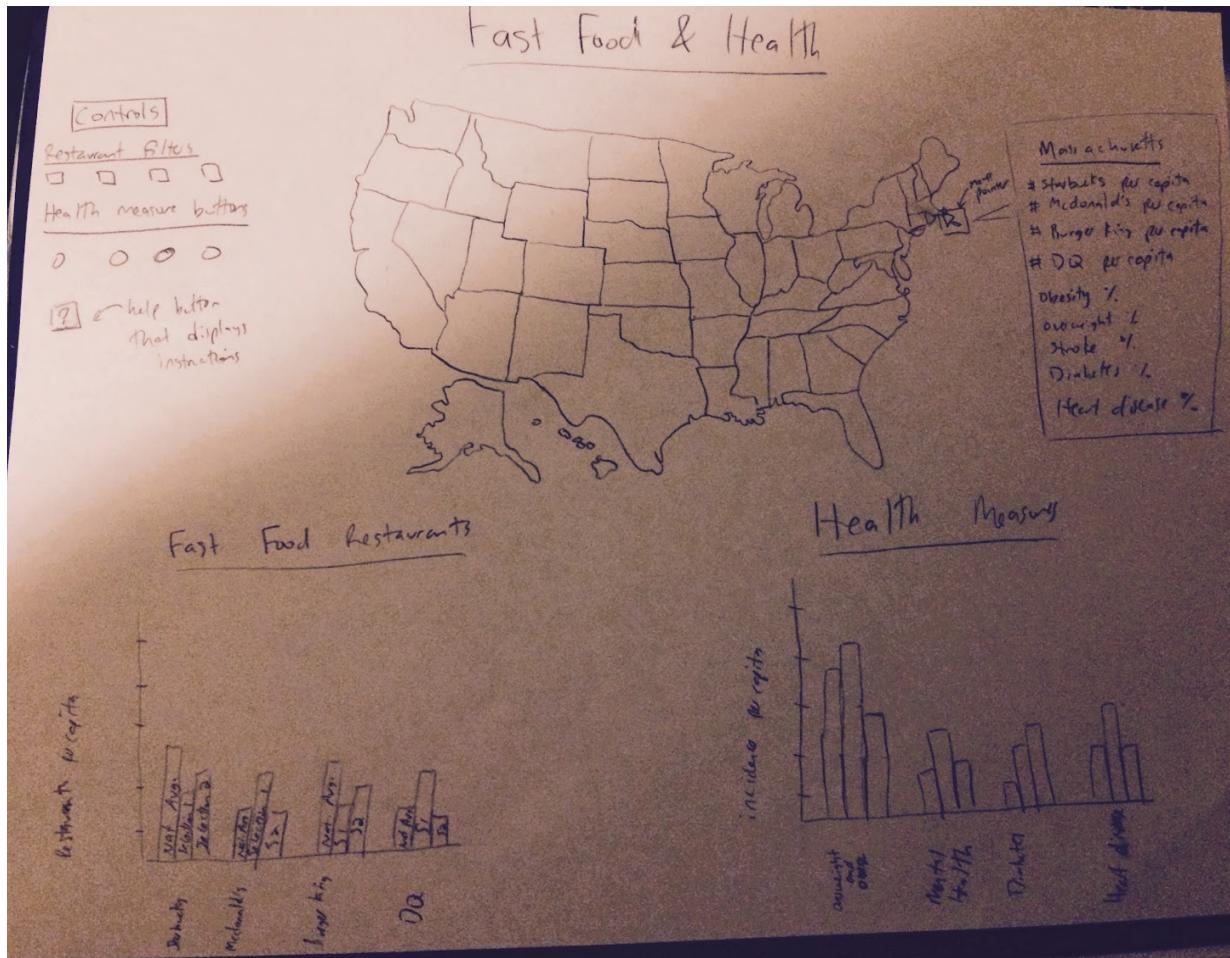


In addition to this central map, we are planning on a number of other features that allow users to better quantify the data we are presenting on our map.

Specifically, we are going to add two bar graphs below our visualization. These bar graphs will redundantly encode the information on the map, but they will encode the information in a way that is easier for the viewer to compare and quantify. In this way, health data will not be represented by color (which is very hard to accurately compare), but rather by the length of a bar (which is relatively easy to quantify and compare). We will generate two bar graphs, one for health data and one for restaurant data. Users will be able to make selections on the map, and these selections will be shown in our bar graphs below the map.

Other than these bar graphs, we also plan to implement a small pop-up that comes up when a user hovers over a state. This pop-up should give the actual number of restaurants in a state, some basic health facts, and metadata, such as the state's name.

A sketch of the above description is shown below.



We then started work on the project milestone, and implemented a first version of our map. This map followed an [example](#) by Mike Bostock and has a simple zoom that allows viewers to zoom in onto a state by clicking the map.



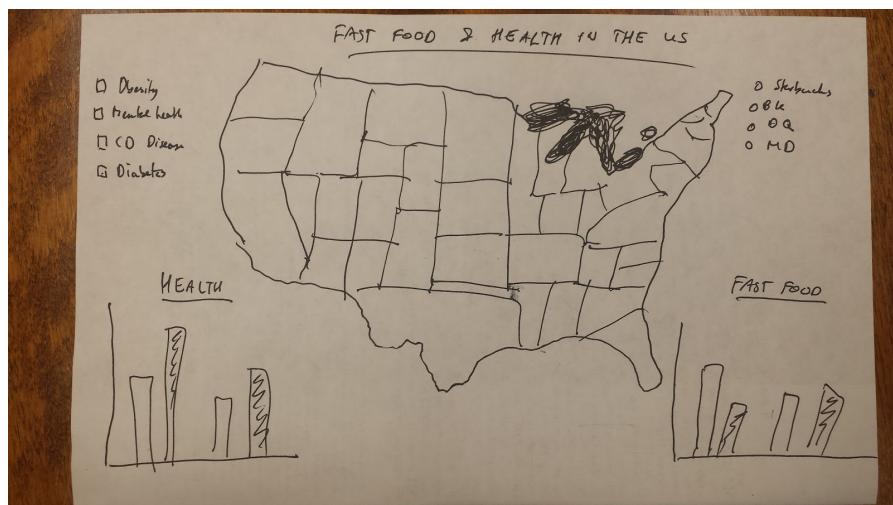
There are some problems with this map:

- It does not take into account the size of states well, so large states like Texas are cut off in visualization (see below). A future implementation will thus have to take into account the size of states to guarantee that a user will be able to find themselves on the map.

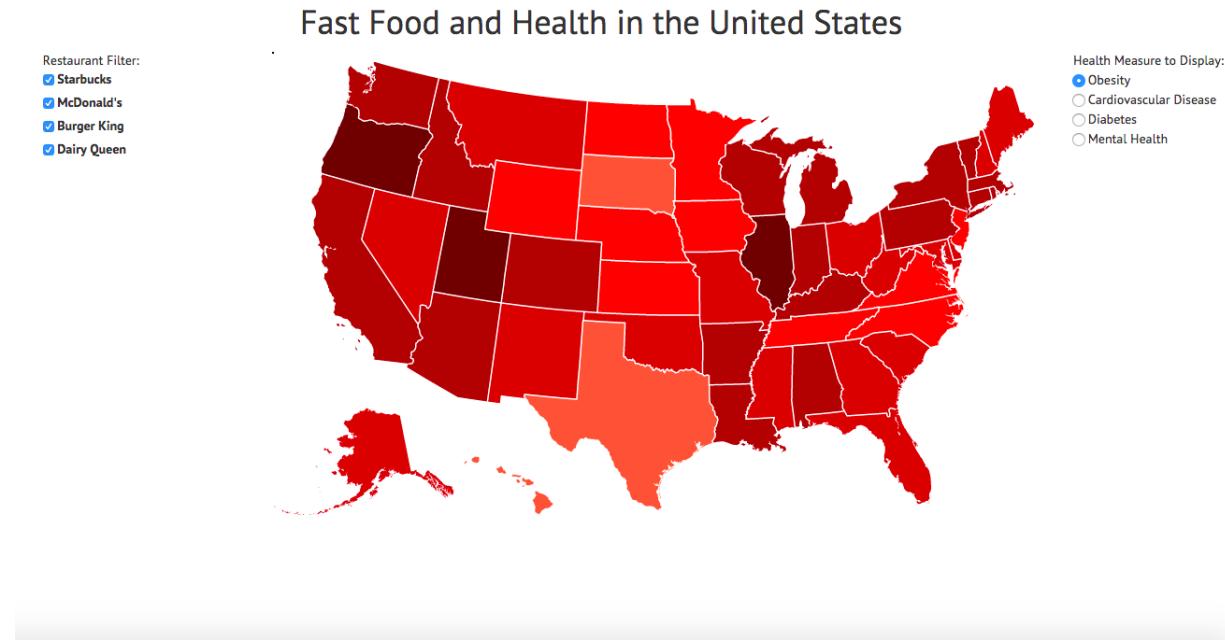


- The map was much larger than we expected. It took up a much larger portion of the screen than we had planned for and made it impossible to stick to our initial design idea of putting two bar charts below the graph. If we want to stick to the dimensions given to us by the Mike Bostock implementation we will have to make design changes. After some discussion, we agreed to move all fast food data to the left of the map and all health data to the right of the map.

Depicted below is a sketch of our new layout for the entire page. We decided to move those bar charts to the bottom corners of the page and the map down to center it under the title. The controls have been moved such that there are some filter options on the right and others on the left of the graph.



We then implemented the heat map portion of our map. The intention of the heat map is to allow users to visually see trends in the specified health measure throughout the country. We currently use a color scale that relies on colorbrewer, and we use the color of each state to represent the proportion of people who fit into the category specified by the health measure. The interactive functionality of the heat map has not been implemented, but once implemented, the user will be able to specify the health measure he or she would like to visualize with the heat map. There will be four possible selections (obesity/overweightness, diabetes, mental health, and cardiovascular disease). Upon a change in health measure, the heat map will recalibrate the scale used to determine the color.



In the photo above, we see the map with encoding information about mental health. States that are a darker shade of red have higher incidences of mental health diagnoses than states with lighter shades of red. The interactive element will be the radio buttons on the right side of the map. These buttons will allow the user to change the data to be encoded in the heat map. We still need to definitively choose the color scale that we would like to use.

We have partially implemented the scatterplot that will allow the user to see the geographical distribution of four major fast food restaurants. We currently have circles that correspond to each restaurant, but are working on making the scales work correctly to update the position of circles. At this point, all of the circles are bunched together to the bottom left of the title.