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 < BMI < 29.9) and another 35% of the US population is obese (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 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. We ultimately ended up using data directly from the CDC for the obesity levels, seeing as the Kaiser Family Foundation groups overweightness and obesity together, and we were only interested in obesity.

The United States Census Bureau regularly collects population estimates. We will be using their population data for 2013 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.

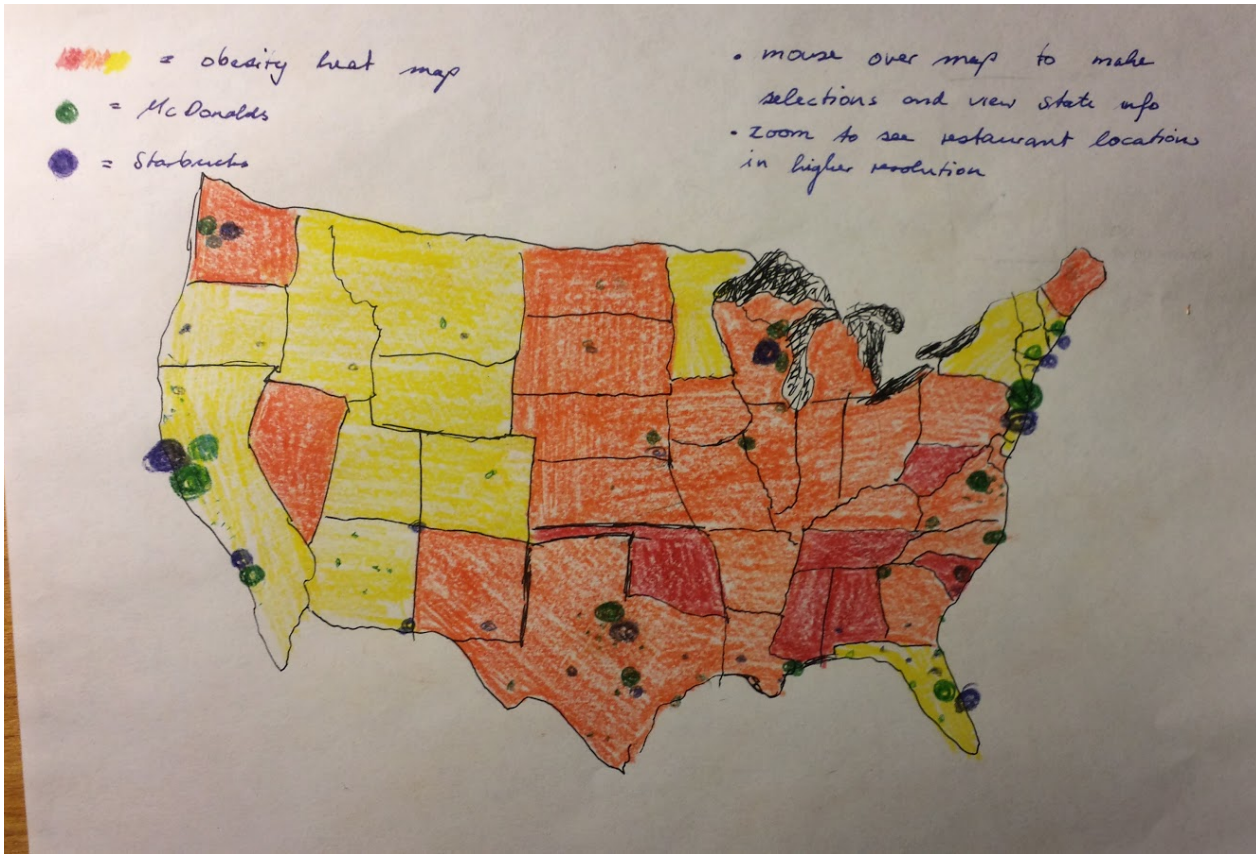
In our story for the website, we were interested in showing a trend of increasing obesity in the US over the last decade. We found time series data and implemented a simple bar chart using it.

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 and Design Evolution

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.

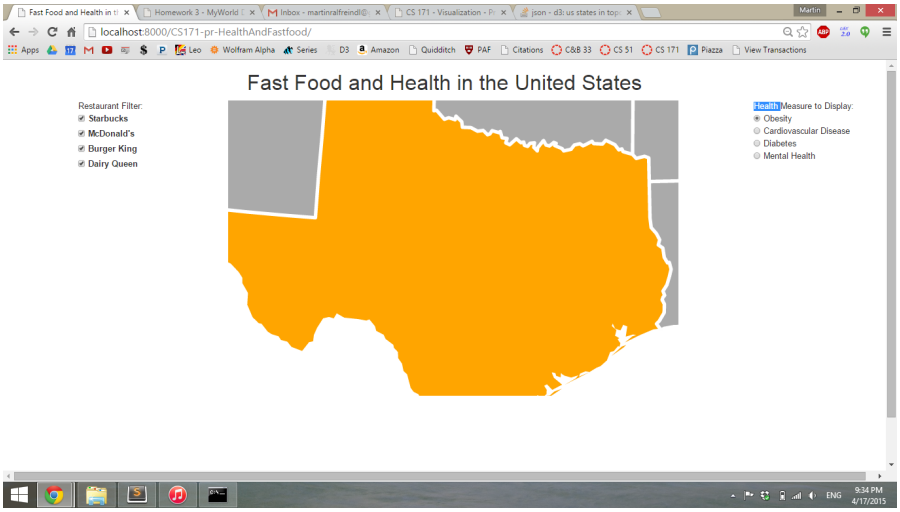
Other than these bar graphs, we also plan to implement a small pop-up that comes up when a user hovers over a state. This pop-up should give the actual number of restaurants in a state, some basic health facts, and metadata, such as the state's name.

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:

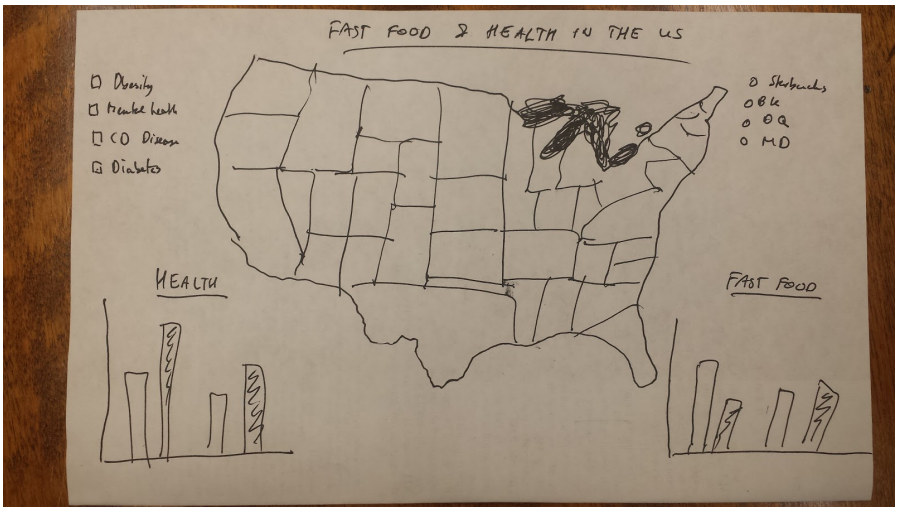
- It does not take into account the size of states well, so large states like Texas are cut off in visualization (see below). A future implementation will thus have to take into account the size of states to guarantee that a user will be able to find themselves on the map.



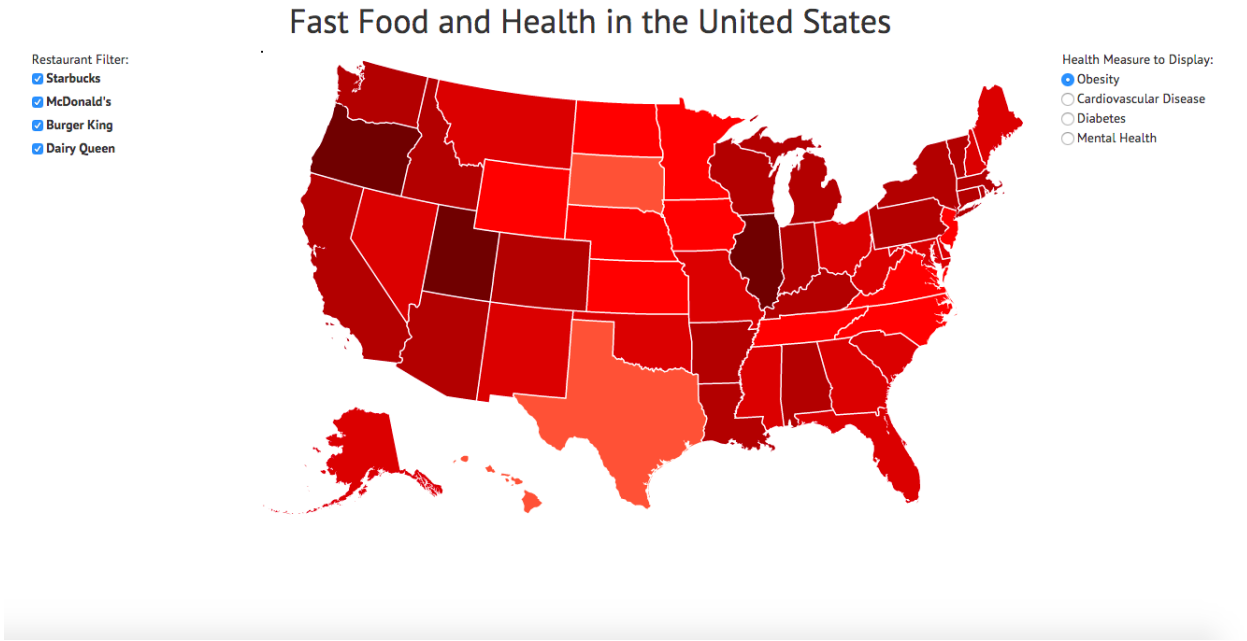
- The map was much larger than we expected. It took up a much larger portion of the screen than we had planned for and made it impossible to stick to our initial design idea of putting two bar charts below the graph. If we want to stick to the dimensions given to us by the Mike Bostock implementation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

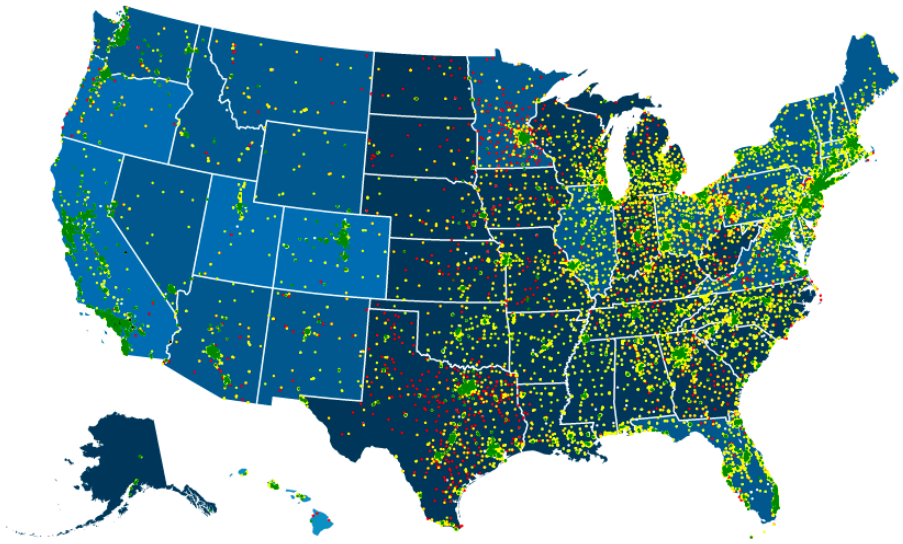


In the photo above, we see the map with encoding information about mental health. States that are a darker shade of red have higher incidences of mental health diagnoses than states with lighter shades of red. The interactive element will be the radio buttons on the right side of the map. These buttons 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.

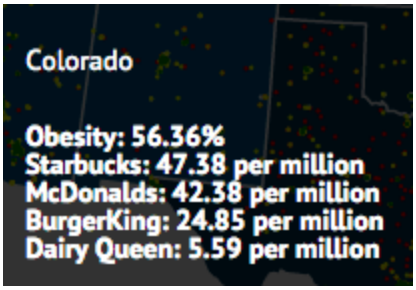
We have partially implemented the scatterplot that will allow the user to see the geographical distribution of four major fast food restaurants. We currently have circles that correspond to each restaurant, but are working on making the scales work correctly to update the position of circles. At this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

Next we focused on creating the supporting bar graphs to the left and right of our graphs. We started with the graph displaying fast food information. This graph is to be put in the lower left corner of the screen according to our current design for the visualization. During the design studio we got feedback from our peers that indicated we should adjust our measure of fast food consumption to reflect the population of a given state. We thus gathered population data from the US census bureau and used this to get the number of fast foods per capita for each chain in each state.

Once the heat map aspect of the map visualization was complete, we focused on rendering the points on the map that encode the locations of restaurants. Below is a screenshot of the map with the points rendered (the color scales are still to be determined).

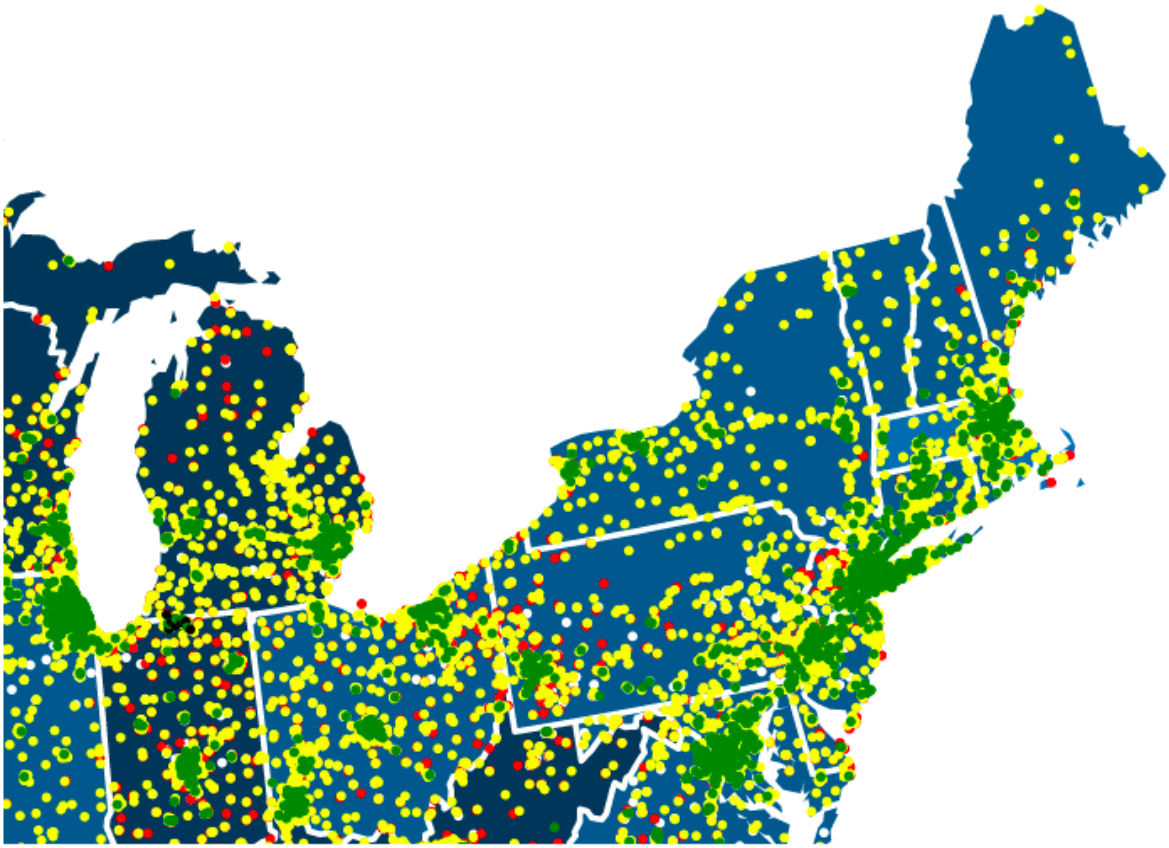


To give users the ability to probe the map visualization for more information about a specific state, we implemented a hover feature. When hovering over each state, a text box appears that gives more detailed information about the health and restaurant data associated with that state. A screenshot of one of these text boxes is shown below.



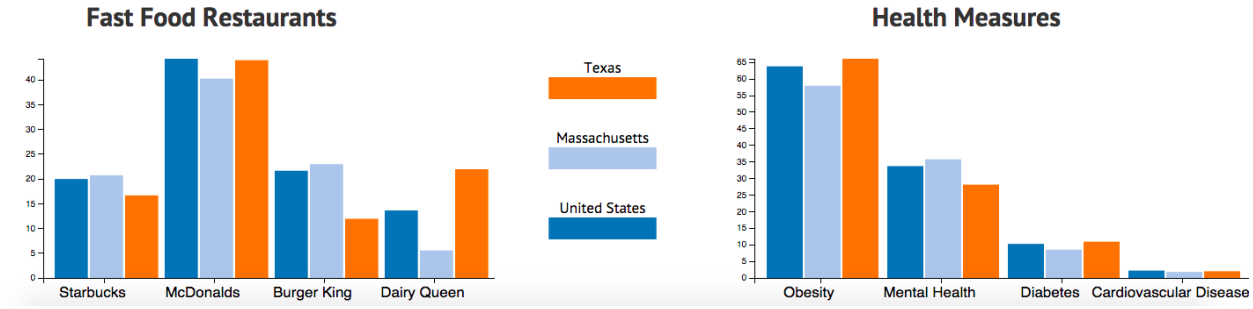
It gives the state name, the percentage of adults diagnosed in that state with the health measure that is encoded in the heat map, and the number of restaurants (for each restaurant) per million people in that state.

The goal of the combined heat map and location-based scatterplot is to allow the user to explore the relationship between measures of poor health and locations of restaurants. To allow a greater freedom in exploration, we implemented a zoom and drag feature using the d3.behavior.zoom() function. The screenshot below shows a zoomed in image of the northeast.

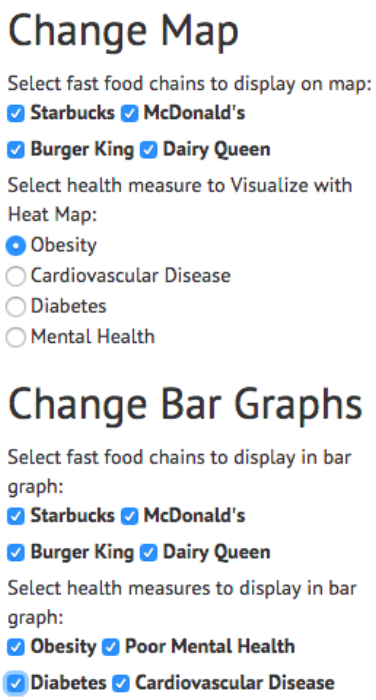


It is important to note here that this visualization does not account for population when displaying the locations of the restaurants. This means that it is difficult to get a sense of the per capita density of restaurants in this visualization.

To address these concerns, we combined the main map visualization with secondary visualizations to allow users to explore this correlation further. We decided to try to show bar charts that compare selected states against the national averages for health measures and per capita number of restaurants. In the screenshot below, there are two side-by-side bar charts. The barchart on the left compares the per capita numbers of restaurants. The barchart on the right compares the percentage of people in the state that have been diagnosed with a certain health disorder. The legend in the middle shows the colors corresponding to each state and the national average.



To allow users to explore relationships between specific health measures and specific restaurants, we implemented various options to change the visualization. We have two major categories of controls for changing the visualization. The first is to change the map visualization by changing the health measure displayed in the heat map or by filtering the points shown on the map by restaurant. The other category is to change the barcharts by filtering which health measures will be shown on one graph and which restaurants will be shown on the other graph. An image of the controls is shown below.

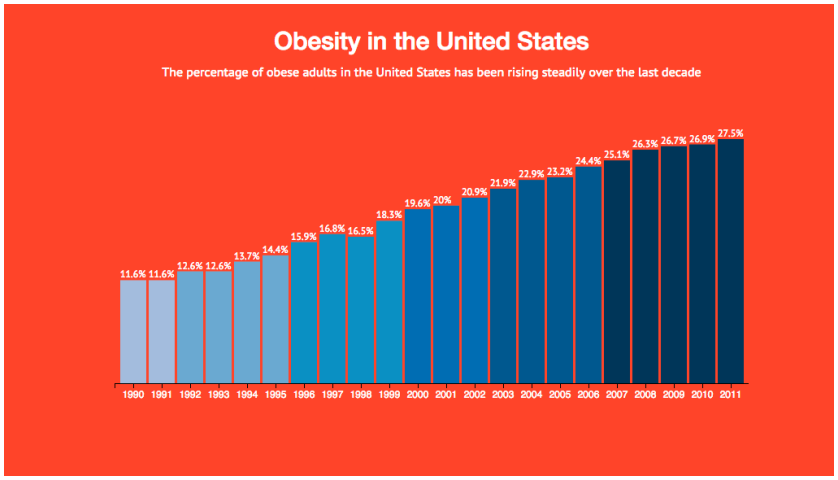


An important aspect of our visualization is the interaction between the main map visualization and the barchart visualizations. We allow the user to use the map visualization to select states to be compared using the bottom visualizations.

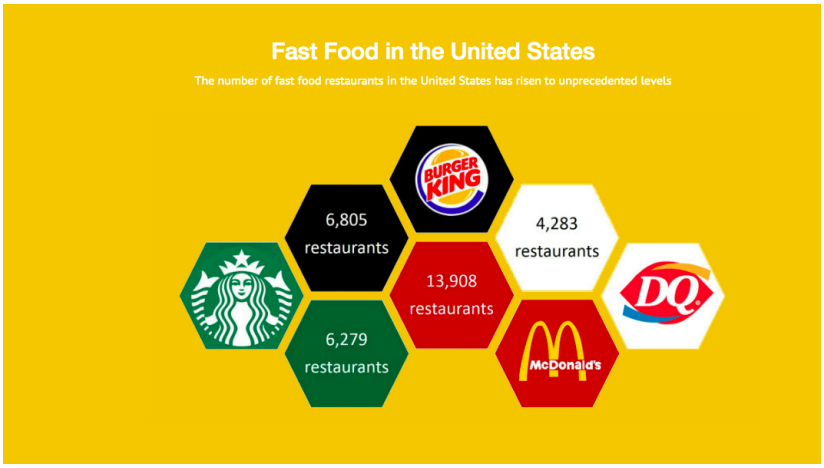
At this point we felt that we had completed the basics of the main part of our visualization, and so we started to start working on the story we wanted to tell (as requested by Mohammad during our TF-meeting). We came up with the following story.

We wanted to add a story that captured the motivation behind our visualization. To accomplish this, we wanted to do three main things. We wanted to try to show the increase in obesity in the United States over the past few years. We also wanted to connect that with an increasing number of fast food restaurants. Lastly, we wanted to show that there may be a potential connection between this and other measures of poor health.

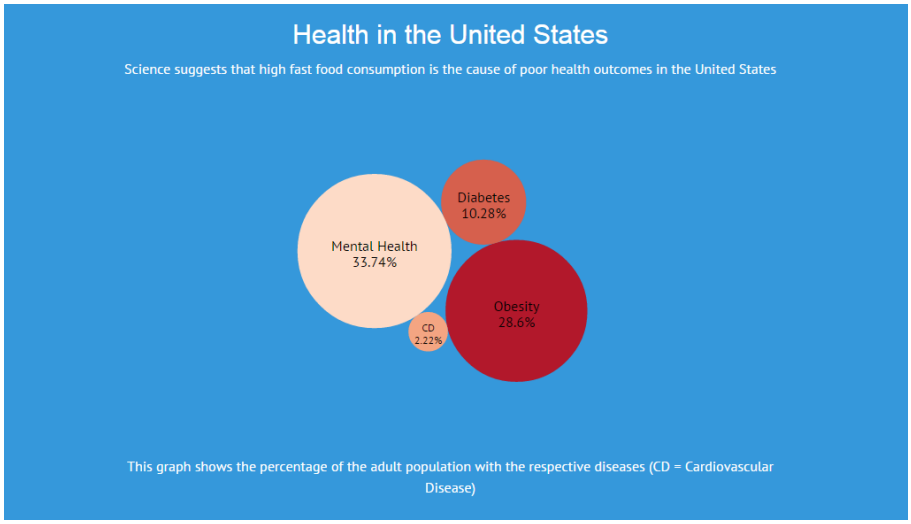
Having, come up with this idea, we started looking for extra data and found time series data for obesity. The data on Obesity allowed us to implement the visualization shown below. It is a simple bar chart that shows the percentage of adults diagnosed with obesity each year from 1990-2011. This visualization accomplished the first goal that we had for our story.



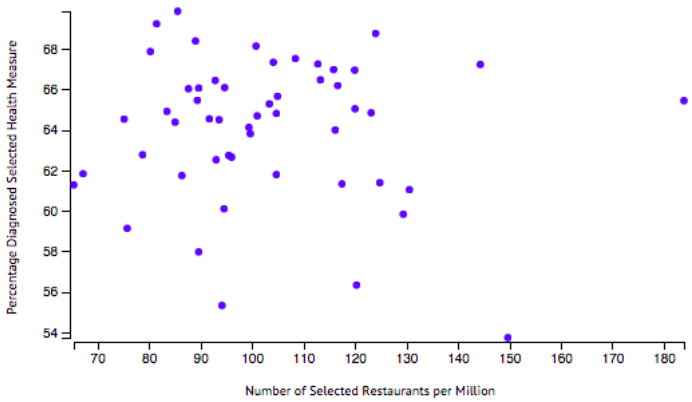
Our original dataset allowed us to create a secondary visualization that captured the number of fast food restaurants that existed when the Real World Data Series collected the data. Since we could not find a reliable source of data that had data spanning a range of time, we relied on this visualization to show the current magnitude of fast food restaurants in the country. The visualization is shown below.



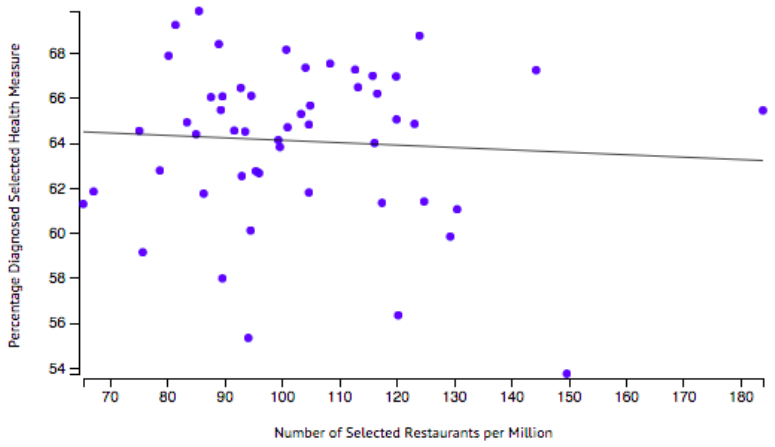
The final part of our story sought to place this information (which is likely well explored) in the context of other measures of poor health. Since we weren't able to find time series data for the health measures in our visualization, we chose to show national averages for the most recent year of CDC data, which is 2013. We implemented a force directed layout in which nodes have areas corresponding to the percent values of people affected with a given disease. The force directed layout can be dragged just because we think that's cool and it's kind of fun play with force layouts. We had major issues with overlapping and getting the text to fit into the bubbles well, but were ultimately able to overcome them with Mike Bostock's help.



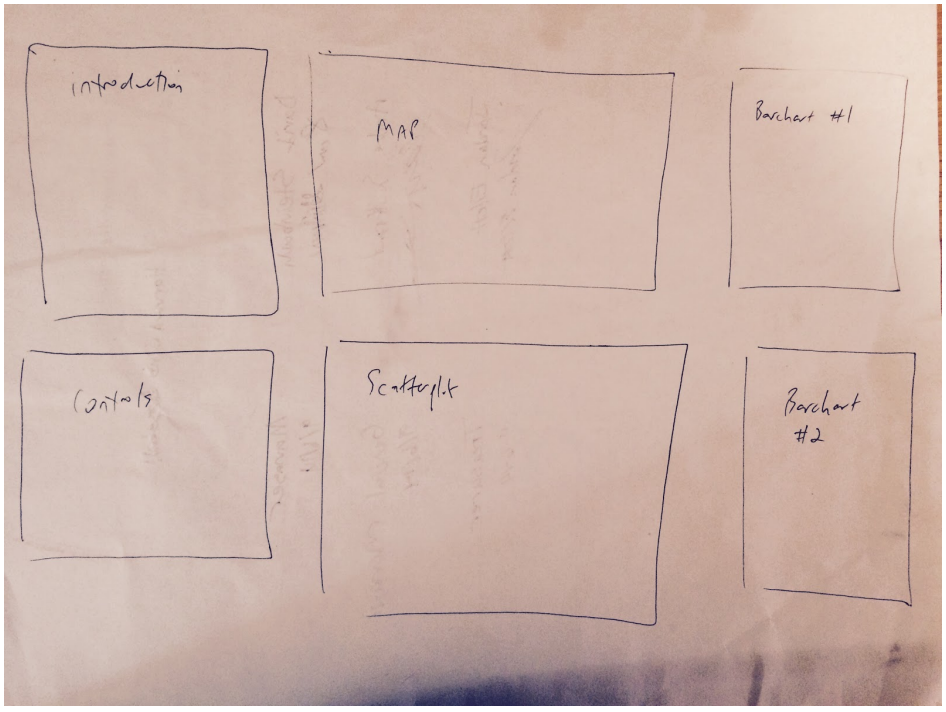
Having implemented a story for our visualization that connects back to our ultimate motivation for the project, we realized that our core visualization did not offer the functionality that we wanted it to. Our idea was to let users explore the relationships and correlations between fast food locations and measures of poor health. Since the restaurant data displayed on the map is not normalized for population, most of the concentrated areas of restaurants appear to be due to increased population in those areas. The bar charts allow users to compare the data from two states (and the national averages) at the same time, but it is difficult to use this to establish any clear correlation. We decided to try to implement a scatterplot that plotted the number of restaurants per million people vs. the percentage of adults diagnosed with the specified health measure for each state. An image of such a scatterplot is shown below. The health measure selected is obesity and all of the 4 restaurants chains were included in the number of restaurants per million.



We decided that this kind of scatterplot does not give the user an intuitive way to figure out what trend exists. The easiest way that we could think about showing the trend was to run a linear regression through the scatterplot. An image of this visualization is shown below.

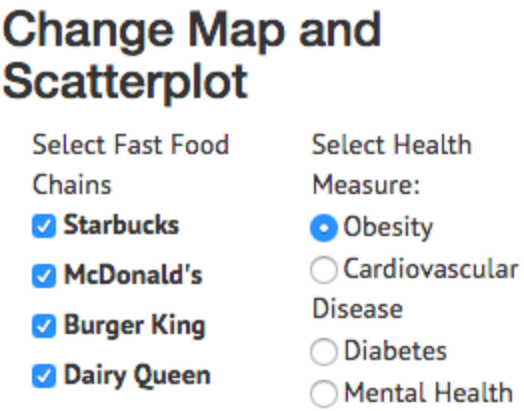


To fit the scatter plot along with the rest of our visualization, we had to rethink our design layout. We ended up going with the layout that is sketched in the following image.

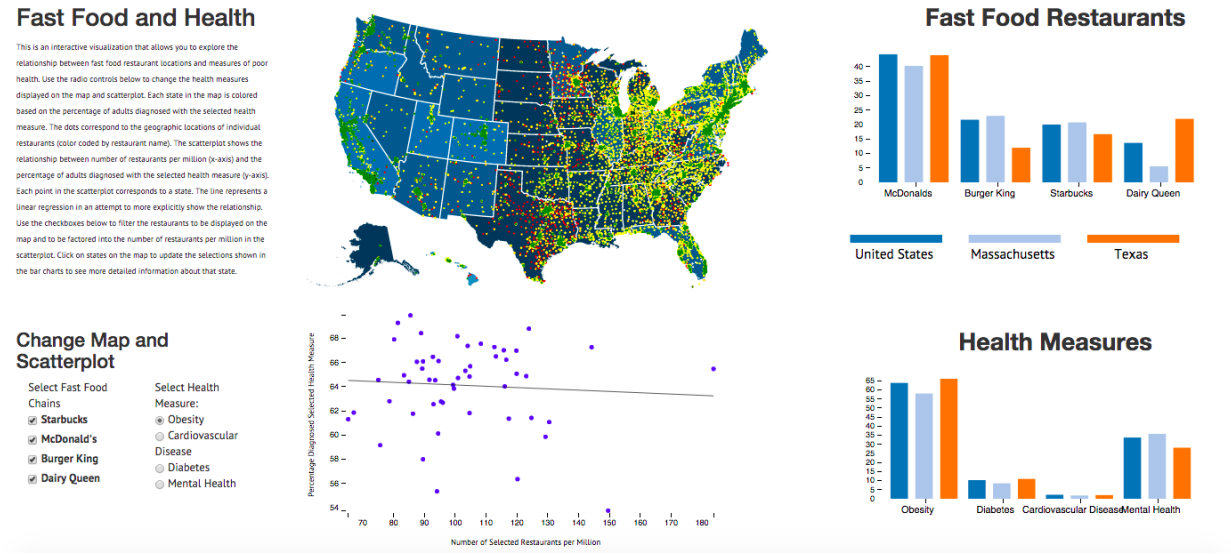


This allowed us to keep the map and scatterplot the main aspects of our visualization with more detailed data showing up to the right. We kept each visualization separated in terms of space but they still interact with each other.

To fit the controls in the bottom ⅛ of the screen (roughly), we had to rethink their implementation. The barcharts now have no filtering available and the controls simultaneously change the map and scatterplot visualizations. An image of the new controls is shown below.



The following is an image of our final visualization layout and implementation.



Having completed this final version of our visualization, we applied some minor graphical fixes (overlapping texts and such) and moved our visualization online (www.healthandfastfood.website).

Implementation

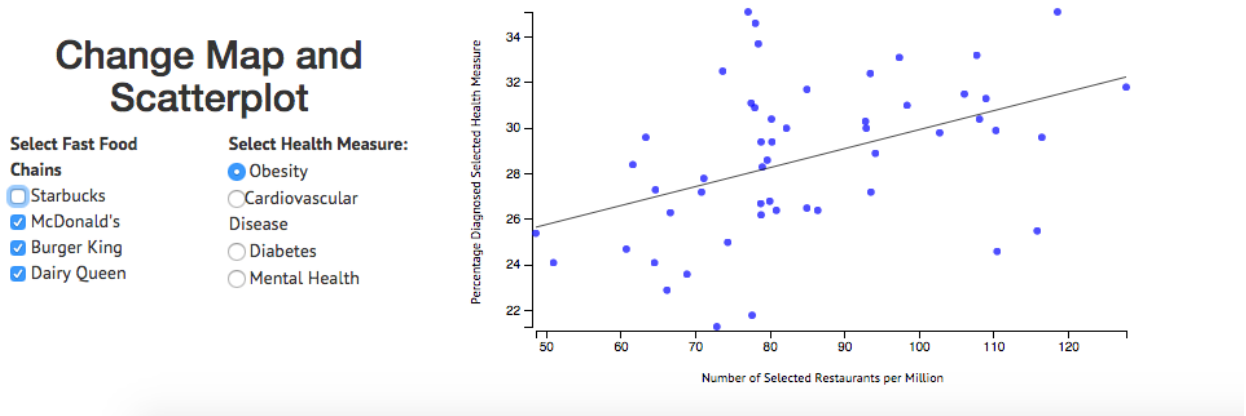
We ultimately created a main visualization with a supporting storyline.

- The supporting storyline has minor interactivity. It is mostly static (except for the force directed layout, which is only interactive in the sense that it's fun to drag around the bubbles).
- Our main visualization has multiple interactive aspects. It is an interactive visualization that allows the user to explore the relationship between fast food restaurant locations and measures of health.
    - The **radio controls** below to change the health measures displayed on the map and scatterplot.
    - The **checkboxes** filter the restaurants to be displayed on the map and to be factored into the number of restaurants per million in the scatterplot.
    - Each state in the **map** is colored based on the percentage of adults diagnosed with the selected health measure. The dots correspond to individual restaurants. Click on states to update the selections shown in the bar charts.
    - The **scatterplot** shows the relationship between number of restaurants and the percentage of adults diagnosed with the selected health measure. Each point in the scatterplot corresponds to a state. The line represents a linear regression to show overall trends.
    - The **bar charts** show health and restaurant data for the state selections

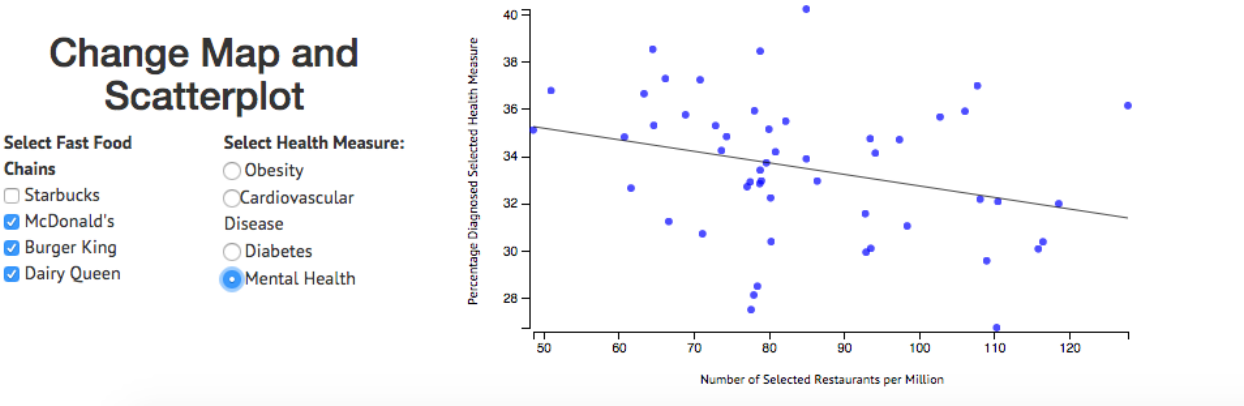
Evaluation

We learned a number of things about our data in the process of our visualization. Most of those things became clear to us, once we followed Mohammad's advice and included a trendline within a scatterplot. This trendline showed us a number of interesting trends, such as the fact that the number of McDonald's, BurgerKing, and DairyQueen restaurants in a state is positively correlated with several health measures, such as diabetes, obesity, and cardiovascular disease. We also learned that adding Starbucks data to the visualization essentially erases the positive correlation.

Obesity:



Mental Health:



However, these restaurants apparently have a positive impact on happiness, since we found that mental health with high numbers of these chains have fewer people with mental health issues. Apart from that it was also interesting to see, how our data showed population centers. Major population centers easily show up on our map as a collection of dots.

Going into this project we had one major question: does the number of fast food chains in a state have an effect on the health of that state. Using our visualization, we were able to see that there is indeed a correlation between the number of fast food restaurants in a state and health outcomes. The exception here was Starbucks that showed a negative correlation for most of our major health issues. Apparently drinking lots of coffee makes you healthier.

Overall our visualization works remarkably well. We were able to implement all of our core features and even a number of our extensions. In addition to this, our understanding of the project grew so much during our work that we added visualizations we had previously not even considered as options (such as the scatterplot, which we basically started thinking about after our TF meeting). The visualization is currently a little bit slow in updating restaurant locations on the map upon clicking a checkbox. This is because we are manipulating thousands of values at once. We were wondering if it might be possible to speed the process up in some way, but did not have time to work on this issue. In terms of data analysis, we feel as though there may be some confounding factors that we were unable to visualize, such as income. If given more time we might be able to implement the ability to view a heat map of per capita income by state. This would allow the user to also explore the relationship between income and measures of poor health, which might underlie some of the relationships shown in our visualization. Our main visualization is also a little bit cramped, and we wish we had more time to struggle and fight with HTML and CSS to get it to a more clear place. However, compared to our earliest versions the current implementation is very clear, and so we are confident that users will be able to use our visualization well.