**CPSC 583 – Assignment #2 – Report**

**Description of Data**

 The data I used for this assignment was the Cereal Obesity Poverty excel file, which was provided in class. It provides insight into the obesity problems plaguing the United States, broken down into individual states. It is useful to visualize this data because it is a large data set that contains a lot of different trends that would be hard to identify by looking at it as a whole. By visualizing the data we can see if certain age ranges, income levels, races, education levels, or genders tend to be more obese than others and then try and determine a cause for it and figure out a solution. Also, by looking at their favorite types of cereals, this can give us insight into why some states may have a higher percentage of individuals that are overweight or obese than others. By creating this visualization, I learnt that although we think that sugary cereal would be large part of the obesity problem, there are some states that have a very sugary cereal as their favourite cereal, but their poverty rate is low and their population is mostly in the good weight category (Colorado as an example). So cereal must not be the only factor contributing to the obesity problem.

**Directions and Change**

I produced 10 distinct sketches in order to determine how I could best visualize the data. I found it more difficult to create a data summary this time because the data set was so large. I decided to do 4 data summaries that were a little bit smaller in order to try and incorporate a little bit of data from each of the data set sheets in the excel file, but also make sure that my data summary was representative of the full data. I had a difficult time coming up with a way to visualize the cereal nutritional data. As I sketched, a lot of my ideas for the cereal data came about in the form of a tooltip. I found that a lot of the sketches worked well for the data summary I was using, but when I was thinking of the final visualization, I thought it would need a lot more interaction in order to not make it cluttered. A lot of the sketches I completed could have been used with multiple views in order to help visualize all of the data. In the end, I chose to complete my visualization based on my sketch of the US map, so that the user could easily see the differences between each state. Then using a second and third view, the user could interact with the different states and the different buttons in order to access the information they want to see at that time.

**Representation**

The mapping from the data to the visual representation that I used was in the form of each state containing a circle, which represented the poverty rate of that state, which was then represented as the radius of the circle. Then, the colour of the circle represents the weight category that is the highest for that state.



I chose to do it this way in order to show any trends that may become self-evident between the poverty rate and the obesity rate of each state. The colours of the sideways bar graph represent the three weight categories of Good, Obese, and Overweight which equal 100%. I used a bar graph so that the user could easily see the differences between the three categories for each state. I also created a bar for the exercise category, but kept it separated from the other bar graph, since it is not part of the weight category and changed its colour to signify that it is separate data.

Each scatterplot graph I created has a unique colour scheme. The age category uses shades of orange, the income category uses shades of green, the education category uses shades of red, and the race category uses shades of blue. In the gender scatterplot, the blue nodes represent obesity information for the male population and the pink nodes represent obesity information for the female population.

Each of these scatterplots use darker colours for higher numbers and the colours become lighter as you make your way to the lower numbers they are representing. I also used different shades of colours to represent the categories that didn’t contain numbers.

**Presentation**

 I created a map graph(see end of report for sources of images) in order to show any trends that may be visible between the poverty rate of each state and their highest weight category.

 I decided this would be the easiest way for the user to see any correlations between these two sets of data.

 I then created a bar graph that showed the weight breakdown of each state out of 100%, so that it is easy to compare the different weight groups for each state.

I also decided to create multiple scatterplots that the user can view based on the year and subcategory they want to see.



This gives the user the freedom to choose which information they can see at any given time and doesn’t overwhelm the user with too much information at a single time, as they explore the obesity data. I also created tooltips that give the cereal information for each state in the map. I only included the cereal information for each of the states favourite cereal. I did not include the rest of the cereal data because they were not attached to state data like all of the other data was.

**Interaction**

For user interaction, I decided to use hovering over the circles of each state in order to give the user more information. Also, when hovering over a circle in the state, the bar graph changes to display that states weight breakdown.

Also, when hovering over the circle, the favourite cereal of that state is shown along with the cereal information for that particular cereal. There is a lot of cereal information, so I decided this would be a good way to only show it when the user wants to see it.

When the user moves off of the circle, the cereal information disappears, but the weight breakdown stays until the user hovers over another circle. I thought this way the user could see the breakdown and examine it before moving onto the next state.

I created buttons for each of the 5 years that obesity data was collected and a button for each subcategory as well.

This way the user could pick the year they want to see information about and pick the specific category as well. The year must be picked first and then the subcategory.

**Positive Features**

 There are many positive features to the visualization of this data set. By taking a glance at the map, it is easy to see the poverty rate and highest obesity rate for each state. For example, Colorado has a low poverty rate and most of their population is in the good category whereas Mississippi has a high poverty rate and most of their population is in the overweight category. The interactivity of the visualization is helpful in order to guide the user through the large amount of data that the data set contained and be able to see the trends that occur between the favourite cereal of the state, their exercise rates and their obesity rates. This could help the user quickly identify alarming trends with weight and certain types of cereals or help identify harder to see correlations with different category types.

**Map image sources:**

Source for map of US:

[https://commons.wikimedia.org/wiki/File:Usa-state-boundaries-lower48%2B2.png](https://commons.wikimedia.org/wiki/File%3AUsa-state-boundaries-lower48%2B2.png)

Source for map of extra US States:

https://commons.wikimedia.org/wiki/File:Blank\_map\_of\_the\_United\_States.PNG