CS 4460 - Intro to Information Visualization
P5: Putting It All Together
Anna Riehle

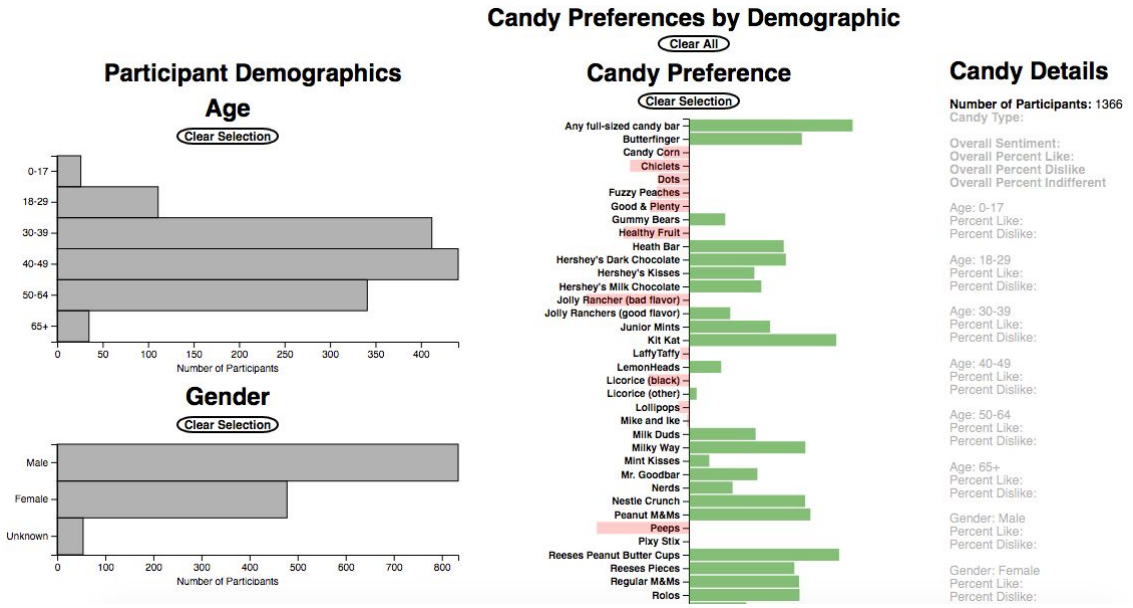
This project uses the candy dataset from the Science Creative Quarterly of U. of British Columbia. I did this project individually. For this project, I wanted to answer the following questions: What percent of people like or dislike certain types of candy? How does candy preference vary by age, gender, or nationality?

I started by bracketing, bucketing, and encoding the data by doing the following:

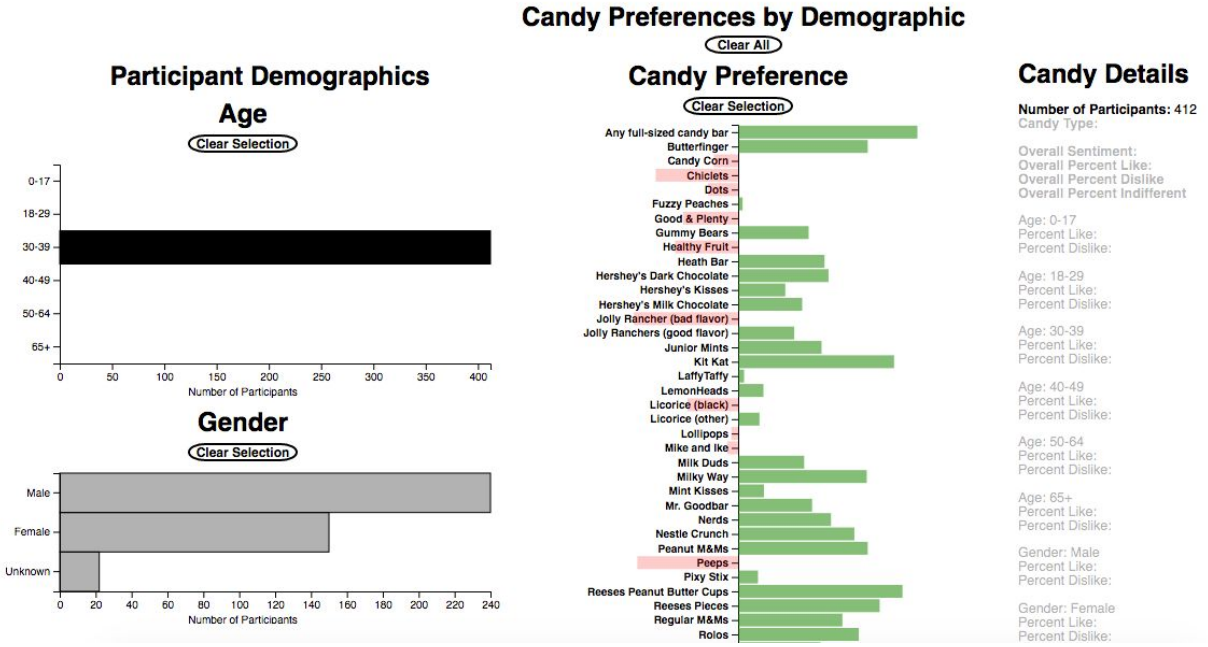
- Bracketed the age data into the following bins:
    - Age 0-17
    - Age 18-29
    - Age 30-39
    - Age 40-49
    - Age 50-64
    - Age 65+
- Encoded cells in the Gender column that were neither Male nor Female as “Unknown”
- Encoded cells in the Country column that were neither the United States nor Canada as “Other”
- Encoded candy columns with the following values:
    - JOY: 1
    - MEH or Blank: 0
    - DESPAIR: -1

I then calculated sentiment for each candy column as that column’s average value, taking inspiration from Sentiment Analysis in Natural Language Processing.

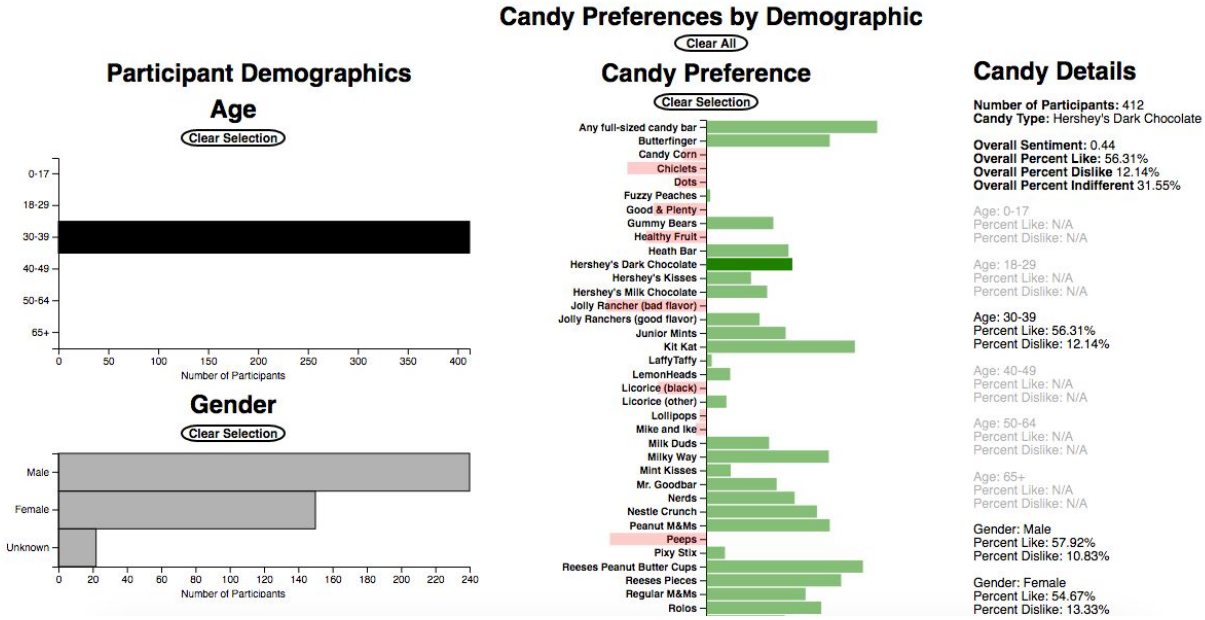
My visualization begins with an overview of the data, showing the distribution of each age bracket, gender value, and country value as well as the overall sentiment of each candy. The Candy Details column is largely grayed out as it is not yet in use.



The visualization is composed of a series of linked charts. Hovering over a bar in the left column highlights that bar and changes the values in the center chart, reflecting how candy preferences change for that demographic. Clicking on a bar in the left column also changes the appearance of the charts in the left column, reflecting how the distributions of age, gender, and nationality vary when filtered by the selected value. A single bar can be selected for each demographic category with a maximum of three possible filters applied simultaneously. Filters can be removed by clicking on the selected bar, clicking the “Clear Selected” button for the relevant chart, or clicking the “Clear All” button. The changes in the charts in the left column are animated.



Hovering over a bar in the center column highlights that bar and provides details about candy preferences - both overall and by demographic - in the right column. When filled in, the text in the right column becomes more prominent and is no longer grayed out. If a value has been filtered out of the data, that value will remain grayed out and receive the value “N/A” in the right column. Clicking a bar in the center column keeps that candy selected, allowing the user to explore how candy preferences vary for different demographics by using the demographic charts to apply filters. As with bars in the left column, bars in the center chart can be deselected by clicking the bar, clicking the “Clear Selected” button for the chart, or clicking the “Clear All” button.



Given more time, I would like to do the following:

- Sort the bars in the Candy Preference chart based on selected demographic values
- Use scatter plots to show correlations between preferences for multiple types of candy
- Create choropleths showing candy preferences by country or by state/province for the United States and Canada
- Apply responsive design principles to the visualization