P5: Candies and Who Likes Them

Thomas Farid, Brent Peterson

CS 4460

4/18/2019

Overview:

As a team, we decided that the most interesting datasets of the given list were movies and candies. We looked at the movies dataset and found that it did not provide easily accessible questions that a viewer may want to answer by using the vis. When we saw the data for candies, we saw that the individual rows were reviews, with information such as age and location of the reviewer and values from the ordinal set "'Despair', 'Meh', 'Joy'" for a set of 47 candies.

Analytic Tasks: *what can you do with our project?*

- View which candies are most popular
- Find which candies are preferred by which ages
- Learn the popularity of a selected candy in all 50 states
- See the number of reviews that led to the popularity score
- Search for a candy of interest

Design Overview

The first vis is a scatterplot which contains each of the individual candies. They are arranged through their values for two attributes: 1. How many people gave that candy a rating of 'Joy' on the Y-axis, and 2. The average age of the people who like that candy on the X-axis. This vis provides an easily accessible view of the different levels of popularity of candy for different age groups. Given that it was an average calculation for the x-axis, most values are in the middle of the range for the x-values, which makes it easier to see immediately what candies were preferred by younger, older, or all ages of people. Moreover, you can search by a specific name of candy which makes irrelevant points disappear until you click the clear button.

The second vis, located to its right, is initially blank, because the user has not yet selected a point in the scatterplot by clicking on it. Once they do, the point clicked becomes green and the second vis loads up as a heatmap of the United States for the popularity of the selected candy across all the states. By passing the mouse over the different states you are given details on demand about what the average review score was [0,1] and how many reviews make up that given state's response. This was put in to allow for understanding the context of the data. It discourages conclusions that are not supported by the data or cannot be logically made.

Technical Details

In order to translate from three ordinal values, we decided to assign them the values of 0, for despair, 0.5 for meh and 1 for joy. This made it easy to find the popularity rankings for the

map and the joy count for the scatterplot. Values for the popularity in the map come from taking the total score given for a given candy for a state and dividing it by the total number of reviews for that state. The ages for the scatterplot were found using traditional average methods (a + b + c + … /n).

Screenshots and Descriptions: To better understand our vis, we have taken screenshots of key frames.

- Figure 1 shows the description section of the site. Here users are given the basic details of what the site does.

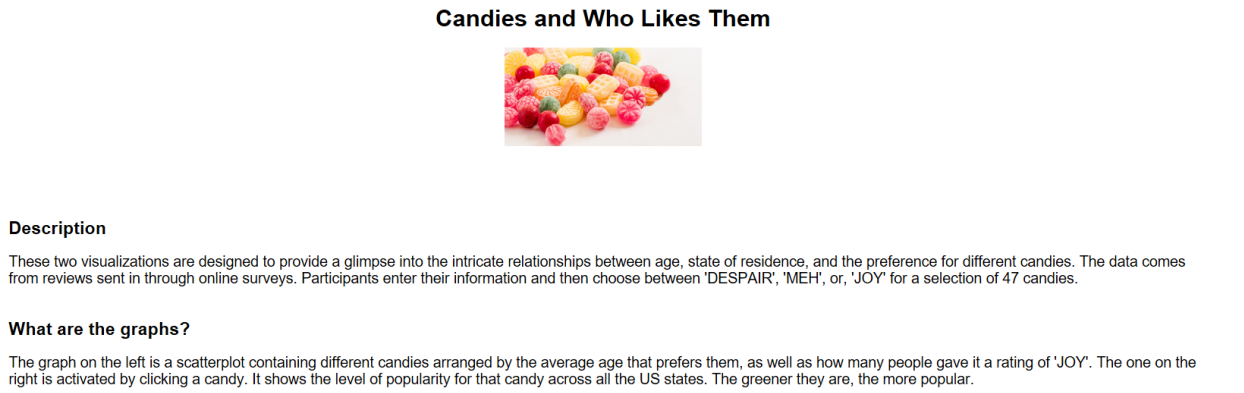

Figure 1 Description Section

- Figure 2 shows the initial data that is loaded – the scatterplot with the candies arranged as described before.

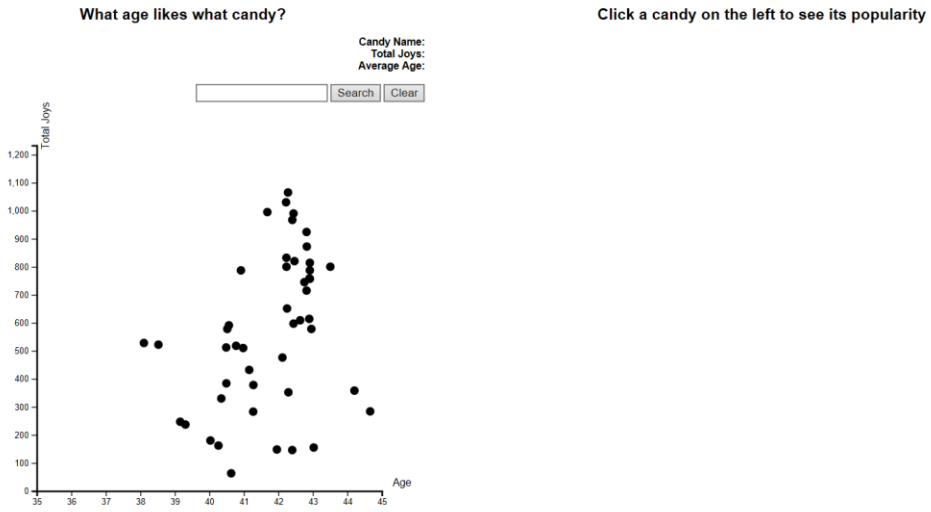

Figure 2 Initial Screen

- Figure 3 shows the mouseover functionality – the dot becomes red and its details appear next to the bold labels at the top right corner.

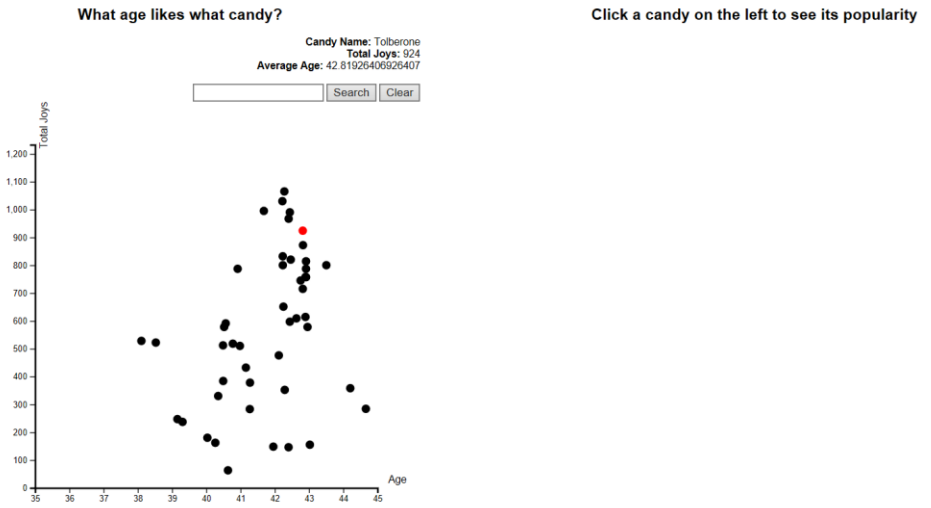

Figure 3 Mouse Over ScatterPlot

- Figure 4 shows the click functionality – the dot becomes green and its heatmap appears.

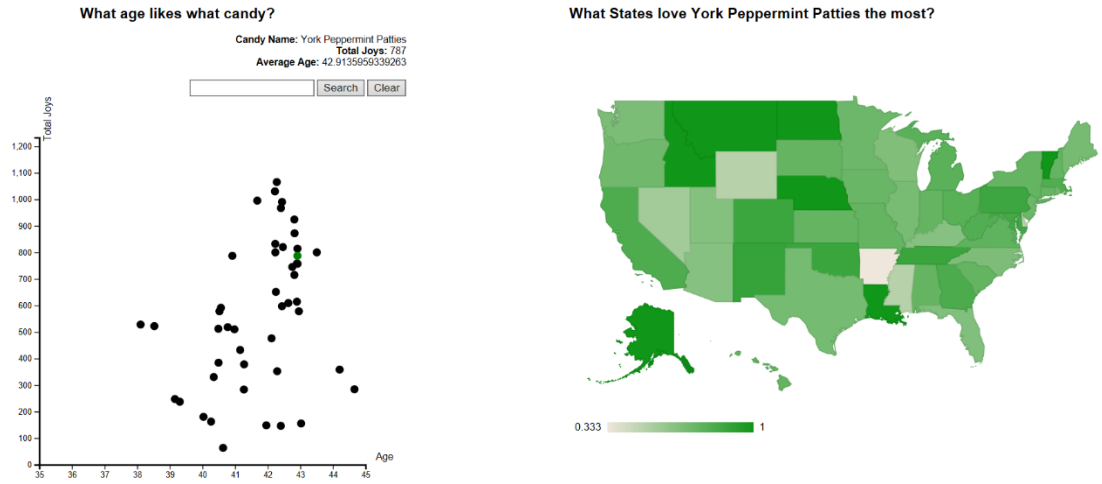

Figure 4 Clicked Candy

- Figure 5 shows the mouseover functionality for the map – a small window appears above the state giving its full name, its popularity value for the clicked candy and how many reviews were used to find that value.

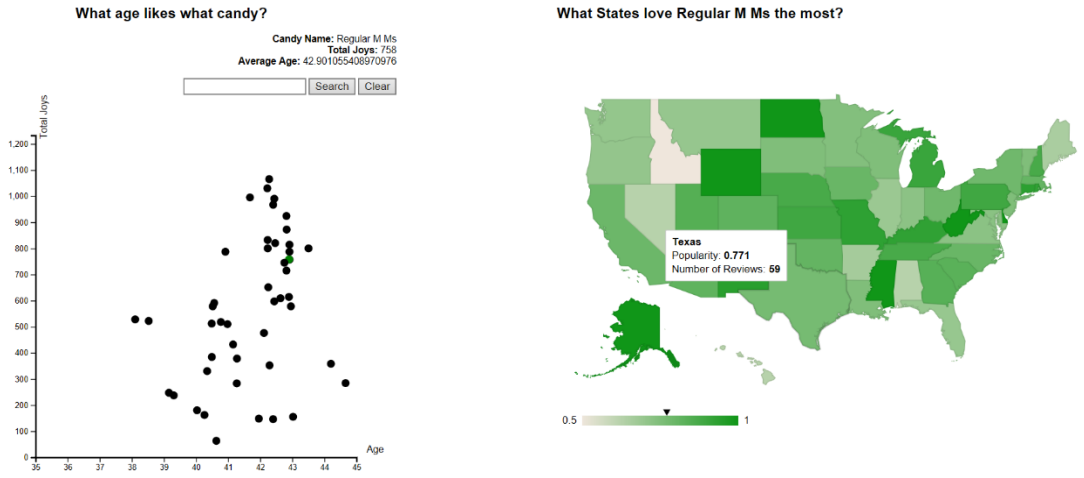

Figure 5 Mouse Over Map

- Figure 6 shows the search functionality for the scatterplot- only the dots that match the search query are shown. Pressing the clear button clears the search and shows all candies. The same as figure 2.

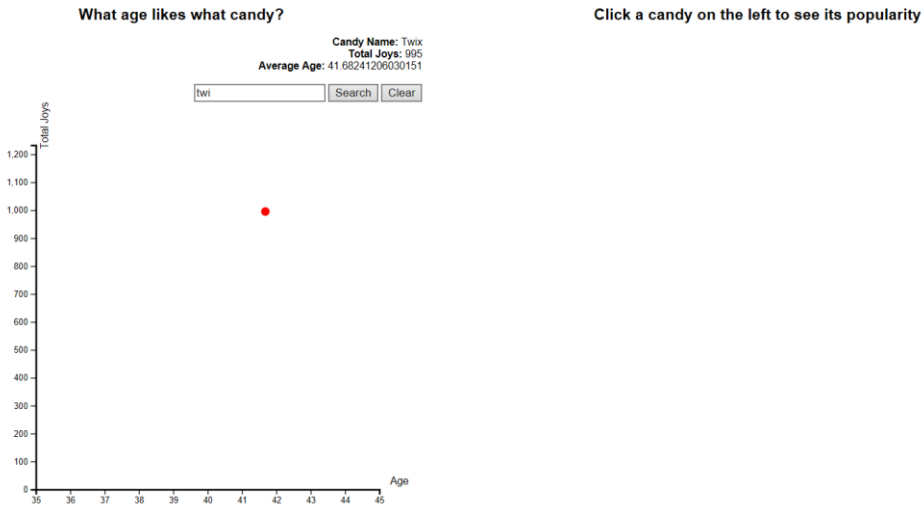

Figure 6 Search Functionality