

Gabriella Chu
INFO 4310
Jeff Rzeszutarski

Homework 1

When I looked at the data, the information that intrigued me the most was the species data for the trees. I wanted to create a visualization based on the different species of trees in San Francisco, as provided by the filtered dataset. To start, I used Python to get summary statistics and a sense of the range of numbers involved with the data. After looking at the data, I saw that many of the trees did not have many occurrences; many of the species had only one occurrence. Because I wanted to group them by species, I decided to subset out the data to only include the ten species with the most occurrences. I made two csv files to load into JavaScript. One was a subset of the given filtered data that included only the instances of the trees whose species data matched the tree species from the ten most common species. The other csv file included the names of the most common species with the counts of each species. To make the labels more concise and coherent, I split the species name column to only include the common name of the tree and not the scientific name.

To link data to visual channels, the main visual encoding I used was color and spatial position. For the first visualization, I made a bar chart. I decided that a bar graph was the best option for displaying the counts of the tree species because it is visually simple making it easy to understand and conveys the ratio of the counts well. I used the built-in d3 color library, `schemeTableau10` because it provided ten non-sequential colors, so it was well-suited to the categorical data for tree species and makes the data easy to differentiate. Inherent to a bar graph, the height of each bar, or the vertical position, is used as a visual channel. It allows the user to easily distinguish and compare value amounts, which was the goal for this graph.

For the second visualization, I created a map with only the trees from the dataset I created that includes only trees that belong to the top ten most common species. Both to relate the map to the previous bar graph and to allow the user to identify which type of tree is plotted, I kept the colors the same from the bar graph to the map. I used circles as marks because they are visually simple and understandable. The goal with this visualization was to be able to see if certain trees

are more common in certain areas or if there is a pattern in how the trees are planted in San Francisco. Here, spatial position is being used as a visual channel as the trees are being plotted based on longitude and latitude data on top of a map of the city. Looking at this map, I noticed a couple of interesting spots with high species density. I was curious about what these areas were, so I thought that users of the visualization would also be interested in seeing what these areas are. Therefore, I looked up which streets they were and if there was any importance related to them. I added text to these spots to annotate the street name and a small description of the area. I thought this would give the audience more information about the map and answer the question of what these areas were. Furthermore, I thought added some text would help to better communicate information about the map and tree species because I thought just having the colored dots on the map was not conveying information efficiently enough.