Cameron Kull, cak264

For creating my visualization, I began by using the raw data because I noticed through pre-processing that there were a lot of blank fields, so I wanted to make sure I was able to use as many data points as possible after I filtered those out. The filters I applied to the tree data removed any trees that did not have a Legal Status or a DBH, and I also removed any trees that had a DBH less than 5 so the visualization could be less crowded from the small trees and focus more on the areas on interest. I also added an attribute to each of the trees, color. I assigned a color, either blue, red, or yellow, to each of the trees based on their Legal Status so the visualization could show three generalized categories of Legal Status, DPW Maintained/Permitted Site, Landmark/Significant Tree, and all other Legal Statuses classified as Other.

To represent the data visually, the marks I used for each tree were circles, each exhibiting a variety of visual channels. I used horizontal and vertical aligned position to place the trees on the map according to their latitude and longitude. I also used hue to represent Legal Status, with 3 hues representing the DPW Maintained/Permitted Site, Landmark/Significant Tree, and all Other Legal Statuses. Lastly, I used area to represent DBH.

During my design process, the first design choice I made was to exclude trees with a DBH less than 5 because the map of San Francisco is geographically compact, and keeping the small trees in the data cluttered the map, making it harder to see the data points of significance, so I chose to remove them to help the Landmark/Significant tress and Other trees stand out more. Next, for the hue of the three categories of Legal Status I was representing, I wanted to choose colors that contrasted and would stand out in comparison to one another, but also were color blind friendly, so red, yellow, and blue were colors that fit both conditions. Another aspect I wanted to emphasize in this visualization was the viewer being able to geographically locate the data point accurately on the map, so I made the background of the San Francisco map light gray and the neighborhood borders black to help them stand out in comparison to the datapoints. In addition, I moved the neighborhood lines in front of the data points so that the viewer could see them despite there being many datapoints towards the center and upper parts of the map. Lastly, I chose to move the datapoints of Landmark/Significant Legal Status and Other Legal Status to the front of the map so they could stand out with the large numbers of blue datapoints surrounding them. Another way that I made these datapoints stand out was by making the DPW Maintained/Permitted Site trees have a lower opacity, so these blended more into the background.

As I explored the San Francisco Trees dataset, one of the data attributes that I noticed was the tree's Legal Status. Most of the trees had a Legal Status of "Permitted Site" or "DPW Maintained", but the trees which had alternative Legal Statuses, like "Landmark tree" or

“Private”, intrigued me as to why their status differed from the majority, so this was this is what inspired me to create this visualization. One thing I thought that could be related to these alternative legal statuses was the tree’s size, measured by the Diameter at Breast Height, or geographic location, so I created a map detailing all three of these variables to see if there were any correlations. Looking at this visualization, we can see that the trees with Legal Status “Permitted Site” or “DPW Maintained” are the overwhelming majority, and they appear to be located all across San Francisco with a large variation in DBH. As for the trees with Legal Statuses differing from the majority, there appears to be no geographic association to their altered status, or association to size either, meaning there is more context needed as to why these trees differ from their neighbors and receive an alternative Legal Status. Some particularly interesting findings that can be seen in this visualization are that trees larger in diameter appear to have been planted in rows or on the same street, and that Landmark and Significant trees appear to be sporadically placed throughout San Francisco, and are medium to smaller size in diameter. This shows viewers that the significance assigned to these trees is not necessarily because they are in a protected area, or that they are on the larger side suggesting a tree older in age, but rather exhibit some other characteristic that would need observing.