

Jennifer Russell
INFO 4310
9/22/20

Homework 1 Write Up

When I first went to explore the San Francisco tree dataset, a couple of things immediately came to mind. First of all, to me, the location data wasn't super interesting because I have no prior knowledge of SF. I wouldn't be able to create a context of the city to draw further connections between trees and the city landscape. However, what I did want to know were things like: How many of trees are planted every year? What kinds of trees are planted? Are there some trees that fare really well in San Francisco? In the end, I ultimately decided that I really wanted to share a story of time. Would San Francisco plant more trees over time since it is a fairly liberal city that cares for the environment? Would people be able to see any defining trends, such as more trees being planted under democratic presidencies? The passage of time holds its own meaning for every viewer, and I really wanted to share the story of how the urban canopy of SF was formed and how modern day society compares to life even just 50 years ago.

I started by looking at the processed data in a Jupyter Notebook, a tool I find incredibly helpful for exploring and analyzing the data. I was especially curious about the different types of trees and what species were the most popular to plant on a year-to-year bases. However, graphing the amount of trees planted by species in a static visualization lead to many problems. For one, many species had many years with 0 planted and maybe a few years with large spikes. There were also over a hundred different species, so it would've only been plausible to pick the top 5 or so species and many of the trees were missing the species label. Lastly, showing just a few species wouldn't allow the user to truly understand the scope of the city planting initiatives and I decided that the graphic would be much more powerful if I stuck to the overall totals of trees planted per year.

This is when I realized I had made the grave mistake of working with the de-densified data. Although using the full dataset would lead to longer loading times, this was a crucial factor for my graphic since I needed the accurate tree counts. I went back and downloaded the full dataset. The only datapoints that were filtered out were only trees that didn't have a date. That lead to a pretty comprehensive dataset from 1970 onward. The data from the 60s had some dates, but only small amounts, leading me to think the data recording was extremely variable and I

didn't want to share inaccurate data with the viewer, so I used 1970 as the start date for the graphic. I didn't use 2020 since the data only accounts for half the year. To get the counts by year, I made a second csv, treesPlanted, which contains every year between 1970- and 2019 and a count of the number of trees planted that year.

I decided to keep my design pretty simple so that the main point, a story of SFs history of planting trees, could truly be the focal point. I chose to use a line graph since that would allow users to easily identify changes over time. I used **aligned horizontal position** to show the years, since it lines up easily with our human perception of moving through time and **aligned vertical position** to show the number of trees planted. To make the data points prominent to the user, I put circle marks at every data point that were fairly larger than the line. This allows the user to easily pick out the exact values represented by the line graph. I also made these points more obvious by lowering the opacity of the line but keeping full opacity of the circles. After creating the graph, I wanted to really drive home a story of how San Francisco has thought about planting trees over the years. I did quite a bit of research into tree planning initiatives/laws/funds and came up with little, but I did find some information that I thought would be really cool to know. I used **text** annotations to add more information for the user to process, with pointers linking those text descriptions with the specific years data point on the graph, so the user can put into context exactly when these milestone moments happened. For my color schema, I wanted to use earthy tones that celebrated arboriculture. I settled on a dark forest green, used for the lettering & axis, medium green used for the title to immediately grab the viewers' attention, and a lighter bright green used to make the graph stand out and present the information clearly. All in all, my design choices were made with the values of simplicity, clarity, and a focus to bring to light an analysis of how SF has done in terms of building their urban canopy over the years.