Visualizing Tree Distribution in San Fransisco Over Years since 1970

Data Preprocessing

I am specifically interested in the distribution of trees in San Francisco over years and how it changes over time. All data processing processes were done in Python jupyter notebook with pandas package.

Because this project only uses static visualizations, I need to group the data of years into decades so that the chronological information shown in the visualizations would not be too much (if not grouped into bins). Because many rows were missing the timing information if the trees were planted before 1955, I removed all NA values in the time column as well as all invalid dates (I noticed that there were some trees planted in 2065 according to the data). Then I extracted the year from each datetime and grouped them into decades. In addition, I am also interested in the caretaker of trees over time. I noticed that a majority of the trees is under the care of Private organizations or directorate of public works (DPW), so I set all other entities to be "Others." I also calculated the number of trees planted by each category of caretaker in each decade by grouping the data by decade and then caretaker.

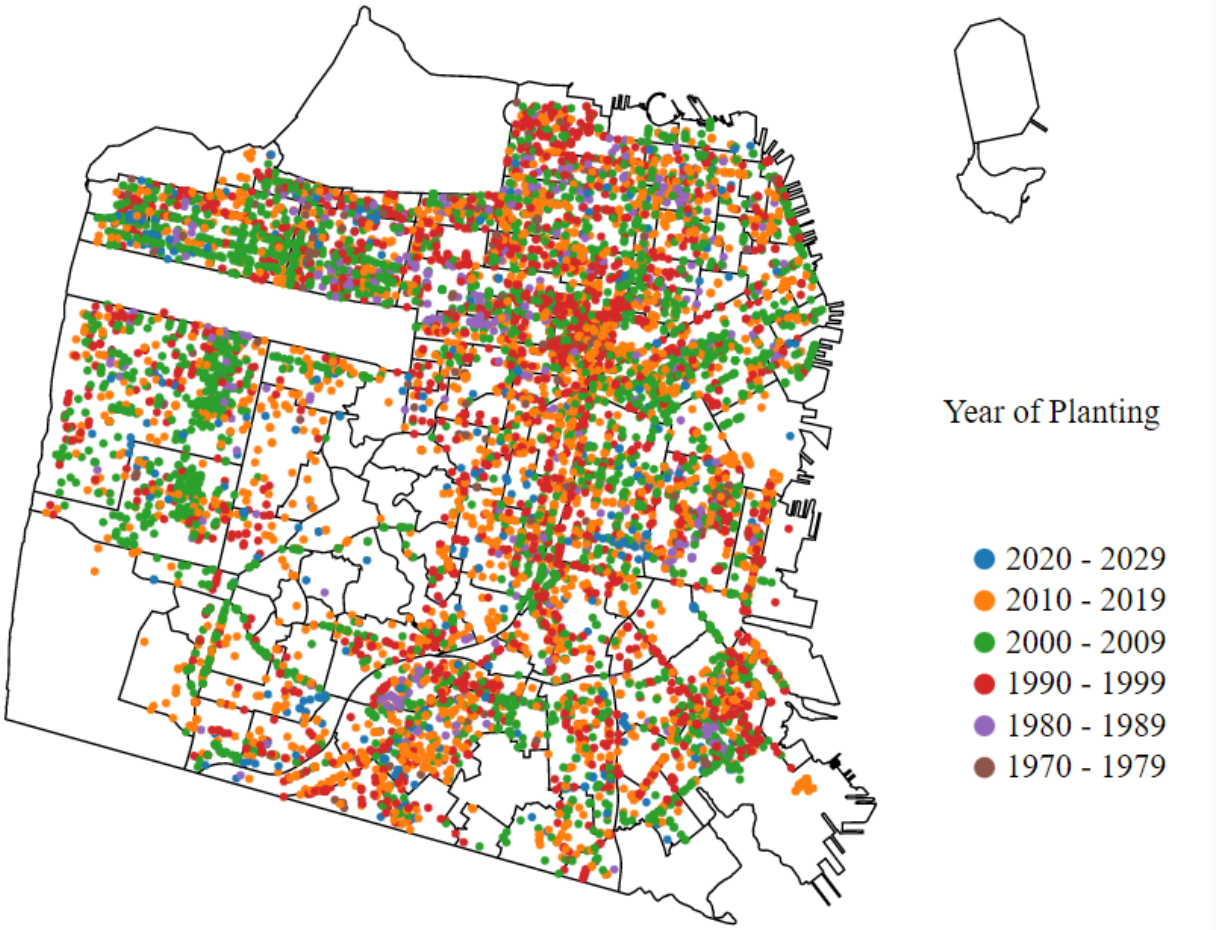
Design Rationale

The first visualization is a map implemented with the provided topojson file. The marks are the lines of the districts in San Francisco and the points that represents trees, and the visual channels include the color hue of the points representing the year of the planting of the tree. The visual encoding is mainly the color values, and I used darker hues (dark red, purple, brown) to represent older years and lighter hues (blue, orange, green) to represent the more recent years. It can help viewers intuitively understand the change in time.

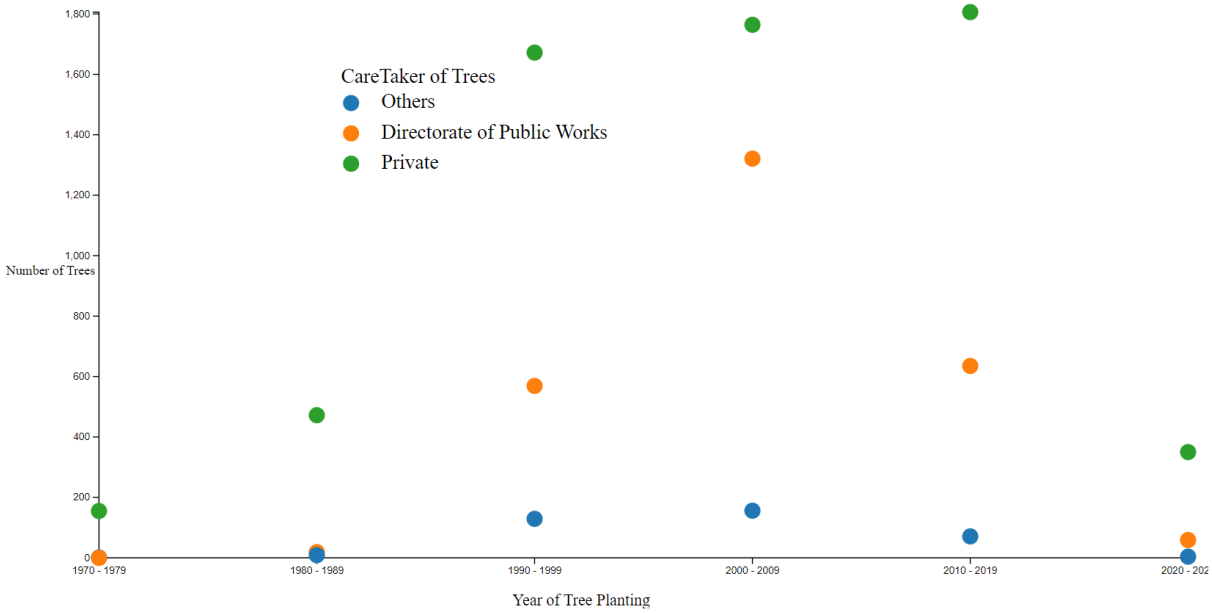
The second visualization is a scatterplot used to reflect the change of caretakers over years in San Francisco. The marks are the points representing a data point in a specific decade, and the visual channels include the color hues that represent the caretaker, the aligned horizontal position, and the vertical position representing the number of planted trees in a specific decade. The visual encodings are mainly the color values and the shape. I chose to use circle shapes on the plot to help viewers connect the data points here with the trees in the previous map visualization and build a synthesizing understanding of the topic.

The Story

The two visualizations combined show the change in tree distribution in San Francisco over time. From the map, we can see that there was a drastic increase in tree planting in the city since 1990, representing by the red, green, and orange dots. In addition, the scheme of tree planning has been laid by the works done in the 1990s to be concentrating in the northern and the eastern parts of San Francisco, as we can see on the map that later trees were planted mostly near the regions of red dots concentration.



However, We can notice that there is an increase in tree planting in the western part of the city since 2000s. This coincides with the observation in the second visualization that although the number of trees taken care of by private holders always outnumbered the other entities, there was an increase of caretaking by DPW in the 2000s.



The coincidence actually reflected a government effort of development and redevelopment in western San Francisco in the 2000s in which one of the accompanying effects was the increasing number of government-planted trees.