

HW1 - Sonia Sunil (ss3587)

The story

This static visualization illustrates the distribution of the top three tree species in San Francisco. The purpose is to synthesize patterns in where different species are located across the area and on how they are planted/distributed (scattered or clustered, straight or uneven). From the visualization created, we can see that for the three most common species of trees in the area, the *Platanus x hispanica* is distributed largely around the North-East and is planted in a straight pattern for most. It is clustered in this straight pattern generally, other than a few exceptions. The *Metrosideros excelsa* is mostly distributed well over the whole area and also has many of its trees planted in a straight pattern. However, there are also some that are planted unevenly. As for the *Lophostemon confertus*, it is highly concentrated in a specific spot in the North-East and there is a scattered planting of this species over the whole area.

Processing of data

The data was processed to extract the species name for the "qSpecies" attribute in the filtered dataset, by trimming at the ":" symbol. After which, a count of each type of species present was calculated. The top three species were only extracted to be evaluated as this would give a clearer picture of distribution on the geojson map and it would be more useful in understanding the distribution of a few species than many. The dataset was thus filtered to retrieve only data items that belonged to the top three species in San Francisco, to scale down the data plotted.

Visual encodings used

As there are three species to display, we needed to make their representation distinct on the visualization. Hence, the use of color as a visual channel. Shape as a visual channel was considered, however, it was not easy to see on the visual at areas where data points were clustered. Three specific colors were chosen based on their likeness to trees and plants. As we want to visualize the distribution by area and its patterns in planting, a map was used, plotting at Latitude and Longitude. For each data point a circle of low opacity was used to plot. This was to allow interpretation of cluster locations on the map - darker circles showed greater clustering as opposed to lighter circles.

Other rationale

A legend was added at the top left corner of the visualization to allow users to understand the color-coding in the map - which color represented which tree species. To clarify that this is only a visual representation of the top three species, a legend title was added. A title was placed as well to give users an idea of what they are visualizing.