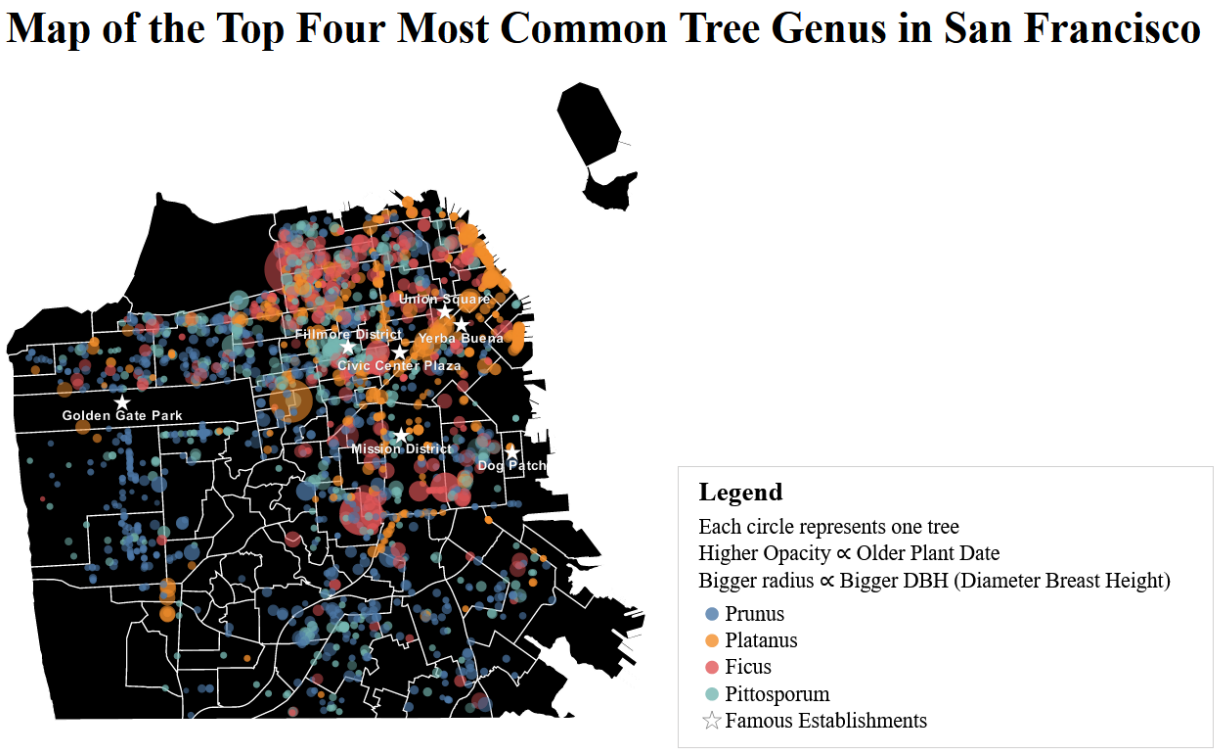
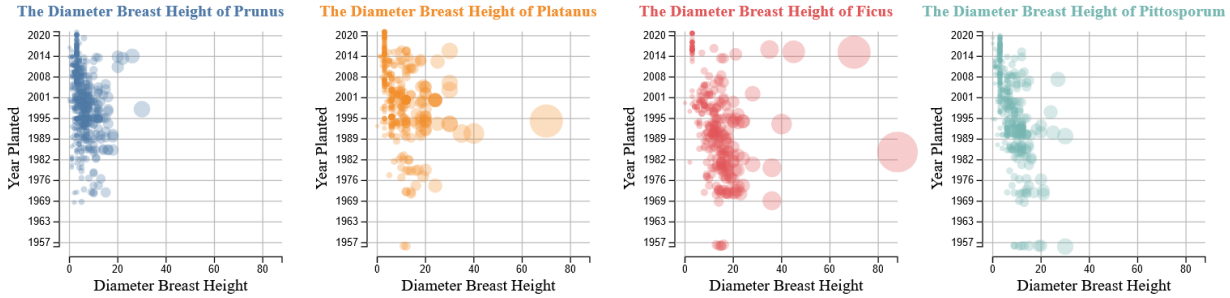
Design Document

Picture of visualization https://sf-treemapvis.onrender.com/static/ :





Species in genus: Prunus

Prunus cerasifera
Prunus serrulata 'Kwanzan'
Prunus serrulata
Prunus x yedoensis
Prunus blireiana
Prunus spp 'Purpurea'
Prunus subhirtella 'autumnalis'
Prunus caroliniana
Prunus domestica 'Mariposa'
Prunus cerasifera 'Atropurpurea'
Prunus spp
Prunus cerasifera 'Krauter vesuvius'
Prunus laurocerasus
Prunus subhirtella 'Pendula'
Prunus serrulata 'Mt. Fuji'
Prunus sargentii 'Columnaris'
Prunus serrulata 'Amanagawa'
Prunus x yedoensis 'Akebono'

Species in genus: Platanus

Platanus x hispanica
Platanus x hispanica 'Yarwood'
Platanus x hispanica 'Columbia'
Platanus racemosa
Platanus x hispanica 'Bloodgood'

Species in genus: Ficus

Ficus nitida
Ficus microcarpa nitida 'Green Gem'
Ficus retusa nitida
Ficus macrophylla
Ficus microcarpa
Ficus Spp.
Ficus laurel

Species in genus: Pittosporum

Pittosporum undulatum
Pittosporum tenuifolium
Pittosporum crassifolium
Pittosporum tobira
Pittosporum spp
Pittosporum eugenioides

Story

The story I'm trying to tell is that the size of the diameter breast height (DBH) is affected by various factors such as location, the year the tree was planted, and the type of genus. In terms of location, denser urban traffic (potentially from tourism) would require more shade. This may lead to more trees in bustling urban areas – to balance out buildings to make an area more habitable. There also seems to be a trend with the year the tree was planted and the DBH. For example, in all four of the genera's scatterplots, the dots shifted from the top right to the bottom and slightly towards the left. While the variation isn't super big probably since the trees are comprised of similar genes, it's possible that the year that the tree was planted is somewhat proportional to the tree's age.

Processing

I further filtered the previously given filtered file, Street_Tree_List-2022-01-30_FILTERED.csv, by the top four genera and whether the row has a nonempty PlantDate. The new csv is called Street_Tree_Filtered.csv. To filter by genus, I first found the four most common genera. The genera was identified by taking the first word in the qSpecies column of the csv file. Then I filtered out all the rows that have an empty string for the PlantDate column.

Visualization Encodings

I created a map of San Francisco where the horizontally aligned position represents the latitude, and the vertically aligned position represents the longitude. Each circle (mark) is used to represent a tree. Each star (mark) is used to represent a special establishments in San Francisco's culture (taken from https://www.sftravel.com/article/7-arts-culture-neighborhood-maps-to-help-you-explore-san-francisco). The color channel represents each of the four most common genus in the csv. The size of the radius (channel) represents the DBH (Diameter of Breast Height). The opacity of each circular mark (channel) represents the year the tree was planted; the darker the earlier the plant was planted relative to today. The Diameter Breast Height of genus charts uses circles as marks as well. It uses two visual channels to portray the DBH: horizontally aligned position and the radius of the circle. The visual channel is vertically aligned position which represents the year the tree was planted. The color visual channel when looking at all four charts together represents the genus again.

Rationale of Features

Using two visual channels to communicate the same idea in the graphs is a bit strange. However, I justify that by using both the vertically aligned position and circle radius to show DBH, the user is able connect the circle's radius from the map to the scatterplot. This may make it easier to understand how size would correlate with a tree's numerical DBH. It also serves to help draw the viewers' attention from the top left to the bottom right, so they understand that the year a tree was planted is slightly connected with the DBH. I also decided to put year in the y-axis instead of the x-axis because it's easier to view more years without being scrunched up on the bottom axis. Additionally, while I originally wanted to include plot size as the bottom axis, I realized that there were some strange inconsistencies*.

For the map of San Francisco, I decided to stay consistent with the circle mark to represent a tree and radius for ease of understanding. I decided to use opacity to show where older trees reside. While this metric in hindsight is hard to notice due to overlapping circles, we're still able to see that most of the older trees are in the same areas where most trees are. I also chose to use a mark of a star to differentiate the popular establishments from trees. The stars are white which are high contrast with the background and the circles. As a result of plotting on the map, we're able to see the density of the platanus trees growing along the division of the SoMa neighborhood (where Yerba Buena is). Additionally, most of the trees seem to be growing on the upper northeast side of the map which seems to align with where most famous establishments are located. Color choices were made with red-green blindness kept in mind. While this doesn't really pass the blue-blind/tritanopia test, I decided to keep these colors because for people without blindness, I find these colors pop out bright and contrast against the black map.

*I didn't know how to convey including that some of the measurements only had the width and most of the measurements were in a form of dxd where d is a size. After filtering that out, I realized that some of the trees DBH is bigger than the plot size (assuming the plot size is measured in feet or meter and DBH is measured in feet or meter). It was also harder to see variations and trends since most of the circles overlapped. So, I just abandoned it for the time being. To help the user understand that there's multiple species in a genus that contributes to variability in the DBH, I added a list of all the tree species on the map and chart that belong to a certain genus.