Using the dataset, I want to dig deeper into the insights of the "PlantDate" column. Specifically, I want to speak to the proportion of newer versus older trees as well as their geographic distributions. This pokes at the idea of walking around San Francisco areas and getting a general sense of how old trees are around there.

In the data processing stage, I projected the latitude and longitude of each tree into a pixel on canva. Then, to work with the time data, I needed to convert each item in the PlantDate column into Date objects using formatted parsing. This poses the challenge of correcting the century since the data presented year as only the last two digits. Thus, knowing that no tree should be planted in the future, I conditioned that the starting two digits of the year is "20" if and only if the last two digits are less than or equal to "24". Furthermore, prior to any date parsing, I needed to remedy the null data from the PlantDate column. By generalizing the specification of the column, I made all null to be in the year of 1954, which does not cause inaccurate representation as will be described later. Lastly, I made counts of the number of trees in each year and aggregate them as categorized by D3 quantize scale.

My initial design was to implement each tree as a dot and color them based on the PlantDate year. However, this soon proved useless as the points overlapped each other and adjusting the radius only made the map messier despite giving a small black boundary line for each dot. Furthermore, in my original idea, since I colored the dot using a time scale of the PlantDate that assigned color from black to blue accordingly, the amount of year present in the dataset and the skew of data to have year in the earlier range (especially 1954) made all dataset look similar in colors. Particularly, in addition to the misrepresentation of many 1954, I found that darker colors were especially hard to distinguish between and the addition of dots on top of one another just made all dots look like a single color.

Hence, in the final design, I made the trade-off between precision and interpretability. Specifically, the purpose of the hexbins was to group points from the same area to represent them collectively. With such, the amount of overlapping between the hexagons became a design decision based on radius and the colors of the hexagons became easier to see. However, there was a loss of some information presented since the outliers became hidden when only the median of the collective hexagons were considered. Additionally, due to both a correspondence to the colors of the pie chart and the difficulty of distinguishing specific years out of gradient colors, I chose to bin together the years instead of using D3 time scale on a range between two colors. This is a trade-off between precision and readability of time. However, since 5 years is often insignificant in the world of trees, limiting to only five colors can actually help readers gain information about the trees in that neighborhood. Furthermore, this can remedy the change of null data in data processing such that trees planted prior to 1954 are not generalized into the single year. Also on the topic of colors, my initial design was to use shades of green to resemble a bird's-eye view of the city in its richness from the trees. However, I soon realized the issue of accessibility and decided to use shades of blue instead. In terms of radius of the hexagons, I chose a square root scale ranging from 2 to 12 because it is easier for users to correspond area, rather than radius, to the number of trees in the region and the area of the hexagon squares as the radius grows. The decision to set 2 as the lower bound of the radius was a battle between seeing the color of the hexagon in regions of sparse trees versus making differences in tree counts more apparent between different regions by maximizing the gap between lower and upper radius bounds. On the other hand, the decision for 12 as the upper range of radius was for the overall idea of not smudging different regions together nor hiding smaller regions behind bigger regions despite using a small black boundary to better distinguish the hexagons.

Moving to the next chart, the design of the pie chart was to complement the map and to better display the proportion of the tree age in San Francisco as a whole. Specifically, because the hexbin map had information loss in the trade-off between using dots for each data versus using hexagons to represent a region, the direct ratio between the age bins cannot be interpreted accurately from the map. Hence, the use of the pie chart helps encode the number of trees in each year bin out of the whole of San Francisco trees and the label in each region clarifies the specifics in those numbers. Lastly, knowing that the pie chart is meant to complement the map above it, the radius was chosen such that the information is quick to read while not force grabbing the attention when users are looking at the map.