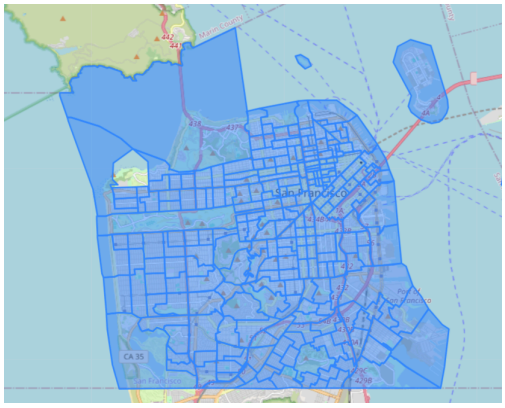
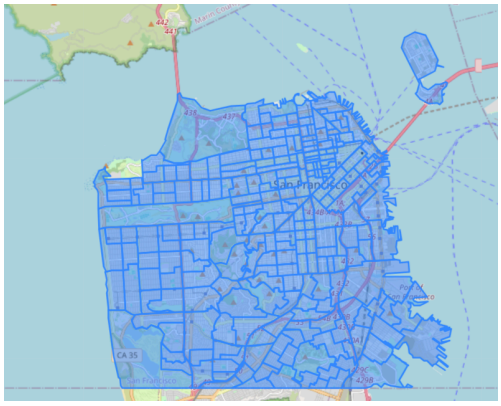
INFO4310 Homework 1
Desai Wang (dw584)

Aim:
Trees are non-uniformly distributed throughout cities, and areas of higher income often have more trees. Inspired by a NYT article investigating tree equity, I wanted to explore the relationship between number, size, and planted date of trees and income in San Francisco.

Data Processing:
I first found 2022 income survey data from census.gov and linked it to corresponding shapefiles for census tracts. Using pandas and geopandas, I extracted the median and mean income data for census tract polygons, then found the tree points that fall within each census tract. Since the tracts are different sizes, I wanted to also calculate the number of trees per area. I realized that the census tracts were not so precise (see left image, which includes a lot of area on water), so I used an outline of San Francisco to intersect with each tract, and used the union of the shapes (shown on right). However, the number of trees per sq meter had a weaker correlation with median income than total number of trees, so I didn't use this data for my final visualizations. I think the weaker correlation might have resulted from areas on the shore that are industrial (don't have trees, but also don't have residents).





	est household mean income
est household mean income	1.000000
AvgDBH	0.165795
trees per sqm	0.239028
TreeCount	0.386717

Design
I placed the overview visualizations at the top, and the more detailed visualizations on the bottom. For the overview, I wanted to demonstrate the correlation between wealth and number of trees. Thus I chose to place two choropleth maps of San Francisco side by side, one showing household income and the other showing the number of trees. Both have aligned horizontal and vertical position as visual channels, which correspond to latitude and longitude. Their visual channels also include saturation, which correspond to mean household income in the left map and number of trees in the right map. I chose a green color scale for the tree map due to the

common association of trees and the color green, and a blue color scale for the income map because it’s similar enough to the green color scale to allow pattern recognition between the two maps, but also different enough to allow the two maps to be distinguishable.

The first two maps disregard the position of individual trees, so I wanted to make another visualization that showed the difference in trees between relatively high-income and relatively low-income census tracts. I chose two groups of tracts that are similar sized, and drew trees within that area as circles. Again, vertically and horizontally aligned positions encode latitude and longitude. Further, I used circle radius as a visual channel for tree diameter at breast height, and circle saturation to represent tree plant date. I set those without a plant date to a darker shade of green, as the data suggest that trees planted before 1955 have no plant dates. Lastly, I made a key to the left that shows where the two census tract groups are located within San Francisco. I colored the groups that I’m examining based on their mean household income to highlight connections to the previous section. Lastly, I wanted to display the correlation between trees and income in another way, so I decided to remove the geographical information and create a scatter plot. Each circle represents a census tract. The aligned horizontal position represents mean household income, and aligned vertical position represents the number of trees. I chose to color the circles gray as I didn't want to overload the page by using a new color.