Alsa Khan| ak2328
INFO4310: HMW1

San Francisco's urban canopy provides crucial environmental and ecological benefits, but not all trees contribute equally. Larger-diameter trees (DBH > 20) are typically older and more established, offering increased shade, carbon sequestration, and habitat value. This project aims to answer two key questions: Where do large trees tend to cluster? Who maintains them- private owners, the Department of Public Works (DPW), or other entities? By analyzing species composition and caretaker responsibility, this visualization reveals patterns in how San Francisco's largest trees are planted and managed.

To prepare the data for visualization, I first filtered the dataset to include only trees with a DBH greater than 20 inches. I then extracted and standardized species names by isolating the botanical classification from the qSpecies field. The five most common species were identified, while all others were grouped under an "Other" category to simplify visualization. Initial mapping of tree locations revealed significant clustering along neighborhood *borders*. This observation led me to classify trees as either "Border" (within ~15 pixels of a neighborhood boundary) or "Inner" (further away) by comparing their coordinates to neighborhood boundaries from a GeoJSON file. Finally, I categorized trees based on their caretaker type, consolidating them into three groups: "Private," "DPW," and "Other." These steps ensured that the dataset was structured effectively to answer the research questions.

The visualization consists of two main components: a map of large trees by species and a bar chart comparing trees near borders vs. within neighborhoods by caretaker type.

1. The map places each tree at its geographic location using latitude and longitude, using color to indicate species type while maintaining a neutral gray basemap to provide spatial context without overwhelming the viewer. By limiting the color encoding to the five most common species, the visualization remains readable while still conveying important clustering patterns. The legend includes species percentages, adding on to the clear visual feedback on which species dominate the city's large-tree population. The results specifically demonstrate that Platanus x hispanica, Pinus radiata, Cupressus macrocarpa, and Phoenix canariensis are some of the most common species. Interestingly, these species tend to cluster along horizontal and vertical borders, suggesting large-scale planting efforts along major roads or neighborhood boundaries. This could indicate that these trees were planted as aesthetic buffers, street greenery, or part of city-wide planting initiatives.
2. Building on the patterns observed in the map, the grouped bar chart categorizes large trees based on their proximity to neighborhood borders and breaks them down by caretaker type. The x-axis separates Border and Inner trees, while the colored bars represent the *number of trees* maintained by Private caretakers, DPW, or Other entities. This encoding makes it easy to compare how tree maintenance responsibilities shift between different locations. The results suggest that private caretakers manage the majority of large trees, though DPW still appears to have a stronger presence along neighborhood boundaries. This pattern could reflect city policies prioritizing maintenance along major public corridors or differences in how trees on private versus public land are managed.

Overall, this project effectively communicates key patterns in San Francisco's tree distribution by pairing spatial and categorical analysis. The map provides an intuitive, geographic perspective on species distribution, while the bar chart makes caretaker comparisons straightforward. By structuring the dataset carefully and choosing visual encodings that emphasize the most relevant details, this visualization makes it easier to understand where large trees are located, what species dominate, and who is responsible for their maintenance.