# INFO4310: HW1 - Static Visualization Design

**DUE BEFORE: February 8, 2024 at 8:00PM eastern time**

In this assignment, you will work independently to develop a *static* (non-interactive) visualization for a dataset and explain the design choices you made during its development. Theoretically you should be able to explain why every single pixel of the display looks the way it does and how it relates to the goal you have in your visualization.

You must use HTML, Javascript, and D3 to complete this assignment. While you may use any programming language you like to pre-process the data, your visualization must not make use of any other external libraries. My aim in constraining the specification is to force you to focus on the specific design elements of your visualization.

This assignment is 1.5 weeks in duration. Please be sure to create a submission that reflects at least one week of work on the visualization and short design rationale document.

## Dataset

DataSF makes available a wide variety of data on the City of San Francisco. One of the more interesting items on the site is a dataset of municipal trees. As of 2017, San Francisco reclaimed ownership and maintenance of nearly 200,000 trees occupying city sidewalks. The dataset outlines each individual tree, providing species, size, and location information. In INFO3300, we created a simple map visualization using a small subset of this dataset. Within the assignment ZIP folder you will find several files:

| | |
|---|---|
| DATASET_INFO.pdf | - a description of useful data columns |
| Street_Tree_List-2022-01-30_RAW.csv | - a 50,000 point subset of the raw tree dataset with some columns removed |
| Street_Tree_List-2022-01-30_FILTERED.csv | - a subset of the RAW dataset containing points with species and lat/lng |
| post-process.py | - the Python3 script used to filter |
| SF-Neighborhoods.geo.json | - a geojson file with SF outlines/poly |

Please make use of either the RAW or FILTERED dataset in your visualization. You'll notice many missing rows in this dataset, as information is rarely complete in real world data. **You are welcome to further process the data** (e.g. by counting species, summing columns, computing averages, removing more missing data). Should you want to represent a dimension/measure with missing data, you can choose how to handle this issue.

# Task

**Design a *static* (non-interactive) visualization that effectively communicates some interesting aspect(s) of this dataset.** Do not just represent a few columns. Instead, think about what unexpected or meaningful thing someone could find in the data and *communicate* that using your visualization. You are free to manipulate the dataset as you wish, so long as you remain faithful to the original data and avoid deception. Recalling the simple map visualization demonstrated in the Monday lecture, note how the static map reveals density information about trees. While interesting, this is a relatively shallow attempt at investigating the data. You are welcome to create another map view with dots on it, but you must think about how you can show something more interesting than that.

There are many, many, many ways to represent this dataset. In thinking about how to communicate your insights, choose a visual metaphor that most effectively communicates the information. It's important to note that **people still 'interact' with a static visualization** by moving their eyes, pointing, and scrolling. While you may not be able to add interactive filters, think about how you can help someone filter data or see different layers of an attribute through techniques such as small multiples and annotation.

**As you design, document your design choices and rationales**. Document the story you want to tell and how the design choices you make help to strengthen it. If you processed the data, describe how you did it. Were there alternate ways of representing you considered? Did you encounter any trade-offs in showing what you wanted to show? What potential messages or areas of the data are sacrificed or hidden by the design choices you made?

# Deliverable

Before the deadline, please submit to CMS a ZIP file containing:

- The working source code for your visualization
- A 2-4 paragraph write-up in PDF form that:
    - Outlines the story you want to tell with the data
    - Explains any processing you did of the data
    - Identifies the *visual encodings* you use in your image to link data to visual channels
    - Provides a rigorous rationale for your design choices and explain how they help to facilitate the communication of the story you want to tell.

Late assignments will not be accepted. Upload early and then re-download to verify.

## Preparing for In-class Critique

In addition to posting on CMS, you must also make your project publicly available so that others can access it during in-class critique. I outlined in class one general strategy for making projects available. You can also use free providers such as Render, Github Pages, and Railway.

Once you have made your project accessible to the public, please submit a link to your project using this Google Form: https://forms.gle/8LNLBpaADREXCT23A

You must submit a link before the deadline to receive credit for this portion of the assignment.

## Grading

This assignment will be graded both on the soundness of your design and the quality of your write-up. I will also be looking for how you think about your audience, the story you are telling, and how your design choices align with those two factors. Some examples for point deductions include misleading, unmotivated, or unnecessary graphic elements; incomplete write-ups; poor choices for encoding data dimensions; choices that don't align with your intended message; and not adding your submission to this document for discussion.