



ASSIGNMENT 1

Comparing Categories

Sahar Harati
Forough Majidi

1. Bar chart: Tree Type Abundance

- The bar chart represents the abundance of tree types.
- Sorting is applied to showcase the most abundant tree types.
- Tooltip functionality is planned to provide additional information on hover.

2. Stacked Bar Chart: Cities vs Tree Type

- City Selection: Limited to a few cities, from the Hawaii state, for focused analysis.
- Tree Type Inclusion: Top-5 tree types are considered, with an aggregated "Other" category for less common types.

3. Stacked Bar Chart with Small Multiple: City vs Tree Type

- Small multiples of stacked bar charts are created, each representing a tree type distribution for each city.
- Sorting ensures a clear comparison based on total tree abundance.

4. Stacked Bar Chart 100%

This chart visualize the same data as the second chart but the x-axis is based on percentage from 0 to 100.

5. Compare Tree Species Abundance - Heatmap

- The heatmap displays tree abundance using a color gradient, emphasizing variations.
- Tree types and cities are represented on the X and Y axes, respectively.
- The color intensity signifies the abundance level, enabling quick comparisons.

Data preprocessing:

The raw dataset has a huge amount of data. We just kept scientific name, common name, city, state and height_M. The other columns were not necessary to keep because of the time it takes to load on the website.

Some of the data had null values in common or scientific names or state name which got filled using mapping them based on the correct data.

The records which were without both common and scientific name, and the ones without city name are dropped.

Due to having better statistics, we also decided to drop nulls in the height_m columns.

During the grouping process we faced a problem which was dividing the records with the same name but difference in lower and uppercase letters, by this way we make it all lower case.

At the end we used the file "output.json" as our raw data and for each plot we did simple processes like group by, count, and selecting n top records on this data.

HTML / JavaScript:

For this part we used some Json files to display the data within each Bar Chart. We fetched the data and extracted the needed data like city name and common name to display it on the bar chart. Also we used Plotly to create plots.

Python:

We used Python to process the data from the dataset to visualize it on the website. We did some processing like counting the number of trees for each city and creating the desired Json file to be used in the HTML part.

Conclusion:

The chosen visualizations aim to provide a comprehensive understanding of tree species abundance data. By employing various chart types, we enable users to explore both individual cities and comparative analyses. The designs prioritize clarity, interactivity, and informative value, enhancing the user experience and facilitating data-driven insights.