Streetside Trees in NYC

By Amber Zhen, Martin Herrera, and Yoo Kim

In the early stages of this project when our group was discussing which dataset to build our data visualization from, we felt particularly compelled by the New York City trees data. This data gave a comprehensive and thorough census of thousands of trees catalogued from streets throughout all five boroughs in New York City. Due to the availability of so much data and the visual opportunity of doing our project on trees, we chose to pursue this option and began to brainstorm what information and story we wanted to convey in our data visualization.

The data utilized in this visualization comes from a census conducted by the New York City Park Service of streetside trees in 2015. The name of the census is "TreesCount! 2015" and can be found in the references below. The census was conducted by individuals who recorded trees currently on streetsides in NYC, this means that the data excludes trees within parks or other forested areas. The census itself was extensive and comprehensive. Over 600,000 trees were recorded and for each tree the species, borough location, latitude and longitude coordinates, trunk diameter, health, location relative to the street, and alive or dead status were recorded. Additionally, each tree received its own ID.

On its own this dataset was extremely large and cumbersome to work with, causing long load times with no digestible breakdown. The primary resource used to manipulate and analyze the dataset was Excel. To begin we used Excel's plot functions to plot several aspects of the data such as tree health breakdown by borough, trees in each borough, and species breakdown for each borough, etc. We found that the only seemingly interesting varying data between boroughs was the tree population size and the species breakdown. With this in mind we proceeded to create "lean" spreadsheets of the needed data to be processed by our page. We identified the top five most populous species in all of NYC by means of counting instances of each species and then proceeded to create a spreadsheet of only those trees, their species, and their positions so they could be plotted on a map and processed easily. Additionally, using lookup functions in Excel, we were able to count the number of instances of each species in a borough and create a spreadsheet with the total number of trees in each borough and the number of each species within the borough. This data would be used to create the percentage breakdown of total trees by borough and the species breakdown by borough visualizations.

Our project consists of 3 different graphics: the leafy part of the tree, the tree roots, and the map. The tree graphic was generated using svg rectangles whose heights were sized by the ratio of the number of trees in that borough to the total number of trees in all of NYC. The leafy part is easily scaled to various sizes simply by changing the max height variable. The tree roots were created by referencing a previously made stacked bar chart (linked below). The csv file that it accesses was generated by dividing the number of trees of a certain species by the total number of trees in that borough. The 'Other' category encompasses every other species type (over 200 of them!) to ensure each bar represents 100% of trees in the borough. The roots graphic utilizes a band scale which evenly spaces out the bars horizontally over a given width. A linear scale was used for the y-axis. An ordinal scale was used to map each bar segment to its species color. Lastly, the map visual was made from a NYC json file, containing feature maps for the five boroughs, and by referencing previous code (linked below). Two ordinal scales were created for the dot colors for the species and for the fill colors for the borough backgrounds to more easily distinguish the five boroughs. The csv file contains latitude and longitude values for every tree, so trees could easily be plotted as circles using the projection variable.

Although, there were many variables we could have chosen, such as tree diameter and tree health, we chose to refine our focus to tree species data. We found that there was more variation and depth in the tree species data, which we felt could provide a more interesting visualization and help our audience learn something new. The data we provide through our graphics is meant to give the audience a sense of what species of trees exist on the streets of New York City and their breakdown by borough. From the first tree graphic, it is easy to see what percentage of all street trees in NYC come from each borough. The roots of the tree graphic convey a breakdown of the top five species and all other species by borough to show whether or not there is variation on most popular trees depending on location. The map gives a more contextual view of the boroughs in relation to each other since it shows the top five species for every borough on one map.

Surprising results from the data include the fact that Queens has such a majority percentage of all New York City street trees and Manhattan has the least amount of street-side trees. If park trees (like Central Park) were included in the data, perhaps there would be different conclusion. In addition, it's interesting to see that all five boroughs have the same top five species of trees, but the breakdown of which of the five is most popular in each borough has a lot of variation. For example, Brooklyn has a clear majority of London Plane Tree, Staten Island has Callery Pear, and Manhattan has Thornless Honeylocust. It's also surprising that out of over 200 species, just five of them make up roughly 50% of all trees. Seeing the variety of trees planted on streets through New York City creates a new image of New York as having a wider range of nature than its popular nickname "Concrete Jungle" would suggest.

References:

Original Census Data: https://www.kaggle.com/nycparks/tree-census
Stacked Bar Chart: https://bl.ocks.org/mbostock/3886394
NYC Json Map: http://bl.ocks.org/phil-pedruco/6646844