Packages Used: shiny, stringr, dplyr, leaflet, RColorBrewer.

Project Name: DC\_trees

Files:

1. app.R: Shiny script;
2. styles.css: the file making my app look good;
3. DC\_trees.csv: the dataset used in my app;
4. Urban\_Forestry\_Street\_Trees.csv: the original dataset;
5. wrangling.R: the script processing the original dataset to be the dataset in my app.

Data source: <http://opendata.dc.gov/datasets/f6c3c04113944f23a7993f2e603abaf2_23>

My question: what is the distribution of tree species like in DC?

Usage of the answer to my question:

1. To city municipality, knowing the distribution of tree species can be helpful when doing Urban Landscape Design.
2. To environmental science researchers, this data can be used in research.
3. To educators and students, my app can be a class material and a guide of outdoor study.

How my app answers my question:

1. My app features an interactive map that shows the location of the trees, color-maps the families of the trees, and provides detailed information of each tree in its corresponding popup.
2. Users can input a common name of trees to filter other trees out to get a clear view of their targets.
3. I added a bar plot showing the statistical distribution of trees in DC.

My work flow:

1. I used random sampling in both the app.R and wrangling.R so the result may be different each time these 2 scripts are run.
2. Data wrangling: I first subseted the original dataset to 9 columns (X, Y, OBJECTID, VICINITY, SCI\_NM, CMMN\_NM, TREE\_NOTES, FAM\_NAME, and GENUS\_NAME) and trimmed the spaces at the beginning and ending of the strings. Then I replaced all the empty strings, unknowns, Nulls, and other equivalent values by NAs. I only kept the observations that had complete X, Y, OBJECTID, SCI\_NM, CMMN\_NM, FAM\_NAME, and GENUS\_NAME. I randomly sampled 35,000 observations from the dataset and then renamed X to longitude and Y to latitude to prevent myself from being confused later.
3. Building the app: I loaded the wrangled dataset, added some noise to the locations, and randomly sampled 10,000 observations outside ui and server sections so that those steps are done in the global environment. I decided to color-map the FAM\_NAME, which is the family names of the trees, because it has reasonable number of levels compared to other columns. Detailed documentation of this step is in the app.R file.

YouTube video: https://youtu.be/8yR1DdHPv5k