THE NEW YORK CITY TREE EXPLORER: AN IMMERSIVE EXPLORATION OF OPEN GOVERNMENT DATA

Author: Sankalp (Sonny) Chauhan
Associated Faculty:
James A. Hendler
John S. Erickson

Abstract:

The study conducted in this paper was to determine the effectiveness of open government data in creating an application that everybody can use. The application created shows a lot of information all at once. Using Multi Window Shiny in R is an efficient means to display a bunch of windows all at once. Open government data can be useful as a means of helping people in a community. Anyone who lives in New York City can use our application to inform them about the health of trees in their area, as well as information about the trees through Wikipedia.

Introduction:

The goal of our project was to create a compelling interactive application on the Rensselaer IDEA Campfire. The Rensselaer IDEA Campfire is a multi-user, collaborative, immersive computing interface. The Campfire is an ideal platform for our project as it has a multiuser capacity and allows many people to work at once, it also helps show many different aspects of our project at once as well. Campfire is a desk-height, 10-foot panoramic screen (the Wall) and floor projection (Floor) that users gather around and investigate, maintaining contact with one another with no artificial or virtual barriers

between themselves as they observe and engage with presentations and applications. Two large monitors adjacent to the Campfire complement the integrated Wall and Floor visualizations with appropriate content, enabling investigators to be fully immersed in their exploratory tasks.

Another objective was to learn how to use open government data to learn about the kinds of data that are available from government sources and the many ways it can be applied. The open government data online [1] contains many datasets by agency and datasets by category. In this website underneath the datasets by agency, there is data by the office or agency that maintains that dataset. For our project, the data was utilized from the Department of Parks and Recreation. After going on the New York City Parks and Recreation subpage, the dataset used was the 2015 Street Tree Census- Tree Data. The 2015 Street tree data contains information about trees in New York City from the five boroughs of New York City (Bronx, Brooklyn, Queens, Staten Island, and Manhattan) such as tree species, health of the tree, Latin name of the tree, common name of the tree, latitude and longitude of the tree, and other important information. The dataset for the trees comes from the website [2], as shown in Figure 1, which describes the columns of the dataset and what they represent. This website also contains a Tree Data Dictionary which contains further notes on each of the columns of the dataset. From this website anybody can export the dataset onto their computer. With the open government data, our next task was to create a multi window shiny app in R Studio to show the many different aspects of government data and to link our work with the campfire. Our use of multi window Shiny was based on a package developed by Hannah De Los Santos, a graduate student in Rensselaer Polytechnic Institute (RPI). The pdf description for the multiwindow shiny package is on [3]. The pdf description includes a function mwsApp that runs the Shiny app in many specified windows. Multiwindow shiny is useful for my project because

it helps me show the multifaceted nature of my project and helps create an interactive app that can be shown on the Campfire. Before I used multiwindow Shiny for my project, I did a bunch of tutorials online to familiarize myself with Shiny. Each tutorial contained lessons as well as practice exercises to help familiarize myself with what I was learning. The tutorials are from [4] and are under written tutorials subsection.

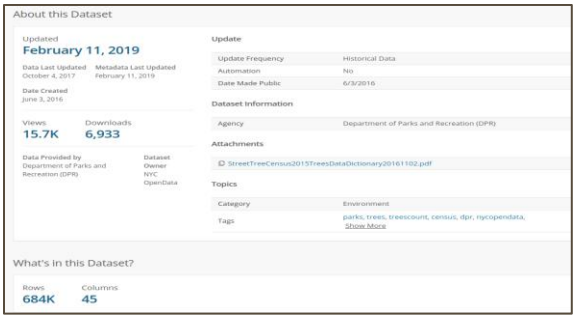

Figure 1: Tree Census Data via NYC Open Data

Materials and Methods:

The dataset used for the project as described earlier came from a Street Tree Census data conducted in 2015, conducted by volunteers and staff organized by the New York City Parks and Recreation department and partner organizations. The dataset was read into my R server and then the dataset was separated by specific boroughs. From the big dataset, smaller datasets were made that information related to each of the boroughs of New York (Staten Island, Brooklyn, Bronx, Queens, and Manhattan). The specific borough datasets did not contain all the features (columns) that were present in the original dataset. The new borough datasets only contained the features that were important in the study: unique tree identifier (tree id), latitude, longitude, borough of the tree, the health of the tree, the Latin and English names of the tree, and zip code of the tree. Afterwards, with the whole dataset the unique Latin names of the trees were made into a data frame. From this new data frame, Wikipedia pages were added corresponding with the Latin names of the tree, making a second column in the data frame. For each of

the borough data frames the Latin name of each row was compared and if there was a match, then that Wikipedia page was added. All the Wikipedia pages for the Latin name of the trees were known as tour text. These Wikipedia URLs were then shown in one of the external monitors in the Campfire. The dataset contained unique latitude/longitude points for each tree. With these unique latitude/longitude points, valid Google Street View URLS were created. These were shown as panoramic views on the Campfire wall. Afterwards leaflet was used to create an interactive map surface on the Campfire Floor. The map contained the data points of all the trees present in the dataset colored by the health of the tree. If a tree has good health it is colored by a green dot on the map. If a tree has fair health, then there is a yellow dot on the map. If a tree has poor health, then there is a red dot on the map. An example of how the map looks like with all the data points is shown in Figure 3. In order to create this multiwindow Shiny app, the CampusTour.R was used as a basis. This was a previous project done which utilized Shiny as well and we used that as a basis for our project. The code for the Campus Tour project is found on the website [5]. The code was modified for our project, but the structure remained the same. This previous project was also shown on the IDEA Campfire as well, and the code was separated into many parts needed such as Controller, Wall, Floor, Left External Monitor, and Right External Monitor.

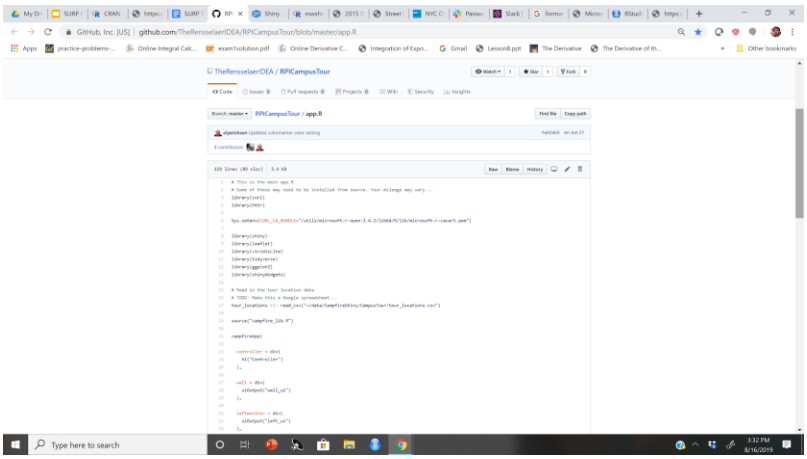

Figure 2: Display of Campus Tour.R code

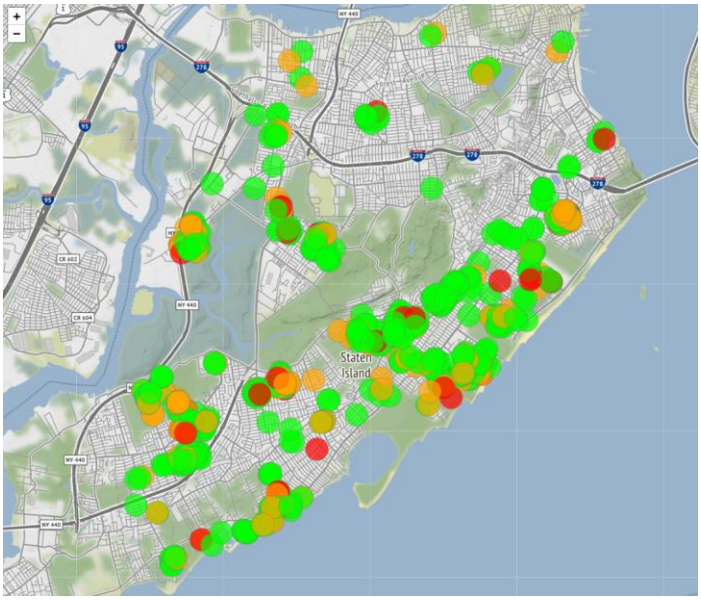

Figure 3: map of New York City with data points representing health of trees

Results and Discussion:

The results of our project include a multiwindow shiny app shown in the Rensselaer Idea Campfire. Our app helps us do many things at once. When one clicks on a datapoint shown on the map displayed on the Campfire Floor, then it displays the panoramic view of the tree using Google Street View shown in Figures 5 and 6. In Figure 7, there is a more detailed panoramic view of the street. There also is a Wikipedia page

about the tree on one of the external monitors. The other external monitor contains a controller panel that can filter our map based on either tree health, tree species, or tree zip code. For example with respect to tree health if one checks off only good for tree health, then only the dots on the map that have green which represents good tree health will come up, and the trees that have poor or fair health will not be shown. The controller also shows the statistics of the trees that are present on the map. For example, if one filters to only have good trees on the map, then the statistics on the controller only have the species of the good trees. The x axis has the species of the tree, and the numbers on the y axis that represents the number of trees. For each species of the trees on the x axis there are three bars, one representing good health (green color), one representing fair health (yellow color), and one representing poor health (red color). Also, on the controller the user can select the borough's dataset and then hit the update button called Load boroughs data to display points from that borough on the map from the Campfire Floor, shown in Figure 4. With this, a user can visualize what tree species are the healthiest from each borough. Knowing which tree species are the healthiest, it helps to plant new trees in order to increase vegetation and wildlife in the area. Also knowing the species of the trees that is not as healthy, it helps to avoid planting such trees.

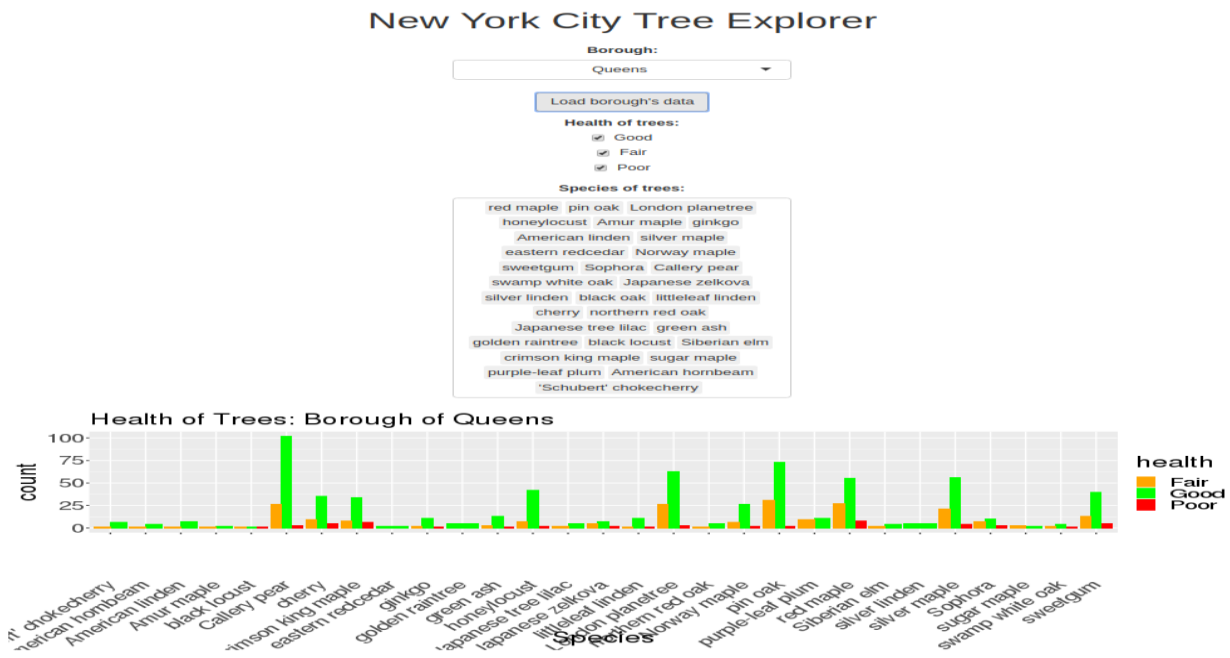

Figure 4: Controller View (one of the external monitors)

Figure 5: Map of datapoints colored by health of tree along with panoramic street view

Figure 6: View of the Campfire map along with panoramic street view and wikipedia page

Figure 7: Street level panoramic view of the tree

Conclusion:

This project helped show the value of open government data as a basis for creating relevant and useful applications for communities. Using open government data from New York City (NYC), we created an app that anybody from the community can understand. Our app can be used to inform people about the conditions of the trees near where they live. The significance of people knowing about the conditions of the trees near where they

live is that the people can either improve the conditions where they live and/or make sure to keep their areas green. If somebody lives in an area with a lot of poor trees, they might want to make sure to take better care of the trees near where they live and try to plant more trees. If somebody, on the other hand, lives in an area with a lot of good trees, they want to make sure that they keep taking good care of the trees. Also, one could investigate which tree species are healthier than others, and increase planting healthier tree species to increase the amount of vegetation and wildlife where they live. If somebody lives in an area with a lot of poor trees, they may want to plant these healthier tree species, over some of the unhealthy ones they planted. Our goal for this project was to create an app that displays as much information about trees as possible all at once. Multiwindow Shiny is powerful for displaying a lot of information all at once at the same time. From the NYC we constructed URIs (uniform resource identifiers) to link to Wikipedia pages and physical location-based web APIs such as Google Street (the street-level panaroma). The IDEA Campfire is a great environment for simultaneously showing different aspects of our application and data. The IDEA Campfire helps show the power of multi window Shiny to show a lot of information at the same time, and to create an application that anybody can use.

Future Work:

There are many possible extensions for our project. One of the many uses is to integrate a search bar on our map. If one wants to search up a famous landmark in New York City such as the Statue of Liberty, then one can type that location in the search bar and investigate the health of the trees near it. One aspect to investigate is how population affects the health of trees in certain areas. Is there any correlation between the health of tree species and population? Another idea is to try to experiment with other

geographically linked open-datasets, and to link other datasets into our app as well. Also, for the controller, one idea is to fix the filtering for the zip codes, as some of the reactivity is still not working 100 percent.

References:

1. New York City (NYC) Open Data: https://opendata.cityofnewyork.us/data/
2. 2015 Street Tree Census - Tree Data: https://opendata.cityofnewyork.us/data/
3. Hannah De los Santos, John Erickson, Kristin Bennett, and Nicholas Thomspon (2019). Package 'mwshiny'.https://cran.r-project.org/web/packages/mwshiny/mwshiny.pdf.
4. Learn Shiny Tutorials: https://shiny.rstudio.com/tutorial/
5. TheRensselaerIDEA/RPICampusTour:https://github.com/TheRensselaerIDEA/RPICampusTour/blob/master/app.R.