Jack Martin and Shoaib Jakvani

CS424 Project 2

14 March, 2022

CS424 Project 2 Documentation

0) Shiny App URL:

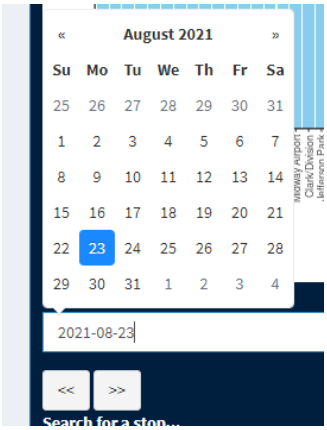
https://sjakva.shinyapps.io/cs424project2/

YouTube URL:

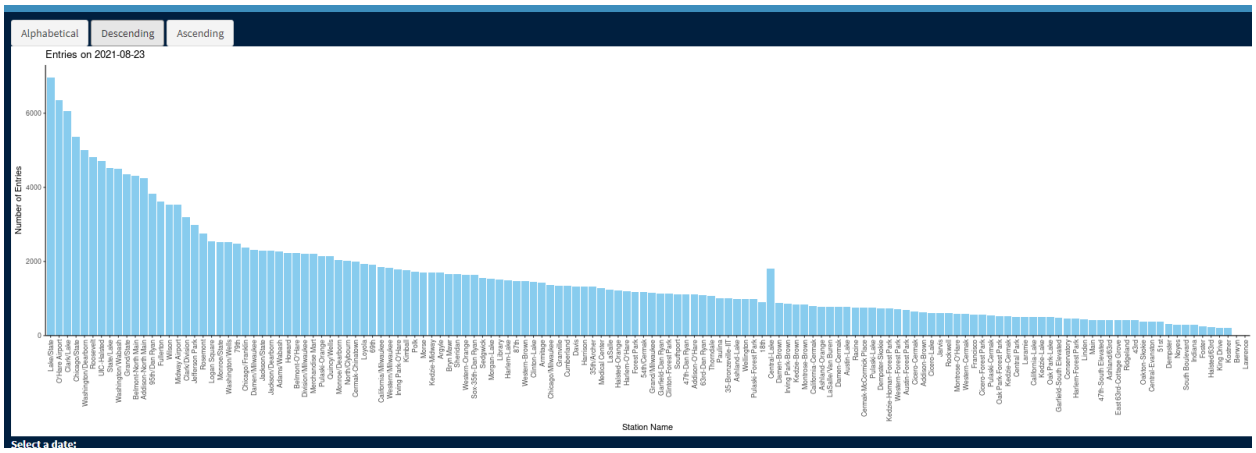
https://youtu.be/61MwgSfPpBE

1) Introduction to app

In this project, we will focus on using R to expand on the visualizations in Project 1 to add in a geographic component, so we can see how the location of the stations in the city may affect what we were seeing in Project 1, which just looked at UIC-Halsted, O'Hare, and one other station of our choosing. To use our project, you would simply need to visit the link and you would be presented with the entries per day on each station in the CTA. Initially, the date that the stations are reporting on is the first day of class for UIC August 23, 2021. Users will be able to select a date through the built-in calendar date selection input section to choose a specific date that they would like to visualize in the bar chart above. Another thing that users will be able to utilize is the helpful arrows that indicate previous or next date.



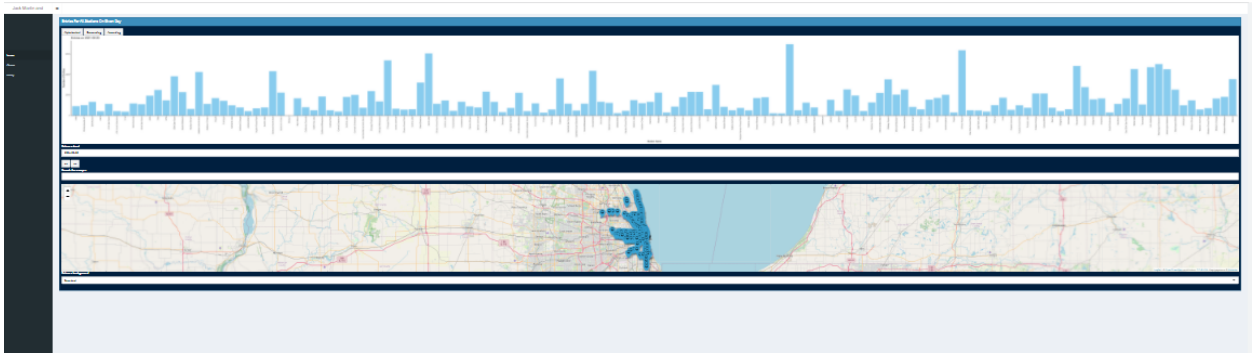
If they push either of these buttons, the graph being plotted will be updated because it's looking for inputs that the user has changed. This allows for user controllability and for them to be able to explore whatever data they wish to view. The graph is also able to be ordered in a variety of ways. By default, this will be alphabetically from A-Z, but there are also options for it to be sorted from minimum entries from that day or from maximum entries per day.



The user will be able to view the data in the form of a data table instead of a graph as well. From another file, we were able to find the latitude and longitude of the various stations we are interested in. Using this, we are able to make a leaflet focused on Chicago's public transportation system. This leaflet would contain markers which indicate where specific stations are in relation

to the world. There is also a drop down menu that lets the user up to three kinds of backgrounds they want depending on how they feel.

Users can visit the shiny app hosted at the link above; the web application is designed to be used (dimensionally) according to the classroom lab, which had dimensions of ***5,760*** x ***1,620*** as a touch screen application with no scrolling. Due to unimplemented window-resizing-responsiveness, the application will be enlarged on normal/common devices.



2) Data manipulation and analysis

We had seven columns for our dataset - ***station_id, stationname, daytype, rides, lat, long, date***. Here is a figure of what a row looks like (RStudio view):

station_id	stationname	daytype	rides	lat	long	date
40830	18th	A	790	41.85791	-87.66915	2017-12-30

Most of the data is available from the Chicago Data Portal with the addition of the location data that includes the latitude and longitude for each station in the original dataset (going by station ID instead of the name). Ideally, you would want to load in all the primary dataset into a dataframe and then go about appending the additional location data from the second dataset onto the primary, joining/merging on station ID. We tried using *merge()* or *rbind()* to append the data but the dataset containing the location only had 300 rows compared to the original dataset that contained 1,000,000+ rows. We weren't able to look up a solution that worked for us aside from

manually pasting in the data into the original dataset (although we had split it by station names to make it easier to keep track of). We also "cleaned" the *date* column by reformatting it from "month/day/year" to "month-day-year". Furthermore, for the leaflet markers, we wanted to add a list of station names (and ideally create a map or dictionary with the key-value pair being the station-location) for the user to select and look up, so we took a subset of just the column names from the original dataset to then use for later, which was unimplemented.

3) Github repo and instructions / Software

https://github.com/sjakva/cs424project2

The preferred method of running this application is to follow the Shiny application link. Users can click on the repository link above to the GitHub repository if they wish to then download the project files and run the application locally (our application was developed using version RStudio 2021.09.2 Build 382). You can click on the green **Code** button → **Download ZIP**. After extracting the file, you can now open and run the application in that directory using *app.r*, which contains all our code. The Xstation_*nameOfStation*.txt files are TSV text files that took fairly long to parse and split through. We attempted to load in the two data files and append the location columns using the station_id and MAP_ID columns to "merge" the data; this attempt failed due to the difference in rows between the datasets. We weren't able to find a solution that worked so we resorted to manually pasting in the two columns for every single station file (we had wrote a function to split each stations data in its own file which made things easier) by looking up the ID column. Later on, we realized that the application was not deploying at all on Shiny due to some unencoded character and that it needed to be UTF-8 encoded otherwise it

wouldn’t work. We didn’t receive any error messages that specified where the error was and even though we had code to go through the data and encode it, it wouldn’t work. We used an online converter to convert all the station files and after replacing the original TSV’s with the encoded-TSV’s, the application deployed. The *backups* folder contains the original TSV files in case the converter ended up failing.

4) Interesting notes

Unfortunately, we were unable to complete the part of the project that asked us to compare and contrast different aspects of the data in comparison to the latitude and longitude coordinates that we associate with each station.