

Shoaib Jakvani

CS424 Project 1

14 January, 2022

CS424 Project 1 Documentation

0) Shiny App URL:

<https://sjakva.shinyapps.io/cs424project1/>

1) Introduction:

This is the documentation for Project 1 for CS 424 that focuses on using R to visualize data using Shiny/Shiny Dashboard and ggplot/lubridate to integrate an interactive display for users. The dataset is on Chicago L riders over the past twenty years (2001 to 2021) and can be found from the Chicago Data Portal. It should be noted that this project is unfinished as it required just a few more hours to implement the interactivity of the Shiny application fully and functionally. For now, the data from 2021 allows on deploying the application.

This project required importing and filtering data, using R to manipulate it, and then using the Shiny dashboard to host an interactive application. Shiny is an R package that allows one to build and host interactive applications straight from R. The other important package we used was ggplot; another powerful package that aids users in producing graphs using three components: a dataset/dataframe, aesthetics (labels, colors, shape, etc), and geometry (bar plot,

line, histogram, etc). Additionally, we also used lubridate to parse through dates in our dataset to more easily work with and manipulate data.

For our data over the Chicago L riders, we focused on the UIC-Halsted and O'Hare stations, as well as a station of our interest/choosing. It was interesting to analyze and visualize this dataset as a majority of UIC students use the L station often and it brings a better sense of understanding the dataset within its context. For example, it made sense to us that tickets/ride entries very drastically dropped towards 2020-2021 as the COVID-19 pandemic shut down a lot of public facilities and services, with the CTA being one of them. For the station of my choosing, I chose the UIC Medical District / Medical Center station as that's another stop I often take for commuting between work and school. I wasn't able to implement the last portion of the project (after we were supposed to implement the interactivity) but that station should have also been available as one of the stations to look over and compare with UIC-Halsted and O'Hare.

2) Application Overview:

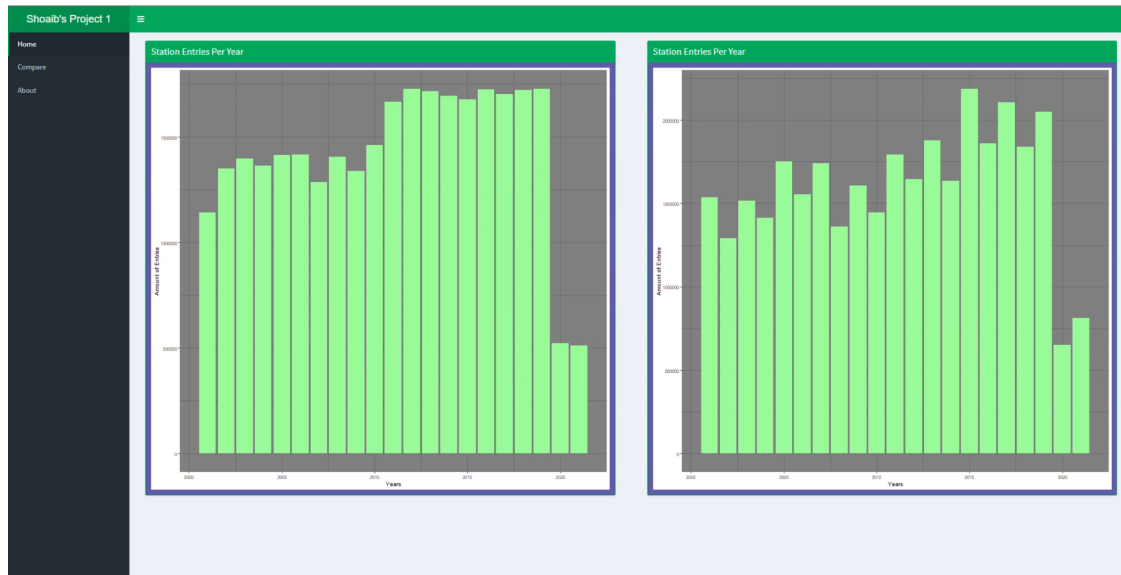


Figure 2.1 Home Page - Landing page of application displaying number of station entries over the years for UIC-Halsted and O'Hare.

Figure 1.1 displays the landing page of the Project 1 application. Upon loading, the home page displays station entries, or rides, over the past twenty years ranging from 2001 to 2021 for UIC-Halsted and O'Hare on the left and right, respectively. The final touches after all the implementations were to be added was to fix the resizability of the Shiny widgets to fill the empty space upon different browser dimensions.

I chose to go with a green color scheme and was going to implement final touches in a complementary color like orange or tan to make it easier on the eyes of the user. I initially added the purple background to distinctly identify the boundaries of the container I was working with, and because purple and green tend to work together.

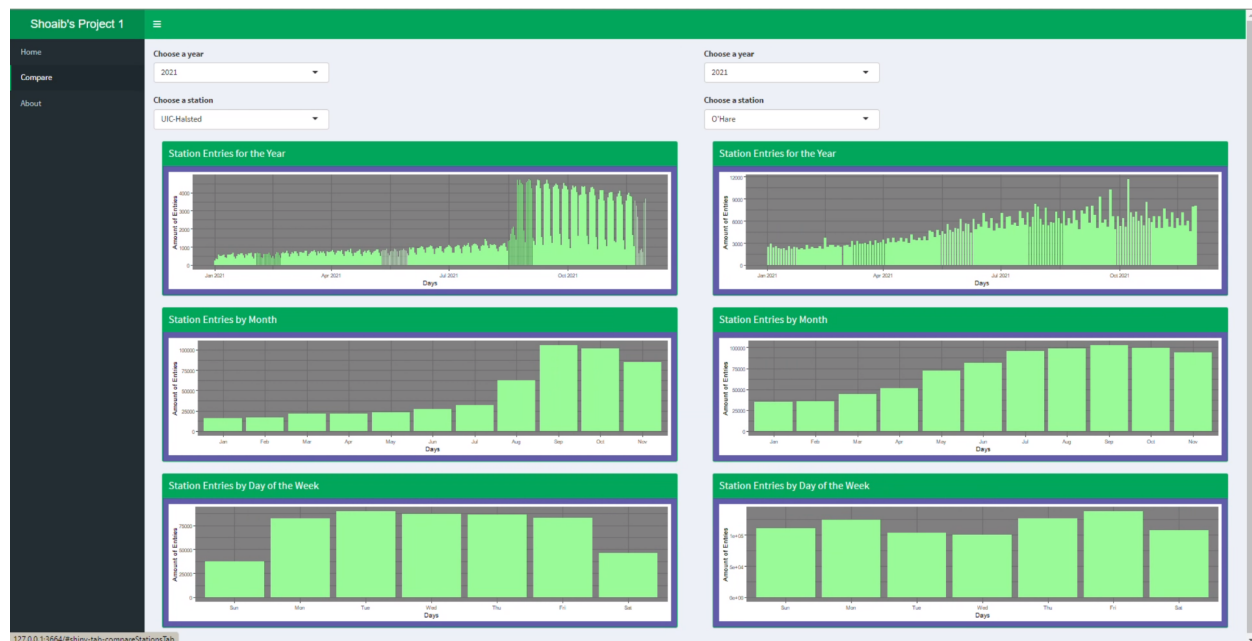


Figure 2.2 Compare Page – Displays a split comparison page allowing users to filter year and station name.

Figure 1.2 displays the Comparison Page consisting of two divisions of data; users can choose which year and station to filter through to then display charts for. The only dataset I could get working was the initial data we worked with from 2021, as implementing the reactive Shiny variables was what I got stuck on.

Given a functional implementation of this interface, the user would be able to choose any year from 2001 to 2021 and either the UIC-Halsted, O'Hare, and/or the UIC Medical District, and then compare the results side by side. Three bar charts are generated for each year and station coupling: the first shows station entries over the year, the second shows them divided by month, and lastly we have station entries per day.

I implemented a consistent color scheme across all the bar charts although there could have been more filtering/selecting of data through colors/visualizations to make data analysis easier/more efficient.

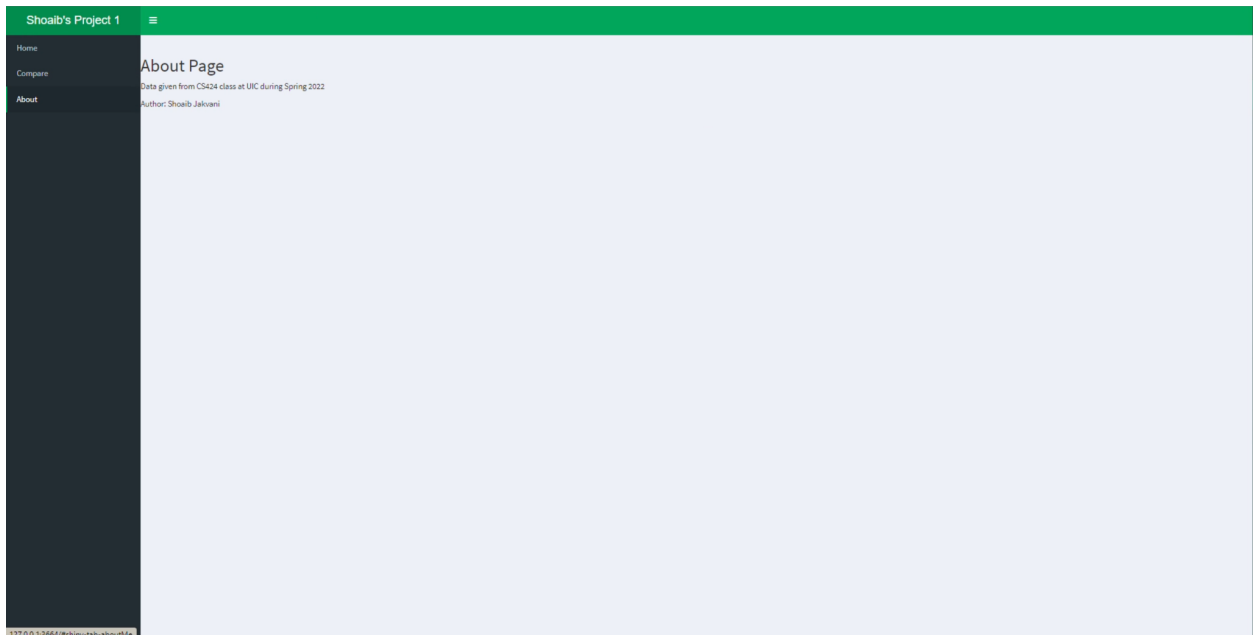


Figure 2.3 About Page – Displays information about the application.

We were to add an about page displaying information about the application, its use, where the data was derived from, etc. I was going to implement this last after finishing the interactivity but I was unable to get to it. The data we worked with was given from the [Chicago Data Portal](#).

3) Data Analysis:

Given that I wasn't able to fully implement the interactive component and analyze my data through that lens, my data analysis section of this project is quite bare.

	station_id	stationname	daytype	rides	nDate
1	40350	UIC-Halsted	W	3664	2021-11-30
2	40350	UIC-Halsted	W	3459	2021-11-29
3	40350	UIC-Halsted	U	715	2021-11-28
4	40350	UIC-Halsted	A	914	2021-11-27
5	40350	UIC-Halsted	W	812	2021-11-26
6	40350	UIC-Halsted	U	636	2021-11-25
7	40350	UIC-Halsted	W	2667	2021-11-24
8	40350	UIC-Halsted	W	3261	2021-11-23
9	40350	UIC-Halsted	W	3540	2021-11-22
10	40350	UIC-Halsted	U	1047	2021-11-21
11	40350	UIC-Halsted	A	1668	2021-11-20
12	40350	UIC-Halsted	W	3755	2021-11-19
13	40350	UIC-Halsted	W	3795	2021-11-18
14	40350	UIC-Halsted	W	3971	2021-11-17
15	40350	UIC-Halsted	W	3865	2021-11-16

Figure 3.1 UIC-Halsted data – Snapshot of the data tables used for the UIC-Halsted and O'Hare L Stations over the years.

Figure 3.1 displays the typical data structure used throughout our data manipulation. The columns given were *station_id*, *stationname*, *daytype*, *rides*, and the *date*. These were consistent across all the datasets used (for UIC-Halsted, O'Hare, Medical District) and were going to be selected/sorted from the overarching datasets by year to then output data to the bar charts on the Comparison page.

The **station_id** is the identification number of the L station. The **stationname** is the name of the station. The **daytype** is the type of day, denoted by W, U, or A (W = Weekday, U = Saturday, and A = Sunday or a holiday). The **rides** column identifies the number of entries that day. Lastly, the **date** column denotes the day of the records.

4) Conclusion:

This project will be hosted on [Github](#) where users can then use the Shiny application. Users who wish to download the repository and deploy it through R should note that the application was built and deployed on a system using R 4.1.2 and RStudio 2021.09.