EDA of MBTA data

Tianjian Xie

2022-12-13

Abstract

Nowadays, Google map is a widely used app. People can use this app to check how to reach to their destinations and how long will it takes. Usually, the app will provides several transportation methods to the users to choose and give how long will each methods take. It’s pretty convenient and becomes very popular in a short time. However, when we are using Google map, we can always meet some problems. The most common problem I met is the cost time the app provides me always inaccurate. It’s too hard for individuals to fix the problem, but we can study the data and see if we can find if there is any ways we can change the code to make improvement.

Introduction

This project is aimed to build a shiny app that ask the users to select their preferred transportation and give them a virtual result that how long will the chosen route take based on the Boston map. To reach the goal, first, we need to check the data provided by MBTA.

Input Data

Data Cleaning

Select one week from each month, combine the data sets, and clean the useless variables.

EDA

Now have some data visualization based on the selected data.

Travel time between two stops of different routes in seconds

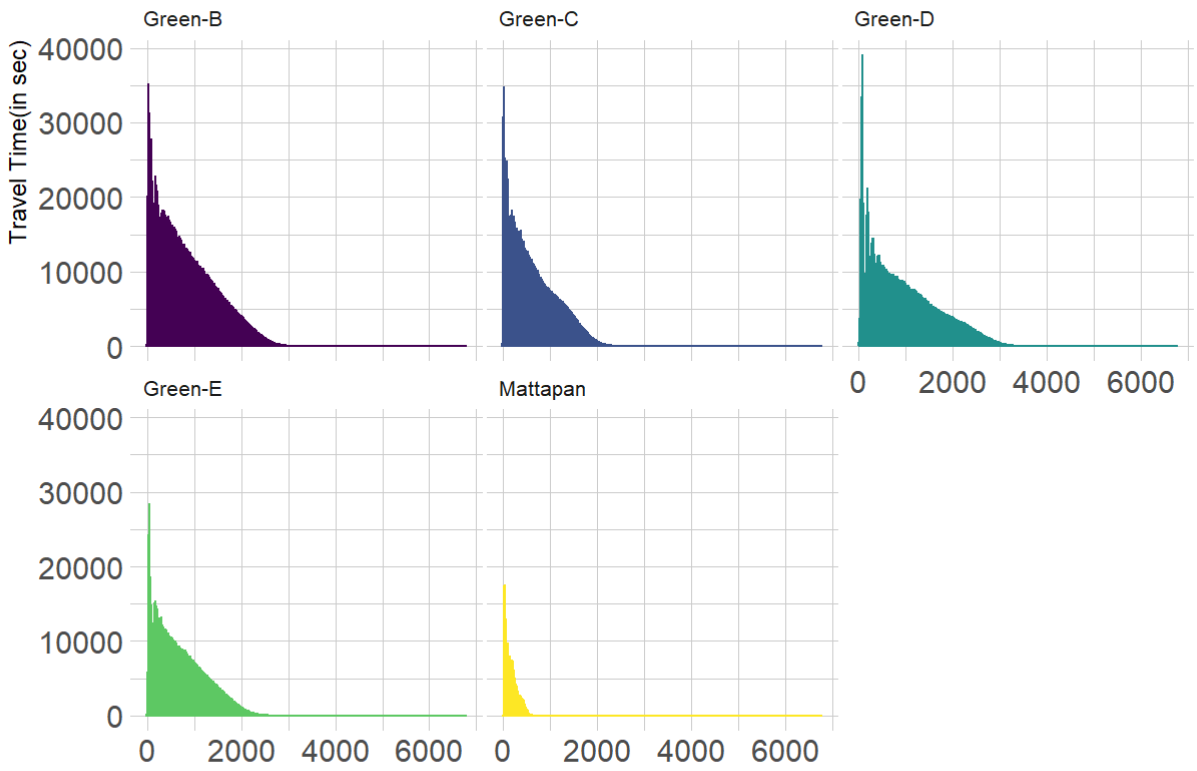

Figure 1: Travel time between two stops of LR routes in seconds

Figure 1 shows how long will each routes in LR system take to travel between two steps. This can help us to see if any routes has very large distance between two stops on its path.

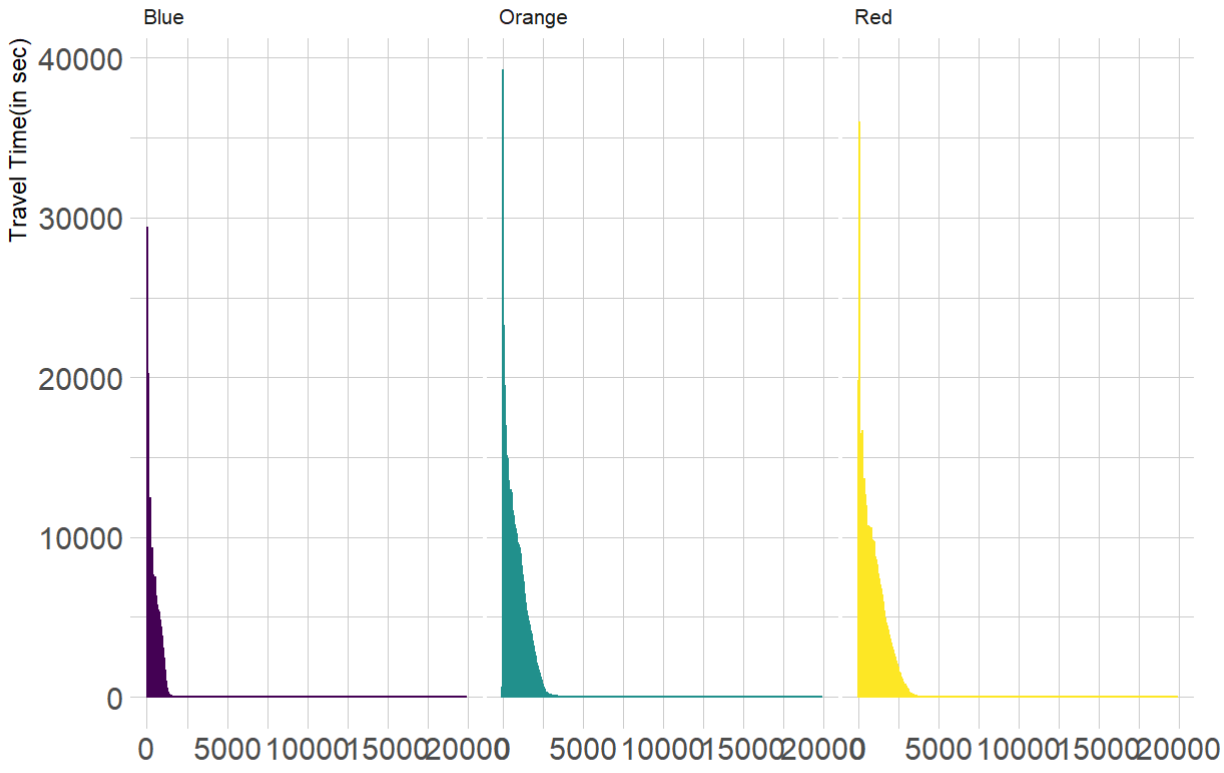

Figure 2: Travel time between two stops of HR routes in seconds

Figure 2 shows how long will each routes in HR system take to travel between two steps. The feature is similar with the figure 1.

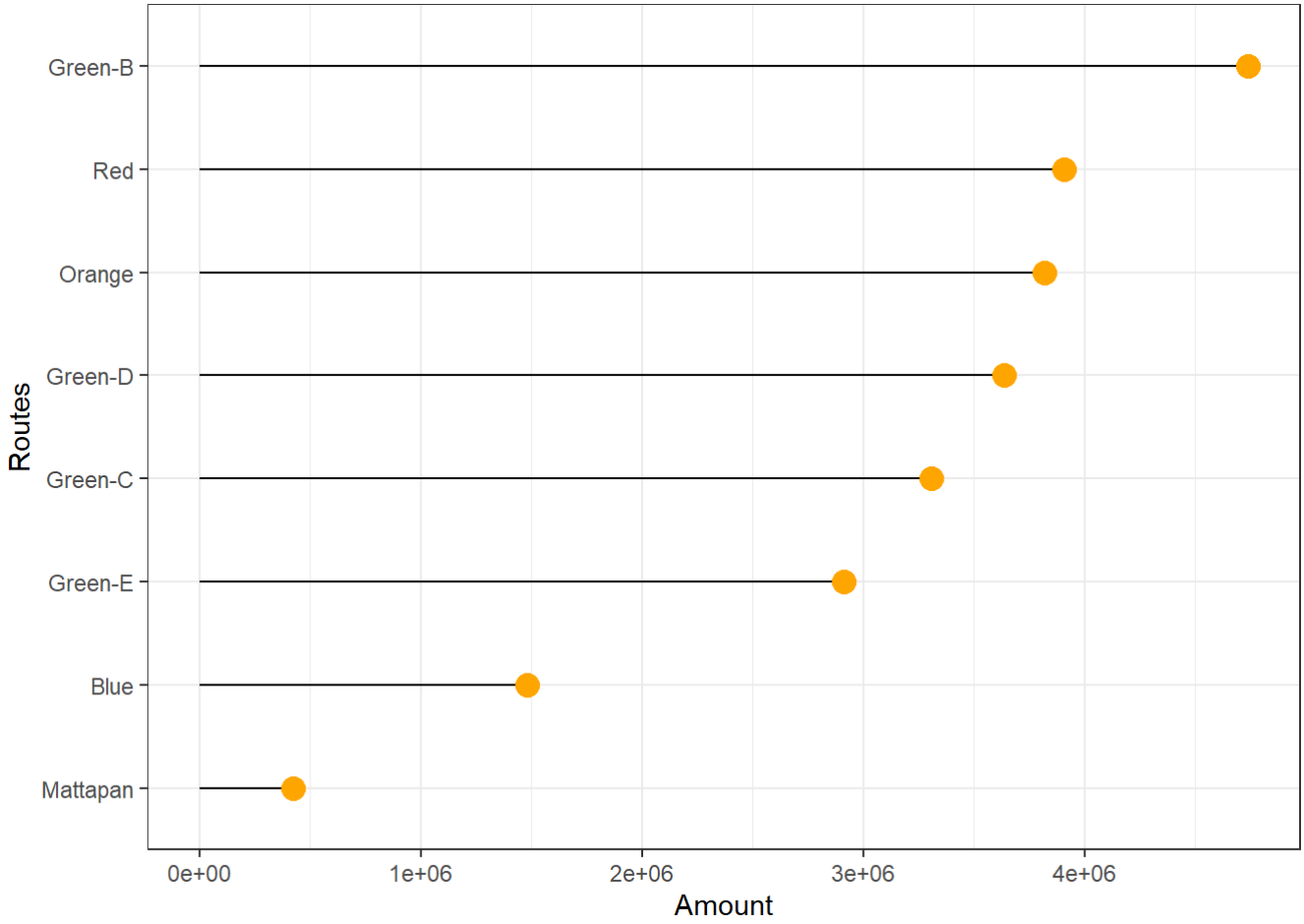

Figure 3: Amount of different routes ranked

Figure 3 shows the amount of different routes operated in selected days during passed year. This can show us how busy the route is, for some route that running frequently, it might means the estimated passenger flow volume is high.

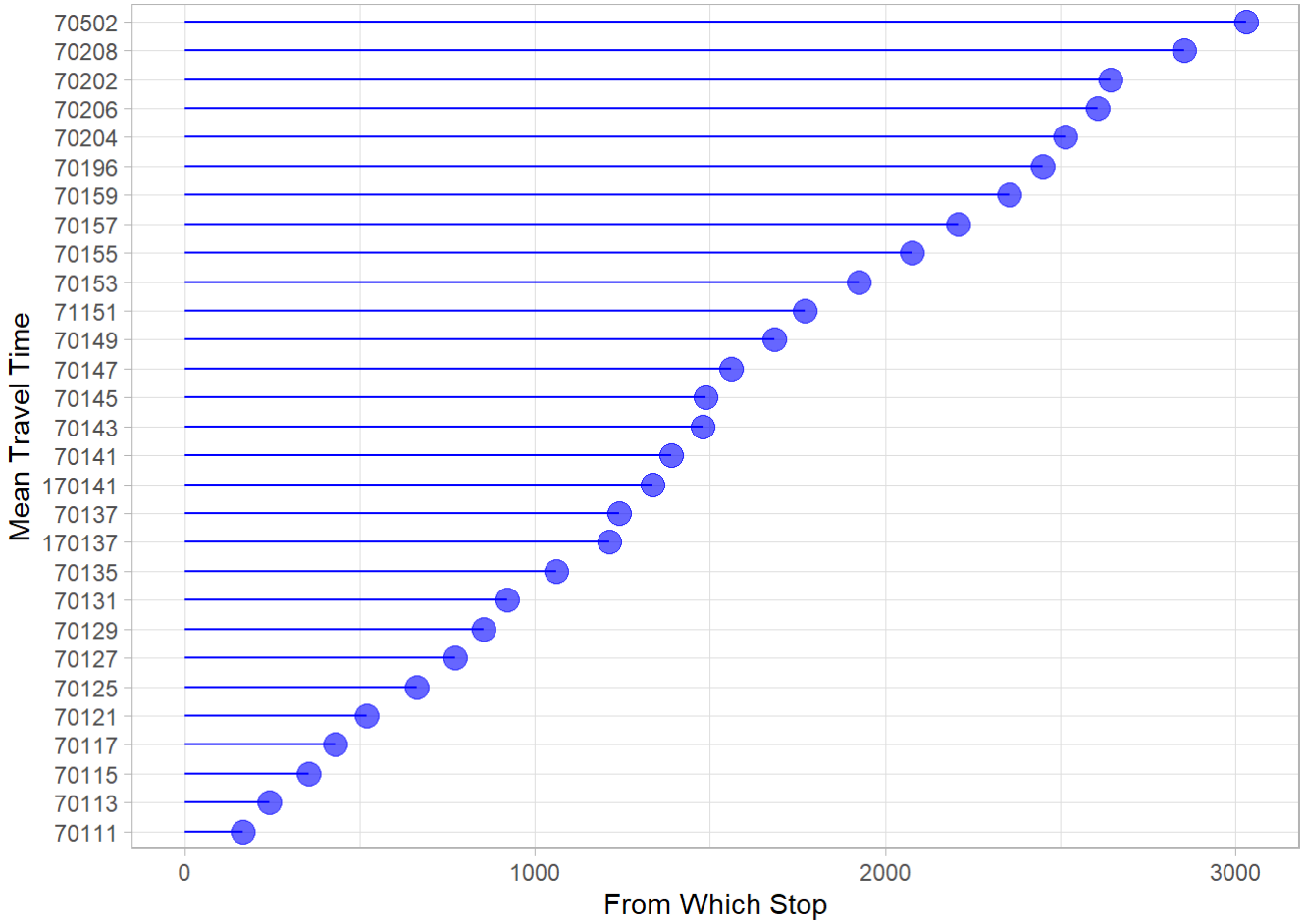

Figure 4: Average spending time from different stations to 70107

Figure 4 shows the average spending time from different stations to stop 70107. This can shows that the distance between stop 70107 and the connected stations.

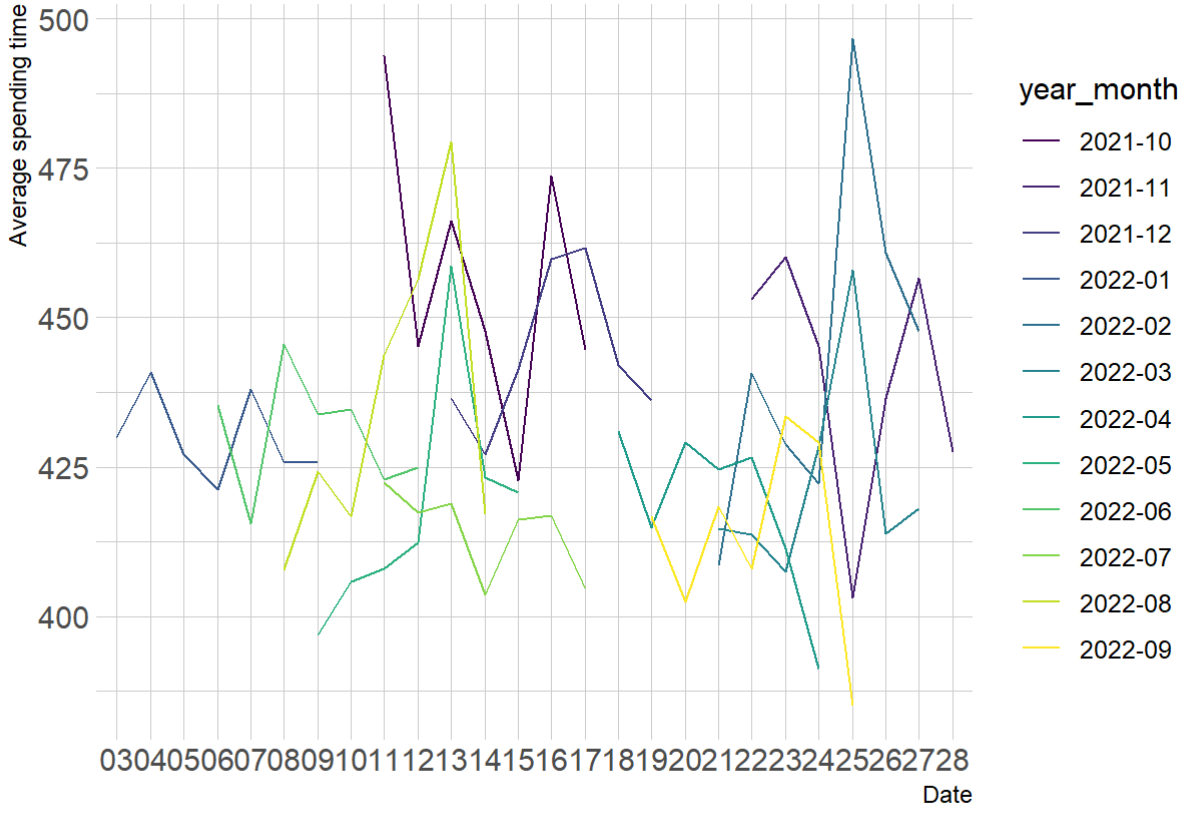

Figure 5: Average spending time between stop 70117 and 70107 on different date

Figure 5 shows the average spending time between stop 70107 and 70117 on different date. The feature is to find the time spent between stop 70117 and 70107 so we can see if there is large difference between different days. From the result we can see that 2021-10-11 and 2022-02-25 are the two days that the train took longest time to travel between this two stations. From the weather forecast(World Weather), 2021-10-11 is a regular day, and 2022-02-25 is the only day in that week that has shower sleet. So we can say 2022-02-25's delay is reasonable. However, for 2021-10-11, we still need to find what caused the problem.

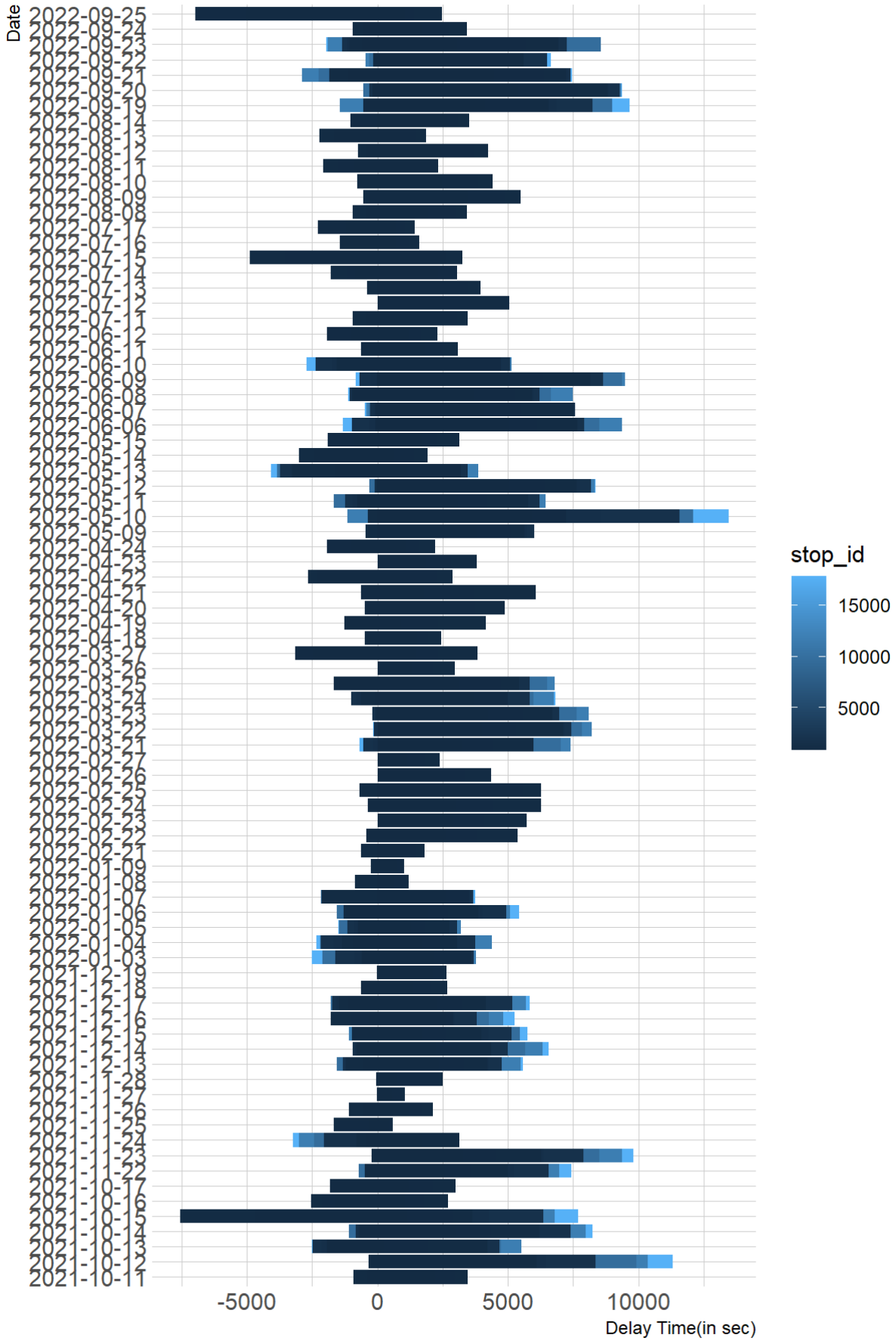

Figure 6: Average delay time between different stops of route 57 at different date

Figure 6 shows the average delay time between two stops of route 57 at different date. Positive means there is a delay, and negative means the bus arrives early. As the figure shows, mostly, the bus arrives late, which fits my personal experiences, for some days, there is extremely late or early.

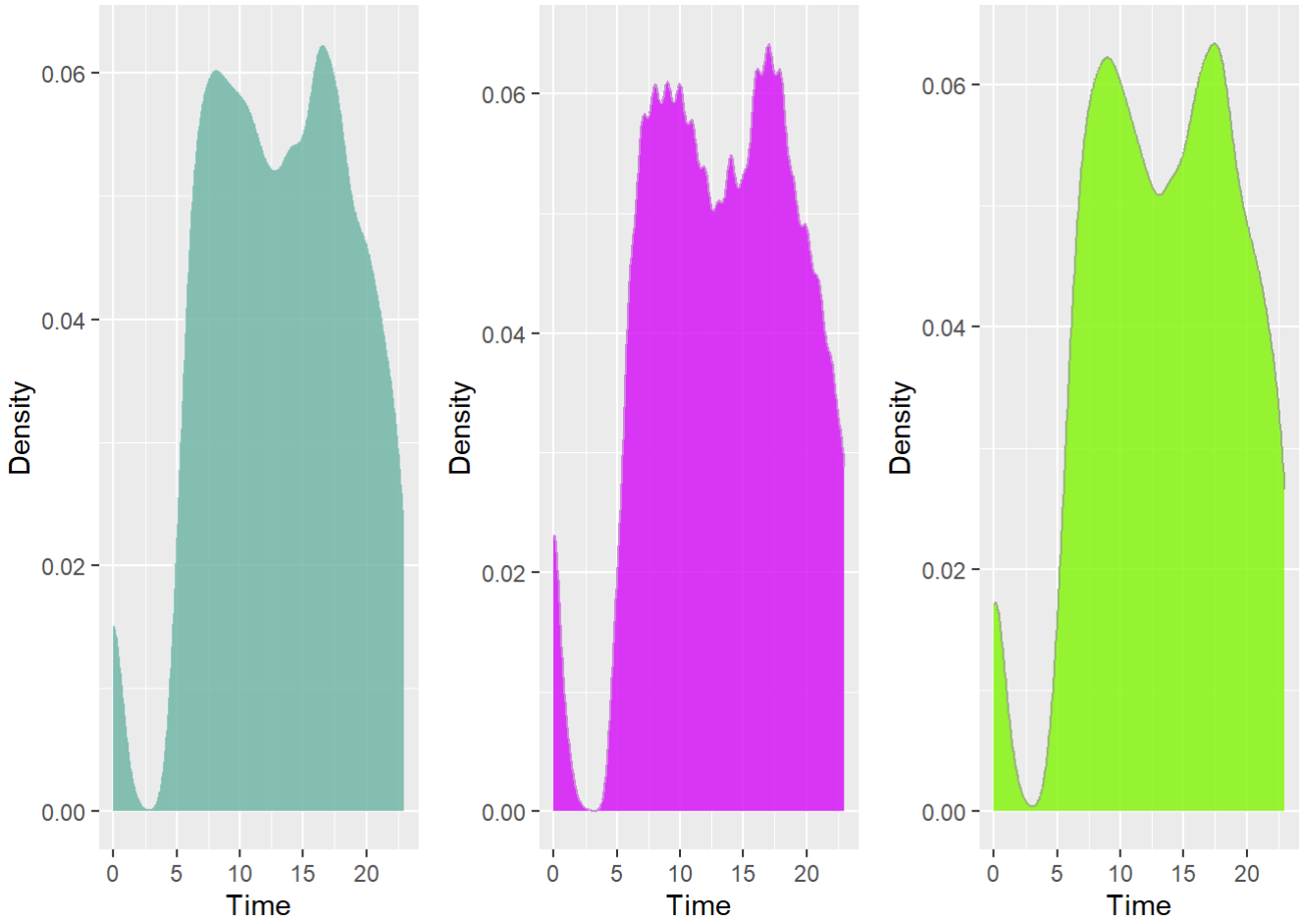

Figure 7: Start time, time enter the midpoint stop, and end time of route 57

Figure 7 shows the density of starting time, time the route 57 enter the midpoint, and the ending time. We can see the graphs are basically obeying the similar trending. Which means there is not lots of critical situation of the running of route 57.

Conclusion

So from the data visualizations, we can see the figures have already explained some of the questions that confusing us. The extreme weather can be a affect influence to the delay of the transportation. However, sometimes when the weather is common, it still can be late. The reason can be the traffic jam, or car accidents for the bus route, and can be the railway conditions for the trains. Based on my personal experiences, I think the flow amount of passengers is also a big deal in causing the delay. If we can consider about all these influence factors during writing the code of the app, we can have a huge improvement on the final present result.

Citation

World Weather https://world-weather.info/forecast/usa/boston/october-2021/ (https://world-weather.info/forecast/usa/boston/october-2021/) https://world-weather.info/forecast/usa/boston/february-2022/ (https://world-weather.info/forecast/usa/boston/february-2022/)