

# EDA report for MBTA data

Jing Wu

2022-12-17

## **Instruction**

### **MBTA**

MBTA(Massachusetts Bay Transportation Authority) is a public transporation services in Greater Boston. It contains rapid transits, bus, commuter rail and ferry routes.

### **Main ideal**

The main ideal of the report is EDA towards the MBTA data.

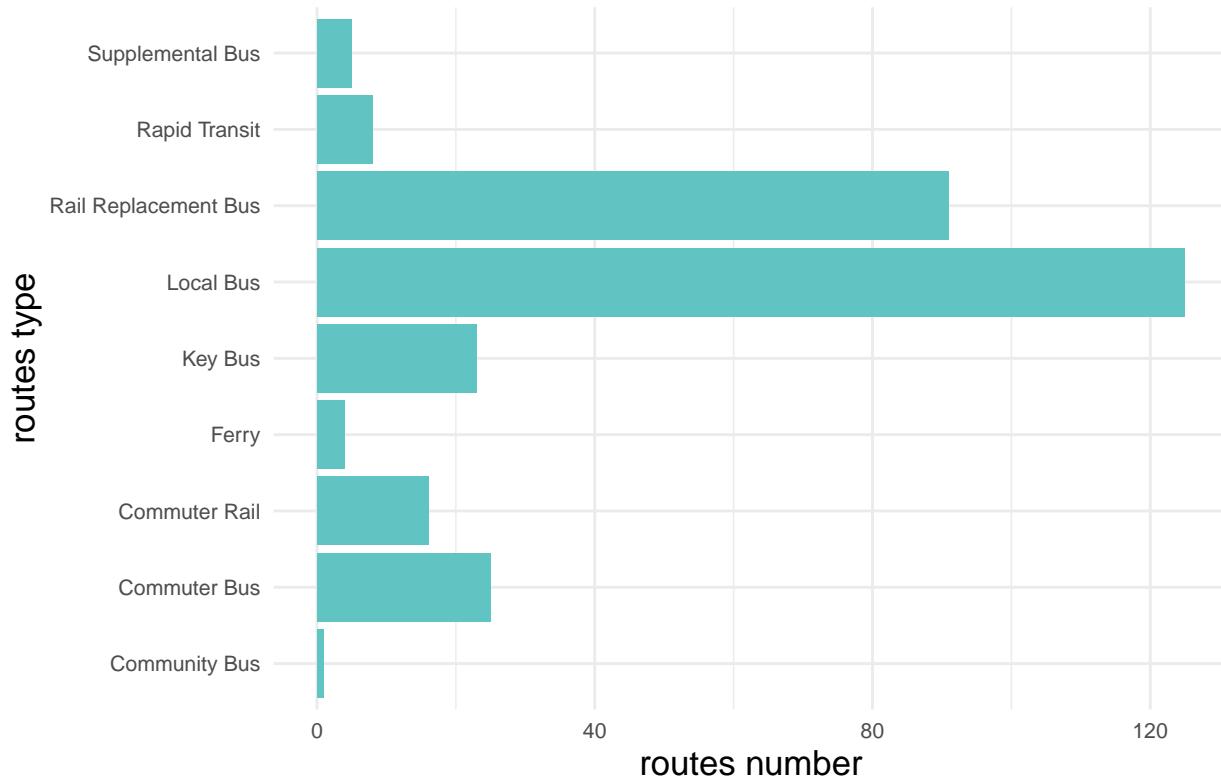
### **Data resource**

Data used in the report is data cleaned from raw data of MBTA archive and MBTA Open Data Portal. The data is all filtered by date(1 random week for each month), and the range of months used is from Nov. 2021 to Oct. 2022.

### **Routes and calendar**

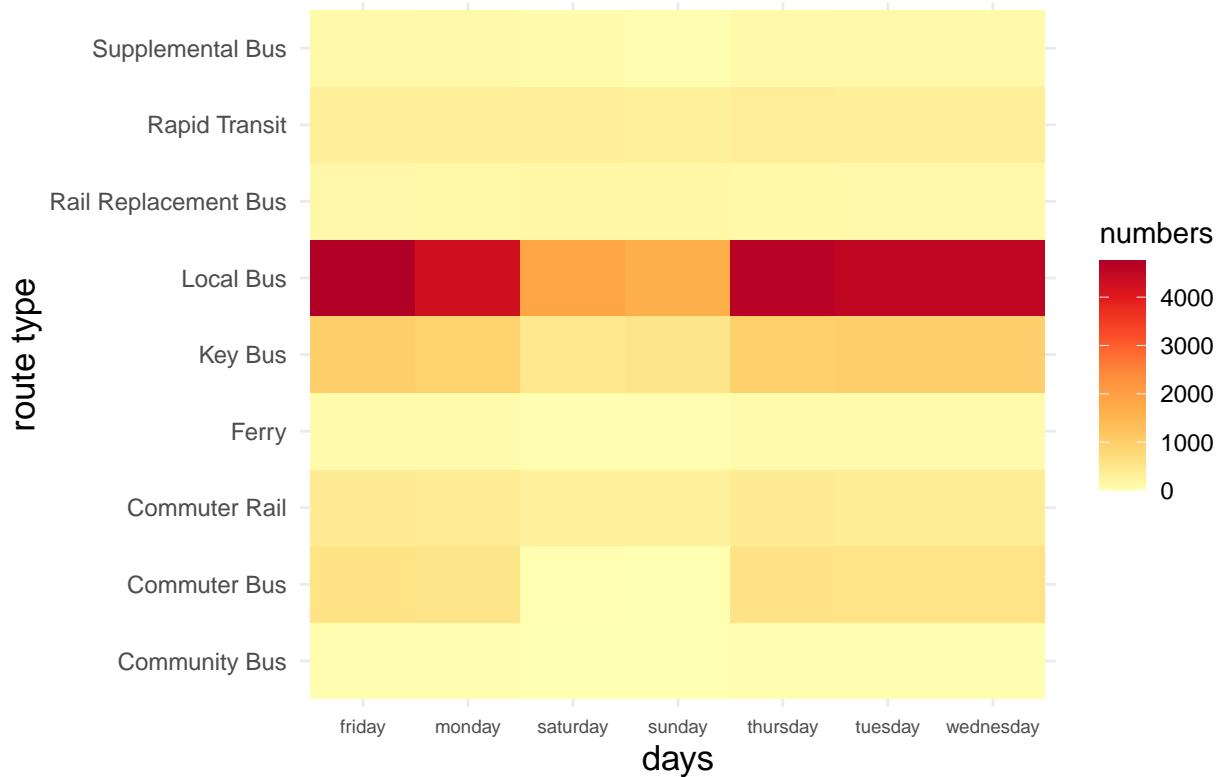
First, we plot the stops and routes for the data.

## Routes number vs Routes type



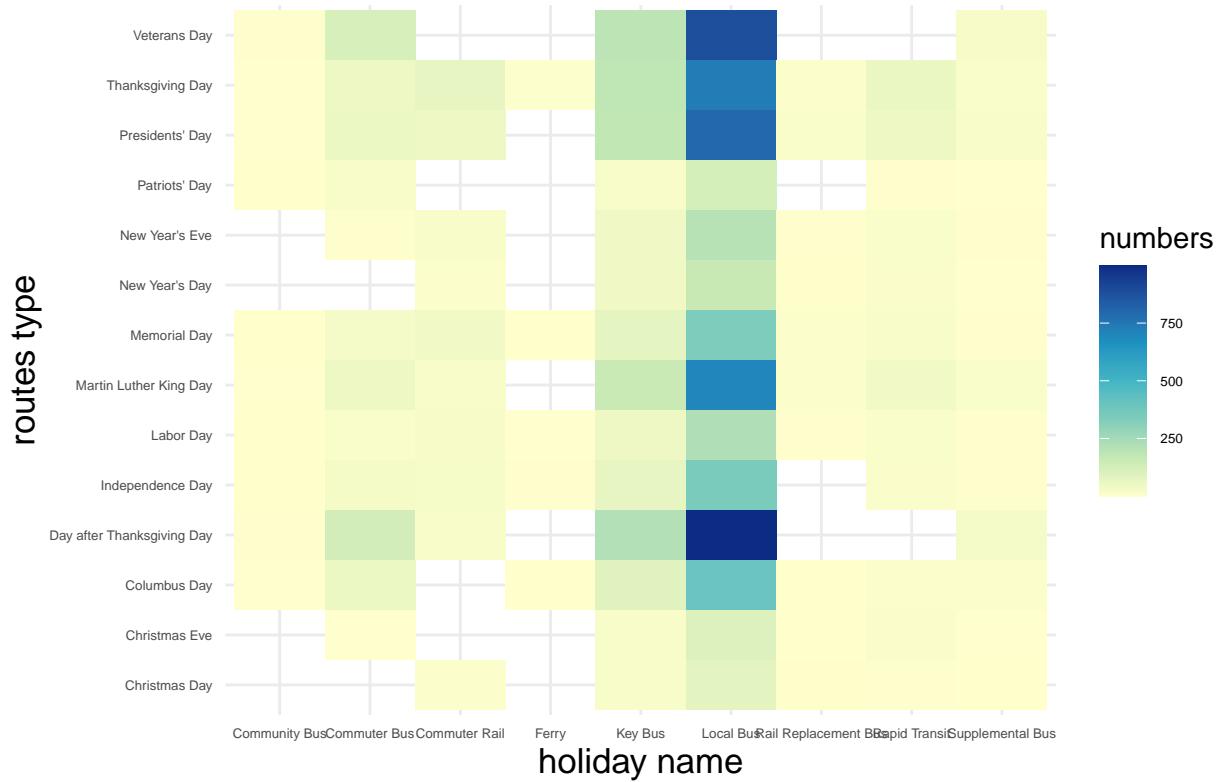
From the plot we can find that the number of local bus and replacement bus seem to be most, and there are also a lot of key bus routes.

## Routes type numbers vs days



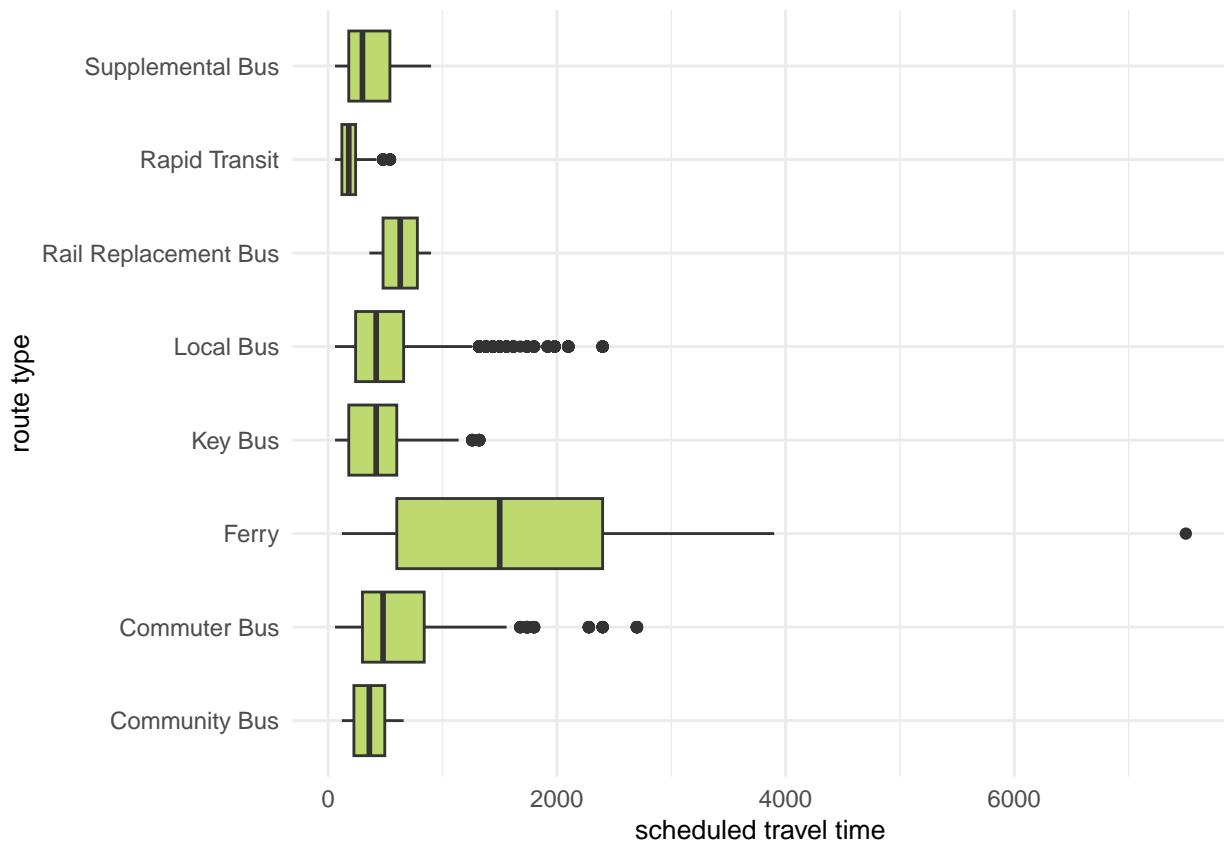
The heat map shows the information saved in calendar.txt. From the plot, we can find that the number of routes scheduled on weekdays is significantly higher than the number on weekends. Also, for weekdays, there are more scheduled routes on Thursday and Friday.

## Routes in holiday vs routes type



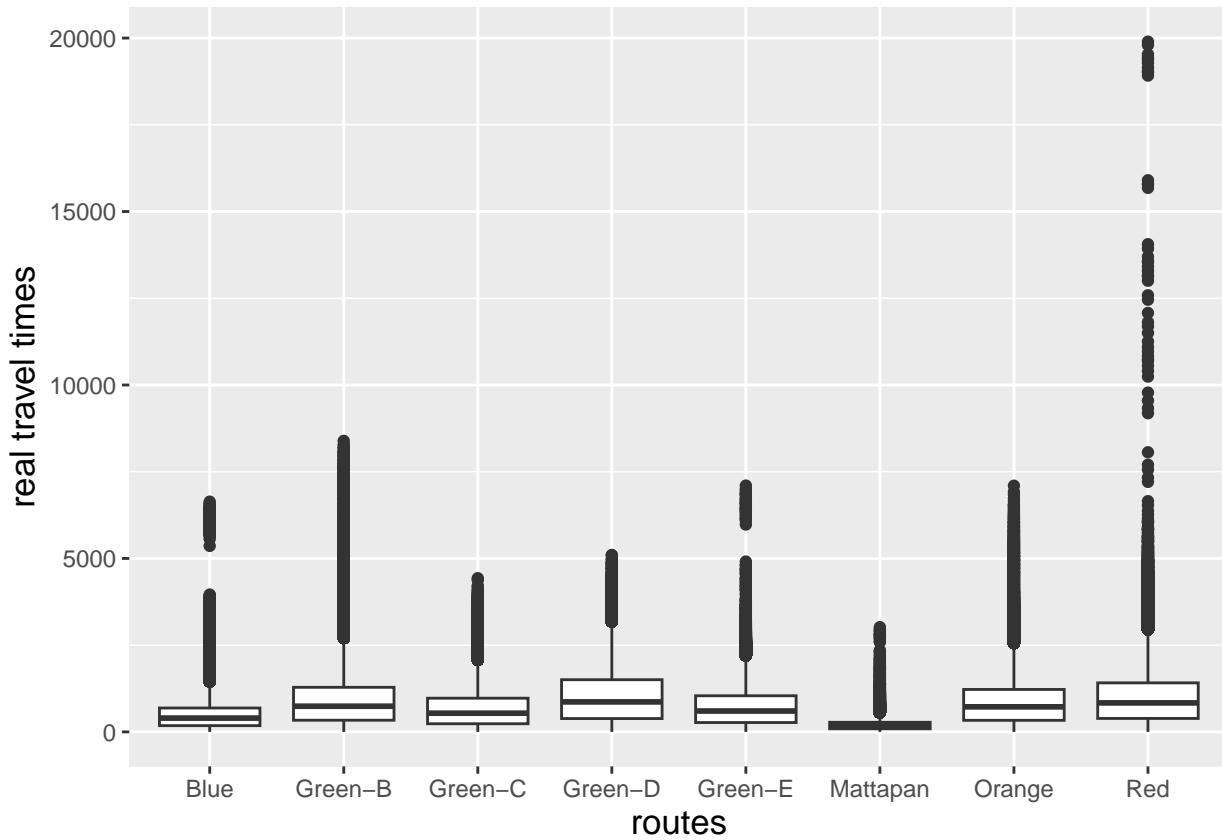
From the plot, we can observe the number of routes for holiday. In general, the routes in Veterans Day and Day after Thanksgiving day are most in data sample, and () do not provide service in most holidays.

## Scheduled times



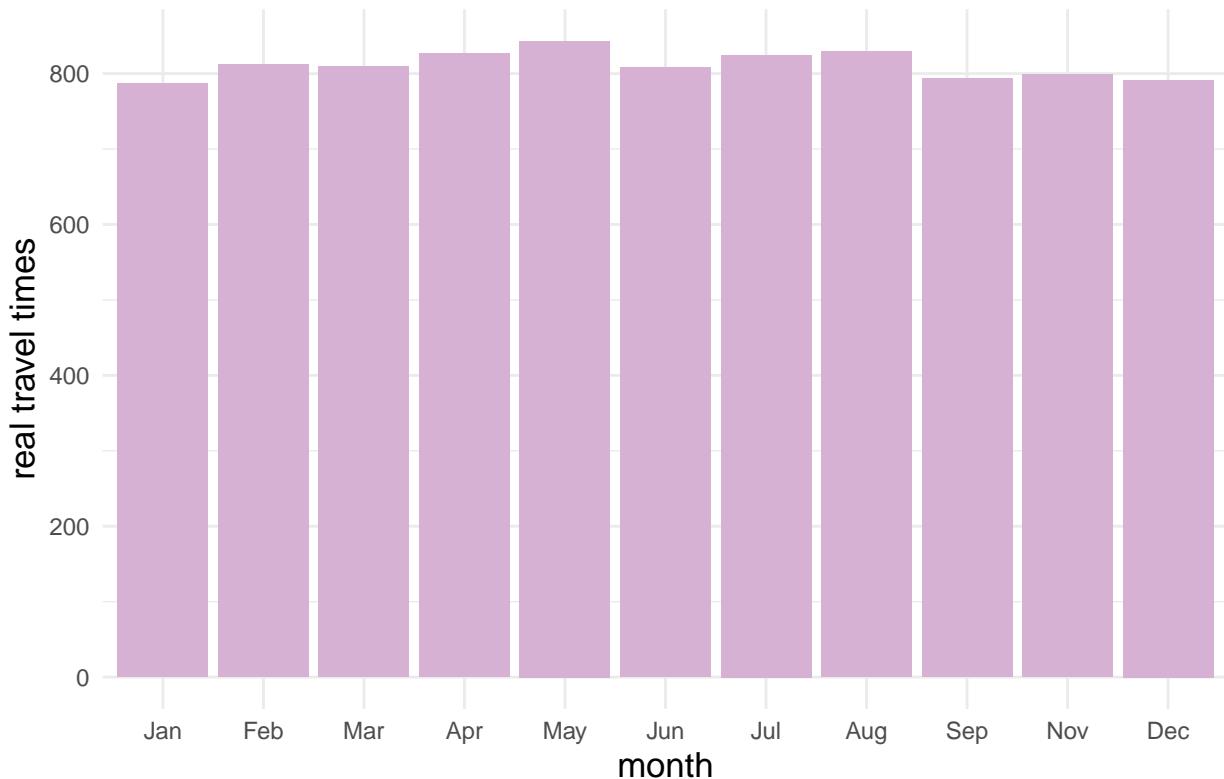
From the plot, we can find the travel time between two adjacent stations. The ferry travel time will take much longer than others, and commuter buses tend to have longer travel time than key and local buses.

## Reliability of rapid transit

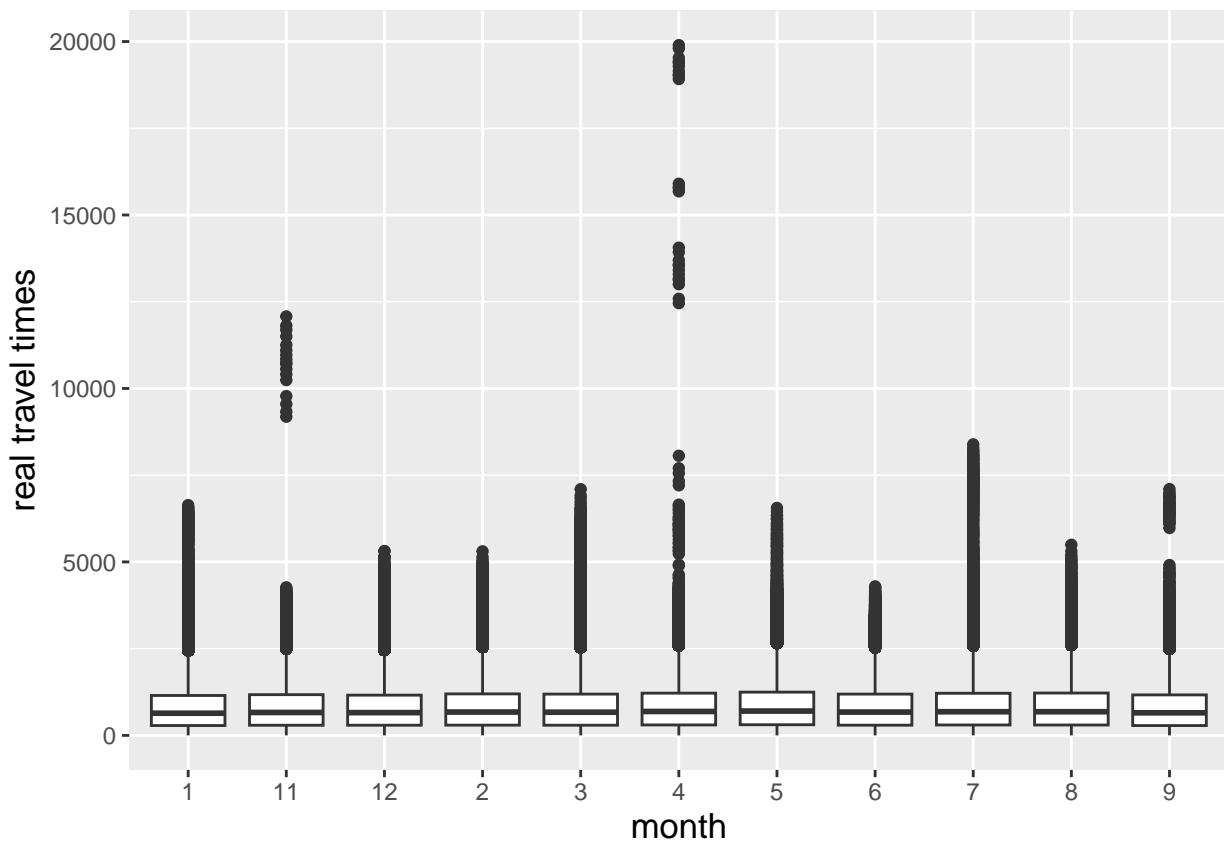


The plot is about the real travel time distribution of different lines in rapid transit. Although the mean values of green, red and orange lines are pretty close, the red line has more high value than other subway lines.

## Travel time for rapid transit vs month

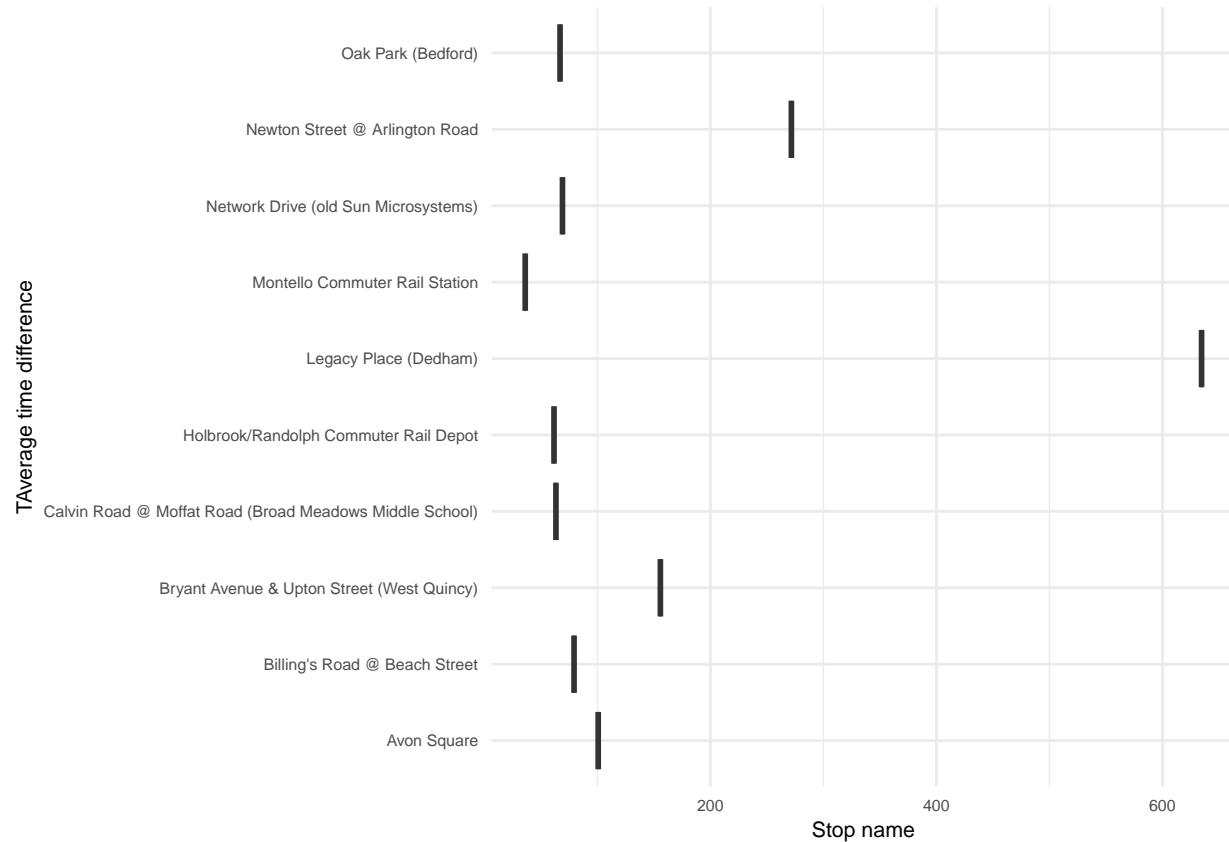


From the plot, we can find that the average travel time is about 800 seconds, and in May, rapid transit tends to have higher average travel times.



From the plot we can find that April had more unusually high values than any other month.

## Reliability of bus



The plot shows the 10 bus stops with the highest average lag from the scheduled time.

## Conclusion

In conclusion, most of the routes in MBTA are bus routes, and in general, the travel times for the ferry are much longer than other transportation services. For rapid transit, the red lines take longer time than other routes, and for buses, some stops (especially Legacy Place) usually come with a longer delay.