# 615_final_eda

Yingnan Lyu

2022-12-17

## Introduction

The Massachusetts Bay Transportation Authority (abbreviated MBTA and known colloquially as "the T")is the public agency responsible for operating most public transportation services in Greater Boston, Massachusetts. The MBTA transit network includes the MBTA subway with three metro lines (the Blue, Orange, and Red lines), two light rail lines (the Green and Ashmont–Mattapan lines), and a five-line bus rapid transit system (the Silver Line); MBTA bus local and express service; the twelve-line MBTA Commuter Rail system, and several ferry routes.

In this report, I will do some exploratory data analysis to answer the question that how reliable is the MBTA service. In order to finish the analysis, I pick one week in July as my data.

```
##
##     'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

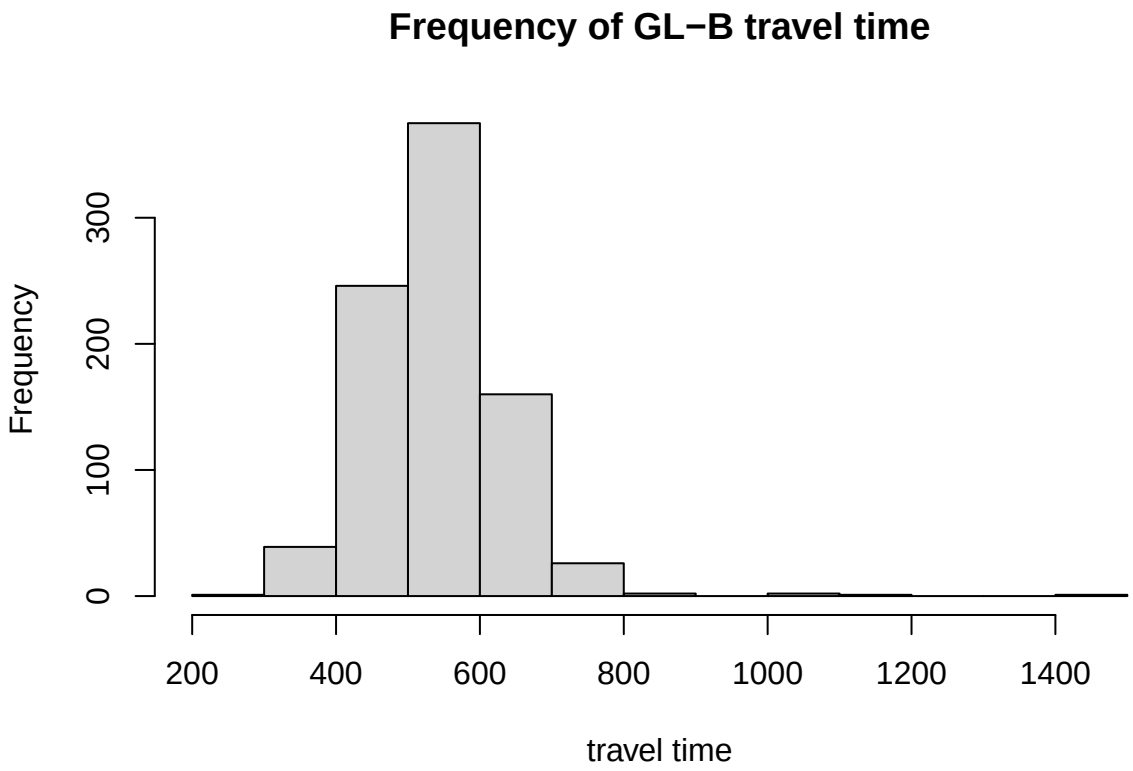
```
##
##     'qqplotr'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     stat_qq_line, StatQqLine
```

## Data cleaning

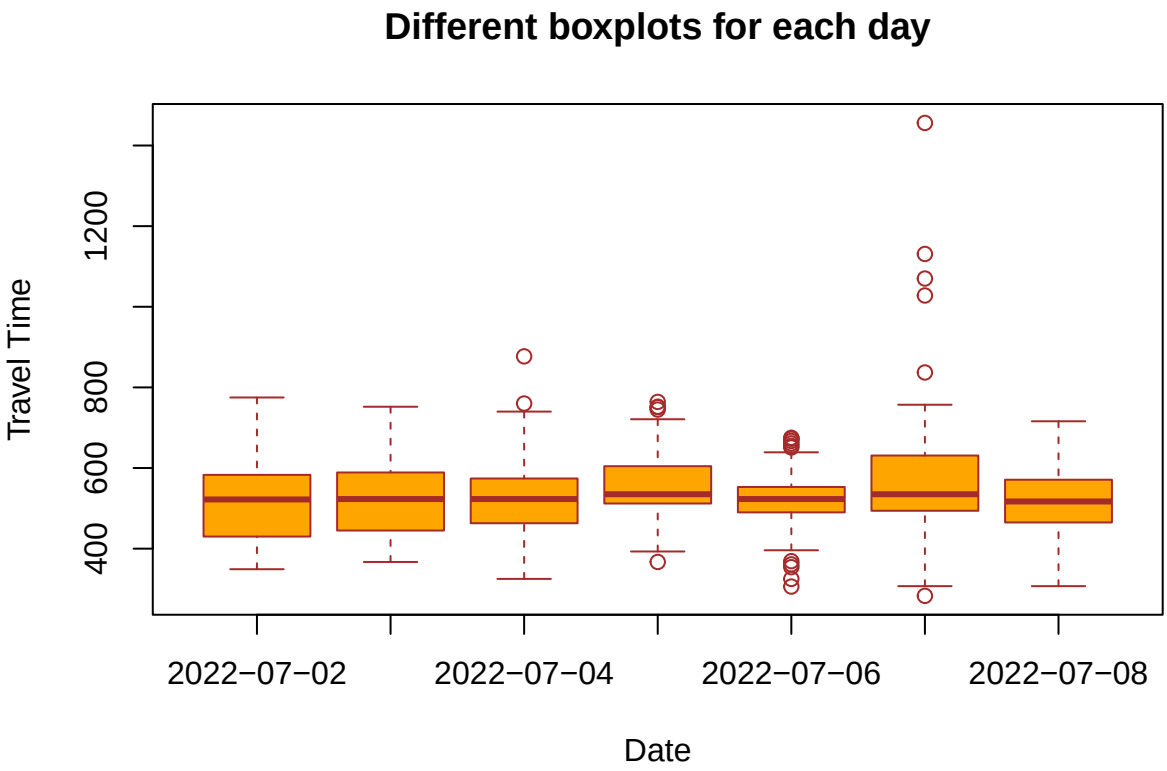
```
data <- read.csv("/Users/roselv/Downloads/TravelTimes_2022/2022-Q3_LRTravelTimes.csv")
df1 <- data %>%
  mutate(service_date = lubridate::ymd(service_date)) %>%
  filter(service_date >= as.Date('2022-07-02') &
           service_date <= as.Date('2022-07-08') &
           from_stop_id == 70130 &
           to_stop_id == 70144 &
           route_id == "Green-B")
drop <- c("direction_id","route_id")
df = df1[,!(names(df1) %in% drop)]
```

**Histogram**



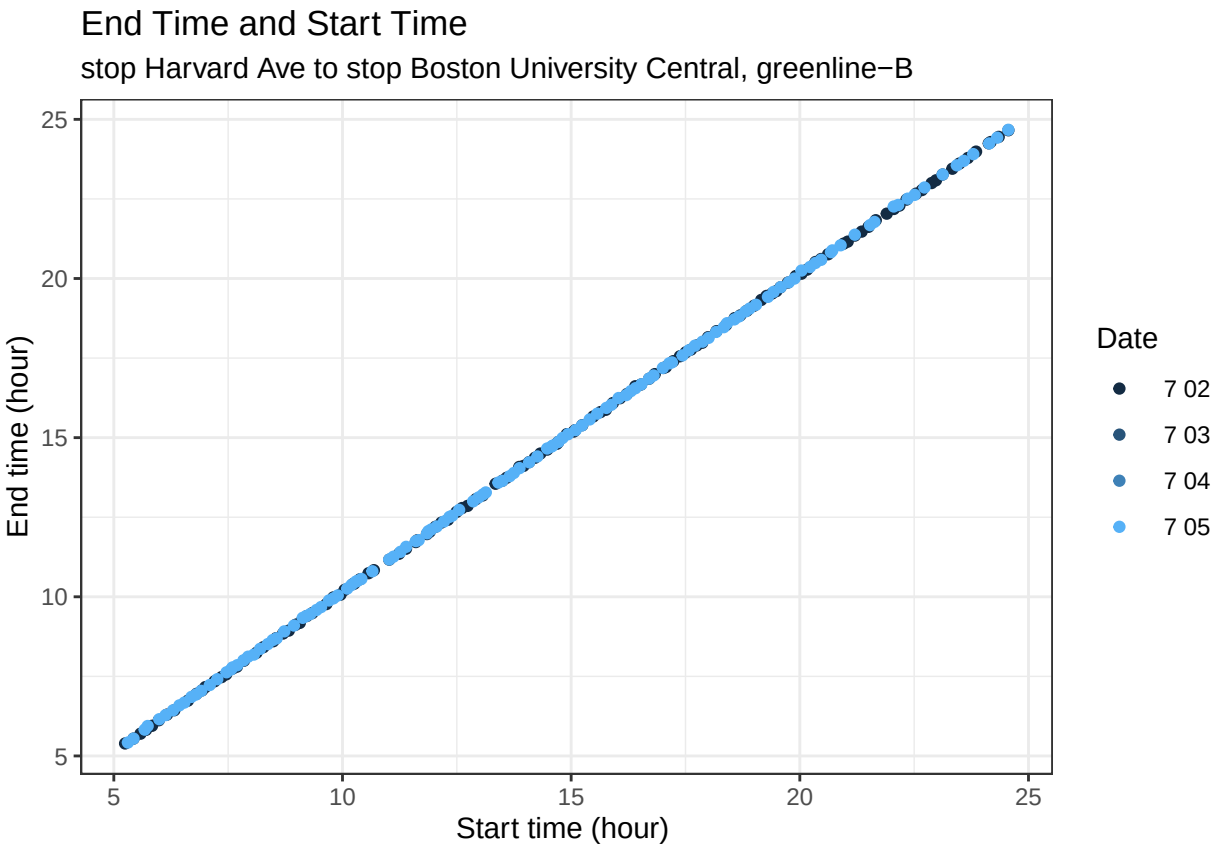

This figure is a histogram which is used to see the distribution of travel time. From this histogram, we can see that the distribution of this data is roughly Gaussian. Also, the travel time from the Harvard Ave to Boston University Central usually won't take too much time,most of them takes 6-12mins.

## Boxplot




Here is a box-plot to show the travel times on different dates. We can tell from the plot that the medians are roughly on the same horizontal level, which means the travel times are generally stable. In this case, greenline-B is reliable and people can trust the MBTA system.

### Scatter plot




In this plot, we can see that the scatters from different day of the week are almost on the same line. In this case, the start time and end time during the weekdays and weekends are stable, which means the MBTA system is kind of reliable for people to choose as an option of going out.

## Conclusion

I draw three plots in this report to see the reliability of the MBTA system. Through my exploratory data analysis, I found that the travel time between stop Harvard Ave and stop Boston University is relatively stable. Even though the traffic pattern is different between weekdays and weekends, there is barely a difference in the travel time of greenline-B between the weekdays and weekends. Therefore, people can trust the MBTA.