MBTA project EDA

runze

2022-12-18

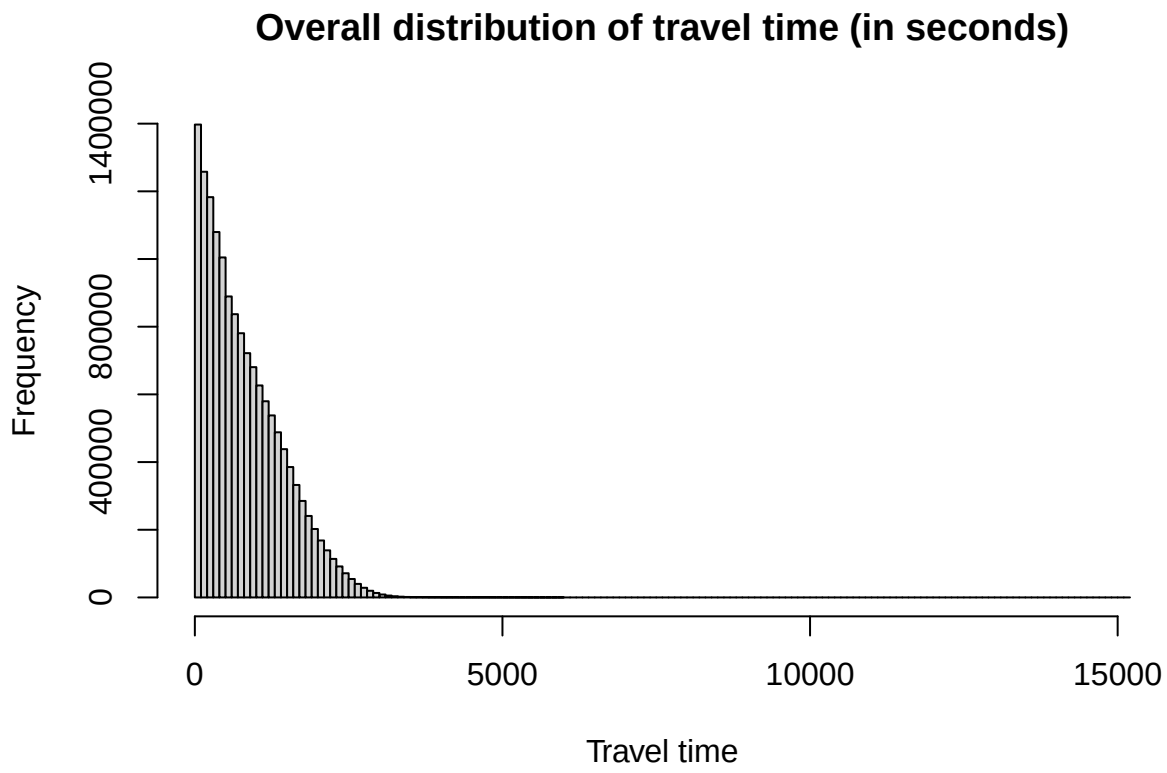
The light rail data EDA

Check the overall distribution of travel times

```
summary(smp_T$travel_time_sec)
```

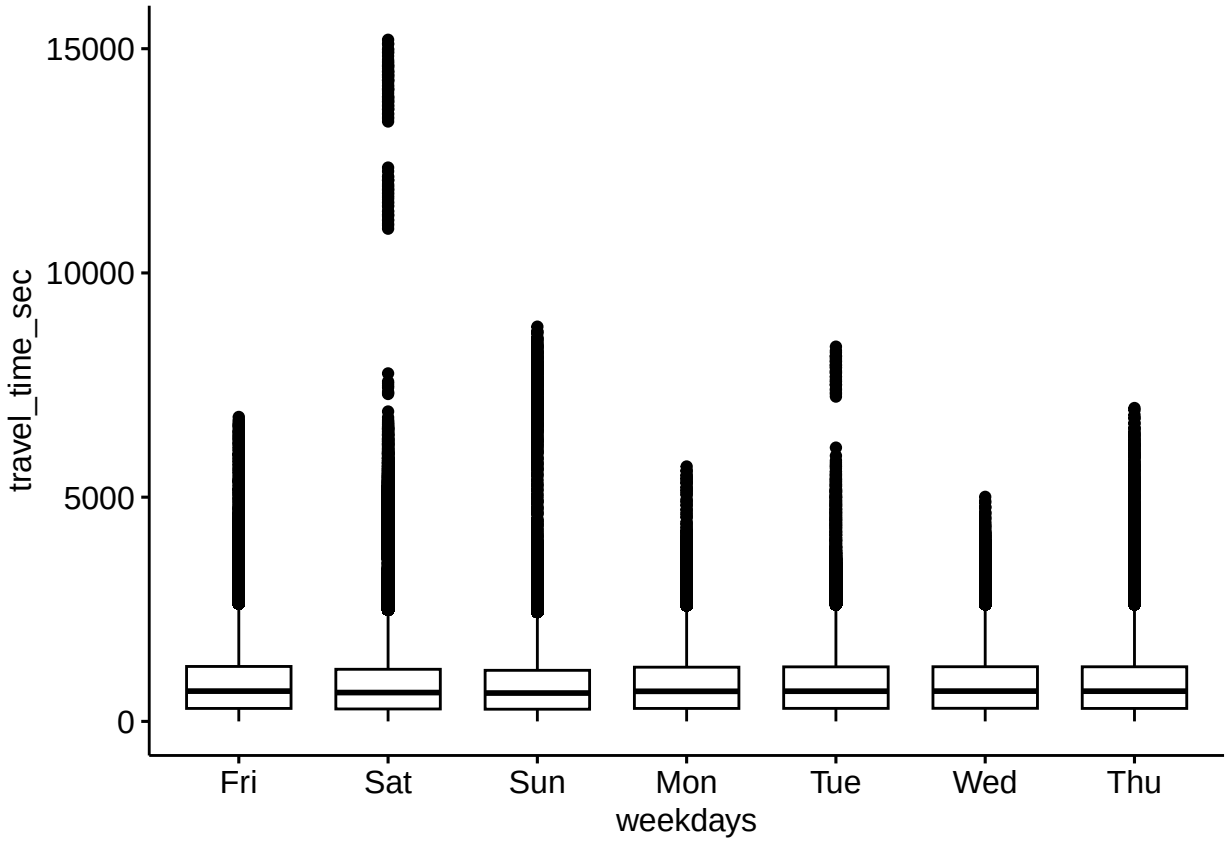
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   287.0   665.0   805.9  1201.0 15199.0
```

```
hist(smp_T$travel_time_sec,breaks = 120,main = "Overall distribution of travel time (in seconds)",xlab =
```



Check the distribution of travel times for each weekdays

```
ggboxplot(smp_T,x = "weekdays",y="travel_time_sec")
```



We can check many kinds of trips are there

```
trip<-levels(smp_T$trip)

length(trip)
```

```
## [1] 2023
```

There are total 2023 kinds of trips for light rails in boston

We can check which trip has the highest frequency

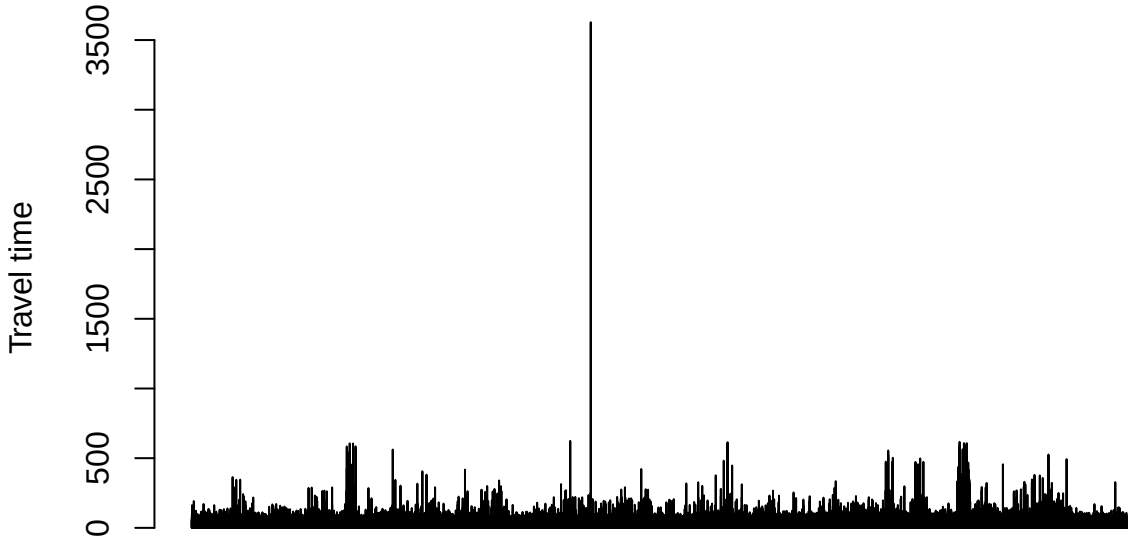
```
table(smp_T$trip)[table(smp_T$trip)==max(table(smp_T$trip))]
```

```
## From 70157 to 70155
##               39928
```

The most busy trip is from station 70157 to station 70155, which has a frequency of 39928

Then check how this trip's travel times are distributed

```
barplot(smp_T$travel_time_sec[smp_T$trip=="From 70157 to 70155"],ylab = "Travel time")
```



We can see that for most cases, this trip finishes within 500 seconds, but for some extreme situation, the travel time can up to 3500 seconds.