

mbta eda

xiaoyanbin Cai

2022-12-15

loading libraries

```
library(shiny)
library(googleway)
library(dplyr)

## 
## Attaching package: 'dplyr'

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

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

library(leaflet)
library(knitr)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --

## v ggplot2 3.4.0      v purrr    0.3.5
## v tibble   3.1.8      v stringr  1.4.1
## v tidyverse 1.2.1     vforcats  0.5.2
## v readr    2.1.3
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()

library(lubridate)

## Loading required package: timechange
## 
## Attaching package: 'lubridate'
## 
## The following objects are masked from 'package:base':
## 
##     date, intersect, setdiff, union
```

```
import data
```

```
traveltime_1 <- read.csv("~/Desktop/615-final-/2022-Q1_HRTravelTimes.csv")
head(traveltime_1)
```

```
##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-01-01      70004     70001 Orange          0        45822
## 2 2022-01-01      70004     70001 Orange          0        45355
## 3 2022-01-01      70004     70001 Orange          0        58824
## 4 2022-01-01      70004     70001 Orange          0        57792
## 5 2022-01-01      70004     70001 Orange          0        44557
## 6 2022-01-01      70004     70001 Orange          0        44187
##   end_time_sec travel_time_sec
## 1           46027            205
## 2           45556            201
## 3           59108            284
## 4           58021            229
## 5           44776            219
## 6           44396            209
```

```
traveltime_1.2 <- read.csv("~/Desktop/615-final-/2022-Q1_LRTravelTimes.csv")
head(traveltime_1.2)
```

```
##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-01-01      70134    170136 Green-B          1        75133
## 2 2022-01-01      70134    170136 Green-B          1        74501
## 3 2022-01-01      70134    170136 Green-B          1        69630
## 4 2022-01-01      70134    170136 Green-B          1        69173
## 5 2022-01-01      70134    170136 Green-B          1        68386
## 6 2022-01-01      70134    170136 Green-B          1        25360
##   end_time_sec travel_time_sec
## 1           75193            60
## 2           74591            90
## 3           69762           132
## 4           69257            84
## 5           68452            66
## 6           25432            72
```

Extract one week data

```
traveltime_1$service_date <- as.Date(traveltime_1$service_date)
data1 <- traveltime_1[traveltime_1$service_date >= '2022-01-01' & traveltime_1$service_date <= '2022-01-07'
head(data1)
```

```
##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-01-01      70004     70001 Orange          0        45822
## 2 2022-01-01      70004     70001 Orange          0        45355
## 3 2022-01-01      70004     70001 Orange          0        58824
## 4 2022-01-01      70004     70001 Orange          0        57792
```

```

## 5 2022-01-01 70004 70001 Orange 0 44557
## 6 2022-01-01 70004 70001 Orange 0 44187
##   end_time_sec travel_time_sec
## 1        46027      205
## 2        45556      201
## 3        59108      284
## 4        58021      229
## 5        44776      219
## 6        44396      209

```

Extract Jan data

```

traveltime_1.2$service_date <- as.Date(traveltime_1.2$service_date)
data1.2 <- traveltime_1.2[traveltime_1.2$service_date>='2022-01-01' & traveltime_1.2$service_date<='2022-01-31',]
head(data1.2)

```

```

##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-01-01       70134     170136 Green-B           1        75133
## 2 2022-01-01       70134     170136 Green-B           1        74501
## 3 2022-01-01       70134     170136 Green-B           1        69630
## 4 2022-01-01       70134     170136 Green-B           1        69173
## 5 2022-01-01       70134     170136 Green-B           1        68386
## 6 2022-01-01       70134     170136 Green-B           1        25360
##   end_time_sec travel_time_sec
## 1        75193      60
## 2        74591      90
## 3        69762     132
## 4        69257      84
## 5        68452      66
## 6        25432      72

```

Put two data set together

```

total_1 = dplyr::bind_rows(data1,data1.2)
head(total_1)

```

```

##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-01-01       70004     70001 Orange           0        45822
## 2 2022-01-01       70004     70001 Orange           0        45355
## 3 2022-01-01       70004     70001 Orange           0        58824
## 4 2022-01-01       70004     70001 Orange           0        57792
## 5 2022-01-01       70004     70001 Orange           0        44557
## 6 2022-01-01       70004     70001 Orange           0        44187
##   end_time_sec travel_time_sec
## 1        46027      205
## 2        45556      201
## 3        59108      284
## 4        58021      229
## 5        44776      219
## 6        44396      209

```

import data

```
traveltime_2 <- read.csv("~/Desktop/615-final-/2022-Q1_HRTravelTimes.csv")  
head(traveltime_2)
```

```

##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-01-01          70004     70001 Orange            0        45822
## 2 2022-01-01          70004     70001 Orange            0        45355
## 3 2022-01-01          70004     70001 Orange            0        58824
## 4 2022-01-01          70004     70001 Orange            0        57792
## 5 2022-01-01          70004     70001 Orange            0        44557
## 6 2022-01-01          70004     70001 Orange            0        44187
##   end_time_sec travel_time_sec
## 1          46027           205
## 2          45556           201
## 3          59108           284
## 4          58021           229
## 5          44776           219
## 6          44396           209

```

```
traveltime_3 <- read.csv("~/Desktop/615-final-/2022-Q1_LRTravelTimes.csv")  
tail(traveltime_3)
```

```

##          service_date from_stop_id to_stop_id route_id direction_id
## 15879906    2022-03-31        70196    71151 Green-B         0
## 15879907    2022-03-31        70196    71151 Green-B         0
## 15879908    2022-03-31        70196    71151 Green-B         0
## 15879909    2022-03-31        70196    71151 Green-B         0
## 15879910    2022-03-31        70196    71151 Green-B         0
## 15879911    2022-03-31        70196    71151 Green-B         0
##          start_time_sec end_time_sec travel_time_sec
## 15879906           81673       82299            626
## 15879907           25803       26544            741
## 15879908           25003       25533            530
## 15879909           32512       33123            611
## 15879910           32944       33517            573
## 15879911           33507       34125            618

```

extract Fab data

```
traveltimes_2$service_date <- as.Date(traveltimes_2$service_date)
data2 <- traveltimes_2[traveltimes_2$service_date >= '2022-02-01' & traveltimes_2$service_date <= '2022-02-07'
tail(data2)

##           service_date from_stop_id to_stop_id route_id direction_id
## 4312071    2022-02-07        70049     70838   Blue          0
## 4312072    2022-02-07        70049     70838   Blue          0
## 4312073    2022-02-07        70049     70838   Blue          0
## 4312074    2022-02-07        70045     70838   Blue          0
## 4312075    2022-02-07        70045     70838   Blue          0
## 4312076    2022-02-07        70045     70838   Blue          0
##           start_time_sec end_time_sec travel_time_sec
## 4312071            60112       60728             616
## 4312072            59548       60137             589
## 4312073            57167       57793             626
## 4312074            57419       57793             374
## 4312075            58757       59109             352
## 4312076            58237       58615             378
```

extarct March data

```
traveltimes_3$service_date <- as.Date(traveltimes_3$service_date)
data3 <- traveltimes_3[traveltimes_3$service_date >= '2022-03-01' & traveltimes_3$service_date <= '2022-03-07'
tail(data3)

##           service_date from_stop_id to_stop_id route_id direction_id
## 11571146    2022-03-07        70202     71151 Green-B          0
## 11571147    2022-03-07        70202     71151 Green-B          0
## 11571148    2022-03-07        70202     71151 Green-B          0
## 11571149    2022-03-07        70202     71151 Green-B          0
## 11571150    2022-03-07        70202     71151 Green-B          0
## 11571151    2022-03-07        70202     71151 Green-B          0
##           start_time_sec end_time_sec travel_time_sec
## 11571146            73673       74845             1172
## 11571147            74689       75421              732
## 11571148            74382       75151              769
## 11571149            78745       79464              719
## 11571150            79635       80316              681
## 11571151            80075       80818              743
```

extarct April data

```
traveltimes_4 <- read.csv("~/Desktop/615-final-/2022-Q2_HRTravelTimes.csv")
head(traveltimes_4)
```

```

##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-04-01          70008     70001 Orange        0      76930
## 2 2022-04-01          70008     70001 Orange        0      75987
## 3 2022-04-01          70008     70001 Orange        0      75442
## 4 2022-04-01          70008     70001 Orange        0      76495
## 5 2022-04-01          70008     70001 Orange        0      43463
## 6 2022-04-01          70008     70001 Orange        0      42923
##   end_time_sec travel_time_sec
## 1           77368        438
## 2           76442        455
## 3           75903        461
## 4           76906        411
## 5           43843        380
## 6           43354        431

```

```

traveltime_4.2 <- read.csv("~/Desktop/615-final-/2022-Q2_LRTravelTimes.csv")
head(traveltime_4.2)

```

```

##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-04-01          70112    170136 Green-B        1      52208
## 2 2022-04-01          70112    170136 Green-B        1      52759
## 3 2022-04-01          70112    170136 Green-B        1      51655
## 4 2022-04-01          70112    170136 Green-B        1      65670
## 5 2022-04-01          70112    170136 Green-B        1      66646
## 6 2022-04-01          70112    170136 Green-B        1      34409
##   end_time_sec travel_time_sec
## 1           53176        968
## 2           53716        957
## 3           52654        999
## 4           66693       1023
## 5           67580        934
## 6           35485       1076

```

put 2 data together

```

traveltime_4$service_date <- as.Date(traveltime_4$service_date)
data4 <- traveltime_4[traveltime_4$service_date>='2022-04-01' & traveltime_4$service_date<='2022-04-07'
tail(data4)

```

```

##   service_date from_stop_id to_stop_id route_id direction_id
## 808582 2022-04-07          70053    70838 Blue        0
## 808583 2022-04-07          70053    70838 Blue        0
## 808584 2022-04-07          70053    70838 Blue        0
## 808585 2022-04-07          70053    70838 Blue        0
## 808586 2022-04-07          70053    70838 Blue        0
## 808587 2022-04-07          70053    70838 Blue        0
##   start_time_sec end_time_sec travel_time_sec
## 808582          29351      30291        940
## 808583          28736      29608        872
## 808584          28130      28932        802
## 808585          29030      29878        848

```

```

## 808586      29618      30478      860
## 808587      28436      29341      905

traveltime_4.2$service_date <- as.Date(traveltime_4.2$service_date)
data4.2 <- traveltime_4.2[traveltime_4.2$service_date>='2022-04-01' & traveltime_4.2$service_date<='2022-04-30',]
tail(data4.2)

##   service_date from_stop_id to_stop_id route_id direction_id
## 1151656    2022-04-07      70153     71151 Green-B        0
## 1151657    2022-04-07      70153     71151 Green-B        0
## 1151658    2022-04-07      70153     71151 Green-B        0
## 1151659    2022-04-07      70153     71151 Green-B        0
## 1151660    2022-04-07      70153     71151 Green-B        0
## 1151661    2022-04-07      70153     71151 Green-B        0
##   start_time_sec end_time_sec travel_time_sec
## 1151656          86789       86880           91
## 1151657          87460       87546           86
## 1151658          31046       31155          109
## 1151659          31966       32080          114
## 1151660          31380       31477           97
## 1151661          30792       30895          103

total_4 = dplyr::bind_rows(data4,data4.2)
head(total_4)

##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-04-01      70008     70001 Orange        0       76930
## 2 2022-04-01      70008     70001 Orange        0       75987
## 3 2022-04-01      70008     70001 Orange        0       75442
## 4 2022-04-01      70008     70001 Orange        0       76495
## 5 2022-04-01      70008     70001 Orange        0       43463
## 6 2022-04-01      70008     70001 Orange        0       42923
##   end_time_sec travel_time_sec
## 1          77368        438
## 2          76442        455
## 3          75903        461
## 4          76906        411
## 5          43843        380
## 6          43354        431

```

extract May data

```

traveltime_4$service_date <- as.Date(traveltime_4$service_date)
data5 <- traveltime_4[traveltime_4$service_date>='2022-05-01' & traveltime_4$service_date<='2022-05-07',]
tail(data5)

##   service_date from_stop_id to_stop_id route_id direction_id
## 4298861    2022-05-07      70027     70279 Orange        1
## 4298862    2022-05-07      70027     70279 Orange        1
## 4298863    2022-05-07      70027     70279 Orange        1

```

```

## 4298864 2022-05-07 70027 70279 Orange 1
## 4298865 2022-05-07 70027 70279 Orange 1
## 4298866 2022-05-07 70027 70279 Orange 1
## start_time_sec end_time_sec travel_time_sec
## 4298861 88080 88457 377
## 4298862 89319 89701 382
## 4298863 46503 46888 385
## 4298864 47293 47661 368
## 4298865 47846 48217 371
## 4298866 48157 48537 380

traveltime_4.2$service_date <- as.Date(traveltime_4.2$service_date)
data5.2 <- traveltime_4.2[traveltime_4.2$service_date>='2022-05-01' & traveltime_4.2$service_date<='2022-05-07',]
tail(data5.2)

## service_date from_stop_id to_stop_id route_id direction_id
## 7082773 2022-05-07 70155 71151 Green-B 0
## 7082774 2022-05-07 70155 71151 Green-B 0
## 7082775 2022-05-07 70155 71151 Green-B 0
## 7082776 2022-05-07 70155 71151 Green-B 0
## 7082777 2022-05-07 70155 71151 Green-B 0
## 7082778 2022-05-07 70155 71151 Green-B 0
## start_time_sec end_time_sec travel_time_sec
## 7082773 42194 42436 242
## 7082774 50869 51138 269
## 7082775 50385 50587 202
## 7082776 51616 51873 257
## 7082777 74212 74458 246
## 7082778 73918 74155 237

total_5 = dplyr::bind_rows(data5,data5.2)
head(total_5)

## service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-05-01 70278 70001 Orange 0 51849
## 2 2022-05-01 70278 70001 Orange 0 50518
## 3 2022-05-01 70278 70001 Orange 0 51168
## 4 2022-05-01 70278 70001 Orange 0 37209
## 5 2022-05-01 70278 70001 Orange 0 36362
## 6 2022-05-01 70278 70001 Orange 0 63065
## end_time_sec travel_time_sec
## 1 53809 1960
## 2 52540 2022
## 3 53215 2047
## 4 39315 2106
## 5 38189 1827
## 6 65093 2028

```

extract June data

```

traveltime_4$service_date <- as.Date(traveltime_4$service_date)
data6 <- traveltime_4[traveltime_4$service_date>='2022-06-01' & traveltime_4$service_date<='2022-06-07']
tail(data6)

##           service_date from_stop_id to_stop_id route_id direction_id
## 7728232    2022-06-07        70045     70838   Blue          0
## 7728233    2022-06-07        70045     70838   Blue          0
## 7728234    2022-06-07        70045     70838   Blue          0
## 7728235    2022-06-07        70045     70838   Blue          0
## 7728236    2022-06-07        70045     70838   Blue          0
## 7728237    2022-06-07        70045     70838   Blue          0
##           start_time_sec end_time_sec travel_time_sec
## 7728232            63268       63611            343
## 7728233            62833       63226            393
## 7728234            64272       64671            399
## 7728235            65694       66047            353
## 7728236            66188       66572            384
## 7728237            65236       65613            377

traveltime_4.2$service_date <- as.Date(traveltime_4.2$service_date)
data6.2 <- traveltime_4.2[traveltime_4.2$service_date>='2022-06-01' & traveltime_4.2$service_date<='2022-06-07']
tail(data6.2)

##           service_date from_stop_id to_stop_id route_id direction_id
## 13196904   2022-06-07        70196     71151 Green-B          0
## 13196905   2022-06-07        70196     71151 Green-B          0
## 13196906   2022-06-07        70196     71151 Green-B          0
## 13196907   2022-06-07        70196     71151 Green-B          0
## 13196908   2022-06-07        70196     71151 Green-B          0
## 13196909   2022-06-07        70196     71151 Green-B          0
##           start_time_sec end_time_sec travel_time_sec
## 13196904            51454       52197            743
## 13196905            50803       51438            635
## 13196906            50098       50780            682
## 13196907            75140       75685            545
## 13196908            76172       76808            636
## 13196909            75817       76457            640

total_6 = dplyr::bind_rows(data6,data6.2)
head(total_6)

##           service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1   2022-06-01        70002     70001 Orange          0        21862
## 2   2022-06-01        70002     70001 Orange          0        21514
## 3   2022-06-01        70002     70001 Orange          0        23160
## 4   2022-06-01        70002     70001 Orange          0        22639
## 5   2022-06-01        70002     70001 Orange          0        90136
## 6   2022-06-01        70002     70001 Orange          0        57334
##           end_time_sec travel_time_sec
## 1            21977           115
## 2            21641           127

```

```

## 3      23286      126
## 4      22723       84
## 5      90256      120
## 6      57519      185

```

extract July data

```

traveltime_7 <- read.csv("~/Desktop/615-final-/2022-Q3_HRTravelTimes.csv")
head(traveltime_7)

```

```

##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-07-01        70010     70001 Orange          0      70841
## 2 2022-07-01        70010     70001 Orange          0      83557
## 3 2022-07-01        70010     70001 Orange          0      21951
## 4 2022-07-01        70010     70001 Orange          0      89478
## 5 2022-07-01        70010     70001 Orange          0      22979
## 6 2022-07-01        70010     70001 Orange          0      23842
##   end_time_sec travel_time_sec
## 1           71409         568
## 2            84167         610
## 3            22493         542
## 4            90026         548
## 5            23516         537
## 6            24337         495

```

```

traveltime_7.2 <- read.csv("~/Desktop/615-final-/2022-Q3_LRTravelTimes.csv")
head(traveltime_7.2)

```

```

##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-07-01        70228     70150 Green-C          1      19551
## 2 2022-07-01        70228     70150 Green-C          1      20187
## 3 2022-07-01        70228     70150 Green-C          1      24691
## 4 2022-07-01        70228     70150 Green-C          1      25679
## 5 2022-07-01        70228     70150 Green-C          1      26023
## 6 2022-07-01        70228     70150 Green-C          1      31223
##   end_time_sec travel_time_sec
## 1           20202         651
## 2            20926         739
## 3            25487         796
## 4            26355         676
## 5            26771         748
## 6            32191         968

```

```

traveltime_7$service_date <- as.Date(traveltime_7$service_date)
data7 <- traveltime_7[traveltime_7$service_date>='2022-07-01' & traveltime_7$service_date<='2022-07-07'
tail(data7)

```

```

##   service_date from_stop_id to_stop_id route_id direction_id
## 686708 2022-07-07        70057     70838 Blue          0
## 686709 2022-07-07        70057     70838 Blue          0

```

```

## 686710 2022-07-07 70057 70838 Blue 0
## 686711 2022-07-07 70057 70838 Blue 0
## 686712 2022-07-07 70057 70838 Blue 0
## 686713 2022-07-07 70057 70838 Blue 0
## start_time_sec end_time_sec travel_time_sec
## 686708 23998 25103 1105
## 686709 22657 23782 1125
## 686710 34488 35594 1106
## 686711 34017 35262 1245
## 686712 33163 34304 1141
## 686713 33583 34801 1218

traveltime_7.2$service_date <- as.Date(traveltime_7.2$service_date)
data7.2 <- traveltime_7.2[traveltime_7.2$service_date>='2022-07-01' & traveltime_7.2$service_date<='2022-07-07',]
tail(data7.2)

## service_date from_stop_id to_stop_id route_id direction_id
## 1290858 2022-07-07 70153 71151 Green-B 0
## 1290859 2022-07-07 70153 71151 Green-B 0
## 1290860 2022-07-07 70153 71151 Green-B 0
## 1290861 2022-07-07 70153 71151 Green-B 0
## 1290862 2022-07-07 70153 71151 Green-B 0
## 1290863 2022-07-07 70153 71151 Green-B 0
## start_time_sec end_time_sec travel_time_sec
## 1290858 35506 35712 206
## 1290859 35436 35568 132
## 1290860 34676 34814 138
## 1290861 34437 34559 122
## 1290862 43699 43860 161
## 1290863 44561 44748 187

```

put two data set together

```

total_7 = dplyr::bind_rows(data7,data7.2)
head(total_7)

```

```

## service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-07-01 70010 70001 Orange 0 70841
## 2 2022-07-01 70010 70001 Orange 0 83557
## 3 2022-07-01 70010 70001 Orange 0 21951
## 4 2022-07-01 70010 70001 Orange 0 89478
## 5 2022-07-01 70010 70001 Orange 0 22979
## 6 2022-07-01 70010 70001 Orange 0 23842
## end_time_sec travel_time_sec
## 1 71409 568
## 2 84167 610
## 3 22493 542
## 4 90026 548
## 5 23516 537
## 6 24337 495

```

Put 2022 data set all together

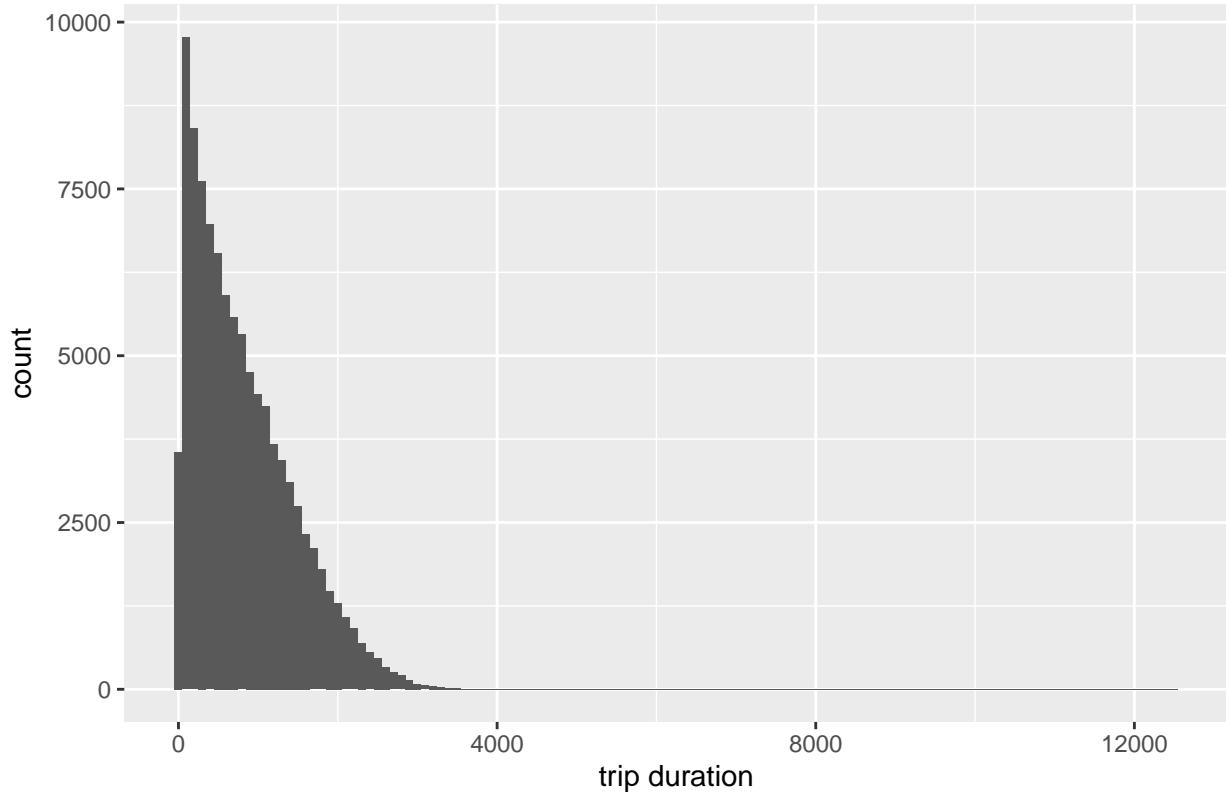
```
total_2022 = dplyr::bind_rows(total_1,data2,data3,total_4,total_5,total_6,total_7)
head(total_2022)

##   service_date from_stop_id to_stop_id route_id direction_id start_time_sec
## 1 2022-01-01        70004     70001 Orange            0        45822
## 2 2022-01-01        70004     70001 Orange            0        45355
## 3 2022-01-01        70004     70001 Orange            0        58824
## 4 2022-01-01        70004     70001 Orange            0        57792
## 5 2022-01-01        70004     70001 Orange            0        44557
## 6 2022-01-01        70004     70001 Orange            0        44187
##   end_time_sec travel_time_sec
## 1        46027          205
## 2        45556          201
## 3        59108          284
## 4        58021          229
## 5        44776          219
## 6        44396          209
```

Trip durations and their distributions

```
sample <- sample_n(total_2022,100000)#randomly choose 100000 data from total data set
plot1 <- ggplot(sample)+geom_histogram(binwidth =100, aes(travel_time_sec))+ ggttitle("Fig 1.trip duration")
plot1
```

Fig 1.trip durations and their distribution



Trip durations and their distributions transformation

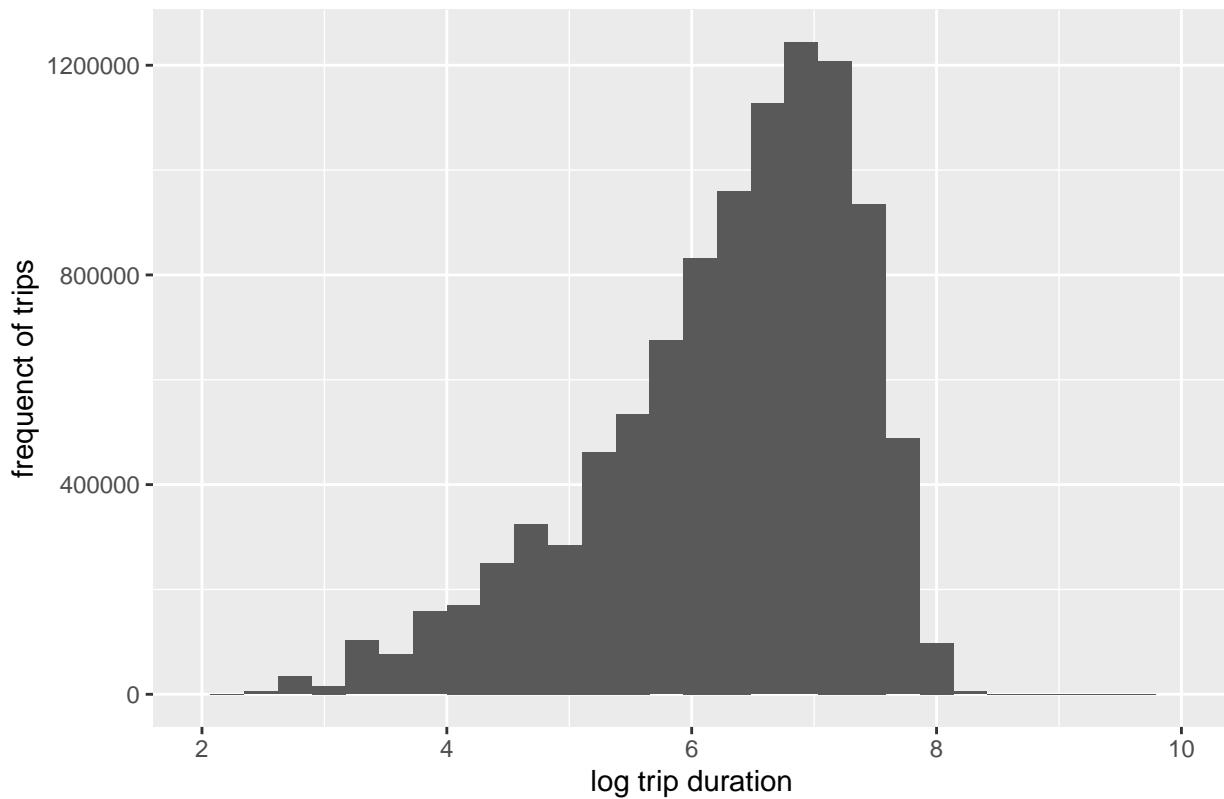
```
sample <- sample_n(total_2022, 10000000) #randomly choose 100000 data from total data set
plot2 <- ggplot(sample)+geom_histogram(aes(log(travel_time_sec)))+ ggtitle("Fig 2.trip durations and the")
plot2

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 7274 rows containing non-finite values (`stat_bin()`).

## Warning: Removed 2 rows containing missing values (`geom_bar()`).
```

Fig 2.trip durations and their distribution after transformation



To evaluate how travel time vary over months

```
total_2022 %>%
  group_by(service_date) %>%
  summarise(mean_travel = mean(travel_time_sec)) %>%
  arrange(desc(mean_travel))
```

```
## # A tibble: 49 x 2
##   service_date    mean_travel
##   <date>            <dbl>
## 1 2022-02-02      869.
## 2 2022-02-04      859.
## 3 2022-05-06      854.
## 4 2022-05-05      850.
## 5 2022-06-07      845.
## 6 2022-02-03      844.
## 7 2022-06-04      842.
## 8 2022-05-07      840.
## 9 2022-06-06      837.
## 10 2022-06-01     837.
## # ... with 39 more rows
```

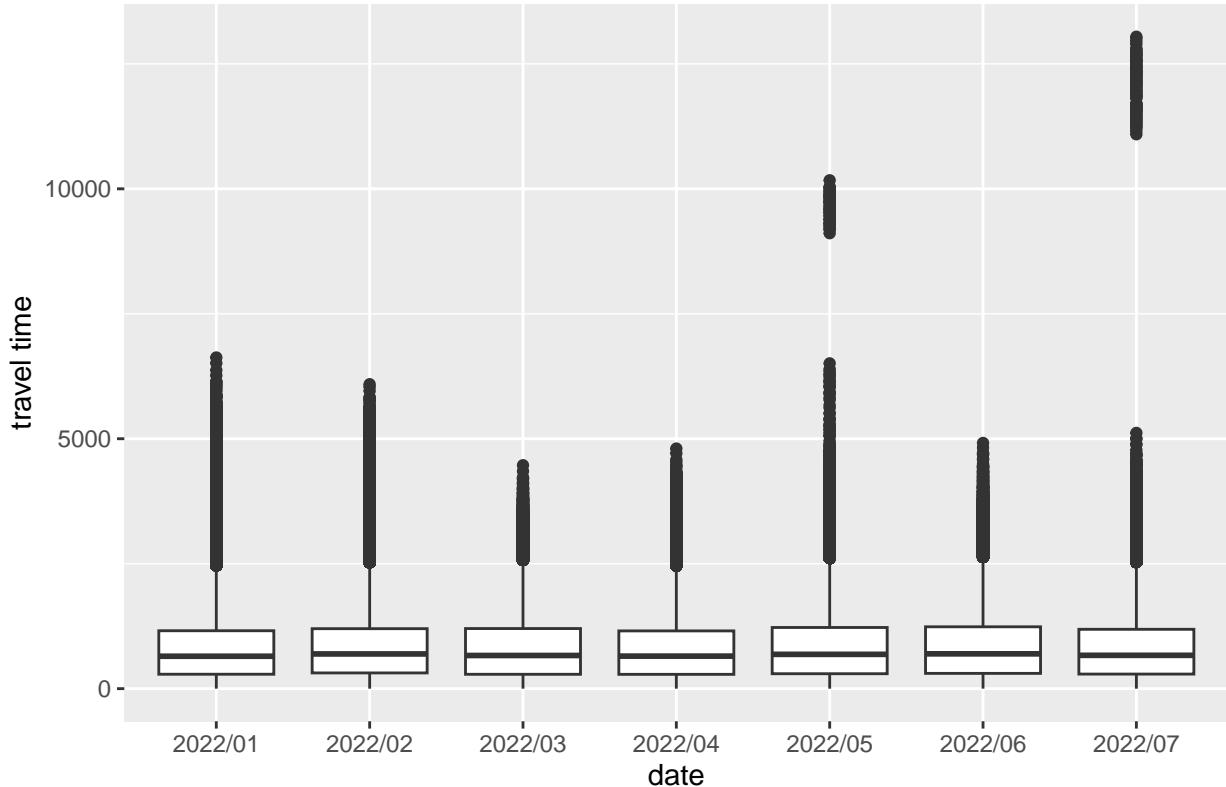
- As you can see from the result, on 2022-02-02 there's highest average travel time of subway from 2022 MBTA data frame.

```
service_date_1 <- format(as.Date(total_2022$service_date), "%Y/%m")
head(service_date_1)
```

```
## [1] "2022/01" "2022/01" "2022/01" "2022/01" "2022/01" "2022/01"
```

```
ggplot(total_2022,aes(x=factor(service_date_1),y=travel_time_sec)) + geom_boxplot() + labs(x="date",y="travel time")
```

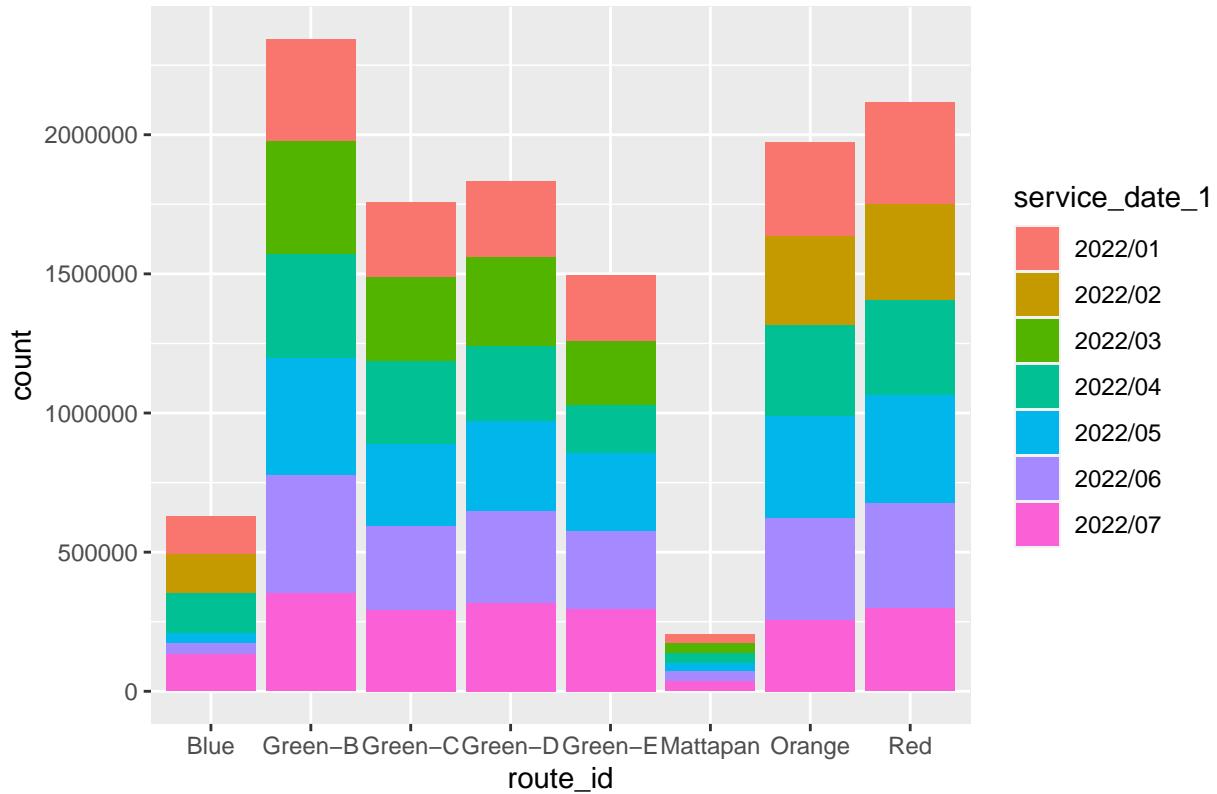
Fig 3.boxplot of service date of different train lines



* This boxplot has given us a clear vision about the most reliable measure and mean.

```
ggplot(data=total_2022,aes(x=route_id,fill = service_date_1)) + geom_bar() + ggtitle("Fig 4.histogram of route id")
```

Fig 4.histogram of number of different lines corporation in 2022



- This chart shows the operation of different metro lines in different months of the 22 years *