615 Final EDA

Ruiyi Feng

2022-12-16

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readr)
library(ggplot2)
library(dplyr)
library(tidyr)
```

EDA

Data Processing

```
setwd("~/Downloads")
dataq4 <- read.csv("HRTravelTimesQ4_21.csv")
# Data for orange line in the fourth quarter of 2021
dataq4_orange <- filter(dataq4,route_id == "Orange")
```

```
setwd("~/Downloads/TravelTimes_2022")
dataq1 <- read.csv("2022-Q1_HRTravelTimes.csv")
# Data for orange line in the first quarter of 2022
dataq1_orange <- filter(dataq1,route_id == "Orange")

dataq2 <- read.csv("2022-Q2_HRTravelTimes.csv")
# Data for orange line in the second quarter of 2022
dataq2_orange <- filter(dataq2,route_id == "Orange")

dataq3 <- read.csv("2022-Q3_HRTravelTimes.csv")
# Data for orange line in the third quarter of 2022
dataq3_orange <- filter(dataq3,route_id == "Orange")
```

In order to make sense for the analysis, I chose direction_id to be 0 and to_stop_id to be 70001. Then I selected the lines which had the daily latest arrival time. Dates are randomly selected. For the three months in 2021, I used the first seven days in each month. I filtered the data using same criteria for the data in 2022. I tried selecting the first seven days in each month but the first seven days does not work for data in September. Then I looked through the data for September and found seven days which could be used in the analysis.

```
# Data for October 2021
data1001 <- filter(dataq4_orange, service_date == "2021-10-01")
data1001_1 <- filter(data1001,direction_id == 0 & to_stop_id == 70001)
data1001_2 <- filter(data1001_1,end_time_sec == max(end_time_sec))

data1002 <- filter(dataq4_orange, service_date == "2021-10-02")
data1002_1 <- filter(data1002,direction_id == 0 & to_stop_id == 70001)
data1002_2 <- filter(data1002_1,end_time_sec == max(end_time_sec))

data1003 <- filter(dataq4_orange, service_date == "2021-10-03")
data1003_1 <- filter(data1003,direction_id == 0 & to_stop_id == 70001)
data1003_2 <- filter(data1003_1,end_time_sec == max(end_time_sec))

data1004 <- filter(dataq4_orange, service_date == "2021-10-04")
data1004_1 <- filter(data1004,direction_id == 0 & to_stop_id == 70001)
data1004_2 <- filter(data1004_1,end_time_sec == max(end_time_sec))

data1005 <- filter(dataq4_orange, service_date == "2021-10-05")
data1005_1 <- filter(data1005,direction_id == 0 & to_stop_id == 70001)
data1005_2 <- filter(data1005_1,end_time_sec == max(end_time_sec))

data1006 <- filter(dataq4_orange, service_date == "2021-10-06")
data1006_1 <- filter(data1006,direction_id == 0 & to_stop_id == 70001)
data1006_2 <- filter(data1006_1,end_time_sec == max(end_time_sec))
```

```
## Warning in max(end_time_sec): no non-missing arguments to max; returning -Inf
```

```
data1007 <- filter(dataq4_orange, service_date == "2021-10-07")
data1007_1 <- filter(data1007,direction_id == 0 & to_stop_id == 70001)
data1007_2 <- filter(data1007_1,end_time_sec == max(end_time_sec))

dataoct <- rbind(data1001_2,data1002_2,data1003_2,data1004_2,data1005_2,data1006_2,data1007_2)
```

```
# Data for November 2021
data1101 <- filter(dataq4_orange, service_date == "2021-11-01")
data1101_1 <- filter(data1101,direction_id == 0 & to_stop_id == 70001)
data1101_2 <- filter(data1101_1,end_time_sec == max(end_time_sec))

data1102 <- filter(dataq4_orange, service_date == "2021-11-02")
data1102_1 <- filter(data1102,direction_id == 0 & to_stop_id == 70001)
data1102_2 <- filter(data1102_1,end_time_sec == max(end_time_sec))

data1103 <- filter(dataq4_orange, service_date == "2021-11-03")
data1103_1 <- filter(data1103,direction_id == 0 & to_stop_id == 70001)
data1103_2 <- filter(data1103_1,end_time_sec == max(end_time_sec))
```

```
data1104 <- filter(dataq4_orange, service_date == "2021-11-04")
data1104_1 <- filter(data1104,direction_id == 0 & to_stop_id == 70001)
data1104_2 <- filter(data1104_1,end_time_sec == max(end_time_sec))

data1105 <- filter(dataq4_orange, service_date == "2021-11-05")
data1105_1 <- filter(data1105,direction_id == 0 & to_stop_id == 70001)
data1105_2 <- filter(data1105_1,end_time_sec == max(end_time_sec))

data1106 <- filter(dataq4_orange, service_date == "2021-11-06")
data1106_1 <- filter(data1106,direction_id == 0 & to_stop_id == 70001)
data1106_2 <- filter(data1106_1,end_time_sec == max(end_time_sec))

data1107 <- filter(dataq4_orange, service_date == "2021-11-07")
data1107_1 <- filter(data1107,direction_id == 0 & to_stop_id == 70001)
data1107_2 <- filter(data1107_1,end_time_sec == max(end_time_sec))

datanov <- rbind(data1101_2,data1102_2,data1103_2,data1104_2,data1105_2,data1106_2,data1107_2)
```

```
# Data for December 2021
data1201 <- filter(dataq4_orange, service_date == "2021-12-01")
data1201_1 <- filter(data1201,direction_id == 0 & to_stop_id == 70001)
data1201_2 <- filter(data1201_1,end_time_sec == max(end_time_sec))

data1202 <- filter(dataq4_orange, service_date == "2021-12-02")
data1202_1 <- filter(data1202,direction_id == 0 & to_stop_id == 70001)
data1202_2 <- filter(data1202_1,end_time_sec == max(end_time_sec))

data1203 <- filter(dataq4_orange, service_date == "2021-12-03")
data1203_1 <- filter(data1203,direction_id == 0 & to_stop_id == 70001)
data1203_2 <- filter(data1203_1,end_time_sec == max(end_time_sec))

data1204 <- filter(dataq4_orange, service_date == "2021-12-04")
data1204_1 <- filter(data1204,direction_id == 0 & to_stop_id == 70001)
data1204_2 <- filter(data1204_1,end_time_sec == max(end_time_sec))

data1205 <- filter(dataq4_orange, service_date == "2021-12-05")
data1205_1 <- filter(data1205,direction_id == 0 & to_stop_id == 70001)
data1205_2 <- filter(data1205_1,end_time_sec == max(end_time_sec))

data1206 <- filter(dataq4_orange, service_date == "2021-12-06")
data1206_1 <- filter(data1206,direction_id == 0 & to_stop_id == 70001)
data1206_2 <- filter(data1206_1,end_time_sec == max(end_time_sec))

data1207 <- filter(dataq4_orange, service_date == "2021-12-07")
data1207_1 <- filter(data1207,direction_id == 0 & to_stop_id == 70001)
data1207_2 <- filter(data1207_1,end_time_sec == max(end_time_sec))

datadec <- rbind(data1201_2,data1202_2,data1203_2,data1204_2,data1205_2,data1206_2,data1207_2)
```

```
# Data for January 2022
data0101 <- filter(dataq1_orange, service_date == "2022-01-01")
data0101_1 <- filter(data0101,direction_id == 0 & to_stop_id == 70001)
data0101_2 <- filter(data0101_1,end_time_sec == max(end_time_sec))
```

```r
data0102 <- filter(dataq1_orange, service_date == "2022-01-02")
data0102_1 <- filter(data0102,direction_id == 0 & to_stop_id == 70001)
data0102_2 <- filter(data0102_1,end_time_sec == max(end_time_sec))

data0103 <- filter(dataq1_orange, service_date == "2022-01-03")
data0103_1 <- filter(data0103,direction_id == 0 & to_stop_id == 70001)
data0103_2 <- filter(data0103_1,end_time_sec == max(end_time_sec))

data0104 <- filter(dataq1_orange, service_date == "2022-01-04")
data0104_1 <- filter(data0104,direction_id == 0 & to_stop_id == 70001)
data0104_2 <- filter(data0104_1,end_time_sec == max(end_time_sec))

data0105 <- filter(dataq1_orange, service_date == "2022-01-05")
data0105_1 <- filter(data0105,direction_id == 0 & to_stop_id == 70001)
data0105_2 <- filter(data0105_1,end_time_sec == max(end_time_sec))

data0106 <- filter(dataq1_orange, service_date == "2022-01-06")
data0106_1 <- filter(data0106,direction_id == 0 & to_stop_id == 70001)
data0106_2 <- filter(data0106_1,end_time_sec == max(end_time_sec))

data0107 <- filter(dataq1_orange, service_date == "2022-01-07")
data0107_1 <- filter(data0107,direction_id == 0 & to_stop_id == 70001)
data0107_2 <- filter(data0107_1,end_time_sec == max(end_time_sec))

datajan <- rbind(data0101_2,data0102_2,data0103_2,data0104_2,data0105_2,data0106_2,data0107_2)
```

```r
# Data for February 2022
data0201 <- filter(dataq1_orange, service_date == "2022-02-01")
data0201_1 <- filter(data0201,direction_id == 0 & to_stop_id == 70001)
data0201_2 <- filter(data0201_1,end_time_sec == max(end_time_sec))

data0202 <- filter(dataq1_orange, service_date == "2022-02-02")
data0202_1 <- filter(data0202,direction_id == 0 & to_stop_id == 70001)
data0202_2 <- filter(data0202_1,end_time_sec == max(end_time_sec))

data0203 <- filter(dataq1_orange, service_date == "2022-02-03")
data0203_1 <- filter(data0203,direction_id == 0 & to_stop_id == 70001)
data0203_2 <- filter(data0203_1,end_time_sec == max(end_time_sec))

data0204 <- filter(dataq1_orange, service_date == "2022-02-04")
data0204_1 <- filter(data0204,direction_id == 0 & to_stop_id == 70001)
data0204_2 <- filter(data0204_1,end_time_sec == max(end_time_sec))

data0205 <- filter(dataq1_orange, service_date == "2022-02-05")
data0205_1 <- filter(data0205,direction_id == 0 & to_stop_id == 70001)
data0205_2 <- filter(data0205_1,end_time_sec == max(end_time_sec))

data0206 <- filter(dataq1_orange, service_date == "2022-02-06")
data0206_1 <- filter(data0206,direction_id == 0 & to_stop_id == 70001)
data0206_2 <- filter(data0206_1,end_time_sec == max(end_time_sec))

data0207 <- filter(dataq1_orange, service_date == "2022-02-07")
data0207_1 <- filter(data0207,direction_id == 0 & to_stop_id == 70001)
```

```
data0207_2 <- filter(data0207_1,end_time_sec == max(end_time_sec))

datafeb <- rbind(data0201_2,data0202_2,data0203_2,data0204_2,data0205_2,data0206_2,data0207_2)
```

```
# Data for March 2022
data0301 <- filter(dataq1_orange, service_date == "2022-03-01")
data0301_1 <- filter(data0301,direction_id == 0 & to_stop_id == 70001)
data0301_2 <- filter(data0301_1,end_time_sec == max(end_time_sec))

data0302 <- filter(dataq1_orange, service_date == "2022-03-02")
data0302_1 <- filter(data0302,direction_id == 0 & to_stop_id == 70001)
data0302_2 <- filter(data0302_1,end_time_sec == max(end_time_sec))

data0303 <- filter(dataq1_orange, service_date == "2022-03-03")
data0303_1 <- filter(data0303,direction_id == 0 & to_stop_id == 70001)
data0303_2 <- filter(data0303_1,end_time_sec == max(end_time_sec))

data0304 <- filter(dataq1_orange, service_date == "2022-03-04")
data0304_1 <- filter(data0304,direction_id == 0 & to_stop_id == 70001)
data0304_2 <- filter(data0304_1,end_time_sec == max(end_time_sec))

data0305 <- filter(dataq1_orange, service_date == "2022-03-05")
data0305_1 <- filter(data0305,direction_id == 0 & to_stop_id == 70001)
data0305_2 <- filter(data0305_1,end_time_sec == max(end_time_sec))

data0306 <- filter(dataq1_orange, service_date == "2022-03-06")
data0306_1 <- filter(data0306,direction_id == 0 & to_stop_id == 70001)
data0306_2 <- filter(data0306_1,end_time_sec == max(end_time_sec))

data0307 <- filter(dataq1_orange, service_date == "2022-03-07")
data0307_1 <- filter(data0307,direction_id == 0 & to_stop_id == 70001)
data0307_2 <- filter(data0307_1,end_time_sec == max(end_time_sec))

datamar <- rbind(data0301_2,data0302_2,data0303_2,data0304_2,data0305_2,data0306_2,data0307_2)
```

```
# Data for April 2022
data0401 <- filter(dataq2_orange, service_date == "2022-04-01")
data0401_1 <- filter(data0401,direction_id == 0 & to_stop_id == 70001)
data0401_2 <- filter(data0401_1,end_time_sec == max(end_time_sec))

data0402 <- filter(dataq2_orange, service_date == "2022-04-02")
data0402_1 <- filter(data0402,direction_id == 0 & to_stop_id == 70001)
data0402_2 <- filter(data0402_1,end_time_sec == max(end_time_sec))

data0403 <- filter(dataq2_orange, service_date == "2022-04-03")
data0403_1 <- filter(data0403,direction_id == 0 & to_stop_id == 70001)
data0403_2 <- filter(data0403_1,end_time_sec == max(end_time_sec))

data0404 <- filter(dataq2_orange, service_date == "2022-04-04")
data0404_1 <- filter(data0404,direction_id == 0 & to_stop_id == 70001)
data0404_2 <- filter(data0404_1,end_time_sec == max(end_time_sec))

data0405 <- filter(dataq2_orange, service_date == "2022-04-05")
```

```
data0405_1 <- filter(data0405,direction_id == 0 & to_stop_id == 70001)
data0405_2 <- filter(data0405_1,end_time_sec == max(end_time_sec))

data0406 <- filter(dataq2_orange, service_date == "2022-04-06")
data0406_1 <- filter(data0406,direction_id == 0 & to_stop_id == 70001)
data0406_2 <- filter(data0406_1,end_time_sec == max(end_time_sec))

data0407 <- filter(dataq2_orange, service_date == "2022-04-07")
data0407_1 <- filter(data0407,direction_id == 0 & to_stop_id == 70001)
data0407_2 <- filter(data0407_1,end_time_sec == max(end_time_sec))

dataapr <- rbind(data0401_2,data0402_2,data0403_2,data0404_2,data0405_2,data0406_2,data0407_2)
```

```
# Data for May 2022
data0501 <- filter(dataq2_orange, service_date == "2022-05-01")
data0501_1 <- filter(data0501,direction_id == 0 & to_stop_id == 70001)
data0501_2 <- filter(data0501_1,end_time_sec == max(end_time_sec))

data0502 <- filter(dataq2_orange, service_date == "2022-05-02")
data0502_1 <- filter(data0502,direction_id == 0 & to_stop_id == 70001)
data0502_2 <- filter(data0502_1,end_time_sec == max(end_time_sec))

data0503 <- filter(dataq2_orange, service_date == "2022-05-03")
data0503_1 <- filter(data0503,direction_id == 0 & to_stop_id == 70001)
data0503_2 <- filter(data0503_1,end_time_sec == max(end_time_sec))

data0504 <- filter(dataq2_orange, service_date == "2022-05-04")
data0504_1 <- filter(data0504,direction_id == 0 & to_stop_id == 70001)
data0504_2 <- filter(data0504_1,end_time_sec == max(end_time_sec))

data0505 <- filter(dataq2_orange, service_date == "2022-05-05")
data0505_1 <- filter(data0505,direction_id == 0 & to_stop_id == 70001)
data0505_2 <- filter(data0505_1,end_time_sec == max(end_time_sec))

data0506 <- filter(dataq2_orange, service_date == "2022-05-06")
data0506_1 <- filter(data0506,direction_id == 0 & to_stop_id == 70001)
data0506_2 <- filter(data0506_1,end_time_sec == max(end_time_sec))

data0507 <- filter(dataq2_orange, service_date == "2022-05-07")
data0507_1 <- filter(data0507,direction_id == 0 & to_stop_id == 70001)
data0507_2 <- filter(data0507_1,end_time_sec == max(end_time_sec))

datamay <- rbind(data0501_2,data0502_2,data0503_2,data0504_2,data0505_2,data0506_2,data0507_2)
```

```
# Data for June 2022
data0601 <- filter(dataq2_orange, service_date == "2022-06-01")
data0601_1 <- filter(data0601,direction_id == 0 & to_stop_id == 70001)
data0601_2 <- filter(data0601_1,end_time_sec == max(end_time_sec))

data0602 <- filter(dataq2_orange, service_date == "2022-06-02")
data0602_1 <- filter(data0602,direction_id == 0 & to_stop_id == 70001)
data0602_2 <- filter(data0602_1,end_time_sec == max(end_time_sec))
```

```
data0603 <- filter(dataq2_orange, service_date == "2022-06-03")
data0603_1 <- filter(data0603,direction_id == 0 & to_stop_id == 70001)
data0603_2 <- filter(data0603_1,end_time_sec == max(end_time_sec))

data0604 <- filter(dataq2_orange, service_date == "2022-06-04")
data0604_1 <- filter(data0604,direction_id == 0 & to_stop_id == 70001)
data0604_2 <- filter(data0604_1,end_time_sec == max(end_time_sec))

data0605 <- filter(dataq2_orange, service_date == "2022-06-05")
data0605_1 <- filter(data0605,direction_id == 0 & to_stop_id == 70001)
data0605_2 <- filter(data0605_1,end_time_sec == max(end_time_sec))

data0606 <- filter(dataq2_orange, service_date == "2022-06-06")
data0606_1 <- filter(data0606,direction_id == 0 & to_stop_id == 70001)
data0606_2 <- filter(data0606_1,end_time_sec == max(end_time_sec))

data0607 <- filter(dataq2_orange, service_date == "2022-06-07")
data0607_1 <- filter(data0607,direction_id == 0 & to_stop_id == 70001)
data0607_2 <- filter(data0607_1,end_time_sec == max(end_time_sec))

datajun <- rbind(data0601_2,data0602_2,data0603_2,data0604_2,data0605_2,data0606_2,data0607_2)
```

```
# Data for July 2022
data0701 <- filter(dataq3_orange, service_date == "2022-07-01")
data0701_1 <- filter(data0701,direction_id == 0 & to_stop_id == 70001)
data0701_2 <- filter(data0701_1,end_time_sec == max(end_time_sec))

data0702 <- filter(dataq3_orange, service_date == "2022-07-02")
data0702_1 <- filter(data0702,direction_id == 0 & to_stop_id == 70001)
data0702_2 <- filter(data0702_1,end_time_sec == max(end_time_sec))

data0703 <- filter(dataq3_orange, service_date == "2022-07-03")
data0703_1 <- filter(data0703,direction_id == 0 & to_stop_id == 70001)
data0703_2 <- filter(data0703_1,end_time_sec == max(end_time_sec))

data0704 <- filter(dataq3_orange, service_date == "2022-07-04")
data0704_1 <- filter(data0704,direction_id == 0 & to_stop_id == 70001)
data0704_2 <- filter(data0704_1,end_time_sec == max(end_time_sec))

data0705 <- filter(dataq3_orange, service_date == "2022-07-05")
data0705_1 <- filter(data0705,direction_id == 0 & to_stop_id == 70001)
data0705_2 <- filter(data0705_1,end_time_sec == max(end_time_sec))

data0706 <- filter(dataq3_orange, service_date == "2022-07-06")
data0706_1 <- filter(data0706,direction_id == 0 & to_stop_id == 70001)
data0706_2 <- filter(data0706_1,end_time_sec == max(end_time_sec))

data0707 <- filter(dataq3_orange, service_date == "2022-07-07")
data0707_1 <- filter(data0707,direction_id == 0 & to_stop_id == 70001)
data0707_2 <- filter(data0707_1,end_time_sec == max(end_time_sec))

datajul <- rbind(data0701_2,data0702_2,data0703_2,data0704_2,data0705_2,data0706_2,data0707_2)
```

```r
# Data for August 2022
data0801 <- filter(dataq3_orange, service_date == "2022-08-01")
data0801_1 <- filter(data0801,direction_id == 0 & to_stop_id == 70001)
data0801_2 <- filter(data0801_1,end_time_sec == max(end_time_sec))

data0802 <- filter(dataq3_orange, service_date == "2022-08-02")
data0802_1 <- filter(data0802,direction_id == 0 & to_stop_id == 70001)
data0802_2 <- filter(data0802_1,end_time_sec == max(end_time_sec))

data0803 <- filter(dataq3_orange, service_date == "2022-08-03")
data0803_1 <- filter(data0803,direction_id == 0 & to_stop_id == 70001)
data0803_2 <- filter(data0803_1,end_time_sec == max(end_time_sec))

data0804 <- filter(dataq3_orange, service_date == "2022-08-04")
data0804_1 <- filter(data0804,direction_id == 0 & to_stop_id == 70001)
data0804_2 <- filter(data0804_1,end_time_sec == max(end_time_sec))

data0805 <- filter(dataq3_orange, service_date == "2022-08-05")
data0805_1 <- filter(data0805,direction_id == 0 & to_stop_id == 70001)
data0805_2 <- filter(data0805_1,end_time_sec == max(end_time_sec))

data0806 <- filter(dataq3_orange, service_date == "2022-08-06")
data0806_1 <- filter(data0806,direction_id == 0 & to_stop_id == 70001)
data0806_2 <- filter(data0806_1,end_time_sec == max(end_time_sec))

data0807 <- filter(dataq3_orange, service_date == "2022-08-07")
data0807_1 <- filter(data0807,direction_id == 0 & to_stop_id == 70001)
data0807_2 <- filter(data0807_1,end_time_sec == max(end_time_sec))

dataaug <- rbind(data0801_2,data0802_2,data0803_2,data0804_2,data0805_2,data0806_2,data0807_2)
```

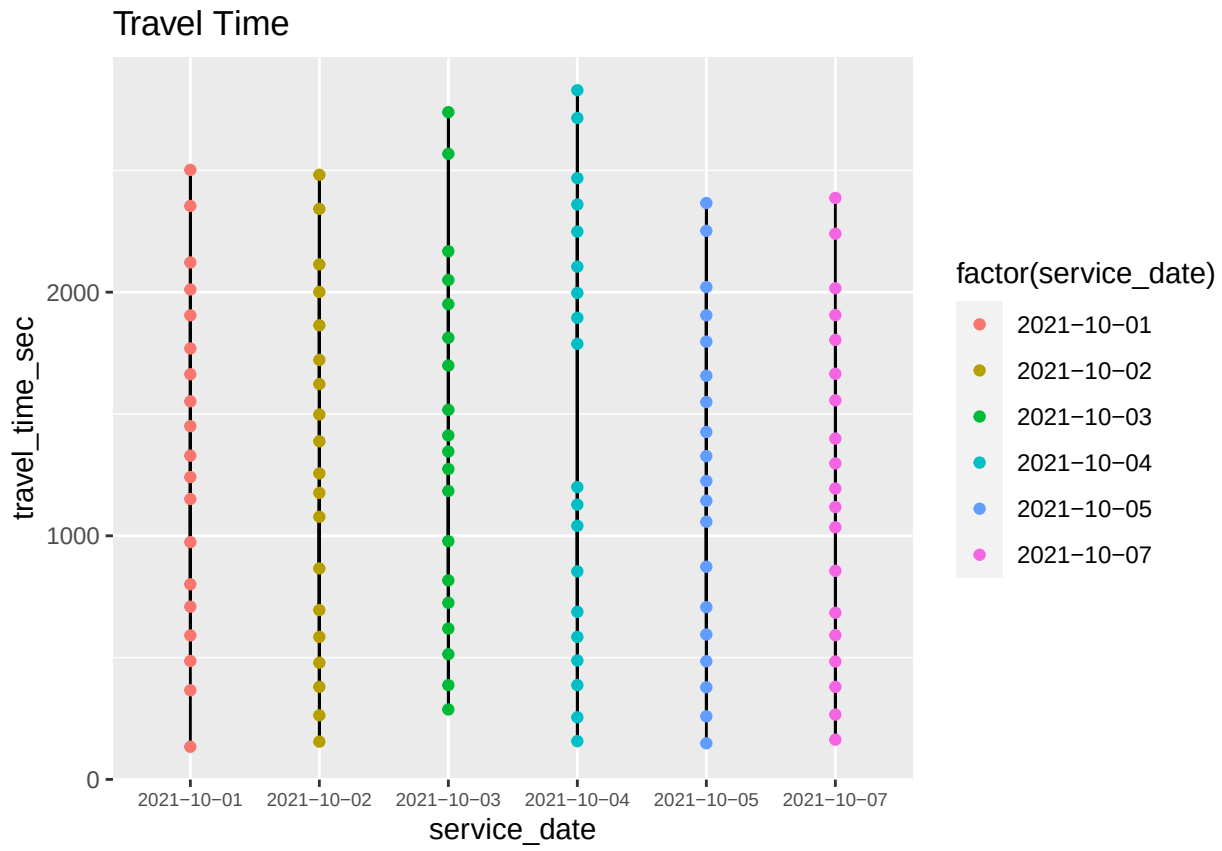
```r
# Data for September 2022
data0919 <- filter(dataq3_orange, service_date == "2022-09-19")
data0919_1 <- filter(data0919,direction_id == 0 & to_stop_id == 70001)
data0919_2 <- filter(data0919_1,end_time_sec == max(end_time_sec))

data0920 <- filter(dataq3_orange, service_date == "2022-09-20")
data0920_1 <- filter(data0920,direction_id == 0 & to_stop_id == 70001)
data0920_2 <- filter(data0920_1,end_time_sec == max(end_time_sec))

data0921 <- filter(dataq3_orange, service_date == "2022-09-21")
data0921_1 <- filter(data0921,direction_id == 0 & to_stop_id == 70001)
data0921_2 <- filter(data0921_1,end_time_sec == max(end_time_sec))

data0922 <- filter(dataq3_orange, service_date == "2022-09-22")
data0922_1 <- filter(data0922,direction_id == 0 & to_stop_id == 70001)
data0922_2 <- filter(data0922_1,end_time_sec == max(end_time_sec))

data0923 <- filter(dataq3_orange, service_date == "2022-09-23")
data0923_1 <- filter(data0923,direction_id == 0 & to_stop_id == 70001)
data0923_2 <- filter(data0923_1,end_time_sec == max(end_time_sec))

data0924 <- filter(dataq3_orange, service_date == "2022-09-24")
```

```
data0924_1 <- filter(data0924,direction_id == 0 & to_stop_id == 70001)
data0924_2 <- filter(data0924_1,end_time_sec == max(end_time_sec))

data0925 <- filter(dataq3_orange, service_date == "2022-09-25")
data0925_1 <- filter(data0925,direction_id == 0 & to_stop_id == 70001)
data0925_2 <- filter(data0925_1,end_time_sec == max(end_time_sec))

datasep <- rbind(data0919_2,data0920_2,data0921_2,data0922_2,data0923_2,data0924_2,data0925_2)
```

Plotting and Brief Explanations

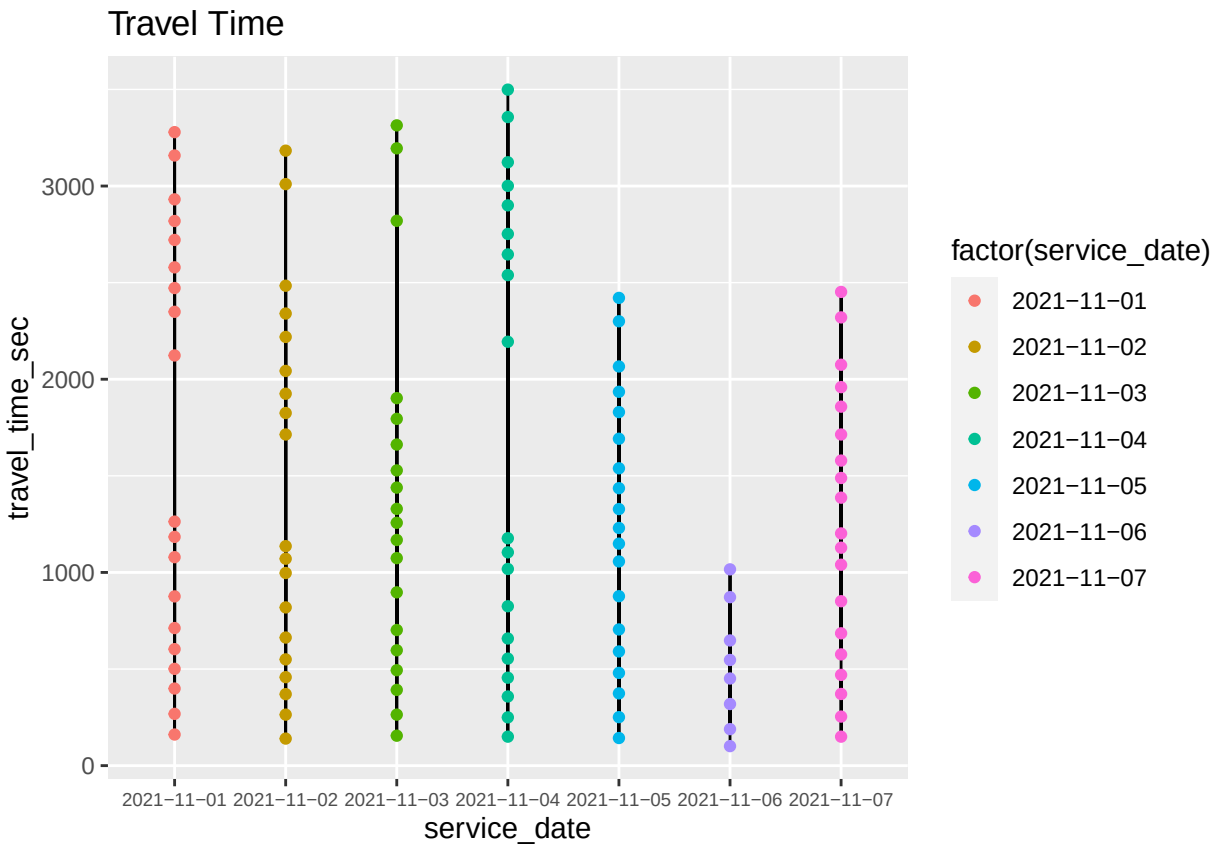
Here are the plots showing the travel time for selected dates in each month.

```
# Plot for October 2021
ggplot(data = dataoct) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```

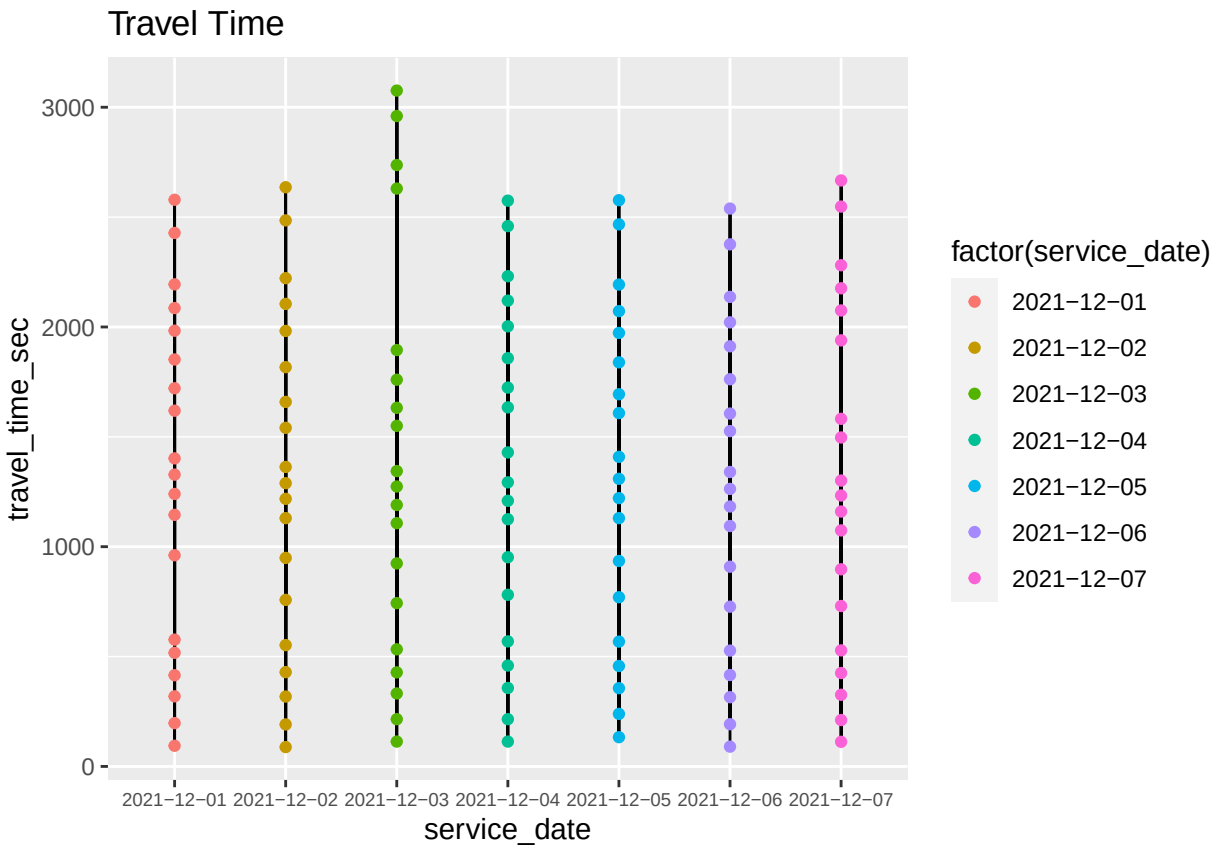


There are only six dates contained in this plot. The reason is that there were no lines which satisfied my selection for the missing day.

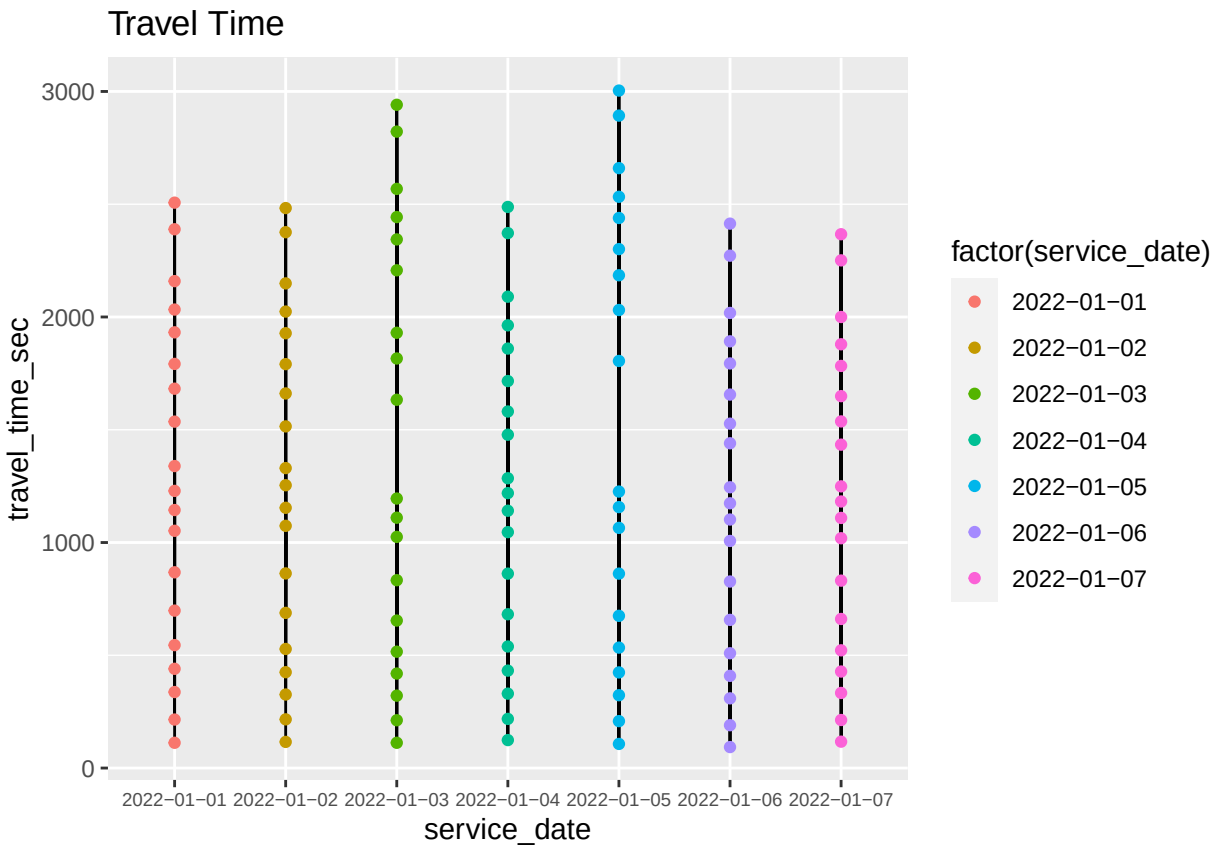
```
#Plot for November 2021
ggplot(data = datanov) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```



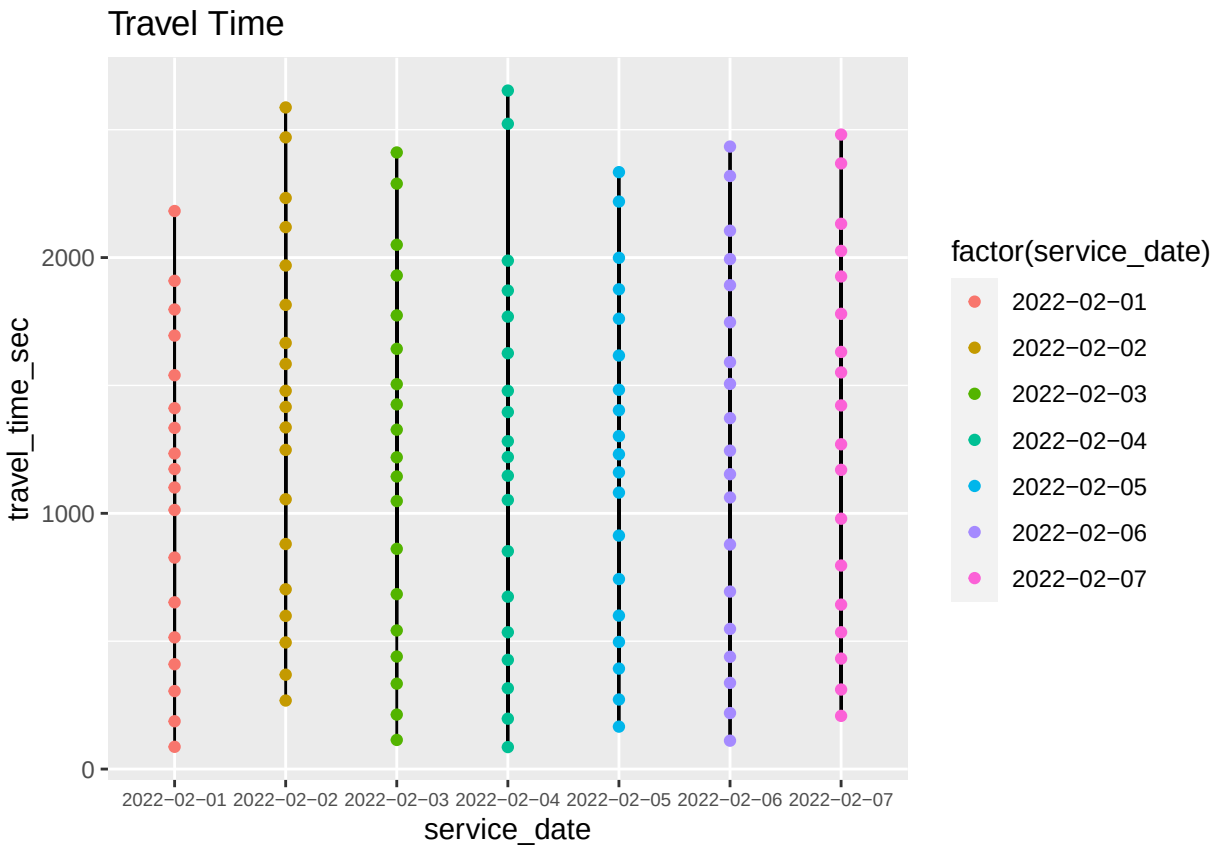

```
# Plot for December 2021
ggplot(data = datadec) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```



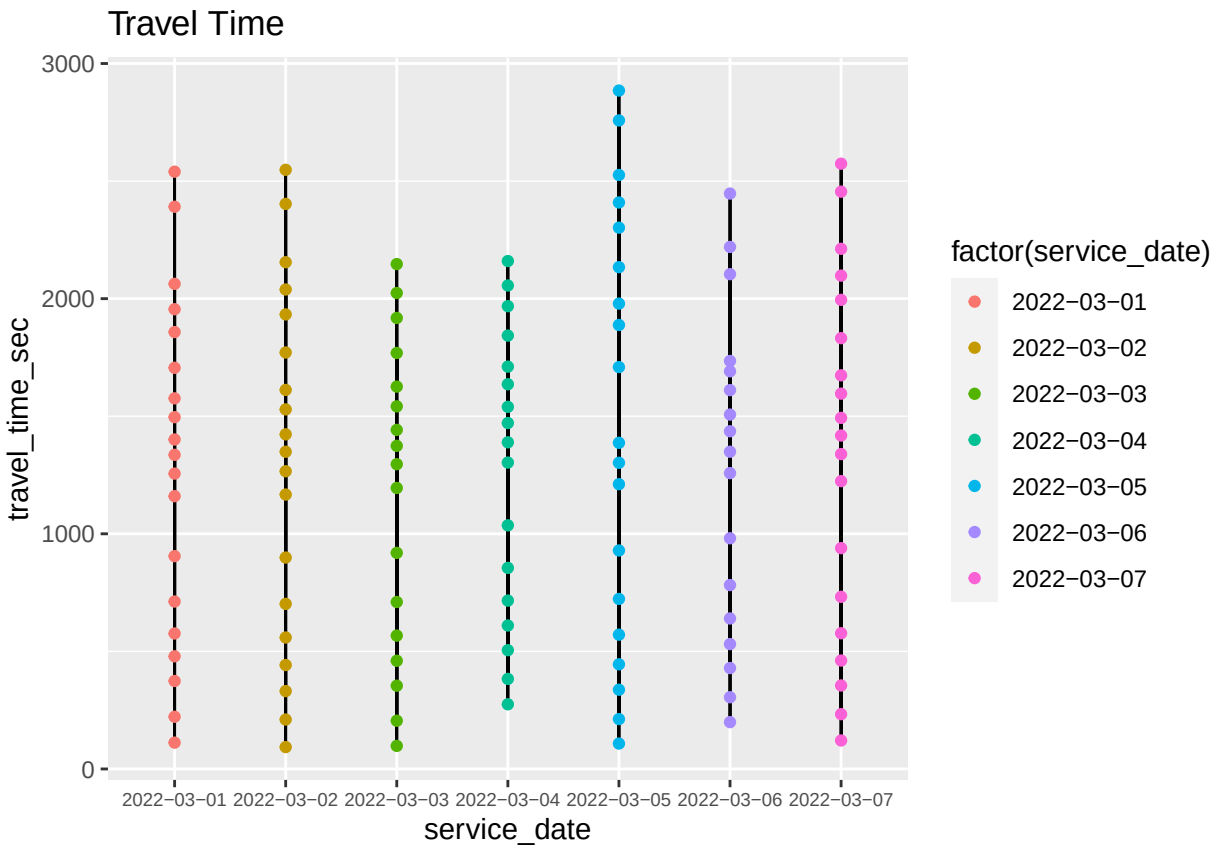

```
# Plot for January 2022
ggplot(data = datajan) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```



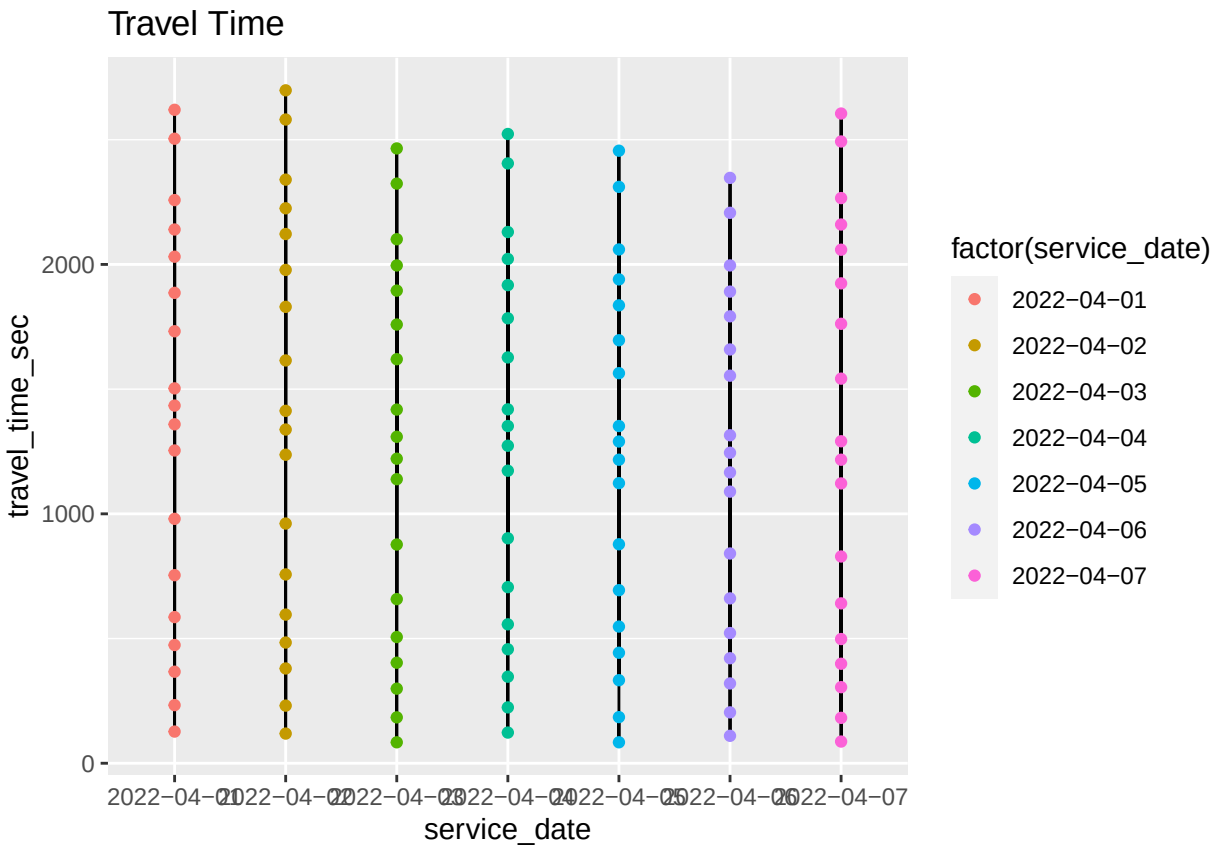

```
# Plot for Febuary 2022
ggplot(data = datafeb) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```



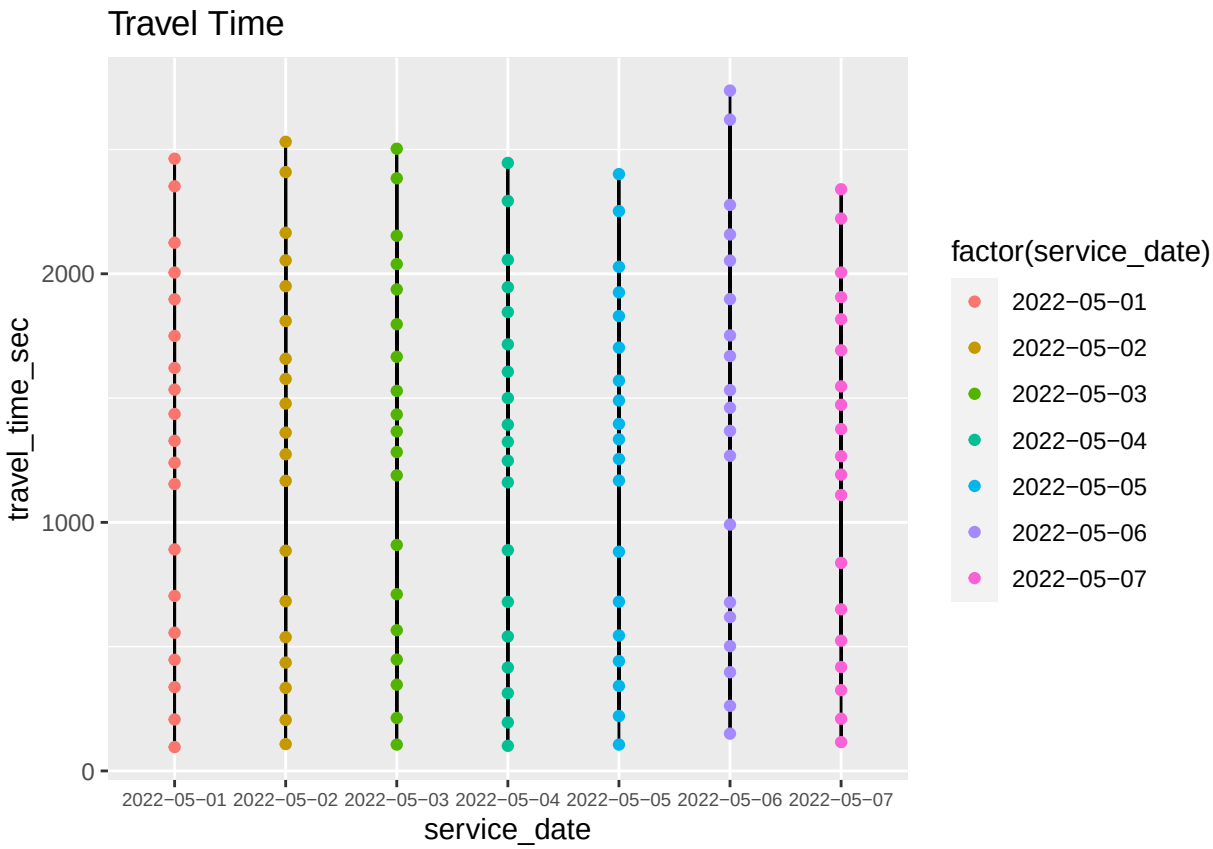

```
# Plot for March 2022
ggplot(data = datamar) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```



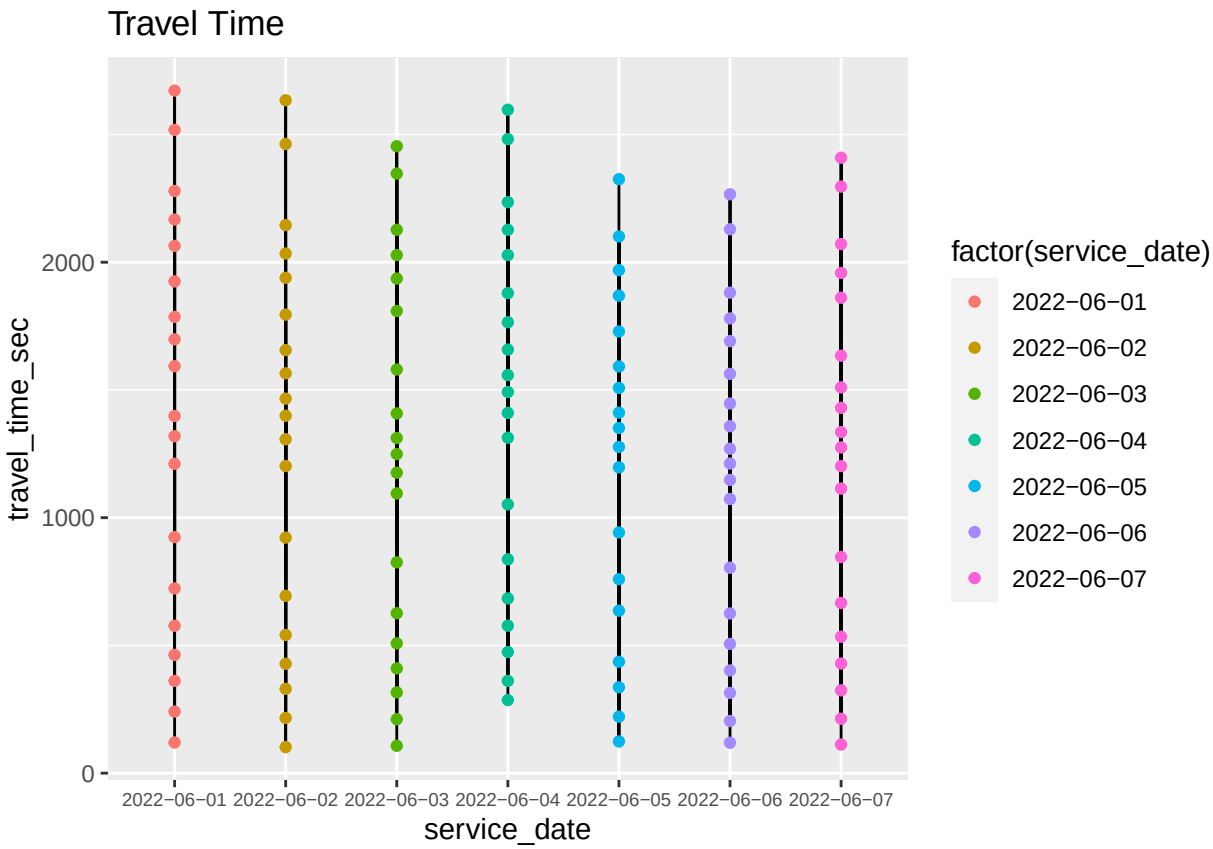

```
# Plot for April 2022
ggplot(data = dataapr) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time")
```



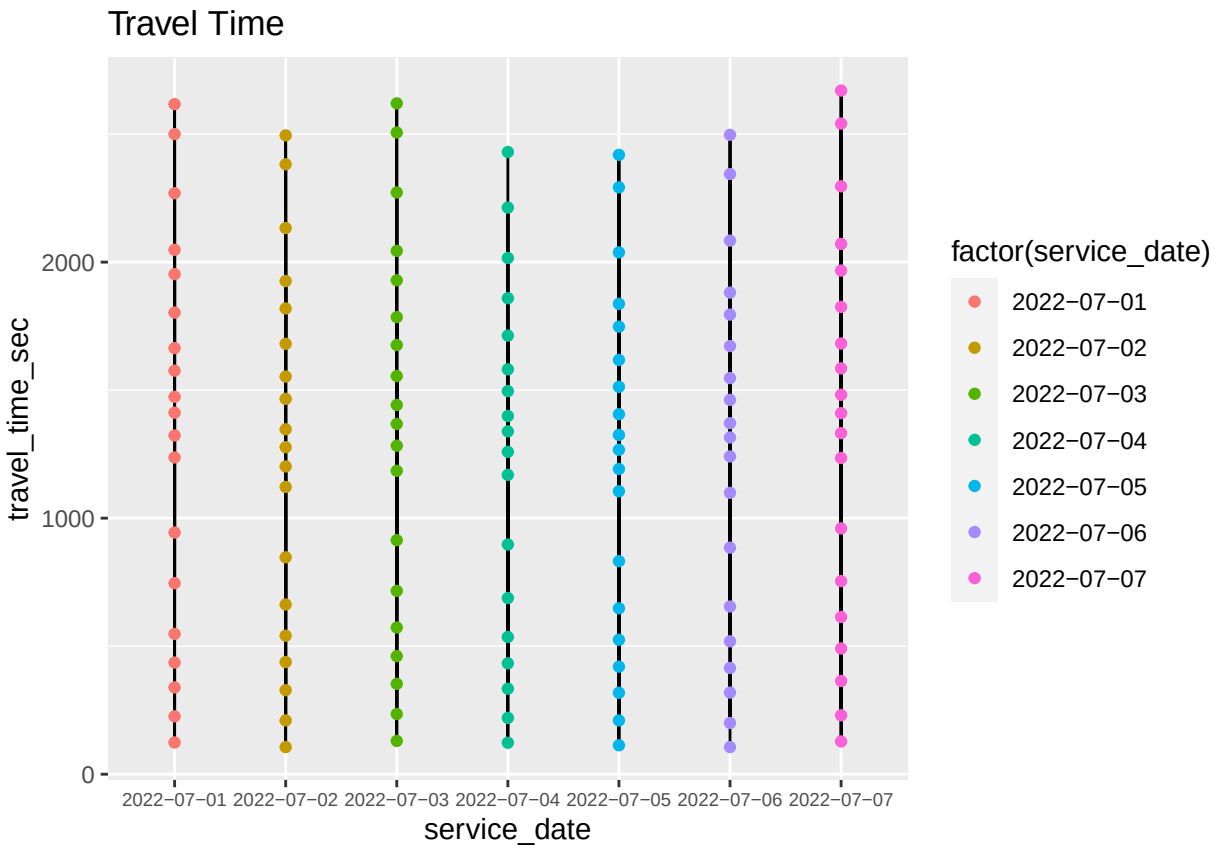

```
# Plot for May 2022
ggplot(data = datamay) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```



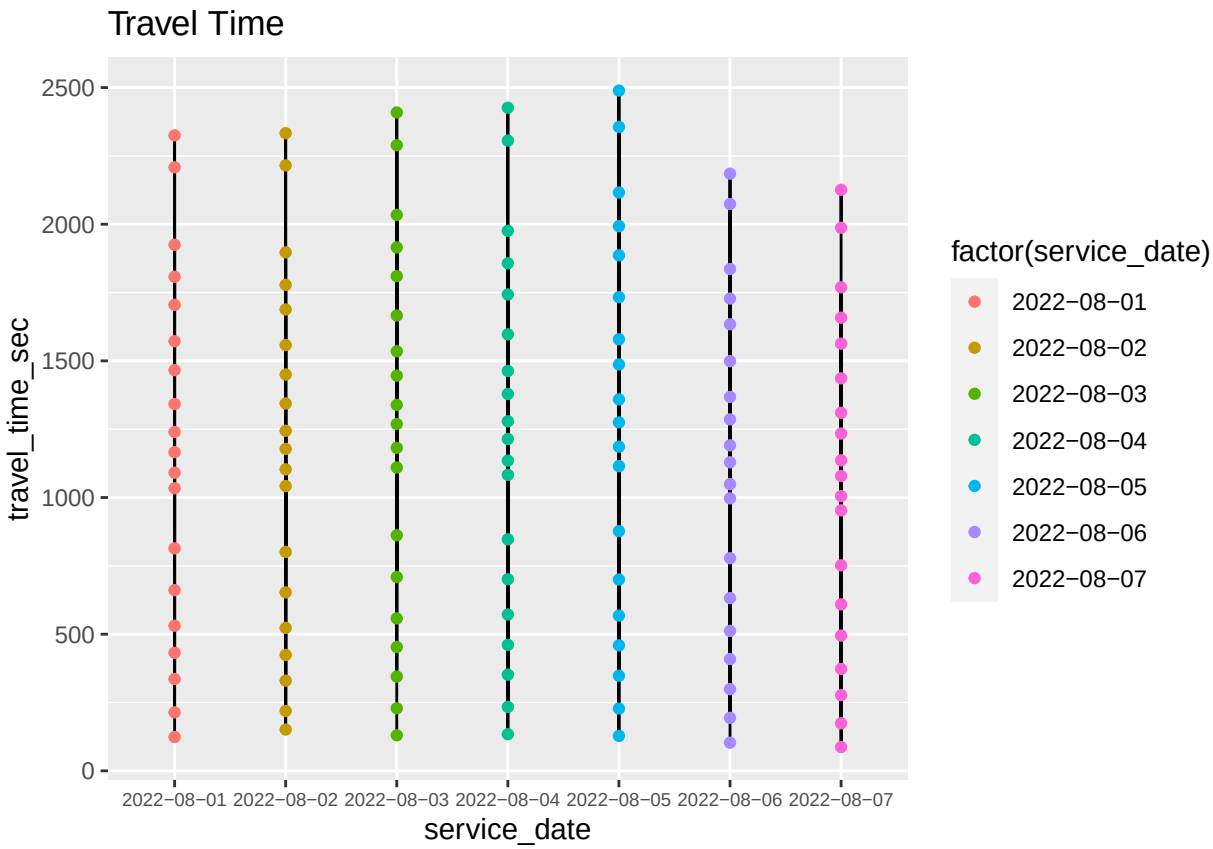

```
# Plot for June 2022
ggplot(data = datajun) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```



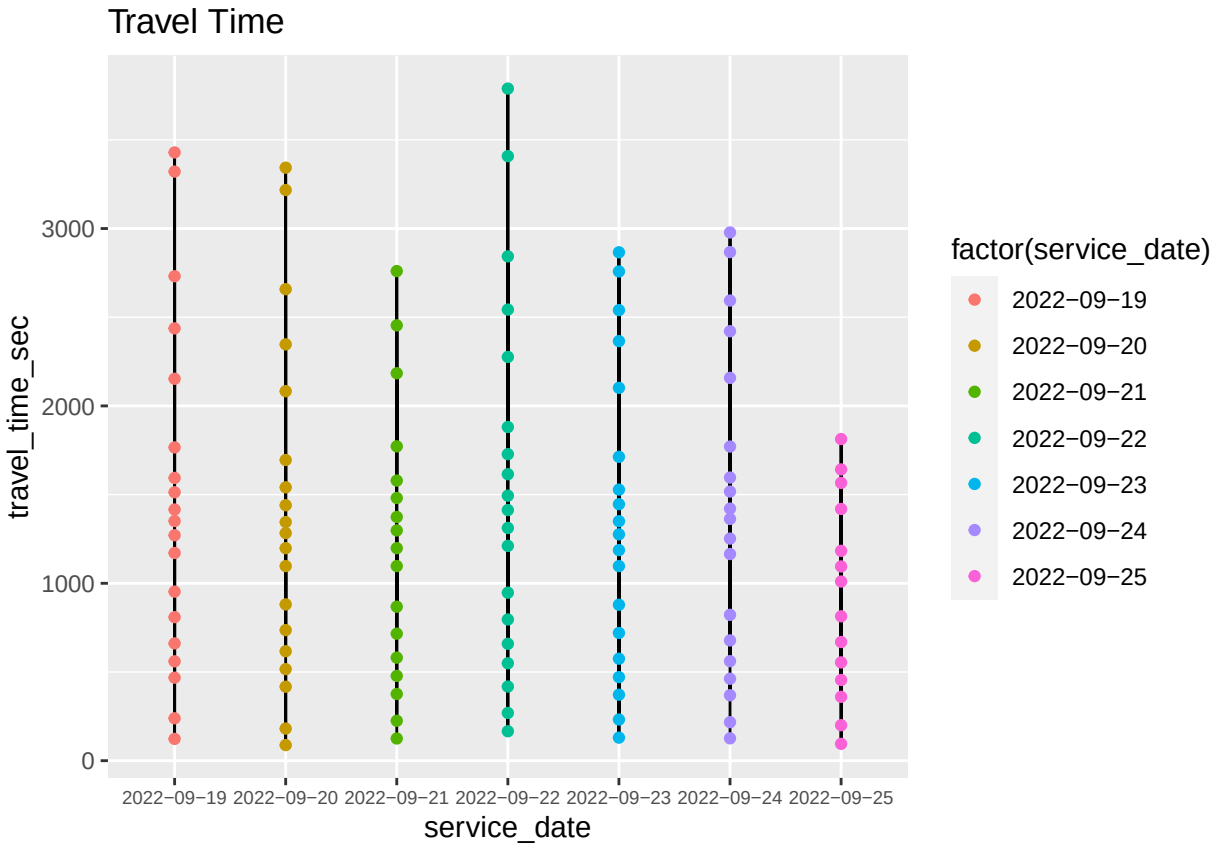

```
# Plot for July 2022
ggplot(data = datajul) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```




```
# Plot for August 2022
ggplot(data = dataaug) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```




```
# Plot for September 2022
ggplot(data = datasep) +
  aes(x = service_date, y = travel_time_sec) +
  geom_line() +
  geom_point(aes(x = service_date, y = travel_time_sec,color = factor(service_date),group = 1)) +
  ggtitle(label = "Travel Time") +
  theme(axis.text.x = element_text(size = 7))
```



The points on the same line had the same direction and the same destination. Different starting stops led to different travel times. Each line showed the distribution of travel time on the same day and each plot showed weekly distribution of travel time based on my selection.

To avoid the situation of having too many points on one line so that all points could not be seen clearly, I filtered the data with some criteria to show the distribution under one specific situation. Other situations can be shown using similar methods.