MA615 Final Project

Tong Sun

2022-12-17

Data

```
data <- rbind(Nov8_14, Dec6_12, Jan3_9, Feb7_13, Mar7_13, Apr4_10, May2_8, Jul4_10, Jul4_10, Aug1_7, Se
head(data)
```

```
## # A tibble: 6 x 8
##   service_date from_stop_id to_stop_id route_id direct~1 start~2 end_t~3 trave~4
##   <date>              <dbl>      <dbl> <chr>       <dbl>   <dbl>   <dbl>   <dbl>
## 1 2021-11-08          70002      70001 Orange          0   48100   48250     150
## 2 2021-11-08          70004      70001 Orange          0   65300   65595     295
## 3 2021-11-08          70008      70001 Orange          0   35073   35973     900
## 4 2021-11-08          70010      70001 Orange          0   37539   38174     635
## 5 2021-11-08          70012      70001 Orange          0   73012   73634     622
## 6 2021-11-08          70032      70001 Orange          0   84890   86852    1962
## # ... with abbreviated variable names 1: direction_id, 2: start_time_sec,
## #   3: end_time_sec, 4: travel_time_sec
```

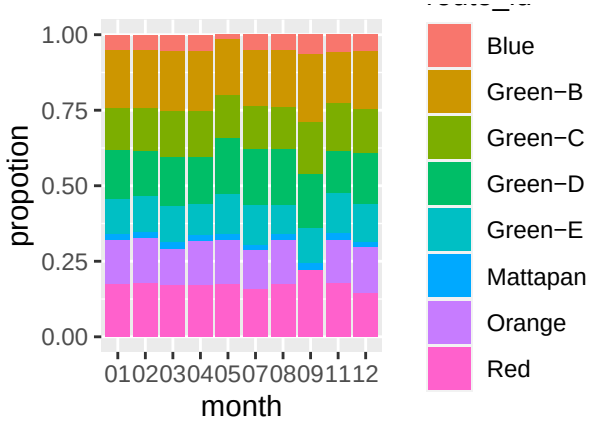
Distribution

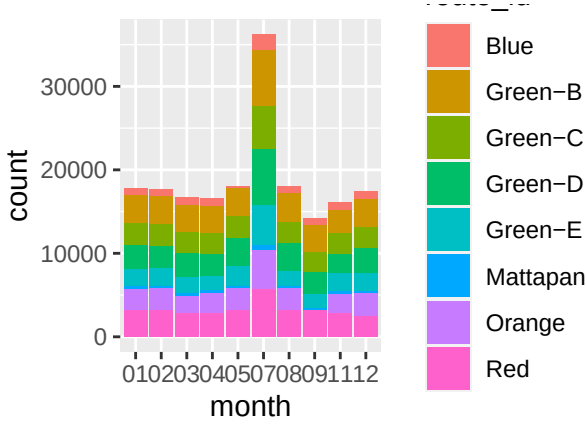
```
unique(data$route_id)
```

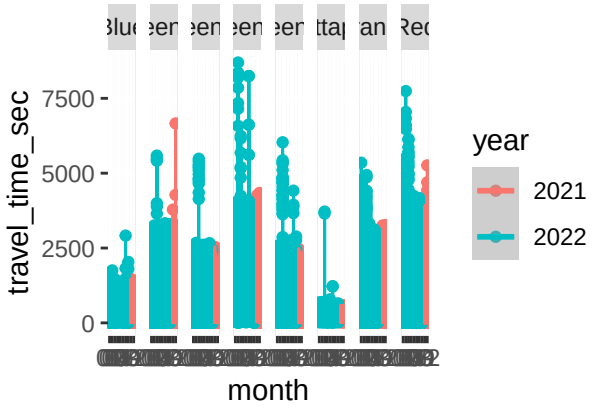
```
## [1] "Orange"   "Blue"     "Red"      "Green-B"  "Green-C"  "Green-D"  "Green-E"
## [8] "Mattapan"
```

```
## "Orange"
Orange <- data %>%
  filter(route_id == "Orange")

## "Blue"
Blue <- data %>%
  filter(route_id == "Blue")

## "Red"
Red <- data %>%
  filter(route_id == "Red")
```

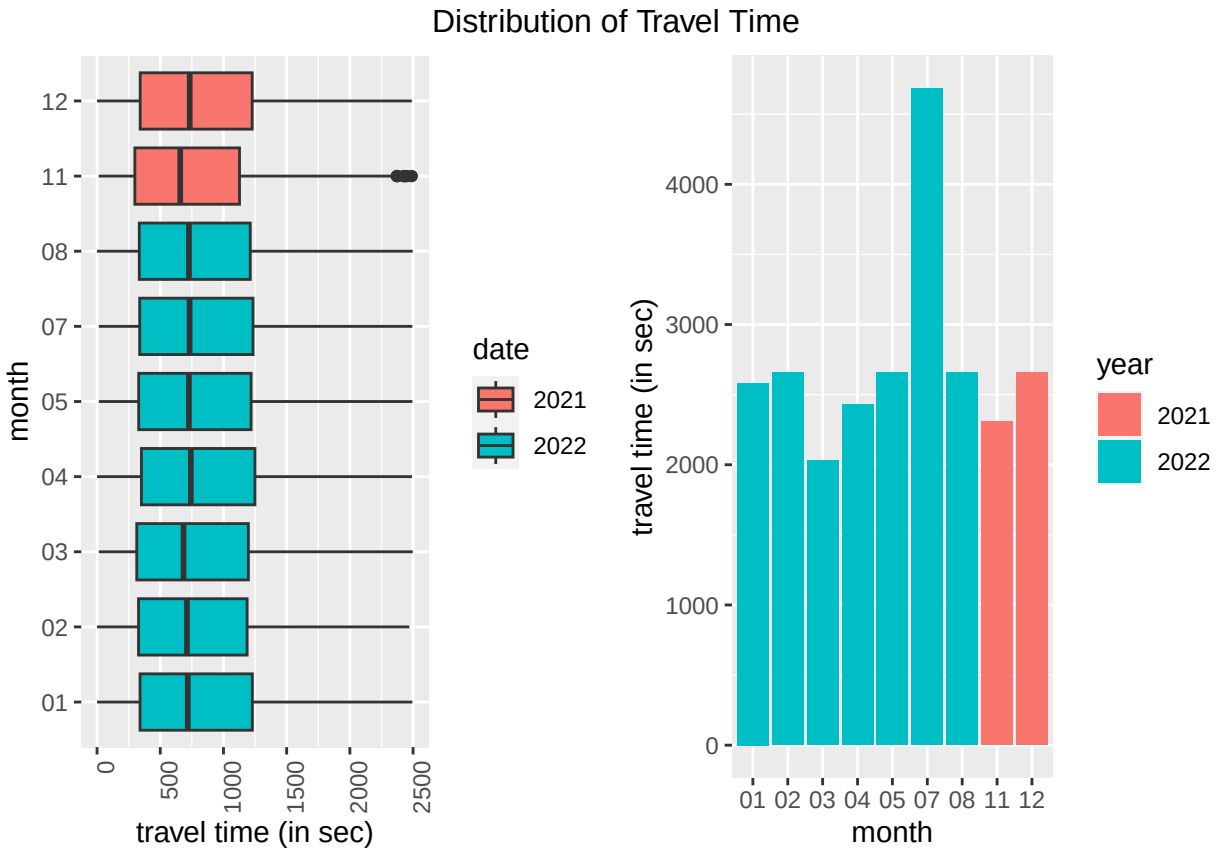
```
# ## "Green-B"
# Green-B <- data %>%
#   filter(route_id == "Green-B")
#
# ## "Green-C"
# Green-C <- data %>%
#   filter(route_id == "Green-C")
#
# ## "Green-D"
# Green-D <- data %>%
#   filter(route_id == "Green-D")
#
# ## "Green-E"
# Green-E <- data %>%
#   filter(route_id == "Green-E")

## "Mattapan"
Mattapan <- data %>%
  filter(route_id == "Mattapan")

df <- data %>%
  mutate(month = format(service_date, "%m"), year = format(service_date, "%Y")) %>%
  group_by(month, year) %>%
  select(month, year, travel_time_sec, route_id)
plot0 <- ggplot(df) +
  geom_bar(aes(x = month, fill = route_id), position = "fill") + labs(y='propotion')
plot01 <- ggplot(df) +
  geom_bar(aes(x = month, fill = route_id))
plot02 <- ggplot(data = df, mapping = aes(x = month, y = travel_time_sec, color=year)) +
  geom_point() +
  geom_line()+
  geom_smooth(method = 'loess',formula=y~x)+
  facet_grid(~route_id)
grid.arrange(plot0, plot01, plot02, ncol = 2)
```

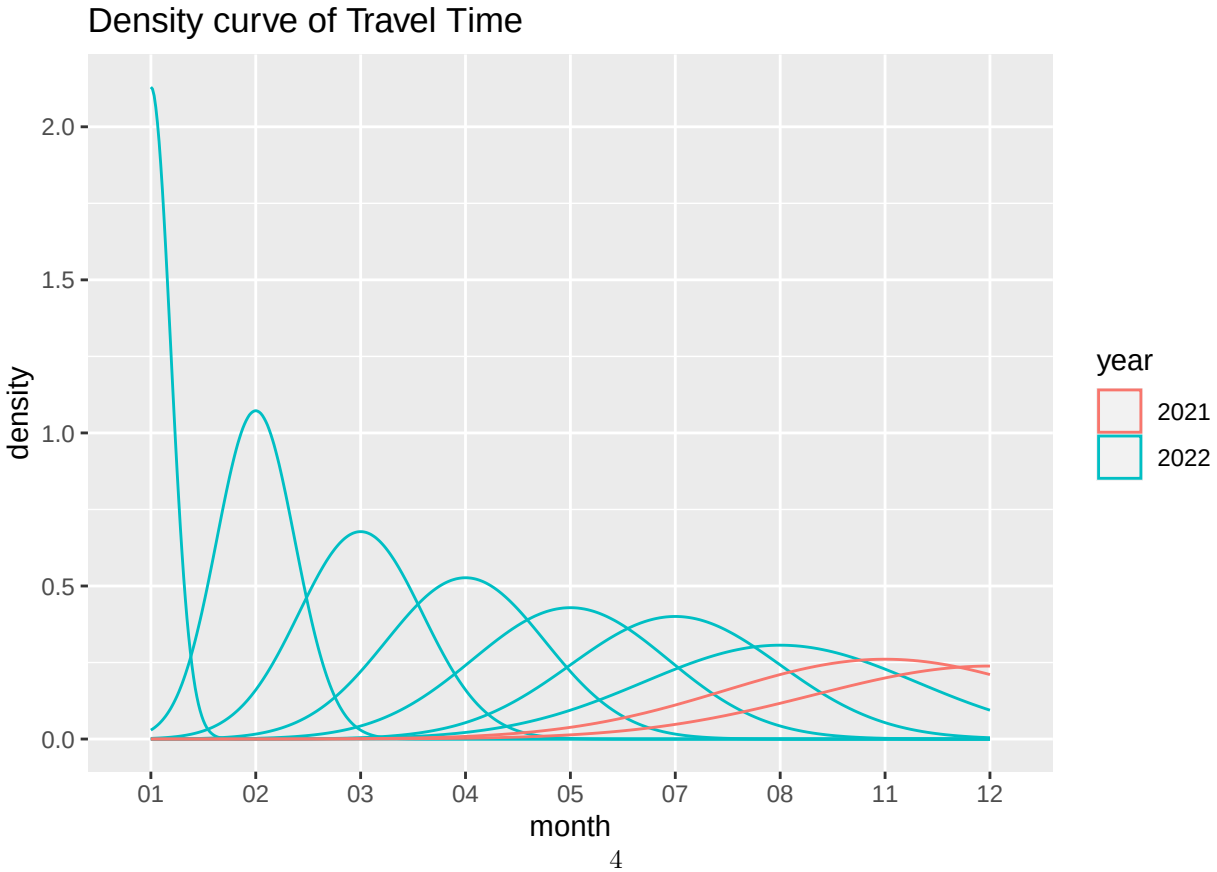

1.00
0.75
0.50
0.25
0.00
propotion
0102030405070809 1112
month
Blue
Green−B
Green−C
Green−D
Green−E
Mattapan
Orange
Red


30000
20000
10000
0
count
0102030405070809 1112
month
Blue
Green−B
Green−C
Green−D
Green−E
Mattapan
Orange
Red

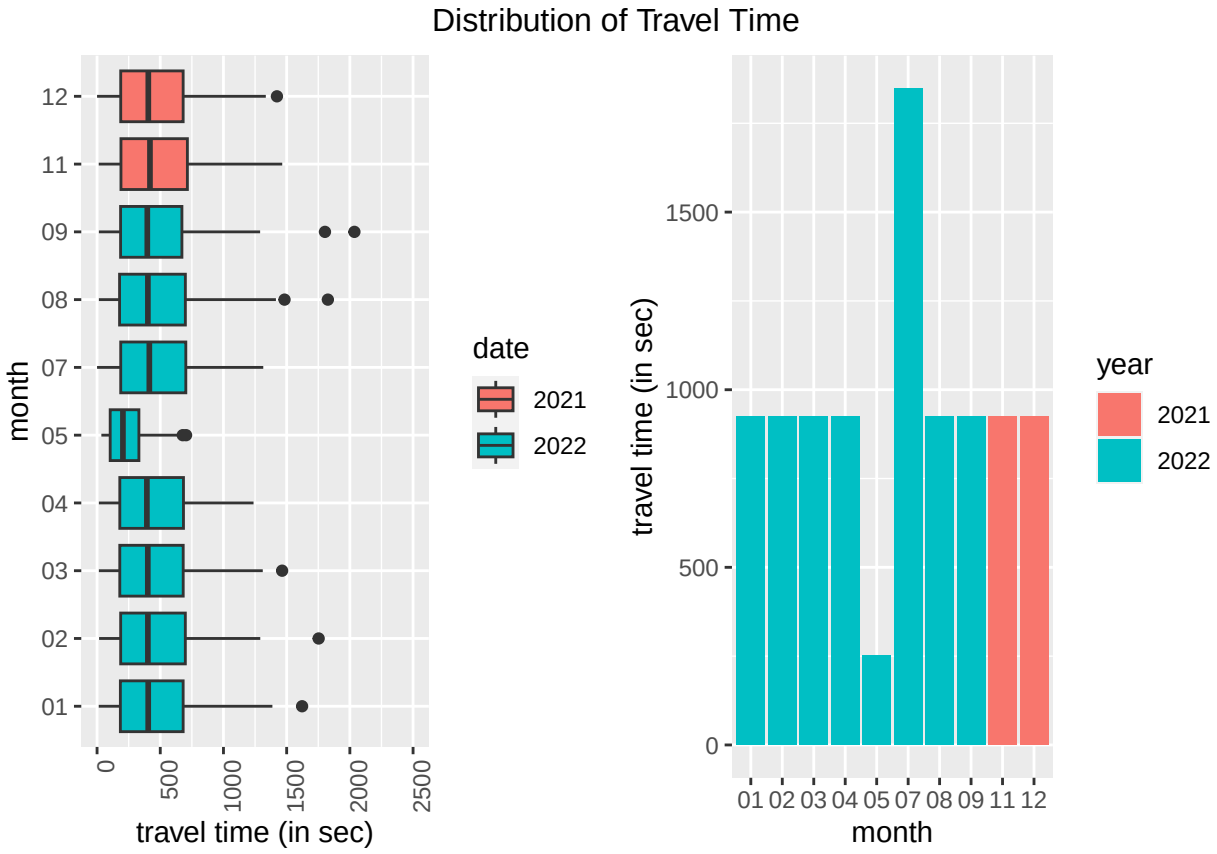

7500
5000
2500
0
travel_time_sec
month
year
2021
2022

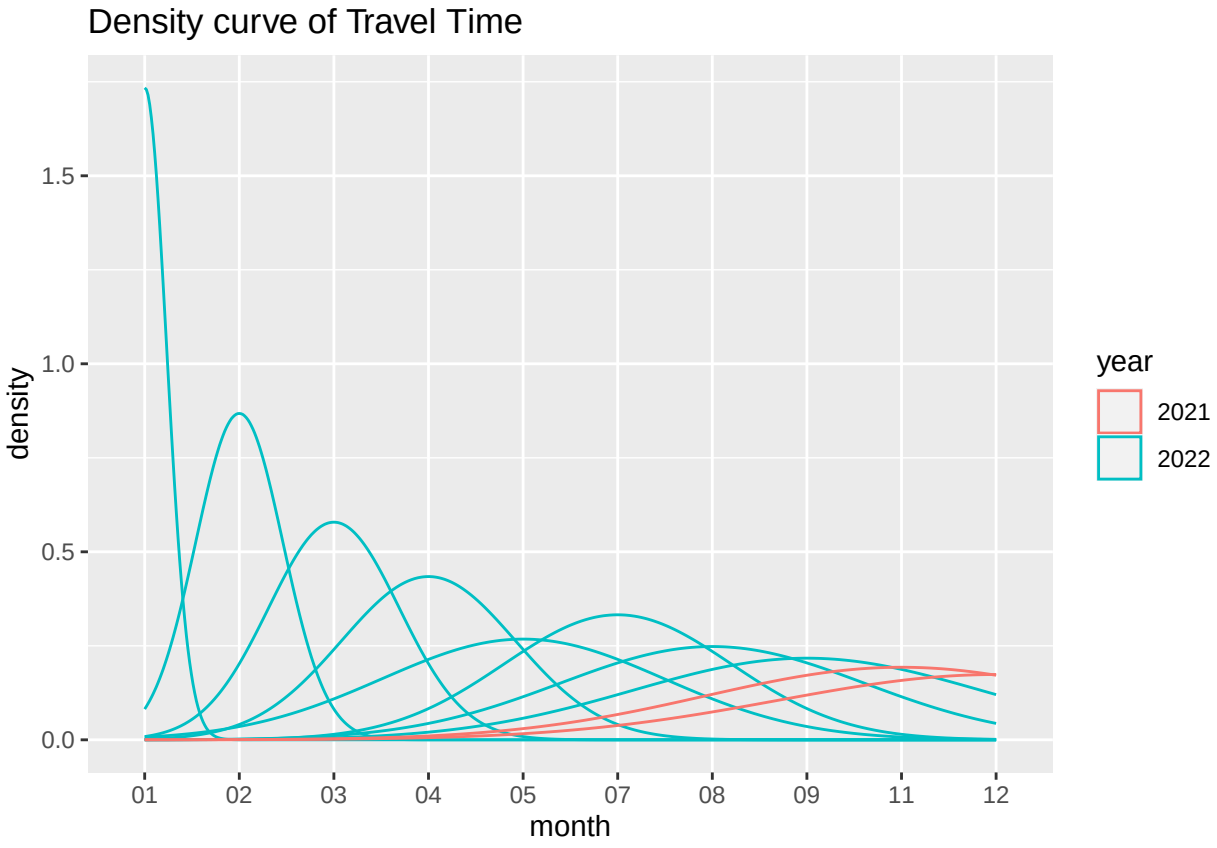
Orange


Distribution of Travel Time
month
travel time (in sec)
0
500
1000
1500
2000
2500
01
02
03
04
05
07
08
11
12
date
2021
2022
travel time (in sec)
month
year
2021
2022

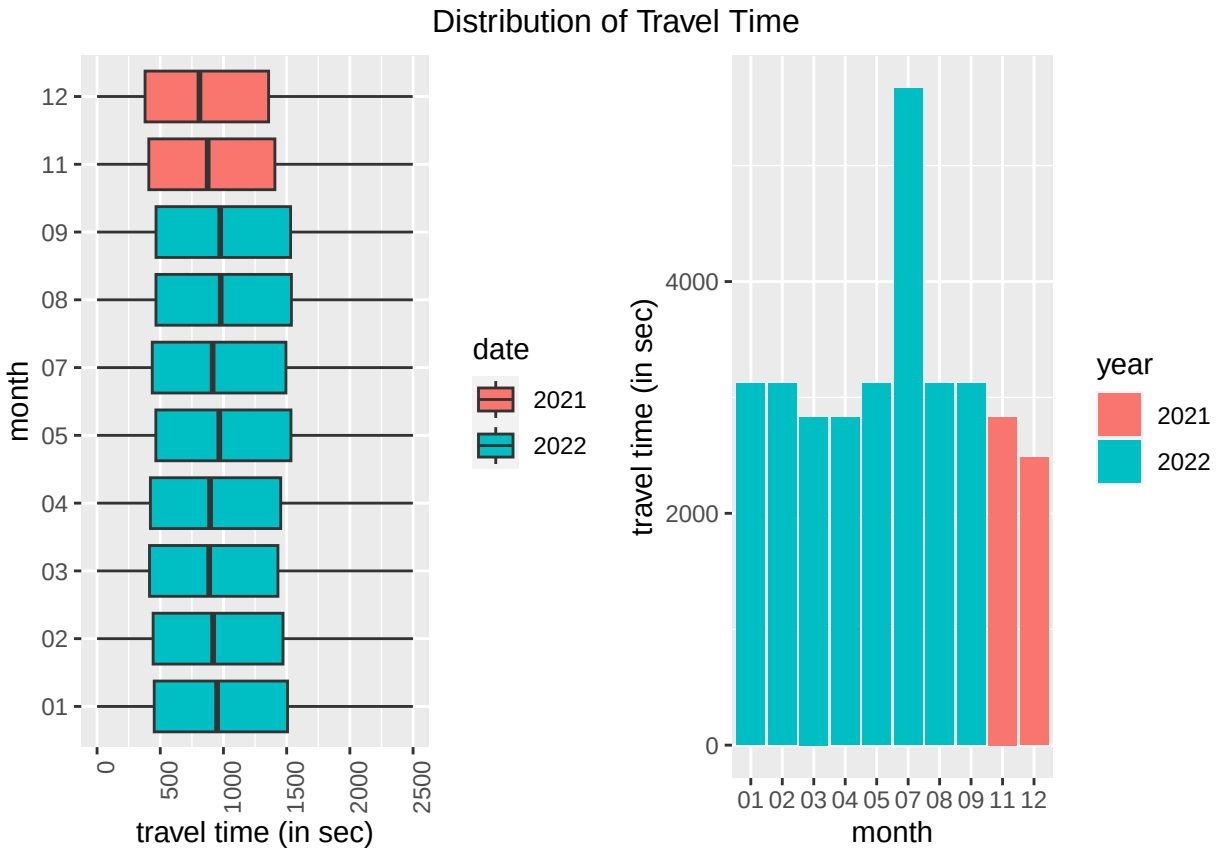

Density curve of Travel Time
density
month
year
2021
2022

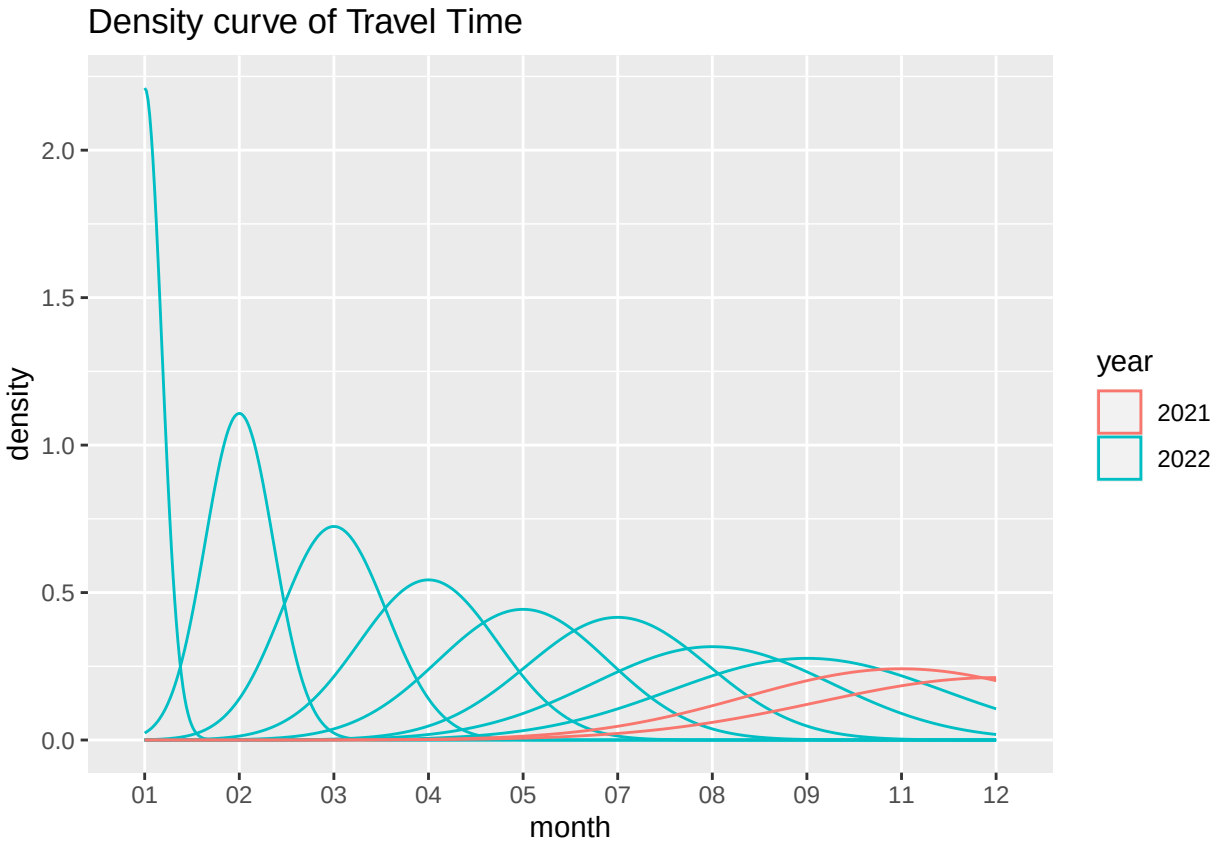
Blue


Distribution of Travel Time
month
travel time (in sec)
date
2021
2022
travel time (in sec)
month
year
2021
2022

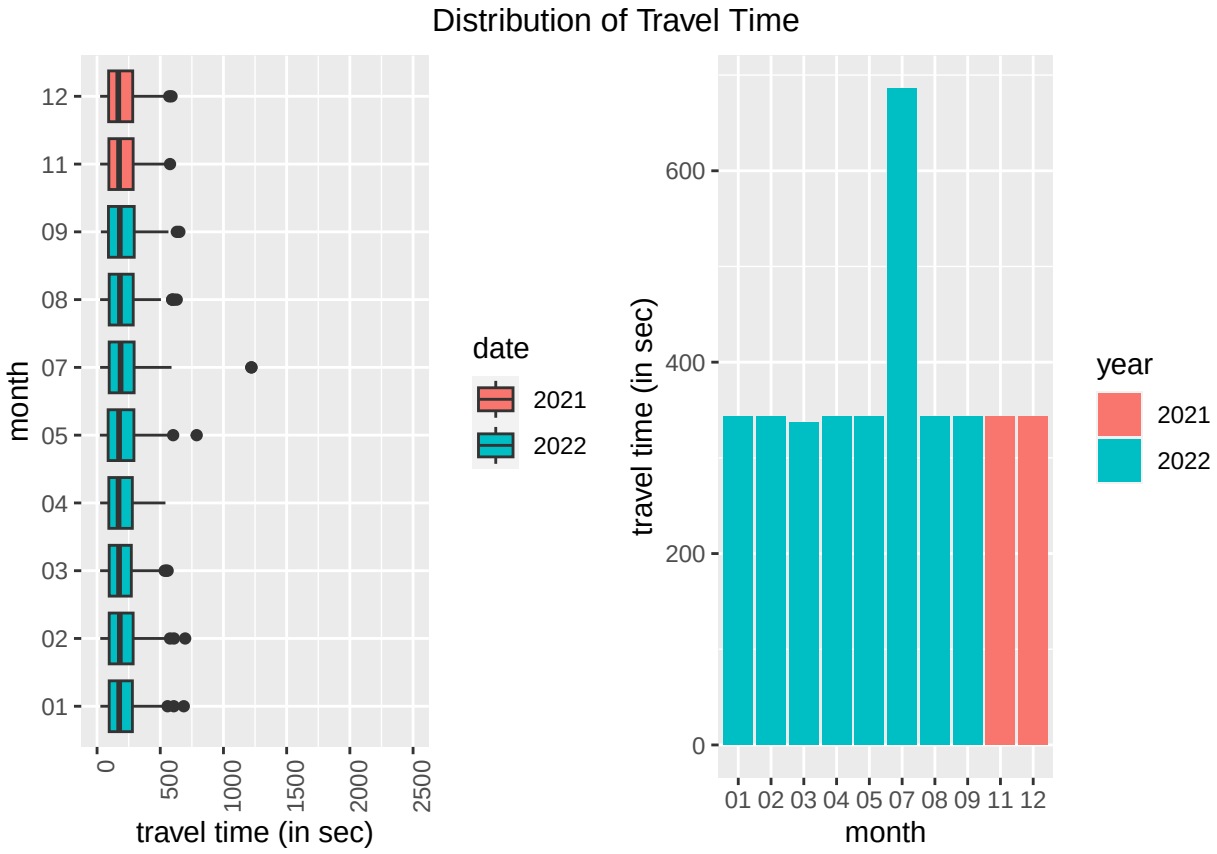

Density curve of Travel Time
density
month
year
2021
2022

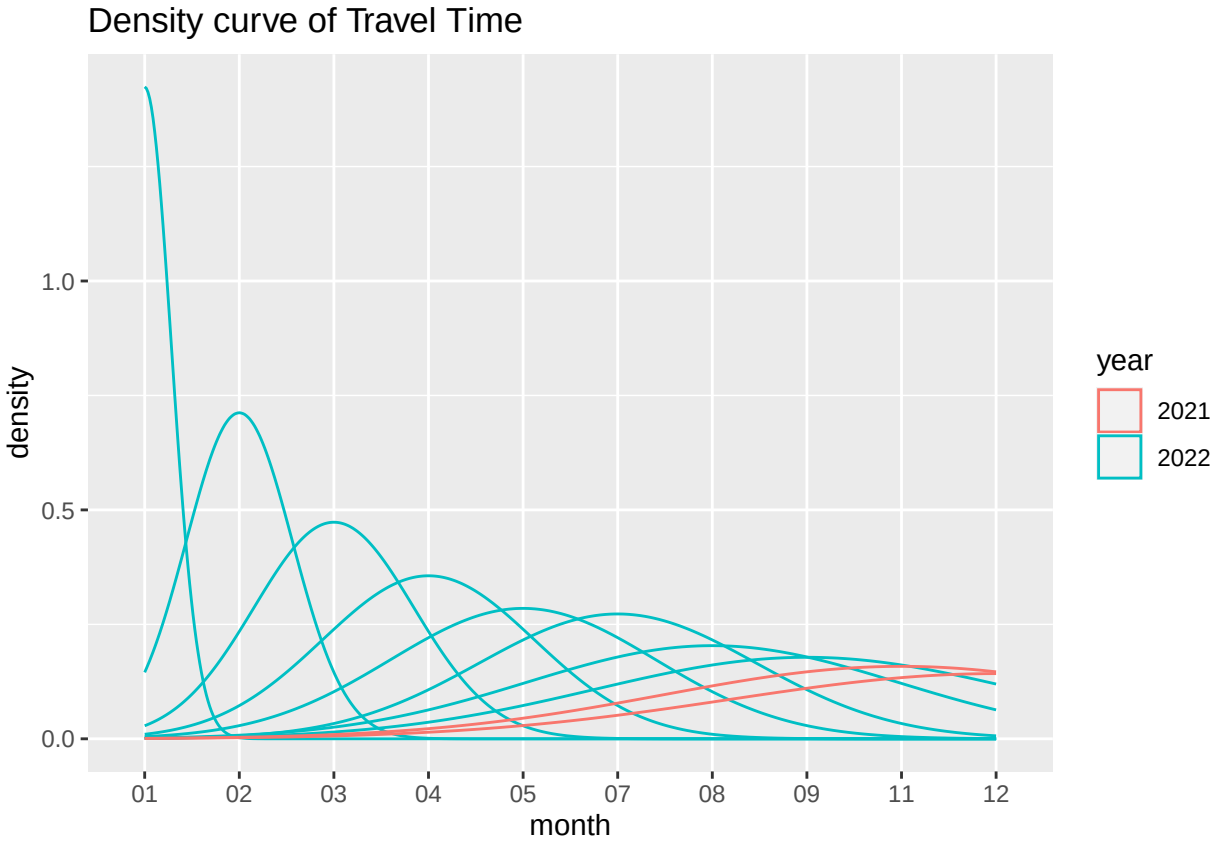
Red


Distribution of Travel Time
month
travel time (in sec)
0
500
1000
1500
2000
2500
01
02
03
04
05
07
08
09
11
12
date
2021
2022
travel time (in sec)
month
year
2021
2022


Density curve of Travel Time
density
month
year
2021
2022

Mattapan


Distribution of Travel Time
month
travel time (in sec)
date
2021
2022
travel time (in sec)
month
year
2021
2022


Density curve of Travel Time
density
month
year
2021
2022