MA615_final

Hanzhang Song

2022-12-17

Loading data

Since we are looking at MBTA data from November 2021 to October 2022, the data sets we need are 4th quarter in 2021 and the latest provided from 3rd quarter in 2022, for both light rail lines(Green) and heavy rail lines(Blue, Orange, and Red). It ended up with eight data sets.

Sorting by week

Then, before selecting weeks, I assigned week numbers according to dates.

Selecting weeks

Now, we can select weeks randomly from 11/21 to 10/22. Twelve weeks are needed in total from 8 data sets.

Extracting data

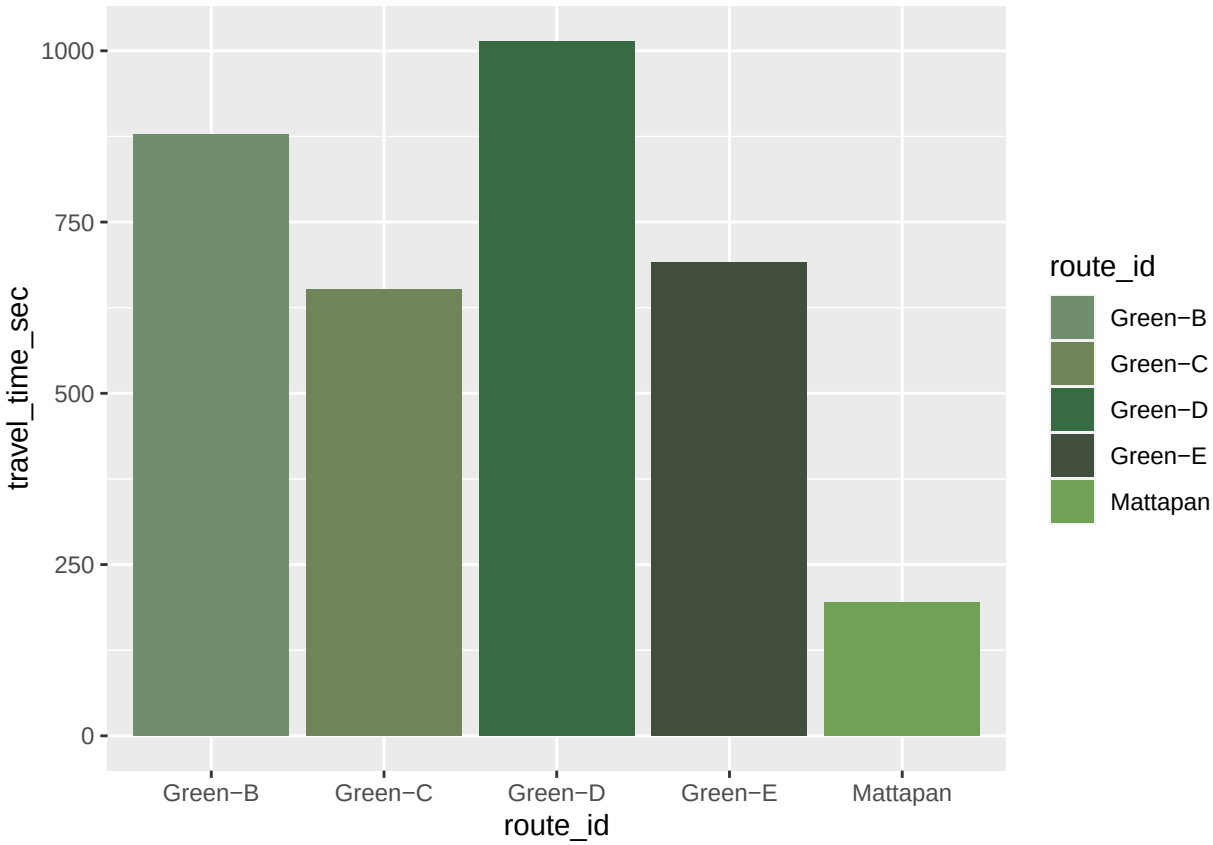
After having week numbers, we can extract them from the data:

EDA

I'm curious about how travel times differ between routes, and if they are also different for light and heavy rails.

Light rails in 2021 4th quarter

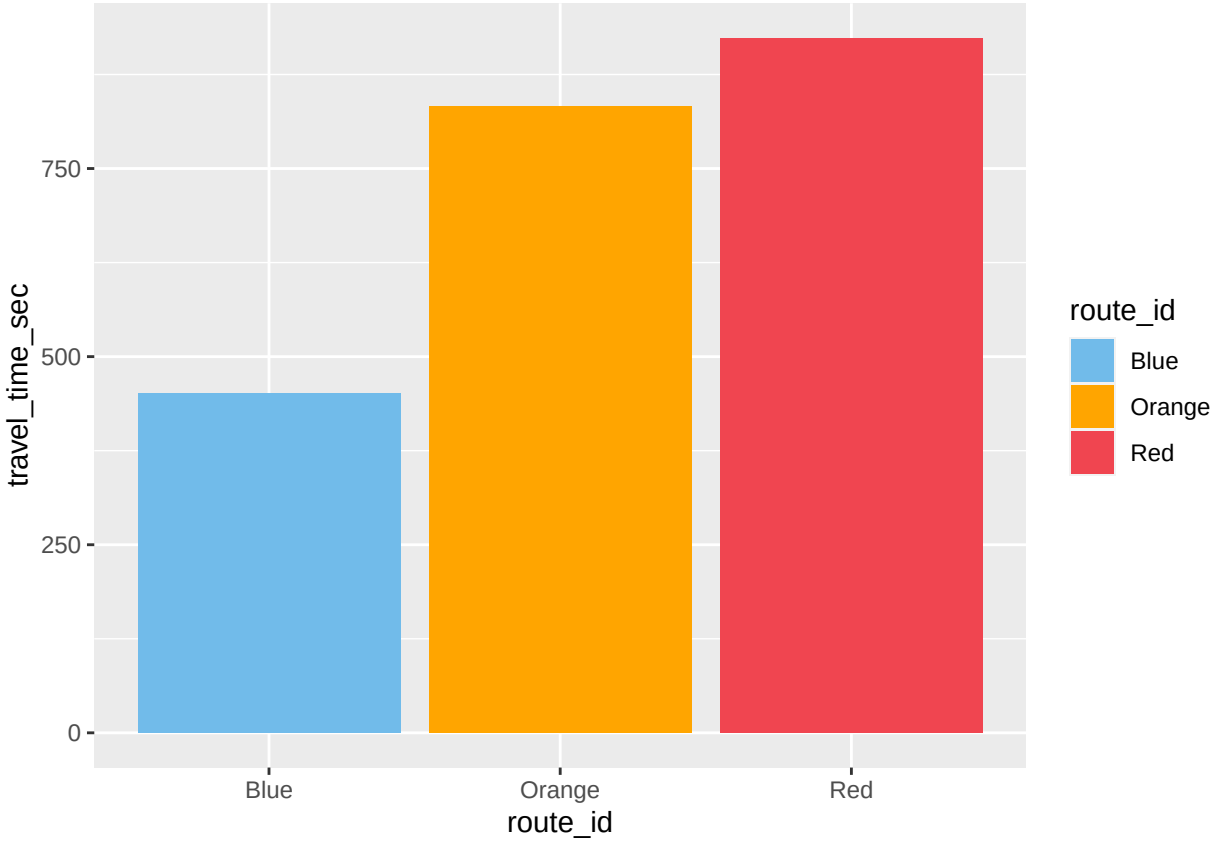
```
## # A tibble: 5 x 2
##   route_id mean_travel_time
##   <chr>               <dbl>
## 1 Green-B              878.
## 2 Green-C              651.
## 3 Green-D             1014.
## 4 Green-E              691.
## 5 Mattapan             195.
```



According to the plots, Green-D had the longest travel time of more than 1000 seconds on average from origin stop to destination stop, converting to approximately 17 minutes. All lines ran more than 10 minutes from origin to destination.

Heavy rails in 2021 4th quarter

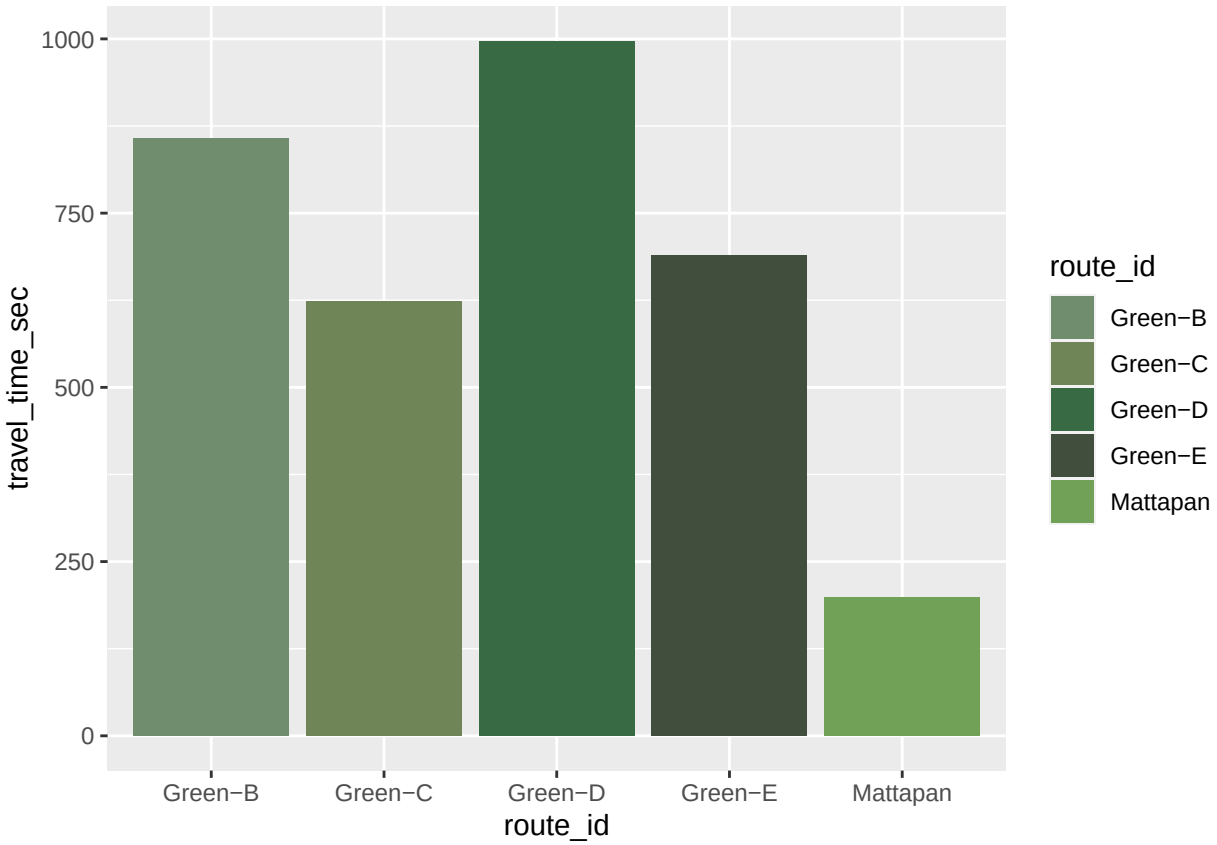
```
## # A tibble: 3 x 2
##   route_id mean_travel_time
##   <chr>               <dbl>
## 1 Blue                 451.
## 2 Orange               832.
## 3 Red                  923.
```



Red line had the longest travel time. It took about 931 seconds(15.5 minutes) travelling from origin to destination. The shortest lines, Blue line, needed 463 seconds(7.71 minutes) to ran the whole ride.

Light Rails in 2022 1st quarter

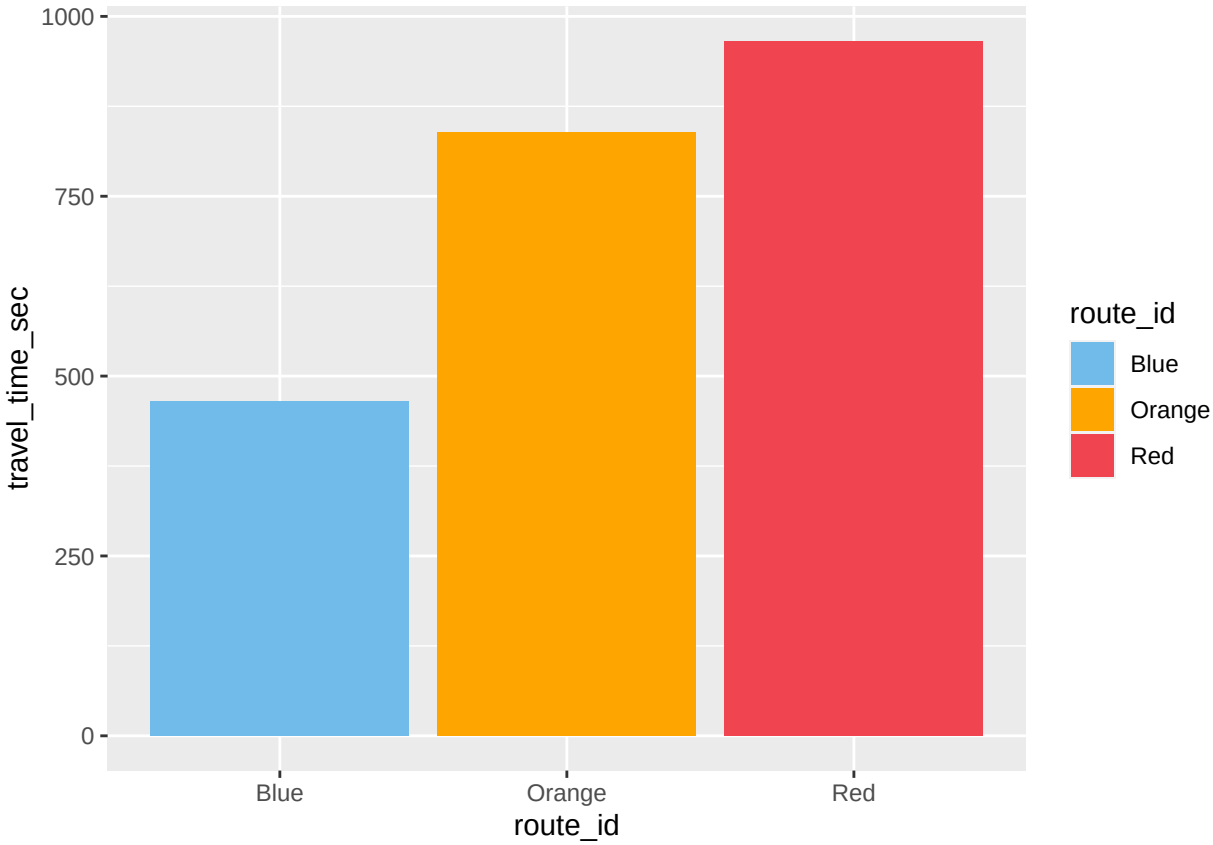
```
## # A tibble: 5 x 2
##   route_id mean_travel_time
##   <chr>               <dbl>
## 1 Green-B              857.
## 2 Green-C              623.
## 3 Green-D              997.
## 4 Green-E              689.
## 5 Mattapan             199.
```



We can see that Green-C and Green-D had close frequencies, but their travel time vary a lot. Green-C needed 623 seconds(10.38 minutes) and Green-D needed 997 seconds(16.62 minutes) to ran the whole ride. Noticeably, the journeys were longer in 2021 4th quarter. The reason could be weather or temperature since it took longer if there was snow in winter.

Heavy Rails in 2022 1st quarter

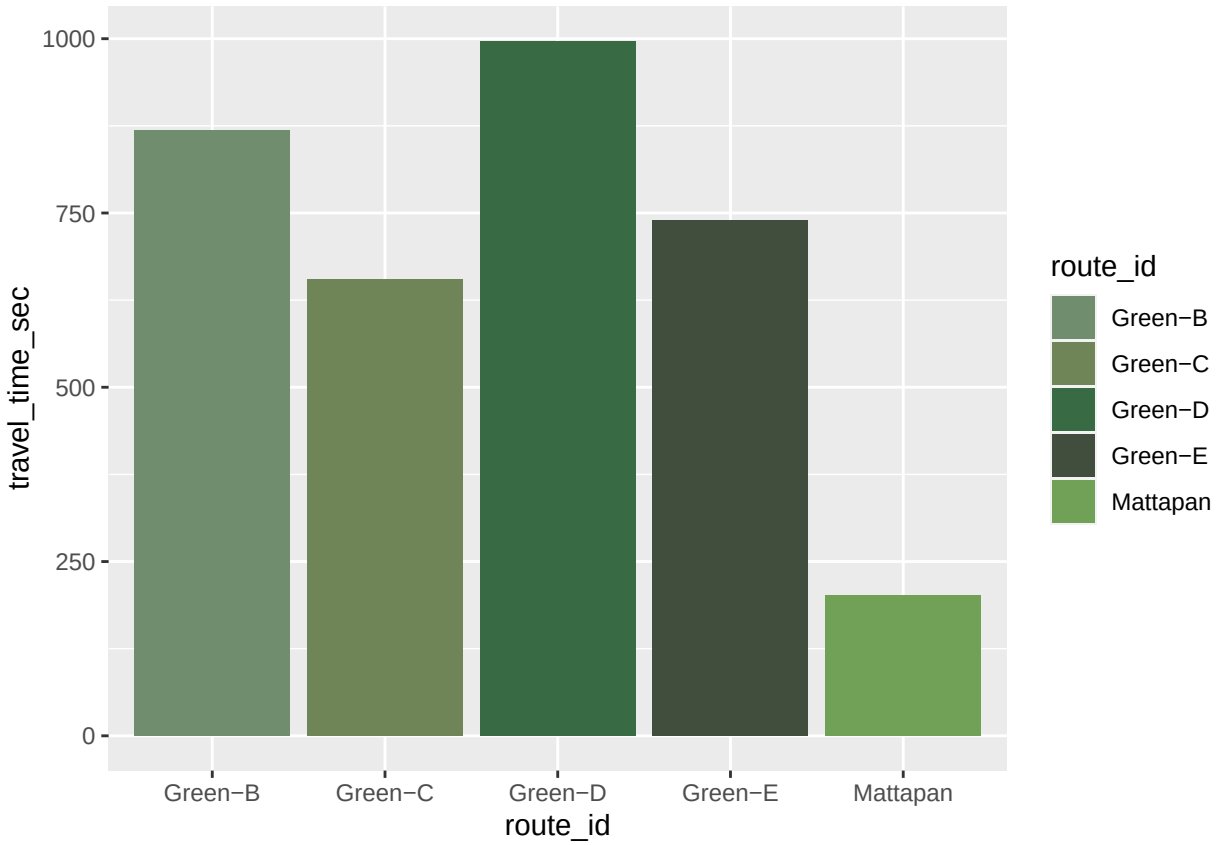
```
## # A tibble: 3 x 2
##   route_id mean_travel_time
##   <chr>               <dbl>
## 1 Blue                 464.
## 2 Orange               839.
## 3 Red                  966.
```



Comparing with 2021 4th quarter, Red Line was still the most common and longest line. However, unlike light rails, the traveling time increased for all lines.

Light Rails in 2022 2nd quarter

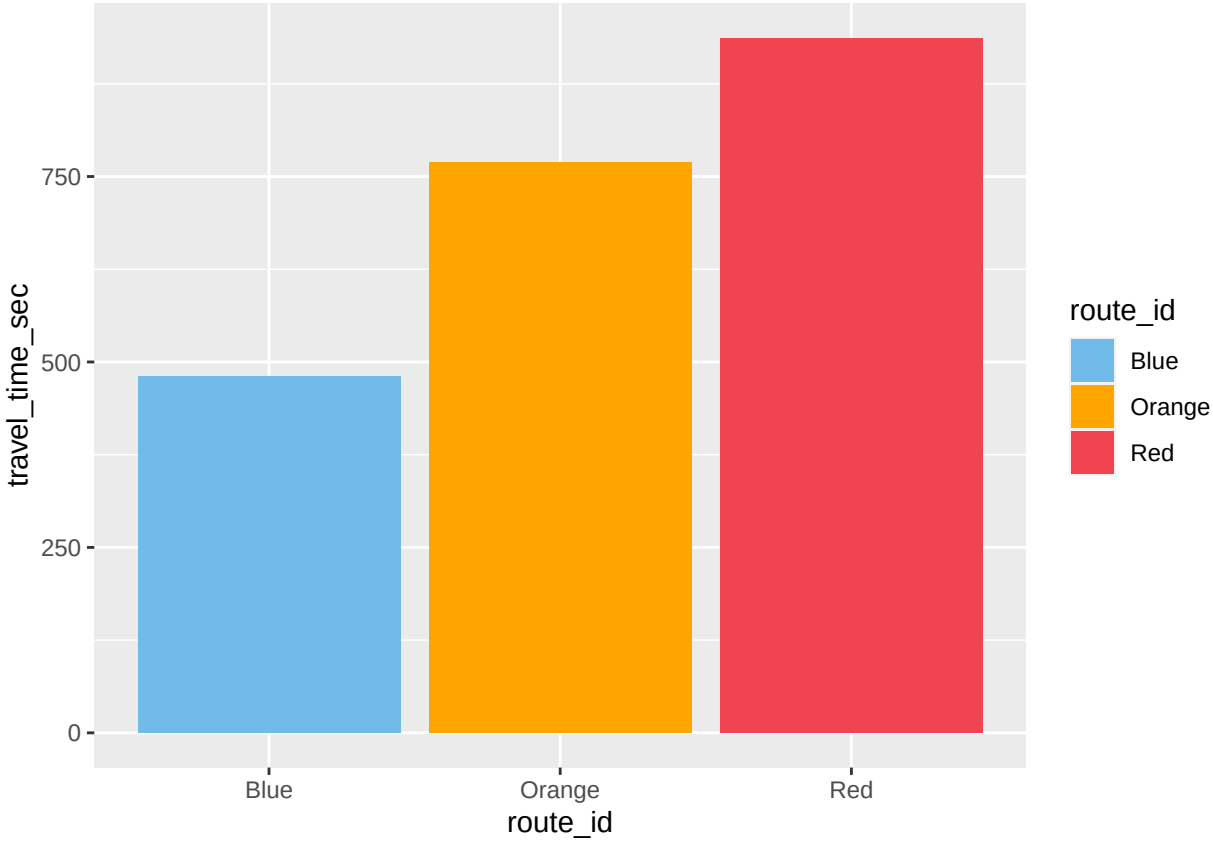
```
## # A tibble: 5 x 2
##   route_id mean_travel_time
##   <chr>               <dbl>
## 1 Green-B              869.
## 2 Green-C              655.
## 3 Green-D              997.
## 4 Green-E              740.
## 5 Mattapan             202.
```



The average travel time was longer in this quarter than last for light rails, so my earlier prediction of weather might not be the case.

Heavy Rails in 2022 2nd quarter

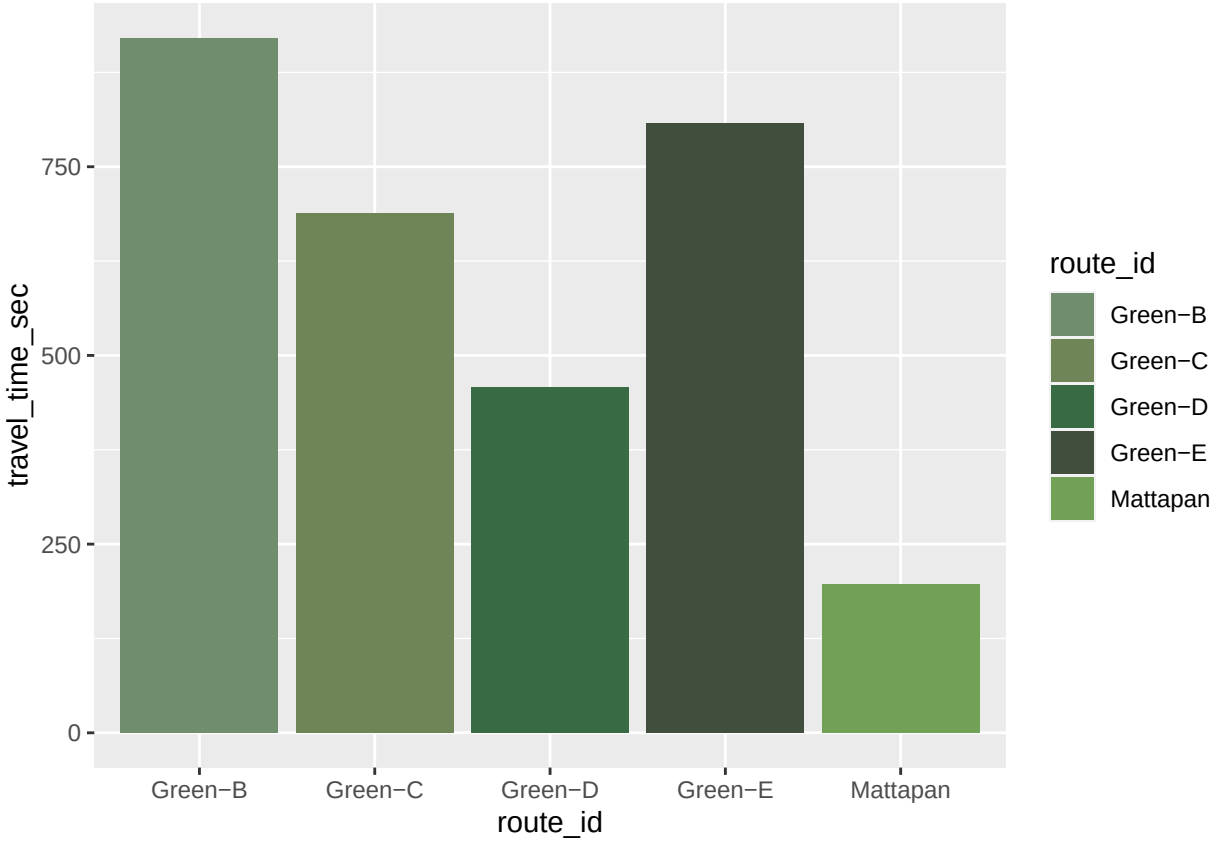
```
## # A tibble: 3 x 2
##   route_id mean_travel_time
##   <chr>               <dbl>
## 1 Blue                 480.
## 2 Orange               769.
## 3 Red                  937.
```



For heavy rails, only Blue line had longer travel time in 2022 2nd quarter.

Light Rails in 2022 3rd quarter

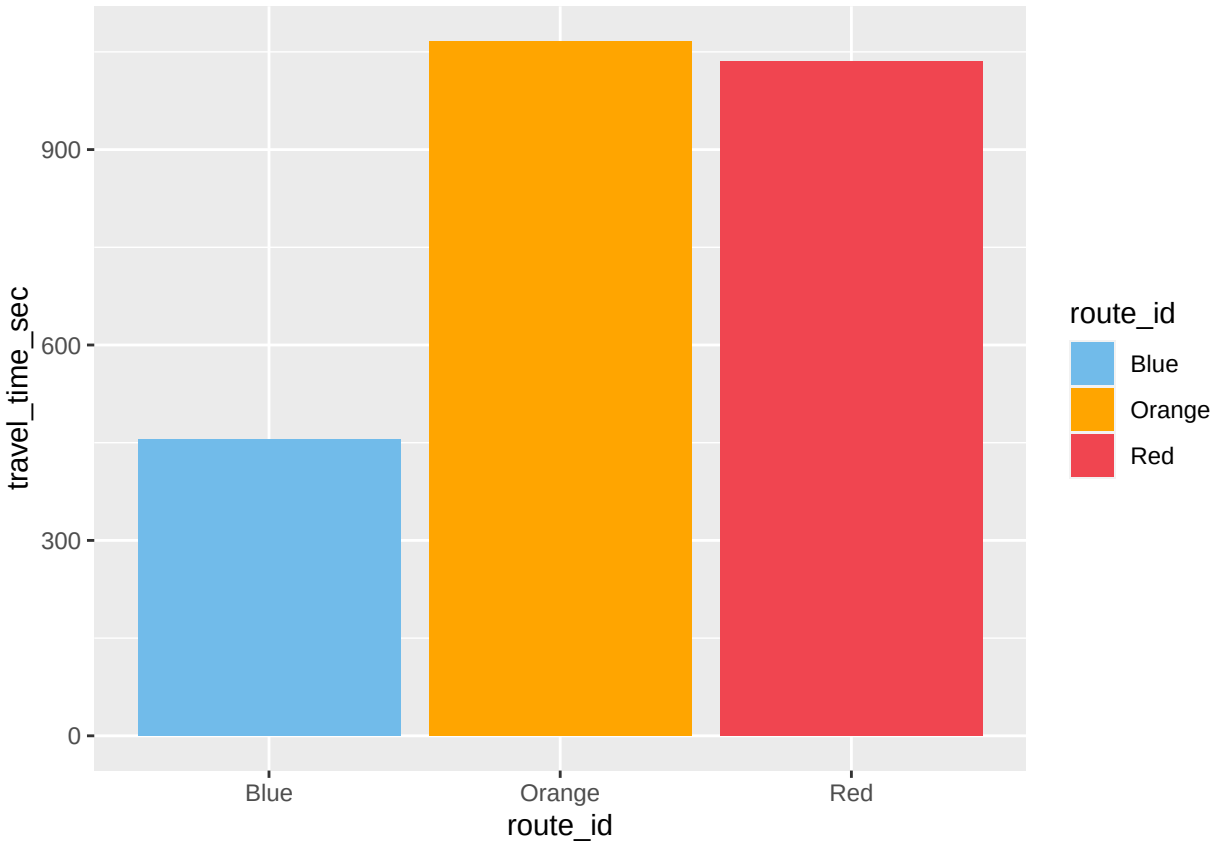
```
## # A tibble: 5 x 2
##   route_id mean_travel_time
##   <chr>               <dbl>
## 1 Green-B              921.
## 2 Green-C              688.
## 3 Green-D              457.
## 4 Green-E              807.
## 5 Mattapan             197.
```



Moving to the year of 2nd quarter in 2022, all Green lines had steadily grown except D line dropped a lot, from 997 seconds to 457 seconds(9 minutes). Green-D had also became the least common line in this quarter.

Heavy Rails in 2022 3rd quarter

```
## # A tibble: 3 x 2
##   route_id mean_travel_time
##   <chr>               <dbl>
## 1 Blue                 455.
## 2 Orange              1067.
## 3 Red                 1035.
```



Orange and Red lines had very close travel times this quarter, only a 20-second difference was observed.