Untitled

Jin Yan

2022-12-13

Data import

According to the data characteristics, I selected the data from November 2021 to October 2022, and randomly selected one week of data each months. The data selected four data sets of route, route_pattern, direction, and trips, and carried out according to route_id merge

data connection

```
data<-rbind(data1,data2,data3,data4,data5,data6,data7,data8,data9,data10,data11,data12)
nrow(data %>% filter(route_id=="Red"))
```

```
## [1] 1008136
```

```
nrow(data %>% filter(route_id=="Orange"))
```

```
## [1] 122448
```

```
nrow(data %>% filter(route_id=="Green-B"))
```

```
## [1] 157992
```

```
nrow(data %>% filter(route_id=="Green-C"))
```

```
## [1] 175056
```

```
nrow(data %>% filter(route_id=="Green-D"))
```

```
## [1] 465496
```

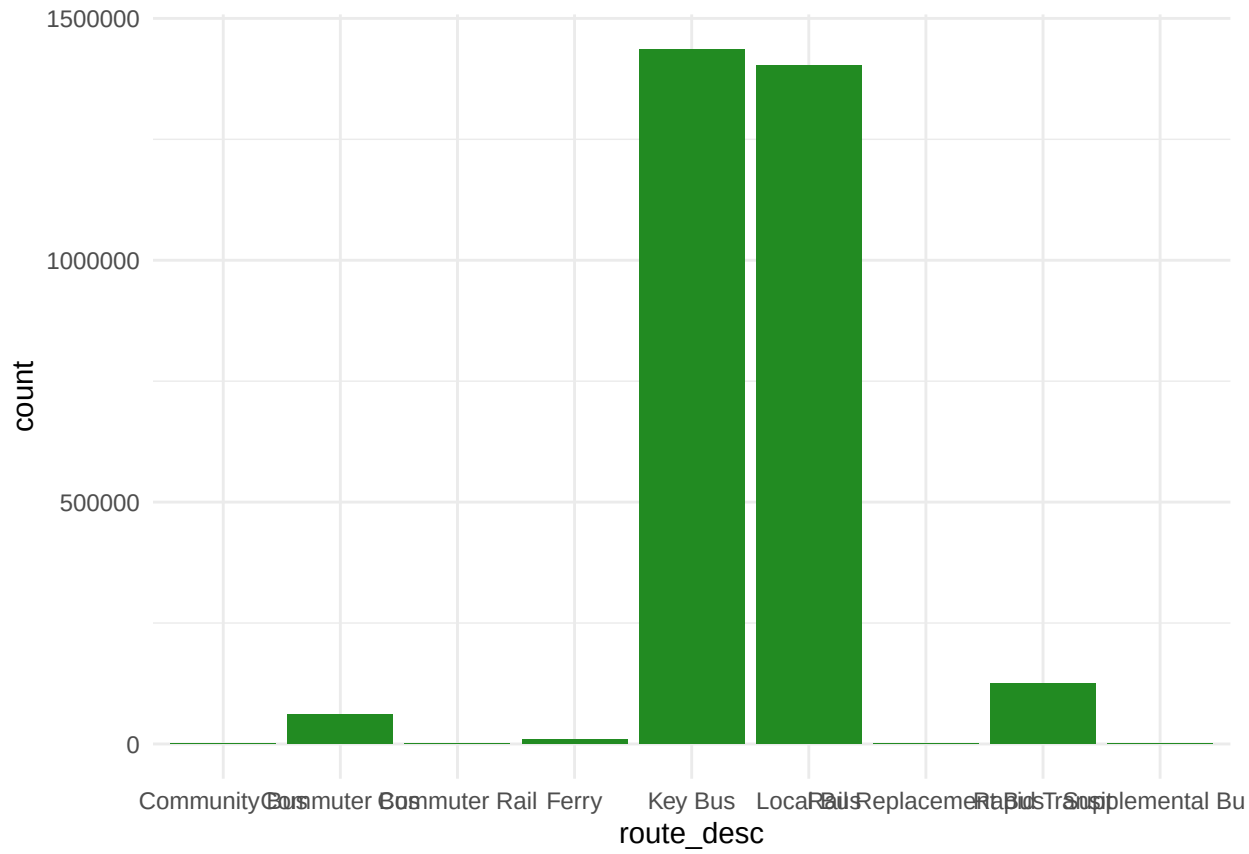
```
nrow(data %>% filter(route_id=="Green-E"))
```

```
## [1] 192064
```

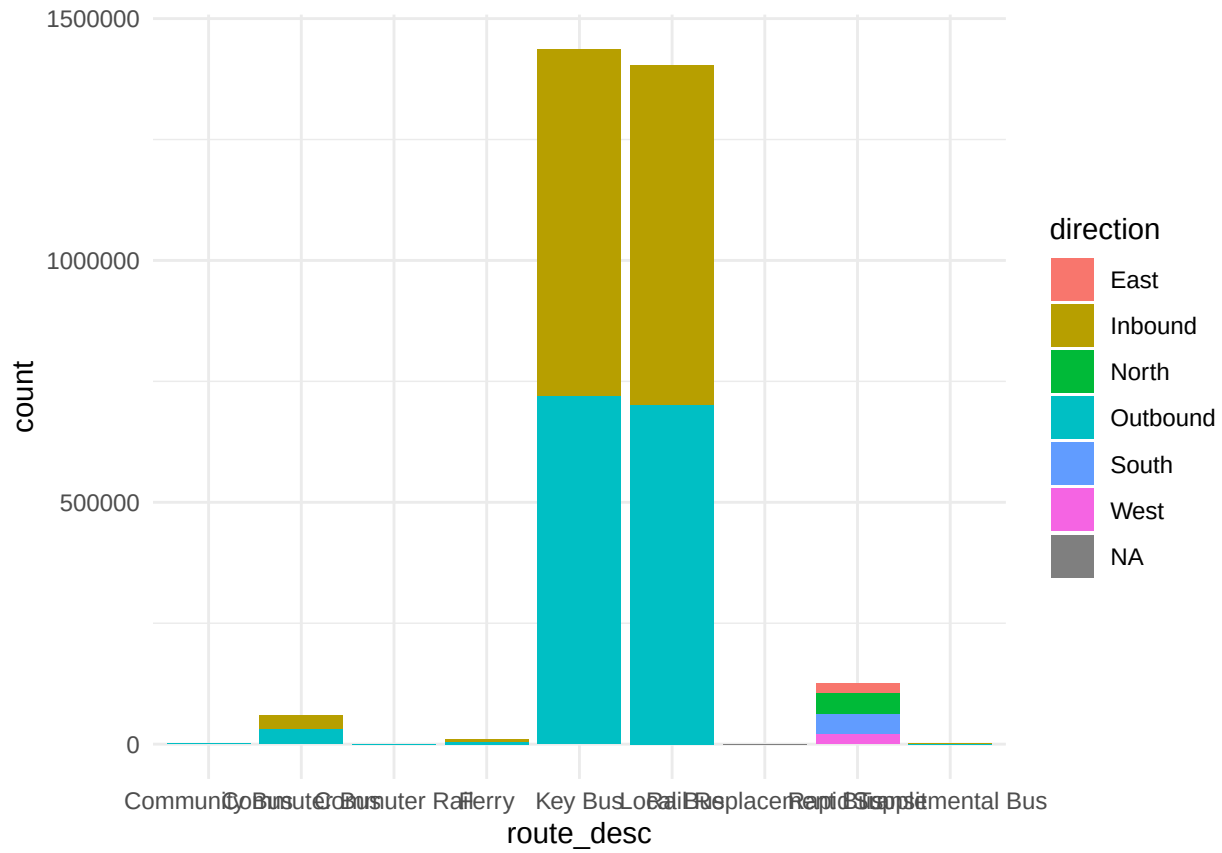
In the data set, there are many kinds of transportation. I chose the most commonly used subway lines in Boston:Red,Orange,Green-B,Green-C,Green-D,Green-E,there are 1008136 data for Red,122448 data for Orange,157992 data for Green-B,175056 data for Green-C,465496 data for Green-D,192064 data for Green-E.

Data present

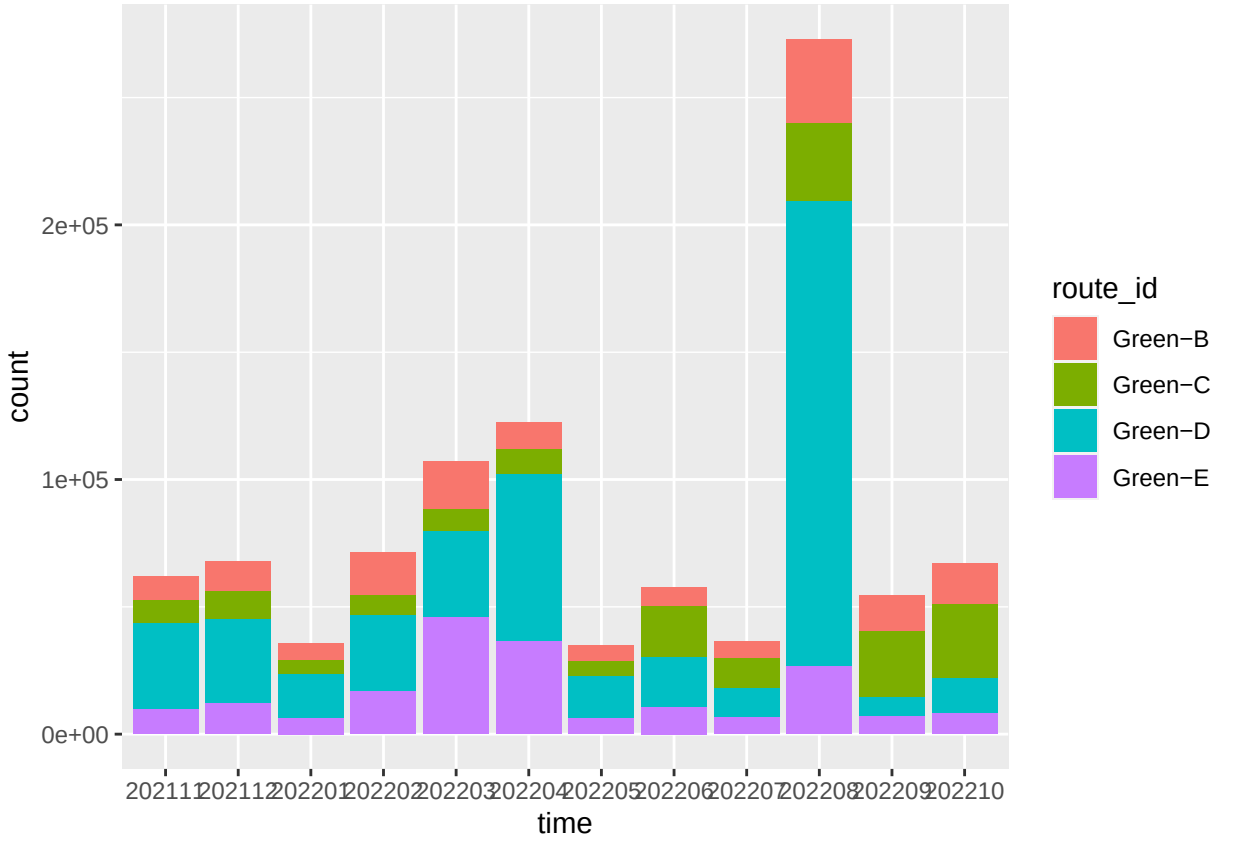
```
data %>%
 filter(!(route_desc %in% "")) %>%
 filter(!(route_pattern_time_desc %in% "") | is.na(route_pattern_time_desc)) %>%
 ggplot() +
  aes(x = route_desc) +
  geom_bar(fill = "#228B22") +
  theme_minimal()
```



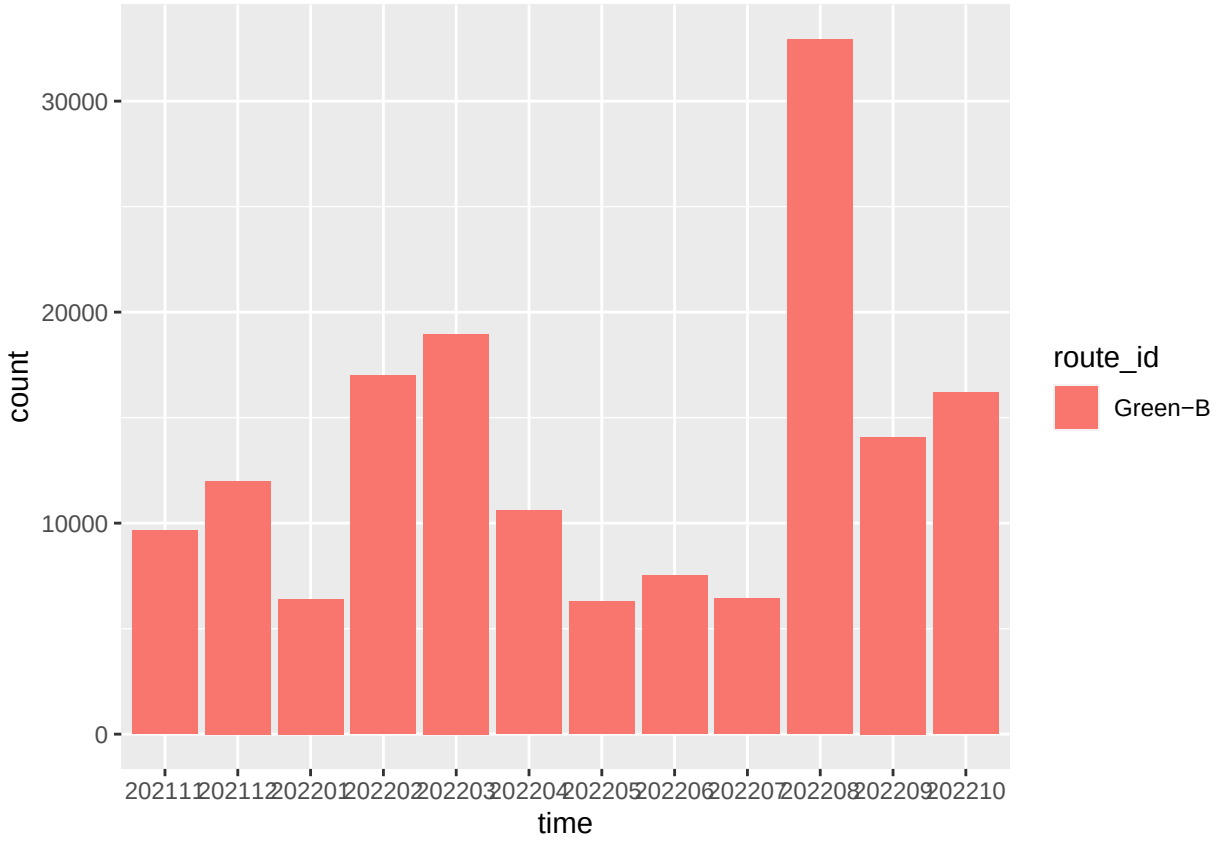

```
data %>%
 filter(!(route_desc %in% "")) %>%
 filter(!(route_pattern_time_desc %in% "") | is.na(route_pattern_time_desc)) %>%
 ggplot() +
  aes(x = route_desc, fill = direction) +
  geom_bar() +
  scale_fill_hue(direction = 1) +
  theme_minimal()
```




```
data %>% filter(route_id=="Green-B"|route_id=="Green-E"|route_id=="Green-C"|route_id=="Green-D") %>%
  ggplot() + aes(x=time,fill=route_id) + geom_bar()
```




```
data %>% filter(route_id=="Green-B") %>%
  ggplot() + aes(x=time,fill=route_id) + geom_bar()
```



In the first picture, we can know that key bus and local bus were the most used means of transportation last year, followed by rapid transit. It can be seen that key bus and local bus are the main pillars of Boston's transportation. It can be seen that Boston's bus system very developed. From the second picture, it can be seen that the key bus and local bus are both outbound and inbound directions, while the rapid transit travels in all directions, indicating that the destination of the key bus and local bus is very clear, but has a strong Purposeful, and rapid transit is the pillar of daily traffic. In terms of time, the subway data reached its peak in August of the past year. It may be that there are many universities in Boston. Busy, because this line passes through many student residential areas, so it is very intuitive to see the distribution of Boston's subway operations

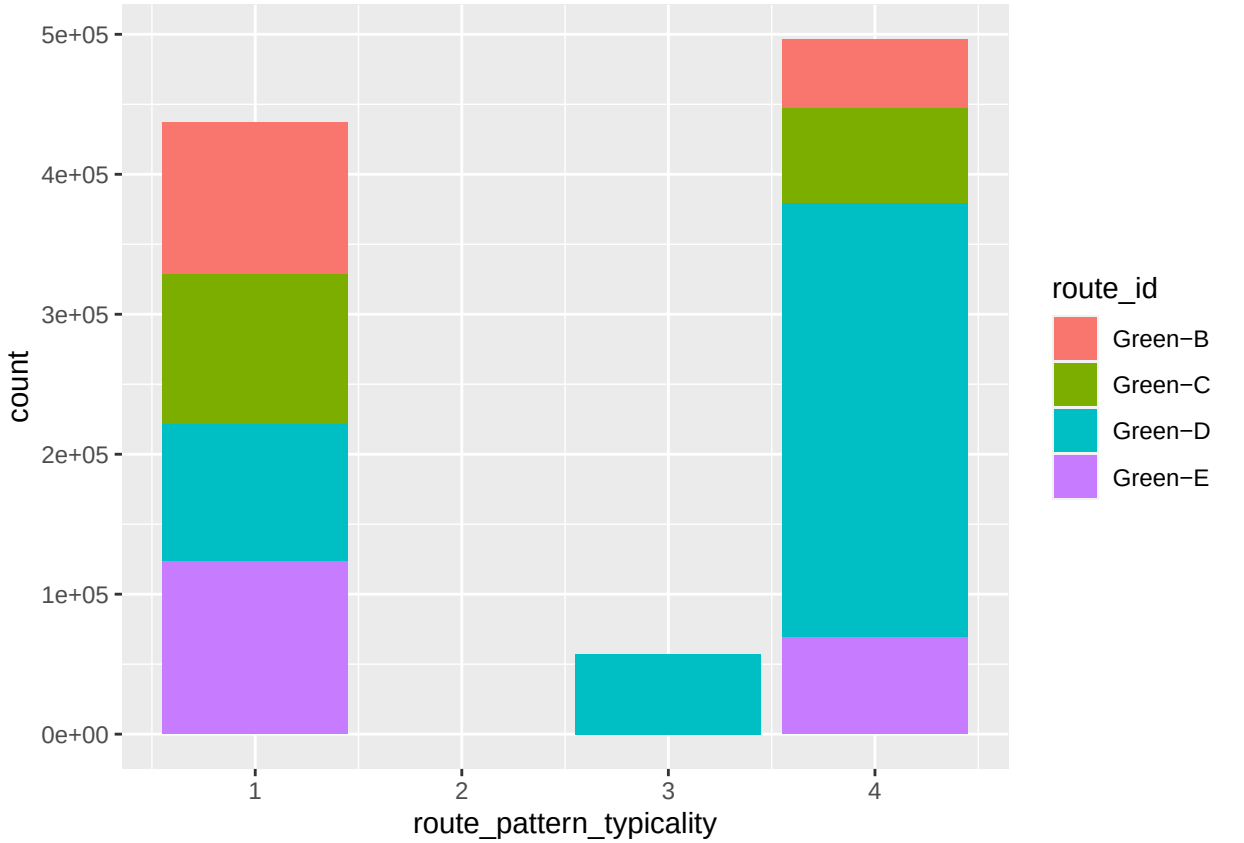
```
unique(data$bikes_allowed)
```

```
## [1]  0  2  1 NA
```

```
sum(!is.na(data$bikes_allowed))
```

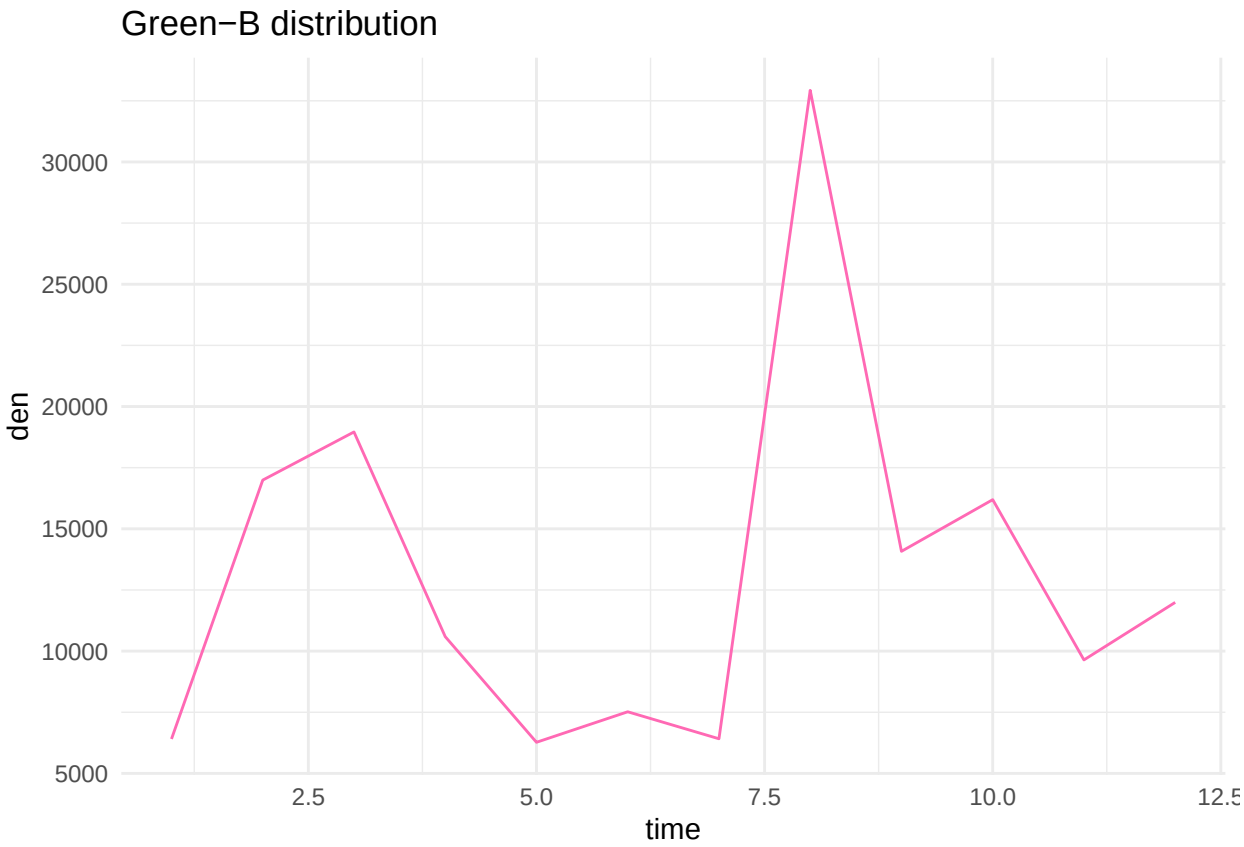
```
## [1] 8787204
```

```
test<-data %>% filter(route_id=="Green-B")

test2<-data %>% filter(route_id=="Green-B"|route_id=="Green-E"|route_id=="Green-C"|route_id=="Green-D")



ggplot(data=test2) + aes(x=route_pattern_typicality,fill=route_id) + geom_bar()
```



Looking at the distribution of the four major green lines in Boston from the route pattern, I found that the four major green lines do not have pattern 2, that is, they deviate from the conventional route, indicating that although the green lines are busy with transportation, they all perform their duties and have not been dispatched to other routes. Case. The most common mode is mode four, which is a deviation from normal service, which may be related to the disrepair or over-business of the Green Line, which often has problems with delays or breakdowns

```
ggplot(ggp2) +
  aes(x = time, y = den) +
  geom_line(colour = "#FF69B4") +
  theme_minimal()+
  ggtitle("Green-B distribution")
```



It can be seen from the above figure that the distribution of green-b lines is seasonal, which may be related to the opening and holiday time of universities in Boston

```
test3<- data %>% select(route_pattern_typicality,route_id) %>% filter(route_pattern_typicality=="1")
nrow(test3)
```

```
## [1] 3737934
```

```
test4<- data %>% select(route_pattern_typicality,route_id) %>% filter(route_pattern_typicality=="2")
nrow(test4)
```

```
## [1] 1095448
```

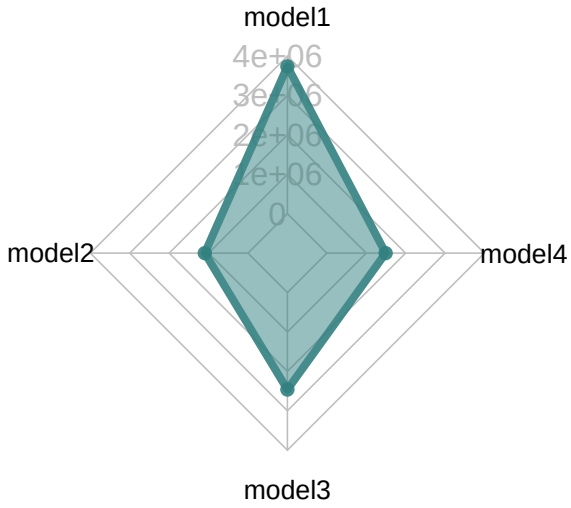
```
test5<- data %>% select(route_pattern_typicality,route_id) %>% filter(route_pattern_typicality=="3")
nrow(test5)
```

```
## [1] 2461150
```

```
test6<- data %>% select(route_pattern_typicality,route_id) %>% filter(route_pattern_typicality=="4")
nrow(test6)
```

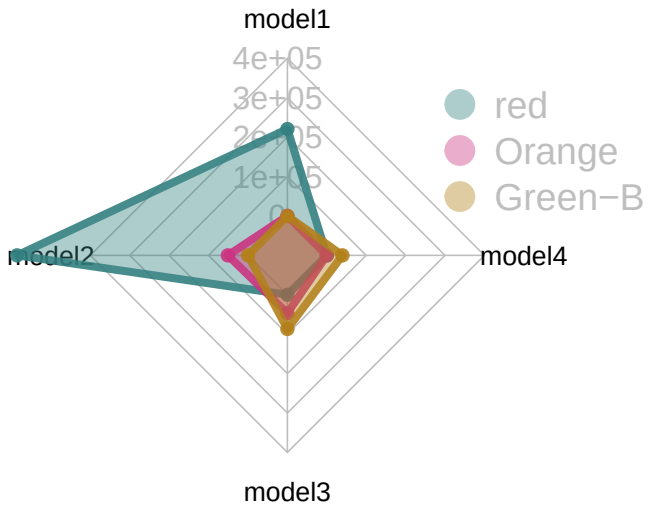
```
## [1] 1492672
```

```
test7<-as.data.frame(cbind(nrow(test3),nrow(test4),nrow(test5),nrow(test6)))
colnames(test7)<-c("model1","model2","model3","model4")
test7<-rbind(rep(4000000,4),rep(0,4),test7)

radarchart( test7  , axistype=1 ,

    #custom polygon
    pcol=rgb(0.2,0.5,0.5,0.9) , pfcol=rgb(0.2,0.5,0.5,0.5) , plwd=4 ,

    #custom the grid
    cglcol="grey", cglty=1, axislabcol="grey", caxislabels=seq(0,4000000,1000000), cglwd=0.8,

    #custom labels
    vlcex=0.8
    )
```




```
radartest1<- data %>% filter(route_id=="Red"&route_pattern_typicality=="1")
radartest2<- data %>% filter(route_id=="Red"&route_pattern_typicality=="2")
radartest3<- data %>% filter(route_id=="Red"&route_pattern_typicality=="3")
radartest4<- data %>% filter(route_id=="Red"&route_pattern_typicality=="4")

radartest5<- data %>% filter(route_id=="Orange"&route_pattern_typicality=="1")
radartest6<- data %>% filter(route_id=="Orange"&route_pattern_typicality=="2")
radartest7<- data %>% filter(route_id=="Orange"&route_pattern_typicality=="3")
radartest8<- data %>% filter(route_id=="Orange"&route_pattern_typicality=="4")
```

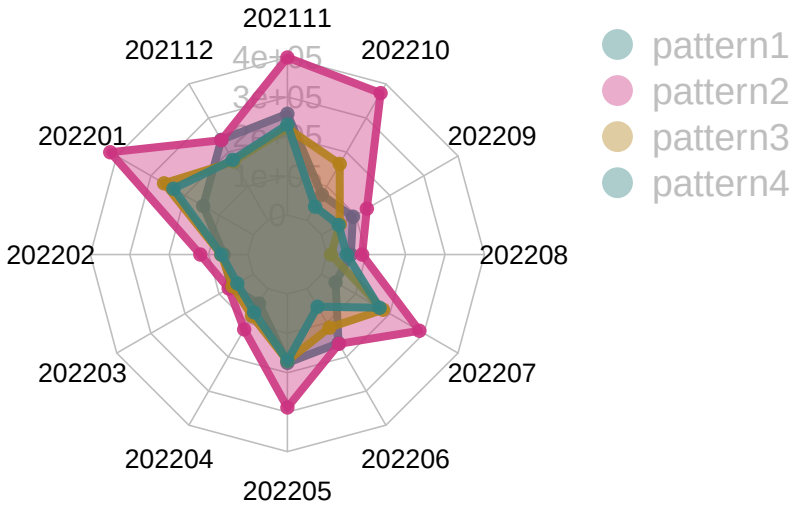
```
radartest9<- data %>% filter(route_id=="Green-B"&route_pattern_typicality=="1")
radartest10<- data %>% filter(route_id=="Green-B"&route_pattern_typicality=="2")
radartest11<- data %>% filter(route_id=="Green-B"&route_pattern_typicality=="3")
radartest12<- data %>% filter(route_id=="Green-B"&route_pattern_typicality=="4")

test8<-as.data.frame(matrix(c(nrow(radartest1),nrow(radartest2),nrow(radartest3),nrow(radartest4),nrow(
colnames(test8)<-c("model1","model2","model3","model4")
rownames(test8)<-c("red","Orange","Green-B")
test8<-rbind(rep(500000,4),rep(0,4),test8)

colors_border=c( rgb(0.2,0.5,0.5,0.9), rgb(0.8,0.2,0.5,0.9) , rgb(0.7,0.5,0.1,0.9) )
colors_in=c( rgb(0.2,0.5,0.5,0.4), rgb(0.8,0.2,0.5,0.4) , rgb(0.7,0.5,0.1,0.4) )

# plot with default options:
radarchart( test8  , axistype=1 ,
    #custom polygon
    pcol=colors_border , pfcol=colors_in , plwd=4 , plty=1,
    #custom the grid
    cglcol="grey", cglty=1, axislabcol="grey", caxislabels=seq(0,500000,100000), cglwd=0.8,
    #custom labels
    vlcex=0.8
    )
legend(x=0.7, y=1, legend = rownames(test8[-c(1,2),]), bty = "n", pch=20 , col=colors_in , text.col = "
```



It can be seen from the above radar chart that the operating mode of the red line is biased toward mode

1 and mode 2, and the end-of-period mode is more typical. But the green-b and orange lines have a more normal operating mode bias

```
test10<-as.data.frame(matrix(c(nrow(radar1),nrow(radar2),nrow(radar3),nrow(radar4),nrow(radar5),nrow(ra
colnames(test10)<-c("202111","202112","202201","202202","202203","202204","202205","202206","202207","2
rownames(test10)<-c("pattern1","pattern2","pattern3","pattern4")
test10<-rbind(rep(500000,12),rep(0,4),test10)

colors_border=c( rgb(0.2,0.5,0.5,0.9), rgb(0.8,0.2,0.5,0.9) , rgb(0.7,0.5,0.1,0.9) )
colors_in=c( rgb(0.2,0.5,0.5,0.4), rgb(0.8,0.2,0.5,0.4) , rgb(0.7,0.5,0.1,0.4) )

# plot with default options:
radarchart( test10  , axistype=1 ,
    #custom polygon
    pcol=colors_border , pfcol=colors_in , plwd=4 , plty=1,
    #custom the grid
    cglcol="grey", cglty=1, axislabcol="grey", caxislabels=seq(0,500000,100000), cglwd=0.8,
    #custom labels
    vlcex=0.8
    )
legend(x=1.5, y=1.3, legend = rownames(test10[-c(1,2),]), bty = "n", pch=20 , col=colors_in , text.col =
```



Judging from the above time and the radar map of the route pattern, mode 2 deviates more from the normal route, indicating that Boston's subway distribution has high flexibility and adjustability, and it is a normal bus system