Case\_study1

Bthyw

9/2/2021

### Combine into a single data.frame

Divvy\_Trips\_2020\_Q1<-Divvy\_Trips\_2020\_Q1 %>%   
 rename(trip\_id= ride\_id,bikeid= rideable\_type,start\_time= started\_at,end\_time= ended\_at,from\_station\_id= start\_station\_id,from\_station\_name= start\_station\_name,to\_station\_id=end\_station\_id,to\_station\_name= end\_station\_name, usertype=member\_casual)  
Divvy\_Trips\_2020\_Q1<-Divvy\_Trips\_2020\_Q1 %>%   
 select(-c(start\_lat,start\_lng,end\_lat,end\_lng))  
  
Divvy\_Trips\_2019<- rbind(Divvy\_Trips\_2019\_Q3,Divvy\_Trips\_2019\_Q4)  
Divvy\_Trips\_2019<-Divvy\_Trips\_2019 %>%   
 select(-c(gender,birthyear,tripduration))  
all\_trips<-rbind(Divvy\_Trips\_2019,Divvy\_Trips\_2020\_Q1)

all\_trips$from\_station\_id<-as.character(all\_trips$from\_station\_id)  
all\_trips$to\_station\_id<-as.character(all\_trips$to\_station\_id)  
  
all\_trips$usertype<-all\_trips$usertype %>%   
 recode("casual"="Customer","member"="Subscriber")

### Time interval

all\_trips$date<-as.Date(all\_trips$start\_time)  
all\_trips$day<-format(as.Date(all\_trips$date),"%d")  
all\_trips$month<-format(as.Date(all\_trips$date),"%m")  
all\_trips$year<-format(as.Date(all\_trips$date),"%Y")  
  
all\_trips$day\_of\_week<-format(as.Date(all\_trips$date),"%A")  
  
all\_trips$ride\_length<-difftime(all\_trips$end\_time,all\_trips$start\_time)  
all\_trips$ride\_length<-as.numeric(as.character(all\_trips$ride\_length))

### Oversee data trends by using table

all\_trips$day\_of\_week<-ordered(all\_trips$day\_of\_week,levels=c("Sunday","Monday","Tuesday","Wednesday","Thursday","Friday","Saturday"))  
#all\_trips$year<-ordered(all\_trips$year,levels=c("2019","2020"))  
aggregate(all\_trips$ride\_length~all\_trips$usertype+all\_trips$day\_of\_week,FUN = mean)

## all\_trips$usertype all\_trips$day\_of\_week all\_trips$ride\_length  
## 1 Customer Sunday 3871.6677  
## 2 Subscriber Sunday 910.7714  
## 3 Customer Monday 3577.4999  
## 4 Subscriber Monday 843.2026  
## 5 Customer Tuesday 3832.1473  
## 6 Subscriber Tuesday 837.5502  
## 7 Customer Wednesday 3964.5198  
## 8 Subscriber Wednesday 822.4797  
## 9 Customer Thursday 3895.5877  
## 10 Subscriber Thursday 824.2099  
## 11 Customer Friday 4144.4606  
## 12 Subscriber Friday 823.7175  
## 13 Customer Saturday 3460.4535  
## 14 Subscriber Saturday 991.1996

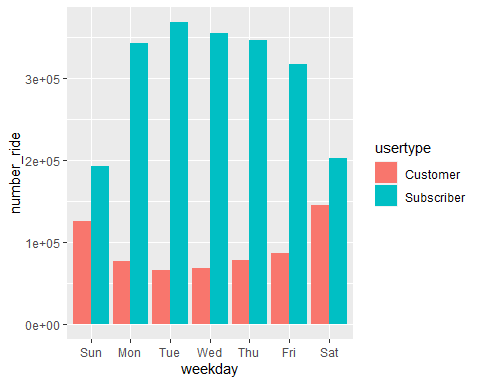
### Analyze ridership data by type and weekday

all\_trips<-all\_trips %>%   
 mutate(weekday=wday(start\_time,label = TRUE)) %>%   
 group\_by(weekday,usertype) %>% #then cal the number of rides&average time duration  
 summarise(number\_ride=n(),average\_duration=mean(ride\_length)) %>%   
 arrange(usertype,weekday)

## `summarise()` has grouped output by 'weekday'. You can override using the `.groups` argument.

### Visualize the number of rides by rider type

ggplot(data = all\_trips,mapping = aes(x=weekday,y=number\_ride,fill=usertype))+  
 geom\_col(position = "dodge")

 ### Visualize the average duration

ggplot(data = all\_trips,mapping = aes(x=weekday,y=average\_duration,fill=usertype))+  
 geom\_col(position = "dodge")

