```r
install.packages("tidyverse")
install.packages("lubridate")
install.packages("ggplot")
install.packages("dplyr")

library(tidyverse)
library(lubridate)
library(ggplot2)
library(dplyr)
getwd()
setwd("E:/RECOVERY/Documents/DIVVY DATA/csv")
Feb_2022<-read.csv("t1.csv")
March_2022<-read.csv("t2.csv")
April_2022<-read.csv("t3.csv")
May_2022<-read.csv("t4.csv")
June_2022<-read.csv("t5.csv")
July_2022<-read.csv("t6.csv")
Aug_2022<-read.csv("t7.csv")
Sept_2022<-read.csv("t8.csv")
Oct_2022<-read.csv("t9.csv")
Nov_2022<-read.csv("t10.csv")
Dec_2022<-read.csv("t11.csv")
Jan_2023<-read.csv("t12.csv")

colnames(Feb_2022)
colnames(March_2022)
colnames(April_2022)
colnames(May_2022)
colnames(June_2022)
colnames(July_2022)
colnames(Aug_2022)
colnames(Sept_2022)
colnames(Oct_2022)
colnames(Nov_2022)
colnames(Dec_2022)
colnames(Jan_2023)

str(Feb_2022)
str(March_2022)
str(April_2022)
str(May_2022)
str(June_2022)
str(July_2022)
str(Aug_2022)
str(Sept_2022)
str(Oct_2022)
str(Nov_2022)
str(Dec_2022)
str(Jan_2023)


Feb_2022 <- Feb_2022 %>% select(-c(start_station_name, start_station_id, end_station_name,
end_station_id))
March_2022 <- March_2022 %>% select(-c(start_station_name, start_station_id,
end_station_name, end_station_id))
April_2022 <- April_2022 %>% select(-c(start_station_name, start_station_id,
end_station_name, end_station_id))
May_2022 <- May_2022 %>% select(-c(start_station_name, start_station_id, end_station_name,
end_station_id))
June_2022 <- June_2022 %>% select(-c(start_station_name, start_station_id,
end_station_name, end_station_id))
July_2022 <- July_2022 %>% select(-c(start_station_name, start_station_id,
end_station_name, end_station_id))
Aug_2022 <- Aug_2022 %>% select(-c(start_station_name, start_station_id, end_station_name,
end_station_id))
```

```r
Sept_2022 <- Sept_2022 %>% select(-c(start_station_name, start_station_id,
end_station_name, end_station_id))
Oct_2022 <- Oct_2022 %>% select(-c(start_station_name, start_station_id, end_station_name,
end_station_id))
Nov_2022 <- Nov_2022 %>% select(-c(start_station_name, start_station_id, end_station_name,
end_station_id))
Dec_2022 <- Dec_2022 %>% select(-c(start_station_name, start_station_id, end_station_name,
end_station_id))
Jan_2023 <- Jan_2023 %>% select(-c(start_station_name, start_station_id, end_station_name,
end_station_id))

str(Feb_2022)
str(March_2022)
str(April_2022)
str(May_2022)
str(June_2022)
str(July_2022)
str(Aug_2022)
str(Sept_2022)
str(Oct_2022)
str(Nov_2022)
str(Dec_2022)
str(Jan_2023)

all_trips<-bind_rows(Feb_2022, March_2022, April_2022, May_2022, June_2022, July_2022,
Aug_2022, Sept_2022, Oct_2022, Nov_2022, Dec_2022, Jan_2023)

str(all_trips)

all_trips$date <- as.Date(all_trips$started_at)
all_trips$month <- format(as.Date(all_trips$date), "%m")
all_trips$day <- format(as.Date(all_trips$date), "%d")
all_trips$year <- format(as.Date(all_trips$date), "%Y")
all_trips$day_of_week <- format(as.Date(all_trips$date), "%A")

all_trips$ride_length <- difftime(all_trips$ended_at, all_trips$started_at)

colnames(all_trips)

nrow(all_trips)
dim(all_trips)

head(all_trips)

str(all_trips)

summary(all_trips)

all_trips_v2 <- all_trips[!(all_trips$rideable_type == "docked_bike" |
all_trips$ride_length<0),]

str(all_trips_v2)

all_trips_v2 %>%
  group_by(member_casual) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length))

all_trips_v2 %>%
  group_by(member_casual, rideable_type) %>%
  summarise(number_of_rides = n())


mean(all_trips_v2$ride_length)
median(all_trips_v2$ride_length)
max(all_trips_v2$ride_length)
```

```
min(all_trips_v2$ride_length)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = mean)
aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = median)
aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = max)
aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = min)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual + all_trips_v2$day_of_week,
FUN = mean)

all_trips_v2$day_of_week <- ordered(all_trips_v2$day_of_week, levels=c("Sunday", "Monday",
"Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))

all_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%  #creates weekday field using wday()
  group_by(member_casual, weekday) %>%  #groups by use rtype and weekday
  summarise(number_of_rides = n()
,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)




++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+++++++++++++++++++++

## VISUALIZATION

# Creating a visualization for the number of rides by rider type for each day of the week
all_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge")

----# Creating a visualization for average duration of ride
all_trips_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(pl
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge")


counts <- aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual +
all_trips_v2$day_of_week, FUN = mean)

write.csv(counts,"E:\\RECOVERY\\Documents\\DIVVY DATA\\csv\\AVG.csv,row.names =FALSE")
```