cyclistic <- list.files(path='C:/Users/gravich1/Downloads/Cyclistic') %>% lapply(read\_csv) %>% bind\_rows

write.csv(cyclistic, "C:\\Users\\gravich1\\Downloads\\Cyclistic.csv", row.names=FALSE)

data = read.csv('C:/Users/gravich1/Downloads/Cyclistic/cleaned\_data.csv')

> library(tidyverse)

> library(lubridate)

> library(ggplot2)

> data <- data %>%  select(-c(start\_lat, start\_lng, end\_lat, end\_lng))

> mean(data$ride\_time)

[1] 19.50667

> median(data$ride\_time)

[1] 12

> max(data$ride\_time)

[1] 10723

> min(data$ride\_time)

[1] 1

> summary(data$ride\_time)

    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.

    1.00     7.00    12.00    19.51    22.00 10723.00

> aggregate(data$ride\_time ~ data$member\_casual, FUN = mean)

  data$member\_casual data$ride\_time

1             casual       26.67587

2             member       13.62515

> aggregate(data$ride\_time ~ data$member\_casual, FUN = median)

  data$member\_casual data$ride\_time

1             casual             16

2             member             10

> aggregate(data$ride\_time ~ data$member\_casual, FUN = max)

  data$member\_casual data$ride\_time

1             casual          10723

2             member           1476

> aggregate(data$ride\_time ~ data$member\_casual, FUN = min)

  data$member\_casual data$ride\_time

1             casual              1

2             member              1

> data$new\_date <- strptime(as.character(data$start\_date), "%m/%d/%Y")

> data$day\_of\_week <- format(as.Date(data$new\_date), "%A")

> data <- data %>% rename("start\_day" = "day\_of\_week")

> data$end\_date\_nf <- strptime(as.character(data$end\_date), "%m/%d/%Y")

> data <- data %>% rename("start\_date\_nf" = "new\_date")

> data$end\_day <- format(as.Date(data$end\_date\_nf), "%A")

> data$start\_day\_od <- ordered(data$start\_day, levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))

> data$end\_day\_od <- ordered(data$end\_day, levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))

> aggregate(data$ride\_time ~ data$member\_casual + data$start\_day\_od, FUN = mean)

   data$member\_casual data$start\_day\_od data$ride\_time

1              casual            Sunday       29.81727

2              member            Sunday       15.14839

3              casual            Monday       28.69409

4              member            Monday       13.33049

5              casual           Tuesday       24.51160

6              member           Tuesday       12.97954

7              casual         Wednesday       22.43897

8              member         Wednesday       12.65968

9              casual          Thursday       24.36905

10             member          Thursday       13.29698

11             casual            Friday       24.49556

12             member            Friday       13.22009

13             casual          Saturday       28.63469

14             member          Saturday       15.18085

> data %>%

+   mutate(weekday = wday(start\_date\_nf, label = TRUE)) %>%

+   group\_by(member\_casual, weekday) %>%  summarise(number\_of\_rides = n(), average\_duration = mean(ride\_time)) %>% arrange(member\_casual, weekday)

`summarise()` has grouped output by 'member\_casual'. You can override using the `.groups` argument.

# A tibble: 14 × 4

# Groups:   member\_casual [2]

   member\_casual weekday number\_of\_rides average\_duration

   <chr>         <ord>             <int>            <dbl>

 1 casual        Sun               75489             29.8

 2 casual        Mon               54176             28.7

 3 casual        Tue               44253             24.5

 4 casual        Wed               39502             22.4

 5 casual        Thu               50977             24.4

 6 casual        Fri               49867             24.5

 7 casual        Sat               72993             28.6

 8 member        Sun               61346             15.1

 9 member        Mon               72169             13.3

10 member        Tue               74341             13.0

11 member        Wed               68888             12.7

12 member        Thu               75648             13.3

13 member        Fri               60685             13.2

14 member        Sat               58965             15.2

We can infer that while casual people's duration is longer consistently across the week, they outperform members only during the weekends.

> data %>%

+   mutate(weekday = wday(start\_date\_nf, label = TRUE)) %>%

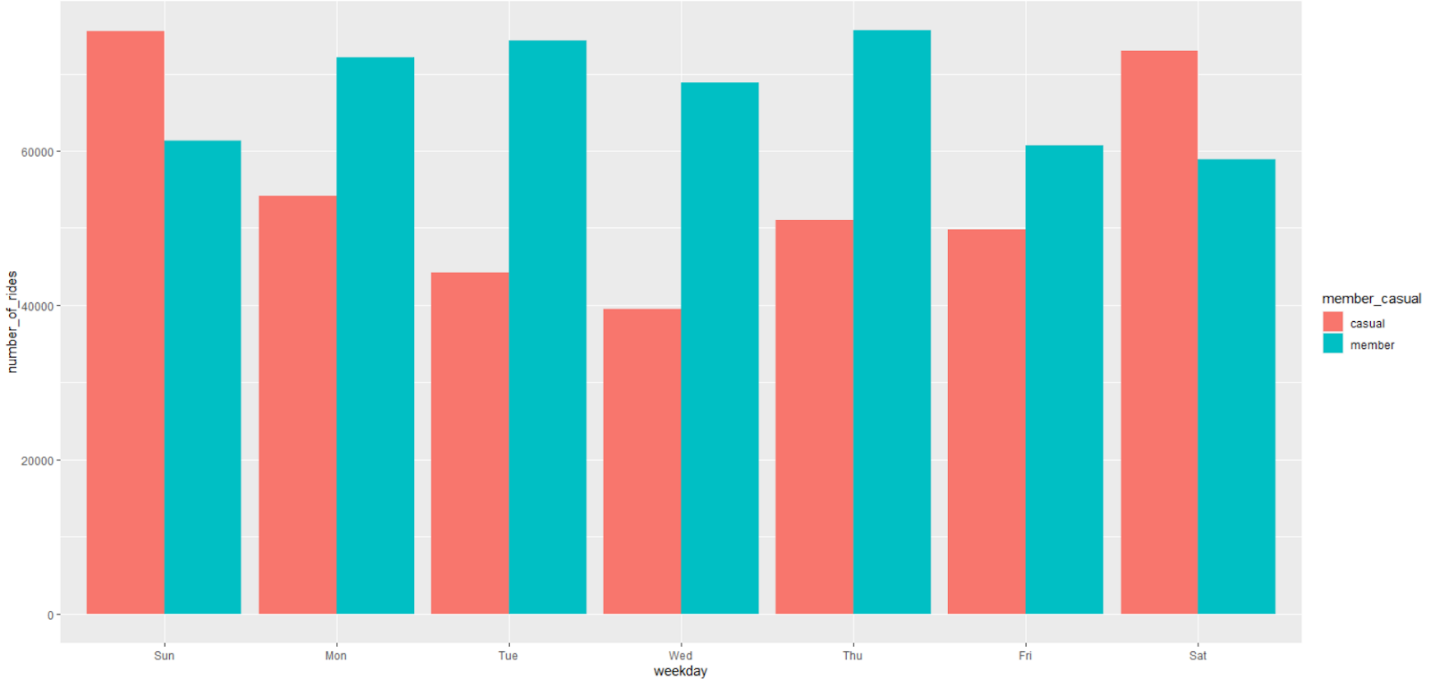
+   group\_by(member\_casual, weekday) %>%

+   summarise(number\_of\_rides = n(), average\_duration = mean(ride\_time)) %>%

+   arrange(member\_casual, weekday)  %>%

+   ggplot(aes(x = weekday, y = number\_of\_rides, fill = member\_casual)) +

+   geom\_col(position = "dodge")



> data%>%

+   mutate(weekday = wday(start\_date\_nf, label = TRUE)) %>%

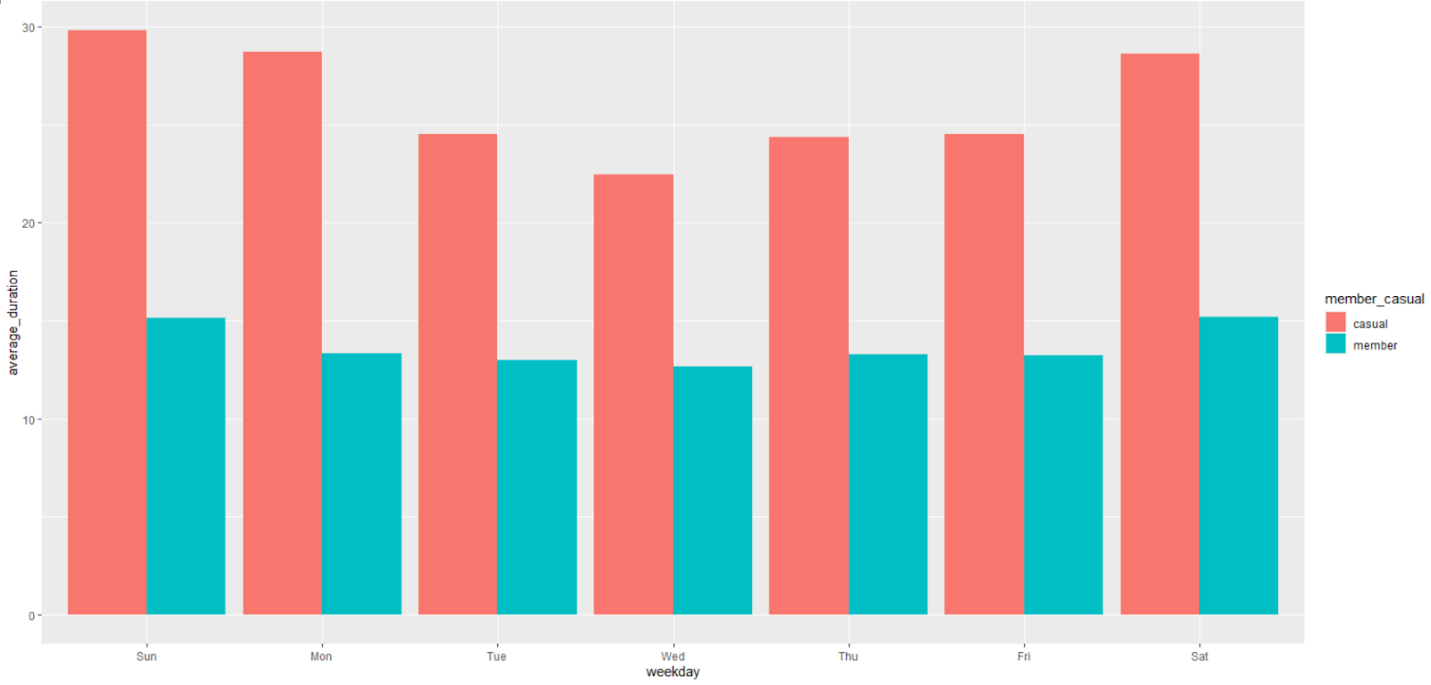
+   group\_by(member\_casual, weekday) %>%

+   summarise(number\_of\_rides = n(), average\_duration = mean(ride\_time)) %>%

+   arrange(member\_casual, weekday)  %>%

+   ggplot(aes(x = weekday, y = average\_duration, fill = member\_casual)) +

+   geom\_col(position = "dodge")



> counts <- aggregate(data$ride\_time ~ data$member\_casual + data$start\_day, FUN = mean)

> write.csv(counts, file = 'C:/Users/gravich1/Downloads/Cyclistic/avg\_ride\_length.csv')