

Cyclistic Data Analysis - Analysis

Gayatri Paul

2023-01-06

4. Analyze

Statistics

```
average_ride_length = round(mean(df$ride_length), 2)
print(paste0("Average ride_length = ", average_ride_length))
```

```
## [1] "Average ride_length = 19.41"
```

```
max_ride_length = round(max(df$ride_length), 2)
print(paste0("Max. ride_length = ", max_ride_length))
```

```
## [1] "Max. ride_length = 41387.25"
```

```
tb_no_of_days <- df %>%
  group_by(rider_type) %>%
  summarise(min_no_of_days=min(no_of_days),max_no_of_days=max(no_of_days),avg_no_of_days=round(mean(no_of_days),digits=2))
tb_no_of_days
```

```
## # A tibble: 2 × 4
##   rider_type min_no_of_days max_no_of_days avg_no_of_days
##   <chr>          <dbl>          <dbl>          <dbl>
## 1 casual             1             30             1.01
## 2 member             1              3              1
```

```
rides_by_day_of_week <- df %>%
  count(day_of_week)

mode_of_day_of_week <- rides_by_day_of_week %>%
  filter(n == max(n)) %>%
  select(day_of_week)

print(paste0("Mode of the day of week = ", mode_of_day_of_week))
```

```
## [1] "Mode of the day of week = saturday"
```

```
df_casual <- df %>%
  filter(rider_type == 'casual')

rides_by_day_of_week_casual <- df_casual%>%
  count(day_of_week)

mode_of_day_of_week_casual <- rides_by_day_of_week_casual %>%
  filter(n == max(n)) %>%
  select(day_of_week)

print(paste0("Mode of the day of week for casual rider = ", mode_of_day_of_week_casual))
```

```
## [1] "Mode of the day of week for casual rider = saturday"
```

```
df_member <- df %>%
  filter(rider_type == 'member')

rides_by_day_of_week_member <- df_member%>%
  count(day_of_week)

mode_of_day_of_week_member <- rides_by_day_of_week_member %>%
  filter(n == max(n)) %>%
  select(day_of_week)

print(paste0("Mode of the day of week for member = ", mode_of_day_of_week_member))
```

```
## [1] "Mode of the day of week for member = wednesday"
```

Money spent by casual riders per ride

```
keep <- c("ride_id", "rideable_type", "rider_type", "ride_length")
df_casual <- df_casual[keep]

df_casual_classic <- df_casual %>%
  filter(rideable_type=="classic_bike" & ride_length>180)
nrow(df_casual_classic)
```

```
## [1] 8595
```

There are significant number of rides which exceeded the time limit.

We do not know whether the casual riders had day pass or single ride pass We don't know if each ride in day pass has been given new ride ID , lets assume considering minimum extra charges: classic bike if ride length ≤ 30 - single ride pass + no extra charge if ride length ≤ 60 single ride pass + extra charge if ride length $60 < x < 180$ - day pass + no extra charge if ride length > 180 - day pass + extra charges electric bike No time limit as bike is charged per minute docked bike pricing same as classic but without extra charge

As we are making assumptions, the value won't be exact, but we can get a general idea.

```
casual_rider_charges <- df_casual %>%
  mutate(ride_length = ceiling(ride_length)) %>%
  mutate(charges=
    ifelse(rideable_type=="electric_bike", 1+0.39*(ride_length),
      ifelse(rideable_type=="docked_bike", 3.30,
        ifelse(ride_length <= 30, 3.30,
          ifelse(ride_length <= 60, 3.30+0.15*(ride_length-30),
            ifelse(ride_length <= 180, 15,
              15+(ride_length-180)*0.16))))))
```

```
avg_charge_per_ride_casual_rider = round(sum(casual_rider_charges$charges)/nrow(casual_rider_charges), digits=2)
print(paste0("avg_charge_per_ride_casual_rider = ", avg_charge_per_ride_casual_rider, "$"))
```

```
## [1] "avg_charge_per_ride_casual_rider = 6.38 $"
```

```
max_charge_per_ride_casual_rider = max(casual_rider_charges$charges)
print(paste0("max_charge_per_ride_casual_rider = ", max_charge_per_ride_casual_rider, "$"))
```

```
## [1] "max_charge_per_ride_casual_rider = 235.8 $"
```

Casual riders who spent more than annual membership charges

```
Target <- casual_rider_charges %>%
  filter(charges > 108) %>%
  nrow()
```

```
Target
```

```
## [1] 4105
```

At least 4105 Casual riders paid more than Annual membership for a single ride. These people will be easy to convert into members. More data will be needed for further analysis.

Casual riders who spent more than annual membership charges

```
Target1 <- casual_rider_charges %>%
  filter(charges > 100) %>%
  nrow()
```

```
Target1
```

```
## [1] 4296
```

Members vs Casual Riders at a glance

Number of rides

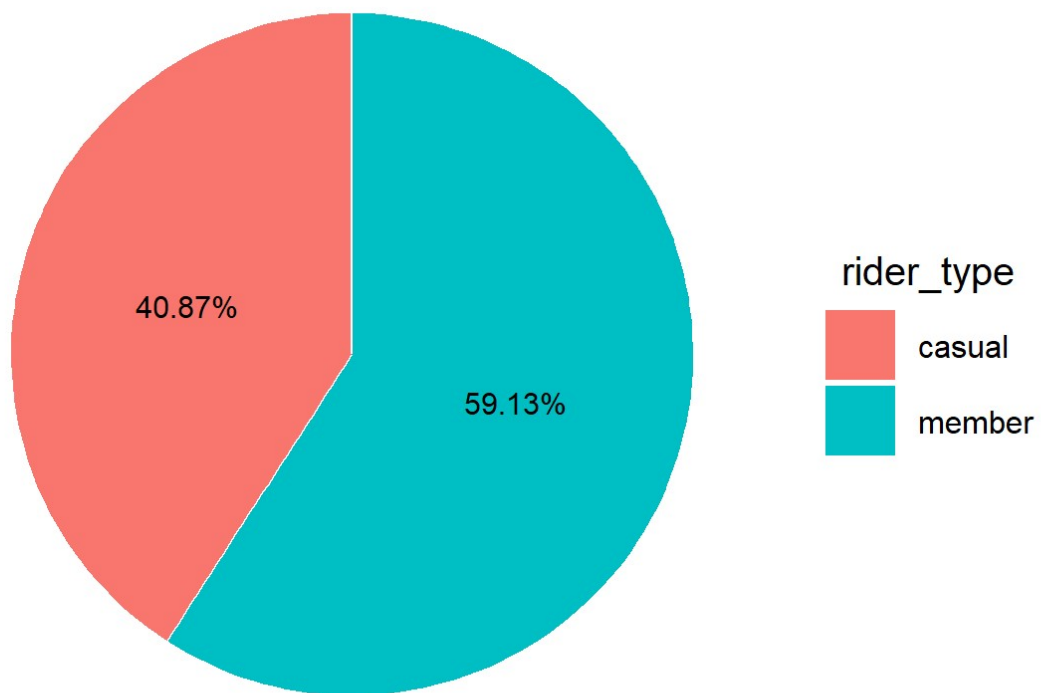
```
riders_count <- df %>%  
  group_by(rider_type) %>%  
  summarize(count = n())  
  
total_rides <- nrow(df)  
  
riders_count$percentage <- round((riders_count$count/ total_rides)*100, digits = 2)
```

```
riders_count_freq_tbl <- riders_count%>%  
  bind_rows(summarise(., across(where(is.numeric), sum),  
                        across(where(is.character), ~'Total')))
```

```
percentage <- round(riders_count$percentage, digits=2)  
rider_type <- riders_count$rider_type  
  
ypos = cumsum(percentage) - 0.5 * percentage  
ypos = 100 - ypos  
  
pie1 = ggplot() + theme_bw() +  
  geom_bar(aes(x = "", y = percentage, fill = rider_type),  
           stat = "identity", color = "white") +  
  coord_polar("y", start = 0) +  
  ggtitle("Riders Count") +  
  theme(plot.title = element_text(hjust = 0.5, size = 20),  
        axis.title = element_blank(),  
        axis.text = element_blank(),  
        axis.ticks = element_blank(),  
        panel.grid=element_blank(),  
        panel.border = element_blank()) +  
  theme(legend.text=element_text(size=12),  
        legend.title = element_text(hjust = 0.5, size=15),  
        legend.key.size = unit(1,"cm")) +  
  geom_text(aes(x = "", y = ypos, label = paste0(percentage, "%")), size = 4)
```

```
pie1
```

Riders Count



```
riders_count_freq_tbl
```

```
## # A tibble: 3 × 3
##   rider_type  count percentage
##   <chr>      <int>      <dbl>
## 1 casual    2609952      40.9
## 2 member    3776332      59.1
## 3 Total     6386284     100
```

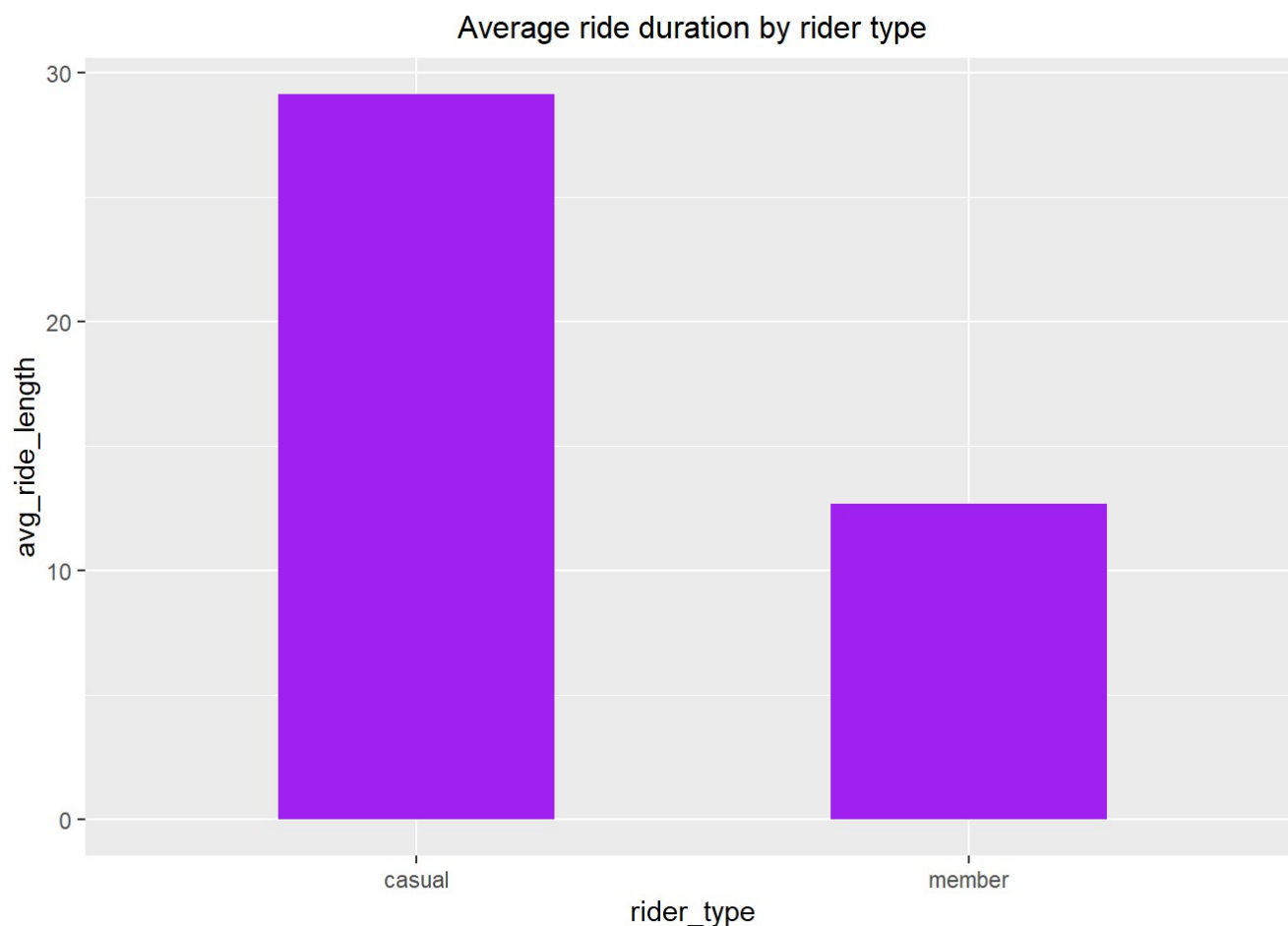
- Bike usage was observed to be more in Annual members compared to casual riders.

Average ride duration

```
avg_ride_length <- df %>%
  group_by(rider_type) %>%
  summarise(avg_ride_length = round(mean(ride_length), digits=2))
```

```
bar1 <- ggplot(data=avg_ride_length, aes(x=rider_type, y=avg_ride_length)) +
  geom_bar(stat="identity", width = 0.5, fill="purple")+
  ggtitle("Average ride duration by rider type")+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

```
bar1
```



avg_ride_length

```
## # A tibble: 2 × 2
##   rider_type avg_ride_length
##   <chr>      <dbl>
## 1 casual      29.1
## 2 member      12.7
```

- Average ride duration for members was 12.69 mins.
- Average ride duration were higher in Casual riders (29.13 mins).

This may indicate that :

- Members used bikes for their daily commute to work or University &
- Casual riders used the bikes for longer distances or leisure activities.

Trends based on day of the week

Number of rides

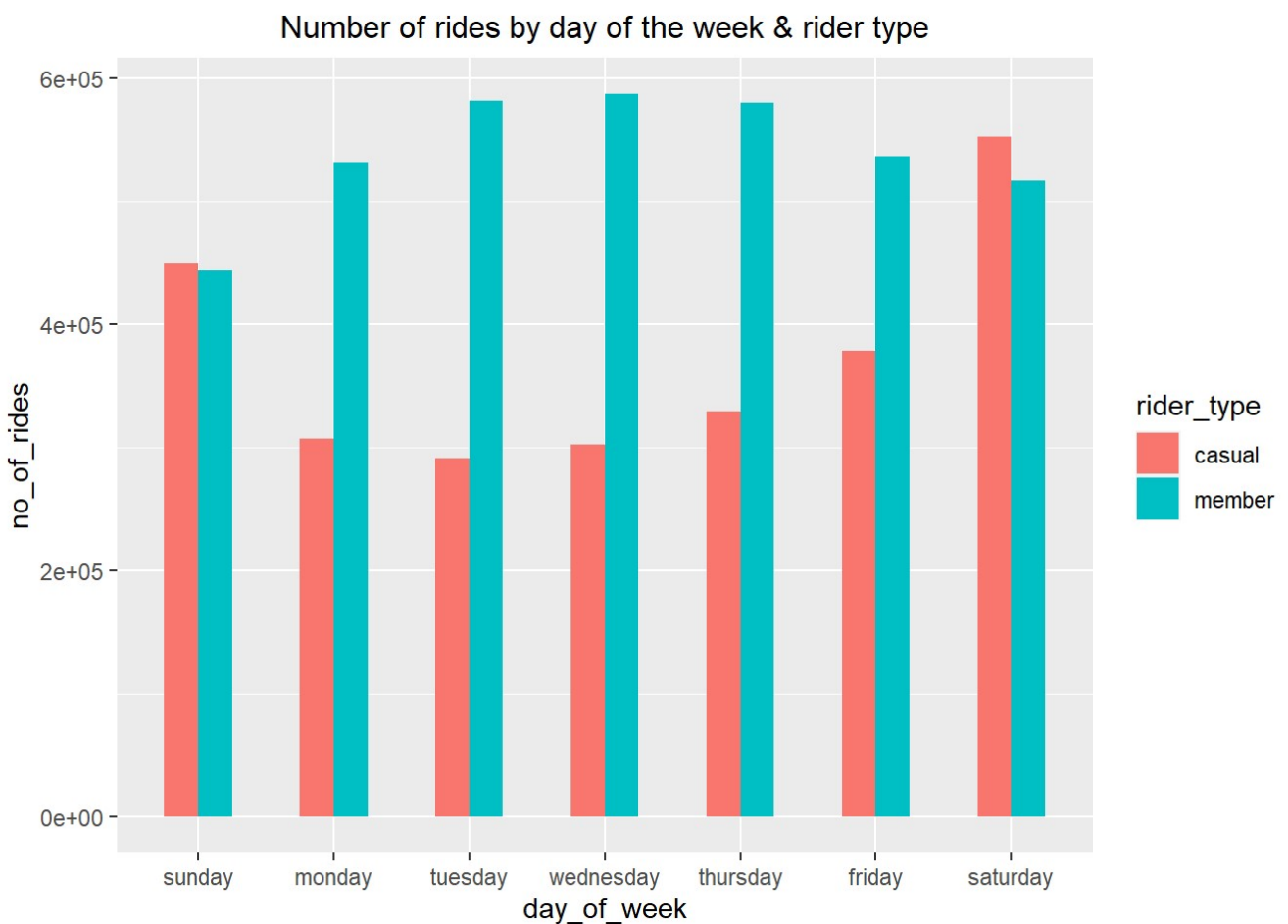
```
no_of_rides_by_day <- df %>%
  group_by(day_of_week, rider_type) %>%
  summarise(no_of_rides = n())
```

```
no_of_rides_by_day_tbl <- spread(no_of_rides_by_day, key= "rider_type", value= "no_of_rides")
```

```
positions <- c("sunday","monday","tuesday","wednesday","thursday","friday","saturday")
```

```
bar2 <- ggplot(data=no_of_rides_by_day, aes(x=day_of_week, y=no_of_rides, fill=rider_type)) +  
  geom_bar(stat="identity", width = 0.5, position = 'dodge')+  
  scale_x_discrete(limits = positions)+  
  ggtitle("Number of rides by day of the week & rider type")+  
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

bar2



no_of_rides_by_day_tbl

```
## # A tibble: 7 × 3
## # Groups:   day_of_week [7]
##   day_of_week casual member
##   <chr>         <int> <int>
## 1 friday        378440 536453
## 2 monday        306995 531311
## 3 saturday      551952 516322
## 4 sunday        449653 443931
## 5 thursday      329245 579743
## 6 tuesday       291437 581226
## 7 wednesday     302230 587346
```

- Number of rides for annual members were seen to be higher on weekdays. This helps us reaffirm our theory that members use the bikes for short distance daily commutes to and from work or University.
- Number of rides were highest on Saturdays and Sundays for casual riders which again confirms our previous assumption that the casual riders used bikes for leisure activities like exploring the city, going to movies, cafes or restaurants.
- Number of rides by casual riders were marginally higher than members on Saturdays and Sundays. On all other days, number of rides by members significantly higher than casual riders. This is also an indication that the significant proportion of casual riders may be tourists.

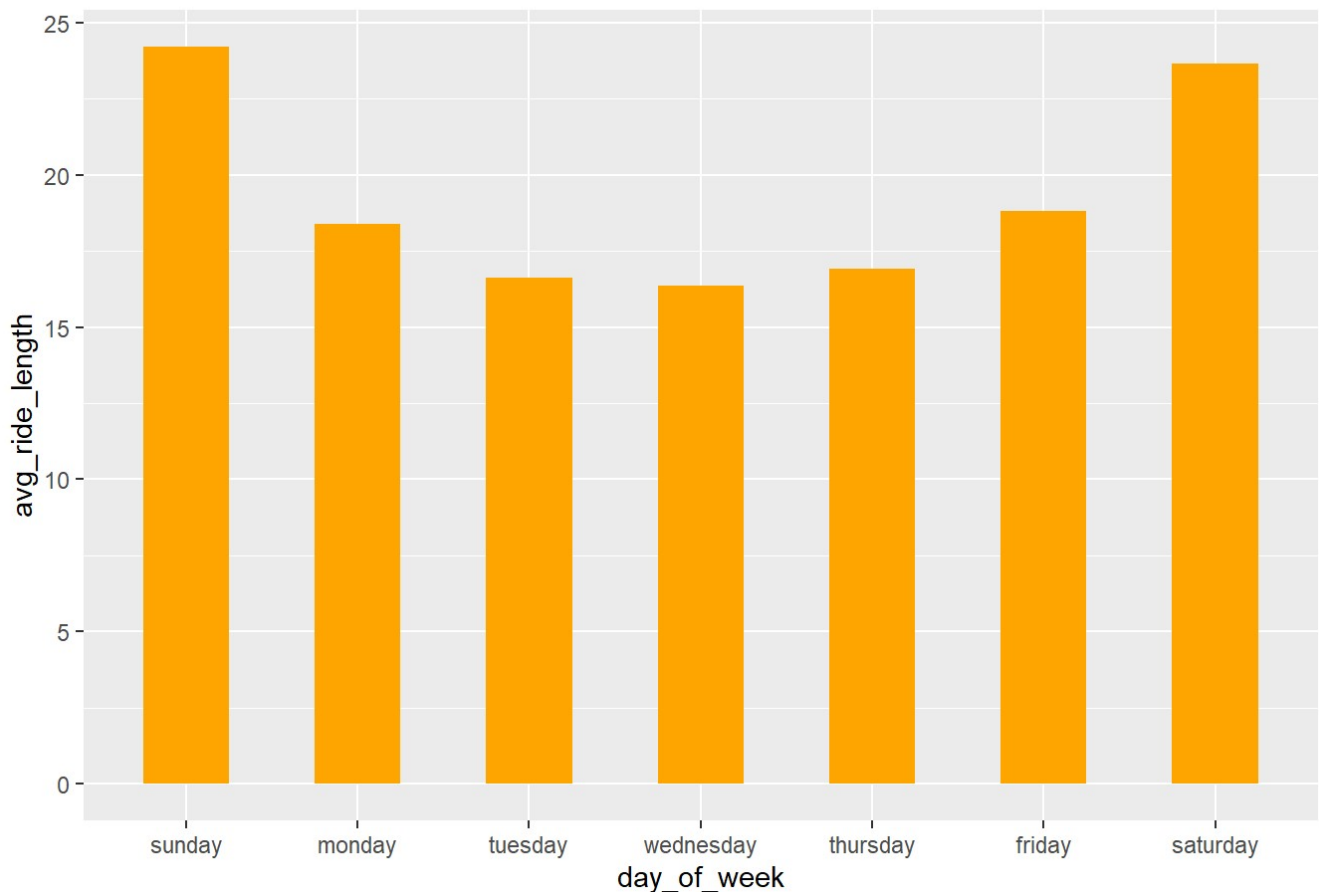
Average ride duration

```
avg_ride_length1 <- df %>%
  group_by(day_of_week) %>%
  summarise(avg_ride_length = round(mean(ride_length), digits=2))
```

```
positions <- c("sunday", "monday", "tuesday", "wednesday", "thursday", "friday", "saturday")
bar3 <- ggplot(data=avg_ride_length1, aes(x=day_of_week, y=avg_ride_length)) +
  geom_bar(stat="identity", width = 0.5, fill="orange")+
  scale_x_discrete(limits = positions)+
  ggtitle("Average ride duration by day of the week")+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

```
bar3
```


Average ride duration by day of the week



```
avg_ride_length1
```

```
## # A tibble: 7 × 2
##   day_of_week avg_ride_length
##   <chr>         <dbl>
## 1 friday          18.8
## 2 monday          18.4
## 3 saturday        23.6
## 4 sunday          24.2
## 5 thursday        16.9
## 6 tuesday         16.6
## 7 wednesday       16.4
```

- Average ride duration was highest on Sundays followed by Saturdays.
- The ride lengths remained in the similar range from Tuesdays to Thursdays but were slightly higher on Mondays and Fridays. This may be due to the long Weekends.

Average ride duration by day of the week & rider type

```
avg_ride_length_by_day <- df %>%
  group_by(day_of_week, rider_type) %>%
  summarise(avg_ride_length = round(mean(ride_length), digits=2))
```

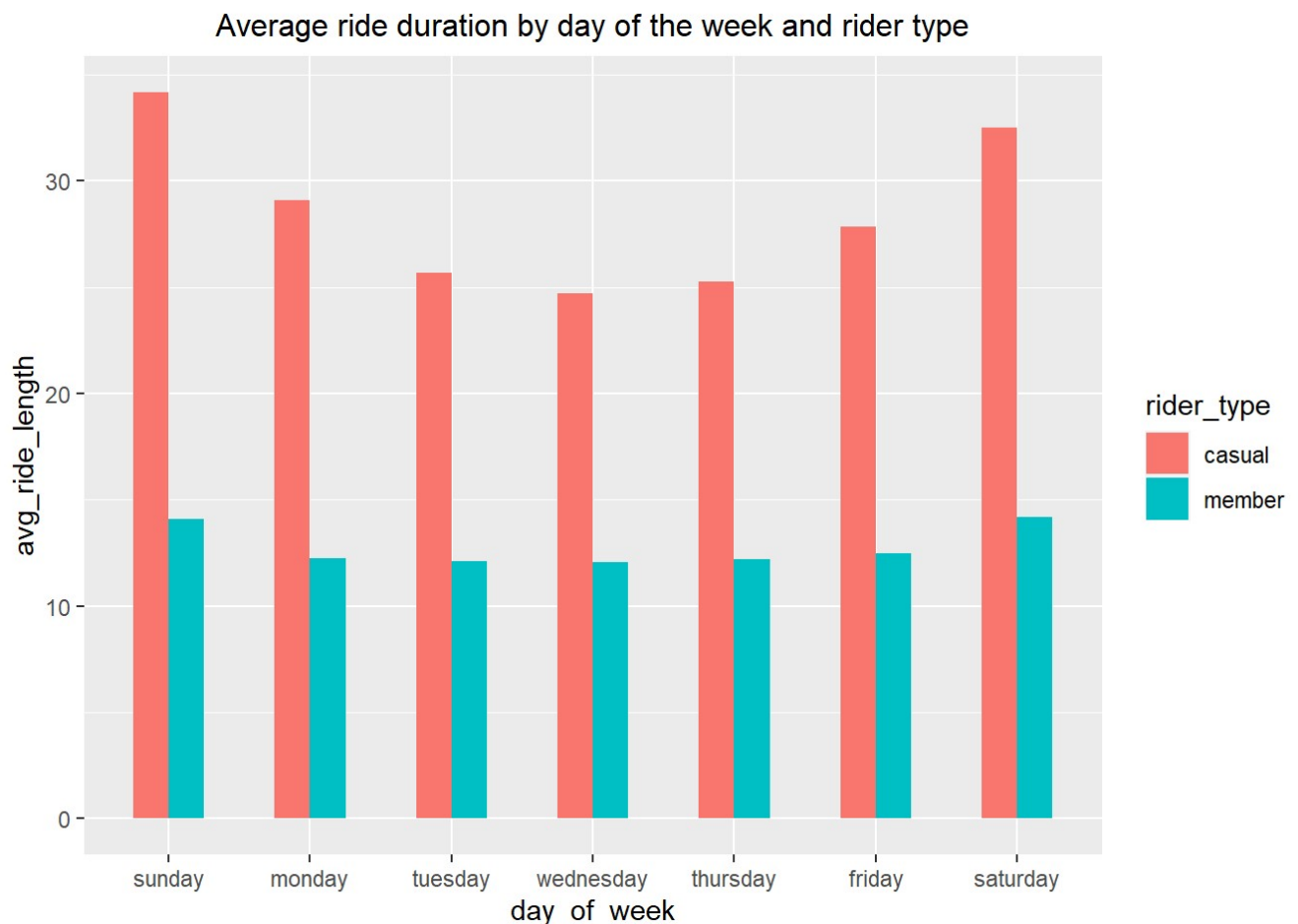
```
library(tidyr)
```

```
avg_ride_length_by_day_tbl <- spread(avg_ride_length_by_day, key= "rider_type", value= "avg_ride_length")
```

```
avg_ride_length_by_day_tbl <- avg_ride_length_by_day_tbl %>%  
  rename( "avg_ride_length_casual" = "casual",  
          "avg_ride_length_member" = "member")
```

```
bar4 <- ggplot(data=avg_ride_length_by_day, aes(x=day_of_week, y=avg_ride_length, fill=rider_type)) +  
  geom_bar(stat="identity", width = 0.5, position = 'dodge')+  
  scale_x_discrete(limits = positions)+  
  ggtitle("Average ride duration by day of the week and rider type")+  
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

bar4



avg_ride_length_by_day_tbl

```
## # A tibble: 7 × 3
## # Groups:   day_of_week [7]
##   day_of_week avg_ride_length_casual avg_ride_length_member
##   <chr>                <dbl>                <dbl>
## 1 friday                27.8                12.5
## 2 monday                29.1                12.2
## 3 saturday              32.5                14.2
## 4 sunday                34.2                14.1
## 5 thursday              25.3                12.2
## 6 tuesday               25.7                12.1
## 7 wednesday             24.7                12.0
```

- Average ride duration for casual riders (25 mins or more) was consistently higher than annual members (less than 15 mins) regardless of the day of the week.
- It was seen to be higher on weekend compared to weekdays in Casual riders as well as members.

Bike-type wise trend

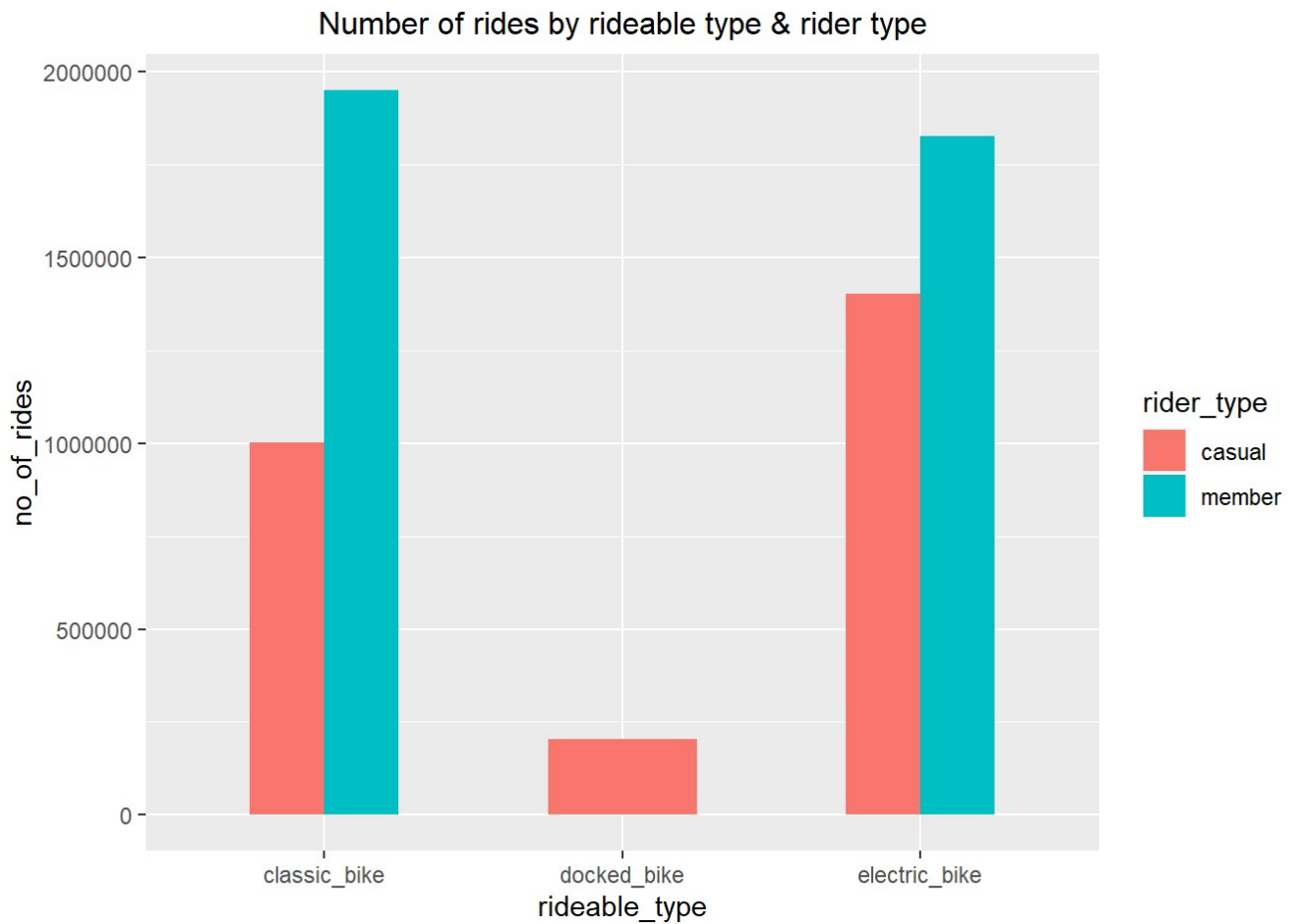
Number of rides by rider type

```
bike_type <- df %>%
  group_by(rideable_type, rider_type) %>%
  summarise(no_of_rides = n())
```

```
bike_type_tbl <- spread(bike_type, key= "rider_type", value= "no_of_rides")
```

```
bar5 <- ggplot(data=bike_type, aes(x=rideable_type, y=no_of_rides, fill=rider_type)) +
  geom_bar(stat="identity", width = 0.5, position = 'dodge')+
  ggtitle("Number of rides by rideable type & rider type")+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

```
bar5
```



```
bike_type_tbl
```

```
## # A tibble: 3 × 3
## # Groups:   rideable_type [3]
##   rideable_type casual member
##   <chr>          <int> <int>
## 1 classic_bike 1002953 1950970
## 2 docked_bike  205078    NA
## 3 electric_bike 1401921 1825362
```

- Classic bike was the bike of preference for members, closely followed by electric bike. Members did not use docked bike even once throughout the year.
- Casual riders preferred electric bikes over classic bikes. Casual riders did opt for docked bikes occasionally.

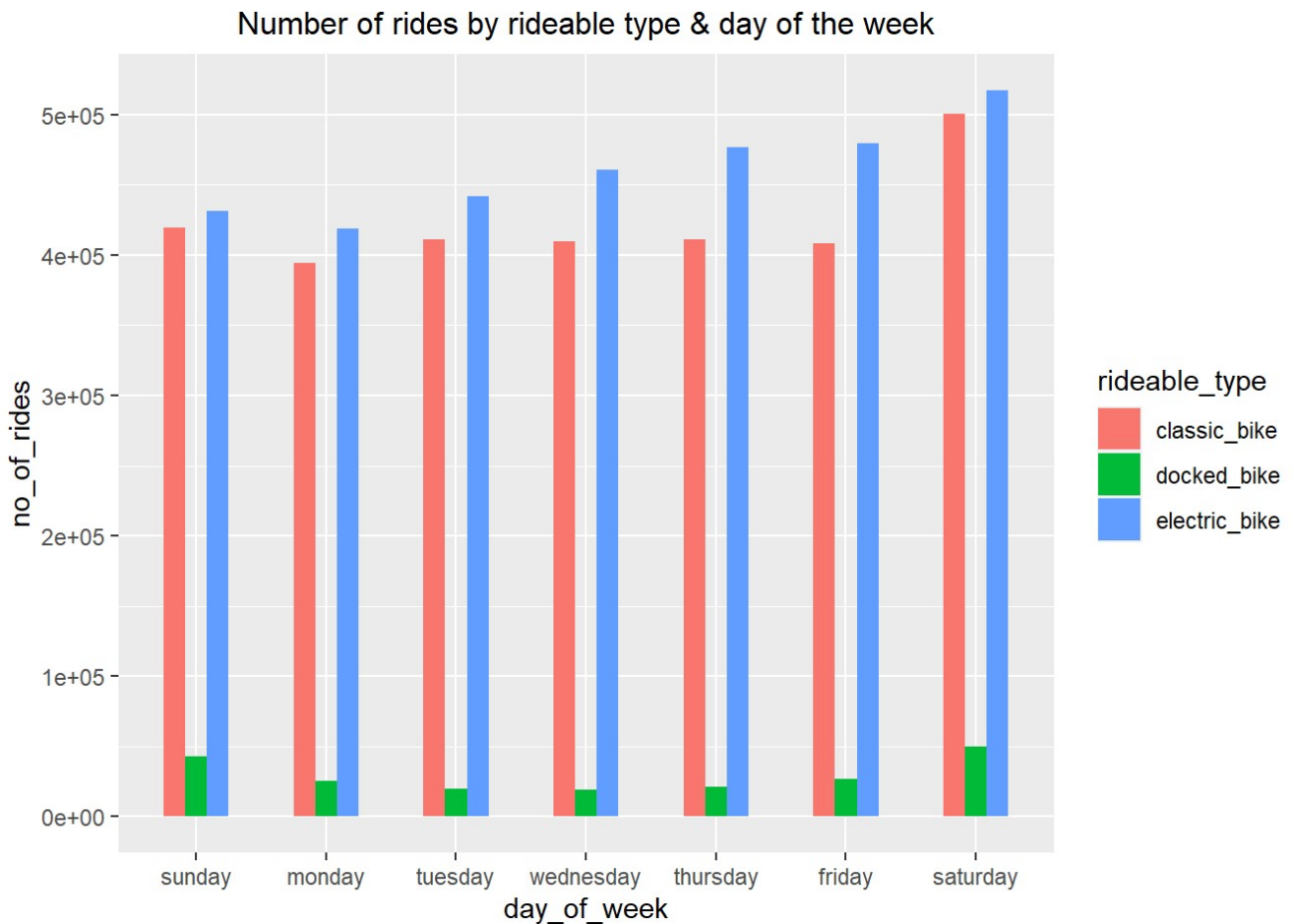
Number of rides by day of the week

```
bike_type2 <- df %>%
  group_by(day_of_week, rideable_type) %>%
  summarise(no_of_rides = n())
```

```
bike_type2_tbl <- spread(bike_type2, key= "rideable_type", value= "no_of_rides")
```

```
bar6 <- ggplot(data=bike_type2, aes(x=day_of_week, y=no_of_rides, fill=rideable_type)) +
  geom_bar(stat="identity", width = 0.5, position = 'dodge')+
  scale_x_discrete(limits = positions)+
  ggtitle("Number of rides by rideable type & day of the week")+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

bar6



bike_type2_tbl

```
## # A tibble: 7 × 4
## # Groups:   day_of_week [7]
##   day_of_week classic_bike docked_bike electric_bike
##   <chr>          <int>      <int>      <int>
## 1 friday         408071      27120      479702
## 2 monday         394140      25235      418931
## 3 saturday       500859      49902      517513
## 4 sunday         419319      42886      431379
## 5 thursday       410990      20938      477060
## 6 tuesday        410890      19720      442053
## 7 wednesday      409654      19277      460645
```

- Number of bike rides on all three bikes was highest on Saturday followed by Sunday.
- Classic and docked bikes usage was less than electric bikes throughout the week.
- This may be due to the fact that, electric bikes are faster and take less effort.

This must have resulted in

- more people choosing electric bikes.
- Faster rides means faster bike returns and in turn more availability.

Average ride duration by rideable type & rider type

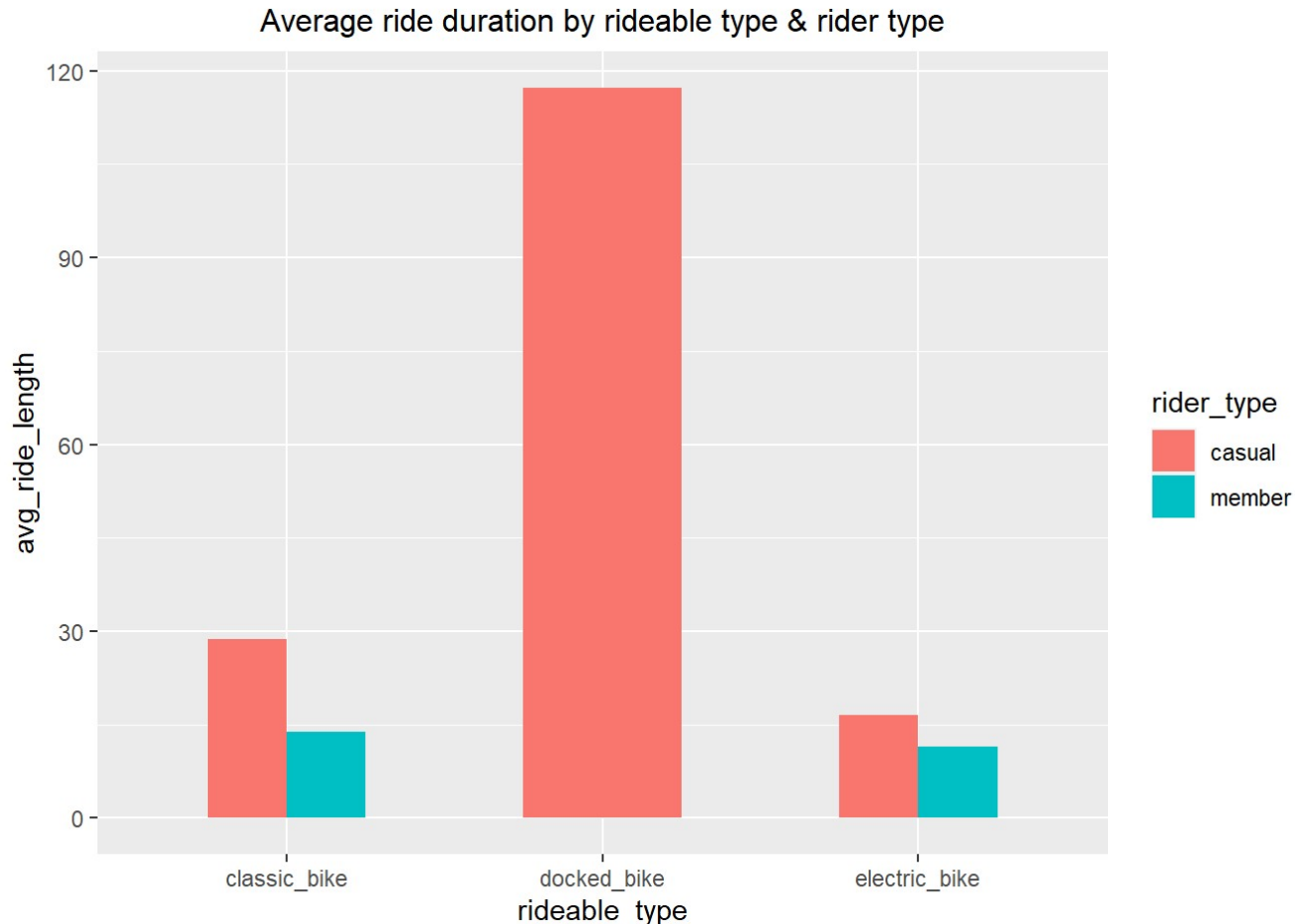
```
avg_ride_length_by_ride_type <- df %>%  
  group_by(rideable_type, rider_type) %>%  
  summarise(avg_ride_length = round(mean(ride_length), digits=2))
```

```
avg_ride_length_by_ride_type_tbl <- spread(avg_ride_length_by_ride_type, key= "rider_type",  
  value= "avg_ride_length")
```

```
avg_ride_length_by_ride_type_tbl <- avg_ride_length_by_ride_type_tbl %>%  
  rename( "avg_ride_length_casual" = "casual",  
    "avg_ride_length_member" = "member")
```

```
bar7 <- ggplot(data=avg_ride_length_by_ride_type, aes(x=rideable_type, y=avg_ride_length,  
  fill=rider_type)) +  
  geom_bar(stat="identity", width = 0.5, position = 'dodge')+  
  ggtitle("Average ride duration by rideable type & rider type")+  
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

bar7



```
avg_ride_length_by_ride_type_tbl
```

```
## # A tibble: 3 × 3
## # Groups:   rideable_type [3]
##   rideable_type avg_ride_length_casual avg_ride_length_member
##   <chr>          <dbl>          <dbl>
## 1 classic_bike      28.8          13.8
## 2 docked_bike      117.           NA
## 3 electric_bike     16.5          11.5
```

- This graph shows that rides were faster on electric bikes compared to classic / docked bikes.

Month-wise Trends

Number of rides by month-year

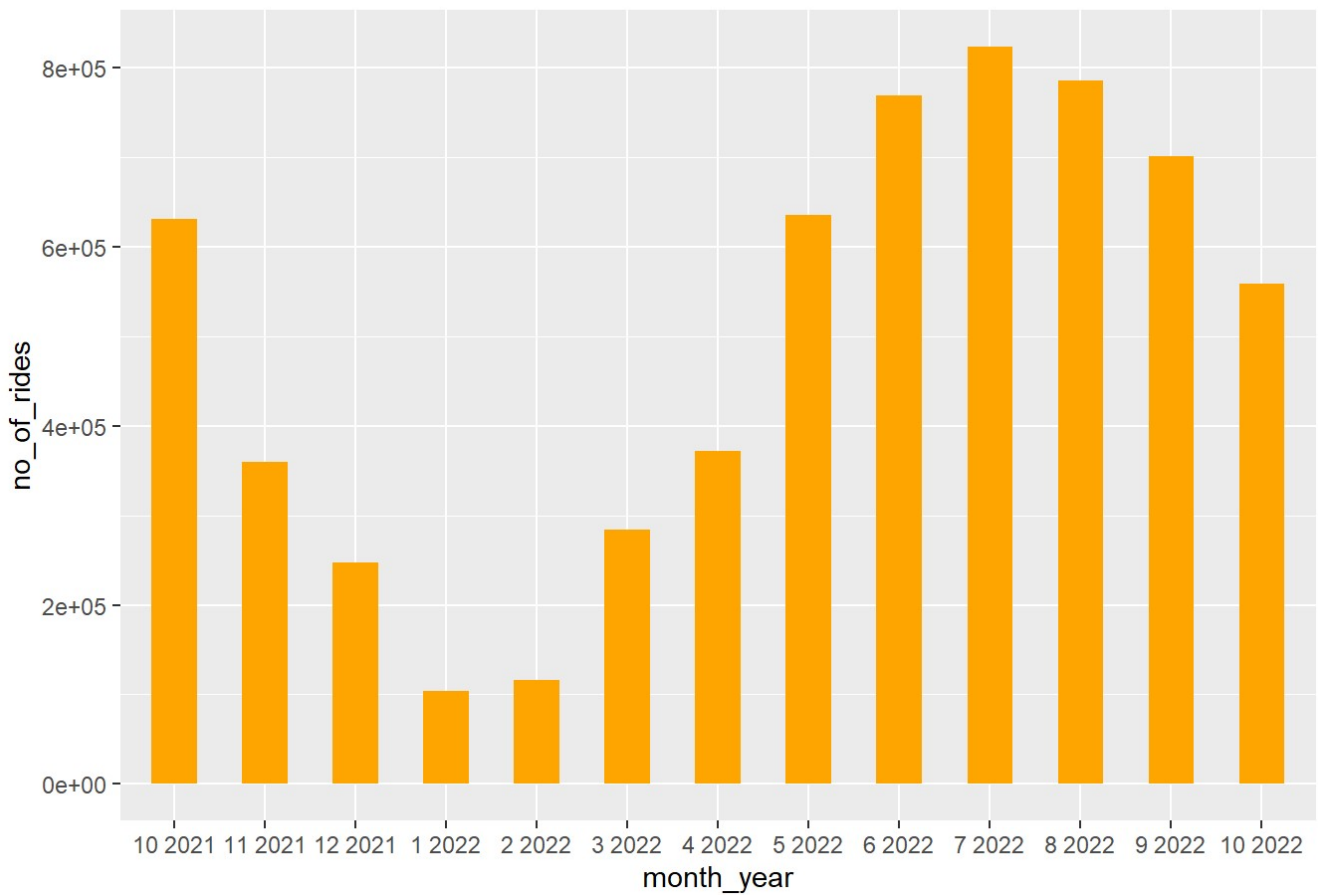
```
monthwise_trend <- df %>%
  group_by(month_year) %>%
  summarise(no_of_rides = n())
```

```
positions3 <- c("10 2021","11 2021","12 2021","1 2022","2 2022","3 2022", "4 2022", "5 2022", "6 2022", "7 2022", "8 2022", "9 2022", "10 2022")
```

```
bar8 <- ggplot(data=monthwise_trend, aes(x=month_year, y=no_of_rides)) +
  geom_bar(stat="identity", width = 0.5, fill="orange")+
  scale_x_discrete(limits = positions3)+
  ggtitle("Number of rides by month-year")+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

```
bar8
```

Number of rides by month-year



monthwise_trend

```
## # A tibble: 13 × 2
##   month_year no_of_rides
##   <chr>      <int>
## 1 1 2022      103765
## 2 10 2021     631156
## 3 10 2022     558620
## 4 11 2021     359892
## 5 12 2021     247519
## 6 2 2022      115604
## 7 3 2022      284024
## 8 4 2022      371218
## 9 5 2022      634810
## 10 6 2022     769138
## 11 7 2022     823416
## 12 8 2022     785855
## 13 9 2022     701267
```

- This graph shows how the number of rides gradually increased around summer and decreased during winters.

Number of rides by month-year & rider type

```
monthwise_trend1 <- df %>%
  group_by(month_year, rider_type) %>%
  summarise(no_of_rides = n())
```



```
monthwise_trend_tbl1 <- spread(monthwise_trend1, key= "rider_type", value= "no_of_rides")
```

```
positions3 <- c("10 2021","11 2021","12 2021","1 2022","2 2022","3 2022", "4 2022", "5 2022", "6 2022", "7 2022", "8 2022", "9 2022", "10 2022")
```

```
bar9 <- ggplot(data=monthwise_trend1, aes(x=month_year, y=no_of_rides, fill=rider_type)) +  
  geom_bar(stat="identity", width = 0.7, position = 'dodge') +  
  scale_x_discrete(limits = positions3) +  
  ggtitle("Number of rides by month-year & rider type") +  
  theme(plot.title = element_text(hjust = 0.5, size = 12),  
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

bar9



monthwise_trend_tbl1

```
## # A tibble: 13 × 3
## # Groups:   month_year [13]
##   month_year casual member
##   <chr>      <int> <int>
## 1 1 2022      18517  85248
## 2 10 2021     257203 373953
## 3 10 2022     208961 349659
## 4 11 2021     106884 253008
## 5 12 2021      69729 177790
## 6 2 2022      21414  94190
## 7 3 2022      89874 194150
## 8 4 2022     126398 244820
## 9 5 2022     280387 354423
## 10 6 2022     369022 400116
## 11 7 2022     406013 417403
## 12 8 2022     358886 426969
## 13 9 2022     296664 404603
```

- Number of rides by members were always higher than casual riders.
- There was marginal difference between number of rides by casual riders and members June and July.
- The difference kept increasing as the winter approached, and during winter even rides by members decreased.

Season-wise Trends

Number of rides

```
season <- df %>%
  group_by(season) %>%
  summarise(no_of_rides = n())

season$percentage <- round((season$no_of_rides/ total_rides)*100, digits=2)
```

```
seasonal_freq_tbl <- season%>%
  bind_rows(summarise(., across(where(is.numeric), sum),
                        across(where(is.character), ~'Total')))
```

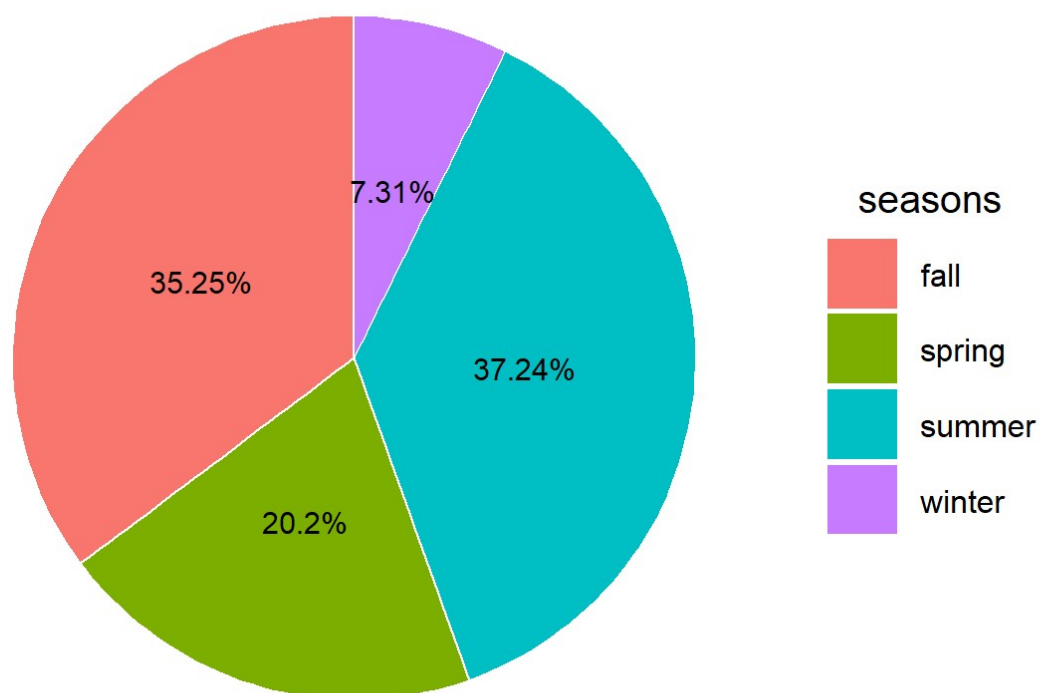
```
percentage2 <- round(season$percentage, digits=2)
seasons <- season$season

ypos1 = cumsum(percentage2) - 0.5 * percentage2
ypos1 = 100 - ypos1

pie2 <- ggplot() + theme_bw() +
  geom_bar(aes(x = "", y = percentage2, fill = seasons),
    stat = "identity", color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("Seasonwise Trend") +
  theme(plot.title = element_text(hjust = 0.5, size = 20),
    axis.title = element_blank(),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid=element_blank(),
    panel.border = element_blank()) +
  theme(legend.text=element_text(size=12),
    legend.title = element_text(hjust = 0.5, size=15),
    legend.key.size = unit(1,"cm")) +
  geom_text(aes(x = "", y = ypos1, label = paste0(percentage2, "%")), size = 4)
```

pie2

Seasonwise Trend



seasonal_freq_tbl

```
## # A tibble: 5 × 3
##   season no_of_rides percentage
##   <chr>      <int>      <dbl>
## 1 fall        2250935      35.2
## 2 spring      1290052      20.2
## 3 summer      2378409      37.2
## 4 winter        466888       7.31
## 5 Total      6386284     100
```

- Highest numbers of rides were observed in Summer followed by fall.
- Lowest number of rides were seen in Winters.
- As the data is for Chicago, which has tremendous amounts of snow and winds during winters, such low numbers of rides during winters are justified.

Number of rides by Season & Rider type

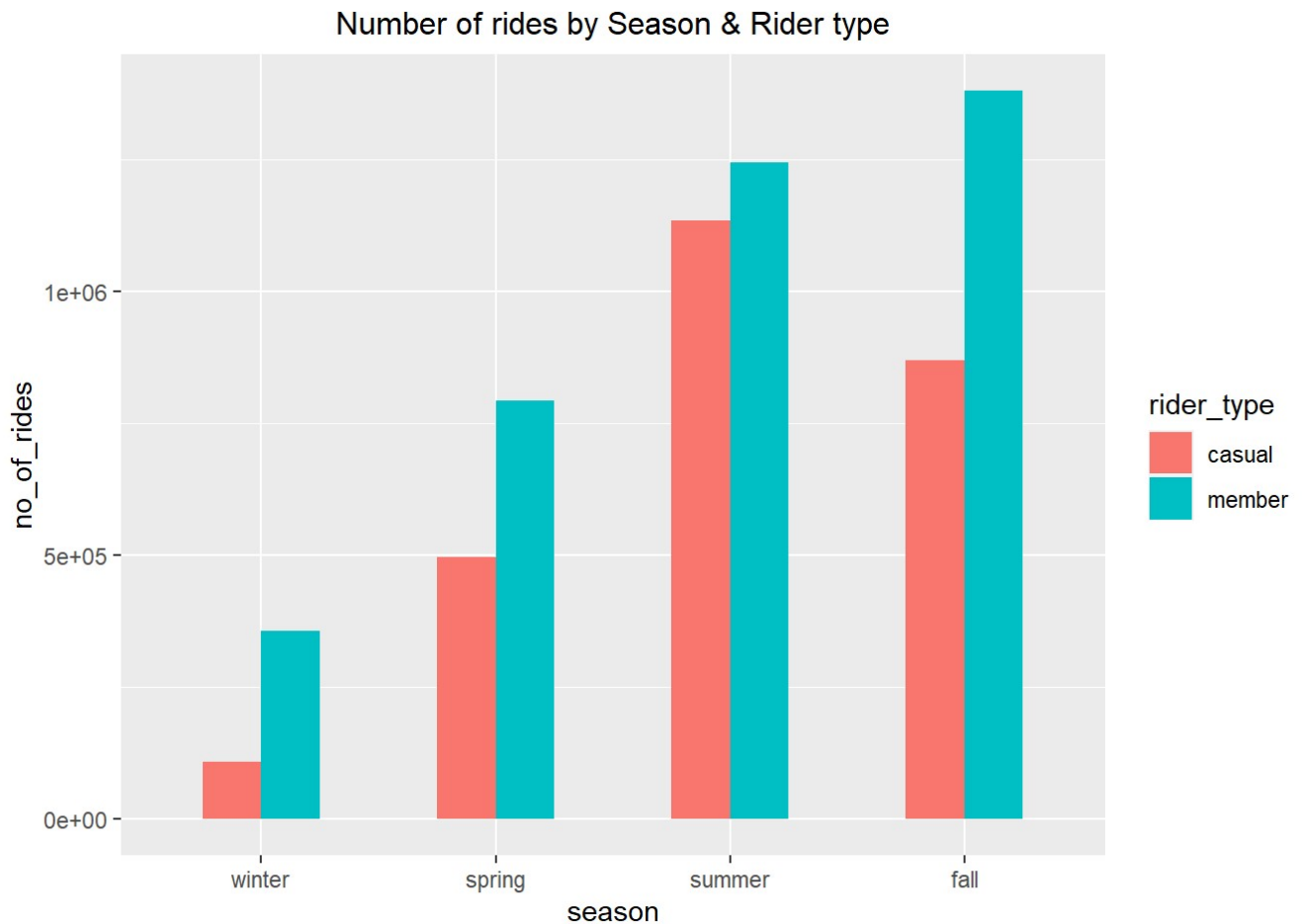
```
seasonwise_rides <- df %>%
  group_by(season, rider_type) %>%
  summarise(no_of_rides = n())
```

```
seasonwise_rides_tbl <- spread(seasonwise_rides, key= "rider_type", value= "no_of_rides")
```

```
positions2 <- c("winter", "spring", "summer", "fall")
```

```
bar10 <- ggplot(data=seasonwise_rides, aes(x=season, y=no_of_rides, fill=rider_type)) +
  geom_bar(stat="identity", width = 0.5, position = 'dodge')+
  scale_x_discrete(limits = positions2)+
  ggtitle("Number of rides by Season & Rider type")+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

```
bar10
```



```
seasonwise_rides_tbl
```

```
## # A tibble: 4 × 3
## # Groups:   season [4]
##   season casual member
##   <chr>   <int> <int>
## 1 fall     869712 1381223
## 2 spring   496659  793393
## 3 summer  1133921 1244488
## 4 winter   109660  357228
```

- Number of rides by members were higher than Casual riders each season.
- Fall was season in which members used maximum number of bike rides.
- Maximum number of rides by casual riders were seen in Summer.

Average ride duration by Season & Rider type

```
avg_ride_length_by_season <- df %>%
  group_by(season, rider_type) %>%
  summarise(avg_ride_length = round(mean(ride_length), digits=2))
```

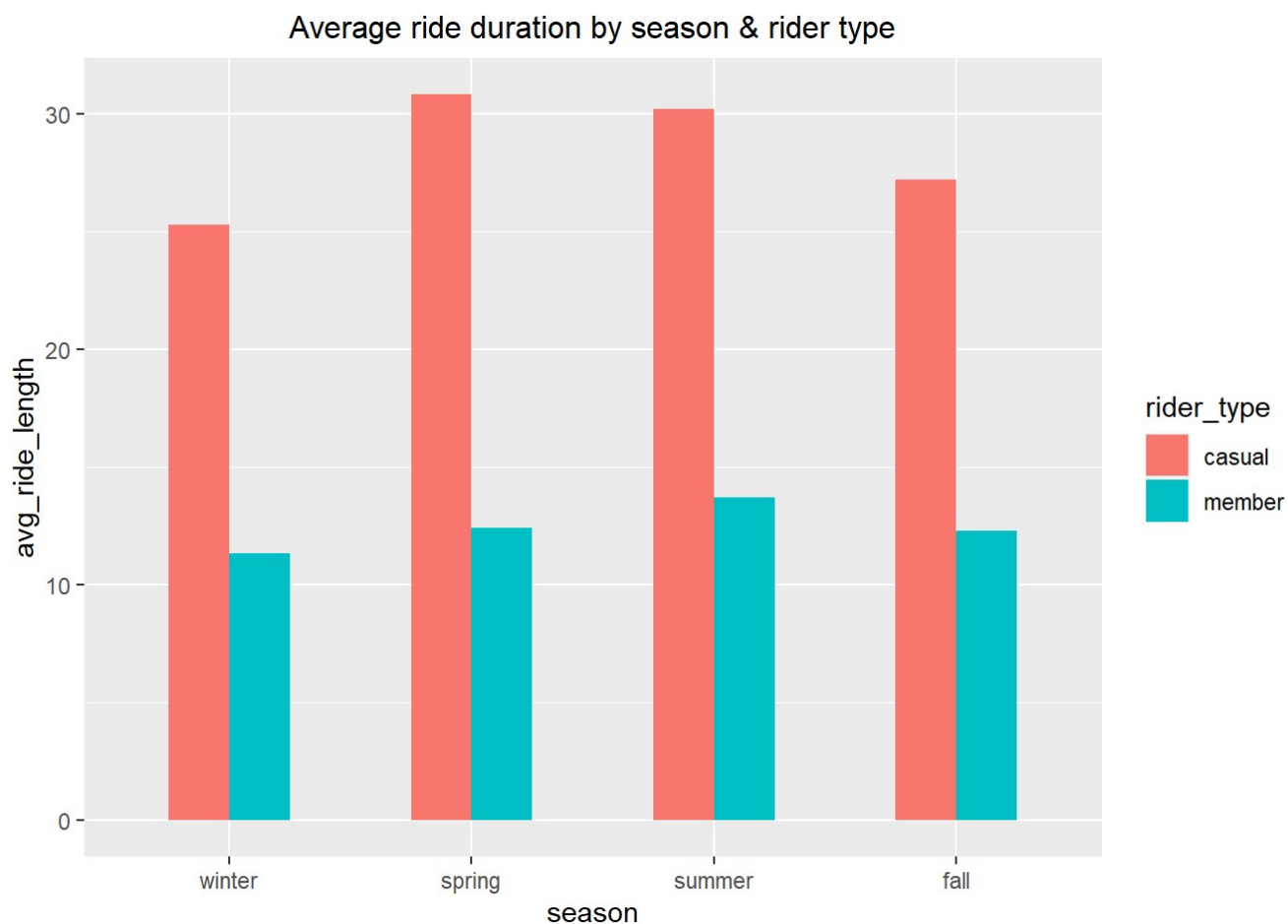
```
avg_ride_length_by_season_tbl <- spread(avg_ride_length_by_season, key= "rider_type", value= "avg_ride_length")
```

```
avg_ride_length_by_season_tbl <- avg_ride_length_by_season_tbl %>%
  rename( "avg_ride_length_casual" = "casual",
          "avg_ride_length_member" = "member")
```

```
positions2 <- c("winter","spring","summer","fall")
```

```
bar11 <- ggplot(data=avg_ride_length_by_season, aes(x=season, y=avg_ride_length, fill=rider_type)) +
  geom_bar(stat="identity", width = 0.5, position = 'dodge')+
  scale_x_discrete(limits = positions2)+
  ggtitle("Average ride duration by season & rider type")+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

bar11



avg_ride_length_by_season_tbl

```
## # A tibble: 4 × 3
## # Groups:   season [4]
##   season avg_ride_length_casual avg_ride_length_member
##   <chr>          <dbl>          <dbl>
## 1 fall           27.2           12.3
## 2 spring         30.8           12.4
## 3 summer         30.2           13.7
## 4 winter         25.3           11.3
```

- Average ride length remained within the range of 25 to 31 mins.
- Average ride duration remained within the range of 11 to 14 mins.

Holiday-wise Trends

```
#find unique dates in data and create a new data frame
```

```
unique_dates <- unique(df$start_date)
unique_dates_df <- as.data.frame(unique_dates)
```

```
#add day column to new data frame
```

```
unique_dates_df$day_num <- wday(unique_dates_df$unique_dates)

unique_dates_df <- unique_dates_df %>%
  mutate(day=
    ifelse(day_num==1,"sunday",
      ifelse(day_num==2, "monday",
        ifelse(day_num==3, "tuesday",
          ifelse(day_num==4, "wednesday",
            ifelse(day_num==5, "thursday",
              ifelse(day_num==6, "friday",
                "saturday" ))))))))
```

```
#add holiday column to new data frame
```

```
unique_dates_df <- unique_dates_df %>%
  mutate(holidays=
    ifelse(unique_dates=="2021-11-11" | unique_dates=="2021-11-25" | unique_dates=="20
21-12-24"
      | unique_dates=="2021-12-31" | unique_dates=="2022-01-01" | unique_dates=="20
22-03-17"
      | unique_dates=="2022-05-30" | unique_dates=="2022-07-04" | unique_dates=="20
22-09-05"
      | day=="sunday" | day=="saturday",
      "holiday",
      "workday"))
```

```
holiday_tbl <- unique_dates_df %>%
  group_by(holidays) %>%
  summarise("no_of_days"=n())

holiday_tbl
```

```
## # A tibble: 2 × 2
##   holidays no_of_days
##   <chr>      <int>
## 1 holiday      122
## 2 workday      274
```

```
no_of_rides_by_holiday <- df %>%
  group_by(holiday,rider_type) %>%
  summarise("no_of_rides"= n())

no_of_rides_by_holiday <- no_of_rides_by_holiday %>%
  spread(key = rider_type, value=no_of_rides)

no_of_rides_by_holiday$total = no_of_rides_by_holiday$casual + no_of_rides_by_holiday$member

no_of_rides_by_holiday$num_of_days <- holiday_tbl$no_of_days

no_of_rides_by_holiday <- as.data.frame(no_of_rides_by_holiday)

no_of_rides_by_holiday
```

```
##   holiday  casual  member  total num_of_days
## 1 holiday 1060024 1019811 2079835      122
## 2 workday 1549928 2756521 4306449      274
```

```
rides_by_holiday <- no_of_rides_by_holiday %>%
  select(holiday,casual,member,num_of_days)

rides_by_holiday <-rides_by_holiday %>%
  gather(key=rider_type, no_of_rides, casual:member)

rides_by_holiday <- rides_by_holiday %>%
  mutate(avg_no_of_rides = round(no_of_rides/num_of_days,digits=2))
```

```
total_rides_by_holiday <- no_of_rides_by_holiday %>%
  select(holiday,total,num_of_days) %>%
  rename("no_of_rides" = "total") %>%
  mutate(avg_rides = no_of_rides/num_of_days)
```

```
all_riders_tbl <- no_of_rides_by_holiday %>%
  select(holiday, total, num_of_days) %>%
  mutate(avg_rides= total/num_of_days)
```


Holidays vs Workdays

```
total_num_of_days = sum(total_rides_by_holiday$num_of_days)

holidays <- total_rides_by_holiday %>%
  mutate(percentage = round(num_of_days/total_num_of_days*100, digits=2)) %>%
  select(holiday, num_of_days, percentage)
```

```
percentage3 <- round(holidays$percentage, digits=2)
```

```
holiday_Class <- holidays$holiday
holiday_Class
```

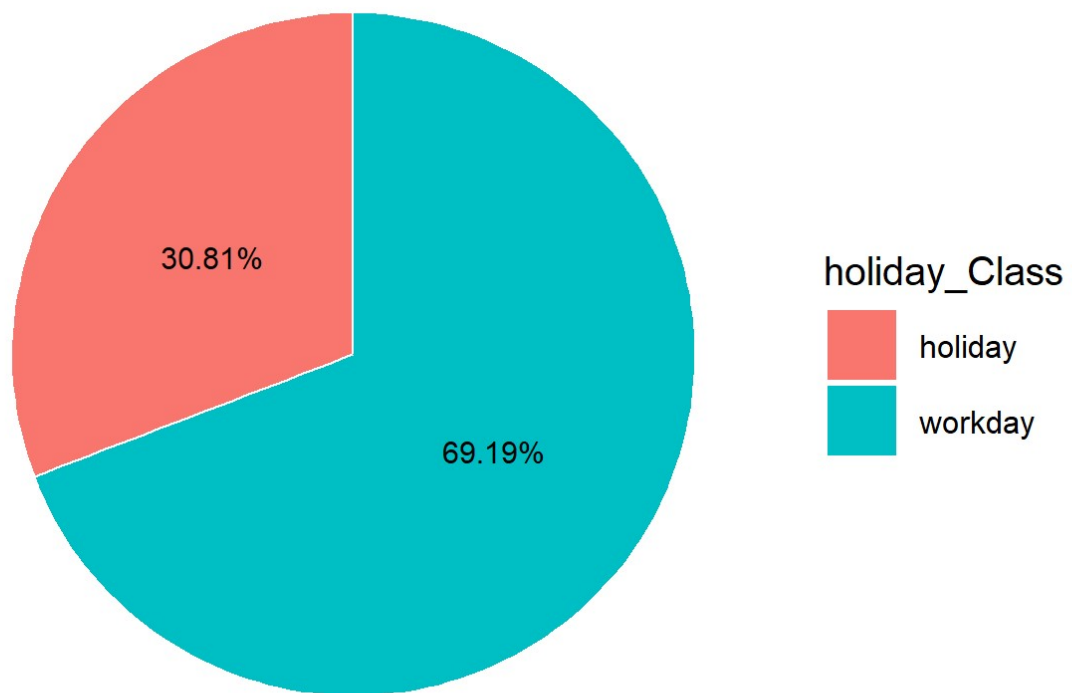
```
## [1] "holiday" "workday"
```

```
ypos = cumsum(percentage3) - 0.5 * percentage3
ypos = 100 - ypos
```

```
pie3 = ggplot() + theme_bw() +
  geom_bar(aes(x = "", y = percentage3, fill = holiday_Class),
    stat = "identity", color = "white") +
  coord_polar("y", start = 0) +
  ggtitle("Holidays vs Workdays") +
  theme(plot.title = element_text(hjust = 0.5, size = 20),
    axis.title = element_blank(),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid=element_blank(),
    panel.border = element_blank()) +
  theme(legend.text=element_text(size=12),
    legend.title = element_text(hjust = 0.5, size=15),
    legend.key.size = unit(1,"cm")) +
  geom_text(aes(x = "", y = ypos, label = paste0(percentage3, "%")), size = 4)
```

```
pie3
```

Holidays vs Workdays



rides_by_holiday

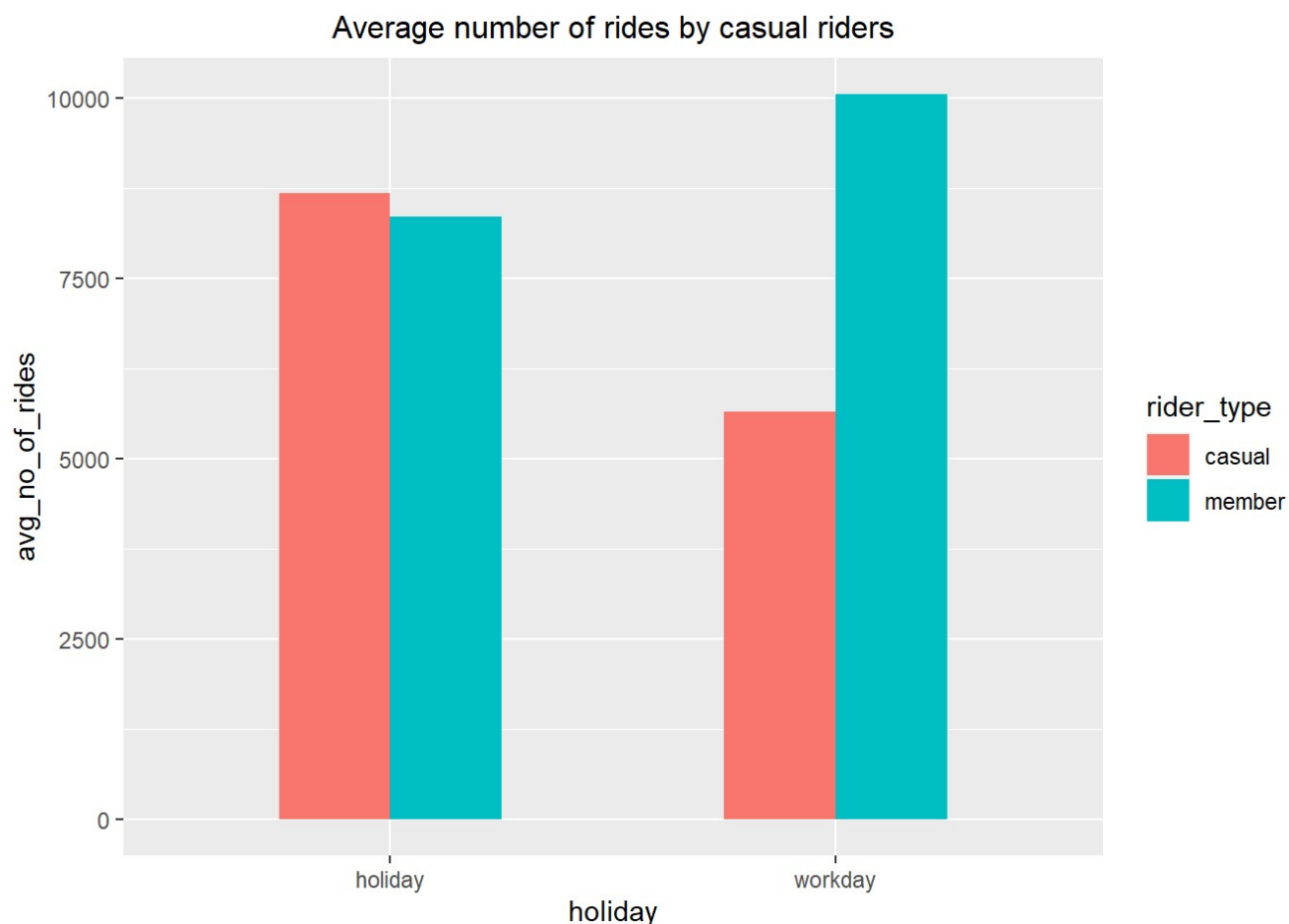
```
##   holiday num_of_days rider_type no_of_rides avg_no_of_rides
## 1 holiday      122      casual   1060024      8688.72
## 2 workday      274      casual  1549928      5656.67
## 3 holiday      122      member   1019811      8359.11
## 4 workday      274      member  2756521     10060.30
```

Average number of rides

```
positions5 <- c("holiday", "workday")

bar12 <- ggplot(data=rides_by_holiday, aes(x=holiday, y=avg_no_of_rides, fill=rider_type))
+
  geom_bar(stat="identity", width = 0.5, position = "dodge")+
  ggtitle("Average number of rides by casual riders")+
  scale_x_discrete(limits = positions5)+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

bar12



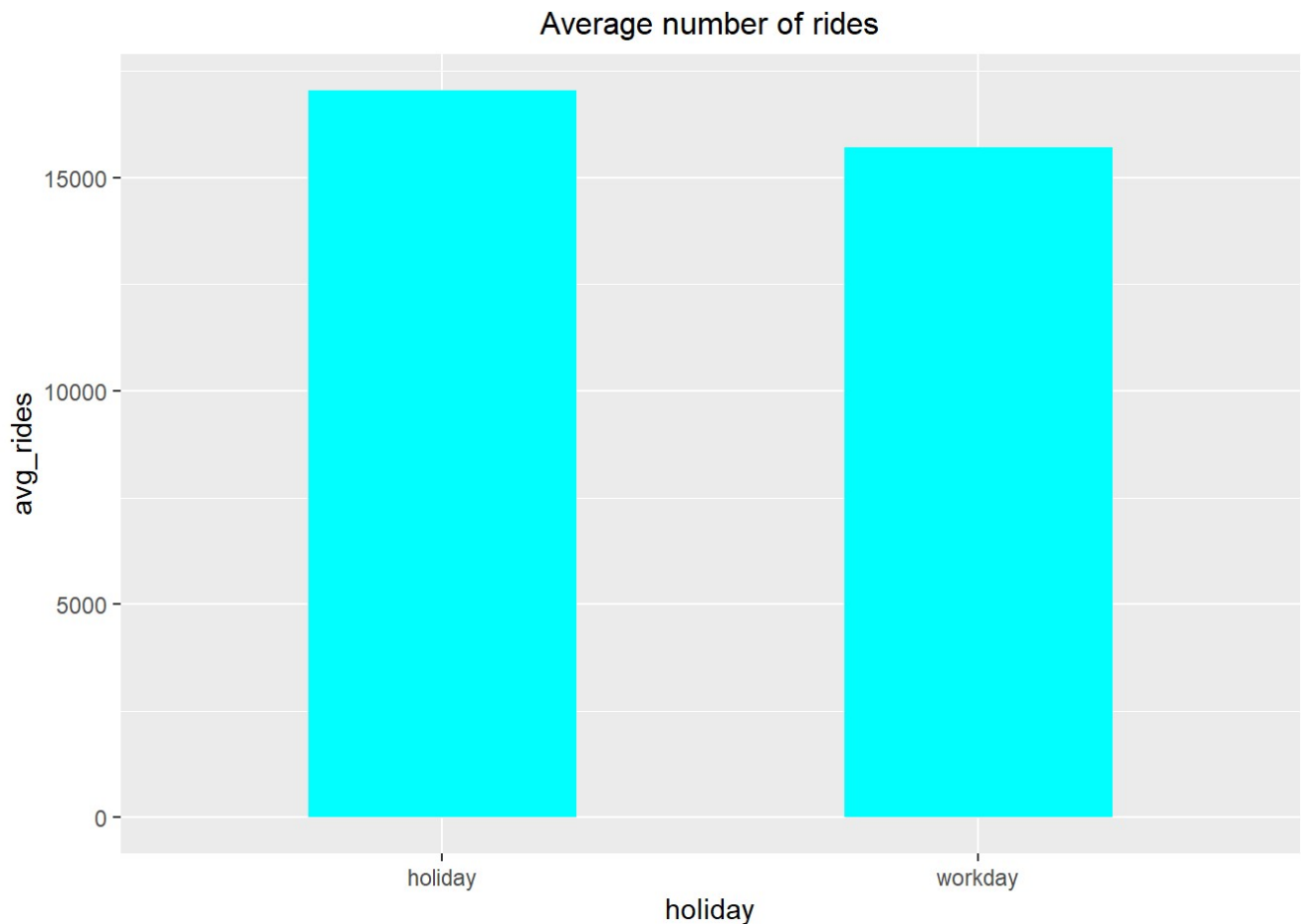
rides_by_holiday

```
##   holiday num_of_days rider_type no_of_rides avg_no_of_rides
## 1 holiday         122   casual   1060024      8688.72
## 2 workday         274   casual   1549928      5656.67
## 3 holiday         122   member   1019811      8359.11
## 4 workday         274   member   2756521     10060.30
```

Casual riders used more numbers of bikes during holidays, members preferred to use more bikes on Workdays. During holidays, number of casual riders was more than members. During non-holiday days, number of members was more than casual riders.

```
bar13 <- ggplot(data=all_riders_tbl, aes(x=holiday, y=avg_rides)) +
  geom_bar(stat="identity", width = 0.5, fill="cyan")+
  ggtitle("Average number of rides")+
  scale_x_discrete(limits = positions5)+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

bar13



```
all_riders_tbl
```

```
##   holiday  total num_of_days avg_rides
## 1 holiday 2079835         122 17047.83
## 2 workday 4306449         274 15716.97
```

In general average number of bikes used on holidays was more than workdays.

Average ride duration

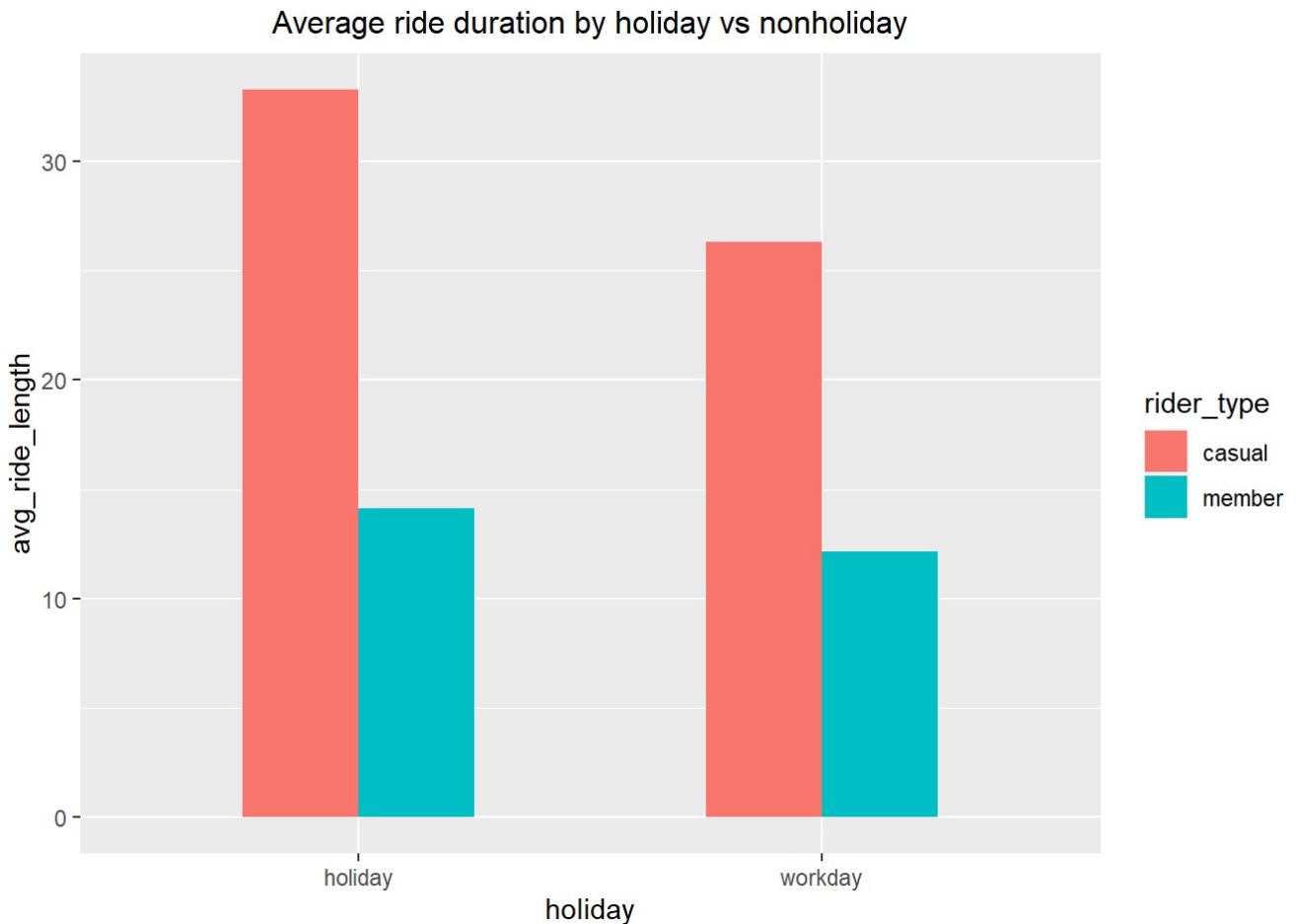
```
avg_ride_length_by_holiday_tbl <- df %>%
  group_by(holiday, rider_type) %>%
  summarise(avg_ride_length = round(mean(ride_length), digits=2))
```

```
avg_ride_length_by_holiday_tbl
```

```
## # A tibble: 4 × 3
## # Groups:   holiday [2]
##   holiday rider_type avg_ride_length
##   <chr>    <chr>         <dbl>
## 1 holiday casual      33.3
## 2 holiday member     14.1
## 3 workday casual      26.3
## 4 workday member     12.2
```

```
bar14 <- ggplot(data=avg_ride_length_by_holiday_tbl, aes(x=holiday, y=avg_ride_length, fill=rider_type)) +
  geom_bar(stat="identity", width = 0.5, position = "dodge")+
  ggtitle("Average ride duration by holiday vs nonholiday")+
  scale_x_discrete(limits = positions5)+
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

bar14



avg_ride_length_by_holiday_tbl

```
## # A tibble: 4 × 3
## # Groups:   holiday [2]
##   holiday rider_type avg_ride_length
##   <chr>    <chr>         <dbl>
## 1 holiday casual          33.3
## 2 holiday member          14.1
## 3 workday casual          26.3
## 4 workday member          12.2
```

Rides duration were seen to be more on holidays compared to work days.