Cyclistic - Data Process

2023-08-06

Import and bind data into data frame

```
# data_all <- rbind(read.csv("dataclean/202207-clean.csv"),
#                   read.csv("dataclean/202208-clean.csv"),
#                   read.csv("dataclean/202209-clean.csv"),
#                   read.csv("dataclean/202210-clean.csv"),
#                   read.csv("dataclean/202211-clean.csv"),
#                   read.csv("dataclean/202212-clean.csv"),
#                   read.csv("dataclean/202301-clean.csv"),
#                   read.csv("dataclean/202302-clean.csv"),
#                   read.csv("dataclean/202303-clean.csv"),
#                   read.csv("dataclean/202304-clean.csv"),
#                   read.csv("dataclean/202305-clean.csv"),
#                   read.csv("dataclean/202306-clean.csv"))

# load save environtment
load("data_all.RData")
```

```
head(data_all)
```

```
##            ride_id rideable_type          started_at            ended_at
## 1 954144C2F67B1932  classic_bike 2022-07-05 08:12:47 2022-07-05 08:24:32
## 2 292E027607D218B6  classic_bike 2022-07-26 12:53:38 2022-07-26 12:55:31
## 3 57765852588AD6E0  classic_bike 2022-07-03 13:58:49 2022-07-03 14:06:32
## 4 B5B6BE44314590E6  classic_bike 2022-07-31 17:44:21 2022-07-31 18:42:50
## 5 A4C331F2A00E79E0  classic_bike 2022-07-13 19:49:06 2022-07-13 20:15:24
## 6 579D73BE2ED880B3 electric_bike 2022-07-01 17:04:35 2022-07-01 17:13:18
##   start_station_id end_station_id start_lat start_lng  end_lat   end_lng
## 1            13224   KA1503000043  41.90707 -87.66725 41.88918 -87.63851
## 2            15541            623  41.86962 -87.62398 41.87277 -87.62398
## 3            15541            623  41.86962 -87.62398 41.87277 -87.62398
## 4            15541   TA1307000164  41.86962 -87.62398 41.79526 -87.59647
## 5    TA1307000117   TA1307000052  41.89147 -87.62676 41.93625 -87.65266
## 6            15535         WL-008  41.88461 -87.64456 41.86712 -87.64109
##   member_casual time_total year month day
## 1        member   705 secs 2022     7   5
## 2        casual   113 secs 2022     7  26
## 3        casual   463 secs 2022     7   3
## 4        casual  3509 secs 2022     7  31
## 5        member  1578 secs 2022     7  13
## 6        member   523 secs 2022     7   1
```

```r
count(data_all)
```

```
##         n
## 1 5773649
```

Format data datetime to correct it data types

format started_at and ended_at from char to datetime or S3: POSIXct

```r
data_all$started_at <- as.POSIXct(data_all$started_at, format="%Y-%m-%d %H:%M:%S", tz="UTC")
data_all$ended_at <- as.POSIXct(data_all$ended_at, format="%Y-%m-%d %H:%M:%S", tz="UTC")

head(data_all)
```

```
##            ride_id rideable_type          started_at            ended_at
## 1 954144C2F67B1932  classic_bike 2022-07-05 08:12:47 2022-07-05 08:24:32
## 2 292E027607D218B6  classic_bike 2022-07-26 12:53:38 2022-07-26 12:55:31
## 3 57765852588AD6E0  classic_bike 2022-07-03 13:58:49 2022-07-03 14:06:32
## 4 B5B6BE44314590E6  classic_bike 2022-07-31 17:44:21 2022-07-31 18:42:50
## 5 A4C331F2A00E79E0  classic_bike 2022-07-13 19:49:06 2022-07-13 20:15:24
## 6 579D73BE2ED880B3 electric_bike 2022-07-01 17:04:35 2022-07-01 17:13:18
##   start_station_id end_station_id start_lat start_lng  end_lat   end_lng
## 1            13224   KA1503000043  41.90707 -87.66725 41.88918 -87.63851
## 2            15541            623  41.86962 -87.62398 41.87277 -87.62398
## 3            15541            623  41.86962 -87.62398 41.87277 -87.62398
## 4            15541   TA1307000164  41.86962 -87.62398 41.79526 -87.59647
## 5    TA1307000117   TA1307000052  41.89147 -87.62676 41.93625 -87.65266
## 6            15535         WL-008  41.88461 -87.64456 41.86712 -87.64109
##   member_casual time_total year month day
## 1        member   705 secs 2022     7   5
## 2        casual   113 secs 2022     7  26
## 3        casual   463 secs 2022     7   3
## 4        casual  3509 secs 2022     7  31
## 5        member  1578 secs 2022     7  13
## 6        member   523 secs 2022     7   1
```

count time duration in each data and join it into table

```r
data_all$time_total <- data_all$ended_at - data_all$started_at

head(data_all)
```

```
##            ride_id rideable_type          started_at            ended_at
## 1 954144C2F67B1932  classic_bike 2022-07-05 08:12:47 2022-07-05 08:24:32
## 2 292E027607D218B6  classic_bike 2022-07-26 12:53:38 2022-07-26 12:55:31
## 3 57765852588AD6E0  classic_bike 2022-07-03 13:58:49 2022-07-03 14:06:32
## 4 B5B6BE44314590E6  classic_bike 2022-07-31 17:44:21 2022-07-31 18:42:50
## 5 A4C331F2A00E79E0  classic_bike 2022-07-13 19:49:06 2022-07-13 20:15:24
## 6 579D73BE2ED880B3 electric_bike 2022-07-01 17:04:35 2022-07-01 17:13:18
```

```
##   start_station_id end_station_id start_lat start_lng  end_lat   end_lng
## 1            13224   KA1503000043  41.90707 -87.66725 41.88918 -87.63851
## 2            15541            623  41.86962 -87.62398 41.87277 -87.62398
## 3            15541            623  41.86962 -87.62398 41.87277 -87.62398
## 4            15541   TA1307000164  41.86962 -87.62398 41.79526 -87.59647
## 5    TA1307000117   TA1307000052  41.89147 -87.62676 41.93625 -87.65266
## 6            15535         WL-008  41.88461 -87.64456 41.86712 -87.64109
##   member_casual time_total year month day
## 1        member   705 secs 2022     7   5
## 2        casual   113 secs 2022     7  26
## 3        casual   463 secs 2022     7   3
## 4        casual  3509 secs 2022     7  31
## 5        member  1578 secs 2022     7  13
## 6        member   523 secs 2022     7   1
```

separate year, month, and day in each data and join it into table

```
data_all <- data_all %>%
  mutate(year = year(started_at),
         month = month(started_at),
         day = day(started_at))

head(data_all)
```

```
##            ride_id rideable_type          started_at            ended_at
## 1 954144C2F67B1932  classic_bike 2022-07-05 08:12:47 2022-07-05 08:24:32
## 2 292E027607D218B6  classic_bike 2022-07-26 12:53:38 2022-07-26 12:55:31
## 3 57765852588AD6E0  classic_bike 2022-07-03 13:58:49 2022-07-03 14:06:32
## 4 B5B6BE44314590E6  classic_bike 2022-07-31 17:44:21 2022-07-31 18:42:50
## 5 A4C331F2A00E79E0  classic_bike 2022-07-13 19:49:06 2022-07-13 20:15:24
## 6 579D73BE2ED880B3 electric_bike 2022-07-01 17:04:35 2022-07-01 17:13:18
##   start_station_id end_station_id start_lat start_lng  end_lat   end_lng
## 1            13224   KA1503000043  41.90707 -87.66725 41.88918 -87.63851
## 2            15541            623  41.86962 -87.62398 41.87277 -87.62398
## 3            15541            623  41.86962 -87.62398 41.87277 -87.62398
## 4            15541   TA1307000164  41.86962 -87.62398 41.79526 -87.59647
## 5    TA1307000117   TA1307000052  41.89147 -87.62676 41.93625 -87.65266
## 6            15535         WL-008  41.88461 -87.64456 41.86712 -87.64109
##   member_casual time_total year month day
## 1        member   705 secs 2022     7   5
## 2        casual   113 secs 2022     7  26
## 3        casual   463 secs 2022     7   3
## 4        casual  3509 secs 2022     7  31
## 5        member  1578 secs 2022     7  13
## 6        member   523 secs 2022     7   1
```

save data into .RData files

```
#save environtment data_all
# save(data_all, file="data_all.RData")
```

Visualize data

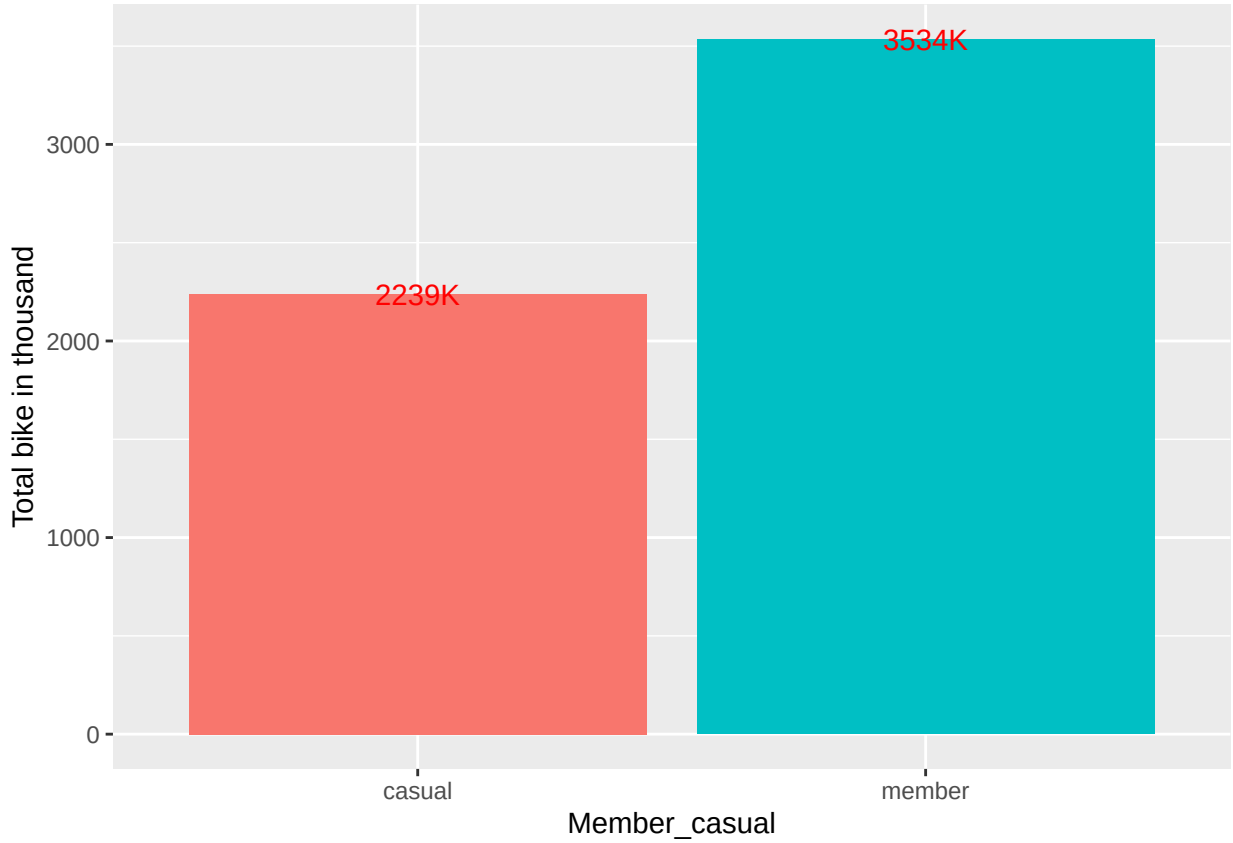
Visualize total ride of casual and member

```
# count total ride for casual and member in bilion

ride_total <- data.frame(label=c("member", "casual"),
                         total=c(count(data_all[data_all$member_casual == "member", ])[1,1],
                                 count(data_all[data_all$member_casual == "casual", ])[1,1]
                                       )
                         )

ride_total
```

a. Total ride in a year (July 2022 - June 2023)

```
##    label   total
## 1 member 3534428
## 2 casual 2239221
```

```
# Visualize into barchart
ride_total %>%
  ggplot(aes(x=label, y=total/1000, fill=label)) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = paste0(sprintf("%.0f", total/1000), "K")), color = "red") +
  ylab("Total bike in thousand") +
  xlab("Member_casual")
```



From visualize we know that member users is bike more often than casual users

```
# initialize data frame
ride_total_month <- data.frame(month = c("jul", "aug", "sep", "oct", "nov", "dec", "jan", "feb",
                                         "march", "apr", "mei", "jun",
                                         "jul", "aug", "sep", "oct", "nov", "dec", "jan", "feb",
                                         "march", "apr", "mei", "jun"))
count_df <- data.frame(member_casual = c(NA), total = c(NA))
```

```
# set function to get month data using loop

count_ride <- function(data = data_all, mem_cas){
  #this function to take the ride data on data frame

  if(mem_cas == "member") {
      count_data <-  count(data[data$member_casual == mem_cas & data$month == 7, ])[1,1]
      count_df <- data.frame(member_casual = c(mem_cas), total = c(count_data))
  }
  else {
    count_data <-  count(data[data$member_casual == mem_cas & data$month == 7, ])[1,1]
      count_df[nrow(count_df) + 1, ] <- c(member_casual = mem_cas, total = count_data)
  }
```

```
  for (i in 8:12) {
      count_data <-  count(data[data$member_casual == mem_cas & data$month == i, ])[1,1]
      count_df[nrow(count_df) + 1, ] <- c(member_casual = mem_cas, total = count_data)
  }
  for (i in 1:6) {
    count_data <-  count(data[data$member_casual == mem_cas & data$month == i, ])[1,1]
    count_df[nrow(count_df) + 1, ] <- c(member_casual = mem_cas, total = count_data)
  }

  return(count_df)
}

#get data ride
count_df <- count_df[nrow(count_df) + 1, ] <- count_ride(mem_cas = "member")
```

b. Total ride monthly (July 2022 - June 2023)

```
## Warning in `[<-.data.frame`(`*tmp*`, nrow(count_df) + 1, , value =
## structure(list(: replacement element 1 has 12 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(count_df) + 1, , value =
## structure(list(: replacement element 2 has 12 rows to replace 1 rows
```

```
count_df <- count_df[nrow(count_df) + 1, ] <- count_ride(mem_cas = "casual")
```

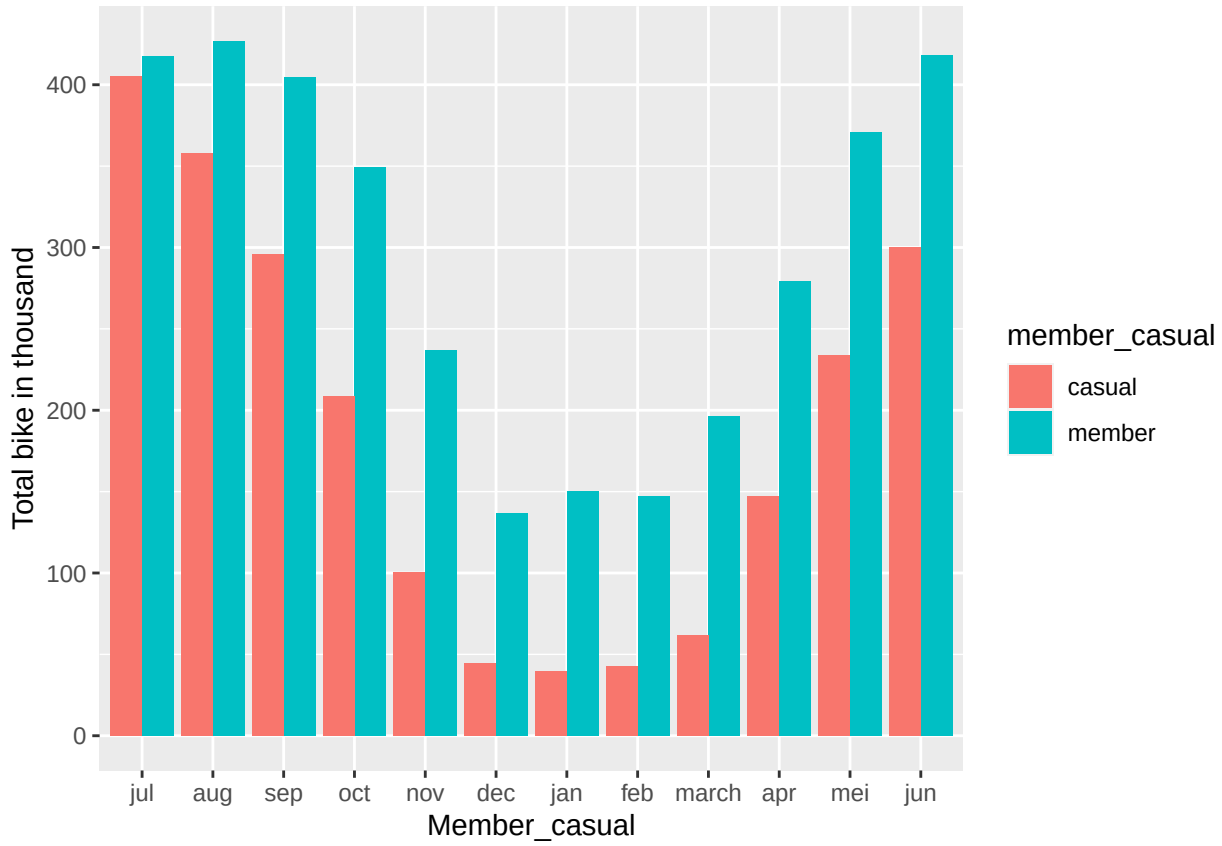
```
## Warning in `[<-.data.frame`(`*tmp*`, nrow(count_df) + 1, , value =
## structure(list(: replacement element 1 has 24 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(count_df) + 1, , value =
## structure(list(: replacement element 2 has 24 rows to replace 1 rows
```

```
# join into table
ride_total_month$member_casual <- count_df$member_casual
ride_total_month$total <-  as.numeric(count_df$total)

ride_total_month
```

```
##    month member_casual  total
## 1    jul        member 417353
## 2    aug        member 426921
## 3    sep        member 404550
## 4    oct        member 349598
## 5    nov        member 236921
## 6    dec        member 136887
## 7    jan        member 150269
## 8    feb        member 147407
## 9  march        member 196446
## 10   apr        member 279261
## 11   mei        member 370526
## 12   jun        member 418289
## 13   jul        casual 405188
```

```
## 14   aug        casual 358168
## 15   sep        casual 296077
## 16   oct        casual 208612
## 17   nov        casual 100584
## 18   dec        casual  44791
## 19   jan        casual  39905
## 20   feb        casual  42922
## 21 march        casual  62049
## 22   apr        casual 146894
## 23   mei        casual 233591
## 24   jun        casual 300440
```

```r
# Visualize into barchart

ride_total_month %>%
  ggplot(aes(fill=member_casual, y=total/1000, x=fct_inorder(month))) +
  geom_bar(position="dodge", stat="identity") +
  ylab("Total bike in thousand") +
  xlab("Member_casual")
```



Form visualize we know that there are more bikers in middle of the year (in US this is spring time)

```r
# initialize data frame
```

```r
ride_total_days <- data.frame(dates = c(NA), member_casual = c(NA), total = c(NA))
```

```r
# set function to get daily data using loop

count_ride <- function(data = data_all, mem_cas, count_df = NA){
  #this function to take the ride data on data frame

  i = 1
  if(mem_cas == "member") {
    count_data <-  count(data[data$member_casual == mem_cas & data$day == i, ])[1,1]
    count_df <- data.frame(dates = c(i), member_casual = c(mem_cas), total = c(count_data))
  }
  else {
    count_data <-  count(data[data$member_casual == mem_cas & data$day == i, ])[1,1]
    count_df[nrow(count_df) + 1, ] <- c(dates = i, member_casual = mem_cas, total = count_data)
  }

  for (i in 2:31) {
    count_data <-  count(data[data$member_casual == mem_cas & data$day == i, ])[1,1]
    count_df[nrow(count_df) + 1, ] <- c(dates = i, member_casual = mem_cas, total = count_data)
    }
  return(count_df)
}
```

```r
#get data ride
ride_total_days <- ride_total_days[nrow(ride_total_days) + 1, ] <- count_ride(mem_cas = "member")
```

c. Total ride daily of all month (July 2022 - June 2023)

```
## Warning in `[<-.data.frame`(`*tmp*`, nrow(ride_total_days) + 1, , value =
## structure(list(: replacement element 1 has 31 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(ride_total_days) + 1, , value =
## structure(list(: replacement element 2 has 31 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(ride_total_days) + 1, , value =
## structure(list(: replacement element 3 has 31 rows to replace 1 rows
```

```r
ride_total_days <- ride_total_days[nrow(ride_total_days) + 1, ] <- count_ride(mem_cas = "casual", count
```

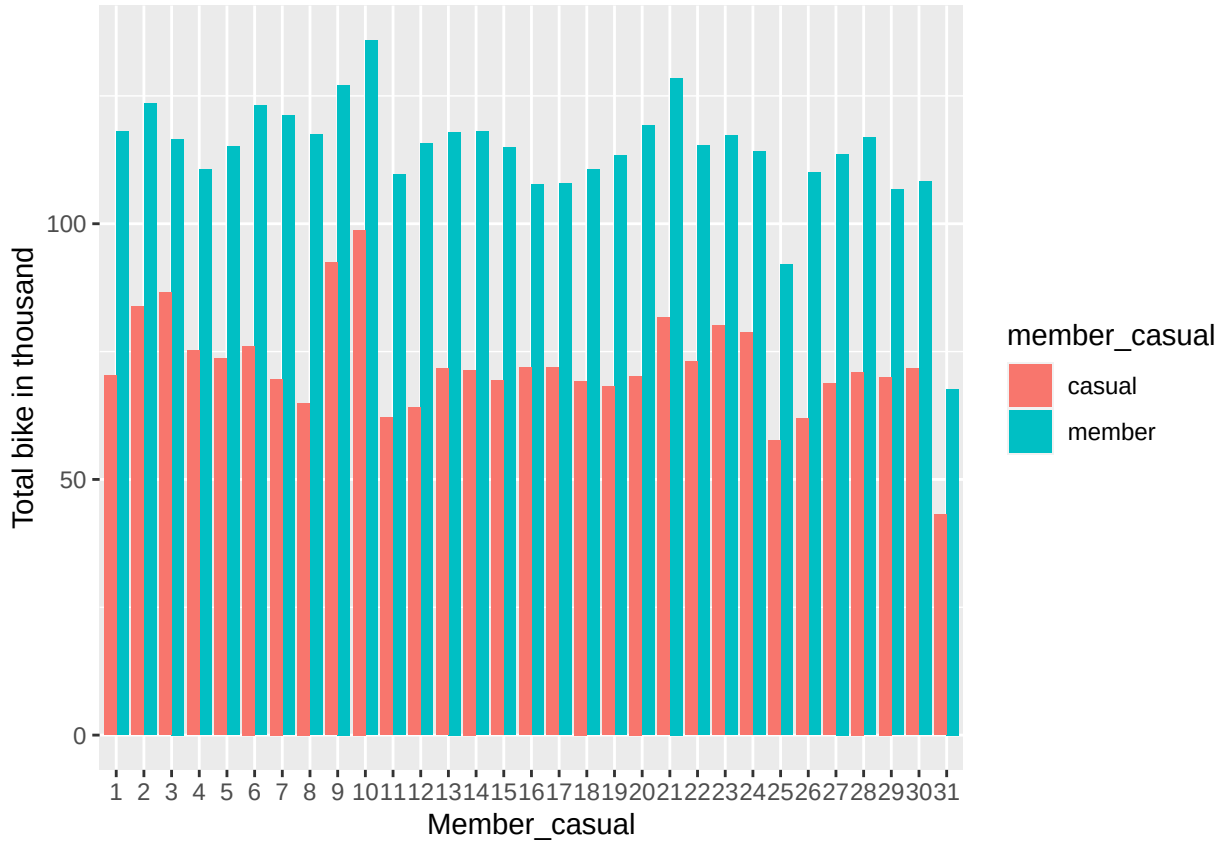
```
## Warning in `[<-.data.frame`(`*tmp*`, nrow(ride_total_days) + 1, , value =
## structure(list(: replacement element 1 has 62 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(ride_total_days) + 1, , value =
## structure(list(: replacement element 2 has 62 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(ride_total_days) + 1, , value =
## structure(list(: replacement element 3 has 62 rows to replace 1 rows
```

```r
ride_total_days$total <-  as.numeric(ride_total_days$total)

ride_total_days
```

```
##    dates member_casual  total
## 1      1        member 118035
## 2      2        member 123573
## 3      3        member 116560
## 4      4        member 110575
## 5      5        member 115137
## 6      6        member 123192
## 7      7        member 121114
## 8      8        member 117517
## 9      9        member 127130
## 10    10        member 135877
## 11    11        member 109632
## 12    12        member 115727
## 13    13        member 117956
## 14    14        member 118053
## 15    15        member 114971
## 16    16        member 107657
## 17    17        member 107919
## 18    18        member 110667
## 19    19        member 113411
## 20    20        member 119178
## 21    21        member 128506
## 22    22        member 115345
## 23    23        member 117266
## 24    24        member 114167
## 25    25        member  92078
## 26    26        member 110074
## 27    27        member 113653
## 28    28        member 116813
## 29    29        member 106696
## 30    30        member 108275
## 31    31        member  67674
## 32     1        casual  70306
## 33     2        casual  83841
## 34     3        casual  86508
## 35     4        casual  75219
## 36     5        casual  73711
## 37     6        casual  76096
## 38     7        casual  69628
## 39     8        casual  64948
## 40     9        casual  92484
## 41    10        casual  98796
## 42    11        casual  62045
## 43    12        casual  64096
## 44    13        casual  71720
## 45    14        casual  71397
## 46    15        casual  69397
## 47    16        casual  71965
## 48    17        casual  71926
```

```
## 49    18         casual  69247
## 50    19         casual  68098
## 51    20         casual  70224
## 52    21         casual  81651
## 53    22         casual  73005
## 54    23         casual  80028
## 55    24         casual  78755
## 56    25         casual  57543
## 57    26         casual  61963
## 58    27         casual  68821
## 59    28         casual  70999
## 60    29         casual  70020
## 61    30         casual  71628
## 62    31         casual  43156
```

```r
# Visualize into bar chart

ride_total_days %>%
  ggplot(aes(fill=member_casual, y=total/1000, x=fct_inorder(dates))) +
  geom_bar(position="dodge", stat="identity") +
  ylab("Total bike in thousand") +
  xlab("Member_casual")
```



From Visualize we know that bike data are distributed evenly for every dates

Visualize duration ride of casual and member

```
data_member <- data_all[data_all$member_casual == "member", ]
data_casual <- data_all[data_all$member_casual == "casual", ]
```

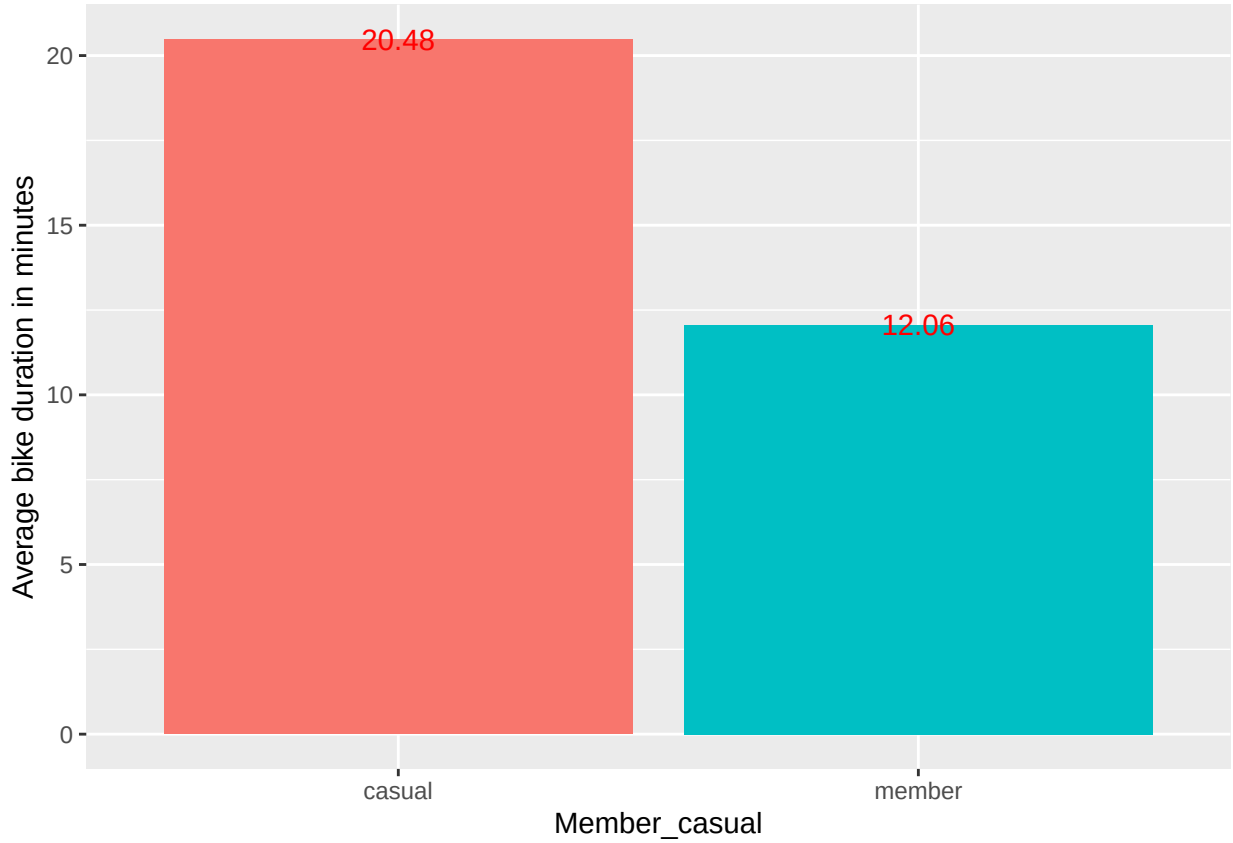
```
# average ride duration for casual and member in bilion

ride_duration <- data.frame(member_casual = c("member", "casual"),
                            duration = c(
                              mean(data_member$time_total),
                              mean(data_casual$time_total)
                            ))

ride_duration
```

a. Average ride duration in a year (July 2022 - June 2023)

```
##   member_casual       duration
## 1        member  723.5959 secs
## 2        casual 1228.6750 secs
```

```
# print(sum(data_member$time_total))
```

```
# Visualize into barchart
ride_duration %>%
  ggplot(aes(x=member_casual, y=as.numeric(duration)/60, fill=member_casual)) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = sprintf("%.2f", as.numeric(duration)/60) ), color = "red") +
  ylab("Average bike duration in minutes") +
  xlab("Member_casual")
```



From visualize we know that casual users bike longer than member users

```
# initialize data frame
ride_dur_month <- data.frame(month = c("jul", "aug", "sep", "oct", "nov", "dec", "jan", "feb",
                                       "march", "apr", "mei", "jun",
                                       "jul", "aug", "sep", "oct", "nov", "dec", "jan", "feb",
                                       "march", "apr", "mei", "jun"))
avg_df <- data.frame(member_casual = c(NA), avg_dur = c(NA))
```

```
# set function to get month data using loop

avg_duration <- function(data, mem_cas, count_df = NA ){
  #this function to take the average ride duration on data frame

  if(mem_cas == "member") {
      count_data <-  mean(data$time_total[data$month == 7])
      count_df <- data.frame(member_casual = c(mem_cas), avg_dur = c(count_data))
  }
  else {
      count_data <-  mean(data$time_total[data$month == 7])
      count_df[nrow(count_df) + 1, ] <- c(member_casual = mem_cas, avg_dur = count_data)
  }
```

```
  for (i in 8:12) {
      count_data <-  mean(data$time_total[data$month == i])
      count_df[nrow(count_df) + 1, ] <- c(member_casual = mem_cas, avg_dur = count_data)
  }
  for (i in 1:6) {
      count_data <-  mean(data$time_total[data$month == i])
      count_df[nrow(count_df) + 1, ] <- c(member_casual = mem_cas, avg_dur = count_data)
  }
  return(count_df)

}

#get data duration
avg_df <- avg_df[nrow(avg_df) + 1, ] <- avg_duration(data = data_member, mem_cas = "member")
```

b. Average ride duration monthly (July 2022 - June 2023)

```
## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_df) + 1, , value =
## structure(list(: replacement element 1 has 12 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_df) + 1, , value =
## structure(list(: replacement element 2 has 12 rows to replace 1 rows
```

```
avg_df <- avg_df[nrow(avg_df) + 1, ] <- avg_duration(data = data_casual, mem_cas = "casual", count_df =
```

```
## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_df) + 1, , value =
## structure(list(: replacement element 1 has 24 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_df) + 1, , value =
## structure(list(: replacement element 2 has 24 rows to replace 1 rows
```

```
avg_df
```

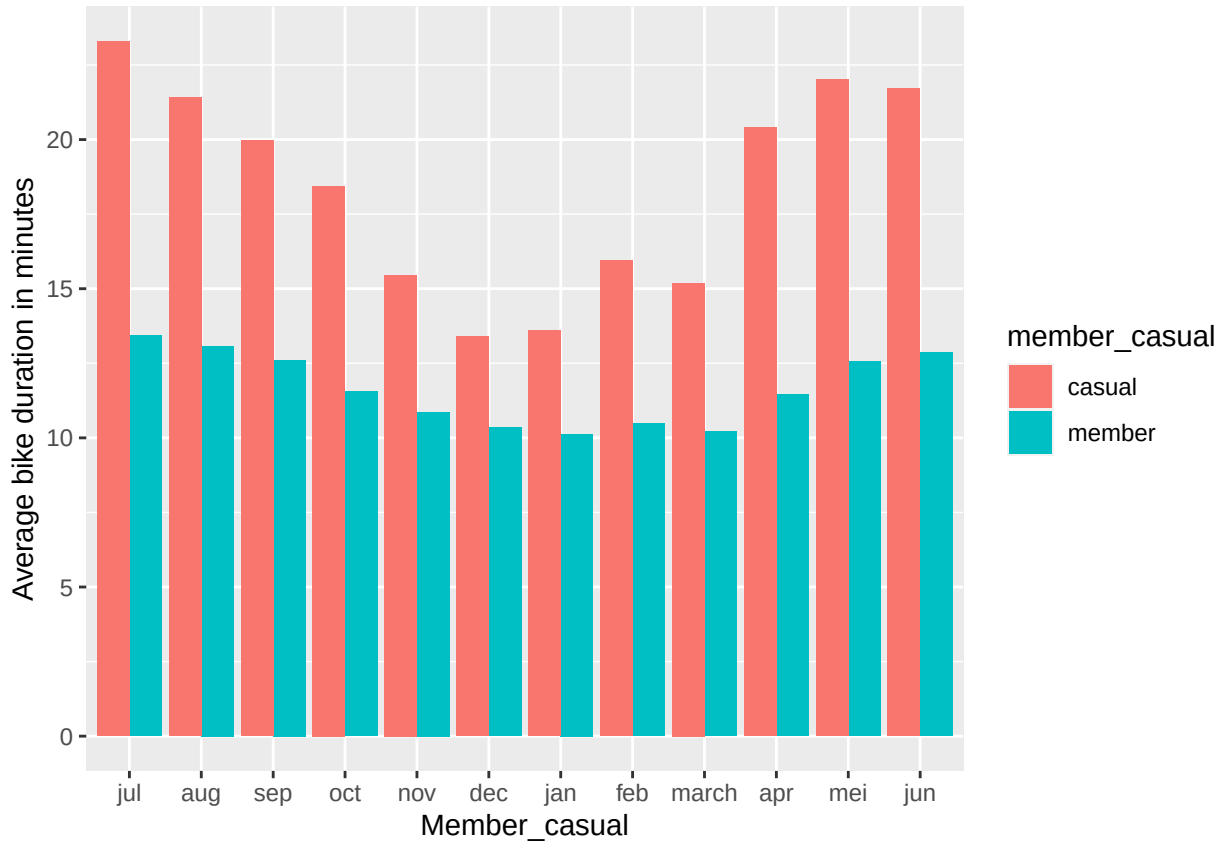
```
##    member_casual               avg_dur
## 1         member 806.008096263834 secs
## 2         member 784.874370199639 secs
## 3         member  756.71946112965 secs
## 4         member 692.699286037105 secs
## 5         member 651.912958327881 secs
## 6         member 621.334166137033 secs
## 7         member 607.432318042976 secs
## 8         member 629.521196415367 secs
## 9         member 612.395070401026 secs
## 10        member 687.556253827065 secs
## 11        member  753.57460744995 secs
## 12        member 771.293041413951 secs
## 13        casual 1397.53622762767 secs
## 14        casual 1285.12612517031 secs
## 15        casual 1198.38650756391 secs
## 16        casual 1106.53513220716 secs
## 17        casual 927.719806331027 secs
```

```
## 18        casual 803.986269563082 secs
## 19        casual 815.432452073675 secs
## 20        casual 956.063860025162 secs
## 21        casual 911.508082322036 secs
## 22        casual 1223.70708129672 secs
## 23        casual 1320.46953007607 secs
## 24        casual 1302.76753428305 secs
```

```
# join into table
ride_dur_month$member_casual <- avg_df$member_casual
ride_dur_month$avg_dur <-  avg_df$avg_dur

ride_dur_month
```

```
##    month member_casual               avg_dur
## 1    jul        member 806.008096263834 secs
## 2    aug        member 784.874370199639 secs
## 3    sep        member  756.71946112965 secs
## 4    oct        member 692.699286037105 secs
## 5    nov        member 651.912958327881 secs
## 6    dec        member 621.334166137033 secs
## 7    jan        member 607.432318042976 secs
## 8    feb        member 629.521196415367 secs
## 9  march        member 612.395070401026 secs
## 10   apr        member 687.556253827065 secs
## 11   mei        member  753.57460744995 secs
## 12   jun        member 771.293041413951 secs
## 13   jul        casual 1397.53622762767 secs
## 14   aug        casual 1285.12612517031 secs
## 15   sep        casual 1198.38650756391 secs
## 16   oct        casual 1106.53513220716 secs
## 17   nov        casual 927.719806331027 secs
## 18   dec        casual 803.986269563082 secs
## 19   jan        casual 815.432452073675 secs
## 20   feb        casual 956.063860025162 secs
## 21 march        casual 911.508082322036 secs
## 22   apr        casual 1223.70708129672 secs
## 23   mei        casual 1320.46953007607 secs
## 24   jun        casual 1302.76753428305 secs
```

```
# Visualize into barchart

ride_dur_month %>%
  ggplot(aes(fill=member_casual, y=as.numeric(avg_dur)/60, x=fct_inorder(month))) +
  geom_bar(position="dodge", stat="identity") +
  ylab("Average bike duration in minutes") +
  xlab("Member_casual")
```



From visualize we know that peoples bike longer in middle of the year (in US this is a spring)

```
# initialize data frame

avg_dur_days <- data.frame(dates = c(NA), member_casual = c(NA), avg_dur = c(NA))
```

```
# set function to get daily data using loop

avg_duration <- function(data, mem_cas, count_df = NA ){
  #this function to take the average ride duration on data frame
  i = 1
  if(mem_cas == "member") {
      count_data <-  mean(data$time_total[data$day == i])
      count_df <- data.frame(dates = c(i), member_casual = c(mem_cas), avg_dur = c(count_data))
  }
  else {
      count_data <-  mean(data$time_total[data$day == i])
      count_df[nrow(count_df) + 1, ] <- c(dates = i, member_casual = mem_cas, avg_dur = count_data)
  }

  for (i in 2:31) {
      count_data <-  mean(data$time_total[data$day == i])
```

```
      count_df[nrow(count_df) + 1, ] <- c(dates = i, member_casual = mem_cas, avg_dur = count_data)
  }
  return(count_df)

}
```

```
#get data duration
avg_dur_days <- avg_dur_days[nrow(avg_dur_days) + 1, ] <- avg_duration(data = data_member, mem_cas = "m
```

c. Average ride duration daily of all month (July 2022 - June 2023)

```
## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_dur_days) + 1, , value =
## structure(list(: replacement element 1 has 31 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_dur_days) + 1, , value =
## structure(list(: replacement element 2 has 31 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_dur_days) + 1, , value =
## structure(list(: replacement element 3 has 31 rows to replace 1 rows
```

```
avg_dur_days <- avg_dur_days[nrow(avg_dur_days) + 1, ] <- avg_duration(data = data_casual, mem_cas = "c
```

```
## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_dur_days) + 1, , value =
## structure(list(: replacement element 1 has 62 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_dur_days) + 1, , value =
## structure(list(: replacement element 2 has 62 rows to replace 1 rows

## Warning in `[<-.data.frame`(`*tmp*`, nrow(avg_dur_days) + 1, , value =
## structure(list(: replacement element 3 has 62 rows to replace 1 rows
```

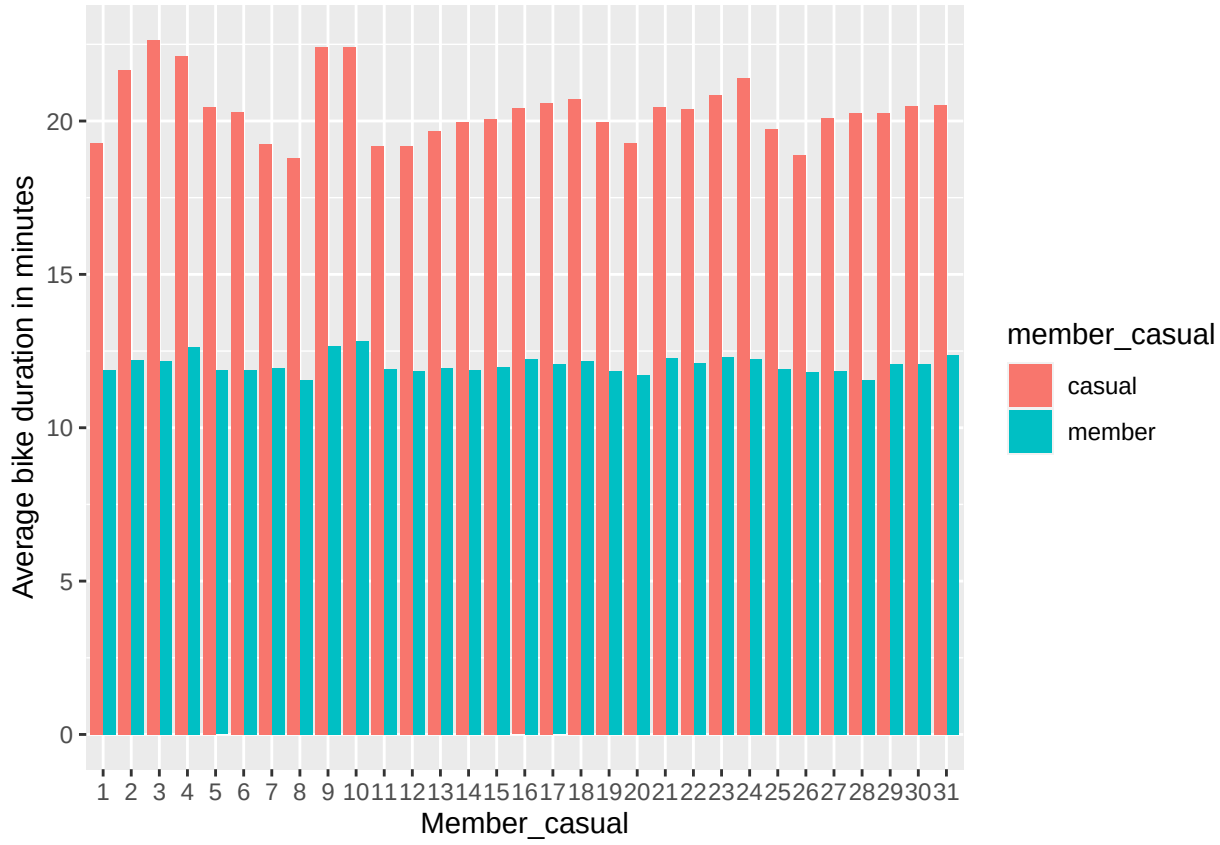
```
avg_dur_days
```

```
##    dates member_casual               avg_dur
## 1      1        member 712.359867835811 secs
## 2      2        member 731.910190737459 secs
## 3      3        member 730.154418325326 secs
## 4      4        member 757.248772326475 secs
## 5      5        member 711.992487210888 secs
## 6      6        member 712.483456717969 secs
## 7      7        member 716.679830572849 secs
## 8      8        member 692.581115923654 secs
## 9      9        member 760.240903012664 secs
## 10    10        member 769.062475621334 secs
## 11    11        member 714.788866389375 secs
## 12    12        member 710.253320314188 secs
## 13    13        member 716.184399267523 secs
## 14    14        member 713.655595368182 secs
## 15    15        member 718.618451609536 secs
```

```
## 16    16        member 734.550163946608 secs
## 17    17        member 723.720716463273 secs
## 18    18        member  729.86905762332 secs
## 19    19        member 711.875065028965 secs
## 20    20        member 703.423282820655 secs
## 21    21        member 737.070743778501 secs
## 22    22        member 726.649269582557 secs
## 23    23        member 737.860044684734 secs
## 24    24        member 734.087293175786 secs
## 25    25        member 715.213482047829 secs
## 26    26        member 709.436342823918 secs
## 27    27        member  710.98069562616 secs
## 28    28        member 692.678237867361 secs
## 29    29        member 725.354530629077 secs
## 30    30        member  724.04783190949 secs
## 31    31        member 743.172636463043 secs
## 32     1        casual 1157.76033339971 secs
## 33     2        casual 1299.94092389165 secs
## 34     3        casual 1359.14344338096 secs
## 35     4        casual  1328.0485781518 secs
## 36     5        casual 1227.17428877644 secs
## 37     6        casual 1217.32468198066 secs
## 38     7        casual 1155.75486873097 secs
## 39     8        casual 1127.57350495781 secs
## 40     9        casual 1345.63459625449 secs
## 41    10        casual 1345.21418883356 secs
## 42    11        casual 1150.44499959707 secs
## 43    12        casual 1151.38740327009 secs
## 44    13        casual 1179.68763245956 secs
## 45    14        casual 1197.65592391837 secs
## 46    15        casual  1204.1634076401 secs
## 47    16        casual 1224.49703328007 secs
## 48    17        casual 1234.97283318967 secs
## 49    18        casual 1242.36634078011 secs
## 50    19        casual 1197.38020206173 secs
## 51    20        casual 1156.68203178401 secs
## 52    21        casual 1227.01325152172 secs
## 53    22        casual 1224.16641325937 secs
## 54    23        casual 1250.90143449793 secs
## 55    24        casual 1283.44103866421 secs
## 56    25        casual 1184.13529013086 secs
## 57    26        casual 1134.39165308329 secs
## 58    27        casual 1205.27041164761 secs
## 59    28        casual 1214.90632262426 secs
## 60    29        casual 1215.49622964867 secs
## 61    30        casual 1229.23701625063 secs
## 62    31        casual 1231.19568542034 secs
```

```
# Visualize into bar chart

avg_dur_days %>%
  ggplot(aes(fill=member_casual, y=as.numeric(avg_dur)/60, x=fct_inorder(dates))) +
  geom_bar(position="dodge", stat="identity") +
  ylab("Average bike duration in minutes") +
```

```
xlab("Member_casual")
```



From Visualize we know that bike data are distributed evenly for every dates