Bike Share Data

Uma Negi

2022-04-24

1. Installing various packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr   1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(readr)
library(ggplot2)
library(repr)
library(plotrix)
```

2.Importing the CSV files to upload the Bike Share data set to R

```
bsd_1 <- read_csv("~/PROJECT/BIKE SHARE/202101-divvy-tripdata.csv")
```

```
## Rows: 96834 Columns: 13
## -- Column specification -------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
bsd_2 <- read_csv("~/PROJECT/BIKE SHARE/202102-divvy-tripdata.csv")
```

```
## Rows: 49622 Columns: 13
## -- Column specification -------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
bsd_3 <- read_csv("~/PROJECT/BIKE SHARE/202103-divvy-tripdata.csv")
```

```
## Rows: 228496 Columns: 13
## -- Column specification -------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
bsd_4 <- read_csv("~/PROJECT/BIKE SHARE/202104-divvy-tripdata.csv")
```

```
## Rows: 337230 Columns: 13
## -- Column specification -------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
bsd_5 <- read_csv("~/PROJECT/BIKE SHARE/202105-divvy-tripdata.csv")
```

```
## Rows: 531633 Columns: 13
## -- Column specification ------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
bsd_6 <- read_csv("~/PROJECT/BIKE SHARE/202106-divvy-tripdata.csv")
```

```
## Rows: 729595 Columns: 13
## -- Column specification ------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
bsd_7 <- read_csv("~/PROJECT/BIKE SHARE/202107-divvy-tripdata.csv")
```

```
## Rows: 822410 Columns: 13
## -- Column specification ------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
bsd_8 <- read_csv("~/PROJECT/BIKE SHARE/202108-divvy-tripdata.csv")
```

```
## Rows: 804352 Columns: 13
## -- Column specification ------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
bsd_9 <- read_csv("~/PROJECT/BIKE SHARE/202109-divvy-tripdata.csv")
```

```
## Rows: 756147 Columns: 13
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
bsd_10 <- read_csv("~/PROJECT/BIKE SHARE/202110-divvy-tripdata.csv")
```

```
## Rows: 631226 Columns: 13
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
bsd_11 <- read_csv("~/PROJECT/BIKE SHARE/202111-divvy-tripdata.csv")
```

```
## Rows: 359978 Columns: 13
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
bsd_12 <- read_csv("~/PROJECT/BIKE SHARE/202112-divvy-tripdata.csv")
```

```
## Rows: 247540 Columns: 13
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

3. Creating Data Frame and binding thee CSV files together

```r
bike_share_data <- rbind(bsd_1, bsd_2, bsd_3, bsd_4, bsd_5, bsd_6, bsd_7, bsd_8, bsd_9, bsd_10, bsd_11,
glimpse(bike_share_data)
```

```
## Rows: 5,595,063
## Columns: 13
## $ ride_id            <chr> "E19E6F1B8D4C42ED", "DC88F20C2C55F27F", "EC45C94683~
## $ rideable_type      <chr> "electric_bike", "electric_bike", "electric_bike", ~
## $ started_at         <dttm> 2021-01-23 16:14:19, 2021-01-27 18:43:08, 2021-01-~
## $ ended_at           <dttm> 2021-01-23 16:24:44, 2021-01-27 18:47:12, 2021-01-~
## $ start_station_name <chr> "California Ave & Cortez St", "California Ave & Cor~
## $ start_station_id   <chr> "17660", "17660", "17660", "17660", "17660", "17660~
## $ end_station_name   <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, "Wood St & Augu~
## $ end_station_id     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, "657", "13258",~
## $ start_lat          <dbl> 41.90034, 41.90033, 41.90031, 41.90040, 41.90033, 4~
## $ start_lng          <dbl> -87.69674, -87.69671, -87.69664, -87.69666, -87.696~
## $ end_lat            <dbl> 41.89000, 41.90000, 41.90000, 41.92000, 41.90000, 4~
## $ end_lng            <dbl> -87.72000, -87.69000, -87.70000, -87.69000, -87.700~
## $ member_casual      <chr> "member", "member", "member", "member", "casual", "~
```

4. Cleaning Data frame for null value

```
bike_share_data <- janitor::remove_empty(bike_share_data, which = c("cols"))
bike_share_data <- janitor::remove_empty(bike_share_data, which = c("rows"))
glimpse(bike_share_data)
```

```
## Rows: 5,595,063
## Columns: 13
## $ ride_id            <chr> "E19E6F1B8D4C42ED", "DC88F20C2C55F27F", "EC45C94683~
## $ rideable_type      <chr> "electric_bike", "electric_bike", "electric_bike", ~
## $ started_at         <dttm> 2021-01-23 16:14:19, 2021-01-27 18:43:08, 2021-01-~
## $ ended_at           <dttm> 2021-01-23 16:24:44, 2021-01-27 18:47:12, 2021-01-~
## $ start_station_name <chr> "California Ave & Cortez St", "California Ave & Cor~
## $ start_station_id   <chr> "17660", "17660", "17660", "17660", "17660", "17660~
## $ end_station_name   <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, "Wood St & Augu~
## $ end_station_id     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, "657", "13258",~
## $ start_lat          <dbl> 41.90034, 41.90033, 41.90031, 41.90040, 41.90033, 4~
## $ start_lng          <dbl> -87.69674, -87.69671, -87.69664, -87.69666, -87.696~
## $ end_lat            <dbl> 41.89000, 41.90000, 41.90000, 41.92000, 41.90000, 4~
## $ end_lng            <dbl> -87.72000, -87.69000, -87.70000, -87.69000, -87.700~
## $ member_casual      <chr> "member", "member", "member", "member", "casual", "~
```

```
summary(bike_share_data)
```

```
##    ride_id           rideable_type        started_at
##  Length:5595063     Length:5595063     Min.   :2021-01-01 00:02:05
##  Class :character    Class :character    1st Qu.:2021-06-06 23:52:40
##  Mode  :character    Mode  :character    Median :2021-08-01 01:52:11
##                                          Mean   :2021-07-29 07:41:02
##                                          3rd Qu.:2021-09-24 16:36:16
##                                          Max.   :2021-12-31 23:59:48
##
##     ended_at                      start_station_name start_station_id
##  Min.   :2021-01-01 00:08:39     Length:5595063      Length:5595063
##  1st Qu.:2021-06-07 00:44:21     Class :character    Class :character
```

```
##   Median :2021-08-01 02:21:55   Mode  :character   Mode  :character
##   Mean   :2021-07-29 08:02:58
##   3rd Qu.:2021-09-24 16:54:05
##   Max.   :2022-01-03 17:32:18
##
##   end_station_name   end_station_id      start_lat       start_lng
##   Length:5595063     Length:5595063    Min.   :41.64   Min.   :-87.84
##   Class :character   Class :character  1st Qu.:41.88   1st Qu.:-87.66
##   Mode  :character   Mode  :character  Median :41.90   Median :-87.64
##                                        Mean   :41.90   Mean   :-87.65
##                                        3rd Qu.:41.93   3rd Qu.:-87.63
##                                        Max.   :42.07   Max.   :-87.52
##
##      end_lat          end_lng        member_casual
##   Min.   :41.39   Min.   :-88.97   Length:5595063
##   1st Qu.:41.88   1st Qu.:-87.66   Class :character
##   Median :41.90   Median :-87.64   Mode  :character
##   Mean   :41.90   Mean   :-87.65
##   3rd Qu.:41.93   3rd Qu.:-87.63
##   Max.   :42.17   Max.   :-87.49
##   NA's   :4771    NA's   :4771
```

```r
clean_data  <- bike_share_data[complete.cases(bike_share_data), ] # remove Null values

table(clean_data$member_casual)      # counting the number of casual and annual member
```

```
##
##   casual  member
## 2048379 2539923
```

5. Processing Data

```r
clean_data$month <- month(clean_data$started_at)  #retrieving the date from  started_at column
clean_data$day <- day(clean_data$started_at)
clean_data$year <- year(clean_data$started_at)
clean_data$day_of_week <- weekdays(clean_data$started_at)
clean_data$month_of_year <- month.abb[clean_data$month]
clean_data$bikeuse_days <- (day(clean_data$ended_at) - day(clean_data$started_at)) #to see how many cus

show <- clean_data %>% count(bikeuse_days) #I found discrepancy in data, some rows are having negative
discrepant_value <- clean_data[clean_data$bikeuse_days < 0,] # I got to find out that the formula that
clean_data$bikeuse_days <- NULL    # removing the column with incorrect value

clean_data$start_date <- as.Date(clean_data$started_at)
clean_data$end_date <- as.Date(clean_data$ended_at)
clean_data$days_bikeuse <- ((clean_data$end_date) - (clean_data$start_date))    # corrected column for

clean_data$start_time <- format(clean_data$started_at, format = "%H:%M")
clean_data$end_time <- format(clean_data$ended_at, format = "%H:%M")
clean_data$time_duration <- difftime(clean_data$ended_at, clean_data$started_at, units = "mins") #getti
```

```r
wrong_duration_value <- clean_data[clean_data$time_duration < 0,] # values with negative time duration
clean_data <- subset(clean_data, clean_data$time_duration > 0)

summary(clean_data)
```

```
##     ride_id          rideable_type         started_at
##  Length:4588104     Length:4588104      Min.    :2021-01-01 00:02:24
##  Class :character    Class :character    1st Qu.:2021-06-04 17:33:46
##  Mode  :character    Mode  :character    Median :2021-07-28 17:33:09
##                                          Mean    :2021-07-25 06:07:47
##                                          3rd Qu.:2021-09-18 15:35:33
##                                          Max.    :2021-12-31 23:59:48
##      ended_at                      start_station_name start_station_id
##  Min.    :2021-01-01 00:08:39    Length:4588104      Length:4588104
##  1st Qu.:2021-06-04 17:54:50    Class :character    Class :character
##  Median :2021-07-28 17:50:45    Mode  :character    Mode  :character
##  Mean    :2021-07-25 06:29:36
##  3rd Qu.:2021-09-18 16:02:12
##  Max.    :2022-01-03 17:32:18
##  end_station_name   end_station_id        start_lat        start_lng
##  Length:4588104     Length:4588104      Min.    :41.65    Min.    :-87.83
##  Class :character    Class :character    1st Qu.:41.88    1st Qu.:-87.66
##  Mode  :character    Mode  :character    Median :41.90    Median :-87.64
##                                          Mean    :41.90    Mean    :-87.64
##                                          3rd Qu.:41.93    3rd Qu.:-87.63
##                                          Max.    :42.06    Max.    :-87.53
##      end_lat          end_lng         member_casual          month
##  Min.    :41.65    Min.    :-87.83    Length:4588104      Min.    : 1.000
##  1st Qu.:41.88    1st Qu.:-87.66    Class :character    1st Qu.: 6.000
##  Median :41.90    Median :-87.64    Mode  :character    Median : 7.000
##  Mean    :41.90    Mean    :-87.64                       Mean    : 7.286
##  3rd Qu.:41.93    3rd Qu.:-87.63                       3rd Qu.: 9.000
##  Max.    :42.17    Max.    :-87.52                       Max.    :12.000
##      day             year         day_of_week        month_of_year
##  Min.    : 1.00    Min.    :2021    Length:4588104      Length:4588104
##  1st Qu.: 8.00    1st Qu.:2021    Class :character    Class :character
##  Median :15.00    Median :2021    Mode  :character    Mode  :character
##  Mean    :15.43    Mean    :2021
##  3rd Qu.:23.00    3rd Qu.:2021
##  Max.    :31.00    Max.    :2021
##     start_date            end_date          days_bikeuse        start_time
##  Min.    :2021-01-01    Min.    :2021-01-01    Length:4588104      Length:4588104
##  1st Qu.:2021-06-04    1st Qu.:2021-06-04    Class :difftime    Class :character
##  Median :2021-07-28    Median :2021-07-28    Mode  :numeric    Mode  :character
##  Mean    :2021-07-24    Mean    :2021-07-24
##  3rd Qu.:2021-09-18    3rd Qu.:2021-09-18
##  Max.    :2021-12-31    Max.    :2022-01-03
##    end_time          time_duration
##  Length:4588104     Length:4588104
##  Class :character    Class :difftime
##  Mode  :character    Mode  :numeric
##
##
```

```
##
```

```
str(clean_data)
```

```
## tibble [4,588,104 x 24] (S3: tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:4588104] "B9F73448DFBE0D45" "457C7F4B5D3DA135" "57C750326F9FDABE" "4D5
##  $ rideable_type     : chr [1:4588104] "classic_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:4588104], format: "2021-01-24 19:15:38" "2021-01-23 12:57:38" ...
##  $ ended_at          : POSIXct[1:4588104], format: "2021-01-24 19:22:51" "2021-01-23 13:02:10" ...
##  $ start_station_name: chr [1:4588104] "California Ave & Cortez St" "California Ave & Cortez St" "Cal
##  $ start_station_id  : chr [1:4588104] "17660" "17660" "17660" "17660" ...
##  $ end_station_name  : chr [1:4588104] "Wood St & Augusta Blvd" "California Ave & North Ave" "Wood St
##  $ end_station_id    : chr [1:4588104] "657" "13258" "657" "657" ...
##  $ start_lat         : num [1:4588104] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:4588104] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:4588104] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:4588104] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:4588104] "member" "member" "casual" "casual" ...
##  $ month             : num [1:4588104] 1 1 1 1 1 1 1 1 1 1 ...
##  $ day               : int [1:4588104] 24 23 9 9 24 22 5 30 27 15 ...
##  $ year              : num [1:4588104] 2021 2021 2021 2021 2021 ...
##  $ day_of_week       : chr [1:4588104] "Sunday" "Saturday" "Saturday" "Saturday" ...
##  $ month_of_year     : chr [1:4588104] "Jan" "Jan" "Jan" "Jan" ...
##  $ start_date        : Date[1:4588104], format: "2021-01-24" "2021-01-23" ...
##  $ end_date          : Date[1:4588104], format: "2021-01-24" "2021-01-23" ...
##  $ days_bikeuse      : 'difftime' num [1:4588104] 0 0 0 0 ...
##   ..- attr(*, "units")= chr "days"
##  $ start_time        : chr [1:4588104] "19:15" "12:57" "15:28" "15:28" ...
##  $ end_time          : chr [1:4588104] "19:22" "13:02" "15:37" "15:37" ...
##  $ time_duration     : 'difftime' num [1:4588104] 7.21666666666667 4.53333333333333 9.78333333333333
##   ..- attr(*, "units")= chr "mins"
```

5. Analyzing Data

```
clean_data %>% group_by(member_casual) %>%
  summarize(max_duration = max(time_duration),
            min_duration = min(time_duration),
            avg_duration = mean(time_duration),
            middle_duration = median(time_duration))
```

```
## # A tibble: 2 x 5
##   member_casual max_duration    min_duration    avg_duration  middle_duration
##   <chr>         <drtn>          <drtn>          <drtn>        <drtn>
## 1 casual          55944.150 mins 0.01666667 mins 32.51015 mins 16.650000 mins
## 2 member           1495.633 mins 0.01666667 mins 13.18401 mins  9.716667 mins
```

```
# Compare ride length between members and casual riders
aggregate(clean_data$time_duration ~ clean_data$member_casual, FUN = mean)
```

```
##   clean_data$member_casual clean_data$time_duration
## 1                   casual             32.51015 mins
## 2                   member             13.18401 mins
```

```
aggregate(clean_data$time_duration ~ clean_data$member_casual, FUN = median)
```

```
##   clean_data$member_casual clean_data$time_duration
## 1                   casual            16.650000 mins
## 2                   member             9.716667 mins
```

```
aggregate(clean_data$time_duration ~ clean_data$member_casual, FUN = max)
```

```
##   clean_data$member_casual clean_data$time_duration
## 1                   casual            55944.150 mins
## 2                   member             1495.633 mins
```

```
aggregate(clean_data$time_duration ~ clean_data$member_casual, FUN = min)
```

```
##   clean_data$member_casual clean_data$time_duration
## 1                   casual           0.01666667 mins
## 2                   member           0.01666667 mins
```

```
# See the average ride length by each day of week for members vs. casual riders
clean_data$day_of_week <- ordered(clean_data$day_of_week,
                           levels = c("Monday", "Tuesday", "Wednesday",
                                      "Thursday", "Friday", "Saturday", "Sunday"))

aggregate(clean_data$time_duration ~ clean_data$member_casual + clean_data$day_of_week, FUN = mean)
```

```
##    clean_data$member_casual clean_data$day_of_week clean_data$time_duration
## 1                    casual                 Monday            32.63724 mins
## 2                    member                 Monday            12.72520 mins
## 3                    casual                Tuesday            28.80335 mins
## 4                    member                Tuesday            12.38802 mins
## 5                    casual              Wednesday            28.26964 mins
## 6                    member              Wednesday            12.45815 mins
## 7                    casual               Thursday            28.01100 mins
## 8                    member               Thursday            12.35370 mins
## 9                    casual                 Friday            30.91907 mins
## 10                   member                 Friday            12.79261 mins
## 11                   casual               Saturday            34.87241 mins
## 12                   member               Saturday            14.81979 mins
## 13                   casual                 Sunday            37.60361 mins
## 14                   member                 Sunday            15.19465 mins
```

```
# See the average ride length by month for members vs. casual riders
clean_data$month_of_year <- ordered(clean_data$month_of_year,
                           levels = c("Jan", "Feb", "Mar","Apr", "May", "Jun",
                                      "Jul", "Aug", "Sep","Oct", "Nov", "Dec"))
aggregate(clean_data$time_duration ~ clean_data$member_casual + clean_data$month_of_year, FUN = mean)
```

```
##   clean_data$member_casual clean_data$month_of_year clean_data$time_duration
## 1                   casual                      Jan            26.37000 mins
## 2                   member                      Jan            12.03113 mins
```

```
## 3              casual         Feb       47.13712 mins
## 4              member         Feb       14.78120 mins
## 5              casual         Mar       38.48048 mins
## 6              member         Mar       13.66530 mins
## 7              casual         Apr       38.44172 mins
## 8              member         Apr       14.26311 mins
## 9              casual         May       39.63541 mins
## 10             member         May       14.34406 mins
## 11             casual         Jun       38.52029 mins
## 12             member         Jun       14.14005 mins
## 13             casual         Jul       33.28582 mins
## 14             member         Jul       13.79144 mins
## 15             casual         Aug       28.56455 mins
## 16             member         Aug       13.55293 mins
## 17             casual         Sep       28.08851 mins
## 18             member         Sep       13.13137 mins
## 19             casual         Oct       26.31429 mins
## 20             member         Oct       12.01463 mins
## 21             casual         Nov       22.48274 mins
## 22             member         Nov       10.95157 mins
## 23             casual         Dec       24.84308 mins
## 24             member         Dec       10.58142 mins
```

```
# Number of rides between members and casual riders for each day of week
clean_data %>%
  group_by(member_casual, day_of_week) %>%
  summarise(number_of_rides = n(), .groups = 'drop') %>%
  arrange(day_of_week)
```

```
## # A tibble: 14 x 3
##    member_casual day_of_week number_of_rides
##    <chr>         <ord>                 <int>
##  1 casual        Monday               228931
##  2 member        Monday               346476
##  3 casual        Tuesday              214932
##  4 member        Tuesday              388120
##  5 casual        Wednesday            218129
##  6 member        Wednesday            397681
##  7 casual        Thursday             224204
##  8 member        Thursday             373469
##  9 casual        Friday               290034
## 10 member        Friday               365774
## 11 casual        Saturday             468318
## 12 member        Saturday             357070
## 13 casual        Sunday               403754
## 14 member        Sunday               311212
```

```
# Number of rides between members and casual riders for each month
clean_data %>%
  group_by(member_casual, month_of_year) %>%
  summarise(number_of_rides = n(), .groups = 'drop') %>%
  arrange(month_of_year)
```

```
## # A tibble: 24 x 3
```

```
##    member_casual month_of_year number_of_rides
##    <chr>          <ord>                   <int>
##  1 casual        Jan                     14690
##  2 member        Jan                     68818
##  3 casual        Feb                      8613
##  4 member        Feb                     34381
##  5 casual        Mar                     75641
##  6 member        Mar                    130046
##  7 casual        Apr                    120418
##  8 member        Apr                    177781
##  9 casual        May                    216823
## 10 member        May                    234155
## # ... with 14 more rows
```

```r
# Comparing general bike type preference between members and casual riders
clean_data %>%
  group_by(rideable_type, member_casual) %>%
  summarize(number_of_rides = n(), .groups = 'drop')
```

```
## # A tibble: 6 x 3
##   rideable_type member_casual number_of_rides
##   <chr>          <chr>                  <int>
## 1 classic_bike  casual              1261508
## 2 classic_bike  member              1980328
## 3 docked_bike   casual               312040
## 4 docked_bike   member                    1
## 5 electric_bike casual               474754
## 6 electric_bike member               559473
```

```r
# Number of rides between members and casual riders for start hour
clean_data$start_hour <- format(clean_data$started_at, format = "%H")
clean_data %>%
  group_by(member_casual, start_hour) %>%
  summarise(number_of_rides = n(), .groups = 'drop') %>%
  arrange(start_hour)
```

```
## # A tibble: 48 x 3
##    member_casual start_hour number_of_rides
##    <chr>          <chr>                 <int>
##  1 casual        00                    42321
##  2 member        00                    25237
##  3 casual        01                    30667
##  4 member        01                    16362
##  5 casual        02                    19579
##  6 member        02                     8976
##  7 casual        03                    10220
##  8 member        03                     5104
##  9 casual        04                     6688
## 10 member        04                     5996
## # ... with 38 more rows
```

```r
# Get the top 10 start and end stations for casual riders
top_start_station <- clean_data %>%
  group_by(member_casual = "Casual",start_station_name) %>%
  summarise(station_count = n()) %>%
  arrange(desc(station_count)) %>%
  slice(1:10)
```

```
## 'summarise()' has grouped output by 'member_casual'. You can override using the
## '.groups' argument.
```
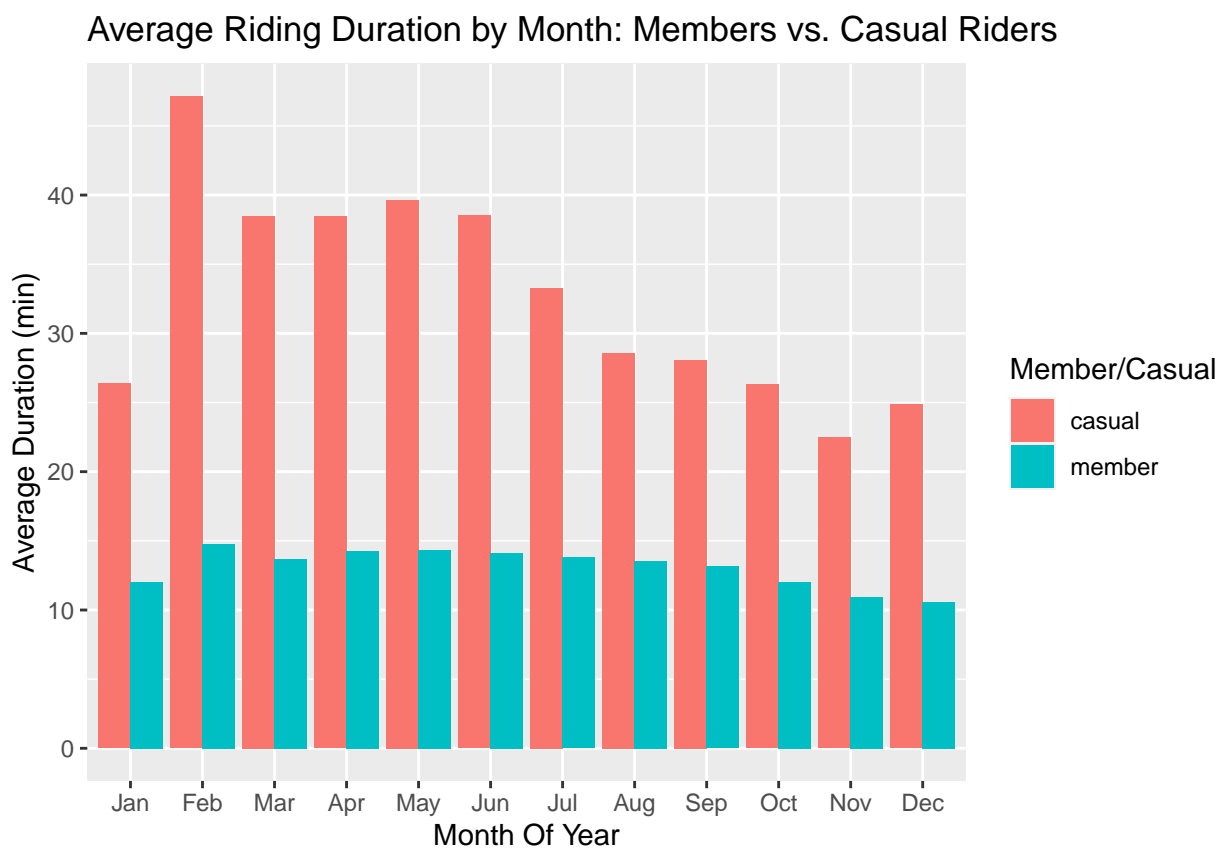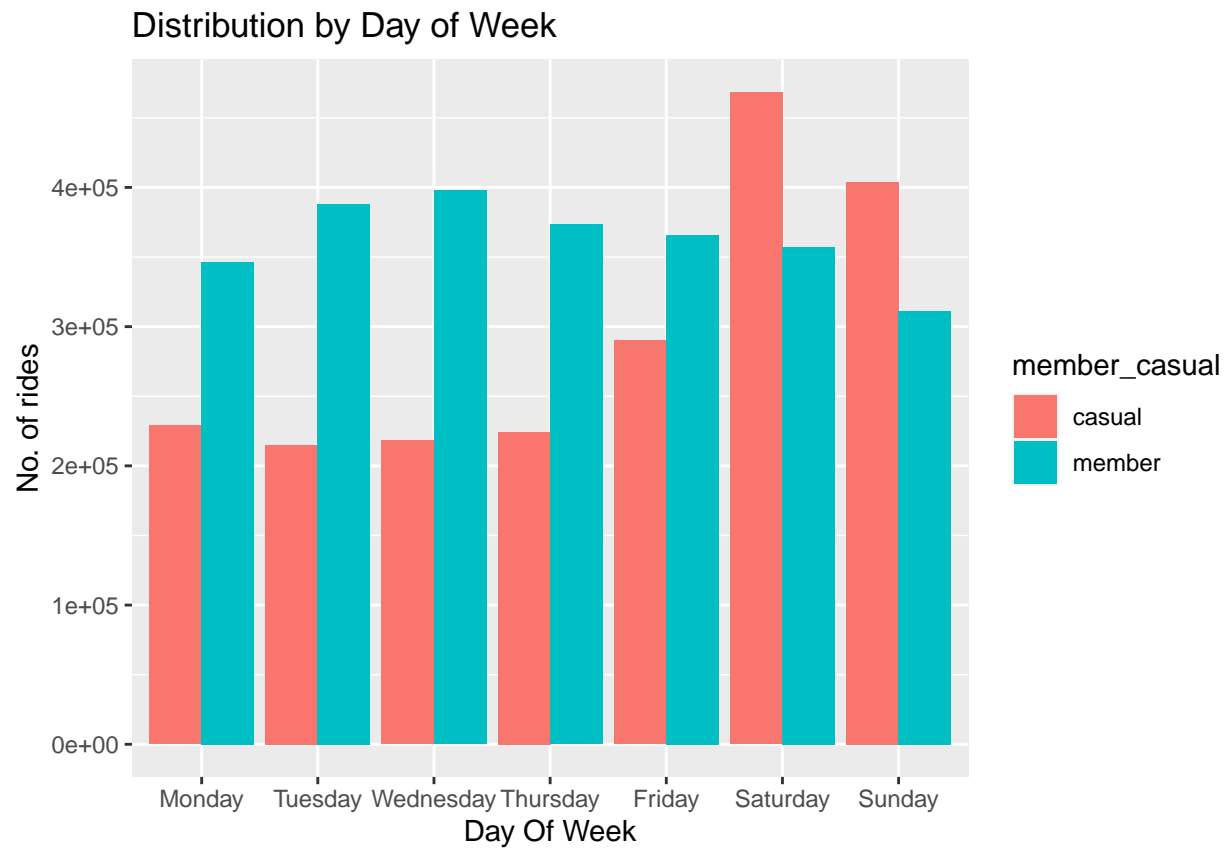
```r
top_end_station <- clean_data %>%
  group_by(member_casual = "Casual",end_station_name) %>%
  summarise(station_count = n()) %>%
  arrange(desc(station_count)) %>%
  slice(1:10)
```

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

Data Visualization

You can also embed plots, for example:
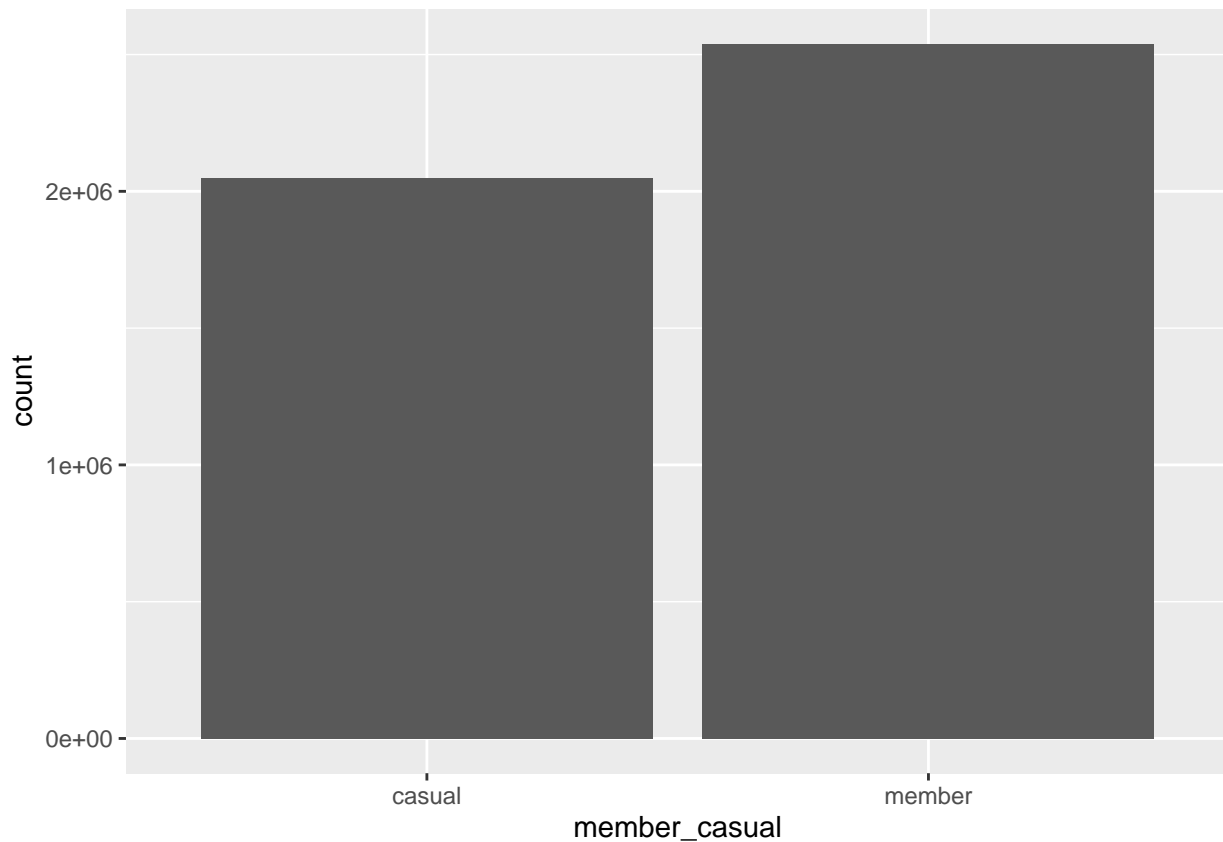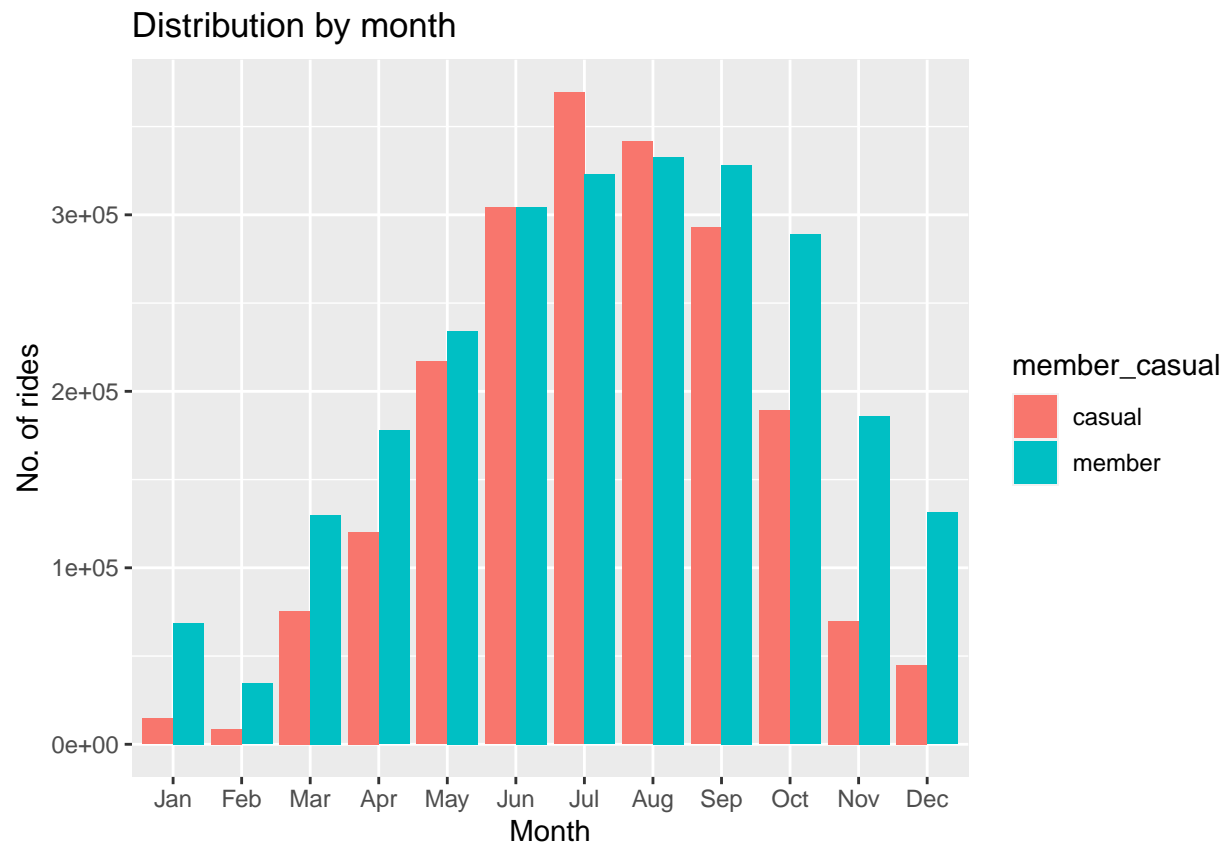
Distribution by month

Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.

Average Riding Duration by Month: Members vs. Casual Riders

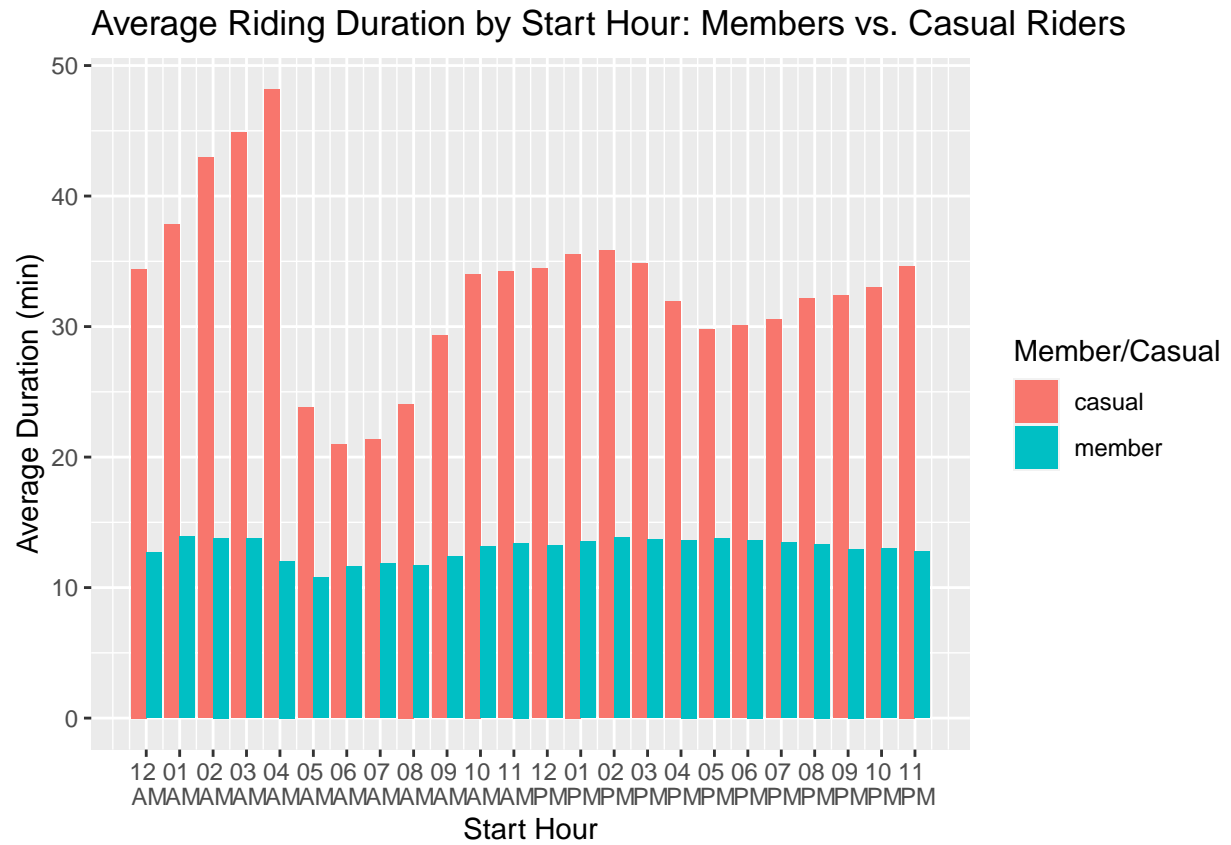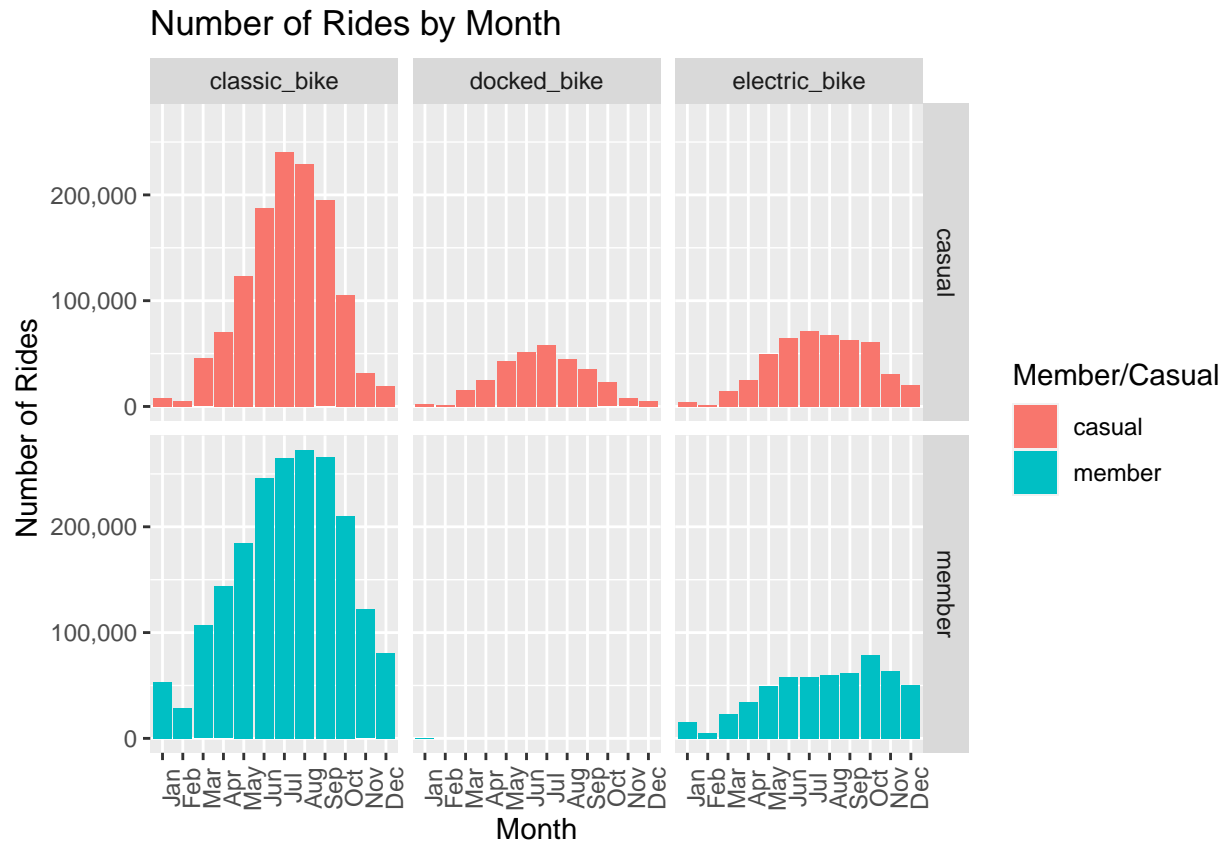## Distribution by Day of Week

Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
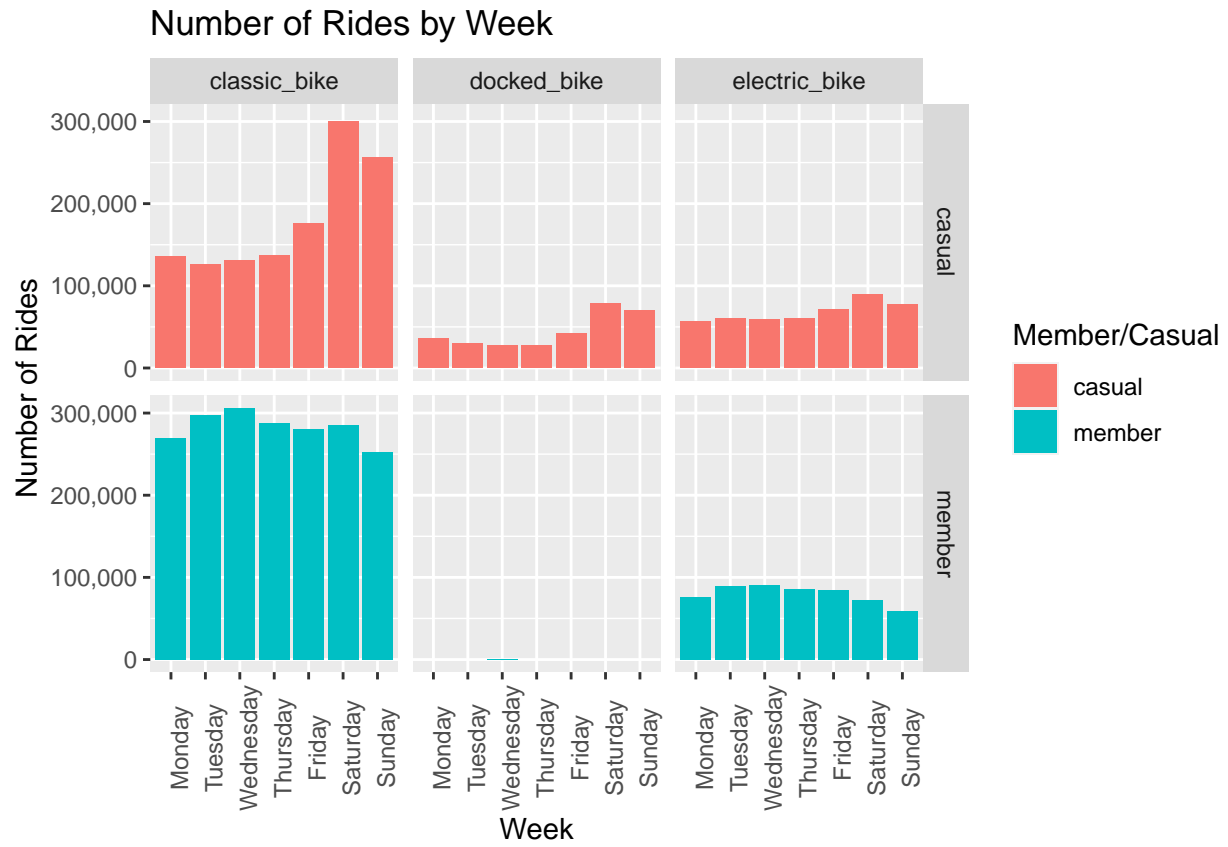
Average Riding Duration by Day: Members vs. Casual Riders
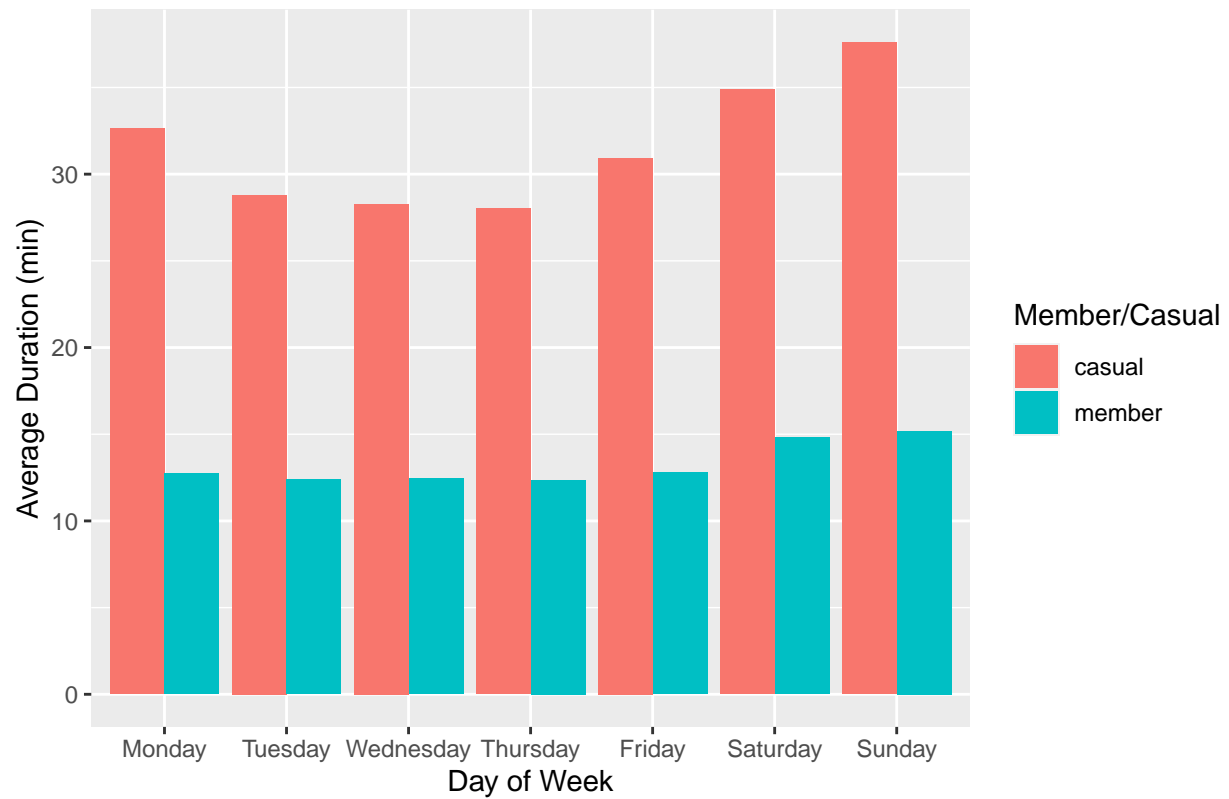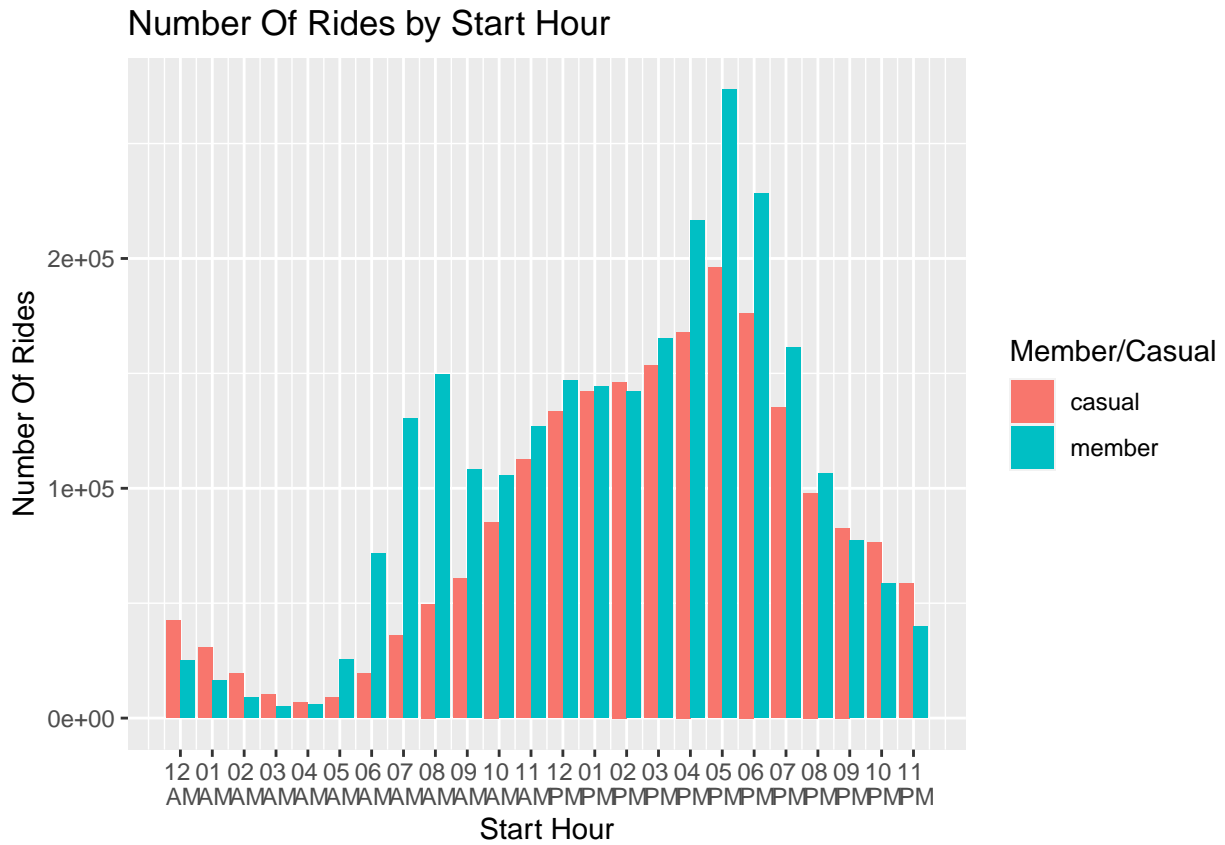
Number Of Rides by Start Hour


```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```

Average Riding Duration by Start Hour: Members vs. Casual Riders

Number of Rides on different cycles: Members vs. Casual Riders

Number of Rides by Month

Number of Rides by Week
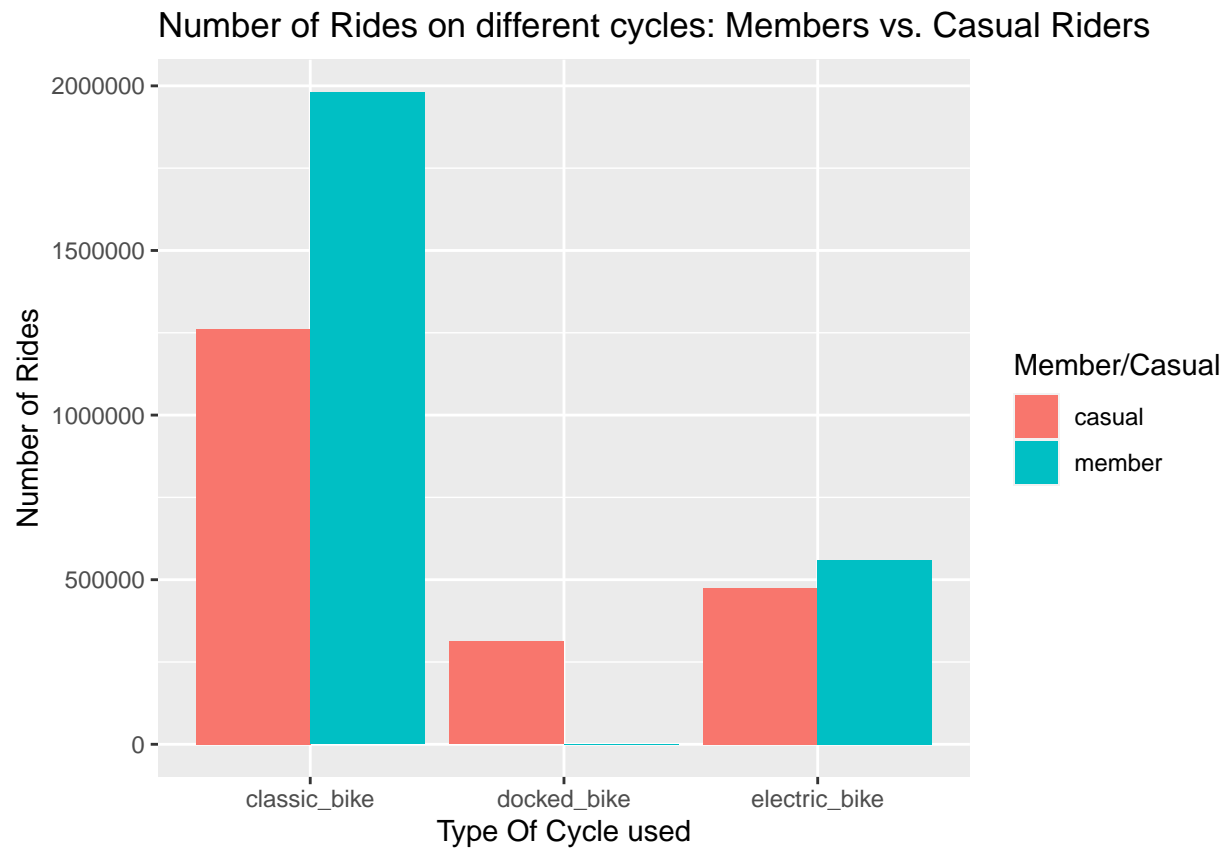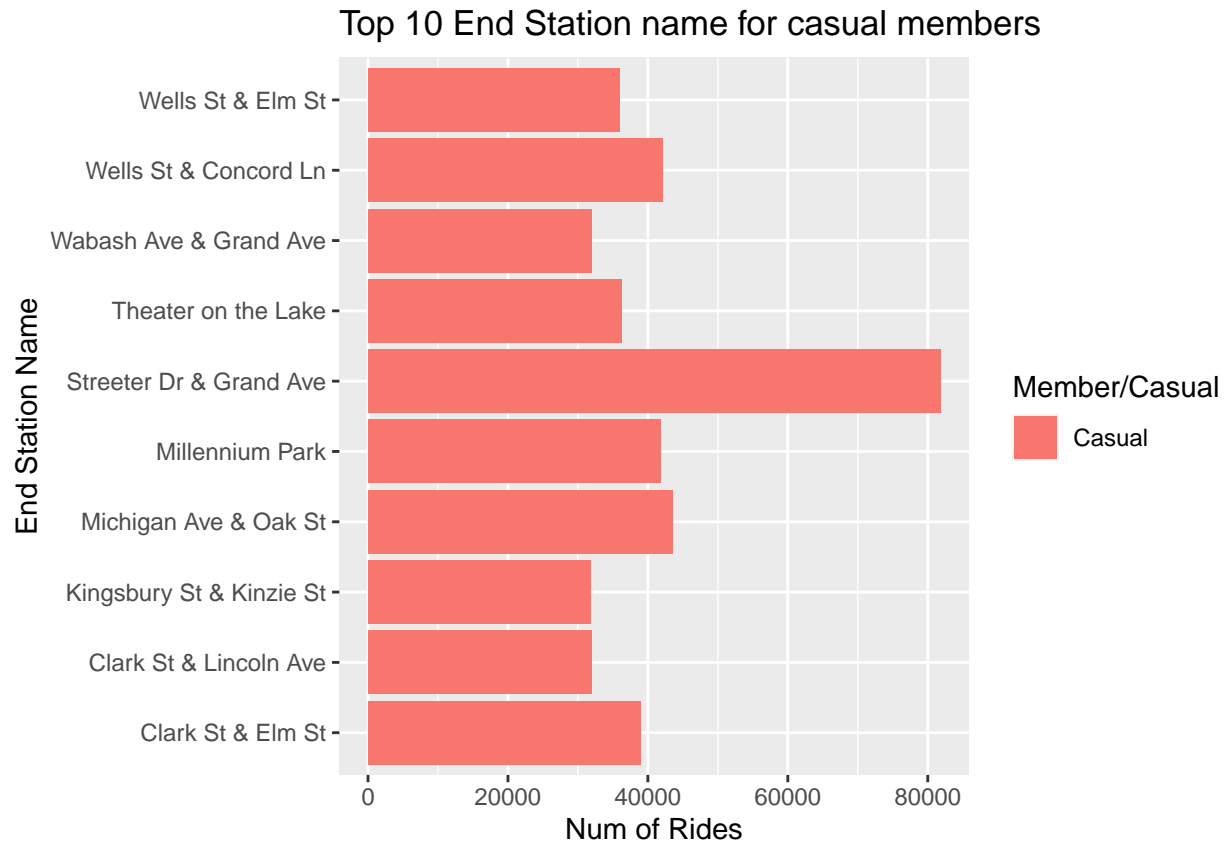
Top 10 Start Station name for casual members

Top 10 End Station name for casual members
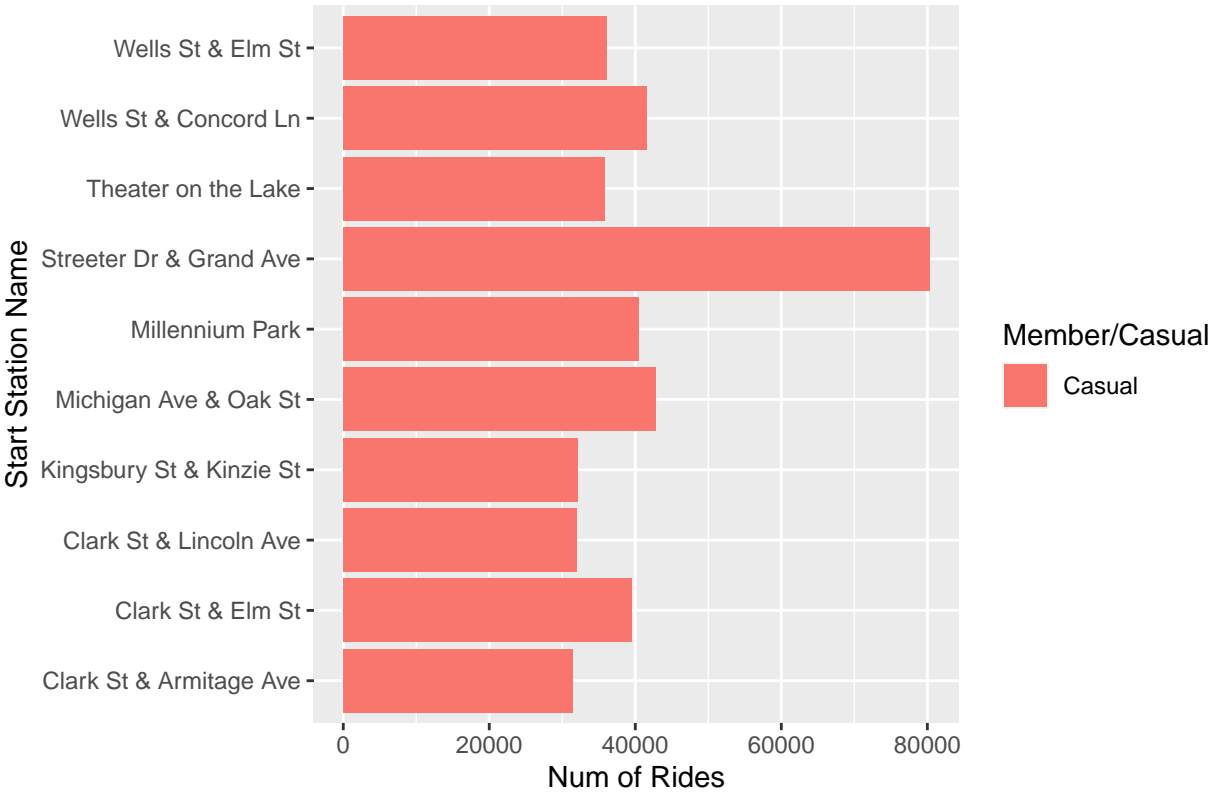
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.