```
# # # # # # # # # # # # # # # # # # # # # # #
# Install required packages
# tidyverse for data import and wrangling
# lubridate for date functions
# ggplot for visualization
# # # # # # # # # # # # # # # # # # # # # # #

> install.packages("tidyverse")
> install.packages("lubridate")
> install.packages("ggplot2")


library(tidyverse)  #helps wrangle data
library(lubridate)  #helps wrangle date attributes
library(ggplot2)    #helps visualize data

> getwd()          #displays your working directory
[1] "C:/Users/Dell/Desktop/bike_sharing_csv"
> setwd("C:/Users/Dell/Desktop/bike_sharing_csv")  #sets your working
directory to simplify calls to data

#=====================
# STEP 1: UPLOAD DATA
#=====================

# Upload Divvy datasets (csv files) here
```

```
> m1_2021 <- read.csv("202101-divvy-tripdata.csv")
> m2_2021 <- read.csv("202102-divvy-tripdata.csv")
> m3_2021 <- read.csv("202103-divvy-tripdata.csv")
> m4_2021 <- read.csv("202104-divvy-tripdata.csv")
> m5_2021 <- read.csv("202105-divvy-tripdata.csv")
> m6_2021 <- read.csv("202106-divvy-tripdata.csv")
> m7_2021 <- read.csv("202107-divvy-tripdata.csv")
> m8_2021 <- read.csv("202108-divvy-tripdata.csv")
> m9_2021 <- read.csv("202109-divvy-tripdata.csv")
> m10_2021 <- read.csv("202110-divvy-tripdata.csv")
> m11_2021 <- read.csv("202111-divvy-tripdata.csv")
> m12_2021 <- read.csv("202112-divvy-tripdata.csv")

#=====================================================
# STEP 2: WRANGLE DATA AND COMBINE INTO A SINGLE FILE
#=====================================================

# Compare column names each of the files
# While the names don't have to be in the same order, they do need to match perfectly before we can use a command to join them into one file

> colnames(m1_2021)
 [1] "ride_id"           "rideable_type"     "started_at"        "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"
"end_station_id"
 [9] "start_lat"         "start_lng"         "end_lat"           "end_lng"
[13] "member_casual"
```

```
> colnames(m2_2021)
 [1] "ride_id"          "rideable_type"     "started_at"        "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"
"end_station_id"
 [9] "start_lat"         "start_lng"         "end_lat"           "end_lng"
[13] "member_casual"
> colnames(m3_2021)
 [1] "ride_id"          "rideable_type"     "started_at"        "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"
"end_station_id"
 [9] "start_lat"         "start_lng"         "end_lat"           "end_lng"
[13] "member_casual"
> colnames(m4_2021)
 [1] "ride_id"          "rideable_type"     "started_at"        "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"
"end_station_id"
 [9] "start_lat"         "start_lng"         "end_lat"           "end_lng"
[13] "member_casual"
> colnames(m5_2021)
 [1] "ride_id"          "rideable_type"     "started_at"        "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"
"end_station_id"
 [9] "start_lat"         "start_lng"         "end_lat"           "end_lng"
[13] "member_casual"
> colnames(m6_2021)
 [1] "ride_id"          "rideable_type"     "started_at"        "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"
"end_station_id"
```

[9] "start_lat" "start_lng" "end_lat" "end_lng"

[13] "member_casual"

> colnames(m7_2021)

[1] "ride_id" "rideable_type" "started_at" "ended_at"

[5] "start_station_name" "start_station_id" "end_station_name" "end_station_id"

[9] "start_lat" "start_lng" "end_lat" "end_lng"

[13] "member_casual"

> colnames(m8_2021)

[1] "ride_id" "rideable_type" "started_at" "ended_at"

[5] "start_station_name" "start_station_id" "end_station_name" "end_station_id"

[9] "start_lat" "start_lng" "end_lat" "end_lng"

[13] "member_casual"

> colnames(m9_2021)

[1] "ride_id" "rideable_type" "started_at" "ended_at"

[5] "start_station_name" "start_station_id" "end_station_name" "end_station_id"

[9] "start_lat" "start_lng" "end_lat" "end_lng"

[13] "member_casual"

> colnames(m10_2021)

[1] "ride_id" "rideable_type" "started_at" "ended_at"

[5] "start_station_name" "start_station_id" "end_station_name" "end_station_id"

[9] "start_lat" "start_lng" "end_lat" "end_lng"

[13] "member_casual"

> colnames(m11_2021)

```
 [1] "ride_id"          "rideable_type"     "started_at"        "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"
"end_station_id"
 [9] "start_lat"        "start_lng"         "end_lat"           "end_lng"
[13] "member_casual"
> colnames(m12_2021)
 [1] "ride_id"          "rideable_type"     "started_at"        "ended_at"
 [5] "start_station_name" "start_station_id"   "end_station_name"
"end_station_id"
 [9] "start_lat"        "start_lng"         "end_lat"           "end_lng"
[13] "member_casual"
```

Inspect the dataframes and look for incongruencies

```
> str(m1_2021)
'data.frame':         96834 obs. of  13 variables:
 $ ride_id          : chr  "E19E6F1B8D4C42ED" "DC88F20C2C55F27F"
"EC45C94683FE3F27" "4FA453A75AE377DB" ...
 $ rideable_type    : chr  "electric_bike" "electric_bike" "electric_bike"
"electric_bike" ...
 $ started_at       : chr  "2021-01-23 16:14:19" "2021-01-27 18:43:08" "2021-
01-21 22:35:54" "2021-01-07 13:31:13" ...
 $ ended_at         : chr  "2021-01-23 16:24:44" "2021-01-27 18:47:12" "2021-
01-21 22:37:14" "2021-01-07 13:42:55" ...
 $ start_station_name: chr  "California Ave & Cortez St" "California Ave &
Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" ...
 $ start_station_id  : chr  "17660" "17660" "17660" "17660" ...
 $ end_station_name  : chr  "" "" "" "" ...
 $ end_station_id    : chr  "" "" "" "" ...
```

$ start_lat : num 41.9 41.9 41.9 41.9 41.9 ...

$ start_lng : num -87.7 -87.7 -87.7 -87.7 -87.7 ...

$ end_lat : num 41.9 41.9 41.9 41.9 41.9 ...

$ end_lng : num -87.7 -87.7 -87.7 -87.7 -87.7 ...

$ member_casual : chr "member" "member" "member" "member" ...

> str(m2_2021)

'data.frame': 49622 obs. of 13 variables:

$ ride_id : chr "89E7AA6C29227EFF" "0FEFDE2603568365" "E6159D746B2DBB91" "B32D3199F1C2E75B" ...

$ rideable_type : chr "classic_bike" "classic_bike" "electric_bike" "classic_bike" ...

$ started_at : chr "2021-02-12 16:14:56" "2021-02-14 17:52:38" "2021-02-09 19:10:18" "2021-02-02 17:49:41" ...

$ ended_at : chr "2021-02-12 16:21:43" "2021-02-14 18:12:09" "2021-02-09 19:19:10" "2021-02-02 17:54:06" ...

$ start_station_name: chr "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Clark St & Lake St" "Wood St & Chicago Ave" ...

$ start_station_id : chr "525" "525" "KA1503000012" "637" ...

$ end_station_name : chr "Sheridan Rd & Columbia Ave" "Bosworth Ave & Howard St" "State St & Randolph St" "Honore St & Division St" ...

$ end_station_id : chr "660" "16806" "TA1305000029" "TA1305000034" ...

$ start_lat : num 42 42 41.9 41.9 41.8 ...

$ start_lng : num -87.7 -87.7 -87.6 -87.7 -87.6 ...

$ end_lat : num 42 42 41.9 41.9 41.8 ...

$ end_lng : num -87.7 -87.7 -87.6 -87.7 -87.6 ...

$ member_casual : chr "member" "casual" "member" "member" ...

> str(m3_2021)

'data.frame': 228496 obs. of 13 variables:

$ ride_id : chr "CFA86D4455AA1030" "30D9DC61227D1AF3" "846D87A15682A284" "994D05AA75A168F2" ...

$ rideable_type : chr "classic_bike" "classic_bike" "classic_bike" "classic_bike" ...

$ started_at : chr "2021-03-16 08:32:30" "2021-03-28 01:26:28" "2021-03-11 21:17:29" "2021-03-11 13:26:42" ...

$ ended_at : chr "2021-03-16 08:36:34" "2021-03-28 01:36:55" "2021-03-11 21:33:53" "2021-03-11 13:55:41" ...

$ start_station_name: chr "Humboldt Blvd & Armitage Ave" "Humboldt Blvd & Armitage Ave" "Shields Ave & 28th Pl" "Winthrop Ave & Lawrence Ave" ...

$ start_station_id : chr "15651" "15651" "15443" "TA1308000021" ...

$ end_station_name : chr "Stave St & Armitage Ave" "Central Park Ave & Bloomingdale Ave" "Halsted St & 35th St" "Broadway & Sheridan Rd" ...

$ end_station_id : chr "13266" "18017" "TA1308000043" "13323" ...

$ start_lat : num 41.9 41.9 41.8 42 42 ...

$ start_lng : num -87.7 -87.7 -87.6 -87.7 -87.7 ...

$ end_lat : num 41.9 41.9 41.8 42 42.1 ...

$ end_lng : num -87.7 -87.7 -87.6 -87.6 -87.7 ...

$ member_casual : chr "casual" "casual" "casual" "casual" ...

> str(m4_2021)

'data.frame': 337230 obs. of 13 variables:

$ ride_id : chr "6C992BD37A98A63F" "1E0145613A209000" "E498E15508A80BAD" "1887262AD101C604" ...

$ rideable_type : chr "classic_bike" "docked_bike" "docked_bike" "classic_bike" ...

$ started_at : chr "2021-04-12 18:25:36" "2021-04-27 17:27:11" "2021-04-03 12:42:45" "2021-04-17 09:17:42" ...

$ ended_at : chr "2021-04-12 18:56:55" "2021-04-27 18:31:29" "2021-04-07 11:40:24" "2021-04-17 09:42:48" ...

$ start_station_name: chr "State St & Pearson St" "Dorchester Ave & 49th St" "Loomis Blvd & 84th St" "Honore St & Division St" ...

$ start_station_id : chr "TA1307000061" "KA1503000069" "20121" "TA1305000034" ...

$ end_station_name : chr "Southport Ave & Waveland Ave" "Dorchester Ave & 49th St" "Loomis Blvd & 84th St" "Southport Ave & Waveland Ave" ...

$ end_station_id : chr "13235" "KA1503000069" "20121" "13235" ...

$ start_lat : num 41.9 41.8 41.7 41.9 41.7 ...

$ start_lng : num -87.6 -87.6 -87.7 -87.7 -87.7 ...

$ end_lat : num 41.9 41.8 41.7 41.9 41.7 ...

$ end_lng : num -87.7 -87.6 -87.7 -87.7 -87.7 ...

$ member_casual : chr "member" "casual" "casual" "member" ...

> str(m5_2021)

'data.frame': 531633 obs. of 13 variables:

$ ride_id : chr "C809ED75D6160B2A" "DD59FDCE0ACACAF3" "0AB83CB88C43EFC2" "7881AC6D39110C60" ...

$ rideable_type : chr "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...

$ started_at : chr "2021-05-30 11:58:15" "2021-05-30 11:29:14" "2021-05-30 14:24:01" "2021-05-30 14:25:51" ...

$ ended_at : chr "2021-05-30 12:10:39" "2021-05-30 12:14:09" "2021-05-30 14:25:13" "2021-05-30 14:41:04" ...

$ start_station_name: chr "" "" "" "" ...

$ start_station_id : chr "" "" "" "" ...

$ end_station_name : chr "" "" "" "" ...

$ end_station_id : chr "" "" "" "" ...

$ start_lat : num 41.9 41.9 41.9 41.9 41.9 ...

$ start_lng : num -87.6 -87.6 -87.7 -87.7 -87.7 ...

$ end_lat : num 41.9 41.8 41.9 41.9 41.9 ...

$ end_lng : num -87.6 -87.6 -87.7 -87.7 -87.7 ...

$ member_casual : chr "casual" "casual" "casual" "casual" ...

> str(m6_2021)

'data.frame': 729595 obs. of 13 variables:

$ ride_id : chr "99FEC93BA843FB20" "06048DCFC8520CAF" "9598066F68045DF2" "B03C0FE48C412214" ...

$ rideable_type : chr "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...

$ started_at : chr "2021-06-13 14:31:28" "2021-06-04 11:18:02" "2021-06-04 09:49:35" "2021-06-03 19:56:05" ...

$ ended_at : chr "2021-06-13 14:34:11" "2021-06-04 11:24:19" "2021-06-04 09:55:34" "2021-06-03 20:21:55" ...

$ start_station_name: chr "" "" "" "" ...

$ start_station_id : chr "" "" "" "" ...

$ end_station_name : chr "" "" "" "" ...

$ end_station_id : chr "" "" "" "" ...

$ start_lat : num 41.8 41.8 41.8 41.8 41.8 ...

$ start_lng : num -87.6 -87.6 -87.6 -87.6 -87.6 ...

$ end_lat : num 41.8 41.8 41.8 41.8 41.8 ...

$ end_lng : num -87.6 -87.6 -87.6 -87.6 -87.6 ...

$ member_casual : chr "member" "member" "member" "member" ...

> str(m7_2021)

'data.frame': 822410 obs. of 13 variables:

$ ride_id : chr "0A1B623926EF4E16" "B2D5583A5A5E76EE" "6F264597DDBF427A" "379B58EAB20E8AA5" ...

$ rideable_type : chr "docked_bike" "classic_bike" "classic_bike" "classic_bike" ...

$ started_at : chr "2021-07-02 14:44:36" "2021-07-07 16:57:42" "2021-07-25 11:30:55" "2021-07-08 22:08:30" ...

$ ended_at : chr "2021-07-02 15:19:58" "2021-07-07 17:16:09" "2021-07-25 11:48:45" "2021-07-08 22:23:32" ...

$ start_station_name: chr "Michigan Ave & Washington St" "California Ave & Cortez St" "Wabash Ave & 16th St" "California Ave & Cortez St" ...

$ start_station_id : chr "13001" "17660" "SL-012" "17660" ...

$ end_station_name : chr "Halsted St & North Branch St" "Wood St & Hubbard St" "Rush St & Hubbard St" "Carpenter St & Huron St" ...

$ end_station_id : chr "KA1504000117" "13432" "KA1503000044" "13196" ...

$ start_lat : num 41.9 41.9 41.9 41.9 41.9 ...

$ start_lng : num -87.6 -87.7 -87.6 -87.7 -87.7 ...

$ end_lat : num 41.9 41.9 41.9 41.9 41.9 ...

$ end_lng : num -87.6 -87.7 -87.6 -87.7 -87.7 ...

$ member_casual : chr "casual" "casual" "member" "member" ...

> str(m8_2021)

'data.frame': 804352 obs. of 13 variables:

$ ride_id : chr "99103BB87CC6C1BB" "EAFCCCFB0A3FC5A1" "9EF4F46C57AD234D" "5834D3208BFAF1DA" ...

$ rideable_type : chr "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...

$ started_at : chr "2021-08-10 17:15:49" "2021-08-10 17:23:14" "2021-08-21 02:34:23" "2021-08-21 06:52:55" ...

$ ended_at : chr "2021-08-10 17:22:44" "2021-08-10 17:39:24" "2021-08-21 02:50:36" "2021-08-21 07:08:13" ...

$ start_station_name: chr "" "" "" "" ...

$ start_station_id : chr "" "" "" "" ...

$ end_station_name : chr "" "" "" "" ...

$ end_station_id : chr "" "" "" "" ...

$ start_lat : num 41.8 41.8 42 42 41.8 ...

$ start_lng : num -87.7 -87.7 -87.7 -87.7 -87.6 ...

$ end_lat : num 41.8 41.8 42 42 41.8 ...

$ end_lng : num -87.7 -87.6 -87.7 -87.7 -87.6 ...

$ member_casual : chr "member" "member" "member" "member" ...

> str(m9_2021)

'data.frame': 756147 obs. of 13 variables:

$ ride_id : chr "9DC7B962304CBFD8" "F930E2C6872D6B32" "6EF72137900BB910" "78D1DE133B3DBF55" ...

$ rideable_type : chr "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...

$ started_at : chr "2021-09-28 16:07:10" "2021-09-28 14:24:51" "2021-09-28 00:20:16" "2021-09-28 14:51:17" ...

$ ended_at : chr "2021-09-28 16:09:54" "2021-09-28 14:40:05" "2021-09-28 00:23:57" "2021-09-28 15:00:06" ...

$ start_station_name: chr "" "" "" "" ...

$ start_station_id : chr "" "" "" "" ...

$ end_station_name : chr "" "" "" "" ...

$ end_station_id : chr "" "" "" "" ...

$ start_lat : num 41.9 41.9 41.8 41.8 41.9 ...

$ start_lng : num -87.7 -87.6 -87.7 -87.7 -87.7 ...

$ end_lat : num 41.9 42 41.8 41.8 41.9 ...

$ end_lng : num -87.7 -87.7 -87.7 -87.7 -87.7 ...

$ member_casual : chr "casual" "casual" "casual" "casual" ...

> str(m10_2021)

'data.frame': 631226 obs. of 13 variables:

$ ride_id : chr "620BC6107255BF4C" "4471C70731AB2E45" "26CA69D43D15EE14" "362947F0437E1514" ...

$ rideable_type : chr "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...

$ started_at : chr "2021-10-22 12:46:42" "2021-10-21 09:12:37" "2021-10-16 16:28:39" "2021-10-16 16:17:48" ...

$ ended_at : chr "2021-10-22 12:49:50" "2021-10-21 09:14:14" "2021-10-16 16:36:26" "2021-10-16 16:19:03" ...

$ start_station_name: chr "Kingsbury St & Kinzie St" "" "" "" ...

$ start_station_id : chr "KA1503000043" "" "" "" ...

$ end_station_name : chr "" "" "" "" ...

$ end_station_id : chr "" "" "" "" ...

$ start_lat : num 41.9 41.9 41.9 41.9 41.9 ...

$ start_lng : num -87.6 -87.7 -87.7 -87.7 -87.7 ...

$ end_lat : num 41.9 41.9 41.9 41.9 41.9 ...

$ end_lng : num -87.6 -87.7 -87.7 -87.7 -87.7 ...

$ member_casual : chr "member" "member" "member" "member" ...

> str(m11_2021)

'data.frame': 359978 obs. of 13 variables:

$ ride_id : chr "7C00A93E10556E47" "90854840DFD508BA" "0A7D10CDD144061C" "2F3BE33085BCFF02" ...

$ rideable_type : chr "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...

$ started_at : chr "2021-11-27 13:27:38" "2021-11-27 13:38:25" "2021-11-26 22:03:34" "2021-11-27 09:56:49" ...

$ ended_at : chr "2021-11-27 13:46:38" "2021-11-27 13:56:10" "2021-11-26 22:05:56" "2021-11-27 10:01:50" ...

$ start_station_name: chr "" "" "" "" ...

$ start_station_id : chr "" "" "" "" ...

$ end_station_name : chr "" "" "" "" ...

$ end_station_id   : chr "" "" "" "" ...

$ start_lat   : num 41.9 42 42 41.9 41.9 ...

$ start_lng   : num -87.7 -87.7 -87.7 -87.8 -87.6 ...

$ end_lat   : num 42 41.9 42 41.9 41.9 ...

$ end_lng   : num -87.7 -87.7 -87.7 -87.8 -87.6 ...

$ member_casual   : chr "casual" "casual" "casual" "casual" ...

> str(m12_2021)

'data.frame':   247540 obs. of 13 variables:

$ ride_id   : chr "46F8167220E4431F" "73A77762838B32FD" "4CF42452054F59C5" "3278BA87BF698339" ...

$ rideable_type   : chr "electric_bike" "electric_bike" "electric_bike" "classic_bike" ...

$ started_at   : chr "2021-12-07 15:06:07" "2021-12-11 03:43:29" "2021-12-15 23:10:28" "2021-12-26 16:16:10" ...

$ ended_at   : chr "2021-12-07 15:13:42" "2021-12-11 04:10:23" "2021-12-15 23:23:14" "2021-12-26 16:30:53" ...

$ start_station_name: chr "Laflin St & Cullerton St" "LaSalle Dr & Huron St" "Halsted St & North Branch St" "Halsted St & North Branch St" ...

$ start_station_id : chr "13307" "KP1705001026" "KA1504000117" "KA1504000117" ...

$ end_station_name : chr "Morgan St & Polk St" "Clarendon Ave & Leland Ave" "Broadway & Barry Ave" "LaSalle Dr & Huron St" ...

$ end_station_id   : chr "TA1307000130" "TA1307000119" "13137" "KP1705001026" ...

$ start_lat   : num 41.9 41.9 41.9 41.9 41.9 ...

$ start_lng   : num -87.7 -87.6 -87.6 -87.6 -87.7 ...

$ end_lat   : num 41.9 42 41.9 41.9 41.9 ...

$ end_lng   : num -87.7 -87.7 -87.6 -87.6 -87.6 ...

```
 $ member_casual     : chr  "member" "casual" "member" "member" ...
```

```
# Convert ride_id and rideable_type to character so that they can stack correctly
> m1_2021 <-  mutate(m1_2021, ride_id = as.character(ride_id)
+                    ,rideable_type = as.character(rideable_type))
> m2_2021 <-  mutate(m2_2021, ride_id = as.character(ride_id)
+                    ,rideable_type = as.character(rideable_type))
> m3_2021 <-  mutate(m3_2021, ride_id = as.character(ride_id)
+                    ,rideable_type = as.character(rideable_type))
> m4_2021 <-  mutate(m4_2021, ride_id = as.character(ride_id)
+                    ,rideable_type = as.character(rideable_type))
> m5_2021 <-  mutate(m5_2021, ride_id = as.character(ride_id)
+                    ,rideable_type = as.character(rideable_type))
> m6_2021 <-  mutate(m6_2021, ride_id = as.character(ride_id)
+                    ,rideable_type = as.character(rideable_type))
> m7_2021 <-  mutate(m7_2021, ride_id = as.character(ride_id)
+                    ,rideable_type = as.character(rideable_type))
> m8_2021 <-  mutate(m8_2021, ride_id = as.character(ride_id)
+                    ,rideable_type = as.character(rideable_type))
> m9_2021 <-  mutate(m9_2021, ride_id = as.character(ride_id)
+                    ,rideable_type = as.character(rideable_type))
> m10_2021 <-  mutate(m10_2021, ride_id = as.character(ride_id)
+                     ,rideable_type = as.character(rideable_type))
> m11_2021 <-  mutate(m11_2021, ride_id = as.character(ride_id)
+                     ,rideable_type = as.character(rideable_type))
```

```
> m12_2021 <-  mutate(m12_2021, ride_id = as.character(ride_id)
+                     ,rideable_type = as.character(rideable_type))


# Stack individual quarter's data frames into one big data frame
> all_rides <-
bind_rows(m1_2021,m2_2021,m3_2021,m4_2021,m5_2021,m6_2021,
+                       m7_2021,m8_2021,m9_2021,m10_2021,m11_2021,m12_2021)


# Remove lat, long fields
> all_rides <- all_rides %>%
+   select(-c(start_lat,start_lng,end_lat,end_lng))


#=====================================================
# STEP 3: CLEAN UP AND ADD DATA TO PREPARE FOR ANALYSIS
#=====================================================

# Inspect the new table that has been created

> colnames(all_rides)    #List of column names
[1] "ride_id"           "rideable_type"     "started_at"        "ended_at"
[5] "start_station_name" "start_station_id"   "end_station_name"
"end_station_id"
[9] "member_casual"
```

```
> nrow(all_rides)       #How many rows are in data frame?
[1] 5595063

> dim(all_rides)        #Dimensions of the data frame?
[1] 5595063       9

> head(all_rides)       #See the first 6 rows of data frame
          ride_id rideable_type         started_at           ended_at
start_station_name
1 E19E6F1B8D4C42ED electric_bike 2021-01-23 16:14:19 2021-01-23 16:24:44
California Ave & Cortez St
2 DC88F20C2C55F27F electric_bike 2021-01-27 18:43:08 2021-01-27 18:47:12
California Ave & Cortez St
3 EC45C94683FE3F27 electric_bike 2021-01-21 22:35:54 2021-01-21 22:37:14
California Ave & Cortez St
4 4FA453A75AE377DB electric_bike 2021-01-07 13:31:13 2021-01-07 13:42:55
California Ave & Cortez St
5 BE5E8EB4E7263A0B electric_bike 2021-01-23 02:24:02 2021-01-23 02:24:45
California Ave & Cortez St
6 5D8969F88C773979 electric_bike 2021-01-09 14:24:07 2021-01-09 15:17:54
California Ave & Cortez St
  start_station_id end_station_name end_station_id member_casual
1            17660                                        member
2            17660                                        member
3            17660                                        member
4            17660                                        member
5            17660                                        casual
6            17660                                        casual
```

```
> str(all_rides)       #See list of columns and data types (numeric, character, etc)
'data.frame':          5595063 obs. of  9 variables:
 $ ride_id          : chr  "E19E6F1B8D4C42ED" "DC88F20C2C55F27F" "EC45C94683FE3F27" "4FA453A75AE377DB" ...
 $ rideable_type    : chr  "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...
 $ started_at       : chr  "2021-01-23 16:14:19" "2021-01-27 18:43:08" "2021-01-21 22:35:54" "2021-01-07 13:31:13" ...
 $ ended_at         : chr  "2021-01-23 16:24:44" "2021-01-27 18:47:12" "2021-01-21 22:37:14" "2021-01-07 13:42:55" ...
 $ start_station_name: chr  "California Ave & Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" ...
 $ start_station_id  : chr  "17660" "17660" "17660" "17660" ...
 $ end_station_name  : chr  "" "" "" "" ...
 $ end_station_id    : chr  "" "" "" "" ...
 $ member_casual     : chr  "member" "member" "member" "member" ...

> summary(all_rides)      #Statistical summary of data. Mainly for numerics
   ride_id          rideable_type       started_at          ended_at       start_station_name
 Length:5595063     Length:5595063     Length:5595063     Length:5595063     Length:5595063
 Class :character   Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```
 start_station_id   end_station_name   end_station_id     member_casual
 Length:5595063     Length:5595063     Length:5595063     Length:5595063
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```
# Begin by seeing how many observations fall under each usertype
> table(all_rides$member_casual)
 casual  member
2529005 3066058
```

```
# Add columns that list the date, month, day, and year of each ride
# This will allow us to aggregate ride data for each month, day, or year
> all_rides$date <- as.Date(all_rides$started_at)          #The default format is yyyy-mm-dd
> all_rides$month <- format(as.Date(all_rides$date), "%m")
> all_rides$day <- format(as.Date(all_rides$date), "%d")
> all_rides$year <- format(as.Date(all_rides$date), "%Y")
> all_rides$day_of_week <- format(as.Date(all_rides$date), "%A")
```

```
# Add a "ride_length" calculation to all_trips (in seconds)
> all_rides$ride_length <- difftime(all_rides$ended_at,all_rides$started_at)
```

```
# Inspect the structure of the columns
> str(all_rides)
```

'data.frame': 5595063 obs. of 15 variables:

$ ride_id : chr "E19E6F1B8D4C42ED" "DC88F20C2C55F27F" "EC45C94683FE3F27" "4FA453A75AE377DB" ...

$ rideable_type : chr "electric_bike" "electric_bike" "electric_bike" "electric_bike" ...

$ started_at : chr "2021-01-23 16:14:19" "2021-01-27 18:43:08" "2021-01-21 22:35:54" "2021-01-07 13:31:13" ...

$ ended_at : chr "2021-01-23 16:24:44" "2021-01-27 18:47:12" "2021-01-21 22:37:14" "2021-01-07 13:42:55" ...

$ start_station_name: chr "California Ave & Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" ...

$ start_station_id : chr "17660" "17660" "17660" "17660" ...

$ end_station_name : chr "" "" "" "" ...

$ end_station_id : chr "" "" "" "" ...

$ member_casual : chr "member" "member" "member" "member" ...

$ date : Date, format: "2021-01-23" "2021-01-27" "2021-01-21" ...

$ month : chr "01" "01" "01" "01" ...

$ day : chr "23" "27" "21" "07" ...

$ year : chr "2021" "2021" "2021" "2021" ...

$ day_of_week : chr "Saturday" "Wednesday" "Thursday" "Thursday" ...

$ ride_length : 'difftime' num 625 244 80 702 ...

Convert "ride_length" from Factor to numeric so we can run calculations on the data

>is.factor(all_trips$ride_length)

> all_rides$ride_length <- as.numeric(as.character(all_rides$ride_length))

> is.numeric(all_rides$ride_length)

```
[1] TRUE
```

```
# Remove "bad" data

# We will create a new version of the dataframe (total_rides) since data is being removed

> total_rides <- all_rides[!(all_rides$start_station_name == "HQ QR" | all_rides$ride_length<0),]
```

```
#=====================================

# STEP 4: CONDUCT DESCRIPTIVE ANALYSIS

#=====================================

# Descriptive analysis on ride_length

> mean(total_rides$ride_length)

[1] 1316.18

> median(total_rides$ride_length)

[1] 720

> max(total_rides$ride_length)

[1] 3356649

> min(total_rides$ride_length)

[1] 0
```

```
# You can condense the four lines above to one line using summary() on the specific attribute
> summary(total_rides$ride_length)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0     405     720    1316    1307 3356649


#=============================================
# STEP 5: EXPORT SUMMARY FILE FOR FURTHER ANALYSIS
#=============================================

# Create a csv file that we will visualize in Excel, Tableau, or my presentation software

write.csv(total_rides, "C:/Users/Dell/Desktop/bike_sharing_csv/data.csv")
```