Case_Study_1_Analysis

John E Case

2022-09-26

Divvy Bikeshare Data Analysis September 2021 to August 2022

The analysis steps are outlined below.

This analysis was completed as part of the Google Data Analytics Certificate. Credit to Kevin Hartman for the steps outlined below. Please see this link for additional information.

Download the data from the Divvy S3 bucket, using this link.

####Install and Load Packages Begin by installing the required R packages.

```
#install.packages("tidyverse")
#install.packages("lubridate")
#install.packages("ggplot2")
```

Then load the packages so you can use them.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

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

```
library(ggplot2)
```

####Import the data int RStudio Next, import your data files into R Studio, and assign each file to a data frame.

```
setwd("/Users/jcase/Documents/Bikeshare/Raw")
Sep_2021 <- read_csv("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.
```

```
Oct_2021 <- read_csv("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.
```

```
Nov_2021 <- read_csv("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.
```

```
Dec_2021 <- read_csv("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.
```

```r
Jan_2022 <- read_csv("202201-divvy-tripdata.csv")
```

```
## Rows: 103770 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
Feb_2022 <- read_csv("202202-divvy-tripdata.csv")
```

```
## Rows: 115609 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
Mar_2022 <- read_csv("202203-divvy-tripdata.csv")
```

```
## Rows: 284042 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
Apr_2022 <- read_csv("202204-divvy-tripdata.csv")
```

```
## Rows: 371249 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
May_2022 <- read_csv("202205-divvy-tripdata.csv")
```

```
## Rows: 634858 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.
```

```
Jun_2022 <- read_csv("202206-divvy-tripdata.csv")
```

```
## Rows: 769204 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.
```

```
Jul_2022 <- read_csv("202207-divvy-tripdata.csv")
```

```
## Rows: 823488 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.
```

```
Aug_2022 <- read_csv("202208-divvy-tripdata.csv")
```

```
## Rows: 785932 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.
```

####Inspect each column Then print out the column names for each data frame.

```
colnames(Sep_2021)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Oct_2021)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Nov_2021)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Dec_2021)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Jan_2022)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Feb_2022)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Mar_2022)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Apr_2022)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(May_2022)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Jun_2022)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Jul_2022)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```
colnames(Aug_2022)
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

Print the structure of each data frame for inspection.

```
str(Sep_2021)
```

```
## spec_tbl_df [756,147 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:756147] "9DC7B962304CBFD8" "F930E2C6872D6B32" "6EF72137900BB910" "78D1
##  $ rideable_type     : chr [1:756147] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:756147], format: "2021-09-28 16:07:10" "2021-09-28 14:24:51" ...
##  $ ended_at          : POSIXct[1:756147], format: "2021-09-28 16:09:54" "2021-09-28 14:40:05" ...
##  $ start_station_name: chr [1:756147] NA NA NA NA ...
##  $ start_station_id  : chr [1:756147] NA NA NA NA ...
##  $ end_station_name  : chr [1:756147] NA NA NA NA ...
##  $ end_station_id    : chr [1:756147] NA NA NA NA ...
##  $ start_lat         : num [1:756147] 41.9 41.9 41.8 41.8 41.9 ...
##  $ start_lng         : num [1:756147] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:756147] 41.9 42 41.8 41.8 41.9 ...
##  $ end_lng           : num [1:756147] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:756147] "casual" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Oct_2021)
```

```
## spec_tbl_df [631,226 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:631226] "620BC6107255BF4C" "4471C70731AB2E45" "26CA69D43D15EE14" "3629
##  $ rideable_type     : chr [1:631226] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:631226], format: "2021-10-22 12:46:42" "2021-10-21 09:12:37" ...
##  $ ended_at          : POSIXct[1:631226], format: "2021-10-22 12:49:50" "2021-10-21 09:14:14" ...
##  $ start_station_name: chr [1:631226] "Kingsbury St & Kinzie St" NA NA NA ...
##  $ start_station_id  : chr [1:631226] "KA1503000043" NA NA NA ...
##  $ end_station_name  : chr [1:631226] NA NA NA NA ...
##  $ end_station_id    : chr [1:631226] NA NA NA NA ...
##  $ start_lat         : num [1:631226] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:631226] -87.6 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:631226] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:631226] -87.6 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:631226] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
```

```
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Nov_2021)
```

```
## spec_tbl_df [359,978 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:359978] "7C00A93E10556E47" "90854840DFD508BA" "0A7D10CDD144061C" "2F3B
##  $ rideable_type     : chr [1:359978] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:359978], format: "2021-11-27 13:27:38" "2021-11-27 13:38:25" ...
##  $ ended_at          : POSIXct[1:359978], format: "2021-11-27 13:46:38" "2021-11-27 13:56:10" ...
##  $ start_station_name: chr [1:359978] NA NA NA NA ...
##  $ start_station_id  : chr [1:359978] NA NA NA NA ...
##  $ end_station_name  : chr [1:359978] NA NA NA NA ...
##  $ end_station_id    : chr [1:359978] NA NA NA NA ...
##  $ start_lat         : num [1:359978] 41.9 42 42 41.9 41.9 ...
##  $ start_lng         : num [1:359978] -87.7 -87.7 -87.7 -87.8 -87.6 ...
##  $ end_lat           : num [1:359978] 42 41.9 42 41.9 41.9 ...
##  $ end_lng           : num [1:359978] -87.7 -87.7 -87.7 -87.8 -87.6 ...
##  $ member_casual     : chr [1:359978] "casual" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Dec_2021)
```

```
## spec_tbl_df [247,540 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
```

```
##  $ ride_id           : chr [1:247540] "46F8167220E4431F" "73A77762838B32FD" "4CF42452054F59C5" "3278
##  $ rideable_type     : chr [1:247540] "electric_bike" "electric_bike" "electric_bike" "classic_bike"
##  $ started_at        : POSIXct[1:247540], format: "2021-12-07 15:06:07" "2021-12-11 03:43:29" ...
##  $ ended_at          : POSIXct[1:247540], format: "2021-12-07 15:13:42" "2021-12-11 04:10:23" ...
##  $ start_station_name: chr [1:247540] "Laflin St & Cullerton St" "LaSalle Dr & Huron St" "Halsted St
##  $ start_station_id  : chr [1:247540] "13307" "KP1705001026" "KA1504000117" "KA1504000117" ...
##  $ end_station_name  : chr [1:247540] "Morgan St & Polk St" "Clarendon Ave & Leland Ave" "Broadway &
##  $ end_station_id    : chr [1:247540] "TA1307000130" "TA1307000119" "13137" "KP1705001026" ...
##  $ start_lat         : num [1:247540] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:247540] -87.7 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:247540] 41.9 42 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:247540] -87.7 -87.7 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:247540] "member" "casual" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Jan_2022)
```

```
## spec_tbl_df [103,770 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:103770] "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB8
##  $ rideable_type     : chr [1:103770] "electric_bike" "electric_bike" "classic_bike" "classic_bike"
##  $ started_at        : POSIXct[1:103770], format: "2022-01-13 11:59:47" "2022-01-10 08:41:56" ...
##  $ ended_at          : POSIXct[1:103770], format: "2022-01-13 12:02:44" "2022-01-10 08:46:17" ...
##  $ start_station_name: chr [1:103770] "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffiel
##  $ start_station_id  : chr [1:103770] "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr [1:103770] "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave &
##  $ end_station_id    : chr [1:103770] "RP-007" "RP-007" "TA1307000001" "TA1309000021" ...
##  $ start_lat         : num [1:103770] 42 42 41.9 42 41.9 ...
##  $ start_lng         : num [1:103770] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num [1:103770] 42 42 41.9 42 41.9 ...
##  $ end_lng           : num [1:103770] -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr [1:103770] "casual" "casual" "member" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
```

```
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Feb_2022)
```

```
## spec_tbl_df [115,609 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:115609] "E1E065E7ED285C02" "1602DCDC5B30FFE3" "BE7DD2AF4B55C4AF" "A178
##  $ rideable_type     : chr [1:115609] "classic_bike" "classic_bike" "classic_bike" "classic_bike" ..
##  $ started_at        : POSIXct[1:115609], format: "2022-02-19 18:08:41" "2022-02-20 17:41:30" ...
##  $ ended_at          : POSIXct[1:115609], format: "2022-02-19 18:23:56" "2022-02-20 17:45:56" ...
##  $ start_station_name: chr [1:115609] "State St & Randolph St" "Halsted St & Wrightwood Ave" "State S
##  $ start_station_id  : chr [1:115609] "TA1305000029" "TA1309000061" "TA1305000029" "13235" ...
##  $ end_station_name  : chr [1:115609] "Clark St & Lincoln Ave" "Southport Ave & Wrightwood Ave" "Cana
##  $ end_station_id    : chr [1:115609] "13179" "TA1307000113" "13011" "13323" ...
##  $ start_lat         : num [1:115609] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:115609] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ end_lat           : num [1:115609] 41.9 41.9 41.9 42 41.9 ...
##  $ end_lng           : num [1:115609] -87.6 -87.7 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:115609] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Mar_2022)
```

```
## spec_tbl_df [284,042 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:284042] "47EC0A7F82E65D52" "8494861979B0F477" "EFE527AF80B66109" "9F44
##  $ rideable_type     : chr [1:284042] "classic_bike" "electric_bike" "classic_bike" "classic_bike" .
##  $ started_at        : POSIXct[1:284042], format: "2022-03-21 13:45:01" "2022-03-16 09:37:16" ...
##  $ ended_at          : POSIXct[1:284042], format: "2022-03-21 13:51:18" "2022-03-16 09:43:34" ...
##  $ start_station_name: chr [1:284042] "Wabash Ave & Wacker Pl" "Michigan Ave & Oak St" "Broadway & Be
```

```
##  $ start_station_id  : chr [1:284042] "TA1307000131" "13042" "13109" "TA1307000131" ...
##  $ end_station_name  : chr [1:284042] "Kingsbury St & Kinzie St" "Orleans St & Chestnut St (NEXT Apts
##  $ end_station_id    : chr [1:284042] "KA1503000043" "620" "15578" "TA1305000025" ...
##  $ start_lat         : num [1:284042] 41.9 41.9 42 41.9 41.9 ...
##  $ start_lng         : num [1:284042] -87.6 -87.6 -87.7 -87.6 -87.6 ...
##  $ end_lat           : num [1:284042] 41.9 41.9 42 41.9 41.9 ...
##  $ end_lng           : num [1:284042] -87.6 -87.6 -87.7 -87.6 -87.7 ...
##  $ member_casual     : chr [1:284042] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Apr_2022)
```

```
## spec_tbl_df [371,249 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:371249] "3564070EEFD12711" "0B820C7FCF22F489" "89EEEE32293F07FF" "84D4
##  $ rideable_type     : chr [1:371249] "electric_bike" "classic_bike" "classic_bike" "classic_bike" .
##  $ started_at        : POSIXct[1:371249], format: "2022-04-06 17:42:48" "2022-04-24 19:23:07" ...
##  $ ended_at          : POSIXct[1:371249], format: "2022-04-06 17:54:36" "2022-04-24 19:43:17" ...
##  $ start_station_name: chr [1:371249] "Paulina St & Howard St" "Wentworth Ave & Cermak Rd" "Halsted S
##  $ start_station_id  : chr [1:371249] "515" "13075" "TA1307000121" "13075" ...
##  $ end_station_name  : chr [1:371249] "University Library (NU)" "Green St & Madison St" "Green St & M
##  $ end_station_id    : chr [1:371249] "605" "TA1307000120" "TA1307000120" "KA1706005007" ...
##  $ start_lat         : num [1:371249] 42 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:371249] -87.7 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:371249] 42.1 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:371249] -87.7 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:371249] "member" "member" "member" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
```

```
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(May_2022)
```

```
## spec_tbl_df [634,858 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:634858] "EC2DE40644C6B0F4" "1C31AD03897EE385" "1542FBEC830415CF" "6FF5
##  $ rideable_type     : chr [1:634858] "classic_bike" "classic_bike" "classic_bike" "classic_bike" ..
##  $ started_at        : POSIXct[1:634858], format: "2022-05-23 23:06:58" "2022-05-11 08:53:28" ...
##  $ ended_at          : POSIXct[1:634858], format: "2022-05-23 23:40:19" "2022-05-11 09:31:22" ...
##  $ start_station_name: chr [1:634858] "Wabash Ave & Grand Ave" "DuSable Lake Shore Dr & Monroe St" "
##  $ start_station_id  : chr [1:634858] "TA1307000117" "13300" "TA1305000032" "TA1305000032" ...
##  $ end_station_name  : chr [1:634858] "Halsted St & Roscoe St" "Field Blvd & South Water St" "Wood S
##  $ end_station_id    : chr [1:634858] "TA1309000025" "15534" "13221" "TA1305000030" ...
##  $ start_lat         : num [1:634858] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:634858] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:634858] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:634858] -87.6 -87.6 -87.7 -87.6 -87.7 ...
##  $ member_casual     : chr [1:634858] "member" "member" "member" "member" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Jun_2022)
```

```
## spec_tbl_df [769,204 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:769204] "600CFD130D0FD2A4" "F5E6B5C1682C6464" "B6EB6D27BAD771D2" "C9C3
##  $ rideable_type     : chr [1:769204] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:769204], format: "2022-06-30 17:27:53" "2022-06-30 18:39:52" ...
##  $ ended_at          : POSIXct[1:769204], format: "2022-06-30 17:35:15" "2022-06-30 18:47:28" ...
##  $ start_station_name: chr [1:769204] NA NA NA NA ...
##  $ start_station_id  : chr [1:769204] NA NA NA NA ...
##  $ end_station_name  : chr [1:769204] NA NA NA NA ...
##  $ end_station_id    : chr [1:769204] NA NA NA NA ...
##  $ start_lat         : num [1:769204] 41.9 41.9 41.9 41.8 41.9 ...
##  $ start_lng         : num [1:769204] -87.6 -87.6 -87.7 -87.7 -87.6 ...
```

```
##  $ end_lat           : num [1:769204] 41.9 41.9 41.9 41.8 41.9 ...
##  $ end_lng           : num [1:769204] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr [1:769204] "casual" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Jul_2022)
```

```
## spec_tbl_df [823,488 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:823488] "954144C2F67B1932" "292E027607D218B6" "57765852588AD6E0" "B5B6
##  $ rideable_type     : chr [1:823488] "classic_bike" "classic_bike" "classic_bike" "classic_bike" ..
##  $ started_at        : POSIXct[1:823488], format: "2022-07-05 08:12:47" "2022-07-26 12:53:38" ...
##  $ ended_at          : POSIXct[1:823488], format: "2022-07-05 08:24:32" "2022-07-26 12:55:31" ...
##  $ start_station_name: chr [1:823488] "Ashland Ave & Blackhawk St" "Buckingham Fountain (Temp)" "Buc
##  $ start_station_id  : chr [1:823488] "13224" "15541" "15541" "15541" ...
##  $ end_station_name  : chr [1:823488] "Kingsbury St & Kinzie St" "Michigan Ave & 8th St" "Michigan A
##  $ end_station_id    : chr [1:823488] "KA1503000043" "623" "623" "TA1307000164" ...
##  $ start_lat         : num [1:823488] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:823488] -87.7 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:823488] 41.9 41.9 41.9 41.8 41.9 ...
##  $ end_lng           : num [1:823488] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:823488] "member" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Aug_2022)
```

```
## spec_tbl_df [785,932 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:785932] "550CF7EFEAE0C618" "DAD198F405F9C5F5" "E6F2BC47B65CB7FD" "F597
##  $ rideable_type     : chr [1:785932] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:785932], format: "2022-08-07 21:34:15" "2022-08-08 14:39:21" ...
##  $ ended_at          : POSIXct[1:785932], format: "2022-08-07 21:41:46" "2022-08-08 14:53:23" ...
##  $ start_station_name: chr [1:785932] NA NA NA NA ...
##  $ start_station_id  : chr [1:785932] NA NA NA NA ...
##  $ end_station_name  : chr [1:785932] NA NA NA NA ...
##  $ end_station_id    : chr [1:785932] NA NA NA NA ...
##  $ start_lat         : num [1:785932] 41.9 41.9 42 41.9 41.9 ...
##  $ start_lng         : num [1:785932] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:785932] 41.9 41.9 42 42 41.8 ...
##  $ end_lng           : num [1:785932] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:785932] "casual" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

####Combine Monthly Data Frames Put monthly data frames into one large data frame.

```
all_trips <- bind_rows(Sep_2021, Oct_2021, Nov_2021, Dec_2021, Jan_2022,
                       Feb_2022, Mar_2022, Apr_2022, May_2022, Jun_2022,
                       Jul_2022, Aug_2022)
```

Inspect the new data frame.

```
colnames(all_trips) #column names
```

```
##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

```r
nrow(all_trips) #number of rows
```

```
## [1] 5883043
```

```r
dim(all_trips) #dimensions
```

```
## [1] 5883043      13
```

```r
head(all_trips) #The first 6 rows
```

```
## # A tibble: 6 x 13
##   ride_id        ridea~1 started_at          ended_at            start~2 start~3
##   <chr>          <chr>   <dttm>              <dttm>              <chr>   <chr>
## 1 9DC7B962304CB~ electr~ 2021-09-28 16:07:10 2021-09-28 16:09:54 <NA>    <NA>
## 2 F930E2C6872D6~ electr~ 2021-09-28 14:24:51 2021-09-28 14:40:05 <NA>    <NA>
## 3 6EF72137900BB~ electr~ 2021-09-28 00:20:16 2021-09-28 00:23:57 <NA>    <NA>
## 4 78D1DE133B3DB~ electr~ 2021-09-28 14:51:17 2021-09-28 15:00:06 <NA>    <NA>
## 5 E03D4ACDCAEF6~ electr~ 2021-09-28 09:53:12 2021-09-28 10:03:44 <NA>    <NA>
## 6 346DE323A2677~ electr~ 2021-09-28 01:53:18 2021-09-28 02:00:02 <NA>    <NA>
## # ... with 7 more variables: end_station_name <chr>, end_station_id <chr>,
## #   start_lat <dbl>, start_lng <dbl>, end_lat <dbl>, end_lng <dbl>,
## #   member_casual <chr>, and abbreviated variable names 1: rideable_type,
## #   2: start_station_name, 3: start_station_id
```

```r
str(all_trips) #List of columns and data types
```

```
## spec_tbl_df [5,883,043 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:5883043] "9DC7B962304CBFD8" "F930E2C6872D6B32" "6EF72137900BB910" "78D
##  $ rideable_type     : chr [1:5883043] "electric_bike" "electric_bike" "electric_bike" "electric_bik
##  $ started_at        : POSIXct[1:5883043], format: "2021-09-28 16:07:10" "2021-09-28 14:24:51" ...
##  $ ended_at          : POSIXct[1:5883043], format: "2021-09-28 16:09:54" "2021-09-28 14:40:05" ...
##  $ start_station_name: chr [1:5883043] NA NA NA NA ...
##  $ start_station_id  : chr [1:5883043] NA NA NA NA ...
##  $ end_station_name  : chr [1:5883043] NA NA NA NA ...
##  $ end_station_id    : chr [1:5883043] NA NA NA NA ...
##  $ start_lat         : num [1:5883043] 41.9 41.9 41.8 41.8 41.9 ...
##  $ start_lng         : num [1:5883043] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:5883043] 41.9 42 41.8 41.8 41.9 ...
##  $ end_lng           : num [1:5883043] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:5883043] "casual" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
```

```
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
summary(all_trips) #Statistical summary.
```

```
##    ride_id          rideable_type        started_at
##  Length:5883043     Length:5883043     Min.   :2021-09-01 00:00:06.00
##  Class :character   Class :character   1st Qu.:2021-11-06 13:47:33.00
##  Mode  :character   Mode  :character   Median :2022-05-07 12:30:47.00
##                                        Mean   :2022-03-22 05:41:41.57
##                                        3rd Qu.:2022-07-06 16:00:29.50
##                                        Max.   :2022-08-31 23:59:39.00
##
##     ended_at                      start_station_name start_station_id
##  Min.   :2021-09-01 00:00:41.00   Length:5883043     Length:5883043
##  1st Qu.:2021-11-06 14:07:58.50   Class :character   Class :character
##  Median :2022-05-07 12:52:55.00   Mode  :character   Mode  :character
##  Mean   :2022-03-22 06:01:26.78
##  3rd Qu.:2022-07-06 16:18:47.00
##  Max.   :2022-09-06 21:49:04.00
##
##  end_station_name   end_station_id       start_lat       start_lng
##  Length:5883043     Length:5883043     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.   :45.64   Max.   :-73.80
##
##     end_lat         end_lng       member_casual
##  Min.   :41.39   Min.   :-88.97   Length:5883043
##  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.   :42.37   Max.   :-87.50
##  NA's   :5727    NA's   :5727
```

####Add Columns for Analysis Preparation Add columns to list the Month, Day, and Year of each ride.

```
all_trips$date <- as.Date(all_trips$started_at)
all_trips$month <- format(as.Date(all_trips$date), "%m")
all_trips$day <- format(as.Date(all_trips$date), "%d")
all_trips$year <- format(as.Date(all_trips$date), "%Y")
all_trips$day_of_week <- format(as.Date(all_trips$date), "%A")
```

Add a column for ride length to the data frame.

```
all_trips$ride_length <- difftime(all_trips$ended_at,all_trips$started_at)
```

Check the data type of the new columns

```
str(all_trips)
```

```
## spec_tbl_df [5,883,043 x 19] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:5883043] "9DC7B962304CBFD8" "F930E2C6872D6B32" "6EF72137900BB910" "78D
##  $ rideable_type     : chr [1:5883043] "electric_bike" "electric_bike" "electric_bike" "electric_bik
##  $ started_at        : POSIXct[1:5883043], format: "2021-09-28 16:07:10" "2021-09-28 14:24:51" ...
##  $ ended_at          : POSIXct[1:5883043], format: "2021-09-28 16:09:54" "2021-09-28 14:40:05" ...
##  $ start_station_name: chr [1:5883043] NA NA NA NA ...
##  $ start_station_id  : chr [1:5883043] NA NA NA NA ...
##  $ end_station_name  : chr [1:5883043] NA NA NA NA ...
##  $ end_station_id    : chr [1:5883043] NA NA NA NA ...
##  $ start_lat         : num [1:5883043] 41.9 41.9 41.8 41.8 41.9 ...
##  $ start_lng         : num [1:5883043] -87.7 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:5883043] 41.9 42 41.8 41.8 41.9 ...
##  $ end_lng           : num [1:5883043] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:5883043] "casual" "casual" "casual" "casual" ...
##  $ date              : Date[1:5883043], format: "2021-09-28" "2021-09-28" ...
##  $ month             : chr [1:5883043] "09" "09" "09" "09" ...
##  $ day               : chr [1:5883043] "28" "28" "28" "28" ...
##  $ year              : chr [1:5883043] "2021" "2021" "2021" "2021" ...
##  $ day_of_week       : chr [1:5883043] "Tuesday" "Tuesday" "Tuesday" "Tuesday" ...
##  $ ride_length       : 'difftime' num [1:5883043] 164 914 221 529 ...
##   ..- attr(*, "units")= chr "secs"
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

If necessary, convert the ride_length column to a numeric value.

```
is.factor(all_trips$ride_length)
```

```
## [1] FALSE
```

```
all_trips$ride_length <- as.numeric(as.character(all_trips$ride_length))
is.numeric(all_trips$ride_length)
```

```
## [1] TRUE
```

####Steps for Analysis Calculations on the ride_length column

```
mean(all_trips$ride_length) #straight average (total ride length / rides)
```

```
## [1] 1185.21
```

```
median(all_trips$ride_length) #midpoint number in the ascending array of ride lengths
```

```
## [1] 643
```

```
max(all_trips$ride_length) #longest ride
```

```
## [1] 2442301
```

```
min(all_trips$ride_length) #shortest ride
```

```
## [1] -8245
```

```
summary(all_trips$ride_length)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -8245     363     643    1185    1160 2442301
```

Compare members versus casual riders

```
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = mean)
```

```
##   all_trips$member_casual all_trips$ride_length
## 1                  casual             1757.7991
## 2                  member              771.2701
```

```
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = median)
```

```
##   all_trips$member_casual all_trips$ride_length
## 1                  casual                   835
## 2                  member                   538
```

```
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = max)
```

```
##   all_trips$member_casual all_trips$ride_length
## 1                  casual               2442301
## 2                  member                 93594
```

```
aggregate(all_trips$ride_length ~ all_trips$member_casual, FUN = min)
```

```
##   all_trips$member_casual all_trips$ride_length
## 1                  casual                -8245
## 2                  member                -7745
```

Average ride time each day for members versus casual riders

```
aggregate(all_trips$ride_length ~ all_trips$member_casual + all_trips$day_of_week, FUN = mean)
```

```
##    all_trips$member_casual all_trips$day_of_week all_trips$ride_length
## 1                   casual                Friday             1673.3297
## 2                   member                Friday              755.6730
## 3                   casual                Monday             1790.5627
## 4                   member                Monday              747.0089
## 5                   casual              Saturday             1922.3602
## 6                   member              Saturday              858.0859
## 7                   casual                Sunday             2050.8559
## 8                   member                Sunday              865.2627
## 9                   casual              Thursday             1561.5749
## 10                  member              Thursday              742.4699
## 11                  casual               Tuesday             1554.6513
## 12                  member               Tuesday              730.6401
## 13                  casual             Wednesday             1502.4619
## 14                  member             Wednesday              731.7435
```

Reorder days of the week

```
all_trips$day_of_week <- ordered(all_trips$day_of_week, levels=c("Sunday", "Monday", "Tuesday", "Wednes
```

Rerun Average ride time by day

```
aggregate(all_trips$ride_length ~ all_trips$member_casual + all_trips$day_of_week, FUN = mean)
```

```
##    all_trips$member_casual all_trips$day_of_week all_trips$ride_length
## 1                   casual                Sunday             2050.8559
## 2                   member                Sunday              865.2627
## 3                   casual                Monday             1790.5627
## 4                   member                Monday              747.0089
## 5                   casual               Tuesday             1554.6513
## 6                   member               Tuesday              730.6401
## 7                   casual             Wednesday             1502.4619
## 8                   member             Wednesday              731.7435
## 9                   casual              Thursday             1561.5749
## 10                  member              Thursday              742.4699
## 11                  casual                Friday             1673.3297
## 12                  member                Friday              755.6730
## 13                  casual              Saturday             1922.3602
## 14                  member              Saturday              858.0859
```

Summarize rides by member type and by day

```
all_trips %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%  #creates weekday field using wday()
  group_by(member_casual, weekday) %>%  #groups by usertype and weekday
  summarise(number_of_rides = n()                            #calculates the number of rides and average
  ,average_duration = mean(ride_length)) %>%        # calculates the average duration
  arrange(member_casual, weekday)                                  # sorts
```

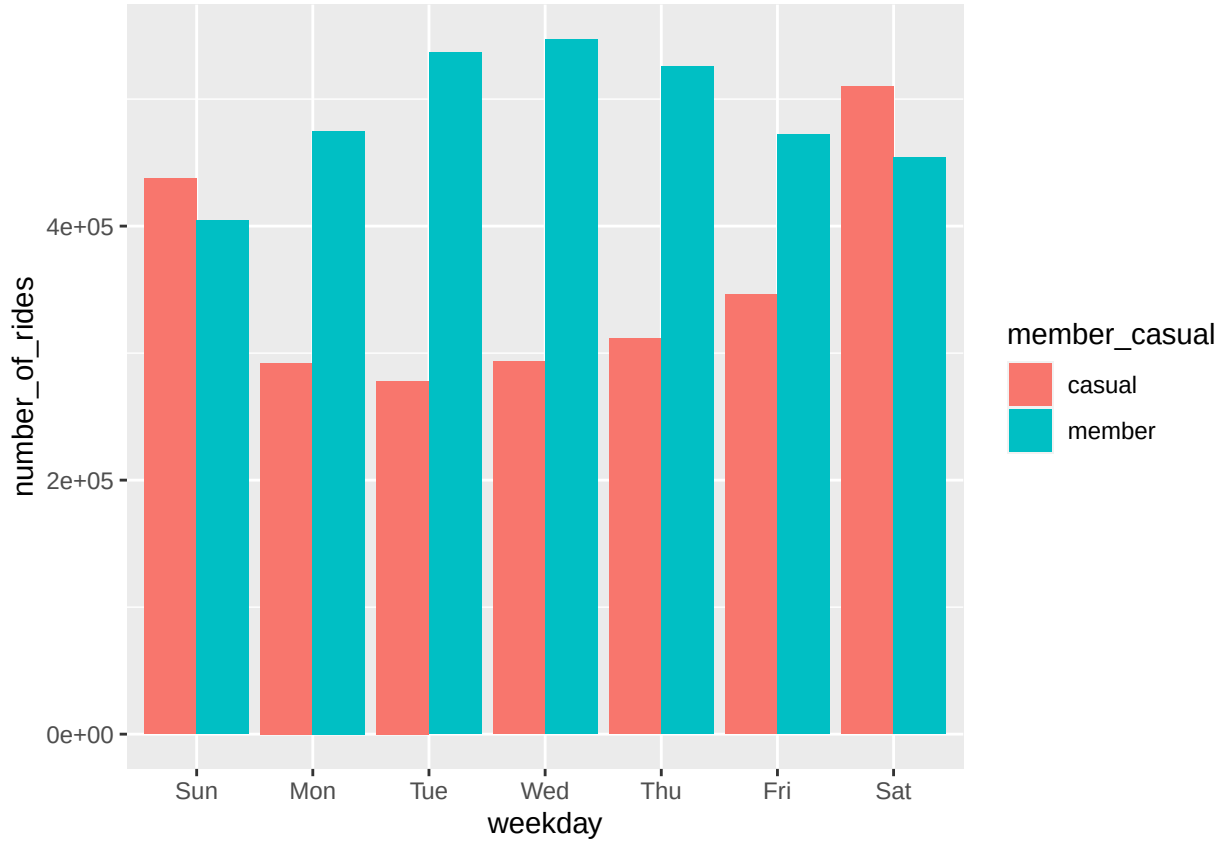
```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 14 x 4
## # Groups:   member_casual [2]
##    member_casual weekday number_of_rides average_duration
##    <chr>         <ord>             <int>            <dbl>
##  1 casual        Sun              437447            2051.
##  2 casual        Mon              292184            1791.
##  3 casual        Tue              278023            1555.
##  4 casual        Wed              293245            1502.
##  5 casual        Thu              311489            1562.
##  6 casual        Fri              346027            1673.
##  7 casual        Sat              510064            1922.
##  8 member        Sun              404525             865.
##  9 member        Mon              474818             747.
## 10 member        Tue              536494             731.
## 11 member        Wed              547038             732.
## 12 member        Thu              525487             742.
## 13 member        Fri              472180             756.
## 14 member        Sat              454022             858.
```

Visualize the number of rides by membership type

```
all_trips %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge")
```

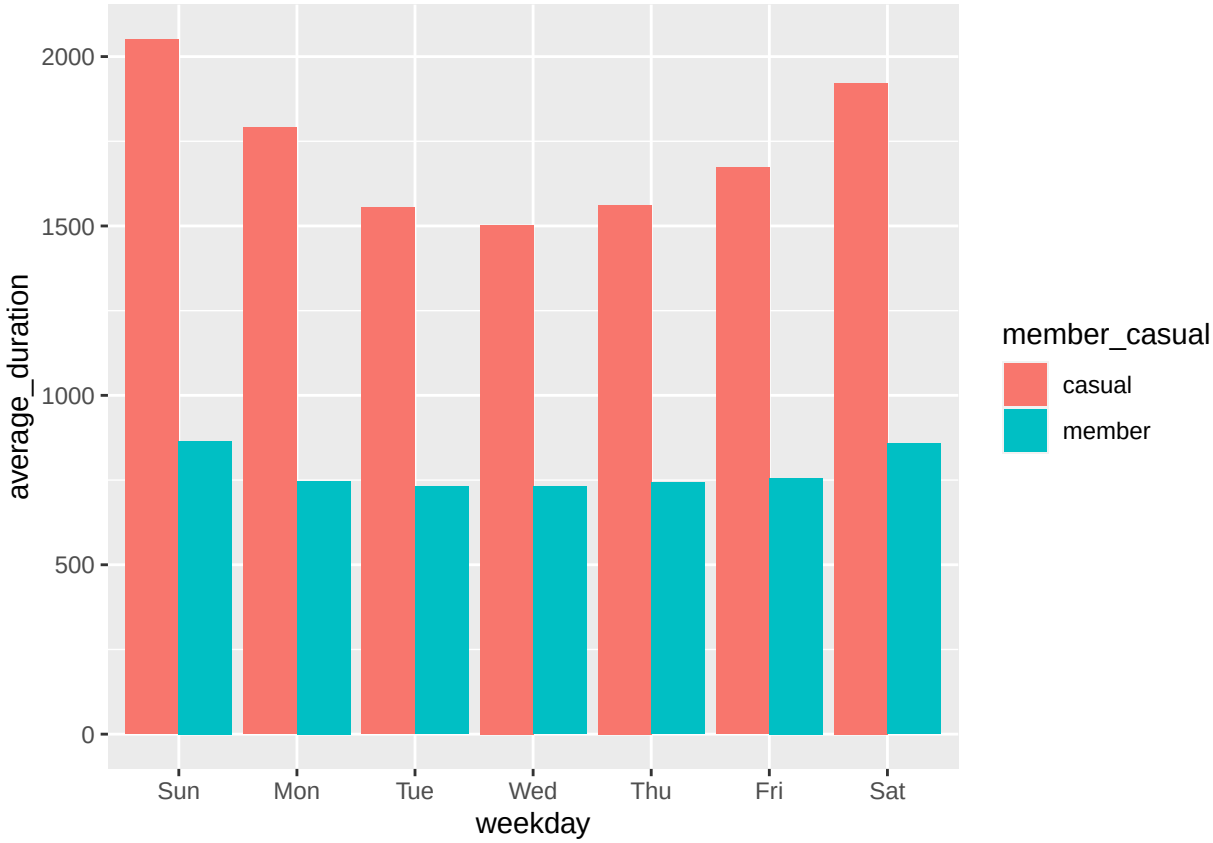
```
## `summarise()` has grouped output by 'member_casual'. You can override using the
## `.groups` argument.
```



Visualize the average duration

```
all_trips %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n()
            ,average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge")
```

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



Create a csv file for visualization

```
counts <- aggregate(all_trips$ride_length ~ all_trips$member_casual + all_trips$day_of_week, FUN = mean)
write.csv(counts, file = '/Users/jcase/Documents/Bikeshare/Raw/avg_ride_length.csv')
```