Trips\_data

Enoch

2023-11-23

# Trips Data : Analysis and Visualization

# Setting up my environment

Notes: setting up my environment by loading ‘tidyverse’ and ‘ggplot2’

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.3.2

## Warning: package 'ggplot2' was built under R version 4.3.2

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(ggplot2)

# Getting the data

q4\_2019 <- read\_csv("Divvy\_Trips\_2019\_Q4.csv")

## Rows: 704054 Columns: 12  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (6): start\_time, end\_time, from\_station\_name, to\_station\_name, usertype,...  
## dbl (5): trip\_id, bikeid, from\_station\_id, to\_station\_id, birthyear  
## num (1): tripduration  
##   
## 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.

q3\_2019 <- read\_csv("Divvy\_Trips\_2019\_Q3.csv")

## Rows: 1640718 Columns: 12  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): from\_station\_name, to\_station\_name, usertype, gender  
## dbl (5): trip\_id, bikeid, from\_station\_id, to\_station\_id, birthyear  
## num (1): tripduration  
## dttm (2): start\_time, end\_time  
##   
## 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.

q2\_2019 <- read\_csv("Divvy\_Trips\_2019\_Q2.csv")

## Rows: 1108163 Columns: 12  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (4): 03 - Rental Start Station Name, 02 - Rental End Station Name, User...  
## dbl (5): 01 - Rental Details Rental ID, 01 - Rental Details Bike ID, 03 - R...  
## num (1): 01 - Rental Details Duration In Seconds Uncapped  
## dttm (2): 01 - Rental Details Local Start Time, 01 - Rental Details Local En...  
##   
## 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.

q1\_2020 <- read\_csv("Divvy\_Trips\_2020\_Q1.csv")

## Rows: 426887 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, started\_at, ended\_at, start\_station\_name, e...  
## dbl (6): start\_station\_id, end\_station\_id, start\_lat, start\_lng, end\_lat, en...  
##   
## 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.

## Rename columns to make them consistent with q1\_2020

q4\_2019 <- rename(q4\_2019  
 ,ride\_id=trip\_id  
 ,rideable\_type = bikeid  
 ,started\_at = start\_time  
 ,ended\_at = end\_time  
 ,start\_station\_name = from\_station\_name  
 ,start\_station\_id = from\_station\_id  
 ,end\_station\_name = to\_station\_name  
 ,end\_station\_id = to\_station\_id  
 ,member\_casual = usertype)  
  
q3\_2019 <- rename(q3\_2019  
 ,ride\_id = trip\_id  
 ,rideable\_type = bikeid  
 ,started\_at = start\_time  
 ,ended\_at = end\_time  
 ,start\_station\_name = from\_station\_name  
 ,start\_station\_id = from\_station\_id  
 ,end\_station\_name = to\_station\_name  
 ,end\_station\_id = to\_station\_id  
 ,member\_casual = usertype)  
  
q2\_2019 <- rename(q2\_2019  
 ,ride\_id = "01 - Rental Details Rental ID"  
 ,rideable\_type = "01 - Rental Details Bike ID"  
 ,started\_at = "01 - Rental Details Local Start Time"  
 ,ended\_at = "01 - Rental Details Local End Time"  
 ,start\_station\_name = "03 - Rental Start Station Name"  
 ,start\_station\_id = "03 - Rental Start Station ID"  
 ,end\_station\_name = "02 - Rental End Station Name"  
 ,end\_station\_id = "02 - Rental End Station ID"  
 ,member\_casual = "User Type")

## inspect the dataframes and look for inconguencies

str(q1\_2020)

## spc\_tbl\_ [426,887 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:426887] "EACB19130B0CDA4A" "8FED874C809DC021" "789F3C21E472CA96" "C9A388DAC6ABF313" ...  
## $ rideable\_type : chr [1:426887] "docked\_bike" "docked\_bike" "docked\_bike" "docked\_bike" ...  
## $ started\_at : chr [1:426887] "21/01/2020 20:06" "30/01/2020 14:22" "09/01/2020 19:29" "06/01/2020 16:17" ...  
## $ ended\_at : chr [1:426887] "21/01/2020 20:14" "30/01/2020 14:26" "09/01/2020 19:32" "06/01/2020 16:25" ...  
## $ start\_station\_name: chr [1:426887] "Western Ave & Leland Ave" "Clark St & Montrose Ave" "Broadway & Belmont Ave" "Clark St & Randolph St" ...  
## $ start\_station\_id : num [1:426887] 239 234 296 51 66 212 96 96 212 38 ...  
## $ end\_station\_name : chr [1:426887] "Clark St & Leland Ave" "Southport Ave & Irving Park Rd" "Wilton Ave & Belmont Ave" "Fairbanks Ct & Grand Ave" ...  
## $ end\_station\_id : num [1:426887] 326 318 117 24 212 96 212 212 96 100 ...  
## $ start\_lat : num [1:426887] 42 42 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:426887] -87.7 -87.7 -87.6 -87.6 -87.6 ...  
## $ end\_lat : num [1:426887] 42 42 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:426887] -87.7 -87.7 -87.7 -87.6 -87.6 ...  
## $ member\_casual : chr [1:426887] "member" "member" "member" "member" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_character(),  
## .. ended\_at = col\_character(),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_double(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_double(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(q2\_2019)

## spc\_tbl\_ [1,108,163 × 12] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : num [1:1108163] 22178529 22178530 22178531 22178532 22178533 ...  
## $ started\_at : POSIXct[1:1108163], format: "2019-04-01 00:02:22" "2019-04-01 00:03:02" ...  
## $ ended\_at : POSIXct[1:1108163], format: "2019-04-01 00:09:48" "2019-04-01 00:20:30" ...  
## $ rideable\_type : num [1:1108163] 6251 6226 5649 4151 3270 ...  
## $ 01 - Rental Details Duration In Seconds Uncapped: num [1:1108163] 446 1048 252 357 1007 ...  
## $ start\_station\_id : num [1:1108163] 81 317 283 26 202 420 503 260 211 211 ...  
## $ start\_station\_name : chr [1:1108163] "Daley Center Plaza" "Wood St & Taylor St" "LaSalle St & Jackson Blvd" "McClurg Ct & Illinois St" ...  
## $ end\_station\_id : num [1:1108163] 56 59 174 133 129 426 500 499 211 211 ...  
## $ end\_station\_name : chr [1:1108163] "Desplaines St & Kinzie St" "Wabash Ave & Roosevelt Rd" "Canal St & Madison St" "Kingsbury St & Kinzie St" ...  
## $ member\_casual : chr [1:1108163] "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...  
## $ Member Gender : chr [1:1108163] "Male" "Female" "Male" "Male" ...  
## $ 05 - Member Details Member Birthday Year : num [1:1108163] 1975 1984 1990 1993 1992 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. `01 - Rental Details Rental ID` = col\_double(),  
## .. `01 - Rental Details Local Start Time` = col\_datetime(format = ""),  
## .. `01 - Rental Details Local End Time` = col\_datetime(format = ""),  
## .. `01 - Rental Details Bike ID` = col\_double(),  
## .. `01 - Rental Details Duration In Seconds Uncapped` = col\_number(),  
## .. `03 - Rental Start Station ID` = col\_double(),  
## .. `03 - Rental Start Station Name` = col\_character(),  
## .. `02 - Rental End Station ID` = col\_double(),  
## .. `02 - Rental End Station Name` = col\_character(),  
## .. `User Type` = col\_character(),  
## .. `Member Gender` = col\_character(),  
## .. `05 - Member Details Member Birthday Year` = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(q3\_2019)

## spc\_tbl\_ [1,640,718 × 12] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : num [1:1640718] 23479388 23479389 23479390 23479391 23479392 ...  
## $ started\_at : POSIXct[1:1640718], format: "2019-07-01 00:00:27" "2019-07-01 00:01:16" ...  
## $ ended\_at : POSIXct[1:1640718], format: "2019-07-01 00:20:41" "2019-07-01 00:18:44" ...  
## $ rideable\_type : num [1:1640718] 3591 5353 6180 5540 6014 ...  
## $ tripduration : num [1:1640718] 1214 1048 1554 1503 1213 ...  
## $ start\_station\_id : num [1:1640718] 117 381 313 313 168 300 168 313 43 43 ...  
## $ start\_station\_name: chr [1:1640718] "Wilton Ave & Belmont Ave" "Western Ave & Monroe St" "Lakeview Ave & Fullerton Pkwy" "Lakeview Ave & Fullerton Pkwy" ...  
## $ end\_station\_id : num [1:1640718] 497 203 144 144 62 232 62 144 195 195 ...  
## $ end\_station\_name : chr [1:1640718] "Kimball Ave & Belmont Ave" "Western Ave & 21st St" "Larrabee St & Webster Ave" "Larrabee St & Webster Ave" ...  
## $ member\_casual : chr [1:1640718] "Subscriber" "Customer" "Customer" "Customer" ...  
## $ gender : chr [1:1640718] "Male" NA NA NA ...  
## $ birthyear : num [1:1640718] 1992 NA NA NA NA ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. trip\_id = col\_double(),  
## .. start\_time = col\_datetime(format = ""),  
## .. end\_time = col\_datetime(format = ""),  
## .. bikeid = col\_double(),  
## .. tripduration = col\_number(),  
## .. from\_station\_id = col\_double(),  
## .. from\_station\_name = col\_character(),  
## .. to\_station\_id = col\_double(),  
## .. to\_station\_name = col\_character(),  
## .. usertype = col\_character(),  
## .. gender = col\_character(),  
## .. birthyear = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(q4\_2019)

## spc\_tbl\_ [704,054 × 12] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : num [1:704054] 25223640 25223641 25223642 25223643 25223644 ...  
## $ started\_at : chr [1:704054] "01/10/2019 00:01" "01/10/2019 00:02" "01/10/2019 00:04" "01/10/2019 00:04" ...  
## $ ended\_at : chr [1:704054] "01/10/2019 00:17" "01/10/2019 00:06" "01/10/2019 00:18" "01/10/2019 00:43" ...  
## $ rideable\_type : num [1:704054] 2215 6328 3003 3275 5294 ...  
## $ tripduration : num [1:704054] 940 258 850 2350 1867 ...  
## $ start\_station\_id : num [1:704054] 20 19 84 313 210 156 84 156 156 336 ...  
## $ start\_station\_name: chr [1:704054] "Sheffield Ave & Kingsbury St" "Throop (Loomis) St & Taylor St" "Milwaukee Ave & Grand Ave" "Lakeview Ave & Fullerton Pkwy" ...  
## $ end\_station\_id : num [1:704054] 309 241 199 290 382 226 142 463 463 336 ...  
## $ end\_station\_name : chr [1:704054] "Leavitt St & Armitage Ave" "Morgan St & Polk St" "Wabash Ave & Grand Ave" "Kedzie Ave & Palmer Ct" ...  
## $ member\_casual : chr [1:704054] "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...  
## $ gender : chr [1:704054] "Male" "Male" "Female" "Male" ...  
## $ birthyear : num [1:704054] 1987 1998 1991 1990 1987 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. trip\_id = col\_double(),  
## .. start\_time = col\_character(),  
## .. end\_time = col\_character(),  
## .. bikeid = col\_double(),  
## .. tripduration = col\_number(),  
## .. from\_station\_id = col\_double(),  
## .. from\_station\_name = col\_character(),  
## .. to\_station\_id = col\_double(),  
## .. to\_station\_name = col\_character(),  
## .. usertype = col\_character(),  
## .. gender = col\_character(),  
## .. birthyear = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

### Convert ride\_id and rideable\_type to character so that they can stack correctly

q4\_2019 <- mutate(q4\_2019,ride\_id = as.character(ride\_id)  
 ,rideable\_type = as.character(rideable\_type)  
 ,ended\_at = as\_datetime(ended\_at)  
 ,started\_at = as\_datetime(started\_at))  
q3\_2019 <- mutate(q3\_2019,ride\_id = as.character(ride\_id)  
 ,rideable\_type = as.character(rideable\_type))  
q2\_2019 <- mutate(q2\_2019,ride\_id = as.character(ride\_id)  
 ,rideable\_type = as.character(rideable\_type))  
q1\_2020 <- mutate(q1\_2020, ended\_at = as\_datetime(ended\_at)  
 ,started\_at = as\_datetime(started\_at))

## Stack individual quarter’s data frames into one big data frames

all\_trips <- bind\_rows(q2\_2019,q3\_2019,q4\_2019,q1\_2020)

## Remove lat,long,birthyear, and gender field

all\_trips <- all\_trips %>%   
 select(-c(start\_lat,start\_lng,end\_lat,end\_lng,birthyear,gender, "01 - Rental Details Duration In Seconds Uncapped","05 - Member Details Member Birthday Year","Member Gender","tripduration"))

## Inspect the new table that has been created

colnames(all\_trips)

## [1] "ride\_id" "started\_at" "ended\_at"   
## [4] "rideable\_type" "start\_station\_id" "start\_station\_name"  
## [7] "end\_station\_id" "end\_station\_name" "member\_casual"

nrow(all\_trips)

## [1] 3879822

dim(all\_trips)

## [1] 3879822 9

head(all\_trips)

## # A tibble: 6 × 9  
## ride\_id started\_at ended\_at rideable\_type start\_station\_id  
## <chr> <dttm> <dttm> <chr> <dbl>  
## 1 221785… 2019-04-01 00:02:22 2019-04-01 00:09:48 6251 81  
## 2 221785… 2019-04-01 00:03:02 2019-04-01 00:20:30 6226 317  
## 3 221785… 2019-04-01 00:11:07 2019-04-01 00:15:19 5649 283  
## 4 221785… 2019-04-01 00:13:01 2019-04-01 00:18:58 4151 26  
## 5 221785… 2019-04-01 00:19:26 2019-04-01 00:36:13 3270 202  
## 6 221785… 2019-04-01 00:19:39 2019-04-01 00:23:56 3123 420  
## # ℹ 4 more variables: start\_station\_name <chr>, end\_station\_id <dbl>,  
## # end\_station\_name <chr>, member\_casual <chr>

str(all\_trips)

## tibble [3,879,822 × 9] (S3: tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:3879822] "22178529" "22178530" "22178531" "22178532" ...  
## $ started\_at : POSIXct[1:3879822], format: "2019-04-01 00:02:22" "2019-04-01 00:03:02" ...  
## $ ended\_at : POSIXct[1:3879822], format: "2019-04-01 00:09:48" "2019-04-01 00:20:30" ...  
## $ rideable\_type : chr [1:3879822] "6251" "6226" "5649" "4151" ...  
## $ start\_station\_id : num [1:3879822] 81 317 283 26 202 420 503 260 211 211 ...  
## $ start\_station\_name: chr [1:3879822] "Daley Center Plaza" "Wood St & Taylor St" "LaSalle St & Jackson Blvd" "McClurg Ct & Illinois St" ...  
## $ end\_station\_id : num [1:3879822] 56 59 174 133 129 426 500 499 211 211 ...  
## $ end\_station\_name : chr [1:3879822] "Desplaines St & Kinzie St" "Wabash Ave & Roosevelt Rd" "Canal St & Madison St" "Kingsbury St & Kinzie St" ...  
## $ member\_casual : chr [1:3879822] "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...

summary(all\_trips)

## ride\_id started\_at   
## Length:3879822 Min. :2001-01-20 20:00:04.00   
## Class :character 1st Qu.:2019-04-24 21:05:19.75   
## Mode :character Median :2019-07-05 17:24:25.50   
## Mean :2018-02-23 20:09:10.21   
## 3rd Qu.:2019-08-26 20:14:26.00   
## Max. :2031-12-20 19:23:57.00   
##   
## ended\_at rideable\_type start\_station\_id  
## Min. :2001-01-20 20:00:04.00 Length:3879822 Min. : 1.0   
## 1st Qu.:2019-04-24 21:24:20.50 Class :character 1st Qu.: 77.0   
## Median :2019-07-05 17:53:07.50 Mode :character Median :174.0   
## Mean :2018-02-23 22:08:31.47 Mean :202.9   
## 3rd Qu.:2019-08-26 20:52:21.75 3rd Qu.:291.0   
## Max. :2031-12-20 19:23:59.00 Max. :675.0   
##   
## start\_station\_name end\_station\_id end\_station\_name member\_casual   
## Length:3879822 Min. : 1.0 Length:3879822 Length:3879822   
## Class :character 1st Qu.: 77.0 Class :character Class :character   
## Mode :character Median :174.0 Mode :character Mode :character   
## Mean :203.8   
## 3rd Qu.:291.0   
## Max. :675.0   
## NA's :1

###How many observation fall under each usertype

table(all\_trips$member\_casual)

##   
## casual Customer member Subscriber   
## 48480 857474 378407 2595461

###Reassign to the desire values(we will go with the current 2020 labels)

all\_trips <- all\_trips %>%   
 mutate(member\_casual = recode(member\_casual  
 ,"Subscriber"= "member"  
 ,"Customer" = "casual"))

###Check to make sure the proper number of oberservations were reassigned

table(all\_trips$member\_casual)

##   
## casual member   
## 905954 2973868

### Add columns that list the date, 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 “ride\_length” calculation to all\_trips (in seconds)

all\_trips$ride\_length <- difftime(all\_trips$ended\_at,all\_trips$started\_at)

#### Inspect the structure of the columns

str(all\_trips)

## tibble [3,879,822 × 15] (S3: tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:3879822] "22178529" "22178530" "22178531" "22178532" ...  
## $ started\_at : POSIXct[1:3879822], format: "2019-04-01 00:02:22" "2019-04-01 00:03:02" ...  
## $ ended\_at : POSIXct[1:3879822], format: "2019-04-01 00:09:48" "2019-04-01 00:20:30" ...  
## $ rideable\_type : chr [1:3879822] "6251" "6226" "5649" "4151" ...  
## $ start\_station\_id : num [1:3879822] 81 317 283 26 202 420 503 260 211 211 ...  
## $ start\_station\_name: chr [1:3879822] "Daley Center Plaza" "Wood St & Taylor St" "LaSalle St & Jackson Blvd" "McClurg Ct & Illinois St" ...  
## $ end\_station\_id : num [1:3879822] 56 59 174 133 129 426 500 499 211 211 ...  
## $ end\_station\_name : chr [1:3879822] "Desplaines St & Kinzie St" "Wabash Ave & Roosevelt Rd" "Canal St & Madison St" "Kingsbury St & Kinzie St" ...  
## $ member\_casual : chr [1:3879822] "member" "member" "member" "member" ...  
## $ date : Date[1:3879822], format: "2019-04-01" "2019-04-01" ...  
## $ month : chr [1:3879822] "04" "04" "04" "04" ...  
## $ day : chr [1:3879822] "01" "01" "01" "01" ...  
## $ year : chr [1:3879822] "2019" "2019" "2019" "2019" ...  
## $ day\_of\_week : chr [1:3879822] "Monday" "Monday" "Monday" "Monday" ...  
## $ ride\_length : 'difftime' num [1:3879822] 446 1048 252 357 ...  
## ..- attr(\*, "units")= chr "secs"

### Convert “ride\_length” from Factor to numeric so we can run calculations on the data

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

### Remove Bad data

all\_trips\_v2 <- all\_trips[!(all\_trips$start\_station\_name == "HQ QR" | all\_trips$ride\_length<0),]

### Descriptive analysis on ride\_length

summary(all\_trips\_v2$ride\_length)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0 23 502 38997 1098 946684785

### Compare members and casual users

aggregate(all\_trips\_v2$ride\_length ~ all\_trips\_v2$member\_casual, FUN = mean)

## all\_trips\_v2$member\_casual all\_trips\_v2$ride\_length  
## 1 casual 105451.53  
## 2 member 18838.77

aggregate(all\_trips\_v2$ride\_length ~ all\_trips\_v2$member\_casual, FUN = median)

## all\_trips\_v2$member\_casual all\_trips\_v2$ride\_length  
## 1 casual 1334  
## 2 member 389

aggregate(all\_trips\_v2$ride\_length ~ all\_trips\_v2$member\_casual, FUN = max)

## all\_trips\_v2$member\_casual all\_trips\_v2$ride\_length  
## 1 casual 946684785  
## 2 member 820454515

aggregate(all\_trips\_v2$ride\_length ~ all\_trips\_v2$member\_casual, FUN = min)

## all\_trips\_v2$member\_casual all\_trips\_v2$ride\_length  
## 1 casual 0  
## 2 member 0

### See the average ride time by each dat for members vs casual users

aggregate(all\_trips\_v2$ride\_length ~ all\_trips\_v2$member\_casual + all\_trips\_v2$day\_of\_week, FUN = mean)

## all\_trips\_v2$member\_casual all\_trips\_v2$day\_of\_week all\_trips\_v2$ride\_length  
## 1 casual Friday 93872.76  
## 2 member Friday 17225.32  
## 3 casual Monday 117681.20  
## 4 member Monday 15409.30  
## 5 casual Saturday 66780.49  
## 6 member Saturday 23916.14  
## 7 casual Sunday 91677.44  
## 8 member Sunday 21894.59  
## 9 casual Thursday 135578.82  
## 10 member Thursday 17586.46  
## 11 casual Tuesday 167255.55  
## 12 member Tuesday 15145.78  
## 13 casual Wednesday 115485.86  
## 14 member Wednesday 22516.84

### Notice that the days of the week are out of order.

all\_trips\_v2$day\_of\_week <- ordered(all\_trips\_v2$day\_of\_week, levels = c("Sunday","Monday","TuesdaY","Wednesday","Thursday","Friday","Saturday"))

### Average ride time by each day for members vs casual users

aggregate(all\_trips\_v2$ride\_length ~ all\_trips\_v2$member\_casual + all\_trips\_v2$day\_of\_week, FUN = mean)

## all\_trips\_v2$member\_casual all\_trips\_v2$day\_of\_week all\_trips\_v2$ride\_length  
## 1 casual Sunday 91677.44  
## 2 member Sunday 21894.59  
## 3 casual Monday 117681.20  
## 4 member Monday 15409.30  
## 5 casual Wednesday 115485.86  
## 6 member Wednesday 22516.84  
## 7 casual Thursday 135578.82  
## 8 member Thursday 17586.46  
## 9 casual Friday 93872.76  
## 10 member Friday 17225.32  
## 11 casual Saturday 66780.49  
## 12 member Saturday 23916.14

### analyze ridership data by type and weekday

all\_trips\_v2 %>%   
 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)

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

## # A tibble: 14 × 4  
## # Groups: member\_casual [2]  
## member\_casual weekday number\_of\_rides average\_duration  
## <chr> <ord> <int> <dbl>  
## 1 casual Sun 164329 91677.  
## 2 casual Mon 107369 117681.  
## 3 casual Tue 96935 167256.  
## 4 casual Wed 97635 115486.  
## 5 casual Thu 109084 135579.  
## 6 casual Fri 122842 93873.  
## 7 casual Sat 203882 66780.  
## 8 member Sun 319287 21895.  
## 9 member Mon 448320 15409.  
## 10 member Tue 473121 15146.  
## 11 member Wed 478824 22517.  
## 12 member Thu 459586 17586.  
## 13 member Fri 435126 17225.  
## 14 member Sat 359537 23916.

### Let’s visualize the number of rides by rider type

all\_trips\_v2 %>%   
 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.

