Data Combination

María del Carmen Martín Rodríguez

2025-01-13

Divvy Exercise Full Year Analysis

This analysis is based on the Divvy case study " 'Sophisticated, Clear, and Polished': Divvy and Data Visualization" written by Kevin Hartman found here. The purpose of this script is to consolidate downloaded Divvy data into a single dataframe and then conduct simple analysis to help answer the key question: "In what ways do members and casual riders use Divvy bikes differently?"

Install required packages

- tidyverse for data import and wrangling
- lubridate for date functions
- ggplot for visualization

```
library(tidyverse)  #helps wrangle data
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

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

STEP 1: COLLECT DATA

Data was collected from this site

```
setwd("/Users/carme/Desktop/Datos_divvy/CSV") #sets your working directory to simplify calls to data ..
# Upload Divvy datasets (csv files) here
jan <- read_csv("202401-divvy-tripdata.csv")
feb <- read_csv("202402-divvy-tripdata.csv")
mar <- read_csv("202403-divvy-tripdata.csv")
apr <- read_csv("202404-divvy-tripdata.csv")
```

```
may <- read_csv("202405-divvy-tripdata.csv")
jun <- read_csv("202406-divvy-tripdata.csv")
jul <- read_csv("202407-divvy-tripdata.csv")
aug <- read_csv("202408-divvy-tripdata.csv")
sep <- read_csv("202409-divvy-tripdata.csv")
oct <- read_csv("202410-divvy-tripdata.csv")
nov <- read_csv("202411-divvy-tripdata.csv")
dec <- read_csv("202412-divvy-tripdata.csv")
```

STEP 2: COMBINE DATA INTO A SINGLE DATAFRAME

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:

```
# Compare column names each of the files
colnames(jan)
```

```
##  [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)
```

```
##  [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)
```

```
##  [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)
```

```
##  [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)
```

```
##  [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)
```

```
##  [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)
```

```
##  [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)
```

```
##  [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(sep)
```

```
##  [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)
```

```
##  [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)
```

```
##  [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)
```

```
##  [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"
```

Inspect the dataframes and look for incongruencies:

```
str(jan)
```

```
## spc_tbl_ [144,873 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:144873] "C1D650626C8C899A" "EECD38BDB25BFCB0" "F4A9CE78061F17F7" "0A0D9
##  $ rideable_type     : chr [1:144873] "electric_bike" "electric_bike" "electric_bike" "classic_bike"
##  $ started_at        : POSIXct[1:144873], format: "2024-01-12 15:30:27" "2024-01-08 15:45:46" ...
##  $ ended_at          : POSIXct[1:144873], format: "2024-01-12 15:37:59" "2024-01-08 15:52:59" ...
##  $ start_station_name: chr [1:144873] "Wells St & Elm St" "Wells St & Elm St" "Wells St & Elm St" "We
##  $ start_station_id  : chr [1:144873] "KA1504000135" "KA1504000135" "KA1504000135" "TA1305000030" ..
##  $ end_station_name  : chr [1:144873] "Kingsbury St & Kinzie St" "Kingsbury St & Kinzie St" "Kingsbu
##  $ end_station_id    : chr [1:144873] "KA1503000043" "KA1503000043" "KA1503000043" "13193" ...
##  $ start_lat         : num [1:144873] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:144873] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:144873] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:144873] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:144873] "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(feb)
```

```
## spc_tbl_ [223,164 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:223164] "FCB05EB1758F85E8" "7FB986AD5D3DE9D6" "40CA13E15B5B470D" "D47A
##  $ rideable_type     : chr [1:223164] "classic_bike" "classic_bike" "electric_bike" "classic_bike" .
##  $ started_at        : POSIXct[1:223164], format: "2024-02-03 14:14:18" "2024-02-05 21:10:06" ...
##  $ ended_at          : POSIXct[1:223164], format: "2024-02-03 14:21:00" "2024-02-05 21:15:44" ...
##  $ start_station_name: chr [1:223164] "Clark St & Newport St" "Michigan Ave & Washington St" "Leavit
##  $ start_station_id  : chr [1:223164] "632" "13001" "TA1309000029" "13235" ...
##  $ end_station_name  : chr [1:223164] "Southport Ave & Waveland Ave" "Wabash Ave & Grand Ave" "Milwa
##  $ end_station_id    : chr [1:223164] "13235" "TA1307000117" "13243" "13229" ...
##  $ start_lat         : num [1:223164] 41.9 41.9 41.9 41.9 41.8 ...
##  $ start_lng         : num [1:223164] -87.7 -87.6 -87.7 -87.7 -87.6 ...
##  $ end_lat           : num [1:223164] 41.9 41.9 41.9 41.9 41.8 ...
##  $ end_lng           : num [1:223164] -87.7 -87.6 -87.7 -87.7 -87.6 ...
##  $ member_casual     : chr [1:223164] "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)
```

```
## spc_tbl_ [301,687 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:301687] "64FBE3BAED5F29E6" "9991629435C5E20E" "E5C9FECD5B71BEBD" "4CEA
##  $ rideable_type     : chr [1:301687] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:301687], format: "2024-03-05 18:33:11" "2024-03-06 17:15:14" ...
##  $ ended_at          : POSIXct[1:301687], format: "2024-03-05 18:51:48" "2024-03-06 17:16:04" ...
##  $ start_station_name: chr [1:301687] NA NA NA NA ...
##  $ start_station_id  : chr [1:301687] NA NA NA NA ...
##  $ end_station_name  : chr [1:301687] NA NA NA NA ...
##  $ end_station_id    : chr [1:301687] NA NA NA NA ...
##  $ start_lat         : num [1:301687] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:301687] -87.7 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:301687] 42 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:301687] -87.7 -87.6 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:301687] "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)
```

```
## spc_tbl_ [415,025 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:415025] "743252713F32516B" "BE90D33D2240C614" "D47BBDDE7C40DD61" "6684
##  $ rideable_type     : chr [1:415025] "classic_bike" "electric_bike" "classic_bike" "classic_bike" .
##  $ started_at        : POSIXct[1:415025], format: "2024-04-22 19:08:21" "2024-04-11 06:19:24" ...
##  $ ended_at          : POSIXct[1:415025], format: "2024-04-22 19:12:56" "2024-04-11 06:22:21" ...
##  $ start_station_name: chr [1:415025] "Aberdeen St & Jackson Blvd" "Aberdeen St & Jackson Blvd" "She
##  $ start_station_id  : chr [1:415025] "13157" "13157" "TA1307000107" "13157" ...
##  $ end_station_name  : chr [1:415025] "Desplaines St & Jackson Blvd" "Desplaines St & Jackson Blvd"
##  $ end_station_id    : chr [1:415025] "15539" "15539" "13249" "15539" ...
##  $ start_lat         : num [1:415025] 41.9 41.9 42 41.9 42 ...
##  $ start_lng         : num [1:415025] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:415025] 41.9 41.9 42 41.9 41.9 ...
##  $ end_lng           : num [1:415025] -87.6 -87.6 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:415025] "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(may)
```

```
## spc_tbl_ [609,493 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:609493] "7D9F0CE9EC2A1297" "02EC47687411416F" "101370FB2D3402BE" "E97E
##  $ rideable_type     : chr [1:609493] "classic_bike" "classic_bike" "classic_bike" "electric_bike" .
##  $ started_at        : POSIXct[1:609493], format: "2024-05-25 15:52:42" "2024-05-14 15:11:51" ...
##  $ ended_at          : POSIXct[1:609493], format: "2024-05-25 16:11:50" "2024-05-14 15:22:00" ...
```

```
##  $ start_station_name: chr [1:609493] "Streeter Dr & Grand Ave" "Sheridan Rd & Greenleaf Ave" "Stree
##  $ start_station_id  : chr [1:609493] "13022" "KA1504000159" "13022" "13022" ...
##  $ end_station_name  : chr [1:609493] "Clark St & Elm St" "Sheridan Rd & Loyola Ave" "Wabash Ave & 9
##  $ end_station_id    : chr [1:609493] "TA1307000039" "RP-009" "TA1309000010" "TA1307000052" ...
##  $ start_lat         : num [1:609493] 41.9 42 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:609493] -87.6 -87.7 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:609493] 41.9 42 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:609493] -87.6 -87.7 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr [1:609493] "casual" "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(jun)
```

```
## spc_tbl_ [710,721 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:710721] "CDE6023BE6B11D2F" "462B48CD292B6A18" "9CFB6A858D23ABF7" "6365
##  $ rideable_type     : chr [1:710721] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:710721], format: "2024-06-11 17:20:06" "2024-06-11 17:19:21" ...
##  $ ended_at          : POSIXct[1:710721], format: "2024-06-11 17:21:39" "2024-06-11 17:19:36" ...
##  $ start_station_name: chr [1:710721] NA NA NA NA ...
##  $ start_station_id  : chr [1:710721] NA NA NA NA ...
##  $ end_station_name  : chr [1:710721] NA NA NA NA ...
##  $ end_station_id    : chr [1:710721] NA NA NA NA ...
##  $ start_lat         : num [1:710721] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:710721] -87.7 -87.7 -87.7 -87.6 -87.6 ...
##  $ end_lat           : num [1:710721] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:710721] -87.7 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:710721] "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)
```

```
## spc_tbl_ [748,962 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:748962] "2658E319B13141F9" "B2176315168A47CE" "C2A9D33DF7EBB422" "8BFE
##  $ rideable_type     : chr [1:748962] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:748962], format: "2024-07-11 08:15:14" "2024-07-11 15:45:07" ...
##  $ ended_at          : POSIXct[1:748962], format: "2024-07-11 08:17:56" "2024-07-11 16:06:04" ...
##  $ start_station_name: chr [1:748962] NA NA NA NA ...
##  $ start_station_id  : chr [1:748962] NA NA NA NA ...
##  $ end_station_name  : chr [1:748962] NA NA NA NA ...
##  $ end_station_id    : chr [1:748962] NA NA NA NA ...
##  $ start_lat         : num [1:748962] 41.8 41.8 41.8 41.9 42 ...
##  $ start_lng         : num [1:748962] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ end_lat           : num [1:748962] 41.8 41.8 41.8 41.9 41.9 ...
##  $ end_lng           : num [1:748962] -87.6 -87.6 -87.6 -87.7 -87.6 ...
##  $ member_casual     : chr [1:748962] "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(aug)
```

```
## spc_tbl_ [755,639 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:755639] "BAA154388A869E64" "8752245932EFF67A" "44DDF9F57A9A161F" "44AA
##  $ rideable_type     : chr [1:755639] "classic_bike" "electric_bike" "classic_bike" "electric_bike"
##  $ started_at        : POSIXct[1:755639], format: "2024-08-02 13:35:14" "2024-08-02 15:33:13" ...
##  $ ended_at          : POSIXct[1:755639], format: "2024-08-02 13:48:24" "2024-08-02 15:55:23" ...
##  $ start_station_name: chr [1:755639] "State St & Randolph St" "Franklin St & Monroe St" "Franklin S
##  $ start_station_id  : chr [1:755639] "TA1305000029" "TA1309000007" "TA1309000007" "TA1307000039" ..
##  $ end_station_name  : chr [1:755639] "Wabash Ave & 9th St" "Damen Ave & Cortland St" "Clark St & El
##  $ end_station_id    : chr [1:755639] "TA1309000010" "13133" "TA1307000039" "TA1306000029" ...
##  $ start_lat         : num [1:755639] 41.9 41.9 41.9 41.9 42 ...
```

```
##  $ start_lng         : num [1:755639] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:755639] 41.9 41.9 41.9 41.9 42 ...
##  $ end_lng           : num [1:755639] -87.6 -87.7 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:755639] "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(sep)
```

```
## spc_tbl_ [821,276 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:821276] "31D38723D5A8665A" "67CB39987F4E895B" "DA61204FD26EC681" "06F1
##  $ rideable_type     : chr [1:821276] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:821276], format: "2024-09-26 15:30:58" "2024-09-26 15:31:32" ...
##  $ ended_at          : POSIXct[1:821276], format: "2024-09-26 15:30:59" "2024-09-26 15:53:13" ...
##  $ start_station_name: chr [1:821276] NA NA NA NA ...
##  $ start_station_id  : chr [1:821276] NA NA NA NA ...
##  $ end_station_name  : chr [1:821276] NA NA NA NA ...
##  $ end_station_id    : chr [1:821276] NA NA NA NA ...
##  $ start_lat         : num [1:821276] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:821276] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:821276] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:821276] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:821276] "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(oct)
```

```
## spc_tbl_ [616,281 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:616281] "4422E707103AA4FF" "19DB722B44CBE82F" "20AE2509FD68C939" "D0F1
##  $ rideable_type     : chr [1:616281] "electric_bike" "electric_bike" "electric_bike" "electric_bike
##  $ started_at        : POSIXct[1:616281], format: "2024-10-14 03:26:04" "2024-10-13 19:33:38" ...
##  $ ended_at          : POSIXct[1:616281], format: "2024-10-14 03:32:56" "2024-10-13 19:39:04" ...
##  $ start_station_name: chr [1:616281] NA NA NA NA ...
##  $ start_station_id  : chr [1:616281] NA NA NA NA ...
##  $ end_station_name  : chr [1:616281] NA NA NA NA ...
##  $ end_station_id    : chr [1:616281] NA NA NA NA ...
##  $ start_lat         : num [1:616281] 42 42 42 42 42 ...
##  $ start_lng         : num [1:616281] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:616281] 42 42 42 42 42 ...
##  $ end_lng           : num [1:616281] -87.7 -87.7 -87.7 -87.7 -87.7 ...
##  $ member_casual     : chr [1:616281] "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)
```

```
## spc_tbl_ [335,075 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:335075] "578DDD7CE1771FFA" "78B141C50102ABA6" "1E794CF36394E2D7" "E5DD
##  $ rideable_type     : chr [1:335075] "classic_bike" "classic_bike" "classic_bike" "classic_bike" ..
##  $ started_at        : POSIXct[1:335075], format: "2024-11-07 19:21:58" "2024-11-22 14:49:00" ...
##  $ ended_at          : POSIXct[1:335075], format: "2024-11-07 19:28:57" "2024-11-22 14:56:15" ...
##  $ start_station_name: chr [1:335075] "Walsh Park" "Walsh Park" "Walsh Park" "Clark St & Elm St" ...
##  $ start_station_id  : chr [1:335075] "18067" "18067" "18067" "TA1307000039" ...
##  $ end_station_name  : chr [1:335075] "Leavitt St & North Ave" "Leavitt St & Armitage Ave" "Damen Ave
##  $ end_station_id    : chr [1:335075] "TA1308000005" "TA1309000029" "13133" "TA1307000142" ...
##  $ start_lat         : num [1:335075] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:335075] -87.7 -87.7 -87.7 -87.6 -87.6 ...
##  $ end_lat           : num [1:335075] 41.9 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:335075] -87.7 -87.7 -87.7 -87.6 -87.6 ...
##  $ member_casual     : chr [1:335075] "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(dec)
```

```
## spc_tbl_ [178,372 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:178372] "6C960DEB4F78854E" "C0913EEB2834E7A2" "848A37DD4723078A" "3FA0
##  $ rideable_type     : chr [1:178372] "electric_bike" "classic_bike" "classic_bike" "electric_bike"
##  $ started_at        : POSIXct[1:178372], format: "2024-12-31 01:38:35" "2024-12-21 18:41:26" ...
##  $ ended_at          : POSIXct[1:178372], format: "2024-12-31 01:48:45" "2024-12-21 18:47:33" ...
##  $ start_station_name: chr [1:178372] "Halsted St & Roscoe St" "Clark St & Wellington Ave" "Sheridan
##  $ start_station_id  : chr [1:178372] "TA1309000025" "TA1307000136" "TA1307000107" "13157" ...
##  $ end_station_name  : chr [1:178372] "Clark St & Winnemac Ave" "Halsted St & Roscoe St" "Broadway &
##  $ end_station_id    : chr [1:178372] "TA1309000035" "TA1309000025" "13137" "chargingstx3" ...
##  $ start_lat         : num [1:178372] 41.9 41.9 42 41.9 41.9 ...
##  $ start_lng         : num [1:178372] -87.6 -87.6 -87.7 -87.7 -87.7 ...
##  $ end_lat           : num [1:178372] 42 41.9 41.9 41.9 41.9 ...
##  $ end_lng           : num [1:178372] -87.7 -87.6 -87.6 -87.6 -87.7 ...
##  $ member_casual     : chr [1:178372] "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>
```

Stack individual data frames into one big data frame:

```r
all_trips <- bind_rows(jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec)
```

Inspect the new table that has been created:

```r
colnames(all_trips)  #List of 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)  #How many rows are in data frame?
```

```
## [1] 5860568
```

```r
dim(all_trips)  #Dimensions of the data frame?
```

```
## [1] 5860568      13
```

```r
head(all_trips)  #See the first 6 rows of data frame.  Also tail(all_trips)
```

```
## # A tibble: 6 x 13
##   ride_id          rideable_type started_at          ended_at
##   <chr>            <chr>         <dttm>              <dttm>
## 1 C1D650626C8C899A electric_bike 2024-01-12 15:30:27 2024-01-12 15:37:59
## 2 EECD38BDB25BFCB0 electric_bike 2024-01-08 15:45:46 2024-01-08 15:52:59
## 3 F4A9CE78061F17F7 electric_bike 2024-01-27 12:27:19 2024-01-27 12:35:19
## 4 0A0D9E15EE50B171 classic_bike  2024-01-29 16:26:17 2024-01-29 16:56:06
## 5 33FFC9805E3EFF9A classic_bike  2024-01-31 05:43:23 2024-01-31 06:09:35
## 6 C96080812CD285C5 classic_bike  2024-01-07 11:21:24 2024-01-07 11:30:03
## # i 9 more variables: start_station_name <chr>, start_station_id <chr>,
## #   end_station_name <chr>, end_station_id <chr>, start_lat <dbl>,
## #   start_lng <dbl>, end_lat <dbl>, end_lng <dbl>, member_casual <chr>
```

```r
str(all_trips)  #See list of columns and data types (numeric, character, etc)
```

```
## spc_tbl_ [5,860,568 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:5860568] "C1D650626C8C899A" "EECD38BDB25BFCB0" "F4A9CE78061F17F7" "0A0
##  $ rideable_type     : chr [1:5860568] "electric_bike" "electric_bike" "electric_bike" "classic_bike
##  $ started_at        : POSIXct[1:5860568], format: "2024-01-12 15:30:27" "2024-01-08 15:45:46" ...
##  $ ended_at          : POSIXct[1:5860568], format: "2024-01-12 15:37:59" "2024-01-08 15:52:59" ...
##  $ start_station_name: chr [1:5860568] "Wells St & Elm St" "Wells St & Elm St" "Wells St & Elm St" "
##  $ start_station_id  : chr [1:5860568] "KA1504000135" "KA1504000135" "KA1504000135" "TA1305000030" .
##  $ end_station_name  : chr [1:5860568] "Kingsbury St & Kinzie St" "Kingsbury St & Kinzie St" "Kingsbu
##  $ end_station_id    : chr [1:5860568] "KA1503000043" "KA1503000043" "KA1503000043" "13193" ...
##  $ start_lat         : num [1:5860568] 41.9 41.9 41.9 41.9 41.9 ...
##  $ start_lng         : num [1:5860568] -87.6 -87.6 -87.6 -87.6 -87.7 ...
##  $ end_lat           : num [1:5860568] 41.9 41.9 41.9 41.9 41.9 ...
```

```
##  $ end_lng           : num [1:5860568] -87.6 -87.6 -87.6 -87.6 -87.6 ...
##  $ member_casual     : chr [1:5860568] "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>
```

```
summary(all_trips)  #Statistical summary of data. Mainly for numerics
```

```
##    ride_id          rideable_type        started_at
##  Length:5860568     Length:5860568     Min.   :2024-01-01 00:00:39.00
##  Class :character   Class :character   1st Qu.:2024-05-20 19:47:53.00
##  Mode  :character   Mode  :character   Median :2024-07-22 20:36:16.27
##                                        Mean   :2024-07-17 07:55:47.61
##                                        3rd Qu.:2024-09-17 20:14:22.56
##                                        Max.   :2024-12-31 23:56:49.84
##
##     ended_at                      start_station_name start_station_id
##  Min.   :2024-01-01 00:04:20.00   Length:5860568     Length:5860568
##  1st Qu.:2024-05-20 20:07:54.75   Class :character   Class :character
##  Median :2024-07-22 20:53:59.16   Mode  :character   Mode  :character
##  Mean   :2024-07-17 08:13:06.54
##  3rd Qu.:2024-09-17 20:27:46.02
##  Max.   :2024-12-31 23:59:55.70
##
##  end_station_name   end_station_id       start_lat       start_lng
##  Length:5860568     Length:5860568     Min.   :41.64   Min.   :-87.91
##  Class :character   Class :character   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Mode  :character   Median :41.90   Median :-87.64
##                                        Mean   :41.90   Mean   :-87.65
##                                        3rd Qu.:41.93   3rd Qu.:-87.63
##                                        Max.   :42.07   Max.   :-87.52
##
##     end_lat         end_lng        member_casual
##  Min.   :16.06   Min.   :-144.05   Length:5860568
##  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.   :87.96   Max.   : 152.53
##  NA's   :7232    NA's   :7232
```

Save data:

```
save(all_trips, file = "C:/Users/carme/Desktop/data_combination.RData")

# Confirmación
cat("El archivo .RData ha sido guardado en el escritorio.")
```

```
## El archivo .RData ha sido guardado en el escritorio.
```