

# Bike Rides Analysis

Pooja Rathee

2022-04-08

## Load Libraries.

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

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

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

##
## Attaching package: 'janitor'

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

##
## Attaching package: 'lubridate'

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

##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##   discard

## The following object is masked from 'package:readr':
##
##   col_factor
```

Read all data files separately.

```
dt1 <- read_csv("202104-divvy-tripdata/202104-divvy-tripdata.csv")
```

```
## Rows: 337230 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dtm   (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.
```

```
dt2 <- read_csv("202105-divvy-tripdata/202105-divvy-tripdata.csv")
```

```
## Rows: 531633 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dtm   (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.
```

```
dt3 <- read_csv("202106-divvy-tripdata/202106-divvy-tripdata.csv")
```

```
## Rows: 729595 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dtm   (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.
```

```
dt4 <- read_csv("202107-divvy-tripdata/202107-divvy-tripdata.csv")
```

```
## Rows: 822410 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dtm   (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.
```

```
dt5 <- read_csv("202108-divvy-tripdata/202108-divvy-tripdata.csv")
```

```
## Rows: 804352 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (7): ride_id, rideable_type, start_station_name, start_station_id, end...
## dbl (4): start_lat, start_lng, end_lat, end_lng
## dtm (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.
```

```
dt6 <- read_csv("202109-divvy-tripdata/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
## dtm (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.
```

```
dt7 <- read_csv("202110-divvy-tripdata/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
## dtm (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.
```

```
dt8 <- read_csv("202111-divvy-tripdata/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
## dtm (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.
```

```
dt9 <- read_csv("202112-divvy-tripdata/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
## dtm (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.
```

```
dt10 <- read_csv("202201-divvy-tripdata/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
## dtm (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.
```

```
dt11 <- read_csv("202202-divvy-tripdata/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
## dtm (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.
```

```
dt12 <- read_csv("202203-divvy-tripdata/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
## dtm (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.
```

Combine All files into bike\_rides

```
bike_rides <- rbind(dt1, dt2, dt3, dt4, dt5, dt6, dt7, dt8, dt9, dt10, dt11, dt12)
dim(bike_rides)
```

```
## [1] 5723532      13
```

```
bike_rides <- janitor::remove_empty(bike_rides, which = c("rows", "cols"))
dim(bike_rides)
```

```
## [1] 5723532      13
```

Below lines parse the hours and dates.

```
bike_rides$start_hour <- lubridate::hour(bike_rides$started_at)
bike_rides$end_hour <- lubridate::hour(bike_rides$ended_at)

bike_rides$start_date <- as.Date(bike_rides$started_at)
bike_rides$end_date <- as.Date(bike_rides$ended_at)
```

#Plot The plot provides an overview that count of Bike Rides increases between 5 AM to 3 PM in previous 12 months.

```
bike_rides %>% count(start_hour, sort = T) %>% ggplot() +
  geom_line(aes(x=start_hour,y=n)) + scale_y_continuous(labels = comma) + labs(title = "Count of Bike Rides by hour")
```

