

Time Analysis of Bikeshare Data

Lucas Adams

2023-10-23

Loading packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## 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(here)
```

```
## here() starts at C:/Users/lucas/OneDrive/Desktop/bikeshare_2023
```

```
library(janitor)
```

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

Loading data

```
df1 <- read_csv(here("data_raw", "202309-capitalbikeshare-tripdata.csv"))
```

```
## Rows: 450090 Columns: 13
## -- Column specification -----
```

```
## Delimiter: ","
## chr (5): ride_id, rideable_type, start_station_name, end_station_name, memb...
## dbl (6): start_station_id, end_station_id, start_lat, start_lng, end_lat, e...
## 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.
```

Creating new dataframe that only looks at first 100 observations

```
df2 = df1 %>%
  slice(1:1000)
```

Combining timestamps into one column

```
df3 = pivot_longer(df2, cols = c("started_at", "ended_at"), values_to = "time")
```

New Variable for Start and End

```
df3 = df3 %>%
  mutate(time_type = case_when(
    time %in% df2$started_at ~ "start",
    time %in% df2$ended_at ~ "end",
  ))
```

Sorting timestamps

```
df3 = df3 %>%
  arrange(time)
```

Assigning value to start and end for cumulative sum calc later

```
df4 = df3 %>%
  mutate(up_down = case_when(
    time_type == "start" ~ 1,
    time_type == "end" ~ -1
  ))
```

Creating cumulative sum variable

```
df5 = df4 %>%  
  mutate(riders = cumsum(up_down))
```

Plotting time vs number of riders

```
ggplot(df5, aes(x = time, y = riders)) + geom_line()
```

