Citibike Data Monthly

Jana Choe

2026-02-19

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.5.2
## v ggplot2   4.0.0     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- 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)
citibike_sum <- read.csv(file = "~/Downloads/citibike_summary_2023.csv", header = T)
```

```
monthly_sum <- citibike_sum %>%
  separate(date, into = c("month", "day"), sep = "-", convert = TRUE) %>%
  group_by(month) %>%
  summarise(
    total_trips = sum(trips),
    total_time = sum(total_time),

    # weighted average trip time
    avg_time = sum(avg_time * trips) / sum(trips),

    # weighted percentages
    member_pct = sum(member_percent * trips) / sum(trips),
    casual_pct = sum(casual_percent * trips) / sum(trips),

    ebike_pct = sum(ebike_percent * trips) / sum(trips),
    regular_pct = sum(regular_percent * trips) / sum(trips)
  )

tail(monthly_sum)
```

```
## # A tibble: 6 x 8
##   month total_trips total_time avg_time member_pct casual_pct ebike_pct
##   <int>       <int>      <dbl>    <dbl>      <dbl>      <dbl>     <dbl>
## 1     7     3629424  51542825.     14.2       77.1       22.9      48.5
```

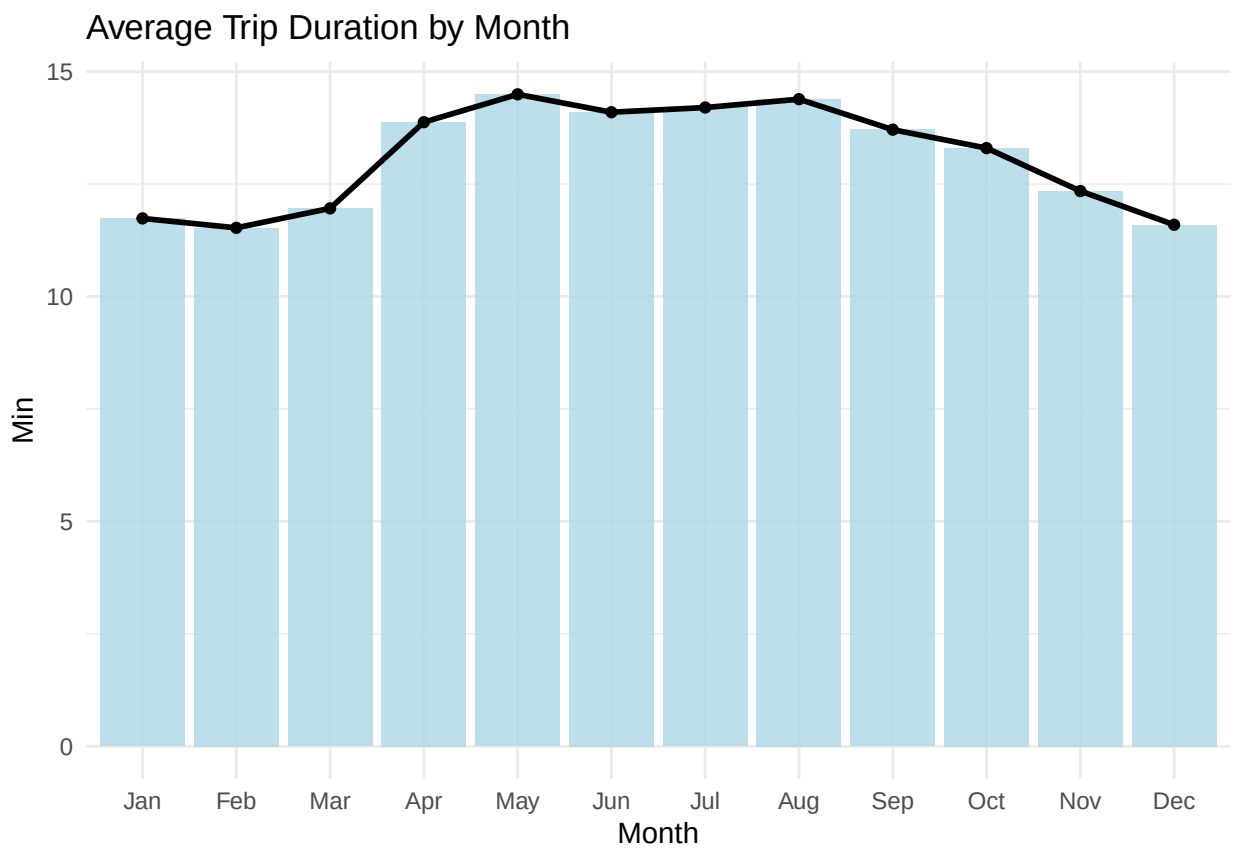
```
## 2     8     3932627  56579344.     14.4       77.6       22.4       50.2
## 3     9     3438816  47144509.     13.7       79.2       20.8       51.3
## 4    10     3691507  49095312.     13.3       81.0       19.0       53.5
## 5    11     2794182  34493111.     12.3       83.7       16.3       59.2
## 6    12     2188927  25381457.     11.6       84.9       15.1       61.8
## # i 1 more variable: regular_pct <dbl>
```

Average Trip Duration by month

```
gg1 <- ggplot(monthly_sum, aes(x = factor(month), y = avg_time)) +
  geom_col(fill = "lightblue", alpha = 0.8) +
  geom_line(aes(group = 1), color = "black", linewidth = 1) +
  geom_point(color = "black", linewidth = 2) +
  scale_x_discrete(labels = month.abb) +
  labs(
    title = "Average Trip Duration by Month",
    x = "Month",
    y = "Min"
  ) +
  theme_minimal()
```

```
## Warning in geom_point(color = "black", linewidth = 2): Ignoring unknown
## parameters: ‘linewidth‘
```

```
gg1
```

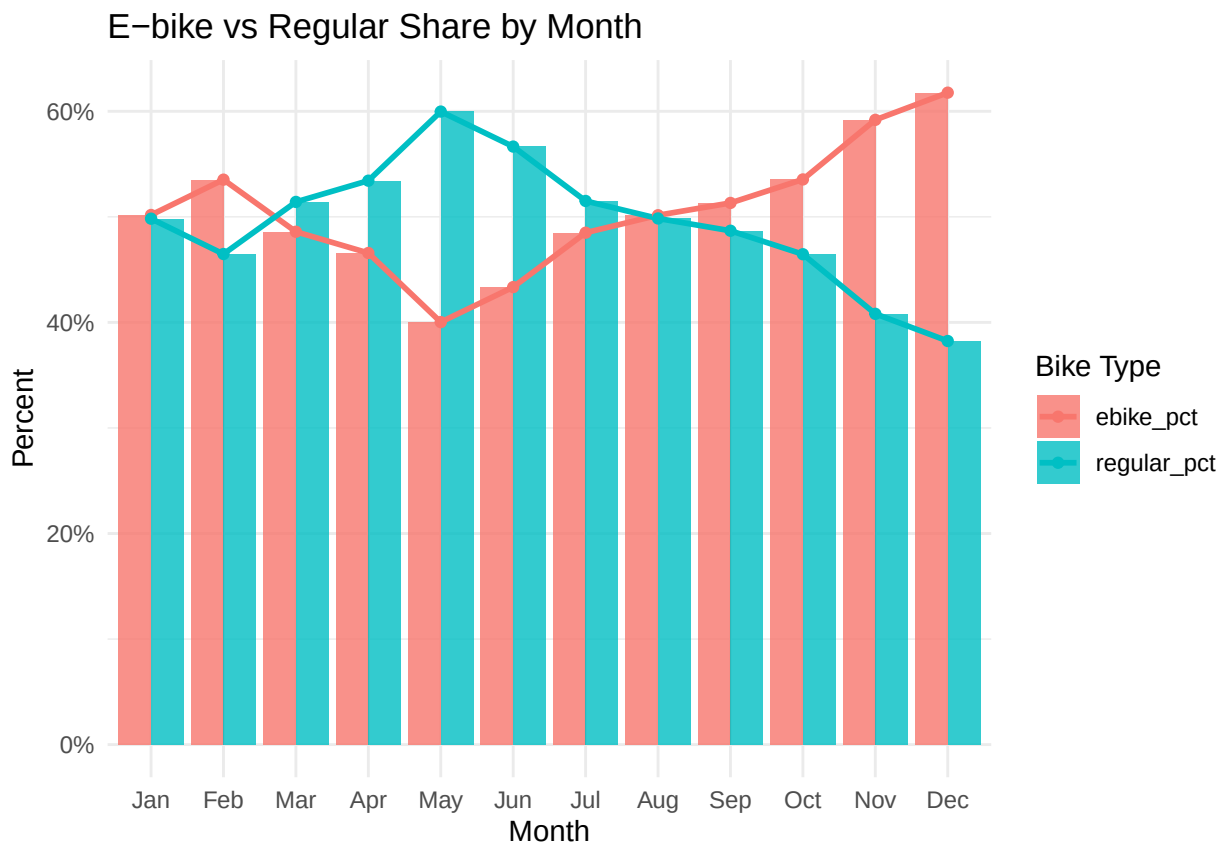


```r
monthly_bike <- monthly_sum %>%
  select(month, ebike_pct, regular_pct) %>%
  pivot_longer(-month, names_to = "bike_type", values_to = "pct")

gg_bike <- ggplot(monthly_bike, aes(x = factor(month), y = pct, fill = bike_type)) +
  geom_col(position = "dodge", alpha = 0.8) +
  geom_line(aes(group = bike_type, color = bike_type), linewidth = 1) +
  geom_point(aes(color = bike_type), linewidth = 2) +
  scale_y_continuous(labels = scales::percent_format(scale = 1)) +
  scale_x_discrete(labels = month.abb) +
  labs(
    title = "E-bike vs Regular Share by Month",
    x = "Month",
    y = "Percent",
    fill = "Bike Type",
    color = "Bike Type"
  ) +
  theme_minimal()
```

```
## Warning in geom_point(aes(color = bike_type), linewidth = 2): Ignoring unknown
## parameters: 'linewidth'
```

```r
gg_bike
```



```r
monthly_member <- monthly_sum %>%
  select(month, member_pct, casual_pct) %>%
  pivot_longer(-month, names_to = "user_type", values_to = "pct")

gg_mem <- ggplot(monthly_member, aes(x = factor(month), y = pct, fill = user_type)) +
  geom_col(position = "dodge", alpha = 0.8) +
  geom_line(aes(group = user_type, color = user_type), linewidth = 1) +
  geom_point(aes(color = user_type),linewidth = 2) +
  scale_y_continuous(labels = scales::percent_format(scale = 1)) +
  scale_x_discrete(labels = month.abb) +
  labs(
    title = "Member vs Nonmember distribution by Month",
    x = "Month",
    y = "Percent"
  ) +
  theme_minimal()
```

```
## Warning in geom_point(aes(color = user_type), linewidth = 2): Ignoring unknown
## parameters: `linewidth`
```

```r
gg_mem
```