Cyclistic Report and Recommendations

Amy Cavender

2022-09-29

Overview

This report analyzes ridership data from June 1, 2021 through May 31, 2022 to uncover differences in the ways annual members and casual riders use Cyclistic bikes.

Data

Source

Data were downloaded from Amazon storage (https://divvy-tripdata.s3.amazonaws.com/index.html) in June 2022. The data are provided under Divvy's Data License Agreement (https://ride.divvybikes.com/data-license-agreement). Divvy is Chicago's bike share service, operated by Lyft.

Each file downloaded contains one month's data.

Cleaning

For each of the twelve data files downloaded, the following steps were taken to clean and prepare the data for analysis:

- Calculated the length of each ride.
- Calculated the day of the week that each ride started.
- Removed rides where the ride length was negative, or the name of the start station was "HQ QR."
- Removed columns containing information about start and end station names and IDs. That information was not available for all rides. Latitude and longitude *were* available for all rides, so locations can still be determined if needed for future analysis.

The twelve individual files were then combined into one csv file and used to create a data frame, `all_rides`.

```
all_rides <- read_csv("~/Desktop/Cyclistic study/2022-07-06-all_trips_v4.csv")
```

```
## New names:
## Rows: 5853109 Columns: 16
## ── Column specification
## ──────────────────────────────────────────────────────── Delimiter: "," chr
## (6): ride_id, rideable_type, day_of_week, member_casual, month, day dbl (7):
## ...1, ride_length, start_lat, start_lng, end_lat, end_lng, year dttm (2):
## started_at, ended_at date (1): date
## 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.
## • `` -> `...1`
```

Analysis

Average ride length, longest ride, shortest ride

All riders

The longest ride was 932 hours, 24 minutes.

```
longest_ride <- hms::hms(max(all_rides$ride_length))
longest_ride
```

```
## 932:24:00
```

That's more than five weeks. It's more likely that someone forgot to return a bike in timely fashion than that their ride actually lasted that long.

The shortest ride was one second.

```
shortest_ride <- hms::hms(min(all_rides$ride_length))
shortest_ride
```

```
## 00:00:01
```

This is also an unlikely ride length. Perhaps someone checked out a bike and immediately changed their mind.

The average ride length was 20 minutes, 43 seconds.

```
average_ride <- hms::hms(mean(all_rides$ride_length))
average_ride
```

```
## 00:20:43.005519
```

Members

The longest ride length for members was just under 26 minutes.

```
members <- filter(all_rides, member_casual=="member")
longest_ride_m <-hms::hms(max(members$ride_length))
longest_ride_m
```

```
## 25:59:54
```

The shortest ride length for members was one second. As noted above, this seems unlikely to be accurate.

```
shortest_ride_m <- hms::hms(min(members$ride_length))
shortest_ride_m
```

```
## 00:00:01
```

The average ride length for members was just over 13 minutes, which is considerably shorter than the overall average of 20 minutes, 43 seconds.

```
average_ride_m <- hms::hms(mean(members$ride_length))
average_ride_m
```

```
## 00:13:03.583933
```

Casual riders

The longest ride length for casual riders was over five weeks. As noted above, this is likely inaccurate.

```
casual <- filter(all_rides, member_casual=="casual")
longest_ride_c <-hms::hms(max(casual$ride_length))
longest_ride_c
```

```
## 932:24:00
```

The shortest ride length for casual riders was one second. Again, this is likely inaccurate.

```
shortest_ride_c <- hms::hms(min(casual$ride_length))
shortest_ride_c
```

```
## 00:00:01
```

The average ride length for casual riders was 30 minutes, 36 seconds.

```
average_ride_c <- hms::hms(mean(casual$ride_length))
average_ride_c
```

```
## 00:30:35.576141
```

Note that there is a clear difference in average ride length. Casual riders use bikes for considerably longer per ride (an average of 30 minutes, 36 seconds) than members (who average only 13 minutes, 4 seconds).

Days of week with the most rides

All riders

Saturday sees the most rides overall.

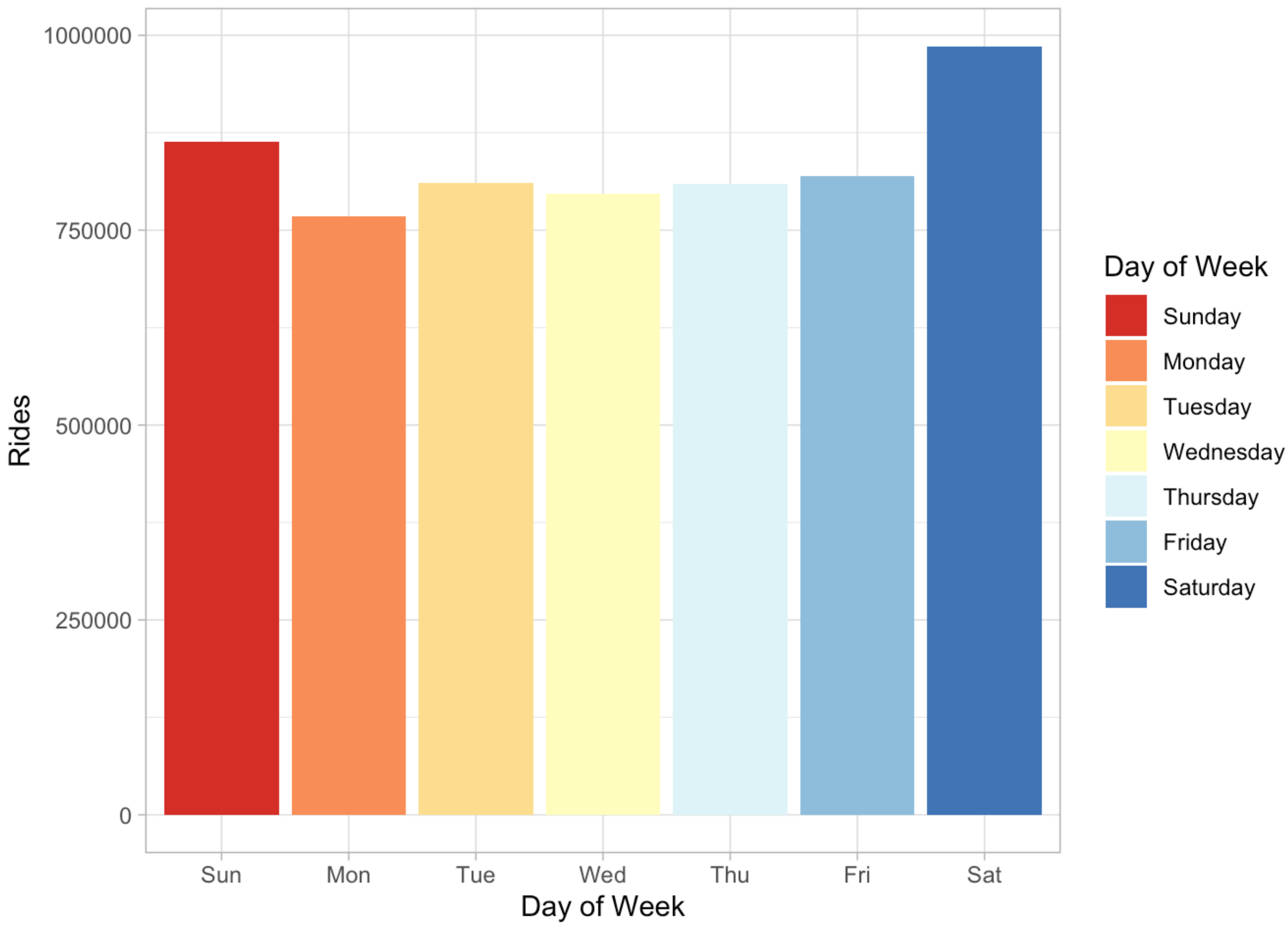
```
busiest_day <- names(which.max(table(all_rides$day_of_week)))
busiest_day
```

```
## [1] "Saturday"
```

```
num_rides <- all_rides %>% mutate(day_of_week = factor(day_of_week, levels = c("Sunda
y", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))) %>% count(d
ay_of_week)
print(num_rides)
```

```
## # A tibble: 7 × 2
##   day_of_week      n
##   <fct>        <int>
## 1 Sunday      863625
## 2 Monday      767423
## 3 Tuesday     810740
## 4 Wednesday   797227
## 5 Thursday    809497
## 6 Friday      818753
## 7 Saturday    985844
```

```
plot_all <- ggplot(data=num_rides, aes(x=day_of_week, y=n, fill=day_of_week)) + geom_
bar(stat="identity") + labs (y="Rides", x="Day of Week", fill="Day of Week") + theme_
light() + scale_fill_brewer(palette = "RdYlBu")
plot_all2 <- plot_all + scale_x_discrete(breaks=c("Sunday", "Monday", "Tuesday", "Wed
nesday", "Thursday", "Friday", "Saturday"), labels=c("Sun","Mon","Tue","Wed","Thu","F
ri","Sat"))
plot_all2
```



Members

For members, Tuesday is the busiest day.

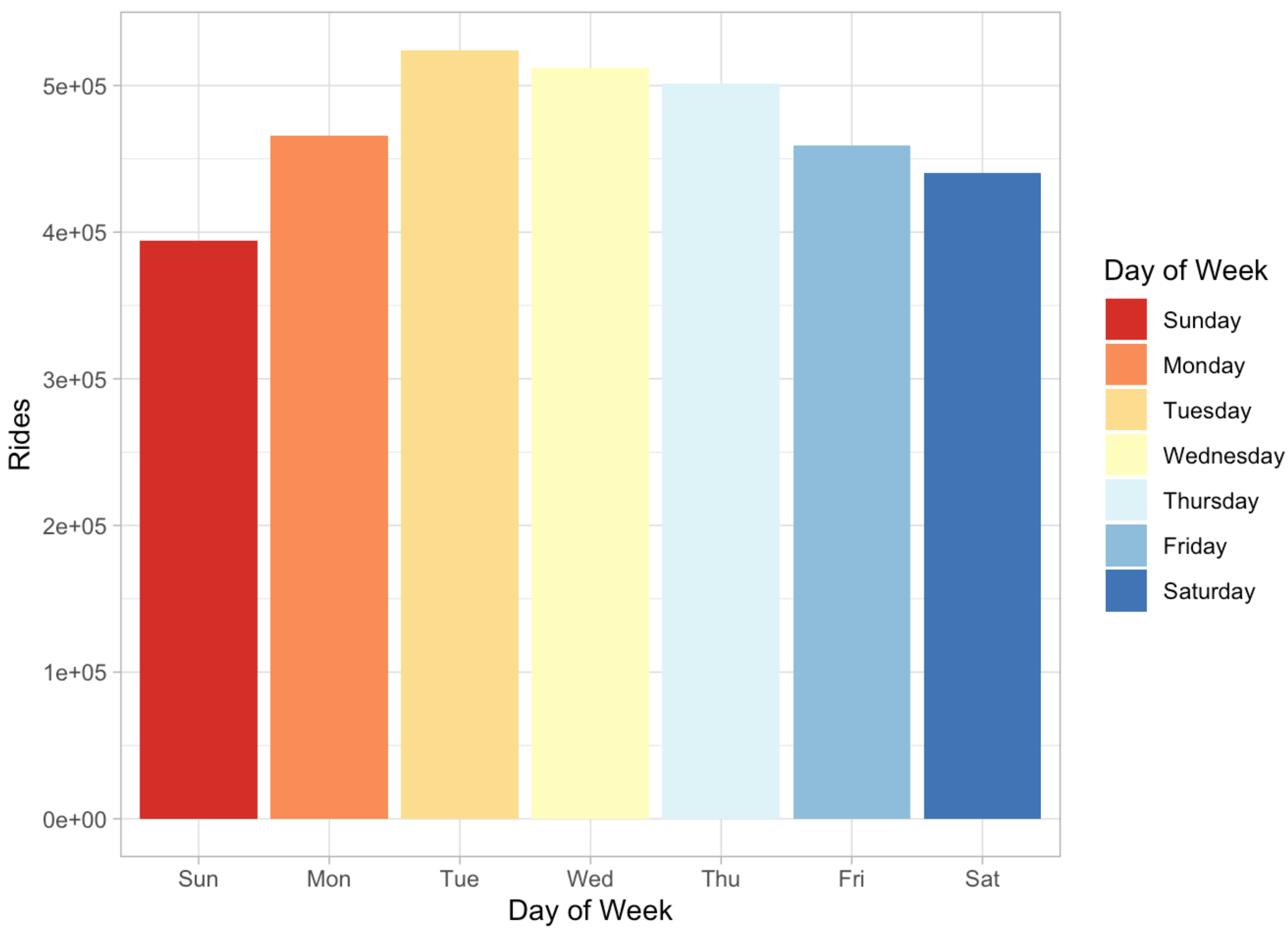
```
members <- filter(all_rides, member_casual=="member")
busiest_day_m <- names(which.max(table(members$day_of_week)))
busiest_day_m
```

```
## [1] "Tuesday"
```

```
num_rides_m <- members %>% mutate(day_of_week = factor(day_of_week, levels = c("Sunda
y", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))) %>% count(d
ay_of_week)
print(num_rides_m)
```

```
## # A tibble: 7 × 2
##   day_of_week      n
##   <fct>        <int>
## 1 Sunday      394125
## 2 Monday      465679
## 3 Tuesday     524202
## 4 Wednesday   511953
## 5 Thursday    501282
## 6 Friday      459273
## 7 Saturday    440450
```

```
plot_all_m <- ggplot(data=num_rides_m, aes(x=day_of_week, y=n, fill=day_of_week)) + g
eom_bar(stat="identity") + labs (y="Rides", x="Day of Week", fill="Day of Week") + th
eme_light() + scale_fill_brewer(palette = "RdYlBu")
plot_all_m2 <- plot_all_m + scale_x_discrete(breaks=c("Sunday", "Monday", "Tuesday",
"Wednesday", "Thursday", "Friday", "Saturday"), labels=c("Sun","Mon","Tue","Wed","Thu
","Fri","Sat"))
plot_all_m2
```



Casual riders

Saturday is the busiest day for casual riders.

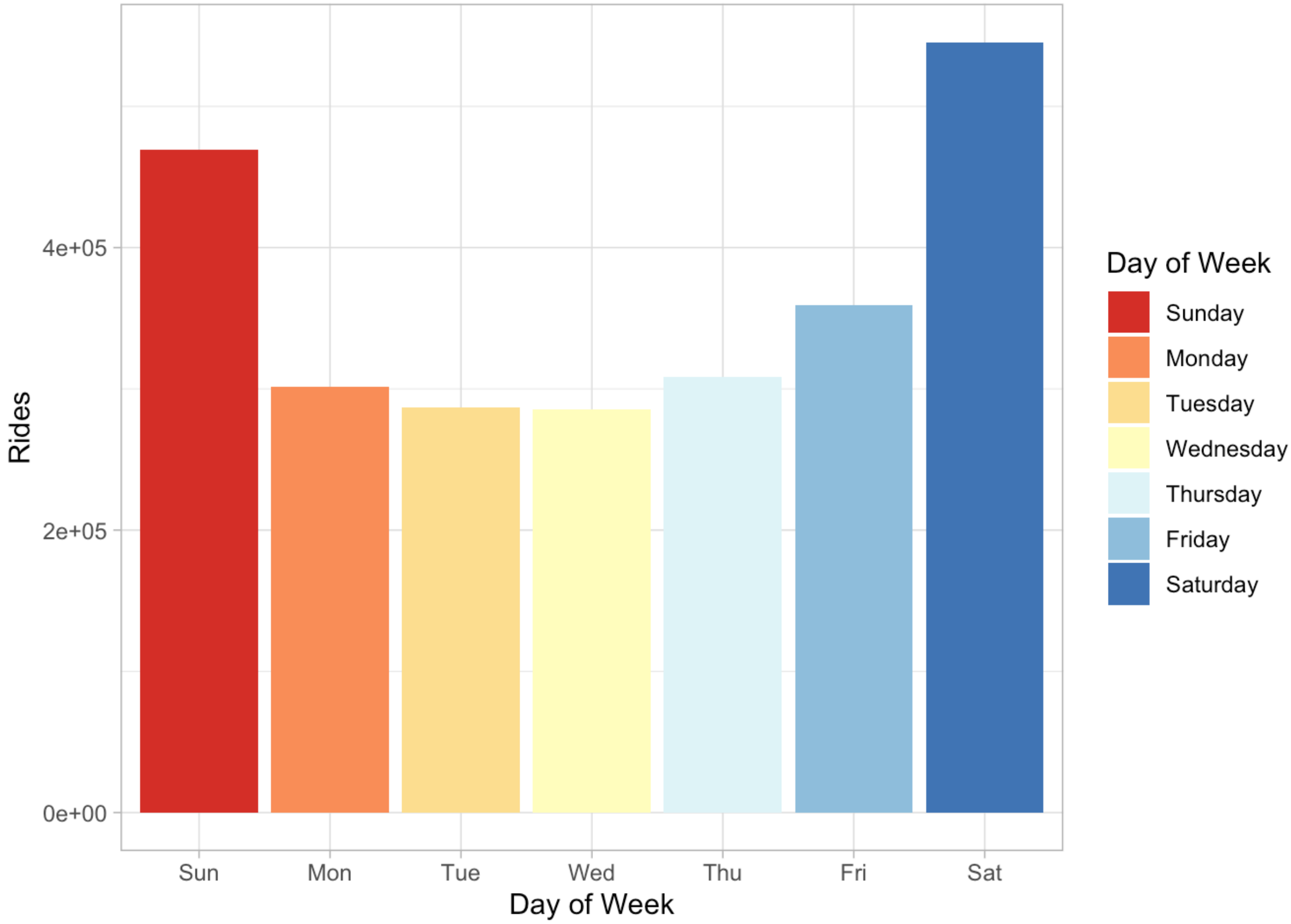
```
casual <- filter(all_rides, member_casual=="casual")
busiest_day_c <- names(which.max(table(casual$day_of_week)))
busiest_day_c
```

```
## [1] "Saturday"
```

```
num_rides_c <- casual %>% mutate(day_of_week = factor(day_of_week, levels = c("Sunday
", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))) %>% count(da
y_of_week)
print(num_rides_c)
```

```
## # A tibble: 7 × 2
##   day_of_week      n
##   <fct>        <int>
## 1 Sunday      469500
## 2 Monday      301744
## 3 Tuesday     286538
## 4 Wednesday   285274
## 5 Thursday    308215
## 6 Friday      359480
## 7 Saturday    545394
```

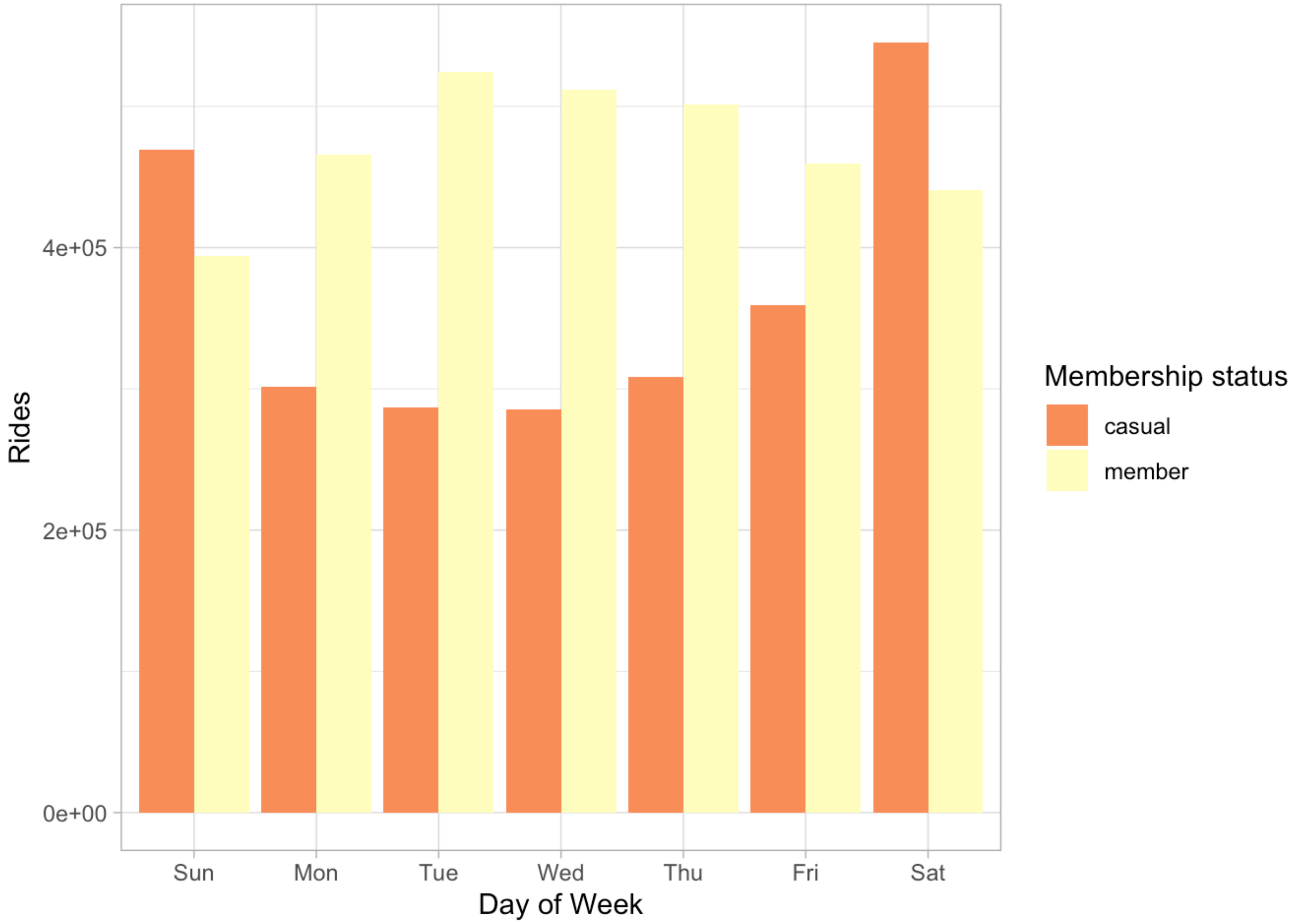
```
plot_all_c <- ggplot(data=num_rides_c, aes(x=day_of_week, y=n, fill=day_of_week)) + g
eom_bar(stat="identity") + labs (y="Rides", x="Day of Week", fill="Day of Week") + th
eme_light() + scale_fill_brewer(palette = "RdYlBu")
plot_all_c2 <- plot_all_c + scale_x_discrete(breaks=c("Sunday", "Monday", "Tuesday",
"Wednesday", "Thursday", "Friday", "Saturday"), labels=c("Sun","Mon","Tue","Wed","Thu
","Fri","Sat"))
plot_all_c2
```



Differences

We can clearly see the difference between the two groups when we plot them together:

```
riders_day <- all_rides %>%
  group_by(member_casual) %>%
  mutate(day_of_week = factor(day_of_week, levels = c("Sunday", "Monday", "Tuesday",
"Wednesday", "Thursday", "Friday", "Saturday"))) %>%
  count(day_of_week)
riders_day_plot <- ggplot(riders_day, aes(x=day_of_week, y=n, fill=member_casual)) +
geom_bar(position="dodge", stat = "identity") + labs(x="Day of Week", y="Rides", fill
="Membership status") + theme_light() + scale_fill_brewer(palette = "RdYlBu")
riders_day_plot2 <- riders_day_plot + scale_x_discrete(breaks=c("Sunday", "Monday", "
Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"), labels=c("Sun","Mon","Tue",
"Wed","Thu","Fri","Sat"))
riders_day_plot2
```

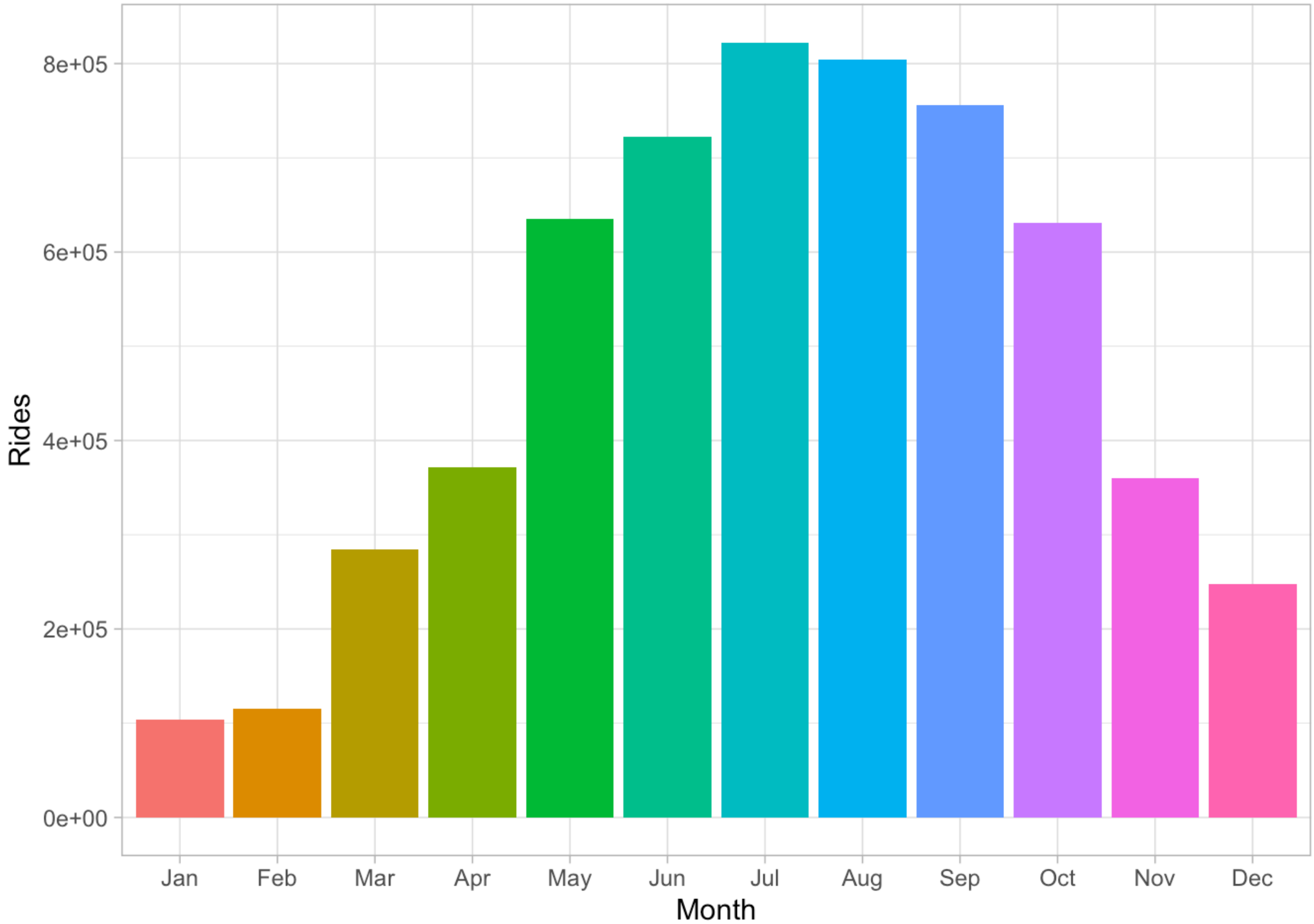


Months with the most rides

All riders

July, August, and September (in that order) are the busiest months overall.

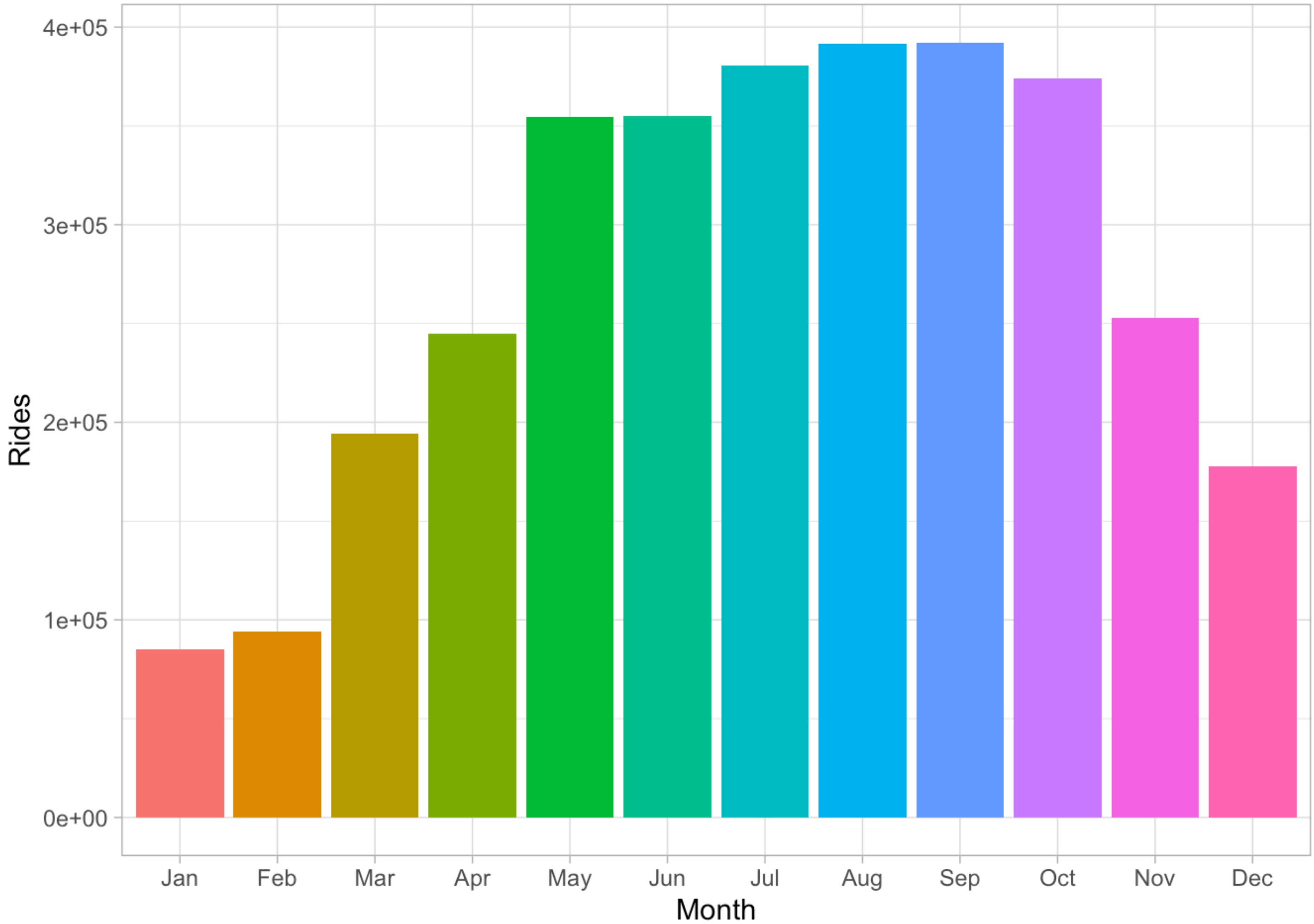
```
month_rides <- all_rides %>% count(month)
plot_month_all <- ggplot(data=month_rides, aes(x=month, y=n, fill=month)) + geom_bar(
stat="identity") + labs(y="Rides", x="Month", fill="Month") + theme_light() + scale_x
_discrete(labels=c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov",
"Dec")) + theme(legend.position = "none")
plot_month_all
```



Members

They are also the busiest months for members, though the order differs. Note that member usage is fairly consistent from May through October.

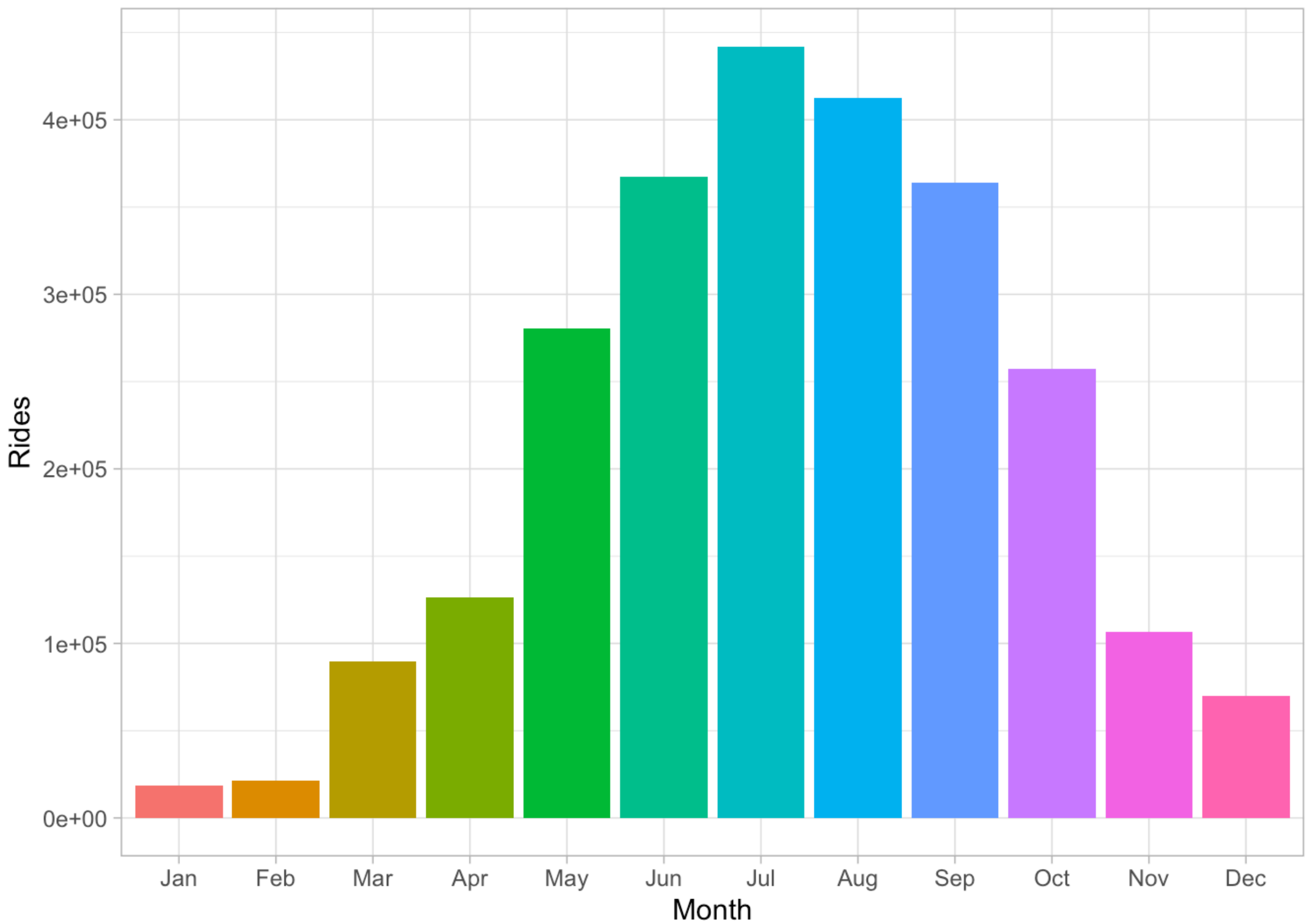
```
month_rides_m <- members %>% count(month)
plot_month_m <- ggplot(data=month_rides_m, aes(x=month, y=n, fill=month)) + geom_bar(
stat="identity") + labs(y="Rides", x="Month", fill="Month") + theme_light() + scale_x
_discrete(labels=c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov",
"Dec")) + theme(legend.position = "none")
plot_month_m
```



Casual riders

July, August, and September are also the busiest months for casual riders, though the order is reversed from the overall pattern. Note that casual riders' use drops dramatically during the colder months of the year.

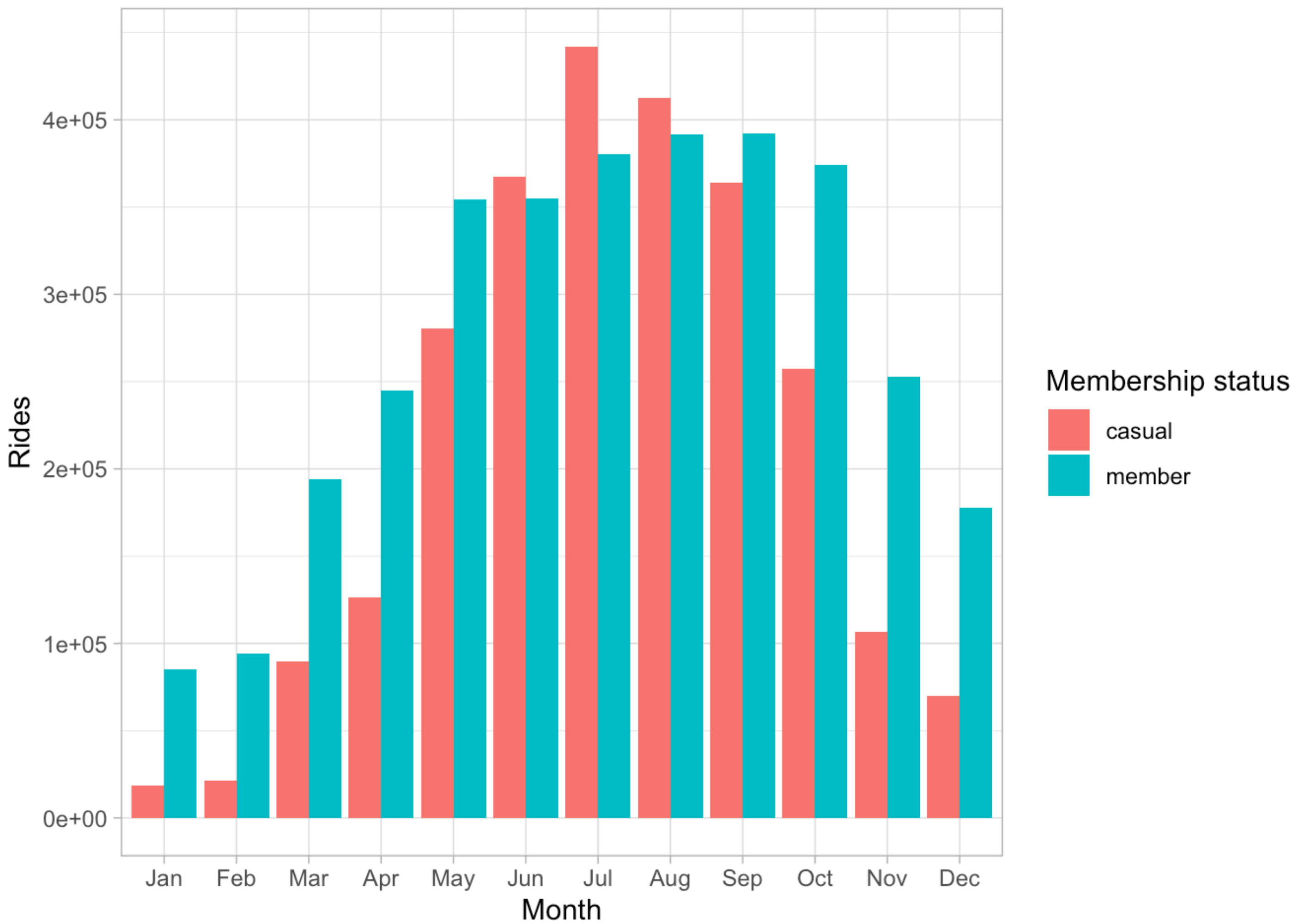
```
month_rides_c <- casual %>% count(month)
plot_month_c <- ggplot(data=month_rides_c, aes(x=month, y=n, fill=month)) + geom_bar(
stat="identity") + labs(y="Rides", x="Month", fill="month") + theme_light() + scale_x
_discrete(labels=c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov",
"Dec")) + theme(legend.position = "none")
plot_month_c
```



Differences

Casual members use bikes more than members during June, July, and August. Members show higher usage in other months; their usage is *much* higher than that of casual members during the colder months of the year (especially January - April and October - December). This may indicate that members are likely to be using Cyclistic bikes to commute to and from work.

```
riders_month <- all_rides %>%
  group_by(member_casual) %>%
  count(month)
riders_month_plot <- ggplot(riders_month, aes(x=month, y=n, fill=member_casual)) + ge
om_bar(position="dodge", stat = "identity") + labs(y="Rides", x="Month", fill="Member
ship status") + theme_light() + scale_x_discrete(labels=c("Jan","Feb","Mar","Apr","Ma
y","Jun","Jul","Aug","Sep","Oct","Nov","Dec"))
riders_month_plot
```



Bike types

All riders

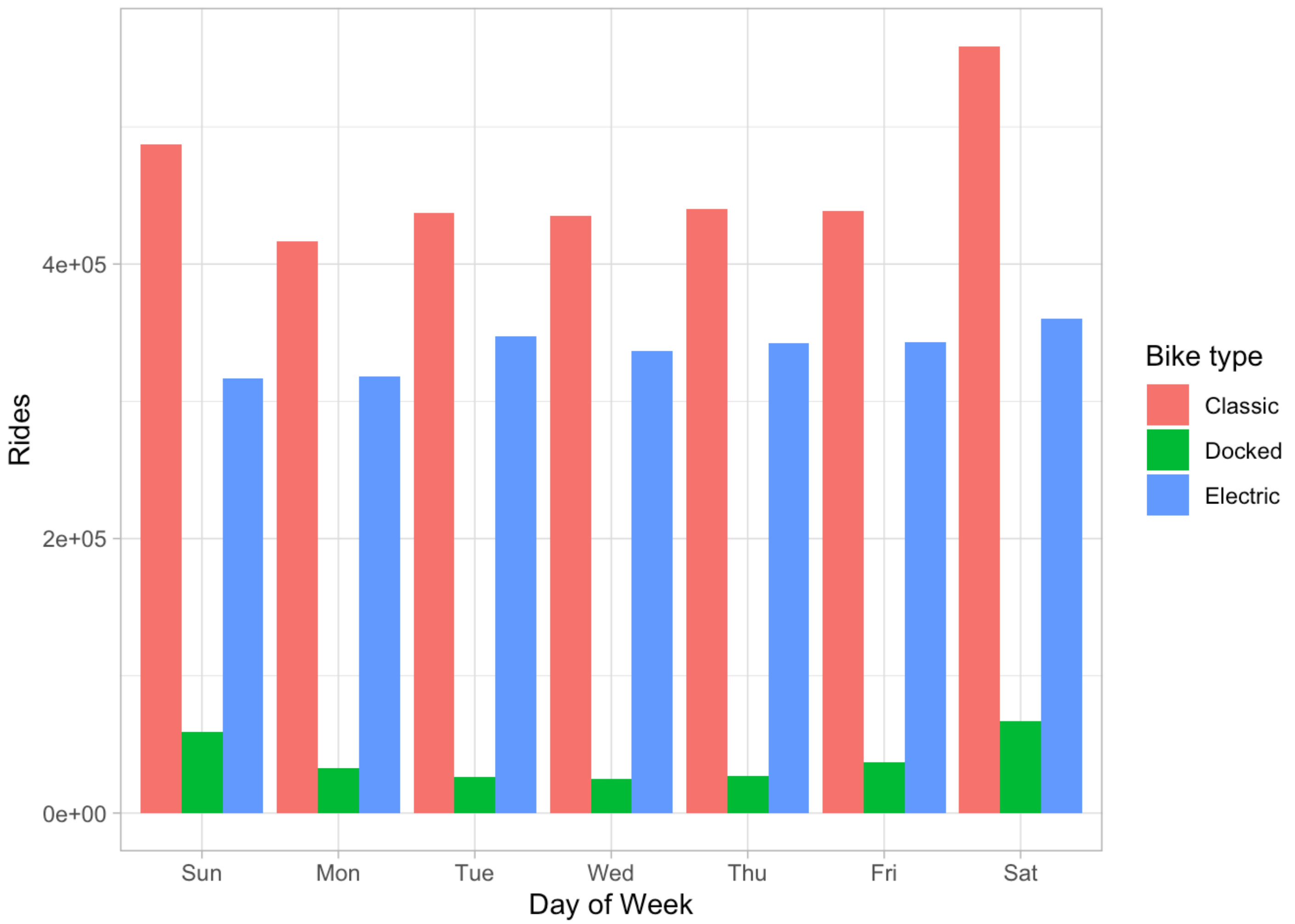
Classic bikes are used most frequently, followed by electric bikes, then docked bikes (which see *much* less frequent use than either of the other two types).

```
num_type <- all_rides %>% count(rideable_type)
num_type
```

```
## # A tibble: 3 × 2
##   rideable_type       n
##   <chr>           <int>
## 1 classic_bike  3213935
## 2 docked_bike    274246
## 3 electric_bike 2364928
```

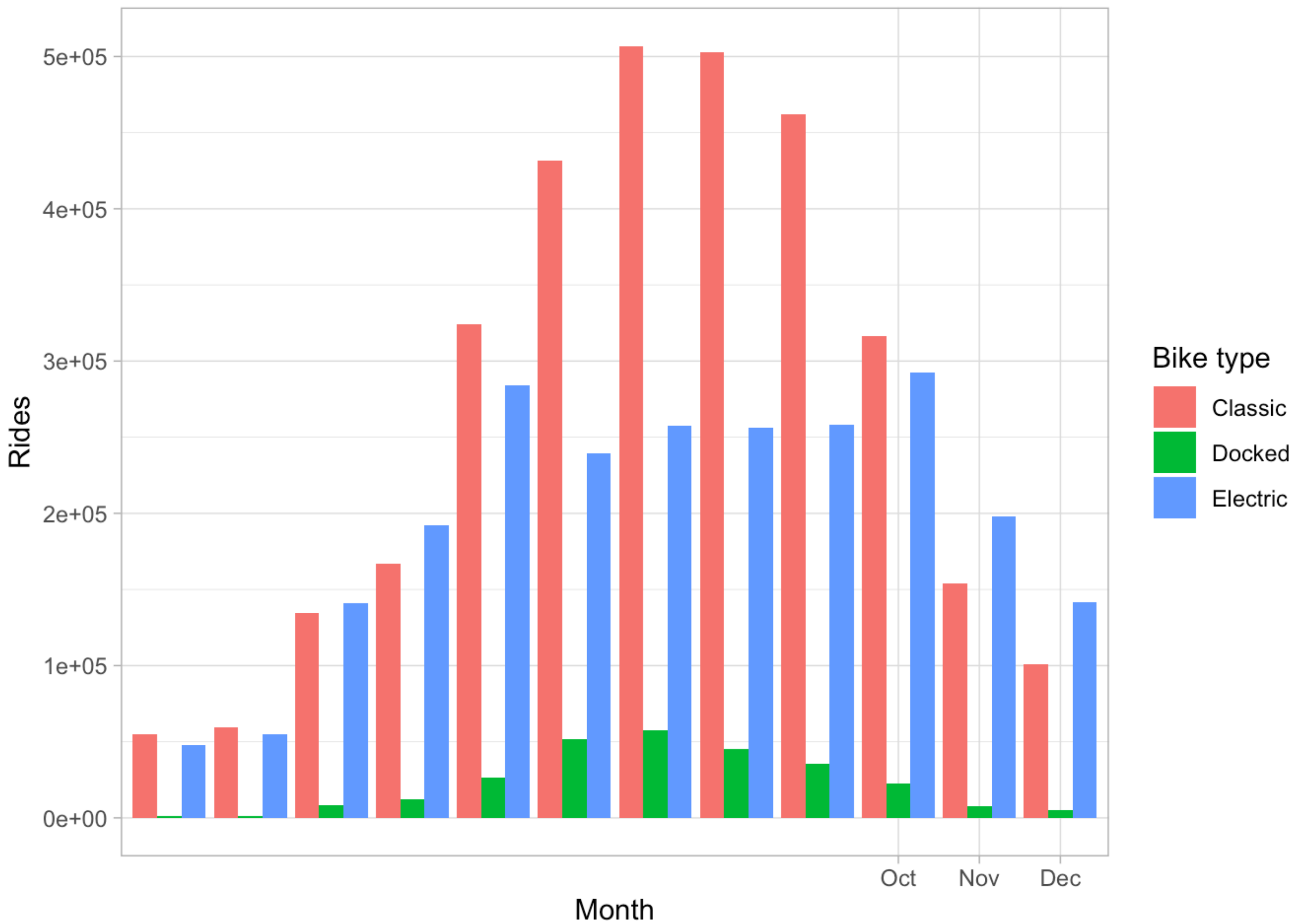
Classic bikes and docked bikes are used more frequently on the weekends than on weekdays. Electric bikes see more even use throughout the week, with the fewest rides on Sundays and the most rides on Saturdays.

```
biketype_day <- all_rides %>%
  group_by(rideable_type) %>%
  mutate(day_of_week = factor(day_of_week, levels = c("Sunday", "Monday", "Tuesday",
"Wednesday", "Thursday", "Friday", "Saturday"))) %>%
  count(day_of_week)
biketype_day_plot <- ggplot(biketype_day, aes(x=day_of_week, y=n, fill=rideable_type)
) + geom_bar(position = "dodge", stat = "identity") + labs(x="Day of Week", y="Rides"
, fill="Bike type") + theme_light()
biketype_day_plot2 <- biketype_day_plot + scale_x_discrete(breaks=c("Sunday", "Monday
", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"), labels=c("Sun","Mon","T
ue","Wed","Thu","Fri","Sat")) + scale_fill_discrete(labels=c("Classic", "Docked", "El
ectric"))
biketype_day_plot2
```



Classic and docked bikes see more use during the summer months, when there's a noticeable increase in activity. Electric bike usage is more even, with relatively consistent use May through October and a dropoff November through April (though usage starts to ramp up in March and April).

```
biketype_month <- all_rides %>%
  group_by(rideable_type) %>%
  count(month)
biketype_month_plot <- ggplot(biketype_month, aes(x=month, y=n, fill=rideable_type))
+ geom_bar(position = "dodge", stat = "identity") + labs(y="Rides", x="Month", fill="
Bike type") + theme_light() + scale_x_discrete(breaks=c(1,2,3,4,5,6,7,8,9,10,11,12),
labels=c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec")) +
scale_fill_discrete(labels=c("Classic", "Docked", "Electric"))
biketype_month_plot
```



Members

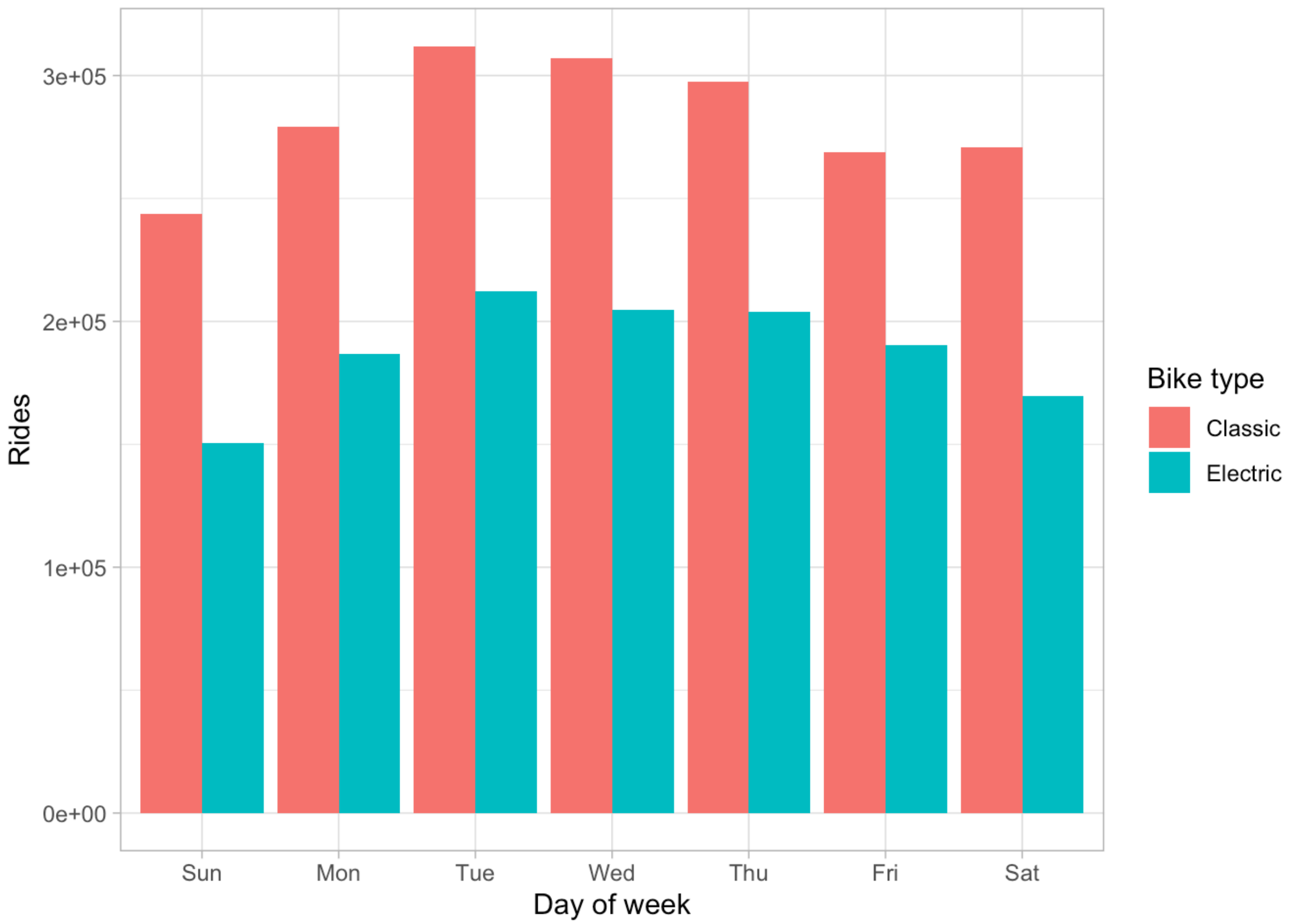
Members are most likely to use classic bikes, though they also make considerable use of electric bikes. Interestingly, they didn't use docked bikes at all during the twelve months under investigation.

```
num_type_m <- members %>% count(rideable_type)
new_row <- c("docked_bike", 0)
num_type_members <- rbind(num_type_m[1:1,], new_row, num_type_m[ 2,])
num_type_members
```

```
## # A tibble: 3 × 2
##   rideable_type n
##   <chr>         <chr>
## 1 classic_bike  1978793
## 2 docked_bike   0
## 3 electric_bike 1318171
```

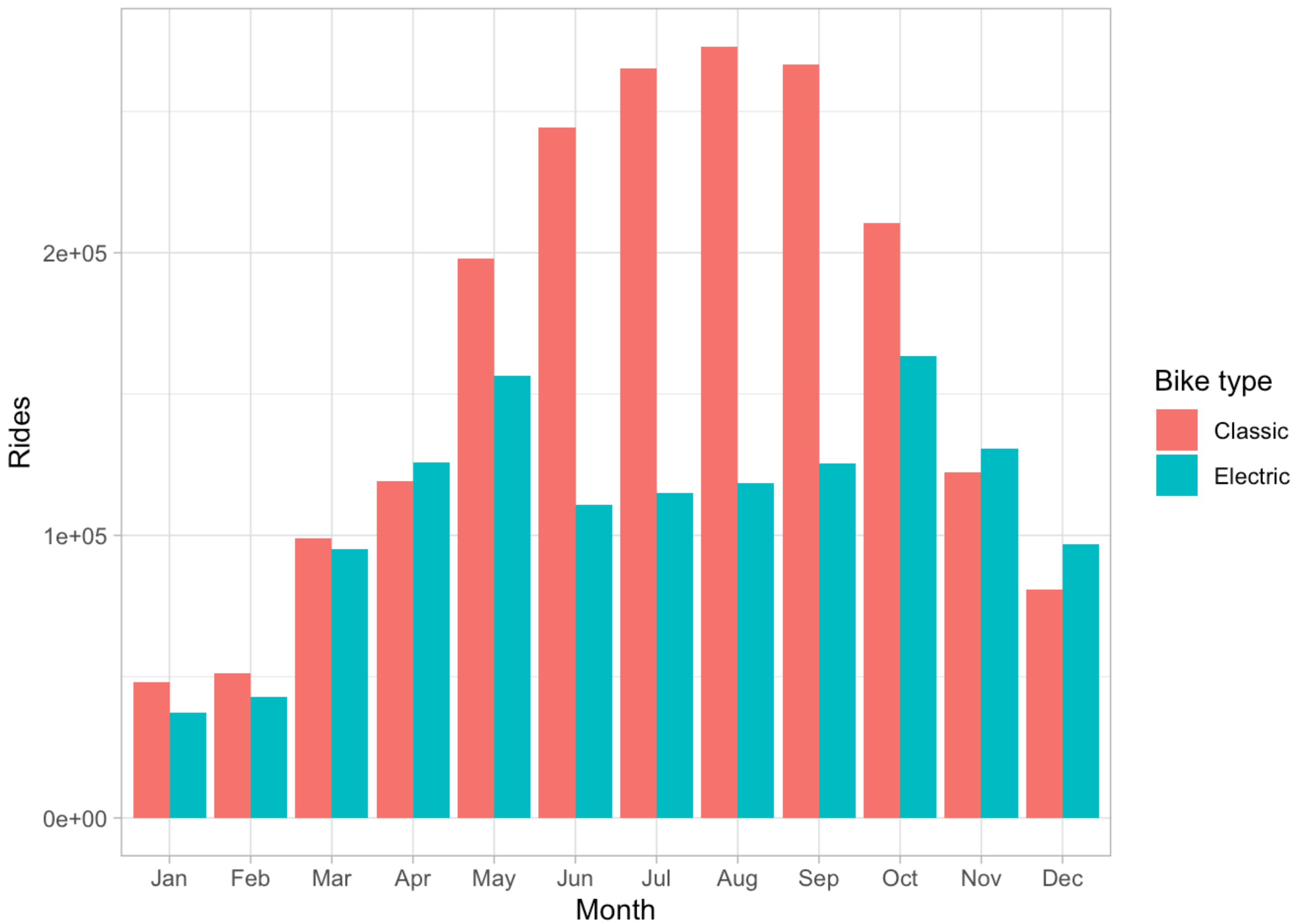
There is no clear difference in day-by-day usage patterns between the two bike types that members favor. Members are more likely to use classic than electric bikes, but they're not more likely to choose one over the other depending on the day of the week.

```
biketype_day_m <- members %>%
  group_by(rideable_type) %>%
  mutate(day_of_week = factor(day_of_week, levels = c("Sunday", "Monday", "Tuesday",
"Wednesday", "Thursday", "Friday", "Saturday"))) %>%
  count(day_of_week)
plot_biketype_day_m <- ggplot(biketype_day_m, aes(x=day_of_week, y=n, fill=rideable_t
ype)) + geom_bar(position = "dodge", stat = "identity") + labs(y="Rides", x="Day of w
eek", fill="Bike type") + theme_light()
plot_biketype_day_m2 <- plot_biketype_day_m + scale_x_discrete(breaks=c("Sunday", "Mo
nday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"), labels=c("Sun","Mon
","Tue","Wed","Thu","Fri","Sat")) + scale_fill_discrete(labels=c("Classic", "Electric
"))
plot_biketype_day_m2
```



Members are more likely to choose classic bikes over electric bikes in the warmer months (June through September).

```
biketype_month_m <- members %>%
  group_by(rideable_type) %>%
  count(month)
plot_biketype_month_m <- ggplot(biketype_month_m, aes(x=month, y=n, fill=rideable_typ
e)) + geom_bar(position = "dodge", stat = "identity") + labs(x="Month", y="Rides", fi
ll="Bike type") + theme_light() + scale_x_discrete(labels=c("Jan","Feb","Mar","Apr","
May","Jun","Jul","Aug","Sep","Oct","Nov","Dec")) + scale_fill_discrete(labels=c("Clas
sic", "Electric"))
plot_biketype_month_m
```



Casual riders

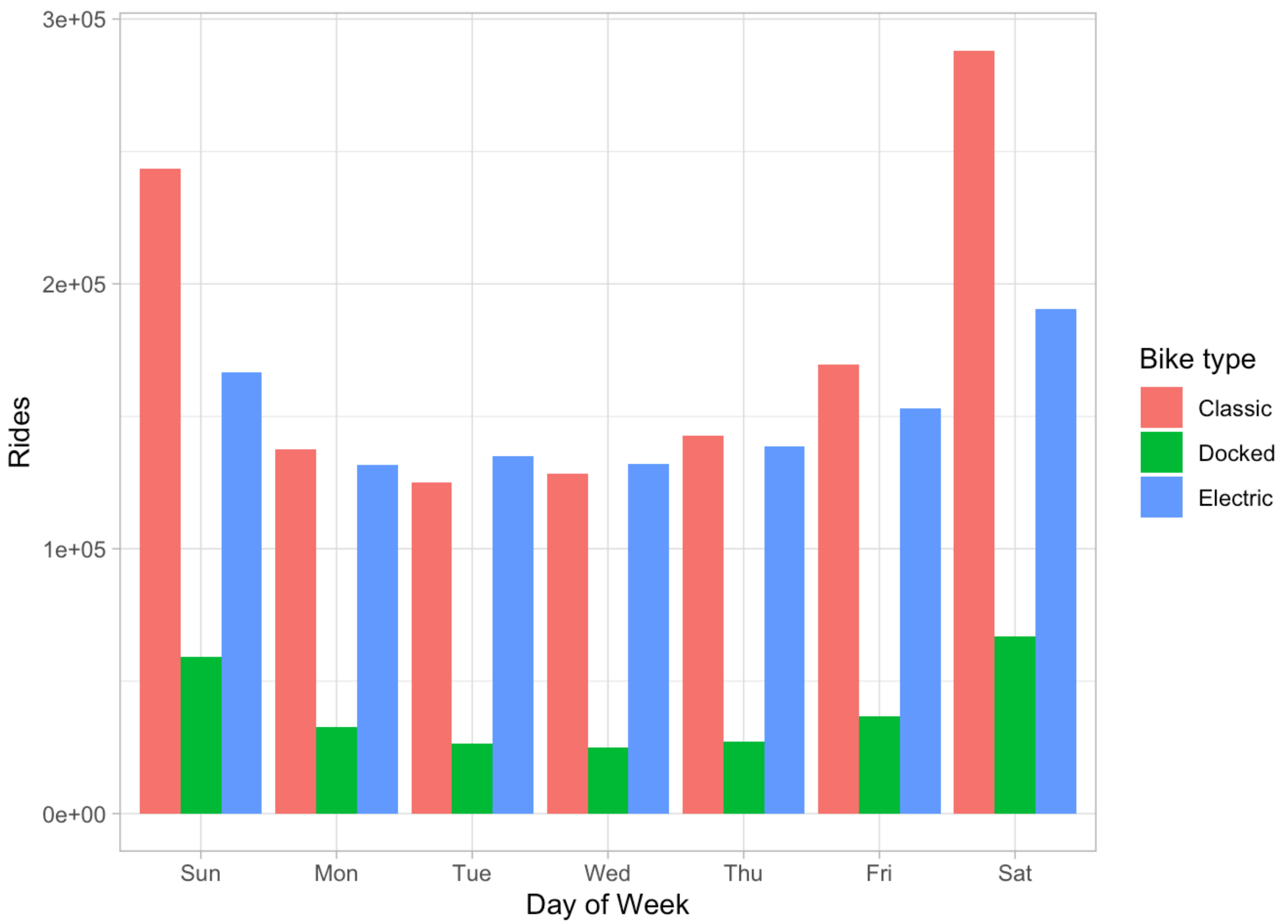
Casual riders use classic bikes most frequently, closely followed by electric bikes. They also use docked bikes, but much less frequently than either of the other two types.

```
num_type_c <- casual %>% count(rideable_type)
num_type_c
```

```
## # A tibble: 3 × 2
##   rideable_type       n
##   <chr>           <int>
## 1 classic_bike  1235142
## 2 docked_bike    274246
## 3 electric_bike 1046757
```

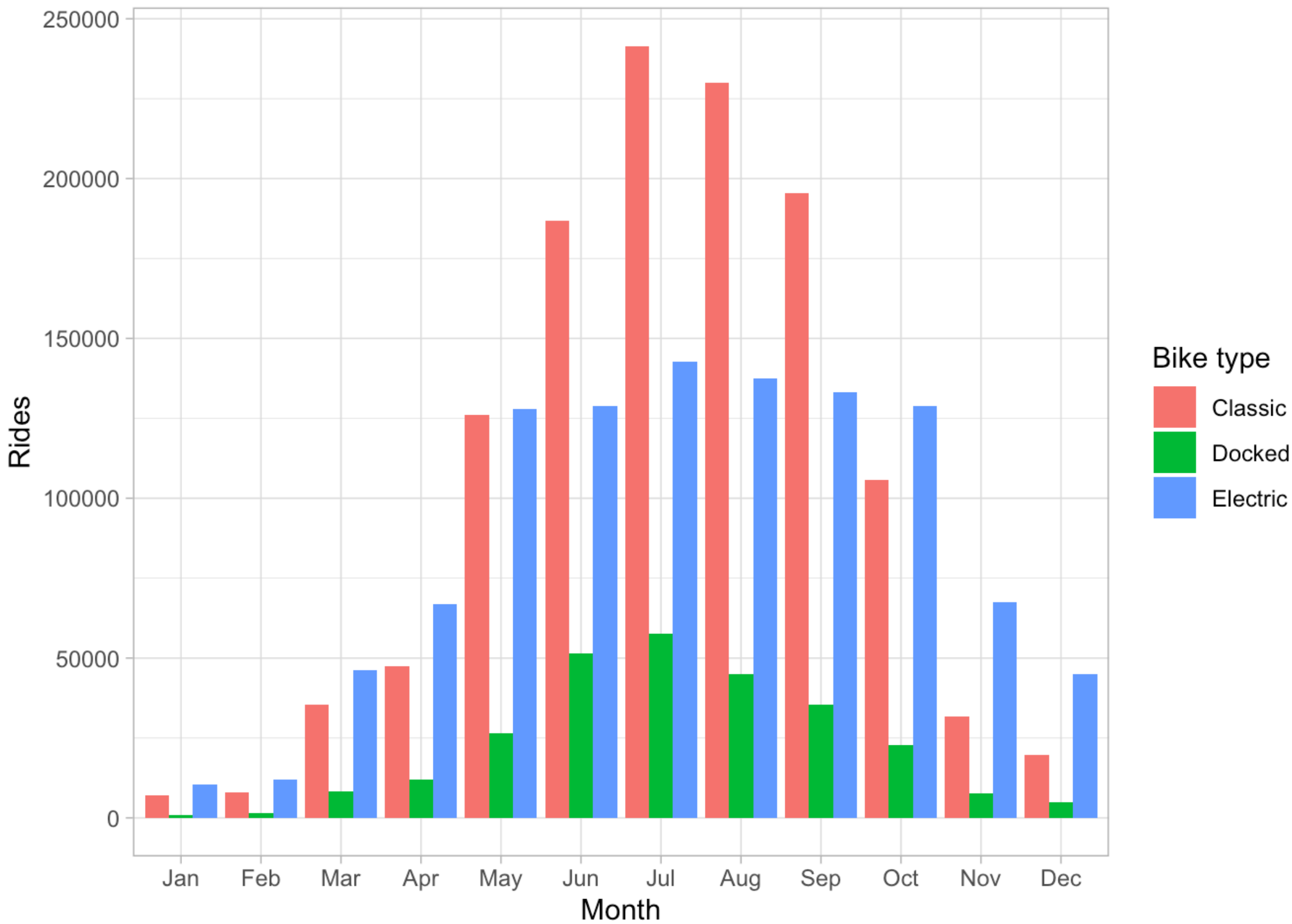
Casual riders use all types of bikes most often on the weekends. The increase in weekend usage is especially strong for classic bikes.

```
biketype_day_c <- casual %>%
  group_by(rideable_type) %>%
  mutate(day_of_week = factor(day_of_week, levels = c("Sunday", "Monday", "Tuesday",
"Wednesday", "Thursday", "Friday", "Saturday"))) %>%
  count(day_of_week)
plot_biketype_day_c <- ggplot(biketype_day_c, aes(x=day_of_week, y=n, fill=rideable_t
ype)) + geom_bar(position = "dodge", stat = "identity") + labs(x="Day of Week", y="Ri
des", fill="Bike type") + theme_light()
plot_biketype_day_c2 <- plot_biketype_day_c + scale_x_discrete(breaks=c("Sunday", "Mo
nday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"), labels=c("Sun","Mon
","Tue","Wed","Thu","Fri","Sat")) + scale_fill_discrete(labels=c("Classic", "Docked",
"Electric"))
plot_biketype_day_c2
```



As already noted, casual riders ride most often during the summer months. The summer months are also when casual riders’ preference is strongly in favor of classic bikes.

```
biketype_month_c <- casual %>%
  group_by(rideable_type) %>%
  count(month)
plot_biketype_month_c <- ggplot(biketype_month_c, aes(x=month, y=n, fill=rideable_typ
e)) + geom_bar(position = "dodge", stat = "identity") + labs(y="Rides", x="Month", fi
ll="Bike type") + theme_light() + scale_fill_discrete(labels=c("Classic", "Docked", "
Electric")) + scale_x_discrete(labels=c("Jan","Feb","Mar","Apr","May","Jun","Jul","Au
g","Sep","Oct","Nov","Dec"))
plot_biketype_month_c
```



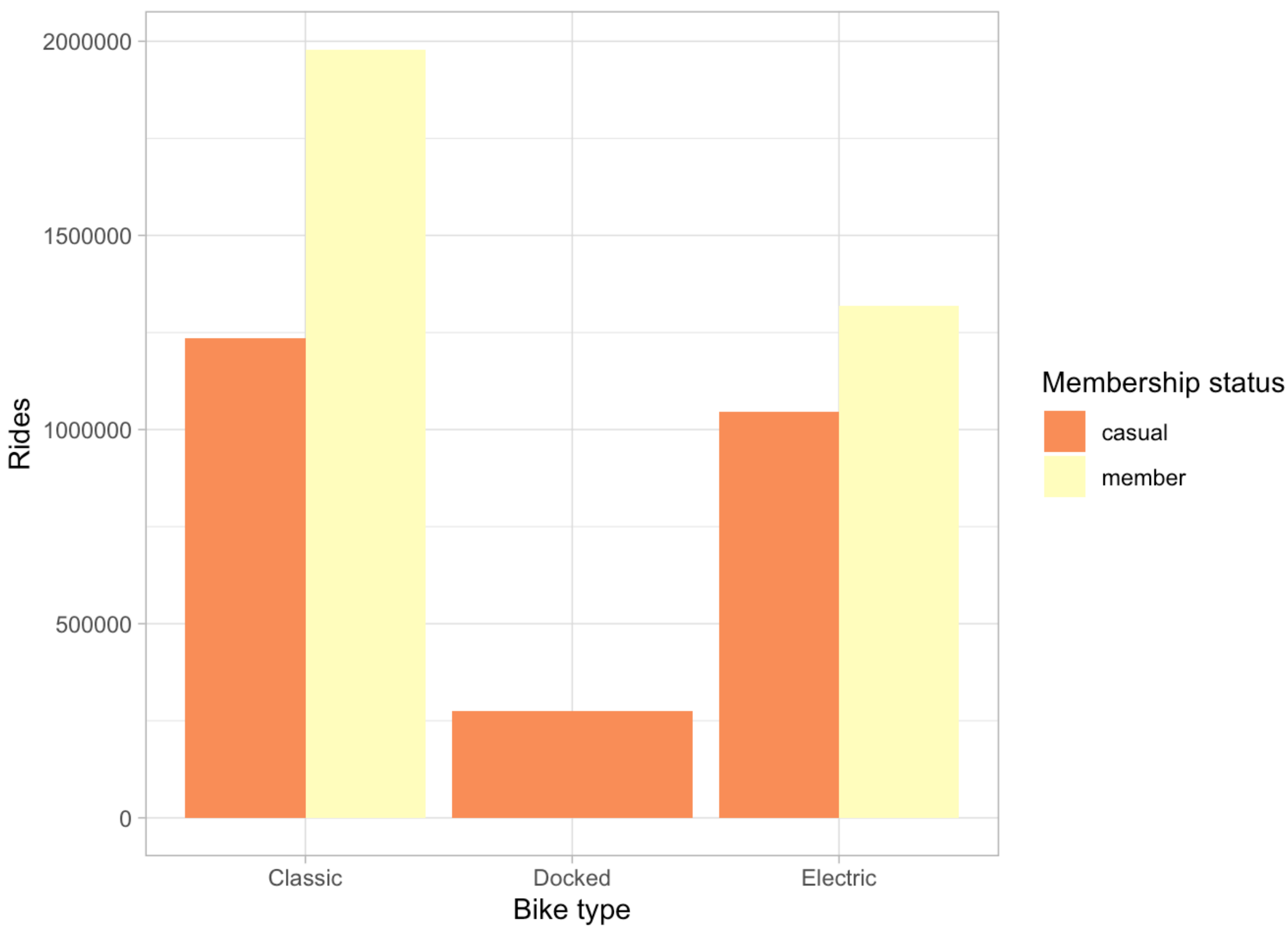
Differences

Only casual riders used docked bikes during the twelve months in question. Members were more likely to use both classic and electric bikes.

```
num_type_compare <- all_rides %>% group_by(member_casual) %>% count(rideable_type)
num_type_compare
```

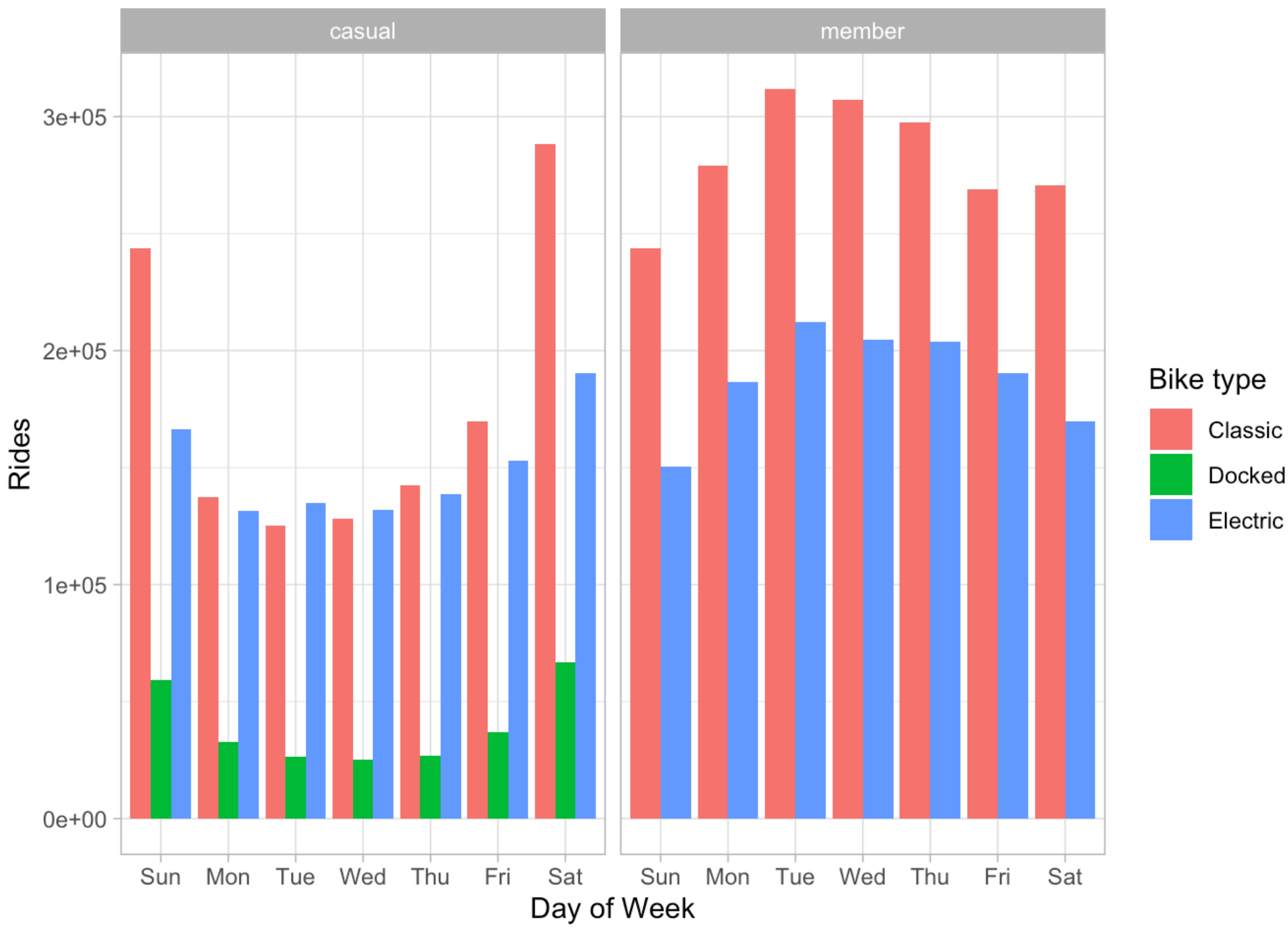
```
## # A tibble: 5 × 3
## # Groups:   member_casual [2]
##   member_casual rideable_type       n
##   <chr>         <chr>           <int>
## 1 casual        classic_bike  1235142
## 2 casual        docked_bike    274246
## 3 casual        electric_bike 1046757
## 4 member        classic_bike  1978793
## 5 member        electric_bike 1318171
```

```
plot_num_type_compare <- ggplot(num_type_compare, aes(x=rideable_type, y=n, fill=memb
er_casual)) + geom_bar(position = "dodge", stat = "identity") + labs(y="Rides", x="Bi
ke type", fill="Membership status") + theme_light() + scale_fill_brewer(palette = "Rd
YlBu") + scale_x_discrete(breaks=c("classic_bike","docked_bike","electric_bike"), lab
els=c("Classic","Docked","Electric"))
plot_num_type_compare
```



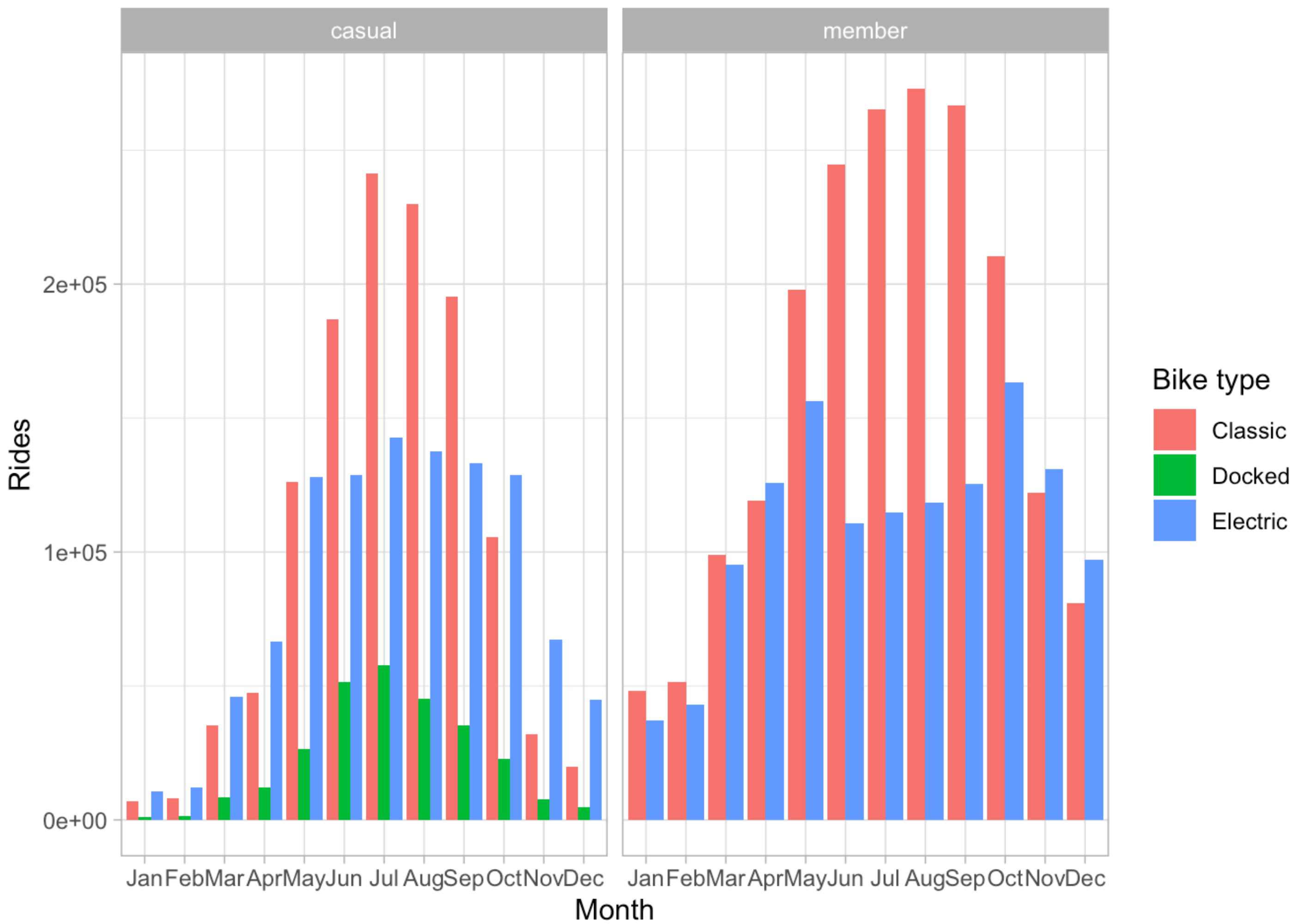
Members use both classic bikes and electric bikes more frequently than casual riders do every day of the week, except for Saturdays.

```
biketype_day_compare <- all_rides %>%
  group_by(member_casual,rideable_type) %>%
  mutate(day_of_week = factor(day_of_week, levels = c("Sunday", "Monday", "Tuesday",
"Wednesday", "Thursday", "Friday", "Saturday"))) %>%
  count(day_of_week)
biketype_day_compare_plot <- ggplot(biketype_day_compare, aes(x=day_of_week, y=n, fil
l=rideable_type)) + geom_col(position = "dodge") + labs(y="Rides", x="Day of Week", f
ill="Bike type") + scale_x_discrete(breaks=c("Sunday", "Monday", "Tuesday", "Wednesda
y", "Thursday", "Friday", "Saturday"), labels=c("Sun","Mon","Tue","Wed","Thu","Fri","
Sat")) + facet_wrap(~member_casual) + theme_light() + scale_fill_discrete(labels=c("C
lassic", "Docked", "Electric"))
biketype_day_compare_plot
```



In general, members use all bike types more frequently in all months than casual riders do (except for docked bikes, which members didn't use at all). The exception: June through September, casual riders used electric bikes more frequently than members did.

```
biketype_month_compare <- all_rides %>%
  group_by(member_casual,rideable_type) %>%
  count(month)
biketype_month_compare_plot <- ggplot(biketype_month_compare, aes(x=month, y=n, fill=
rideable_type)) + geom_col(position = "dodge") + labs(y="Rides", x="Month", fill="Bik
e type") + facet_wrap(~member_casual) + theme_light() + scale_fill_discrete(labels=c(
"Classic", "Docked", "Electric")) + scale_x_discrete(labels=c("Jan","Feb","Mar","Apr"
,"May","Jun","Jul","Aug","Sep","Oct","Nov","Dec"))
biketype_month_compare_plot
```



Key findings

As seen in the above analysis, annual members ride most often during the work week, and least often on weekends. In contrast, casual riders ride least during the work week, and notably more frequently on weekends.

This pattern suggests that annual members likely use Cyclistic bikes to commute to work, while casual riders use them recreationally. Their average ride times (approximately 13 minutes for members and just under 31 minutes for casual riders) also lend support to this idea.

Variances in usage by month also suggests that casual members use bikes recreationally while annual members use them primarily for commuting. Casual members' use increases as the months get warmer, increasing sharply in May, peaking in July, and falling off substantially in November. Annual members also ride more frequently during the warmer months, but their summer usage doesn't show the same sharp peak seen with casual riders, and their uptick in usage begins in March. While annual members do show a decrease in usage during November and December, it is nowhere near as dramatic a drop-off as seen with casual riders.

In general, both casual riders and annual members prefer classic bikes to electric bikes. Unsurprisingly, casual riders are most likely to use classic bikes on the weekends, suggesting recreational use. Their use of electric bikes increases much less on the weekends, suggesting that, when casual riders use electric bikes, they may be using them to commute.

There is a clear difference in casual riders' choice of bike type by month. For classic bikes, there is sharp increase beginning in May, peaking in July, and then falling off in colder months. This is the same pattern seen in casual riders' general usage patterns.

Casual riders' choice of electric bikes, however, follows a different pattern. There is still a sharp increase in May and a clear decline in November, but their use of electric bikes May through October is fairly steady, though it's generally lower than casual riders' classic bike use during those months, particularly in June through September. In May and October, however, casual riders are much more likely to choose electric bikes than classic bikes. Though they ride much less frequently in March, April, November, and December, they're far more likely to choose electric bikes in those months.

Recommendations

Currently, Cyclistic offers only one type of membership: annual. Casual riders' usage patterns suggest that they could be converted to members if additional membership types were introduced.

Recommendation 1

Offer weekend-only memberships. Casual riders' use of Cyclistic bikes spikes on the weekends. They're likely to find a weekend membership attractive.

Recommendation 2

Offer summer-only memberships. The summer months are when casual riders are most likely to make use of Cyclistic bikes.

Recommendation 3

Offer winter memberships for use of electric bikes. During the coldest months of the year, casual riders consistently choose electric bikes over classic bikes.