Portfolio1

Evan O'Keefe

2/20/2022

```
city <- read_csv("https://raw.githubusercontent.com/evanokeefe/Stat-479-Portfolio-1/main/Eco-Totem_Capit
  mutate(Count_Date = parse_date_time(Count_Date, "mdy HM"),
         Date = date(Count_Date),
         Time = hour(Count_Date)) %>%
  select(Date, Time, Count)
```

```
## Parsed with column specification:
## cols(
##   Count_Date = col_character(),
##   Count = col_double(),
##   OBJECTID = col_double()
## )
```

```
sw <- read_csv("https://raw.githubusercontent.com/evanokeefe/Stat-479-Portfolio-1/main/Eco-Totem_Southwe
    mutate(Count_Date = parse_date_time(Count_Date, "mdy HM"),
         Date = date(Count_Date),
         Time = hour(Count_Date)) %>%
  select(Date, Time, Count)
```

```
## Parsed with column specification:
## cols(
##   Count_Date = col_character(),
##   Count = col_double(),
##   OBJECTID = col_double()
## )
```

```
temp <- read_csv("https://raw.githubusercontent.com/evanokeefe/Stat-479-Portfolio-1/main/madison-weathe
```

```
## Parsed with column specification:
## cols(
##   year = col_double(),
##   Jan = col_double(),
##   Feb = col_double(),
##   Mar = col_double(),
##   Apr = col_double(),
##   May = col_double(),
##   Jun = col_double(),
##   Jul = col_double(),
##   Aug = col_double(),
##   Sep = col_double(),
##   Oct = col_double(),
##   Nov = col_double(),
##   Dec = col_double()
## )
```

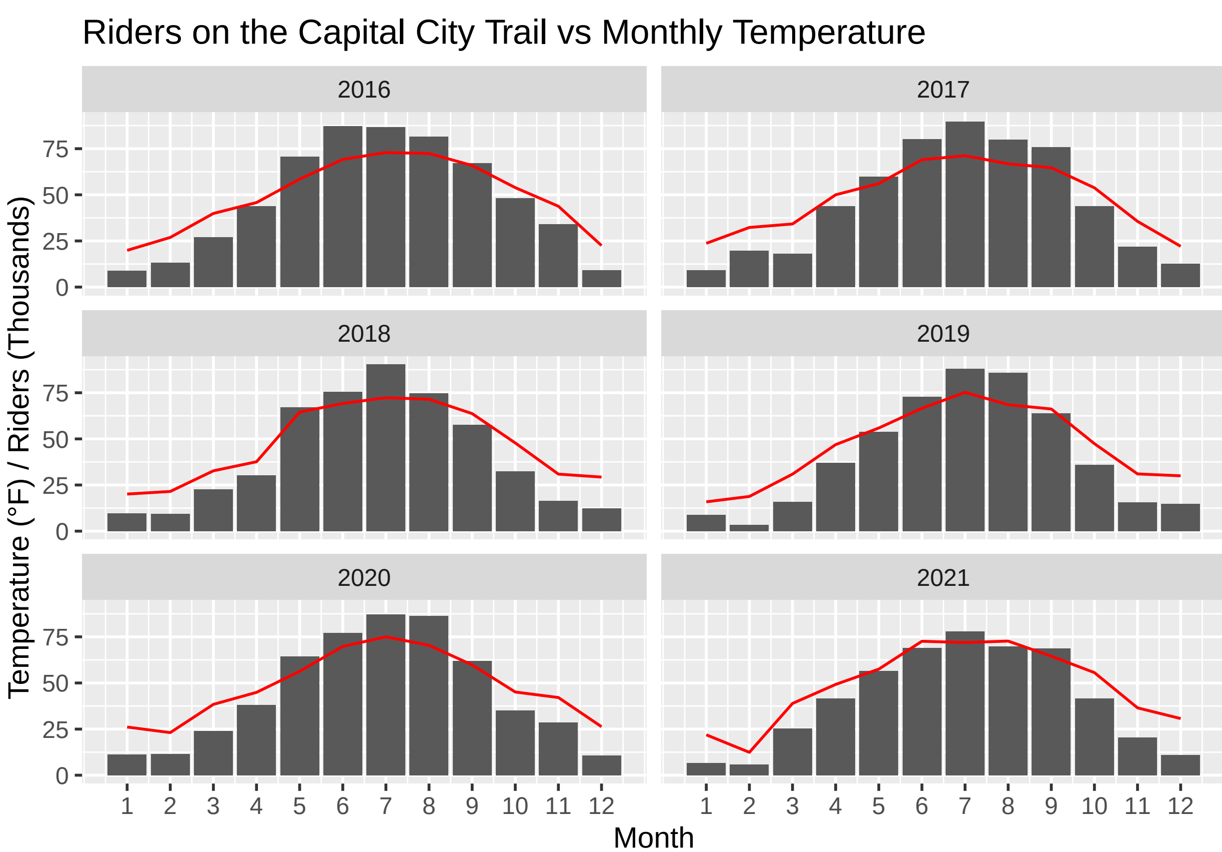
```r
sum_city <- city %>%
  mutate(m = month.abb[month(Date)],
         y = year(Date)) %>%
  group_by(y, m) %>%
  summarise(count = sum(Count), x = mean(month(Date)))
```

```
## `summarise()` regrouping output by 'y' (override with `.groups` argument)
```

```r
temp2 <- temp %>%
  rename(y = year) %>%
  pivot_longer(Jan:Dec, names_to = "m", values_to = "temp")

df <- merge(temp2, sum_city) %>%
  filter(y > 2015) %>%
  arrange(x)


ggplot(data = df) +
  geom_col(aes(x = x, y = count/1000)) +
  geom_line(aes(x = x, y = temp, group = y), color = "red") +
  facet_wrap(~y, ncol = 2) +
  scale_x_continuous(breaks = 1:12) +
  labs(
    title = "Riders on the Capital City Trail vs Monthly Temperature",
    x = "Month",
    y = "Temperature (°F) / Riders (Thousands)"
  )
```

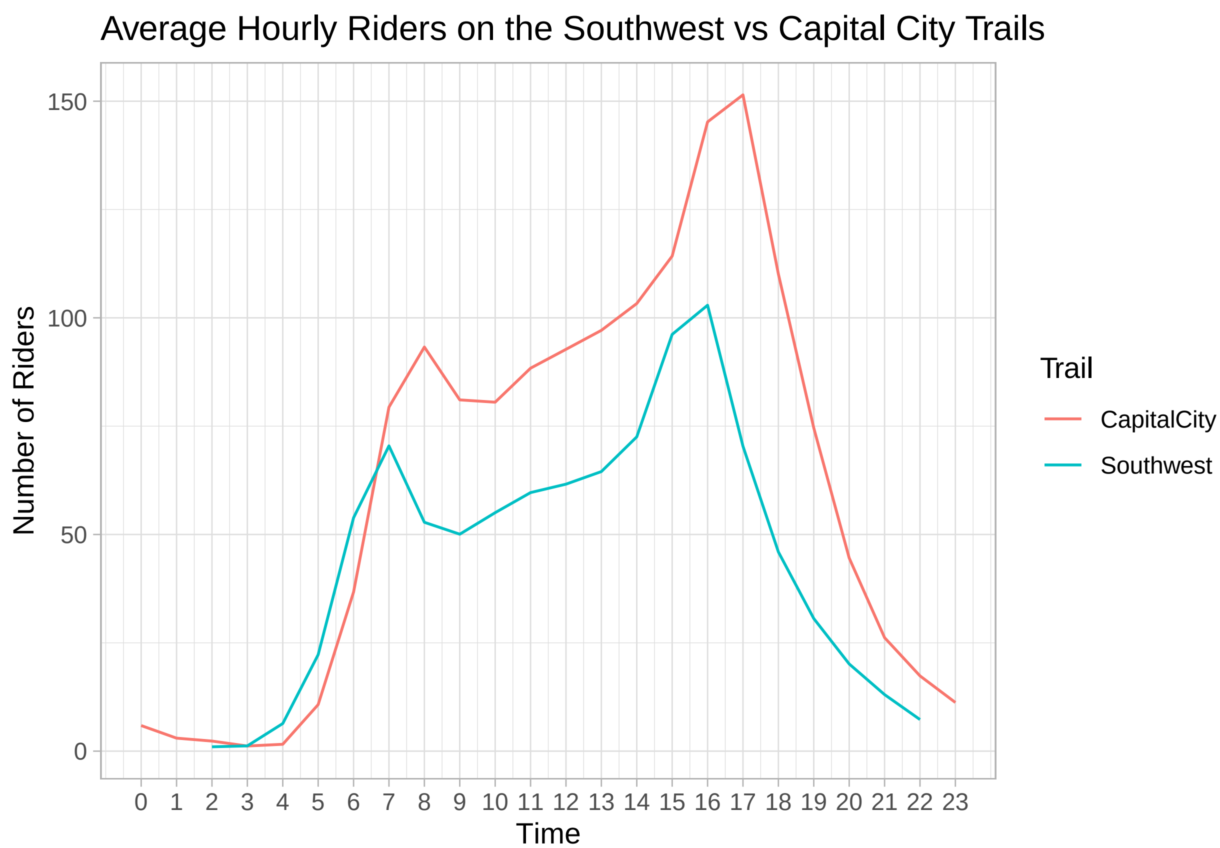


```r
timez <- city %>%
  rename(city_count = Count) %>%
  merge(sw) %>%
  group_by(Time) %>%
  summarise(CapitalCity = mean(city_count), Southwest = mean(Count)) %>%
  pivot_longer(CapitalCity:Southwest, names_to = "trail", values_to = "count")
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
ggplot(timez) +
  geom_line(aes(x = Time, y = count, color = trail)) +
  labs(
    title = "Average Hourly Riders on the Southwest vs Capital City Trails",
    y = "Number of Riders",
    color = "Trail"
  ) +
  scale_x_continuous(breaks = 0:23) +
  theme_light()
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```



I chose to work with Madison bike data because for nearly everything promotional for the city and university, one of the top mentioned items is the number of bike trails. Given how cold it gets in Madison, I chose my first visualization to see how the trail ridership varies month by month. Very much to expectation the ridership increases significantly in the summer. I additionally chose to facet by year to see if there were any interesting outliers that I could pick out. I expected that there would be some increase over the past year or two due to the pandemic and activities such as biking and hiking were said to be good socially distanced activities. However, when looking at the data ridership remained constant through 2020 and declined slightly in 2021. One interesting thing from these plots is the temperature and monthly rider totals work out well on the same scale (when riders are presented in thousands). This neat coincidence allowed the visualization to be much cleaner.

For my second visualization I was curious to see how ridership varied over time on the Capital City Trail vs the Southwest Commuter Path (the two bike paths with bike counters). I initially expected the two graphs to have different peaks since the Southwest path tends to be a primary route to classes, while the Capital City trail is more leisure and commuter. However, when I looked into where the trail counter is located, the results made more sense. The Southwest counter is located on Monroe which is on the edge of where students live, so this data is largely uncaptured. If the counter were located by East Campus Mall or Randall Street I would expect to see a much larger plateau in the middle of the day as students use the path for passing times. On the other hand, the Capital City trail counter in on John Nolan, so it captures much more commuter traffic than I initially thought.