Capital Metro Data

2024-08-18

Visual story telling part 2: Capital Metro data

Chat GPT was employed to create each of the following plots

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

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

```
cap = read.csv('/Users/krummelha/Downloads/capmetro_UT.csv')
cap%>%glimpse
```

```
## Rows: 5,824
## Columns: 8
## $ timestamp   <chr> "2018-09-01 06:00:00", "2018-09-01 06:15:00", "2018-09-01 ~
## $ boarding    <int> 0, 2, 3, 3, 2, 4, 3, 8, 4, 7, 2, 1, 7, 10, 11, 9, 19, 16, ~
## $ alighting   <int> 1, 1, 4, 4, 4, 4, 12, 4, 15, 10, 4, 12, 15, 9, 15, 22, 26,~
## $ day_of_week <chr> "Sat", "Sat", "Sat", "Sat", "Sat", "Sat", "Sat", "Sat", "S~
## $ temperature <dbl> 74.82, 74.82, 74.82, 74.82, 74.39, 74.39, 74.39, 74.39, 75~
## $ hour_of_day <int> 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10~
## $ month       <chr> "Sep", "Sep", "Sep", "Sep", "Sep", "Sep", "Sep", "Sep", "S~
## $ weekend     <chr> "weekend", "weekend", "weekend", "weekend", "weekend", "we~
```

```r
# convert timestamp to an actual datetime
# Convert timestamp to a datetime object if it's not already
cap$timestamp <- ymd_hms(cap$timestamp)

# Create a total ridership variable
cap$total_ridership <- cap$boarding + cap$alighting

# Plot total ridership over time
ggplot(cap, aes(x = timestamp, y = total_ridership)) +
  geom_line(color = "blue", alpha = 0.7) +
  labs(title = "Total Ridership Over Time",
       x = "Time",
       y = "Total Ridership") +
  theme_minimal()
```
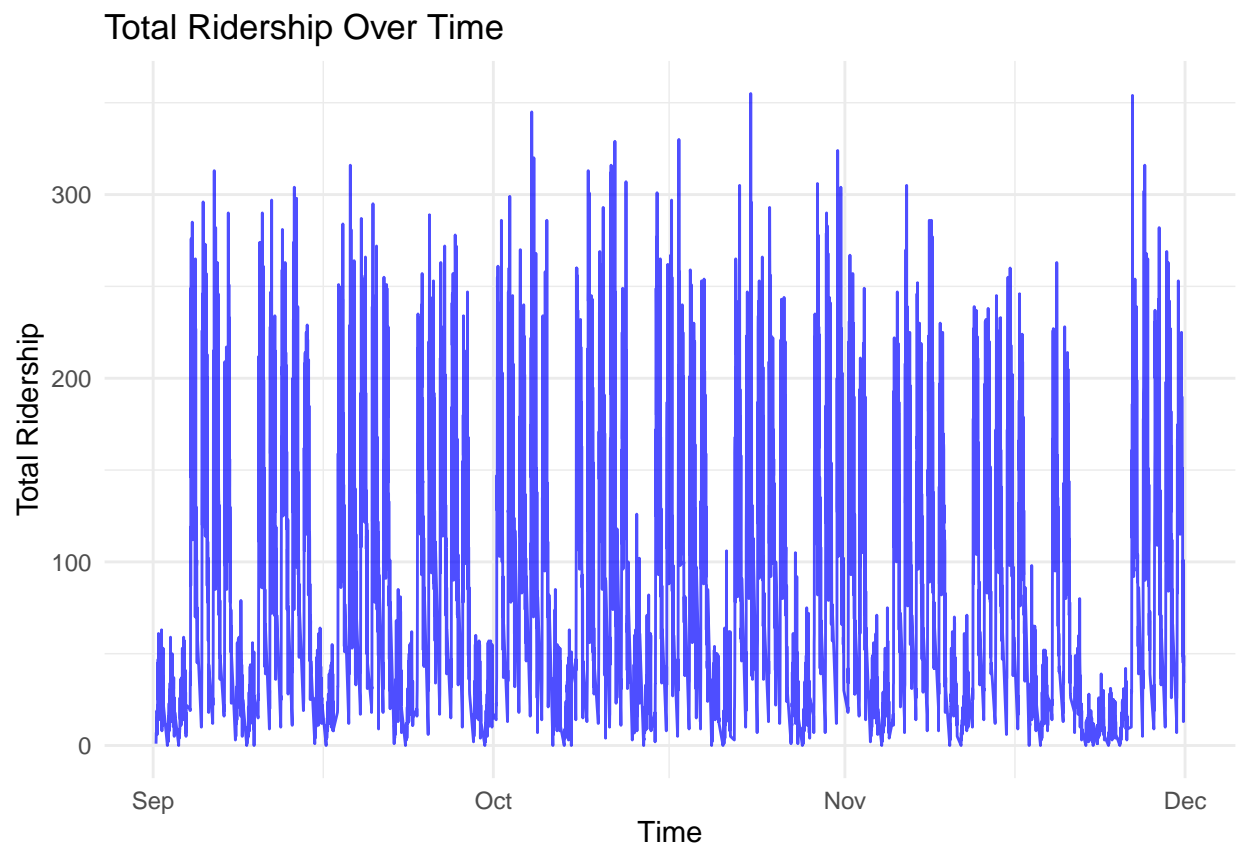

The above plot is interesting in that it demonstrates how the bus ridership decreases immensely at the end of November, during Thanksgiving break. Moreover, it is interesting to note that each of the small valleys within each month represent the low bus ridership on the weekends (to be discussed in more depth below). There is also a noticeable reduction in bus ridership during mid November, which is when the weather becomes nice enough in Texas to walk outside without sweating through your clothes (Temperature to be discussed in more depth later).

```r
# Average ridership by day of the week
ggplot(cap, aes(x = day_of_week, y = total_ridership)) +
  stat_summary(fun = "mean", geom = "bar", fill = "darkorange") +
  labs(title = "Average Ridership by Day of the Week",
```

```
        x = "Day of the Week",
        y = "Average Total Ridership") +
  theme_minimal()
```

Average Ridership by Day of the Week
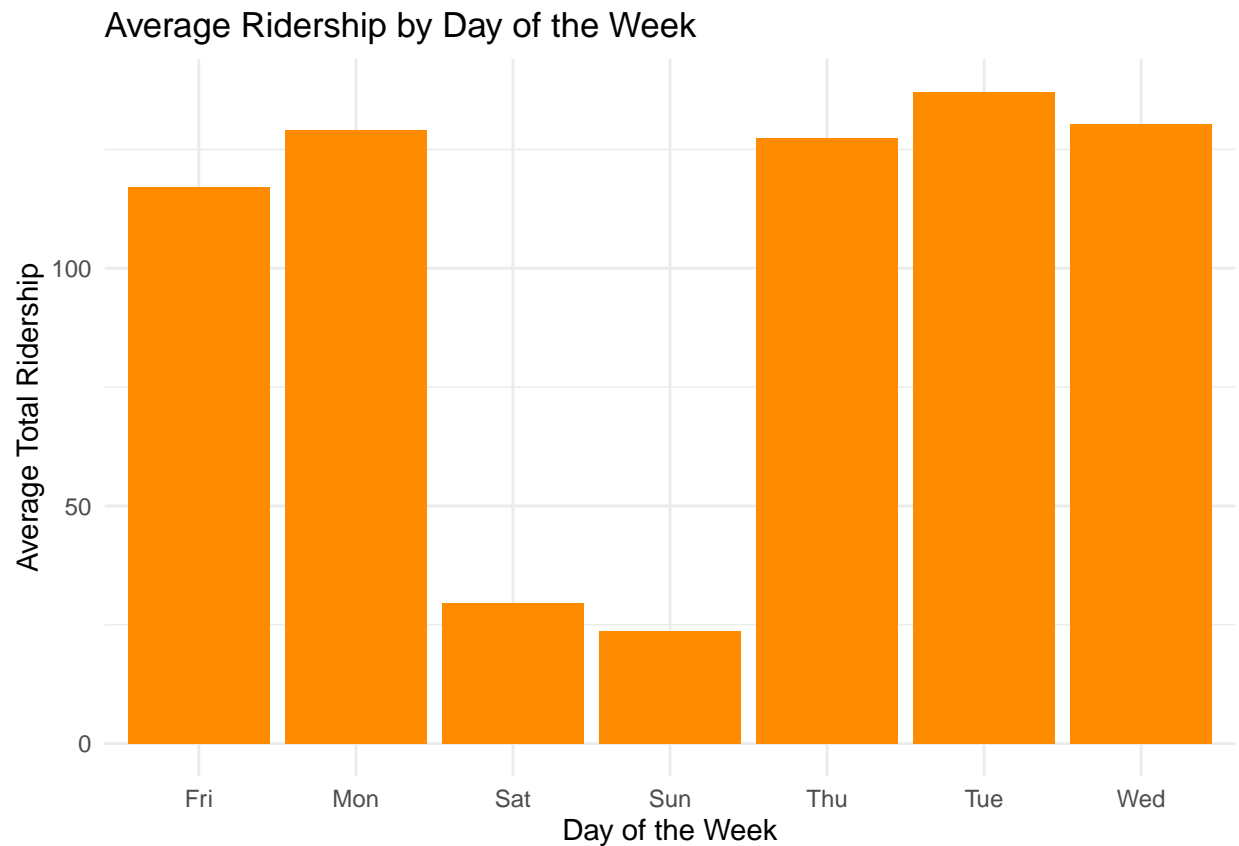
The above graph demonstrates the drastic difference between Ridership on the weekends vs. the weekday. Based on this plot, it seems as though people ride the bus heavily during the weekday, most likely to get to work, and on their free time, choose to stay home or utilize other methods of transportation in lieu of the bus sytem.

```
# Create a heatmap of ridership by hour of day and day of the week
ggplot(cap, aes(x = hour_of_day, y = day_of_week, fill = total_ridership)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  labs(title = "Heatmap of Ridership by Hour and Day",
       x = "Hour of the Day",
       y = "Day of the Week") +
  theme_minimal()
```

Heatmap of Ridership by Hour and Day
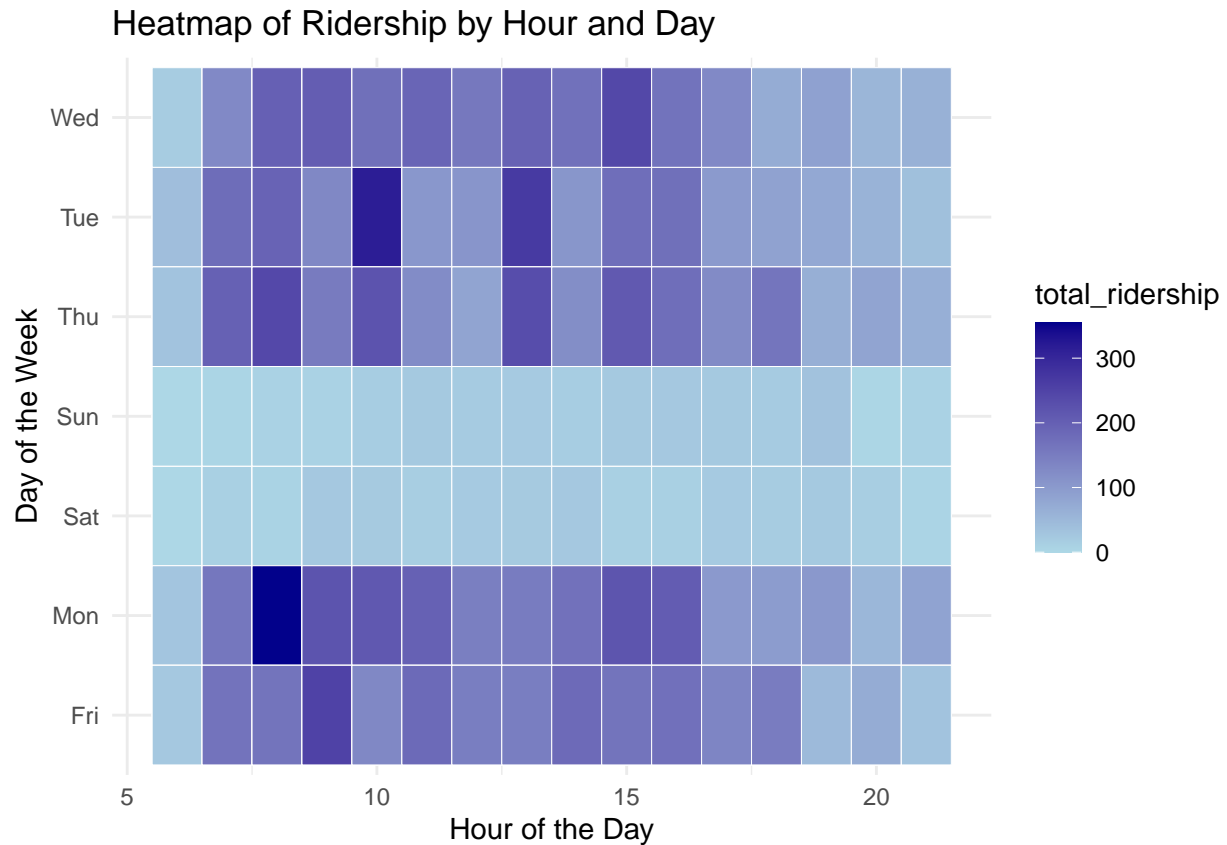
The above graph gives another great example of how drastic the bus ridership is on the weekends vs. the weekdays. It seems as though there is some fluctuation on the weekends with the midday hour having the most bus usage for those days. 7am on Mondays seems to be the highest traffic time of the entire dataset, with 9am on Tuesdays following closely behind. Moreover, this heatmap demonstrates that the bus is used much less often at night across the board.

```
# Plot ridership vs temperature
ggplot(cap, aes(x = temperature, y = total_ridership)) +
  geom_point(alpha = 0.4, color = "purple") +
  geom_smooth(method = "loess", color = "red") +
  labs(title = "Ridership vs. Temperature",
       x = "Temperature (°F)",
       y = "Total Ridership") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Ridership vs. Temperature
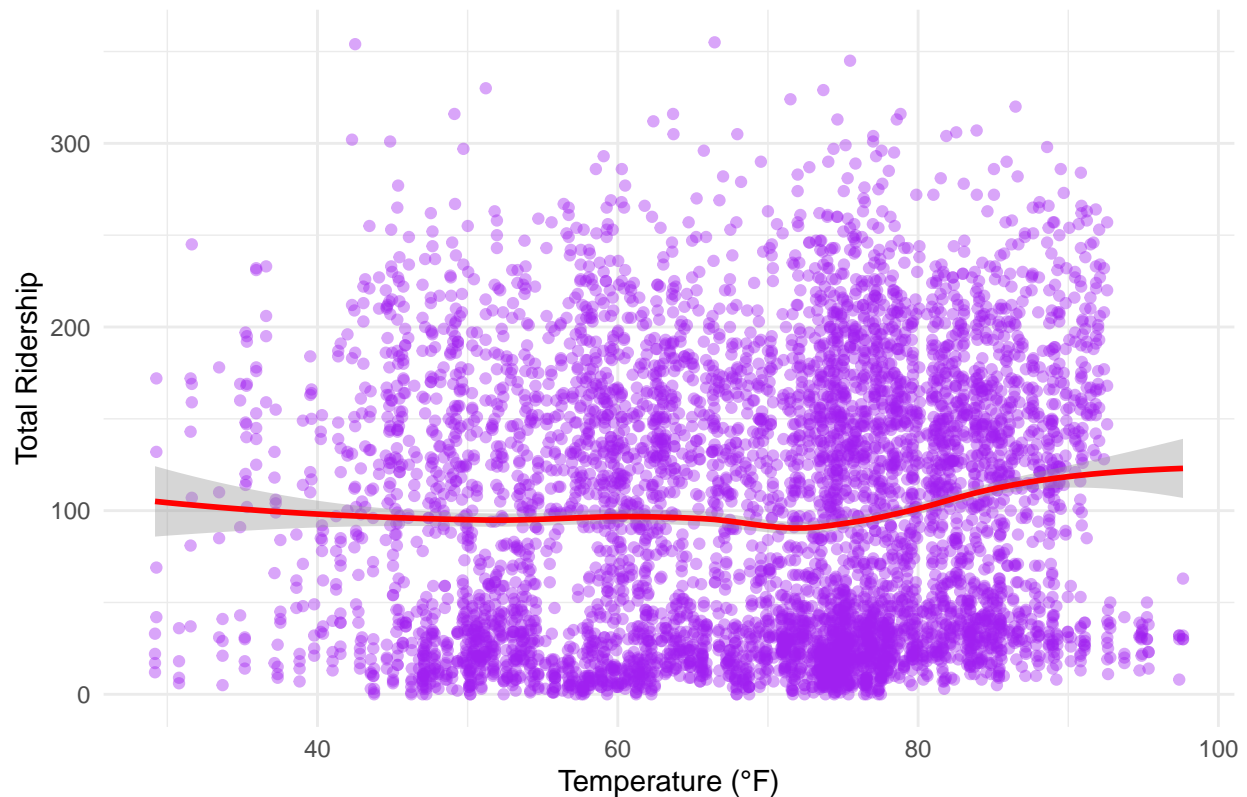
This last graph demonstrates how bus ridership increases with extreme temperature shifts. There is an obvious dip where people choose to walk when it is nice out, around 72 degrees, and there is a moderate peak beyond this point with increasing temperature, and a very slight slope left with decreasing temperature. This demonstrates that people like to ride the bus to escape the extreme traveling conditions that they might otherwise experience walking to their destination. As this is Texas, there is less consistent data during colder temperatures.

Final Conclusions

Based on the plots we provided above, the capital metro experiences a very regular bus ridership pattern with increased ridership during weekdays at the beginning of the week, higher ridership in extreme temperatures, and lower bus ridership over Thanksgiving break.