

SDS315 HW 2

Zachary Brunell (EID - ztb456)

2025-01-28

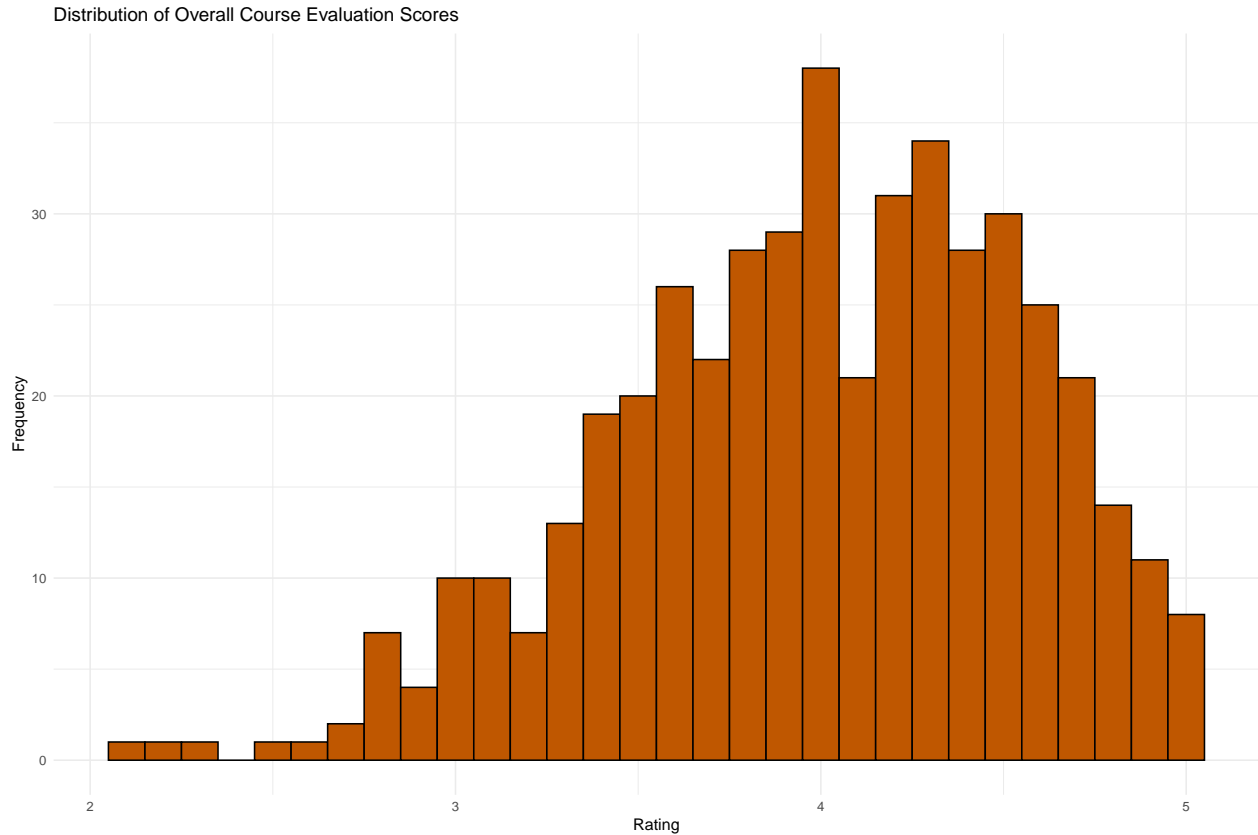
Contents

GitHub Link	2
Problem 1: Beauty, or not, in the classroom	2
Part A	2
Part B	3
Part C	4
Part D	5
Problem 2: bike sharing	6
Part A	6
Part B	7
Part C	8
Problem 3 - Capital Metro UT Ridership	9
Part A	9
Part B	10
Problem 4: Wrangling the Billboard Top 100	11
Part A	11
Part B	12

GitHub Link

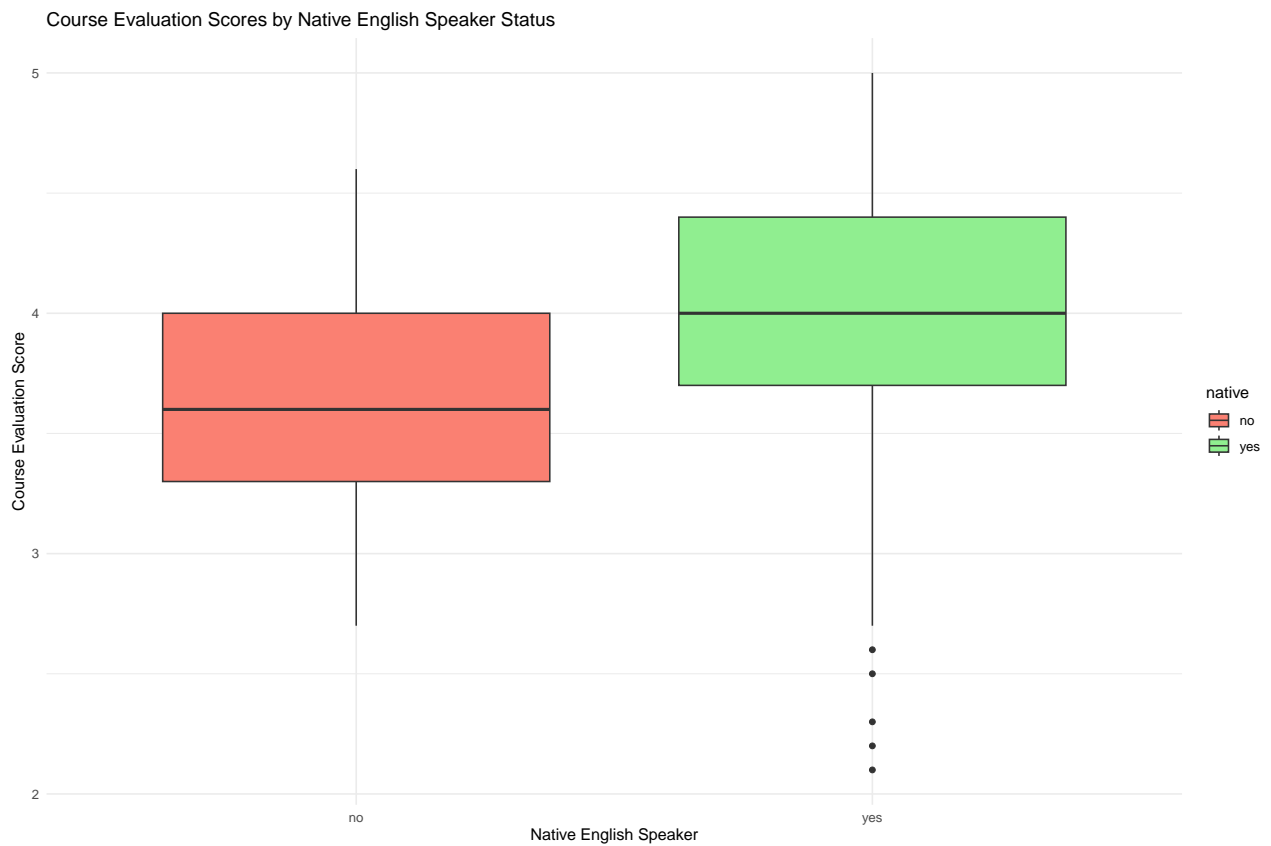
Problem 1: Beauty, or not, in the classroom

Part A



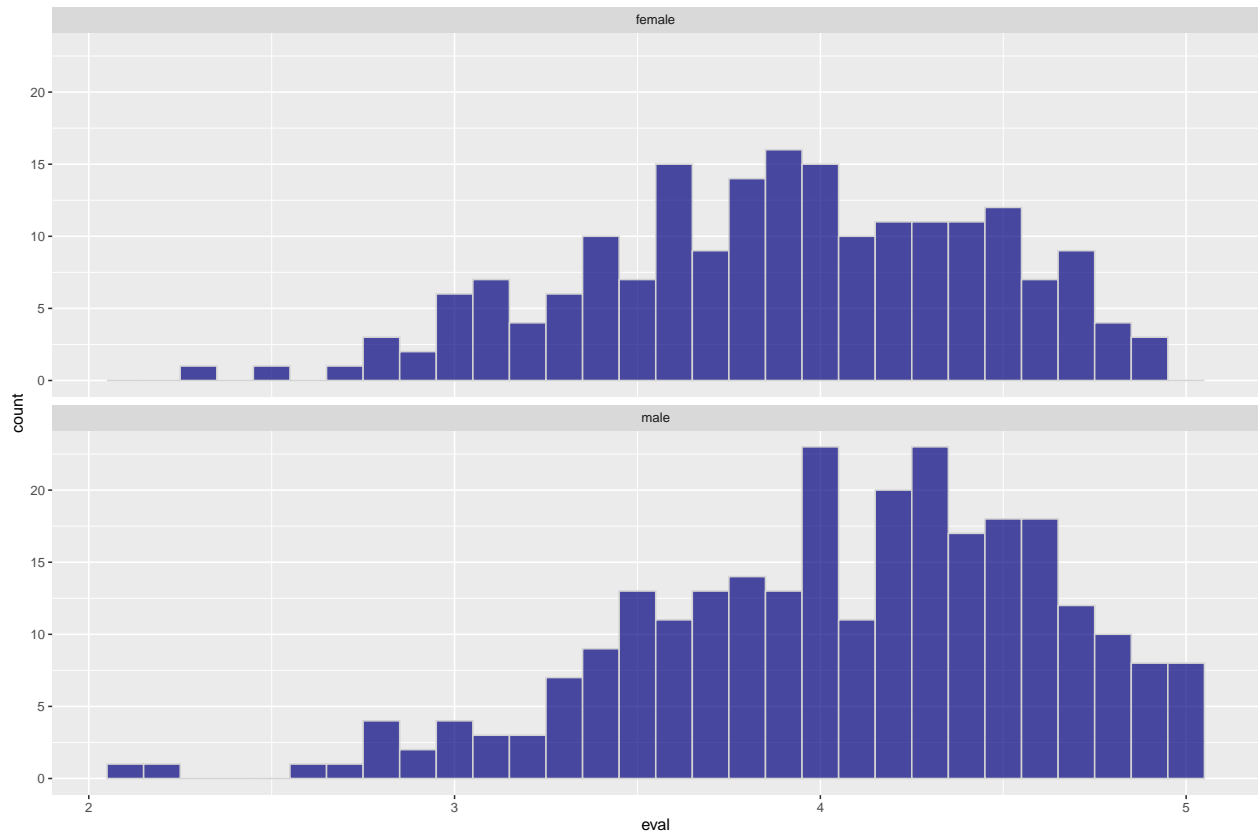
This histogram demonstrates the distribution of course evaluation scores. The majority of courses fall between a rating of 3.0 and 4.5, with a noticeable concentration around the 4.0 mark. This is evidenced by the mean, which is 4. The distribution is somewhat right-skewed, indicating that students generally rated courses positively.

Part B



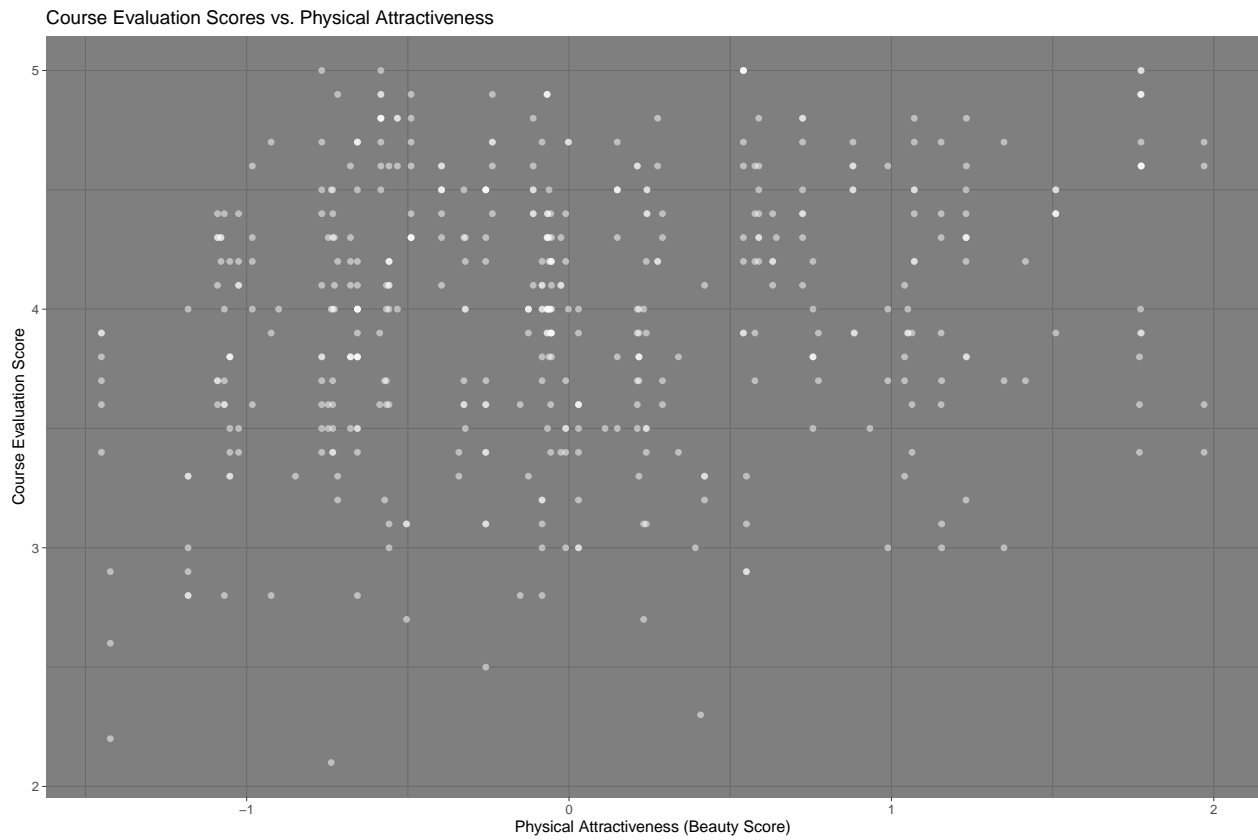
This boxplot compares course evaluation scores for instructors who are native English speakers versus those who are not. It suggests that native English-speaking instructors tend to receive slightly higher scores on average.

Part C



This faceted histogram shows the distribution of course evaluation scores separated by instructor gender. The graph demonstrates that female teachers have more variation than their male counterparts, but both maintain similar a average score of around 4.0

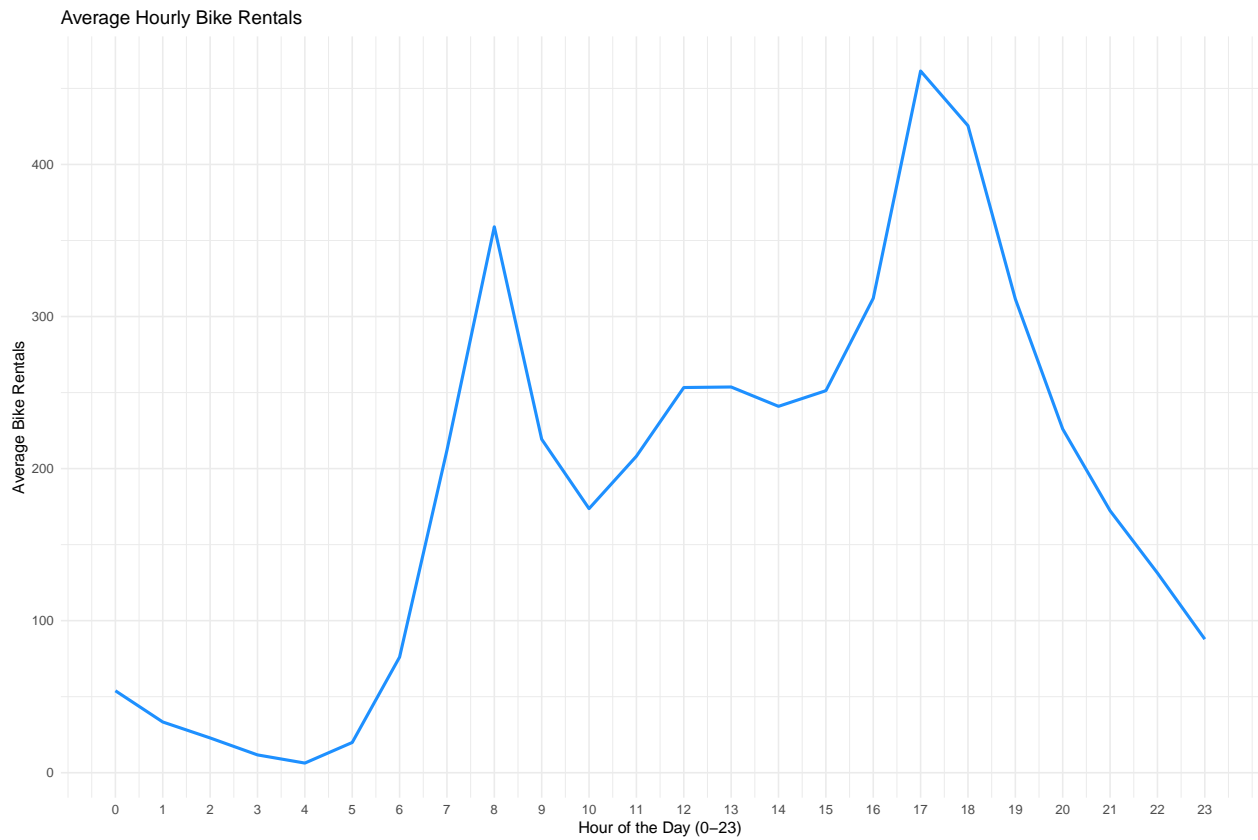
Part D



This scatter plot shows the relationship between physical attractiveness (beauty score) and course evaluation scores. The correlation of 0.1890391 indicates a weak, positive relationship between a professor's attractiveness and their evaluation score.

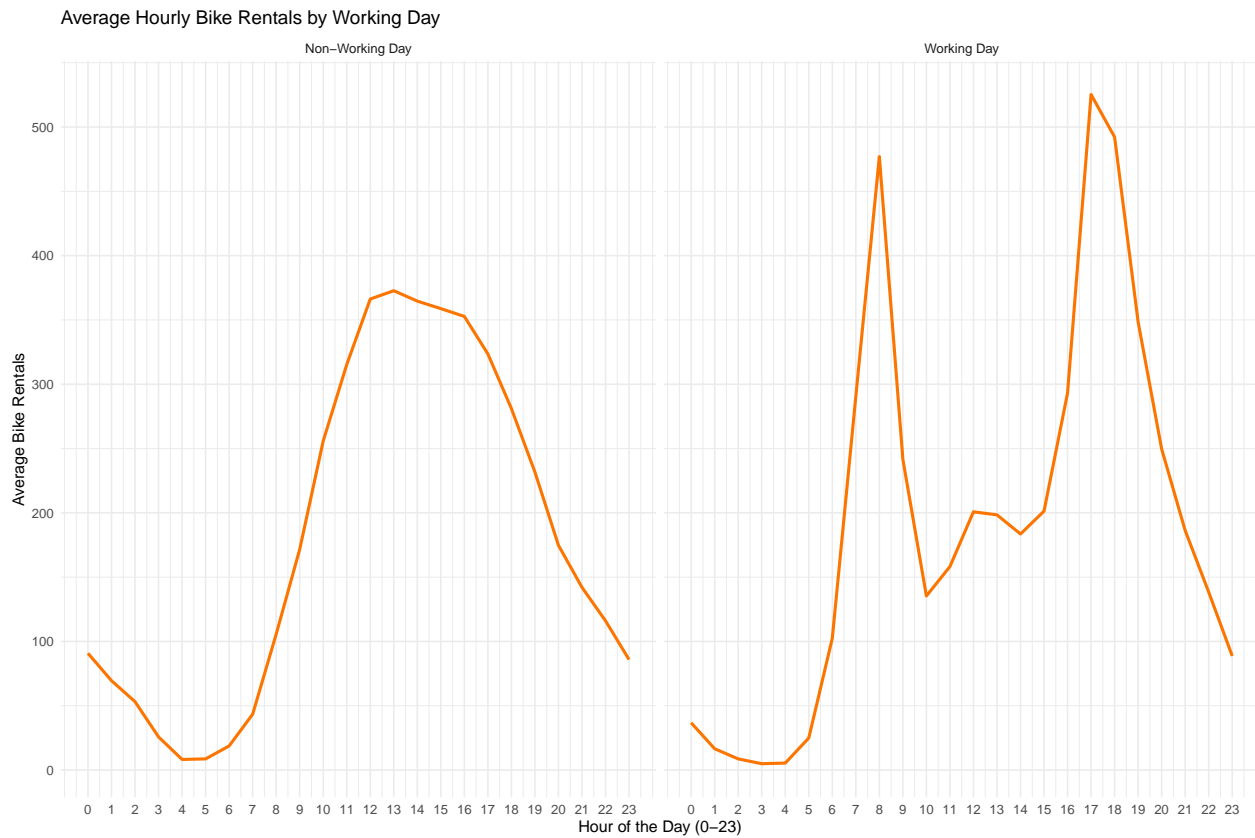
Problem 2: bike sharing

Part A



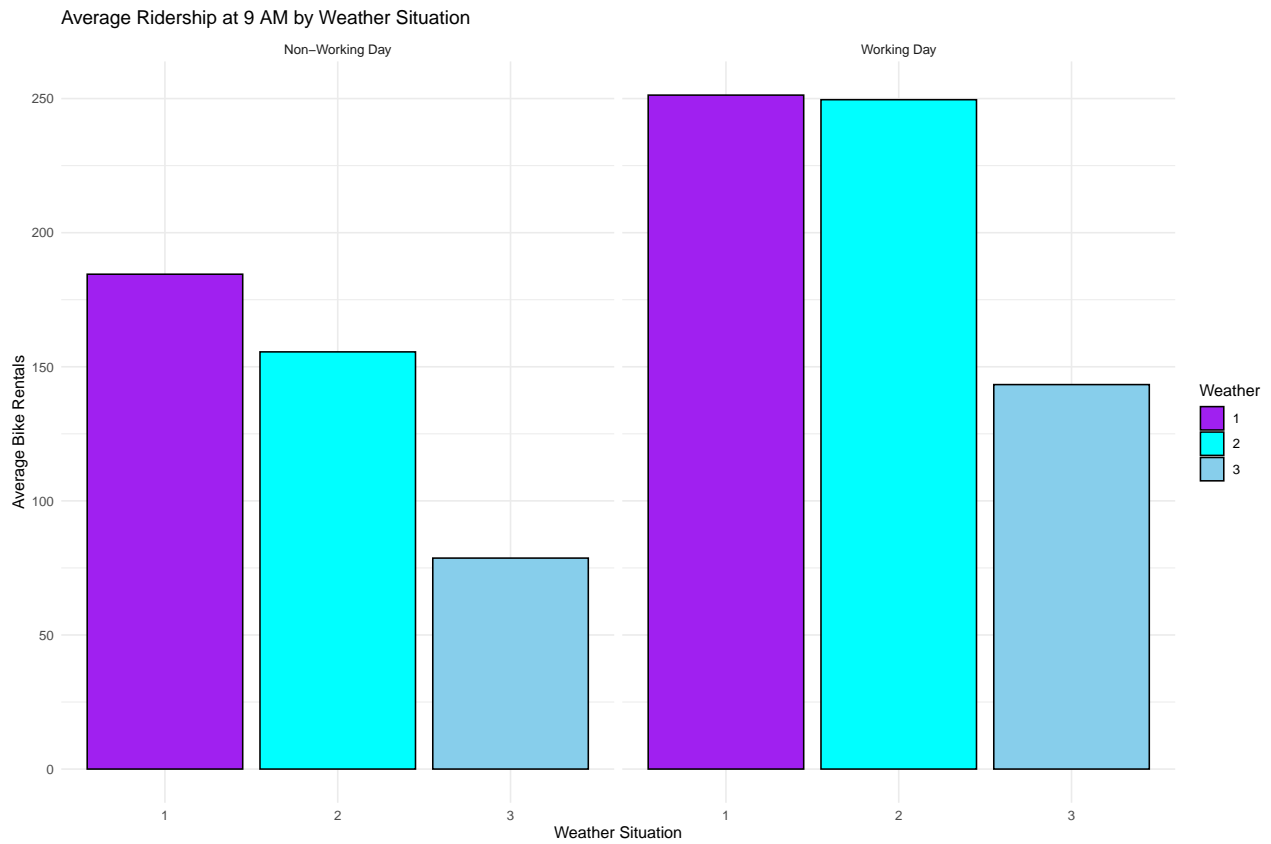
The above line plot illustrates the average number of bike rentals by hour of day. The data reveals significant peaks during the morning (7-9 AM) and evening (4-6 PM) commute times, which lower activity in the night. These patterns reflect commuter behaviors.

Part B



The plot displays bike rentals by hour, separated into working and non-working days. On working days, rental patterns have sharp peaks during morning and evening commutes, while non-working days exhibit a gradual increase as the day progresses to noon, and a gradual decrease afterwards. This suggests that on working days, the bulk of the rentals occur during commute times, while on non-working days, this occurs around (11 AM - 2 PM). The difference highlights the impact of work schedules on bike usage.

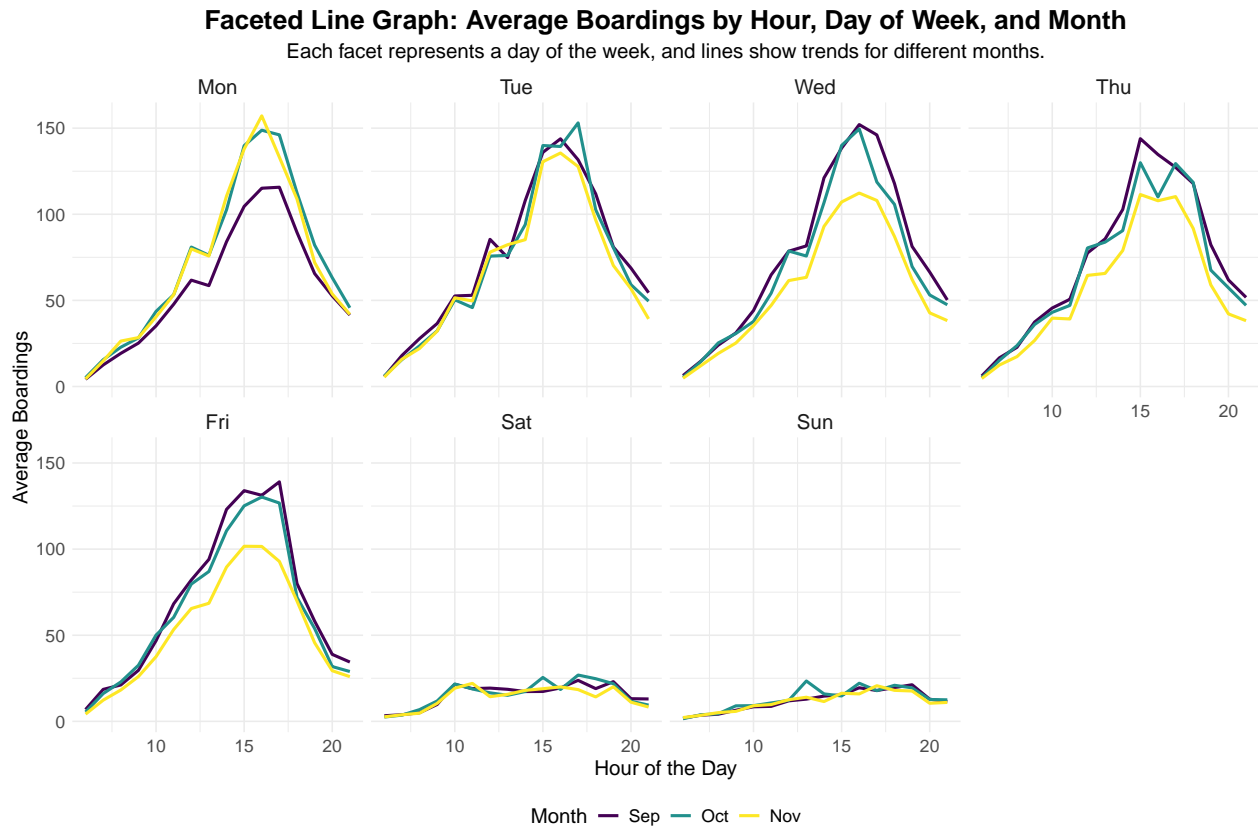
Part C



This bar plot shows the average number of bike rentals at 9 AM under different weather conditions, split by working and non-working days. Favorable weather (clear or partly cloudy) correlates with higher ridership, especially on non-working days, indicating a preference for outdoor activities in good weather. Additionally, poor weather has less of an effect on rentals when it is a working day, as many will still rely on them to commute. This means rentals for working days is less dependent on weather than that of non-working days.

Problem 3 - Capital Metro UT Ridership

Part A

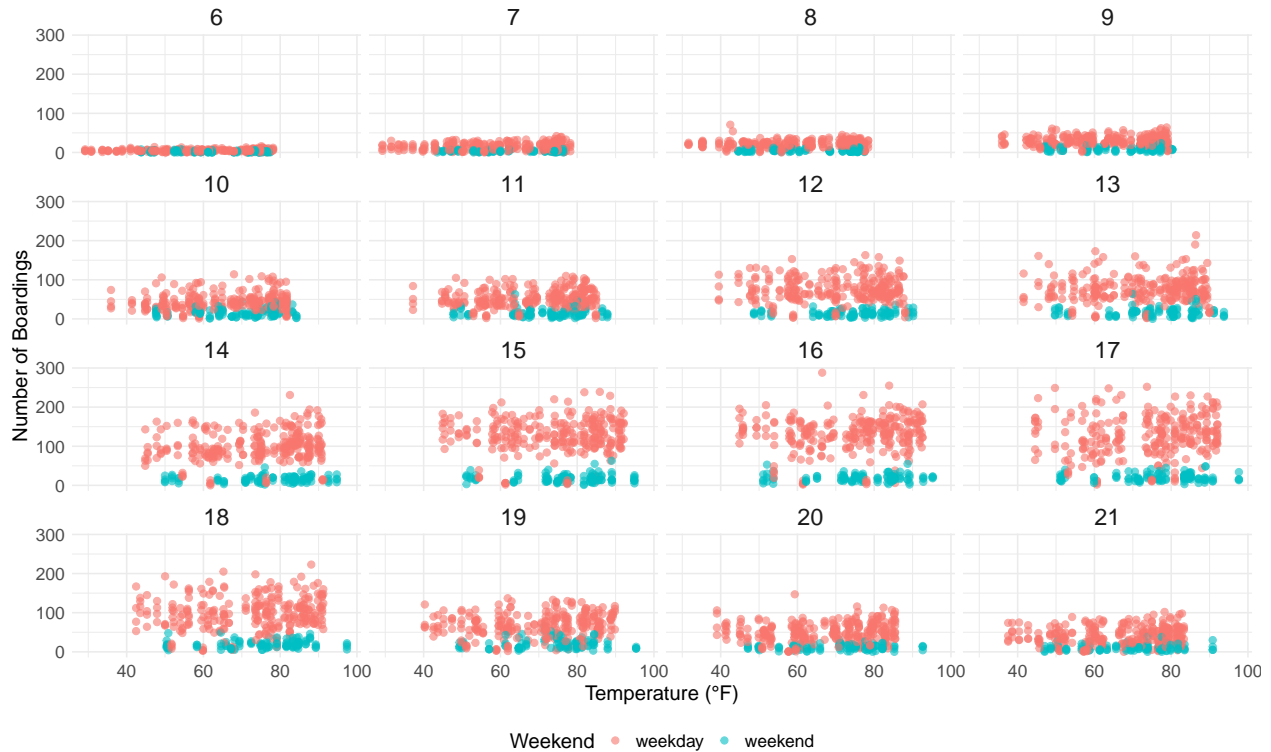


The faceted line graph demonstrates the average boardings per hour at UT Austin for each day of the week across September, October, and November. Boardings peak in the morning (7-9 AM) and early evening (4-6 PM) on weekdays, aligning with common class schedules and commute times. The trends are far less pronounced on the weekends, with peaks around midday. September shows slightly higher boardings overall compared to the other months, likely due to the start of the semester and the weather.

Part B

Faceted Scatter Plot: Boardings vs. Temperature, Colored by Weekday/Weekend

Each facet represents an hour of the day. Points are colored by weekend status.



The scatter plot showcases the relationship between temperature (in degrees Fahrenheit) and boardings, faceted by hour of the day. On weekdays, boardings are higher during peak hours (7-9 AM and 4-6 PM) regardless of the temperature. On weekends, boardings are more varied but decrease significantly at extreme temperatures (very hot/cold). Temperatures around 60-80 degrees correspond to the highest boarding numbers.

Problem 4: Wrangling the Billboard Top 100

Part A

```
## # A tibble: 10 x 3
##   performer      song      count
##   <chr>         <chr>    <int>
## 1 Imagine Dragons Radioactive    87
## 2 AWOLNATION     Sail      79
## 3 Jason Mraz     I'm Yours    76
## 4 The Weeknd     Blinding Lights 76
## 5 LeAnn Rimes    How Do I Live  69
## 6 LMFAO Featuring Lauren Bennett & GoonRock Party Rock Anthem 68
## 7 OneRepublic    Counting Stars 68
## 8 Adele          Rolling In The Deep 65
## 9 Jewel          Foolish Games/You Were Meant~ 65
## 10 Carrie Underwood Before He Cheats 64

## Table: Top 10 Most Popular Songs Since 1958 by Number of Weeks on the Billboard Top 100 Chart.
```

This table lists the top ten most popular songs from 1958 to 2021, ranked by number of weeks on the Billboard Top 100 chart. “Performer” indicates the artist, “Song” is the title of the track, and “Count” represents the total number of weeks the song stayed on the chart.

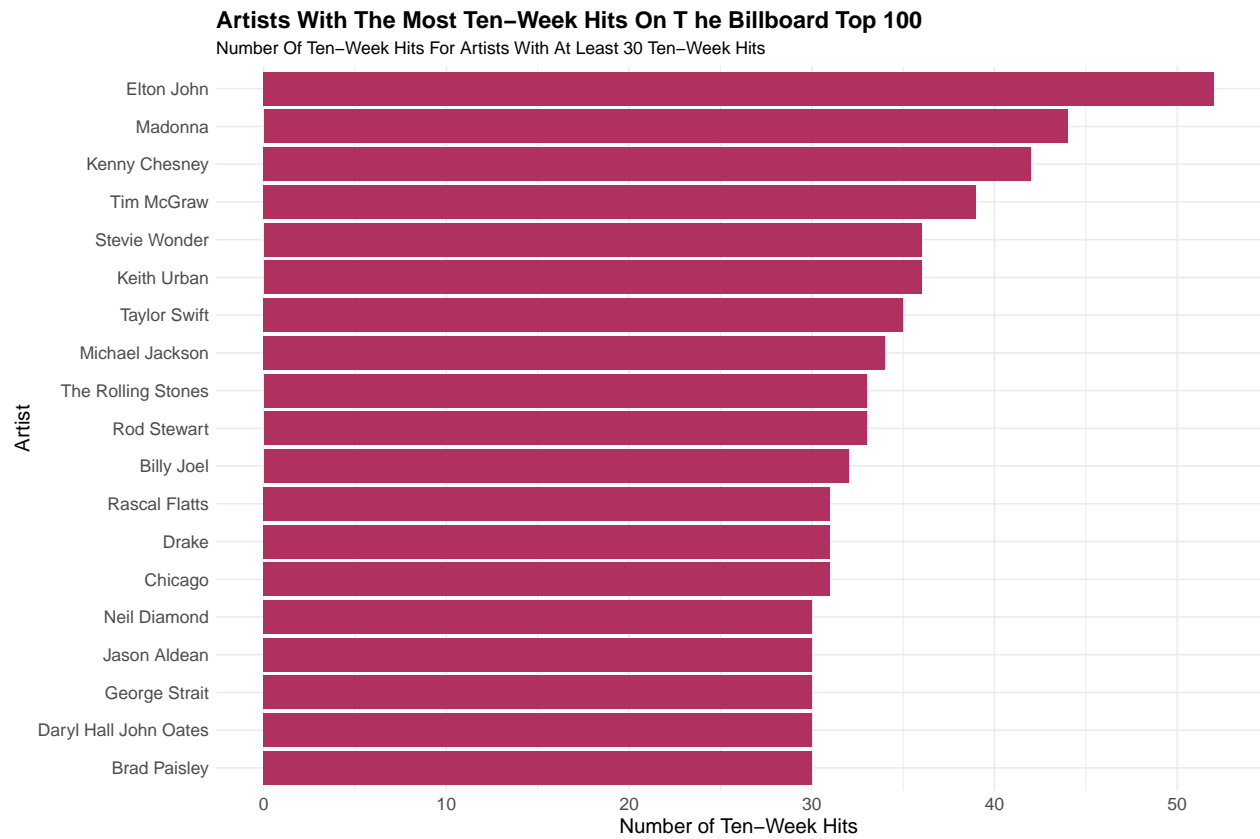
Part B

Musical Diversity of the Billboard Top 100 Over Time

Number Of Unique Songs Appearing on The List Each Year (1959–2020)



The line graph illustrates the number of unique songs that appeared on the Billboard Top 100 each year from 1959 to 2020. The x-axis represents the year, and the y-axis represents the count of unique songs. As is evident from the graph, the amount of unique songs decreased drastically from the late 1950s into the 21st century, hitting a minimum around 2001. This decline likely reflects a concentration of mainstream artists and genres during that period due to various factors. However, from the 2000s onwards, there is a clear resurgence in musical diversity. This reflects a time of dominance by a smaller group of artists (fewer one hit wonders). The strong positive slope in this portion of the graph reflects a growing variety of voices in music that have been able to crack into the Billboard Top 100.



The bar chart highlights artists with at least 30 songs that spent 10 or more weeks on the Billboard Top 100. The x-axis represents the artists, sorted by the number of ten-week hits (y-axis).