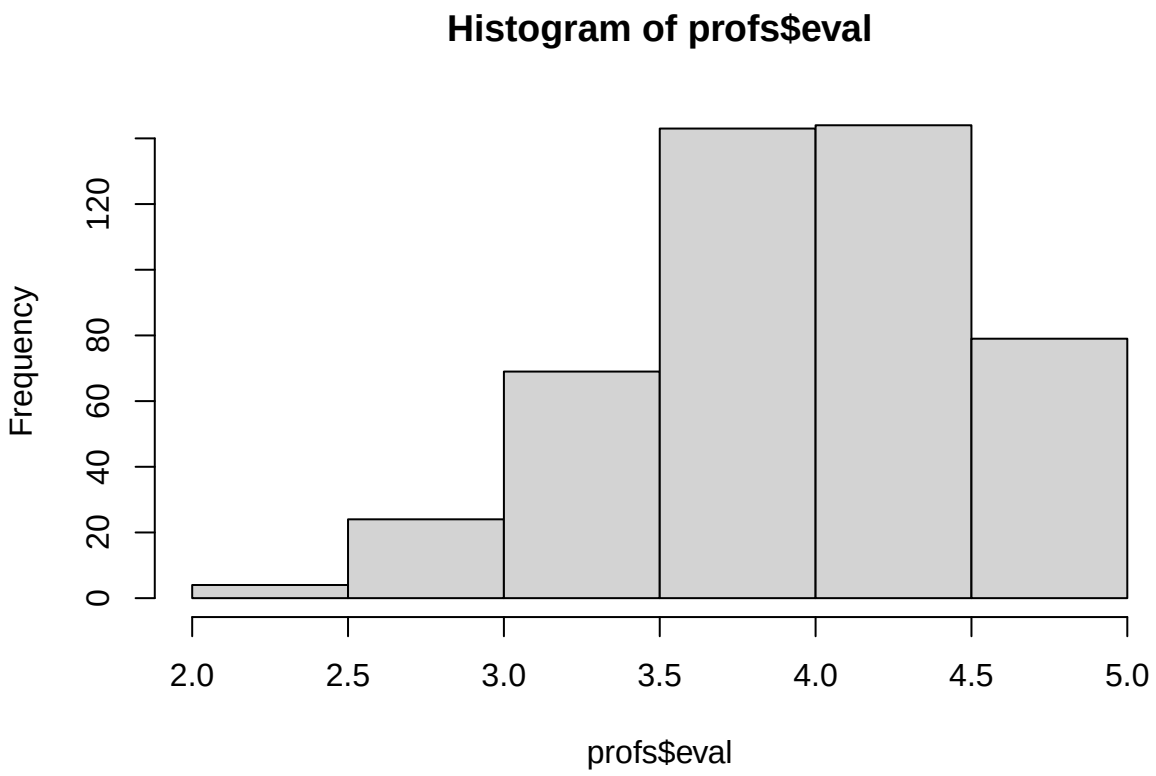
SDS315 HW2 - Palash Pawar ppp625

https://github.com/palashpawar/SDS315-HW2

2025-01-27

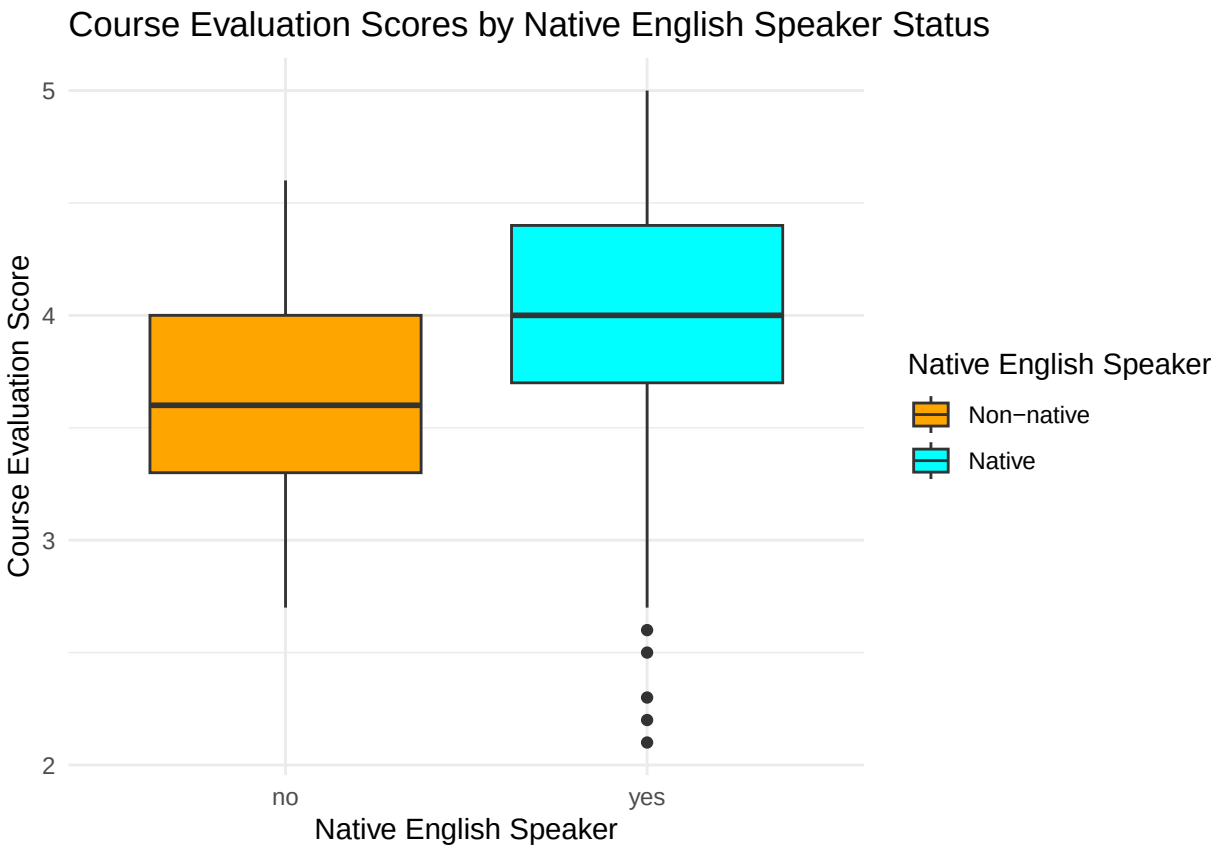
Problem 1: Beauty, or not, in the classroom

PART A



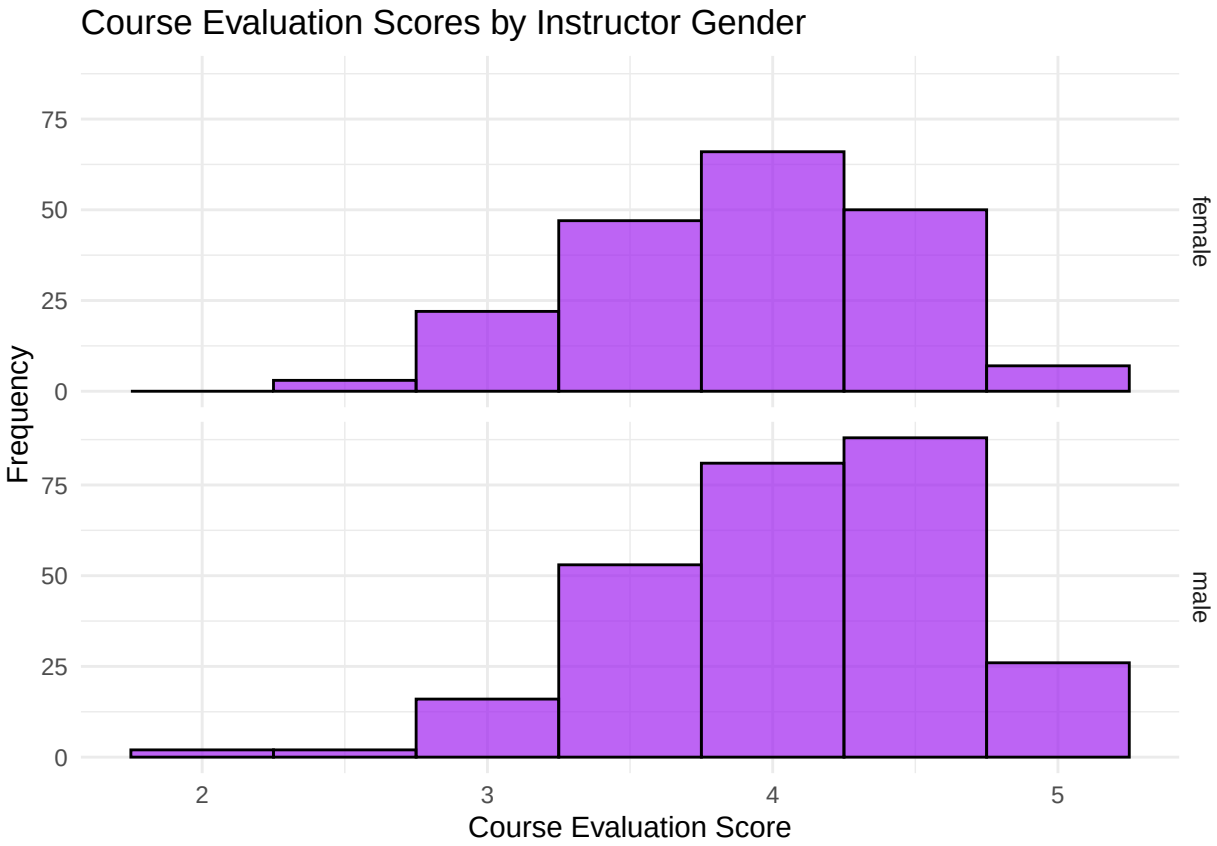
The distribution for the course evaluations is skewed to the left, with the mean score being between 3.5 and 4.5. The x-axis represents the course evaluations scores and the y-axis is the frequency that a professor has received a score in that range.

PART B



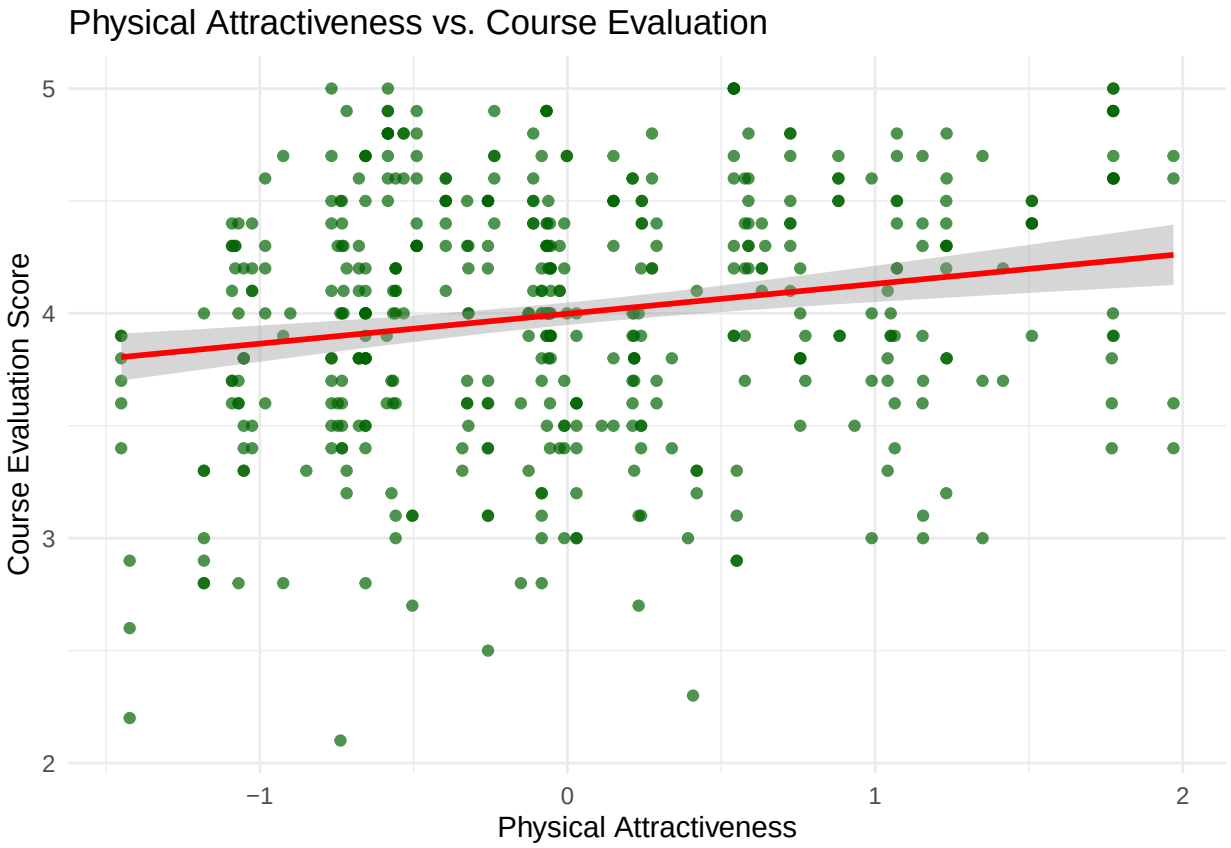
Based on the box plots, it is clear to see that the professors that are native English speakers received higher course evaluation scores than the professors that were non-native English speakers. The x-axis represents if the professor was a native English speaker or not and the y-axis is the course evaluations score received.

PART C



Based the histograms, we can see that male professors overall received higher course evaluation than female professors. The x-axis represents the course evaluations scores and the y-axis is the frequency that a professor has received a score in that range.

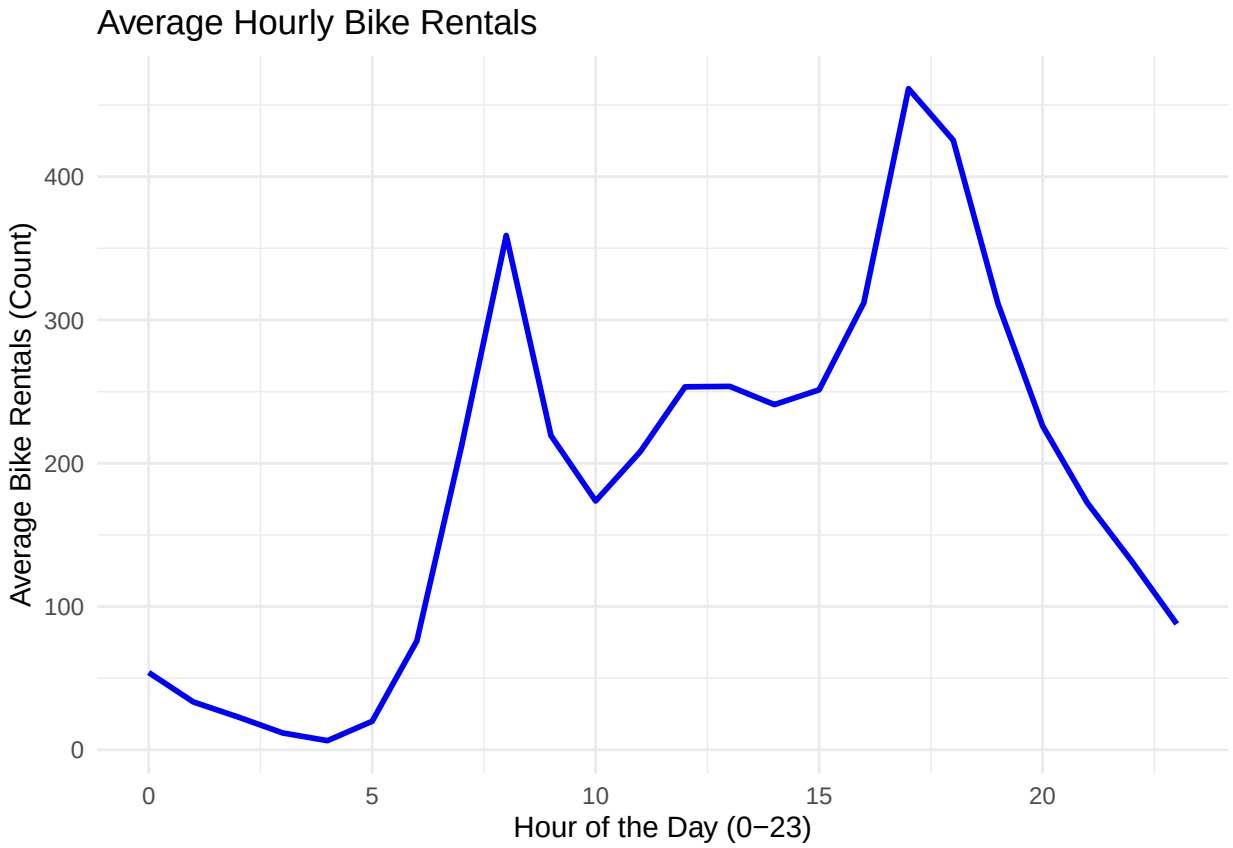
PART D



Based on the scatterplot above, there is a very weak positive correlation between the physical attractiveness of professors and the scores they receive for their course evaluations. The x-axis represents the scale of physical attractiveness and the y-axis represents the respective course evaluation score.

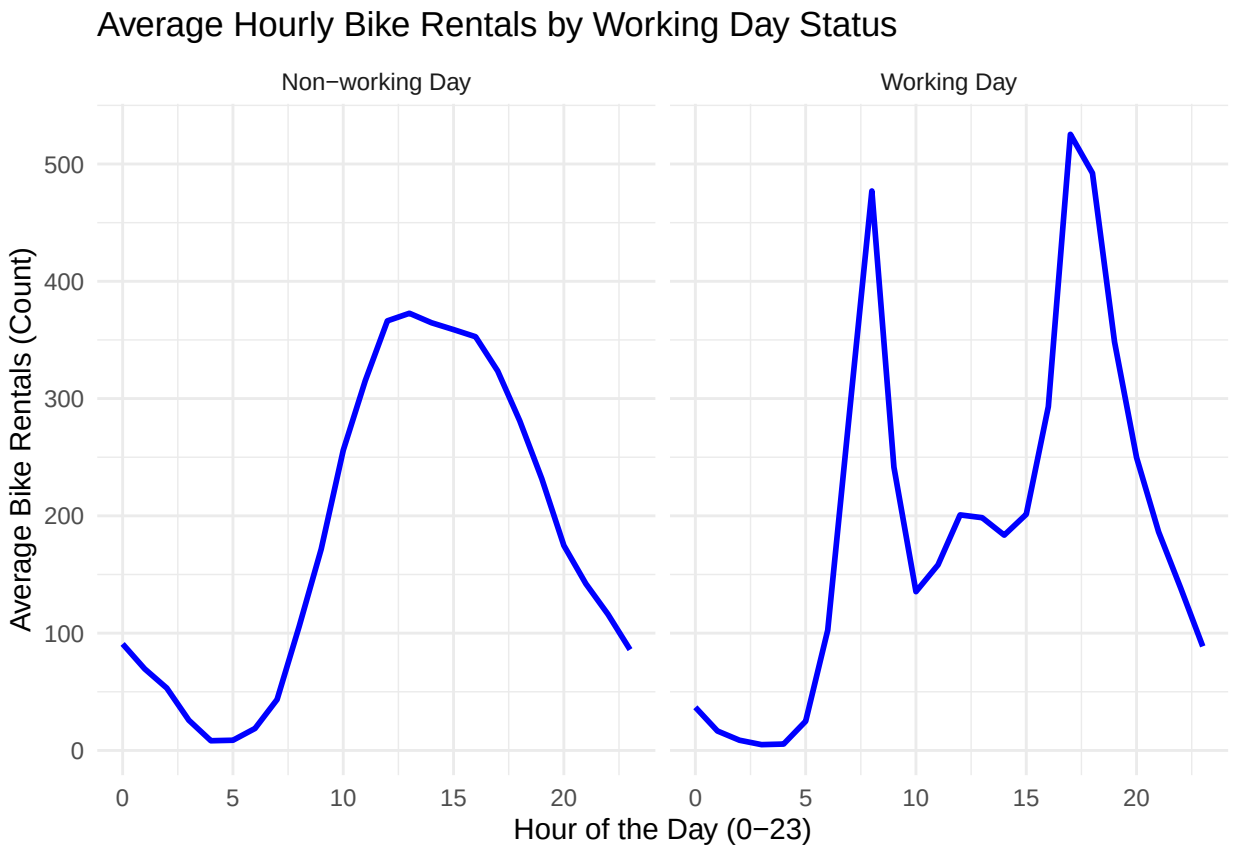
Problem 2: Bike Sharing

PLOT A



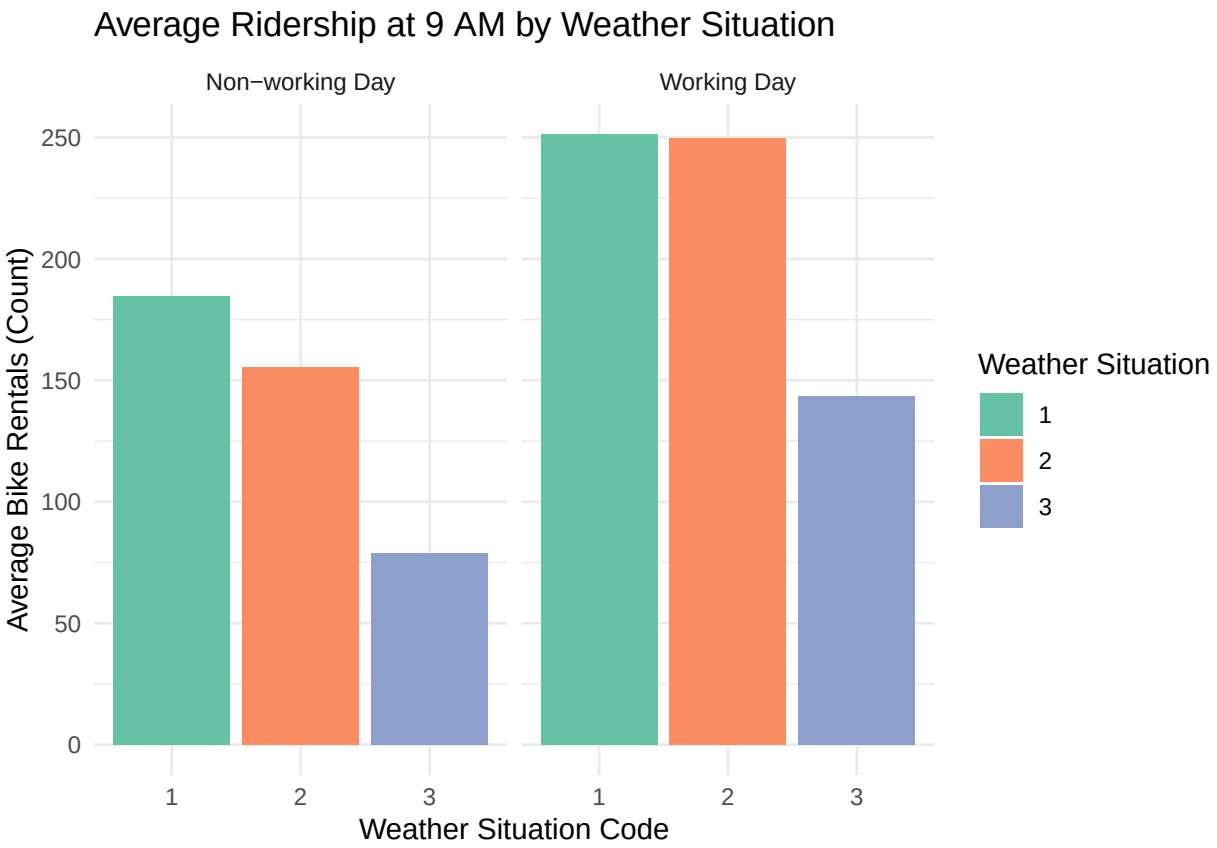
This line graph displays the average number of bike rentals (y-axis, in counts) for each hour of the day (x-axis, 0–23). The graph reveals clear peaks at 8 AM and 5 PM, corresponding to typical commuting hours, highlighting how bike rentals are driven by commuting patterns.

PLOT B



This faceted line graph compares the average number of bike rentals (y-axis, in counts) throughout the day (x-axis, 0–23), with separate panels for working days (top) and non-working days (bottom). On working days, rentals peak during commuting hours (8 AM and 5 PM), while non-working days show a more consistent midday peak, likely due to leisure activities.

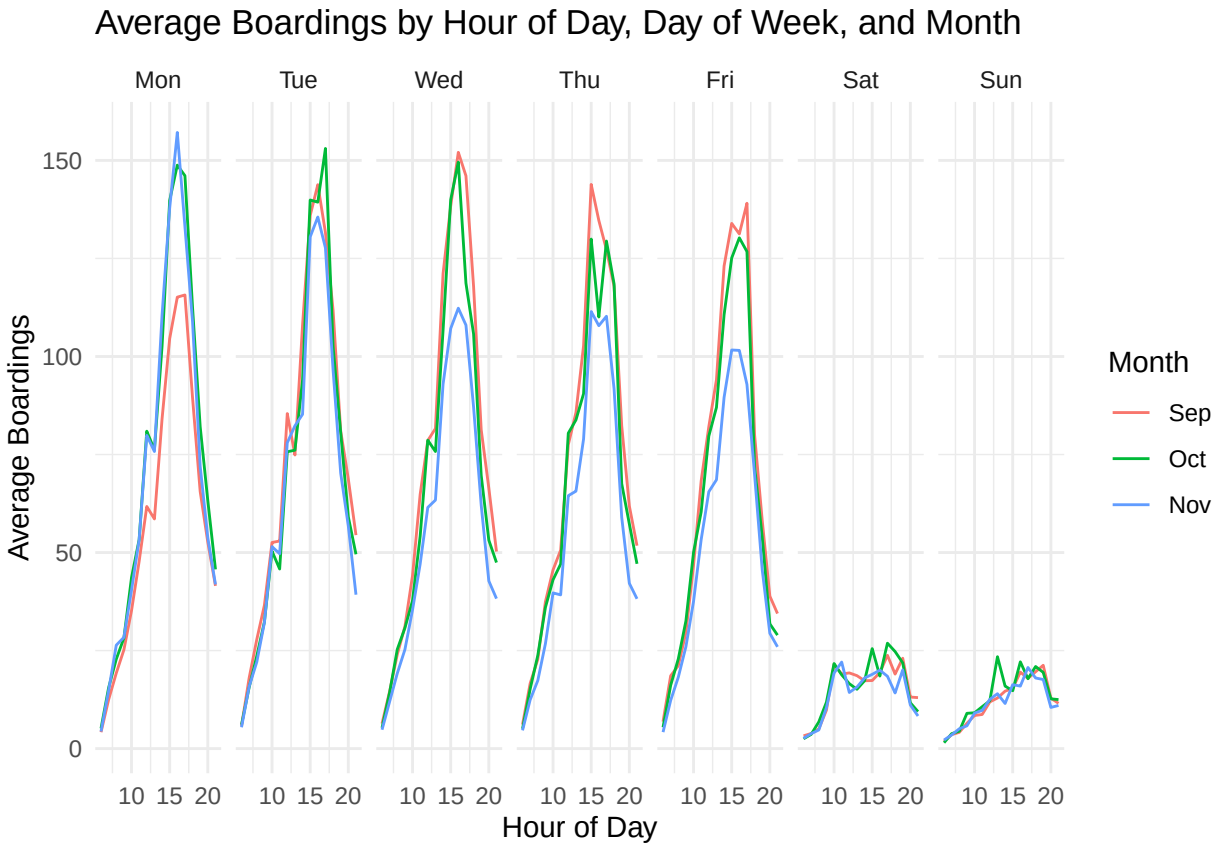
PLOT C



This faceted bar plot shows the average number of bike rentals (y-axis, in counts) at 9 AM for different weather situations (x-axis), with separate panels for working days and non-working days. Ridership decreases as weather conditions worsen (e.g., from clear to rainy), and this effect is more pronounced on working days. This suggests weather significantly impacts morning commutes.

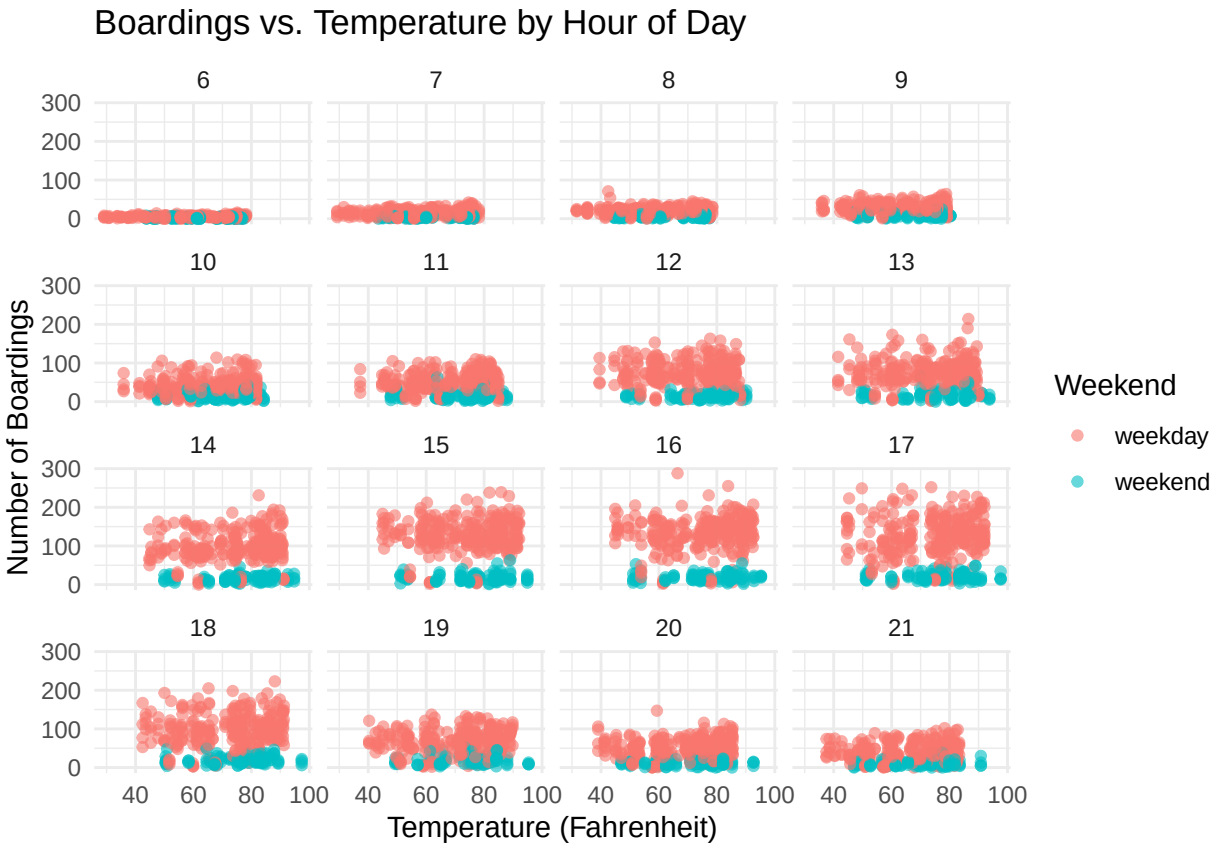
Problem 3 - Capital Metro UT Ridership

1.



This line graph shows the average boardings by hour of the day for each day of the week, with separate lines for each month (September, October, November). The hour of peak boardings is generally consistent across days, occurring in the morning (around 8 AM) and evening (around 5 PM). Lower average boardings on Mondays in September may reflect reduced activity at the start of the school year. Similarly, lower boardings on Wed/Thu/Fri in November might indicate holidays or reduced schedules near Thanksgiving.

2.



This scatter plot shows the relationship between boardings and temperature, faceted by hour of the day. Points are colored based on whether it is a weekend or a weekday. Holding the hour of day and weekend status constant, temperature appears to have a noticeable effect on boardings, with higher temperatures generally correlating with higher ridership, particularly during midday hours

Problem 4: Wrangling the Billboard Top 100

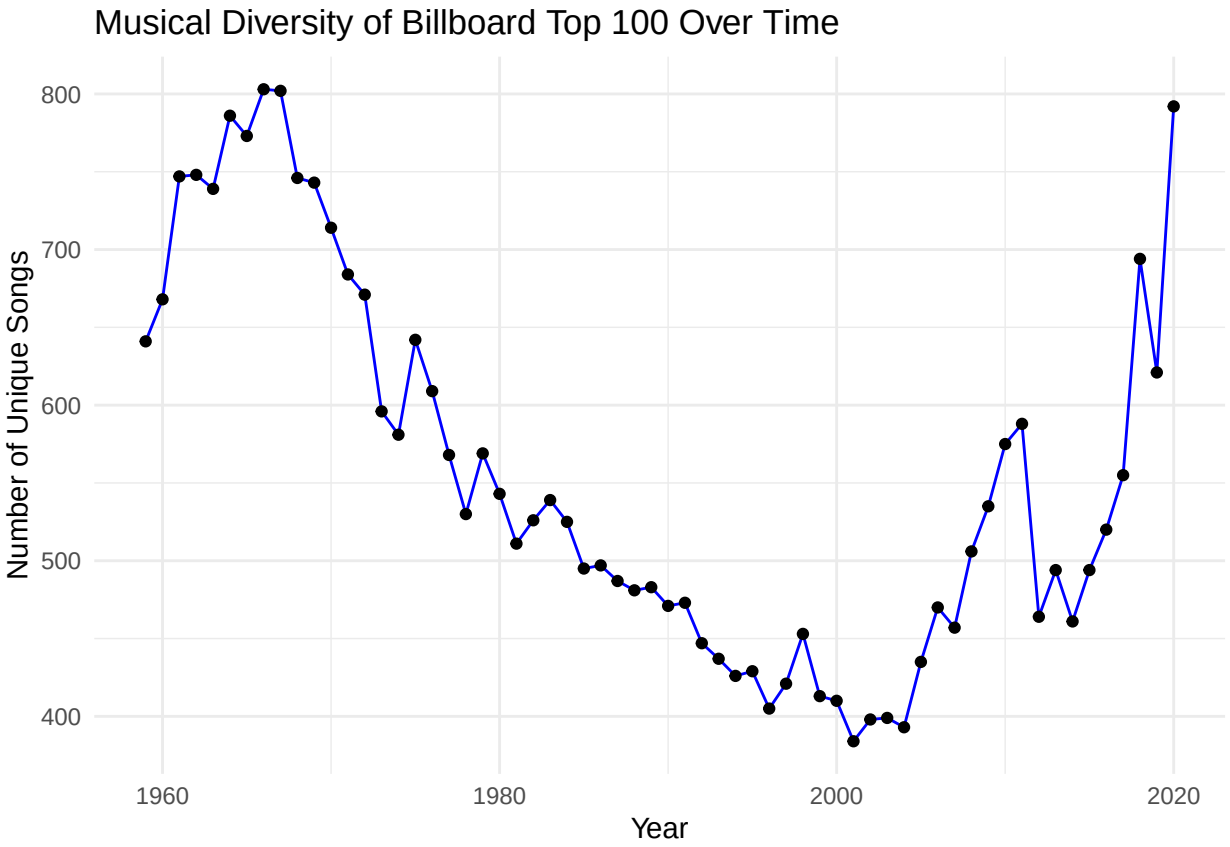
PART A

Table 1: Top 10 Most Popular Songs on Billboard

performer	song	total_weeks
Imagine Dragons	Radioactive	3828
AWOLNATION	Sail	3160
Jason Mraz	I'm Yours	2926
The Weeknd	Blinding Lights	2926
LeAnn Rimes	How Do I Live	2415
LMFAO Featuring Lauren Bennett & GoonRock	Party Rock Anthem	2346
OneRepublic	Counting Stars	2346
Adele	Rolling In The Deep	2145
Jewel	Foolish Games/You Were Meant For Me	2145
Carrie Underwood	Before He Cheats	2080

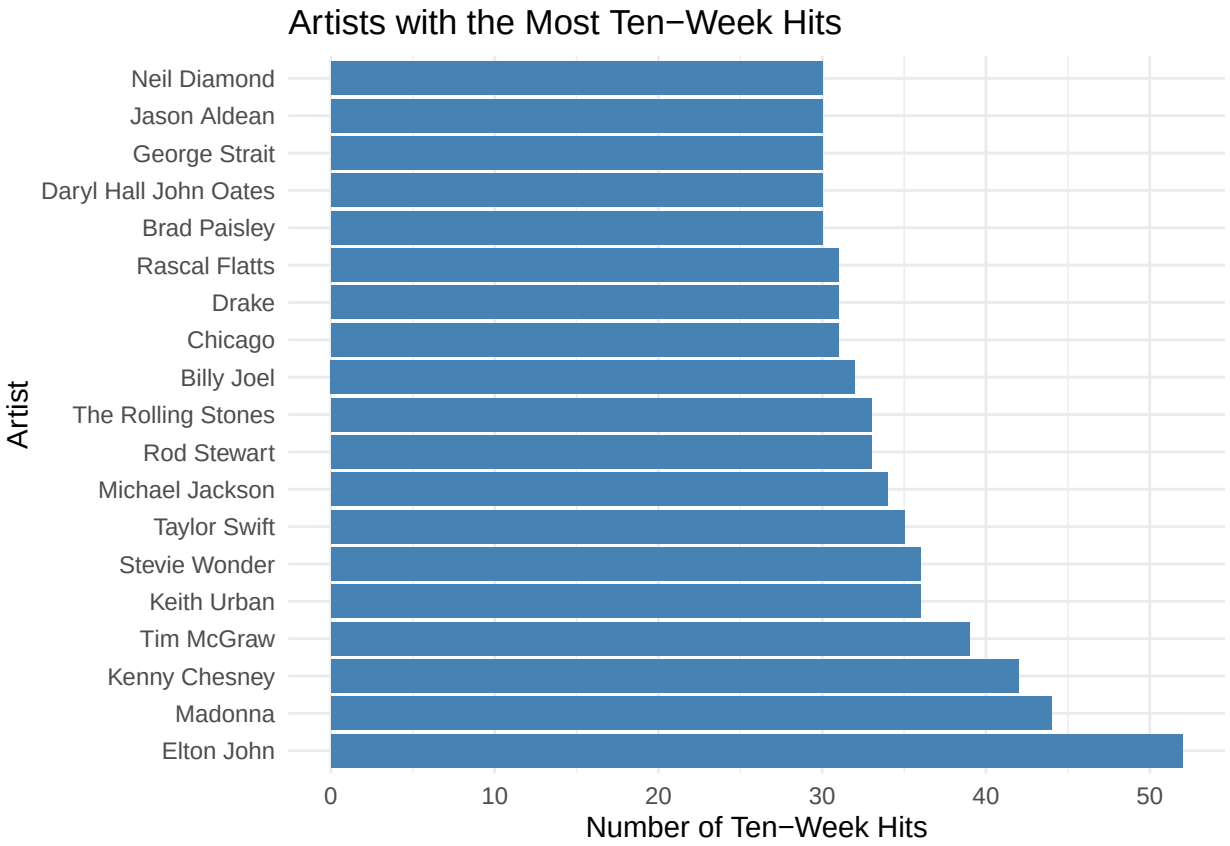
This table shows the top 10 most popular songs since 1958, ranked by the total number of weeks spent on the Billboard Top 100 chart. Columns include the performer, song title, and total count of weeks on the chart.

PART B



This line graph shows the musical diversity of the Billboard Top 100 from 1959 to 2020. Diversity is measured by the number of unique songs appearing in any position on the chart each year. The x-axis represents the year and the y-axis represents the number of unique songs.

PART C



This bar graph shows the 19 artists who have had at least 30 songs that appeared on the Billboard Top 100 for at least 10 weeks. The x-axis represents the number of ten-weeks hits and the y-axis represents the artist.