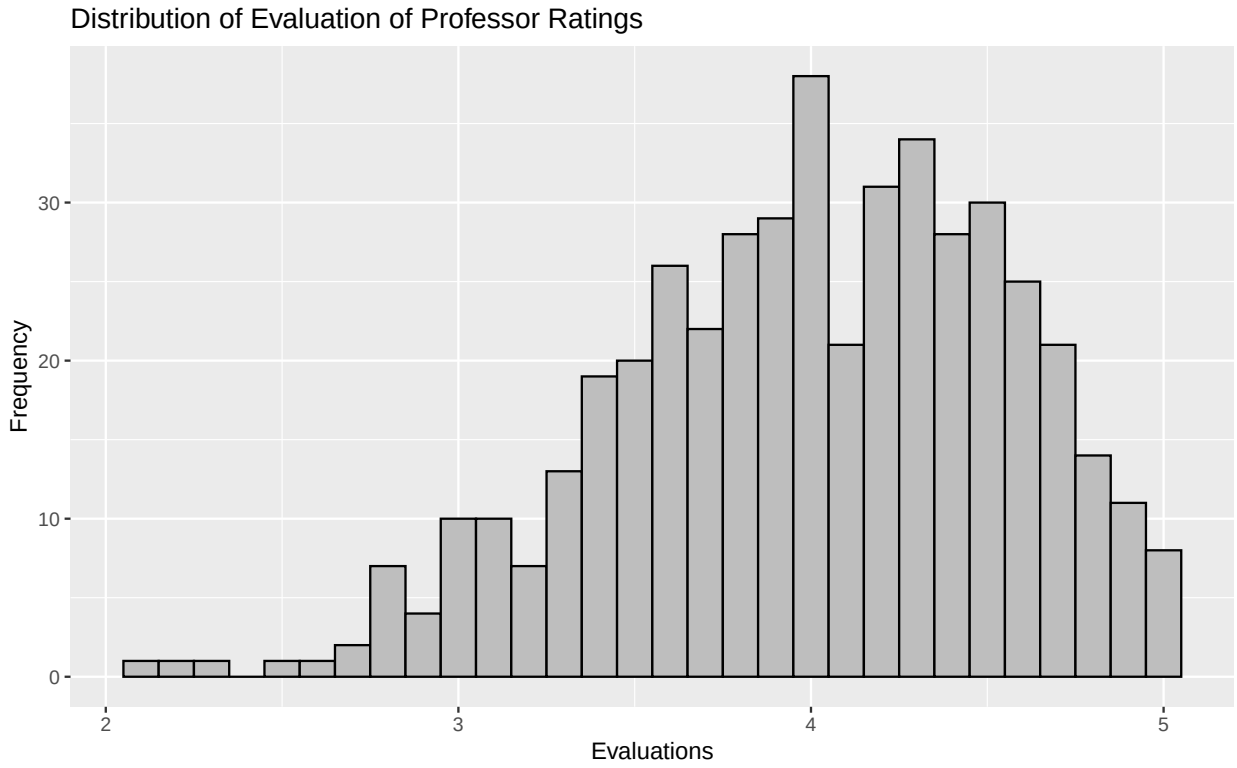
Homework 2

Pranav Belligundu(psb898) - SDS 315 UT Austin

https://github.com/pranav-B21/SDS-315/tree/main

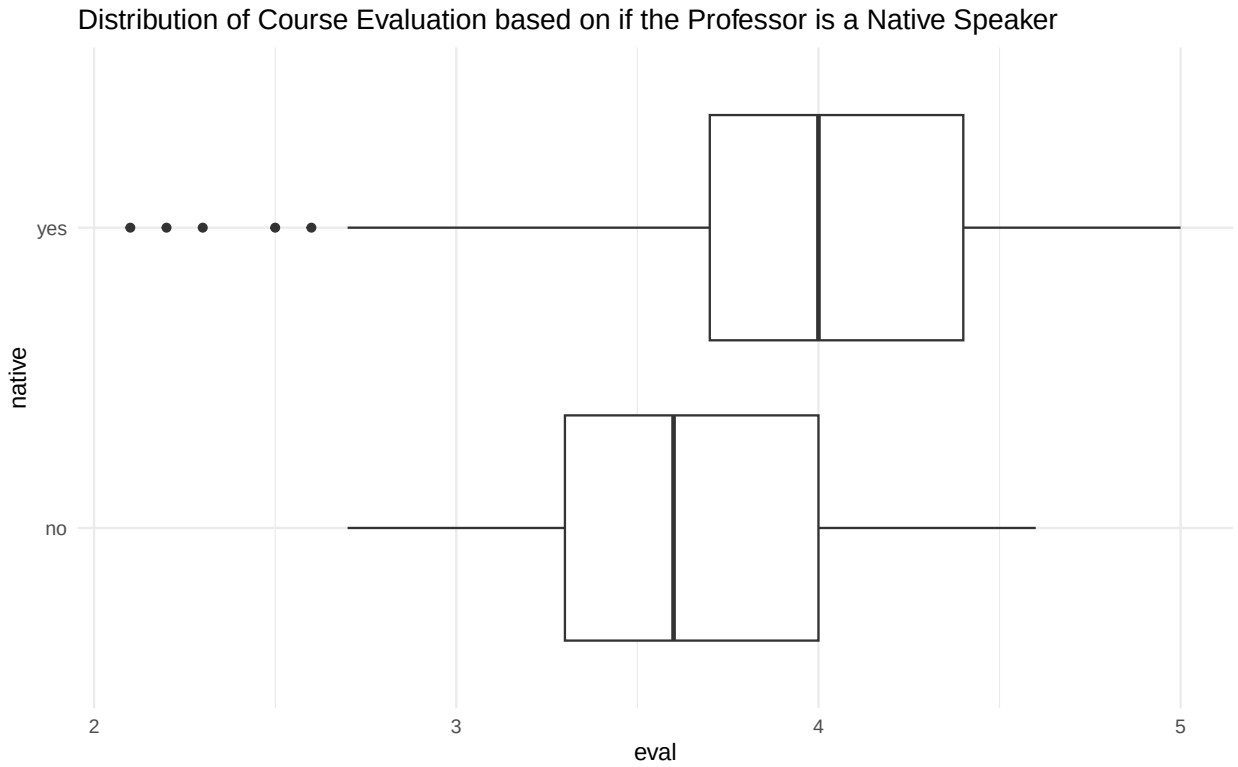
Problem 1: Beauty, or not, in the classroom

Part A



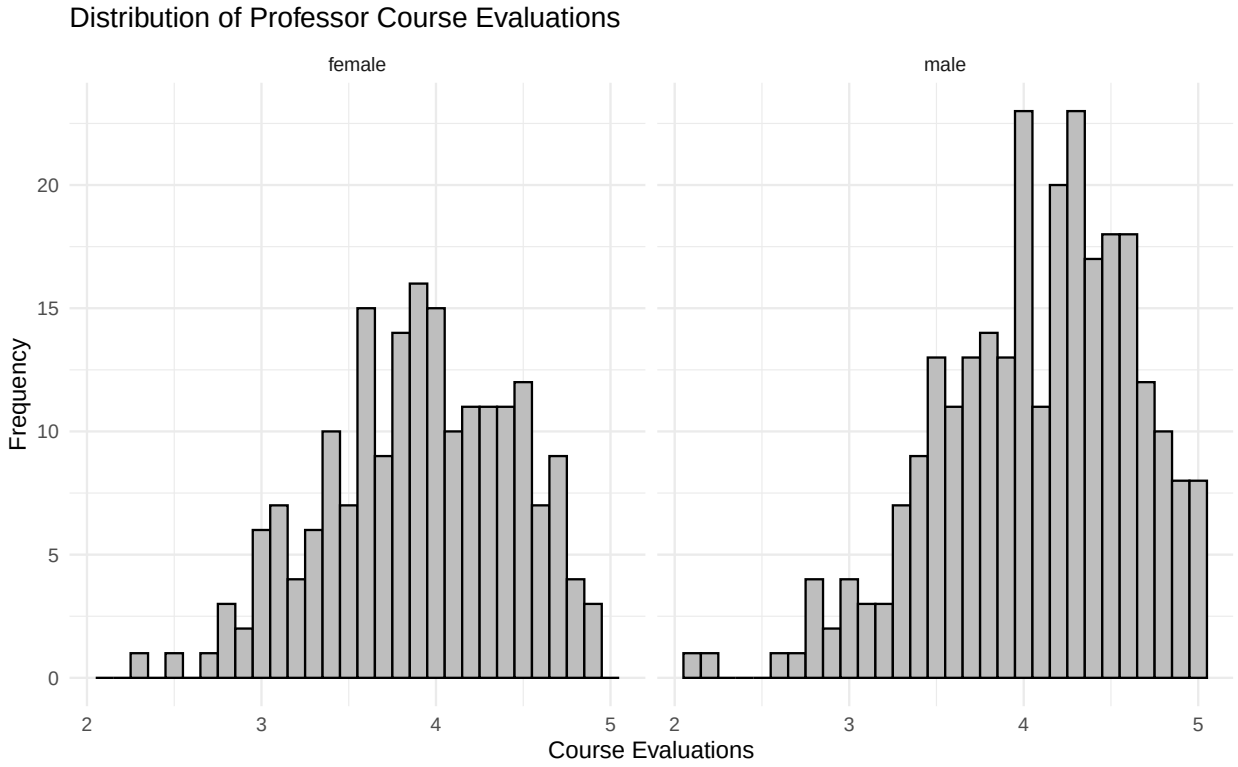
The x-axis represents the course evaluations scores and the y-axis is the frequency that a professor has gotten. The distribution for the course evaluations is skewed a little bit to the left, with the mean score being around 3.9 and the median being 4.

Part B



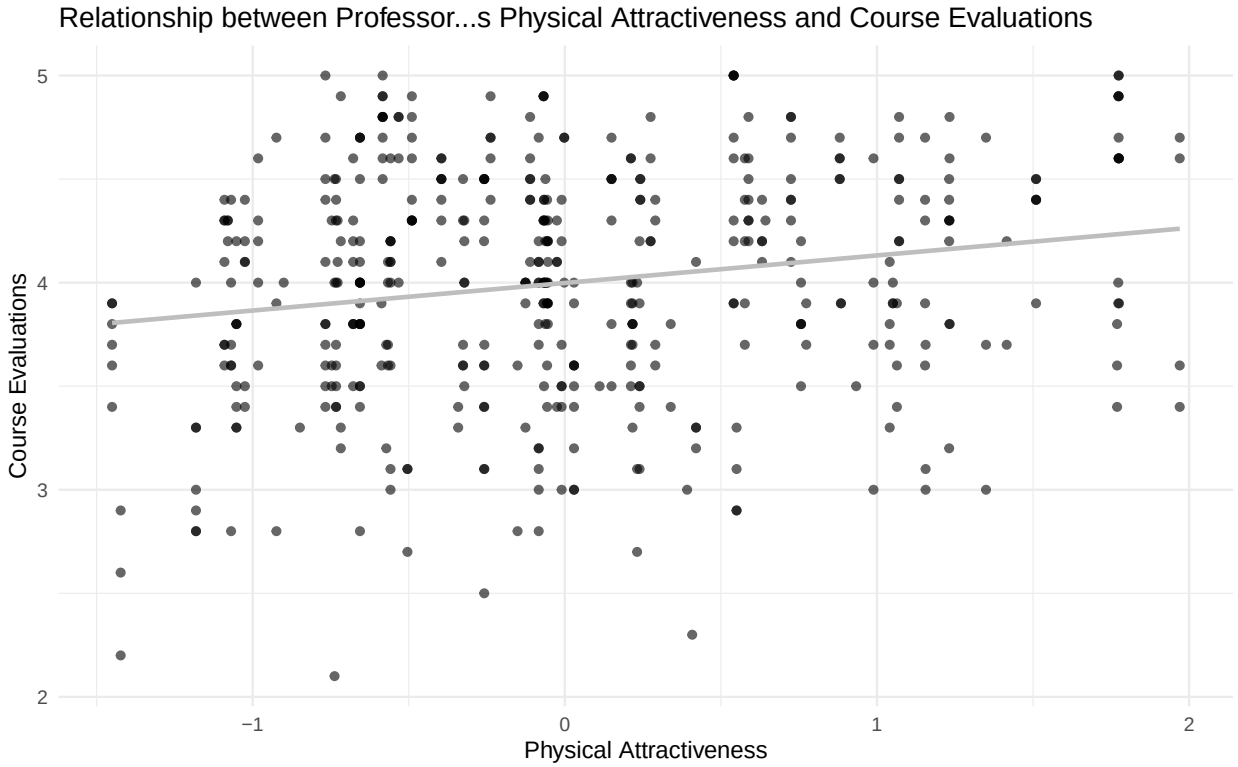
The x-axis represents if the professor was a native English speaker or not and the y-axis is the course evaluations score received. Based on the box plots, we can see that professors who are native English speakers received higher course evaluation scores than the professors that were non-native English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

Part C



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 with the graphs faceted by male and female instructors. Based the histograms, we can see that male professors received higher course evaluation than female professors.

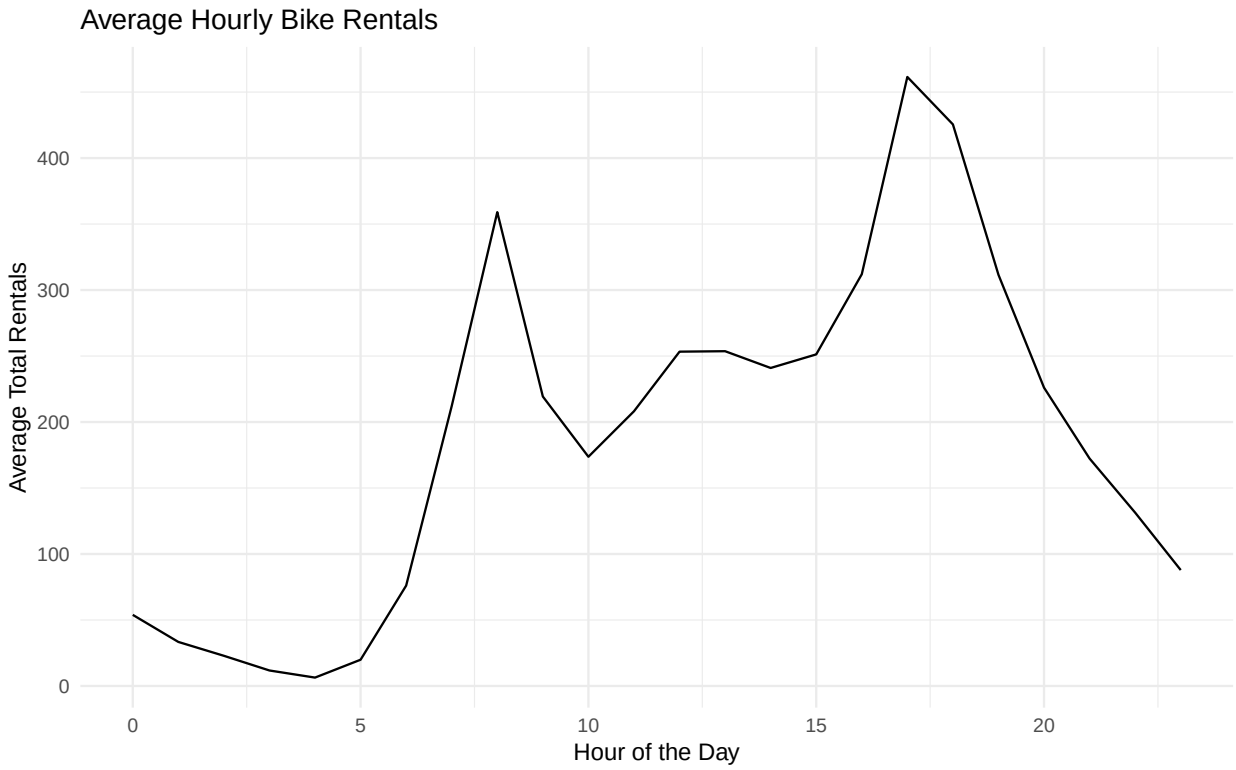
Part D



The correlation between the professor's physical attractiveness and course evaluations is 0.1890391 which is a small positive correlation. This shows that there is no correlation and there is no basis for an invdividual to believe that physical attractiveness does actually correlate with course evaluations.

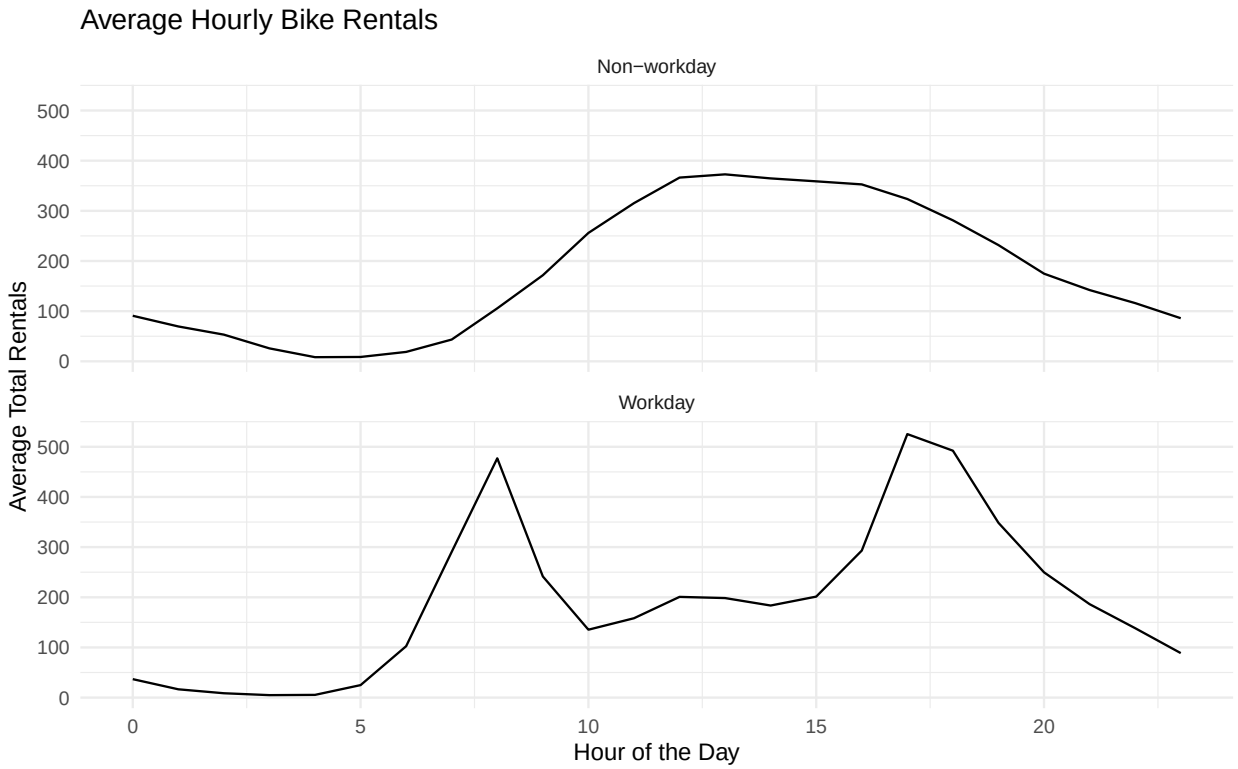
Problem 2: Bike Sharing

Plot A



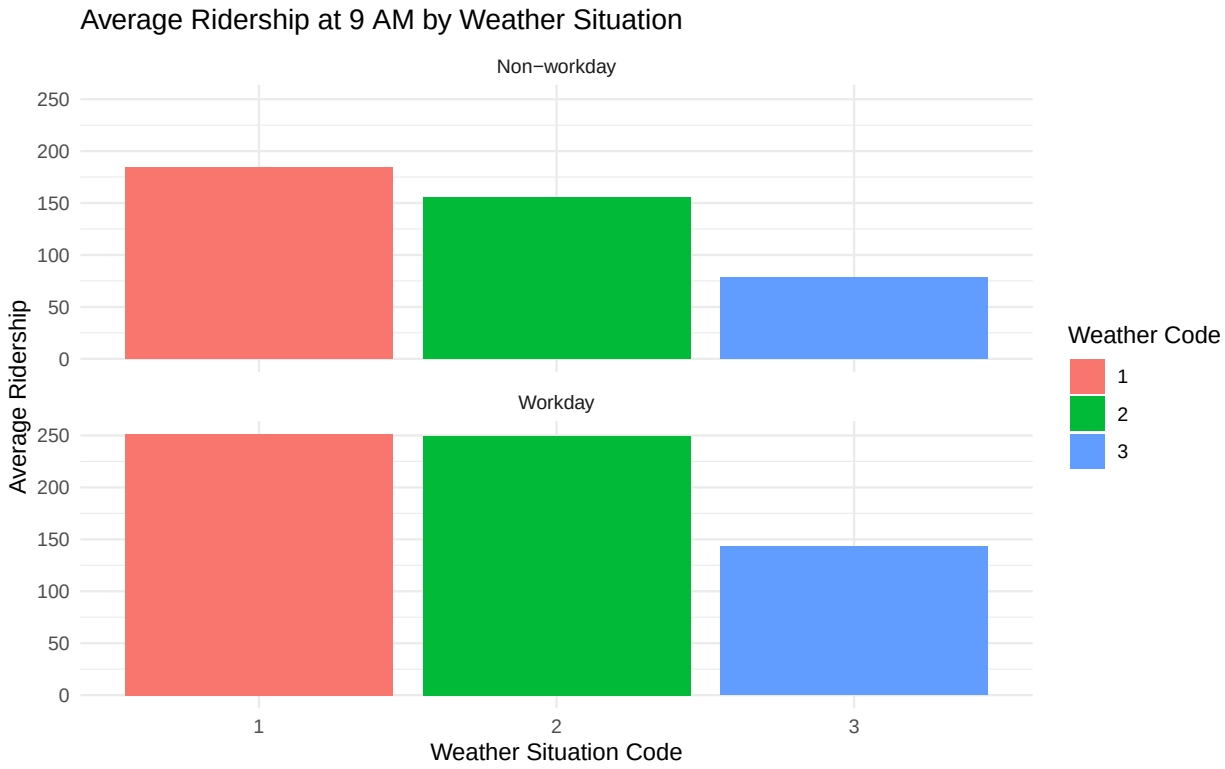
The y-axis represents the counts for the average number of bike rentals and the x-axis represents each hour of the day. The graph shows clear peaks at around 8 AM and 5 PM which corresponds to normal commuting hours for those who go to work. This shows how bike rentals are driven by commuting patterns.

Plot B



The y-axis represents the counts for the average number of bike rentals and the x-axis represents each hour of the day which is faceted by working days (top) and non-working days (bottom). On working days, the rentals peak during commuting hours (8 AM and 5 PM). During non-working days, there is a small peak and decline around midday peak due to users using the rentals for fun.

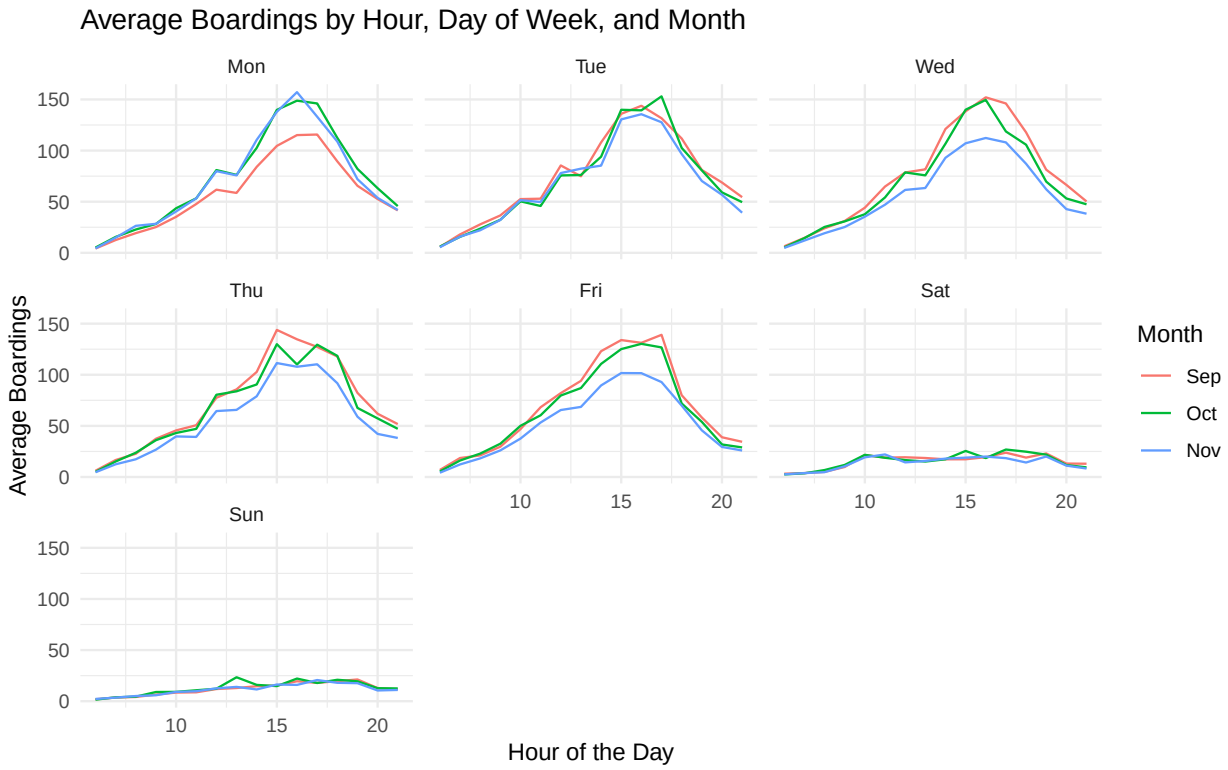
Plot C



This is a bar plot that shows the average number of bike rentals in the y-axis at 9 AM for different weather situations shown by the different colors. which is faceted by working days and non-working days. We can see that ridership decreases as weather conditions worsen and this effect is more evident during working days. This shows that weather significantly impacts morning commutes.

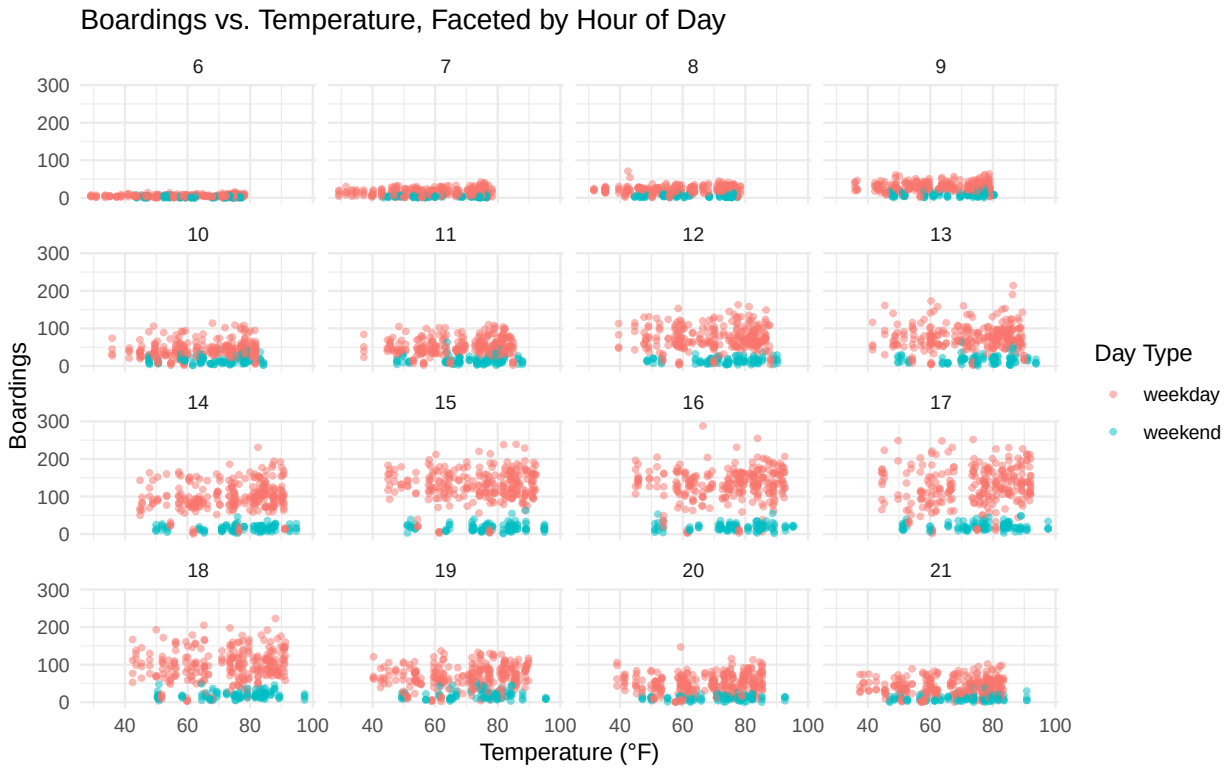
Problem 3: Capital Metro UT Ridership

Plot 1



This line graph shows the average boardings by hour of the day faceted by each day of the week with separate lines for each month. The hour of peak boardings is generally consistent occurring in the morning around 8 AM and evening around 5 PM. Lower average boardings during Mondays in September and Wed/Thu/Fri in November might indiciate holidays and days off.

Plot 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 where the red represents a weekday and blue represents a weekend. We can see that temperature appears to have a noticeable effect on boardings where a higher temperature correlates with higher ridership, especially during midday hours.

Problem 4: Wrangling the Billboard Top 10

Part A

Table 1: Top 10 Songs by Total Weeks on Billboard Top 100

Song	Performer	Total Weeks
Radioactive	Imagine Dragons	87
Sail	AWOLNATION	79
Blinding Lights	The Weeknd	76
I'm Yours	Jason Mraz	76
How Do I Live	LeAnn Rimes	69
Counting Stars	OneRepublic	68
Party Rock Anthem	LMFAO Featuring Lauren Bennett & GoonRock	68
Foolish Games/You Were Meant For Me	Jewel	65
Rolling In The Deep	Adele	65
Before He Cheats	Carrie Underwood	64

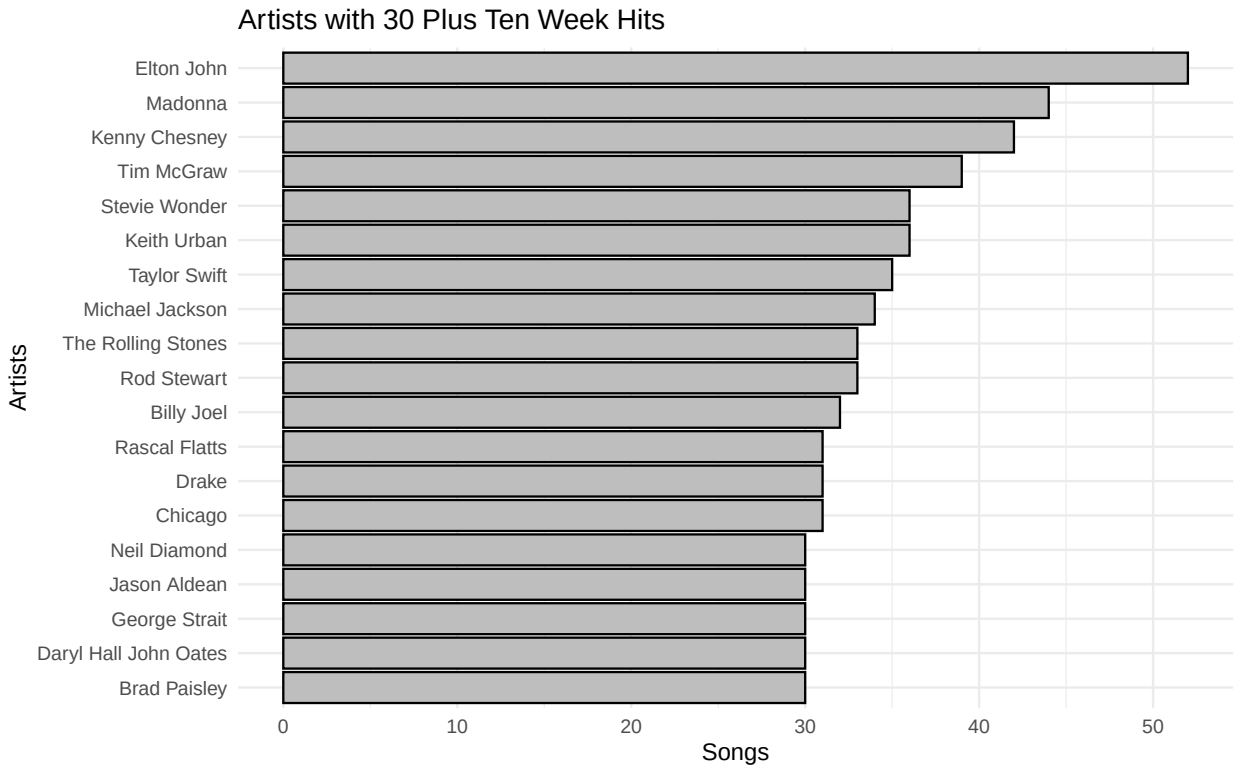
This table above shows the top 10 most popular songs since 1958 which is ranked by the total number of

weeks spent on the Billboard Top 100 chart. Columns include the song title, performer, and total count of weeks on the chart

Part B

The x-axis represents the year and the y-axis represents the number of unique songs. This line graph shows the musical diversity of the Billboard Top 100 from 1959 to 2020 wehre diversity is measured by how many new, unique songs enter any position on the chart each year.

Part C



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