

SDS 315 Homework 2

Avik Malladi

2024-01-28

arm7542

Github Repository

Problem 1: Beauty, or not, in the classroom

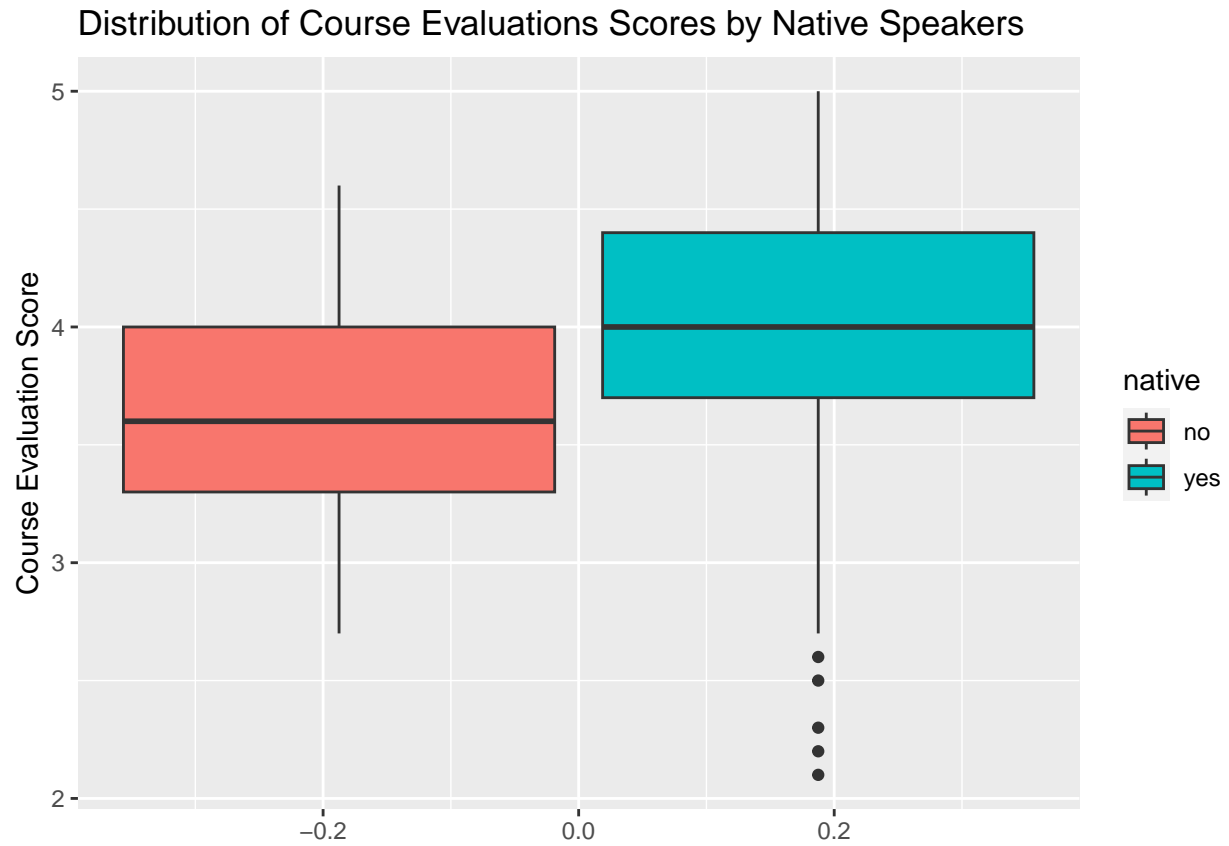
Part A

Distribution of Course Evaluation Scores



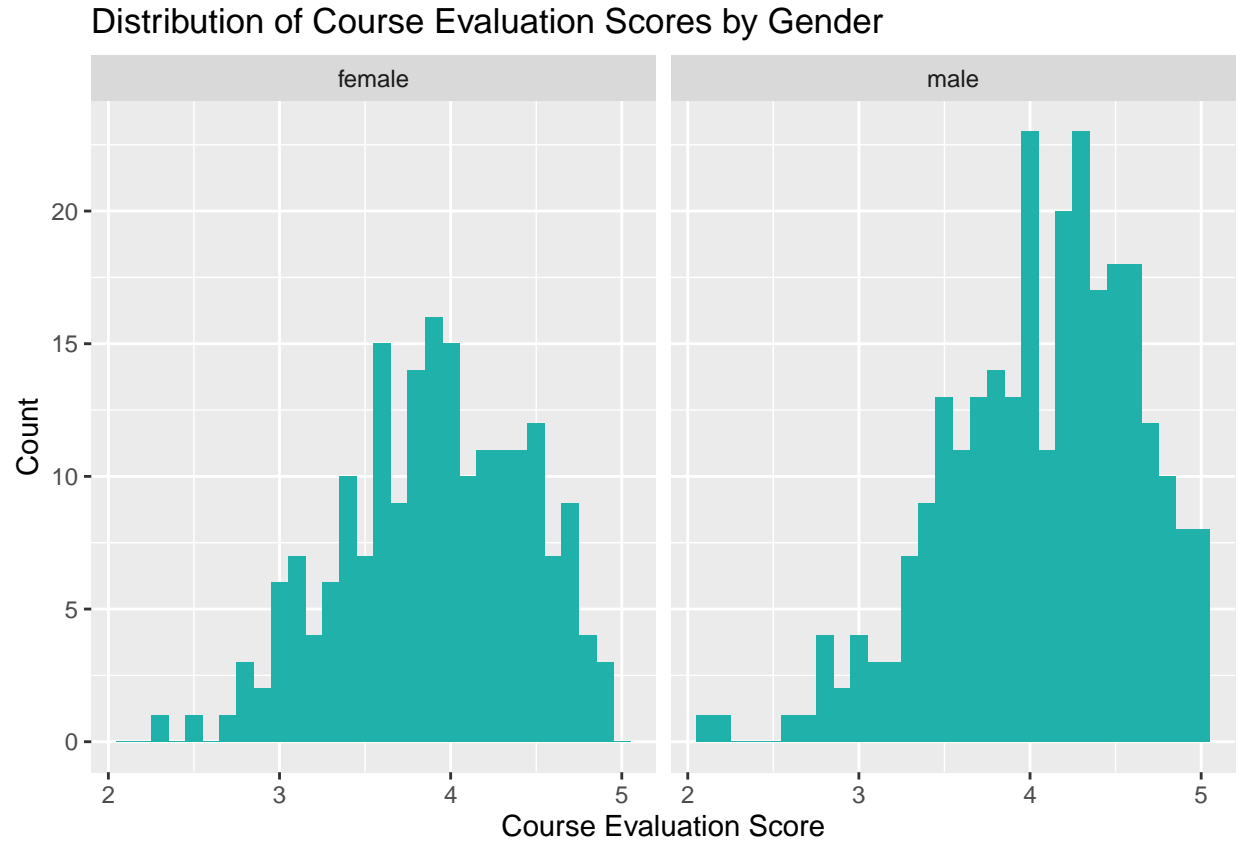
This histogram displays the overall distribution of course evaluation scores. The mean of the scores is 3.99.

Part B



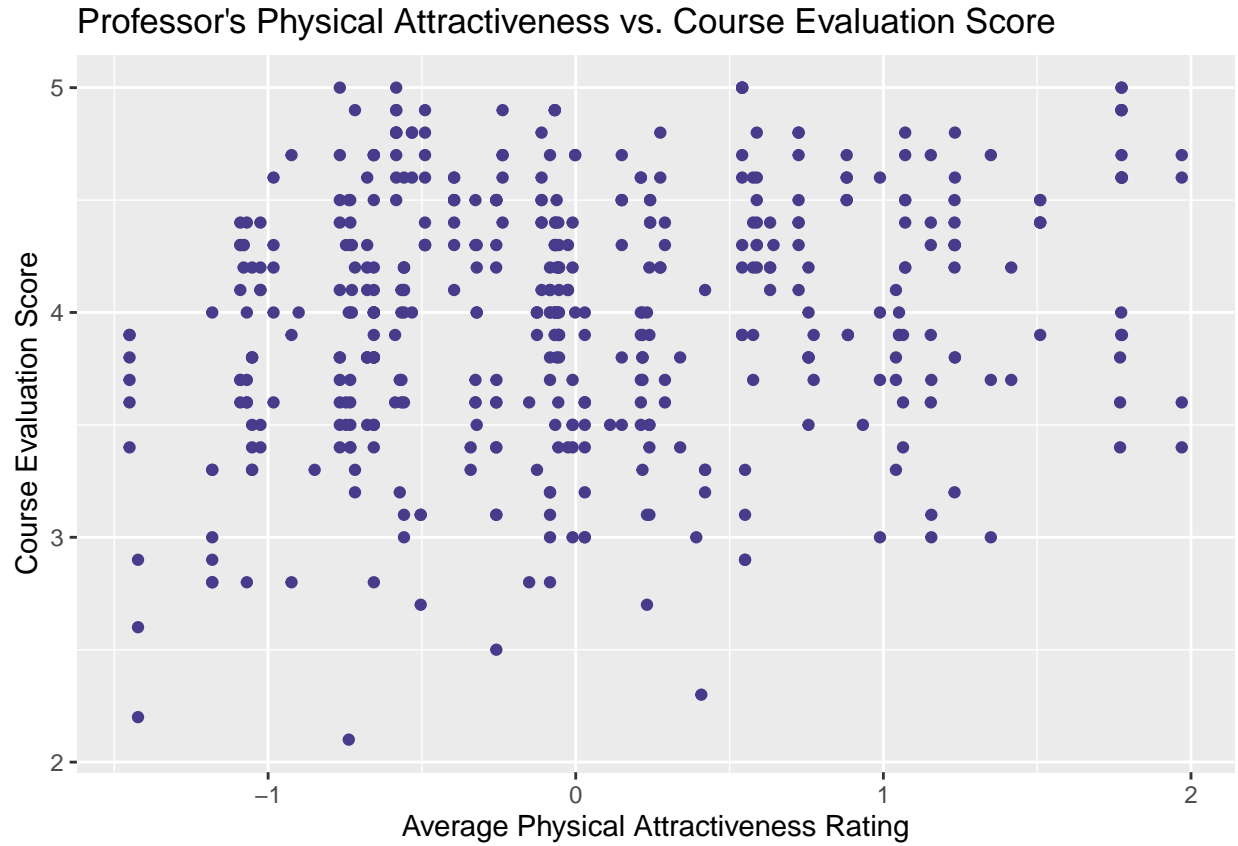
These two boxplots display the distribution of the course evaluation scores based on whether the professor is a native english speaker or not. Professors that are native english speakers earn a score of 0.81 higher on average.

Part C



These two graphs show the distribution of the course evaluation scores split based on gender. There is more data collected for male professors and also a slightly higher mean evaluation score.

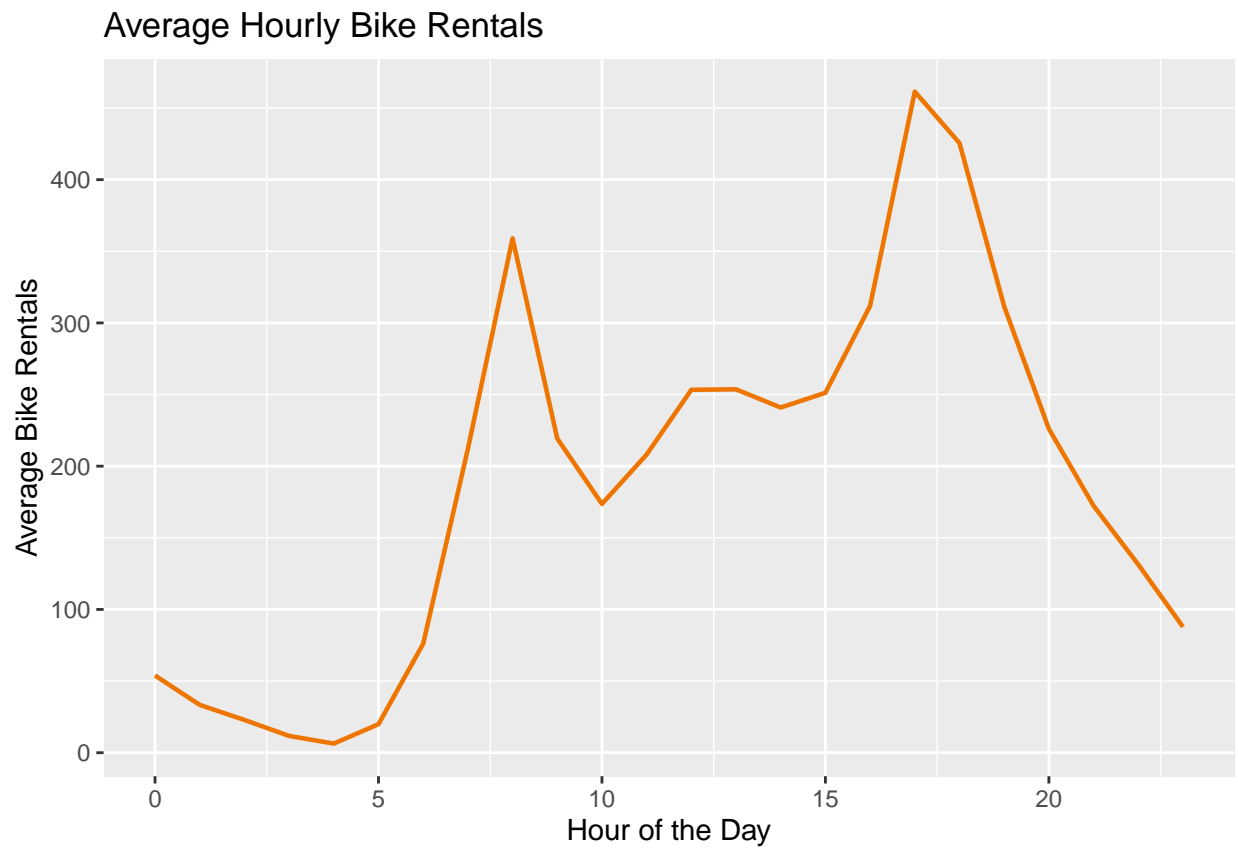
Part D



This dot plot shows the relationship between the professors average physical attractiveness rating and their average course evaluation score. The correlation coefficient between these two metrics is 0.19.

Problem 2: Bike Sharing

Part A



This line graph displays the average bike rentals throughout each hour of the day. From the graph we can see that the most active times of the day for bike rentals are hour 17 (5:00 PM) and hour 8 (8:00 AM).

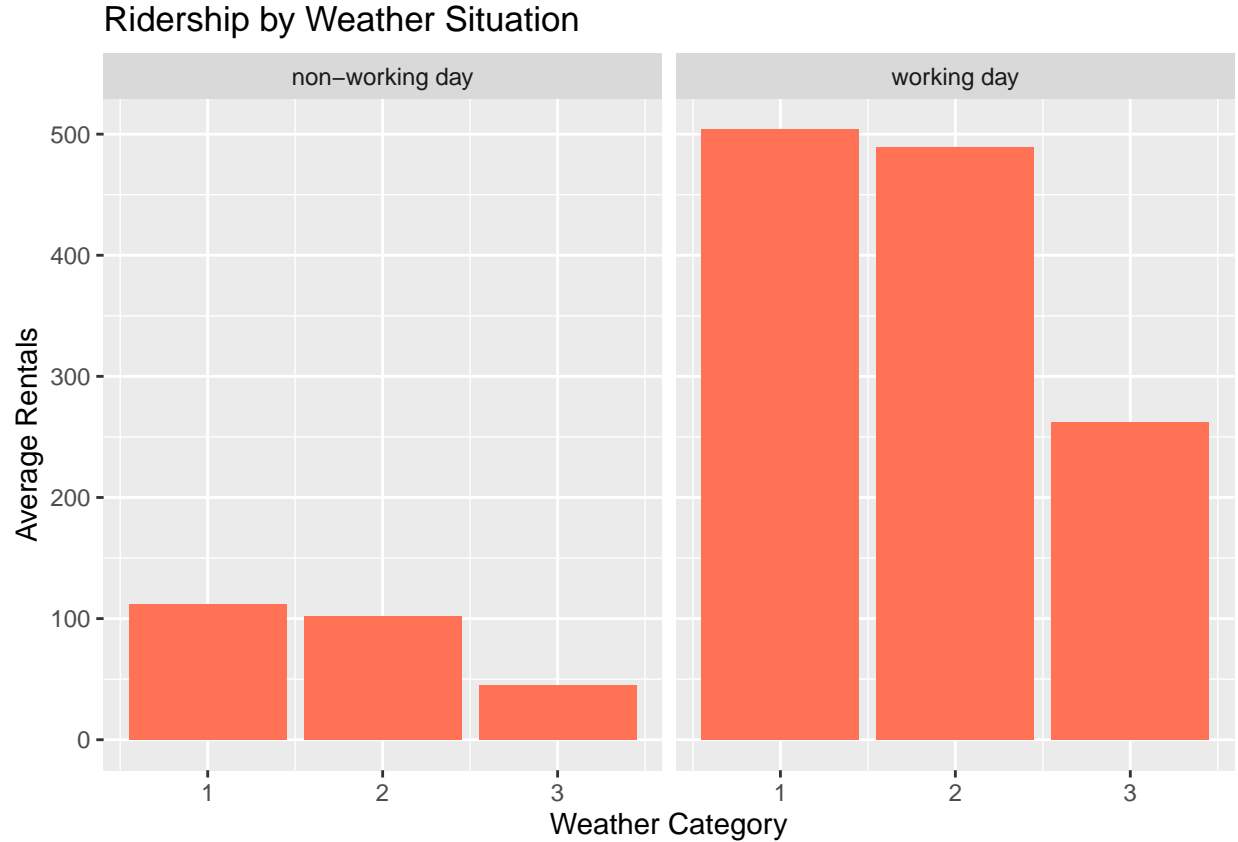
Part B

Average Hourly Bike Rentals by Working Day



The faceted graph above displays the average bike rentals per hour on working days vs. non-working days (weekends and holidays). '0' represents a non-working day, while '1' represents a working day. We notice that working days have two significant peaks at 8:00 AM and 5:00 PM, while the non-working days have a fairly even distribution with a peak at around noon.

Part C



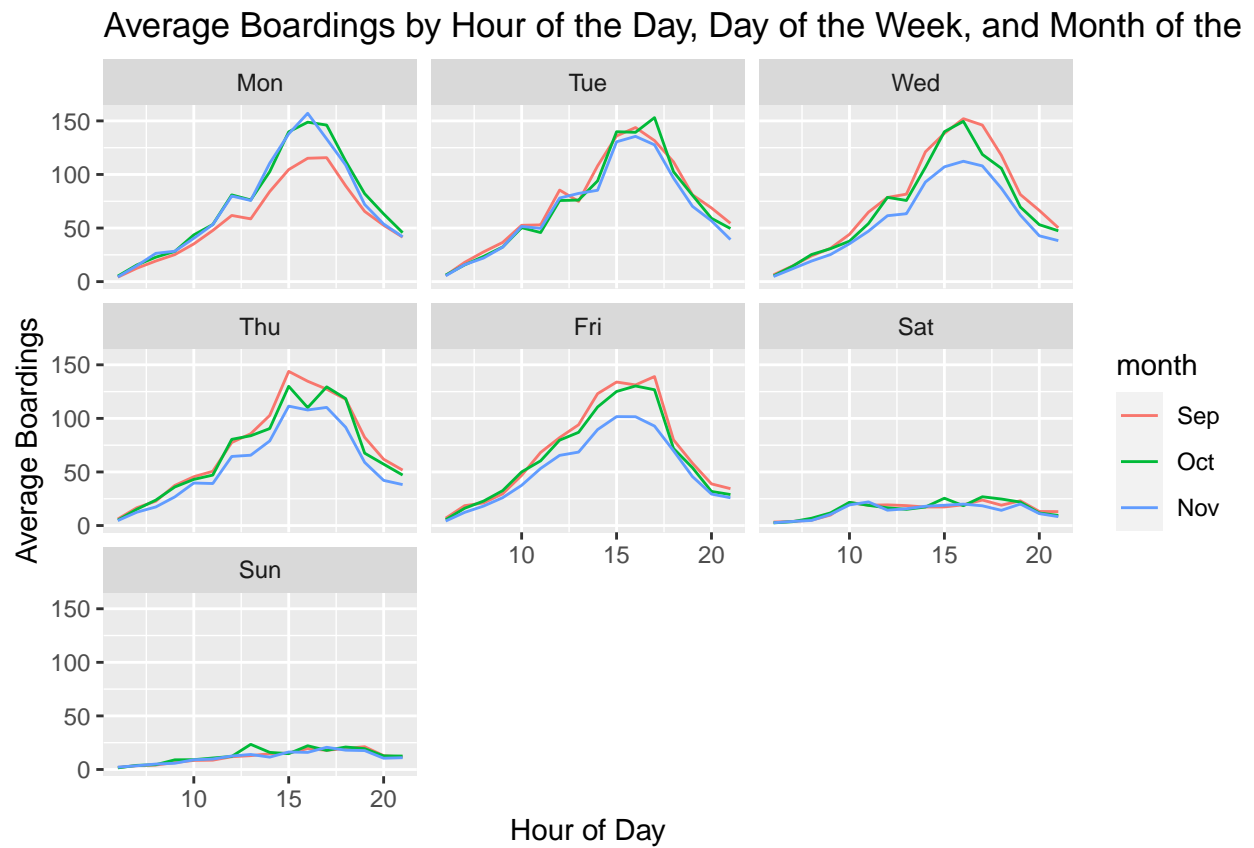
The graph above displays the average rentals based on different weather conditions for working days and non-working days. The weather conditions are defined as:

1: Clear, Few clouds, Partly cloudy, Partly cloudy 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds

We can see that in both working and non-working days, the weather condition 1 has the most ridership.

Problem 3: Captial UT Metro Ridership

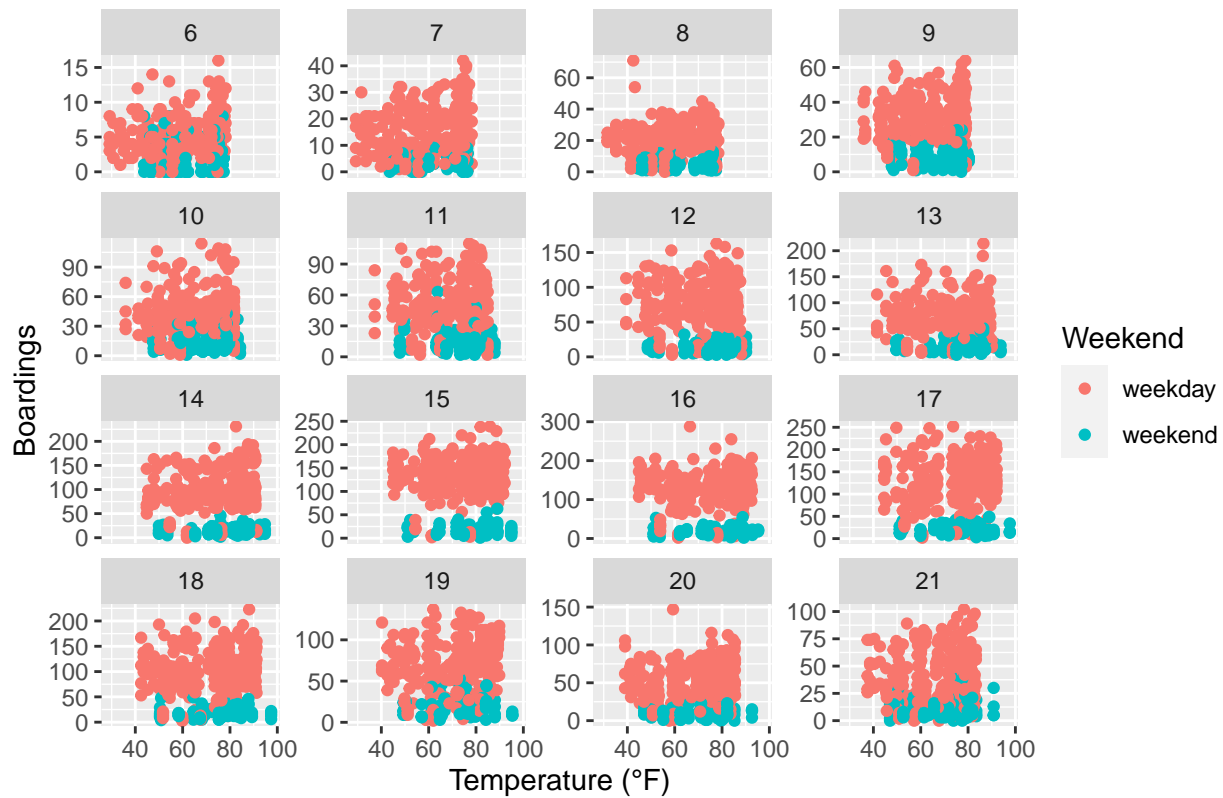
Part A



The graphs above display the average number of boardings separated by month, hour, and day of the week. Weekdays have similar peak times (4:00PM) while weekends do not have apparent peak times. Some significant metrics are the boarding in Mondays in September which appear lower possibly due to a holiday during that month. Similarly, Wednesday, Thursday, and Friday of November seem lower than the other lines as well possibly due to large amounts of people staying at home during Thanksgiving week.

Part B

Boardings vs. Temperature on weekends and weekdays by Hour of the day



The graph above displays the average number of boardings at different temperatures while highlighting points on weekdays and weekends separated by hour of the day. When the hour of the day and the weekend status is held constant, the temperature does have an impact on the number of UT students riding the bus. As you can see, in Hour 19 on weekdays, less people rode the bus when it was 40 degrees than when it was 80.

Problem 4: Wrangling the Billboard Top 100

Part A

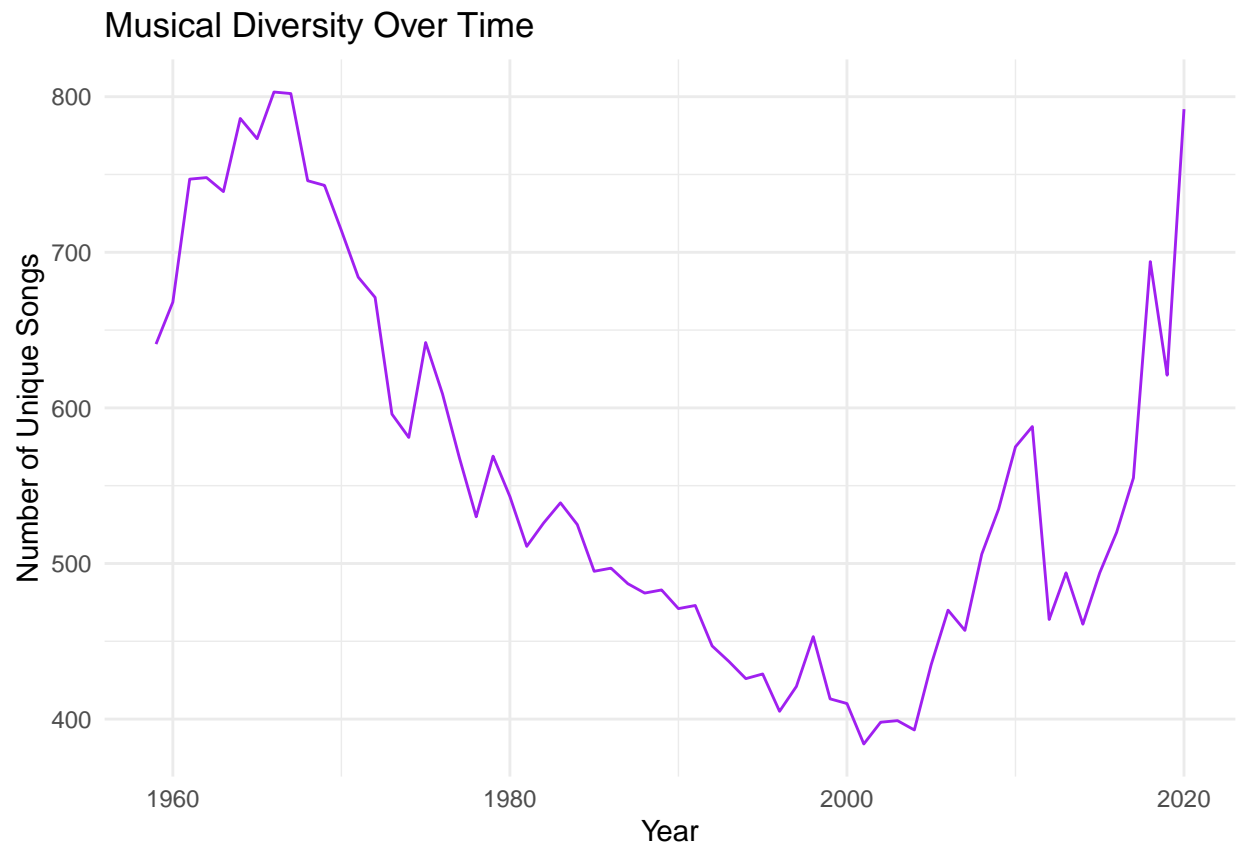
Table 1: Top 10 Most Popular Songs since 1958: Number of Weeks on Billboard Top 100

performer	song	count
Imagine Dragons	Radioactive	87
AWOLNATION	Sail	79
Jason Mraz	I'm Yours	76
The Weeknd	Blinging Lights	76
LeAnn Rimes	How Do I Live	69
LMFAO Featuring Lauren Bennett & GoonRock	Party Rock Anthem	68
OneRepublic	Counting Stars	68
Adele	Rolling In The Deep	65
Jewel	Foolish Games/You Were Meant For Me	65

performer	song	count
Carrie Underwood	Before He Cheats	64

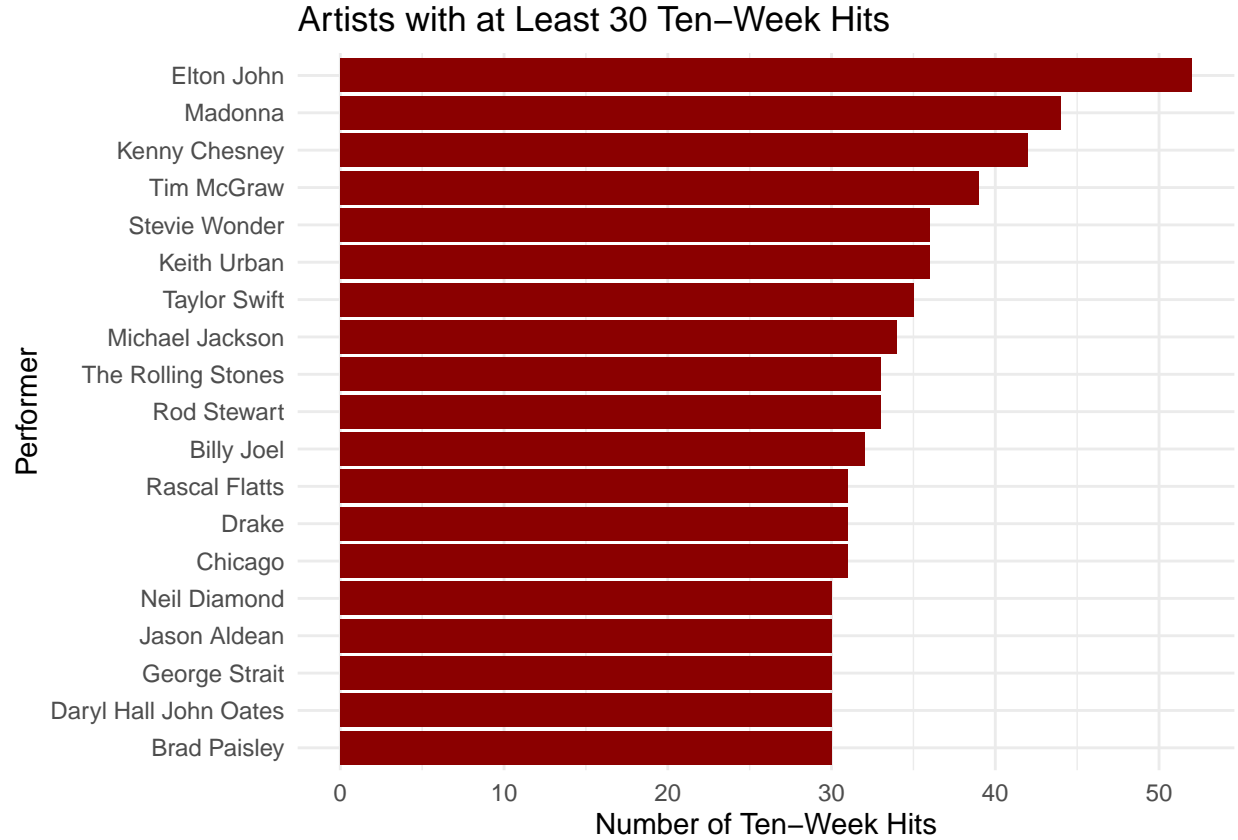
This table displays the top 10 most popular songs since 1958 according to the Billboard Top 100. Provided is the Performer, Song name, and the number of weeks the song appeared in the Billboard top 100.

Part B



The above graph measures the number of unique songs that appeared in the Billboard top 100 since 1958. The graph shows that around the year 2002 was the lowest number of unique songs, a huge dip from the year 1968 which was the highest. The graph seems to descend from 1960-2000 and increase from 2000-present.

Part C



This graph shows the artists that had at least 30 songs that were on the billboard top 100 for ten weeks. Elton John tops the list with 52 ten-week hits.