HW2

Brian Liang

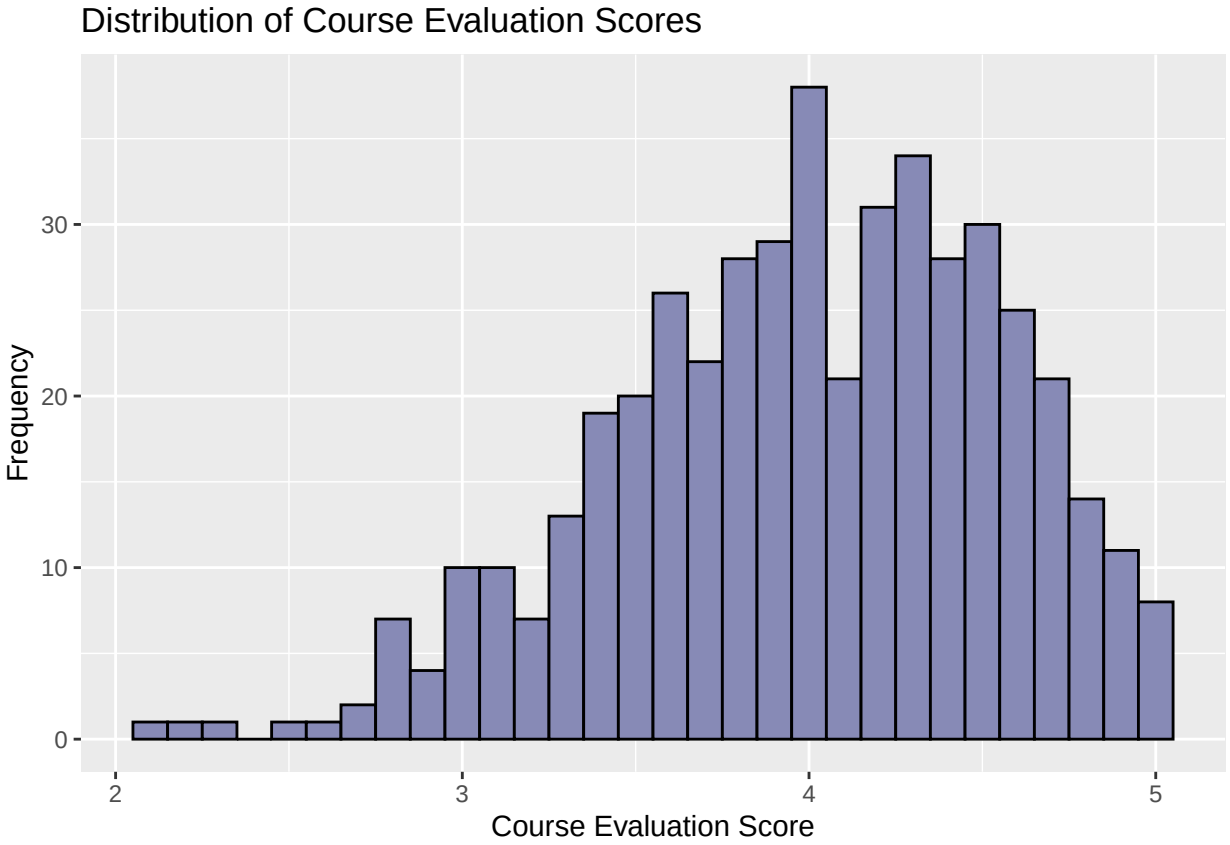
2025-01-22

UT EID: bl29375

GitHub Link: https://github.com/brianliangg/SDS315_HW2

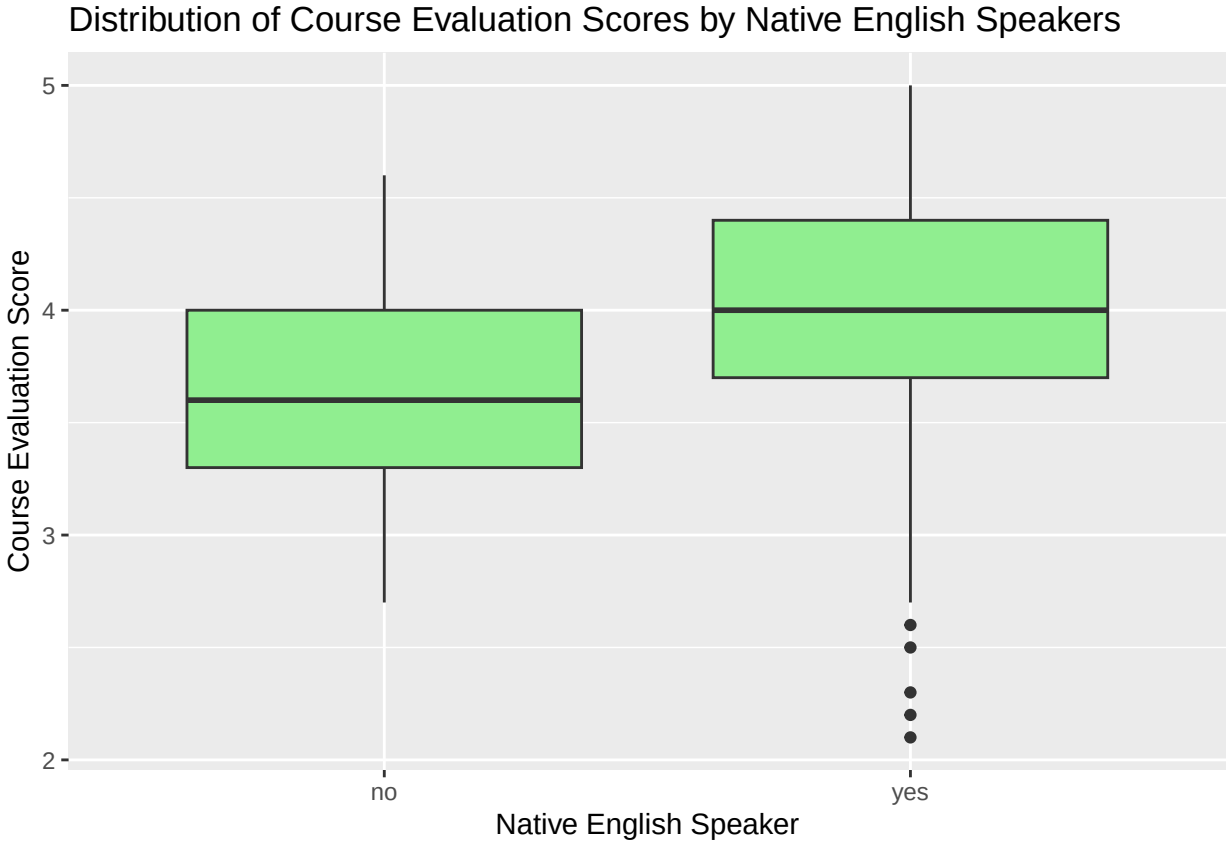
Problem 1

Part A



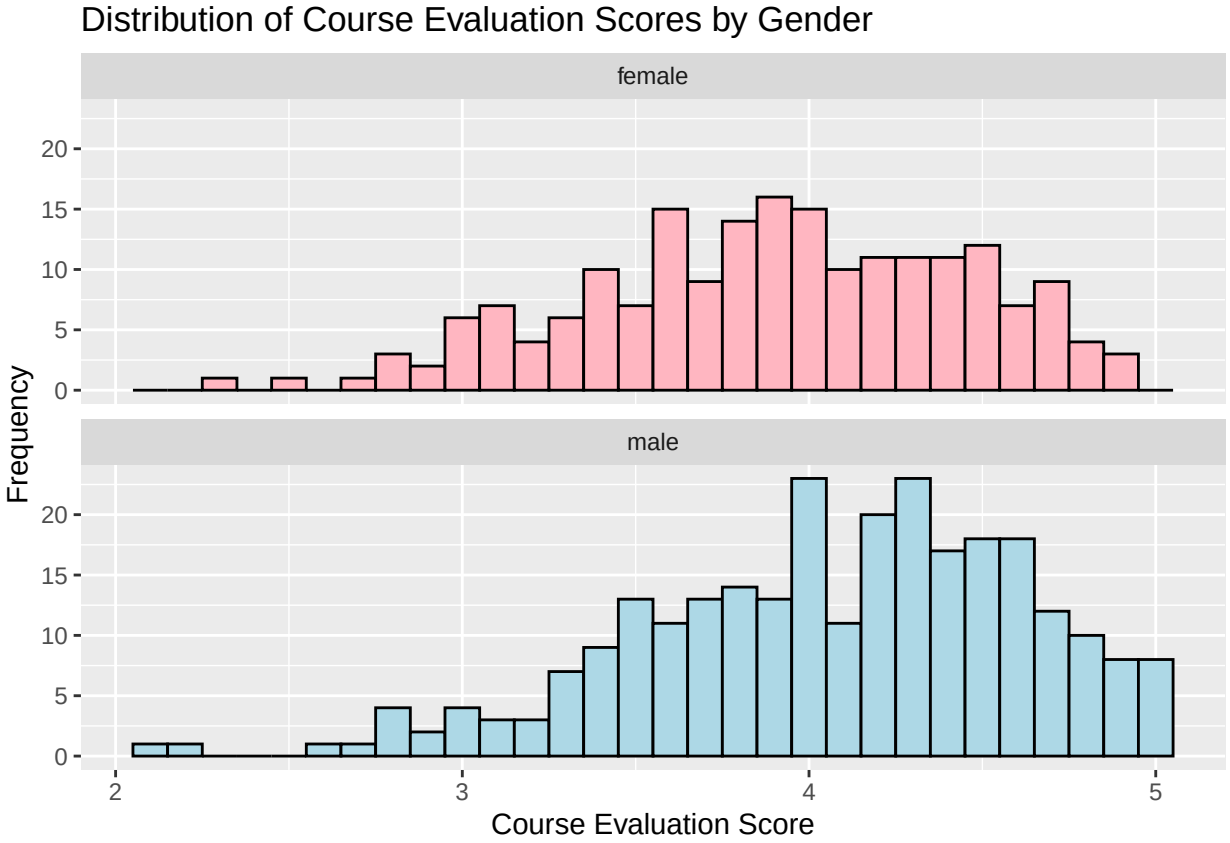
The histogram demonstrates a right skew meaning that a majority of the data lies on the higher end being centered at a Course Evaluation Score of around 4. This means that a majority of professors would have a course evaluation score at around 4 which means that professors from the dataset will have typically received a higher rating.

Part B



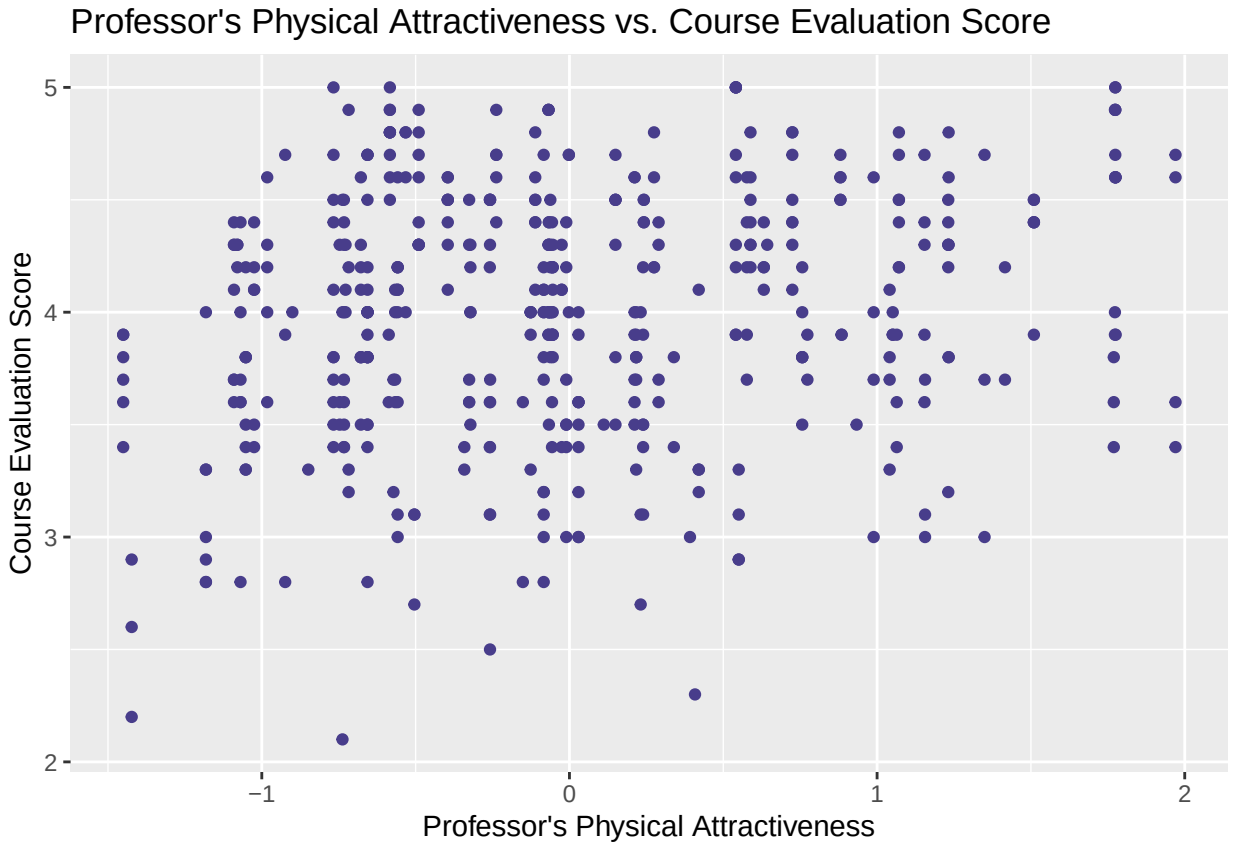
The boxplots show the distribution of Course Evaluation Scores by if the professors are Native English Speakers. Professors who were not Native English Speakers were shown to usually receive a lower Course Evaluation Score in comparison to professors who were Native English Speakers. However, professors who were Native English Speakers were the only ones to have outliers meaning that the professors who received the lowest Course Evaluation Scores were Native English Speakers.

Part C



The faceted histogram shows the distribution of Course Evaluation Scores by the Gender of the professor. Both of the distributions for female and male professors demonstrate a left skew meaning that a majority of the scores lie towards the higher end. These distributions display that male professors typically receive a slightly higher Course Evaluation Score than female professors as the distribution of male professors is centered around 4.2 and at around 4 for female professors.

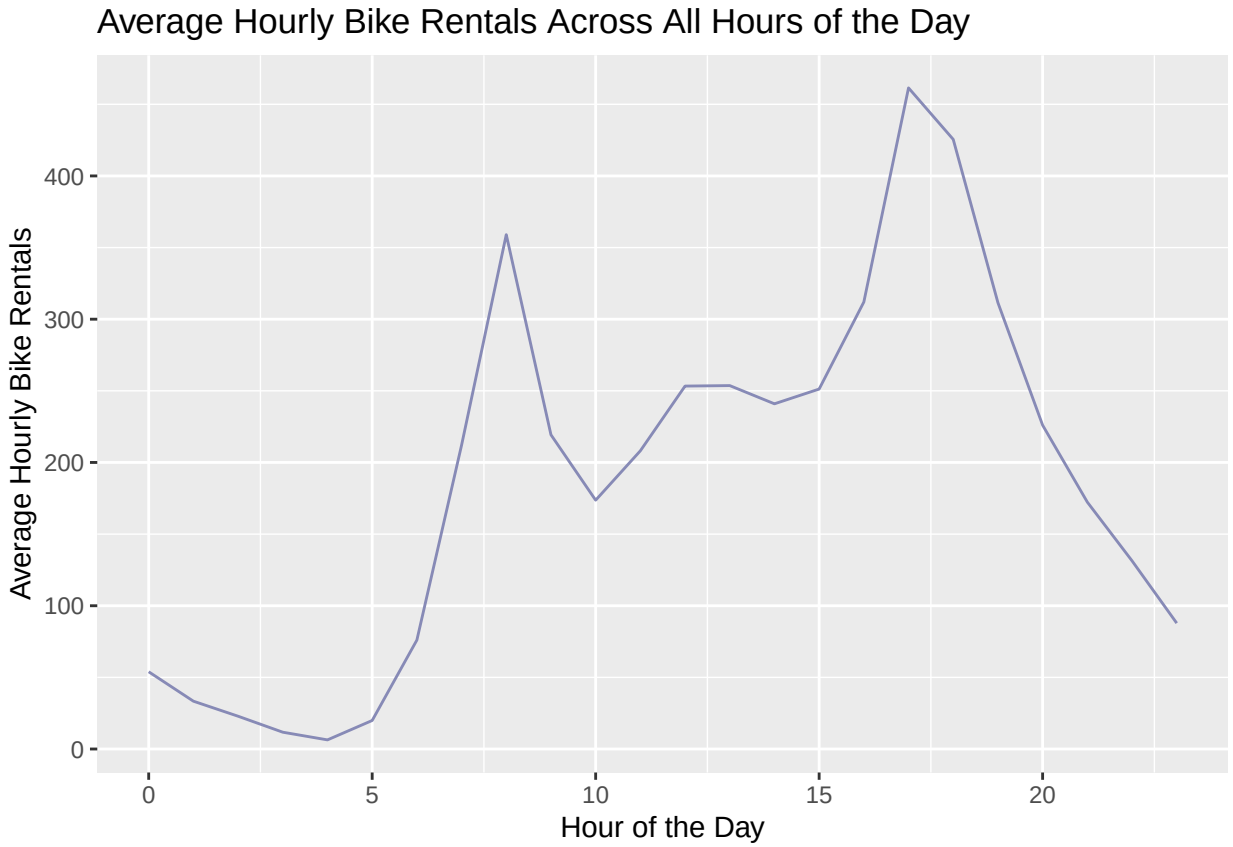
Part D



The scatterplot shows the relationship between a Professor's Physical Attractiveness and Course Evaluation Score. From the plot it is demonstrated that there is not pattern between the two variables meaning that there is no correlation between a Professor's Physical Attractiveness and Course Evaluation Score.

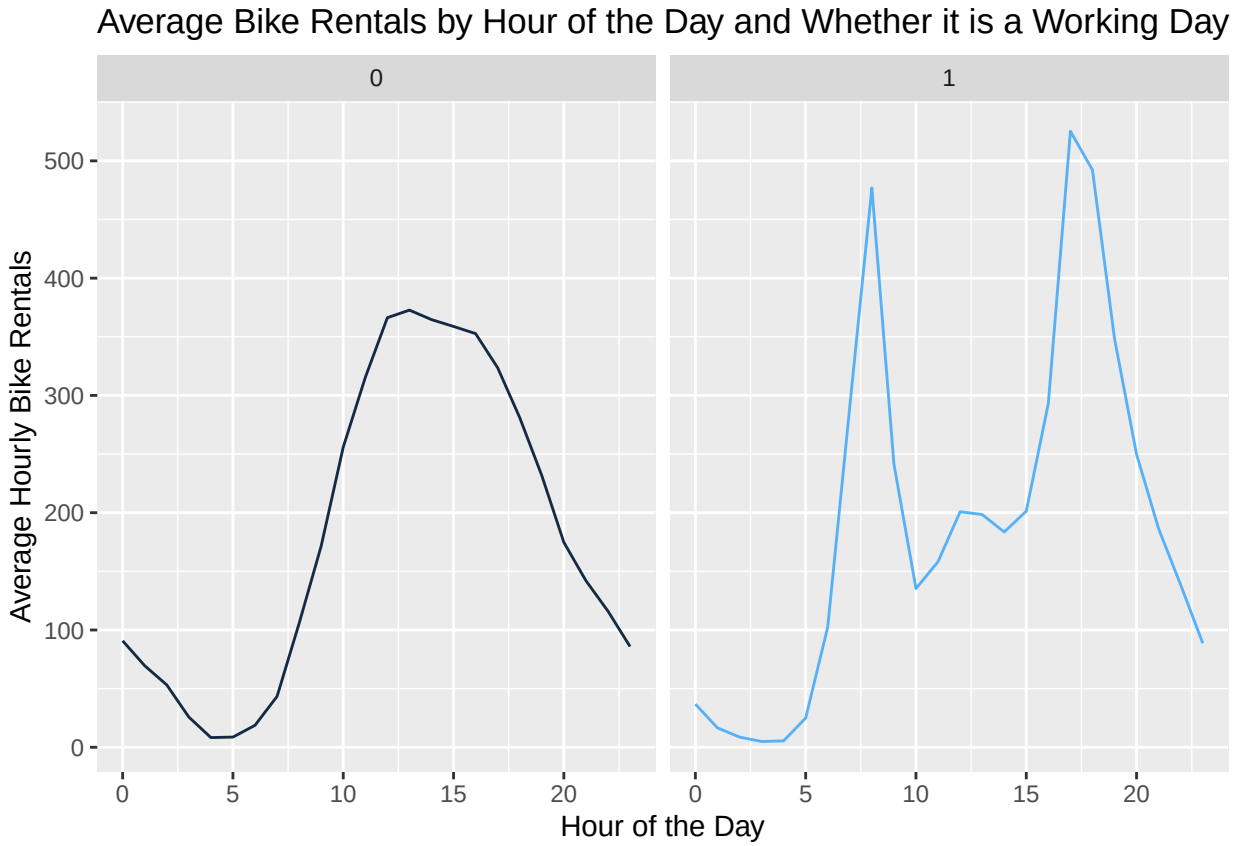
Problem 2

Plot A



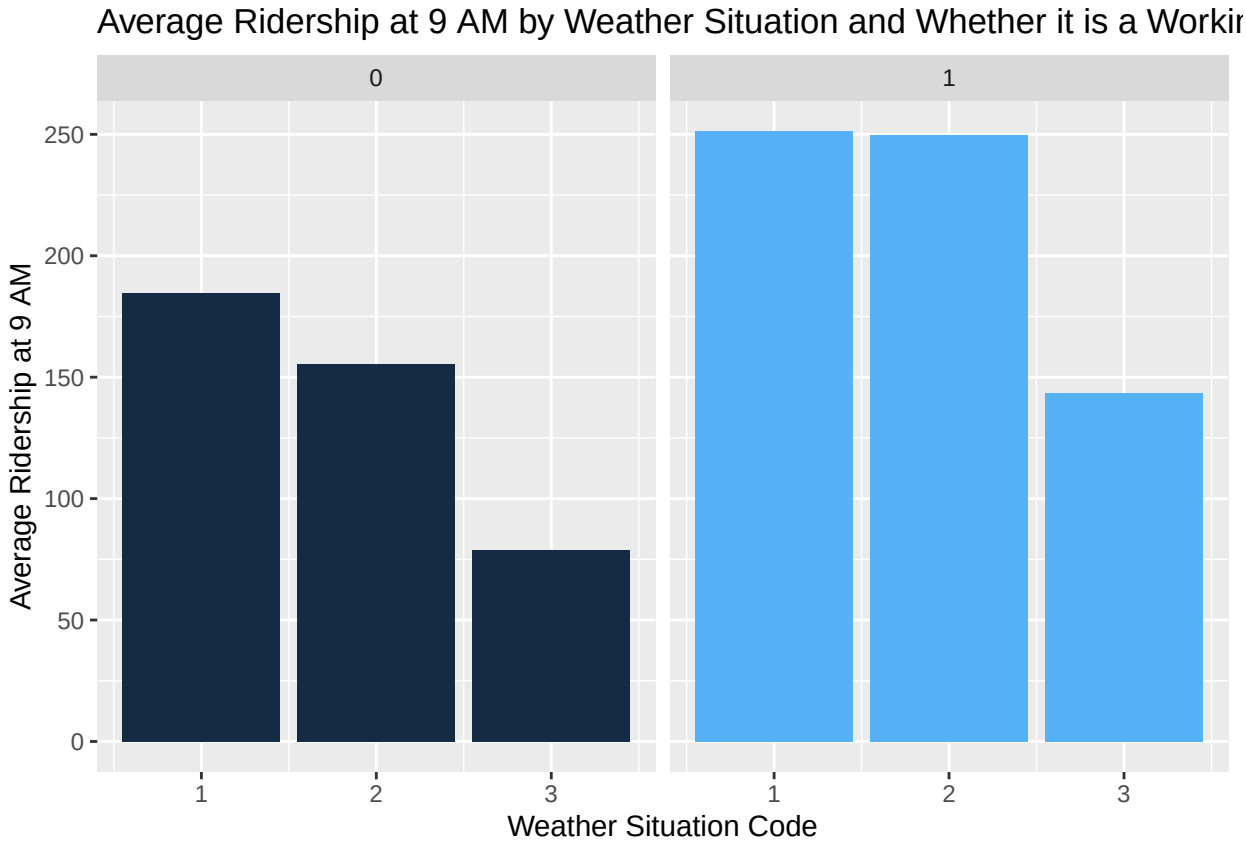
This plot displays the Average Hourly Bike Rentals, on the x-axis, across all Hours of the Day, on the y-axis. The plot starts from midnight being represented as hour 0 and goes to midnight of the next day being represented as hour 24. During the day, bike rentals begin to rise after 5:00 am to hit a peak of an average around 360 bike rentals at 8:00 am, and then drop but level to around 200 to 250 bike rentals from 10:00 am to 3:00 pm. Bike rentals will then begin to rise again to a high for the day of around 460 rentals at 5:00 pm and then decline for the rest of the day. From this pattern, we learn that ridership aligns similarly to rush hour times when people are going to work and then later in the day leaving work.

Plot B



This plot is similar to the previous plot in that it displays the Average Bike Rentals, on the x-axis, by Hour of the day, on the y-axis, but creates two line graphs based on whether it is a working day or not. If the day is neither a weekend nor a holiday, it is considered a working day and is represented by the second panel with a 1. The first panel with a 0 considers days that are not working days. It is demonstrated that the the working days graph is similar to our previous Average Hourly Bike Rentals Across All Hours of the Day plot as there are peaks during rush hour times, however with higher average bike rentals, but lower averages between the two rush hour times. These lower averages are likely explained by the exclusion of non-working days because from the non-working days graph, it is shown that bike rentals rise from 5:00 am to a peak at 1:00 pm and then decline throughout the rest of the day. From both of these plots, it is shown that there is a difference between the Average Hourly Bike Rentals across all Hours of the Day based on whether it is a working day or not.

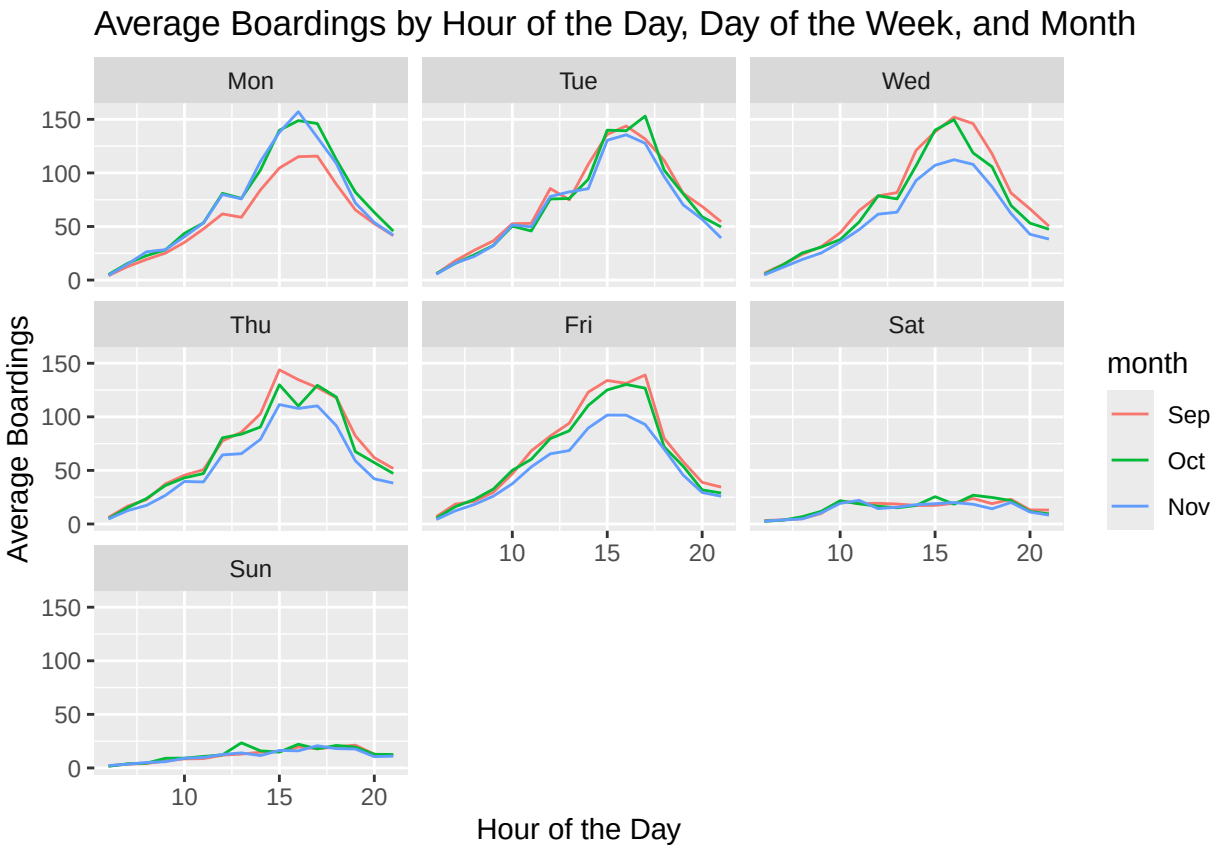
Plot C



This plot displays the Average Ridership during the 9 AM hour by the Weather Situation Code and if it is a Working Day. The Weather Situation Code, on the x-axis, signals what the weather was on that day with 1 being Clear, Few clouds, Partly cloudy, or Partly cloudy; 2 being Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, or Mist; 3 being Light Snow, Light Rain + Thunderstorm + Scattered clouds, or Light Rain + Scattered clouds; and 4 being Heavy Rain + Ice Pallets + Thunderstorm + Mist, or Snow + Fog. However, there were no recorded bike rentals at 9 AM during weather situation code 4 which is why it is not shown in the plots above. Average Ridership, on the y-axis, was determined by the average number of bike rentals at 9 AM, and there are two panels signaling if it is a non-working day, first panel with 0, or a working day, second panel with 1. From the plots, it is demonstrated that the Average Ridership at 9 AM is higher if it is a working day in comparison to if it is not a working during the same weather situation, other than for weather situation 4.

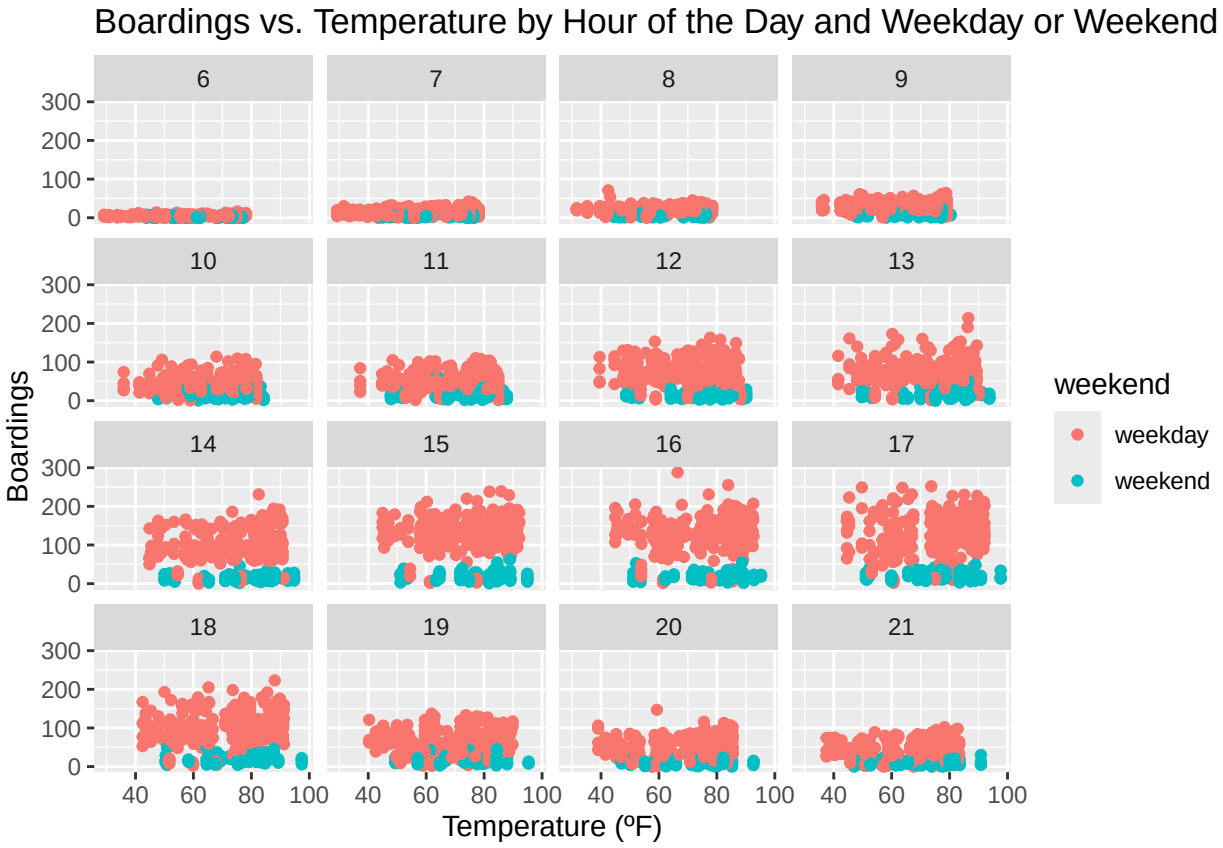
Problem 3

Question 1



The line graphs shows the Average Boardings by Hour of the Day, Day of the Week, and the Month. It is demonstrated that on weekends, Saturday and Sunday, Average Boardings will be lower than compared to the weekdays. And on the weekdays, Average Boardings will look broadly similar and are higher in comparison to weekends. I think that the Average Boardings on Mondays in September look lower compared to other days and month because of Labor Day. This means more people staying home as a result which in turn means less boardings during that day. I think average boardings on Weds/Thurs/Fri in November look lower because of Thanksgiving as people will typically not have to work during those days, leading to less boardings.

Question 2



The figure shows the Boardings vs. Temperature by Hour of the Day and Weekday or Weekend. From the figure, it is shown that boardings will typically be higher on weekdays rather than on weekends. But also when we hold hour of day and weekend status constant, temperature does not seem to have a noticeable effect on the number of UT students riding the bus.

Problem 4

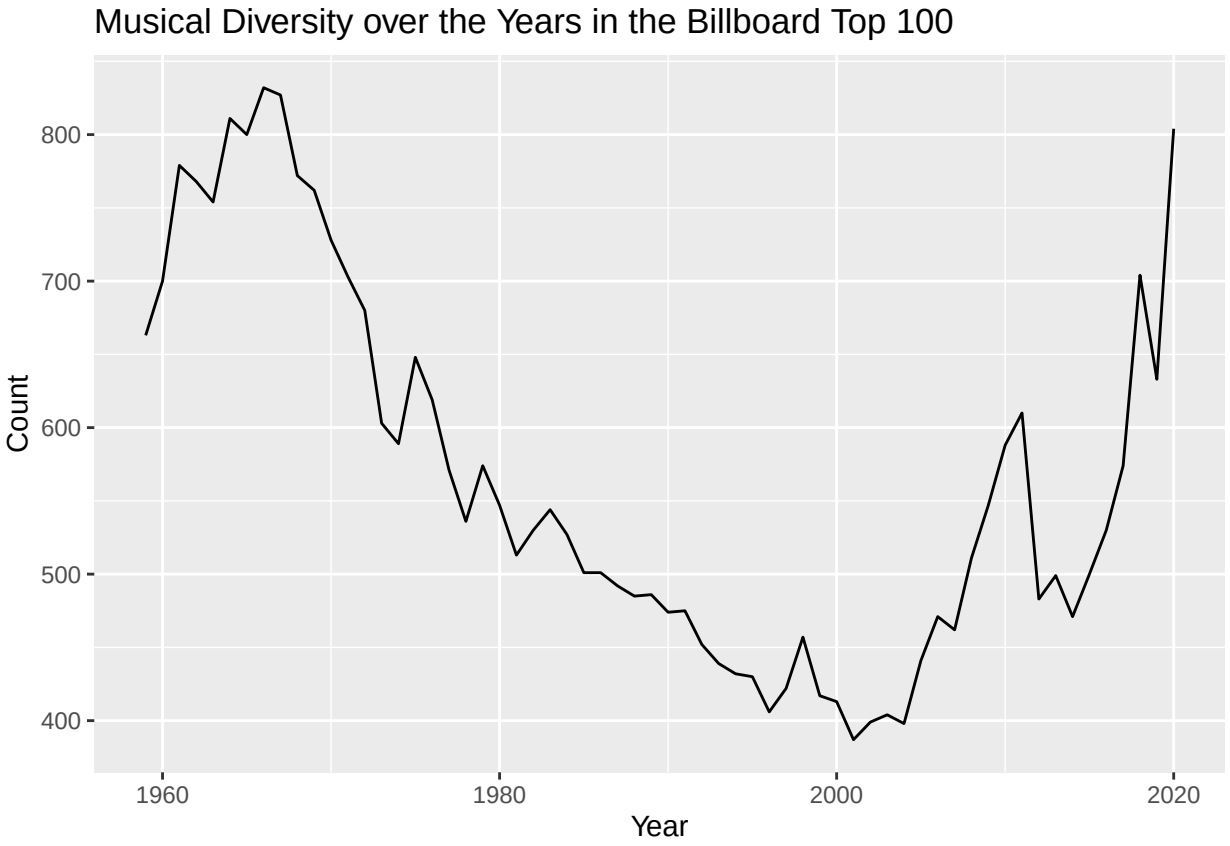
Part A

Table 1: Top 10 Most Popular Songs on Billboard Top 100 Since 1958

Performer	Song	Count
Imagine Dragons	Radioactive	87
AWOLNATION	Sail	79
Jason Mraz	I'm Yours	76
The Weeknd	Blinding 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
Carrie Underwood	Before He Cheats	64

The table shows the top 10 most popular songs since 1958 in descending order, as measured by the total number of weeks that a song spent on the Billboard Top 100.

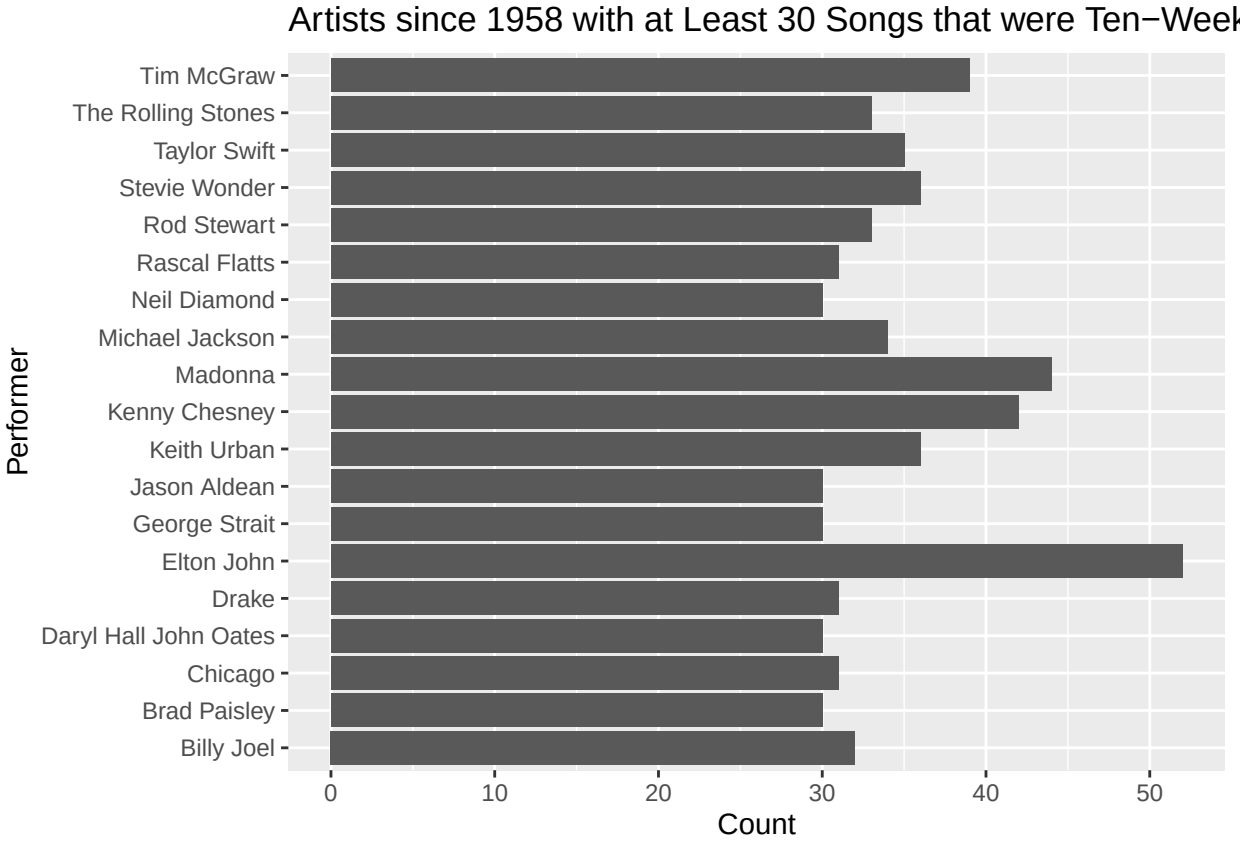
Part B



The figure displays the musical diversity over the years in the Billboard Top 100, which was found by the number of unique songs during that given year. The years 1958 and 2021 were also excluded as the we do not have the complete data for these years. An interesting trend that can be observed is musical diversity is quite

high to hit a peak at around the year 1966, but then decline until around the year 2000. After this, musical diversity rose to another peak at the year of 2020.

Part C



The figure shows the Artists since 1958 with at Least 30 Songs that were Ten-Week Hits. From what is displayed, Elton John has the most songs that were ten-week hits with over 50, followed by Madonna with 44, and then by Kenny Chesney with 42.