HW2 - SDS 315

Ahantya Sharma

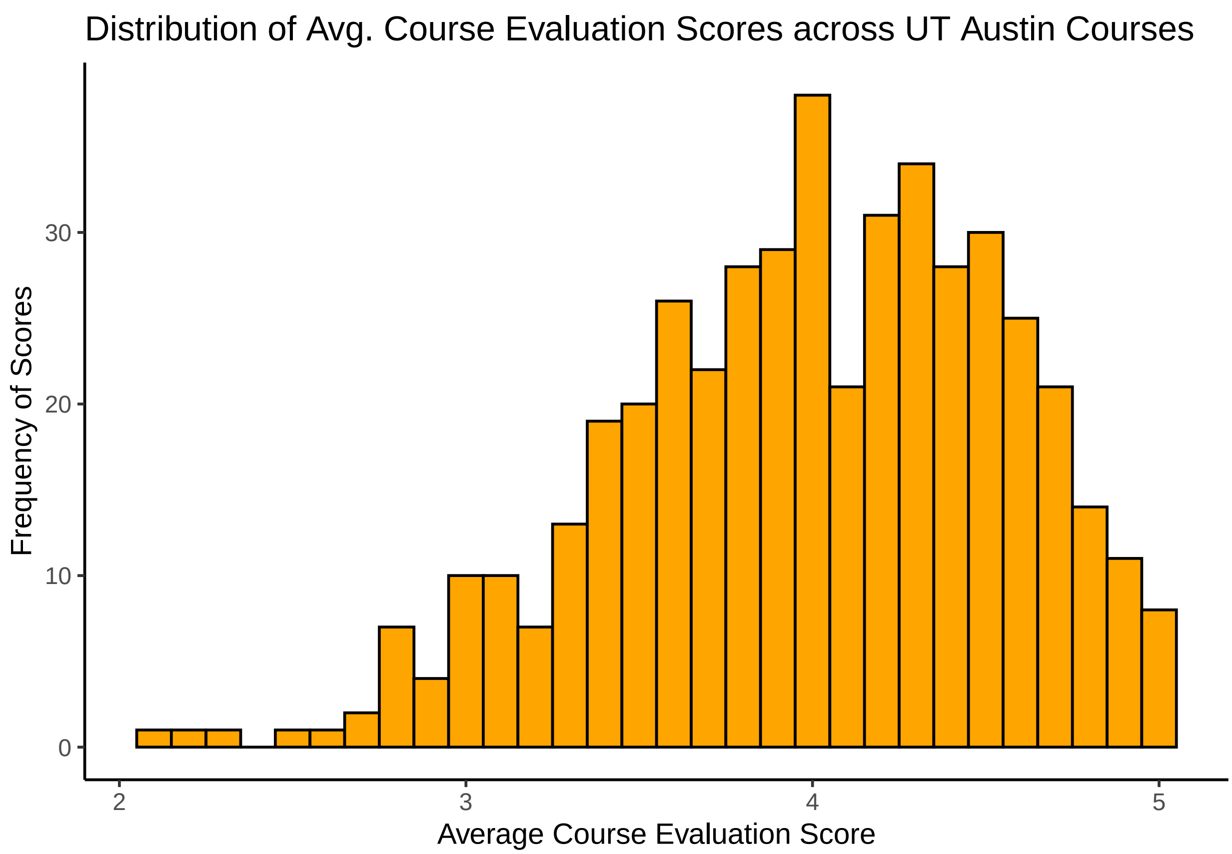
2025-01-18

UTEID: as236366

Github Link: https://github.com/Ahantya/SDS315/blob/main/HW2/HW2Markdown.Rmd

Question 1 - Professor Course Evaluations

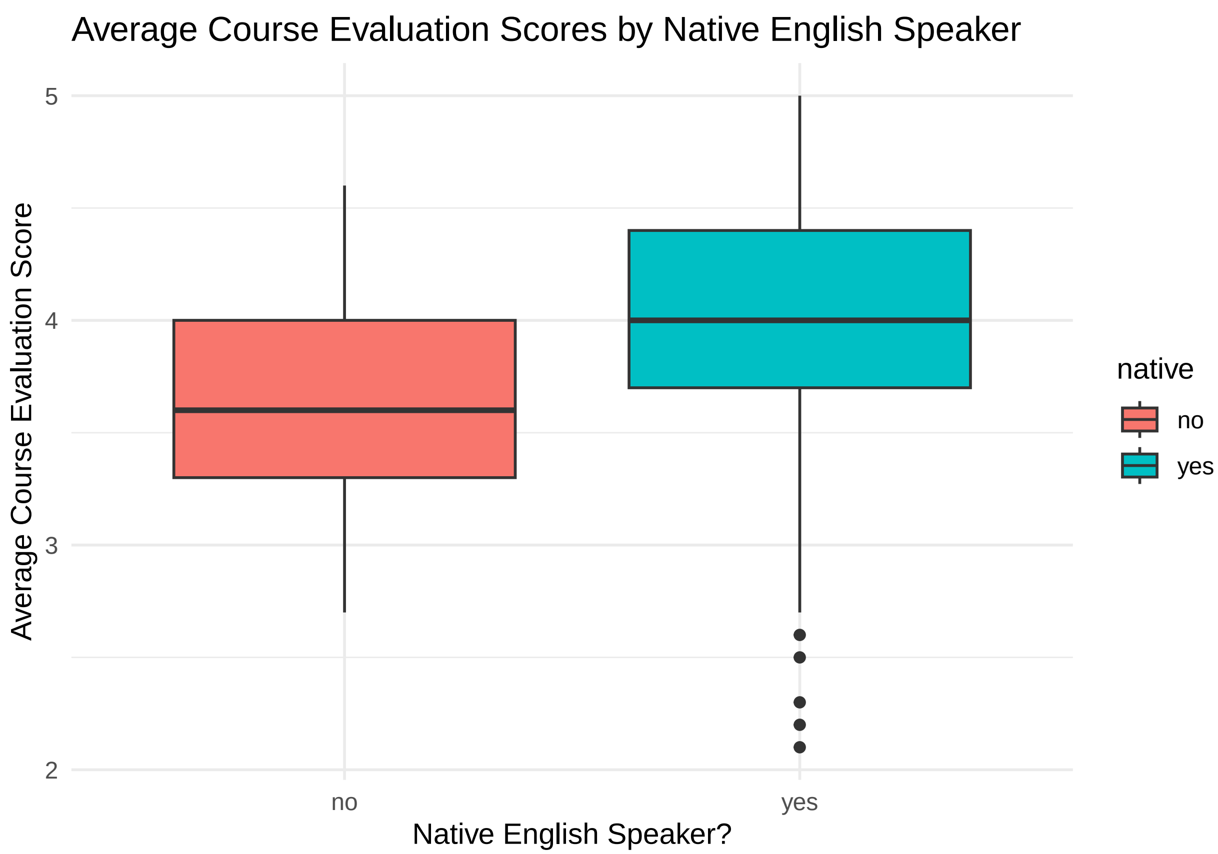
Part A



The distribution of the Average Course Evaluation Scores at UT Austin shows scores that UT Students rated courses from 1 to 5.

Looking at it, it's skewed to the left, and the mean of the distribution is 3.998, implying that a majority of the courses are rated positively. The highest frequency of scores is around 4.0 to 4.5, which explains that professors at UT Austin are well-regarded by students when it comes to handling their classes. Along with this, there are very few courses rated below 3 on average, which suggests that UT Austin is keeping up well in course quality.

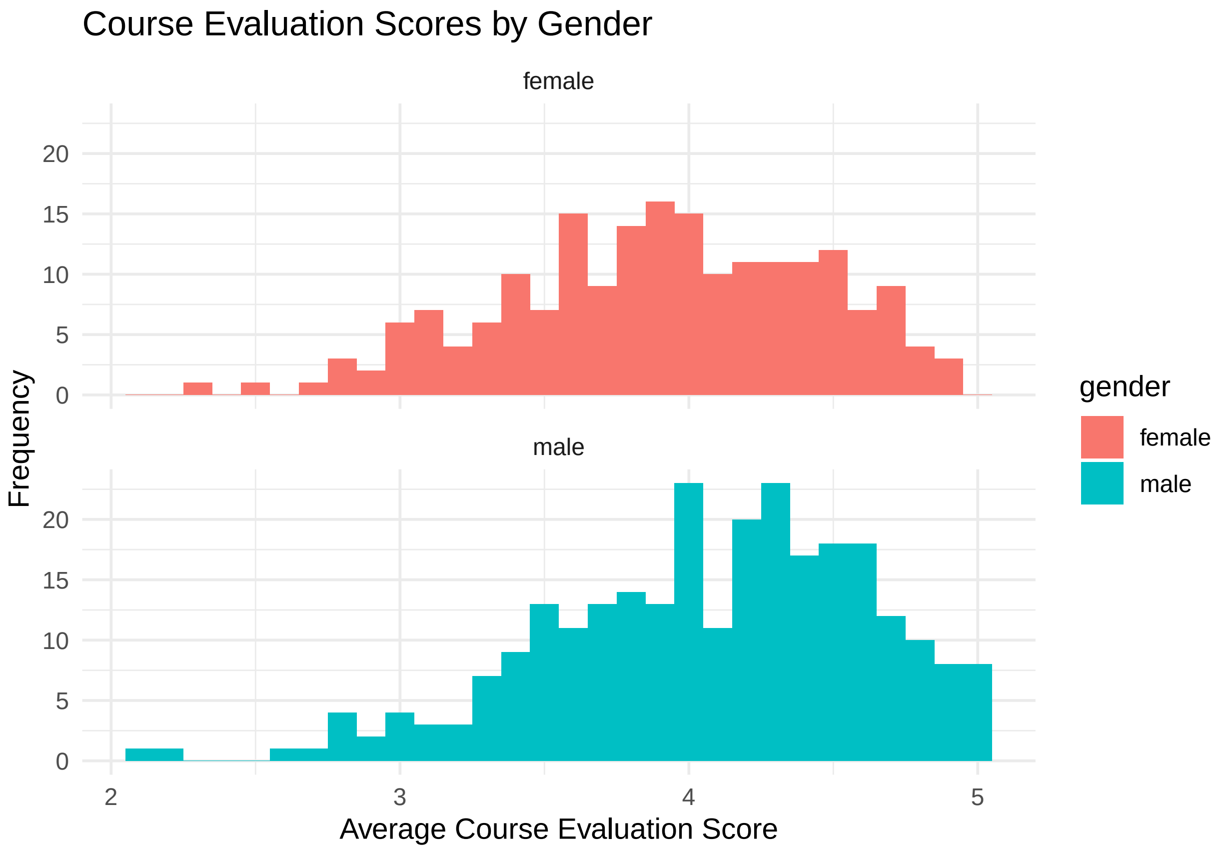
Part B



These side to side boxplots show the distribution of average course evaluation scores split up by whether the professor of the course was a native or non-native English speaker.

They show that while the lowest recorded average scores may come from courses with a professor who is a native English speaker, they also have a higher median score on average at 4 compared to non-native English speakers at 3.6 and contain the highest recorded average scores as well. This is most likely because students may have an easier time understanding the lectures, leading to higher evaluations overall.

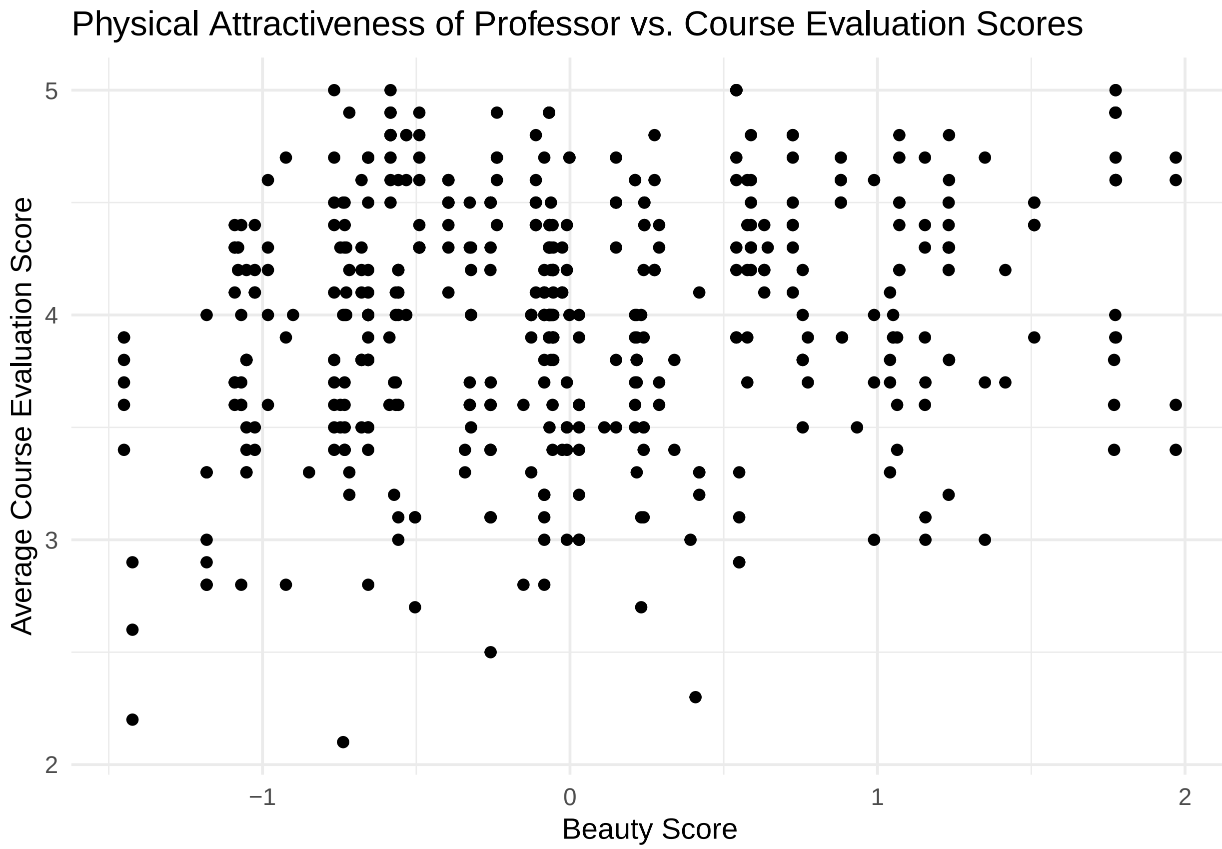
Part C



This graph shows the distribution of the average course evaluation score by gender.

Both distributions are still left-skewed, showing that most professors received higher evaluation scores closer to 5 regardless of gender. The range of scores is similar for both genders, suggesting not a very big difference in the overall evaluation scores based on gender. The mean of the (average) course evaluation scores for female instructors was 3.9 while the mean for male instructors was 4.07. The counts also reveal that there are more male professors represented in the data than female professors, as shown by the higher bars across the male histogram.

Part D

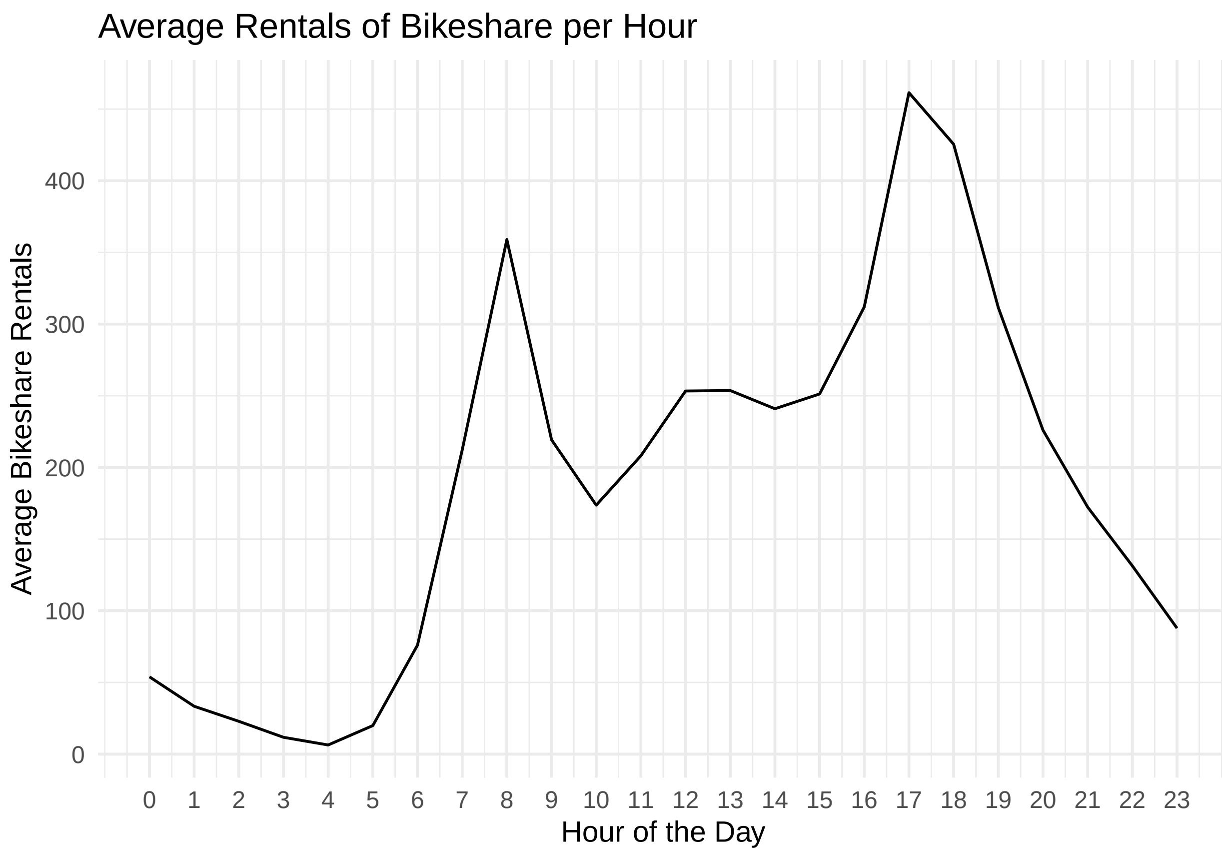


This scatterplot shows the correlation between the physical attractiveness (on a scoring scale) of a instructor vs their average course evaluation score.

Looking at this scatterplot, the points are scattered without a clear linear trend. Specifically, the correlation coefficient score is 0.19, which shows that there is a very weak, positive correlation that indicates that physical attractiveness does not predict course evaluation scores well. Course evaluation scores are instead effected much more by other factors, such as what we found in Part B of whether the professor of the course is a native or non-native English speaker.

Question 2 - Bike Sharing

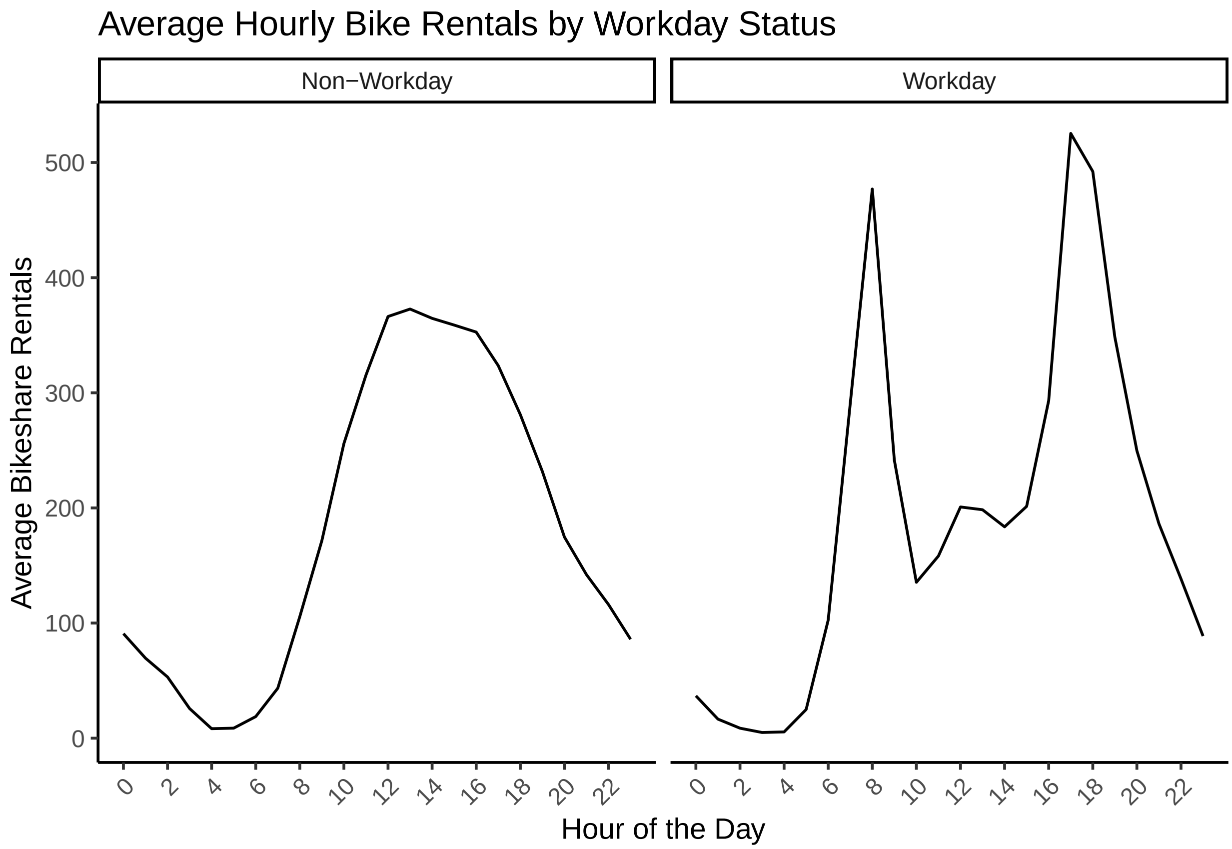
Plot A



This line graphs plot each hour as the x-value and the Average Bikeshare Rentals as the y-value, as a way to show the average amount of bikeshare rentals per hour.

Looking at the line graph, we can see that bikeshares pick up very quick in the early morning and the evening, most likely because people are in need of transportation to get to school or work at those times. As the evening turns into late night, the average rentals per hour slow down as most people are most likely sleeping and are in no need of getting anywhere.

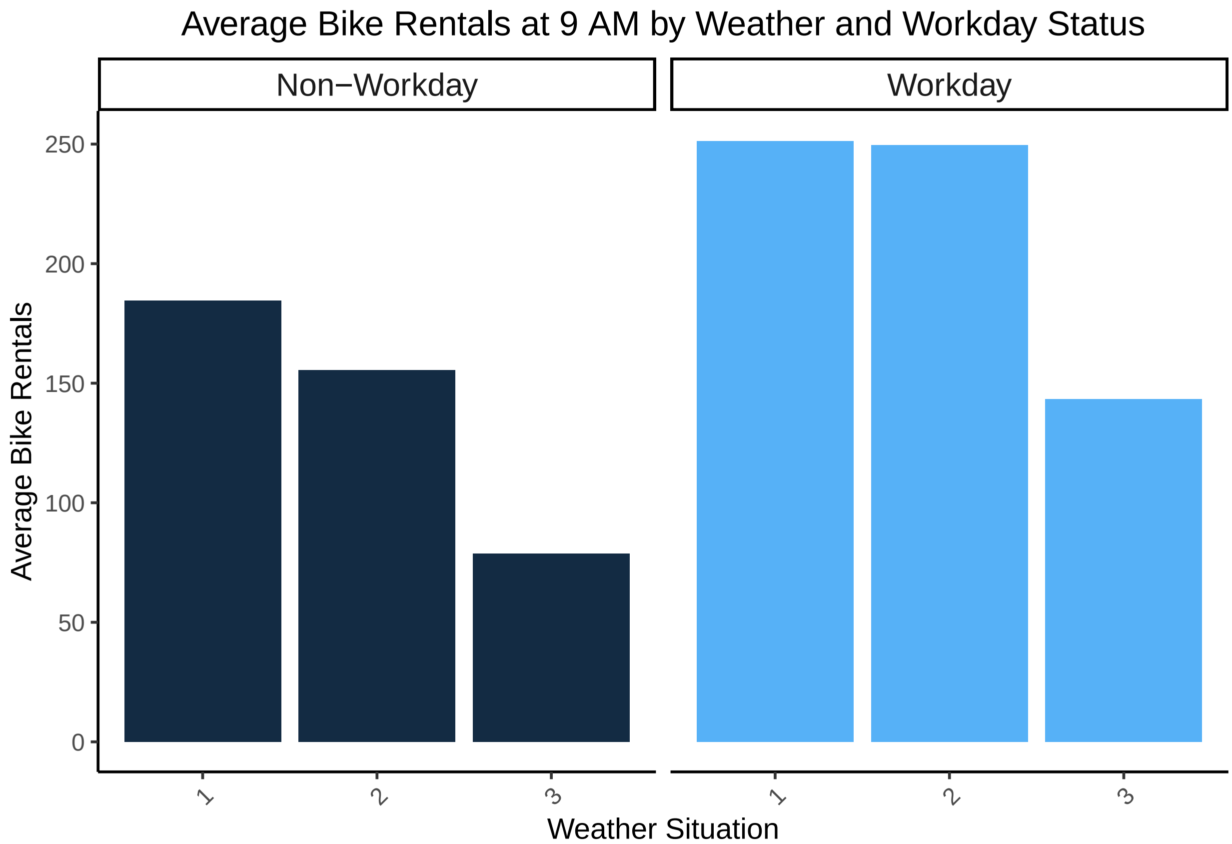
Plot B



This line graph is similar to the one showed in Part A, however it splits the average rental bikeshare data into the working days and non-working days. The x-value still represents the Hour of the Day while the y-value still represents the amount of average bikeshare rentals. The original data had the variable "workingday" listed as 0 or 1, which had 0 represent non-working day and 1 as a working day, which I changed through my R code.

By the figure, there is two bumps on the working day as those are the times when people are going to and leaving from work / school. On a non-working day, this trend doesn't exist and instead bikeshares steadily increase for the first half of the day and decrease for the second half of the day.

Plot C

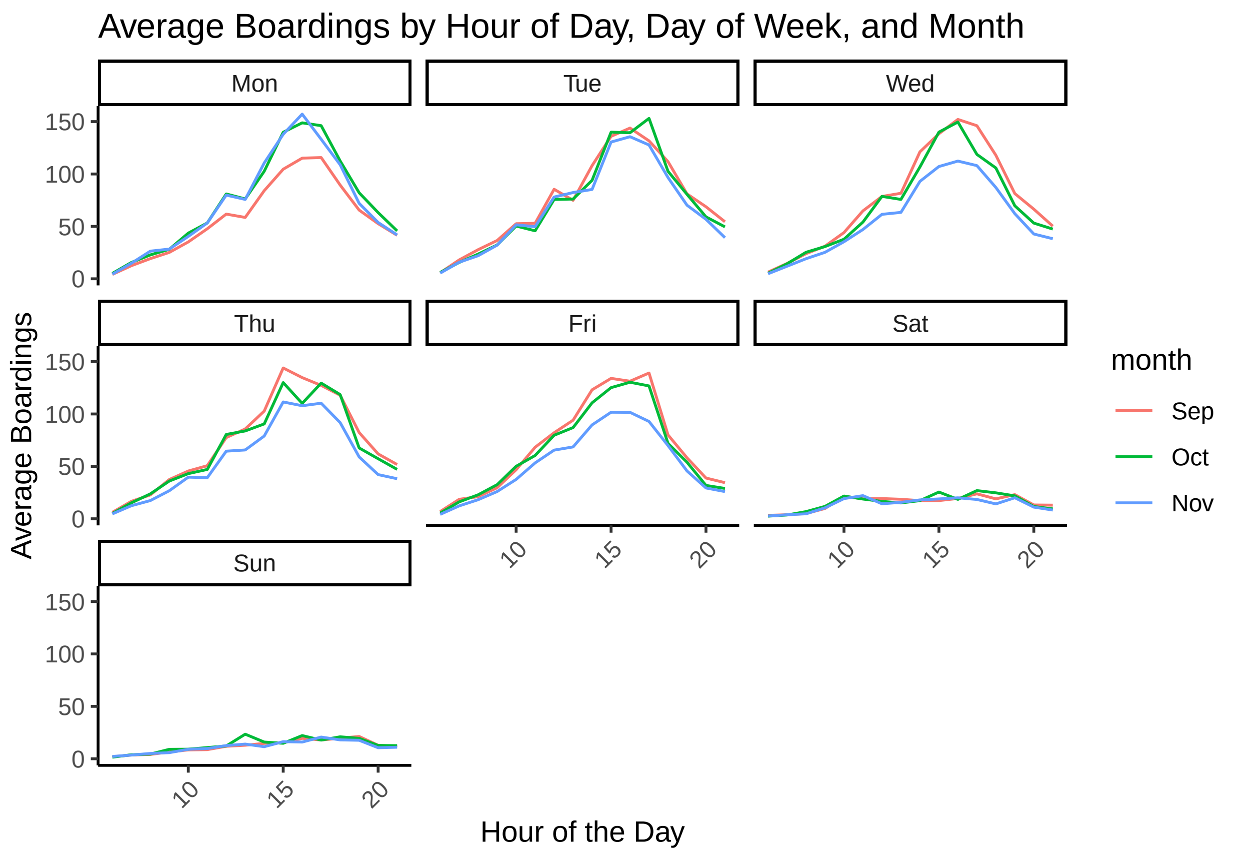


In this bar plot, the average bike share rentals (y) are displayed for different weather situations (x) at 9 AM on workdays and non-workdays. The numbers correspond to specific weather conditions: 1 represents clear days, few clouds, or partly cloudy conditions; 2 represents mist combined with cloudy, broken clouds, or few clouds; 3 represents light snow, light rain with thunderstorms, or scattered clouds; and 4 (which has no data points at 9 AM) represents days with heavy rain, ice pallets with thunderstorms and mist, or snow and fog.

By the barplot, workdays have heavier bike share rentals at 9 AM in each weather situation vs non-workdays. As the weather has worse conditions, the average bike share rentals also decreases as people don't want (or have a need) to go out.

Problem 3 - Capital Metro UT Ridership

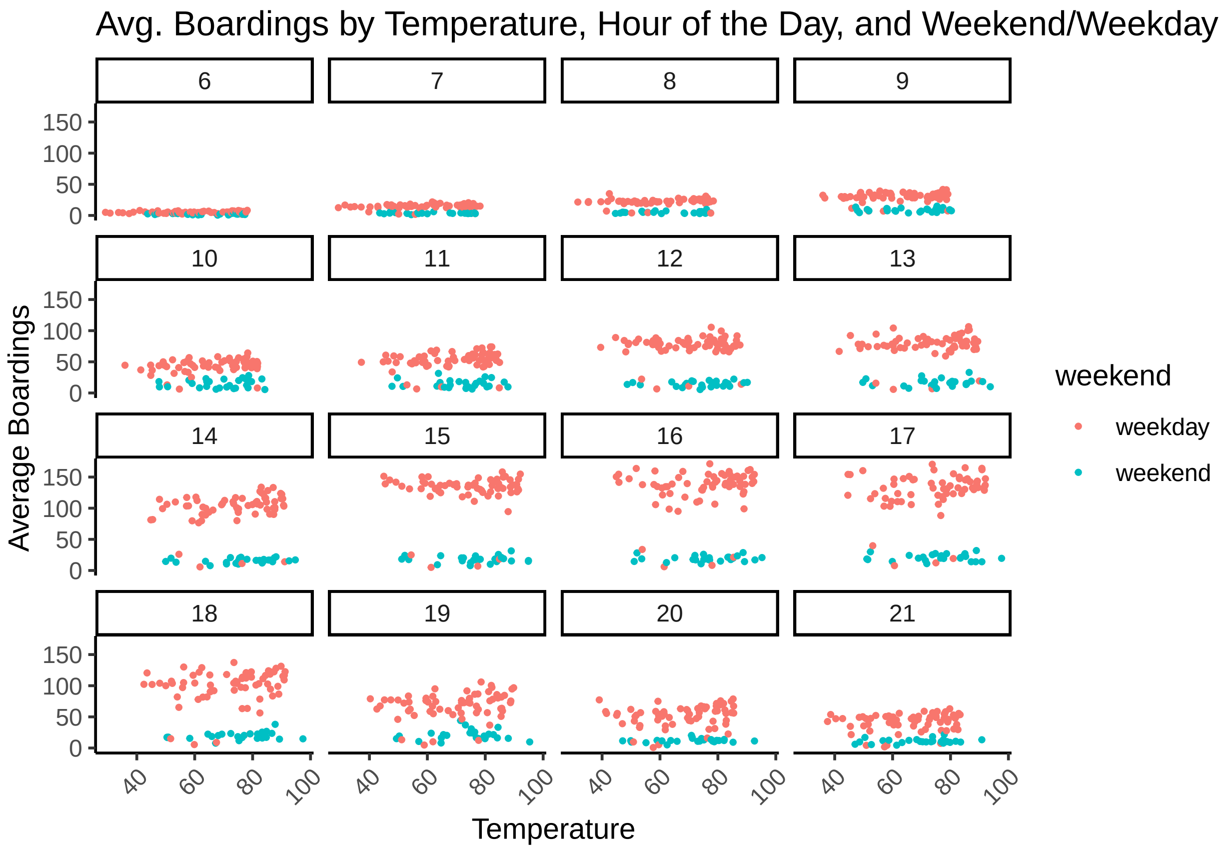
1.



This graph displays the average boardings (y-axis) by the hour of the day (x-axis) for each day of the week, with each facet representing a specific day. Within each facet, three lines correspond to the average boardings for each month (September, October, and November), allowing for a comparison of trends across months.

From the figure, we can tell that the hour of peak boardings is consistent across weekdays. But, in overall, there are apparent changes regarding the hour of peak boarding from day to day because the weekends have a different trend than the weekdays and seem to either barely peak average boardings a little earlier in the day (than weekdays) or seem to stay the same consistently the whole day. The lower average boarding on Mondays in September could be because of the potential holidays in early September like Labor Day and many students are just getting back from summer break so school isn't in full action. The average boundaries on Weds/Thurs/Fri in November look lower most likely because of Thanksgiving break when many students are home and aren't on campus taking buses.

2.



This scatterplot displays the relationship between temperature and average boardings for UT students, segmented by hour of the day and weekend/weekday status. Each panel corresponds to a specific hour of the day, while the color indicates whether the data represents weekdays (red) or weekends (blue). The x-axis shows the temperature, and the y-axis represents the average boardings.

When the hour of the day and weekend/weekday are held constant, the temperature seems to have very little or no effect on the number of UT students who ride the bus. This is most likely because temperature is affected by the hour of the day, so if you hold the hour of the day constant, then the temperature is probably not changing too much, which is a domino effect and also means that the amount of UT students riding the bus isn't too much.

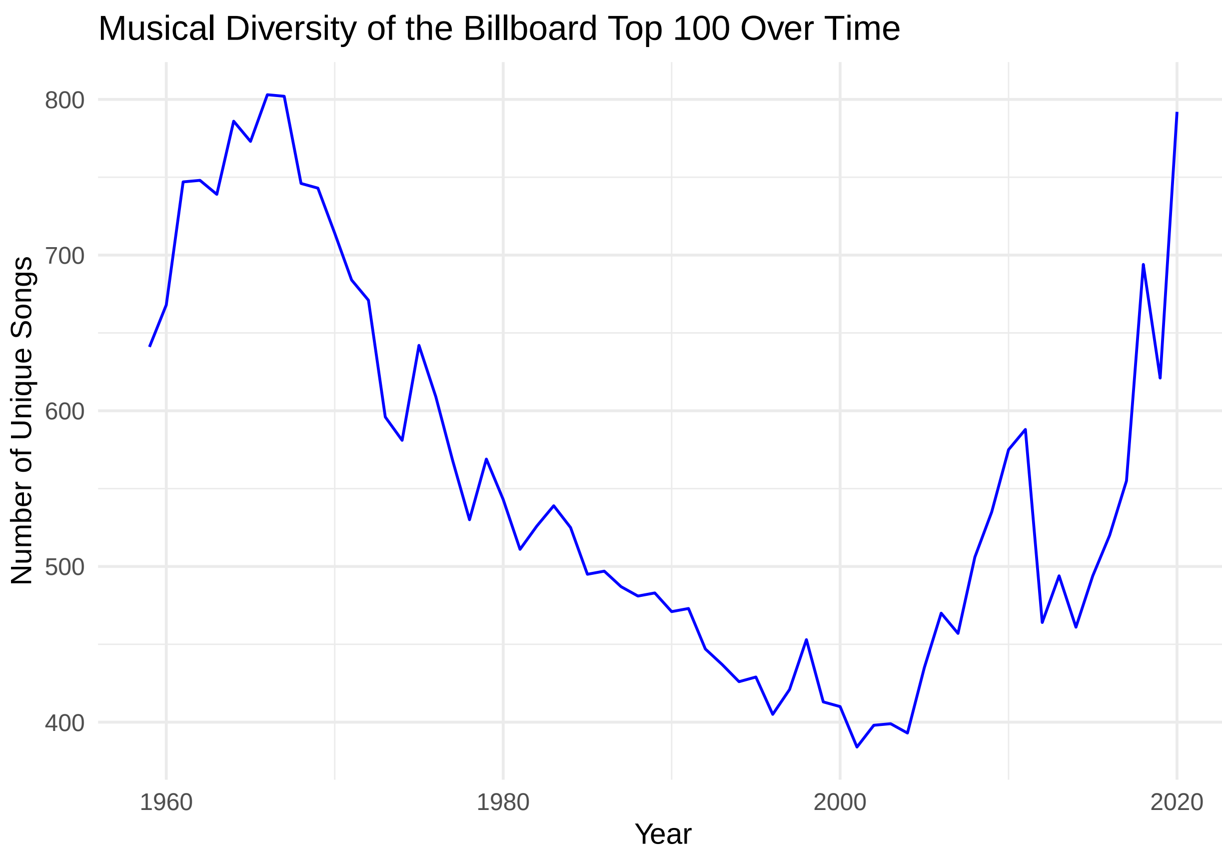
Problem 4: Wrangling the Billboard Top 100

Part A

Performer	Song	Weeks on Chart
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

This table shows the Top 10 most popular songs since 1958 based on the total number of weeks on the Billboard Top 100. The table includes the performer, song title, and the number of weeks each song appeared on the chart.

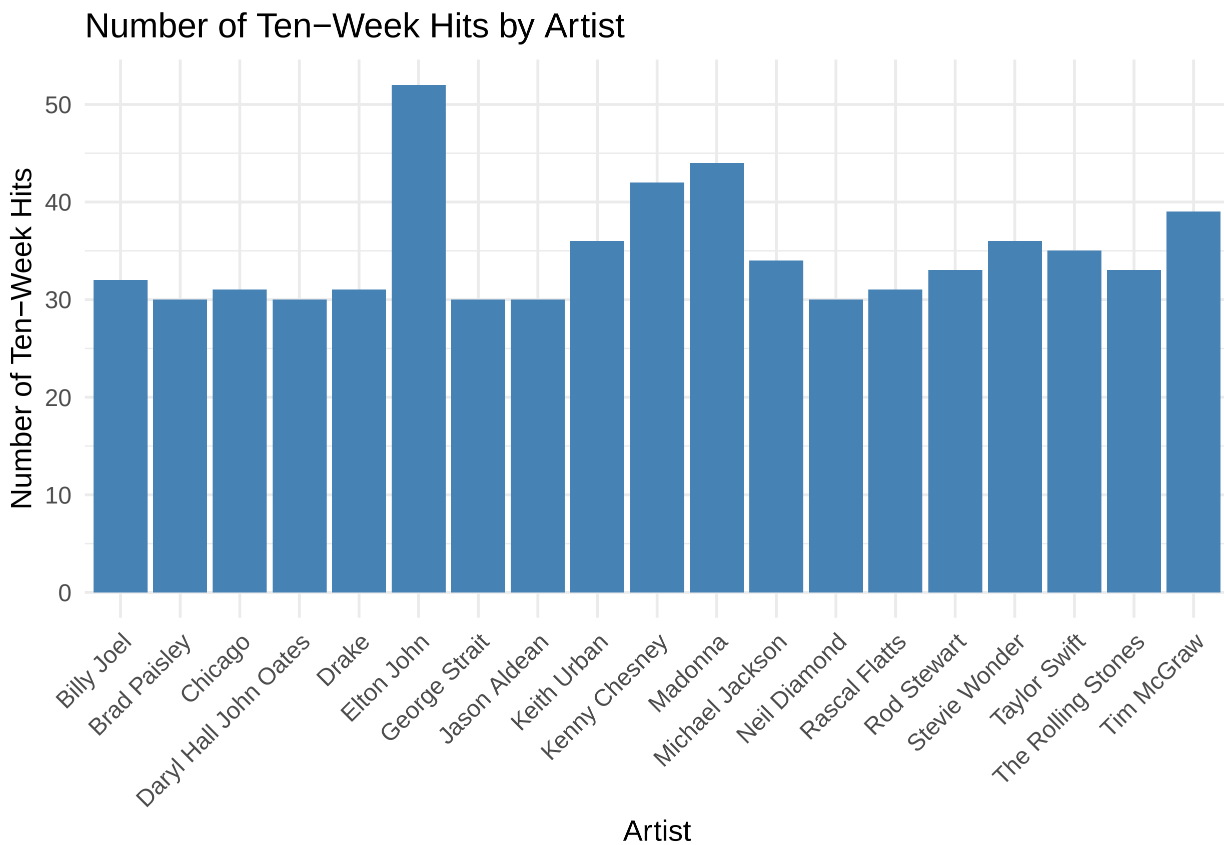
Part B



This graph shows the musical diversity of the Billboard Top 100 over time, measured as the number of unique songs that appeared on the chart each year. The x-axis represents the year, while the y-axis shows the count of unique songs. Data from the years 1958 and 2021 were excluded due to incomplete records.

From the graph, we can see that musical diversity increased from the late 1950s to mid 1960s where it plummeted mostly steadily until around the early 2000s where it has gradually increased minus a dip in the early 2010s. The heavy increase from 2019 to 2020 may have been new songs being discovered during the COVID-19 Lockdown.

Part C



This bar plot illustrates the number of "ten-week hits" achieved by 19 artists in U.S. musical history since 1958. A "ten-week hit" is defined as a song that remained on the Billboard Top 100 for at least ten weeks. The x-axis lists the artists, and the y-axis shows the total count of "ten-week hits" for each artist. The plot is meant to show the popularity of these artists' songs on the Billboard charts.