SDS 315 HW 2

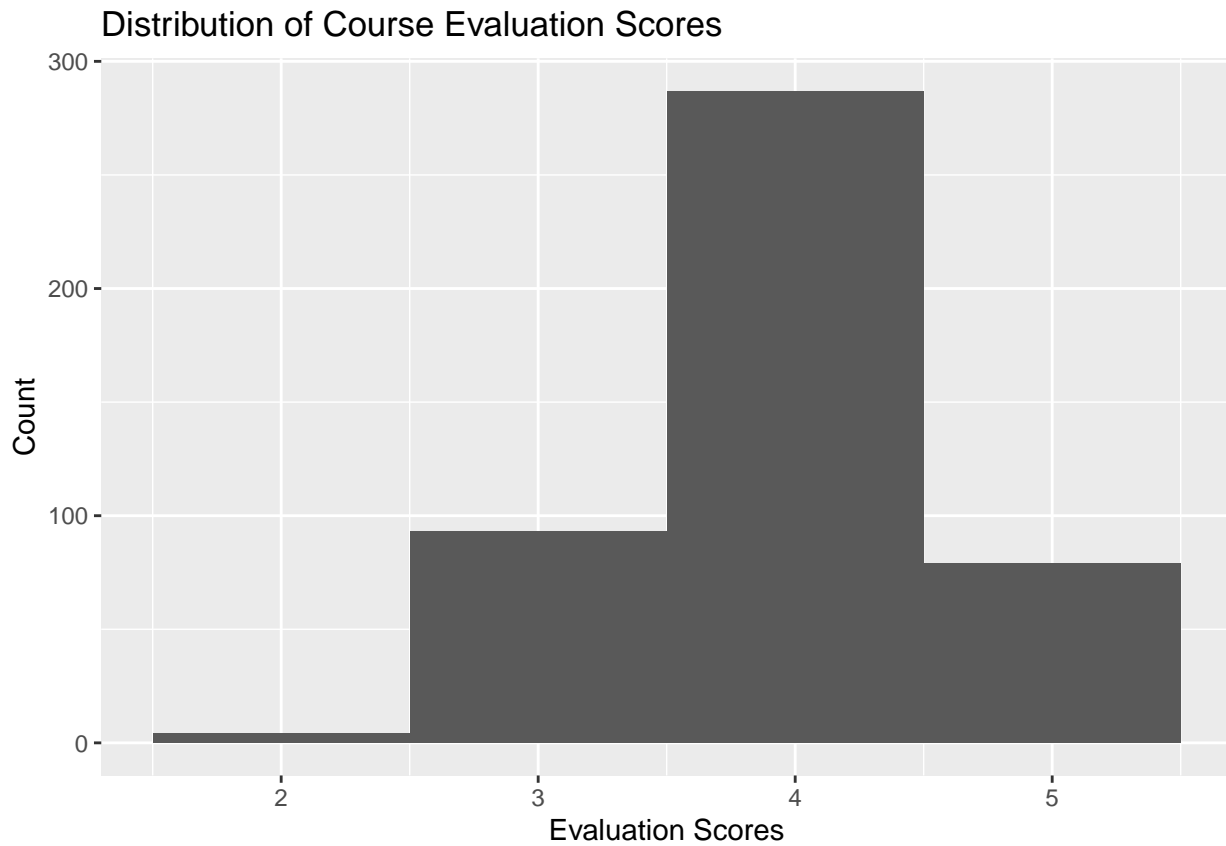https://github.com/sophiayang5/SDS-315-HW-2

Sophia Yang (sy23928)

2025-01-25

Contents

Problem 1: Beauty, or Not, in The Classroom

Part A.

Distribution of Course Evaluation Scores

The above histogram displays the distribution of course evaluation scores given to UT professors. The distribution seems to be relatively normal, with a bell shape. Thus, the mean (3.998) and median (4.000) are also close to each other. The range is 3, with a minimum of 2 and a maximum of 5.
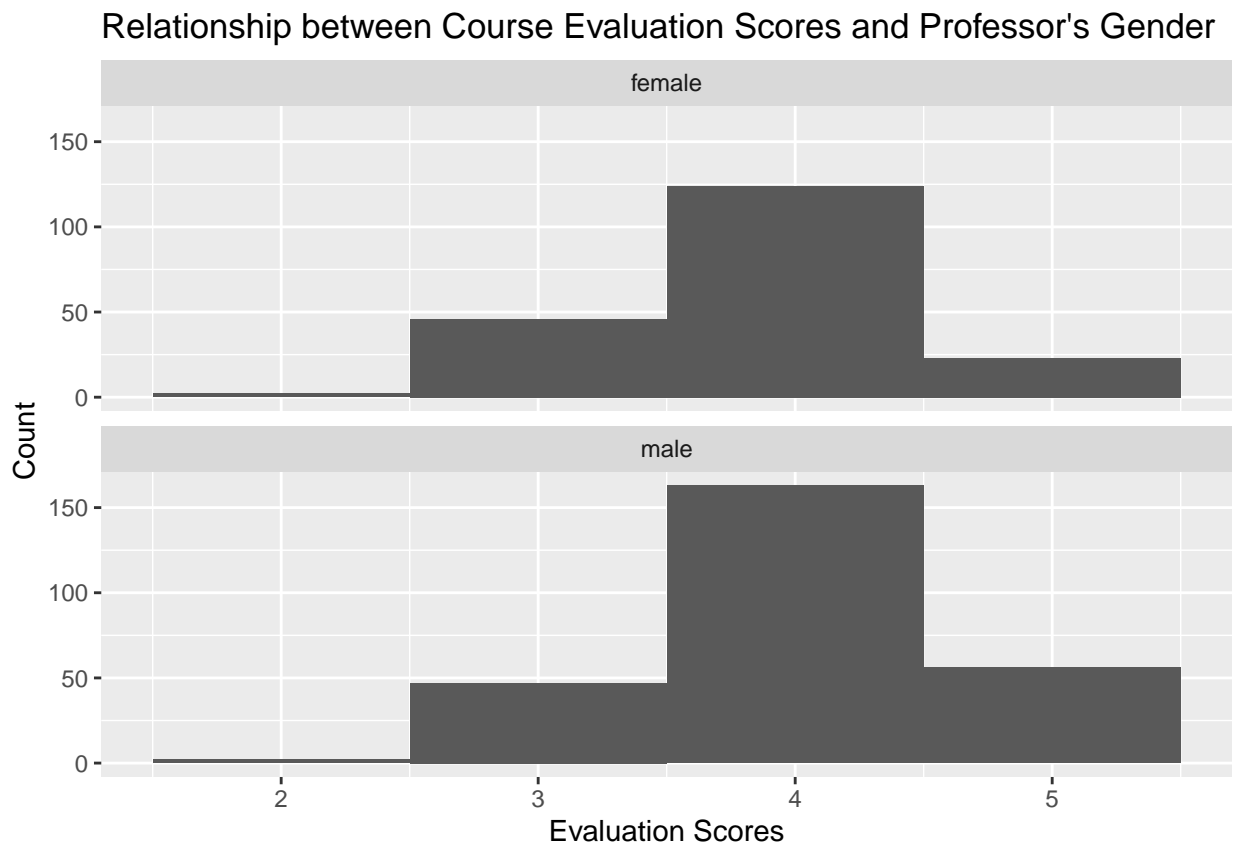
Part B.

Relationship between Course Evaluation Scores and Professor's English
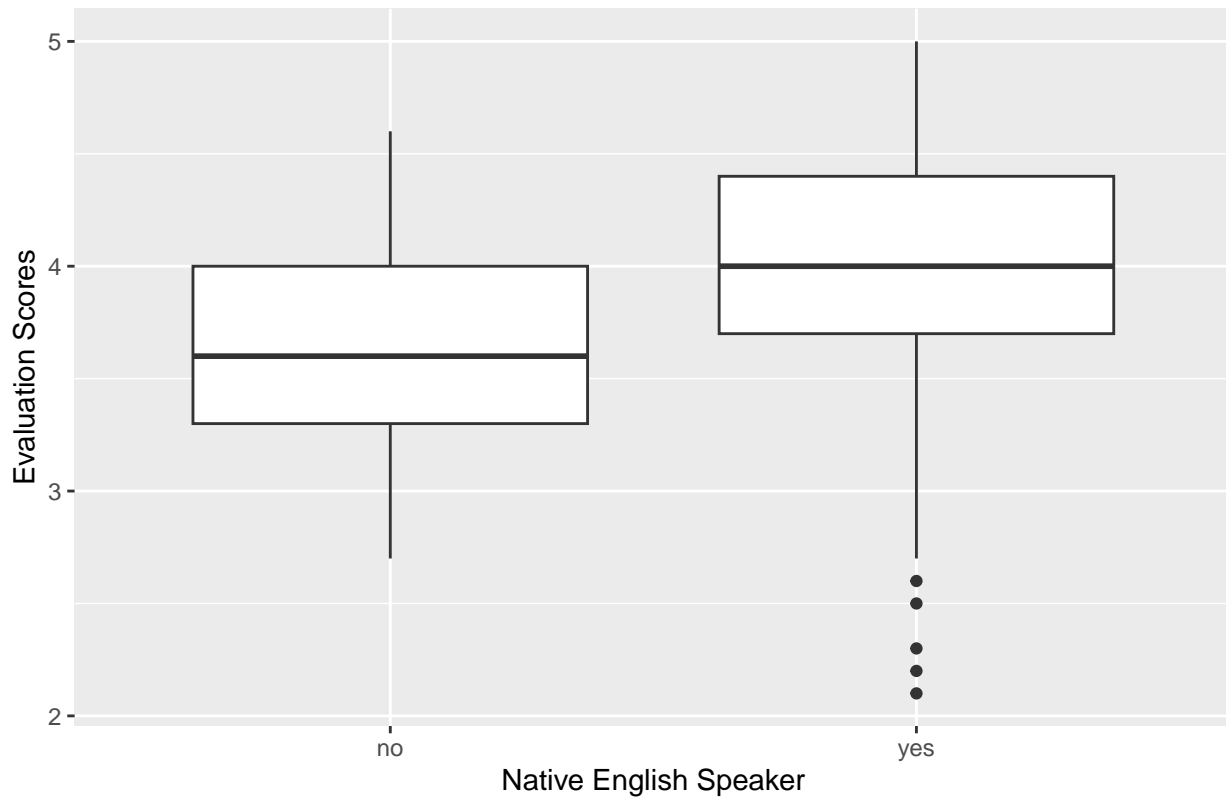
The above boxplot shows the relationship between a professor's course evaluation scores and whether or not English is their native language. The median for professors whose English is their native language (around 4) is higher than the median for professors whose English is not their native language (around 3.6). However, the range for professors who are native speakers (around 3) is higher than the range for professors who are not (around 1.9). Additionally, the distribution for professors who are native speakers seems skewed to the left.
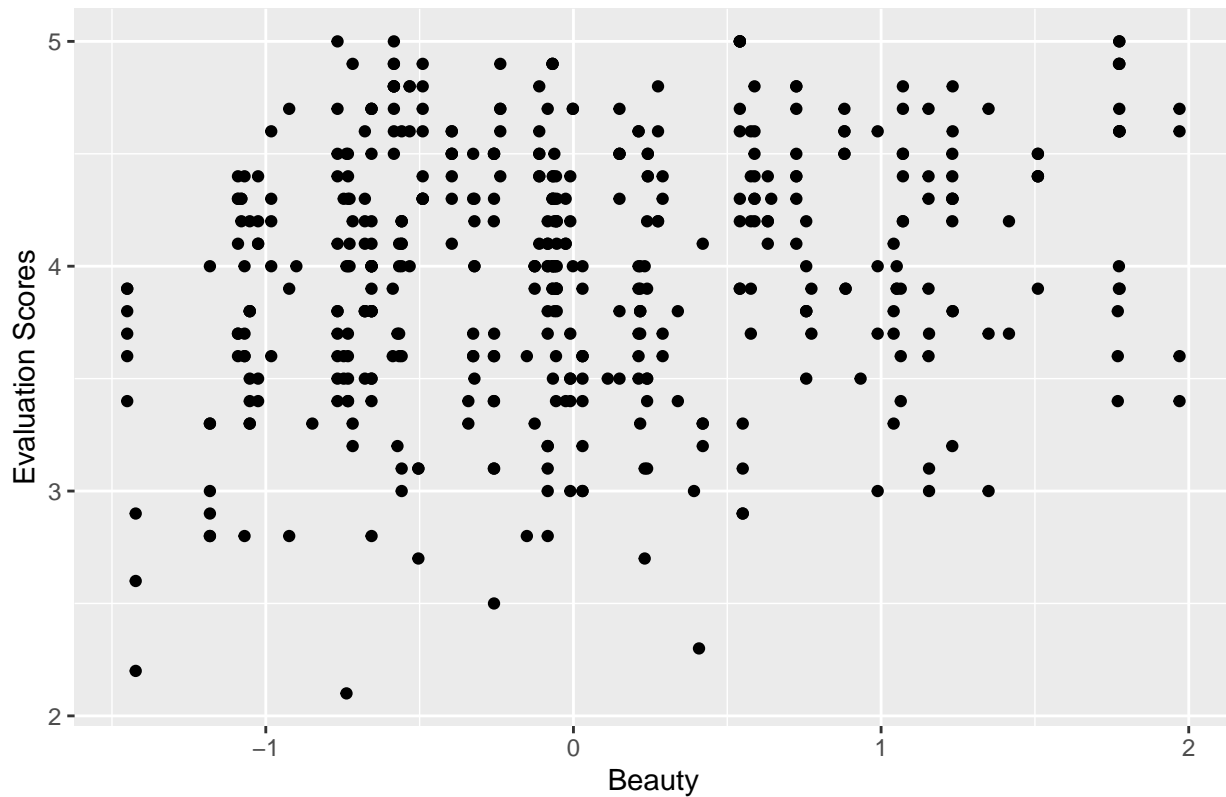
Part C.

Relationship between Course Evaluation Scores and Professor's Gender

The above histogram shows the relationship between a professor's course evaluation scores and their gender (male or female). The graphs for both male and female professors are unimodal, both having only one peak at 4. Additionally, the ranges for both male and female professors are 3, with a minimum of 2 and a maximum of 5. There does not seem to be too big of a difference between the two genders: the median for male professors is similar to the median for female professors.

Part D.

Relationship between Course Evaluation Scores and Beauty

The above scatterplot shows the relationship between a professor's course evaluation scores and their physical attractiveness (rated by 6 panelists). The beauty variable ranges from around -1.4 to around 2. There could be a positive correlation between the two, but overall, the distribution of points seem to be random: there is no correlation between a professor's attractiveness and their evaluation scores. The medians for each beauty rating are roughly the same.
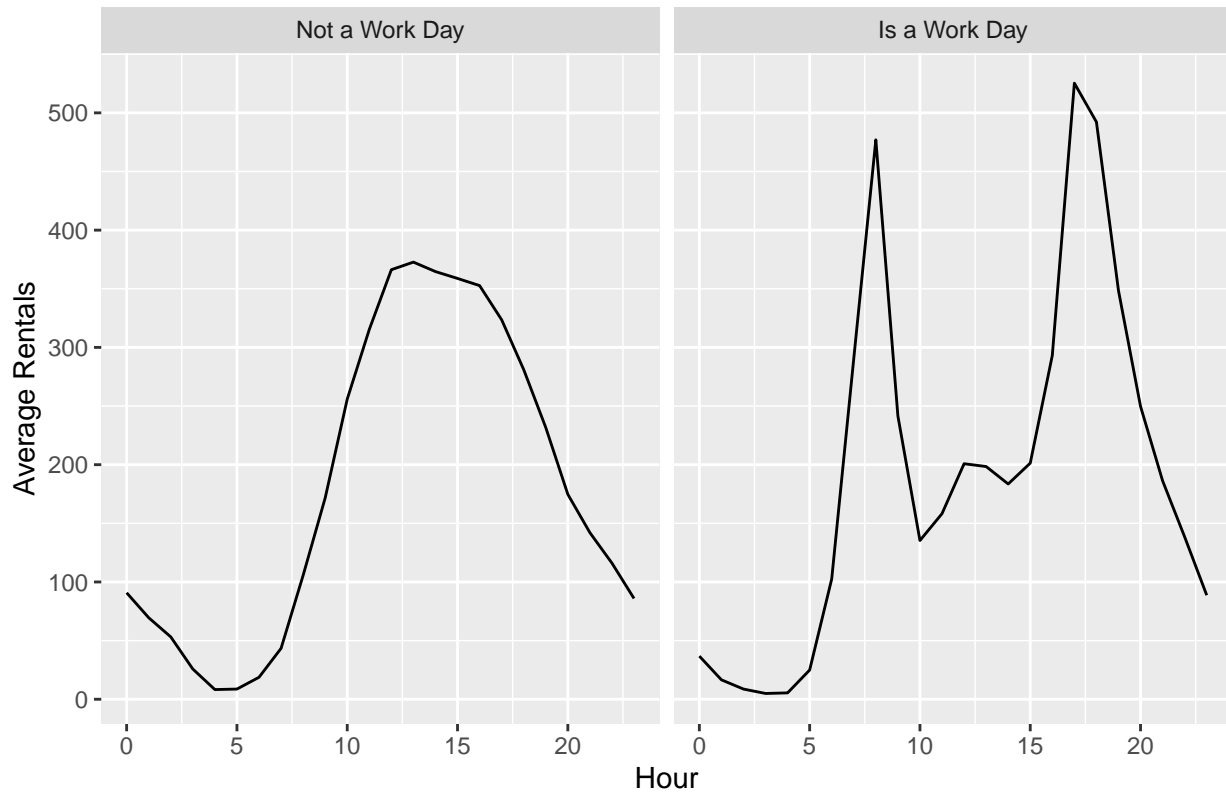
Problem 2: Bike Sharing

Plot A.

Relationship between Average Bike Rentals and Hour
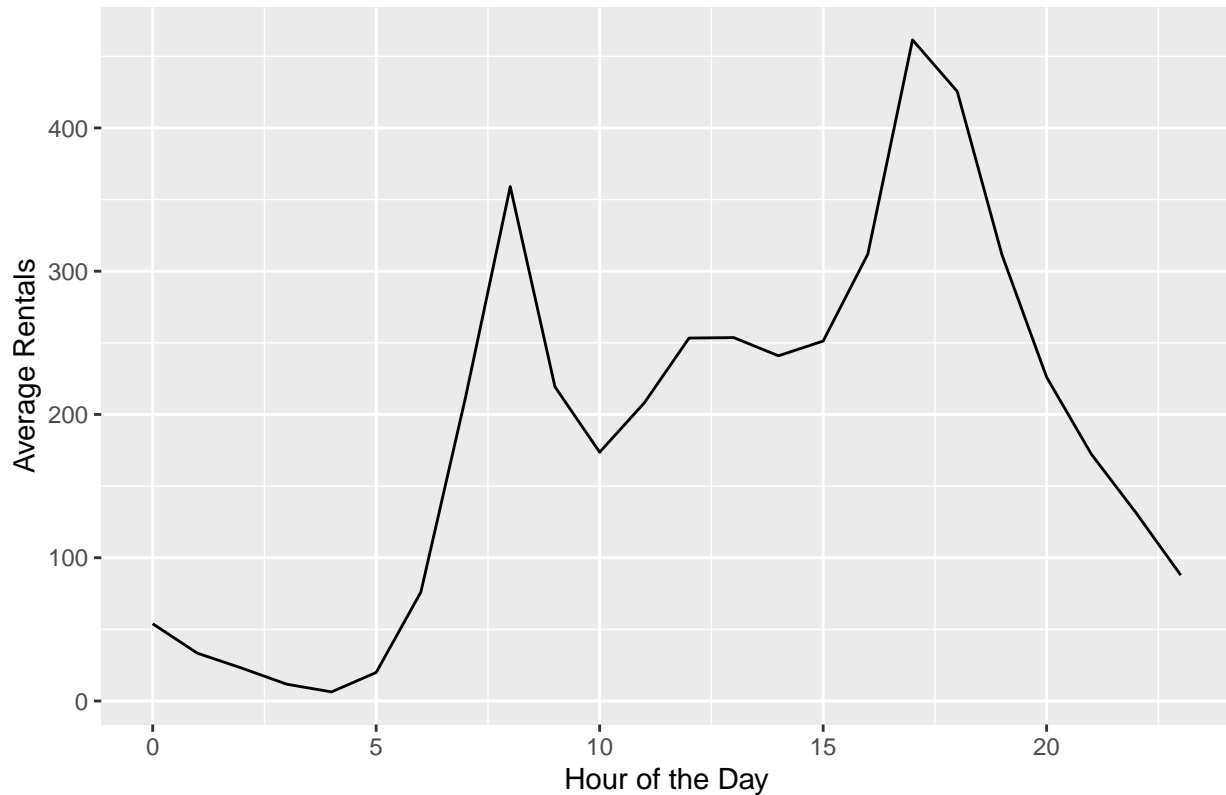
The above line graph shows the relationship between hour of the day (0-23) and average bike rentals. The average number of bike rentals seem to be lowest in the early morning (before 5 A.M.), and highest in the early evening (from 5-7 P.M.). There are two big peaks at around 8-9 A.M. and 5 P.M. (the times when people are going to school or work and going home). Then, after 8 P.M., the average number of bike rentals steadily decreases again. Overall, ridership tends to increase and peak during times when most people are leaving for work or home, and decrease and remain low during early and late hours.
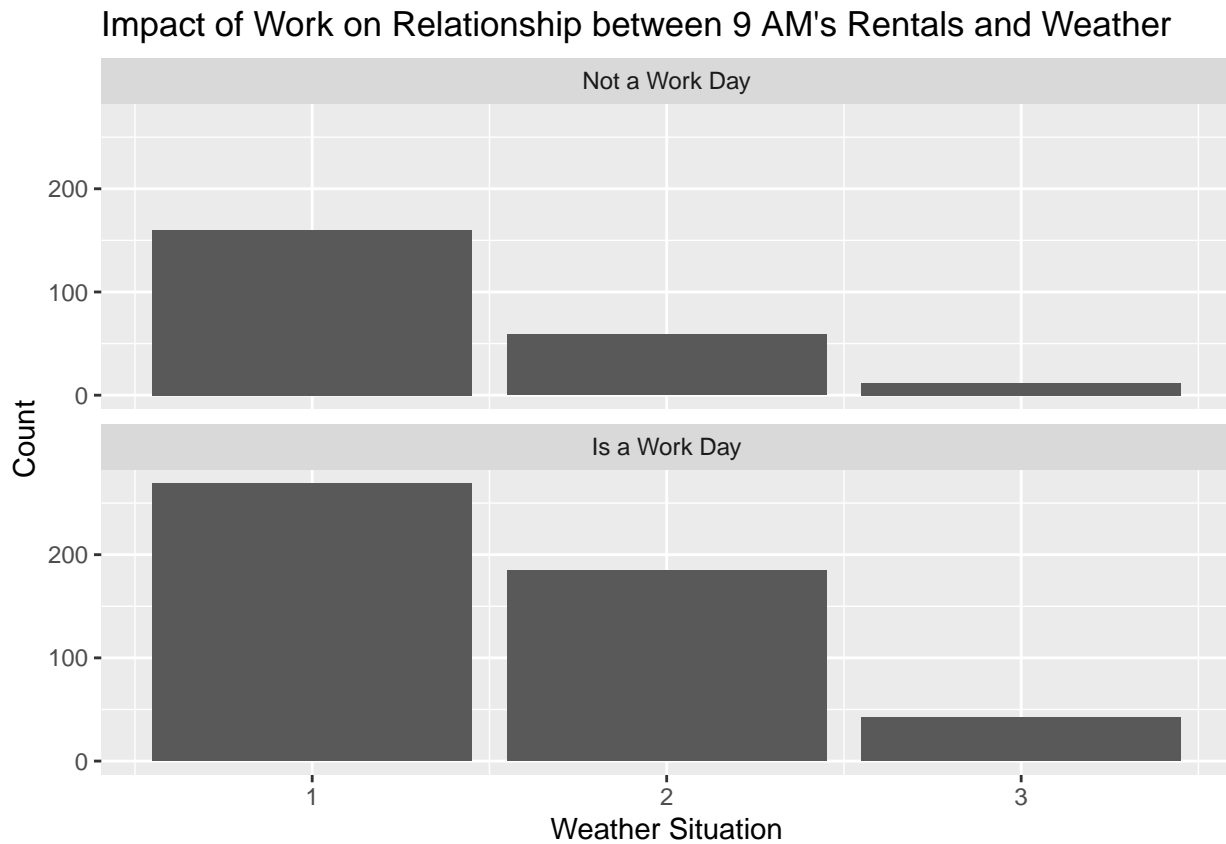
Plot B.

Impact of Work on the Relationship between Bike Rentals and Hour

The above faceted line graph shows the impact of whether or not it's a working day on the relationship between average bike rentals and hour of the day (0-23). When it's not a work day, the distribution is shaped like a rounded hill; in contrast, when it's a work day, the distribution has 2 sharp peaks. Additionally, during non-work days, the average rentals only reaches a maximum of around 375, while during work days, there is a greater maximum of around 525. Overall, there is greater variability in average bike rentals during work days than during non-work days.

Plot C.

Impact of Work on Relationship between 9 AM's Rentals and Weather
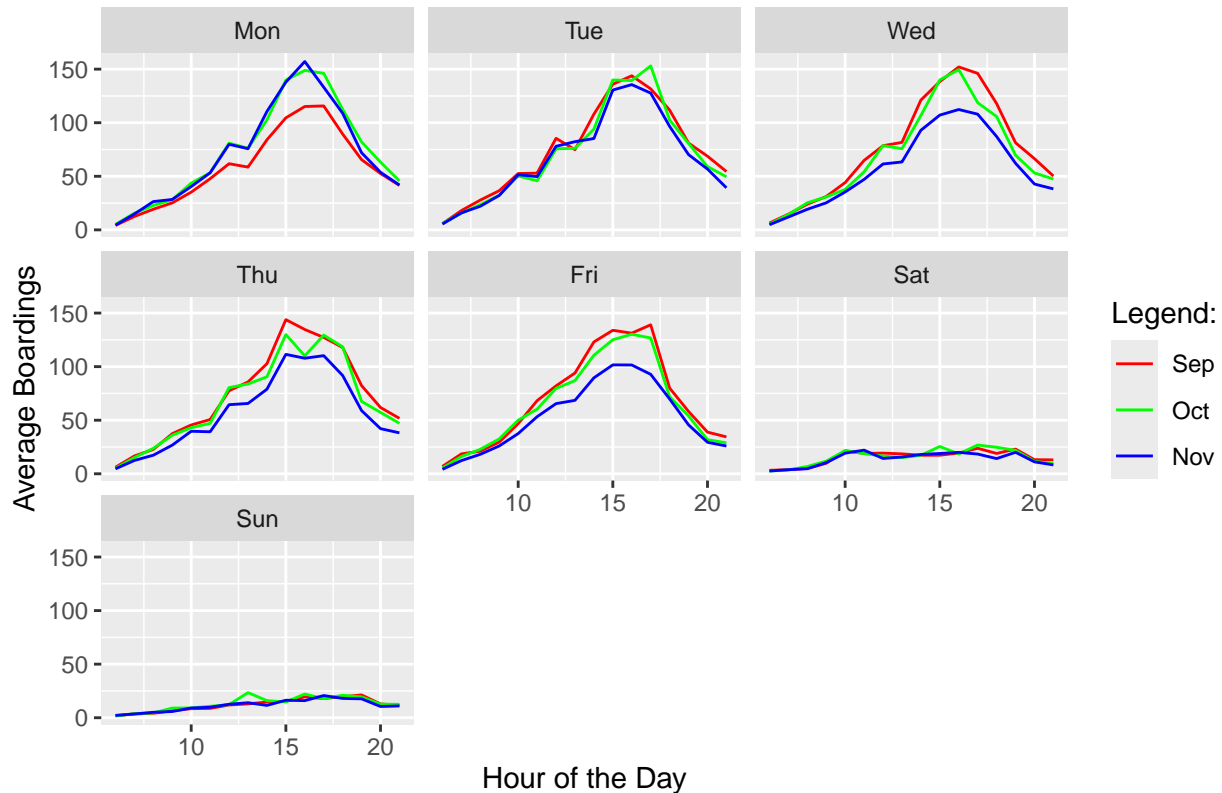

The above faceted bar graphs show the impact of whether or not it's a working day on the relationship between average bike rentals (during 9 A.M.) and the weather situation (1-4 with 1 being clear skies and 4 being stormy). Both work days and non-work days have a distribution that is skewed to the right. However, the range for non-work days (around 150 rentals) is lesser than the range for work days (around 220 rentals). Additionally, during less ideal weather conditions (3), the number of average rentals for non-work days (around 10) is lesser than the minimum number of rentals for work days (around 40). Overall, 9 A.M. work days have greater average rentals across all weather conditions than 9 A.M. non-work days.

Problem 3: Capital Metro UT Ridership

Task 1.

Impact of Month on the Relationship between Boardings and Hour

The above faceted line graph shows the impact of the month on the relationship between the average number of the UT area's bus boardings and the hour of the day. The red lines show September boardings, the green lines show October boardings, and the blue lines show November boardings. The hour of peak boardings changes from day to day: during Saturdays and Sundays, peak boardings shift to later hours, as shown in the graph. I think average boardings on Mondays in September look lower compared to other months; because September is quite early into the school year, UT students are still adjusting to their change in environment, and do not feel as comfortable or safe riding unfamiliar bus lines. I think average boardings on Wednesdays, Thursdays, and Fridays in November look lower; because Thanksgiving break is in November, there are less students on the UT campus and therefore less people boarding the buses.

Task 2.

Impact of Hour on the Relationship between Boarding and Temperature
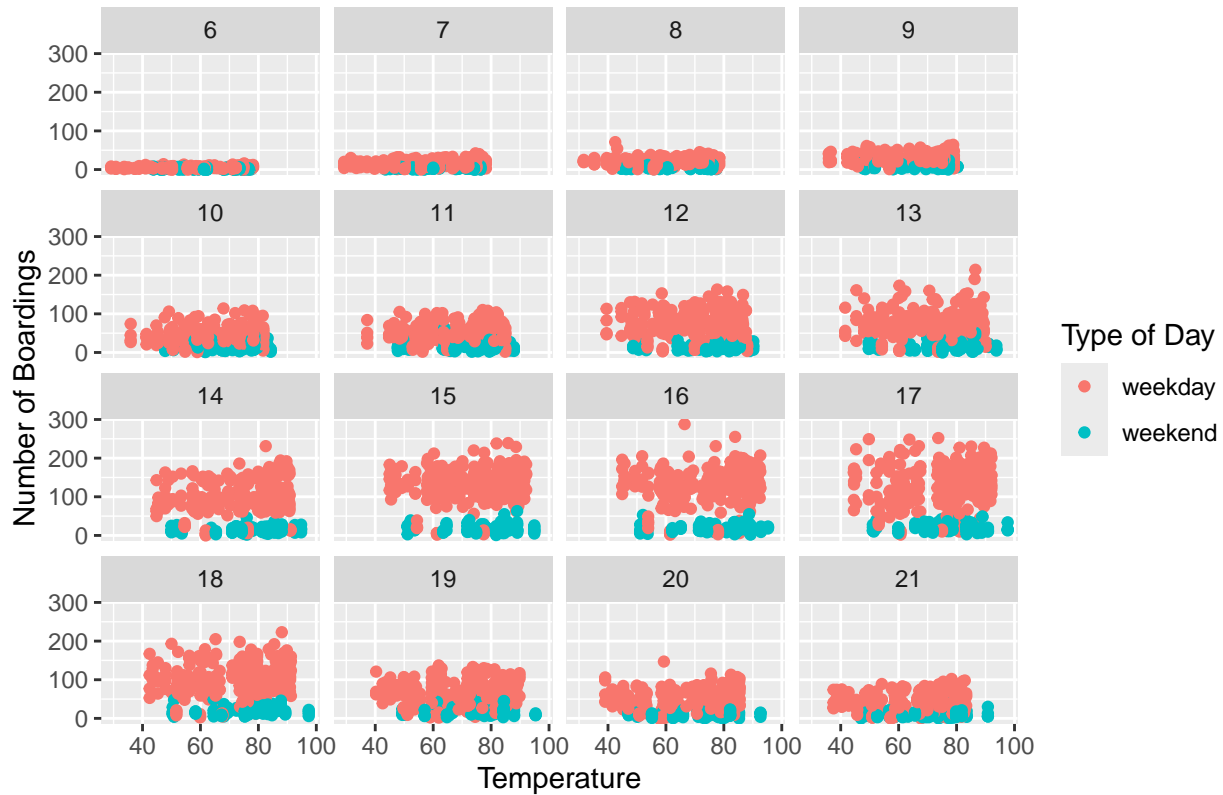
The above faceted scatterplot shows the impact of the hour of the day (0-23), with days color-coded by weekdays (red) and weekends (blue), on the relationship between the number of boardings and the temperature. When we hold the hour and type of day constant, temperature does not seem to have a noticeable effect on the number of UT students riding the bus: there are not any noticeably less numbers of people boarding the bus during lower temperatures than higher temperatures, and the distribution of the points does not have any obvious correlation, seeming to be random across different temperatures.

Problem 4: Wrangling the Billboard Top 100

Part A.

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 above table shows the top 10 songs with the most number of weeks being on the Billboard Top 100. Imagine Dragons' "Radioactive" has the most songs (87).

Part B.

Relationship between Musical Diversity and Year
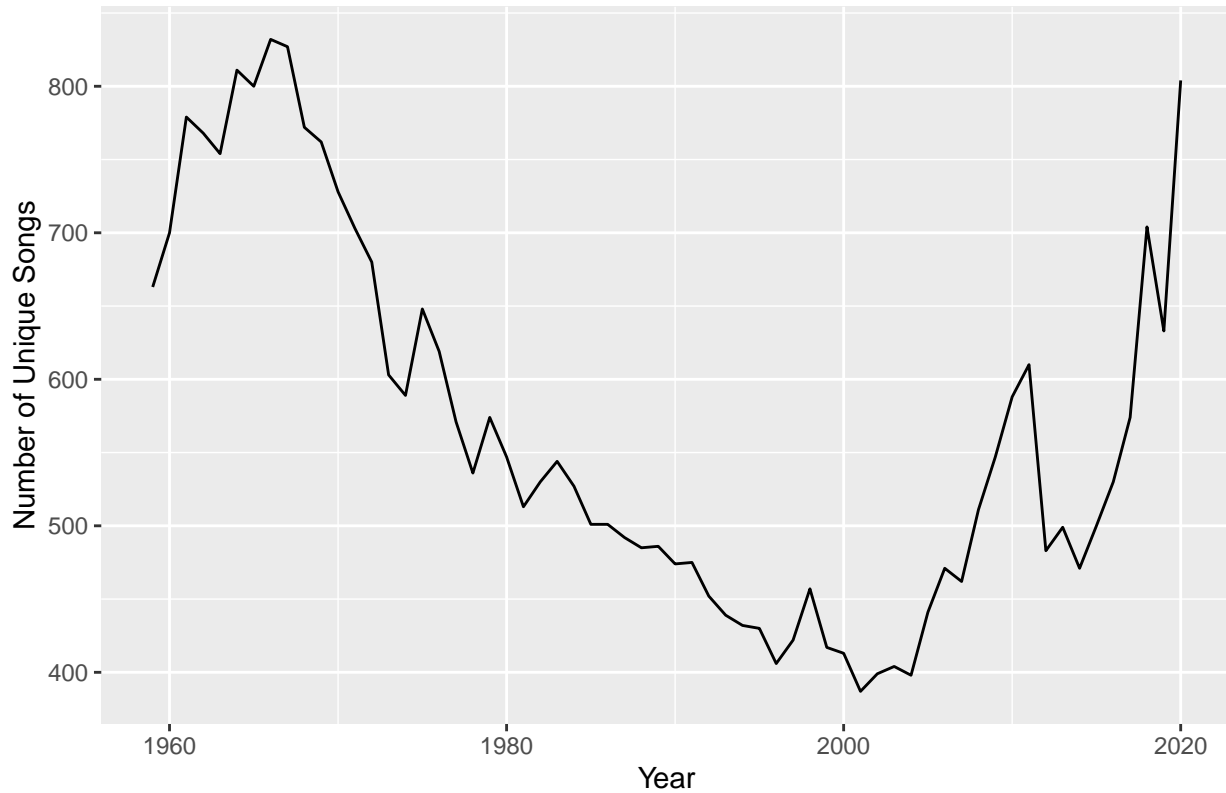
The above line graph shows the relationship between musical diversity (determined by the number of unique songs on the Billboard Top 100 in a year) and year (1958-2021). There is a peak around 1965, a gradual decrease until 2000, and then a rapid increase onward. The peak around 1965 is interesting, and is perhaps due to The Beatles as well as other rock 'n' roll bands gaining popularity among the counterculture movement. Additionally, the upward trend from 2000 onward is also interesting, and might be caused by the rapid growth of the Internet and the creation of Spotify.

Part C.

Distribution of Performers with 30+ Ten−Week Hits
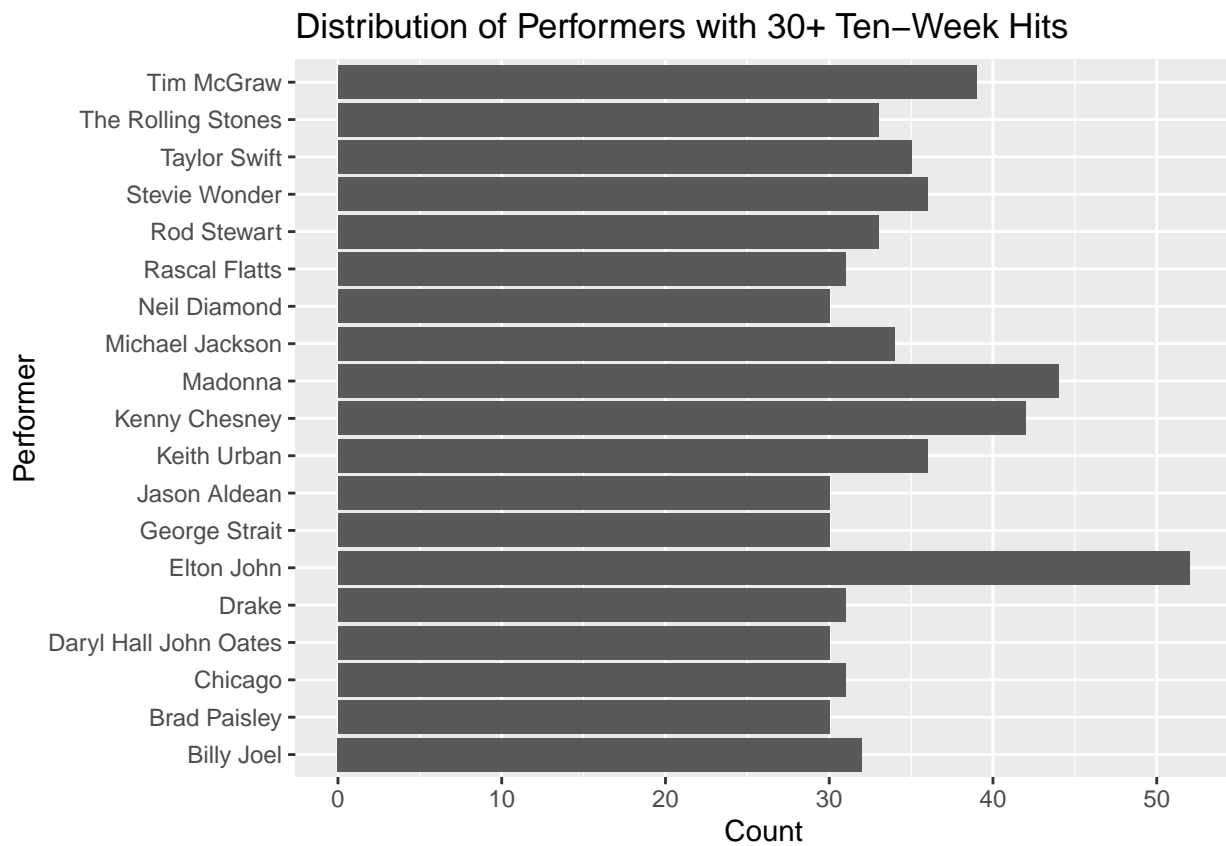
The above bar graph shows the 19 artists in U.S. musical history (1958-2021) who've had at least 30 songs that were 10-week hits (a single song that was on the Billboard Top 100 for at least 10 weeks). Elton John has had the most 10-week hits, while Neil Diamond has had the least.