Homework 2

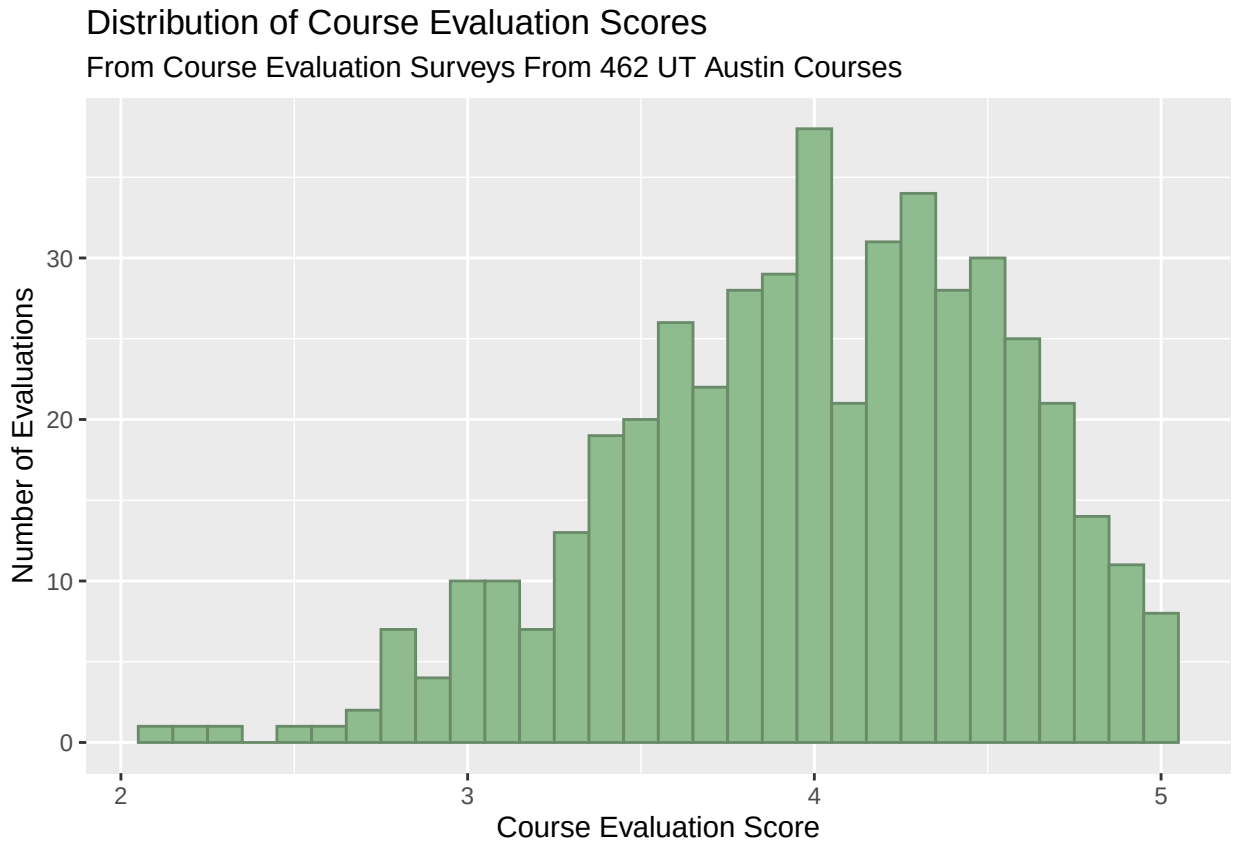
Najia Sarker (Student ID: nms3364)

2025-01-25

Link to GitHub: https://github.com/nsarker2/Homework2

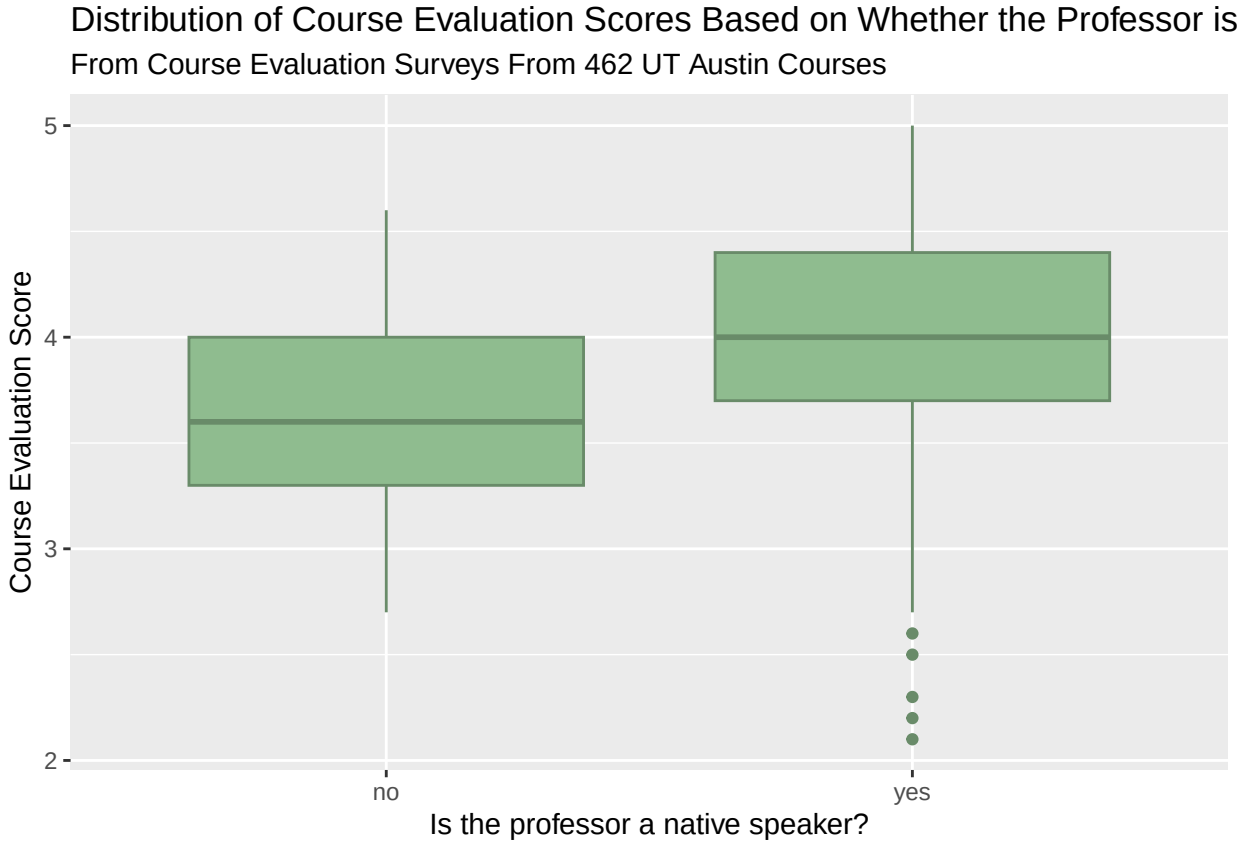
Problem 1: Beauty, or not, in the classroom

Part A



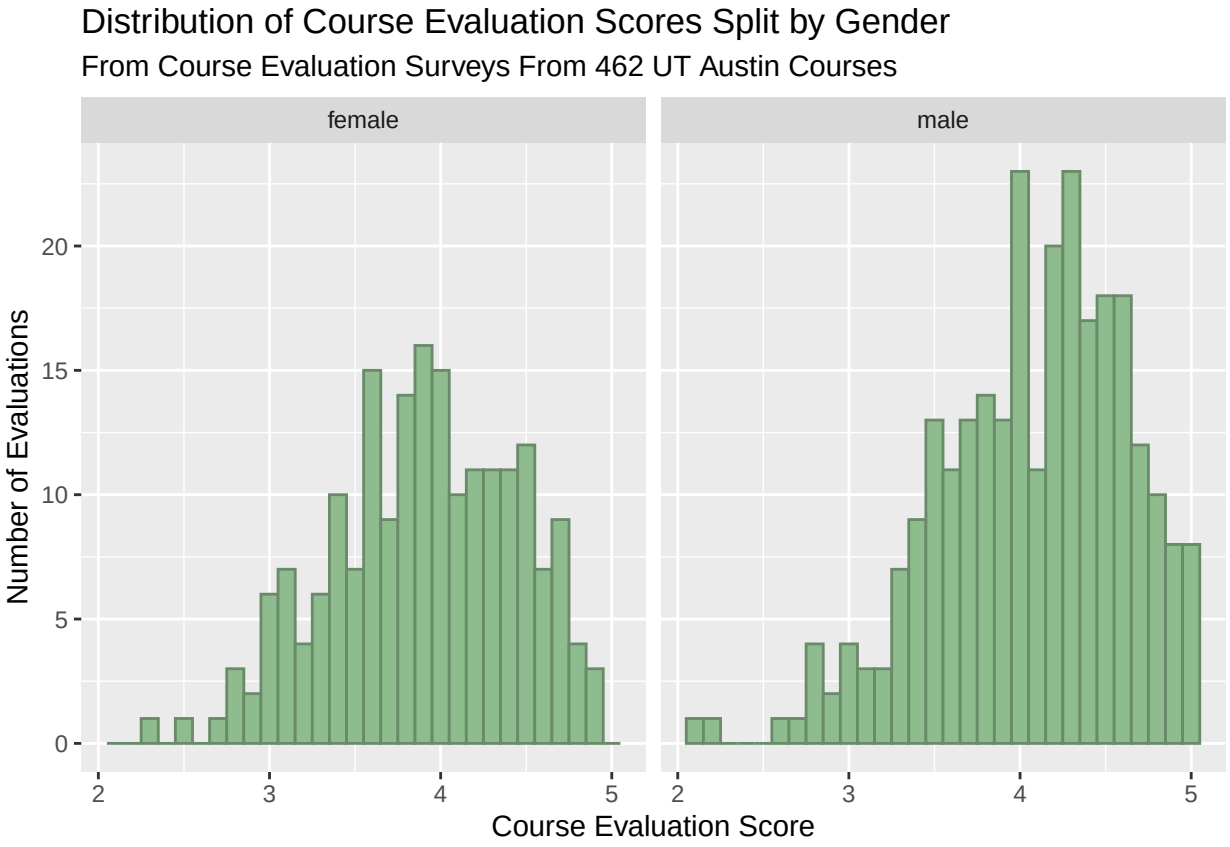
This graph illustrates the distribution of course evaluation scores. The median score is 4, indicating that most professors are rated quite positively. The spread is a minimum of 2.1 to a maximum of 5, which results in a IQR of 2.9. The distribution is approximately normal, though there is a slight leftward skew, suggesting a small number of lower-rated evaluations.

Part B



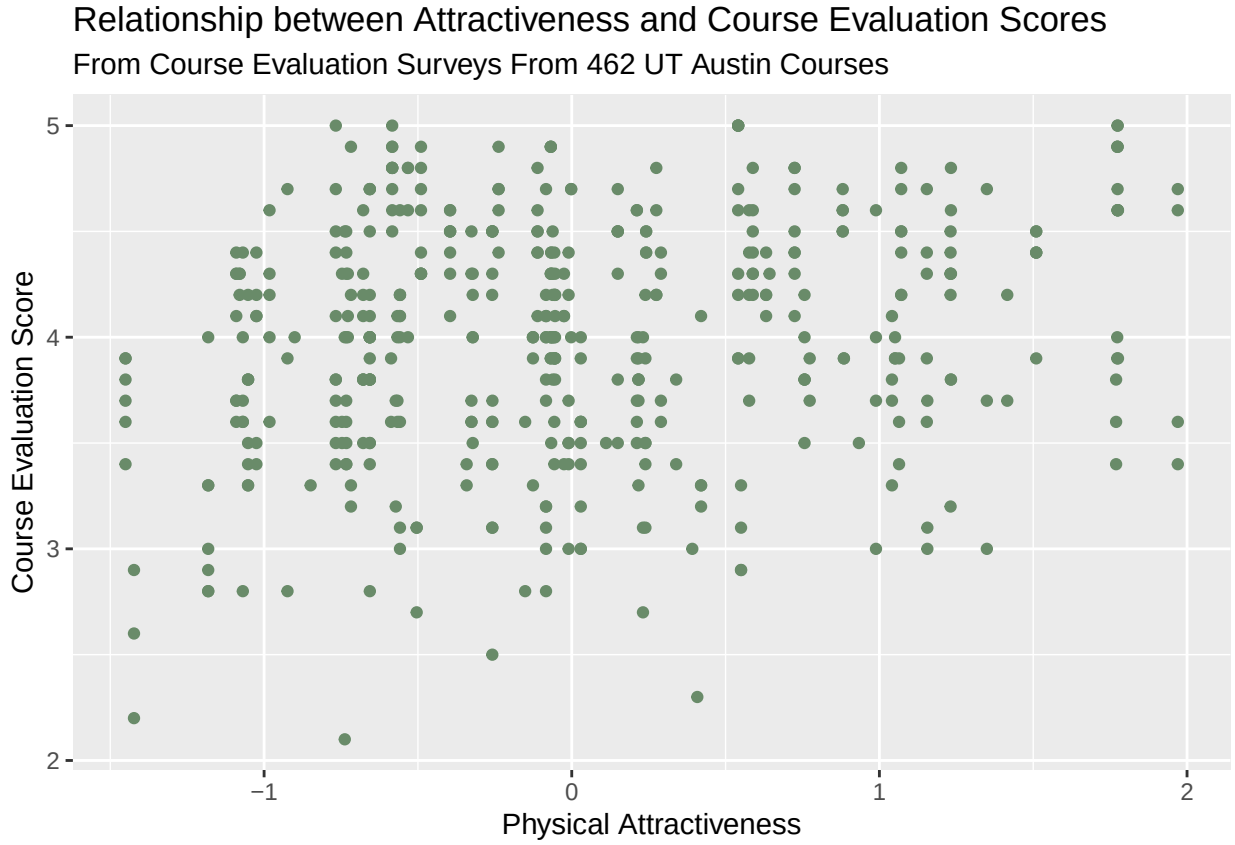
This graph illustrates the distribution of course evaluation scores based on whether the professor is a native English speaker or not. For the non-native English speakers, the median is 3.6, while the median for the native English speakers is 4, which shows that that the native English professors have a higher course evaluation rating. The IQR for both distributions are 0.7, with a few outliers for the native English speaker boxplot.

Part C



This graph illustrates the distribution of course evaluation scores based on whether the professor's gender. For female professors, the median was 3.90, while the male professors had a median of 4.15, suggesting that male professors are rated higher than female professors. The male histogram had a higher IQR of 0.8, while the female histogram had an IQR of 0.7, which explains why the male histogram has a bit more of a skew on the left than the female one.

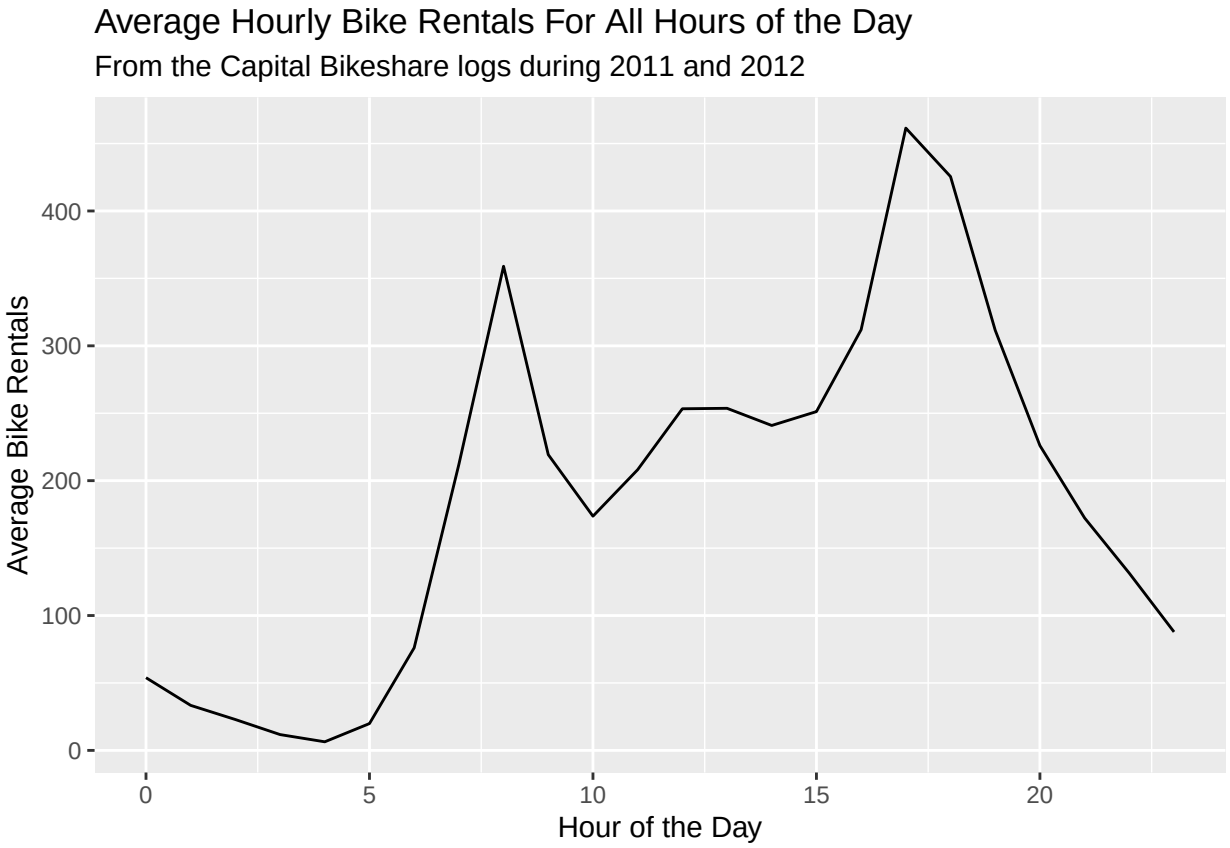
Part D



This distribution illustrates the relationship between physical attractiveness and course evaluation scores. The correlation between attractiveness and evaluation scores is 0.189, indicating a weak, positive, linear relationship. This suggests that while there is some association between physical attractiveness and evaluation scores, physical attractiveness is not a strong predictor of course evaluations.

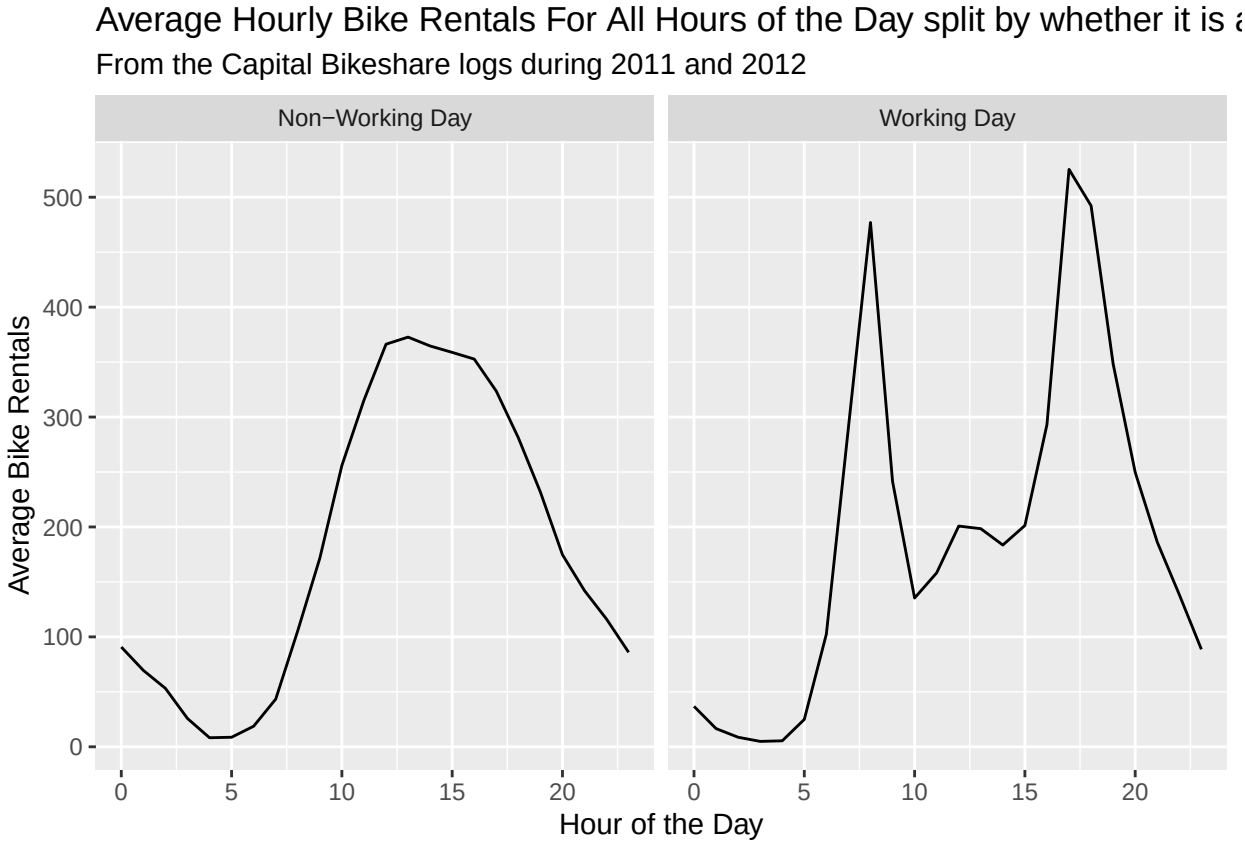
Problem 2: bike sharing

Part A



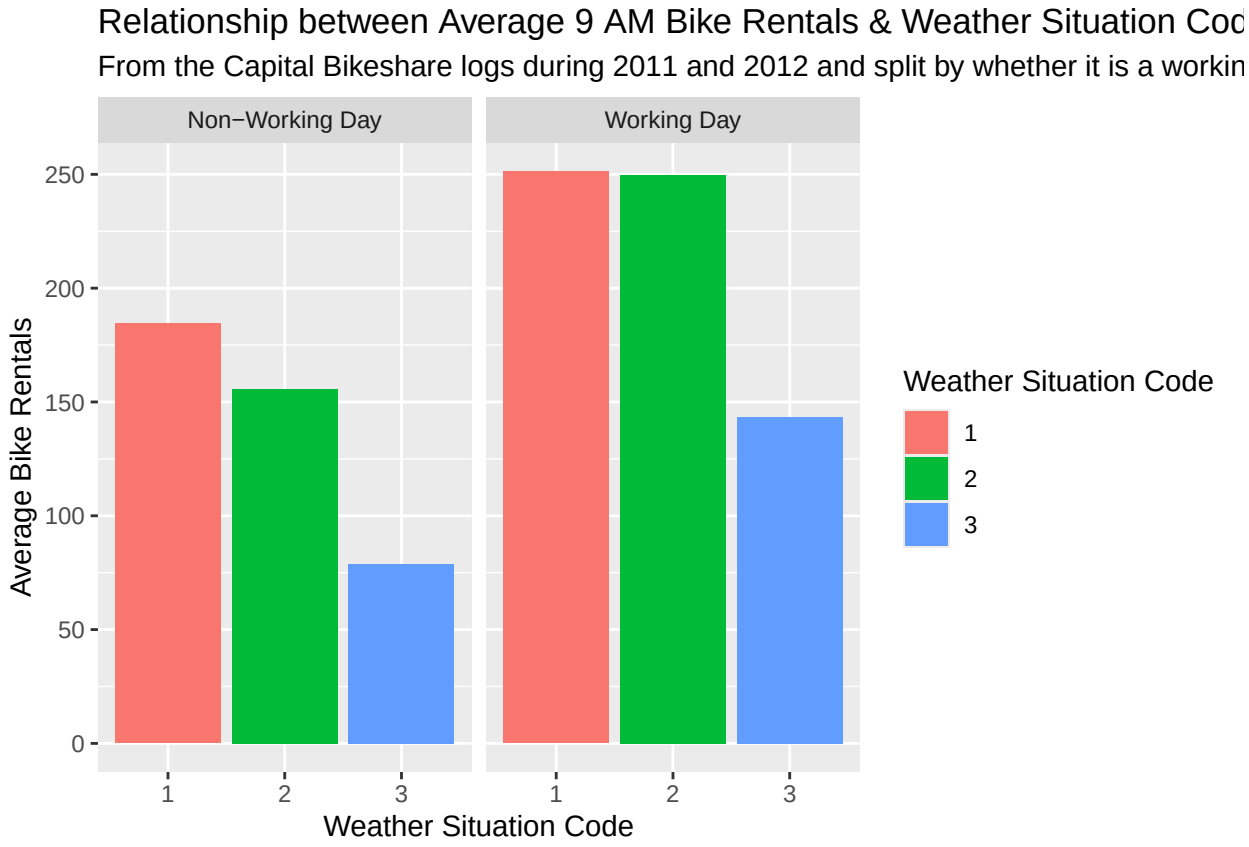
This graph illustrates the relationship between the hour of the day and average bike rentals. The correlation between these variables is 0.56, indicating a moderate, positive linear relationship. This suggests that bike rentals tend to increase during certain hours of the day, such as around 8 AM and around 4 PM, reflecting a noticeable pattern in rental activity.

Part B



These graphs illustrate the relationship between hour of the day and average bike rentals, faceted by whether it is a working day or not. The correlation for non-working days is 0.49, while for working days, it is 0.50, suggesting a moderate, positive, linear relationship in both cases. Despite the similar correlations, the patterns in the graphs differ significantly. On non-working days, there is a single spike in rentals between 12 PM and 4 PM, reflecting mid-day recreational activity. On working days, there are two distinct spikes: one around 8 AM, corresponding to morning commutes, and another around 4 PM, reflecting afternoon commutes.

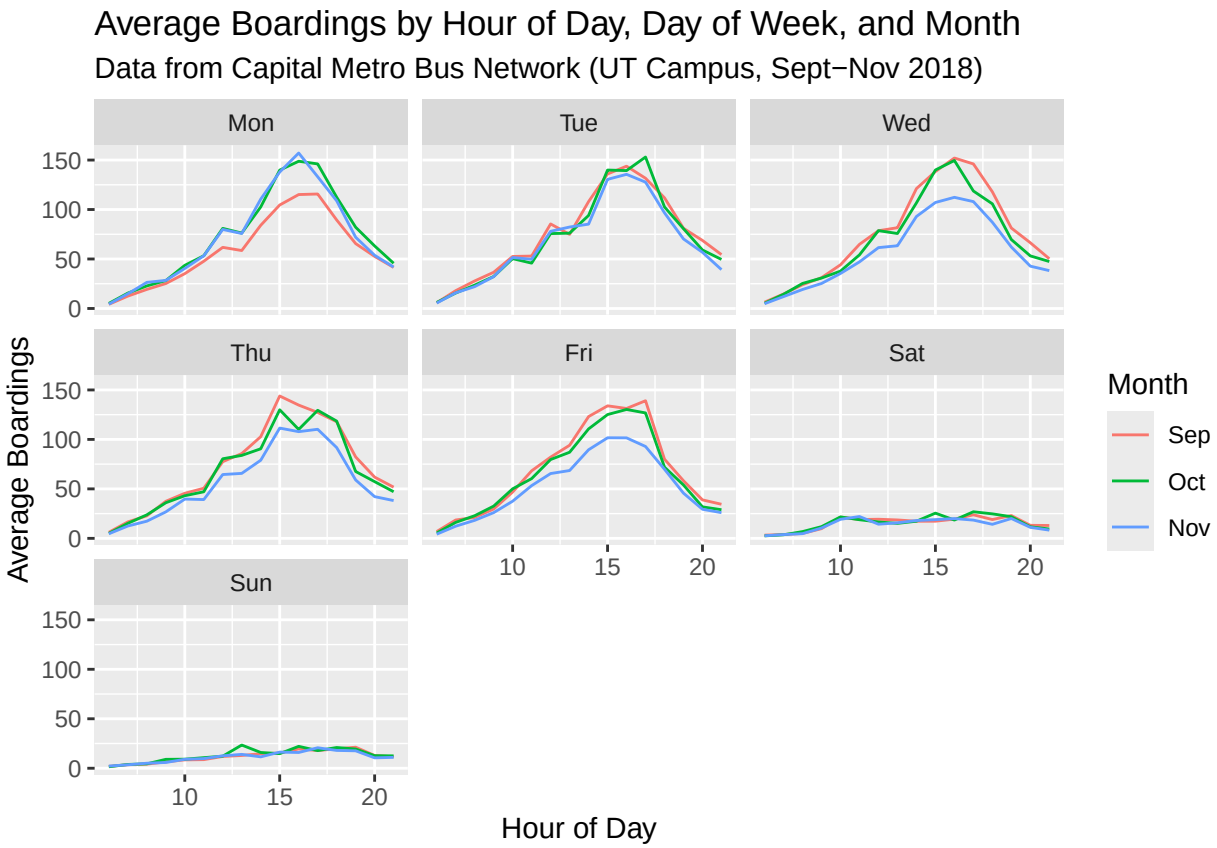
Part C



These graphs illustrate the relationship between the weather situation code and the average bike rentals at 9 AM. On the non-working days, there were the most rentals when the weather situation code was a 1 at around 185, which means the the weather was clear, and there were the least rentals when the weather situation was a 3 at around 75, which means that there could have been light snow, light rain, a thunderstorm, and other light weather. For the working days, there were almost equal rentals when the weather situation code was a 1 and 2 at around 250, 2 meaning that it was either cloudy or misty. Weather situation code 3 still had the least amount of rentals, but almost double the amount of the non-working day rentals at around 140. Overall, these results suggest that most people rent bike at 9 AM when the weather is better, but there are more rentals on working days no matter the weather.

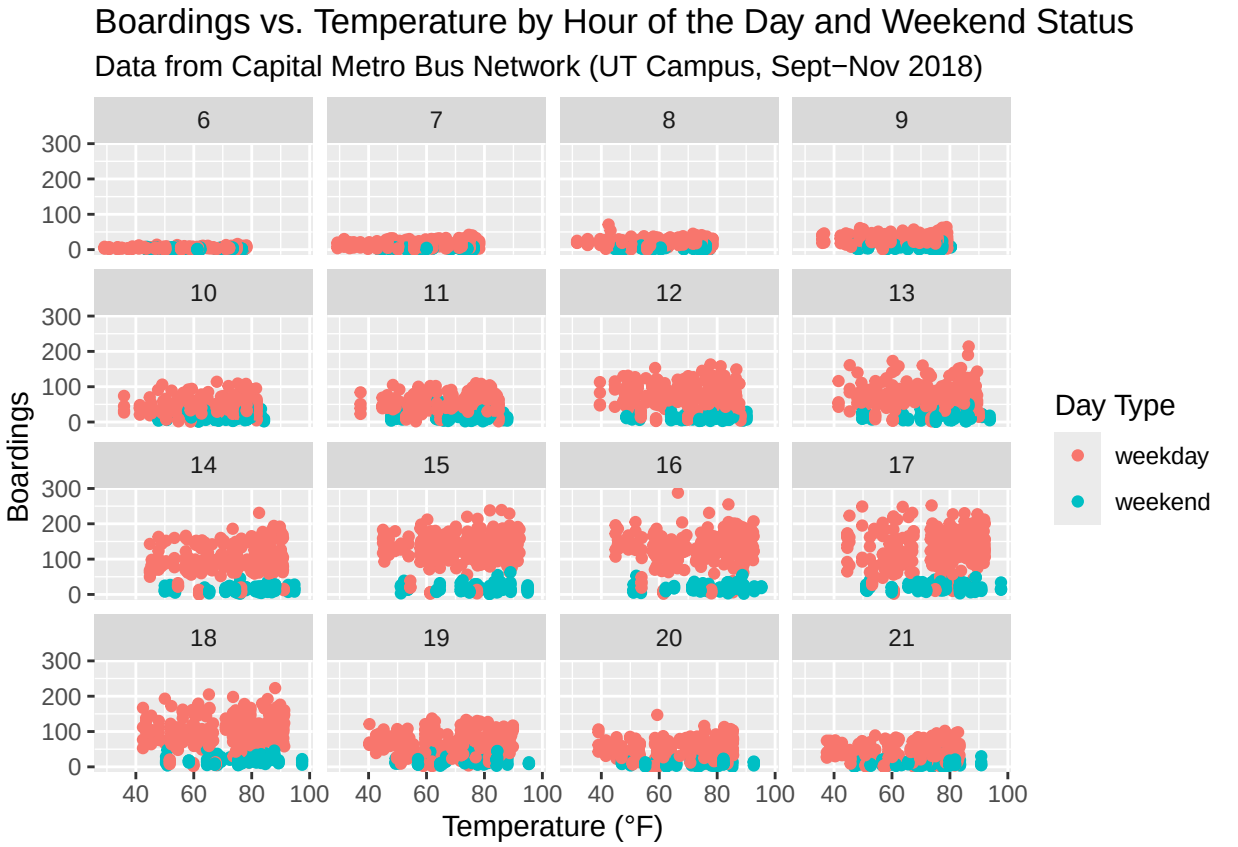
Problem 3: Capital Metro UT Ridership

Part 1



These graphs illustrate the average number of boardings by hour of the day across different days of the week (faceted) and are colored by month (September, October, and November) for the Capital Metro Bus Network. The graph highlights distinct boarding patterns for each month and day, with the hour of peak boardings showing clear differences between weekdays and weekends. On weekdays, the highest boarding rates are concentrated around 3-4 PM, with weekends showing a consistent horizontal line throughout the day. On Mondays in September, average boardings are noticeably lower, which could be due to the start of the academic year or fewer early-morning classes. In November, there is a clear dip in average boardings on Wednesday, Thursday, and Friday afternoons, which could be attributed to several factors, including the approaching holiday season or fewer campus events taking place toward the end of the semester.

Part 2



This figure illustrates the average number of UT students riding the bus at different hours of the day, differentiated by weekend status (weekdays vs. weekends), with the temperature shown as a continuous variable. Evidence from the figure shows that, even with temperature variation, the ridership levels remain consistent across different temperature ranges, suggesting that factors like commuting or classes may play a more prominent role in influencing ridership than temperature.

Problem 4: Wrangling the Billboard Top 100

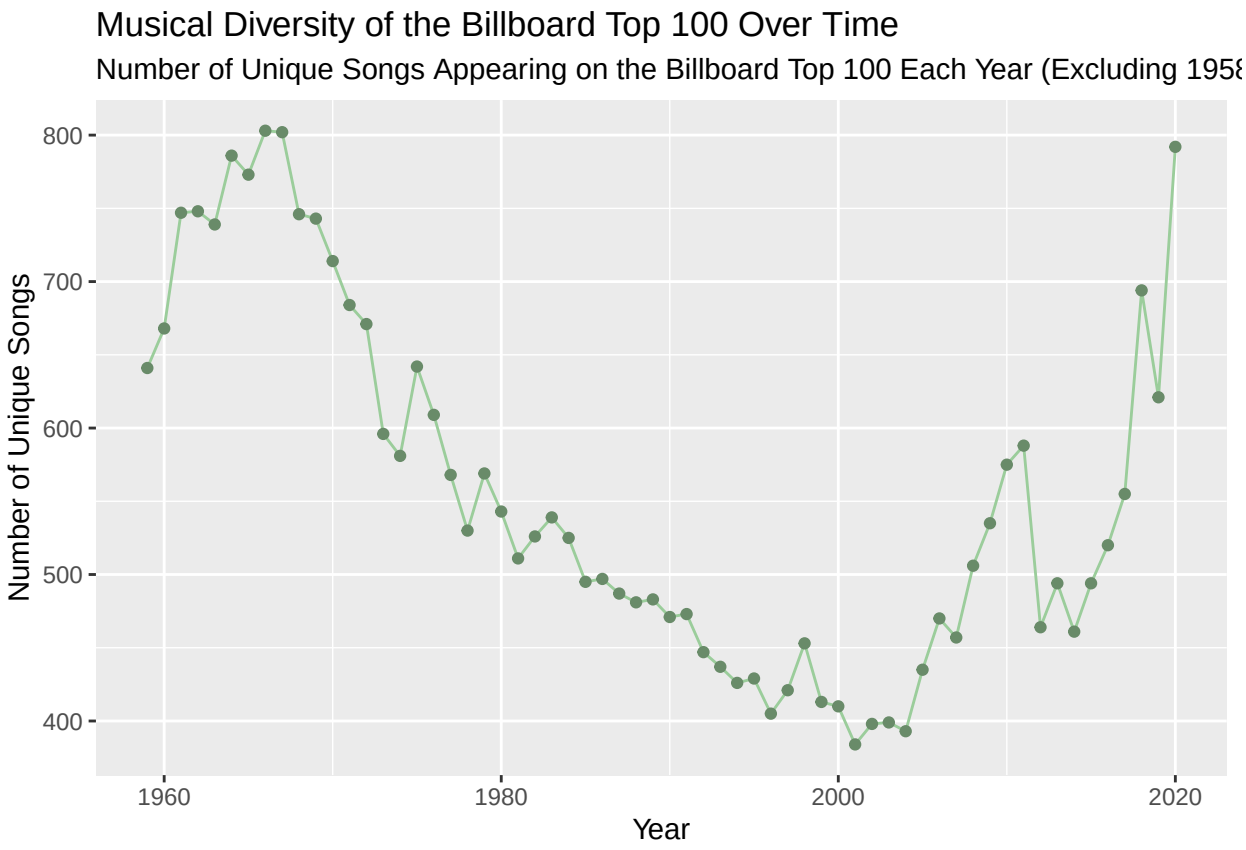
Part A

Table 1: Top 10 Songs by Weeks on Billboard Chart

performer	song	count
3 Doors Down	Kryptonite	53
AWOLNATION	Sail	53
Adele	Rolling In The Deep	53
LMFAO Featuring Lauren Bennett & GoonRock	Party Rock Anthem	53
Lifehouse	You And Me	53
Avicii	Wake Me Up!	52
Bastille	Pompeii	52
Carrie Underwood	Before He Cheats	52
Chris Brown & Young Thug	Go Crazy	52
Dua Lipa	Don't Start Now	52

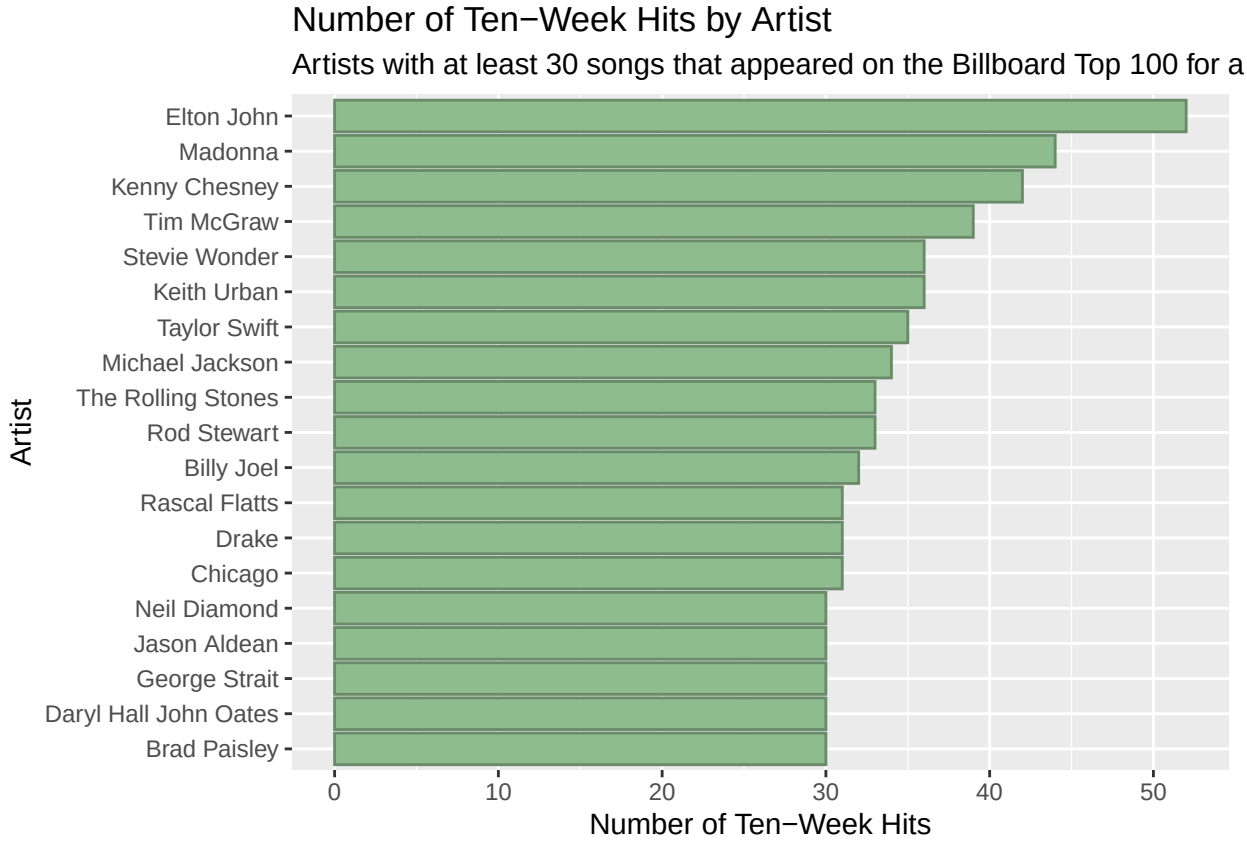
In this table, the top 10 songs based on how many songs have been on the Billboard 100 for more than 10 weeks from 1953-2021 are displayed. The songs are ranked by the total number of weeks each song spent on the chart, with the highest-ranked song appearing first. The table includes the performer and the song title.

Part B



This graph shows how the number of unique songs on the Billboard Top 100 has evolved over time. Excluding the years 1958 and 2021, the chart highlights trends in musical diversity and whether it has increased or decreased over the years. There are spikes in the number of unique songs around 1965 and 2020, which suggest periods of increased diversity in the music industry. These spikes may be indicative of significant shifts in cultural factors influencing the chart at those times.

Part C



This bar plot shows the number of ten-week hits for the top artists who have had at least 30 songs on the Billboard Top 100 for a minimum of 10 weeks. The plot highlights the artists with the most ten-week hits, illustrating their long-term success on the charts. Elton John has the most ten-week hits, with around 52 songs.