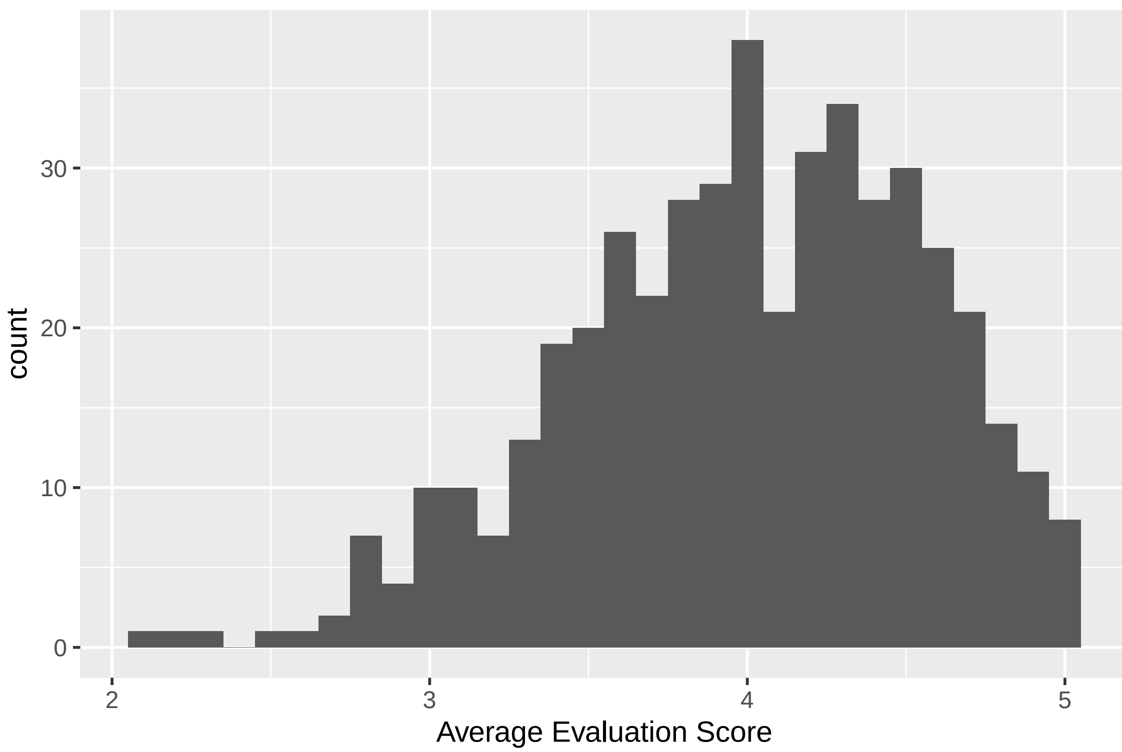
Homework 2

Brian Pham, bp26834, https://github.com/bpham010/SDS315-Homework2

2024-01-18

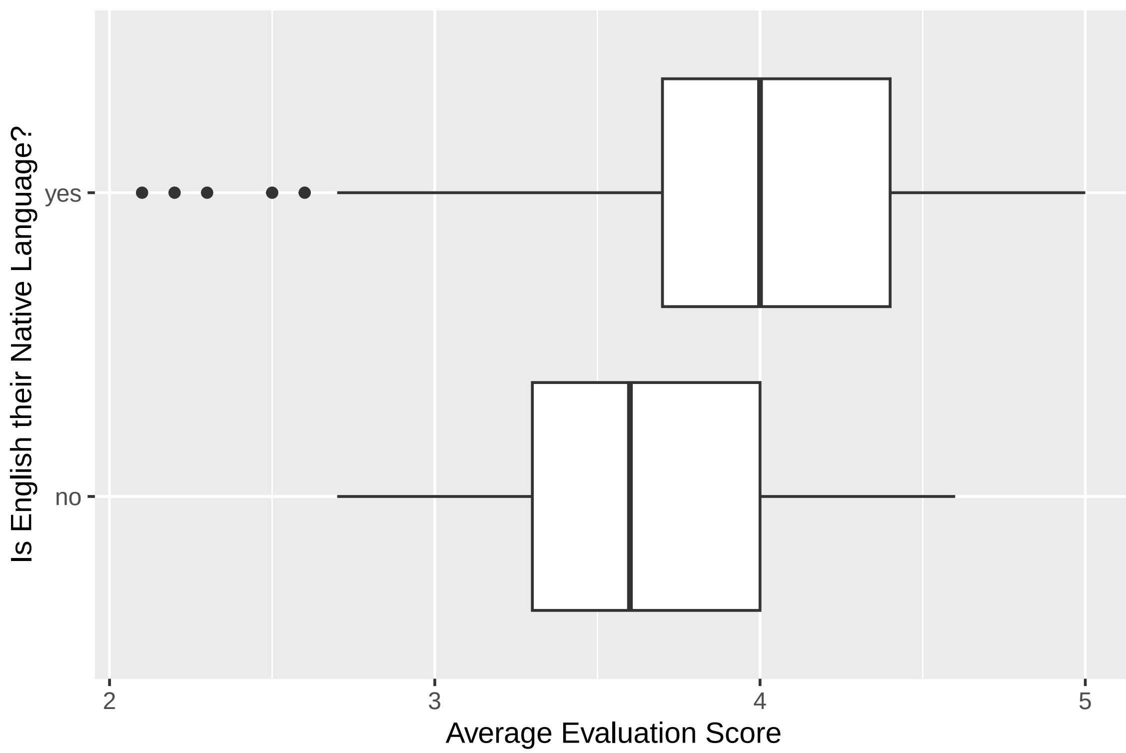
Problem 1: Beauty, or not, in the Classroom

Part A



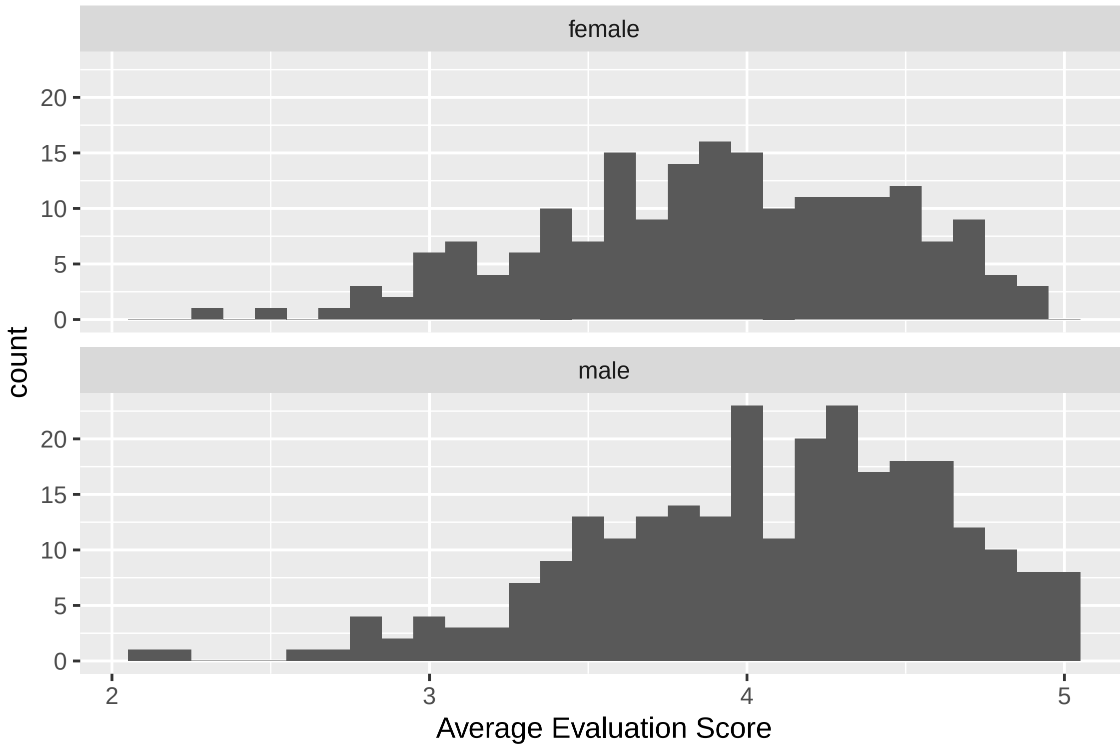
The overall data of course evaluation scores outputs a nicely left-skewed histogram. The average evaluation score is 3.998 while the median evaluation is 4. The abundance in high ratings shows that there are more highly evaluated professors at UT than the latter.

Part B



The first thing many will notice is that professors who are native english speakers on average, have a higher evaluation score than those whose first language isn't english. Another interesting feature is that there are no non-native english professors with a rating higher than 4.75 but they also do not have any professors with ratings below 2.5 while the latter group does for both marks. Overall, the numbers align because professors with difficulty speaking english may struggle to teach english speaking students which may decrease their teaching ability compared to those who speak english since language is not at all an issue.

Part C



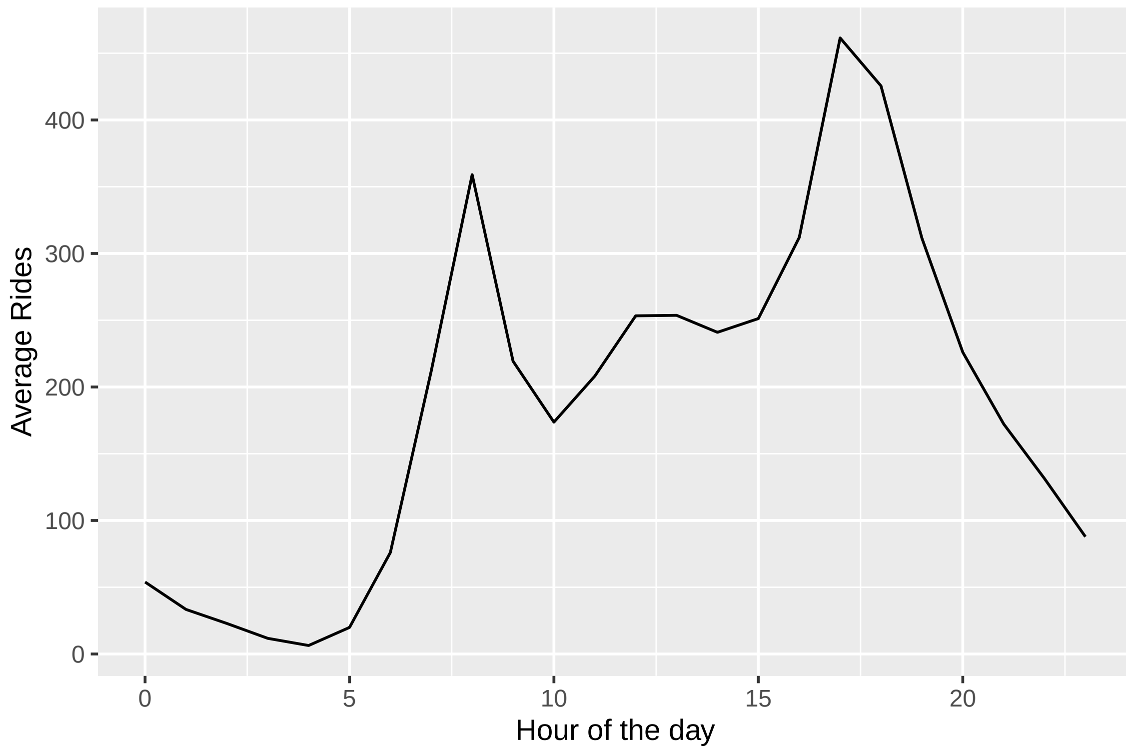
At first glance, the graphs for female and male professor's evaluation scores look even. Both graphs are left-skewed while the male graph is slightly denser. Female professors had an average evaluation score of 3.90 while the male professors had an average of 4.07 evaluation score. Although the male professors have a higher average evaluation score, this doesn't necessary mean that male professors are better professors since there is more data on male professors over female professors.

Part D

At first notice, the scatter plot looks balanced everywhere and appears to have no correlation between a professors beauty rating and evaluation score. Which isn't a surprise to most since a professor's evaluation is usually based on their class and how they teach. But suprisingly, there is a 0.18 correlation constant between the two variables which could possibly mean that students may slightly prefer attractive professors but there are many variables that could prove otherwise.

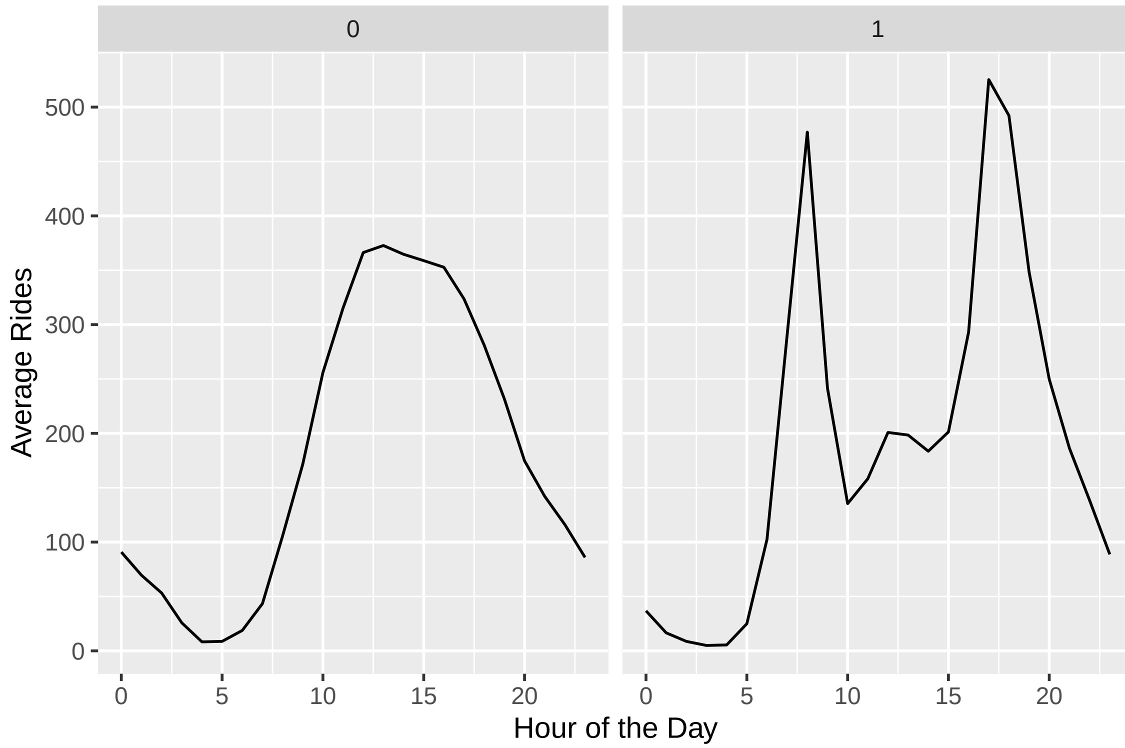
Problem 2: Bike Sharing

Part A



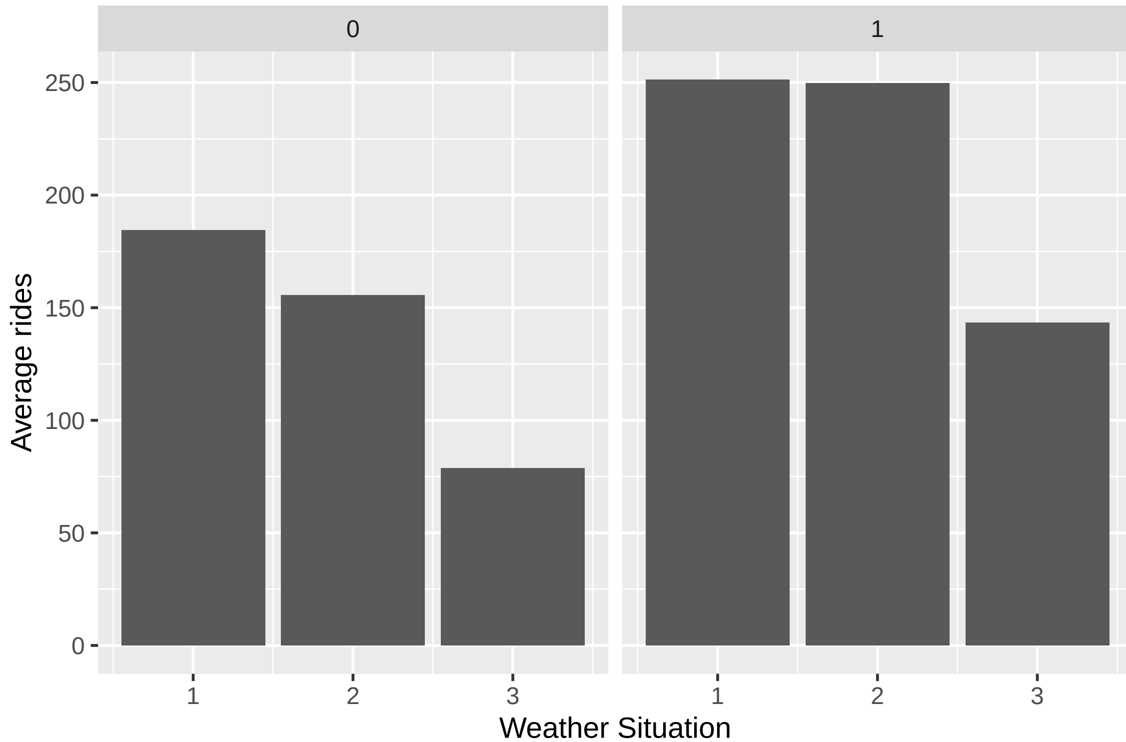
The line graph shows average number of rides during the day throughout the years 2011 and 2012. The average number of rides appears to peak around hour 6 and hour 18 (6:00am and 6:00pm). The increased number of rides may be due to people taking the bikes to get to work and back since the times are around when people start (8-4, 9-5) and may be compensating for time traveled to arrive.

Part B



The above graphs show average number of bike rentals during the day, but this time separated by whether or not they were rented on working days. A working day (represented by 1) is a day that isn't a weekend nor holiday, and it shows the same peak at hours 6 and 18 while the non-working days (represented by 0) appear to peak at hour 12, which may just be people biking leisurely.

Part C



This time, the above graphs show the average number of riders based on whether or not it is a working day, but instead of hours, the riders are grouped by the weather. The first thing some may notice is that many people are less willing to rent bikes on days with inclement weather. This goes for both working days and non-working days. Although people may be less inclined to bike in mist on non-working days, the number of people who biked on sunny days and foggy days during the working day did not change significantly. The weather situation are as follows:

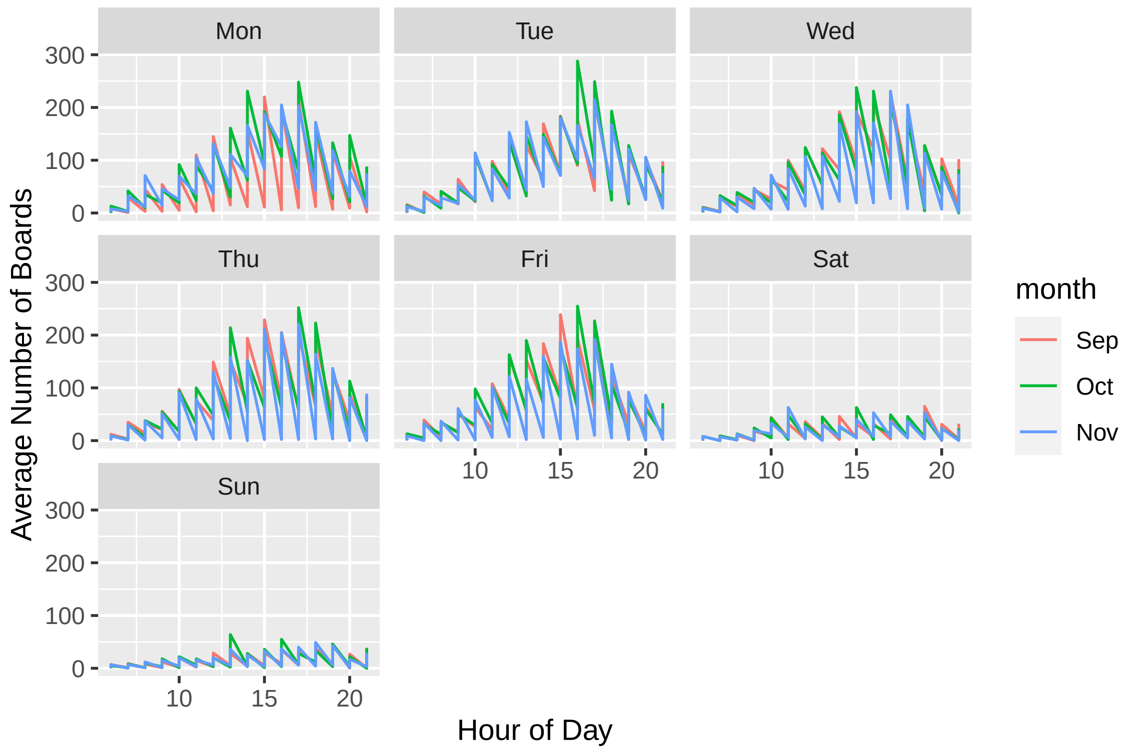
1). Clear, Few clouds, Partly cloudy, Partly cloudy

2). Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist

3). Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds

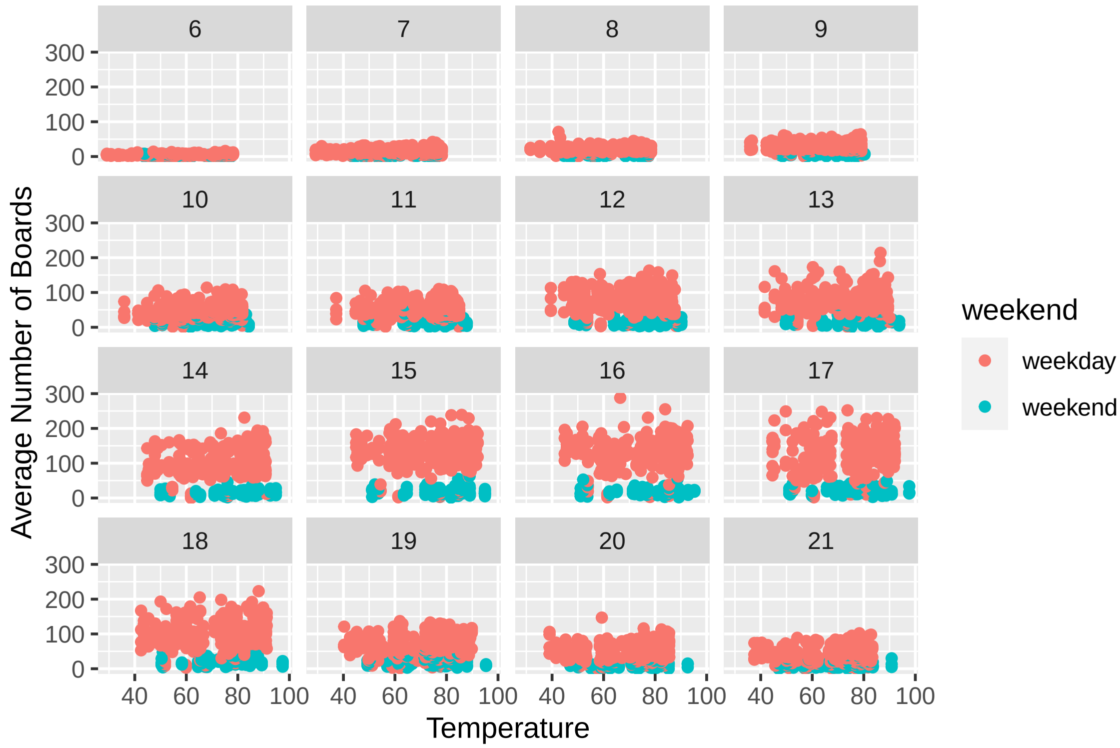
Problem 3: Capital Metro UT Ridership

Part 1



Like the bike rentals seen earlier, the number of rides on the capmetro bus system peak during the 6th and 18th hours of the day, the times where people leave and for and after work and is suprisingly consistent across the weekdays. The weekends face the same situation as seen above in the bike rentals where ridership is not as high. Each graph represents a day of the week showing the time of day (x), and the average number of people who rode the bus (y). Ridership on Mondays in September may be lower than the other months and days because, there may be shifts in people getting or getting laid off jobs in September. Additionally, some people may just not work Mondays in general. Similarly, ridership in the later months may look lower because people are taking off for the holidays, and may not want to work days closer to weekends.

Part 2



The above figure shows the average boarding counts compared to the temperture at the time. Each graph represents an hour of the day starting from 6AM to 9PM. When hour and weekend status is consistent, temperature does not appear to affect the number of boardees on the capmetro bus system. This may be because the bus system is convient is mainly used by those routinely, so temperature may not affect them as much since many students rely on the bus for transportation around campus.

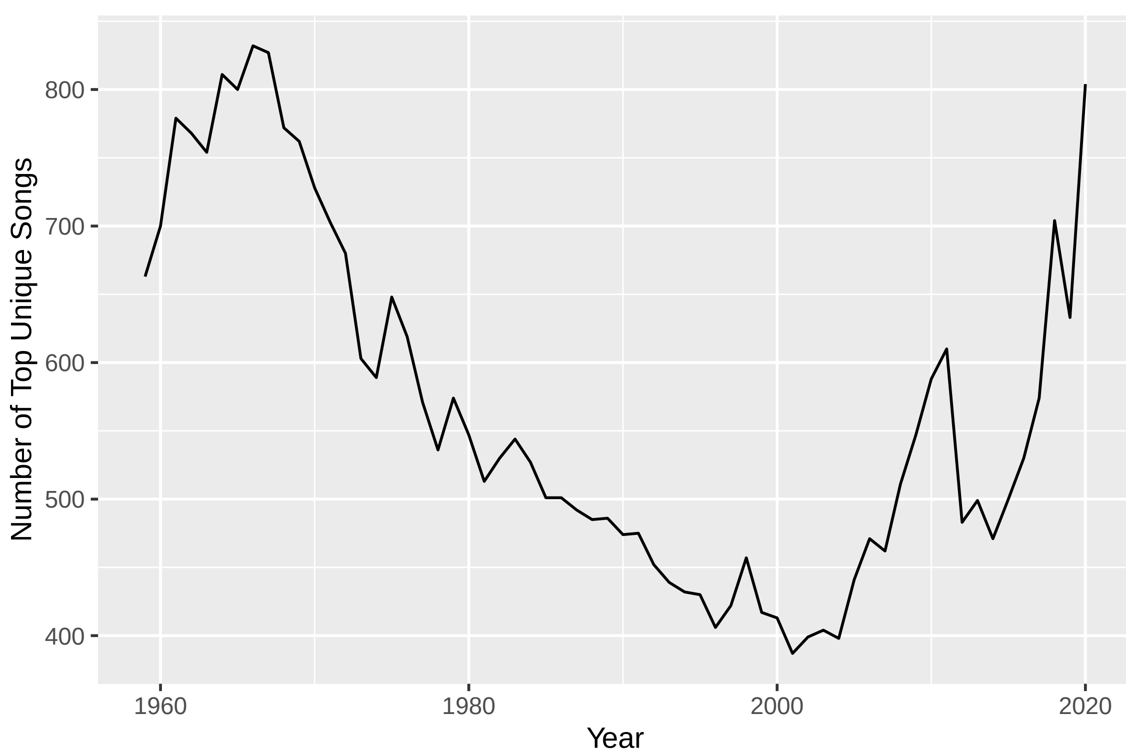
Problem 4: Wrangling the Billboard Top 100

Part A

performer	song	count
Imagine Dragons	Radioactive	87
AWOLNATION	Sail	79
Jason Mraz	I'm Yours	76
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
Lifehouse	You And Me	62

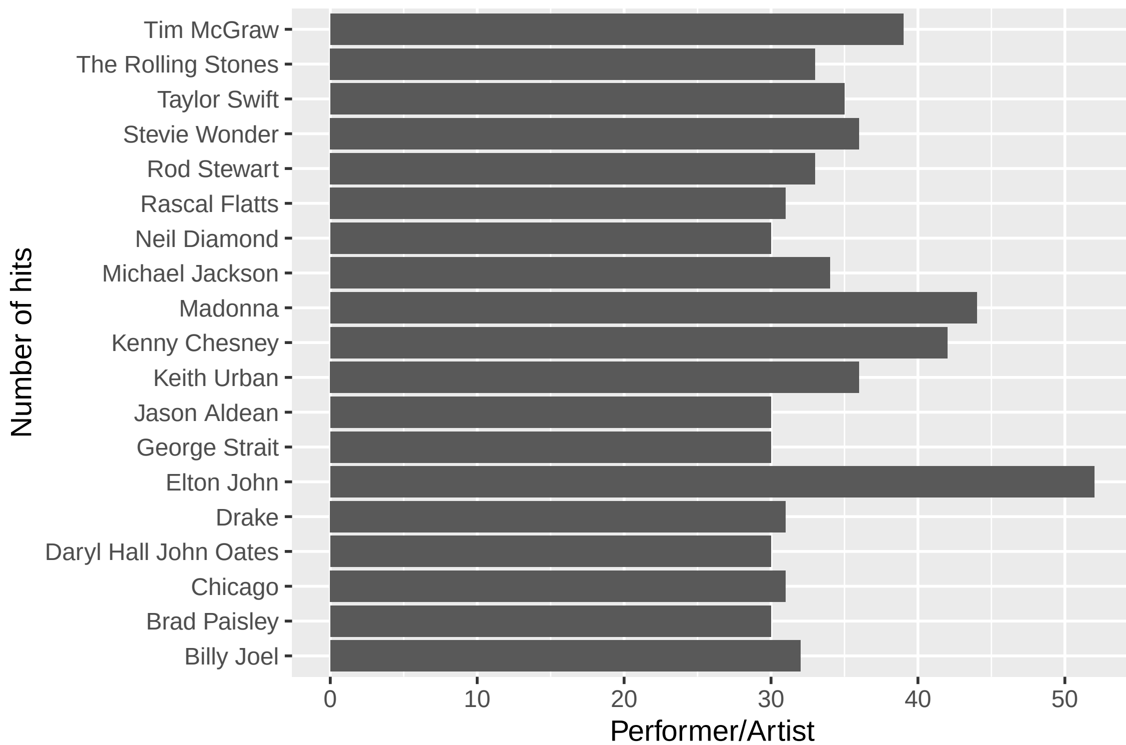
The table above demostrates the top 10 most popular songs since 1985 and their respective artists. Count represents the number of times the respective song as appeared in the Billboard Top 100.

Part B



This figure shows the number of unique songs that made it to the Billboard Top 100 each year. The number of unique songs peaked around the early 1960s and dipped around the early 2000s. This may be due to the changing technology in the modern world because the number of unique songs rises again significantly in the 2020s.

Part C



Last but not least, the above figure shows 19 artists who have at least 30 songs that appeared in the Billboard Top 100 for at least 10 weeks each. Unsuprisingly, Elton John has the most hits with 52 songs making the top charts of the Billboard Top 100 with Madonna coming in close at 2nd with 44 songs.