Homework 2

Mahima Mahajan

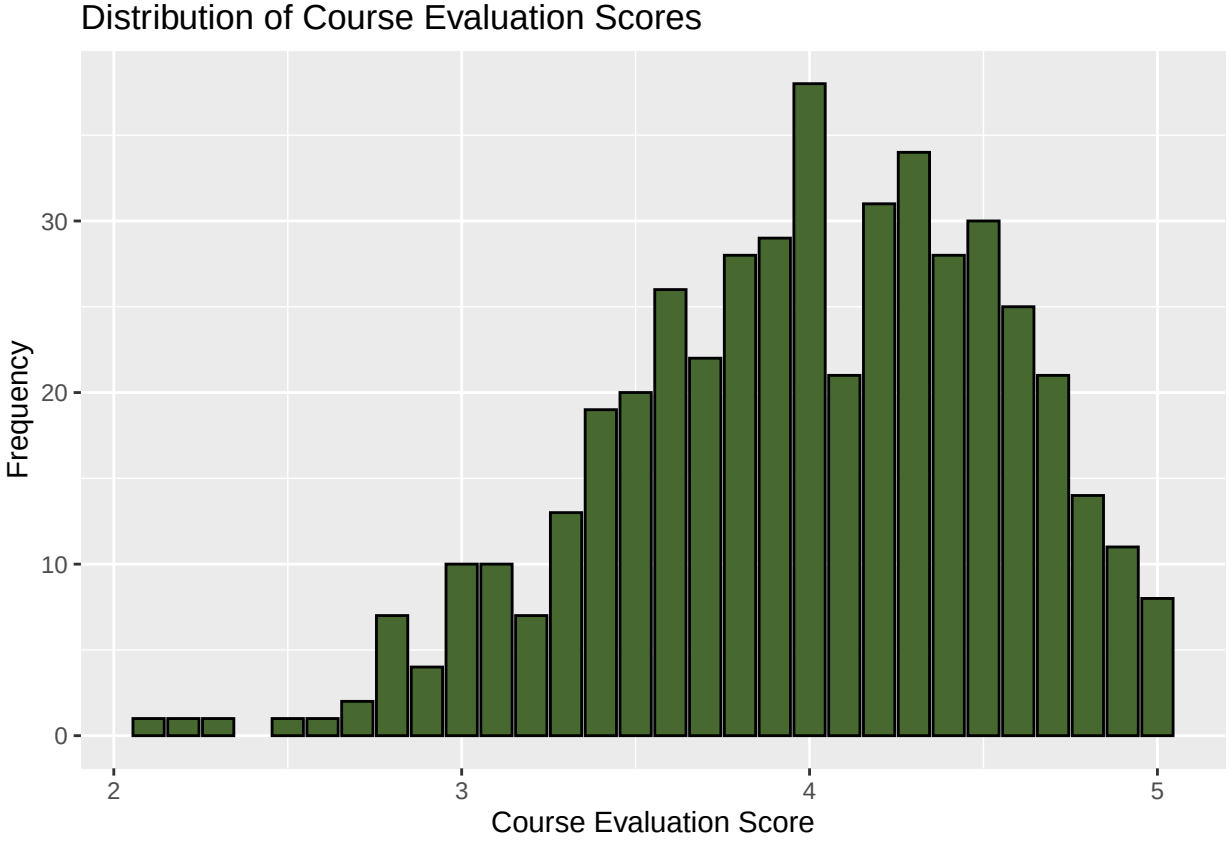
2025-01-18

EID : mkm4582

GitHub repo link (CLICK)

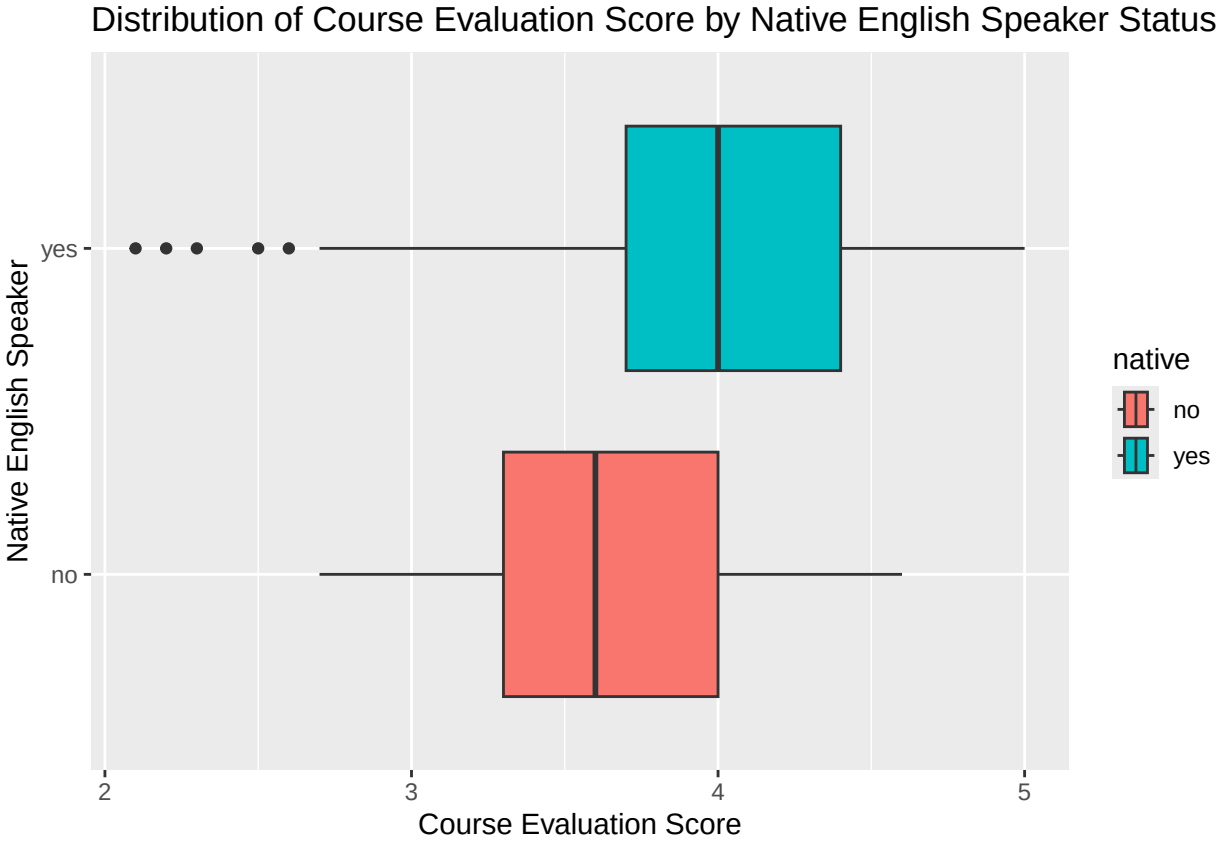
Problem 1. Beauty, or not, in the classroom

Part A



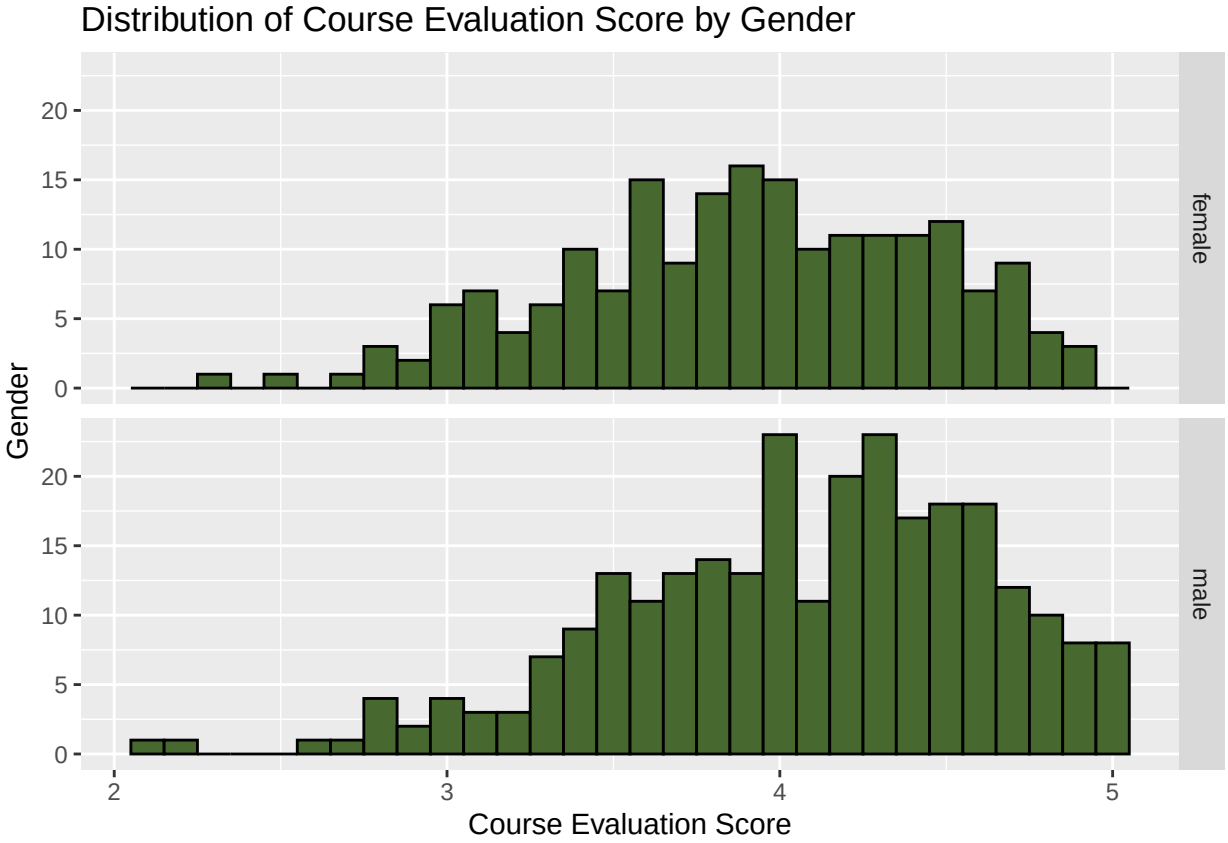
In the graph above, we see that distribution for Course Evaluation Scores is skewed left. The scores with the maximum frequencies given are 4 and 4.3.

Part B



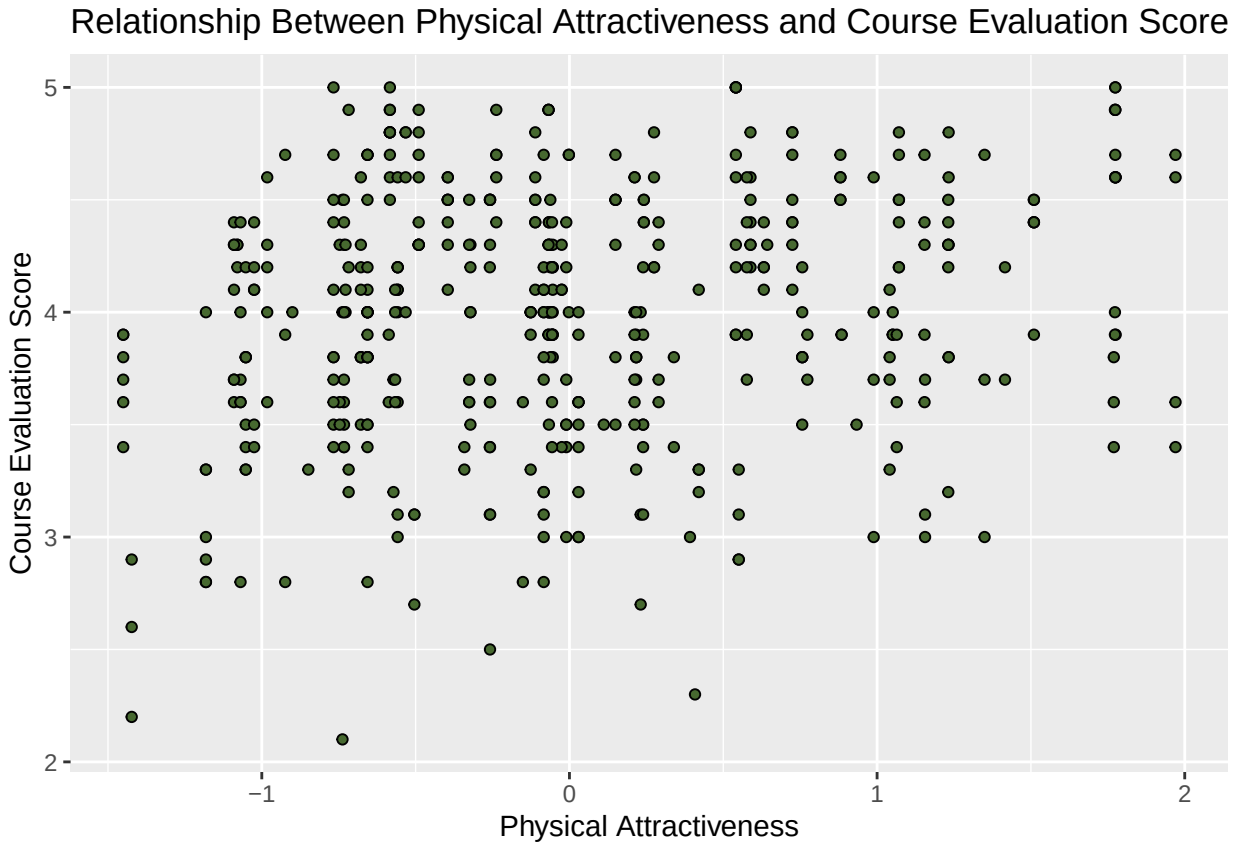
When comparing the descriptive statistics of the graph above, we see that the median score for native English speaking professors is greater than that for non-native speakers. We also see that the Course Evaluation Scores for native English speaking professors has several outliers, all being below a 2.5 evaluation score.

Part C



From the graph of the Distribution of Course Evaluation Score by Gender, we see that both distributions are skewed left. The medians for both graphs are 3.9 and 4.15 for females and males respectively.

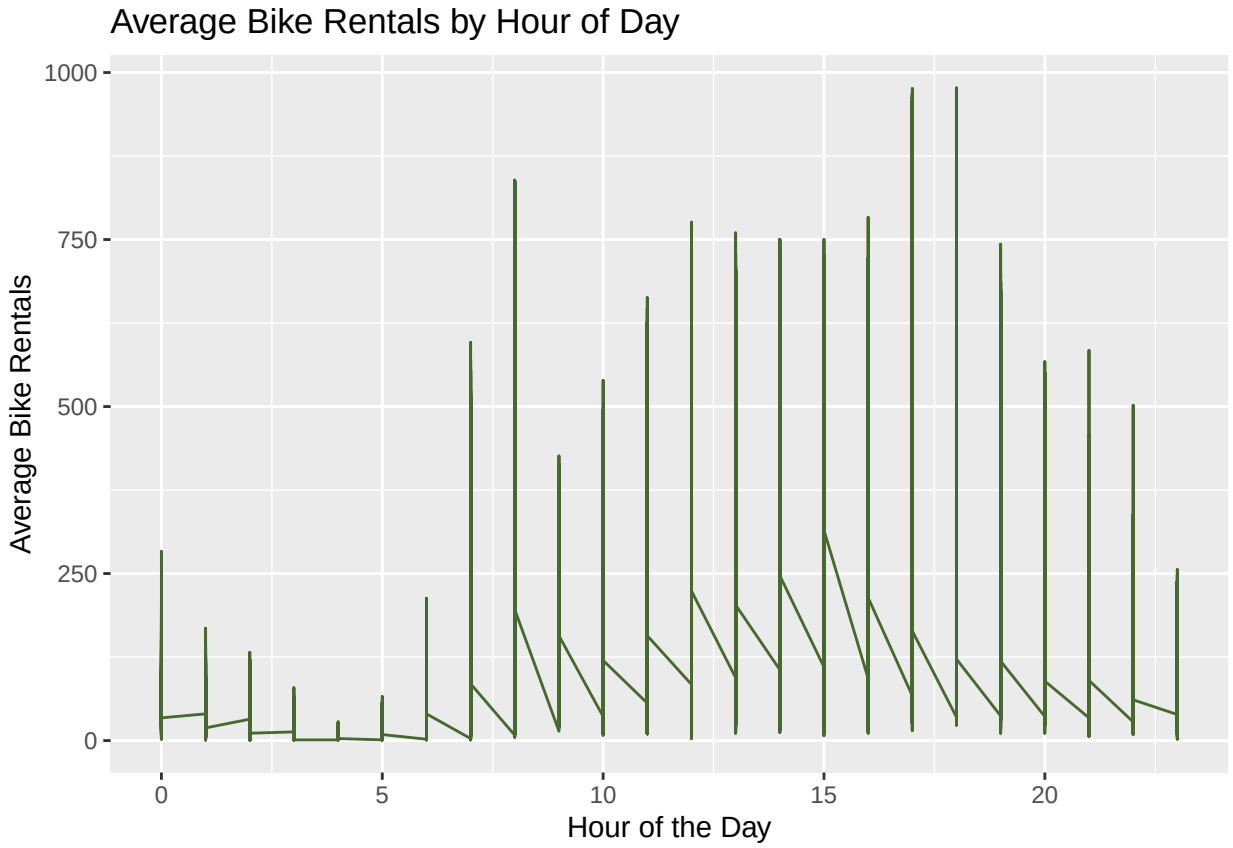
Part D



From the scatterplot above, we see no correlation and a weak strength. There appears to be an equal amount of professors rated with a high course evaluation score who have a negative physical attractiveness score as those who have a positive one.

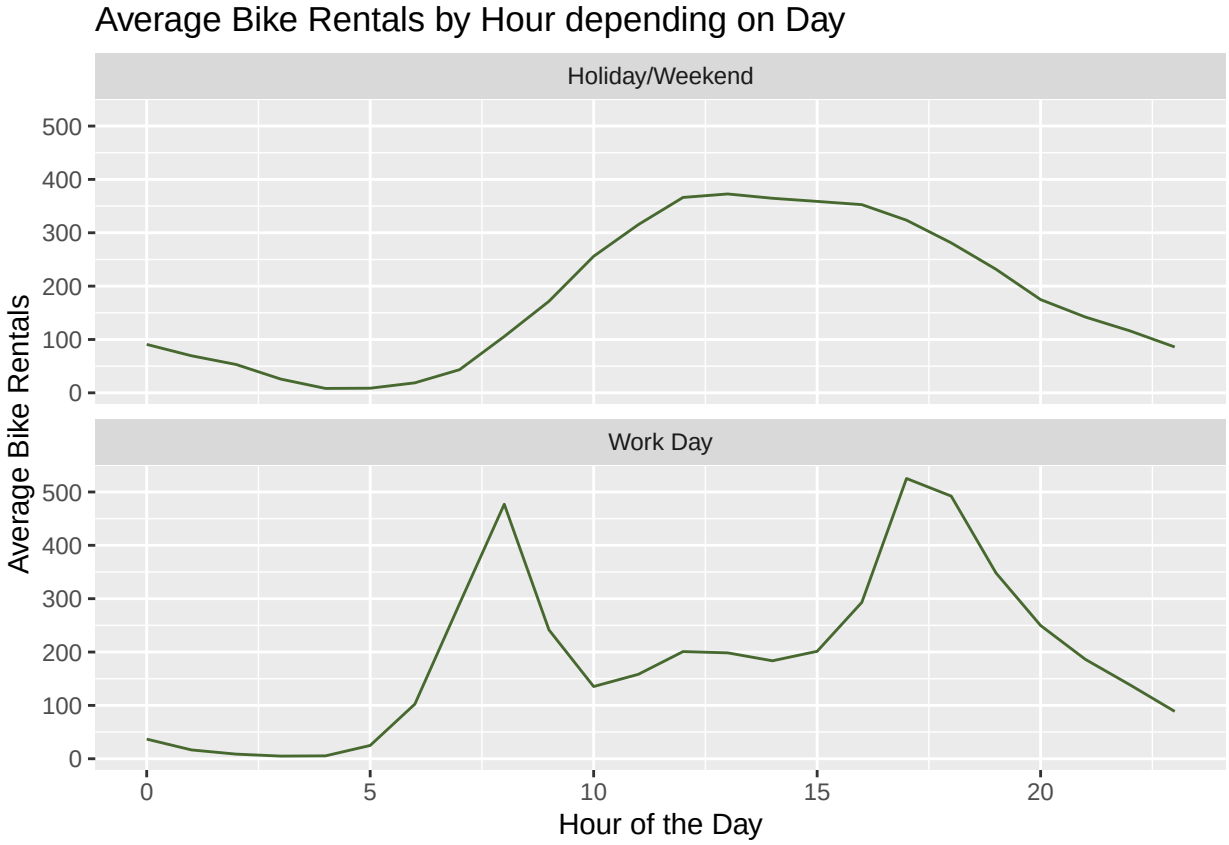
Problem 2. bike sharing

Part A



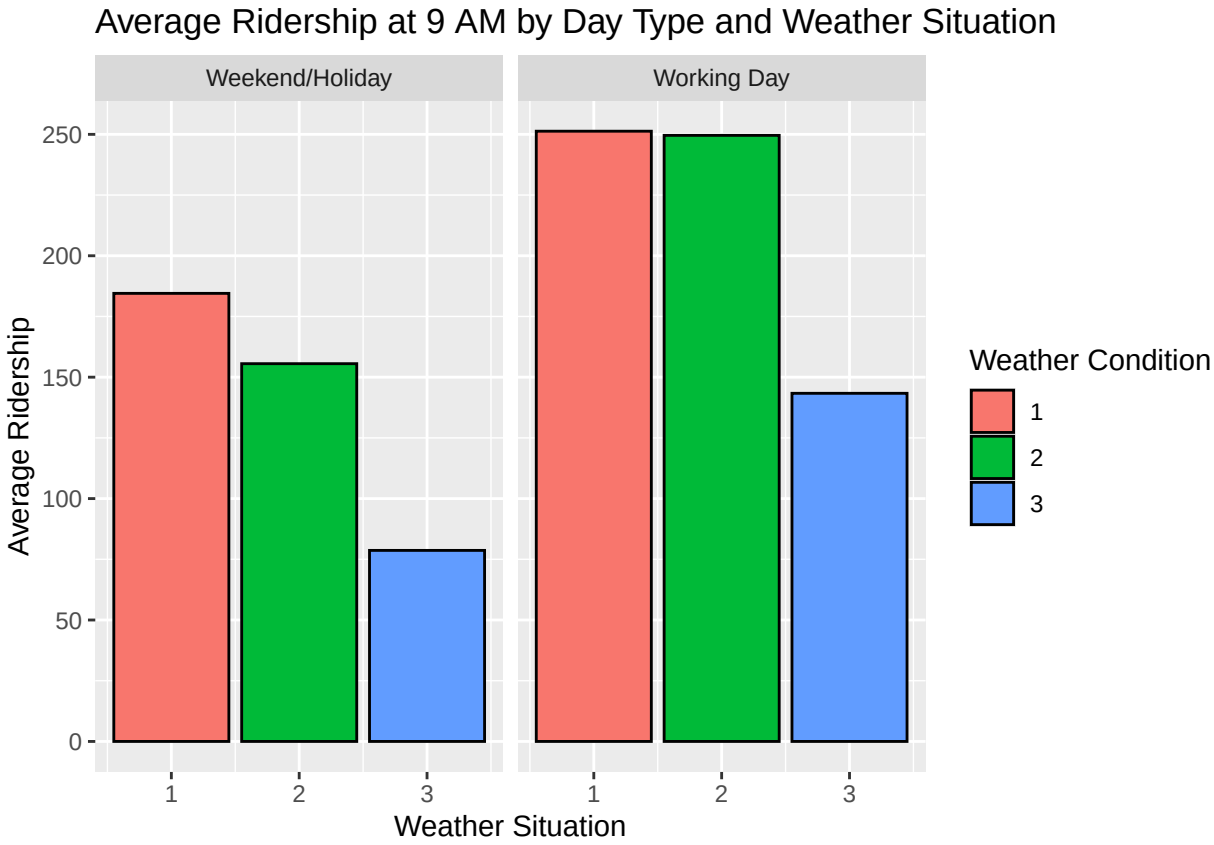
The graph above shows the Average Bike Rentals by Hour of the Day, where the x-axis is the hour (in military time) and the y-axis is the average bike rentals. We see that average bike rentals typically spike in numbers at 6am and 10am and dwindle down towards the end of a day at the 19th Hour (7pm). The majority of average bike rentals are rented between noon and 7pm.

Part B



The visual above shows the Average Bike rentals by Hour, depending on Day. The top graph shows data on a Holiday/Weekday and the bottom graph shows data taken on a Work Day. We see that for Holidays/Weekdays the average bike rentals start to increase at around 7am-8am and gradually decreases towards the end of the day at 6pm-7pm. For Workdays, however, the average bike rentals by hour is clearly the greatest at the start and end of the work day, as we can see two distinct tips at 8am and 5pm. There is little bike rental activity in before, in between, and after those times.

Part C



The bar graph above displays the average ridership at 9am by day type and weather situation.

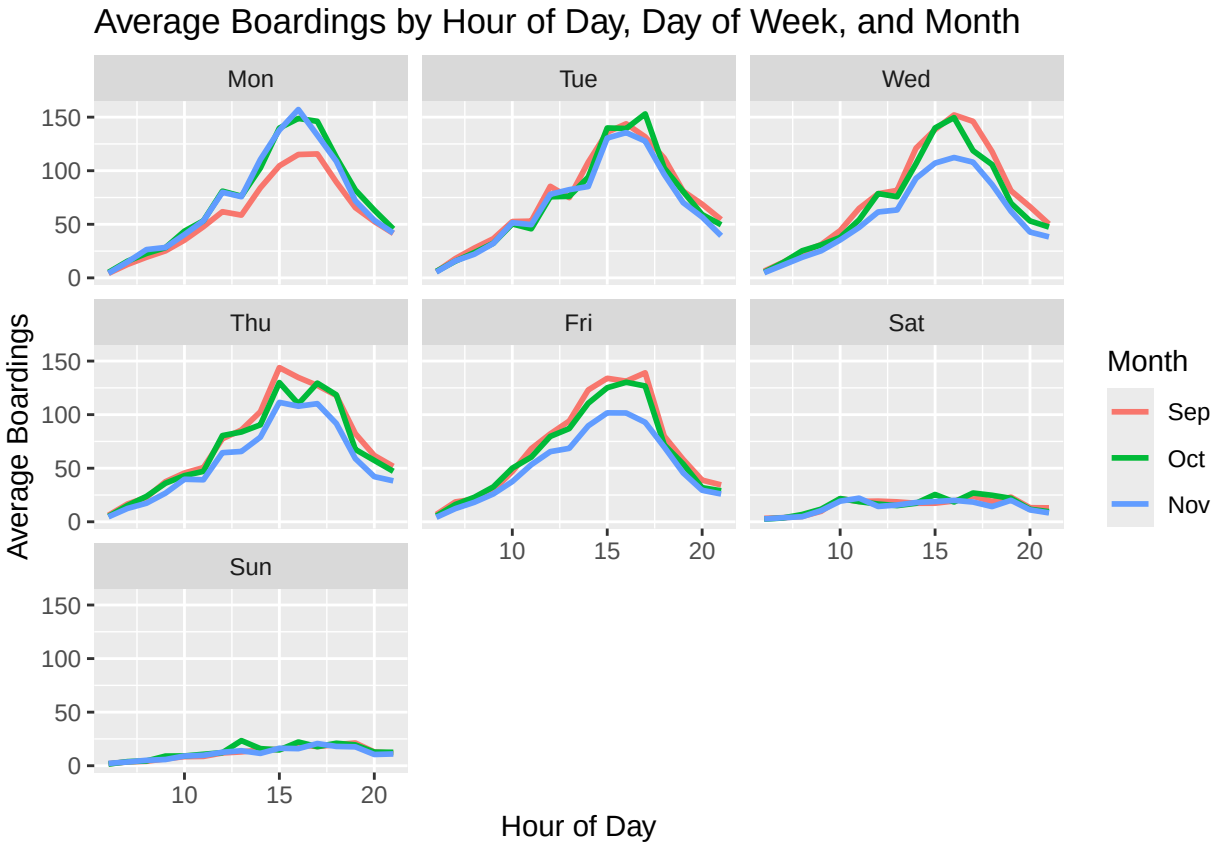
Please note that for the variable Weather Condition:

```
* 1 (moderately light weather):
   + Clear, Few clouds, Partly cloudy, Partly cloudy
* 2 (light weather) :
   + Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
* 3 (moderately severe weather):
   + Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
```

We can see that average ridership on working days weren't impacted unless the weather was moderately severe– the average ridership decreases by almost half in this case. This is likely due to adults/kids staying at home instead of going in-person to work/school. On Weekends/Holidays, the average ridership at 9am steadily decreases as the weather worsens.

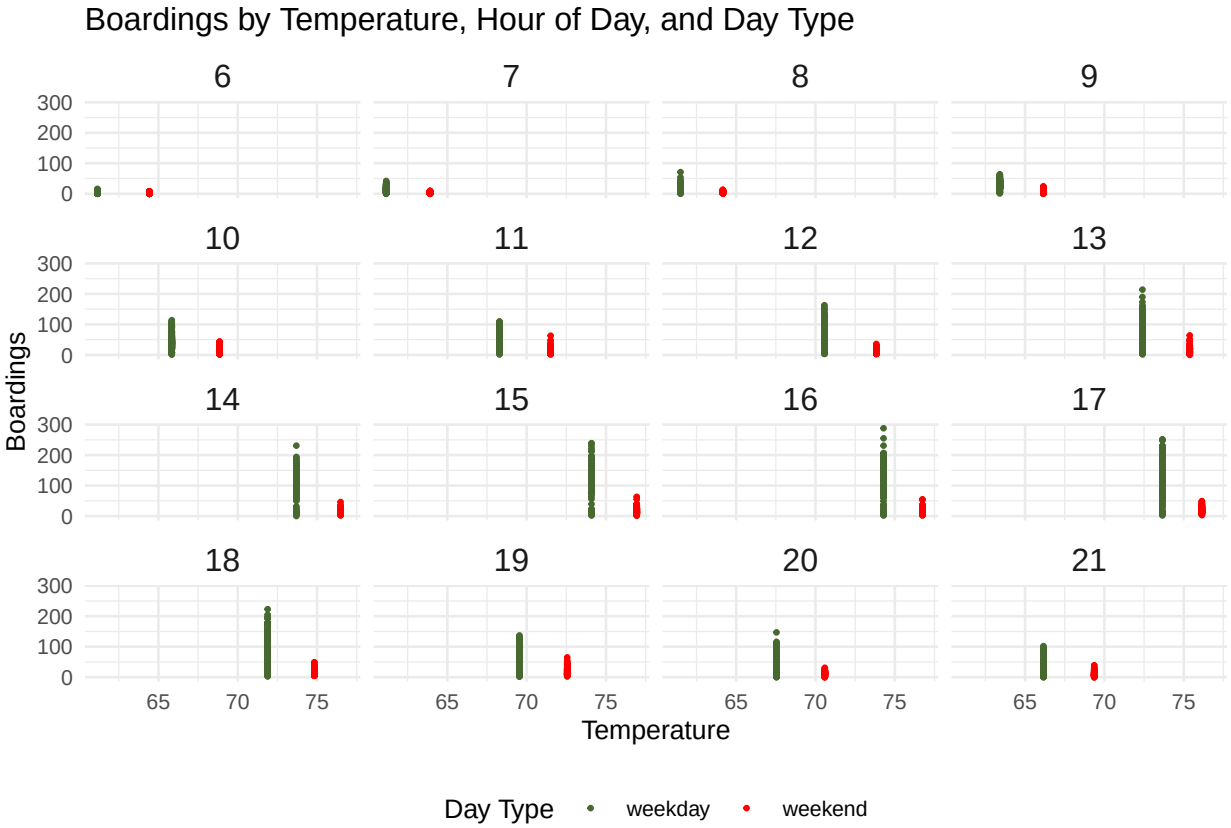
Problem 3. Captial Metro

Part 1



In the figure above, we see 7 graphs of Average Boardings by Hour of Day and Month, each representing a different day of the Week. There are 3 months represented in our data, September, October and November. The hour of peak boardings is broadly similar across the weekdays, maximizing between 3pm-5pm. On weekends (Sat, Sun), we do not see any peaks. The average boardings on Mondays in September is most likely lower because Labor Day is always on a Monday in September. Becuase people don't typically go to work/school on Labor Day, the average boardings look lower. For a similar reason, Weds/Thurs/Fri in November seem to have a lower average number of boardings because of Thanksgiving. Thanksgiving is always on a Thursday, but majority of people take Wednesday and Friday off as well.

Part 2



In the figure above, we see 16 graphs of Boardings by Temperature, where each graph represents a different Hour of day. In each graph we see two colors, blue representing a weekday and red representing a weekend. When we hold hour of the day and weekend status constant, temperature does seem to have a noticeable effect on the number of UT students riding the bus. We see that for temperatures greater than 70 degrees Fahrenheit there are about a 100 students on different hours riding the bus, but when the temperature is below 70 (as we can see in hours 6-10), there are significantly less students riding the bus.

Problem 4. Wrangling the Billboard Top 100

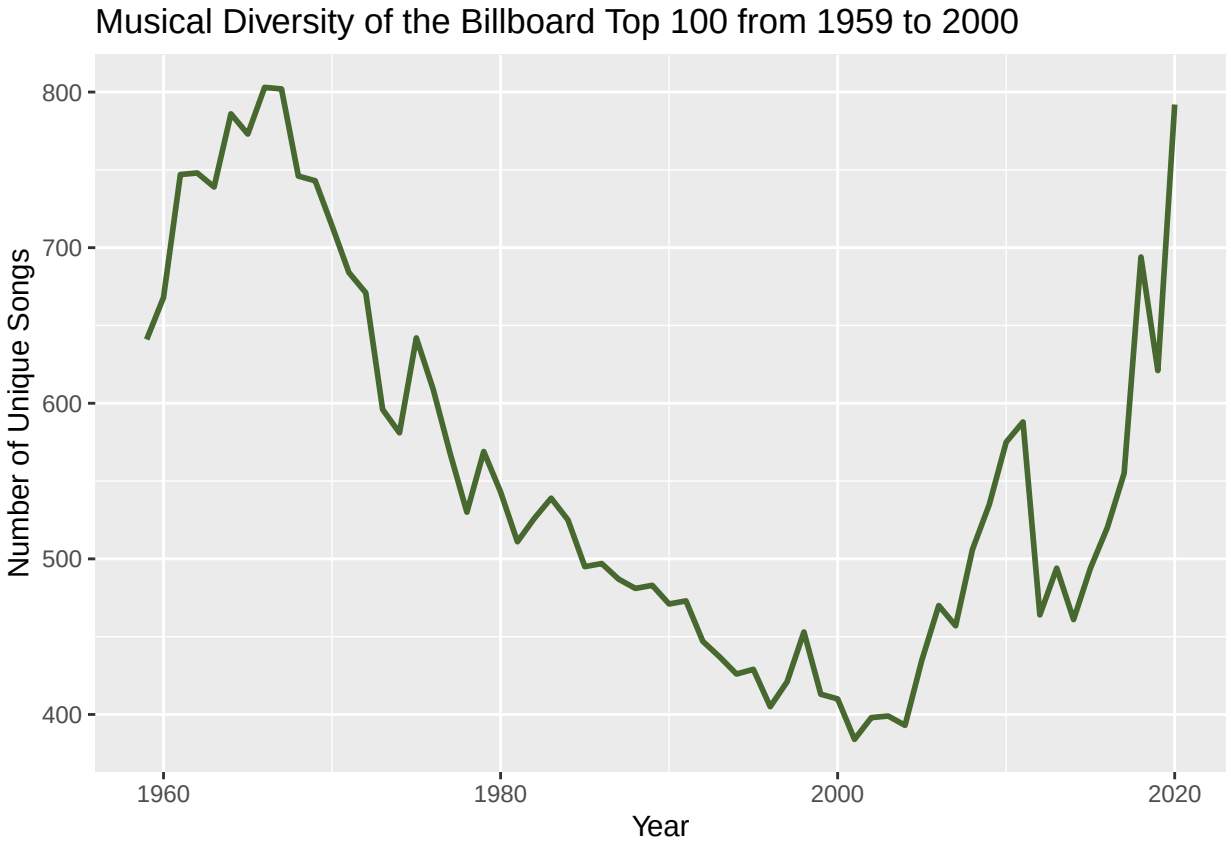
Part A

Table 1: Top 10 most Popular Songs since 1958 based on Billboard Top 100 Chart

Performer	Song	Weeks on Billboard 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

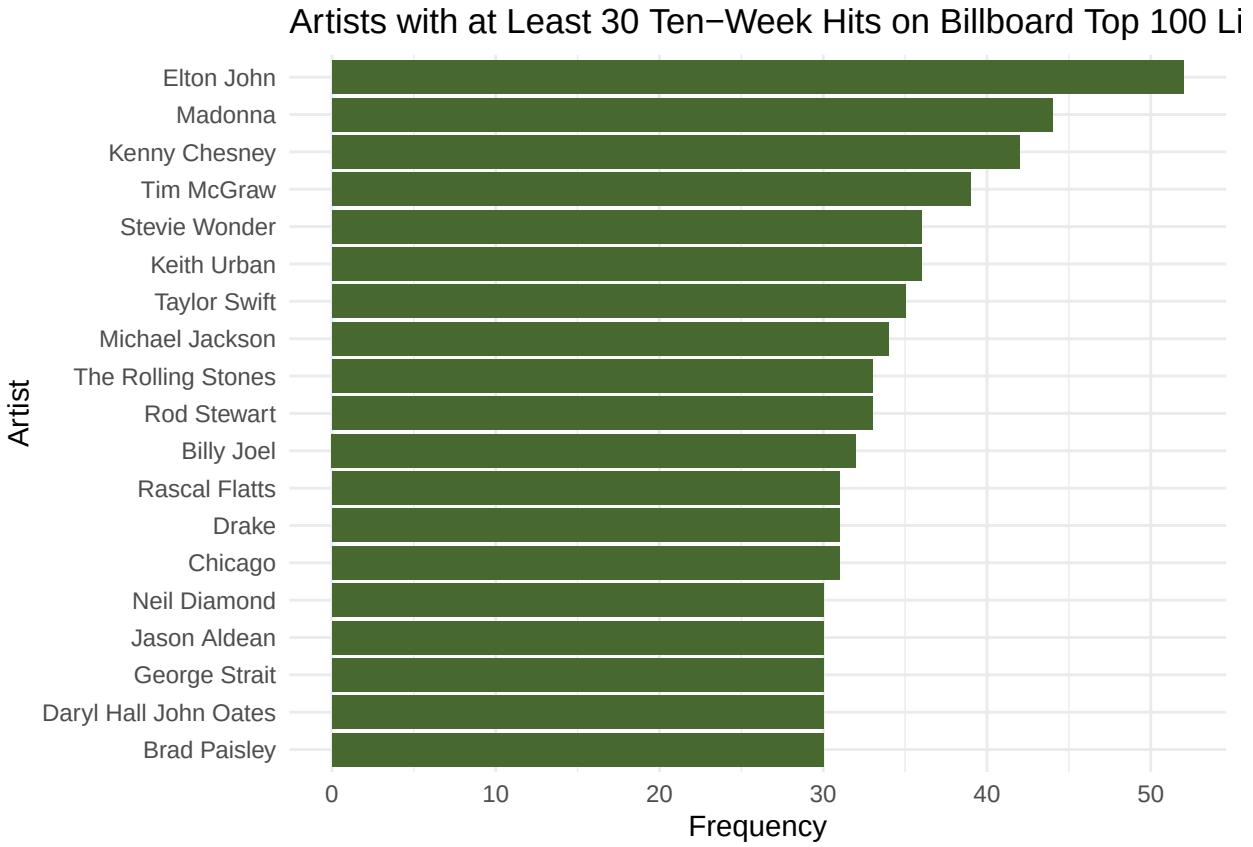
The table above (Table 1) displays the top 10 most popular songs since 1958 (up to late May 2021), as measured by the total number of weeks that a song spent on the Billboard Top 100. We see that there is a 23-week difference between the most popular song, *Radioactive* by Imagine Dragons, and the 10th listed song, *Before He Cheats* by Carrie Underwood. We also see that *Party Rock Anthem* and *Counting Stars* were equally popular at 68 weeks and *Rolling In The Deep* and *Foolish Games/You Were Meant For Me* were also equally popular at 65 weeks.

Part B



The graph above describes the trend of musical diversity from 1959 to 2000 on Billboard's Top 100 list. We see that at the most musical diversity was in 1966 with a diversity of 803 unique songs. There was a drastic decrease in unique songs in the late 1900s, and a sudden increase in unique songs in the early and late 2010s.

Part C



The figure above shows all the artists who had at least 30 ten-week hits on the Billboard Top 100 list, as well how many ten-week hits they each had. We see that Elton John had the most ten-week hits, with a record of 52 hits! There are 5 artists who reached the 30 hits cutoff in our figure: Neil Diamond, Jason Aldean, George Strait, Daryl Hall John Oates, and Brad Paisley.