

Homework 2

Swati Misra (SM83264)

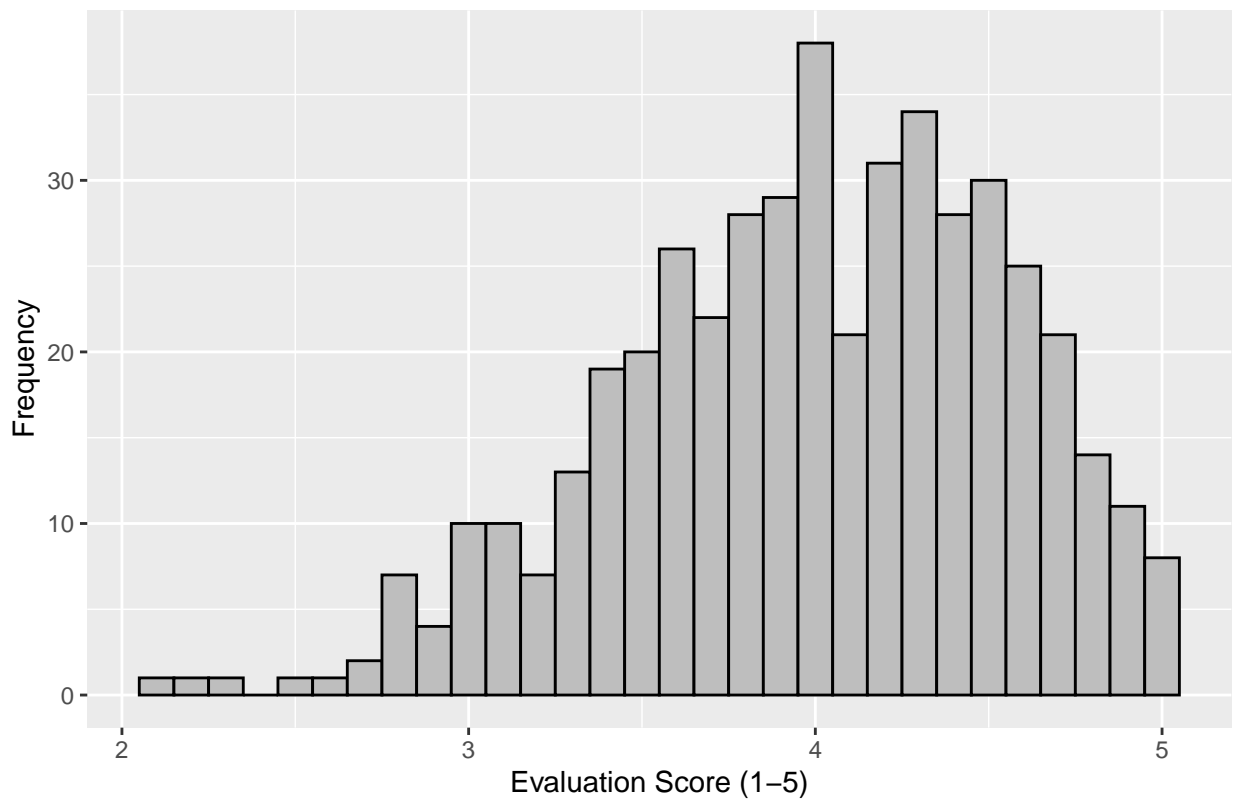
2024-01-24

Link to GitHub Repo: <https://github.com/swatimisra/Homework2>

Problem 1: Beauty, or not, in the classroom

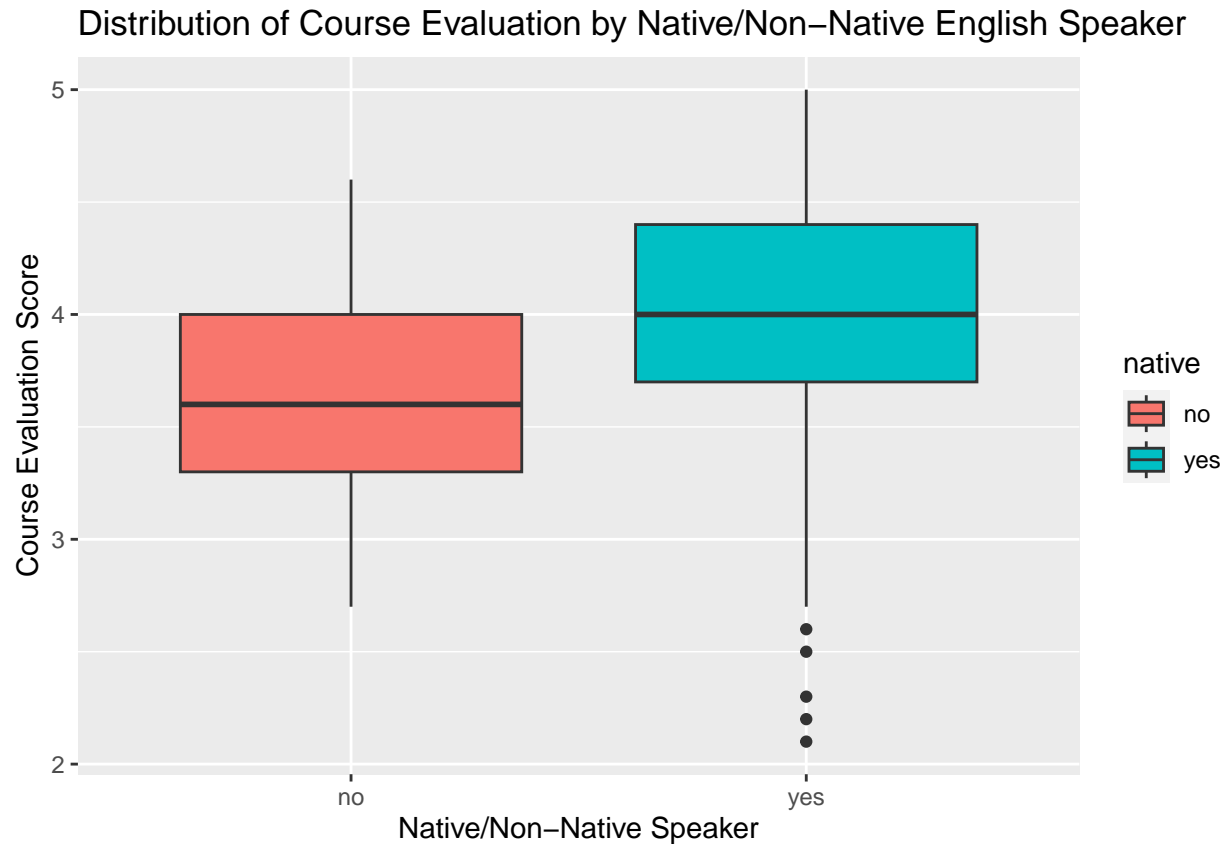
Part A

Distribution of Course Evaluation Scores



This histogram of the distribution of Course Evaluation Scores has the x axis labeled for evaluation scores and y axis labeled for the frequency of each score. This distribution seems to be slightly skewed left. A relatively small number of students evaluated the courses below 3 when compared to 3 and above. The median of this graph is a score of 4. The 0%, 25%, 50%, 75%, and 100% quartiles for evaluation scores are 2.1, 3.6, 4, 4.4, 5 respectively.

Part B



Aggregate Means:

```
## native eval
## 1 no 3.689286
## 2 yes 4.018161
```

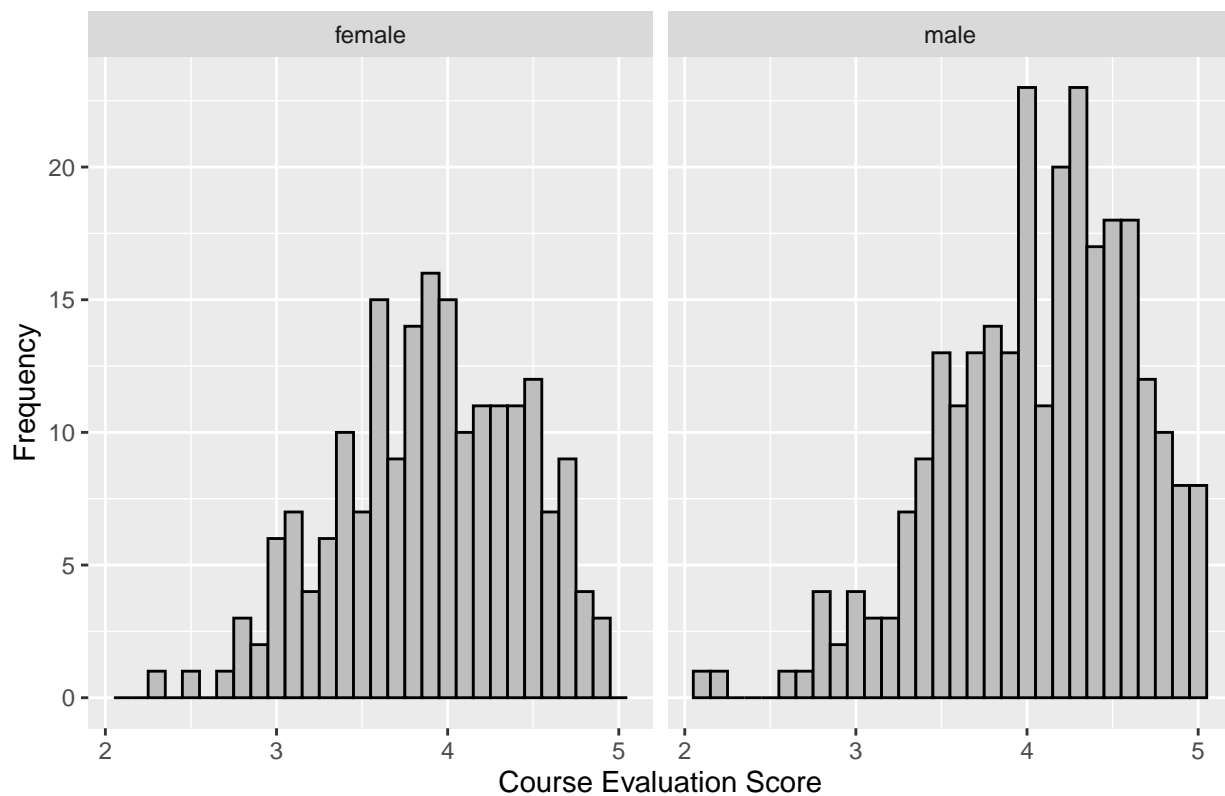
Aggregate Standard Deviations:

```
## native eval
## 1 no 0.4755810
## 2 yes 0.5541767
```

These side-by-side boxplots display the spread of course evaluation scores divided by whether or not a professor is a native English speaker. These plots show that the median of course evaluation scores for native English speaking professors is higher than non-native English speakers. Based on the aggregate means, native English speaking professors had a higher evaluation score on average with 4.018 compared to the non-native speakers score of 3.69. Additionally, there seems to be slightly more variance from the mean in the native English speaking professors scores than the non-native English speaking professors scores as seen in their respective standard deviations of 0.55 and 0.48.

Part C

Distribution of Course Evaluation Score by Gender



Aggregate Means:

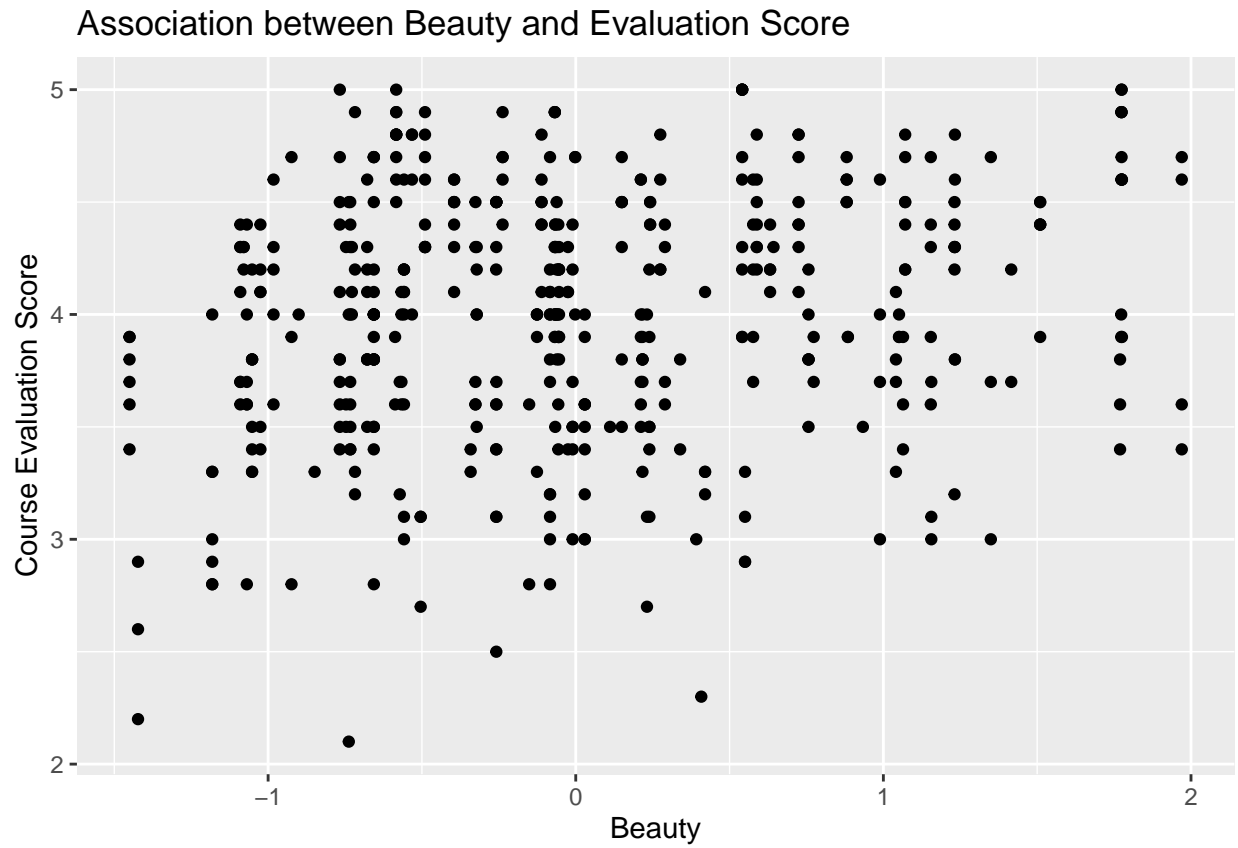
```
##  gender    eval
## 1 female 3.901026
## 2  male 4.069030
```

Aggregate Standard Deviations:

```
##  gender    eval
## 1 female 0.5388026
## 2  male 0.5566518
```

This faceted histogram divides the distribution of course evaluation scores by gender of the professor. Based on this faceted histogram, we can see that there is generally a higher frequency of high course evaluation scores for male professors than female professors. Male professors seem to have a higher average evaluation score than female professors, as seen by their average score of 4.07 and 3.90 respectively. There seems to be approximately equal variance from the mean in their evaluation scores as seen in their standard deviations of 0.54 and 0.56.

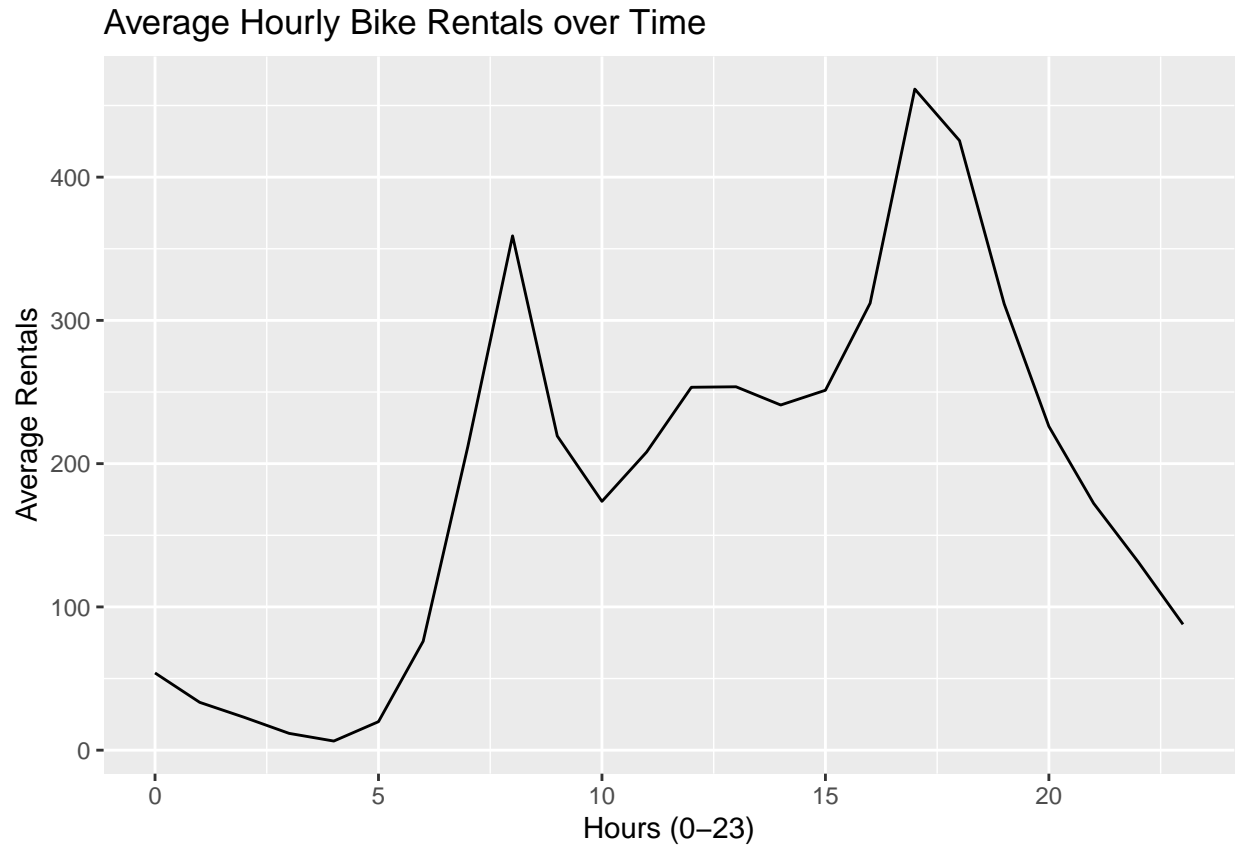
Part D



This scatterplot displays the association between beauty and evaluation score of a professor. This relationship has a correlation coefficient of 0.1890391 which shows a slight positive relationship between beauty score and course evaluation score.

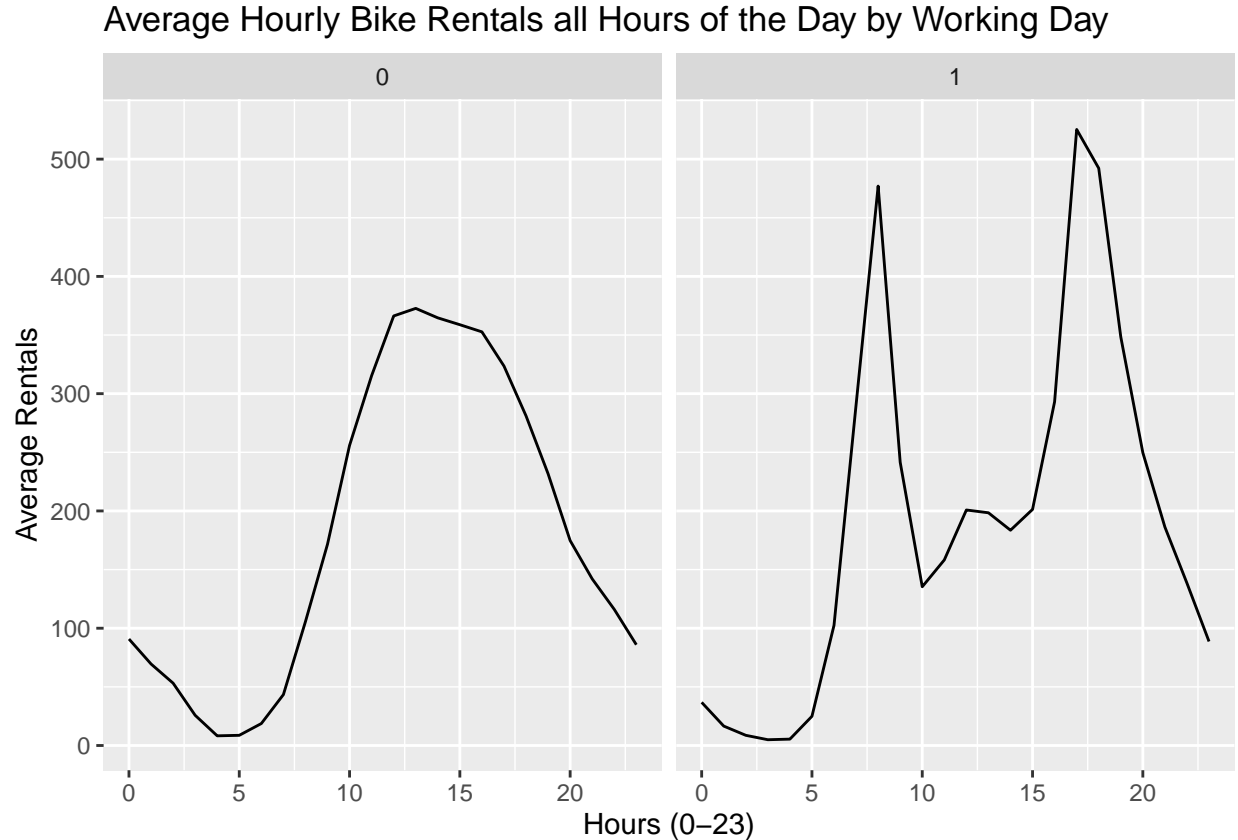
Problem 2: Bike Sharing

Plot A



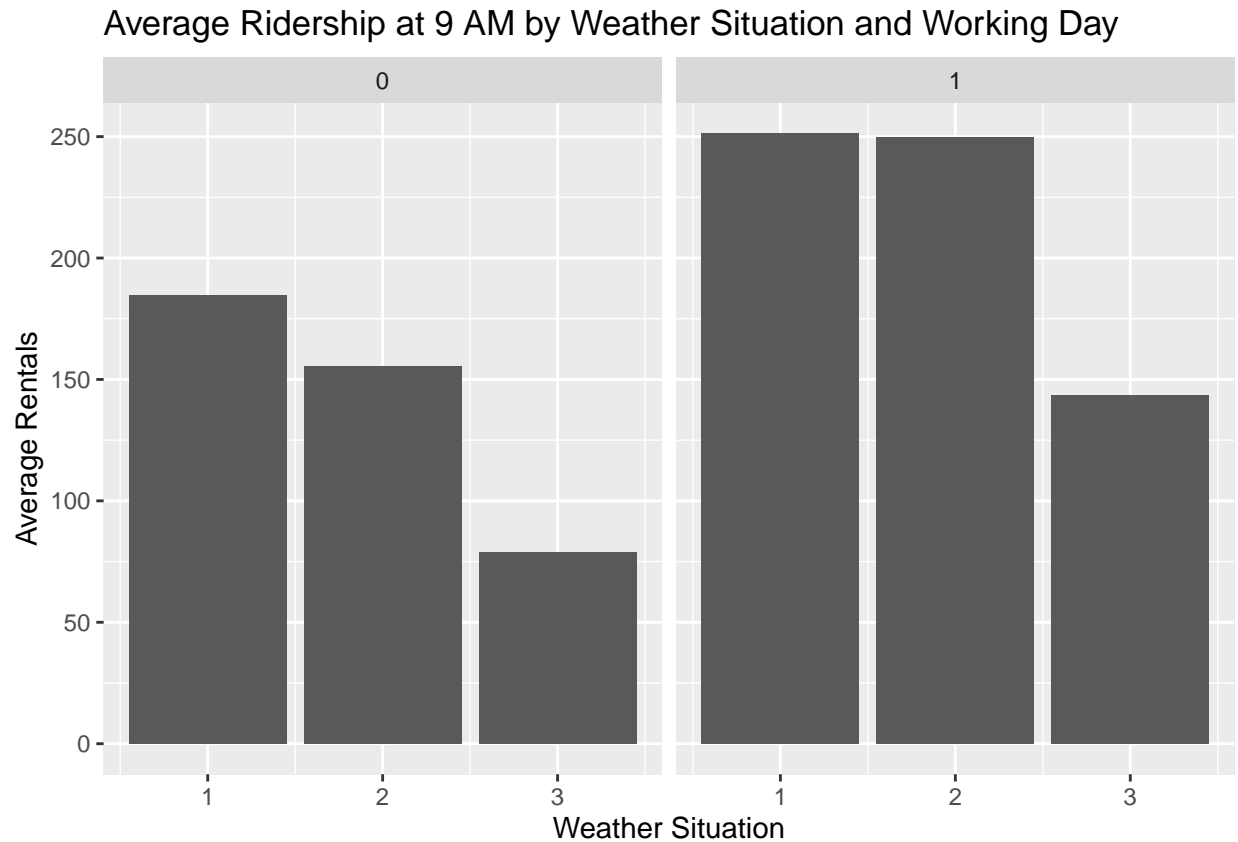
This line graph displays the average bike rentals in Washington D.C. over time (hours measured by 0-23). The x-axis represents the hours (0-23) and the y axis represents the average rentals. Based on this plot, we can see that there is a steep increase in average rentals at two points in time: between the hours of 5 am and 10 am and 3 pm and 8 pm. These modes in the data may be due to the fact that these are the times that people may be commuting to or from work or school.

Plot B



This faceted line graph displays average bike rentals by hour of the day (0-23) and is divided by whether or not it is a working day. The panel labeled 1 displays days when it is neither a weekend nor a holiday and the panel labeled 0 displays days when it is a weekend or a holiday. The x axis is the hours of the day and the y axis is the average rentals. Based on this plot, we can see that the working days have a more dynamic change in average rentals throughout the day (there are 2 modes and a dip between them) while on weekends/holidays, there is only 1 mode between the hours of 10 am and 3 pm. This may be due to the fact that on weekends, people do not have to commute to work or school so they may be leaving around the times of 10 am and 3 pm for lunch hour.

Plot C

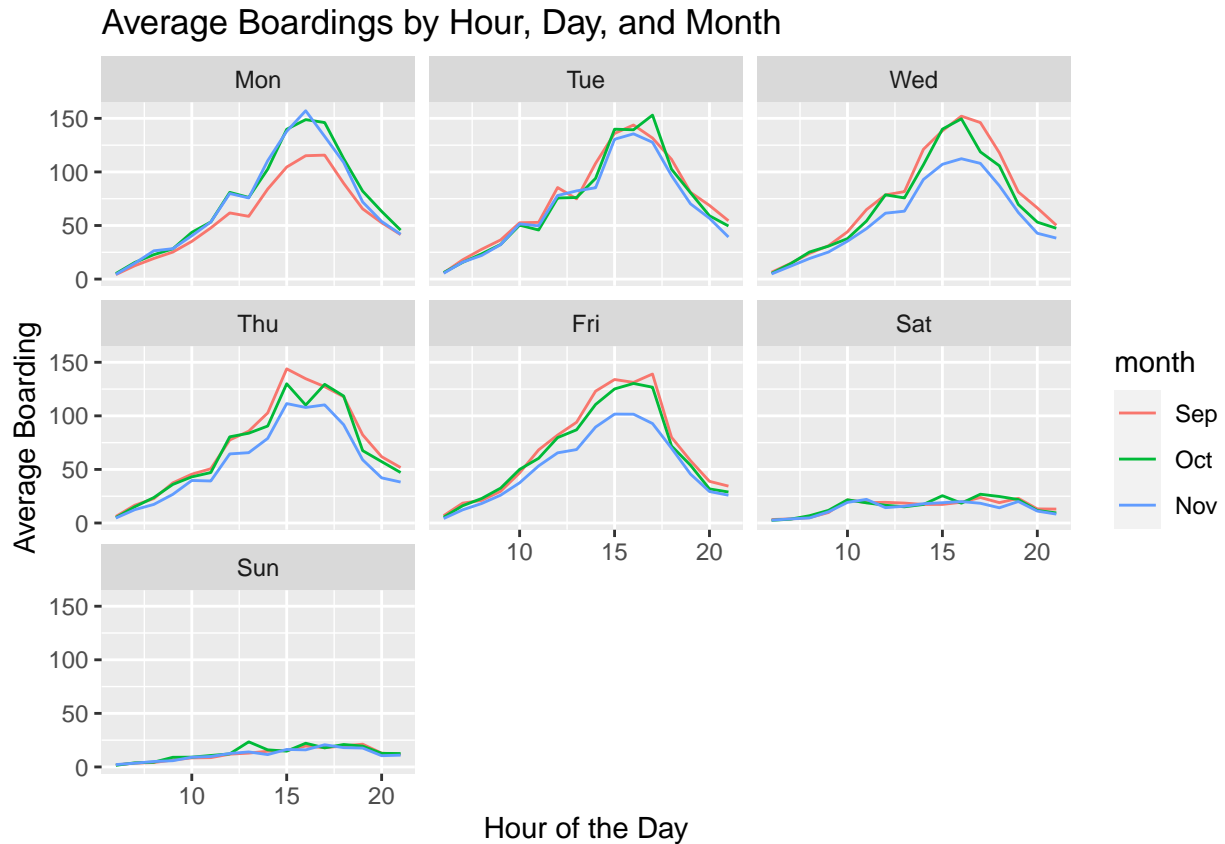


This faceted barplot displays average ridership by weather situation during the 9 AM hour, divided by whether it is a working day or not. The weather situation has been split into number codes with the following values: 1: Clear/few clouds/partly cloudy 2: Misty/Cloudy 3: Light Snow/Light Rain/Scattered clouds with thunderstorms 4: Heavy Rain/Thunderstorm/Snow/Fog.

The panel labeled 1 displays days when it is neither a weekend nor a holiday and the panel labeled 0 displays days when it is a weekend or a holiday. This plot shows that there is a high number of average bike rentals on working days on days with clear/lightly cloudy weather.

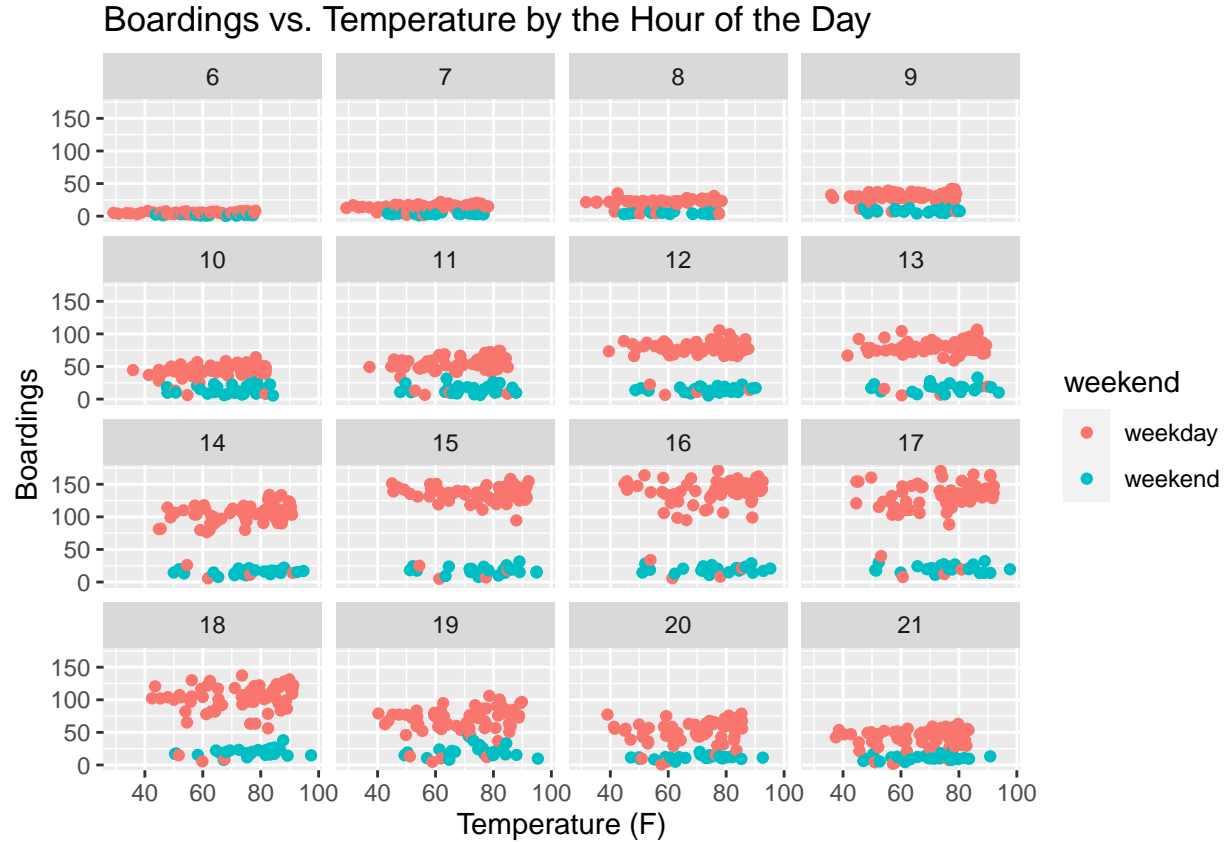
Problem 3: Capital Metro UT Ridership

Part 1



The faceted line graph displays the average boardings on Austin's Capital Metro bus network faceted by day of the week and separated by hour of the day and month. As seen in the plot, the hour of peak boardings is essentially the same throughout the weekdays (M-F), usually between 15 and 20 hours, while the weekend does not share this peak. Average boardings on Mondays in September look lower compared to other days and months because of the Labor Day holiday when less people had to travel to work/school. Similarly, the average boardings on Wednesday, Thursday, and Friday in November look lower because of Thanksgiving break where people got days off school and work.

Part 2



This faceted scatterplot shows Boardings vs. Temperature (Fahrenheit) faceted by hour of the day and colored by whether it is a weekend or weekday. The x axis is the temperature while the y axis is number of boardings on the metro. When we hold hour of the day and weekend status constant, the temperature does not have a noticeable effect on the number of UT students riding the bus. Though the temperature changes, the number of boardings stays approximately constant throughout the graphs.

Problem 4: Wrangling the Billboard Top 100

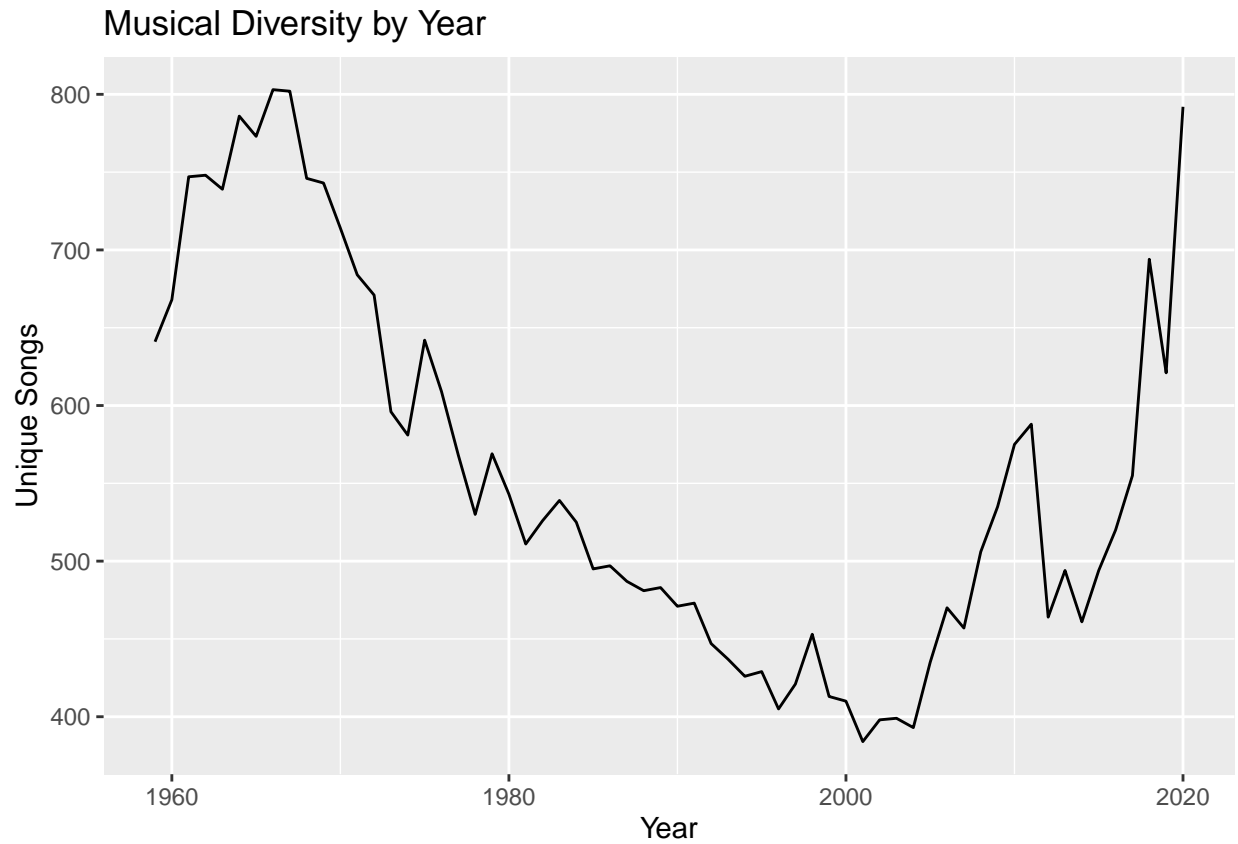
Part A

Table 1: Top 10 Most Popular Songs

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

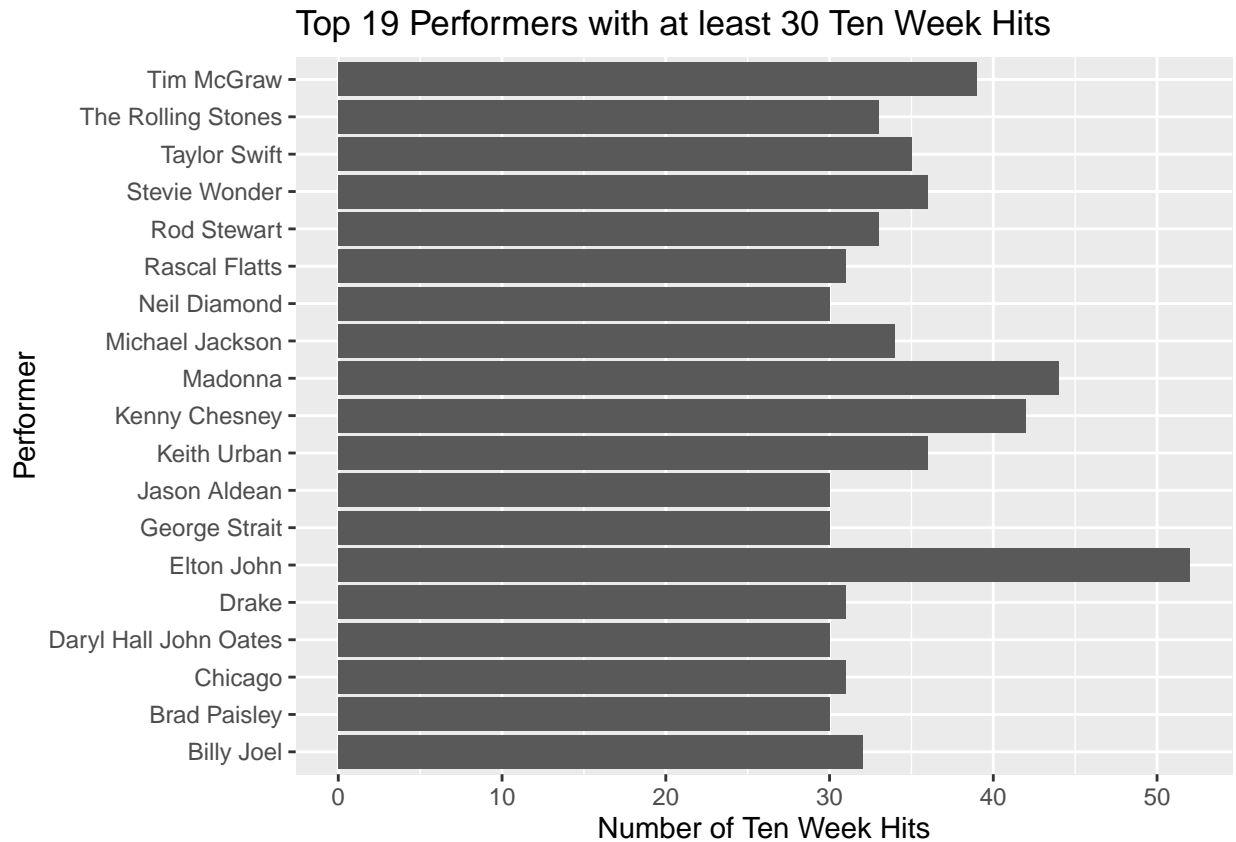
This table displays the top 10 most popular songs since 1985, measured by the total number of weeks a song spent on the Billboard Top 100. For example, the song Radioactive by Imagine Dragons spent 87 weeks on the Billboard Top 100 chart.

Part B



This line graph displays the number of unique songs that appeared in the Billboard Top 100 that year. The x axis is the year while the y axis is the number of unique songs. Between the years of 1980 and 2000, there seems to be a big dip in musical diversity as the number of unique songs decrease sharply. This may be due to the emergence of the “pop” music sound where a specific sound was trending among artists and audiences. There is a mode between 1960 and 1970 where musical diversity is at a high, likely due to an era of experimental music.

Part C



This barplot displays the 19 artists in U.S. musical history since 1958 who have had at least 30 songs that were “ten-week hits”. A ten-week hit is a single song that appeared on the Billboard Top 100 for at least ten weeks. The x axis is the number of ten week hits while the y axis is the name of the performer. Elton John is the performer with the largest number of ten week hits, over 50.