Homework2_SDS 315

Nancy Nakyung Kwak (UT EID: nk24424)
GitHub link: https://github.com/nancy1404/sds315_hw2.git

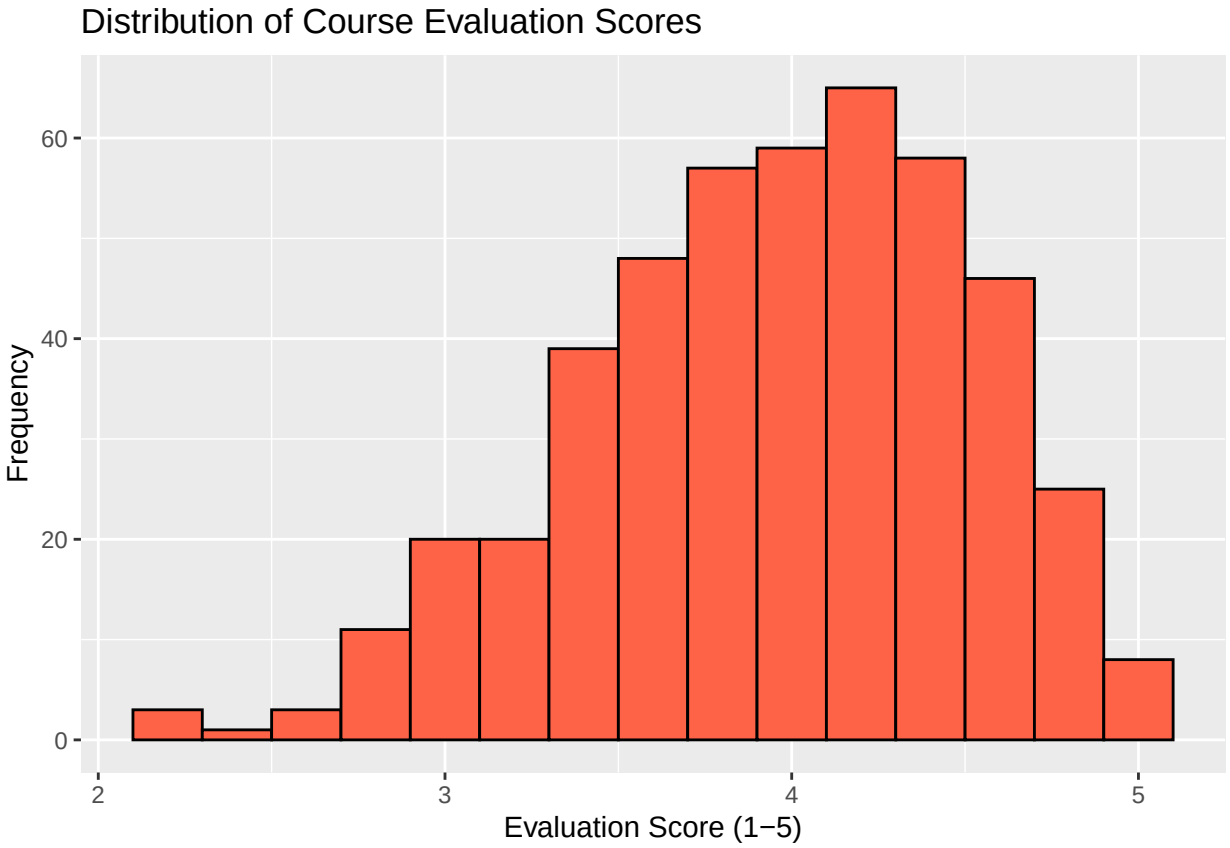
2025-01-23

Contents

Link to My GitHub Repository

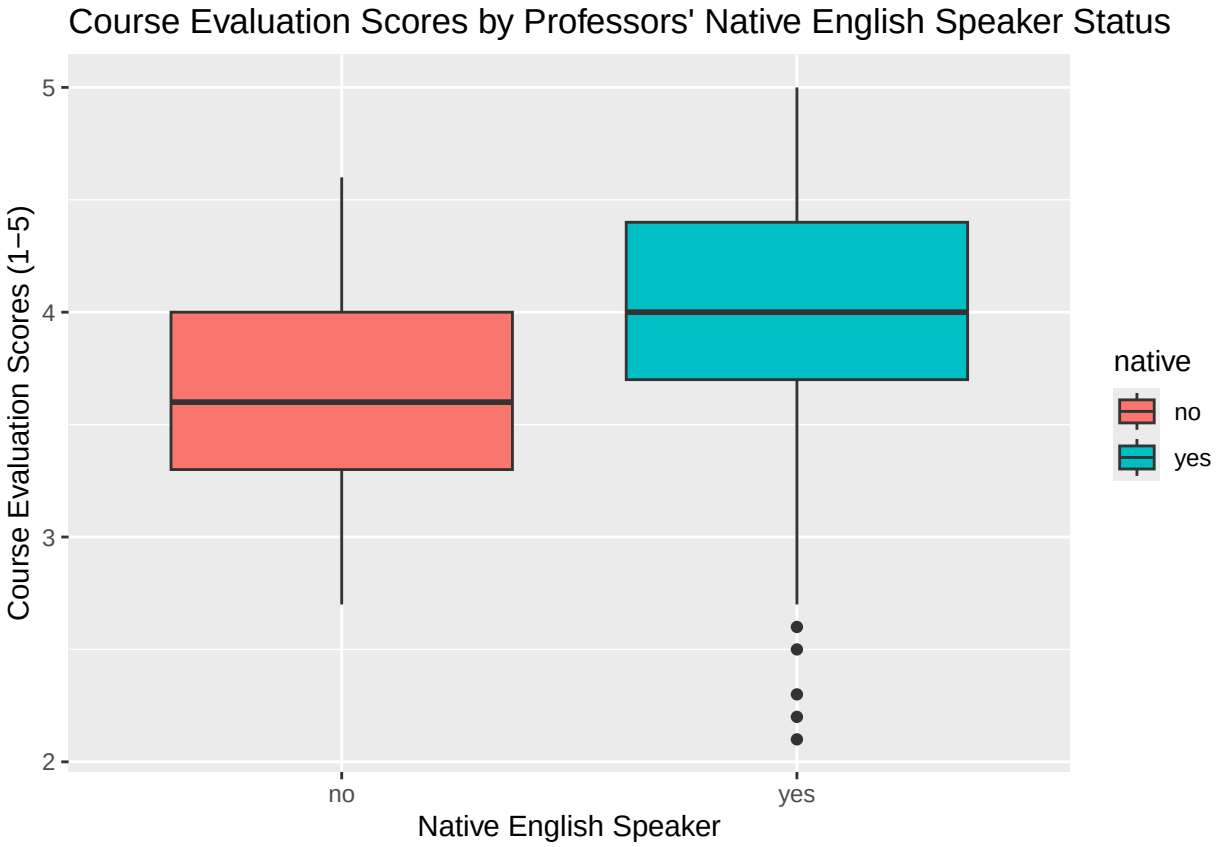
PROBLEM 1: Beauty, or not, in the classroom

PART A



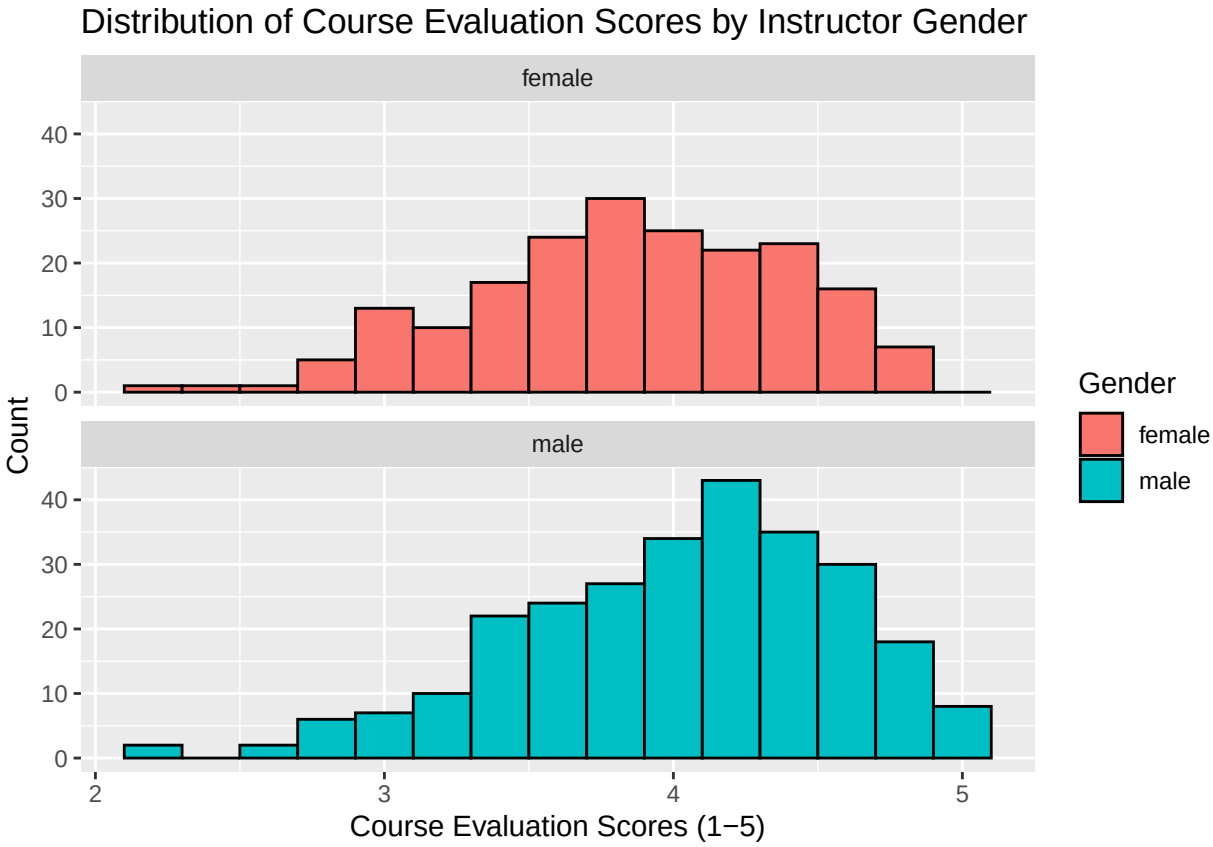
- This histogram displays the distribution of course evaluation scores over the courses in the data set, where x-axis representing the score from 1-5 and y-axis representing frequency of that score.
- Majority of courses are clustered between 4.0-4.5 Evaluation Score, and the plot shows that the students rate their courses more positively than negatively (left-skewed distribution), which displays overall student's high satisfaction for their taken courses.

PART B



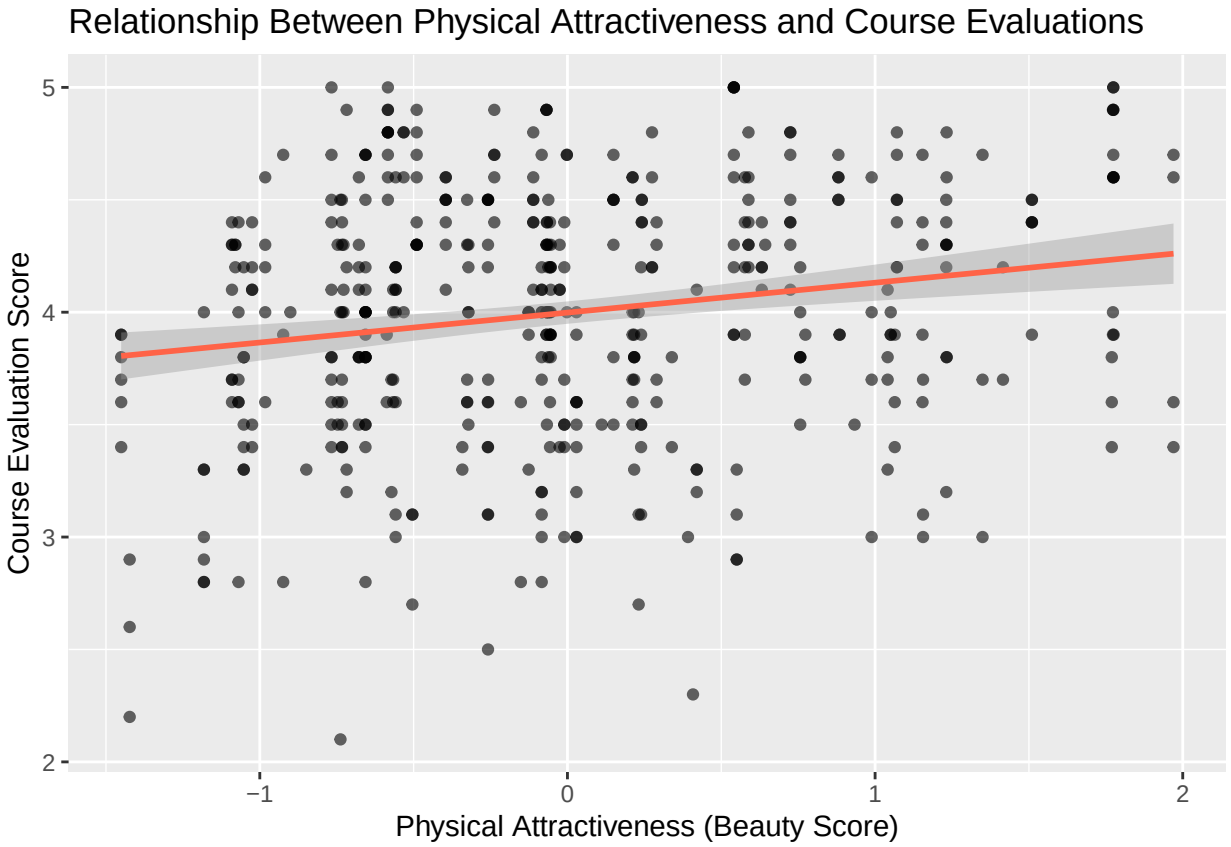
- The side-by-side boxplots show the distribution of course evaluation scores (y-axis) for professors who are native English speakers versus those who are not (x-axis).
- Professors who are native speakers have slightly higher course evaluation scores, mean of 4.0181609,compared to those who are not native speakers with mean of 3.6892857.

PART C



- This faceted histogram compares the distribution of course evaluation scores by gender of the instructors. The histogram shows that both genders have similar distributions, but majority scores for female instructors are clustered more in the middle, while majority scores for male instructors are more left-skewed.

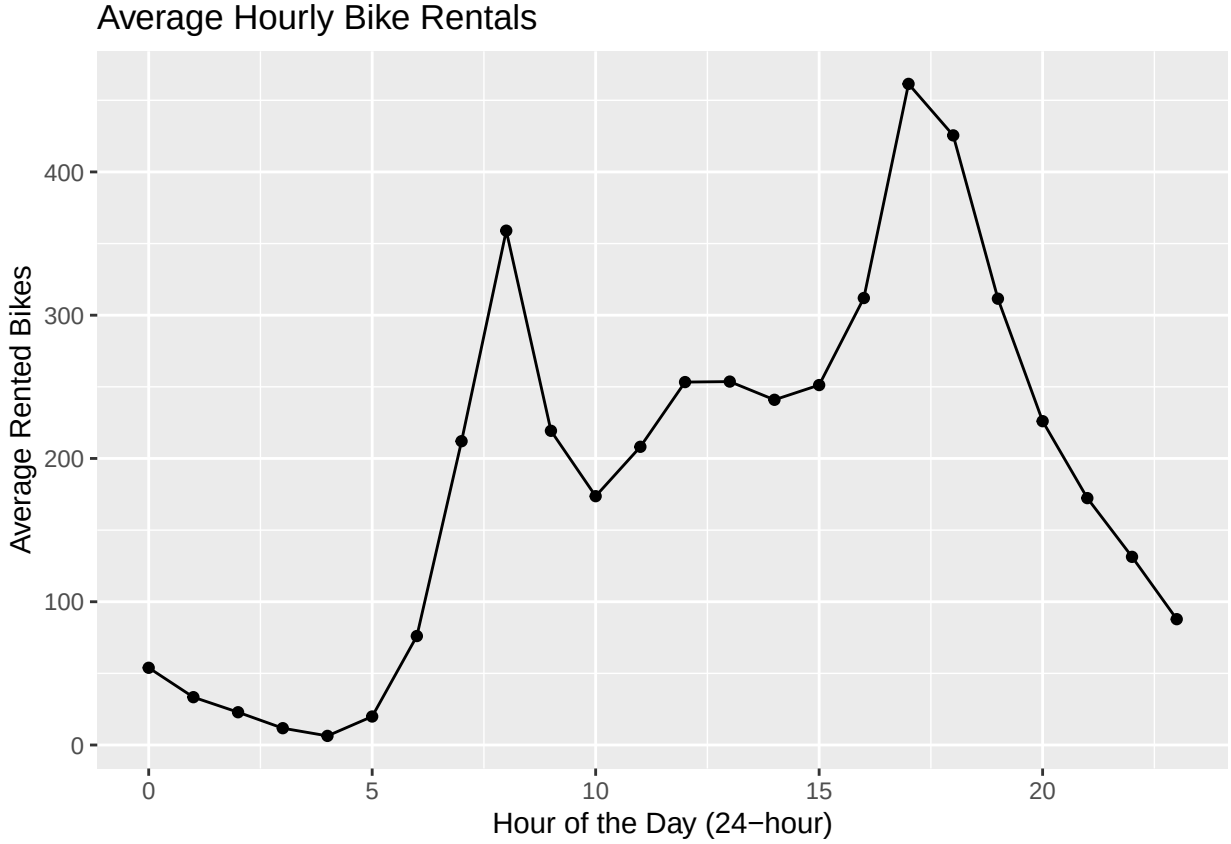
PART D



- This scatterpoint visualizes the relationship between professors' physical attractiveness (x-axis) and their course evaluation scores (y-axis).
- The correlation was 0.19, which suggests while there is some evidence of association, other factors would play a more important role for determining evaluation scores.

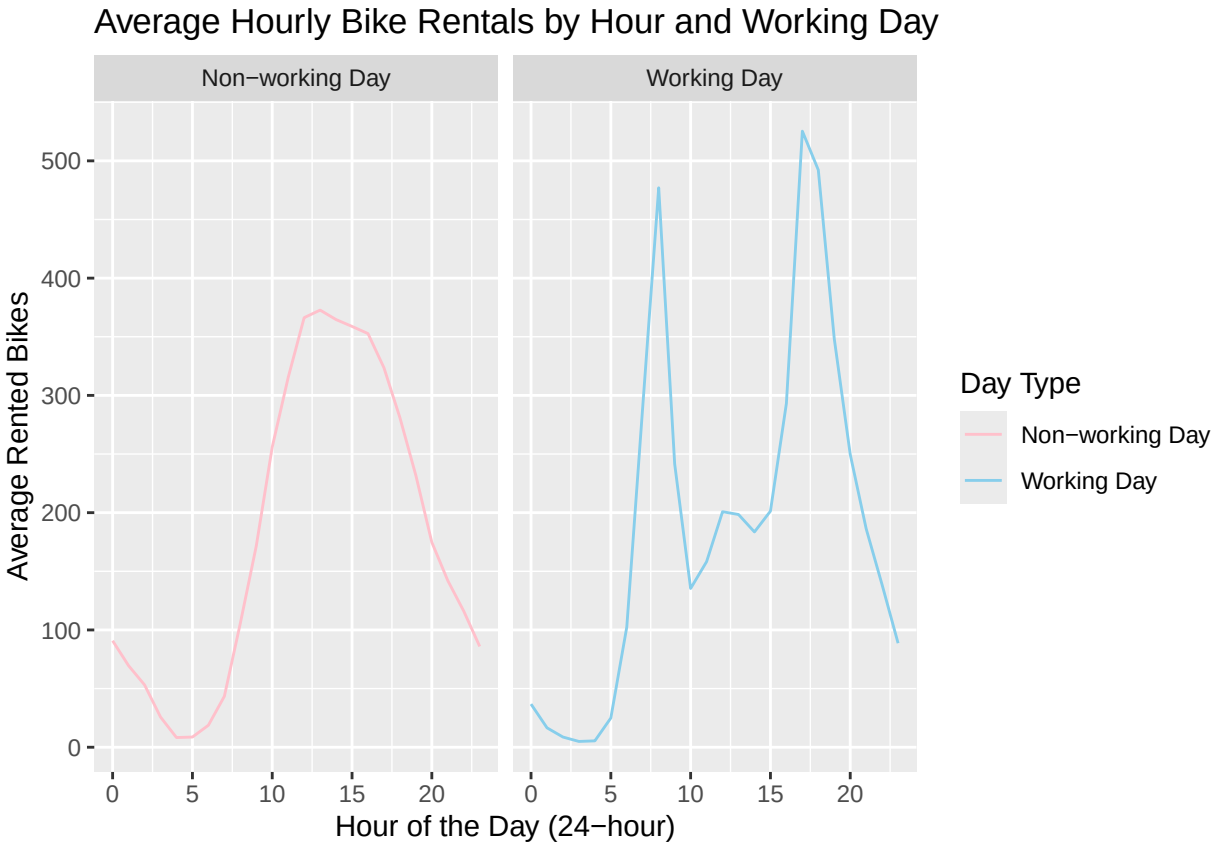
PROBLEM 2: bike sharing

PLOT A



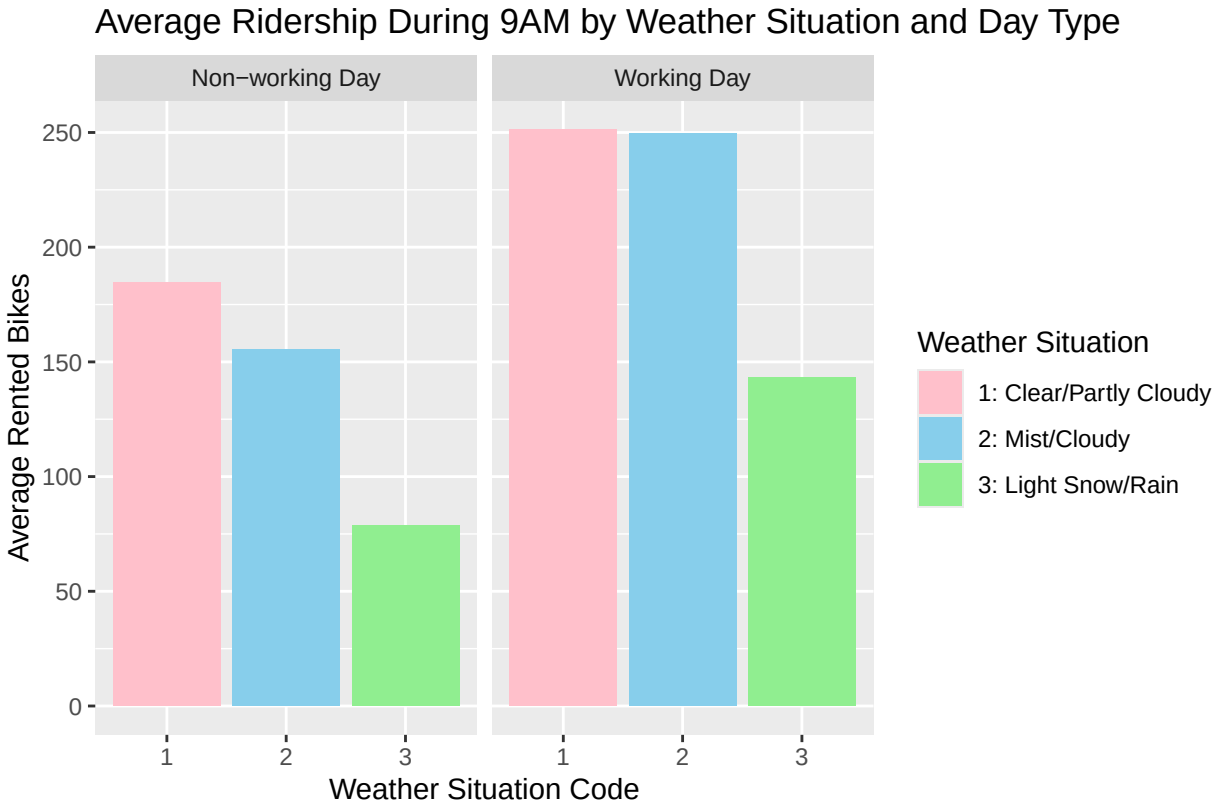
- The line graph demonstrates the average number of bike rentals (y-axis) per hour for the day in 24-hour format (x-axis), where it clearly shows the pattern for bike rentals.
- The graph peaks around 7-8AM and 16-17 or 4-5PM, with relatively low averages around late at night and in the early morning.

PLOT B



- The faceted line graph shows the average bike rentals for the day (24 hours), separated by working day and non-working day. The x-axis shows the hour of the day (24 hours) and y axis shows the average number of rented bikes.
- On working days, people tend to rent bikes more around 6-7AM and 4-5PM while on non-working days, the graph is more steadily increase and peaks around 1-3PM and declines afterwards.

PLOT C

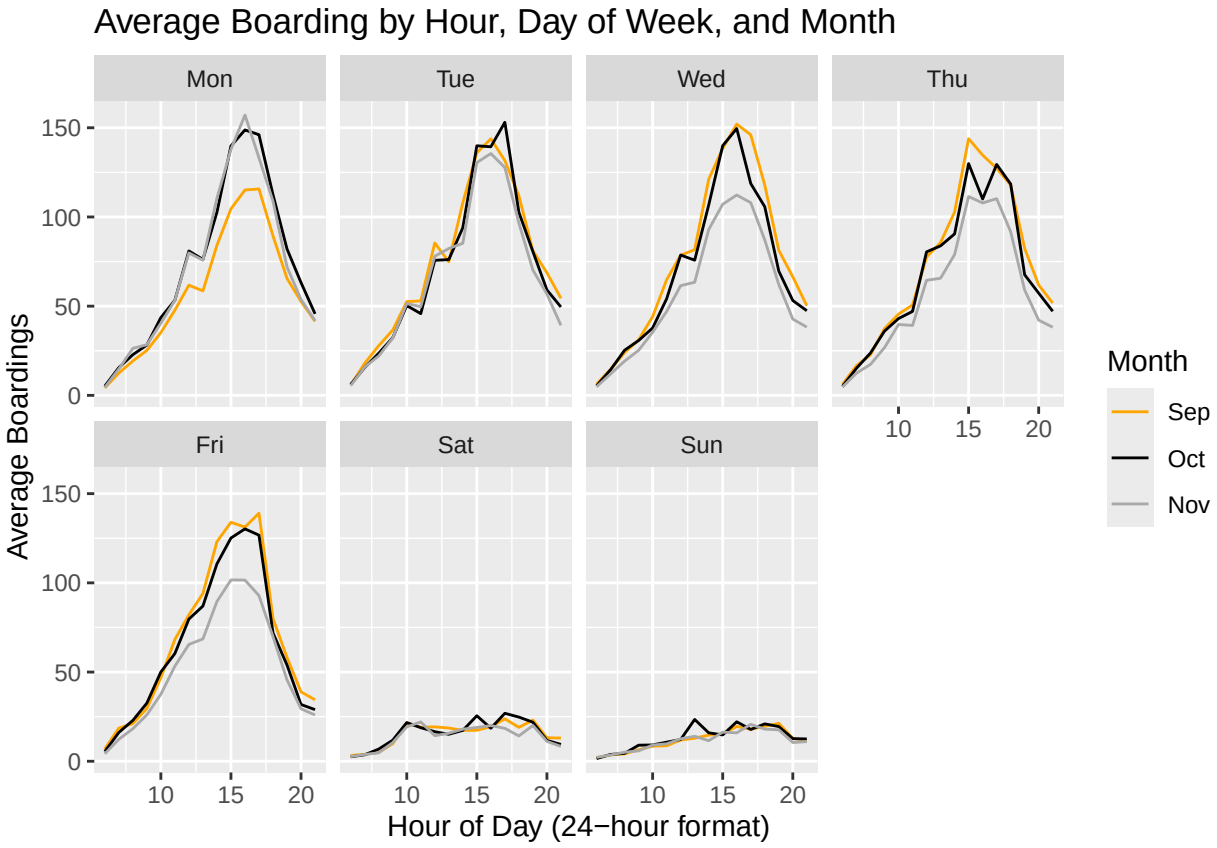


*weathersit code 4 (heavy rain/snow) is not found for filtered dataset.

- This bar plot demonstrates the average rented bikes (y-axis) at 9AM by weather conditions (x-axis) and by day type (working day or non-working day).
- It seems the ridership is higher during clear or party cloudy (code 1) weather for both day types, while ridership drops by a lot when there's light snow or rain (code 4), showing sharper drop especially on non-working days.

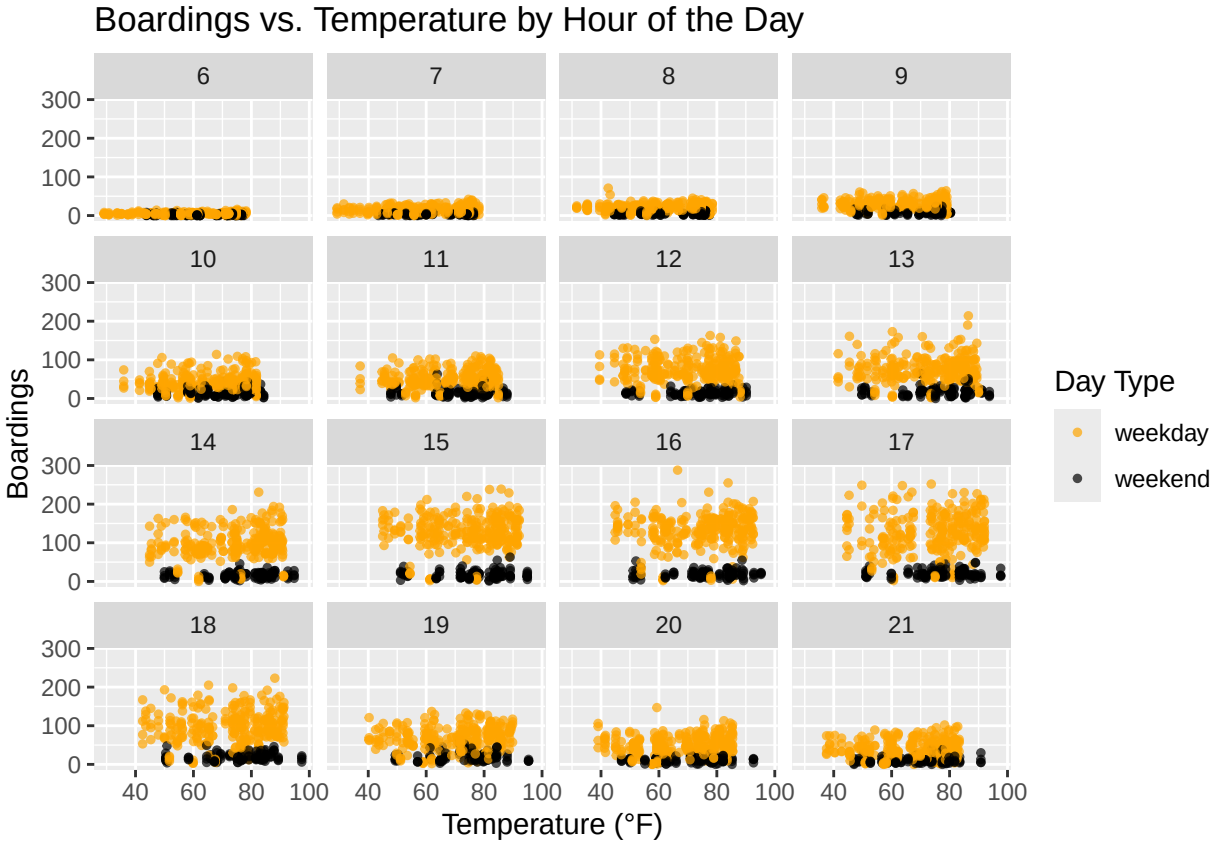
PROBLEM 3: Capital Metro UT Ridership

1ST PlOT



- This line graph demonstrates the average number of boardings (y-axis) on Capital Metro by hour of the day (x-axis) in 24-hour format, separated by day of week and months.
- It seems that boardings peak during 3-5PM for the weekdays and the graph looks much steadier during weekends. Distinctly, average boardings on Mondays in September look lower compared to other days and months, possibly because it overlaps with the beginning of school semester and the boarding patterns are still adjusting for most students. Also the average boardings on Wednesdays, Thursdays, and Fridays in November are relatively lower possibly due to the Thanksgiving break where students' use for Cap Metro decreases.

2ND PLOT



- This scatterplot demonstrates the relationship between the temperature (x-axis) and the number of boardings (y-axis) separated by hour of the day, where orange represents weekday and black represents weekend.
- The scatterplot seems to remain constant for almost every single hours during the day (there is no significant peak or drop), it seems like that temperature is relatively not a significant factor for the Cap Metro Riders, who are mostly students who are considered to be mandatory commuters.

PROBLEM 4: Wrangling the Billboard Top 100

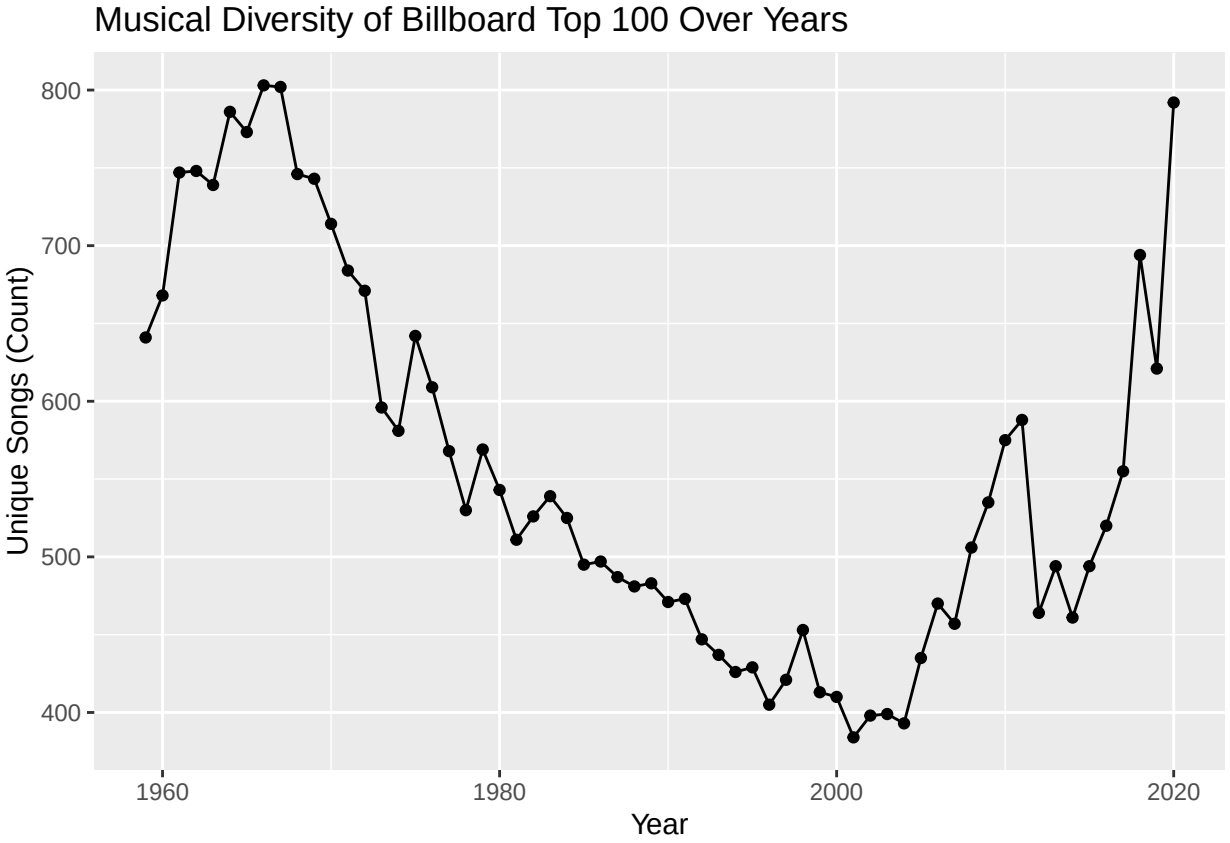
PART A

Table 1: Top 10 most popular songs since 1958

performer	song	count
Imagine Dragons	Radioactive	3828
AWOLNATION	Sail	3160
Jason Mraz	I'm Yours	2926
The Weeknd	Blinding Lights	2926
LeAnn Rimes	How Do I Live	2415
LMFAO Featuring Lauren Bennett & GoonRock	Party Rock Anthem	2346
OneRepublic	Counting Stars	2346
Adele	Rolling In The Deep	2145
Jewel	Foolish Games/You Were Meant For Me	2145
Carrie Underwood	Before He Cheats	2080

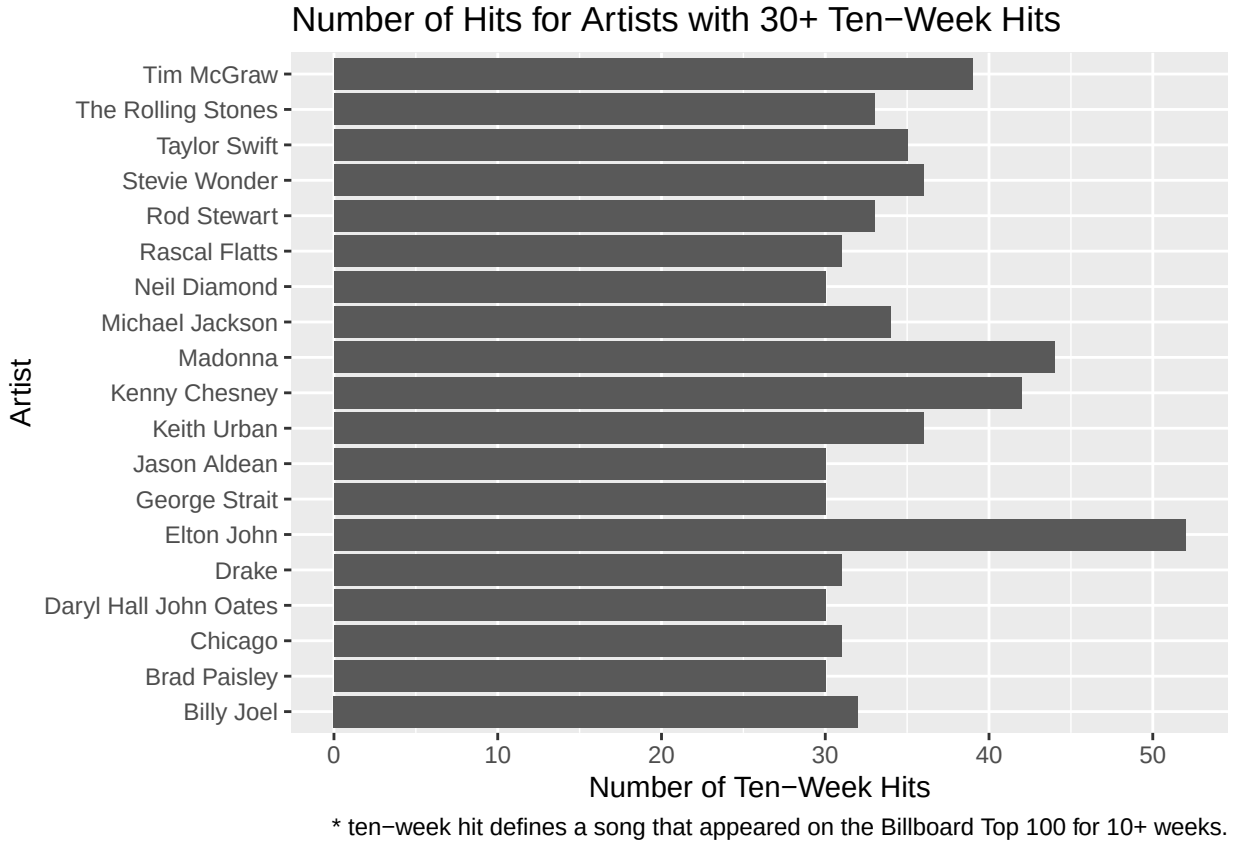
- Above table demonstrates the top 10 most popular songs since 1958 calculated based on the total number of weeks (count column) spent on the Billboard Chart.

PART B



- This line graph demonstrates the total number of unique songs on the Billboard Top 100 Chart each year, which ultimately shows the music diversity over years.
- The trend shows that the music diversity increased until 1967-1968 but started to decrease constantly afterwards until 2000. The music diversity then increased again and is showing the trend of increase until now (2020).

PART C



- This bar chart demonstrates the 19 artists in the U.S. music history with at least 30 ten-week hits, where a 'ten-week hit' is for the songs that stayed on the Billboard Top 100 chart for at least 10 weeks.