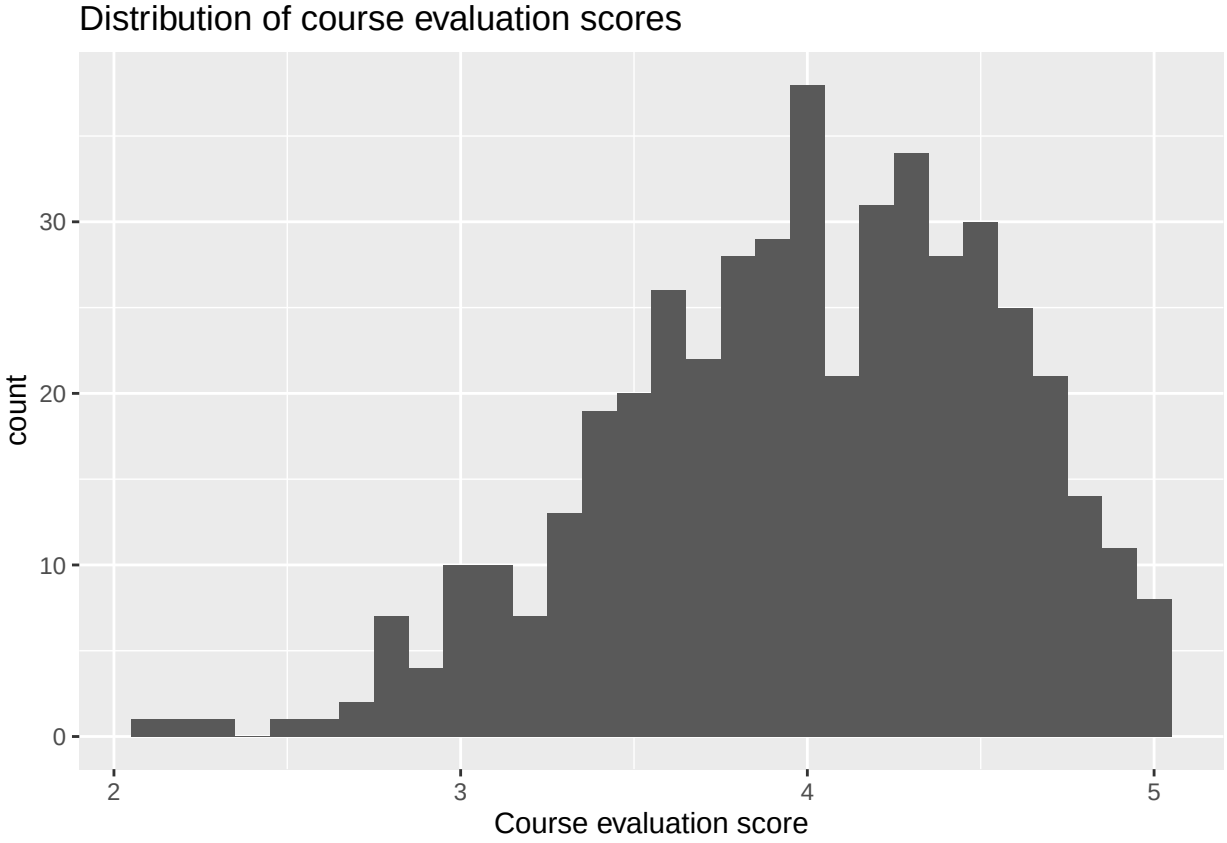
SDS 315 Homework 2

Elizabeth 'Betsy' Sherhart UT EID: eas5778
Click here for link to GitHub repository

January 28, 2025

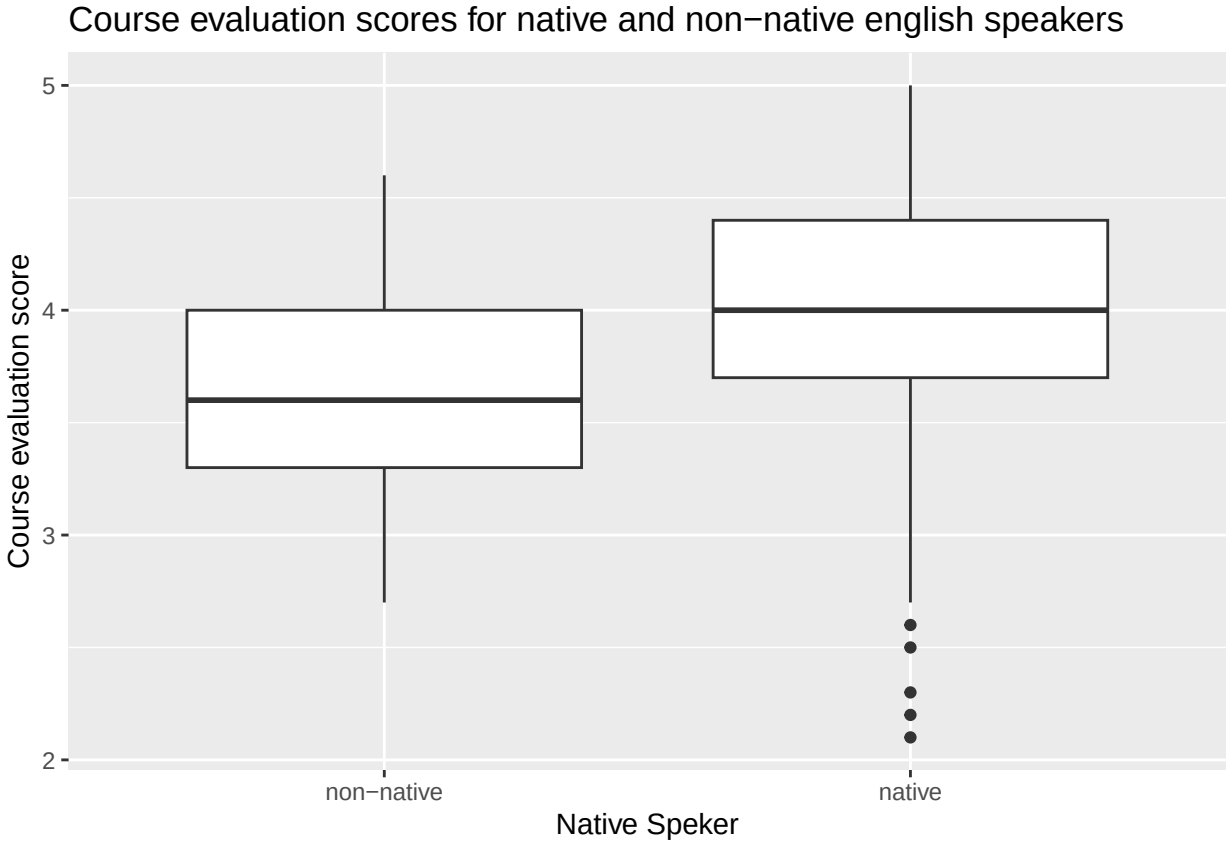
Problem 1: Beauty, or not, in the classroom

Part A



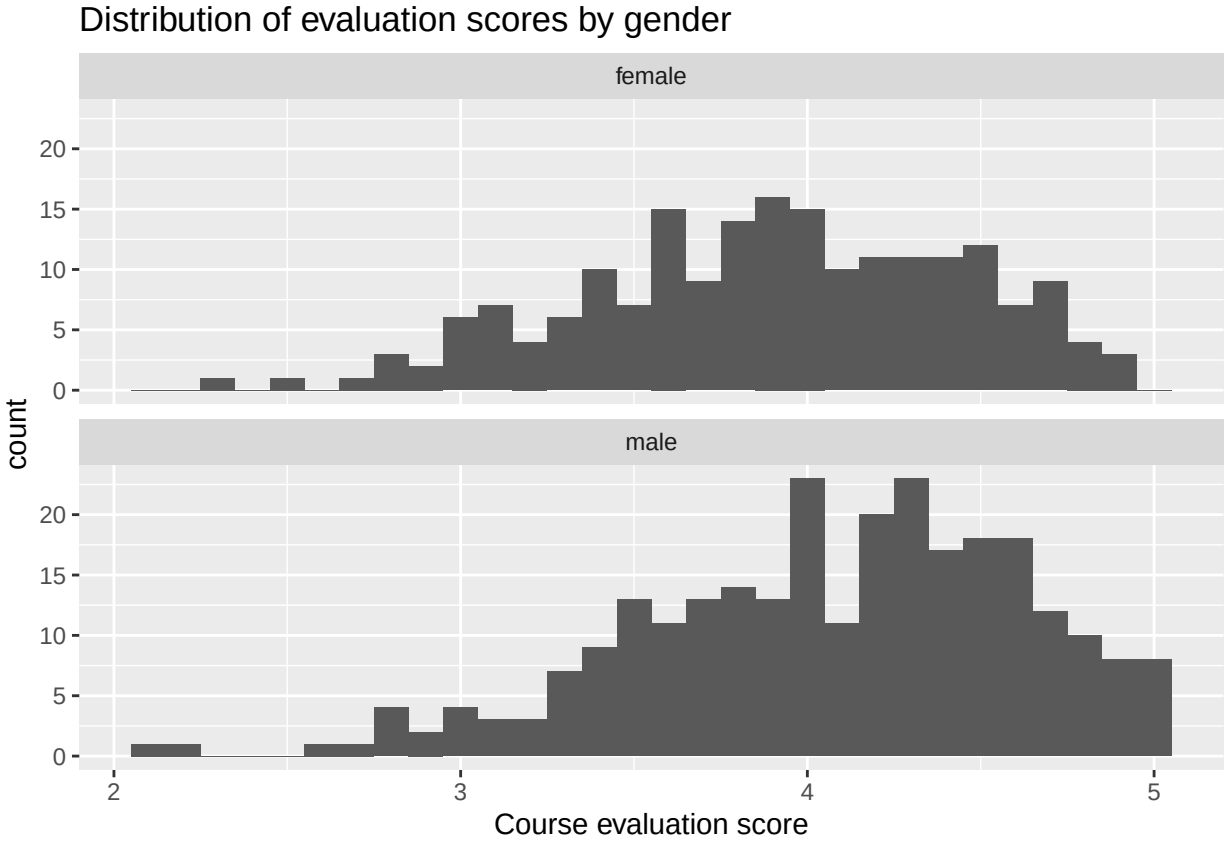
The histogram shows the distribution of course evaluation scores for a sample of 463 UT Austin courses. The distribution reveals that even though students had an option to rank courses on a scale from 1 to 5, the lowest evaluation score given was 2.1 and the highest was 5. The histogram is skewed towards the left slightly, meaning the median, 4, and IQR, 0.8, are used to summarize the data.

Part B



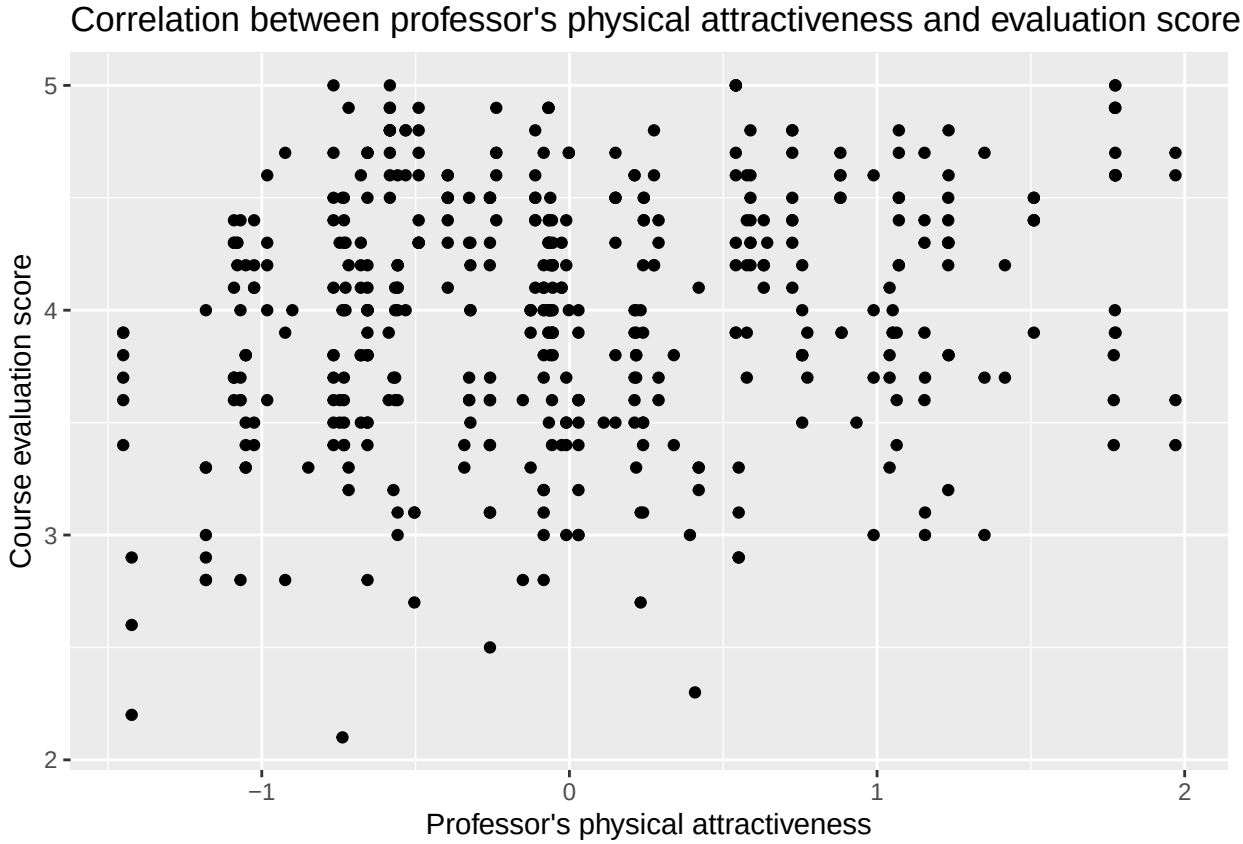
The side-by-side boxplot compares the distribution of course evaluation scores for a sample of 463 UT Austin courses by whether the professor is a native English speaker or not. The boxplot reveals that non-native speakers have a lower average evaluation score than native speakers, but both non-native and native speakers have a similar spread between their 25th and 75th percentiles. However, native speaker professors have a greater spread of lower scores than non-native speakers.

Part C



The histogram shows the distribution of course evaluation scores for a sample of 463 UT Austin courses faceted by the gender of the professor. Both genders have a similar distribution of scores that are slightly skewed towards the left. However, the male professors have a greater spread of evaluation scores for both lower and higher scores.

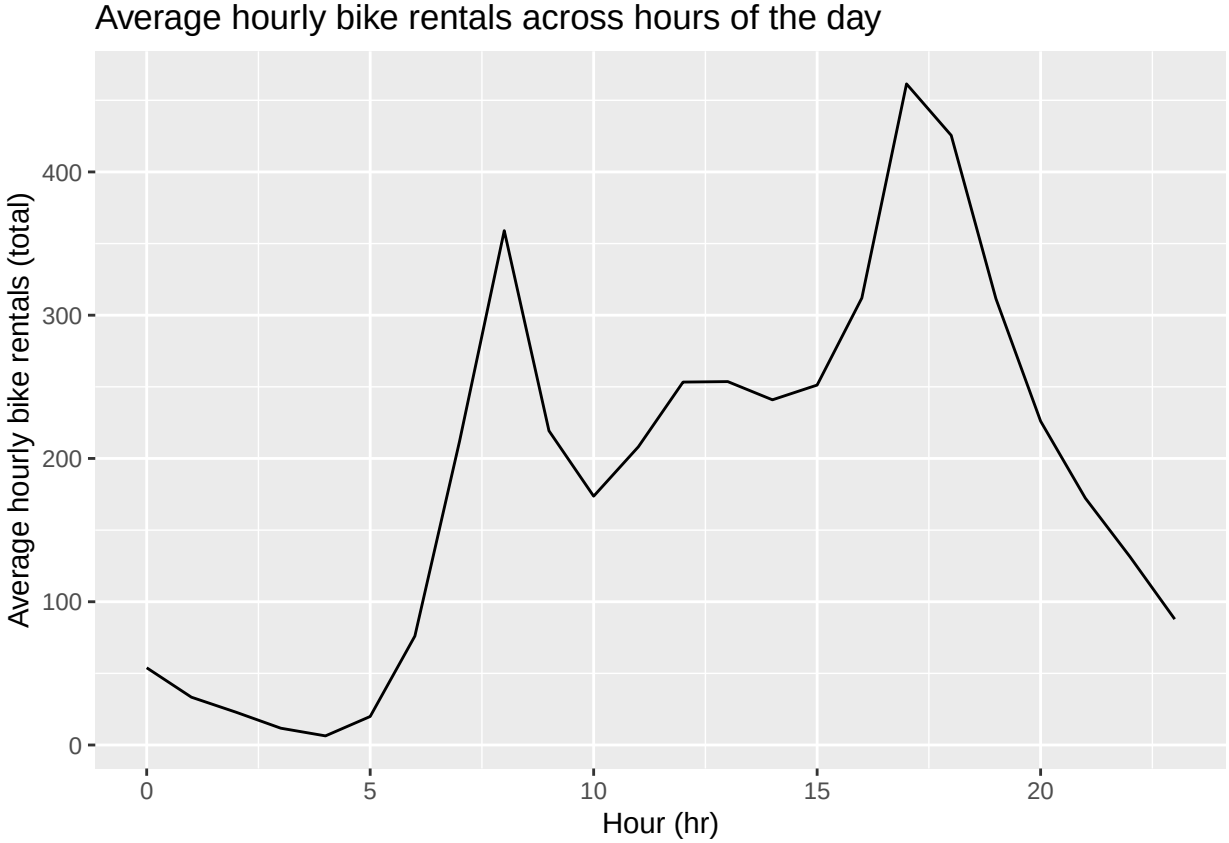
Part D



The scatter plot visualizes the extent to which there may be an association between a professor's physical attractiveness and their course evaluation scores. The professor's physical attractiveness was determined by six student panelists (3 males, 3 males) who were shown pictures and asked to rank their attractiveness. Their rankings were averaged, and the average was shifted to have a mean of zero. This means a positive ranking means an above average physical attractiveness ranking and a negative ranking means a below average physical attractiveness ranking. The correlation coefficient is 0.1890391 indicating that there is a weak positive relationship between physical attractiveness and evaluation scores.

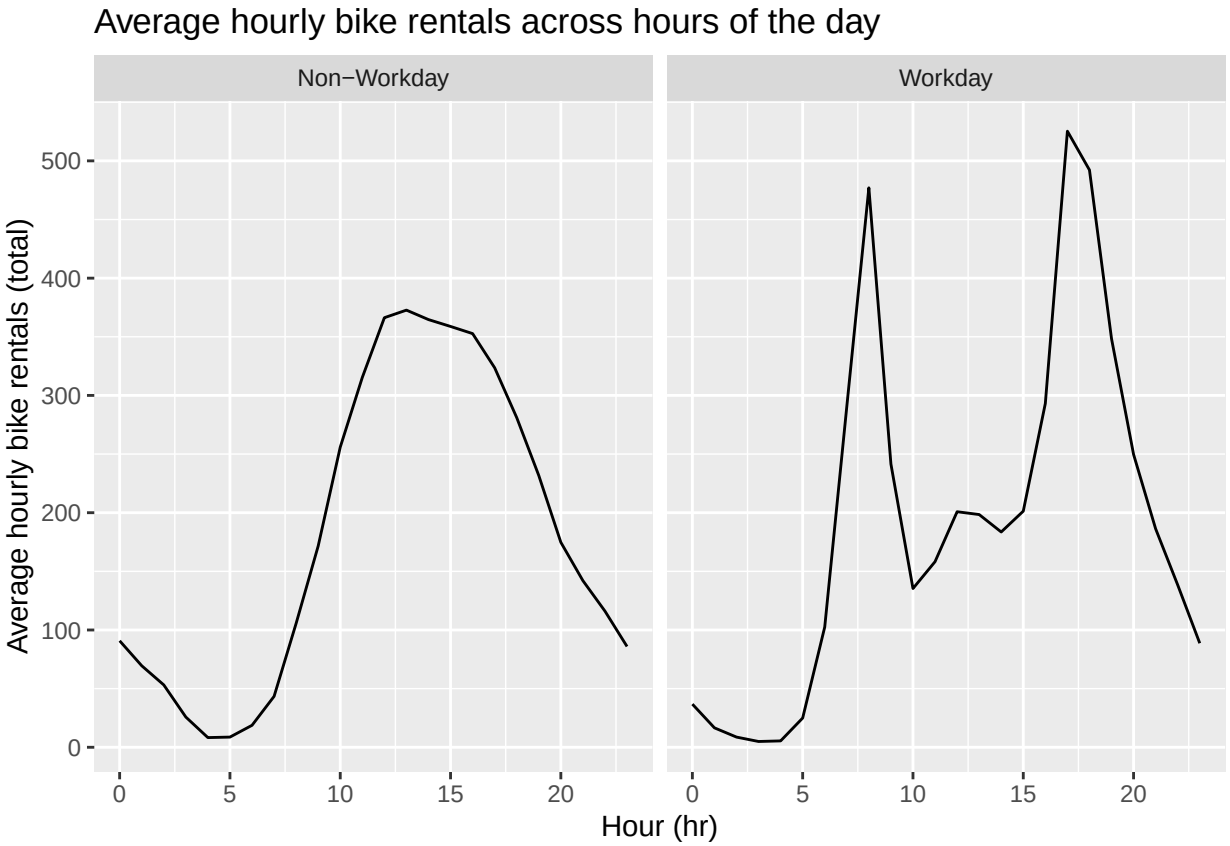
Problem 2: bike sharing

Part A



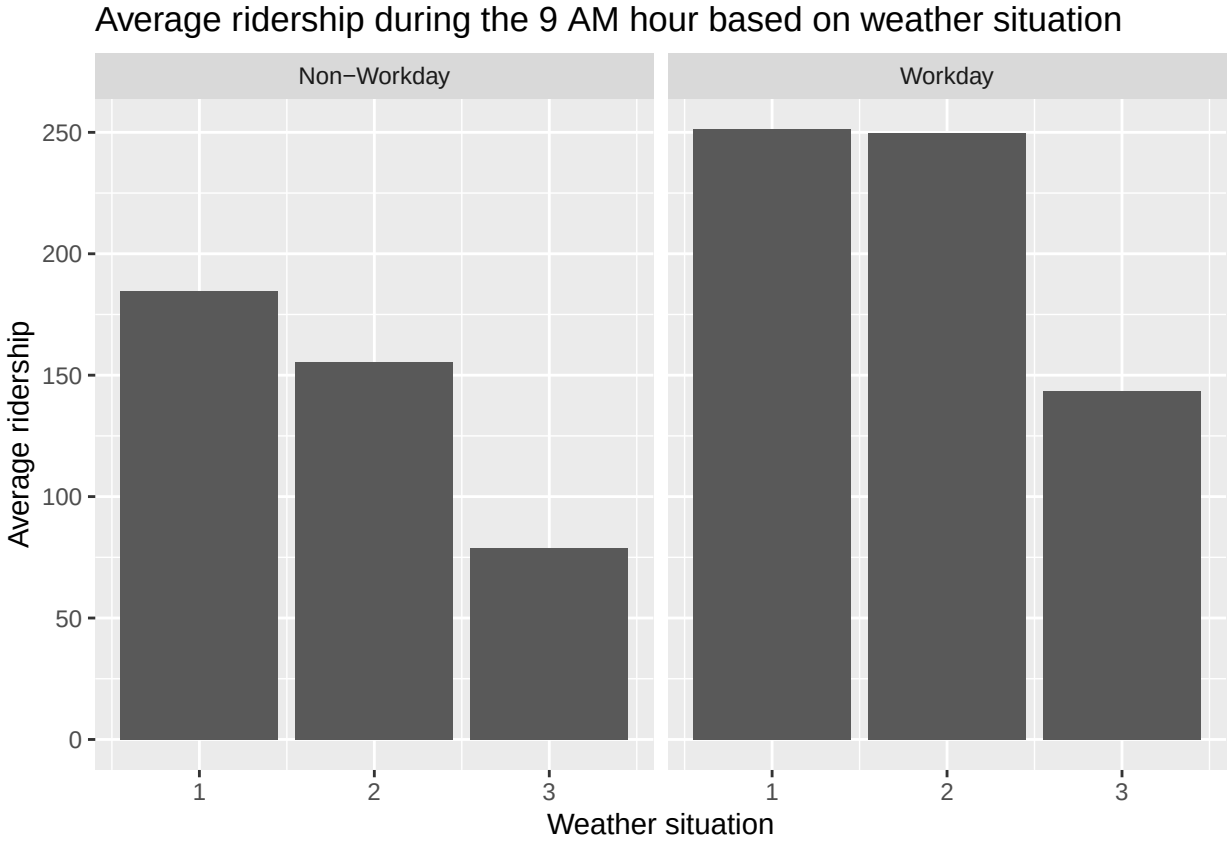
The line graph shows the average hourly bike rentals across the hours of the day from 0 = midnight or 12 AM and 23 = 11 PM for the Capital Bikeshare system in Washington D.C. from 2011 to 2012. The line graph shows spikes in average hourly bike rentals during rush hours, such as 7-9 AM and 4-8 PM, and a slight spike during lunchtime. The graph reflects low bike rentals between the late night and early hours of morning. In conclusion, the line graph reflects the high demand for bikes to commute to and from work and low demand for bikes during the night.

Part B



The line graph shows the average hourly bike rentals across the hours of the day faceted by whether it is a working or non-working day for the Capital Bikeshare system in Washington D.C. from 2011 to 2012. The line graph for the average bike rentals for a non-workday shows a broader plateau of demand in the middle of the day and notably higher demand during the late and early hours. The line graph of the average bike rentals for a workday shows spikes in average hourly bike rentals during the morning and evening rush hours. In conclusion, the faceted line graph reflects the high demand for bikes to commute to and from work and low demand for bikes during the night on workdays and the more constant demand during the day and higher rentals during the wee hours on non-workdays.

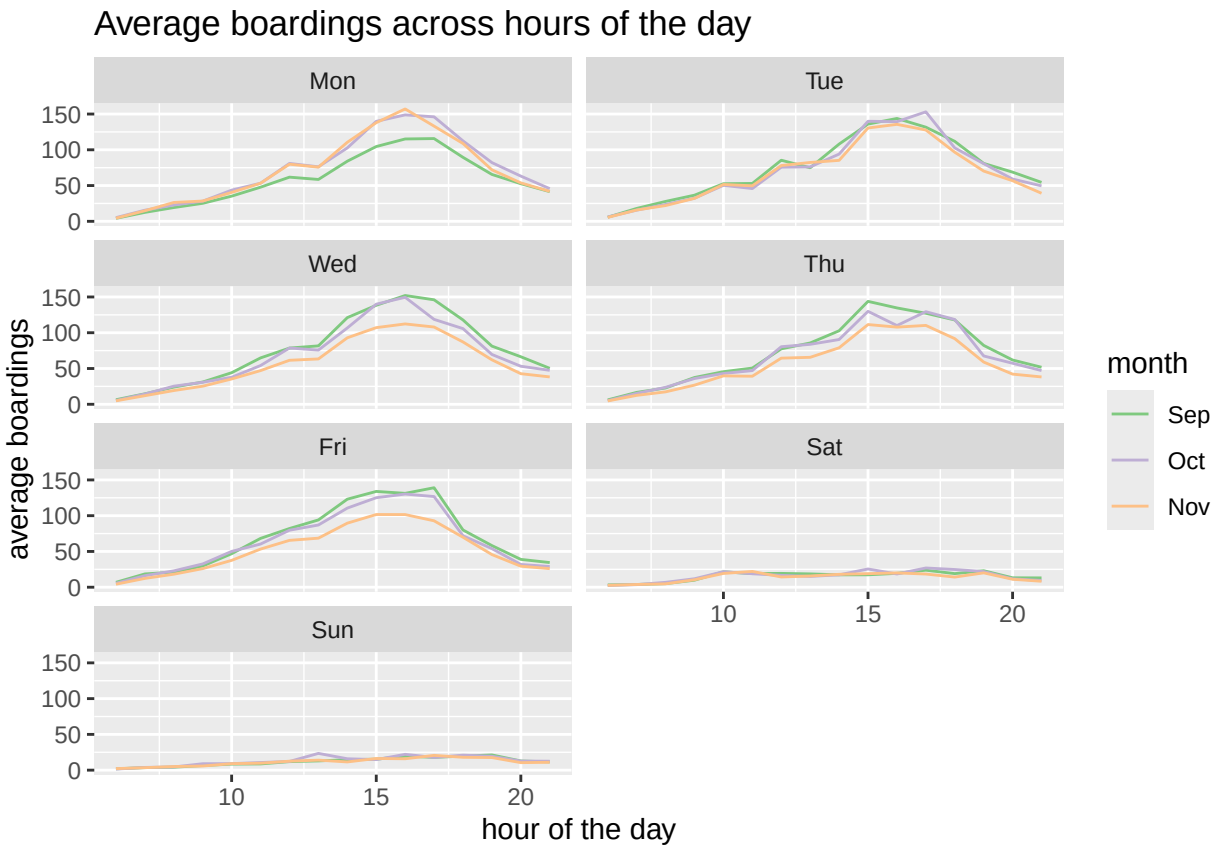
Part C



The faceted bar plot shows the average ridership for the 9 AM hour depending on the weather situation at the time faceted by whether it is a workday or non-workday for the Capital Bikeshare system in Washington D.C. from 2011 to 2012. The weather situation was classified by the numbers 1, 2, 3 and 4 originally. The weather was classified as 1 if it was clear, few clouds, or partly cloudy, 2 if there was mist and clouds, 3 if there was light rain with scattered clouds including or excluding a thunderstorm, and 4 if there was light snow, ice pallets, thunderstorm, or fog. It is important to note that there were only recordings of levels 1, 2, and 3 for the 9 AM hour, so weather situation 4 was excluded from the graph. The bar plot reveals that non-workday average ridership is slightly more affected by weather than workday average ridership at 9 AM. However, the proportion of the bar graph heights are relatively consistent between workdays and non-workdays because of changes in ridership due to weather. The 9 AM hour is a peak morning rush hour, so there is more demand for bikes on work days at 9 AM, but the change in average ridership is similar for workdays and non-workdays. There is a small decrease in ridership from level 1 to level 2 and a larger decrease from level 2 to level 3. This means that the weather has similar affects on average ridership for all days, with the decreases in ridership because of the weather being only slightly more extreme for non-workdays.

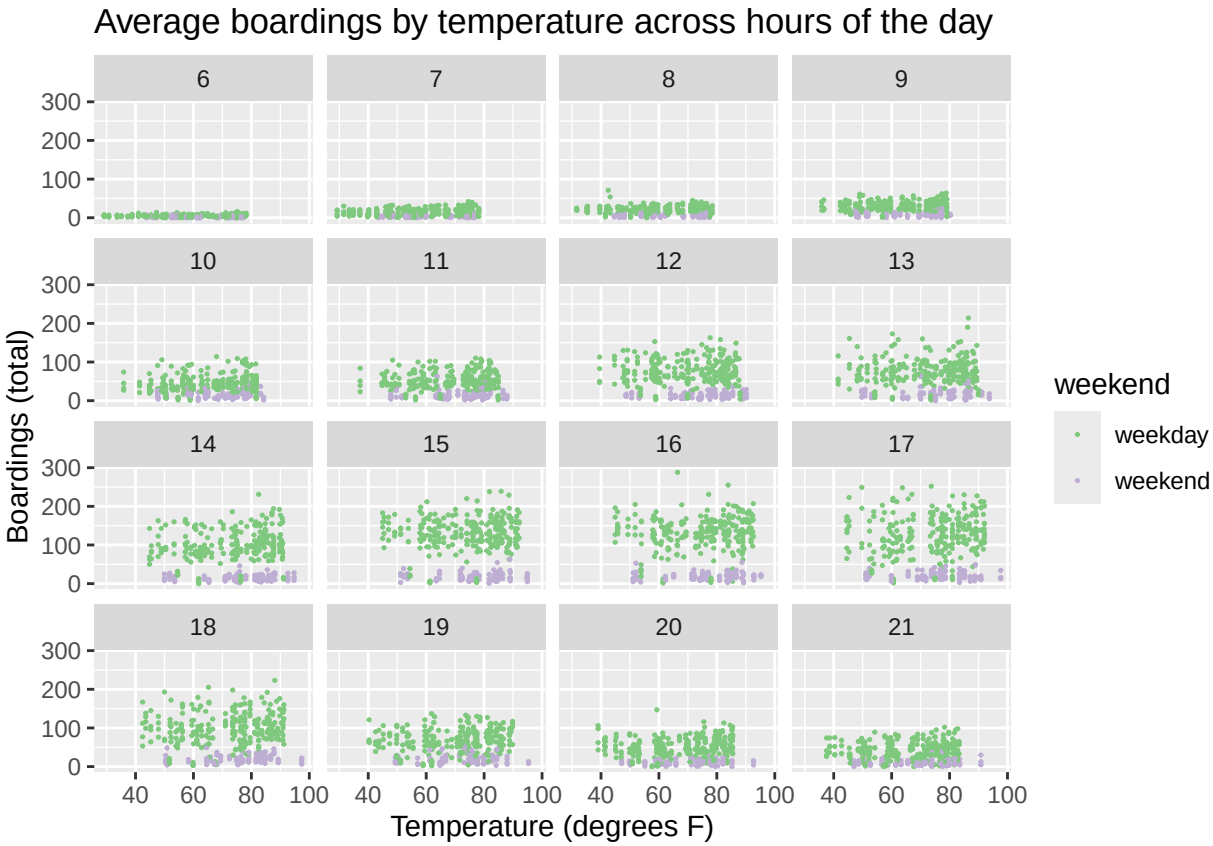
Problem 3: Capital Metro UT Ridership

Question 1



The faceted line graph displays average boarding numbers across the hours of the day for the months of September, October, and November during 2018, separated by the day of the week for UT Capital Metro buses to, from, and around the UT campus. The general trends through the figure are that weekends have the least average boardings, indicating low demand on the weekend when there are no classes, and that weekdays have similar shapes/arcs of peak hours of average boardings during the day. For all the weekdays boardings tend to steadily increase throughout the morning towards a peak boarding between 3-5 PM. The peak boardings are broadly similar across weekdays. The average boardings on Mondays in September are lower compared to other days and months because UT has Labor day, which means no classes on the first Monday in September. Similarly, a holiday lowering average boardings is also reflected for Weds/Thurs/Fri in November because of the Thanksgiving holiday, where again there are no classes.

Question 2



The faceted scatter plot displays boarding numbers across the temperature in degrees F faceted by time and colored by whether it is a weekday or weekend for the months of September, October, and November during 2018 for UT Capital Metro buses to, from, and around the UT campus. The general trends to keep in mind are that weekends have a small number of boardings, and that weekday boardings tend to steadily increase throughout the morning towards a peak boarding between 3-5 PM. The scatter plot indicates that despite temperature the boardings steadily increase from morning to peak hours and then back down. When the hour of the day and weekend status are held constant temperature does not seem to have a noticeable effect on those riding the bus, because the points plateau meaning that boarding numbers remain relatively the same across different temperature values.

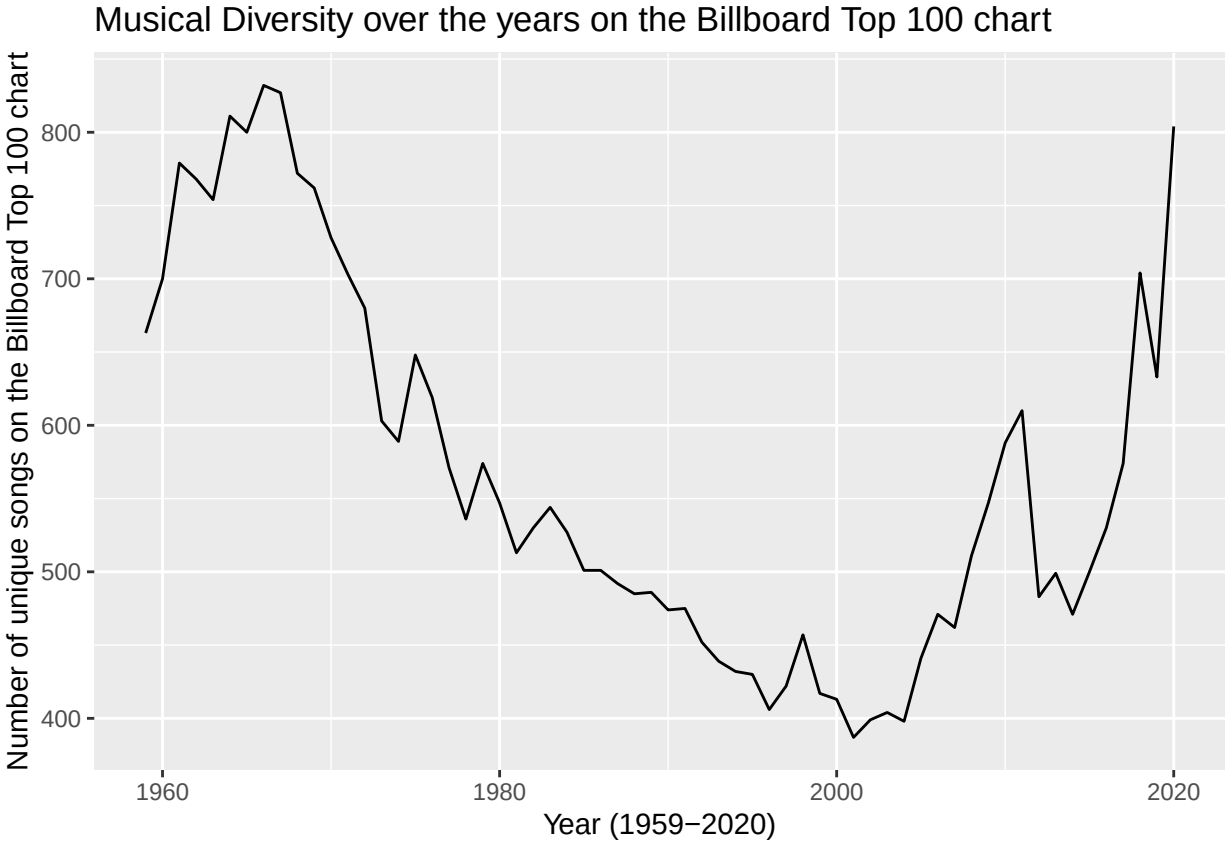
Problem 4: Wrangling the Billboard Top 100

Part A

```
## # A tibble: 10 x 3
## # Groups:   performer [10]
##    performer                                song                          count
##    <chr>                                    <chr>                         <int>
##  1 Imagine Dragons                          Radioactive                      87
##  2 AWOLNATION                               Sail                             79
##  3 Jason Mraz                               I'm Yours                        76
##  4 The Weeknd                               Blinding Lights                  76
##  5 LeAnn Rimes                              How Do I Live                    69
##  6 LMFAO Featuring Lauren Bennett & GoonRock Party Rock Anthem               68
##  7 OneRepublic                              Counting Stars                   68
##  8 Adele                                    Rolling In The Deep              65
##  9 Jewel                                    Foolish Games/You Were Meant~    65
## 10 Carrie Underwood                         Before He Cheats                 64
```

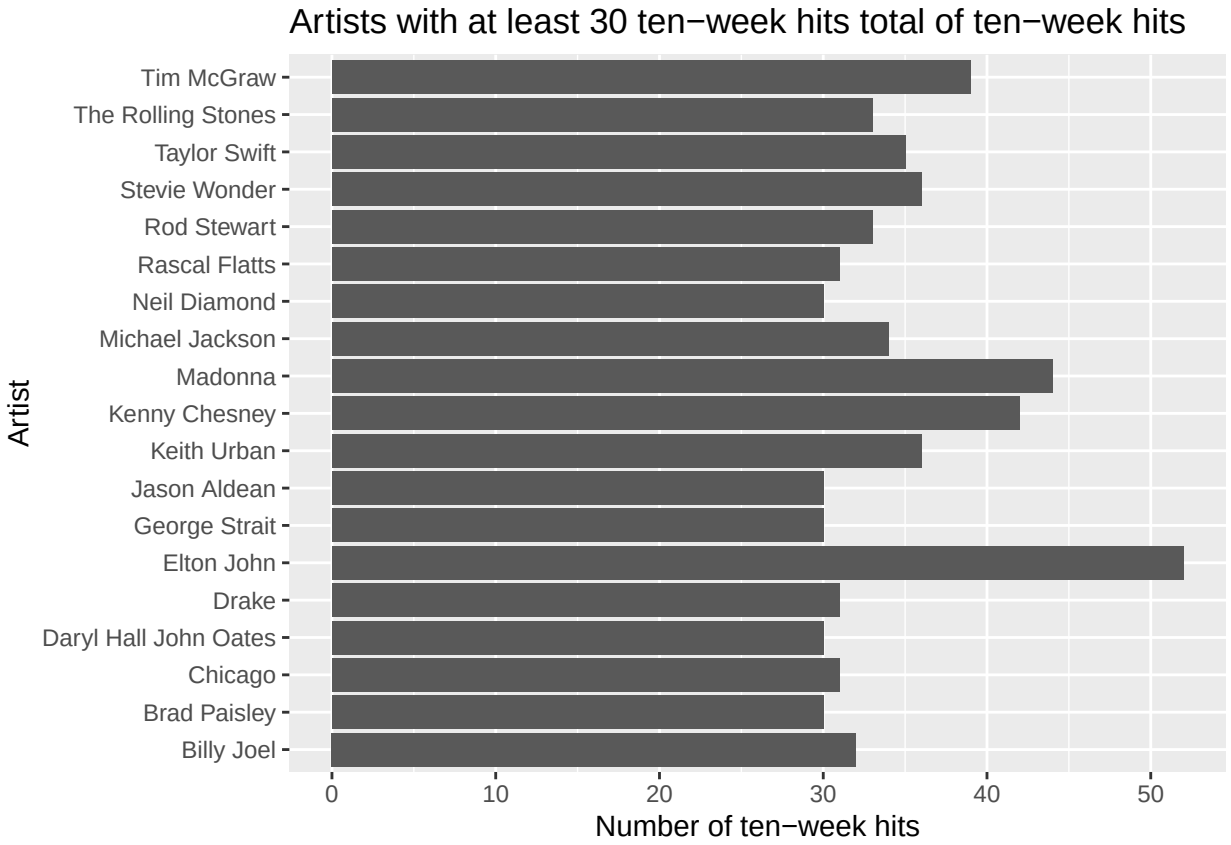
The table above shows the 10 most popular songs since 1958 based on the total number of weeks (count) they spent on the Billboard Top 100. The songs are listed in descending order, with the song that appeared the most weeks on the Billboard Top 100 being first.

Part B



The line graph shows the musical diversity over the years 1959 to 2020 by showing the number of unique songs on the Billboard Top 100 that year. The graph shows musical diversity increasing in the 1960s to over 800 unique songs, then a trend of decreasing from the 1970s to 2000s to under 400 unique songs. From the 2000s, the musical diversity increased back to over 600 unique songs by the 2010s, then decreased to under 500 unique songs in the mid 2010s to a sharp increase back to over 800 unique songs by 2020. The line graph in general shows that musical diversity has fluctuated from 1959 to 2020 between around 850 unique songs to a little less than 400 unique songs.

Part C



The bar plot shows the 19 artists who have had over 30 songs be ten-week hits meaning they appeared on the Billboard Top 100 chart for at least 10 weeks. The bar plot also shows the total number of ten-week hits each artist has had from part of 1958 to part of 2021.