SDS315 Homework 2 Report

Prady Kandi, EID: prk599

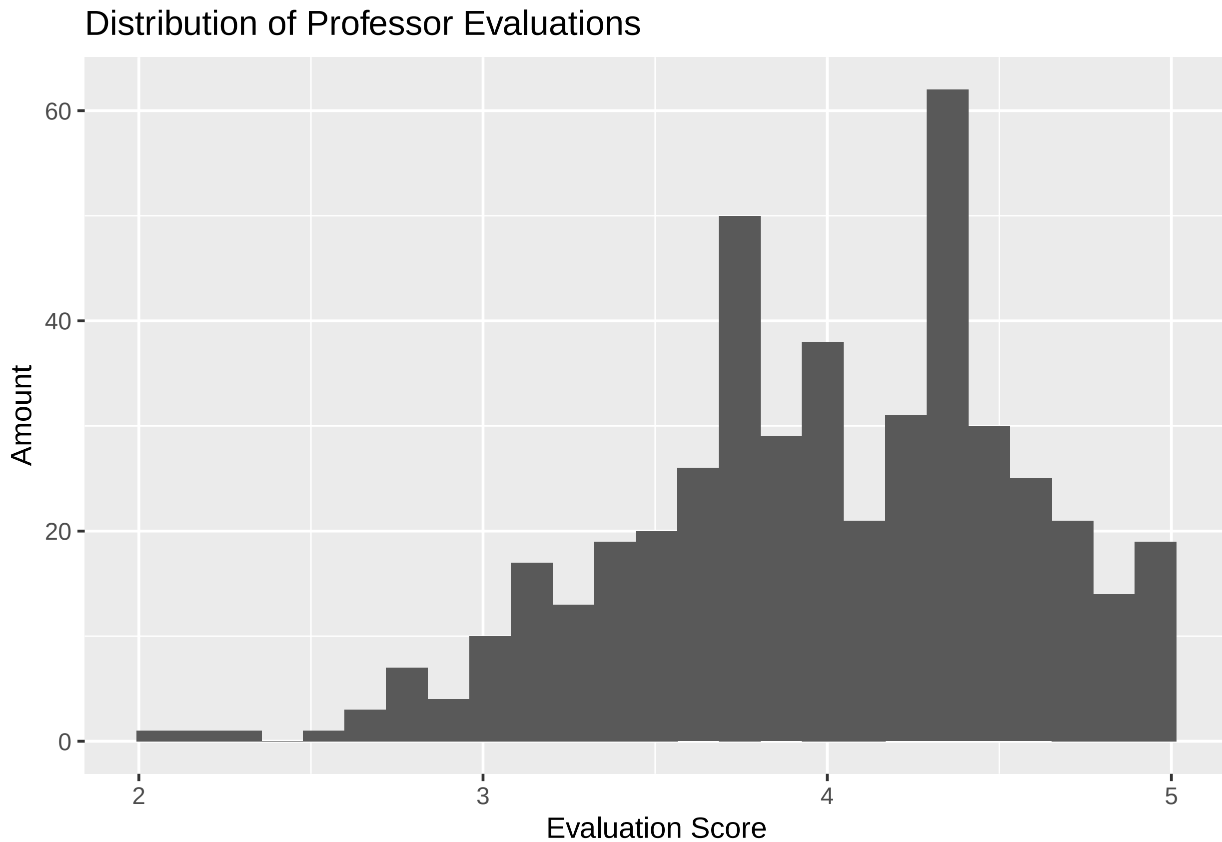
2025-01-28

Contents

The link to the Github repo containing the R file can be found here.

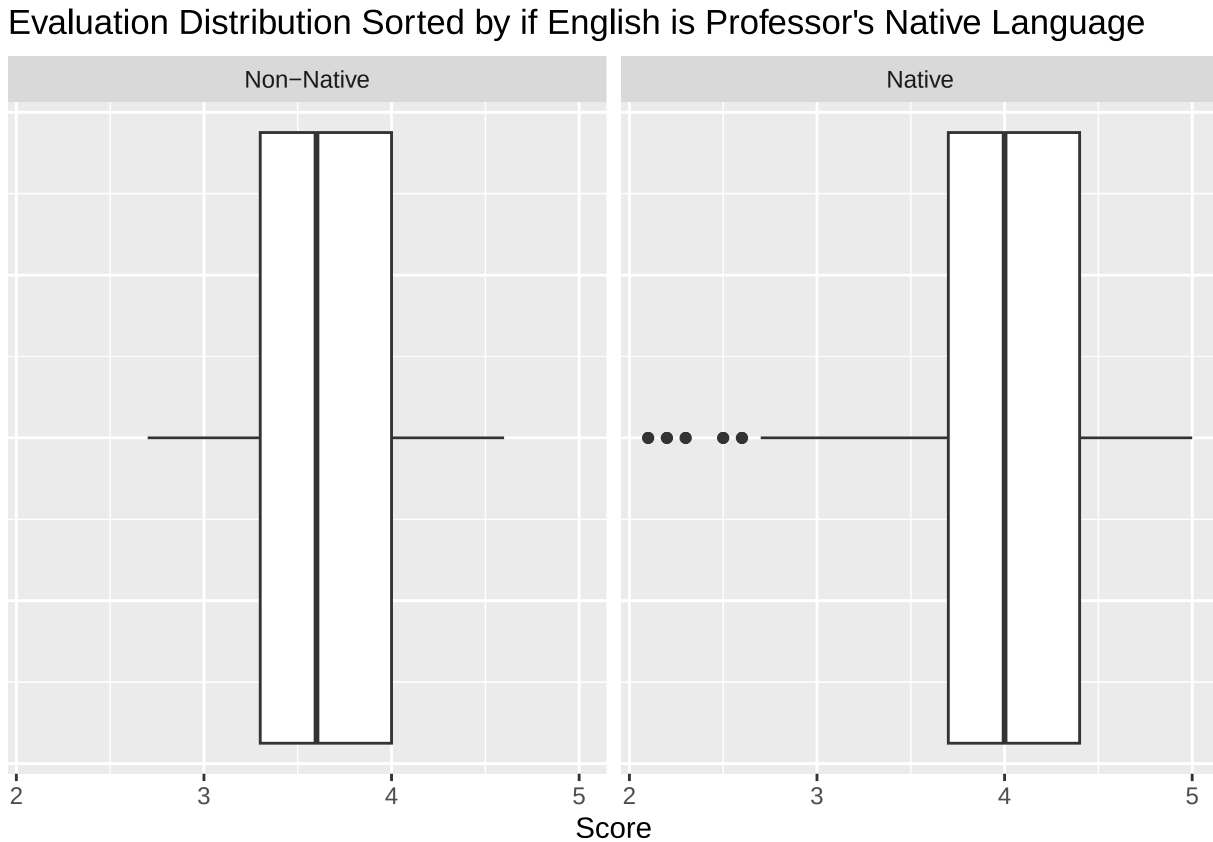
Problem 1

Part A



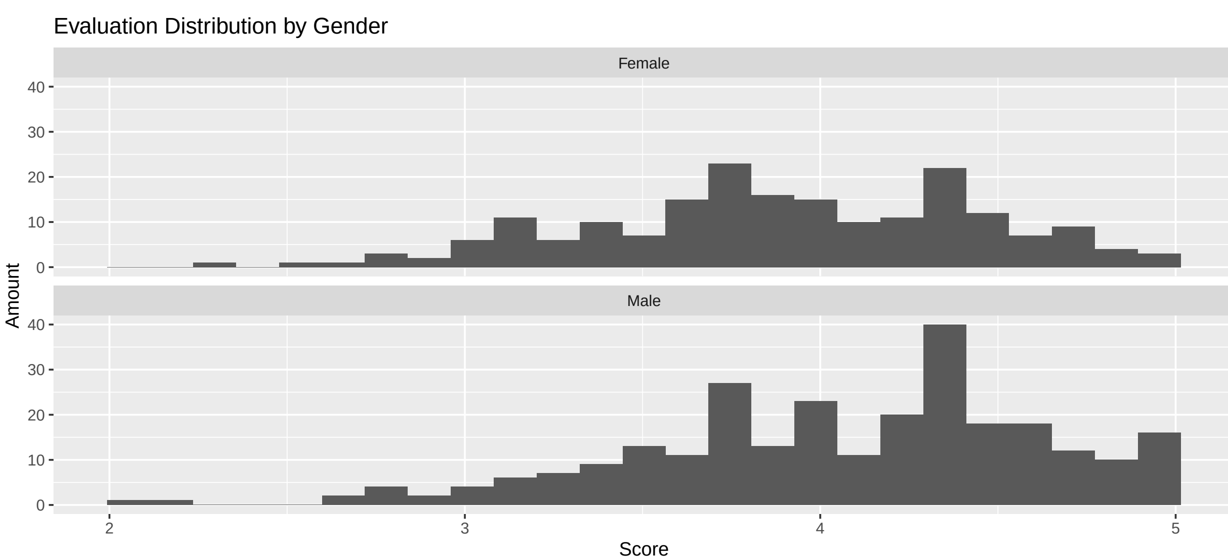
This histogram shows that course evaluation scores follow a roughly binomial distribution, with the histogram being left-skewed. The average course evaluation score appears to be around 4, with the most common scores being 3.8 and 4.3.

Part B



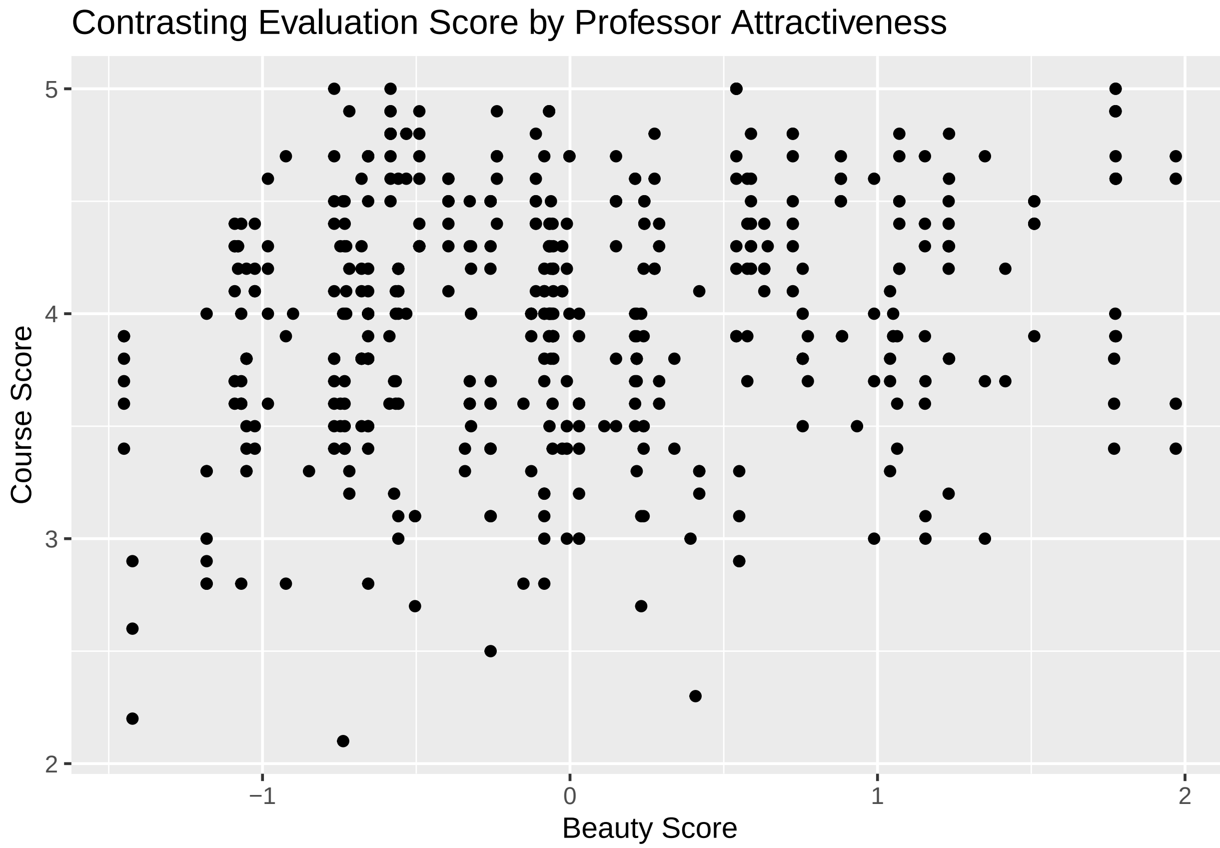
Native English speakers have a higher average score at 4.0, while non-native speakers have a lower average score around 3.6. This is likely indicitave of either inherent bias among students or widespread difficulty in understanding non-native English speakers.

Part C



Female professors have a more uniform distribution than male professors, with around the same average score. However, male professors appear to have more extreme scores, both at the high and low end.

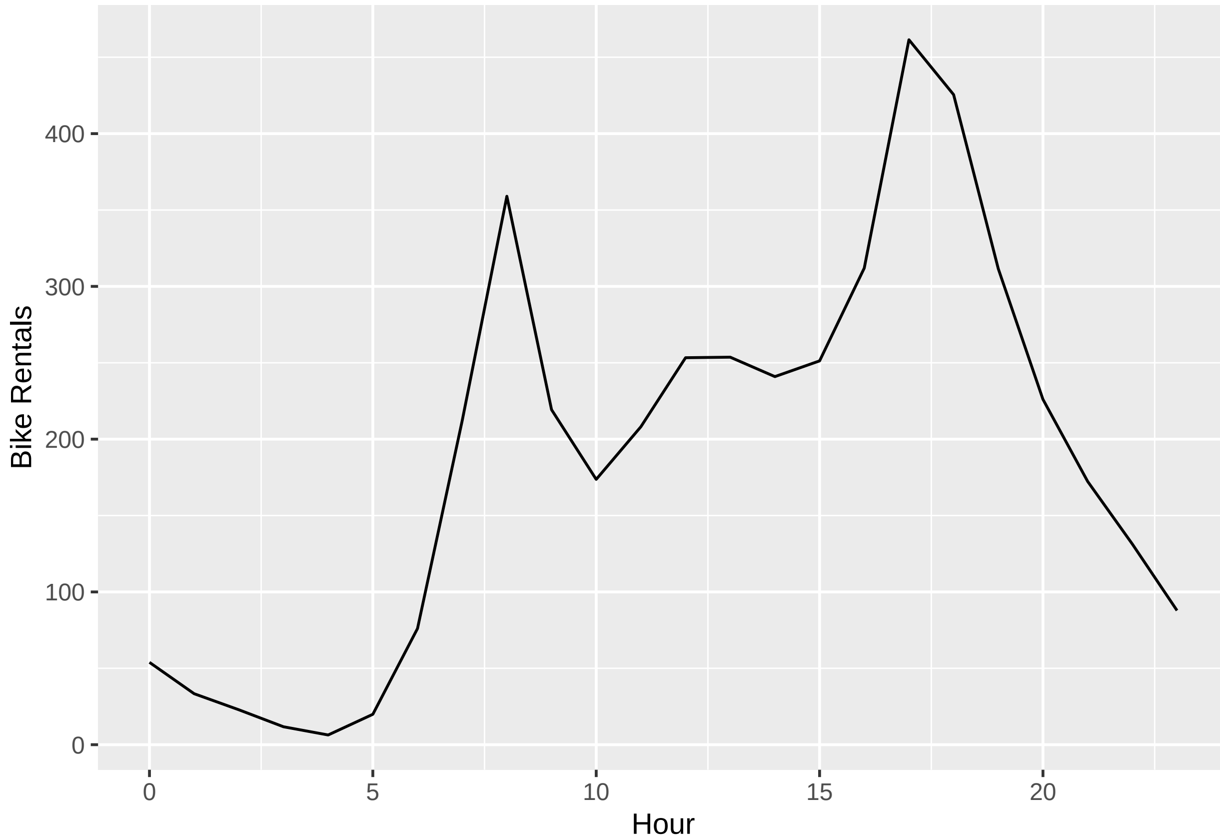
Part D



An association is not immediately clear, but rather, a professor's beauty score seems to serve as a limiter, with professors in certain ranges of beauty being limited to certain scores. With this information, it can be deducted there is a slight correlation, although the majority of a professor's score is not determined by their beauty.

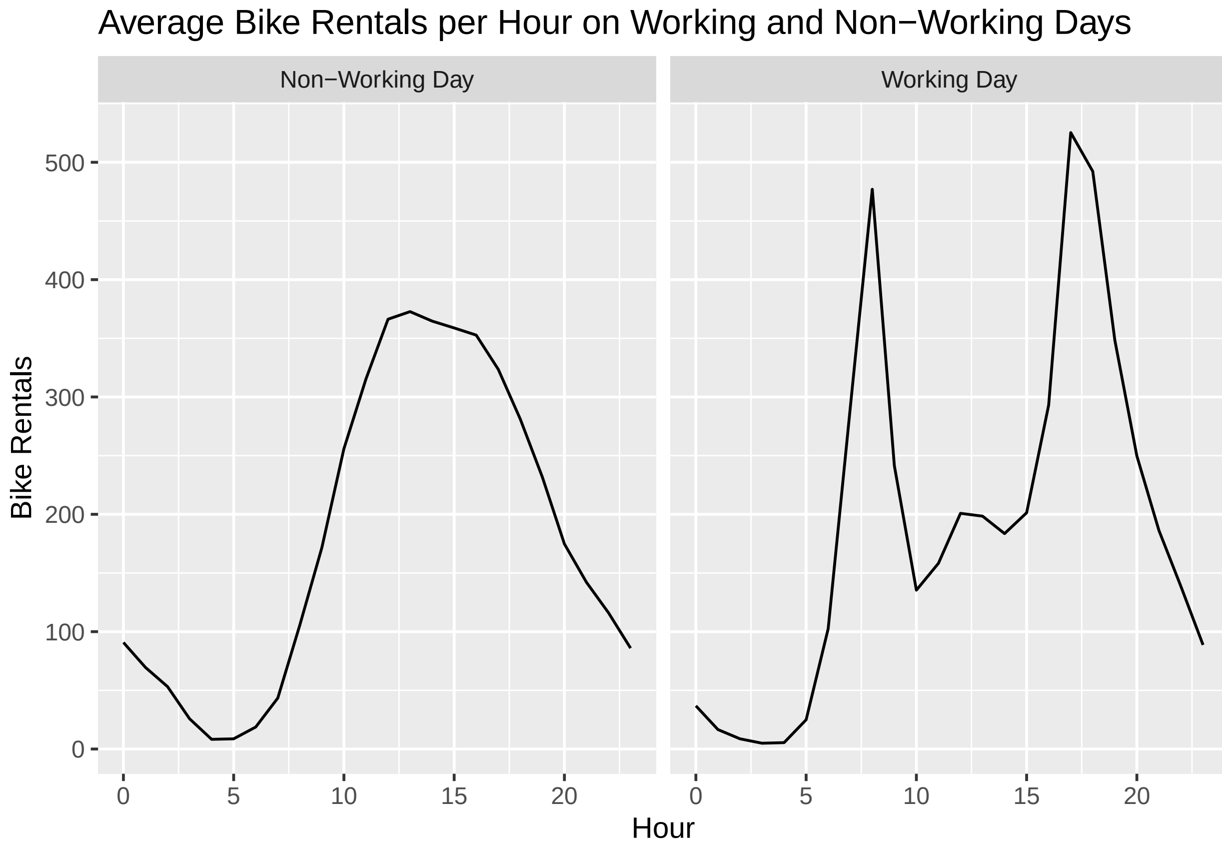
Problem 2

Part A



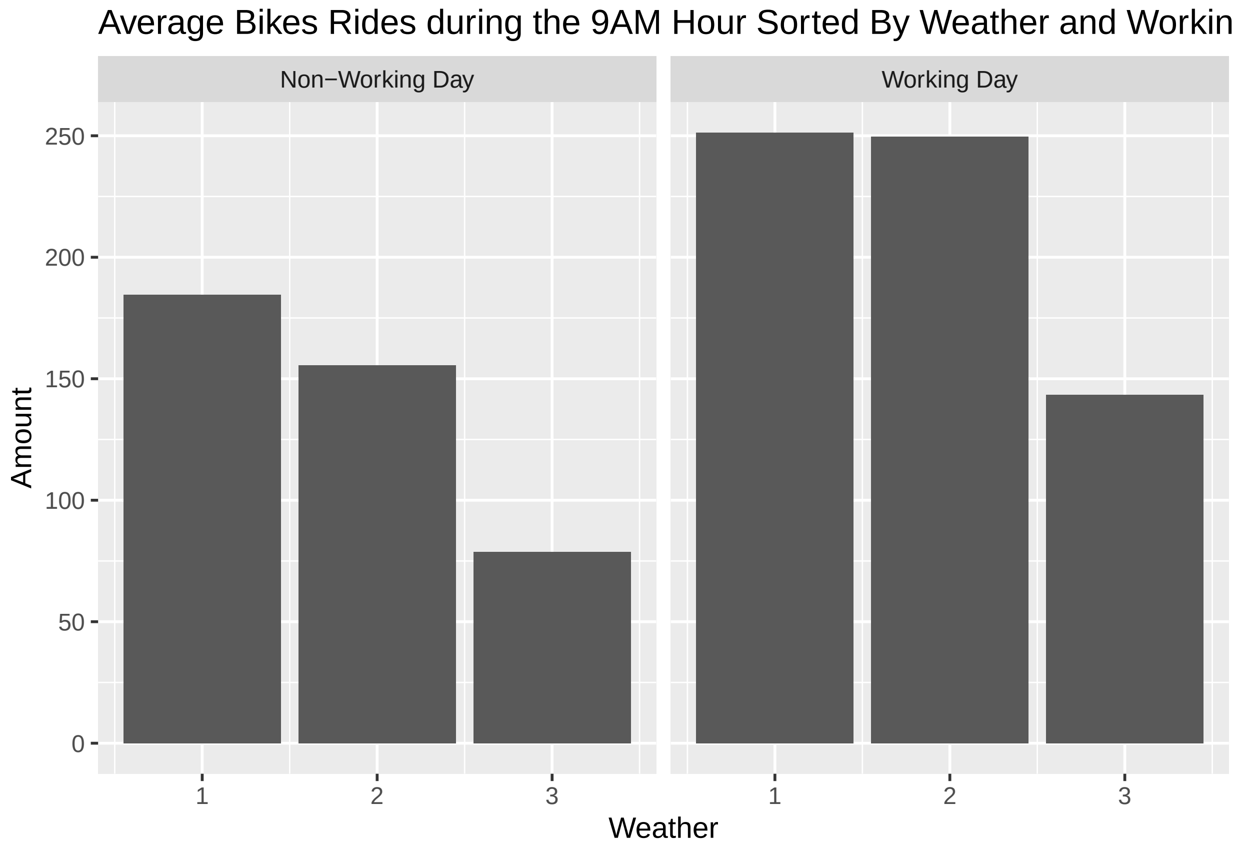
The hour axis represents each individual hour during the day, ranging from 0-23, with the 0 hour representing the time period 12-1 AM. The Bike Rentals axis displays the amount of rentals during each hour's time frame. Bike rentals are very low during the night hours, until a gradual rise with a peak during the 8th hour and another peak during the 17th hour, with continued high use throughout midday. This indicates that most bikes rentals occur between 8-9 AM and 5-6 AM, which is a logical assumption as this is when the most people are commuting to and from work. General activity during the middle of the day is also expected, as the most people are awake and traveling during this time. In summary, bike rentals peaks during the 9AM and 5PM hours, the times that most people are commuting to and from work.

Part B



When splitting the data, the Working Day graph is similar to the aggregate data graph, but with higher extremes and less people renting bikes during the 9-5 hours. As many people will have already arrived at their workplace, there will be less of an available population to rent bikes. On Non-Working Days, there is instead a gradual increase of bike rentals throughout the day, with the peak occurring between 12-2 PM. Since there is no work on these days, this data instead can be explained as when humans are outdoors and the most active during the day, with the peak likely being a direct result of people biking to eat lunch. In summary, the Non-Working Day graph shows a gradual peak of activity at midday, while the Working Day graph has 2 sharp peaks during 9AM and 5PM, the peak hours of commute to and from work.

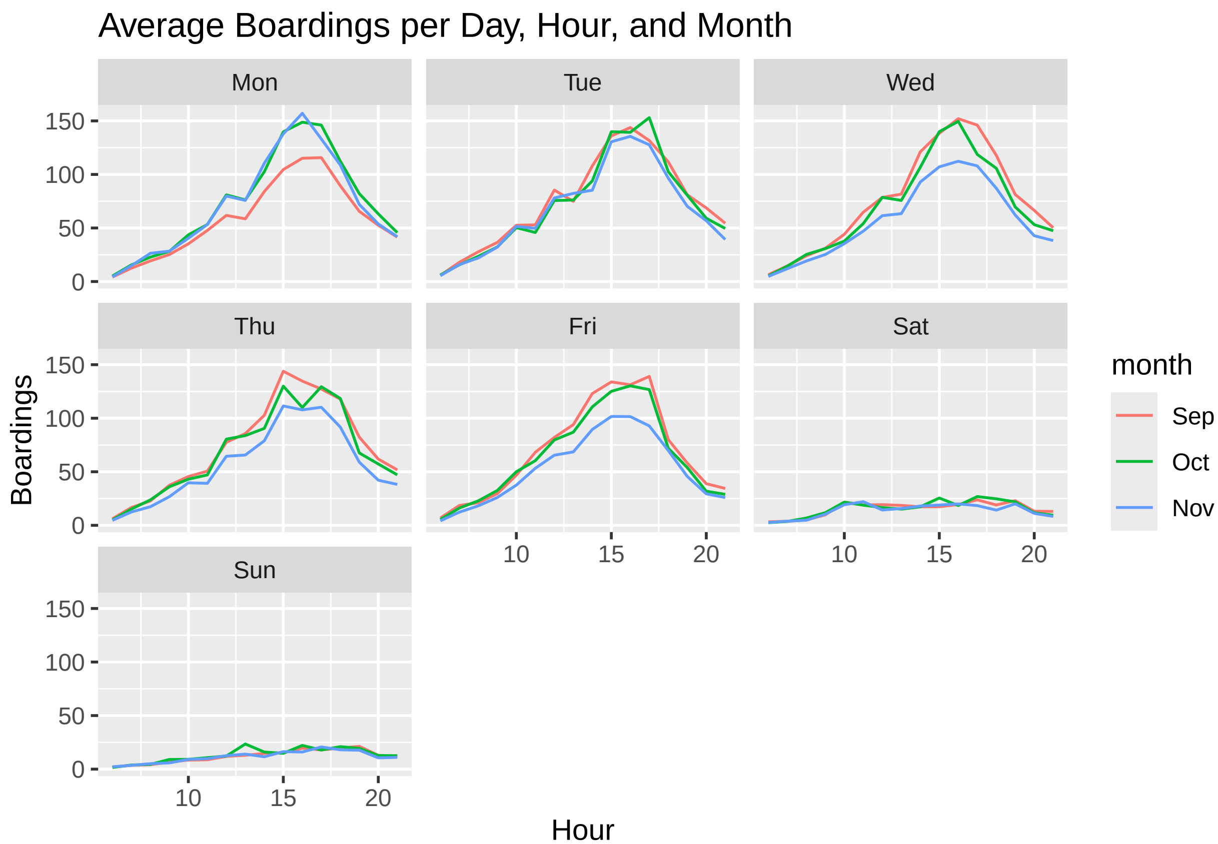
Part C



The two bar graphs demonstrate the average ridership during the 9AM hour based on the weather, with each graph segmented by whether or not the day was a working day. The weather values are numerical from 1-4, with a lower value indicating better and clear weather and the highest values, 3-4, indicating heavy precipitation, strong winds, and/or inclement weather. Observing the Non-Working Day Graph, the number of rentals significantly decreases between each stage of weather, with no recorded cases for the worst weather. For Working Days, there is a very small decrease between good and fair weather, and a significant decrease only when weather includes rain and snow. Since there is a necessity for bike renters to commute during working-days, they will only not rent if the weather directly obstructs their bike rides. Since bike rentals during Non-Working Days are more recreational, the weather plays a bigger factor in if people want to write bikes. In summary, poor weather has a higher effect on rentals during Non-Working Days than Working Days since those who rent bikes during off days need them less.

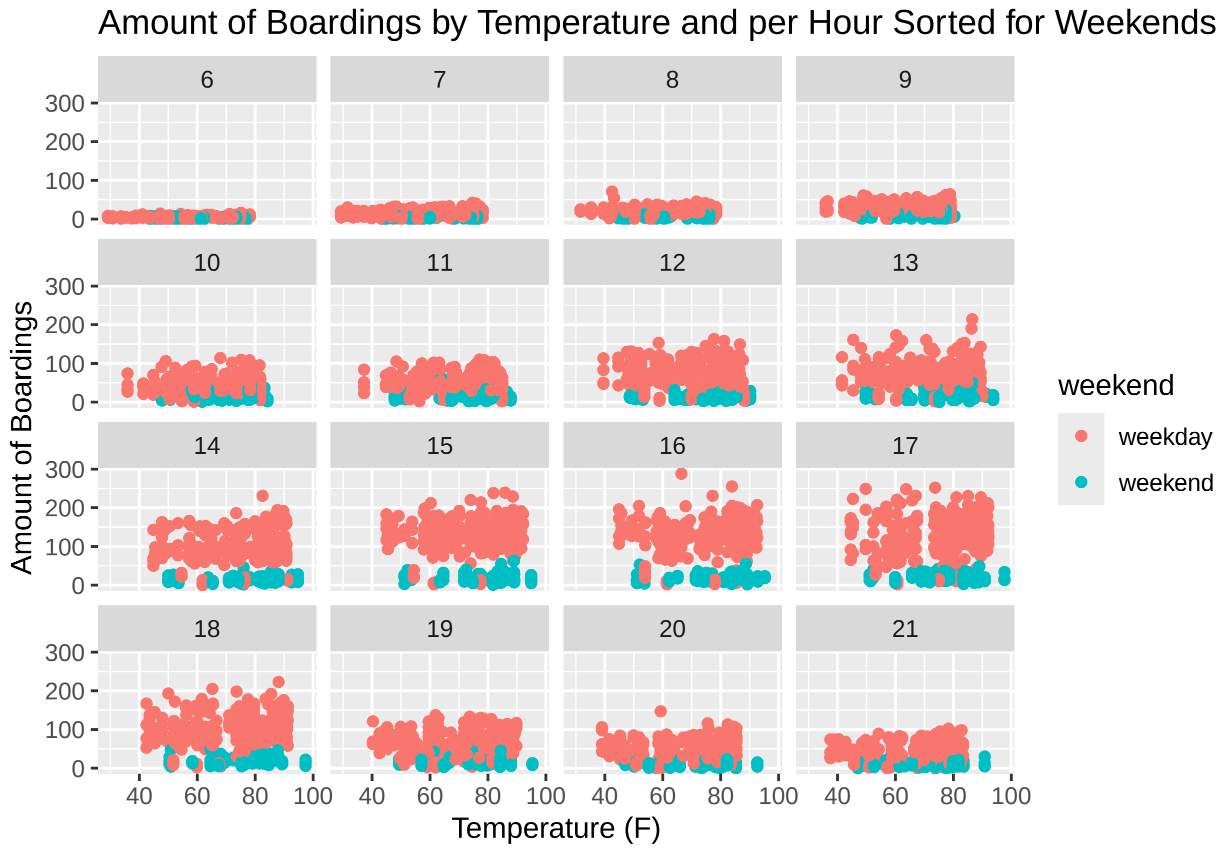
Problem 3

Part A



Each smaller graph demonstrates the average boardings per hour across each of the 7 days, with each line showing the amount per month. The hour of peak boardings appears to be broadly similar across each day, with t falling around either 3PM or 4PM. Decreased use on Monday of September could be attributed to the start of school, when many students have not yet arrived on campus or apartments and so are more likely to traveling by car or plane. November has markedly less boardings on Wednesday, Thursday, and Friday. As November contains Thanksgiving break, many students will be at home celebrating Thanksgiving, and since Thanksgiving falls on a Thursday, many students will already be off campus by Wednesday and likely stay off campus on Friday, leading to decreased boardings around UT. In summary, the university environment of Austin and UT contributes significantly to data trends in the boardings around the city and campus.

Part B



Each smaller graph displays the boardings per hour by each Fahrenheit temperature value, sorted by if it was a weekend or weekday. Boardings appear to be much high on weekdays across most hours, with only the early morning hours showing similar numbers. When holding hour of day weekend/weekday constant, there appears to be a very slight correlation between temperature and boardings. In particular, hours 12, 13, 14, and 18 contain the most noticeable correlations.

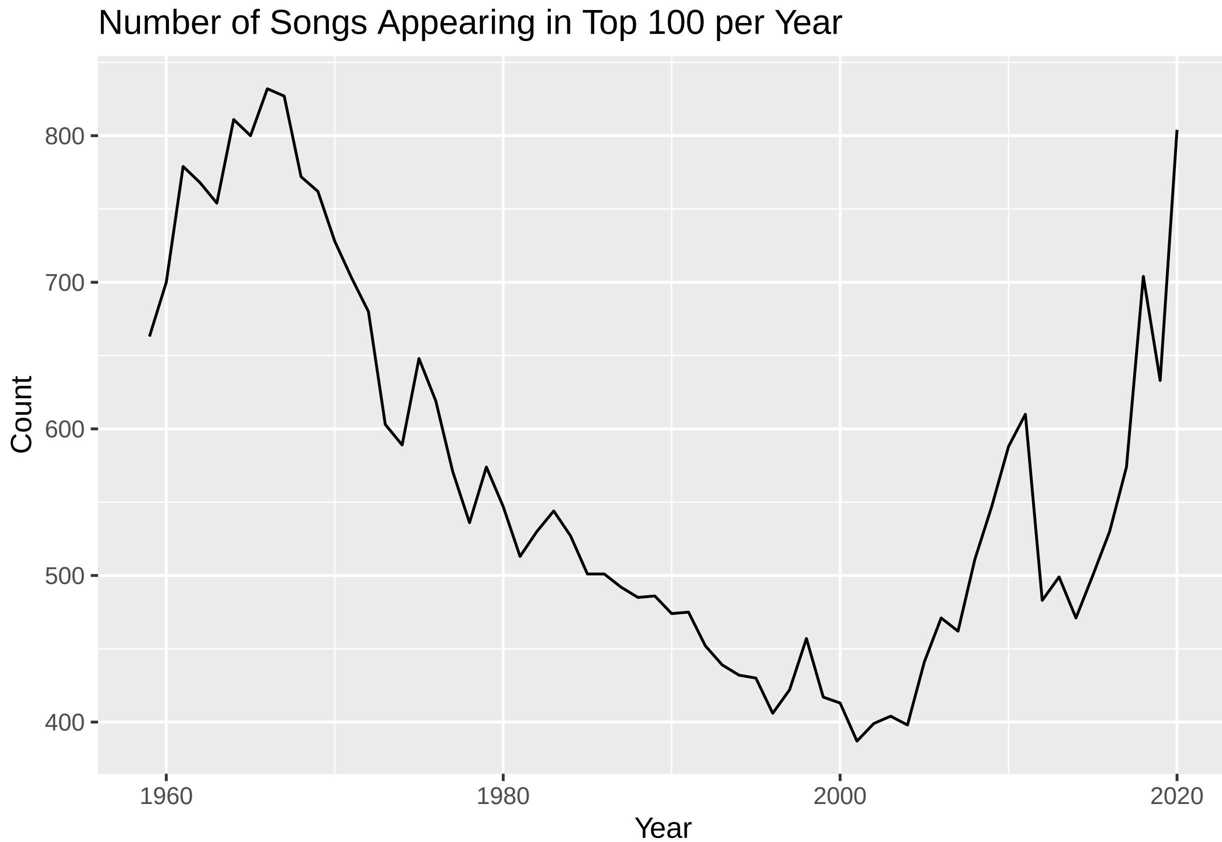
Problem 4

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
```

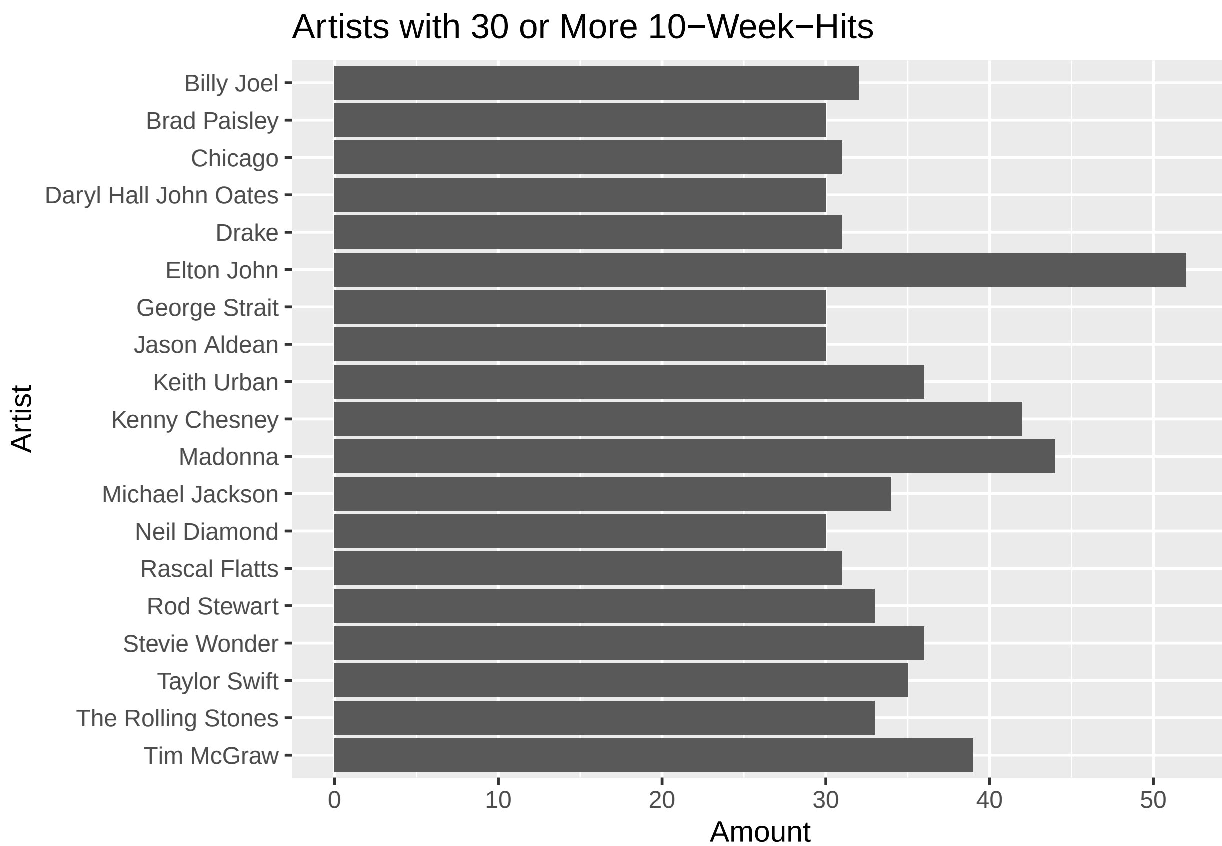
This table displays the 10 songs that spent the longest on the Billboard Top 100, measured by weeks. Radiocative by Imagine Dragons spent the longest at 87 weeks, or around 1.67 years.

Part B



The line graph shows the amount of songs that appeared on the Top 100 in any given year, i.e. any song that spent at least one week on the Top 100 at any point in that year. Uniqueness peaked in the late 1960s, when around 830 songs appeared on the 1966 Top 100. From then until the year 2001, uniqueness decreased, until less than 400 songs appeared on the 2001 Top 100. Since then, there has been a resurgence, with 2020, the last complete year, having just above 800 unique songs, rivaling the peak of the 1960s.

Part C



The bar graph displays artists who had at least 30 “10-Week-Hits”, or songs that spent at least 10 weeks on the Top 100. There are 19 artists in total who meet such criteria, with Elton John recording the most such songs at a monumental 52 songs.