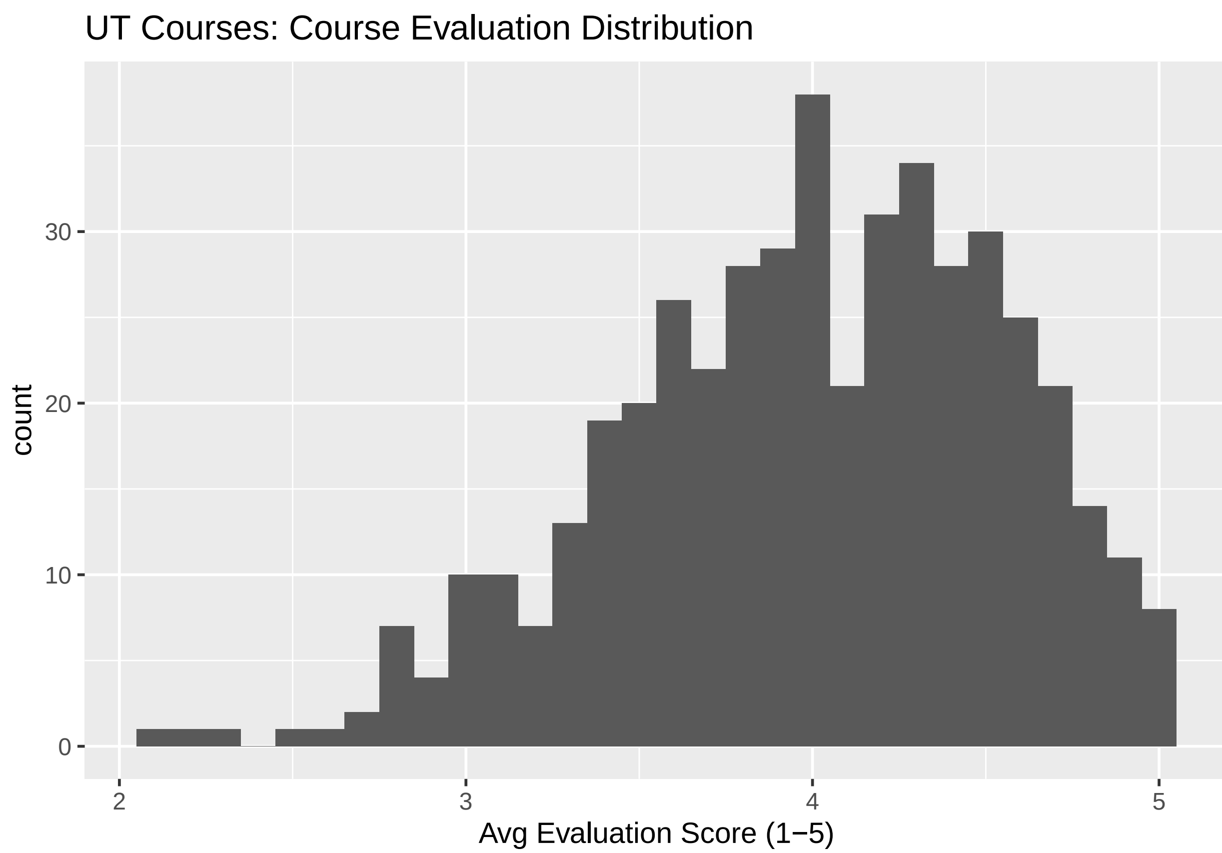
HW2

Danny Pan, EID: dp36627, https://github.com/Panda-nny/HW2_Pan

2025-01-26

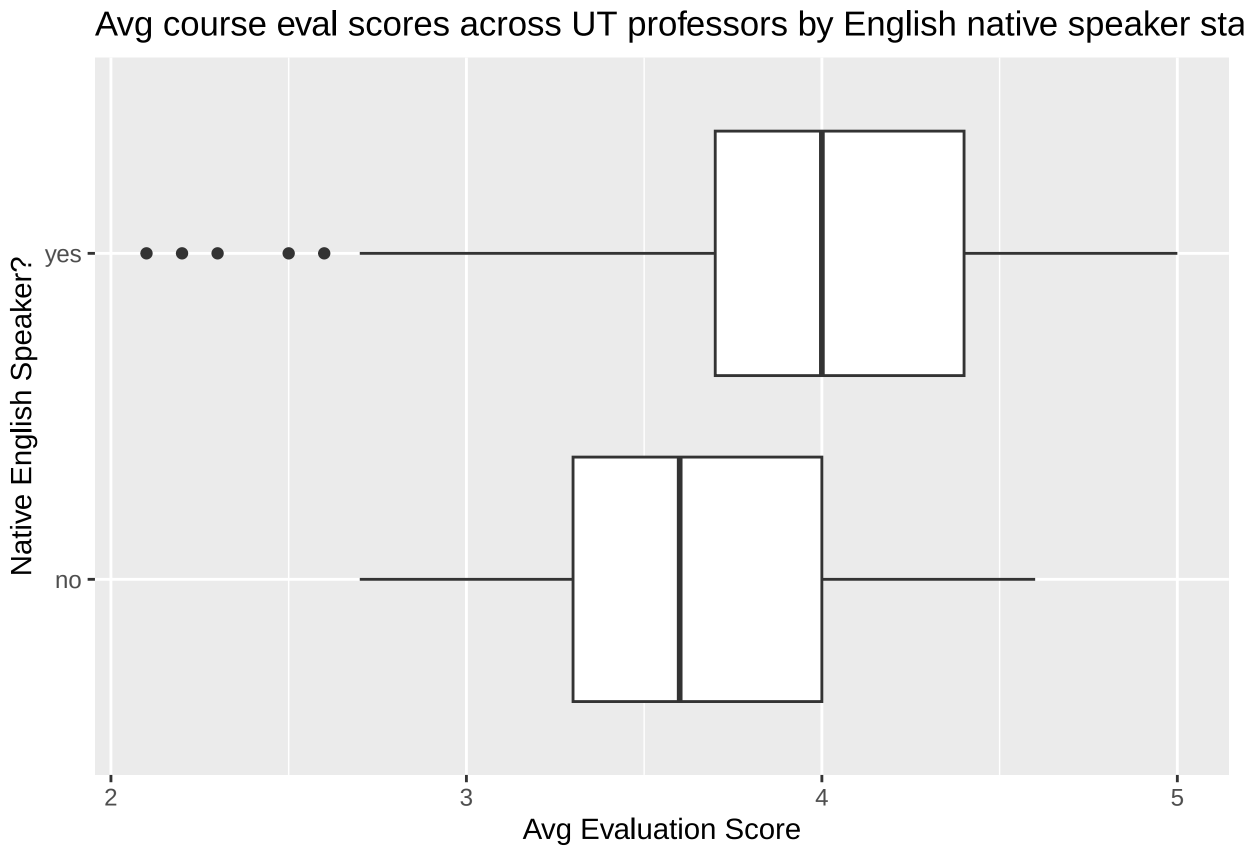
Problem 1:

1.A)



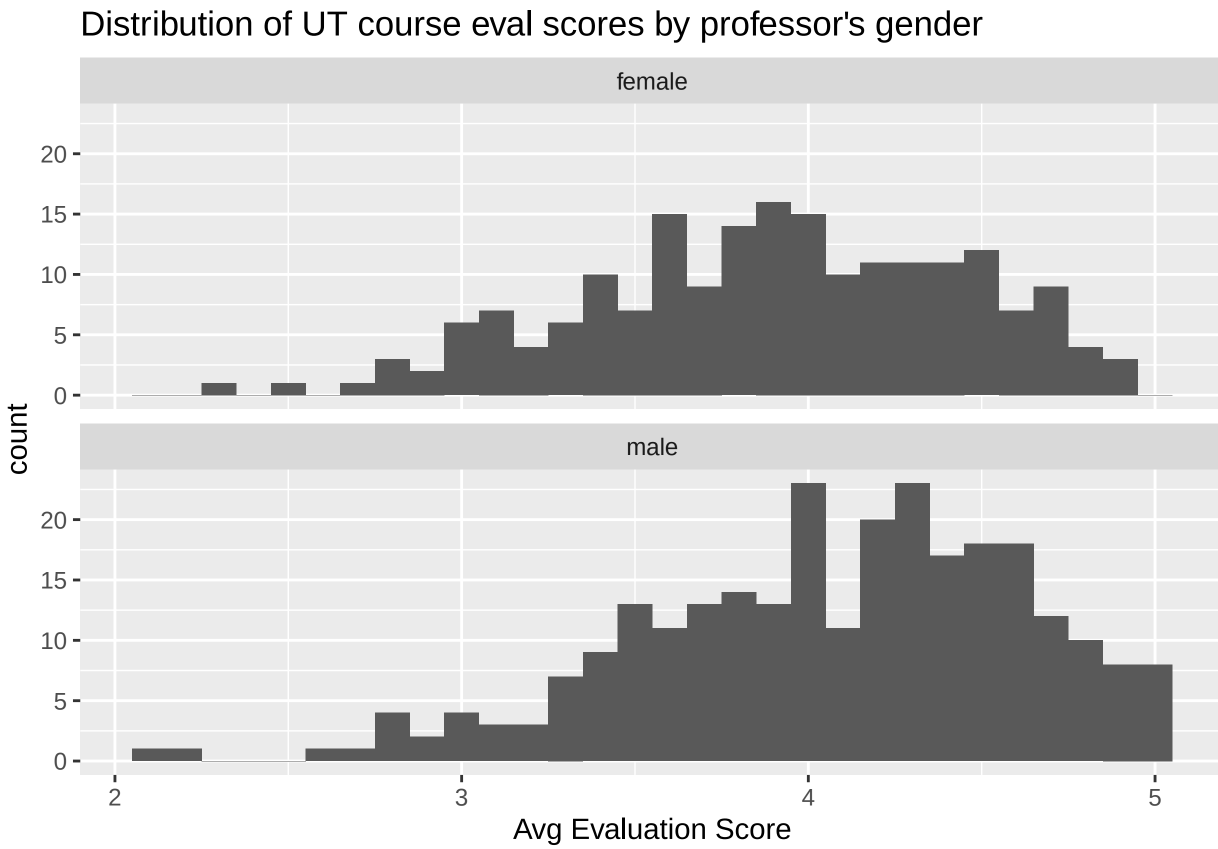
The distribution of course evaluation scores of 463 UT courses appear to be approximately normal, centered around a mean of 4.0 with a standard deviation of 0.55. While multiple courses had the maximum average course rating of 5, no professor's average course rating was lower than 2.1. Note that some professors taught multiple courses that were evaluated in this dataset (a total of 94 unique professors).

1.B)



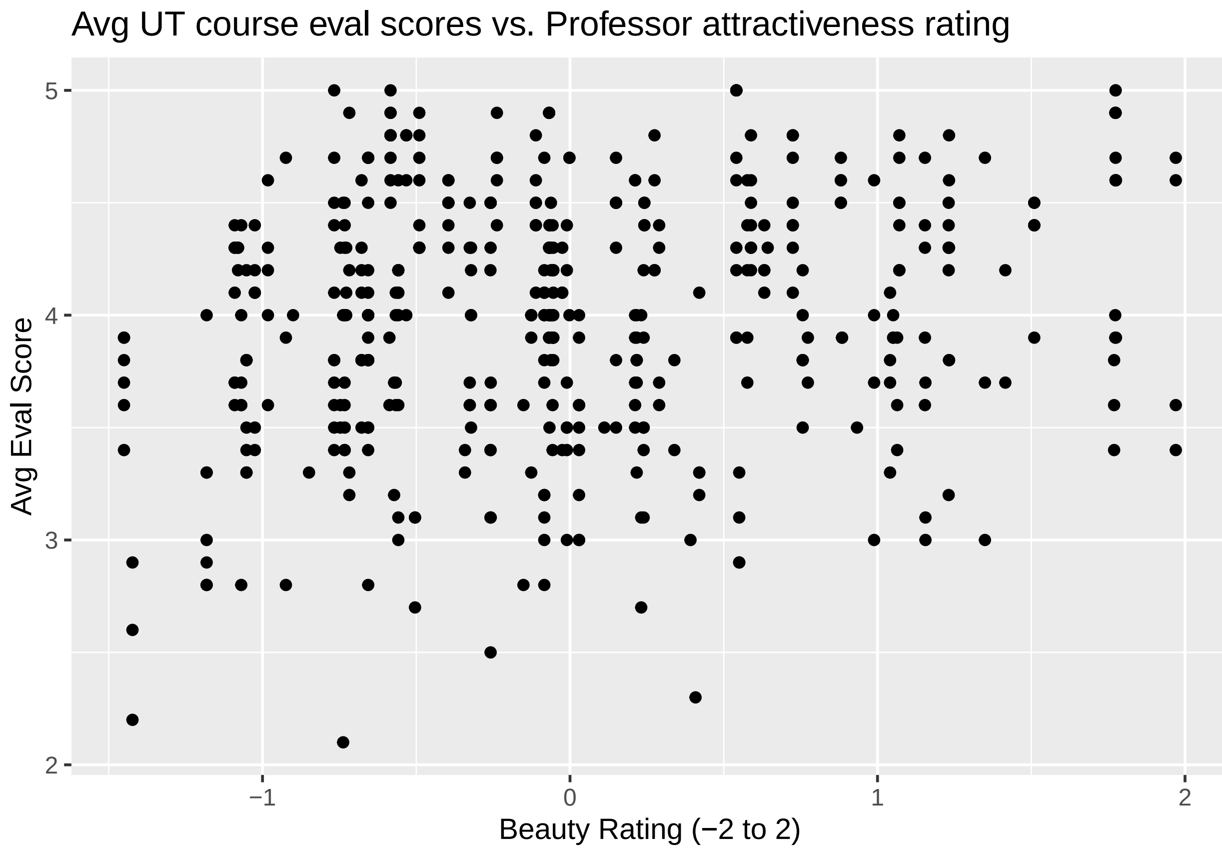
The median course eval rating for the 435 average evaluations for native English speaking UT professors was 4.0, compared to a median rating of 3.6 for the 28 average evaluations of non-native English speaking UT professors. That being said, the range of average evaluation scores was much greater for native English speaking professors, with both a higher maximum and lower minimum scores.

1.C)



The distribution of female and male UT professors' average evaluation scores appear fairly similar, with an average score of 4.07 with a standard deviation of 0.56 for 268 average evaluations of male professors and an average score of 3.90 with a standard deviation of 0.54 for 195 average evaluations of female professors. Male professors were the only professors to be scored an average of 5/5 but also consisted of the professors with the lowest ratings.

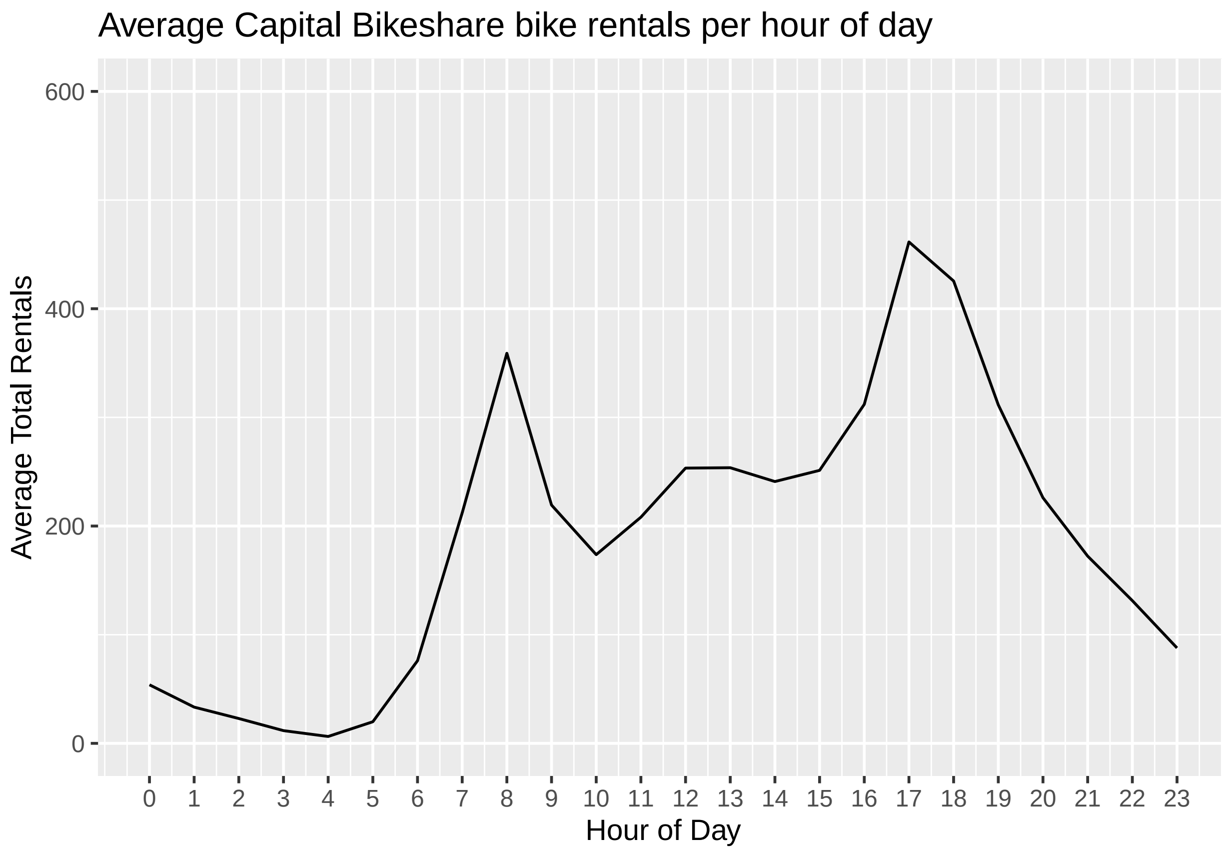
1.D)



There doesn't appear to be a clear association between the average rating of UT course evaluations and the attractiveness rating of the course's professor, which has an low r-squared value of 0.04.

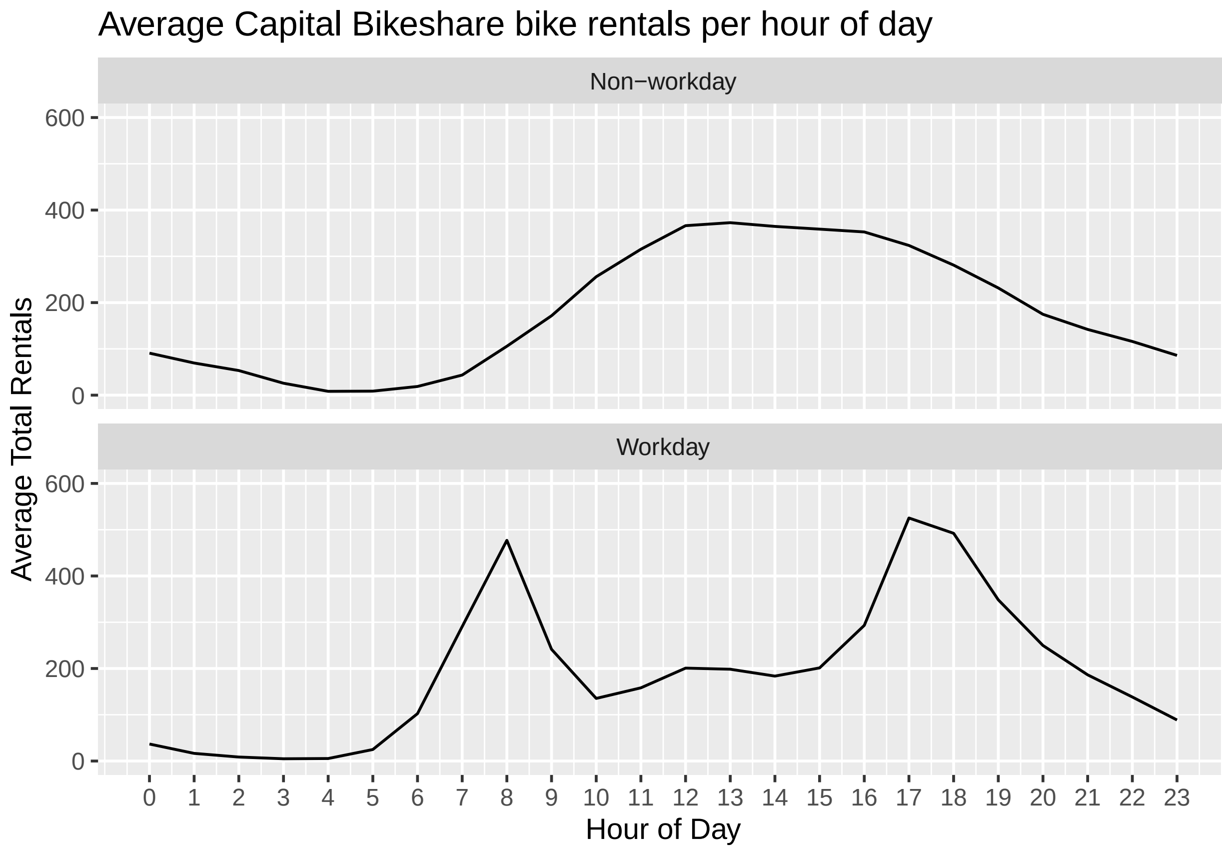
Problem 2:

2.A)



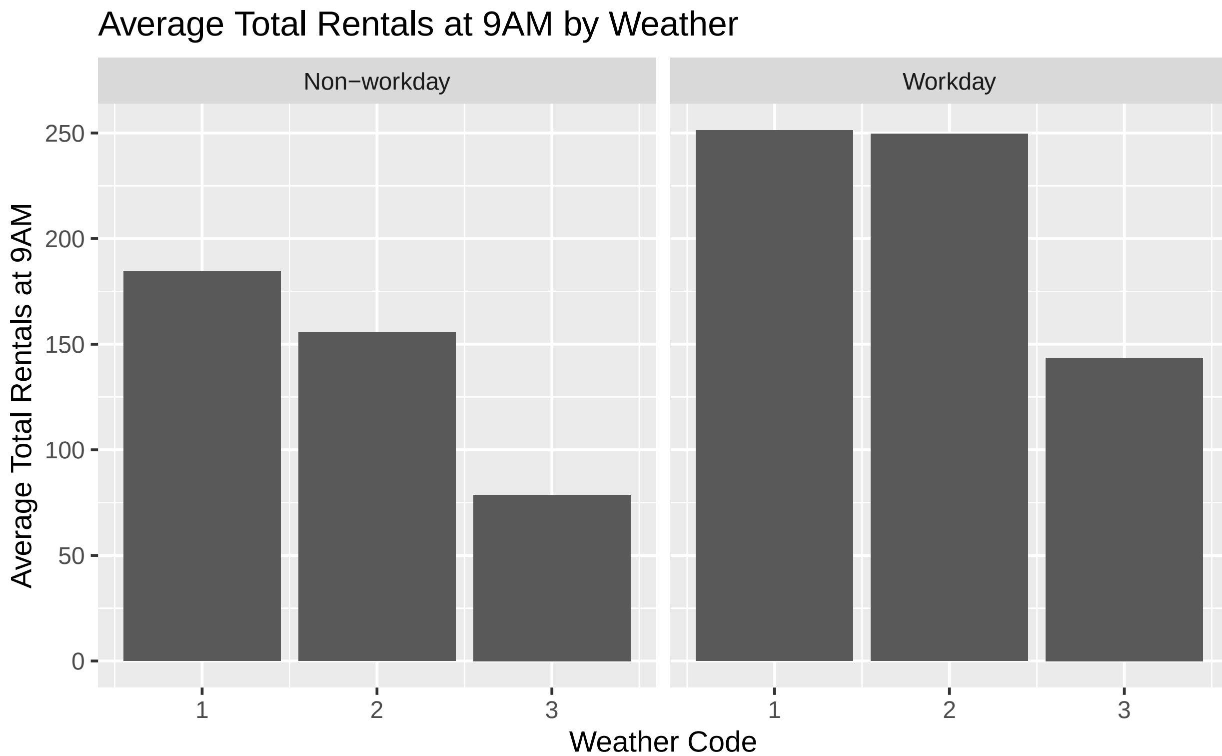
There appears to be two primary peaks in hours for bike rentals in DC: one at 8-9AM and another larger peak of over 450 average rentals each day at 5-6PM.

2.B)



It's interesting to note that the pattern of average total rentals per hour for nonworkdays doesn't resemble the overall pattern at all, with a gradual increase in the morning to a fairly-uniform peak in the afternoon to a gradual decrease afterwards.

2.C)

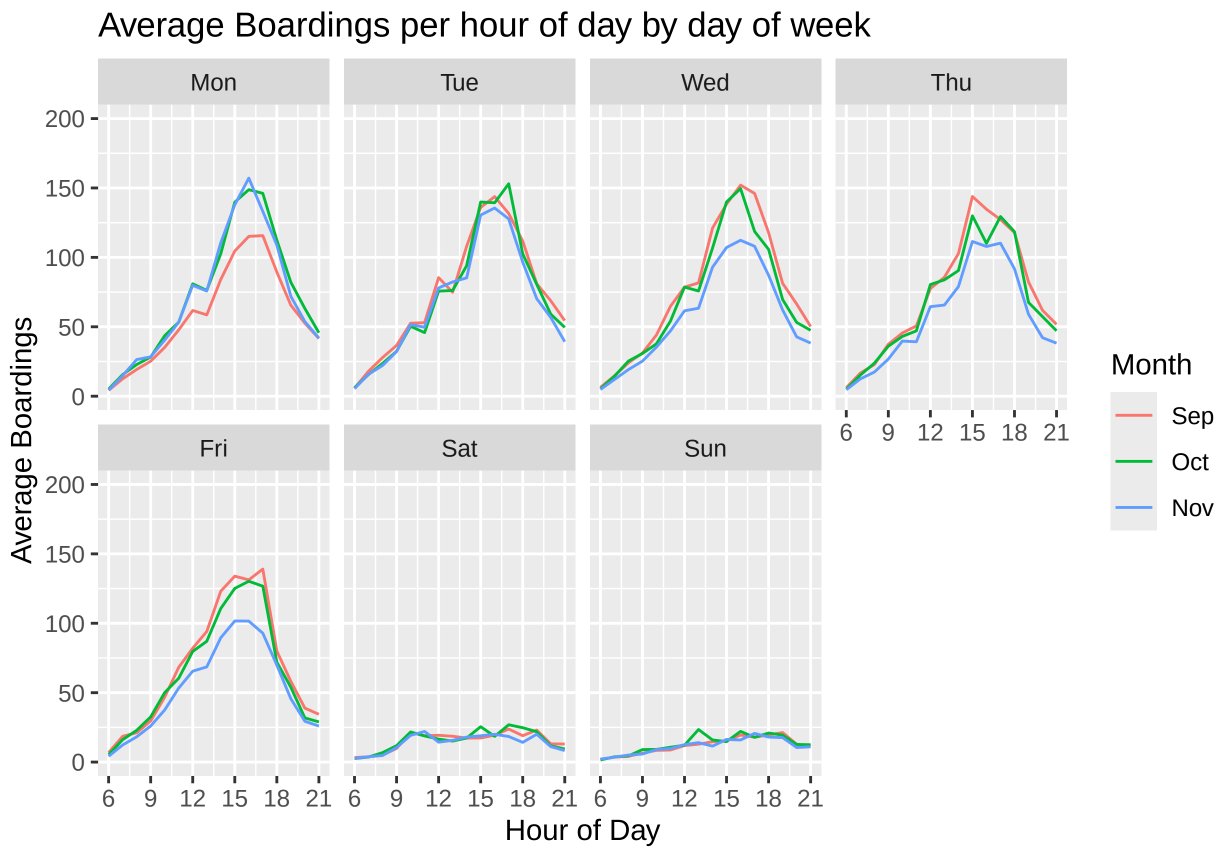


1: Clear, Few clouds, Partly cloudy, Partly cloudy
2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds

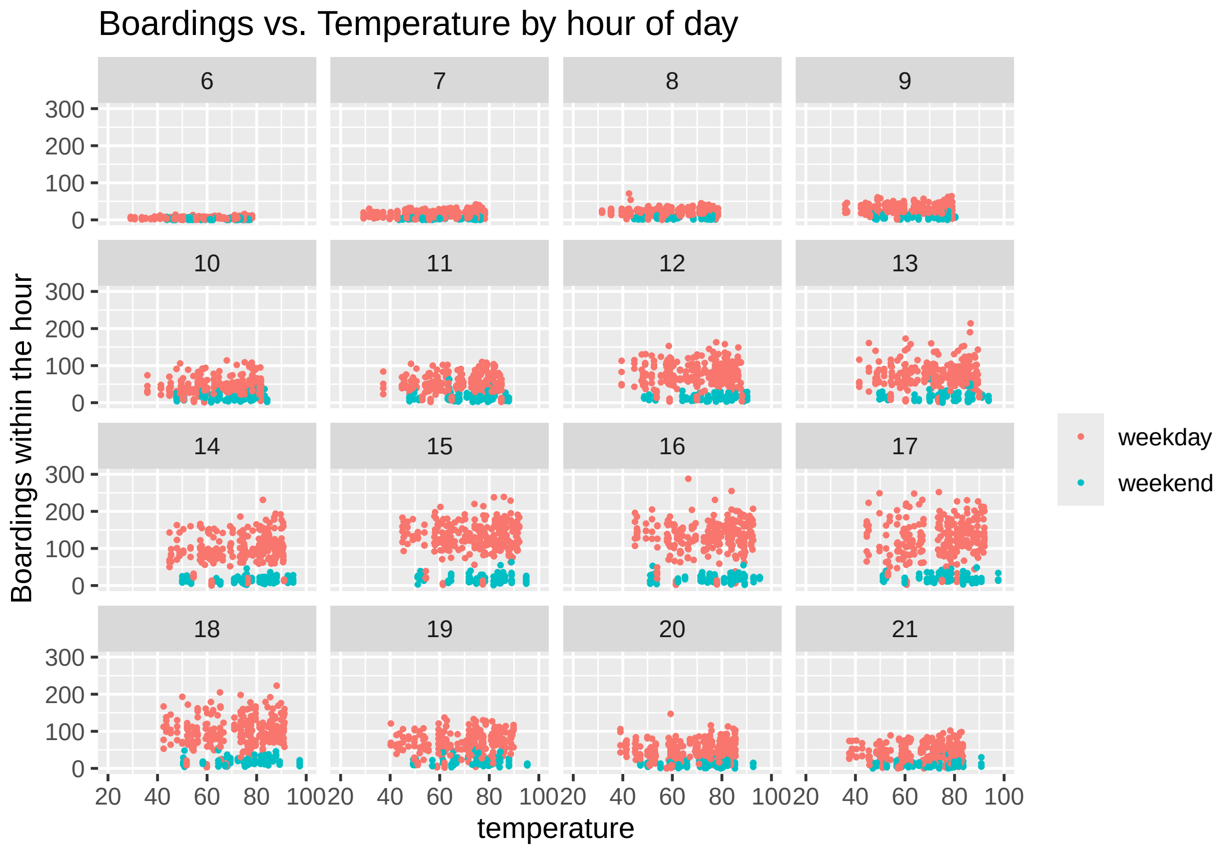
While mist seems to be a factor contributing to less 9AM rentals on non-workdays, mist doesn't affect rates on workdays. There is a noticeable drop in 9AM rentals across all days on average when there is light precipitation.

Problem 3:

3.1)



The above graphs display the average hourly number of Capital Metro bus boardings in the UT area, split by the day of week and differentiated by different Fall months. It appears that bus usage is infrequent on the weekends (which is relatively stagnant, potentially due to reduced service) compared to weekdays (which has a noticeable peak in the afternoon). Additionally, peak hours on weekdays don't vary too much from around 4-6pm. It's interesting to note that there is a significantly lower average number of boardings on September Mondays and Wed/Th/Fri of November, which may be explained by holidays such as Labor Day and Thanksgiving Break.



The above scatterplots do not display a clear correlation between temperature and boardings, and a test for linear correlation between boarding counts and temperature controlling for weekends vs. weekdays yields an r-squared value of less than .03 across all hours of the day. However, a test for linear correlation may not be the most reliable method of determine the effects of temperature on boarding counts as both excessively hotter and colder temperatures are generally considered undesirable (which may or may not lead to an increase of bus usage). Regardless, the scatterplot shows a clear difference in boarding numbers between weekends and weekdays, with the general disparity reaching a high in the late afternoon between weekends and weekdays.

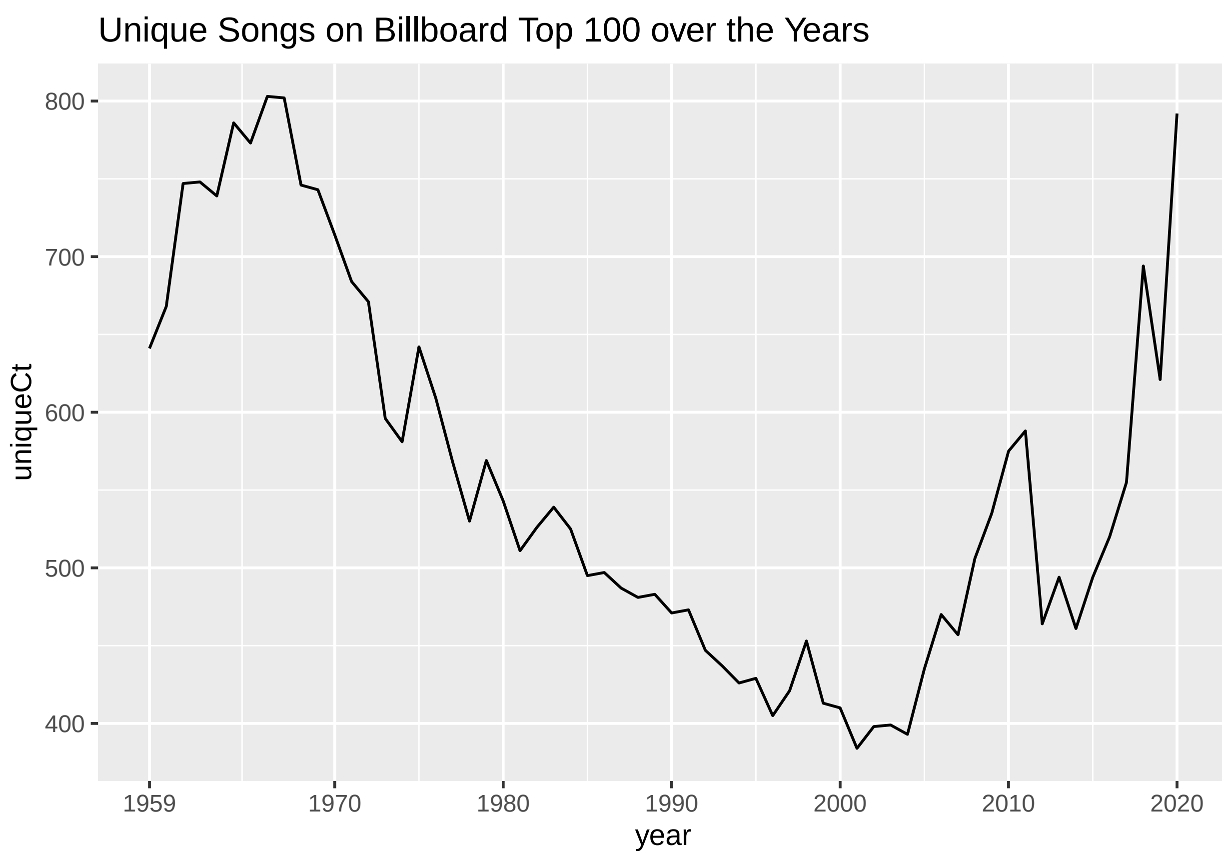
Problem 4:

4.A)

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

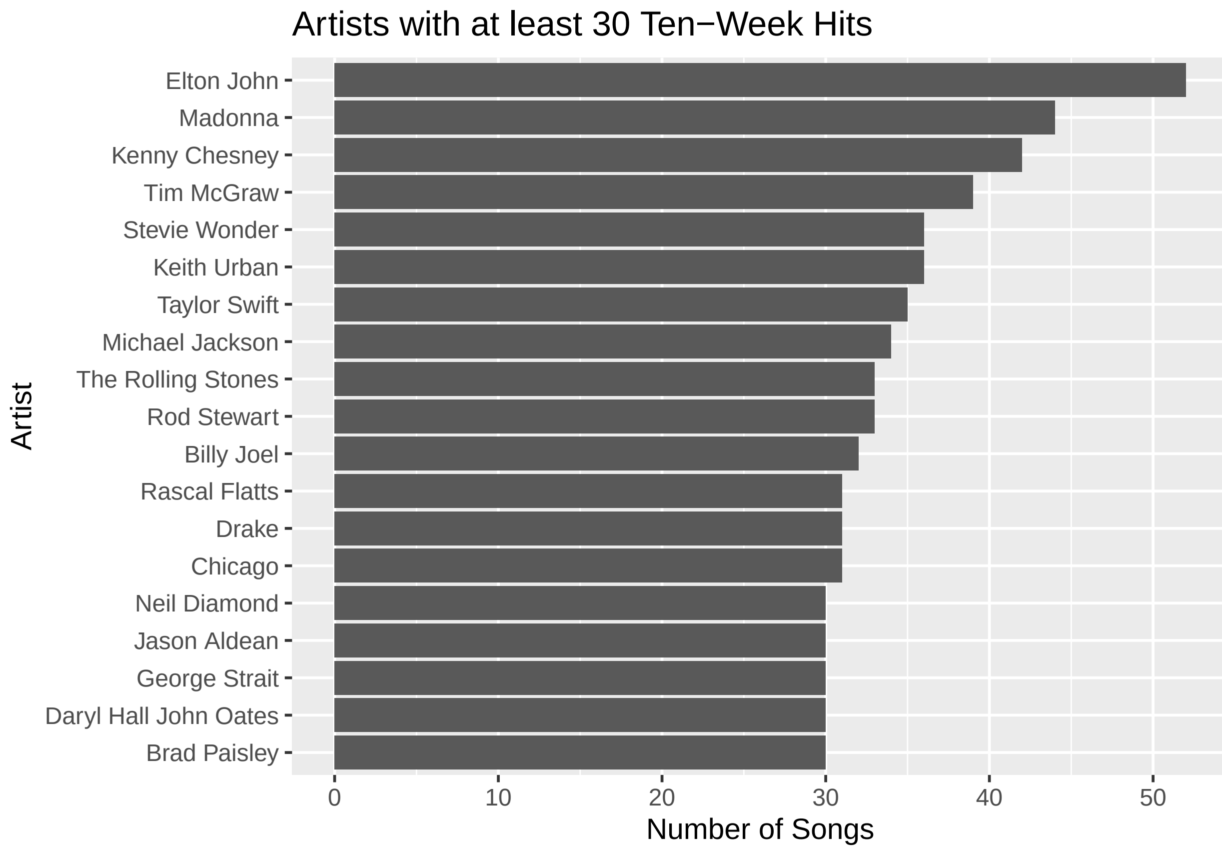
The above table lists the 10 songs by a performer that appeared on the Billboard Top 100 for the longest total number of weeks from 1958 to the 22nd week of 2021. There is yet to be a song to be on the Billboard Top 100 for over 1.5 years.

4.B)



The graph of total unique songs per year from 1959-2020 displays a high at around 1960, then a continual decline of unique songs until the early 2000s, and generally gradual increase since then, with a sharp relative peak in 2010/2011. The presence of more unique songs may indicate both a diversity of songs in the Billboard Top 100 but also potentially less audience retention.

4.3)



The 19 artists with 30 or more songs in the Billboard Top 100 for at least 10 weeks are listed above, with Elton John assuredly topping the list with 32 songs, cementing his legacy as one of the Kings of Pop.