HW2

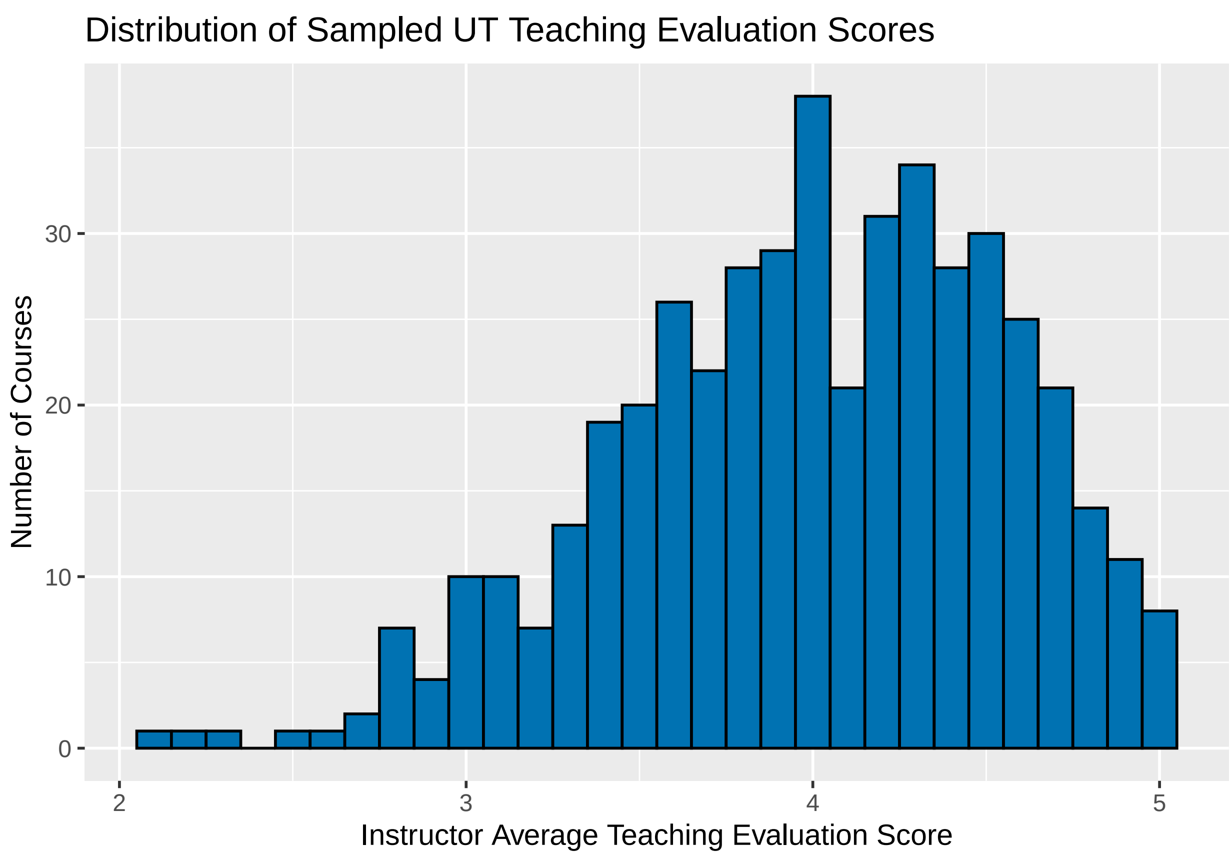
Surafel Geleta

ssg2775

Github: https://github.com/surafelgeleta/SDS315_HW2

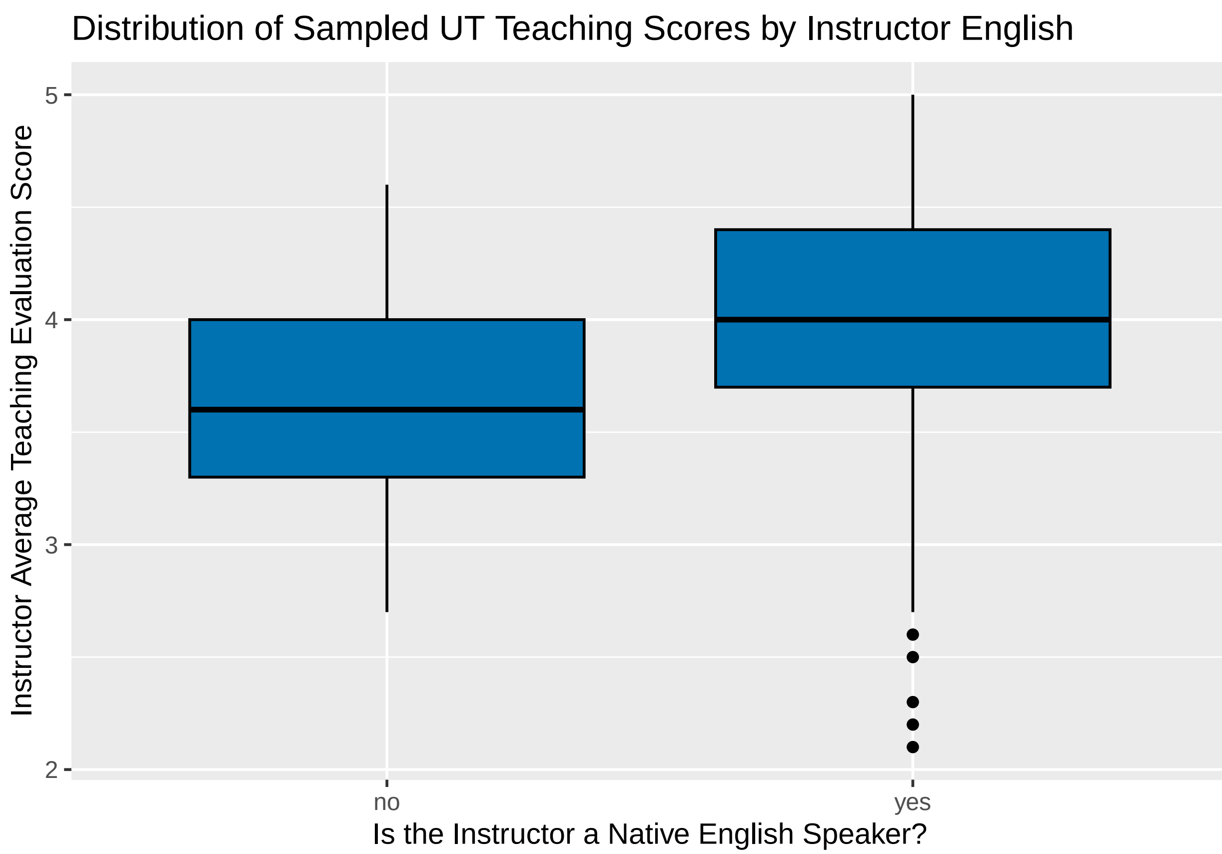
Problem 1: Beauty, or not, in the classroom

Part A



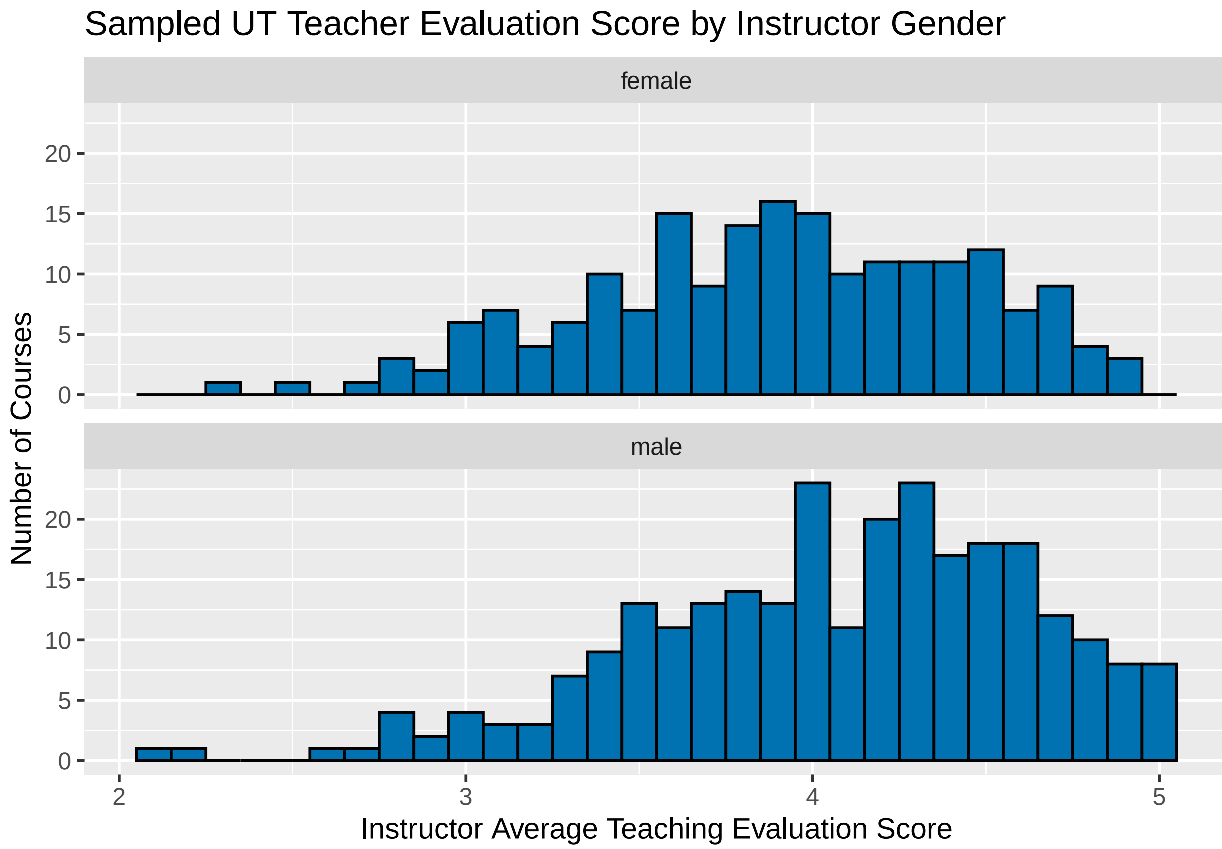
Instructors at UT Austin in the sample appear to have average teaching evaluation scores centered around 4, indicating that most of the sampled instructors are rated generally positively. The distribution of scores is left-skewed, and the range of scores appears to be about 3 points.

Part B



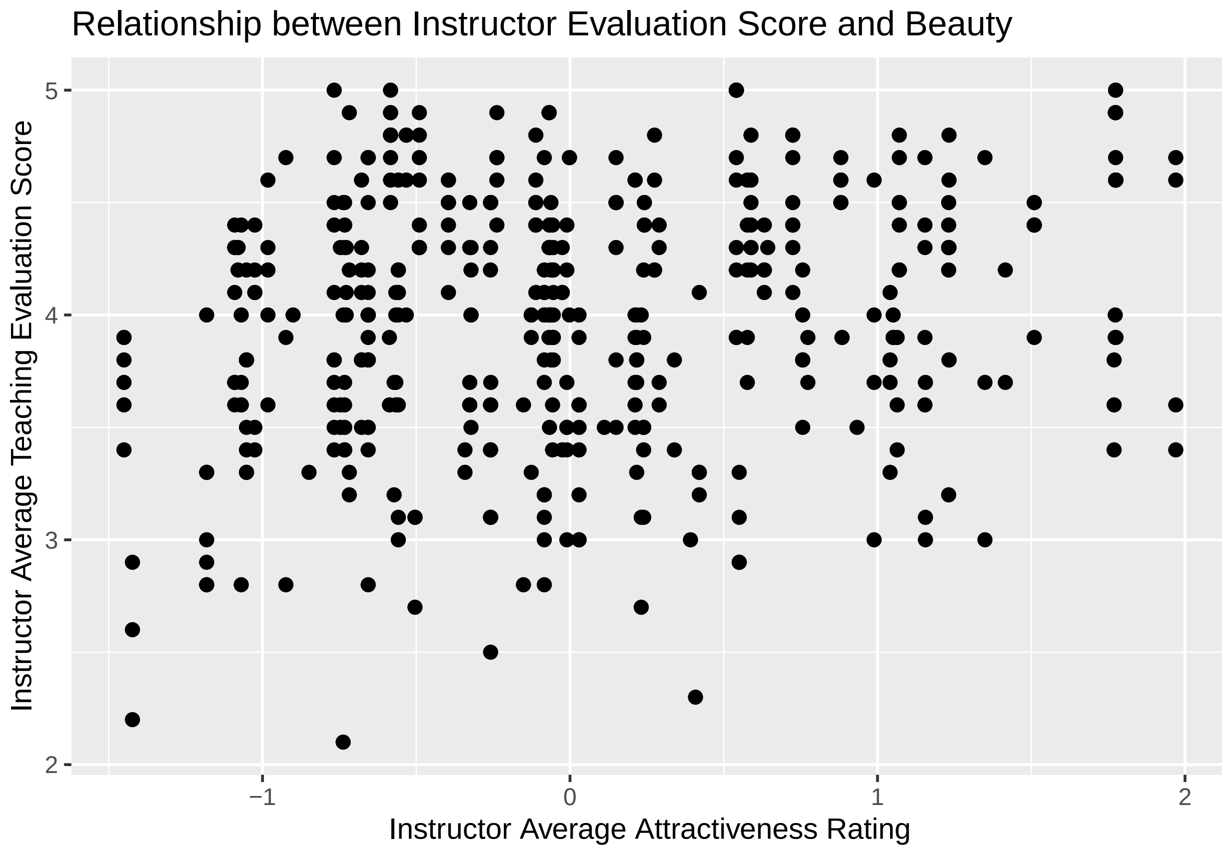
Sampled UT instructors who are native English speakers have a median average evaluation score about 0.4 points higher than non-native English speaking instructors.

Part C



Sampled male instructors appear to have average evaluation scores slightly higher than sampled female instructors. The distributions for both genders are left-skewed, and female instructors seem have a slightly wider range of evaluation scores than female instructors.

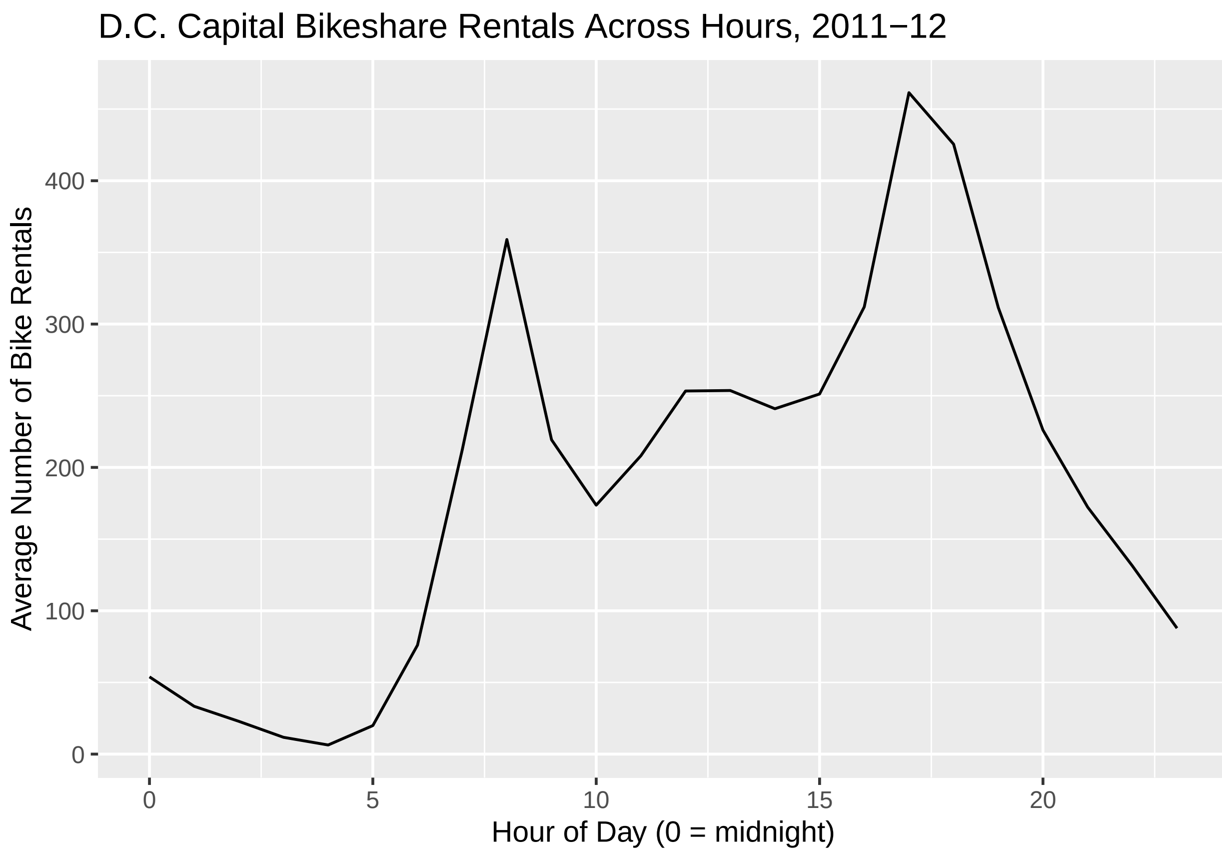
Part D



There is a very weak, positive relationship between sampled instructor attractiveness and teaching evaluation score.

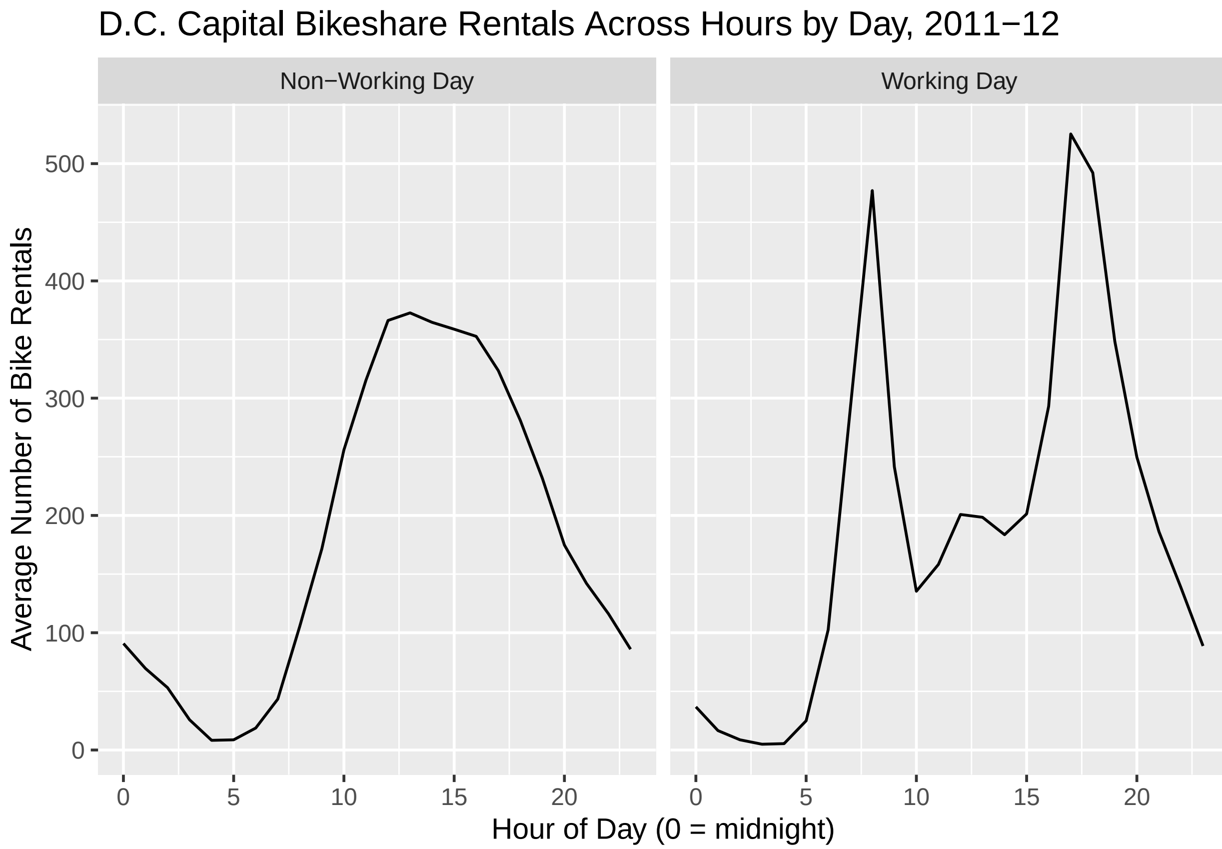
Problem 2: Bike Sharing

Plot A



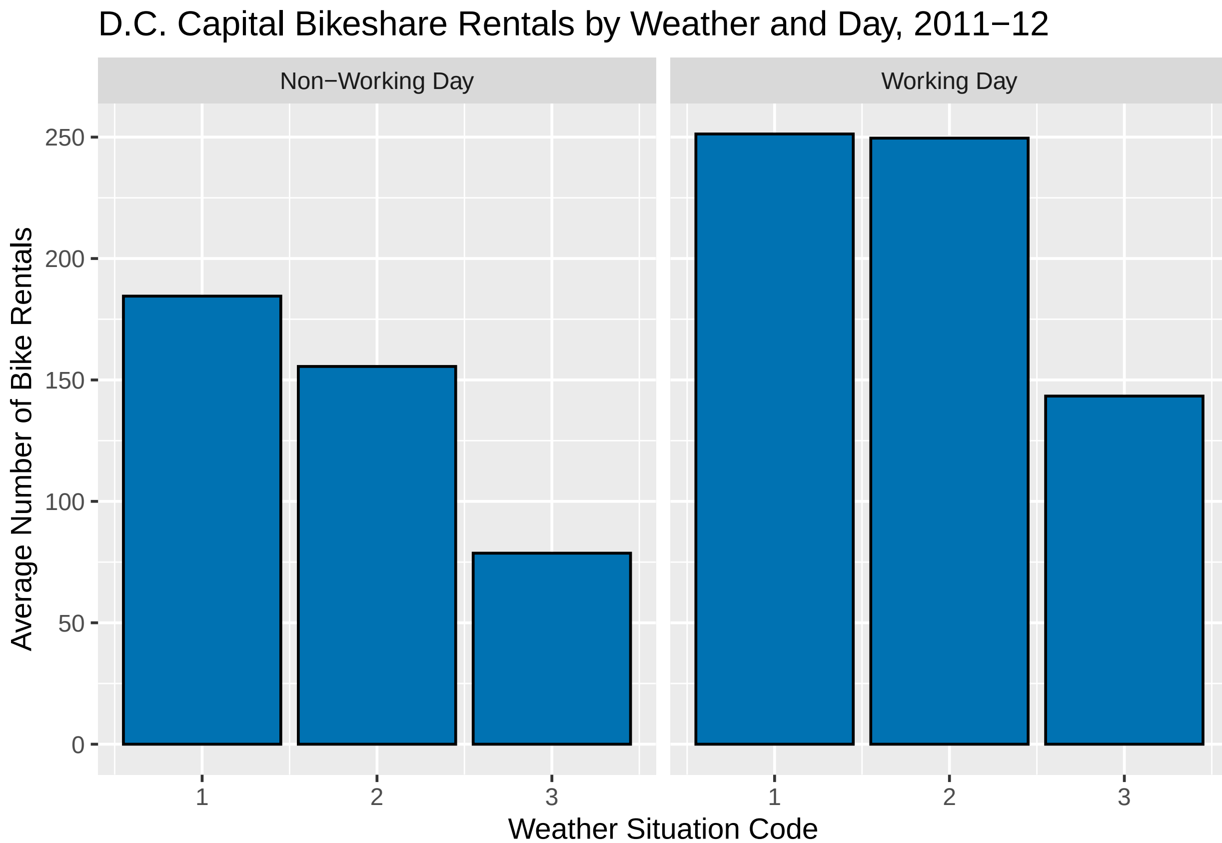
This line graph shows the changes in average D.C. Bikeshare rentals across hours, using data collected between 2011 and 2012. The x-axis displays hours across a day, with 0 representing midnight, 15 representing 3PM, and so on. The y-axis indicates the average number of bike rentals made out, calculated by grouping data by hours and finding the mean number of rentals made within each hour. Average bike rentals start low, with about an average of 50 rentals made on midnight, and the average dropping until 4AM. Bike rentals begin to slowly rise after 4AM, and experience a sharp morning spike at 8AM, with a little over an average of 350 bikes rentals made out during that hour. After 8AM, average bike rentals fall steeply until 10AM, where they experience a brief increase to around an average of 250 bike rentals by noon. Between noon and 3PM, mean bike rentals remain stagnant; however, after 3PM, rentals spike to an all-day high at 5PM, with a little over an average of 450 rentals made during the hour. After 5PM, rentals steadily fall. **Overall, it appears that the Bikeshare bike rental patterns in D.C. during 2011-12 follow standard work and commute hours**. People typically start work in the morning (perhaps at 9AM), so rentals peaking at 8AM may indicate that people are taking these bikes to work. The small increase in rentals occuring after 10AM could be indicative of people leaving work for lunch, while the significant peak in rentals at 5PM may be people taking out bike rentals to return home.

Plot B



This plot displays D.C. Bikeshare bike rental patterns by the type of day when they occurred, with working days referring to non-holiday and non-weekend days, and non-working days referring to the opposite. The x-axis shows hours across the day, while the y-axis displays the average number of bike rentals made during each hour. First, it's important to note that average nighttime bike rentals are higher on non-working days than they are on working days. For instance, at midnight, the average number of bike rentals made on non-working days is about 90, while on working days it is less than 50. However, non-working days appear to experience a more late increase in morning bike rentals than working days due. On working days, average bike rentals begin to increase between 5AM and 6AM, while on non-working days this spike begins to develop sometime at 7AM. In addition, after this morning peak, non-working days do not experience a sharp decline in average bike rentals, while this does occur on working days. Mean bike rentals on non-working days appear to be at their highest at 1PM, and from there fall to their nighttime levels. Given these differences, its likely that bike rental patterns are associated with work hours, as working days appear to have much more different rental patterns than non-working days do. **More importantly with regards to non-working days, average bike rentals may be higher than on working days during night due to nightlife activity and holidays being more common on non-working days.**

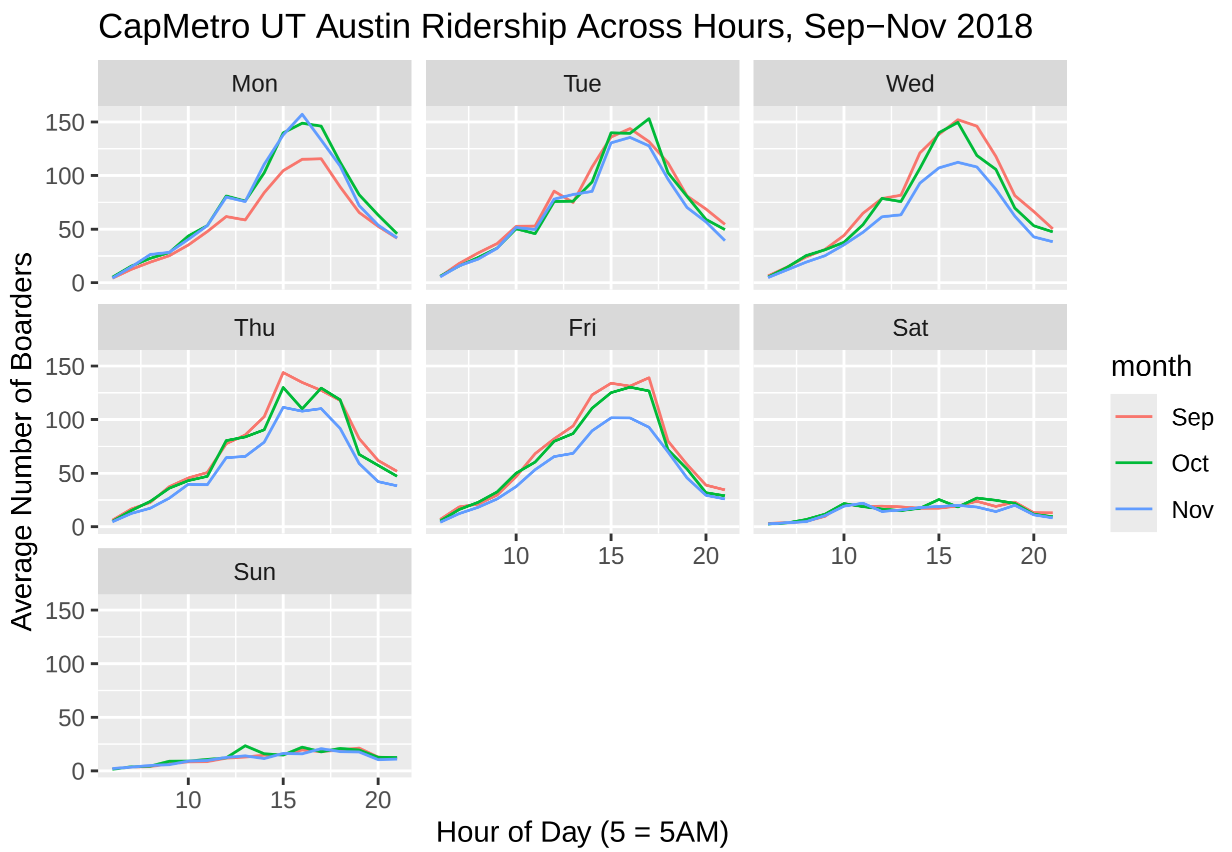
Plot C



This graphic displays the average Bikeshare rentals during 9AM across different weather situations. Moreover, as with the previous graphic, there are two plots by the day status; working day and non-working day. The weather situation codes are assigned as follows: "1" indicates anywhere from clear skies to a partly cloudy day, "2" indicates cloudy days with mist, and "3" indicates precipitous days with thunderstorms and light rain, light snow, or with clouds and light rain. On working days during 9AM, average bike rentals are nearly equal between weather situations 1 and 2 (at an average of ~250 rentals), while weather situation 3 saw far less rentals made, with an average of about 143 rentals during 9AM. In addition to having lower average bike rentals across the bar, bike rentals appear to be more sensitive to weather situations on non-working days than working days during 9AM. For instance, average bike rentals fall by about 30 between weather situations 1 and 2 on non-working days, a large difference when compared to the working day difference. Moreover, on non-working days at 9AM there is a significant reduction in bike rentals between weather situations 2 and 3; working days see a slightly smaller rental reduction between weather situations 2 and 3 during 9AM. **Overall, these results suggest that during 9AM, Bikeshare rental patterns in D.C. between 2011-12 are less responsive to negative weather changes on working days than on non-working days, perhaps due to the necessity behind arriving to work on time compared to the reduced importance of other activites on non-working days.**

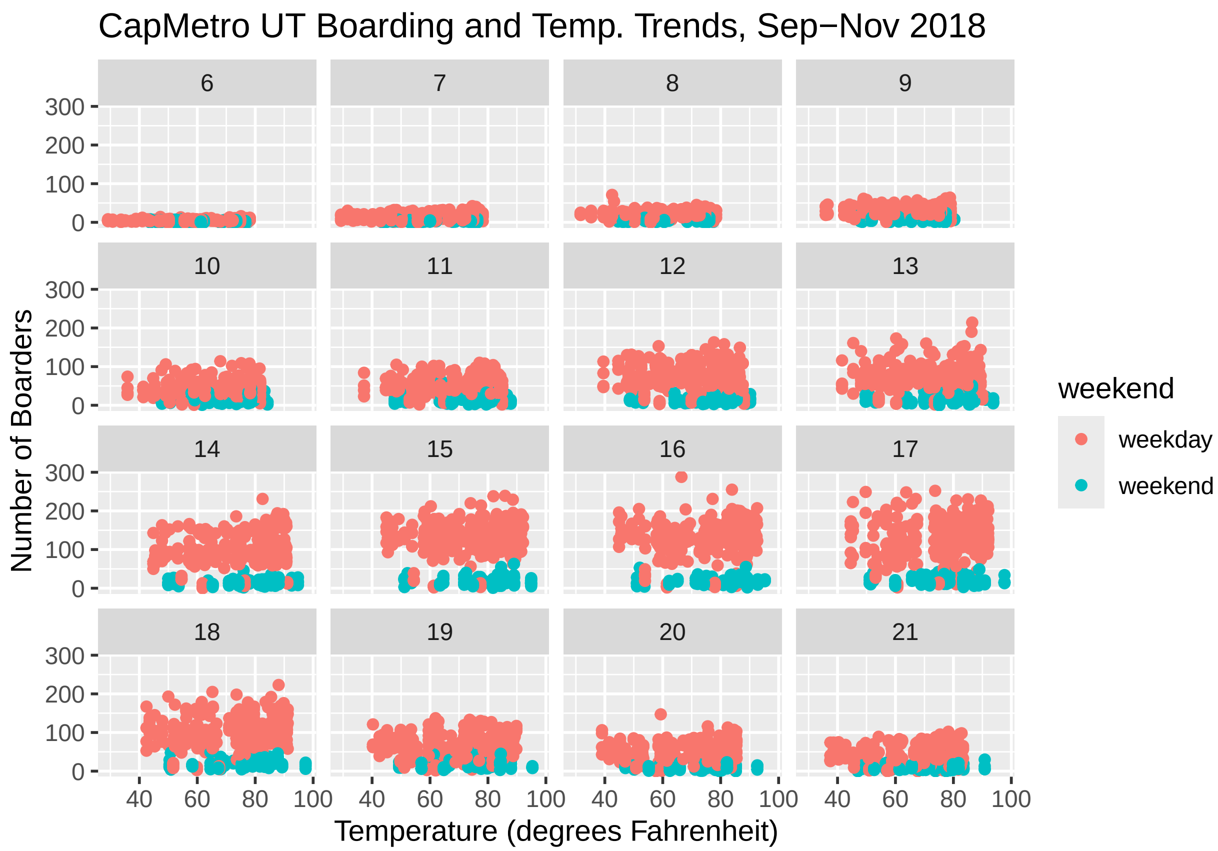
Problem 3: Capital Metro UT Ridership

Plot 1



This graphic displays CapMetro's bus ridership trends across stops around UT Austin between September and November of 2018. The y-axis shows the average, or mean, number of boarders for a given hour (Hours start at 6AM, and end at 9PM), and the x-axis shows bus operating hours within each day. The graphic is faceted, displaying 7 plots corresponding to data from each day of the week, with each plot having three color-coded lines corresponding to a month. Saturday and Sunday ridership is much lower than on weekdays; moreover, weekend ridership does not seem to follow the same patterns as weekday ridership. For instance, on Saturday average ridership peaks at 5PM in October, with an average of about 10 riders within that hour; that same hour and month on Friday sees an average of around 120 riders within the hour. Weekday ridership generally follows a similar pattern, with boarding starting low at 6AM, followed by an increase in boardings until reaching peak ridership, typically between 3-5PM. From there, boardings drop steeply into the night. Peak hours vary slightly depending on the day, as well as the month; on Mondays, September boardings peak at 5PM, and October and November at 4PM. Tuesday September boardings peak at 4PM, October at 5PM, and November at 4PM. September and November Thursdays experience peak boarding during 3PM. Saturday and Sunday stray a little from this trend; in particular, peak boardings during November Saturdays occurred at 11AM, while peaks on September and October Sundays occurring at 7PM and 1PM respectively. September boardings are noticebly than October and November boardings on Mondays; this may be explained by Labor Day, which falls every 1st Monday of September, contributing to off-days from school and work. Similarly, November boardings during Wendesday-Friday are generally lower than the other months; this is likely due to Thanksgiving, which falls on the fourth Thursday of November. Wednesday may experience lower ridership due to people staying home or travelling to prepare for Thanksgiving day on Thursday, and people may rest and recover on Friday.

Plot 2



This graphic displays the relationship between temperature and total boarding in CapMetro's UT Austin routes, faceted to include a plot of the relationship for each operating hour of the day. Temperatures increase as the day approaches mid-afternoon (3-5PM), seen with how the points generally shift right between 6AM and 5PM; temperatures begin to decline after 5PM. In addition, the plots indicate that weekend boarding counts are significantly lower than weekday boardings across all available hours. When holding the hour of the day and day type constant, there is little evidence of an association between temperature and the number of boarders. Temperature does not seem to have any associated change as the number of boarders increases.

Problem 4: Wrangling the Billboard Top 100

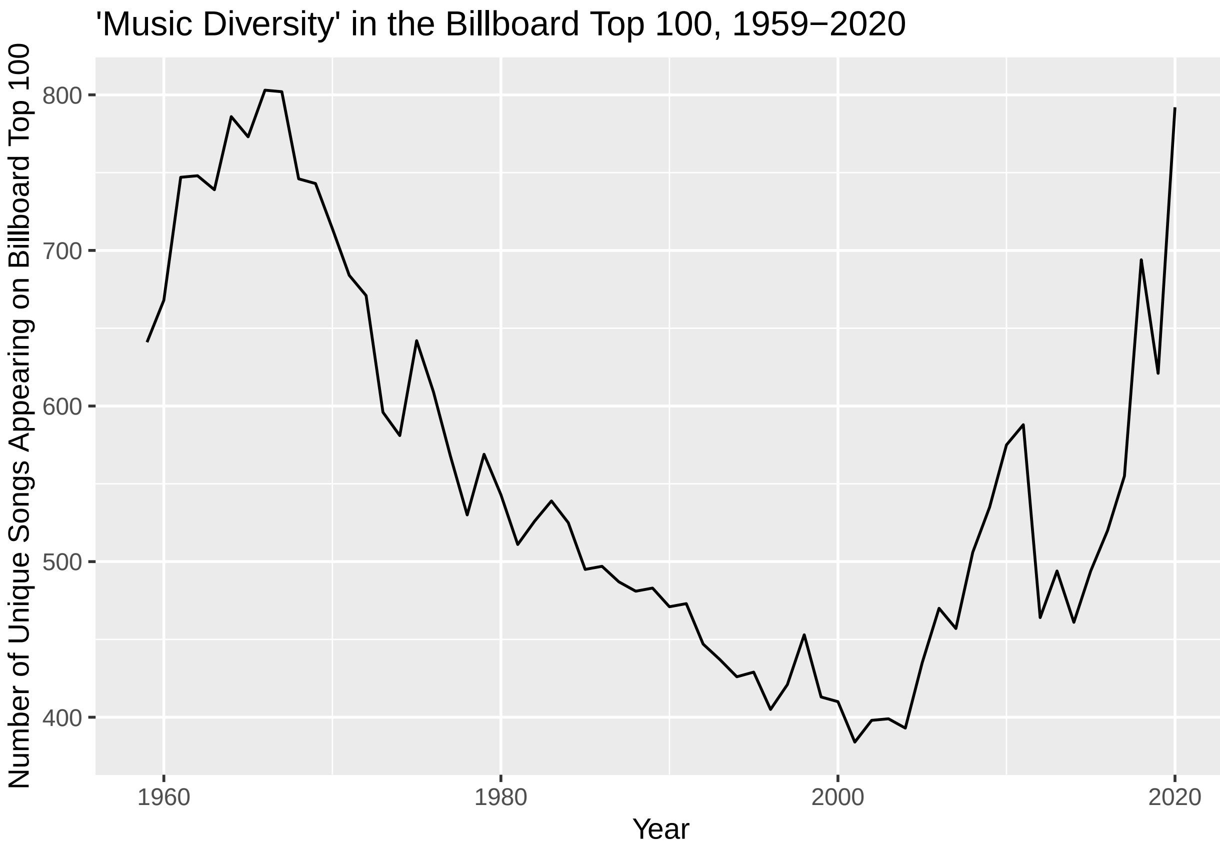
Part A

Table 1: Top 10 Most Popular Songs By Count of Weeks Spent on Billboard Top 100

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

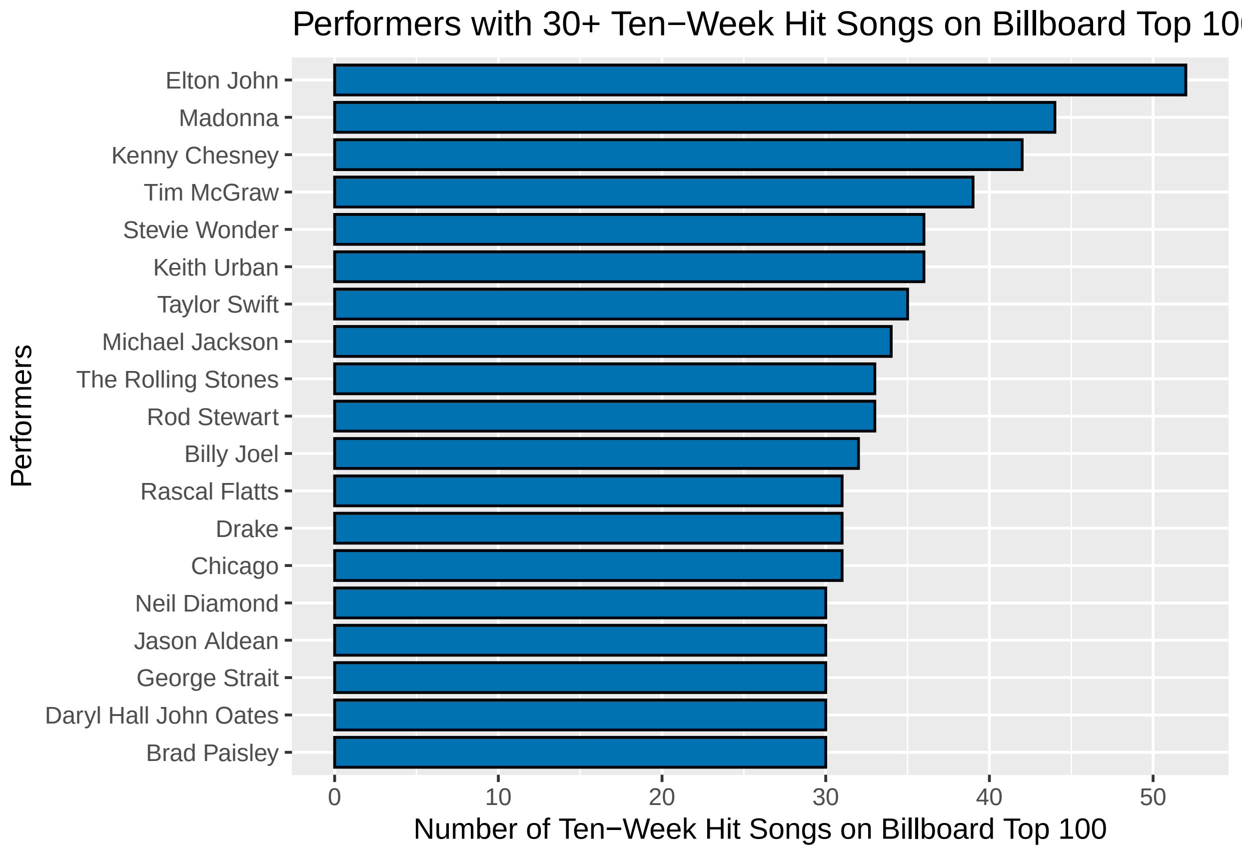
This table features the top 10 most popular songs featured in the Billboard Top 100, with popularity measured by the number of weeks the song remained on the Top 100. The "Performer" column indicates the artist of the song, and the "Count" column represents the number of weeks the song was on the Billboard Top 100. By this metric, the most popular song featured in the Billboard is "Radioactive" by Imagine Dragons.

Part B



This line chart shows changes in “musical diversity” in the Billboard Top 100 spanning between 1959 and 2020. As indicated in the y-axis, “musical diversity” is measured as the number of unique songs appearing on the Billboard Top 100, with the x-axis of the chart displaying years. Musical diversity was at it’s peak during the mid-1960s, with about 800 unique songs featuring on the Billboard Top 100 across those years. After the 60s, musical diversity steadily fell, with less than 400 unique songs appearing on the Billboard Top 100 during the very early 2000s. From there, musical diversity increased to a little over 550 unique songs by the early 2010s, but again decreased. By 2020, musical diversity once again reached a peak nearly rivaling the one it held during the 1960s.

Part C



This barplot displays the 19 performers who have had at least 30 songs that each lasted at least 10 weeks on the Billboard Top 100. The x-axis shows the number of ten-week hit songs featured on the Billboard Top 100, and the y-axis shows the names of the performers. The performer with the most ten-week hit songs is Elton John, with a little less than 55 songs meeting this standard. Notably, the difference between John's and Madonna's ten-week hit song count is fairly significant.