

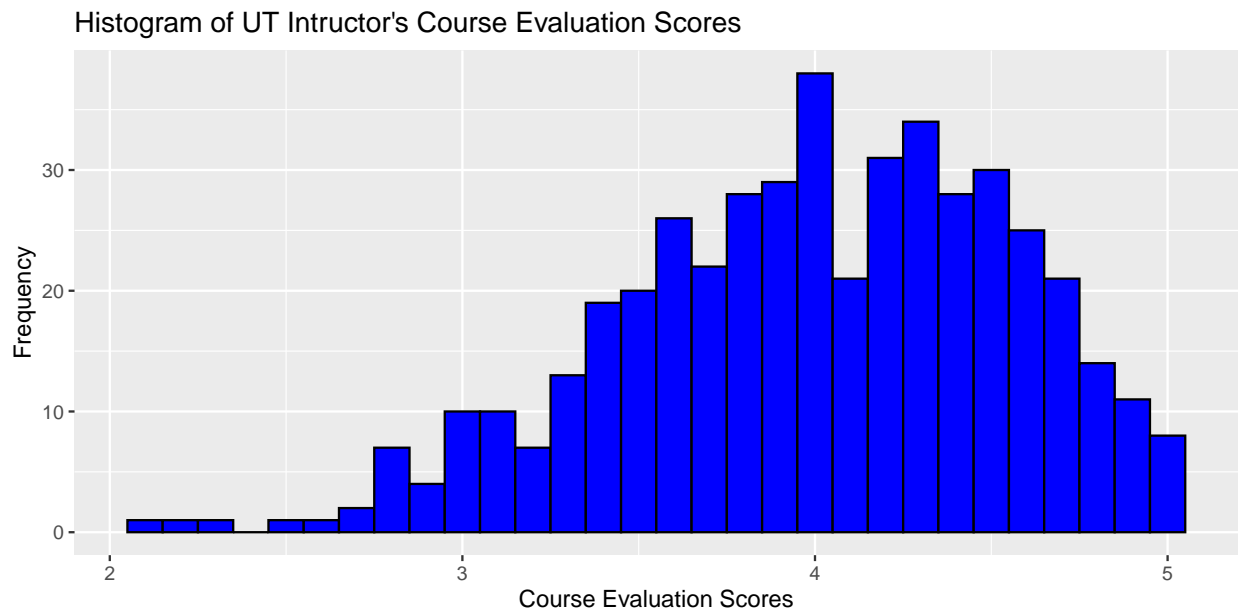
Homework 2

Jayden Polansky - jbp2839

<https://github.com/jaydenpolansky/SDS315.git>

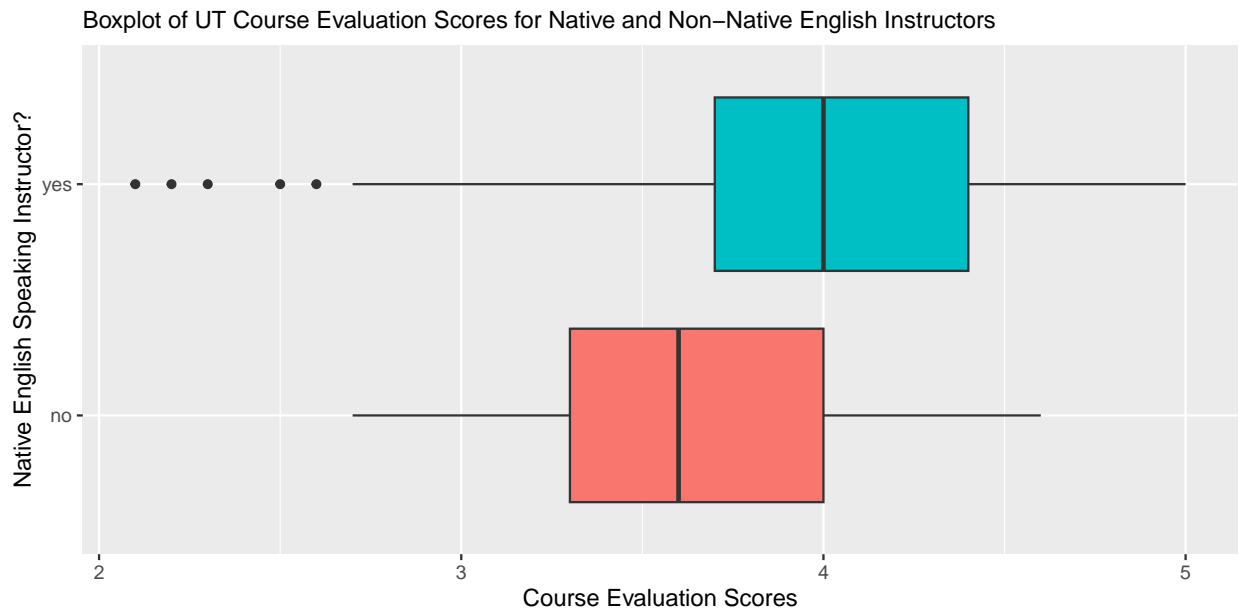
Problem 1: Beauty, or not, in the classroom

Part A



This histogram displays how many instructors received what grade in UT course evaluation scores, with the score received being on the x-axis and the count for each on the y-axis. This graph is skewed to the left, with a mean of 3.9983, and a slightly higher median of 4, meaning most UT instructors, on average, receive a score of a 4 on their course evaluations.

Part B



This is a boxplot of UT course evaluation scores that instructors have received based on if they are native English speakers or not, with the evaluation score on the x-axis and the top boxplot meaning English is their native language and the bottom boxplot being that English is not. It can be seen that the mean course evaluation score is higher for native English speaking instructors at around a score of 4 whereas the mean score for non-native English speaking instructors is lower at about 3.60.

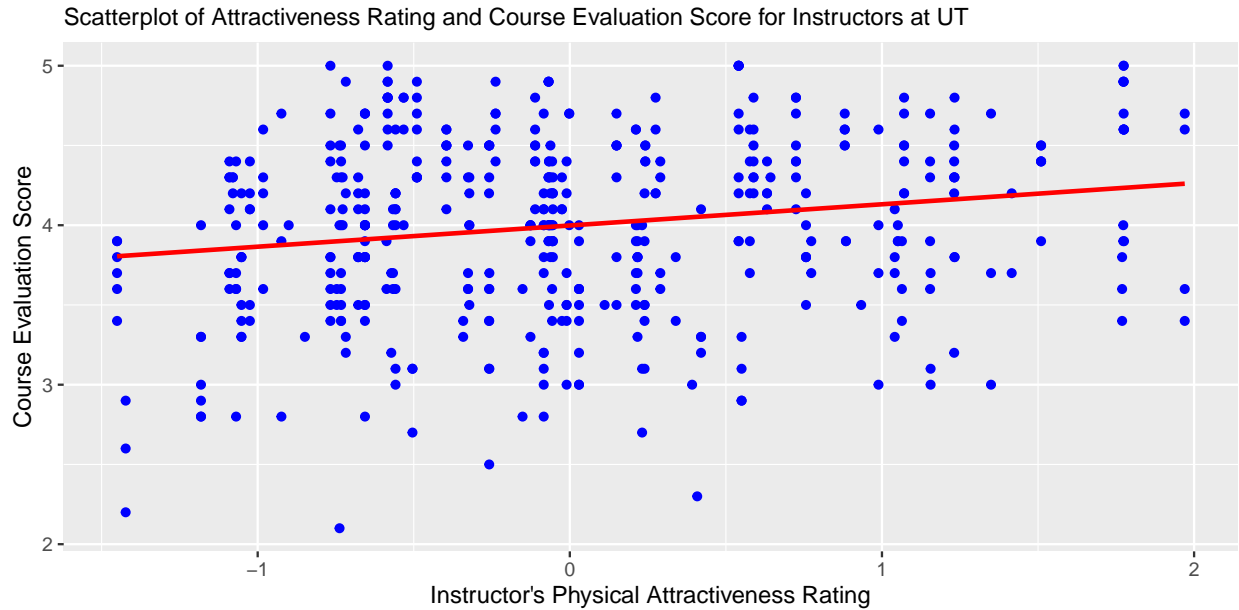
Part C



This is a faceted histogram displaying the course evaluation scores for male and female instructors, with the score distribution for females being on the left and males on the right. The x-axis is the course evaluation score received and the y-axis is the number of instructors that received that score. Both histograms are skewed to the left, but the male histogram has a higher mean score of about 4.07 whereas the female mean

is at about 3.9. The male instructors also had a few lower scores, but in general, received more higher scores than females did, meaning male instructors tended to receive higher course evaluation scores.

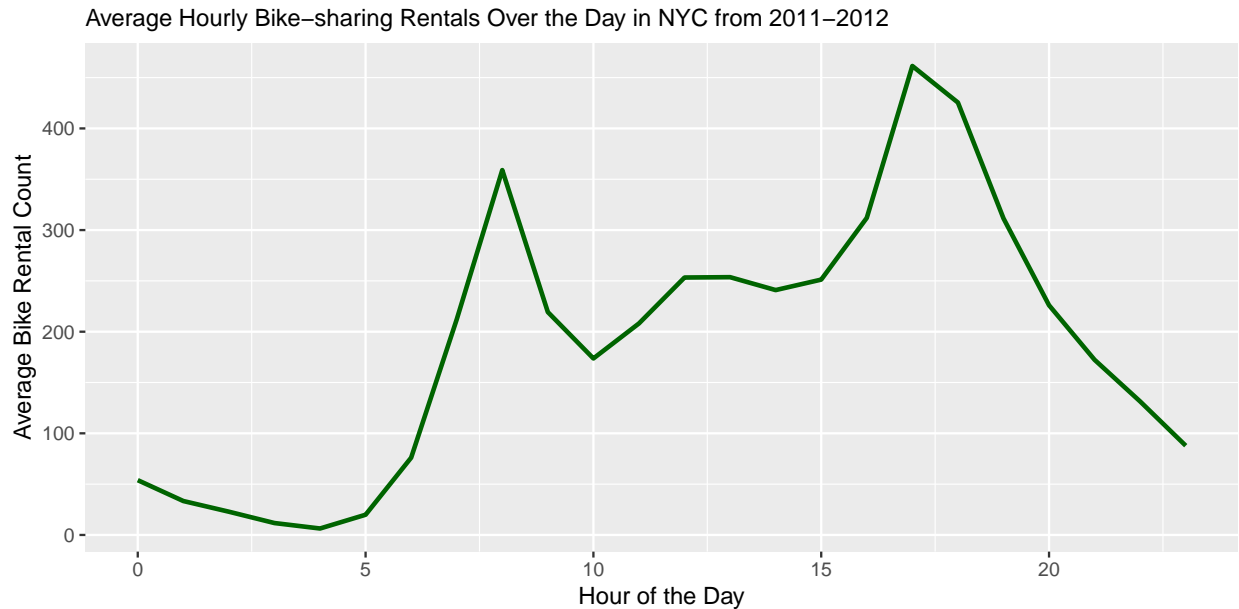
Part D



This is a scatterplot that displays the instructor's physical attractiveness rating (on x-axis) against the same instructor's course evaluation score (on y-axis). It can be seen from the line of best fit (the red line) that there is a very weak, positive correlation between beauty and evaluation score for UT instructors, as the r-value for this scatterplot is 0.19.

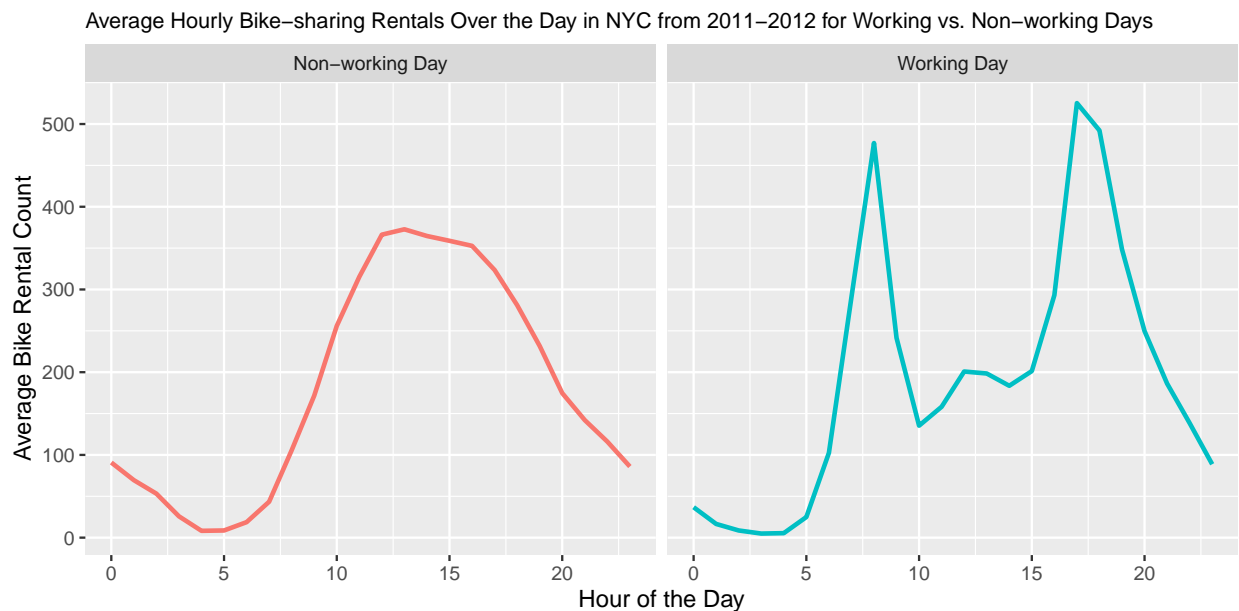
Problem 2: Bike sharing

Part A



This line graph represents the average number of bike rentals for each hour of the day in NYC during all days from 2011-2012, with the x-axis representing what hour of the day it was and the y-axis representing the count of bike rentals. The graph is bimodal, with what appears to be two peaks: one between hours 7-8 and another between hours 16-17, meaning those were the times of day when bike rentals were most popular in NYC.

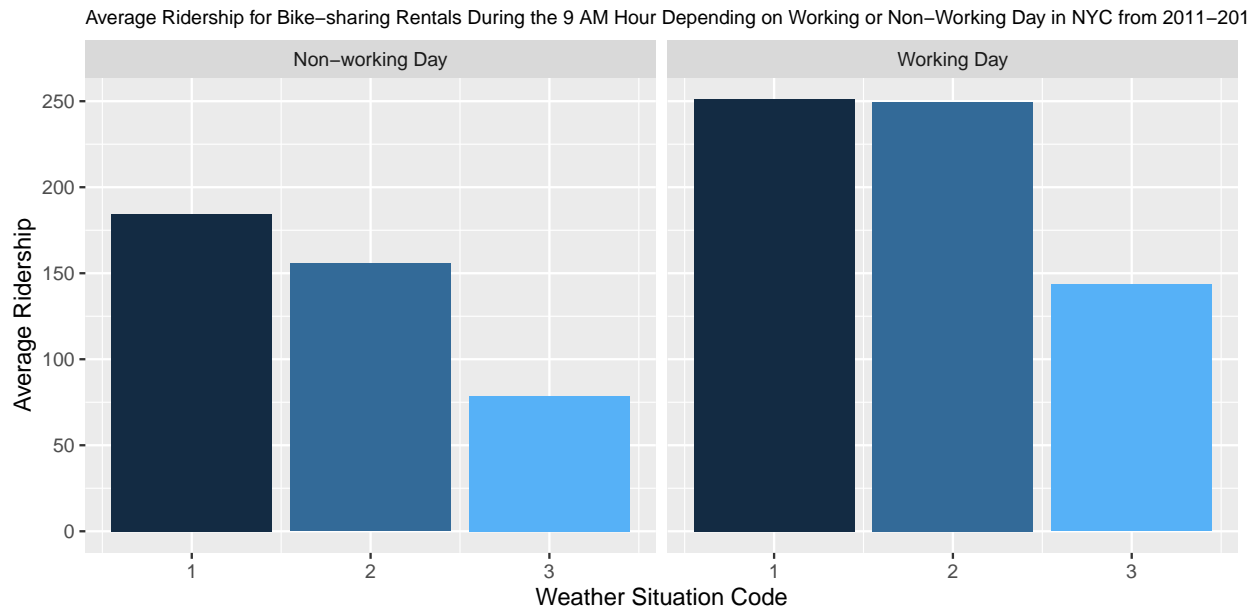
Part B



This is a faceted line graph to show the average bike rental counts (represented by y-axis) across the hours

of the day (represented on x-axis) by if it was a working day (meaning normal weekday) or not (meaning holiday, weekend, etc). It can be seen that non-working days had less average bike rentals, with the max being around 380 rentals and a unimodal distribution with a peak between 11-12. On the other hand, working days typically had more bike rentals with a max of about 530 bike rentals and was bimodal with a peak between hours 7-8 and another between hours 16-17, meaning weekdays around those times were the most popular to rent bikes.

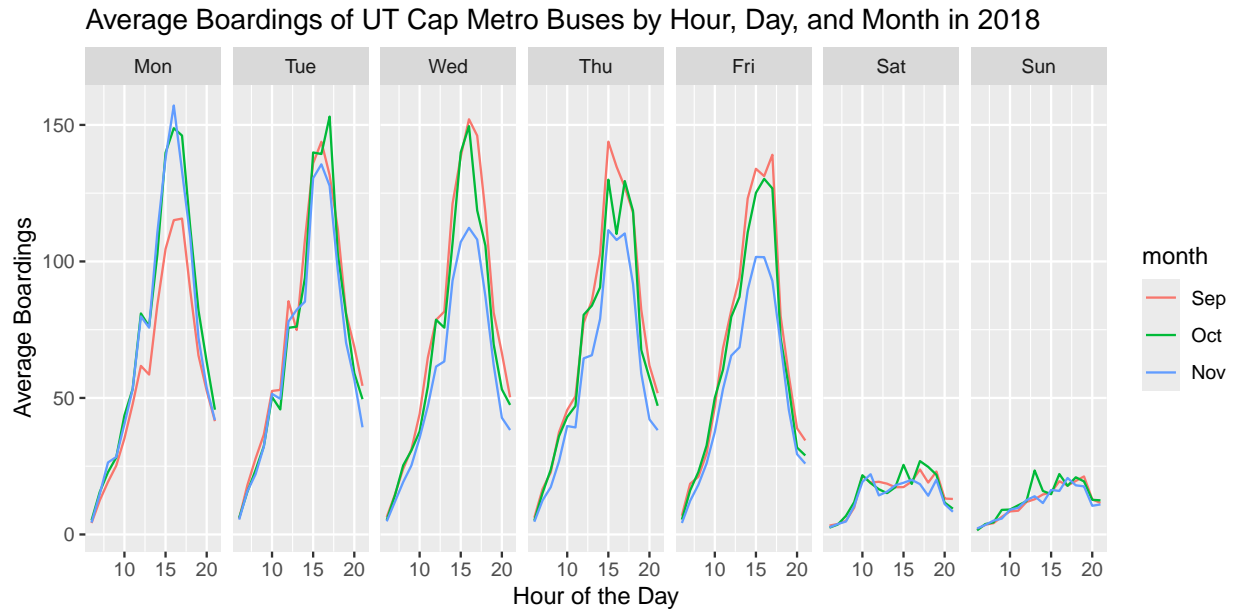
Part C



This faceted bar plot shows the average number of riders during the 9 AM hour depending on if it was a working day or not and how many for each type of weather situation. A weather situation code of 1 means better weather (clear, few clouds, etc.) whereas a 3 meant worse weather (light rain/snow, thunderstorms, etc.). So working days had higher ridership than non-working days, and for both, there was more ridership when the weather was better (when the code was a 1 or 2), and less ridership when the weather was worse (code of a 3).

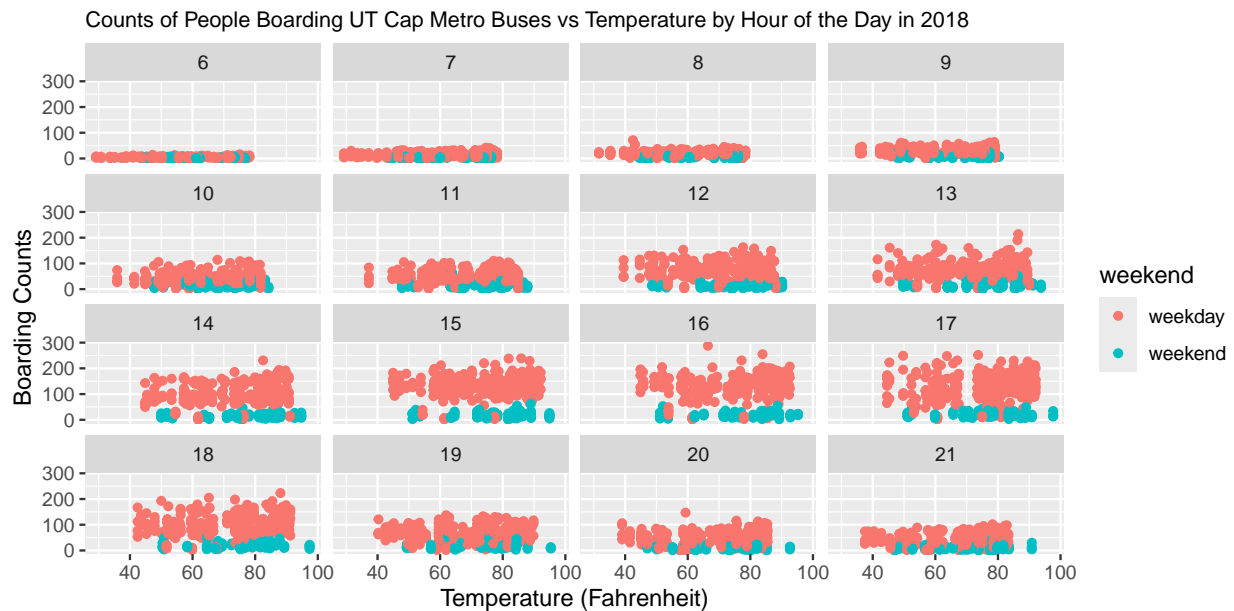
Problem 3: Capital Metro UT Ridership

1.



The faceted line graph displays the average number of boardings onto UT Cap Metro buses according to the hour of the day, colored by the month, and faceted by the day of the week. The hour of peak boarding is generally similar across all the weekdays, but not on the weekends, as it can be seen that on the Mon, Tue, Wed, Thu, and Fri line graphs that the peak boarding time is between hours 17-18, whereas the weekends have different peak boarding times. The average boardings on Mondays in September might be lower because that is generally the start of the school year so new students are not as familiar or comfortable quite yet with using the UT bus system. Average boardings on Wed/Thu/Fri in November might look lower because that is around Thanksgiving so there are most likely less people using the buses then, especially since a lot of people go home for that holiday.

2.



This faceted scatterplot displays the boarding counts of people onto UT buses (y-axis) versus the temperature outside (x-axis), faceted by the hour of the day (from 6 AM to 9 PM) and color coded for if it was a weekend or weekday. When we hold hour of day and weekend status constant, the temperature does not have a very noticeable effect on the number of UT students riding the bus, because for almost every hour of the day, it does not matter what the temperature is at that hour, as there are pretty much an even number of people boarding at that hour whether the temperature is 50 Fahrenheit or if it is 80 Fahrenheit, meaning the temperature does not have a super large role in ridership.

Problem 4: Wrangling the Billboard Top 100

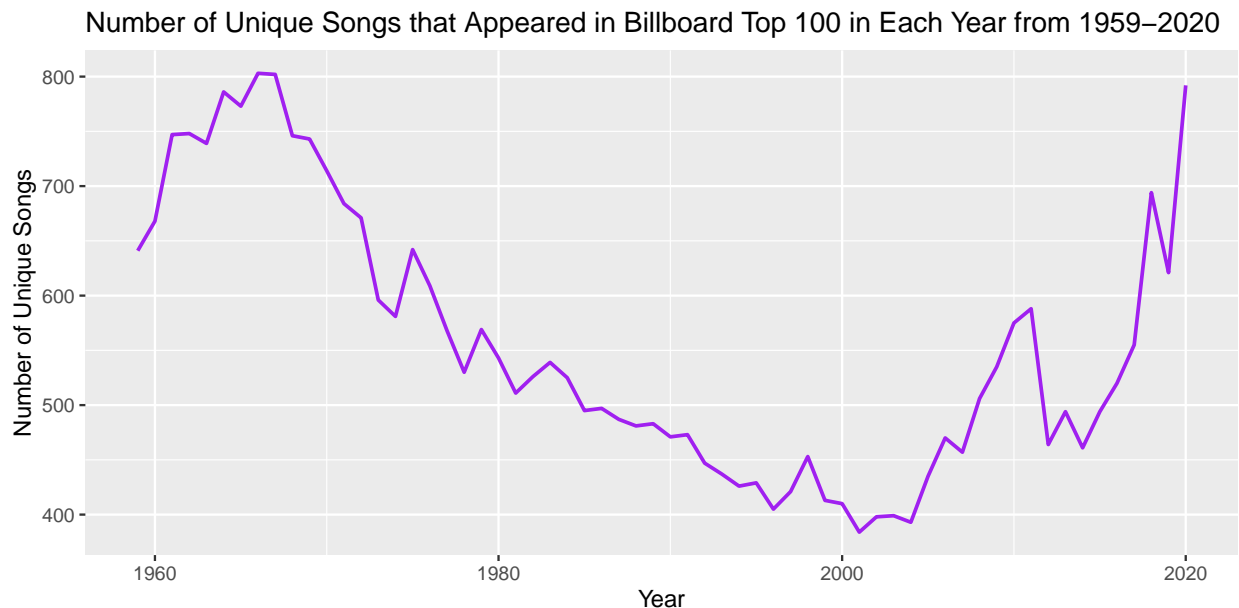
Part A

Table 1: Top 10 Most Popular Songs on Billboard Top 100 Since

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
OneRepublic	Counting Stars	68
LMFAO Featuring Lauren Bennett & GoonRock	Party Rock Anthem	68
Jewel	Foolish Games/You Were Meant For Me	65
Adele	Rolling In The Deep	65
Carrie Underwood	Before He Cheats	64

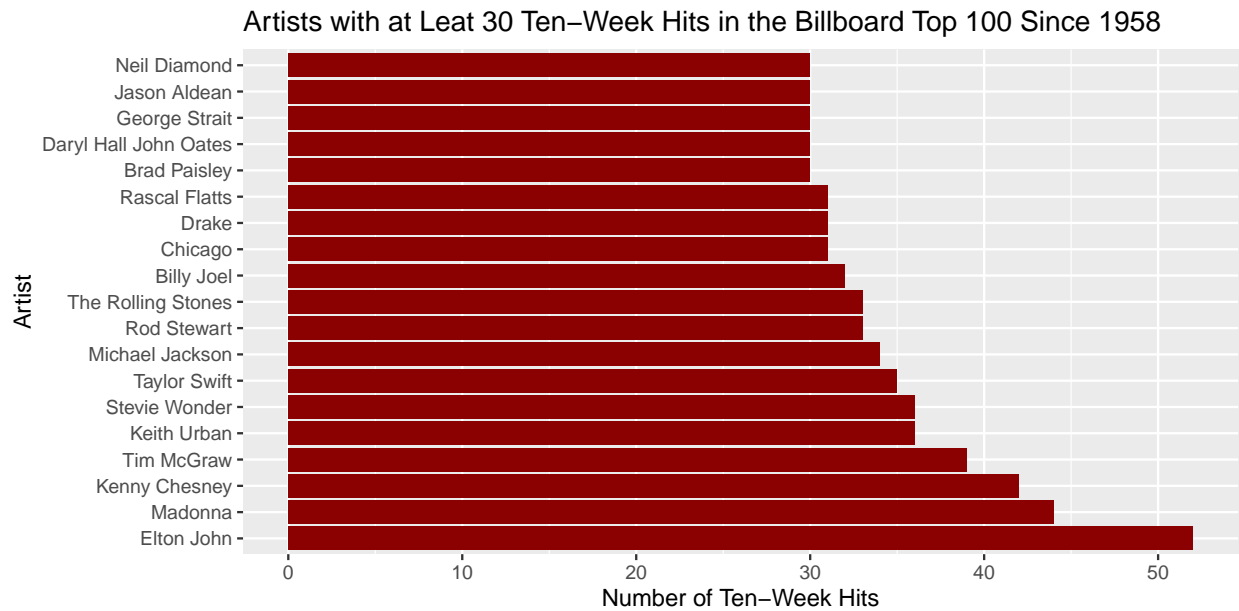
The table shows the top 10 most popular songs on the billboard top 100, which are the songs that have the 10 highest number of weeks in the billboard top 100, as well as the artist of the song and how many weeks it was in the top 100 for.

Part B



The line graph shows the number of unique songs that appeared in the top 100 each year over time, with the x-axis being the year and the y-axis being the count of unique songs that year. It can be seen that between 1960-1970, there were a lot more unique songs in the billboard top 100, with the peak being at 803 unique songs in the billboard top 100 in one year. On the otherhand, around the 2000s, there was an all-time low of 384 unique songs in the top 100 in one year. It then rose again, with 2020 having close to as many as the 1960-1970s did.

Part C



The bar graph shows artists that have at least 30 ten-week hits in the billboard top 100 since 1958, with there being 19 artists total. Elton John has the most by far, with 52 ten-week hits. Madonna is the next closest with 44 ten-week hits in the billboard top 100 since 1958.