

HW 2

Denys Osmak

2024-01-25

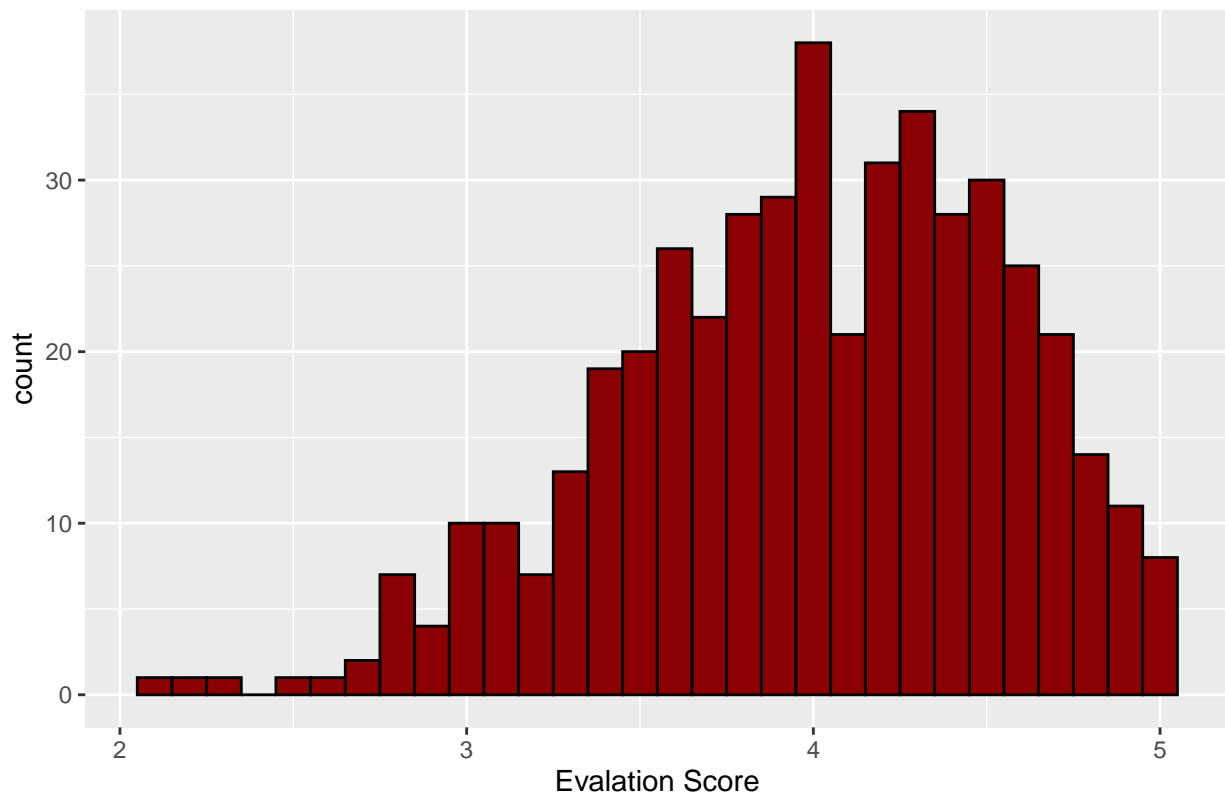
Github link:

<https://github.com/DenysUkr/SDS-315-GitHub-Repo/tree/main/Homework/HW%202>

Problem 1

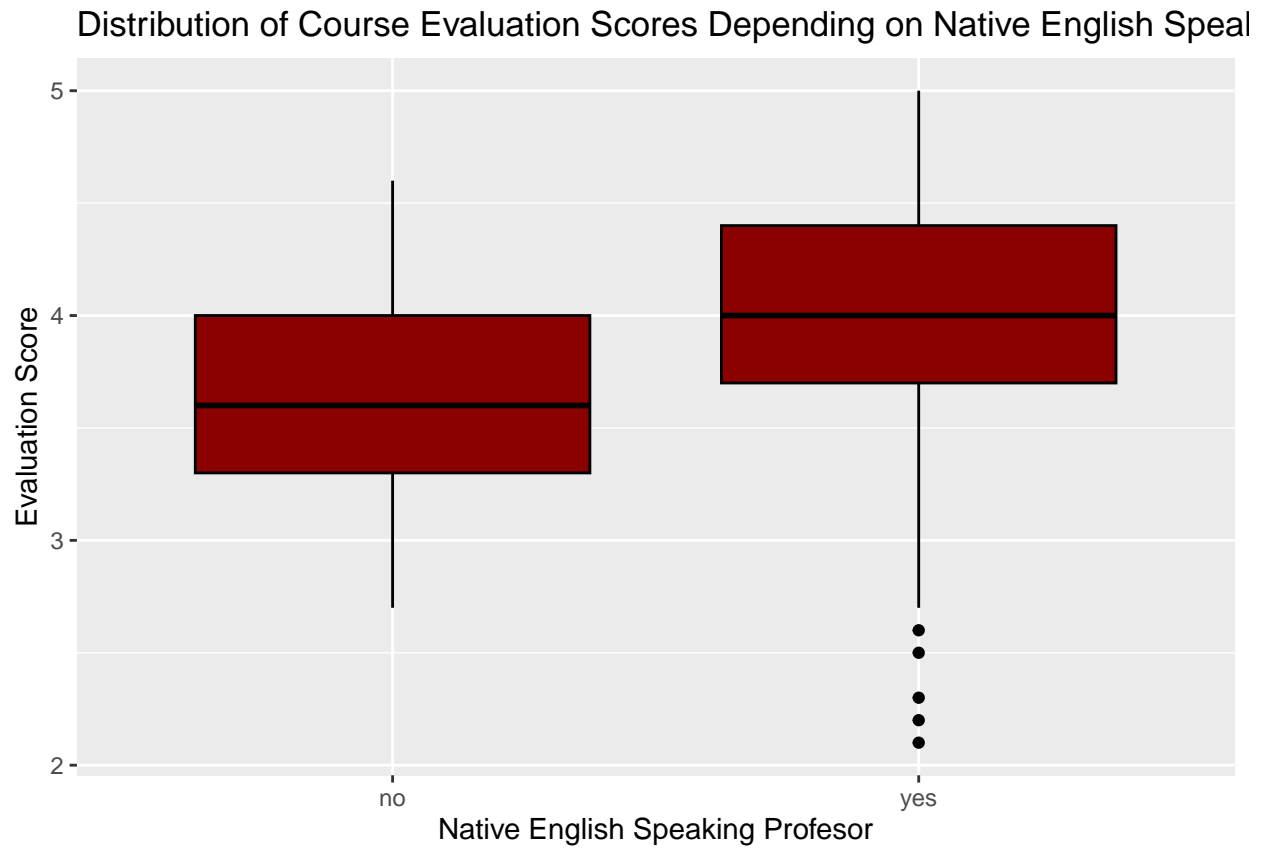
Part A

Distribution of Course Evaluation Scores



We can tell from this graph that the largest majority of the professors are evaluated at around the 4/5 mark. While no professors are marked to be 2/5 and below.

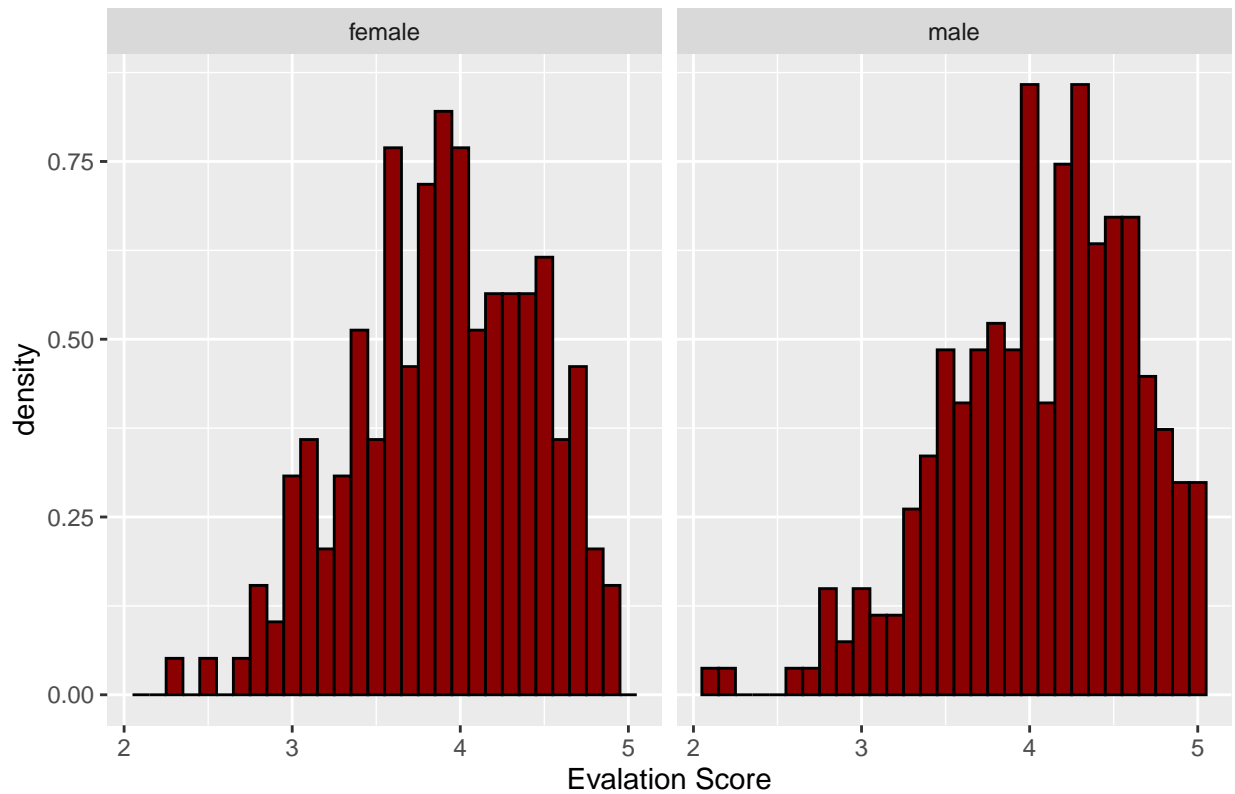
Part B



This box plot shows that there is a significant difference in evaluation scores between professors that are native English speakers vs that are not. From the graph we can see that on average a native English speaking professor gets ~ 0.5 higher score than a non native English speaker.

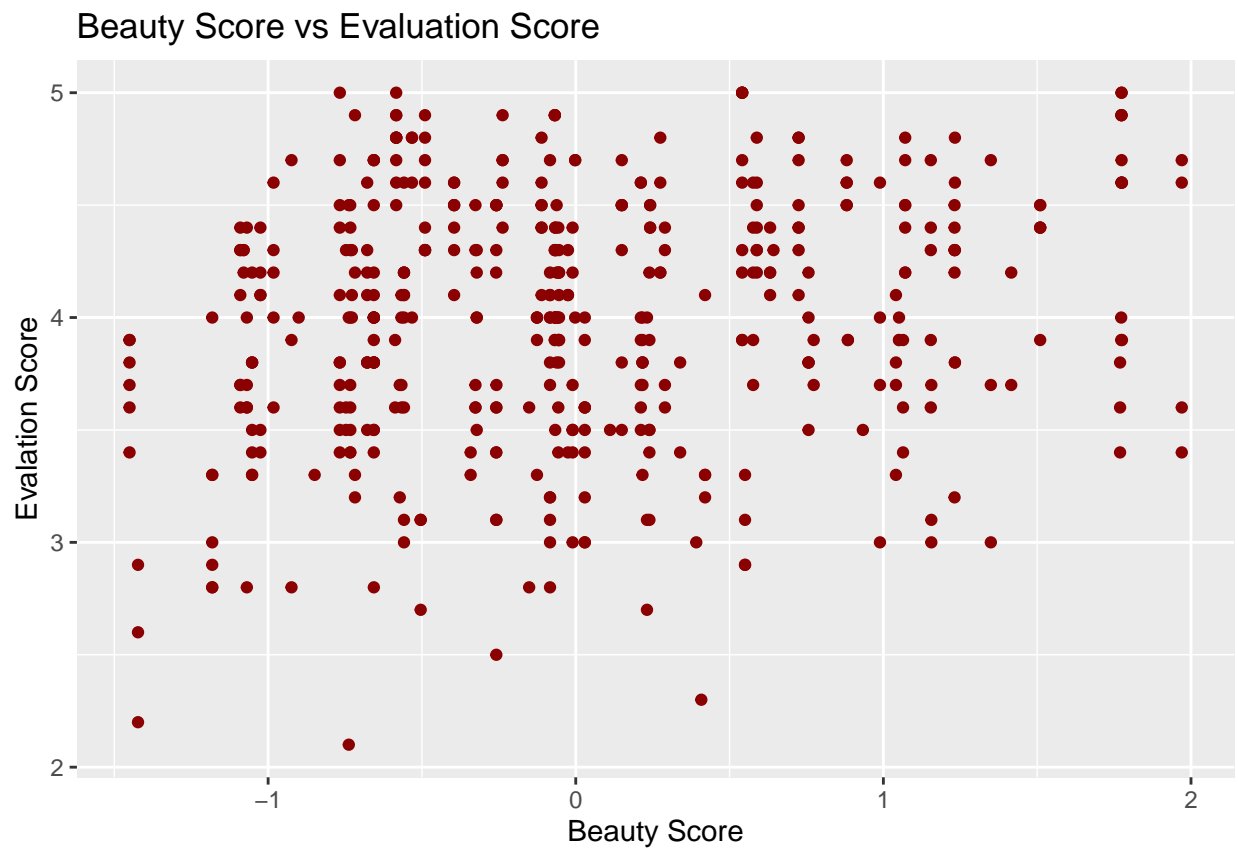
Part C

Desity Histograms of Course Evaluation Scores vs Gender



Due to the fact that there seems to be a larger number of male professors, to see the true difference in the distribution of the evaluation score between male and female professors, the charts were normalized. After normalizing the charts it can be seen that while the male and female professors have relatively same evaluation scores, most of them being between 3.5-4.5, there is an abnormal ammount of female professors that are rated 3.5 and male professors at 4.5.

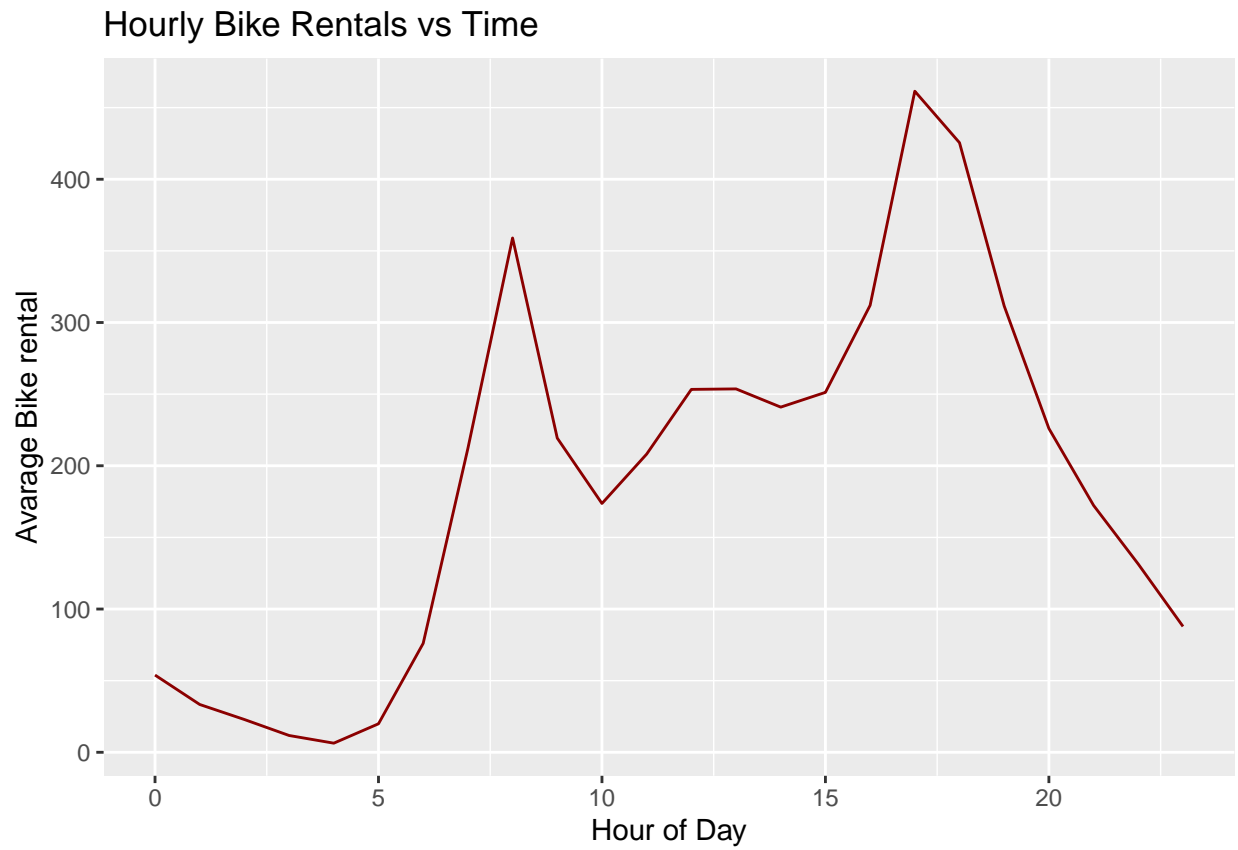
Part D



From the scatter plot there seems to be little to no correlation between beauty score and the evaluation score.

Problem 2

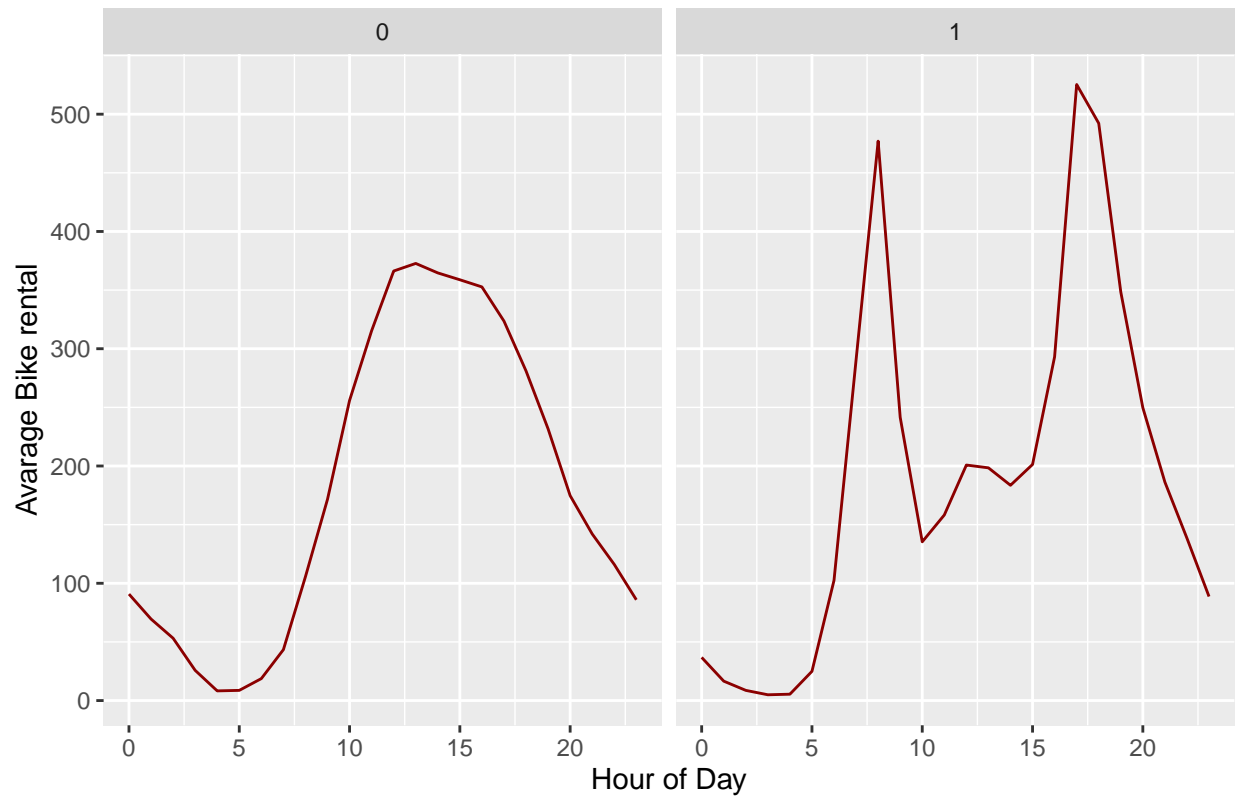
Part A



By looking at this graph we can conclude that there is a large spike of bike rentals during the rush hours around 7 am and 6pm when people are commuting to and from work.

Part B

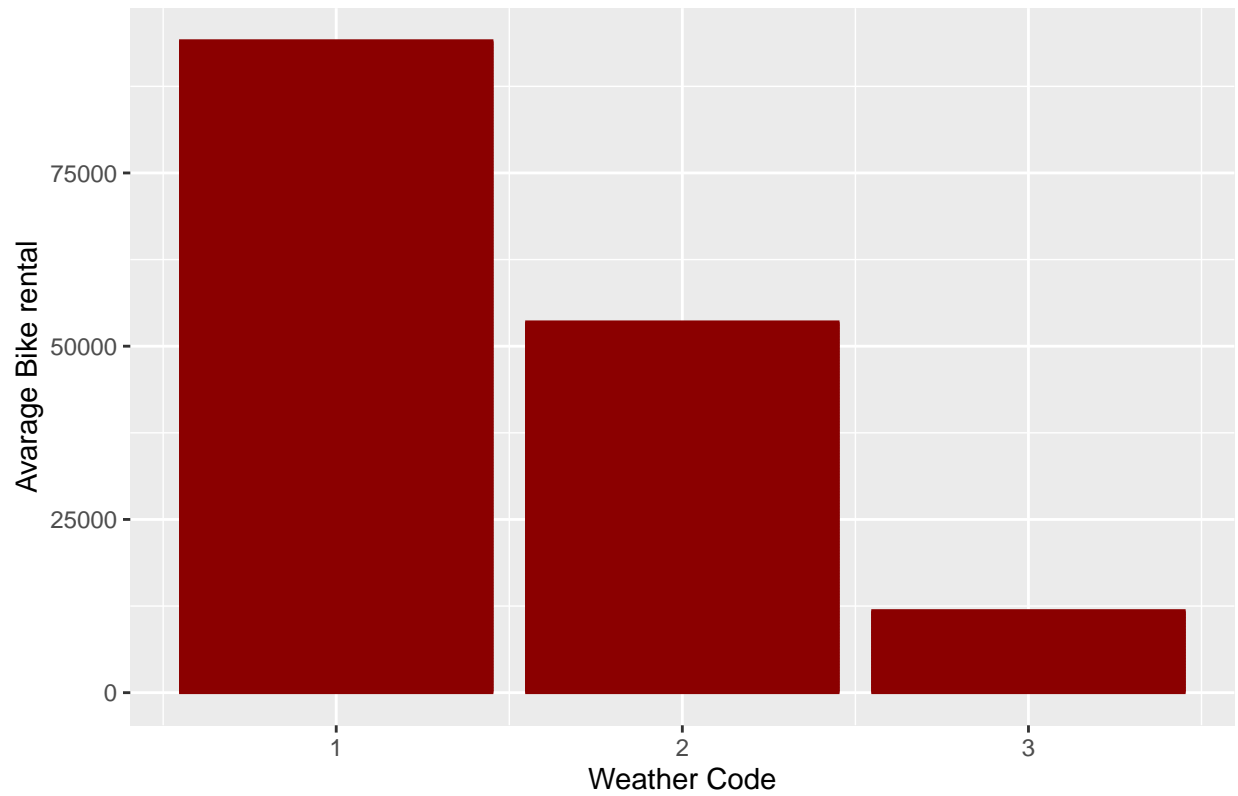
Hourly Bike Rentals vs Time



From the graph we can see that the weekends (identified as 0) have a much more relaxed curve. This is due to the fact that during the weeked the bikes are used for pleasure rather than for commuting to work.

Part C

Hourly Bike Rentals at 9am vs Weather Code

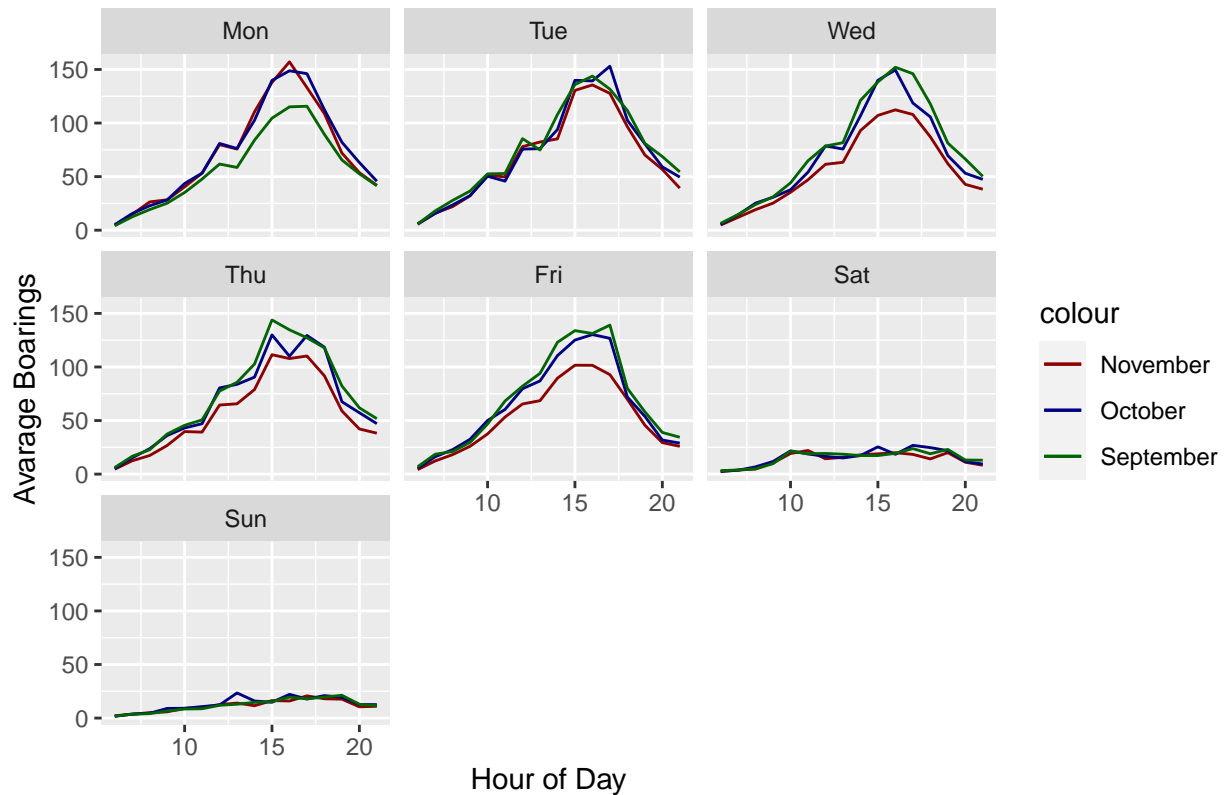


We can see that the average ridership goes down when the weather is cloudy or has light rain (type 2 & 3 weather codes). Additionally there is no recorded ridership at 9am during heavy rains.

Problem 3

Part A

Avarage Hourly Boarding per hour over 3 months



```
##### TESTING NEW MEATHODS #####
```

```
# Tesiting out an all in one solution
```

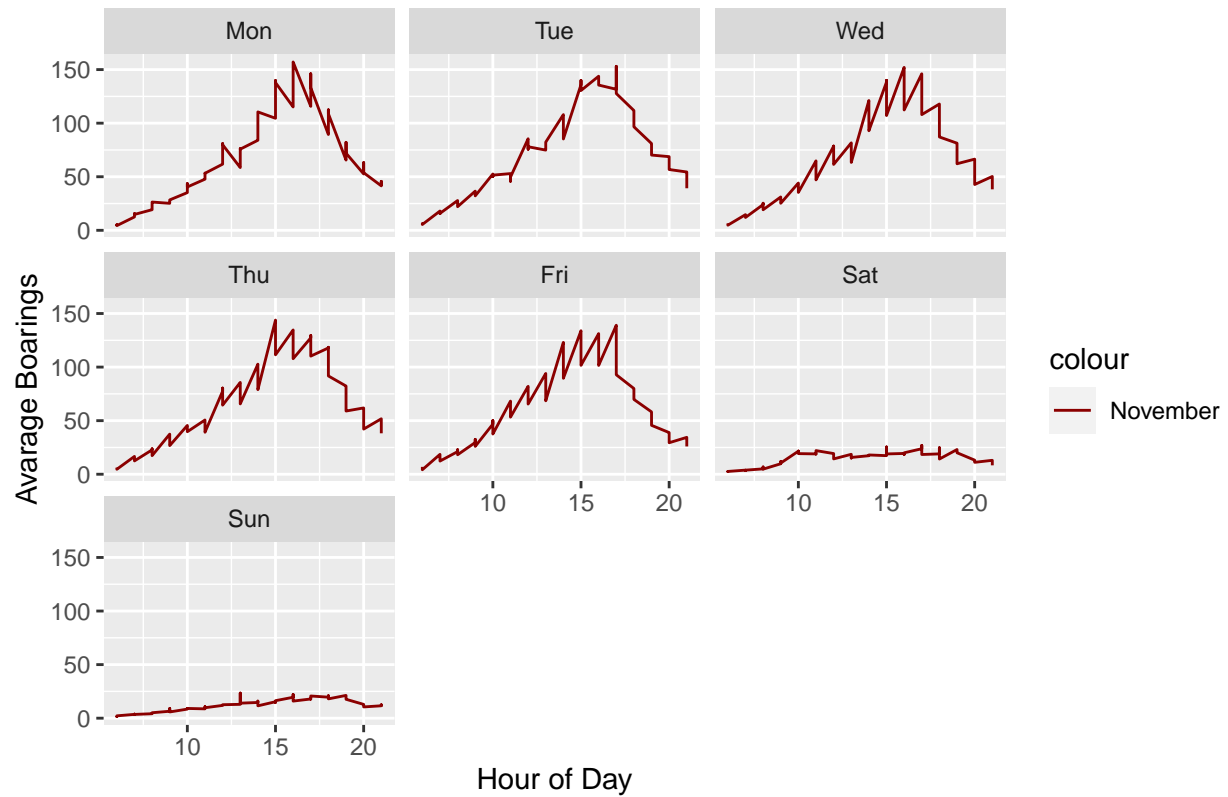
```
pr3DataTesting = pr3Data %>%  
  group_by(day_of_week, hour_of_day, month) %>%  
  summarise(avgBoarding = mean(boarding))
```

```
## `summarise()` has grouped output by 'day_of_week', 'hour_of_day'. You can  
## override using the `.groups` argument.
```

```
# creating a line graph
```

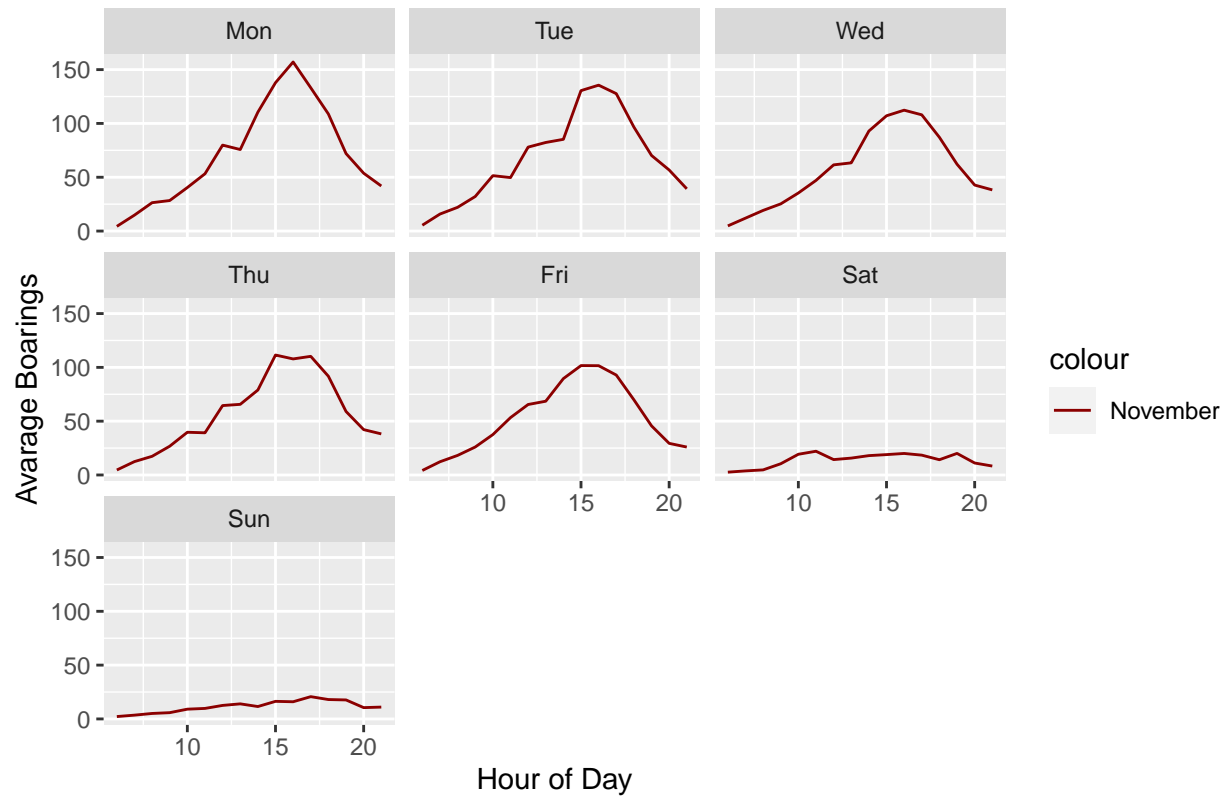
```
ggplot(pr3DataTesting) + geom_line(aes(x=hour_of_day, y=avgBoarding, col="November")) + labs(title="A
```


Avarage Hourly Boarding per hour over 3 months



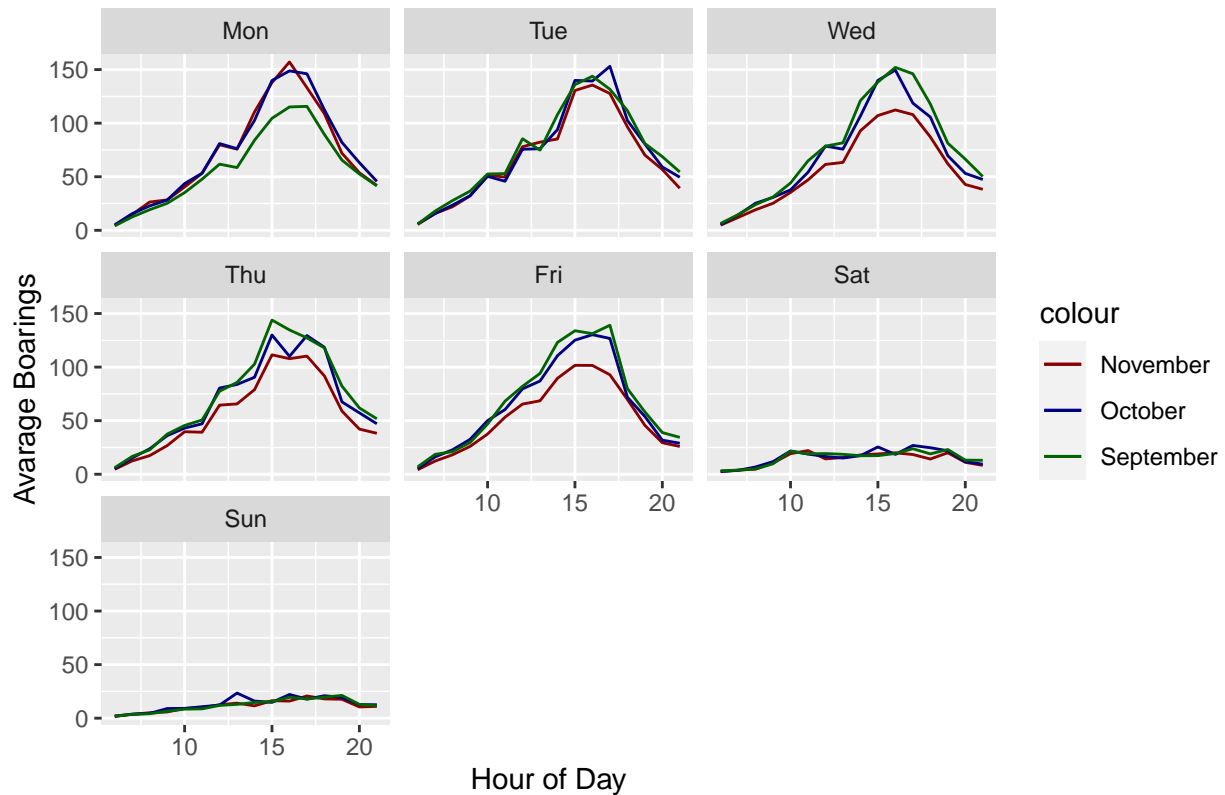
```
ggplot() + geom_line( data=pr3DataN,aes(x=hour_of_day, y=avgBoarding, col="November")) + labs(title="A
```

Average Hourly Boarding per hour over 3 months



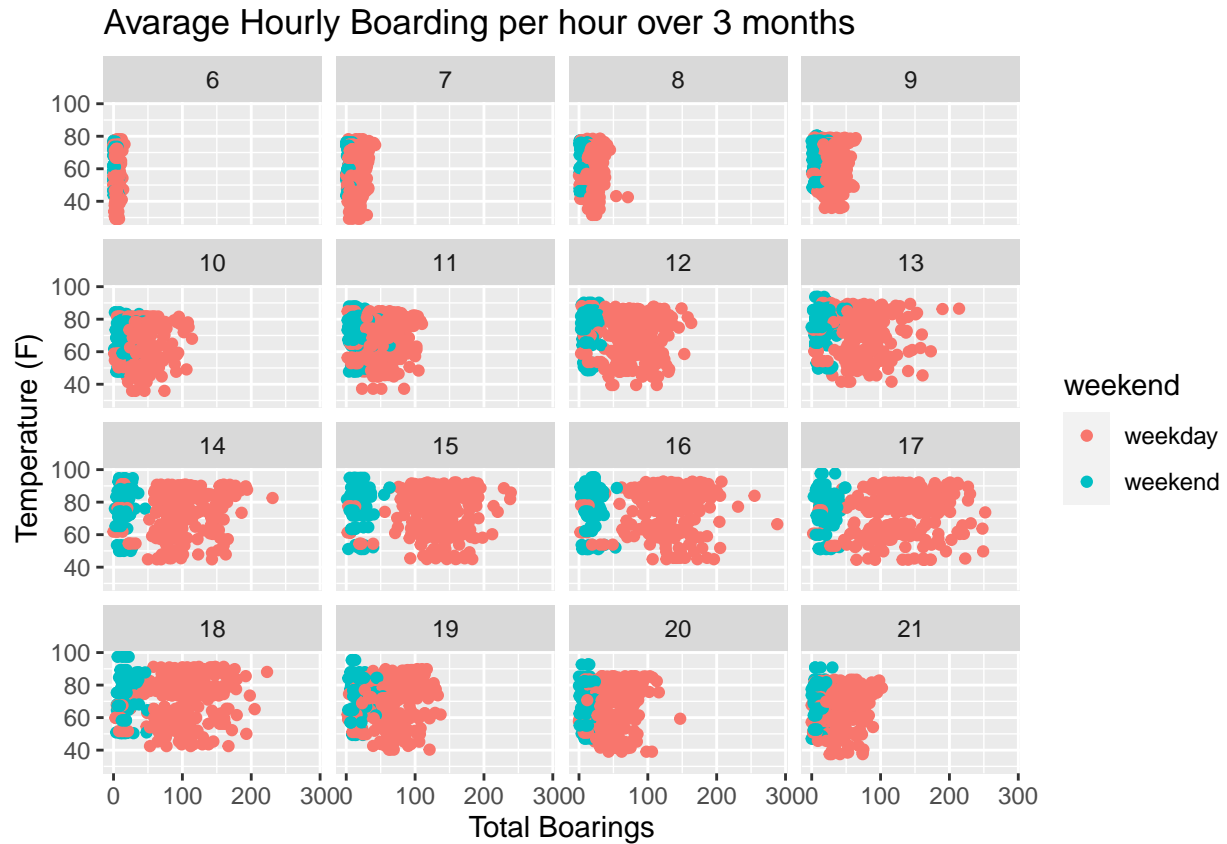
```
ggplot() + geom_line(data=filter(pr3DataTesting, month == "Nov"), aes(x=hour_of_day, y=avgBoarding, colour = "November"))
```

Average Hourly Boarding per hour over 3 months



By looking at the graphs, we can tell that there is the peak for boarding for weekdays are around 3-6pm. The weekdays have very similar curves, no matter the day of the week possibly attributed because the buses are primary used by work commuters. Additional, the reason that November might have the lowest is because it is the coldest month. However the reason why September might have the lowest boarding on Mondays might be due to the fact that there was some sort of holiday on Monday of September of 2018, which lowered the average hourly boarding.

Part B



There is little to no correlation between the temperature and the transit usage, since there is a completely vertical slope as shown on the graph. Interesting enough the weekends tend to have less overall boarding.

Problem 4

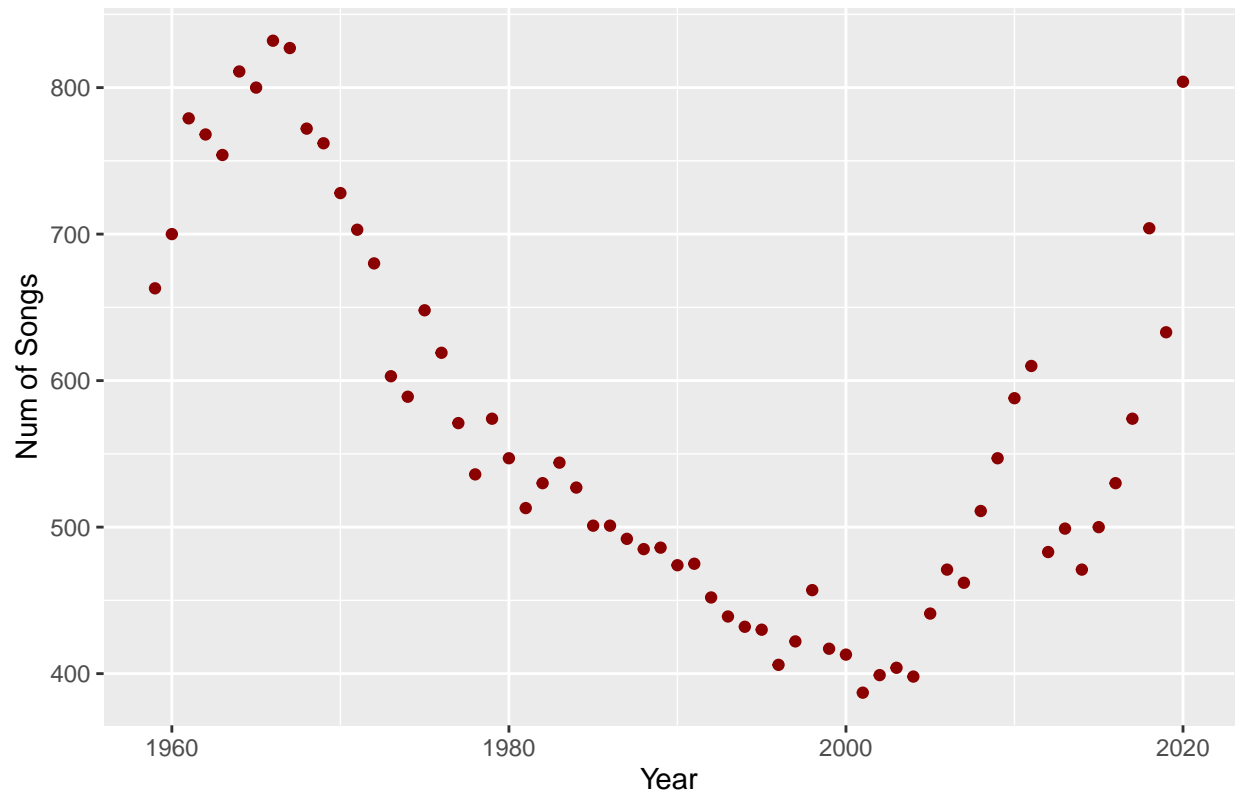
Part A

Table 1: Top Songs

performer	song	totalWeeks
Imagine Dragons	Radioactive	87
AWOLNATION	Sail	79
Jason Mraz	I'm Yours	76
The Weeknd	Blinking 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

Part B

How many unique songs there were each year in top 100



There is an interesting dip how number of unique songs have dropped significantly from 1960 to 2000 and then rapidly picked back up after 2000. This might be to the fact of a music monopoly or some other extreme factor

Part C

Let's define a "ten-week hit" as a single song that appeared on the Billboard Top 100 for at least ten weeks. There are 19 artists in U.S. musical history since 1958 who have had at least 30 songs that were "ten-week hits." Make a bar plot for these 19 artists, showing how many ten-week hits each one had in their musical career. Give the plot an informative caption in which you explain what is shown