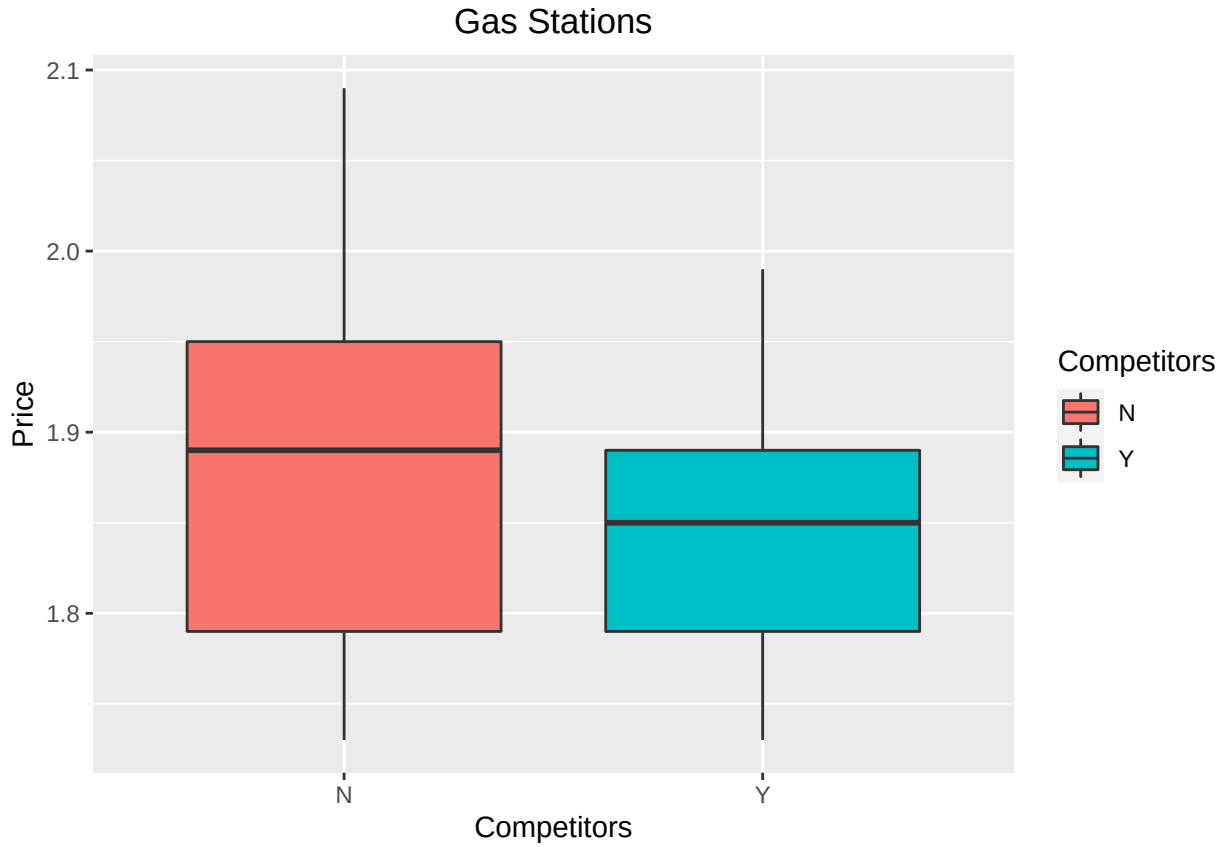
HW 1

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift

##
## Attaching package: 'foreach'

## The following objects are masked from 'package:purrr':
##
##     accumulate, when
```

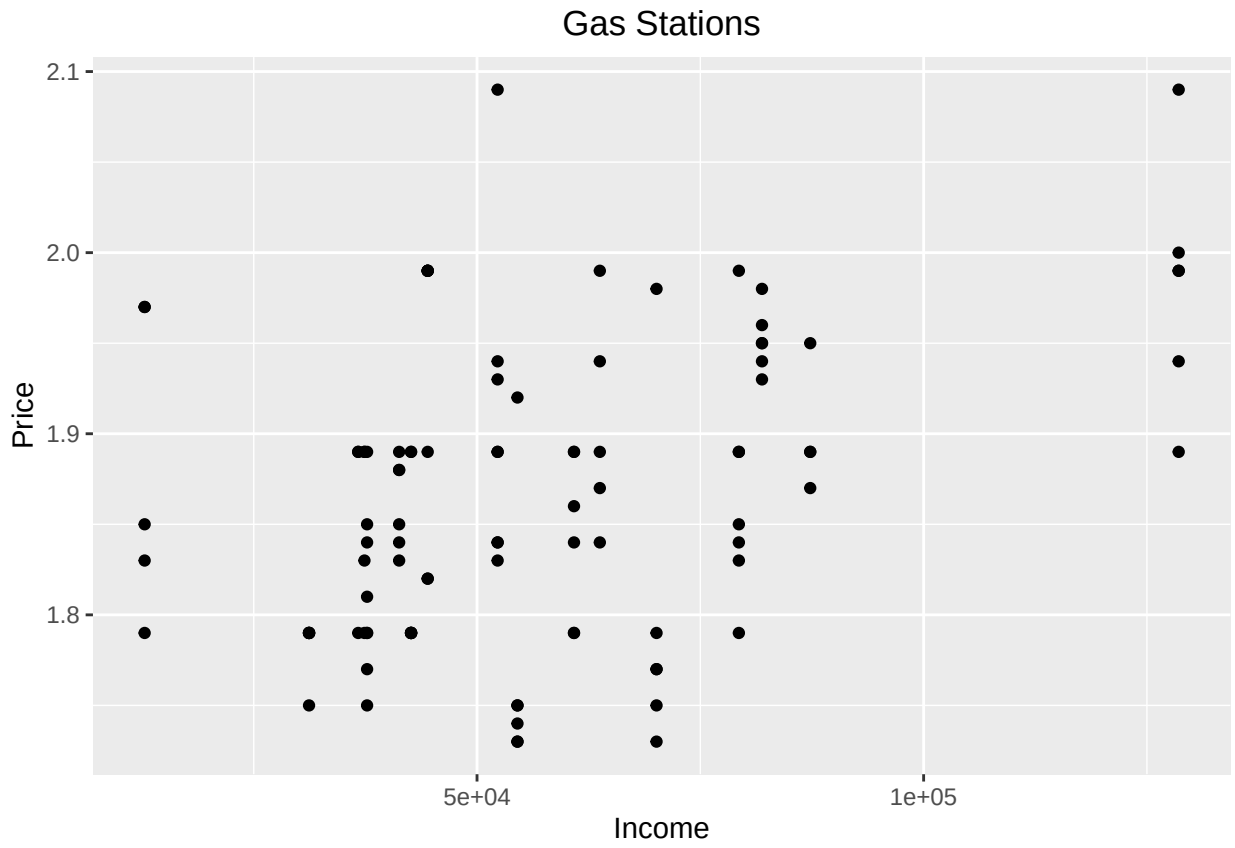
Question 1

Part A



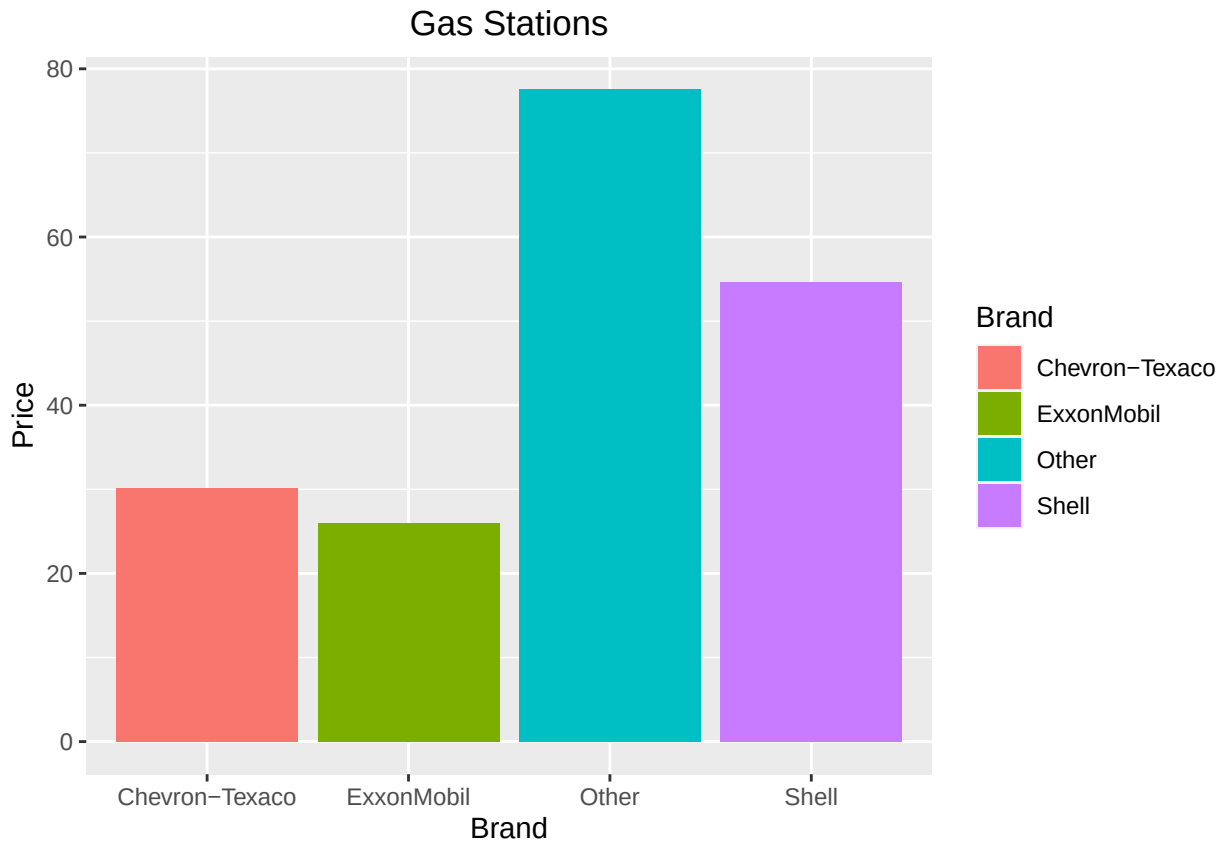
In this boxplot, we are assessing whether the lack of direct competition in sight results in higher gas prices. The boxplot is separated into two categories, yes or no, on whether a competitor is within sight of the observed gas station. Interpreting the data, it becomes clear that the lack of direct competition in sight does result in higher gas prices. The median price of gas stations with no competitors in sight is approximately \$1.89/gallon, while in contrast, the median price of gas stations with competitors in sight is \$1.85/gallon. Moreover, gas stations with no direct competitors in sight have a larger range of prices relative to gas stations with direct competitors in sight.

Part B



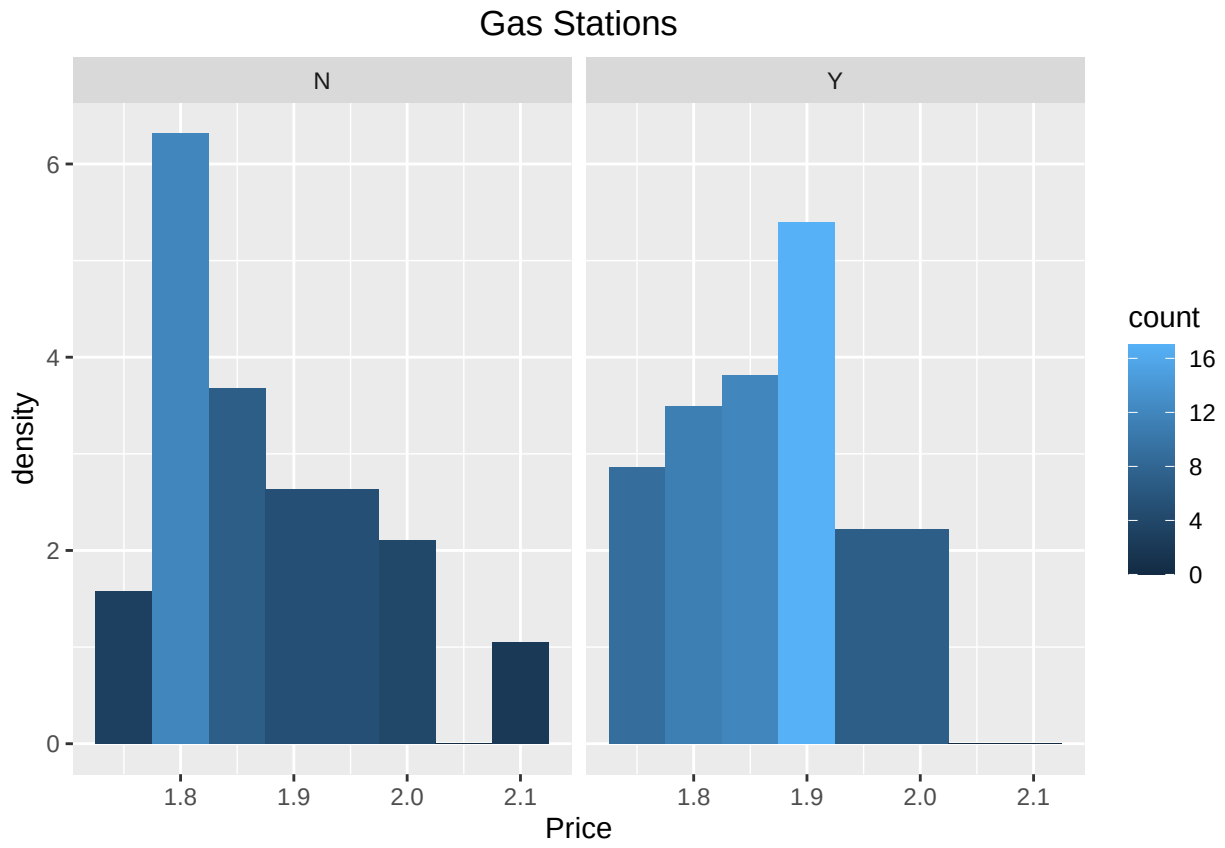
Next, we are determining whether more affluent areas experience higher gas prices. The scatterplot appears to show a positive association between the variables Income and Price, indicating yes, more affluent areas experience higher gas prices. Although, it should be noted that there are several outliers.

Part C



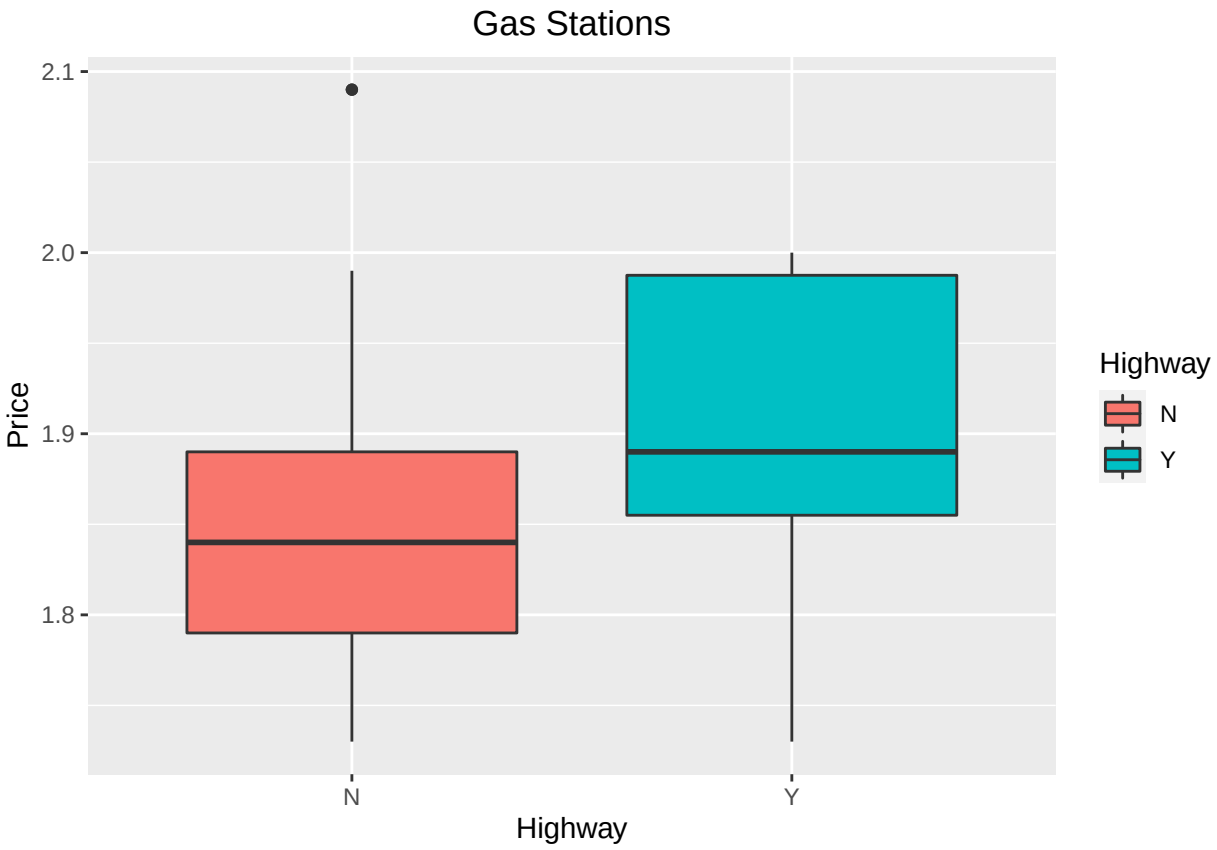
The objective here is to to answer whether Shell charges more for gasoline relative to other brands. Upon observing the barplot, it becomes clear that Shell charges more for gas than either Chevron-Texaco or ExxonMobil, but charges less than Other.

Part D



Next, a faceted histogram is created to assist in determing if whether the precence of stoplights near gas stations increases gas prices. The histogram indicting no stoplights near the gas station has a mean price \$1.80/gallon. Alternatively, gas stations with a stoplight nearby has a mean price of \$1.90/gallon.

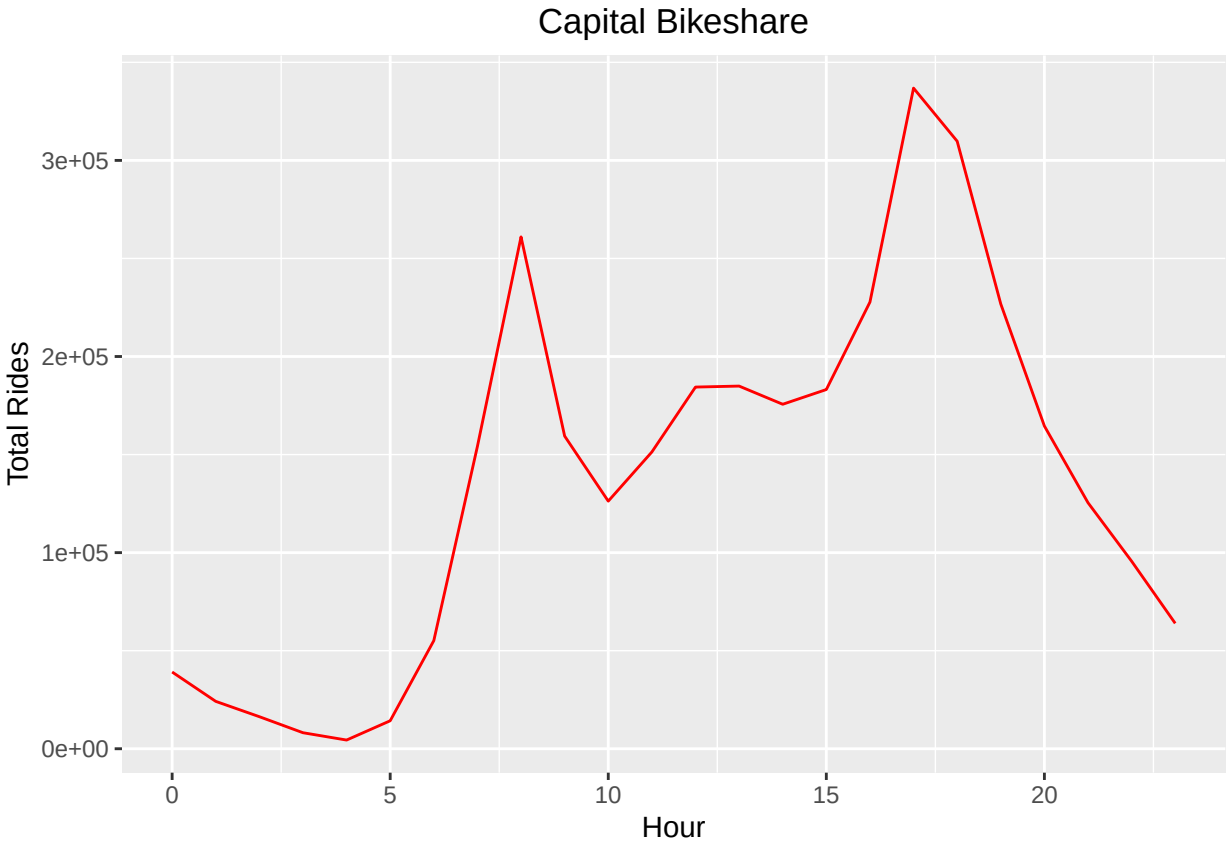
Part E



Lastly, I created boxplots to assist in determining if a gas station has higher gas prices as a consequence of direct highway access. Given the data, it is true that gas stations with direct highway access have higher gas prices. Gas stations with direct highway access have a higher median price of approximately \$1.89/gallon, while in contrast, gas statiosn without a direct highway access have a lower median price of approximately \$1.84/gallon.

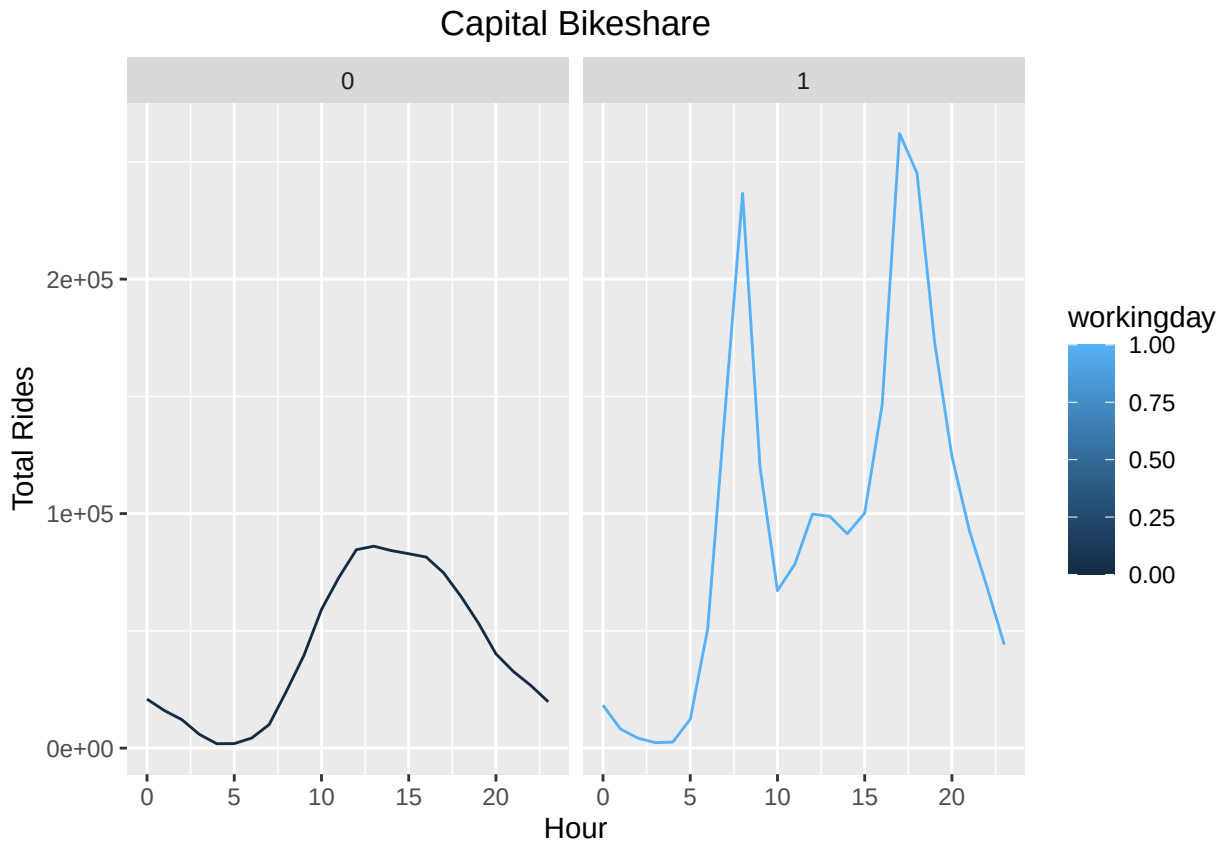
Question 2

Plot A



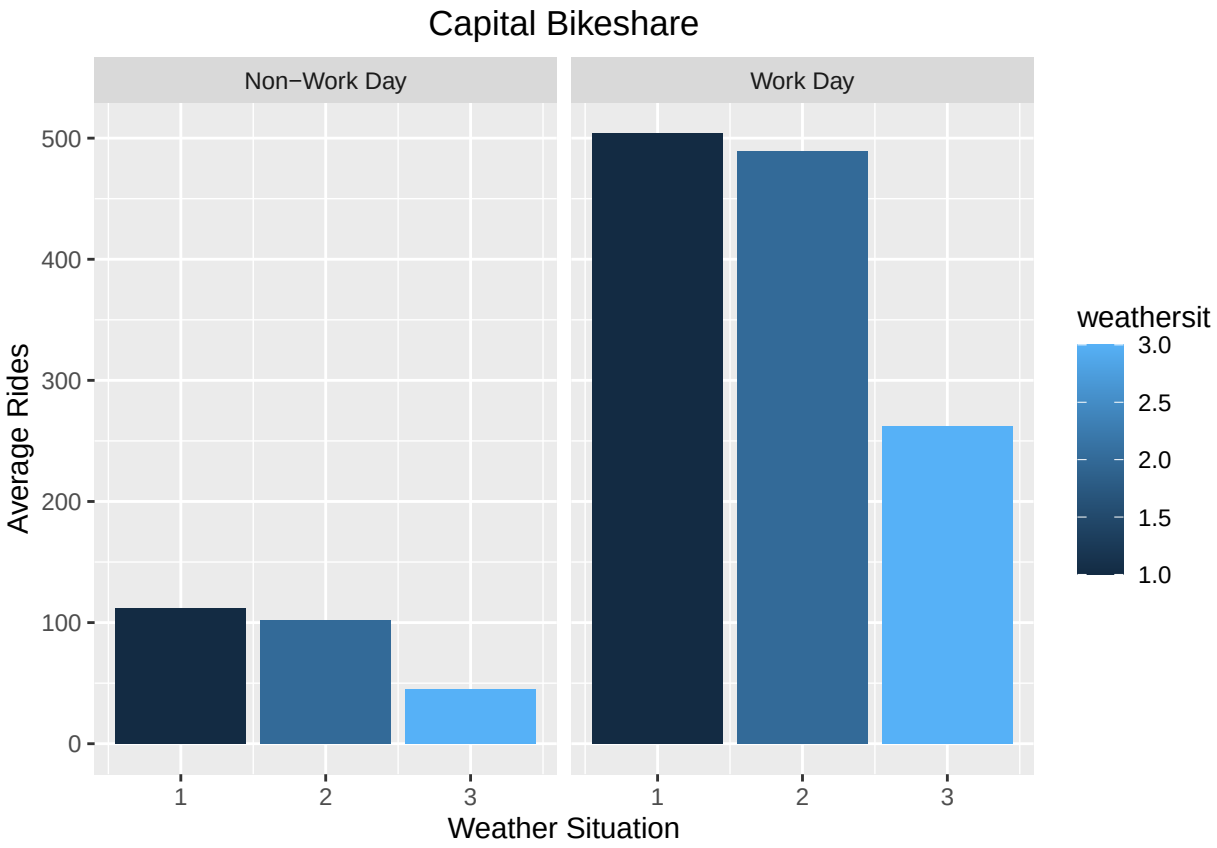
This line graph displays the change in average ridership throughout the day in Washington, D.C. The time of day lays on the x-axis, and the ridership totals lays on the y-axis. The graph displays that the busiest times of day for ridership is during peak morning and evening rush hour.

Plot B



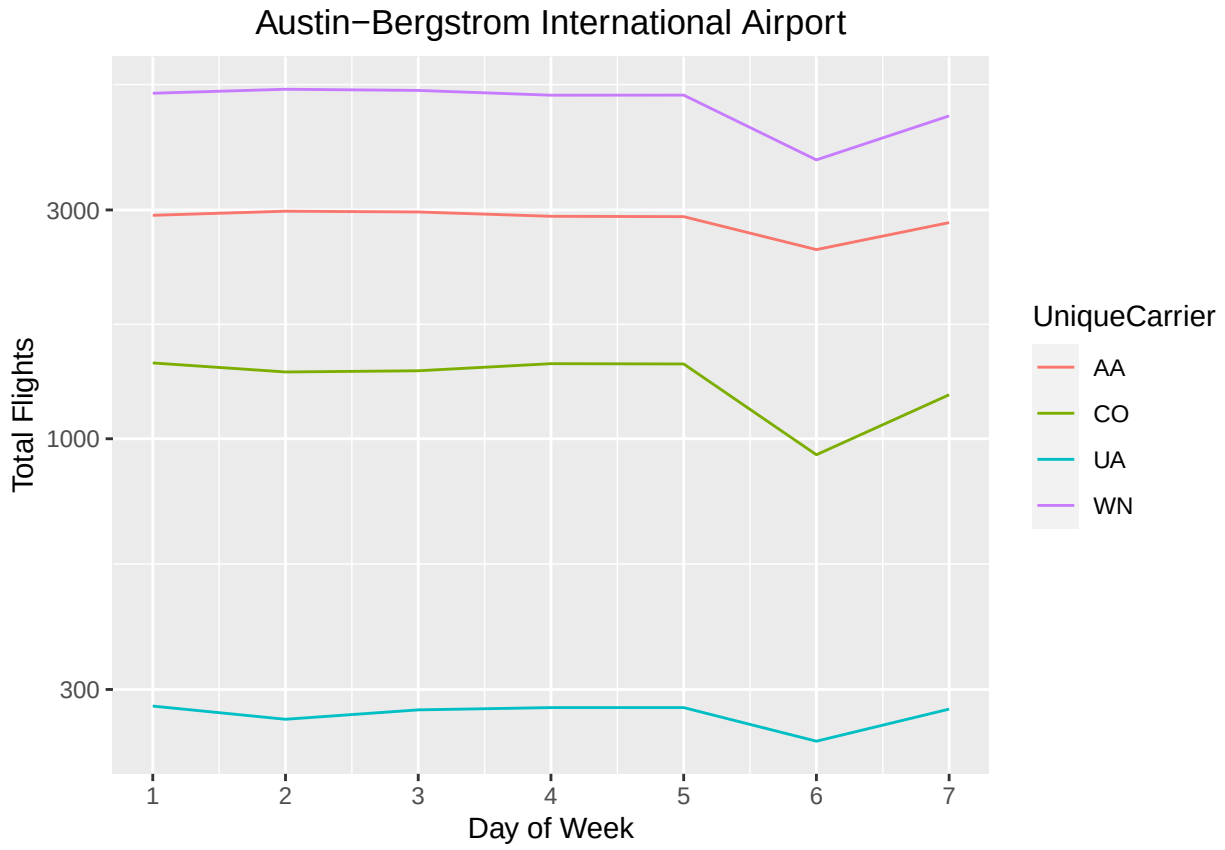
This faceted line graph was made to help determine the differences in ridership between workdays, which have a value of 1, and non-workdays, which have a value of 0. Upon observing either graph, its easy to understand that the highest peaks of ridership occurs during a workday, and at rushhour times. Non-workdays have a peak during mid-afternoons.

Plot C

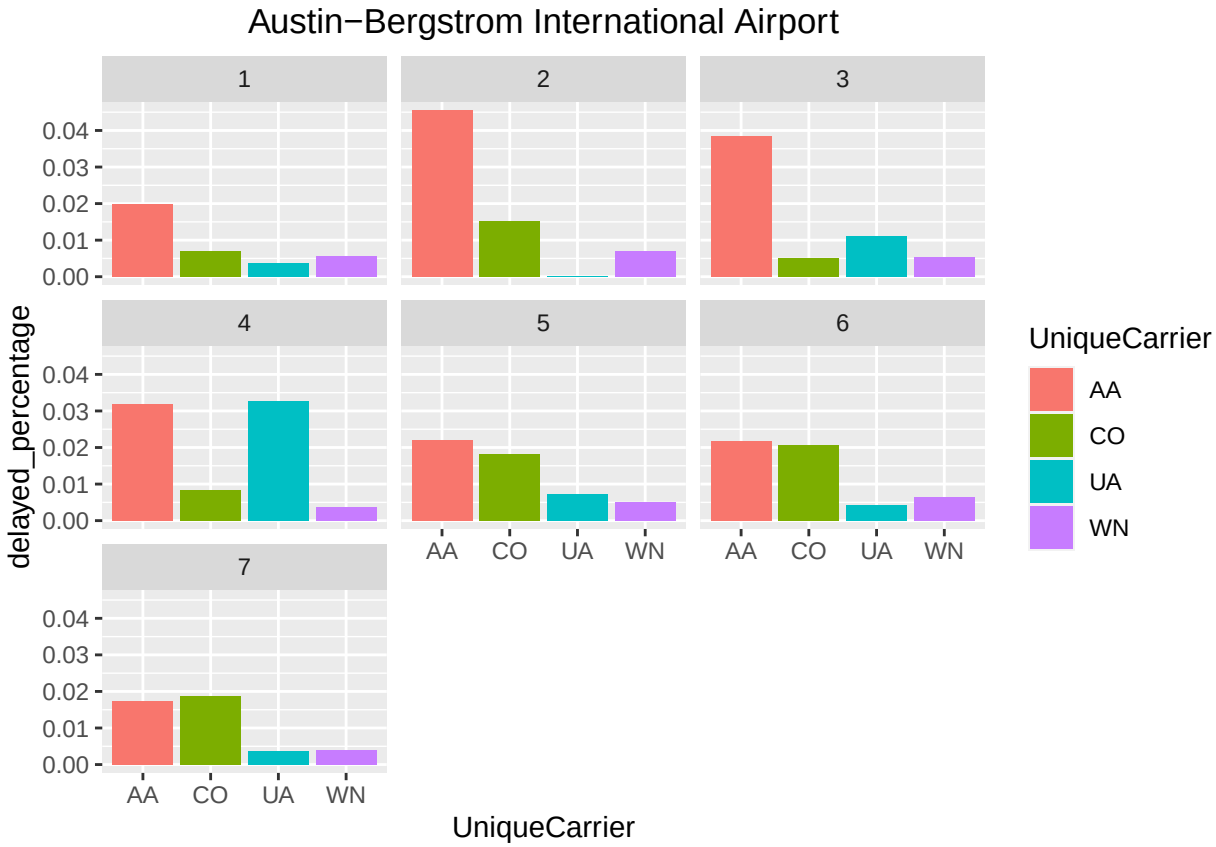


In addition to this faceted bar graph breaking down between non-workdays and workdays, differences of average ridership can also observed in reference to weather conditions. At first glace, its very clear that average ridership is much higher during a workday. In all cases, ridership decreases as the weather worsens.

Question 3

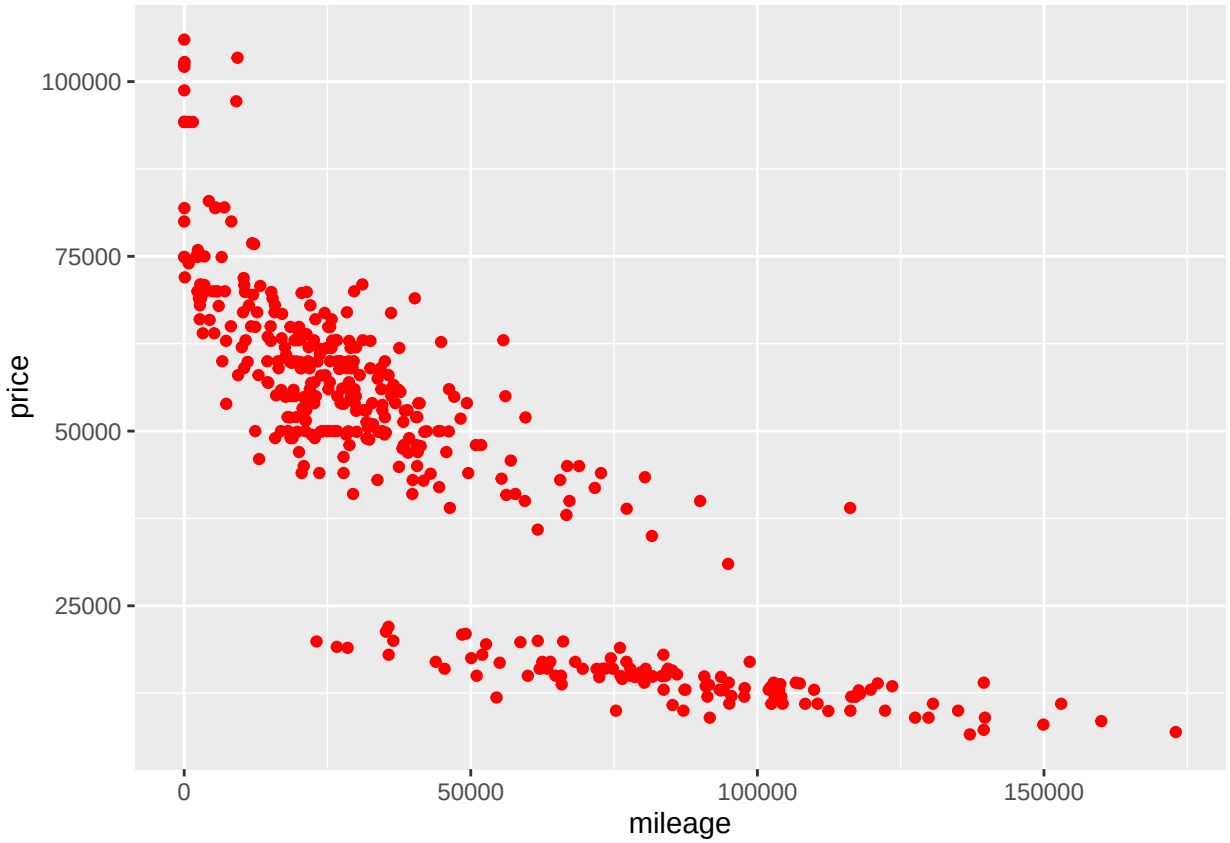


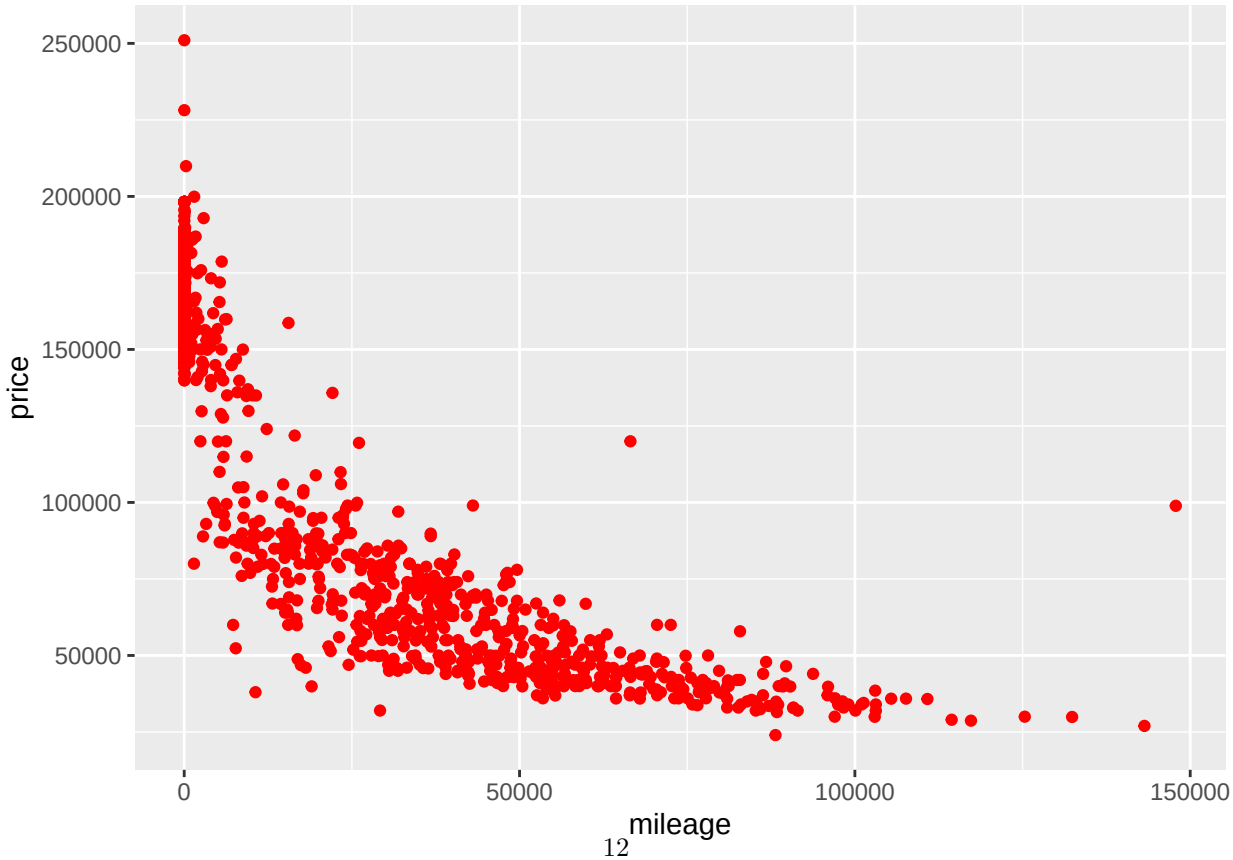
I chose four airlines to observe and compare throughout the week that traveled to and from ABIA. AA (American Airlines), CO (Continental Airlines), UA (United Airlines), and WN (Southwest Airlines). WN has the most flights at ABIA, while UA has, by far, the lowest number of flights. All airlines have a dip in traffic on Saturdays, before increasing again for Sunday. This makes intutive sense because most people tend to fly someplace for the weekend, and return on Sunday for work on Monday.



I then decided to create a bar graph to display the percentages of canceled flights each day of the week for the four airlines listed. One conclusion made is that although WN has the most flights out of ABIA, AA has the highest percentages of cancelled flights on nearly every day of the week, with Tuesday being the worst day to fly on AA, but the best day to fly UA.

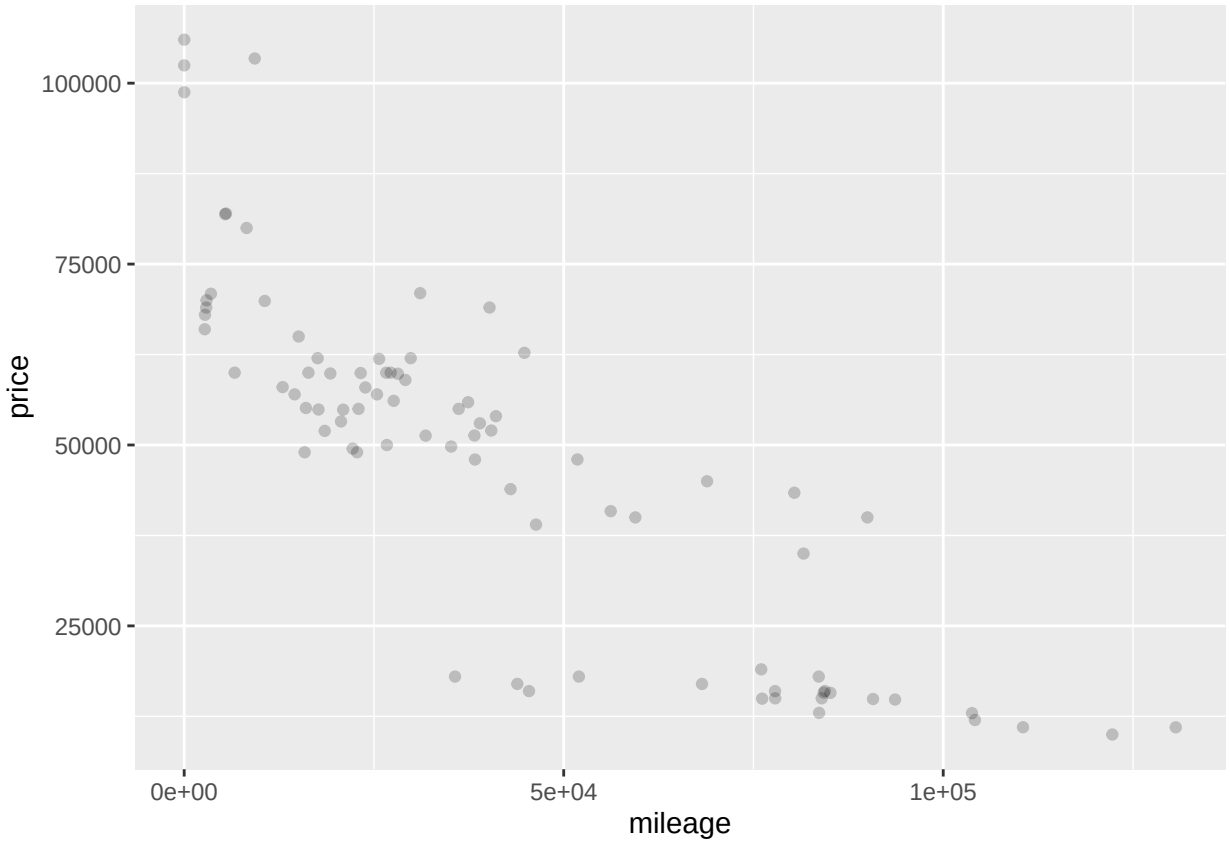
Question 4

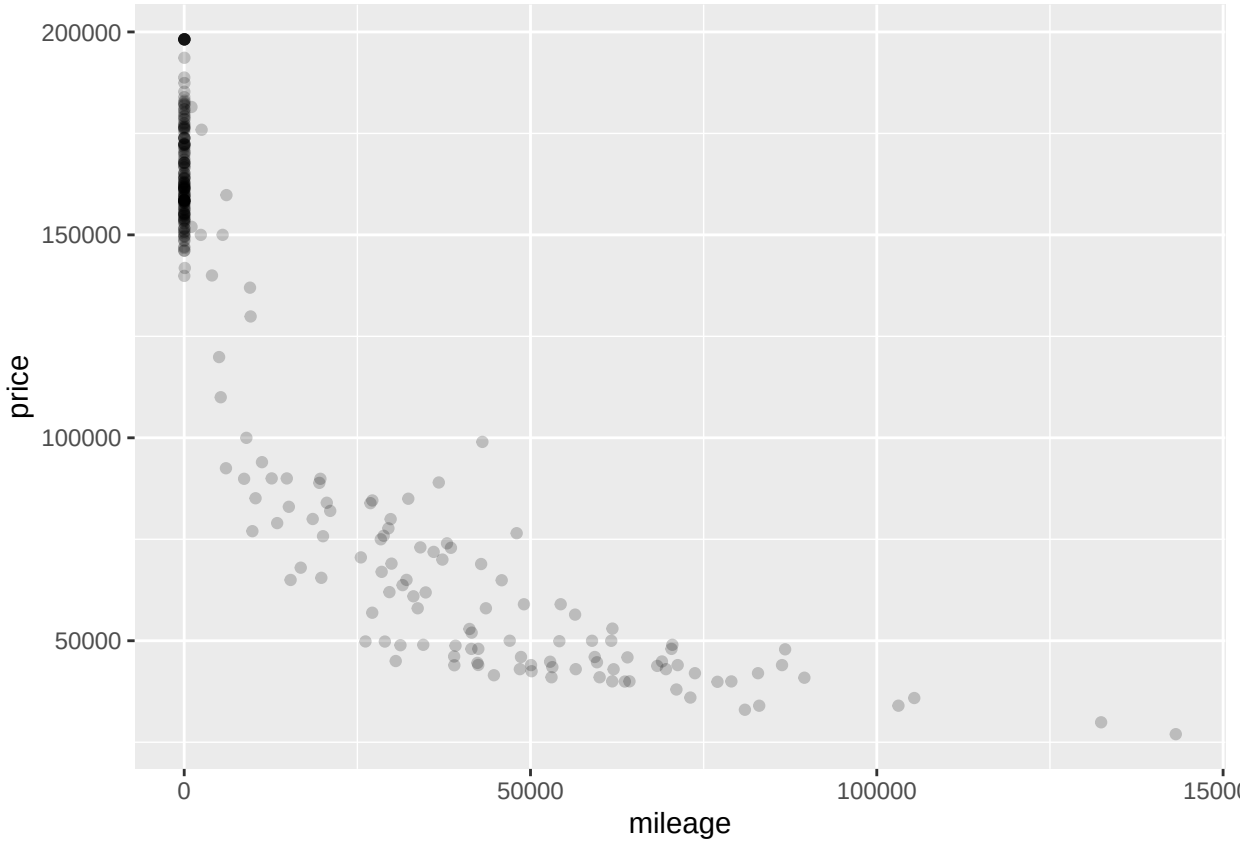

100000
75000
50000
25000
price
0
50000
100000
150000
mileage

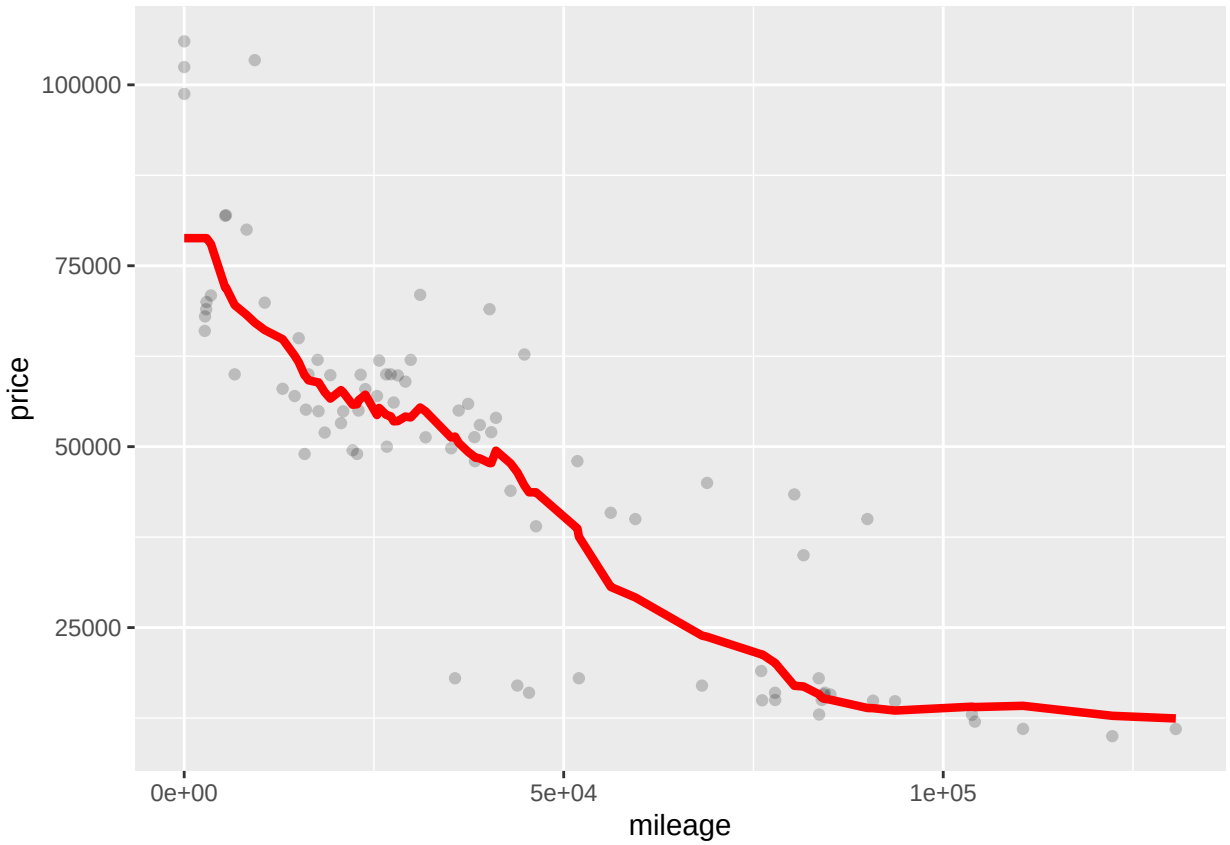

250000
200000
150000
100000
50000
price
0
50000
100000
150000
mileage

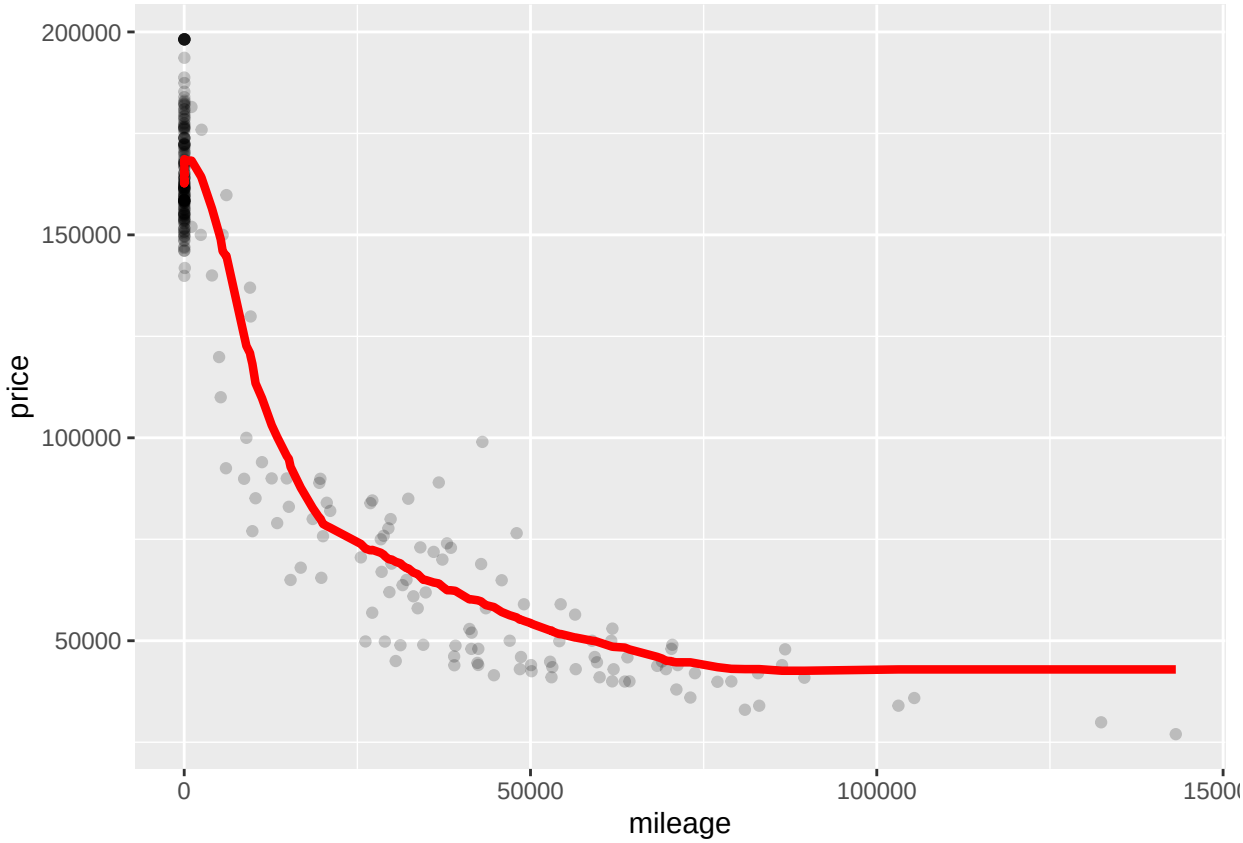
```
## [1] 11842.6
```

```
## [1] 13572.42
```




200000
150000
100000
50000
price
0
50000
100000
mileage


price
100000
75000
50000
25000
0e+00
5e+04
1e+05
mileage



I believe that the trim 63 AMG has a higher optimal value of k because theres a larger concentration of cars valued much higher with mileage close to zero,whereas the 350 trim is much more spread out.