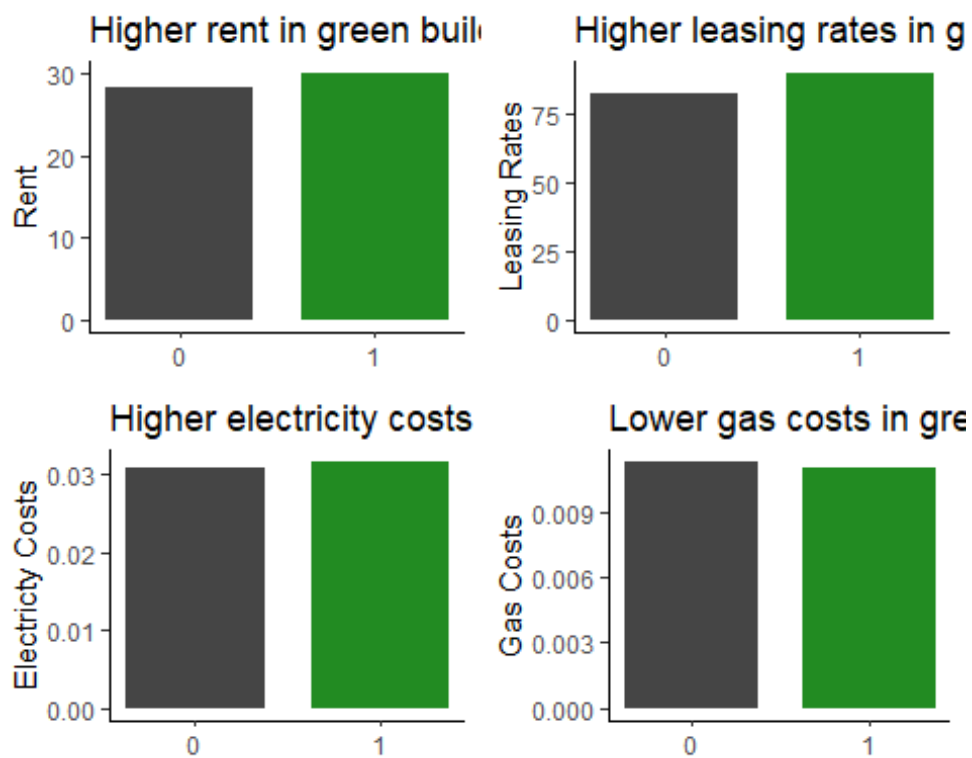# Excercise 1

Nam Tran

Part 1

In this part, I aimed to find a relationship that could dispel the notion that green ratings equate to higher rents, and thus higher revenues.I first started by grouping the buildings by whether or not they attained a green rating. Here we see some quick differences in rent, leasing rate etc.
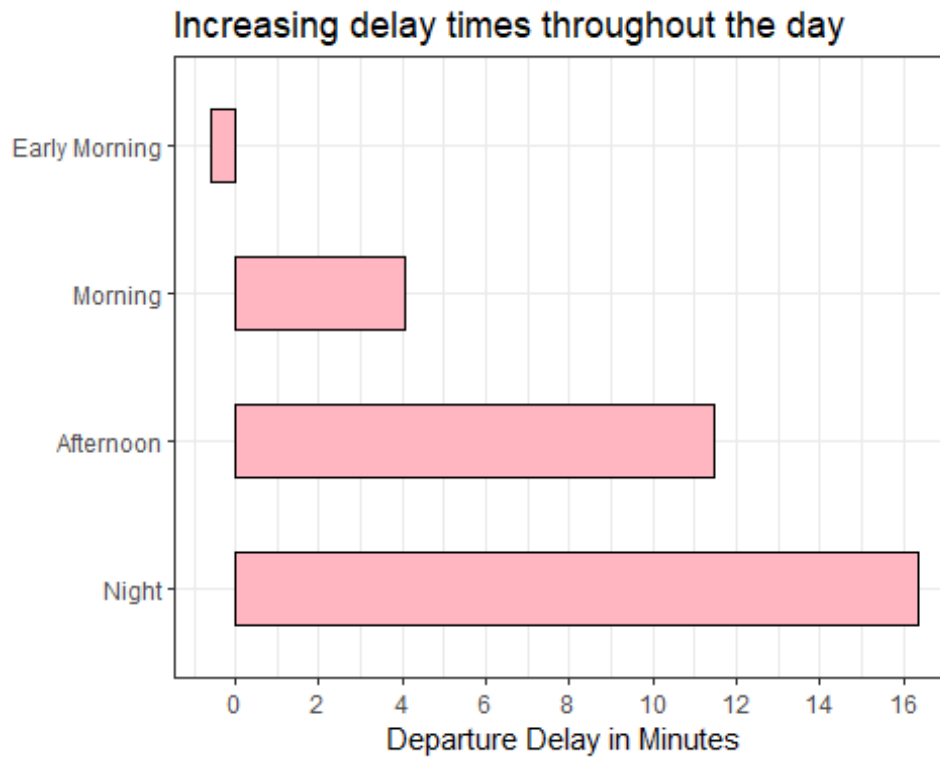

The pros of green buildings are evident here, but are the higher rents due to green ratings or other variables? Next, I examined the relationship between rent and age.

Decreasing rent as age i

Higher distribution of green ratings in lower age range
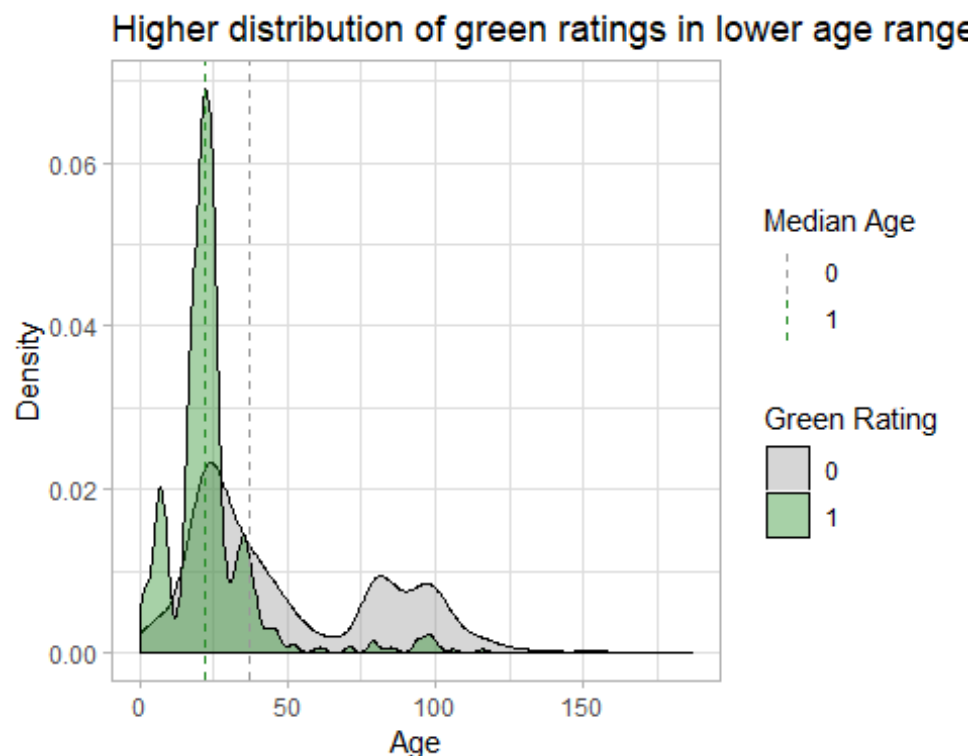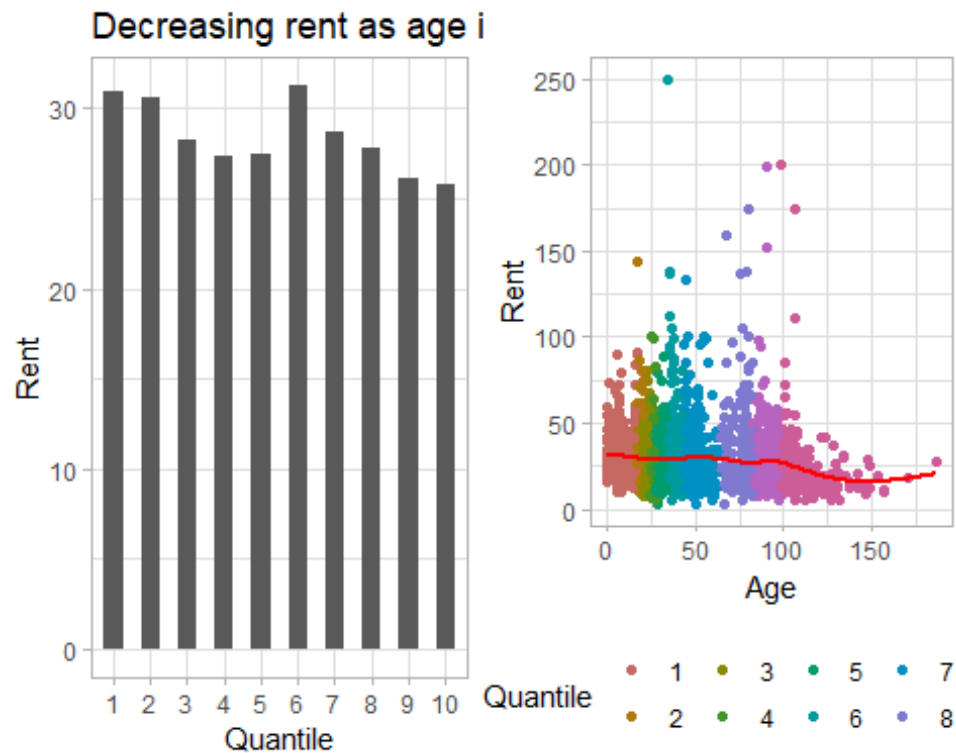
Here, it is clear that rent is lower in the newer buildings. I did this by dividing the age into ten quantiles and finding the average rent in each quantile. The age density graph serves to reinforce the notion that green buildings are much newer than non-green buildings. This certainly doesn't prove much, but it does make the "data guru'"s assertions a bit odd.

Part 2

In this section, I mainly looked to answer two questions: What is the best time of day to fly to minimize delays and What is the best time of year to fly to minimize delays? Here are the results:

Increasing delay times throughout the day

Early Morning

Morning

Afternoon

Night

Departure Delay in Minutes

For the times of day, I divided the 24 hours into 4, 6 hour periods. Early morning flights ran from 12 a.m. to 6 a.m., morning flights from 6.a.m. to 12 p.m. and onwards. Unsurprisingly, the least amount of delays occured in the "early morning" time frame, and in fact, the average flight was ahead of time! When looking at seasons, the results are also rather unsurprisng.

Less delay times during Autumn
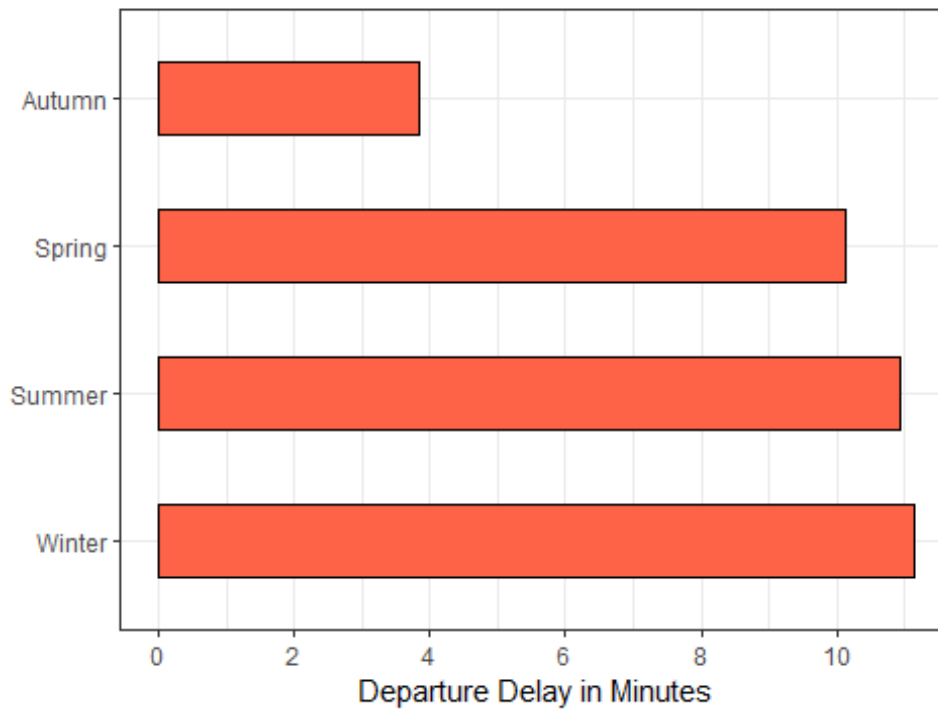
## Seasonal departure delay distribution
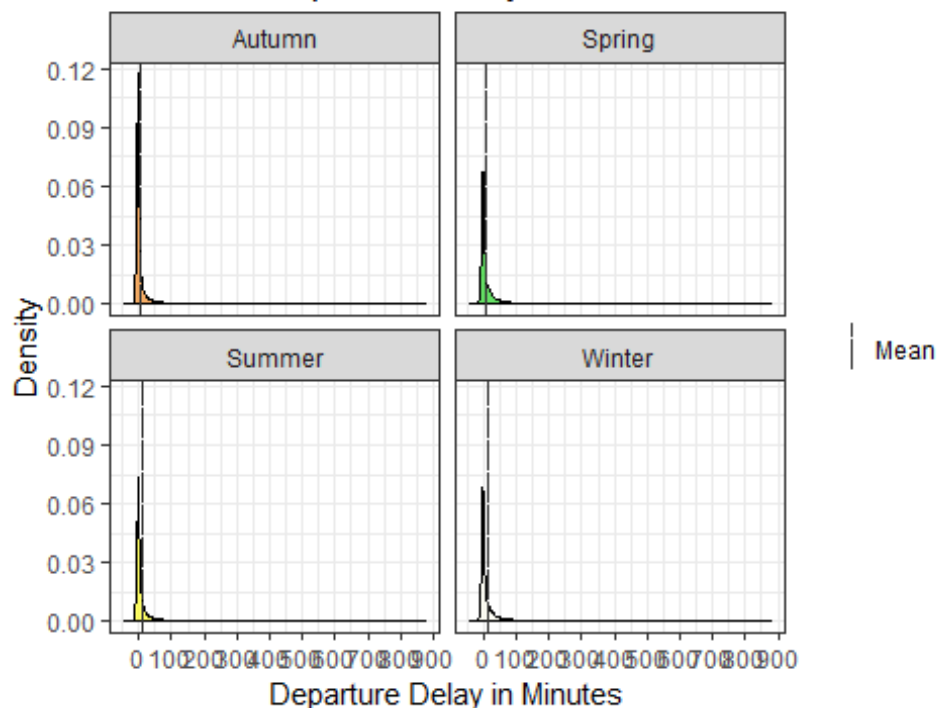
For the seasons, I divided the months by the traditional definitions. Winter runs from December to february, Spring going from March to May, Summer spanning June through August, and Autumn rounding it out from September to November. These results follow traditional intuition. Autumn doesn't span many major holidays, and it is also one of the

more fair weather seasons in this area. I quickly also looked at delay times by destination. The only major surprise is the absurdly high delay time from DSM which is Des Moines International Airport. Perhaps it is just a victim of severe outliers due to extreme weather or other unlikely delays.


Worst delay by destination
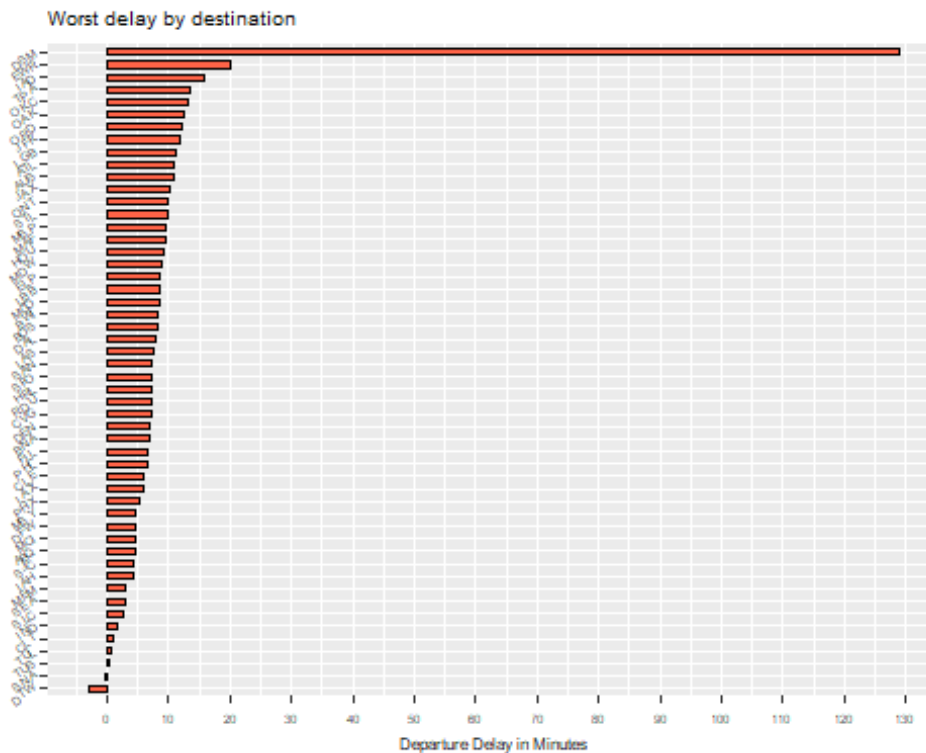
Part 3

In this part, I begin by dividng the data by two specific trims, 350 and 65 AMG. Then after creating the training and testing splits, I use K-nearest neighbors to build a model. I run each trim from K = 3 to K = 100, finding the out of sample rmse for each. I then used the minimum rsme of the set to build the final predictive graph.

Rmse for 65 AMG trim
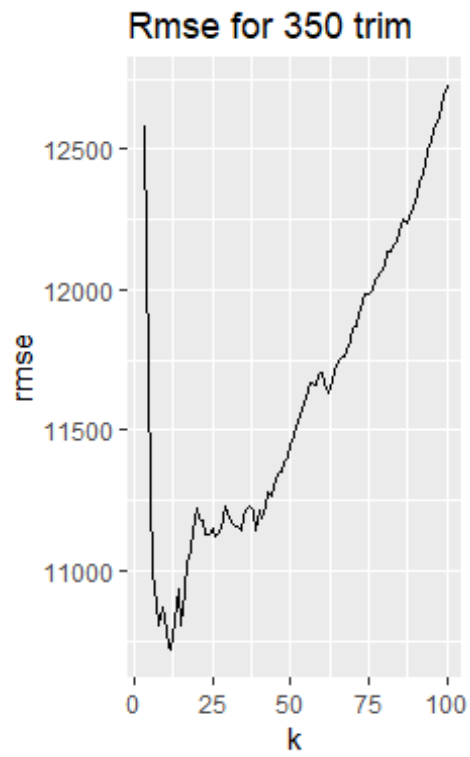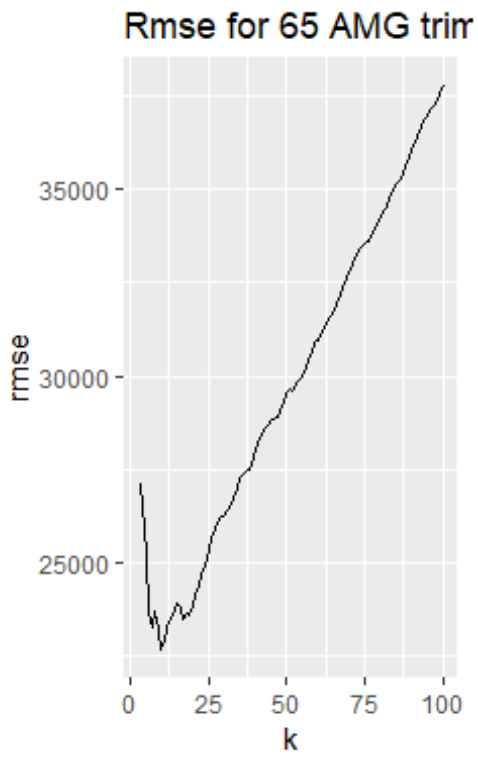
Rmse for 350 trim


```
## Optimal K for 65 AMG trim:  10

## Optimal K for 350 trim:  12
```

Prediction for 65 AMG

Prediction for 350
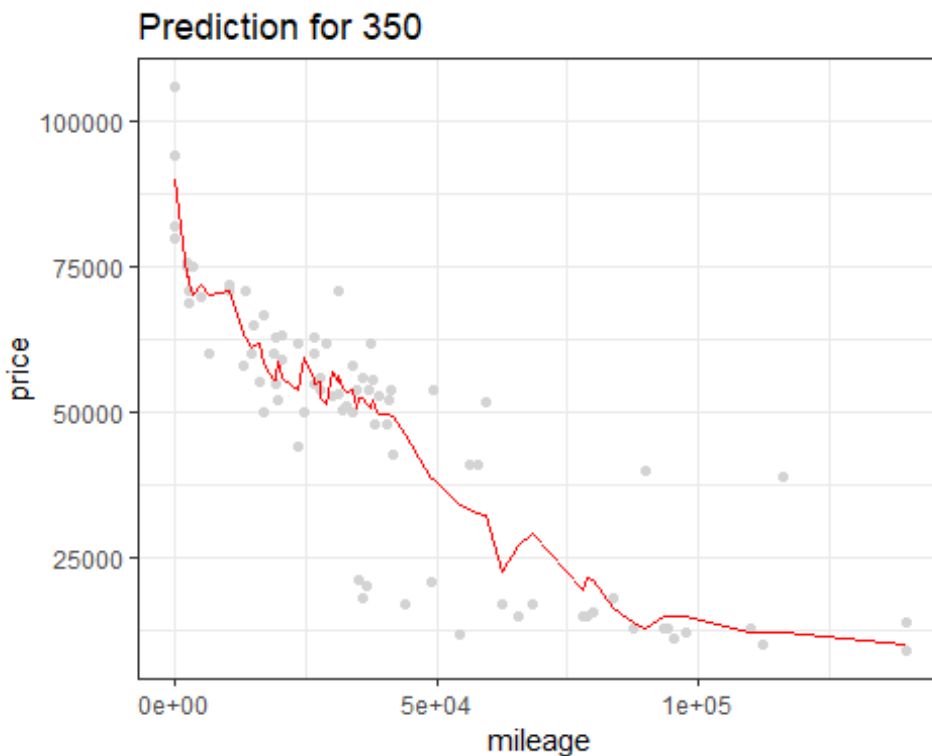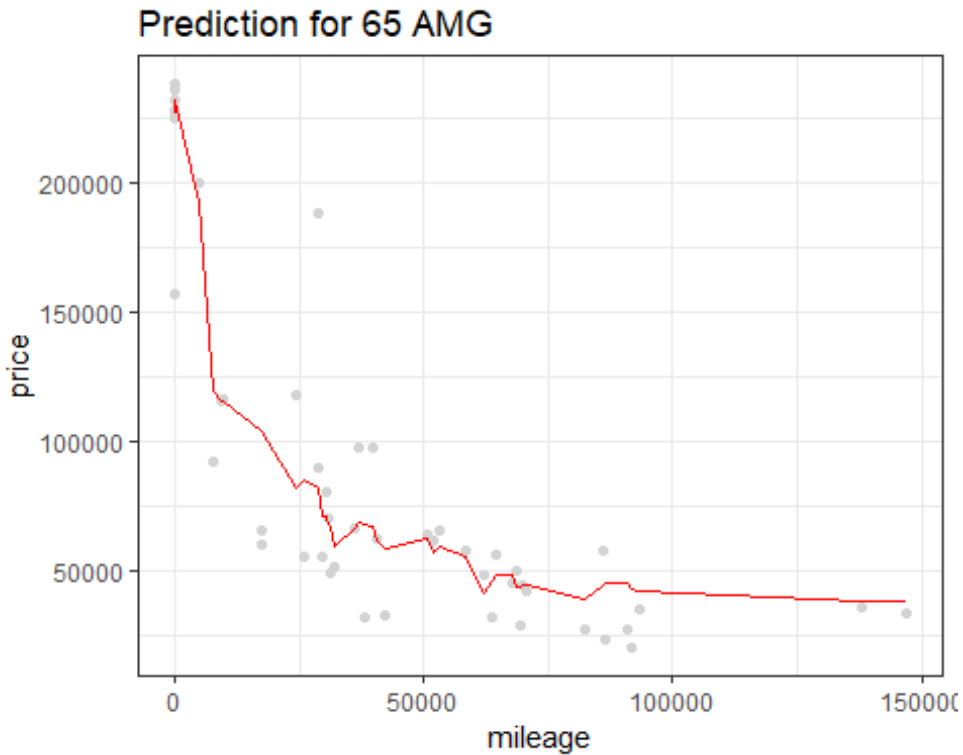
Overall, the 350 trim has a higher optimal K value. I think this is due to the larger sample size available. It enables the model to "explain" more of the variance. However, with this training and test split, the results do vary from run to run.