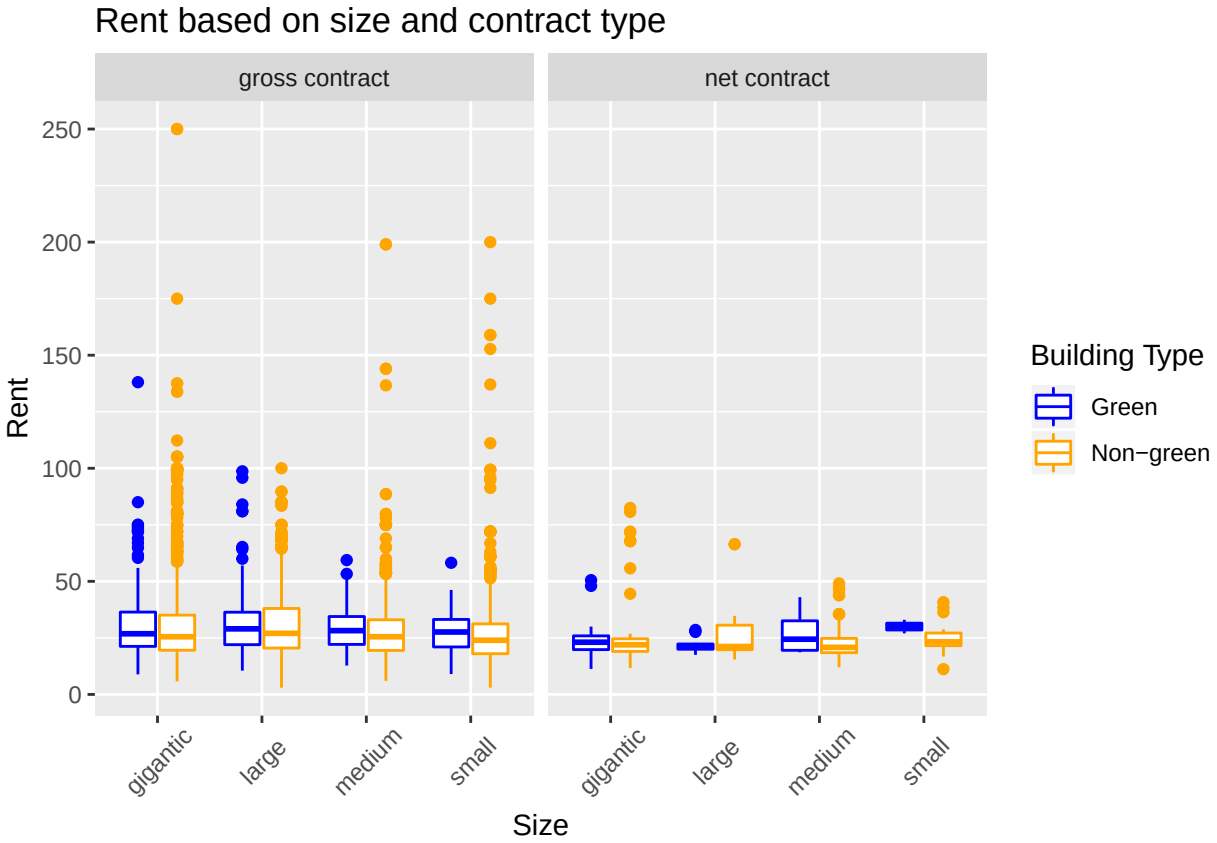
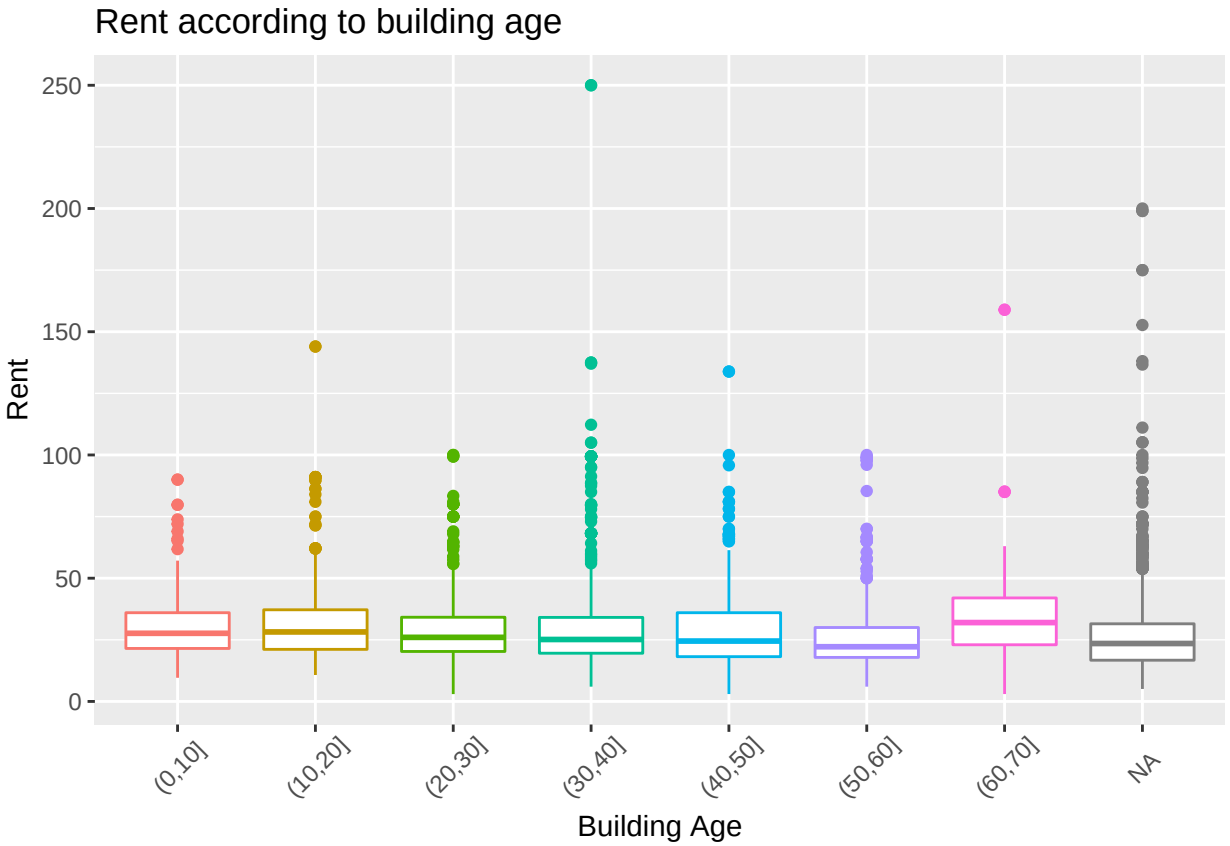
Exercise 1

Question 1

In order to compare results with the data guru, it is essential to reorganize the data in a sensible fashion. To accomplish this, we first reorganize the size of the propoerties into 4 sizes: small, medium, large, and gigantic. Next, we sort the contract types between gross and net. Finally, we cut the data into 10 year intervals. These factors were all chosen due to their suspected effects on rent: larger properties may cost more, individuals with net contracts (pay their own utilities) likely pay more, and the building age may alter rental fees. After these steps, we are ready to proceed with analysis.



In this figure, we observe that gross contracts tend to have higher rents. Generally speaking, this makes the data guru's estimation of the difference between green and non-green rents dubious, as they have not factored in this informaiton.

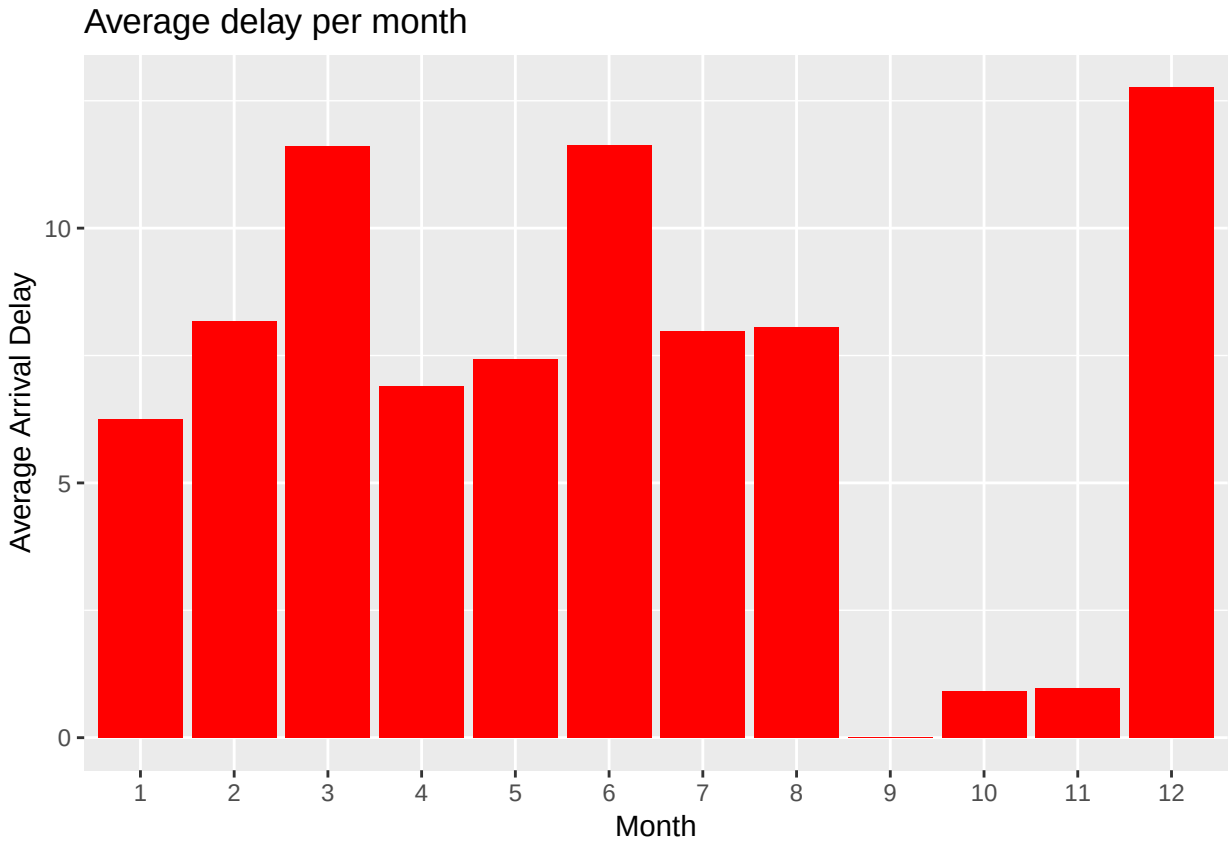


In the second figure, it is clearly established that rent is not constant throughout a buildings lifespan, casting doubt on the guru's estimation for profit margins. In fact, it appears that the median rental cost decreases for the majority of years, with an abberation in 60-70 year old buildings.

In conclusion, it is hard to fully trust the data guru's analysis. His assumptions of differing contract's yielding similar rents and constant rent over time are clearly misguided. In order to improve the analysis, steps must be committed that normalize median rents over the course of contract type, time, and many other factors.

Question 2

In order to figure out which time of year is the best for traveling, I organized the data by month and create a bar chart.



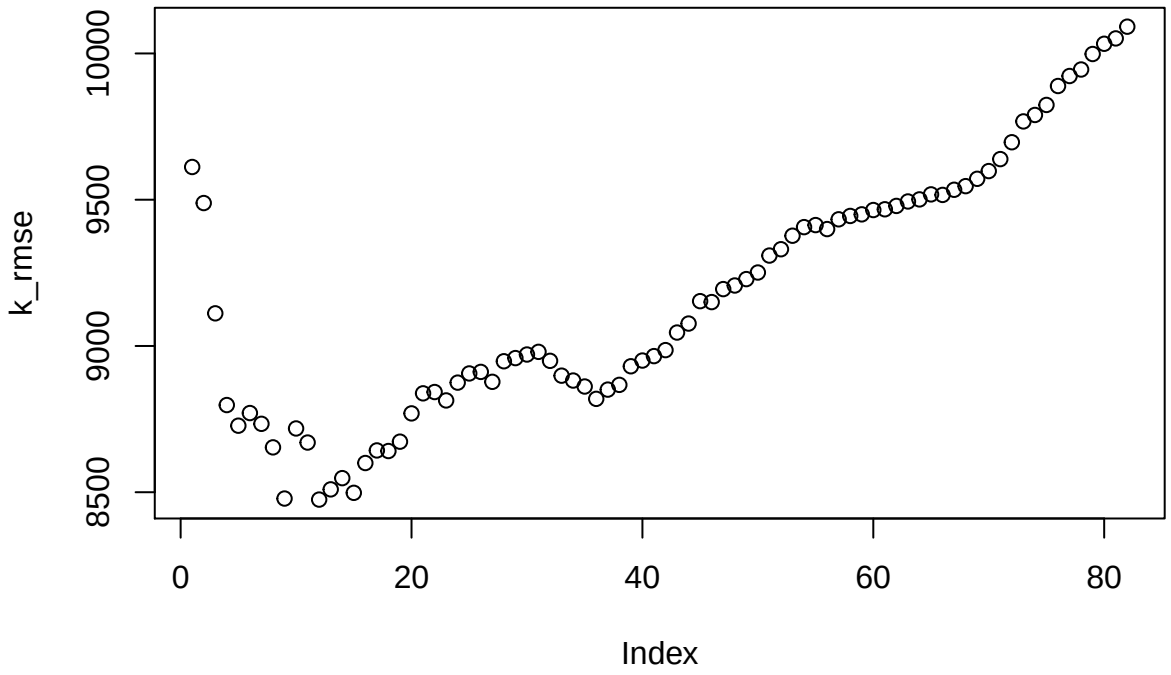
In conclusion, it seems fall is the best time because of less carrier and weather delays. This makes sense, as there are not many vacation and fall and weather is moderate.

Question 3

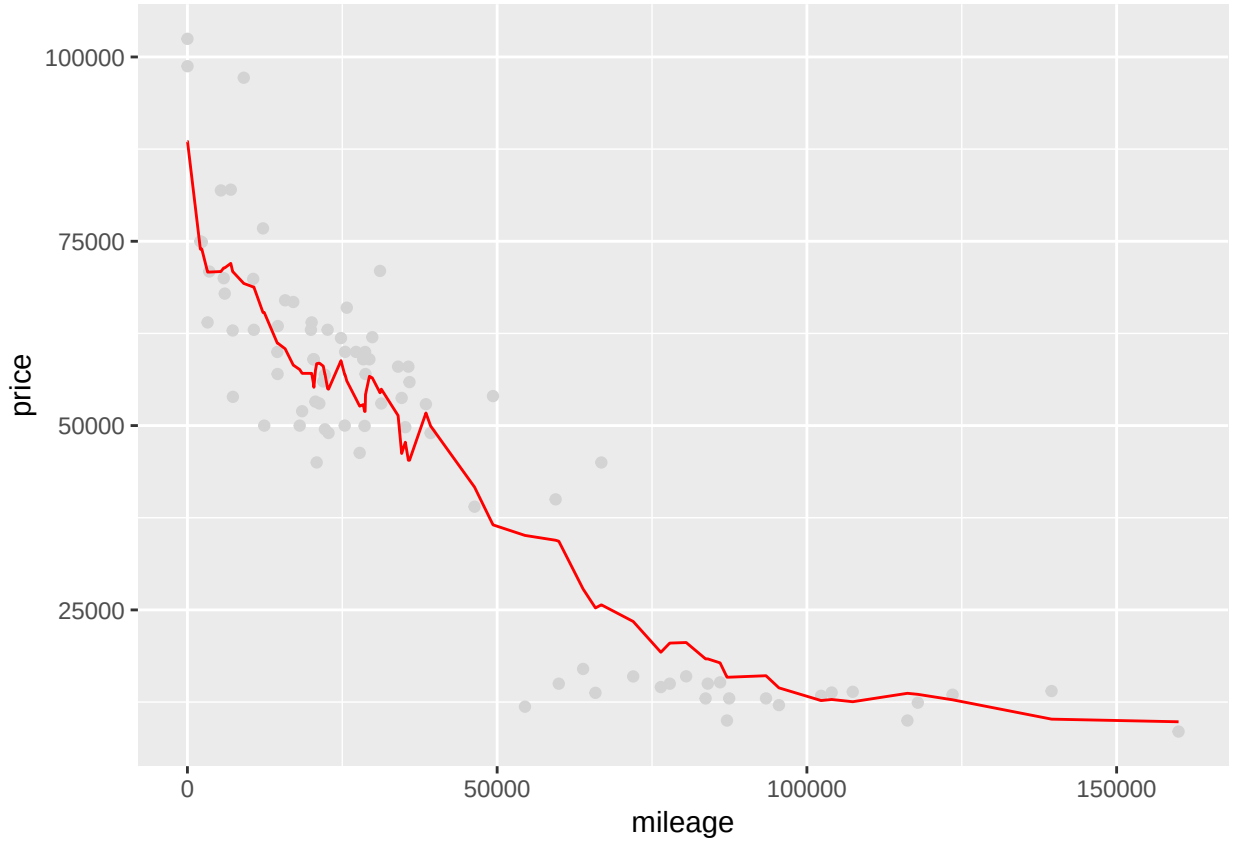
First, we will focus on 350 AMG.

Here, we have generated the optimal K value.

```
## [1] 12
```



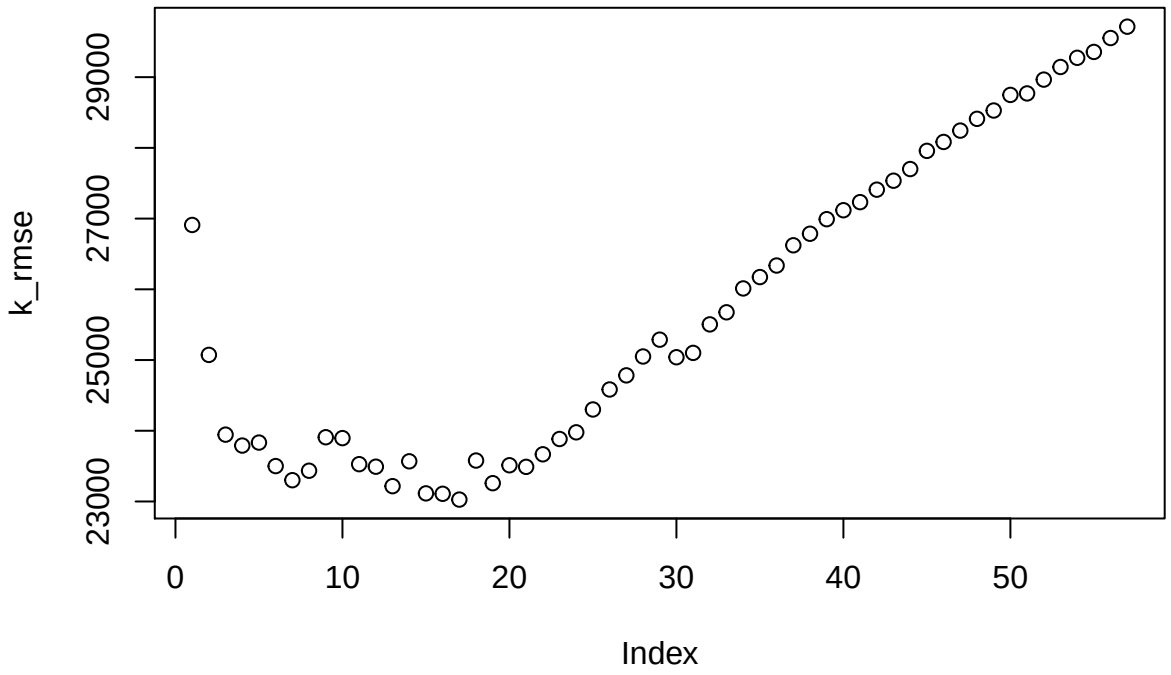
This graph displays the RMSE associated with each K-Value.



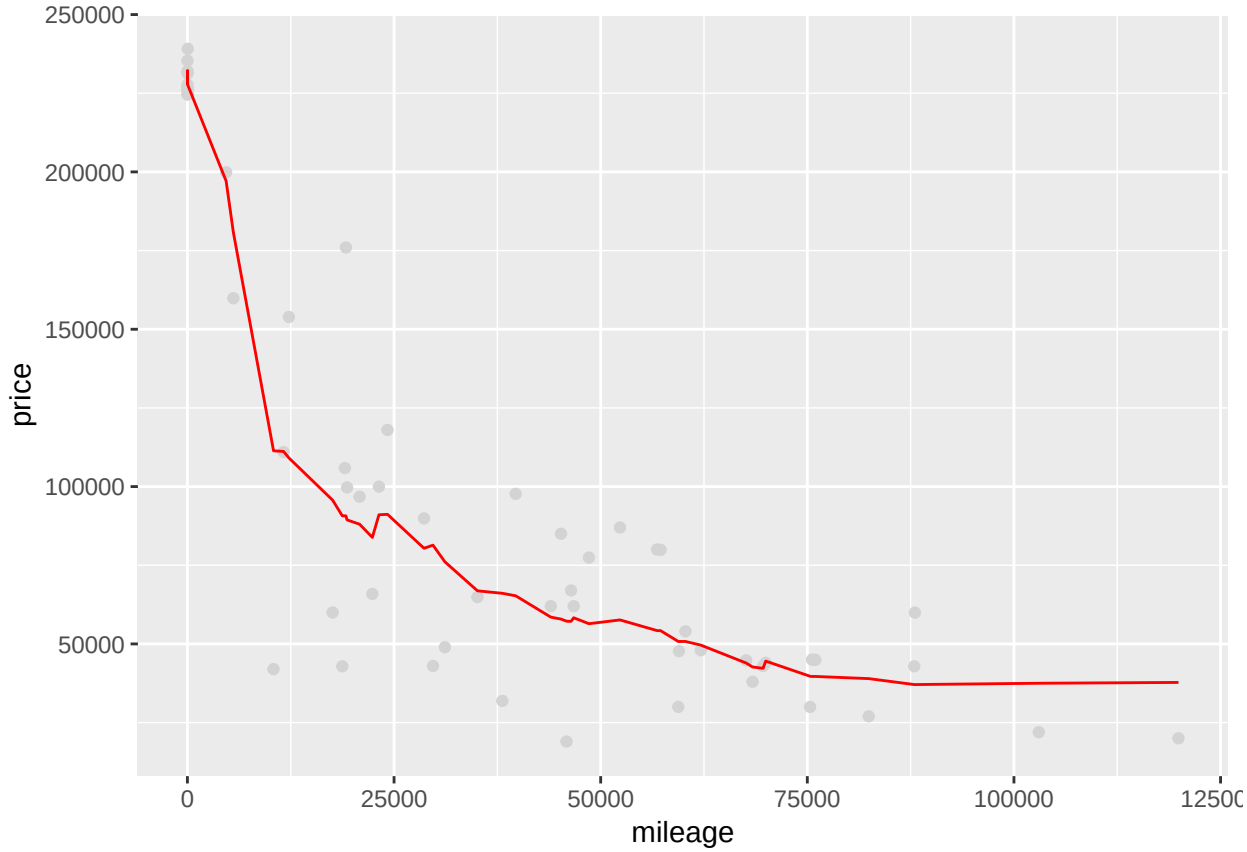
This graph displays the pattern between price and mileage at the optimal K.

We will now do the same thing for 65 AMG.

```
## [1] 17
```



Again, we have generated an optimal K Value as well as the pattern between RMSE and K Values.



Pattern between price and mileage for 65 AMG.

In this instance, the optimal K value for 65 AMG was greater than 350 AMG. Since there is more observations for 65 AMG, this does make sense. We can use a bigger K and not be hurt by variance in bigger data sets.

In general, the K value is subject to change in any given trial, so this will not always be the case.