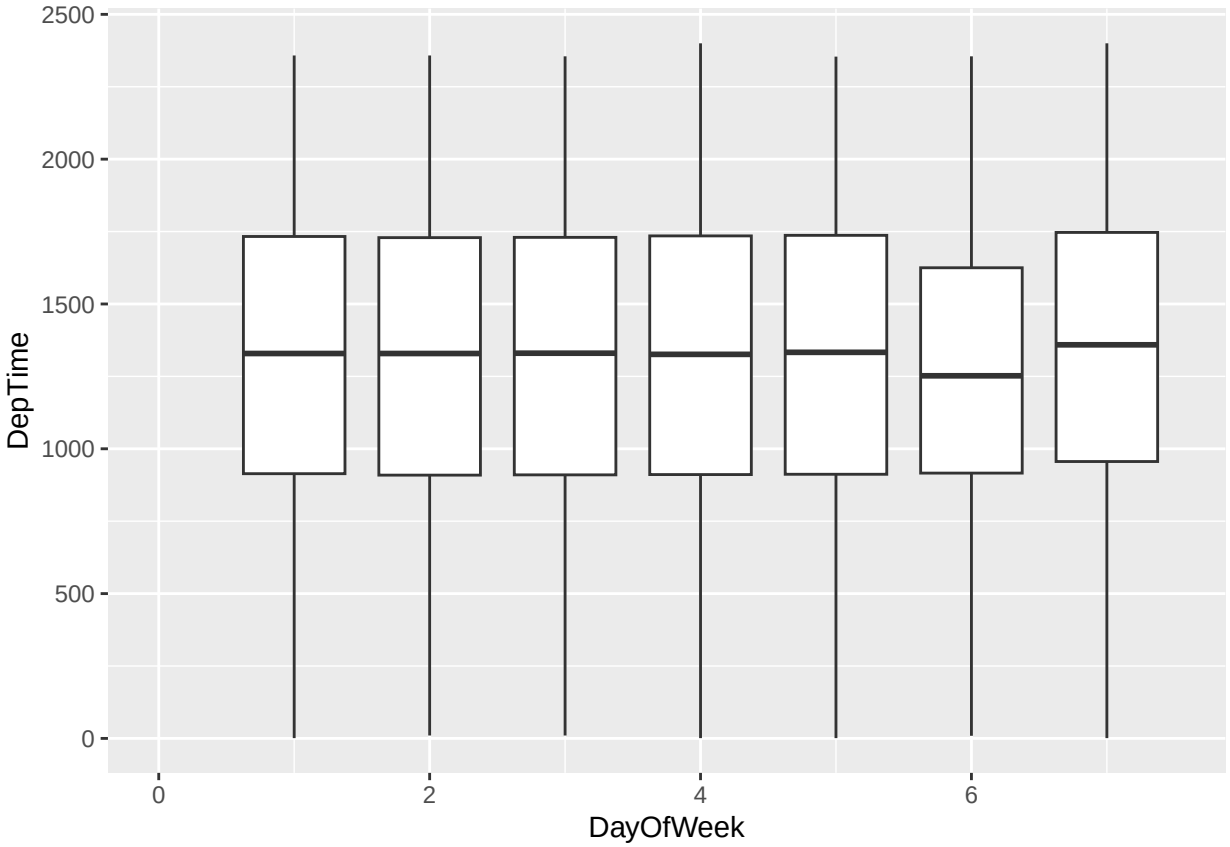
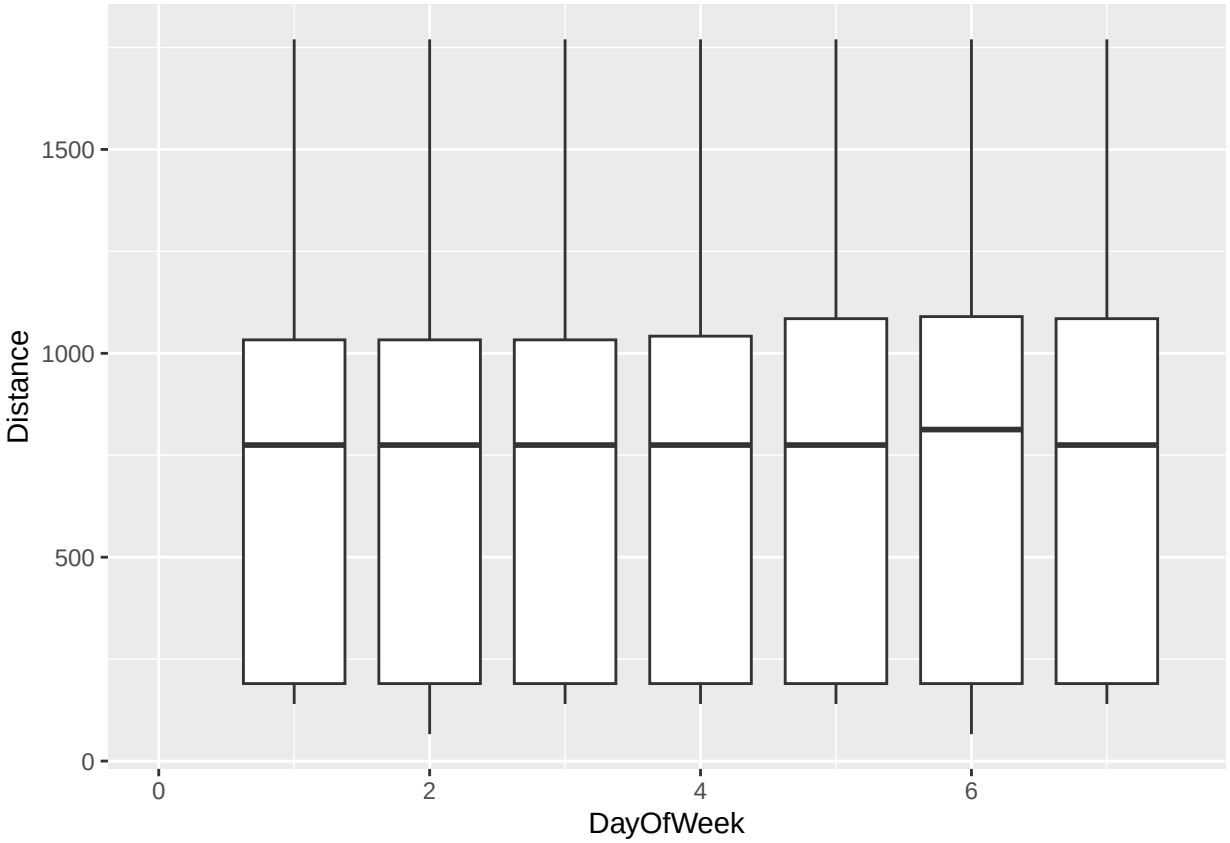
# Problem Set 1

By Jae You (jhy479)

1) One interesting observation from the 2008 flight data for Austin-Bergstrom International Airport is the departure time of flights over the course of the week.




The boxplots show the Departure Time of the flights across the week, where 1~7 represent the days Monday through Sunday. From the boxplots above, we see that the mean departure time is roughly the same for all the flights except on Saturdays, where the flight departs early. With such a large sample size of almost 100,000 flights, we would expect an even spread across the week, but Saturday's departure time is a clear outlier. However, I came up with a potential explanation for this phenomenon from the graph below.




By graphing the distance that flights have to travel across the week, we can see that flights on Saturday have a higher mean distance. So flights would be flying a longer distance than normal on average, which would explain why the flights must depart earlier in order to travel a longer distance and arrive on time.

2)

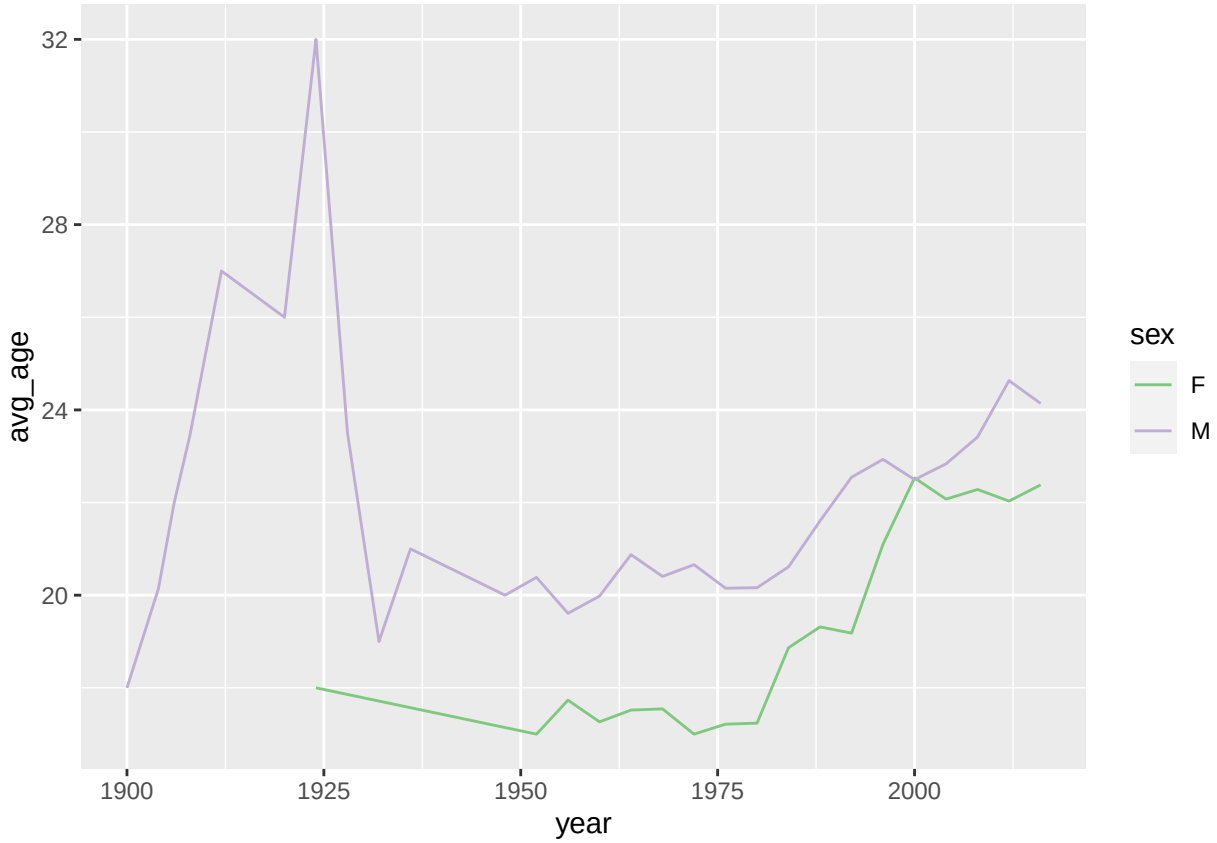
a)

The 95th percentile of heights for female competitors is 186cm.

b)

The event Rowing Women's Coxed Fours had the greatest standard deviation in female competitor's heights.

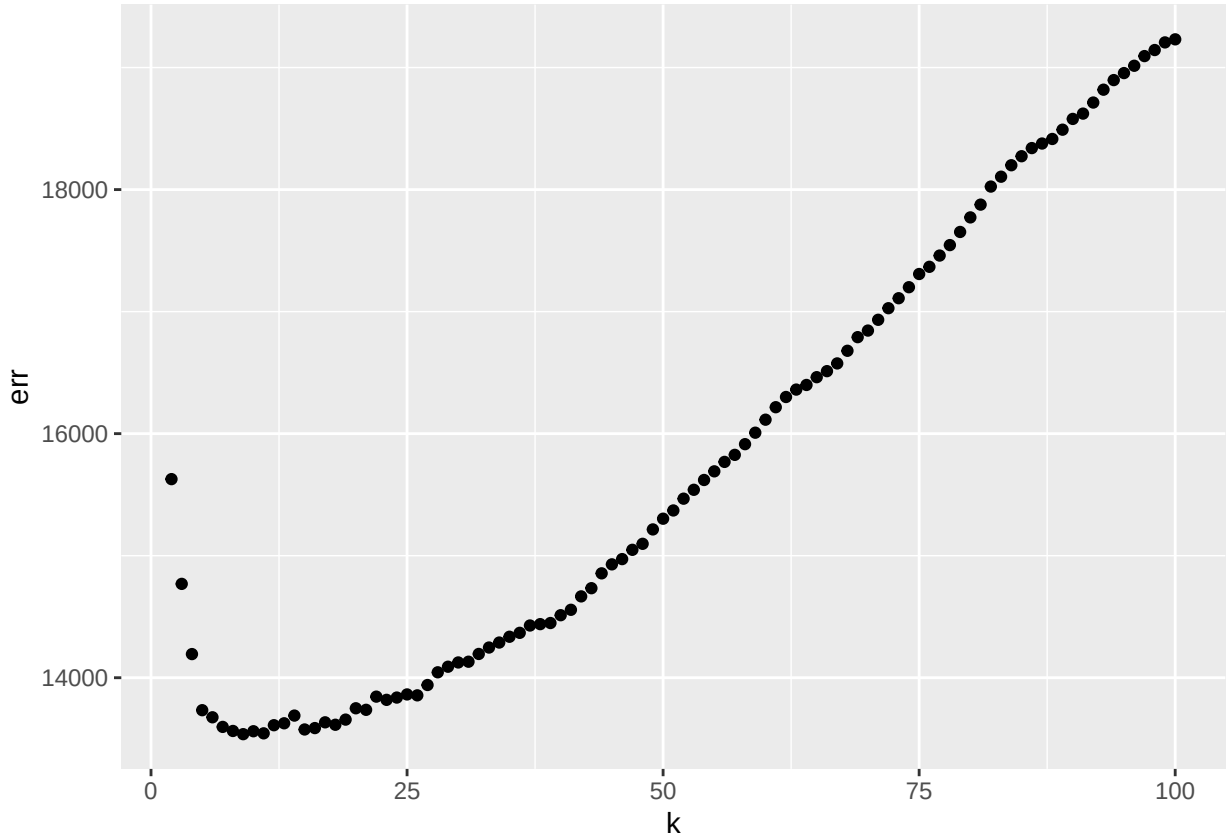
c)



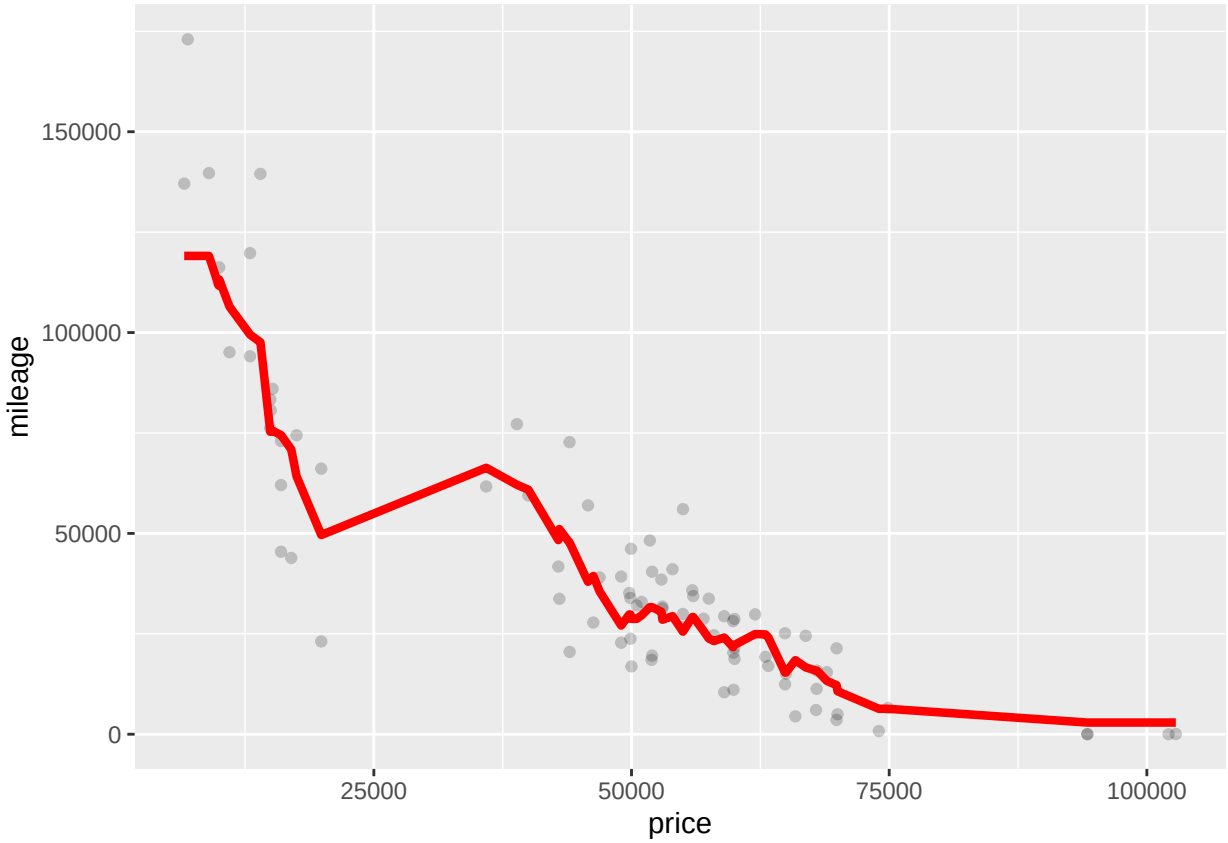

From the 1930s to now, the average age of Olympic Swimmers has increased over time for both males and females. However, there were only male Olympic Swimmers in 1900 to 1925, where the trend was increasing over time until a sharp drop in the average age around 1925.

3)

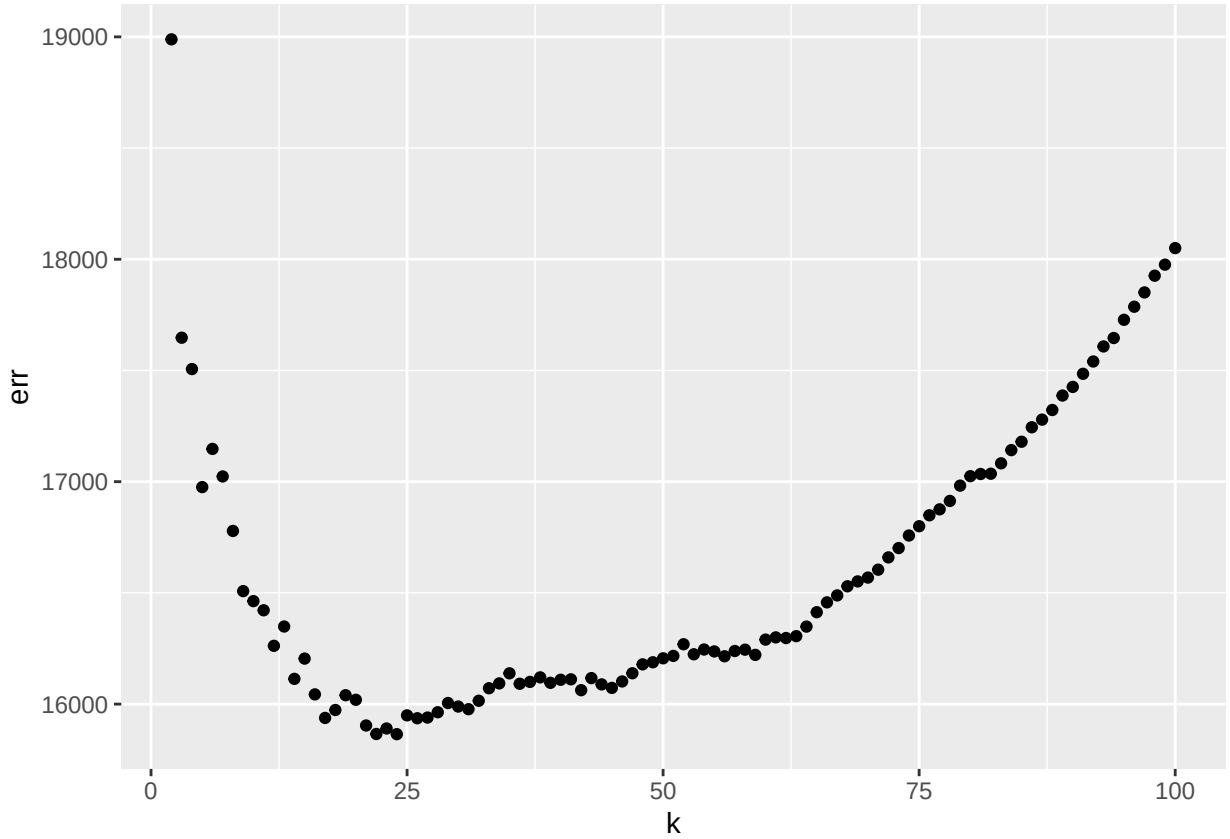
After filtering the two trim levels, I ran KNN for the 350 trim level below:



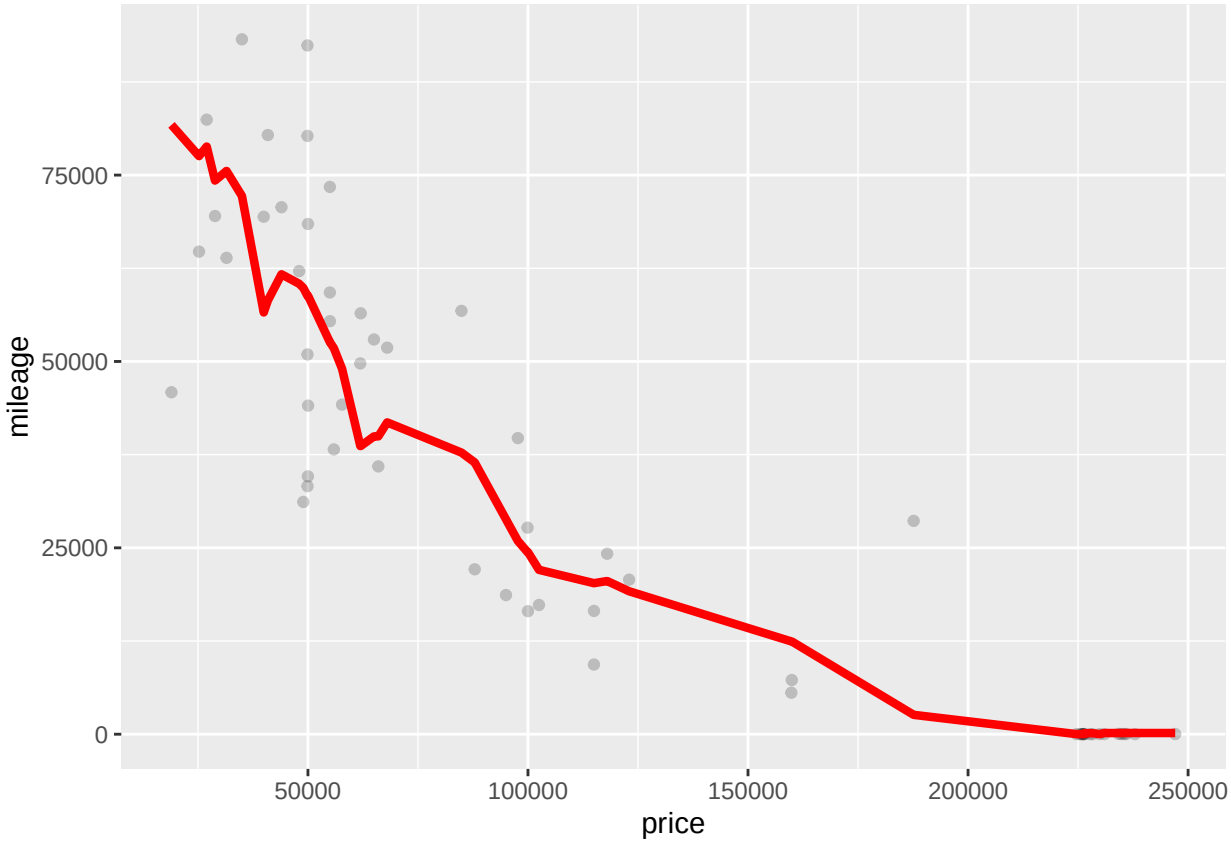

We can see that the lowest rmse is at K=13 for the 350 trim level. The fitted model is below:




Next, I ran the 65 AMG trim level:




We can see that the lowest rmse is at K=18 for the 65AMG trim level. The fitted model is below:


mileage
75000
50000
25000
0
50000
100000
150000
200000
250000
price