E1

Siyuan Liu

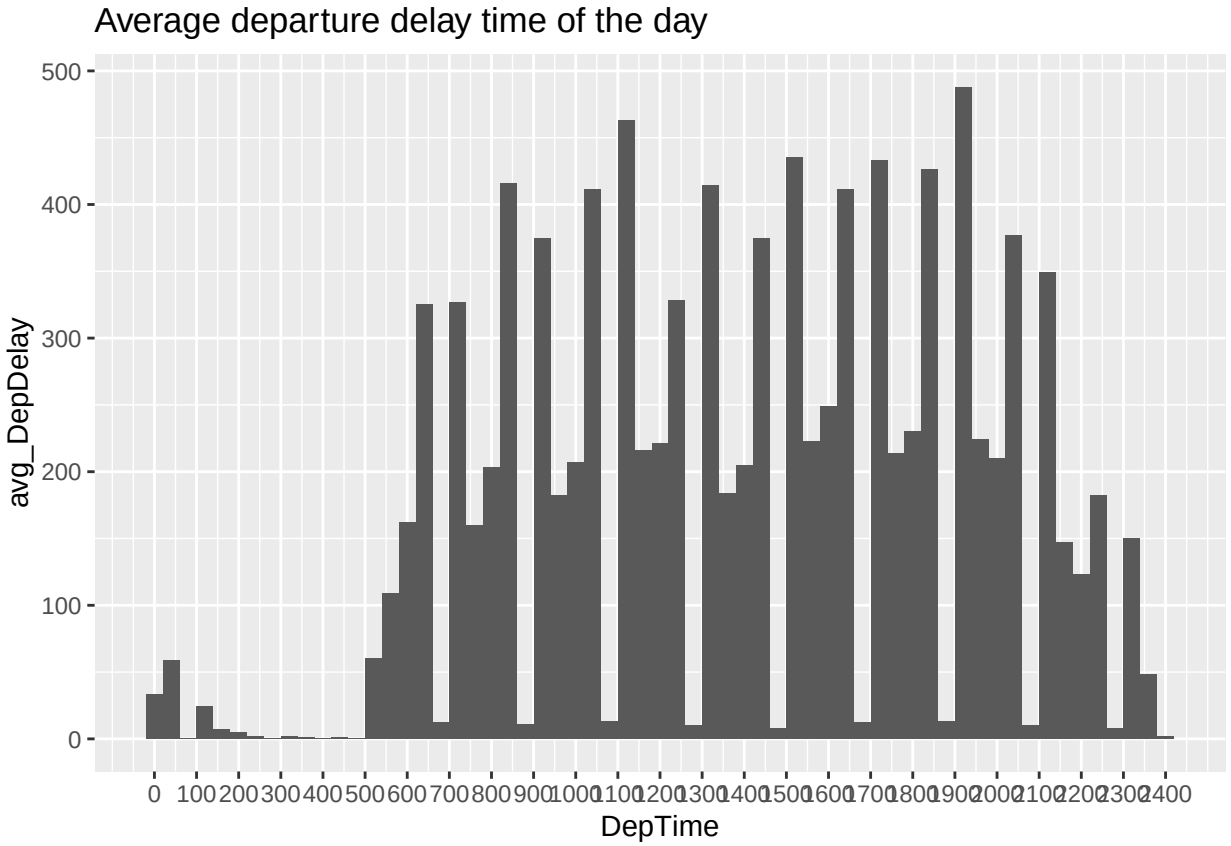
2023-01-27

##problem1

I try to figure out what is the best of day to fly to minimize delays, and does this change by airline. As it shows in plot 1, 4am-5am is the best time of day to fly to minimize delays, and for different airline, the beat time of day looks different. For example the best time for B6 to minimize delays is about 12am-1am, but for XE the best time is about 5am-6am.

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

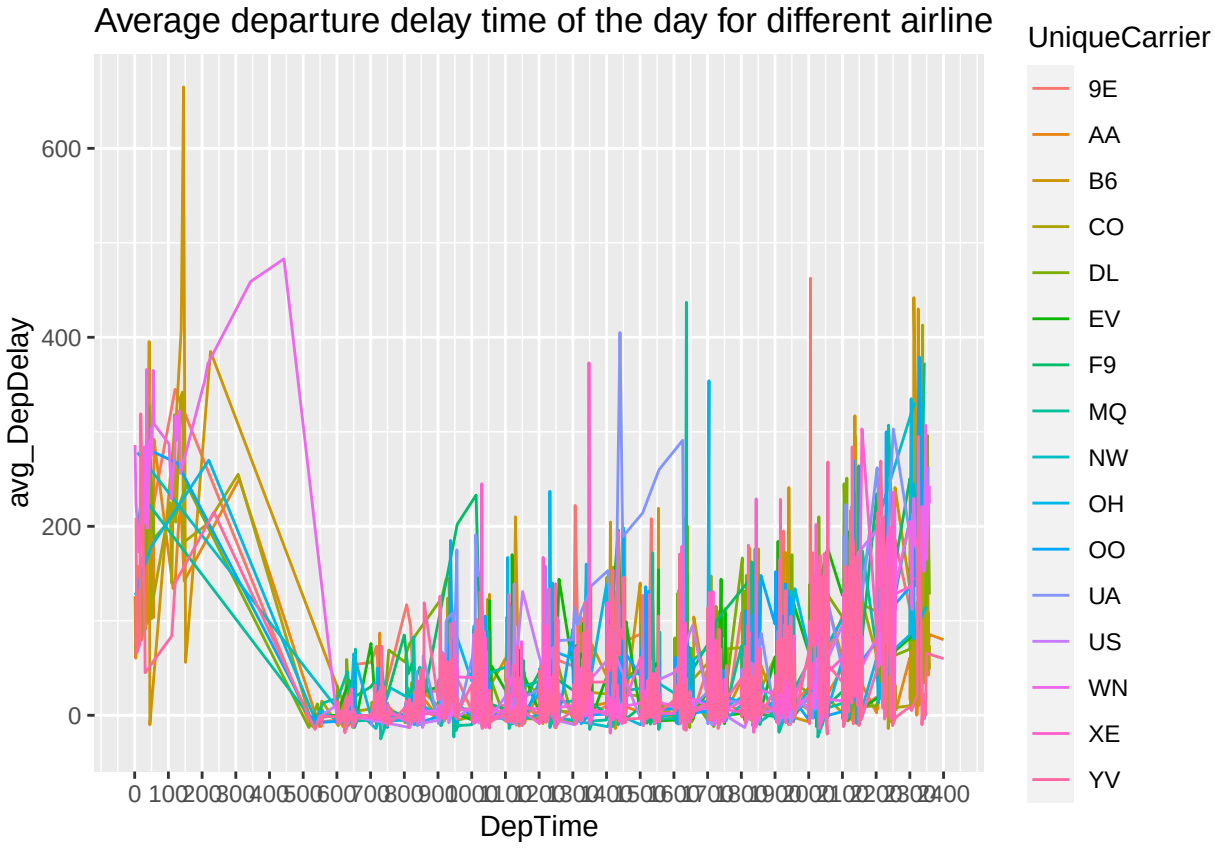
```
## Warning: Removed 15 rows containing non-finite values (`stat_bin()`).
```



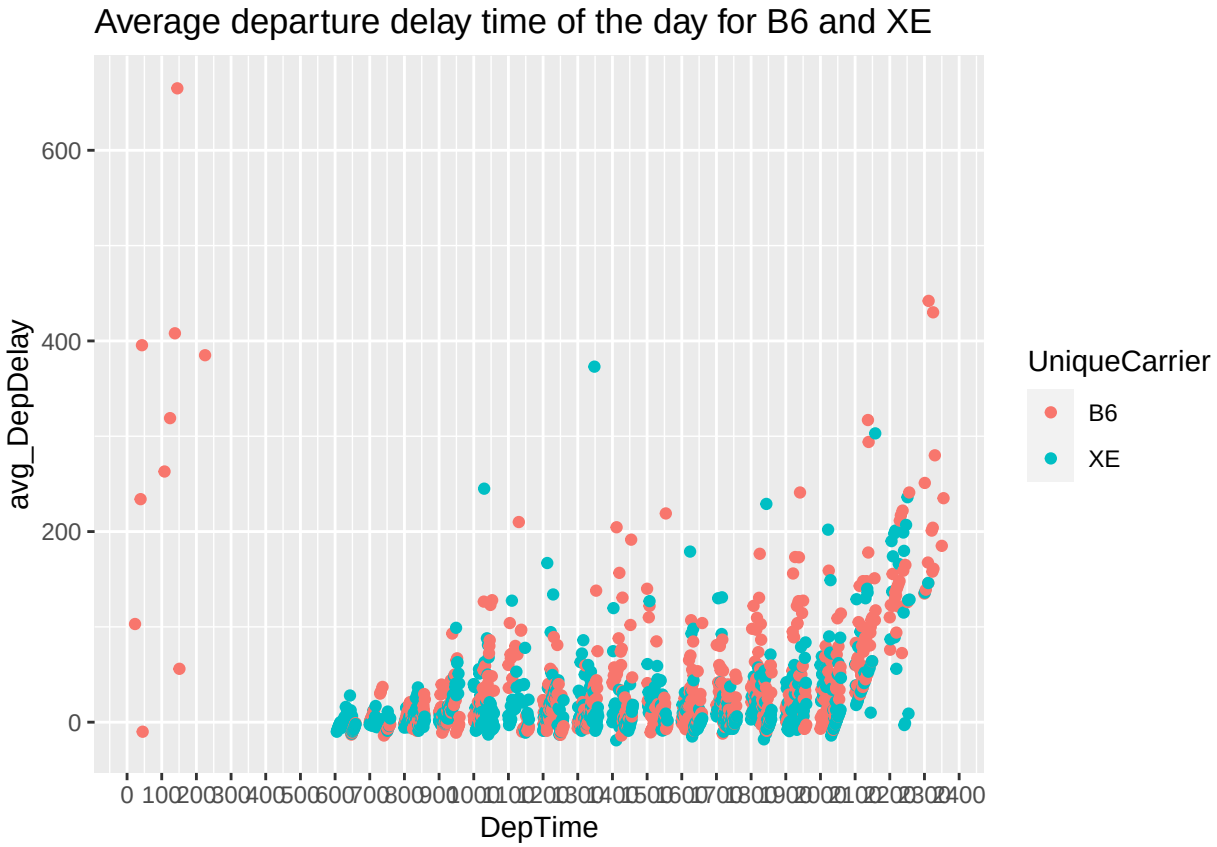

```
## # A tibble: 10 x 3
```

```
## # Groups:   DepTime [10]
##    DepTime UniqueCarrier avg_DepDelay
##      <int> <chr>                <dbl>
##  1     730 MQ                     -25
##  2     947 MQ                     -23
##  3    2027 MQ                     -23
##  4    2029 MQ                     -21
##  5    2055 YV                     -20
##  6    1411 XE                     -19
##  7    2031 MQ                     -19
##  8     624 YV                   -18.5
##  9    1837 XE                     -18
## 10    1912 MQ                     -18
```

```
## Warning: Removed 15 rows containing missing values (‘geom_line()‘).
```




```
## Warning: Removed 2 rows containing missing values (‘geom_point()‘).
```



##problem 2
##A

```
##   q95_temp
## 1      183
```

The 95th percentile of heights for female competitors across all Athletics events is 183.

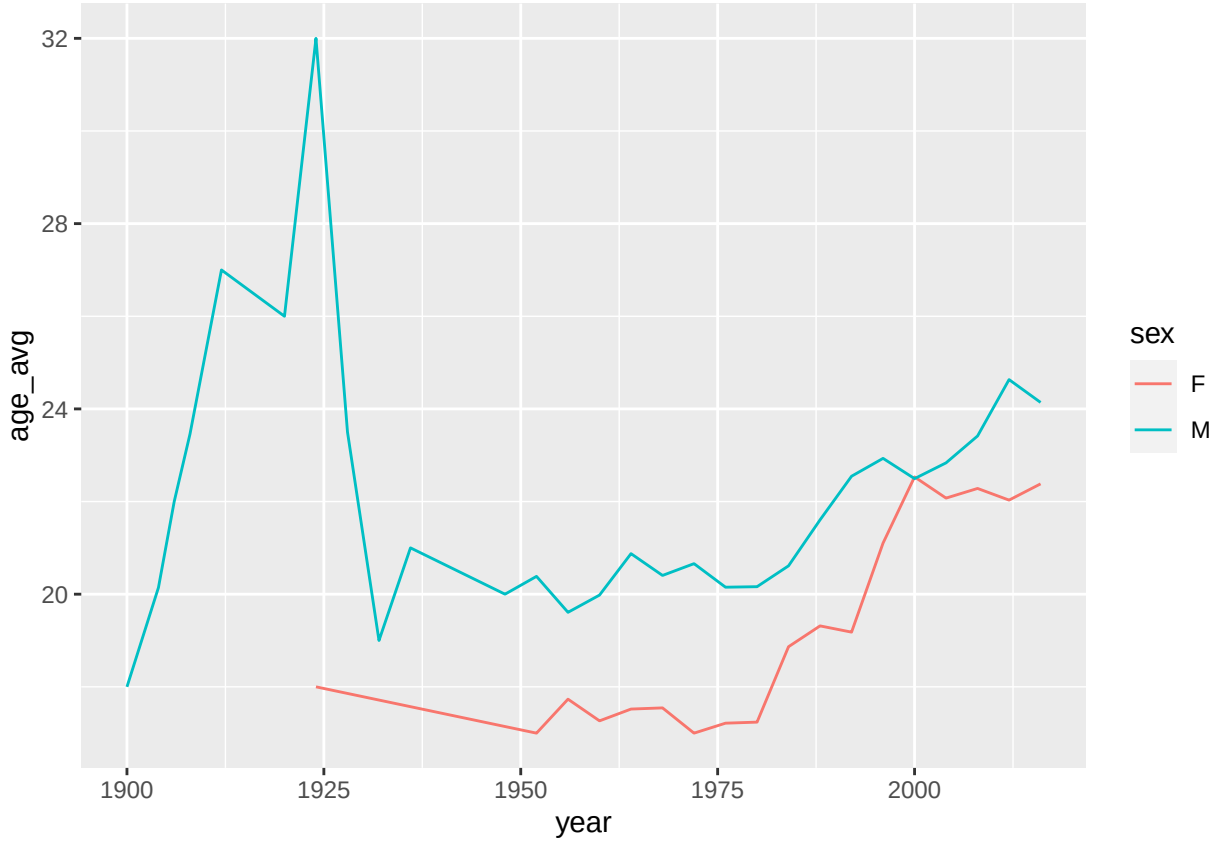
##B

```
## # A tibble: 132 x 2
##    event                                sd_height
##    <chr>                                    <dbl>
##  1 Rowing Women's Coxed Fours               10.9
##  2 Basketball Women's Basketball             9.70
##  3 Rowing Women's Coxed Quadruple Sculls     9.25
##  4 Rowing Women's Coxed Eights               8.74
##  5 Swimming Women's 100 metres Butterfly     8.13
##  6 Volleyball Women's Volleyball             8.10
##  7 Gymnastics Women's Uneven Bars            8.02
##  8 Shooting Women's Double Trap              7.83
##  9 Cycling Women's Keirin                    7.76
## 10 Swimming Women's 400 metres Freestyle     7.62
## # ... with 122 more rows
```

Swimming Women's 100 meters Butterfly had the greatest variability in competitor's heights across the entire history of the Olympics.

##C

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



The trend look different for male swimmers For male swimmers, the average age of Olympic swimmers increase from 1900 to 1925 and the average age of Olympic swimmers decrease until about 1975 for both male and female.

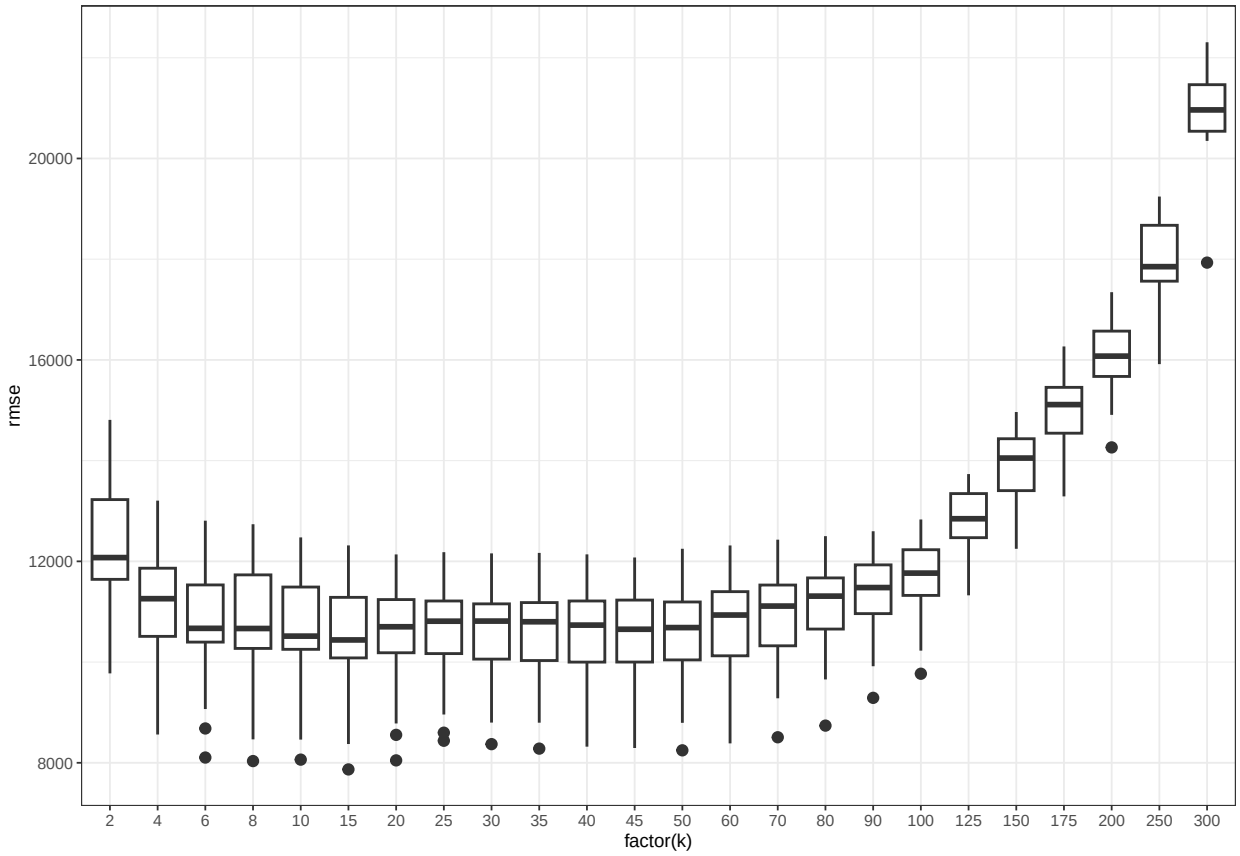
##problem 3

##350

```
## Warning: executing %dopar% sequentially: no parallel backend registered
```

```
## # A tibble: 24 x 2
##        k  armse
##    <dbl>  <dbl>
##  1     2 12348.
##  2     4 11199.
##  3     6 10708.
##  4     8 10705.
##  5    10 10602.
##  6    15 10437.
##  7    20 10532.
##  8    25 10610.
##  9    30 10590.
```

```
## 10    35 10593.
## # ... with 14 more rows
```

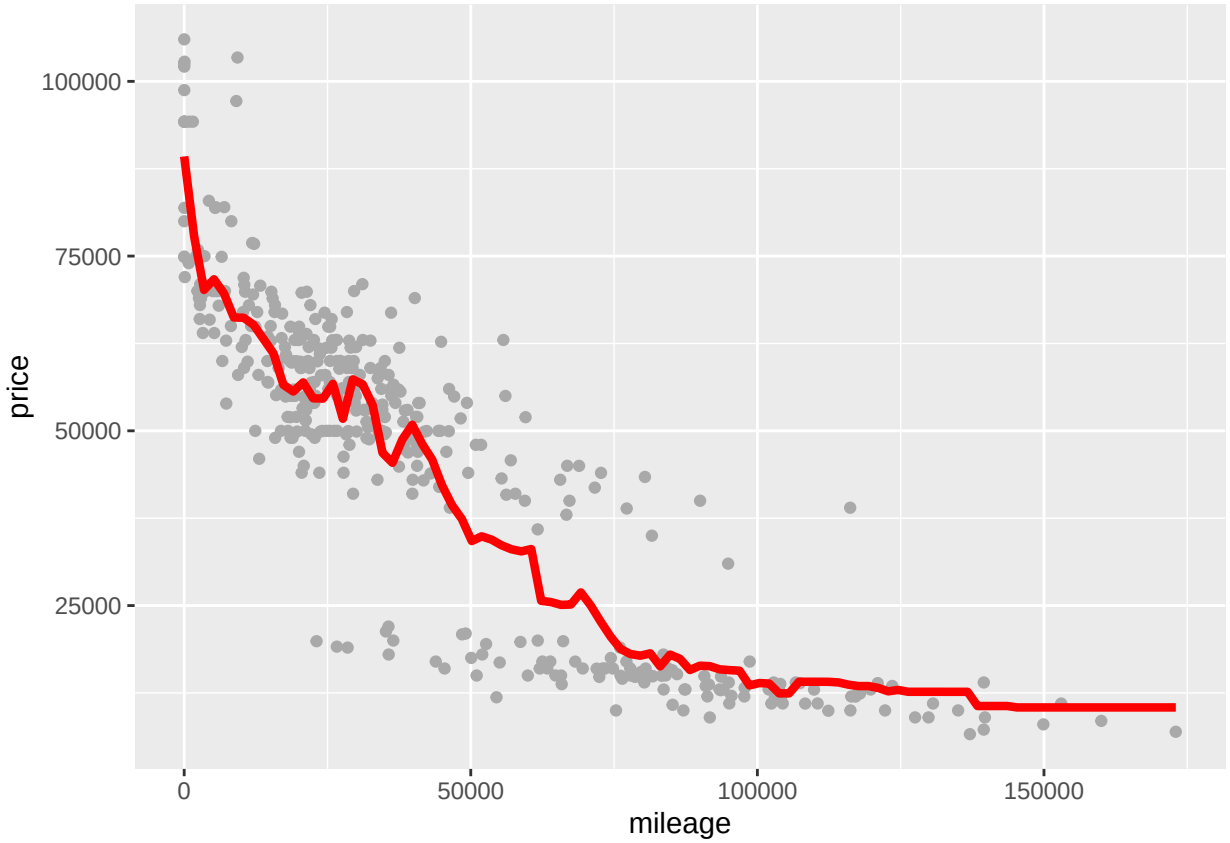


When K = 15 it bottoms out.

For k = 15 the plot of the fitted model shows as follow:

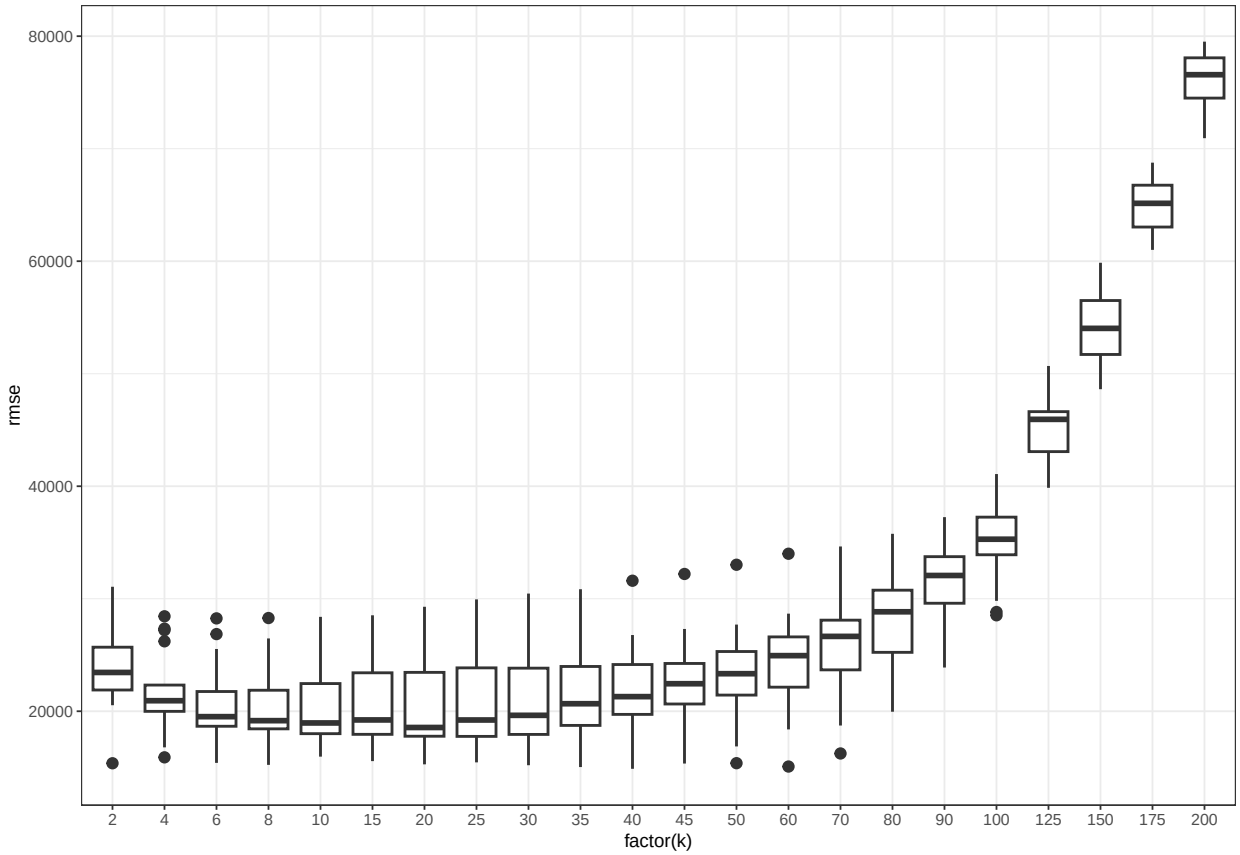
```
## [1] 8994.777
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
```



##65 AMG

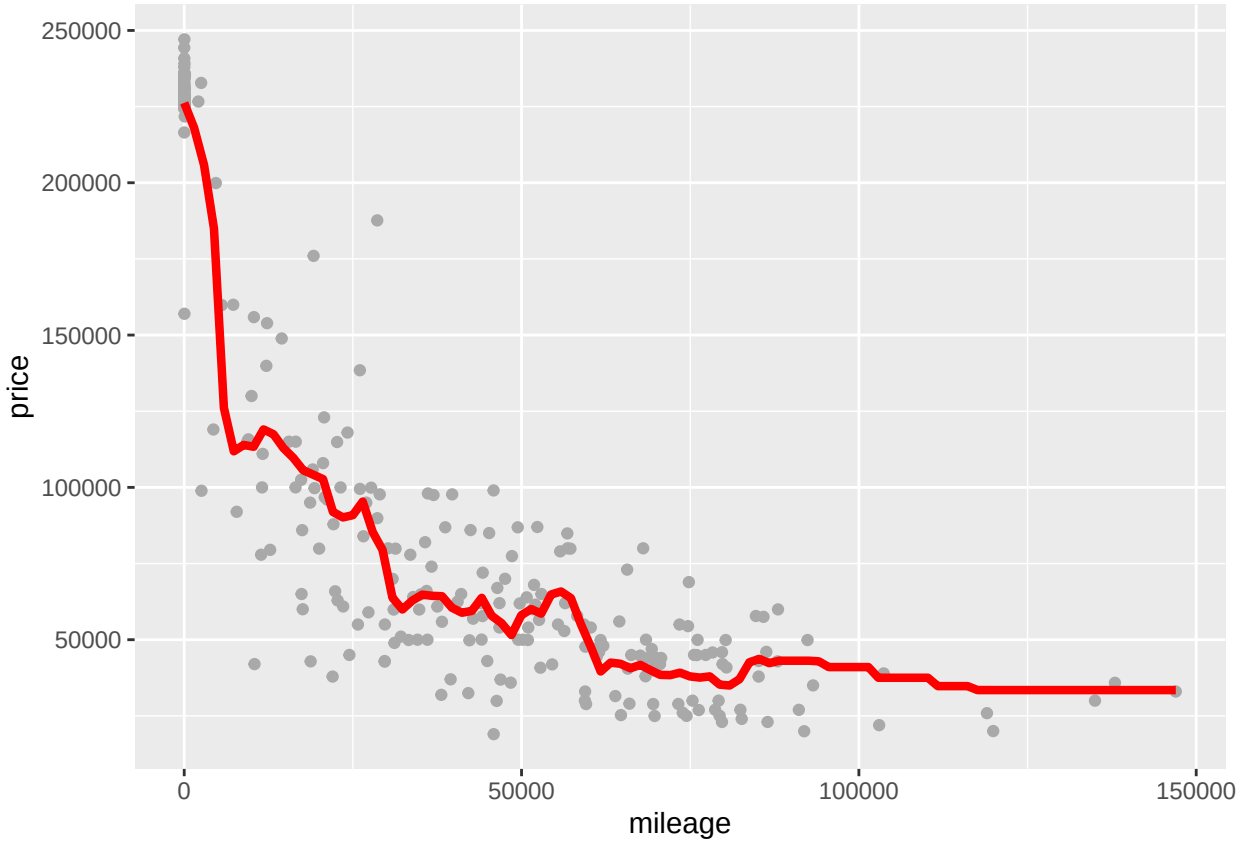
```
## # A tibble: 22 x 2
##        k  armse
##    <dbl>  <dbl>
##  1     2 24025.
##  2     4 21462.
##  3     6 20609.
##  4     8 20438.
##  5    10 20342.
##  6    15 20463.
##  7    20 20361.
##  8    25 20546.
##  9    30 20762.
## 10    35 21192.
## # ... with 12 more rows
```



When K = 10 it bottoms out.

For k = 10 the plot of the fitted model shows as follow:

```
## [1] 19416.18
```



I think 350's yields a larger optimal value of K. Because it has a larger sample size.