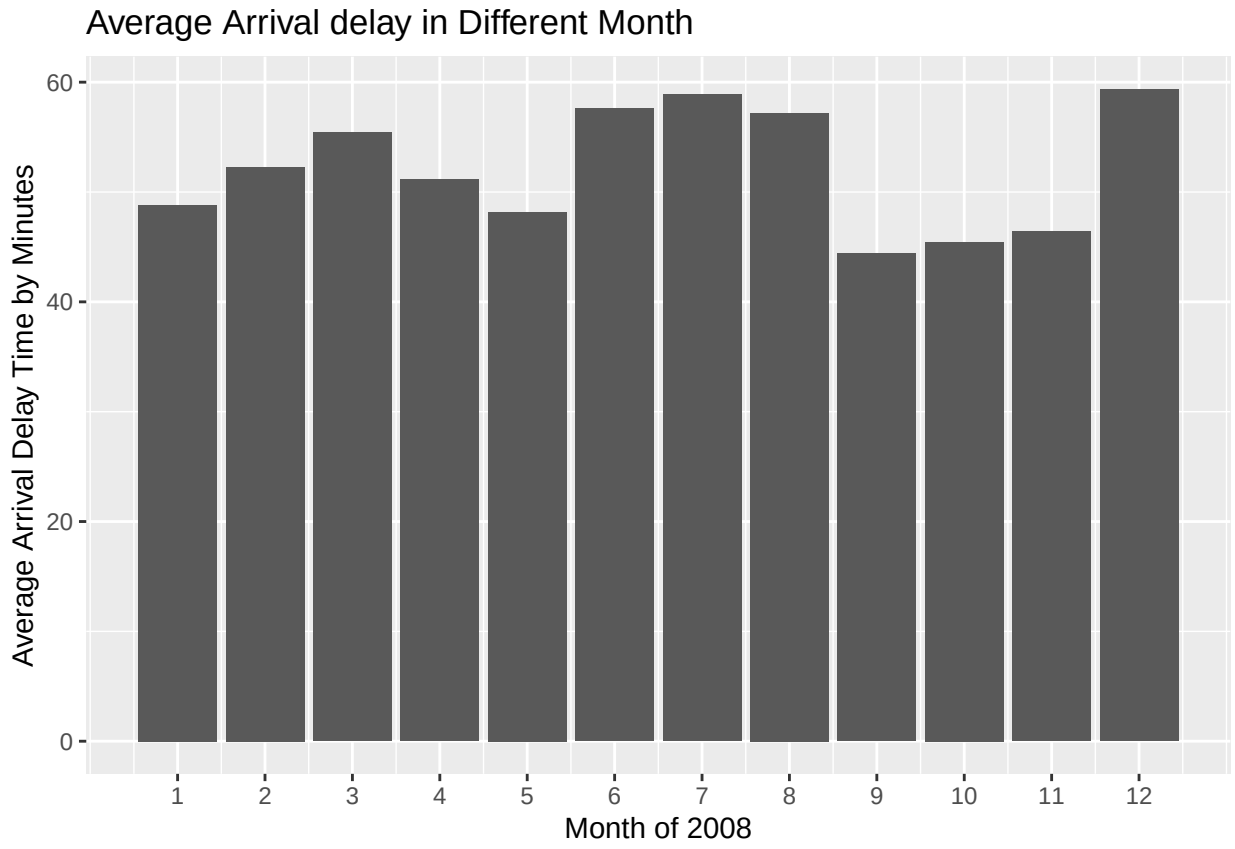
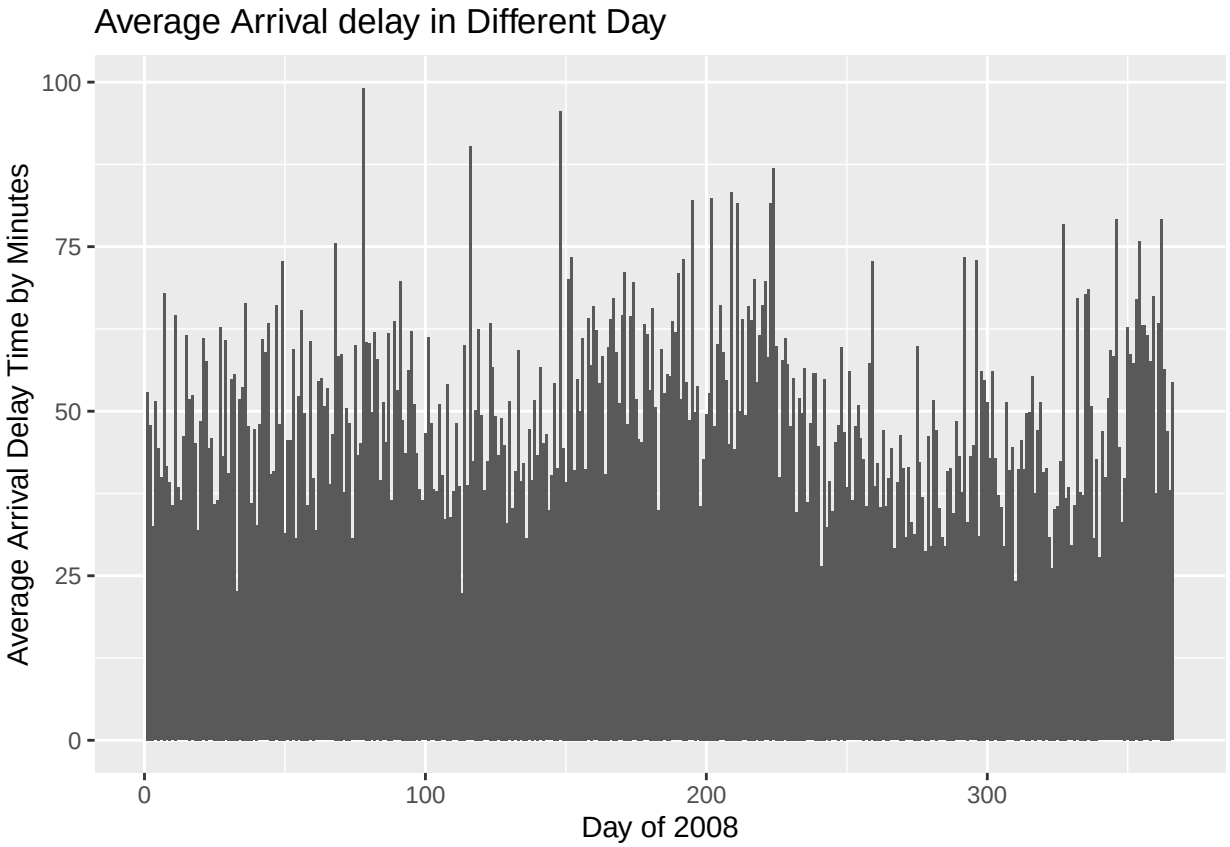
##Problem 1: ABIA We mainly focus on the question: What is the best time of year to fly to minimize delays, and does this change by destination? First, we calculate and plot the "Average Arrival delay in Different Month".

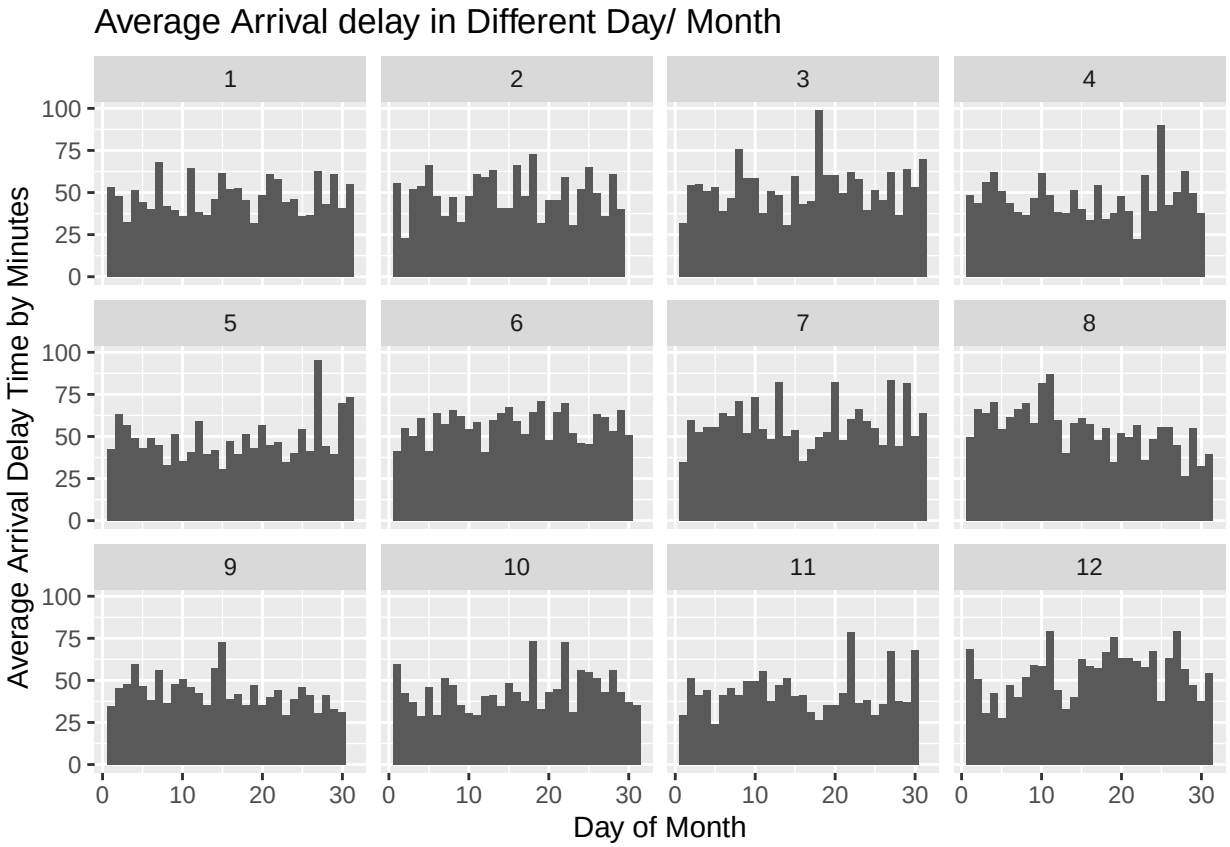


From the plot, we can see that the Arrival delay time of Augest, September and October is relatively lower than other month (lower than 50 minutes). Augest is the lowest.

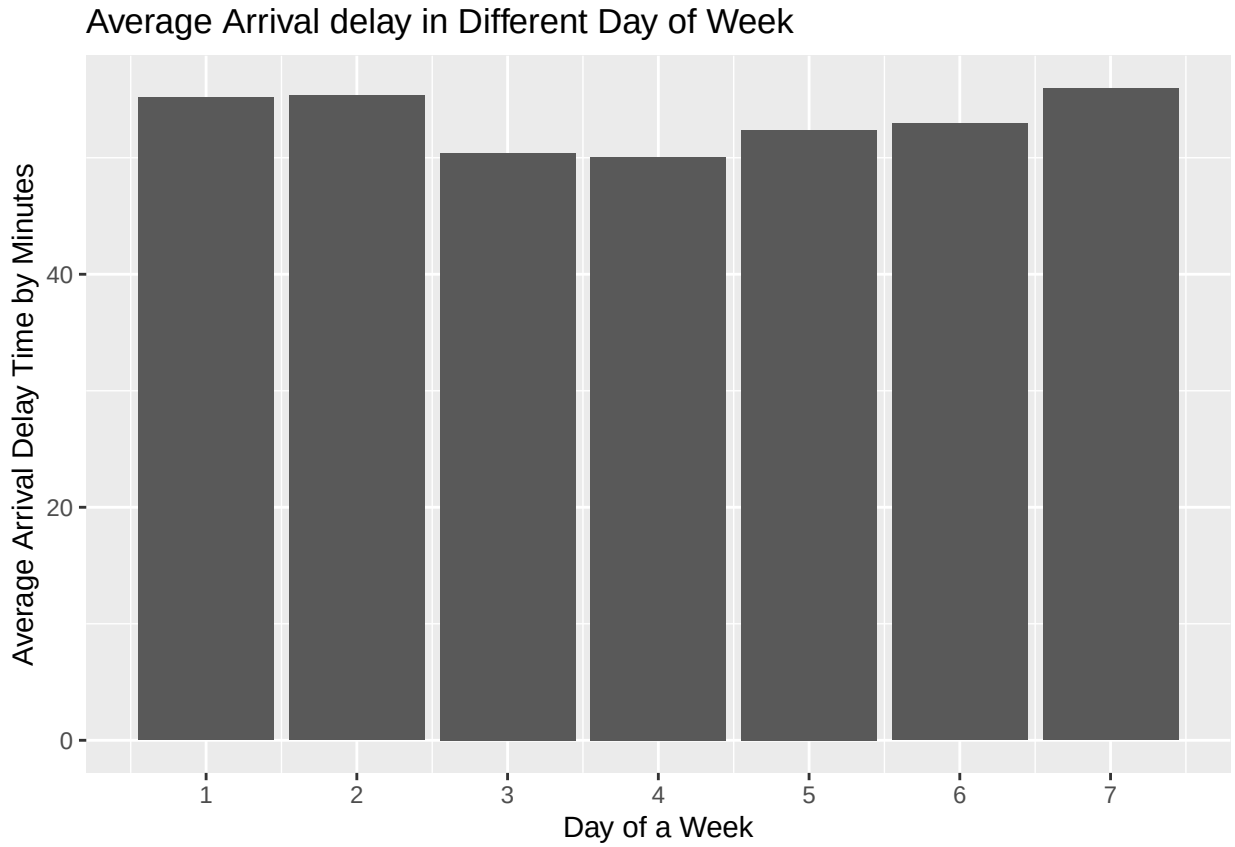
Then what about the different day of 2008? Will there be some trend?

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


Average Arrival delay in Different Day
Average Arrival Delay Time by Minutes
100
75
50
25
0
0
100
200
300
Day of 2008



There is no visible relationship of the delay time with different day but it seems that it has a seasonality. Let's check it.



We can see that Wednesday and Thursday have a relatively low average Arrival delay time.

##We can conclude that the best time of year to fly to minimize delays is Aguest, better to fly on Wednesday or Thursday.

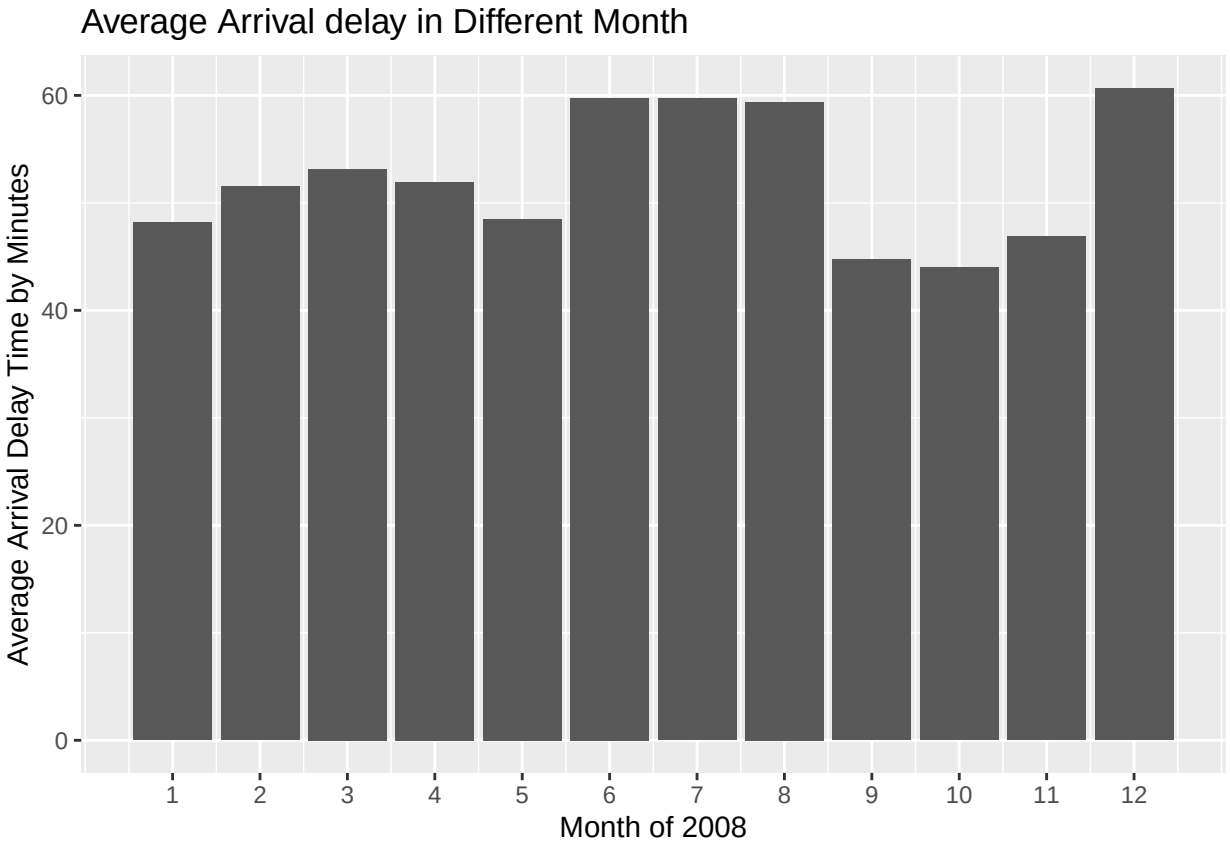
##Check for different destination: First count for the most frequently appeared destination (besides AUS)

```
##    Dest X2008
## 1   ABQ   435
## 2   ATL  2252
## 3   AUS 49637
## 4   BNA   792
## 5   BOS   368
## 6   BWI   730
## 7   CLE   380
## 8   CLT   659
## 9   CVG   653
## 10  DAL  5573
## 11  DEN  2673
## 12  DFW  5506
## 13  DSM     1
## 14  DTW     1
## 15  ELP  1349
## 16  EWR   949
## 17  FLL   481
## 18  HOU  2319
## 19  HRL   367
```

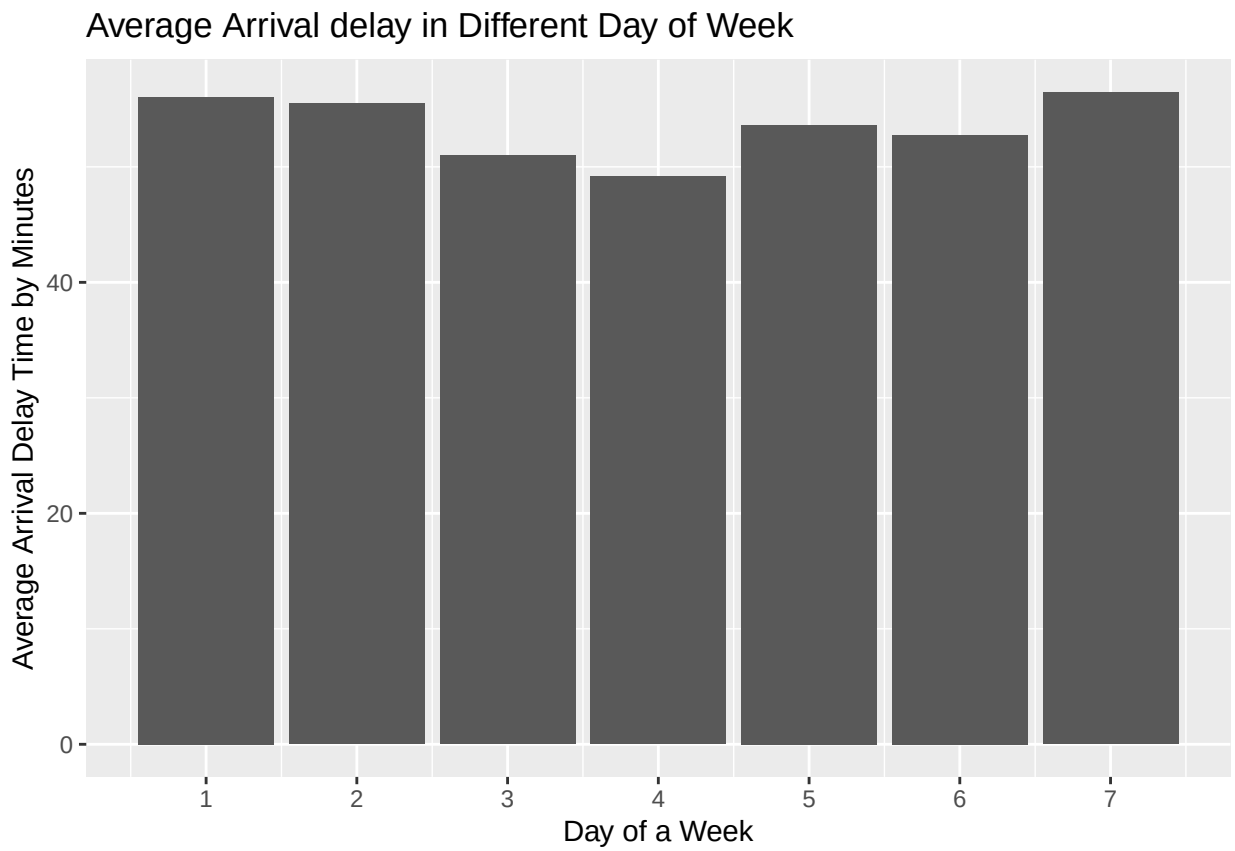
```
## 20  IAD   670
## 21  IAH  3691
## 22  IND   218
## 23  JAX   226
## 24  JFK  1358
## 25  LAS  1231
## 26  LAX  1733
## 27  LBB   692
## 28  LGB   245
## 29  MAF   470
## 30  MCI   459
## 31  MCO   632
## 32  MDW   712
## 33  MEM   834
## 34  MSP    55
## 35  MSY   444
## 36  OAK   236
## 37  OKC    88
## 38  ONT   305
## 39  ORD  2514
## 40  ORF     1
## 41  PHL   290
## 42  PHX  2783
## 43  RDU   231
## 44  SAN   719
## 45  SEA   149
## 46  SFO   610
## 47  SJC   968
## 48  SLC   548
## 49  SNA   245
## 50  STL    95
## 51  TPA   367
## 52  TUL    88
## 53  TUS   228
```

```
##      2008
## AUS 49637
## DAL  5573
## DFW  5506
## IAH  3691
## PHX  2783
## DEN  2673
```

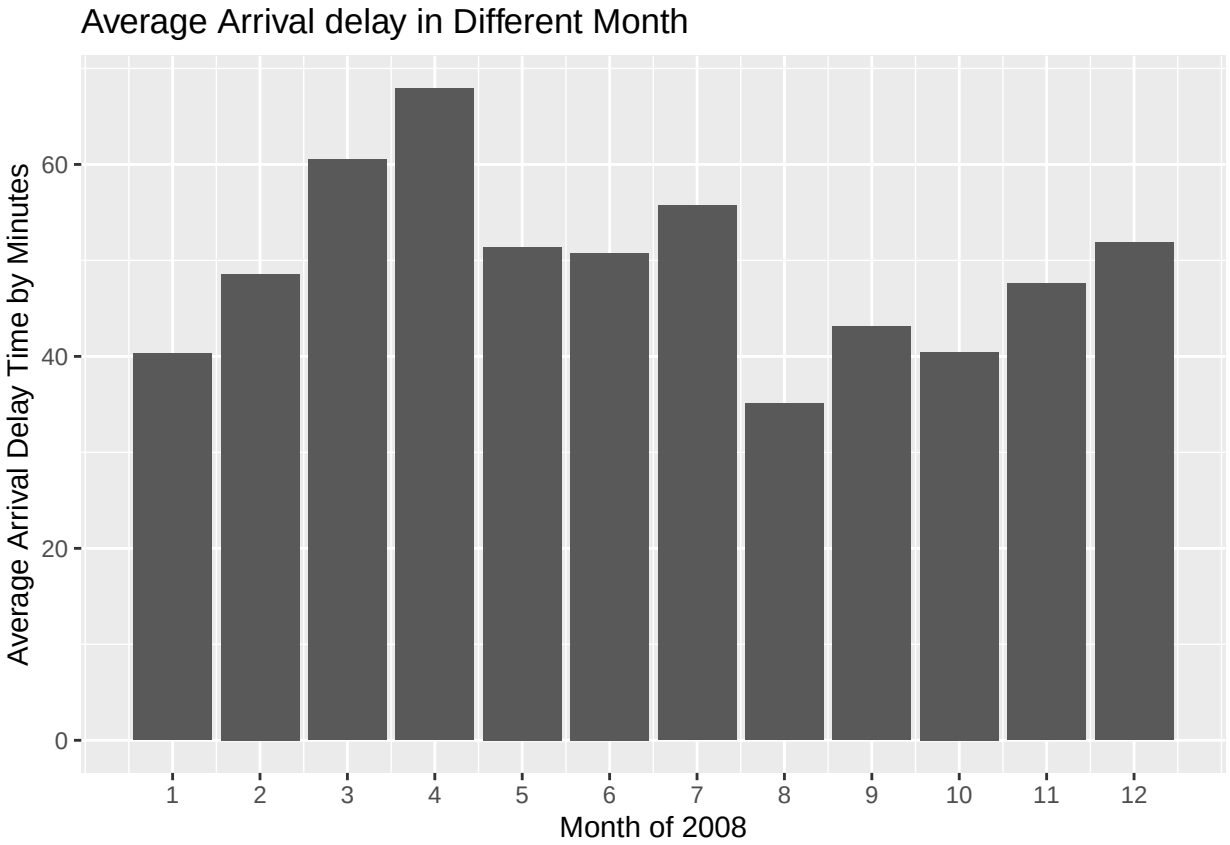
Then we choose the most frequent three destination(above 5000 times): AUS, DAL and DFW to check



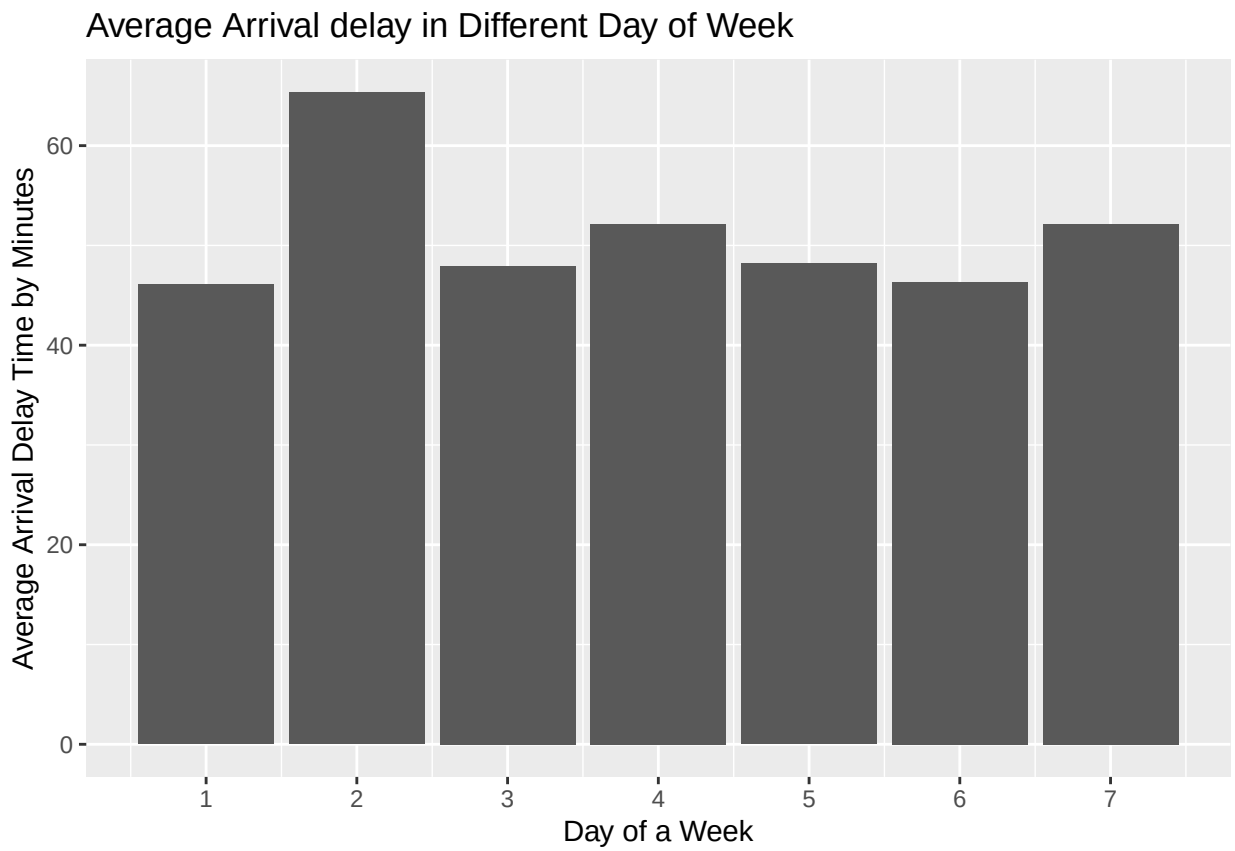
For AUS, October has the smallest average Arrival delay time. Different from the overall situation.



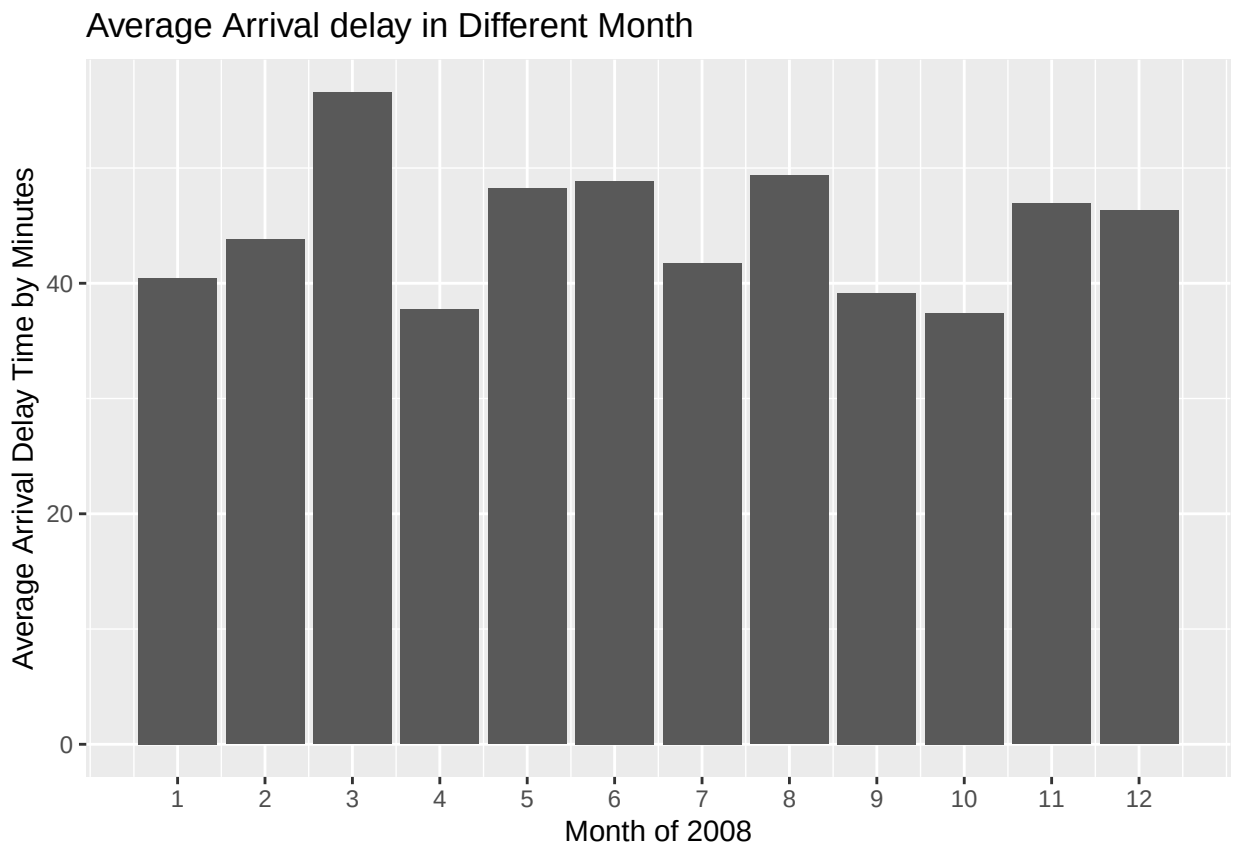
For AUS, roughly same as the overall situation.



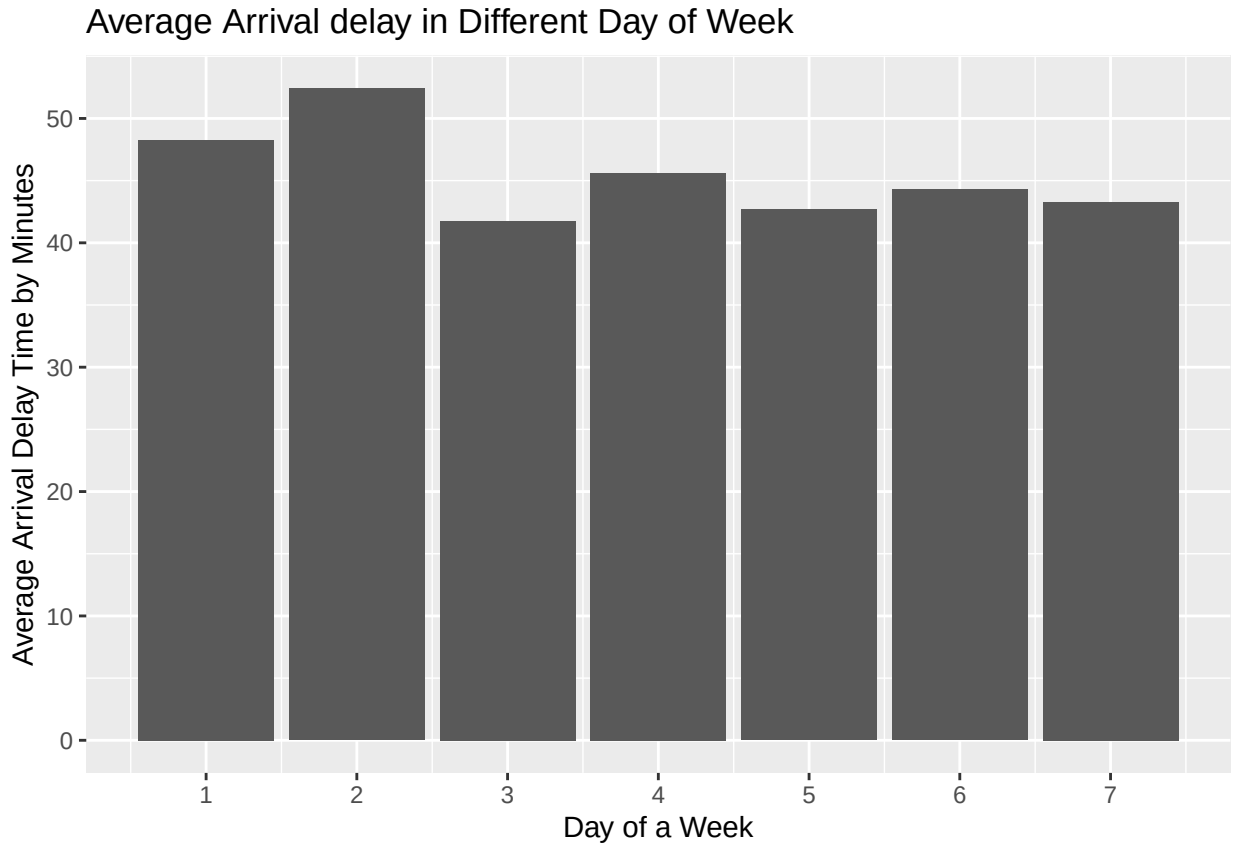
For DAL, Augest has the lowest average Arrival delay time, same as the overall situation.



For DAL, Monday has the lowest average arrival delay time, different from the overall situation.



For DFW, October has the smallest average Arrival delay time. Different from the overall situation.



For DFW, Wednesday has the lowest average arrival delay time, roughly same as the overall situation.

##The best time of year/week changes by destination.

##Problem 2: Olympic A)

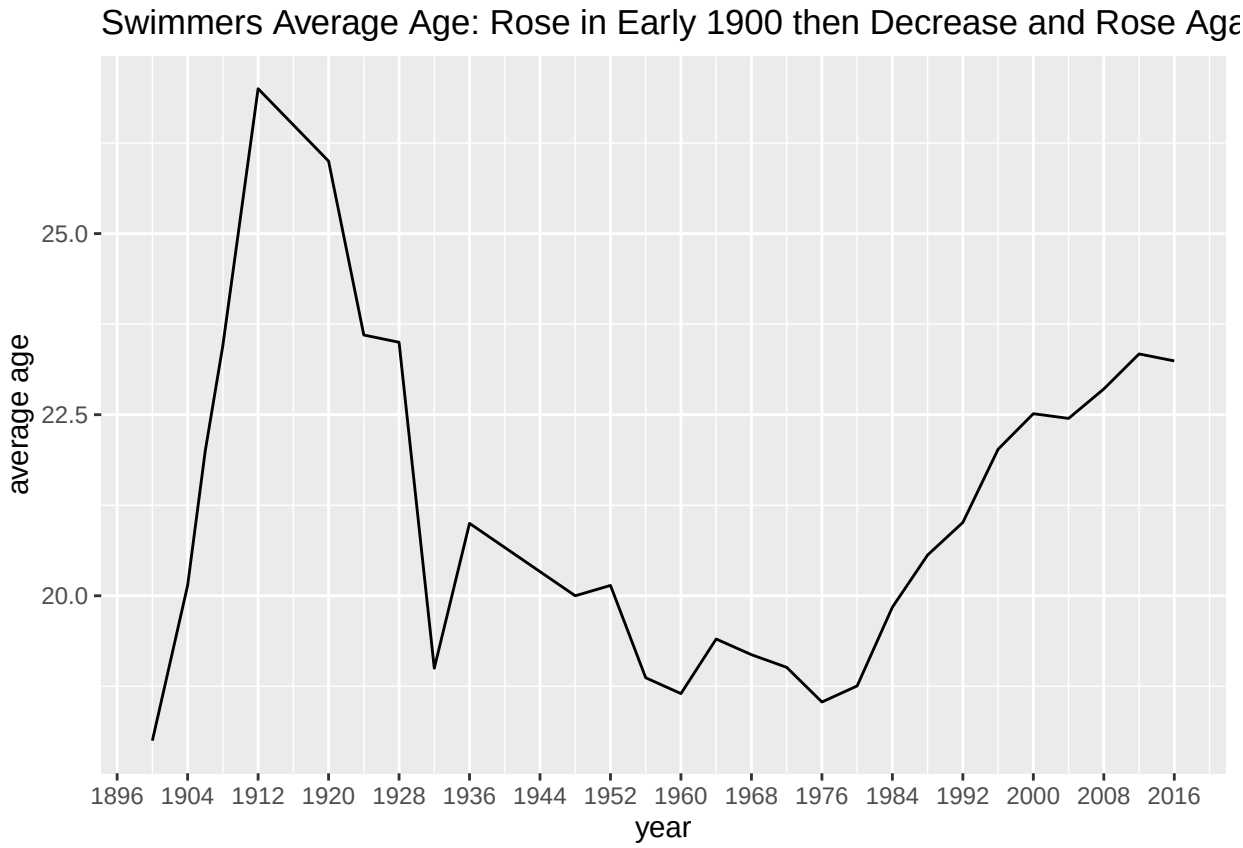
```
## 95%
## 183
```

B)

```
## # A tibble: 10 x 3
##    event                                mean_h  sd_h
##    <chr>                                 <dbl> <dbl>
##  1 Rowing Women's Coxed Fours             173. 10.9
##  2 Basketball Women's Basketball          183.  9.70
##  3 Rowing Women's Coxed Quadruple Sculls  172.  9.25
##  4 Rowing Women's Coxed Eights            178.  8.74
##  5 Swimming Women's 100 metres Butterfly  173.  8.13
##  6 Volleyball Women's Volleyball          180.  8.10
##  7 Gymnastics Women's Uneven Bars         155   8.02
##  8 Shooting Women's Double Trap           169.  7.83
##  9 Cycling Women's Keirin                 170.  7.76
## 10 Swimming Women's 400 metres Freestyle  174.  7.62
```

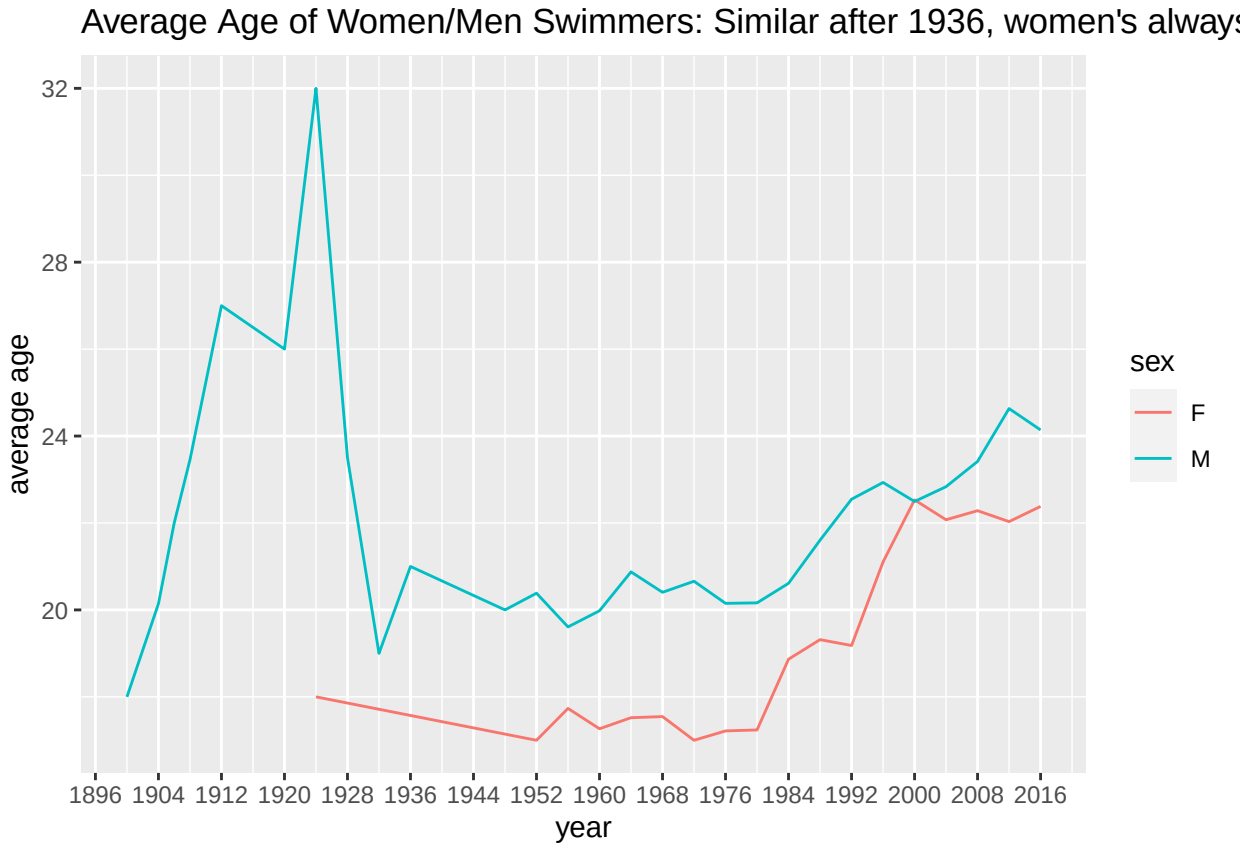
For single women's event, "Swimming Women's 100 metres Butterfly" had the greatest variability in competitor's heights across the entire history of the Olympics, as measured by the standard deviation.

C)



In the begin of 1900, the average age rose rapidly, and then decreased rapidly and roughly remain around 20. After 1980, it rose up again.

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

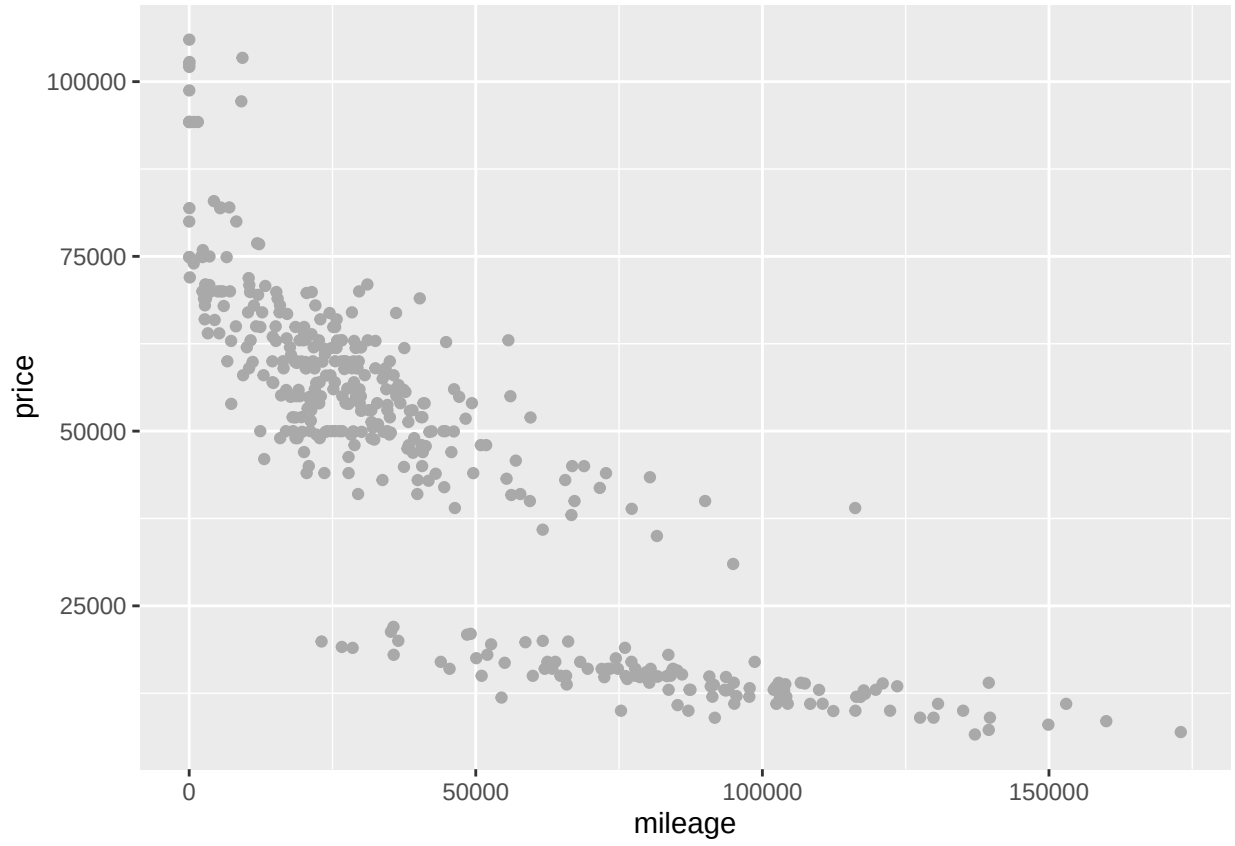


The trend looks similar for male swimmers relative to female swimmers after 1936. But female's average age is always lower than male.

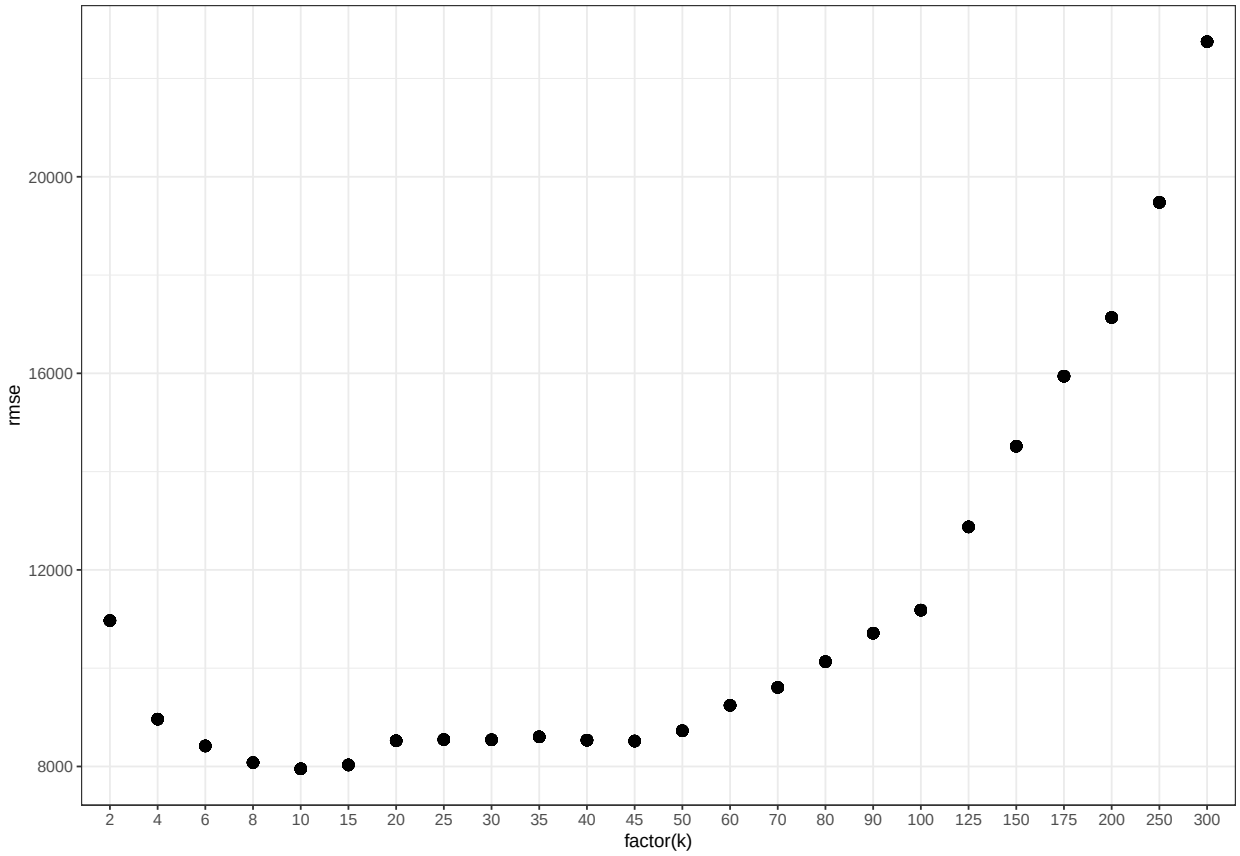
Problem 3: sclass For trim = 350

```
##        id            trim             subTrim           condition
##  Min.   :  282   Length:416         Length:416         Length:416
##  1st Qu.:14290   Class :character   Class :character   Class :character
##  Median :26658   Mode  :character   Mode  :character   Mode  :character
##  Mean   :26520
##  3rd Qu.:39599
##  Max.   :52220
##   isOneOwner           mileage            year         color
##  Length:416         Min.   :     6   Min.   :1994   Length:416
##  Class :character   1st Qu.: 19264   1st Qu.:2006   Class :character
##  Mode  :character   Median : 29998   Median :2012   Mode  :character
##                     Mean   : 42926   Mean   :2010
##                     3rd Qu.: 63479   3rd Qu.:2012
##                     Max.   :173000   Max.   :2013
##  displacement          fuel              state              region
##  Length:416         Length:416         Length:416         Length:416
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  soundSystem         wheelType          wheelSize          featureCount
```

```
##  Length:416         Length:416         Length:416         Min.   :  0.00
##  Class :character   Class :character   Class :character   1st Qu.: 31.75
##  Mode  :character   Mode  :character   Mode  :character   Median : 54.00
##                                                           Mean   : 49.22
##                                                           3rd Qu.: 70.00
##                                                           Max.   :112.00
##      price
##  Min.   :  6600
##  1st Qu.: 19401
##  Median : 52900
##  Mean   : 46854
##  3rd Qu.: 61991
##  Max.   :106010
```

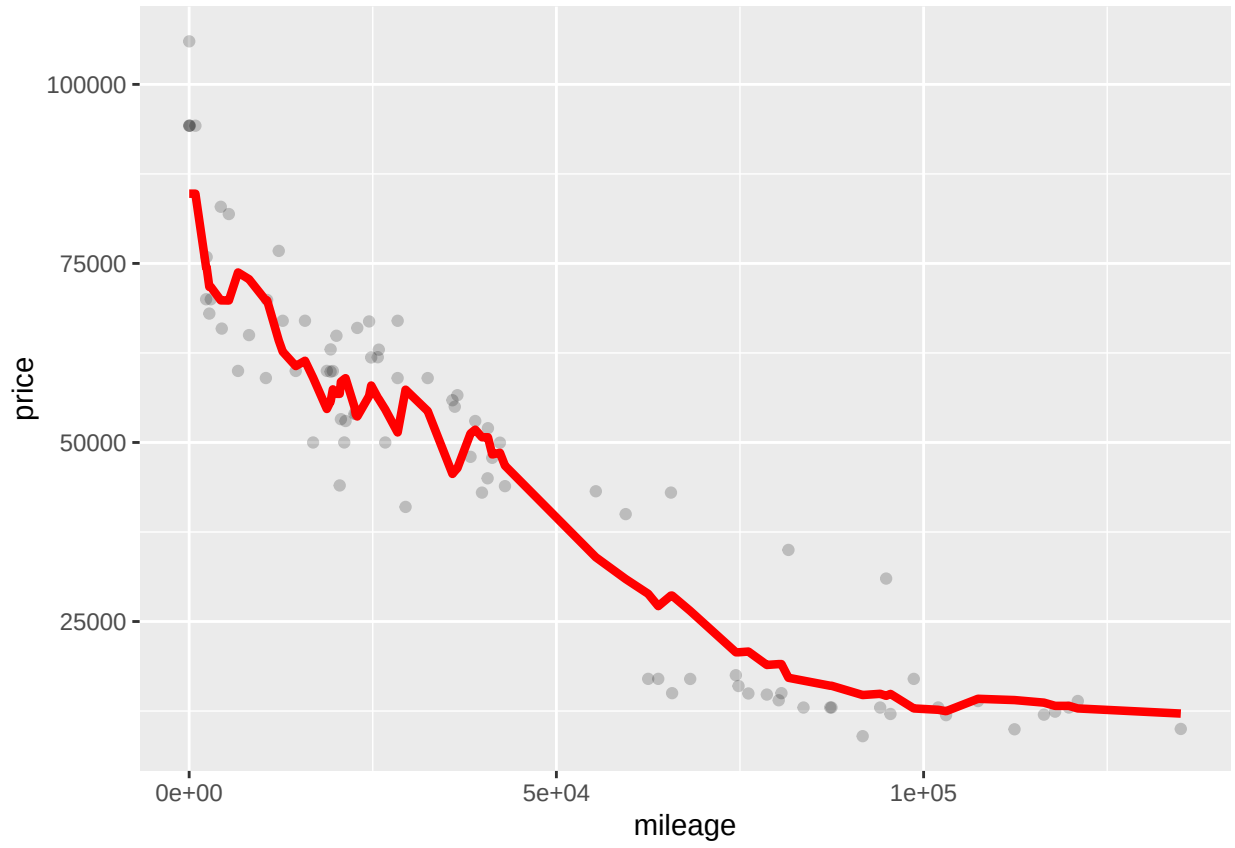



```
## Warning: executing %dopar% sequentially: no parallel backend registered
```



We can see that the bottom is k = 15.

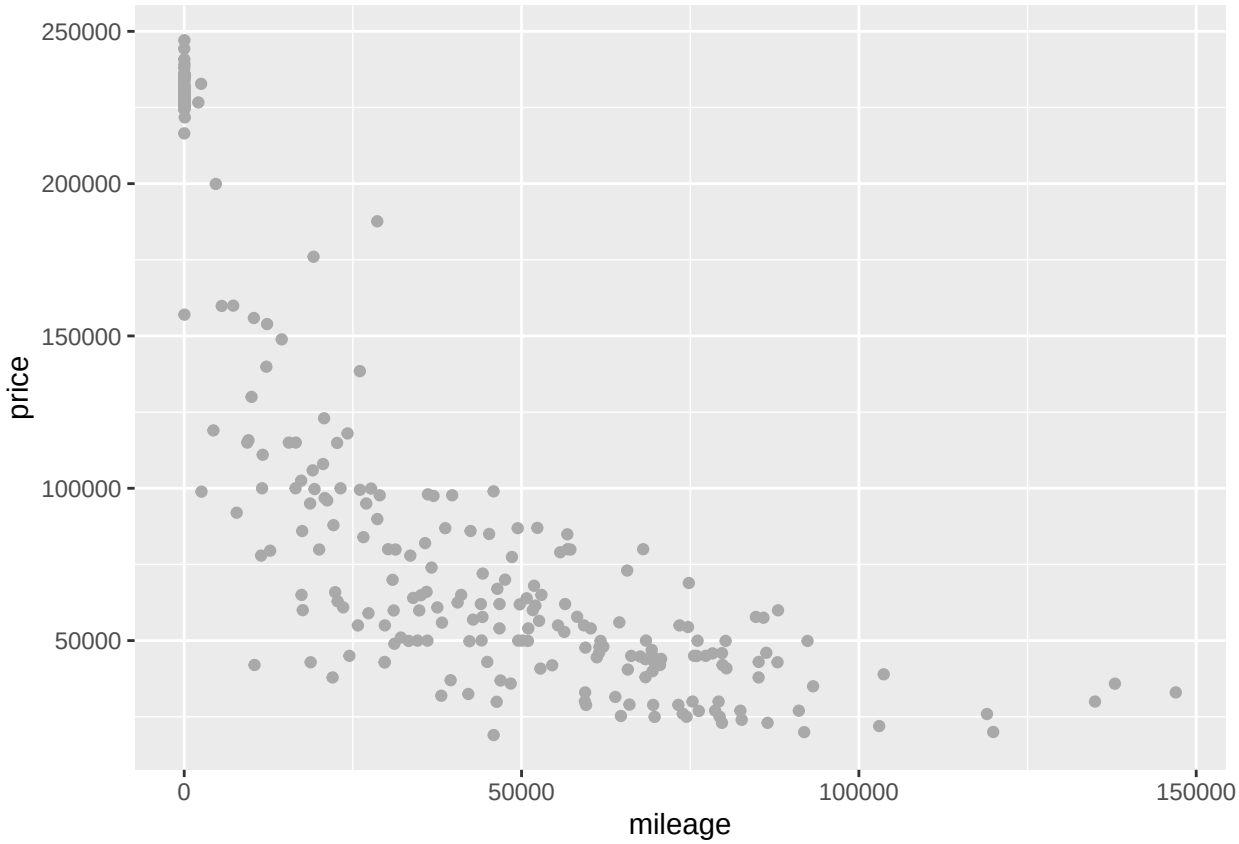
```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
```



For Trim = 65 AGM

```
##        id             trim             subTrim           condition
##  Min.   : 1060   Length:292         Length:292         Length:292
##  1st Qu.:13977   Class :character   Class :character   Class :character
##  Median :26557   Mode  :character   Mode  :character   Mode  :character
##  Mean   :26444
##  3rd Qu.:38687
##  Max.   :52326
##   isOneOwner           mileage            year         color
##  Length:292         Min.   :     1   Min.   :2006   Length:292
##  Class :character   1st Qu.:    20   1st Qu.:2007   Class :character
##  Mode  :character   Median : 28803   Median :2010   Mode  :character
##                     Mean   : 33700   Mean   :2010
##                     3rd Qu.: 58496   3rd Qu.:2015
##                     Max.   :146975   Max.   :2015
##  displacement           fuel              state              region
##  Length:292         Length:292         Length:292         Length:292
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  soundSystem         wheelType          wheelSize          featureCount
##  Length:292         Length:292         Length:292         Min.   :  0.00
##  Class :character   Class :character   Class :character   1st Qu.: 17.00
```

```
##  Mode  :character   Mode  :character   Mode  :character   Median : 58.00
##                                                          Mean   : 48.09
##                                                          3rd Qu.: 72.00
##                                                          Max.   :112.00
##      price
##  Min.   : 18990
##  1st Qu.: 48711
##  Median : 79995
##  Mean   :117121
##  3rd Qu.:225975
##  Max.   :247075
```

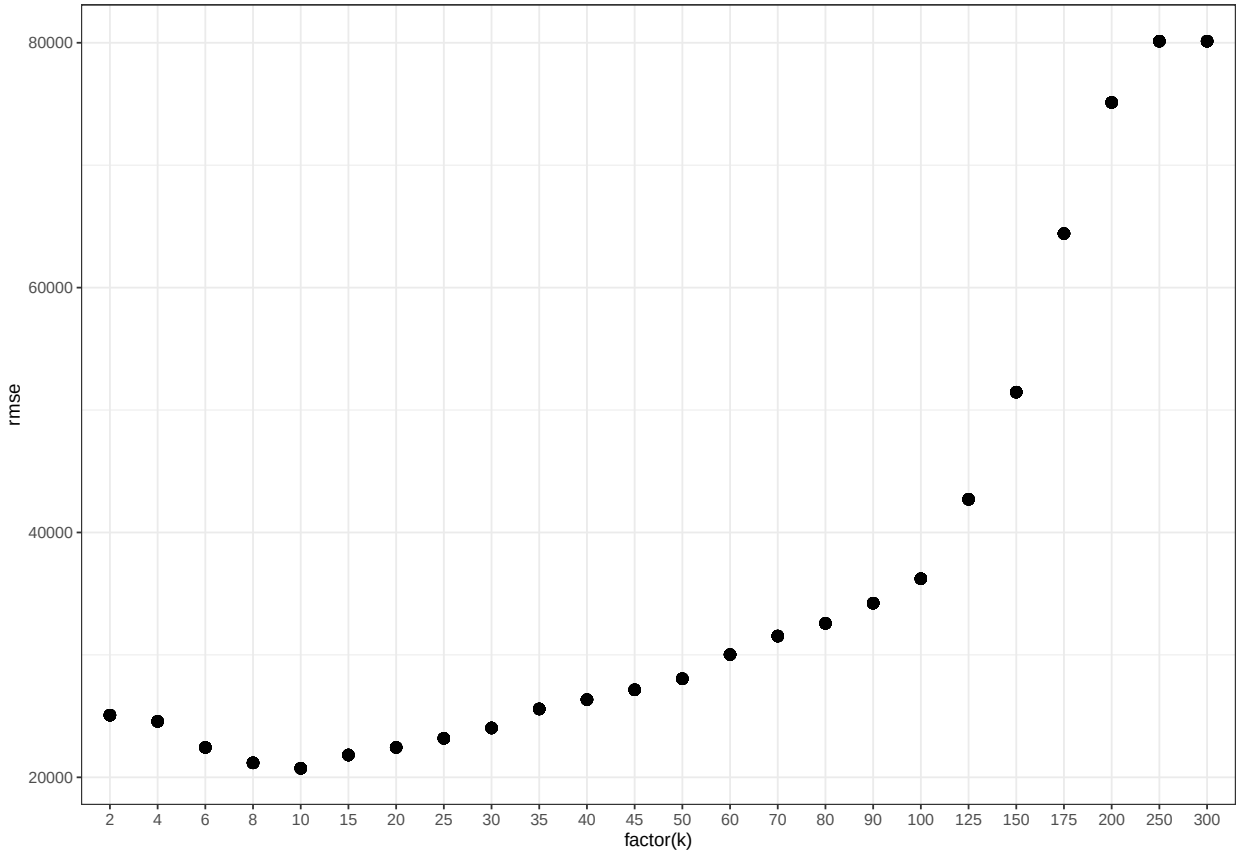



```
## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns
```

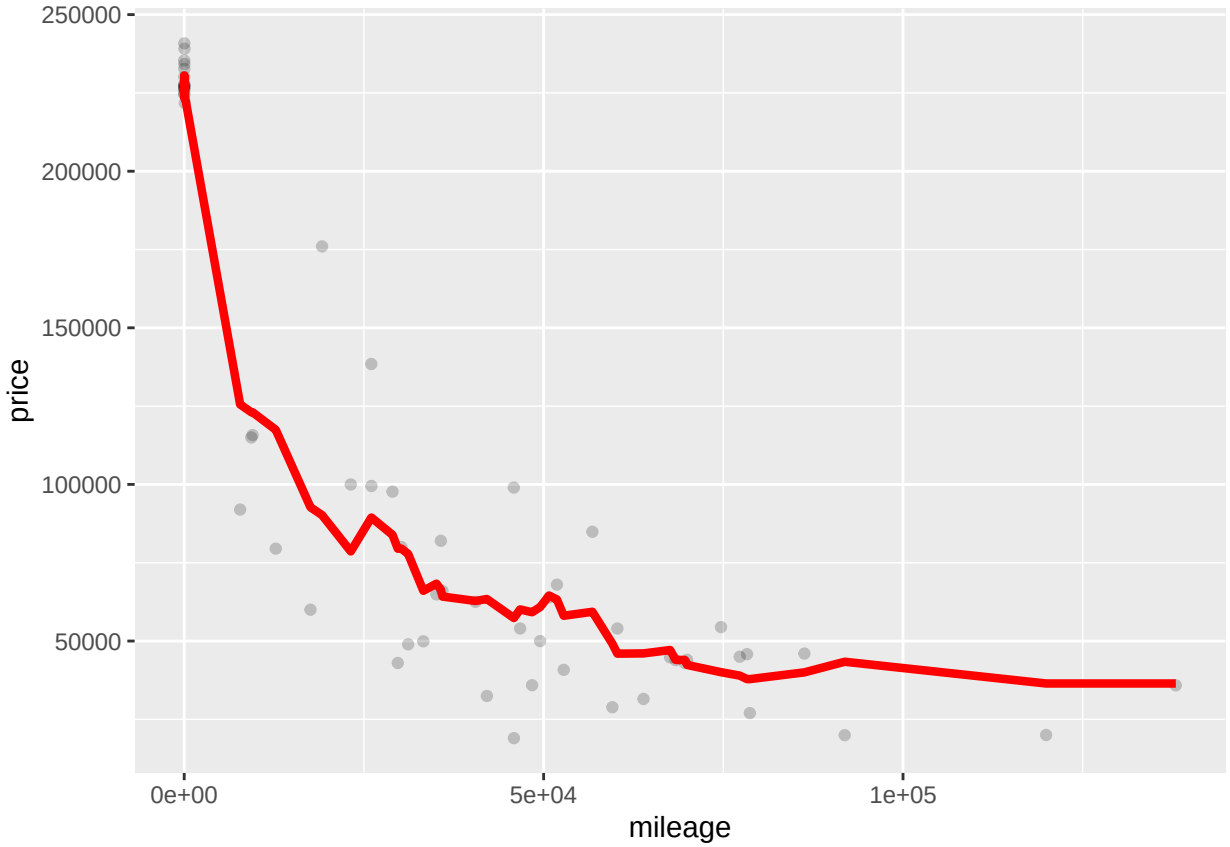
```
## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns
```

```
## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns

## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 250 exceeds number 233 of patterns
```

```
## Warning in knnregTrain(train = structure(c(23, 36052, 91102, 17, 64497, : k =
## 300 exceeds number 233 of patterns
```



We can see that the bottom is k = 10.



The k of trim = 65 AMG is lower than trim = 350. I think it's because 350's sample size is larger than 65 AMG, so k can be a little larger.