HW1

Yuri Lee, Seunghoon Choi

2021 2 7

1-1

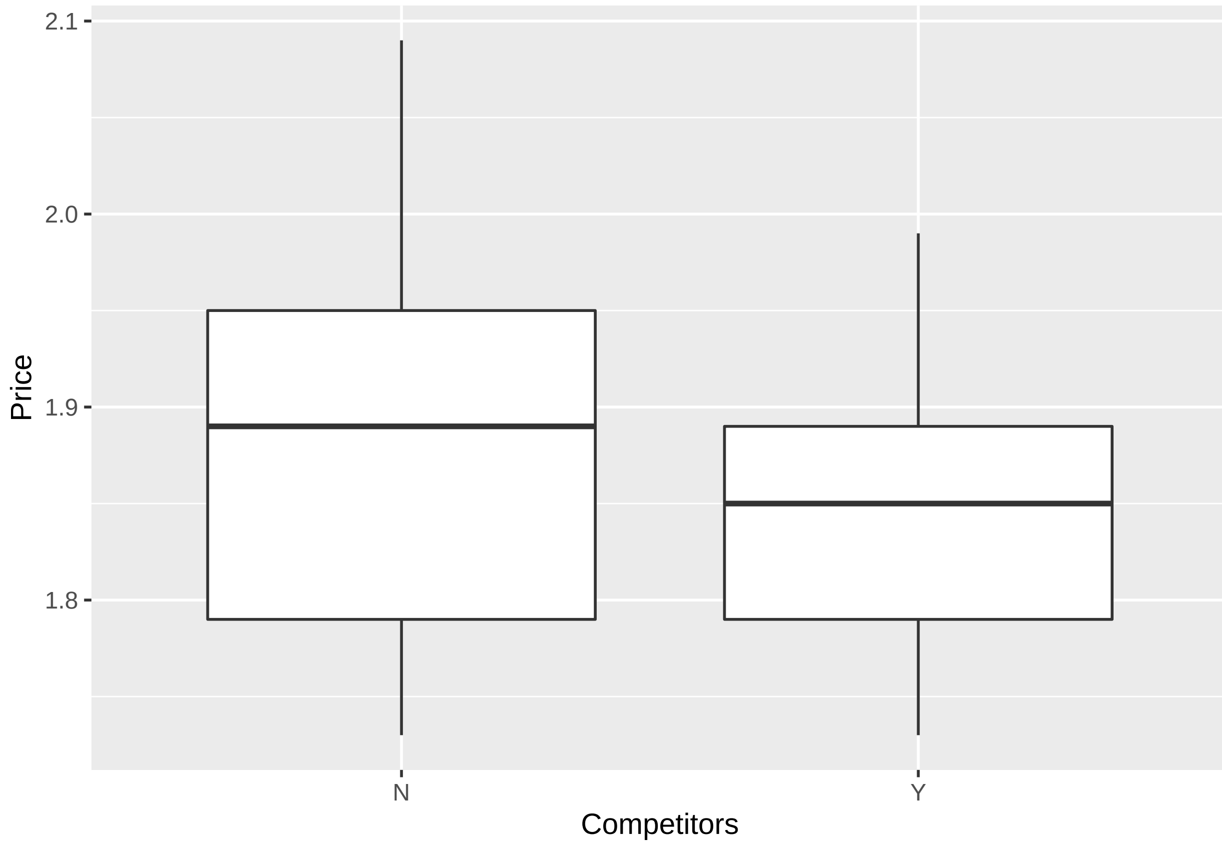
```r
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------------------------------------------

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------------------------------ tid
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)

GasPrices = read.csv('C:/Users/CHOI/Desktop/GasPrices.csv')
```

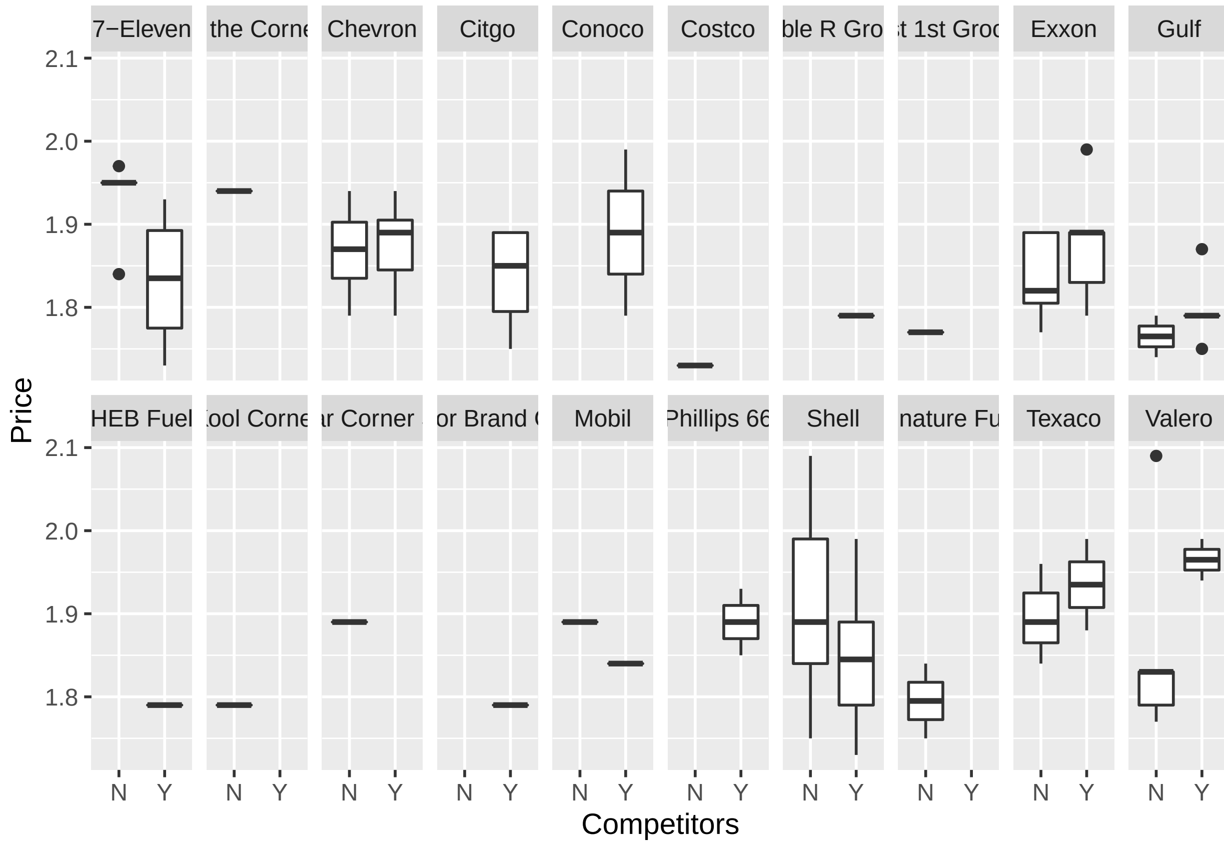
A. Competition & Price

```r
ggplot(data=GasPrices) +
  geom_boxplot(aes(x = Competitors, y=Price))
```



The bar plot above shows the gas price of gas providers which have competitors is lower than the price of the providers which do not have competitors.

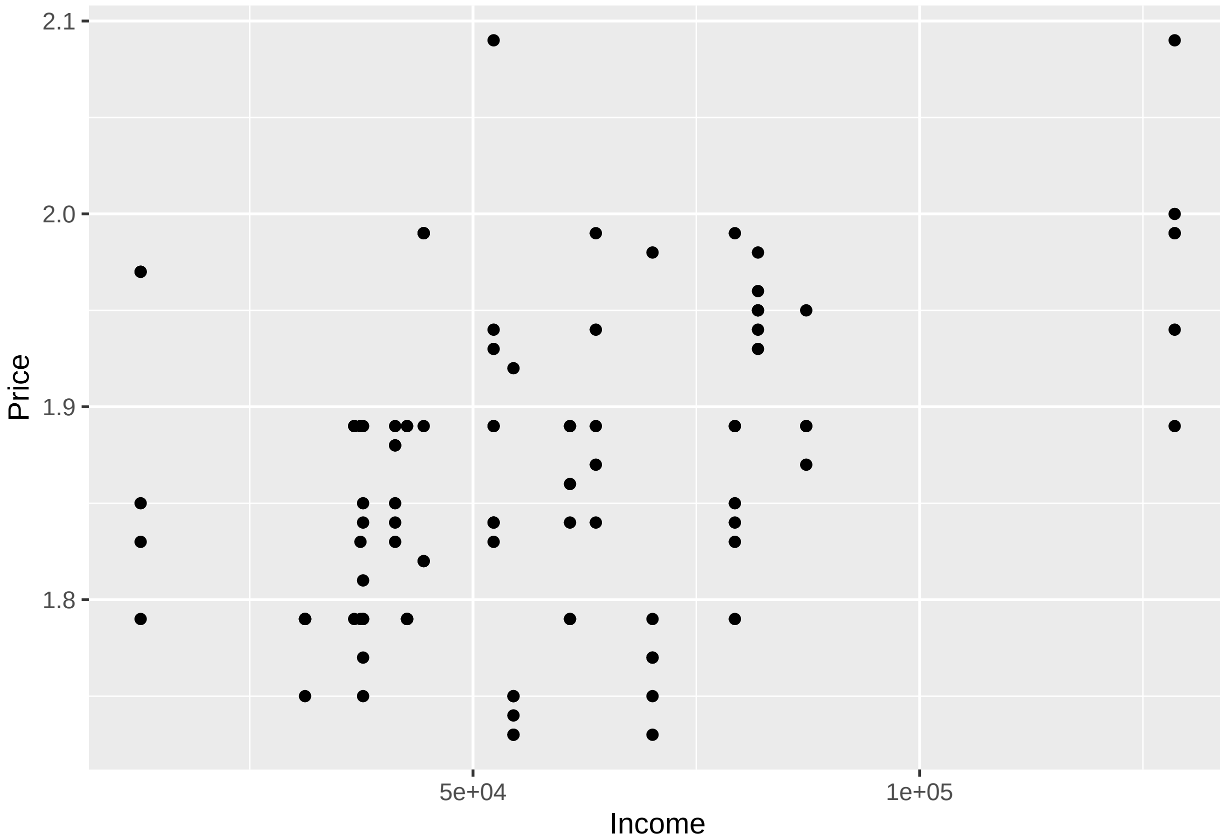
```
ggplot(data=GasPrices) +
  geom_boxplot(aes(x = Competitors, y=Price)) +
  facet_wrap(~Name, nrow=2)
```



However, it would be hard to generalize the relation between the price and the existence of competitors for all providers. Only three providers shows lower price when they have competitors than without-competitors cases among the eight eligible cases out of twenty, whose with-competitor prices and without-copetitors prices can be compared.

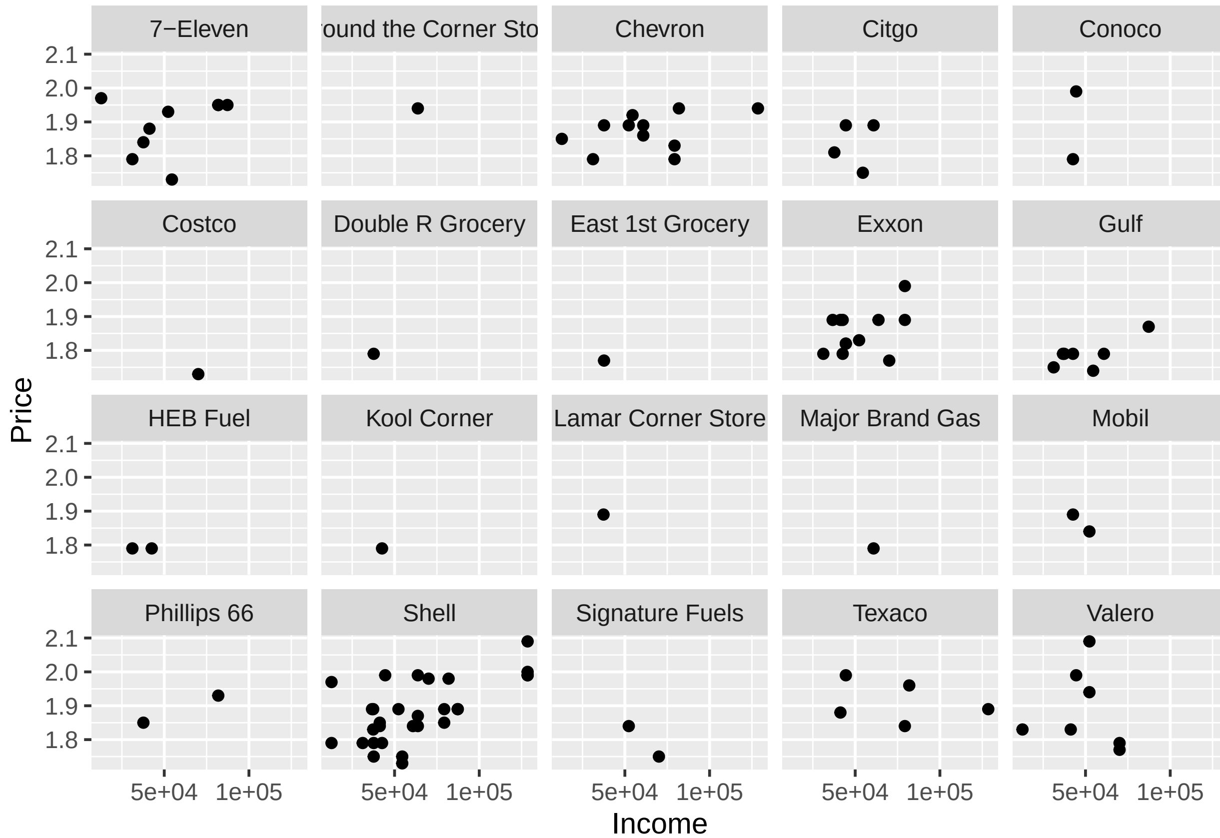
B. Income & Price

```
ggplot(data=GasPrices) +
  geom_point(mapping = aes(x=Income, y=Price))
```



We can see upward shape of dots in this graph, which means the gas prices and income of the area where the gas station is located have a positive relation.

```
ggplot(data=GasPrices) +
  geom_point(mapping = aes(x=Income, y=Price)) +
  facet_wrap(~Name, nrow=4)
```



On the graph of each company, several companies such as 7-Eleven, Exxon, of Shell represent these positive relation obviously. On the contrary, we can see that some companies like Costco sticks to one-price polices.

C. Price of Shell vs Other sellers

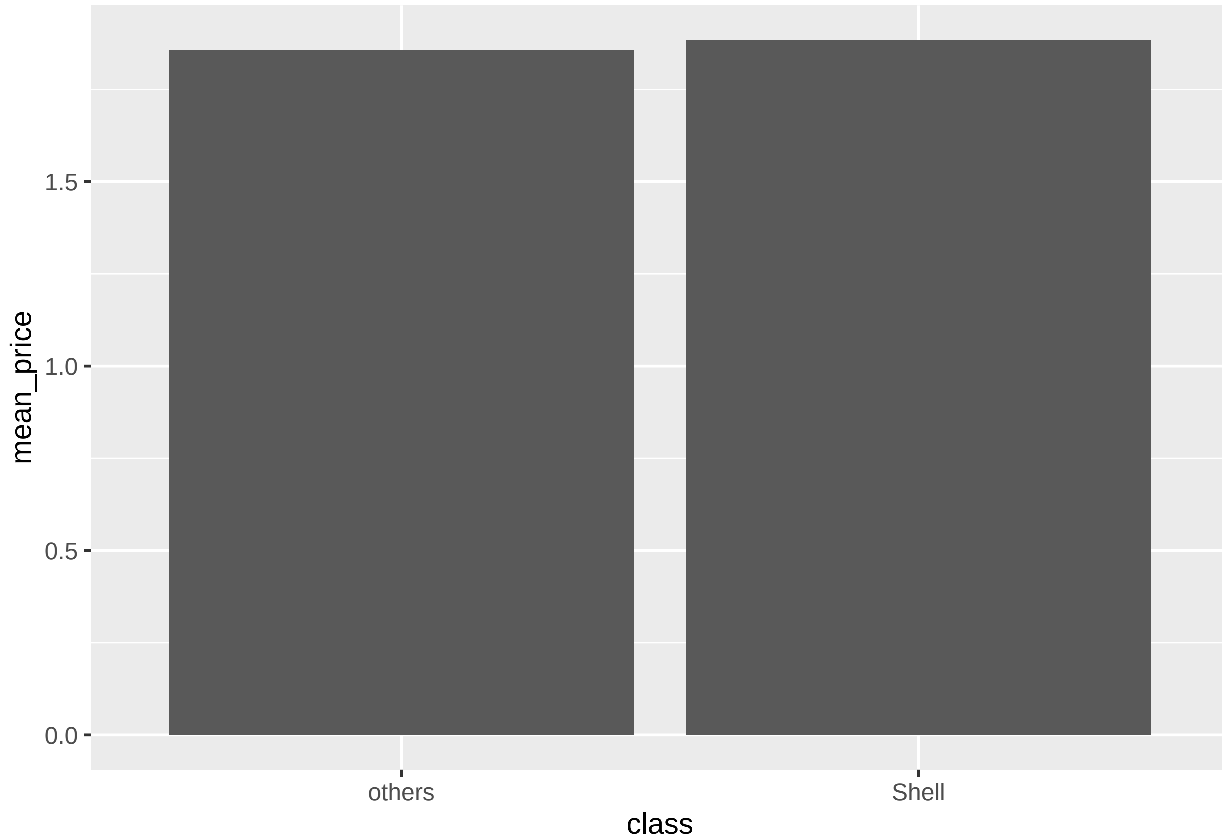
```
GasPrices = GasPrices %>%
  mutate(class = ifelse(Name == 'Shell', 'Shell', 'others'))

d1 = GasPrices %>%
  group_by(class) %>%
  summarise(mean_price = mean(Price))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
d1
```

```
## # A tibble: 2 x 2
##   class  mean_price
##   <chr>       <dbl>
## 1 others       1.86
## 2 Shell        1.88
```

```
ggplot(data=d1) +
  geom_col(mapping=aes(x=class, y=mean_price), position = 'dodge')
```



The average gas price of Shell($1.88) is a little bit higher than that of other providers($1.86).

```
d2 = GasPrices %>%
  group_by(Name) %>%
  summarise(mean_price = mean(Price))
```

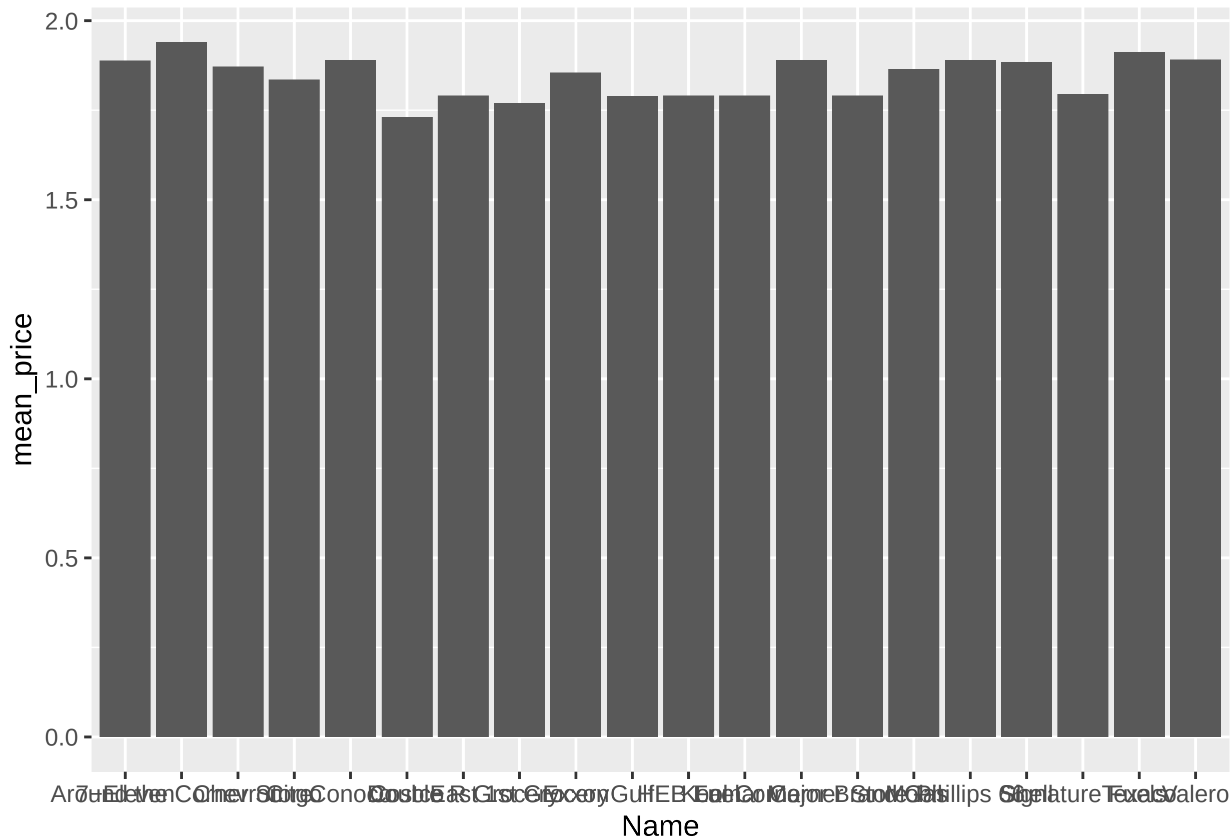
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
d2
```

```
## # A tibble: 20 x 2
##    Name                       mean_price
##    <chr>                           <dbl>
##  1 7-Eleven                         1.89
##  2 Around the Corner Store          1.94
##  3 Chevron                          1.87
##  4 Citgo                            1.84
##  5 Conoco                           1.89
##  6 Costco                           1.73
##  7 Double R Grocery                 1.79
##  8 East 1st Grocery                 1.77
##  9 Exxon                            1.86
## 10 Gulf                             1.79
## 11 HEB Fuel                         1.79
## 12 Kool Corner                      1.79
## 13 Lamar Corner Store               1.89
## 14 Major Brand Gas                  1.79
## 15 Mobil                            1.86
```

```
## 16 Phillips 66                      1.89
## 17 Shell                           1.88
## 18 Signature Fuels                 1.80
## 19 Texaco                          1.91
## 20 Valero                          1.89
```

```r
ggplot(data=d2) +
  geom_col(mapping=aes(x=Name, y=mean_price), position = 'dodge')
```

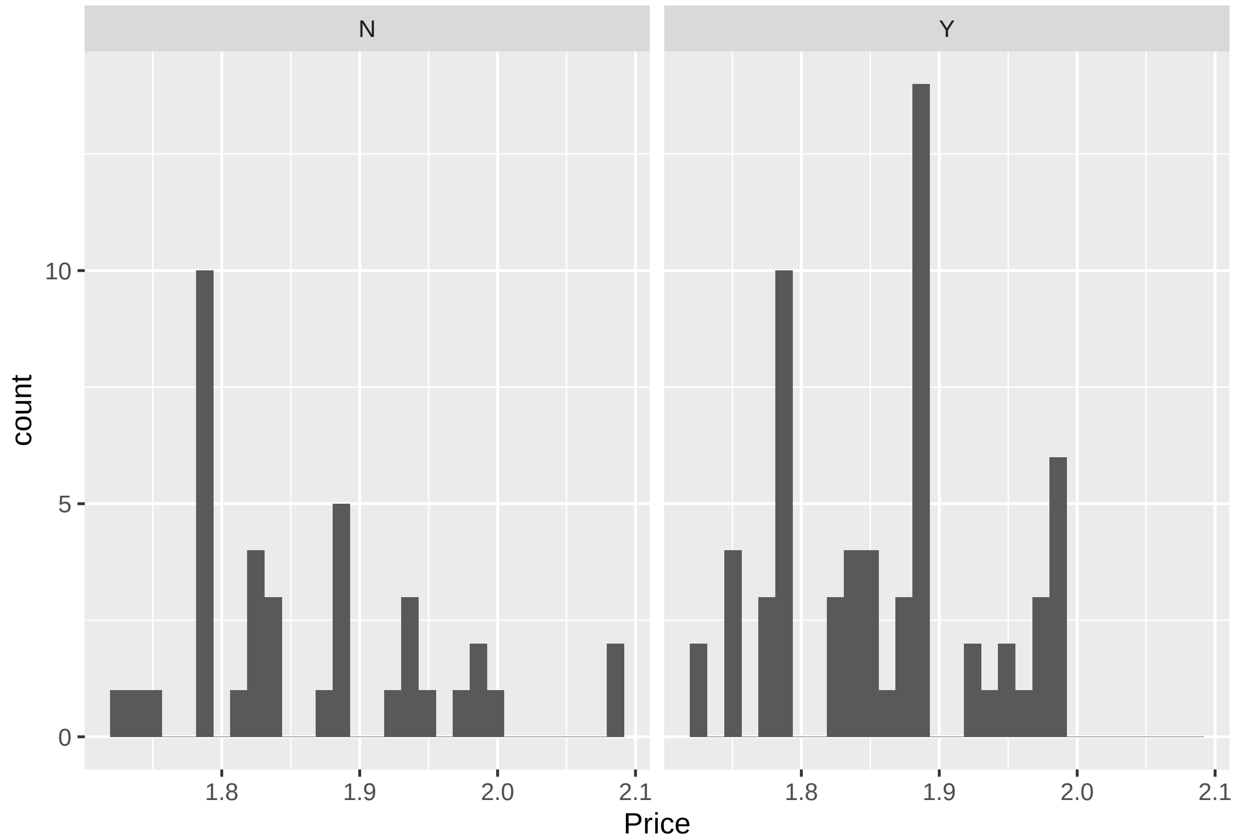


The 12 out of 19 providers have lower average gas prices than that of Shell.

D. stoplights' effects on Price

```r
ggplot(data = GasPrices) +
  geom_histogram(aes(x=Price)) +
  facet_wrap(~ Stoplight)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Gas stations nearby stoplights generally have higher gas prices. Prices of gas stations without stoplight nearby(the left graph) are concentrated around 1.8, while gas stations near stoplight have a lot of prices around 1.8~1.9.

E. The effect of Highway access on Price

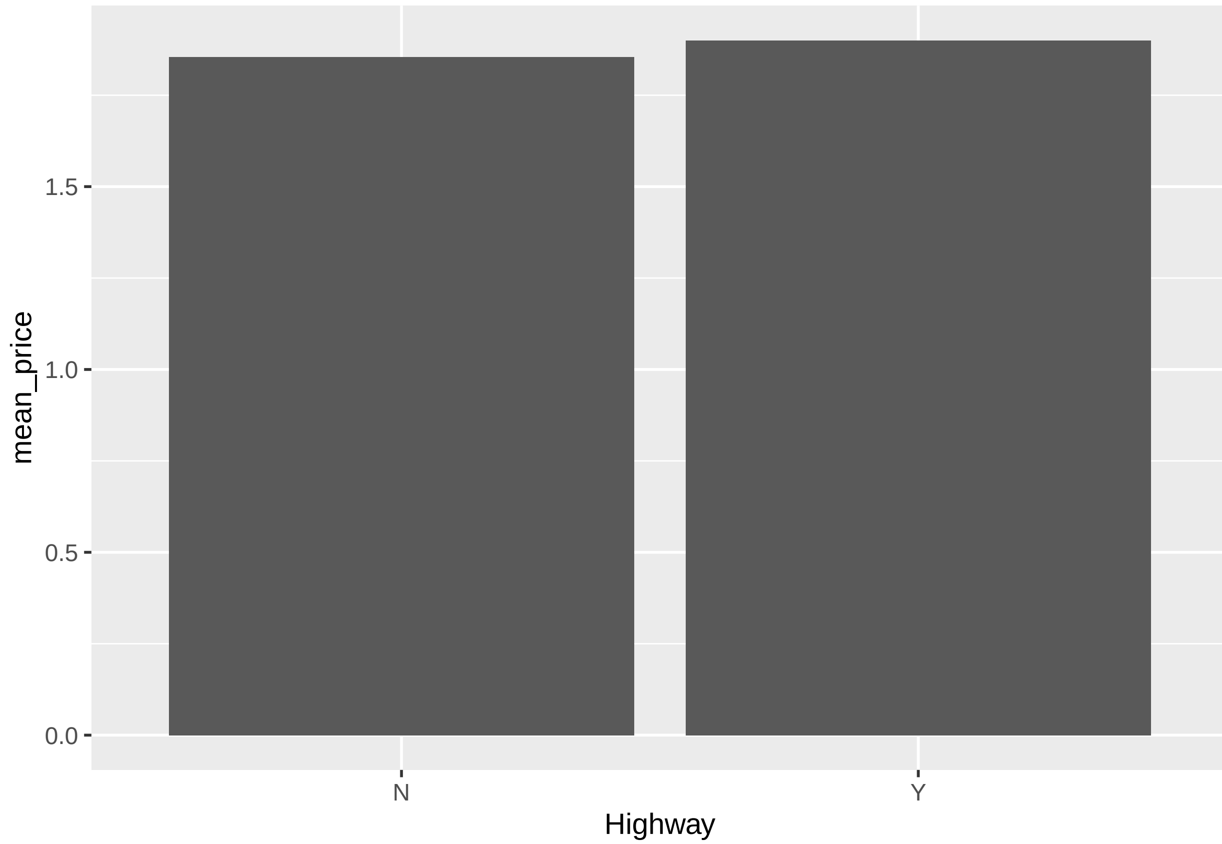
```
d3 = GasPrices %>%
  group_by(Highway) %>%
  summarise(mean_price = mean(Price))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
d3
```

```
## # A tibble: 2 x 2
##   Highway mean_price
##   <chr>        <dbl>
## 1 N             1.85
## 2 Y             1.9
```

```
ggplot(data=d3) +
  geom_col(mapping = aes(x=Highway, y=mean_price))
```



Gas stations which is accessible to highways tend to set gas prices higher than gas stations which is far from highways.

```
d4 = GasPrices %>%
  group_by(Highway, Name) %>%
  summarize(mean_price = mean(Price))
```

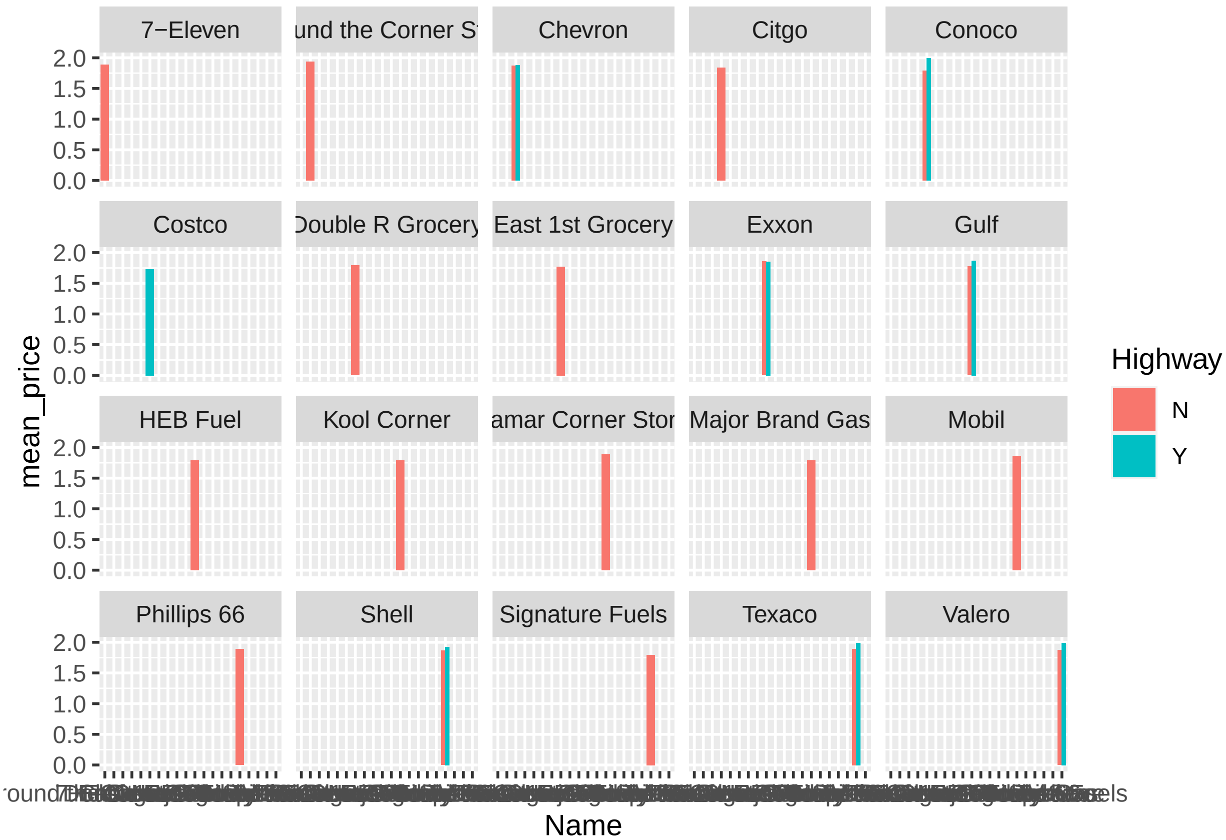
```
## `summarise()` regrouping output by 'Highway' (override with `.groups` argument)
```

```
d4
```

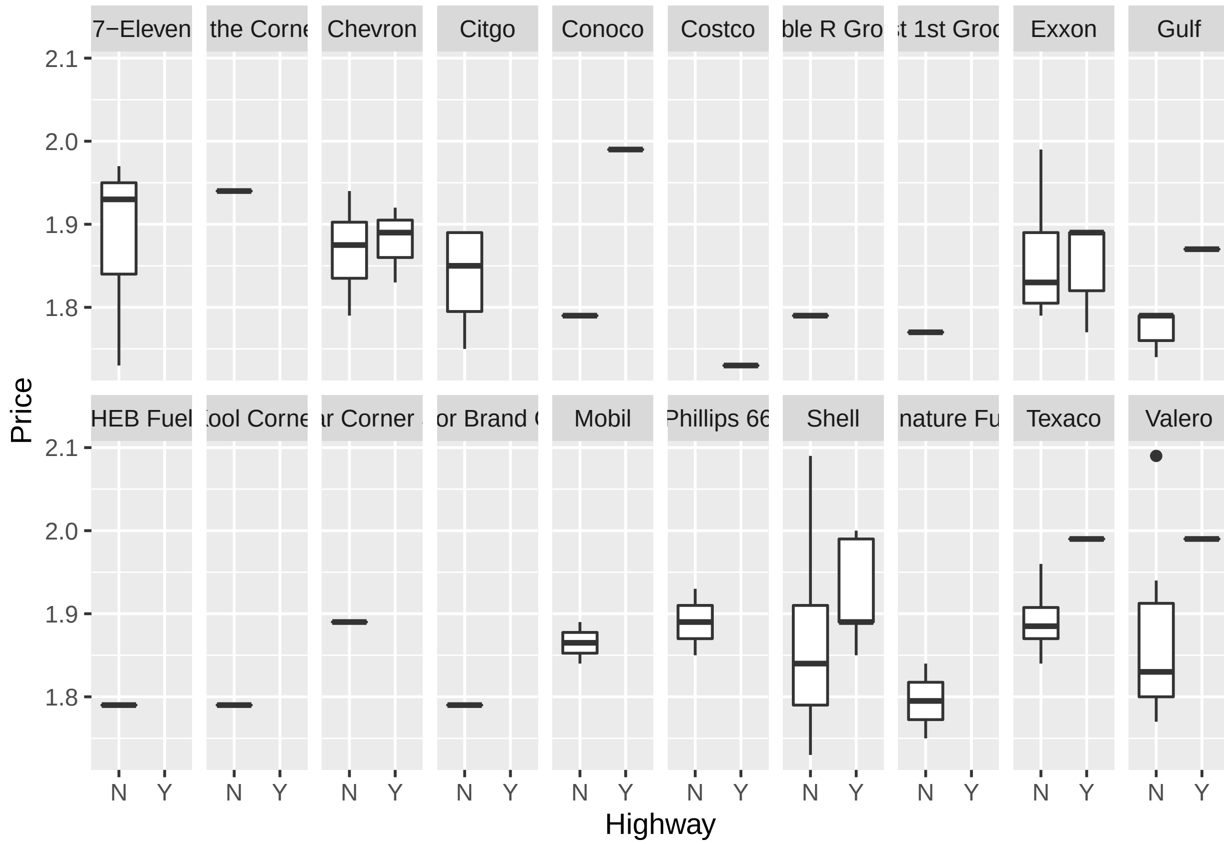
```
## # A tibble: 27 x 3
## # Groups:   Highway [2]
##    Highway Name                   mean_price
##    <chr>   <chr>                       <dbl>
##  1 N       7-Eleven                     1.89
##  2 N       Around the Corner Store      1.94
##  3 N       Chevron                      1.87
##  4 N       Citgo                        1.84
##  5 N       Conoco                       1.79
##  6 N       Double R Grocery             1.79
##  7 N       East 1st Grocery             1.77
##  8 N       Exxon                        1.86
##  9 N       Gulf                         1.78
## 10 N       HEB Fuel                     1.79
## # ... with 17 more rows
```

```
ggplot(data = d4) +
  geom_col(mapping = aes(x = Name,y = mean_price,
```

```
                         fill=Highway), position = 'dodge') +
  facet_wrap(~Name, nrow=4)
```




```
ggplot(data=GasPrices) +
  geom_boxplot(aes(x = Highway, y=Price)) +
  facet_wrap(~Name, nrow=2)
```



These plots show that companies usually set higher gas price for the gas stations which are highway accessible.

1-2

```r
library(tidyverse)
library(ggplot2)

bikeshare = read.csv('C:/Users/CHOI/Desktop/bikeshare.csv')
```

plot A

```r
bikeshare_a = bikeshare %>%
  group_by(hr) %>%
  summarise(average_rental=mean(total))
```

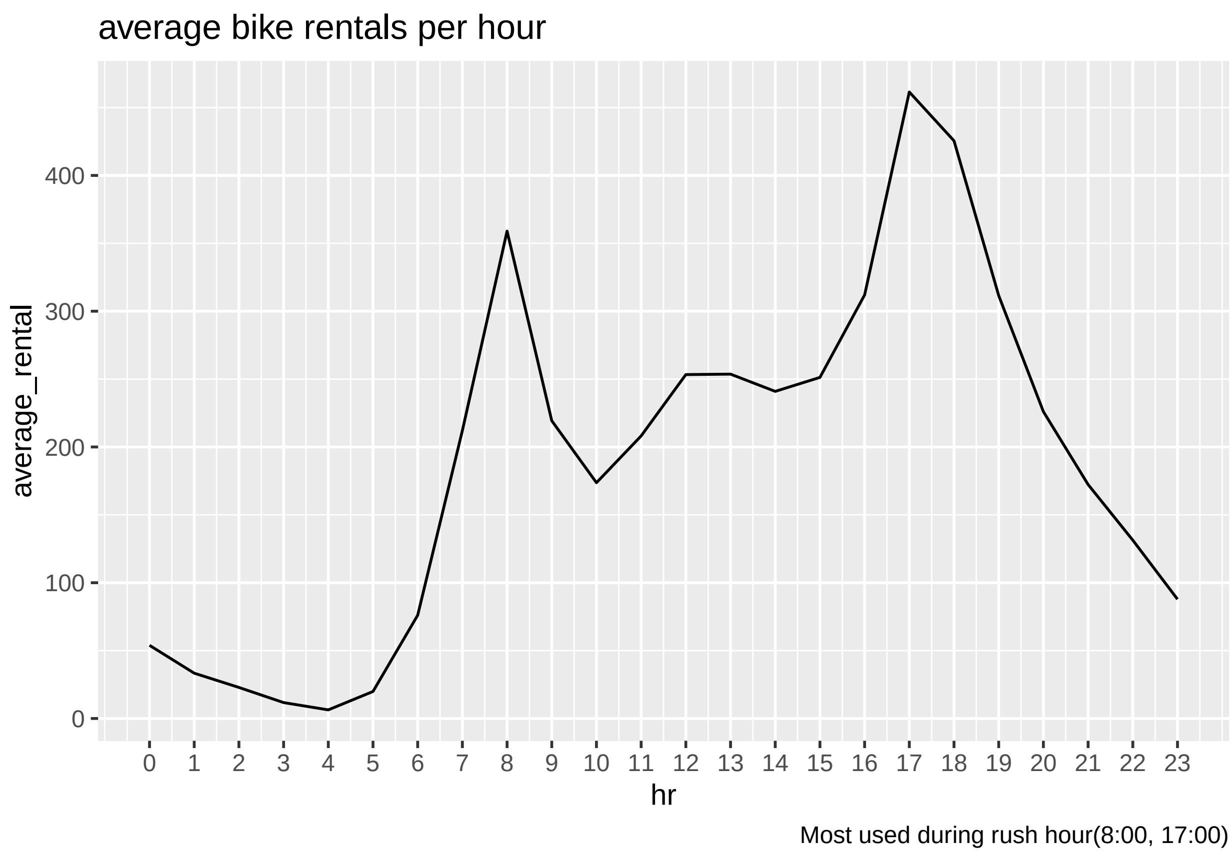
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
head(bikeshare_a)
```

```
## # A tibble: 6 x 2
##      hr average_rental
##   <int>          <dbl>
## 1     0           53.9
## 2     1           33.4
## 3     2           22.9
## 4     3           11.7
```

```
## 5      4           6.35
## 6      5          19.9
```

```
ggplot(data=bikeshare_a) +
  geom_line(aes(x=hr, y=average_rental)) +
  scale_x_continuous(breaks=0:23) +
  labs(title="average bike rentals per hour", caption="Most used during rush hour(8:00, 17:00)")
```



plot B

```
bikeshare_b = bikeshare %>%
  mutate(work = ifelse(workingday==1, "working", "dayoff")) %>%
  group_by(hr, work) %>%
  summarise(average_rental=mean(total))
```

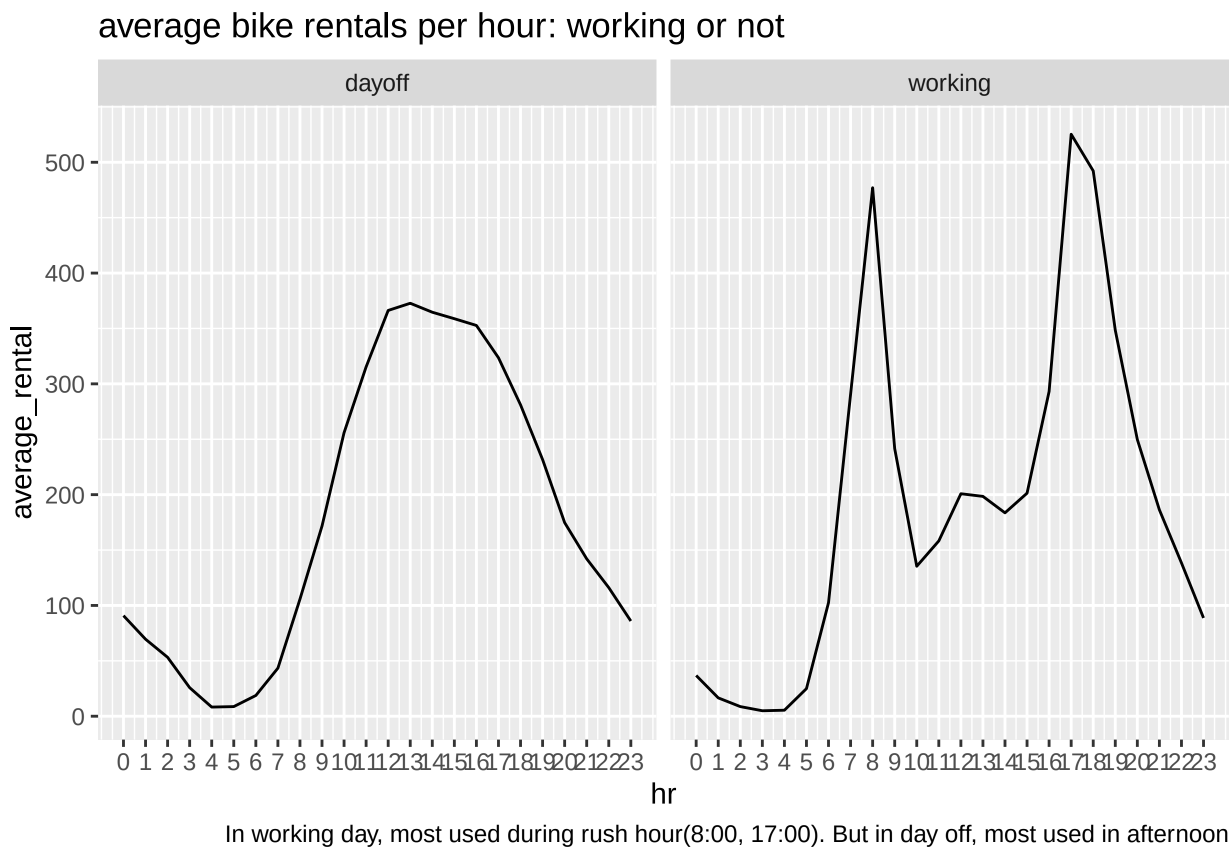
```
## `summarise()` regrouping output by 'hr' (override with `.groups` argument)
```

```
head(bikeshare_b)
```

```
## # A tibble: 6 x 3
## # Groups:   hr [3]
##      hr work    average_rental
##   <int> <chr>            <dbl>
## 1     0 dayoff           90.8
## 2     0 working          36.8
## 3     1 dayoff           69.5
## 4     1 working          16.6
```

```
## 5      2 dayoff            53.2
## 6      2 working            8.68
```

```
ggplot(data=bikeshare_b) +
  geom_line(aes(x=hr, y=average_rental)) + facet_wrap(~work) +
  scale_x_continuous(breaks=0:23) +
  labs(title="average bike rentals per hour: working or not",
       caption="In working day, most used during rush hour(8:00, 17:00). But in day off, most used in a
```

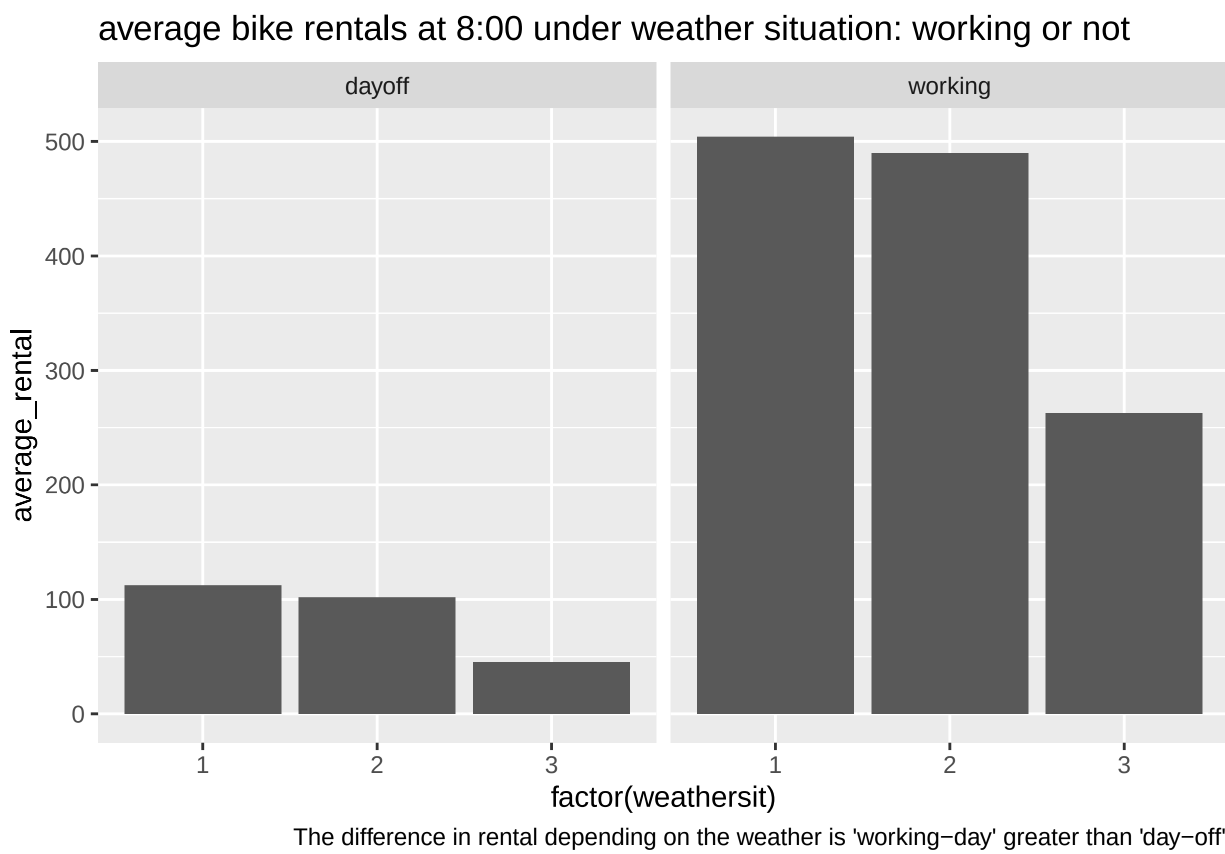


plot C

```
bikeshare_c = bikeshare %>%
  filter(hr==8) %>%
  mutate(work = ifelse(workingday==1, "working", "dayoff")) %>%
  group_by(weathersit, work) %>%
  summarise(average_rental=mean(total))
```

```
## `summarise()` regrouping output by 'weathersit' (override with `.groups` argument)
```

```
ggplot(data=bikeshare_c) +
  geom_col(aes(x=factor(weathersit), y=average_rental)) +
  facet_wrap(~work) +
  labs(title="average bike rentals at 8:00 under weather situation: working or not",
       caption="The difference in rental depending on the weather is 'working-day' greater than 'day-of
```



1-3

```r
library(tidyverse)
library(ggplot2)

ABIA = read.csv('C:/Users/CHOI/Desktop/ABIA.csv')

head(ABIA)
```

```
##   Year Month DayofMonth DayOfWeek DepTime CRSDepTime ArrTime CRSArrTime
## 1 2008     1          1         2     120       1935     309       2130
## 2 2008     1          1         2     555        600     826        835
## 3 2008     1          1         2     600        600     728        729
## 4 2008     1          1         2     601        605     727        750
## 5 2008     1          1         2     601        600     654        700
## 6 2008     1          1         2     636        645     934        932
##   UniqueCarrier FlightNum TailNum ActualElapsedTime CRSElapsedTime AirTime
## 1            9E      5746  84129E               109            115      88
## 2            AA      1614  N438AA               151            155     133
## 3            YV      2883  N922FJ               148            149     125
## 4            9E      5743  89189E                86            105      70
## 5            AA      1157  N4XAAA                53             60      38
## 6            NW      1674   N967N               178            167     145
##   ArrDelay DepDelay Origin Dest Distance TaxiIn TaxiOut Cancelled
## 1      339      345    MEM  AUS      559      3      18         0
```

```
## 2       -9       -5     AUS   ORD       978       7       11          0
## 3       -1        0     AUS   PHX       872       7       16          0
## 4      -23       -4     AUS   MEM       559       4       12          0
## 5       -6        1     AUS   DFW       190       5       10          0
## 6        2       -9     AUS   MSP      1042      11       22          0
##   CancellationCode Diverted CarrierDelay WeatherDelay NASDelay SecurityDelay
## 1                         0          339            0        0             0
## 2                         0           NA           NA       NA            NA
## 3                         0           NA           NA       NA            NA
## 4                         0           NA           NA       NA            NA
## 5                         0           NA           NA       NA            NA
## 6                         0           NA           NA       NA            NA
##   LateAircraftDelay
## 1                 0
## 2                NA
## 3                NA
## 4                NA
## 5                NA
## 6                NA
```

Which day of a week is the worst departure/arrival(long delay) in Austin?

Departure Delay

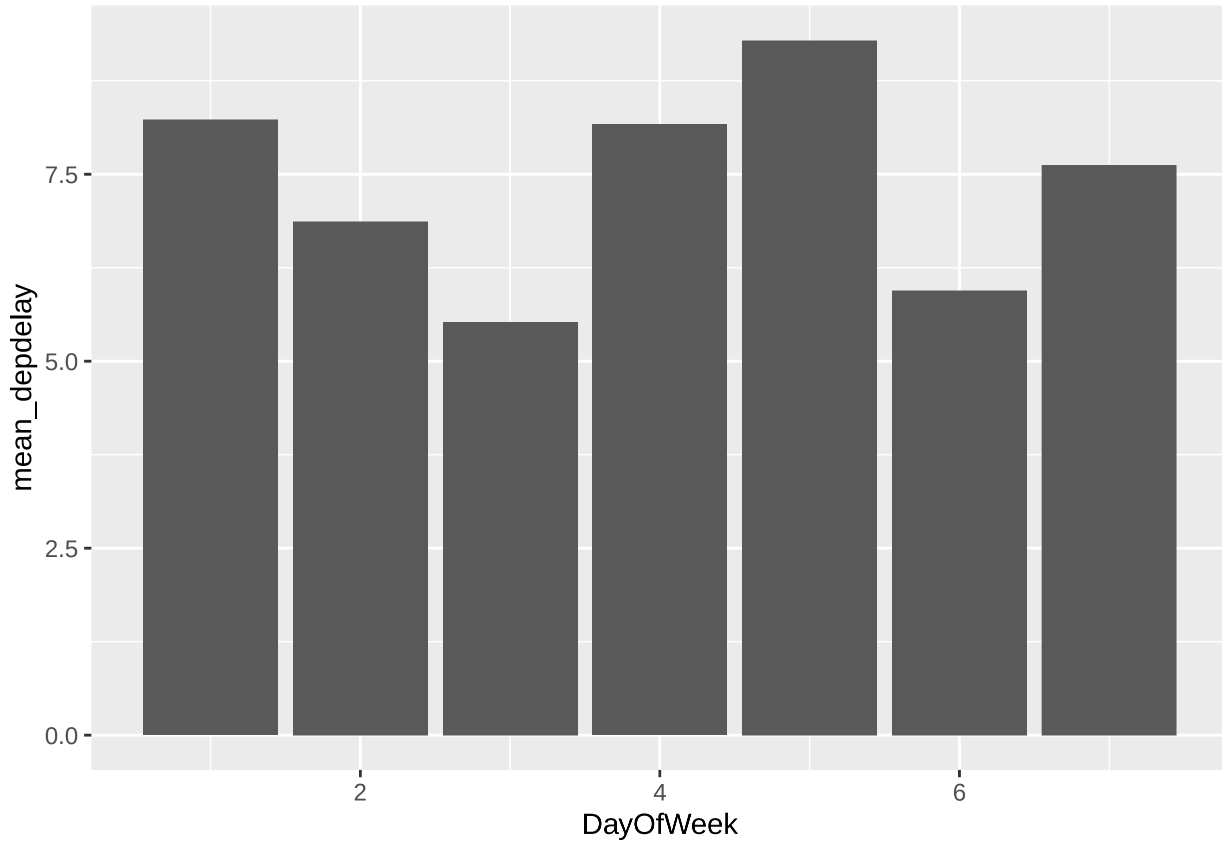
```
d1 = ABIA %>%
  filter(Origin == 'AUS') %>%
  filter(!is.na(DepDelay)) %>%
  group_by(DayOfWeek) %>%
  summarise(mean_depdelay = mean(DepDelay))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
d1
```

```
## # A tibble: 7 x 2
##   DayOfWeek mean_depdelay
##       <int>         <dbl>
## 1         1          8.23
## 2         2          6.87
## 3         3          5.53
## 4         4          8.17
## 5         5          9.29
## 6         6          5.94
## 7         7          7.62
```

```
ggplot(data=d1) +
  geom_col(aes(x=DayOfWeek, y=mean_depdelay), position = 'dodge')
```



If you leave from Austin by airplane, Wednesday is the best choice, which you can minimize your departure delay, the average departure delay is around 5 minutes, while Friday gives the longest delay.

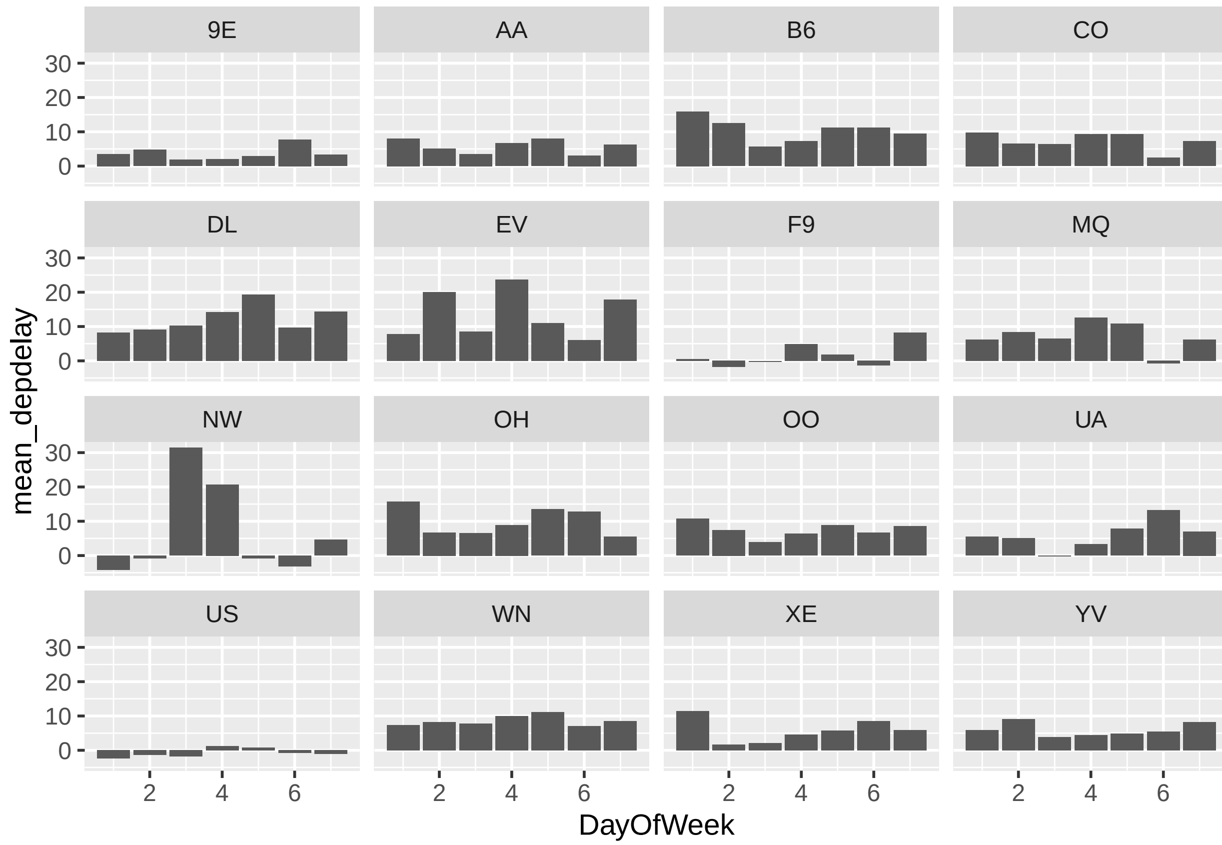
```
d2 = ABIA %>%
  filter(Origin == 'AUS') %>%
  filter(!is.na(DepDelay)) %>%
  group_by(DayOfWeek, UniqueCarrier) %>%
  summarise(mean_depdelay = mean(DepDelay))
```

```
## `summarise()` regrouping output by 'DayOfWeek' (override with `.groups` argument)
```

```
d2
```

```
## # A tibble: 112 x 3
## # Groups:   DayOfWeek [7]
##    DayOfWeek UniqueCarrier mean_depdelay
##        <int> <chr>                 <dbl>
##  1         1 9E                    3.55
##  2         1 AA                    8.04
##  3         1 B6                   15.9
##  4         1 CO                    9.78
##  5         1 DL                    8.23
##  6         1 EV                    7.78
##  7         1 F9                    0.487
##  8         1 MQ                    6.12
##  9         1 NW                   -4.17
## 10         1 OH                   15.7
## # ... with 102 more rows
```

```r
ggplot(data=d2) +
  geom_col(aes(x=DayOfWeek, y=mean_depdelay), position = 'dodge') +
  facet_wrap(~UniqueCarrier)
```



However, each airline has different delay pattern by day of week. So, if you plan airline trip, you might need to consider which day of week is best and worst for your airline.

Arrival Delay

```r
d3 = ABIA %>%
  filter(Dest=='AUS') %>%
  filter(!is.na(ArrDelay)) %>%
  group_by(DayOfWeek) %>%
  summarise(mean_ArrDelay = mean(ArrDelay))
```

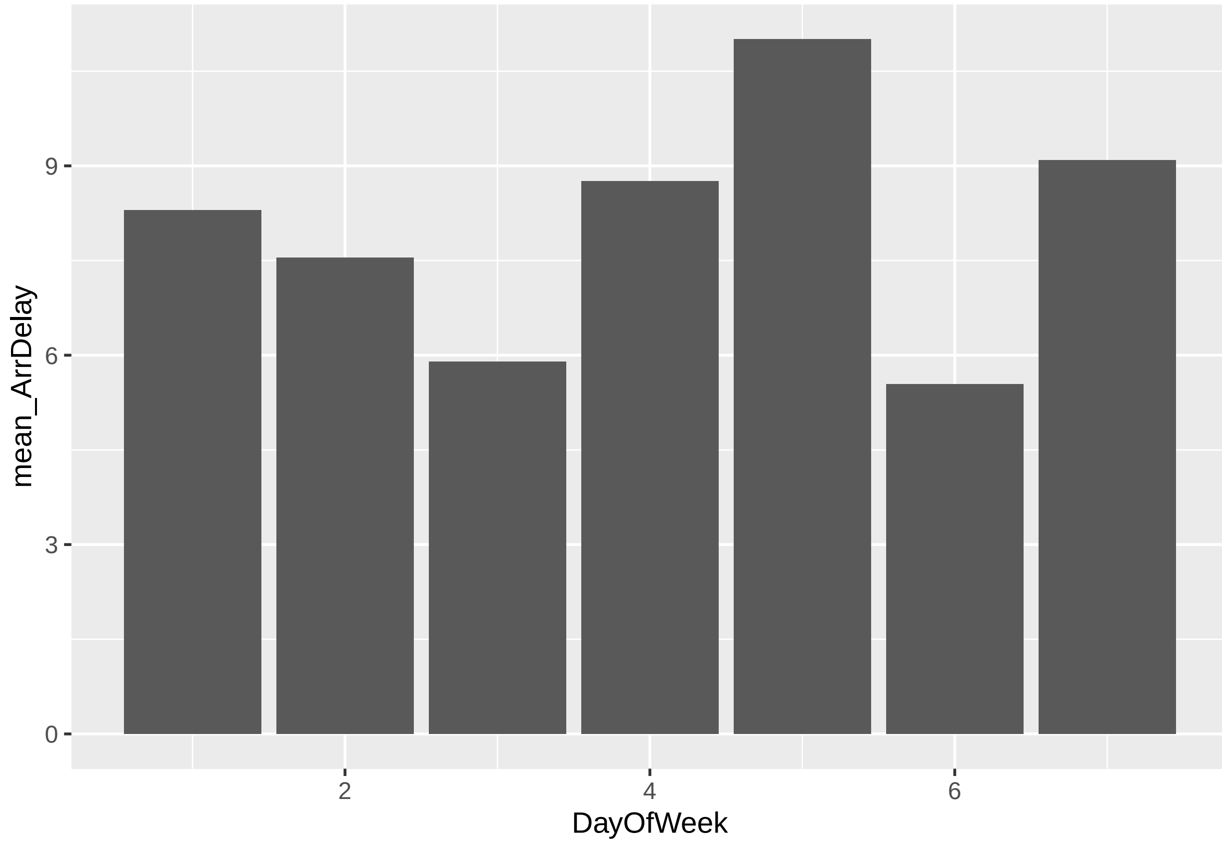
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
d3
```

```
## # A tibble: 7 x 2
##   DayOfWeek mean_ArrDelay
##       <int>         <dbl>
## 1         1          8.30
## 2         2          7.54
## 3         3          5.90
## 4         4          8.75
## 5         5         11.0
## 6         6          5.54
```

```
## 7         7          9.09
```

```
ggplot(data=d3) +
  geom_col(aes(x=DayOfWeek, y=mean_ArrDelay), position = 'dodge')
```



The arrival delay is also the longest on Friday like the departure delay in Austin.

```
d4 = ABIA %>%
  filter(Dest=='AUS') %>%
  filter(!is.na(ArrDelay)) %>%
  group_by(DayOfWeek, UniqueCarrier) %>%
  summarise(mean_ArrDelay = mean(ArrDelay))
```

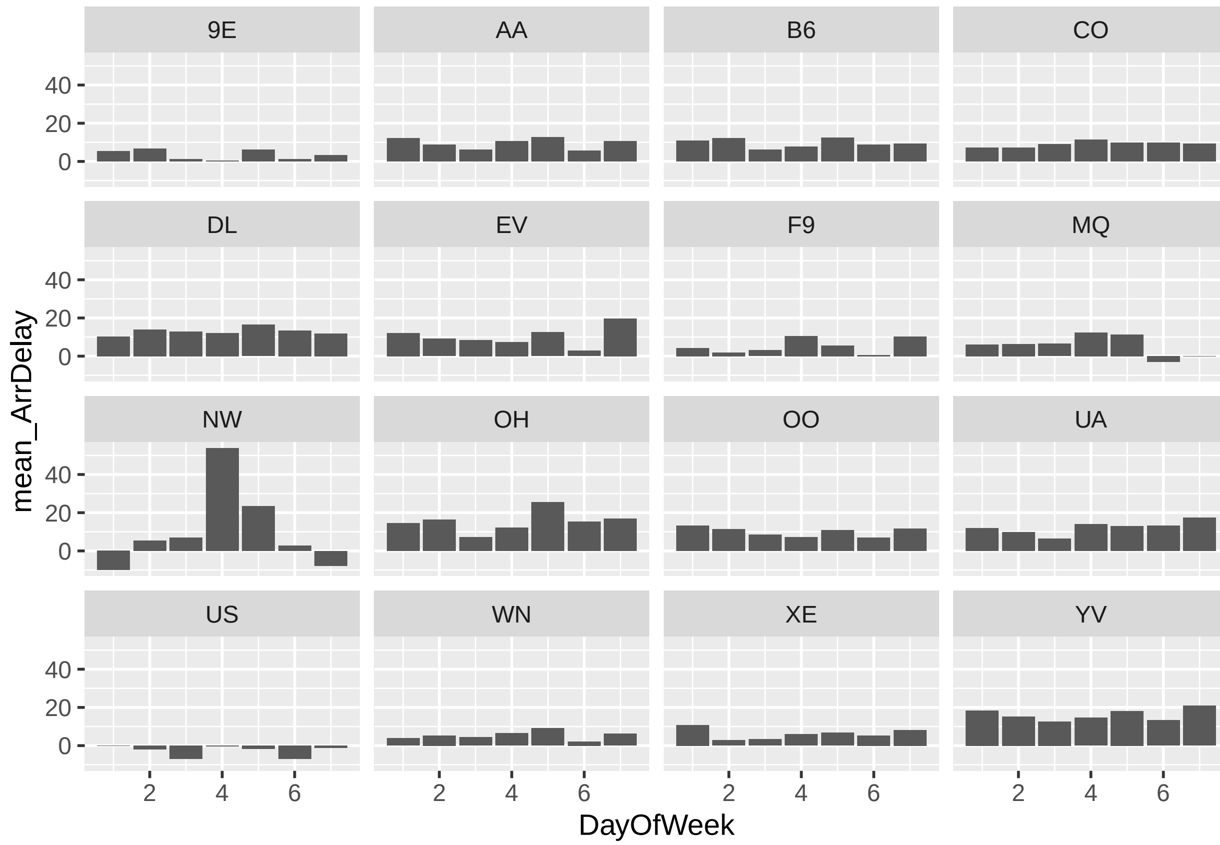
```
## `summarise()` regrouping output by 'DayOfWeek' (override with `.groups` argument)
```

```
d4
```

```
## # A tibble: 112 x 3
## # Groups:   DayOfWeek [7]
##    DayOfWeek UniqueCarrier mean_ArrDelay
##        <int> <chr>                 <dbl>
##  1         1 9E                     5.48
##  2         1 AA                    12.3
##  3         1 B6                    10.9
##  4         1 CO                     7.29
##  5         1 DL                    10.4
##  6         1 EV                    12.1
##  7         1 F9                     4.26
##  8         1 MQ                     6.09
```

```
## 9          1 NW                    -10
## 10         1 OH                     14.4
## # ... with 102 more rows
```

```r
ggplot(data=d4) +
  geom_col(aes(x=DayOfWeek, y=mean_ArrDelay), position = 'dodge') +
  facet_wrap(~UniqueCarrier)
```



Each airline has different shape of arrival delay by day of week. The interesting thing is NW airline shows high peak in departure and arrival delay in the middle of week, while US airline has very low, and stable delay.

1-4

```r
library(tidyverse)
library(ggplot2)
library(rsample)
library(caret)
```

```
## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(modelr)
library(parallel)
library(foreach)
```

```
##
## Attaching package: 'foreach'
## The following objects are masked from 'package:purrr':
##
##     accumulate, when
```

```r
sclass = read.csv('C:/Users/CHOI/Desktop/sclass.csv')
```

To separate data set and make training/testing set

```r
sclass %>%
  filter(trim=="350" | trim=="65 AMG")  %>%
  select(trim, mileage, price)
```

```
##       trim mileage  price
## 1      350   21929  55994
## 2      350   17770  60900
## 3      350   29108  54995
## 4      350   35004  59988
## 5      350   66689  37995
## 6      350   19567  59977
## 7      350   10616  69900
## 8      350    2578  68960
## 9      350   23677  61001
## 10     350   28384  58992
## 11     350   21388  69900
## 12  65 AMG     106 235375
## 13  65 AMG      11 226465
## 14     350   87100   9995
## 15  65 AMG   74461  24995
## 16     350   26183  49990
## 17     350   32800  53999
## 18     350   55683  62997
## 19     350   29044  61900
## 20     350   61676  35900
## 21     350  117683  12900
## 22  65 AMG   73415  54981
## 23  65 AMG   17335 102500
## 24     350   29468  40999
## 25  65 AMG       7 230860
## 26     350   35642  21995
## 27  65 AMG   48398  35888
## 28  65 AMG   61500  45981
## 29  65 AMG   49515  49982
## 30  65 AMG   70692  43990
## 31  65 AMG       5 216510
## 32     350    7342  53900
## 33     350   22751  56991
## 34     350    2384  75900
```

```
## 35        350   21874  58975
## 36        350    5404  81895
## 37        350   12414  64900
## 38        350   15435  68950
## 39        350   41075  53981
## 40        350   11862  76878
## 41        350   31300  53000
## 42    65 AMG       50 226115
## 43    65 AMG       89 221750
## 44        350   68221  16980
## 45        350   52003  17998
## 46        350  104426  10995
## 47    65 AMG   69652  42982
## 48    65 AMG   79795  41995
## 49    65 AMG   55730  78992
## 50        350   11076  59900
## 51        350   21185  51495
## 52        350   32290  48789
## 53        350   38310  47994
## 54        350   40755  46995
## 55    65 AMG        7 244325
## 56    65 AMG       43 224625
## 57    65 AMG   31048  59888
## 58    65 AMG   11632 110995
## 59    65 AMG   31321  79888
## 60        350   31782  52999
## 61        350      14  74900
## 62    65 AMG       11 235365
## 63        350   62028  15991
## 64    65 AMG   45200  85000
## 65    65 AMG   85142  37900
## 66    65 AMG   48579  77444
## 67        350   33720  42999
## 68    65 AMG       17 225681
## 69    65 AMG       10 227715
## 70    65 AMG       12 227685
## 71    65 AMG       10 236125
## 72        350   76146  14950
## 73    65 AMG   52800  40800
## 74    65 AMG   76093  49950
## 75    65 AMG   52951  64999
## 76    65 AMG   49436  86887
## 77        350   18748  59995
## 78        350    9300 103410
## 79        350   19266  62995
## 80    65 AMG       18 240825
## 81        350   80511  15991
## 82    65 AMG   51670  59995
## 83    65 AMG   49735  61900
## 84    65 AMG   52045  61491
## 85    65 AMG   28626  89888
## 86        350   10385  71895
## 87        350    7000  82000
## 88        350   29996  61995
```

```
## 89       350    3524  70900
## 90       350   10721  62988
## 91  65 AMG        3 224625
## 92       350   51026  14991
## 93  65 AMG    15512 114998
## 94       350   25685  61900
## 95       350   38239  51312
## 96       350   20868  44999
## 97       350   48230  51777
## 98       350   38503  52900
## 99  65 AMG       20 231325
## 100 65 AMG       10 226115
## 101 65 AMG        8 224765
## 102 65 AMG       16 228325
## 103      350   65757  14993
## 104 65 AMG    86472  22994
## 105      350   32047  50500
## 106      350   24501  66900
## 107      350   29648  69994
## 108      350   59439  39995
## 109      350    1514  94230
## 110 65 AMG       11 225975
## 111 65 AMG       13 235475
## 112      350   77159  16995
## 113 65 AMG    62114  47995
## 114 65 AMG    61560  47888
## 115 65 AMG    20578 107965
## 116      350   31158  62995
## 117 65 AMG     4685 199900
## 118      350  102919  13895
## 119      350   52670  19495
## 120      350   93400  12995
## 121 65 AMG    50942  49900
## 122 65 AMG    22662 114889
## 123 65 AMG     9982 129995
## 124      350   36851  53993
## 125      350   22634  62995
## 126      350   18516  51950
## 127      350   41913  49888
## 128      350   56981  45777
## 129      350   22893  65995
## 130      350    6538  74895
## 131      350    5246  63991
## 132      350    2747  67995
## 133 65 AMG        6 228325
## 134 65 AMG        4 226710
## 135      350   55046  16851
## 136      350   95011  13998
## 137 65 AMG    24482  44992
## 138 65 AMG    44075  50037
## 139 65 AMG     2563  98888
## 140      350   81600  34995
## 141      350   12018  69500
## 142      350   21596  59991
```

```
## 143     350   21348  63862
## 144     350   79707  15500
## 145     350   48481  20879
## 146     350  123524  13498
## 147 65 AMG   34586  49991
## 148 65 AMG   38694  86900
## 149 65 AMG   27704  99900
## 150     350   35014  51970
## 151     350   50884  47995
## 152     350   26703  49995
## 153     350   22180  56900
## 154 65 AMG       7 227865
## 155 65 AMG       8 226965
## 156     350   91278  11998
## 157 65 AMG   79713  45888
## 158 65 AMG   40527  62498
## 159     350   19259  59881
## 160     350   15129  62900
## 161     350   44492  41977
## 162     350  120997  13900
## 163 65 AMG   59391  29991
## 164 65 AMG   76271  26900
## 165 65 AMG      10 230825
## 166 65 AMG      27 227360
## 167 65 AMG   77288  44995
## 168     350    2902  68991
## 169     350   27806  46298
## 170     350   34577  53002
## 171     350      39  81900
## 172 65 AMG   45869  18990
## 173 65 AMG   67604  44777
## 174 65 AMG   23174  99990
## 175     350    7116  69998
## 176     350   31810  51295
## 177     350   33917  49888
## 178     350   18046  51990
## 179     350  137076   6600
## 180     350   84402  15999
## 181 65 AMG   29744  54995
## 182 65 AMG   32115  50999
## 183 65 AMG   11400  77900
## 184 65 AMG   68444  49999
## 185 65 AMG      30 231565
## 186     350   24094  49990
## 187     350   27797  53900
## 188     350   49329  53999
## 189     350  149895   7999
## 190     350   23097  19899
## 191 65 AMG   46319  29899
## 192 65 AMG   18741  42888
## 193 65 AMG   78695  26995
## 194 65 AMG   20000  79900
## 195 65 AMG   18667  95000
## 196     350   15772  66995
```

```
## 197     350   40460  51999
## 198     350   29946  54982
## 199     350   39035  46900
## 200     350   44822  62739
## 201     350  160000   8500
## 202     350   45433  15994
## 203     350   81707  14900
## 204 65 AMG  118986  25887
## 205 65 AMG   61154  44495
## 206     350   37470  44877
## 207     350   41754  42900
## 208     350   19571  51995
## 209     350   46352  38999
## 210     350   19086  55900
## 211     350   25599  61971
## 212 65 AMG       7 224625
## 213     350   61686  19988
## 214     350   36483  19995
## 215 65 AMG   41054  64990
## 216 65 AMG   43967  61991
## 217 65 AMG   24209 117993
## 218     350       8  79987
## 219     350   23258  59941
## 220     350   19305  54995
## 221     350   34548  49994
## 222     350   64785  14992
## 223 65 AMG  119910  19999
## 224 65 AMG   69507  28882
## 225 65 AMG   42443  85981
## 226     350   28149  59850
## 227     350   28717  49900
## 228     350   35848  55900
## 229     350    2930  68991
## 230     350   40231  68995
## 231 65 AMG      15 226925
## 232     350  102292  13350
## 233 65 AMG   79710  22999
## 234 65 AMG   12737  79500
## 235 65 AMG   56337  52850
## 236 65 AMG   22123  87900
## 237     350   94900  30995
## 238     350     100  94230
## 239     350   36095  66900
## 240     350    7308  62900
## 241 65 AMG      10 232035
## 242     350   85141  15750
## 243 65 AMG      28 156992
## 244     350   22597  54536
## 245     350   21191  51500
## 246 65 AMG      17 234815
## 247     350   63347  15999
## 248 65 AMG   73240  28900
## 249 65 AMG  137951  35850
## 250 65 AMG   75973  44950
```

```
## 251     350   32482  58991
## 252 65 AMG       3 224625
## 253 65 AMG   74676  54450
## 254     350   25159  55993
## 255     350   37728  55600
## 256     350   32952  50985
## 257     350   24638  57991
## 258     350   13280  70770
## 259     350   21101  54556
## 260 65 AMG      12 232475
## 261     350   98665  16988
## 262     350  104175  11999
## 263     350  101985  12995
## 264 65 AMG   87945  42888
## 265 65 AMG   93210  35000
## 266     350    5465  69991
## 267 65 AMG      16 226575
## 268 65 AMG       8 237935
## 269     350   76426  14532
## 270     350   90994  13488
## 271 65 AMG   65992  28995
## 272 65 AMG   28614 187645
## 273 65 AMG   42813  56888
## 274 65 AMG   69966  43998
## 275     350   22503  53995
## 276 65 AMG      10 229315
## 277     350   75335   9995
## 278     350   66120  19900
## 279 65 AMG    7780  91989
## 280 65 AMG    7267 159950
## 281 65 AMG   19049 105888
## 282     350   27813  43995
## 283     350   25464  64900
## 284     350   28774  62900
## 285     350   28605  49955
## 286 65 AMG      12 228465
## 287     350  108345  10977
## 288     350   73002  15995
## 289 65 AMG   59567  28871
## 290 65 AMG   59471  47690
## 291 65 AMG   78303  45800
## 292 65 AMG   16502  99999
## 293 65 AMG   16536 115000
## 294     350      18  94230
## 295     350   90009  39999
## 296     350   29377  58990
## 297     350   39256  48995
## 298     350   35851  55900
## 299     350   72720  43985
## 300 65 AMG      30 227675
## 301     350   84263  15770
## 302     350   80645  14995
## 303     350   71963  15977
## 304 65 AMG   69302  46917
```

```
## 305 65 AMG   60261  53995
## 306     350     57 102810
## 307     350   39870  42995
## 308     350   18906  48991
## 309     350  117119  11980
## 310     350   83649  12989
## 311 65 AMG   35092  64888
## 312     350   25402  56992
## 313     350   71639  41873
## 314     350   42315  49970
## 315     350    2325  74899
## 316     350   27214  59995
## 317 65 AMG       5 225425
## 318 65 AMG       6 224975
## 319 65 AMG   64734  25247
## 320 65 AMG   82436  26995
## 321 65 AMG  103682  38930
## 322 65 AMG   17570  59994
## 323 65 AMG   52340  86991
## 324 65 AMG      17 228410
## 325 65 AMG   39721  97699
## 326     350   40639  44995
## 327     350   44343  49995
## 328     350   29755  53999
## 329     350   18738  59763
## 330     350   18540  64900
## 331 65 AMG      23 226375
## 332 65 AMG      10 230985
## 333     350   43894  16988
## 334 65 AMG   47559  69995
## 335 65 AMG   35705  81999
## 336 65 AMG  146975  32995
## 337     350   29715  55990
## 338     350  117907  12412
## 339     350   58662  19795
## 340 65 AMG   46738  61995
## 341     350   40682  51993
## 342     350   14670  56900
## 343     350   46213  55992
## 344     350   17088  66766
## 345     350   25448  59999
## 346     350    3559  69895
## 347     350   16928  55881
## 348     350   91536  13700
## 349     350   91705   8990
## 350     350   83310  14951
## 351 65 AMG   75569  44995
## 352 65 AMG   38201  55888
## 353 65 AMG   56828  79990
## 354     350   10055  61995
## 355     350   36523  56600
## 356     350   17584  61995
## 357     350   12403  49995
## 358     350       6 102460
```

```
## 359     350   50085  17511
## 360     350  110511  10995
## 361     350   78656  14799
## 362 65 AMG   30892  69949
## 363 65 AMG   27310  58980
## 364     350   20325  63093
## 365     350   34008  57995
## 366     350   54489  11875
## 367     350   85214  10800
## 368 65 AMG  103007  21950
## 369 65 AMG   74800  68900
## 370 65 AMG   26561  83950
## 371 65 AMG   11539  99995
## 372     350   10447  58998
## 373     350   21130  59734
## 374     350   27300  53991
## 375     350   49560  43987
## 376     350   57773  40999
## 377     350   40548  47999
## 378     350   10488  70895
## 379     350  119774  12988
## 380 65 AMG   80374  40890
## 381 65 AMG   65751  40500
## 382     350   31825  48955
## 383     350   28749  56993
## 384 65 AMG       5 228825
## 385 65 AMG   29712  42990
## 386 65 AMG   69415  39950
## 387     350   23775  49900
## 388     350   56045  54991
## 389     350     791  73995
## 390     350   20055  64900
## 391     350     851  94230
## 392 65 AMG       5 225420
## 393     350  106705  13995
## 394     350  102440  10980
## 395 65 AMG   91102  26980
## 396 65 AMG   50761  63895
## 397     350   34271  58991
## 398     350    4308  82910
## 399     350   14515  59988
## 400     350    9099  97184
## 401     350   27604  56100
## 402 65 AMG       5 230085
## 403 65 AMG      17 226465
## 404 65 AMG      17 226465
## 405     350  135020   9998
## 406     350   73429  15995
## 407 65 AMG   42107  32455
## 408 65 AMG   17392  64991
## 409 65 AMG   44920  42995
## 410 65 AMG    9343 114989
## 411 65 AMG   57242  79888
## 412     350   20498  43999
```

```
## 413     350    4978  69990
## 414     350   45750  46995
## 415     350    2814  70995
## 416     350    2951  69995
## 417     350   12201  76751
## 418     350   22199  49500
## 419     350   34387  55991
## 420 65 AMG       27 227375
## 421     350   72426  14791
## 422     350   83604  17995
## 423     350  139500  13995
## 424 65 AMG   69724  24950
## 425 65 AMG   31156  48950
## 426     350    3541  74988
## 427     350   26733  54990
## 428     350   40863  53981
## 429     350   17698  54900
## 430     350  116193  39000
## 431 65 AMG       16 234910
## 432 65 AMG       14 229870
## 433 65 AMG        8 226135
## 434 65 AMG        7 226620
## 435 65 AMG   79212  29991
## 436 65 AMG   33515  77889
## 437 65 AMG   68369  43888
## 438     350   51809  47999
## 439     350   32493  62900
## 440     350   15873  48995
## 441     350   18575  48995
## 442 65 AMG        7 224625
## 443     350  122279   9995
## 444 65 AMG   79370  25000
## 445 65 AMG   46417  66991
## 446 65 AMG   44245  71995
## 447     350    2086  74991
## 448     350   59552  51950
## 449     350   37406  55911
## 450     350   80276  14000
## 451     350   81757  14888
## 452 65 AMG   30223  80000
## 453 65 AMG   12280 153900
## 454     350   31091  70981
## 455     350   28259  49495
## 456 65 AMG        1 226235
## 457 65 AMG       18 228935
## 458 65 AMG        9 226625
## 459     350   35681  17995
## 460     350   26634  19120
## 461 65 AMG   46890  36900
## 462 65 AMG   54540  41892
## 463     350   44685  49991
## 464     350   19938  62988
## 465     350   20414  58990
## 466     350       10 106010
```

```
## 467 65 AMG      10 228875
## 468 65 AMG      11 225525
## 469     350  65869  13750
## 470     350  93603  12888
## 471     350  77861  14991
## 472     350  95106  10995
## 473 65 AMG  26985  94995
## 474 65 AMG  45856  99000
## 475 65 AMG  68026  79999
## 476     350  29239  55950
## 477 65 AMG  58243  57800
## 478 65 AMG  59253  54983
## 479 65 AMG  56788  84887
## 480 65 AMG  65654  72999
## 481     350  20663  53250
## 482     350   6655  59991
## 483 65 AMG       5 230685
## 484 65 AMG      10 225915
## 485     350  77841  15989
## 486     350  62500  16991
## 487     350  74439  17500
## 488 65 AMG  63900  31500
## 489 65 AMG  39488  36995
## 490 65 AMG  75341  29981
## 491 65 AMG  20834  96777
## 492     350  20033  47000
## 493     350  16877  49995
## 494     350   5864  69991
## 495 65 AMG  86281  45999
## 496 65 AMG  22739  62900
## 497     350   4444  65900
## 498     350  11682  64991
## 499     350  25730  65991
## 500 65 AMG      31 226135
## 501     350 153001  10980
## 502 65 AMG  19176 175999
## 503 65 AMG  19316  99740
## 504 65 AMG  36663  73980
## 505 65 AMG   5562 159850
## 506     350  55372  43188
## 507     350  28685  59982
## 508     350    111  71995
## 509     350  15841  67991
## 510     350  67229  39995
## 511     350  16038  55112
## 512     350  66821  44995
## 513 65 AMG       6 233625
## 514 65 AMG       7 226485
## 515     350 102843  13995
## 516     350  74801  15991
## 517 65 AMG  34836  59900
## 518 65 AMG  46732  53999
## 519     350  41268  47855
## 520     350   2714  65995
```

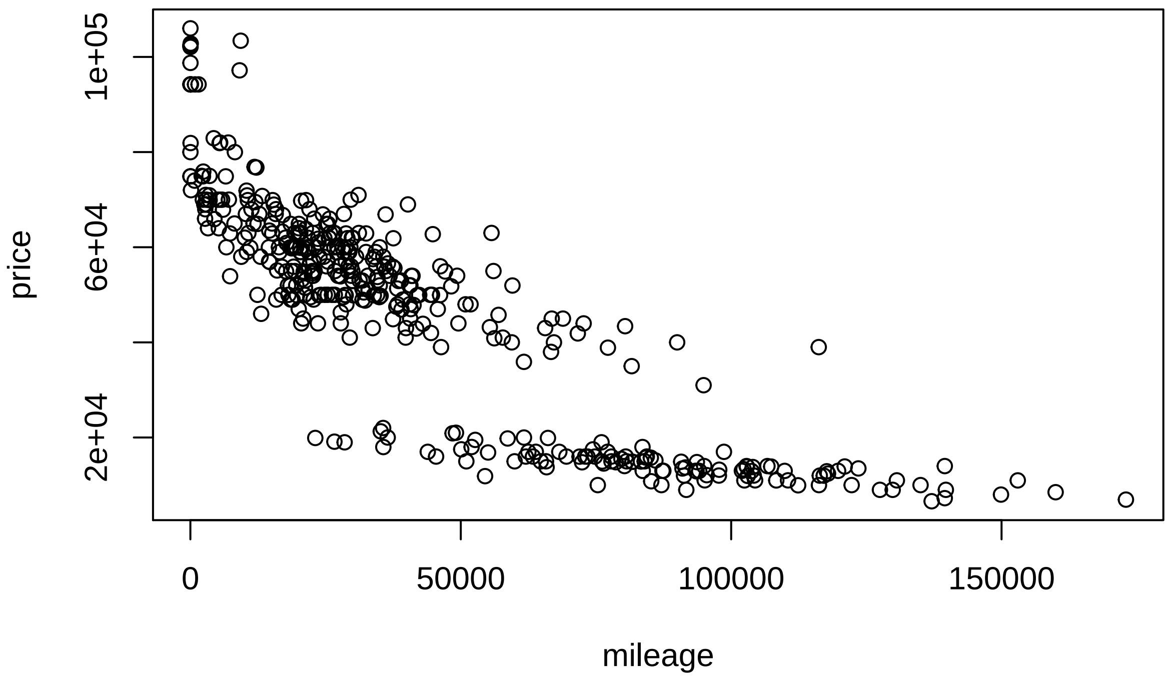
```
## 521     350      10  94230
## 522     350   23871  57950
## 523 65 AMG   21998  37895
## 524 65 AMG   56458  61994
## 525 65 AMG    9519 115770
## 526     350      25  94230
## 527     350   22972  54990
## 528     350   26652  62995
## 529     350   68891  44990
## 530     350  103060  11922
## 531 65 AMG   36052  49981
## 532     350  107403  13888
## 533     350   94087  12988
## 534 65 AMG   36123  97995
## 535 65 AMG   68375  37994
## 536     350   30639  58001
## 537 65 AMG      12 225910
## 538     350  127512   8999
## 539 65 AMG   61718  49890
## 540 65 AMG   51854  68000
## 541     350   33761  57498
## 542     350   36168  54991
## 543     350   27199  59991
## 544     350  104000  13790
## 545     350   93654  14826
## 546 65 AMG   22373  65881
## 547 65 AMG   23560  60900
## 548 65 AMG   26048  99500
## 549     350   26620  59988
## 550     350   16474  58985
## 551     350   47107  54900
## 552 65 AMG      13 226480
## 553     350  112381   9950
## 554     350  109890  12975
## 555 65 AMG   35927  65991
## 556 65 AMG   25748  54981
## 557     350   65602  43000
## 558     350   15089  64988
## 559     350   17003  63271
## 560     350   23614  61768
## 561 65 AMG      12 226570
## 562     350   84000  14995
## 563 65 AMG   37514  60882
## 564 65 AMG   55403  54991
## 565 65 AMG   12175 139900
## 566     350   22690  53994
## 567 65 AMG      44 239115
## 568     350  139497   7250
## 569 65 AMG   64497  55978
## 570 65 AMG   21195  95999
## 571     350   11297  67962
## 572     350      17 102110
## 573     350   26485  62991
## 574     350    8154  64991
```

```
## 575     350   20952  54883
## 576     350    3257  63995
## 577     350   25825  62960
## 578 65 AMG       11 235170
## 579 65 AMG       20 230625
## 580 65 AMG       10 229810
## 581 65 AMG        9 226465
## 582 65 AMG        8 224351
## 583     350  139700   8999
## 584     350   69536  15995
## 585 65 AMG   82636  24000
## 586     350   12741  66990
## 587     350    5508  81991
## 588     350  129867   8995
## 589     350   76025  18995
## 590     350   28516  18987
## 591 65 AMG   66259  44980
## 592     350      13  74900
## 593     350   22779  48995
## 594     350   22020  67988
## 595     350   15211  69900
## 596     350   24788  61873
## 597     350   18637  54995
## 598     350   19710  49900
## 599 65 AMG   50142  49995
## 600 65 AMG   51000  53995
## 601 65 AMG   17488  85950
## 602     350   43002  43900
## 603     350   12974  57999
## 604     350   16373  59995
## 605 65 AMG        1 226615
## 606     350  103788  12950
## 607     350  130642  10985
## 608 65 AMG   10420  41988
## 609 65 AMG   73900  26000
## 610 65 AMG   88015  59950
## 611 65 AMG   84736  57777
## 612 65 AMG    4324 119000
## 613     350   38087  47500
## 614     350    6029  67900
## 615     350   37533  61883
## 616     350   35670  57995
## 617     350   24845  49995
## 618     350   39806  40999
## 619 65 AMG       10 226465
## 620 65 AMG     2097 226662
## 621 65 AMG       22 225975
## 622 65 AMG        5 224625
## 623 65 AMG       36 234215
## 624     350   95487  12085
## 625     350   85990  15178
## 626 65 AMG   29722  42808
## 627 65 AMG   33940  64000
## 628 65 AMG   85858  57500
```

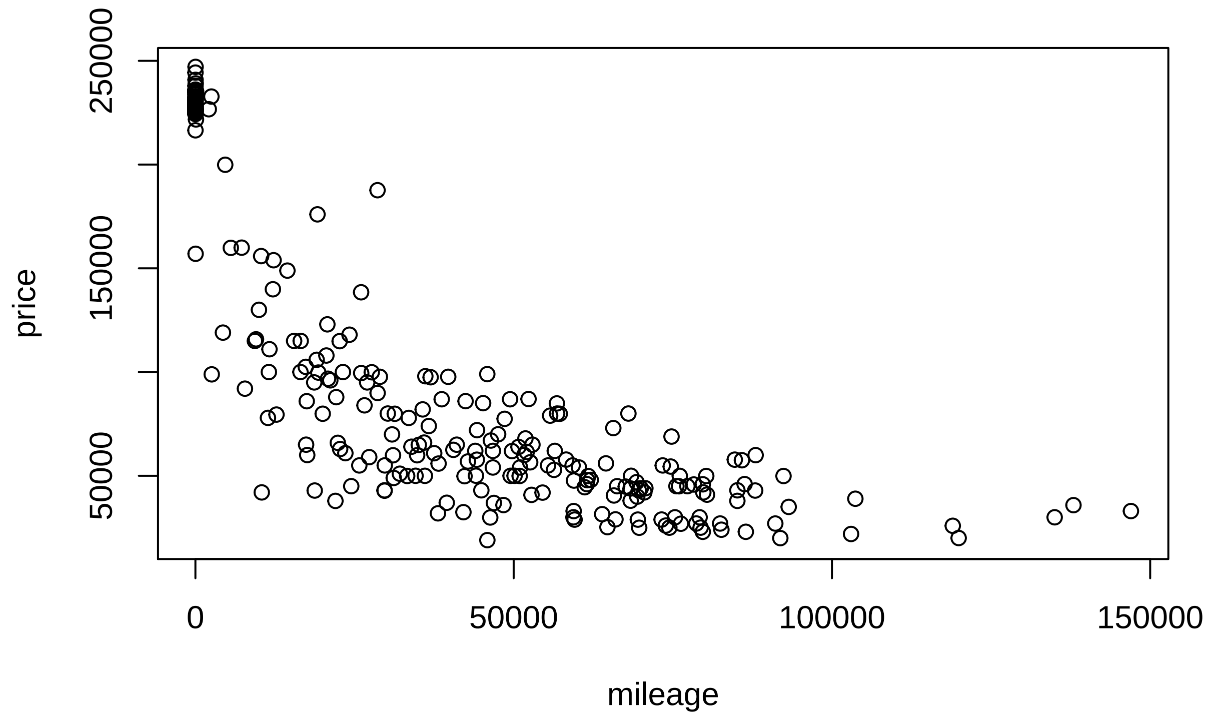
```
## 629 65 AMG       6 224625
## 630 65 AMG      20 234465
## 631     350 173000   6942
## 632 65 AMG  134997  29995
## 633 65 AMG   80256  49888
## 634     350   9405  57995
## 635     350  30000  52900
## 636     350  35176  49781
## 637     350  38956  52995
## 638 65 AMG      24 247075
## 639     350 116385  11990
## 640 65 AMG   52596  56490
## 641     350  25140  64892
## 642     350  14588  63500
## 643     350  18126  49994
## 644     350  27079  58881
## 645     350  21730  61990
## 646     350  25392  49999
## 647     350 116233   9995
## 648 65 AMG   38106  31900
## 649 65 AMG   33267  49900
## 650 65 AMG   42278  49800
## 651 65 AMG   28980  97700
## 652     350  20291  59888
## 653     350  30091  49900
## 654     350  34928  49510
## 655     350  29846  61995
## 656     350  22554  54995
## 657     350  10270  66991
## 658 65 AMG    2509 232775
## 659 65 AMG   85149  42999
## 660 65 AMG   70527  42000
## 661 65 AMG   36953  97493
## 662     350  13086  45999
## 663     350  21122  49999
## 664     350  21296  52998
## 665     350     14  98745
## 666 65 AMG      25 235980
## 667 65 AMG      16 236020
## 668 65 AMG      10 226485
## 669 65 AMG   59426  32995
## 670 65 AMG   44210  57750
## 671 65 AMG   92393  49888
## 672 65 AMG   20731 122981
## 673     350  20327  58991
## 674 65 AMG      17 234875
## 675     350  63874  16990
## 676 65 AMG   91893  19950
## 677 65 AMG   10329 155888
## 678 65 AMG   26035 138450
## 679     350  34588  53750
## 680 65 AMG      17 232675
## 681     350  97789  13226
## 682     350  90756  14900
```

```
## 683    350   97724  11995
## 684 65 AMG   14453 148888
## 685    350   46171  49950
## 686    350   56201  40860
## 687    350   77207  38884
## 688    350   18427  59955
## 689    350   23584  43999
## 690    350   28842  47995
## 691    350    2309  69995
## 692 65 AMG      25 229135
## 693 65 AMG      11 226465
## 694    350   49112  20995
## 695    350   59950  14990
## 696    350    8227  79990
## 697    350   20485  69759
## 698    350   80386  43400
## 699    350   14554  56995
## 700    350   28391  66990
## 701    350   35200  21290
## 702    350   87291  12994
## 703    350   87458  12995
## 704    350   20056  64000
## 705    350   27730  53900
## 706    350   29143  58990
## 707    350   29583  59995
## 708 65 AMG      10 226465
```

```
s3 = subset(sclass, trim=="350")
s6 = subset(sclass, trim=="65 AMG")

plot(price ~ mileage, data = s3)
```




```r
plot(price ~ mileage, data = s6)
```




```
s3_split = initial_split(s3, prop=0.8)
s3_train = training(s3_split)
s3_test = testing(s3_split)

s6_split = initial_split(s6, prop=0.8)
s6_train = training(s6_split)
s6_test = testing(s6_split)
```

350 trim of sclass

k=2,5,10,15,20,25,50,100

```
s3_knn2 = knnreg(price ~ mileage, data=s3_train, k=2)
s3_knn5 = knnreg(price ~ mileage, data=s3_train, k=5)
s3_knn10 = knnreg(price ~ mileage, data=s3_train, k=10)
s3_knn15 = knnreg(price ~ mileage, data=s3_train, k=15)
s3_knn20 = knnreg(price ~ mileage, data=s3_train, k=20)
s3_knn25 = knnreg(price ~ mileage, data=s3_train, k=25)
s3_knn50 = knnreg(price ~ mileage, data=s3_train, k=50)
s3_knn100 = knnreg(price ~ mileage, data=s3_train, k=100)

s3_test = s3_test %>%
  mutate(price_pred = predict(s3_knn2, s3_test)) %>%
  mutate(price_pred = predict(s3_knn5, s3_test)) %>%
  mutate(price_pred = predict(s3_knn10, s3_test)) %>%
  mutate(price_pred = predict(s3_knn15, s3_test)) %>%
```

```r
  mutate(price_pred = predict(s3_knn20, s3_test)) %>%
  mutate(price_pred = predict(s3_knn25, s3_test)) %>%
  mutate(price_pred = predict(s3_knn50, s3_test)) %>%
  mutate(price_pred = predict(s3_knn100, s3_test))
```

Calculating RMSE

```r
modelr::rmse(s3_knn2, s3_test)
```

```
## [1] 11485.82
```

```r
modelr::rmse(s3_knn5, s3_test)
```

```
## [1] 10166.73
```

```r
modelr::rmse(s3_knn10, s3_test)
```

```
## [1] 10054.07
```

```r
modelr::rmse(s3_knn15, s3_test)
```

```
## [1] 9993.88
```

```r
modelr::rmse(s3_knn20, s3_test)
```

```
## [1] 9989.683
```

```r
modelr::rmse(s3_knn25, s3_test)
```

```
## [1] 9971.978
```

```r
modelr::rmse(s3_knn50, s3_test)
```

```
## [1] 9876.5
```

```r
modelr::rmse(s3_knn100, s3_test)
```

```
## [1] 10898.11
```

When k=15, RMSE minimized

65 AMG trim of sclass

k=2,5,10,15,20,25,50,100

```r
s6_knn2 = knnreg(price ~ mileage, data=s6_train, k=2)
s6_knn5 = knnreg(price ~ mileage, data=s6_train, k=5)
s6_knn10 = knnreg(price ~ mileage, data=s6_train, k=10)
s6_knn15 = knnreg(price ~ mileage, data=s6_train, k=15)
s6_knn20 = knnreg(price ~ mileage, data=s6_train, k=20)
s6_knn25 = knnreg(price ~ mileage, data=s6_train, k=25)
s6_knn50 = knnreg(price ~ mileage, data=s6_train, k=50)
s6_knn100 = knnreg(price ~ mileage, data=s6_train, k=100)

s6_test = s6_test %>%
  mutate(price_pred = predict(s6_knn2, s6_test)) %>%
  mutate(price_pred = predict(s6_knn5, s6_test)) %>%
  mutate(price_pred = predict(s6_knn10, s6_test)) %>%
  mutate(price_pred = predict(s6_knn15, s6_test)) %>%
```

```
  mutate(price_pred = predict(s6_knn20, s6_test)) %>%
  mutate(price_pred = predict(s6_knn25, s6_test)) %>%
  mutate(price_pred = predict(s6_knn50, s6_test)) %>%
  mutate(price_pred = predict(s6_knn100, s6_test))
```

Calculating RMSE

```
modelr::rmse(s6_knn2, s6_test)
```

```
## [1] 24363.51
```

```
modelr::rmse(s6_knn5, s6_test)
```

```
## [1] 22762.85
```

```
modelr::rmse(s6_knn10, s6_test)
```

```
## [1] 22557.7
```

```
modelr::rmse(s6_knn15, s6_test)
```

```
## [1] 23175.97
```

```
modelr::rmse(s6_knn20, s6_test)
```

```
## [1] 23384.51
```

```
modelr::rmse(s6_knn25, s6_test)
```

```
## [1] 24151.82
```

```
modelr::rmse(s6_knn50, s6_test)
```

```
## [1] 26499.37
```

```
modelr::rmse(s6_knn100, s6_test)
```

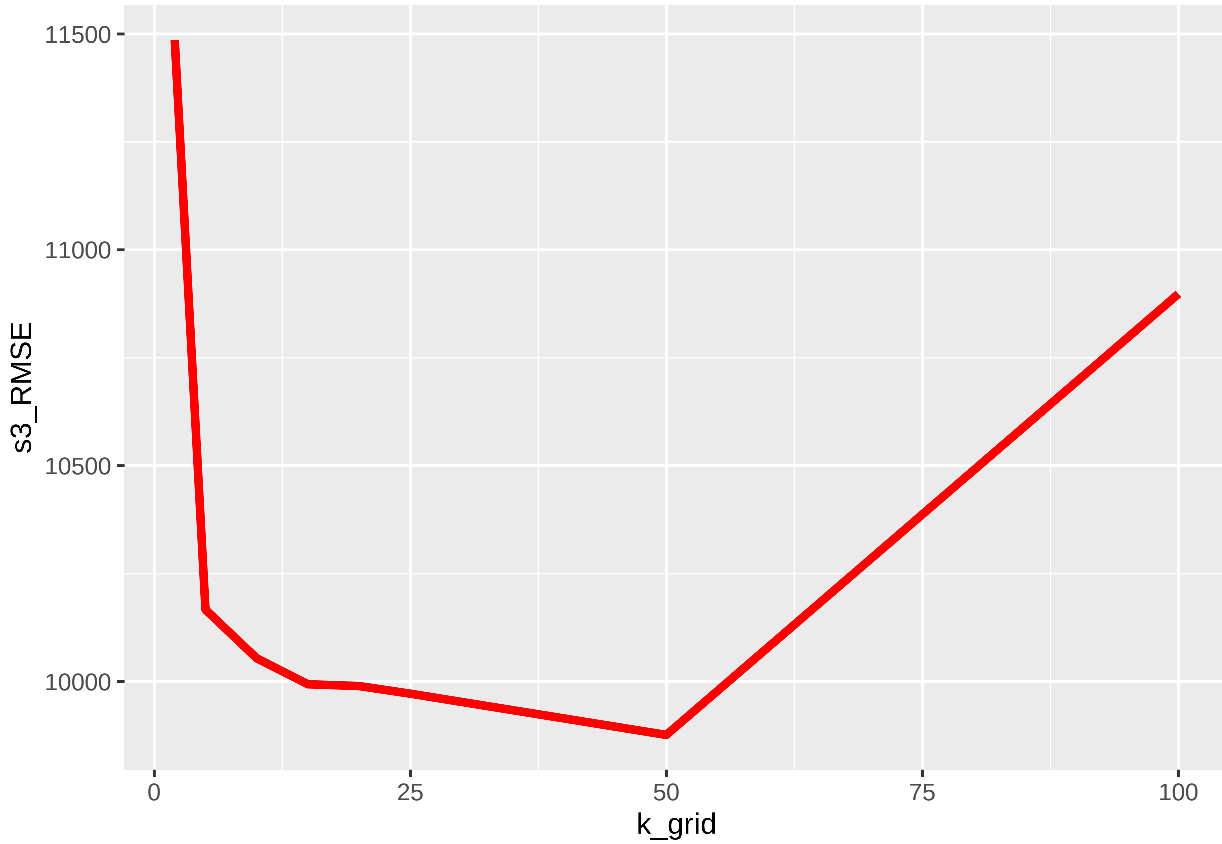
```
## [1] 34639.92
```

When k=20, RMSE minimized

K vs RMSE

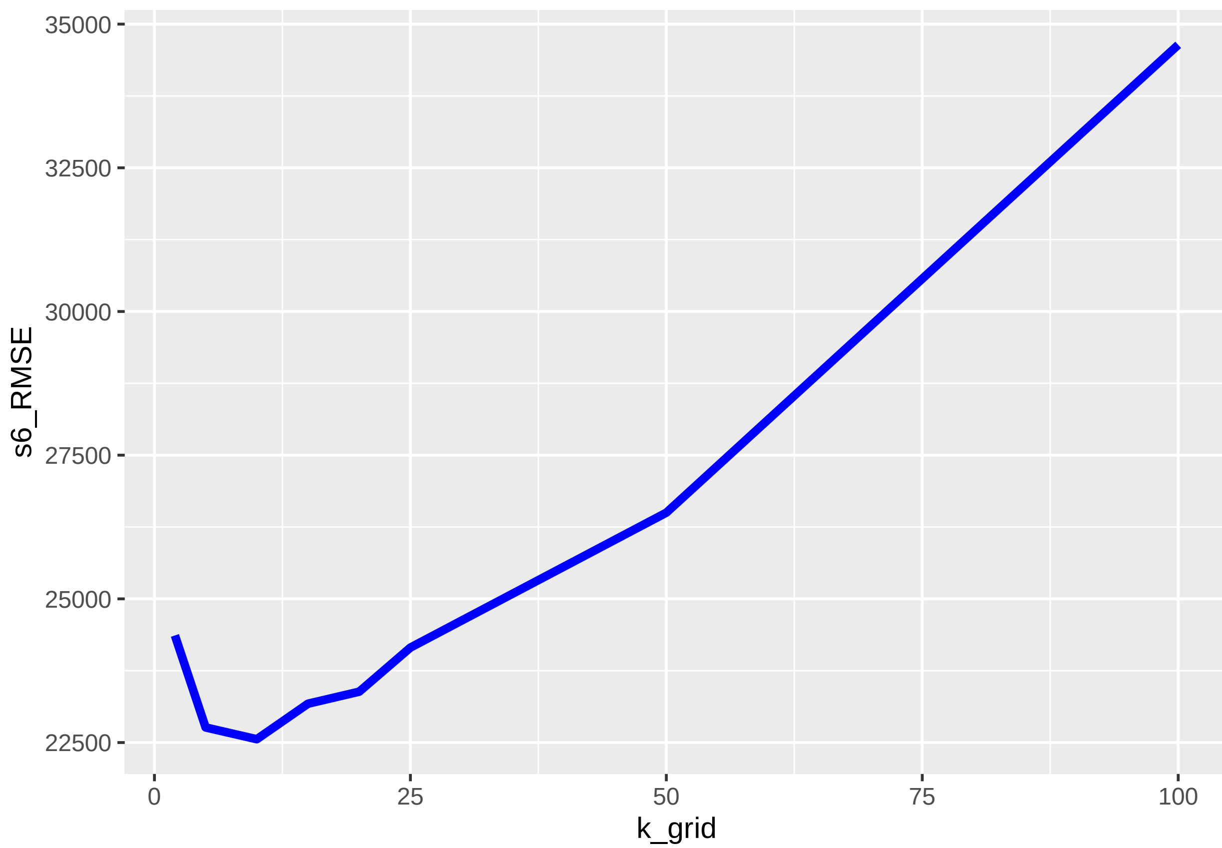
350 trim of sclass

```
k_grid = c(2,5,10,15,20,25,50,100)

s3_RMSE = foreach(k=k_grid, .combine='c') %do% {
  s3_knn_model = knnreg(price~mileage, data=s3_train, k=k)
  modelr::rmse(s3_knn_model, s3_test)}

ggplot() +
  geom_line(aes(x = k_grid, y = s3_RMSE), color='red', size=1.5)
```



65 AMG trim of sclass

```
k_grid = c(2,5,10,15,20,25,50,100)

s6_RMSE = foreach(k=k_grid, .combine='c') %do% {
  s6_knn_model = knnreg(price~mileage, data=s6_train, k=k)
  modelr::rmse(s6_knn_model, s6_test)}

ggplot() +
  geom_line(aes(x = k_grid, y = s6_RMSE), color='blue', size=1.5)
```



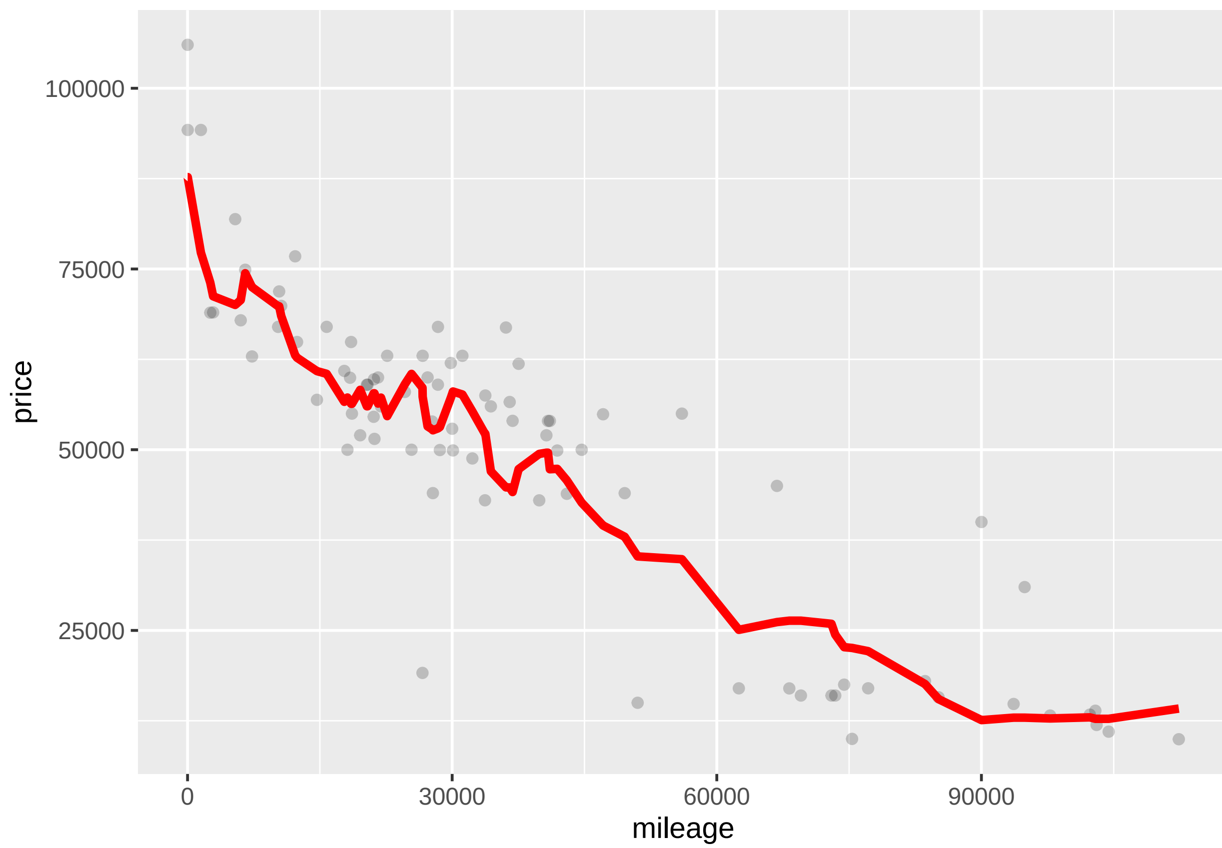
plot 2 models

350 trim with K = 15

```
s3_knn15_plot = knnreg(price ~ mileage, data=s3_train, k=15)
s3_test = s3_test %>%
  mutate(price_pred = predict(s3_knn15_plot, s3_test))
modelr::rmse(s3_knn15_plot, s3_test)
```

```
## [1] 9993.88
```

```
s3_plot = ggplot(data = s3_test) +
  geom_point(mapping = aes(x = mileage, y = price), alpha=0.2) +
  geom_line(aes(x = mileage, y = price_pred), color='red', size=1.5)

s3_plot
```

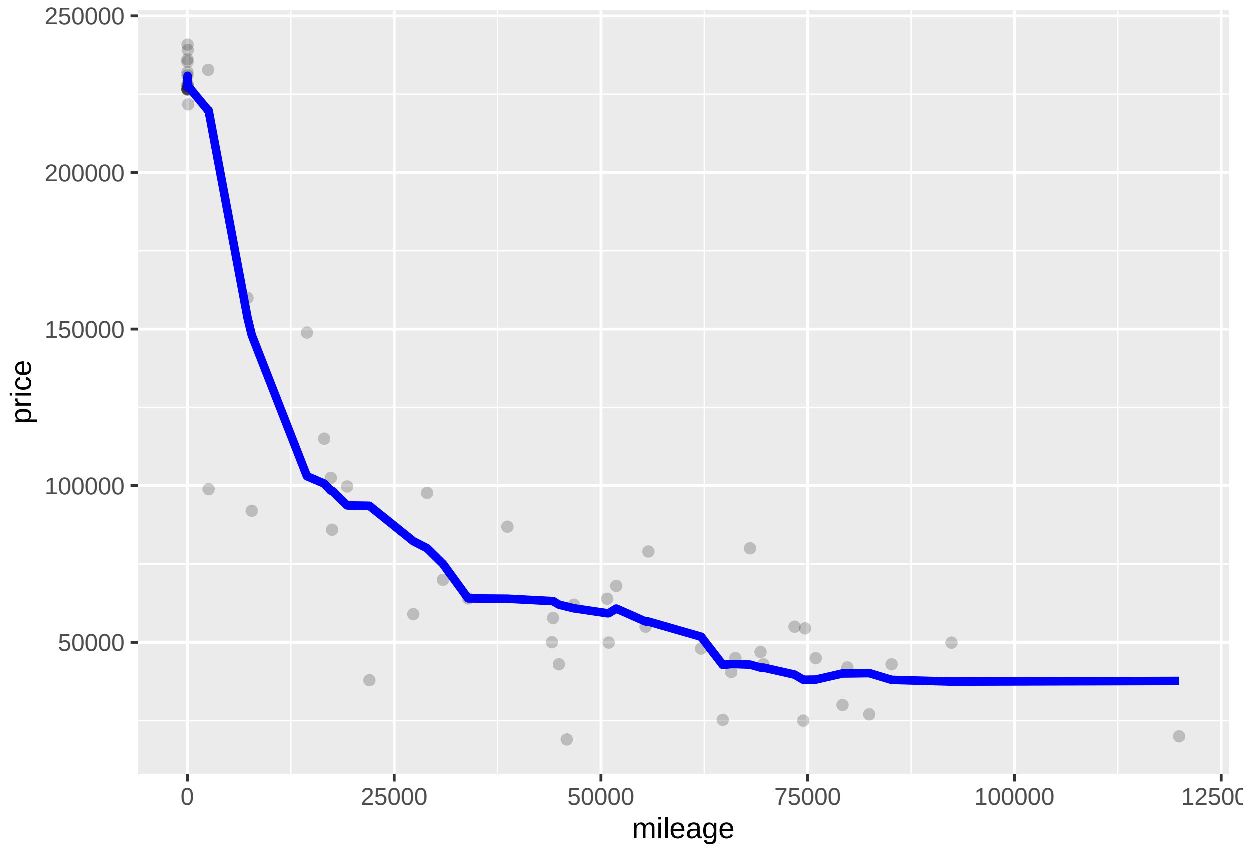


65 AMG trim with K = 20

```
s6_knn20_plot = knnreg(price ~ mileage, data=s6_train, k=20)
s6_test = s6_test %>%
  mutate(price_pred = predict(s6_knn20_plot, s6_test))
modelr::rmse(s6_knn20_plot, s6_test)
```

```
## [1] 23384.51
```

```
s6_plot = ggplot(data = s6_test) +
  geom_point(mapping = aes(x = mileage, y = price), alpha=0.2) +
  geom_line(aes(x = mileage, y = price_pred), color='blue', size=1.5)

s6_plot
```



RMSE of ‘350 trim’ is smaller than ‘65 AMG Trim’ in optimal ‘K’. So, ‘350 trim’ yields a larger optimal value of ‘K’