

Car Price Prediction

Jiachen_Liu and Thinh_Nguyen

4/15/2021

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v tibble 3.1.0      v dplyr    1.0.5  
## v tidyr   1.1.3      v stringr  1.4.0  
## v readr   1.4.0      vforcats  0.5.1  
## v purrr   0.3.4
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'readr' was built under R version 4.0.5
```

```
## Warning: package 'purrr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## Warning: package 'forcats' was built under R version 4.0.5
```

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

```
library(dplyr)
```

EDA

```
car = read.csv('raw_data.csv')
```

```
head(car)
```

```
##   X model year price transmission mileage fuelType tax mpg engineSize brand
## 1 0   A1 2017 12500      Manual  15735 Petrol 150 55.4       1.4 audi
## 2 1   A6 2016 16500     Automatic 36203 Diesel 20 64.2       2.0 audi
## 3 2   A1 2016 11000      Manual  29946 Petrol 30 55.4       1.4 audi
## 4 3   A4 2017 16800     Automatic 25952 Diesel 145 67.3       2.0 audi
## 5 4   A3 2019 17300      Manual  1998 Petrol 145 49.6       1.0 audi
## 6 5   A1 2016 13900     Automatic 32260 Petrol 30 58.9       1.4 audi
```

```
## choose subset of bmw data
bmw = car[which( car$brand == 'bmw' ),]
bmw = subset(bmw, select = -1 )
head(bmw)
```

```
##           model year price transmission mileage fuelType tax mpg engineSize
## 10669 5 Series 2014 11200     Automatic 67068 Diesel 125 57.6       2.0
## 10670 6 Series 2018 27000     Automatic 14827 Petrol 145 42.8       2.0
## 10671 5 Series 2016 16000     Automatic 62794 Diesel 160 51.4       3.0
## 10672 1 Series 2017 12750     Automatic 26676 Diesel 145 72.4       1.5
## 10673 7 Series 2014 14500     Automatic 39554 Diesel 160 50.4       3.0
## 10674 5 Series 2016 14900     Automatic 35309 Diesel 125 60.1       2.0
##   brand
## 10669 bmw
## 10670 bmw
## 10671 bmw
## 10672 bmw
## 10673 bmw
## 10674 bmw
```

```
## summary of bmw data
summary(bmw)
```

```
##   model          year      price      transmission
## Length:10781  Min.   :1996  Min.   : 1200  Length:10781
## Class  :character 1st Qu.:2016  1st Qu.: 14950  Class  :character
## Mode   :character Median :2017  Median : 20462  Mode   :character
##                  Mean   :2017  Mean   : 22733
##                  3rd Qu.:2019  3rd Qu.: 27940
##                  Max.   :2020  Max.   :123456
##   mileage      fuelType      tax      mpg
##   Min.   :    1  Length:10781  Min.   : 0.0  Min.   : 5.5
##   1st Qu.: 5529  Class  :character  1st Qu.:135.0  1st Qu.: 45.6
##   Median :18347  Mode   :character  Median :145.0  Median : 53.3
##   Mean   :25497                           Mean   :131.7  Mean   : 56.4
##   3rd Qu.:38206                           3rd Qu.:145.0  3rd Qu.: 62.8
##   Max.   :214000                          Max.   :580.0  Max.   :470.8
##   engineSize      brand
##   Min.   :0.000  Length:10781
##   1st Qu.:2.000  Class  :character
##   Median :2.000  Mode   :character
##   Mean   :2.168
##   3rd Qu.:2.000
##   Max.   :6.600
```

```
## check missing value for each column  
library(mice)
```

```
## Warning: package 'mice' was built under R version 4.0.5
```

```
##  
## Attaching package: 'mice'
```

```
## The following object is masked from 'package:stats':  
##  
##     filter
```

```
## The following objects are masked from 'package:base':  
##  
##     cbind, rbind
```

```
library(VIM)
```

```
## Warning: package 'VIM' was built under R version 4.0.5
```

```
## Loading required package: colorspace
```

```
## Warning: package 'colorspace' was built under R version 4.0.5
```

```
## Loading required package: grid
```

```
## VIM is ready to use.
```

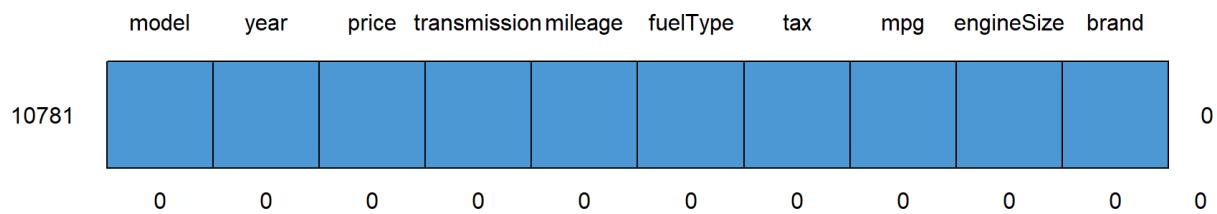
```
## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues
```

```
##  
## Attaching package: 'VIM'
```

```
## The following object is masked from 'package:datasets':  
##  
##     sleep
```

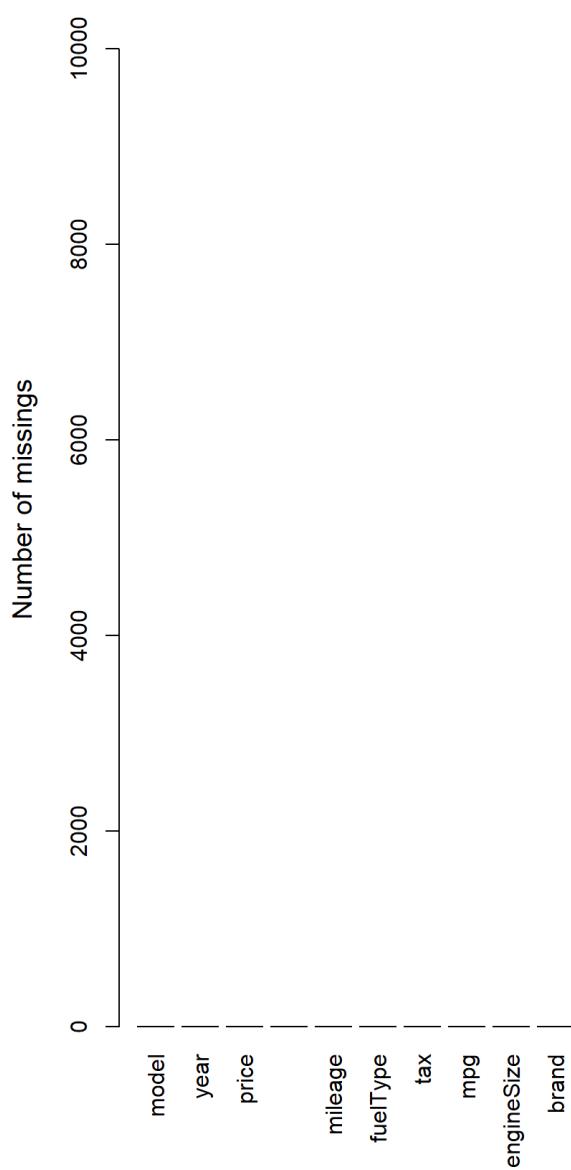
```
md.pattern(bmw)
```

```
##  /\  /\  
## { `---, }  
## { 0 0 }  
## ==> V <== No need for mice. This data set is completely observed.  
##  \|/ /  
##    `-----,
```



```
##      model year price transmission mileage fuelType tax mpg engineSize brand
## 10781     1    1     1           1       1       1     1   1     1     1  0
##          0    0     0           0       0       0     0   0     0     0  0
```

```
aggr(bmw, prop=FALSE, numbers=TRUE)
```



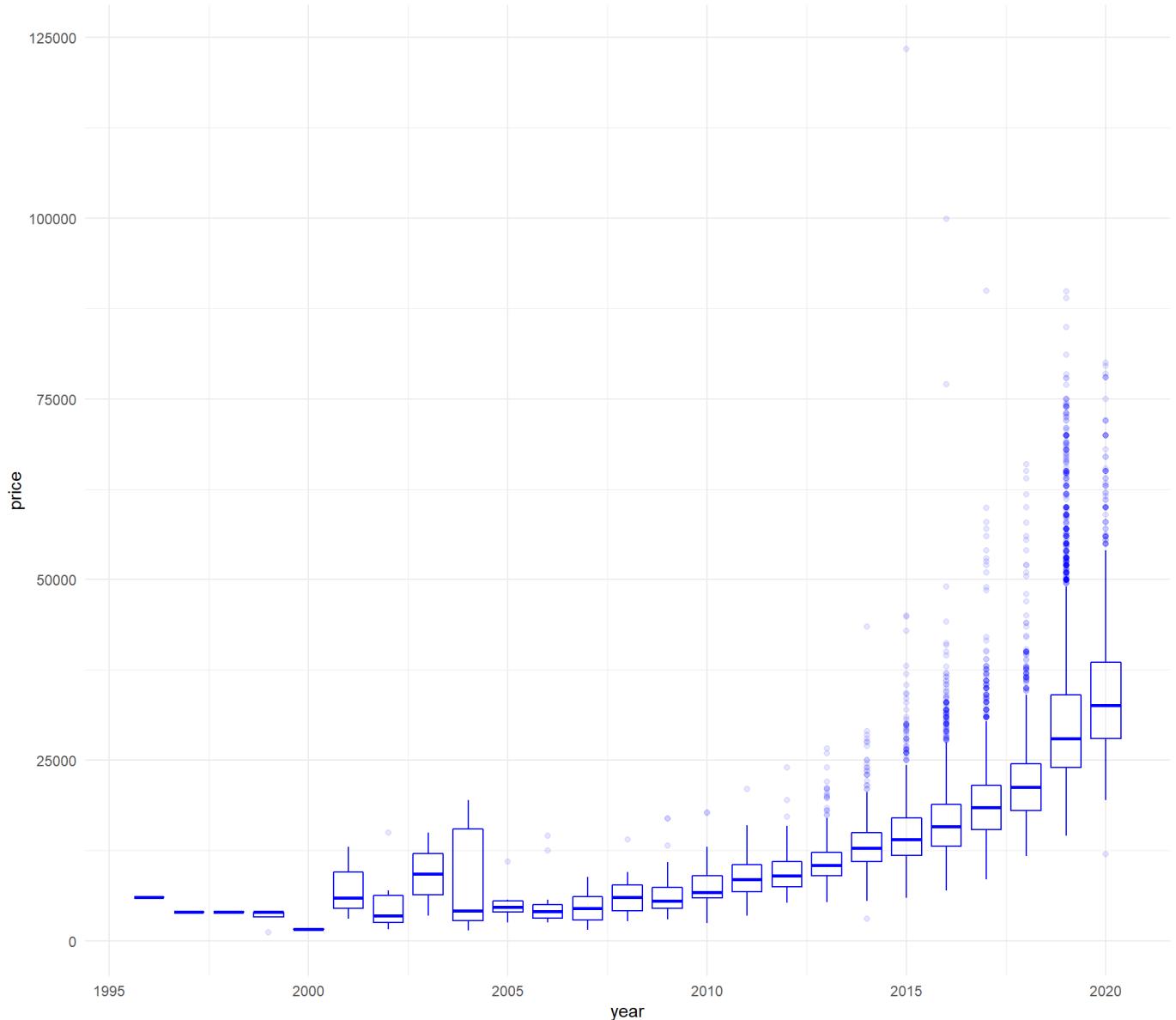
```
colnames(bmw)
```

```
## [1] "model"          "year"           "price"          "transmission"   "mileage"
## [6] "fuelType"        "tax"            "mpg"            "engineSize"     "brand"
```

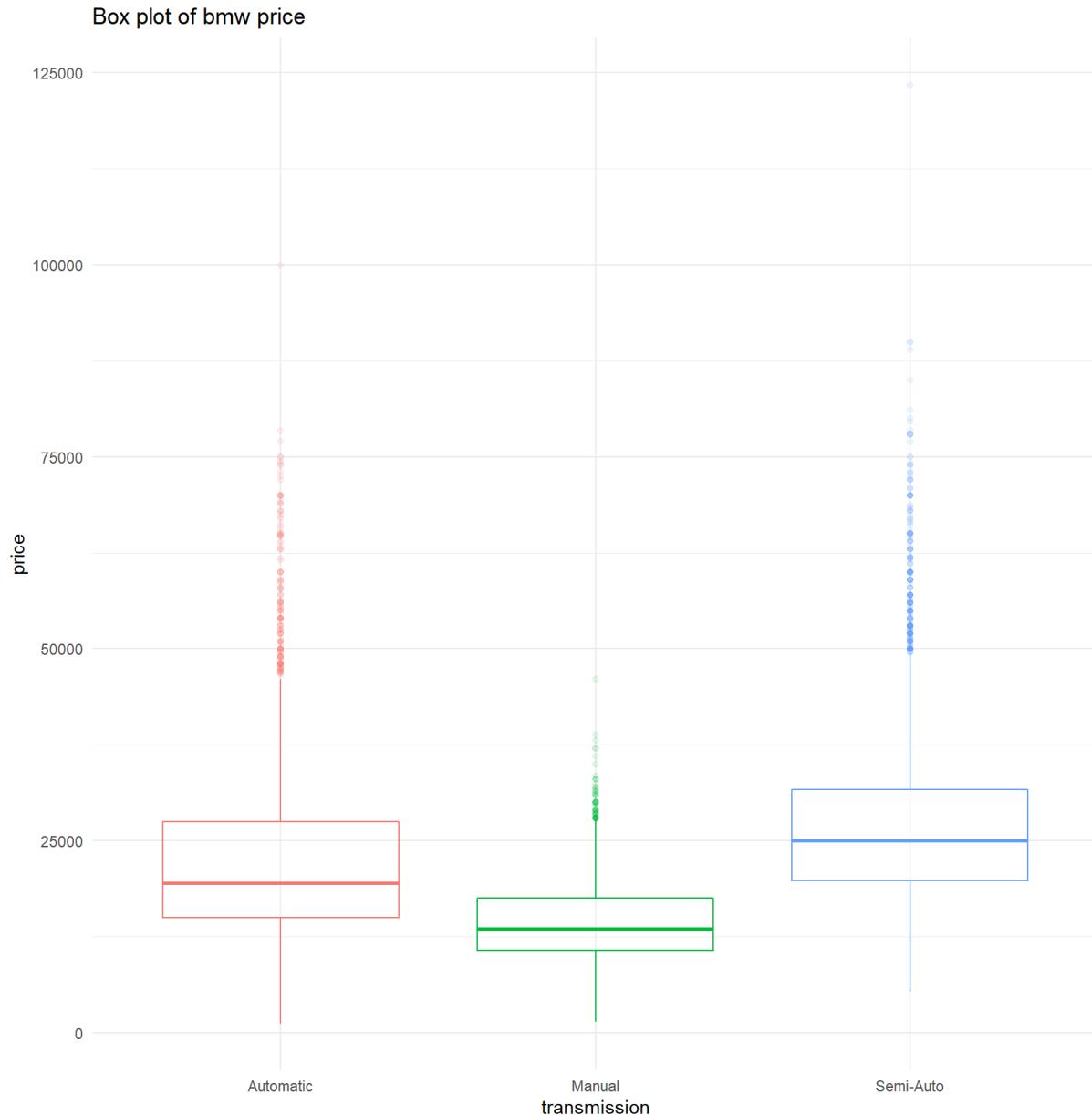
Exploratory Data Analysis

```
# box plot for the bmw year
ggplot(data = bmw, aes( x = year, y = price, group = year ) ) + geom_boxplot(color = "blue", alpha = 0.1 ) + ggttitle("Box plot of bmw year") + theme_minimal()
```

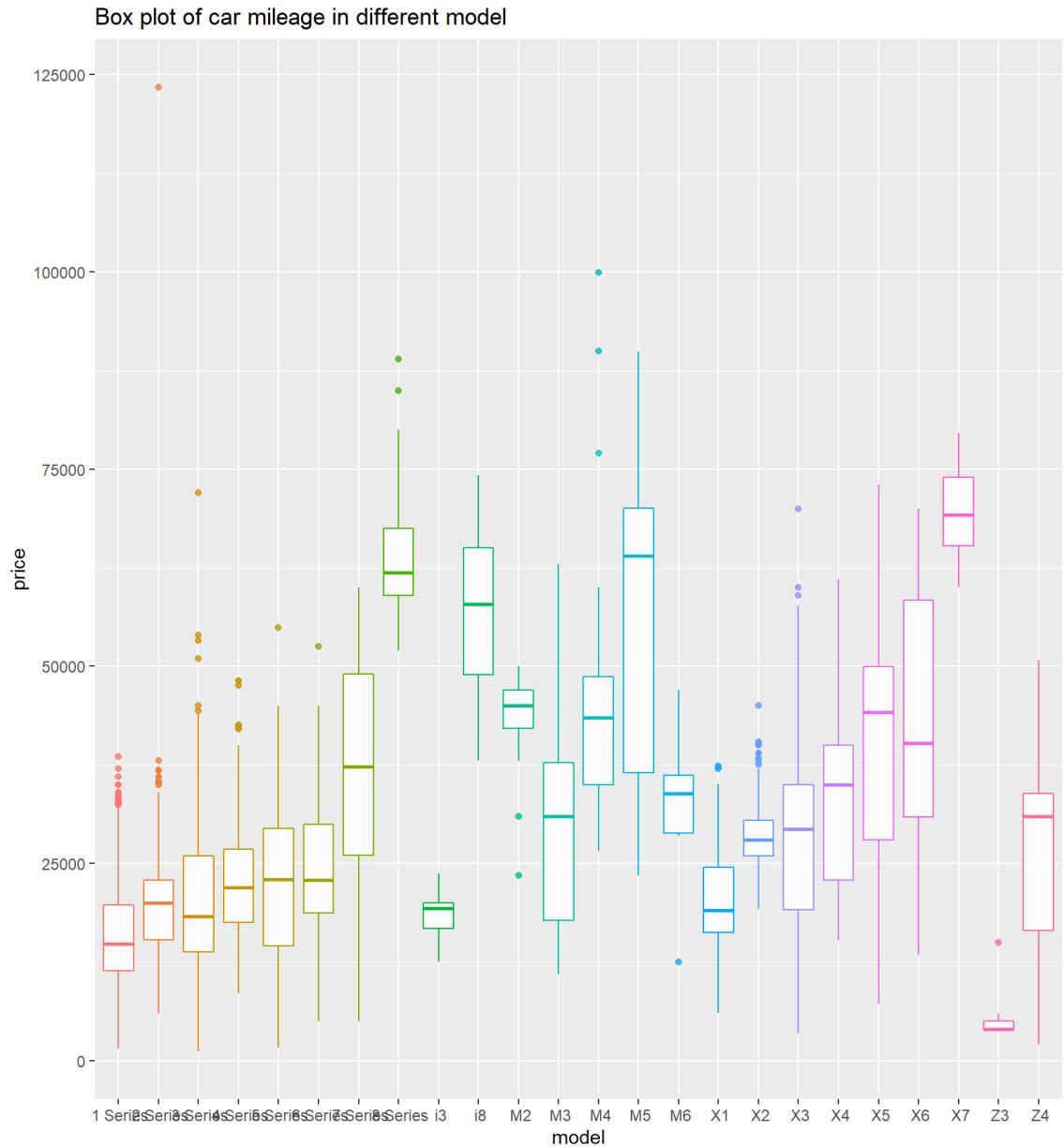
Box plot of bmw year



```
# box plot of bmw price
ggplot(data = bmw, aes(x = transmission, y = price )) + geom_boxplot( aes(color = transmission), alpha = 0.1) + ggttitle("Box plot of bmw price") + theme_minimal()
```



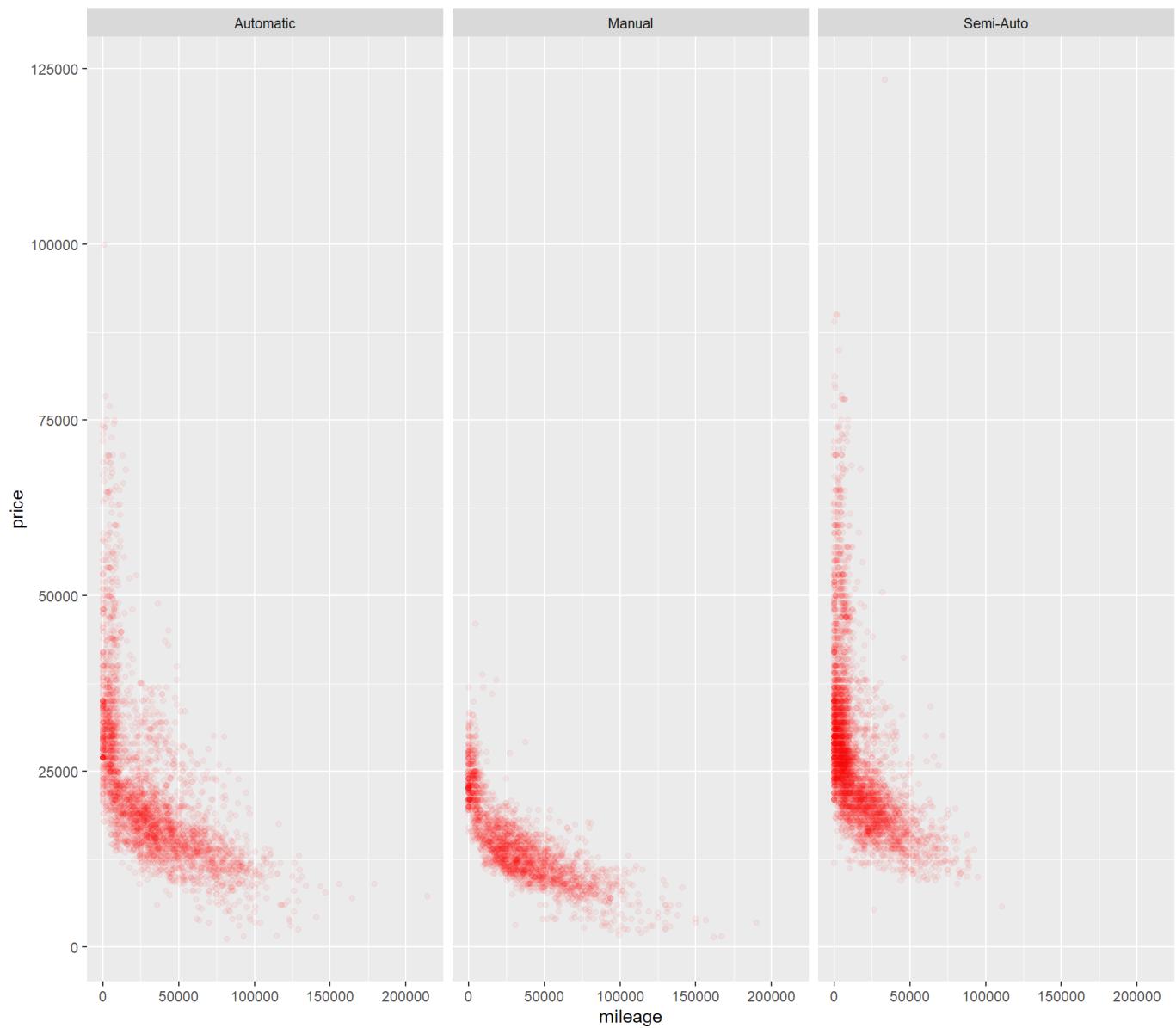
```
# boxplot for car model
ggplot(data = bmw, aes( y = price, x = model, group = model ) ) + geom_boxplot( aes(color = model), alpha = 0.8 ) + ggtitle("Box plot of car mileage in different model")
```



```
# scatter plot of millage and price
```

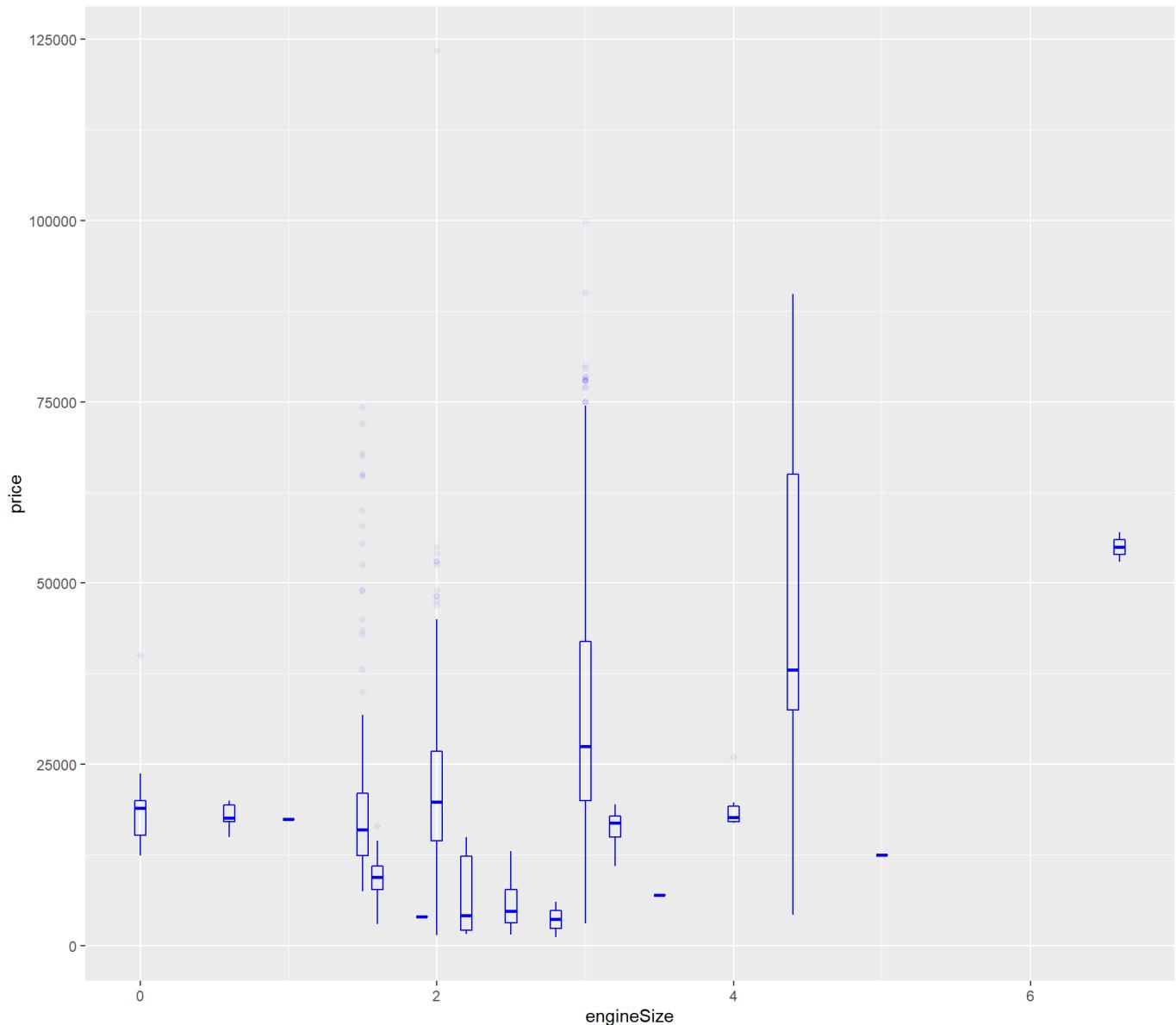
```
ggplot(data = bmw, aes(x = mileage , y = price)) + geom_point(color ="red", alpha= 0.05) + ggtitle("The relationship between millage and price in different transmission") + facet_wrap(~ transmission)
```

The relationship between millage and price in different transmission



```
#scatter plot for car mpg and price  
ggplot(data = bmw, aes(x = engineSize , y = price , group = engineSize)) + geom_boxplot(color = "blue", alpha= 0.05) + ggtitle("The relationship between engine size and price")
```

The relationship between engine size and price



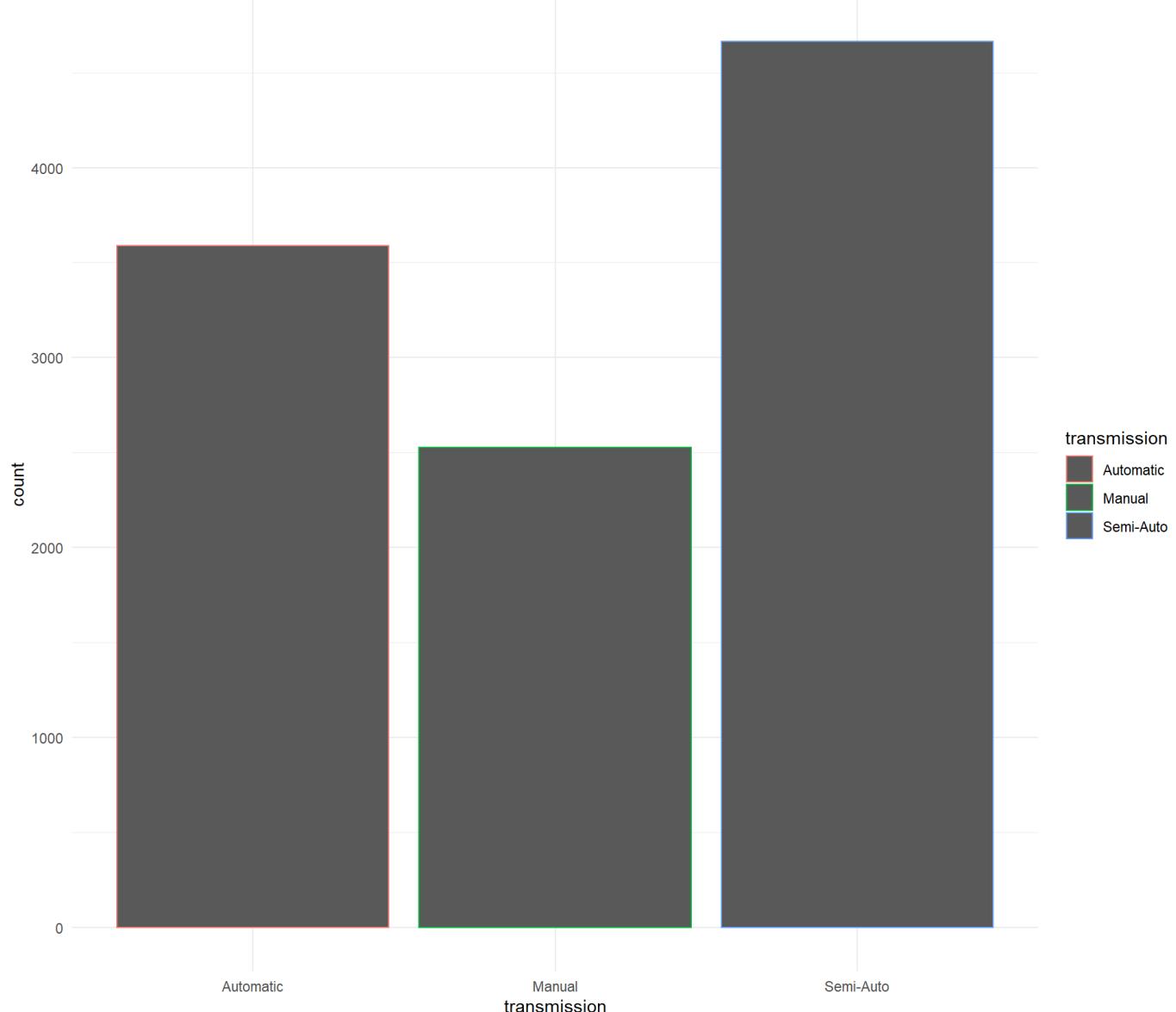
```
ggplot(data = bmw[which(bmw$year>2014), ], aes(x = mileage , y = price)) + geom_point(color ="red", alpha= 0.05) + ggtitle("The relationship between millage and price in different transmission") + facet_wrap(~ year )
```

The relationship between millage and price in different transmission



```
# bar plot of "Total Number of each transmission"
ggplot( data = bmw , aes(x = transmission ), color = "red" ) + geom_bar(aes(color = transmission)) + ggtitle( "Total Number of each transmission") + theme_minimal()
```

Total Number of each transmission



Feature Engineering

encoding categorical data

```
# Feature Engineering  
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.0.5
```

```
## Loading required package: lattice
```

```
##  
## Attaching package: 'caret'
```

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

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.0.5
```

```
##  
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
```

```
##  
##     %+%, alpha
```

```
library(fastDummies)
```

```
## Warning: package 'fastDummies' was built under R version 4.0.5
```

```
bmw_data = dummy_cols(bmw, select_columns =c('year', 'model', 'transmission', 'fuelType', 'engineSize' ) )  
head(bmw_data)
```

```

##      model year price transmission mileage fuelType tax mpg engineSize brand
## 1  5 Series 2014 11200     Automatic   67068 Diesel 125 57.6      2.0 bmw
## 2  6 Series 2018 27000     Automatic  14827 Petrol 145 42.8      2.0 bmw
## 3  5 Series 2016 16000     Automatic  62794 Diesel 160 51.4      3.0 bmw
## 4  1 Series 2017 12750     Automatic  26676 Diesel 145 72.4      1.5 bmw
## 5  7 Series 2014 14500     Automatic  39554 Diesel 160 50.4      3.0 bmw
## 6  5 Series 2016 14900     Automatic  35309 Diesel 125 60.1      2.0 bmw
##   year_1996 year_1997 year_1998 year_1999 year_2000 year_2001 year_2002
## 1          0          0          0          0          0          0          0
## 2          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 4          0          0          0          0          0          0          0
## 5          0          0          0          0          0          0          0
## 6          0          0          0          0          0          0          0
##   year_2003 year_2004 year_2005 year_2006 year_2007 year_2008 year_2009
## 1          0          0          0          0          0          0          0
## 2          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 4          0          0          0          0          0          0          0
## 5          0          0          0          0          0          0          0
## 6          0          0          0          0          0          0          0
##   year_2010 year_2011 year_2012 year_2013 year_2014 year_2015 year_2016
## 1          0          0          0          0          1          0          0
## 2          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          1
## 4          0          0          0          0          0          0          0
## 5          0          0          0          0          1          0          0
## 6          0          0          0          0          0          0          1
##   year_2017 year_2018 year_2019 year_2020 model_1 Series model_2 Series
## 1          0          0          0          0          0          0          0
## 2          0          1          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 4          1          0          0          0          0          1          0
## 5          0          0          0          0          0          0          0
## 6          0          0          0          0          0          0          0
##   model_3 Series model_4 Series model_5 Series model_6 Series
## 1          0          0          1          0          0
## 2          0          0          0          0          1
## 3          0          0          1          0          0
## 4          0          0          0          0          0
## 5          0          0          0          0          0
## 6          0          0          1          0          0
##   model_7 Series model_8 Series model_i3 model_i8 model_M2 model_M3
## 1          0          0          0          0          0          0
## 2          0          0          0          0          0          0
## 3          0          0          0          0          0          0
## 4          0          0          0          0          0          0
## 5          1          0          0          0          0          0
## 6          0          0          0          0          0          0
##   model_M4 model_M5 model_M6 model_X1 model_X2 model_X3 model_X4
## 1          0          0          0          0          0          0          0
## 2          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 4          0          0          0          0          0          0          0
## 5          0          0          0          0          0          0          0
## 6          0          0          0          0          0          0          0
##   model_X5 model_X6 model_X7 model_Z3 model_Z4 transmission_Automatic

```

```

## 1      0      0      0      0      0      1
## 2      0      0      0      0      0      1
## 3      0      0      0      0      0      1
## 4      0      0      0      0      0      1
## 5      0      0      0      0      0      1
## 6      0      0      0      0      0      1
##   transmission_Manual transmission_Semi-Auto fuelType_Diesel fuelType_Electric
## 1            0                  0          1          0
## 2            0                  0          0          0
## 3            0                  0          1          0
## 4            0                  0          1          0
## 5            0                  0          1          0
## 6            0                  0          1          0
##   fuelType_Hybrid fuelType_Other fuelType_Petrol engineSize_0 engineSize_0.6
## 1            0          0          0          0          0
## 2            0          0          1          0          0
## 3            0          0          0          0          0
## 4            0          0          0          0          0
## 5            0          0          0          0          0
## 6            0          0          0          0          0
##   engineSize_1 engineSize_1.5 engineSize_1.6 engineSize_1.9 engineSize_2
## 1            0          0          0          0          1
## 2            0          0          0          0          1
## 3            0          0          0          0          0
## 4            0          1          0          0          0
## 5            0          0          0          0          0
## 6            0          0          0          0          1
##   engineSize_2.2 engineSize_2.5 engineSize_2.8 engineSize_3 engineSize_3.2
## 1            0          0          0          0          0
## 2            0          0          0          0          0
## 3            0          0          0          1          0
## 4            0          0          0          0          0
## 5            0          0          0          1          0
## 6            0          0          0          0          0
##   engineSize_3.5 engineSize_4 engineSize_4.4 engineSize_5 engineSize_6.6
## 1            0          0          0          0          0
## 2            0          0          0          0          0
## 3            0          0          0          0          0
## 4            0          0          0          0          0
## 5            0          0          0          0          0
## 6            0          0          0          0          0

```

```

bmw_df = subset(bmw_data , select =c(-year,-model, -transmission, - fuelType, -brand , -engineSi
ze) )
head(bmw_df)

```

```

##   price mileage tax  mpg year_1996 year_1997 year_1998 year_1999 year_2000
## 1 11200    67068 125 57.6        0        0        0        0        0
## 2 27000    14827 145 42.8        0        0        0        0        0
## 3 16000    62794 160 51.4        0        0        0        0        0
## 4 12750    26676 145 72.4        0        0        0        0        0
## 5 14500    39554 160 50.4        0        0        0        0        0
## 6 14900    35309 125 60.1        0        0        0        0        0
##   year_2001 year_2002 year_2003 year_2004 year_2005 year_2006 year_2007
## 1        0        0        0        0        0        0        0
## 2        0        0        0        0        0        0        0
## 3        0        0        0        0        0        0        0
## 4        0        0        0        0        0        0        0
## 5        0        0        0        0        0        0        0
## 6        0        0        0        0        0        0        0
##   year_2008 year_2009 year_2010 year_2011 year_2012 year_2013 year_2014
## 1        0        0        0        0        0        0        1
## 2        0        0        0        0        0        0        0
## 3        0        0        0        0        0        0        0
## 4        0        0        0        0        0        0        0
## 5        0        0        0        0        0        0        1
## 6        0        0        0        0        0        0        0
##   year_2015 year_2016 year_2017 year_2018 year_2019 year_2020 model_1 Series
## 1        0        0        0        0        0        0        0
## 2        0        0        0        1        0        0        0
## 3        0        1        0        0        0        0        0
## 4        0        0        1        0        0        0        1
## 5        0        0        0        0        0        0        0
## 6        0        1        0        0        0        0        0
##   model_2 Series model_3 Series model_4 Series model_5 Series
## 1        0        0        0        0        1
## 2        0        0        0        0        0
## 3        0        0        0        0        1
## 4        0        0        0        0        0
## 5        0        0        0        0        0
## 6        0        0        0        0        1
##   model_6 Series model_7 Series model_8 Series model_i3 model_i8 model_M2
## 1        0        0        0        0        0        0        0
## 2        1        0        0        0        0        0        0
## 3        0        0        0        0        0        0        0
## 4        0        0        0        0        0        0        0
## 5        0        1        0        0        0        0        0
## 6        0        0        0        0        0        0        0
##   model_M3 model_M4 model_M5 model_M6 model_X1 model_X2 model_X3
## 1        0        0        0        0        0        0        0
## 2        0        0        0        0        0        0        0
## 3        0        0        0        0        0        0        0
## 4        0        0        0        0        0        0        0
## 5        0        0        0        0        0        0        0
## 6        0        0        0        0        0        0        0
##   model_X4 model_X5 model_X6 model_X7 model_Z3 model_Z4
## 1        0        0        0        0        0        0
## 2        0        0        0        0        0        0
## 3        0        0        0        0        0        0
## 4        0        0        0        0        0        0
## 5        0        0        0        0        0        0
## 6        0        0        0        0        0        0
##   transmission_Automatic transmission_Manual transmission_Semi-Auto

```

```

## 1          1          0          0
## 2          1          0          0
## 3          1          0          0
## 4          1          0          0
## 5          1          0          0
## 6          1          0          0
##   fuelType_Diesel fuelType_Electric fuelType_Hybrid fuelType_Other
## 1          1          0          0          0
## 2          0          0          0          0
## 3          1          0          0          0
## 4          1          0          0          0
## 5          1          0          0          0
## 6          1          0          0          0
##   fuelType_Petrol engineSize_0 engineSize_0.6 engineSize_1 engineSize_1.5
## 1          0          0          0          0          0
## 2          1          0          0          0          0
## 3          0          0          0          0          0
## 4          0          0          0          0          1
## 5          0          0          0          0          0
## 6          0          0          0          0          0
##   engineSize_1.6 engineSize_1.9 engineSize_2 engineSize_2.2 engineSize_2.5
## 1          0          0          1          0          0
## 2          0          0          1          0          0
## 3          0          0          0          0          0
## 4          0          0          0          0          0
## 5          0          0          0          0          0
## 6          0          0          1          0          0
##   engineSize_2.8 engineSize_3 engineSize_3.2 engineSize_3.5 engineSize_4
## 1          0          0          0          0          0
## 2          0          0          0          0          0
## 3          0          1          0          0          0
## 4          0          0          0          0          0
## 5          0          1          0          0          0
## 6          0          0          0          0          0
##   engineSize_4.4 engineSize_5 engineSize_6.6
## 1          0          0          0
## 2          0          0          0
## 3          0          0          0
## 4          0          0          0
## 5          0          0          0
## 6          0          0          0

```

Normalize data and prepare for PCA

```

bmw_scale = as.data.frame(scale(bmw_df[, -2]))
head(bmw_scale)

```

```

##      price      tax      mpg year_1996 year_1997 year_1998
## 1 -1.0103263 -0.1089577  0.03832423 -0.00963098 -0.00963098 -0.00963098
## 2  0.3737533  0.2161887 -0.43396157 -0.00963098 -0.00963098 -0.00963098
## 3 -0.5898465  0.4600485 -0.15952523 -0.00963098 -0.00963098 -0.00963098
## 4 -0.8745464  0.2161887  0.51061002 -0.00963098 -0.00963098 -0.00963098
## 5 -0.7212464  0.4600485 -0.19143643 -0.00963098 -0.00963098 -0.00963098
## 6 -0.6862064 -0.1089577  0.11810223 -0.00963098 -0.00963098 -0.00963098
##      year_1999 year_2000 year_2001 year_2002 year_2003 year_2004
## 1 -0.01926464 -0.01362089 -0.01668289 -0.02359646 -0.01362089 -0.03337973
## 2 -0.01926464 -0.01362089 -0.01668289 -0.02359646 -0.01362089 -0.03337973
## 3 -0.01926464 -0.01362089 -0.01668289 -0.02359646 -0.01362089 -0.03337973
## 4 -0.01926464 -0.01362089 -0.01668289 -0.02359646 -0.01362089 -0.03337973
## 5 -0.01926464 -0.01362089 -0.01668289 -0.02359646 -0.01362089 -0.03337973
## 6 -0.01926464 -0.01362089 -0.01668289 -0.02359646 -0.01362089 -0.03337973
##      year_2005 year_2006 year_2007 year_2008 year_2009 year_2010
## 1 -0.02359646 -0.03605758 -0.03855075 -0.04623576 -0.05282215 -0.06178309
## 2 -0.02359646 -0.03605758 -0.03855075 -0.04623576 -0.05282215 -0.06178309
## 3 -0.02359646 -0.03605758 -0.03855075 -0.04623576 -0.05282215 -0.06178309
## 4 -0.02359646 -0.03605758 -0.03855075 -0.04623576 -0.05282215 -0.06178309
## 5 -0.02359646 -0.03605758 -0.03855075 -0.04623576 -0.05282215 -0.06178309
## 6 -0.02359646 -0.03605758 -0.03855075 -0.04623576 -0.05282215 -0.06178309
##      year_2011 year_2012 year_2013 year_2014 year_2015 year_2016 year_2017
## 1 -0.06893902 -0.1056414 -0.1850533  4.5295760 -0.3057939 -0.459853 -0.435819
## 2 -0.06893902 -0.1056414 -0.1850533 -0.2207507 -0.3057939 -0.459853 -0.435819
## 3 -0.06893902 -0.1056414 -0.1850533 -0.2207507 -0.3057939  2.174406 -0.435819
## 4 -0.06893902 -0.1056414 -0.1850533 -0.2207507 -0.3057939 -0.459853  2.294318
## 5 -0.06893902 -0.1056414 -0.1850533  4.5295760 -0.3057939 -0.459853 -0.435819
## 6 -0.06893902 -0.1056414 -0.1850533 -0.2207507 -0.3057939  2.174406 -0.435819
##      year_2018 year_2019 year_2020 model_1 Series model_2 Series
## 1 -0.2921713 -0.6910967 -0.2700798      -0.4726781      -0.3586812
## 2  3.4223321 -0.6910967 -0.2700798      -0.4726781      -0.3586812
## 3 -0.2921713 -0.6910967 -0.2700798      -0.4726781      -0.3586812
## 4 -0.2921713 -0.6910967 -0.2700798      2.1154085      -0.3586812
## 5 -0.2921713 -0.6910967 -0.2700798      -0.4726781      -0.3586812
## 6 -0.2921713 -0.6910967 -0.2700798      -0.4726781      -0.3586812
##      model_3 Series model_4 Series model_5 Series model_6 Series
## 1   -0.5412659   -0.3188517    3.0345389     -0.1005885
## 2   -0.5412659   -0.3188517   -0.3295088     9.9405688
## 3   -0.5412659   -0.3188517    3.0345389     -0.1005885
## 4   -0.5412659   -0.3188517   -0.3295088     -0.1005885
## 5   -0.5412659   -0.3188517   -0.3295088     -0.1005885
## 6   -0.5412659   -0.3188517    3.0345389     -0.1005885
##      model_7 Series model_8 Series model_i3 model_i8 model_M2
## 1   -0.09964347  -0.06025174 -0.06327795 -0.03973905 -0.04417569
## 2   -0.09964347  -0.06025174 -0.06327795 -0.03973905 -0.04417569
## 3   -0.09964347  -0.06025174 -0.06327795 -0.03973905 -0.04417569
## 4   -0.09964347  -0.06025174 -0.06327795 -0.03973905 -0.04417569
## 5   10.03484957 -0.06025174 -0.06327795 -0.03973905 -0.04417569
## 6   -0.09964347  -0.06025174 -0.06327795 -0.03973905 -0.04417569
##      model_M3 model_M4 model_M5 model_M6 model_X1 model_X2 model_X3
## 1   -0.0501045  -0.1083023 -0.0519319 -0.02724937 -0.2838624 -0.1656633 -0.2320694
## 2   -0.0501045  -0.1083023 -0.0519319 -0.02724937 -0.2838624 -0.1656633 -0.2320694
## 3   -0.0501045  -0.1083023 -0.0519319 -0.02724937 -0.2838624 -0.1656633 -0.2320694
## 4   -0.0501045  -0.1083023 -0.0519319 -0.02724937 -0.2838624 -0.1656633 -0.2320694
## 5   -0.0501045  -0.1083023 -0.0519319 -0.02724937 -0.2838624 -0.1656633 -0.2320694
## 6   -0.0501045  -0.1083023 -0.0519319 -0.02724937 -0.2838624 -0.1656633 -0.2320694
##      model_X4 model_X5 model_X6 model_X7 model_Z3 model_Z4

```

```

## 1 -0.1299309 -0.213015 -0.09964347 -0.07160483 -0.02548827 -0.1005885
## 2 -0.1299309 -0.213015 -0.09964347 -0.07160483 -0.02548827 -0.1005885
## 3 -0.1299309 -0.213015 -0.09964347 -0.07160483 -0.02548827 -0.1005885
## 4 -0.1299309 -0.213015 -0.09964347 -0.07160483 -0.02548827 -0.1005885
## 5 -0.1299309 -0.213015 -0.09964347 -0.07160483 -0.02548827 -0.1005885
## 6 -0.1299309 -0.213015 -0.09964347 -0.07160483 -0.02548827 -0.1005885
##   transmission_Automatic transmission_Manual transmission_Semi-Auto
## 1           1.415822      -0.5532867      -0.8734821
## 2           1.415822      -0.5532867      -0.8734821
## 3           1.415822      -0.5532867      -0.8734821
## 4           1.415822      -0.5532867      -0.8734821
## 5           1.415822      -0.5532867      -0.8734821
## 6           1.415822      -0.5532867      -0.8734821
##   fuelType_Diesel fuelType_Electric fuelType_Hybrid fuelType_Other
## 1          0.7308732     -0.01668289     -0.1685952     -0.05787992
## 2         -1.3680995     -0.01668289     -0.1685952     -0.05787992
## 3          0.7308732     -0.01668289     -0.1685952     -0.05787992
## 4          0.7308732     -0.01668289     -0.1685952     -0.05787992
## 5          0.7308732     -0.01668289     -0.1685952     -0.05787992
## 6          0.7308732     -0.01668289     -0.1685952     -0.05787992
##   fuelType_Petrol engineSize_0 engineSize_0.6 engineSize_1 engineSize_1.5
## 1         -0.6811542    -0.066168     -0.02548827    -0.00963098    -0.3962239
## 2          1.4679601    -0.066168     -0.02548827    -0.00963098    -0.3962239
## 3         -0.6811542    -0.066168     -0.02548827    -0.00963098    -0.3962239
## 4         -0.6811542    -0.066168     -0.02548827    -0.00963098    2.5235915
## 5         -0.6811542    -0.066168     -0.02548827    -0.00963098    -0.3962239
## 6         -0.6811542    -0.066168     -0.02548827    -0.00963098    -0.3962239
##   engineSize_1.6 engineSize_1.9 engineSize_2 engineSize_2.2 engineSize_2.5
## 1         -0.1010579    -0.02153952    0.7997728     -0.02359646    -0.03605758
## 2         -0.1010579    -0.02153952    0.7997728     -0.02359646    -0.03605758
## 3         -0.1010579    -0.02153952   -1.2502392     -0.02359646    -0.03605758
## 4         -0.1010579    -0.02153952   -1.2502392     -0.02359646    -0.03605758
## 5         -0.1010579    -0.02153952   -1.2502392     -0.02359646    -0.03605758
## 6         -0.1010579    -0.02153952    0.7997728     -0.02359646    -0.03605758
##   engineSize_2.8 engineSize_3 engineSize_3.2 engineSize_3.5 engineSize_4
## 1        -0.01362089   -0.543414     -0.02153952    -0.00963098    -0.02359646
## 2        -0.01362089   -0.543414     -0.02153952    -0.00963098    -0.02359646
## 3        -0.01362089    1.840047     -0.02153952    -0.00963098    -0.02359646
## 4        -0.01362089   -0.543414     -0.02153952    -0.00963098    -0.02359646
## 5        -0.01362089    1.840047     -0.02153952    -0.00963098    -0.02359646
## 6        -0.01362089   -0.543414     -0.02153952    -0.00963098    -0.02359646
##   engineSize_4.4 engineSize_5 engineSize_6.6
## 1        -0.0859134   -0.00963098    -0.01362089
## 2        -0.0859134   -0.00963098    -0.01362089
## 3        -0.0859134   -0.00963098    -0.01362089
## 4        -0.0859134   -0.00963098    -0.01362089
## 5        -0.0859134   -0.00963098    -0.01362089
## 6        -0.0859134   -0.00963098    -0.01362089

```

Apply PCA on scale data set and choose appropriate number of components

```
fa.parallel(bmw_scale, fa = "pc", n.iter = 100, show.legend = TRUE )
```

```
## Warning in cor.smooth(R) : Matrix was not positive definite, smoothing was done
```

```
## In smc, smcs < 0 were set to .0
```

```
## Warning in cor.smooth(R) : Matrix was not positive definite, smoothing was done
```

```
## In smc, smcs < 0 were set to .0
```

```
## Warning in cor.smooth(R) : Matrix was not positive definite, smoothing was done
```

```
## In smc, smcs < 0 were set to .0
```

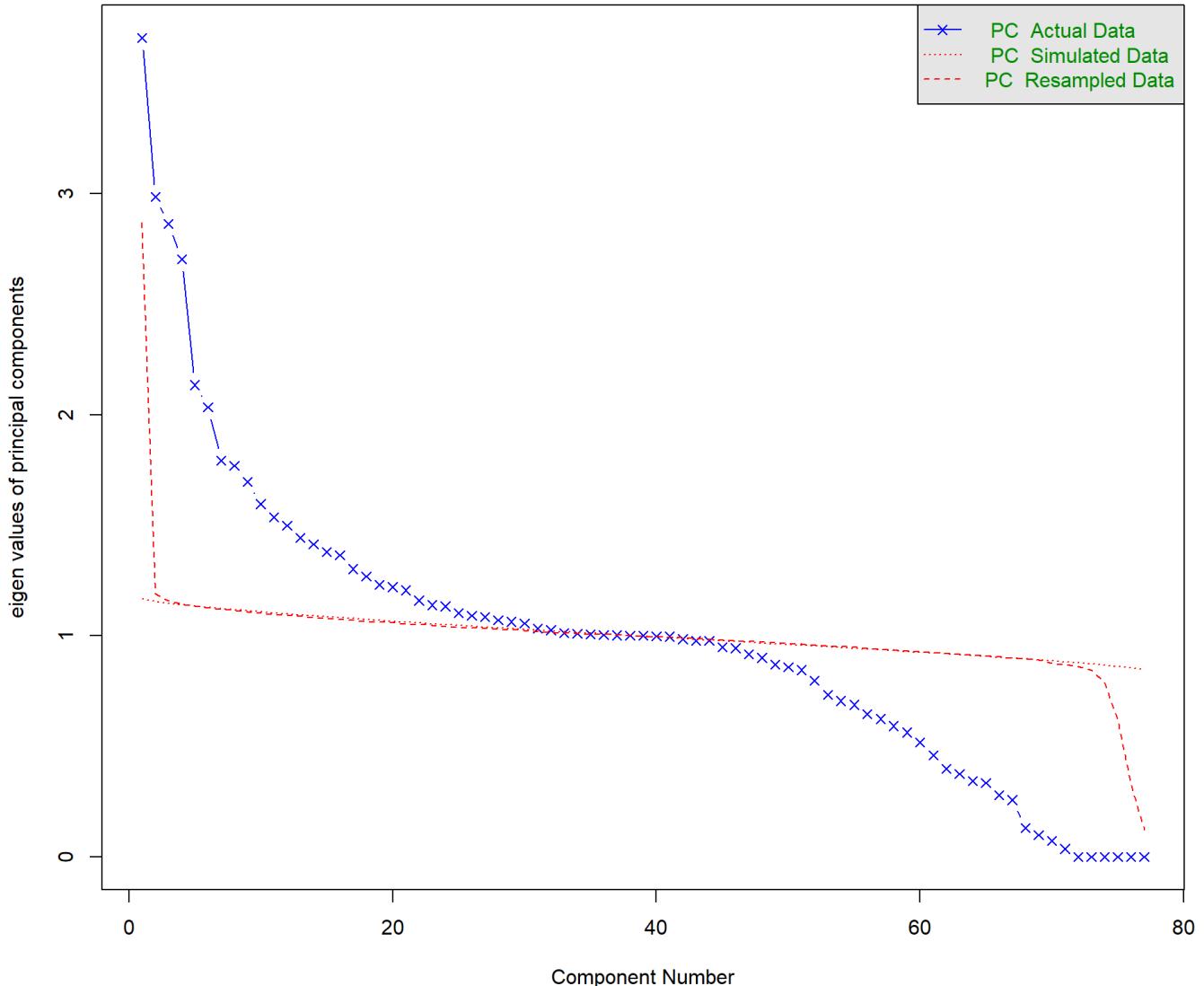
```
## Warning in cor.smooth(r) : Matrix was not positive definite, smoothing was done
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :  
## The estimated weights for the factor scores are probably incorrect. Try a  
## different factor score estimation method.
```

```
## In factor.scores, the correlation matrix is singular, an approximation is used
```

```
## Warning in cor.smooth(r) : Matrix was not positive definite, smoothing was done
```

Parallel Analysis Scree Plots



```
## Parallel analysis suggests that the number of factors = NA and the number of components = 31
```

Choose components number is 31 and create pca data frame for modeling part

```
# From the result from above , we choose components number = 31
pc = principal(bmw_scale, nfactors = 31 , rotate = "none", scores = TRUE )
```

```
## Warning in cor.smooth(r) : Matrix was not positive definite, smoothing was done
```

```
## Warning in principal(bmw_scale, nfactors = 31, rotate = "none", scores = TRUE):
## The matrix is not positive semi-definite, scores found from Structure loadings
```

```
pca_df = cbind(bmw_df$price , as.data.frame(pc$scores))%>%
  rename(price = "bmw_df$price")
head(pca_df)
```

```

##   price      PC1      PC2      PC3      PC4      PC5      PC6
## 1 11200 -2.7592364 1.14913095 -2.17292034 1.87097779 1.528301 0.9079794
## 2 27000 1.8180832 -0.09179234 0.97774878 -0.3148008 1.469252 0.8083152
## 3 16000 -0.3554297 1.50818980 -0.49584106 1.1338361 3.542515 -1.2293824
## 4 12750 -2.8365247 -2.07815817 0.85636937 -1.5179296 2.209667 -1.8333380
## 5 14500 1.1362679 1.72564242 -0.07781286 1.0412928 5.138287 -1.9136126
## 6 14900 -2.8138309 1.24804691 -1.83579681 1.6176241 1.191650 0.3376283
##          PC7      PC8      PC9      PC10     PC11     PC12
## 1 1.404097573 0.11109876 -0.062077428 -0.4826857 -0.2559175 -1.46104486
## 2 1.717857184 0.05421004 -0.201203966 -0.4117159 -1.7307477 -1.68791863
## 3 0.370981885 0.11795142 0.415964489 -0.8518581 0.3202836 -0.84255909
## 4 -0.006346546 0.02009673 0.048072346 -0.6488114 0.3539405 0.17301404
## 5 0.488422758 -0.17308463 -0.014521616 0.4829784 -0.9139513 -0.04204794
## 6 0.949574275 0.14690256 -0.009503119 -1.1888732 0.7980028 -1.15387691
##          PC13     PC14     PC15     PC16     PC17     PC18
## 1 0.24706139 -0.82797749 0.6267995 0.2042523 -1.0983395 -1.28641995
## 2 -0.28306846 -0.51480055 -2.0682258 -1.2563315 -1.6360491 -0.83127773
## 3 0.83413058 -0.14498030 -1.0706652 -0.5137191 -0.4426571 0.12430718
## 4 0.35106059 0.57869243 -0.5152109 -0.6336214 0.6115357 -1.64518090
## 5 -0.03380889 -0.78270925 -0.2302812 0.7905958 0.4208583 -1.94746398
## 6 0.75767803 0.01422884 -0.7737714 -0.4680817 -0.4709486 0.02817485
##          PC19     PC20     PC21     PC22     PC23     PC24
## 1 1.51904313 1.15983450 -0.7161846 1.16334234 0.4486476 0.38729159
## 2 -0.02182948 -0.53745825 -0.7506409 -0.20492583 -1.2182768 0.42077202
## 3 2.04968743 0.05289594 -0.4673725 0.12815320 0.5040385 1.08139440
## 4 -0.56399172 1.33375052 0.5212263 0.72865970 -0.5628010 0.05652515
## 5 -0.58184264 1.26007340 0.6492400 0.09520865 6.6620116 -1.69850285
## 6 2.23948558 0.12569335 -0.2512388 0.22205700 0.5757520 1.01389160
##          PC25     PC26     PC27     PC28     PC29     PC30     PC31
## 1 1.0650357 -0.2243290 1.3161403 -1.0915712 -1.4546581 -2.0956721 -2.1178922
## 2 2.9718762 0.6885659 2.4730448 4.8611117 2.3895521 -0.3418099 -0.1963465
## 3 1.1278403 -1.3599156 0.2982725 -0.9989323 0.3476459 0.2224365 -0.1649775
## 4 -1.0825040 0.1737113 0.1697216 0.1161050 0.8100524 -0.2007155 -0.2315985
## 5 2.4841038 1.8972341 0.3454195 -0.3095847 -1.7226081 -2.0659016 -1.4300980
## 6 0.9881169 -1.2213456 0.2713121 -0.8767083 0.3682526 0.3418935 -0.2304752

```

Split dataset as train and test data set

```
# split train and test data , which 80% train data and 20% test data
library(caTools)
```

```
## Warning: package 'caTools' was built under R version 4.0.5
```

```
pca_split_data = sample.split(pca_df, SplitRatio = 0.8)
pca_train_data = subset(pca_df, pca_split_data==TRUE)
pca_test_data = subset( pca_df, pca_split_data ==FALSE)
```

implementing model

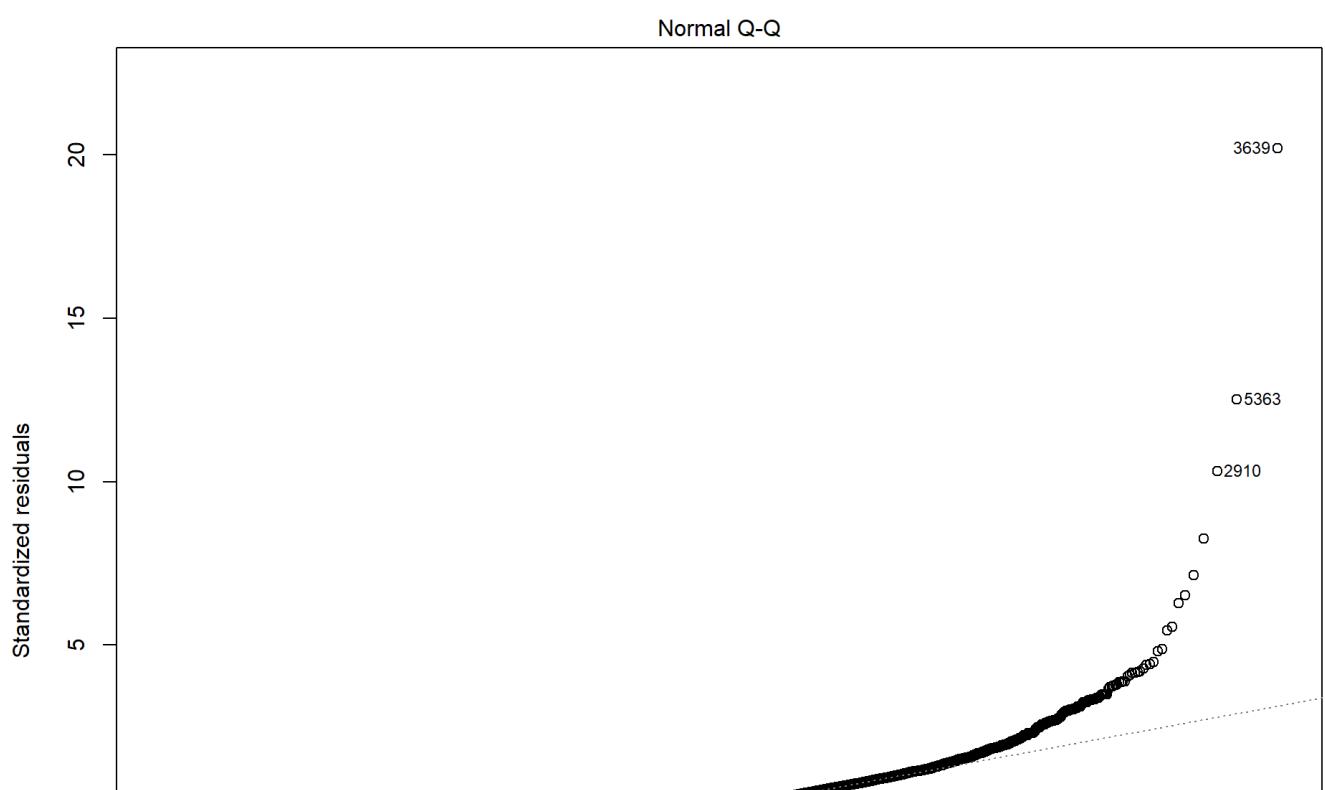
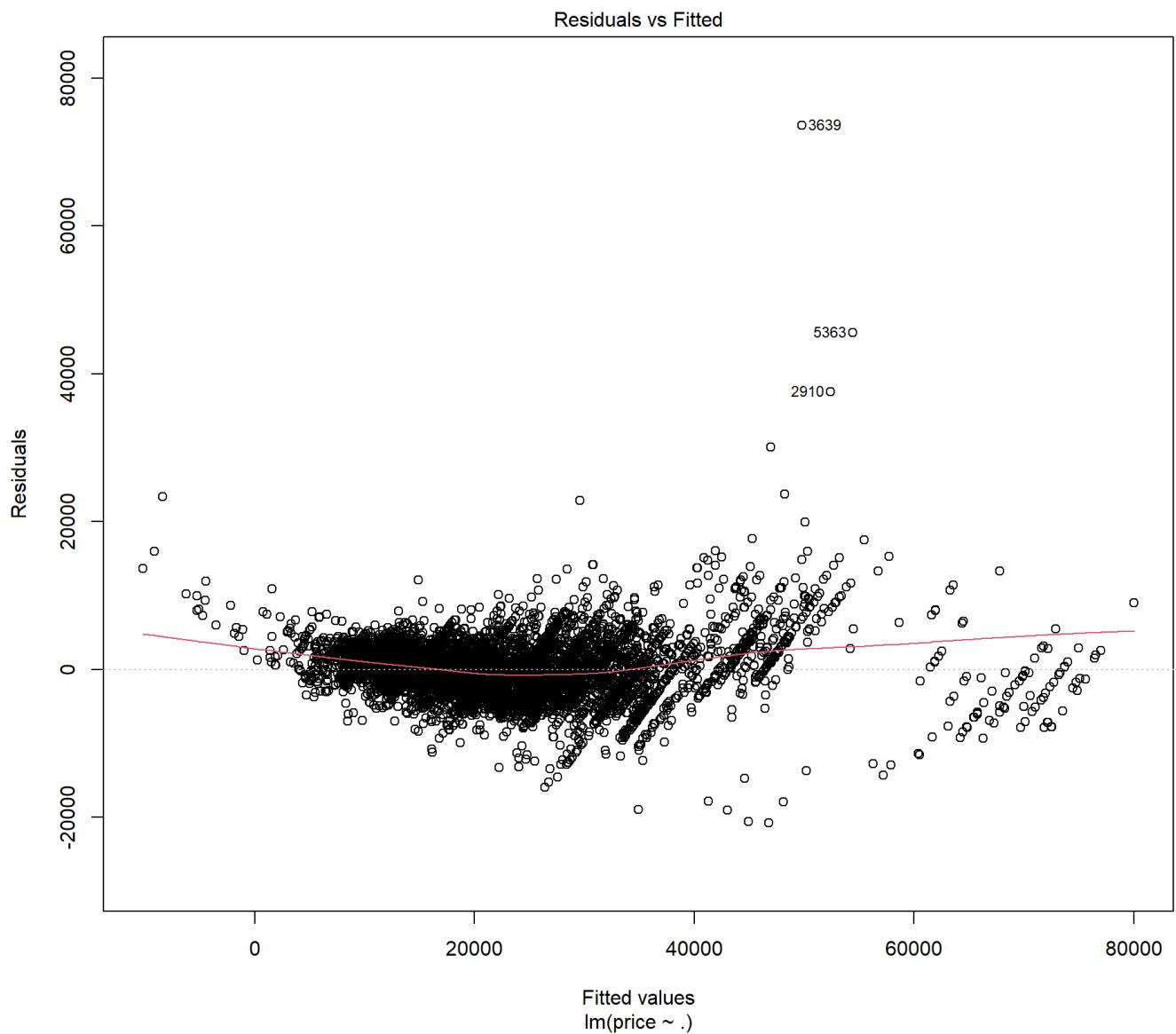
linear regression with PCA

```
## linear regression
linear = lm(price ~ ., data = pca_train_data)
summary(linear)
```

```
##
## Call:
## lm(formula = price ~ ., data = pca_train_data)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -20769  -2077     31    1956   73628 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 22737.80    39.77 571.682 < 2e-16 ***
## PC1         2124.82    10.83 196.247 < 2e-16 ***
## PC2         1766.52    13.43 131.515 < 2e-16 ***
## PC3         868.48     14.30 60.747 < 2e-16 ***
## PC4          55.16     14.62  3.772 0.000163 ***
## PC5        -676.06    18.69 -36.164 < 2e-16 ***
## PC6        -953.21    19.79 -48.164 < 2e-16 ***
## PC7        -648.42    21.97 -29.510 < 2e-16 ***
## PC8        -34.48     20.54 -1.678 0.093343 .
## PC9        -189.38    23.40 -8.093 6.62e-16 ***
## PC10       -489.52    40.16 -12.189 < 2e-16 ***
## PC11       1382.27    27.53 50.219 < 2e-16 ***
## PC12       319.12     30.81 10.359 < 2e-16 ***
## PC13       -50.47    26.20 -1.926 0.054140 .
## PC14       103.12     29.97  3.440 0.000584 ***
## PC15       900.41     29.68 30.341 < 2e-16 ***
## PC16       600.78     29.73 20.206 < 2e-16 ***
## PC17       226.99     30.58  7.423 1.26e-13 ***
## PC18       619.40     32.29 19.185 < 2e-16 ***
## PC19       544.67     32.32 16.850 < 2e-16 ***
## PC20       365.94     34.76 10.529 < 2e-16 ***
## PC21      1293.34     33.67 38.409 < 2e-16 ***
## PC22      -145.73     33.92 -4.297 1.75e-05 ***
## PC23        42.33     37.66  1.124 0.261097
## PC24       143.17     35.54  4.028 5.66e-05 ***
## PC25      -1039.04    36.20 -28.704 < 2e-16 ***
## PC26       686.64     36.51 18.805 < 2e-16 ***
## PC27      -11.53     36.66 -0.315 0.753064
## PC28       180.31     36.96  4.879 1.09e-06 ***
## PC29      -95.96     37.13 -2.584 0.009773 **
## PC30       569.36     37.91 15.019 < 2e-16 ***
## PC31      -330.95    37.44 -8.839 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

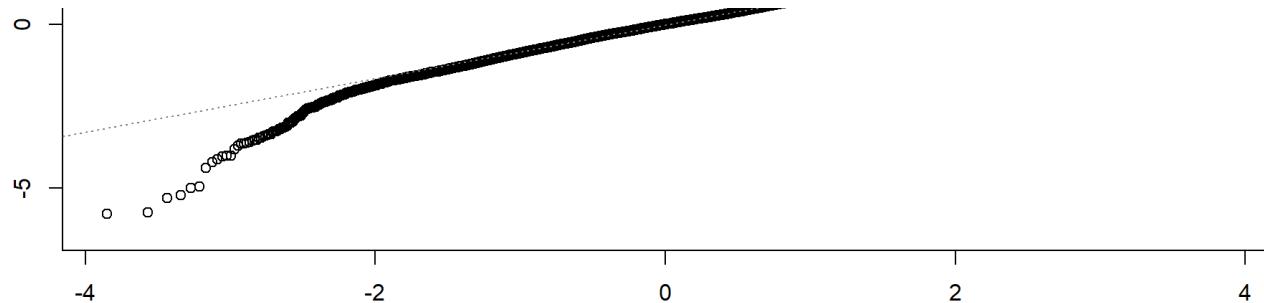
Residual standard error: 3649 on 8390 degrees of freedom
Multiple R-squared: 0.897, Adjusted R-squared: 0.8966
F-statistic: 2356 on 31 and 8390 DF, p-value: < 2.2e-16

```
plot(linear)
```

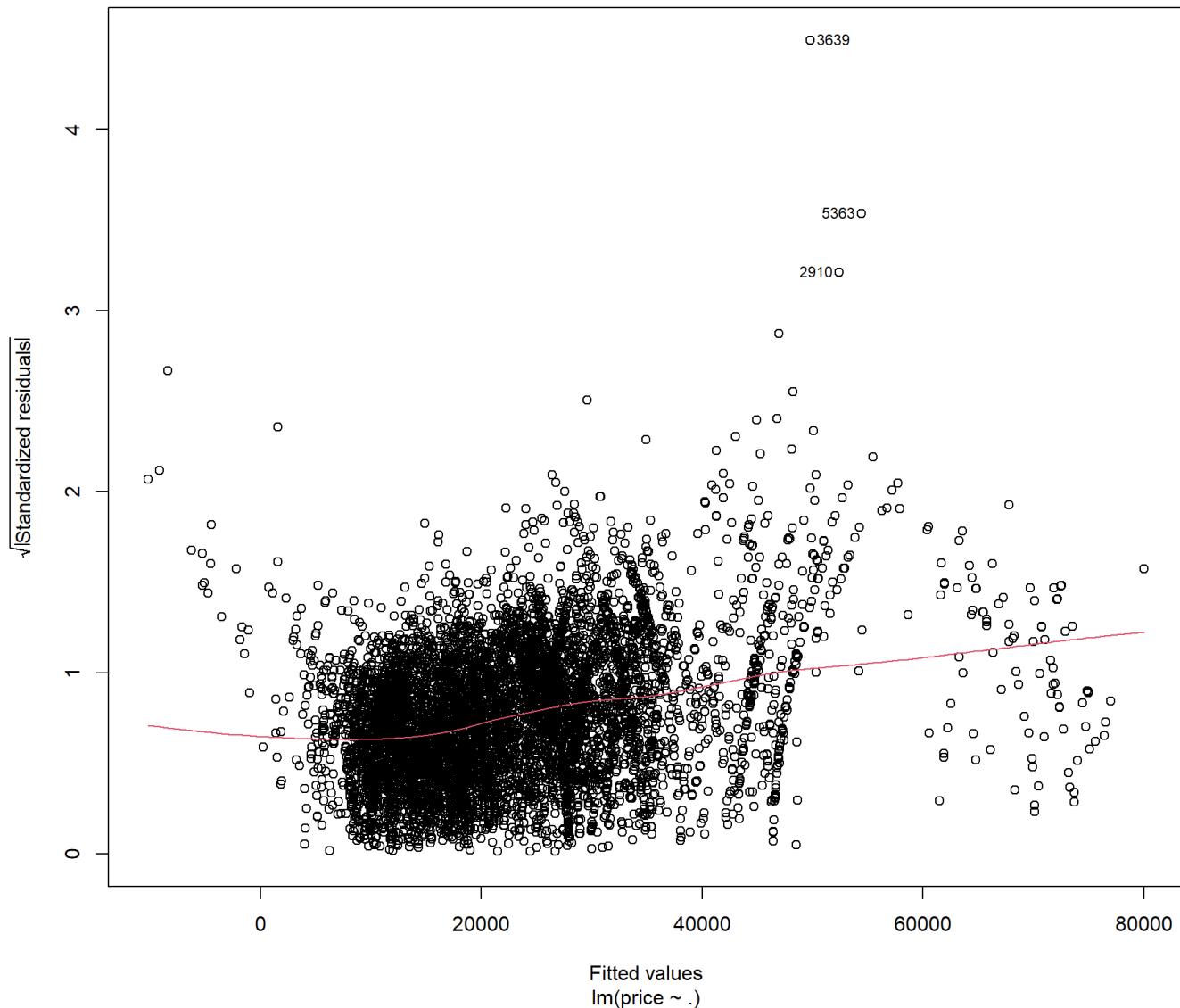
2021/4/24

Car Price Prediction

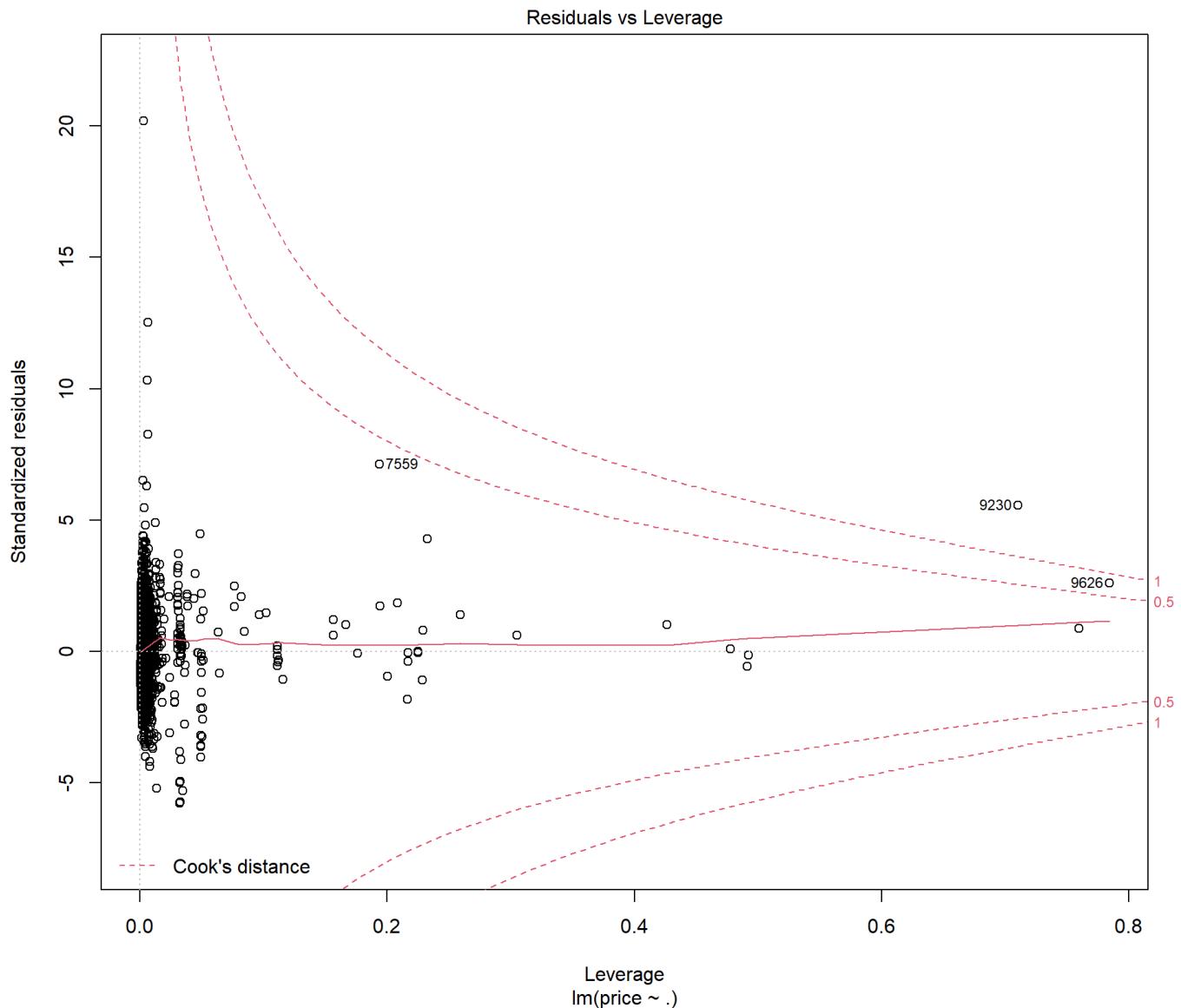


Theoretical Quantiles
 $\text{Im}(\text{price} \sim .)$

Scale-Location



Fitted values
 $\text{Im}(\text{price} \sim .)$

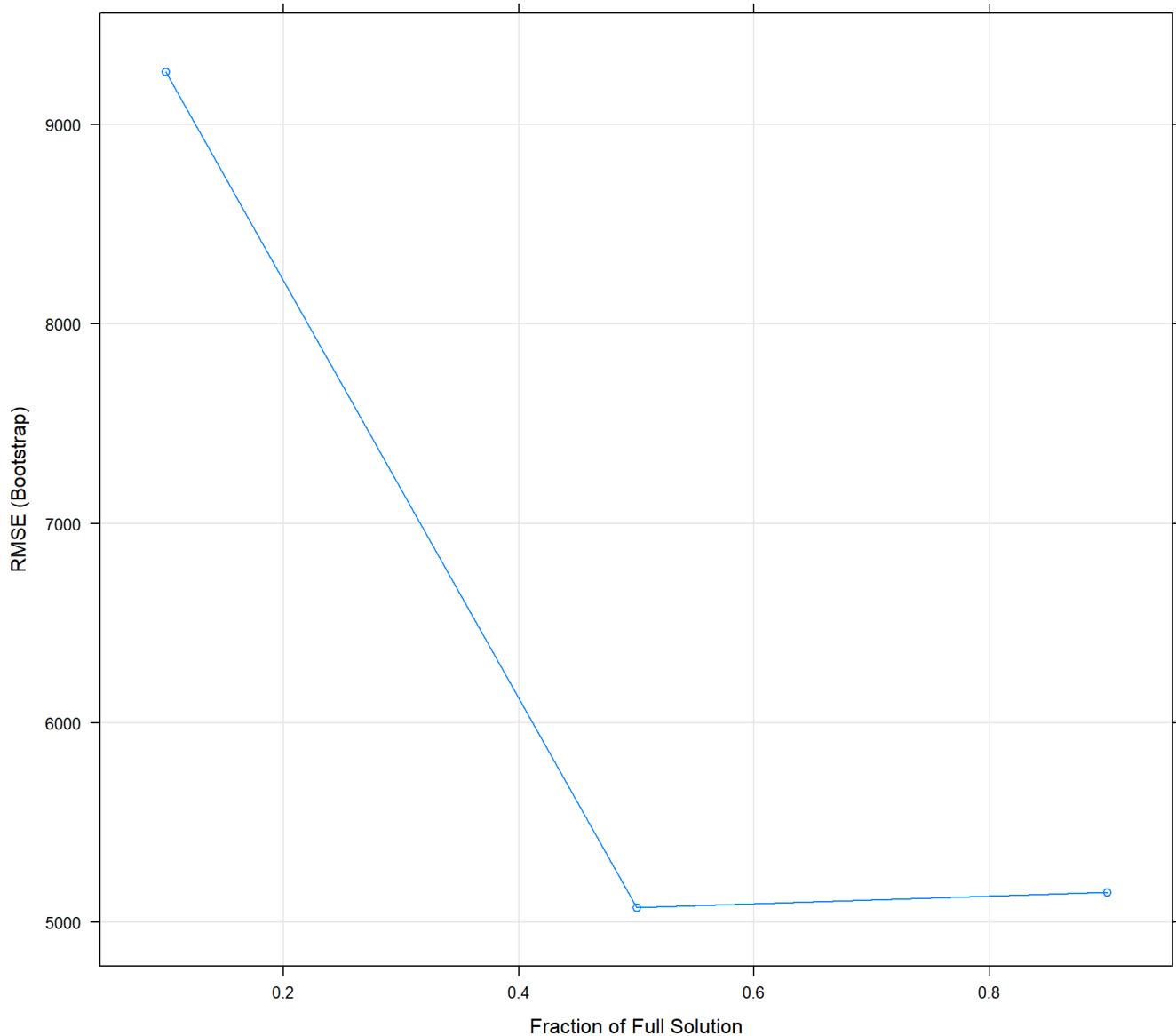


Lasso Regression with PCA

```
## Lasso Regression
lasso_pca <- train(price ~ . ,
                     data = pca_train_data,
                     method = "lasso")
lasso_pca
```

```
## The lasso
##
## 8422 samples
##   31 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 8422, 8422, 8422, 8422, 8422, 8422, ...
## Resampling results across tuning parameters:
##
##   fraction  RMSE      Rsquared     MAE
##   0.1        9266.520  0.5580142  6736.382
##   0.5        5073.275  0.8236085  3533.407
##   0.9        5151.493  0.8585858  2688.957
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was fraction = 0.5.
```

```
plot(lasso_pca)
```

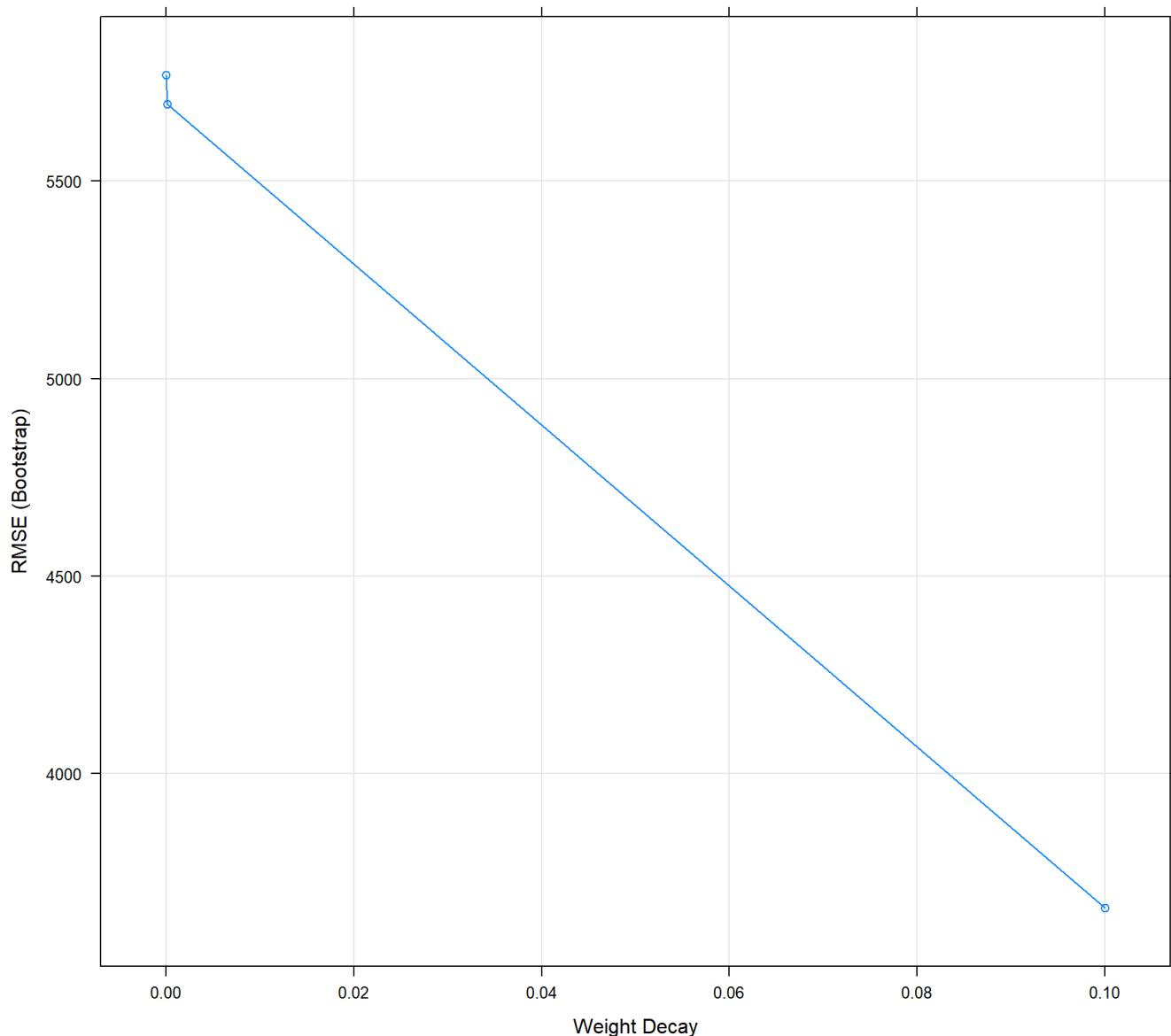


Ridge Regression with PCA

```
## ridge regression
ridge_pca <- train(price ~ . ,
                     data = pca_train_data,
                     method = "ridge")
ridge_pca
```

```
## Ridge Regression
##
## 8422 samples
##   31 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 8422, 8422, 8422, 8422, 8422, 8422, ...
## Resampling results across tuning parameters:
##
##   lambda   RMSE     Rsquared    MAE
##   0e+00    5769.355  0.8297025  2678.747
##   1e-04    5695.735  0.8303368  2676.700
##   1e-01    3659.588  0.8951695  2613.910
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was lambda = 0.1.
```

```
plot(ridge_pca)
```

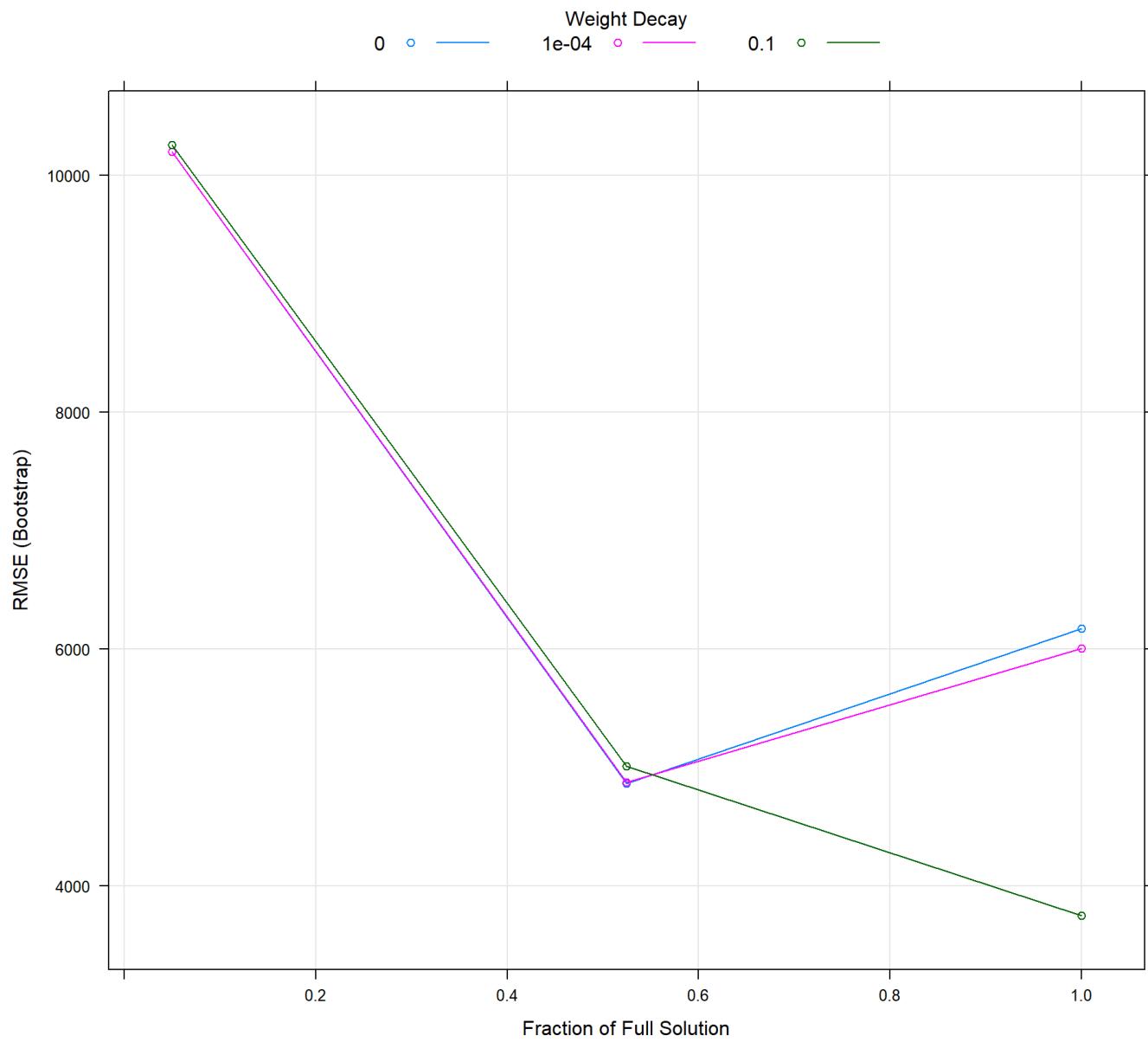


Elasticnet with PCA

```
## Elasticnet
enet_pca <- train(price ~ . ,
                     data = pca_train_data,
                     method = "enet")
enet_pca
```

```
## Elasticnet
##
## 8422 samples
##   31 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 8422, 8422, 8422, 8422, 8422, ...
## Resampling results across tuning parameters:
##
##     lambda  fraction    RMSE      Rsquared      MAE
## 0e+00    0.050    10198.447  0.4622664  7482.738
## 0e+00    0.525    4867.566  0.8348269  3395.957
## 0e+00    1.000    6172.367  0.8051181  2732.327
## 1e-04    0.050    10202.661  0.4613966  7486.154
## 1e-04    0.525    4874.250  0.8345093  3400.482
## 1e-04    1.000    6009.335  0.8067995  2725.737
## 1e-01    0.050    10260.675  0.4600051  7532.487
## 1e-01    0.525    5011.316  0.8272891  3490.770
## 1e-01    1.000    3749.853  0.8907413  2629.293
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were fraction = 1 and lambda = 0.1.
```

```
plot(enet_pca)
```

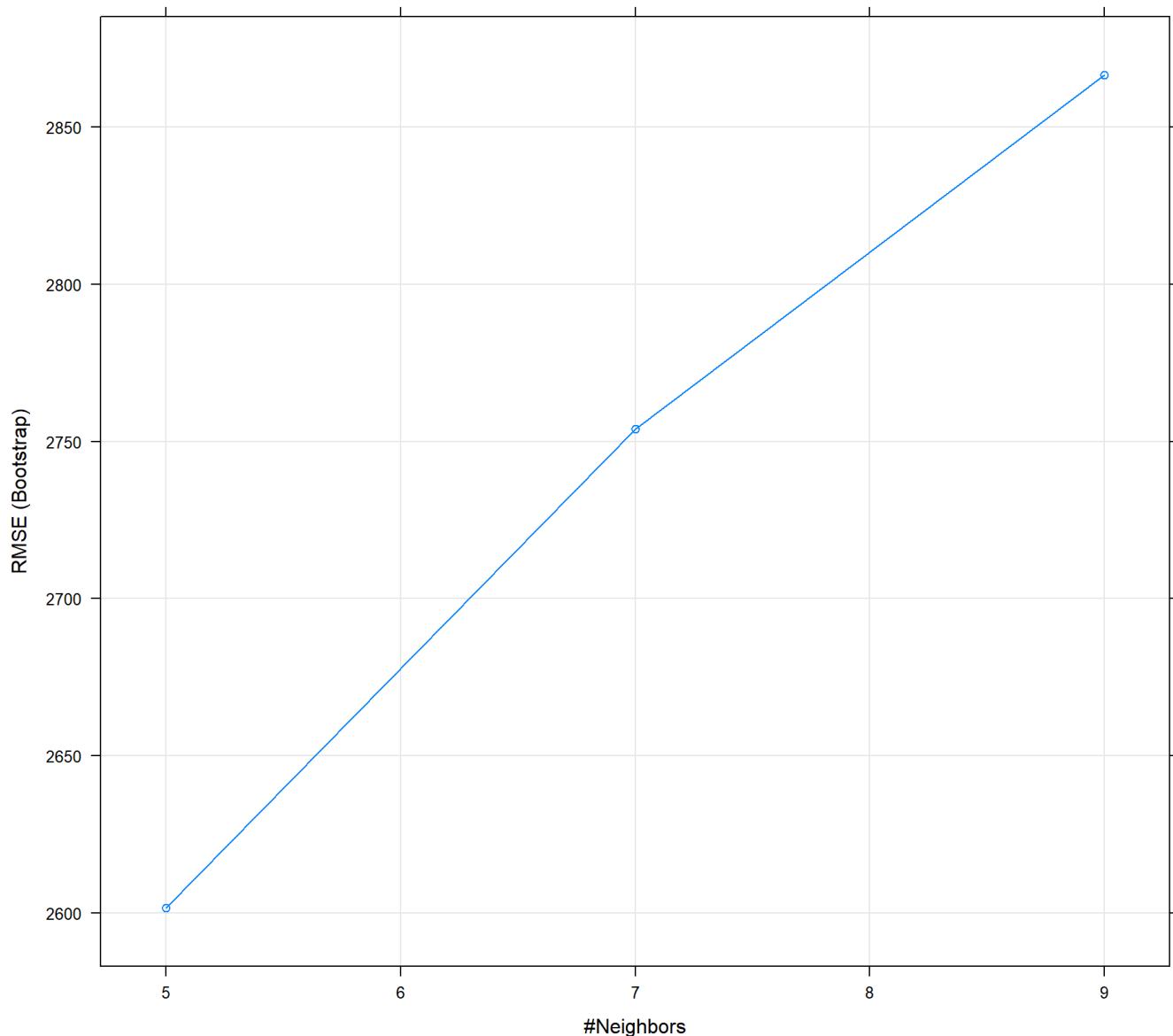


KNN regression with PCA

```
## KNN regression
knn_pca <- train(price ~ . ,
                    data = pca_train_data,
                    method = "knn")
knn_pca
```

```
## k-Nearest Neighbors
##
## 8422 samples
##   31 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 8422, 8422, 8422, 8422, 8422, 8422, ...
## Resampling results across tuning parameters:
##
##     k    RMSE      Rsquared     MAE
##     5   2601.653  0.9470779  1231.675
##     7   2753.984  0.9408119  1348.401
##     9   2866.505  0.9359094  1446.883
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was k = 5.
```

```
plot(knn_pca)
```



XGBoost with PCA

```
## xgboost
xgboost_pca<- train(price ~ . ,
                      data = pca_train_data,
                      method = "xgbTree")
```



```
eprecated in favor of reg:squarederror.  
## [01:40:49] WARNING: amalgamation/../src/objective/regression_obj.cu:170: reg:linear is now d  
eprecated in favor of reg:squarederror.  
## [01:40:50] WARNING: amalgamation/../src/objective/regression_obj.cu:170: reg:linear is now d  
eprecated in favor of reg:squarederror.
```

```
xgboost_pca
```

```

## eXtreme Gradient Boosting
##
## 8422 samples
##   31 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 8422, 8422, 8422, 8422, 8422, 8422, ...
## Resampling results across tuning parameters:

##     eta  max_depth  colsample_bytree  subsample  nrounds    RMSE    Rsquared
## 0.3    1          0.6              0.50       50      4000.314  0.8814553
## 0.3    1          0.6              0.50       100     3455.125  0.9093303
## 0.3    1          0.6              0.50       150     3226.293  0.9205794
## 0.3    1          0.6              0.75       50      3991.132  0.8826993
## 0.3    1          0.6              0.75       100     3451.972  0.9099624
## 0.3    1          0.6              0.75       150     3223.180  0.9209560
## 0.3    1          0.6              1.00       50      4000.677  0.8823141
## 0.3    1          0.6              1.00       100     3471.295  0.9090026
## 0.3    1          0.6              1.00       150     3248.282  0.9197856
## 0.3    1          0.8              0.50       50      3931.094  0.8856262
## 0.3    1          0.8              0.50       100     3413.038  0.9115041
## 0.3    1          0.8              0.50       150     3201.082  0.9218638
## 0.3    1          0.8              0.75       50      3909.941  0.8876948
## 0.3    1          0.8              0.75       100     3400.621  0.9123729
## 0.3    1          0.8              0.75       150     3187.636  0.9225545
## 0.3    1          0.8              1.00       50      3952.974  0.8854365
## 0.3    1          0.8              1.00       100     3435.826  0.9108969
## 0.3    1          0.8              1.00       150     3214.600  0.9214072
## 0.3    2          0.6              0.50       50      3160.332  0.9237883
## 0.3    2          0.6              0.50       100     2771.427  0.9411186
## 0.3    2          0.6              0.50       150     2583.284  0.9487897
## 0.3    2          0.6              0.75       50      3149.901  0.9244698
## 0.3    2          0.6              0.75       100     2745.525  0.9423036
## 0.3    2          0.6              0.75       150     2558.333  0.9498450
## 0.3    2          0.6              1.00       50      3124.366  0.9257031
## 0.3    2          0.6              1.00       100     2749.328  0.9421131
## 0.3    2          0.6              1.00       150     2565.262  0.9494999
## 0.3    2          0.8              0.50       50      3067.798  0.9282311
## 0.3    2          0.8              0.50       100     2705.059  0.9439552
## 0.3    2          0.8              0.50       150     2521.339  0.9512095
## 0.3    2          0.8              0.75       50      3068.857  0.9282807
## 0.3    2          0.8              0.75       100     2695.007  0.9444175
## 0.3    2          0.8              0.75       150     2511.684  0.9516162
## 0.3    2          0.8              1.00       50      3064.210  0.9284703
## 0.3    2          0.8              1.00       100     2696.732  0.9443346
## 0.3    2          0.8              1.00       150     2520.406  0.9512924
## 0.3    3          0.6              0.50       50      2739.017  0.9425005
## 0.3    3          0.6              0.50       100     2418.999  0.9550465
## 0.3    3          0.6              0.50       150     2260.642  0.9606662
## 0.3    3          0.6              0.75       50      2672.999  0.9453732
## 0.3    3          0.6              0.75       100     2356.553  0.9573862
## 0.3    3          0.6              0.75       150     2210.628  0.9624330
## 0.3    3          0.6              1.00       50      2688.899  0.9447360
## 0.3    3          0.6              1.00       100     2371.003  0.9568787
## 0.3    3          0.6              1.00       150     2223.311  0.9620247
## 0.3    3          0.8              0.50       50      2672.577  0.9452045

```

Car Price Prediction

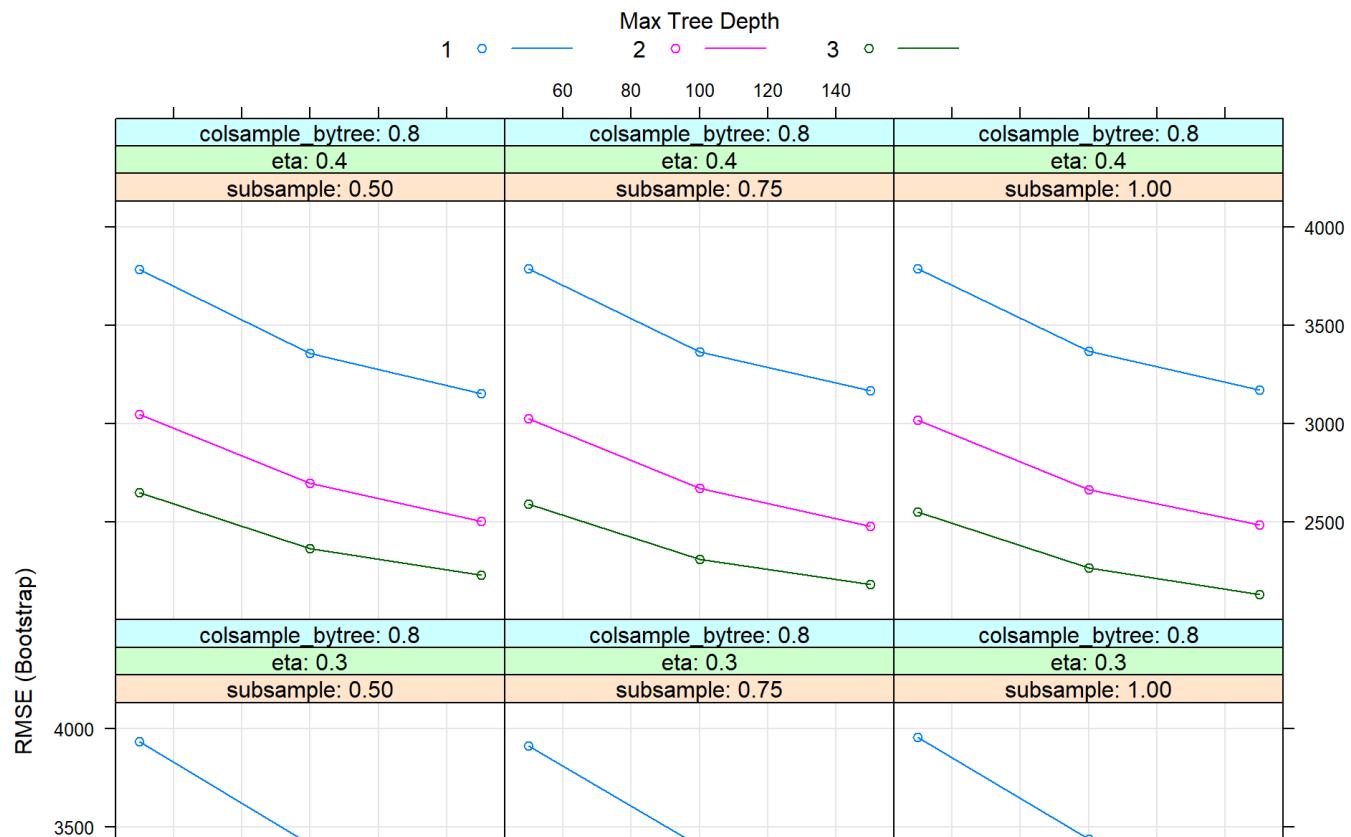
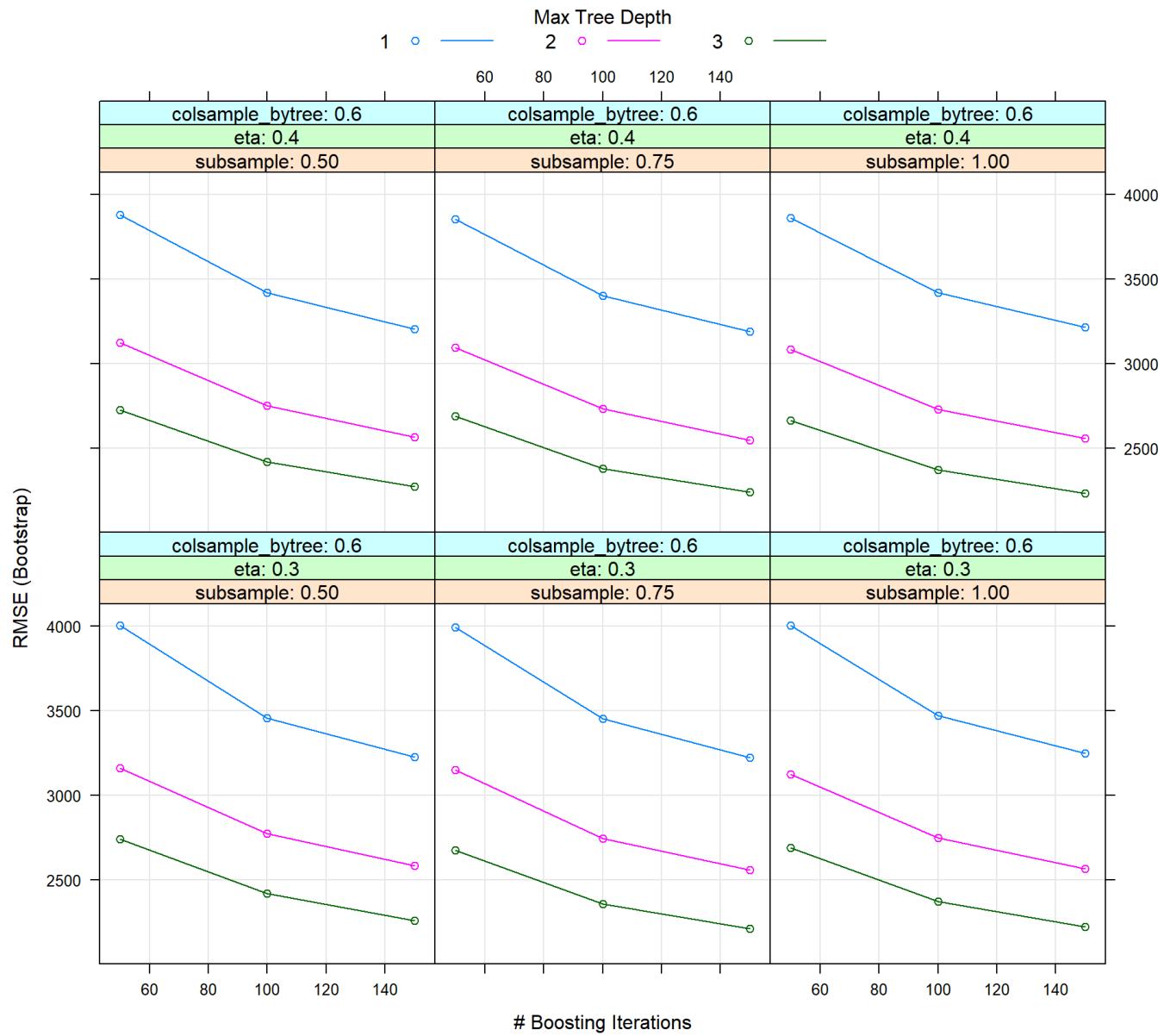
##	0.3	3	0.8	0.50	100	2369.483	0.9567580
##	0.3	3	0.8	0.50	150	2224.292	0.9617979
##	0.3	3	0.8	0.75	50	2603.437	0.9481051
##	0.3	3	0.8	0.75	100	2300.576	0.9593708
##	0.3	3	0.8	0.75	150	2154.787	0.9642752
##	0.3	3	0.8	1.00	50	2599.214	0.9482755
##	0.3	3	0.8	1.00	100	2300.546	0.9593499
##	0.3	3	0.8	1.00	150	2157.883	0.9641697
##	0.4	1	0.6	0.50	50	3877.630	0.8857341
##	0.4	1	0.6	0.50	100	3418.725	0.9106084
##	0.4	1	0.6	0.50	150	3203.659	0.9213972
##	0.4	1	0.6	0.75	50	3853.192	0.8877139
##	0.4	1	0.6	0.75	100	3400.845	0.9116940
##	0.4	1	0.6	0.75	150	3187.859	0.9223218
##	0.4	1	0.6	1.00	50	3860.475	0.8872869
##	0.4	1	0.6	1.00	100	3418.849	0.9108055
##	0.4	1	0.6	1.00	150	3215.040	0.9209766
##	0.4	1	0.8	0.50	50	3785.447	0.8912139
##	0.4	1	0.8	0.50	100	3357.863	0.9138576
##	0.4	1	0.8	0.50	150	3154.579	0.9239156
##	0.4	1	0.8	0.75	50	3788.590	0.8914286
##	0.4	1	0.8	0.75	100	3364.159	0.9135638
##	0.4	1	0.8	0.75	150	3166.043	0.9233151
##	0.4	1	0.8	1.00	50	3785.996	0.8917166
##	0.4	1	0.8	1.00	100	3368.930	0.9134391
##	0.4	1	0.8	1.00	150	3172.361	0.9231180
##	0.4	2	0.6	0.50	50	3124.975	0.9251930
##	0.4	2	0.6	0.50	100	2750.979	0.9419818
##	0.4	2	0.6	0.50	150	2564.426	0.9495024
##	0.4	2	0.6	0.75	50	3093.382	0.9267050
##	0.4	2	0.6	0.75	100	2734.842	0.9426101
##	0.4	2	0.6	0.75	150	2545.865	0.9502124
##	0.4	2	0.6	1.00	50	3084.982	0.9269206
##	0.4	2	0.6	1.00	100	2729.117	0.9426885
##	0.4	2	0.6	1.00	150	2558.090	0.9495869
##	0.4	2	0.8	0.50	50	3045.990	0.9289865
##	0.4	2	0.8	0.50	100	2696.624	0.9442211
##	0.4	2	0.8	0.50	150	2503.150	0.9518984
##	0.4	2	0.8	0.75	50	3023.714	0.9299916
##	0.4	2	0.8	0.75	100	2670.428	0.9453214
##	0.4	2	0.8	0.75	150	2480.137	0.9527555
##	0.4	2	0.8	1.00	50	3019.847	0.9302364
##	0.4	2	0.8	1.00	100	2663.763	0.9455966
##	0.4	2	0.8	1.00	150	2487.186	0.9524967
##	0.4	3	0.6	0.50	50	2724.786	0.9429382
##	0.4	3	0.6	0.50	100	2418.820	0.9549461
##	0.4	3	0.6	0.50	150	2275.699	0.9600475
##	0.4	3	0.6	0.75	50	2691.390	0.9445324
##	0.4	3	0.6	0.75	100	2380.697	0.9565507
##	0.4	3	0.6	0.75	150	2240.595	0.9614770
##	0.4	3	0.6	1.00	50	2665.395	0.9454794
##	0.4	3	0.6	1.00	100	2374.388	0.9566143
##	0.4	3	0.6	1.00	150	2232.598	0.9615750
##	0.4	3	0.8	0.50	50	2649.201	0.9461128
##	0.4	3	0.8	0.50	100	2366.767	0.9568566
##	0.4	3	0.8	0.50	150	2230.412	0.9615915
##	0.4	3	0.8	0.75	50	2591.729	0.9484363
##	0.4	3	0.8	0.75	100	2311.249	0.9588582

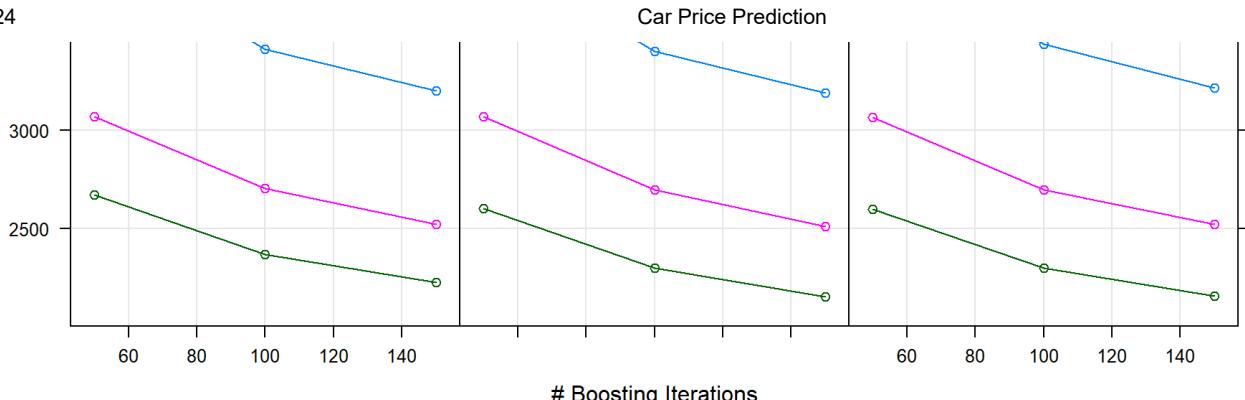
##	0.4	3	0.8	0.75	150	2184.521	0.9631652
##	0.4	3	0.8	1.00	50	2552.596	0.9499950
##	0.4	3	0.8	1.00	100	2265.974	0.9605302
##	0.4	3	0.8	1.00	150	2133.722	0.9649395
##	MAE						
##	2731.191						
##	2331.978						
##	2161.434						
##	2725.369						
##	2328.208						
##	2156.664						
##	2716.496						
##	2326.513						
##	2158.342						
##	2692.286						
##	2309.899						
##	2146.894						
##	2650.725						
##	2280.243						
##	2118.513						
##	2673.842						
##	2292.994						
##	2131.055						
##	2123.558						
##	1818.563						
##	1665.808						
##	2110.962						
##	1804.210						
##	1647.656						
##	2087.112						
##	1789.867						
##	1637.275						
##	2054.750						
##	1774.613						
##	1627.876						
##	2060.680						
##	1772.835						
##	1618.163						
##	2042.768						
##	1757.318						
##	1609.105						
##	1789.820						
##	1526.156						
##	1375.814						
##	1737.375						
##	1471.830						
##	1331.762						
##	1737.656						
##	1468.109						
##	1328.888						
##	1732.734						
##	1477.260						
##	1338.168						
##	1701.113						
##	1443.787						
##	1303.303						
##	1679.860						
##	1423.908						

```
## 1289.677
## 2665.204
## 2328.315
## 2164.567
## 2641.356
## 2300.574
## 2137.957
## 2622.489
## 2294.335
## 2139.198
## 2599.069
## 2274.807
## 2117.690
## 2575.368
## 2258.177
## 2104.808
## 2570.391
## 2257.056
## 2110.593
## 2107.425
## 1806.807
## 1646.232
## 2062.853
## 1771.151
## 1610.460
## 2060.860
## 1757.618
## 1604.168
## 2047.645
## 1768.827
## 1609.973
## 2020.215
## 1731.288
## 1571.416
## 2015.756
## 1723.474
## 1568.109
## 1758.620
## 1491.499
## 1348.936
## 1735.540
## 1462.506
## 1318.805
## 1712.339
## 1446.298
## 1304.778
## 1720.370
## 1459.392
## 1323.781
## 1666.442
## 1414.754
## 1275.816
## 1644.426
## 1386.288
## 1254.251
##
## Tuning parameter 'gamma' was held constant at a value of 0
## Tuning
```

```
## parameter 'min_child_weight' was held constant at a value of 1  
## RMSE was used to select the optimal model using the smallest value.  
## The final values used for the model were nrounds = 150, max_depth = 3, eta  
## = 0.4, gamma = 0, colsample_bytree = 0.8, min_child_weight = 1 and subsample  
## = 1.
```

```
plot(xgboost_pca)
```





Model Evaluation

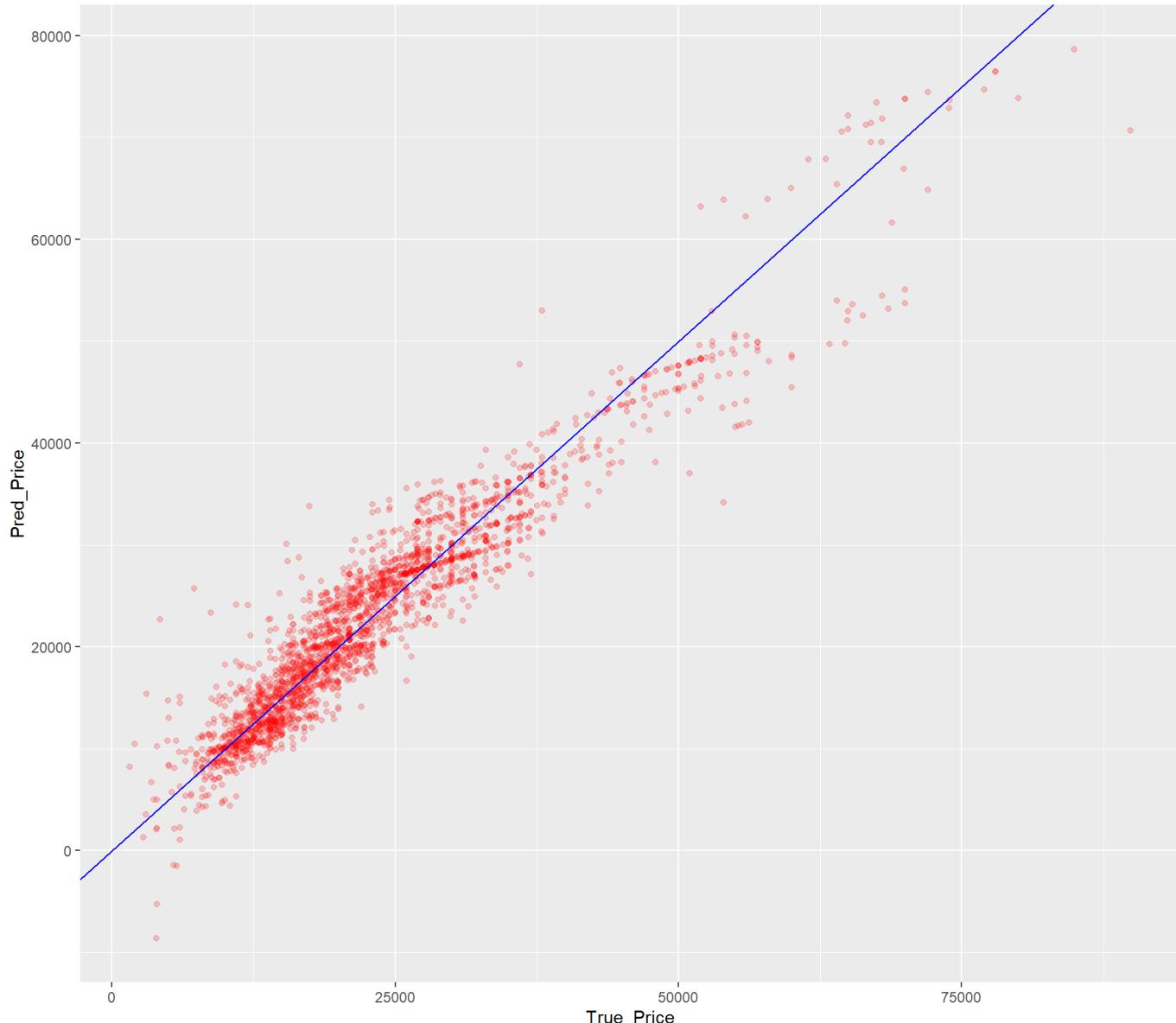
Linear Regression on test set

```
## RMSE, R2 and MAE in test data set of Linear Regression
linear_pca_pred <- predict(linear, pca_test_data)
as.table(postResample(pred = linear_pca_pred, obs = pca_test_data$price))
```

```
##           RMSE      Rsquared        MAE
## 3535.5402315  0.9081011 2583.9095914
```

```
ggplot() + geom_point(aes(x = pca_test_data$price , y = linear_pca_pred), color = "red" ,alpha =0.2) + xlab("True_Price") + ylab("Pred_Price") + ggtitle("The scatter plot of true price and predict price(Linear Regression)")+geom_abline (slope= 1, color = "blue")
```

The scatter plot of true price and predict price(Linear Regression)



Lasso Regression on test set

```
## RMSE, R2 and MAE in test data set of Linear Regression
lasso_pca_pred <- predict(lasso_pca, pca_test_data)
as.table(postResample(pred = lasso_pca_pred, obs = pca_test_data$price))
```

```
##           RMSE      Rsquared        MAE
## 5256.0043747 0.8292439 3593.1494320
```

```
ggplot() + geom_point(aes(x = pca_test_data$price , y = lasso_pca_pred), color = "red" ,alpha = 0.2) + xlab("True_Price") + ylab("Pred_Price") + ggtitle("The scatter plot of true price and predict price(Lasso Regression)")+geom_abline (slope= 1, color = "blue")
```

The scatter plot of true price and predict price(Lasso Regression)



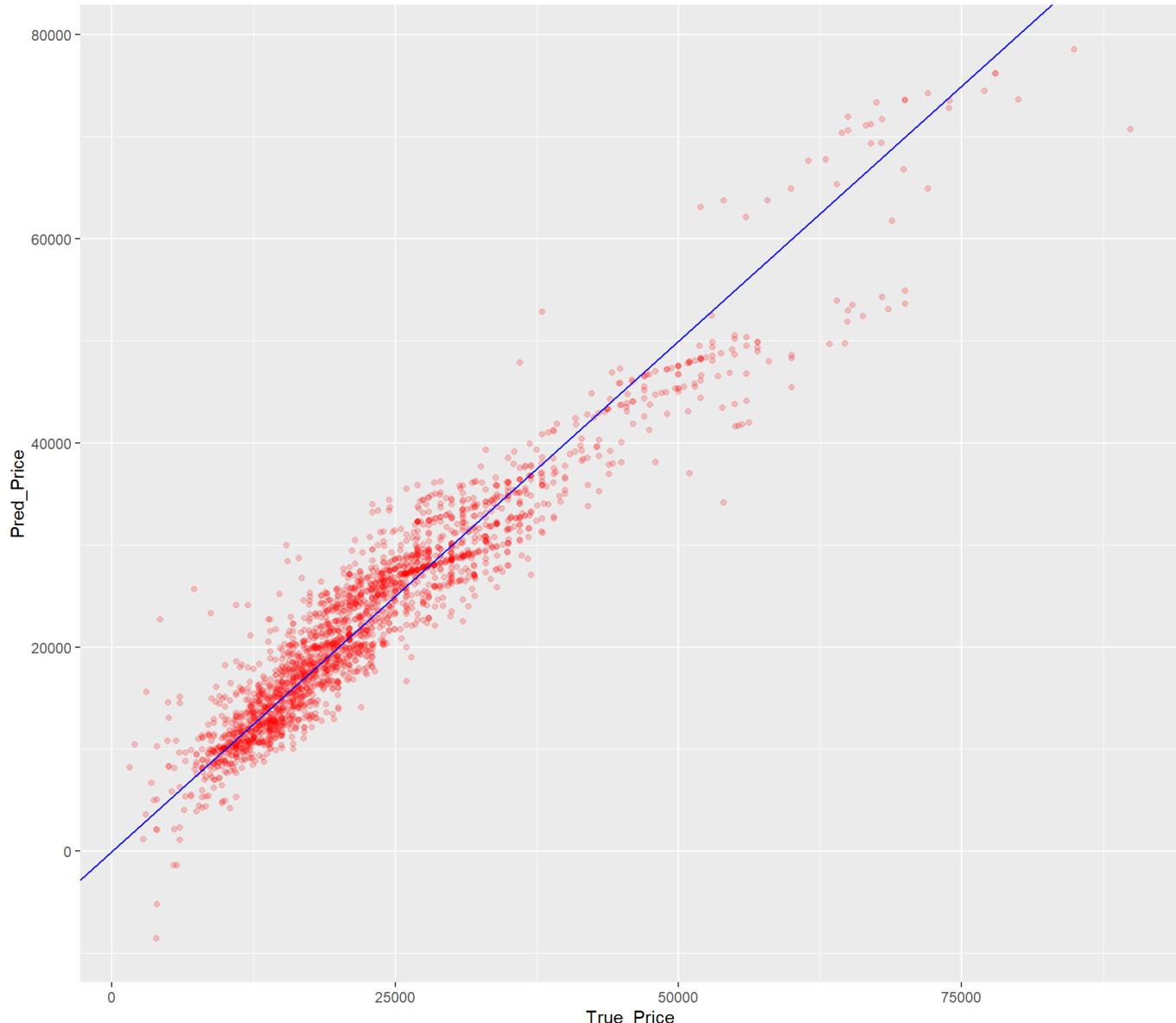
Ridge Regression on test set

```
## RMSE, R2 and MAE in test data set of Ridge Regression
ridge_pca_pred <- predict(ridge_pca, pca_test_data)
as.table(postResample(pred = ridge_pca_pred, obs = pca_test_data$price))
```

```
##           RMSE      Rsquared        MAE
## 3521.6345848  0.9088323 2578.9292709
```

```
ggplot() + geom_point(aes(x = pca_test_data$price , y = ridge_pca_pred), color = "red" ,alpha = 0.2) + xlab("True_Price") + ylab("Pred_Price") + ggtitle("The scatter plot of true price and predict price(Ridge Regression)")+geom_abline (slope= 1, color = "blue")
```

The scatter plot of true price and predict price(Ridge Regression)



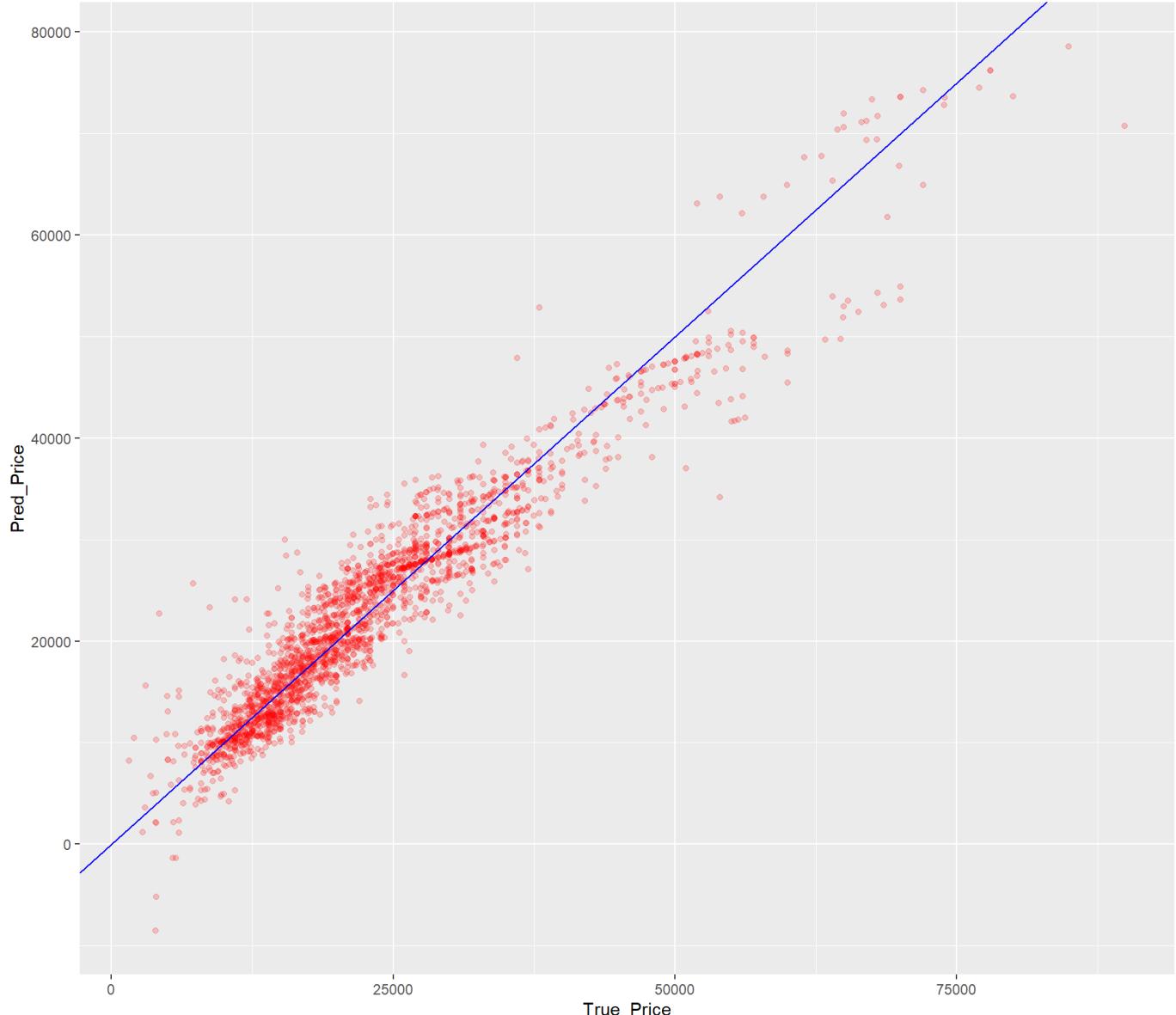
Elasticnet on test set

```
## RMSE, R2 and MAE in test data set of Elastic net
enet_pca_pred <- predict(enet_pca, pca_test_data)
as.table(postResample(pred = enet_pca_pred, obs = pca_test_data$price))
```

```
##           RMSE      Rsquared        MAE
## 3521.6345848  0.9088323 2578.9292709
```

```
ggplot() + geom_point(aes(x = pca_test_data$price , y = enet_pca_pred), color = "red" ,alpha = 0.2) + xlab("True_Price") + ylab("Pred_Price") + ggtitle("The scatter plot of true price and predict price(Elastic net Regression)")+geom_abline (slope= 1, color = "blue")
```

The scatter plot of true price and predict price(Elastic net Regression)



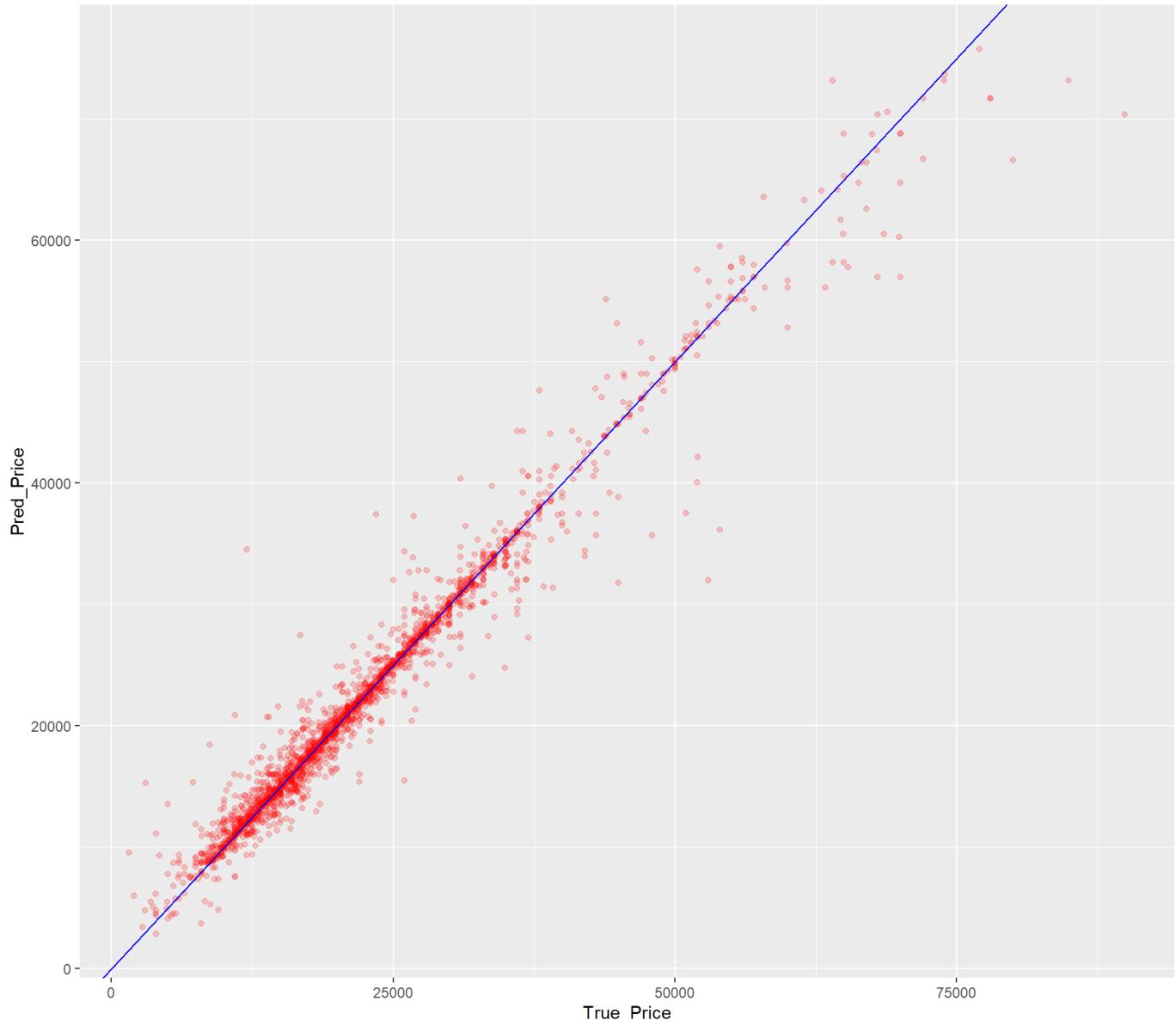
KNN Regression on test set

```
## RMSE, R2 and MAE in test data set
knn_pca_pred <- predict(knn_pca, pca_test_data)
as.table(postResample(pred = knn_pca_pred, obs = pca_test_data$price))
```

```
##           RMSE      Rsquared        MAE
## 2107.9561276  0.9675403 1056.1253734
```

```
ggplot() + geom_point(aes(x = pca_test_data$price , y = knn_pca_pred), color = "red" ,alpha =0.2) + xlab("True_Price") + ylab("Pred_Price") + ggtitle("The scatter plot of true price and predict price(KNN Regression)")+geom_abline (slope= 1, color = "blue")
```

The scatter plot of true price and predict price(KNN Regression)



XGboost on test set

```
## xgboost evaluation
xgboost_pca_pred <- predict(xgboost_pca, pca_test_data)
as.table(postResample(pred = xgboost_pca_pred, obs = pca_test_data$price))
```

```
##           RMSE      Rsquared        MAE
## 1848.3558484 0.9749145 1179.9157517
```

```
ggplot() + geom_point(aes(x = pca_test_data$price , y = xgboost_pca_pred), color = "red" ,alpha =0.2) + xlab("True_Price") + ylab("Pred_Price") + ggtitle("The scatter plot of true price and predict price(XGBoost)")+geom_abline (slope= 1, color = "Blue")
```

The scatter plot of true price and predict price(XGBoost)

