Stat350-Final Project

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(stats)
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 3.6.2
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```r
library(car)
```

```
## Warning: package 'car' was built under R version 3.6.2
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```r
# import data

df1 <- read.csv('Real estate.csv', header = T)
names(df1)
```

```
## [1] "No"
## [2] "X1.transaction.date"
## [3] "X2.house.age"
## [4] "X3.distance.to.the.nearest.MRT.station"
## [5] "X4.number.of.convenience.stores"
## [6] "X5.latitude"
## [7] "X6.longitude"
## [8] "Y.house.price.of.unit.area"
```

```r
# create year
df1$Year <- floor(df1$X1.transaction.date)
names(df1)
```

```
## [1] "No"
## [2] "X1.transaction.date"
## [3] "X2.house.age"
## [4] "X3.distance.to.the.nearest.MRT.station"
## [5] "X4.number.of.convenience.stores"
## [6] "X5.latitude"
## [7] "X6.longitude"
## [8] "Y.house.price.of.unit.area"
## [9] "Year"
```

```
table(df1$Year)
```

```
##
## 2012 2013
##  126  288
```

```
df1$X1.transaction.date <- df1$Year
names(df1)[2] <- 'Year'
df1 <- df1[, -9]
names(df1)
```

```
## [1] "No"
## [2] "Year"
## [3] "X2.house.age"
## [4] "X3.distance.to.the.nearest.MRT.station"
## [5] "X4.number.of.convenience.stores"
## [6] "X5.latitude"
## [7] "X6.longitude"
## [8] "Y.house.price.of.unit.area"
```

```
# transform of the response
df1$Y2 <- sqrt(df1$Y.house.price.of.unit.area)

df2 <- df1 %>%
  filter(Year == '2012')

# introduce new data point
df1[415, 1] <- 415
df1[415, 2] <- '2012'
N <- 414
for (i in 3: 8){
  set.seed(i)
  SRS.index <- sample.int(N, 1, replace = FALSE)
  df1[415, i] <- df1[SRS.index, i]
}
df1[415, ]
```

```
##      No Year X2.house.age X3.distance.to.the.nearest.MRT.station
## 415 415 2012           17                               379.5575
##     X4.number.of.convenience.stores X5.latitude X6.longitude
## 415                               4    24.98203     121.5458
##     Y.house.price.of.unit.area Y2
## 415                       28.6 NA
```

```
# Potential outlier
df1[which.max(df1$Y2),]
```

```
##      No Year X2.house.age X3.distance.to.the.nearest.MRT.station
## 271 271 2013         10.8                               252.5822
##     X4.number.of.convenience.stores X5.latitude X6.longitude
## 271                               1     24.9746     121.5305
##     Y.house.price.of.unit.area      Y2
## 271                      117.5 10.83974
```

```
# Figure 2
par(mfrow=c(2,2))
plot(df1$X2.house.age, df1$Y2, col = df1$Year, pch = 16,
     main = 'A. house age and house price',
     xlab = 'house.age', ylab = 'sqrt.house.price.of.unit.area')
plot(df1$X3.distance.to.the.nearest.MRT.station, df1$Y2, col = df1$Year, pch = 16,
     main = 'B. distance to MRT and house price',
     xlab = 'distance.to.MRT', ylab = 'sqrt.house.price.of.unit.area')
plot(df1$X4.number.of.convenience.stores, df1$Y2, col = df1$Year, pch = 16,
     main = 'C. number of stores and house price',
     xlab = 'number.of.convenience.stores', ylab = 'sqrt.house.price.of.unit.area')
plot(df1$X5.latitude, df1$X6.longitude, col = df1$Year, pch = 16,
     main = 'D. latitude vs. longitude',
     xlab = 'latitude', ylab = 'longitude')
```
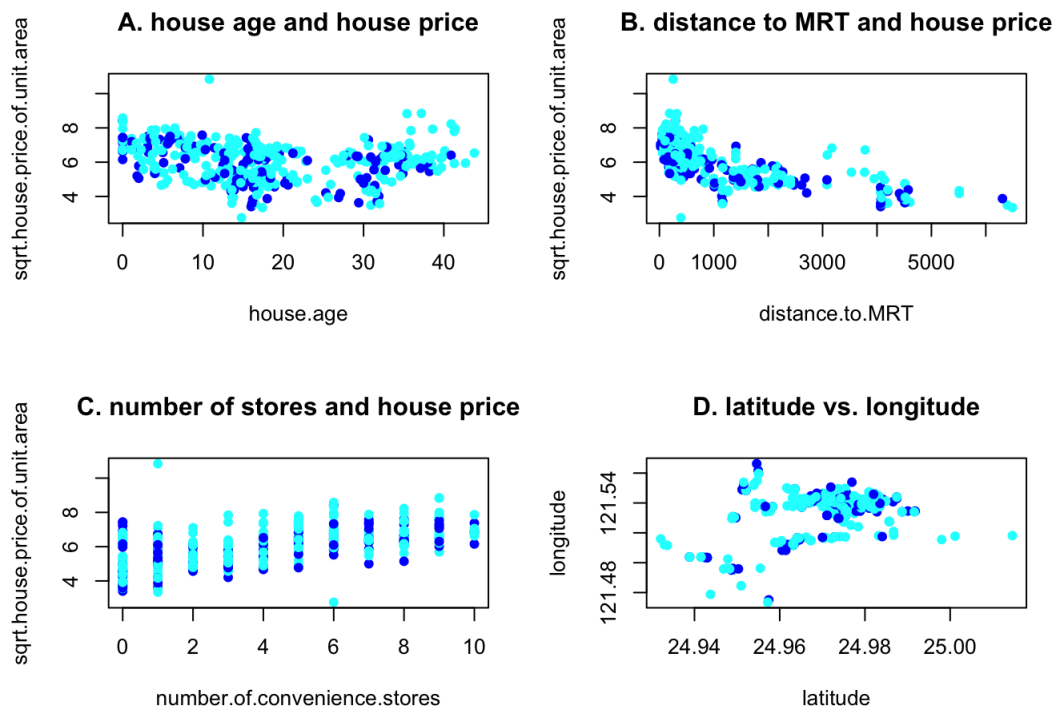


```
# Figure 1
par(mfrow=c(1,2))
hist(df1$Y.house.price.of.unit.area, main = 'A. Original scale')
hist(df1$Y2, main = 'B. Square-root scale')
```

A. Original scale

df1$Y.house.price.of.unit.area

B. Square-root scale

df1$Y2
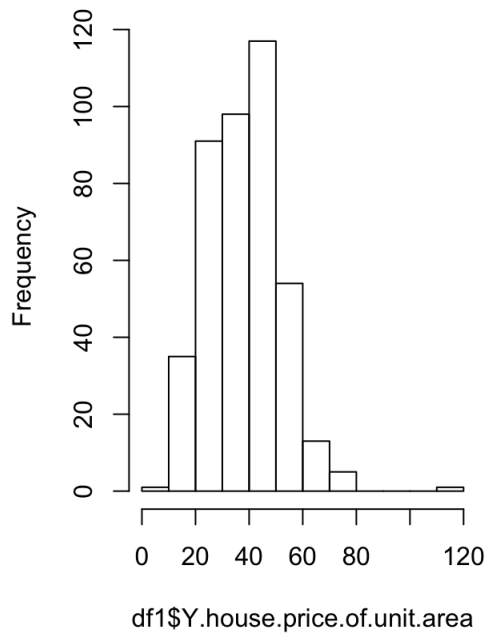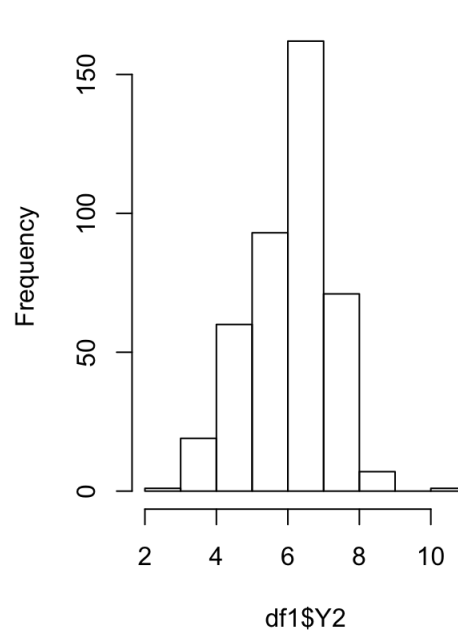
```
hist(df1$X4.number.of.convenience.stores)
fivenum(df1$X4.number.of.convenience.stores)
```

```
## [1]  0  1  4  6 10
```

```
#hist(log(df1$Y.house.price.of.unit.area))
hist(sqrt(df1$Y.house.price.of.unit.area))
```

gram of df1$X4.number.of.convenienogram of sqrt(df1$Y.house.price.of.u

df1$X4.number.of.convenience.stores

sqrt(df1$Y.house.price.of.unit.area)
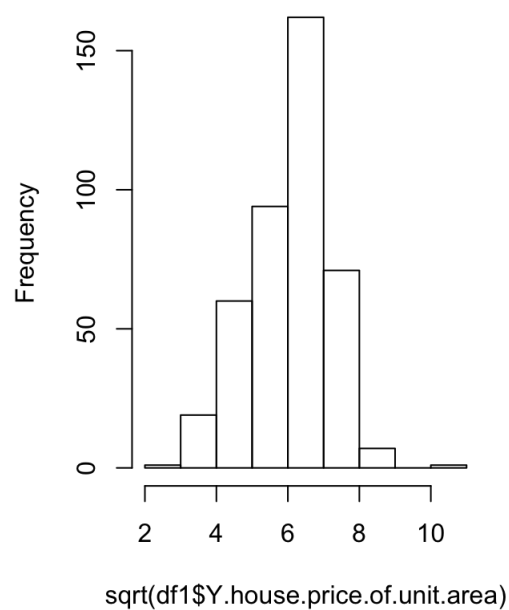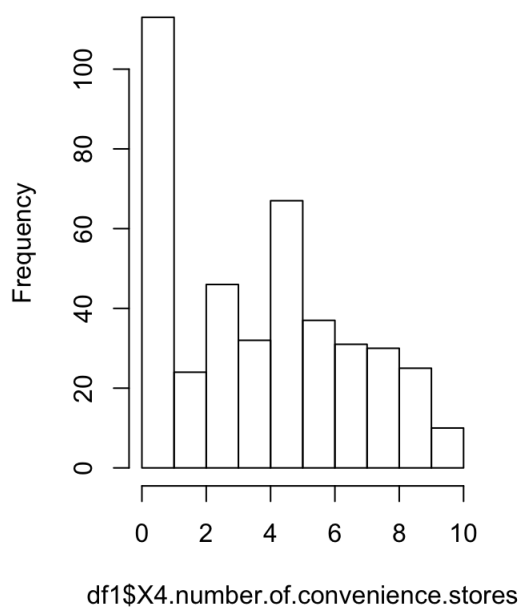
```
# test for normality
shapiro.test(sqrt(df1$Y.house.price.of.unit.area))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  sqrt(df1$Y.house.price.of.unit.area)
## W = 0.98833, p-value = 0.002112
```

```
plot(df1$X5.latitude, df1$X6.longitude, col = df1$Year)
plot(df1$X2.house.age, df1$Y.house.price.of.unit.area)
```
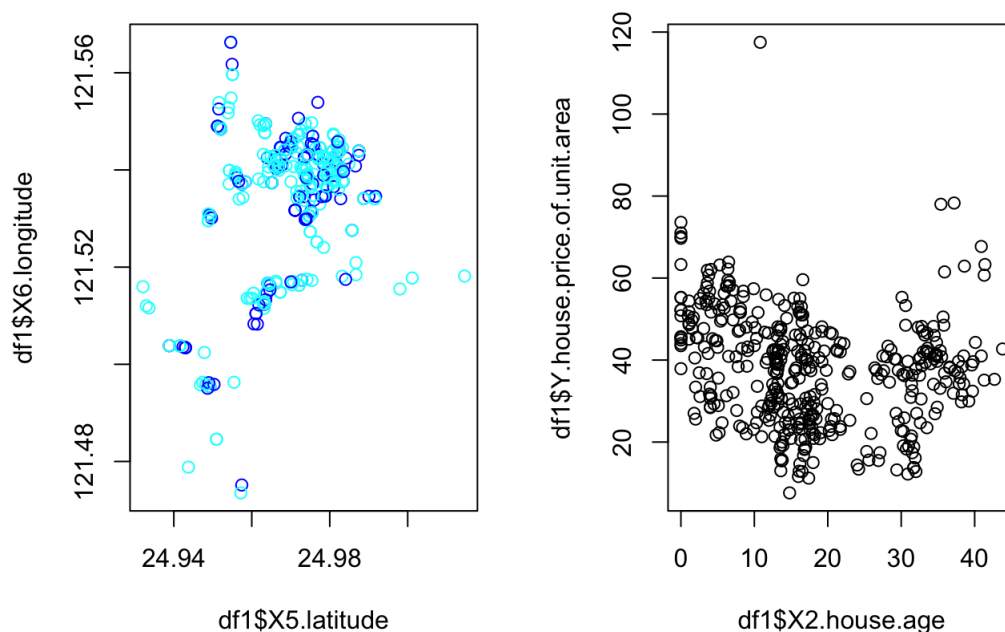


```
# Figure 3
# check multi-colineariry
pairs(df1[,3:5], col= "blue", pch=18,
      main= "Relationship between variables")
```

Relationship between variables
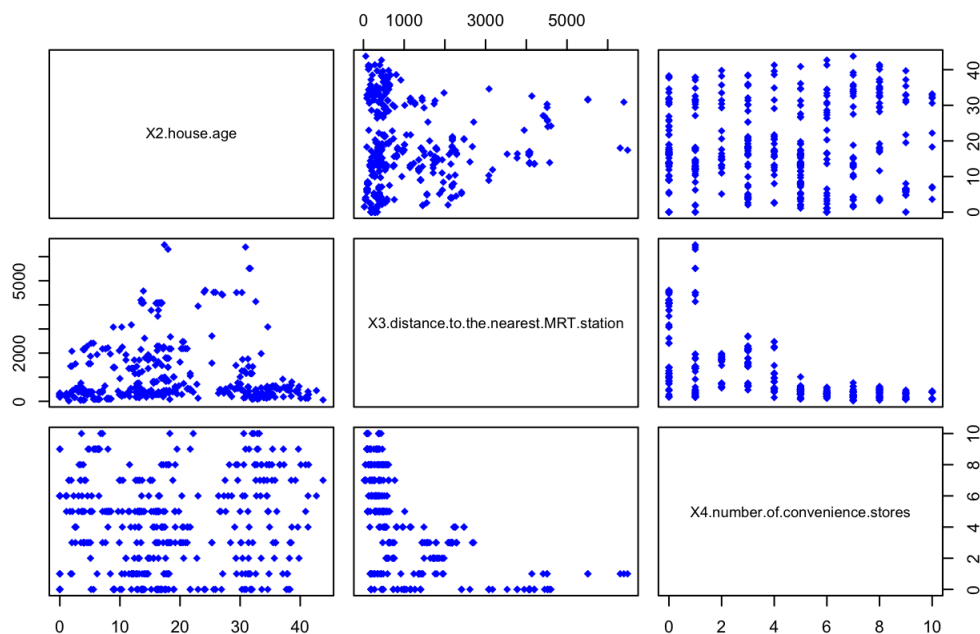
```
# variable selection

# PCA
df1$Year <- as.factor(df1$Year)
df1.pca <- prcomp(df1[,c(4, 6, 7)], center = TRUE,scale. = TRUE)
summary(df1.pca)
```

```
## Importance of components:
##                           PC1    PC2     PC3
## Standard deviation     1.4908 0.7833 0.40468
## Proportion of Variance 0.7409 0.2045 0.05459
## Cumulative Proportion  0.7409 0.9454 1.00000
```

```
par(mfrow=c(1,1))
plot(df1.pca$x[,1:2], col = df1$Year)
```
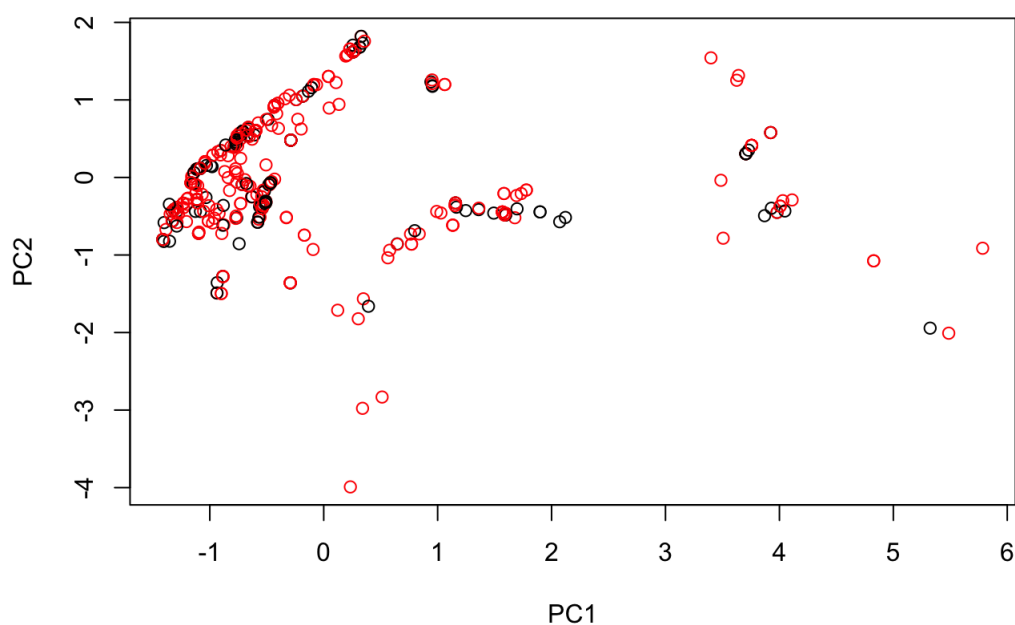


```
# M0: Baseline model
#df1$Y2 <- sqrt(df1$Y.house.price.of.unit.area)
my.lm0 <- lm(Y2 ~ ., data = df1[, -c(1, 8)])
vif(my.lm0)
```

```
##                               Year                         X2.house.age
##                           1.005502                             1.016652
## X3.distance.to.the.nearest.MRT.station  X4.number.of.convenience.stores
##                           4.293775                             1.613346
##                         X5.latitude                         X6.longitude
##                           1.599937                             2.931216
```

```
summary(my.lm0)
```

```
##
## Call:
## lm(formula = Y2 ~ ., data = df1[, -c(1, 8)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7196 -0.3947 -0.0600  0.3361  4.3982
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        -5.017e+02  4.659e+02  -1.077  0.28223
## Year2013                            2.323e-01  7.193e-02   3.230  0.00134
## X2.house.age                       -2.178e-02  2.925e-03  -7.447 5.74e-13
## X3.distance.to.the.nearest.MRT.station -3.870e-04  5.426e-05  -7.133 4.53e-12
## X4.number.of.convenience.stores     9.123e-02  1.425e-02   6.402 4.23e-10
## X5.latitude                         2.157e+01  3.368e+00   6.403 4.20e-10
## X6.longitude                       -2.510e-01  3.687e+00  -0.068  0.94576
##
## (Intercept)
## Year2013                           **
## X2.house.age                       ***
## X3.distance.to.the.nearest.MRT.station ***
## X4.number.of.convenience.stores    ***
## X5.latitude                        ***
## X6.longitude
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6716 on 407 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.6472, Adjusted R-squared:  0.642
## F-statistic: 124.4 on 6 and 407 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(my.lm0)
```
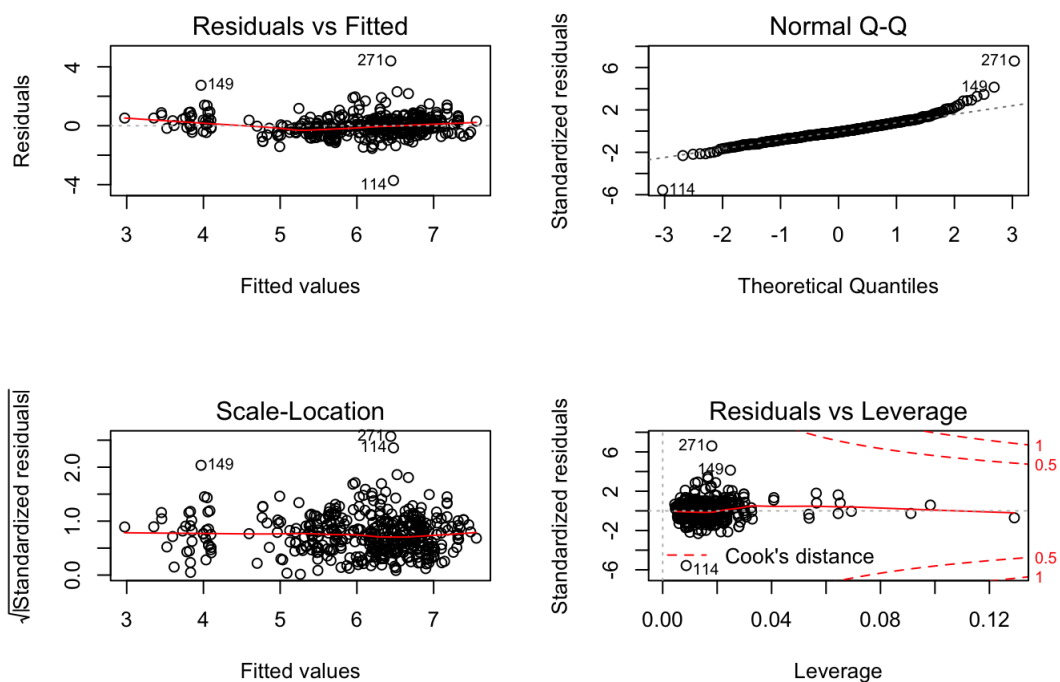


```
# M1: full model
df1$Y2 <- sqrt(df1$Y.house.price.of.unit.area)
my.lm1 <- lm(Y2 ~ (.)^2, data = df1[, -c(1, 8)])
#vif(my.lm1)
summary(my.lm1)
```

```
##
```

```
## Call:
## lm(formula = Y2 ~ (.)^2, data = df1[, -c(1, 8)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7674 -0.3246 -0.0498  0.2796  4.1371
##
## Coefficients: (1 not defined because of singularities)
##                                                                   Estimate
## (Intercept)                                                      -2.295e+03
## Year2013                                                         -5.142e+02
## X2.house.age                                                      4.077e+01
## X3.distance.to.the.nearest.MRT.station                           3.894e-01
## X4.number.of.convenience.stores                                  2.694e+02
## X5.latitude                                                      7.676e+01
## X6.longitude                                                     3.166e+00
## Year2013:X2.house.age                                            1.387e-02
## Year2013:X3.distance.to.the.nearest.MRT.station                 -1.867e-06
## Year2013:X4.number.of.convenience.stores                        -8.755e-03
## Year2013:X5.latitude                                             2.181e+00
## Year2013:X6.longitude                                            3.783e+00
## X2.house.age:X3.distance.to.the.nearest.MRT.station             3.911e-07
## X2.house.age:X4.number.of.convenience.stores                     5.525e-04
## X2.house.age:X5.latitude                                         1.573e-03
## X2.house.age:X6.longitude                                       -3.360e-01
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores -1.141e-04
## X3.distance.to.the.nearest.MRT.station:X5.latitude             -1.744e-02
## X3.distance.to.the.nearest.MRT.station:X6.longitude             3.752e-04
## X4.number.of.convenience.stores:X5.latitude                    -8.764e+00
## X4.number.of.convenience.stores:X6.longitude                   -4.149e-01
## X5.latitude:X6.longitude                                               NA
##                                                                 Std. Error
## (Intercept)                                                      1.314e+03
## Year2013                                                         8.964e+02
## X2.house.age                                                     5.419e+01
## X3.distance.to.the.nearest.MRT.station                           2.402e-01
## X4.number.of.convenience.stores                                  2.993e+02
## X5.latitude                                                      1.319e+01
## X6.longitude                                                     1.021e+01
## Year2013:X2.house.age                                            5.801e-03
## Year2013:X3.distance.to.the.nearest.MRT.station                 1.061e-04
## Year2013:X4.number.of.convenience.stores                        2.783e-02
## Year2013:X5.latitude                                             7.441e+00
## Year2013:X6.longitude                                            7.030e+00
## X2.house.age:X3.distance.to.the.nearest.MRT.station             6.008e-06
## X2.house.age:X4.number.of.convenience.stores                     1.065e-03
## X2.house.age:X5.latitude                                         3.350e-01
## X2.house.age:X6.longitude                                         4.400e-01
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores 2.961e-05
## X3.distance.to.the.nearest.MRT.station:X5.latitude             3.370e-03
## X3.distance.to.the.nearest.MRT.station:X6.longitude             2.073e-03
## X4.number.of.convenience.stores:X5.latitude                     1.649e+00
## X4.number.of.convenience.stores:X6.longitude                    2.380e+00
## X5.latitude:X6.longitude                                               NA
##                                                                    t value
## (Intercept)                                                       -1.747
## Year2013                                                          -0.574
## X2.house.age                                                       0.752
## X3.distance.to.the.nearest.MRT.station                            1.621
## X4.number.of.convenience.stores                                   0.900
## X5.latitude                                                       5.820
## X6.longitude                                                      0.310
## Year2013:X2.house.age                                             2.391
## Year2013:X3.distance.to.the.nearest.MRT.station                  -0.018
## Year2013:X4.number.of.convenience.stores                         -0.315
## Year2013:X5.latitude                                              0.293
## Year2013:X6.longitude                                             0.538
## X2.house.age:X3.distance.to.the.nearest.MRT.station              0.065
## X2.house.age:X4.number.of.convenience.stores                      0.519
## X2.house.age:X5.latitude                                          0.005
## X2.house.age:X6.longitude                                        -0.764
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores -3.854
## X3.distance.to.the.nearest.MRT.station:X5.latitude              -5.173
```

```
## X3.distance.to.the.nearest.MRT.station:X6.longitude                          0.181
## X4.number.of.convenience.stores:X5.latitude                                 -5.315
## X4.number.of.convenience.stores:X6.longitude                                -0.174
## X5.latitude:X6.longitude                                                        NA
##                                                                            Pr(>|t|)
## (Intercept)                                                                0.081421
## Year2013                                                                   0.566525
## X2.house.age                                                               0.452275
## X3.distance.to.the.nearest.MRT.station                                     0.105765
## X4.number.of.convenience.stores                                            0.368655
## X5.latitude                                                                1.22e-08
## X6.longitude                                                               0.756710
## Year2013:X2.house.age                                                      0.017248
## Year2013:X3.distance.to.the.nearest.MRT.station                            0.985973
## Year2013:X4.number.of.convenience.stores                                   0.753206
## Year2013:X5.latitude                                                       0.769600
## Year2013:X6.longitude                                                      0.590792
## X2.house.age:X3.distance.to.the.nearest.MRT.station                        0.948132
## X2.house.age:X4.number.of.convenience.stores                               0.604252
## X2.house.age:X5.latitude                                                   0.996257
## X2.house.age:X6.longitude                                                  0.445451
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores 0.000136
## X3.distance.to.the.nearest.MRT.station:X5.latitude                         3.68e-07
## X3.distance.to.the.nearest.MRT.station:X6.longitude                        0.856448
## X4.number.of.convenience.stores:X5.latitude                                1.79e-07
## X4.number.of.convenience.stores:X6.longitude                               0.861705
## X5.latitude:X6.longitude                                                        NA
##
## (Intercept)                                                                .
## Year2013
## X2.house.age
## X3.distance.to.the.nearest.MRT.station
## X4.number.of.convenience.stores
## X5.latitude                                                                ***
## X6.longitude
## Year2013:X2.house.age                                                      *
## Year2013:X3.distance.to.the.nearest.MRT.station
## Year2013:X4.number.of.convenience.stores
## Year2013:X5.latitude
## Year2013:X6.longitude
## X2.house.age:X3.distance.to.the.nearest.MRT.station
## X2.house.age:X4.number.of.convenience.stores
## X2.house.age:X5.latitude
## X2.house.age:X6.longitude
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores ***
## X3.distance.to.the.nearest.MRT.station:X5.latitude                         ***
## X3.distance.to.the.nearest.MRT.station:X6.longitude
## X4.number.of.convenience.stores:X5.latitude                                ***
## X4.number.of.convenience.stores:X6.longitude
## X5.latitude:X6.longitude
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6042 on 394 degrees of freedom
## Multiple R-squared:  0.7238, Adjusted R-squared:  0.7098
## F-statistic: 51.62 on 20 and 394 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(my.lm1)
```
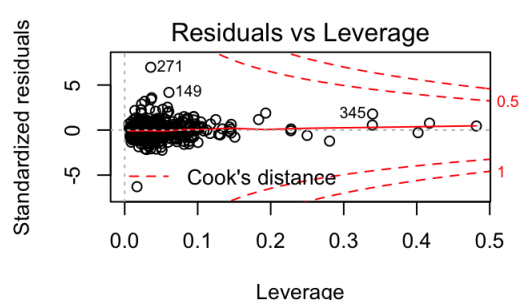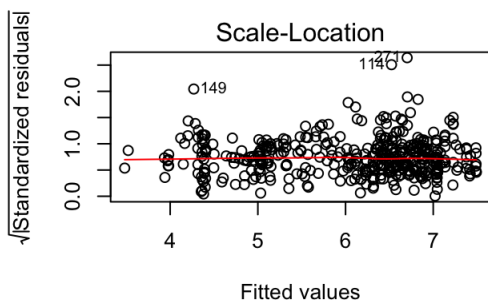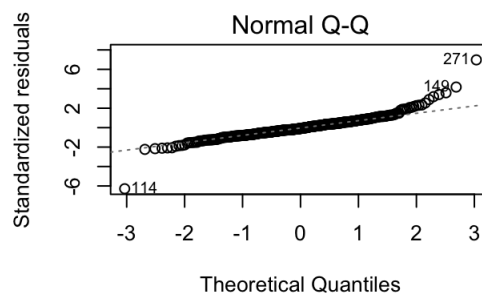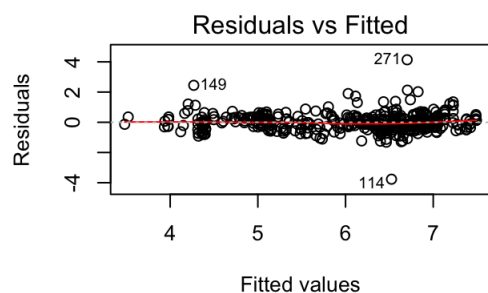
```
# M2: stepwise
fit.start <- lm(Y2 ~1, data = df1[, -c(1, 8)])
fit.end <- lm(Y2 ~ (.)^2, data = df1[, -c(1, 8)])
step.BIC = step(fit.start, list(upper = fit.end),
                K = log(nrow(df1[, -c(1, 8)])), trace = 0)
summary(step.BIC)
```

```
##
## Call:
## lm(formula = Y2 ~ X3.distance.to.the.nearest.MRT.station + X4.number.of.convenience.stores +
##     X2.house.age + X5.latitude + Year + X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.s
tores +
##     X2.house.age:Year + X3.distance.to.the.nearest.MRT.station:X5.latitude +
##     X4.number.of.convenience.stores:X5.latitude, data = df1[,
##     -c(1, 8)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7651 -0.3264 -0.0508  0.2851  4.1582
##
## Coefficients:
##                                                                   Estimate
## (Intercept)                                                      -1.926e+03
## X3.distance.to.the.nearest.MRT.station                            4.237e-01
## X4.number.of.convenience.stores                                   2.157e+02
## X2.house.age                                                     -3.068e-02
## X5.latitude                                                       7.740e+01
## Year2013                                                         -1.352e-02
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores -1.114e-04
## X2.house.age:Year2013                                             1.404e-02
## X3.distance.to.the.nearest.MRT.station:X5.latitude              -1.698e-02
## X4.number.of.convenience.stores:X5.latitude                     -8.633e+00
##                                                                   Std. Error
## (Intercept)                                                       1.965e+02
## X3.distance.to.the.nearest.MRT.station                            6.587e-02
## X4.number.of.convenience.stores                                   3.423e+01
## X2.house.age                                                      4.702e-03
## X5.latitude                                                       7.870e+00
## Year2013                                                          1.160e-01
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores 1.823e-05
## X2.house.age:Year2013                                             5.626e-03
## X3.distance.to.the.nearest.MRT.station:X5.latitude               2.639e-03
## X4.number.of.convenience.stores:X5.latitude                      1.371e+00
##                                                                   t value
```

```
## (Intercept)                                                                    -9.801
## X3.distance.to.the.nearest.MRT.station                                          6.432
## X4.number.of.convenience.stores                                                 6.301
## X2.house.age                                                                   -6.526
## X5.latitude                                                                     9.834
## Year2013                                                                       -0.117
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores         -6.111
## X2.house.age:Year2013                                                           2.496
## X3.distance.to.the.nearest.MRT.station:X5.latitude                             -6.435
## X4.number.of.convenience.stores:X5.latitude                                    -6.298
##                                                                                Pr(>|t|)
## (Intercept)                                                                    < 2e-16
## X3.distance.to.the.nearest.MRT.station                                         3.56e-10
## X4.number.of.convenience.stores                                                7.72e-10
## X2.house.age                                                                   2.02e-10
## X5.latitude                                                                    < 2e-16
## Year2013                                                                        0.907
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores 2.32e-09
## X2.house.age:Year2013                                                           0.013
## X3.distance.to.the.nearest.MRT.station:X5.latitude                             3.49e-10
## X4.number.of.convenience.stores:X5.latitude                                    7.87e-10
##
## (Intercept)                                                                    ***
## X3.distance.to.the.nearest.MRT.station                                         ***
## X4.number.of.convenience.stores                                                ***
## X2.house.age                                                                   ***
## X5.latitude                                                                    ***
## Year2013
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores ***
## X2.house.age:Year2013                                                          *
## X3.distance.to.the.nearest.MRT.station:X5.latitude                             ***
## X4.number.of.convenience.stores:X5.latitude                                    ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5981 on 405 degrees of freedom
## Multiple R-squared:  0.7218, Adjusted R-squared:  0.7157
## F-statistic: 116.8 on 9 and 405 DF,  p-value: < 2.2e-16
```
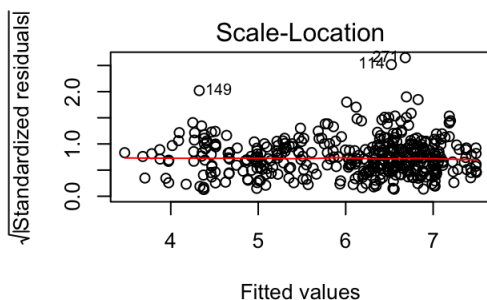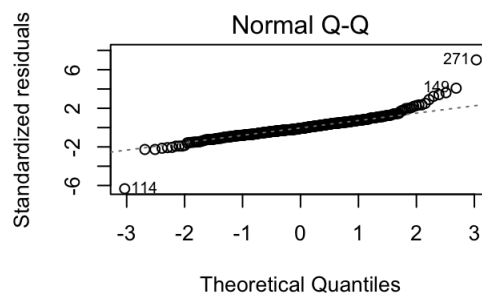
```
plot(step.BIC)
```
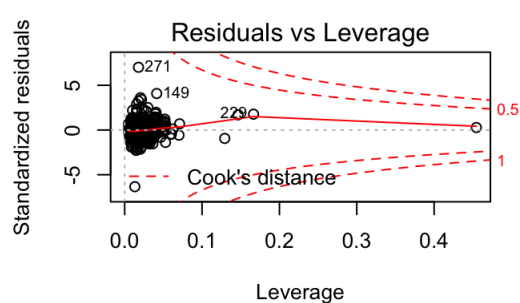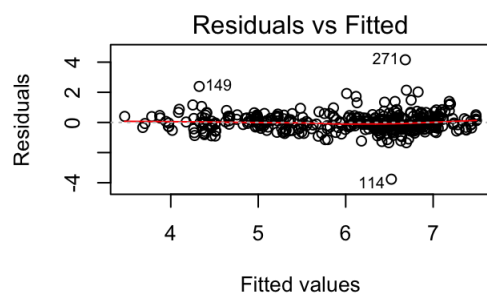
```
df1$Y2 <- sqrt(df1$Y.house.price.of.unit.area)
my.lm1 <- lm(Y2 ~., data = df1[, -c(1, 8)])
vif(my.lm1)
```

```
##                              Year                      X2.house.age
##                          1.005202                          1.016661
## X3.distance.to.the.nearest.MRT.station   X4.number.of.convenience.stores
##                          4.294917                          1.612241
##                      X5.latitude                        X6.longitude
##                          1.599704                          2.931653
```

```
summary(my.lm1)
```

```
##
## Call:
## lm(formula = Y2 ~ ., data = df1[, -c(1, 8)])
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -3.7233 -0.4009 -0.0593  0.3415  4.3993
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -4.666e+02  4.664e+02  -1.000 0.317680
## Year2013                        2.413e-01  7.187e-02   3.357 0.000862
## X2.house.age                   -2.179e-02  2.931e-03  -7.433 6.29e-13
## X3.distance.to.the.nearest.MRT.station -3.890e-04  5.435e-05  -7.156 3.89e-12
## X4.number.of.convenience.stores  9.184e-02  1.427e-02   6.434 3.49e-10
## X5.latitude                     2.128e+01  3.370e+00   6.313 7.16e-10
## X6.longitude                   -4.798e-01  3.691e+00  -0.130 0.896656
##
## (Intercept)
## Year2013                          ***
## X2.house.age                      ***
## X3.distance.to.the.nearest.MRT.station ***
## X4.number.of.convenience.stores   ***
## X5.latitude                       ***
## X6.longitude
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.673 on 408 degrees of freedom
## Multiple R-squared:  0.6452, Adjusted R-squared:   0.64
## F-statistic: 123.7 on 6 and 408 DF,  p-value: < 2.2e-16
```

```
# K-means clutering
set.seed(2020)
km.res <- kmeans(df1[,c(6, 7)], 3, nstart = 25)

fviz_cluster(km.res, data = df1[,c(6, 7)],
             palette = c("#2E9FDF", "#00AFBB", "#E7B800"),
             geom = "point",
             ellipse.type = "convex",
             ggtheme = theme_bw()
)
```

Cluster plot
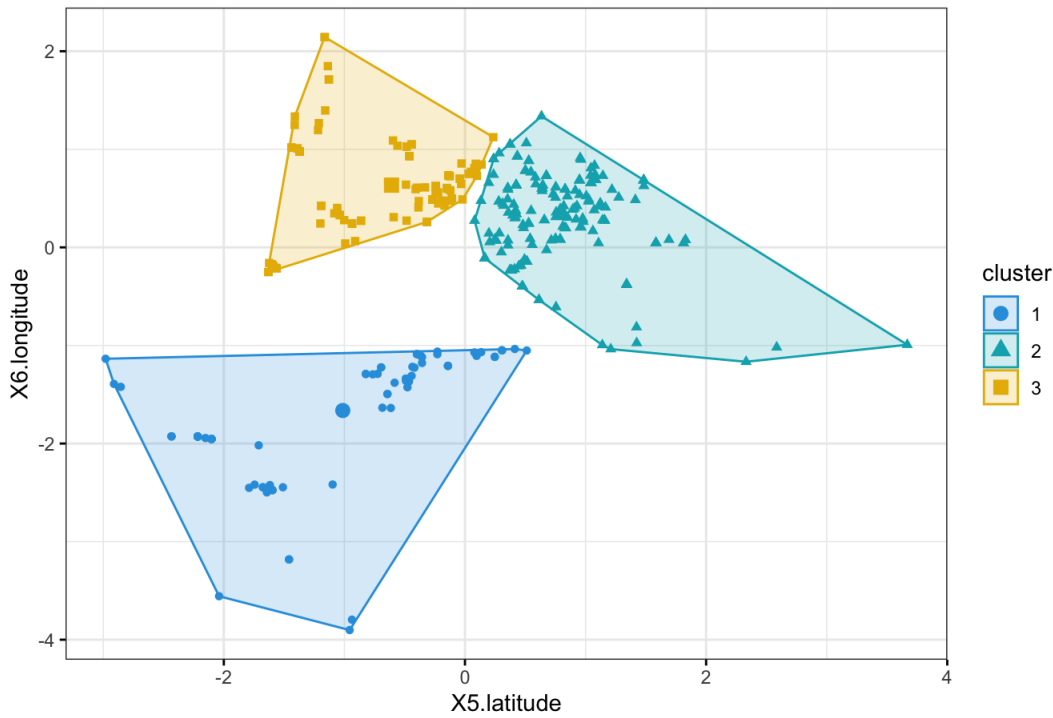

```
df3 <- cbind(df1, cluster = km.res$cluster)
head(df3)
```

```
##   No Year X2.house.age X3.distance.to.the.nearest.MRT.station
## 1  1 2012         32.0                               84.87882
## 2  2 2012         19.5                              306.59470
## 3  3 2013         13.3                              561.98450
## 4  4 2013         13.3                              561.98450
## 5  5 2012          5.0                              390.56840
## 6  6 2012          7.1                             2175.03000
##   X4.number.of.convenience.stores X5.latitude X6.longitude
## 1                              10    24.98298     121.5402
## 2                               9    24.98034     121.5395
## 3                               5    24.98746     121.5439
## 4                               5    24.98746     121.5439
## 5                               5    24.97937     121.5425
## 6                               3    24.96305     121.5125
##   Y.house.price.of.unit.area       Y2 cluster
## 1                       37.9 6.156298       2
## 2                       42.2 6.496153       2
## 3                       47.3 6.877500       2
## 4                       54.8 7.402702       2
## 5                       43.1 6.565059       2
## 6                       32.1 5.665686       1
```

```
names(df3)
```

```
##  [1] "No"
##  [2] "Year"
##  [3] "X2.house.age"
##  [4] "X3.distance.to.the.nearest.MRT.station"
##  [5] "X4.number.of.convenience.stores"
##  [6] "X5.latitude"
##  [7] "X6.longitude"
##  [8] "Y.house.price.of.unit.area"
##  [9] "Y2"
## [10] "cluster"
```

```
df3$cluster <- as.factor(df3$cluster)
my.lm2 <- lm(Y2 ~., data = df3[, c(2:5, 9:10)])
vif(my.lm2)
```

```
##                                          GVIF Df GVIF^(1/(2*Df))
## Year                                 1.003309  1        1.001653
## X2.house.age                         1.026710  1        1.013267
## X3.distance.to.the.nearest.MRT.station 3.557034  1      1.886010
## X4.number.of.convenience.stores      1.620469  1        1.272976
## cluster                              2.716769  2        1.283847
```

```
summary(my.lm2)
```

```
##
## Call:
## lm(formula = Y2 ~ ., data = df3[, c(2:5, 9:10)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5898 -0.3796 -0.0341  0.2872  4.2499
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       5.781e+00  1.742e-01  33.179  < 2e-16
## Year2013                          2.352e-01  6.813e-02   3.453 0.000613
## X2.house.age                     -2.325e-02  2.795e-03  -8.320 1.33e-15
## X3.distance.to.the.nearest.MRT.station -3.304e-04  4.693e-05  -7.039 8.25e-12
## X4.number.of.convenience.stores   1.072e-01  1.358e-02   7.894 2.73e-14
## cluster2                          8.013e-01  1.305e-01   6.138 1.98e-09
## cluster3                          1.616e-01  1.309e-01   1.235 0.217524
##
## (Intercept)                       ***
## Year2013                          ***
## X2.house.age                      ***
## X3.distance.to.the.nearest.MRT.station ***
## X4.number.of.convenience.stores   ***
## cluster2                          ***
## cluster3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6385 on 408 degrees of freedom
## Multiple R-squared:  0.6806, Adjusted R-squared:  0.6759
## F-statistic: 144.9 on 6 and 408 DF,  p-value: < 2.2e-16
```

```
fit.start <- lm(Y2 ~., data = df3[, c(2:5, 9:10)])
fit.end <- lm(Y2 ~ (.)^2, data = df3[, c(2:5, 9:10)])
step.BIC2 = step(fit.start, list(upper = fit.end),
               K = log(nrow(df1[, -c(1, 8)])), trace = 0)
summary(step.BIC2)
```

```
##
## Call:
## lm(formula = Y2 ~ Year + X2.house.age + X3.distance.to.the.nearest.MRT.station +
##     X4.number.of.convenience.stores + cluster + X4.number.of.convenience.stores:cluster +
##     X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores +
##     X2.house.age:X3.distance.to.the.nearest.MRT.station + Year:X2.house.age,
##     data = df3[, c(2:5, 9:10)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6374 -0.3397 -0.0325  0.2727  4.0549
##
## Coefficients:
##                                                       Estimate
## (Intercept)                                          5.489e+00
## Year2013                                             5.327e-02
## X2.house.age                                        -3.289e-02
## X3.distance.to.the.nearest.MRT.station              -2.863e-04
## X4.number.of.convenience.stores                      4.560e-01
## cluster2                                             1.467e+00
## cluster3                                             9.378e-02
## X4.number.of.convenience.stores:cluster2            -3.479e-01
## X4.number.of.convenience.stores:cluster3            -1.992e-01
```

```
## X4.number.of.convenience.stores:cluster3                                     -1.992e-01
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores       -1.618e-04
## X2.house.age:X3.distance.to.the.nearest.MRT.station                           6.173e-06
## Year2013:X2.house.age                                                         1.097e-02
##                                                                             Std. Error
## (Intercept)                                                                   3.179e-01
## Year2013                                                                      1.151e-01
## X2.house.age                                                                  4.940e-03
## X3.distance.to.the.nearest.MRT.station                                        8.856e-05
## X4.number.of.convenience.stores                                               1.086e-01
## cluster2                                                                      2.743e-01
## cluster3                                                                      2.403e-01
## X4.number.of.convenience.stores:cluster2                                      1.010e-01
## X4.number.of.convenience.stores:cluster3                                      9.612e-02
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores        3.209e-05
## X2.house.age:X3.distance.to.the.nearest.MRT.station                           2.838e-06
## Year2013:X2.house.age                                                         5.573e-03
##                                                                                 t value
## (Intercept)                                                                      17.267
## Year2013                                                                          0.463
## X2.house.age                                                                     -6.658
## X3.distance.to.the.nearest.MRT.station                                           -3.234
## X4.number.of.convenience.stores                                                   4.197
## cluster2                                                                          5.349
## cluster3                                                                          0.390
## X4.number.of.convenience.stores:cluster2                                         -3.443
## X4.number.of.convenience.stores:cluster3                                         -2.072
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores           -5.041
## X2.house.age:X3.distance.to.the.nearest.MRT.station                               2.175
## Year2013:X2.house.age                                                             1.968
##                                                                                Pr(>|t|)
## (Intercept)                                                                     < 2e-16
## Year2013                                                                        0.643659
## X2.house.age                                                                    9.09e-11
## X3.distance.to.the.nearest.MRT.station                                         0.001324
## X4.number.of.convenience.stores                                                3.33e-05
## cluster2                                                                        1.48e-07
## cluster3                                                                        0.696528
## X4.number.of.convenience.stores:cluster2                                       0.000636
## X4.number.of.convenience.stores:cluster3                                       0.038886
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores         7.02e-07
## X2.house.age:X3.distance.to.the.nearest.MRT.station                            0.030185
## Year2013:X2.house.age                                                          0.049802
##
## (Intercept)                                                                     ***
## Year2013
## X2.house.age                                                                    ***
## X3.distance.to.the.nearest.MRT.station                                          **
## X4.number.of.convenience.stores                                                 ***
## cluster2                                                                        ***
## cluster3
## X4.number.of.convenience.stores:cluster2                                        ***
## X4.number.of.convenience.stores:cluster3                                        *
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores          ***
## X2.house.age:X3.distance.to.the.nearest.MRT.station                             *
## Year2013:X2.house.age                                                           *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5905 on 403 degrees of freedom
## Multiple R-squared:  0.7301, Adjusted R-squared:  0.7228
## F-statistic: 99.13 on 11 and 403 DF,  p-value: < 2.2e-16
```

```
plot(step.BIC2)
```
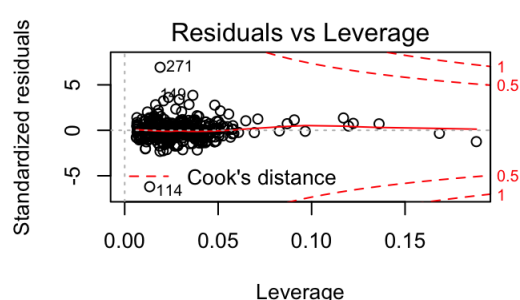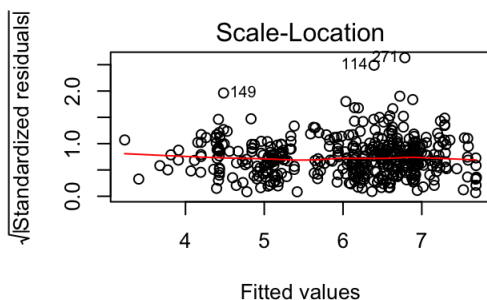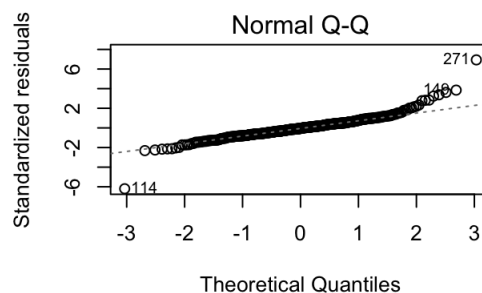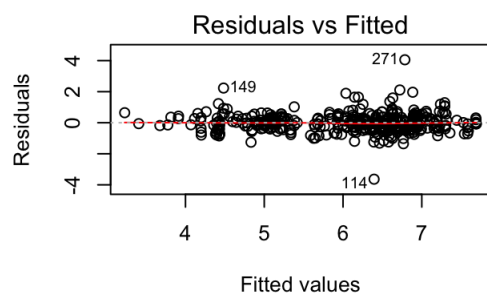
```
names(step.BIC2)
```

```
## [1] "coefficients"  "residuals"     "effects"       "rank"
## [5] "fitted.values" "assign"        "qr"            "df.residual"
## [9] "contrasts"     "xlevels"       "call"          "terms"
## [13] "model"        "anova"
```

```
step.BIC2$coefficients
```

```
##                                                                  (Intercept)
##                                                                 5.489256e+00
##                                                                     Year2013
##                                                                 5.327209e-02
##                                                                 X2.house.age
##                                                                -3.288955e-02
##                                  X3.distance.to.the.nearest.MRT.station
##                                                                -2.863489e-04
##                                     X4.number.of.convenience.stores
##                                                                 4.559547e-01
##                                                                     cluster2
##                                                                 1.467409e+00
##                                                                     cluster3
##                                                                 9.377991e-02
##                              X4.number.of.convenience.stores:cluster2
##                                                                -3.478953e-01
##                              X4.number.of.convenience.stores:cluster3
##                                                                -1.991656e-01
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores
##                                                                -1.617496e-04
##              X2.house.age:X3.distance.to.the.nearest.MRT.station
##                                                                 6.172852e-06
##                                                         Year2013:X2.house.age
##                                                                 1.096547e-02
```

```
confint(step.BIC2)
```

```
##                                                                           2.5 %
## (Intercept)                                                        4.864288e+00
## Year2013                                                          -1.729483e-01
## X2.house.age                                                      -4.260016e-02
## X3.distance.to.the.nearest.MRT.station                           -4.604399e-04
## X4.number.of.convenience.stores                                   2.423712e-01
## cluster2                                                           9.281539e-01
## cluster3                                                          -3.785827e-01
## X4.number.of.convenience.stores:cluster2                         -5.465367e-01
## X4.number.of.convenience.stores:cluster3                         -3.881154e-01
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores -2.248320e-04
## X2.house.age:X3.distance.to.the.nearest.MRT.station               5.944025e-07
## Year2013:X2.house.age                                             9.540238e-06
##                                                                          97.5 %
## (Intercept)                                                        6.114225e+00
## Year2013                                                           2.794925e-01
## X2.house.age                                                      -2.317895e-02
## X3.distance.to.the.nearest.MRT.station                           -1.122579e-04
## X4.number.of.convenience.stores                                   6.695381e-01
## cluster2                                                           2.006664e+00
## cluster3                                                           5.661425e-01
## X4.number.of.convenience.stores:cluster2                         -1.492538e-01
## X4.number.of.convenience.stores:cluster3                         -1.021575e-02
## X3.distance.to.the.nearest.MRT.station:X4.number.of.convenience.stores -9.866711e-05
## X2.house.age:X3.distance.to.the.nearest.MRT.station               1.175130e-05
## Year2013:X2.house.age                                             2.192140e-02
```

```
my.ci <- cbind(step.BIC2$coefficients, confint(step.BIC2))
write.csv(my.ci, 'ci.csv')
```