Home Price Model

```
#This section of the code focused on looking at correlations between predictors, doing different regres
rm(list = ls())
library(readxl)
library(olsrr)
```

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

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

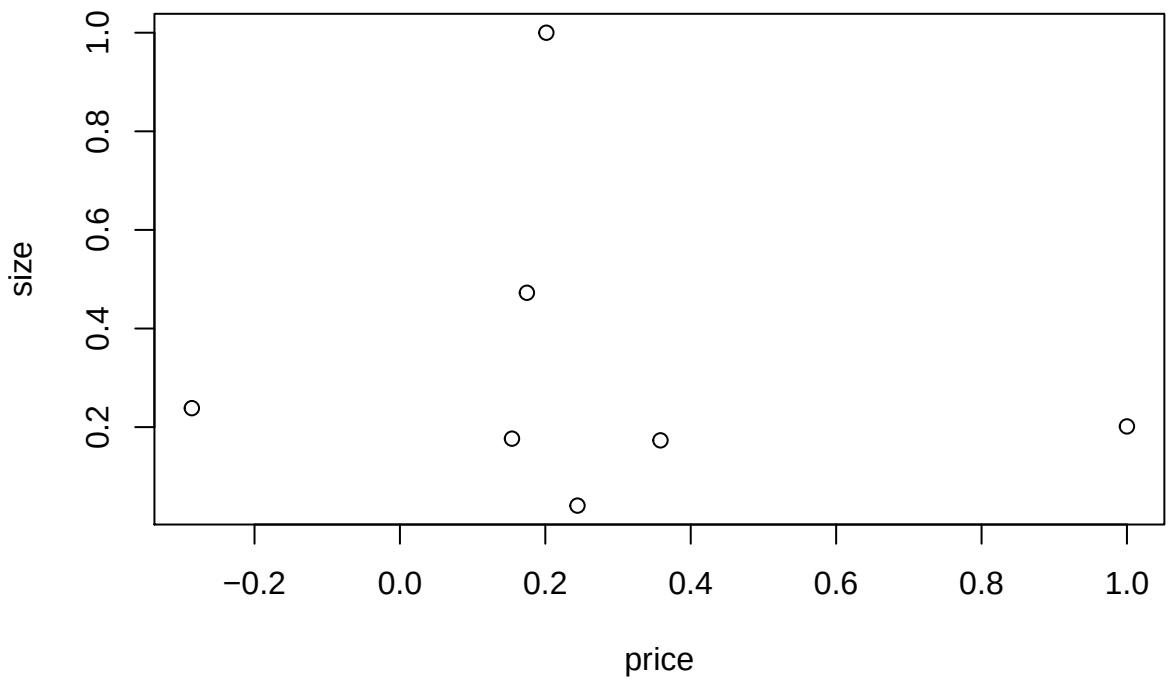
```
##
## Attaching package: 'olsrr'
```

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

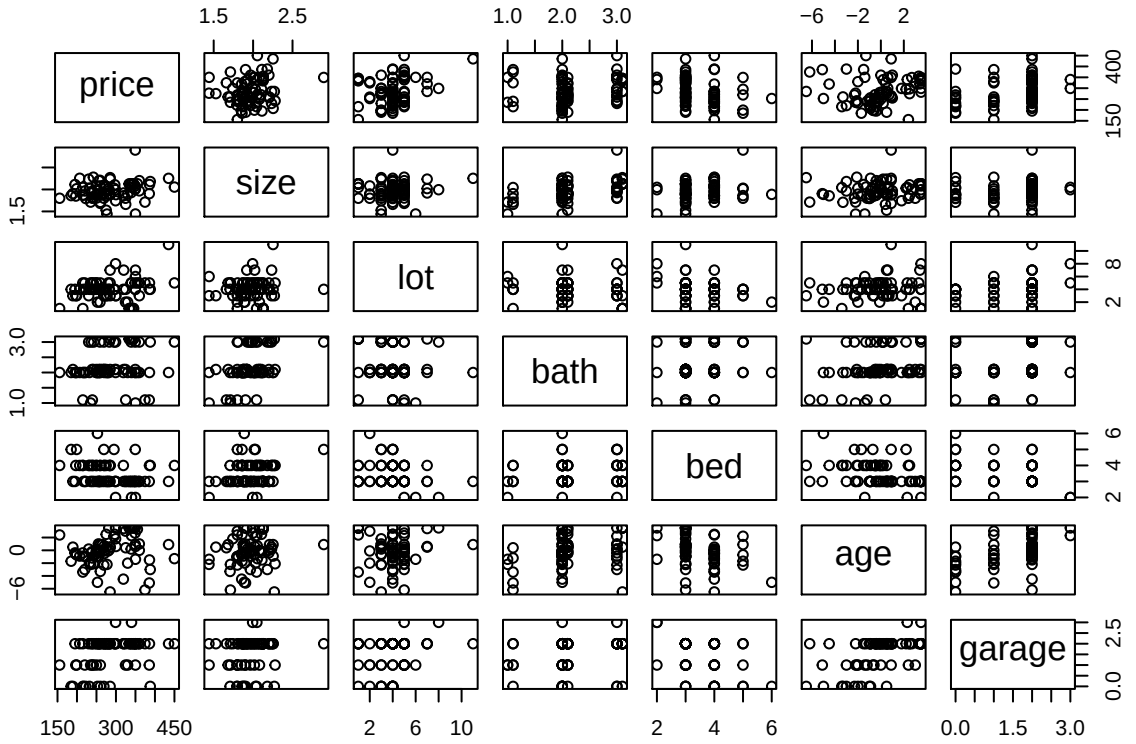
```
homeprice <- read_excel("C:/Users/Daniel Huang/Desktop/Junior/IE Project/homeprice (1).xls")
homeprice2 <- subset(homeprice, select = -c(id, elem, status, year, agesq, bathbed))
View(homeprice)

#Correlation between predictors
numeric_data <- subset(homeprice, select = c(price, size, lot, bath, bed, age, garage))
cor(numeric_data)
```

```
##             price       size         lot       bath        bed         age
## price   1.0000000 0.20143783  0.24423228  0.1746578 -0.2861975  0.15412476
## size    0.2014378 1.00000000  0.04079199  0.4725406  0.2384530  0.17656934
## lot     0.2442323 0.04079199  1.00000000 -0.1709961 -0.2138298 -0.03933975
## bath    0.1746578 0.47254061 -0.17099609  1.0000000  0.1468258  0.33452385
## bed    -0.2861975 0.23845303 -0.21382977  0.1468258  1.0000000 -0.36315062
## age     0.1541248 0.17656934 -0.03933975  0.3345239 -0.3631506  1.00000000
## garage  0.3583861 0.17315776  0.23581664  0.2017635 -0.4038039  0.53718464
##            garage
## price   0.3583861
## size    0.1731578
## lot     0.2358166
## bath    0.2017635
## bed    -0.4038039
## age     0.5371846
## garage  1.0000000
```

```r
plot(cor(numeric_data))
```




```r
pairs(numeric_data)
```




```
#Forward Stepwise Regression with R^2
model <- lm(price ~ ., data = homeprice2)
summary(ols_step_both_p(model))
```

```
##            Length Class      Mode
## orders      6     -none-     character
## method      6     -none-     character
## steps       1     -none-     numeric
## predictors  6     -none-     character
## rsquare     6     -none-     numeric
## aic         6     -none-     numeric
## sbc         6     -none-     numeric
## sbic        6     -none-     numeric
## adjr        6     -none-     numeric
## rmse        6     -none-     numeric
## mallows_cp  6     -none-     numeric
## indvar     12     -none-     character
## betas      27     -none-     numeric
## lbetas      6     -none-     numeric
## pvalues    27     -none-     numeric
## beta_pval   4     data.frame list
## model      12     lm         list
```

```
#Backward Regression
ols_step_backward_p(model, details = TRUE)
```

```
## Backward Elimination Method
```

```
## ---------------------------
##
## Candidate Terms:
##
## 1 . size
## 2 . lot
## 3 . bath
## 4 . bed
## 5 . age
## 6 . garage
## 7 . active
## 8 . edison
## 9 . harris
## 10 . adams
## 11 . crest
## 12 . parker
##
## We are eliminating variables based on p value...
##
## - crest
##
## Backward Elimination: Step 1
##
##  Variable crest Removed
##
##                           Model Summary
## ----------------------------------------------------------------
## R                       0.728       RMSE                44.757
## R-Squared               0.530       Coef. Var           15.661
## Adj. R-Squared          0.450       MSE               2003.210
## Pred R-Squared          0.273       MAE                 33.325
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F        Sig.
## -------------------------------------------------------------------
## Regression    144797.033        11      13163.367    6.571    0.0000
## Residual      128205.438        64       2003.210
## Total         273002.471        75
## -------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)     96.323        57.615                1.672    0.099    -18.776    211.422
##        size     69.119        30.164        0.243   2.291    0.025      8.858    129.379
##         lot     10.687         3.598        0.293   2.970    0.004      3.499     17.874
##        bath      4.739        11.608        0.045   0.408    0.684    -18.450     27.929
```

```
##         bed    -12.488         9.046       -0.153    -1.380    0.172    -30.561      5.584
##         age      1.619         3.298        0.063     0.491    0.625     -4.970      8.207
##      garage     10.292         8.957        0.132     1.149    0.255     -7.602     28.186
##      active     30.264        11.691        0.237     2.589    0.012      6.909     53.619
##      edison     80.121        17.164        0.487     4.668    0.000     45.833    114.410
##      harris     46.740        15.760        0.302     2.966    0.004     15.257     78.224
##       adams     -6.938        28.469       -0.023    -0.244    0.808    -63.810     49.935
##      parker    -18.737        14.550       -0.124    -1.288    0.202    -47.805     10.331
## ----------------------------------------------------------------------------------------
##
##
## - adams
##
## Backward Elimination: Step 2
##
##  Variable adams Removed
##
##                           Model Summary
## ----------------------------------------------------------------
## R                       0.728       RMSE                44.432
## R-Squared               0.530       Coef. Var           15.547
## Adj. R-Squared          0.458       MSE               1974.222
## Pred R-Squared          0.329       MAE                 33.406
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression    144678.066        10      14467.807    7.328    0.0000
## Residual      128324.405        65       1974.222
## Total         273002.471        75
## --------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)     92.232        54.714                1.686    0.097    -17.040    201.504
##        size     69.758        29.832        0.246     2.338    0.022     10.179    129.336
##         lot     10.709         3.571        0.293     2.999    0.004      3.578     17.840
##        bath      4.993        11.477        0.047     0.435    0.665    -17.928     27.915
##         bed    -12.189         8.898       -0.149    -1.370    0.175    -29.959      5.581
##         age      1.609         3.274        0.063     0.492    0.625     -4.929      8.147
##      garage     10.636         8.781        0.136     1.211    0.230     -6.902     28.173
##      active     30.267        11.606        0.237     2.608    0.011      7.088     53.445
##      edison     80.856        16.774        0.492     4.820    0.000     47.357    114.356
##      harris     47.322        15.464        0.306     3.060    0.003     16.438     78.207
##      parker    -18.154        14.248       -0.121    -1.274    0.207    -46.609     10.301
## ----------------------------------------------------------------------------------------
```

```
##
##
## - bath
##
## Backward Elimination: Step 3
##
##  Variable bath Removed
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.727       RMSE                44.158
## R-Squared               0.529       Coef. Var           15.451
## Adj. R-Squared          0.464       MSE               1949.971
## Pred R-Squared          0.354       MAE                 33.234
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F        Sig.
## -------------------------------------------------------------------------
## Regression    144304.369         9      16033.819    8.223    0.0000
## Residual      128698.102        66       1949.971
## Total         273002.471        75
## -------------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)     90.505        54.234                1.669    0.100    -17.776    198.786
##        size     75.051        27.070        0.264   2.773    0.007     21.005    129.098
##         lot     10.540         3.528        0.289   2.988    0.004      3.497     17.584
##         bed    -11.421         8.667       -0.140  -1.318    0.192    -28.725      5.883
##         age      1.938         3.166        0.075   0.612    0.542     -4.382      8.259
##      garage     10.875         8.710        0.139   1.249    0.216     -6.516     28.265
##      active     30.638        11.503        0.240   2.663    0.010      7.671     53.605
##      edison     80.965        16.669        0.493   4.857    0.000     47.685    114.245
##      harris     47.939        15.305        0.310   3.132    0.003     17.382     78.495
##      parker    -19.489        13.828       -0.129  -1.409    0.163    -47.097      8.118
## ----------------------------------------------------------------------------------------------
##
##
## - age
##
## Backward Elimination: Step 4
##
##  Variable age Removed
##
##                          Model Summary
## ----------------------------------------------------------------
```

```
## R                       0.725       RMSE                 43.952
## R-Squared               0.526       Coef. Var            15.379
## Adj. R-Squared          0.469       MSE                1931.781
## Pred R-Squared          0.377       MAE                  33.358
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F        Sig.
## ---------------------------------------------------------------------
## Regression    143573.123         8      17946.640     9.29    0.0000
## Residual      129429.348        67       1931.781
## Total         273002.471        75
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)     87.956        53.821                 1.634    0.107    -19.471    195.384
##        size     77.850        26.557        0.274    2.931    0.005     24.843    130.857
##         lot     10.116         3.443        0.277    2.938    0.005      3.244     16.988
##         bed    -12.788         8.335       -0.156   -1.534    0.130    -29.425      3.849
##      garage     13.096         7.882        0.168    1.661    0.101     -2.638     28.829
##      active     30.960        11.437        0.243    2.707    0.009      8.131     53.789
##      edison     76.789        15.139        0.467    5.072    0.000     46.572    107.006
##      harris     49.560        15.003        0.321    3.303    0.002     19.613     79.507
##      parker    -18.973        13.737       -0.126   -1.381    0.172    -46.393      8.447
## ----------------------------------------------------------------------------------------------
##
##
##
## No more variables satisfy the condition of p value = 0.3
##
##
## Variables Removed:
##
## - crest
## - adams
## - bath
## - age
##
##
## Final Model Output
## ------------------
##
##                           Model Summary
## ----------------------------------------------------------------
## R                       0.725       RMSE                 43.952
## R-Squared               0.526       Coef. Var            15.379
```

```
## Adj. R-Squared          0.469       MSE                1931.781
## Pred R-Squared          0.377       MAE                  33.358
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F        Sig.
## ---------------------------------------------------------------------
## Regression    143573.123         8      17946.640     9.29    0.0000
## Residual      129429.348        67       1931.781
## Total         273002.471        75
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ------------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ------------------------------------------------------------------------------------------------
## (Intercept)     87.956        53.821                1.634    0.107    -19.471     195.384
##        size     77.850        26.557        0.274   2.931    0.005     24.843     130.857
##         lot     10.116         3.443        0.277   2.938    0.005      3.244      16.988
##         bed    -12.788         8.335       -0.156  -1.534    0.130    -29.425       3.849
##      garage     13.096         7.882        0.168   1.661    0.101     -2.638      28.829
##      active     30.960        11.437        0.243   2.707    0.009      8.131      53.789
##      edison     76.789        15.139        0.467   5.072    0.000     46.572     107.006
##      harris     49.560        15.003        0.321   3.303    0.002     19.613      79.507
##      parker    -18.973        13.737       -0.126  -1.381    0.172    -46.393       8.447
## ------------------------------------------------------------------------------------------------

##
##
##                            Elimination Summary
## ---------------------------------------------------------------------------
##         Variable                  Adj.
## Step    Removed     R-Square    R-Square     C(p)        AIC         RMSE
## ---------------------------------------------------------------------------
##    1    crest         0.5304      0.4497    11.0134    806.4085    44.7572
##    2    adams          0.530      0.4576     9.0719    804.4790    44.4322
##    3    bath          0.5286      0.4643     7.2556    802.7000    44.1585
##    4    age           0.5259      0.4693     5.6150    801.1306    43.9520
## ---------------------------------------------------------------------------
```

```
#Forward Regression
model2 <- lm(price ~ ., data = homeprice2)
ols_step_forward_p(model2, details = TRUE)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. size
```

```
## 2. lot
## 3. bath
## 4. bed
## 5. age
## 6. garage
## 7. active
## 8. edison
## 9. harris
## 10. adams
## 11. crest
## 12. parker
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - garage
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.358       RMSE                56.704
## R-Squared               0.128       Coef. Var           19.841
## Adj. R-Squared          0.117       MSE               3215.377
## Pred R-Squared          0.082       MAE                 45.622
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     35064.597         1      35064.597    10.905    0.0015
## Residual      237937.874        74       3215.377
## Total         273002.471        75
## --------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)    241.927        14.791                16.356    0.000    212.455    271.399
##      garage     28.017         8.484        0.358     3.302    0.001     11.112     44.922
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 2
##
## - edison
##
```

```
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.514       RMSE                 52.462
## R-Squared               0.264       Coef. Var            18.357
## Adj. R-Squared          0.244       MSE                2752.304
## Pred R-Squared          0.195       MAE                  42.180
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------------
## Regression     72084.282         2      36042.141    13.095    0.0000
## Residual      200918.189        73       2752.304
## Total         273002.471        75
## -------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    224.044        14.527                15.422    0.000    195.091    252.997
##      garage     33.235         7.977        0.425     4.166    0.000     17.336     49.134
##      edison     61.512        16.772        0.374     3.667    0.000     28.085     94.940
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 3
##
## - harris
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.590       RMSE                 49.696
## R-Squared               0.349       Coef. Var            17.389
## Adj. R-Squared          0.322       MSE                2469.687
## Pred R-Squared          0.273       MAE                  37.529
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------------
## Regression     95184.994         3      31728.331    12.847    0.0000
## Residual      177817.477        72       2469.687
```

```
## Total         273002.471        75
## ----------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    217.842        13.910                15.661    0.000    190.113    245.571
##      garage     30.831         7.597        0.394     4.058    0.000     15.686     45.976
##      edison     70.718        16.171        0.430     4.373    0.000     38.483    102.954
##      harris     46.205        15.108        0.299     3.058    0.003     16.088     76.322
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 4
##
## - lot
##
##                          Model Summary
## ------------------------------------------------------------------
## R                       0.648       RMSE                47.232
## R-Squared               0.420       Coef. Var           16.526
## Adj. R-Squared          0.387       MSE               2230.851
## Pred R-Squared          0.309       MAE                 37.651
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## ----------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression    114612.033         4      28653.008    12.844    0.0000
## Residual      158390.438        71       2230.851
## Total         273002.471        75
## ----------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    182.243        17.897                10.183    0.000    146.557    217.929
##      garage     24.594         7.524        0.315     3.269    0.002      9.592     39.596
##      edison     72.519        15.381        0.441     4.715    0.000     41.850    103.187
##      harris     61.082        15.218        0.395     4.014    0.000     30.739     91.426
##         lot     10.620         3.599        0.291     2.951    0.004      3.444     17.796
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 5
```

```
##
## - active
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.673       RMSE                46.178
## R-Squared               0.453       Coef. Var           16.158
## Adj. R-Squared          0.414       MSE               2132.372
## Pred R-Squared          0.320       MAE                 36.059
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------------
## Regression    123736.401         5      24747.280    11.606    0.0000
## Residual      149266.070        70       2132.372
## Total         273002.471        75
## -------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t       Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    176.818        17.693                 9.994    0.000    141.531    212.106
##      garage     21.838         7.475        0.279    2.921    0.005      6.929     36.747
##      edison     78.438        15.307        0.477    5.124    0.000     47.909    108.968
##      harris     60.278        14.883        0.390    4.050    0.000     30.594     89.962
##         lot     10.855         3.520        0.297    3.083    0.003      3.833     17.876
##      active     24.376        11.784        0.191    2.069    0.042      0.874     47.879
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 6
##
## - size
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.703       RMSE                44.735
## R-Squared               0.494       Coef. Var           15.653
## Adj. R-Squared          0.450       MSE               2001.184
## Pred R-Squared          0.360       MAE                 34.954
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
```

```
## ----------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression    134920.786         6      22486.798    11.237    0.0000
## Residual      138081.685        69       2001.184
## Total         273002.471        75
## ----------------------------------------------------------------------
##
##                                  Parameter Estimates
## ------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t       Sig       lower      upper
## ------------------------------------------------------------------------------------------
## (Intercept)    62.336        51.369                1.213    0.229    -40.143    164.816
##      garage    18.674         7.364        0.239    2.536    0.013      3.983     33.366
##      edison    82.059        14.908        0.499    5.504    0.000     52.318    111.799
##      harris    59.962        14.419        0.388    4.159    0.000     31.197     88.727
##         lot    10.877         3.410        0.298    3.189    0.002      4.073     17.680
##      active    29.305        11.605        0.230    2.525    0.014      6.154     52.455
##        size    59.487        25.163        0.209    2.364    0.021      9.288    109.685
## ------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 7
##
## - bed
##
##                           Model Summary
## ----------------------------------------------------------------
## R                        0.716       RMSE                44.244
## R-Squared                0.512       Coef. Var           15.481
## Adj. R-Squared           0.462       MSE               1957.562
## Pred R-Squared           0.368       MAE                 33.751
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## ----------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression    139888.226         7      19984.032    10.209    0.0000
## Residual      133114.245        68       1957.562
## Total         273002.471        75
## ----------------------------------------------------------------------
##
##                                  Parameter Estimates
## ------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ------------------------------------------------------------------------------------------
## (Intercept)     91.950        54.101                1.700    0.094    -16.006    199.907
```

```
##      garage     13.712         7.922         0.175     1.731    0.088     -2.096     29.520
##      edison     82.074        14.745         0.499     5.566    0.000     52.651    111.496
##      harris     54.513        14.665         0.353     3.717    0.000     25.248     83.777
##         lot      9.703         3.453         0.266     2.810    0.006      2.813     16.592
##      active     29.783        11.481         0.233     2.594    0.012      6.872     52.694
##        size     74.562        26.626         0.263     2.800    0.007     21.432    127.693
##         bed    -13.350         8.381        -0.163    -1.593    0.116    -30.074      3.373
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 8
##
## - parker
##
##                           Model Summary
## ------------------------------------------------------------------
## R                        0.725       RMSE                  43.952
## R-Squared                0.526       Coef. Var             15.379
## Adj. R-Squared           0.469       MSE                 1931.781
## Pred R-Squared           0.377       MAE                   33.358
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F        Sig.
## ---------------------------------------------------------------------
## Regression    143573.123         8      17946.640     9.29    0.0000
## Residual      129429.348        67       1931.781
## Total         273002.471        75
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)     87.956        53.821                 1.634    0.107    -19.471    195.384
##      garage     13.096         7.882         0.168     1.661    0.101     -2.638     28.829
##      edison     76.789        15.139         0.467     5.072    0.000     46.572    107.006
##      harris     49.560        15.003         0.321     3.303    0.002     19.613     79.507
##         lot     10.116         3.443         0.277     2.938    0.005      3.244     16.988
##      active     30.960        11.437         0.243     2.707    0.009      8.131     53.789
##        size     77.850        26.557         0.274     2.931    0.005     24.843    130.857
##         bed    -12.788         8.335        -0.156    -1.534    0.130    -29.425      3.849
##      parker    -18.973        13.737        -0.126    -1.381    0.172    -46.393      8.447
## ----------------------------------------------------------------------------------------------
##
##
##
## No more variables to be added.
```

```
##
## Variables Entered:
##
## + garage
## + edison
## + harris
## + lot
## + active
## + size
## + bed
## + parker
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.725       RMSE                43.952
## R-Squared               0.526       Coef. Var           15.379
## Adj. R-Squared          0.469       MSE               1931.781
## Pred R-Squared          0.377       MAE                 33.358
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression    143573.123         8      17946.640     9.29    0.0000
## Residual      129429.348        67       1931.781
## Total         273002.471        75
## --------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)     87.956        53.821                 1.634    0.107    -19.471    195.384
##      garage     13.096         7.882        0.168     1.661    0.101     -2.638     28.829
##      edison     76.789        15.139        0.467     5.072    0.000     46.572    107.006
##      harris     49.560        15.003        0.321     3.303    0.002     19.613     79.507
##         lot     10.116         3.443        0.277     2.938    0.005      3.244     16.988
##      active     30.960        11.437        0.243     2.707    0.009      8.131     53.789
##        size     77.850        26.557        0.274     2.931    0.005     24.843    130.857
##         bed    -12.788         8.335       -0.156    -1.534    0.130    -29.425      3.849
##      parker    -18.973        13.737       -0.126    -1.381    0.172    -46.393      8.447
## ----------------------------------------------------------------------------------------
##
##                            Selection Summary
```

```
## ----------------------------------------------------------------------
##         Variable                 Adj.
## Step    Entered     R-Square    R-Square     C(p)        AIC        RMSE
## ----------------------------------------------------------------------
##    1    garage        0.1284      0.1167    44.9473    833.4051    56.7043
##    2    edison        0.2640      0.2439    28.7520    822.5526    52.4624
##    3    harris        0.3487      0.3215    19.3979    815.2699    49.6959
##    4    lot           0.4198      0.3871    11.8495    808.4771    47.2319
##    5    active        0.4532      0.4142     9.3648    805.9678    46.1776
##    6    size          0.4942      0.4502     5.8676    802.0486    44.7346
##    7    bed           0.5124      0.4622     5.4261    801.2641    44.2443
##    8    parker        0.5259      0.4693     5.6150    801.1306    43.9520
## ----------------------------------------------------------------------
```

```
#Preliminary Model
model2 <- lm(price~garage+edison+harris+lot+active+size+age,data=homeprice2)
additive_mod <- lm(price~garage+edison+harris+lot+active+size+age+agesq,data=homeprice)
anova(additive_mod)
```

```
## Analysis of Variance Table
##
## Response: price
##           Df Sum Sq Mean Sq F value    Pr(>F)
## garage     1  35065   35065 18.4348 5.801e-05 ***
## edison     1  37020   37020 19.4627 3.813e-05 ***
## harris     1  23101   23101 12.1449 0.0008728 ***
## lot        1  19427   19427 10.2135 0.0021274 **
## active     1   9124    9124  4.7970 0.0319937 *
## size       1  11184   11184  5.8801 0.0180160 *
## age        1   1638    1638  0.8613 0.3567183
## agesq      1   9004    9004  4.7336 0.0331109 *
## Residuals 67 127440    1902
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#F-test with age and age^2
age_only <- lm(price~age,data=homeprice)
withagesq <- lm(price~age + agesq,data=homeprice)
anova(withagesq)
```

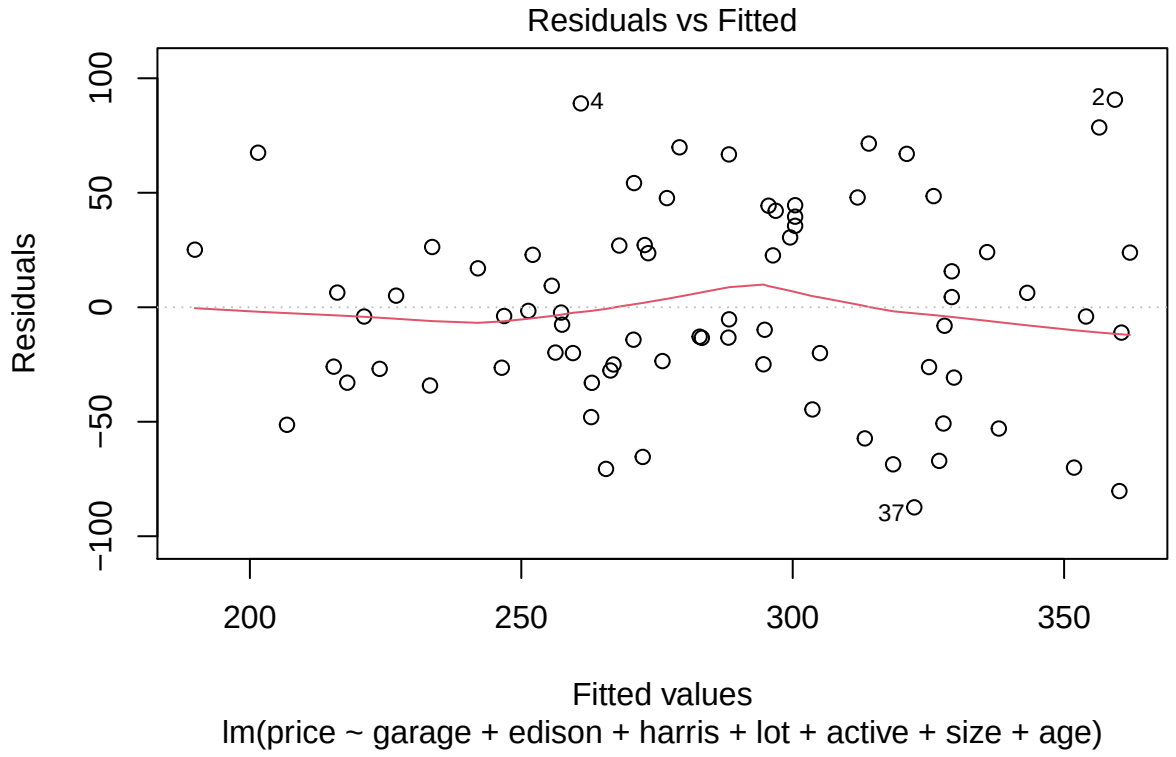
```
## Analysis of Variance Table
##
## Response: price
##           Df Sum Sq Mean Sq F value   Pr(>F)
## age        1   6485    6485  2.0325 0.158231
## agesq      1  33598   33598 10.5299 0.001774 **
## Residuals 73 232920    3191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

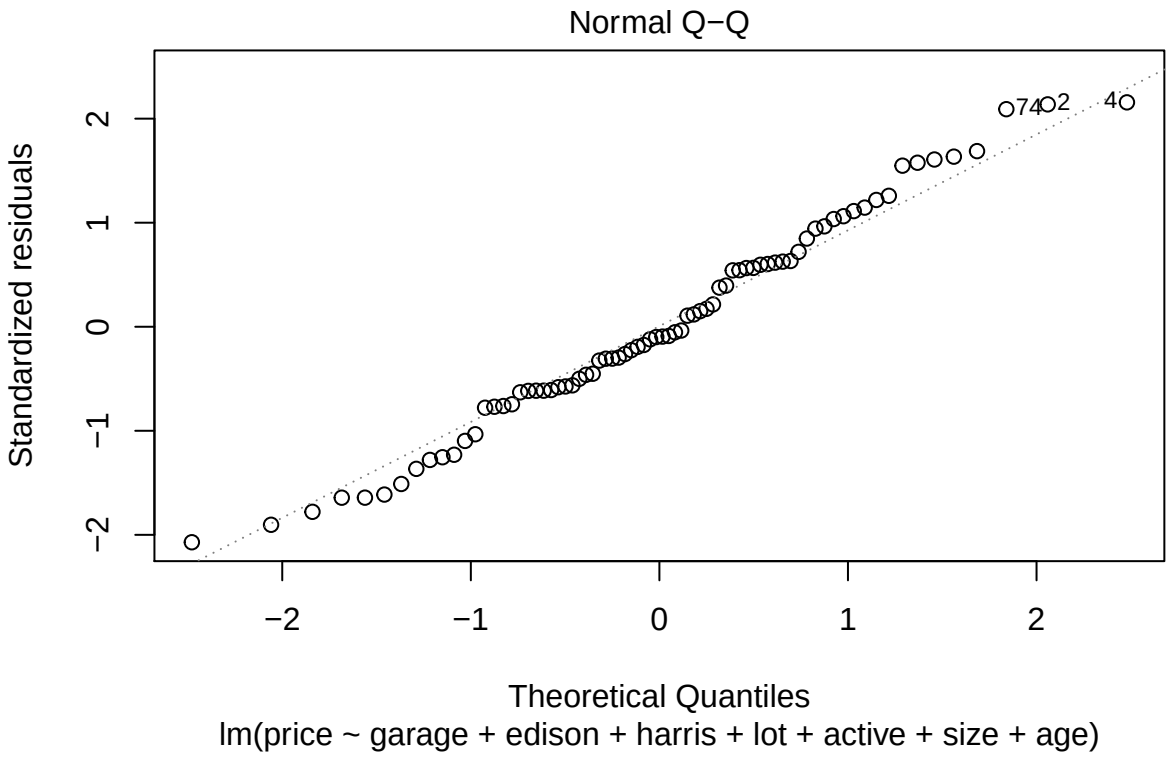
```
Fcrit <- qf(0.95, 1, 73)
summary(model2)
```

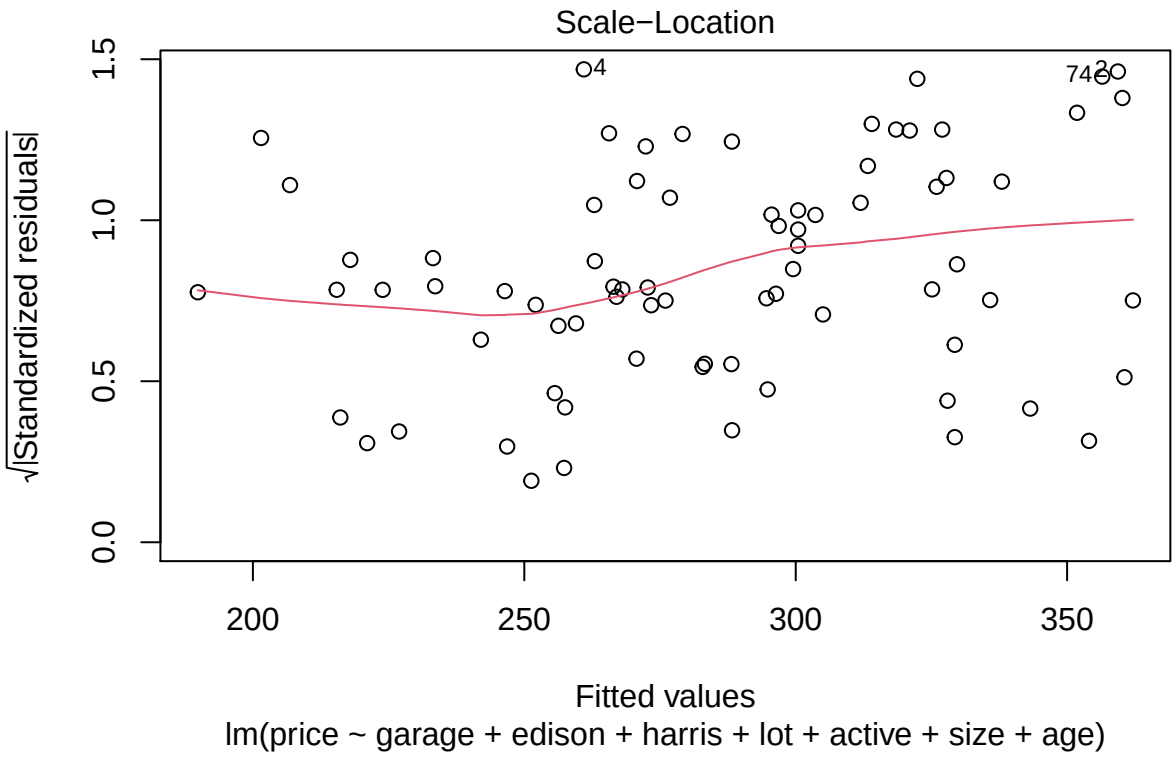
```
##
## Call:
## lm(formula = price ~ garage + edison + harris + lot + active +
```

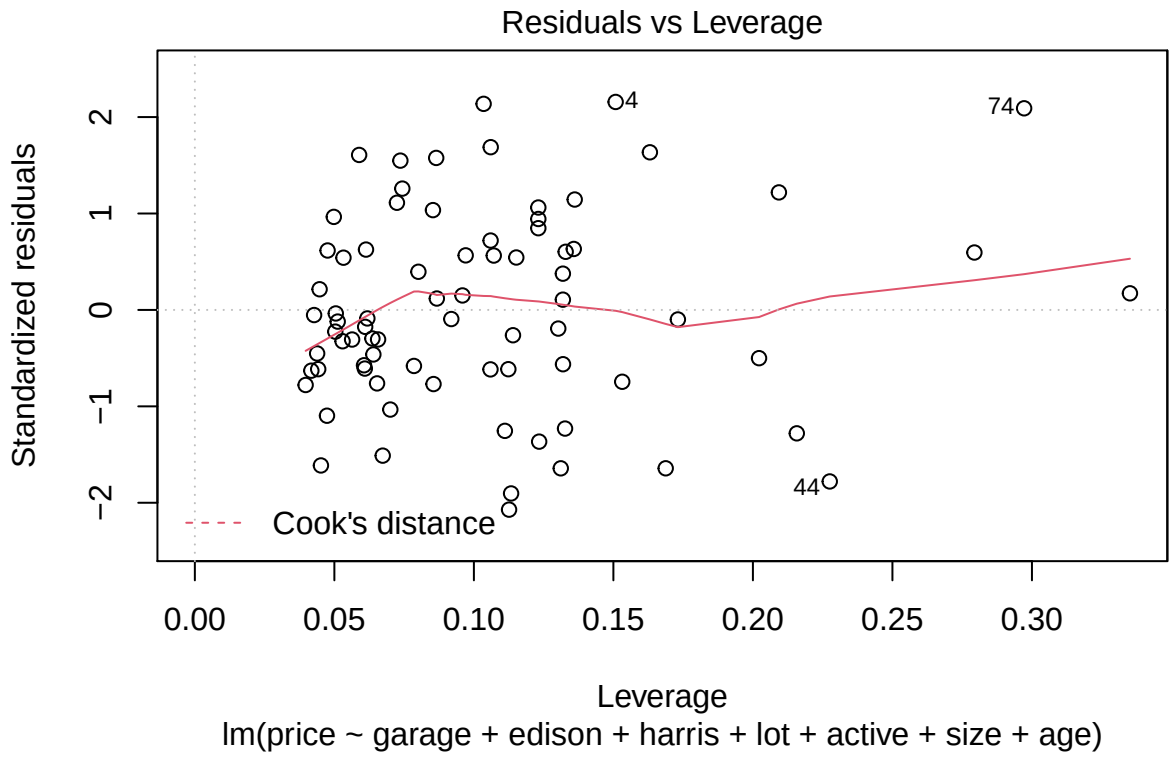
```
##     size + age, data = homeprice2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -87.396 -26.532  -4.041  26.999  90.647
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   70.507     52.227   1.350 0.181485
## garage        14.765      8.550   1.727 0.088740 .
## edison        88.302     16.449   5.368 1.04e-06 ***
## harris        57.017     14.801   3.852 0.000262 ***
## lot           11.302      3.447   3.279 0.001646 **
## active        28.863     11.630   2.482 0.015552 *
## size          57.520     25.290   2.274 0.026102 *
## age            2.801      3.099   0.904 0.369416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.79 on 68 degrees of freedom
## Multiple R-squared:  0.5002, Adjusted R-squared:  0.4488
## F-statistic: 9.723 on 7 and 68 DF,  p-value: 2.507e-08
```

```
plot(model2)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(price ~ garage + edison + harris + lot + active + size + age)
74
2
4


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
200
250
300
350
4
74
3
Fitted values
lm(price ~ garage + edison + harris + lot + active + size + age)

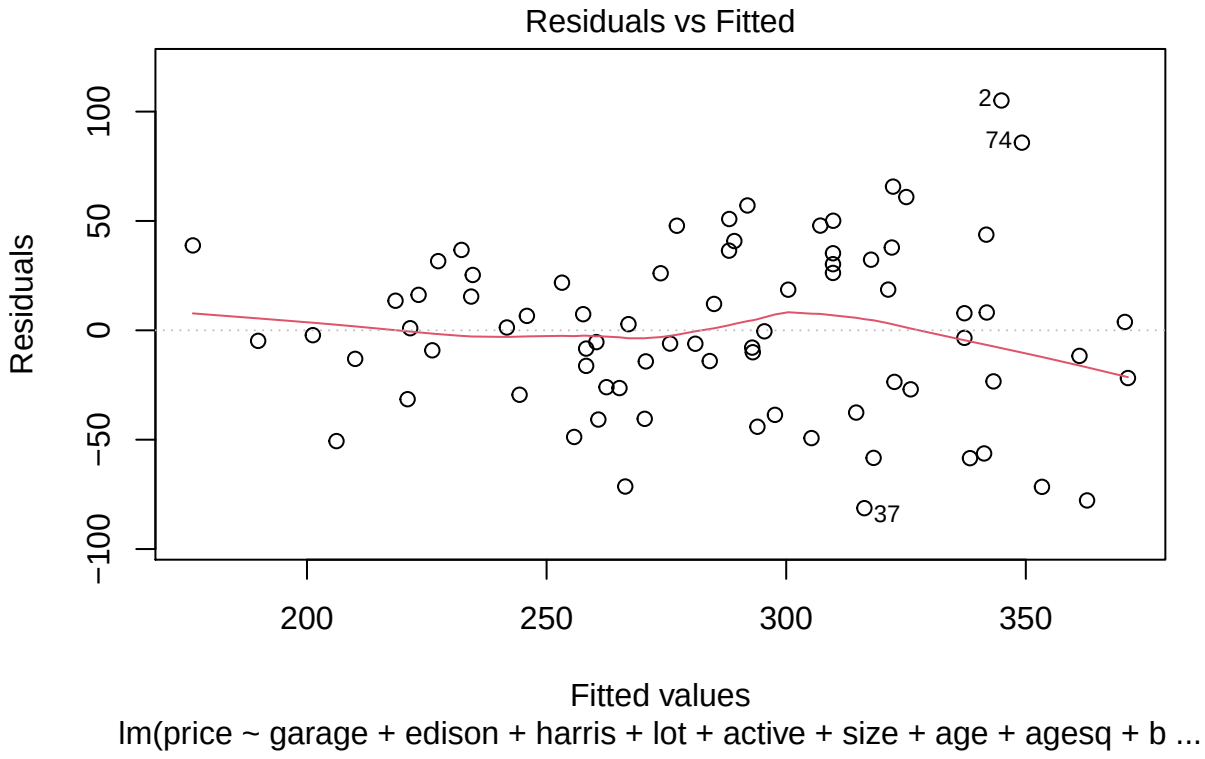


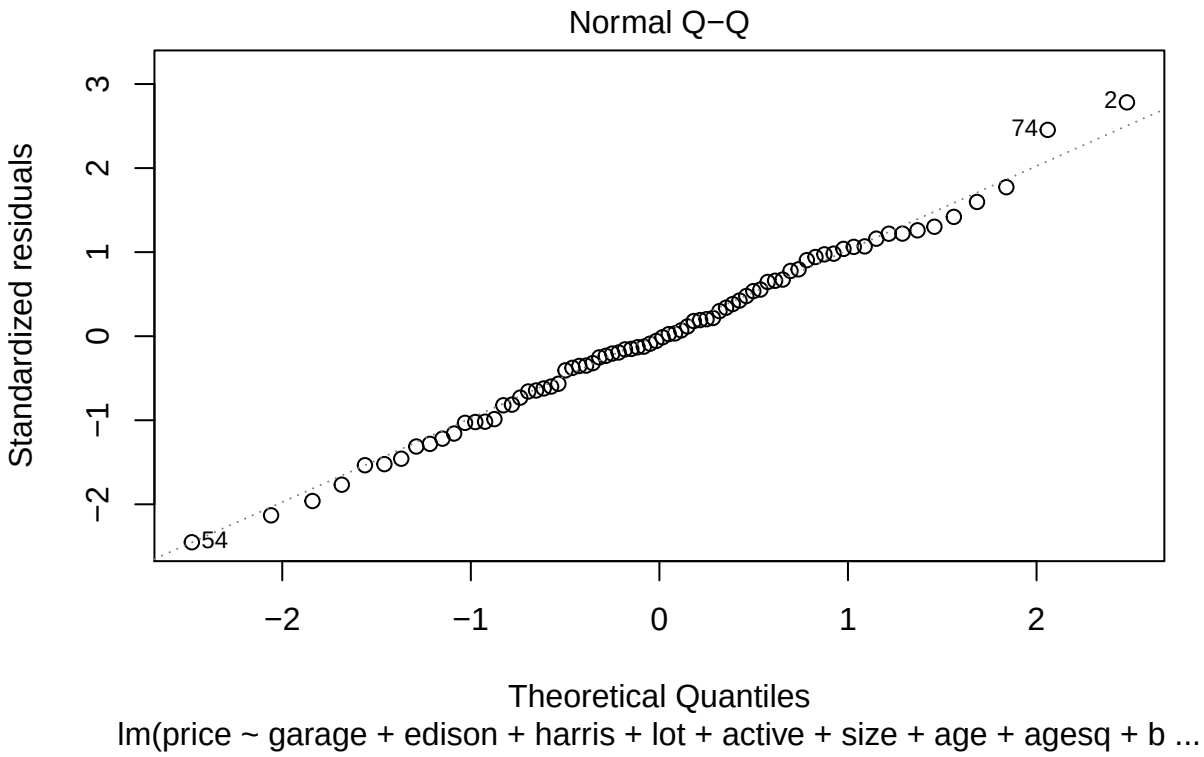

```
#Final Model
final_mod <- lm(price~garage+edison+harris+lot+active+size+age+agesq+bath+bed+bathbed,data=homeprice)
summary(final_mod)
```

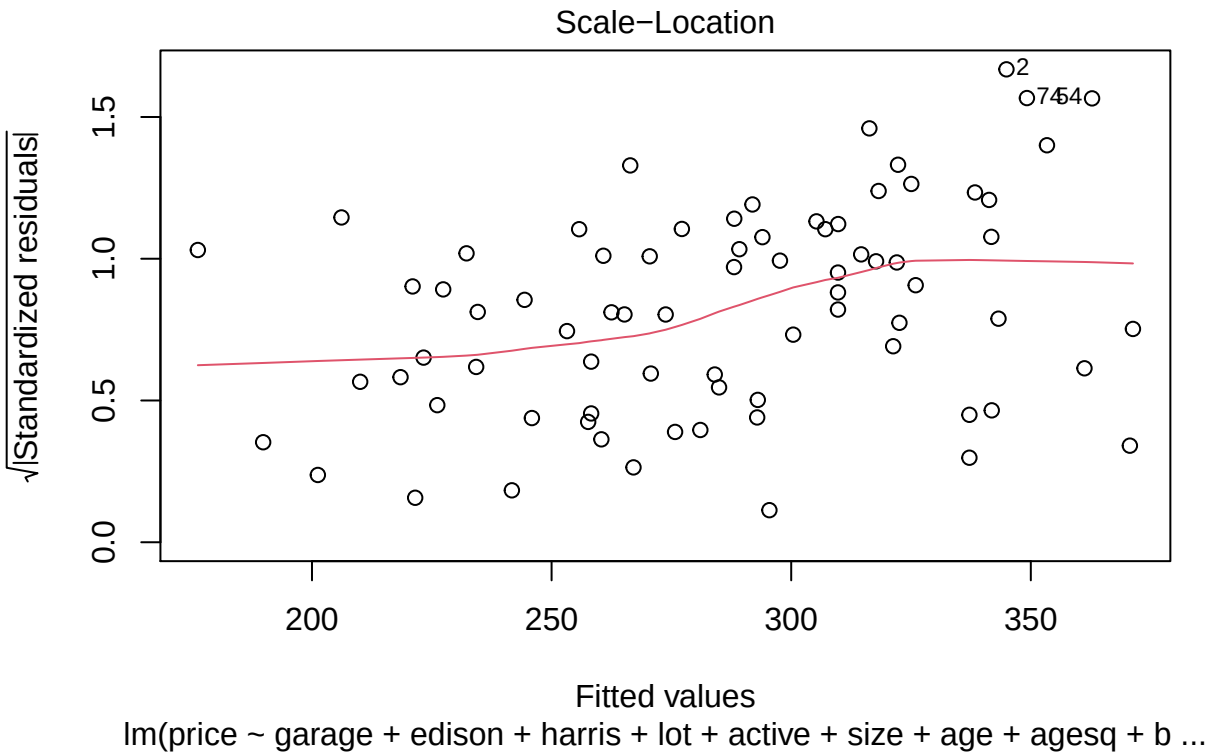
```
##
## Call:
## lm(formula = price ~ garage + edison + harris + lot + active +
##     size + age + agesq + bath + bed + bathbed, data = homeprice)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -81.310 -26.101  -1.337  27.236 105.083
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 332.4781   106.5992   3.119 0.002720 **
## garage       13.1185     8.2846   1.583 0.118239
## edison       67.0619    16.8220   3.987 0.000175 ***
## harris       47.2732    14.8443   3.185 0.002240 **
## lot           9.9168     3.4379   2.885 0.005335 **
## active       27.4236    10.9885   2.496 0.015160 *
## size         56.7190    27.9738   2.028 0.046772 *
## age           3.3014     3.1695   1.042 0.301507
## agesq         1.6413     0.7333   2.238 0.028683 *
## bath        -98.1564    42.6665  -2.301 0.024686 *
## bed         -78.9099    27.7519  -2.843 0.005985 **
```

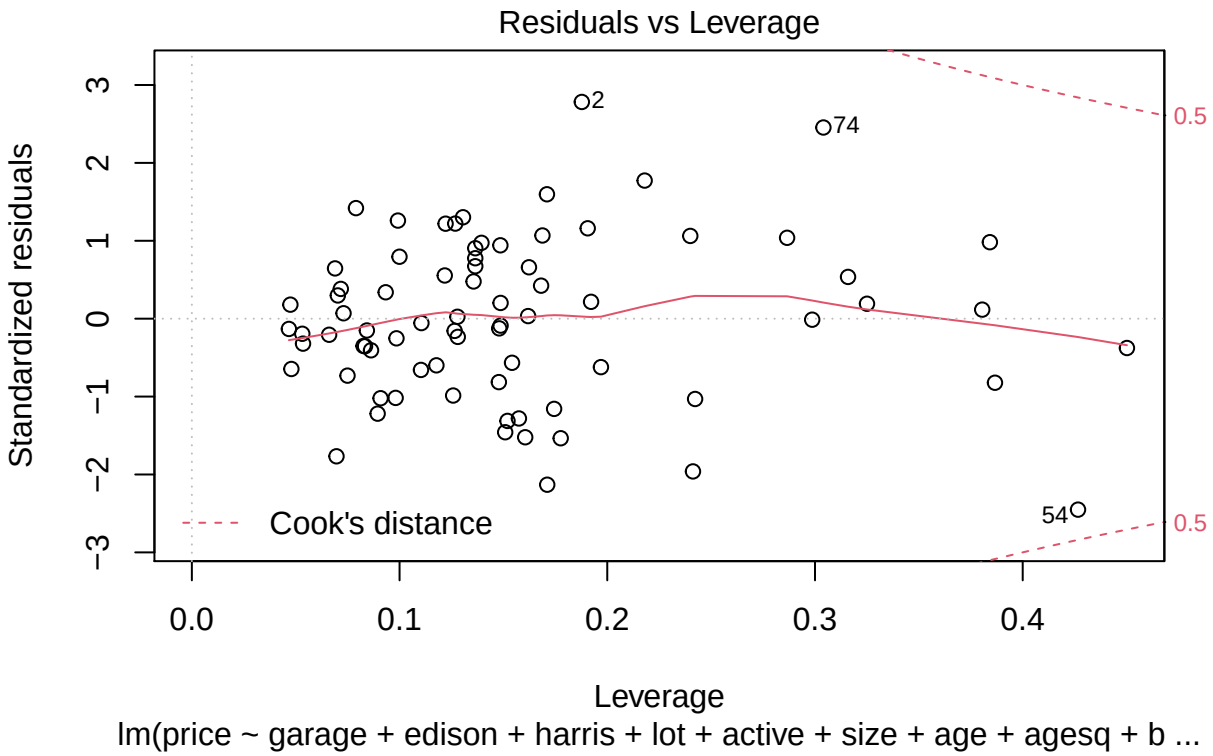
```
## bathbed      30.3901    11.8781   2.559 0.012889 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.91 on 64 degrees of freedom
## Multiple R-squared:  0.5883, Adjusted R-squared:  0.5175
## F-statistic: 8.314 on 11 and 64 DF,  p-value: 7.698e-09
```

```
plot(final_mod)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(price ~ garage + edison + harris + lot + active + size + age + agesq + b ...
54
74
2
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
200
250
300
350
2
74
54
Fitted values
lm(price ~ garage + edison + harris + lot + active + size + age + agesq + b ...



The code in this box is the code used to plot the residuals vs fitted values of our final model, the code run for our outlier analysis, and the code used to run our AIC and BIC model selection.