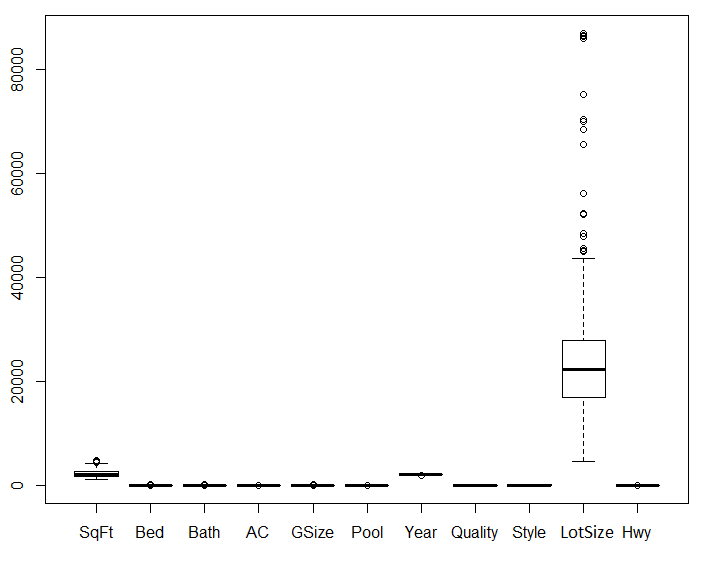
PALLAVI KARAN

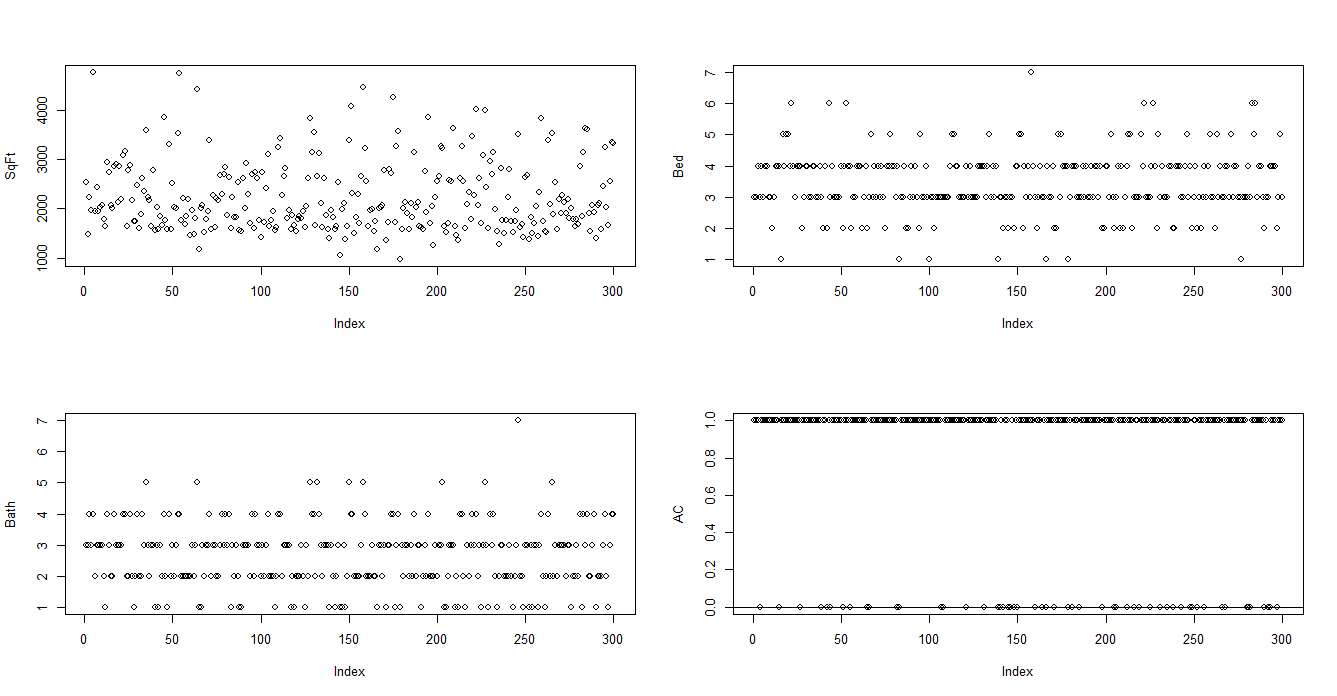
ASSIGNMENT NO: 6

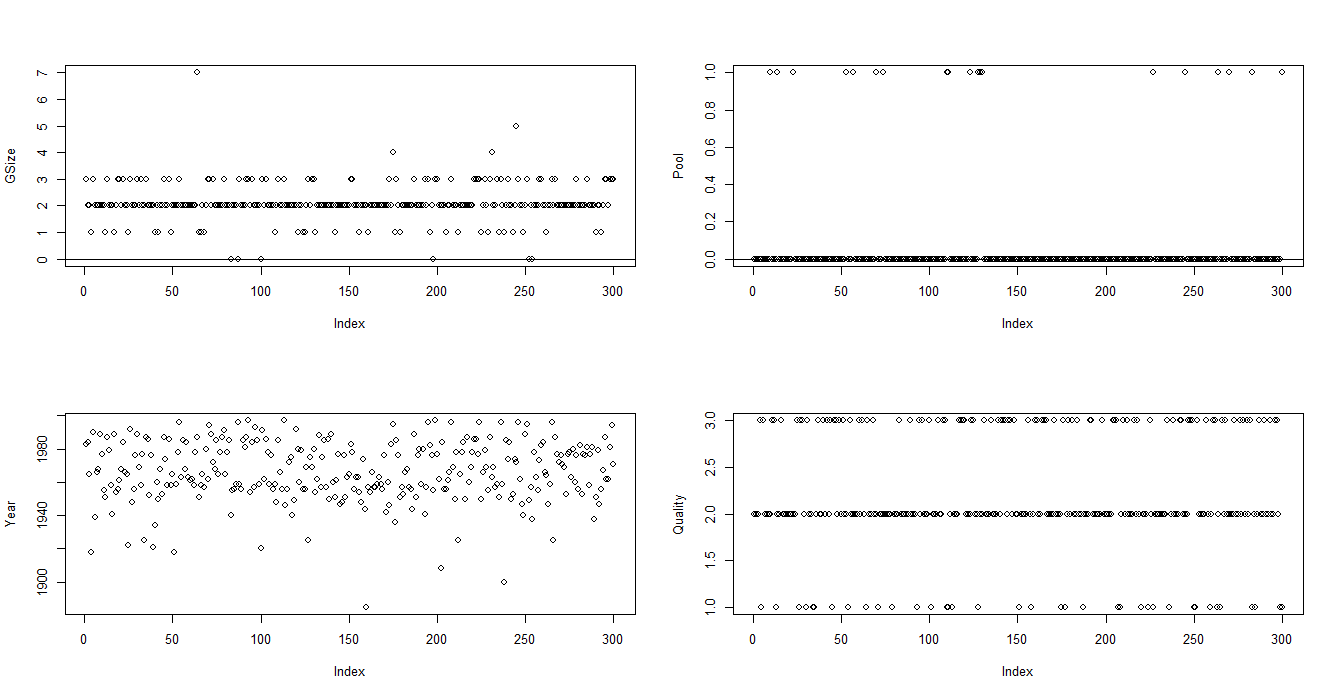
BOXPLOT

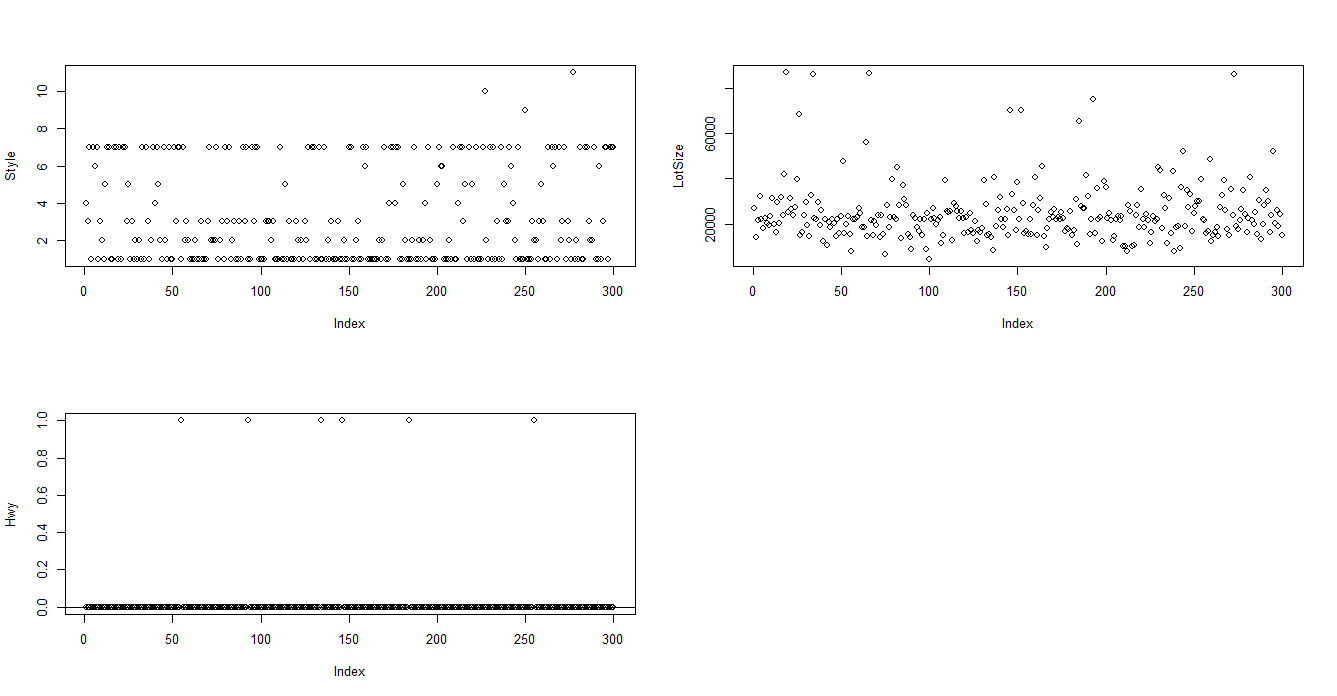


SqFt has an extreme outlier, but LotSize has a lot of extreme outliers.

SCATTERPLOT







The data points for any predictor doesn’t look linear.

Co-RELATION MATRIX

> cor(sampled.data)

Price SqFt Bed Bath AC GSize Pool Year

Price 1.0000000 0.82417050 0.49997728 0.7249379 0.29250376 0.56545649 0.19326417 0.526215516

SqFt 0.8241705 1.00000000 0.58235262 0.7805503 0.29923014 0.52777327 0.18991496 0.417445248

Bed 0.4999773 0.58235262 1.00000000 0.6031824 0.27363533 0.35569922 0.11128635 0.320982948

Bath 0.7249379 0.78055032 0.60318245 1.0000000 0.35920388 0.49190273 0.22429507 0.503674723

AC 0.2925038 0.29923014 0.27363533 0.3592039 1.00000000 0.32598379 0.12182964 0.396122050

GSize 0.5654565 0.52777327 0.35569922 0.4919027 0.32598379 1.00000000 0.13742376 0.444711141

Pool 0.1932642 0.18991496 0.11128635 0.2242951 0.12182964 0.13742376 1.00000000 0.095520519

Year 0.5262155 0.41744525 0.32098295 0.5036747 0.39612205 0.44471114 0.09552052 1.000000000

Quality -0.7538074 -0.68813462 -0.44846559 -0.7038145 -0.42204109 -0.53214802 -0.20449395 -0.570281611

Style 0.3659703 0.64286579 0.35403915 0.4941841 0.14216908 0.22919584 0.08541718 0.197986928

LotSize 0.2648146 0.13896230 0.11488214 0.1381109 -0.09292171 0.17351628 -0.08645407 -0.092818821

Hwy -0.0557220 -0.06529436 0.01119307 -0.0407855 -0.05701582 0.01455651 -0.03714716 0.002770024

Quality Style LotSize Hwy

Price -0.75380740 0.36597031 0.264814601 -0.055721999

SqFt -0.68813462 0.64286579 0.138962299 -0.065294357

Bed -0.44846559 0.35403915 0.114882144 0.011193070

Bath -0.70381446 0.49418412 0.138110863 -0.040785497

AC -0.42204109 0.14216908 -0.092921710 -0.057015818

GSize -0.53214802 0.22919584 0.173516276 0.014556506

Pool -0.20449395 0.08541718 -0.086454065 -0.037147159

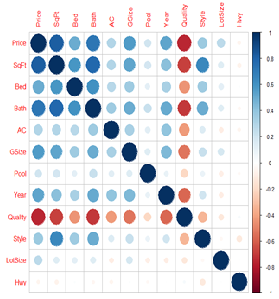
Year -0.57028161 0.19798693 -0.092818821 0.002770024

Quality 1.00000000 -0.33986808 -0.121548682 0.035741093

Style -0.33986808 1.00000000 -0.038570630 -0.115077220

LotSize -0.12154868 -0.03857063 1.000000000 0.008606401

Hwy 0.03574109 -0.11507722 0.008606401 1.000000000



There is max co-relation is between Price(dependent) and SqFt(independent). The is moderate co-relation between Bath-SqFt, style-SqFt.

FULL MODEL:

Call:

lm(formula = Price ~ SqFt + Bed + Bath + factor(AC) + GSize +

factor(Pool) + Year + factor(Quality) + factor(Style) + LotSize +

factor(Hwy), data = train)

Residuals:

Min 1Q Median 3Q Max

-216133 -28539 -2336 23716 276335

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.180e+06 5.441e+05 -4.006 7.92e-05 \*\*\*

SqFt 9.717e+01 1.035e+01 9.393 < 2e-16 \*\*\*

Bed -6.679e+03 4.636e+03 -1.441 0.15078

Bath 7.282e+03 5.840e+03 1.247 0.21351

factor(AC)1 -9.349e+02 1.084e+04 -0.086 0.93130

GSize 1.687e+04 6.791e+03 2.484 0.01359 \*

factor(Pool)1 1.755e+04 1.381e+04 1.271 0.20469

Year 1.168e+03 2.772e+02 4.212 3.42e-05 \*\*\*

factor(Quality)2 -1.306e+05 1.440e+04 -9.068 < 2e-16 \*\*\*

factor(Quality)3 -1.405e+05 1.930e+04 -7.284 3.33e-12 \*\*\*

factor(Style)2 -2.480e+04 1.239e+04 -2.001 0.04633 \*

factor(Style)3 -2.287e+04 1.243e+04 -1.839 0.06696 .

factor(Style)4 2.498e+04 2.364e+04 1.057 0.29165

factor(Style)5 -4.373e+04 2.295e+04 -1.905 0.05782 .

factor(Style)6 -1.980e+03 2.432e+04 -0.081 0.93519

factor(Style)7 -3.171e+04 1.204e+04 -2.633 0.00894 \*\*

factor(Style)9 -7.719e+04 6.065e+04 -1.273 0.20419

factor(Style)10 -7.471e+04 6.266e+04 -1.192 0.23419

factor(Style)11 -1.032e+05 6.079e+04 -1.698 0.09063 .

LotSize 1.753e+00 3.163e-01 5.544 6.86e-08 \*\*\*

factor(Hwy)1 -5.984e+04 3.040e+04 -1.969 0.04997 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 59330 on 279 degrees of freedom

Multiple R-squared: 0.8361, Adjusted R-squared: 0.8243

F-statistic: 71.16 on 20 and 279 DF, p-value: < 2.2e-16

> anova(lm.fit.full)

Analysis of Variance Table

Response: Price

Df Sum Sq Mean Sq F value Pr(>F)

SqFt 1 4.0554e+12 4.0554e+12 1152.2268 < 2.2e-16 \*\*\*

Bed 1 9.7688e+09 9.7688e+09 2.7755 0.096838 .

Bath 1 7.2424e+10 7.2424e+10 20.5771 8.516e-06 \*\*\*

factor(AC) 1 5.3206e+08 5.3206e+08 0.1512 0.697718

GSize 1 1.5815e+11 1.5815e+11 44.9337 1.126e-10 \*\*\*

factor(Pool) 1 2.0812e+09 2.0812e+09 0.5913 0.442568

Year 1 7.4445e+10 7.4445e+10 21.1515 6.443e-06 \*\*\*

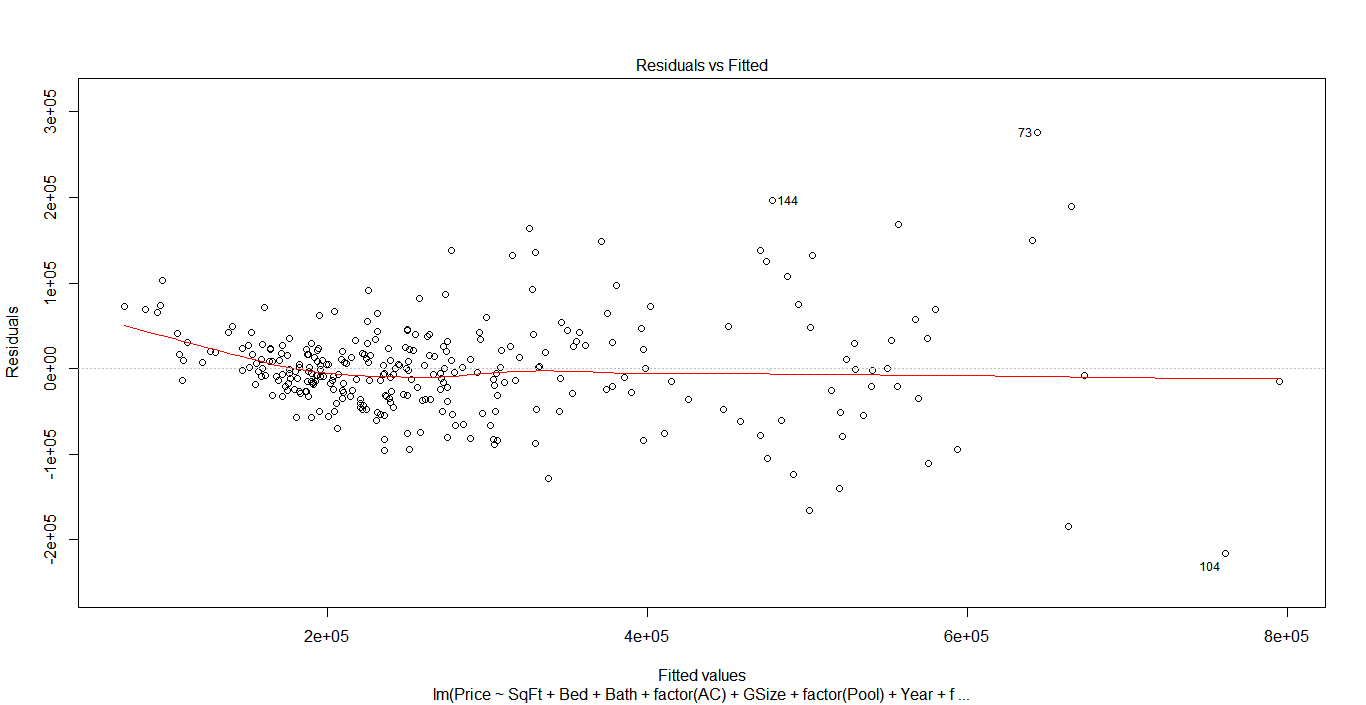
factor(Quality) 2 4.3218e+11 2.1609e+11 61.3951 < 2.2e-16 \*\*\*

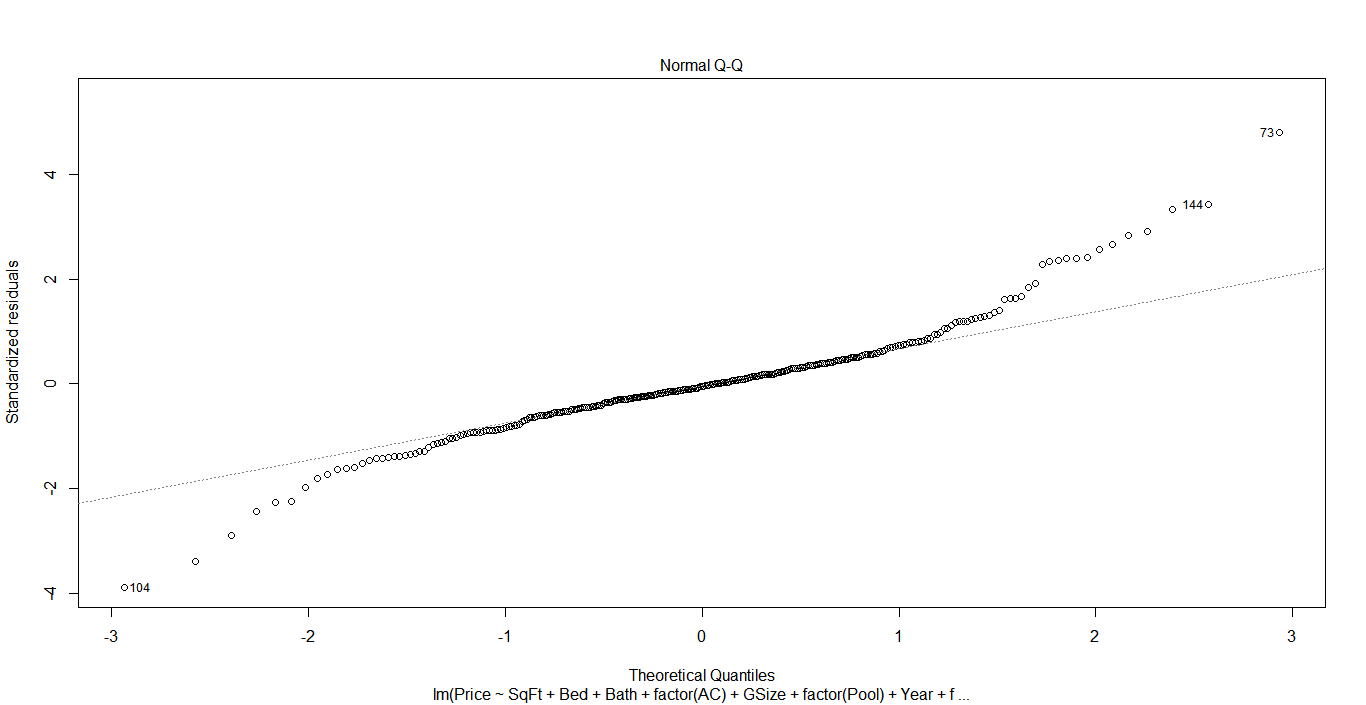
factor(Style) 9 8.5783e+10 9.5314e+09 2.7081 0.004879 \*\*

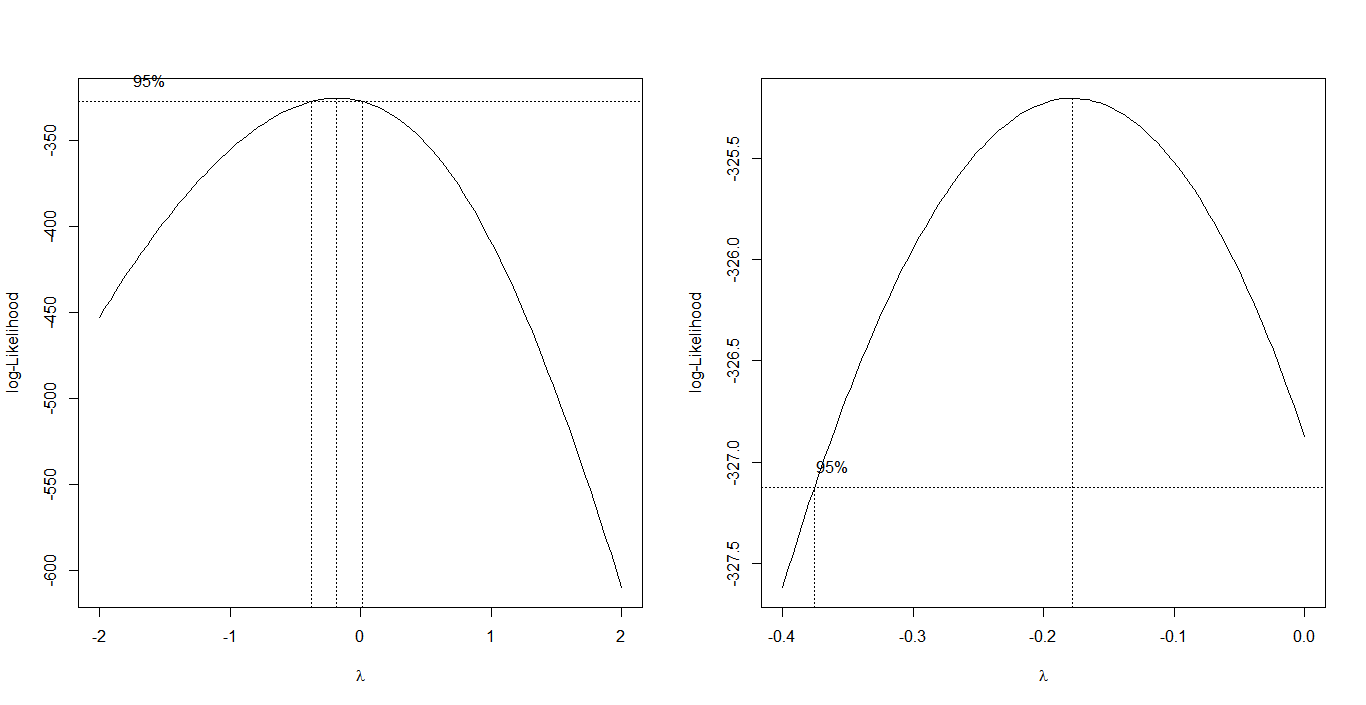
LotSize 1 1.0453e+11 1.0453e+11 29.6981 1.110e-07 \*\*\*

factor(Hwy) 1 1.3642e+10 1.3642e+10 3.8760 0.049969 \*

Residuals 279 9.8198e+11 3.5196e+09







MODEL 1:

Call:

lm(formula = sqrt(Price) ~ SqFt + Bed + Bath + factor(AC) + GSize +

factor(Pool) + Year + factor(Quality) + factor(Style) + LotSize +

factor(Hwy), data = train)

Residuals:

Min 1Q Median 3Q Max

-189.788 -27.225 -3.271 25.789 147.695

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.561e+03 4.463e+02 -3.497 0.000548 \*\*\*

SqFt 7.656e-02 8.485e-03 9.023 < 2e-16 \*\*\*

Bed -2.139e+00 3.803e+00 -0.562 0.574246

Bath 8.778e+00 4.790e+00 1.832 0.067956 .

factor(AC)1 4.965e+00 8.887e+00 0.559 0.576867

GSize 1.256e+01 5.570e+00 2.255 0.024878 \*

factor(Pool)1 1.946e+01 1.133e+01 1.719 0.086811 .

Year 9.754e-01 2.274e-01 4.289 2.47e-05 \*\*\*

factor(Quality)2 -9.851e+01 1.181e+01 -8.341 3.41e-15 \*\*\*

factor(Quality)3 -1.181e+02 1.583e+01 -7.460 1.10e-12 \*\*\*

factor(Style)2 -2.065e+01 1.016e+01 -2.031 0.043175 \*

factor(Style)3 -1.619e+01 1.020e+01 -1.587 0.113647

factor(Style)4 2.489e+01 1.939e+01 1.284 0.200318

factor(Style)5 -3.303e+01 1.883e+01 -1.754 0.080481 .

factor(Style)6 6.032e+00 1.995e+01 0.302 0.762602

factor(Style)7 -1.940e+01 9.878e+00 -1.964 0.050479 .

factor(Style)9 -3.754e+01 4.975e+01 -0.755 0.451144

factor(Style)10 -6.419e+01 5.140e+01 -1.249 0.212760

factor(Style)11 -9.632e+01 4.986e+01 -1.932 0.054415 .

LotSize 1.555e-03 2.594e-04 5.995 6.27e-09 \*\*\*

factor(Hwy)1 -4.643e+01 2.493e+01 -1.862 0.063602 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 48.66 on 279 degrees of freedom

Multiple R-squared: 0.8473, Adjusted R-squared: 0.8364

F-statistic: 77.41 on 20 and 279 DF, p-value: < 2.2e-16

> anova(lm.fit.mod1)

Analysis of Variance Table

Response: sqrt(Price)

Df Sum Sq Mean Sq F value Pr(>F)

SqFt 1 3050127 3050127 1288.0917 < 2.2e-16 \*\*\*

Bed 1 55 55 0.0232 0.87917

Bath 1 79543 79543 33.5918 1.835e-08 \*\*\*

factor(AC) 1 4884 4884 2.0627 0.15206

GSize 1 92363 92363 39.0055 1.571e-09 \*\*\*

factor(Pool) 1 213 213 0.0898 0.76459

Year 1 54210 54210 22.8933 2.779e-06 \*\*\*

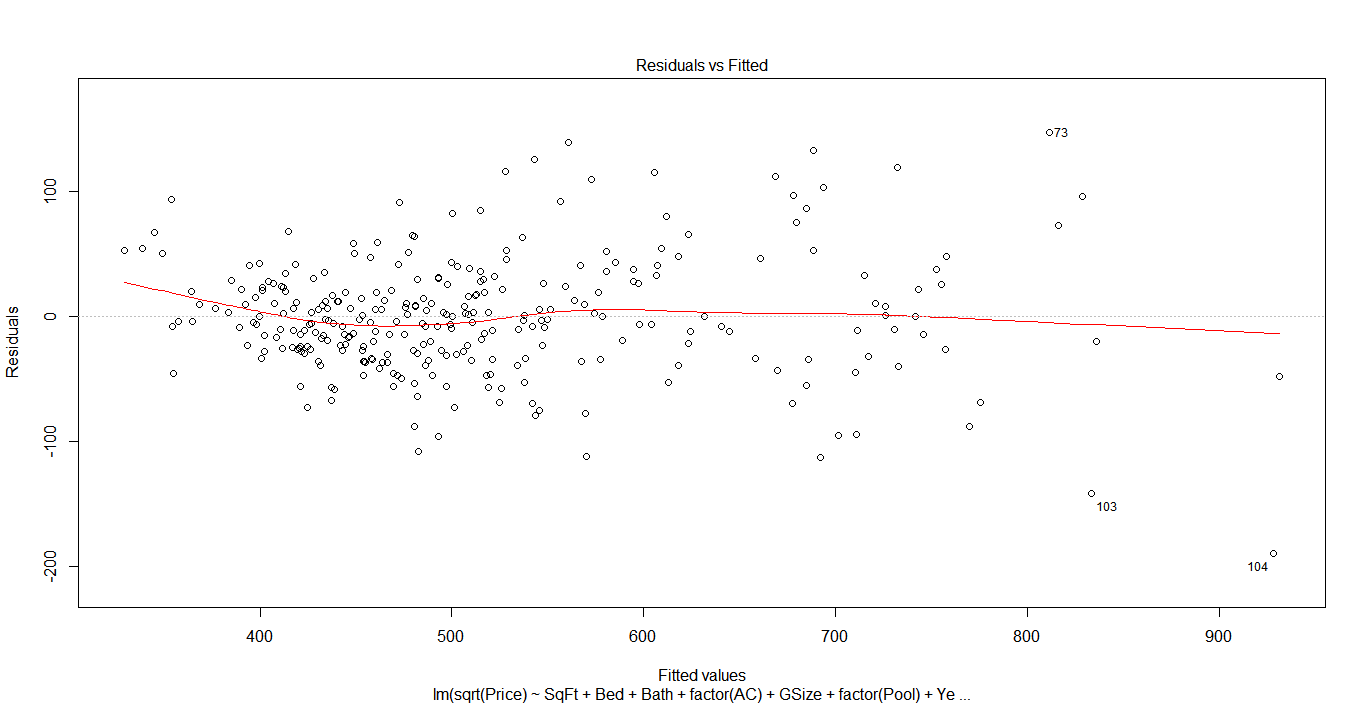
factor(Quality) 2 241684 120842 51.0325 < 2.2e-16 \*\*\*

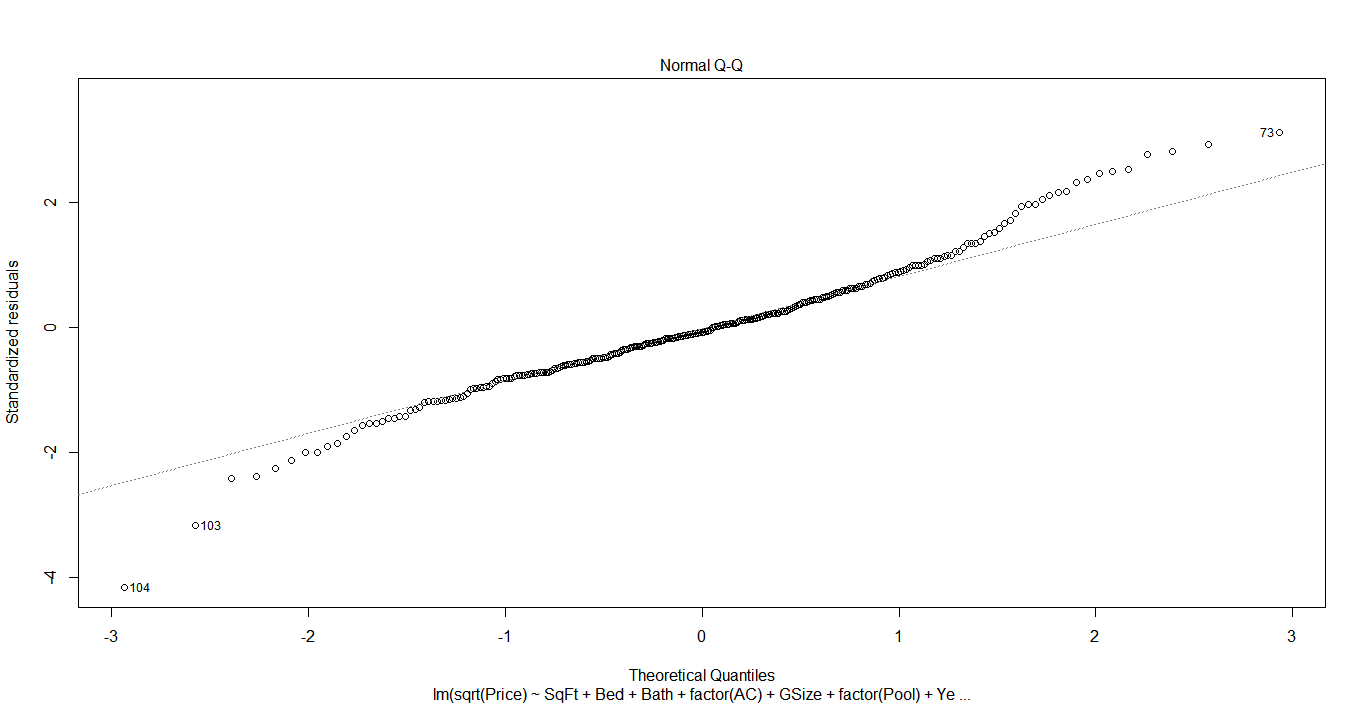
factor(Style) 9 52315 5813 2.4548 0.01051 \*

LotSize 1 82628 82628 34.8946 1.011e-08 \*\*\*

factor(Hwy) 1 8213 8213 3.4684 0.06360 .

Residuals 279 660656 2368





Call:

lm(formula = log(Price) ~ SqFt + Bed + Bath + factor(AC) + GSize +

factor(Pool) + Year + factor(Quality) + factor(Style) + LotSize +

factor(Hwy), data = train)

Residuals:

Min 1Q Median 3Q Max

-0.69904 -0.10529 -0.00473 0.10579 0.48298

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.189e+00 1.632e+00 3.179 0.00165 \*\*

SqFt 2.523e-04 3.104e-05 8.129 1.42e-14 \*\*\*

Bed 5.382e-03 1.391e-02 0.387 0.69911

Bath 4.030e-02 1.752e-02 2.300 0.02217 \*

factor(AC)1 4.215e-02 3.251e-02 1.297 0.19585

GSize 4.082e-02 2.038e-02 2.003 0.04611 \*

factor(Pool)1 8.014e-02 4.143e-02 1.935 0.05406 .

Year 3.358e-03 8.318e-04 4.037 6.99e-05 \*\*\*

factor(Quality)2 -2.946e-01 4.319e-02 -6.820 5.63e-11 \*\*\*

factor(Quality)3 -4.048e-01 5.789e-02 -6.993 1.99e-11 \*\*\*

factor(Style)2 -6.765e-02 3.718e-02 -1.820 0.06988 .

factor(Style)3 -4.457e-02 3.730e-02 -1.195 0.23319

factor(Style)4 1.004e-01 7.092e-02 1.416 0.15785

factor(Style)5 -1.030e-01 6.887e-02 -1.496 0.13580

factor(Style)6 4.007e-02 7.297e-02 0.549 0.58332

factor(Style)7 -4.753e-02 3.613e-02 -1.316 0.18940

factor(Style)9 -3.716e-02 1.820e-01 -0.204 0.83835

factor(Style)10 -2.397e-01 1.880e-01 -1.275 0.20341

factor(Style)11 -3.749e-01 1.824e-01 -2.055 0.04077 \*

LotSize 5.693e-06 9.489e-07 6.000 6.11e-09 \*\*\*

factor(Hwy)1 -1.516e-01 9.120e-02 -1.662 0.09762 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.178 on 279 degrees of freedom

Multiple R-squared: 0.842, Adjusted R-squared: 0.8307

F-statistic: 74.33 on 20 and 279 DF, p-value: < 2.2e-16

> anova(lm.fit.mod2)

Analysis of Variance Table

Response: log(Price)

Df Sum Sq Mean Sq F value Pr(>F)

SqFt 1 39.646 39.646 1251.4207 < 2.2e-16 \*\*\*

Bed 1 0.066 0.066 2.0912 0.14927

Bath 1 1.424 1.424 44.9367 1.125e-10 \*\*\*

factor(AC) 1 0.212 0.212 6.7012 0.01014 \*

GSize 1 0.946 0.946 29.8603 1.030e-07 \*\*\*

factor(Pool) 1 0.000 0.000 0.0137 0.90675

Year 1 0.677 0.677 21.3780 5.773e-06 \*\*\*

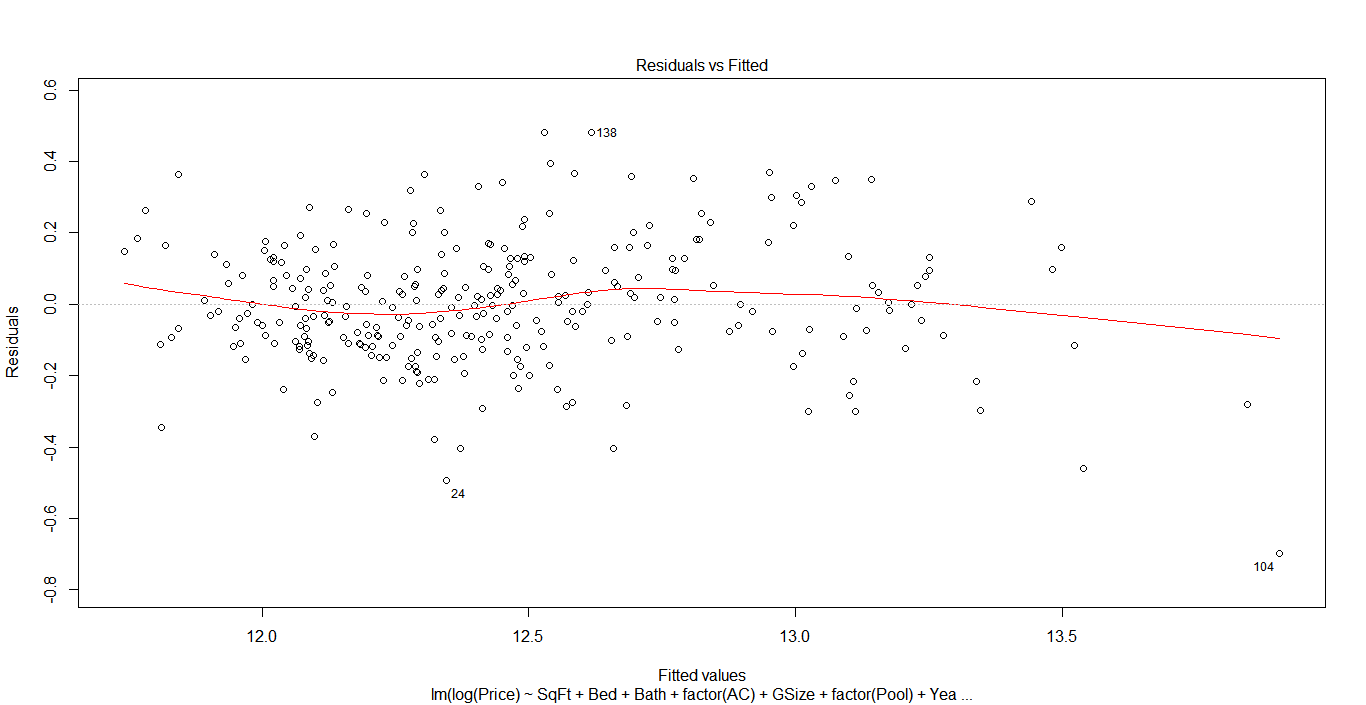
factor(Quality) 2 2.332 1.166 36.8026 6.531e-15 \*\*\*

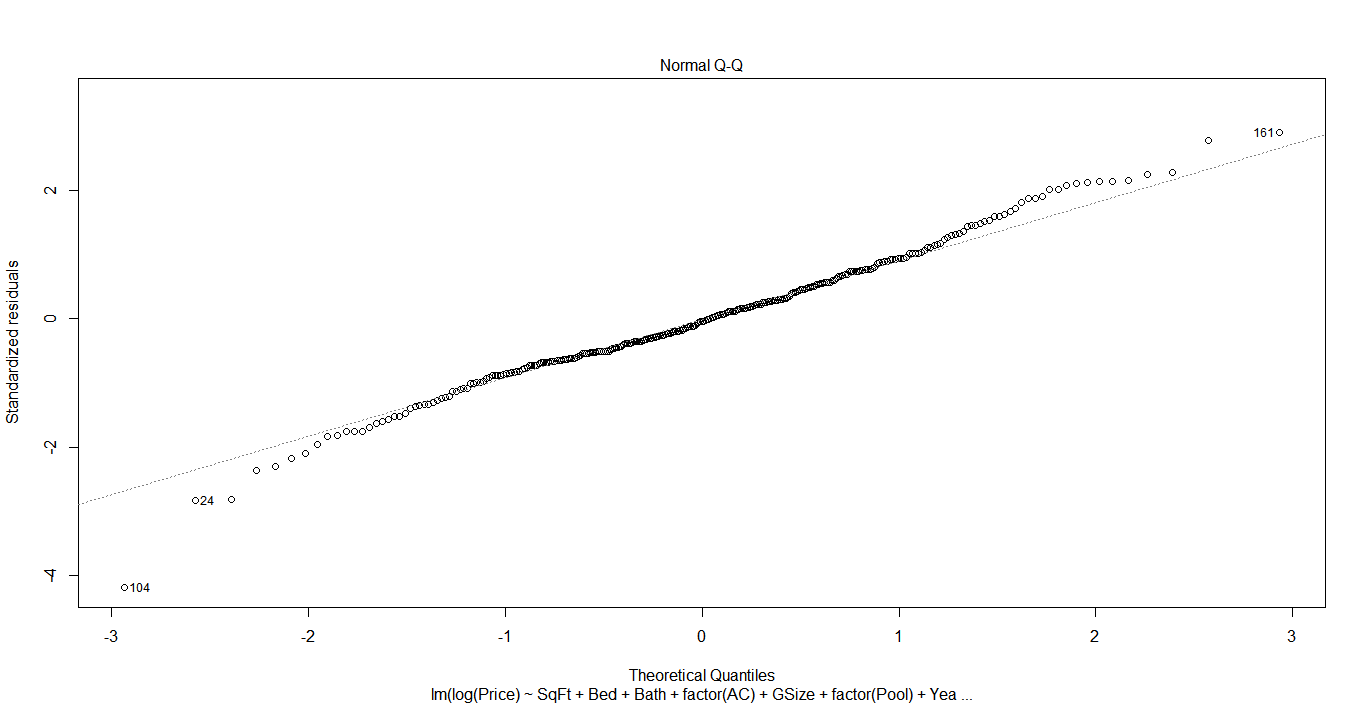
factor(Style) 9 0.594 0.066 2.0830 0.03112 \*

LotSize 1 1.111 1.111 35.0732 9.316e-09 \*\*\*

factor(Hwy) 1 0.088 0.088 2.7625 0.09762 .

Residuals 279 8.839 0.032





maxadjr(checks.adjr2, best=10)

1,3,4,5,6,7,8,9,10,11,12,13,15,17,18,19,20 1,3,4,5,6,7,8,9,10,12,13,15,17,18,19,20

0.832 0.832

1,3,4,5,6,7,8,9,10,11,12,13,15,18,19,20 1,3,4,5,6,7,8,9,10,11,12,13,14,15,17,18,19,20

0.832 0.832

1,3,5,6,7,8,9,10,11,12,13,15,17,18,19,20 1,3,4,5,6,7,8,9,10,12,13,17,18,19,20

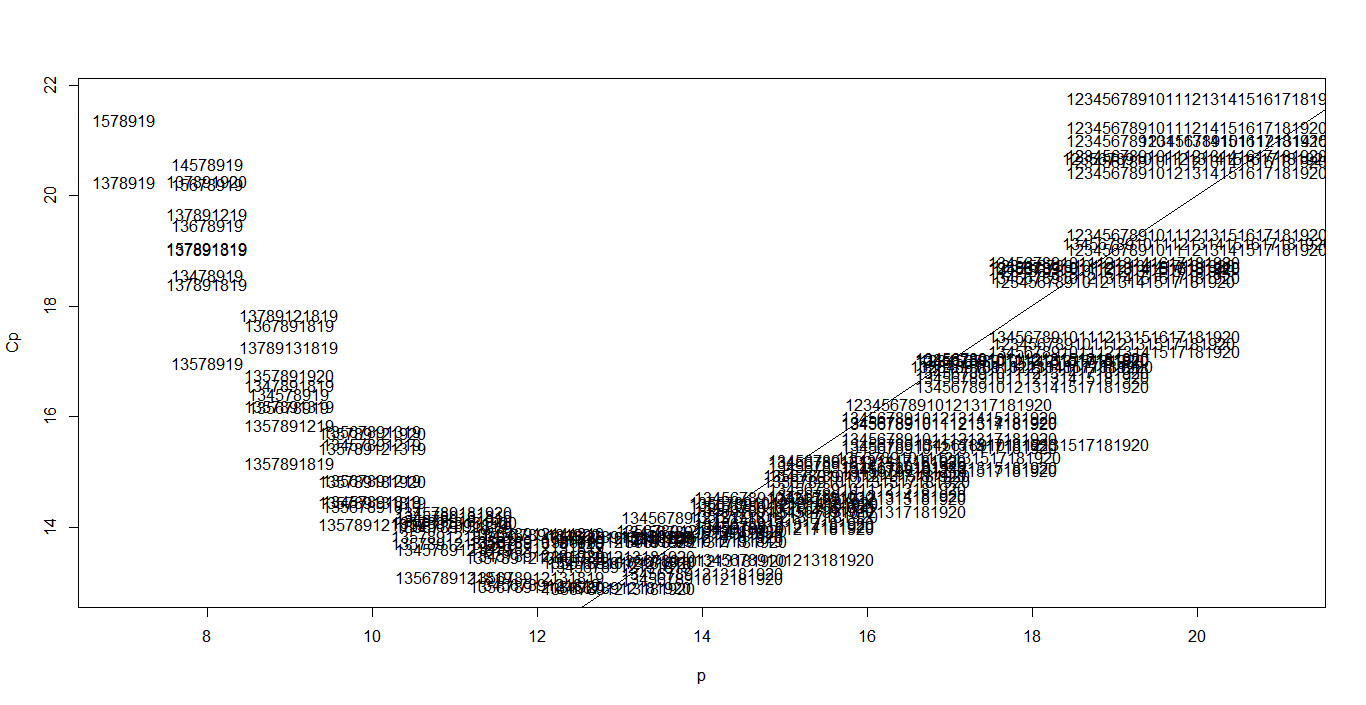
0.832 0.832

1,2,3,4,5,6,7,8,9,10,11,12,13,15,17,18,19,20 1,3,4,5,6,7,8,9,10,12,13,18,19,20

0.832 0.832

1,3,4,5,6,7,8,9,10,11,12,13,15,16,17,18,19,20 1,3,4,5,6,7,8,9,10,12,13,14,17,18,19,20

0.832 0.832



Predictors chosen: 1 ,3,5,7,8,9,10,12,13,15,17,18,19

Call:

lm(formula = log(Price) ~ SqFt + Bath + GSize + Year + factor(Style,

levels = c(2, 3, 4, 5, 7, 10, 11)) + LotSize, data = train)

Residuals:

Min 1Q Median 3Q Max

-0.53136 -0.08606 -0.00214 0.09064 0.39625

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 9.278e+00 1.912e+00 4.853 2.9e-06 \*\*\*

SqFt 3.207e-04 2.935e-05 10.929 < 2e-16 \*\*\*

Bath 6.533e-02 1.906e-02 3.428 0.000778 \*\*\*

GSize 7.984e-02 2.845e-02 2.806 0.005648 \*\*

Year 1.013e-03 9.826e-04 1.031 0.304013

factor(Style, levels = c(2, 3, 4, 5, 7, 10, 11))3 2.292e-03 3.987e-02 0.057 0.954230

factor(Style, levels = c(2, 3, 4, 5, 7, 10, 11))4 7.186e-02 7.262e-02 0.990 0.323874

factor(Style, levels = c(2, 3, 4, 5, 7, 10, 11))5 -1.640e-01 7.092e-02 -2.313 0.022024 \*

factor(Style, levels = c(2, 3, 4, 5, 7, 10, 11))7 -5.263e-03 4.127e-02 -0.128 0.898687

factor(Style, levels = c(2, 3, 4, 5, 7, 10, 11))10 -1.046e-02 1.741e-01 -0.060 0.952150

factor(Style, levels = c(2, 3, 4, 5, 7, 10, 11))11 -2.753e-01 1.674e-01 -1.645 0.102032

LotSize 4.047e-06 1.359e-06 2.978 0.003359 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1644 on 157 degrees of freedom

(131 observations deleted due to missingness)

Multiple R-squared: 0.8297, Adjusted R-squared: 0.8178

F-statistic: 69.54 on 11 and 157 DF, p-value: < 2.2e-16

> anova(lm.fit.mod3)

Analysis of Variance Table

Response: log(Price)

Df Sum Sq Mean Sq F value Pr(>F)

SqFt 1 19.1458 19.1458 708.6448 < 2.2e-16 \*\*\*

Bath 1 0.6769 0.6769 25.0546 1.481e-06 \*\*\*

GSize 1 0.2866 0.2866 10.6064 0.001381 \*\*

Year 1 0.0565 0.0565 2.0918 0.150084

factor(Style, levels = c(2, 3, 4, 5, 7, 10, 11)) 6 0.2623 0.0437 1.6178 0.145533

LotSize 1 0.2397 0.2397 8.8703 0.003359 \*\*

Residuals 157 4.2417 0.0270

Call:

lm(formula = log(Price) ~ SqFt + factor(Quality) + LotSize +

Year + Bath + GSize + factor(Pool) + factor(AC) + factor(Hwy),

data = train)

Coefficients:

(Intercept) SqFt factor(Quality)2 factor(Quality)3 LotSize Year

5.861e+00 2.503e-04 -3.081e-01 -3.984e-01 5.933e-06 3.013e-03

Bath GSize factor(Pool)1 factor(AC)1 factor(Hwy)1

3.893e-02 4.245e-02 6.233e-02 4.724e-02 -1.301e-01

> stepAIC(lm.fit.mod2, scope=list(lower=lm.fit.mod.null, upper=lm.fit.mod2), direction="backward", trace=F)

Call:

lm(formula = log(Price) ~ SqFt + Bath + factor(AC) + GSize +

factor(Pool) + Year + factor(Quality) + LotSize + factor(Hwy),

data = train)

Coefficients:

(Intercept) SqFt Bath factor(AC)1 GSize factor(Pool)1

5.861e+00 2.503e-04 3.893e-02 4.724e-02 4.245e-02 6.233e-02

Year factor(Quality)2 factor(Quality)3 LotSize factor(Hwy)1

3.013e-03 -3.081e-01 -3.984e-01 5.933e-06 -1.301e-01

> stepAIC(lm.fit.mod2, scope=list(lower=lm.fit.mod.null, upper=lm.fit.mod2), direction="both", trace=F)

Call:

lm(formula = log(Price) ~ SqFt + Bath + factor(AC) + GSize +

factor(Pool) + Year + factor(Quality) + LotSize + factor(Hwy),

data = train)

Coefficients:

(Intercept) SqFt Bath factor(AC)1 GSize factor(Pool)1

5.861e+00 2.503e-04 3.893e-02 4.724e-02 4.245e-02 6.233e-02

Year factor(Quality)2 factor(Quality)3 LotSize factor(Hwy)1

3.013e-03 -3.081e-01 -3.984e-01 5.933e-06 -1.301e-01

Subset selection object

Call: regsubsets.formula(log(Price) ~ SqFt + Bed + Bath + factor(AC) +

GSize + factor(Pool) + Year + factor(Quality) + factor(Style) +

LotSize + factor(Hwy), data = train)

20 Variables (and intercept)

Forced in Forced out

SqFt FALSE FALSE

Bed FALSE FALSE

Bath FALSE FALSE

factor(AC)1 FALSE FALSE

GSize FALSE FALSE

factor(Pool)1 FALSE FALSE

Year FALSE FALSE

factor(Quality)2 FALSE FALSE

factor(Quality)3 FALSE FALSE

factor(Style)2 FALSE FALSE

factor(Style)3 FALSE FALSE

factor(Style)4 FALSE FALSE

factor(Style)5 FALSE FALSE

factor(Style)6 FALSE FALSE

factor(Style)7 FALSE FALSE

factor(Style)9 FALSE FALSE

factor(Style)10 FALSE FALSE

factor(Style)11 FALSE FALSE

LotSize FALSE FALSE

factor(Hwy)1 FALSE FALSE

1 subsets of each size up to 8

Selection Algorithm: exhaustive

SqFt Bed Bath factor(AC)1 GSize factor(Pool)1 Year factor(Quality)2 factor(Quality)3 factor(Style)2

1 ( 1 ) "\*" " " " " " " " " " " " " " " " " " "

2 ( 1 ) "\*" " " " " " " " " " " "\*" " " " " " "

3 ( 1 ) "\*" " " " " " " " " " " " " "\*" "\*" " "

4 ( 1 ) "\*" " " " " " " " " " " " " "\*" "\*" " "

5 ( 1 ) "\*" " " " " " " " " " " "\*" "\*" "\*" " "

6 ( 1 ) "\*" " " "\*" " " " " " " "\*" "\*" "\*" " "

7 ( 1 ) "\*" " " "\*" " " "\*" " " "\*" "\*" "\*" " "

8 ( 1 ) "\*" " " "\*" " " "\*" " " "\*" "\*" "\*" " "

factor(Style)3 factor(Style)4 factor(Style)5 factor(Style)6 factor(Style)7 factor(Style)9

1 ( 1 ) " " " " " " " " " " " "

2 ( 1 ) " " " " " " " " " " " "

3 ( 1 ) " " " " " " " " " " " "

4 ( 1 ) " " " " " " " " " " " "

5 ( 1 ) " " " " " " " " " " " "

6 ( 1 ) " " " " " " " " " " " "

7 ( 1 ) " " " " " " " " " " " "

8 ( 1 ) " " " " " " " " " " " "

factor(Style)10 factor(Style)11 LotSize factor(Hwy)1

1 ( 1 ) " " " " " " " "

2 ( 1 ) " " " " " " " "

3 ( 1 ) " " " " " " " "

4 ( 1 ) " " " " "\*" " "

5 ( 1 ) " " " " "\*" " "

6 ( 1 ) " " " " "\*" " "

7 ( 1 ) " " " " "\*" " "

8 ( 1 ) " " "\*" "\*" " "

> summary(all.subset)$bic

[1] -358.6988 -391.5936 -445.8284 -466.4982 -483.9477 -485.6233 -485.1277 -483.2586

With BIC, predictors selected are, SqFt, Year, factor(Quality 2), factor(Quality 3),Style(11), LotSize.

Call:

lm(formula = log(Price) ~ SqFt + Year + I(Quality == 2) + I(Quality ==

3) + LotSize, data = train)

Residuals:

Min 1Q Median 3Q Max

-0.63558 -0.10494 0.00379 0.09929 0.43960

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.486e+00 1.543e+00 2.907 0.00393 \*\*

SqFt 3.062e-04 2.101e-05 14.578 < 2e-16 \*\*\*

Year 3.777e-03 7.777e-04 4.857 1.94e-06 \*\*\*

I(Quality == 2)TRUE -3.171e-01 3.953e-02 -8.022 2.50e-14 \*\*\*

I(Quality == 3)TRUE -4.448e-01 5.374e-02 -8.278 4.44e-15 \*\*\*

LotSize 5.871e-06 9.232e-07 6.360 7.71e-10 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1839 on 294 degrees of freedom

Multiple R-squared: 0.8222, Adjusted R-squared: 0.8192

F-statistic: 272 on 5 and 294 DF, p-value: < 2.2e-16

> anova(lm.fit.mod4)

Analysis of Variance Table

Response: log(Price)

Df Sum Sq Mean Sq F value Pr(>F)

SqFt 1 39.646 39.646 1172.167 < 2.2e-16 \*\*\*

Year 1 1.967 1.967 58.144 3.393e-13 \*\*\*

I(Quality == 2) 1 0.078 0.078 2.314 0.1293

I(Quality == 3) 1 2.932 2.932 86.700 < 2.2e-16 \*\*\*

LotSize 1 1.368 1.368 40.448 7.705e-10 \*\*\*

Residuals 294 9.944 0.034

Call:

lm(formula = log(Price) ~ SqFt + Bath + Year + I(Quality == 2) +

I(Quality == 3) + LotSize, data = train)

Residuals:

Min 1Q Median 3Q Max

-0.65873 -0.11121 -0.00037 0.09798 0.43385

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.116e+00 1.545e+00 3.312 0.00104 \*\*

SqFt 2.676e-04 2.523e-05 10.604 < 2e-16 \*\*\*

Bath 4.591e-02 1.699e-02 2.701 0.00731 \*\*

Year 3.433e-03 7.800e-04 4.402 1.51e-05 \*\*\*

I(Quality == 2)TRUE -3.094e-01 3.922e-02 -7.890 6.07e-14 \*\*\*

I(Quality == 3)TRUE -4.117e-01 5.457e-02 -7.545 5.74e-13 \*\*\*

LotSize 5.753e-06 9.145e-07 6.290 1.15e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.182 on 293 degrees of freedom

Multiple R-squared: 0.8265, Adjusted R-squared: 0.823

F-statistic: 232.7 on 6 and 293 DF, p-value: < 2.2e-16

> anova(lm.fit.mod5)

Analysis of Variance Table

Response: log(Price)

Df Sum Sq Mean Sq F value Pr(>F)

SqFt 1 39.646 39.646 1197.2705 < 2.2e-16 \*\*\*

Bath 1 1.488 1.488 44.9419 1.042e-10 \*\*\*

Year 1 1.204 1.204 36.3447 4.964e-09 \*\*\*

I(Quality == 2) 1 0.222 0.222 6.6902 0.01018 \*

I(Quality == 3) 1 2.363 2.363 71.3750 1.398e-15 \*\*\*

LotSize 1 1.310 1.310 39.5688 1.150e-09 \*\*\*

Residuals 293 9.702 0.033

Call:

lm(formula = log(Price) ~ SqFt + Bath + GSize + Year + I(Quality ==

2) + I(Quality == 3) + LotSize, data = train)

Residuals:

Min 1Q Median 3Q Max

-0.63404 -0.11454 -0.00291 0.10281 0.52844

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.876e+00 1.570e+00 3.742 0.000220 \*\*\*

SqFt 2.580e-04 2.541e-05 10.150 < 2e-16 \*\*\*

Bath 4.221e-02 1.696e-02 2.489 0.013362 \*

GSize 4.567e-02 2.020e-02 2.261 0.024475 \*

Year 3.012e-03 7.967e-04 3.781 0.000189 \*\*\*

I(Quality == 2)TRUE -2.966e-01 3.936e-02 -7.537 6.09e-13 \*\*\*

I(Quality == 3)TRUE -3.981e-01 5.452e-02 -7.302 2.71e-12 \*\*\*

LotSize 5.486e-06 9.158e-07 5.990 6.17e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1807 on 292 degrees of freedom

Multiple R-squared: 0.8295, Adjusted R-squared: 0.8254

F-statistic: 203 on 7 and 292 DF, p-value: < 2.2e-16

> anova(lm.fit.mod6)

Analysis of Variance Table

Response: log(Price)

Df Sum Sq Mean Sq F value Pr(>F)

SqFt 1 39.646 39.646 1214.0797 < 2.2e-16 \*\*\*

Bath 1 1.488 1.488 45.5729 7.924e-11 \*\*\*

GSize 1 1.072 1.072 32.8246 2.509e-08 \*\*\*

Year 1 0.743 0.743 22.7515 2.915e-06 \*\*\*

I(Quality == 2) 1 0.164 0.164 5.0078 0.02599 \*

I(Quality == 3) 1 2.116 2.116 64.7945 2.117e-14 \*\*\*

LotSize 1 1.172 1.172 35.8757 6.171e-09 \*\*\*

Residuals 292 9.535 0.033

> press.stats

[1] 10.55659 10.43428 10.37855

> mod.ss

[1] 9.943862 9.702250 9.535266