Project_PSTAT_126_Inwoong_Bae

Inwoong Bae

06/07/2019

part 1

```
MyData <- read.table("RealEstateValuation.txt", header = TRUE)
MyData
```

```
##        TDate  Age      Metro Stores Latitude Longitude Price
## 1   2012.917 32.0   84.87882     10 24.98298  121.5402  37.9
## 2   2012.917 19.5  306.59470      9 24.98034  121.5395  42.2
## 3   2013.583 13.3  561.98450      5 24.98746  121.5439  47.3
## 4   2013.500 13.3  561.98450      5 24.98746  121.5439  54.8
## 5   2012.833  5.0  390.56840      5 24.97937  121.5425  43.1
## 6   2012.667  7.1 2175.03000      3 24.96305  121.5125  32.1
## 7   2012.667 34.5  623.47310      7 24.97933  121.5364  40.3
## 8   2013.417 20.3  287.60250      6 24.98042  121.5423  46.7
## 9   2013.500 31.7 5512.03800      1 24.95095  121.4846  18.8
## 10  2013.417 17.9 1783.18000      3 24.96731  121.5149  22.1
## 11  2013.083 34.8  405.21340      1 24.97349  121.5337  41.4
## 12  2013.333  6.3   90.45606      9 24.97433  121.5431  58.1
## 13  2012.917 13.0  492.23130      5 24.96515  121.5374  39.3
## 14  2012.667 20.4 2469.64500      4 24.96108  121.5105  23.8
## 15  2013.500 13.2 1164.83800      4 24.99156  121.5341  34.3
## 16  2013.583 35.7  579.20830      2 24.98240  121.5462  50.5
## 17  2013.250  0.0  292.99780      6 24.97744  121.5446  70.1
## 18  2012.750 17.7  350.85150      1 24.97544  121.5312  37.4
## 19  2013.417 16.9  368.13630      8 24.96750  121.5445  42.3
## 20  2012.667  1.5   23.38284      7 24.96772  121.5410  47.7
## 21  2013.417  4.5 2275.87700      3 24.96314  121.5115  29.3
## 22  2013.417 10.5  279.17260      7 24.97528  121.5454  51.6
## 23  2012.917 14.7 1360.13900      1 24.95204  121.5484  24.6
## 24  2013.083 10.1  279.17260      7 24.97528  121.5454  47.9
## 25  2013.000 39.6  480.69770      4 24.97353  121.5388  38.8
## 26  2013.083 29.3 1487.86800      2 24.97542  121.5173  27.0
## 27  2012.667  3.1  383.86240      5 24.98085  121.5439  56.2
## 28  2013.250 10.4  276.44900      5 24.95593  121.5391  33.6
## 29  2013.500 19.2  557.47800      4 24.97419  121.5380  47.0
## 30  2013.083  7.1  451.24380      5 24.97563  121.5469  57.1
## 31  2013.500 25.9 4519.69000      0 24.94826  121.4959  22.1
## 32  2012.750 29.6  769.40340      7 24.98281  121.5341  25.0
## 33  2012.750 37.9  488.57270      1 24.97349  121.5345  34.2
## 34  2013.250 16.5  323.65500      6 24.97841  121.5428  49.3
## 35  2012.750 15.4  205.36700      7 24.98419  121.5424  55.1
```

```
## 36  2013.500 13.9 4079.41800      0 25.01459  121.5182  27.3
## 37  2012.917 14.7 1935.00900      2 24.96386  121.5146  22.9
## 38  2013.167 12.0 1360.13900      1 24.95204  121.5484  25.3
## 39  2012.667  3.1  577.96150      6 24.97201  121.5472  47.7
## 40  2013.167 16.2  289.32480      5 24.98203  121.5435  46.2
## 41  2013.000 13.6 4082.01500      0 24.94155  121.5038  15.9
## 42  2013.500 16.8 4066.58700      0 24.94297  121.5034  18.2
## 43  2013.417 36.1  519.46170      5 24.96305  121.5376  34.7
## 44  2012.750 34.4  512.78710      6 24.98748  121.5430  34.1
## 45  2013.583  2.7  533.47620      4 24.97445  121.5477  53.9
## 46  2013.083 36.6  488.81930      8 24.97015  121.5449  38.3
## 47  2013.417 21.7  463.96230      9 24.97030  121.5446  42.0
## 48  2013.583 35.9  640.73910      3 24.97563  121.5371  61.5
## 49  2013.417 24.2 4605.74900      0 24.94684  121.4958  13.4
## 50  2012.667 29.4 4510.35900      1 24.94925  121.4954  13.2
## 51  2013.417 21.7  512.54870      4 24.97400  121.5384  44.2
## 52  2013.083 31.3 1758.40600      1 24.95402  121.5528  20.7
## 53  2013.583 32.1 1438.57900      3 24.97419  121.5175  27.0
## 54  2013.083 13.3  492.23130      5 24.96515  121.5374  38.9
## 55  2013.083 16.1  289.32480      5 24.98203  121.5435  51.7
## 56  2012.833 31.7 1160.63200      0 24.94968  121.5301  13.7
## 57  2013.417 33.6  371.24950      8 24.97254  121.5406  41.9
## 58  2012.917  3.5   56.47425      7 24.95744  121.5371  53.5
## 59  2013.500 30.3 4510.35900      1 24.94925  121.4954  22.6
## 60  2013.083 13.3  336.05320      5 24.95776  121.5344  42.4
## 61  2013.417 11.0 1931.20700      2 24.96365  121.5147  21.3
## 62  2013.500  5.3  259.66070      6 24.97585  121.5452  63.2
## 63  2012.917 17.2 2175.87700      3 24.96303  121.5125  27.7
## 64  2013.583  2.6  533.47620      4 24.97445  121.5477  55.0
## 65  2013.333 17.5  995.75540      0 24.96305  121.5491  25.3
## 66  2013.417 40.1  123.74290      8 24.97635  121.5433  44.3
## 67  2013.000  1.0  193.58450      6 24.96571  121.5409  50.7
## 68  2013.500  8.5  104.81010      5 24.96674  121.5407  56.8
## 69  2013.417 30.4  464.22300      6 24.97964  121.5380  36.2
## 70  2012.833 12.5  561.98450      5 24.98746  121.5439  42.0
## 71  2013.583  6.6   90.45606      9 24.97433  121.5431  59.0
## 72  2013.083 35.5  640.73910      3 24.97563  121.5371  40.8
## 73  2013.583 32.5  424.54420      8 24.97587  121.5391  36.3
## 74  2013.167 13.8 4082.01500      0 24.94155  121.5038  20.0
## 75  2012.917  6.8  379.55750     10 24.98343  121.5376  54.4
## 76  2013.500 12.3 1360.13900      1 24.95204  121.5484  29.5
## 77  2013.583 35.9  616.40040      3 24.97723  121.5377  36.8
## 78  2012.833 20.5 2185.12800      3 24.96322  121.5124  25.6
## 79  2012.917 38.2  552.43710      2 24.97598  121.5338  29.8
## 80  2013.000 18.0 1414.83700      1 24.95182  121.5489  26.5
## 81  2013.500 11.8  533.47620      4 24.97445  121.5477  40.3
## 82  2013.000 30.8  377.79560      6 24.96427  121.5396  36.8
## 83  2013.083 13.2  150.93470      7 24.96725  121.5425  48.1
## 84  2012.917 25.3 2707.39200      3 24.96056  121.5083  17.7
```

```
## 85  2013.083 15.1  383.28050      7 24.96735  121.5446  43.7
## 86  2012.750  0.0  338.96790      9 24.96853  121.5441  50.8
## 87  2012.833  1.8 1455.79800      1 24.95120  121.5490  27.0
## 88  2013.583 16.9 4066.58700      0 24.94297  121.5034  18.3
## 89  2012.917  8.9 1406.43000      0 24.98573  121.5276  48.0
## 90  2013.500 23.0 3947.94500      0 24.94783  121.5024  25.3
## 91  2012.833  0.0  274.01440      1 24.97480  121.5306  45.4
## 92  2013.250  9.1 1402.01600      0 24.98569  121.5276  43.2
## 93  2012.917 20.6 2469.64500      4 24.96108  121.5105  21.8
## 94  2012.917 31.9 1146.32900      0 24.94920  121.5308  16.1
## 95  2012.917 40.9  167.59890      5 24.96630  121.5403  41.0
## 96  2012.917  8.0  104.81010      5 24.96674  121.5407  51.8
## 97  2013.417  6.4   90.45606      9 24.97433  121.5431  59.5
## 98  2013.083 28.4  617.44240      3 24.97746  121.5330  34.6
## 99  2013.417 16.4  289.32480      5 24.98203  121.5435  51.0
## 100 2013.417  6.4   90.45606      9 24.97433  121.5431  62.2
## 101 2013.500 17.5  964.74960      4 24.98872  121.5341  38.2
## 102 2012.833 12.7  170.12890      1 24.97371  121.5298  32.9
## 103 2013.083  1.1  193.58450      6 24.96571  121.5409  54.4
## 104 2012.750  0.0  208.39050      6 24.95618  121.5384  45.7
## 105 2012.667 32.7  392.44590      6 24.96398  121.5425  30.5
## 106 2012.833  0.0  292.99780      6 24.97744  121.5446  71.0
## 107 2013.083 17.2  189.51810      8 24.97707  121.5431  47.1
## 108 2013.333 12.2 1360.13900      1 24.95204  121.5484  26.6
## 109 2013.417 31.4  592.50060      2 24.97260  121.5356  34.1
## 110 2013.583  4.0 2147.37600      3 24.96299  121.5128  28.4
## 111 2013.083  8.1  104.81010      5 24.96674  121.5407  51.6
## 112 2013.583 33.3  196.61720      7 24.97701  121.5422  39.4
## 113 2013.417  9.9 2102.42700      3 24.96044  121.5146  23.1
## 114 2013.333 14.8  393.26060      6 24.96172  121.5381   7.6
## 115 2012.667 30.6  143.83830      8 24.98155  121.5414  53.3
## 116 2013.083 20.6  737.91610      2 24.98092  121.5474  46.4
## 117 2013.000 30.9 6396.28300      1 24.94375  121.4788  12.2
## 118 2013.000 13.6 4197.34900      0 24.93885  121.5038  13.0
## 119 2013.500 25.3 1583.72200      3 24.96622  121.5171  30.6
## 120 2013.500 16.6  289.32480      5 24.98203  121.5435  59.6
## 121 2013.167 13.3  492.23130      5 24.96515  121.5374  31.3
## 122 2013.500 13.6  492.23130      5 24.96515  121.5374  48.0
## 123 2013.250 31.5  414.94760      4 24.98199  121.5446  32.5
## 124 2013.417  0.0  185.42960      0 24.97110  121.5317  45.5
## 125 2012.917  9.9  279.17260      7 24.97528  121.5454  57.4
## 126 2013.167  1.1  193.58450      6 24.96571  121.5409  48.6
## 127 2013.083 38.6  804.68970      4 24.97838  121.5348  62.9
## 128 2013.250  3.8  383.86240      5 24.98085  121.5439  55.0
## 129 2013.083 41.3  124.99120      6 24.96674  121.5404  60.7
## 130 2013.417 38.5  216.83290      7 24.98086  121.5416  41.0
## 131 2013.250 29.6  535.52700      8 24.98092  121.5365  37.5
## 132 2013.500  4.0 2147.37600      3 24.96299  121.5128  30.7
## 133 2013.167 26.6  482.75810      5 24.97433  121.5386  37.5
```

```
## 134 2012.833 18.0  373.39370      8 24.98660  121.5408  39.5
## 135 2012.667 33.4  186.96860      6 24.96604  121.5421  42.2
## 136 2012.917 18.9 1009.23500      0 24.96357  121.5495  20.8
## 137 2012.750 11.4  390.56840      5 24.97937  121.5425  46.8
## 138 2013.500 13.6  319.07080      6 24.96495  121.5428  47.4
## 139 2013.167 10.0  942.46640      0 24.97843  121.5241  43.5
## 140 2012.667 12.9  492.23130      5 24.96515  121.5374  42.5
## 141 2013.250 16.2  289.32480      5 24.98203  121.5435  51.4
## 142 2013.333  5.1 1559.82700      3 24.97213  121.5163  28.9
## 143 2013.417 19.8  640.60710      5 24.97017  121.5465  37.5
## 144 2013.500 13.6  492.23130      5 24.96515  121.5374  40.1
## 145 2013.083 11.9 1360.13900      1 24.95204  121.5484  28.4
## 146 2012.917  2.1  451.24380      5 24.97563  121.5469  45.5
## 147 2012.750  0.0  185.42960      0 24.97110  121.5317  52.2
## 148 2012.750  3.2  489.88210      8 24.97017  121.5449  43.2
## 149 2013.500 16.4 3780.59000      0 24.93293  121.5120  45.1
## 150 2012.667 34.9  179.45380      8 24.97349  121.5425  39.7
## 151 2013.250 35.8  170.73110      7 24.96719  121.5427  48.5
## 152 2013.500  4.9  387.77210      9 24.98118  121.5379  44.7
## 153 2013.333 12.0 1360.13900      1 24.95204  121.5484  28.9
## 154 2013.250  6.5  376.17090      6 24.95418  121.5371  40.9
## 155 2013.500 16.9 4066.58700      0 24.94297  121.5034  20.7
## 156 2013.167 13.8 4082.01500      0 24.94155  121.5038  15.6
## 157 2013.583 30.7 1264.73000      0 24.94883  121.5295  18.3
## 158 2013.250 16.1  815.93140      4 24.97886  121.5346  35.6
## 159 2013.000 11.6  390.56840      5 24.97937  121.5425  39.4
## 160 2012.667 15.5  815.93140      4 24.97886  121.5346  37.4
## 161 2012.917  3.5   49.66105      8 24.95836  121.5376  57.8
## 162 2013.417 19.2  616.40040      3 24.97723  121.5377  39.6
## 163 2012.750 16.0 4066.58700      0 24.94297  121.5034  11.6
## 164 2013.500  8.5  104.81010      5 24.96674  121.5407  55.5
## 165 2012.833  0.0  185.42960      0 24.97110  121.5317  55.2
## 166 2012.917 13.7 1236.56400      1 24.97694  121.5539  30.6
## 167 2013.417  0.0  292.99780      6 24.97744  121.5446  73.6
## 168 2013.417 28.2  330.08540      8 24.97408  121.5401  43.4
## 169 2013.083 27.6  515.11220      5 24.96299  121.5432  37.4
## 170 2013.417  8.4 1962.62800      1 24.95468  121.5548  23.5
## 171 2013.333 24.0 4527.68700      0 24.94741  121.4963  14.4
## 172 2013.083  3.6  383.86240      5 24.98085  121.5439  58.8
## 173 2013.583  6.6   90.45606      9 24.97433  121.5431  58.1
## 174 2013.083 41.3  401.88070      4 24.98326  121.5446  35.1
## 175 2013.417  4.3  432.03850      7 24.98050  121.5378  45.2
## 176 2013.083 30.2  472.17450      3 24.97005  121.5376  36.5
## 177 2012.833 13.9 4573.77900      0 24.94867  121.4951  19.2
## 178 2013.083 33.0  181.07660      9 24.97697  121.5426  42.0
## 179 2013.500 13.1 1144.43600      4 24.99176  121.5346  36.7
## 180 2013.083 14.0  438.85130      1 24.97493  121.5273  42.6
## 181 2012.667 26.9 4449.27000      0 24.94898  121.4962  15.5
## 182 2013.167 11.6  201.89390      8 24.98489  121.5412  55.9
```

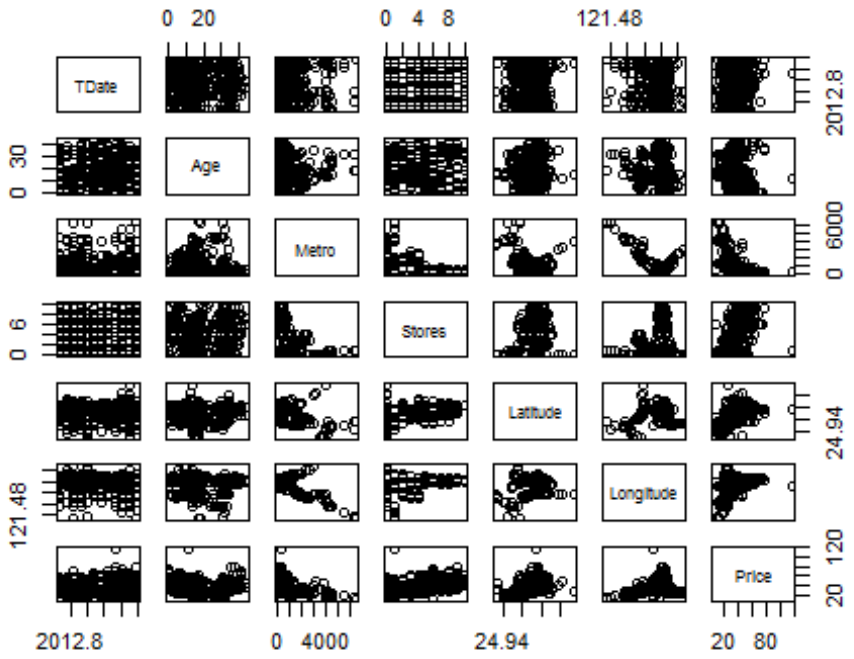
```
## 183 2013.500 13.5 2147.37600      3 24.96299  121.5128  23.6
## 184 2013.500 17.0 4082.01500      0 24.94155  121.5038  18.8
## 185 2012.750 14.1 2615.46500      0 24.95495  121.5617  21.8
## 186 2012.750 31.4 1447.28600      3 24.97285  121.5173  21.5
## 187 2013.167 20.9 2185.12800      3 24.96322  121.5124  25.7
## 188 2013.000  8.9 3078.17600      0 24.95464  121.5663  22.0
## 189 2012.917 34.8  190.03920      8 24.97707  121.5431  44.3
## 190 2012.917 16.3 4066.58700      0 24.94297  121.5034  20.5
## 191 2013.500 35.3  616.57350      8 24.97945  121.5364  42.3
## 192 2013.167 13.2  750.07040      2 24.97371  121.5495  37.8
## 193 2013.167 43.8   57.58945      7 24.96750  121.5407  42.7
## 194 2013.417  9.7  421.47900      5 24.98246  121.5448  49.3
## 195 2013.500 15.2 3771.89500      0 24.93363  121.5116  29.3
## 196 2013.333 15.2  461.10160      5 24.95425  121.5399  34.6
## 197 2013.000 22.8  707.90670      2 24.98100  121.5471  36.6
## 198 2013.250 34.4  126.72860      8 24.96881  121.5409  48.2
## 199 2013.083 34.0  157.60520      7 24.96628  121.5420  39.1
## 200 2013.417 18.2  451.64190      8 24.96945  121.5449  31.6
## 201 2013.417 17.4  995.75540      0 24.96305  121.5491  25.5
## 202 2013.417 13.1  561.98450      5 24.98746  121.5439  45.9
## 203 2012.917 38.3  642.69850      3 24.97559  121.5371  31.5
## 204 2012.667 15.6  289.32480      5 24.98203  121.5435  46.1
## 205 2013.000 18.0 1414.83700      1 24.95182  121.5489  26.6
## 206 2013.083 12.8 1449.72200      3 24.97289  121.5173  21.4
## 207 2013.250 22.2  379.55750     10 24.98343  121.5376  44.0
## 208 2013.083 38.5  665.06360      3 24.97503  121.5369  34.2
## 209 2012.750 11.5 1360.13900      1 24.95204  121.5484  26.2
## 210 2012.833 34.8  175.62940      8 24.97347  121.5427  40.9
## 211 2013.500  5.2  390.56840      5 24.97937  121.5425  52.2
## 212 2013.083  0.0  274.01440      1 24.97480  121.5306  43.5
## 213 2013.333 17.6 1805.66500      2 24.98672  121.5209  31.1
## 214 2013.083  6.2   90.45606      9 24.97433  121.5431  58.0
## 215 2013.583 18.1 1783.18000      3 24.96731  121.5149  20.9
## 216 2013.333 19.2  383.71290      8 24.97200  121.5448  48.1
## 217 2013.250 37.8  590.92920      1 24.97153  121.5356  39.7
## 218 2012.917 28.0  372.62420      6 24.97838  121.5412  40.8
## 219 2013.417 13.6  492.23130      5 24.96515  121.5374  43.8
## 220 2012.750 29.3  529.77710      8 24.98102  121.5366  40.2
## 221 2013.333 37.2  186.51010      9 24.97703  121.5426  78.3
## 222 2013.333  9.0 1402.01600      0 24.98569  121.5276  38.5
## 223 2013.583 30.6  431.11140     10 24.98123  121.5374  48.5
## 224 2013.250  9.1 1402.01600      0 24.98569  121.5276  42.3
## 225 2013.333 34.5  324.94190      6 24.97814  121.5417  46.0
## 226 2013.250  1.1  193.58450      6 24.96571  121.5409  49.0
## 227 2013.000 16.5 4082.01500      0 24.94155  121.5038  12.8
## 228 2012.917 32.4  265.06090      8 24.98059  121.5399  40.2
## 229 2013.417 11.9 3171.32900      0 25.00115  121.5178  46.6
## 230 2013.583 31.0 1156.41200      0 24.94890  121.5310  19.0
## 231 2013.500  4.0 2147.37600      3 24.96299  121.5128  33.4
```

```
## 232 2012.833 16.2 4074.73600      0 24.94235  121.5036  14.7
## 233 2012.917 27.1 4412.76500      1 24.95032  121.4959  17.4
## 234 2013.333 39.7  333.36790      9 24.98016  121.5393  32.4
## 235 2013.250  8.0 2216.61200      4 24.96007  121.5136  23.9
## 236 2012.750 12.9  250.63100      7 24.96606  121.5430  39.3
## 237 2013.167  3.6  373.83890     10 24.98322  121.5376  61.9
## 238 2013.167 13.0  732.85280      0 24.97668  121.5252  39.0
## 239 2013.083 12.8  732.85280      0 24.97668  121.5252  40.6
## 240 2013.500 18.1  837.72330      0 24.96334  121.5477  29.7
## 241 2013.083 11.0 1712.63200      2 24.96412  121.5167  28.8
## 242 2013.500 13.7  250.63100      7 24.96606  121.5430  41.4
## 243 2012.833  2.0 2077.39000      3 24.96357  121.5133  33.4
## 244 2013.417 32.8  204.17050      8 24.98236  121.5392  48.2
## 245 2013.083  4.8 1559.82700      3 24.97213  121.5163  21.7
## 246 2013.417  7.5  639.61980      5 24.97258  121.5481  40.8
## 247 2013.417 16.4  389.82190      6 24.96412  121.5427  40.6
## 248 2013.333 21.7 1055.06700      0 24.96211  121.5493  23.1
## 249 2013.000 19.0 1009.23500      0 24.96357  121.5495  22.3
## 250 2012.833 18.0 6306.15300      1 24.95743  121.4752  15.0
## 251 2013.167 39.2  424.71320      7 24.97429  121.5392  30.0
## 252 2012.917 31.7 1159.45400      0 24.94960  121.5302  13.8
## 253 2012.833  5.9   90.45606      9 24.97433  121.5431  52.7
## 254 2012.667 30.4 1735.59500      2 24.96464  121.5162  25.9
## 255 2012.667  1.1  329.97470      5 24.98254  121.5439  51.8
## 256 2013.417 31.5 5512.03800      1 24.95095  121.4846  17.4
## 257 2012.667 14.6  339.22890      1 24.97519  121.5315  26.5
## 258 2013.250 17.3  444.13340      1 24.97501  121.5273  43.9
## 259 2013.417  0.0  292.99780      6 24.97744  121.5446  63.3
## 260 2013.083 17.7  837.72330      0 24.96334  121.5477  28.8
## 261 2013.250 17.0 1485.09700      4 24.97073  121.5170  30.7
## 262 2013.167 16.2 2288.01100      3 24.95885  121.5136  24.4
## 263 2012.917 15.9  289.32480      5 24.98203  121.5435  53.0
## 264 2013.417  3.9 2147.37600      3 24.96299  121.5128  31.7
## 265 2013.167 32.6  493.65700      7 24.96968  121.5452  40.6
## 266 2012.833 15.7  815.93140      4 24.97886  121.5346  38.1
## 267 2013.250 17.8 1783.18000      3 24.96731  121.5149  23.7
## 268 2012.833 34.7  482.75810      5 24.97433  121.5386  41.1
## 269 2013.417 17.2  390.56840      5 24.97937  121.5425  40.1
## 270 2013.000 17.6  837.72330      0 24.96334  121.5477  23.0
## 271 2013.333 10.8  252.58220      1 24.97460  121.5305 117.5
## 272 2012.917 17.7  451.64190      8 24.96945  121.5449  26.5
## 273 2012.750 13.0  492.23130      5 24.96515  121.5374  40.5
## 274 2013.417 13.2  170.12890      1 24.97371  121.5298  29.3
## 275 2013.167 27.5  394.01730      7 24.97305  121.5399  41.0
## 276 2012.667  1.5   23.38284      7 24.96772  121.5410  49.7
## 277 2013.000 19.1  461.10160      5 24.95425  121.5399  34.0
## 278 2013.417 21.2 2185.12800      3 24.96322  121.5124  27.7
## 279 2012.750  0.0  208.39050      6 24.95618  121.5384  44.0
## 280 2013.417  2.6 1554.25000      3 24.97026  121.5164  31.1
```

```
## 281 2013.250  2.3  184.33020      6 24.96581  121.5409  45.4
## 282 2013.333  4.7  387.77210      9 24.98118  121.5379  44.8
## 283 2012.917  2.0 1455.79800      1 24.95120  121.5490  25.6
## 284 2013.417 33.5 1978.67100      2 24.98674  121.5184  23.5
## 285 2012.917 15.0  383.28050      7 24.96735  121.5446  34.4
## 286 2013.167 30.1  718.29370      3 24.97509  121.5364  55.3
## 287 2012.917  5.9   90.45606      9 24.97433  121.5431  56.3
## 288 2013.000 19.2  461.10160      5 24.95425  121.5399  32.9
## 289 2013.583 16.6  323.69120      6 24.97841  121.5428  51.0
## 290 2013.333 13.9  289.32480      5 24.98203  121.5435  44.5
## 291 2013.083 37.7  490.34460      0 24.97217  121.5347  37.0
## 292 2012.833  3.4   56.47425      7 24.95744  121.5371  54.4
## 293 2013.083 17.5  395.67470      5 24.95674  121.5340  24.5
## 294 2012.667 12.6  383.28050      7 24.96735  121.5446  42.5
## 295 2013.500 26.4  335.52730      6 24.97960  121.5414  38.1
## 296 2013.167 18.2 2179.59000      3 24.96299  121.5125  21.8
## 297 2012.750 12.5 1144.43600      4 24.99176  121.5346  34.1
## 298 2012.833 34.9  567.03490      4 24.97003  121.5458  28.5
## 299 2013.333 16.7 4082.01500      0 24.94155  121.5038  16.7
## 300 2013.167 33.2  121.72620     10 24.98178  121.5406  46.1
## 301 2013.083  2.5  156.24420      4 24.96696  121.5399  36.9
## 302 2012.750 38.0  461.78480      0 24.97229  121.5345  35.7
## 303 2013.500 16.5 2288.01100      3 24.95885  121.5136  23.2
## 304 2013.500 38.3  439.71050      0 24.97161  121.5342  38.4
## 305 2013.417 20.0 1626.08300      3 24.96622  121.5167  29.4
## 306 2013.083 16.2  289.32480      5 24.98203  121.5435  55.0
## 307 2013.500 14.4  169.98030      1 24.97369  121.5298  50.2
## 308 2012.833 10.3 3079.89000      0 24.95460  121.5663  24.7
## 309 2013.417 16.4  289.32480      5 24.98203  121.5435  53.0
## 310 2013.250 30.3 1264.73000      0 24.94883  121.5295  19.1
## 311 2013.583 16.4 1643.49900      2 24.95394  121.5517  24.7
## 312 2013.167 21.3  537.79710      4 24.97425  121.5381  42.2
## 313 2013.583 35.4  318.52920      9 24.97071  121.5407  78.0
## 314 2013.333  8.3  104.81010      5 24.96674  121.5407  42.8
## 315 2013.250  3.7  577.96150      6 24.97201  121.5472  41.6
## 316 2013.083 15.6 1756.41100      2 24.98320  121.5181  27.3
## 317 2013.250 13.3  250.63100      7 24.96606  121.5430  42.0
## 318 2012.750 15.6  752.76690      2 24.97795  121.5345  37.5
## 319 2013.333  7.1  379.55750     10 24.98343  121.5376  49.8
## 320 2013.250 34.6  272.67830      5 24.95562  121.5387  26.9
## 321 2012.750 13.5 4197.34900      0 24.93885  121.5038  18.6
## 322 2012.917 16.9  964.74960      4 24.98872  121.5341  37.7
## 323 2013.000 12.9  187.48230      1 24.97388  121.5298  33.1
## 324 2013.417 28.6  197.13380      6 24.97631  121.5444  42.5
## 325 2012.667 12.4 1712.63200      2 24.96412  121.5167  31.3
## 326 2013.083 36.6  488.81930      8 24.97015  121.5449  38.1
## 327 2013.500  4.1   56.47425      7 24.95744  121.5371  62.1
## 328 2013.417  3.5  757.33770      3 24.97538  121.5497  36.7
## 329 2012.833 15.9 1497.71300      3 24.97003  121.5170  23.6
```

```
## 330 2013.000 13.6 4197.34900      0 24.93885  121.5038  19.2
## 331 2013.083 32.0 1156.77700      0 24.94935  121.5305  12.8
## 332 2013.333 25.6 4519.69000      0 24.94826  121.4959  15.6
## 333 2013.167 39.8  617.71340      2 24.97577  121.5348  39.6
## 334 2012.750  7.8  104.81010      5 24.96674  121.5407  38.4
## 335 2012.917 30.0 1013.34100      5 24.99006  121.5346  22.8
## 336 2013.583 27.3  337.60160      6 24.96431  121.5406  36.5
## 337 2012.833  5.1 1867.23300      2 24.98407  121.5175  35.6
## 338 2012.833 31.3  600.86040      5 24.96871  121.5465  30.9
## 339 2012.917 31.5  258.18600      9 24.96867  121.5433  36.3
## 340 2013.333  1.7  329.97470      5 24.98254  121.5439  50.4
## 341 2013.333 33.6  270.88950      0 24.97281  121.5327  42.9
## 342 2013.000 13.0  750.07040      2 24.97371  121.5495  37.0
## 343 2012.667  5.7   90.45606      9 24.97433  121.5431  53.5
## 344 2013.000 33.5  563.28540      8 24.98223  121.5360  46.6
## 345 2013.500 34.6 3085.17000      0 24.99800  121.5155  41.2
## 346 2012.667  0.0  185.42960      0 24.97110  121.5317  37.9
## 347 2013.417 13.2 1712.63200      2 24.96412  121.5167  30.8
## 348 2013.583 17.4 6488.02100      1 24.95719  121.4735  11.2
## 349 2012.833  4.6  259.66070      6 24.97585  121.5452  53.7
## 350 2012.750  7.8  104.81010      5 24.96674  121.5407  47.0
## 351 2013.000 13.2  492.23130      5 24.96515  121.5374  42.3
## 352 2012.833  4.0 2180.24500      3 24.96324  121.5124  28.6
## 353 2012.833 18.4 2674.96100      3 24.96143  121.5083  25.7
## 354 2013.500  4.1 2147.37600      3 24.96299  121.5128  31.3
## 355 2013.417 12.2 1360.13900      1 24.95204  121.5484  30.1
## 356 2013.250  3.8  383.86240      5 24.98085  121.5439  60.7
## 357 2012.833 10.3  211.44730      1 24.97417  121.5300  45.3
## 358 2013.417  0.0  338.96790      9 24.96853  121.5441  44.9
## 359 2013.167  1.1  193.58450      6 24.96571  121.5409  45.1
## 360 2013.500  5.6 2408.99300      0 24.95505  121.5596  24.7
## 361 2012.667 32.9   87.30222     10 24.98300  121.5402  47.1
## 362 2013.083 41.4  281.20500      8 24.97345  121.5409  63.3
## 363 2013.417 17.1  967.40000      4 24.98872  121.5341  40.0
## 364 2013.500 32.3  109.94550     10 24.98182  121.5409  48.0
## 365 2013.417 35.3  614.13940      7 24.97913  121.5367  33.1
## 366 2012.917 17.3 2261.43200      4 24.96182  121.5122  29.5
## 367 2012.750 14.2 1801.54400      1 24.95153  121.5525  24.8
## 368 2012.833 15.0 1828.31900      2 24.96464  121.5153  20.9
## 369 2013.417 18.2  350.85150      1 24.97544  121.5312  43.1
## 370 2012.667 20.2 2185.12800      3 24.96322  121.5124  22.8
## 371 2012.750 15.9  289.32480      5 24.98203  121.5435  42.1
## 372 2013.500  4.1  312.89630      5 24.95591  121.5396  51.7
## 373 2013.000 33.9  157.60520      7 24.96628  121.5420  41.5
## 374 2013.083  0.0  274.01440      1 24.97480  121.5306  52.2
## 375 2013.250  5.4  390.56840      5 24.97937  121.5425  49.5
## 376 2013.250 21.7 1157.98800      0 24.96165  121.5501  23.8
## 377 2013.417 14.7 1717.19300      2 24.96447  121.5165  30.5
## 378 2013.333  3.9   49.66105      8 24.95836  121.5376  56.8
```

```
## 379 2013.333 37.3  587.88770      8 24.97077  121.5463  37.4
## 380 2013.333  0.0  292.99780      6 24.97744  121.5446  69.7
## 381 2013.333 14.1  289.32480      5 24.98203  121.5435  53.3
## 382 2013.417  8.0  132.54690      9 24.98298  121.5398  47.3
## 383 2013.000 16.3 3529.56400      0 24.93207  121.5160  29.3
## 384 2012.667 29.1  506.11440      4 24.97845  121.5389  40.3
## 385 2012.750 16.1 4066.58700      0 24.94297  121.5034  12.9
## 386 2013.000 18.3   82.88643     10 24.98300  121.5403  46.6
## 387 2012.833  0.0  185.42960      0 24.97110  121.5317  55.3
## 388 2013.250 16.2 2103.55500      3 24.96042  121.5146  25.6
## 389 2013.500 10.4 2251.93800      4 24.95957  121.5135  27.3
## 390 2013.250 40.9  122.36190      8 24.96756  121.5423  67.7
## 391 2013.500 32.8  377.83020      9 24.97151  121.5435  38.6
## 392 2013.583  6.2 1939.74900      1 24.95155  121.5539  31.3
## 393 2013.083 42.7  443.80200      6 24.97927  121.5387  35.3
## 394 2013.000 16.9  967.40000      4 24.98872  121.5341  40.3
## 395 2013.500 32.6 4136.27100      1 24.95544  121.4963  24.7
## 396 2012.917 21.2  512.54870      4 24.97400  121.5384  42.5
## 397 2012.667 37.1  918.63570      1 24.97198  121.5506  31.9
## 398 2013.417 13.1 1164.83800      4 24.99156  121.5341  32.2
## 399 2013.417 14.7 1717.19300      2 24.96447  121.5165  23.0
## 400 2012.917 12.7  170.12890      1 24.97371  121.5298  37.3
## 401 2013.250 26.8  482.75810      5 24.97433  121.5386  35.5
## 402 2013.083  7.6 2175.03000      3 24.96305  121.5125  27.7
## 403 2012.833 12.7  187.48230      1 24.97388  121.5298  28.5
## 404 2012.667 30.9  161.94200      9 24.98353  121.5397  39.7
## 405 2013.333 16.4  289.32480      5 24.98203  121.5435  41.2
## 406 2012.667 23.0  130.99450      6 24.95663  121.5376  37.2
## 407 2013.167  1.9  372.13860      7 24.97293  121.5403  40.5
## 408 2013.000  5.2 2408.99300      0 24.95505  121.5596  22.3
## 409 2013.417 18.5 2175.74400      3 24.96330  121.5124  28.1
## 410 2013.000 13.7 4082.01500      0 24.94155  121.5038  15.4
## 411 2012.667  5.6   90.45606      9 24.97433  121.5431  50.0
## 412 2013.250 18.8  390.96960      7 24.97923  121.5399  40.6
## 413 2013.000  8.1  104.81010      5 24.96674  121.5407  52.5
## 414 2013.500  6.5   90.45606      9 24.97433  121.5431  63.9

plot(MyData)
```




```
According to the plot among variables in the dataset from Real Estate Valuati
on, there seems no significant relationship between price and each factor.
mod <- lm(Price ~ TDate + Age + Stores + Latitude, data = MyData)
summary(mod)

##
## Call:
## lm(formula = Price ~ TDate + Age + Stores + Latitude, data = MyData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.620  -5.601  -0.714   4.207  80.465
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.742e+04  3.524e+03  -4.944 1.12e-06 ***
## TDate        3.613e+00  1.686e+00   2.143   0.0327 *
## Age         -3.020e-01  4.178e-02  -7.227 2.44e-12 ***
## Stores       1.929e+00  1.801e-01  10.712  < 2e-16 ***
## Latitude     4.078e+02  4.278e+01   9.534  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.654 on 409 degrees of freedom
## Multiple R-squared:  0.5015, Adjusted R-squared:  0.4966
## F-statistic: 102.8 on 4 and 409 DF,  p-value: < 2.2e-16
```

let Y be price, x1 be TDate, x2 be Age, x3 be Stores, x4 be Latitude. then the equation for the fitted regression line is Y = -1.742e+04 + 3.613e+00*x1* - *3.020e-01*x2 + 1.929e+00*x3 + 4.078e+02 + error. By summary, the r-value of each variable except TDate is lower than 0.01. This means that the all variables except TDate are significant for this model. TDate is not significant for this model, so it can be removed to make the model become better model.

```
#Suppose we add Metro or longitude on the previous model.
mod2 = lm(Price ~ TDate + Age + Stores + Latitude + Metro, data = MyData )
mod3 = lm(Price ~ TDate + Age + Stores + Latitude + Longitude, data = MyData)

#The null hypothesis for the original model and the the model that adds Metro
 is betha of Metro equals zero and the alternative hypothesis is the betha of
 Metro is nonzero.
anova(mod,mod2) #anova table for the original model and the the model with Me
tro
```

```
## Analysis of Variance Table
##
## Model 1: Price ~ TDate + Age + Stores + Latitude
## Model 2: Price ~ TDate + Age + Stores + Latitude + Metro
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1    409 38119
## 2    408 31938  1    6181.8 78.972 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#The null hypothesis for the original model and the the model that adds Longi
tude is betha of Longitude equals zero and the alternative hypothesis is the
betha of Longitude is nonzero.
anova(mod,mod3) #anova table for the original model and the the model with Lo
ngitude
```

```
## Analysis of Variance Table
##
## Model 1: Price ~ TDate + Age + Stores + Latitude
## Model 2: Price ~ TDate + Age + Stores + Latitude + Longitude
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1    409 38119
## 2    408 34997  1    3122.5 36.402 3.605e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#by the anova tables, p-values of Metro and Longitude are much lower than alp
ha = 0.05. Therefore, both are available to be added on the original model.

#Suppose we have another possible model
modSec <- lm(Price ~ TDate + Age + Metro + Latitude, data = MyData)
summary(modSec)
```

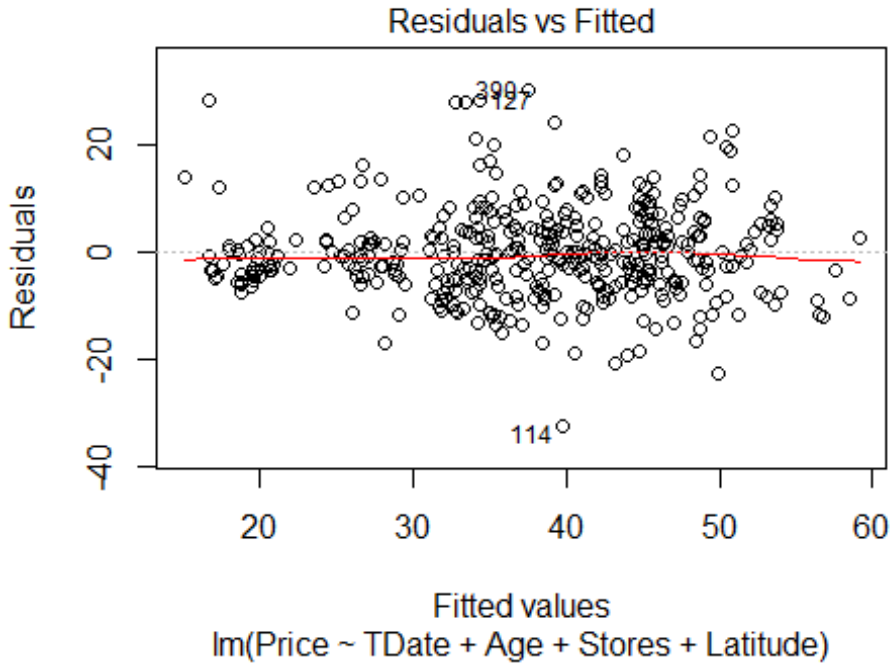
```
##
## Call:
## lm(formula = Price ~ TDate + Age + Metro + Latitude, data = MyData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.218  -5.269  -0.700   4.433  70.502
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.767e+04  3.359e+03  -5.262 2.30e-07 ***
## TDate        5.570e+00  1.619e+00   3.440 0.000642 ***
## Age         -2.530e-01  4.001e-02  -6.323 6.71e-10 ***
## Metro       -5.764e-03  4.493e-04 -12.829  < 2e-16 ***
## Latitude     2.607e+02  4.569e+01   5.705 2.23e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.225 on 409 degrees of freedom
## Multiple R-squared:  0.5448, Adjusted R-squared:  0.5403
## F-statistic: 122.4 on 4 and 409 DF,  p-value: < 2.2e-16
```

```
#The equation for the regression line is Price = -1.767e+4 + 5.570e+00 * TDat
e - 2.530e-01 * Age - 5.764e-03 * Metro + 2.607e+02 * Latitude.
summary(mod)
```

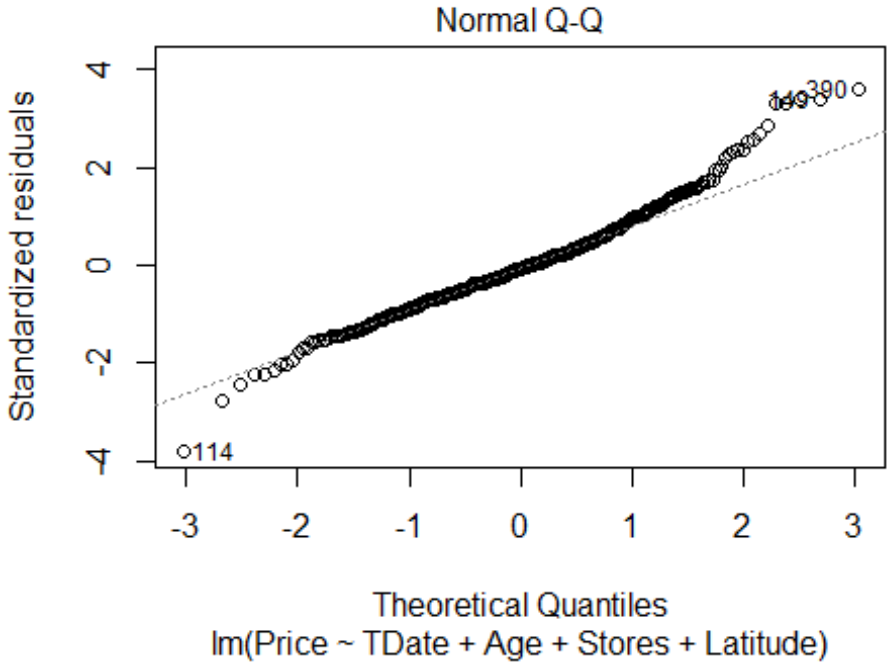
```
##
## Call:
## lm(formula = Price ~ TDate + Age + Stores + Latitude, data = MyData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.620  -5.601  -0.714   4.207  80.465
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.742e+04  3.524e+03  -4.944 1.12e-06 ***
## TDate        3.613e+00  1.686e+00   2.143   0.0327 *
## Age         -3.020e-01  4.178e-02  -7.227 2.44e-12 ***
## Stores       1.929e+00  1.801e-01  10.712  < 2e-16 ***
## Latitude     4.078e+02  4.278e+01   9.534  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.654 on 409 degrees of freedom
## Multiple R-squared:  0.5015, Adjusted R-squared:  0.4966
## F-statistic: 102.8 on 4 and 409 DF,  p-value: < 2.2e-16
```

In both model, global p-values are same, but p-values for each individual variable are different. In the first model, p-value for TDate is relatively high and it results in being insignificant for the model by some significance levels. However, in the second model, all p-values are low enough to be significant for all significance levels. Therefore, we prefer the second model.

```
library(car)

## Warning: package 'car' was built under R version 3.5.3

## Loading required package: carData

## Warning: package 'carData' was built under R version 3.5.2

attach(MyData)
outlierTest(modSec) #find outliers on original model

##     rstudent unadjusted p-value Bonferonni p
## 271 8.270876         1.8999e-15   7.8657e-13
## 313 4.147533         4.0910e-05   1.6937e-02
## 221 4.108887         4.8069e-05   1.9900e-02

newData <- MyData[c(1:220,222:270,272:312, 314:414),] #delete outliers
newmodSec <- lm(Price ~ TDate + Age + Stores + Latitude, data = newData) #new
 dataset without outliers
plot(newmodSec, which = 1)
```



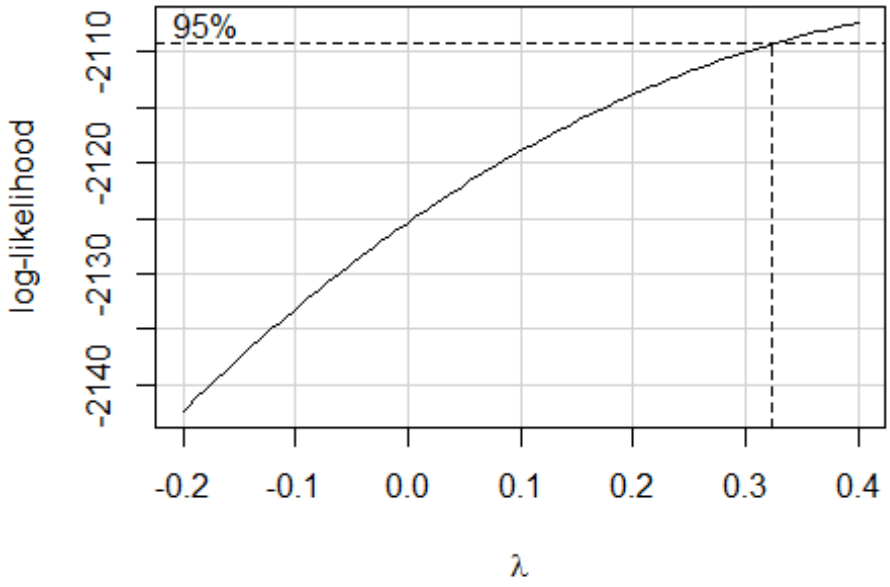
```
plot(newmodSec, which = 2)
```




```
#When we see the Residuals vs Fitted, the line is not parallel at 0 and all p
oints in Normal Q-Q plot are not in the line
```

Using the Box-Cox method, we see that λ = 0 is both in the interval, and extremely close to the maximum, which suggests a transformation of the form log(Price)

```
boxCox(newmodSec, lambda = seq(-0.2, 0.4, by = 0.05), plotit = TRUE)
```




```
modSec_cox <- lm(log(Price) ~ TDate + Age + Stores + Latitude, data = newData)
summary(modSec_cox)
```

```
##
## Call:
## lm(formula = log(Price) ~ TDate + Age + Stores + Latitude, data = newData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.57120 -0.13585  0.00979  0.14493  0.94143
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.770e+02  9.104e+01  -5.240 2.59e-07 ***
## TDate        6.565e-02  4.359e-02   1.506    0.133
## Age         -8.331e-03  1.083e-03  -7.693 1.10e-13 ***
## Stores       5.465e-02  4.679e-03  11.678  < 2e-16 ***
## Latitude     1.395e+01  1.104e+00  12.638  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2486 on 406 degrees of freedom
## Multiple R-squared:  0.588,  Adjusted R-squared:  0.5839
## F-statistic: 144.8 on 4 and 406 DF,  p-value: < 2.2e-16
```

```
#After modifying the model by Box-Cox method, we conclude that the model is n
ot needed to be changed by Box-Cox method because the modified model has insi
gnificant variable, TDate.
```

We now apply the powertransform method to the model.

```
summary(Price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7.60   27.70   38.45   37.98   46.60  117.50
```

```
summary(TDate)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2013    2013    2013    2013    2013    2014
```

```
summary(Age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   9.025  16.100  17.713  28.150  43.800
```

```
summary(Metro)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   23.38  289.32  492.23 1083.89 1454.28 6488.02
```

```
summary(Latitude)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   24.93   24.96   24.97   24.97   24.98   25.01
```

```
#Since Age has 0 value and it cannot be transformed by powerTransform. We the
refore need to add a small constant to transform
newData$Age1 <- with(newData, (Age*TDate + 1)/TDate)
pt <- powerTransform(Price ~ cbind(TDate, Age1, Stores, Latitude), newData)
summary(pt)
```

```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1    0.5854         0.5       0.3941       0.7768
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                             LRT df       pval
## LR test, lambda = (0) 39.79588  1 2.8194e-10
##
## Likelihood ratio test that no transformation is needed
##                             LRT df       pval
## LR test, lambda = (1) 16.71345  1 4.3472e-05
```

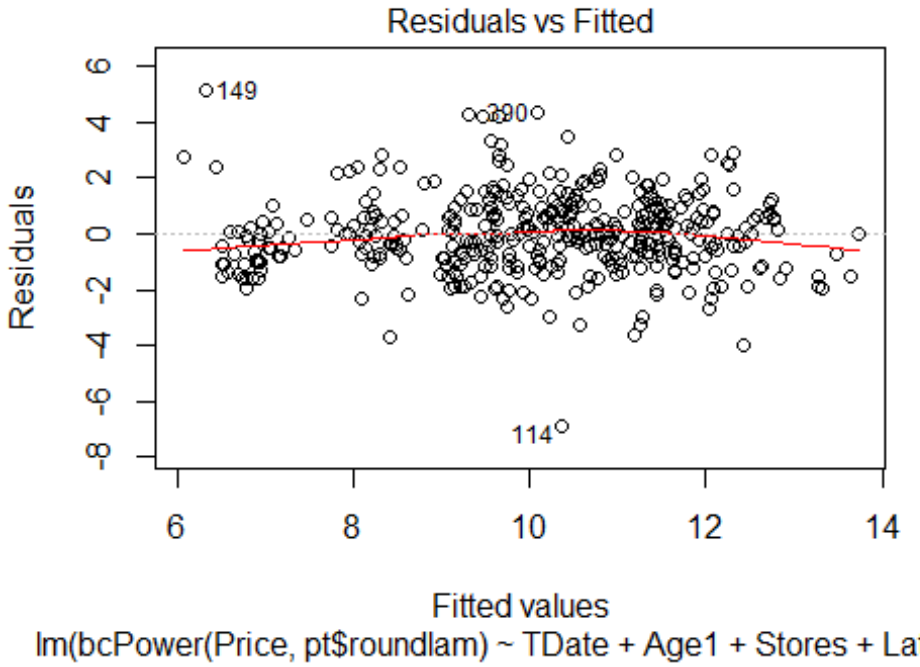
```
pt$roundlam
```

```
##  Y1
## 0.5
```

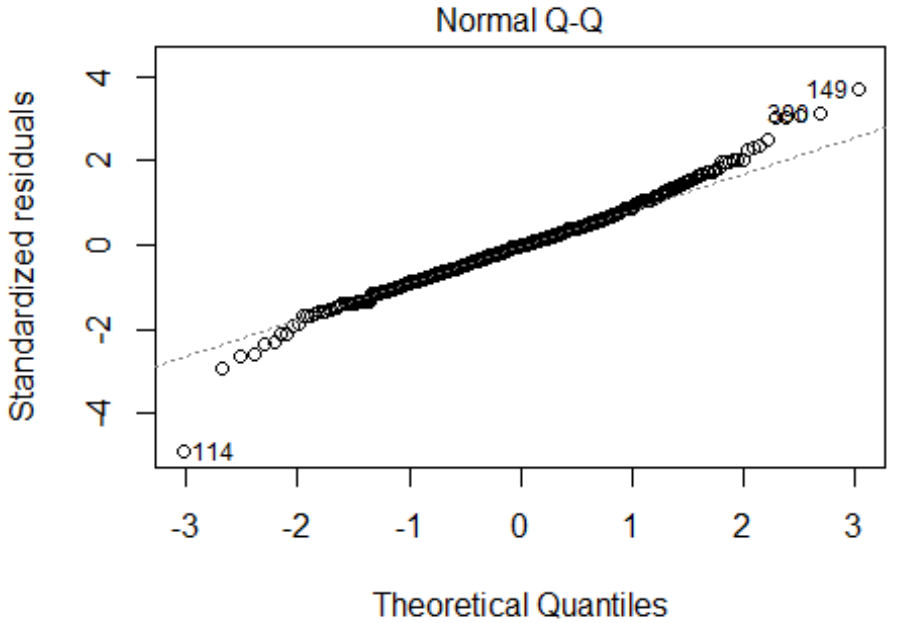
```
summary(newMod2 <- lm(bcPower(Price, pt$roundlam) ~ TDate + Age1 + Stores + L
atitude, data = newData))
```

```
##
## Call:
## lm(formula = bcPower(Price, pt$roundlam) ~ TDate + Age1 + Stores +
##     Latitude, data = newData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8564 -0.8779 -0.0252  0.7632  5.1160
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.626e+03  5.148e+02  -5.100 5.21e-07 ***
## TDate        3.983e-01  2.465e-01   1.616    0.107
## Age1        -5.040e-02  6.124e-03  -8.230 2.57e-15 ***
## Stores       3.211e-01  2.646e-02  12.135  < 2e-16 ***
## Latitude     7.343e+01  6.242e+00  11.763  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.406 on 406 degrees of freedom
## Multiple R-squared:  0.5836, Adjusted R-squared:  0.5795
## F-statistic: 142.2 on 4 and 406 DF,  p-value: < 2.2e-16
```

```
plot(newMod2, which = 1)
```




```
plot(newMod2, which = 2)
```



We apply polynomial fits

```
quadratic.lm1 <- lm(Price ~ TDate  + Age + I(Age^2) + Stores + Latitude, data
 = newData)
summary(quadratic.lm1)
```

```
##
## Call:
## lm(formula = Price ~ TDate + Age + I(Age^2) + Stores + Latitude,
##     data = newData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.1291  -4.6118  -0.5942   4.7345  29.6598
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.629e+04  2.874e+03  -5.666 2.78e-08 ***
## TDate        3.425e+00  1.378e+00   2.485   0.0133 *
## Age         -1.367e+00  1.284e-01 -10.646  < 2e-16 ***
## I(Age^2)     2.663e-02  3.130e-03   8.506 3.51e-16 ***
## Stores       1.688e+00  1.502e-01  11.234  < 2e-16 ***
## Latitude     3.778e+02  3.488e+01  10.832  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.838 on 405 degrees of freedom
## Multiple R-squared:  0.6278, Adjusted R-squared:  0.6232
## F-statistic: 136.6 on 5 and 405 DF,  p-value: < 2.2e-16
```

```
quadratic.lm2 <- lm(Price ~ TDate  + Age + I(Age^2) +I(Age^3)+ Stores + Latit
ude, data = newData)
summary(quadratic.lm2)
```

```
##
## Call:
## lm(formula = Price ~ TDate + Age + I(Age^2) + I(Age^3) + Stores +
##     Latitude, data = newData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.3748  -4.5603  -0.4862   4.5653  29.5234
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.607e+04  2.880e+03  -5.581 4.39e-08 ***
## TDate        3.345e+00  1.380e+00   2.424   0.0158 *
## Age         -1.099e+00  2.754e-01  -3.990 7.86e-05 ***
## I(Age^2)     8.544e-03  1.672e-02   0.511   0.6096
## I(Age^3)     3.117e-04  2.831e-04   1.101   0.2715
## Stores       1.695e+00  1.503e-01  11.275  < 2e-16 ***
```

```
## Latitude     3.758e+02  3.492e+01  10.763  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.836 on 404 degrees of freedom
## Multiple R-squared:  0.6289, Adjusted R-squared:  0.6234
## F-statistic: 114.1 on 6 and 404 DF,  p-value: < 2.2e-16
```

```
quadratic.lm3 <- lm(Price ~ TDate  + Age  + Stores + Latitude + I(Latitude^2),
 data = newData)
summary(quadratic.lm3)
```

```
##
## Call:
## lm(formula = Price ~ TDate + Age + Stores + Latitude + I(Latitude^2),
##     data = newData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.557  -5.488  -0.608   4.170  32.646
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.878e+06  1.237e+06  -2.327   0.0205 *
## TDate          2.824e+00  1.487e+00   1.898   0.0584 .
## Age           -3.120e-01  3.684e-02  -8.469 4.63e-16 ***
## Stores         1.813e+00  1.668e-01  10.870  < 2e-16 ***
## Latitude       2.297e+05  9.907e+04   2.319   0.0209 *
## I(Latitude^2) -4.592e+03  1.984e+03  -2.315   0.0211 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.454 on 405 degrees of freedom
## Multiple R-squared:  0.567,  Adjusted R-squared:  0.5617
## F-statistic: 106.1 on 5 and 405 DF,  p-value: < 2.2e-16
```

```
#All polynomial fits does not get better than original one.
```

After modifying the model by transform methods, we cannot see any significantly positive changes and improvements.There are some influencial points such as outliers in this models and the influencial points distract us when we find the appropriate model. We conclude that transformation is unnecessary to apply because none of method affects to make it clear.

Part 2

```
concrete<-read.table('Concrete.txt')
summary(concrete)
```

```
##        X1              X2              X3              X4
##  Min.   :102.0   Min.   :  0.0   Min.   :  0.00   Min.   :121.8
##  1st Qu.:192.4   1st Qu.:  0.0   1st Qu.:  0.00   1st Qu.:164.9
##  Median :272.9   Median : 22.0   Median :  0.00   Median :185.0
##  Mean   :281.2   Mean   : 73.9   Mean   : 54.19   Mean   :181.6
##  3rd Qu.:350.0   3rd Qu.:142.9   3rd Qu.:118.27   3rd Qu.:192.0
##  Max.   :540.0   Max.   :359.4   Max.   :200.10   Max.   :247.0
##        X5              X6              X7              X8
##  Min.   : 0.000   Min.   : 801.0   Min.   :594.0   Min.   :  1.00
##  1st Qu.: 0.000   1st Qu.: 932.0   1st Qu.:731.0   1st Qu.:  7.00
##  Median : 6.350   Median : 968.0   Median :779.5   Median : 28.00
##  Mean   : 6.203   Mean   : 972.9   Mean   :773.6   Mean   : 45.66
##  3rd Qu.:10.160   3rd Qu.:1029.4   3rd Qu.:824.0   3rd Qu.: 56.00
##  Max.   :32.200   Max.   :1145.0   Max.   :992.6   Max.   :365.00
##        Y
##  Min.   : 2.332
##  1st Qu.:23.707
##  Median :34.443
##  Mean   :35.818
##  3rd Qu.:46.136
##  Max.   :82.599
```

```
names(concrete)
```

```
## [1] "X1" "X2" "X3" "X4" "X5" "X6" "X7" "X8" "Y"
```

```
head(concrete)
```

```
##      X1    X2 X3  X4  X5     X6    X7  X8        Y
## 1 540.0   0.0  0 162 2.5 1040.0 676.0  28 79.98611
## 2 540.0   0.0  0 162 2.5 1055.0 676.0  28 61.88737
## 3 332.5 142.5  0 228 0.0  932.0 594.0 270 40.26954
## 4 332.5 142.5  0 228 0.0  932.0 594.0 365 41.05278
## 5 198.6 132.4  0 192 0.0  978.4 825.5 360 44.29608
## 6 266.0 114.0  0 228 0.0  932.0 670.0  90 47.02985
```

```
mod.full<-lm(Y~.,data=concrete)
anova(mod.full)
```

```
## Analysis of Variance Table
##
## Response: Y
##             Df Sum Sq Mean Sq  F value    Pr(>F)
## X1           1  71172   71172 658.0463 < 2.2e-16 ***
## X2           1  22957   22957 212.2606 < 2.2e-16 ***
## X3           1  21636   21636 200.0464 < 2.2e-16 ***
## X4           1  11459   11459 105.9488 < 2.2e-16 ***
## X5           1   1360    1360  12.5785 0.0004079 ***
## X6           1    253     253   2.3435 0.1261178
## X7           1      1       1   0.0058 0.9393393
```

```
## X8           1  47905   47905 442.9232 < 2.2e-16 ***
## Residuals 1021 110428     108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the anova table above, 1021=n-8-1. So n=1030

a) forward selection with BIC

```
mod.0<-lm(Y~1, data=concrete) #linear model with only intercepts
mod.full<-lm(Y~.,data=concrete) #full model
step(mod.0, scope = list(lower = mod.0, upper = mod.full), direction = 'forwa
rd', k = log(1030))
```

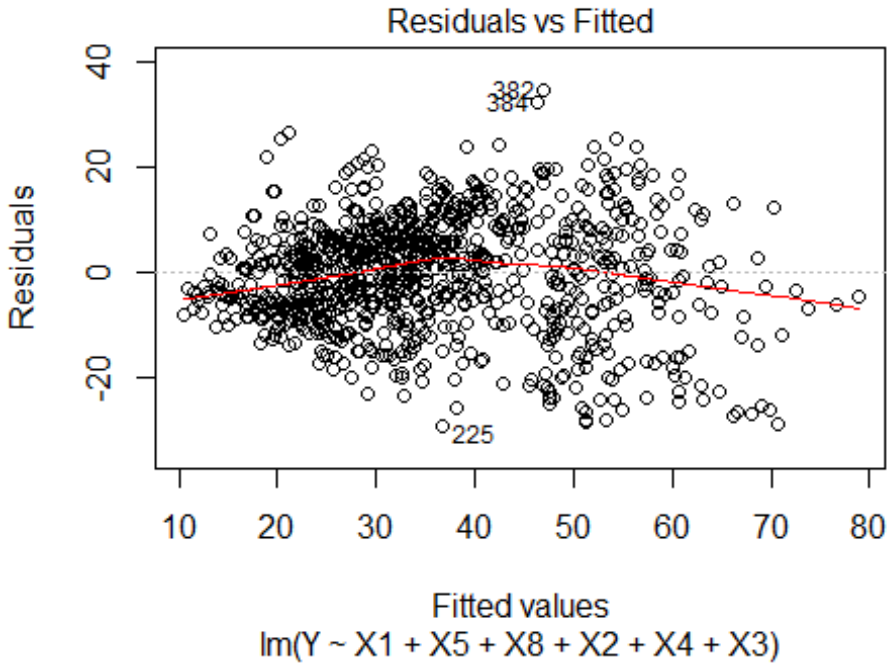
```
## Start:  AIC=5806.38
## Y ~ 1
##
##        Df Sum of Sq    RSS    AIC
## + X1    1     71172 216001 5520.0
## + X5    1     38490 248683 5665.1
## + X8    1     31061 256112 5695.4
## + X4    1     24087 263086 5723.1
## + X7    1      8033 279140 5784.1
## + X6    1      7811 279362 5784.9
## + X2    1      5220 281953 5794.4
## + X3    1      3212 283961 5801.7
## <none>              287173 5806.4
##
## Step:  AIC=5519.97
## Y ~ X1
##
##        Df Sum of Sq    RSS    AIC
## + X5    1   29646.5 186354 5374.8
## + X8    1   23993.8 192007 5405.6
## + X2    1   22957.4 193043 5411.2
## + X4    1   17926.8 198074 5437.7
## + X6    1    3548.0 212453 5509.8
## + X3    1    2894.4 213106 5513.0
## <none>              216001 5520.0
## + X7    1     960.2 215041 5522.3
##
## Step:  AIC=5374.85
## Y ~ X1 + X5
##
##        Df Sum of Sq    RSS    AIC
## + X8    1     37498 148857 5150.4
## + X2    1     19456 166898 5268.2
## + X7    1      5862 180493 5348.9
## <none>              186354 5374.8
## + X4    1       782 185572 5377.5
```

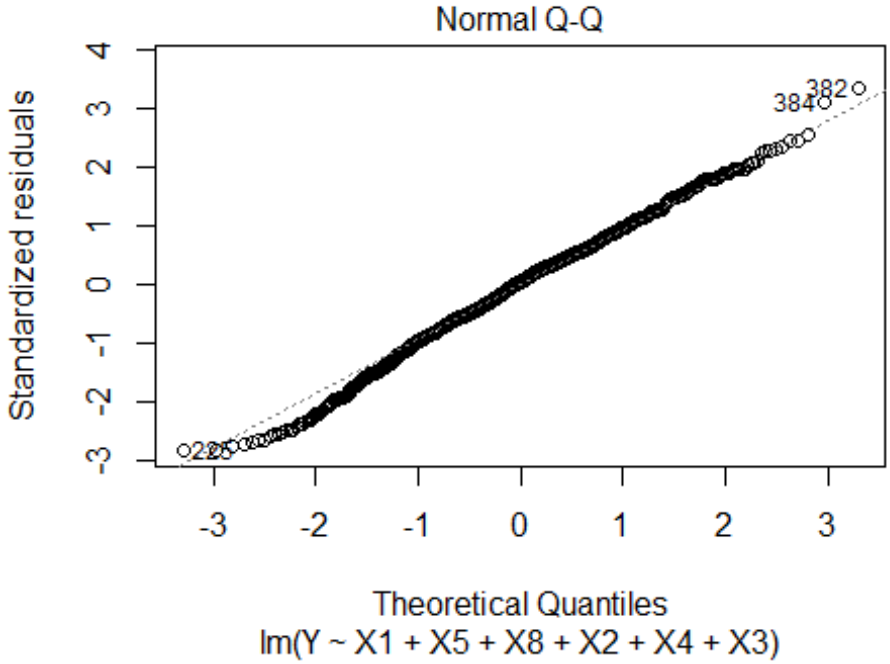
```
## + X3    1       741 185613 5377.7
## + X6    1       241 186113 5380.4
##
## Step:  AIC=5150.38
## Y ~ X1 + X5 + X8
##
##        Df Sum of Sq    RSS    AIC
## + X2    1   19908.5 128948 5009.4
## + X4    1    4868.8 143988 5123.1
## + X7    1    3385.5 145471 5133.6
## <none>              148857 5150.4
## + X3    1     323.9 148533 5155.1
## + X6    1      36.9 148820 5157.1
##
## Step:  AIC=5009.43
## Y ~ X1 + X5 + X8 + X2
##
##        Df Sum of Sq    RSS    AIC
## + X4    1    9544.7 119403 4937.2
## + X3    1    6524.7 122423 4962.9
## + X6    1    1737.0 127211 5002.4
## <none>              128948 5009.4
## + X7    1       3.5 128945 5016.3
##
## Step:  AIC=4937.16
## Y ~ X1 + X5 + X8 + X2 + X4
##
##        Df Sum of Sq    RSS    AIC
## + X3    1    8547.4 110856 4867.6
## + X7    1    1895.7 117508 4927.6
## <none>              119403 4937.2
## + X6    1      24.1 119379 4943.9
##
## Step:  AIC=4867.59
## Y ~ X1 + X5 + X8 + X2 + X4 + X3
##
##        Df Sum of Sq    RSS    AIC
## <none>              110856 4867.6
## + X6    1    44.271 110812 4874.1
## + X7    1    29.398 110827 4874.3

##
## Call:
## lm(formula = Y ~ X1 + X5 + X8 + X2 + X4 + X3, data = concrete)
##
## Coefficients:
## (Intercept)           X1           X5           X8           X2
##    29.03022      0.10543      0.23900      0.11349      0.08649
```

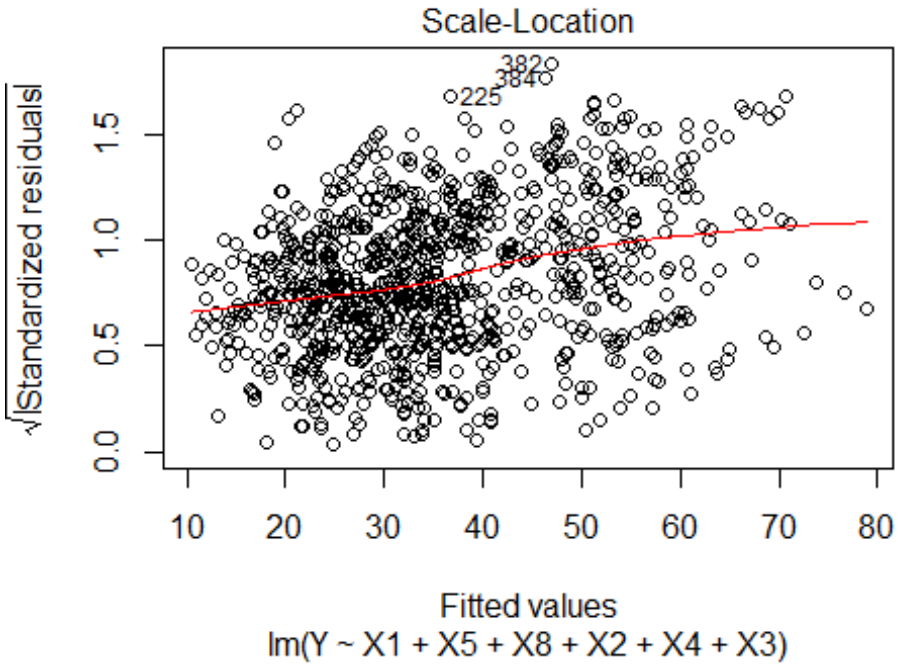
```
##          X4          X3
##    -0.21829     0.06871
```

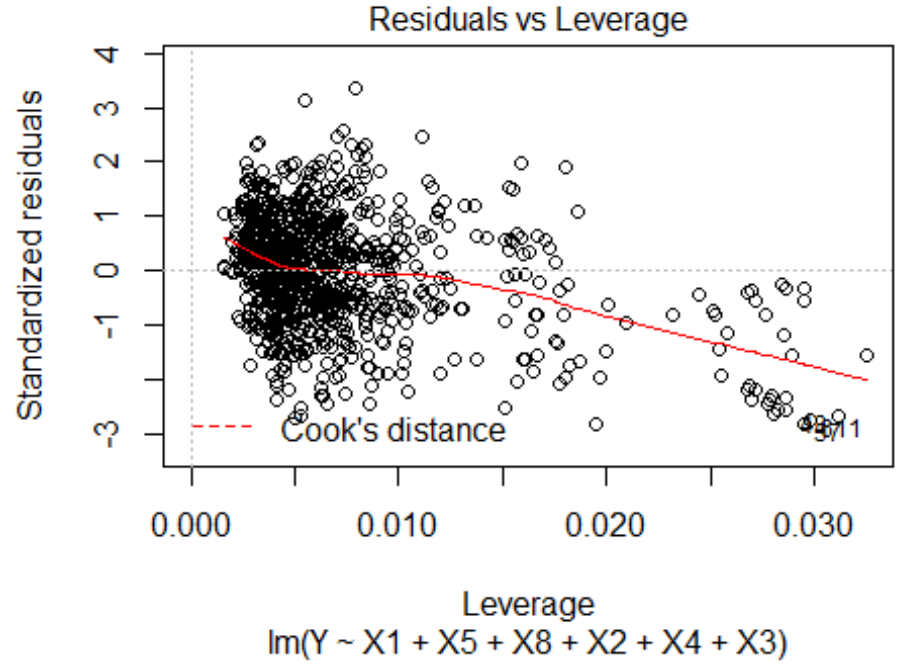
Do diagnostic checks

```
mod.for<-lm(Y ~ X1 + X5 + X8 + X2 + X4 + X3, data = concrete)
plot(mod.for)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Y ~ X1 + X5 + X8 + X2 + X4 + X3)
382
384
225
40
20
0
-20
10
20
30
40
50
60
70
80


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Y ~ X1 + X5 + X8 + X2 + X4 + X3)
382
384
225
4
3
2
1
0
-1
-2
-3

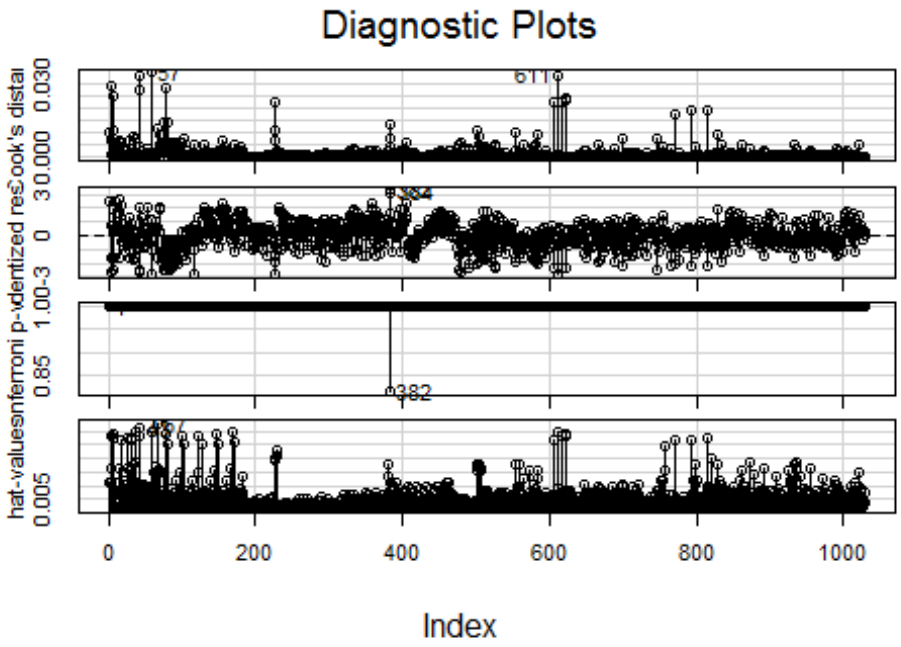




According to the plots, model by forward selection pretty fufills noramlity, linearity, and constant variance.

Use influenceIndexPlot to find influential points to remove

```
library(car)
infIndexPlot(mod.for)
```




```
predict(mod.for, data.frame(X1= 200, X5=10, X8=100, X2=150, X4=180, X3=85),
interval = 'confidence', level = 0.95)

##        fit      lwr      upr
## 1 43.37686 42.12569 44.62803
```

We are 95% confident that true mean response is between 42.126 and 44.628

```
predict(mod.for, data.frame(X1= 200, X5=10, X8=100, X2=150, X4=180, X3=85),
interval = 'prediction', level = 0.95)

##        fit      lwr      upr
## 1 43.37686 22.91161 63.84211
```

We are 95% confident that concrete compressive strength for and individual value of each predictor values is between 22.912 and 63.842

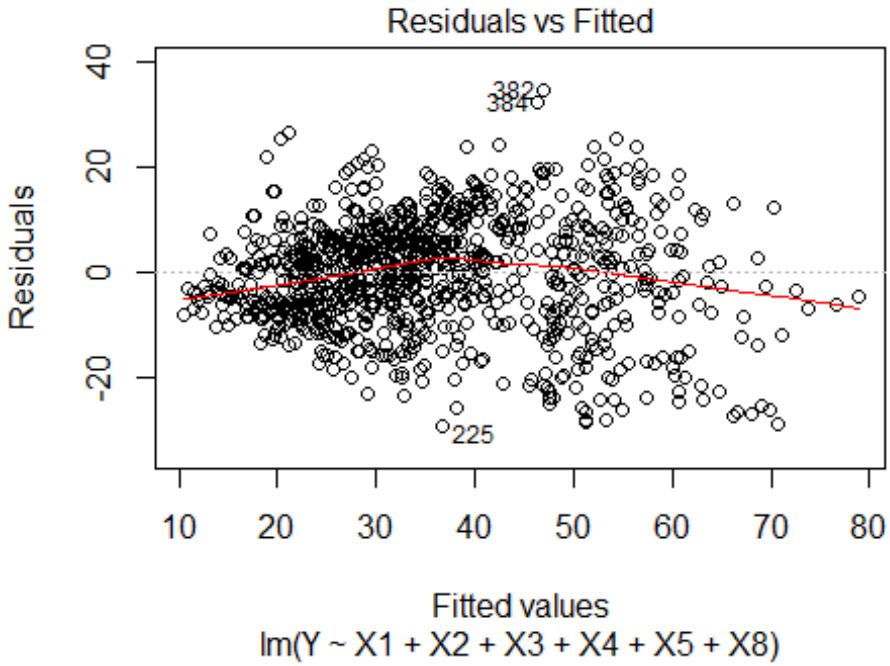
(b) Backward elimination with BIC

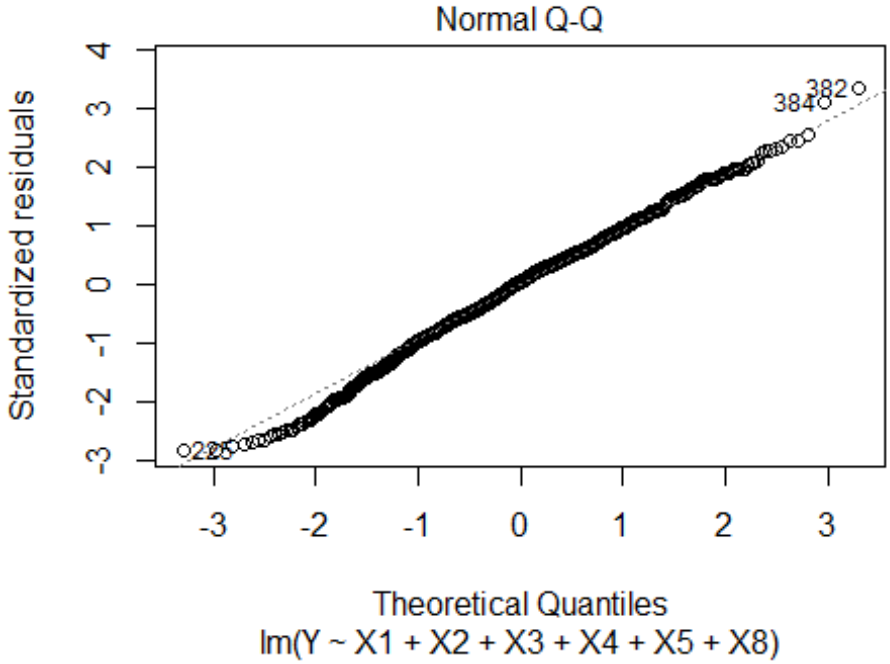
```
step(mod.full, scope = list(lower = mod.0, upper = mod.full), direction = 'ba
ckward', k = log(1030))
```

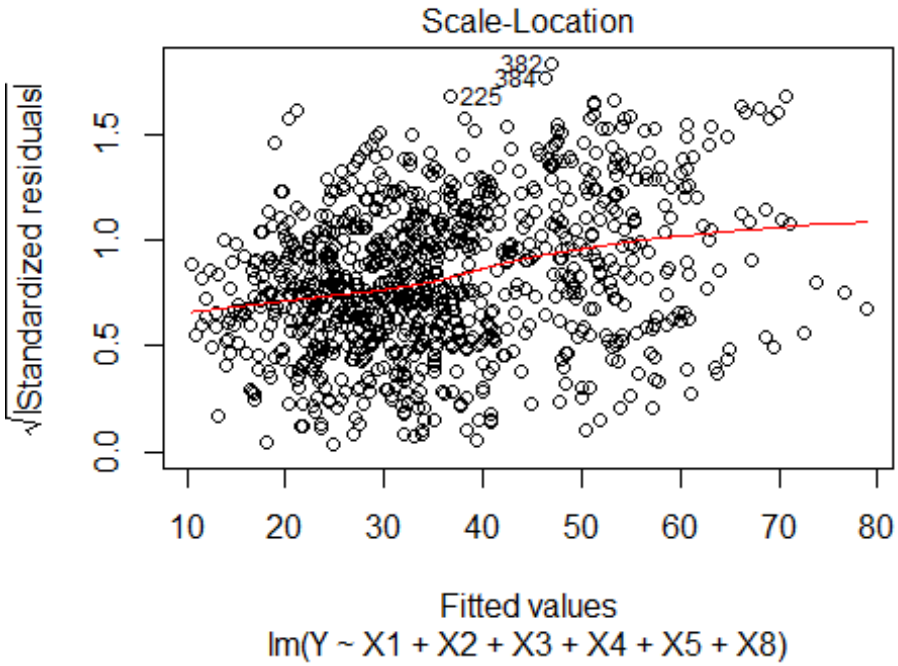
```
## Start:  AIC=4877.49
## Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8
##
##        Df Sum of Sq    RSS    AIC
## - X7    1       384 110812 4874.1
## - X6    1       398 110827 4874.3
## <none>              110428 4877.5
## - X5    1      1046 111474 4880.3
## - X4    1      1513 111942 4884.6
## - X3    1      5281 115709 4918.7
## - X2    1     11353 121781 4971.3
## - X1    1     21533 131961 5054.0
## - X8    1     47905 158333 5241.7
##
## Step:  AIC=4874.12
## Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X8
##
##        Df Sum of Sq    RSS    AIC
## - X6    1        44 110856 4867.6
## <none>              110812 4874.1
## - X5    1       877 111688 4875.3
## - X4    1      8526 119338 4943.5
## - X3    1      8568 119379 4943.9
## - X2    1     30693 141505 5119.0
## - X8    1     47522 158334 5234.8
## - X1    1     64008 174819 5336.8
##
## Step:  AIC=4867.59
## Y ~ X1 + X2 + X3 + X4 + X5 + X8
##
##        Df Sum of Sq    RSS    AIC
## <none>              110856 4867.6
## - X5    1       865 111721 4868.7
## - X3    1      8547 119403 4937.2
## - X4    1     11567 122423 4962.9
## - X2    1     32757 143613 5127.3
## - X8    1     47731 158587 5229.5
## - X1    1     66760 177616 5346.2

##
## Call:
## lm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X8, data = concrete)
##
## Coefficients:
## (Intercept)           X1           X2           X3           X4
##    29.03022      0.10543      0.08649      0.06871     -0.21829
##          X5           X8
##     0.23900      0.11349
```

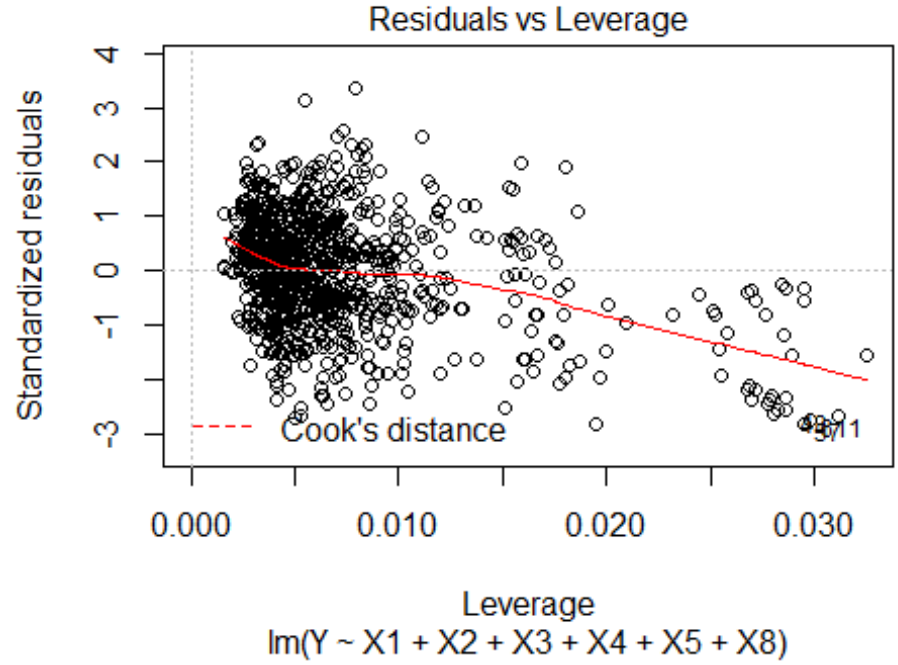
Do diagnostic checks

```
mod.back<-lm(Y ~ X1 + X2 + X3 + X4 + X5 + X8, data=concrete)
plot(mod.back)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Y ~ X1 + X2 + X3 + X4 + X5 + X8)
382
384
225


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Y ~ X1 + X2 + X3 + X4 + X5 + X8)
382
384
225

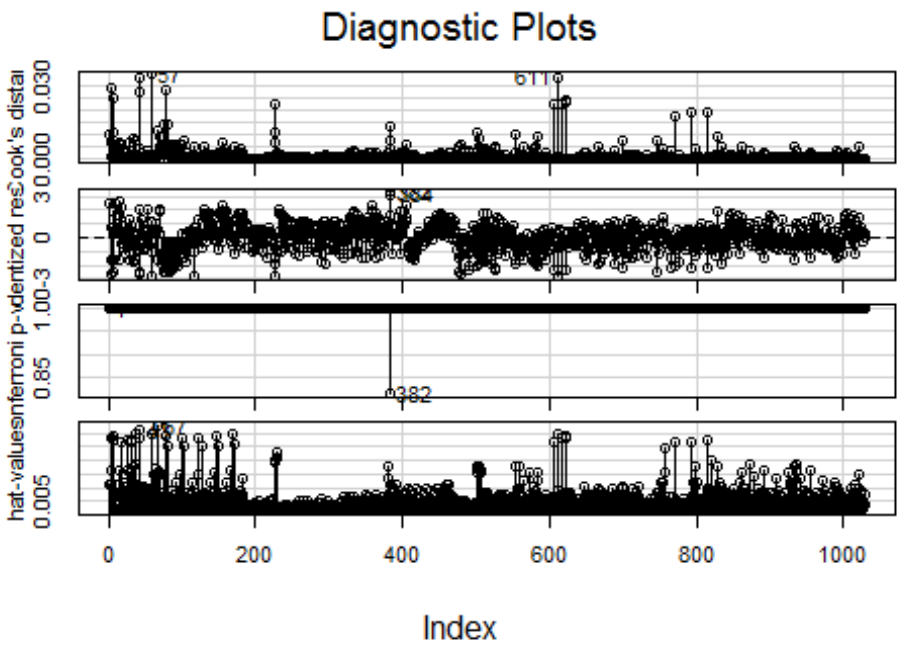




According to the plots, model by forward selection pretty fufills noramlity, linearity, and constant variance.

Use influenceIndexPlot to find influential points to remove

```
library(car)
infIndexPlot(mod.back)
```




```
anova(mod.for, mod.back)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X5 + X8 + X2 + X4 + X3
## Model 2: Y ~ X1 + X2 + X3 + X4 + X5 + X8
##   Res.Df    RSS Df  Sum of Sq F Pr(>F)
## 1   1023 110856
## 2   1023 110856  0 5.8208e-11
```

The models derived by two differnt methods are same except for the order of predictors. And the result of ANOVA also same.