# Group1_Project

*Group_1*

*April 29, 2020*

## Introduction:

The background of this data comes from ...

## Read In Data:

```
## Loading required package: ggplot2

## Loading required package: carData

##
## Attaching package: 'olsrr'

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

## Loading required package: KernSmooth

## KernSmooth 2.23 loaded
## Copyright M. P. Wand 1997-2009

##
## Attaching package: 'MPV'

## The following object is masked from 'package:olsrr':
##
##     cement
```

## Performing variable selection:

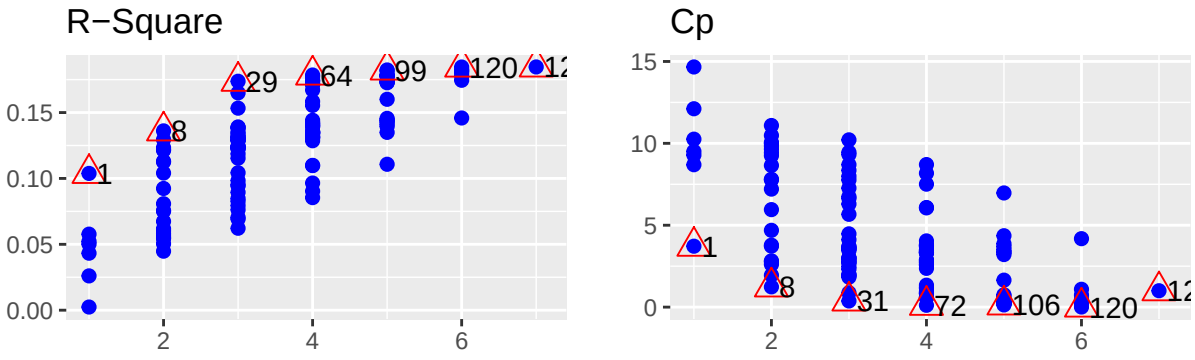
```
##     Index N           Predictors   R-Square Adj. R-Square Mallow's Cp
## 1       1 1                   x1 0.10381011   0.094276177    4.719250
## 2       2 1                   x2 0.05771212   0.047687785    9.694270
## 5       3 1                   x5 0.05223762   0.042155039   10.285094
## 7       4 1                   x7 0.05016001   0.040055334   10.509314
## 6       5 1                   x6 0.04324888   0.033070675   11.255183
## 4       6 1                   x4 0.02603395   0.015672614   13.113064
## 3       7 1                   x3 0.00239633  -0.008216475   15.664101
## 12      8 2               x1 x6 0.13609838   0.117519847    3.234613
## 13      9 2               x1 x7 0.12950341   0.110783053    3.946359
## 11     10 2               x1 x5 0.12339323   0.104541471    4.605787
## 27     11 2               x5 x7 0.12181247   0.102926715    4.776387
## 8      12 2               x1 x2 0.12135810   0.102462571    4.825424
## 26     13 2               x5 x6 0.11302532   0.093950600    5.724719
## 10     14 2               x1 x4 0.11249746   0.093411385    5.781688
## 9      15 2               x1 x3 0.10408452   0.084817518    6.689635
## 16     16 2               x2 x5 0.09237954   0.072860824    7.952868
## 23     17 2               x4 x5 0.08077039   0.061002009    9.205760
```

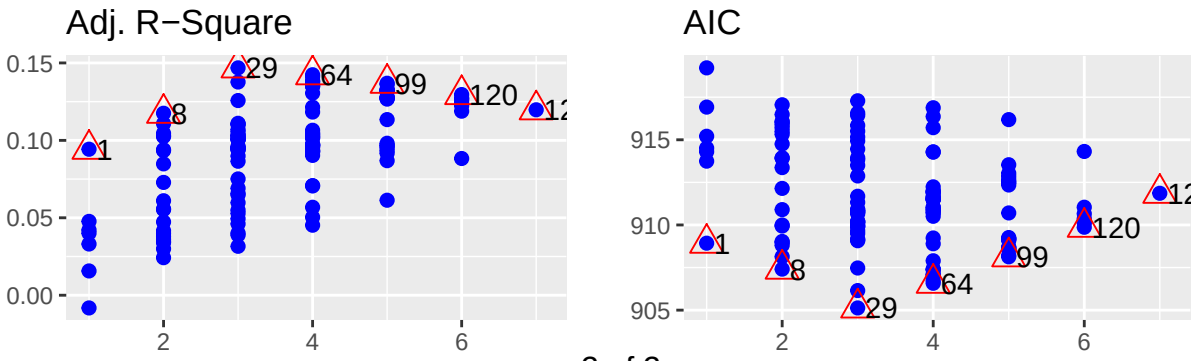
```
## 18     18 2                x2 x7 0.07548215   0.055600043    9.776481
## 21     19 2                x3 x6 0.07523058   0.055343063    9.803631
## 17     20 2                x2 x6 0.06739014   0.047334010   10.649793
## 14     21 2                x2 x3 0.06184974   0.041674465   11.247728
## 15     22 2                x2 x4 0.06017702   0.039965774   11.428252
## 19     23 2                x3 x4 0.05812812   0.037872815   11.649374
## 28     24 2                x6 x7 0.05636761   0.036074437   11.839374
## 20     25 2                x3 x5 0.05494962   0.034625957   11.992407
## 22     26 2                x3 x7 0.05435293   0.034016434   12.056804
## 25     27 2                x4 x7 0.05042085   0.029999793   12.481164
## 24     28 2                x4 x6 0.04477933   0.024236954   13.090012
## 36     29 3             x1 x3 x6 0.17375716   0.146814459    1.170374
## 57     30 3             x3 x5 x6 0.16492514   0.137694443    2.123549
## 39     31 3             x1 x4 x6 0.15332894   0.125720104    3.375043
## 43     32 3             x1 x6 x7 0.13908854   0.111015343    4.911906
## 41     33 3             x1 x5 x6 0.13863843   0.110550554    4.960483
## 32     34 3             x1 x2 x6 0.13819982   0.110097641    5.007819
## 33     35 3             x1 x2 x7 0.13525894   0.107060865    5.325207
## 37     36 3             x1 x3 x7 0.13378473   0.105538584    5.484307
## 31     37 3             x1 x2 x5 0.13170370   0.103389686    5.708899
## 40     38 3             x1 x4 x7 0.13131963   0.102993099    5.750348
## 63     39 3             x5 x6 x7 0.13122129   0.102891547    5.760961
## 42     40 3             x1 x5 x7 0.12979032   0.101413916    5.915395
## 34     41 3             x1 x3 x4 0.12906703   0.100667039    5.993455
## 52     42 3             x2 x5 x7 0.12900997   0.100608120    5.999613
## 58     43 3             x3 x5 x7 0.12886586   0.100459314    6.015165
## 61     44 3             x4 x5 x7 0.12478309   0.096243412    6.455789
## 38     45 3             x1 x4 x5 0.12385098   0.095280903    6.556385
## 35     46 3             x1 x3 x5 0.12369898   0.095123946    6.572789
## 29     47 3             x1 x2 x3 0.12350282   0.094921386    6.593960
## 30     48 3             x1 x2 x4 0.12273957   0.094133250    6.676332
## 60     49 3             x4 x5 x6 0.12237581   0.093757629    6.715589
## 51     50 3             x2 x5 x6 0.11863478   0.089894607    7.119332
## 54     51 3             x3 x4 x5 0.11536571   0.086518943    7.472138
## 46     52 3             x2 x3 x6 0.10430631   0.075098911    8.665699
## 48     53 3             x2 x4 x5 0.09846917   0.069071426    9.295659
## 44     54 3             x2 x3 x4 0.09529396   0.065792677    9.638336
## 45     55 3             x2 x3 x5 0.09429931   0.064765588    9.745682
## 47     56 3             x2 x3 x7 0.08944975   0.059757897   10.269059
## 59     57 3             x3 x6 x7 0.08485610   0.055014456   10.764818
## 55     58 3             x3 x4 x6 0.08273400   0.052823147   10.993842
## 50     59 3             x2 x4 x7 0.07942053   0.049401637   11.351440
## 53     60 3             x2 x6 x7 0.07619390   0.046069787   11.699667
## 49     61 3             x2 x4 x6 0.07049820   0.040188359   12.314362
## 62     62 3             x4 x6 x7 0.06917229   0.038819216   12.457458
## 56     63 3             x3 x4 x7 0.06218192   0.031600897   13.211878
## 97     64 4          x3 x5 x6 x7 0.17848419   0.142373608    2.660220
## 66     65 4          x1 x2 x3 x6 0.17726370   0.141099463    2.791939
## 79     66 4          x1 x3 x6 x7 0.17492850   0.138661621    3.043960
## 77     67 4          x1 x3 x5 x6 0.17404349   0.137737707    3.139473
## 75     68 4          x1 x3 x4 x6 0.17397322   0.137664349    3.147057
## 98     69 4          x4 x5 x6 x7 0.17157653   0.135162314    3.405714
## 87     70 4          x2 x3 x5 x6 0.17154573   0.135130155    3.409038
## 82     71 4          x1 x4 x6 x7 0.17140111   0.134979177    3.424646
```

```
## 94     72 4          x3 x4 x5 x6 0.16711214   0.130501683    3.887523
## 80     73 4          x1 x4 x5 x6 0.15840802   0.121414967    4.826895
## 69     74 4          x1 x2 x4 x6 0.15542635   0.118302235    5.148685
## 67     75 4          x1 x2 x3 x7 0.14413821   0.106517911    6.366932
## 64     76 4          x1 x2 x3 x4 0.14232843   0.104628578    6.562249
## 88     77 4          x2 x3 x5 x7 0.14199863   0.104284284    6.597841
## 73     78 4          x1 x2 x6 x7 0.14085571   0.103091131    6.721188
## 71     79 4          x1 x2 x5 x6 0.14043845   0.102655525    6.766220
## 70     80 4          x1 x2 x4 x7 0.13945870   0.101632710    6.871957
## 83     81 4          x1 x5 x6 x7 0.13934523   0.101514250    6.884204
## 72     82 4          x1 x2 x5 x7 0.13547819   0.097477231    7.301545
## 91     83 4          x2 x4 x5 x7 0.13511568   0.097098785    7.340668
## 65     84 4          x1 x2 x3 x5 0.13505935   0.097039977    7.346748
## 84     85 4          x2 x3 x4 x5 0.13501751   0.096996305    7.351262
## 76     86 4          x1 x3 x4 x7 0.13493486   0.096910016    7.360183
## 93     87 4          x2 x5 x6 x7 0.13396382   0.095896297    7.464980
## 78     88 4          x1 x3 x5 x7 0.13378493   0.095709538    7.484287
## 68     89 4          x1 x2 x4 x5 0.13182026   0.093658510    7.696319
## 81     90 4          x1 x4 x5 x7 0.13139466   0.093214208    7.742250
## 74     91 4          x1 x3 x4 x5 0.13104014   0.092844106    7.780511
## 95     92 4          x3 x4 x5 x7 0.13071638   0.092506115    7.815452
## 90     93 4          x2 x4 x5 x6 0.12860550   0.090302449    8.043264
## 85     94 4          x2 x3 x4 x6 0.10990566   0.070780635   10.061402
## 89     95 4          x2 x3 x6 x7 0.10973595   0.070603461   10.079718
## 86     96 4          x2 x3 x4 x7 0.09650171   0.056787504   11.507992
## 92     97 4          x2 x4 x6 x7 0.09033265   0.050347271   12.173774
## 96     98 4          x3 x4 x6 x7 0.08541813   0.045216728   12.704163
## 117    99 5       x2 x3 x5 x6 x7 0.18231009   0.136882869    4.247319
## 119   100 5       x3 x4 x5 x6 x7 0.18160687   0.136140585    4.323212
## 112   101 5       x1 x3 x5 x6 x7 0.17867047   0.133041053    4.640116
## 104   102 5       x1 x2 x3 x6 x7 0.17814076   0.132481910    4.697285
## 111   103 5       x1 x3 x4 x6 x7 0.17784127   0.132165790    4.729606
## 102   104 5       x1 x2 x3 x5 x6 0.17778880   0.132110400    4.735269
## 100   105 5       x1 x2 x3 x4 x6 0.17739636   0.131696157    4.777622
## 109   106 5       x1 x3 x4 x5 x6 0.17409726   0.128213775    5.133670
## 114   107 5       x2 x3 x4 x5 x6 0.17351105   0.127594995    5.196935
## 113   108 5       x1 x4 x5 x6 x7 0.17324210   0.127311108    5.225961
## 118   109 5       x2 x4 x5 x6 x7 0.17316521   0.127229943    5.234259
## 107   110 5       x1 x2 x4 x6 x7 0.17262053   0.126654999    5.293043
## 105   111 5       x1 x2 x4 x5 x6 0.16007770   0.113415354    6.646699
## 101   112 5       x1 x2 x3 x4 x7 0.14560024   0.098133582    8.209146
## 103   113 5       x1 x2 x3 x5 x7 0.14430710   0.096768602    8.348705
## 115   114 5       x2 x3 x4 x5 x7 0.14410617   0.096556515    8.370390
## 99    115 5       x1 x2 x3 x4 x5 0.14277288   0.095149151    8.514282
## 108   116 5       x1 x2 x5 x6 x7 0.14107797   0.093360085    8.697201
## 106   117 5       x1 x2 x4 x5 x7 0.13945973   0.091651937    8.871846
## 110   118 5       x1 x3 x4 x5 x7 0.13494599   0.086887438    9.358981
## 116   119 5       x2 x3 x4 x6 x7 0.11074722   0.061344289   11.970578
## 126   120 6    x2 x3 x4 x5 x6 x7 0.18455359   0.129579671    6.005195
## 123   121 6    x1 x2 x3 x5 x6 x7 0.18233077   0.127207007    6.245087
## 125   122 6    x1 x3 x4 x5 x6 x7 0.18183780   0.126680797    6.298290
## 122   123 6    x1 x2 x3 x4 x6 x7 0.18023945   0.124974691    6.470788
## 120   124 6    x1 x2 x3 x4 x5 x6 0.17778881   0.122358838    6.735268
## 124   125 6    x1 x2 x4 x5 x6 x7 0.17450663   0.118855394    7.089489
```

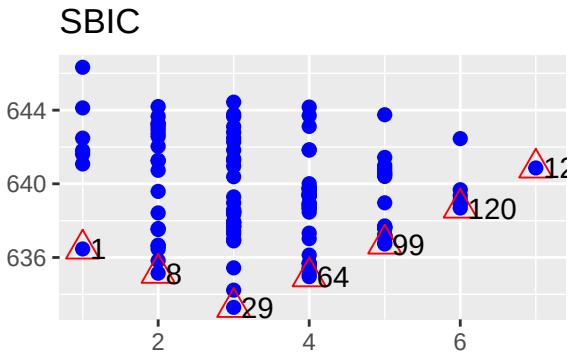
```
## 121   126 6    x1 x2 x3 x4 x5 x7 0.14587398   0.088292448   10.179603
## 127   127 7 x1 x2 x3 x4 x5 x6 x7 0.18460172   0.119740493    8.000000
```

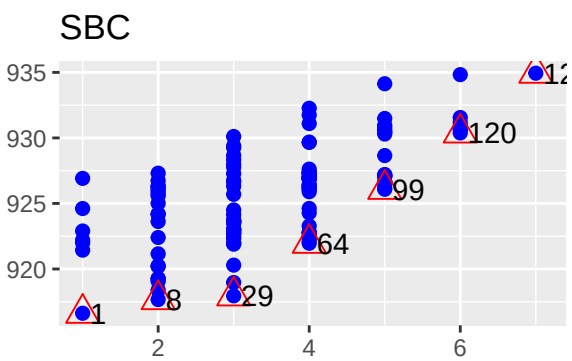
page 1 of 2




page 2 of 2




# Interpretation of variable Selection: In looking for the "best" model, certain criteria must be met inorder for proper variable selection of the regressor

equation. These criteria help us to be able to explain the data in the simpliest way with redundant predictors removed inorder minimize cost and to avoid multi-collinearity in our regression model.

The criteria for our variable selection include: 1) Large R^2 value 2) Maximum Adjusted R^2 value 3) Minimum MSres 4) Minimum Mallow's Cp Statistic value

Based on the above criteria, the "best" candidate models are: 1) Model 29: y ~ x1 + x3 + x6 2) Model 64: Y ~ x3 + x5 + x6 + x7 3) Model 99: Y ~ x2 + x3 + x5 + x6 + x7 4) Model 120: Y ~ x2 + x3 + x4 + x5 + x6 + x7

Once we identified the "best"" candidate models, we compare its predicted residual error sum of squares (PRESS) statistic with other candidate models and selected the model with the smallest value. We also compare candidate models by performing a variance inflation factor (VIF) in order to quantify the severity of multicollinearity in the model.

```
## [1] 68704.26
```

```
## [1] 69789.7
```

```
## [1] 71110.33
```

```
## [1] 77659.85
```

```
##       x1       x3       x6
## 1.010615 2.528279 2.523045
```

```
##       x3       x5       x6       x7
## 2.617428 1.060150 3.650110 1.820798
```

```
##       x2       x3       x5       x6       x7
## 1.564314 2.622358 1.184723 3.883741 1.878902
```

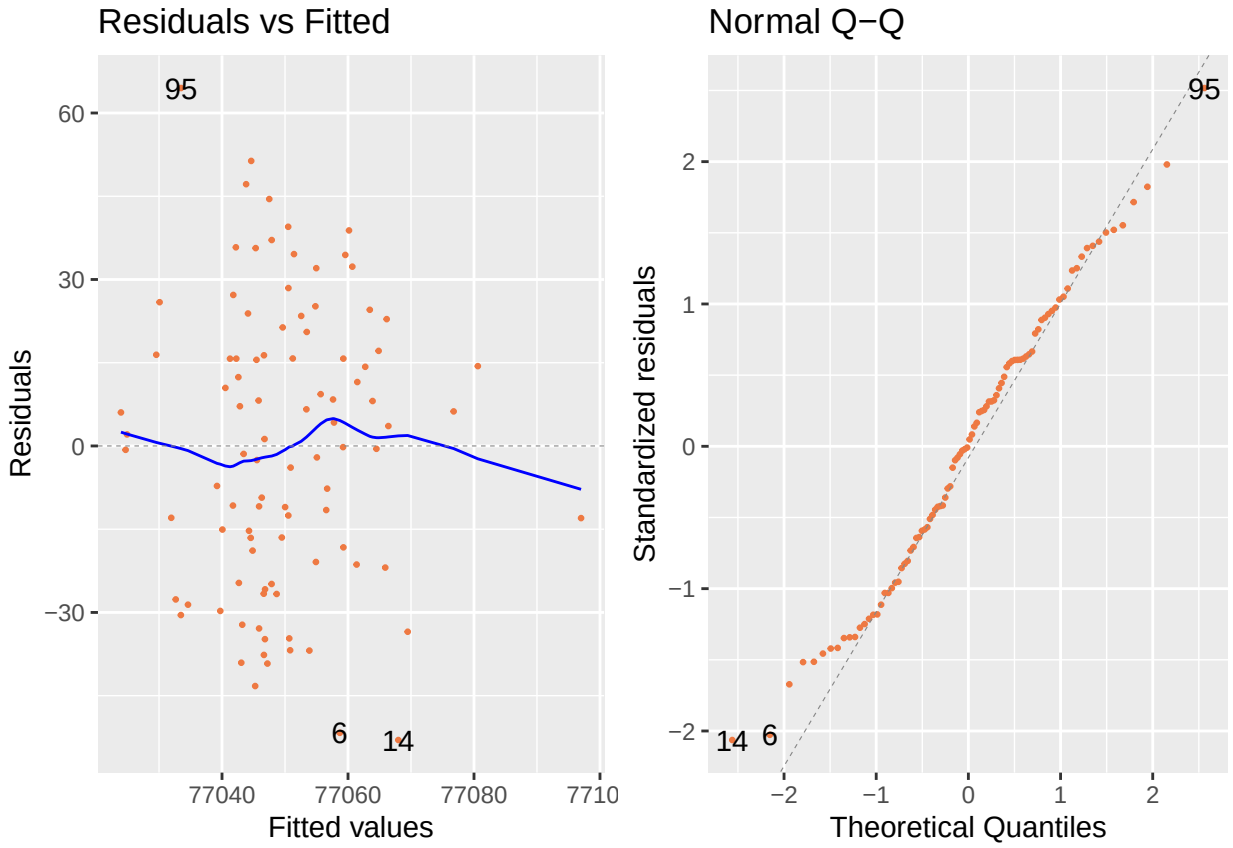
```
##        x2        x3        x4        x5        x6        x7
##  1.598317  6.222163 14.323819  1.290660  4.818502  4.102786
```

## Interpretation of PRESS and Vif of candidate models:

The model with the lowest PRESS value in model 29, Y = x1 + x3 + x6 and the same model doesnt show any evidence of multicolinearity in the variance inflation factor test of each regressor.

## Plot of model:

```
autoplot(fit29,size = 0.5, colour = 'sienna2')[1:2]
```




**other CSV links**

#TX_gz_num <- read.csv("CSV/DATA/texas_Gazetteer_numeric.csv") #products <- read.csv("CSV/OTHER/products.csv #test_data <- read.csv("CSV/OTHER/test_data.csv") #TX_gz_obj <- read.csv("CSV/OTHER/texas_Gazetteer_object.cs #train_dat <- read.csv("CSV/OTHER/train_data.csv")