HW8

11.1

Stepwise Regression: Though the homework specifically asks for Stepwise regression I performed all three types: Stepwise, Forward Selection and Backward Elimination mainly to be able to compare and contrast among the three approaches.

After Stepwise regression we can see that the following attributes are deemed to be the most important ones: M, Ed, Po1, M.F, U1, U2, Ineq, Prob. Also, the R-squared value is 0.788, and a standard error is 195.5.

Next, backward propagation produces similar results - M, Ed, Po1, M.F, U1, U2, Ineq, Prob are the important attributes with the same standard error and R-squared as stepwise regression.

However, when I performed forward selection, I obtained different set of important attributes which are as follows: Po1, Ineq, Ed, M, Prob, U2. The standard error is 200.7 and R-squared is 0.765, both are worse than Stepwise and Backward elimination.

The AKAIC scores for the three models are as follows: Stepwise 639.31, Backward elimination is 639.31 and for Forward selection is 640.16. This is consistent with the above observation that both Stepwise and Backward elimination are good for the dataset.

Lasso and Elastic Net

First, I split the data into training and test set and scaled the data as needed. I then ran the regression model for varying values of alpha - i.e. when alpha is 0 the model would be a Ridge regression, when it is 1 then it is a Lasso regression and anything in between it would be an Elastic net regression.

As shown in the results and the graph, the MSE for Elastic Net Regression is the lowest - 1,009,750 so we can conclude that Elastic Net regression is best suited for this dataset.

```r
crime_data = read.table("uscrime.txt", header=TRUE)
#crime_data

mod = lm(Crime ~., data=crime_data)
summary(mod)
```

```
##
## Call:
## lm(formula = Crime ~ ., data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -395.74  -98.09   -6.69  112.99  512.67
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.984e+03  1.628e+03  -3.675 0.000893 ***
## M            8.783e+01  4.171e+01   2.106 0.043443 *
## So          -3.803e+00  1.488e+02  -0.026 0.979765
```

```
## Ed           1.883e+02  6.209e+01   3.033 0.004861 **
## Po1          1.928e+02  1.061e+02   1.817 0.078892 .
## Po2         -1.094e+02  1.175e+02  -0.931 0.358830
## LF          -6.638e+02  1.470e+03  -0.452 0.654654
## M.F          1.741e+01  2.035e+01   0.855 0.398995
## Pop         -7.330e-01  1.290e+00  -0.568 0.573845
## NW           4.204e+00  6.481e+00   0.649 0.521279
## U1          -5.827e+03  4.210e+03  -1.384 0.176238
## U2           1.678e+02  8.234e+01   2.038 0.050161 .
## Wealth       9.617e-02  1.037e-01   0.928 0.360754
## Ineq         7.067e+01  2.272e+01   3.111 0.003983 **
## Prob        -4.855e+03  2.272e+03  -2.137 0.040627 *
## Time        -3.479e+00  7.165e+00  -0.486 0.630708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209.1 on 31 degrees of freedom
## Multiple R-squared:  0.8031, Adjusted R-squared:  0.7078
## F-statistic: 8.429 on 15 and 31 DF,  p-value: 3.539e-07
```

```
## Stepwise Regression
step_mod = step(mod, direction="both")
```

```
## Start:  AIC=514.65
## Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 +
##     U2 + Wealth + Ineq + Prob + Time
##
##          Df Sum of Sq     RSS    AIC
## - So      1        29 1354974 512.65
## - LF      1      8917 1363862 512.96
## - Time    1     10304 1365250 513.00
## - Pop     1     14122 1369068 513.14
## - NW      1     18395 1373341 513.28
## - M.F     1     31967 1386913 513.74
## - Wealth  1     37613 1392558 513.94
## - Po2     1     37919 1392865 513.95
## <none>                1354946 514.65
## - U1      1     83722 1438668 515.47
## - Po1     1    144306 1499252 517.41
## - U2      1    181536 1536482 518.56
## - M       1    193770 1548716 518.93
## - Prob    1    199538 1554484 519.11
## - Ed      1    402117 1757063 524.86
## - Ineq    1    423031 1777977 525.42
##
## Step:  AIC=512.65
## Crime ~ M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 +
##     Wealth + Ineq + Prob + Time
##
##          Df Sum of Sq     RSS    AIC
## - Time    1     10341 1365315 511.01
## - LF      1     10878 1365852 511.03
## - Pop     1     14127 1369101 511.14
## - NW      1     21626 1376600 511.39
```

```
## - M.F     1     32449 1387423 511.76
## - Po2     1     37954 1392929 511.95
## - Wealth  1     39223 1394197 511.99
## <none>                1354974 512.65
## - U1      1     96420 1451395 513.88
## + So      1        29 1354946 514.65
## - Po1     1    144302 1499277 515.41
## - U2      1    189859 1544834 516.81
## - M       1    195084 1550059 516.97
## - Prob    1    204463 1559437 517.26
## - Ed      1    403140 1758114 522.89
## - Ineq    1    488834 1843808 525.13
##
## Step:  AIC=511.01
## Crime ~ M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 +
##     Wealth + Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## - LF      1     10533 1375848 509.37
## - NW      1     15482 1380797 509.54
## - Pop     1     21846 1387161 509.75
## - Po2     1     28932 1394247 509.99
## - Wealth  1     36070 1401385 510.23
## - M.F     1     41784 1407099 510.42
## <none>                1365315 511.01
## - U1      1     91420 1456735 512.05
## + Time    1     10341 1354974 512.65
## + So      1        65 1365250 513.00
## - Po1     1    134137 1499452 513.41
## - U2      1    184143 1549458 514.95
## - M       1    186110 1551425 515.01
## - Prob    1    237493 1602808 516.54
## - Ed      1    409448 1774763 521.33
## - Ineq    1    502909 1868224 523.75
##
## Step:  AIC=509.37
## Crime ~ M + Ed + Po1 + Po2 + M.F + Pop + NW + U1 + U2 + Wealth +
##     Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## - NW      1     11675 1387523 507.77
## - Po2     1     21418 1397266 508.09
## - Pop     1     27803 1403651 508.31
## - M.F     1     31252 1407100 508.42
## - Wealth  1     35035 1410883 508.55
## <none>                1375848 509.37
## - U1      1     80954 1456802 510.06
## + LF      1     10533 1365315 511.01
## + Time    1      9996 1365852 511.03
## + So      1      3046 1372802 511.26
## - Po1     1    123896 1499744 511.42
## - U2      1    190746 1566594 513.47
## - M       1    217716 1593564 514.27
## - Prob    1    226971 1602819 514.54
```

```
## - Ed      1    413254 1789103 519.71
## - Ineq    1    500944 1876792 521.96
##
## Step:  AIC=507.77
## Crime ~ M + Ed + Po1 + Po2 + M.F + Pop + U1 + U2 + Wealth + Ineq +
##     Prob
##
##          Df Sum of Sq     RSS    AIC
## - Po2     1     16706 1404229 506.33
## - Pop     1     25793 1413315 506.63
## - M.F     1     26785 1414308 506.66
## - Wealth  1     31551 1419073 506.82
## <none>                1387523 507.77
## - U1      1     83881 1471404 508.52
## + NW      1     11675 1375848 509.37
## + So      1      7207 1380316 509.52
## + LF      1      6726 1380797 509.54
## + Time    1      4534 1382989 509.61
## - Po1     1    118348 1505871 509.61
## - U2      1    201453 1588976 512.14
## - Prob    1    216760 1604282 512.59
## - M       1    309214 1696737 515.22
## - Ed      1    402754 1790276 517.74
## - Ineq    1    589736 1977259 522.41
##
## Step:  AIC=506.33
## Crime ~ M + Ed + Po1 + M.F + Pop + U1 + U2 + Wealth + Ineq +
##     Prob
##
##          Df Sum of Sq     RSS    AIC
## - Pop     1     22345 1426575 505.07
## - Wealth  1     32142 1436371 505.39
## - M.F     1     36808 1441037 505.54
## <none>                1404229 506.33
## - U1      1     86373 1490602 507.13
## + Po2     1     16706 1387523 507.77
## + NW      1      6963 1397266 508.09
## + So      1      3807 1400422 508.20
## + LF      1      1986 1402243 508.26
## + Time    1       575 1403654 508.31
## - U2      1    205814 1610043 510.76
## - Prob    1    218607 1622836 511.13
## - M       1    307001 1711230 513.62
## - Ed      1    389502 1793731 515.83
## - Ineq    1    608627 2012856 521.25
## - Po1     1   1050202 2454432 530.57
##
## Step:  AIC=505.07
## Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Wealth + Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## - Wealth  1     26493 1453068 503.93
## <none>                1426575 505.07
## - M.F     1     84491 1511065 505.77
```

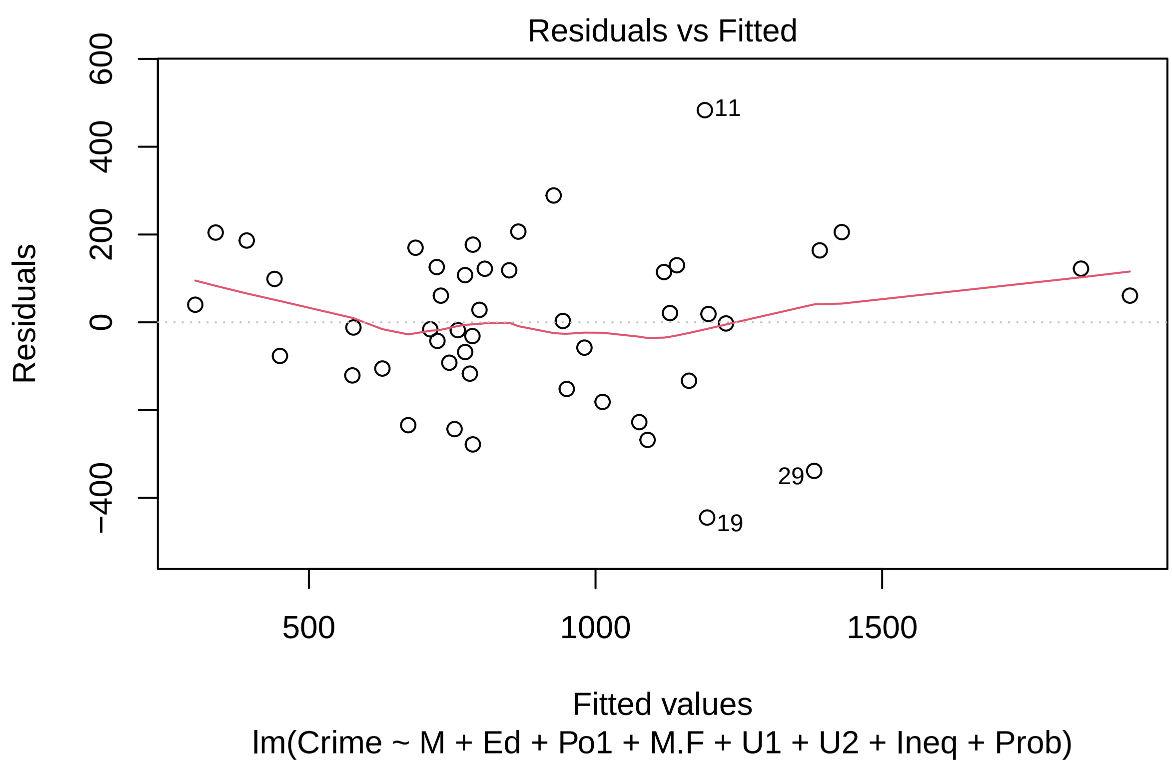
```
## - U1      1     99463 1526037 506.24
## + Pop     1     22345 1404229 506.33
## + Po2     1     13259 1413315 506.63
## + NW      1      5927 1420648 506.87
## + So      1      5724 1420851 506.88
## + LF      1      5176 1421398 506.90
## + Time    1      3913 1422661 506.94
## - Prob    1    198571 1625145 509.20
## - U2      1    208880 1635455 509.49
## - M       1    320926 1747501 512.61
## - Ed      1    386773 1813348 514.35
## - Ineq    1    594779 2021354 519.45
## - Po1     1   1127277 2553852 530.44
##
## Step:  AIC=503.93
## Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## <none>                 1453068 503.93
## + Wealth  1     26493 1426575 505.07
## - M.F     1    103159 1556227 505.16
## + Pop     1     16697 1436371 505.39
## + Po2     1     14148 1438919 505.47
## + So      1      9329 1443739 505.63
## + LF      1      4374 1448694 505.79
## + NW      1      3799 1449269 505.81
## + Time    1      2293 1450775 505.86
## - U1      1    127044 1580112 505.87
## - Prob    1    247978 1701046 509.34
## - U2      1    255443 1708511 509.55
## - M       1    296790 1749858 510.67
## - Ed      1    445788 1898855 514.51
## - Ineq    1    738244 2191312 521.24
## - Po1     1   1672038 3125105 537.93
```

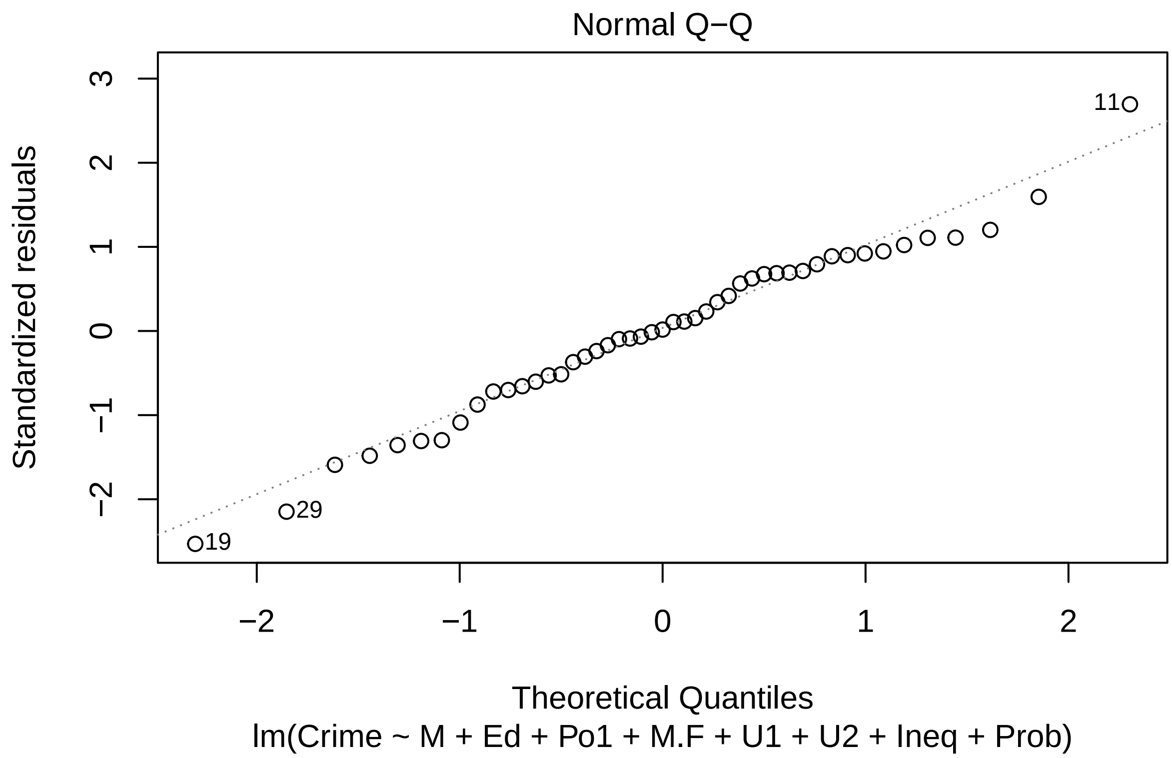
```r
summary(step_mod)
```

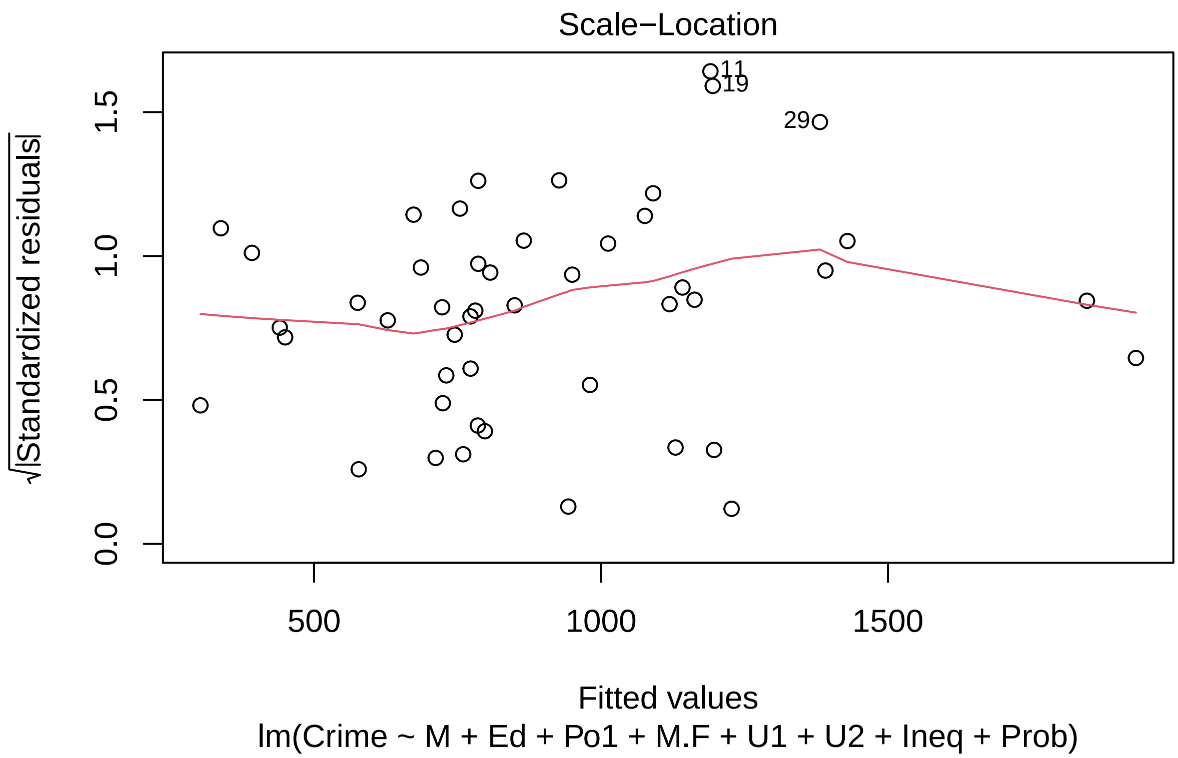
```
##
## Call:
## lm(formula = Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob,
##     data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -444.70 -111.07    3.03  122.15  483.30
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6426.10    1194.61  -5.379 4.04e-06 ***
## M              93.32      33.50   2.786  0.00828 **
## Ed            180.12      52.75   3.414  0.00153 **
## Po1           102.65      15.52   6.613 8.26e-08 ***
## M.F            22.34      13.60   1.642  0.10874
## U1          -6086.63    3339.27  -1.823  0.07622 .
```

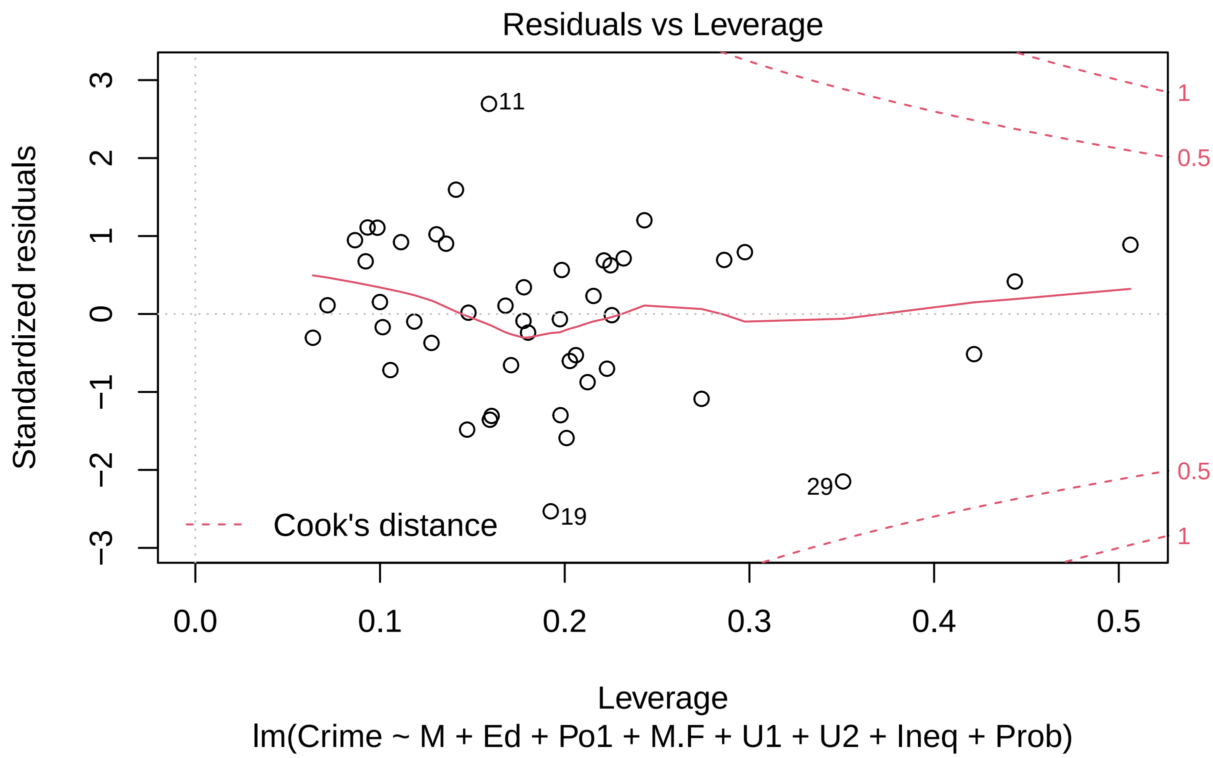
```
## U2            187.35      72.48   2.585  0.01371 *
## Ineq           61.33      13.96   4.394 8.63e-05 ***
## Prob        -3796.03    1490.65  -2.547  0.01505 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 195.5 on 38 degrees of freedom
## Multiple R-squared:  0.7888, Adjusted R-squared:  0.7444
## F-statistic: 17.74 on 8 and 38 DF,  p-value: 1.159e-10
```

```r
plot(step_mod)
```




Normal Q−Q
Standardized residuals
−2
−1
0
1
2
3
11
29
19
Theoretical Quantiles
lm(Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
500
1000
1500
11
19
29
Fitted values
lm(Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob)




```r
cat("The Akaic score for Stepwise is ", AIC(step_mod))
```

```
## The Akaic score for Stepwise is  639.3151
```

```r
########################################################################################################
#Backward Elimination

back_mod = step(mod, direction="backward")
```

```
## Start:  AIC=514.65
## Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 +
##     U2 + Wealth + Ineq + Prob + Time
##
##          Df Sum of Sq     RSS    AIC
## - So      1        29 1354974 512.65
## - LF      1      8917 1363862 512.96
## - Time    1     10304 1365250 513.00
## - Pop     1     14122 1369068 513.14
## - NW      1     18395 1373341 513.28
## - M.F     1     31967 1386913 513.74
## - Wealth  1     37613 1392558 513.94
## - Po2     1     37919 1392865 513.95
## <none>                1354946 514.65
## - U1      1     83722 1438668 515.47
## - Po1     1    144306 1499252 517.41
```

```
## - U2      1    181536 1536482 518.56
## - M       1    193770 1548716 518.93
## - Prob    1    199538 1554484 519.11
## - Ed      1    402117 1757063 524.86
## - Ineq    1    423031 1777977 525.42
##
## Step:  AIC=512.65
## Crime ~ M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 +
##     Wealth + Ineq + Prob + Time
##
##          Df Sum of Sq     RSS    AIC
## - Time    1     10341 1365315 511.01
## - LF      1     10878 1365852 511.03
## - Pop     1     14127 1369101 511.14
## - NW      1     21626 1376600 511.39
## - M.F     1     32449 1387423 511.76
## - Po2     1     37954 1392929 511.95
## - Wealth  1     39223 1394197 511.99
## <none>                1354974 512.65
## - U1      1     96420 1451395 513.88
## - Po1     1    144302 1499277 515.41
## - U2      1    189859 1544834 516.81
## - M       1    195084 1550059 516.97
## - Prob    1    204463 1559437 517.26
## - Ed      1    403140 1758114 522.89
## - Ineq    1    488834 1843808 525.13
##
## Step:  AIC=511.01
## Crime ~ M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 +
##     Wealth + Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## - LF      1     10533 1375848 509.37
## - NW      1     15482 1380797 509.54
## - Pop     1     21846 1387161 509.75
## - Po2     1     28932 1394247 509.99
## - Wealth  1     36070 1401385 510.23
## - M.F     1     41784 1407099 510.42
## <none>                1365315 511.01
## - U1      1     91420 1456735 512.05
## - Po1     1    134137 1499452 513.41
## - U2      1    184143 1549458 514.95
## - M       1    186110 1551425 515.01
## - Prob    1    237493 1602808 516.54
## - Ed      1    409448 1774763 521.33
## - Ineq    1    502909 1868224 523.75
##
## Step:  AIC=509.37
## Crime ~ M + Ed + Po1 + Po2 + M.F + Pop + NW + U1 + U2 + Wealth +
##     Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## - NW      1     11675 1387523 507.77
## - Po2     1     21418 1397266 508.09
```

```
## - Pop     1     27803 1403651 508.31
## - M.F     1     31252 1407100 508.42
## - Wealth  1     35035 1410883 508.55
## <none>                1375848 509.37
## - U1      1     80954 1456802 510.06
## - Po1     1    123896 1499744 511.42
## - U2      1    190746 1566594 513.47
## - M       1    217716 1593564 514.27
## - Prob    1    226971 1602819 514.54
## - Ed      1    413254 1789103 519.71
## - Ineq    1    500944 1876792 521.96
##
## Step:  AIC=507.77
## Crime ~ M + Ed + Po1 + Po2 + M.F + Pop + U1 + U2 + Wealth + Ineq +
##     Prob
##
##          Df Sum of Sq     RSS    AIC
## - Po2     1     16706 1404229 506.33
## - Pop     1     25793 1413315 506.63
## - M.F     1     26785 1414308 506.66
## - Wealth  1     31551 1419073 506.82
## <none>                1387523 507.77
## - U1      1     83881 1471404 508.52
## - Po1     1    118348 1505871 509.61
## - U2      1    201453 1588976 512.14
## - Prob    1    216760 1604282 512.59
## - M       1    309214 1696737 515.22
## - Ed      1    402754 1790276 517.74
## - Ineq    1    589736 1977259 522.41
##
## Step:  AIC=506.33
## Crime ~ M + Ed + Po1 + M.F + Pop + U1 + U2 + Wealth + Ineq +
##     Prob
##
##          Df Sum of Sq     RSS    AIC
## - Pop     1     22345 1426575 505.07
## - Wealth  1     32142 1436371 505.39
## - M.F     1     36808 1441037 505.54
## <none>                1404229 506.33
## - U1      1     86373 1490602 507.13
## - U2      1    205814 1610043 510.76
## - Prob    1    218607 1622836 511.13
## - M       1    307001 1711230 513.62
## - Ed      1    389502 1793731 515.83
## - Ineq    1    608627 2012856 521.25
## - Po1     1   1050202 2454432 530.57
##
## Step:  AIC=505.07
## Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Wealth + Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## - Wealth  1     26493 1453068 503.93
## <none>                1426575 505.07
## - M.F     1     84491 1511065 505.77
```

```
## - U1      1     99463 1526037 506.24
## - Prob    1    198571 1625145 509.20
## - U2      1    208880 1635455 509.49
## - M       1    320926 1747501 512.61
## - Ed      1    386773 1813348 514.35
## - Ineq    1    594779 2021354 519.45
## - Po1     1   1127277 2553852 530.44
##
## Step:  AIC=503.93
## Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob
##
##        Df Sum of Sq     RSS    AIC
## <none>              1453068 503.93
## - M.F   1    103159 1556227 505.16
## - U1    1    127044 1580112 505.87
## - Prob  1    247978 1701046 509.34
## - U2    1    255443 1708511 509.55
## - M     1    296790 1749858 510.67
## - Ed    1    445788 1898855 514.51
## - Ineq  1    738244 2191312 521.24
## - Po1   1   1672038 3125105 537.93
```

```r
summary(back_mod)
```

```
##
## Call:
## lm(formula = Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob,
##     data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -444.70 -111.07    3.03  122.15  483.30
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6426.10    1194.61  -5.379 4.04e-06 ***
## M              93.32      33.50   2.786  0.00828 **
## Ed            180.12      52.75   3.414  0.00153 **
## Po1           102.65      15.52   6.613 8.26e-08 ***
## M.F            22.34      13.60   1.642  0.10874
## U1          -6086.63    3339.27  -1.823  0.07622 .
## U2            187.35      72.48   2.585  0.01371 *
## Ineq           61.33      13.96   4.394 8.63e-05 ***
## Prob        -3796.03    1490.65  -2.547  0.01505 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 195.5 on 38 degrees of freedom
## Multiple R-squared:  0.7888, Adjusted R-squared:  0.7444
## F-statistic: 17.74 on 8 and 38 DF,  p-value: 1.159e-10
```

```r
cat("The Akaic score for Backward elimination is ", AIC(back_mod))
```

```
## The Akaic score for Backward elimination is  639.3151
```

```r
#######################################################################################################
mod2 = lm(Crime ~ 1, data=crime_data)
f = formula(lm(Crime ~., data=crime_data))
f_mod = step(mod2, direction = "forward", scope = f)
```

```
## Start:  AIC=561.02
## Crime ~ 1
##
##          Df Sum of Sq     RSS    AIC
## + Po1     1   3253302 3627626 532.94
## + Po2     1   3058626 3822302 535.39
## + Wealth  1   1340152 5540775 552.84
## + Prob    1   1257075 5623853 553.54
## + Pop     1    783660 6097267 557.34
## + Ed      1    717146 6163781 557.85
## + M.F     1    314867 6566061 560.82
## <none>                6880928 561.02
## + LF      1    245446 6635482 561.32
## + Ineq    1    220530 6660397 561.49
## + U2      1    216354 6664573 561.52
## + Time    1    154545 6726383 561.96
## + So      1     56527 6824400 562.64
## + M       1     55084 6825844 562.65
## + U1      1     17533 6863395 562.90
## + NW      1      7312 6873615 562.97
##
## Step:  AIC=532.94
## Crime ~ Po1
##
##          Df Sum of Sq     RSS    AIC
## + Ineq    1    739819 2887807 524.22
## + M       1    616741 3010885 526.18
## + M.F     1    250522 3377104 531.57
## + NW      1    232434 3395192 531.82
## + So      1    219098 3408528 532.01
## + Wealth  1    180872 3446754 532.53
## <none>                3627626 532.94
## + Po2     1    146167 3481459 533.00
## + Prob    1     92278 3535348 533.72
## + LF      1     77479 3550147 533.92
## + Time    1     43185 3584441 534.37
## + U2      1     17848 3609778 534.70
## + Pop     1      5666 3621959 534.86
## + U1      1      2878 3624748 534.90
## + Ed      1       767 3626859 534.93
##
## Step:  AIC=524.22
## Crime ~ Po1 + Ineq
##
##          Df Sum of Sq     RSS    AIC
## + Ed      1    587050 2300757 515.53
## + M.F     1    454545 2433262 518.17
## + Prob    1    280690 2607117 521.41
```

```
## + LF      1    260571 2627236 521.77
## + Wealth  1    213937 2673871 522.60
## + M       1    181236 2706571 523.17
## + Pop     1    130377 2757430 524.04
## <none>                2887807 524.22
## + NW      1     36439 2851369 525.62
## + So      1     33738 2854069 525.66
## + Po2     1     30673 2857134 525.71
## + U1      1      2309 2885498 526.18
## + Time    1       497 2887310 526.21
## + U2      1       253 2887554 526.21
##
## Step:  AIC=515.53
## Crime ~ Po1 + Ineq + Ed
##
##          Df Sum of Sq     RSS    AIC
## + M       1    239405 2061353 512.37
## + Prob    1    234981 2065776 512.47
## + M.F     1    117026 2183731 515.08
## <none>                2300757 515.53
## + Wealth  1     79540 2221218 515.88
## + U2      1     62112 2238646 516.25
## + Time    1     61770 2238987 516.26
## + Po2     1     42584 2258174 516.66
## + Pop     1     39319 2261438 516.72
## + U1      1      7365 2293392 517.38
## + LF      1      7254 2293503 517.39
## + NW      1      4210 2296547 517.45
## + So      1      4135 2296622 517.45
##
## Step:  AIC=512.37
## Crime ~ Po1 + Ineq + Ed + M
##
##          Df Sum of Sq     RSS    AIC
## + Prob    1    258063 1803290 508.08
## + U2      1    200988 1860365 509.55
## + Wealth  1    163378 1897975 510.49
## <none>                2061353 512.37
## + M.F     1     74398 1986955 512.64
## + U1      1     50835 2010518 513.20
## + Po2     1     45392 2015961 513.32
## + Time    1     42746 2018607 513.39
## + NW      1     16488 2044865 513.99
## + Pop     1      8101 2053251 514.19
## + So      1      3189 2058164 514.30
## + LF      1      2988 2058365 514.30
##
## Step:  AIC=508.08
## Crime ~ Po1 + Ineq + Ed + M + Prob
##
##          Df Sum of Sq     RSS    AIC
## + U2      1    192233 1611057 504.79
## + Wealth  1     86490 1716801 507.77
## + M.F     1     84509 1718781 507.83
```

```
## <none>                1803290 508.08
## + U1       1     52313 1750977 508.70
## + Pop      1     47719 1755571 508.82
## + Po2      1     37967 1765323 509.08
## + So       1     21971 1781320 509.51
## + Time     1     10194 1793096 509.82
## + LF       1       990 1802301 510.06
## + NW       1       797 1802493 510.06
##
## Step:  AIC=504.79
## Crime ~ Po1 + Ineq + Ed + M + Prob + U2
##
##          Df Sum of Sq     RSS    AIC
## <none>                1611057 504.79
## + Wealth  1     59910 1551147 505.00
## + U1      1     54830 1556227 505.16
## + Pop     1     51320 1559737 505.26
## + M.F     1     30945 1580112 505.87
## + Po2     1     25017 1586040 506.05
## + So      1     17958 1593098 506.26
## + LF      1     13179 1597878 506.40
## + Time    1      7159 1603898 506.58
## + NW      1       359 1610698 506.78
```

```r
summary(f_mod)
```

```
##
## Call:
## lm(formula = Crime ~ Po1 + Ineq + Ed + M + Prob + U2, data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -470.68  -78.41  -19.68  133.12  556.23
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5040.50     899.84  -5.602 1.72e-06 ***
## Po1           115.02      13.75   8.363 2.56e-10 ***
## Ineq           67.65      13.94   4.855 1.88e-05 ***
## Ed            196.47      44.75   4.390 8.07e-05 ***
## M             105.02      33.30   3.154  0.00305 **
## Prob        -3801.84    1528.10  -2.488  0.01711 *
## U2             89.37      40.91   2.185  0.03483 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 200.7 on 40 degrees of freedom
## Multiple R-squared:  0.7659, Adjusted R-squared:  0.7307
## F-statistic: 21.81 on 6 and 40 DF,  p-value: 3.418e-11
```

```r
cat("The Akaic score for Forward selection is ", AIC(f_mod))
```

```
## The Akaic score for Forward selection is  640.1661
```

```
################################################################################################
```

```r
library(glmnet)
```

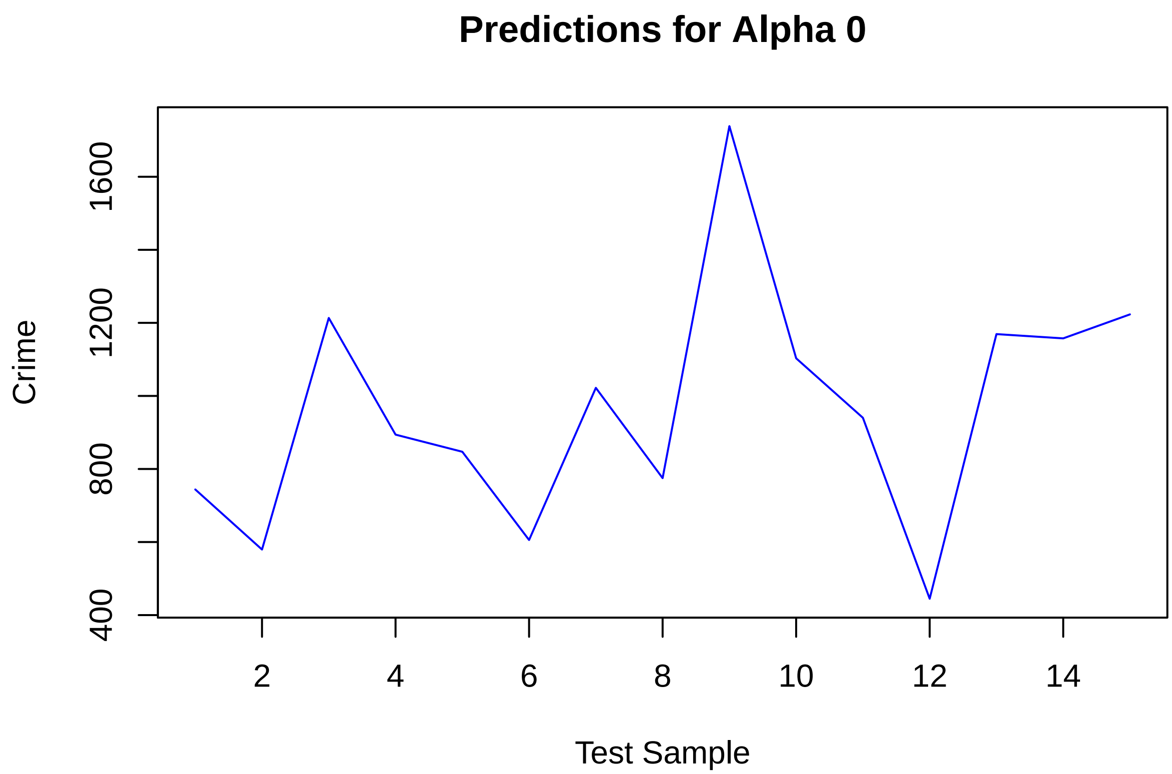
```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.0-2
```

```r
library(Metrics)
library(scales)
set.seed(100)
samples = sort(sample(nrow(crime_data), nrow(crime_data)*0.70))
train_data = crime_data[samples,]
test_data = crime_data[-samples,]

scaled_train_data = scale(train_data, center = TRUE, scale = TRUE)
scaled_test_data = scale(test_data, center = TRUE, scale = TRUE)

merr = c()
alp = c()
for (i in seq(0.0, 1.0, by = 0.1))
  {
    mod = cv.glmnet(as.matrix(scaled_train_data[,1:15]), train_data[,16], type.measure = "mse", alpha =
    print(mod)
    pred = predict(mod, s=mod$lambda.1se, newx = scaled_test_data[,1:15])
    cat("\n")
    print(pred)

    plot(pred, type="l", col="blue", xlab = "Test Sample", ylab = "Crime", main = paste("Predictions fo
    #mean((pred - scaled_test_data[,16])^2)
    if (i == 0.0)
      var = "Ridge Regression"
    else if (i == 1)
      var = "Lasso Regression"
    else
      var = "Elastic Net Regression"
    res = mse(pred,scaled_test_data[,16])
    cat("The MSE for ", var , "with an alpha value of ", i, " is ", comma_format() (res), "\n")
    merr = c(merr,mse(pred,scaled_test_data[,16]))
    alp = c(alp, i)
}
```

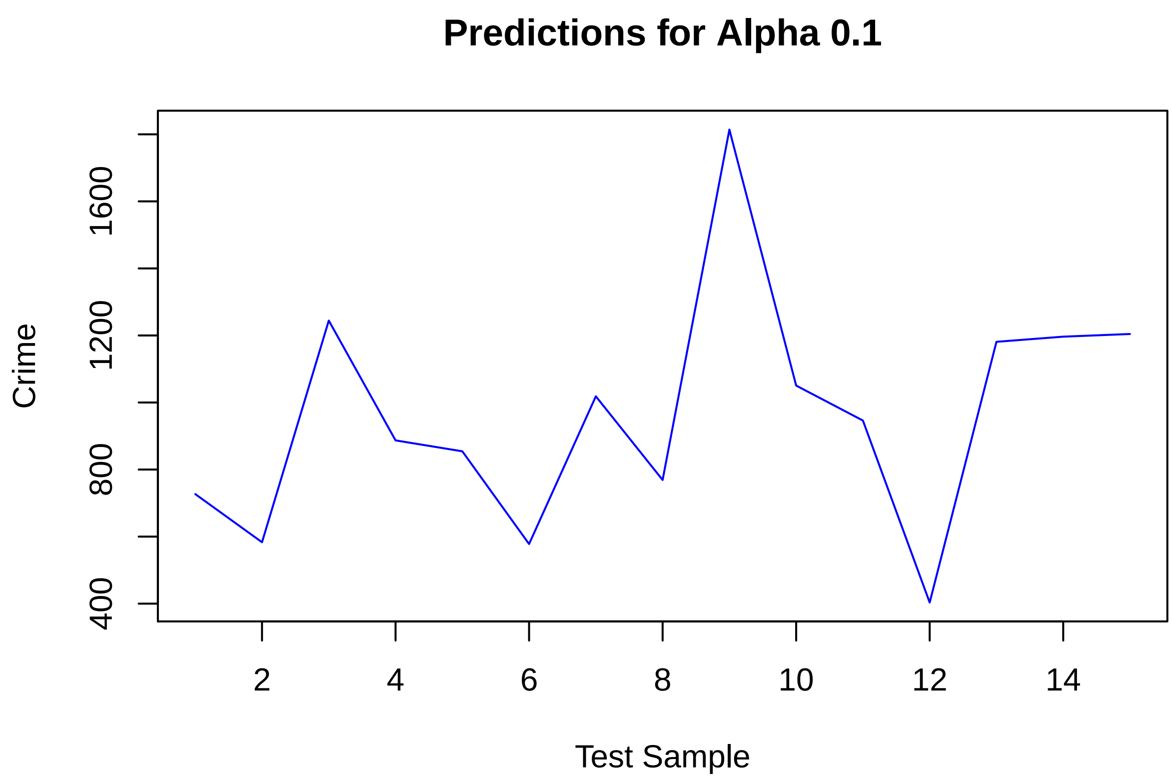
```
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
## Measure: Mean-Squared Error
##
##     Lambda Measure    SE Nonzero
## min  43.33   59852 10935      15
## 1se 191.97   70032  9901      15
##
##             1
```

```
## 1   743.9640
## 3   579.4643
## 5  1213.4217
## 9   894.0522
## 15  847.1115
## 17  605.5916
## 24 1022.1599
## 27  774.9623
## 29 1738.5602
## 32 1102.9555
## 33  940.1110
## 42  444.7918
## 43 1169.2410
## 46 1157.5618
## 47 1223.2386
```



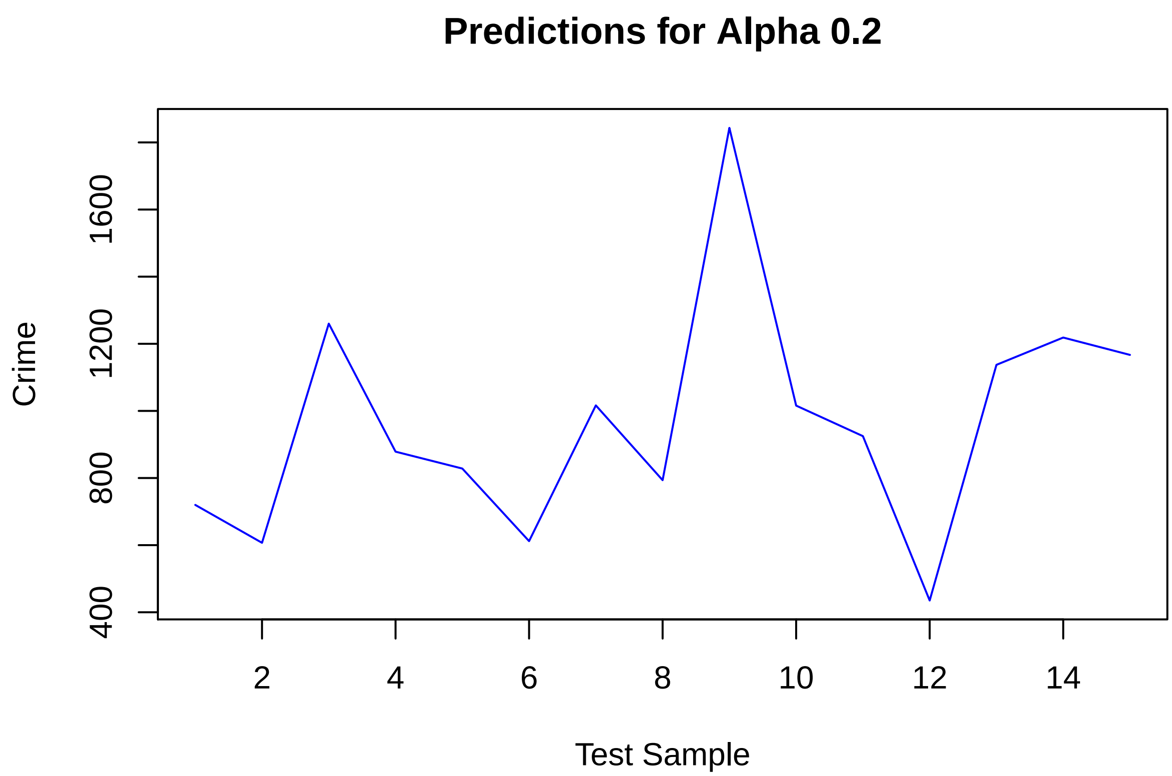

```
## The MSE for  Ridge Regression with an alpha value of  0  is  1,027,300
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
## Measure: Mean-Squared Error
##
##     Lambda Measure    SE Nonzero
## min  28.51   63313 12635      14
## 1se 104.86   75293 15559      15
##
```

```
##             1
## 1   726.8674
## 3   583.1571
## 5  1244.3453
## 9   886.9528
## 15  854.4632
## 17  577.8360
## 24 1018.4628
## 27  768.9234
## 29 1814.0506
## 32 1050.5610
## 33  946.2756
## 42  403.4494
## 43 1181.0857
## 46 1196.4196
## 47 1204.3376
```



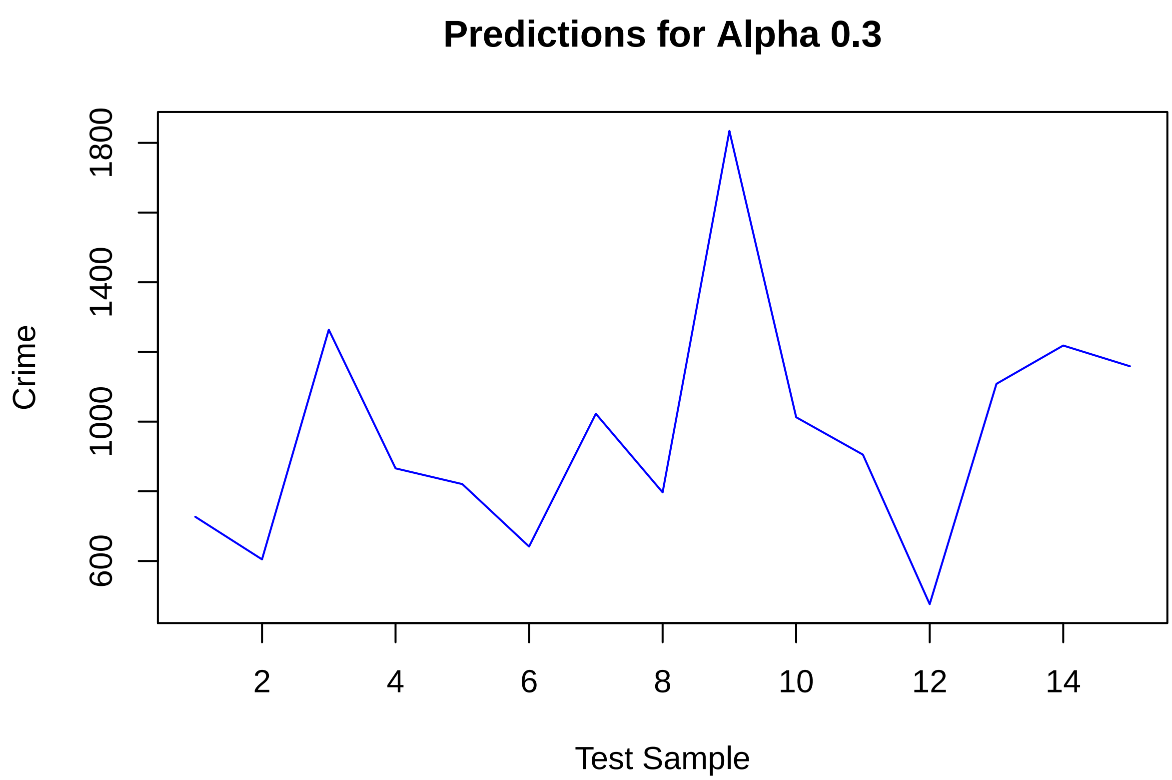

```
## The MSE for  Elastic Net Regression with an alpha value of  0.1  is  1,041,319
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
## Measure: Mean-Squared Error
##
##     Lambda Measure    SE Nonzero
## min  27.34   56823 13867      14
## 1se  91.62   69250 18308      12
```

```
##
##              1
## 1   719.8774
## 3   606.9597
## 5  1260.0343
## 9   878.5675
## 15  828.1789
## 17  611.9270
## 24 1016.4200
## 27  793.5715
## 29 1843.2693
## 32 1015.7518
## 33  924.8657
## 42  435.0643
## 43 1137.2366
## 46 1218.5443
## 47 1166.9190
```



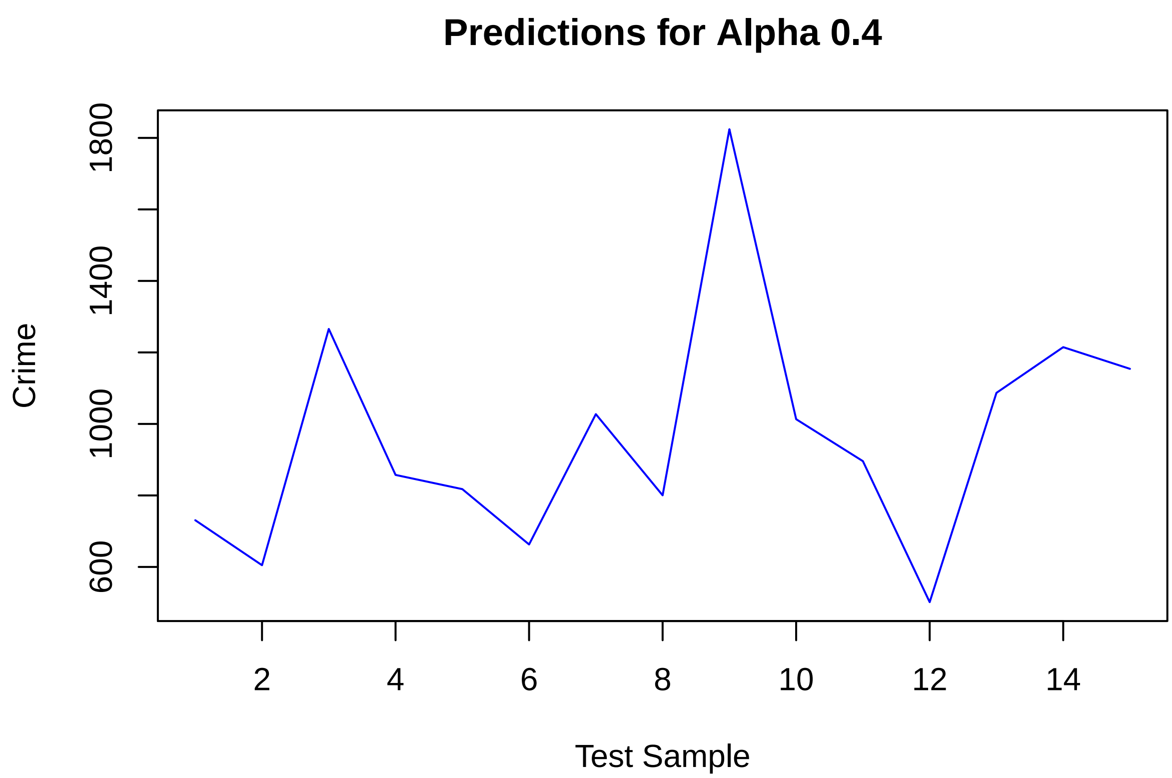

```
## The MSE for  Elastic Net Regression with an alpha value of  0.2  is  1,038,526
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
## Measure: Mean-Squared Error
##
##     Lambda Measure    SE Nonzero
## min  26.44   56092 12146      14
```

```
## 1se  80.75   66511 15748      10
##
##              1
## 1   726.9247
## 3   604.6501
## 5  1263.8631
## 9   865.8851
## 15  820.6747
## 17  641.5095
## 24 1022.6656
## 27  797.0425
## 29 1834.2355
## 32 1012.6141
## 33  905.3309
## 42  476.1613
## 43 1108.3991
## 46 1218.3799
## 47 1158.8514
```



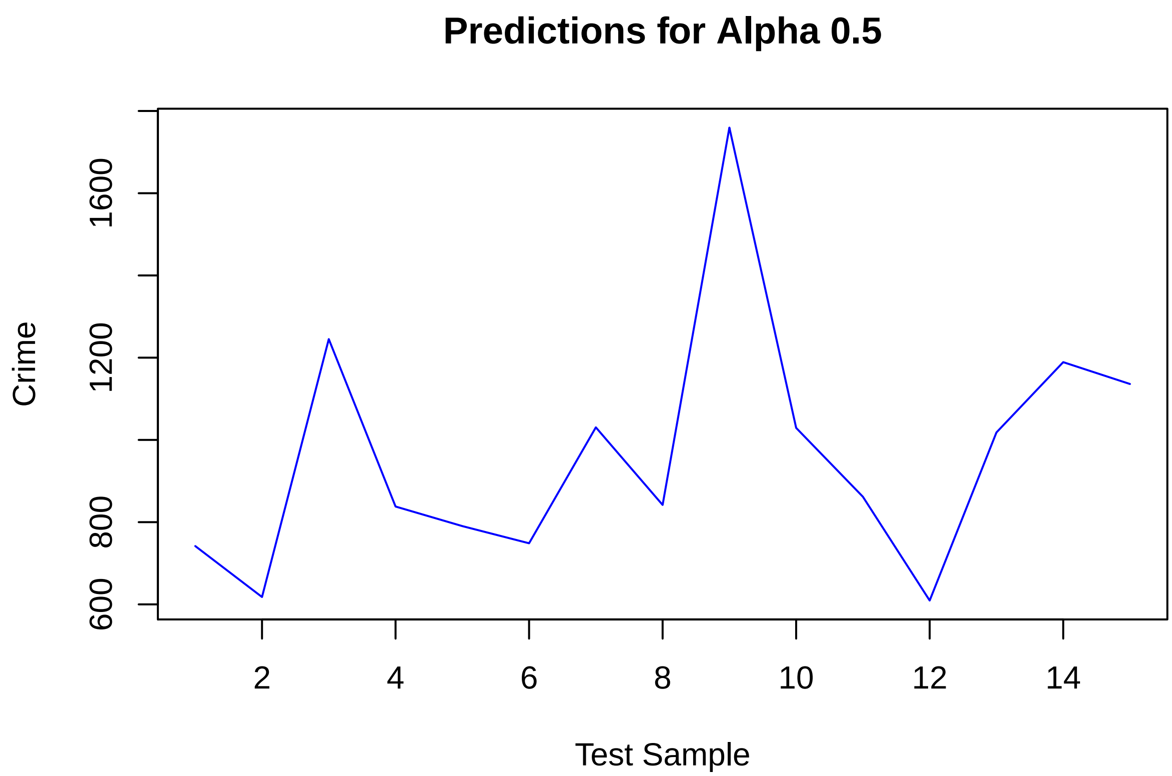

```
## The MSE for  Elastic Net Regression with an alpha value of  0.3  is  1,032,946
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
## Measure: Mean-Squared Error
##
##     Lambda Measure    SE Nonzero
```

```
## min  19.83   62451 15629      14
## 1se  72.94   75316 20839       9
##
##             1
## 1   730.6884
## 3   604.8393
## 5  1265.6640
## 9   857.3463
## 15  817.6397
## 17  662.8301
## 24 1027.3817
## 27  800.3439
## 29 1824.2146
## 32 1013.1714
## 33  895.8977
## 42  501.5530
## 43 1086.8524
## 46 1214.7146
## 47 1154.0505
```



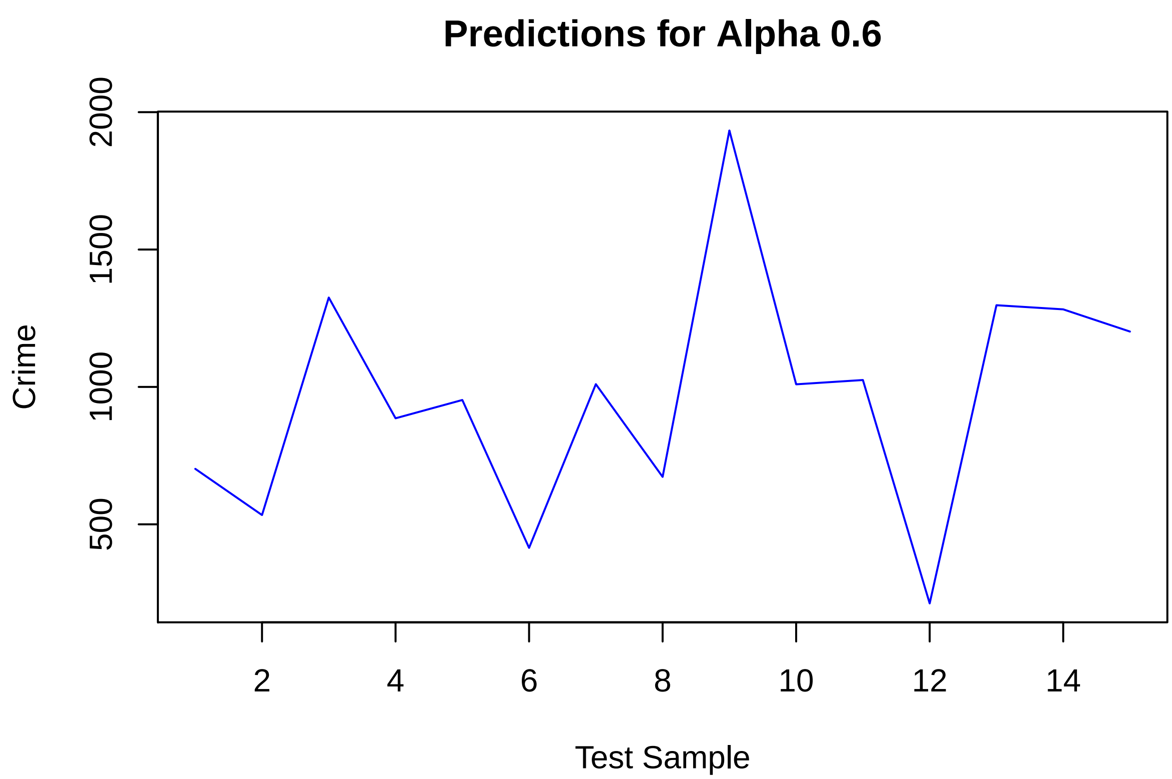

```
## The MSE for  Elastic Net Regression with an alpha value of  0.4  is  1,028,843
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
## Measure: Mean-Squared Error
##
```

```
##     Lambda Measure    SE Nonzero
## min  30.43   78144 28445      13
## 1se  84.66  103105 33451       8
##
##            1
## 1   741.8200
## 3   617.9026
## 5  1245.1406
## 9   837.9608
## 15  790.3749
## 17  748.5887
## 24 1030.5443
## 27  841.8705
## 29 1759.5124
## 32 1029.1626
## 33  861.7128
## 42  609.3737
## 43 1018.1452
## 46 1189.1015
## 47 1135.9769
```



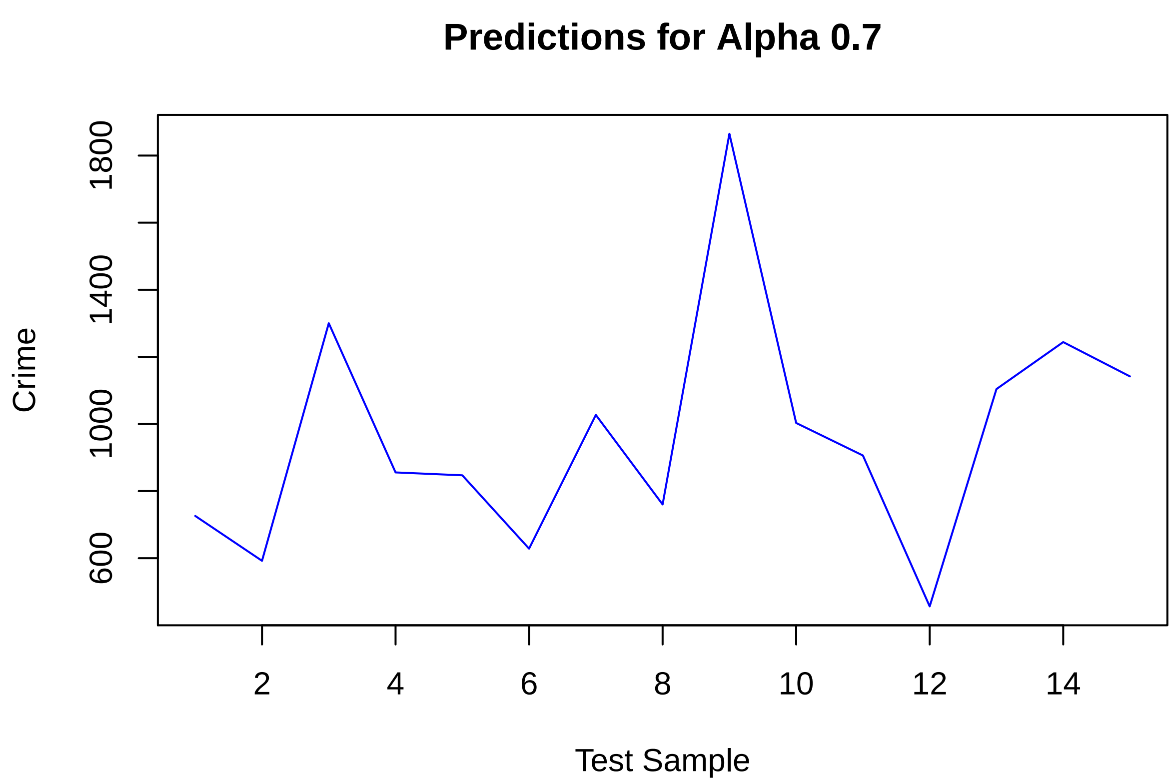

```
## The MSE for  Elastic Net Regression with an alpha value of  0.5  is  1,009,750
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
## Measure: Mean-Squared Error
```

```
##
##     Lambda Measure    SE Nonzero
## min  8.303   60520 10231      14
## 1se 17.477   70121 12954      14
##
##            1
## 1   701.8321
## 3   533.7197
## 5  1325.3125
## 9   885.8346
## 15  952.4455
## 17  414.1668
## 24 1010.0402
## 27  672.6724
## 29 1933.1842
## 32 1009.5447
## 33 1025.1164
## 42  212.1784
## 43 1297.3274
## 46 1282.3128
## 47 1201.5001
```



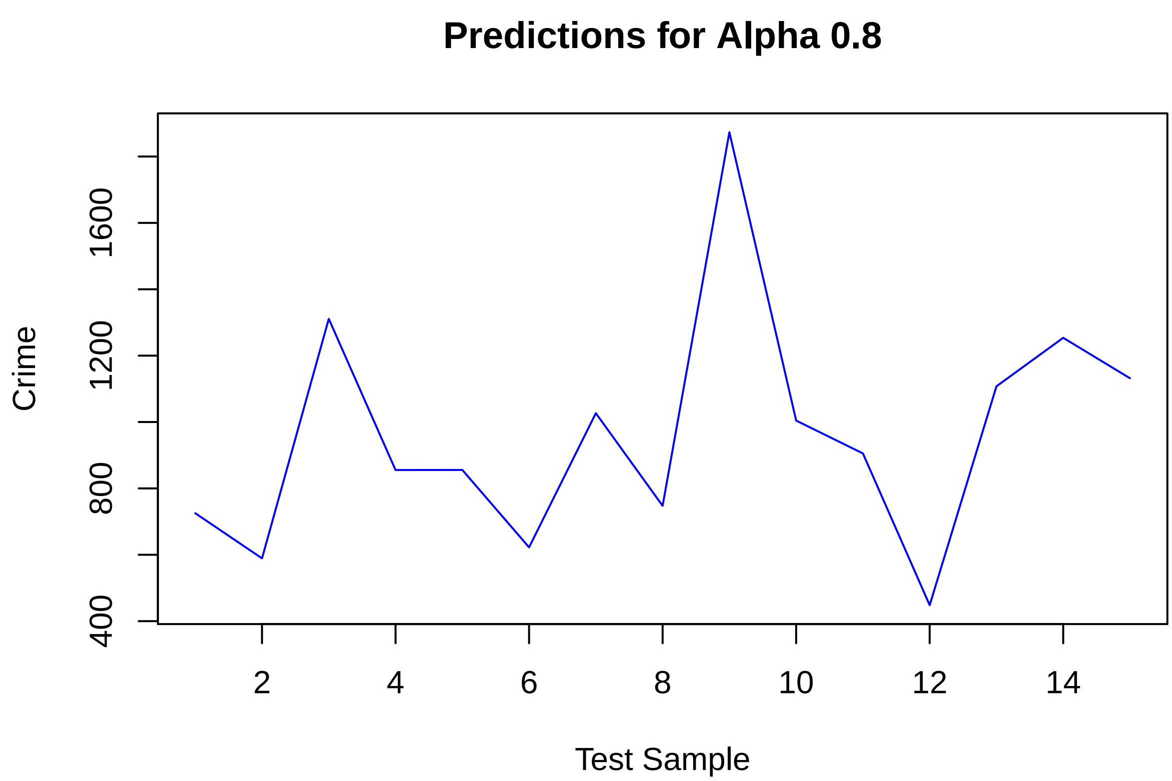

```
## The MSE for  Elastic Net Regression with an alpha value of  0.6  is  1,099,051
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
```

```
## Measure: Mean-Squared Error
##
##     Lambda Measure    SE Nonzero
## min  11.33   59128 12731      13
## 1se  41.68   70015 19270       9
##
##             1
## 1   725.9085
## 3   592.2676
## 5  1300.3265
## 9   855.6134
## 15  847.0497
## 17  628.5165
## 24 1027.0054
## 27  760.1304
## 29 1864.8258
## 32 1003.0575
## 33  906.1292
## 42  456.4044
## 43 1104.0880
## 46 1244.0614
## 47 1141.8031
```



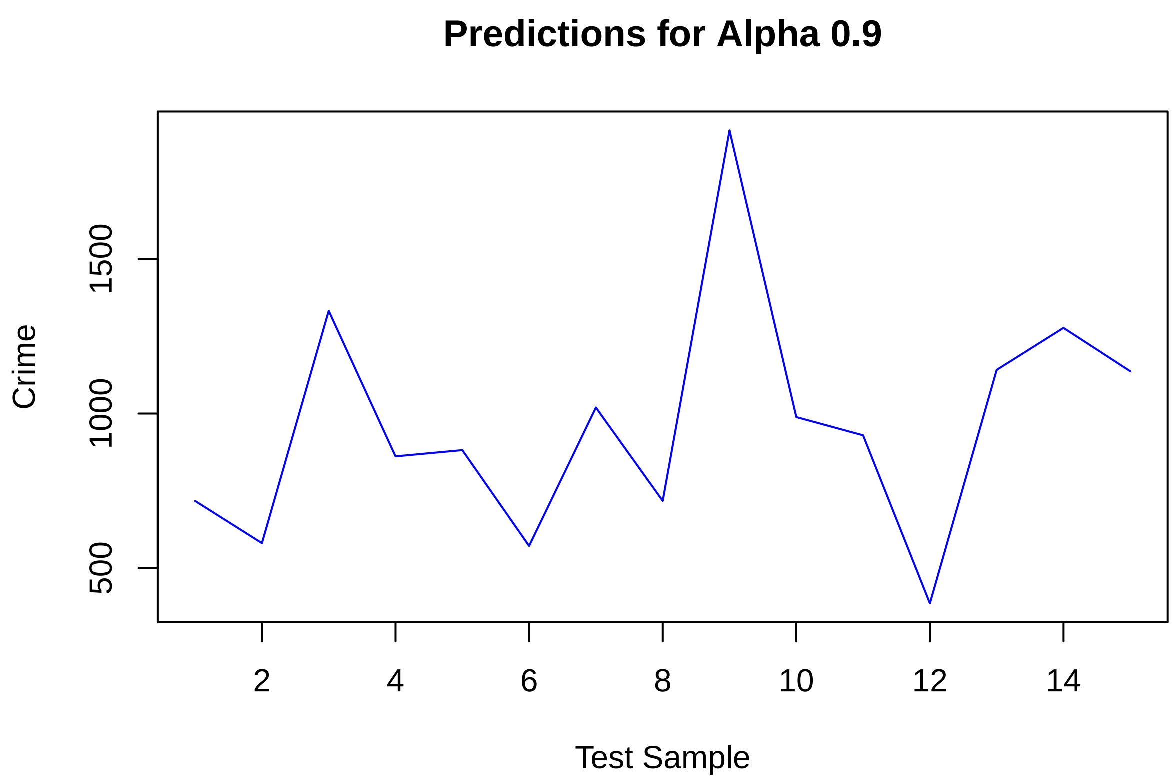

```
## The MSE for  Elastic Net Regression with an alpha value of  0.7  is  1,041,576
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
```

```
##
## Measure: Mean-Squared Error
##
##     Lambda Measure    SE Nonzero
## min   9.92   56745 13300      13
## 1se  36.47   68806 16143       8
##
##              1
## 1   725.1977
## 3   589.2293
## 5  1310.9859
## 9   855.3102
## 15  855.6010
## 17  622.4434
## 24 1026.5768
## 27  747.5247
## 29 1873.0934
## 32 1004.3393
## 33  905.4526
## 42  448.1520
## 43 1107.5801
## 46 1253.7890
## 47 1131.9122
```



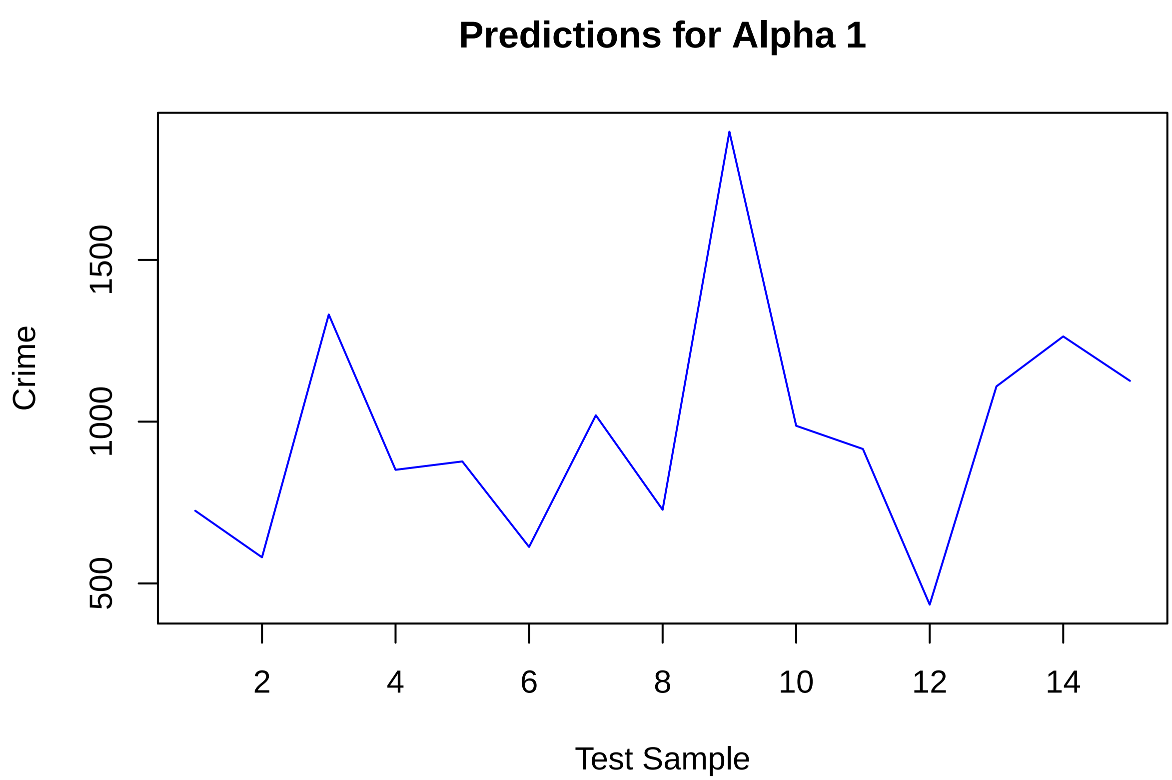

```
## The MSE for  Elastic Net Regression with an alpha value of  0.8  is  1,044,508
##
```

```
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
## Measure: Mean-Squared Error
##
##     Lambda Measure    SE Nonzero
## min  4.595   50717  8545      14
## 1se 26.915   57857 14217       8
##
##             1
## 1   717.1027
## 3   580.5862
## 5  1332.3597
## 9   861.4534
## 15  881.6668
## 17  571.6704
## 24 1019.4850
## 27  717.4605
## 29 1916.1107
## 32  988.5573
## 33  929.5331
## 42  385.8854
## 43 1141.2249
## 46 1277.2605
## 47 1136.8308
```




```
## The MSE for  Elastic Net Regression with an alpha value of  0.9  is  1,060,565
```

```
##
## Call:  cv.glmnet(x = as.matrix(scaled_train_data[, 1:15]), y = train_data[,      16], type.measure =
##
## Measure: Mean-Squared Error
##
##     Lambda Measure    SE Nonzero
## min  7.228   51989  8828      13
## 1se 29.178   59631 10900       8
##
##             1
## 1   724.6228
## 3   580.6821
## 5  1331.2752
## 9   851.1236
## 15  877.0752
## 17  612.7630
## 24 1019.6291
## 27  727.6867
## 29 1896.2894
## 32  987.2477
## 33  915.5558
## 42  434.4813
## 43 1109.1227
## 46 1263.5131
## 47 1126.1197
```



```
## The MSE for  Lasso Regression with an alpha value of  1  is  1,050,621
```

```r
plot(alp, merr, xlab = "Alpha", ylab = "MSE", main="Error vs Alpha", type="l", col="red")
```