ISL Fall 2015 Assignment II. 100 pts.

NAME(s):

**LOKESHWAR REDDY INJA (16207239/ LIRK9)**
**SANTHOSH KUMAR GATTU (16211118/ SG6N6)**
**SANTOSH KUMAR (16207971/ SK7Z9)**

You may submit this assignment in groups of upto three each. Write your names on this sheet and include it as the cover page for your submission.
The objective of this assignment is to practice using R, and gain a fundamental understanding of linear regression. Your submission should include both your code as well your answers to the questions.
Electronic submission on Blackboard is due **latest by 11 pm on Wed, Sep 30th**.
You may upload upto **three** submissions **before** the deadline – only the last submission will be graded. Submissions received after the deadline will be graded only for effort for a maximum of 70% of the total grade (Refer to class syllabus for detailed grading policy).
**State any assumptions you make, justify your answers, show intermediate steps and explain your results for maximum credit.** All answers should be in your own words with any sources you refer to cited at the appropriate places. Any knowledge you acquire from the Internet should be written in your own words and be appropriately referenced. Copying and pasting from the Internet, each other or any other source will not count as your effort (Refer to class syllabus for detailed policy on plagiarism).
**Remember that answers need to be word-processed (NOT handwritten) and should use R.**

Answer the following questions from Chapter 3.

*Undergraduate Students:*
Questions 4, 8, 10

*Graduate Students:*
Questions 6, 10, 15

Question 6.

Using (3.4), argue that in the case of simple linear regression, the least squares line always passes through the point ($\bar{x}$, $\bar{y}$).

Solution:

We know that the Simple linear regression is used to predict the value of response Y on the basis of predictor X. It also assumes the relationship between X and Y to be approximately linear.

In mathematical terms we can represent Y as a function of X.

$Y=f(X) \cong \hat{\beta}_0+\hat{\beta}_1 X$ ----------------→ equation 1
Substitute ($\bar{X}$, $\bar{Y}$) in equation 1

$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ --------------------→equation 2

Now again substitute $\hat{\beta}_0$ value in equation 1

We get, $Y = \bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X$ ---→ equation 3

Now from the equation 3 it is clear that the point ($\bar{X}$, $\bar{Y}$) will always pass through it.

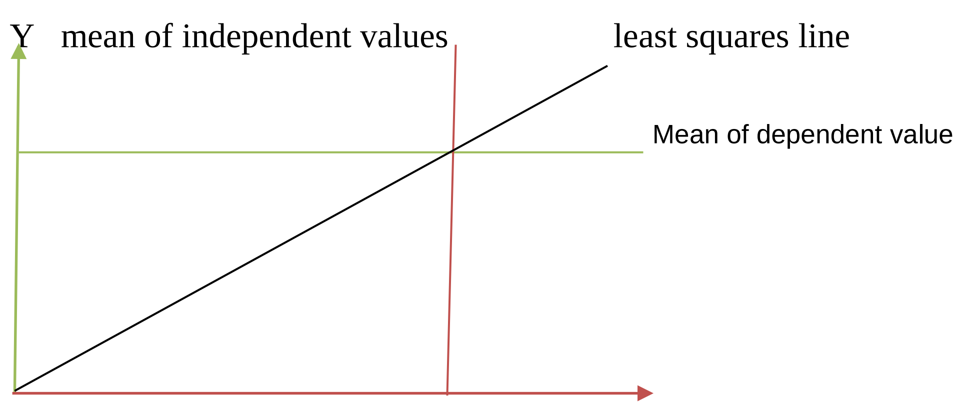
If we pass the point through the equation, we get

$\bar{Y} = \bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 \bar{X}$

$\bar{Y} = \bar{Y}$ -----------------------------→ equation 4

0=0

Hence, with the help of above argument we can argue that in the case of simple linear regression, the least squares line will always pass through the point ($\bar{X}$ , $\bar{Y}$).



X

Question 10.

This question should be answered using the Carseats data set.

(a) Fit a multiple regression model to predict Sales using Price, Urban, and US.

isl10a.R

(b)

Interpretation of each coefficient from the below data.

```
> summary(saleslm)

Call:
lm(formula = Sales ~ Price + Urban + US, data = Carseats)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.9206 -1.6220 -0.0564  1.5786  7.0581 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 13.043469   0.651012  20.036  < 2e-16 ***
Price       -0.054459   0.005242 -10.389  < 2e-16 ***
UrbanYes    -0.021916   0.271650  -0.081    0.936    
USYes        1.200573   0.259042   4.635 4.86e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.472 on 396 degrees of freedom
Multiple R-squared:  0.2393,	Adjusted R-squared:  0.2335 
F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

- Price: On an average when price increases by 1 dollar the sales decreases by 54.459 units when all other predictors remain fixed because the coefficient of the predictor price is -0.054459. The standard error is also less in the model.
- Urban: the unit sales in urban areas are 21.916 units less when compared to the unit sales in rural areas when all other predictors remain fixed. Urban predictor cannot be used as it did not reject the null hypothesis which is identified from the large ‘p-value’ (0.936)
- US : the unit sales in US region are 1200.573 units greater than the sales in non-US regions when all other predictors remain fixed.

**(C)**

Model in Equation form:

Sales = B0 + B1 * (price) + B2 * (Urban) + B3 * (US) + error

Where B0 = 13.043469, B1 = -0.054459    B2 = -0.021916    B3 = 1.200573

Equation for Urban and non US is:

Sales = B0 + B1 * price + B2 * Urban + error

Equation for rural and non US is:

Sales = B0 + B1 * price + error

Equation for Urban and US is:

Sales = B0 + B1 * price + B2 * Urban + B3 * US + error

Equation for rural and US is:

Sales = B0 + B1 * price + B3 * US + error

**(D)**

Price predictor rejects the Null hypothesis as p-value (2e-16) is very less.
US predictor rejects the Null hypothesis as p-value (2e-16) is very less.

**(E)**

Please find below attached the .R file

isl10e.R

**(F)**

Summary for the fit in (10 a) is

```
> summary(saleslm)

Call:
lm(formula = Sales ~ Price + Urban + US, data = Carseats)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.9206 -1.6220 -0.0564  1.5786  7.0581 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 13.043469   0.651012  20.036  < 2e-16 ***
Price       -0.054459   0.005242 -10.389  < 2e-16 ***
UrbanYes    -0.021916   0.271650  -0.081    0.936    
USYes        1.200573   0.259042   4.635 4.86e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.472 on 396 degrees of freedom
Multiple R-squared:  0.2393,   Adjusted R-squared:  0.2335 
F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

**Summary for (10.e) is**

```
> summary(saleslm1)

Call:
lm(formula = Sales ~ Price + US, data = Carseats)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.9269 -1.6286 -0.0574  1.5766  7.0515 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 13.03079    0.63098  20.652  < 2e-16 ***
Price       -0.05448    0.00523 -10.416  < 2e-16 ***
USYes        1.19964    0.25846   4.641 4.71e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.469 on 397 degrees of freedom
Multiple R-squared:  0.2393,   Adjusted R-squared:  0.2354 
F-statistic: 62.43 on 2 and 397 DF,  p-value: < 2.2e-16
```

Both the models fit the data very poorly as the R-squared value is 0.2393.
In both models 10.e model is better than 10.a model because RSE is less for 10.e model.

**(G)**

Please find below attached the .R file

isl10g.R

```
                  2.5 %      97.5 %
(Intercept) 11.79032020 14.27126531
Price       -0.06475984 -0.04419543
USYes        0.69151957  1.70776632
```

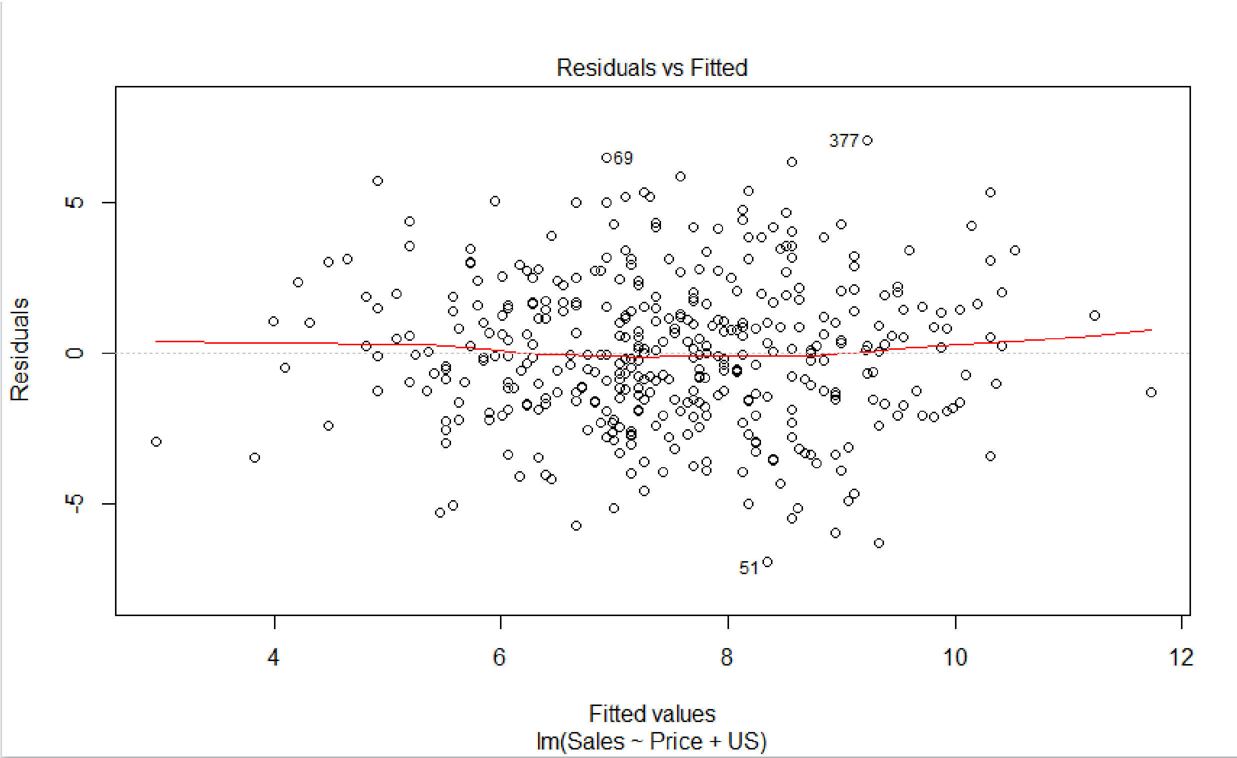
**(H)**

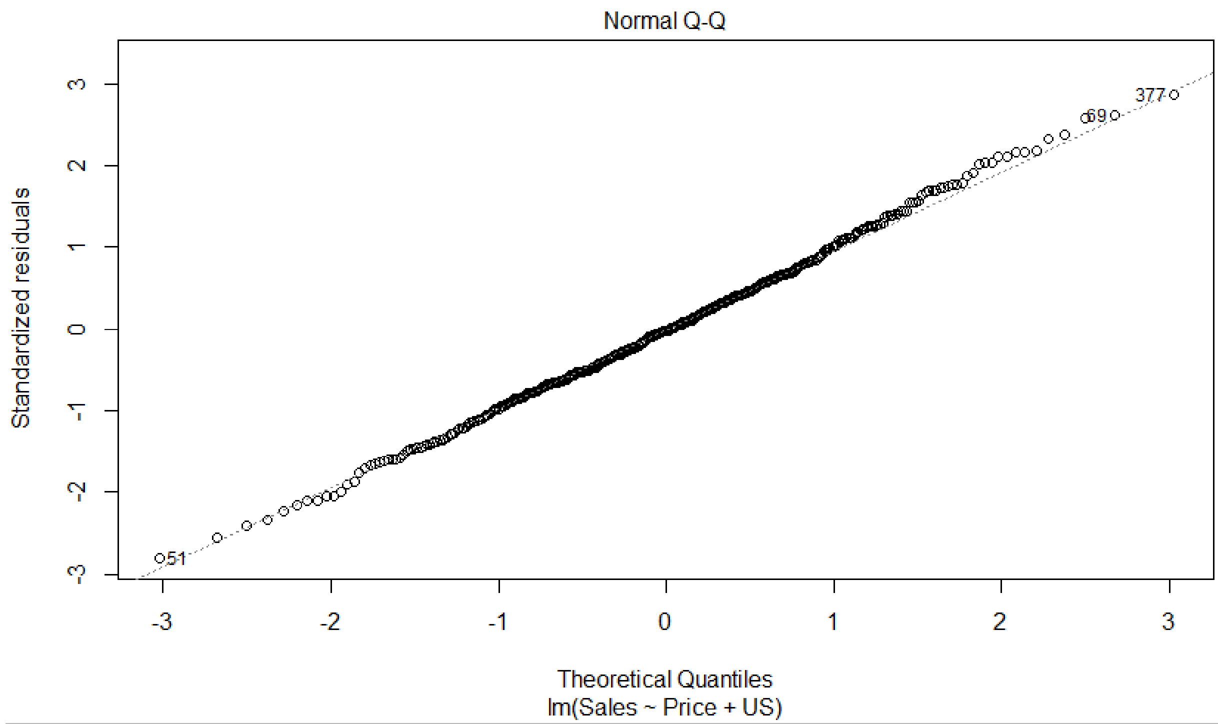
Please find below attached the .R file

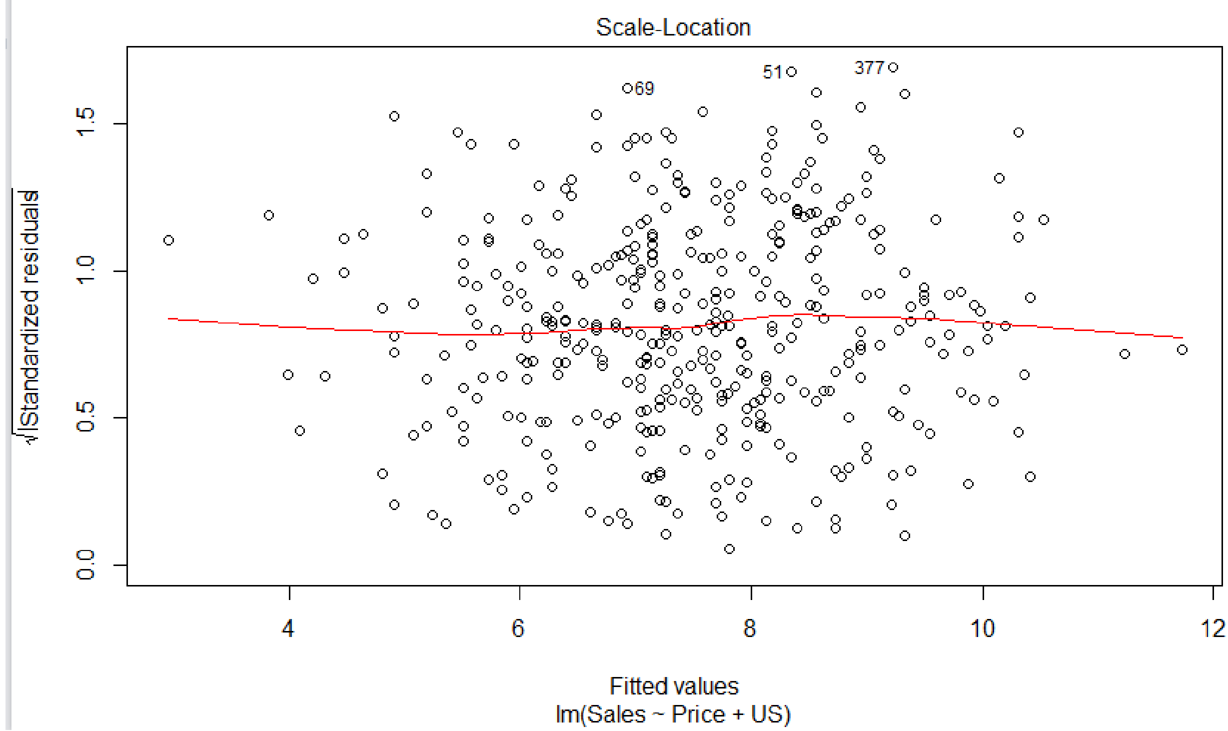
isl10h.R

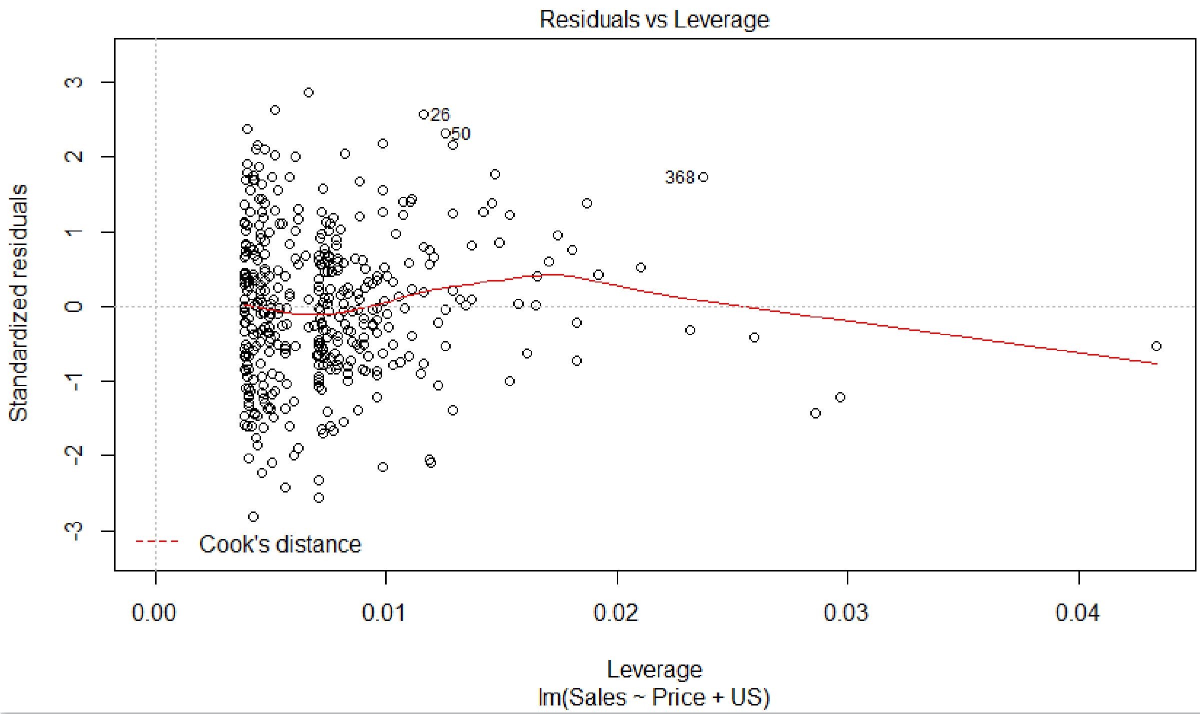
Standardized residuals versus leverage plots show the presence of a few outliers and some leverage points.

<u>The plots for the model 10.e:</u>


Residuals vs Fitted
Residuals
Fitted values
lm(Sales ~ Price + US)
377
69
51
5
0
-5
4
6
8
10
12


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Sales ~ Price + US)
51
69
377
-3
-2
-1
0
1
2
3


Scale-Location
√|Standardized residuals|
Fitted values
lm(Sales ~ Price + US)
69
51
377
0.0
0.5
1.0
1.5
4
6
8
10
12




**Question 15.**

(a)

```
Expect 'chas' all the other predictors have less p-value, so all predictors
are significant except chas because chas did not reject Null hypothesis.Also,
chas has very low R-squared (0.003124) value, so it is not a good predictor.

'Rad' and 'tax' are very good predictor among all predictors because they
have very high R-squared (for rad-0.3913 and for tax-0.3396) value.
```

Please find attached below the .R file.

isl15a.R

```
> lmplot(crim,zn)

Call:
lm(formula = y ~ x, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-4.429 -4.222 -2.620  1.250 84.523 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.45369    0.41722  10.675  < 2e-16 ***
x           -0.07393    0.01609  -4.594 5.51e-06 ***
---
```

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

Residual standard error: 8.435 on 504 degrees of freedom
Multiple R-squared:  0.04019,  Adjusted R-squared:  0.03828
F-statistic:  21.1 on 1 and 504 DF,  p-value: 5.506e-06
```

```
> lmplot(crim,indus)

Call:
lm(formula = y ~ x, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-11.972  -2.698  -0.736   0.712  81.813

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.06374    0.66723  -3.093  0.00209 **
x            0.50978    0.05102   9.991  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.866 on 504 degrees of freedom
Multiple R-squared:  0.1653,   Adjusted R-squared:  0.1637
F-statistic: 99.82 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
> lmplot(crim,chas)

Call:
lm(formula = y ~ x, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max
-3.738 -3.661 -3.435  0.018 85.232

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.7444     0.3961   9.453   <2e-16 ***
x            -1.8928     1.5061  -1.257    0.209
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.597 on 504 degrees of freedom
Multiple R-squared:  0.003124, Adjusted R-squared:  0.001146
F-statistic: 1.579 on 1 and 504 DF,  p-value: 0.2094
```

```
> lmplot(crim,nox)

Call:
```

```
lm(formula = y ~ x, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-12.371  -2.738  -0.974   0.559  81.728 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -13.720      1.699  -8.073 5.08e-15 ***
x             31.249      2.999  10.419  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.81 on 504 degrees of freedom
Multiple R-squared:  0.1772,	Adjusted R-squared:  0.1756 
F-statistic: 108.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
> lmplot(crim,rm)

Call:
lm(formula = y ~ x, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-6.604 -3.952 -2.654  0.989 87.197 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   20.482      3.365   6.088 2.27e-09 ***
x             -2.684      0.532  -5.045 6.35e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.401 on 504 degrees of freedom
Multiple R-squared:  0.04807,	Adjusted R-squared:  0.04618 
F-statistic: 25.45 on 1 and 504 DF,  p-value: 6.347e-07
```

```
> lmplot(crim,age)

Call:
```

```
lm(formula = y ~ x, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-6.789 -4.257 -1.230  1.527 82.849 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.77791    0.94398  -4.002 7.22e-05 ***
x            0.10779    0.01274   8.463 2.85e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.057 on 504 degrees of freedom
Multiple R-squared:  0.1244,	Adjusted R-squared:  0.1227 
F-statistic: 71.62 on 1 and 504 DF,  p-value: 2.855e-16
```

```
> lmplot(crim,dis)

Call:
lm(formula = y ~ x, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-6.708 -4.134 -1.527  1.516 81.674 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   9.4993     0.7304  13.006   <2e-16 ***
x            -1.5509     0.1683  -9.213   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.965 on 504 degrees of freedom
Multiple R-squared:  0.1441,	Adjusted R-squared:  0.1425 
F-statistic: 84.89 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
> lmplot(crim,rad)

Call:
```

```
lm(formula = y ~ x, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-10.164  -1.381  -0.141   0.660  76.433 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.28716    0.44348  -5.157 3.61e-07 ***
x            0.61791    0.03433  17.998  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.718 on 504 degrees of freedom
Multiple R-squared:  0.3913,	Adjusted R-squared:   0.39 
F-statistic: 323.9 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
> lmplot(crim,tax)

Call:
lm(formula = y ~ x, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-12.513  -2.738  -0.194   1.065  77.696 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -8.528369   0.815809  -10.45   <2e-16 ***
x            0.029742   0.001847   16.10   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.997 on 504 degrees of freedom
Multiple R-squared:  0.3396,	Adjusted R-squared:  0.3383 
F-statistic: 259.2 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
> lmplot(crim,ptratio)

Call:
lm(formula = y ~ x, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-7.654 -3.985 -1.912  1.825 83.353 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -17.6469     3.1473  -5.607 3.40e-08 ***
x             1.1520     0.1694   6.801 2.94e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.24 on 504 degrees of freedom
Multiple R-squared:  0.08407,	Adjusted R-squared:  0.08225 
F-statistic: 46.26 on 1 and 504 DF,  p-value: 2.943e-11
```

```
> lmplot(crim,black)

Call:
lm(formula = y ~ x, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-13.756  -2.299  -2.095  -1.296  86.822 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 16.553529   1.425903  11.609   <2e-16 ***
x           -0.036280   0.003873  -9.367   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.946 on 504 degrees of freedom
Multiple R-squared:  0.1483,	Adjusted R-squared:  0.1466 
F-statistic: 87.74 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
> lmplot(crim,lstat)
```

```
Call:
lm(formula = y ~ x, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-13.925  -2.822  -0.664   1.079  82.862 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.33054    0.69376  -4.801 2.09e-06 ***
x            0.54880    0.04776  11.491  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.664 on 504 degrees of freedom
Multiple R-squared:  0.2076,	Adjusted R-squared:  0.206 
F-statistic:   132 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
> lmplot(crim,medv)

Call:
lm(formula = y ~ x, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-9.071 -4.022 -2.343  1.298 80.957 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11.79654    0.93419   12.63   <2e-16 ***
x           -0.36316    0.03839   -9.46   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.934 on 504 degrees of freedom
Multiple R-squared:  0.1508,	Adjusted R-squared:  0.1491 
F-statistic: 89.49 on 1 and 504 DF,  p-value: < 2.2e-16
```

(15.b)

Please find attached below the .R file.

isl15b.R

```
> summary(lmplotall)

Call:
lm(formula = crim ~ ., data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-9.924 -2.120 -0.353  1.019 75.051 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  17.033228   7.234903   2.354 0.018949 *  
zn            0.044855   0.018734   2.394 0.017025 *  
indus        -0.063855   0.083407  -0.766 0.444294    
chas         -0.749134   1.180147  -0.635 0.525867    
nox         -10.313535   5.275536  -1.955 0.051152 .  
rm            0.430131   0.612830   0.702 0.483089    
age           0.001452   0.017925   0.081 0.935488    
dis          -0.987176   0.281817  -3.503 0.000502 ***
rad           0.588209   0.088049   6.680 6.46e-11 ***
tax          -0.003780   0.005156  -0.733 0.463793    
ptratio      -0.271081   0.186450  -1.454 0.146611    
black        -0.007538   0.003673  -2.052 0.040702 *  
lstat         0.126211   0.075725   1.667 0.096208 .  
medv         -0.198887   0.060516  -3.287 0.001087 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.439 on 492 degrees of freedom
Multiple R-squared:  0.454,	Adjusted R-squared:  0.4396 
F-statistic: 31.47 on 13 and 492 DF,  p-value: < 2.2e-16
```

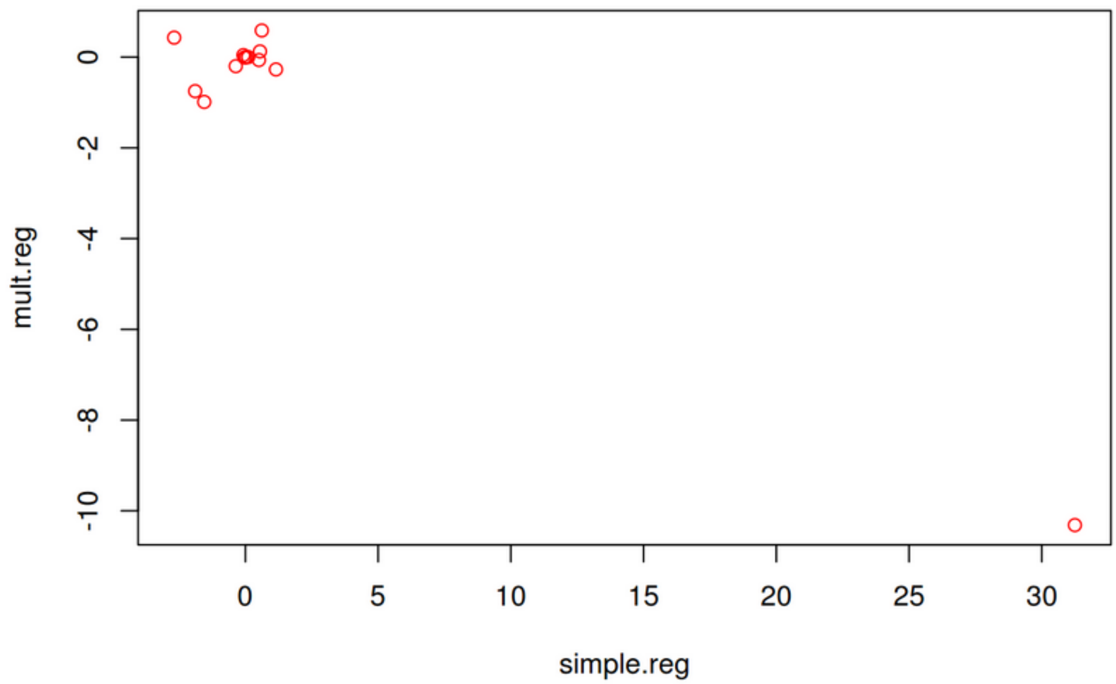
**The following predictors have p-value as:**

Rad – 6.46e-11
Dis – 0.000502
Med – 0.001087
Zn – 0.017025
Black – 0.040702

The above predictors rejects the Null hypothesis because they have low p-value.

15.c)

isl15c.R




There is a difference between the simple and multiple regression coefficients. This difference is due to the fact that in the simple regression case, the slope term represents the average effect of an increase in the predictor, ignoring other predictors. In contrast, in the multiple regression case, the slope term represents the average effect of an increase in the predictor, while holding other predictors fixed. It does make sense for the multiple regression to suggest no relationship between the response and some of the predictors while the simple linear regression implies the opposite because the correlation between the predictors show some strong relationships between some of the predictors.

>cor(Boston[-c(1,4)])

```
##                 zn      indus        nox         rm        age        dis
## zn       1.0000000 -0.5338282 -0.5166037  0.3119906 -0.5695373  0.6644082
## indus   -0.5338282  1.0000000  0.7636514 -0.3916759  0.6447785 -0.7080270
## nox     -0.5166037  0.7636514  1.0000000 -0.3021882  0.7314701 -0.7692301
## rm       0.3119906 -0.3916759 -0.3021882  1.0000000 -0.2402649  0.2052462
## age     -0.5695373  0.6447785  0.7314701 -0.2402649  1.0000000 -0.7478805
## dis      0.6644082 -0.7080270 -0.7692301  0.2052462 -0.7478805  1.0000000
## rad     -0.3119478  0.5951293  0.6114406 -0.2098467  0.4560225 -0.4945879
## tax     -0.3145633  0.7207602  0.6680232 -0.2920478  0.5064556 -0.5344316
## ptratio -0.3916785  0.3832476  0.1889327 -0.3555015  0.2615150 -0.2324705
## black    0.1755203 -0.3569765 -0.3800506  0.1280686 -0.2735340  0.2915117
## lstat   -0.4129946  0.6037997  0.5908789 -0.6138083  0.6023385 -0.4969958
## medv     0.3604453 -0.4837252 -0.4273208  0.6953599 -0.3769546  0.2499287
##                rad        tax    ptratio      black      lstat       medv
## zn      -0.3119478 -0.3145633 -0.3916785  0.1755203 -0.4129946  0.3604453
## indus    0.5951293  0.7207602  0.3832476 -0.3569765  0.6037997 -0.4837252
## nox      0.6114406  0.6680232  0.1889327 -0.3800506  0.5908789 -0.4273208
## rm      -0.2098467 -0.2920478 -0.3555015  0.1280686 -0.6138083  0.6953599
## age      0.4560225  0.5064556  0.2615150 -0.2735340  0.6023385 -0.3769546
## dis     -0.4945879 -0.5344316 -0.2324705  0.2915117 -0.4969958  0.2499287
## rad      1.0000000  0.9102282  0.4647412 -0.4444128  0.4886763 -0.3816262
## tax      0.9102282  1.0000000  0.4608530 -0.4418080  0.5439934 -0.4685359
## ptratio  0.4647412  0.4608530  1.0000000 -0.1773833  0.3740443 -0.5077867
## black   -0.4444128 -0.4418080 -0.1773833  1.0000000 -0.3660869  0.3334608
## lstat    0.4886763  0.5439934  0.3740443 -0.3660869  1.0000000 -0.7376627
## medv    -0.3816262 -0.4685359 -0.5077867  0.3334608 -0.7376627  1.0000000
```

So for example, when “age” is high there is a tendency in “dis” to be low, hence in simple linear regression which only examines “crim” versus “age”, we observe that higher values of “age” are associated with higher values of “crim”, even though “age” does not actually affect “crim”. So “age” is a surrogate for “dis”; “age” gets credit for the effect of “dis” on “crim”.

15.d)

Please find attached below the .R file

isl15d.R

‘Chas’ does not support

```
> lmpoly(crim,zn)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
```

```
   Min     1Q Median     3Q    Max 
-4.821 -4.614 -1.294  0.473 84.130 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    3.6135     0.3722   9.709  < 2e-16 ***
poly(x, 3)1 -38.7498     8.3722  -4.628  4.7e-06 ***
poly(x, 3)2  23.9398     8.3722   2.859  0.00442 ** 
poly(x, 3)3 -10.0719     8.3722  -1.203  0.22954    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.372 on 502 degrees of freedom
Multiple R-squared:  0.05824,	Adjusted R-squared:  0.05261 
F-statistic: 10.35 on 3 and 502 DF,  p-value: 1.281e-06
```

```
> lmpoly(crim,indus)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-8.278 -2.514  0.054  0.764 79.713 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)    3.614      0.330  10.950  < 2e-16 ***
poly(x, 3)1   78.591      7.423  10.587  < 2e-16 ***
poly(x, 3)2  -24.395      7.423  -3.286  0.00109 ** 
poly(x, 3)3  -54.130      7.423  -7.292  1.2e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.423 on 502 degrees of freedom
Multiple R-squared:  0.2597,	Adjusted R-squared:  0.2552 
F-statistic: 58.69 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
> lmpoly(crim,nox)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-9.110 -2.068 -0.255  0.739 78.302 
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   3.6135     0.3216  11.237  < 2e-16 ***
poly(x, 3)1  81.3720     7.2336  11.249  < 2e-16 ***
poly(x, 3)2 -28.8286     7.2336  -3.985 7.74e-05 ***
poly(x, 3)3 -60.3619     7.2336  -8.345 6.96e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.234 on 502 degrees of freedom
Multiple R-squared:  0.297,	Adjusted R-squared:  0.2928 
F-statistic: 70.69 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
> lmpoly(crim,rm)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-18.485  -3.468  -2.221  -0.015  87.219 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   3.6135     0.3703   9.758  < 2e-16 ***
poly(x, 3)1 -42.3794     8.3297  -5.088 5.13e-07 ***
poly(x, 3)2  26.5768     8.3297   3.191  0.00151 ** 
poly(x, 3)3  -5.5103     8.3297  -0.662  0.50858    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.33 on 502 degrees of freedom
Multiple R-squared:  0.06779,	Adjusted R-squared:  0.06222 
F-statistic: 12.17 on 3 and 502 DF,  p-value: 1.067e-07
```

```
> lmpoly(crim,age)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-9.762 -2.673 -0.516  0.019 82.842 
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   3.6135     0.3485  10.368  < 2e-16 ***
poly(x, 3)1  68.1820     7.8397   8.697  < 2e-16 ***
poly(x, 3)2  37.4845     7.8397   4.781 2.29e-06 ***
poly(x, 3)3  21.3532     7.8397   2.724  0.00668 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.84 on 502 degrees of freedom
Multiple R-squared:  0.1742,	Adjusted R-squared:  0.1693 
F-statistic: 35.31 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
> lmpoly(crim,dis)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-10.757  -2.588   0.031   1.267  76.378 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   3.6135     0.3259  11.087  < 2e-16 ***
poly(x, 3)1 -73.3886     7.3315 -10.010  < 2e-16 ***
poly(x, 3)2  56.3730     7.3315   7.689 7.87e-14 ***
poly(x, 3)3 -42.6219     7.3315  -5.814 1.09e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.331 on 502 degrees of freedom
Multiple R-squared:  0.2778,	Adjusted R-squared:  0.2735 
F-statistic: 64.37 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
> lmpoly(crim,rad)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-10.381  -0.412  -0.269   0.179  76.217 
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    3.6135     0.2971  12.164  < 2e-16 ***
poly(x, 3)1 120.9074     6.6824  18.093  < 2e-16 ***
poly(x, 3)2  17.4923     6.6824   2.618  0.00912 ** 
poly(x, 3)3   4.6985     6.6824   0.703  0.48231    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.682 on 502 degrees of freedom
Multiple R-squared:    0.4,	Adjusted R-squared:  0.3965 
F-statistic: 111.6 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
> lmpoly(crim,tax)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-13.273  -1.389   0.046   0.536  76.950 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    3.6135     0.3047  11.860  < 2e-16 ***
poly(x, 3)1 112.6458     6.8537  16.436  < 2e-16 ***
poly(x, 3)2  32.0873     6.8537   4.682 3.67e-06 ***
poly(x, 3)3  -7.9968     6.8537  -1.167    0.244    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.854 on 502 degrees of freedom
Multiple R-squared:  0.3689,	Adjusted R-squared:  0.3651 
F-statistic:  97.8 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
> lmpoly(crim,ptratio)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-6.833 -4.146 -1.655  1.408 82.697 
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    3.614      0.361  10.008  < 2e-16 ***
poly(x, 3)1   56.045      8.122   6.901 1.57e-11 ***
poly(x, 3)2   24.775      8.122   3.050  0.00241 ** 
poly(x, 3)3  -22.280      8.122  -2.743  0.00630 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.122 on 502 degrees of freedom
Multiple R-squared:  0.1138,	Adjusted R-squared:  0.1085 
F-statistic: 21.48 on 3 and 502 DF,  p-value: 4.171e-13
```

```
> lmpoly(crim,black)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-13.096  -2.343  -2.128  -1.439  86.790 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    3.6135     0.3536  10.218   <2e-16 ***
poly(x, 3)1 -74.4312     7.9546  -9.357   <2e-16 ***
poly(x, 3)2   5.9264     7.9546   0.745    0.457    
poly(x, 3)3  -4.8346     7.9546  -0.608    0.544    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.955 on 502 degrees of freedom
Multiple R-squared:  0.1498,	Adjusted R-squared:  0.1448 
F-statistic: 29.49 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
> lmpoly(crim,lstat)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-15.234  -2.151  -0.486   0.066  83.353 
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    3.6135     0.3392  10.654   <2e-16 ***
poly(x, 3)1   88.0697     7.6294  11.543   <2e-16 ***
poly(x, 3)2   15.8882     7.6294   2.082   0.0378 *  
poly(x, 3)3  -11.5740     7.6294  -1.517   0.1299    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.629 on 502 degrees of freedom
Multiple R-squared:  0.2179,	Adjusted R-squared:  0.2133 
F-statistic: 46.63 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
> lmpoly(crim,medv)

Call:
lm(formula = y ~ poly(x, 3), data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-24.427  -1.976  -0.437   0.439  73.655 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)    3.614      0.292  12.374  < 2e-16 ***
poly(x, 3)1  -75.058      6.569 -11.426  < 2e-16 ***
poly(x, 3)2   88.086      6.569  13.409  < 2e-16 ***
poly(x, 3)3  -48.033      6.569  -7.312 1.05e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.569 on 502 degrees of freedom
Multiple R-squared:  0.4202,	Adjusted R-squared:  0.4167 
F-statistic: 121.3 on 3 and 502 DF,  p-value: < 2.2e-16
```

The below table shows which predictors have non-linear association based on p-value.

| Predictor | Linear | B0+B1*x+B2*x^2 (Non-linear) | B0+B1*x+B2*x^2+BX^3 (Non-linear) |
|---|---|---|---|
| Zn | NO | YES | NO |
| Indus | YES | YES | YES |
| Nox | YES | YES | YES |
| Rm | YES | YES | NO |
| Age | YES | YES | YES |
| Dis | YES | YES | YES |
| Rad | YES | YES | NO |
| Tax | YES | YES | NO |
| Ptratio | YES | YES | YES |
| Black | YES | NO | NO |
| Lstat | YES | YES | NO |
| medv | YES | YES | YES |