

Assignment5

October 31, 2018

```
In [1]: setwd("/Users/chengaoxiang/Desktop/18_Fall/Data_Analytics/Assignment/Assignment5")
```

```
In [2]: getwd()
```

```
"/Users/chengaoxiang/Desktop/18_Fall/Data_Analytics/Assignment/Assignment5"
```

```
In [3]: collection_data_ori <- read.csv("summary.csv", header = TRUE)
```

```
In [4]: head(collection_data_ori)
```

Entity	Laureates10million	cocoaconsump2010	alcoholconsump2015	fishconsump2013	Smoke
Algeria	0.476	0.575	0.6	3.92	312310
Argentina	1.119	0.785	7.6	7.05	598769
Australia	4.844	2.874	12.6	26.09	296126
Austria	23.995	3.800	8.5	13.88	210904
Azerbaijan	1.008	NA	2.1	2.13	162218
Bangladesh	0.060	NA	0.2	19.21	240137

```
In [5]: summary(collection_data_ori)
```

Entity	Laureates10million	cocoaconsump2010	alcoholconsump2015
Algeria : 1	Min. : 0.0470	Min. :0.027	Min. : 0.100
Argentina : 1	1st Qu.: 0.3407	1st Qu.:0.785	1st Qu.: 5.250
Australia : 1	Median : 2.0795	Median :1.792	Median : 9.050
Austria : 1	Mean : 7.6136	Mean :2.055	Mean : 8.096
Azerbaijan: 1	3rd Qu.: 8.6252	3rd Qu.:3.022	3rd Qu.:11.200
Bangladesh: 1	Max. :111.3170	Max. :5.883	Max. :17.100
(Other) :64		NA's :29	

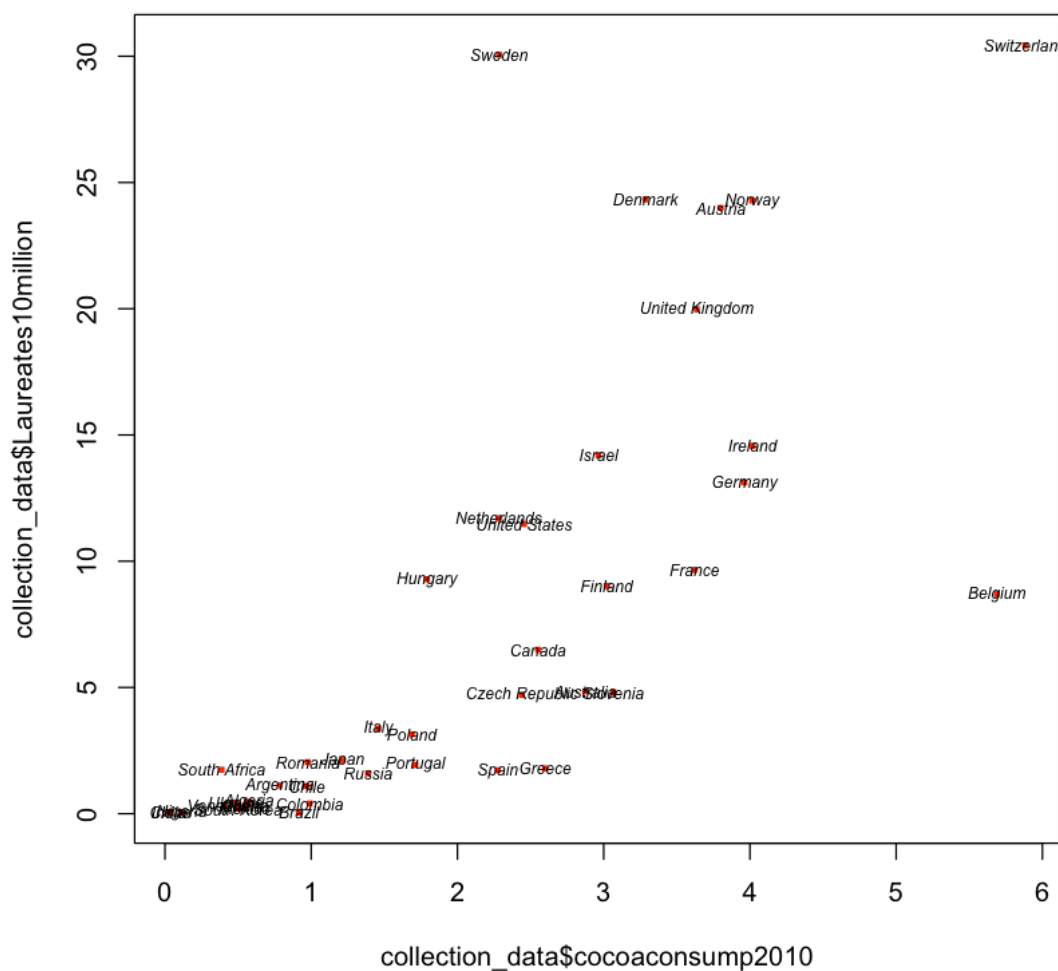
fishconsump2013	Smoker_Num2012
Min. : 1.290	Min. : 14130
1st Qu.: 7.438	1st Qu.: 901824
Median :19.095	Median : 2911105
Mean :20.461	Mean : 11418155
3rd Qu.:25.003	3rd Qu.: 9565081
Max. :91.920	Max. :281714540

```
In [6]: collection_data <- na.omit(collection_data_ori)
```

```
In [7]: summary(collection_data)
```

Entity	Laureates10million	cocoaconsump2010	alcoholconsump2015
Algeria : 1	Min. : 0.047	Min. :0.027	Min. : 0.600
Argentina: 1	1st Qu.: 0.476	1st Qu.:0.785	1st Qu.: 7.600
Australia: 1	Median : 3.149	Median :1.792	Median :10.300
Austria : 1	Mean : 7.317	Mean :2.055	Mean : 9.534
Belgium : 1	3rd Qu.:11.476	3rd Qu.:3.022	3rd Qu.:11.500
Brazil : 1	Max. :30.431	Max. :5.883	Max. :14.500
(Other) :35			
fishconsump2013	Smoker_Num2012		
Min. : 3.92	Min. : 393481		
1st Qu.:10.48	1st Qu.: 2109044		
Median :20.76	Median : 3874289		
Mean :21.38	Mean : 16557207		
3rd Qu.:26.09	3rd Qu.: 10355707		
Max. :53.76	Max. :281714540		

```
In [31]: plot(collection_data$cocoaconsump2010,collection_data$Laureates10million,col='red',ce  
text(collection_data$cocoaconsump2010,collection_data$Laureates10million,collection_d
```



```
In [32]: lm.nobel <- lm(Laureates10million ~ cocoaconsump2010, data = collection_data)
```

```
In [33]: summary(lm.nobel)
```

Call:

```
lm(formula = Laureates10million ~ cocoaconsump2010, data = collection_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-14.6144	-2.5895	-0.5259	1.7659	21.7208

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.7275	1.5907	-1.086	0.284
cocoaconsump2010	4.4021	0.6274	7.016	2.04e-08 ***

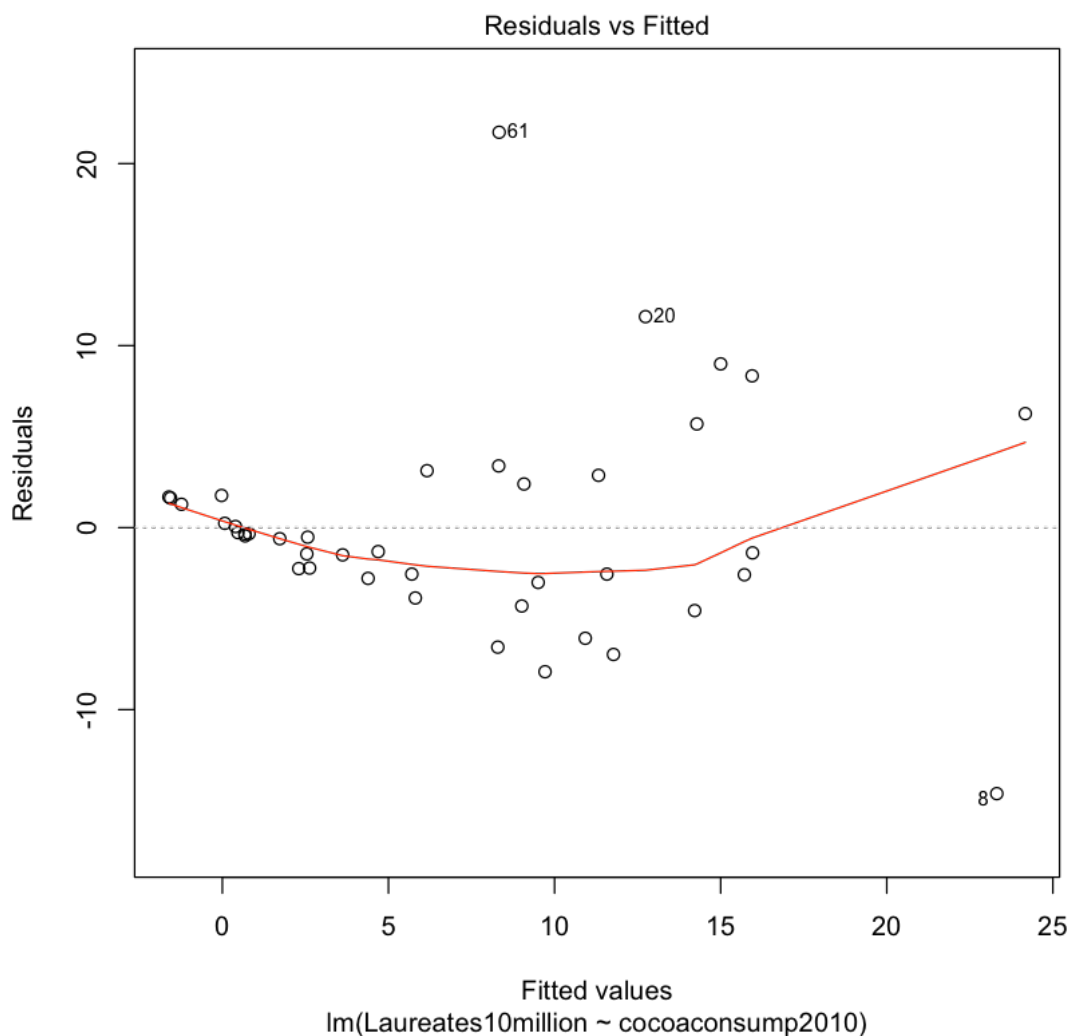
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

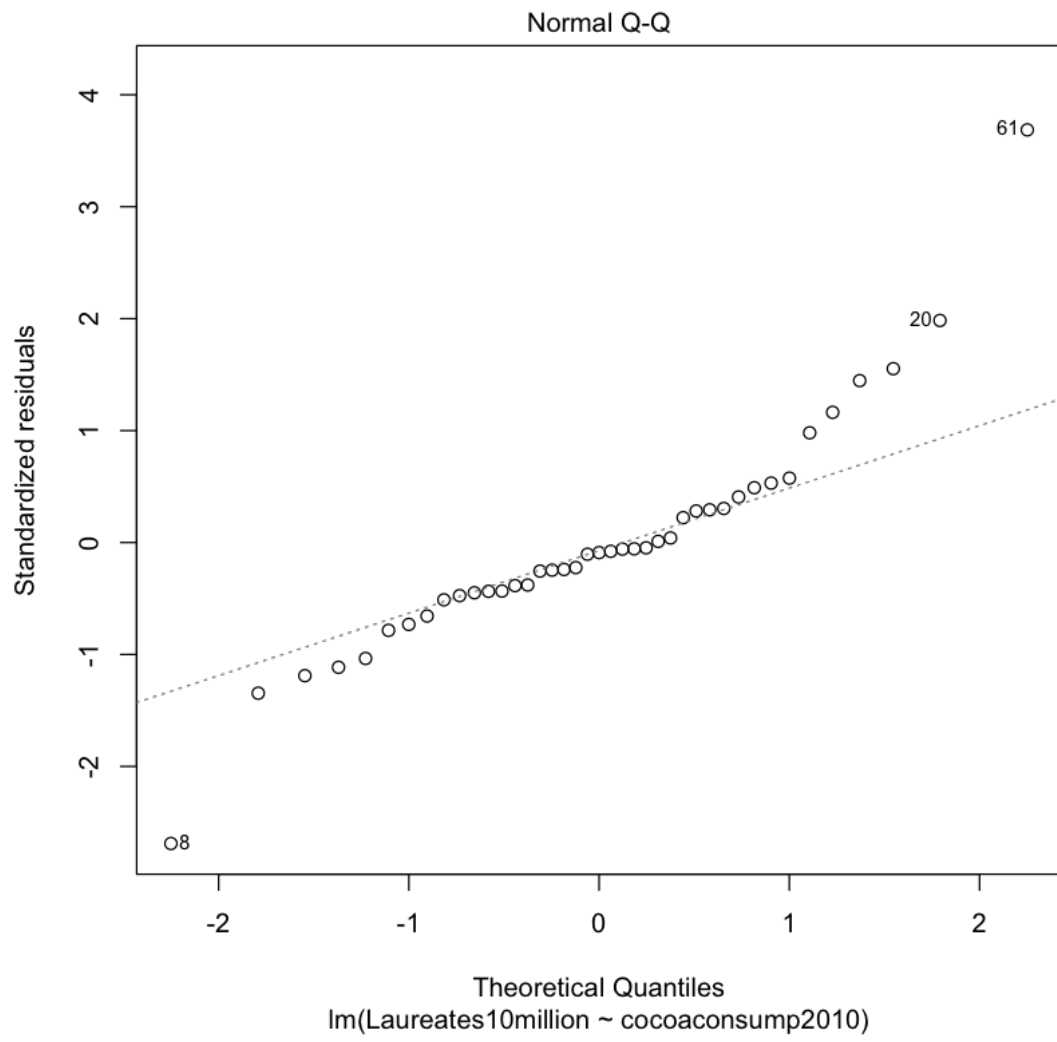
Residual standard error: 5.967 on 39 degrees of freedom

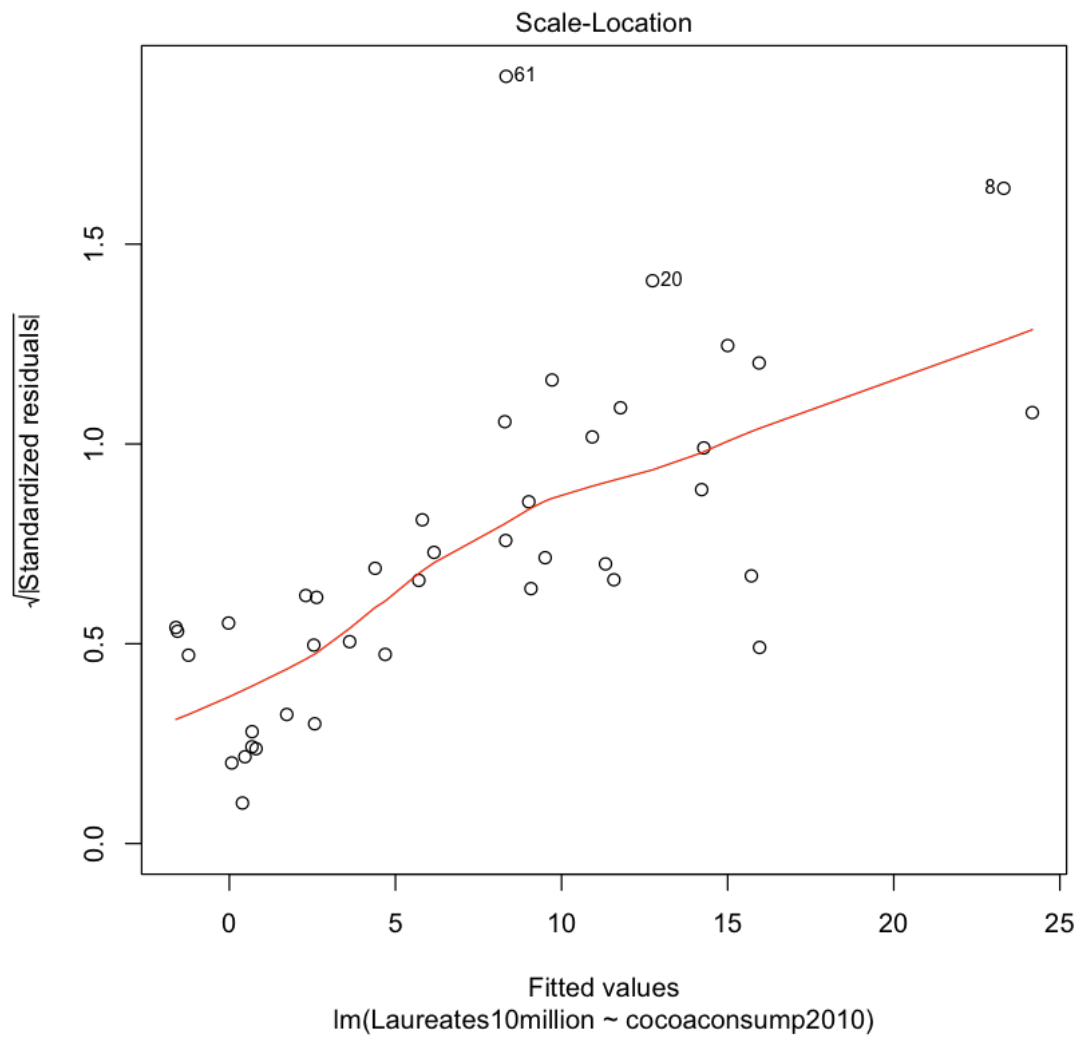
Multiple R-squared: 0.558, Adjusted R-squared: 0.5466

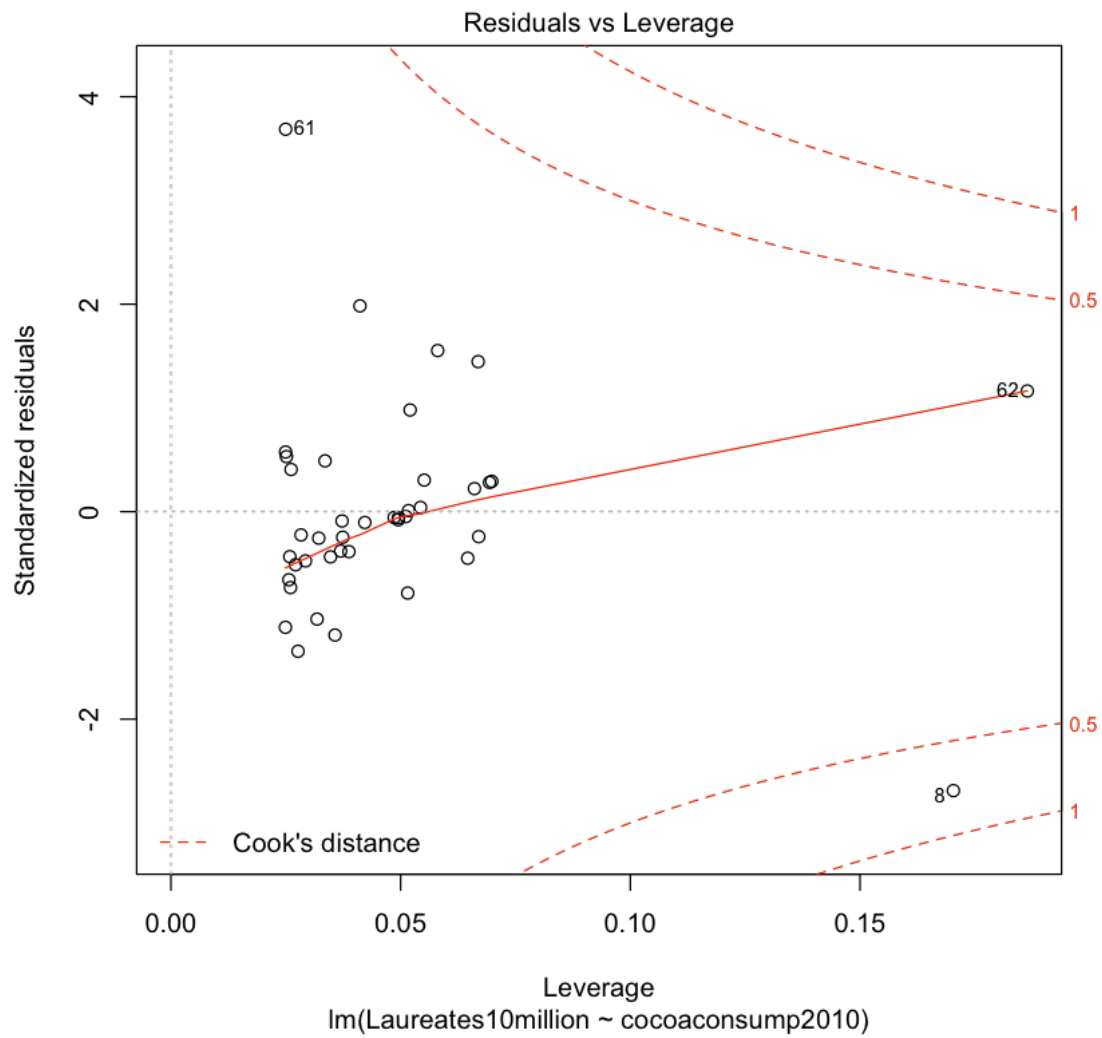
F-statistic: 49.23 on 1 and 39 DF, p-value: 2.037e-08

In [34]: plot(lm.nobel)









```
In [35]: lm.all_nobel <- lm(Laureates10million ~ ., data = collection_data[,2:6])
```

```
In [36]: summary(lm.all_nobel)
```

Call:

```
lm(formula = Laureates10million ~ ., data = collection_data[,
  2:6])
```

Residuals:

	Min	1Q	Median	3Q	Max
	-14.795	-2.821	-1.144	2.061	21.094

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.730e+00	3.616e+00	0.478	0.635
cocoaconsump2010	4.591e+00	7.008e-01	6.551	1.28e-07 ***
alcoholconsump2015	-4.507e-01	3.424e-01	-1.317	0.196
fishconsump2013	2.056e-02	7.155e-02	0.287	0.775
Smoker_Num2012	7.489e-10	2.217e-08	0.034	0.973

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 6.06 on 36 degrees of freedom

Multiple R-squared: 0.5792, Adjusted R-squared: 0.5325

F-statistic: 12.39 on 4 and 36 DF, p-value: 1.955e-06

```
In [37]: final.lm <- step(lm.all_nobel)
```

Start: AIC=152.41

Laureates10million ~ cocoaconsump2010 + alcoholconsump2015 +
fishconsump2013 + Smoker_Num2012

	Df	Sum of Sq	RSS	AIC
- Smoker_Num2012	1	0.04	1322.0	150.41
- fishconsump2013	1	3.03	1325.0	150.50
- alcoholconsump2015	1	63.65	1385.6	152.33
<none>			1322.0	152.41
- cocoaconsump2010	1	1576.08	2898.1	182.59

Step: AIC=150.41

Laureates10million ~ cocoaconsump2010 + alcoholconsump2015 +
fishconsump2013

	Df	Sum of Sq	RSS	AIC
- fishconsump2013	1	3.24	1325.3	148.51
- alcoholconsump2015	1	64.74	1386.8	150.37
<none>			1322.0	150.41
- cocoaconsump2010	1	1727.99	3050.0	182.68

Step: AIC=148.51

Laureates10million ~ cocoaconsump2010 + alcoholconsump2015

	Df	Sum of Sq	RSS	AIC
- alcoholconsump2015	1	63.48	1388.8	148.43
<none>			1325.3	148.51
- cocoaconsump2010	1	1808.79	3134.1	181.80

Step: AIC=148.43


```
Laureates10million ~ cocoaconsump2010
```

	Df	Sum of Sq	RSS	AIC
<none>			1388.8	148.43
- cocoaconsump2010	1	1752.9	3141.6	179.90

```
In [38]: summary(final.lm)
```

Call:

```
lm(formula = Laureates10million ~ cocoaconsump2010, data = collection_data[,  
  2:6])
```

Residuals:

Min	1Q	Median	3Q	Max
-14.6144	-2.5895	-0.5259	1.7659	21.7208

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.7275	1.5907	-1.086	0.284
cocoaconsump2010	4.4021	0.6274	7.016	2.04e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.967 on 39 degrees of freedom

Multiple R-squared: 0.558, Adjusted R-squared: 0.5466

F-statistic: 49.23 on 1 and 39 DF, p-value: 2.037e-08