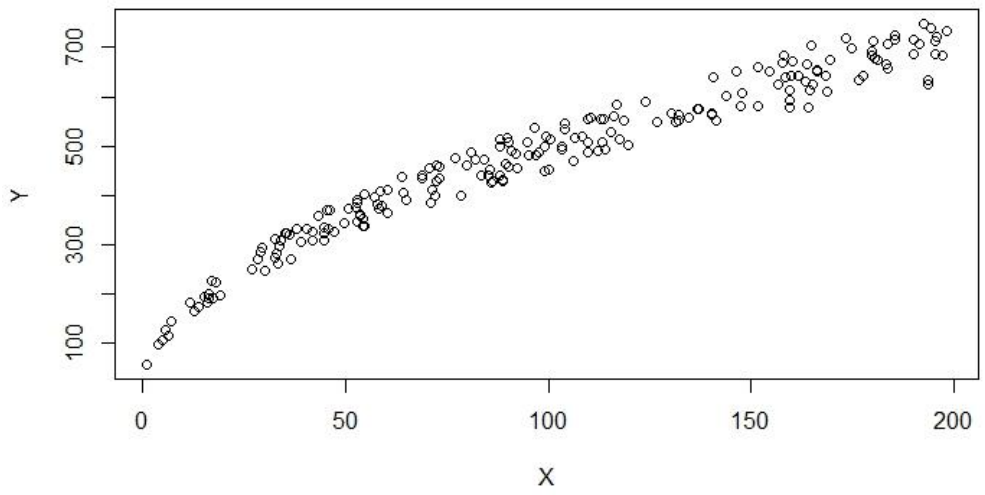
ISL Assignment 1-Q2

Q2. For this dataset,

i. **Plot the dataset**

**Answer:**

```
plot(Y~X,data=Assignment_Dataset_A)
```




ii. **Apply linear regression model on the orginal data set and on its transformations such log of the data**

**Answer:**

**Linear regression model on the original data set:**

```
> dataset.mod1=lm(Y~X,data=Assignment_Dataset_A)
> dataset.mod1

Call:
lm(formula = Y ~ X, data = Assignment_Dataset_A)

Coefficients:
(Intercept)            X
199.517              2.769
```

**Transformations (log of the data):**

```
> dataset.mod2=lm(Y~log(X),data=Assignment_Dataset_A)
> dataset.mod2

Call:
lm(formula = Y ~ log(X), data = Assignment_Dataset_A)

Coefficients:
(Intercept)       log(X)
 -289.3          175.6
```

**iii. Report on the summary results on your models.**

**Answer:**

**Original Data set summary() :**

```
> summary(dataset.mod1)

Call:
lm(formula = Y ~ X, data = Assignment_Dataset_A)

Residuals:
 Min       1Q   Median       3Q      Max
-147.224  -22.183    1.705   29.137   71.576

Coefficients:
     Estimate Std. Error t value Pr(>|t|)
(Intercept) 199.5172     5.7231   34.86   <2e-16 ***
X             2.7689     0.0506   54.72   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 39.87 on 198 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.938,    Adjusted R-squared:  0.9377
F-statistic:  2994 on 1 and 198 DF,  p-value: < 2.2e-16
```

**Transformation (log of the data) summary ():**

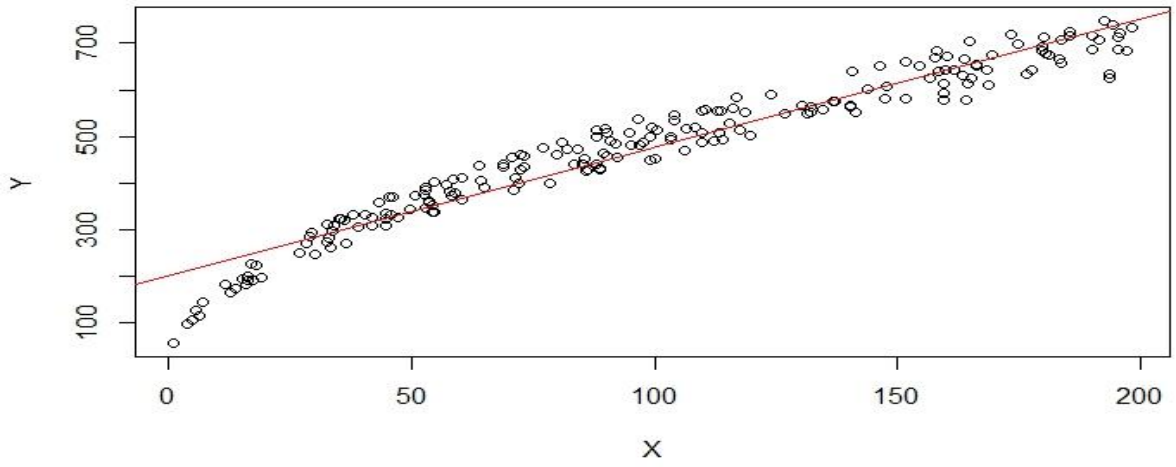
```
> summary(dataset.mod2)

Call:
lm(formula = Y ~ log(X), data = Assignment_Dataset_A)

Residuals:
Min     1Q Median     3Q    Max
-77.93 -33.58 -12.30  23.00 330.00

Coefficients:
     Estimate Std. Error t value Pr(>|t|)
(Intercept) -289.257     19.427  -14.89   <2e-16 ***
log(X)       175.574      4.396   39.94   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 53.2 on 198 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.8896,   Adjusted R-squared:  0.889
F-statistic:  1595 on 1 and 198 DF,  p-value: < 2.2e-16
```

**iv. Show the linear regression fit for the original and the transformed data through plots.**

**Answer:**

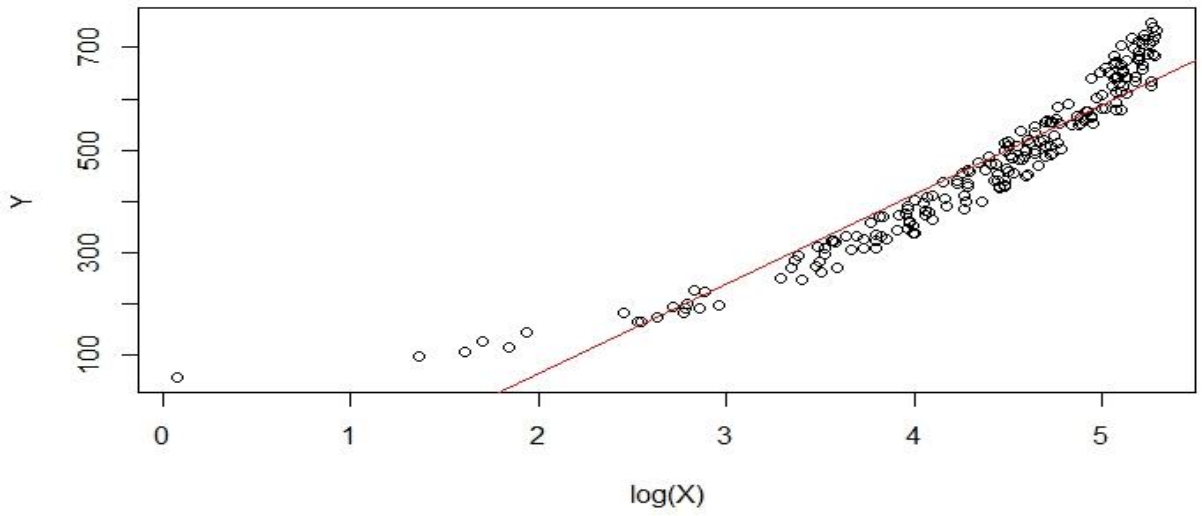
Linear regression model for the original data:

```
> plot(Y~X,data=Assignment_Dataset_A)
> abline(dataset.mod1,col="red")
```




**Linear regression fit for transformed data:**

```
> plot(Y~log(X),data=Assignment_Dataset_A)
> abline(dataset.mod2,col="red")
```




V. **Compare your model and comment them on their fit by considering factors such as p-value, $R^2$ etc.**

**Answer**:

By comparing the simple linear model of the two data models (Original, Transformed) data, we have

```
p-value: < 2.2e-16
```

As the value of is very small, we can reject the null hypothesis for both data models and infer that there is a relation among predictor and response.

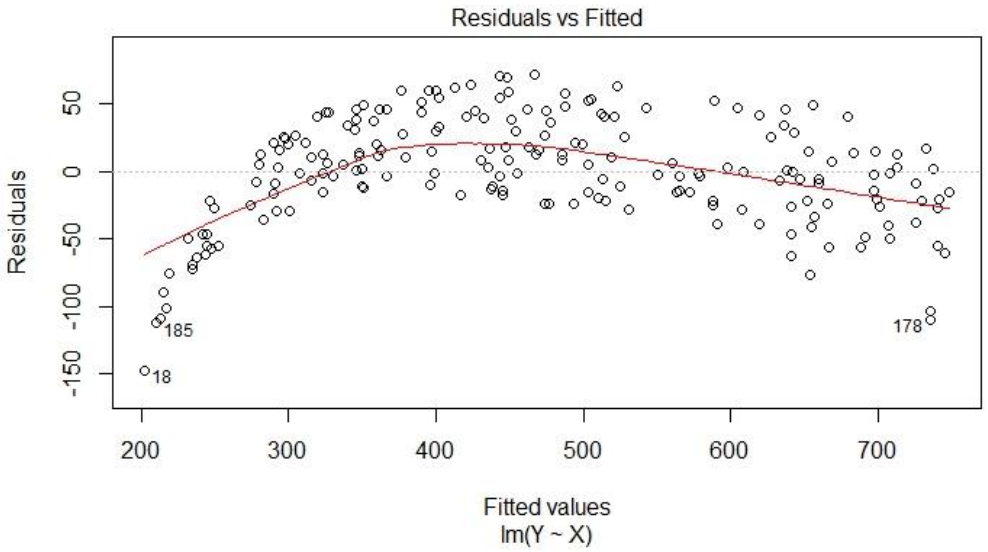
$R^2$ generally measures how close the data is fitted to the regression line. And it also represents the scatter around the fitted regression line.

Higher the value of the $R^2$ the better the model fits the data.
Original data $R^2$ : 0.938(~ 94%)
Transformed data $R^2$ : 0.8896 (~89%)

Thus, the original data model has strong relationship between predictor and response than the transformed data model.

Below graphs shows fitted regression for Original dataset & Transformed data set

Original dataset:




Transformeddataset: