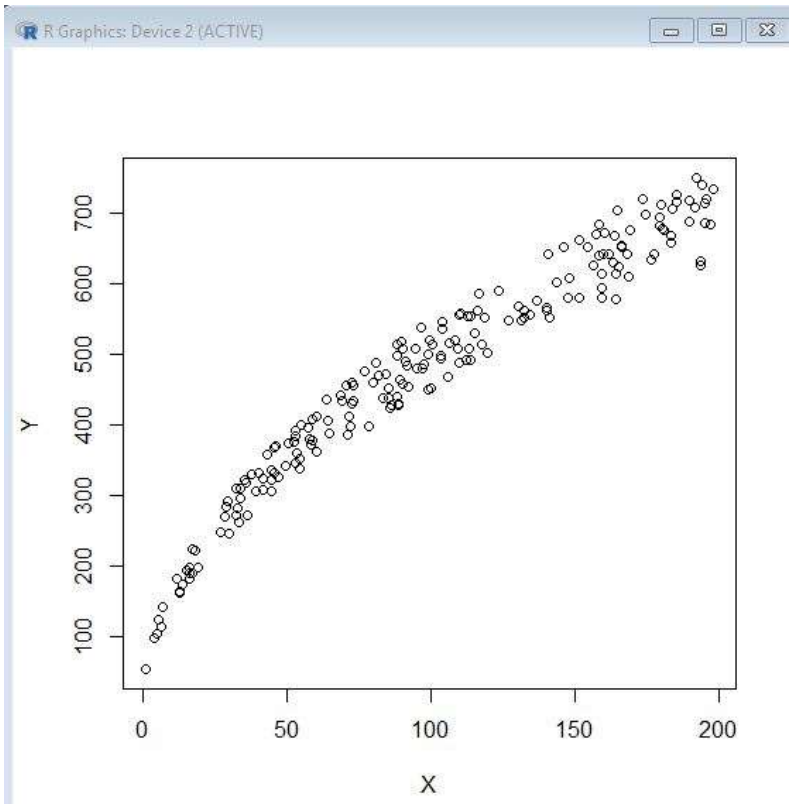
Question 2

i) Plot the dataset

```
>
> plot(Y~X,data=ISLAs12)
>
```



II) summary lmfit original

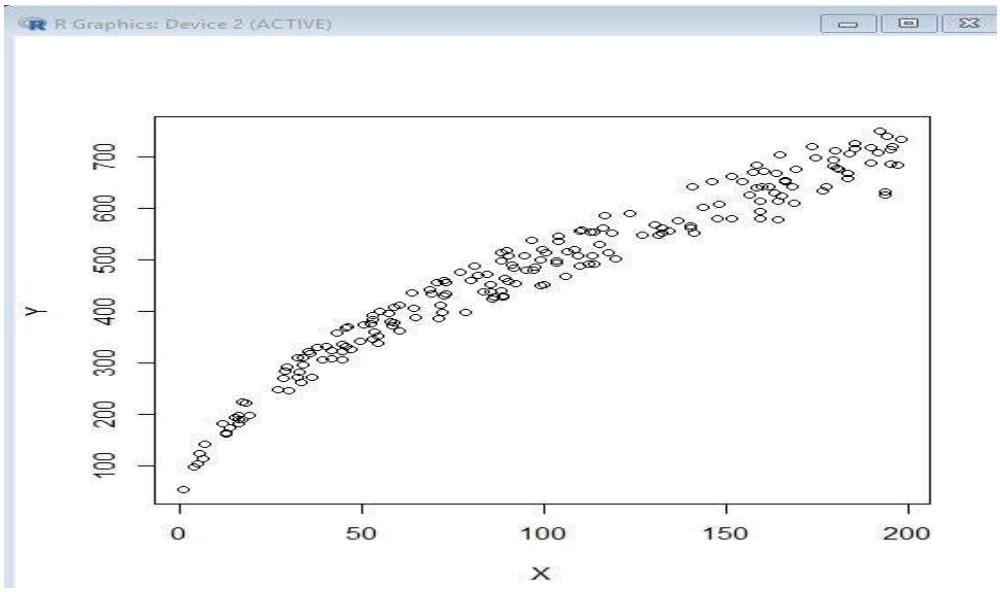
```
> plot(Y~X,data=ISLAs12)
> lmfit2<-lm(Y~X,data=ISLAs12)
> summary(lmfit2);

Call:
lm(formula = Y ~ X, data = ISLAs12)

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
Multiple R-squared:  0.938,     Adjusted R-squared:  0.9377
F-statistic:  2994 on 1 and 198 DF,  p-value: < 2.2e-16
```

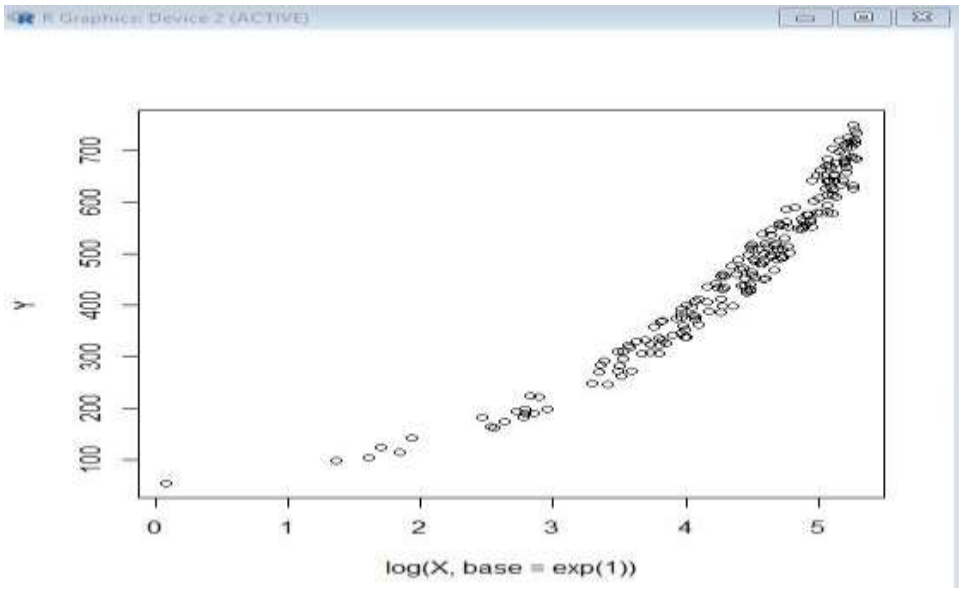
On transformed data

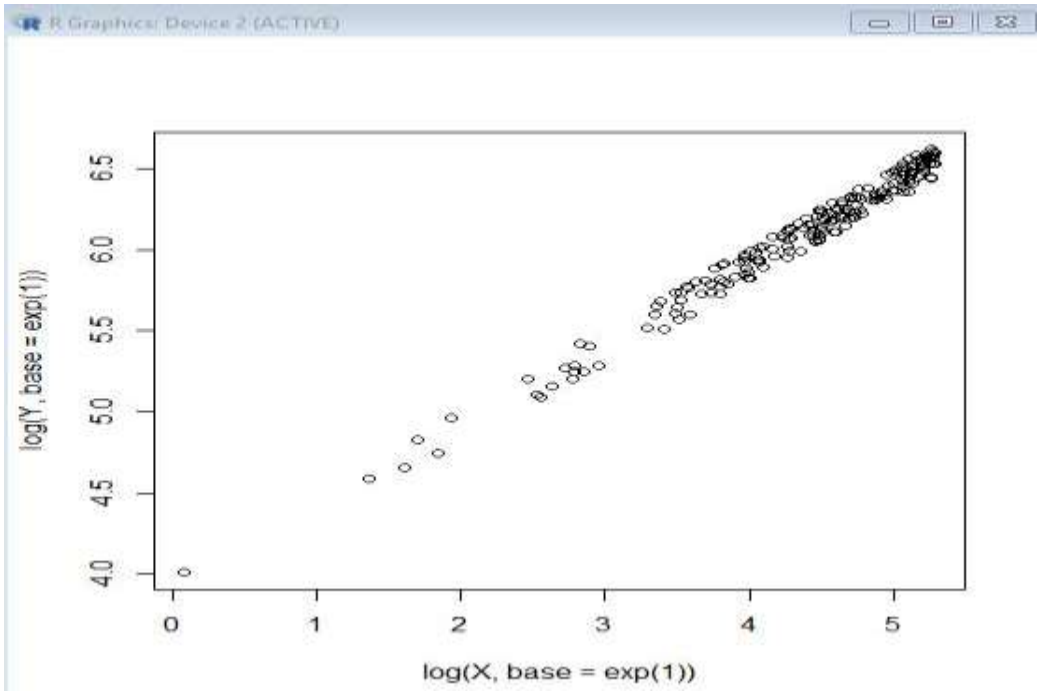
```
> lmdata4<-lm(Y~log(X,base=exp(1)),data=ISLAs12)
> summary(lmdata4)

Call:
lm(formula = Y ~ log(X, base = exp(1)), data = ISLAs12)

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, base = exp(1))  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
Multiple R-squared:  0.8896,    Adjusted R-squared:  0.889
F-statistic:  1595 on 1 and 198 DF,  p-value: < 2.2e-16


> lmdata5<-lm(log(Y,base=exp(1))~log(X,base=exp(1)),data=ISLAs12)
> summary(lmdata5)

Call:
lm(formula = log(Y, base = exp(1)) ~ log(X, base = exp(1)), data = ISLAs12)

Residuals:
      Min        1Q    Median        3Q       Max
-0.107212 -0.046358 -0.000565  0.052449  0.092625

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.903721   0.021283   183.4   <2e-16 ***
log(X, base = exp(1)) 0.501951   0.004816   104.2   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05828 on 198 degrees of freedom
Multiple R-squared:  0.9821,    Adjusted R-squared:  0.982
F-statistic: 1.086e+04 on 1 and 198 DF,  p-value: < 2.2e-16


> lmdata5<-lm(log(Y,base=exp(1))~X,data=ISLAs12)
> summary(lmdata5)

Call:
lm(formula = log(Y, base = exp(1)) ~ X, data = ISLAs12)

Residuals:
     Min       1Q   Median       3Q      Max
-1.39858 -0.08071  0.04548  0.12597  0.23619

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.4039242  0.0293150  184.34   <2e-16 ***
X           0.0068710  0.0002592   26.51   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2042 on 198 degrees of freedom
Multiple R-squared:  0.7802,    Adjusted R-squared:  0.779
F-statistic: 702.6 on 1 and 198 DF,  p-value: < 2.2e-16
```

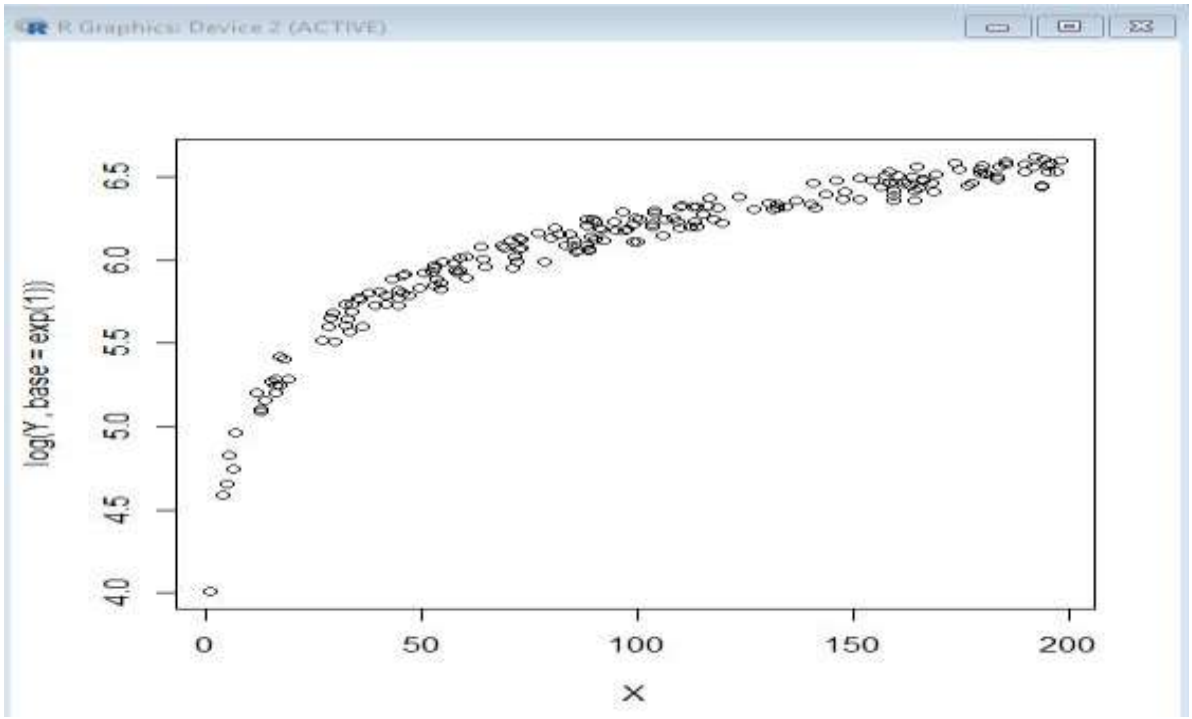
iv. Plot the linear and transformed graphs

```
> plot(Y~X,data=ISLAs12)
> plot(log(Y,base=exp(1))~X,data=ISLAs12)
> plot(log(Y,base=exp(1))~log(X,base=exp(1)),data=ISLAs12)
> plot(log(Y,base=exp(1))~X,data=ISLAs12)
>
```









v. we need to compare

- Plots number is more for the transformed data when compared to that of the original data
- By observing p-values on f-statistics it can be said that transformed data is better fit then the original data set

Original data

Residual standard error: 39.87 on 198 degrees of freedom
Multiple R-squared: 0.938, Adjusted R-squared: 0.9377
F-statistic: 2994 on 1 and 198 DF, p-value: < 2.2e-16

Transforming data

Residual standard error: 0.05828 on 198 degrees of freedom
Multiple R-squared: 0.9821, Adjusted R-squared: 0.982
F-statistic: 1.086e+04 on 1 and 198 DF, p-value: < 2.2e-16

Conclusions: P is same in both base.

Transformed dataset has the low Residual standard error value and good F-statistic value which is a strong sign for better fit.

R-square value in the transformed data varies by 98% where as in original data it varies by 93%