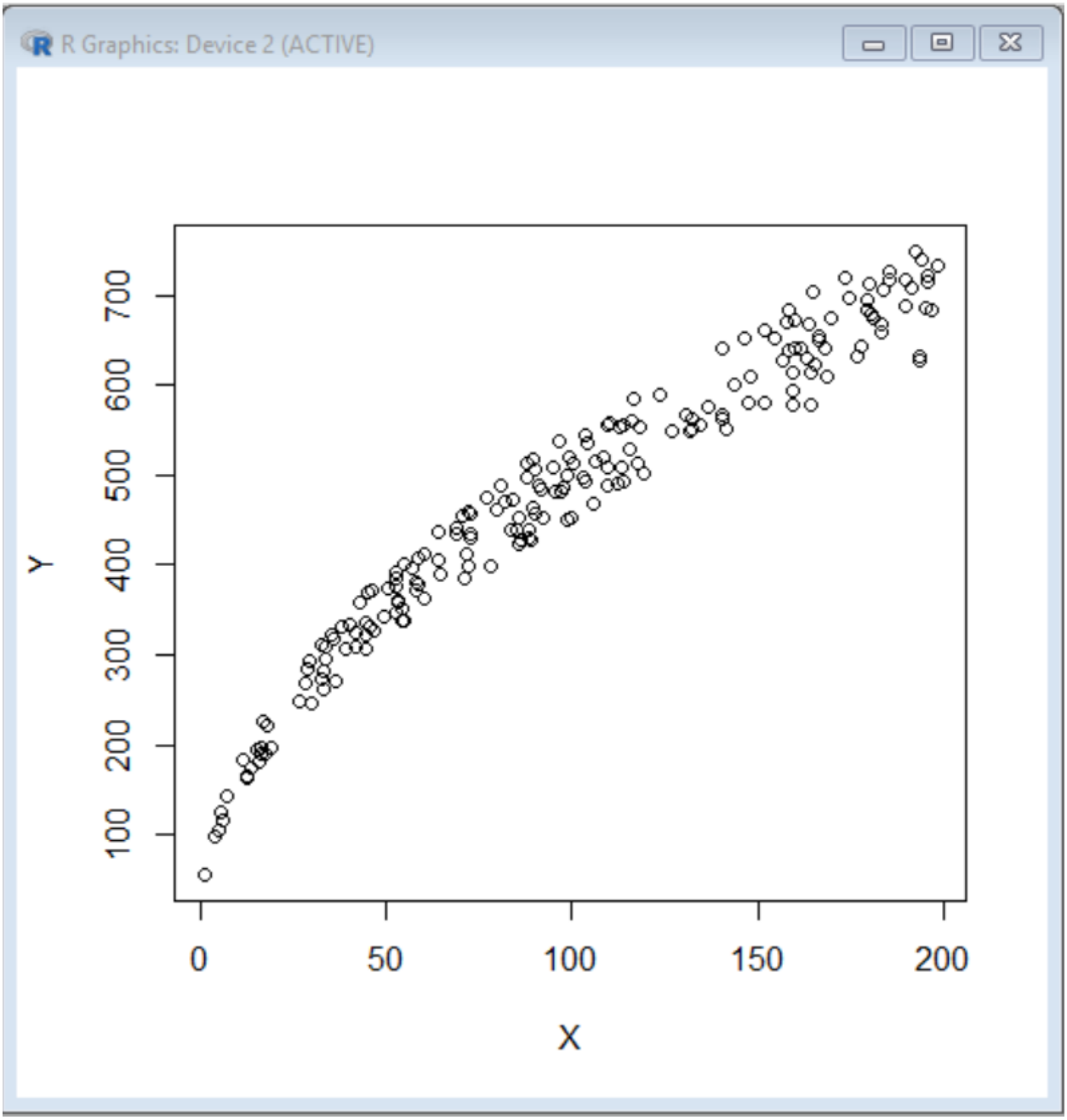
Q-2 For this dataset,

i. Plot the dataset

```
> Q2_data= read.table("Assil_Q2.txt",header = T,na.strings = "?")
> plot(Q2_data)
```



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

```
Q2_mod1=lm(X~Y, data=Q2_data)
Q2_mod2=lm(log(X)~log(Y), data=Q2_data)
```

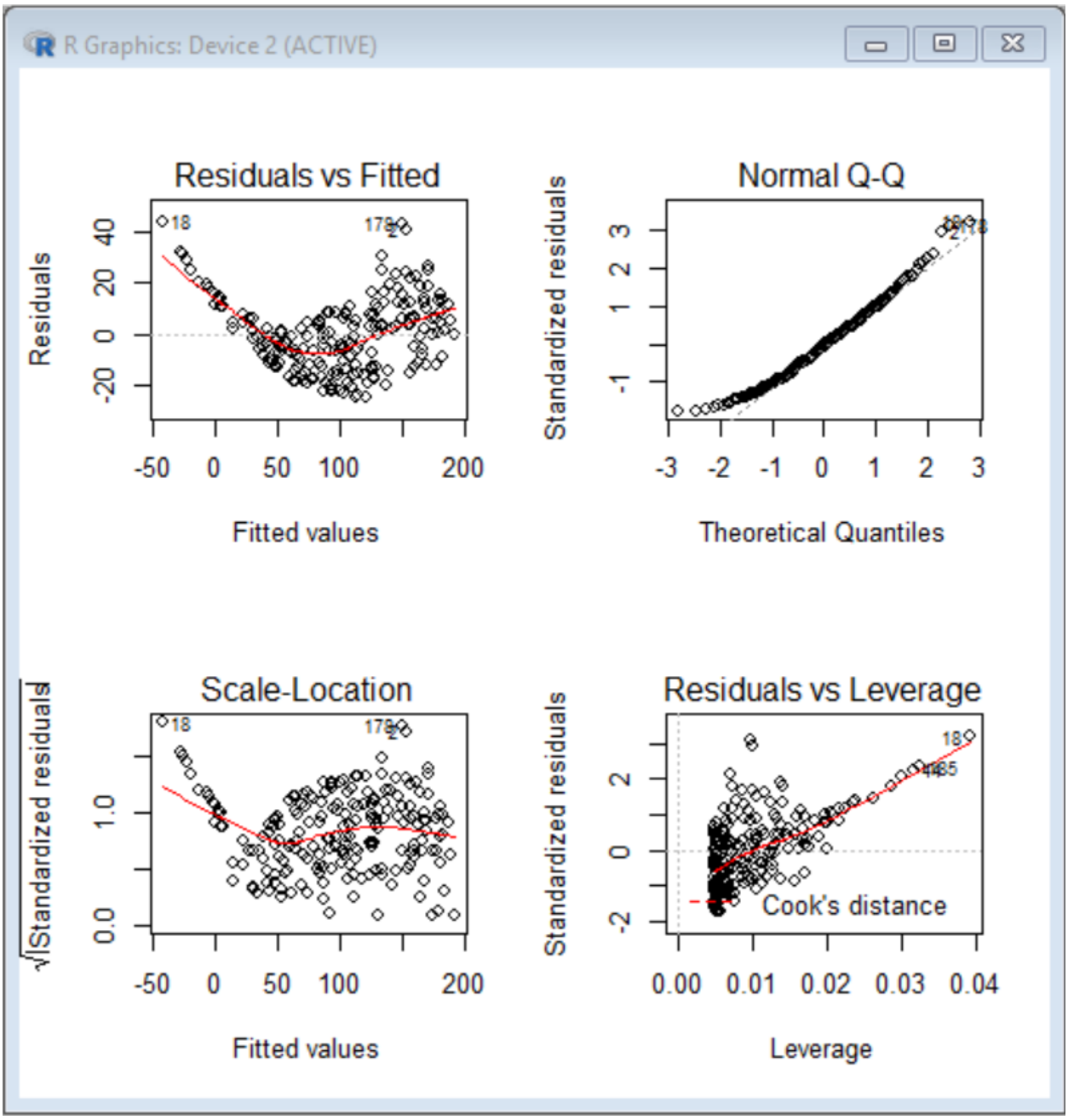
iii. Report on the summary results on your models.

```
> summary(Q2_mod1)

Call:
lm(formula = X ~ Y, data = Q2_data)

Residuals:
    Min      1Q  Median      3Q     Max
-24.363 -11.262  -0.889   8.385  43.835

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -61.481835   3.084263  -19.93   <2e-16 ***
Y             0.338753   0.006191   54.72   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.95 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
```

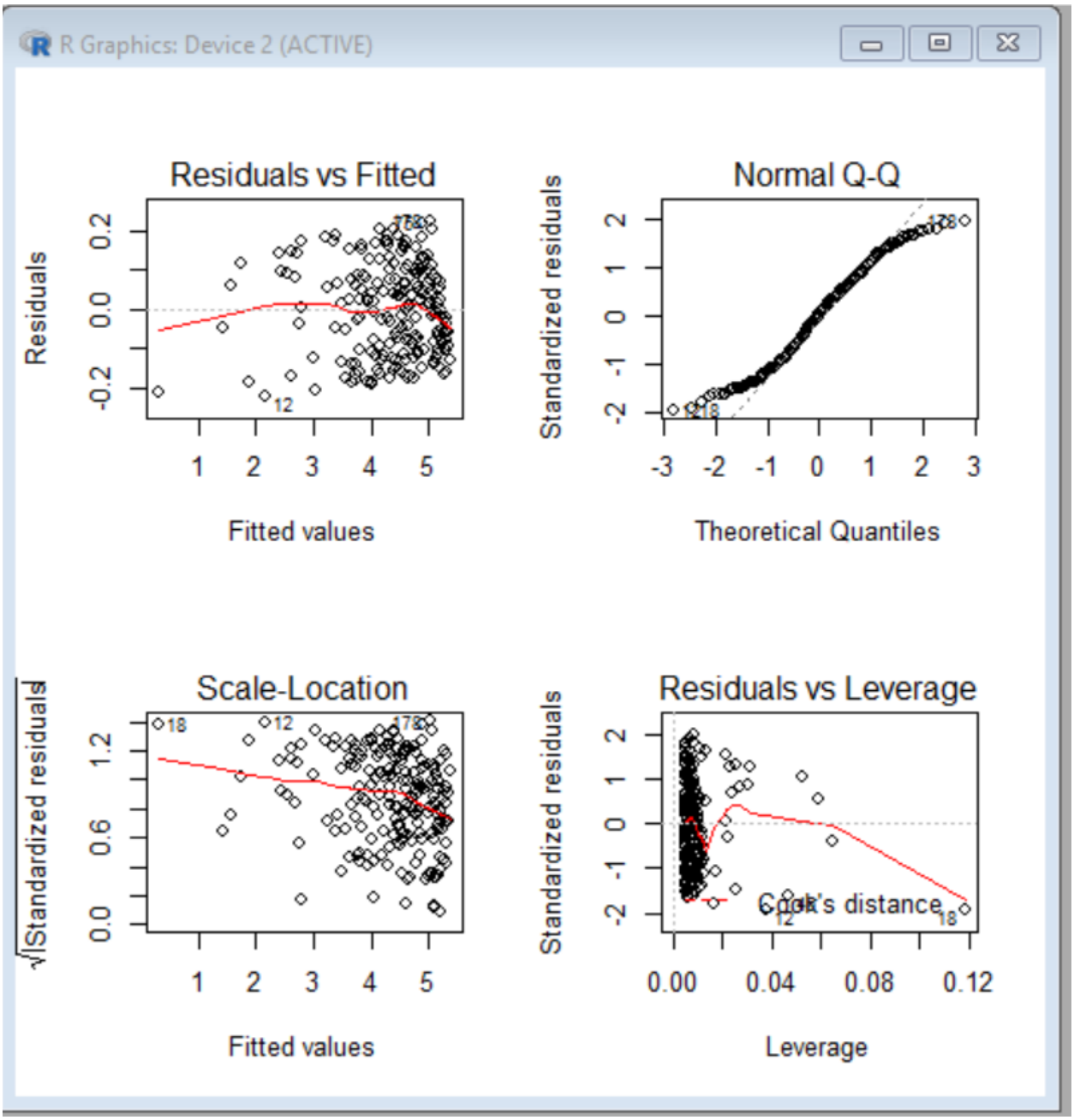
```
> summary(Q2_mod2)

Call:
lm(formula = log(X) ~ log(Y), data = Q2_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.219268 -0.100641  0.007125  0.087996  0.226412

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.56029    0.11442  -66.08   <2e-16 ***
log(Y)       1.95657    0.01877  104.23   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1151 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
```

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

```
> par(mfrow = c(2, 2))
> plot(Q2_mod1)
```




```
> par(mfrow = c(2, 2))
> plot(Q2_mod2)
```



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

Quantity	Original Model	Transformed Model
P value	2.2e-16	2.2e-16
R^2 Value	93.8 %	98.2%
Residual Standard Error	13.95	0.1151
F Statistics	2994	1.086e+04

- **P value** of both the model is the same and good. So, both the models contribute the same for the Y value with respect to X.
- **F-statistic** which represents the significance of overall model. F-statistic for the original model is far higher than 1 => Its against null hypothesis.
- The **Residual Standard Error** represents the standard deviation of the residuals. It's a measure of how close the fit is to the points. Original model is close to the fit than the transformed data.
- **Mutual R Squared** i.e., interpreted value of the percentage of variance. The variance of transformed model is higher than original model.