

TP: Is Batman Somewhere

Olaniyan Folajimi and Carlos Vargas

December 9, 2021

```
library(ggplot2)

myData <- read.table(file="bats.csv", sep=";", skip=3, header=T)
names(myData)

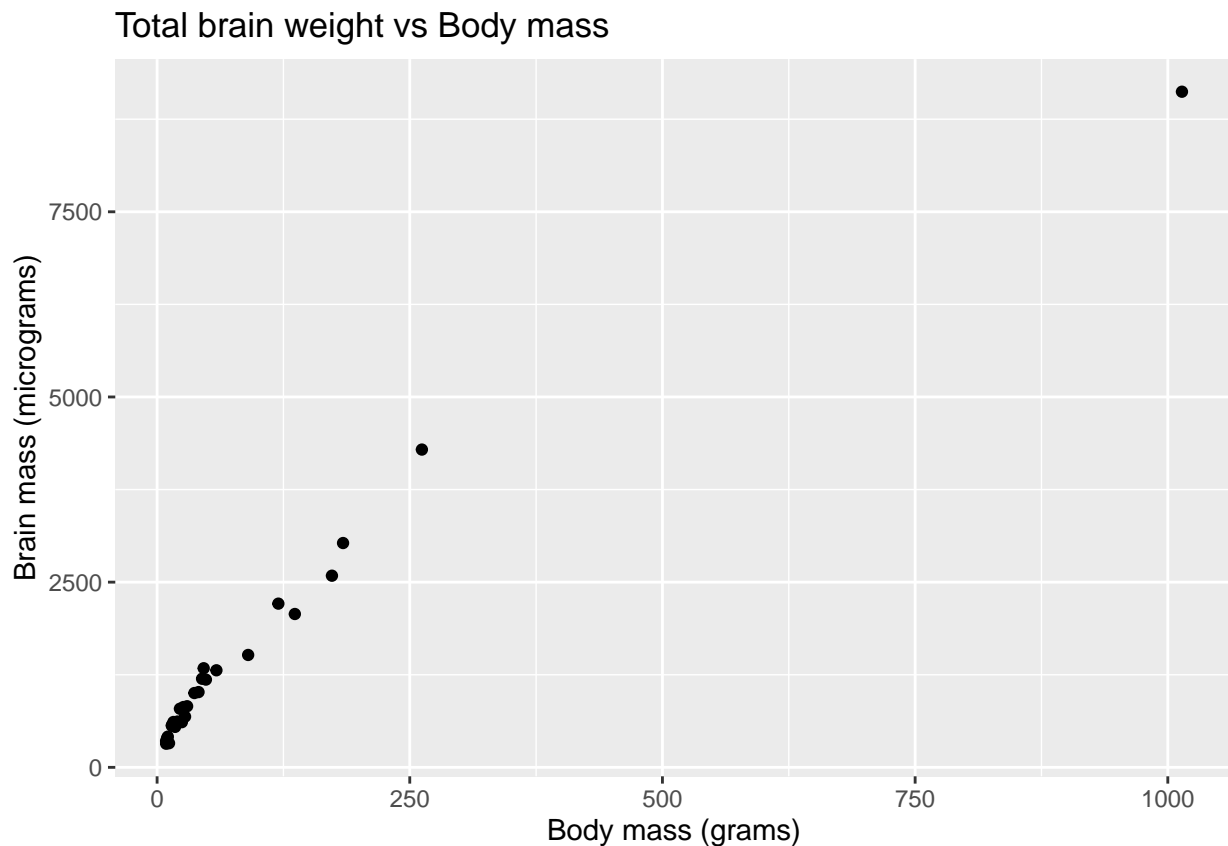
## [1] "Species" "Diet" "Clade" "BOW" "BRW" "AUD" "MOB"
## [8] "HIP"

dim(myData)

## [1] 63 8
```

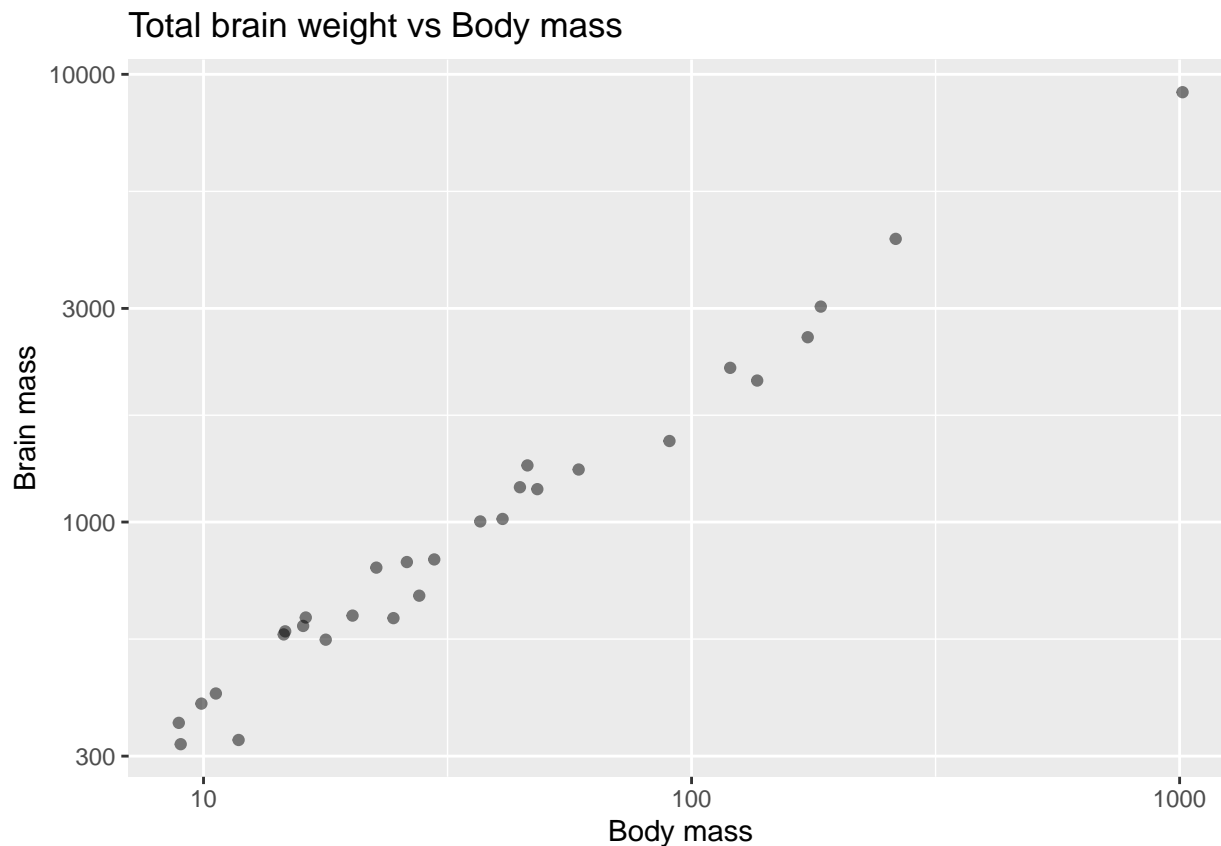
Study of the relationship between brain weight and body mass

```
phyto=myData[(myData$Diet==1),]
```



The initial graph shows a linear relationship between the brain mass and body mass. The data also appear

to be clustered as 95% of the data points have a body mass below 250. In the following graph, we try to observe a more spreadout representation by taking the logarithm of the data.

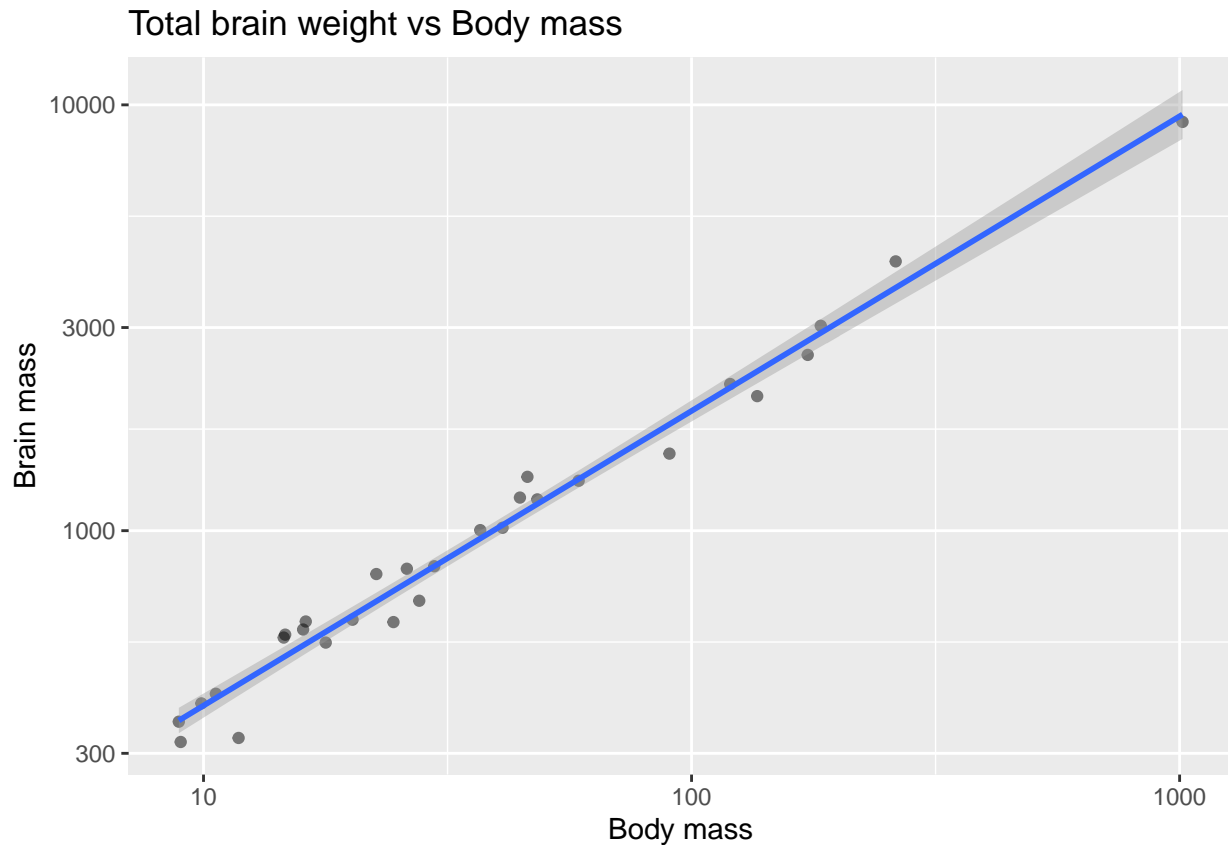


We can also include fit regression lines to both graphs as below:

```
## `geom_smooth()` using formula 'y ~ x'
```



```
## `geom_smooth()` using formula 'y ~ x'
```



```
reg1 = lm(BRW ~ BOW, data=phyto)
```

Mathematical form

In general, the expression of the model estimated for linear regression is: $Y = \hat{\beta}_1 + \hat{\beta}_2 X + \epsilon$. Specifically, R computes

$$BRW = \hat{\beta}_1 + \hat{\beta}_2 BOW + \epsilon$$

```
summary(reg1)
```

```
##
## Call:
## lm(formula = BRW ~ BOW, data = phyto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -628.32 -233.94  -65.74  158.26 1308.59
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  623.4469    81.4762   7.652 3.14e-08 ***
## BOW           8.9999     0.3972  22.659 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 396.9 on 27 degrees of freedom
## Multiple R-squared:  0.95, Adjusted R-squared:  0.9482
```

```
## F-statistic: 513.4 on 1 and 27 DF, p-value: < 2.2e-16
```

```
anova(reg1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: BRW
```

```
##           Df    Sum Sq Mean Sq F value    Pr(>F)
```

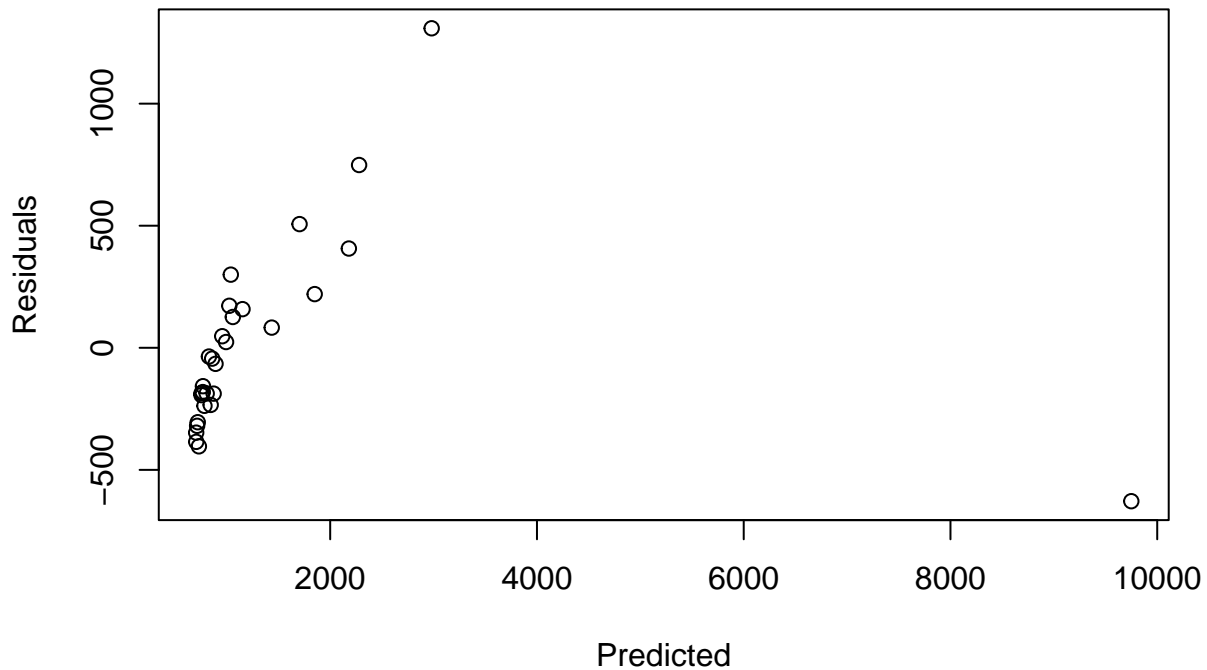
```
## B0W         1 80888380 80888380  513.42 < 2.2e-16 ***
```

```
## Residuals  27  4253838   157550
```

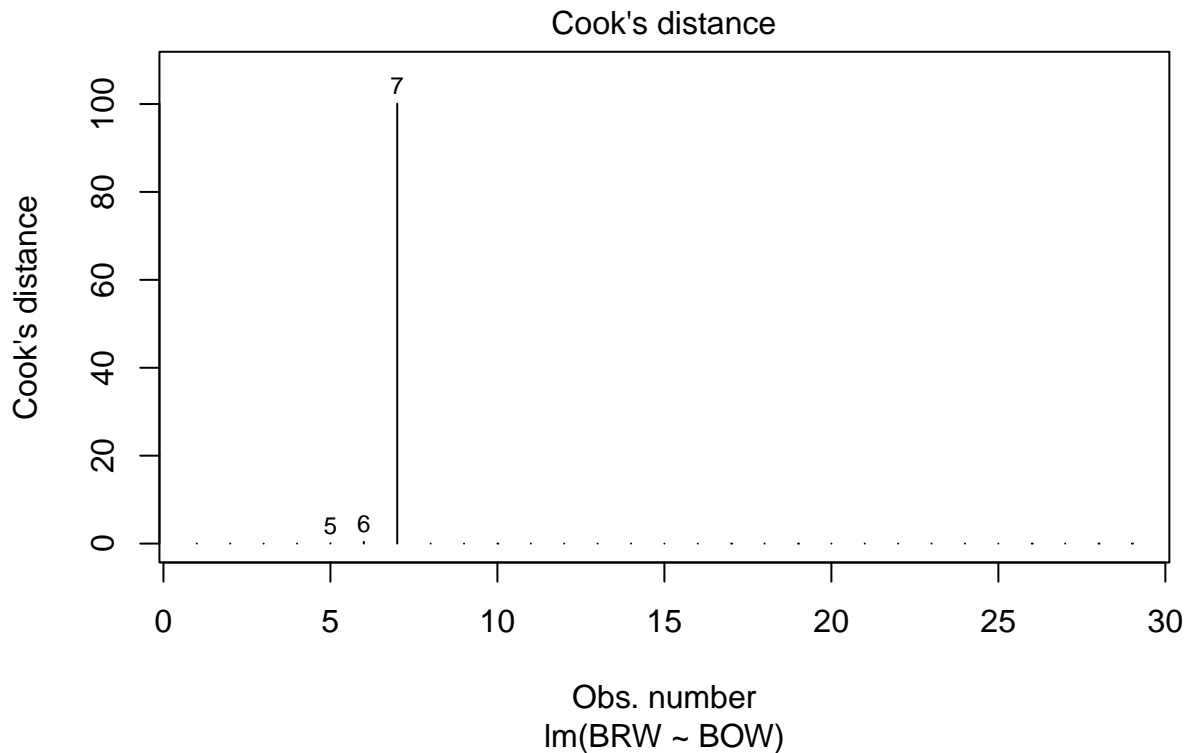
```
## ---
```

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

```
plot(reg1$fitted.values, reg1$residuals, xlab="Predicted", ylab="Residuals")
```



```
plot(reg1,4)
```



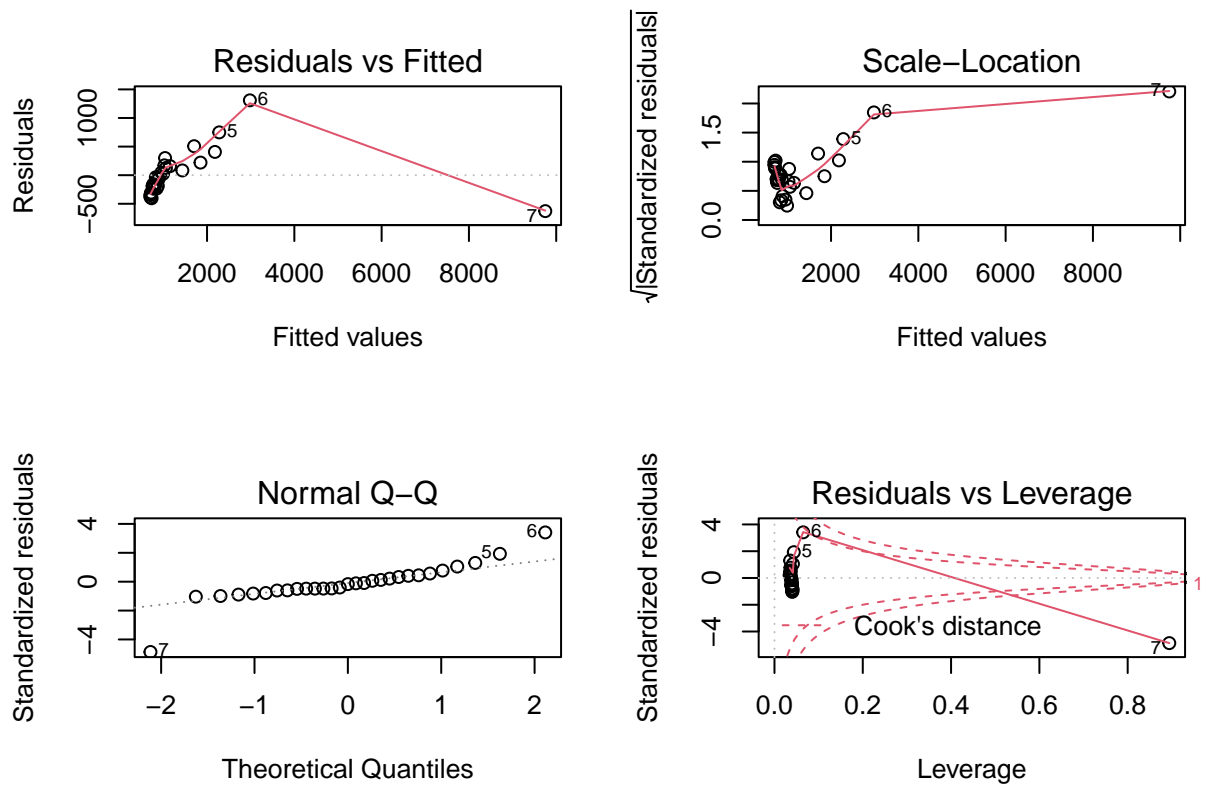
```
which(phyto$BRW>8000)
```

```
## [1] 7
```

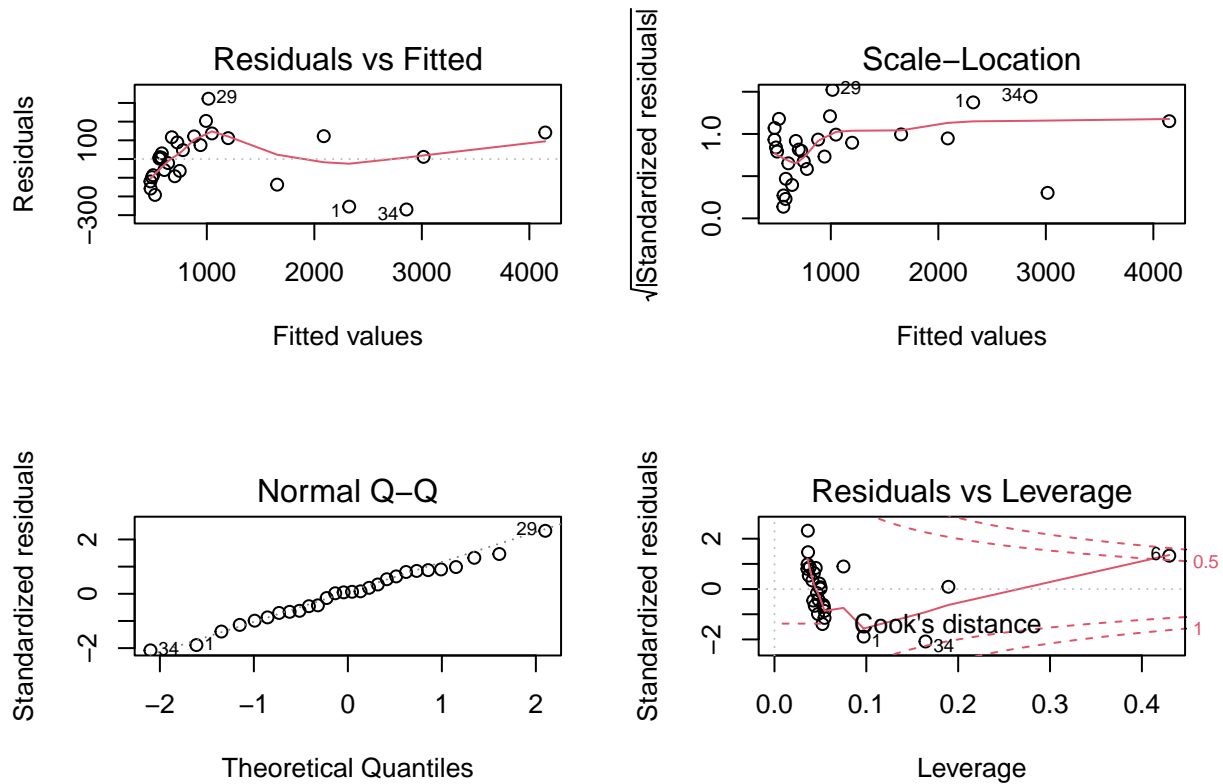
```
phytobis=phyto[which(phyto$BRW<8000),]
reg2 = lm(BRW ~ BOW, data=phytobis)
summary(reg2)
```

```
##
## Call:
## lm(formula = BRW ~ BOW, data = phytobis)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -269.76  -93.33    8.73   112.93   322.55
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  346.5452    35.4920   9.764 3.48e-10 ***
## BOW          14.5099     0.4285  33.860 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 141.8 on 26 degrees of freedom
## Multiple R-squared:  0.9778, Adjusted R-squared:  0.977
## F-statistic: 1147 on 1 and 26 DF, p-value: < 2.2e-16

par(mfcol=c(2,2))
plot(reg1)
```



```
par(mfcol=c(2,2))
plot(reg2)
```



By observing diagnostic graphs 2 and 3 for reg1 and reg2, we can see that in the case of graph 2 (Scale-Location),

the data points are grouped describe a pattern.

graph 3 comparison: More symmetric wrt zero (reg2). reg1 not that symmetric.

3. Study of the contribution to the total weight of each part of the brain