

# KJPipho\_PS\_9

Krista Pipho

November 8, 2018

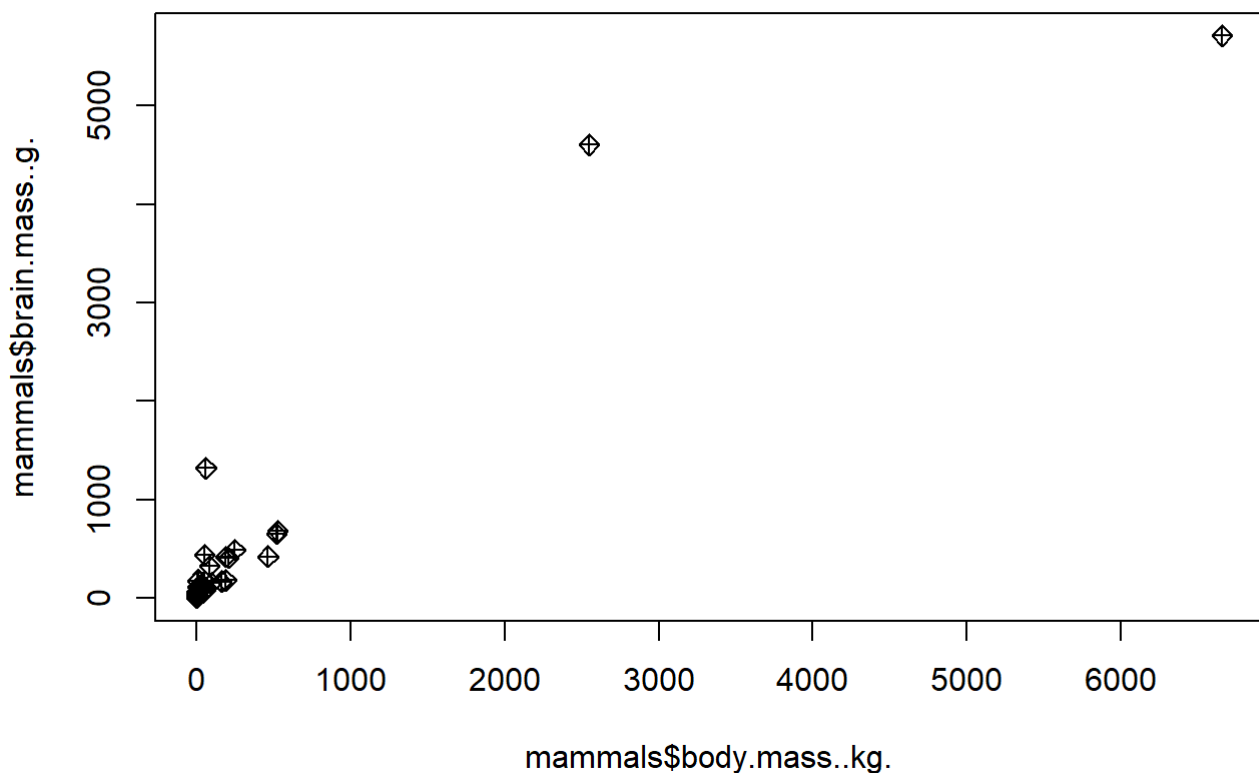
Question 1.

Do mammals obey a common law of allometric scaling for brain size?

a.

```
# First we read in the mammal data set and save it as "mammal"
mammals<-read.csv( "C:\\Users\\xenon\\Desktop\\R Studio 2018\\mammals.csv")
# Now we look over the data
#mammals
# Next we plot brain size against body size to see if they have a linear relationship to each other
plot(mammals$body.mass..kg.,mammals$brain.mass..g., main= "Scaling of brain size with body size", pch=9)
```

**Scaling of brain size with body size**



The relationship of the two variables is probably not linear, but its hard to tell visually with all of the points clustered at the bottom.

b. What if we log transform the data?

```
# Here we log transform both sets of weights
log_br = log(mammals$brain.mass..g.)
log_bo = log(mammals$body.mass..kg.)
# Here we formally test the normality of each transformed variable
shapiro.test(log_bo)
```

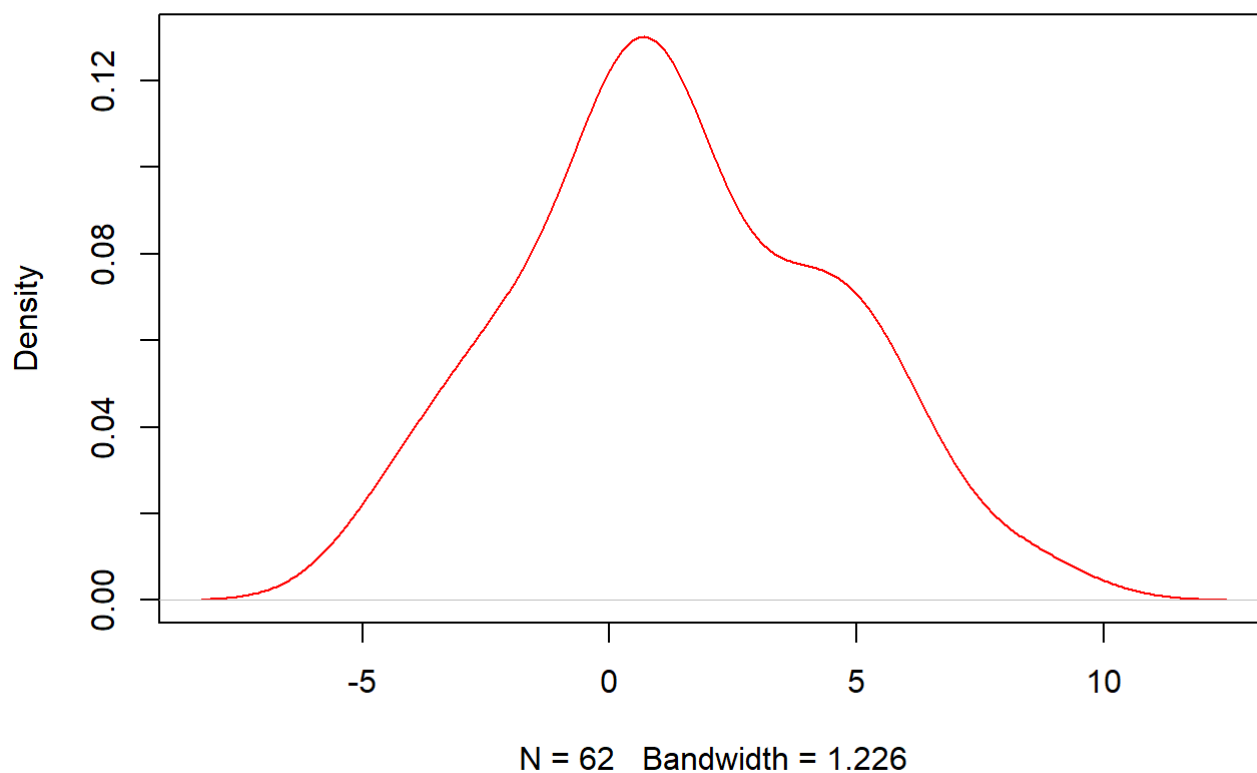
```
##
## Shapiro-Wilk normality test
##
## data: log_bo
## W = 0.98248, p-value = 0.5192
```

```
shapiro.test(log_br)
```

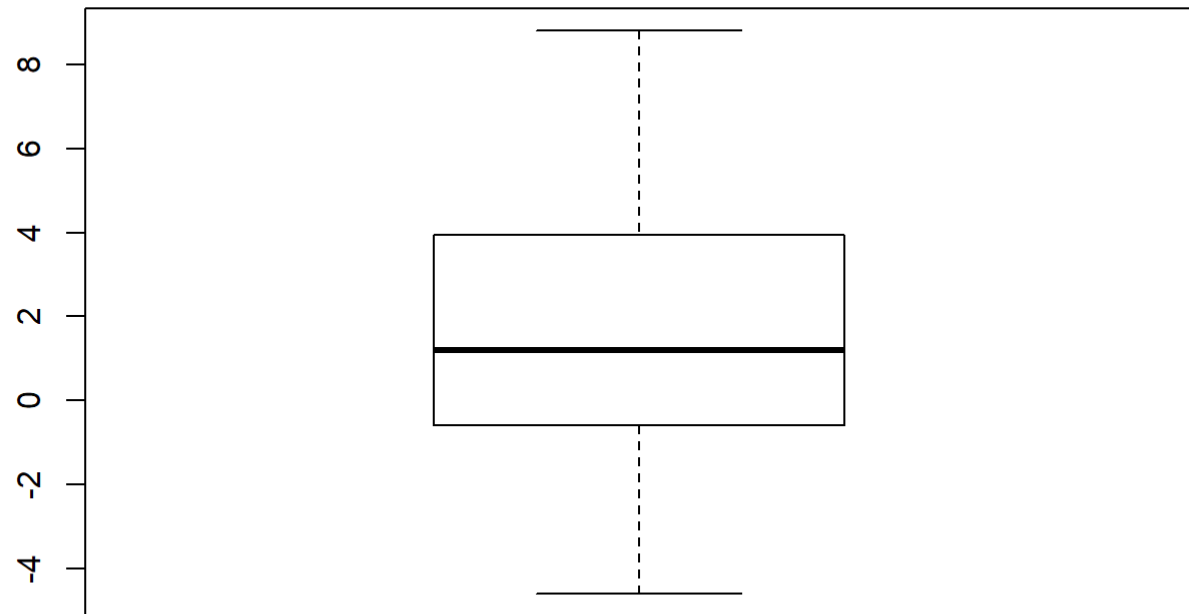
```
##
## Shapiro-Wilk normality test
##
## data: log_br
## W = 0.98589, p-value = 0.697
```

```
# We can visualize the new shape of the distribution using a density plot
plot(density(log_bo),col="red")
```

**density.default(x = log\_bo)**

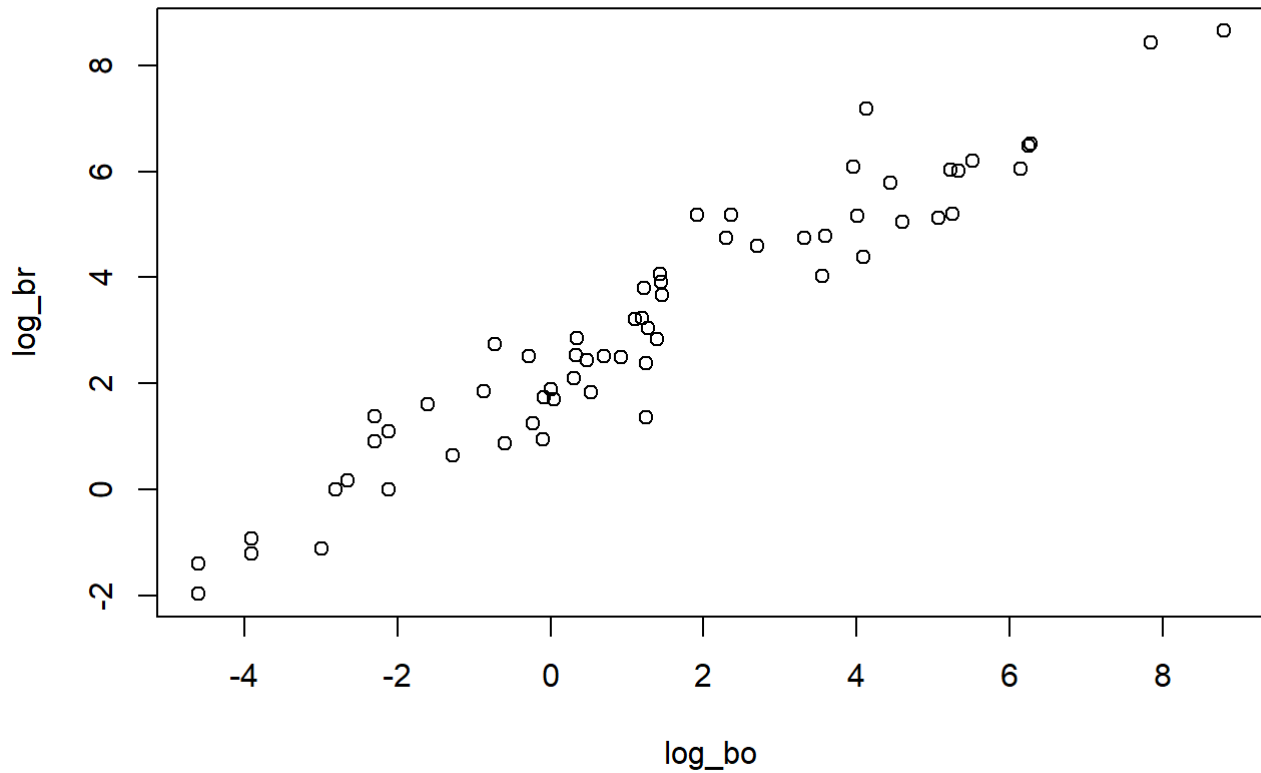


```
# This appears normal enough
# We check for outliers using a boxplot
boxplot(log_bo)
```



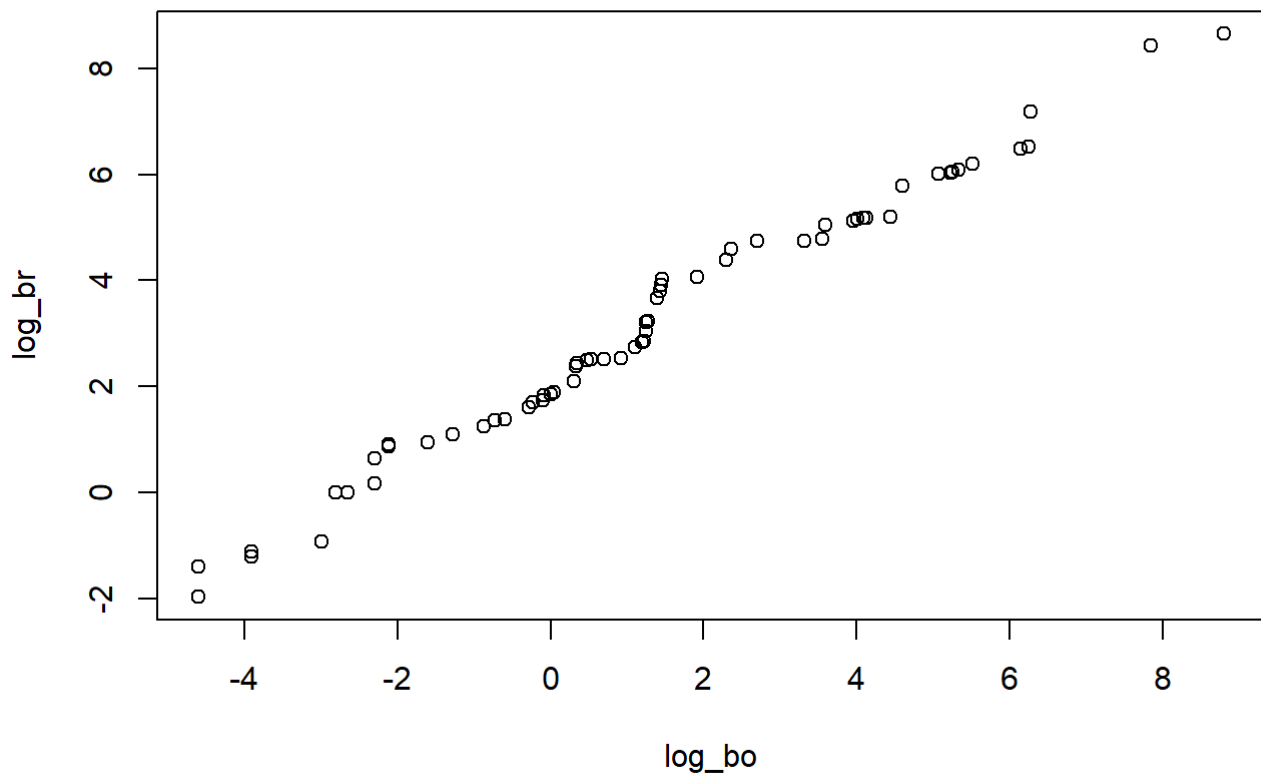
```
# With this transformation there are no outliers in the proposed predictor variable  
# Now we can replot with the transformed data  
plot(log_bo, log_br, main= "Log transformed values")
```

## Log transformed values



```
# We can also visualize this even better with a qq plot  
qqplot(log_bo, log_br, main = "qq plot of Body mass vs Brain mass")
```

## qq plot of Body mass vs Brain mass



The correspondance between brain and body size certainly looks linear when they are both log transformed!

c. Now we will examine the correaltion between the two variables

```
# Here we will use a correlation test to get a preliminary idea of the correlational relation  
ship between thesetransformed variables  
cor.test(log_bo,log_br)
```

```
##  
## Pearson's product-moment correlation  
##  
## data: log_bo and log_br  
## t = 26.216, df = 60, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9326419 0.9751927  
## sample estimates:  
## cor  
## 0.9590134
```

```
cor.test(log_bo,log_br, method = "spearman")
```

```
## Warning in cor.test.default(log_bo, log_br, method = "spearman"): Cannot  
## compute exact p-value with ties
```

```
##  
## Spearman's rank correlation rho  
##  
## data: log_bo and log_br  
## S = 1847.7, p-value < 2.2e-16  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho  
## 0.9534717
```

```
# The two seem to have a very strong (.96 out of a totoal of 1 possible) positive correlation
```

With pearsons, the correlatin found is slightly higher (.959) than with spearman (.953). Because we meet the assumptions for pearson we can use this more accurate method, while spearman is designed for dealing with various violations and would be expected to give a more conservative estimate of the correlation.

d. Now we will make a linear model for this correlation using regression

```
line_mod_bobr <- lm(log_br~log_bo)  
print(line_mod_bobr)
```

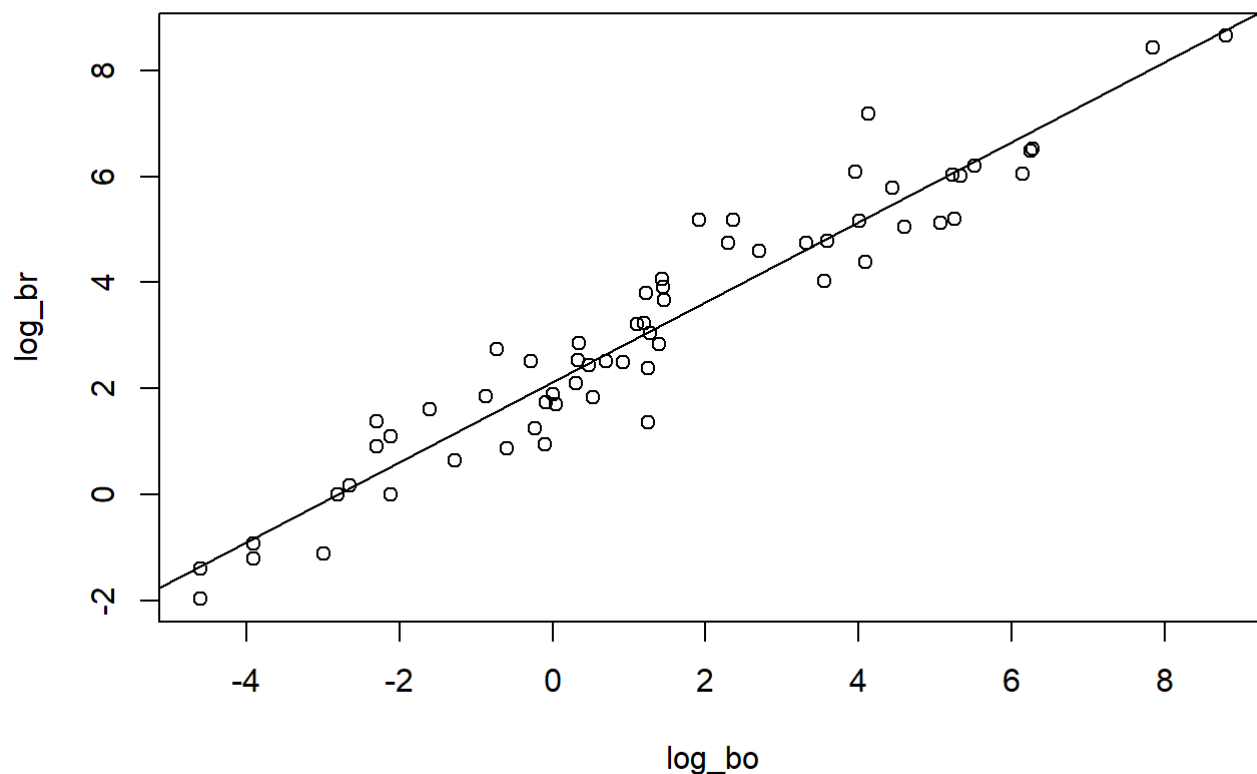
```
##
## Call:
## lm(formula = log_br ~ log_bo)
##
## Coefficients:
## (Intercept)      log_bo
##      2.1272      0.7545
```

```
summary(line_mod_bobr)
```

```
##
## Call:
## lm(formula = log_br ~ log_bo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71143 -0.50667 -0.05606  0.43833  1.94425
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.12719    0.09682   21.97  <2e-16 ***
## log_bo       0.75451    0.02878   26.22  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.699 on 60 degrees of freedom
## Multiple R-squared:  0.9197, Adjusted R-squared:  0.9184
## F-statistic: 687.3 on 1 and 60 DF,  p-value: < 2.2e-16
```

```
plot(log_bo,log_br,main= "Log transformed values with line")
abline(line_mod_bobr)
```

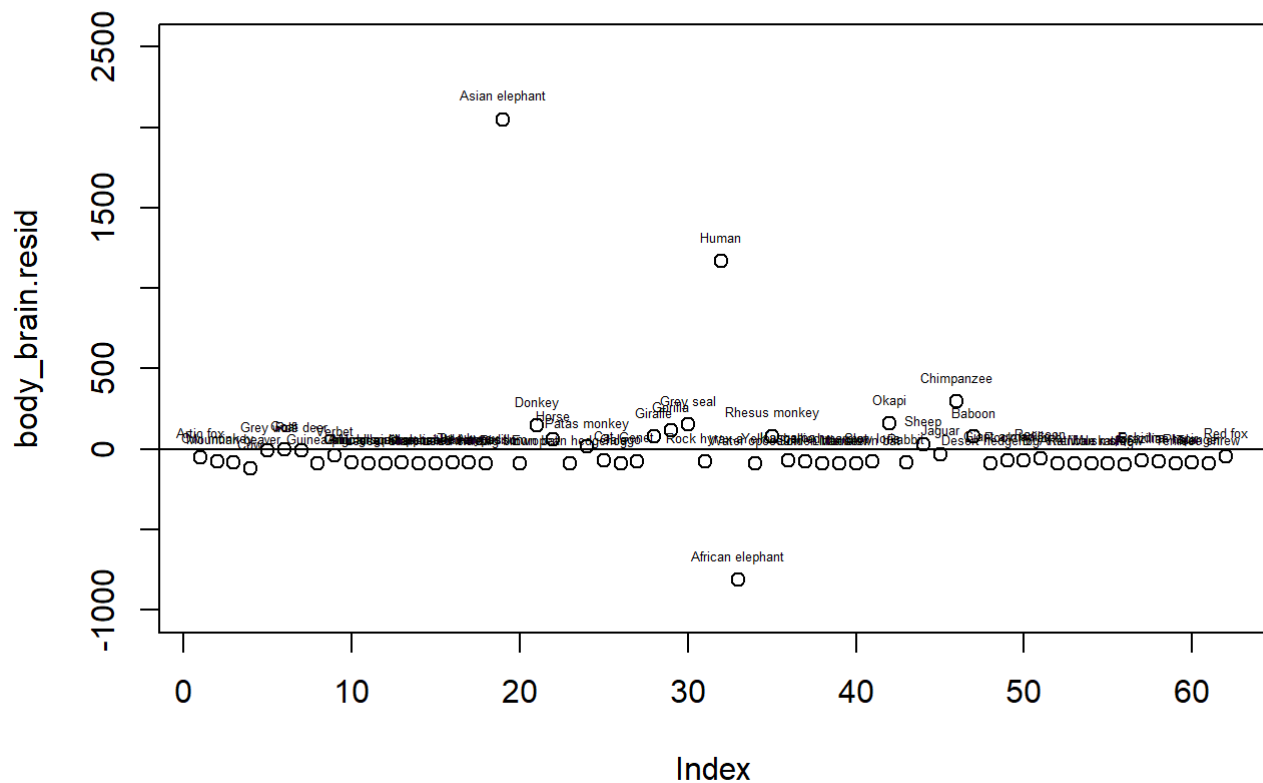
## Log transformed values with line



The line that best describes the correlation between the log of brain and body size in mammals is  $y = .75x + 2.13$

e. Lets look at the residual plots now

```
# We will make a comparasin of brain size to body side
body_brain_lm <-lm(mammals$brain.mass..g.~ mammals$body.mass..kg.)
# Now we will look at the residuals using the resid function
body_brain.resid<-resid(body_brain_lm)
plot(body_brain.resid, ylim = c(-1000,2500))
text(body_brain.resid, labels=mammals$name, cex=0.4, pos=3)
# we expect no slope with residuals so we plot a plain, straight line for comparasin
abline(0,0)
```



From this we can discern that the mammal with the largest brain for its body are Asian elephants, and the mammal with the smallest brain for its body are African elephants.

- f. Humans have a significantly positive residual, thus have brains larger than expected for our body size. The African and Asian elephants are also outliers.