HW3

2025-04-25

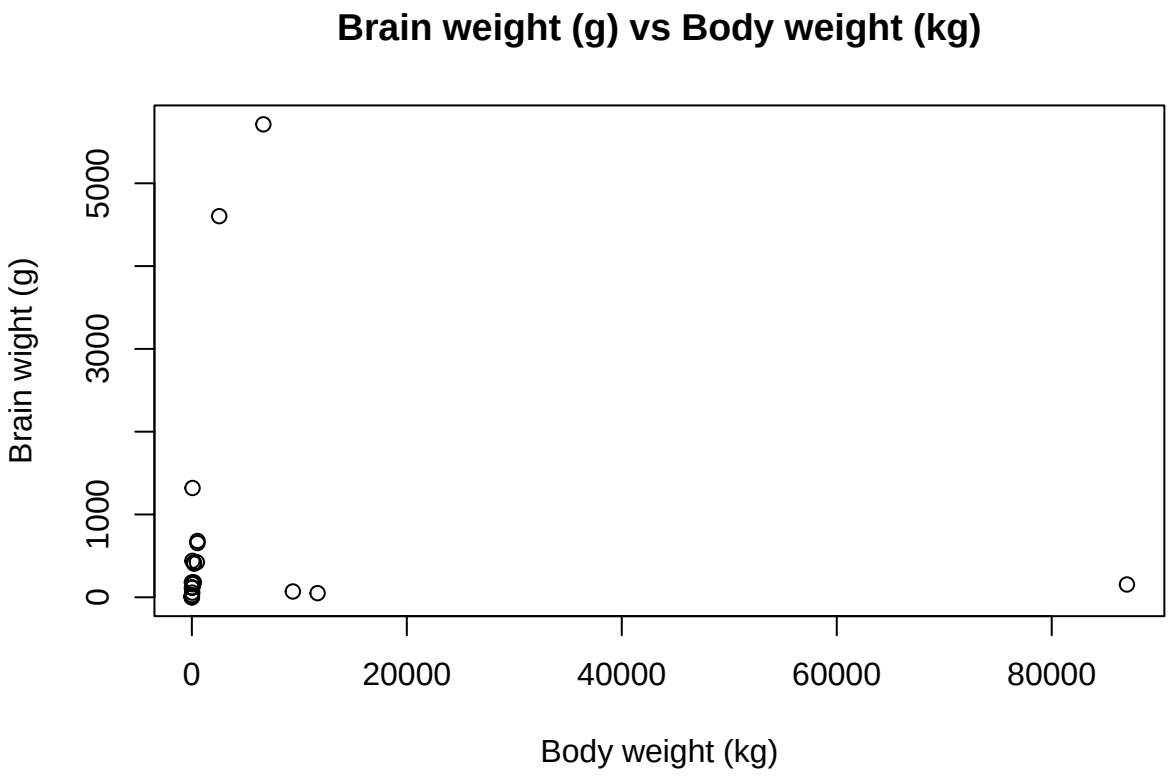
Question 3

```
library(MASS)
dat <- Animals
head(dat, 2)
```

```
##                   body brain
## Mountain beaver   1.35   8.1
## Cow             465.00 423.0
```

a

```
plot(dat$brain ~ dat$body,
     xlab = "Body weight (kg)",
     ylab = "Brain wight (g)",
     main = "Brain weight (g) vs Body weight (kg)")
```

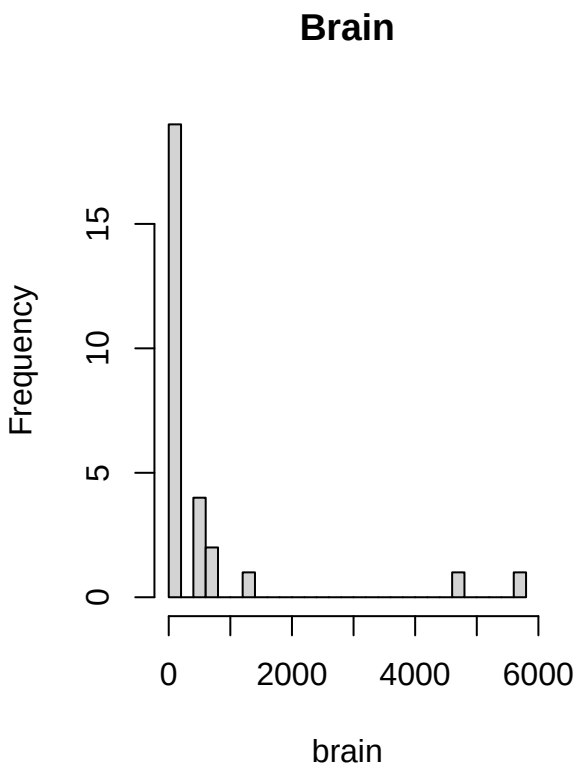


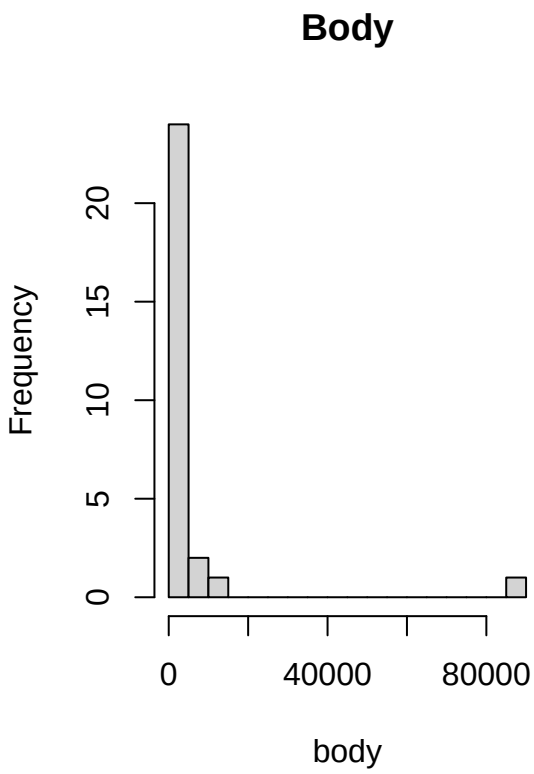
b

```
par(mfrow = c(1, 2))

hist(dat$brain,
     main = "Brain",
     xlab = "brain",
     breaks = 25)

hist(dat$body,
     main = "Body",
     xlab = "body",
     breaks = 25)
```





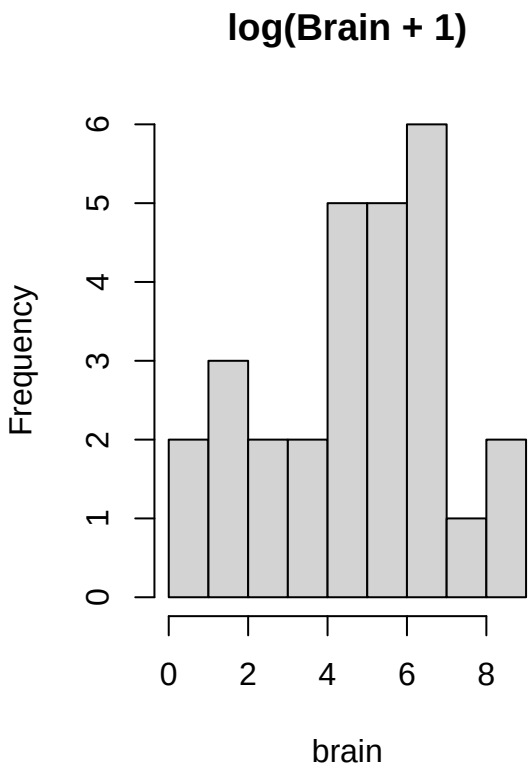

```
par(mfrow = c(1, 1))
```

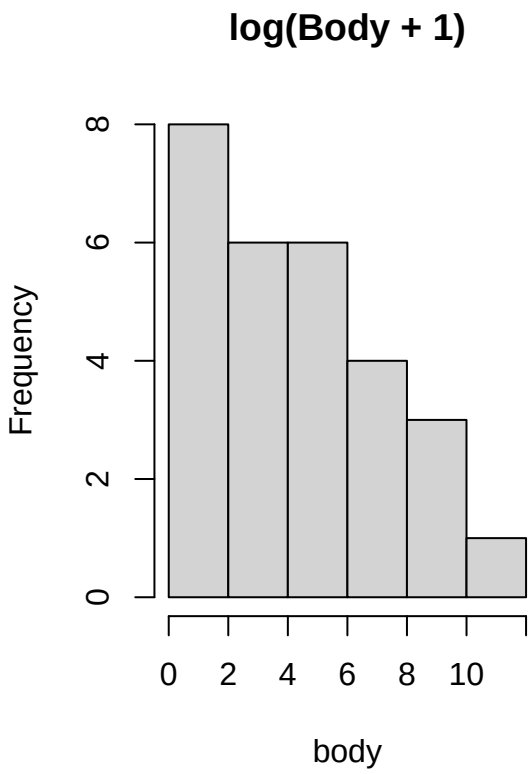
It seems preferable to apply a log transformation.

```
par(mfrow = c(1, 2))
log_brain = log(dat$brain + 1)
log_body = log(dat$body + 1)

hist(log_brain,
     main = "log(Brain + 1)",
     xlab = "brain")

hist(log_body,
     main = "log(Body + 1)",
     xlab = "body")
```



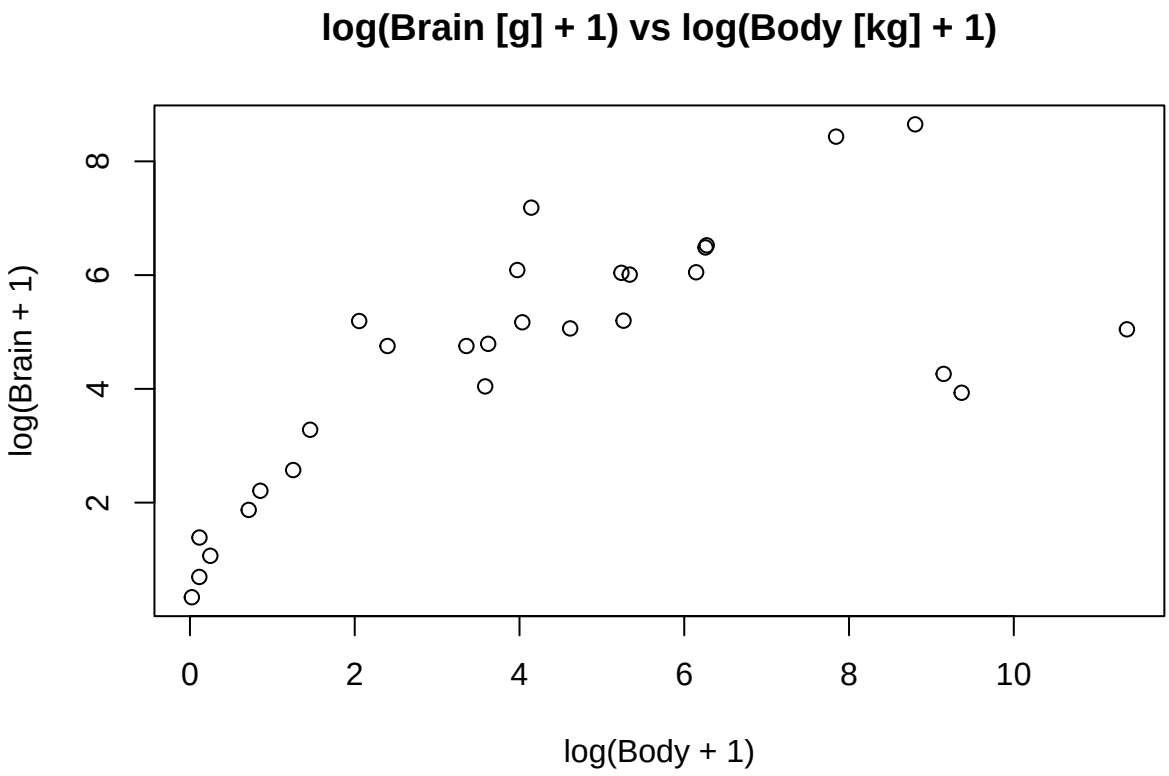



```
par(mfrow = c(1, 1))
```

c

```
log_body = log(dat$body + 1)
log_brain = log(dat$brain + 1)

plot(log_brain ~ log_body,
     xlab = "log(Body + 1)",
     ylab = "log(Brain + 1)",
     main = "log(Brain [g] + 1) vs log(Body [kg] + 1)")
```



The plot after the log transformation appears approximately linear.

d

```
log_lm = lm(log_brain ~ log_body)
coef(log_lm)
```

```
## (Intercept)    log_body
##   2.4854049   0.4889349
```

The regression coefficient for log_body is 0.489, indicating a positive linear relationship between log_body and log_brain.

e

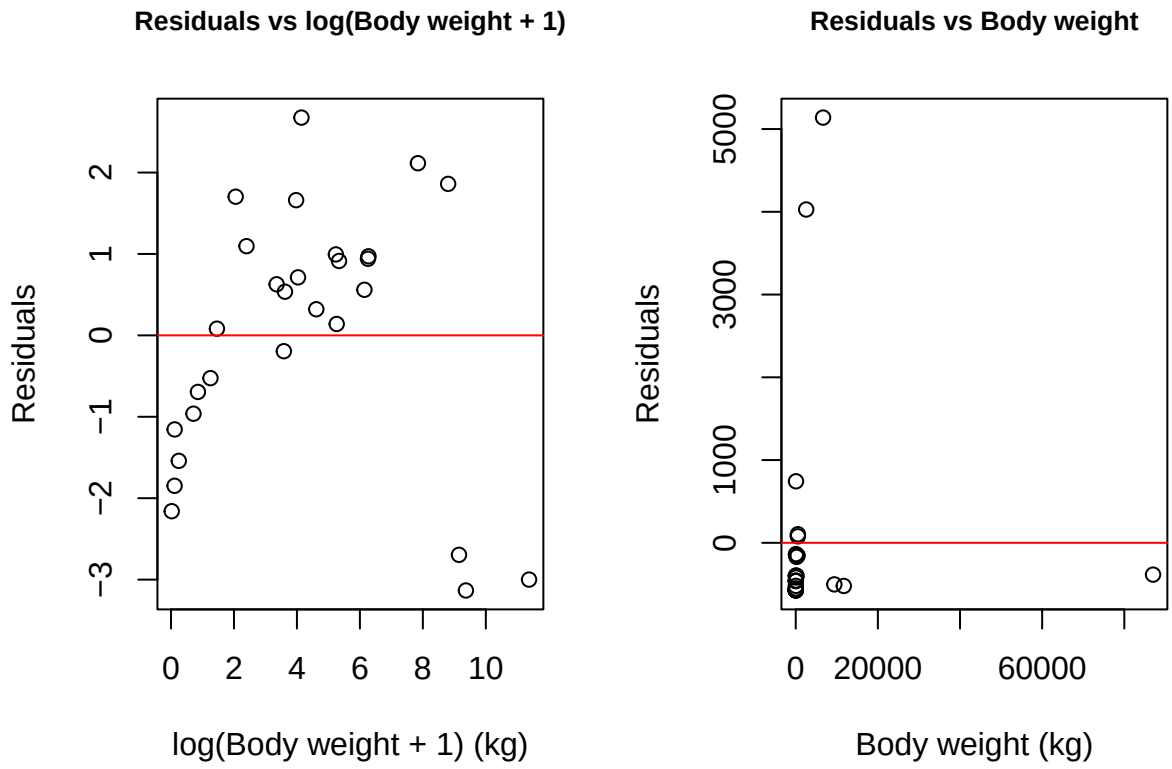
```
par(mfrow = c(1, 2))

plot(log_lm$residuals ~ log_body,
     xlab = "log(Body weight + 1) (kg)",
     ylab = "Residuals",
     main = "Residuals vs log(Body weight + 1)",
     cex.main = 0.8)
```

```r
abline(h = 0, col = "red")

raw_lm = lm(dat$brain ~ dat$body)

plot(raw_lm$residuals ~ dat$body,
     xlab = "Body weight (kg)",
     ylab = "Residuals",
     main = "Residuals vs Body weight",
     cex.main = 0.8)
abline(h = 0, col = "red")
```

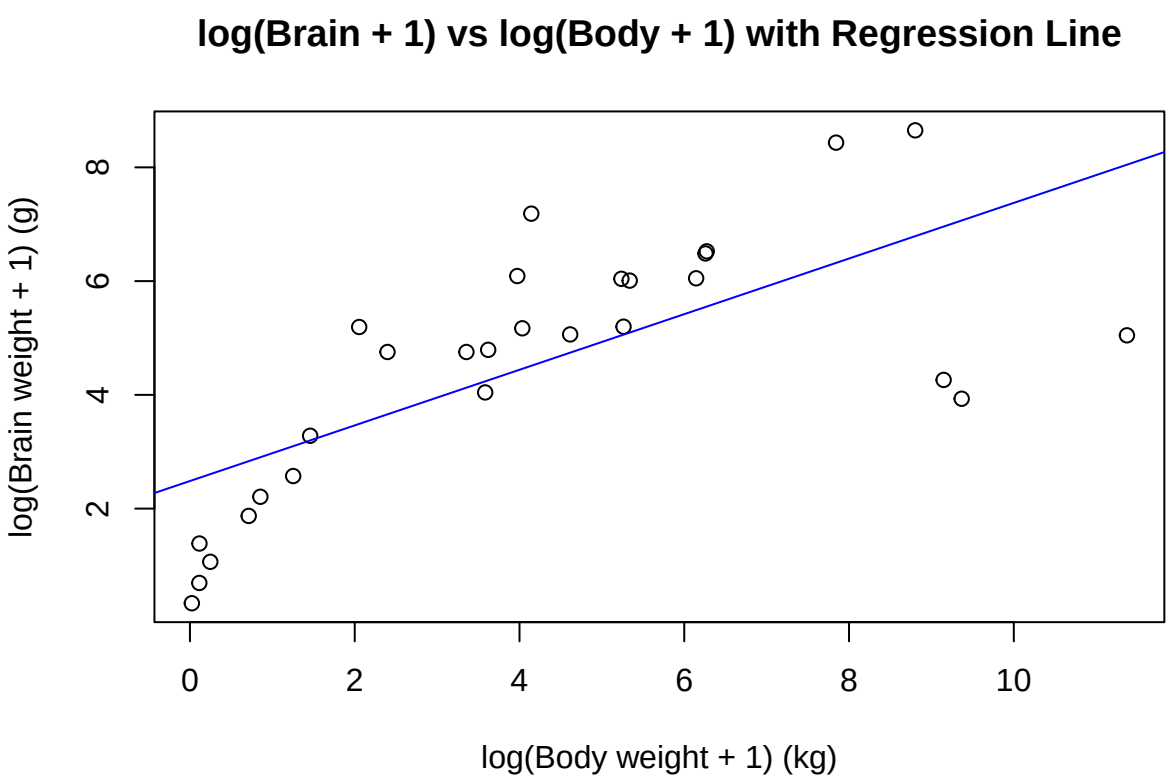



```r
par(mfrow = c(1, 1))
```

The transformed model fits, while the untransformed model does not, due to residual concentration below zero.

f

```r
plot(log_brain ~ log_body,
     xlab = "log(Body weight + 1) (kg)",
     ylab = "log(Brain weight + 1) (g)",
     main = "log(Brain + 1) vs log(Body + 1) with Regression Line")
abline(log_lm, col = "blue")
```


log(Brain + 1) vs log(Body + 1) with Regression Line
log(Brain weight + 1) (g)
2
4
6
8
0
2
4
6
8
10
log(Body weight + 1) (kg)