

Homework 3

```
data = read.csv('sleep.csv', header=TRUE, sep = ',')
```

```
head(data)
```

```
##           Species  BodyWt BrainWt NonDreaming Dreaming TotalSleep
## 1 Africangiantpouchedrat    1.000    6.6         6.3      2.0        8.3
## 2      Asiatic elephant 2547.000 4603.0         2.1      1.8        3.9
## 3           Baboon    10.550   179.5         9.1      0.7        9.8
## 4       Bigbrownbat    0.023    0.3        15.8      3.9       19.7
## 5      Braziliantapir  160.000   169.0         5.2      1.0        6.2
## 6             Cat     3.300   25.6        10.9      3.6       14.5
##  LifeSpan Gestation Predation Exposure Danger
## 1      4.5         42         3         1      3
## 2     69.0        624         3         5      4
## 3     27.0        180         4         4      4
## 4     19.0         35         1         1      1
## 5     30.4        392         4         5      4
## 6     28.0         63         1         2      1
```

```
#Response Variables
```

```
NonDreaming <- data$NonDreaming
```

```
Dreaming <- data$Dreaming
```

```
#Continuous Variables
```

```
BodyWt <- data$BodyWt
```

```
BrainWt <- data$BrainWt
```

```
LifeSpan <- data$LifeSpan
```

```
Gestation <- data$Gestation
```

```
#Categorical Variables
```

```
Predation <- data$Predation
```

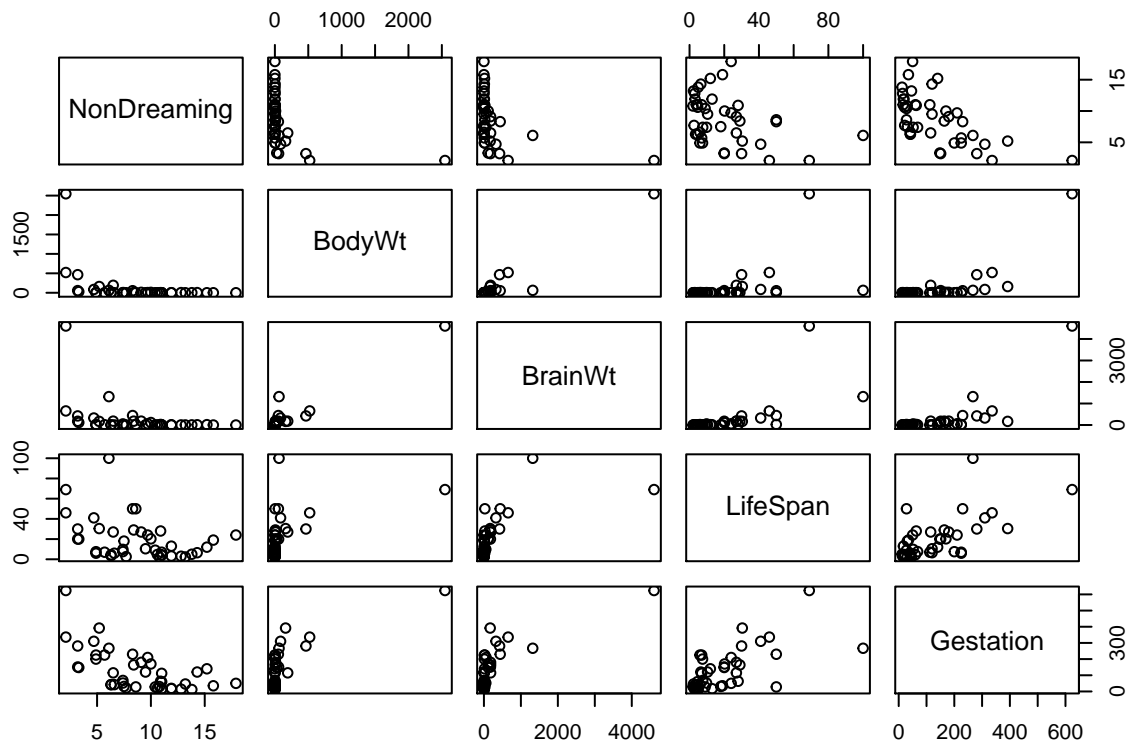
```
Exposure <- data$Exposure
```

```
Danger <- data$Danger
```

Question 1: Exploratory Data Analysis.

1a. Using scatterplots, describe the relationship between NonDreaming and the continuous independent variables.

```
pairs(NonDreaming~BodyWt+BrainWt+LifeSpan+Gestation) #or plot(data)
```



There is a weak negative correlation between non-dreaming and gestation. There is possibly a weak negative correlation between non-dreaming and lifespan that could improve with a transformation. There is not any discernible linear relationship between non-dreaming and body weight or brain weight.

1b. Calculate and interpret the correlation coefficients for continuous variables

```
NonDreamingcor <- cor(data[c(2,3,7,8)],data[4])
```

```
NonDreamingcor
```

```
##          NonDreaming
## BodyWt    -0.3936373
## BrainWt   -0.3867947
## LifeSpan  -0.3722345
## Gestation -0.6061048
```

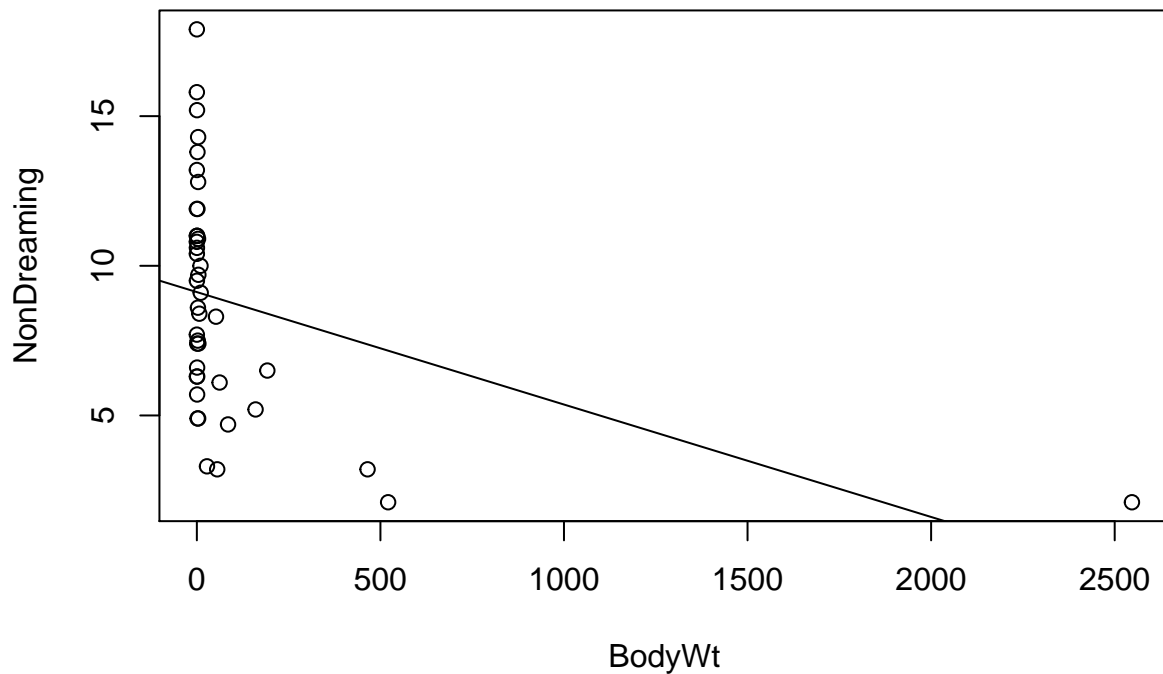
All correlations between non-dreaming and the continuous variables are negative. Negative correlation means that as any of these continuous variables increase the non-dreaming will decrease, on average by some amount. Gestation has the strongest correlation with non-dreaming at -.61. Interesting body weight and brain weight have a stronger correlation than lifespan.

1c. Improving linearity

Using the initial scatterplots, are you able to visually validate the direction and strength of the correlation coefficients? If you see clusters of data points, try adding a directional line (abline) to the scatterplot by individually inspecting each predicting variable. You may need to transform the predicting continuous variable(s) to improve the linearity of the data. You can also transform the predicting variable NonDreaming, to improve linearity, although not required.

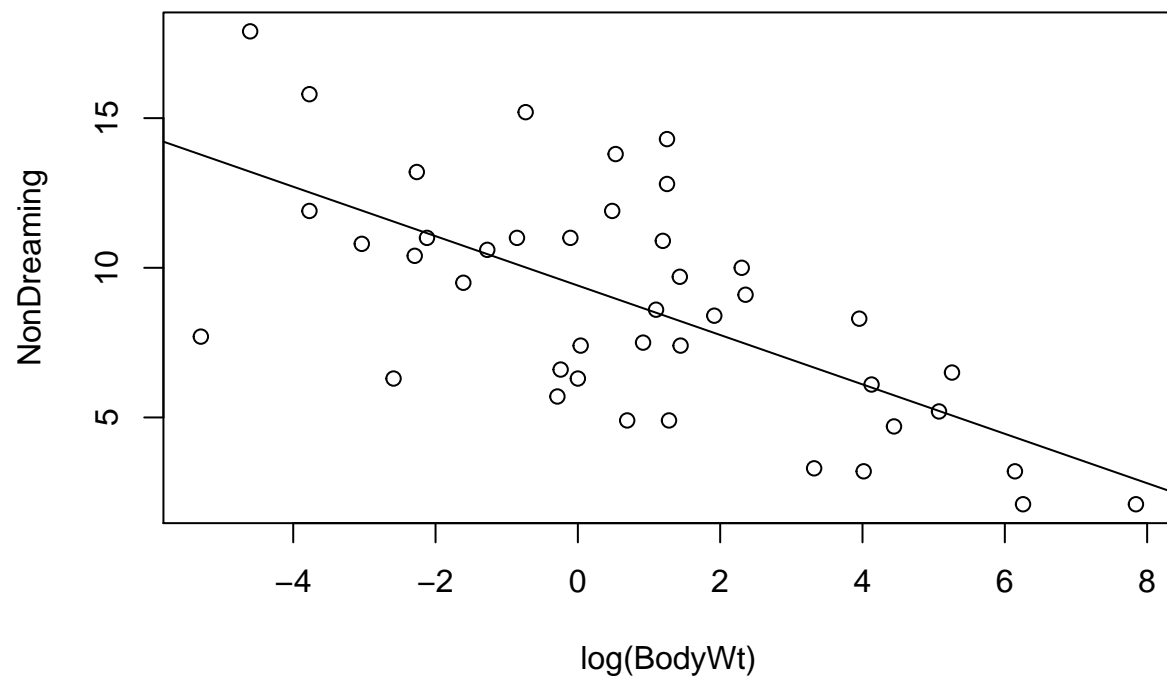
```
#Inspect BodyWt
```

```
plot(BodyWt, NonDreaming)
abline(lm(NonDreaming ~ BodyWt, data = data))
```

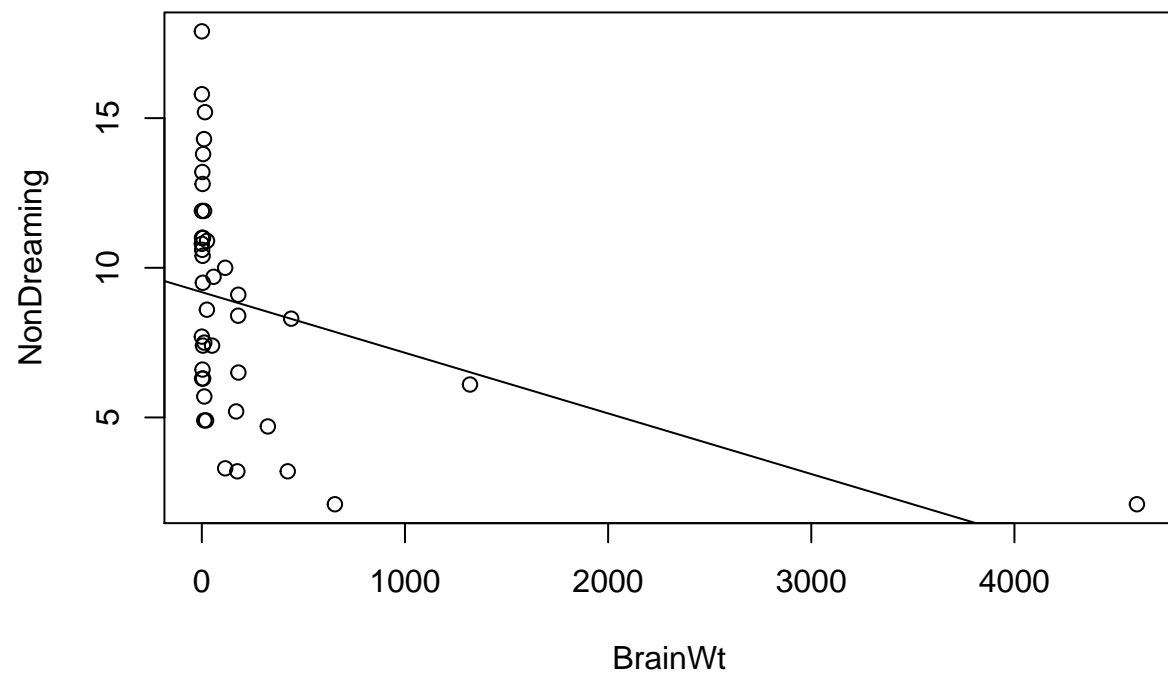


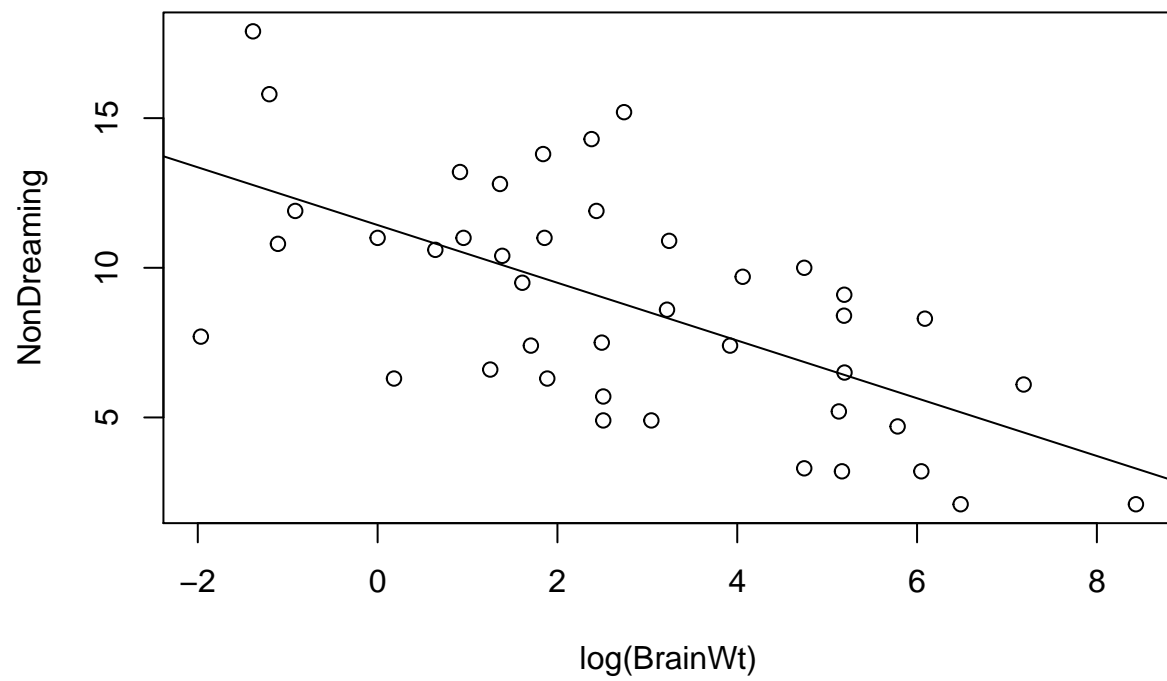
```
# Transform BodyWt
```

```
plot(log(BodyWt), NonDreaming)
abline(lm(NonDreaming ~ log(BodyWt), data = data))
```

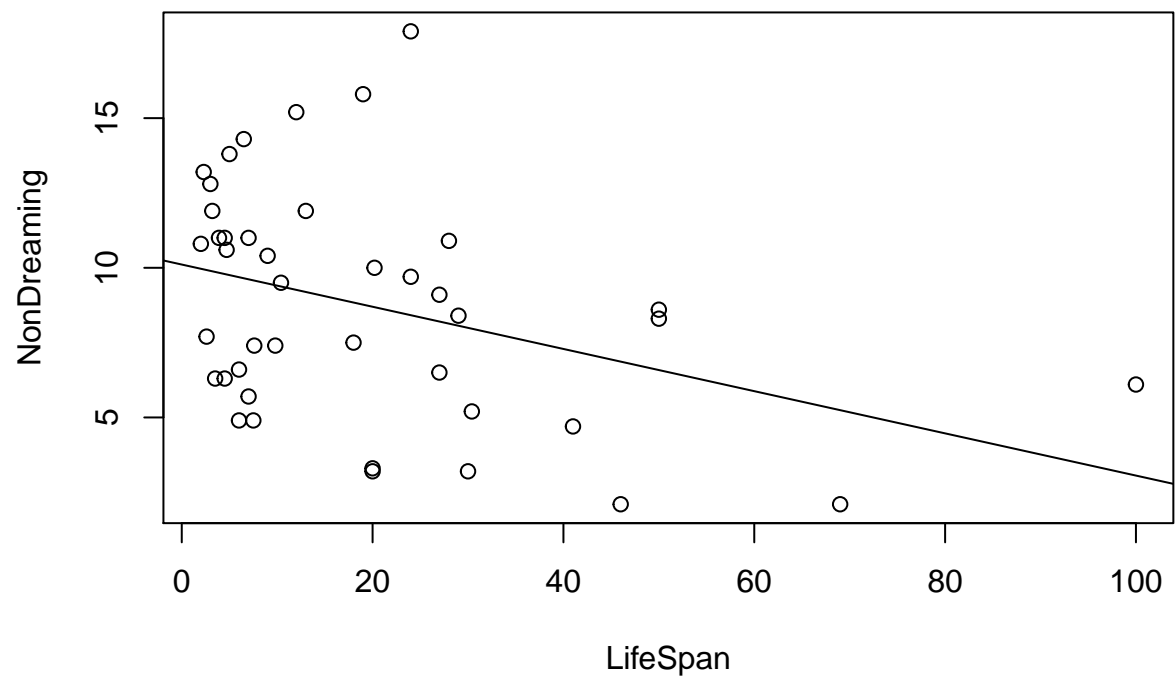


```
#Inspect BrainWt  
plot(BrainWt, NonDreaming)  
abline(lm(NonDreaming ~ BrainWt, data = data))
```

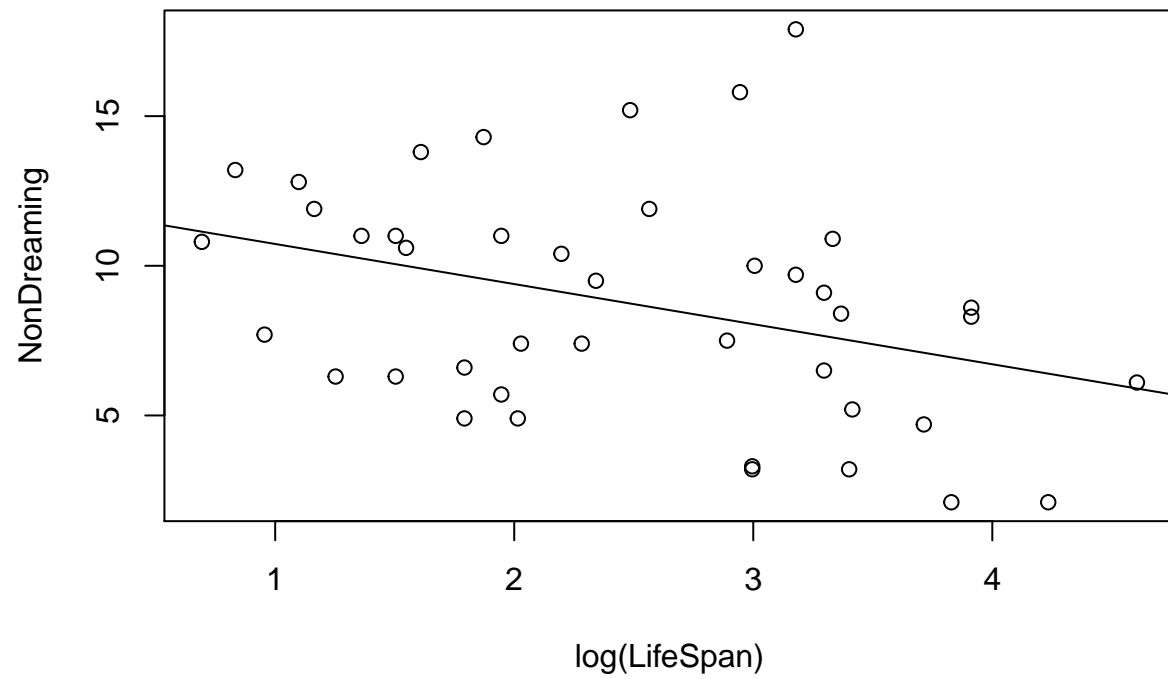




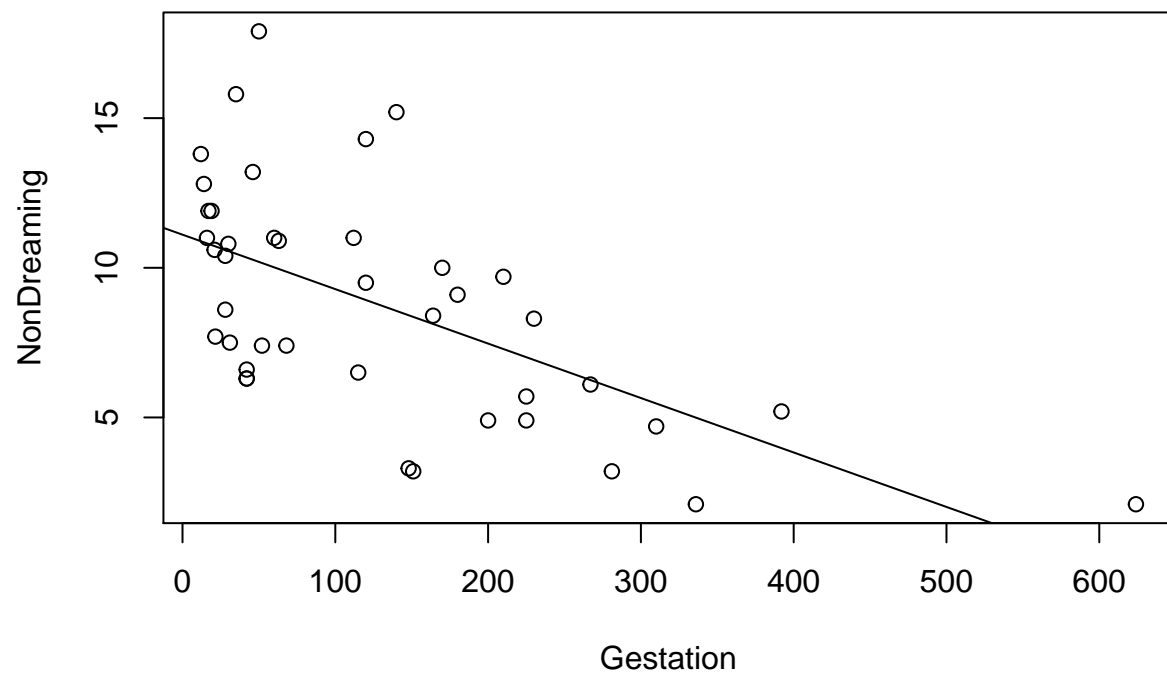
```
#Inspect LifeSpan  
plot(LifeSpan, NonDreaming)  
abline(lm(NonDreaming ~ LifeSpan, data = data))
```



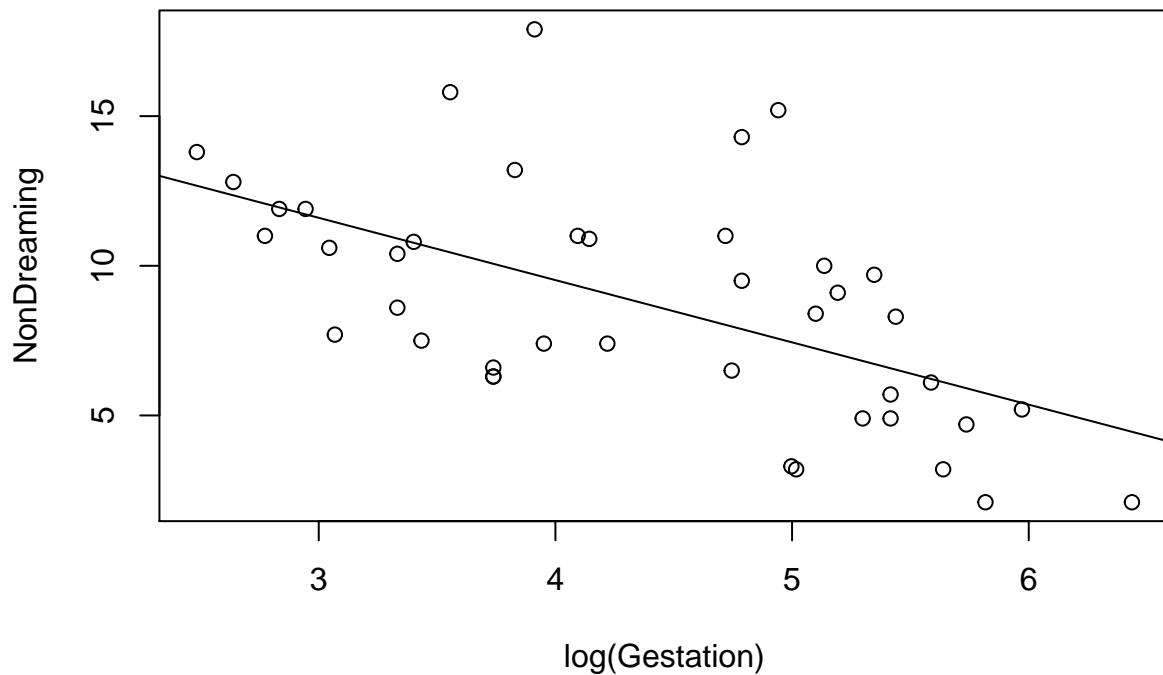
```
# Transform LifeSpan  
plot(log(LifeSpan), NonDreaming)  
abline(lm(NonDreaming ~log(LifeSpan), data = data))
```



```
#Inspect Gestation  
plot(Gestation, NonDreaming)  
abline(lm(NonDreaming ~ Gestation, data = data))
```

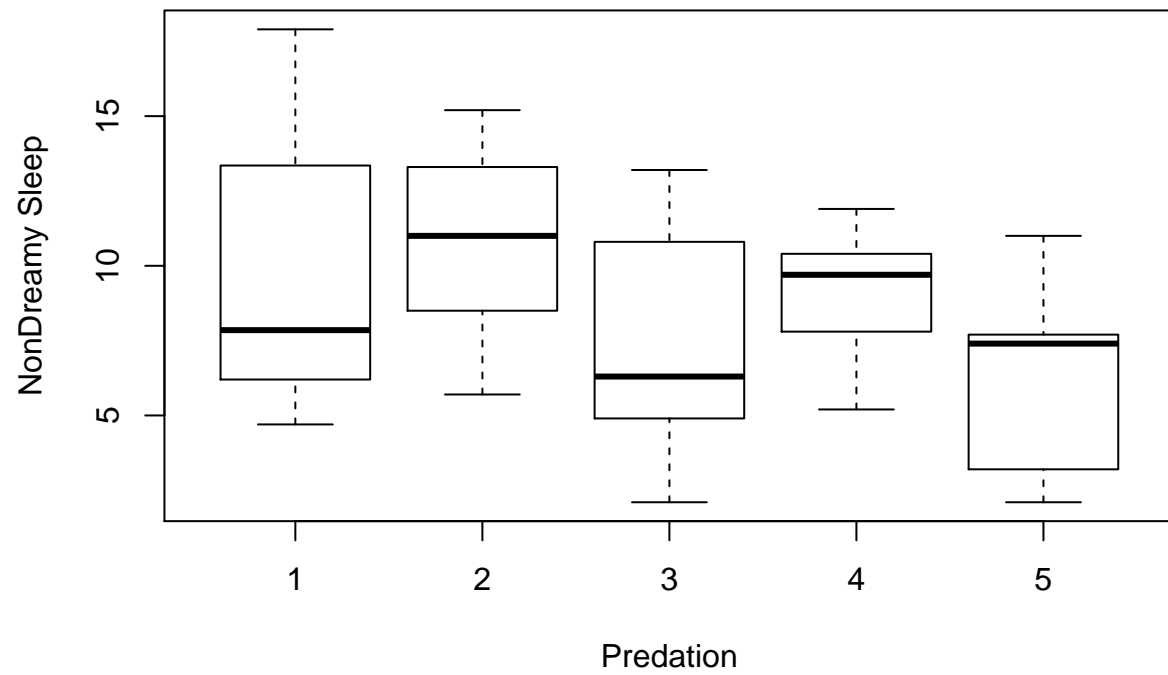
```
# Transform Gestation  
plot(log(Gestation), NonDreaming)  
abline(lm(NonDreaming ~log(Gestation), data = data))
```



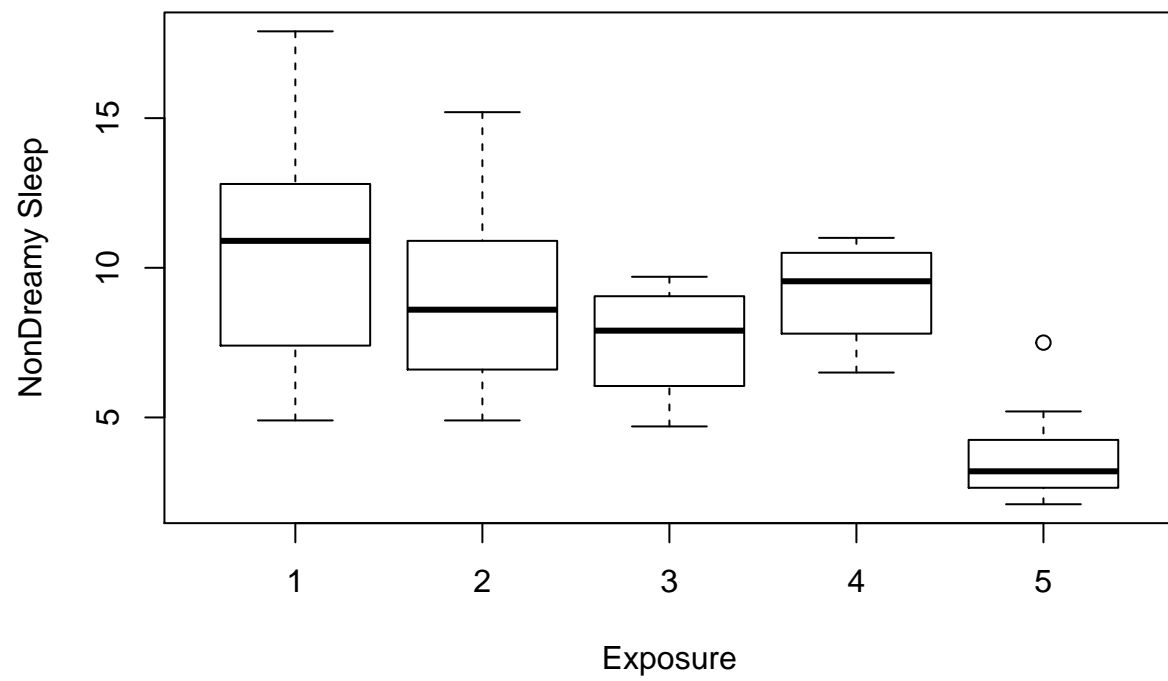
All scatterplots improved after the transformation most dramatically in body and brain weight. This validates the correlation coefficient because all scatterplots show a negative slope as represented by the abline.

1d. Using boxplots, describe the relationship between **NonDreaming** and the categorical independent variables **Predation**, **Exposure**, and **Danger**. Does **NonDreaming** vary with the categorical variables?

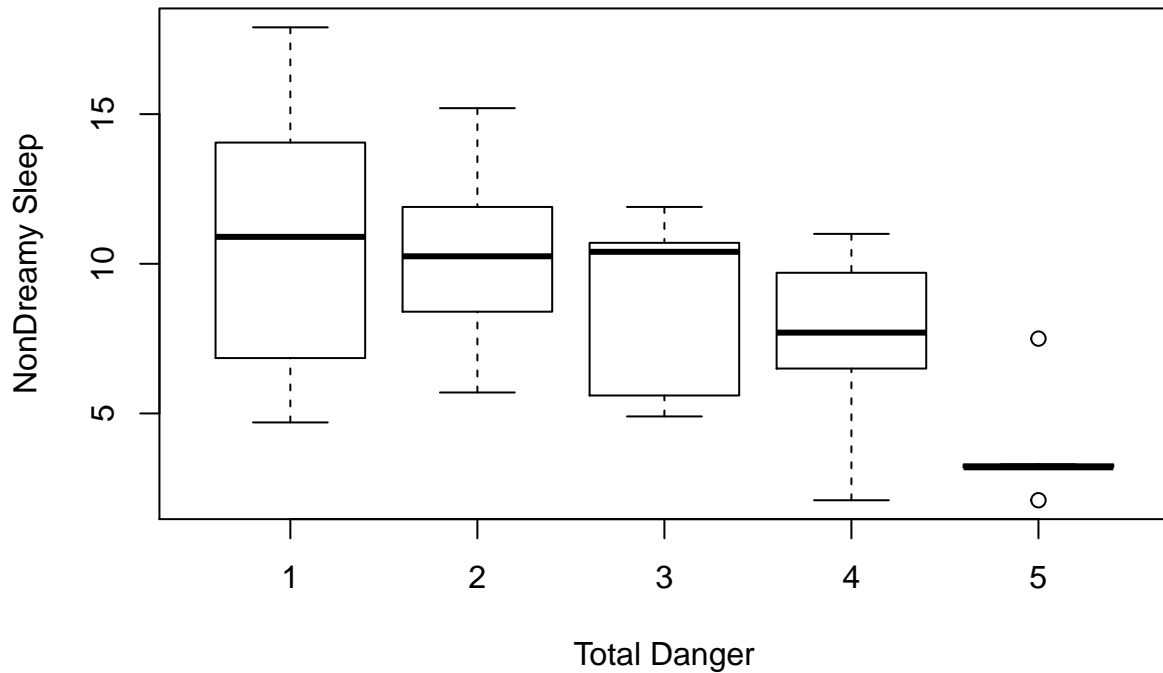
```
boxplot(NonDreaming ~ Predation, xlab = "Predation", ylab = "NonDreamy Sleep")
```



```
boxplot(NonDreaming ~ Exposure, xlab = "Exposure", ylab = "NonDreamy Sleep")
```



```
boxplot(NonDreaming ~ Danger, xlab = "Total Danger", ylab = "NonDreamy Sleep")
```



Predation

predation index (1-5) 1 = minimum (least likely to be preyed upon);

5 = maximum (most likely to be preyed upon)

The animals least likely to be preyed upon have the highest spread and one of the lower median values for sleep. The first and second most preyed upon animals (groups 4 and 5) have the lowest spread and have higher levels of non-dreamy sleep. This seems logical, the most likely to be eaten alive do not rest easy!

Exposure

sleep exposure index (1-5) 1 = least exposed (e.g. animal sleeps in a well-protected den);

5 = most exposed

The animals that are least exposed have the highest spread and have higher levels of non-dreamy sleep, compared to the least exposed that have lower amounts of non-dreamy sleep with a smaller spread. This seems counterintuitive, I would expect animals that have the most exposure to have less dreamy sleep.

Danger

overall danger index (1-5) (based on the above two indices and other information) 1 = least danger (from other animals);

5 = most danger (from other animals)

The animals that are in the most danger get the least amount of non-dreamy sleep with very little spread for group 5. The animals least in danger get the most.

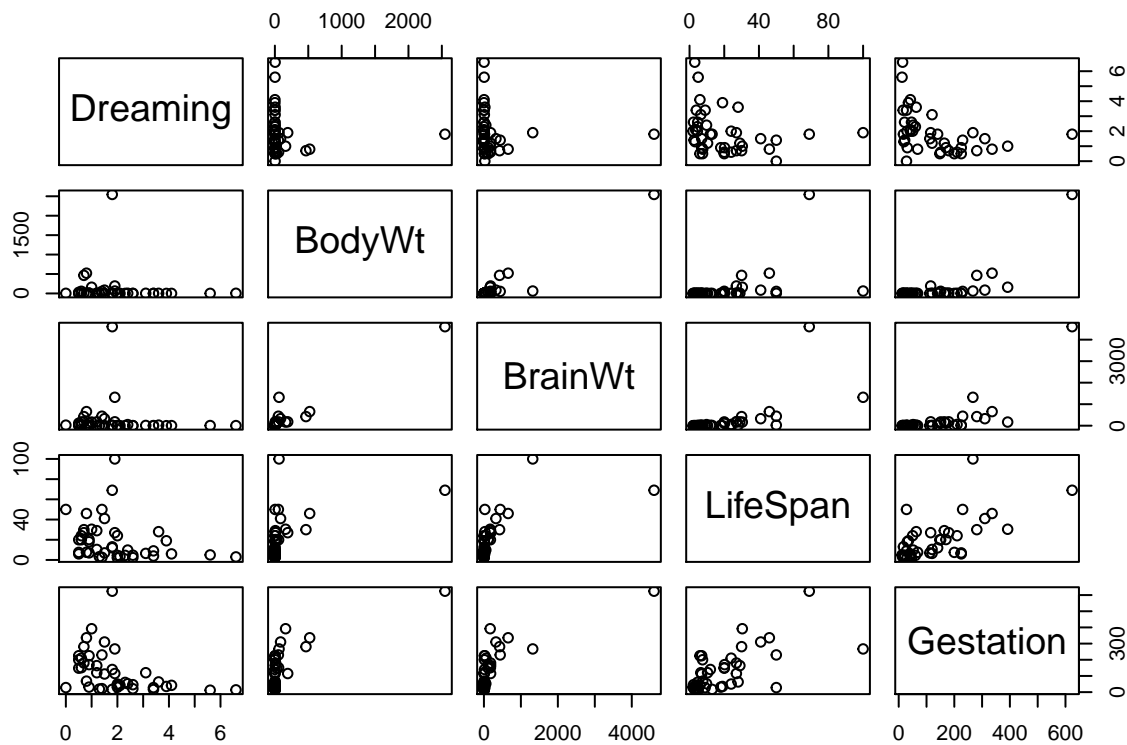
1e. Based on this section for exploratory analysis, is it reasonable to assume a linear regression model? Would you suggest that NonDreaming varies with all or only some of the independent variables? Would you recommend using the categorical variables Predation, Exposure, and Danger in the model? Why?

I think it would be reasonable to assume a linear regression model using the Predation, Exposure and Danger categorical variables. All of the variables seemed to have some level of negative correlation by looking at the boxplots. Therefore, I think a linear regression model would fit this data and I would add them to the model.

Question 2: Exploratory Data Analysis.

2a. Using scatterplots, describe the relationship between Dreaming and the continuous independent vari

```
pairs(Dreaming~BodyWt+BrainWt+LifeSpan+Gestation) #or plot(data)
```



There is a weak negative correlation between dreaming and gestation. There is possibly a weak negative correlation between dreaming and lifespan that could improve with a transformation. There is not any discernible linear relationship between dreaming and body weight or brain weight. This looks very similar to the non-dreaming data.

2b. Calculate and interpret the correlation coefficients for continuous variables

```
Dreamingcor <- cor(data[c(2,3,7,8)],data[5])
```

```
Dreamingcor
```

```
##           Dreaming
## BodyWt      -0.07488845
## BrainWt     -0.07427740
## LifeSpan    -0.26834006
## Gestation  -0.40893177
```

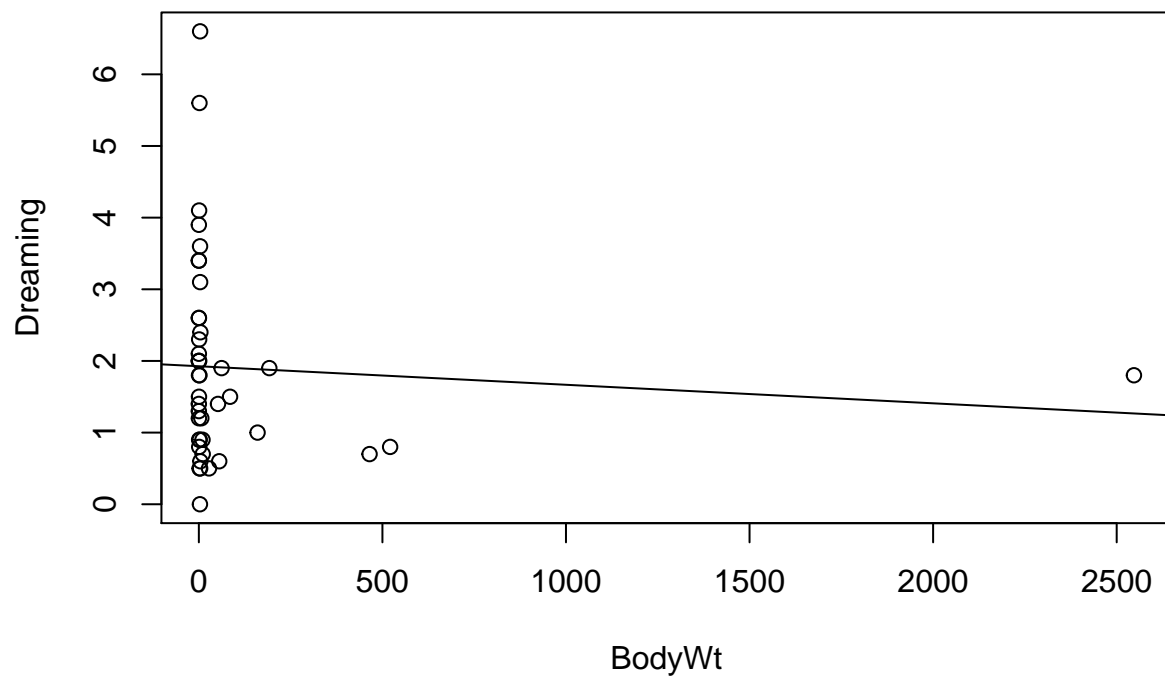
All correlations between dreaming and the continuous variables are negative. Negative correlation means that as any of these continuous variables increase the dreaming will decrease, on average by some amount. Gestation has the strongest correlation with dreaming at -.41, followed by lifespan at -.27 with body and brainweight having a low correlation of -.07.

2c. Improving linearity

Using the initial scatterplots, are you able to visually validate the direction and strength of the correlation coefficients? If you see clusters of data points, try adding a directional line (abline) to the scatterplot by individually inspecting each predicting variable. You may need to transform the predicting continuous variable(s) to improve the linearity of the data. You can also transform the predicting variable NonDreaming, to improve linearity, although not required.

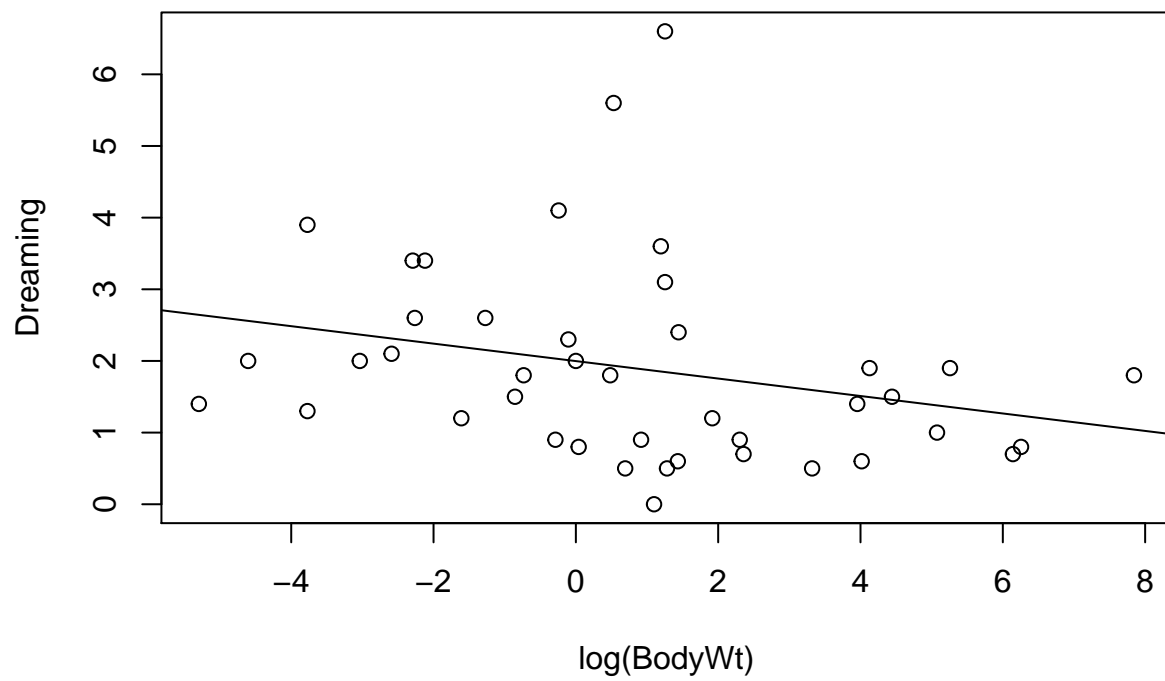
```
#Inspect BodyWt
```

```
plot(BodyWt, Dreaming)
abline(lm(Dreaming ~ BodyWt, data = data))
```

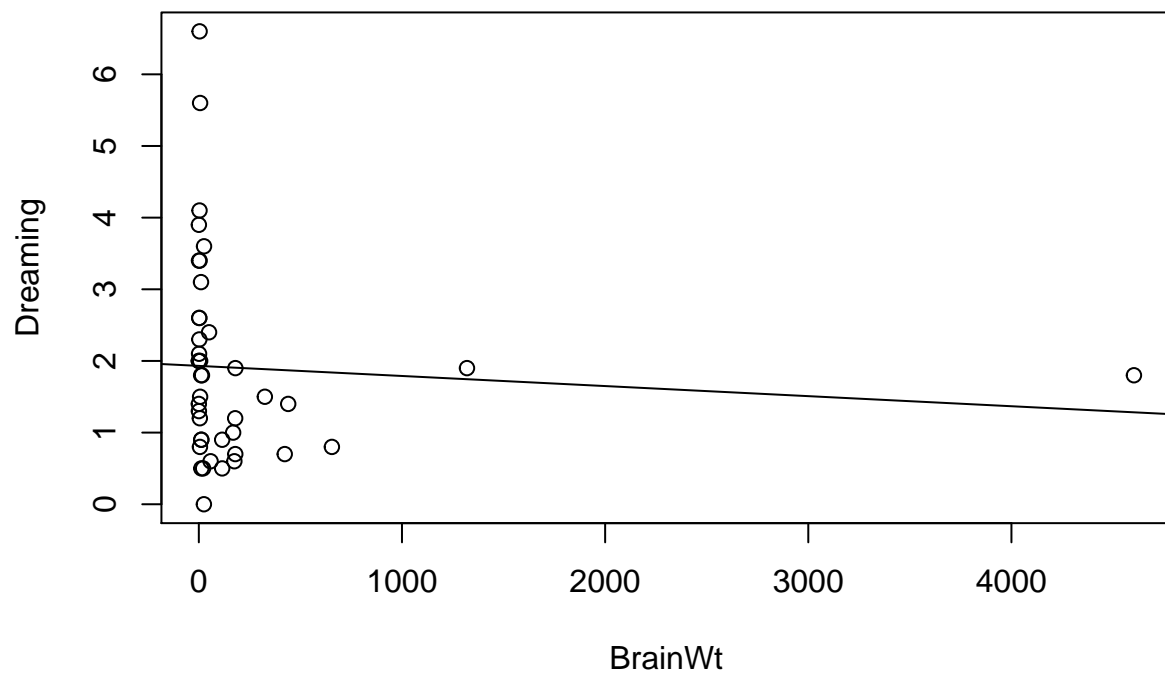


```
# Transform BodyWt
```

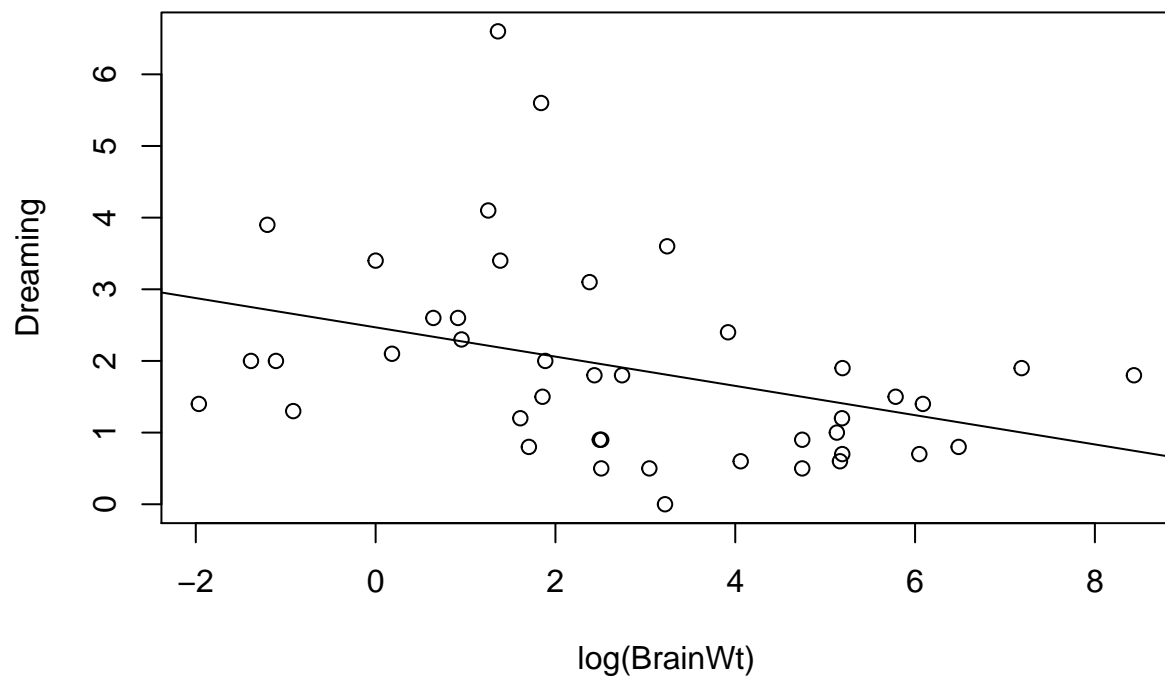
```
plot(log(BodyWt), Dreaming)  
abline(lm(Dreaming ~ log(BodyWt), data = data))
```

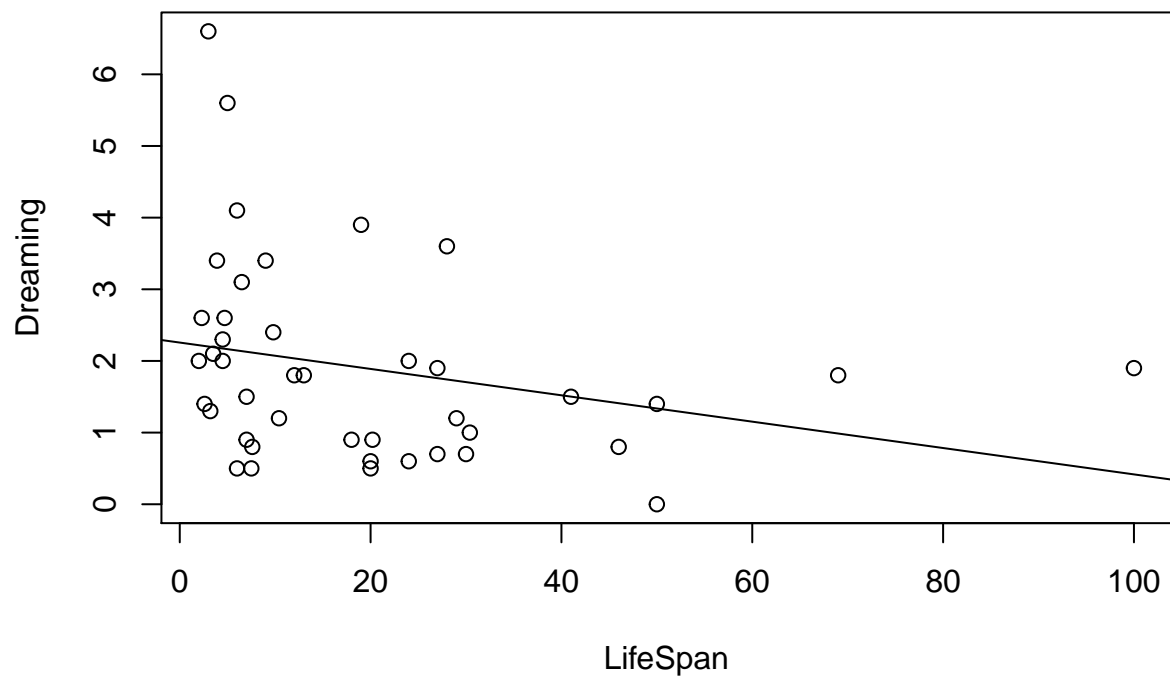
```
#Inspect BrainWt  
plot(BrainWt, Dreaming)  
abline(lm(Dreaming ~ BrainWt, data = data))
```



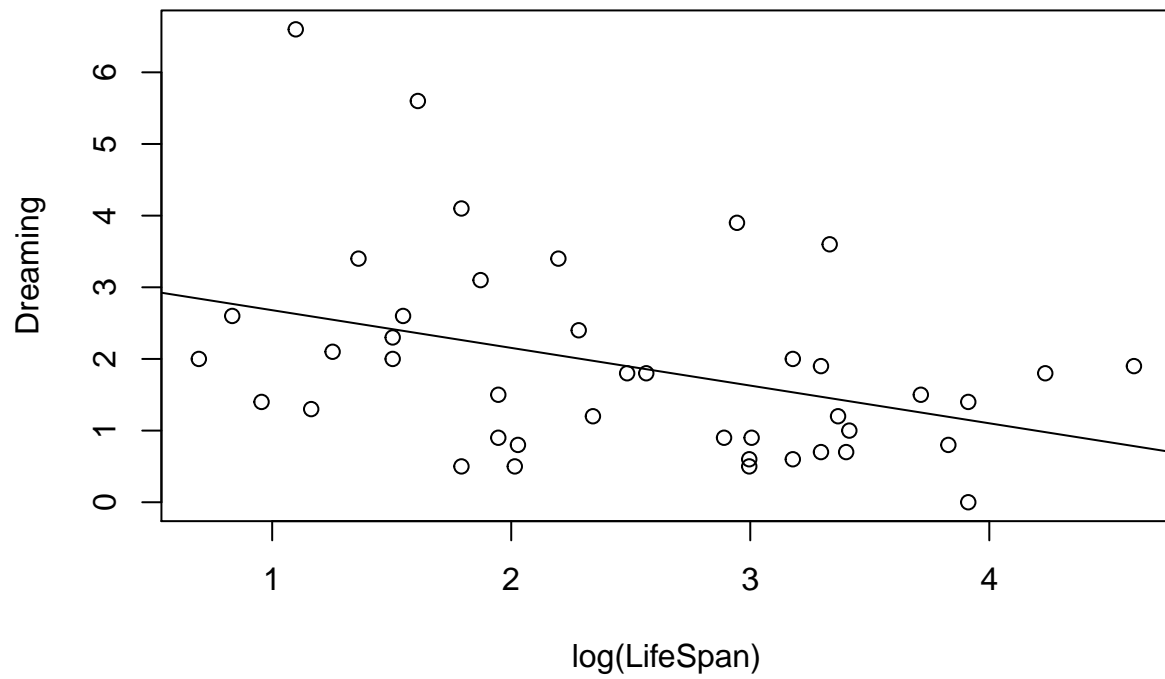
```
# Transform BrainWt  
plot(log(BrainWt), Dreaming)  
abline(lm(Dreaming ~ log(BrainWt), data = data))
```



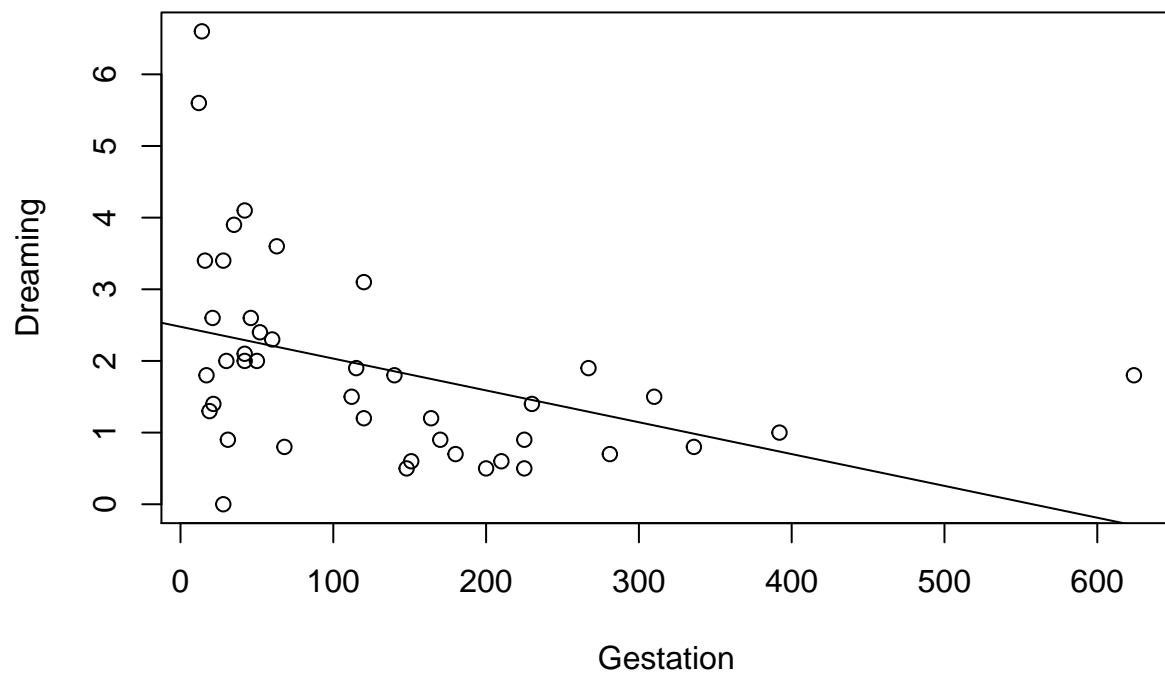
```
#Inspect LifeSpan  
plot(LifeSpan, Dreaming)  
abline(lm(Dreaming ~ LifeSpan, data = data))
```



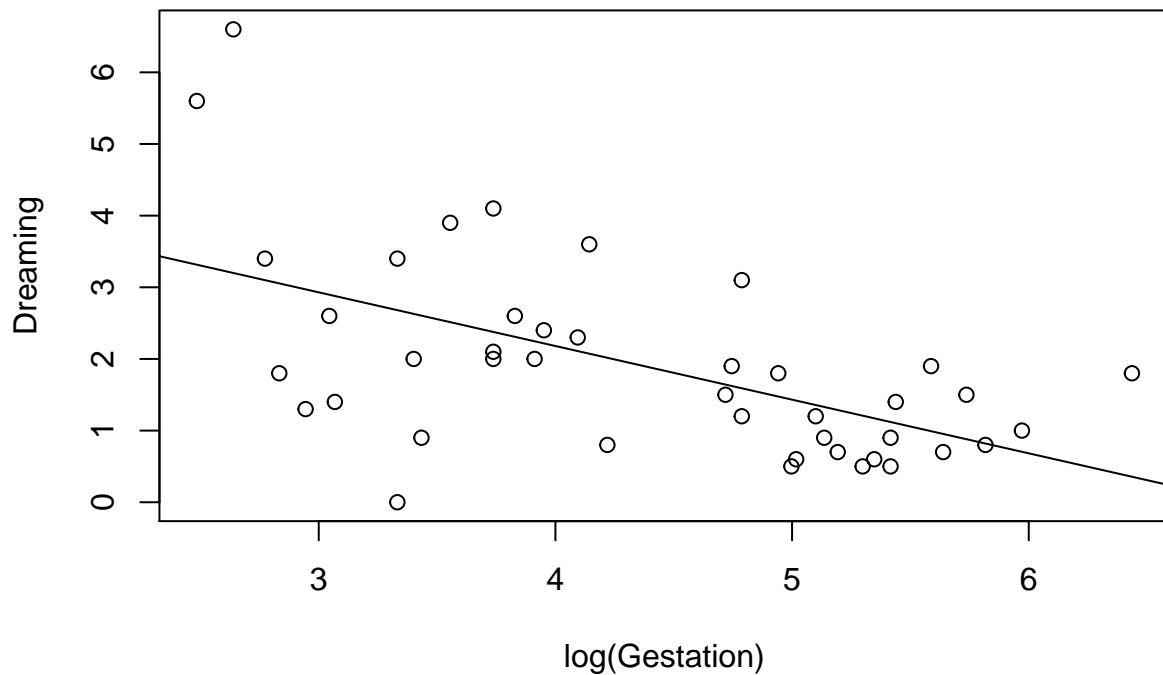
```
# Transform LifeSpan
plot(log(LifeSpan), Dreaming)
abline(lm(Dreaming ~ log(LifeSpan), data = data))
```



```
#Inspect Gestation  
plot(Gestation, Dreaming)  
abline(lm(Dreaming ~ Gestation, data = data))
```



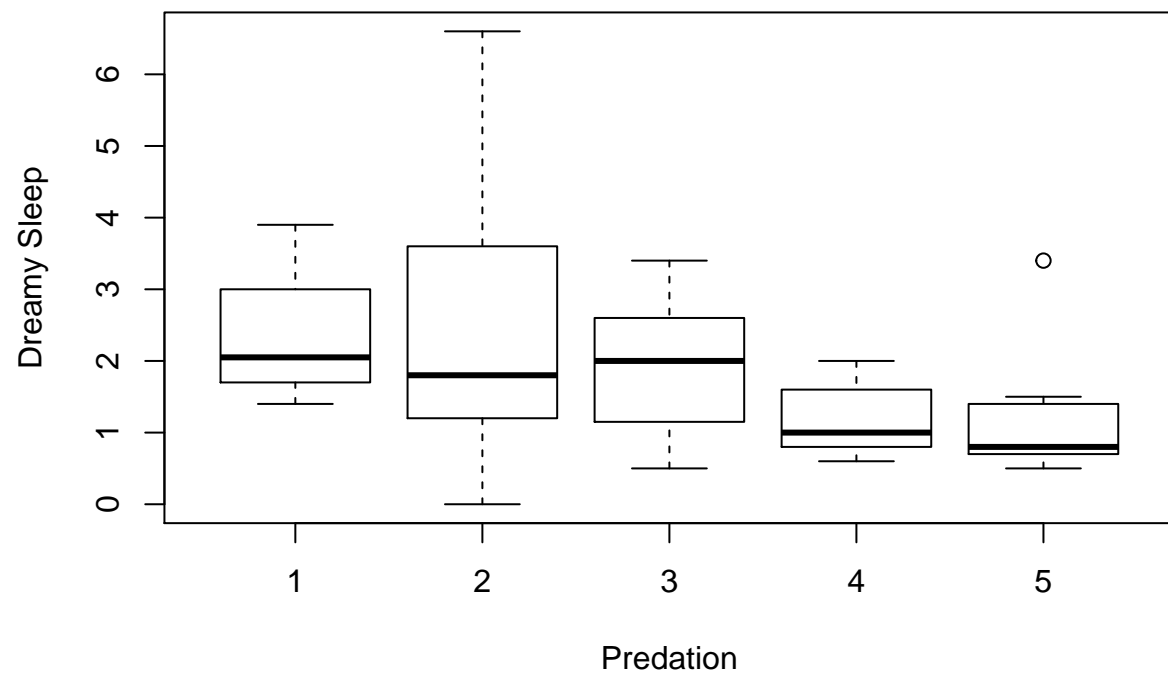
```
# Transform Gestation
plot(log(Gestation), Dreaming)
abline(lm(Dreaming ~ log(Gestation), data = data))
```



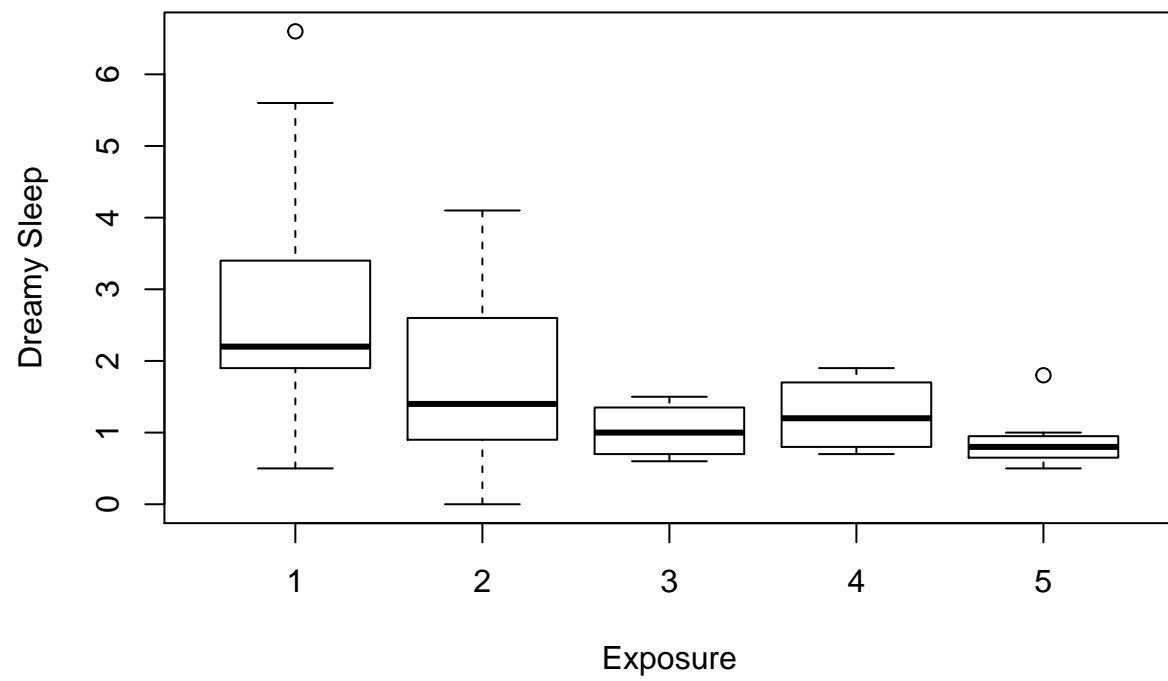
All scatterplots improved after the transformation most dramatically in body and brain weight. This validates the correlation coefficient because all scatterplots show a negative slope as represented by the abline.

2d. Using boxplots, describe the relationship between Dreaming and the categorical independent variables Predation, Exposure, and Danger. Does Dreaming vary with the categorical variables?

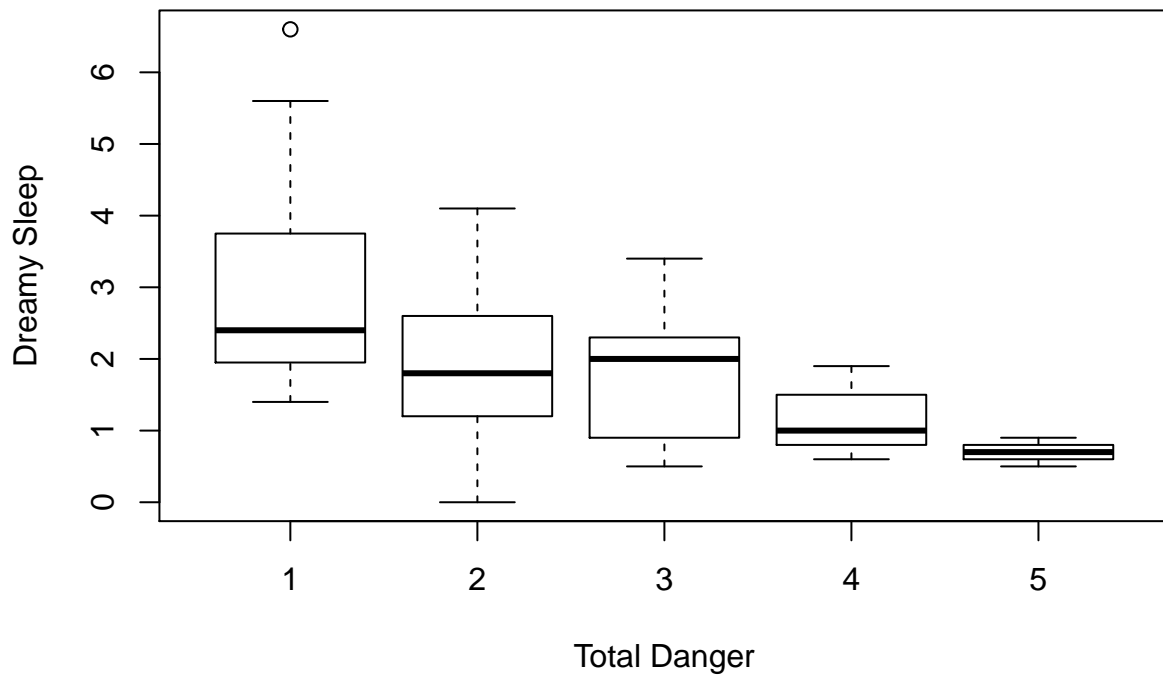
```
boxplot(Dreaming ~ Predation, xlab = "Predation", ylab = "Dreamy Sleep")
```



```
boxplot(Dreaming ~ Exposure, xlab = "Exposure", ylab = "Dreamy Sleep")
```

```
boxplot(Dreaming ~ Danger, xlab = "Total Danger", ylab = "Dreamy Sleep")
```



Predation

predation index (1-5) 1 = minimum (least likely to be preyed upon);

5 = maximum (most likely to be preyed upon)

The predation boxplot show that dreamy sleep varies with predation level. There is a weak negative correlation between dreamy sleep and predation level.

Exposure

sleep exposure index (1-5) 1 = least exposed (e.g. animal sleeps in a well-protected den);

5 = most exposed

The exposure boxplot show that dreamy sleep varies with exposure level. There is a weak negative correlation between dreamy sleep and exposure level.

Danger

overall danger index (1-5) (based on the above two indices and other information) 1 = least danger (from other animals);

5 = most danger (from other animals)

The danger boxplot show that dreamy sleep varies with danger level. There is a moderate negative correlation between dreamy sleep and danger level.

2e. Based on this section for exploratory analysis, is it reasonable to assume a linear regression model? Would you suggest that Dreaming varies with all or only some of the independent variables? Would you recommend using the categorical variables Predation, Exposure, and Danger in the model? Why?

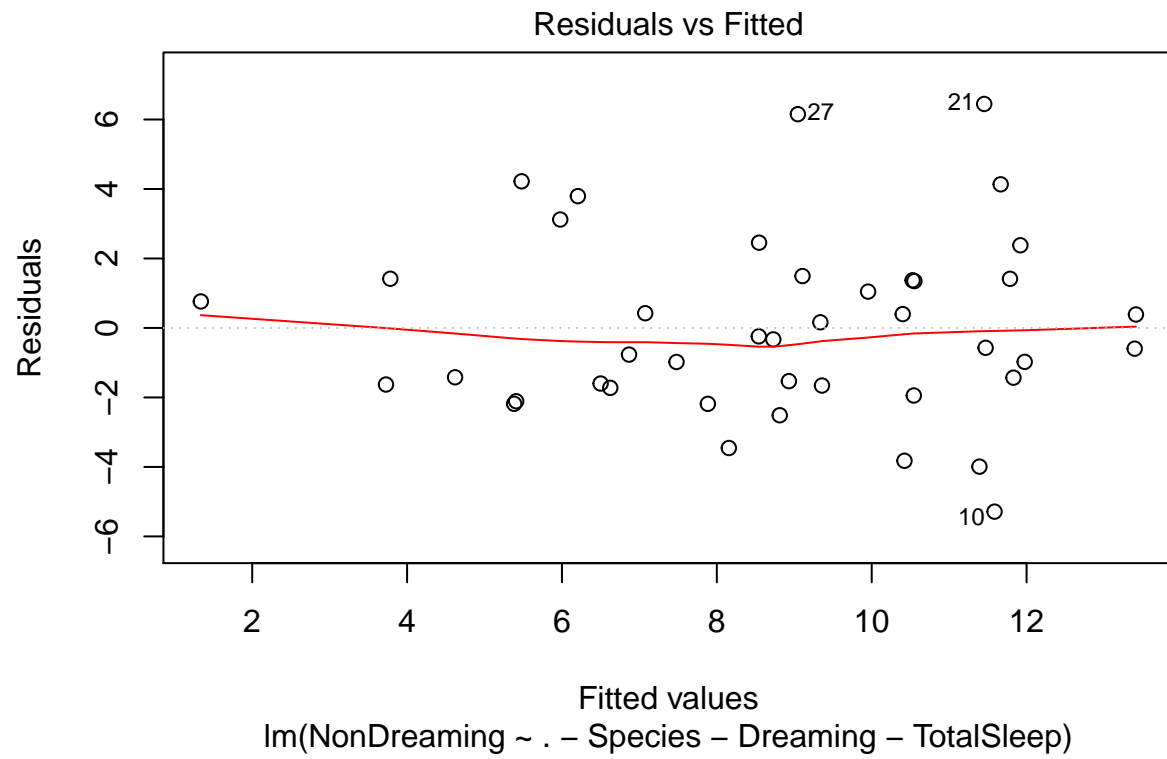
I think it would be reasonable to assume a linear regression model using the Predation, Exposure and Danger categorical variables. All of the variables seemed to have some level of negative correlation by looking at the boxplots with the danger variable having the strongest correlation. Therefore, I think a linear regression model would fit this data and I would add them to the model.

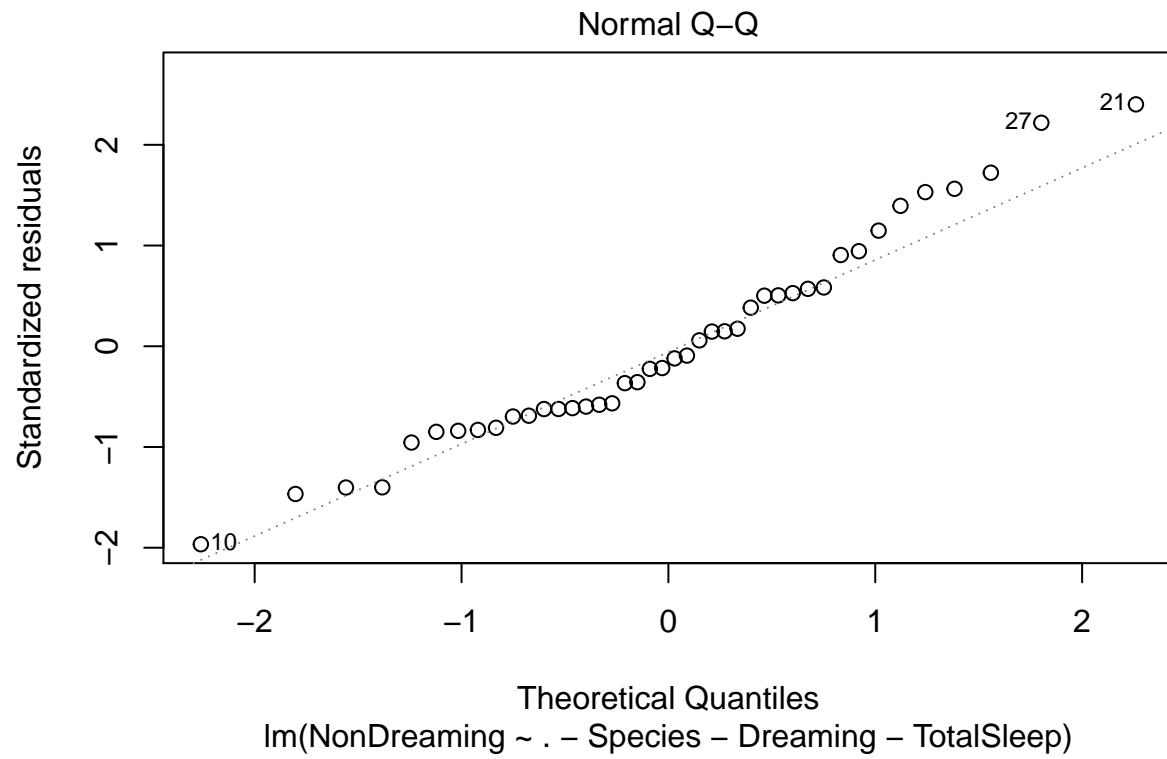
Question 3: Fitting the Linear Regression Model.

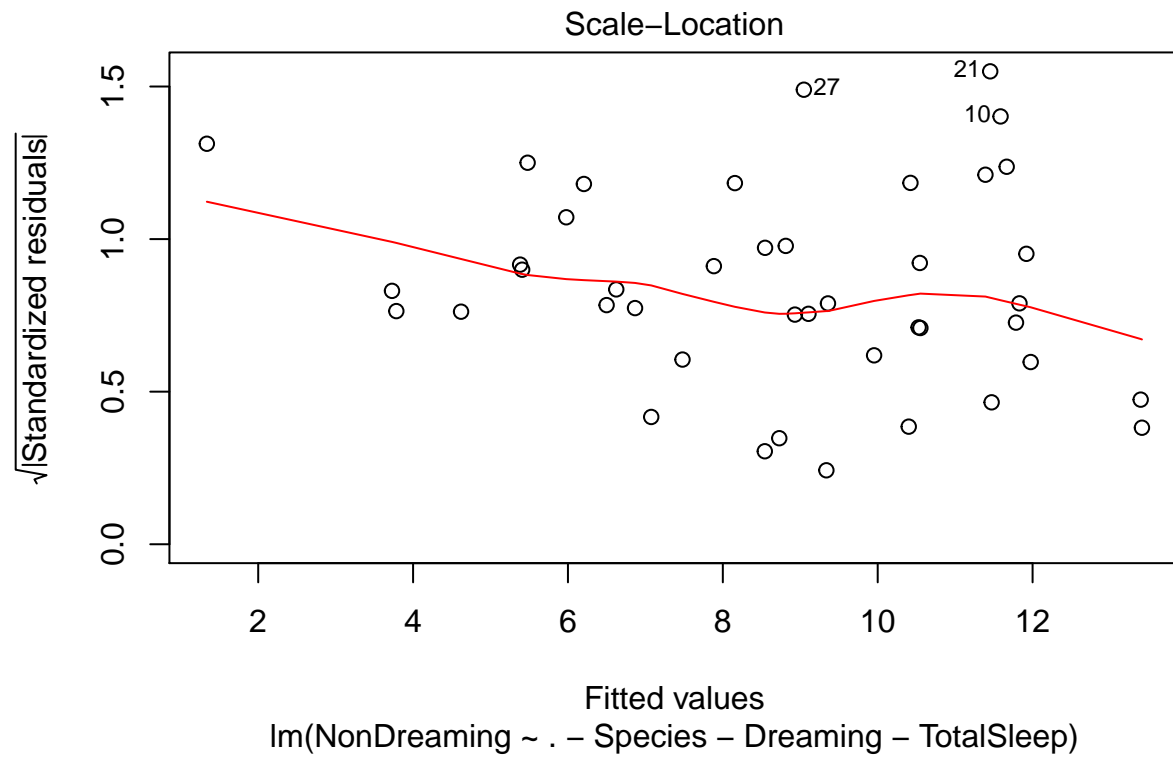
```
model3 <- lm(NonDreaming ~. -Species -Dreaming -TotalSleep, data = data)
```

```
summary(model3)
```

```
##
## Call:
## lm(formula = NonDreaming ~ . - Species - Dreaming - TotalSleep,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.2864 -1.6503 -0.4501  1.4037  6.4473
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.331488   1.256475  10.610 2.5e-12 ***
## BodyWt       0.003332   0.005568   0.598  0.5535
## BrainWt     -0.001294   0.003342  -0.387  0.7010
## LifeSpan    -0.001181   0.043509  -0.027  0.9785
## Gestation   -0.013804   0.006563  -2.103  0.0429 *
## Predation    1.414774   1.027350   1.377  0.1775
## Exposure     0.224418   0.643644   0.349  0.7295
## Danger     -2.799115   1.275630  -2.194  0.0351 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.857 on 34 degrees of freedom
## Multiple R-squared:  0.5404, Adjusted R-squared:  0.4458
## F-statistic: 5.711 on 7 and 34 DF,  p-value: 0.0002014
plot(model3, cook.levels = c(4/42,0.5,1))
```



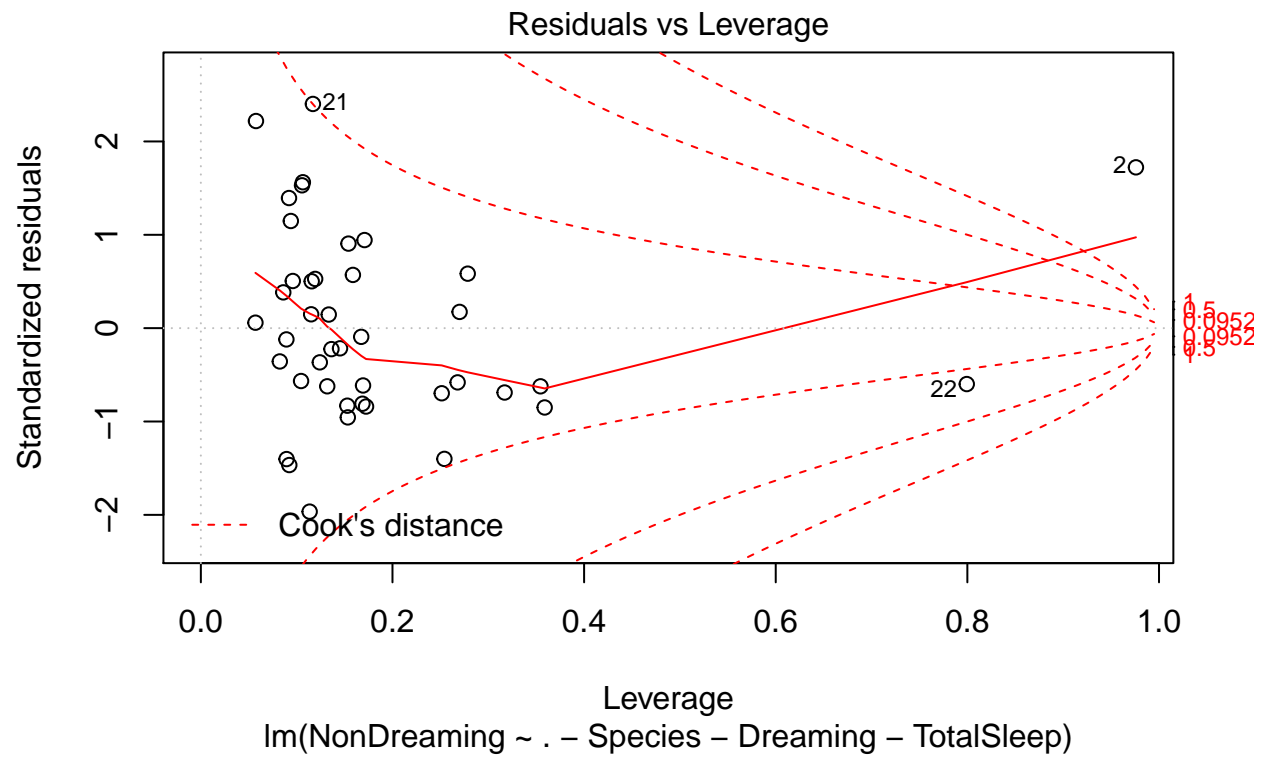




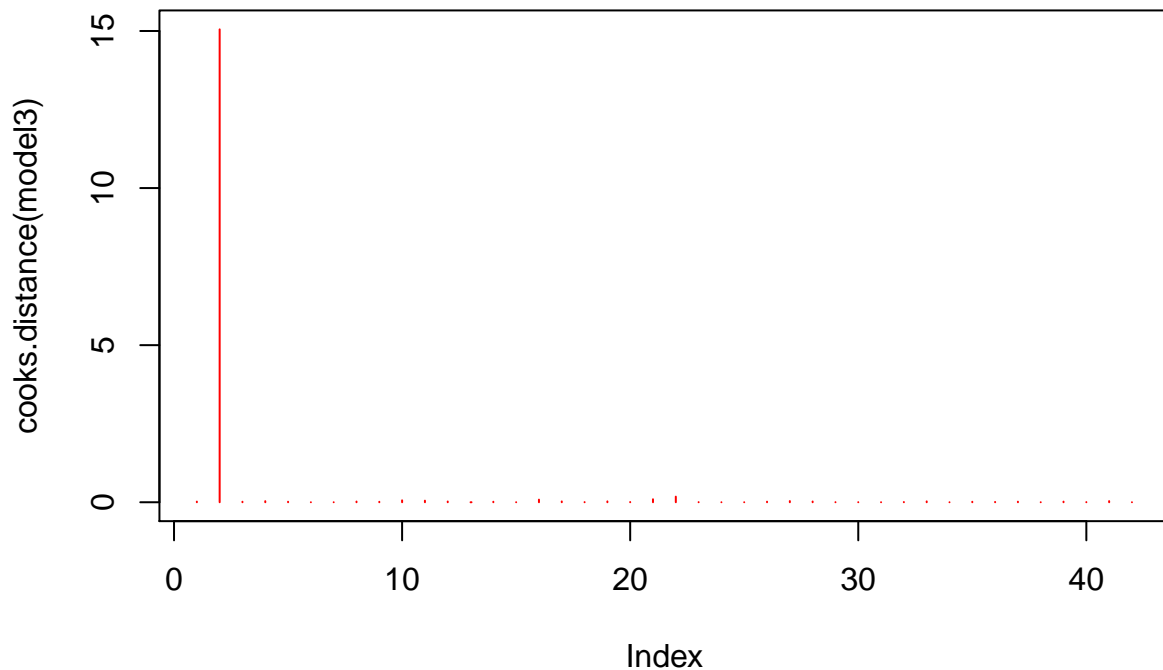
```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



```
plot(cooks.distance(model3), type = 'h', col = 'red')
```



1. What are the model parameters and what are their estimates?

The model parameters with their estimates are BodyWt (0.003332), BrainWt(-0.001294), LifeSpan(-0.001181), Gestation(-0.013804), Predation(1.414774), Exposure(0.224418), and Danger (-2.799115)

2. What is the equation for the regression line?

$y = 13.33 + 0.003332\text{BodyWt} + -0.001294\text{BrainWt} + -0.001181\text{LifeSpan} + -0.013804\text{Gestation} + 1.414774\text{Predation} + 0.224418\text{Exposure} - 2.799115\text{Danger}$

3. Which predicting variable(s) are significant at $\alpha = 0.05$? What are their p-values?

Gestation and Danger are the only predictors with p values $< .05$ at .0429 and .0351 respectively.

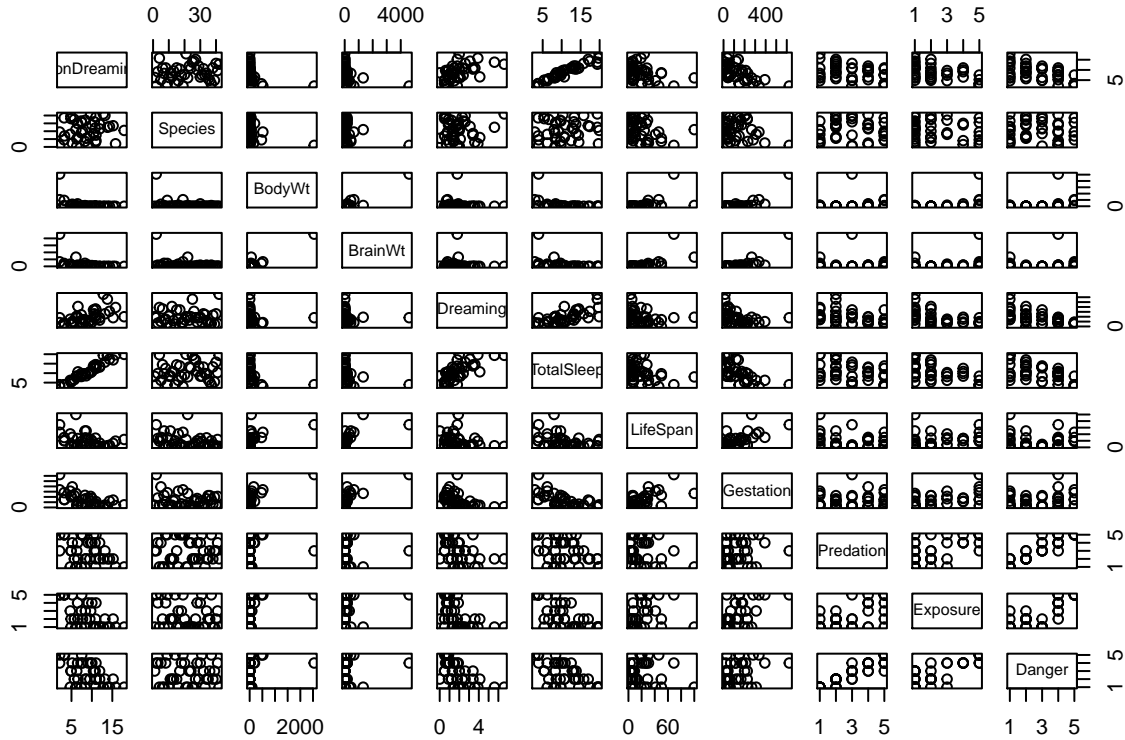
4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The BodyWt parameter is 0.003. This represents the estimated expected change in nonDreaming sleep with one unit of change in BodyWt, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for BodyWt is .0056. So our prediction, on average will be off by .0056.

The Predation parameter is 1.41. This represents the estimated expected change in nonDreaming sleep with one unit of change in Predation, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for Predation is 1.03. So our prediction, on average will be off by 1.03.

5. Check the assumptions of the model through plotting. Note potential outliers, if any.

```
pairs(NonDreaming ~. -Species -Dreaming -TotalSleep, data = data) #or plot(data)
```



Linearity Focusing on the first row in the matrix plot above, we can see a weak negative correlation with nonDreaming and gestation, predation and danger. All other variables associated with the model do not appear to show any linear relationship with nondreaming, the relationship does not hold for all variables.

Constant variance and uncorrelated errors

The residual vs fitted values above shows the errors scattered around the zero line, which indicates that the constant variance and uncorrelated errors assumption holds.

Normality

The q-q plot of the standardized residuals shown above shows that the normality assumption holds with two outliers on the outer right tail. An outlier is also noted on the lower left tail, but it falls within the normal range.

Outliers

The last plot, the cook's distance notes one outlier.

Question 3a: Change model3 to log transform the response variable, NonDreaming.

```
model3a <- lm(log(NonDreaming) ~. -Species -Dreaming -TotalSleep, data = data)
```

```
summary(model3a)
```

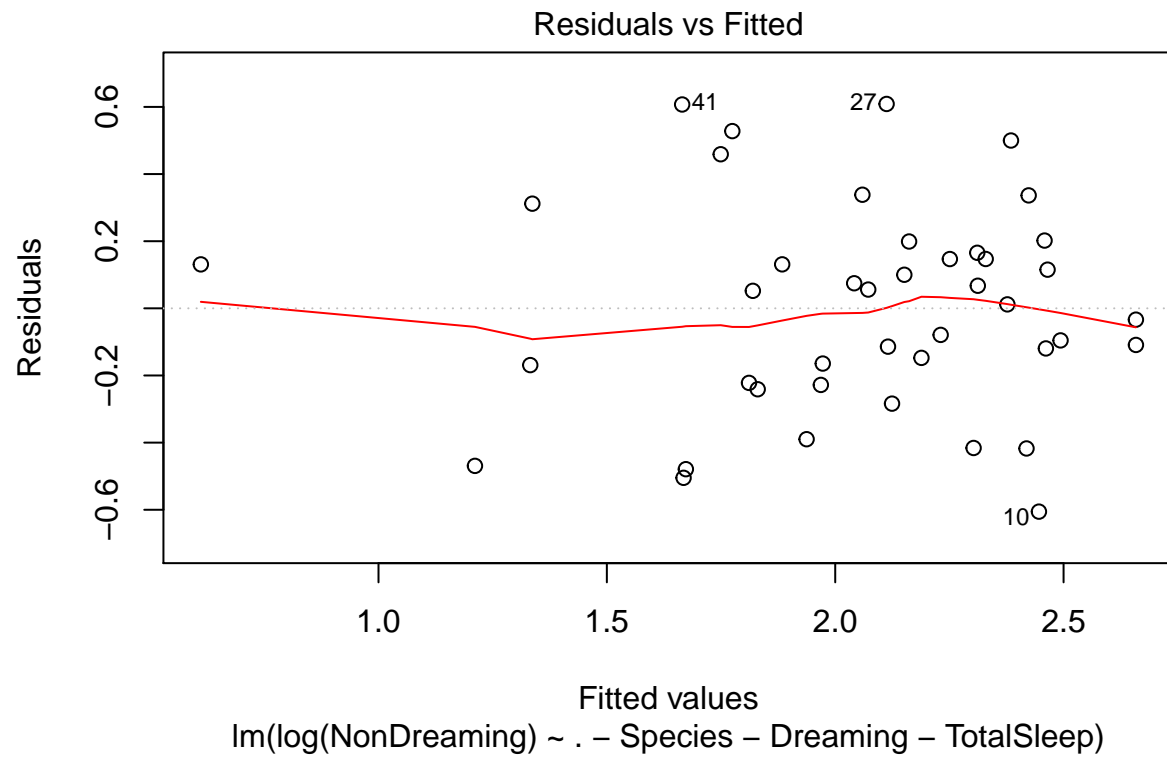
```
##
## Call:
## lm(formula = log(NonDreaming) ~ . - Species - Dreaming - TotalSleep,
##     data = data)
##
## Residuals:
```

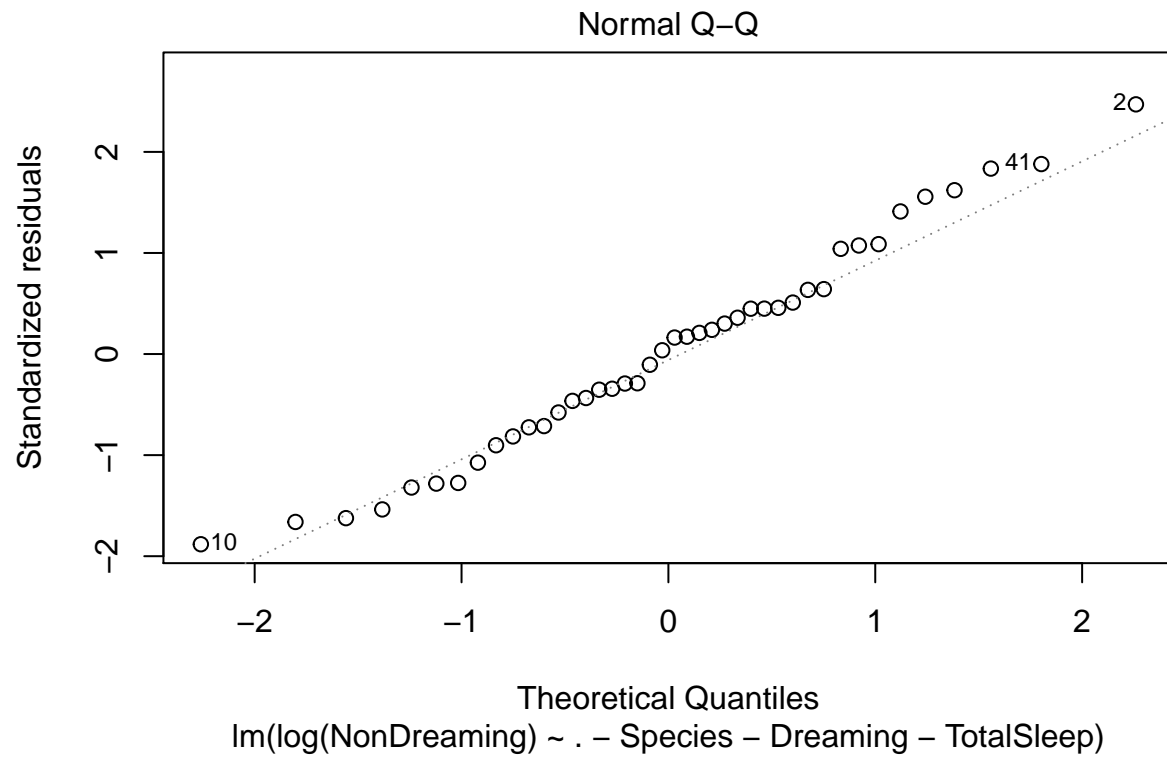
	Min	1Q	Median	3Q	Max
	-0.60553	-0.20880	0.03197	0.16087	0.60877

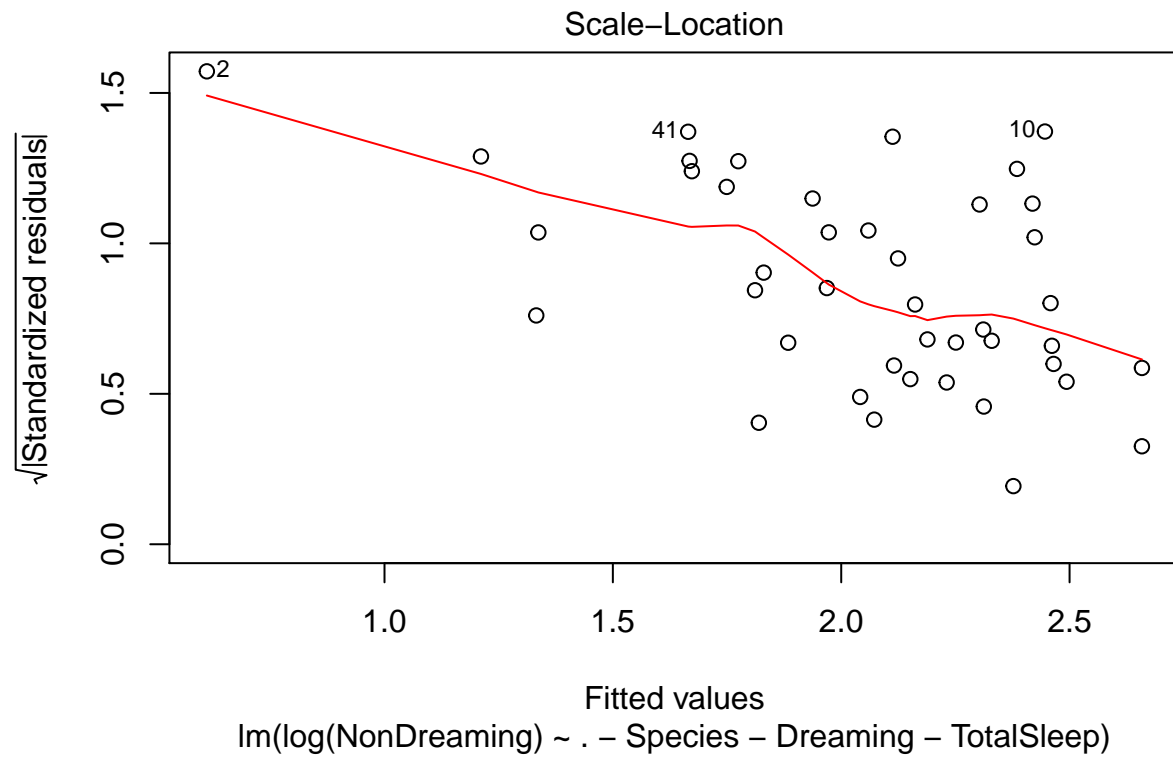
```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.6679033	0.1503207	17.748	<2e-16 ***
BodyWt	-0.0002651	0.0006661	-0.398	0.6931
BrainWt	0.0001303	0.0003998	0.326	0.7464
LifeSpan	-0.0022691	0.0052053	-0.436	0.6657
Gestation	-0.0018214	0.0007851	-2.320	0.0265 *
Predation	0.1608420	0.1229089	1.309	0.1994
Exposure	0.0224126	0.0770035	0.291	0.7728
Danger	-0.3207756	0.1526124	-2.102	0.0430 *

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3418 on 34 degrees of freedom
## Multiple R-squared:  0.6418, Adjusted R-squared:  0.5681
## F-statistic: 8.703 on 7 and 34 DF,  p-value: 4.29e-06
plot(model3a, cook.levels = c(4/42,0.5,1))
```



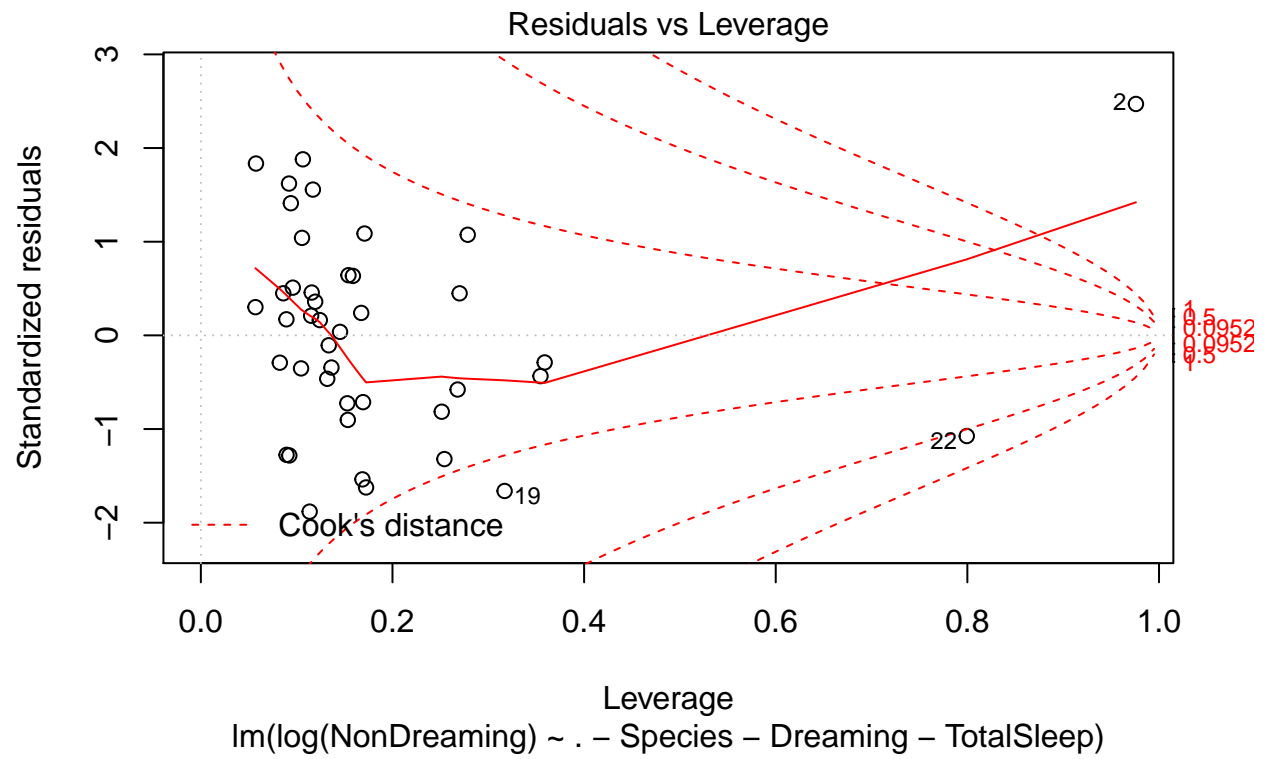




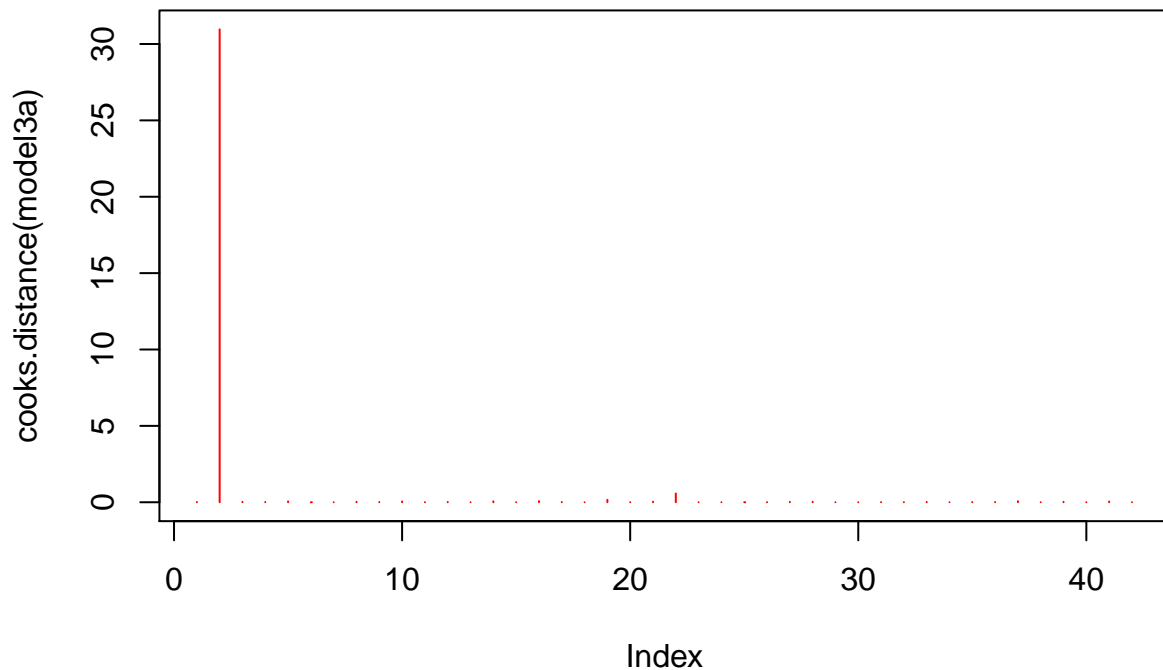
```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



```
plot(cooks.distance(model3a), type='h', col='red')
```



1. What are the model parameters and what are their estimates?

The model parameters with their estimates are BodyWt (0-0.0002651), BrainWt(0.0001303), LifeSpan(-0.0022691), Gestation(-0.0018214), Predation(0.1608420), Exposure(0.0224126), and Danger (-0.3207756)

2. What is the equation for the regression line?

$$\log(y) = 2.6679033 + -0.0002651\text{BodyWt} + 0.0001303\text{BrainWt} + -0.0022691\text{LifeSpan} + -0.0018214\text{Gestation} + 0.1608420\text{Predation} + 0.0224126\text{Exposure} - 0.3207756\text{Danger}$$

3. Which predicting variable(s) are significant at alpha = 0.05? What are their p-values?

Gestation and Danger are the only predictors with p values < .05 at 0.0265 and 0.0430 respectively.

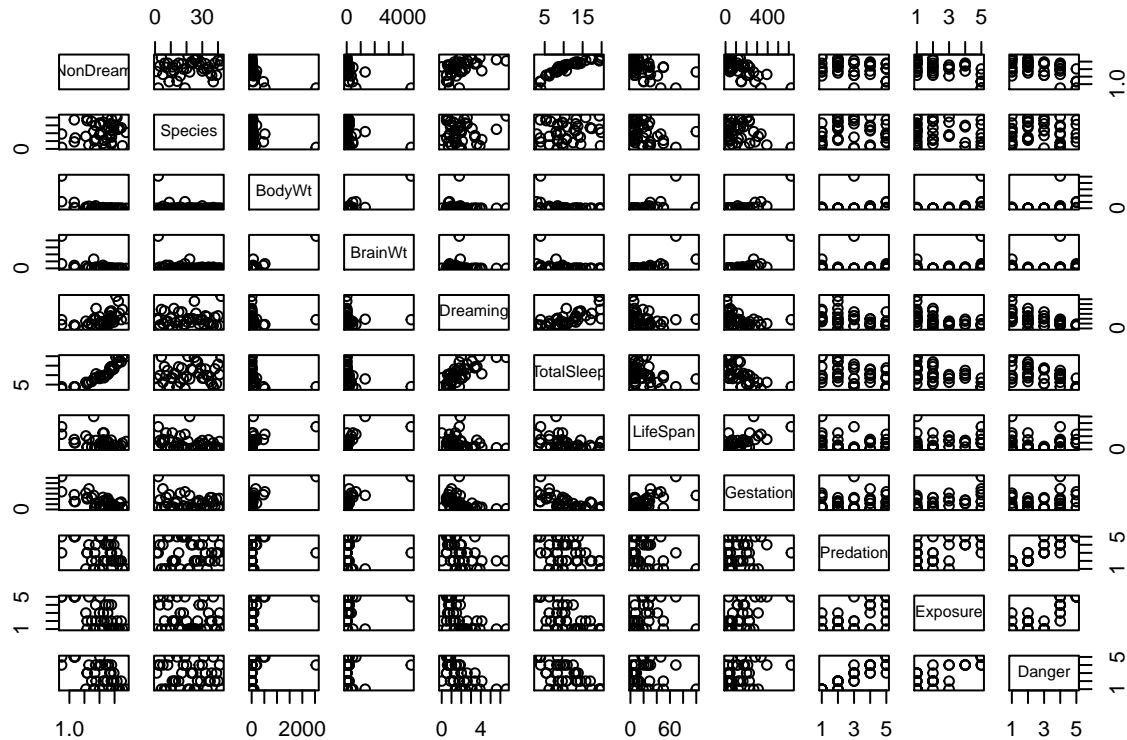
4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The BodyWt parameter is 0.0001303. As BodyWt increases by 1 unit, nondreaming increases by 0.0001303*100%, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for BodyWt is 0.0006661. So our prediction, on average will be off by 0.0006661.

The Predation parameter is 0.1608420. This represents the estimated expected change in nonDreaming sleep with one unit of change in Predation, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for Predation is 0.1229089. So our prediction, on average will be off by 0.1229089.

5. Check the assumptions of the model through plotting. Note potential outliers, if any.

```
pairs(log(NonDreaming) ~. -Species -Dreaming -TotalSleep, data = data) #or plot(data)
```



Linearity

Focusing on the first row in the matrix plot above we can see a weak negative correlation with nonDreaming and gestation, predation and danger. All other variables associated with the model do not appear to show any linear relationship with nondreaming, the relationship does not hold for all variables.

Constant variance and uncorrelated errors

The residual vs fitted values above shows the errors scattered around the zero line, which indicates that the constant variance and uncorrelated errors assumption holds.

Normality

The q-q plot of the standardized residuals shown above shows that the normality assumption holds with two outliers noted on the outer right tail and one on the lower left tail, but all appear to fall within the normal range.

Outliers

The last plot, the cook's distance notes one outlier.

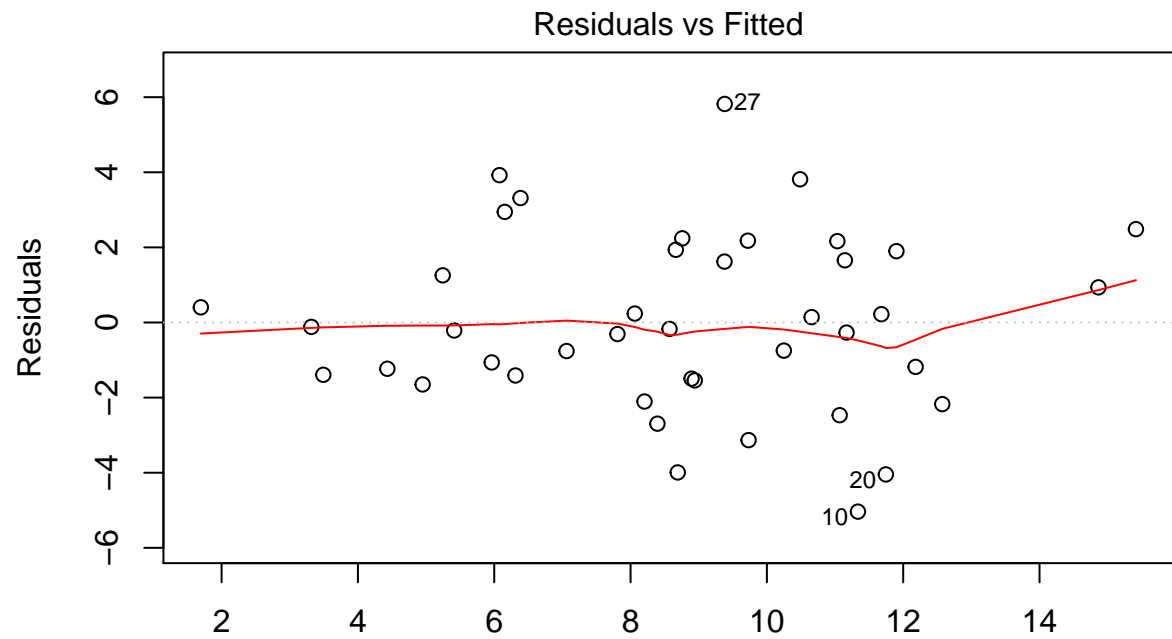
Question 3b: Change model3a to remove the log transform of NonDreaming, and add the log transformation of numeric response variables BrainWt, BodyWt, LifeSpan and Gestation

```
model3b <- lm(NonDreaming ~ log(BodyWt) +log(BrainWt) +log(LifeSpan) +log(Gestation) +Exposure +Predation)

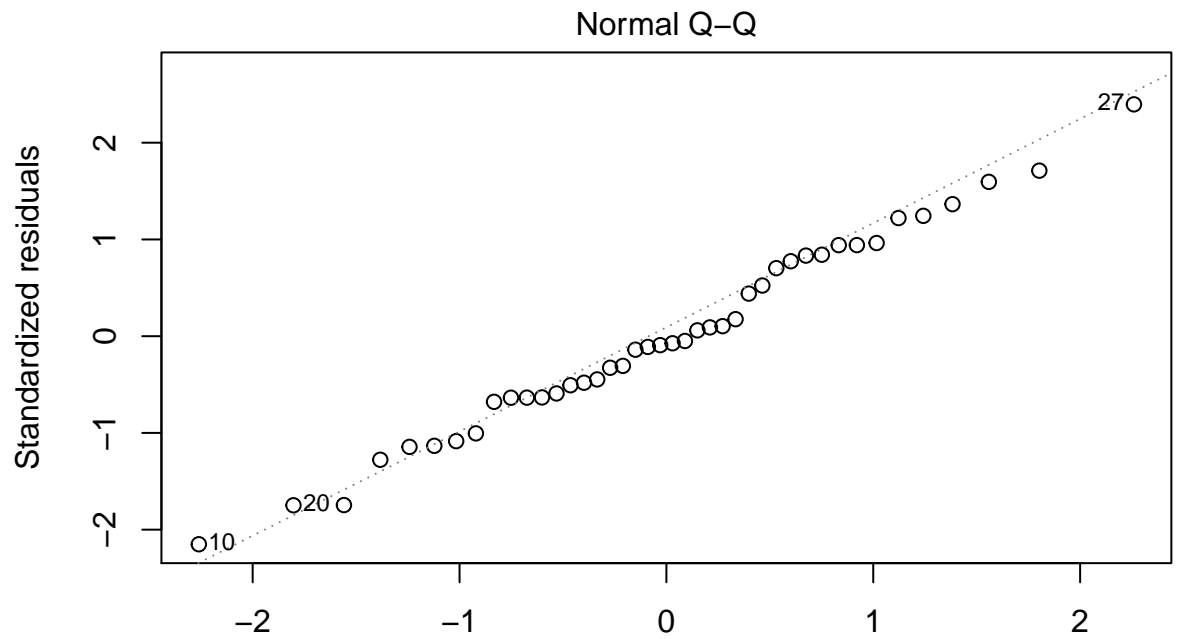
summary(model3b)

##
## Call:
## lm(formula = NonDreaming ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) +
##     log(Gestation) + Exposure + Predation + Danger, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.0359 -1.4743 -0.1921  1.8385  5.8191
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.4819     2.8885   3.975 0.000348 ***
## log(BodyWt)     -0.3917     0.4787  -0.818 0.418974
## log(BrainWt)    -0.5591     0.7019  -0.797 0.431228
## log(LifeSpan)     1.2991     0.7492   1.734 0.091993 .
## log(Gestation)  -0.5083     0.6613  -0.769 0.447419
## Exposure         0.5310     0.6196   0.857 0.397516
## Predation        1.6141     0.9812   1.645 0.109177
## Danger          -2.9313     1.1628  -2.521 0.016562 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.58 on 34 degrees of freedom
## Multiple R-squared:  0.6253, Adjusted R-squared:  0.5481
## F-statistic: 8.105 on 7 and 34 DF,  p-value: 8.706e-06

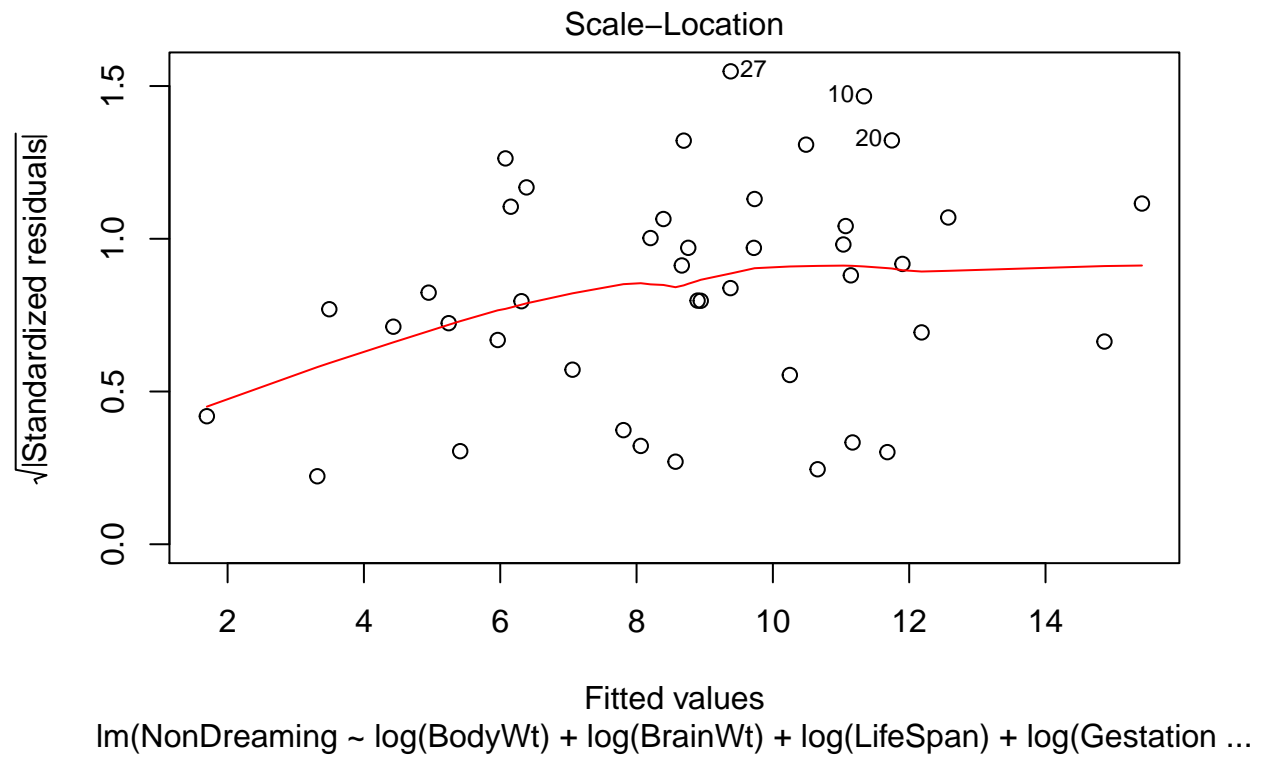
plot(model3b, cook.levels = c(4/42,0.5,1))
```

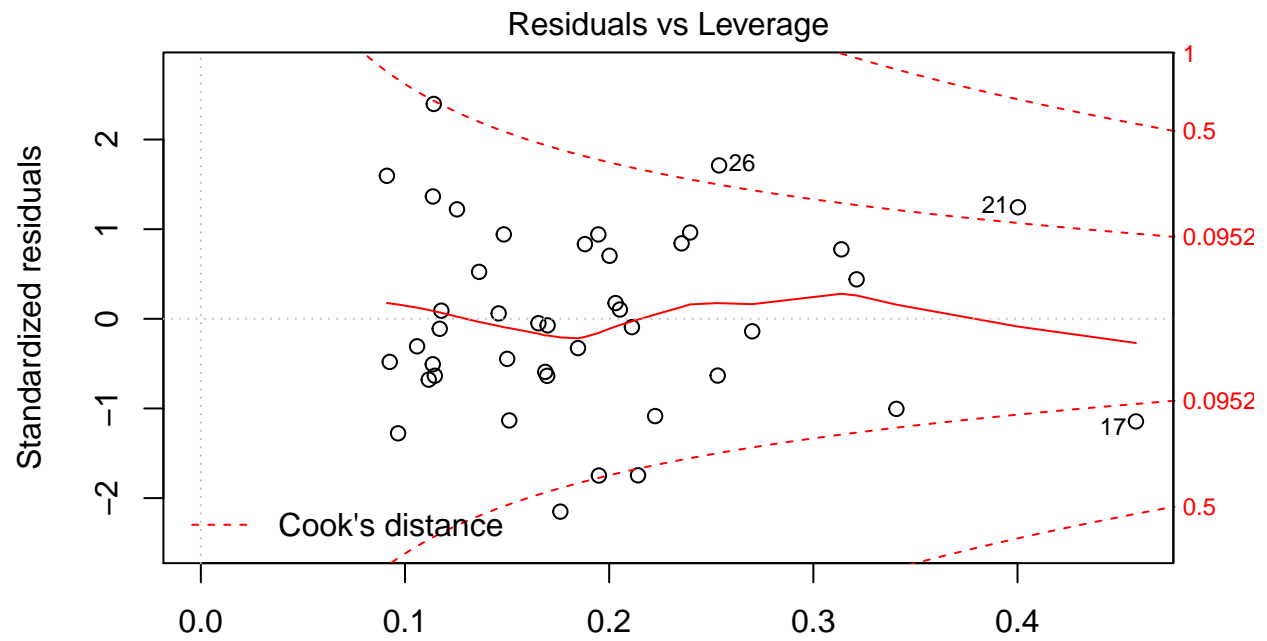


Fitted values
 $\text{lm}(\text{NonDreaming} \sim \log(\text{BodyWt}) + \log(\text{BrainWt}) + \log(\text{LifeSpan}) + \log(\text{Gestation} \dots$



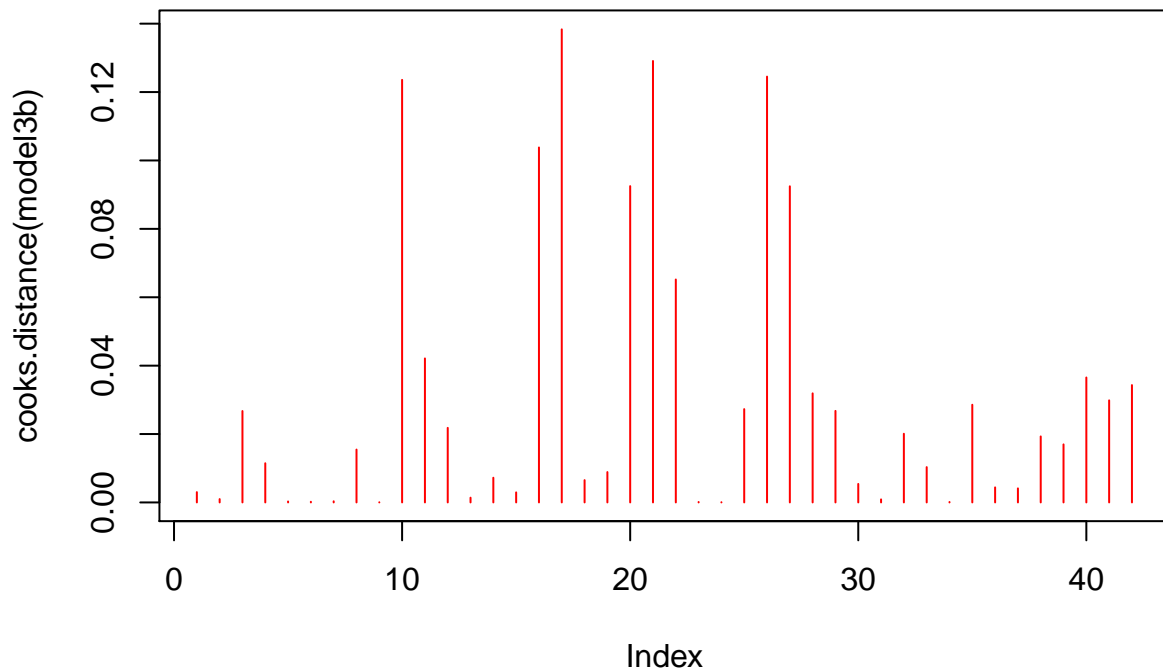
Im(NonDreaming ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) + log(Gestation ...





Im(NonDreaming ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) + log(Gestation ...

```
plot(cooks.distance(model3b), type = 'h', col = 'red')
```



1. What are the model parameters and what are their estimates?

The model parameters with their estimates are $\log(\text{BodyWt})$ (-0.3917), $\log(\text{BrainWt})$ (-0.5591), $\log(\text{LifeSpan})$ (1.2991), $\log(\text{Gestation})$ (-0.5083), Predation (1.6141), Exposure (0.5310), and Danger (-2.9313)

2. What is the equation for the regression line?

$$y = 11.4819 + \log(\text{BodyWt})(-0.3917) + \log(\text{BrainWt})(-0.5591) + \log(\text{LifeSpan})(1.2991) + \log(\text{Gestation})(-0.5083) + \text{Predation}(1.6141) + \text{Exposure}(0.5310) + \text{Danger}(-2.9313)$$

3. Which predicting variable(s) are significant at $\alpha = 0.05$? What are their p-values?

Danger is the only predictor with a p value $< .05$ at 0.016562.

4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The $\log(\text{BodyWt})$ parameter is -0.3917. As BodyWt increases by 1% nondreaming will decrease by -0.3917/100, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for $\log(\text{BodyWt})$ is 0.4787. So our prediction, on average will be off by 0.0006661.

The Predation parameter is 1.6141. This represents the estimated expected change in nonDreaming sleep with one unit of change in Predation , holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for Predation is 0.9812. So our prediction, on average will be off by 0.9812.

5. Did model3b improve over model3a? Explain how you determined if the model improved or not.

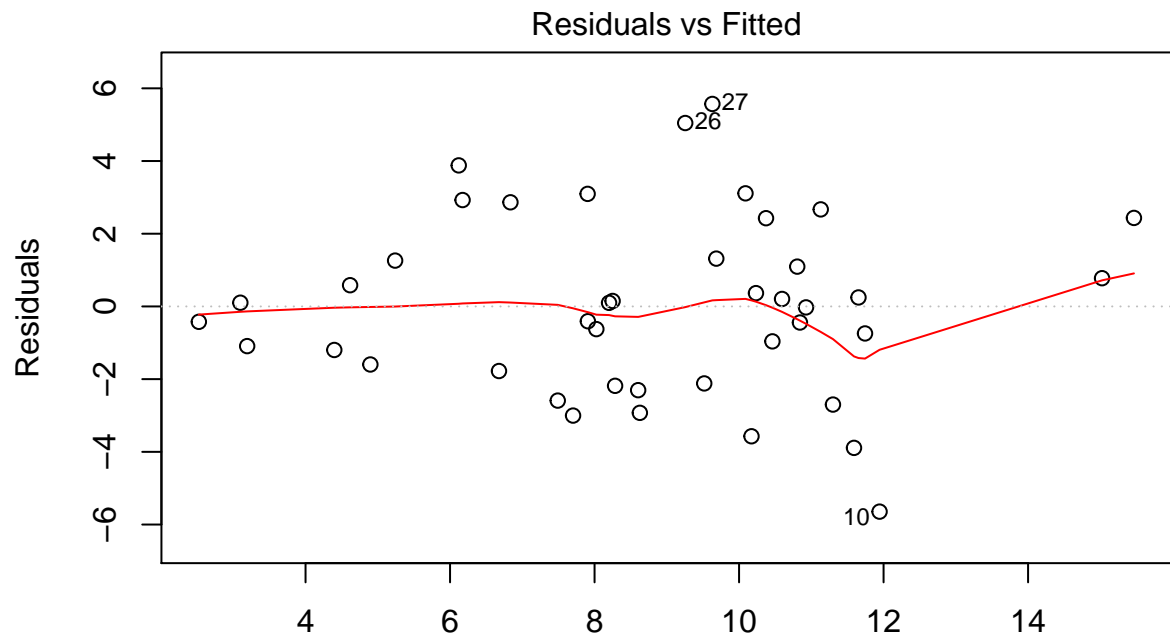
I would say that it did not improve for 2 reasons. First the R^2 value decreased, and second adding a transformation makes the model more difficult to explain and should gain some predictive power if used.

Question 3c: Because the Danger variable is an interpolation of the Exposure and Predation variables, let's keep Danger and remove the other two from the model using model3b as your baseline.

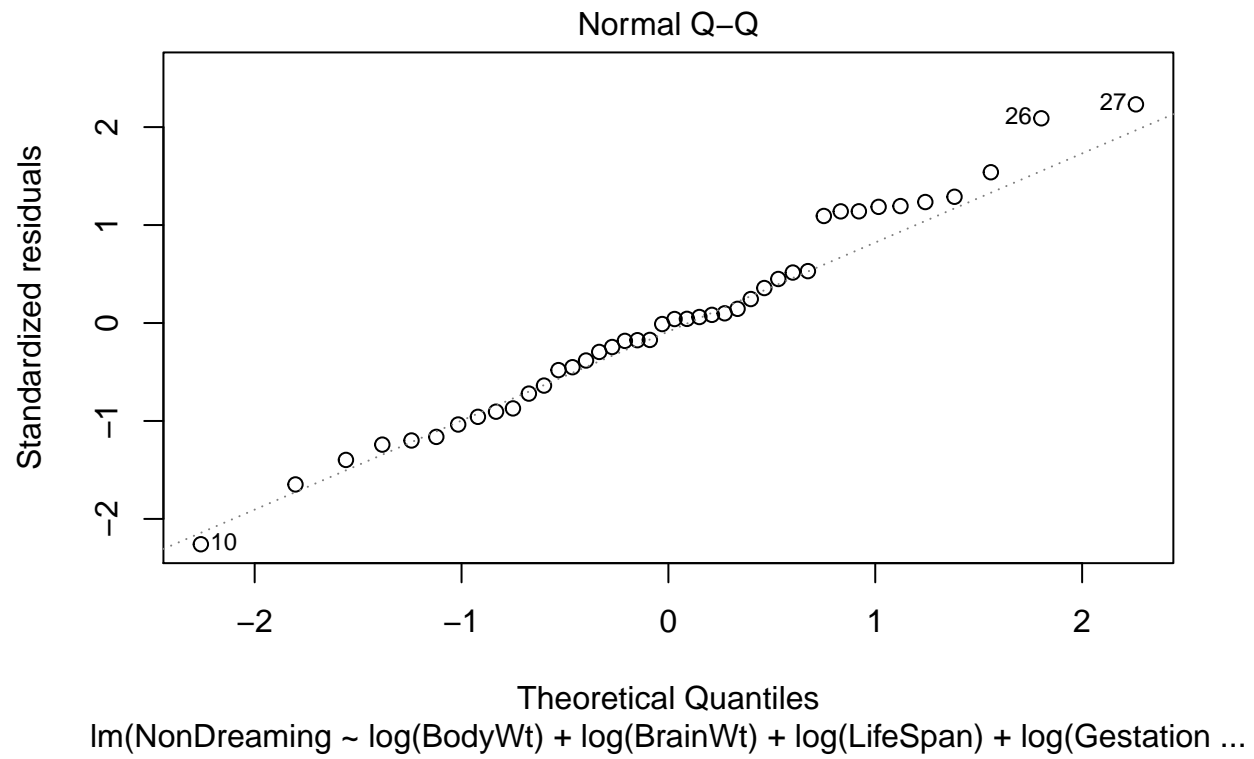
```
model3c <- lm(NonDreaming ~ log(BodyWt) +log(BrainWt) +log(LifeSpan) +log(Gestation) +Danger, data = da

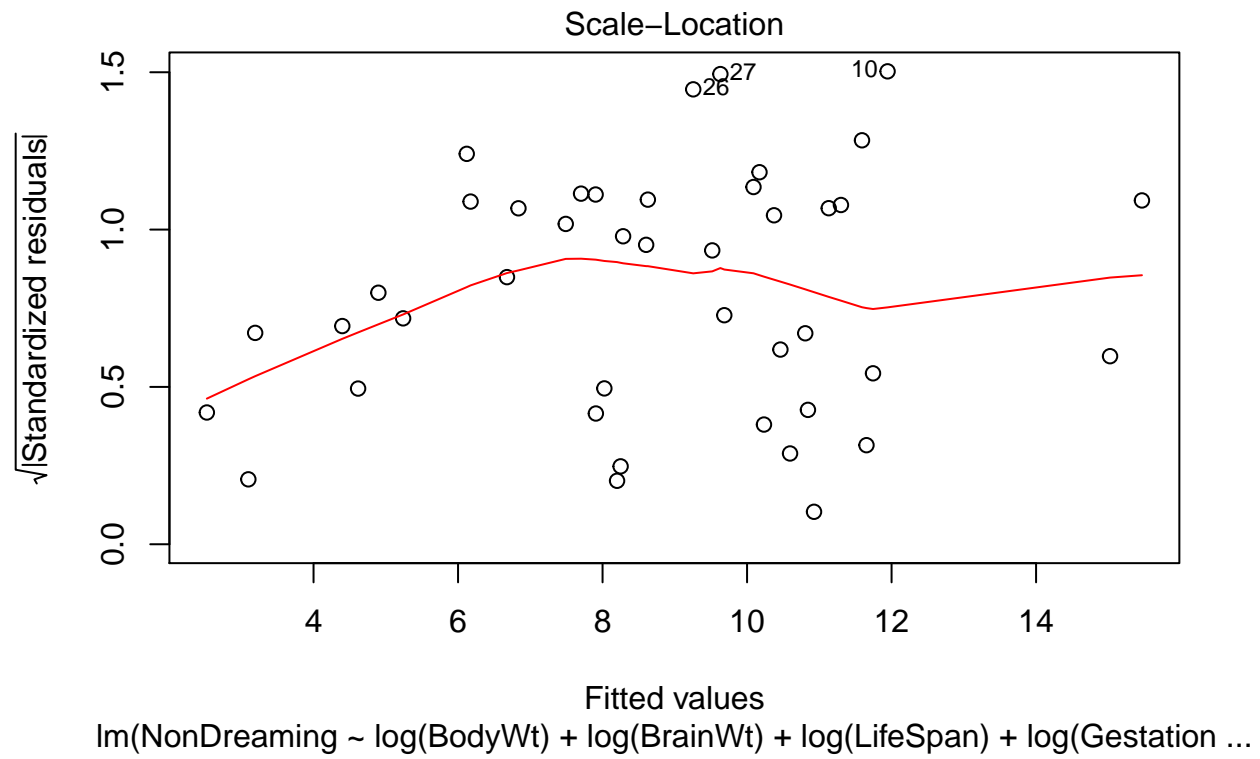
summary(model3c)

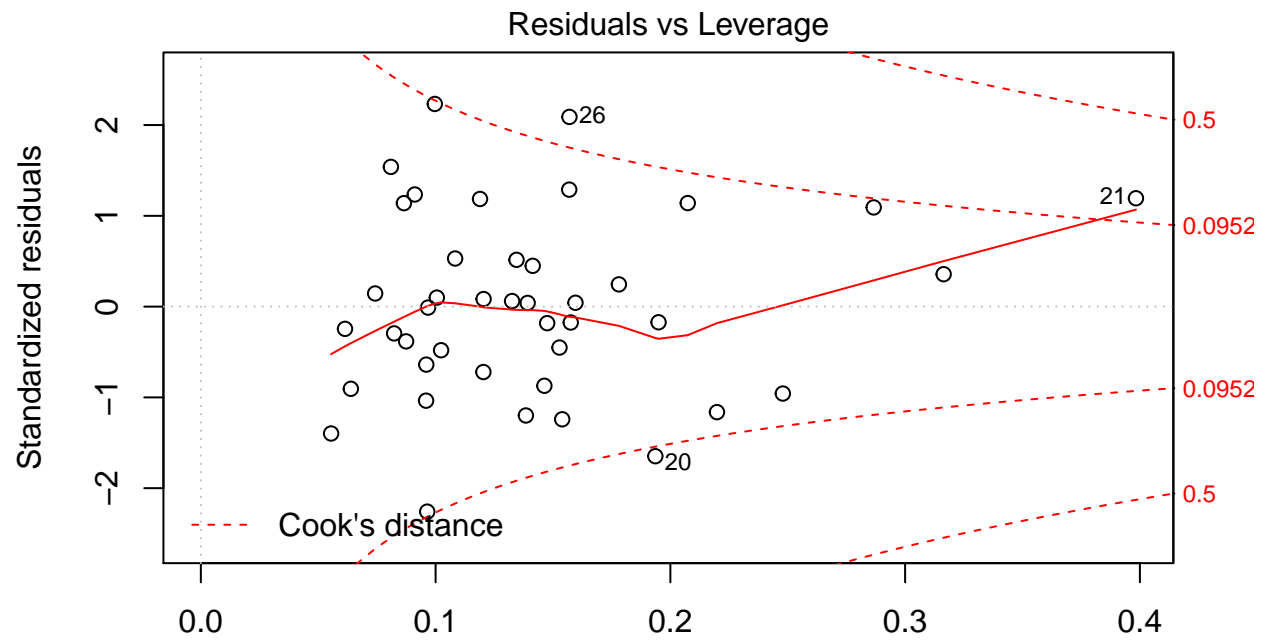
##
## Call:
## lm(formula = NonDreaming ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) +
##     log(Gestation) + Danger, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.6447 -1.7321  0.0363  1.3016  5.5696
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    13.0446     2.5383   5.139 9.82e-06 ***
## log(BodyWt)     -0.4294     0.4663  -0.921  0.3633
## log(BrainWt)    -0.4437     0.7008  -0.633  0.5307
## log(LifeSpan)    1.0811     0.6867   1.574  0.1241
## log(Gestation)  -0.6992     0.6362  -1.099  0.2790
## Danger          -0.8723     0.3249  -2.685  0.0109 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.629 on 36 degrees of freedom
## Multiple R-squared:  0.5878, Adjusted R-squared:  0.5305
## F-statistic: 10.27 on 5 and 36 DF,  p-value: 3.573e-06
plot(model3c, cook.levels = c(4/42,0.5,1))
```



Fitted values
 $\text{lm}(\text{NonDreaming} \sim \log(\text{BodyWt}) + \log(\text{BrainWt}) + \log(\text{LifeSpan}) + \log(\text{Gestation} \dots$

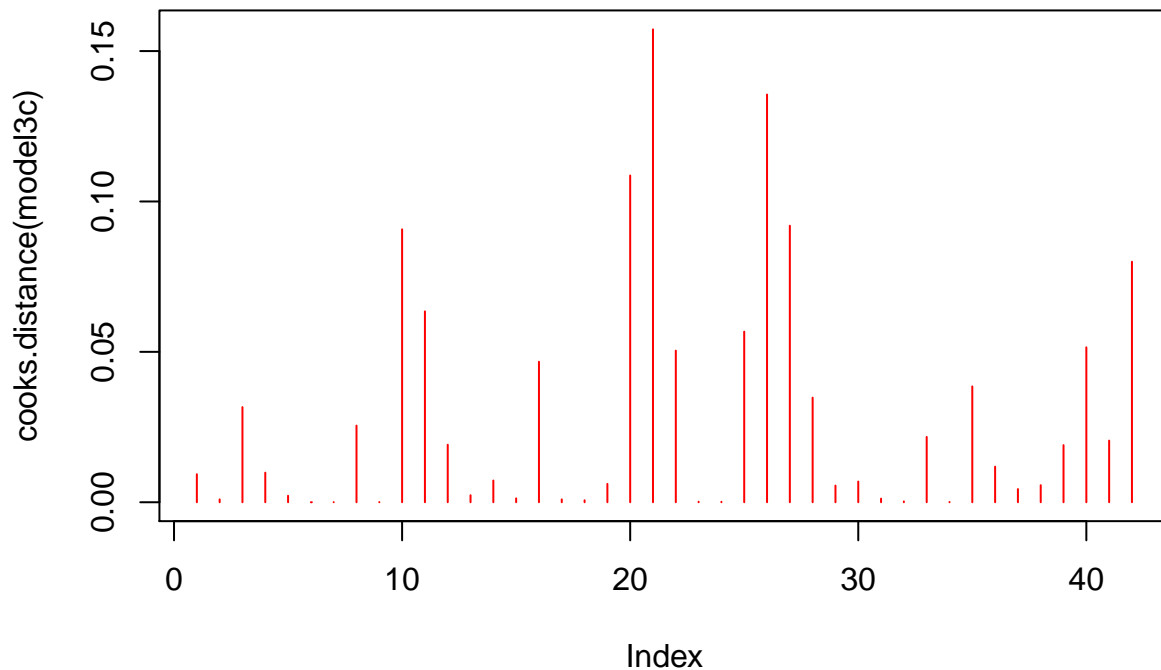






Im(NonDreaming ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) + log(Gestation ...

```
plot(cooks.distance(model3c), type='h', col='red')
```



1. What are the model parameters and what are their estimates?

The model parameters with their estimates are $\log(\text{BodyWt})$ (-0.4294), $\log(\text{BrainWt})$ (-0.4437), $\log(\text{LifeSpan})$ (1.0811), $\log(\text{Gestation})$ (-0.6992), and Danger (-0.8723)

2. What is the equation for the regression line?

$y = 13.0446 + \log(\text{BodyWt})(-0.4294) + \log(\text{BrainWt})(-0.4437) + \log(\text{LifeSpan})(1.0811) + \log(\text{Gestation})(-0.6992) + \text{Danger}$
 ### 3. Which predicting variable(s) are significant at $\alpha = 0.05$? What are their p-values?

Danger is the only predictor with a p value $< .05$ at 0.0109.

4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The $\log(\text{BodyWt})$ parameter is -0.4294. As BodyWt increases by 1% nondreaming will decrease by $-0.4294/100$, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for $\log(\text{BodyWt})$ is 0.4663. So our prediction, on average will be off by 0.0006661.

The Danger parameter is -0.8723. This represents the estimated expected change in nonDreaming sleep with one unit of change in Danger , holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for Predation is 0.3249. So our prediction, on average will be off by 0.3249.

5. Did model3c improve over model3b? Explain how you determined if the model improved or not.

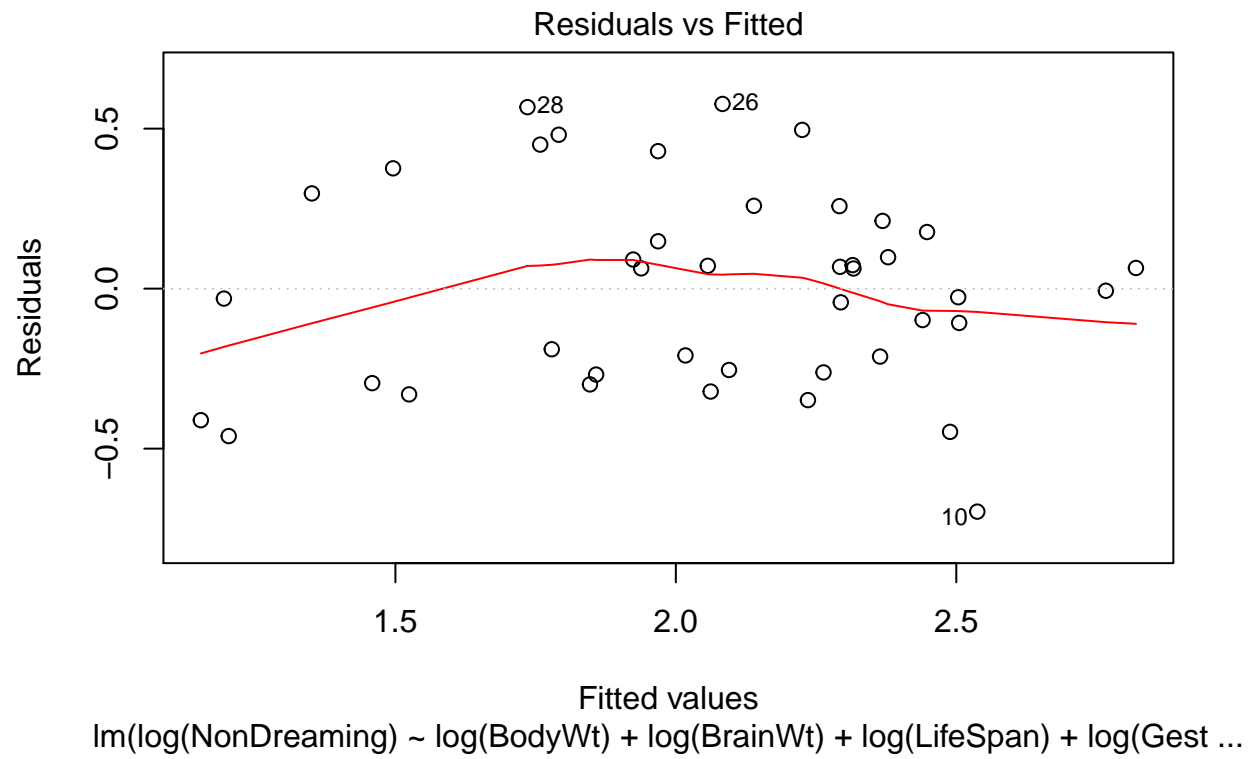
I would say that it is a toss up. The model is simpler because it has less predictors, which I think is a good thing in general, but the adjusted R^2 decreased, which is an indication that model3c may have less predictive power. There is also only one significant predictor in this model where the model3b had 2.

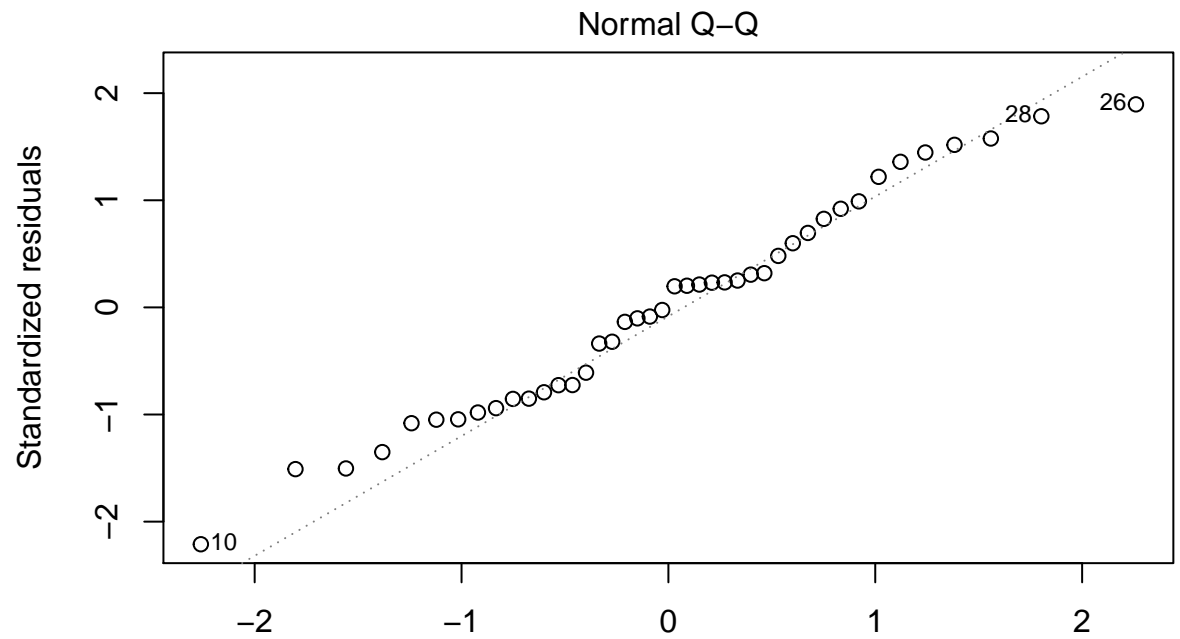
Question 3d: For our final model, let's attempt to improve the data assumptions and model predictability by adding back the transformation of the response variable, NonDreaming, using model3c as your baseline.

```
finalmodel <- lm(log(NonDreaming) ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) + log(Gestation) + Danger, data = data)

summary(finalmodel)

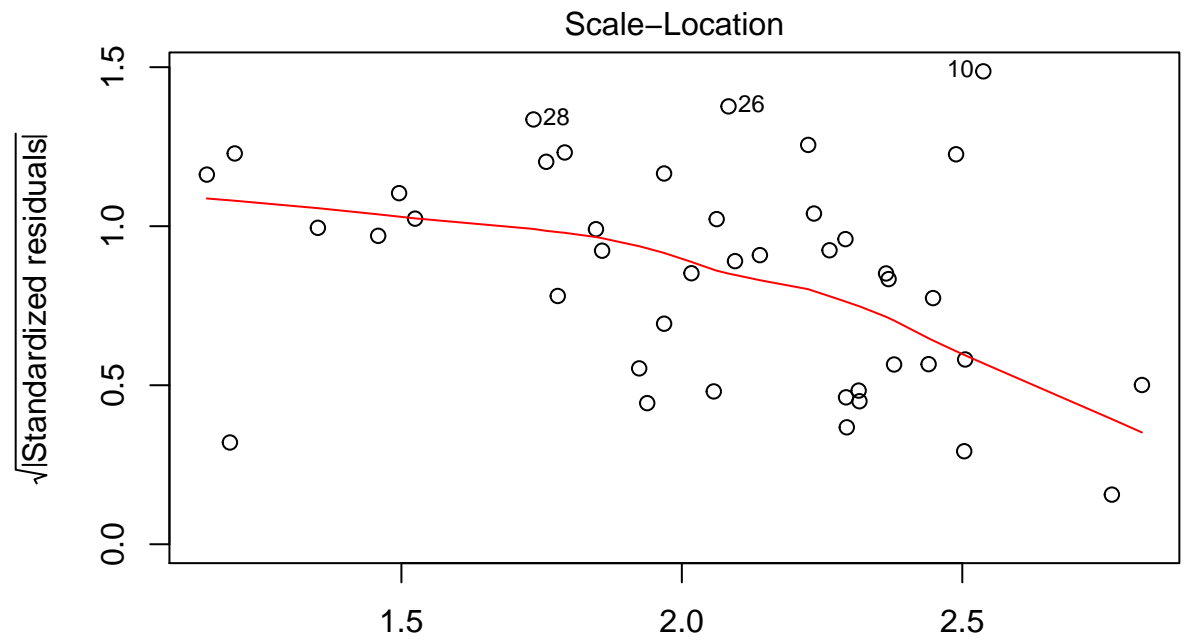
##
## Call:
## lm(formula = log(NonDreaming) ~ log(BodyWt) + log(BrainWt) +
##     log(LifeSpan) + log(Gestation) + Danger, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.69689 -0.25990  0.02811  0.20292  0.57709
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.78927    0.32012   8.713 2.15e-10 ***
## log(BodyWt)  -0.11203    0.05881  -1.905  0.06478 .
## log(BrainWt)   0.03583    0.08838   0.405  0.68756
## log(LifeSpan)  0.07153    0.08660   0.826  0.41425
## log(Gestation) -0.13977    0.08023  -1.742  0.09003 .
## Danger        -0.11576    0.04097  -2.825  0.00766 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3316 on 36 degrees of freedom
## Multiple R-squared:  0.643, Adjusted R-squared:  0.5934
## F-statistic: 12.97 on 5 and 36 DF, p-value: 3.036e-07
plot(finalmodel, cook.levels = c(4/42,0.5,1))
```



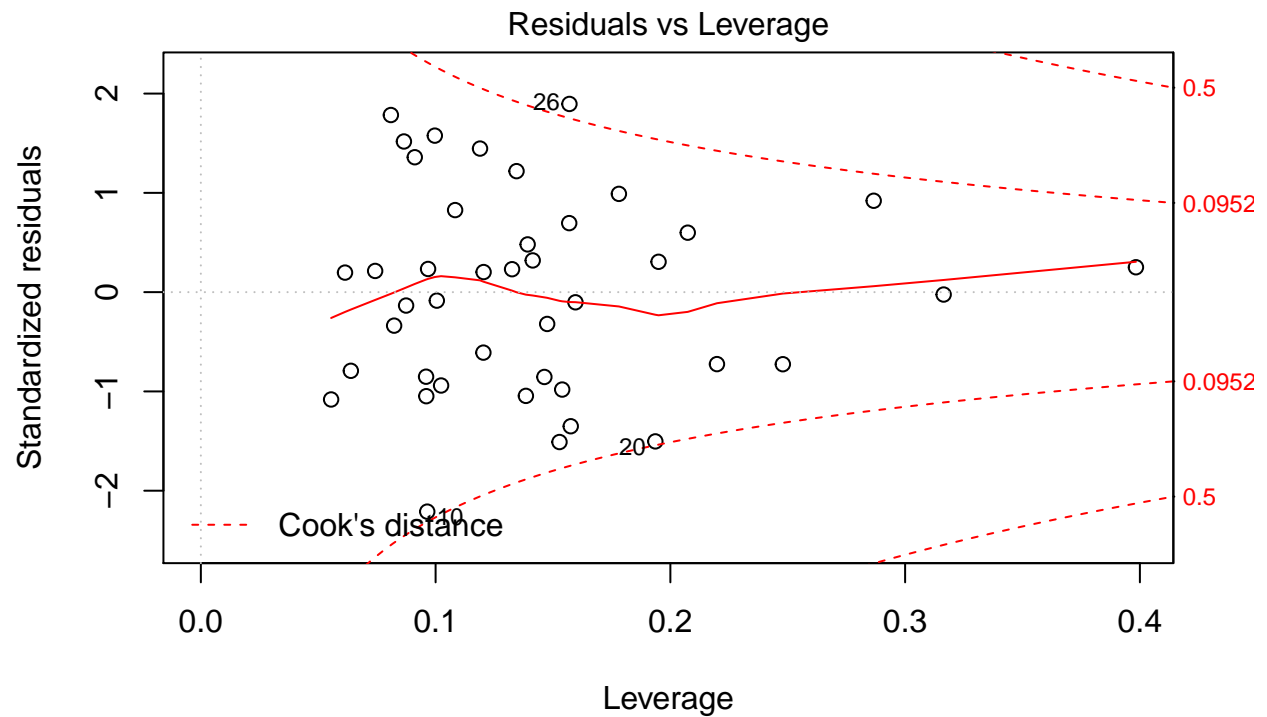


Theoretical Quantiles

$\ln(\log(\text{NonDreaming})) \sim \log(\text{BodyWt}) + \log(\text{BrainWt}) + \log(\text{LifeSpan}) + \log(\text{Gest ...})$

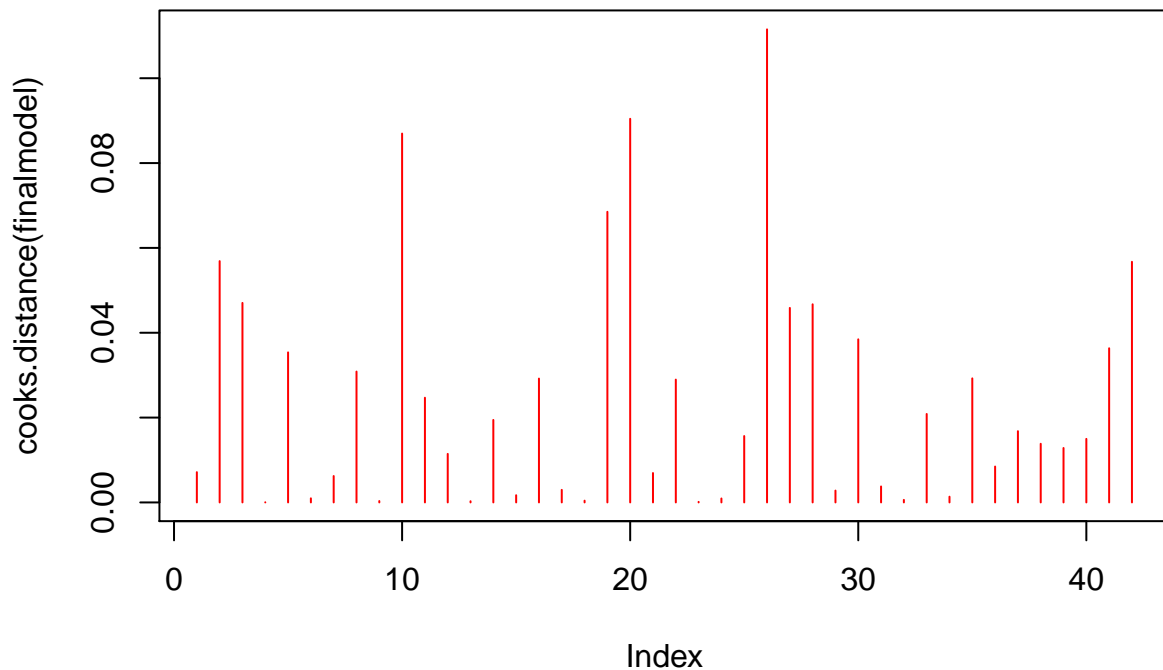


Fitted values
 $\text{lm}(\log(\text{NonDreaming}) \sim \log(\text{BodyWt}) + \log(\text{BrainWt}) + \log(\text{LifeSpan}) + \log(\text{Gest ...})$



$\text{lm}(\log(\text{NonDreaming}) \sim \log(\text{BodyWt}) + \log(\text{BrainWt}) + \log(\text{LifeSpan}) + \log(\text{Gest ...})$

```
plot(cooks.distance(finalmodel), type='h', col='red')
```



1. What are the model parameters and what are their estimates?

The model parameters with their estimates are $\log(\text{BodyWt})$ (-0.11203), $\log(\text{BrainWt})$ (0.03583), $\log(\text{LifeSpan})$ (0.07153), $\log(\text{Gestation})$ (-0.13977), and Danger (-0.11576)

2. What is the equation for the regression line?

$$\log(y) = 2.78927 + \log(\text{BodyWt})(-0.11203) + \log(\text{BrainWt})(0.03583) + \log(\text{LifeSpan})(0.07153) + \log(\text{Gestation})(-0.13977) + \text{Danger}(-0.11576)$$

3. Which predicting variable(s) are significant at $\alpha = 0.05$? What are their p-values?

Danger is the only predictor with a p value $< .05$ at 0.00766.

4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The $\log(\text{BodyWt})$ parameter is -0.11203. As BodyWt increases by 1% nondreaming will decrease by -0.11203%, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for $\log(\text{BodyWt})$ is 0.05881. So our prediction, on average will be off by 0.05881.

The Danger parameter is -0.11576. As Danger increases by 1 unit non-dreaming decreases by $-0.11576 \times 100\%$, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for Danger is 0.3249. So our prediction, on average will be off by 0.3249.

5. Did finalmodel improve over model3c? Explain how you determined if the model improved or not.

I would say that the final model did improve over the model3c. The model has 3 parameters significant at 0.1 and the R^2 value is much greater at 0.64 vs 0.59.

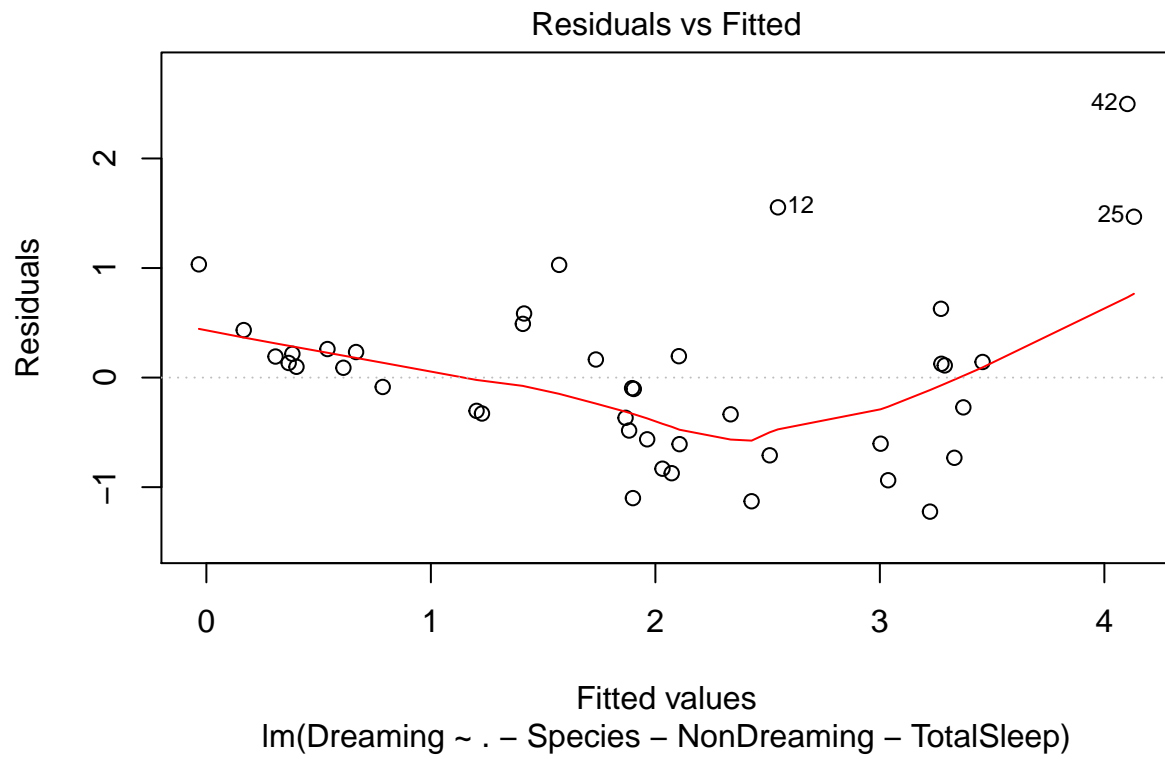
```
data2 = data[-11, ]
```

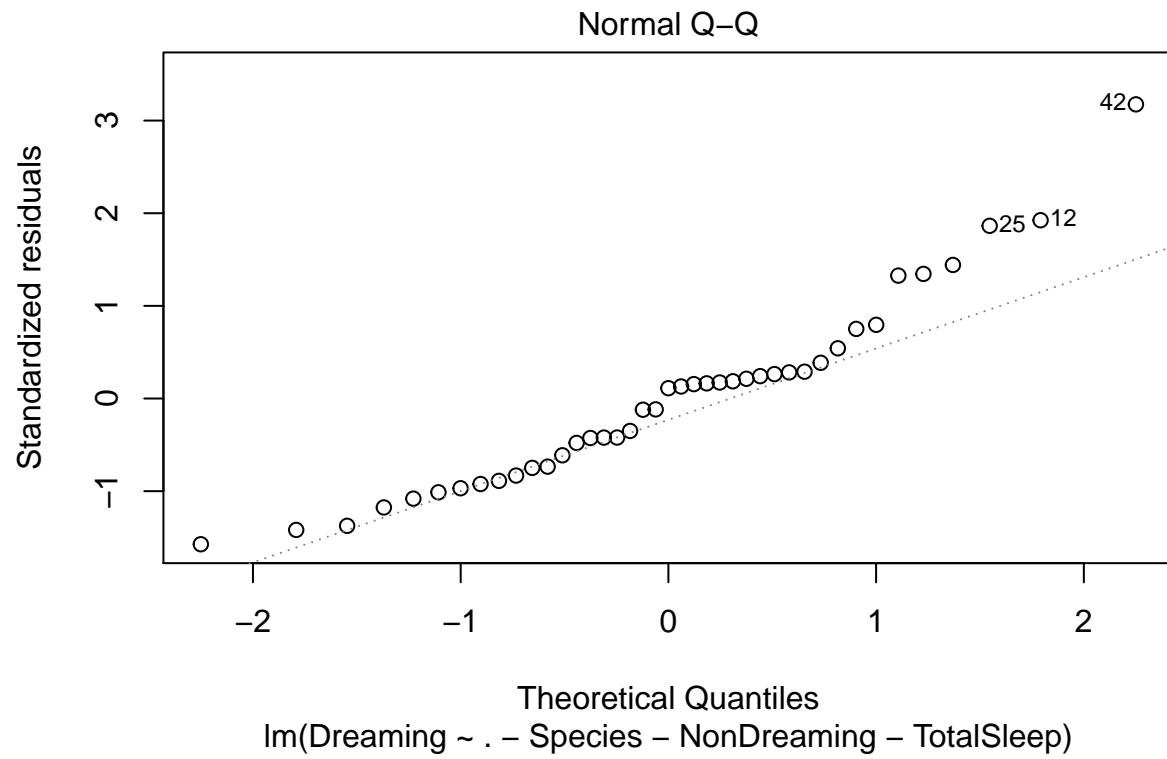
Question 4: Fitting the Linear Regression Model.

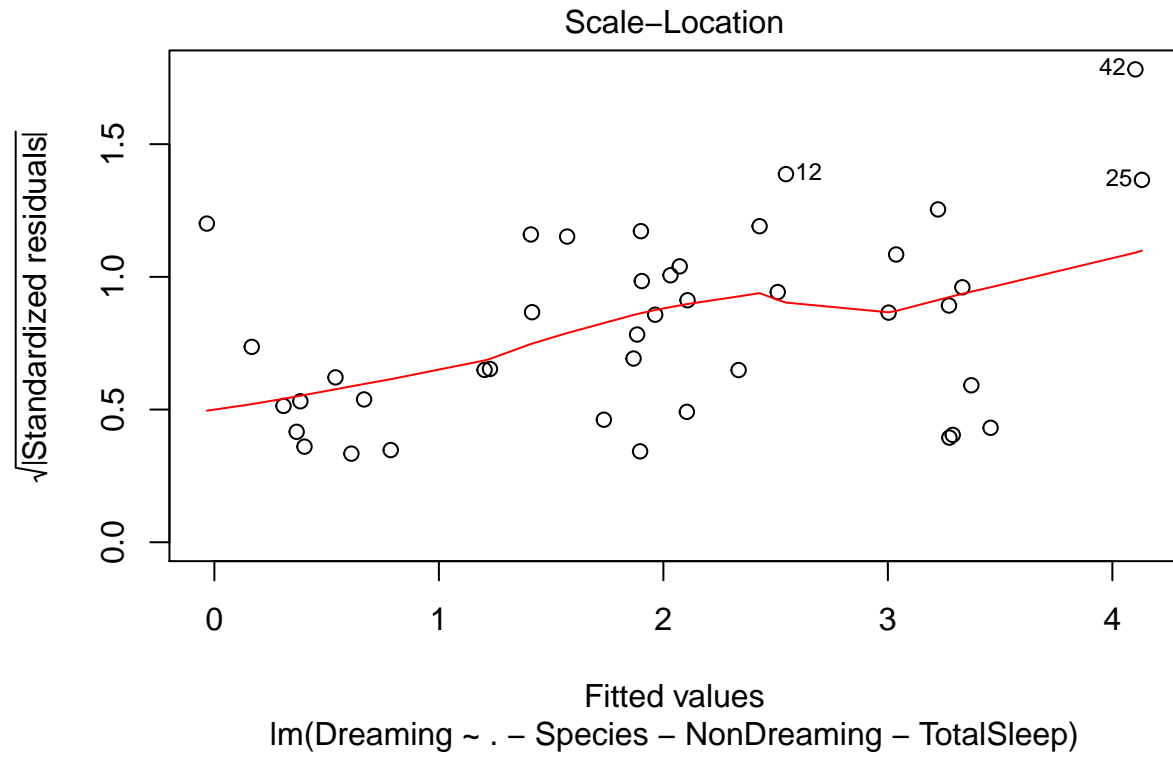
```
model4 <- lm(Dreaming ~. -Species -NonDreaming -TotalSleep, data = data2)
```

```
summary(model4)
```

```
##
## Call:
## lm(formula = Dreaming ~ . - Species - NonDreaming - TotalSleep,
##     data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22331 -0.56344  0.08977  0.23318  2.49739
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.820083   0.372492  10.255 8.57e-12 ***
## BodyWt       0.003615   0.001813   1.994  0.05444 .
## BrainWt      -0.001041   0.001086  -0.958  0.34505
## LifeSpan     0.011875   0.015420   0.770  0.44674
## Gestation    -0.007219   0.002081  -3.470  0.00147 **
## Predation     0.859964   0.304644   2.823  0.00800 **
## Exposure     0.295094   0.191910   1.538  0.13367
## Danger       -1.675645   0.378480  -4.427 9.86e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8463 on 33 degrees of freedom
## Multiple R-squared:  0.6865, Adjusted R-squared:  0.6199
## F-statistic: 10.32 on 7 and 33 DF,  p-value: 8.706e-07
plot(model4, cook.levels = c(4/42,0.5,1))
```



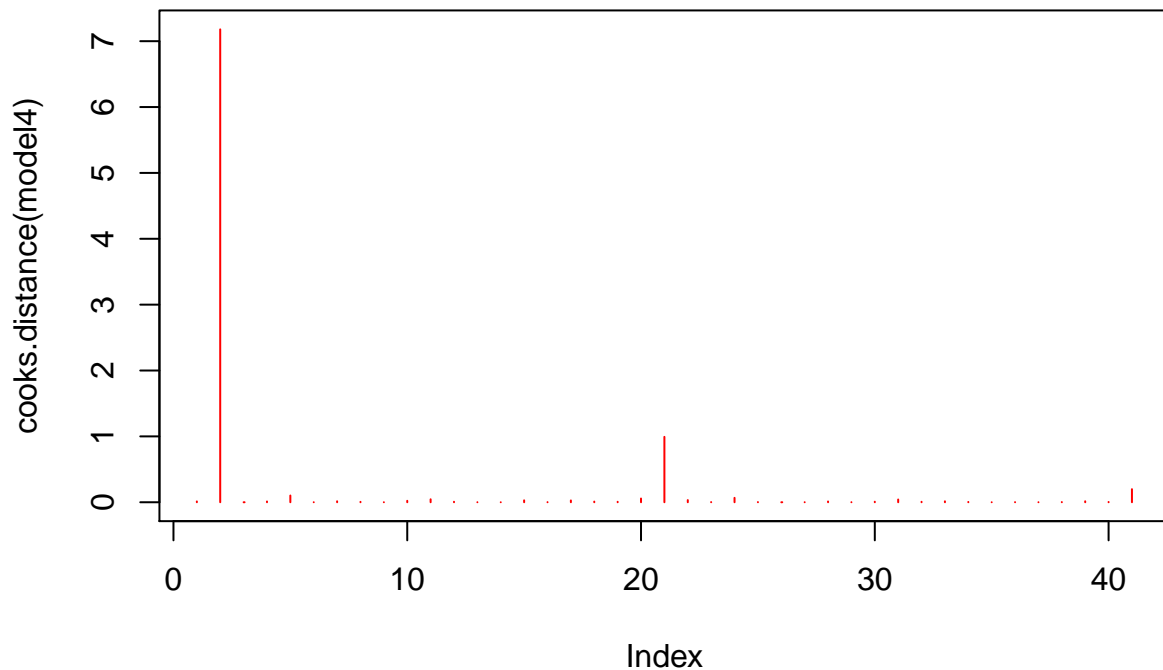




```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

1. What are the model parameters and what are their estimates?

The model parameters with their estimates are intercept 3.820083, BodyWt (0.003615), BrainWt(-0.001041), LifeSpan(0.011875), Gestation(-0.007219), Predation(0.859964), Exposure(0.295094), and Danger (-1.675645)

2. What is the equation for the regression line?

$$y = 3.820083 + 0.003615 \text{BodyWt} - 0.001041 \text{BrainWt} + 0.011875 \text{LifeSpan} - 0.007219 \text{Gestation} + 0.859964 \text{Predation} + 0.295094 \text{Exposure} - 1.675645 \text{Danger}$$

3. Which predicting variable(s) are significant at $\alpha = 0.05$? What are their p-values?

Gestation Predation and Danger are the only predictors with p values $< .05$ at 0.00147, 0.00800 and 9.86×10^{-5} respectively.

4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The BodyWt parameter is 0.003615. This represents the estimated expected change in nonDreaming sleep with one unit of change in BodyWt, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for BodyWt is 0.001813. So our prediction, on average will be off by 0.001813.

The Predation parameter is 0.859964. This represents the estimated expected change in nonDreaming sleep with one unit of change in Predation, holding all other values constant. The standard error represents the

average distance that the observed values fall from the regression line. The standard error for Predation is 0.304644. So our prediction, on average will be off by 0.304644.

5. Check the assumptions of the model through plotting. Note potential outliers, if any.

Constant variance and uncorrelated errors

The residual vs fitted values above shows the errors scattered around the zero line, which indicates that the constant variance and uncorrelated errors assumption holds.

Normality

The q-q plot of the standardized residuals shown above shows that the normality assumption skews to the right, but I would say it still holds with

Outliers

The last plot, the cook's distance notes one outlier.

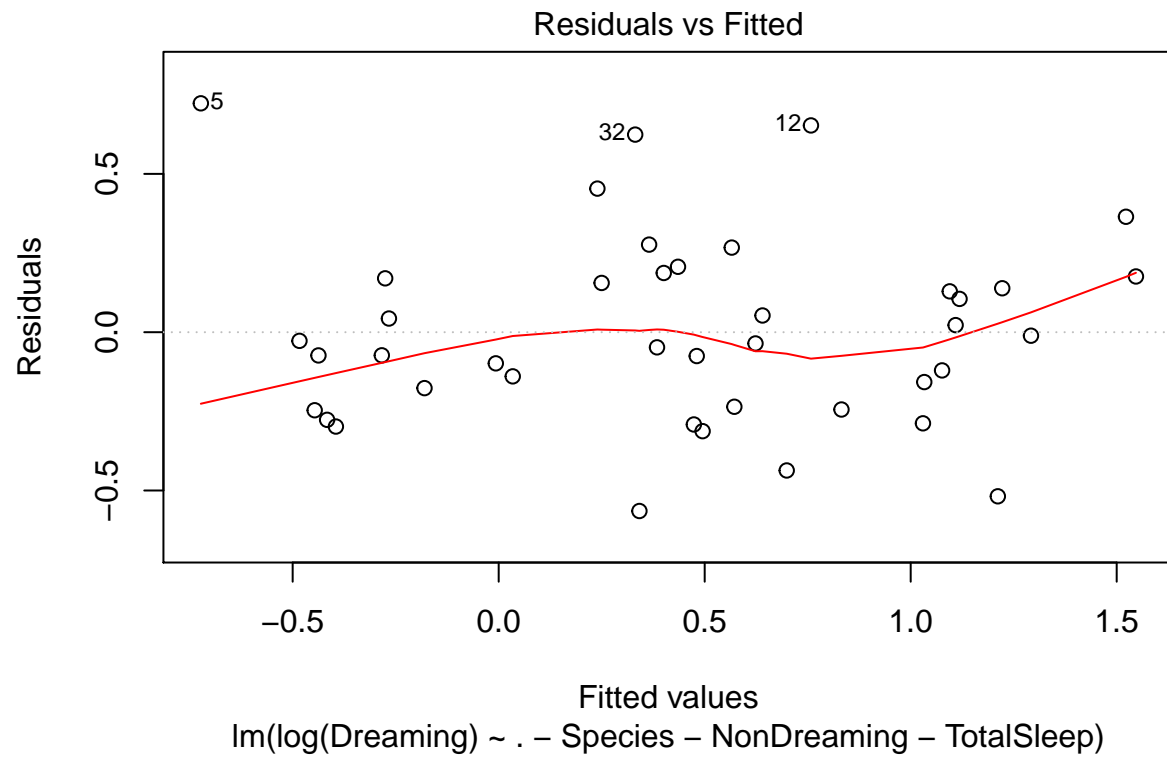
Question 4a: Change model4 to log transform the response variable, NonDreaming.

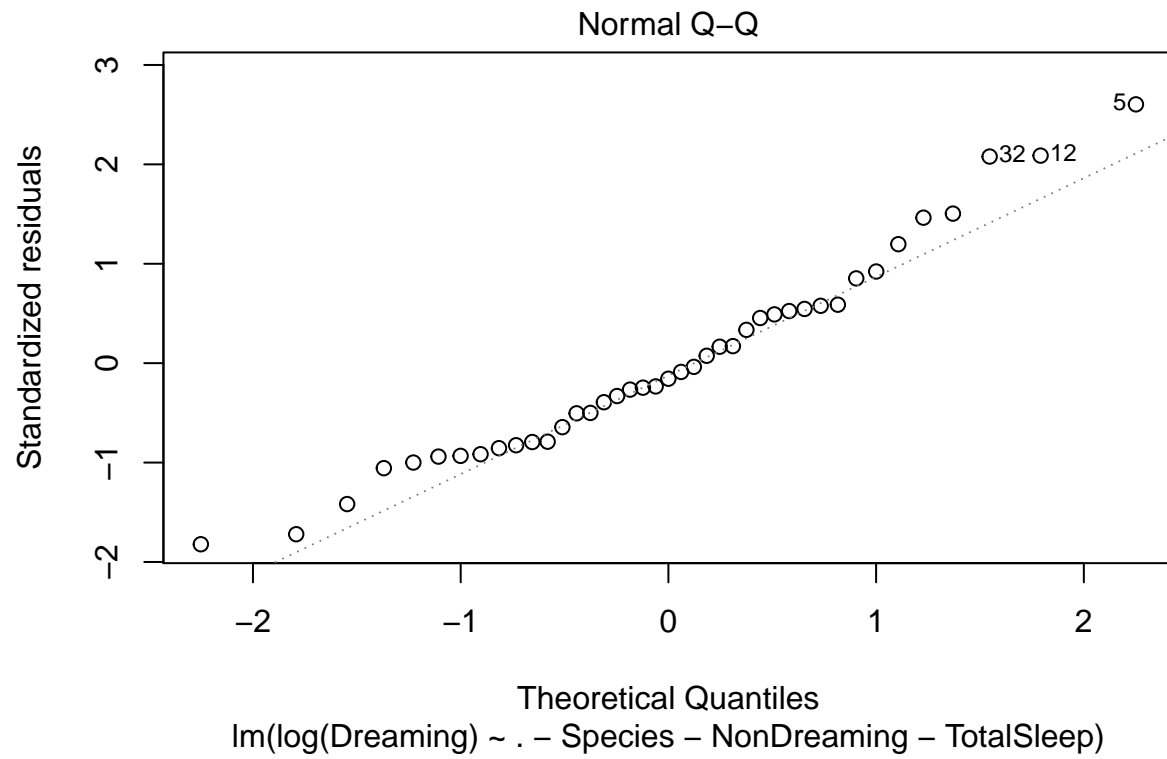
```
model4a <- lm(log(Dreaming) ~. -Species -NonDreaming -TotalSleep, data = data2)
```

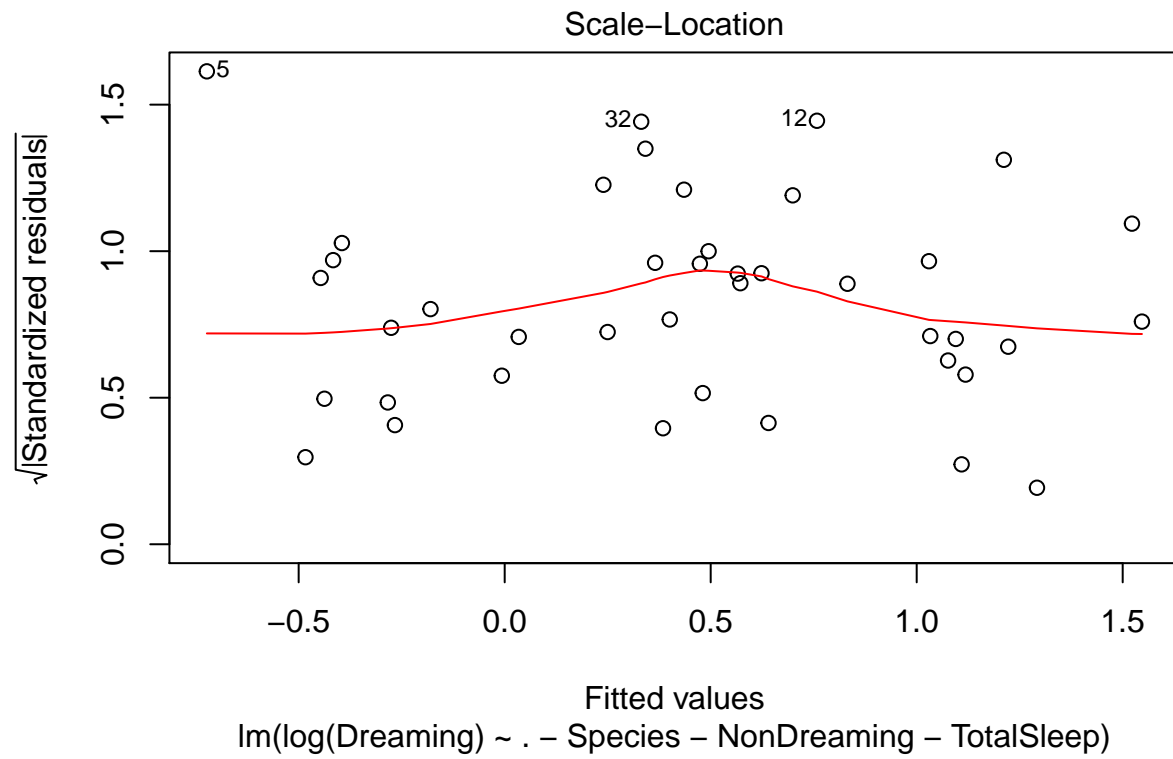
```
summary(model4a)
```

```
##
## Call:
## lm(formula = log(Dreaming) ~ . - Species - NonDreaming - TotalSleep,
##     data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56473 -0.23549 -0.03553  0.17015  0.72288
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.4684458  0.1442539  10.180 1.03e-11 ***
## BodyWt       0.0020701  0.0007020   2.949  0.00582 **
## BrainWt      -0.0005974  0.0004208  -1.420  0.16503
## LifeSpan     0.0104864  0.0059717   1.756  0.08837 .
## Gestation    -0.0042079  0.0008057  -5.223 9.57e-06 ***
## Predation     0.3747033  0.1179789   3.176  0.00323 **
## Exposure     0.1015576  0.0743207   1.366  0.18103
## Danger       -0.7743707  0.1465730  -5.283 8.00e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3277 on 33 degrees of freedom
## Multiple R-squared:  0.8066, Adjusted R-squared:  0.7656
## F-statistic: 19.67 on 7 and 33 DF,  p-value: 4.342e-10
```

```
plot(model4a, cook.levels = c(4/42,0.5,1))
```



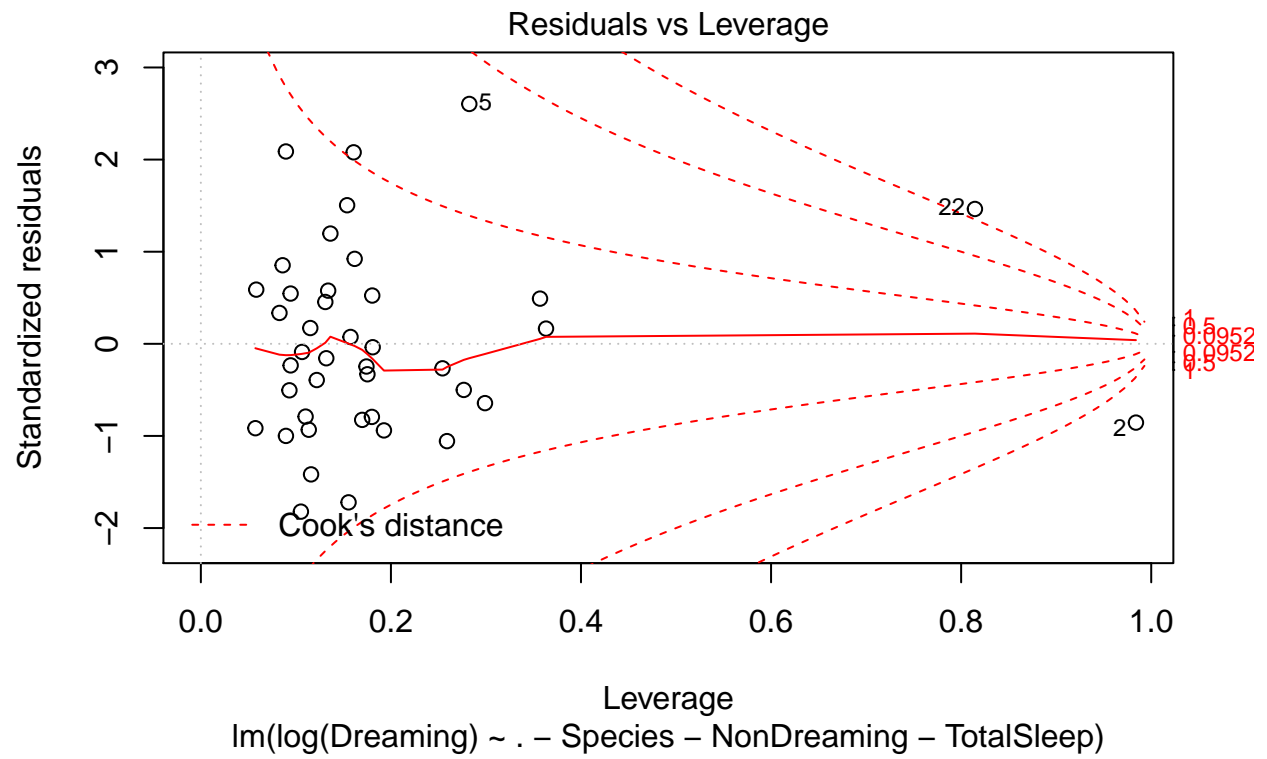




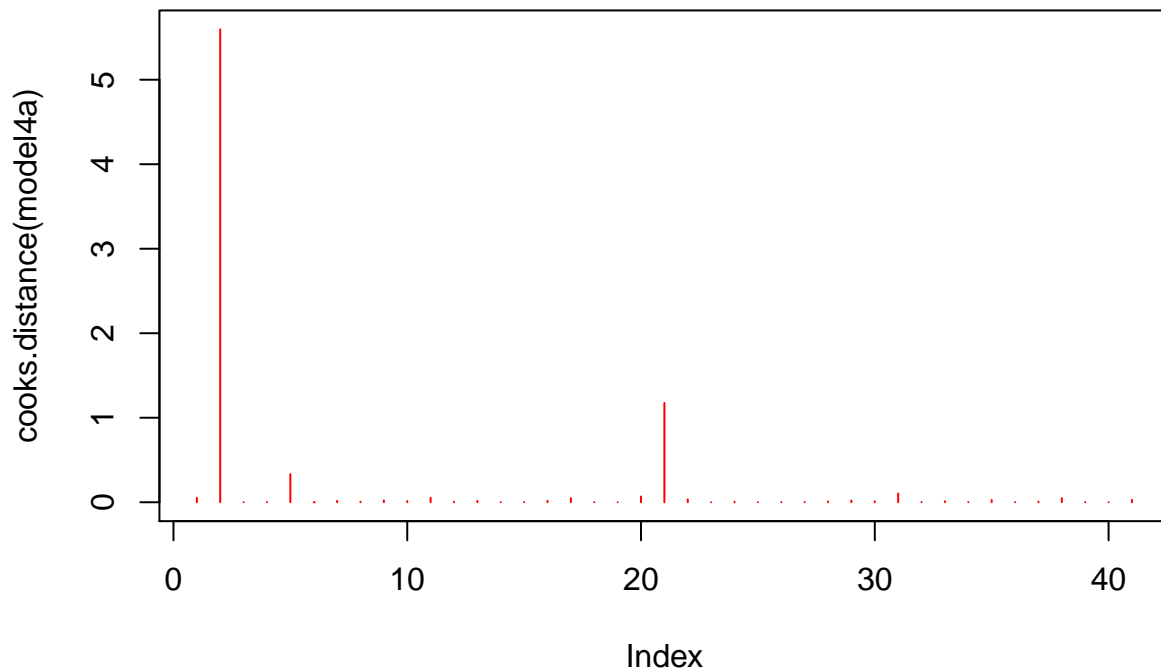
```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



```
plot(cooks.distance(model4a), type='h', col='red')
```



1. What are the model parameters and what are their estimates?

The model parameters with their estimates are intercept 1.4684458 BodyWt (0.0020701), BrainWt(-0.0005974), LifeSpan(0.0104864), Gestation(-0.0042079), Predation(0.3747033), Exposure(0.1015576), and Danger (0.7743707)

2. What is the equation for the regression line?

$$\log(y) = 1.4684458 + 0.0020701 \text{BodyWt} + -0.0005974 \text{BrainWt} + 0.0104864 \text{LifeSpan} + -0.0042079 \text{Gestation} + 0.3747033 \text{Predation} + 0.1015576 \text{Exposure} + 0.7743707 \text{Danger}$$

3. Which predicting variable(s) are significant at alpha = 0.05? What are their p-values?

BodyWt, gestation, Predation and Danger are all significant at the .05 level with p values at 0.00582, 9.57e-06, 0.00323, and 8.00e-06 respectively.

4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The BodyWt parameter is 0.0020701. As BodyWt increases by 1 unit, nondreaming increases by 0.0020701*100%, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for BodyWt is 0.0007020. So our prediction, on average will be off by 0.0007020.

The Predation parameter is 0.3747033. This represents the estimated expected change in nonDreaming sleep with one unit of change in Predation, holding all other values constant. The standard error represents the

average distance that the observed values fall from the regression line. The standard error for Predation is 0.1179789. So our prediction, on average will be off by 0.1179789.

5. Check the assumptions of the model through plotting. Note potential outliers, if any.

Constant variance and uncorrelated errors

The residual vs fitted values above shows the errors scattered around the zero line, which indicates that the constant variance and uncorrelated errors assumption holds.

Normality

The q-q plot of the standardized residuals shown above shows that the normality assumption holds walthough it is right skewed.

Outliers

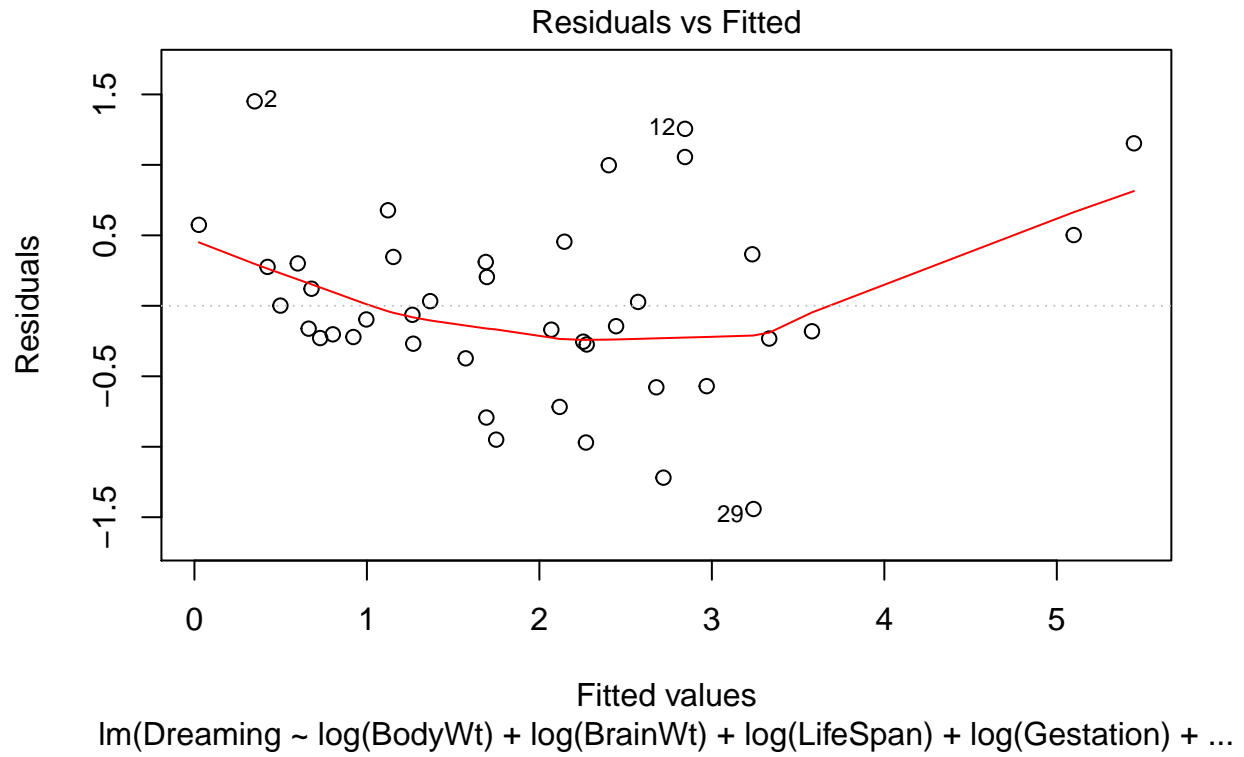
The last plot, the cook's distance notes one outlier.

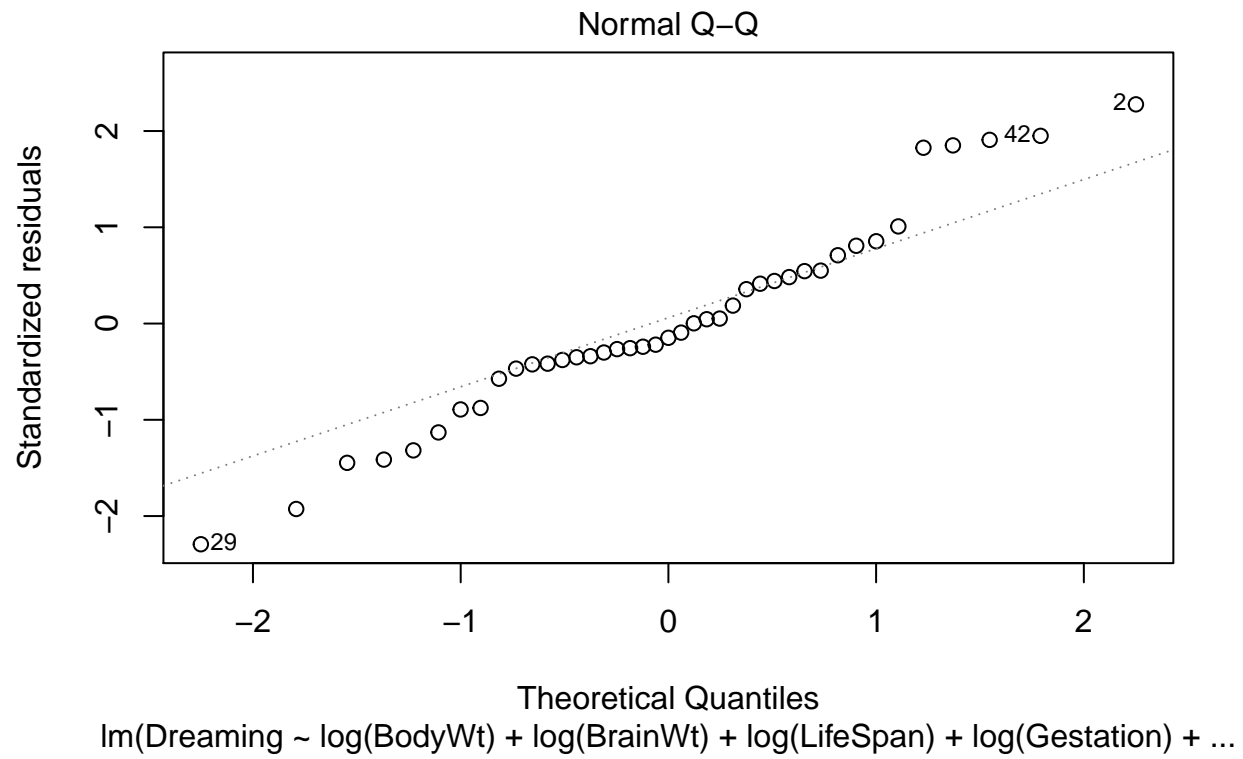
Question 4b: Change model4a to remove the log transform of NonDreaming, and add the log transformation of numeric response variables BrainWt, BodyWt, LifeSpan and Gestation

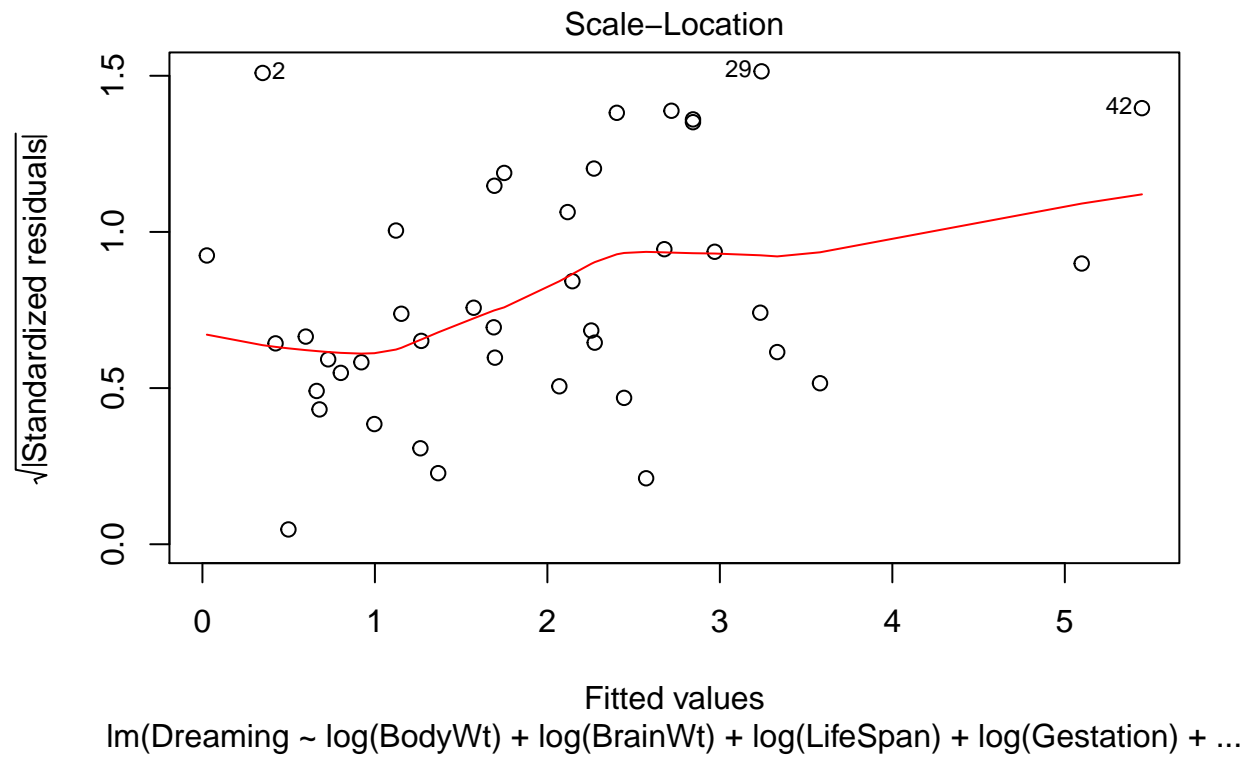
```
model4b <- lm(Dreaming ~ log(BodyWt) +log(BrainWt) +log(LifeSpan) +log(Gestation) +Exposure +Predation +
summary(model4b)

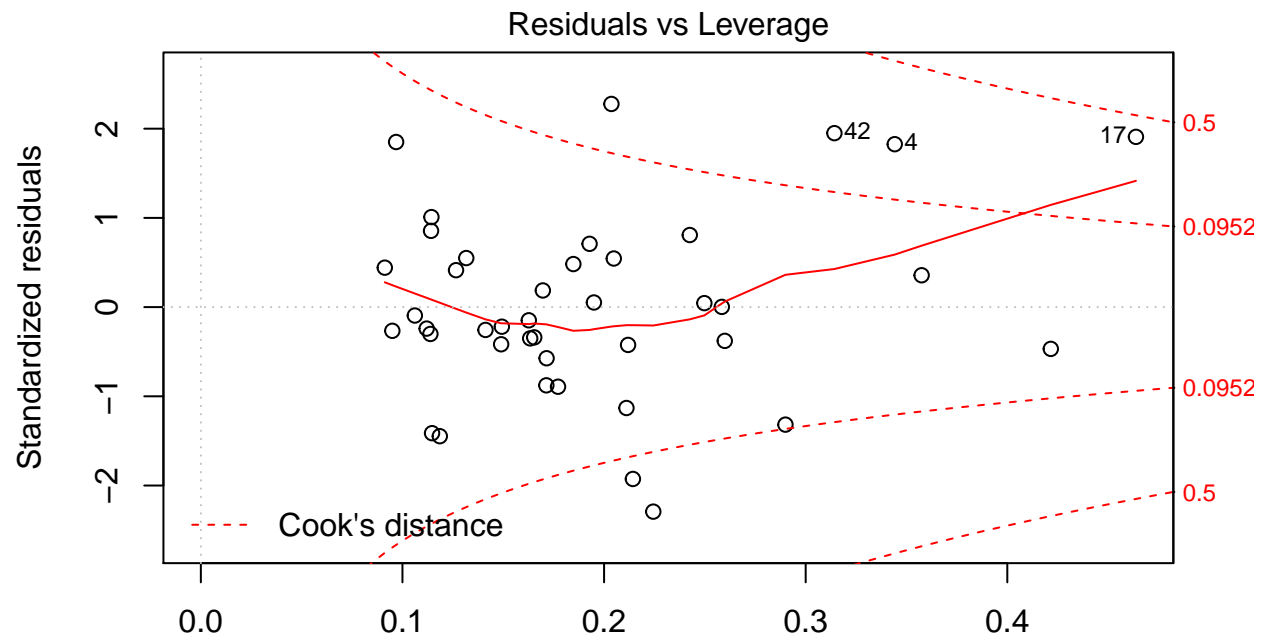
##
## Call:
## lm(formula = Dreaming ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) +
##     log(Gestation) + Exposure + Predation + Danger, data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44161 -0.26871 -0.09673  0.34675  1.45088
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.43127    0.80648   9.214 1.21e-10 ***
## log(BodyWt)     0.44017    0.13251   3.322 0.002195 **
## log(BrainWt)   -0.35662    0.19429  -1.836 0.075449 .
## log(LifeSpan)   0.02462    0.21895   0.112 0.911153
## log(Gestation) -0.82406    0.19193  -4.294 0.000145 ***
## Exposure        0.26488    0.17167   1.543 0.132367
## Predation       0.59634    0.27182   2.194 0.035392 *
## Danger        -1.36005    0.32232  -4.220 0.000180 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.714 on 33 degrees of freedom
## Multiple R-squared:  0.7768, Adjusted R-squared:  0.7295
## F-statistic: 16.41 on 7 and 33 DF,  p-value: 4.24e-09
```

```
plot(model4b, cook.levels = c(4/42,0.5,1))
```



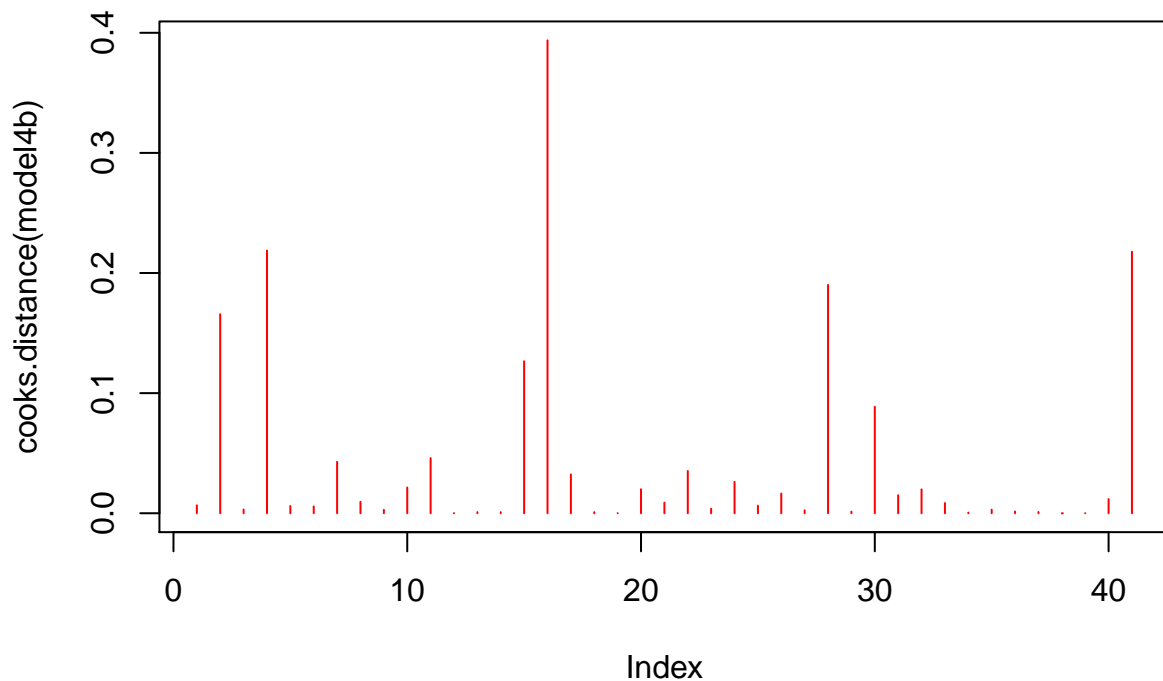






Im(Dreaming ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) + log(Gestation) + ...

```
plot(cooks.distance(model4b), type='h', col='red')
```



1. What are the model parameters and what are their estimates?

The model parameters with their estimates are intercept 7.43127 $\log(\text{BodyWt})$ (0.44017), $\log(\text{BrainWt})$ (-0.35662), $\log(\text{LifeSpan})$ (0.02462), $\log(\text{Gestation})$ (-0.82406), Predation(0.59634), Exposure(0.26488), and Danger (-1.36005)

2. What is the equation for the regression line?

$$y = 7.43127 + \log(\text{BodyWt})(0.44017) + \log(\text{BrainWt})(-0.35662) + \log(\text{LifeSpan})(0.02462) + \log(\text{Gestation})(-0.82406) + \text{Predation}(0.59634) + \text{Exposure}(0.26488) + \text{Danger}(-1.36005)$$

3. Which predicting variable(s) are significant at $\alpha = 0.05$? What are their p-values?

$\log(\text{BodyWt})$, $\log(\text{Gestation})$, Predation and Danger are all significant at the .05 level with p-values of 0.002195, 0.000145, 0.035392, and 0.000180 respectively

4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The $\log(\text{BodyWt})$ parameter is 0.44017. As BodyWt increases by 1% nondreaming will increase by 0.44017/100, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for $\log(\text{BodyWt})$ is 0.13251. So our prediction, on average will be off by 0.13251.

The Predation parameter is 1.6141. This represents the estimated expected change in nonDreaming sleep with one unit of change in Predation, holding all other values constant. The standard error represents the

average distance that the observed values fall from the regression line. The standard error for Predation is 0.9812. So our prediction, on average will be off by 0.9812.

5. Did model3b improve over model3a? Explain how you determined if the model improved or not.

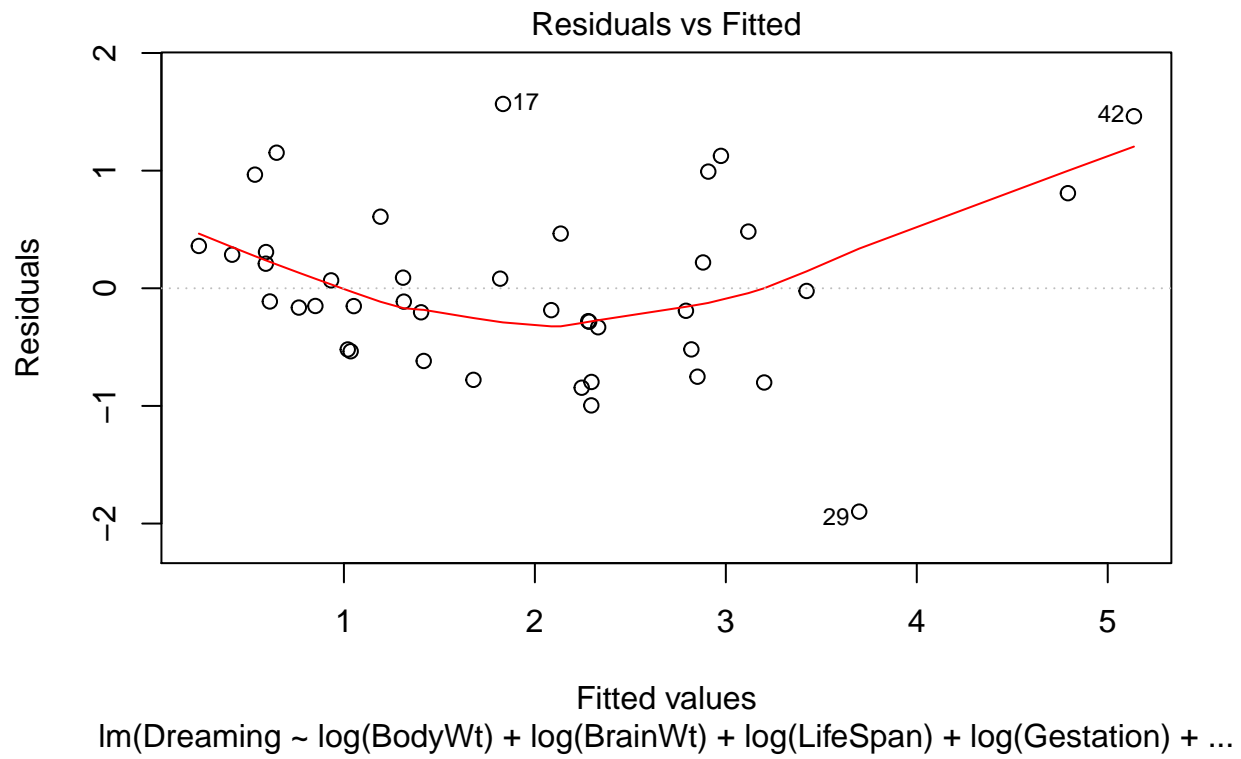
I would say that it did not improve because the R^2 value decreased 0.7768 compared to 0.8066

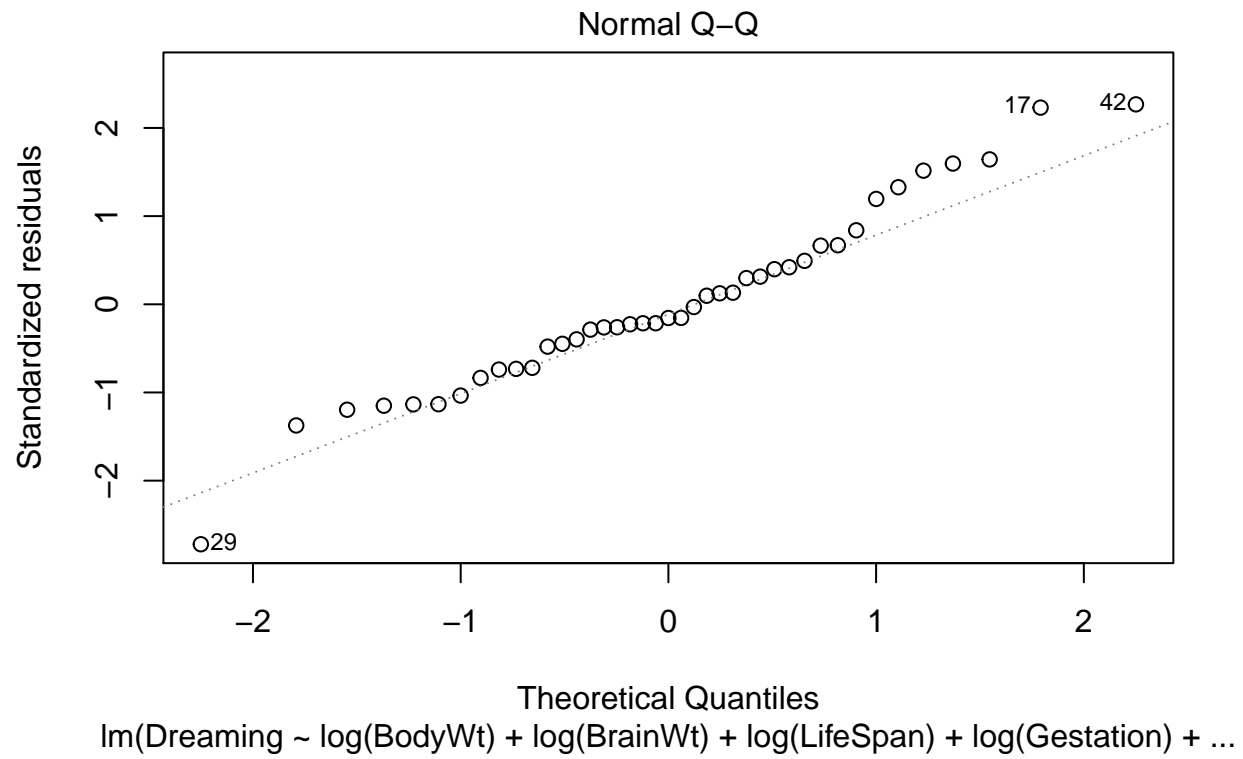
Question 4c: Because the Danger variable is an interpolation of the Exposure and Predation variables, let's keep Danger and remove the other two from the model using model3b as your baseline.

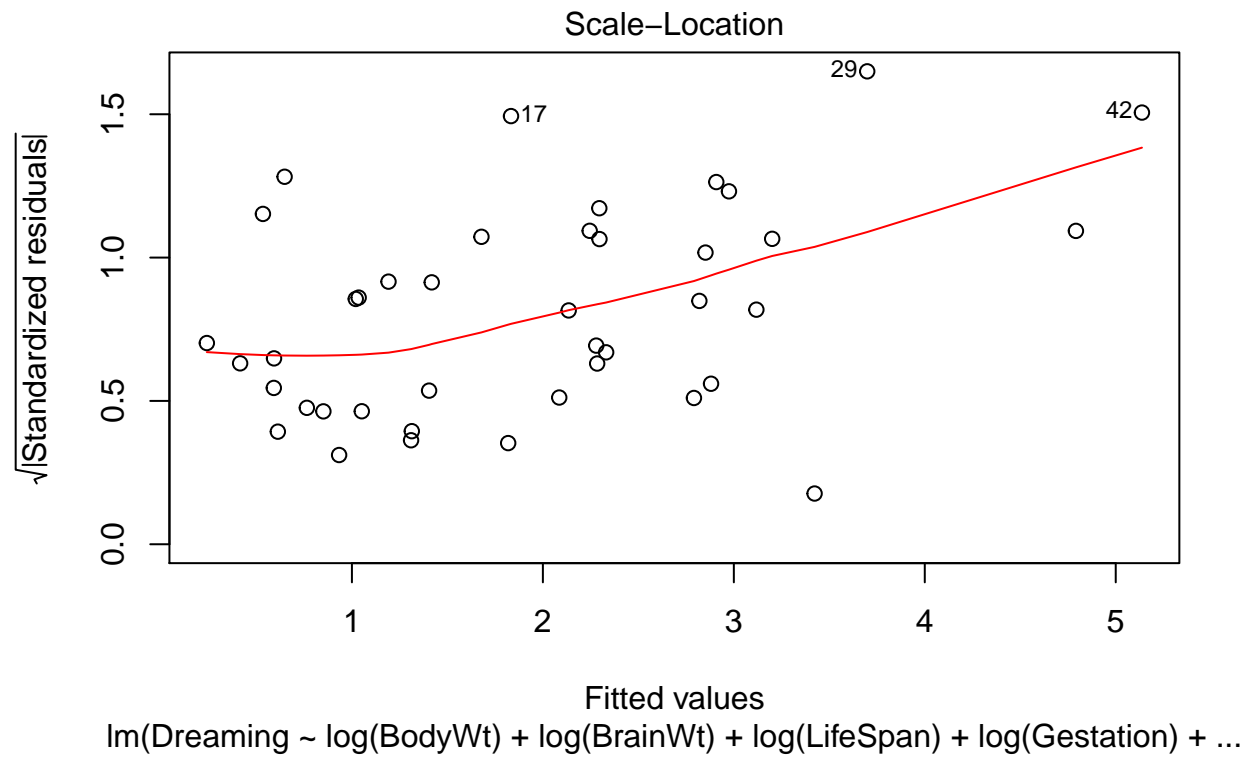
```
model4c <- lm(Dreaming ~ log(BodyWt) +log(BrainWt) +log(LifeSpan) +log(Gestation) +Danger, data = data2)

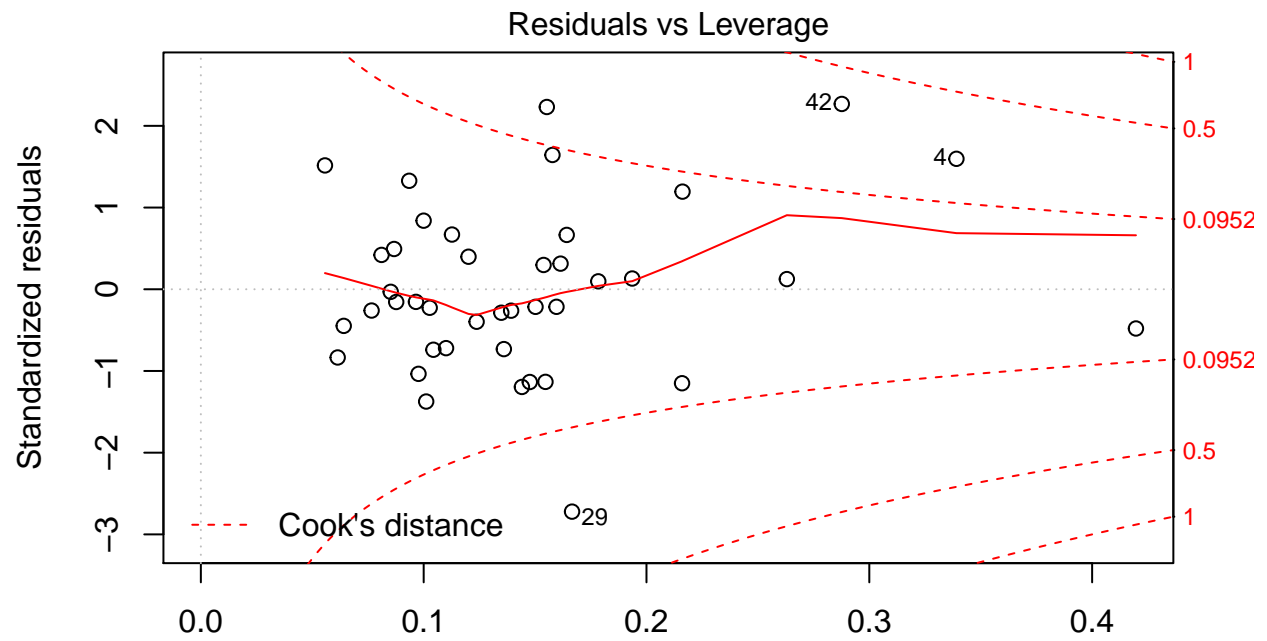
summary(model4c)

##
## Call:
## lm(formula = Dreaming ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) +
##     log(Gestation) + Danger, data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8985 -0.5195 -0.1136  0.3598  1.5668
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.94819    0.74532  10.664 1.53e-12 ***
## log(BodyWt)     0.43897    0.13553   3.239 0.00263 **
## log(BrainWt)   -0.32370    0.20370  -1.589 0.12103
## log(LifeSpan)  -0.02089    0.21328  -0.098 0.92252
## log(Gestation) -0.88941    0.19472  -4.568 5.88e-05 ***
## Danger         -0.55009    0.09443  -5.825 1.31e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7642 on 35 degrees of freedom
## Multiple R-squared:  0.7288, Adjusted R-squared:  0.6901
## F-statistic: 18.82 on 5 and 35 DF,  p-value: 4.721e-09
plot(model4c, cook.levels = c(4/42,0.5,1))
```



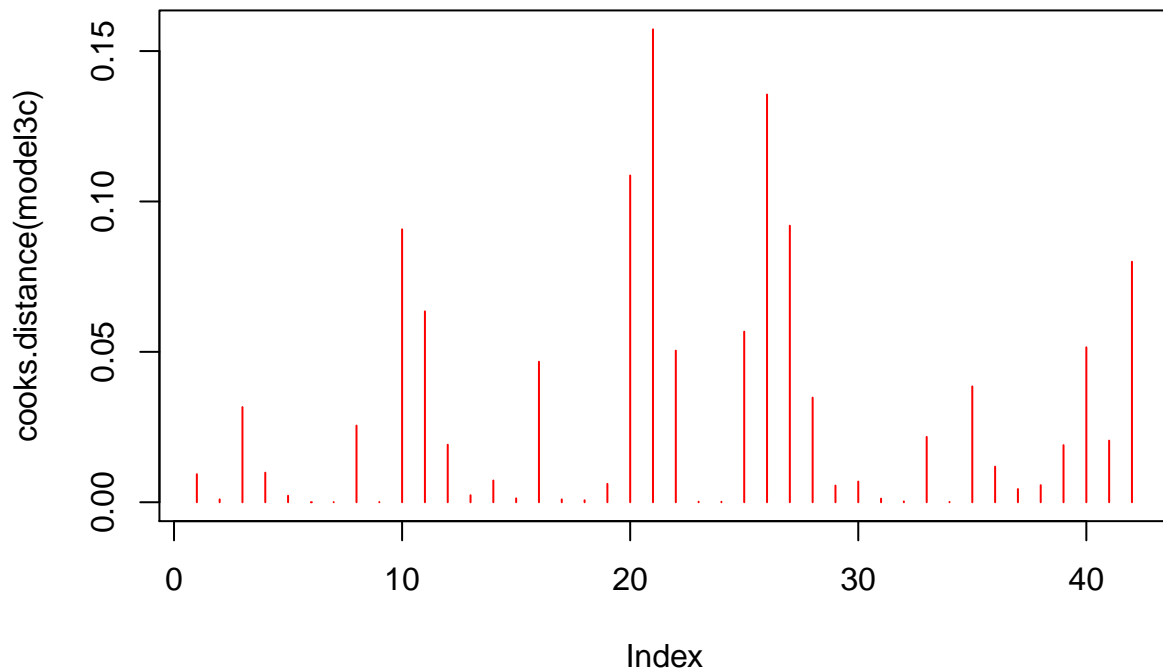






Im(Dreaming ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) + log(Gestation) + ...

```
plot(cooks.distance(model3c), type='h', col='red')
```



1. What are the model parameters and what are their estimates?

The model parameters with their estimates are intercept 7.94819, log(BodyWt) (0.43897), log(BrainWt) (-0.32370), log(LifeSpan) (-0.02089), log(Gestation) (-0.88941), and Danger (-0.55009)

2. What is the equation for the regression line?

$$y = 7.94819 + \log(\text{BodyWt})(0.43897) + \log(\text{BrainWt})(-0.32370) + \log(\text{LifeSpan})(-0.02089) + \log(\text{Gestation})(-0.88941) + \text{Danger}(-0.55009)$$

3. Which predicting variable(s) are significant at $\alpha = 0.05$? What are their p-values?

log(BodyWt), log(LifeSpan), and Danger are all significant at the .05 level with p-values at 0.00263, 5.88e-05, 1.31e-06.

4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The log(BodyWt) parameter is 0.43897 . As BodyWt increases by 1% nondreaming will decrease by 0.43897/100, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for log(BodyWt) is 0.13553. So our prediction, on average will be off by 0.13553.

The Danger parameter is -0.55009. This represents the estimated expected change in nonDreaming sleep with one unit of change in Danger, holding all other values constant. The standard error represents the average

distance that the observed values fall from the regression line. The standard error for Danger is 0.09443. So our prediction, on average will be off by 0.09443.

5. Did model3c improve over model3b? Explain how you determined if the model improved or not.

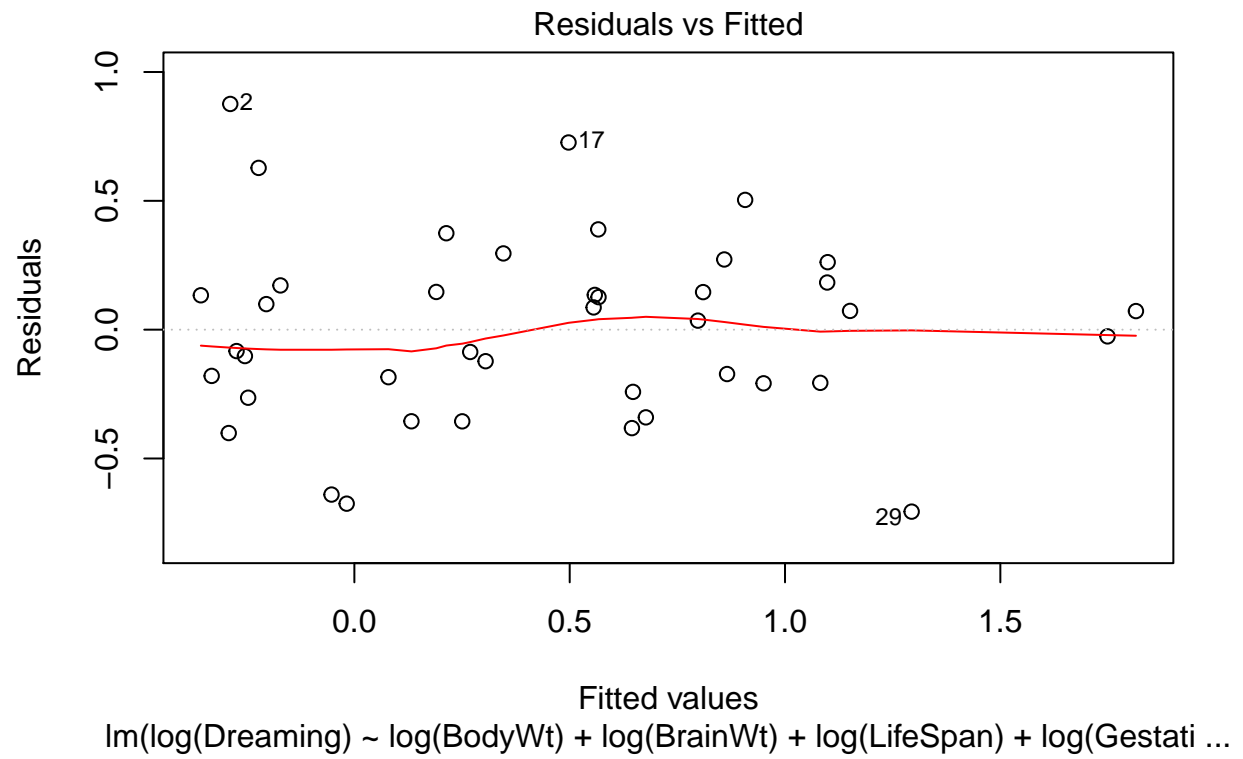
No because the adjusted R^2 decreased

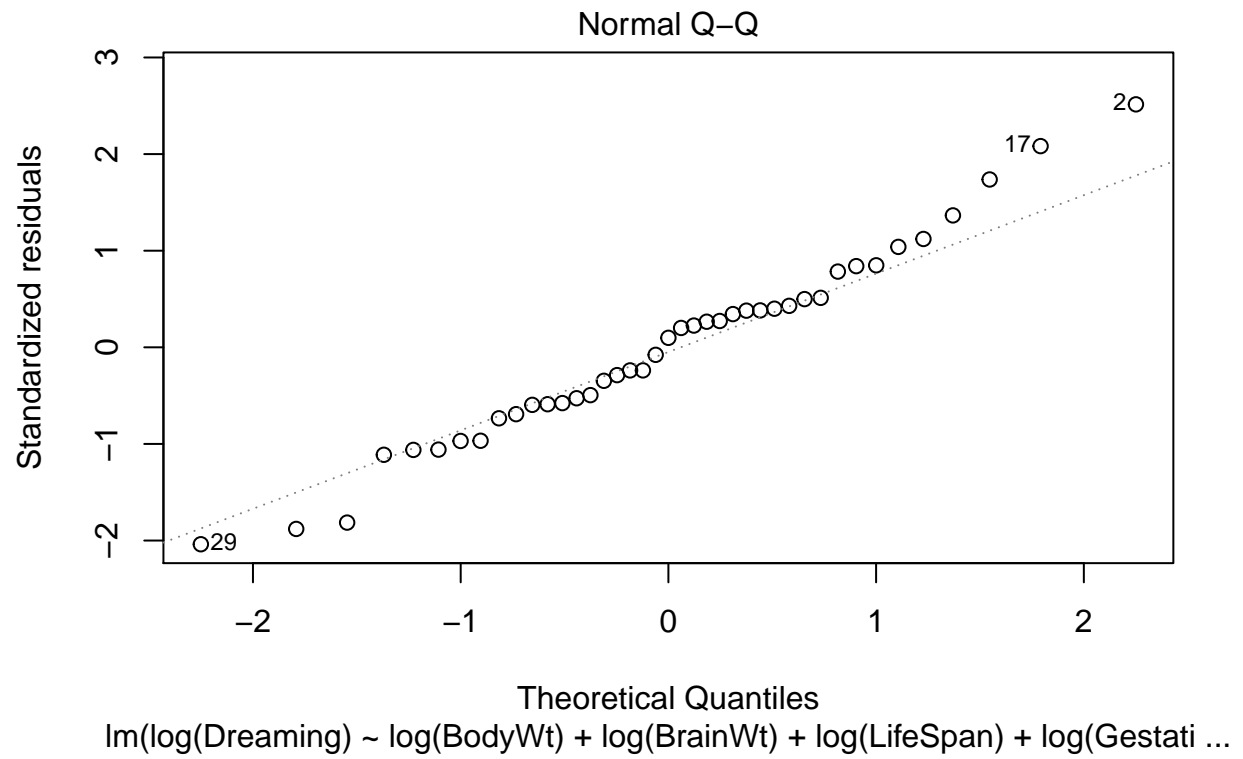
Question 4d: For our final model, let's attempt to improve the data assumptions and model predictability by adding back the transformation of the response variable, NonDreaming, using model3c as your baseline.

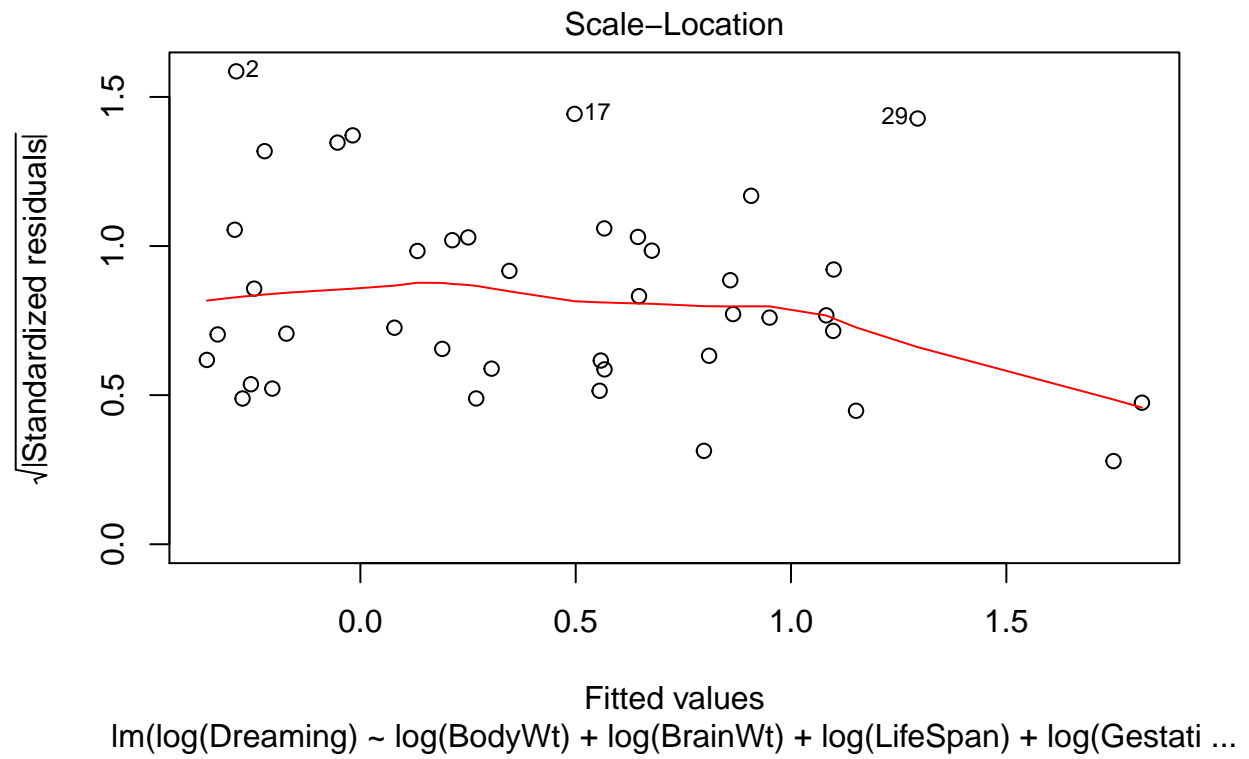
```
finalmodel <- lm(log(Dreaming) ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) + log(Gestation) + Danger, data = data2)

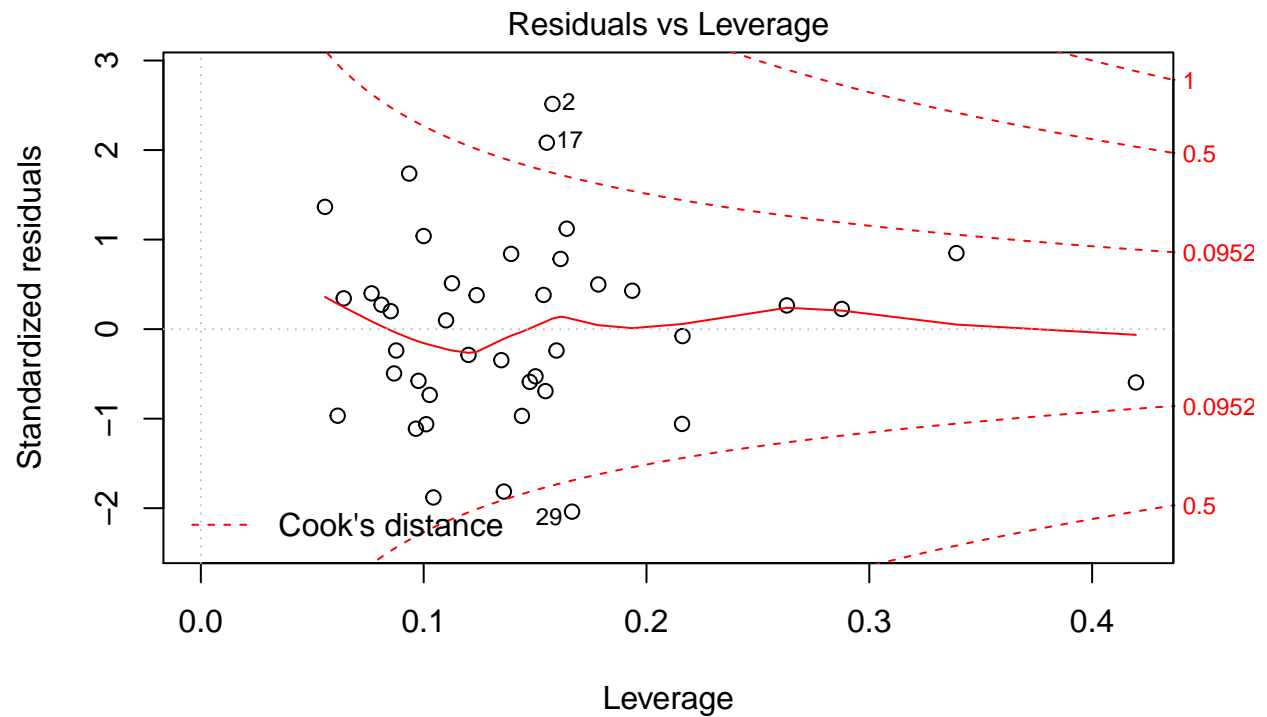
summary(finalmodel)

##
## Call:
## lm(formula = log(Dreaming) ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) +
##     log(Gestation) + Danger, data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70627 -0.20824  0.03513  0.17167  0.87600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.10346    0.37019   8.383 6.90e-10 ***
## log(BodyWt)    0.14340    0.06732   2.130 0.040252 *
## log(BrainWt)  -0.11064    0.10117  -1.094 0.281612
## log(LifeSpan)  0.05317    0.10593   0.502 0.618862
## log(Gestation) -0.41163    0.09671  -4.256 0.000148 ***
## Danger        -0.28973    0.04690  -6.177 4.51e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3796 on 35 degrees of freedom
## Multiple R-squared:  0.7249, Adjusted R-squared:  0.6856
## F-statistic: 18.45 on 5 and 35 DF,  p-value: 6.022e-09
plot(finalmodel, cook.levels = c(4/42,0.5,1))
```



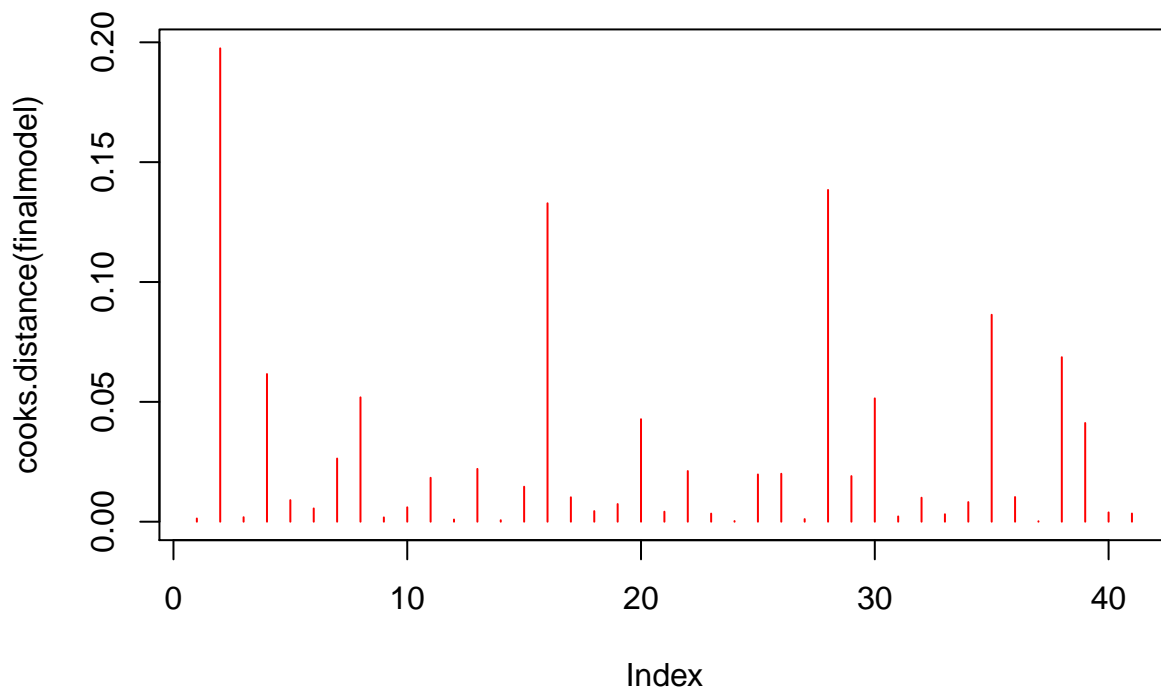






$\text{lm}(\log(\text{Dreaming}) \sim \log(\text{BodyWt}) + \log(\text{BrainWt}) + \log(\text{LifeSpan}) + \log(\text{Gestati} \dots$

```
plot(cooks.distance(finalmodel), type = 'h', col = 'red')
```



1. What are the model parameters and what are their estimates?

The model parameters with their estimates are intercept 3.10346, $\log(\text{BodyWt})$ (0.14340), $\log(\text{BrainWt})$ (-0.11064), $\log(\text{LifeSpan})$ (0.05317), $\log(\text{Gestation})$ (-0.41163), and Danger (-0.28973)

2. What is the equation for the regression line?

$$\log(y) = 3.10346 + \log(\text{BodyWt})(0.14340) + \log(\text{BrainWt})(-0.11064) + \log(\text{LifeSpan})(0.05317) + \log(\text{Gestation})(-0.41163) +$$

3. Which predicting variable(s) are significant at $\alpha = 0.05$? What are their p-values?

$\log(\text{BodyWt})$, $\log(\text{Gestation})$ and Danger all are significant at the .05 level with p-values of 0.040252, 0.000148, and 4.51e-07 respectively.

4. Interpret the estimated value of the parameters, including the error term, corresponding to BodyWt and Predation in the context of the problem

The $\log(\text{BodyWt})$ parameter is 0.14340. As BodyWt increases by 1% nondreaming will decrease by 0.14340%, holding all other values constant. The standard error represents the average distance that the observed values fall from the regression line. The standard error for $\log(\text{BodyWt})$ is 0.06732. So our prediction, on average will be off by 0.06732.

The Danger parameter is -0.28973. As Danger increases by 1 unit non-dreaming decreases by $-0.28973 \times 100\%$, holding all other values constant. The standard error represents the average distance that the observed values

fall from the regression line. The standard error for Danger is 0.04690 . So our prediction, on average will be off by 0.04690 .

5. Did finalmodel improve over model4c? Explain how you determined if the model improved or not.

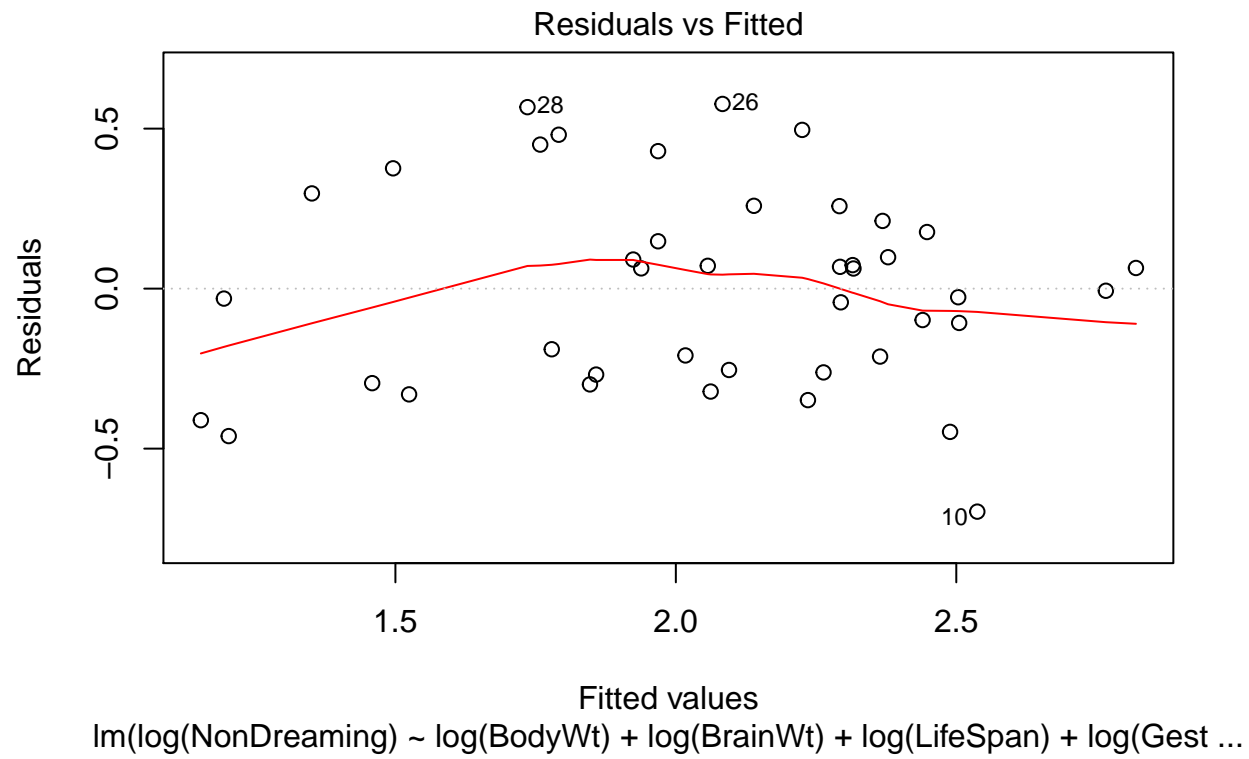
The final model did not improve upon model 4c because the R^2 value did not increase.

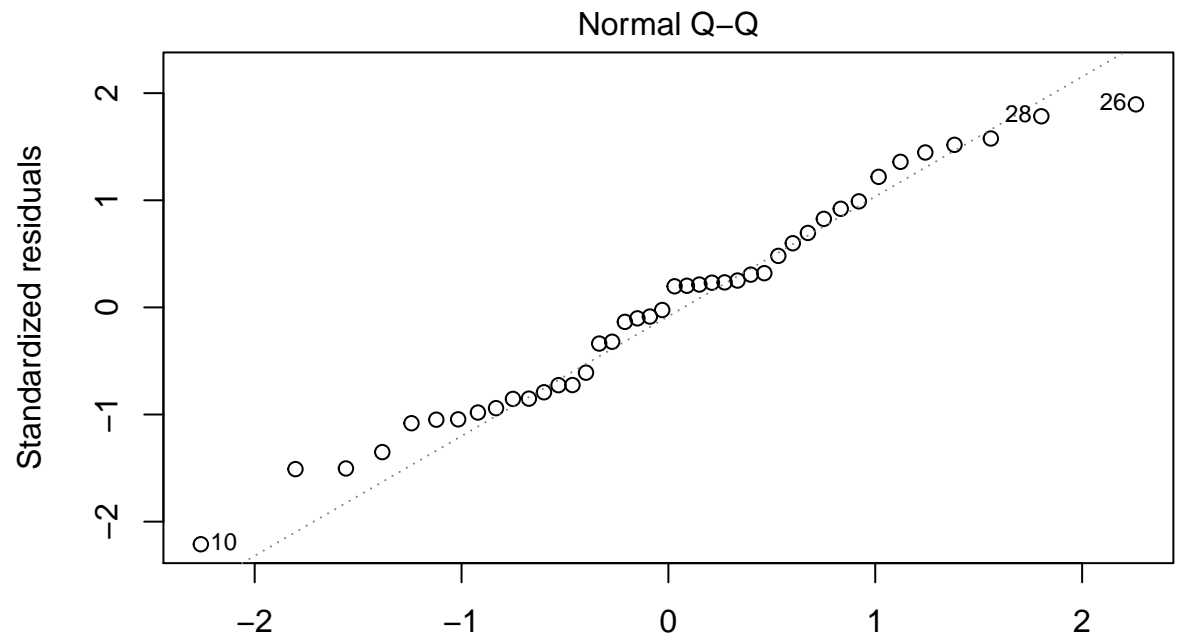
5 Question 5 - Checking the Assumptions of the Model

```
finalmodel <- lm(log(NonDreaming) ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) + log(Gestation) + Danger, data = data)

summary(finalmodel)

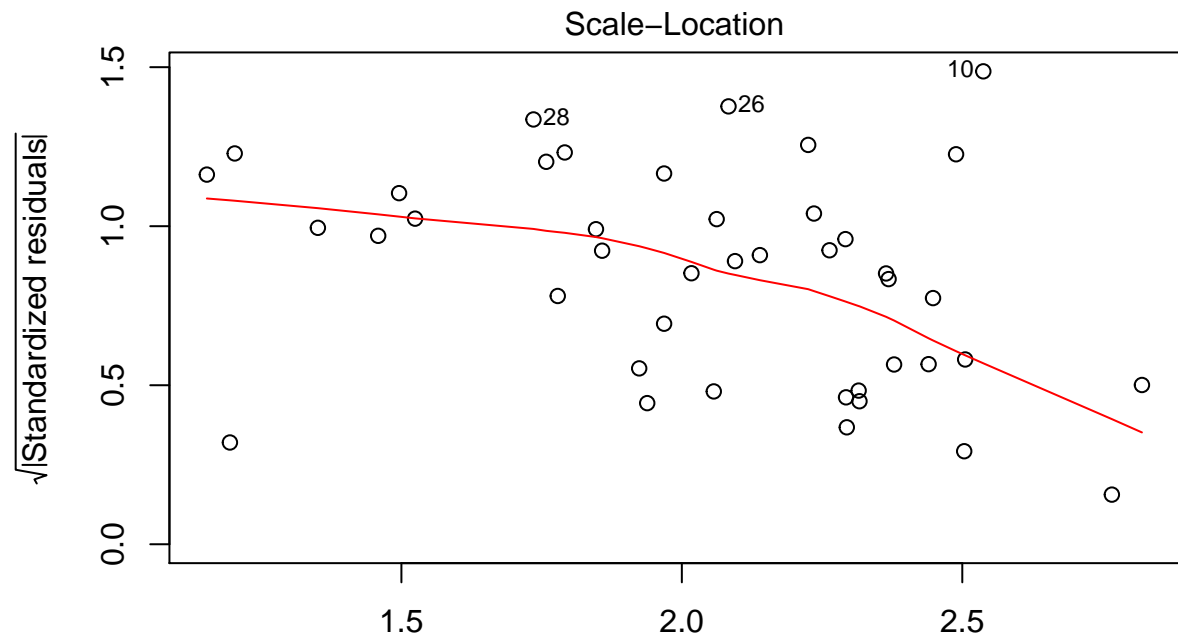
##
## Call:
## lm(formula = log(NonDreaming) ~ log(BodyWt) + log(BrainWt) +
##     log(LifeSpan) + log(Gestation) + Danger, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.69689 -0.25990  0.02811  0.20292  0.57709
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.78927    0.32012   8.713 2.15e-10 ***
## log(BodyWt)   -0.11203    0.05881  -1.905  0.06478 .
## log(BrainWt)   0.03583    0.08838   0.405  0.68756
## log(LifeSpan)  0.07153    0.08660   0.826  0.41425
## log(Gestation) -0.13977    0.08023  -1.742  0.09003 .
## Danger        -0.11576    0.04097  -2.825  0.00766 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3316 on 36 degrees of freedom
## Multiple R-squared:  0.643, Adjusted R-squared:  0.5934
## F-statistic: 12.97 on 5 and 36 DF, p-value: 3.036e-07
plot(finalmodel, cook.levels = c(4/42,0.5,1))
```



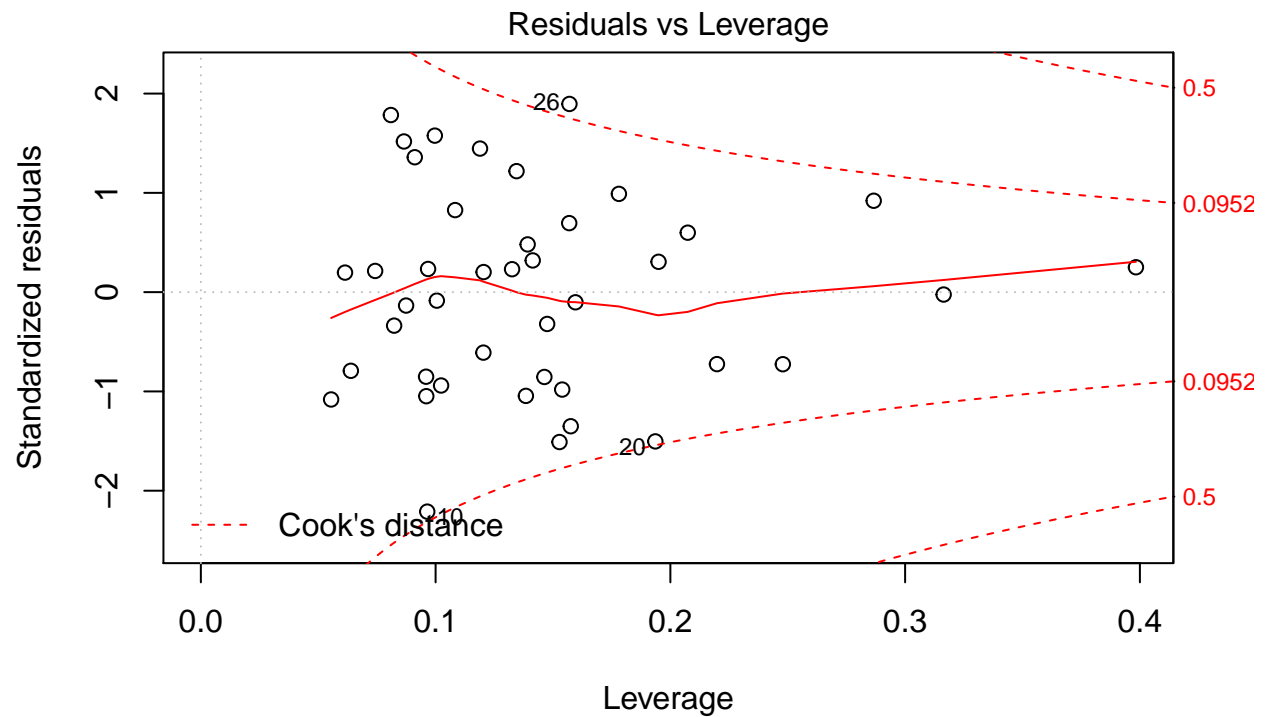


Theoretical Quantiles

$\ln(\log(\text{NonDreaming})) \sim \log(\text{BodyWt}) + \log(\text{BrainWt}) + \log(\text{LifeSpan}) + \log(\text{Gest ...})$

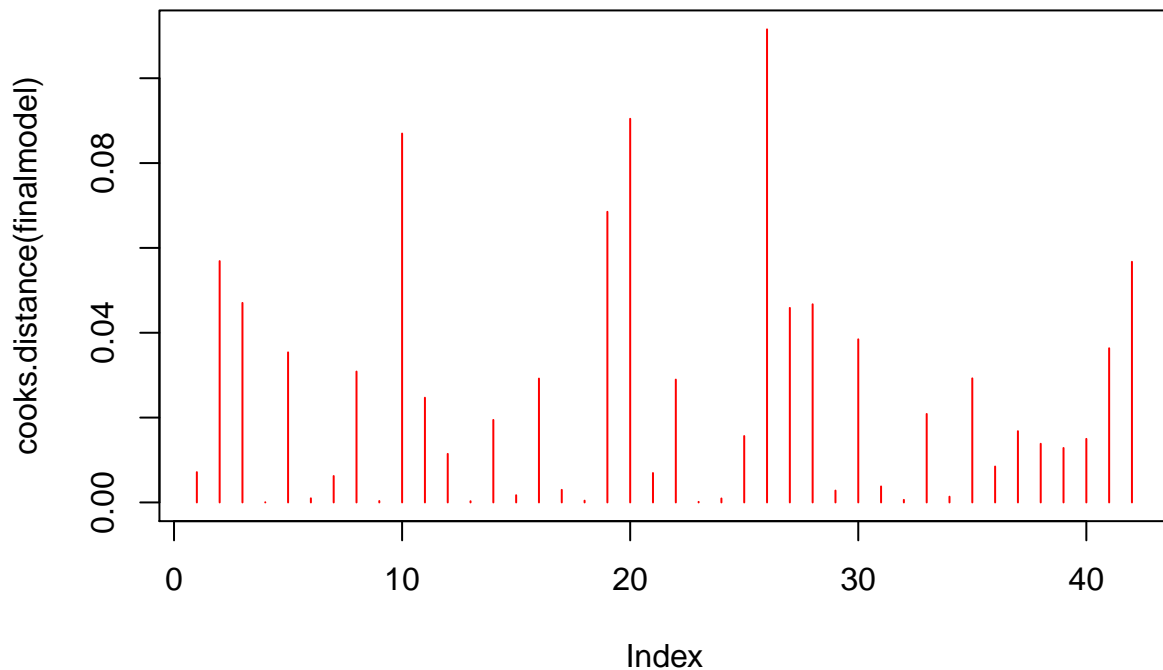


Fitted values
 $\text{lm}(\log(\text{NonDreaming}) \sim \log(\text{BodyWt}) + \log(\text{BrainWt}) + \log(\text{LifeSpan}) + \log(\text{Gest ...})$



$\text{lm}(\log(\text{NonDreaming}) \sim \log(\text{BodyWt}) + \log(\text{BrainWt}) + \log(\text{LifeSpan}) + \log(\text{Gest ...})$

```
plot(cooks.distance(finalmodel), type='h', col='red')
```



Constant variance and uncorrelated errors

The residual vs fitted values above shows the errors scattered around the zero line, which indicates that the constant variance and uncorrelated errors assumption holds.

Normality

The q-q plot of the standardized residuals shown above shows that the normality assumption holds with some deviations in the tails.

Outliers

The last plot, the cook's distance notes multiple outliers.

6 Question 6 - Checking the Assumptions of the Model

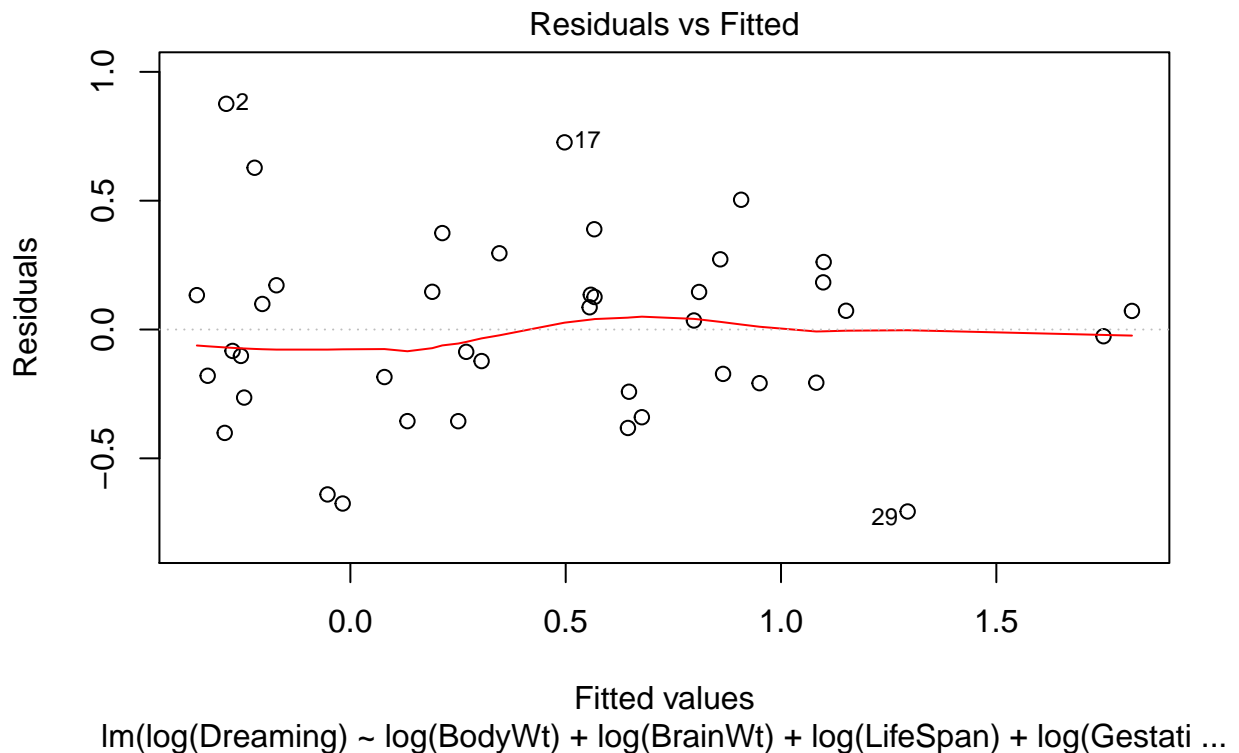
```
finalmodel <- lm(log(Dreaming) ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) + log(Gestation) + Danger, c
```

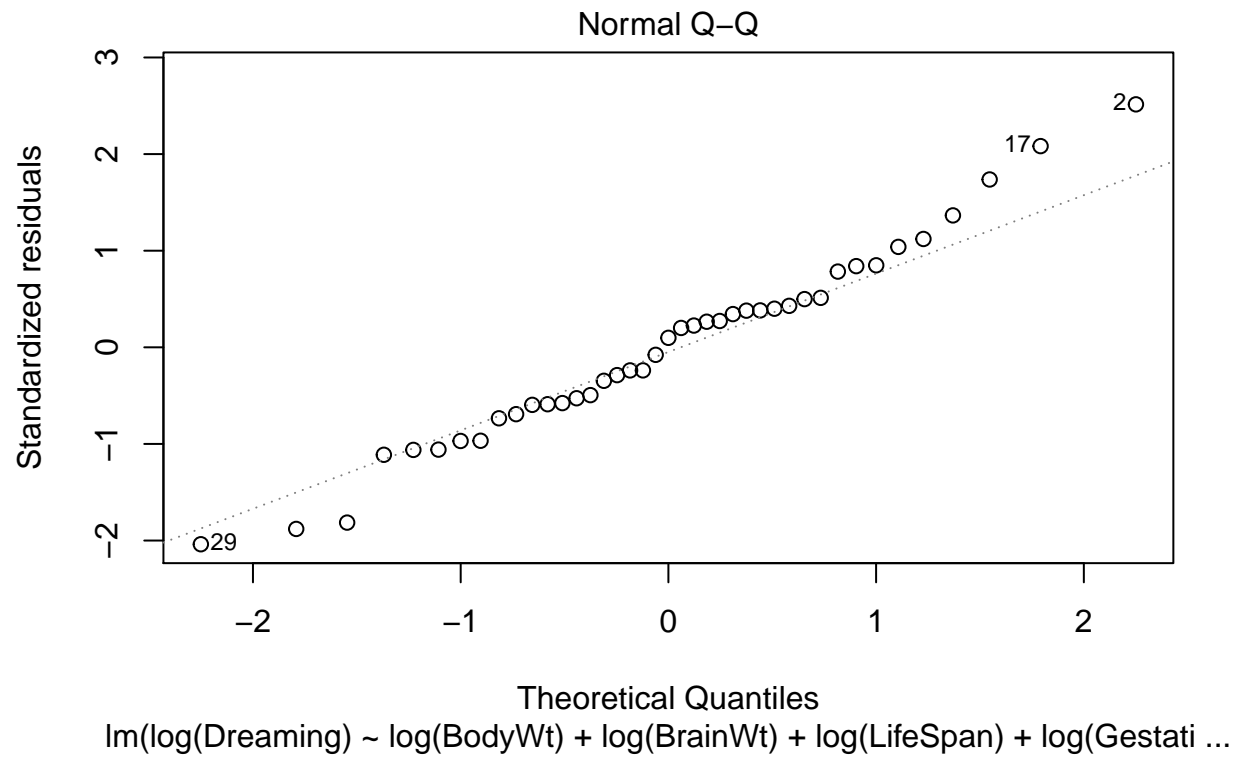
```
summary(finalmodel)
```

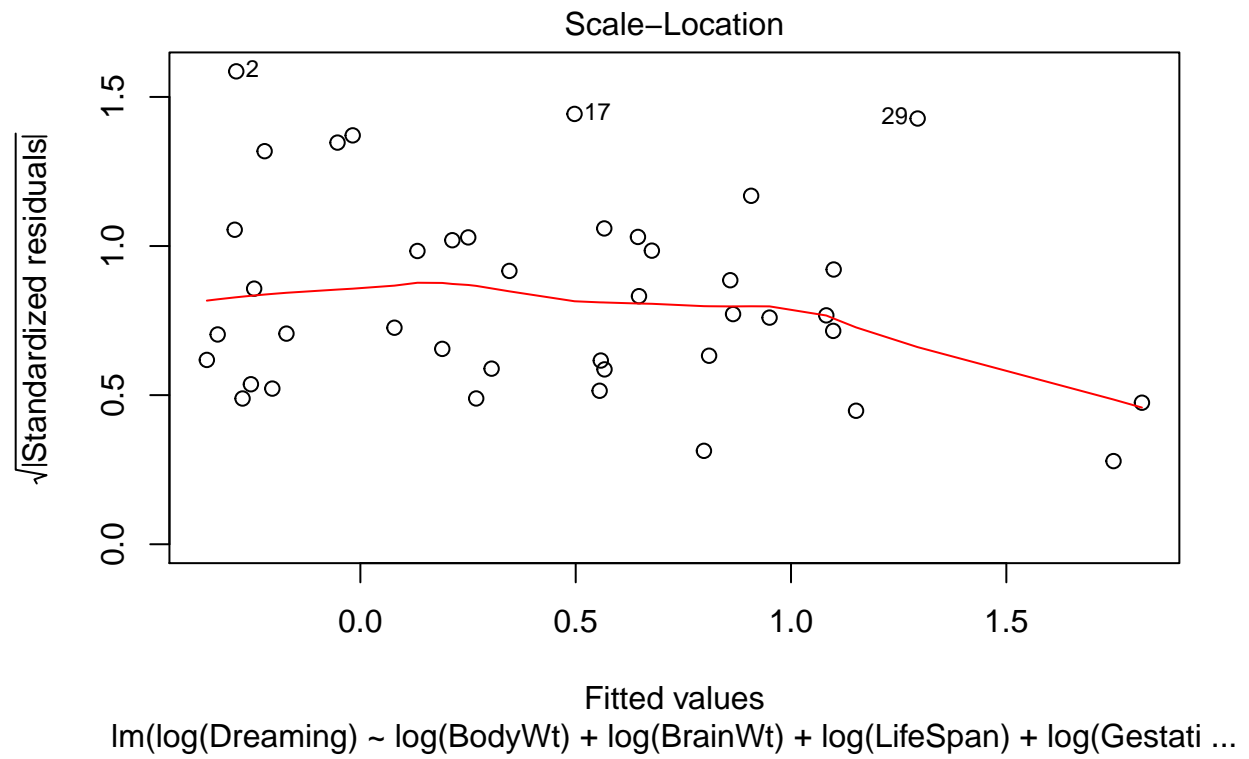
```
##
```

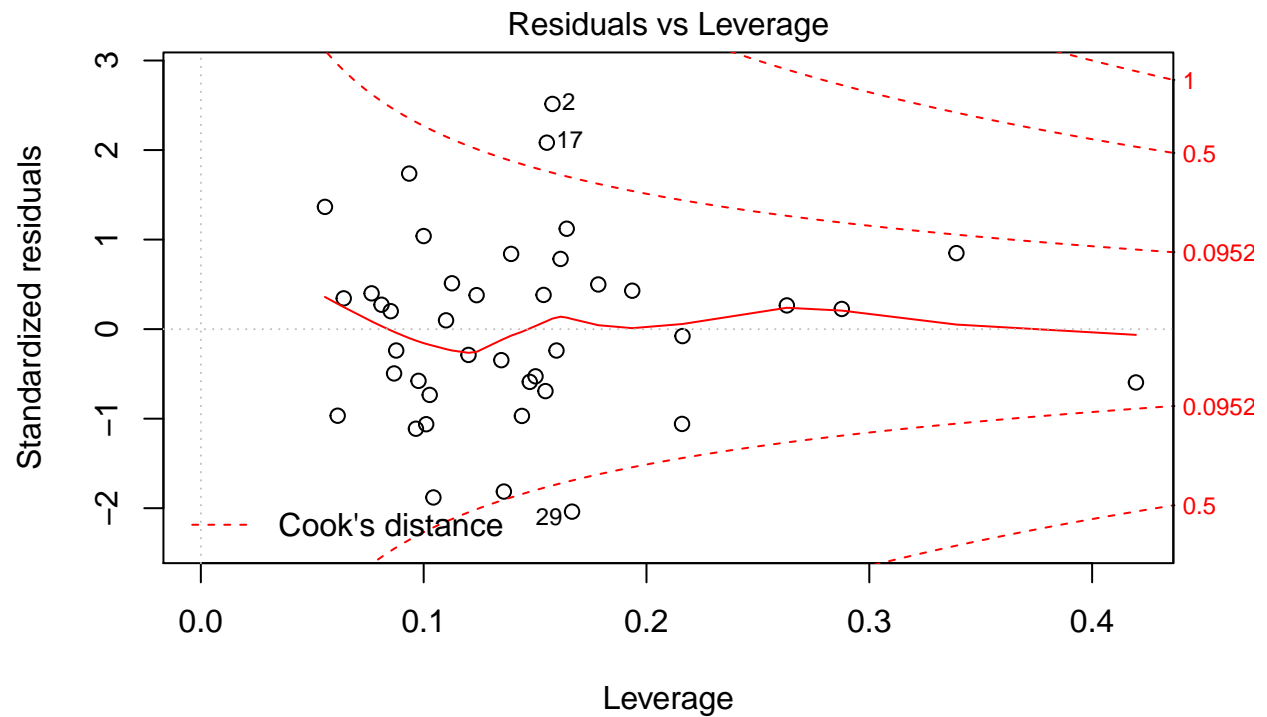
```
## Call:
```

```
## lm(formula = log(Dreaming) ~ log(BodyWt) + log(BrainWt) + log(LifeSpan) +
##     log(Gestation) + Danger, data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70627 -0.20824  0.03513  0.17167  0.87600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.10346    0.37019   8.383 6.90e-10 ***
## log(BodyWt)     0.14340    0.06732   2.130 0.040252 *
## log(BrainWt)   -0.11064    0.10117  -1.094 0.281612
## log(LifeSpan)   0.05317    0.10593   0.502 0.618862
## log(Gestation) -0.41163    0.09671  -4.256 0.000148 ***
## Danger         -0.28973    0.04690  -6.177 4.51e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3796 on 35 degrees of freedom
## Multiple R-squared:  0.7249, Adjusted R-squared:  0.6856
## F-statistic: 18.45 on 5 and 35 DF,  p-value: 6.022e-09
plot(finalmodel, cook.levels = c(4/42,0.5,1))
```

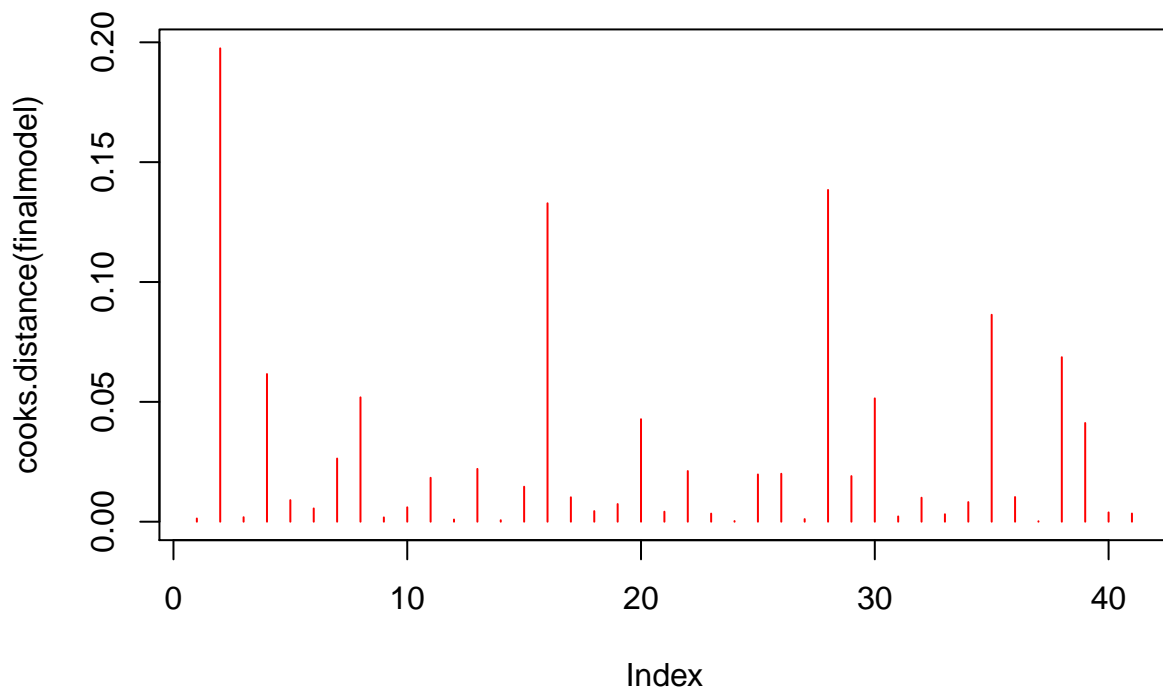








```
plot(cooks.distance(finalmodel), type='h', col='red')
```



Constant variance and uncorrelated errors

The residual vs fitted values above shows the errors scattered around the zero line, which indicates that the constant variance and uncorrelated errors assumption holds.

Normality

The q-q plot of the standardized residuals shown above shows that the normality assumption holds with some deviations in the tails.

Outliers

The last plot, the cook's distance notes multiple outliers.