Analyzing Ingestion Rates of Aquatic Organisms: A Multivariate Exploration of Weight, Organic Content, and Bivalve Status

Code ▾

Hide

```
library(ggplot2)
df <- read.csv(file.choose(), stringsAsFactors = T)
```

The DepositFeeders dataset provides information on the ingestion rates of snails and worms based on their weight. The variables are the dry weight of the organism in milligrams, the percentage of organic material at the bottom if ponds, lakes, and oceans, and the ingestion rate. This data set aims to explore the relationships between the ingestion rates of these organisms and various factors such as their weight and the organic content of the sediment. By analyzing this data set, we seek to understand how these factors influence the ingestion rates of the snails and worms.

Hide

```
summary(df)
```

```
                 Species        Weight          Ingestion          Organic
 Abarenicola claparedi: 1   Min.   :   0.20   Min.   :   0.430   Min.   : 0.40
 Abarenicola pacifica : 1   1st Qu.:   0.68   1st Qu.:   1.175   1st Qu.: 1.65
 Arenicola grubii     : 1   Median :  20.40   Median :  20.200   Median :13.00
 Arenicola marina     : 1   Mean   : 215.30   Mean   :1300.982   Mean   :21.94
 Hyalella azteca      : 1   3rd Qu.:  72.50   3rd Qu.: 983.500   3rd Qu.:26.65
 Hydrobia neglecta    : 1   Max.   :2050.00   Max.   :9400.000   Max.   :93.00
 (Other)              :13
```

The summary statistics show that there are 19 total species of snails and worms. We can see that the minimum weight between these 19 different species is 0.20mg, the mean is 215.30mg, and the max is 2050.0mg. Additionally, the minimum ingestion rate is 0.43mg the mean is 1300.982mg, and the max is 9400.0mg. Lastly, the mimimum amount of organic matter in the bottom of the ocean is 0.40, the mean is 21.94, and the max is 93.0.
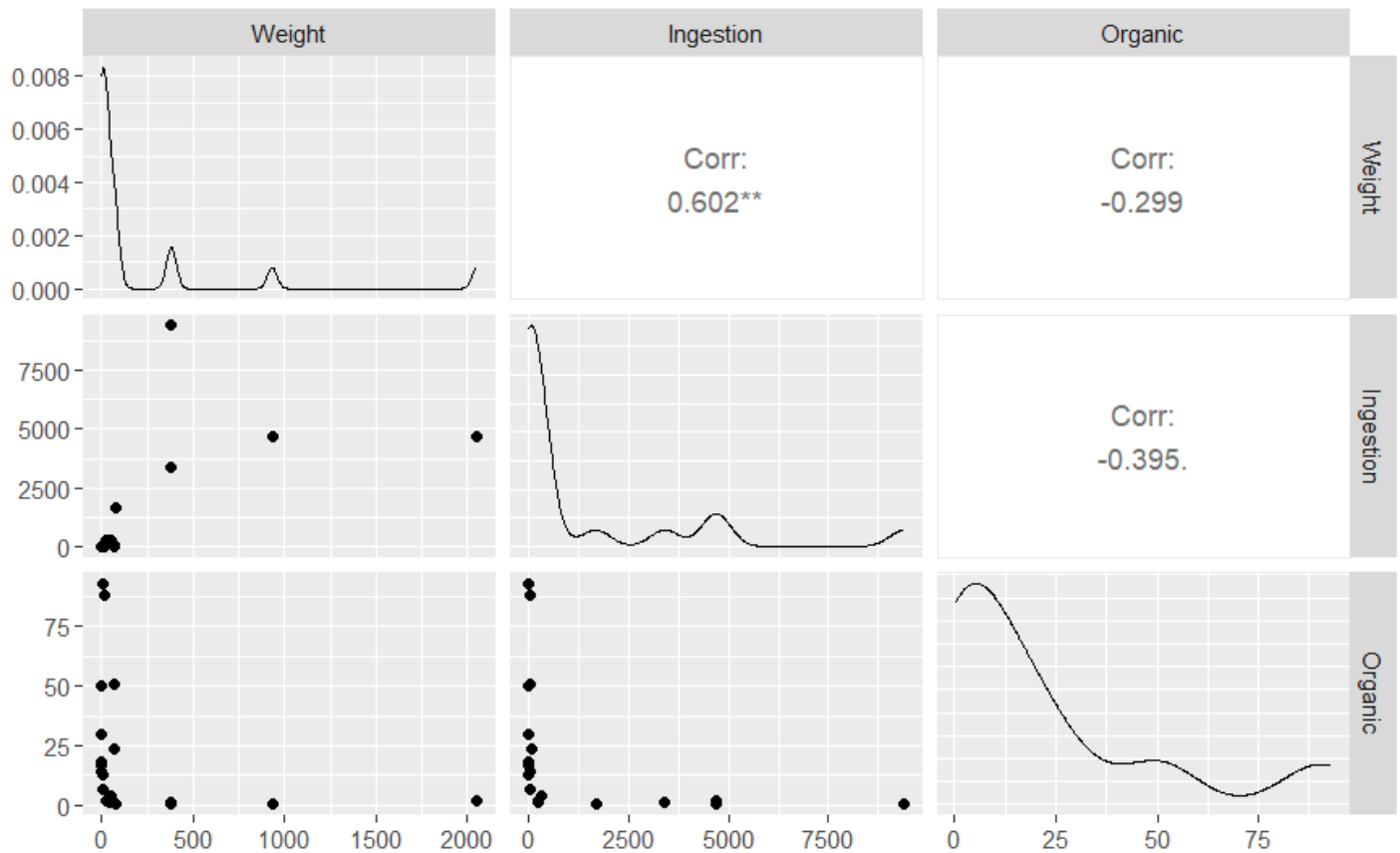
Hide

```
library(ggplot2)
library(GGally)
```

```
Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2
```
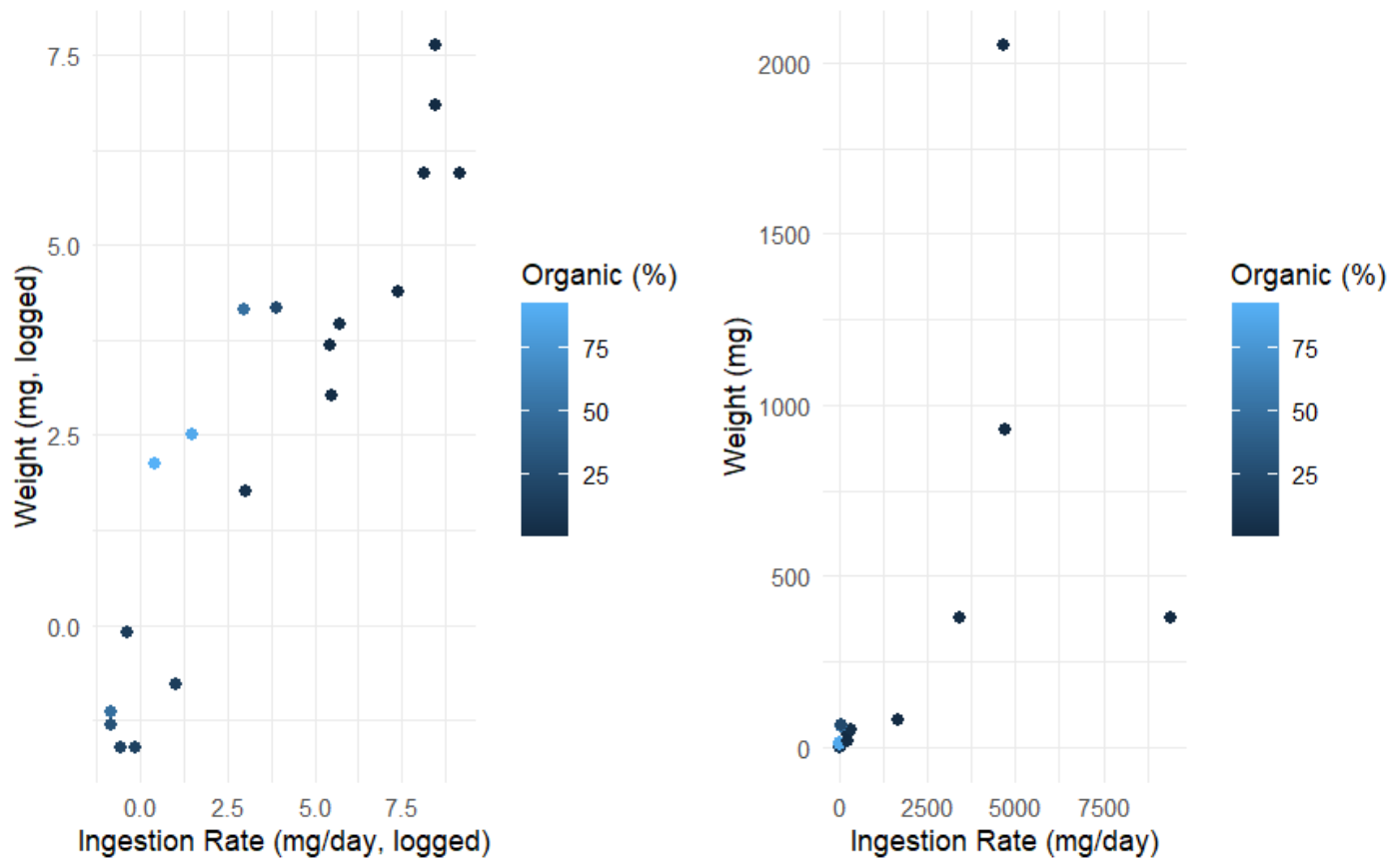
Hide

```
ggpairs(df[,-1])
```


Hide

```
library(gridExtra)
# Scatter plot with logged variables
logged_plot <- ggplot(df, aes(x = log(Ingestion), y = log(Weight), color = Organic)) +
  geom_point(size = 2) +
  labs(x = "Ingestion Rate (mg/day, logged)",
       y = "Weight (mg, logged)",
       color = "Organic (%)") +
  theme_minimal()

# Scatter plot with unlogged variables
unlogged_plot <- ggplot(df, aes(x = Ingestion, y = Weight, color = Organic)) +
  geom_point(size = 2) +
  labs(x = "Ingestion Rate (mg/day)",
       y = "Weight (mg)",
       color = "Organic (%)") +
  theme_minimal()

# Display the scatter plots side by side
grid.arrange(logged_plot, unlogged_plot, ncol = 2)
```

The relationship between the Weight and Ingestion is clearly more visible when they are both logged, therefore, I will use the logged Weight variable when creating the linear model.

Hide

```r
library(ggplot2)
library(gridExtra)

# Scatter plot with logged variables
logged_plot <- ggplot(df, aes(x = log(Organic), y = log(Ingestion), color = Weight)) +
  geom_point(size = 2) +
  labs(x = "Organic (%) (Logged)",
       y = "Ingestion Rate (mg/day, logged)") +
  theme_minimal()

# Scatter plot with unlogged variables
unlogged_plot <- ggplot(df, aes(x = Organic, y = Ingestion, color = Weight)) +
  geom_point(size = 2) +
  labs(x = "Organic (%)",
       y = "Ingestion Rate (mg/day)") +
  theme_minimal()

# Display the scatter plots side by side
grid.arrange(logged_plot, unlogged_plot, ncol = 2)
```
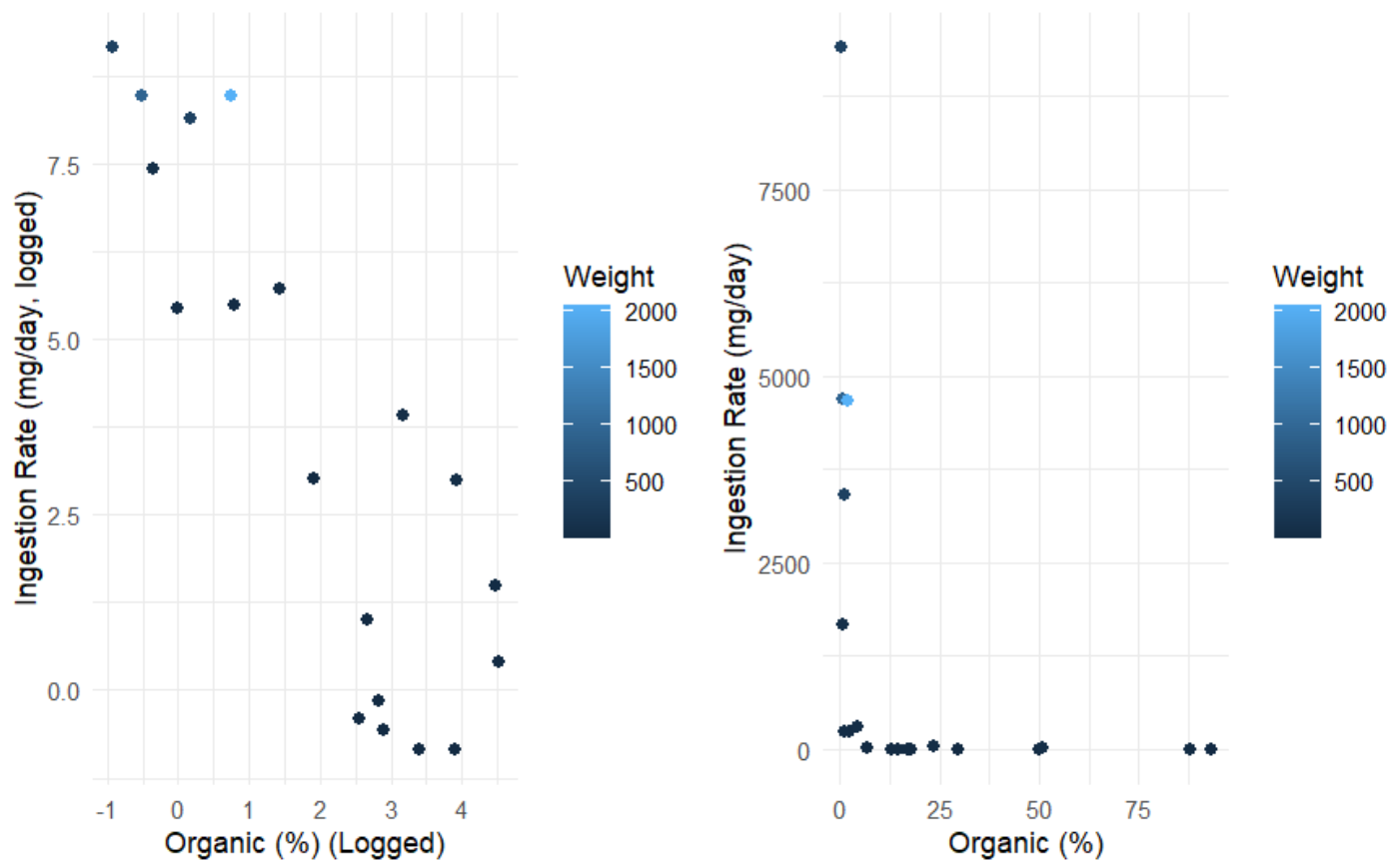
The relationship between the Organic and Ingestion is clearly more visible when they are both logged, therefore, I will use the logged Organic variable when creating the linear model.

a. After controlling for the weight of the animal, is the ingestion rate associated with the percentage of organic material? If so, interpret the relationship between organic and ingestion after controlling for weight.

Hide

```
iwo.lm <- lm( log(Ingestion) ~ log(Weight) + log(Organic), data = df)
summary(iwo.lm)
```

```
Call:
lm(formula = log(Ingestion) ~ log(Weight) + log(Organic), data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-1.3627 -0.1313  0.1975  0.3218  0.6421

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.39507    0.32447  10.463 1.46e-08 ***
log(Weight)   0.77083    0.05553  13.882 2.43e-10 ***
log(Organic) -0.91681    0.09377  -9.777 3.75e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5448 on 16 degrees of freedom
Multiple R-squared:  0.9797,     Adjusted R-squared:  0.9771
F-statistic: 385.5 on 2 and 16 DF,  p-value: 2.917e-14
```

This linear model shows strong statistical significance for both log(Weight) and log(Organic) as predictors of log(Ingestion), based on the p-value of 2.917e-14. Log(Weight) has an estimated coefficient of 0.77083 and a p-value of 2.43e-10. This means that there is strong evidence of a relationship between log(Weight) and log(Ingestion). Additionally log(Organic) has an estimated coefficient of -0.91681 and a p-value of 3.75e-08. This indicates that there is strong evidence of a relationship between log(Organic) and log(Ingestion). The overall model has a high multiple R-squared value of 0.9797, which means that it accounts for approximately 97.97% of the variability in the log(Ingestion) data. The adjusted R-squared is 0.9771, showing a good fit of the model to the data. In conclusion, the linear model using log-transformed variables shows strong statistical significance for both log(Weight) and log(Organic) in predicting log(Ingestion).

Hide

```
par(mfrow = c(2,2))
plot(iwo.lm, c(1,2,4,5))
```
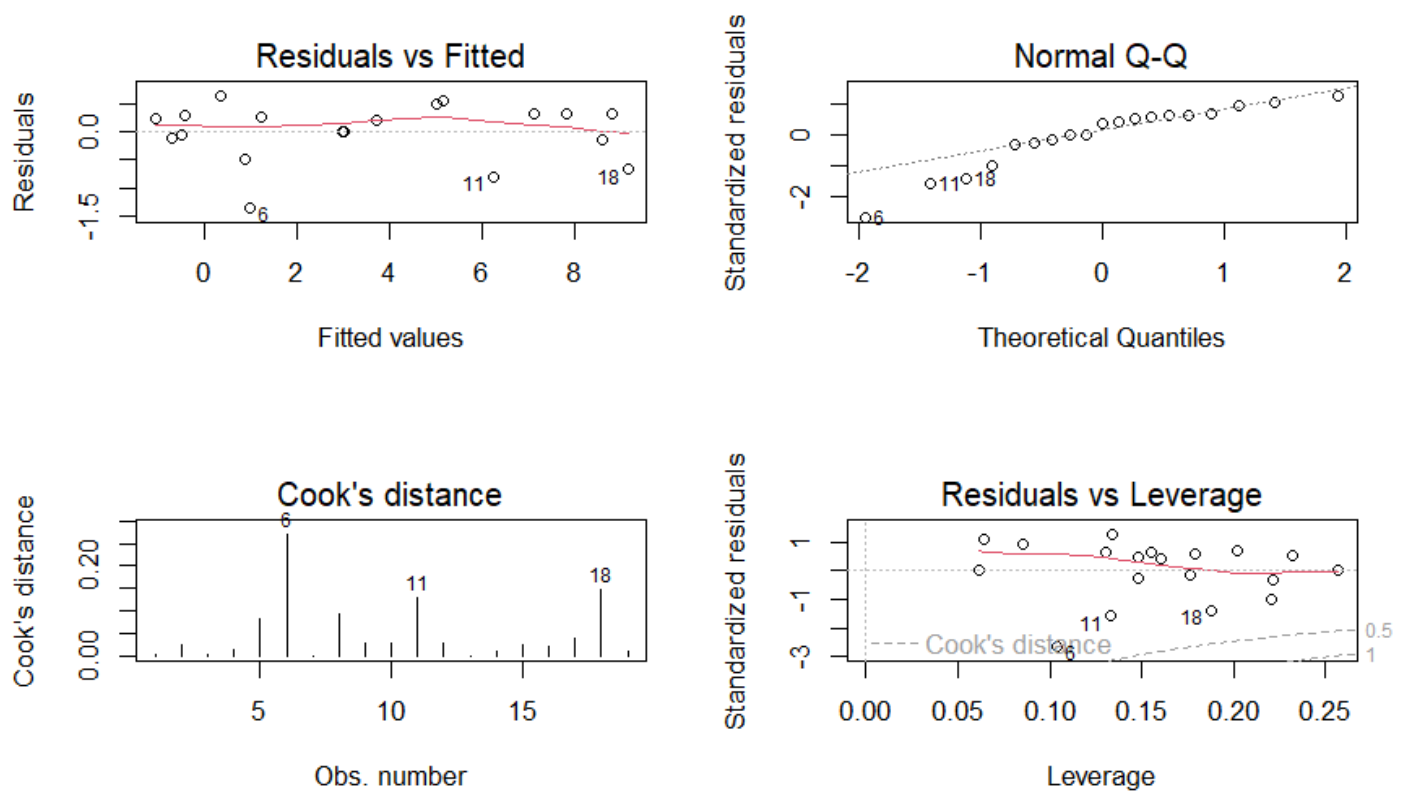
b. The same researcher was interested in whether bivalves should be included in the same analysis. Use DepositFeeders2.csv, which contains 3 new bivalves, to explore that question.

Hide

```
df2 <- read.csv(file.choose(), stringsAsFactors = T)
```

The DepositFeeders2 data set is the same as the original, as it aims to explore the relationships between the ingestion rates of these organisms. It includes the factors of the snail's and worm's weight, the organic content of the sediment, and adds whether the organism is a bivalve or not.
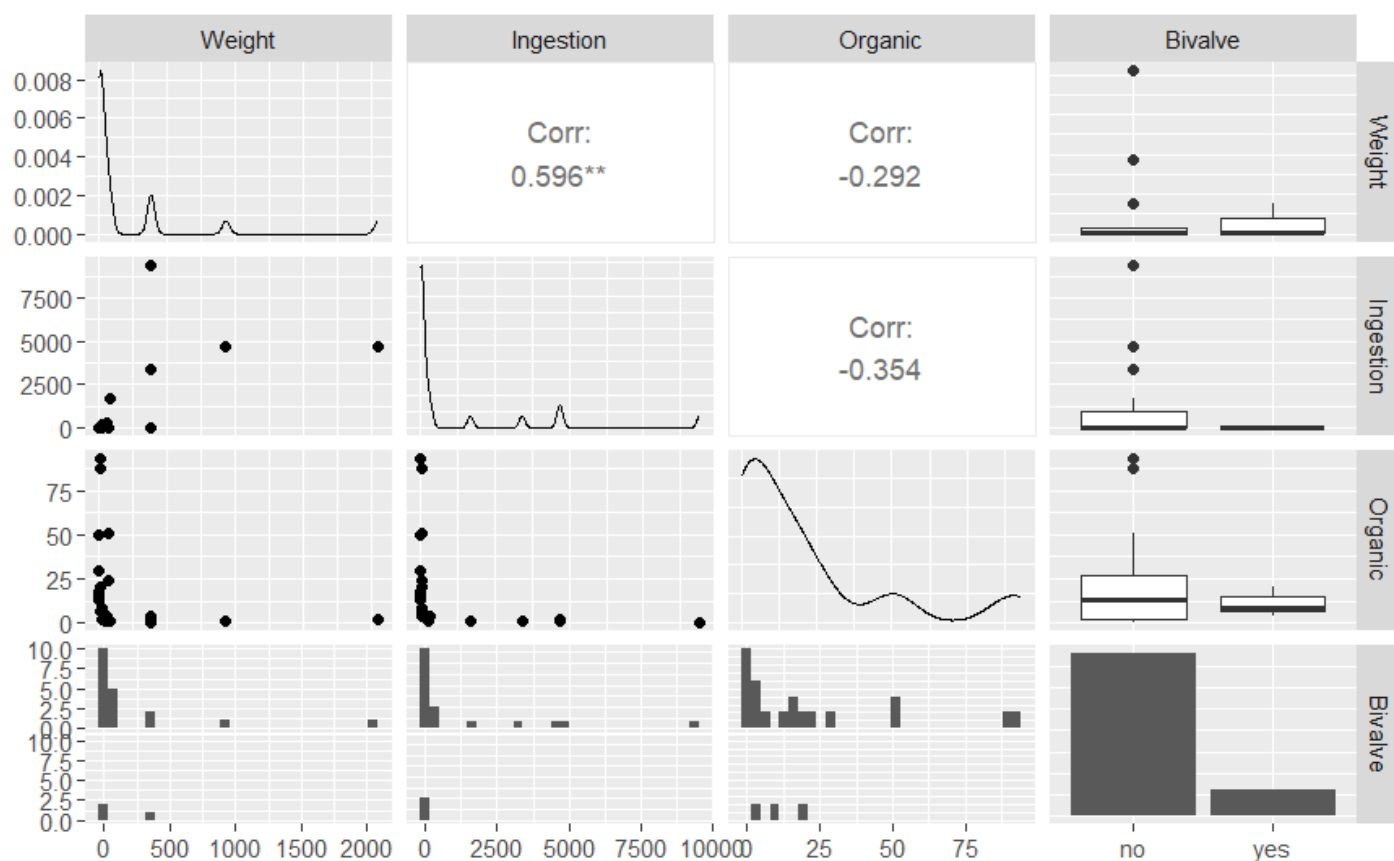
Hide

```
summary(df2)
```

```
                 Species      Weight          Ingestion         Organic          Bivalve
 Abarenicola claparedi: 1   Min.   :   0.20   Min.   :   0.430   Min.   : 0.400   no :19
 Abarenicola pacifica : 1   1st Qu.:   1.95   1st Qu.:   1.792   1st Qu.: 2.125   yes: 3
 Arenicola grubii     : 1   Median :  20.15   Median :  20.050   Median :10.900
 Arenicola marina     : 1   Mean   : 204.35   Mean   :1125.895   Mean   :20.415
 Hyalella azteca      : 1   3rd Qu.:  76.25   3rd Qu.: 285.000   3rd Qu.:22.700
 Hydrobia neglecta    : 1   Max.   :2050.00   Max.   :9400.000   Max.   :93.000
 (Other)              :16
```
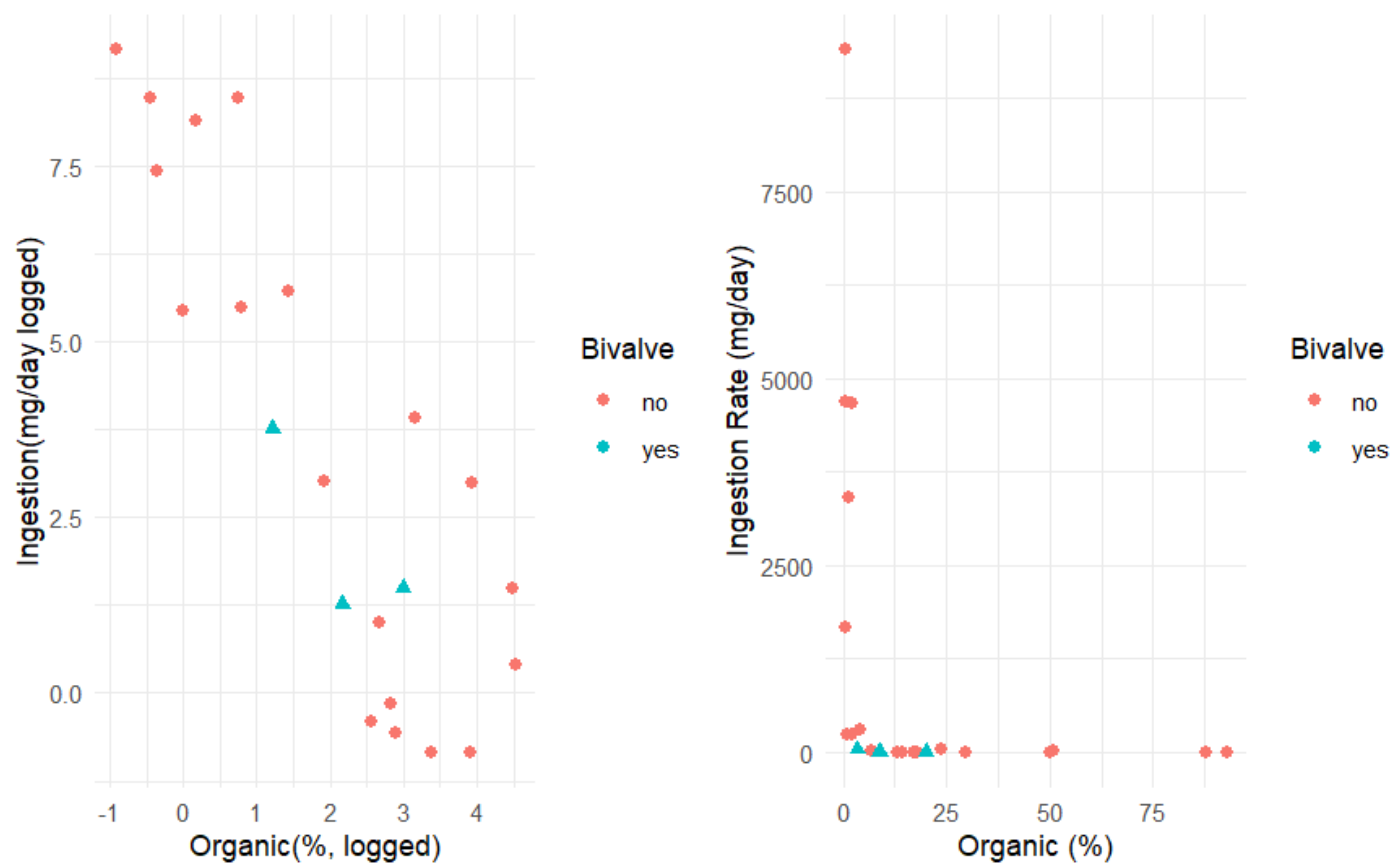
Hide

```
ggpairs(df2[,-1])
```

```
logged_plot <- ggplot(df2, aes(x = log(Organic), y = log(Ingestion), color = Bivalve)) +
  geom_point(size = 2, pch = as.numeric(df2$Bivalve) + 15) +
  labs(x = "Organic(%, logged)",
       y = "Ingestion(mg/day logged)",
       color = "Bivalve") +
  theme_minimal()

# Scatter plot with unlogged variables
unlogged_plot <- ggplot(df2, aes(x = Organic, y = Ingestion, color = Bivalve)) +
  geom_point(size = 2, pch = as.numeric(df2$Bivalve) + 15) +
  labs(x = "Organic (%)",
       y = "Ingestion Rate (mg/day)") +
  theme_minimal()

# Display the scatter plots side by side
grid.arrange(logged_plot, unlogged_plot, ncol = 2)
```

As we did with the 'DepositFeeders' data set, Ingestion, Organic, and Weight were all logged in order to make the scatter plot more linear.

Hide

```
iwob.lm <- lm( log(Ingestion) ~ log(Organic) + log(Weight) + Bivalve, data = df2)
summary(iwob.lm)
```

```
Call:
lm(formula = log(Ingestion) ~ log(Organic) + log(Weight) + Bivalve,
    data = df2)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4452 -0.4197  0.0679  0.4147  1.5177

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.45776    0.40720   8.491 1.04e-07 ***
log(Organic)  -0.91025    0.11804  -7.711 4.12e-07 ***
log(Weight)    0.74300    0.06863  10.826 2.60e-09 ***
Bivalveyes    -1.95732    0.43219  -4.529  0.00026 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6852 on 18 degrees of freedom
Multiple R-squared:  0.9652,    Adjusted R-squared:  0.9594
F-statistic: 166.2 on 3 and 18 DF,  p-value: 2.619e-13
```

Based on the linear model summary, we can see that there is strong evidence for a statistically significant relationship between the predictors log(Weight), log(Organic), and Bivalve, and the response variable, log(Ingestion), at the $p < 0.01$ significance level (p-value of 2.619e-13). The log(Weight) variable has a p-value of 2.60e-09, which is much smaller than 0.01, indicating a statistically significant relationship between log(Weight) and log(Ingestion). The log(Organic) has a p-value of 4.12e-07, which is also smaller than 0.01, demonstrating a statistically significant relationship between log(Organic) and log(Ingestion). The Bivalve variable has a p-value for of 0.00026, which is smaller than 0.01. This suggests that the Bivalve status of the species has a statistically significant relationship with log(Ingestion). Overall, the model's adjusted R-squared value is 0.9594, indicating that about 95.94% of the variation in the log-transformed Ingestion rate can be explained by the log-transformed Weight, log-transformed Organic percentage, and Bivalve status. Although this R-squared value is slightly smaller than the previous model that did not account for Bivalve status, it's not by a significant amount, so the Bivalve variable should be included in the same analysis since the variable has a statistically significant relationship with the Ingestion rate of the various worms and snails and since the model still maintains a very high R-squared value.