3. Worksheet: Basic R

Student Name; Z620: Quantitative Biodiversity, Indiana University

22 January, 2025

OVERVIEW

This worksheet introduces some of the basic features of the R computing environment (http://www.r-project.org). It is designed to be used along side the **3. RStudio** handout in your binder. You will not be able to complete the exercises without the corresponding handout.

Directions:

1. In the Markdown version of this document in your cloned repo, change "Student Name" on line 3 (above) with your name.
2. Complete as much of the worksheet as possible during class.
3. Use the handout as a guide; it contains a more complete description of data sets along with examples of proper scripting needed to carry out the exercises.
4. Answer questions in the worksheet. Space for your answers is provided in this document and is indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">". You should notice that the answer is highlighted in green by RStudio (color may vary if you changed the editor theme).
5. Before you leave the classroom today, you must **push** this file to your GitHub repo, at whatever stage you are. This will enable you to pull your work onto your own computer.
6. When you have completed the worksheet, **Knit** the text and code into a single PDF file by pressing the `Knit` button in the RStudio scripting panel. This will save the PDF output in your '3.RStudio' folder.
7. After Knitting, please submit the worksheet by making a **push** to your GitHub repo and then create a **pull request** via GitHub. Your pull request should include this file (**3.RStudio_Worksheet.Rmd**) with all code blocks filled out and questions answered) and the PDF output of `Knitr` (**3.RStudio_Worksheet.pdf**).

The completed exercise is due on **Wednesday, January 22nd, 2025 before 12:00 PM (noon)**.

1) HOW WE WILL BE USING R AND OTHER TOOLS

You are working in an RMarkdown (.Rmd) file. This allows you to integrate text and R code into a single document. There are two major features to this document: 1) Markdown formatted text and 2) "chunks" of R code. Anything in an R code chunk will be interpreted by R when you *Knit* the document.

When you are done, you will *knit* your document together. However, if there are errors in the R code contained in your Markdown document, you will not be able to knit a PDF file. If this happens, you will need to review your code, locate the source of the error(s), and make the appropriate changes. Even if you are able to knit without issue, you should review the knitted document for correctness and completeness before you submit the Worksheet. Next to the `Knit` button in the RStudio scripting panel there is a spell checker button (`ABC`) button.

2) SETTING YOUR WORKING DIRECTORY

In the R code chunk below, please provide the code to: 1) clear your R environment, 2) print your current working directory, and 3) set your working directory to your '3.RStudio' folder.

```
## Clear the R environment
rm(list = ls())

### Print working directory
getwd()
```

```
## [1] "C:/Users/ADMIN/OneDrive/Documents/GitHub/QB2025_Nambiar/Week1-RStudio"
```

```
##set working directory
setwd("C:/Users/ADMIN/OneDrive/Documents/GitHub/QB2025_Nambiar/Week1-RStudio/data")  # Replace "path/to/
```

3) USING R AS A CALCULATOR

To follow up on the pre-class exercises, please calculate the following in the R code chunk below. Feel free to reference the **1. Introduction to version control and computing tools** handout.

1) the volume of a cube with length, l, = 5 (volume = l^3)
2) the area of a circle with radius, r, = 2 (area = pi * r^2).
3) the length of the opposite side of a right-triangle given that the angle, theta, = pi/4. (radians, a.k.a. 45°) and with hypotenuse length sqrt(2) (remember: sin(theta) = opposite/hypotenuse).
4) the log (base e) of your favorite number.

```
### Volume of  cube
l <- 5
volume_cube <- l^3
volume_cube
```

```
## [1] 125
```

```
# #Area of a circle
r <- 2
area_circle <- pi * r^2
area_circle
```

```
## [1] 12.56637
```

```
# Length of the opposite side of  right-triangle
theta <- pi / 4  # unit-readians
hypotenuse <- sqrt(2)
opposite_length <- sin(theta) * hypotenuse
opposite_length
```

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

```r
# Log (base e) of my favorite number (n)
n <- 9
log_n <- log(n)
log_n
```

```
## [1] 2.197225
```

4) WORKING WITH VECTORS

To follow up on the pre-class exercises, please perform the requested operations in the R-code chunks below.

Basic Features Of Vectors

In the R-code chunk below, do the following: 1) Create a vector `x` consisting of any five numbers. 2) Create a new vector `w` by multiplying `x` by 14 (i.e., "scalar"). 3) Add `x` and `w` and divide by 15.

```r
#random numbers within a limit
x<-sample(2:30, 5)
x
```

```
## [1]  5 13 14  8 18
```

```r
w<-14*x
w
```

```
## [1]  70 182 196 112 252
```

```r
final<-(x + w)/15
final
```

```
## [1]  5 13 14  8 18
```

Now, do the following: 1) Create another vector (`k`) that is the same length as `w`. 2) Multiply `k` by `x`. 3) Use the combine function to create one more vector, `d` that consists of any three elements from `w` and any four elements of `k`.

```r
k<-sample(47:99, 5)
k
```

```
## [1] 58 74 99 73 52
```

```r
mult<-k*x
mult
```

```
## [1]  290  962 1386  584  936
```

```r
d <- c(sample(w, 3), sample(k, 4))
d
```

```
## [1]  70 112 196  73  58  74  52
```

Summary Statistics of Vectors

In the R-code chunk below, calculate the **summary statistics** (i.e., maximum, minimum, sum, mean, median, variance, standard deviation, and standard error of the mean) for the vector (v) provided.

```r
v <- c(16.4, 16.0, 10.1, 16.8, 20.5, NA, 20.2, 13.1, 24.8, 20.2, 25.0, 20.5, 30.5, 31.4, 27.1)

summary(v)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   10.10   16.50   20.35   20.90   24.95   31.40       1
```

```r
#variance

var(x)
```

```
## [1] 26.3
```

```r
#standard deviation

sd(x)
```

```
## [1] 5.128353
```

```r
#create function
se <- function(x) {
  sd_x <- sd(x)#standard deviation
  #number of observations
  n <- length(x)
  # Standard error
  se <- sd_x / sqrt(n)
  return(se)
}

se(x)
```

```
## [1] 2.293469
```

5) WORKING WITH MATRICES

In the R-code chunk below, do the following: Using a mixture of Approach 1 and 2 from the **3. RStudio** handout, create a matrix with two columns and five rows. Both columns should consist of random numbers. Make the mean of the first column equal to 8 with a standard deviation of 2 and the mean of the second column equal to 25 with a standard deviation of 10.

```r
x<-rnorm(5, mean=8, sd=2)
y<-rnorm (5, mean=5, sd=10)
x
```

```
## [1]  7.710587  5.389593  6.550673  8.873335 10.168281
```

```r
y
```

```
## [1] 26.5379495 -0.4552569  1.1181477 21.5390662 26.7406287
```

```r
dataframe<-data.frame(x,y)
dataframe
```

```
##           x          y
## 1  7.710587 26.5379495
## 2  5.389593 -0.4552569
## 3  6.550673  1.1181477
## 4  8.873335 21.5390662
## 5 10.168281 26.7406287
```

Question 1: What does the `rnorm` function do? What do the arguments in this function specify? Remember to use `help()` or type `?rnorm`.

> Answer 1: rnorm is a function generate random numbers of a specific length which can be loaded as a vector of a specific length, mean and standard deviation. Most importantly this data will be nroamlly distributed

In the R code chunk below, do the following: 1) Load `matrix.txt` from the **3.RStudio** data folder as matrix `m`. 2) Transpose this matrix. 3) Determine the dimensions of the transposed matrix.

```r
m<-read.table("C:/Users/ADMIN/OneDrive/Documents/GitHub/QB2025_Nambiar/Week1-RStudio/data/matrix.txt", s
m
```

```
##   X8 X1 X7 X6 X1.1
## 1  5  5  2  4    1
## 2  2  5  4  3    3
## 3  3  2  5  1    4
## 4  9  9  1  1    2
## 5 11  8  1  8    8
## 6  2  2  5  8    5
## 7  3  3  6  7    6
## 8  5  5  1  3    6
## 9  6  5  9  2    2
```

```r
#transposed matrix

m_trans<-t(m)
m
```

```
##   X8 X1 X7 X6 X1.1
## 1  5  5  2  4    1
## 2  2  5  4  3    3
## 3  3  2  5  1    4
## 4  9  9  1  1    2
## 5 11  8  1  8    8
## 6  2  2  5  8    5
## 7  3  3  6  7    6
## 8  5  5  1  3    6
## 9  6  5  9  2    2
```

```
#dimensions

dim(m_trans)
```

```
## [1] 5 9
```

Question 2: What are the dimensions of the matrix you just transposed?

Answer 2:5 rows and 9 columns

###Indexing a Matrix

In the R code chunk below, do the following: 1) Index matrix `m` by selecting all but the third column. 2) Remove the last row of matrix `m`.

```
m1<-m[,-3] #remove column 3

m2<-m[-9,]
m2
```

```
##   X8 X1 X7 X6 X1.1
## 1  5  5  2  4    1
## 2  2  5  4  3    3
## 3  3  2  5  1    4
## 4  9  9  1  1    2
## 5 11  8  1  8    8
## 6  2  2  5  8    5
## 7  3  3  6  7    6
## 8  5  5  1  3    6
```

6) BASIC DATA VISUALIZATION AND STATISTICAL ANALYSIS

Load Zooplankton Data Set

In the R code chunk below, do the following: 1) Load the zooplankton data set from the **3.RStudio** data folder. 2) Display the structure of this data set.

```
zoops<-read.table("C:/Users/ADMIN/OneDrive/Documents/GitHub/QB2025_Nambiar/Week1-RStudio/data/zoop_nuts
zoops
```

```
##    TANK NUTS     TP     TN   SRP     TIN  CHLA     ZP
## 1    34    L  20.31  720.1  4.02  131.62  1.52 1.7808
## 2    14    L  25.55  750.5  1.56  141.10  4.00 0.4090
## 3    23    L  14.22  610.1  4.97  107.70  0.61 1.2014
## 4    16    L  39.11  760.9  2.89   71.28  0.53 3.3598
## 5    21    L  20.09  570.4  5.11   80.40  1.44 0.7332
## 6     5    L  15.75  680.5  4.68  135.77  1.19 0.9773
## 7    25    L  19.55  665.5  5.00   79.40  0.37 1.0999
## 8    27    L  16.19  660.8  0.10  100.91  0.72 2.2714
## 9    30    M  29.46 1770.4  7.90 1329.26  6.93 3.1633
## 10   28    M  37.88 2590.3  3.92 1163.64  0.94 1.8747
## 11   35    M  30.26 2110.9  4.45 1850.18  1.36 4.3802
## 12   36    M  36.94 2060.9  5.14  249.93 38.38 2.4051
## 13   12    M  34.73 1370.1  4.69  420.01 15.99 1.7079
## 14   22    M  26.00 2110.3  5.35 1466.70  0.95 4.0999
## 15   18    M  28.50 1760.4  7.15 1351.83  1.36 5.4430
## 16   15    M  35.33 1360.8  5.96 1036.27  2.13 4.2677
## 17   17    H  41.56 4130.1 20.34 3421.43  1.44 8.2084
## 18   10    H  53.50 4530.4 33.57 4042.10  0.93 4.2273
## 19   29    H  99.07 4410.9 11.57 3307.05  0.61 6.2381
## 20    6    H 128.04 4750.4 26.27 3686.17  1.27 8.5713
## 21   24    H  33.47 3410.4  9.32 2791.52  1.11 1.4240
## 22   19    H  52.41 3710.3  3.23 2890.73 17.59 2.9714
## 23    4    H  42.21 3690.4 12.71 3041.75  1.08 8.1509
## 24   11    H  77.65 4380.6 21.86 3041.75  1.08 8.3868
```

```
str(zoops) #structure
```

```
## 'data.frame':    24 obs. of  8 variables:
##  $ TANK: int  34 14 23 16 21 5 25 27 30 28 ...
##  $ NUTS: chr  "L" "L" "L" "L" ...
##  $ TP  : num  20.3 25.6 14.2 39.1 20.1 ...
##  $ TN  : num  720 750 610 761 570 ...
##  $ SRP : num  4.02 1.56 4.97 2.89 5.11 4.68 5 0.1 7.9 3.92 ...
##  $ TIN : num  131.6 141.1 107.7 71.3 80.4 ...
##  $ CHLA: num  1.52 4 0.61 0.53 1.44 1.19 0.37 0.72 6.93 0.94 ...
##  $ ZP  : num  1.781 0.409 1.201 3.36 0.733 ...
```

Correlation

In the R-code chunk below, do the following: 1) Create a matrix with the numerical data in the `zoop` dataframe. 2) Visualize the pairwise **bi-plots** of the numerical variables. 3) Conduct a simple **Pearson's correlation** analysis.

```
zoop1<-as.data.frame(zoops[,-(1:2)])
zoop1
```

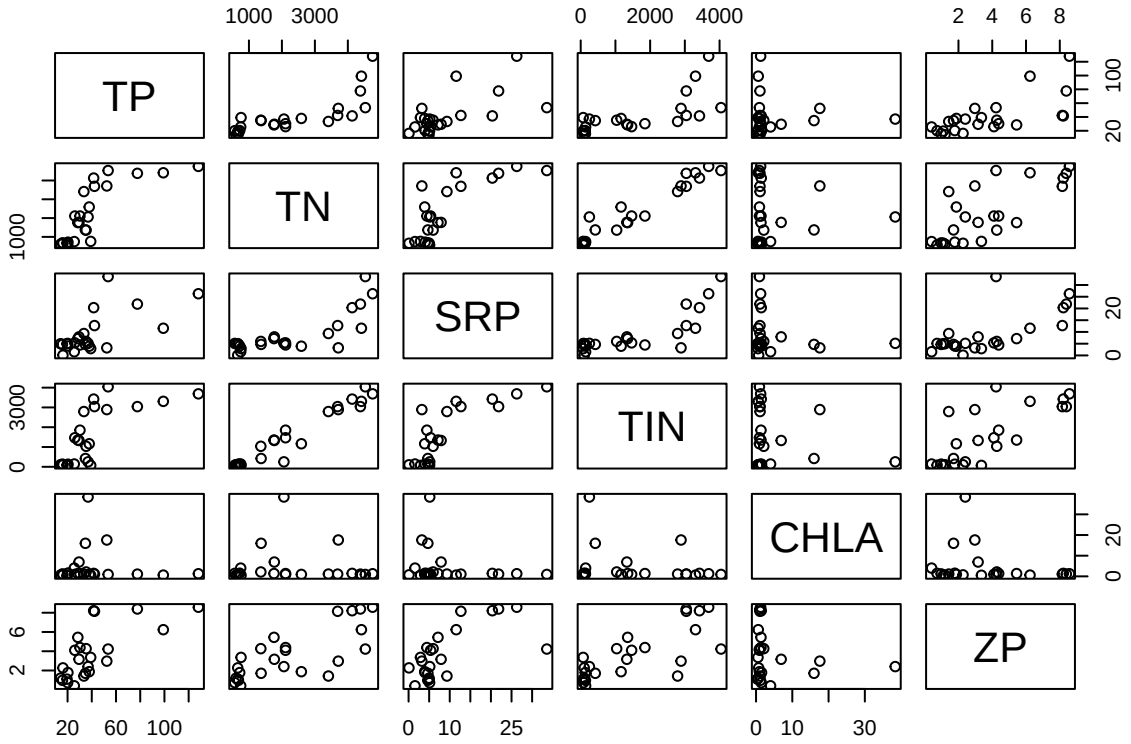
```
##        TP     TN   SRP     TIN  CHLA     ZP
## 1   20.31  720.1  4.02  131.62  1.52 1.7808
## 2   25.55  750.5  1.56  141.10  4.00 0.4090
## 3   14.22  610.1  4.97  107.70  0.61 1.2014
## 4   39.11  760.9  2.89   71.28  0.53 3.3598
```

```
## 5   20.09  570.4  5.11   80.40  1.44 0.7332
## 6   15.75  680.5  4.68  135.77  1.19 0.9773
## 7   19.55  665.5  5.00   79.40  0.37 1.0999
## 8   16.19  660.8  0.10  100.91  0.72 2.2714
## 9   29.46 1770.4  7.90 1329.26  6.93 3.1633
## 10  37.88 2590.3  3.92 1163.64  0.94 1.8747
## 11  30.26 2110.9  4.45 1850.18  1.36 4.3802
## 12  36.94 2060.9  5.14  249.93 38.38 2.4051
## 13  34.73 1370.1  4.69  420.01 15.99 1.7079
## 14  26.00 2110.3  5.35 1466.70  0.95 4.0999
## 15  28.50 1760.4  7.15 1351.83  1.36 5.4430
## 16  35.33 1360.8  5.96 1036.27  2.13 4.2677
## 17  41.56 4130.1 20.34 3421.43  1.44 8.2084
## 18  53.50 4530.4 33.57 4042.10  0.93 4.2273
## 19  99.07 4410.9 11.57 3307.05  0.61 6.2381
## 20 128.04 4750.4 26.27 3686.17  1.27 8.5713
## 21  33.47 3410.4  9.32 2791.52  1.11 1.4240
## 22  52.41 3710.3  3.23 2890.73 17.59 2.9714
## 23  42.21 3690.4 12.71 3041.75  1.08 8.1509
## 24  77.65 4380.6 21.86 3041.75  1.08 8.3868
```

```
zoops
```

```
##    TANK NUTS     TP     TN   SRP     TIN  CHLA     ZP
## 1    34    L  20.31  720.1  4.02  131.62  1.52 1.7808
## 2    14    L  25.55  750.5  1.56  141.10  4.00 0.4090
## 3    23    L  14.22  610.1  4.97  107.70  0.61 1.2014
## 4    16    L  39.11  760.9  2.89   71.28  0.53 3.3598
## 5    21    L  20.09  570.4  5.11   80.40  1.44 0.7332
## 6     5    L  15.75  680.5  4.68  135.77  1.19 0.9773
## 7    25    L  19.55  665.5  5.00   79.40  0.37 1.0999
## 8    27    L  16.19  660.8  0.10  100.91  0.72 2.2714
## 9    30    M  29.46 1770.4  7.90 1329.26  6.93 3.1633
## 10   28    M  37.88 2590.3  3.92 1163.64  0.94 1.8747
## 11   35    M  30.26 2110.9  4.45 1850.18  1.36 4.3802
## 12   36    M  36.94 2060.9  5.14  249.93 38.38 2.4051
## 13   12    M  34.73 1370.1  4.69  420.01 15.99 1.7079
## 14   22    M  26.00 2110.3  5.35 1466.70  0.95 4.0999
## 15   18    M  28.50 1760.4  7.15 1351.83  1.36 5.4430
## 16   15    M  35.33 1360.8  5.96 1036.27  2.13 4.2677
## 17   17    H  41.56 4130.1 20.34 3421.43  1.44 8.2084
## 18   10    H  53.50 4530.4 33.57 4042.10  0.93 4.2273
## 19   29    H  99.07 4410.9 11.57 3307.05  0.61 6.2381
## 20    6    H 128.04 4750.4 26.27 3686.17  1.27 8.5713
## 21   24    H  33.47 3410.4  9.32 2791.52  1.11 1.4240
## 22   19    H  52.41 3710.3  3.23 2890.73 17.59 2.9714
## 23    4    H  42.21 3690.4 12.71 3041.75  1.08 8.1509
## 24   11    H  77.65 4380.6 21.86 3041.75  1.08 8.3868
```

```
pairs(zoop1) #pairwise plots
```




```
cor_mat<- cor(zoop1, method = "pearson")
cor_mat
```

```
##               TP           TN        SRP        TIN         CHLA         ZP
## TP    1.00000000  0.786510407  0.6540957  0.7171143 -0.016659593  0.6974765
## TN    0.78651041  1.000000000  0.7841904  0.9689999 -0.004470263  0.7562474
## SRP   0.65409569  0.784190400  1.0000000  0.8009033 -0.189148017  0.6762947
## TIN   0.71711434  0.968999866  0.8009033  1.0000000 -0.156881463  0.7605629
## CHLA -0.01665959 -0.004470263 -0.1891480 -0.1568815  1.000000000 -0.1825999
## ZP    0.69747649  0.756247384  0.6762947  0.7605629 -0.182599904  1.0000000
```

Question 3: Describe some of the general features based on the visualization and correlation analysis above?

> Answer 3: I have calculated an extra column mean_bio which is the mean biomass of all the species ina tank (row). CHYD seems to have the strongest correlation with the mean_bio suggesting that most the mean value is determinedby CHYD. Overall, we can see the pairwise correlations of the different species of zooplanktons.

In the R code chunk below, do the following: 1) Redo the correlation analysis using the `corr.test()` function in the `psych` package with the following options: method = "pearson", adjust = "BH". 2) Now, redo this correlation analysis using a non-parametric method. 3) Use the print command from the handout to see the results of each correlation analysis.

```r
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.4.2
```

```r
zoop1_para <- corr.test(zoop1, method = "pearson", adjust = "BH")
zoop1_para
```

```
## Call:corr.test(x = zoop1, method = "pearson", adjust = "BH")
## Correlation matrix
##         TP   TN   SRP   TIN  CHLA    ZP
## TP    1.00 0.79  0.65  0.72 -0.02  0.70
## TN    0.79 1.00  0.78  0.97  0.00  0.76
## SRP   0.65 0.78  1.00  0.80 -0.19  0.68
## TIN   0.72 0.97  0.80  1.00 -0.16  0.76
## CHLA -0.02 0.00 -0.19 -0.16  1.00 -0.18
## ZP    0.70 0.76  0.68  0.76 -0.18  1.00
## Sample Size
## [1] 24
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##        TP   TN  SRP  TIN CHLA   ZP
## TP   0.00 0.00 0.00 0.00 0.98 0.00
## TN   0.00 0.00 0.00 0.00 0.98 0.00
## SRP  0.00 0.00 0.00 0.00 0.49 0.00
## TIN  0.00 0.00 0.00 0.00 0.54 0.00
## CHLA 0.94 0.98 0.38 0.46 0.00 0.49
## ZP   0.00 0.00 0.00 0.00 0.39 0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

```r
zoop1_nonpara<-corr.test(zoop1, method = "spearman", adjust = "BH") #using spearman for the non para te
zoop1_nonpara
```

```
## Call:corr.test(x = zoop1, method = "spearman", adjust = "BH")
## Correlation matrix
##        TP   TN   SRP  TIN  CHLA    ZP
## TP   1.00 0.89  0.54 0.76  0.04  0.74
## TN   0.89 1.00  0.65 0.94  0.02  0.75
## SRP  0.54 0.65  1.00 0.73 -0.06  0.63
## TIN  0.76 0.94  0.73 1.00  0.09  0.74
## CHLA 0.04 0.02 -0.06 0.09  1.00 -0.07
## ZP   0.74 0.75  0.63 0.74 -0.07  1.00
## Sample Size
## [1] 24
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##        TP   TN  SRP  TIN CHLA   ZP
## TP   0.00 0.00 0.01 0.00 0.91 0.00
## TN   0.00 0.00 0.00 0.00 0.92 0.00
## SRP  0.01 0.00 0.00 0.00 0.88 0.00
## TIN  0.00 0.00 0.00 0.00 0.88 0.00
## CHLA 0.85 0.92 0.77 0.68 0.00 0.88
## ZP   0.00 0.00 0.00 0.00 0.74 0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

Question 4: Describe what you learned from `corr.test`. Specifically, are the results sensitive to whether you use parametric (i.e., Pearson's) or non-parametric methods? When should one use non-parametric methods instead of parametric methods? With the Pearson's method, is there evidence for false discovery rate due to multiple comparisons? Why is false discovery rate important?

> Answer 4: With non-parametric tests we appear to see similar result in terms of the directions of the trends but the strength of correlation is stronger in the spearman test vs pearson test (parametric). Parametric tests are used when the variables arenormally distributed, numerical adn linear (along with homoskedasticity), if these requirements are not satisfied we used non parametrics statistics. We also use non parametric statistics when we need to analyse ordinal data. Im not very clear about the relationship between Pearosn corelations and false discovery rates.

Linear Regression

In the R code chunk below, do the following: 1) Conduct a linear regression analysis to test the relationship between total nitrogen (TN) and zooplankton biomass (ZP). 2) Examine the output of the regression analysis. 3) Produce a plot of this regression analysis including the following: categorically labeled points, the predicted regression line with 95% confidence intervals, and the appropriate axis labels.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.1
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x ggplot2::%+%()   masks psych::%+%()
## x ggplot2::alpha() masks psych::alpha()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

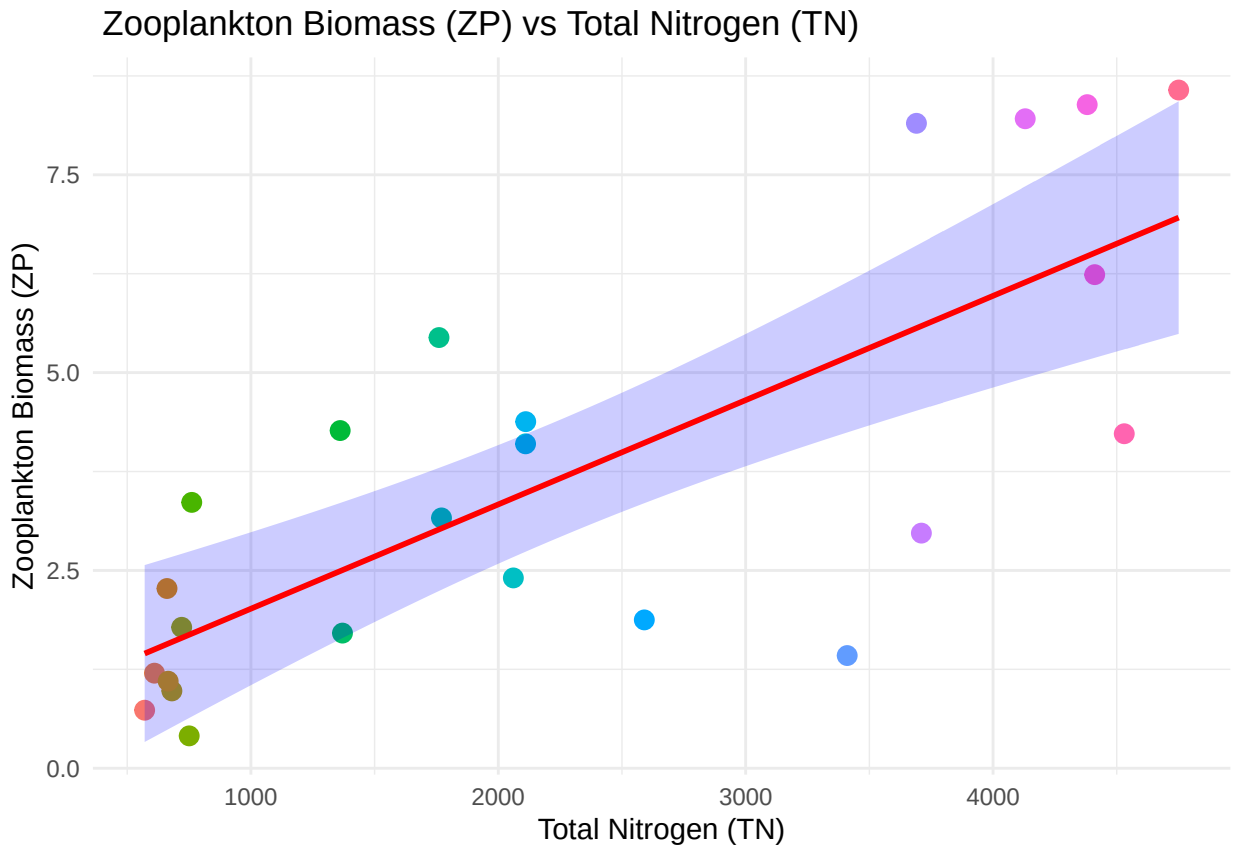
```
zoops<-read.table("C:/Users/ADMIN/OneDrive/Documents/GitHub/QB2025_Nambiar/Week1-RStudio/data/zoop_nuts

lm_model <- lm(ZP ~ TN, data = zoops)

summary(lm_model)
```

```
##
## Call:
## lm(formula = ZP ~ TN, data = zoops)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7690 -0.8491 -0.0709  1.6238  2.5888
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6977712  0.6496312   1.074    0.294
## TN          0.0013181  0.0002431   5.421 1.91e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.75 on 22 degrees of freedom
## Multiple R-squared:  0.5719, Adjusted R-squared:  0.5525
## F-statistic: 29.39 on 1 and 22 DF,  p-value: 1.911e-05
```

```
ggplot(zoops, aes(x = TN, y = ZP)) +
  geom_point(aes(color = factor(TN)), size = 3) +  # Points with categorical labels by TN
  geom_smooth(method = "lm", se = TRUE, color = "red", fill = "blue", alpha = 0.2) +  # Regression line
  labs(title = " Zooplankton Biomass (ZP) vs Total Nitrogen (TN)",
       x = "Total Nitrogen (TN)",
       y = "Zooplankton Biomass (ZP)") +
  theme_minimal() +
  theme(legend.position = "none")
```

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



Question 5: Interpret the results from the regression model

Answer 5: Based on the output and figure we see that the interescept is at 0.69 (non-seignficant)

and the slope is 0.0012(significant becuase p<0.05). This means that increase in 1 unit of TN there is increase in 0.0013 units increase in the total biomass.

Analysis of Variance (ANOVA)

Using the R code chunk below, do the following: 1) Order the nutrient treatments from low to high (see handout). 2) Produce a barplot to visualize zooplankton biomass in each nutrient treatment. 3) Include error bars (+/- 1 sem) on your plot and label the axes appropriately. 4) Use a one-way analysis of variance (ANOVA) to test the null hypothesis that zooplankton biomass is affected by the nutrient treatment.

```
zoop1<-as.data.frame(zoops)
zoop1
```

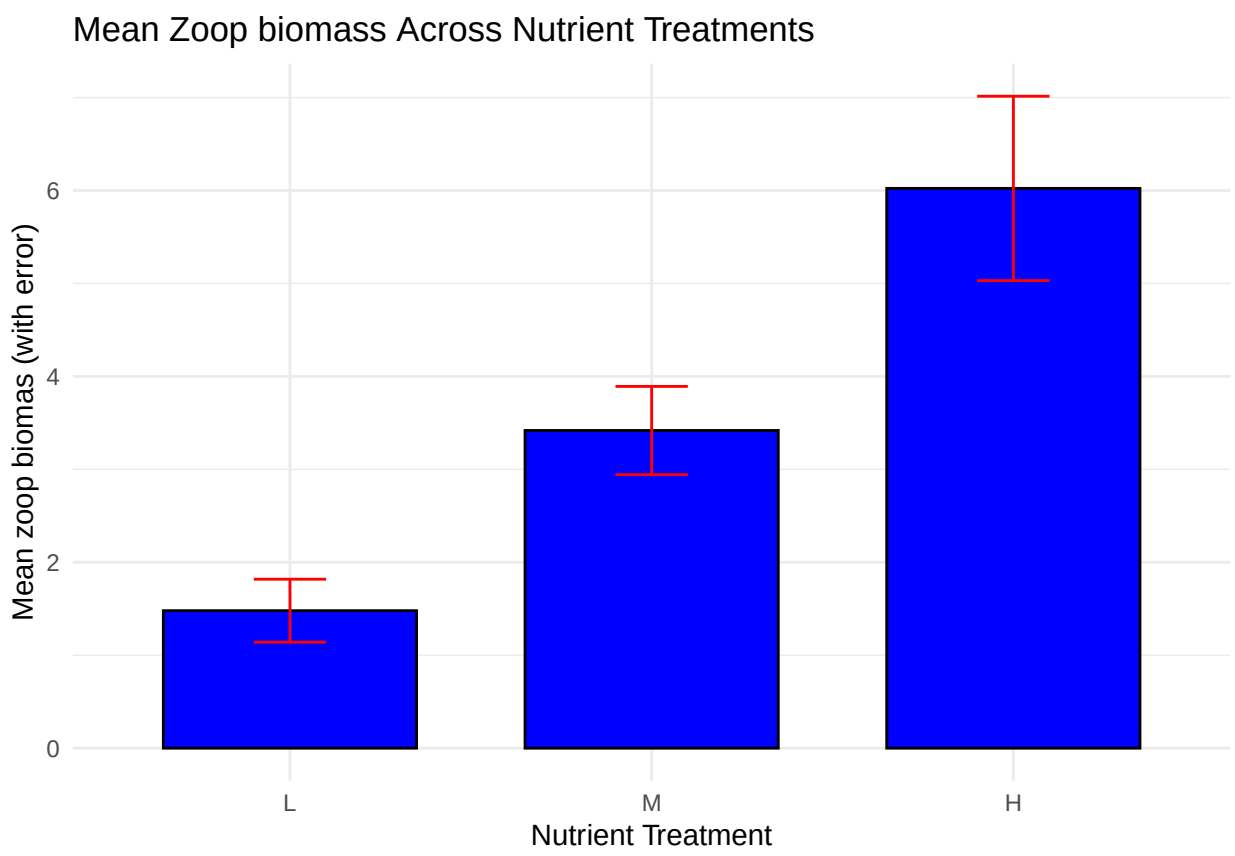
```
##    TANK NUTS     TP     TN   SRP     TIN  CHLA     ZP
## 1    34    L  20.31  720.1  4.02  131.62  1.52 1.7808
## 2    14    L  25.55  750.5  1.56  141.10  4.00 0.4090
## 3    23    L  14.22  610.1  4.97  107.70  0.61 1.2014
## 4    16    L  39.11  760.9  2.89   71.28  0.53 3.3598
## 5    21    L  20.09  570.4  5.11   80.40  1.44 0.7332
## 6     5    L  15.75  680.5  4.68  135.77  1.19 0.9773
## 7    25    L  19.55  665.5  5.00   79.40  0.37 1.0999
## 8    27    L  16.19  660.8  0.10  100.91  0.72 2.2714
## 9    30    M  29.46 1770.4  7.90 1329.26  6.93 3.1633
## 10   28    M  37.88 2590.3  3.92 1163.64  0.94 1.8747
## 11   35    M  30.26 2110.9  4.45 1850.18  1.36 4.3802
## 12   36    M  36.94 2060.9  5.14  249.93 38.38 2.4051
## 13   12    M  34.73 1370.1  4.69  420.01 15.99 1.7079
## 14   22    M  26.00 2110.3  5.35 1466.70  0.95 4.0999
## 15   18    M  28.50 1760.4  7.15 1351.83  1.36 5.4430
## 16   15    M  35.33 1360.8  5.96 1036.27  2.13 4.2677
## 17   17    H  41.56 4130.1 20.34 3421.43  1.44 8.2084
## 18   10    H  53.50 4530.4 33.57 4042.10  0.93 4.2273
## 19   29    H  99.07 4410.9 11.57 3307.05  0.61 6.2381
## 20    6    H 128.04 4750.4 26.27 3686.17  1.27 8.5713
## 21   24    H  33.47 3410.4  9.32 2791.52  1.11 1.4240
## 22   19    H  52.41 3710.3  3.23 2890.73 17.59 2.9714
## 23    4    H  42.21 3690.4 12.71 3041.75  1.08 8.1509
## 24   11    H  77.65 4380.6 21.86 3041.75  1.08 8.3868
```

```
colnames(zoop1)
```

```
## [1] "TANK" "NUTS" "TP"   "TN"   "SRP"  "TIN"  "CHLA" "ZP"
```

```
zoop_summary <- zoop1 %>%
  group_by(NUTS) %>%
  summarize(
    mean_ZP = mean(ZP, na.rm = TRUE),
    sem_ZP = se(ZP))%>%
  mutate(NUTS = factor(NUTS, levels = c("L", "M", "H")))

####----- plots
```

```r
ggplot(zoop_summary, aes(x = NUTS, y = mean_ZP)) +
  geom_bar(stat = "identity", fill = "blue", color = "black", width = 0.7) +
  geom_errorbar(aes(ymin = mean_ZP - sem_ZP, ymax = mean_ZP + sem_ZP),
                width = 0.2, color = "red") +
  labs(
    title = "Mean Zoop biomass Across Nutrient Treatments",
    x = "Nutrient Treatment",
    y = "Mean zoop biomas (with error)"
  ) +
  theme_minimal()
```




```r
# One-way ANOVA
zoop_anova <- aov(ZP ~ NUTS, data = zoop1)
summary(zoop_anova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## NUTS         2  83.15   41.58   11.77 0.000372 ***
## Residuals   21  74.16    3.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SYNTHESIS: SITE-BY-SPECIES MATRIX

In the R code chunk below, load the zoops.txt data set in your **3.RStudio** data folder. Create a site-by-species matrix (or dataframe) that does *not* include TANK or NUTS. The remaining columns of data refer

to the biomass (µg/L) of different zooplankton taxa:

- CAL = calanoid copepods
- DIAP = *Diaphanasoma* sp.
- CYL = cyclopoid copepods
- BOSM = *Bosmina* sp.
- SIMO = *Simocephallus* sp.
- CERI = *Ceriodaphnia* sp.
- NAUP = naupuli (immature copepod)
- DLUM = *Daphnia lumholtzi*
- CHYD = *Chydorus* sp.

Question 6: With the visualization and statistical tools that we learned about in the **3. RStudio** handout, use the site-by-species matrix to assess whether and how different zooplankton taxa were responsible for the total biomass (ZP) response to nutrient enrichment. Describe what you learned below in the "Answer" section and include appropriate code in the R chunk.

```
zoops<-read.table("C:/Users/ADMIN/OneDrive/Documents/GitHub/QB2025_Nambiar/Week1-RStudio/data/zoops.txt
zoops
```

```
##    TANK NUTS   CAL  DIAP  CYCL BOSM   SIMO  CERI NAUP DLUM   CHYD
## 1     5    L  70.5   0.0  66.1  2.2  417.8 159.8  0.0  0.0  266.9
## 2    14    L  27.1  19.2 129.6  0.0    0.0  79.4  0.0  0.0  158.7
## 3    16    L   5.3   8.8  12.7  0.0   73.1 107.5  1.2  0.0 3158.2
## 4    21    L  79.2  17.9 141.3  3.4    0.0 199.0  0.0  0.0  298.5
## 5    23    L  31.4   0.0  11.0  0.0  482.0 101.9  0.0  0.0  580.2
## 6    25    L  22.7 285.1 153.0  0.0  241.5 135.5  1.2  6.6  262.4
## 7    27    L   0.0   2.3  11.0  0.0   73.1 185.0  1.6  0.0 2004.4
## 8    34    L  35.7  65.9 102.9  0.0    0.0 318.5  3.1  0.0 1260.7
## 9    12    M  74.8 178.7 266.5  0.0    0.0   1.9  0.0  0.0 1190.9
## 10   15    M   5.3   4.9  87.8  0.0 1099.2 136.4  1.4  0.0 2939.6
## 11   18    M  18.4   2.3  29.4  0.0  393.8 147.6  1.2  0.0 4857.3
## 12   22    M  14.0   2.3  37.7  0.0 1251.5  74.8  0.0  0.0 2725.5
## 13   28    M  14.0   2.3 132.9  0.0  818.6  98.1  1.2  0.0  814.5
## 14   30    M  48.8   2.3 107.9  2.2    9.0 132.7  0.0  0.0 2867.5
## 15   35    M   0.0   0.0  17.7  0.0  145.3  19.7  0.0  0.0 4201.6
## 16   36    M 292.0 269.5 373.4 10.7    0.0   8.5  1.2  0.0 1456.8
## 17    4    H   9.7   0.0  41.1  0.0 2397.8   9.4  0.0  0.0 5697.9
## 18    6    H   0.0   2.3   0.0  0.0  225.5  24.3  0.0  0.0 8323.2
## 19   10    H   5.3   0.0  86.2  0.0  465.9 527.7  1.2  0.0 3146.9
## 20   11    H  14.0   7.5  69.5  0.0  594.2  78.5  0.0  0.0 7629.2
## 21   17    H   0.0  24.4 101.2  0.0  313.6 176.6  0.0  0.0 7597.6
## 22   19    H   0.0   7.5 253.2  8.3    0.0 112.1  1.6  0.0 2594.8
## 23   24    H   5.3   2.3  96.2  0.0  786.6  76.6  0.0  0.0  463.0
## 24   29    H   0.0   2.3  66.1  0.0  826.7  85.1  0.0  0.0 5263.0
```

```r
#get a new column for row means
zoops <- zoops %>%
  mutate(mean_bio = rowMeans(select(., CAL:CHYD), na.rm = TRUE))

zoops
```

```
##    TANK NUTS   CAL  DIAP  CYCL BOSM   SIMO  CERI NAUP DLUM   CHYD  mean_bio
## 1     5    L  70.5   0.0  66.1  2.2  417.8 159.8  0.0  0.0  266.9 109.25556
## 2    14    L  27.1  19.2 129.6  0.0    0.0  79.4  0.0  0.0  158.7  46.00000
## 3    16    L   5.3   8.8  12.7  0.0   73.1 107.5  1.2  0.0 3158.2 374.08889
## 4    21    L  79.2  17.9 141.3  3.4    0.0 199.0  0.0  0.0  298.5  82.14444
## 5    23    L  31.4   0.0  11.0  0.0  482.0 101.9  0.0  0.0  580.2 134.05556
## 6    25    L  22.7 285.1 153.0  0.0  241.5 135.5  1.2  6.6  262.4 123.11111
## 7    27    L   0.0   2.3  11.0  0.0   73.1 185.0  1.6  0.0 2004.4 253.04444
## 8    34    L  35.7  65.9 102.9  0.0    0.0 318.5  3.1  0.0 1260.7 198.53333
## 9    12    M  74.8 178.7 266.5  0.0    0.0   1.9  0.0  0.0 1190.9 190.31111
## 10   15    M   5.3   4.9  87.8  0.0 1099.2 136.4  1.4  0.0 2939.6 474.95556
## 11   18    M  18.4   2.3  29.4  0.0  393.8 147.6  1.2  0.0 4857.3 605.55556
## 12   22    M  14.0   2.3  37.7  0.0 1251.5  74.8  0.0  0.0 2725.5 456.20000
## 13   28    M  14.0   2.3 132.9  0.0  818.6  98.1  1.2  0.0  814.5 209.06667
## 14   30    M  48.8   2.3 107.9  2.2    9.0 132.7  0.0  0.0 2867.5 352.26667
## 15   35    M   0.0   0.0  17.7  0.0  145.3  19.7  0.0  0.0 4201.6 487.14444
## 16   36    M 292.0 269.5 373.4 10.7    0.0   8.5  1.2  0.0 1456.8 268.01111
## 17    4    H   9.7   0.0  41.1  0.0 2397.8   9.4  0.0  0.0 5697.9 906.21111
## 18    6    H   0.0   2.3   0.0  0.0  225.5  24.3  0.0  0.0 8323.2 952.81111
## 19   10    H   5.3   0.0  86.2  0.0  465.9 527.7  1.2  0.0 3146.9 470.35556
## 20   11    H  14.0   7.5  69.5  0.0  594.2  78.5  0.0  0.0 7629.2 932.54444
## 21   17    H   0.0  24.4 101.2  0.0  313.6 176.6  0.0  0.0 7597.6 912.60000
## 22   19    H   0.0   7.5 253.2  8.3    0.0 112.1  1.6  0.0 2594.8 330.83333
## 23   24    H   5.3   2.3  96.2  0.0  786.6  76.6  0.0  0.0  463.0 158.88889
## 24   29    H   0.0   2.3  66.1  0.0  826.7  85.1  0.0  0.0 5263.0 693.68889
```

```r
zoop_nuts<-read.table("C:/Users/ADMIN/OneDrive/Documents/GitHub/QB2025_Nambiar/Week1-RStudio/data/zoop_n

zoop_nuts
```

```
##    TANK NUTS     TP     TN   SRP     TIN  CHLA     ZP
## 1    34    L  20.31  720.1  4.02  131.62  1.52 1.7808
## 2    14    L  25.55  750.5  1.56  141.10  4.00 0.4090
## 3    23    L  14.22  610.1  4.97  107.70  0.61 1.2014
## 4    16    L  39.11  760.9  2.89   71.28  0.53 3.3598
## 5    21    L  20.09  570.4  5.11   80.40  1.44 0.7332
## 6     5    L  15.75  680.5  4.68  135.77  1.19 0.9773
## 7    25    L  19.55  665.5  5.00   79.40  0.37 1.0999
## 8    27    L  16.19  660.8  0.10  100.91  0.72 2.2714
## 9    30    M  29.46 1770.4  7.90 1329.26  6.93 3.1633
## 10   28    M  37.88 2590.3  3.92 1163.64  0.94 1.8747
## 11   35    M  30.26 2110.9  4.45 1850.18  1.36 4.3802
## 12   36    M  36.94 2060.9  5.14  249.93 38.38 2.4051
## 13   12    M  34.73 1370.1  4.69  420.01 15.99 1.7079
## 14   22    M  26.00 2110.3  5.35 1466.70  0.95 4.0999
## 15   18    M  28.50 1760.4  7.15 1351.83  1.36 5.4430
```

```
## 16   15    M  35.33 1360.8  5.96 1036.27  2.13 4.2677
## 17   17    H  41.56 4130.1 20.34 3421.43  1.44 8.2084
## 18   10    H  53.50 4530.4 33.57 4042.10  0.93 4.2273
## 19   29    H  99.07 4410.9 11.57 3307.05  0.61 6.2381
## 20    6    H 128.04 4750.4 26.27 3686.17  1.27 8.5713
## 21   24    H  33.47 3410.4  9.32 2791.52  1.11 1.4240
## 22   19    H  52.41 3710.3  3.23 2890.73 17.59 2.9714
## 23    4    H  42.21 3690.4 12.71 3041.75  1.08 8.1509
## 24   11    H  77.65 4380.6 21.86 3041.75  1.08 8.3868
```

```
merged_zoops <- left_join(zoops, zoop_nuts, by = "TANK")

merged_zoops <- merged_zoops %>%
  select(-NUTS.y) %>%      # Remove the NUTS.y column
  rename(NUTS = NUTS.x)

merged_zoops
```

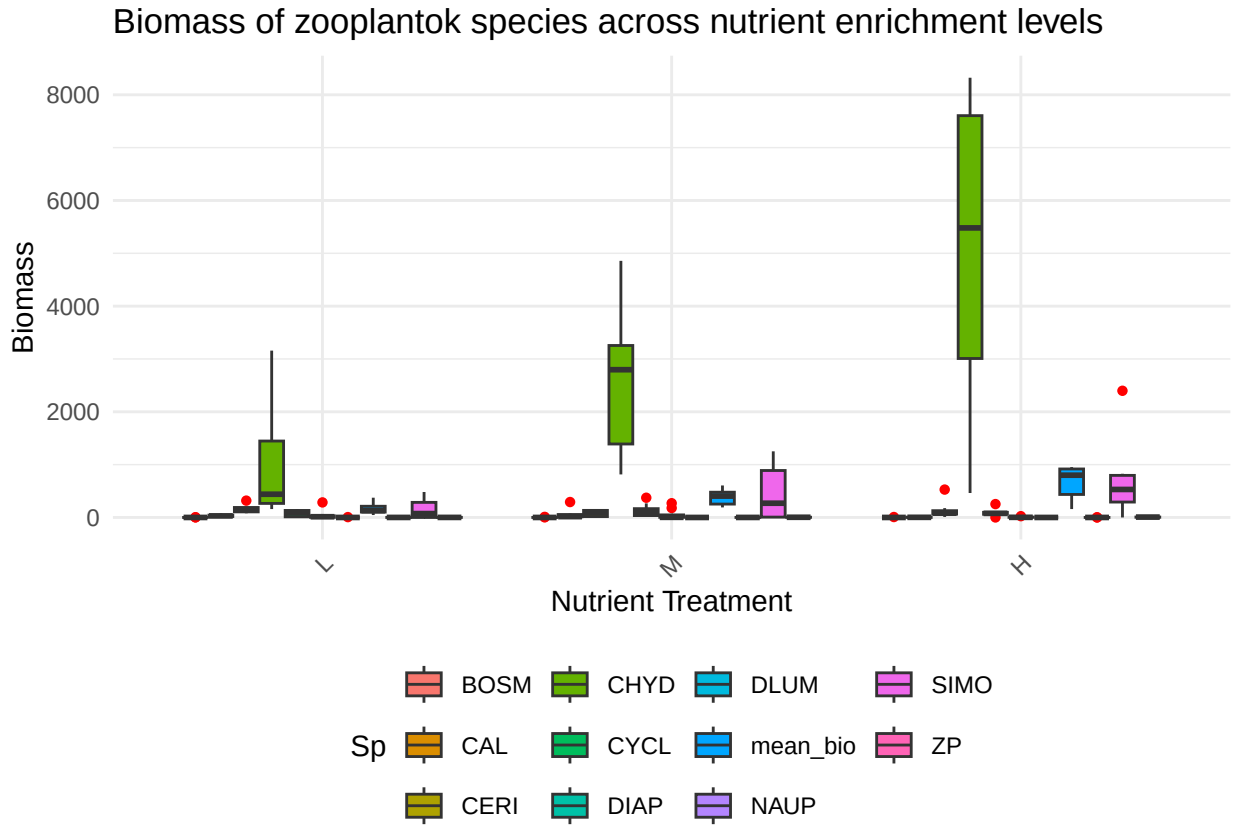
```
##    TANK NUTS   CAL  DIAP  CYCL BOSM   SIMO  CERI NAUP DLUM   CHYD  mean_bio
## 1     5    L  70.5   0.0  66.1  2.2  417.8 159.8  0.0  0.0  266.9 109.25556
## 2    14    L  27.1  19.2 129.6  0.0    0.0  79.4  0.0  0.0  158.7  46.00000
## 3    16    L   5.3   8.8  12.7  0.0   73.1 107.5  1.2  0.0 3158.2 374.08889
## 4    21    L  79.2  17.9 141.3  3.4    0.0 199.0  0.0  0.0  298.5  82.14444
## 5    23    L  31.4   0.0  11.0  0.0  482.0 101.9  0.0  0.0  580.2 134.05556
## 6    25    L  22.7 285.1 153.0  0.0  241.5 135.5  1.2  6.6  262.4 123.11111
## 7    27    L   0.0   2.3  11.0  0.0   73.1 185.0  1.6  0.0 2004.4 253.04444
## 8    34    L  35.7  65.9 102.9  0.0    0.0 318.5  3.1  0.0 1260.7 198.53333
## 9    12    M  74.8 178.7 266.5  0.0    0.0   1.9  0.0  0.0 1190.9 190.31111
## 10   15    M   5.3   4.9  87.8  0.0 1099.2 136.4  1.4  0.0 2939.6 474.95556
## 11   18    M  18.4   2.3  29.4  0.0  393.8 147.6  1.2  0.0 4857.3 605.55556
## 12   22    M  14.0   2.3  37.7  0.0 1251.5  74.8  0.0  0.0 2725.5 456.20000
## 13   28    M  14.0   2.3 132.9  0.0  818.6  98.1  1.2  0.0  814.5 209.06667
## 14   30    M  48.8   2.3 107.9  2.2    9.0 132.7  0.0  0.0 2867.5 352.26667
## 15   35    M   0.0   0.0  17.7  0.0  145.3  19.7  0.0  0.0 4201.6 487.14444
## 16   36    M 292.0 269.5 373.4 10.7    0.0   8.5  1.2  0.0 1456.8 268.01111
## 17    4    H   9.7   0.0  41.1  0.0 2397.8   9.4  0.0  0.0 5697.9 906.21111
## 18    6    H   0.0   2.3   0.0  0.0  225.5  24.3  0.0  0.0 8323.2 952.81111
## 19   10    H   5.3   0.0  86.2  0.0  465.9 527.7  1.2  0.0 3146.9 470.35556
## 20   11    H  14.0   7.5  69.5  0.0  594.2  78.5  0.0  0.0 7629.2 932.54444
## 21   17    H   0.0  24.4 101.2  0.0  313.6 176.6  0.0  0.0 7597.6 912.60000
## 22   19    H   0.0   7.5 253.2  8.3    0.0 112.1  1.6  0.0 2594.8 330.83333
## 23   24    H   5.3   2.3  96.2  0.0  786.6  76.6  0.0  0.0  463.0 158.88889
## 24   29    H   0.0   2.3  66.1  0.0  826.7  85.1  0.0  0.0 5263.0 693.68889
##        TP     TN   SRP     TIN  CHLA     ZP
## 1   15.75  680.5  4.68  135.77  1.19 0.9773
## 2   25.55  750.5  1.56  141.10  4.00 0.4090
## 3   39.11  760.9  2.89   71.28  0.53 3.3598
## 4   20.09  570.4  5.11   80.40  1.44 0.7332
## 5   14.22  610.1  4.97  107.70  0.61 1.2014
## 6   19.55  665.5  5.00   79.40  0.37 1.0999
## 7   16.19  660.8  0.10  100.91  0.72 2.2714
## 8   20.31  720.1  4.02  131.62  1.52 1.7808
## 9   34.73 1370.1  4.69  420.01 15.99 1.7079
## 10  35.33 1360.8  5.96 1036.27  2.13 4.2677
```

```
## 11  28.50 1760.4  7.15 1351.83  1.36 5.4430
## 12  26.00 2110.3  5.35 1466.70  0.95 4.0999
## 13  37.88 2590.3  3.92 1163.64  0.94 1.8747
## 14  29.46 1770.4  7.90 1329.26  6.93 3.1633
## 15  30.26 2110.9  4.45 1850.18  1.36 4.3802
## 16  36.94 2060.9  5.14  249.93 38.38 2.4051
## 17  42.21 3690.4 12.71 3041.75  1.08 8.1509
## 18 128.04 4750.4 26.27 3686.17  1.27 8.5713
## 19  53.50 4530.4 33.57 4042.10  0.93 4.2273
## 20  77.65 4380.6 21.86 3041.75  1.08 8.3868
## 21  41.56 4130.1 20.34 3421.43  1.44 8.2084
## 22  52.41 3710.3  3.23 2890.73 17.59 2.9714
## 23  33.47 3410.4  9.32 2791.52  1.11 1.4240
## 24  99.07 4410.9 11.57 3307.05  0.61 6.2381
```

```
#wranling dataset for plot
long_zoops <- merged_zoops %>%
  pivot_longer(cols = c(CAL:mean_bio, ZP), names_to = "Sp", values_to = "biomass") %>%
  mutate(NUTS = factor(NUTS, levels = c("L", "M", "H")))  # needed for proper plot
long_zoops
```

```
## # A tibble: 264 x 9
##     TANK NUTS     TP    TN   SRP   TIN  CHLA Sp       biomass
##    <int> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>      <dbl>
##  1     5 L      15.8  680.  4.68  136.  1.19 CAL         70.5
##  2     5 L      15.8  680.  4.68  136.  1.19 DIAP         0
##  3     5 L      15.8  680.  4.68  136.  1.19 CYCL        66.1
##  4     5 L      15.8  680.  4.68  136.  1.19 BOSM         2.2
##  5     5 L      15.8  680.  4.68  136.  1.19 SIMO       418.
##  6     5 L      15.8  680.  4.68  136.  1.19 CERI       160.
##  7     5 L      15.8  680.  4.68  136.  1.19 NAUP         0
##  8     5 L      15.8  680.  4.68  136.  1.19 DLUM         0
##  9     5 L      15.8  680.  4.68  136.  1.19 CHYD       267.
## 10     5 L      15.8  680.  4.68  136.  1.19 mean_bio   109.
## # i 254 more rows
```

```
#lot to see difference in response to nutrient enrichment


#
ggplot(long_zoops, aes(x = NUTS, y = biomass, fill = Sp)) +
  geom_boxplot(outlier.color = "red", outlier.shape = 16, position = position_dodge(width = 0.8)) +
  labs(
    title = "Biomass of zooplantok species across nutrient enrichment levels",
    x = "Nutrient Treatment",
    y = "Biomass",
    fill = "Sp"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),  # Rotate x-axis labels
    legend.position = "bottom"  # Position the legend below
  )
```




```
##anova to see difference in the response of biomass to nutrient enrichmeent

anova <- aov(biomass ~ NUTS * Sp, data = long_zoops)

summary(anova)
```

```
##              Df    Sum Sq  Mean Sq F value   Pr(>F)
## NUTS          2   9346068  4673034  12.686 5.93e-06 ***
## Sp           10 176206725 17620673  47.834  < 2e-16 ***
## NUTS:Sp      20  60762829  3038141   8.247  < 2e-16 ***
## Residuals   231  85094585   368375
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
merged_zoops <- merged_zoops %>%
  mutate(NUTS_numeric = as.numeric(NUTS))  # L = 1, M = 2, H = 3 #already ordered
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `NUTS_numeric = as.numeric(NUTS)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
# Separate columns for correlation test
cor_data <- merged_zoops %>%
```

```
  select(NUTS_numeric, CAL:ZP)

# Perform Spearman correlation
spearman <- corr.test(cor_data, method = "spearman", adjust = "BH")
```

```
## Warning in sqrt(n - 2): NaNs produced
```

```
## Warning in corr.test(cor_data, method = "spearman", adjust = "BH"): Number of
## subjects must be greater than 3 to find confidence intervals.
```

```
## Warning in sqrt(n[lower.tri(n)] - 3): NaNs produced
```

```
spearman
```

```
## Call:corr.test(x = cor_data, method = "spearman", adjust = "BH")
## Correlation matrix
##              NUTS_numeric   CAL  DIAP  CYCL  BOSM  SIMO  CERI  NAUP  DLUM  CHYD
## NUTS_numeric           NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## CAL                    NA  1.00  0.31  0.47  0.42 -0.33 -0.01 -0.12  0.11 -0.59
## DIAP                   NA  0.31  1.00  0.66  0.15 -0.57  0.02  0.27  0.35 -0.24
## CYCL                   NA  0.47  0.66  1.00  0.48 -0.43  0.00  0.16  0.26 -0.48
## BOSM                   NA  0.42  0.15  0.48  1.00 -0.47  0.10  0.05 -0.11 -0.28
## SIMO                   NA -0.33 -0.57 -0.43 -0.47  1.00 -0.13 -0.22 -0.02  0.32
## CERI                   NA -0.01  0.02  0.00  0.10 -0.13  1.00  0.46  0.11 -0.15
## NAUP                   NA -0.12  0.27  0.16  0.05 -0.22  0.46  1.00  0.17 -0.09
## DLUM                   NA  0.11  0.35  0.26 -0.11 -0.02  0.11  0.17  1.00 -0.32
## CHYD                   NA -0.59 -0.24 -0.48 -0.28  0.32 -0.15 -0.09 -0.32  1.00
## mean_bio               NA -0.61 -0.24 -0.44 -0.30  0.41 -0.19 -0.08 -0.26  0.98
## TP                     NA -0.53 -0.02  0.00 -0.12  0.23 -0.32 -0.05 -0.26  0.73
## TN                     NA -0.57 -0.23 -0.07 -0.16  0.36 -0.34 -0.20 -0.26  0.70
## SRP                    NA -0.16 -0.22 -0.14 -0.13  0.49 -0.09 -0.40 -0.05  0.59
## TIN                    NA -0.55 -0.36 -0.14 -0.18  0.42 -0.24 -0.30 -0.32  0.69
## CHLA                   NA  0.31  0.38  0.55  0.53 -0.56 -0.13  0.02 -0.35 -0.07
## ZP                     NA -0.61 -0.24 -0.44 -0.30  0.41 -0.19 -0.08 -0.26  0.98
##              mean_bio    TP    TN   SRP   TIN  CHLA    ZP
## NUTS_numeric       NA    NA    NA    NA    NA    NA    NA
## CAL             -0.61 -0.53 -0.57 -0.16 -0.55  0.31 -0.61
## DIAP            -0.24 -0.02 -0.23 -0.22 -0.36  0.38 -0.24
## CYCL            -0.44  0.00 -0.07 -0.14 -0.14  0.55 -0.44
## BOSM            -0.30 -0.12 -0.16 -0.13 -0.18  0.53 -0.30
## SIMO             0.41  0.23  0.36  0.49  0.42 -0.56  0.41
## CERI            -0.19 -0.32 -0.34 -0.09 -0.24 -0.13 -0.19
## NAUP            -0.08 -0.05 -0.20 -0.40 -0.30  0.02 -0.08
## DLUM            -0.26 -0.26 -0.26 -0.05 -0.32 -0.35 -0.26
## CHYD             0.98  0.73  0.70  0.59  0.69 -0.07  0.98
## mean_bio         1.00  0.74  0.75  0.63  0.74 -0.07  1.00
## TP               0.74  1.00  0.89  0.54  0.76  0.04  0.74
## TN               0.75  0.89  1.00  0.65  0.94  0.02  0.75
## SRP              0.63  0.54  0.65  1.00  0.73 -0.06  0.63
## TIN              0.74  0.76  0.94  0.73  1.00  0.09  0.74
## CHLA            -0.07  0.04  0.02 -0.06  0.09  1.00 -0.07
## ZP               1.00  0.74  0.75  0.63  0.74 -0.07  1.00
```

```
## Sample Size
##              NUTS_numeric CAL DIAP CYCL BOSM SIMO CERI NAUP DLUM CHYD mean_bio
## NUTS_numeric            0   0    0    0    0    0    0    0    0    0        0
## CAL                     0  24   24   24   24   24   24   24   24   24       24
## DIAP                    0  24   24   24   24   24   24   24   24   24       24
## CYCL                    0  24   24   24   24   24   24   24   24   24       24
## BOSM                    0  24   24   24   24   24   24   24   24   24       24
## SIMO                    0  24   24   24   24   24   24   24   24   24       24
## CERI                    0  24   24   24   24   24   24   24   24   24       24
## NAUP                    0  24   24   24   24   24   24   24   24   24       24
## DLUM                    0  24   24   24   24   24   24   24   24   24       24
## CHYD                    0  24   24   24   24   24   24   24   24   24       24
## mean_bio                0  24   24   24   24   24   24   24   24   24       24
## TP                      0  24   24   24   24   24   24   24   24   24       24
## TN                      0  24   24   24   24   24   24   24   24   24       24
## SRP                     0  24   24   24   24   24   24   24   24   24       24
## TIN                     0  24   24   24   24   24   24   24   24   24       24
## CHLA                    0  24   24   24   24   24   24   24   24   24       24
## ZP                      0  24   24   24   24   24   24   24   24   24       24
##              TP TN SRP TIN CHLA ZP
## NUTS_numeric  0  0   0   0    0  0
## CAL          24 24  24  24   24 24
## DIAP         24 24  24  24   24 24
## CYCL         24 24  24  24   24 24
## BOSM         24 24  24  24   24 24
## SIMO         24 24  24  24   24 24
## CERI         24 24  24  24   24 24
## NAUP         24 24  24  24   24 24
## DLUM         24 24  24  24   24 24
## CHYD         24 24  24  24   24 24
## mean_bio     24 24  24  24   24 24
## TP           24 24  24  24   24 24
## TN           24 24  24  24   24 24
## SRP          24 24  24  24   24 24
## TIN          24 24  24  24   24 24
## CHLA         24 24  24  24   24 24
## ZP           24 24  24  24   24 24
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##              NUTS_numeric  CAL DIAP CYCL BOSM SIMO CERI NAUP DLUM CHYD mean_bio
## NUTS_numeric          NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN      NaN
## CAL                   NaN 0.00 0.29 0.06 0.12 0.26 0.99 0.76 0.77 0.01     0.01
## DIAP                  NaN 0.14 0.00 0.00 0.66 0.02 0.97 0.37 0.21 0.43     0.42
## CYCL                  NaN 0.02 0.00 0.00 0.06 0.10 1.00 0.63 0.40 0.06     0.09
## BOSM                  NaN 0.04 0.48 0.02 0.00 0.06 0.79 0.90 0.77 0.36     0.29
## SIMO                  NaN 0.12 0.00 0.03 0.02 0.00 0.72 0.47 0.97 0.28     0.12
## CERI                  NaN 0.98 0.93 0.99 0.64 0.56 0.00 0.07 0.77 0.66     0.55
## NAUP                  NaN 0.59 0.20 0.44 0.82 0.30 0.02 0.00 0.63 0.81     0.83
## DLUM                  NaN 0.62 0.09 0.23 0.62 0.94 0.62 0.43 0.00 0.28     0.40
## CHYD                  NaN 0.00 0.26 0.02 0.19 0.13 0.47 0.67 0.13 0.00     0.00
## mean_bio              NaN 0.00 0.25 0.03 0.15 0.05 0.36 0.72 0.23 0.00     0.00
## TP                    NaN 0.01 0.94 1.00 0.59 0.28 0.13 0.83 0.23 0.00     0.00
## TN                    NaN 0.00 0.28 0.73 0.46 0.08 0.10 0.36 0.23 0.00     0.00
## SRP                   NaN 0.44 0.31 0.53 0.53 0.02 0.68 0.06 0.83 0.00     0.00
## TIN                   NaN 0.01 0.08 0.51 0.41 0.04 0.26 0.16 0.13 0.00     0.00
```

```
## CHLA                  NaN 0.14 0.07 0.01 0.01 0.00 0.55 0.93 0.10 0.75     0.74
## ZP                    NaN 0.00 0.25 0.03 0.15 0.05 0.36 0.72 0.23 0.00     0.00
##                 TP   TN  SRP  TIN CHLA   ZP
## NUTS_numeric   NaN  NaN  NaN  NaN  NaN  NaN
## CAL           0.03 0.02 0.63 0.02 0.29 0.01
## DIAP          0.97 0.45 0.48 0.20 0.18 0.42
## CYCL          1.00 0.83 0.71 0.69 0.02 0.09
## BOSM          0.76 0.64 0.71 0.61 0.03 0.29
## SIMO          0.45 0.20 0.05 0.12 0.02 0.12
## CERI          0.28 0.24 0.81 0.43 0.72 0.55
## NAUP          0.90 0.55 0.15 0.31 0.97 0.83
## DLUM          0.40 0.40 0.90 0.28 0.23 0.40
## CHYD          0.00 0.00 0.01 0.00 0.84 0.00
## mean_bio      0.00 0.00 0.01 0.00 0.83 0.00
## TP            0.00 0.00 0.03 0.00 0.91 0.00
## TN            0.00 0.00 0.00 0.00 0.97 0.00
## SRP           0.01 0.00 0.00 0.00 0.85 0.01
## TIN           0.00 0.00 0.00 0.00 0.81 0.00
## CHLA          0.85 0.92 0.77 0.68 0.00 0.83
## ZP            0.00 0.00 0.00 0.00 0.74 0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

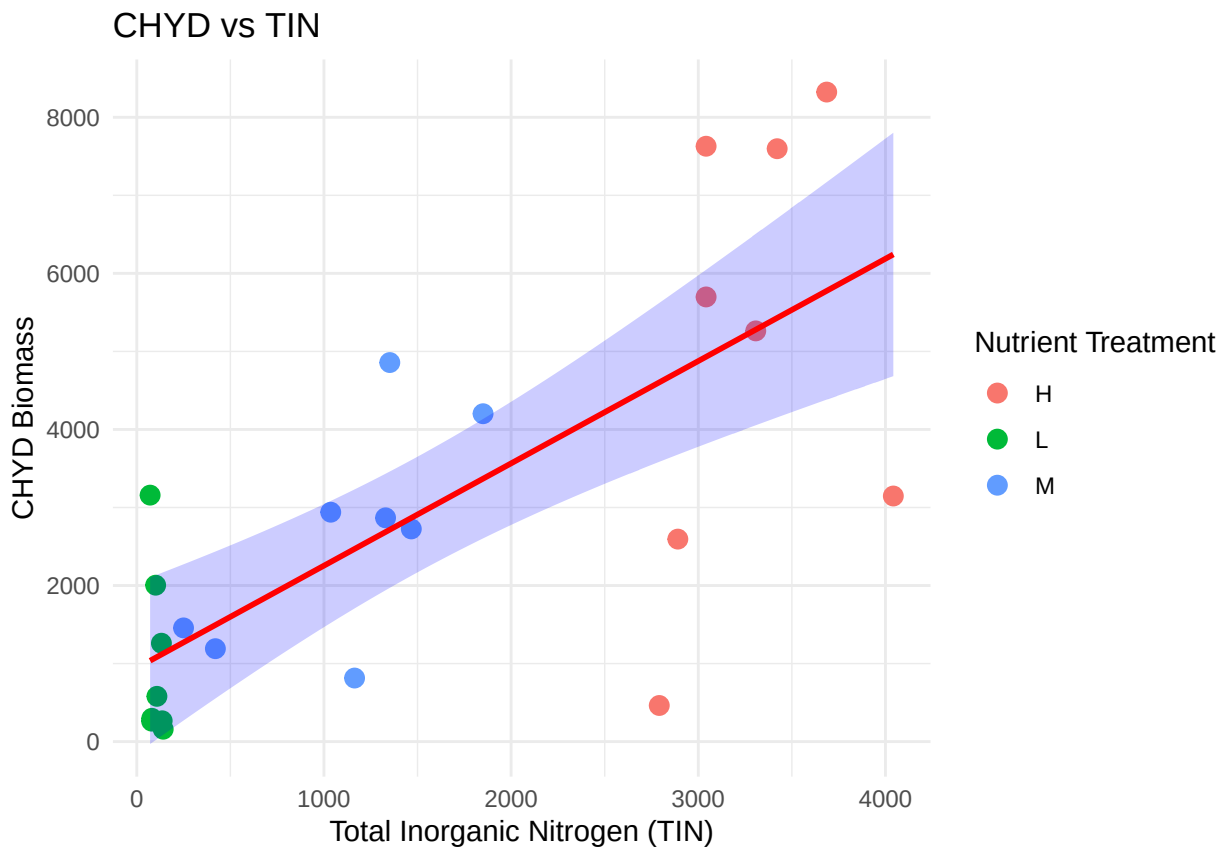
```
#pearson correlations

# remove unecessary data for correlations
cor_data <- merged_zoops %>%
  select(-TANK, -NUTS, -NUTS_numeric)

pearson <- corr.test(cor_data, method = "pearson", adjust = "BH")
pearson
```

```
## Call:corr.test(x = cor_data, method = "pearson", adjust = "BH")
## Correlation matrix
##             CAL  DIAP  CYCL  BOSM  SIMO  CERI  NAUP  DLUM  CHYD mean_bio    TP
## CAL        1.00  0.64  0.71  0.73 -0.27 -0.19  0.06 -0.03 -0.32    -0.31 -0.19
## DIAP       0.64  1.00  0.69  0.38 -0.29 -0.17  0.22  0.64 -0.31    -0.30 -0.17
## CYCL       0.71  0.69  1.00  0.75 -0.32 -0.13  0.19  0.13 -0.37    -0.36 -0.12
## BOSM       0.73  0.38  0.75  1.00 -0.31 -0.14  0.18 -0.09 -0.21    -0.21 -0.04
## SIMO      -0.27 -0.29 -0.32 -0.31  1.00 -0.18 -0.24 -0.08  0.26     0.43  0.11
## CERI      -0.19 -0.17 -0.13 -0.14 -0.18  1.00  0.47  0.02 -0.14    -0.14 -0.17
## NAUP       0.06  0.22  0.19  0.18 -0.24  0.47  1.00  0.15 -0.24    -0.24 -0.22
## DLUM      -0.03  0.64  0.13 -0.09 -0.08  0.02  0.15  1.00 -0.22    -0.21 -0.16
## CHYD      -0.32 -0.31 -0.37 -0.21  0.26 -0.14 -0.24 -0.22  1.00     0.98  0.72
## mean_bio -0.31 -0.30 -0.36 -0.21  0.43 -0.14 -0.24 -0.21  0.98     1.00  0.70
## TP        -0.19 -0.17 -0.12 -0.04  0.11 -0.17 -0.22 -0.16  0.72     0.70  1.00
## TN        -0.25 -0.26 -0.04  0.00  0.35  0.00 -0.25 -0.22  0.73     0.76  0.79
## SRP       -0.21 -0.19 -0.21 -0.21  0.19  0.37 -0.22 -0.10  0.66     0.68  0.65
## TIN       -0.40 -0.38 -0.19 -0.11  0.38  0.10 -0.27 -0.22  0.73     0.76  0.72
## CHLA       0.83  0.60  0.87  0.86 -0.31 -0.29  0.14 -0.10 -0.18    -0.18 -0.02
## ZP        -0.31 -0.30 -0.36 -0.21  0.43 -0.14 -0.24 -0.21  0.98     1.00  0.70
##               TN   SRP   TIN  CHLA    ZP
## CAL       -0.25 -0.21 -0.40  0.83 -0.31
## DIAP      -0.26 -0.19 -0.38  0.60 -0.30
```

```
## CYCL     -0.04 -0.21 -0.19  0.87 -0.36
## BOSM      0.00 -0.21 -0.11  0.86 -0.21
## SIMO      0.35  0.19  0.38 -0.31  0.43
## CERI      0.00  0.37  0.10 -0.29 -0.14
## NAUP     -0.25 -0.22 -0.27  0.14 -0.24
## DLUM     -0.22 -0.10 -0.22 -0.10 -0.21
## CHYD      0.73  0.66  0.73 -0.18  0.98
## mean_bio  0.76  0.68  0.76 -0.18  1.00
## TP        0.79  0.65  0.72 -0.02  0.70
## TN        1.00  0.78  0.97  0.00  0.76
## SRP       0.78  1.00  0.80 -0.19  0.68
## TIN       0.97  0.80  1.00 -0.16  0.76
## CHLA      0.00 -0.19 -0.16  1.00 -0.18
## ZP        0.76  0.68  0.76 -0.18  1.00
## Sample Size
## [1] 24
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##           CAL DIAP CYCL BOSM SIMO CERI NAUP DLUM CHYD mean_bio   TP   TN  SRP
## CAL      0.00 0.00 0.00 0.00 0.43 0.54 0.85 0.91 0.33     0.35 0.54 0.49 0.52
## DIAP     0.00 0.00 0.00 0.22 0.39 0.55 0.52 0.00 0.35     0.36 0.55 0.46 0.54
## CYCL     0.00 0.00 0.00 0.00 0.33 0.63 0.54 0.65 0.23     0.25 0.67 0.89 0.52
## BOSM     0.00 0.07 0.00 0.00 0.35 0.61 0.54 0.75 0.52     0.52 0.89 0.99 0.52
## SIMO     0.20 0.17 0.12 0.14 0.00 0.54 0.50 0.78 0.45     0.13 0.68 0.27 0.54
## CERI     0.37 0.42 0.54 0.51 0.39 0.00 0.07 0.96 0.63     0.61 0.56 1.00 0.23
## NAUP     0.79 0.31 0.39 0.40 0.27 0.02 0.00 0.61 0.50     0.49 0.52 0.48 0.52
## DLUM     0.88 0.00 0.56 0.69 0.72 0.93 0.49 0.00 0.52     0.52 0.58 0.52 0.72
## CHYD     0.13 0.14 0.08 0.33 0.22 0.53 0.26 0.29 0.00     0.00 0.00 0.00 0.00
## mean_bio 0.15 0.16 0.09 0.31 0.04 0.51 0.25 0.33 0.00     0.00 0.00 0.00 0.00
## TP       0.37 0.42 0.58 0.85 0.60 0.44 0.30 0.45 0.00     0.00 0.00 0.00 0.00
## TN       0.25 0.22 0.84 0.99 0.10 1.00 0.23 0.30 0.00     0.00 0.00 0.00 0.00
## SRP      0.33 0.37 0.34 0.33 0.38 0.08 0.29 0.65 0.00     0.00 0.00 0.00 0.00
## TIN      0.05 0.07 0.38 0.60 0.07 0.66 0.20 0.31 0.00     0.00 0.00 0.00 0.00
## CHLA     0.00 0.00 0.00 0.00 0.14 0.17 0.50 0.65 0.40     0.39 0.94 0.98 0.38
## ZP       0.15 0.16 0.09 0.31 0.04 0.51 0.25 0.33 0.00     0.00 0.00 0.00 0.00
##           TIN CHLA   ZP
## CAL      0.18 0.00 0.35
## DIAP     0.22 0.01 0.36
## CYCL     0.54 0.00 0.25
## BOSM     0.68 0.00 0.52
## SIMO     0.22 0.35 0.13
## CERI     0.72 0.38 0.61
## NAUP     0.43 0.61 0.49
## DLUM     0.52 0.72 0.52
## CHYD     0.00 0.54 0.00
## mean_bio 0.00 0.54 0.00
## TP       0.00 0.96 0.00
## TN       0.00 0.99 0.00
## SRP      0.00 0.54 0.00
## TIN      0.00 0.59 0.00
## CHLA     0.46 0.00 0.54
## ZP       0.00 0.39 0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

```
#####-CHYD the most important species
# specific nutrient effects on biomass--TIN

ggplot(merged_zoops, aes(x = TIN, y = CHYD)) +
  geom_point(aes(color = NUTS), size = 3) +  # Scatter plot with points colored by NUTS
  geom_smooth(method = "lm", se = TRUE, color = "red", fill = "blue", alpha = 0.2) +  # Regression line
  labs(
    title = "CHYD vs TIN",
    x = "Total Inorganic Nitrogen (TIN)",
    y = "CHYD Biomass",
    color = "Nutrient Treatment"
  ) +
  theme_minimal()
```

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



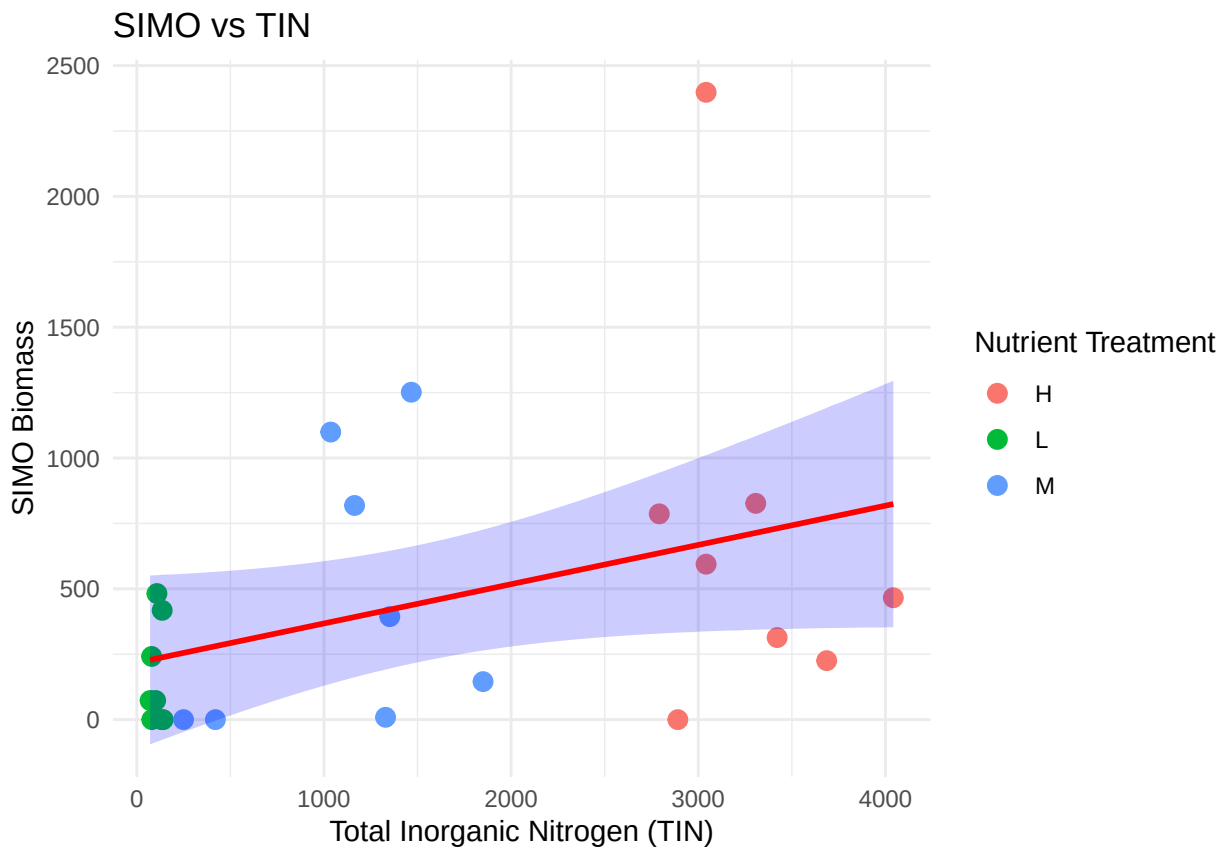

```
# Fit a linear model
lm_CHYD <- lm(CHYD ~ TIN, data = merged_zoops)


summary(lm_CHYD)
```

```
##
## Call:
```

```
## lm(formula = CHYD ~ TIN, data = merged_zoops)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4140.2  -802.5    64.3   856.5  2698.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  943.490    528.549   1.785    0.088 .
## TIN            1.311      0.260   5.042 4.76e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1751 on 22 degrees of freedom
## Multiple R-squared:  0.536,  Adjusted R-squared:  0.5149
## F-statistic: 25.42 on 1 and 22 DF,  p-value: 4.764e-05
```

```
#####-SIMO
# specific nutrient effects on biomass--TIN

ggplot(merged_zoops, aes(x = TIN, y = SIMO)) +
  geom_point(aes(color = NUTS), size = 3) +  # Scatter plot with points colored by NUTS
  geom_smooth(method = "lm", se = TRUE, color = "red", fill = "blue", alpha = 0.2) +  # Regression line
  labs(
    title = "SIMO vs TIN",
    x = "Total Inorganic Nitrogen (TIN)",
    y = "SIMO Biomass",
    color = "Nutrient Treatment"
  ) +
  theme_minimal()
```

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

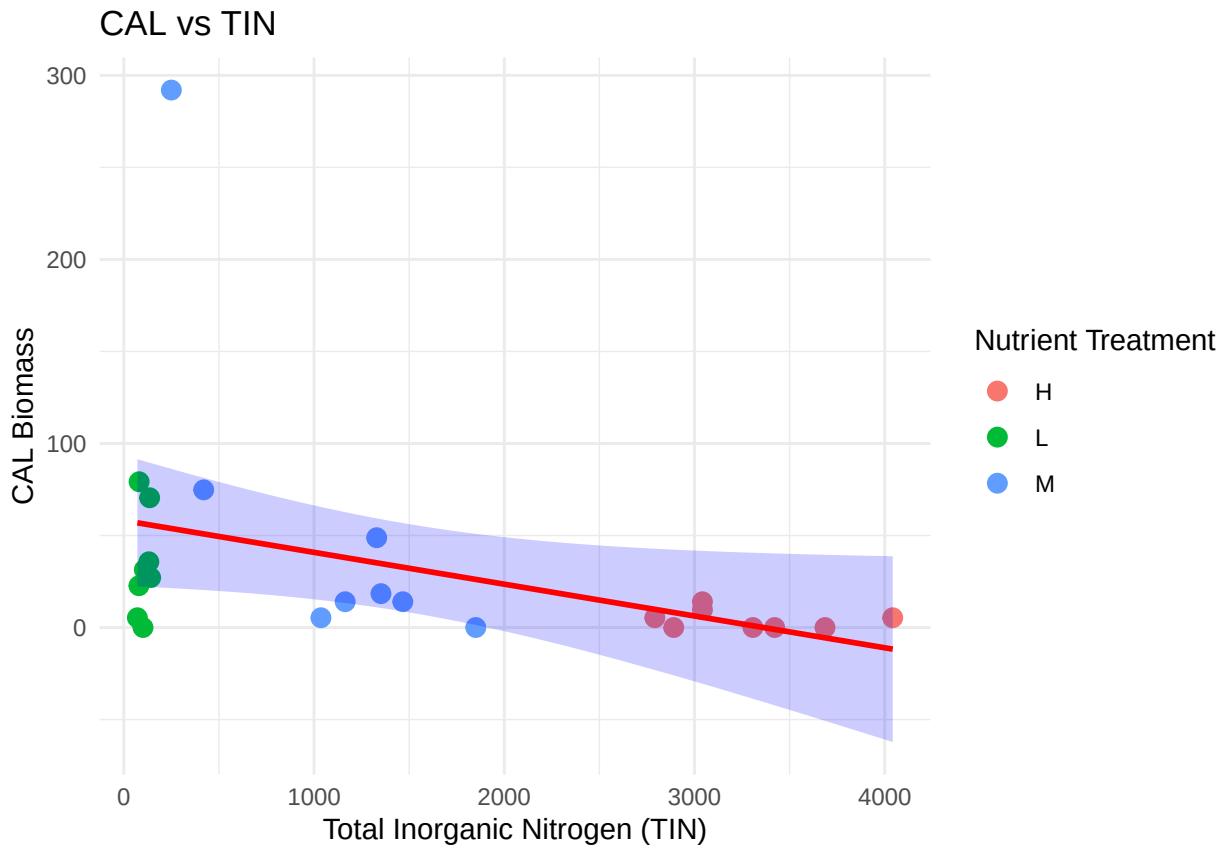



```
# Fit a linear model
lm_SIMO<- lm(SIMO ~ TIN, data = merged_zoops)


summary(lm_SIMO)
```

```
##
## Call:
## lm(formula = SIMO ~ TIN, data = merged_zoops)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -651.4 -297.9 -157.4  157.5 1723.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 217.60013  159.62684   1.363   0.1866
## TIN           0.15006    0.07853   1.911   0.0692 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 528.8 on 22 degrees of freedom
## Multiple R-squared:  0.1423, Adjusted R-squared:  0.1033
## F-statistic: 3.651 on 1 and 22 DF,  p-value: 0.06916
```

```
#####-CAL
# specific nutrient effects on biomass--TIN

ggplot(merged_zoops, aes(x = TIN, y = CAL)) +
  geom_point(aes(color = NUTS), size = 3) +  # Scatter plot with points colored by NUTS
  geom_smooth(method = "lm", se = TRUE, color = "red", fill = "blue", alpha = 0.2) +  # Regression line
  labs(
    title = "CAL vs TIN",
    x = "Total Inorganic Nitrogen (TIN)",
    y = "CAL Biomass",
    color = "Nutrient Treatment"
  ) +
  theme_minimal()
```

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




```
# Fit a linear model
lm_CAL<- lm(CAL ~ TIN, data = merged_zoops)


summary(lm_CAL)
```

```
##
## Call:
```

```
## lm(formula = CAL ~ TIN, data = merged_zoops)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -56.363 -25.167  -6.359   9.760 238.213
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 58.106371  17.107531   3.397  0.00259 **
## TIN         -0.017281   0.008417  -2.053  0.05213 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.67 on 22 degrees of freedom
## Multiple R-squared:  0.1608, Adjusted R-squared:  0.1227
## F-statistic: 4.216 on 1 and 22 DF,  p-value: 0.05213
```

Answer 6: Through ANOVA we see that both treatment type and taxonomic identity predict biomass. In the first figure we see that from Low to High nutrient enrichment treatments there is increase in biomass predominantly for the species "CHYD". Through the correlation matrix we see that CHYD responds most positively and strongly to nutrient enrichments for all variables. Since CHYD shows the most variation we can also compare CHYD biomass and how it is predicted by the nutrient TIN. We see that the nutrient TIN predicts CHYD biomass positively (estimate =1.311) significantly ($p<0.05$) while with other species it is weak, or non significant relationship for other species such as SIMO and CAL.

SUBMITTING YOUR WORKSHEET

Use Knitr to create a PDF of your completed **3.RStudio_Worksheet.Rmd** document, push the repo to GitHub, and create a pull request. Please make sure your updated repo include both the PDF and RMarkdown files.

This assignment is due on **Wednesday, January 22nd, 2025 at 12:00 PM (noon)**.