3. Worksheet: Basic R

Joshua; Z620: Quantitative Biodiversity, Indiana University

22 March, 2021

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, it is *imperative* that you **push** this file to your GitHub repo, at whatever stage you are. Ths 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, March 24th, 2021 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.

```
rm(list=ls())
getwd()
```

```
## [1] "C:/Users/joshu/quantbio/QB2021_Jones/2.Worksheets/3.RStudio"
```

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.

```
l <- 5
volume <- l^3

r <- 2
area = pi * r^2

theta <- pi/4
hyp <- sqrt(2)
opposite <- (sin(theta)*hyp)

favnum <- log(1996)
```

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.

```
x <- c(10,18,24,29,30)
w <- x*14
y <- (w+x)*15
```

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 <- c(10,20,30,40,50)
z <- k*x
d <- c(140,252,336,10,20,30,40)
```

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
sum(na.omit(v))
```

```
## [1] 292.6
```

```r
var(v,na.rm = TRUE)
```

```
## [1] 39.44
```

```r
sd <- sd(v,na.rm = TRUE)
print(sd)
```

```
## [1] 6.280127
```

```r
sem <- function(x){
  sd(na.omit(x))/sqrt(length(na.omit(x)))
}
sem(v)
```

```
## [1] 1.678435
```

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
col1 <- c(rnorm(5, mean = 8, sd = 2))
col2 <- c(rnorm(5, mean = 25, sd = 10))
mat1 <- cbind(col1, col2)
```

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 creates a normal distribution with mean and standard deviation defined by the inputs the first value controls the number of values in the distribution mean controls the mean fo the distribution sd controls the standard deviation of the distribution

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.

```
m <- as.matrix(read.table("data/matrix.txt", sep = "\t", header = FALSE))
dim(m)
```

```
## [1] 10  5
```

```
n <- t(m)
dim(n)
```

```
## [1]  5 10
```

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

Answer 2: The original matrix (m) was 10c x 5r and the transposed matrix (n) is 5c x 10r

###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`.

```
m <- m[, c(1,2,4,5)]
m <- m[1:9,]
dim(m)
```

```
## [1] 9 4
```

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.

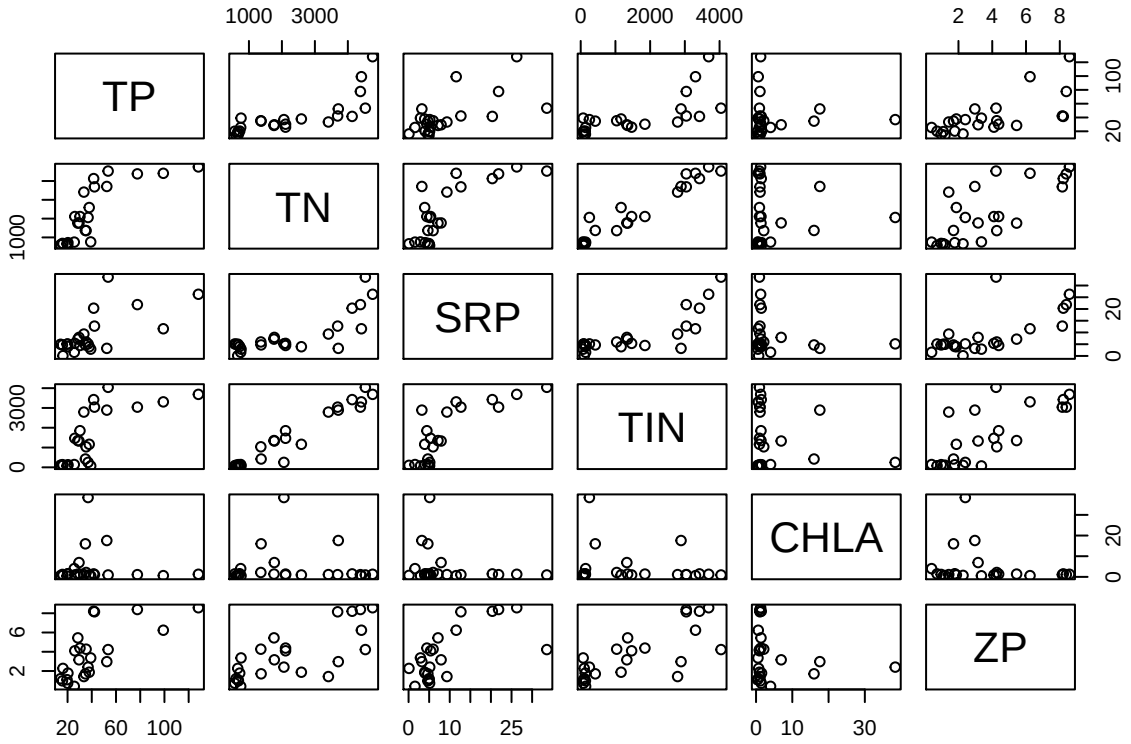
```
meso <- read.table("data/zoop_nuts.txt", sep = "\t", header = TRUE)
str(meso)
```

```
## '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 `meso` dataframe. 2) Visualize the pairwise **bi-plots** of the six numerical variables. 3) Conduct a simple **Pearson's correlation** analysis.

```
meso.matrix <- meso[,3:8]
pairs(meso.matrix)
```




```
cor1 <- cor(meso.matrix)
```

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

> Answer 3: There seems to be a positive correlation between total zooplankton biomass and total TIN, TP, TN, and SRP, also between total inorganic nutrient concentration and SRP, TN, and TP, and between SRP and TP and TN.

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.

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.0.4
```

```
cor2 <- corr.test(meso.matrix, method = "pearson", adjust = "BH")
print(cor2, digits = 3)
```

```
## Call:corr.test(x = meso.matrix, method = "pearson", adjust = "BH")
## Correlation matrix
##          TP     TN    SRP    TIN   CHLA     ZP
## TP    1.000  0.787  0.654  0.717 -0.017  0.697
## TN    0.787  1.000  0.784  0.969 -0.004  0.756
## SRP   0.654  0.784  1.000  0.801 -0.189  0.676
## TIN   0.717  0.969  0.801  1.000 -0.157  0.761
## CHLA -0.017 -0.004 -0.189 -0.157  1.000 -0.183
## ZP    0.697  0.756  0.676  0.761 -0.183  1.000
## Sample Size
## [1] 24
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##         TP    TN   SRP   TIN  CHLA    ZP
## TP   0.000 0.000 0.001 0.000 0.983 0.000
## TN   0.000 0.000 0.000 0.000 0.983 0.000
## SRP  0.001 0.000 0.000 0.000 0.491 0.000
## TIN  0.000 0.000 0.000 0.000 0.536 0.000
## CHLA 0.938 0.983 0.376 0.464 0.000 0.491
## ZP   0.000 0.000 0.000 0.000 0.393 0.000
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

```
cor3 <-corr.test(meso.matrix, method = "kendall", adjust = "BH")
print(cor3, digits = 3)
```

```
## Call:corr.test(x = meso.matrix, method = "kendall", adjust = "BH")
## Correlation matrix
##         TP    TN    SRP   TIN   CHLA     ZP
## TP   1.000 0.739  0.391 0.577  0.044  0.536
## TN   0.739 1.000  0.478 0.809  0.015  0.551
## SRP  0.391 0.478  1.000 0.563 -0.066  0.449
## TIN  0.577 0.809  0.563 1.000  0.044  0.548
## CHLA 0.044 0.015 -0.066 0.044  1.000 -0.051
## ZP   0.536 0.551  0.449 0.548 -0.051  1.000
## Sample Size
## [1] 24
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##         TP    TN   SRP   TIN  CHLA    ZP
## TP   0.000 0.000 0.088 0.014 0.899 0.015
## TN   0.000 0.000 0.034 0.000 0.946 0.014
## SRP  0.059 0.018 0.000 0.014 0.899 0.046
## TIN  0.003 0.000 0.004 0.000 0.899 0.014
## CHLA 0.839 0.946 0.760 0.839 0.000 0.899
## ZP   0.007 0.005 0.028 0.006 0.813 0.000
##
##  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: Yes, results seem to be sensitive to whether the methods are parametric or non-parametric. Non-parametric methods should be used whenever the researchers believe that the data may not be normally distributed. No, there are minorly differeces in p-values when adjusting for false discovery rate but not to the point of changing any of the results from significance. Factoring in these false discovery rates is important because our uncorrected p-values give us the probability that our null hypothesis is being falsly rejected but as more hypotheses are tested the chance of falsly rejecting a null increases and has to be accounted for in order to correctly evaluate results.

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.

```
fitreg <- lm(ZP ~ TN, data = meso)
summary(fitreg)
```

```
##
## Call:
## lm(formula = ZP ~ TN, data = meso)
##
## 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
```

```
library(ggplot2)
```

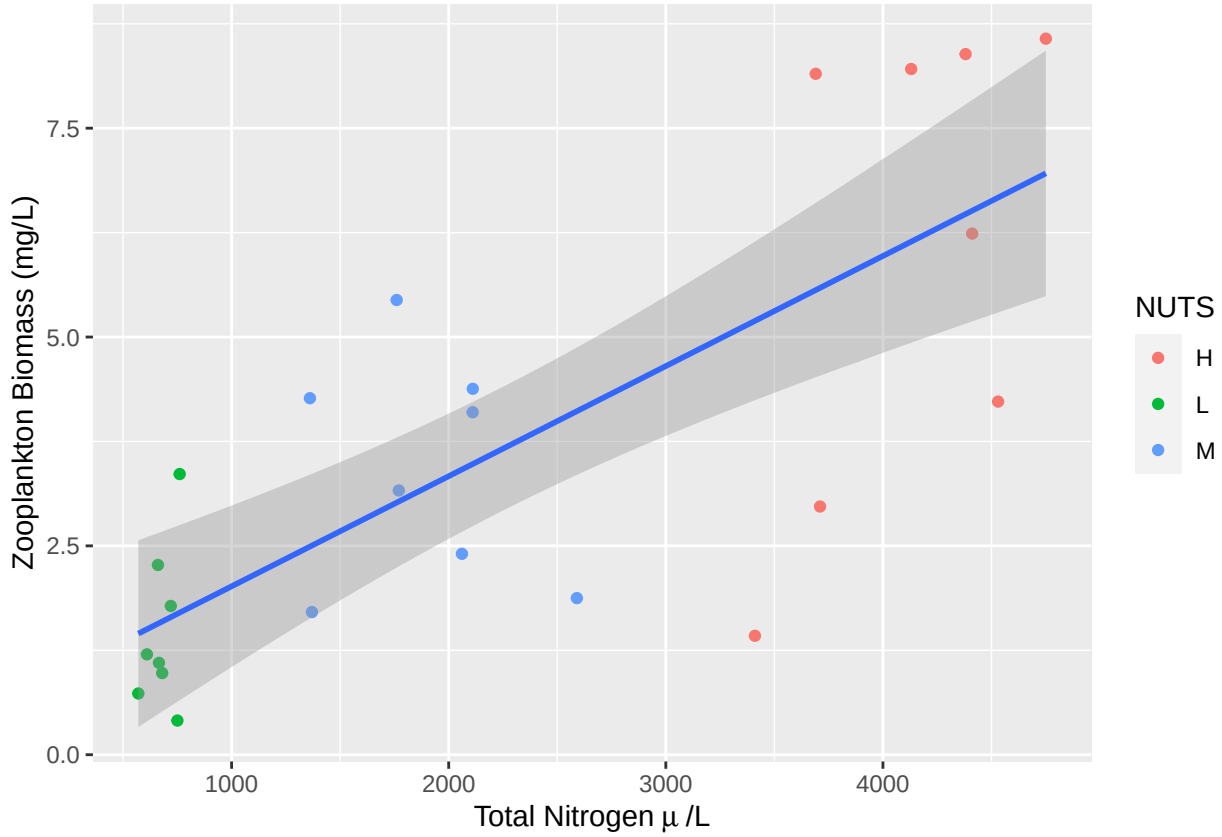
```
##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```
ZPTNplot <- ggplot(data = meso, mapping = aes(y=ZP, x=TN))
ZPTNplot <- ZPTNplot + geom_point(aes(color = NUTS))
ZPTNplot <- ZPTNplot + labs(y = "Zooplankton Biomass (mg/L)", x = expression("Total Nitrogen"~mu~"/L"))
ZPTNplot <- ZPTNplot + geom_smooth(method = "lm", se = TRUE)


print(ZPTNplot)
```

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



Question 5: Interpret the results from the regression model

> Answer 5: These results suggest that higher total nitrogren is positively correlated with higher total zooplankton biomass and that the difference between the medium and the low seems to be smaller than the difference between the medium and the high

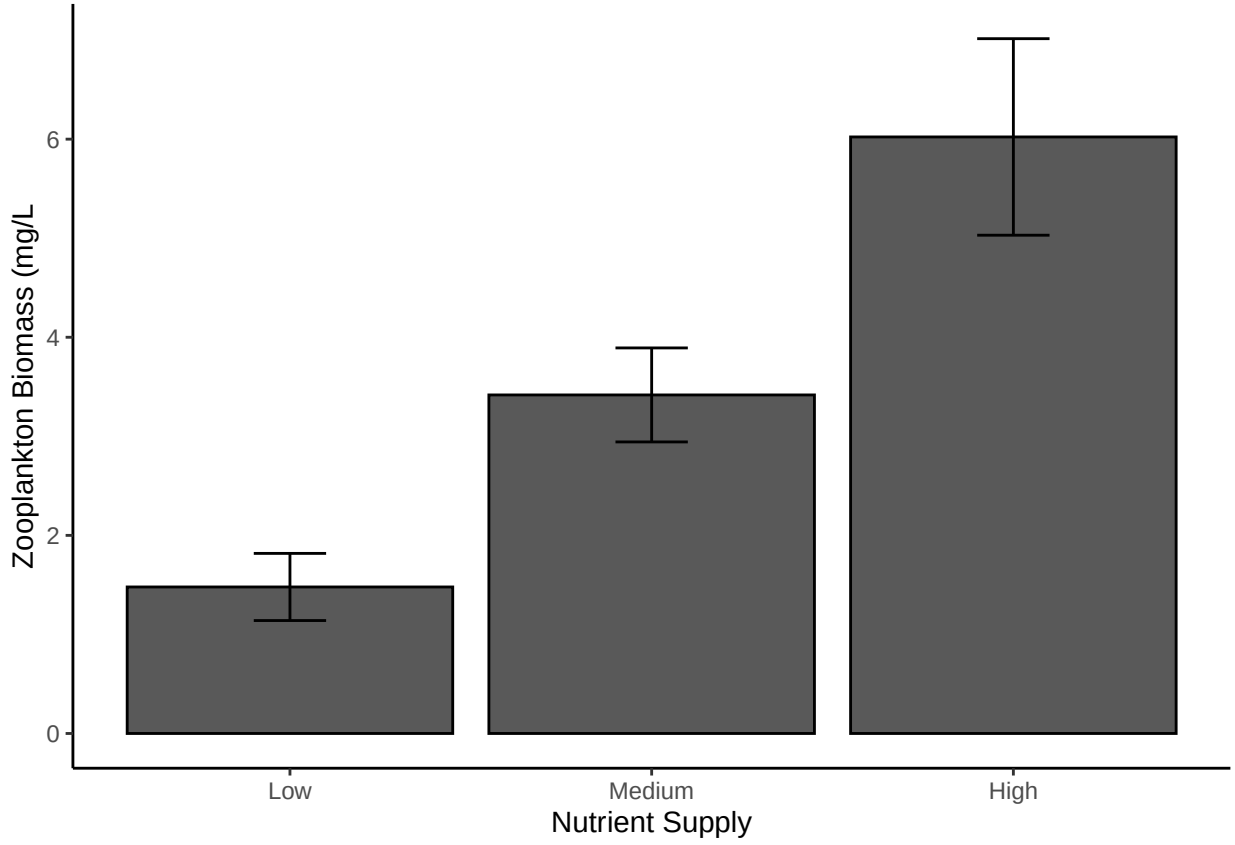
Analysis of Variance (ANOVA)

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.

```r
NUTS <- factor(meso$NUTS, levels = c('L','M','H'))

zp.means <- tapply(meso$ZP, NUTS, mean)
sem <- function(x){
  sd(na.omit(x))/sqrt(length(na.omit(x)))
}
zp.sem <- tapply(meso$ZP, NUTS, sem)
zp.means <- as.data.frame(zp.means)
zp.means <- data.frame(zp.means, nuts.labels = c("Low","Medium","High"))
zp.means <- data.frame(zp.means, std.error = zp.sem)
zp.means$nuts.labels <- factor(zp.means$nuts.labels, levels = c("Low", "Medium", "High"))

barplot <- ggplot(data = zp.means, aes(x = nuts.labels, y = zp.means))
barplot <- barplot + geom_bar(stat = "identity", color = "black", position = position_dodge())
barplot <- barplot + geom_errorbar(aes(ymin = zp.means-std.error, ymax = zp.means+std.error, width=.2))
barplot <- barplot + theme_classic()
barplot <- barplot + labs(y = "Zooplankton Biomass (mg/L", x = "Nutrient Supply")
print(barplot)
```




```r
anova <- aov(ZP ~ NUTS, data = meso)
summary(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.
- CYCL = 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.

```
# Importing dataframe
zoops <- read.csv("data/zoops.txt", sep = "\t", header = TRUE)

#Removing Tank and NUTS
zoops <- zoops[,(3:11)]

#Adding column for total biomass (ZP)
zoops$ZP <- rowSums(zoops)

#Correlation searching
corr <- corr.test(zoops, method = "pearson", adjust = "BH")
print(corr)
```

```
## Call:corr.test(x = zoops, method = "pearson", adjust = "BH")
## Correlation matrix
##        CAL  DIAP  CYCL  BOSM  SIMO  CERI  NAUP  DLUM  CHYD    ZP
## CAL   1.00  0.64  0.71  0.73 -0.27 -0.19  0.06 -0.03 -0.32 -0.31
## DIAP  0.64  1.00  0.69  0.38 -0.29 -0.17  0.22  0.64 -0.31 -0.30
## CYCL  0.71  0.69  1.00  0.75 -0.32 -0.13  0.19  0.13 -0.37 -0.36
## BOSM  0.73  0.38  0.75  1.00 -0.31 -0.14  0.18 -0.09 -0.21 -0.21
## SIMO -0.27 -0.29 -0.32 -0.31  1.00 -0.18 -0.24 -0.08  0.26  0.43
## CERI -0.19 -0.17 -0.13 -0.14 -0.18  1.00  0.47  0.02 -0.14 -0.14
## NAUP  0.06  0.22  0.19  0.18 -0.24  0.47  1.00  0.15 -0.24 -0.24
```

```
## DLUM -0.03  0.64  0.13 -0.09 -0.08  0.02  0.15  1.00 -0.22 -0.21
## CHYD -0.32 -0.31 -0.37 -0.21  0.26 -0.14 -0.24 -0.22  1.00  0.98
## ZP   -0.31 -0.30 -0.36 -0.21  0.43 -0.14 -0.24 -0.21  0.98  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   ZP
## CAL  0.00 0.01 0.00 0.00 0.45 0.55 0.83 0.90 0.38 0.38
## DIAP 0.00 0.00 0.00 0.30 0.41 0.56 0.52 0.01 0.38 0.39
## CYCL 0.00 0.00 0.00 0.00 0.38 0.62 0.55 0.63 0.31 0.33
## BOSM 0.00 0.07 0.00 0.00 0.38 0.62 0.55 0.76 0.52 0.52
## SIMO 0.20 0.17 0.12 0.14 0.00 0.55 0.50 0.77 0.46 0.18
## CERI 0.37 0.42 0.54 0.51 0.39 0.00 0.11 0.93 0.62 0.62
## NAUP 0.79 0.31 0.39 0.40 0.27 0.02 0.00 0.62 0.50 0.50
## DLUM 0.88 0.00 0.56 0.69 0.72 0.93 0.49 0.00 0.52 0.52
## CHYD 0.13 0.14 0.08 0.33 0.22 0.53 0.26 0.29 0.00 0.00
## ZP   0.15 0.16 0.09 0.31 0.04 0.51 0.25 0.33 0.00 0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

```
corr2 <- corr.test(zoops, method = "kendall", adjust = "BH")
print(corr2)
```

```
## Call:corr.test(x = zoops, method = "kendall", adjust = "BH")
## Correlation matrix
##        CAL  DIAP  CYCL  BOSM  SIMO  CERI  NAUP  DLUM  CHYD    ZP
## CAL   1.00  0.23  0.34  0.38 -0.24  0.02 -0.11  0.09 -0.39 -0.43
## DIAP  0.23  1.00  0.50  0.13 -0.42  0.04  0.22  0.31 -0.18 -0.18
## CYCL  0.34  0.50  1.00  0.41 -0.29  0.01  0.13  0.21 -0.31 -0.27
## BOSM  0.38  0.13  0.41  1.00 -0.40  0.08  0.04 -0.10 -0.22 -0.26
## SIMO -0.24 -0.42 -0.29 -0.40  1.00 -0.11 -0.18 -0.01  0.25  0.31
## CERI  0.02  0.04  0.01  0.08 -0.11  1.00  0.38  0.09 -0.09 -0.14
## NAUP -0.11  0.22  0.13  0.04 -0.18  0.38  1.00  0.16 -0.10 -0.08
## DLUM  0.09  0.31  0.21 -0.10 -0.01  0.09  0.16  1.00 -0.26 -0.21
## CHYD -0.39 -0.18 -0.31 -0.22  0.25 -0.09 -0.10 -0.26  1.00  0.91
## ZP   -0.43 -0.18 -0.27 -0.26  0.31 -0.14 -0.08 -0.21  0.91  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   ZP
## CAL  0.00 0.59 0.46 0.35 0.59 0.96 0.80 0.80 0.35 0.35
## DIAP 0.29 0.00 0.30 0.80 0.35 0.92 0.59 0.48 0.68 0.68
## CYCL 0.10 0.01 0.00 0.35 0.53 0.97 0.80 0.59 0.48 0.58
## BOSM 0.07 0.55 0.05 0.00 0.35 0.80 0.92 0.80 0.59 0.58
## SIMO 0.25 0.04 0.17 0.05 0.00 0.80 0.68 0.97 0.58 0.48
## CERI 0.92 0.84 0.97 0.70 0.60 0.00 0.35 0.80 0.80 0.80
## NAUP 0.62 0.29 0.54 0.86 0.40 0.07 0.00 0.73 0.80 0.80
## DLUM 0.67 0.14 0.31 0.63 0.95 0.68 0.45 0.00 0.58 0.59
## CHYD 0.06 0.41 0.14 0.29 0.23 0.66 0.65 0.21 0.00 0.00
## ZP   0.03 0.39 0.20 0.22 0.14 0.52 0.71 0.32 0.00 0.00
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

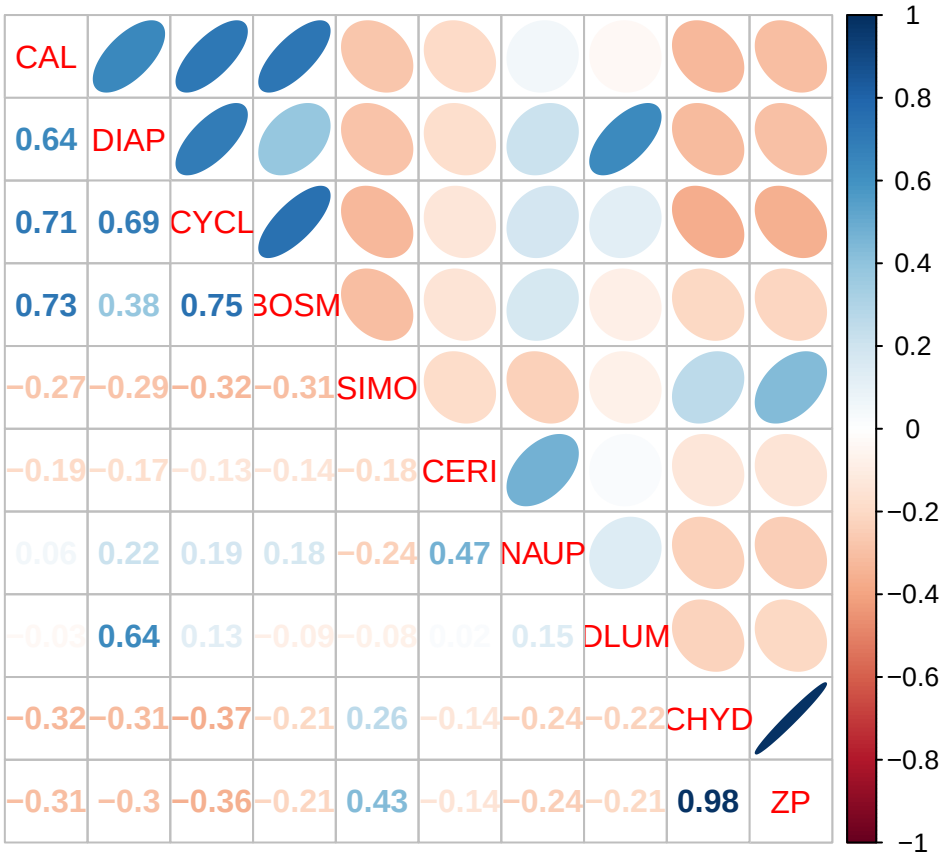
```
# Correlation visualization
corr3 <- cor(zoops)

library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.0.4
```

```
## corrplot 0.84 loaded
```

```
corrplot.mixed(corr = corr3, upper = "ellipse")
```




```
# Testing for significance taxa in determining total biomass
popanova <- aov(ZP ~ CAL + DIAP + CYCL + BOSM + SIMO + CERI + NAUP + DLUM + CHYD, data = zoops)
#popanova <- aov(ZP ~ CAL, data = zoops)
summary(popanova)
```

```
##              Df    Sum Sq  Mean Sq    F value Pr(>F)
## CAL           1 14780937 14780937 1.121e+31 <2e-16 ***
## DIAP          1  2770390  2770390 2.101e+30 <2e-16 ***
## CYCL          1  3650968  3650968 2.768e+30 <2e-16 ***
## BOSM          1  1838467  1838467 1.394e+30 <2e-16 ***
## SIMO          1 18735570 18735570 1.421e+31 <2e-16 ***
## CERI          1  2004138  2004138 1.520e+30 <2e-16 ***
## NAUP          1  2182195  2182195 1.655e+30 <2e-16 ***
```

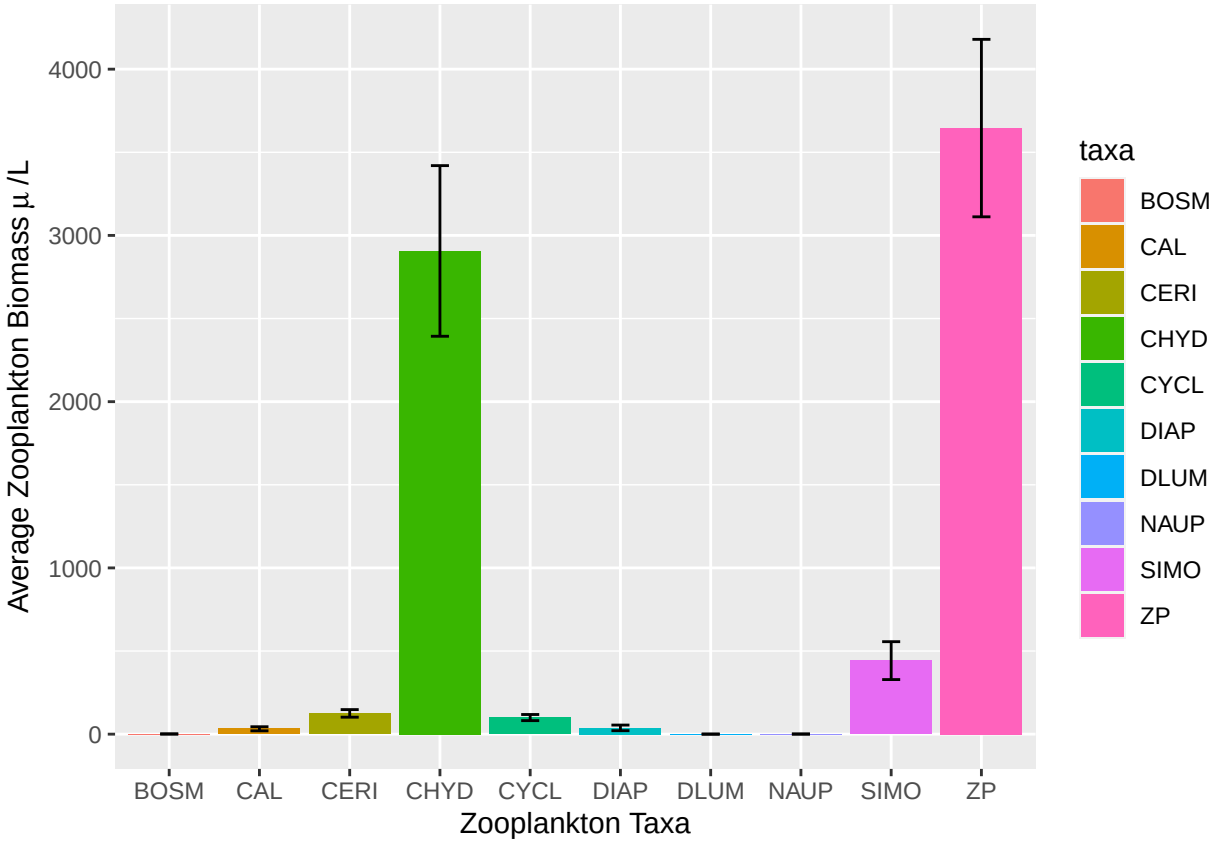
```
## DLUM         1 33871747 33871747 2.568e+31 <2e-16 ***
## CHYD         1 77434861 77434861 5.872e+31 <2e-16 ***
## Residuals   14        0        0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Producing a data frame with data for a box plot
mean <- function(x){
  sum(x)/length(x)
}
zoops.means <- sapply(zoops, mean)
zoops.error <- sapply(zoops, sem)

bardata <- as.data.frame(zoops.means)
bardata <- data.frame(bardata, error = zoops.error)
bardata$taxa <- row.names(bardata)
print(bardata)
```

```
##       zoops.means       error taxa
## CAL    32.2291667  12.3503557  CAL
## DIAP   37.8250000  16.8858006 DIAP
## CYCL   99.7666667  18.4860941 CYCL
## BOSM    1.1166667   0.5625597 BOSM
## SIMO  442.3000000 113.9905815 SIMO
## CERI  124.8583333  22.8856965 CERI
## NAUP    0.6208333   0.1706742 NAUP
## DLUM    0.2750000   0.2750000 DLUM
## CHYD 2906.6375000 513.1769982 CHYD
## ZP   3645.6291667 533.7678386   ZP
```

```
# Producing boxplot to visualize relative biomass
abundance.plot <- ggplot(data = bardata, aes(x = taxa, y = zoops.means, fill = taxa))
abundance.plot <- abundance.plot + geom_bar(stat = "identity")
abundance.plot <- abundance.plot + geom_errorbar(aes(ymin = zoops.means - error, ymax = zoops.means + e
adundance.plot <- abundance.plot + theme_classic()
abundance.plot <- abundance.plot + labs(y = expression("Average Zooplankton Biomass"~mu~"/L"), x = "Zoo
print(abundance.plot)
```



Answer: It seems as if the grand majority of the total zooplankton biomass comes from the Chydorus sp. but without the treatment information it isn't possible to draw conclusions on how different treatments affects this pattern. Additionally, I realized that since I calculated the total biomass using all of the taxa that they will obviously all be significantly correlated with it.

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 24th, 2021 at 12:00 PM (noon)**.