3. Worksheet: Basic R

Thomas Zambiasi; Z620: Quantitative Biodiversity, Indiana University

18 January, 2023

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, January 18th, 2023 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.

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

```
## [1] "C:/Users/tmzam/GitHub/QB2023_Zambiasi/2.Worksheets/3.RStudio"
```

```r
setwd("C:/Users/tmzam/GitHub/QB2023_Zambiasi/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.

```r
5^3
```

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

```r
pi*(2^2)
```

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

```r
sqrt(2)*(sin(pi/4))
```

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

```r
log(35)
```

```
## [1] 3.555348
```

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
x<-c(7,14,31,54,78)
w<-14*x
(x+w)/15
```

```
## [1]  7 14 31 54 78
```

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(13,24,45,48,64)
k*x
```

```
## [1]   91  336 1395 2592 4992
```

```r
d<-c(w[2:4],k[c(1,2,4,5)])
```

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)
max(na.omit(v))
```

```
## [1] 31.4
```

```r
min(na.omit(v))
```

```
## [1] 10.1
```

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

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

```r
mean(na.omit(v))
```

```
## [1] 20.9
```

```r
median(na.omit(v))
```

```
## [1] 20.35
```

```r
var(na.omit(v))
```

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

```
sd(na.omit(v))
```

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

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

```
## [1] 0.4485805
```

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.

```
col1<-c(rnorm(5,mean=8,sd=2))
col2<-c(rnorm(5,mean=25,sd=10))
mtx1<-matrix(c(col1,col2),nrow=5,ncol=2,byrow=FALSE)
mtx1
```

```
##            [,1]     [,2]
## [1,] 10.724773 21.12599
## [2,]  6.387244 10.47212
## [3,]  6.918919 32.30475
## [4,]  5.965773 14.36597
## [5,]  7.410633 26.06613
```

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: The rnorm function creates a normal distribution of random values that fit parameters set through arguments in the function (ex: # of values, mean of distribution, standard deviation of distribution, etc.)

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

```
##    [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## V1    8    5    2    3    9   11    2    3    5    6
## V2    1    5    5    2    9    8    2    3    5    5
## V3    7    2    4    5    1    1    5    6    1    9
## V4    6    4    3    1    1    8    8    7    3    2
## V5    1    1    3    4    2    8    5    6    6    2
```

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

> Answer 2: This matrix has 5 rows and 10 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`.

```r
index<-n[1:4,c(1,2,4:10)]
index
```

```
##    [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
## V1    8    5    3    9   11    2    3    5    6
## V2    1    5    2    9    8    2    3    5    5
## V3    7    2    5    1    1    5    6    1    9
## V4    6    4    1    1    8    8    7    3    2
```

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.

```r
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.

```r
meso.numeric<-meso[,3:8]
pairs(meso.numeric)
```
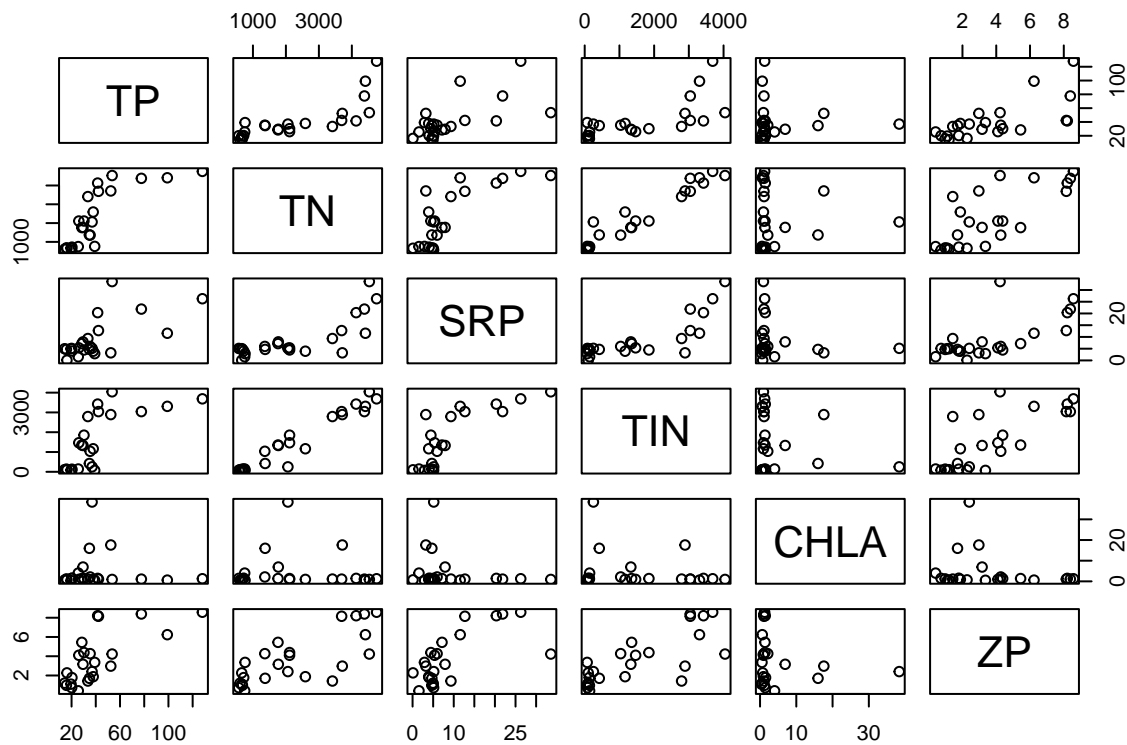
```
cor1<-cor(meso.numeric)
cor1
```

```
##                  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: From looking at the visualization and the correlation analysis; one of the first features that stood out to me was that concentration of chlorophyll a was not strongly correlated with any nutrient concentrations or zooplankton biomass (furthest of these values from 0 was -0.189). It also appears that zooplankton biomass generally had a positive correlation with nutrient concentrations across the board (with the exception of chlorophyll at -0.8125; others fell between 0.676 - 0.761). All nutrient concentrations generally showed positive correlations with other nutrients and zooplankton biomass, and weak negative correlations with chlorophyll concentration (values range from 0.654 - 0.969).

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.

```
require("psych")
```

```
## Loading required package: psych
```

```
## Warning: package 'psych' was built under R version 4.2.2
```

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

```
## Call:corr.test(x = meso.numeric, 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
```

```
print(cor2.np,digits=3)
```

```
## Call:corr.test(x = meso.numeric, method = "spearman", adjust = "BH")
## Correlation matrix
##         TP    TN    SRP   TIN   CHLA     ZP
## TP   1.000 0.895  0.539 0.761  0.040  0.741
## TN   0.895 1.000  0.647 0.942  0.021  0.748
## SRP  0.539 0.647  1.000 0.726 -0.064  0.627
## TIN  0.761 0.942  0.726 1.000  0.088  0.738
## CHLA 0.040 0.021 -0.064 0.088  1.000 -0.072
## ZP   0.741 0.748  0.627 0.738 -0.072  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 0.010 0.000 0.914 0.000
## TN   0.000 0.000 0.001 0.000 0.923 0.000
```

```
## SRP  0.007 0.001 0.000 0.000 0.884 0.002
## TIN  0.000 0.000 0.000 0.000 0.884 0.000
## CHLA 0.853 0.923 0.767 0.683 0.000 0.884
## ZP   0.000 0.000 0.001 0.000 0.737 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: The results do change whether the correlation tests are conducted using parametric or nonparametric methods. The digit limit does not allow for direct comparison on many of the corrected probability values, but these probability values tend to be larger when the correlation test is run using nonparametric methods. Non-parametric methods should be used instead of parametric methods when sets of data are being compared to multiple other datasets (like in this case). This should decrease the likelihood of a correlation being labeled as significant when it really isn't. The Pearson's method of the correlation test does not show any evidence for false discovery rate here; the p-values do not change following the correction. False discovery rate is important because it is a measure of Type-1 error, which is when multiple analyses result in a higher likelihood of returning a significant result regardless of whether it is actually significant. Correcting for this false discovery rate is important for determining whether the results of statistical tests are actually significant.

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.

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

```
##
## 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
```

```
plot(meso$TN,meso$ZP,ylim=c(0,10),xlim=c(500,5000),xlab=expression(paste("Total Nigtrogen (", mu,"g/L)")
text(meso$TN,meso$ZP,meso$NUTS,pos=3,cex=0.8)
newTN<-seq(min(meso$TN),max(meso$TN),10)
regline<-predict(ZPTNreg,newdata=data.frame(TN=newTN))
lines(newTN,regline)
conf95<-predict(ZPTNreg,newdata=data.frame(TN=newTN),interval=c("confidence"),level=0.95,type="response
matlines(newTN,conf95[,c("lwr","upr")],type="l",lty=2,lwd=1,col="black")
```
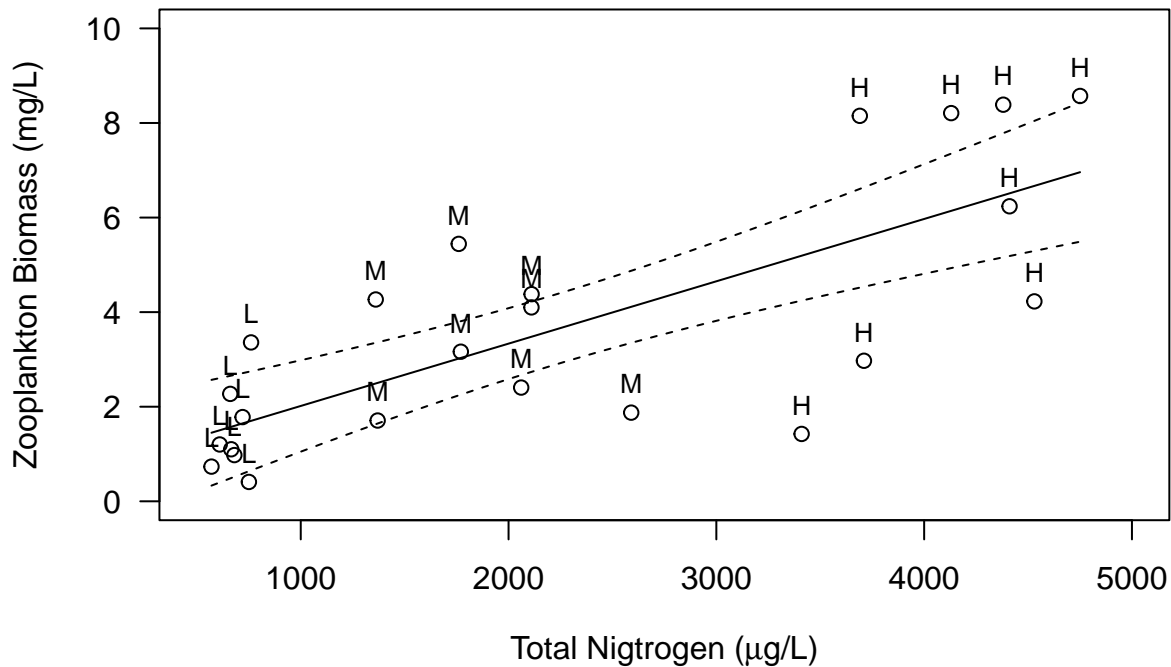

Question 5: Interpret the results from the regression model

Answer 5: The regression model shows that treatments of higher total nitrogen (ug/L) resulted in greater zooplankton biomass (mg/L) (R^2=0.5719, F(1,22)=29.39, p=1.911e-05). The positive correlation shown in the model indicates that increases of 1 ug/L of nitrogen correspond to a 0.0013 mg/L increase in zooplankton biomass. These results can also be seen clearly in the graph, with high (H) nutrient amounts generally resulting in the highest zooplankton biomasses, followed by medium (M) nutrient amounts and low (L) nutrients.

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.

```
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$ZP,NUTS,sem)
bp<-barplot(zp.means,ylim=c(0,round(max(meso$ZP),digits=0)),pch=15,cex=1.25,las=1,cex.lab=1.4,cex.axis=
arrows(x0=bp,y0=zp.means,y1=zp.means-zp.sem,angle=90,length=0.1,lwd=1)
arrows(x0=bp,y0=zp.means,y1=zp.means+zp.sem,angle=90,length=0.1,lwd=1)
```
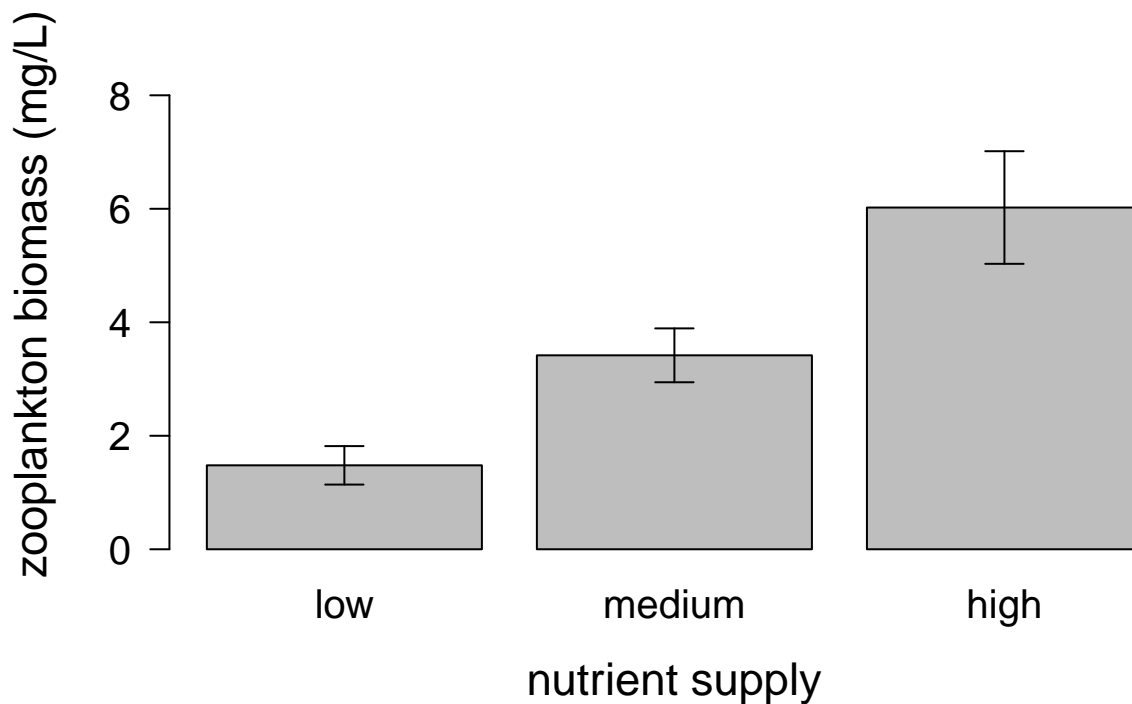


```
fitanova<-aov(ZP~NUTS,data=meso)
summary(fitanova)
```

```
##             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
```

```
TukeyHSD(fitanova)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
```

```
## Fit: aov(formula = ZP ~ NUTS, data = meso)
##
## $NUTS
##           diff        lwr        upr      p adj
## L-H -4.543175 -6.9115094 -2.1748406 0.0002512
## M-H -2.604550 -4.9728844 -0.2362156 0.0294932
## M-L  1.938625 -0.4297094  4.3069594 0.1220246
```

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.

```r
zoops<-read.table("data/zoops.txt",sep="\t",header=TRUE)
#removing TANK and NUTS columns and turning into matrix
zoops2<-as.matrix(zoops[,3:11])
#Need to compare species biomass to determine which were most involved in the greater biomass response
#will create biplot to assess correlation of individual species biomass with total zooplankton biomass
#first, isolating zooplankton biomass as its own vector
zp<-c(meso.numeric$ZP)
#need to convert to ug/L from mg/L
zp.ug<-zp*(10^3)
#now, adding total zooplankton biomass (zp.ug) to siteXspp matrix
zoops3<-matrix(c(zoops2,zp.ug),ncol=10,byrow=FALSE)
#fixing column names
colnames(zoops3)<-c("CAL","DIAP","CYCL","BOSM","SIMO","CERI","NAUP","DLUM","CHYD","ZP")
#creating biplot comparing individual sp. biomass to all zooplankton biomass
pairs(zoops3)
```
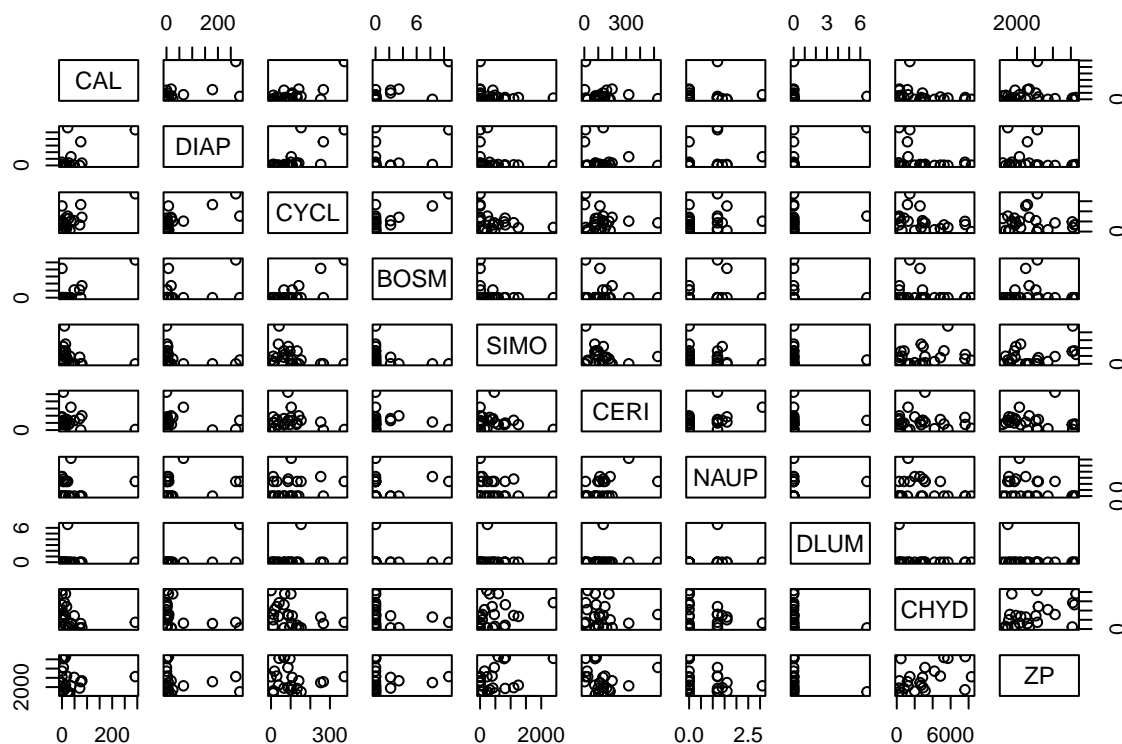
```
#Now will use a correlation test to look for significant relationships between individual species bioma
require("psych")
cor.zoops<-corr.test(zoops3,method="pearson",adjust="BH")
print(cor.zoops,digits=3)
```

```
## Call:corr.test(x = zoops3, method = "pearson", adjust = "BH")
## Correlation matrix
##          CAL   DIAP   CYCL   BOSM   SIMO   CERI   NAUP   DLUM   CHYD     ZP
## CAL    1.000  0.643  0.712  0.728 -0.271 -0.191  0.058 -0.034 -0.322 -0.048
## DIAP   0.643  1.000  0.694  0.381 -0.287 -0.172  0.217  0.637 -0.314 -0.175
## CYCL   0.712  0.694  1.000  0.747 -0.325 -0.132  0.186  0.125 -0.369 -0.066
## BOSM   0.728  0.381  0.747  1.000 -0.308 -0.141  0.179 -0.086 -0.206 -0.017
## SIMO  -0.271 -0.287 -0.325 -0.308  1.000 -0.183 -0.237 -0.077  0.262  0.426
## CERI  -0.191 -0.172 -0.132 -0.141 -0.183  1.000  0.475  0.020 -0.135 -0.096
## NAUP   0.058  0.217  0.186  0.179 -0.237  0.475  1.000  0.148 -0.238 -0.309
## DLUM  -0.034  0.637  0.125 -0.086 -0.077  0.020  0.148  1.000 -0.224 -0.217
## CHYD  -0.322 -0.314 -0.369 -0.206  0.262 -0.135 -0.238 -0.224  1.000  0.463
## ZP    -0.048 -0.175 -0.066 -0.017  0.426 -0.096 -0.309 -0.217  0.463  1.000
## 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    0.000 0.006 0.001 0.001 0.499 0.611 0.866 0.917 0.401 0.884
## DIAP   0.001 0.000 0.002 0.298 0.462 0.611 0.579 0.006 0.401 0.611
## CYCL   0.000 0.000 0.000 0.001 0.401 0.692 0.611 0.700 0.313 0.855
## BOSM   0.000 0.066 0.000 0.000 0.401 0.692 0.611 0.815 0.601 0.936
```

```
## SIMO 0.199 0.175 0.122 0.143 0.000 0.611 0.568 0.833 0.510 0.189
## CERI 0.371 0.421 0.538 0.510 0.393 0.000 0.123 0.936 0.692 0.797
## NAUP 0.789 0.309 0.385 0.403 0.265 0.019 0.000 0.691 0.568 0.401
## DLUM 0.876 0.001 0.560 0.688 0.722 0.925 0.491 0.000 0.579 0.579
## CHYD 0.125 0.136 0.076 0.334 0.216 0.528 0.263 0.293 0.000 0.129
## ZP   0.825 0.413 0.760 0.936 0.038 0.655 0.142 0.309 0.023 0.000
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

Answer 6: Based on the biplot and the correlation analysis, *Chydorus* sp. and *Simocephallus* sp. are the only species that had a moderately strong correlation with overall zooplankton biomass (in µg/L) (respective correlation coefficients: 0.463, 0.426). These results seem to show that *Chydorus* sp. and *Simocephallus* sp. may have had a greater influence on total zooplankton biomass following nutrient enrichment than the other species that were studied. However, these correlations were not quite significant following Benjamini/Hochberg correction (*Chydorus* sp. p=0.129, *Simocephallus* sp. p=0.189). Because of the relative strength of the correlations (and proximity to the significance threshold), I would advocate for collecting more biomass data across these taxa and would predict that a larger sample size may result in stronger, significant correlations of *Chydorus* sp. and *Simocephallus* sp. biomass with overall zooplankton biomass.

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 18th, 2021 at 12:00 PM (noon)**.