3. Worksheet: Basic R

Maddy Spencer; 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.

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

```
## [1] "/cloud/project/QB2025_Spencer/Week1-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.

```
5^3
```

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

```
pi *2^2
```

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

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

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

```
log(25)
```

```
## [1] 3.218876
```

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(1, 2, 3, 4, 5)
w <- x * 14
(x + w) / 15
```

```
## [1] 1 2 3 4 5
```

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`.

```
x <- c(1, 2, 3, 4, 5)
w <- x * 14
k <- c(2, 4, 6, 8, 10)
y <- k * x
```

```
d <- c(w[1:3], k[1:4])
print(d)
```

[1] 14 28 42 2 4 6 8

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.

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

[1] 31.4

```
min(v)
```

[1] 10.1

```
sum(v)
```

[1] 292.6

```
mean(v)
```

[1] 20.9

```
median(v)
```

[1] 20.35

```
var(v)
```

[1] 39.44

```
sd(v)
```

[1] 6.280127

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.

```
a <- c(rnorm(5, mean = 8, sd = 2))
b <- c(rnorm(5, mean = 25, sd = 10))
c <- matrix(c(a, b), nrow = 5, ncol = 2, byrow = FALSE)
print(c)
```

```
##            [,1]     [,2]
## [1,]   4.784094 30.69108
## [2,]   5.656038 22.28206
## [3,]  12.072286 26.41306
## [4,]   8.672160 28.67487
## [5,]   7.808388 39.45347
```

```
help(rnorm)
```

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 generates data that follows a normal distribution. The arguments in this specifies the number of elements/numbers generated (n), the mean of the data generated (mean), and the standard deviation of the numbers generated (sd).

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

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

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

Answer 2: The matrix is 5 columns and 10 rows.

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

```
## [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.

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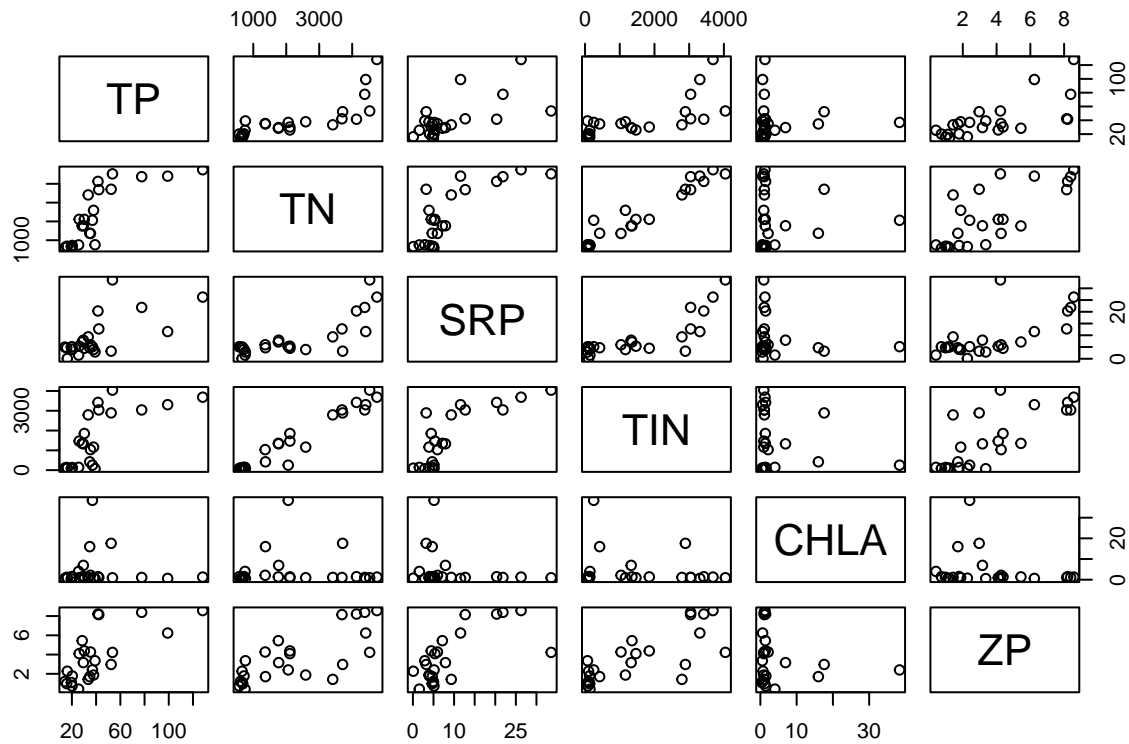
```
cor1 <- cor(meso.num)
print(cor1)
```

```
##                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
```