**Microbial Diversity and Ecology Fall 2020**

**Lab 2, Introduction to R - assignment**

**Due**: Friday September 4, 11:59 pm

**Task**: upload this document in the assignment on WyoCourse; week 2 lab 2 Intro to R, with your answers embedded and rename the file by adding your last name at the end of the file name, e.g. Week2\_Lab2\_Intro\_to\_R\_Assignment\_vanDiepen. You can save your file as a Word document or convert to a pdf.

**Points**: 4

**Objective**: In this exercise you will work through a tutorial in R. The goal of this exercise is to familiarize you with the basics of R and R Studio as well as introductory data manipulation and visualization.

**Instructions:** From WyoCourse you downloaded a folder named “Introduction\_to\_R”. Move this folder to your course folder you created when you installed RStudio in last week’s lab prep, and rename this folder by adding your last name (i.e. Introduction\_to\_R\_YOURLASTNAME). You will need to save all your files in this folder for this tutorial. Saving your work will allow you to revisit your code when you need it in the future.

Open the “Introduction\_to\_R” folder, and double click “Introduction\_to\_R\_Fall2020\_final.Rmd”. This will open the program R studio and the tutorial.

The tutorial contains 19 exercises. You can begin reading in the tutorial and refer back to this sheet when you are ready to answer a question or complete one of the exercises.

Open the R script and start reading.

**#Opening a script and saving**

Exercise 1 (~Line 24): 0.25 pt

Question 1: What do you notice about the initial appearances of the R script vs. the R notebook?

Answer:

Figure is embedded. The script you write directly in and the notebook you have to write in chunks.

Question 2: What is different between the two file types with respect to where the figure appears?

Answer:

Figure is embedded in the notebook.

**#Installing a package**

Exercise 2 (~Line 53) 0.1 pt

Question 1: What can you do with this package?

Answer:

“Do more with date and times”

**#R as a calculator**

Exercise 3: (~Line 79) 0.1 pt

Question 1: Using R find the answer to "(9\*7)/15 + 6". Assign that to a value called "answer1" and multiple it by 15. What is your final product?

Answer: 10.2 and 153

Copy the code you used to do this from R into this Word Document here:

((9\*7)/15 + 6)\*15

**#Familiarizing yourself with your working environment.**

Exercise 4 (~Line 105) 0.25 pt

Code used to change your working directory to your course (e.g. MOLB4540) folder on your computer and display what your current working directory is:

setwd(“/…./…”)

**#Loading data**

Exercise 5 (~Line 128) 0.2 pt

Question 1: What is the difference between the 2 codes and what does it do?

Answer: The assignment operator. <- saves whatever is to the left to the words you type on the right.

Question 2: What happened when you forget to include the "" around the file name?

Answer: It won’t find it.

Exercise 6 (~Line 135) 0.1 pt

Error message from code in lines 136-138 and your code if you were able to get the .csv read in successfully.

Answer: Error in file(file, "rt") : cannot open the connection

Correct code to read in the Pima.csv file using the entire pathway to your folder in your Document/course folder: “/Users/….”

Exercise 7 (~Line 136): 0.1 pt

Description of "read.csv" in help window:

Answer: Reads a file in table format and creates a data frame from it, with cases corresponding to lines and variables to fields in the file.

**#Familiarizing yourself with different data types and structures**

Exercise 8 (~Line 167): 0.2 pt

Question 1: What are some other types of data in R according to your Google search?

Answer: dataframe, matrix, data.table etc.

Exercise 9 (~Line 171): 0.1 pt

Question 1: What does the structure looks like when it is a matrix?

Answer:

num [1:768, 1:10] 1 2 3 4 5 6 7 8 9 10 ...

- attr(\*, "dimnames")=List of 2

..$ : NULL

..$ : chr [1:10] "X" "pregnant" "glucose" "diastolic" ...

**#Creating your own data frame**

Exercise 10 (~Line 215): 0.25 pt

Create your own data frame. Copy code below.

Answer:

There are many ways to do this.

combined4<- data.frame(numbers = c(2, 3, 5, 7, 9), alphabet = c("aa","bb","cc","dd", "ee"))

**#Working with data**

Exercise 11 (~Line 231): 0.25 pt

Code to create a dataset with a subset of columns (columns 1-5):

Answer: subsetPima<-Pima[,1:5]

First line of the resulting display of the data structure using the 'str' command:

'data.frame': 768 obs. of 5 variables:

Exercise 12 (~Line 249): 0.1 pt

Question 1: What does the code “subsetPima<-Pima[,colnames(Pima) %in% columns\_to\_keep]” do? Describe in your own words.

Answer: It keeps only those columns in Pima that match the names in the vector “columns to keep”

Exercise 13 (~Line 251): 0.3 pt

Code that will pull only the first two rows of the columns "age" and "triceps" from the Pima data set and name it subsetPima3:

Answer: couple of different ways to go about this but

subsetpima3<-pima[1:2, c("age", "triceps")]

or you could make a vector like the previous example and have “age” and “triceps” stored there.

subsetPima3<-Pima[1:2,colnames(Pima) %in% columns\_to\_keep\_for\_this\_examples]

Exercise 14 (~Line 262): 0.25 pt

Calculate the mean and standard deviation individually for each of the three columns of "subsetPima3".

Answer:

mean(subsetPima3$column\_name)

sd(subsetPima3$column\_name)

Exercise 15 (~Line 269): 0.1 pt

Code you used to load the data “metadata” into R:

Answer: metadata <-read.csv(“/…/…/”)

Exercise 16 (~Line 296): 0.25 pt

Code you used to create a new column "Infestation\_Stage", and add values to this column for each sample based on the sample ID:

Answer: metadata$Infestation\_Stage <- substr(metadata$Sample\_ID, start = 1, stop = 1)

index <- c("H", "D")

values <- c("Healthy", "Dead")

metadata$Infestation\_Stage <- values[match(metadata$Infestation\_Stage, index)]

Exercise 17 (~Line 306): 0.1 pt

Code you used to save the metadata file as a new file in your folder on the Desktop:

Answer:

Write.csv(metadata\_final, “/User../…../”)

Exercise 18 (~Line 314): Empty and reload of your environment, and add the name of your environment file here: 0.1 pt

Save using the save icon in the environment panel, clean out using the broom and load using the load icon in the same panel.

**#Final exercise: 1.0 pt**

Paste the code you used to make the graph as well as the graph itself here:

soildat<-read.csv("/Users/gordoncuster/Desktop/MicroDiv/Tutorials/Introduction\_to\_R/Soil\_4120.csv")

plot(soildat$soil.pH, soildat$depth, xlab = "pH", ylab= "depth", main = "pH by soil depth")

A close up of a map

Description automatically generated