**Microbial Diversity and Ecology Lab, Fall 2018**

**Lab 10: Intro to R**

In this lab you will go through a tutorial in R. The goal of this exercise is to familiarize you with the basics of R and R Studio.

On the Desktop you will see a folder named “ECOL5540”. Rename this folder by adding your last name (i.e. ECOL5540\_Custer). You will need to save all your files in this folder for the remainder of the semester.

Open the ECOL5540 folder, and double click “Intro to R.Rmd”. This will open the program R studio and the tutorial.

The tutorial contains 16 exercises.

Open the R script

**#Opening a script and saving**

Exercise 1 (Line 19):

Question 1: What do you notice about initial appearances?

Answer:

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

Answer:

**#Installing a package**

Exercise 2 (Line 40)

Question 1: What can you do with this package?

Answer:

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

Exercise 3 (Line 69)

Code used to change your working directory to your ECOL5540 folder on your Desktop and display what your current working directory is:

Exercise 4 (Line 87)

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

Answer:

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

Answer:

Question 3: What happens when your working directory is changed back to the "Downloads" file?

Answer:

**#Loading data**

Exercise 5 (Line 97)

Error message from code in lines 98-100:

Correct code to read in the Pima.csv file using the entire pathway to your folder on your Desktop:

Exercise 6 (line 110):

Description of "read.csv" in help window:

Exercise 7 (line 128):

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

Answer:

Exercise 8 (line 131):

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

Answer:

**#Working with data**

Exercise 9 (line 155):

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

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

Exercise 10 (line 172):

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

Answer:

Exercise 11 (line 174):

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

Exercise 12 (line 198):

Code that will filter the Pima dataset to only include the individuals that are over 25, and then get the average B.M.I. for both the positive and negative test groups:

Fill out the values in the table

|  |  |
| --- | --- |
| Test | Mean Glucose |
| 0 |  |
| 1 |  |

Exercise 13 (line 210):

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

Exercise 14 (line 234):

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

Exercise 15 (line 254):

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

Exercise 16 (line 263): Practice empty and reload of your datasets.