**Introduction to R worksheet and exercises**

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.

From WyoCloud you downloaded a folder named “Introduction\_to\_R”. Move this folder to your Desktop 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.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 21):

Question 1: What do you notice about initial appearances?

Answer: \*Looking for something about grey chunks etc. or how the script only doesn’t contain the header etc.

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

Answer: \* here students should recognize where the figure appears. In the script, the figure is plotted in the plots tab. In the notebook, the figure appears below.

**#Installing a package**

Exercise 2 (~Line 44)

Question 1: What can you do with this package?

Answer: \*manipulate dates or something of the sort

**#R as a calculator**

Exercise 3: (~Line 72)

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:

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

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

Exercise 4 (~Line 98)

Code used to change your working directory to your Life2100 folder on your Desktop and display what your current working directory is:

Exercise 5 (~Line 115)

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

Answer: \*“<-” allows you to save things in your environment

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

Answer: \*it doesn’t work

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

Answer: \*it doesn’t work because the pima.csv file doesn’t exist in that location.

**#Loading data**

Exercise 6 (~Line 127)

Error message from code in lines 124-126 and your code if you were able to get the .csv read in successfully.

Answer:\*doesn’t exist

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

Exercise 7 (~Line 136):

Description of "read.csv" in help window:

Answer: \*paste in description

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

Exercise 8 (~Line 155):

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

Answer: \*matrix, etc

Exercise 9 (~Line 159):

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

Answer: \*only four lines show up and it’s not a line for each variable anymore

**#Creating your own data frame**

Exercise 10 (~Line 200):

Create your own data frame. Copy code below.

Answer:

**#Working with data**

Exercise 11 (~Line 212):

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

Answer: \*however students find to subset code

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

Exercise 12 (~Line 230):

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

Answer: \*subset the data to only include the columns found in the vector.

Exercise 13 (~Line 230):

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

Answer: \* code used to subset

Exercise 14 (~Line 240):

Calculate the mean and standard deviation individually for each of the three columns of "subset\_Pima".

Answer:

Exercise 15 (~Line 250):

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

Answer:

Exercise 16 (~Line 270):

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 17 (~Line 280):

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

Exercise 18 (~Line 290): Empty and reload of your environment.

**#Final exercise:**

Answer: