**­­Microbial Diversity and Ecology Lab, Fall 2020**

**Lab 4: Introduction to data analysis in R**

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

**Task**: upload this document in the assignment on WyoCourse; Week 3 lab 4 Intro Data Analysis in R, with your answers embedded and rename the file by adding your last name at the end of the file name, e.g. Week3\_Lab4\_IntroDataAnalysis\_Assignment\_vanDiepen. You can save your file as a Word document or convert to a pdf.

**Points**: 2

**Objective**: In this tutorial you will be become familiar with basic data analysis in R and the countless options R has to offer. In this tutorial you will use a data set included in base R and soil data from the 2019 sampling from the Happy Grass grazing experimental plots.

**Instructions:** From WyoCourse you downloaded a folder named “IntroToDataAnalysis”. Move this folder to your course folder you created when you installed RStudio in week 1, and rename this folder by adding your last name (i.e. IntroToDataAnalysis). 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 “IntroToDataAnalysis” folder, and double click “Intro\_to\_DataAnalysis\_Fall2020\_final.Rmd”. This will open the program R studio and the tutorial.

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

**Exercise 1** (~ Lines 10-15): Install and load the necessary packages for today's lab. Record the code you used here:

**Exercise 2** (~ Lines 22-29): Rename the mtcars data set as "cars\_data". Record the code you used here:

**Exercise 3** (~ Lines 33-37): Please read in the Happy Grass metadata file ("HappyGrassMetadata.csv") that you downloaded from WyoCourses. Save it as HG\_Data. You will use this for the on your own exercises which applies your new R data analysis knowledge from working through the example dataset *mtcars*. Record the code you used here:

HG\_Data<-read\_csv("HappyGrassMetadata.csv")

**Exercise 4** (~Lines 42-49): Convert the Number\_of\_cylinders, Engine\_v\_or\_inline, and Transmission\_M\_or\_A columns to factors. The first one has been done for you. Take a moment to think about the difference between a categorical vs. continuous variables? Please record the code you used to accomplish this conversion here:

Then give a brief explanation of the difference between categorical and continuous variables here:

**Hypothesis Testing**

**Exercise 5** (~ Lines 69-75): Please provide a brief explanation of each assumption of ANOVA in your handout.

**Exercise 6** (~ Lines 148 - 162): On your own, you are to compare 2 of the variables from the Happy Grass data file you read in for Exercise 3 above among the grazing treatment groups using a test of your choice and pairwise comparison. You will use the Happy Grass metadata as your data sheet. One of the variables to compare among Treatments should be the one that you selected in the "Week 2 Soil sampling/analysis lab, Question 2", which you expected to be most affected by grazing.

Provide the code and brief write up on your thought process as to why you chose the specific test and pairwise comparison that you did:

In addition, please provide a brief write up that indicates what your statistical results mean in the context of the Happy Grass experiment.

**Correlation Testing**

**Exercise 7** (~ Lines 183-193): Examine the correlation between **two different sets of variables** (e.g. set 1: variable 1 and variable 2, set 2: variable 5 and variable 3) from the Happy Grass data set. Record the pair examined, the direction and strength of correlation, which test you used and why you chose this specific correlation coefficient.