**OEB 137 – Assignment 2**

**Due September 23, 2020 at 11:59 am (Eastern)**

**Introduction to Coding in R**

The following tasks are meant to establish your understanding of basic operations in R. In some cases, completing the task will require several steps of coding not explicitly stated in the task prompt. This is meant to prepare you for real use of R software in which you spend time troubleshooting to find a workable solution to the problem at hand. Please save your code as an R script and upload it to the Canvas site by the beginning of Lab on Sept. 23. Also, **\*\*\*annotate your code as much as possible\*\*\*** describing what you are doing with each step. This is best practice for your actual research (it has become commonplace to publish code as a supplement to journal articles) and will help us provide partial credit for any places where your code fails to work.

The dataset provided describes a fictitious census of barnacles in the Narragansett Bay. This census consists of 5 sites, and at each site two individual 0.25 m2 plots were sampled. In each plot, each barnacle was recorded with an individual number and the diameter was measured in mm. Site-level water chemistry measurements were also taken including salinity (parts per thousand – ppt), and dissolved oxygen concentration (mg/L).

1. Set your working directory but annotate this line out (place a hashtag at the beginning) so that it does not try to run on the grader’s computer. Load in the supplied data frame on barnacles in the Narragansett Bay and assign it to an object. Print the header of your new object. **(1 pt)**
2. Calculate the total number of barnacles found at each site, the arithmetic mean and standard deviation for barnacle diameter for each site. **(1 pt)**
3. Write your own function, called “gmean” that calculates the geometric mean, and use this function to calculate the geometric mean for barnacle diameter for each site. **(2 pt)**
4. Create a new column that contains a unique identifier for each plot at each site. Based on eyeballing (not statistics), which site has the most variation between plots in number of barnacles per .25 m2? **(1 pt)**
5. Plot a histogram of barnacle size. Do these data follow a normal distribution? If not, create a new column called “diam.norm” that holds the transformed data (common transformations include log transformation, square root transformation, inverse transformation, among others). If you deemed a transformation necessary, plot a histogram of the transformed data, “diam.norm.” **(2 pt)**
6. Create a new data frame that has (at minimum) a column for each plot at each site, the salinity for that site, and the barnacle density for the site. **(1 pt)**
7. Use your newly created data frame to plot the density of barnacles for each plot (on the y axis) as it relates to salinity (on the x axis). **(1 pt)**
8. Plot barnacle diameter as it relates to salinity. **(1 pt)**

Bonus Point: Describe, in words, how salinity, dissolved oxygen, and barnacle density each relate to barnacle diameter.