R Project Part 1

Use an R Notebook to perform analyses on the **RTestData.txt**. Use the R Cookbook to figure out how to do all the analyses. Use the markup language to indicate what analyses have been done. RENAME your notebook, ***YourLastName*RProjectPt1** and load the finished notebook (the actual notebook) to Canvas. *Let me repeat that*: TURN IN THE ".Rmd" FILE, the text file itself. Not a pdf or anything else.

NOTE: You only have to import a dataset and make a dataframe one time in the file. Also, make sure only to read in the file using only the filename "RTestData.txt", not a path to the file that only exist on your computer. More information on the RTestData.txt can be found on this pdf, page 5: https://kelleybioinfo.org/algorithms/basics/programming/r.pdf

Part A: Univariate statistics with R

Normality check & data transformation

(1) Make histograms, run qqnorm and qqline of all bacteria and "deepest".
(2) Transform variables if non-normal and repeat histogram, qqnorm, qqline on transformed data. For transformations I suggest trying either sqrt() or log().
(3) Example: lepto=sqrt(lepto)

For the following, used the normalized (transformed) values.

One-way ANOVA and summary and boxplots:

(1) Lepto by time
(2) Strep by time

Correlation analysis (cor.test)

(1) Strep vs. prev
(2) Strep vs. fuso
(3) Fuso vs. lepto

Linear regressions

Use the lm command and use plot and abline to produce graphs.

(1) Strep vs. deepest
(2) Fuso vs. deepest

Part B: Use ggplot2 to make pretty graphs
First, load the ggplot2 library and dependencies. It may require installation.

Helpful tutorial:
https://www.tutorialspoint.com/ggplot2/index.htm

Produce the following graphs:

(1) ggplot histogram of prev variable.
(2) ggplot scatterplot of strep by deepest, colored by time
(3) ggplot scatterplot of fuso by deepest
(4) Same as 3 but use smoothing algorithm

Part C: Vegan analysis
First, load the vegan library and dependencies. It may require installation.

The following tutorials will be helpful for the exercises. You will need to create a new dataset that only has the bacterial abundance variables. I suggest a new data frame with just the columns needed. (Use the bacterial abundances from **RTestData.txt** for all the problems below.)

Example: newdata=olddata[3:5]

Helpful tutorial:
https://peat-clark.github.io/BIO381/veganTutorial.html

(1) Calculate alpha diversity (Shannon) for the dataset.
(2) Rarefy the data and make a plot using rarecurve.
(3) Calculate Bray-Curtis dissimilarity matrix, show the matrix and plot a histogram.

For this next part, this tutorial should be helpful:
http://environmentalcomputing.net/multidimensional-scaling/

(1) Make a multidimensional scaling plot using the Bray-Curtis matrix.
(2) Color the plot by status, then time.