Elemental Ecology
LDA Exercise
Due by the end of class, November 21st. Email to Seth, Emma, and Alana.

The Mega Primary Producer Library: Using multivariate techniques to find ecological patterns.
For this assignment, you will be using the file "Producer_Library.csv". This file contains:

- $\delta^{13}C$ data for 12 different amino acids from a wide variety of **marine primary producers**. These data were generated by Emma Elliott Smith as part of her dissertation investigating how different physiologies among marine algae may imprint on their isotopic compositions.
- $\delta^{13}C$ data for 12 different amino acids from a wide variety of **terrestrial and freshwater primary producers**. These data were generated by Alexi Besser as part of her dissertation exploring the mechanisms by which allochthonous (terrestrial) material is incorporated into freshwater food webs.

In this assignment you will help Emma and Alexi use multivariate statistical techniques to find differences, or similarities, among these taxa! *As always, remember to check that your working directory, the names of your files, and the names of the columns in the database match what you are telling R to look for!*

1) **Let's start in the marine realm**. We want to see how well a linear discriminant analysis (LDA) can help us in distinguishing among red algae, green algae, kelps, and particulate organic matter (POM). Let's try an LDA using ***all 12*** amino acids. Refer to the code from the in-class demo to get started. **Address the following:**

    i. Attach a pretty figure of producer amino acid $\delta^{13}C$ fingerprints for these marine producers. Plot your producer groups along the first and second linear discriminant axes, LD_1 and LD_2, and include a 95% confidence ellipse around each group).

    ii. Which four amino acids are most informative for distinguishing among groups along LD_1? And, which four are the most informative for LD_2?

    iii. What is the overall successful reclassification rate of this LDA model?

    iv. Which groups are classifying well? Which are classifying poorly? Why do you think this might be?

2) Now, we'd like you to explore how well LDA using AA $\delta^{13}C$ data can distinguish among different primary producer clades. **Choose 4 groups from the dataset and run an LDA**. You may pick any subset of amino acids. These groups could be functional (i.e., column "Type") or taxonomic (i.e., column "TaxonomicID") in nature, but you must have a sample size of >5 for each group*.

    i. Which groups did you choose? Why?

    ii. Which amino acids did you use? Give a statistically, or biochemically, robust justification for why you chose these aminos.

    iii. What is the overall successful reclassification rate of your LDA model?

    iv. Attach a new, pretty LDA figure!

    v. Are there any groups that are still not separating out from one another? Provide a reasonable hypothesis as to why.

    vi. How would you go about testing the hypothesis you provided above? Briefly (1-2 sentences) outline an experimental design.

* A note on sample size: though we do not worry about it here, LDA requires a sample size (per group) of ≥ number of tracers +1. In other words, if you are using data for **6 amino acids**, LDA requires that you have measured **7 samples of each group** you are trying to distinguish between in multivariate space.