Homework 6

## Question 2

1. The Hellinger transformation was used on the zooplankton as the data set contained many insistences where species abundance equalled zero. The Hellinger transformation decreases the weight assigned to variables with low values or any zeros, decreasing their impact on the final results.
2. The first two axis of my PCA explain 43% and 16% of the total variation respectively.
3. The biplot shows that neutral and recovered lakes in general have more similar zooplankton communities than acidic lakes. This suggests that certain species of zooplankton are more abundant in lakes that are acidic, and others in lakes that are neutral. As is indicated by the plot arrows, D. mnts appears to be more associated with acidic lakes while several other species including D cycle, and M. Scyle are associated with neutral lakes. Most neutral and recovered lakes are in the top left quadrant and the plot, suggesting that lake types are separating according to both the first and second axis.
4. Yes, the zooplankton communities appear to differ by lake type, as acidic and neutral lakes tend to have more similar communities to other lakes in their class than to each other. However, there is some overlap between acidic and recovered lakes.
5. Yes, according to the DiagPlot, there are multiple outliers in this data set.
6. For my project, I have a variety of sites with the average depth of different lichen species recorded. The sites are from sites of a variety of ages following fire. A PCA could be used on this data to determine if sites of specified age ranges are different from eachother in terms of lichen depth for different species. This data would be suitable for a PCA as lichen depth is a continuous measurement variable and, given that multiple species of lichen were measured, is multivariate. However, a Hellinger transformation should be performed on the data set before analysis as several sites did not contain lichen, and thus had depth measurements of zero.

## Question 3

1. It is important to standardize the variables for an RDA as they may be recorded in different units/scales. Therefor, a unit that has large numbers associated with it (eg. mL) may have more influence on the results than a variable recorded in a scale that has relatively small numbers (eg. pH). Standardizing effectively grants more equal weight to different variables.
2. Based on the step wise procedure, lake pH, the presence or absence of fish, and the level of dissolved organic carbon were significant predictor variables.
3. The permutation test yielded a p value of 9.999 x 10-5 which is much less than 0.05, indicating that the final RDA is significant.
4. VIF is less than 4 for all three predictor variables indicating that there is not significant collinearity in the RDA
5. The first and second RDA axes explained ~36% and ~7% of the variation in the zooplankton communities respectively.
6. Acidic lakes appear to be associated with lower levels of pH, dissolved organic carbon and fish presence as they consistently plotted along the side of the axis that the arrows do not point towards. Neutral and recovered lakes do not appear to track the predictor variables, but occur at intermediate levels of all three.
7. It appears that acidic lakes have lower pH (logical since that’s what makes them acidic), dissolved organic carbon levels, and fish presence. Neutral and recovered lakes generally appear to have similar physical properties.
8. It appears that B.smn may be associated with fish presence in a lake. However, most other species appear to be associated with intermediate levels of the three predictor variables.

## Question 4: cluster analysis

1. Based on the elbow method, I would select three clusters for this analysis.
2. (why are some means negative? And why are they so low?) There are noticeable differences in the means for different characteristics of the groups. Group 1 has the greatest mean pH, Conductivity, TP, DOC, Ca, and area of the three.
3. The three clusters are clearly separated. However, the cluster analysis is not very effective in separating lakes of different categories. The exception to this is the acidic cluster, which only contains acidic lakes, indicating that most share similar physical characteristics. However, several acidic lakes also fall in the “recovered cluster.” Most recovered lakes do fall in the recovered cluster. However, there are also numerous lakes classified as both neutral and acidic included. This suggests that, while most recovered lakes share similar characteristics, they are not unique to this lake category. There are the same number of neutral lakes in both the neutral and recovered clusters.
4. Several lakes fall into the “wrong” clusters according to lake status categories. The most apparent are recovered lakes Frood and Evangeline falling in the Neutral cluster. The “recovered” cluster includes numerous lakes classified as both acidic and neutral.