Final Report 2.3.3 Primary Paper

Charles Bond

11/01/2020

For this assignment, I attempted to replicate a non-metric multidimensional scaling (NMDS) analysis and plot from a paper examining the effects of saltwater intrusion on diatom community structure (Mazzei et al. 2018) (I finally got EndNote to make a working .bib file, but I had to change the extension manually). This meant making two separate x-y plots (for fresh and brackish water sites) that are an ordination of a Bray-Curtis dissimilarity of diatom relative abundance data from the two sets of sites. Neither the data nor the R code were publicly available, so I had to contact the corresponding author. They gave me the data, but they had lost track of the code. So, I had to try and replicate the analysis by knowing what packages were used, reportedly `vegan` and `MASS`.

Reading the data into R proved tricky at several points. First off, special characters (the greek letter denoting 'micro') in column headers prevented me from importing the data. Later, after having imported the data, I realized after several errors that the data also included several totally blank rows at the end of the dataset, as if in the initial conversion by the author from excel to the .csv file some blank rows at the end were selected, or maybe to rows had been deleted? They said they removed extremely rare species, so maybe the space leftover from the rare species is what did it. So in the end, I had to manually alter headers and exclude the blank rows to get the data to work.

With the original code missing, I was at least able to reproduce the bare bones of the analysis and plots after much trial and error. Part of the confusion early on was that the paper referenced using the `vegan` package, which is consistent with what I've seen online on Bray-Curtis dissimilarity and NMDS plots. However, the author told me they used the `labdsv` package's `nmds` function, which I could not get to produce a similar-looking result. Ultimately, it was the `vegan` functions for dissimilarity and mds plots (`vegdist` and `metaMDS`, respectively) which gave me plots that a 180 degree rotation of the published version, and with different but proportionally correct scaling of the axes (Figure 1 and Figure 2). This is likely because, without the original code, I was left using the default values for the various parameters in these functions. With the nmds being an iterative function, it makes sense that the plot would be the same shape but rotated, as the start state would determine what a positive or negative value on the summary axes is, but without the original code I don't know what they did differently that altered where the function starts, or how it scales the axes.

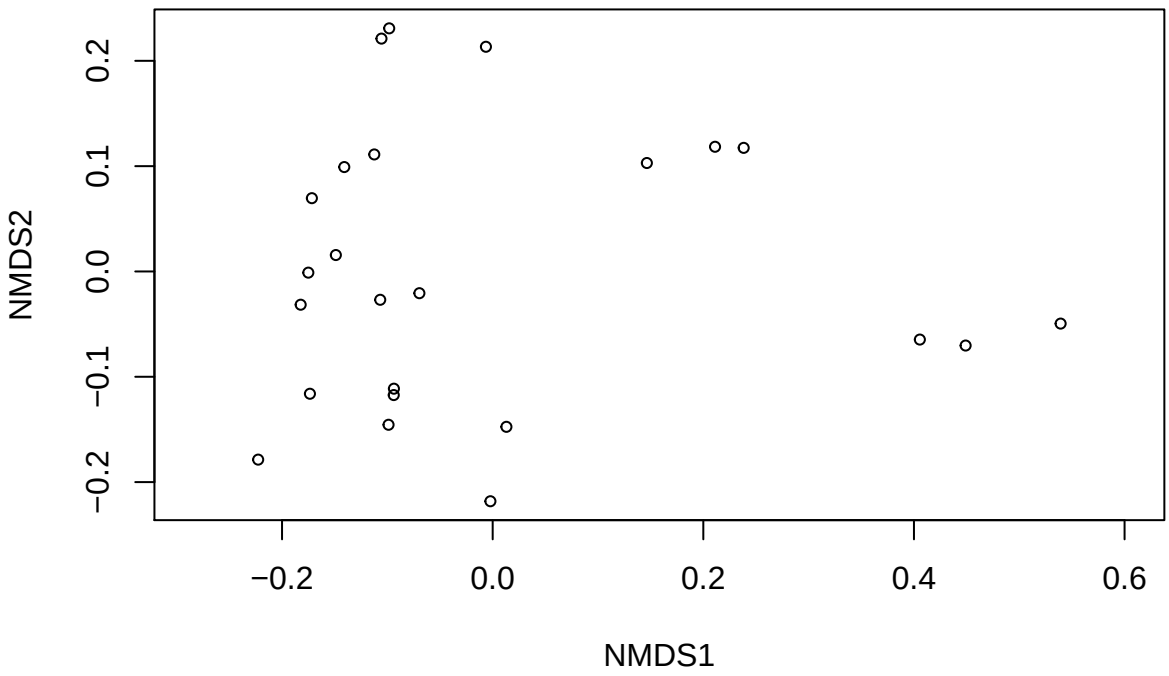

Figure 1: *Ordination of Bray-Curtis dissimilarity of diatom relative abundance from freshwater sites.*

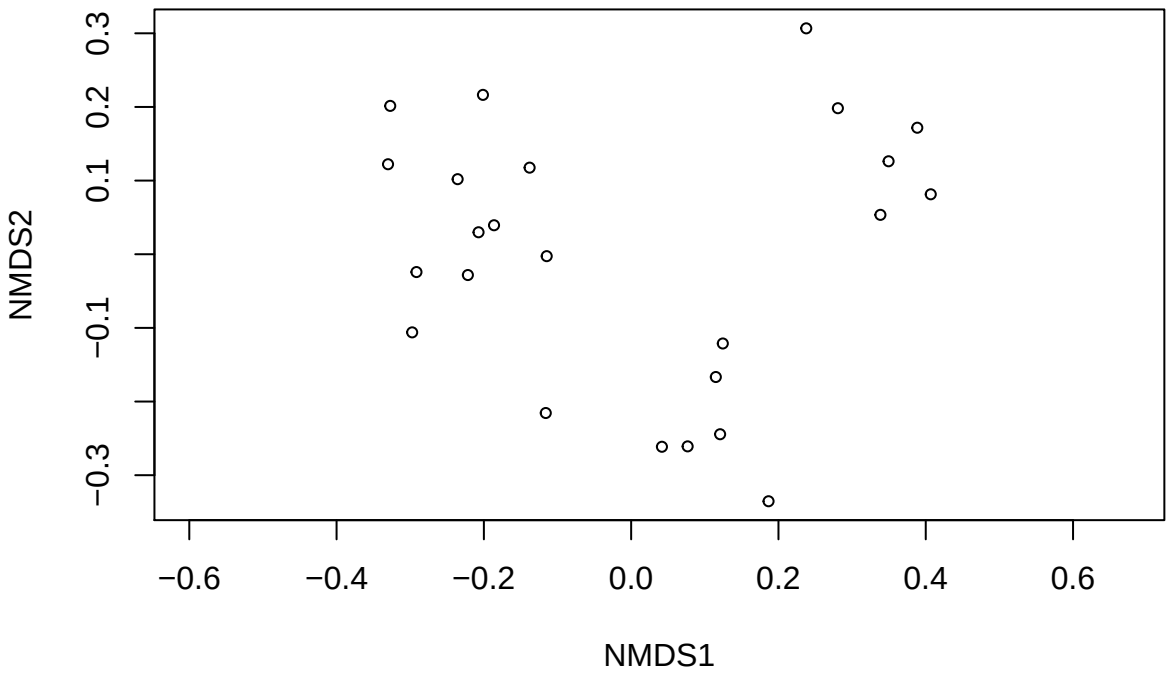

Figure 2: *Ordination of Bray-Curtis dissimilarity of diatom relative abundance from brackish water sites.*

The above figures came only after a lot of figuring out what parts of the data to run through which function from which package. In an earlier version, I ran the brackish and freshwater sites together through the dissimilarity and NMDS functions, which yielded an interesting result, showing clustering by site (Figure 3), but ultimately I realized this was not what was done in the paper. This was useful at least to prove the concept that we could use the function `cbind` to connect the nmds product with the sample metadata and then label the points in ggplot.

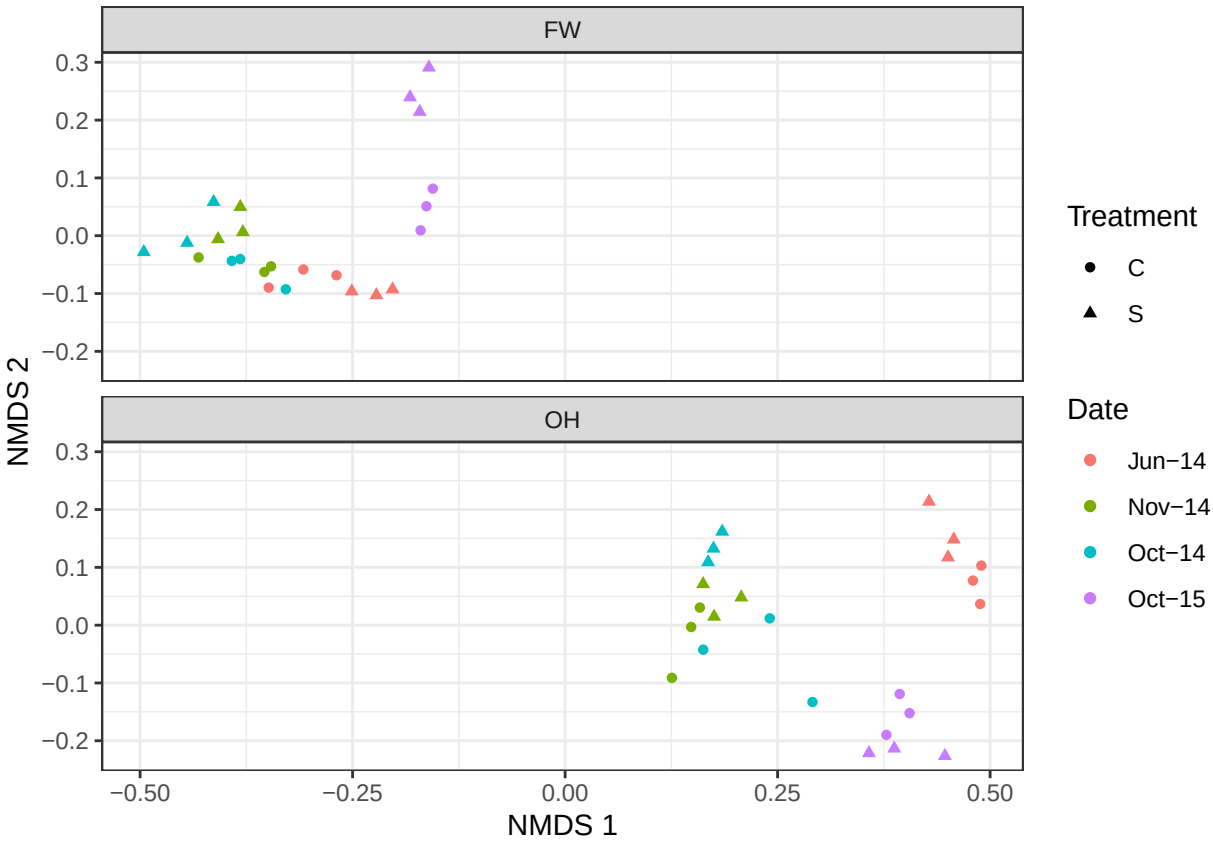

Figure 3: *Earlier attempt at nmds plot, running brackish and fresh sites together and seperating them post-facto. This looks quite different from the published figures, but does show off how sites cluster based on type of site, experimental treatment, and date.*

That is where I have decided to stop this reproduction. I could go on to rotate Figures 1 and 2 by 180 degrees and try to figure out scaling, but it feels arbitrary to try and make the appearance the same if I am missing the code that would allow me to analyze the results. So, rather than spending my time trying to replicate the aesthetics of the plots, I will move on to look for a paper with publicly available data *and* code so that I can replicate an ordination of microbial community data *and* hopefully be able to learn from more complete, well-commented code.

References

Mazzei, Viviana, Evelyn Gaiser, John Kominoski, Benjamin Wilson, Shelby Servais, Laura Bauman, Stephen Davis, et al. 2018. "Functional and Compositional Responses of Periphyton Mats to Simulated Saltwater Intrusion in the Southern Everglades." Journal Article. *Journal of the Coastal and Estuarine Research Federation* 41 (7): 2105–19. https://doi.org/10.1007/s12237-018-0415-6.