EART 101 Paper 1: Analyzing species data

Your goals are to infer the natural environmental conditions, prior to construction of the shipping channel, and to determine biological measures (such as species richness or community composition) that can be used to assess the success of future remediation efforts. You have chosen the sites and determined the sampling strategy, so now you must analyze and interpret the data.

Species composition

As you learned in the video, this type of species count data can be difficult to interpret just from looking at the table of numbers. You will use a technique called "ordination" to create a graph showing the similarity of the different samples (samples that have similar species counts will plot close to each other while samples with very different species will be far apart). Specifically, you will use non-metric multidimensional scaling ordination; the details aren't important but the name will be useful for your methods section.

You can use the graph to assess the similarity of the study site to all of the other samples. You can then combine that result with your data on the environmental conditions (salinity, oxygen, nutrients) to infer the original conditions at the study site.

Open RStudio and type this command at the > prompt:

```
source("https://people.ucsc.edu/~mclapham/eart101/papers/ordination.R")
```

You will need a few of the many add-on packages written for R, so type this at the prompt:

```
install.packages(c("vegan", "ggplot2", "ggrepel", "tidyr"))
```

This only has to be done once, as long as it is successful (no error message). If you get a warning/prompt about problems writing to a directory, let me know!

To run the ordination, you will need to know your group number, which is on your data request handout.

You can view the file at: https://people.ucsc.edu/~mclapham/eart101/papers/group_x_data.csv (where "x" is your group number).

Type this at the prompt:

```
ordination(x) #change x to your group number, e.g. ordination(1)
```

The function will create an ordination plot with the sites in black and the species in red. Refer to the video, lecture slides on Canvas, or ask if you are unsure how to use the plot for your interpretations.

To save the plot, click on the "export" link and select "save as image" or you can take a screenshot.

In addition to the broad results about site similarity that you can determine from the ordination plot, you should discuss specific organisms that you hypothesize may have been abundant in the lagoon prior to modification. The environmental preferences of the different species (whether eurytopic or stenotopic and their specific tolerance of salinity, oxygen, etc.) are important pieces of information.

Species richness

You will also use species richness (the number of species found in a sample) as an indicator of environmental "health". Pay attention to the differences between life and death assemblages. If you collected unequal numbers of specimens from different sites or from life and death assemblages, there will be an additional complication because additional counting of specimens invariably increases the number of observed species.

If your group has unequal count sizes, you should use rarefaction so you can compare richness values at a consistent sample size. If all of your samples have the same number of specimens, you don't have a problem, but the rarefaction will produce a nice-looking graph.

Use this command at the > prompt to load the rarefaction function:

```
source("https://people.ucsc.edu/~mclapham/eart101/papers/rarefaction.R")
```

Type this command at the > prompt to perform rarefaction and plot the results:

```
rarefaction(x) #change x to your group number, e.g. rarefaction(1)
```

Each curve corresponds to one of your sample sites (purple for death assemblage, orange for life, black for the study site).

If you need to adjust for unequal sample sizes, just compare the richness at size of your smallest sample. For example, if some of your samples contain 50 specimens, read the species richness from the position of all the lines at an x-axis value of 50 specimens.