Lab 4: Diversity of Phytoplankton Communities

Lab Overview
In this lab you will examine plankton counts and species diversity to examine plankton dynamics and water column characteristics in the Strait of Georgia.

Learning Goals
After completing this lab successfully, you will be able to:
- use MATLAB more independently, and write your own functions in MATLAB
- appreciate the functionality of log-scales on a plot
- calculate the Shannon-Wiener diversity index using real plankton data
- connect phytoplankton abundance and phytoplankton diversity and their seasonal variability

To hand in
1. The code for your `mymean()` function.
2. A plot with two subfigures of the Total Phytoplankton Count time series. The first subfigure should use a linear y-axis and the other should use a log y-axis.
3. The code for your Shannon-Wiener Index function.
4. A time series plot of the Shannon-Wiener diversity index.
5. Two plots, each with two subfigures, showing the Phytoplankton Count and the SW index for the two years 2003 and 2004 respectively

This week, you will have greater freedom than you've had before to choose your own methods for accomplishing the outlined tasks. Be creative, and remember that entering
```
help function_name
```
will provide you with a description of any built-in MATLAB function.

0. Warmup

Start by creating the function `mymean()` as outlined in the pre-lab. Remember that it needs to be written in its own m-file. Test your function by calculating the mean of the following two vectors:

```
A = [3 5 NaN 7];
B = [2.1 9.4 NaN 7.5 NaN 5.2];
```

You can compare the results of your function with MATLAB's built-in `nanmean()` function. *Once you're satisfied that your function works, please submit the code for your function.*

1. Load the STRATOGEM data

The STRATOGEM project (http://www.stratogem.ubc.ca/index.html) was an interdisciplinary study of the Strait of Georgia's biological and physical processes, attempting to understand how the physical conditions in the Strait of Georgia affected the growth of the plankton. We will be using the phytoplankton count data from this project, courtesy of Dr. Rich Pawlowicz.

From the class website, download the file `STRATOGEM_plankton.xls`. This is a MS Excel file containing the phytoplankton counts, broken up by species. Load the data from this file into MATLAB variables. There are a number of options available to you
- read the Excel file directly using the function `xlsread()`. This can be a bit tricky, but is fantastic once you've got it sorted.
- export to a CSV (text) file and read this using `importdata()`. This worked well for me.
- export to a CSV (text) file and then use `textscan()` as you've done before. This will be cumbersome for a file with so many columns, and it won't give you the header lines.

You should end up with the following variables in your workspace:
1. A 2D array containing the numeric plankton counts. This should have size `51 x 39`.
2. A 1D array containing the corresponding date information, in datetime format. This should have `51` elements.
3. **(Optional)** A 1D cell array made of `39` cells, containing the names of the phytoplankton species. You won't need this today, but it might come in handy when you do lab 5.

2. Plot the STRATOGEM data

Once you have the data loaded into MATLAB, create a time-series plot of the total phytoplankton count of all species. That is, if your matrix containing the phytoplankton counts has the same shape as the original Excel file did, you need to sum the elements in every row. You can use one of the functions `sum()` or `nansum()` to accomplish this, but you need to specify that you want your sum to be taken *along the second dimension*, i.e. along each row, because by default MATLAB will sum along each column.

When you plot your data, you will notice that there is a huge range in phytoplankton abundance, and that it likely makes more sense to plot your data on a log-axis. To do this, you can include the following in your code:

```
set(gca, 'yscale', 'log')
```

Make a figure with two subplots, each showing the time series of Total Phytoplankton Count. Let one figure have a regular linear scale on the y-axis, and let the other have a log-scale on the y-axis. Prepare the figure for submission.

3. Create a function to calculate the Shannon-Wiener Index

Now create a function that will calculate the Shannon-Wiener (SW) diversity index. The input

argument should be a 1D array that contains the counts of a number of different phytoplankton species in a sample. The output will be one number, the SW index. That is, once you have your function, you want to be able to choose a date (by choosing a row in your data matrix) and calculate the diversity on that day by using the row as the input argument to the SW function.

Your function header could look something like this, but of course you can choose your own variable names and your own name for the function:

```
function H = shannonWiener(phytoplanktonCounts);
```

To create the summation in your function, you'll probably want to use a for-loop that runs whatever the length of your input variable is, like this:

```
H=0;  f
for i=1:length(phytoplanktonCounts)
      % Calculate the individual p_i here. Then calculate
      H_i = p_i * log(p_i);
      H = H + H_i;
end
```

You will likely find that the calculation will go more smoothly if you remove any `NaN` or `0` values before the loop.

Now that your function is finished, test it. Use the following species counts to test your function:

`shannonWiener([1 1 1])` should equal `1.0986`
`shannonWiener([1 2 3])` should equal `1.0114`
`shannonWiener([0 0 0])` should equal `0`
`shannonWiener([1 NaN 1])` should equal `0.6931`

If your function returns the correct values, you can continue with the lab. *Please print and submit the code for your function.* For the purpose of submission, to avoid wasting paper, you can probably put it on the same page as the function you created in Part 0.

4. Calculate and plot the Shannon-Wiener index for the STRATOGEM data

Now that you have a function that calculates the SW index, use your phytoplankton data to calculate a time series of the SW index. Because your function only calculates the index for one sample (on one date) you will need a loop that runs over all the rows (dates) and calculates the SW index for the phytoplankton counts on each date.

Make a plot of the SW diversity index in the Straight of Georgia over the time span of the

STRATOGEM data. Prepare this figure for submission.

5. Calculate the SW index and total phytoplankton by year

Make a figure with two subplots. The first subplot should contain the time series of the total phytoplankton count for the year 2003 only; the second should contain the time series of the SW diversity index for the same year. Choose an appropriate type of scale for the y-axis.

Then make another figure with the same two subplots, but for the year 2004.

Prepare both figures for submission.