Lab 1: Data analysis in R: Oceanographic profile
ES 383 September 2016

This lab serves as an introduction to the data analysis, visualization, and interpretation that you will conduct throughout the semester. The tools that are introduced here can be used to graph and analyze data in your labs, data you collect in the field, and data from your independent project. You can use the exercises here as a template for your future analyses.

Moreover, this lab provides experience in writing code—a skill that is becoming increasingly needed in almost any field.

You will use files from sampling done along the Damariscotta Estuary. The data set used here is: Lab01example.csv – CTD data from the deep station

Before coming to lab, please install R and R-studio on your computer.

<u>Installing R and R-studio</u>

1. Install R program at: http://cran.rstudio.com/
This site has options for download for Linux, Mac, and Windows

2. Install r-studio (R interface)
http://www.rstudio.com/products/rstudio/download/
Under “installers” you should see options for Mac and Windows

GRAPHING

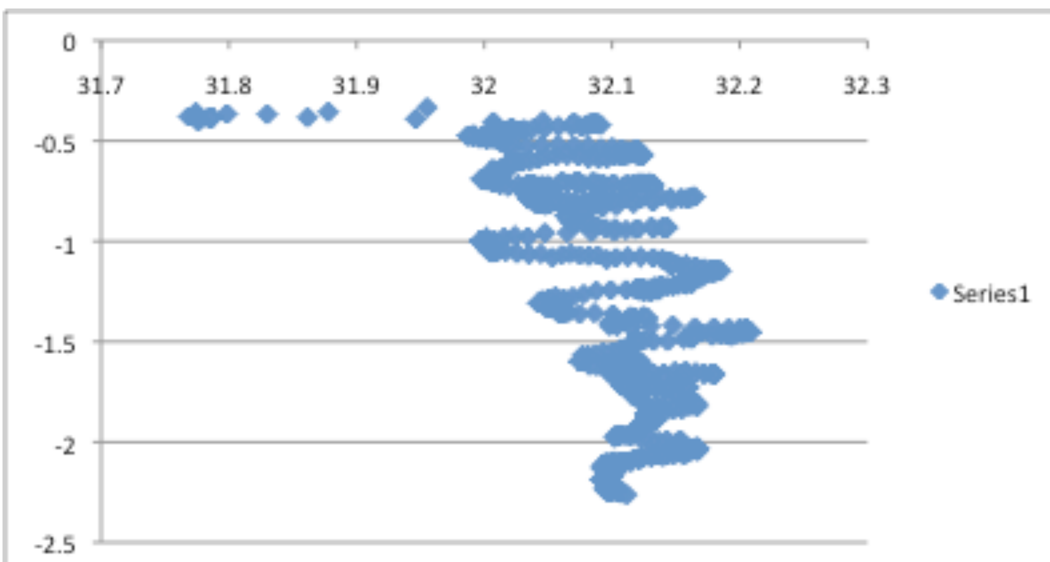
Some basic tips on scientific figures

1. Display information as clearly and simply as possible
2. Label axes (with units in parentheses) and any other critical information
3. Provide details in the figure caption. Don't restate information that should be clear from the figure. However, sometimes it might be helpful to call out a key feature.
4. A title above the figure us optional, but usually not needed.

Here's a nice reference:

Rougier, Nicolas P., Michael Droettboom, and Philip E. Bourne. "Ten simple rules for better figures." (2014): *PLoS Computational Biology e1003833*.
http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

Here's an example of **what to watch out for**:



Problems: Axes are not labeled.
Numbers overlap with axes.
What does "Series 1" mean?
The line of data is difficult to follow.
One side of the rectangle has no line.

Note: Using the literature as a guide is very helpful, but be careful: there are lots of bad figures in the literature. If a figure is difficult to follow, it's probably not a good figure. Make a mental note of what is wrong with it so you can avoid that.

REFERENCE SHEET – PLOTTING IN R

This is a step-by step quick reference for loading in data and making a basic plot. These commands should work mostly as-is, except for the working directory in step 1. There are lots of details that are brushed over here. Those details are covered in the accompanying presentation.

1. Launch R-studio and set the directory you want to work in.

```
setwd('~/Work/Teaching/Colby/ColbyAtBigelow2016/Labs/Lab01/')
```

This command, "`setwd`", is for "set working directory". This can be any directory on your computer, but you should think about how you want to organize your data, your labs, etc. For this example, make sure that the file you want to plot is in this directory.

2. Load the data you want to plot. Here, the file name is "`Lab01example.csv`". The data contained in that file are loaded into a variable called "`DATA`". ***Note***: Different file formats will be read in differently. We'll cover this.

```
DATA <- read.csv('Lab01example.csv')
```

In the command above, we didn't have to call the variable "`DATA`". We could have called it something else, like "`STATION_4`", "`X`", or "`Pinocchio`".

3. Look at a summary of what is in this variable.

```
summary(DATA)
```

Something like this will show on the screen, summarizing the data in "`DATA`":

```
Depth..M.        Temp..deg.C.    Salinity..PSU.    Density..kg.m.3.
 Min.   :  0.166   Min.   : 8.871   Min.   : 0.8708   Min.   : 0.2984
 1st Qu.: 17.636   1st Qu.:10.526   1st Qu.:32.9609   1st Qu.:25.1233
 Median : 44.685   Median :10.892   Median :33.2506   Median :25.4006
 Mean   : 47.196   Mean   :10.857   Mean   :32.3029   Mean   :24.7033
 3rd Qu.: 74.919   3rd Qu.:11.271   3rd Qu.:33.4374   3rd Qu.:25.6468
 Max.   :101.451   Max.   :11.380   Max.   :33.4619   Max.   :25.7321
```

4. To plot temperature (column 2) against depth (column 1):

```
plot(DATA[,2],-DATA[,1])
```

Another option is to use the column headings instead of the column numbers:

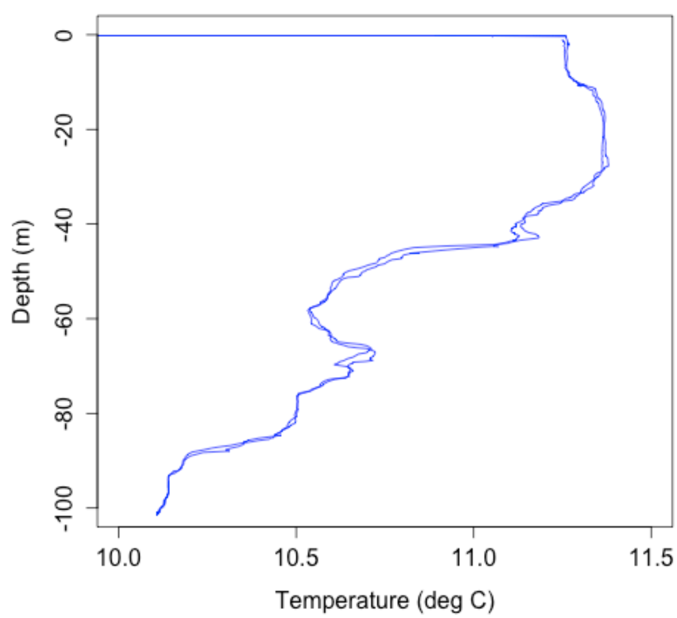
```
plot(DATA[['Temp..deg.C.']],-DATA[['Depth..M.']])
```

And finally, here is an example of some of the options to make the figure nicer (note the options for labeling axes (`xlab`, `ylab`), setting the axis limits (`xlim`, `ylim`) and making it a blue line. There are lots of other options to customize the figure.

```
plot(DATA[['Temp..deg.C.']],-DATA[['Depth..M.']],xlab='Temperature (deg
C)',ylab='Depth (m)',xlim=c(10,11.5),ylim=c(-100,0),type='l',col='blue')
```

5. Save the figure using these two commands:

```
dev.copy(png,'ExampleFigure.png')
dev.off()
```



We will address the following statements/questions using this data set:

1- Describe the physical and chemical characteristics of this site. Plot the vertical profiles (i.e., versus depth) of temperature, salinity, and oxygen.
Once plotted, we will examine the vertical changes in these parameters to infer water column stability and nutrient availability or limitation and discuss what our expectation may be for the profiles in the next points.

2- Plot the vertical profiles of chlorophyll *a* (indicator of phytoplankton biomass) and light (E-par or PAR, photosynthetically available radiation) vs. depth. *Describe the profiles*. Can these 2 parameters be related in another graph, for ex, by plotting chl a vs. light? Try it. How about the logarithm of the data?
Is the relationship easier to understand with this last plot?

3- Use the CTD data to describe different habitats within the water column. What causes the system to be organized in this way?