**Advanced Introduction to R Exercise**

**Explanation**

This exercise will help you to learn how to read in files and install/ use packages as explained in the PowerPoint presentations. This exercise can also be used for you to explore the differences between continuous and discrete variables and you can practice making some basics graphs to explore the data.

This is an advanced example of using R in different ways and working with a real dataset, so don’t worry if some of these ideas and tasks are new to you. This is just supposed to give you an idea of the type of analysis you will learn how to do and to get more familiar with the interface of R.

If you already know how to do some or even all of the tasks described below that’s really great, but no problem if you can’t – just try to understand the main ideas and see this as an example of how you can use R. There will be some more introductory exercises at the beginning of the course, but it’s great if you can already familiarize yourself a little bit with R beforehand.

**Background Information**

A dataset was made, investigating benthic interactions at two locations (coastal vs. offshore). Biological components include the phytoplankton, zooplankton and macrobenthos. We want to know the potential relationships between these components and check the effects of location or the presence of predators on their distribution. The predators were measured as count data to see if they are present or absent.

“Benthicfluxes.txt”



**Research Questions**

Here are examples of research questions you could ask about this dataset.

Which type of statistical analysis would you perform to answer each of those questions?

1. Does the amount of phytoplankton differ between locations (coastal *vs*. offshore)?
2. What is the relationship between zooplankton and phytoplankton?
3. Relationship between macrobenthos and phyto-and zooplankton?
4. Is the amount of macrobenthos influenced by the presence of predators and does this depend on the location?
5. Does the relationship between macrobenthos and zooplankton depend on location?

**Tasks in R**

1. Read in the file and attach the variable names, so that R recognizes them.
2. Load the packages “plyr”, “tidyr” and “car” that you will need to re-arrange your dataset.
3. Load the package “ggplot2” for plotting your graphs.
4. Calculate the average counts for each biotic component (zoo, phyto and macro) per location. You can use the function “tapply()” to calculate the mean values for each component and location separately.
5. Plot the relationships between the different biotic components in 3 plots (zooplankton ~ phytoplankton, macrobenthos ~ zooplankton, macrobenthos ~ phytoplankton).

**Q:** What kind of graph is appropriate here?

1. Plot the densities per location to compare the 2 different locations in a plot for each biotic component (zooplankton ~ location, phytoplankton ~ location, macrobenthos ~ location).

**Q:** What kind of graph is appropriate here?

1. Load the library “lattice” and plot the macrofauna densities per location and the influence of predators (so create a plot for 2 different variables).