**Biodiversity and Macroecology Research Paper**

*Due on Wednesday, December 2nd*

You will conduct an independent research project in which you obtain macroecological datasets (on abundance, species richness, body size, geographic range size, traits, etc.) for a taxonomic group you are interested in, conduct statistical analyses in R on at least three macroecological relationships we have learned about in class, and write a paper contextualizing your analyses and reporting your results. In addition to the paper, you will turn in an R script which documents every step of your analysis from reading in the raw data, to conducting analyses and plotting graphs.

**Timeline**

*Oct 21*: Turn in a proposal of the taxonomic group you will focus on, the datasets you will use, and the three macroecological patterns you will examine (see below). Provide web links to the datasets. For each dataset, describe the relevant data fields you will use. **Submit this info in the Assignment textbox, not as an attachment.**

*Oct 28:* Revised proposal due.

*Nov 2*: Provide a summary of the final datasets you will use for your project.

*Nov 9*: Rough draft of Introduction due; in class peer review

*Nov 9-13*: Schedule a meeting outside class with Allen or Peter to go over analysis plan

*Nov 16-23*: Schedule a meeting outside class with Allen or Peter to go over initial analyses

*Dec 2*: Paper due, uploaded to Sakai before class along with R script and data files. **USE THE FOLLOWING FILE NAMING CONVENTION:**

**[LAST NAME]\_paper.doc**

**[LAST NAME]\_code.R**

***Data files (.csv or .txt) should have whatever name they were downloaded under.***

The paper will be composed of the following sections:

Introduction: (20 pts)

Introduce the three macroecological relationships you plan to investigate. Provide some background and context for examining them. What can examining them tell us? Are the patterns related to each other in some way? Refer to some key papers in the primary literature that have examined these patterns in the past. Are there reasons why the taxonomic group you have chosen is particularly well-suited to examining these relationships? Have other studies examined these relationships with your taxonomic group? If so, then those studies provide an expectation for what you expect to find in your analysis. If not, then are there studies on other taxonomic groups that can set up such an expectation?

*A draft of the Introduction is due by class on November 3, uploaded to Sakai.*

Methods: (10 pts)

Describe the datasets you are using in these analyses. How were the data collected? What is the spatial resolution and extent of the data? When was it collected? How is the dataset delineated taxonomically? For example, if your dataset deals with birds, are there data for all bird species? Just one particular family of birds? Describe your analytical methods and the statistical analyses performed. For example, did you use linear regression? Multiple regression? Variance partitioning? Quantile regression?

Results: (20 pts)

For the heart of your paper, you may (1) analyze 3 macroecological patterns from the list below, or (2) analyze 2 macroecological patterns for each of two taxonomic groups. If there is an analysis you are interested in doing that is not on the list, then check with me. It will probably be ok.

* Use linear models to evaluate potential drivers of species richness
* Examine the relationship between occupancy (# of sites occupied) and abundance (mean abundance at a local site when present)
* Examine range size or occupancy as a function of species traits (body size, dispersal ability, etc)
* Examine variation in beta-diversity between taxonomic groups, or between regions, or between spatial scales
* Conduct a distance decay analysis
* Examine the relationship between local richness and regional richness
* Partition the variance of any macroecological variable between two predictor variables
* Examine invasive species richness as a function of an environmental variable and native species richness
* Examine how metabolic rate varies with body size
* Examine how the total energy use, or total abundance, or species richness of an ecological community varies through time.
* Describe constraints on the relationship between two macroecological variables using quantile regression.

**The analyses you choose will be a factor of both which questions you are interested in, but also which datasets you find to work with.** In the results section, you should describe the results in words, making appropriate statistical and ecological inferences. You should also include tables of statistical output and/or graphs that illustrate your results. Use the class readings for examples of how to format tables and figures. Be careful with your language! In describing the results of statistical analyses, words have specific meanings so keep language simple and don’t use complex synonyms unless you know it’s appropriate. If you are unsure about how to word your interpretation of the results, consult with me.

Note: Your analysis must involve at least two independent datasets (e.g. one might be a dataset on a species’ trait like body size, while another might be survey data on the relative abundance of different species), and you will have to show in your code how to properly use the **merge()** function to integrate information from one with the other.

Graphs and Tables: (10 pts)

**Place graphs and tables of your results into the body of your paper.** Note that in RStudio when you make a graph, you can click on Export > Save Plot as Image or Export > Copy Plot to Clipboard, and then open or paste that image in your document. In R, click on the plot window to make it active, then click on File > Save as in the program’s top menu. Graphs should all have legible axis labels (e.g., “Species richness” and not “HGdata$richness”) set using **xlab** and **ylab** in the plotting statement. Type “?par” to see a list of arguments you can specify when making plots in R, including specifying the symbol type (**pch**), the color of lines or points (**col**), and the size of symbols (**cex**) or graph elements (e.g., **cex.lab** or **cex.axis**). All graphs and tables should be labeled (Figure 1, Table 1) and have appropriate captions.

Discussion: (20 pts)

Whereas the Results section is for describing the results, the Discussion section is for interpreting them. How did the observed results compare to the expectations set out in the Introduction? Does the taxonomic group you examined differ from other taxonomic groups for which these patterns have been examined? In what ways? How do either similarities or differences in the results you found relative to other studies and taxonomic groups shed light on how ecological processes shape these patterns? What additional research and/or data are needed to answer questions prompted by your findings?

References: (5 pts)

Your paper should cite 5-10 peer-reviewed studies ***in addition to*** the data citations. References should be cited in the text using the author and year in parentheses like this (Naya et al. 2013) or this (Swenson and Enquist 2007) or this (Brown 1995). They should be listed in alphabetical order in the References section with the following general format:

[Author names separated by commas]. [Year]. [Title]. [Journal name in italics] [volume number]: [page numbers].

Brown, J. H. 1995. *Macroecology*. University of Chicago Press.

Naya, D. E., L. Spangenberg, H. Naya, and F. Bozinovic. 2013. How does evolutionary variation in basal metabolic rates arise? A statistical assessment and a mechanistic model. *Evolution* 67: 1463–1476.

Swenson, N. G. and B. J. Enquist. 2007. Ecological and evolutionary determinants of a key pant functional trait: wood density and its community-wide variation across latitude and elevation. *American Journal of Botany* 94: 451-459.

Appendix: (15 pts)

In addition to the paper, you will turn in a .R file with your script for reading in, analyzing, and plotting the data, as well as the .csv or .txt data files that you used. This file should be based on the project template file provided during in class exercises (BIOL465\_codetemplate.R) and should allow me to completely duplicate your analysis from start to finish by simply running the entire file in R at once. **Make sure that works!**

Good coding practice is to make sure that you provide lots of comments using the # so that someone else (e.g. me, but also you looking back 2 months from now) knows what each part of your script is doing. For example,

#Read in data

data = read.csv(‘datafile.csv’, header=T)

#Calculate species richness at each site

richness = aggregate(data$Species, by = list(data$site), length)

names(richness) = c(‘site’, ‘richness’)

#Merge environmental variables and richness data

mergeddata = merge(richness, envdata, by = ‘site’, all.x = T)

#Plot richness versus temperature

plot(mergeddata$temp, mergeddata$richness, xlab = “Mean annual temperature”, ylab = “Species richness”)

For a general style guide to writing good, legible R code, see <http://r-pkgs.had.co.nz/style.html>