

# **BIOLOGY 664: Integrated Bioinformatics Using R for Both Wet and Dry Scientists**

## **Problem Set 7: Cluster Analysis and Trees Using data.frames**

**Due: At the beginning of class Tuesday, April 8th.**

### **Chapter 7 Exercises Using data.frames**

Answer questions 1, 2, 3, 4, 5 in the Chapter 7 Exercises using a data.frame for every answer. You should never refer to a matrix in your solutions. Also, create named rows and named columns, and use named indexing whenever possible.

At the very least, you need to hand in an R Markdown File (.Rmd) that logically delineates and presents your R code as well as the R output and plots generated by your code for each exercise. For more information on how to create an Rmd file look here:

[http://www.rstudio.com/ide/docs/authoring/using\\_markdown](http://www.rstudio.com/ide/docs/authoring/using_markdown)

### **Hardcopy**

Also, in addition to submitting a softcopy through blackboard I'd also like to get a printed hardcopy of your Problem Set. Please print double-sided and use Arial font size 7 or 8 in order to save trees.

### **Extra Credit**

I'll be giving extra credit for those Problem Sets that use Knitr and LaTeX to create printed PDF reports that include your R code as well as the R output and plots generated by your code. I'm also going to give extra credit for plots that have been "fancied up" to produce more publication-like figures. These plotting enhancements include colors, titles, legends, axis labels, p-values, and R-squared's. Being able to create publication-quality figures in R is an important skill, and this is an opportunity to earn extra points while learning to do so. Look at the "Quick-R" and "Producing Simple Graphs with R" links on the online syllabus for good online references for how to fancy-up R plots.