

Multivariate Applications in Ecology (BSC 747)

- Ecological datasets are very often large and complex
- Modern integrative approaches have allowed for collection of more data, challenge is proper integration of disparate datasets
- Univariate analyses are often inadequate
- Goal of the class is to cover the most common multivariate analyses used in ecology

Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*)

Multivariate Applications in Ecology

- For each technique:
 - Know the proper use, data assumptions and method of reporting results.
 - Know how to interpret analyses presented in the literature and apply that interpretation to the research question being asked.
 - Be able to compile a dataset, format it appropriately, analyze it in R and extract the components necessary for reporting.

Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*)

“Big Data” in Science

- Spatial data
- *omics
- LTER/NEON/STREON
- Challenges:
 - Computation
 - Analysis
 - Synthesis
 - Visualization

Breeding Bird Survey

DRYAD Data available in Dryad

NCBI GenBank

Structure of class

- I assume you have a basic understanding of univariate statistics (probability, hypothesis testing etc.).
- Generally, one multivariate analysis covered per week
 - Monday:** Introduction to analysis, data structure and assumptions, R functions used, how to interpret results, distribution of paper that utilizes technique, distribution of sample dataset.
 - Homework:** readings, try to perform analysis, interpret output
 - Wednesday:** Discussion of analysis, troubleshooting, discussion of output, discussion of paper and application of analysis.
 - End of week:** Assignments due by the end of the day Friday
- Sometimes multiple similar analyses covered together.

What you turn in

- Due by the end of each week (Friday)
 - Your R code for the analysis. This has to be a self-contained script (file extension of .R) in a separate file ready to run.
 - **One page synthesis of output**, similar to what you might publish in a “Results” section. This should be synthesis, not raw data or output from analyses.
 - **Do not:**
 - copy/paste R output, I want synthesis.
 - Put R code or synthesis in the body of your email. Do not put R code in a word file with your synthesis. **I need two separate files.**
 - Submit via email, no late assignments.
- Graded on a scale of 0-5:
 - 0= Not attempted, or incorrect code and synthesis
 - ...
 - 5= Code works, synthesis correct

R Workflow

- I strongly suggest you keep data you are working with in .csv files, and associated .R code in the same directory.
- To work on something, simply double click the .R script. This ensures the working directory is set as the one where your data is located.
- If you start Rstudio any other way, you will need to manually set the proper working directory.

Your grade

- Details on syllabus.
- There are 14 assignments due, 70 points possible.
- 3 Mulligans
 - A = 49.5+
 - B = 44-48.4
 - C = 38.5-43.9
 - D = 33-38.4
- Working together is fine, turn in your own work.
- Don't procrastinate!

Course Schedule (<http://ichthyology.usm.edu/courses/multivariate/>)

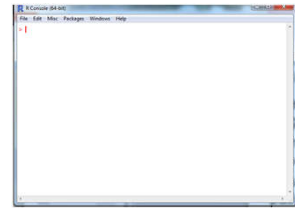
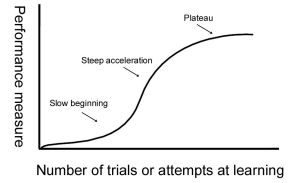
- Jan 17, 26-28 – Course introduction
- Jan 22-24 Introduction to R, Data management
- Jan 29-30 - Analyses of diversity, diversity and similarity indices
- Feb 5-7 - Distance matrices, ANOSIM, MRPP
- Feb 14 - Clustering (no class Monday the 12th)
- Feb 19-21 - Matrix comparisons (Mantel test)
- Feb 26-28 - Ordination I - Principal coordinates analysis
- Mar 5-7 - Ordination II- Principal components analysis
- Mar 19-21 - Ordination III - Non-metric multidimensional scaling
- Mar 26-28 - 62- Constrained Ordination I - Canonical correspondence analysis (CCA) and Redundancy analysis (RA)
- Apr 2-4 - Indicator Species Analysis, TWINSpan and SIMPER
- Apr 9-11 - Regression trees, AIC
- Apr 16-18 - Geometric morphometrics
- Apr 23-25- Phylogenetic Independent Contrasts
- Apr 30-May 2 TBD
- **These can change depending on your interest...let me know.**

Files on course webpage

- Jan 20, 26-28- Course Introduction, software review
- [Download and install R](#)
- You will want some kind of R interface, I suggest Rstudio: <http://www.rstudio.com/ide/>
- [Some useful R manuals](#)
- To install packages, select the Packages tab in Rstudio, then click "Install packages" and type the package name in the box
- Install the following packages:
 - vegan
 - labdsv
 - BiodiversityR
 - car
 - psych
- Some of these may require other packages, these will be automatically downloaded (say yes if it asks for permission).
- [Dataset for first assignment](#) (right click, save as)
- Reading - Chapter 1 in Borcard et al.
- Reading - Chapters 1 & 2 in An Introduction to R (enter "help.start()") in R command line.

What is R?

- Open source statistical language
- Command-line (scripts)
- Base program + 100's of packages for specific analyses
- Everything is free. R comes with the basic packages, additional ones downloaded.



	PC-Ord	CANOCO	NT-SYS	Primer	SPSS	R
Unconstrained Ordinations						
PCA						
PCO						
DCA						
CA (Reciprocal Averaging)						
NMDS						
Factor Analysis						
Partial PCA/DCA/CA						
Constrained Ordinations						
CCA						
RA						
Partial CCA/RADCA						
Clustering					some	
Procrustes/GM						
Permutive Tests						
ANOSIM						
MRPP						
MANTEL						
Indicator Species/TWINSpan						
MANOVA						
Multiple Regression						
AIC/BIC Model Selection						
DiversityIndices						
SimilarityIndices						
Abundance curves						
Rarefaction						
PC and Mac						
Ease of use, data entry and manipulation	Fair	Average (?)	Fair	Excellent	Good	Fair/Poor
Cost	\$325	\$678	\$230	\$500		\$0

R community

HOME PROGRAMME - REGISTRATION NEWS TRAVEL -

useR! 2018

THE CONFERENCE FOR USERS OF R

JULY 10-13, 2018.

BRISBANE, AUSTRALIA.

- Program and packages under active development
- Searchable help archives:
 - <http://tolstoy.newcastle.edu.au/R/>

Journal of Animal Ecology
Journal of Animal Ecology 2015, 84, 1111-1124, doi:10.1111/1365-2656.12302

Measuring β -diversity with species abundance data
Louise J. Barrett^{1,2*}, Nick J. B. Isaac² and William E. Kunin¹

¹Yusuf K. Ibrahim and Cooperative Biology, University of Leeds, L.C. Muf Building, Leeds LS2 9JT, UK; and ²Biological Records Centre, Centre for Ecology and Hydrology, Mansel Building, Mansel Lane, Dorsetland Office, Walsingham DN11 9SS, UK

obtain median β for that combination. All simulations were carried out in R v.3.0.3 (R Core Team 2014). Formulae for evaluating β -diversity metrics for each of the properties can be found in Appendix S2.

MOLECULAR ECOLOGY RESOURCES
Molecular Ecology Resources 2018, 18, 102-103
doi:10.1111/1755-0999.12288

HZAR: hybrid zone analysis using an R software package
ELIZABETH P. HERRYBERRY¹, GRAHAM E. HERRYBERRY¹, JAMES M. MALEY¹ and ROSE E. BIRCHFIELD¹

¹Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA 70118, USA; ²Missouri Botanical Garden and Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA; ³Missouri Botanical Garden, Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA; ⁴Missouri Laboratory of Zoology, Oxford Road, 9009, George Road, Los Angeles, CA 90031, USA

Abstract
We present a new software package (HZAR) that provides functions for fitting molecular genetic and morphological data from hybrid zones to classic equilibrium cline models using the Maximum Entropy Method (MEM) and Bayesian methods. The software applies likelihood functions appropriate for different types of data, including diploid and haploid genetic markers and quantitative morphological traits. The available design allows flexibility for future data types and marker combinations. To facilitate benchmarking, we provide functions to simulate data sets

R and S+

- S+ - commercially available version (~\$900), company provides graphical interface and technical support
- Anything you can do in S+, you can do in R

TIBCO Spotfire S+

With its professional and statistical programs for data analysis and visualization, Spotfire S+ is the only commercial software that offers the power of a programming language, the ability to analyze gigabyte-class data sets on the desktop, and a package system for creating, saving and deployment of analysis to the enterprise across multiple environments.

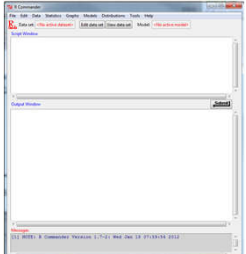
It also integrates with TIBCO Spotfire through TIBCO Spotfire Statistics Services, enabling the creation of powerful analytic applications when using the power of S+ and R's statistics combined with the interactive visualization of TIBCO Spotfire for visualization.

TIBCO Spotfire S+ Value Proposition

- Provides rich analytical tools for using Spotfire S+
- Package analysis for local or cloud-based environments
- Leverage the stability and reliability of R's ecosystem
- Streamline or modernize production environments (e.g., 32-bit OS, 64-bit OS)
- Easily integrate with existing graphics, reports, and dashboarding tools through interactive Spotfire applications

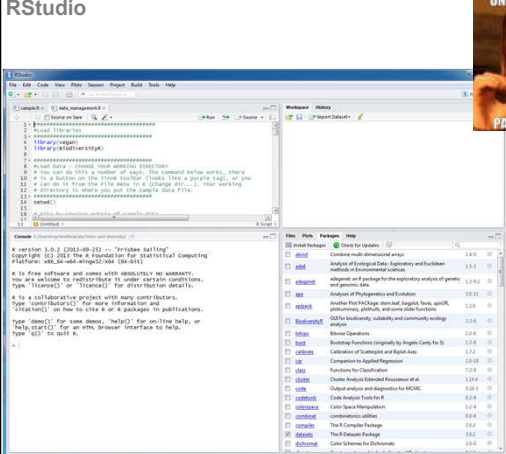
R basics

- Various tools available
 - R commander
 - Point and click GUI, records script
 - About 50% of what we do can be done "point-click" style
 - Rstudio – easiest way to write, edit and run code
 - Free (www.rstudio.com)
 - Works on any operating system
 - Command completion (tab) and history (arrow keys)



RStudio

ONE DOES NOT SIMPLY PASTE CODE INTO IR!



Sample R Code

```
#####
#Load libraries
library(vegan)

# load data from file, save in fish_size object
fish_size<-read.csv("A_mitchilli.csv", header=TRUE)

#####
# View the matrix and some properties
# To view something, just enter the name
fish_size
# dimensions (rows x columns)
dim(fish_size)
# average
mean(fish_size$length)
```

Comments (lines starting with #)
Loading libraries
Loading data from a file

Syntax highlighting – In Rstudio (or other tools), text is formatted based on language specifics. Here, comments are green, variables blue, commands purple.

R command

```
fish_size<-read.csv("A_mitchilli.csv", header=TRUE)
```

Object Created R function File name Information about the file

```
# dimensions (rows x columns)
dim(community)
# total sum
sum(community)

# column and row names
colnames(community)
rownames(community)

#log+1 transform community data
community_log <- log(community+1)
```

R data types

- **Variable** – single value
- **Vector** – one dimensional series of values
- **Matrix** – two (or more) dimensional series of values
- **Dataframe** – multidimensional, may be composed of various types of data
 - Variables within a dataframe referenced with \$
 - E.g. data\$Species1 for the first column here

site	Species1	Species2	Species3
A_08	1	6	2
B_08	2	17	1
C_08	2	22	2
D_08	0	11	1
E_08	1	25	2
F_08	1	64	0
A_09	4	63	1
B_09	0	39	1
C_09	4	14	2
D_09	5	74	1
E_09	4	71	0
F_09	2	32	2

Loading data

- read.csv command will read data from a text file (export from excel as csv).
- **fish_size<-read.csv("A_mitchilli.csv", header=TRUE)**
 - A_mitchilli.csv: file name, must be in quotes
 - header=TRUE: first row contains variable names (FALSE if not names)
 - row.names=If one of your columns has names for each sample, it goes here. You can also use row.names=1 to use the first column as row names. You can't have duplicate row names! Note that our first assignment does not have row names.
- The file to be loaded must be in your working directory (directory where your R script is located)
 - In Rstudio, look at the Files tab in the lower right panel
 - Can change your working directory various ways

Help!

- `help.start()` gives you access to a variety of R manuals
- To get help on any function enter "?" plus the function name
 - For example: "?read.csv" will give you a summary of the function used to load data
- ?? – use to search help for a specific term