

Advanced R programming for biological sciences

Spring 2018

BIONB 7200, Lec 002 (#18319)

Instructor: Marcelo Araya Salas

Permission of Instructor only: Contact Dr. Araya Salas (araya-salas@cornell.edu) for a questionnaire regarding your programming/ coding experience. If appropriate, he will provide you with the registration code. Intended for graduate students and upper level undergraduates.

Course description

R has emerged as a preferred programming language in a wide variety of data intensive disciplines, including the biological sciences, becoming a critical tool in research. This course will introduce students into basic computer programming, database manipulation, customized graphing, and advanced statistical and modeling techniques under the R programming environment. The course is aimed to graduate students with basic knowledge in R. Students will learn how to program in R and how to use R for effective data manipulation, data analysis and presentation of results.

Credits: 2

Grading Procedures: S/U. There will be three programming assignments. Each assignment must be submitted in Rmarkdown and html formats. In addition, students will give a presentation on an R package potentially useful for their research.

Learning outcomes

After completing the course the student is expected to use R to:

- Import and manipulate typical biological data sets
- Understand the documentation of R packages
- Understand function structure and create simple functions
- Create dynamic reports of data analyses
- Use loops to run repetitive computing tasks
- Create elaborated (e.g. multipanel) graphics

Course format

Weekly two hour long lectures in which R code is interspersed with computer demonstrations run by students. Each topic is also followed by biologically meaningful examples and in-class practices.

Tentative course Schedule

Class	Week	Theme
1	01/24/18	Class introduction
2	01/31/18	Introduction to R language
3	02/07/18	Importing and formatting (tidy) data
4	02/14/18	Advanced data manipulation (“dplyr”, “data.table”, “reshape”)
5	02/21/18	Functions

Class	Week	Theme
6	02/28/18	Efficient programming and parallelization (“parallel”) (1st assignment due)
7	03/07/18	Conventional statistics (e.g. linear models, logistic regression)
8	03/14/18	Multivariate statistics (ordination and hypothesis testing) & predictive modelling (“caret”)
9	03/21/18	Randomization and bootstrapping techniques (2nd assignment due)
10	03/28/18	Basic R graphs
12	04/04/18	<i>*Spring Break*</i>
11	04/11/18	Ggplots (“ggplot2”)
13	04/18/18	Maps (“maps”, “ggmap”) (3rd assignment due)
15	04/25/18	Dynamic reports (“knitr”, “Rmarkdown”)
16	05/02/18	Student presentations (session 1)
16	05/09/18	Student presentations (session 2)

** R package names are shown in quotes*

Additional learning resources

- Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59:390–412
- Bates D, Maechler M, Bolker BM, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. <http://cran.r-project.org/package=lme4>.
- Bolker BM (2008) *Ecological Models and Data in R*. Princeton University Press, London
- Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24:127–35
- Borcard D, Gillet F, Legendre P (2011) *Numerical Ecology with R*. Springer, London
- Crawley MJ (2007) *The R Book*, 2nd Editio. Wiley, Southern Gate
- James G, Witten D, Hastie T, Tibshirani R (2013) *An Introduction to Statistical Learning with Applications in R*. Springer
- Maindonald J, Braun WJ (210AD) *Data Analysis and Graphics Using R - an Example-Based Approach* (J Maindonald and W. Braum, Eds.), Third edit. Cambridge University Press, London
- Reimann C, Filzmoser P, Garrett RG (2008) *Statistical Data Analysis Explained*.
- Touchon JC, McCoy MW (2016) The mismatch between current statistical practice and doctoral training in ecology. *Ecosphere* 7:e01394
- Warton DI, Lyons M, Stoklosa J, Ives AR (2016) Three points to consider when choosing a LM or GLM test for count data. *Methods Ecol Evol*:n/a-n/a
- Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol* 1:3–14

Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) Mixed Effects Models and Extensions in Ecology with R. Springer, New York, NY

Online Resources

- <https://cran.r-project.org>
- <https://cran.r-project.org/manuals.html>
- <https://cran.r-project.org/other-docs.html>
- <https://www.r-project.org/doc/bib/R-books.html>
- <https://www.stat.auckland.ac.nz/~paul/Talks/Rgraphics.pdf>
- <https://www.r-bloggers.com/a-fast-intro-to-plyr-for-r/>
- <https://www.r-statistics.com/tag/visualization/>
- <http://blog.revolutionanalytics.com/2014/04/some-r-resources-for-glms.html>
- <https://www.datacamp.com/community/tutorials/r-tutorial-apply-family>