

Ecological Analytics

Location: Harvard Forest June 2015 9-10am EDT

Instructor: Matthew Lau (<http://people.fas.harvard.edu/~matthewklau>)

Goal: Students will learn how to use the R programming language for ecological analyses and gain experience:

- Managing analytical aspects of a project
- Inputting, manipulating and exporting data
- Statistical functions
- Plotting data patterns
- Coding and software best practices
- Getting more help, experience and practice

Although statistics will be introduced very briefly, this will not be a statistics class.

Pre-requisites: experience with using basic computer software. *Required Materials:* laptop or access to some computing device.

Calendar

June 1 (Monday), 7:45-8PM: brief intro to ecological analyses. - Why program? - Brief introduction to project management - Connect to syllabus and course materials

June 2 (Tuesday), 9-10AM: Analytics and programming basics. **Before Class**

- Connect to Harvard Forest Guest wireless*
- Install R on your computer: <http://lib.stat.cmu.edu/R/CRAN/>*
- Download the example project: <https://github.com/HarvardForest/myProject/archive/master.zip>

In Class

- Introduction to analytics, the usefulness of programming and R
- Present the overarching analytics framework (Noble 2009 = document organization, task scheduling, data management, analytics management, keeping an analytics notebook)
- Basic functions/operations and tasks (e.g., do simple calculations on scalars and vectors)
- Objects = way to keep data in your memory
- Functions = name(arguments)
- Scripting and annotation = save your code and notes about code

Challenge

- What's the area of a square with a side = 10 cm?
- What's the circumference of a circle with diameter 10 cm?
- Find a cool equation and play with it in R (e.g. $E=mc^2$)!

June 4 (Thursday): Data management

In Class

- Data structure and spreadsheets
- Inputting and managing data overview
- Manipulating vectors (sorting, ordering)
- Manipulating matrices = sorting and appending
- Data reading functions (read.csv, read.table)
- Advanced data (mode/class, data frames and lists)

Challenge

- How does a circle's circumference change as it's diameter changes?

June 9 (Tuesday): Data analysis

In Class

- Getting data from the HF archives
- Calculating basic statistics (mean and variance)
- Scatterplots, Barplots and Histograms
- Components of the plot (axis, labels, colors, cex, etc. . .)
- Writing your own functions (se: input, process, output)
- Packages = gplot
- Barplot with error bars

Challenge

- Analyze an HF dataset

June 16 (Tuesday): Data, Programming and Research Tips and Best Practices

- Modular Code = just make the wheel once
- Versioning = github
- Loops and applys
- Simulating data = runif and rnorm
- Data and software ethics
- Data provenance
- Hackathon?

Readings and Resources

- R Cheat Sheet – <http://cran.r-project.org/doc/contrib/Short-refcard.pdf>
- Version Control – <https://help.github.com/articles/good-resources-for-learning-git-and-github>
- Code for America – <https://www.codeforamerica.org>
- Learning Statistics
- *Primer of Ecological Statistics* by Ellison and Gotelli
- *The Ecological Detective* by Hillborn and Mangel