Open Source for Open Science – an EEB Summer Workshop
July 10 – 12, 2014
Building/Room TBA
College Station, TX 77843

Course Details
Open Source for Open Science (OSOS) 2014 is a free summer workshop featuring an introduction to R, *nix commands, and open source GIS. This two and a half day course will provide tools for reading in, manipulating and analyzing your data in R, as well as other open-source tools for visualizing and manipulating spatial data. It will provide a basic introduction to scientific programming, creating publication quality graphics in R, as well as wrangling messy data with *nix commands in bash. Although the course is aimed at the EEB graduate student, you need not be in EEB or a graduate student to attend. The modular design of the course will allow participants to attend sections of the course that are useful to them and skip the parts that are not. We encourage you to attend the first day if you are not an avid R user.

Instructors
Rebecca Clark, Thomas Olszewski, Claudio Casola, Mike Treglia, and Michelle Lawing.

Requirements
No previous knowledge of R is required. The course is open to all students, postdocs, and faculty. **A laptop is required**. If you do not have a laptop, please contact Michelle (alawing@tamu.edu) and we will set you up with a loaner.

Schedule
The workshop will start at 9:00 am July 10th and wrap up around noon on July 12th. Snacks and coffee will be provided during morning and afternoon breaks and participants will have an hour for lunch.

Day 1 - July 10th

9:00 am
Session 1 – Introduction to the R environment (Rebecca Clark)

1. Overview - (Presentation)
2. Basic interface (Handout)
    a. Simple calculations
    b. Variables and assignments
    c. Functions
3. Loading data (Handout)
    a. Directories, paths, and finding files
    b. .csv and .txt
    c. Headers

    d. Formats (vectors and data frames)
    e. Group activity ("fixing" an input file for R)
4. Manipulating data (Handout) (R File)
    a. Subsetting
    b. Indexing
    c. Handling missing values

Lunch Break

1:00 pm
Session 2 – Introduction to the R environment continued (Michelle Lawing)

(presentation)

1. Organization of projects
    a. File hierarchy
    b. Reference cards
    c. R style guide
    d. Well documented scripts
2. Base plotting
3. Some simple statistics (Exercise Folder)
    a. Regression example with some plotting
    b. ANOVA example with sums of squares explanation
4. Packages and CRAN
5. Where to go for help
6. R studio
7. Wrap up

Day 2 – July 11th

9:00 am
Session 3 – R for Scientific Programming (Tom Olszewski)

1. Introduction to programming
    a. Control structures
    b. Vector notation
    c. How to write instructions as a script
2. Random walk example
3. Randomization with Monte Carlo and bootstrapping

Lunch Break

1:00 pm
Session 4 - *nix commands in bash and R (Claudio Casola)

1. *nix commands
2. Regular expressions
3. Invoking R from command line

3:00 pm

Session 5 – Publication Quality Graphics (Rebecca Clark)

1. base graphics details
2. plotting with lattice
3. plotting with ggplot2

Day 3 – July 12th

9:00 am

Session 6 – Open Source GIS (Mike Treglia)

1. Some GIS Basics
    a. Projections
    b. Data types and formats
2. GIS in R (Packages "sp", "raster", "rgdal", "rgeos")
    a. Working with vector data in R
    b. Working with raster data in R
    c. Visualizing data and making basic maps
3. QGIS
    a. Importing Data and Viewing Attributes, Settings, etc.
    b. Create new Vector datasets (digitizing, or creating Regions of Interest [ROIs])
    c. Make basic maps
    d. Plug-ins for QGIS
4. A Brief Overview and Demo of SAGA GIS (Macs need boot camp, parallels)
    a. Highlight Some Capabilities, Particularly for Raster Data

Websites

http://cran.r-project.org
http://www.statmethods.net
http://stackoverflow.com/questions/tagged/r

R Users Group

Email listserv and archive

To establish collaboration
To get in touch with others working on the same analyses or packages
Shared code editing