EEB 201 – Introduction to R for Ecology and Evolutionary Biology

Course description

Lecture, six hours; discussion, six hours. 1 unit

Requisites: Enrollment in EEB doctoral program or permission of instructors.

Introduction to basic concepts and practice of scientific programming, using the language R. Topics include working at the command line, writing scripts and functions, flow control, graphics, and conducting basic simulations in discrete and continuous time.

Offered as intensive two-day course at beginning of quarter. S/U grading.

Background

No previous experience in computer programming is assumed. Preliminary readings will be assigned to give all students a common foundation in basic principles of using R. The course will require access to a computer with R installed; students can use their own computers, or borrow a laptop from the library.

Installing R

R is completely free and installs easily on Windows, Mac and Linux systems. Please go to

http://www.r-project.org/

and download and install R on your computer. There are lots of manuals and other resources available on the R website as well. R-Studio is a free package that creates a nicer user environment for R – it is recommended, but optional. It can be downloaded from:

http:// http://www.rstudio.com/

Course aims and format

As sophisticated modeling and analyses of large datasets continue to permeate ecology and evolutionary biology, students need to be equipped with sufficient computational tools. This course aims to give all students a basic foundation in scientific programming in the R environment. Topics include working at the command line, writing scripts and functions, flow control, graphics, input and output of large datasets, basic statistical analyses, and conducting basic simulations in discrete and continuous time. The format of the class is a mix of lectures, demonstrations, discussion, and hands-on practice in executing various tasks in R. Students will be assigned a set of exercises to complete on their own, and will have the opportunity to discuss this work with the instructors. The skills learned in this course will be directly applicable to later EEB courses and to students' own research projects.

Learning objectives

By the end of the course students will:

- Be comfortable executing basic commands in the R environment.
- Be able to load packages in R and make use of their added functionality.
- Be able to read in data files, manipulate data, and perform simple analyses in R.
- Be capable of plotting curves, scatter-plots, histograms, and other graphic outputs in R.
- Be able to write their own computer programs to simulate population models in discrete or continuous time.

Evaluation

The course is graded on an S/U basis. Students will be assigned preparatory work and three sets of programming exercises, which will require them to apply what they have learned during the course. Each of these assignments will be graded out of 25 points, for a total of 100 possible points. Students who attend all sessions, participate actively in class, and complete all assigned exercises (with mean grade of 70% or higher) will be judged to have performed satisfactorily.

Detailed syllabus

Before the course begins, students will be asked to complete online pre-training to develop basic skills in the R environment. The completed training record must be submitted before students can participate in the course.

Morning of Day 1:

- Introduction and review of pre-training work
- Executing basic commands in the R environment
- R data types
- Constructing R functions
- Installing and using R packages
- Flow control

Afternoon Day 1:

- Programming style
- Pseudo-code
- Writing and working with R scripts
- Simulation
- Students complete Problem Set 1 and begin Problem Set 2 (with help from instructors and instuctional assistants)

Morning Day 2:

- Review of Day 1 and time for questions
- Dynamic models
- Basic plotting functions
- Basic statistical analyses
- Analysis of large datasets

Afternoon Day 2:

- Students complete Problem Sets 2 and 3 (with help from instructors and instuctional assistants)

All course materials must be submitted to the instructors no later than one week after Day 2. Materials submitted later than this will not be accepted.

Justification for course

Research in the natural sciences has been transformed by computers. From data analysis to visualization to simulation, computational tools have become essential to conduct and communicate scientific research. This means that computer programming is a vital, foundational skill for any new scientist, yet many students do not receive adequate training in programming in their undergraduate programs. This course aims to introduce new graduate students to scientific programming in the R environment, giving them a foundation in skills and concepts such as working at the command line, writing scripts and functions, flow control, graphics, input and output of large datasets, basic statistical analyses, and conducting basic simulations in discrete and continuous time. After completing the course, students will be empowered to use computational approaches in their research and coursework, and to extend their programming skills to advance their career aims.