MINI-COURSE BIOS248

Scientific Computing for Ecologists, Biologists and Environmental Scientists

Units: 2
Date: Oct. 5-9 2015
Location: Hopkins Marine Station
Target audience: graduate students in Biology and E-IPER.
Post-docs will also be eligible if space available
Prerequisite: None
Enrolment: through Axess for graduate students, by email for post-docs.
Contact email: hms-scicom@lists.stanford.edu
Deadline for enrollment: Sept. 20 2015
Coordination: Giulio De Leo
Instructors: Stefano Allesina (University of Chicago)
Francesco Ferretti (Hopkins Marine Station, Stanford University)

Introduction
Today, the average graduate student in biology needs to become far more computer-savvy than she/he needed to be just 10 years ago, largely irrespective of the specific field of investigation, whether biomechanics, bio-tracking, genomics, neurobiology, ecology, evolutionary biology, biological conservation, fishery management, immunology, or molecular biology. The quantity, quality and complexity of the data a successful PhD in biology has to deal with has been growing exponentially, the basic computational skills that were useful a decade ago are now completely inadequate. This data flood requires better data organization and flow. Moreover, agencies and publishers are increasingly requiring scientists to publish data and source code of the analyses used in submitted manuscripts. Thus, the organization of code and data should be approached with the same emphasis on clarity, reproducibility and style that is characteristic of scientific articles.

Data management and computing are challenges for conventionally trained biologists, ecologists and environmental scientists. By definition, scientists are trying to do something no one has attempted before. As such, no off-the-shelf software or analytic protocol is typically available for the analysis they want to perform. Hence the need to develop more sophisticated programming skills. For a new assistant professor, managing many projects involving many collaborators can be a daunting task, if he/she has not mastered the skills needed to organize the data, code, figures and manuscripts in a transparent, logical and reproducible workflow. Therefore, we at Stanford need to take the lead and train the next generation of graduate students and post-docs in biology and environmental sciences in the most advanced techniques needed for big data processing and management, providing them with the technical expertise required to be successful in the present and future highly competitive scientific market.

Recently, Software Carpentry workshops have been held on campuses all over the world, including Stanford University, to increase scientists computer literacy and workflow efficiency. The program, developed by the Mozilla Foundation and Sloan Foundation, organizes short workshops to introduce graduate students to scientific computing. While extremely useful, these workshops are meant to serve a very wide scientific audience, and are thus sometimes perceived as too generic and basic for students in ecology, biology and environmental science, who would instead need being taught biological-specific material, with a depth and a level of sophistication that is difficult to achieve over the span of two days.
COURSE DESCRIPTION

With this mini-course, we propose to raise the bar set by Software Carpentry by presenting a much more advanced course in scientific computing that illustrates the methodological approach and tools to address problems typically encountered by graduate students in biology, ecology and environmental science, and allow students to train and work on case studies published in peer-reviewed scientific journals. We propose a full-time, weeklong course, striving to achieve a balance between lectures (mornings) and hands-on activities (afternoons), The course will cover:

(a) mining, structuring and management of large biological datasets. Use of open source relational database management systems such as PostgreSQL and MySQL.

(b) Unix/Linux operating systems
(c) Bash programming and regular expressions
(d) Good programming practices in R and Python (languages that are popular among biologists)
(e) Integration of multiple and diverse computing tools to create transparent and reproducible workflow from collection of data through analysis to publication:
   - Revision control systems such as Git, and the use of online repositories using git such as Github or Bitbucket for tracking scientific workflow development and increasing efficiency of collaborative analyses;
   - Creation of R-CRAN packages; plotting with ggplot and other R graphical packages; elements of data visualization.
   - open source software for dynamic report generation such as LaTex and markdown and associated packages and programs for the creation and conversion between multiple markup languages (RStudio, Rmarkdown, Sweave, pandoc and knitr).

The course will have plenty of hands-on and interactive learning opportunities both during lectures and in lab sessions. A distinctive feature is that all examples and exercises will be based on real world case studies, extracted from peer reviewed publications, or scientific problems on which either the instructors or the students are actively working. Q&A sessions will be organized systematically through the day and to foster class discussion. Case studies will be presented in the morning, students will work on them in the afternoon and, by the end of the working day, they will present and discuss the outcome of their projects.

INSTRUCTORS
In addition to Giulio De Leo, course coordinator, lectures and training sessions will be held by

Stefano Allesina, Professor, Department of Ecology & Evolution and Computation Institute at University of Chicago, author of the book “Introduction to Scientific Computing for Biologists” to be published next year by Princeton University Press

with the support of Francesco Ferretti, post-doc in fishery science, statistical modeling and bioinformatics at Hopkins Marine Station.

LOCATION
Hopkins Marine Station, Monterey Bay
Please, contact us at hms-scicom@lists.stanford.edu to discuss accommodation if you come from campus