R Bootcamp 2016

Overview

Introduction to the R programming languange with a focus on using it for biological data analysis. The purpose of this course is to teach scientists (students, postdocs, PIs) in the biological and medical sciences to use R for typical data analyis tasks they might encounter routinely. This includes sequence analysis and other bioinformatics tasks. No prior knowledge of R is expected and workshop attendees can expect to come away with a skill set that is immediately translatable to their respective data tasks.

Workshop Structure

This is a three week course that will consist of a two hour lecture on Monday, Wednesday, and Friday mornings followed by a one hour lab session in the afternoon. There will be a total of 5 assignments which will be handed out during the labs, however, it is expected that attendees will need to spend time outside of the course in order to complete the assignments.

Learning Objectives

At the end of the workshop you will able to:

- Install and update R

- Use the Rstudio IDE

- Understand what CRAN and Bioconductor are and what the differences are between them

- Install and update R packages from CRAN and Bioconductor

- Import a wide variety of data types into R

- Understand the basic data types: integer, numeric, logical, character

- Understand R's basic data structures: vector, matrix, list, data.frame

- Understand basic programming concepts: functions, objects, loops, vecotrization, conditionals

- Manipulate data structures by subsetting and indexing

- Understand key base R functions: seq, apply (and friends)

- Manipulate data with dplyr

- Make plots with ggplot

- Find help about any function

- Understand some common R errors and how to deal with them

- Find and evaluate R packages needed for a particular analysis

- Understand the difference between <- and = and make your own choice about which one to use

-

Outline

Preparation

Attendees are expected to come with their own laptops and have already installed R and RStudio as well as completed at least one of the following online tutorials.

- https://www.datacamp.com/courses/free-introduction-to-r
- http://tryr.codeschool.com/
- http://swirlstats.com/

This small bit of preparation will allow us to move quickly through the basics and get to the good stuff.

Day 1: The basics

- Introductions and course objectives
- Make sure everyone has R and Rstudio running
- Brief overview of the basic material
  - Data types
  - Data structures
- How to get help and deal with errors
- Lab 1: Working with basic data types

Day 2: The not-so basics

- Continue programming concepts
- Environments
- How to work with objects: subsetting, indexing, apply functions
- Other important key R functions including basic statistics
- Lab 2: Working with dataframes and lists

Day 3: Practical data management

- dplyr and friends
- Restructuring data and doing stuff to it
- Bonus: stringr and regular expressions
- Lab 3: Get your data into R and do some stuff to it

Day 4: Data visualization

- ggplot
- heatmaps
- Lab 4: Building up a complex visualization

Day 5: Introduction to Bioconductor

- Finding and installing biocondutor packages
- Learning what packages do and how to evaluate them
- Intro to some key data structures: XStringSet, IRanges, expressionset, etc.
- Lab 5: How do I. . . ?

Day 6:

- Rapid introduction to managing and reproducing your analysis: Rmarkdown, git and github, best practices
- Writing your own functions
- Review and catchup

References