Introduction to R

Sebastian Duchene
Molecular Ecology, Evolution and Phylogenetics Laboratory
School of Biological Sciences
University of Sydney

Genetics Society of Australasia Conference 2014

Friday 11 July, 2014

Overview

R is a powerful programming language for data analysis, with a wide array of applications available for biological sciences. In this workshop we will cover the fundamentals of scientific computing in R. We will start with some basic programming and data exploration, building up to more advanced tools for data visualisation and basic statistical analysis. The workshop assumes no prior programming experience, and the material is applicable across any discipline. The attendees will gain enough knowledge to apply specific tools for their own research.

The workshop will be held in a computing lab in the Charles Perkins Centre at the University of Sydney. It will be run by Sebastian Duchene (sebastian.duchene@sydney.edu.au), with assistance from A/Prof Simon Ho (simon.ho@sydney.edu.au) and Frank Jia (fangzhi.jia@sydney.edu.au). The workshop will involve a series of lectures and tutorials. The tutorials will involve basic analyses and data manipulation using desktop PCs. We will be using free software that can be installed on PC, Mac, and UNIX platforms.

Required software

- R
- RStudio

Optional software

- TextWrangler for MacOSX.
- Notepad++ for Windows.
- Emacs Speaks Statistics for any platform.

Programme

- 09.00 – 09.15 Arrival and set-up
- 09.15 – 09.30 Lecture: Introduction to scientific computing and R
- 09.30 – 09.45 Tutorial 0: Setting up R on different operating systems
- 09.45 – 10.00 Lecture: Syntax and objects
- 10.00 – 10.30 Tutorial 1: Syntax and object types
- 10.30 – 11.00 Tutorial 2. Vectors, matrices, data frames, and indexing
- *Optional Tutorial: Lists*

- *Tea break*
- 11.15 – 12.00 Tutorial 3: Numeric and string operations
- 12.00 – 12.15 Tutorial 4: Help in R
- 12.15 - 12.30 Solutions to exercises and questions
- *Lunch break*
- 13.30 – 13.45 Lecture: Data in R
- 13.45 – 14.15 Tutorial 5: Reading and writing data
- 14.15 – 14.30 Lecture: Graphics in R
- 14.30 – 15.15 Tutorial 6: Plotting data and graphical output
- *Optional Tutorial: Custom graphics*
- *Tea break*
- 15.30 – 15.45 Lecture: Statistical tests and linear models in R
- 15.45 – 16.15 Tutorial 7: Statistical tests and linear models
- 16.15 – 16.30 Solutions to exercises and questions
- 16.30 – 17.00 Tutorial 8: Tools for data analysis

Useful references

General textbooks and references for learning R

simpler – Using R for introductory statistics Verzani (2002) CRC Press.

The R Book Crawley, M. J. (2012). The R book. John Wiley & Sons.

Online resources

- r-tutor
- statmethods
- bioconductor
- Rentrez. This is a great package to access internet data bases, such as NCBI.
- Sequence data in R. This is a tutorial and some resources to process FASTQ files and NGS output data.

Online forums

There are many good forums to find answers for specific R applications. I recommend the following for statistics and programming:

- stack overflow
- cross validated