

STAT588/BIOL588: Genomic Data Science

Lecture 1: Introduction to R

Dr. Yen-Yi Ho (hoyen@stat.sc.edu)

Objectives

- ▶ Learn to read, extract, and create datasets in R
- ▶ Learn to perform a variety of operations on datasets using R
- ▶ Learn to write your own functions/sub-routines in R

Data Analysts Captivated by R's Power



Left, Stuart Isett for The New York Times; right, Kieran Scott for The New York Times

R first appeared in 1996, when the statistics professors Robert Gentleman, left, and Ross Ihaka released the code as a free software package.

By ASHLEE VANCE

Published: January 6, 2009

To some people R is just the 18th letter of the alphabet. To others, it's the rating on racy movies, a measure of an attic's insulation or what pirates in movies say.

Related

R is also the name of a popular programming language used by a

 FACEBOOK

 TWITTER

 GOOGLE+

 SAVE

 EMAIL

R computing & graphics package

- ▶ R is a powerful, free statistical computing and graphics package.
- ▶ Popular with many researchers due to contributed packages: R functions to do specialized, advanced, & often complex statistical analysis.
- ▶ R can also do many important, routine calculations, analysis, and provide common graphical displays used in this course.
- ▶ Installed in several of the computing labs across campus, e.g. Sloan 108 & 109, Gambrell 003.
- ▶ You can download it and install it from CRAN:
<http://cran.r-project.org/>

R: Pros and Cons

Pros

- + Free
- + Available for all major platforms
- + Powerful graphics
- + Comprehensive
- + Easy interface with other languages (such as C, Fortran)
- + Well-designed programming language (object-oriented)
- + Unlimited extensibility
- + Widely used by statisticians
- + Increasingly used for genomic analyses (Bioconductor)

Cons

- No dedicated support
- Complex Syntax
- Not point-and-click
- No warranty
- Relatively slow

Bioconductor: a collection of R packages for genomic data analysis

- ▶ Started by Robert Gentleman housing R packages for genomic data analysis.

The screenshot shows the Bioconductor website homepage. The header includes the Bioconductor logo and navigation links: Home, Install, Help, Developers, and About. A search bar is located in the top right corner. The main content area is divided into several sections:

- About Bioconductor:** A paragraph describing Bioconductor as a collection of high-throughput genomic data analysis tools, developed in R, and open source. It mentions that it has two releases each year, 1200 software packages, and an active user community.
- Install:** A section titled "Get started with Bioconductor" with a list of links: Install Bioconductor, Custom packages, Get started, Latest newsletter, Follow us on Twitter, and Install R.
- Learn:** A section titled "Master Bioconductor tools" with a list of links: Courses, Support site, Package vignettes, Literature citations, Common arch files, FAQ, and Community resources.
- Use:** A section titled "Create bioinformatic solutions with Bioconductor" with a list of links: Software, Annotations, and Bioconductor packages, Amazon Machine Image, Latest release announcement, and Support site.
- Develop:** A section titled "Contribute to Bioconductor" with a list of links: Developer resources, Use the devel, DevEl Software, Annotations and Diagnostic packages, Package authors, New package submission, and Build reports.

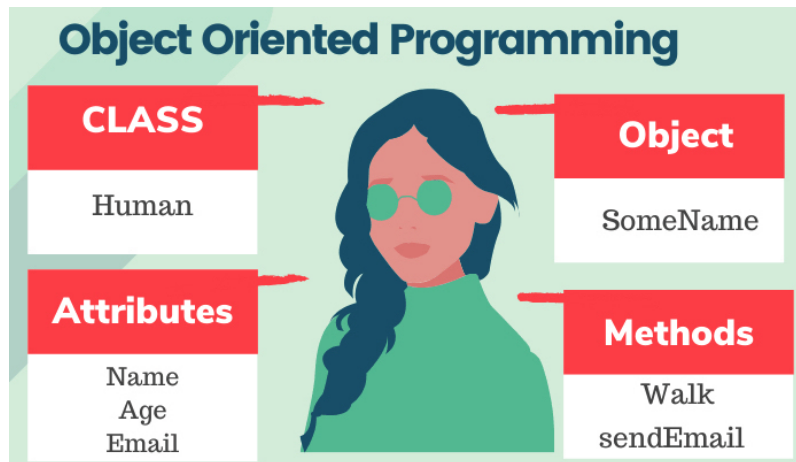
At the bottom of the page, there are sections for Support, Events, and Tweets by @Bioconductor. The Support section lists various issues and their dates. The Events section mentions the BioC2017 conference. The Tweets section shows recent tweets from the Bioconductor account.

Bioconductor installation

- ▶ Installing a specific package from Bioconductor:

```
if (!requireNamespace("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")  
BiocManager::install("LiquidAssociation")
```

Objects



R examples (Lab1.R)

- ▶ Simple objects
- ▶ Functions
- ▶ Data types

Next Lab: R Topics Outline

- ▶ Get Started
- ▶ R as a calculator
- ▶ Vectors
- ▶ Matrices, Arrays, Factors, List, Data Frame
- ▶ Import/Export Data
- ▶ R Graphics
- ▶ Random number generating
- ▶ Writing R function
- ▶ for loops
- ▶ rep, seq, which, match