# Data Carpentry July 2015

## CSE/HPCBio Data Carpentry Workshop

### Home Page

HPCBio Workshop materials are licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](http://creativecommons.org/licenses/by-sa/4.0/).

## Day 1 - Intro to R and Bioconductor

**Date:** July 13, 2015, 9 am – 4 pm  
**Instructors:** Jenny Drnevich (HPCBio), Jessica Kirkpatrick (HPCBio), Neal Davis (CSE)  
**Location:** 607 IGB

**Workshop Materials:**
- Materials for Day 1  
- R Installation guide  
- Continued Use of R and Bioconductor guide  
- R swirl guide  
- Understanding R code cheat sheet  
- CSAMA2015 Labs - Day 1  
- Morning video  
- Afternoon video

**Additional Resources:**
- Why learn R and Bioconductor:
  - [Nature Toolbox: Adventures with R](#)  
  - Why you should learn R first for data science

## Day 2 - Intro to Linux & Cluster Usage; Data Management

**Date:** July 14, 2015, 9 am – 4:30 pm  
**Instructors:** Jenny Drnevich (HPCBio), Jessica Kirkpatrick (HPCBio), Neal Davis (CSE)  
**Location:** 607 IGB

**Linux Materials:**
- Presentation  
- Linux cheat sheet  
- Getting on a UNIX-based terminal  
- Linux / R comparison

**Git Materials/Resources:**
- Data management github repository  
- Git Tutorial  
- Git cheat sheet

**Additional Linux Resources:**
- Beginner Linux - self-guided tutorials for extra practice
  - Tutorial from the book *UNIX and Perl to the RESCUE!*  
  - Parts 1 and 2 cover UNIX  
  - [http://korlab.ucdavis.edu/Unix_and_Perl.html](http://korlab.ucdavis.edu/Unix_and_Perl.html)
- Orchestrating high-throughput genomic analysis with Bioconductor

How to learn R:

- The R manuals
- swirl interactive lessons (see R homework above for guide on how to get started)
- UIUC Intro to R tutorial
- Jared Knowles' R Bootcamp
- IDRE, UCLA: Resources to help you learn and use R
- Software Carpentry's Programming with R
- John Hopkins University's MOOCs on Coursera.org:
  - The Data Scientist's Toolbox
  - R programming
- Code School: Try R (warning: free but must register; they send a lot of spam e-mails)
- Thomas Girke's R & Bioconductor Manual
- Learning Statistics with R
- R reference card (good to print out)
- Writing an R package from scratch (intermediate R users)
- RStudio's cheatsheets and online learning resources

R blogs:

- R-bloggers
- hipsteR
- RStudio blog
- R/Bioconductor blog by Sean Davis

Linux Books:

- UNIX and Perl to the RESCUE! by Keith Bradnam & Ian Knorff
- Developing Bioinformatics Computer Skills by Cynthia Gibas & Per Jambeck
- Linux Pocket Guide by Danial J. Barrett
- Beginning Shell Scripting by Eric Foster-Johnson, John C. Welch, & Micah Anderson
- Learning the bash Shell by Cameron Newham
- bash Cookbook by Carl Albing, JP Vossen, & Cameron Newham

Linux-related Software:

- PuTTY: Allows remote access from a Windows computer to a Linux based terminal.
  - http://www.putty.org/
- Mac already has a UNIX-based terminal pre-installed. You can find it in Applications -> Utilities -> Terminal

Cluster Resources:

- Biocluster help page
- Campus cluster help page
- How to submit a cluster job
- qsub script generator