Introduction Reproducible Research (RR) – Grunwald 2015-02-05

Many thanks to students or graduates Mikaela Miller, Weiming Zhang and Evan Carey for very useful materials in putting this together.

Overview

RR software produces documents that combine text, code, and output and can be re-generated easily if changes are needed. They are most easily used through Rstudio. Roughly there are two kinds of RR files in common use: .Rmd (R markdown) and .Rnw (R noweb). In this document we describe Rmarkdown since it seems to have a few advantages for our purposes, including 1) it is easier to learn and use than .Rnw, 2) it seems less platform-dependent, 3) it does not require LaTeX, and 4) it can create various formats including Word, which is easier for investigators to edit, track changes, make comments, etc. If you would like to insert equations, .Rnw with LaTeX is better.

Setting up software

Step 1.1: Download and install the statistical language **R**, from http://www.r-project.org

Step 1.2: Download and install the **Rstudio** software for conveniently running R, Rmarkdown, and knitr, from http://www.rstudio.com

Step 1.3: In R, install the Rmarkdown package: > install.packages("rmarkdown") and if prompted select a CRAN server (cloud should work fine). You need to do this only once. In some versions of Rstudio this may be included.

Step 1.4: In R, > library(rmarkdown), you need to do this each time you enter Rstudio. Again, some versions of Rstudio may not require this.

Step 1.5: To run a previously created Rmarkdown file, go to the .Rmd file in your folders outside of R and open that file. It should open Rstudio to that file. You can edit and save that under a different name to create other .Rmd files.

Rmarkdown

Good sites for help:
http://rmarkdown.rstudio.com
https://www.youtube.com/watch?v=YcJb1HBc-1Q

R packages:

Some examples and analyses use R packages. An R package is a collection of R functions that carry out particular types of analyses. They are user-submitted and stored on a CRAN server. Procedures sometimes differ but generally a package needs to be installed just once (e.g. Step 1.3 above for package rmarkdown), then each time you enter R it needs to be activated (e.g. Step 1.4 above).

Example:

The file BIOS_6622_RMarkdown_example.Rmd, adapted from one by fellow Biostatistics MS student Mikaela Miller with further additions from MS graduate Evan Carey, gives an excellent annotated example explaining the basics, giving helpful sites, etc. Thanks so much Mikaela and Evan.