Reproducible Research (RR) Course Goals

Keith Baggerly

Things we want people to take away from the course

- an understanding of how RR can help in both the short and the long run
- an understanding of the importance of RR with big data, and what can happen when it breaks down
- an appreciation that the time for planning for reproducibility is ideally when the project starts, not when it ends, and some habits that can help with this
- knowledge of how to use literate programming with R/Rstudio/knitr/rmarkdown to produce reports with code and text interleaved to produce reports in html/pdf/word which others can read, run and check as desired
- knowledge of R packages for easier sharing, reuse, and potential publication
- knowledge of some issues currently hindering replication of many studies, and how some of these points can be addressed (batch effects, p-values, big data, sanity checks)
- knowledge of how to use git for version control
- knowledge of how to share git repositories with others
- an understanding of how to structure reports to improve clarity, utility, and likely progress
- knowledge of some places to go to learn more

Background material (note live links!)

- Baggerly and Coombes (2009)
- Broman Tools for RR Course
- Gandrud 2e (2015)
- NIH Rigor and Reproducibility Guidelines (2016)
- Peng Coursera course and notes (2013)
- Retraction Watch
- Tabak and Collins (2014)
- Wickham R Packages book (2015)
- Xie 2e (2015)

These references have different goals.

Probably the best prep for the tools we hope folks will take away is Karl Broman's RR Tools course. More extensive discussions of many of these tools are given in the books by Gandrud, Peng, Xie and Wickham. Some discussions of sanity checks and breakdowns are given by Baggerly and Coombes. The NIH viewpoint and concerns are discussed by Tabak and Collins, and on the NIH Rigor and Reproducibility web page. Retraction Watch supplies more info on the sociological context.