Documentation and reproducibility with R and LaTeX

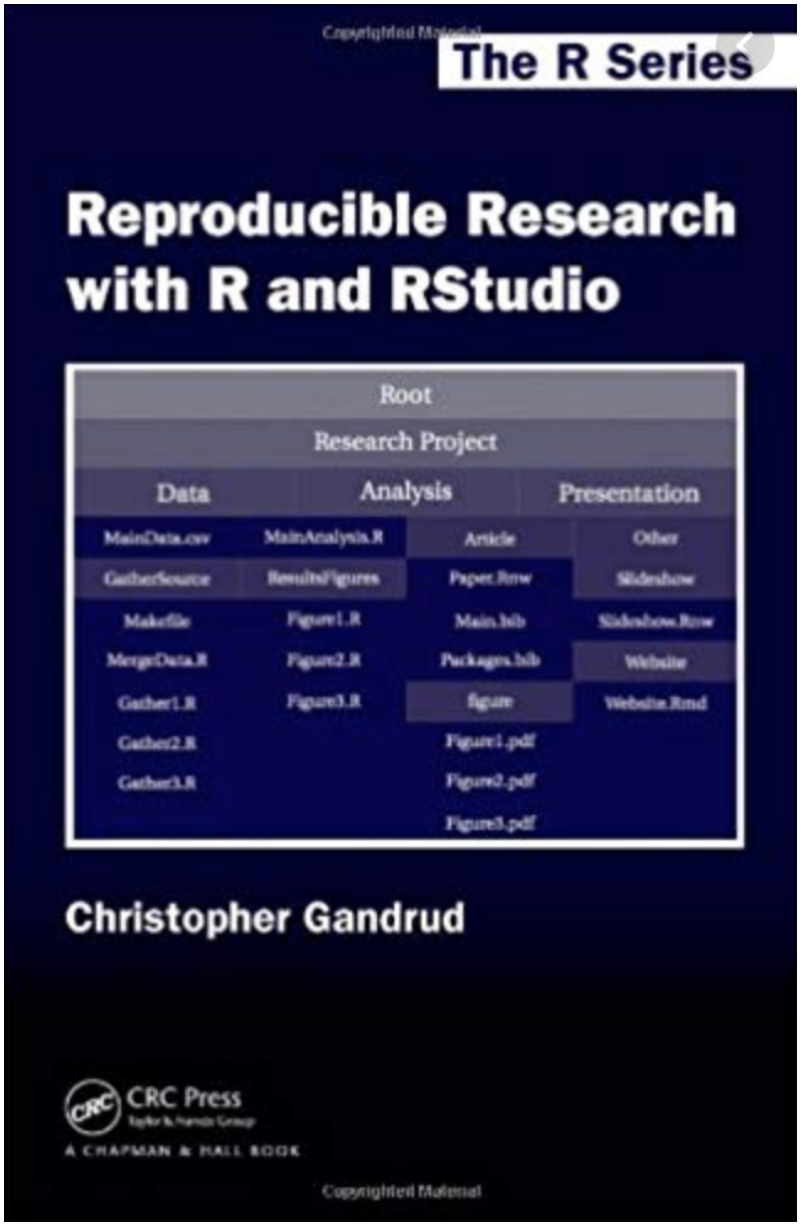
Perry J. Williams

Ecological Statistician
Department of Natural Resources and Environmental Science
University of Nevada, Reno

09/12/2020

Background & Motivation

Background & Motivation



Background & Motivation

Research is often presented in very abridged packages

- Slide shows
- Journal articles
- Books
- Web sites

Background & Motivation

These presentation documents announce a project's findings

Background & Motivation

These documents are not necissarily the research

- Sometimes considered the "advertising"

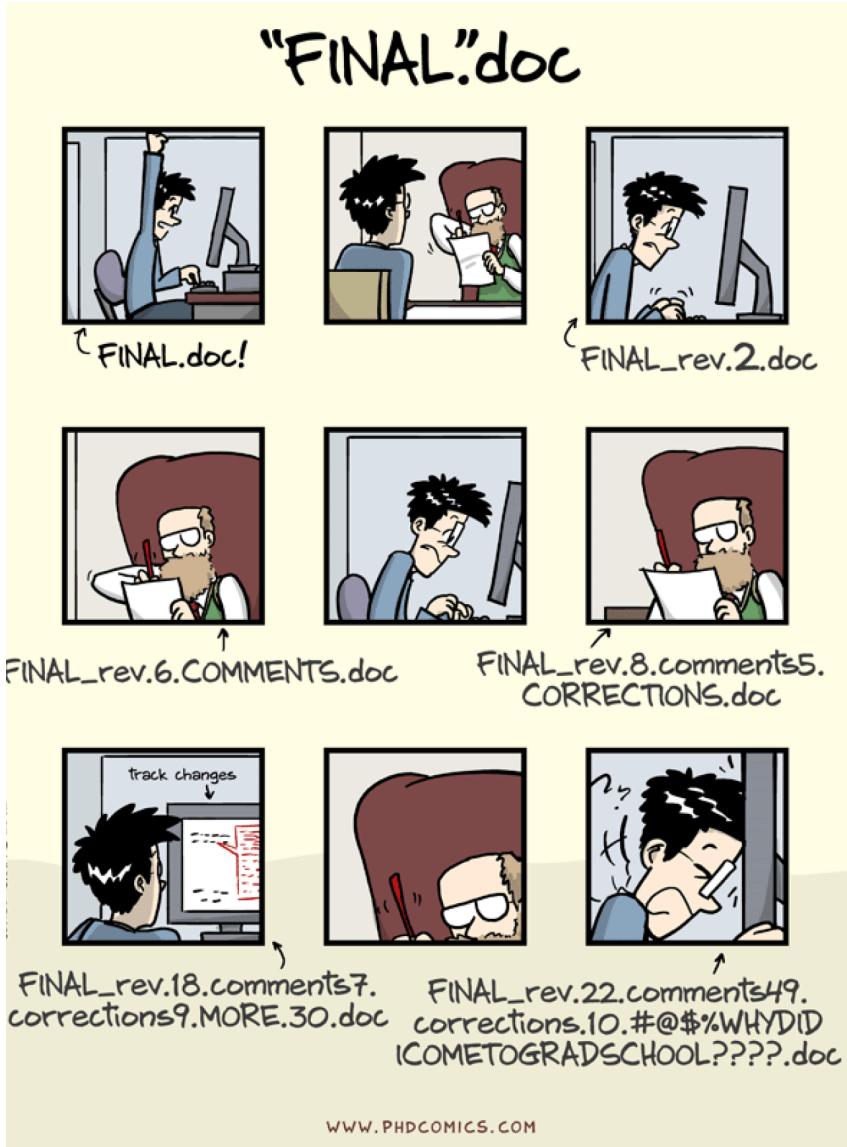
- Especially true in computational and statistical sciences

Background & Motivation

The research also includes:

- Full software environment
- Code
- Data

Background & Motivation



Background & Motivation

This workshop will introduce:

- The tools to dynamically combine research with presentation of findings,
- The `R` statistical language for data analysis,
- the LaTeX mark-up language for documents, slide shows, articles, books, and web-pages,
- the `knitr` package for `R`,
- **RStudio**, a program that brings all of these tools together in one place.

Objective

The objective of this workshop is to:

- Introduce the tools to develop a work-flow to maximize reproducible-ness, collaborations, and research impact.
- Provide templates that can be modified for your own research.

The objective of this workshop is NOT to:

- Become well-versed in `R`, `RStudio`, LaTeX, or `knitr` - that takes repetition (starting with the basic building blocks that are provided).

Background & Motivation

Additional topics include:

- Version control with Git hub,
- Data gathering,
- R markdown,
- File management,
- Projects in RStudio,
- Using LaTeXto make presentations with Beamer.

All are covered in the book: *Reproducible Research with R and RStudio*

Why R?

- Open Source and free
- Very active development community
- Interfaces with LaTeXor other mark-up languages
- Explicitly write down analyses steps as source code

Why `knitr`?

- Literate programming is a crucial part of reproducible quantitative research
- Highlights `R` code in presentation documents making it easier for readers to follow
- Provides control over inclusion of graphics
- Can cache (save output for later)

Why RStudio?

- Stand alone editor for TEXand Markdown
- Many shortcuts
- Works with C++, CSS, JavaScript, and a few other programming languages
- Integrated with version control of Git and SVN
- Simple compiling of .Rnw files
- **Easier to learn than Emacs or vi!**

What is Reproducible Research?

Research results are replicable if there is sufficient information available for independent researchers to make the same findings using the same procedures (King, 1995, 444).

What is Reproducible Research?

Research results are replicable if there is sufficient information available for independent researchers to make the same findings using the same procedures (King, 1995, 444).

In computational sciences, this means:
The data and code used to make a finding are available and they are sufficient for an independent researcher to recreate the finding.