

## 13: Summary + Extras

[bit.ly/SISBID3](https://bit.ly/SISBID3)

The most important tool is the **mindset**,  
when starting, that the end product  
will be reproducible.

– Keith Baggerly

# Steps toward reproducible research

- ▶ Slow down
- ▶ Organize; document
- ▶ Everything with code
- ▶ Scripts → RMarkdown
- ▶ Code → functions → packages
- ▶ Version control with Git
- ▶ Automation with Make
- ▶ Choose a license
- ▶ Share your work with others

# Challenges

- ▶ Daily maintenance
  - READMEs up to date?
  - Documentation match code?
- ▶ Cleaning up the junk
  - Move defunct stuff into an old/ subdirectory?
- ▶ Start over from the beginning, nicely?

# Sharing your work

- ▶ Why share?
  - Funding agency or journal requirement
  - Increased visibility
  - So that others can build on your work
- ▶ When?
  - Continuously and instantaneously
  - When you submit a paper
  - When your paper appears
- ▶ Risks?

# What to share?

- ▶ For sure
  - Primary dataset
  - Metadata
  - Data cleaning scripts
  - Analysis scripts
- ▶ It could help
  - Very-raw data
  - Processed/clean data
  - Intermediate results
- ▶ No
  - Confidential data (e.g. HIPAA data)
  - Passwords, private keys

# Where to share?

- ▶ Domain-specific repository
  - Genbank, dbGaP, etc.
  - See [re3data.org](https://re3data.org)
- ▶ Figshare, Dryad, Zenodo
- ▶ Institutional repository
- ▶ GitHub, BitBucket

# Resources

## ► R Markdown

- [rmarkdown.rstudio.com](https://rmarkdown.rstudio.com)

## ► R Packages

- Releasing to CRAN: [r-pkgs.had.co.nz/release.html](https://r-pkgs.had.co.nz/release.html)
- Leek group: [github.com/jtleek/rpackages](https://github.com/jtleek/rpackages)
- When to trust an R package: [bit.ly/trust\\_r\\_pkg](https://bit.ly/trust_r_pkg)

## ► Make

- [kbroman.org/minimal\\_make](https://kbroman.org/minimal_make)

## ► Git

- Git branches: [nicercode.github.io/git/branches.html](https://nicercode.github.io/git/branches.html)
- Hadley on Git/GitHub: [r-pkgs.had.co.nz/git.html](https://r-pkgs.had.co.nz/git.html)
- Git subtrees: [bit.ly/git\\_subtree](https://bit.ly/git_subtree)



# Some of the things we didn't cover

- ▶ Command-line, including 'R CMD BATCH'
- ▶ Software testing (and debugging)
- ▶ Code review / paired programming
- ▶ Capturing versions of dependent software (e.g. [packrat](#))
- ▶ Containers (e.g. [docker.com](#))
- ▶ Coding conventions, e.g. Hadley's [adv-r.had.co.nz/Style.html](#) and Google's [google.github.io/styleguide/Rguide.xml](#)
- ▶ Report templates, [rmarkdown.rstudio.com/developer\\_document\\_templates.html](#)