Scientific Computing & Consulting
Novartis Analytics

Using R in a Regulated Environment

Xiao Ni and Douglas Robinson
Rutgers University NCB conference
June 2019

NOVARTIS

Disclaimer

The views and opinions expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of Novartis or any of its officers.

Acknowledgement

- R Validation Hub
- Anthony Rossini
- Joe Cheng
- Max Kuhn
- Phil Bowsher

NOVARTIS

Outline

- R use cases
- Regulatory requirements
- R Validation Hub
- Good practices of reproducible research

Outline

- **R use cases**
- Regulatory requirements
- R Validation Hub
- Good practices of reproducible research

NOVARTIS

Use cases of R in Pharma

Our mission: discover and develop medicines to treat patients with serious medical conditions.

Use cases

- Data analysis
- Graphics & visualizations
- Modeling & Simulations
- Machine learning
- Rmarkdown reports
- Shiny web applications

Regulated industry

- Protect safety/privacy of human subjects
- Preserve scientific integrity
- GxP
  - Clinical research
  - GLP Tox
  - Submission

NOVARTIS

Use cases of R at the FDA

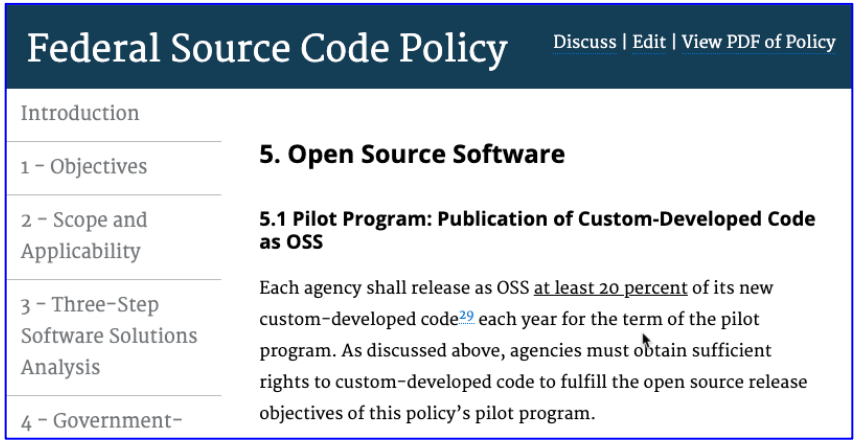
- Review of data analysis in clinical trial submissions. "Can we, on our own, replicate the conclusions of the sponsor?"
- Methodology development, innovation and evaluation
- Graphics (R graphics in labels)
- Simulations
- Shiny Application, e.g. LRT Signal Analysis for the openFDA initiative

Similar goals make R a good solution for both industry and FDA

Using R in a regulatory environment: FDA experiences. (Paul Schuette). useR! 2016 R user conference.

Federal Source Code Policy: Federal Government supports open-source

- Achieving Efficiency, Transparency, and Innovation through Reusable and Open Source Software (OSS)
- "Each agency shall release as OSS at least 20 percent of its new custom-developed code[29] each year for the term of the pilot program."



https://sourcecode.cio.gov

NOVARTIS

Open source software - a double edge sword

Pros

- Reusable code reducing duplicate work
- Transparency and better chance of finding bugs with broad access
- Support from a large community of developers
- Cultivate innovation from large community

Challenges

- No on-demand support
- Uneven quality – everyone can contribute, not following good SDLC
- Lack of documentation and (documented) unit testing
- Adding to challenges in validation

NOVARTIS

Outline

- R use cases
- **Regulatory requirements**
- R Validation Hub
- Good practices of reproducible research

NOVARTIS

Regulations

GxP for clinical research

- Good Clinical Practice (GCP)
- Good Laboratory Practice (GLP)
- Current Good Manufacturing Practice (cGMP)
- ICH E6 Good Clinical Practice Consolidated Guideline
- 21 CFR Part 11 – Electronic Records & Signatures

Software guidance

- Guidance for Industry - Computerized Systems Used in Clinical Investigations (2007)
- General Principles of Software Validation; Final Guidance for Industry and FDA Staff (2002)

Statistical guidance

- ICH E9 - Statistical Principles for Clinical Trials
- Guidance for Industry and FDA Staff - Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials (2010)

R: Regulatory Compliance and Validation Issues. A Guidance Document for the Use of R in Regulated Clinical Trial Environments. The R Foundation March 25, 2018

Regulations do NOT prevent use of R

- FDA Statistical Software Clarifying Statement
  - "FDA does not require use of any specific software for statistical analyses"
- Matt Soukup in useR! 2007
  - "Regulations do NOT prevent use of R."
  - Sponsors accountable for **validation** and documentation of R
  - Lists of validated R functions would benefit FDA reviewers and industry
  - Data driven conclusions require **accuracy**, **reliability**, and **consistency**
  - Critical need for **reproducibility** - independent of software
- FDA definition of "Validation" on next slide

FDA definition of Validation

"Establishing **documented evidence** which provides a high degree of **assurance** that a specific process will **consistently** produce a product **meeting its predetermined specifications** and quality attributes."*

- Documented evidence: traceability
- Consistently: reproducibility
- Meeting its predetermined specifications: fit-for-purpose
- Emphasis on PROCESS, what is it doing or intended to deliver as output

* Glossary Of Computerized System and Software Development Terminology

R Foundation's guidance on Base R

- Base R (14) and Recommended Packages (15) can be used for regulatory purposes, "when used in a qualified fashion".
- However obligation to enforce SOPs to
  - Ensure proper installation, validation and utilization of R
  - Mange end-user related risk

R: Regulatory Compliance and Validation Issues. A Guidance Document for the Use of R in Regulated Clinical Trial Environments. The R Foundation March 25, 2018

What about 14K+ packages on CRAN?

- Unlike Base R
  - R packages can be written by anyone
  - may not follow software development best practices
  - No testing required, <26% CRAN packages have tests
  - No obligation to maintain bug reports
  - Less used packages have less user exposure
- More packages on Github
- How can one trust these packages?

Outline

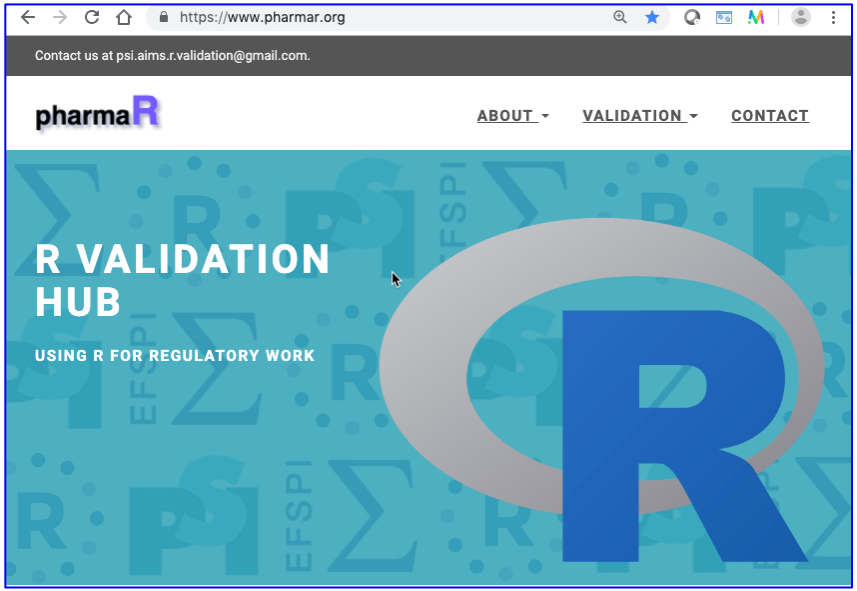
- R use cases
- Regulatory requirements
- **R Validation Hub**
- Good practices of reproducible research

R validation - a community effort

- Need harmonization/collaboration among agencies, industry, and academia (like CDISC)
- Shared sentiment at 2018 R in Pharma conference
- Led to formation of R Validation Hub, aiming to harmonize community validation efforts through
  - Risk assessment framework for R packages
  - Concerted testing framework
  - Online hub for sharing testing scripts
  - Accumulating quality evidence for R packages

R Validation Hub

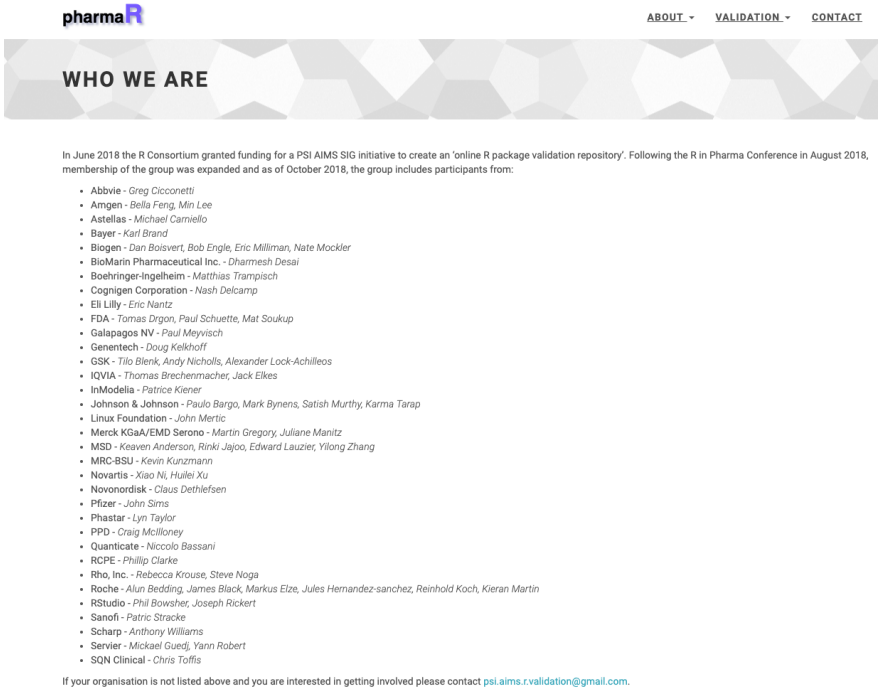
- Started from PSI AIMS SIG for online repository of R package validation
- Received funding support from R-Consortium
- Meeting monthly via TC
  - Website: http://pharmar.org
  - Github: https://github.com/pharmaR
  - Slack workspace: rvalidationhub.slack.com
  - You are welcome to join! psi.aims.r.validation@gmail.com



NOVARTIS

Organization

- Members from 30+ organizations (FDA, industry and academia)
- Executive Committee
  - Andy Nicholls @ GSK, Chair
  - Lyn Taylor @ Phastar, Secretary
  - Joe Rickert @ Rstudio, R-Consortium
  - Min Lee @ Amgen, Transcelerate
  - Yilong Zhang @ Merck, Metrics subteam lead
  - Keaven Anderson @ Merck, Testing subteam lead
- You are welcome to join!



pharmaR

ABOUT ▾ VALIDATION ▾ CONTACT

WHO WE ARE

In June 2018 the R Consortium granted funding for a PSI AIMS SIG initiative to create an 'online R package validation repository'. Following the R in Pharma Conference in August 2018, membership of the group was expanded and as of October 2018, the group includes participants from:

- **Abbvie** - *Greg Cicconetti*
- **Amgen** - *Bella Feng, Min Lee*
- **Astellas** - *Michael Carniello*
- **Bayer** - *Karl Brand*
- **Biogen** - *Dan Boisvert, Bob Engle, Eric Milliman, Nate Mockler*
- **BioMarin Pharmaceutical Inc.** - *Dharmesh Desai*
- **Boehringer-Ingelheim** - *Matthias Trampisch*
- **Cognigen Corporation** - *Nash Delcamp*
- **Eli Lilly** - *Eric Nantz*
- **FDA** - *Tomas Drgon, Paul Schuette, Mat Soukup*
- **Galapagos NV** - *Paul Meyvisch*
- **Genentech** - *Doug Kelkhoff*
- **GSK** - *Tilo Blenk, Andy Nicholls, Alexander Lock-Achilleos*
- **IQVIA** - *Thomas Brechenmacher, Jack Elkes*
- **InModelia** - *Patrice Kiener*
- **Johnson & Johnson** - *Paulo Bargo, Mark Bynens, Satish Murthy, Karma Tarap*
- **Linux Foundation** - *John Mertic*
- **Merck KGaA/EMD Serono** - *Martin Gregory, Juliane Manitz*
- **MSD** - *Keaven Anderson, Rinki Jajoo, Edward Lauzier, Yilong Zhang*
- **MRC-BSU** - *Kevin Kunzmann*
- **Novartis** - *Xiao Ni, Huilei Xu*
- **Novonordisk** - *Claus Dethlefsen*
- **Pfizer** - *John Sims*
- **Phastar** - *Lyn Taylor*
- **PPD** - *Craig McIlloney*
- **Quanticate** - *Niccolo Bassani*
- **RCPE** - *Phillip Clarke*
- **Rho, Inc.** - *Rebecca Krouse, Steve Noga*
- **Roche** - *Alun Bedding, James Black, Markus Elze, Jules Hernandez-sanchez, Reinhold Koch, Kieran Martin*
- **RStudio** - *Phil Bowsher, Joseph Rickert*
- **Sanofi** - *Patric Stracke*
- **Scharp** - *Anthony Williams*
- **Servier** - *Mickael Guedj, Yann Robert*
- **SQN Clinical** - *Chris Toffis*

If your organisation is not listed above and you are interested in getting involved please contact psi.aims.r.validation@gmail.com.

NOVARTIS

Subteams and ongoing effects

- PharmaR.org website development
- Risk assessment/metrics of R packages
- Requirements / testing framework
- White paper development

Risk assessment of R packages: risk scores based on various metrics

- Unit testing metrics
  - Code coverage: % measuring degree of source code executed by unit tests
  - Code coverage of all dependent packages
- Documentation metrics
  - Vignette, Source control, Formal bug tracking, Size of codebase
  - Website, News, Release rate, License
- Community engagement metrics
  - Author reputation, Package maturity, Avg # downloads in past 6 mo.
  - Number of active contributors, Number of authors / maintainers

R packages for metrics

- `packagemetrics` a rOpenSci package
  - https://docs.ropensci.org/packagemetrics
- `packageMetrics2` by Mango solutions
  - https://github.com/MangoTheCat/packageMetrics2
- `riskmetric` under development at R Validation Hub
  - https://github.com/pharmaR/riskmetric

```
library(packagemetrics)
library(dplyr)

dplyr_and_dt <- package_list_metrics(c("dplyr", "data.table"))
glimpse(dplyr_and_dt)
```

```
## Observations: 2
## Variables: 18
## $ package            <chr> "dplyr", "data.table"
## $ published          <chr> "2017-09-28", "2017-10-27"
## $ title              <chr> "A Grammar of Data Manipulation", "Extensio...
## $ depends_count      <int> 1, 1
## $ suggests_count     <int> 17, 18
## $ tidyverse_happy    <lgl> TRUE, FALSE
## $ has_vignette_build <lgl> TRUE, TRUE
## $ has_tests          <lgl> TRUE, TRUE
## $ reverse_count      <int> 677, 377
## $ dl_last_month      <dbl> 393520, 286717
## $ ci                 <chr> "Travis, Appveyor", "NONE"
## $ test_coverage      <chr> "CodeCov", NA
## $ forks              <dbl> 836, 650
## $ stars              <dbl> 2097, 1253
## $ watchers           <dbl> 236, 156
## $ last_commit        <dbl> 0.66666667, 0.06666667
## $ last_issue_closed  <dbl> 0.06666667, 0.10000000
## $ contributors       <dbl> 136, 43
```

R in Pharma 2019 conference

Andy Nicholls R Validation Hub (past, current and future state)

GSK R validation hub website

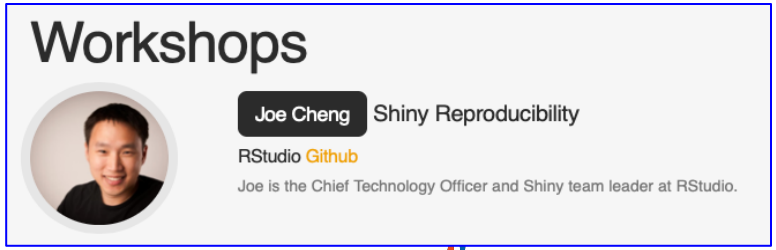
The R-Consortium in June 2018 awarded funding to create an online repository for R package validation in accordance with regulatory standards. Since the main hurdle for widespread use of R in late phase trials is ensuring adequate validation documentation, we are now focused on designing a framework which will specify a set of requirements, including metadata and examples of tests, which together would form evidence of the quality of an R package.

- http://rinpharma.com
- August 21-23, Harvard University
- August 21 pre-conference workshops

NOVARTIS

What about Shiny apps?

- Not yet discussed within the R Validation Hub
- Added challenges
  - Dynamic user interface
  - Web deployment and maintenance
  - Dependency management
- Joe Cheng 2018 R/Pharma keynote talk
- Joe's upcoming workshop at 2019 R/Pharma conference

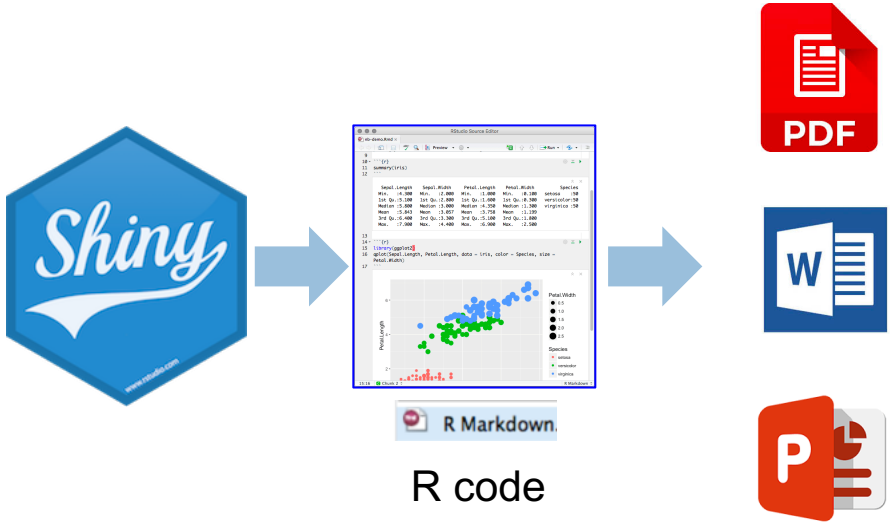


Joe's 2018 R/Pharma Demo

Features

- Save/restore states
- Generate R code
- Generate stand alone reports

Pathway towards validated output for submission



R code

https://github.com/jcheng5/rpharma-demo

What about non-clinical research?

- Some non-clinical research is non-GxP
- But NCB = **N**ever **C**ompromise **B**est practices!
- Quality principle and mindset
- Good practices lead to
  - Reproducible research (see CRAN Task view)
  - Preserving scientific integrity
  - Quality products

NOVARTIS

Outline

- R use cases
- Regulatory requirements
- R Validation Hub
- **Good practices of reproducible research**

Some good practices

- Organizing projects for reproducibility
- Use good R programming style

```
# Good          # Bad
x == y          x==y
```

- Reproducible reports through literature programming
  - Rmarkdown and Rnotebook
- Version control
  - Git and Github
- An excellent resource
  - A Guide to Reproducible Code. (2019). British Ecological Society.

Reproducible reports

- CRAN Task View: Reproducible Research
  - Tie scientific claims + data analysis + data
  - Allows reproducibility and verification by others
- Rmarkdown / Rnotebook for literate programming
  - A single document defines content and analysis
  - Blend subject matter natural language + R code

Summary

- R (and open-source tools) can be used for regulatory use
- Common uses cases make R a good choice for Pharma + FDA
- R validation poses challenges
- R Validation Hub facilitates a community effort

  Your contribution is welcome!
- Good practices lead to good science and quality
- GxP is all about documenation

References

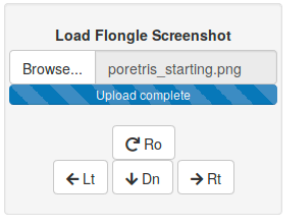
- R: Regulatory Compliance and Validation Issues. A Guidance Document for the Use of R in Regulated Clinical Trial Environments **(2018) R Foundation.**
- R validation hub, http://pharmar.org
- Use of R in Clinical Trials and Industry-Sponsored Medical Research. (2007). Marc Schwartz. useR! 2007.
- Using R in a regulatory environment: some FDA perspective. (2018). Paul Schuette, FDA.
- Using R: Perspectives of a FDA Statistical RevieweR. (2007). Mat Soukup, FDA. useR! 2007.
- A Guide to Reproducible Code. (2019). British Ecological Society.

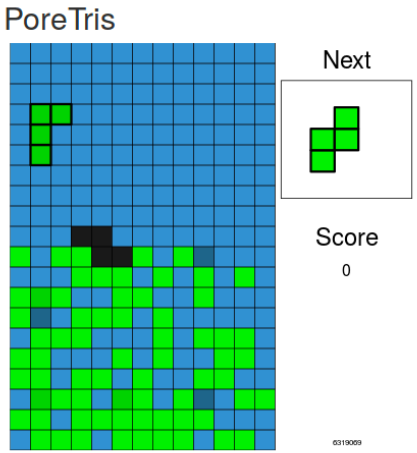
Backup

NOVARTIS

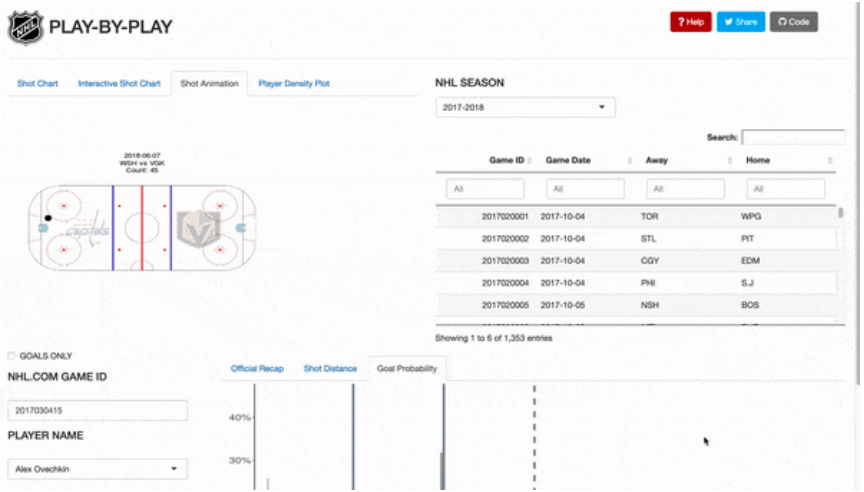
R has many more use cases

Games (link)





NHL play-by-play analysis (link)



See Shiny contest
https://blog.rstudio.com/2019/04/05/first-shiny-contest-winners/

FDA Statistical Software Clarifying Statement

"FDA **does not require use of any specific software for statistical analyses**, and statistical software is not explicitly discussed in Title 21 of the Code of Federal Regulations [e.g., in 21CFR part 11]. However, the software package(s) used for statistical analyses should be **fully documented** in the submission, including version and build identification."

As noted in the FDA guidance, *E9 Statistical Principles for Clinical Trials,* "The computer software used for data management and statistical analysis should be **reliable**, and **documentation of appropriate software testing procedures should be available**." Sponsors are encouraged to **consult with FDA review teams** and especially with FDA statisticians regarding the choice and suitability of statistical software packages at an early stage in the product development process."

https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm587506.pdf

Validation ⊃ Qualification

Validation

- In a business process context
- Assures quality for a process/product throughout its lifecycle
- *Quality assurance*: the maintenance of a desired level of quality in a service or product, especially by means of attention to every stage of the process of delivery or production

Qualification

- A phase of verification and/or testing within an overall validation program
- Verification: an activity performed during and/or between phases of the overall lifecycle
- Installation (IQ), operational (OQ) and Performance Qualification (PQ)

21 CFR part 11

- Issued by the FDA to provide regulatory requirements for processes and controls that must be applied to **electronic records** and **electronic signatures**.
- i.e. **output** from validated system

21 CFR part 11

- If there are *predicate rules that require records* to be maintained, and these *records* are managed electronically, then Part 11 controls apply to those *records*.
- If there are *predicate rules that require a signature* to be applied, and this signature is applied electronically or digitally, then Part 11 controls apply.

- "**Records** submitted to FDA, under predicate rules in electronic format [are Part 11 records]. However, a record that is *not itself submitted*, but is used in generating a *submission*, is not a part 11 record unless it is otherwise required to be maintained under a predicate rule and it is maintained in electronic format."

British Ecological Society: Reproducible Code

Contents

BRITISH ECOLOGICAL SOCIETY

Editors
Natalie Cooper, Natural History Museum, UK and **Pen-Yuan Hsing**, Durham University, UK
Contributors
Mike Croucher, University of Sheffield, UK
Laura Graham, University of Southampton, UK
Tamora James, University of Sheffield, UK
Anna Krystalli, University of Sheffield, UK
François Michonneau, University of Florida, USA

A Guide to Reproducible Code. (2019). British Ecological Society.

NOVARTIS

Good programming styles

- Google's R Style Guide
  - https://google.github.io/styleguide/Rguide.xml
- Hadley Wickham's R style guide in *Advanced R* book
  - http://adv-r.had.co.nz/Style.html
- Some useful R packages for styling and formatting
  - `lintr`
  - `styler`
  - `formatR`

Example:

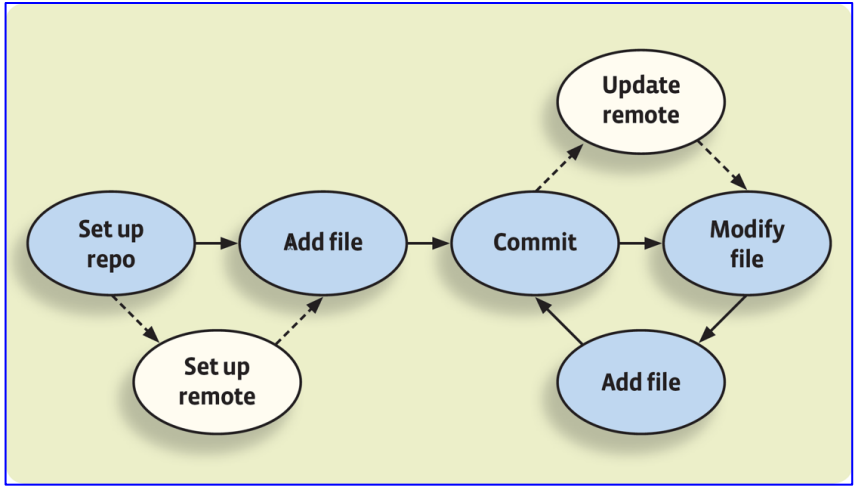
Put spaces around all infix binary operators (=, +, *, ==, &&, <-, %*%, etc.).

```
# Good
x == y

# Bad
x==y
```

Version control

- Keeping track of changes as code evolve
- Reproduce analysis through a previous version
- Facilitate collaboration within teams or contributing to open-source
- Git is a free and open-source version control system
- Github is a web-based hosting service for version control using Git.



A typical version control workflow
Source: BES Guide: Reproducible Code

NOVARTIS