

Principles and Practice of Reproducible Research with R

Author: Andrea Foulkes, Gregory Matthews, Nicholas Reich

Biostatistics in Practice: Research Training in High-Performance
Computing with R

*This material is part of the **statsTeachR** project*

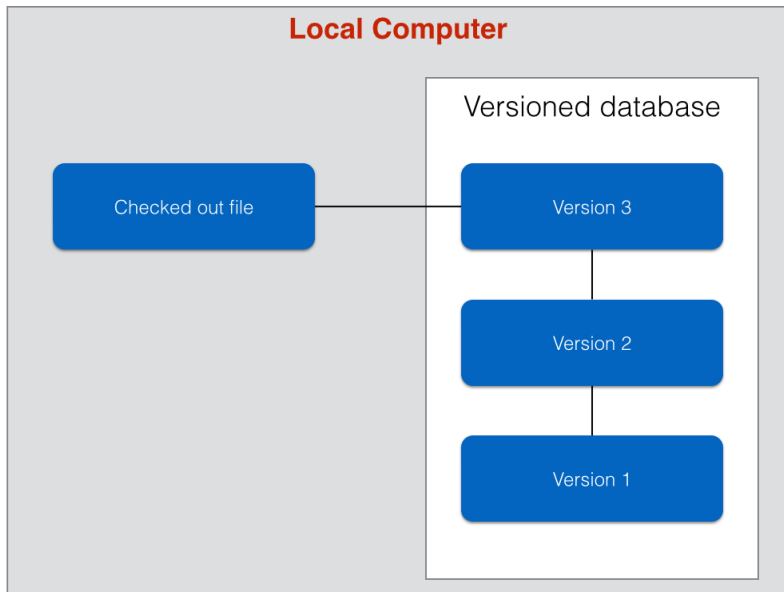
*Made available under the Creative Commons Attribution-ShareAlike 3.0 Unported
License: http://creativecommons.org/licenses/by-sa/3.0/deed.en_US*

Introduction and welcome...

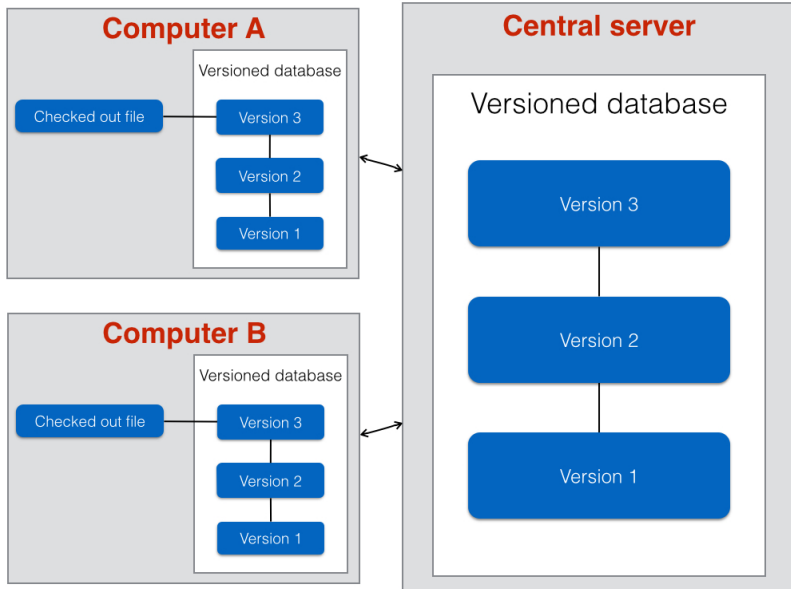
Tools for reproducible data analysis with R

- ▶ Version control: git & GitHub.com
- ▶ Dynamic documents: knitr, RMarkdown, Sweave
- ▶ RStudio
- ▶ ggplot2

Version control systems



Version control systems



Version control systems

Common VCS

- ▶ git
- ▶ subversion (svn)
- ▶ mercurial
- ▶ ...

git

Key command-line operations

- ▶ `git init`: initializes a repository locally
- ▶ `git clone`: clones a repository from a remote source (i.e. GitHub.com)
- ▶ `git add`, `git rm`: manipulating files
- ▶ `git commit`: commits changes you have made

Version control and reproducibility

Dyanamic Documents in R

- ▶ Dynamic R documents allow a user to combine text, R code, and R output, including tables and figures, into one document.
- ▶ Why is this useful?
 - ▶ Writing code and producing reports are now one document rather than many.
 - ▶ When the analysis changes, the results in the report change automatically.
 - ▶ What else?
- ▶ There are several options for how to do this.
 - ▶ R Markdown
 - ▶ Sweave
 - ▶ knitr

Dyanamic R Reports: Summary

- ▶ R Markdown: creates HTML document
- ▶ Sweave: creates pdf document AND incorporates LaTeX
- ▶ knitr: \approx Sweave + cacheSweave + pgfSweave + weaver + animation::saveLatex + R2HTML::RweaveHTML + highlight::HighlightWeaveLatex + 0.2 * brew + 0.1 * SweaveListingUtils + more

R Markdown

- ▶ R Markdown creates HTML files
- ▶ Reference:
http://www.rstudio.com/ide/docs/authoring/using_markdown
- ▶ Markdown files (.Rmd) act just like text files, except they allow a user to embed R code in chunks
- ▶ The syntax for a chunk in R Markdown:

Regular text

```
""{r}
```

Code goes here

```
""
```

Sweave

- ▶ Sweave creates pdf files as output
- ▶ Reference: <http://leisch.userweb.mwn.de/Sweave/>
- ▶ Sweave Manual: <http://www.stat.uni-muenchen.de/~leisch/Sweave/Sweave-manual.pdf>
- ▶ Sweave not only integrates R code, but also LaTeX!
- ▶ The syntax for a chunk in Sweave:

Regular text with $\$LaTeX\$$ if you want it.

```
<<OPTIONS >>=
```

```
Code goes here
```

```
@
```

Sweave: Options

- ▶ `fig=TRUE` (or `FALSE`): This indicates that the code in the chunk will print the figure to the output pdf document
- ▶ `echo=TRUE` (or `FALSE`): Should the R input code be displayed in the output pdf document
- ▶ `eval=TRUE` (or `FALSE`): Should the R input code be evaluated

knitr

- ▶ Created by Ph.D. student Yihui Xie (what have you done?)
- ▶ knitr creates pdf files as output.
- ▶ It also allows the use of LaTeX, like Sweave, whereas R Markdown does not.
- ▶ Syntax for knitr is largely the same as Sweave.
- ▶ Xie describes knitr \approx Sweave + cacheSweave + pgfSweave + weaver + animation::saveLatex + R2HTML::RweaveHTML + highlight::HighlightWeaveLatex + 0.2 * brew + 0.1 * SweaveListingUtils + more