Using  for reproducible science

Thibaut Jombart, Marie-Pauline Beugin

MRC Centre for Outbreak Analysis and Modelling
Imperial College London

Genetic data analysis with 
PR∼Statistics, Millport Field Station
20 Aug 2016

Do analyses need to be reproducible?

- "*Ideally, yes, but we don't have time for this.*"
- *"If it gets published, yes."*
- *"If it gets published, yes; unless it is in PLoS One..."*
- *"No need: I work on my own."*
- *"For others to copy us? You crazy?!"*
- *"No way! We rigged the data, the method does not work, and we ran the analyses in Excel."*

Yes but... science is about reproducible results

Do analyses need to be reproducible?

- "*Ideally, yes, but we don't have time for this.*"
- "*If it gets published, yes.*"
- "*If it gets published, yes; unless it is in PLoS One...*"
- "*No need: I work on my own.*"
- "*For others to copy us? You crazy?!*"
- "*No way! We rigged the data, the method does not work, and we ran the analyses in Excel.*"

Yes but... science is about reproducible results

Do analyses need to be reproducible?

- "*Ideally, yes, but we don't have time for this.*"
- "*If it gets published, yes.*"
- "*If it gets published, yes; unless it is in PLoS One...*"
- "*No need: I work on my own.*"
- "*For others to copy us? You crazy?!*"
- "*No way! We rigged the data, the method does not work, and we ran the analyses in Excel.*"

Yes but... science is about reproducible results

Do analyses need to be reproducible?

- "*Ideally, yes, but we don't have time for this.*"
- "*If it gets published, yes.*"
- "*If it gets published, yes; unless it is in PLoS One...*"
- "*No need: I work on my own.*"
- "*For others to copy us? You crazy?!*"
- "*No way! We rigged the data, the method does not work, and we ran the analyses in Excel.*"

Yes but... science is about reproducible results

Do analyses need to be reproducible?

- "*Ideally, yes, but we don't have time for this.*"
- "*If it gets published, yes.*"
- "*If it gets published, yes; unless it is in PLoS One...*"
- "*No need: I work on my own.*"
- "*For others to copy us? You crazy?!*"
- "*No way! We rigged the data, the method does not work, and we ran the analyses in Excel.*"

Yes but... science is about reproducible results

Do analyses need to be reproducible?

- "*Ideally, yes, but we don't have time for this.*"
- "*If it gets published, yes.*"
- "*If it gets published, yes; unless it is in PLoS One...*"
- "*No need: I work on my own.*"
- "*For others to copy us? You crazy?!*"
- "*No way! We rigged the data, the method does not work, and we ran the analyses in Excel.*"

Yes but... science is about reproducible results

Do analyses need to be reproducible?

- "*Ideally, yes, but we don't have time for this.*"
- "*If it gets published, yes.*"
- "*If it gets published, yes; unless it is in PLoS One...*"
- "*No need: I work on my own.*"
- "*For others to copy us? You crazy?!*"
- "*No way! We rigged the data, the method does not work, and we ran the analyses in Excel.*"

Yes but... science is about reproducible results

Main obstacles to reproducibility

- lack of time

Main obstacles to reproducibility

- ~~lack of time~~
  Ultimately, faster.

Main obstacles to reproducibility

- ~~lack of time~~
  Ultimately, faster.

- fear of plagiarism

Main obstacles to reproducibility

- ~~lack of time~~
  Ultimately, faster.

- ~~fear of plagiarism~~
  Low risks in practice.

Main obstacles to reproducibility

- ~~lack of time~~
  Ultimately, faster.
- ~~fear of plagiarism~~
  Low risks in practice.
- internal work, no need to share

Main obstacles to reproducibility

- ~~lack of time~~
  Ultimately, faster.
- ~~fear of plagiarism~~
  Low risks in practice.
- ~~internal work, no need to share~~
  Almost never true.

Working with others, even when you don't

Be nice to your future self!

Working with others, even when you don't

Be nice to your future self!

Two aspects of reproducible science with 

- making transparent and reproducible analyses
- sharing procedures (package development)

Two aspects of reproducible science with

- making transparent and reproducible analyses
- sharing procedures (package development)

Two aspects of reproducible science with 

- making transparent and reproducible analyses
- sharing procedures (package development)

Two aspects of reproducible science with 

- **making transparent and reproducible analyses**
- sharing procedures (package development)

Literate programming

Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to humans what we want the computer to do.

(Donald E. Knuth, Literate Programming, 1984)

Literate programming

Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to humans what we want the computer to do.

(Donald E. Knuth, Literate Programming, 1984)

Implementing literate programming for

Aim: include code inputs and outputs in a report, article, or book.

 offers several options

- **Sweave** (.Rnw): LaTeX+  → .tex document
  e.g.: `Sweave('foo.Rnw')`

- **knitr**:
  - '**Sweave**' (.Rnw) → pdf with large improvements on Sweave
    `knit2pdf('foo.Rnw')`
  - **rmarkdown** (.Rmd): markdown +  → html document
    e.g.: `knit2html('foo.Rmd')`

- **rmarkdown** (.Rmd) extends knitr to various output formats such as pdf
  e.g.: `render('foo.Rmd')`

Implementing literate programming for 

Aim: include code inputs and outputs in a report, article, or book.

 offers several options

- **Sweave** (.Rnw): LaTeX+  → .tex document
  e.g.: `Sweave('foo.Rnw')`

- **knitr**:
  - '**Sweave**' (.Rnw) → pdf with large improvements on Sweave
    `knit2pdf('foo.Rnw')`
  - **rmarkdown** (.Rmd): markdown +  → html document
    e.g.: `knit2html('foo.Rmd')`

- **rmarkdown** (.Rmd) extends knitr to various output formats such as pdf
  e.g.: `render('foo.Rmd')`

Implementing literate programming for

Aim: include code inputs and outputs in a report, article, or book.

 offers several options

- **Sweave** (.Rnw): LaTeX+  → .tex document
  e.g.: `Sweave('foo.Rnw')`

- **knitr**:
  - '**Sweave**' (.Rnw) → pdf with large improvements on Sweave
    `knit2pdf('foo.Rnw')`
  - **rmarkdown** (.Rmd): markdown +  → html document
    e.g.: `knit2html('foo.Rmd')`

- **rmarkdown** (.Rmd) extends knitr to various output formats such as pdf
  e.g.: `render('foo.Rmd')`

Rationale for Sweave

Sweave is a normal LaTeXdocument where  code is included as:

```
<<chunkTitle, ...>>=
a <- rnorm(1000)
hist(a)
@
```

Where '...' are options for the chunk
(see: `http://yihui.name/knitr/options/`)

Rationale for Rmarkdown

A .Rmd file is a normal markdown (.md) file where  code is included as:

````
```{r, chunkTitle, ...}
a <- rnorm(1000)
hist(a)
```
````

Where '...' are options for the chunk
(see: `http://yihui.name/knitr/options/`)

(Show example now)