

An introduction to R markdown for reproducible research

Stephen J Eglen
<https://sje30.github.io>
sje30@cam.ac.uk

Cambridge Computational Biology Institute
University of Cambridge
@StephenEglen

Slides: <http://bit.ly/eglen2018-rmd> (CC-BY license)

Acknowledgements

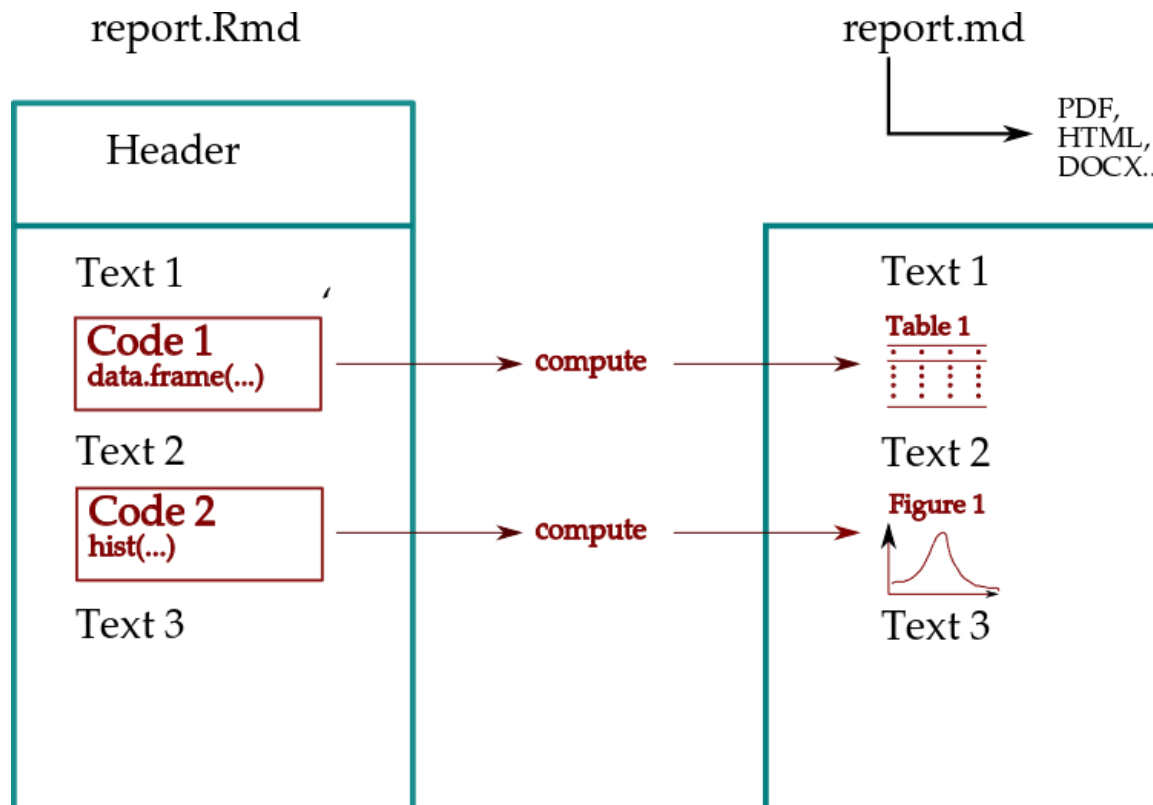
Laurent Gatto, Mike Smith.

(Last update Thu Jul 4 22:46:14 2019)

Literate programming

Donald Knuth invented the literate programming environment to combine code and documentation into one file.

In our context, we interweave R code and markdown. R code is executed and results (text, figures, tables) fed back into markdown. Markdown then translated into html (or pdf or Word or slides...).



Simple example of reproducible research

Eglen SJ (2016) Bivariate spatial point patterns in the retina: a reproducible review. *Journal de la Société Française de Statistique* 157:33–48.

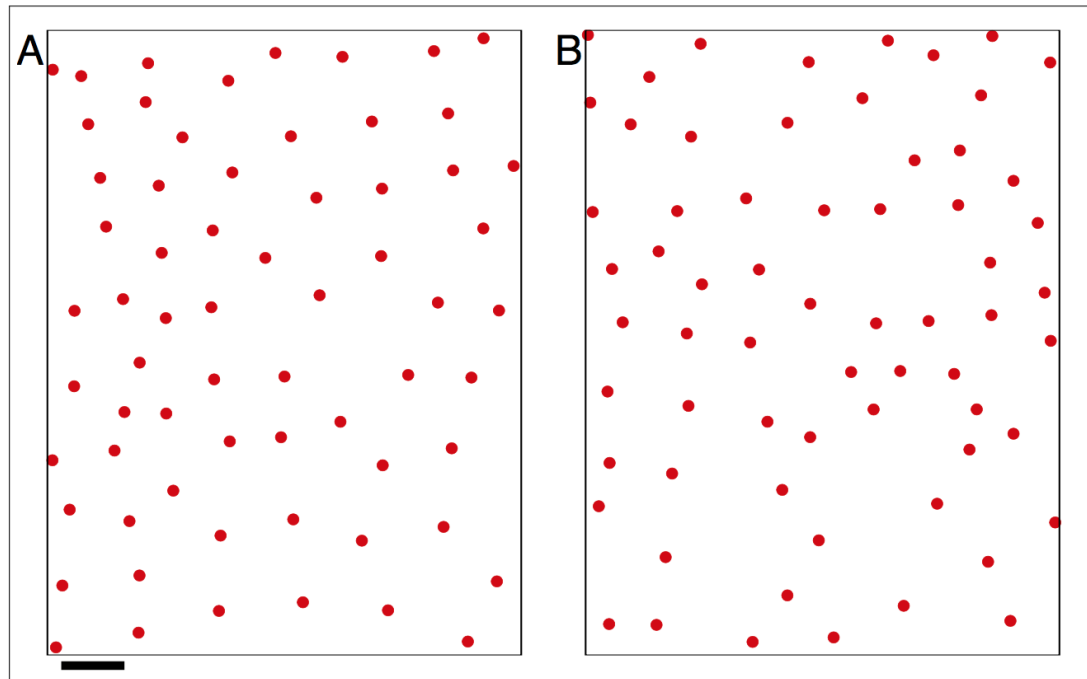


FIGURE 2. An example retinal mosaic : beta on-centre retinal ganglion cells (Wässle *et al.*, 1981). On the left is the observed map, and the right is an example univariate simulation with matching field and density of points. Scale bar is $100\text{ }\mu\text{m}$; soma are drawn to scale with a radius of $9\text{ }\mu\text{m}$.

See [paper](#) or [code](#). [Docker image](#).

See other examples at: <https://rmarkdown.rstudio.com/gallery.html>

What is markdown?

A light-weight markup language for generating HTML.

Example

Here is some `*markup text*` with `**bold**` and a
[link](http://www.rstudio.org).
Maths can be included $x^2 + y^2 = z^2$.

Example

Here is some *markup text* with **bold** and a [link](http://www.rstudio.org). Maths can be included
 $x^2 + y^2 = z^2$.

Embedding code: inline

Take a chunk like:

Simple snippets of code can be embedded like `'r sqrt(144)'` is the square root of 144.

(NB: the backtick character above.)

which makes:

Simple snippets of code can be embedded like 12 is the square root of 144.

Embedding code: chunks

Or a chunk of code

```
```{r eval=TRUE}  
n = 10
rnorm(n)
```
```

which makes this output:

```
n = 10  
rnorm(n)
```

```
## [1] 0.8227833 2.6746684 0.5456923 1.1488790 -0.5290906 0.8224489  
## [7] -0.3450447 1.5241996 0.2439829 0.9147421
```

Figures

If your code generates a figure, it is saved into `figures/` folder and included in output. (See bookdown for captions and references).

```
““{r eval=TRUE, echo=TRUE, fig.height=4}  
x = rnorm(300, mean=4, sd=2)  
hist(x)  
““
```

```
x = rnorm(300, mean=4, sd=2)  
hist(x)
```

Reproducible figures

Never mind making a whole paper reproducible, how about just one figure?

```
library(ggplot2); library(patchwork) # github.com/thomasp85/patchwork
p1 = ggplot(mtcars) + geom_point(aes(mpg, disp)) + labs(tag="A")
p2 = ggplot(mtcars) +
  geom_boxplot(aes(gear, disp, group = gear)) + labs(tag="B")
p3 = ggplot(mtcars) + geom_smooth(aes(displacement, qsec)) + labs(tag="C")
p4 = ggplot(mtcars) + geom_bar(aes(carb)) + labs(tag="D")
((p1 | p2 | p3) / p4) + theme_bw()
```


Won't this be slow?

Q: "If R is always re-running the analysis every time I edit my paper, won't this be slow?"

A: Yes. In which case, use the **cache**.

Rstudio demonstration

Create a new document with a simple structure. See header and options for compiling document.

Bookdown

Bookdown sits atop knitr and provides extra features for writing a book (captioning figures and tables, labels to refer to sections).

<https://bookdown.org/> <https://bookdown.org/yihui/bookdown/>

Challenges

- Long computation times (CODE CHECK).
- licensed software complicates everything.
- Can journals handle reproducible documents?
- When is the best time to think reproducibly?
 - Too early (explore first)
 - Too late (paper now out)?
- Technical challenges << Societal challenges

Extended example

Let's work through the document to estimate π .

You will need `estimatek.Rmd` and `refs.bib`

(References are in BiBTeX format; you can get them from google scholar, or e.g. from <https://www.bioinformatics.org/textmed/> which is pubmed with BiBTeX.)

rticles

The `rticles` package allows you to make Rmd templates for a growing number of journals, easing the adoption of formatting for journals.

<https://github.com/rstudio/rticles>

Next steps

- Material from today at: <http://github.com/sje30/2018-12-07-rmd>
- Jupyter as alternative format for notebook computation.
- Two-way Rmd <-> docx conversion: <https://noamross.github.io/redoc/>
- officedown <https://github.com/davidgohele/officedown>
- pandoc toolkit <https://pandoc.org/> drives a lot of conversion from markdown into other formats (EPUB, Word, PDF, XML...)