Reproducible dynamic documents with Rmarkdown

Francisco Rodríguez-Sánchez

https://frodriguezsanchez.net



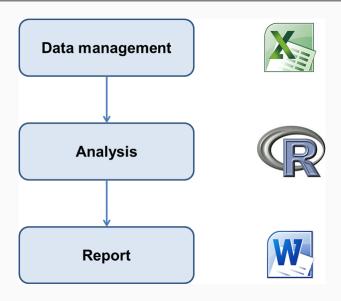
https://youtu.be/s3JldKoA0zw

A typical research workflow

- 1. Prepare data (spreadsheet)
- 2. Analyse data (R)
- 3. Write report/paper (Word)
- 4. Start the email attachments nightmare...



This workflow is broken



4

Problems of a broken workflow

- How did you do this? What analysis is behind this figure? Did you account for ...?
- What dataset was used? Which individuals were left out? Where is the clean dataset?
- Oops, there is an error in the data. Can you repeat the analysis? And update figures/tables in Word!

Manual copy-paste is tedious & problematic

'Transcribing numbers from stats software by hand was the largest source of errors'

(Eubank 2016)



≗ Follow

My rule of thumb: every analysis you do on a dataset will have to be redone 10–15 times before publication. Plan accordingly. #Rstats

Your **closest collaborator** is you 6 months ago, and you don't respond to emails.

(P. Wilson)

Even you will struggle to reproduce
your own results from a few weeks/months ago.

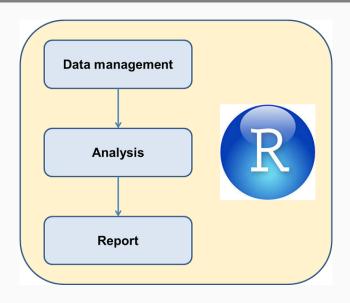
Writing reproducible manuscripts is hard

Revising non-reproducible manuscripts is even harder

•

Also, please note that because rev#1 asked to re-calculate effect sizes (...) we need to change every single number in the main text.

Dynamic reports

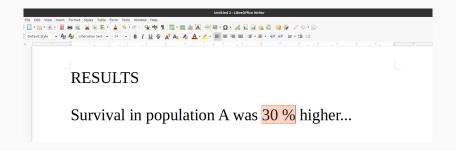


Rmarkdown documents

- · Fully reproducible (trace all results inc. tables and plots)
- · Dynamic (regenerate with 1 click)
- · Multiple outputs:
 - · documents (HTML, Word, PDF)
 - · presentations (HTML, PDF, PowerPoint)
 - · books
 - · websites...



Where does this value come from?



Dynamic documents with Rmarkdown

Rmarkdown:

Survival in population A was $r surv.diff \$ % higher

Output:

Survival in population A was 30 % higher

Dynamic documents with Rmarkdown

mydata <- read.csv('data.txt')</pre>

Rmarkdown:

We measured `r nrow(mydata)` individuals

Output:

We measured 100 individuals

Much better than copy-paste!

Rmarkdown: code (R, Python, etc) + text (Markdown)

title: "Does sunshine make people happy?" author: "FRS" output: word_document	Metadata (YAML)
## Introduction It is well known that individual well-being can be influenced by climatic condition ## Methods	ns. Text (Markdown)
<pre>"" {r echo=FALSE} ## Read data data <- read.table("data.txt", header = TRUE) # Fit linear model model <- lm(happiness ~ sunshine, data = data)</pre>	Code (R, Python)
We collected data on `r $nrow(data)$ ` individuals and fitted a linear model.	

Code chunk options

```
'''{r echo=FALSE, eval=TRUE, fig.height=3}
plot(iris)
'''
https://yihui.org/knitr/options/
```

Code chunk options

```
'''{r}
#| echo = FALSE
#| eval = TRUE
#| fig.cap = 'My figure caption'
plot(iris)
'''
```

Naming chunks helps debugging

```
processing file: test.Rmd
 1......
                                                         14%
 ordinary text without R code
                                                         29%
 1.......
label: setup (with options)
List of 1
$ include: logi FALSE
 1.........
                                                         43%
 ordinary text without R code
 1......
                                                         57%
label: read.data
 1.....
                                                         71%
 ordinary text without R code
 1.....
                                                         86%
label: plot (with options)
List of 1
$ echo: logi FALSE
Ouitting from lines 28-29 (test.Rmd)
Error in eval(predvars, data, env) : object 'specie' not found
Calls: <Anonymous> ... plot.formula -> eval -> eval -> <Anonymous> -> eval -> eval
Execution halted
```

Naming chunks helps navigating long docs

```
1 -
  2 title: "My Analysis"
  3 author: "FRS"
  4 output: html_document
  5 - - - -
  6
  7 ```{r setup, include=FALSE}
                                                           € b
     knitr::opts chunk$set(echo = TRUE)
  9 -
 10
 11
     This is an R Markdown document. Markdown is a simple
      My Analysis
                         for authoring HTML, PDF, and MS Word
      Chunk 1: setup
                        re details on using R Markdown see
      Chunk 2: read.data
                        .rstudio.com>.
      Chunk 3: plot
 12
11:60
      (Top Level) $
                                                         R Markdown 3
```

Naming chunks: figure files take chunk name

- 🔣 unnamed-chunk-1-1.png
- unnamed-chunk-1-2.png
- unnamed-chunk-1-3.png
- unnamed-chunk-1-4.png

¡Not only R! Python, Julia, C++, SQL, Stan, etc

$\verb"knitr" engines:$

[1]	"asis"	"asy"	"awk"	"bash"	"block"	"block2"
[7]	"bslib"	"c"	"cat"	"cc"	"coffee"	"comment"
[13]	"css"	"ditaa"	"dot"	"embed"	"eviews"	"exec"
[19]	"fortran"	"fortran95"	"gawk"	"go"	"groovy"	"haskell"
[25]	"highlight"	"js"	"julia"	"lein"	"mysql"	"node"
[31]	"octave"	"perl"	"php"	"psql"	"python"	"R"
[37]	"Rcpp"	"Rscript"	"ruby"	"sas"	"sass"	"scala"
[43]	"scss"	"sed"	"sh"	"sql"	"stan"	"stata"
[49]	"targets"	"tikz"	"verbatim"	"zsh"		

Markdown: easy text formatting

```
# Header
## Subheader
*italic*
**bold**
[a link](https://example.com)
Handy: https://thinkr-open.github.io/remedy/
Or use Visual Markdown Editor
```

Regenerate Word/PDF/HTML with one click



Does sunshine make people happy?

b

Introduction

Climate influences individual well-being (Rehdanz and Maddison 2005). However, \dots

Methods

We collected data on 100 individuals and fitted a linear model.

Results

We found that...

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0986	0.4271	-0.23	0.8180
sunshine	0.0101	0.0004	23.75	0.0000



Discussion

Our results confirm that happiness is related to sunshine (slope = 0.0100652).

References

Rehdanz, Katrin, and David Maddison. 2005. "Climate and Happiness." Ecological Economics 52 (1). Elsevier BV: 111-25. doi:10.1016/j.ecolecon.2004.06.015.

Spotted error in the data? No problem!

- Make changes in Rmarkdown document
- · Click Knit in Rstudio
- Report will update automatically!

Why Rmarkdown?





Darren_Dahly

- 1. Start using R Markdown to generate reports of your data analyses.
- 2. If the data changes, rerun the report with a click of the mouse.
- 3. Take 3 days off of work.
- On the 4th day, tell your collaborators that the re-analysis is complete.
 Be hailed as a hero.

o. Be nalled as a nero.

https://community.rstudio.com/t/convince-me-to-start-using-r-markdown/1636/12

Your turn

Create, edit and share Rmarkdown document

File > New File > Rmarkdown

Write text

Insert code chunks

Change chunk options (echo, eval, etc)

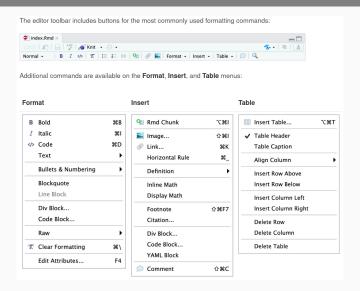
HTML/Word/PDF output

PDF generation requires LaTeX

```
library('tinytex')
install_tinytex()
```

Rmarkdown bells and whistles

'Visual Rmarkdown': Rmd as in word processor



https://rstudio.github.io/visual-markdown-editing

Automatic table generation

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0652	0.4265	-0.15	0.8789
sunshine	0.0100	0.0004	23.68	0.0000

Many alternatives: gtsummary, modelsummary, huxtable, etc

equatiomatic describes model structure

We fitted a linear model:

```
library('equatiomatic')
model <- lm(happiness ~ sunshine, data = mydata)
extract_eq(model)</pre>
```

$$happiness = \alpha + \beta_1(sunshine) + \epsilon$$
 (1)

Models that describe themselves!

```
library('report')
model <- lm(happiness ~ sunshine, data = mydata)
report(model)</pre>
```

We fitted a linear model (estimated using OLS) to predict happiness with sunshine (formula: happiness \sim sunshine). The model explains a statistically significant and substantial proportion of variance (R2 = 0.85, F(1, 98) = 560.90, p < .001, adj. R2 = 0.85). The model's intercept, corresponding to sunshine = 0, is at -0.07 (95% CI [-0.91, 0.78], t(98) = -0.15, p = 0.879). Within this model:

The effect of sunshine is statistically significant and positive (beta = 0.01, 95% CI [9.18e-03, 0.01], t(98) = 23.68,
 p < .001; Std. beta = 0.92, 95% CI [0.85, 1.00])

Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using a Wald t-distribution approximation.

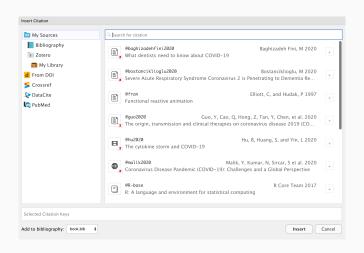
Insert equations with LaTeX

Using LaTeX:

$$y \sim N(\mu, \sigma^2)$$

Mathpix: https://github.com/jonocarroll/mathpix

Citing bibliography



https://rstudio.github.io/visual-markdown-editing/#/citations

Using BibTeX file with references

```
title: "My awesome Rmd"
output: html_document
bibliography: references.bib
```

Format bibliography for any journal

```
title: "Does sunshine make people happy?"
author: "FRS"
output: word_document
bibliography: myrefs.bib
csl: ecology-letters.csl
```

Thousands of Citation Styles:

https://www.zotero.org/styles

https://github.com/citation-style-language/styles

Rmarkdown templates

- rticles
- papaja
- rrtools
- pinp
- rmdTemplates
- pagedreport
- · GitHub!

My cool paper written in Rmarkdown

F. Rodriguez-Genchez^{1,1,2} and And Friends^{1,1}

*Some Institute of Sectioning, Department, Street, City, State, Zip; *Another University Department, Street, City, State, Zip

Please provide an absence of no more than 250 words in a single paragraph. Abstracts should explain to the general reader the major ions of the article. References in the abstract must be ched in full within the abstract lead and cleed in the east.

This PNAS journal template is provided to help you write your work in the correct journal format. Instructions for use

Note: please start year introduction without including the word "Introduction" as a section heading (except for math articles in the Physical Sciences section); this heading is implied in the first paragraphs.

Guide to using this template Please note that whilet this template provides a preview of the

typoset manuscript for solutionion, to help in this preparation, it will not necessarily be the final publication layout. For more detailed information please see the PNAS Information Author Affiliations, Include department, institution, and com-

plete address, with the ZIP/postal code, for each author. Use in the example. Authors with an ORCID ID may supply this

Submitting Manuscripts. All authors must submit their articlos at PNAScentral. If you are using Overleaf to write your article, you can use the "Subset to PNAS" option is the top but of the office window.

Formet Many authors find it sanful to preserve their manuscripts with the following order of sections; Title, Author ing are encouraged to use a language-officing service (see list at Affiliation, Keywords, Abstract, Significance Statement, Results, Discussion, Materials and methods, Acknowledgments, does not take responsibility for or endone these services, and and References. Other orders and headings are permitted. Manuscript Length, PNAS generally uses a two-column for-

mat averaging 67 characters, including spaces, per line. The pages and a PNAS PLUS research article is ten pages including all text, spaces, and the number of characters displaced by figures, tables, and equations. When submitting tables, figures, and/or equations in addition to text, keep the text for your manuscript under 20,000 characters (including spaces) for Direct Submissions and 72,000 characters (including spaces)

Returences, References should be cited in numerical order as they appear in test; this will be done automatically via bibter, u.e. (1) and (2, 3). All references, including for the SI, should sled in the main manuscript file. References appearing in both sections should not be deplicated. SI references



Fig. 1. Peopletian image of a frequent a language caption to show publication writing.

included in tables should be included with the main reference

to a published article. Where such archiving is not possible, deposition of data in public databases, such as GenBank, Ar rayEspense, Protein Data Bank, Unidata, and others outlined

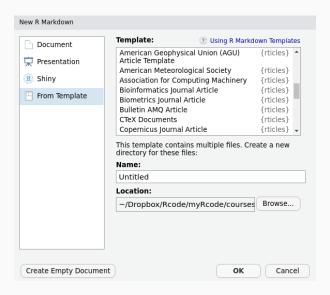
Language-Editing Services, Prior to submission, authors who their use has no Dearing on acceptance of a manuscript for

Significance Statement

Authors must submit a 120-word maximum statement about the paper itself and is required for all research papers. Place provide details of earlier problems has

FMS | September 49,000 | vol.XXX | no.XX | 4-2

Accessing Rmd templates



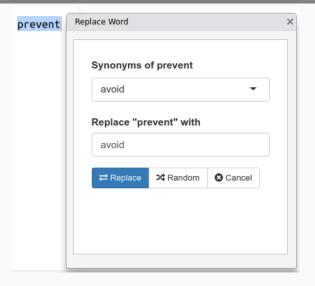
Revise writing style: gramr



https://github.com/ropenscilabs/gramr

https://github.com/nevrome/wellspell.addin

Find synonyms



https://github.com/gadenbuie/synamyn

Word count and readability

Method	koRpus	stringi
:	- :	- :
Word count	107	104
Character count	604	603
Sentence count	10	Not available
Reading time	0.5 minutes	0.5 minutes

https://github.com/benmarwick/wordcountaddin

Automated reproducibility checks

https://github.com/brandmaier/reproducibleRchunks

Here is a computation: my_sum <- x + 1

Code Chunk Reproduction Report

■ my_sum: REPRODUCTION SUCCESSFUL

Here is a computation:

my sum <- x + 1

Code Chunk Reproduction Report

Xmy_sum: REPRODUCTION FAILED Fingerprints are not identical.

BOOKDOWN

Write HTML, PDF, ePub, and Kindle books with R Markdown

The **bookdown** package is an <u>open-source R package</u> that facilitates writing books and long-form articles/reports with R Markdown. Features include:

- · Generate printer-ready books and ebooks from R Markdown documents.
- A markup language easier to learn than LaTeX, and to write elements such as section headers, lists, quotes, figures, tables, and citations.
- . Multiple choices of output formats: PDF, LaTeX, HTML, EPUB, and Word.
- · Possibility of including dynamic graphics and interactive applications (HTML widgets and Shiny apps).
- Support a wide range of languages: R, C/C++, Python, Fortran, Julia, Shell scripts, and SOL, etc.
- . LaTeX equations, theorems, and proofs work for all output formats.
- · Can be published to GitHub, bookdown.org, and any web servers.
- Integrated with the RStudio IDE.
- One-click publishing to https://bookdown.org.



https://bookdown.org/

Presentation Ninja

Yihui Xie

RStudio, PBC

https://slides.yihui.org/xaringan/

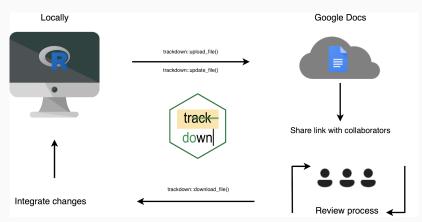
Parameterised reports

```
title: "My template report"
output: html_document
params:
  sp: Adelie
````{r}
 @ × >
library(palmerpenguins)
data("penguins")
mydata <- subset(penguins, species == params$sp)</pre>
plot(mydata$bill_length_mm, mydata$bill_depth_mm,
 main = paste0("Species: ", params$sp))
 Species: Adelie
 mydata$bill_depth_mm
 0
 0
 8
 16
 mydata$bill length mm
```

## Render thousands of individual reports from Rmd template

## Collaborative writing

- · GitHub, GitLab, etc
- · Google Docs (trackdown)
- redoc



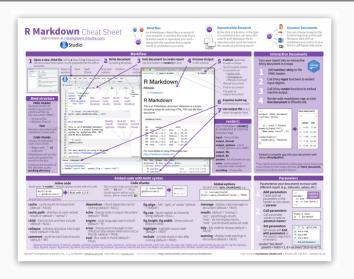
## Rmarkdown resources

#### Rmarkdown website

#### http://rmarkdown.rstudio.com/

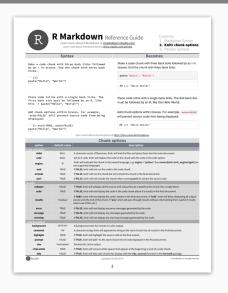


#### Rmarkdown cheat sheet



https://www.rstudio.org/links/r\_markdown\_cheat\_sheet

## Rmarkdown reference guide



https://github.com/rstudio/cheatsheets/blob/main/old/pdfs/rmarkdown-reference.pdf



by Yihui Xie, J. J. Allaire, Garrett Grolemund

2018-09-11



The first official book authored by the core R Markdown developers that provides a comprehensive and accurate reference to the R Markdown ecosystem. With R Markdown, you can easily create reproducible data analysis reports, presentations, dashboards, interactive applications, books, dissertations, websites, and journal articles, while enjoying the

Star < 239

simplicity of Markdown and the great power of R and other languages. Read more  $\rightarrow$ 

https://bookdown.org/yihui/rmarkdown/

https://bookdown.org/yihui/rmarkdown-cookbook/

## Quarto

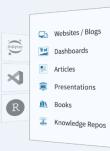
#### Quarto: 2nd generation Rmarkdown

## Welcome to Quarto

#### An open-source scientific and technical publishing system

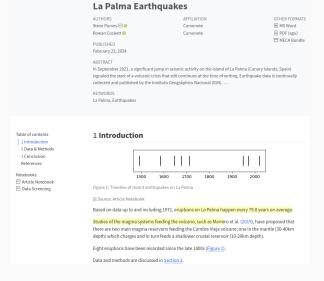
- · Author using Jupyter notebooks or with plain text markdown in your favorite editor.
- · Create dynamic content with Python, R, Julia, and Observable.
- Publish reproducible, production quality articles, presentations, dashboards, websites, blogs, and books in HTML, PDF, MS Word, ePub, and more.
- Share knowledge and insights organization-wide by publishing to Posit Connect, Confluence, or other publishing systems.
- Write using Pandoc markdown, including equations, citations, crossrefs, figure panels, callouts, advanced layout, and more.

Analyze. Share. Reproduce. You have a story to tell with data—tell it with Quarto.



https://quarto.org/

#### Quarto manuscripts



https://quarto-ext.github.io/manuscript-template-jupyter/

#### **Hundreds of Quarto extensions**

```
https://m.canouil.dev/quarto-extensions/
Journal templates:
https://quarto.org/docs/extensions/listing-journals.html
```

## Your turn

#### Your turn

- · Try visual markdown editor
- Add bibliography
- Try templates (rticles, rmdTemplates)
- Parameterised reports (e.g. different iris or penguin species)
- · Quarto manuscript