# Reproducible dynamic documents with Rmarkdown

Francisco Rodríguez-Sánchez

https://frodriguezsanchez.net
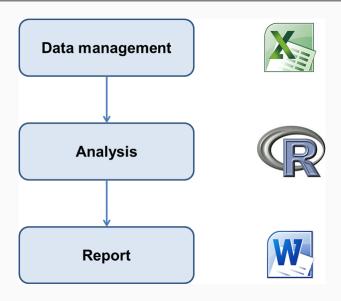
A NON REPRODUCIBLE WORKFLOW

https://youtu.be/s3JldKoA0zw

1. Prepare data (**spreadsheet**)

2. Analyse data (**R**)

3. Write report/paper (**Word**)

4. Start the email attachments
   nightmare...

- How did you do this? What analysis is behind this figure? Did you account for …?

- 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?

- 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!

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0651657  0.4264970  -0.153    0.879
sunshine     0.0100228  0.0004232  23.683  <2e-16
```

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

(Eubank 2016)

6

**Trevor A. Branch**
@TrevorABranch

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

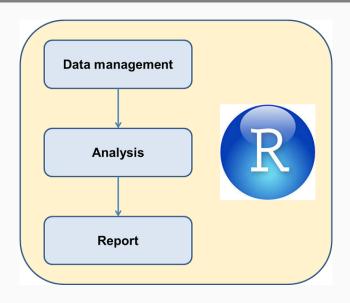**your own results** from a few weeks/months ago.

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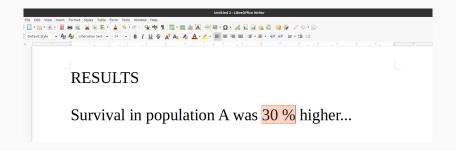
- **Fully reproducible** (trace all results inc. tables and plots)
- **Dynamic** (regenerate with 1 click)
- **Multiple outputs**:
    - documents (HTML, Word, PDF)
    - presentations (HTML, PDF, PowerPoint)
    - books
    - websites...

*Rmarkdown*:

Survival in population A was `` `r surv.diff` `` % higher

*Output*:

Survival in population A was **30** % higher

```
mydata <- read.csv('data.txt')
```

*Rmarkdown:*

> We measured `r nrow(mydata)` individuals

*Output:*

> We measured **100** individuals

> *Much better than copy-paste!*

```
---
title: "Does sunshine make people happy?"
author: "FRS"
output: word_document
---

## Introduction

It is well known that individual well-being can be influenced by climatic conditions.

## Methods

```{r echo=FALSE}
## Read data
data <- read.table("data.txt", header = TRUE)

# Fit linear model
model <- lm(happiness ~ sunshine, data = data)
```

We collected data on `r nrow(data)` individuals and fitted a linear model.
```

Metadata
(YAML)

Text
(Markdown)

Code
(R, Python...)

# Code chunk options

```r
```{r echo=FALSE, eval=TRUE, cache=TRUE, fig.height=3}
plot(iris)
```
```

https://yihui.org/knitr/options/

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

plot(iris)
```
```

```
processing file: test.Rmd
  |..........                                        |  14%
  ordinary text without R code


  |....................                              |  29%
label: setup (with options)
List of 1
 $ include: logi FALSE


  |.............................                     |  43%
  ordinary text without R code


  |......................................            |  57%
label: read.data
  |..............................................    |  71%
  ordinary text without R code


  |..............................................................|  86%
label: plot (with options)
List of 1
 $ echo: logi FALSE


Quitting 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
```

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

```
My Analysis
Chunk 1: setup
Chunk 2: read.data
Chunk 3: plot
```

11:60    (Top Level) ⇕                                    R Markdown ⇕

19

unnamed-chunk-1-1.png

unnamed-chunk-1-2.png

unnamed-chunk-1-3.png

unnamed-chunk-1-4.png

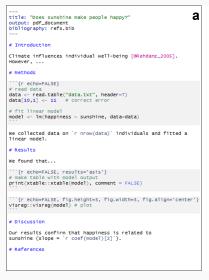knitr engines:

```
 [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"
```

```
# Header

## Subheader

*italic*

**bold**

[a link](https://example.com)
```
.

Handy: https://thinkr-open.github.io/remedy/

Or use Visual Markdown Editor

**a**

```
---
title: "Does sunshine make people happy?"
output: pdf_document
bibliography: refs.bib
---

# Introduction

Climate influences individual well-being [@Rehdanz_2005].
However, ...

# Methods

```{r echo=FALSE}
# read data
data <- read.table("data.txt", header=T)
data[10,1] <- 11    # correct error

# fit linear model
model <- lm(happiness ~ sunshine, data=data)
```

We collected data on `r nrow(data)` individuals and fitted a
linear model.

# Results

We found that...

```{r echo=FALSE, results='asis'}
# make table with model output
print(xtable::xtable(model), comment = FALSE)
```

```{r echo=FALSE, fig.height=3, fig.width=3, fig.align='center'}
visreg::visreg(model) # plot
```

# Discussion

Our results confirm that happiness is related to
sunshine (slope = `r coef(model)[2]`).

# References
```

**b**

### Does sunshine make people happy?

**Introduction**

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

**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.

23

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

🏠 🔒 **Convince me to start using R Markdown**

R Markdown    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.
4. On the 4th day, tell your collaborators that the re-analysis is complete.
5. Be hailed as a hero.

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

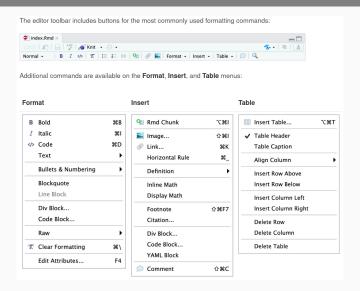Insert code chunks

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



The editor toolbar includes buttons for the most commonly used formatting commands:

Additional commands are available on the **Format**, **Insert**, and **Table** menus:

**Format**

| | | |
|---|---|---|
| **B** | Bold | ⌘B |
| *I* | Italic | ⌘I |
| </> | Code | ⌘D |
| | Text | ▶ |
| | **Bullets & Numbering** | ▶ |
| | Blockquote | |
| | Line Block | |
| | Div Block... | |
| | Code Block... | |
| | Raw | ▶ |
| ✕ | Clear Formatting | ⌘\ |
| | Edit Attributes... | F4 |

**Insert**

| | | |
|---|---|---|
| | Rmd Chunk | ⌥⌘I |
| | Image... | ⇧⌘I |
| | Link... | ⌘K |
| | Horizontal Rule | ⌘_ |
| | Definition | ▶ |
| | Inline Math | |
| | Display Math | |
| | Footnote | ⇧⌘F7 |
| | Citation... | |
| | Div Block... | |
| | Code Block... | |
| | YAML Block | |
| | Comment | ⇧⌘C |

**Table**

| | | |
|---|---|---|
| | Insert Table... | ⌥⌘T |
| ✓ | Table Header | |
| | Table Caption | |
| | Align Column | ▶ |
| | Insert Row Above | |
| | Insert Row Below | |
| | Insert Column Left | |
| | Insert Column Right | |
| | Delete Row | |
| | Delete Column | |
| | Delete Table | |

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

29

```
model <- lm(happiness ~ sunshine, data = mydata)
xtable(model)
```

|             | 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

We fitted a linear model:

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

$$\text{happiness} = \alpha + \beta_1(\text{sunshine}) + \epsilon \tag{1}$$

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

We fitted a linear model (estimated using OLS) to predict happiness with sunshine (formula: happiness ~ 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.

Using LaTeX:

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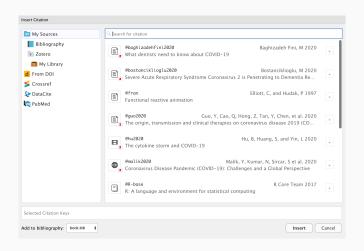
$$y \sim N(\mu, \sigma^2)$$

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

# Citing bibliography

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

```
---
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

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

# Accessing Rmd templates



New R Markdown

- Document
- Presentation
- Shiny
- **From Template**

**Template:**   ? Using R Markdown Templates

| | |
|---|---|
| American Geophysical Union (AGU) Article Template | {rticles} |
| American Meteorological Society | {rticles} |
| Association for Computing Machinery | {rticles} |
| Bioinformatics Journal Article | {rticles} |
| Biometrics Journal Article | {rticles} |
| Bulletin AMQ Article | {rticles} |
| CTeX Documents | {rticles} |
| Copernicus Journal Article | {rticles} |

This template contains multiple files. Create a new directory for these files:

**Name:**

Untitled

**Location:**

~/Dropbox/Rcode/myRcode/courses   Browse...

Create Empty Document     **OK**   Cancel

https://github.com/ropenscilabs/gramr

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

https://github.com/gadenbuie/synamyn

```
|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

https://github.com/brandmaier/reproducibleRchunks

```
15 ▾  ## Some Computation
16
17    Here is a computation:
18
19 ▾  ```{reproducibleR addition}
20    my_sum <- x + 1
21 ▴  ```
22
```

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
- ❌ my_sum: **REPRODUCTION FAILED** Fingerprints are not identical.

## BOOKDOWN

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

The **bookdown** package is an open-source R package 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 SQL, 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/

https://slides.yihui.org/xaringan/
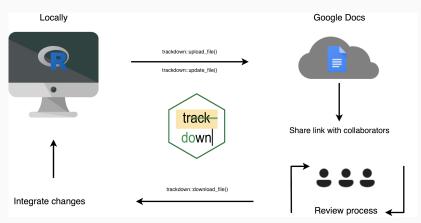
```
---
title: "My template report"
output: html_document
params:
  sp: Adelie
---

```{r}
library(palmerpenguins)
data("penguins")

mydata <- subset(penguins, species == params$sp)

plot(mydata$bill_length_mm, mydata$bill_depth_mm,
    main = paste0("Species: ", params$sp))
```
```



https://bookdown.org/yihui/rmarkdown/parameterized-reports.html

46

# Render thousands of individual reports from Rmd template

```r
library('rmarkdown')

for (i in unique(penguins$species)) {

  render('template_report.Rmd',
         params = list(sp = i))

}
```

- GitHub, GitLab, etc
- Google Docs (trackdown)
- redoc
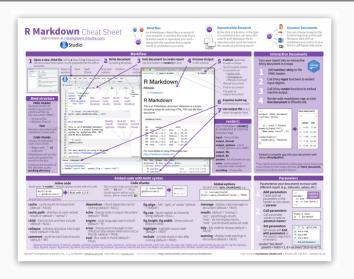
# Rmarkdown resources

http://rmarkdown.rstudio.com/
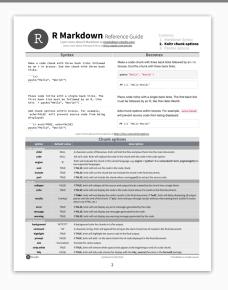
https://www.rstudio.org/links/r_markdown_cheat_sheet

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

## R Markdown: The Definitive Guide

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 simplicity of Markdown and the great power of R and other languages. *Read more →*

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

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

## Welcome to Quarto®

### An open-source scientific and technical publishing system
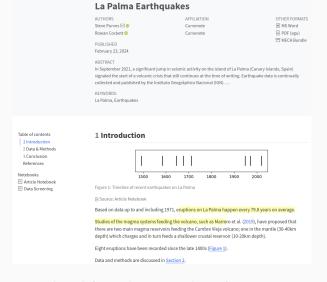
- Author using Jupyter notebooks or with plain text markdown in your favorite editor.
- Create dynamic content with Python, R, Julia, and Observable.
- Publish reproducible, production quality articles, presentations, dashboards, websites, blogs, and books in HTML, PDF, MS Word, ePub, and more.
- Share 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.**

Websites / Blogs
Dashboards
Articles
Presentations
Books
Knowledge Repos

https://quarto.org/

## La Palma Earthquakes

AUTHORS
Steve Purves ✉ ◉
Rowan Cockett ◉

AFFILIATION
Curvenote
Curvenote

OTHER FORMATS
📄 MS Word
📄 PDF (agu)
📄 MECA Bundle

PUBLISHED
February 23, 2024

ABSTRACT
In September 2021, a significant jump in seismic activity on the island of La Palma (Canary Islands, Spain) signaled the start of a volcanic crisis that still continues at the time of writing. Earthquake data is continually collected and published by the Instituto Geográphico Nacional (IGN). …

KEYWORDS
La Palma, Earthquakes

### 1 Introduction



Figure 1: Timeline of recent earthquakes on La Palma

⊞ Source: Article Notebook

Based on data up to and including 1971, eruptions on La Palma happen every 79.8 years on average.

Studies of the magma systems feeding the volcano, such as Marrero et al. (2019), have proposed that there are two main magma reservoirs feeding the Cumbre Vieja volcano; one in the mantle (30-40km depth) which charges and in turn feeds a shallower crustal reservoir (10-20km depth).

Eight eruptions have been recorded since the late 1400s (Figure 1).

Data and methods are discussed in Section 2.

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

https://quarto.org/docs/manuscripts/

55

https://m.canouil.dev/quarto-extensions/

Journal templates:
https://quarto.org/docs/extensions/listing-journals.html

# Your turn

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