

Rmarkdown

Towards reproducibility

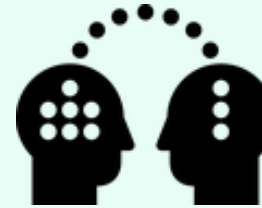


Roland Krause | rworkshop | 2021-09-10

Learning objectives

You will learn to:

- Use the **markdown** syntax
- Create **Rmarkdown** documents
- Define the output format you expect to render
- Use the interactive RStudio interface to
 - Create your documents
 - Insert R code
 - Insert bibliography
 - Build your final document



Typical flow of data

Source data →

- Experimental data
- External data sets
- Manually collected data and meta data

Intermediate →

- Derived data
- Computation
- Manual curation
- Tidy data



Analysis →

- Exploratory analysis
- Statistical models
- Hypothesis testing



Manuscript →

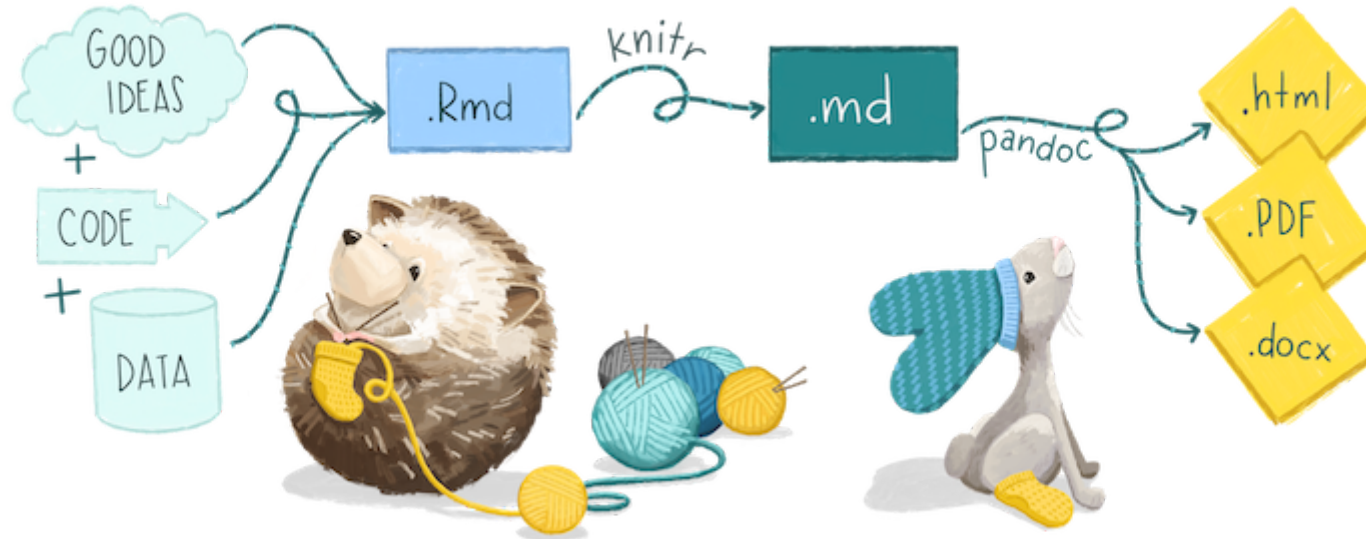
- Can you reproduce your work?
 - All numbers
 - Summaries
 - Images



One workflow

- No editing of data at any step
- All code needed to reproduce from one ingestion to manuscript coded and repeatable

Rmarkdown



Credit: Artwork by [Allison Horst](#)

Rmarkdown

Why using Rmarkdown?

- Write detailed reports
- Ensure reproducibility
- Keep track of your analyses
- Comment/describe each step of your analysis
- Export a single (Rmd) document to various formats (PDF, HTML...)
- Text file that can be managed by a version control system (like [git](#))



Rmarkdown:  +  + 

Markdown

Markdown is used to **format text**.

Markup language

- Such as [XML](#), [HTML](#)
- A coding system used to structure text
- Uses markup tags (e.g. `<h1></h1>` in [HTML](#))

HTML

```
<!DOCTYPE html>
<html>
<body>

<h1>This is a heading</h1>

<p>This is some text in a paragraph.</p>

</body>
</html>
```

Lightweight markup language

- Easy to read and write as it uses simple tags (e.g. `#`)

MD example

```
# This is a heading

This is some text in a paragraph
```

Common text formatting tags

Headers

- 6 levels are defined using #, ##, ### ...
- From BIG to small

Text style

- **bold** (****This will be bold****)
- *italic* (**This will be italic**)

Links and images

- <http://example.com> is auto-linked
- [\[description\]\(http://example.com\)](http://example.com)
- [!\[\]\(path/to/image.jpg\)](path/to/image.jpg)
- [!\[desc\]\(path/to/image.jpg\)](path/to/image.jpg) for alternative description

Verbatim code

- `code(`coding stuff`)`
- Triple backticks are delimiting code blocks:

```
```  
This is *verbatim* code
Even headers are not interpreted
```
```

rendered as:

```
This is *verbatim* code  
# Even headers are not interpreted
```

Including R code for Rmarkdown



Rmarkdown

- Extends markdown
- Place **R code** in **chunks**
- **Chunks** will be **evaluated**
- Can also handle bash; python; css; ...



Knitr

- Extracts R chunks
- Interprets them
- Formats results as markdown
- Reintegrates them into the main document (md)



Pandoc

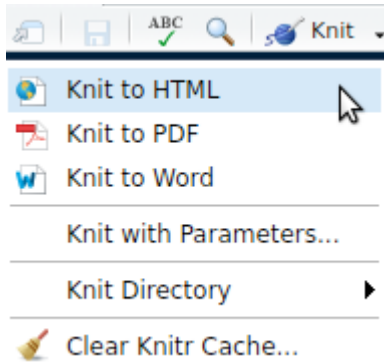
- [pandoc](#) converts markdown to the desired document (Pdf, HTML, ...)

Rmd creation: step 1

Rmd creation: step 2

Rmd creation: step 3

Generate your first HTML file



Use the **knit button** in RStudio

Rmarkdown document: Structure

```
---  
title: "Example"  
author: "John Doe"  
date: "21 September 2016"  
output: pdf_document  
---
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
```{r cars}  
summary(cars)
```
```

Including Plots

You can also embed plots, for example:

```
```{r pressure, echo=FALSE}  
plot(pressure)
```
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

YAML header

- To define document wide options
- *title, name, ...*

Markdown

- Markdown syntax to write your descriptions, remarks
- Literate programming

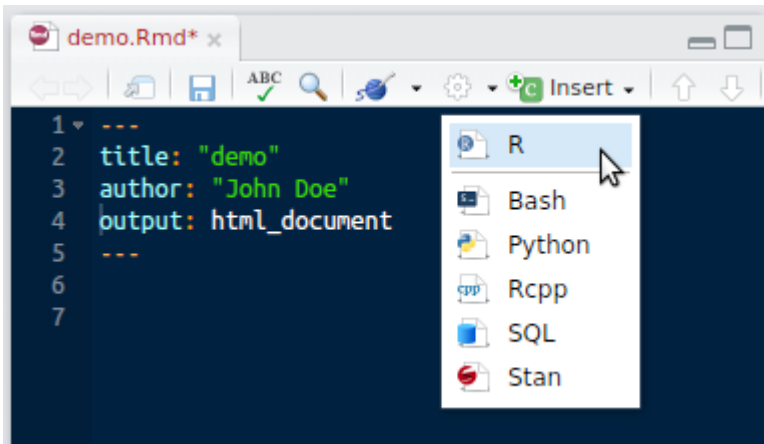
Chunks

- Code to be interpreted by R



R code chunks

Insert a chunk

shortcut: CTRL + Alt + I:



Basic Markup

- Delimited by **triple backticks tags** (`` ` ``)
- Options in **curly brackets**
 - engine evaluating the code  but also python, bash, ...
 - `` `{r}` is the minimum to define a starting  chunk
 - name of chunk (optional)
 - **show** or hide the source code (`echo = TRUE`)
 - evaluate it or **not** (`eval = FALSE`)
 - figure size (inches) ...

```
` `{r example, echo = TRUE, eval = TRUE}
# This is not markdown but R code
# Is interpreted as a comment by R (not evaluated)

# We can use R to perform calulcations:
50*10+2
` ` `
```

Navigation through chunk names

Chunk names allow you to quickly navigate code, automatically name figures, and troubleshoot errors.

- Chunk names must be **unique**. By default a numbered chunk name will be assigned

Inline R code

Integrate small pieces of *R* code

Use backticks (```) followed by the keyword `r`:

```
`r <your R code>`
```

Example

Type in `1 + 1 = `r 1 + 1`` to render **1+1=2**.

Popular output formats

HTML

- Fast rendering
- No need for extra install
- By default embeds binaries (pictures, libraries etc.)

PDF

- Single file
- Requires *L^AT_EX*, have a look at the [TinyTeX](#) package for minimal install

Word

- Widely used
- Easily editable
- **Collaborate** with people not using Rmarkdown
- Prepare **scientific manuscripts** suitable for submission

Styling your tables

If you prefer that data be displayed with additional formatting you can use the `knitr::kable` function.

Add a table caption

```
sw_db <- as_tibble(swiss, rownames = "Province") %>%
  select(Province:Catholic, -starts_with("E")) %>%
  slice_head(n = 5)

knitr::kable(sw_db ,caption = "Swiss Fertility and Socioeconomic Indicators (1888)")
```

Table: Swiss Fertility and Socioeconomic Indicators (1888)

| Province | Fertility | Agriculture | Catholic |
|--------------|-----------|-------------|----------|
| Courtelary | 80.2 | 17.0 | 9.96 |
| Delemont | 83.1 | 45.1 | 84.84 |
| Franches-Mnt | 92.5 | 39.7 | 93.40 |
| Moutier | 85.8 | 36.5 | 33.77 |
| Neuveville | 76.9 | 43.5 | 5.16 |

Specify column alignment

Change the alignment of the table columns with a vector of characters `l` (left), `c` (center), and `r` (right) or a single multi-character string for alignment

```
knitr::kable(sw_db , align = "lccc",
  caption = "Swiss Fertility and Socioeconomic Indicators (1888)")
```

Table: Swiss Fertility and Socioeconomic Indicators (1888)

| Province | Fertility | Agriculture | Catholic |
|--------------|-----------|-------------|----------|
| Courtelary | 80.2 | 17.0 | 9.96 |
| Delemont | 83.1 | 45.1 | 84.84 |
| Franches-Mnt | 92.5 | 39.7 | 93.40 |
| Moutier | 85.8 | 36.5 | 33.77 |
| Neuveville | 76.9 | 43.5 | 5.16 |

Tables

Format numeric columns

Set the maximum number of decimal places via the `digits`, which will be passed to the `round()` function

```
knitr::kable(sw_db , digits = 1)
```

| Province | Fertility | Agriculture | Catholic |
|--------------|-----------|-------------|----------|
| Courtelay | 80.2 | 17.0 | 10.0 |
| Delemont | 83.1 | 45.1 | 84.8 |
| Franches-Mnt | 92.5 | 39.7 | 93.4 |
| Moutier | 85.8 | 36.5 | 33.8 |
| Neuveville | 76.9 | 43.5 | 5.2 |

Beautiful tables with R

Several packages are being developed currently, rapidly moving field.

- [huxtable](#)
- [flextable](#)
- [xtable](#)
- [stargazer](#)
- [pander](#)
- [tables](#)
- [gt](#)
- etc.

Scientific publishing with Markdown

Equations with MathJax

- Enclose in `$` for in line equations, e.g. ``\((a^2+b^2=c^2)\)``
renders as $(a^2+b^2=c^2)$.
- Double (`$$`) for separate equations.

```
$$G_{\mu\nu}=8\pi G(T_{\mu\nu}+\rho_\Lambda g_{\mu\nu})$$
```

yields

$$G_{\mu\nu}=8\pi G(T_{\mu\nu}+\rho_\Lambda g_{\mu\nu})$$

How does it work

- No need for direct interaction with *L^AT_EX*, `pandoc` is taking care.
- Numbering is requiring bookdown pages and PDF output.
- Complex *L^AT_EX* arrangements can be used as alternatives for builtin tables.

```
$$\begin{array}{ccc}x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \end{array}$$
```

$$x_{11} \quad x_{12} \quad x_{13}$$
$$x_{21} \quad x_{22} \quad x_{23}$$

Lessons learned

Experience

Two manuscripts published, computed rendered using Rmarkdown

- Initially, we kept text and analysis code together
- Hard to organize - abstract already contains conclusions
- Eventually, all code was one big chunk in the Rmarkdown doc

Our **standard** setup

Code follows the data life cycle, e.g. using scripts to

1. **Import**
2. **Transform**
3. **Model**

Controlled by another script (e.g. using Make, or runner script)

Proposed organization

- Explore your dataset in the context of an Rmd document
- Move production code to a script, e.g. in a directory called **R**

Better: use the **targets** that extends the concept of reproducible workflows.

Bibliography

Supported formats

- Use it with your EndNote or Zotero database:
- **BibLaTeX**, **BibTeX**, **EndNote**, **EndNote XML**, MEDLINE, ISI, MODS, RIS, Copac, JSON citeproc

Styles

- uses citation style language (**csl**) files
- have a look at:
 - <https://www.zotero.org/styles>
 - <https://github.com/citation-style-language/styles>

Bibliography

How to

- setup in the yaml header
- insert citations using the [pandoc](#) syntax: \ [@citation-key]

```
---  
title: "Sample Document"  
output: html_document  
bibliography: bibliography.bib  
csl: nature.csl  
---
```

Insert your reference [@my-reference] like I did.

Tips for Zotero ☐

- install the [Better Bib\(La\)TeX](#) plugin
- **adjust the preferences** (for better integration)
- export your database as **bibtex**
- **drag and drop** pandoc keys to your Rmarkdown document

Before we stop

You learned to:

- What is [Rmarkdown](#) ([Rmd](#))
- Basic syntax of [Markdown](#)
- [knit](#) your [Rmd](#) to different output formats
- Styling tables
- Bibliography integration

Acknowledgments ☐ ☐

- Eric Koncina (initial content)
- Veronica Codoni (major overhaul)
- Xie Yihuie
- Alison Hill
- Hadley Wickham
- Artwork by [Allison Horst](#)
- Jenny Bryan

Further reading ☐

- [Rmarkdown, the definitive book](#)
- [Rmarkdown website](#)

Thank you for your attention!