

Rmarkdown for Reproducible Research

Andrew Caines | Linguistics, Cambridge | apc38

27 November 2015

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R
- markdown = plain-text authoring syntax, increasingly popular (e.g. GitHub)

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R
- markdown = plain-text authoring syntax, increasingly popular (e.g. GitHub)
- R Markdown = implementation of markdown, can take R code chunks

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R
- markdown = plain-text authoring syntax, increasingly popular (e.g. GitHub)
- R Markdown = implementation of markdown, can take R code chunks
- uses `knitr` package (successor to **Sweave**) and ‘Pandoc’ text file conversion program

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R
- markdown = plain-text authoring syntax, increasingly popular (e.g. GitHub)
- R Markdown = implementation of markdown, can take R code chunks
- uses `knitr` package (successor to **Sweave**) and ‘Pandoc’ text file conversion program
- can produce HTML, PDF and Word documents

What is R Markdown?

- source file / output file paradigm (cp LaTeX, HTML, markdown)
- plain text file (.Rmd), rendered by R
- markdown = plain-text authoring syntax, increasingly popular (e.g. GitHub)
- R Markdown = implementation of markdown, can take R code chunks
- uses `knitr` package (successor to **Sweave**) and ‘Pandoc’ text file conversion program
- can produce HTML, PDF and Word documents
- maintained by R Studio, now R Markdown v2

What is R Markdown?

- **simple syntax:** shallow learning curve

What is R Markdown?

- **simple syntax:** shallow learning curve
- **human-readable:** emphasis on communication and clarity, rather than technique and complexity

What is R Markdown?

- **simple syntax:** shallow learning curve
- **human-readable:** emphasis on communication and clarity, rather than technique and complexity
- **transparent formatting:** wysiwyg

What is R Markdown?

- **simple syntax:** shallow learning curve
- **human-readable:** emphasis on communication and clarity, rather than technique and complexity
- **transparent formatting:** wysiwyg
- **embedded computation:** share with collaborators, reviewers, colleagues, students, critics

What is R Markdown?

*... integrating source code, statistical output, and text in R Markdown is **a model of reproducibility**.*

Such transparency facilitates comprehension, defensibility, and further research or testing.

R Markdown helps to bring the vision for reproducibility in statistical analysis articulated by Gentleman & Temple Lang [2004] to reality.

- Baumer & Udwin 2015, 'R Markdown', *WIRES: Computational Statistics*

How to use R Markdown

installation

- `install.packages('rmarkdown')`

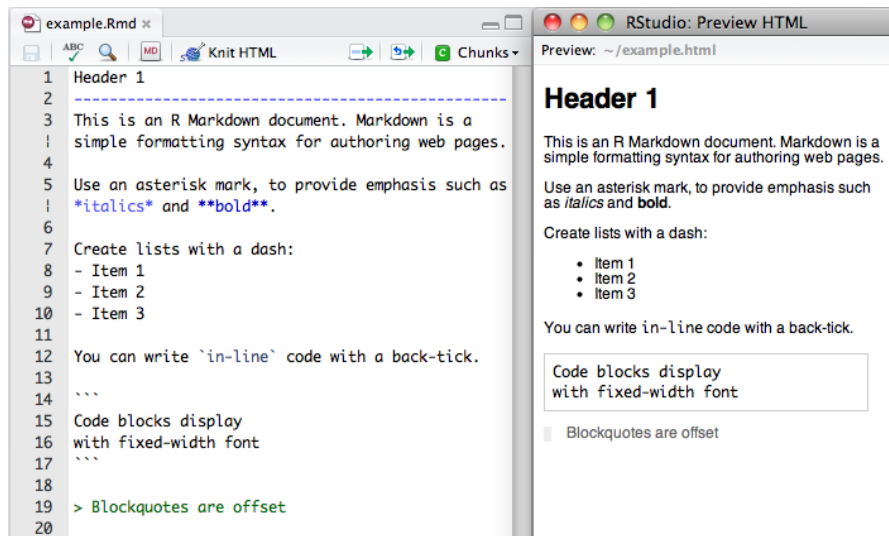
How to use R Markdown

installation

- `install.packages('rmarkdown')`
- `library(rmarkdown)`

How to use R Markdown

simple syntax, human-readable, transparent formatting



The screenshot displays the RStudio interface with a document titled 'example.Rmd' on the left and its rendered HTML preview on the right.

Left Panel (Source Editor):

```
1 Header 1
2 -----
3 This is an R Markdown document. Markdown is a
4 | simple formatting syntax for authoring web pages.
5 Use an asterisk mark, to provide emphasis such as
6 | italics and bold.
7 Create lists with a dash:
8 - Item 1
9 - Item 2
10 - Item 3
11
12 You can write `in-line` code with a back-tick.
13
14 ```
15 Code blocks display
16 with fixed-width font
17 ```
18
19 > Blockquotes are offset
20
```

Right Panel (Preview HTML):

Header 1

This is an R Markdown document. Markdown is a simple formatting syntax for authoring web pages.

Use an asterisk mark, to provide emphasis such as *italics* and **bold**.

Create lists with a dash:

- Item 1
- Item 2
- Item 3

You can write in-line code with a back-tick.

```
Code blocks display
with fixed-width font
```

Blockquotes are offset

How to use R Markdown

embedded computation

chunks.Rmd

Knit HTML

Chunks

```
1 R Code Chunks
2 =====
3
4 With R Markdown, you can insert R code
  chunks including plots:
5
6 ```{r qplot, fig.width=4, fig.height=3,
  | message=FALSE}
7 # quick summary and plot
8 library(ggplot2)
9 summary(cars)
10 qplot(speed, dist, data=cars) +
11   geom_smooth()
12 ```
13
```

RStudio: Preview HTML

Preview: ~/chunks.html Save As Publish

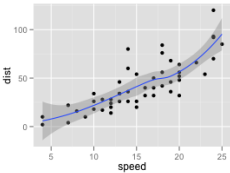
R Code Chunks

With R Markdown, you can insert R code chunks including plots:

```
# quick summary and plot
library(ggplot2)
summary(cars)
```

##	speed	dist
##	Min. : 4.0	Min. : 2
##	1st Qu.:12.0	1st Qu.: 26
##	Median :15.0	Median : 36
##	Mean :15.4	Mean : 43
##	3rd Qu.:19.0	3rd Qu.: 56
##	Max. :25.0	Max. :120

```
qplot(speed, dist, data = cars) + geom_smooth()
```



R Markdown and Reproducibility

- User experience?

R Markdown and Reproducibility

- User experience?
- disappointing :-/

R Markdown and Reproducibility

- User experience?
- disappointing :-/
- Computational linguistics >(open) Linguistics

R Markdown and Reproducibility

- User experience?
- disappointing :-/
- Computational linguistics >(open) Linguistics
- culture of Shared Tasks, for example

R Markdown and Reproducibility

- User experience?
- disappointing :-/
- Computational linguistics >(open) Linguistics
- culture of Shared Tasks, for example
- Current project: proprietary data + 4p paper + no shared code + authors moved on

R Markdown and Reproducibility

- User experience?
- disappointing :-/
- Computational linguistics >(open) Linguistics
- culture of Shared Tasks, for example
- Current project: proprietary data + 4p paper + no shared code + authors moved on
- Martijn Wieling's statistics course: Rmd, HTML

End

useful links

- RStudio: R Markdown
- Udwin & Baumer arXiv.org paper (= WIRES paper)
- Shiny & R Markdown
- my GitHub ‘replication’ repo

contact me

- apc38
- my website
- [@cainesap](<https://twitter.com/cainesap>)