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')`
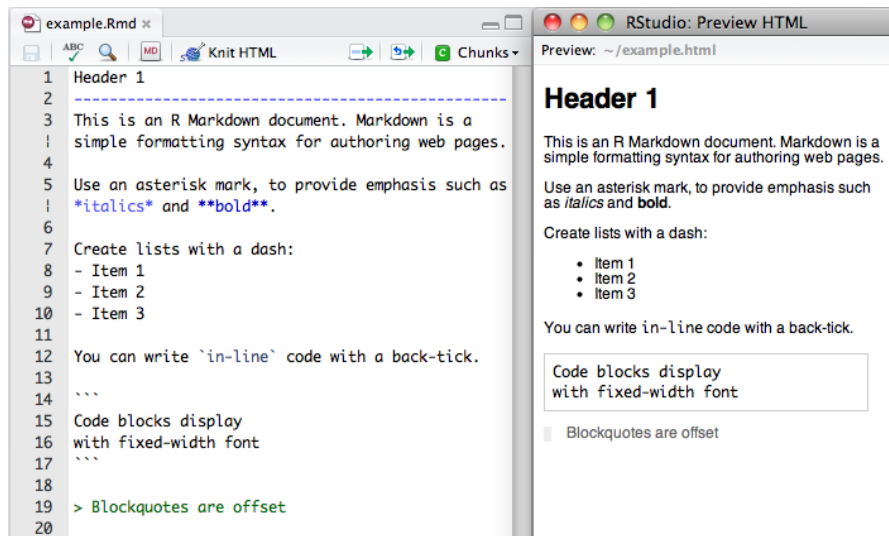
How to use R Markdown

installation

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

How to use R Markdown

simple syntax, human-readable, transparent formatting

How to use R Markdown

embedded computation
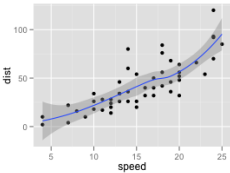
R Markdown and Reproducibility

- 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)