

# R Markdown

## Fundamentals of Computing and Data Display

---

Christoph Kern   Ruben Bach

(c.kern, r.bach)@uni-mannheim.de

# Outline

- 1 Introduction
- 2 Markdown
  - R Markdown
  - Pandoc
- 3  $\text{\LaTeX}$
- 4 Resources
- 5 References

# Introduction

## A non-reproducible workflow<sup>1</sup>

- ① Conduct analysis in statistical package
- ② Copy-and-paste results to a word processor
- ③ Update data or change pre-processing code
- ④ Start all over again

→ Separation of analysis and presentation of results

- Prone to errors
- Selective presentation of findings
- Results are **not readily reproducible**

---

<sup>1</sup>[https://www.youtube.com/watch?time\\_continue=2&v=s3JldKoA0zw](https://www.youtube.com/watch?time_continue=2&v=s3JldKoA0zw)

# Introduction

*“By reproducible research, we mean research papers with accompanying software tools that allow the reader to directly reproduce the results and employ the methods that are presented in the research paper”* (Gentleman and Lang 2007)

## Workflow Components<sup>2</sup>

- Transparency (e.g., by using dynamic documents)
- Reproducibility (e.g., by using dynamic documents)
- Efficiency (a good workflow saves you time, by automating as much of the process as possible)
- Maintainability (standardized script names, good commenting practices, README files)
- Modularity (discrete tasks into separate components (e.g. scripts))
- Portability (e.g., by using relative (not absolute) pathnames)

---

<sup>2</sup><http://blog.revolutionanalytics.com/2010/10/a-workflow-for-r.html>

# Introduction

→ Knitting source code together with presentation of results into **dynamic reproducible documents**

Some (additional) benefits...

- Avoids effort duplication
- Enables to easily update code
- Results are clearly linked to source
- Streamlines collaboration in teams
- Automates recurring tasks
- ...

# Markdown

- 1 Introduction
- 2 **Markdown**
  - R Markdown
  - Pandoc
- 3  $\text{\LaTeX}$
- 4 Resources
- 5 References

# Markdown

A lightweight markup language

- Plain text format for writing structured documents
- Easy-to-write and easy-to-read
- Originally supported only .html as output
  - Syntax description and a Perl script for converting Markdown to HTML
- Many implementations and extensions available



<https://spec.commonmark.org/>

# R Markdown

## Markdown + R code

- An authoring framework for data science
- (Pre-)installed and loaded by RStudio if needed
  - Heavily integrated into the RStudio IDE
- Implements an extension of the original markdown syntax
- Bundles other packages for extensive functionality
  - knitr, e.g. `.Rmd` → `.md`
  - pandoc, e.g. `.md` → `.html`



<https://rmarkdown.rstudio.com/>



# R Markdown

Figure: `rmarkdown::render("example.Rmd")`<sup>3</sup>



---

<sup>3</sup>Wickham and Golemund (2017)

# Text formatting

## Pandoc's Markdown<sup>4</sup>

```
<!-- Text formatting -->
*italic* or _italic_
**bold**  __bold__
`code`
superscript^2^ and subscript~2~
~~strikethrough~~
LaTeX equations inside $ $ or $$ $$

<!-- Headings -->
# 1st Level Header
## 2nd Level Header
### 3rd Level Header
```

---

<sup>4</sup><http://pandoc.org/MANUAL.html#pandocs-markdown>

# Text formatting

```
<!-- Lists -->
```

1. the first item
2. the second item
  - one unordered item

```
<!-- Links and images -->
```

```
<http://example.com>
```

```
[linked phrase](http://example.com)
```

```
![optional caption text](path/to/img.png)
```

```
<!-- Tables -->
```

```
First Header | Second Header
```

```
----- | -----
```

```
Content Cell | Content Cell
```

```
Content Cell | Content Cell
```

# Code chunks

```
```${r, options}
code
```
```

- Include R code via (self-contained) chunks
  - Embed code in text with ``r``
  - Other language engines
    - `names(knitr::knit_engines$get())`
- Chunks can be named with ````${r name}`
- Control chunk output with options
  - Set locally in chunk header
  - Set globally with `knitr::opts_chunk$set()`
  - <https://yihui.name/knitr/options/>
- Enable caching with `cache = TRUE`

# Code chunks

Table: Controlling output with chunk options<sup>5</sup>

| Option                         | Run code | Show code | Output | Plots | Messages | Warnings |
|--------------------------------|----------|-----------|--------|-------|----------|----------|
| <code>eval = FALSE</code>      | —        |           | —      | —     | —        | —        |
| <code>include = FALSE</code>   |          | —         | —      | —     | —        | —        |
| <code>echo = FALSE</code>      |          | —         |        |       |          |          |
| <code>results = "hide"</code>  |          |           | —      |       |          |          |
| <code>fig.show = "hide"</code> |          |           |        | —     |          |          |
| <code>message = FALSE</code>   |          |           |        |       | —        |          |
| <code>warning = FALSE</code>   |          |           |        |       |          | —        |

<sup>5</sup>Wickham and Golemund (2017)

# YAML header

```
---  
title: "Yet Another Markup Language"  
output: html_document  
---
```

- Specify metadata with **YAML syntax**
  - A set of tag-value pairs
  - Structured with indentation
- Set basic **document information** (title, author, date, ...)
  - <http://pandoc.org/MANUAL.html#variables-set-by-pandoc>
- Declare **parameters** with `params` field
  - Accessible with `params$name` in code chunks
- Set the default **output format**

## YAML header

Table: Output formats<sup>6</sup>

| Output value                       | creates              |
|------------------------------------|----------------------|
| <code>html_notebook</code>         | R Notebook           |
| <code>html_document</code>         | HTML                 |
| <code>pdf_document</code>          | PDF                  |
| <code>word_document</code>         | Microsoft Word       |
| <code>odt_document</code>          | OpenDocument Text    |
| <code>rtf_document</code>          | Rich Text Format     |
| <code>md_document</code>           | Markdown             |
| <code>ioslides_presentation</code> | ioslides HTML slides |
| <code>slidy_presentation</code>    | slidy HTML slides    |
| <code>beamer_presentation</code>   | Beamer pdf slides    |
| ...                                | ...                  |

<sup>6</sup><https://rmarkdown.rstudio.com/formats.html>

# YAML header

## Output options

- Each output format is associated with a set of options
- For `html_document`, see `?rmarkdown::html_document`
  - `toc`
  - `highlight`
  - `df_print`
  - `theme`
  - ...
- `pdf_document`
  - `toc`
  - `keep_tex`
  - ...



# Citations

## ① Set bibliography file in YAML header

- Various formats supported, e.g. BibTeX, endnote, ...
- Control style with a CSL file (optional)
- Alternative: Use inline references with YAML's `references` field

```
---  
bibliography: refs.bib  
csl: style.csl  
---
```

## ② Use citation keys in text

```
Smith cited [@smith04].  
Smith cited without author [-@smith04].  
@smith04 cited in line.
```

## ③ Render

# More formats, templates, themes

- bookdown
  - <http://www.bookdown.org>
- prettydoc
  - <https://github.com/yixuan/prettydoc/>
- rticles
  - <https://github.com/rstudio/rticles>
- rmdformats
  - <https://github.com/juba/rmdformats>

# Pandoc

<https://pandoc.org/>

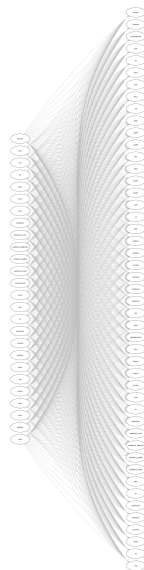
- Converter used by rmarkdown to create output file
- Can be run on its own from your OS's terminal
- Numerous markup and document formats supported

## # Pandoc examples

```
pandoc -s -o output.html input.txt
```

```
pandoc -f markdown -t latex file.txt
```

```
pandoc -f html -t markdown hello.html
```



- 1 Introduction
- 2 Markdown
  - R Markdown
  - Pandoc
- 3 L<sup>A</sup>T<sub>E</sub>X**
- 4 Resources
- 5 References

- L<sup>A</sup>T<sub>E</sub>X is a markup language (like Markdown, HTML, ...)
- Allows to create high-quality documents
- Takes some time to learn
- The typical workflow
  - ① Create L<sup>A</sup>T<sub>E</sub>X source code file (.tex)
  - ② Compile it via L<sup>A</sup>T<sub>E</sub>X or pdfL<sup>A</sup>T<sub>E</sub>X
  - ③ Use a viewer (PDF, DVI) to view the compiled file
- Needs a installed L<sup>A</sup>T<sub>E</sub>X-distribution (e.g. MikT<sub>E</sub>X, T<sub>E</sub>X Live, TinyT<sub>E</sub>X)
  - Also needed to generate **PDF output from a .Rmd file**

# L<sup>A</sup>T<sub>E</sub>X example

*% Preamble*

```
\documentclass[11pt, a4paper]{scrartcl}
\usepackage[utf8]{inputenc}
\usepackage[T1]{fontenc}
\usepackage[english]{babel}
```

```
\title{Title}
\author{Author}
```

*% Body*

```
\begin{document}
```

```
\maketitle
```

```
\section{Heading}
```

```
A \LaTeX{} example document.
```

```
\begin{equation}
\mathcal{T}(x; \Theta) = \sum_{m=1}^M \gamma_m \mathbb{I}\{(x \in \tau_m)\}.
\end{equation}
```

```
\end{document}
```

## Assignment 1

TST-101 – Test course

The name

July 22, 2018

# Sweave

*“Sweave is a tool that allows to embed the R code for complete data analyses in latex documents. The purpose is to create dynamic reports, which can be updated automatically if data or analysis change.”<sup>7</sup>*

## R + L<sup>A</sup>T<sub>E</sub>X

- ① Create a file.Rnw
- ② Use L<sup>A</sup>T<sub>E</sub>X as in .tex files
- ③ Include R code chunks

```
<>=
R-code
@
```

- ④ Create .tex and PDF with Sweave('file.Rnw')
  - `help("Sweave", package = "utils")`

<sup>7</sup><https://leisch.userweb.mwn.de/Sweave/>



# Resources

- R Markdown template paper
  - Bauer, P. C. (2018). Writing a Reproducible Paper in R Markdown.  
<http://dx.doi.org/10.2139/ssrn.3175518>
- R Task View
  - <https://cran.r-project.org/web/views/ReproducibleResearch.html>
- GitHub Pages
  - <https://pages.github.com/>
- Cheatsheets
  - <https://github.com/rstudio/cheatsheets/raw/master/rmarkdown-2.0.pdf>
  - <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>

# References

- Gandrud, C. (2015). *Reproducible Research With R and R Studio*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Gentleman, R. and Lang, D. (2007). Statistical Analyses and Reproducible Research. *Journal of Computational and Graphical Statistics* 16(1), 1–23.
- Wickham, H. and Grolemund, G. (2017). *R for Data Science*. Sebastopol, CA: O'Reilly.
- Xie, Y., Allaire, J.J., Grolemund, G. (2018). *R Markdown: The Definitive Guide*. Boca Raton, FL: Chapman & Hall/CRC.
- Xie, Y. (2016). *bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, FL: Chapman & Hall/CRC.
- Xie, Y. (2015). *Dynamic Documents with R and knitr*. Boca Raton, FL: Chapman & Hall/CRC.