Digital Tools for Reproducible Research

Materials https://bolibaugh.github.io/DigitalTools/ https://osf.io/jrxyw/

Contact cylcia.bolibaugh@york.ac.uk | @CBolibaugh

Research increasingly reliant on computational and data skills...

“Some other time”

Too many options

Inefficient to learn on your own

Shaming?

Thanks to Rochelle Terman for the motivation.

Research increasingly reliant on computational and data skills...

"Some other time"

Reproducible workflows (including code) save time once set up

"Too many options"

Learn the logic now, specialise later

Inefficient to learn on your own

True -- do it here

Shaming? Everyone started somewhere; very few are experts

Week 5 Reproducible research 15 May 2019

Week 6 Preregistration 22 May 2019

Week 8 Open data 5 June 2019

Week 9 Reproducible analyses, power analysis and simulation 12 June 2019

Week 10 Writing a reproducible manuscript 19 June 2019

Thanks to Rochelle Terman for the motivation.

05 Reproducible manuscripts

Today

Reproducible documents, Rmarkdown syntax, and output

Learn about reproducible documents, Rmarkdown syntax, and the Papaja package for APA manuscripts.

Tasks

1. Learn about what is meant by reproducible documents, and Rmark
2. Understand basic Rmarkdown formatting, syntax, and output options
3. Create a manuscript squib in Papaja, reporting an independent samples t-test.

By the end of the today, you should be able to create and format an Rmarkdown report, and have a draft template for an APA manuscript.

What is a reproducible manuscript?

Code, results, and text in a single document.

Results and plots are automatically generated from your data, so documents are updated if data changes

A bit like magic :)

@CBolibaugh

Code, results, & text in a single document.

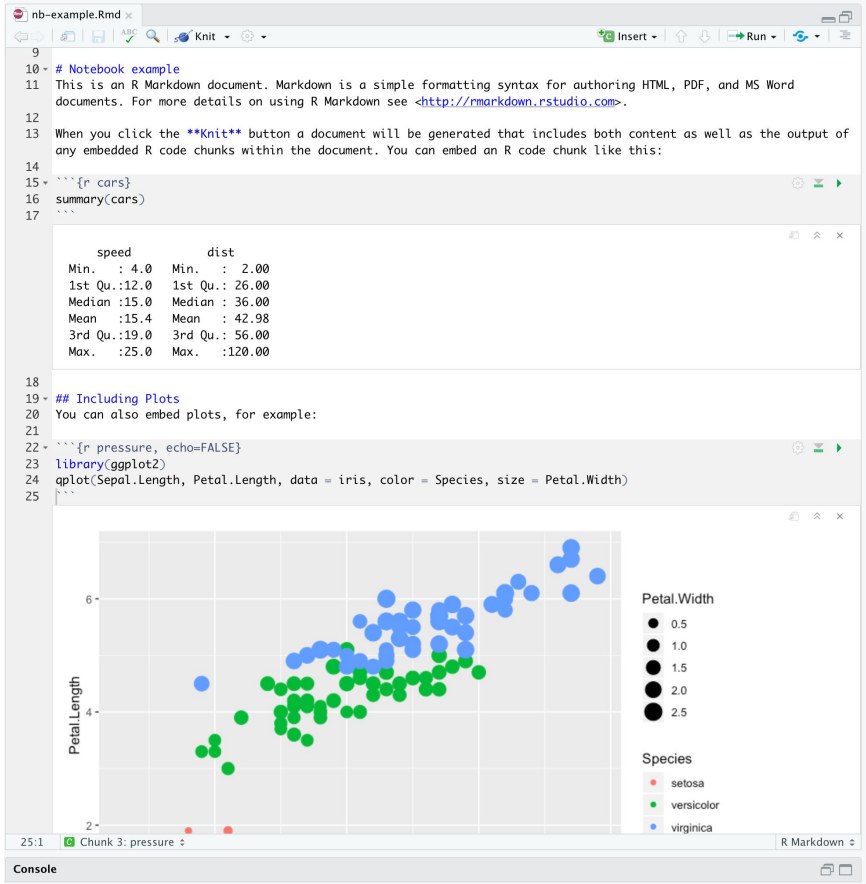
Every analysis I've ever done I've had to repeat, sometimes years later. It saves time in the long run to invest in making a reproducible analysis first time around.
@tomstafford

@CBolibaugh

Code, results, & text in a single document.

Every analysis I've ever done I've had to repeat, sometimes years later. It saves time in the long run to invest in making a reproducible analysis first time around.

@tomstafford



@CBolibaugh

````
---
title: "Notebook example"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

# Notebook example
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}
library(ggplot2)
qplot(Sepal.Length, Petal.Length, data = iris, color = Species,
size = Petal.Width)
```
````

Rmarkdown is a plain text file, with .rmd extension

@CBolibaugh

Rmarkdown YAML metadata

Defaults given:

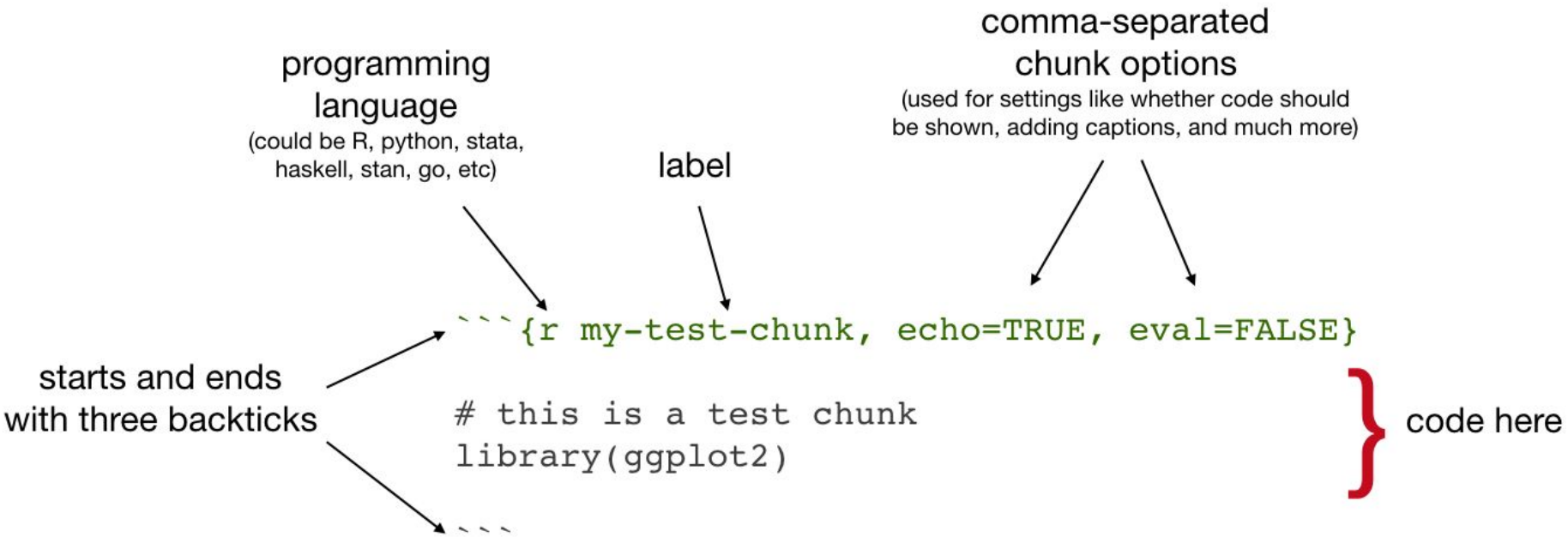
- Title of the document
- Author of the document
- Date: system date at time of creation
- Output: HTML

````
---
title: "Notebook example"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

# Notebook example
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}
library(ggplot2)
qplot(Sepal.Length, Petal.Length, data = iris, color = Species,
size = Petal.Width)
```
````

Rmarkdown YAML metadata

include = TRUE/FALSE	Include code and results in the final output or not. Code is run and outputs created, but is displayed or hidden according to TRUE or FALSE
echo = TRUE/FALSE	Display the code in the final output or not. Results, such as graphs, not affected.
message = TRUE/FALSE	Display output messages (such as package loading info) in the final output or not. Code and outputs not affected.
warning = TRUE/FALSE	Display warning messages in the final output or not.
fig.cap= “…”	Attach a figure caption to graphical results. Put the caption text in the “ ”.

Rmarkdown chunks



Rmarkdown chunk settings

include = TRUE/FALSE	Include code and results in the final output or not. Code is run and outputs created, but is displayed or hidden according to TRUE or FALSE
echo = TRUE/FALSE	Display the code in the final output or not. Results, such as graphs, not affected.
message = TRUE/FALSE	Display output messages (such as package loading info) in the final output or not. Code and outputs not affected.
warning = TRUE/FALSE	Display warning messages in the final output or not.
fig.cap= “…”	Attach a figure caption to graphical results. Put the caption text in the “ ”.

@CBolibaugh

Rmarkdown syntax

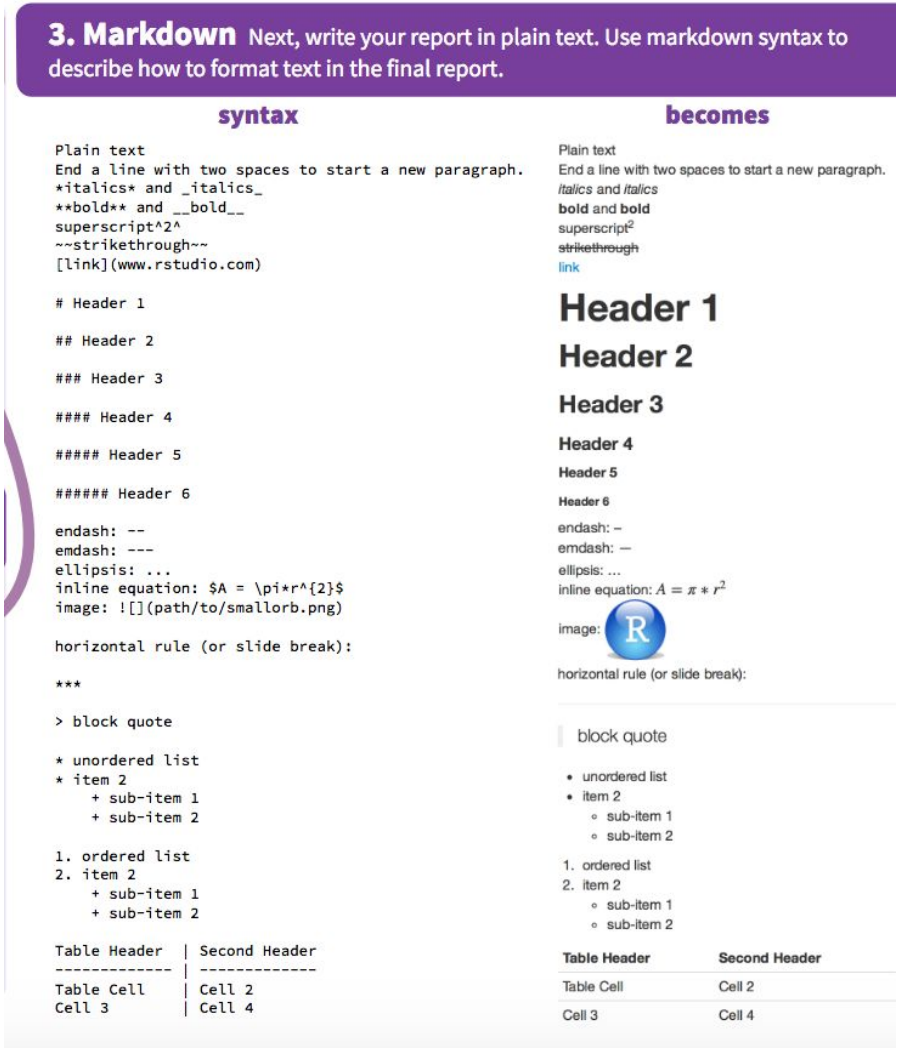
R Markdown Cheat Sheet

learn more at rmarkdown.rstudio.com

rmarkdown 0.2.50 Updated: 8/14

RStudio

https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf



3. Markdown Next, write your report in plain text. Use markdown syntax to describe how to format text in the final report.

syntax

```
Plain text
End a line with two spaces to start a new paragraph.
*italics* and _italics_
**bold** and __bold__
superscript^2^
~~strikethrough~~
[link](www.rstudio.com)

# Header 1

## Header 2

### Header 3

#### Header 4

##### Header 5

###### Header 6

endash: --
emdash: ---
ellipsis: ...
inline equation: $A = \pi*r^{2}$
image: ![](path/to/smallorb.png)

horizontal rule (or slide break):

***

> block quote

* unordered list
* item 2
    + sub-item 1
    + sub-item 2

1. ordered list
2. item 2
    + sub-item 1
    + sub-item 2

Table Header  | Second Header
------------- | -------------
Table Cell    | Cell 2
Cell 3        | Cell 4
```

becomes

Plain text
End a line with two spaces to start a new paragraph.
italics and *italics*
bold and **bold**
superscript^2
~~strikethrough~~
link

Header 1
Header 2
Header 3
Header 4
Header 5
Header 6

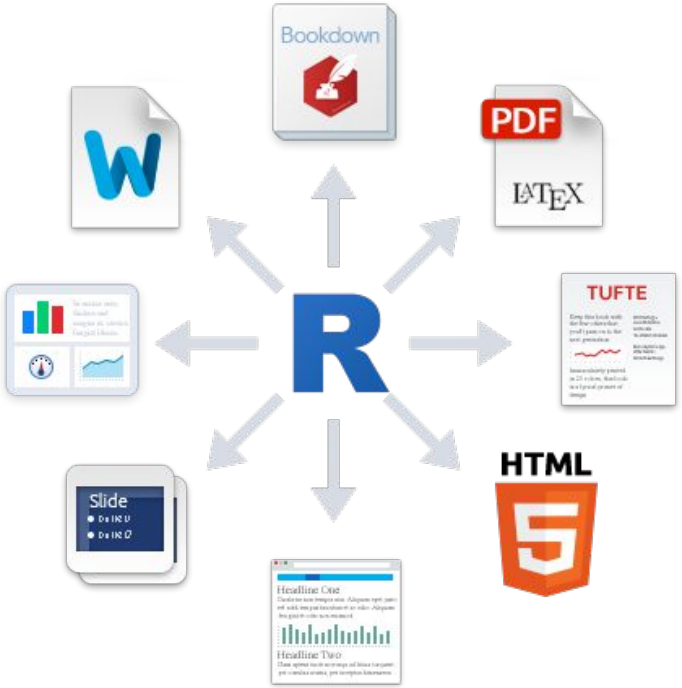
endash: –
emdash: —
ellipsis: …
inline equation: $A = \pi * r^2$
image:
horizontal rule (or slide break):

> block quote

- unordered list
- item 2
    - sub-item 1
    - sub-item 2

1. ordered list
2. item 2
    - sub-item 1
    - sub-item 2

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4



Rmarkdown output formats

‘Knitting’ Rmd files

Knitting a document creates a specific type of output from the .Rmd file:

- Default is HTML
- Word (requires Microsoft Word)
- PDF (requires LaTex installation)

Practicals

Open the RmdWeek10.rmd file in RStudio.

- Add your name and today's date to YAML header
- Add a paragraph, containing a header, bold, and italics
- Change the chunk options to make the code (dis)appear.
- Try writing a new sentence including the mean difference in scores between groups,

Knit the file and see what your changes have accomplished.

Reproducible APA manuscripts

Papaja Reproducible APA Manuscripts

Install papaja

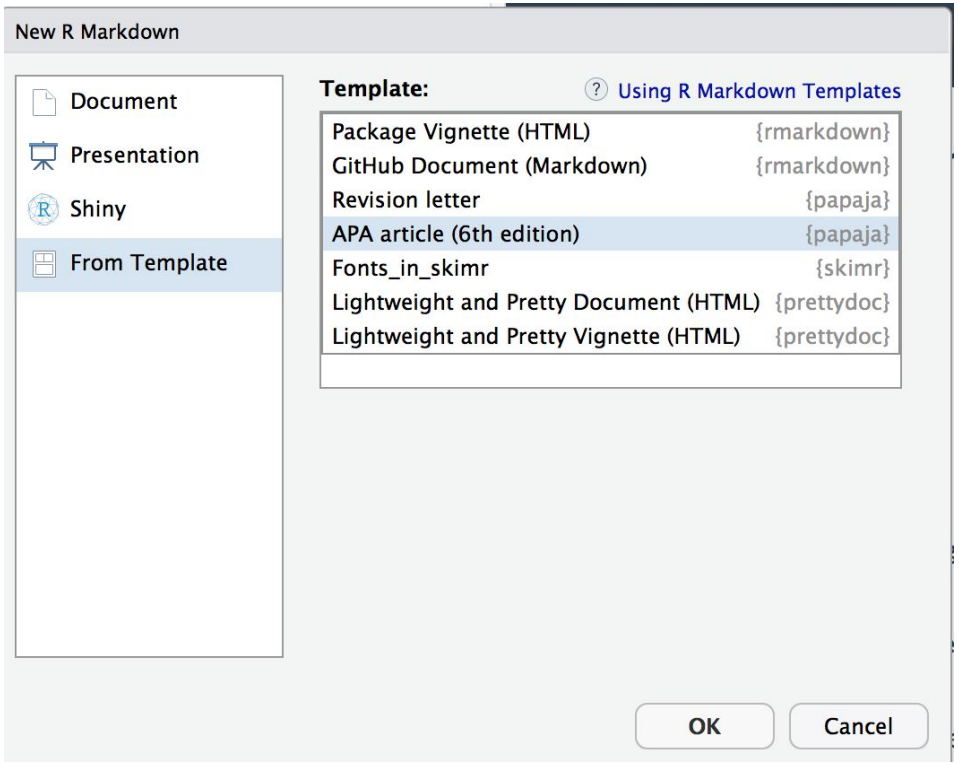
install papaja from this GitHub repository:

```
install.packages("devtools")
```

```
devtools::install_github("crsh/papaja")
```

Create a manuscript

Once you have installed the papaja package you can select the APA template when creating a new Markdown file through the menus in RStudio.



Practicals 2

Open a new .rmd from the Papaja template

- Add your name and details to the YAML
- Add the data simulation, and analysis to the ‘Results’ section of the manuscript.
- Make sure that your code is not outputted.

Knit the file and see what your changes have accomplished.

Resources

- R Markdown Cheat Sheet
- R Markdown reference Guide
- R Markdown Tutorial
- R Markdown: The Definitive Guide by Yihui Xie, J. J. Allaire, & Garrett Grolemund

  Chapter 27: R Markdown in *R for Data Science*