浙江大学爱丁堡大学联合学院

ZJU-UoE Institute

ADS2 Lecture 12
R Markdown – Why and How?

Dr Duncan MacGregor duncan.macgregor@ed.ac.uk

Semester 1, Week 12

2023-24

The Reproducibility Crisis

This lecture is about . . .

. . .how to make your data analysis reproducible by using R Markdown

Learning Objectives

After this lecture you should be able to . . .

- Explain what a markdown language is
- Discuss reasons for using R Markdown

Outline

1. Basic definitions and reminders
2. What is Markdown?
3. Idea behind R Markdown

Reproducibility vs Replicability

General Idea

We want scientific discoveries to reprepresent some form of truth.

Specifically, if a scientist discovers a result, they need to get the same result by running again the experiment.

More importantly, another scientist (with appropriate training, tools, and information) should be able to discover the same result.

"we can never be sure we're right, we can only ever be sure we're wrong"

- Richard Feynman

Reproducibility vs Replicability

General Idea

We want scientific discoveries to reprepresent some form of truth.

Specifically, if a scientist discovers a result, they need to get the same result by running again the experiment.

More importantly, another scientist (with appropriate training, tools, and information) should be able to discover the same result.

Replicate

Discover the same result by designing and running the experiment from scratch: **collect and analyse** data

Reproducibility vs Replicability

General Idea

We want scientific discoveries to reprepresent some form of truth.

Specifically, if a scientist discovers a result, they need to get the same result by running again the experiment.

More importantly, another scientist (with appropriate training, tools, and information) should be able to discover the same result.

Replicate

Discover the same result by designing and running the experiment from scratch: **collect and analyse** data

Reproduce

Draw the same conclusions from an existing dataset: analyse **already existing** data

Reproducibility vs Replicability

General Idea

We want scientific discoveries to reprepresent some form of truth.

Specifically, if a scientist discovers a result, they need to get the same result by running again the experiment.

More importantly, another scientist (with appropriate training, tools, and information) should be able to discover the same result.

Replicate

Discover the same result by designing and running the experiment from scratch: **collect and analyse** data

Reproduce

Draw the same conclusions from an existing dataset: analyse **already existing** data

What differences do you see between replicability and reproducibility? What do they have in common? Can you imagine obstacles to achieving replicability or reproducibility?

Reproducible research

Have you come across techniques or ideas that are designed to make research more reproducible?

Reproducible research

Have you come across techniques or ideas that are designed to make research more reproducible?

- Tidy data
- Publishing data
- Documenting code
- Publishing code
- Methods sections in publications
- Peer review (?)
- Preregistration
- Standards and protocols
- …

Outline

1. Basic definitions and reminders
2. What is Markdown?
3. Idea behind R Markdown

But first ...

What is Markup?

Markup languages

A Markup Language is a system for creating documents that contain both text and instructions on what should be done with the text (e.g. how it should be displayed).

Markup languages

A Markup Language is a system for creating documents that contain both text and instructions on what should be done with the text (e.g. how it should be displayed).

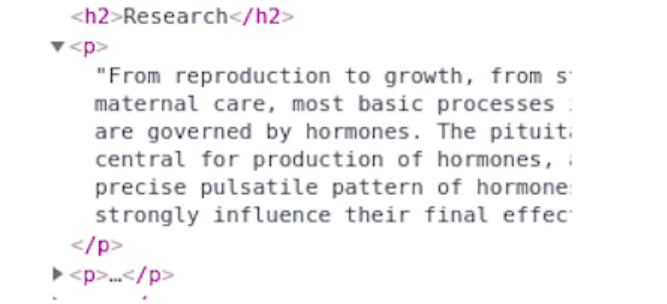
Example: HTML (Hyper Text Markup Language)

Top: an excerpt of Nico's webpage on the UoE website

Bottom: the corresponding HTML

The browser interprets the HTML and renders it according to the instructions within.

Can you think of other examples of markup languages?



Markdown

Markdown is a “light-weight” markup language.

- simple syntax
- easy to read and write
- can be converted into markup (e.g. HTML) or rendered into other document formats (e.g. PDF, DOC)

R Markdown

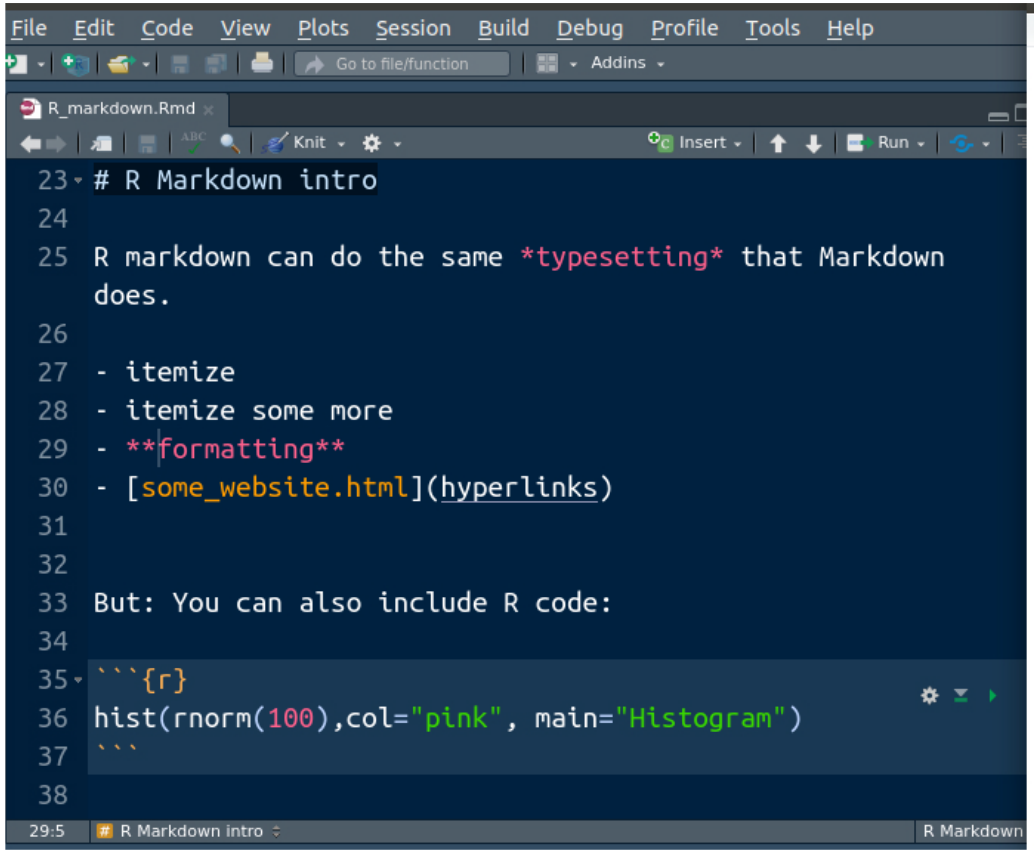
R Markdown is an extended version of Markdown, that allows running R (or other computing language) code directly inside a Markdown document! Which is awesome!

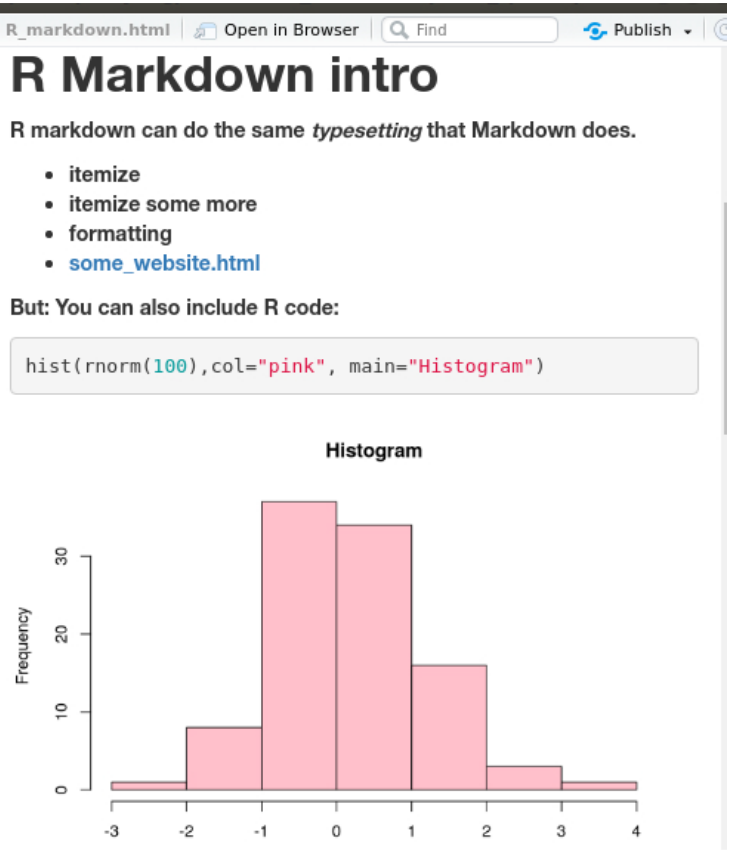
Outline

1. Basic definitions and reminders
2. What is Markdown?
3. Idea behind R Markdown

Main idea of R markdown

Code and documentation in one document. This is done through a combination of formatting (using Markdown) and the inclusion of “code blocks”, which are both reported and executed in the final document





Knitting Markdown

Once you have written your Markdown code you can use software such as RStudio to *knit* the R Markdown file into a document (HTML, PDF or DOC)

Why use R Markdown?

You work in biomedical informatics. In what situations would it be useful to use R Markdown, and why?

What questions do you have?

Now you should be able to . . .

- Explain what a markdown language is
- Discuss reasons for using R Markdown

Optional further reading

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature News, 533(7604), 452.

Acknowledgements and Image credits

This lecture uses materials from ADS2 lectures by Melanie Stefan and Nicola Romano. Where not otherwise indicated, images are also from those lectures.

- Example of HTML code and the resulting webpage (screenshot). Nicola Romano (2020), CC-BY-SA 3.0.
- Example of markdown code (screenshot). Melanie Stefan (2019), CC-BY-SA 3.0.
- Example of R Markdown code and resulting file (screenshot). Melanie Stefan (2019), CC-BY-SA 3.0
- Reproducibility crisis screenshot. From: Baker, M. (2016). 11,500 scientists lift the lid on reproducibility
- Three students working on computers. By Yuuki Guzman and Agoston Tyll (Okinawa Institute of Science and Technology), 2015.
- Two students working on an experiment together. By Yuuki Guzman (Okinawa Institute of Science and Technology), 2016.
- Knitting needles and thread - ulricaloeb, Flickr, 2017. CC-BY 2.0