浙江大学爱丁堡大学联合学院

ZJU-UoE Institute

R Markdown - Why and How?

ADS 2, Lecture 12

Melanie Stefan - melanie.stefan@ed.ac.uk

Semester 1, 2019/20

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

Reproducibility vs Replicability

General idea

Wa want scientific discoveries to represent some form of truth. Specifically, if a scientist discovers a result, another scientist (with appropriate training, tools, and information) should be able to discover the same result.

Reproducibility vs Replicability

General idea

Wa want scientific discoveries to represent some form of truth. Specifically, if a scientist discovers a result, 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

Wa want scientific discoveries to represent some form of truth. Specifically, if a scientist discovers a result, 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

Wa want scientific discoveries to represent some form of truth. Specifically, if a scientist discovers a result, 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

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

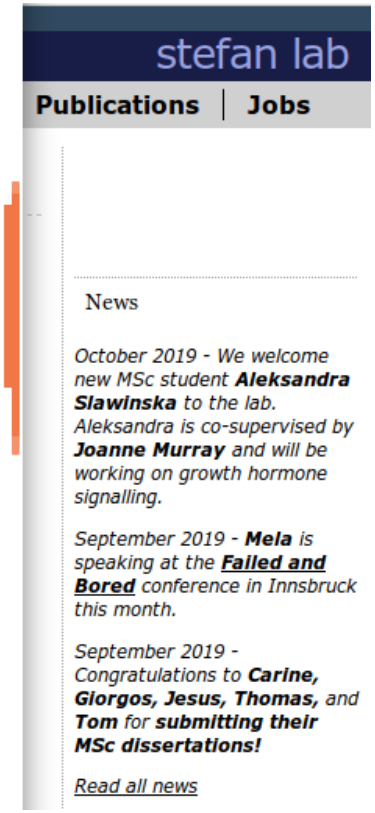
A Markup Language is a system for creating documents that contain both text and instructions on what should be done with the text (e.g. how it should be displayed).
Example: html (hyper text markup language)

```
    <dt>News</dt>
 <dd>
 <p>
 <em>October 2019<em> - We welcome new MSc student <b>Aleksandra Slawinska</b> to the lab. Al
eksandra is co-supervised by <b>Joanne Murray</b> and will be working on growth hormone sign
alling.
 </p>
 </dd>

 <dd>
 <p>
 <em>September 2019<em> - <b>Mela</b> is speaking at the <b><a href="https://www.mariecurieal
umni.eu/mcaa-events/failed-bored-conference">Failed and Bored</a></b> conference in Innsbruc
k this month.
 </p>
 </dd>

 <dd>
 <p>
 <em>September 2019<em> - Congratulations to <b>Carine, Giorgos, Jesus, Thomas,</b> and <b>To
m</b> for <b>submitting their MSc dissertations!</b>
 </p>
 </dd>

 <p>
 <a href="http://stefanlab.net/news.html">Read all news</a>
 </p>
```



Markup languages

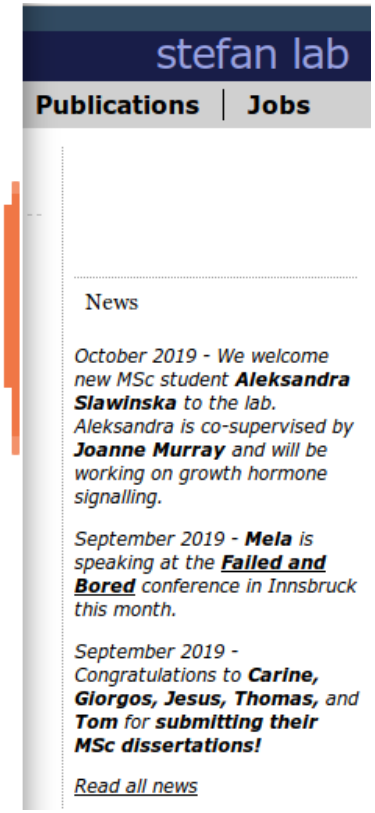
A Markup Language is a system for creating documents that contain both text and instructions on what should be done with the text (e.g. how it should be displayed).
Example: html (hyper text markup language)

```
    <dt>News</dt>
 <dd>
 <p>
 <em>October 2019<em> - We welcome new MSc student <b>Aleksandra Slawinska</b> to the lab. Al
eksandra is co-supervised by <b>Joanne Murray</b> and will be working on growth hormone sign
alling.
 </p>
 </dd>

 <dd>
 <p>
 <em>September 2019<em> - <b>Mela</b> is speaking at the <b><a href="https://www.mariecurieal
umni.eu/mcaa-events/failed-bored-conference">Failed and Bored</a></b> conference in Innsbruc
k this month.
 </p>
 </dd>

 <dd>
 <p>
 <em>September 2019<em> - Congratulations to <b>Carine, Giorgos, Jesus, Thomas,</b> and <b>To
m</b> for <b>submitting their MSc dissertations!</b>
 </p>
 </dd>

 <p>
 <a href="http://stefanlab.net/news.html">Read all news</a>
 </p>
```



Other examples?

Markdown

Markdown is a "light-weight" markup language.

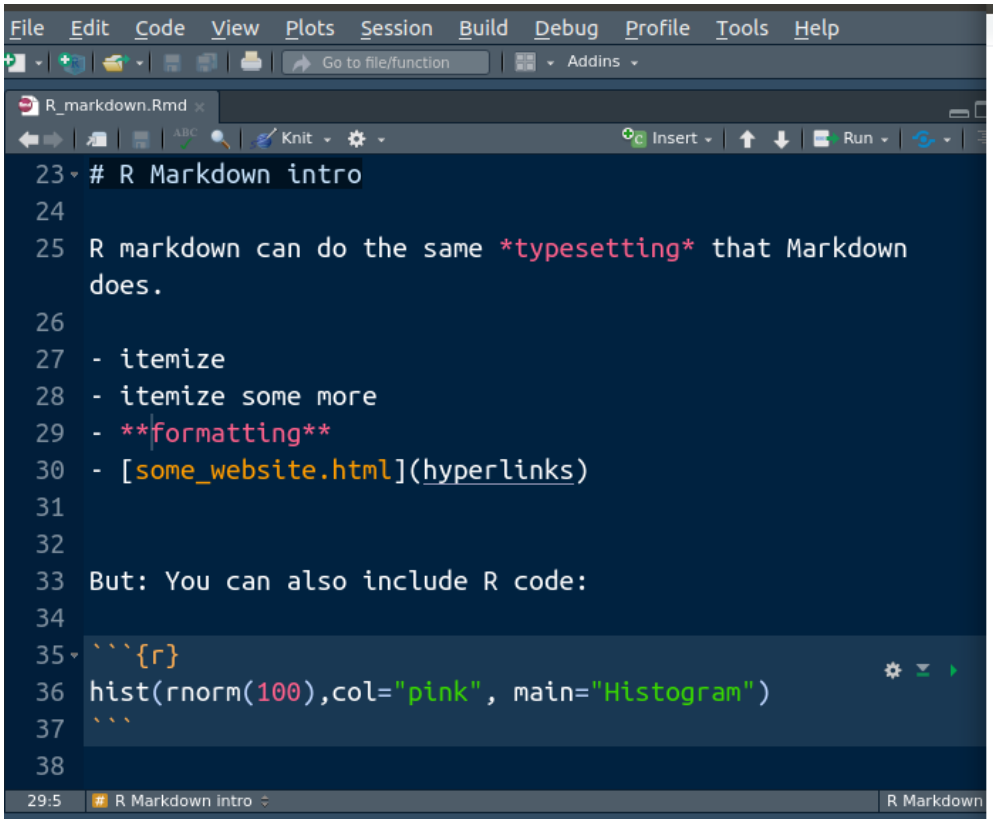
- simple syntax
- easy to read and write
- can be converted into markup (e.g. html)

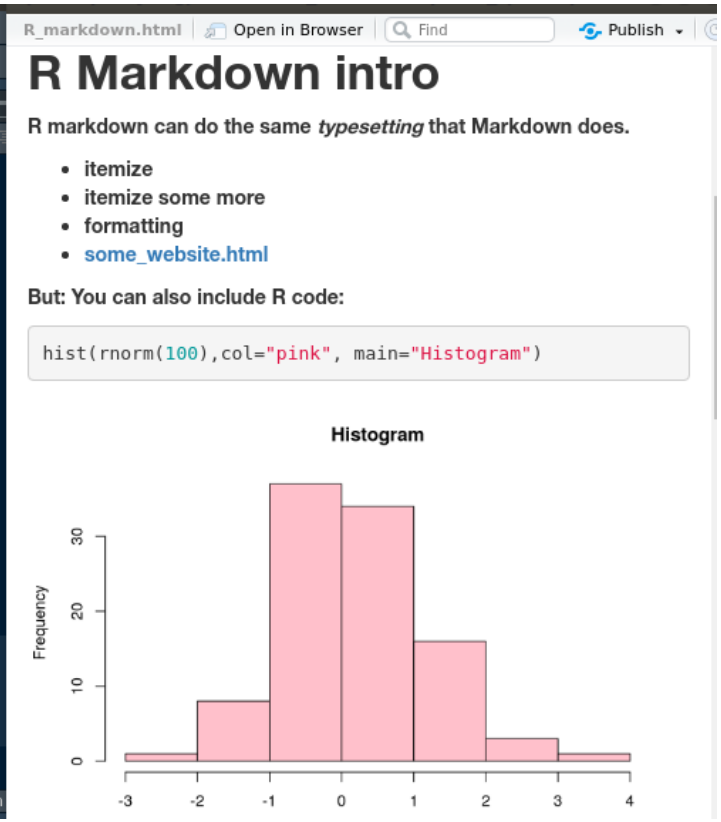
```
11
12
13 # Markdown intro
14
15 Markdown is *awesome*, look at what it can
   do, for instance:
16
17 - itemize
18 - itemize some more
19 - **formatting**
20 - [some_website.html](hyperlinks)
21
22
23
```

Outline

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





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. https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

Image credits

- Example of html code and the resulting webpage (screenshot). My own work (2019), CC-BY-SA 3.0.
- Example of markdown code (screenshot). My own work (2019), CC-BY-SA 3.0.
- Example of R Markdown code and resulting file (screenshot). My own work (2019), CC-BY-SA 3.0
- Reproducibility crisis screenshot. From: Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature News, 533(7604), 452. `https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970`
- 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.