Workshop: Data science with R

ZEW - Session #4

Obryan Poyser

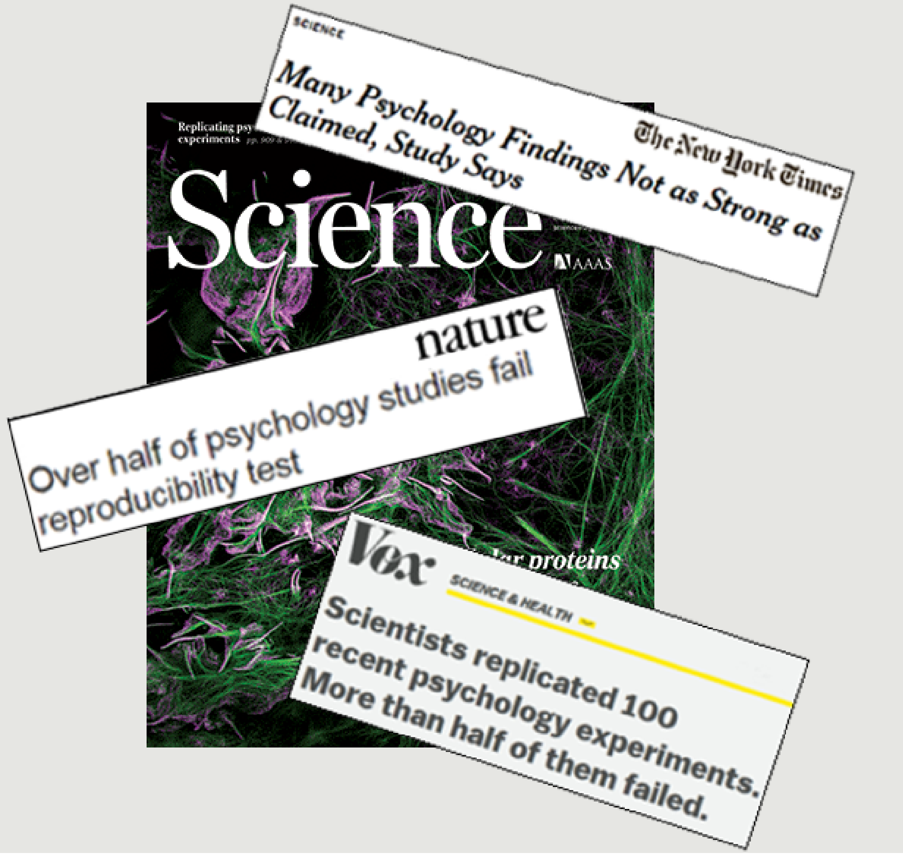
2019-03-20

Outline

1. Reproducible research
    1. LaTeX and Markdown
    2. `knitr` and `rmarkdown`
2. Creating documents in R
    1. PDF, html, beamer, xaringan
    2. Static blogs

Further reading: Gandrud, C. (2016). Reproducible research with R and R studio. Chapman and Hall/CRC. Xie, Y. (2016). Bookdown: Authoring Books and Technical Documents with R Markdown. Chapman and Hall/CRC. Xie, Y., Allaire, J. J., & Grolemund, G. (2018). R markdown: The definitive guide. CRC Press.

Reproducible research


SCIENCE
The New York Times
Replicating psy experiments
Science
AAAS
nature
Vox
SCIENCE & HEALTH
...lar proteins

Many Psychology Findings Not as Strong as Claimed, Study Says

Over half of psychology studies fail reproducibility test

Scientists replicated 100 recent psychology experiments. More than half of them failed.

Source: Hardvard press on Estimating the reproducibility of psychological science

Definition 1: Comprenhensive process of interaction with information that is certified to be reliable, of traceability and provenance, accountable reuse, recycling and re-sampling of pre-existing sources, leading to better practices overall. Source

---

Definition within R: Having sufficient information available that allow third party researchers to find the same results following a given process. Replication open our study to scrutiny. Source

Reproducibility promotes betters individual habits and team work.

Reproducible research

1. Articles and presentations are meant to convince the audience (editors) that the hypothesis you are working on is proven to be true/false.
2. Why R?: It is an all-in-one statistical platform to include markup languages and step-by-step code.
    1. Defining cleaning and transformation process
        1. Imputation?
        2. Transformation such as BoxCox; IHS, or Logs?
    2. Modelling
        1. Is someone doing data fishing or p-hacking?

            > If you torture the data enough, nature will always confess. R. Coase.

    3. Embed the results

Markup languages

Def: a markup language is a system for annotating a document in a way that is syntactically distinguishable from the text. Examples:

TeX:

The standard for writing articles/presentations in academia. LaTeX is a documented package to write plain text opposed to formatted text (Word style).

```
\usepackage[utf8]{inputenc}
\usepackage{mathtools}

\begin{document}
  \begin{equation}\label{eqn:einstein}
    E=mc^2\tag{*}
  \end{equation}
  \eqref{eqn:einstein}
\end{document}
```

HTML

Standard ML for creating web pages and web applications.

```
<HTML>
<HEAD>
<TITLE>Your Title Here</TITLE>
</HEAD>
<CENTER><IMG SRC="clouds.jpg" ALIGN="BOTTOM"> </CENTER>
<a href="http://somegreatsite.com">Link Name</a>
is a link to another nifty site
<H1>This is a Header</H1>
<H2>This is a Medium Header</H2>
Send me mail at <a href="mailto:support@yourcompany.com">
support@yourcompany.com</a>.
<P> This is a new paragraph!
<P> <B>This is a new paragraph!</B>
<BR> <B><I>This is a new sentence without a paragraph break, i
<HR>
</BODY>
</HTML>
```

Linguistics and grammar:
https://github.com/github/linguist/blob/master/lib/linguist/languages.yml

Markup languages

Markdown

1. Lightweight markup language, "to write using an easy-to-read and easy-to-write plain text format, optionally convert it to structurally valid XHTML (or HTML)" J. Gruber
2. Markdown’s syntax is intended for one purpose: to be used as a format for writing for the web.
3. The idea for Markdown is to make it easy to read, write, and edit prose.

Markdown

Block elements

Headers are defined with hashs "#"

```
# This is an H1
## This is an H2
###### This is an H6
```

Blockquote are defined with ">"

```
> This is the first level of quoting.
>
> > This is nested blockquote.
>
> Back to the first level.
```

Lists could be ordered and unordered. For unordered - + and * are interchangeable. Ordered lists admits any sequential numbered lines and interprets them as a list.

```
# unordered
-   Red
-   Green
-   Blue
    # is the same as:
-   Red
-   Green
-   Blue
# ordered
1.   Red
1.   Green
1.   Blue
# is the same as:
5.   Red
9.   Green
2.   Blue
```

Markdown

Block elements

Codeblocks are defined with three consecutive backticks "`"

1. Highlighting is defined right after the backticks, for instance:
2. In line code can be obtained with a single backtick closing the text `css

Links are formated, first surrounding the text with squared brackets then inserting the link inside parentheses.

```
This is [an example](http://example.com/ "Title") inline link.
[This link](http://example.net/) has no title attribute.
```

Or by reference:

```
This is [an example][id] reference-style link.

[foo]: http://example.com/  "Optional Title Here"
[foo]: http://example.com/  'Optional Title Here'
[foo]: http://example.com/  (Optional Title Here)
```

Markdown treats emphasis with asterisks and underscores

```
(*)single asterisks(*)
_single underscores_
(**)double asterisks(**)
__double underscores__
```

Images can be inserted as:

```
![Alt text](/path/to/img.jpg)
![Alt text](/path/to/img.jpg "Optional title")
```

Footnotes are formated as:

```
I have more [^1] to say up here.

[^1]: To say down here.
```

Equations: in line are formated with "$" closing the latex math code. Big equations goes with double dollar sign

RMarkdown

1. The basic idea behind dynamic documents stems from literate programming, a programming paradigm conceived by Donald Knuth (Knuth, 1984).
2. The original idea was mainly for writing software: mix the source code and documentation together; we can either extract the source code out.
3. The document format “R Markdown” was first introduced in the knitr package (Xie 2015, 2019b) in early 2012.
4. The idea was to embed code chunks (of R or other languages) in Markdown documents.
5. Markdown has been considered overly simplistic, nonetheless John McFarlane created a Pandoc to convert Markdown documents (and many other types of documents) to a large variety of output formats.
6. R Markdown stands on the shoulders of knitr and Pandoc. The former executes the computer code embedded in Markdown, and converts R Markdown to Markdown. The latter renders Markdown to the output format you want (such as PDF, HTML, Word, and so on).

What can I write with an RMarkdown document?

```
1. Journal articles
1. Dashboards
1. Websites
1. Blogs
1. Much more!
```

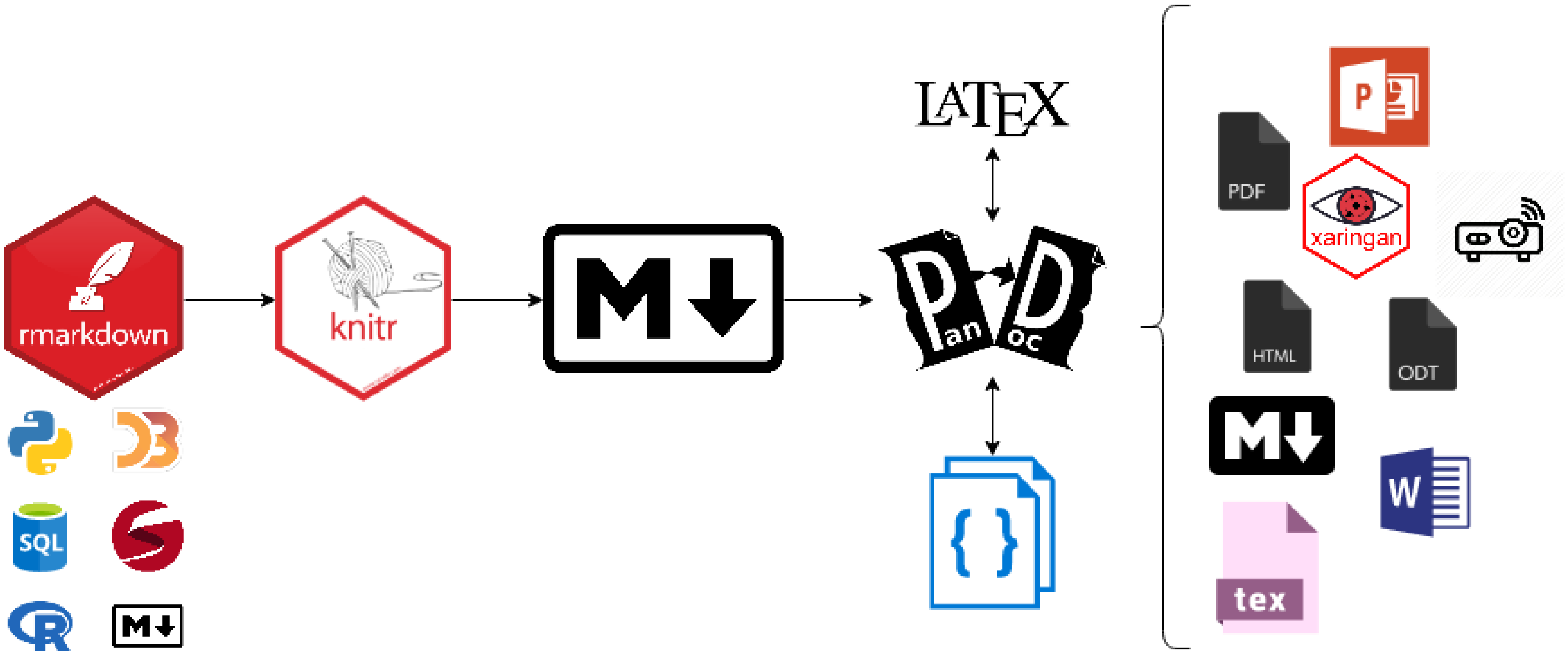
RMarkdown

Let's get our hands dirty! We will need the following packages:

```r
install.packages("rmarkdown", dependencies = T)
install.packages("blogdown", dependencies = T)
install.packages("bookdown", dependencies = T)
install.packages("knitr", dependencies = T)
install.packages("citr", dependencies = T)
install.packages('tinytex', dependencies = T)
```

```r
blogdown::install_hugo()
tinytex::install_tinytex()
```

RMarkdown



Process to convert RMarkdown to other text formats

RMarkdown

Useful websites:

Tinytex: https://yihui.name/tinytex/

Pandoc: https://pandoc.org/MANUAL.html#variables-for-latex

Xaringan: https://github.com/yihui/xaringan/wiki

Knitr: https://yihui.name/knitr/

Bookdown: https://bookdown.org/yihui/rmarkdown/html-document.html#mathjax-equations

RMarkdown: basic metadata

Metadata is defined in the header of any Rmd document. It defines the desired aspects regarding the structure, format, items, etc.

```
---
title:  'This is the title'
subtitle: "This is the subtitle"
author:
- Author One
- Author Two
description: |
    This is a long
    description.

    It consists of two paragraphs
abstract: "This is a abstract"
---
```

Besides the aforementioned metadata we have:

1. `classoption`: option for document class, e.g. oneside; repeat for multiple options
2. `documentclass`: document class: usually one of the standard classes, article, report, and book
3. `geometry`: option for geometry package, e.g. margin=1in
4. `linestretch`: adjusts line spacing using the setspace package, e.g. 1.25, 1.5
5. `margin-left, margin-right, margin-top, margin-bottom`: sets margins if geometry is not used (otherwise geometry overrides these)
6. `paper` size, e.g. letter, a4

RMarkdown: blogs

Example->

RMarkdown: presentation