

USING L^AT_EX AND MARKDOWN FOR REPRODUCIBLE RESEARCH

ERIK BECK

EMILY Y. LI

US EPA

R USER GROUP WORKSHOP

13 AUGUST 2019

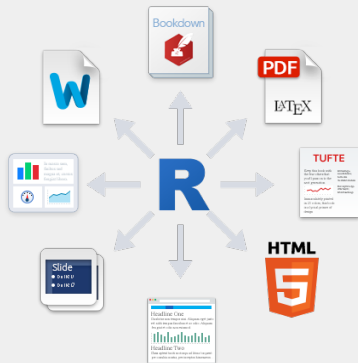


TABLE OF CONTENTS

- 1 Agenda
- 2 Introduction to Markdown
- 3 Introduction to LaTeX
- 4 Markdown & LaTeX examples
- 5 Reproducible Research
- 6 Dynamic documents with Sweave and knitr
- 7 Markdown & LaTeX with R
- 8 Additional Resources

TABLE OF CONTENTS

- 1 Agenda
- 2 Introduction to Markdown
- 3 Introduction to LaTeX
- 4 Markdown & LaTeX examples
- 5 Reproducible Research
- 6 Dynamic documents with Sweave and knitr
- 7 Markdown & LaTeX with R
- 8 Additional Resources

Using LaTeX and Markdown for Reproducible Research

Day	Time	Room
Tues, Aug 13	8:00-11:30am	C111C

- This 1/2-day workshop will provide attendees with hands-on experience using the basics of LaTeX, Markdown, and the R package knitr.
- After attending this workshop, you will be able to use these tools to facilitate reproducible reports and research with R.

We will try to use our three hours as effectively as possible.

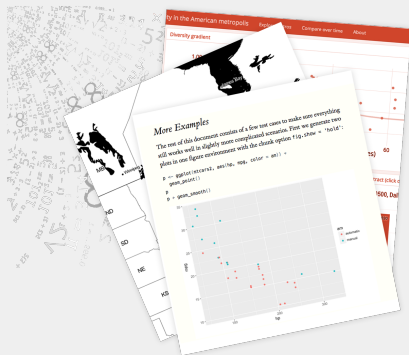
Rough Agenda

#	Time	Topic
1	8:00	System checks & agenda
2	8:20	Intro to Markdown
3	8:40	Intro to LaTeX
4	9:00	Markdown & LaTeX examples
5	9:30	Reproducible Research
	9:50	<i>BREAK</i>
6	10:00	Dynamic documents with knitr
7	10:30	Markdown & LaTeX with R
8	11:20	Wrap-up & additional resources

TABLE OF CONTENTS

- 1 Agenda
- 2 Introduction to Markdown**
- 3 Introduction to LaTeX
- 4 Markdown & LaTeX examples
- 5 Reproducible Research
- 6 Dynamic documents with Sweave and knitr
- 7 Markdown & LaTeX with R
- 8 Additional Resources

INTRO TO MARKDOWN



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

- An **.Rmd file** is an R Markdown file
- Contains the code that a scientist needs to reproduce your work, along with the narration that a reader needs to understand it.
- Choose to export the finished report in a variety of formats, including HTML, PDF, or MS Word.

- Markdown allows us to write using an easy-to-read, easy-to-write plain text format.
- As long as you know how to write emails, you can learn it in a few minutes.
- `https://en.wikipedia.org/wiki/Markdown#Example`

Limitations of Markdown

Markdown was primarily designed to be simple.
For more complicated typesetting, LaTeX may be preferred.

- Maximal focus on content, minimal focus on formatting while writing.
- Quick layout of documents
- Can be converted to a variety of formats easily
- Format durability (records)

MARKDOWN — A SHORT EXAMPLE

First level header

Sup universe!

Second level header

This is **bold**, and *italic*.

- list item
- list item

You can write an ordered list:

1. item 1
1. item 2 # this line will render as "2."

WORKFLOW IN MARKDOWN

Using RStudio:

Open a new .Rmd file, which pre-populates with a template

Write a document by editing the template

Knit the document to create a report; use the knitr button or `render()` to knit

Preview output in IDE window

Publish to web server (optional)

Use output file that is saved alongside .Rmd

Helpful link:

<https://rmarkdown.rstudio.com/lesson-2.html>

TABLE OF CONTENTS

- 1 Agenda
- 2 Introduction to Markdown
- 3 Introduction to LaTeX**
- 4 Markdown & LaTeX examples
- 5 Reproducible Research
- 6 Dynamic documents with Sweave and knitr
- 7 Markdown & LaTeX with R
- 8 Additional Resources

LaTeX is a tool for high-quality typesetting, designed for the production of technical and scientific documentation.

How do you pronounce “LaTeX”?

TeX is usually pronounced tech, making 'lah-teck, lah-'teck, and 'lay-teck the logical choices; but language is not always logical, so 'lay-'tecks is also possible.

— Leslie B. Lamport, original developer of \LaTeX

LaTeX is widely used in academia for the publication of scientific documents in many fields, including mathematics, statistics, computer science, engineering, chemistry, physics, economics, and political science.

1. **TeX engines have excellent quality output.** This especially holds for complex documents such as those with many tables, figures, equations, cross-references or hyperlinks—or just with many pages.
2. **TeX is fast.**
3. **TeX is consistent.**
4. **TeX is stable.** It will never eat your document. *Ever.*

— <https://www.ctan.org/tex/>

LATEX: MINIMAL EXAMPLE

Here is a minimal example of a full document written in LaTeX.

```
\documentclass{article}  
\title{A Minimal LaTeX Example}  
\author{Emily Li}  
  
\begin{document}  
\maketitle  
  
Yer a wizard, Harry.  
  
\end{document}
```

LEVELS OF TEX: A DISAMBIGUATION

Help! There are too many words with “TeX” in them!

If you are wondering, “*Should I use LaTeX or MiKTeX?*”, allow us to clear that up. These two slides will cover four types of TeX-related terms: distributions, editors, engines, and formats.

1. **Distributions:** *MiKTeX, TeX Live, etc.* This is TeX-related software to be downloaded and installed. When someone says, “I need to install TeX on my machine,” they’re usually looking for a distribution.
2. **Editors:** *Emacs, TeXworks, TeXShop, TeXStudio, etc.* These editors are what you use to create a document file. Some (e.g., TeXShop) are devoted specifically to TeX, while others (e.g., Emacs) can be used to edit any sort of file.

— <http://www.tug.org/levels.html>

LEVELS OF TEX: A DISAMBIGUATION

A quick note on editors

You can also use Notepad to edit plaintext, including LaTeX code.

3. **Engines:** *TeX*, *pdfTeX*, *XeTeX*, *LuaTeX*, etc. These are the executable binaries which implement different TeX variants. When someone says, “TeX can’t find my fonts,” they usually mean an engine.
4. **Formats:** *LaTeX*, *plain TeX*, etc. These are the TeX-based languages in which one actually writes documents. When someone says, “TeX is giving me a mysterious error,” they usually mean a format. (Incidentally, “LaTeX” has meant “LaTeX2e” for many years now.)

— <http://www.tug.org/levels.html>

TeX DISTRIBUTIONS

To compile LaTeX, your computer needs one of these TeX distributions installed:

TeX Distributions

Distribution	Operating System
MiKTeX	Windows OS
TeX Live	Linux and other UNIX-like systems
MacTeX	Mac OS X

You can also use an online ready-to-use option like [Overleaf.com](https://overleaf.com).

Try compiling this LaTeX

```
\documentclass{article}  
\title{A Minimal LaTeX Example}  
\author{Emily Li}  
  
\begin{document}  
\maketitle  
  
Yer a wizard , Harry.  
  
\end{document}
```

TABLE OF CONTENTS

- 1 Agenda
- 2 Introduction to Markdown
- 3 Introduction to LaTeX
- 4 Markdown & LaTeX examples**
- 5 Reproducible Research
- 6 Dynamic documents with Sweave and knitr
- 7 Markdown & LaTeX with R
- 8 Additional Resources

EXAMPLES & DEMOS

A MODEST EXERCISE

Let's try a little markup:

- One** Use Markdown to write a .html file with your name and favorite hobby.
- Two** Use LaTeX to write a .pdf file with the name of your first school.

Thoughts?

Markdown is super easy to learn (*i.e.* there's nothing TO learn), but is very limited in typesetting controls.

LaTeX has more commands to learn, but always looks better—even with minimal coding.

Each serves a different purpose. If you're writing a journal article, thesis, textbook, or resume, you may want the extra precision.

TABLE OF CONTENTS

- 1 Agenda
- 2 Introduction to Markdown
- 3 Introduction to LaTeX
- 4 Markdown & LaTeX examples
- 5 Reproducible Research**
- 6 Dynamic documents with Sweave and knitr
- 7 Markdown & LaTeX with R
- 8 Additional Resources

REPRODUCIBLE RESEARCH

IS THERE A REPRODUCIBILITY CRISIS?

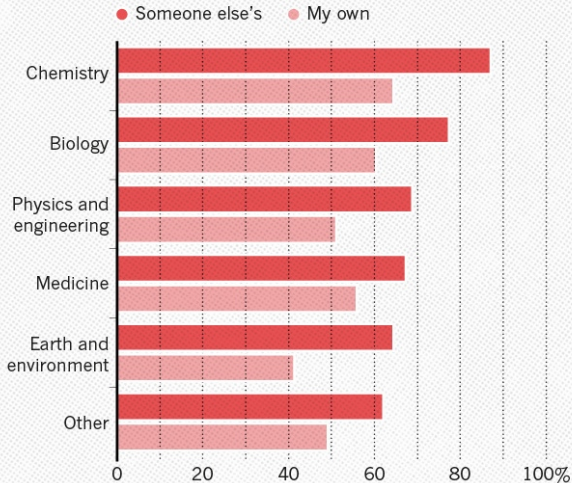


©nature

REPRODUCIBLE RESEARCH

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



Results must be reproducible to be trustworthy.

*An article about computational science in a scientific publication is not the scholarship itself, it is merely the advertising of the scholarship. The actual scholarship is the complete software development environment **and the complete set of instructions which generated the figures.***

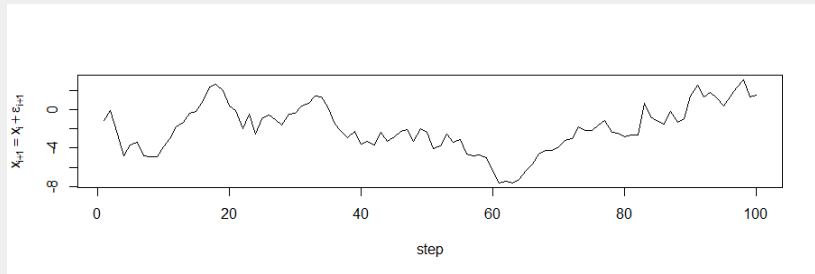
— 1995, David L. Donoho, professor of statistics at Stanford University

This chunk of R code produces a figure that illustrates a simulation of Brownian motion for 100 steps.

Try running this in RStudio

```
set.seed(1213) # for reproducibility
x <- cumsum(rnorm(100))
plot(x, type = "l",
      ylab = expression(x[i+1]==
                        x[i]+epsilon[{i+1}]),
      xlab = "step")
```

REPRODUCIBLE RESEARCH



To put this into a document by hand, we would have to open RStudio, compile the code to draw the plot, save it as an image, then insert it into a document with `\includegraphics{}`.

Then what if we want to change the random seed in `set.seed()`, or the y-axis label?

DYNAMIC REPORT GENERATION

- Instead of separating the results from the computation, we can put everything in one document.
- When we compile this document, the computation will be executed, giving us the results directly.
- Integrating code with narratives is not only easier, but also provides details needed for reproducibility.

Dynamic documents are easier than cut-and-paste

It is fairly common to see student homework and exercises among the countless user contributions on RPubS. Once students are trained, we may expect more reproducible scientific research in the future.

DYNAMIC DOCUMENTS—A SHORT EXAMPLE

This table shows a subset of the *mtcars* dataset.

	mpg	cyl	disp	hp	drat	wt
Mazda RX4	21.0	6	160	110	3.90	2.620
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875
Datsun 710	22.8	4	108	93	3.85	2.320
Hornet 4 Drive	21.4	6	258	110	3.08	3.215
Hornet Sportabout	18.7	8	360	175	3.15	3.440
Valiant	18.1	6	225	105	2.76	3.460

Have you ever included a table in your document? In LaTeX, the code would look like this...

DYNAMIC DOCUMENTS—A SHORT EXAMPLE

Manual table in LaTeX

```
\begin{tabular}{lrrrrrrr}
    & mpg & cyl & disp & hp & drat & wt & \\
Mazda RX4 & 21.0 & 6 & 160 & 110 & 3.90 & 2.620 & \\
Mazda RX4 Wag & 21.0 & 6 & 160 & 110 & 3.90 & 2.875 & \\
Datsun 710 & 22.8 & 4 & 108 & 93 & 3.85 & 2.320 & \\
Hornet 4 Drive & 21.4 & 6 & 258 & 110 & 3.08 & 3.215 & \\
Hornet Sportabout & 18.7 & 8 & 360 & 175 & 3.15 & 3.440 & \\
Valiant & 18.1 & 6 & 225 & 105 & 2.76 & 3.460 & \\
\end{tabular}
```

Even if you maintain this table in Excel, there are manual steps and human intervention required to update your document.

DYNAMIC DOCUMENTS—A SHORT EXAMPLE

If you wanted to include this same table in your dynamic document, you would only need this code:

Table generated in R code

```
library(knitr)  
kable(head(mtcars[, 1:6]))
```

If you ever update the `mtcars` dataframe, the table will update in the document on its own.

BREAK FOR 10 MINUTES



WELCOME BACK!



TABLE OF CONTENTS

- 1 Agenda
- 2 Introduction to Markdown
- 3 Introduction to LaTeX
- 4 Markdown & LaTeX examples
- 5 Reproducible Research
- 6 Dynamic documents with Sweave and knitr**
- 7 Markdown & LaTeX with R
- 8 Additional Resources

Sweave and **knitr** can compile narrative documents with R code.

Sweave is included in base R, and specifically compiles R code integrated into LaTeX or LyX documents.

knitr compiles any input language (R, Python, SAS...) inside any markup language (LaTeX, Markdown, HTML...)

Sweave has been a prominent, longstanding tool for dynamic documents since 2002.

- Combines the power of R with the production value of LaTeX
- Part of base R (as the `Sweave()` function)
 - ▶ From your R session: `Sweave("your_file.Rnw")`
 - ▶ From the command line: `R CMD Sweave your_file.Rnw`

However...

- Development has plateaued in recent years
- Not modular enough; extensions may become incompatible. Some packages are no longer synchronized.
- A PDF produced from LaTeX looks great, but is often not the format we need when collaborating.

knitr was largely motivated by Sweave, and designed to be easier to maintain and extend.

*The design of knitr allows any input languages and **any output markup languages**.*

— Yihui Xie, creator of knitr

- Can compile dynamic documents in Markdown—which are more intuitive and human-readable.
- Works very well with Pandoc, so creating a Word document or OpenDocument format is just as easy as creating a PDF.

Check out <https://www.rdocumentation.org/packages/knitr> for more info on installation, motivation, usage, and functions.

Install knitr

```
install.packages('knitr', dependencies = TRUE)
```

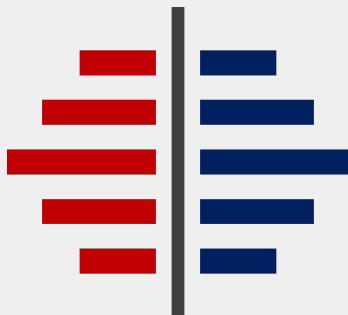
Go to the menu *Tools* > *Options* > *Sweave* and switch the default option for weaving (compiling) to **knitr**.

Visit <https://yihui.name/knitr/faq/> for frequently asked questions.

TABLE OF CONTENTS

- 1 Agenda
- 2 Introduction to Markdown
- 3 Introduction to LaTeX
- 4 Markdown & LaTeX examples
- 5 Reproducible Research
- 6 Dynamic documents with Sweave and knitr
- 7 Markdown & LaTeX with R**
- 8 Additional Resources

MARKDOWN & LATEX SIDE-BY-SIDE



We'll look at a side-by-side comparison of Markdown and LaTeX using the exact same analysis.

We will be using RStudio to edit and compile the source code.

DEMO *Rmd* – *input.Rmd*

- “`{r}`” opens a code chunk and “” terminates a code chunk
- Inline R code is written in backticks ‘ ‘
- Chunk options are written before the closing brace } in the chunk header

You had me at “keyboard shortcut”

In RStudio, for either *.Rnw* or *.Rmd*, quickly insert code chunks with the keyboard shortcut *Ctrl + Alt + I*.

- Files that mix R code and LaTeX have file extension **.Rnw**
- When embedding R code in an Rnw document, start a code chunk with `<<>>=` and terminate it with `@`.
- Write chunk options in between `<<` and `>>=`

Oh, so now the file isn't "TeX" anymore?

"Rnw" stands for "R no web". NoWeb is an old but still active simple system for mixing code and narratives.

QUICK REPORTING IN MARKDOWN

It is also possible to generate a quick report from R script using **knitr**'s `stitch()` function.

Stitch a quick report

```
library(knitr)
stitch("your-script.R")
```

- `stitch()` provides a template so the user only feeds the template with one R script and knitr will compile the template to a report.
- Currently it has built-in templates for LaTeX (default), HTML, and Markdown.

See `?stitch` for details.

LONGER EXAMPLE OF MARKDOWN AND R

Markdown (or LaTeX) can be combined with R using knitr to make longer reports that can be easily updated with a tidbit of new code. Here we will show an example using data from US EPA and US Dept. of State on PM_{2.5} air pollution.

- Particulate pollution is often in the news
- Especially lately with the West Coast forest fires
- CAA primary air pollutant (NAAQS)

Switch over to Rstudio.....

LET'S TRY SOME EXAMPLES
TOGETHER...

Air Example One

Use Markdown to create an R script that will perform summary statistics on a different variable than the concentration value presented.

Air Example Two

Use **LaTeX** to create an R script that will perform summary statistics on a different variable than the concentration value presented.

Air Example Three

Use Markdown to create a plot of your choice from the air data.

TABLE OF CONTENTS

- 1 Agenda
- 2 Introduction to Markdown
- 3 Introduction to LaTeX
- 4 Markdown & LaTeX examples
- 5 Reproducible Research
- 6 Dynamic documents with Sweave and knitr
- 7 Markdown & LaTeX with R
- 8 Additional Resources**

ADDITIONAL RESOURCES

Erik Beck
beck.erik@epa.gov

Emily Y. Li
li.emily@epa.gov

- <https://www.rstudio.com/resources/cheatsheets/>
- <https://support.rstudio.com/hc/en-us/articles/200552056-Using-Sweave-and-knitr>
- Documents source + output examples
<https://yihui.name/knitr/demos/>
- To Markdown, or LaTeX: that is the question...
<https://yihui.name/en/2013/10/markdown-or-latex/>
- https://rmarkdown.rstudio.com/authoring_basics.html

THANKS FOR COMING!