Using LaTeX and Markdown for Reproducible Research

Erik Beck
Emily Y. Li

2018 US EPA
R User Group Workshop

11 Sept 2018

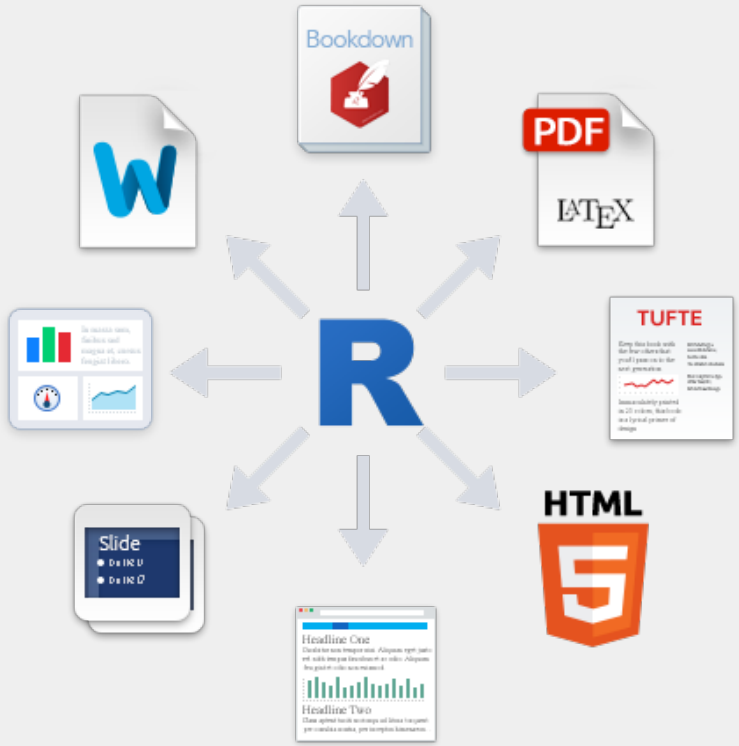


Table of Contents

Table of Contents

Time Frame & Objectives

Using LaTeX and Markdown for Reproducible Research

Day	Time	Room
Tues, Sept 11	**8:30-11:30am**	**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.

Steps/Agenda

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

Rough Agenda

#	Time	Topic
1	8:30	System checks & agenda
2	8:45	Intro to LaTeX
3	9:00	Intro to Markdown
4	9:15	Markdown & LaTeX examples
5	9:40	Reproducible Research
	9:50	*BREAK*
6	10:00	Dynamic documents with Sweave and knitr
7	10:30	Markdown & LaTeX with R
8	11:20	Wrap-up & additional resources

Table of Contents

LaTeX is a tool for high-quality typesetting based on the idea that it is better to leave document design to document designers, and to let authors get on with writing documents.

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 mathematics, with many tables, or many cross-references or hyperlinks, or just with many pages.
2. **TeX is fast.**
3. **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`

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 ShareLaTeX or Overleaf.

LaTeX: Revisiting the minimal example

Try compiling this LaTeX

```
\documentclass{article}
\title{A Minimal LaTeX Example}
\author{Emily Li}

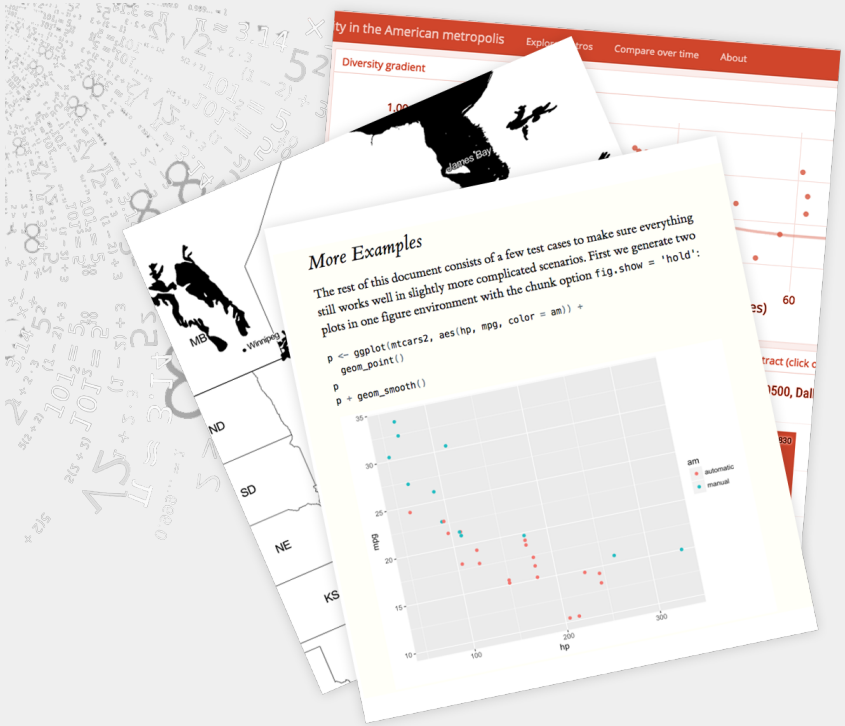
\begin{document}
\maketitle

Yer a wizard, Harry.

\end{document}
```

Table of Contents

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.

Intro to Markdown

- 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.

A short example of Markdown

```
# 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

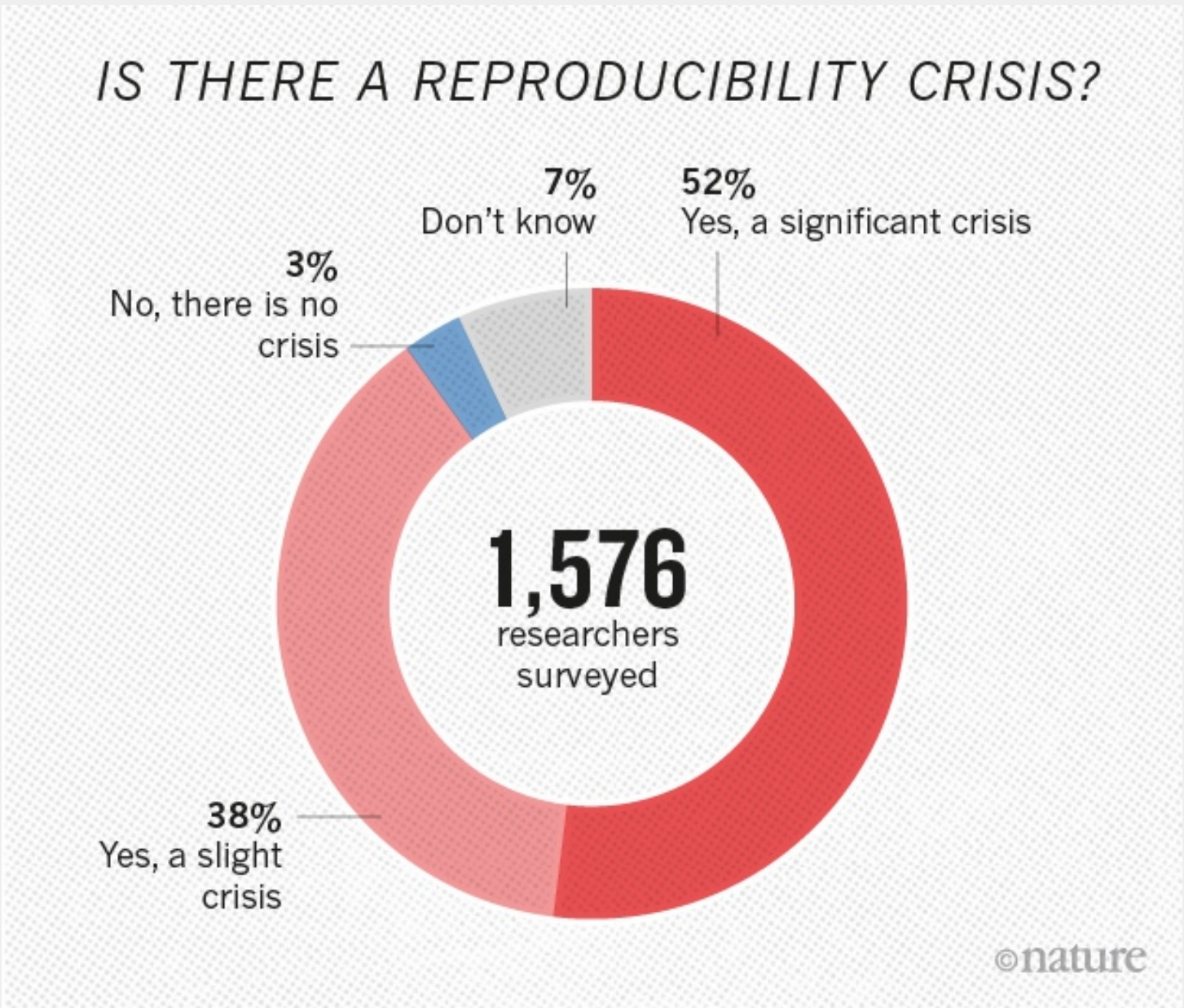
Helpful link:

`https://rmarkdown.rstudio.com/lesson-2.html`

Table of Contents

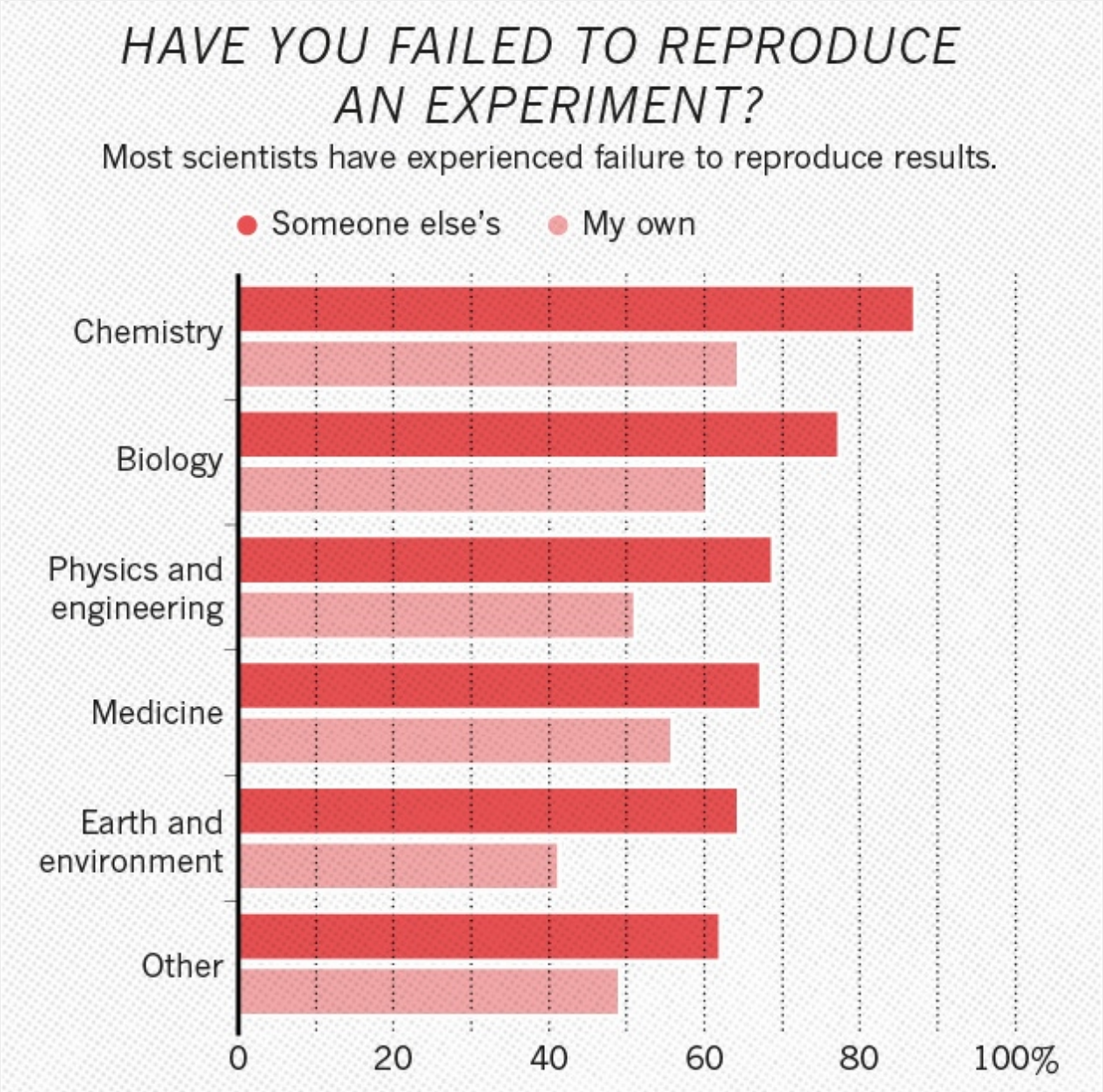
Table of Contents

Reproducible Research



Nature **533**, 452-454 (26 May 2016) | doi:10.1038/533452a

Reproducible Research



Nature **533**, 452-454 (26 May 2016) | doi:10.1038/533452a

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 = ''$x_{i+1}=x_i+\\epsilon_{i+1}$'',
    xlab = ''step'')
```

```
set.seed(1213)
x <- cumsum(rnorm(100))
plot(x, type = ''l'',
    ylab = ''$x_{i+1}=x_i+\\epsilon_{i+1}$'',
    xlab = ''step'')
```

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{}` in LaTeX or 'Insert Image' in Word.

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.

Break for 10 minutes

Image source https://getcubefit.com/. Not an endorsement of the company.

Welcome back!

Table of Contents

Sweave

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

- Deals with Rnw documents, combining the power of R with the production value of LaTeX to enable reproducible research.
- Part of base R (in the **utils** package as the `Sweave()` function)
- Two ways to run Sweave:
  - 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
- Extensions may become incompatible. Some packages are no longer synchronized.

knitr was largely motivated by Sweave First of all, knitr uses Rmarkdown, a set of intuitive human-readable code to do the formatting. While LaTeX is by no means as complicated as its reputation seems to suggest, Rmarkdown is actually easy. By human-readable I mean that anyone who has never even heard of Rmarkdown can understand what is happening to some extent. Sweave is great for producing PDF, but thatâĂŹs one of the biggest drawbacks of LaTeX in the social sciences: while the PDF may look good, they are not the format we need when collaborating with Word-only colleagues, and with rare exceptions when submitting a manuscript to journals. Knitr works very well with Pandoc, so creating a Word document or an ODF is just as easy as creating a PDF. The other day I had to submit a supplementary file as a *.doc file, even though itâĂŹll end up as a PDF on Dataverse or so. With knitr this didnâĂŹt take long.

Why knitr beats Sweave

- LaTeX (to produce Rnw documents) is more complicated than RMarkdown (Rmd), and documents rarely need to be produced as PDFs unless submitting a manuscript to journals.

Table of Contents

Markdown & LaTeX side-by-side

Apples & Oranges?

The next few slides show a side-by-side comparison of Markdown and LaTeX using the exact same analysis.

Example LaTeX with analysis

```
\documentclass{article}
\begin{document}
\title{Speed and Stopping Distance}
\author{Yihui Xie, creator of knitr}

\maketitle

We examine the relationship between speed and stoppi
$Y = \beta_0 + \beta_1 x + \epsilon$.

<<model, fig.width=4, fig.height=3, fig.align='cente
par(mar = c(4, 4, 1, 1), mgp = c(2, 1, 0), cex = 0.8
plot(cars, pch = 20, col = 'darkgray')
fit <- lm(dist ~ speed, data = cars)
abline(fit, lwd = 2)
@

The slope of a simple linear regression is \Sexpr{co
```

R in LaTeX

When embedding R code in LaTeX, start a code chunk with «»= and terminate it with @.

Example Markdown with analysis

```
---
title: Speed and Stopping Distance
---

We examine the relationship between speed and stoppi
$Y = \beta_0 + \beta_1 x + \epsilon$.

```{r fig.width=4, fig.height=3, fig.align='center'}
par(mar = c(4, 4, 1, 1), mgp = c(2, 1, 0), cex = 0.8
plot(cars, pch = 20, col = 'darkgray')
fit <- lm(dist ~ speed, data = cars)
abline(fit, lwd = 2)

The slope of a simple linear regression is `r coef(f
```
```

R in Markdown

- Quickly insert chunks with the keyboard shortcut Ctrl + Alt + I (OS X: Cmd + Option + I).
- By comparison, Markdown has simpler commands.

R in Markdown

- Write code chunks between ```` ```{r} ```` and ```` ``` ````
- Inline R code is written in `` `r  ` ``
- Chunk options are written before closing brace in the chunk header.

Quick reporting in Markdown

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

Stitching quick reports

```
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, HTML, and Markdown.

Table of Contents

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
- knitr document 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/

Thanks for coming!