

Dynamic Reports

With RMarkdown

The R Bootcamp
Twitter: [@therbootcamp](#)

September 2017

What is a dynamic report?

A dynamic report is a document that contains key elements which are automatically accessed and/or created by code:

- Access Data
- Wrangling / Statistics
- Plots
- Regular Text
- Formatting

A screenshot of the RPubs website. The header features the RPubs logo in orange and grey, followed by the text "brought to you by RStudio". On the right side of the header are "Sign in" and "Register" buttons. Below the header, a message says "Last updated 22 days ago". At the bottom of the page are three buttons: "Comments (-)", "Share", and "Hide Toolbars". The main content area is completely blank, indicating no data has been published.

<http://rpubs.com/fabiorocha5150/decisiontreemodel>

Why are dynamic reports important?

1. Efficiency
2. Reproducibility
3. Scaleability

Over half of psychology studies fail reproducibility test

Largest replication study to date casts doubt on many published positive results.

Monya Baker

27 August 2015

According to the replicators' qualitative assessments, as previously reported by *Nature*, only 39 of the 100 replication attempts were successful. (There were 100 completed replication attempts on 98 papers, as in two cases replication efforts were duplicated by separate teams.) But literature. In fact, two thirds of it should probably be distrusted.

In the biggest project of its kind, Brian Nosek, a social psychologist and head of the Center for Open Science in Charlottesville, Virginia, and 269 co-authors repeated work reported in 98 original papers from three psychology journals, to see if



Science AAAS

RESEARCH

RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration¹

Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and observational studies published in three psychology journals using high-powered designs and original methods. The replication attempts were conducted by 270 contributing authors, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of replications had no significant effect. Ninety-four percent of the original effects of effects were subjectively rated to have replicated the original result; and no bias in original results was detected. Correlation coefficients between the strength of original evidence and significant effects. Correlation tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

Sampling frame and study selection

We constructed a sampling frame and selected



FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION



What are the benefits of dynamic reports?

- Data, code, analyses, and plots are all interconnected.
- The code *must* be correct because it is being run as the report is created.
- Anyone, even your future self, can replicate your analyses.
- If something changes in the data, the changes are automatically reflected in the results.
- If you get new data, you can easily create new analyses.

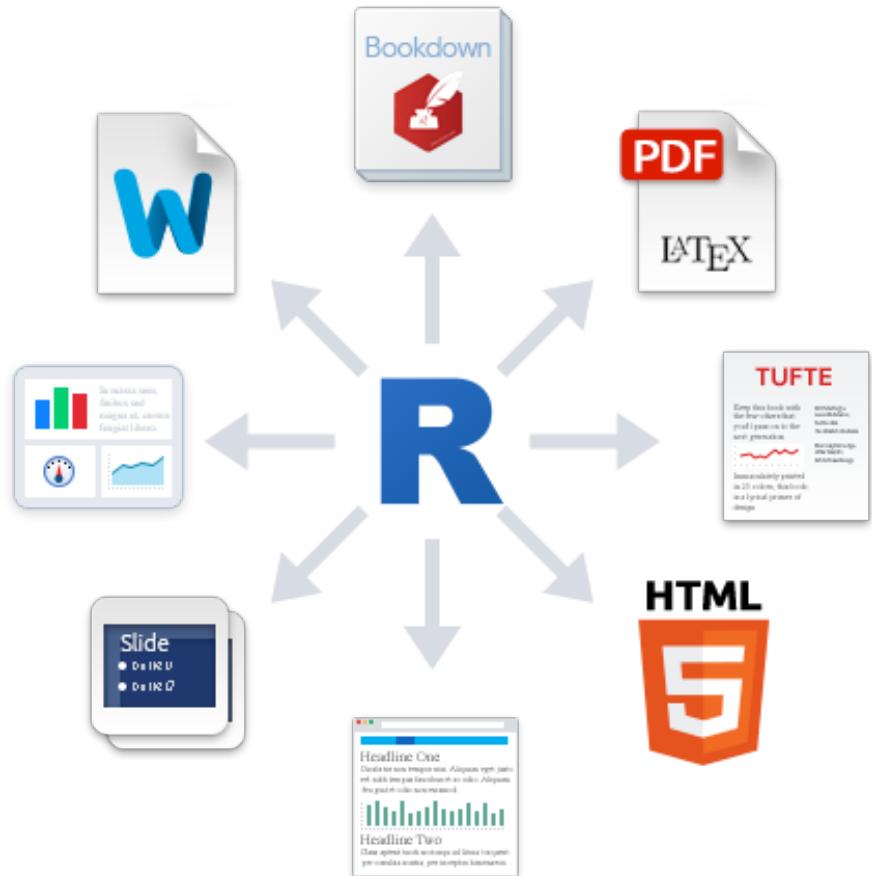
A screenshot of an RPubs page. At the top, there's a navigation bar with 'Sign in' and 'Register' buttons. The main header is 'RPubs' in a stylized orange and grey font, followed by the text 'brought to you by RStudio'. Below that, it says 'Last updated 22 days ago'. At the bottom of the header area are three buttons: 'Comments (-)', 'Share', and 'Hide Toolbars'. The main content area is mostly blank, suggesting the report has been removed or is loading.

<http://rpubs.com/fabiorocha5150/decisiontreemodel>

What dynamic reports can I make in RStudio?

- Web (HTML) pages
 - Websites, blog posts, single documents to share
- PDF documents
- Slideshows (like this one)
- Shiny Apps
- Books
- R Packages

How? Answer: R Markdown



What is Markdown?

- Markdown is a simple, readable **markup language** that allows you to easily write text with special formatting **tags**, which are then converted to formatted outputs.
- **R Markdown** is a combination of Markdown and R code.

The image shows a side-by-side comparison of an R Markdown file and its generated HTML. On the left, the 'example.Rmd' file contains the following content:

```
1 # Header 1
2
3 This is an R Markdown document. Markdown is a
4 simple formatting syntax for authoring webpages.
5 Use an asterisk mark to provide emphasis, such
6 as *italics* or **bold**.
7 Create lists with a dash:
8
9 - Item 1
10 - Item 2
11 - Item 3
12
13 ``
14 Use back ticks to
15 create a block of code
16 ``
17
18 Embed LaTeX or MathML equations,
19 $\frac{1}{n} \sum_{i=1}^n x_i$
20
21 Or even footnotes, citations, and a
22 bibliography. [^1]
23
24 [^1]: Markdown is great.
```

On the right, the generated 'example.html' page displays the same content in a browser:

Header 1

This is an R Markdown document. Markdown is a simple formatting syntax for authoring web pages.

Use an asterisk mark to provide emphasis, such as *italics* or **bold**.

Create lists with a dash:

- Item 1
- Item 2
- Item 3

Use back ticks to create a block of code

Embed LaTeX or MathML equations, $\frac{1}{n} \sum_{i=1}^n x_i$

Or even footnotes, citations, and a bibliography.¹

1. Markdown is great. ↵

Cheatsheet

R Markdown Reference Guide

Learn more about R Markdown at rmarkdown.rstudio.com
Learn more about Interactive Docs at shiny.rstudio.com/articles

Contents:

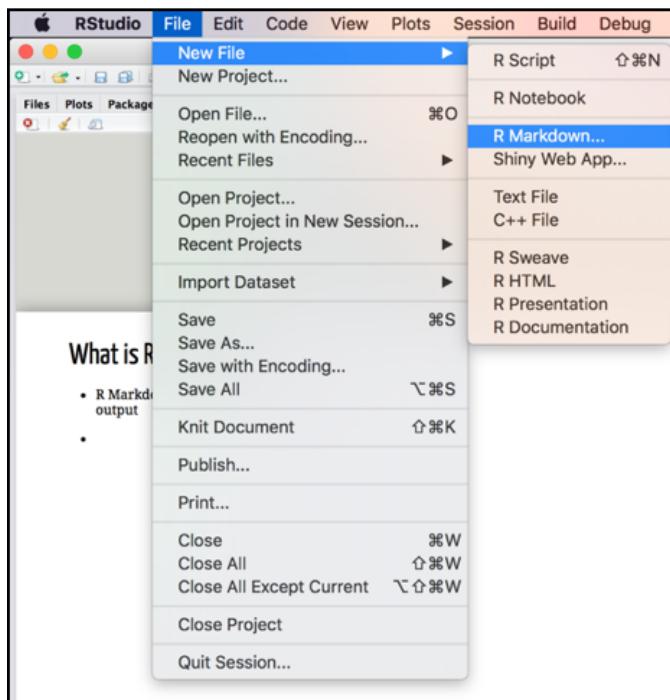
- 1. Markdown Syntax**
- 2. Knitr chunk options
- 3. Pandoc options

Syntax	Becomes
Plain text	Plain text
End a line with two spaces to start a new paragraph.	End a line with two spaces to start a new paragraph.
italics and _italics_	<i>italics</i> and <i>italics</i>
bold and __bold__	bold and bold
superscript ²	superscript ²
~~strikethrough~~	strikethrough
[link](www.rstudio.com)	link
# Header 1	Header 1
## Header 2	Header 2
### Header 3	Header 3
#### Header 4	Header 4
##### Header 5	Header 5
##### Header 6	Header 6
endash: --	endash: –
emdash: ---	emdash: —
ellipsis: ...	ellipsis: ...
inline equation: \$A = \pi r^2\$	inline equation: $A = \pi r^2$
image:	image: 
horizontal rule (or slide break):	horizontal rule (or slide break):

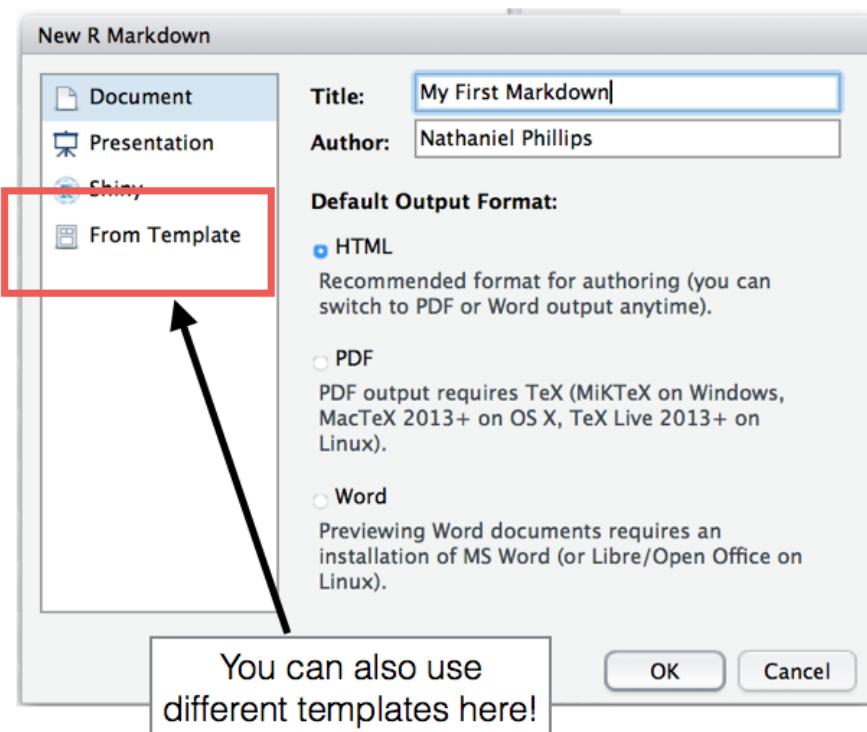
<https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>

How do I write an R Markdown Document

Open New Markdown
File in RStudio



Give it a name, and select
an optional output format



R Markdown = Markdown + R code

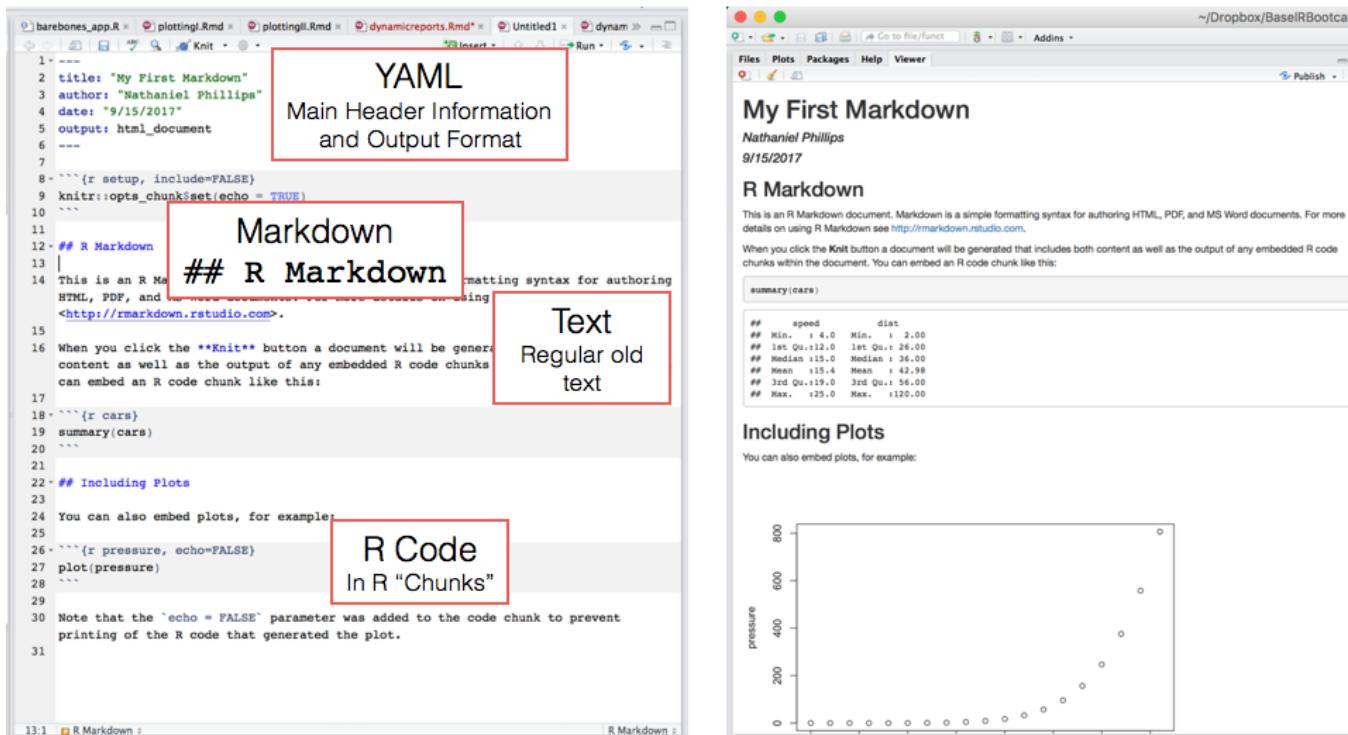
The screenshot shows the RStudio interface with the title bar "Clinical Study - RStudio". The main window displays an R Markdown document with the following content:

```
1 ---  
2 title: "Clinical Data Analysis"  
3 author: "Nathaniel Phillips"  
4 date: "9/15/2017"  
5 output:  
6   html_document: default  
7 ---  
8  
9 ## Clinical Study  
10  
11 ```{r, echo = FALSE, message = FALSE}  
12 # Load packages  
13 library(tidyverse)  
14 library(BayesFactor)  
15 library(lme4)  
16 ```  
17  
18 The study data is stored as a comma-separated text file called  
`clinical_study.csv`.  
19  
20 ```{r, echo = FALSE}  
21 # Load the data from the data folder  
22 clinical_study <- read_csv(file = "data/clinical_study.csv")  
23 ```  
24  
25 ### Primary Measures  
26  
27 The primary measure was whether patient's quality of life significantly  
improved. The following plot shows how quality of life changed over time,  
separated by different treatment groups:  
28  
29 ```{r, echo = FALSE}  
30 # Line graph showing QOL over time
```

The status bar at the bottom left shows "9:18" and "Clinical Study". The status bar at the bottom right shows "R Markdown".

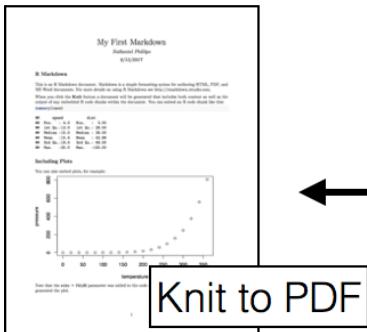
R Markdown = Markdown + R code

R Markdown “Knit” Output
.Rmd File .HTML File



R Markdown = Markdown + R code

example.pdf



Knit to PDF



example.Rmd

```
My First Markdown
Nathaniel Phillips
1/15/2017

R Markdown
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com. When you click the "Knit" button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```{r}
summary(mtcars)
```

# R Markdown
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com. When you click the "Knit" button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

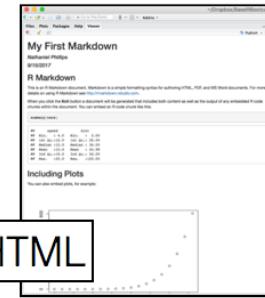
```{r}
summary(mtcars)
```

## Including Plots
You can also embed plots, for example:
```{r}
ggplot(mtcars, aes(wt, mpg))
 geom_point()
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```

example.html



Knit to HTML

Knit to Word



example.doc

R Markdown Gallery

<http://rmarkdown.rstudio.com/gallery.html>

R Markdown (<http://rmarkdown.rstudio.com/index.html>)

from (<https://www.rstudio.com/>)

Get Started (<http://rmarkd...>)

Gallery

Check out the range of outputs and formats you can create using R Markdown.

Documents

With R Markdown, you write a single .Rmd file and then use it to render finished output in a variety of formats.

Great NYT Interactive -- Now Reusable with rCharts

Disclaimer and Attribution

I claim absolute no credit for the visualization, which I consider one of the most beautiful I have ever seen. All credit belongs to the original author(s) of the visualization, whom I do not know. I will take it down immediately if I am explicitly given permission to link back to the data source(s) mentioned.

Another Favorite from NYT

I wrote all the code for this visualization myself. It is easily reusable. Earlier this year in my post <[A with rCharts and shiny](#)> I adapted and renamed the NYT code to be a shiny module to work with shiny. Unfortunately, I was not creative enough to think of other ideas and to pay for the visualization, when Scott Murray tweeted.

A Pandoc Markdown Article Starter and Template*

Steven V. Miller · Clemson University

This document provides an introduction to R Markdown, argue for its benefits, and presents a sample manuscript template intended for an academic audience. I include basic syntax to R Markdown, and provide a working example of how the analysis itself can be conducted within R with the knitr package.

Keywords: `pandoc`, `rmarkdown`, knitr

A Microsoft Word document

RStudio · June 3, 2016

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

What you see in **Knitr** when a document will be generated that includes both content as well as chunks of executable R code chunks within the document. You can control an R code chunk like this:

```
#> knitr::opts_chunk$set(echo = TRUE)
```

Introduction

The **Tufte** handout style is a style that Edward Tufte uses in his books and handouts. Tufte's style is known for its economy of space, and for the way it applies this principle with text, and well-set typography. This style has been implemented in LaTeX and HTML/CSS respectively. We have provided a **Tufte** template for Microsoft Word. If you want LaTeX/HTML output, you may use the **tufte-handout** format for handouts, and **tufte-note** for books. Both of these formats are based on the **Tufte** style. These formats can be either specified in the YAML metadata at the beginning of an R Markdown document (see an example below), or via the **knitr::opts_chunk\$set(tufte)** command. See Allaire et al. (2015) more information about handbooks.

*See GitHub repository [rmarkdown](https://github.com/rstudio/rmarkdown).

Including Plots

You can also embed plots, for example:

```
#> library(ggplot2)
```

Tufte Handout

An implementation in R Markdown

JJ Allaire and Yihui Xie · 2016-02-02

The **Tufte** handout style is a style that Edward Tufte uses in his books and handouts. Tufte's style is known for its economy of space, and for the way it applies this principle with text, and well-set typography. This style has been implemented in LaTeX and HTML/CSS respectively. We have provided a **Tufte** template for handouts, and **tufte-note** for books. Both of these formats are based on the **Tufte** style. These formats can be either specified in the YAML metadata at the beginning of an R Markdown document (see an example below), or via the **knitr::opts_chunk\$set(tufte)** command. See Allaire et al. (2015) more information about handbooks.

*See GitHub repository [rmarkdown](https://github.com/rstudio/rmarkdown).

(http://timelyportfolio.github.io/rmarkdown_gallery.html)

(<https://rstudio.github.io/tufte/>)

R Chunks in R Markdown

R Chunks look like this:

```
```{r, ARGUMENTS}
Code
1 + 1
```
```

You can include many arguments in a chunk

`eval = TRUE`

“Do evaluate this code”

`echo = FALSE`

“Do **NOT** print the code”

```
21
22 - ```{r, eval = TRUE, echo = FALSE, fig.width = 6, fig.height = 4}
23   library(tidyverse)
24
25 # Create line graph
26 ggplot(data = economics,
27         mapping = aes(x = date, y = unemploy)) +
28   geom_line() +
29   theme_bw()
30 ```
31
```

`fig.width = 6, fig.height = 4`
Figure size in output

R Chunks in R Markdown

- You can set default chunk options with `ops_chunk$set()`. When you do, all future chunks will have these settings (unless you specify otherwise)

```
# Example: Set default values for ALL future chunks
#   with knitr:::ops_chunk$set

knitr:::opts_chunk$set(fig.width = 6,           # Figure width (in)
                      fig.height = 6,        # Figure height (in)
                      echo = TRUE,          # Repeat code
                      eval = TRUE,           # Evaluate chunks
                      message = FALSE,       # Don't print messages
                      warning = FALSE,       # Don't print warnings
                      fig.align = 'center') # Center figures
```

- We recommend setting chunk options at the beginning of each document!

Inline chunks

- You can also include *inline chunks* where R code is included in a sentence. This allows you to include R output in your text!

39

40 Text text `r CODE` text text. Text text
text `r CODE` text text text.

41

Inline chunks

R Markdown with inline-chunks → **Knit!** → HTML Output!

```
31 ## Key Results
32
33 There were `r nrow(study_df)` patients in the
  clinical trial. Of these, `r sum(study_df$sex ==
  "f")` were female. The median age of the
  patients was `r median(study_df$age)` and the
  standard deviation was `r
  round(sd(study_df$age), 2)`.

34
35 The primary measure of treatment success was
  defined as patients who indicated that their
  quality of life had significantly improved as a
  result of treatment. Across all conditions, `r
  sum(study_df$success)` patients were successes.
```

Key Results

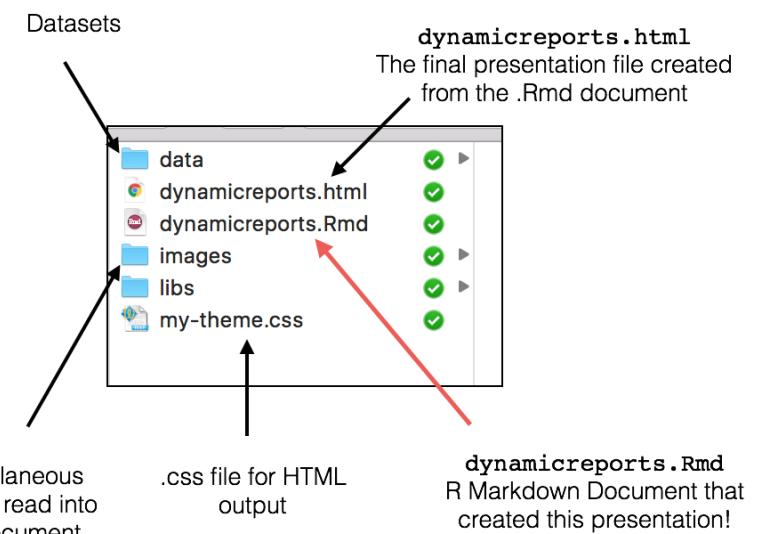
There were 100 patients in the clinical trial. Of these, 52 were female. The median age of the patients was 33 and the standard deviation was 9.77.

The primary measure of treatment success was defined as patients who indicated that their quality of life had significantly improved as a result of treatment. Across all conditions, 20 patients were successes.

Key points in R Markdown

1. You can easily read in external files like images, R code, or datasets.
2. All external files should be in a folder in the "Root" Directory containing the Markdown (.Rmd) file
3. All data necessary for your report *must* be explicitly read into the document. So keep them close by!

Here is the "Root" directory of the R Markdown file `dynamicreports.Rmd` that created *this presentation*



Rendering output with knitr



- To render output like images and tables into R Markdown, use the `knitr` package.
- The two most useful `knitr` functions are `include_graphics()` and `kable()`

| Function | Output |
|-------------------------------------|--|
| <code>include_graphics(path)</code> | Include an external image (e.g.; .png, .jpg) |
| <code>kable(df, format)</code> | Include a dataframe as a table |

Print a dataframe as a table with `kable()`

```
kable(economics[1:3, c("date", "pop")],  
      format = 'markdown')
```

| date | pop |
|------------|--------|
| 1967-07-01 | 198712 |
| 1967-08-01 | 198911 |
| 1967-09-01 | 199113 |

Include an image

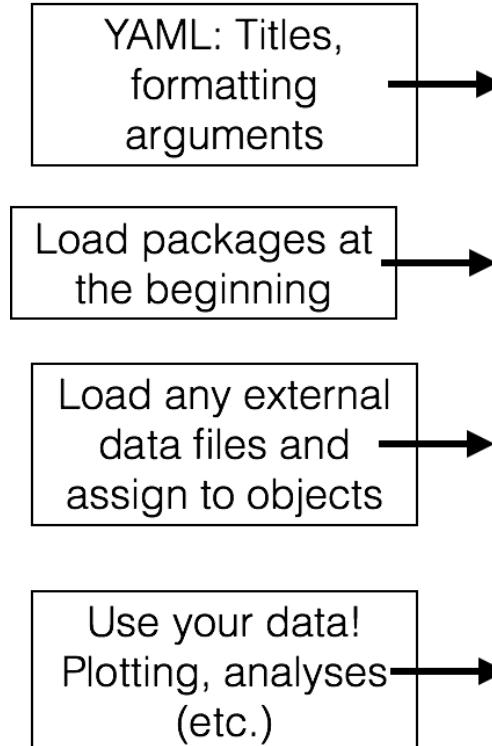
```
include_graphics(path = ".../rlogo.png")
```



Key points in R Markdown

1. When you "Knit" an R markdown document, it will start with an *empty* workspace (ie. it will forget everything!)
2. You must explicitly load all packages with `library()` and load in external datasets (e.g.;
`clinical_data <- read_csv("data/data.csv")`)
3. If you have any typos, errors, or missing code, the document will *not* knit (this is a *good* thing!)

Well formatted R Markdown document



```
BaselRBootcamp2017
md x plottingII.Rmd x dynamicreports.Rmd x myfirstmarkdown.Rmd x dynamicreports_prac >
1 --- 
2 title: "Clinical Data Analysis"
3 author: "Nathaniel Phillips"
4 date: "9/15/2017"
5 output:
6   html_document: default
7 ---
8
9 ## Clinical Study
10
11 ```{r, echo = FALSE, message = FALSE}
12 # Load packages
13 library(tidyverse)
14 library(BayesFactor)
15 library(lme4)
16
17
18 The study data is stored as a comma-separated text file called
`clinical_study.csv`.
19
20 ```{r, echo = FALSE}
21 # Load the data from the data folder
22 clinical_study <- read_csv(file = "data/clinical_study.csv")
23
24
25 ### Primary Measures
26
27 The primary measure was whether patient's quality of life significantly
improved. The following plot shows how quality of life changed over time,
separated by different treatment groups:
28
29 ```{r, echo = FALSE}
30 # Line graph showing QOL over time
31
32 ggplot(data = clinical_study,
33         mapping = aes(x = time, y = qol, col = arm)) +
34     geom_line()
35
36
37
```

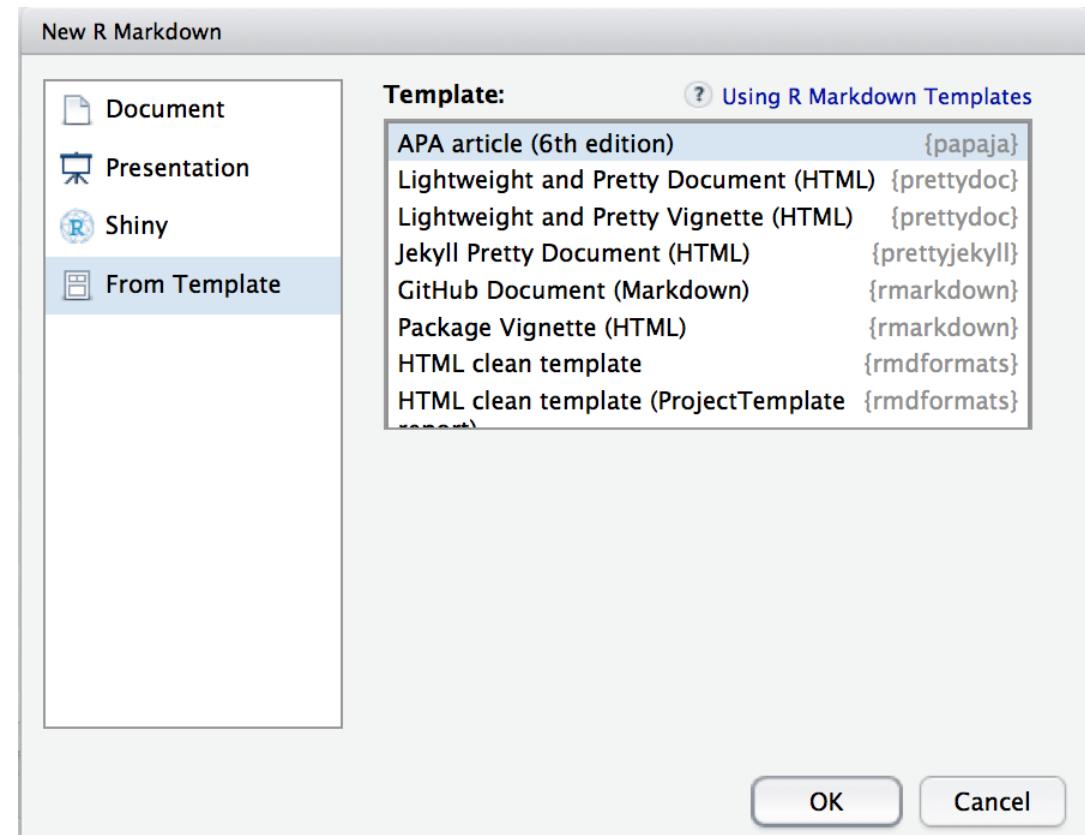
Output types

- There are *many* different output formats you can create from an R Markdown document
- Many come with RStudio, many are distributed in packages:

| Package | Description |
|------------|-----------------------------|
| xaringan* | Slideshows (like this one!) |
| papaja* | APA Manuscripts |
| rmdformats | Many templates |
| prettydoc | Many templates |

- = On GitHub

R Markdown templates in R Studio



To write to PDF, you need TeX

In order to knit your document to a PDF, you'll need a TeX installation



\documentclass[12pt]{article}
\title{\LaTeX: The better way}
\begin{document}
\maketitle

\section{Introduction}
\LaTeX is a robust typesetting language in which one can prepare publication-quality documents with ease. It allows you to shift your focus from the formatting of your document to the content. Results are consistent, compatible across a multitude of operating systems, and best of all, the programs are open source. Below are two examples of equations as generated by \LaTeX.

%This is comment text, it won't be visible in the final document, but is useful for annotation.

\section{Equations}
\subsection{pH equation}
\begin{equation}
pH = pK_a + log \frac{[R^-]}{[RH]}
\end{equation}
\subsection{Enzyme kinetics equation}

\begin{equation}
V = \frac{V_{max} [S]}{K_m + [S]}
\end{equation}

\LaTeX: The better way

1 Introduction

\LaTeX is a robust typesetting language in which one can prepare publication-quality documents with ease. It allows you to shift your focus from the formatting of your document to the content. Results are consistent, compatible across a multitude of operating systems, and best of all, the programs are open source. Below are two examples of equations as generated by L^AT_EX.

2 Equations

2.1 pH equation

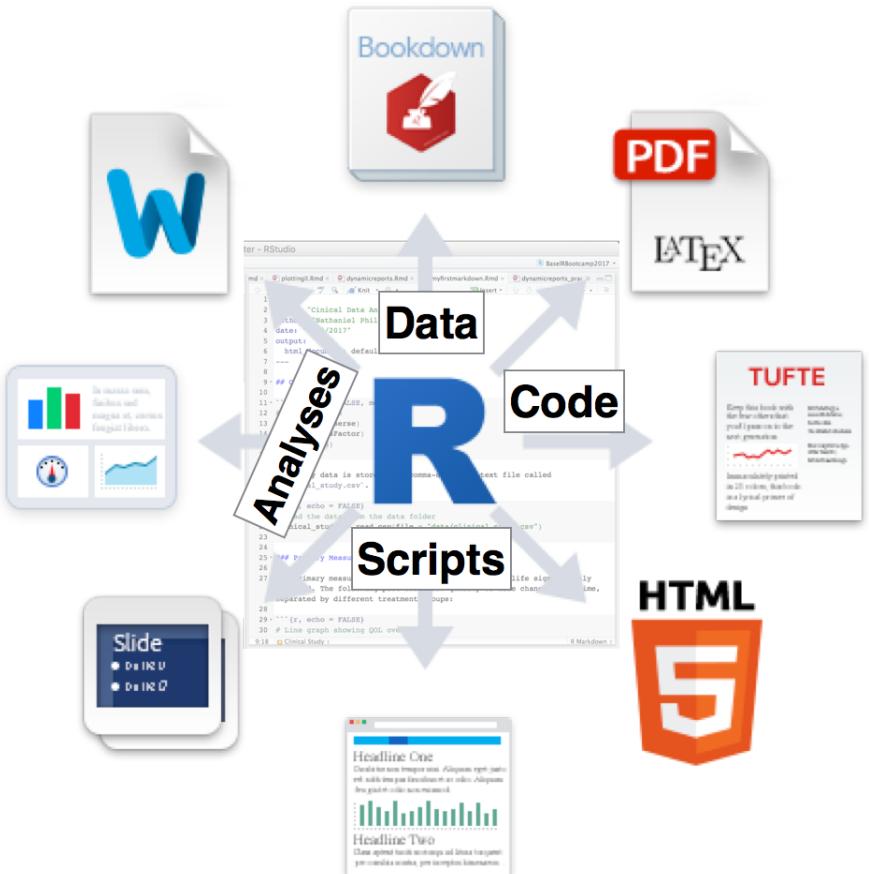
$$pH = pK_a + \log \frac{[R^-]}{[RH]} \quad (1)$$

2.2 Enzyme kinetics equation

$$\frac{1}{v} = \left(\frac{K_m}{V_{max}} \right) \left(1 + \frac{[I]}{K_i} \right) \left(\frac{1}{[S]} \right) + \frac{1}{V_{max}} \quad (2)$$

Why are RMarkdown documents so great?

1. The data, code, and output are all in the same place.
2. Everything works and is replicable! (If it wasn't, the document wouldn't *Knit*!)
3. You can produce great looking documents, from simple PDFs, to webpages, to presentations (like this one), to books.



R Packages - The ultimate dynamic report?

If you want a fully contained, dynamic report that contains data, code, statistics, plots, text, help files and documentation that you can easily share, your answer may be creating an **R Package**.

Installing an example package from GitHub

```
# Install cStudy2017 from github
install_github("bootcamp/cStudy2017")

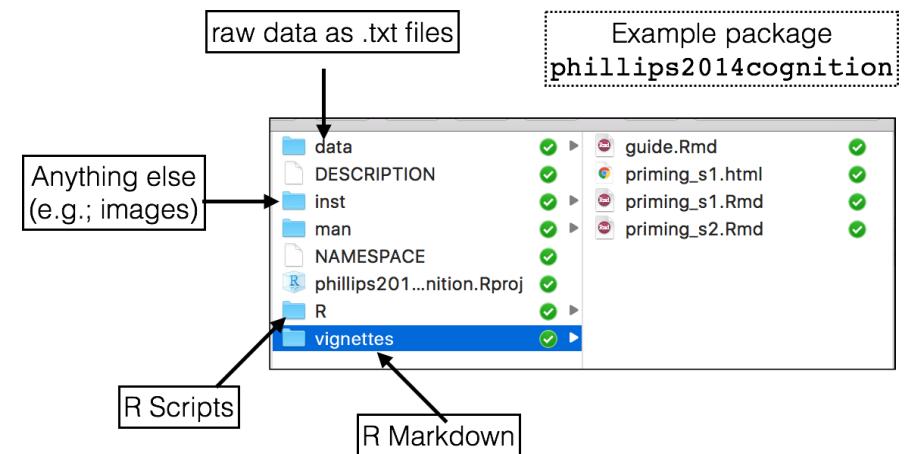
# Load cStudy2017 package
library("cStudy2017")

# Open cStudy2017 package help
?cStudy2017

# Oh study1_A contains study 1...
ggplot(data = study1_A,
       ...)

FFTrees(success ~ .,
        data = study1_A)
```

Folder structure for a simple package



Read Wickham's book "**R Packages**" (also available for free online) to learn how to write your own package

Dynamic Reports Pratical

[Link to Dynamic Reports practical](#)

