

R Markdown

Andrew Ba Tran

Contents

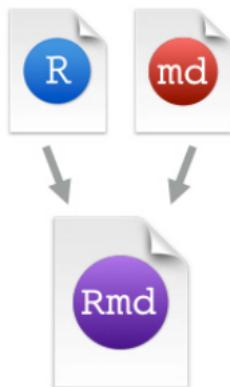
What is Markdown?	1
R Markdown / .Rmd files	2
Formats useful to journalists	3
Links not attachments!	7
Creating R Markdown	7

This is from the sixth chapter of learn.r-journalism.com.

By now you've inadvertently witnessed the different content that R Markdown can help generate:

- HTML documents that are in each chapter subfolder
- Notebooks in which you've run code chunks individually
- PDFs that you can print out to follow along physically with the course
- This entire learn.r-journalism.com website

R Markdown is R code + **Markdown**.



What is Markdown?

A very easy way to add formatting to plain text created by John Gruber (of Daring Fireball) as a simple way for non-programming types to write in an easy-to-read format that could be converted directly into HTML.

In the image below, the text on the left was interpreted as HTML on the right.

The screenshot shows a split-screen interface for editing Markdown. On the left, the 'MARKDOWN' tab is active, displaying the following code:

```

1 # This is a big headline (h1)
2
3 ## This is like h2
4
5 ### An h3 header
6
7 ## NICAR 2017 [Schedule](http://www.ire.org)
8 Friday - March 3
9
10 ----
11
12 ##### Homework
13 | | | |
14 |Type|Where|Details|
15 |---|---|---|
16 |Reading|The Data Journalism Handbook|[What is Data Journalism](http://datajournalismhandbook.org/1.0/en/introduction_0.html)|
17 |Reading|The Data Journalism Handbook|[Why journalists should use data](http://datajournalismhandbook.org/1.0/en/introduction_1.html)|
18 |Reading|The Data Journalism Handbook|[Why Data Journalism is important](http://datajournalismhandbook.org/1.0/en/introduction_2.html)|
19 |Reading|The Data Journalism Handbook|[Data journalism in perspective](http://datajournalismhandbook.org/1.0/en/introduction_4.html)|
20 |Reading|The Truthful Art|[Introduction: What is a journalist?]|()
21
22 **[Next class](class2.html)**

```

On the right, the 'PREVIEW' tab is active, showing the rendered content:

This is a big headline (h1)

This is like h2

An h3 header

NICAR 2017 [Schedule](http://www.ire.org)

Friday - March 3

Homework

Type	Where	Details
Reading	The Data Journalism Handbook	What is Data Journalism
Reading	The Data Journalism Handbook	Why journalists should use data
Reading	The Data Journalism Handbook	Why Data Journalism is important
Reading	The Data Journalism Handbook	Data journalism in perspective
Reading	The Truthful Art	Introduction: What is a journalist?

[Next class](#)

Github has a nice guide on [Markdown](#).

R Markdown / .Rmd files

It's a simple way to embed chunks of R code (or other languages like Python) in **Markdown** documents.

On one level, it allows you to present your analysis process and results in a format that doesn't require R to run.

But the code is still there should a researcher or fellow journalist be compelled to reproduce or add to your work.

It's a version of **literate programming**. By combining your R code with documentation, it makes your programming more robust, portable, and easier to maintain.

At the click of a button, or the type of a command, you can rerun the code in an R Markdown file to reproduce your work and export the results as a finished report.

R Markdown works thanks to the **knitr** package, which runs code embedded in Markdown, and Pandoc, which then converts Markdown into a bunch of different output formats, like Word, PDF, HTML, etc.

R Markdown supports dozens of static and dynamic output formats including

- HTML
- PDF
- MS Word
- Beamer
- HTML5 slideshows
- Tufte-style handouts
- Books
- Dashboards
- Scientific articles (white pages)

- Entire websites
-

Formats useful to journalists

Produce slick-looking PDFs

Exporting your work into PDFs can be effective.

For this to work, be sure to get LaTex installed first.

Sharing analyses with RMarkdown

Reporters are **very** impressed by this output and consider it very official.

*My theory is that it's the font. Doesn't it look **amazing**?*

IPython Notebooks

Notebooks are very popular.

This is how pandas and ipython notebooks render when uploaded to Github.

Branch: master · [homeless-arrests-analysis](#) · analysis.ipynb

P palewire Update analysis.ipynb · 0080ed9 3 days ago

1 contributor

1260 lines (1259 sloc) | 138 KB

[Raw](#) [Blame](#) [History](#)

L.A. homeless arrests analysis

By [Christine Zhang](#)

The Los Angeles Times analyzed daily arrest logs between January 1, 2011 and December 31, 2016 from the LAPD to determine yearly trends in arrests of homeless people.

The results were reported in a February 4, 2018, Los Angeles Times story titled "["L.A. leaders oppose 'criminalizing' homeless people. But thousands are jailed for minor offenses"](#)".

Here are the key findings of the data analysis, which is documented below with code written in the programming language R:

- The LAPD made 14,000 arrests of homeless people last year, a 30% increase over 2011
- LAPD arrests overall went down 15% from 2011 to 2016
- Two-thirds of those arrested were black or Latino
- In 2011, one in 10 people arrests citywide were of homeless people; in 2016, it was 1 in 6
- The 14,000 arrests of homeless people in 2016 included more than 500 unique charges
- By far the most common was failing to appear in court on an unpaid infraction or misdemeanor citation
- The top five charges were for non-violent or minor offenses

Read more about the methodology [here](#).

How we did it

Import R data analysis libraries

```
In [20]: library('dplyr')
library('feather')
library('ggplot2')
```

The file `arrests.feather.zip` must first be unzipped to run this notebook. This is a file that has been prepared outside of this notebook as part of an unpublished processing script. The raw LAPD data includes names and other identifying information about arrestees that The Times has chosen to withhold.

It renders well and matches the experience of someone coding in that environment.

.Rmd files rendered on Github

How does an R Markdown file look on Github?

Branch: master [▼](#) NICAR / 2017 / reproducibility-rmarkdown / rmarkdown.Rmd [Find file](#) [Copy path](#)

 andrewbtran trimmed decf6f9 on Mar 2, 2017

1 contributor

228 lines (132 sloc) 6.59 KB

[Raw](#) [Blame](#) [History](#)   

title	author	output
Reproducibility and R Markdown	Andrew Tran	<pre>html_document toc toc_float true true</pre>

- Creating functions and packages [\[link\]](#)
- Intro to R Markdown [\[link\]](#)

Introduction to R Markdown

Markdown

Super simple way to add formating to plain text

- headers
- bold
- bullet lists

Not well.

Each .Rmd file has its own custom YAML section at the top. These are keywords that when combined with the right packages, lets **knitr** know how to output the .Rmd file. Like `toc: toc_float` creates an HTML file with a self-generated Table of Contents based on the header titles.

Github doesn't have a way to interpret that so it creates that nested image above and doesn't even try to make a table of contents.

That's fine, though

Output .Rmd files to HTML

R Markdown will let you output as HTML, which you can still host on Github Pages (which we'll go over later). You may have to include links to the actual page but doing so can be more effective than having a .Rmd file "render" in a Github repo.

Rendering in HTML lets you add CSS and Javascript, allowing for the inclusion of content like the interactive table below.

```
library(DT)
library(dplyr)
library(readr)
payroll <- read_csv("../data/bostonpayroll2013.csv") %>%
  select(NAME, TITLE, DEPARTMENT, REGULAR, OVERTIME) %>%
  filter(row_number()<100)
```

```
datatable(payroll, extensions = 'Buttons', options = list(
  dom = 'Bfrtip',
  buttons = c('copy', 'csv', 'excel', 'pdf', 'print')
)
)
```

[Copy](#)
[CSV](#)
[Excel](#)
[PDF](#)
[Print](#)
Search:

	NAME	TITLE	DEPARTMENT	REGULAR	OVERTIME
1	Darosa, Baltazar	Police Officer	Boston Police Department	\$66,829	\$10,473
2	Creaven, Jacqueline D	Police Lieutenant	Boston Police Department	\$104,661	\$47,572
3	Cawley, Stephen C	Police Lieutenant	Boston Police Department	\$44,632	\$10,158
4	Danilecki, John H	Police Captain	Boston Police Department	\$121,244	\$71,198
5	Kervin, Timothy M.	Police Lieutenant/Hdq Dispatch	Boston Police Department	\$104,588	\$88,906
6	Hosein, Haseeb	Police Lieutenant	Boston Police Department	\$105,261	\$47,437
7	Lee, Thomas F	Police Captain/DDC	Boston Police Department	\$106,048	\$56,418
8	Assad, Mark L	Police Sergeant (Det)	Boston Police Department	\$94,641	\$79,482
9	Eversley, Eric V	Police Lieutenant (Det)	Boston Police Department	\$106,564	\$66,888
10	Crossen, Patrick J	Police Captain/DDC	Boston Police Department	\$122,705	\$90,319

This table was rendered with the **DT** package.

And we don't have to create the code to display it.

It just worked because the folks who ported the DataTables jquery plugin over to R wanted to make it seamless.

Pass certain arguments to the function **datatable()** and you can include buttons that allow reporters to download your tables as CSVs.

Can you imagine the power of that?

You've transformed the data collected it and then instead of sending them a huge spreadsheet, you send them a link to your report in which they can filter things out and download the table that they've come up with themselves.

You can render other interactives, as well, like leaflet maps.

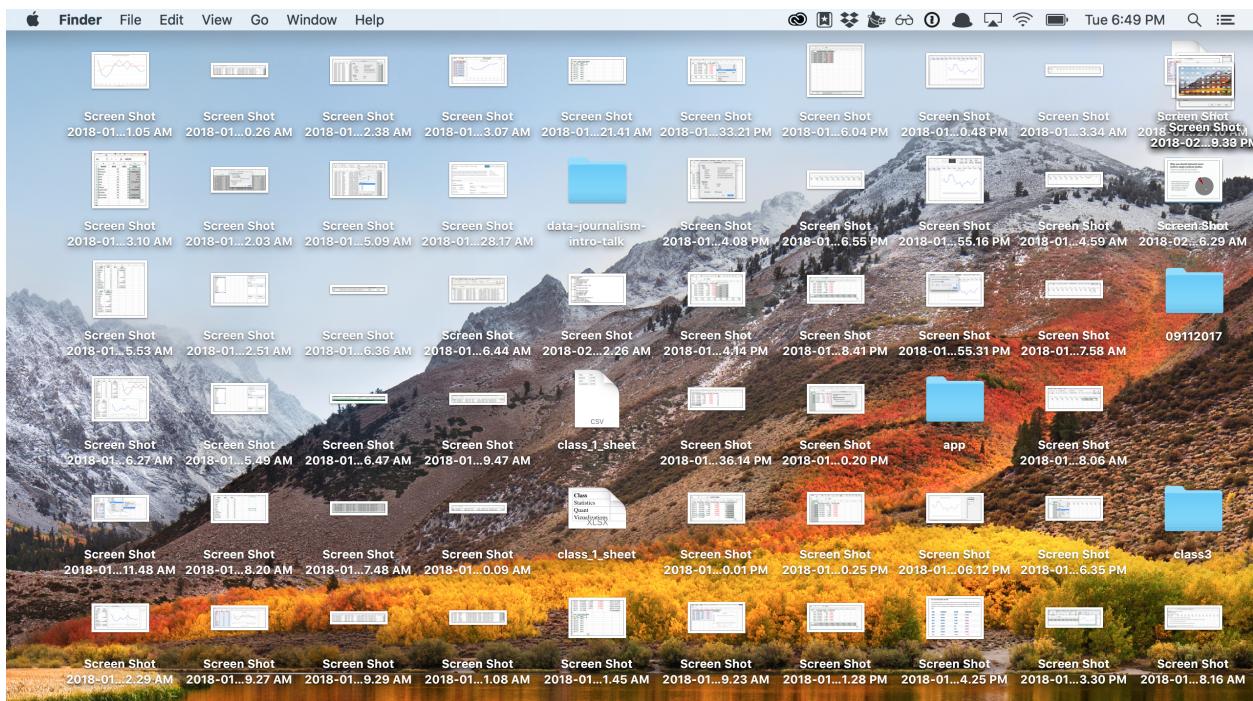
Links not attachments!

This is a very important development as a journalist.

You see, reporters sometimes aren't very organized.

How often have you had someone email you asking if you could resend your summarized spreadsheet to them again?

This is what their downloads or desktop folder might look like. It's easy to see why they might have lost track of it.



Send them links instead!

Host the HTML of your analysis

- On your internal work network or
- On Github pages (for free)
- Share with reporters and editors, let them explore your analysis further
- Let them download customized spreadsheets with buttons in the DT (datatable) package
- If your analysis gets updated, keep the file name then they only have to refresh their link

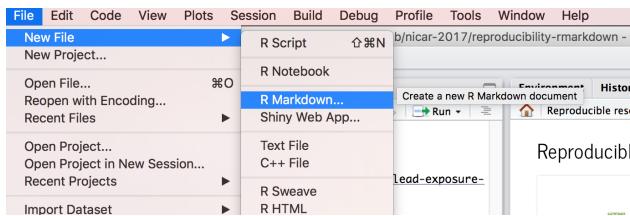
Repeat this process enough and maybe you can start learning how to create Shiny apps with R.

Creating R Markdown

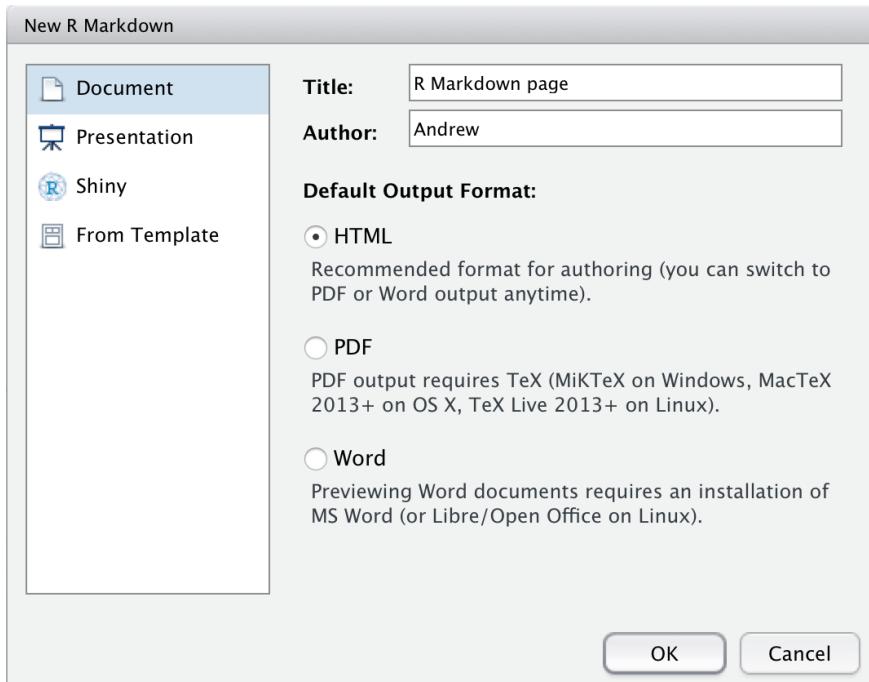
Follow the instructions below.

1. Open a new .Rmd file

at **File > New File > R Markdown**.



Title the R Markdown file and select **HTML** as the output for now.



The default file created comes with code already included to give you a guidance on syntax and how it should be structured.

.Rmd structure

```

1 ---  

2 title: "R Markdown page"  

3 author: "Andrew"  

4 date: "2/23/2017"  

5 output: html_document  

6 ---  

7  

8 ````{r setup, include=FALSE}  

9 knitr::opts_chunk$set(echo = TRUE)  

10 ````  

11  

12 ## R Markdown  

13  

14 This is an R Markdown document. Markdown is  

authoring HTML, PDF, and MS Word documents.  

see <http://rmarkdown.rstudio.com>.  

15  

16 When you click the **Knit** button a docume  

content as well as the output of any embedc  

You can embed an R code chunk like this:  

17  

18 ````{r cars}  

19 summary(cars)  

20 ````  

21

```

Every .Rmd file starts out the same at the top.

YAML HEADER

Optional section of render options written as **key:value** pairs.

- At start of file
- Between lines of — (3 dashes)

This is how R knows how to render your files. In the example above, the line `output: html_document` means it will create an HTML file. If it was `output: pdf` then you'd get a PDF instead.

You can output more than one type at a time, if you want.

CODE CHUNKS

In the image above, the R code is in the gray area.

These are chunks of embedded R code. Each chunk:

- begins with “`{r}`” (the back tick key is to the left of 1)
- ends with “`‘`”

TEXT

The rest of it is just the narrative of markdown interspersed in between the R code chunks.

You can also use HTML code instead of Markdown if that's what you're more comfortable with.

2. Write document

Edit the default template by putting in your own code and text.

Intersperse the text with your code to tell a story.

2b. Label your chunks of code

It'll be easier to navigate to in the future with the pulldown menu at the bottom of the script window.

It also will be useful to narrow down where code errors might occur.

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## 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 cars}
summary(cars)
```

## Including Plots

You can also embed plots, for example:

```{r pressure, echo=FALSE}
plot(pressure)
```

```

2c. Notebooks style

You can run individual chunks of code before generating the full report to see how it looks.

Click the green arrow next to each chunk.

Hopefully, you've already been doing this in our coding walkthroughs in class.

```
8+ ```{r cars}
9 summary(cars)
```
speed dist
Min. :4.0 Min. : 2.00
1st Qu.:12.0 1st Qu.: 26.00
Median :19.0 Median : 36.00
Mean :15.4 Mean : 36.98
3rd Qu.:19.0 3rd Qu.: 56.00
Max. :25.0 Max. :120.00
```

## 3. Knit document to create report

Use knit button (the yarn ball!) at the top left or type **render()** to create the output file(s).

```

1 -> tit Knit to HTML
2 aut Knit to PDF
3 dat Knit to Word
4 out
5 --- Knit with Parameters...
6 ...
7 ...
8 ... Clear Knitr Cache...
9 knitr::opts_chunk$set(echo = TRUE)
.0
.1
.2 ## R Markdown
.3
.4 This is an R Markdown document. Markdown is a simple formatting syntax for au
documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>

```

### 3b. Check out the build log

Warnings and errors will appear in the console below.

Also measures progress by chunks, which is why it's important to label them.

```

Console R Markdown x
.../reproducibility-rmarkdown/testpage.Rmd

processing file: testpage.Rmd
logi FALSE

|..... ordinary text without R code | 43%
label: cars
|..... ordinary text without R code | 57%
|..... ordinary text without R code | 71%
label: pressure (with options)
List of 1
$ echo: logi FALSE

|..... ordinary text without R code | 100%

output file: testpage.knit.md

/Applications/RStudio.app/Contents/MacOS/pandoc +RTS -K512m -RTS testpage.utf8.md --to html --from markdown+tolink_bare_uris+asci_identities+tex_math_single_backslash --output testpage.html --smart --email-obfuscation nor--self-contained --standalone --section-divs --template /Library/Frameworks/R.framework/Versions/3.3/Resources/libr/y/rmarkdown/rmd/h/default.html --no-highlight --variable highlightjs=1 --variable 'theme:bootstrap' --include-in-he er /var/folders/7g/5fxswp_n6092hwr8n4wkqcfw0000gq/T//RtmpwaRzML/rmarkdown-str120fb28c68d6d.html --mathjax --variabl 'mathjax-url:https://cdn.mathjax.org/mathjax/latest/MathJax.js?config=TeX-AMS-MML_HTMLorMML'

Output created: testpage.html

```

#### 4. Preview output in IDE window

The screenshot shows a web browser window displaying an R Markdown document. The title is "R Markdown page". Below it, the author is listed as "Andrew" and the date as "2/23/2017". A section titled "R Markdown" follows. A note states: "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>". Another note says: "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:".

```
summary(cars)
```

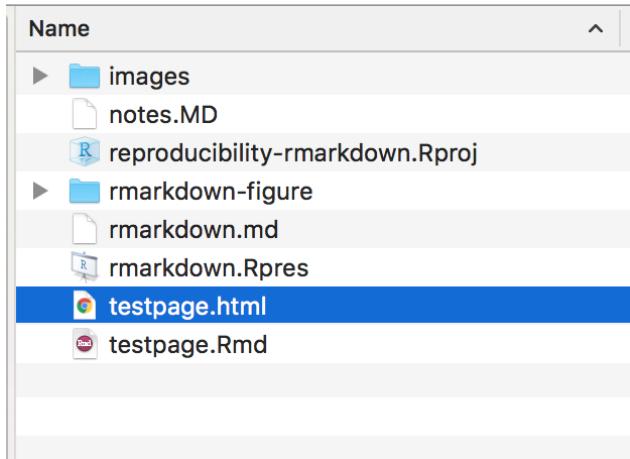
```
speed dist
Min. : 4.0 Min. : 2.00
1st Qu.:12.0 1st Qu.: 26.00
Median :15.0 Median : 36.00
Mean :15.4 Mean : 42.98
3rd Qu.:19.0 3rd Qu.: 56.00
Max. :25.0 Max. :120.00
```

### Including Plots

You can also embed plots, for example:

#### 5. Output file

You started with a .Rmd file and clicking knit HTML also generated a .html file.



It's all there in your project folder now.

Next, we'll go over more specific features in R Markdown files that will give you more control on how its presented.