

Markdown and R

Roger D. Peng

UW SISBID

July 2015

What is Markdown?

“Markdown is a text-to-HTML conversion tool for web writers. Markdown allows you to write using an easy-to-read, easy-to-write plain text format, then convert it to structurally valid XHTML (or HTML).”

John Gruber, creator of Markdown
daringfireball.net

What is Markdown?

“The idea is that a Markdown-formatted document should be publishable as-is, as plain text, without looking like it’s been marked up with tags or formatting instructions.”

John Gruber, creator of Markdown
daringfireball.net

What is Markdown?

- Markdown was originally a Perl script written by John Gruber, but now there many processors
- RStudio has built-in support for Markdown documents
- R package vignettes can be built using Markdown
- GitHub documents automatically processed, many other services support Markdown docs

Why Markdown is Awesome

- Text based, like writing an email
- A simplified version of “markup” languages
- Allows one to focus on **writing** as opposed to formatting
- Simple/minimal intuitive formatting elements; manual is easy to memorize
- Easily converted to valid HTML (and other formats) using existing tools

Title Heading

The Massive Future of Statistics Education

Data are eating the world, but our collective ability to analyze data is going on a starvation diet.

Everywhere you turn, data are being generated somehow. By the time you read this piece, you'll probably have collected some data. (For example this piece has 2,072 words). You can't avoid data---it's coming from all directions.

Text!

Title Heading

The Massive Future of Statistics Education

Data are eating the world, but our collective ability to analyze data is going on a starvation diet.

Section Heading

Everywhere you turn, data are being generated somehow. By the time you bly have collected some data. (For example this can't avoid data---it's coming from all directions.

Massive Open Online Courses (MOOCs)

In April of 2014, Jeff Leek, Brian Caffo, and I launched the [Johns Hopkins Data Science Specialization](<https://www.coursera.org/specialization/jhudatascience/1>) on the Coursera platform. This is a sequence of nine courses that intends to provide a "soup-to-nuts" training in data science for people who are highly motivated and have some basic mathematical and computing background. The sequence of the nine courses follow what we believe is the essential "data science process", which is

1. Formulating a question that can be answered with data
2. Assembling, cleaning, tidying data relevant to a question
3. Exploring data, checking, eliminating hypotheses
4. Developing a statistical model
5. Making statistical inference
6. Communicating findings
7. Making the work reproducible

Text!

Web link

Ordered list

Markdown Mindset

- Designed for writing on the Web, so many features of other markup languages are not available
- No notion of a “page”, or of page layout in general
- Let go of your controlling side
- Most processors (e.g. pandoc, RStudio) can integrate other markup, like LaTeX
- Formatting can be controlled with CSS

R Markdown (.Rmd)

- R markdown is the integration of R code with markdown
- Allows one to create documents containing "live" R code
- R code is evaluated as part of the processing of the markdown
- Results from R code are inserted into markdown document
- A core tool in literate statistical programming

R Markdown (.Rmd)

- R markdown can be converted to standard markdown using the knitr package in R
- Markdown can be converted to HTML using the markdown package in R
- Any basic text editor can be used to create a markdown document; no special editing tools needed
- The R markdown --> markdown --> HTML work flow can be easily managed using R Studio

Headings

This is a primary heading

This is a secondary heading

This is a tertiary heading

Text

Here is some text!

Italics

This text will be italicized
so will this text

Bold

****This text will appear bold!****
__if you like underscores__

Unordered Lists

Here's an unordered list of items:

- * The first item
- * The second item
- * The third item
- * etc.

Ordered Lists

Here's an ordered list of things:

1. The most important thing
2. The second most important thing
4. Whoops, forgot one!
3. The fourth most important thing

Don't worry about the ordering of the numbers!

Links

The #1 school of public health is the [\[Johns Hopkins Bloomberg School of Public Health\]](http://www.jhsph.edu)

Go here to [\[Download R\]](http://cran.r-project.org)

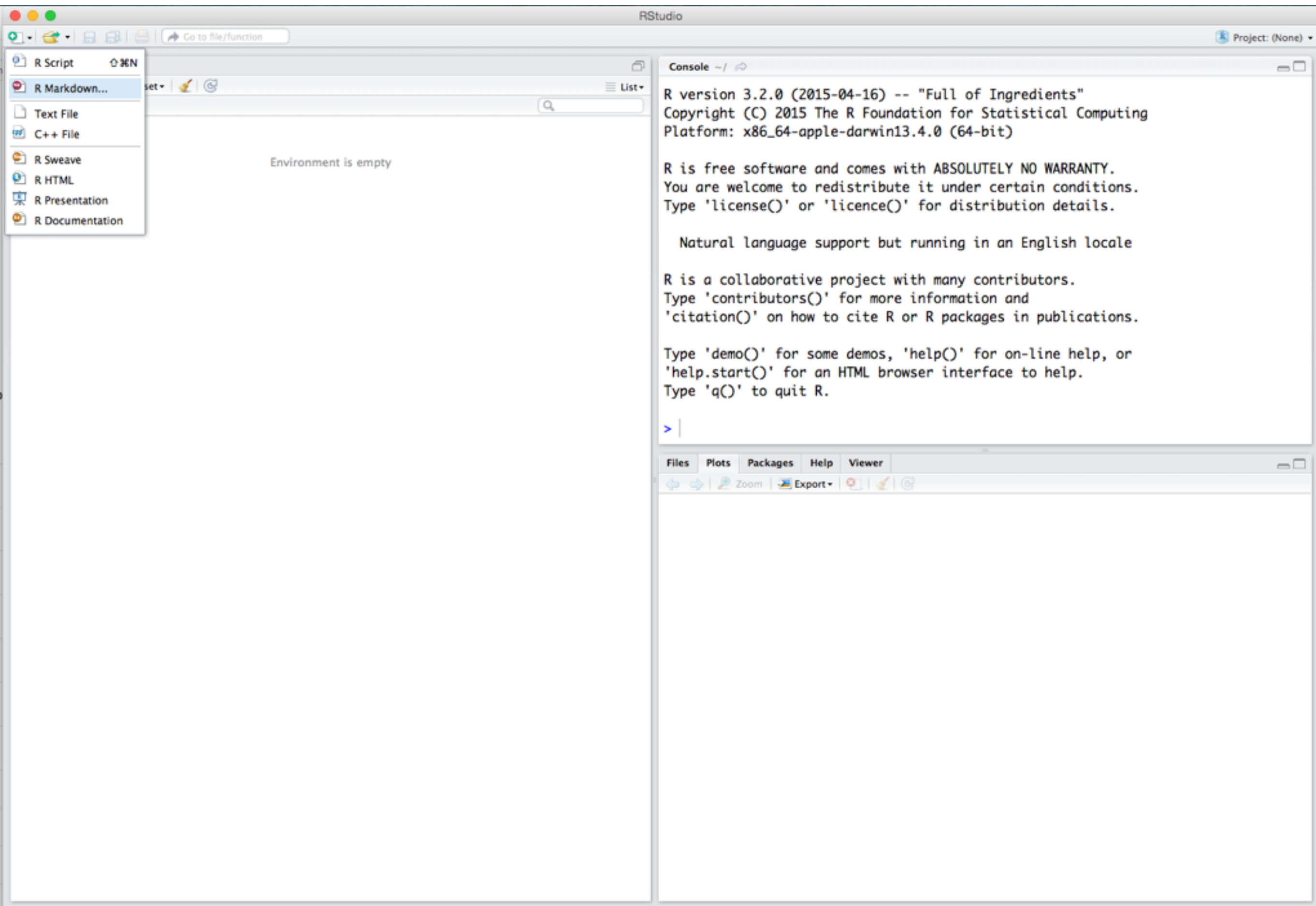
(Sometimes useful to use a link shortener for readability)

Blockquote

Here are some words from an important person:

> Always remember to eat your peas

Demo



New R Markdown



Document



Presentation



Shiny



From Template

Title:

Untitled

Author:

Roger Peng

Default Output Format:

☒ **HTML**

Recommended format for authoring (you can switch to PDF or Word output anytime).

☐ **PDF**

PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

☐ **Word**

Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

OK

Cancel

RStudio

Go to file/function

Project: (None)

Untitled1 x

Knit HTML Run Chunks

```
1 ---
2 title: "My First R Markdown Document"
3 author: "Roger Peng"
4 date: "July 3, 2015"
5 output: html_document
6 ---
7
8 This is an R Markdown document. Markdown is a simple formatting syntax for
9 authoring HTML, PDF, and MS Word documents. For more details on using R
10 Markdown see <http://rmarkdown.rstudio.com>.
11
12 When you click the Knit button a document will be generated that
13 includes both content as well as the output of any embedded R code chunks
14 within the document. You can embed an R code chunk like this:
15
16 ```{r}
17 summary(cars)
18 ```
19
20 You can also embed plots, for example:
21
22 ```{r, echo=FALSE}
23 plot(cars)
24 ```
25
26 Note that the `echo = FALSE` parameter was added to the code chunk to
27 prevent printing of the R code that generated the plot.
```

Console

```
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

Files Plots Packages Help Viewer

Zoom Export

2:1 (Top Level) R Markdown

Environment History

document.Rmd

Knit HTML

Run

Chunks

```
1 ---
2 title: "My First R Markdown Document"
3 author: "Roger Peng"
4 date: "July 3, 2015"
5 output: html_document
6 ---
7
8 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>.
9
10 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:
11
12 ```{r}
13 summary(cars)
14 ```
15
16 You can also embed plots, for example:
17
18 ```{r, echo=FALSE}
19 plot(cars)
20 ```
21
22 Note that the `echo = FALSE` parameter was added to the code chunk to
prevent printing of the R code that generated the plot.
23
```

2:1 (Top Level)

R Markdown

Environment

History

Console

Files Plots Packages Help Viewer

My First R Markdown Document

Roger Peng

July 3, 2015

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:

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

You can also embed plots, for example:

