

Markdown and R

Roger D. Peng

Biostatistics 140.776

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 are 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

Section Heading

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 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 **rmarkdown** 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
This is a ??? heading

Text

Here is some text!

Italics

This text will be italicized

Bold

****This text will appear bold!****

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

Summary

- Markdown is a writing system that is focused on plain text
- Content is king; formatting is someone else's job
- Many web site automatically render Markdown to HTML
- RStudio has built-in support via knitr and rmarkdown packages (other editors are available for plain Markdown)
- There are many “flavors” of Markdown (i.e. GitHub, Markua, pandoc)