

Simple Reproducible Analysis with knitr, R Markdown, and RStudio

Melbourne R Users Group (melbURN)

Jeromy Anglim

Melbourne Business School

18th July 2012

- <http://jeromyanglim.blogspot.com>
- <https://github.com/jeromyanglim/rmarkdown-rmeetup-2012>

Outline

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

1 Introduction

2 Markdown

3 knitr and R Markdown

4 LaTeX

5 Conclusion

Motivation: How to create documents?

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

■ Types and distinctions

- Formal Documents: Journal articles, books, book chapters, theses, consulting reports, etc.
- Informal documents: preliminary analyses, statistical homework,
- Online content: web pages, blog posts, forum posts
- Browser metaphor versus page/slide-based metaphor

■ Context

- When to use reproducible analysis?
- When to use knitr with R Markdown or LaTeX?

What is *reproducible analysis*?

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- Reproducibility varies on a continuum
- One particular form:
 - code transforms raw data and meta-data into processed data,
 - code runs analyses on the data, and
 - code incorporates analyses into a report
- Ideally, the process involves a one-click build
- Public sharing of document, code, and data is optional, but forms part of gold standard of scientific openness
- Goes by many names, particularly “reproducible research”, but I prefer “reproducible analysis”.

See also: <http://stats.stackexchange.com/a/15006/183>
<https://github.com/jeromyanglim/rmarkdown-rmeetup-2012/issues/11>

Aims of reproducible analysis

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- Ability to reproduce analysis
- Increase accuracy
 - Ability to verify analyses are consistent with intentions
 - Ability to review analysis choices
- Increase clarity of communication
- Increased trustworthiness
 - Increased accuracy +
 - Ability for others to verify
- Extensibility
 - Ability to easily modify or re-use existing analyses

Reproducible analysis in R

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

Typically:

- Combine R and plain text file format to produce documents (e.g., pdfs, HTML documents, etc.)

Popular Instances

- Sweave
- brew
- knitr

see also <http://cran.r-project.org/web/views/ReproducibleResearch.html>

Installation of software used in this talk

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- R: <http://www.r-project.org/>
- R Studio: <http://rstudio.org/>
- In R:
 - `install.packages("knitr")`
 - `install.packages("markdown")`
 - `install.packages("xtable")`
 - `install.packages("ggplot2")`
 - `install.packages("lattice")`
- pandoc:
 - <http://johnmacfarlane.net/pandoc/>
- LaTeX distribution:
 - E.g., TeXLive, MikTeX
<http://www.latex-project.org/ftp.html>

What is markdown?

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- Simple, readable, intuitive, light-weight markup
- Convert to HTML
- Raw HTML can be interspersed to add functionality
- Various extensions and flavours of markdown
- Popular on websites: e.g., StackOverflow, GitHub, Reddit

see also: <http://daringfireball.net/projects/markdown/>

Headings

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

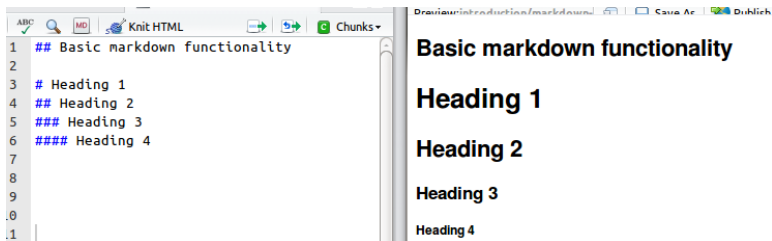
Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion



Basic formatting

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

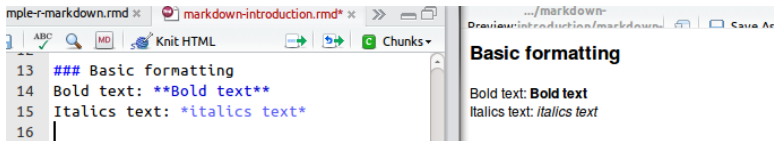
Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion



Paragraphs

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

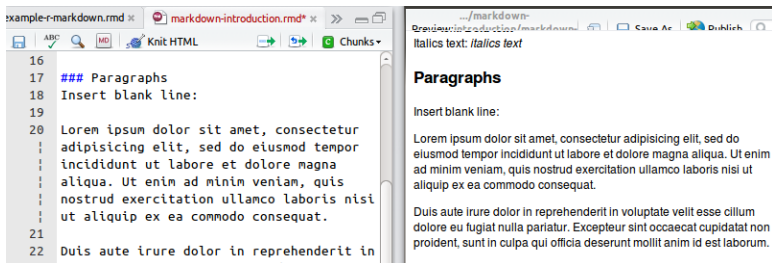
Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion



Dot points

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

```
5 ### Dot Points
6 Simple dot points:
7
8 * Point 1
9 * Point 2
0 * Point 3
1
2 and numeric dot points:
3
4 1. Number 1
5 2. Number 2
6 3. Number 3
7
8 and nested dot points:
9
0 * A
1     * A.1
2     * A.2
3 * B
4     * B.1
5     * B.2
6
7
```

Dot Points

Simple dot points:

- Point 1
- Point 2
- Point 3

and numeric dot points:

1. Number 1
2. Number 2
3. Number 3

and nested dot points:

- A
 - A.1
 - A.2
- B
 - B.1
 - B.2

Equations

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

```
9 ### Equations
9 Uses Mathjax to support LaTeX equations.
1
2 Inline equations: e.g.,  $y_i = \alpha + \beta x_i + e_i$ .
3
4 Displayed equations:
5
6 
$$\frac{1}{1 + \exp(-x)}$$

7
8 
$$\frac{1}{1 + \exp(-x)}$$

9
```

Equations

Uses Mathjax to support LaTeX equations.

Inline equations: e.g., $y_i = \alpha + \beta x_i + e_i$.

Displayed equations:

$$\frac{1}{1 + \exp(-x)}$$

- Uses MathJaX to render LaTeX (and other) equations
- Inserts MathJaX script reference into HTML header

getting started: <http://jeromyanglim.blogspot.com.au/2010/10/getting-started-with-writing.html>

Hyperlinks

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

```
2 ### Hyperlinks
3
4 * [my RSS feed](http://feeds.feedburner
5 .com/jeromyanglim).
6
7 * <http://www.r-project.org/>
```



Hyperlinks

- [my RSS feed.](http://feeds.feedburner.com/jeromyanglim)
- <http://www.r-project.org/>

Images

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

```
1  ### Images|
2
3  (figure/building
4  s.jpg)
```

Images



Code

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

Code

Inline code between backticks: e.g.,
``print('hello world!')``.

Displayed code can be tab indented or
four space indented:

```
```{r}
x <- 1:10
x
```
```

Code

Inline code between backticks: e.g.,
`print('hello world!')`.

Displayed code can be tab indented or four
space indented:

```
```{r}
x <- 1:10
x
```
```


Quotes

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

```
### Quote  
Quotes by adding greater than to start of  
each line.
```

```
> To be, or not to be, that is the  
question:  
> Whether 'tis nobler in the mind to  
suffer  
> The slings and arrows of outrageous  
fortune,
```

Quote

Quotes by adding greater than to start of each line.

To be, or not to be, that is the
question:
Whether 'tis nobler in the mind to
suffer
The slings and arrows of
outrageous fortune,

Tables

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

```
0 ### Tables
1 Extended github table functionality:
2
3 A | B | C
4 ---|---|---
5 1 | Male | Blue
6 2 | Female | Pink
7
8 Or just write HTML:
9
10 <table border="1">
11   <tr><td>Cell A1</td>|
12     <td>Cell B1</td></tr>
13   <tr><td>Cell A2</td>
14     <td>Cell B2</td></tr>
15 </table>
```

Tables

Extended github table functionality:

| A | B | C |
|---|--------|------|
| 1 | Male | Blue |
| 2 | Female | Pink |

Or just write HTML:

| | |
|---------|---------|
| Cell A1 | Cell B1 |
| Cell A2 | Cell B2 |

Raw HTML

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

```
### HTML is passed through  
Hyperlink  
<a href="http://jeromyanglim.blogspot  
.com">My website</a>
```

E.g., new line
<hr />

HTML Symbol Entities

α β ™

HTML is passed through

Hyperlink
[My website](http://jeromyanglim.blogspot.com)

E.g., new line

HTML Symbol Entities

α β TM

knitr, R Markdown, and R Studio

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- knitr: R Package developed by Yihui Xie for weaving R (and other languages) with various markup languages
- R Markdown: A file format that combines R code chunks and markdown text which is converted by knitr into markdown, and other formats (e.g., HTML, pdf, etc.).
- R Studio: Open source, cross-platform IDE for R.

Benefits of knitr

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- knitr supports many markups: LaTeX, Markdown, HTML, reStructuredText
- knitr has really nice defaults
- Tidy placement of generated files
- Simplified figure production
 - automatically print ggplot2 and lattice figures
 - print figures by default
 - permit interspersing of figures and console output
- Greater extensibility:
 - output options
 - supports languages other than R
- Simplified caching
- And more: <http://yihui.name/slides/2012-knitr-RStudio.html>

- Benefits of Rstudio as IDE for R
 - Open source
 - Works on Linux, Mac, and Windows
 - Many useful features
 - It just works
 - Tight integration with knitr
- But many other options
 - Emacs with ESS
 - Vim with R plugin
 - Eclipse with StatET
 - etc.

RMarkdown Examples

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

- *Introduction to R Markdown*
- *Statistics homework example*
- *Analysis of Winter Olympic Medals Example*

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

Rstudio screenshot

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

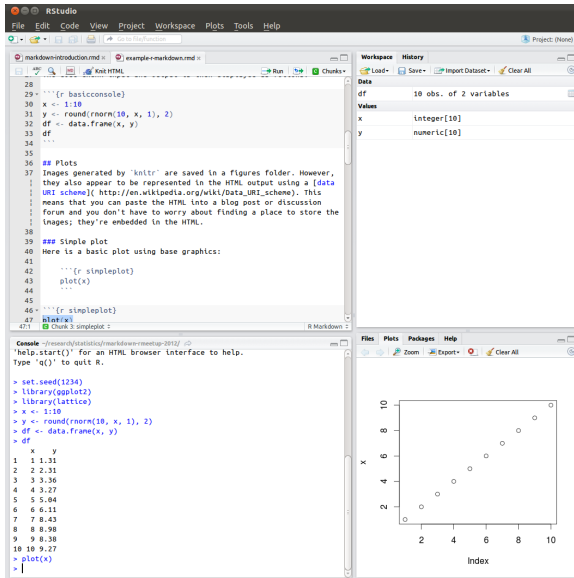
Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion



R Code chunks

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

see <http://yihui.name/knitr/options>

```
```{r my_chunk_name, some_option='some_value'}  
some_r_code
```
```

R Code chunks options

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

Global options:

```
`r opts_chunk$set(opt = value)` # general form  
`r opts_chunk$set(cache=TRUE)` # e.g, global cache
```

Some useful local options

- Hide console input: `echo=FALSE`
- Hide assorted messages: `warning=FALSE`,
`error=FALSE`, `message=FALSE`
- Hide console output: `results="hide"`
- Display console input as is: `tidy=FALSE`
- Output raw markup: `results="asis"`

Inline R Code

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

R Markdown

Markdown

HTML

``r 2 + 2``

``4``

`<code>4</code>`

``r I(2+2)``

4

4

Figures

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- Support for multiple figures in a code block
 - also see e.g., `par(mfrow=c(2,2))` or `grid.arrange`
- Figures and console output can be interspersed in a code chunk
- Various code chunk options
 - see <http://yihui.name/knitr/options>
 - `fig.width` and `fig.height`
 - dev defaults to pdf for LaTeX and png for HTML/markdown

Tables

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- Many options for creating HTML Tables:
 - R packages: xtable, googleVis, R2HTML, hwriter
 - markdown extensions: github, pandoc
 - Custom R code
- xtable is a reasonable option
- For informal reports just use console output
- css can be added later to control table appearance
- If you require sophisticated tables, you may want to switch to LaTeX

xtable example

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

```
print(xtable(my_data_frame, caption = "My Caption",
  digits = 3), type = "html",
  caption.placement = "top",
  html.table.attributes =
    "style=\"border: 1px solid black;\"")
```

My Caption

| | Mean | SD |
|----|-------|-------|
| A1 | 2.413 | 1.408 |
| A2 | 4.802 | 1.172 |
| A3 | 4.604 | 1.302 |
| A4 | 4.700 | 1.480 |
| A5 | 4.560 | 1.259 |

Caching

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

Basic workflow:

- If knitting is quick, don't cache.
- If knitting takes more than ten seconds add ``r opts_chunk$set(cache=TRUE)`` to the top of R Markdown file.
- If caching is causing problems, delete contents of cache folder,
- But if caching is causing problems and knitting takes a long time, name R code chunks and use the `dependson` option in knitr (see <http://yihui.name/knitr/options>). Naming also permits selective deletion of named R code chunks in the cache directory.

R package: markdown

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- Maintained by Jeffrey Horner; Developed by developed JJ Allaire, Jeffrey Horner, Vicent Marti, and Natacha Porte
- R Package that creates more options for converting Markdown to HTML
- `markdownToHTML("file.md", "file.html", options=c(...))`
- The default options are "hard_wrap", "use_xhtml", "smartypants", "base64_images"

Replicating R Studio's Knit to HTML

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

```
require(knitr) # for knitting from rmd to md
require(markdown) # for md to html
knit('test.rmd', 'test.md') # creates md
markdownToHTML('test.md', 'test.html') # create html
browseURL(paste('file://',
  file.path(getwd(), 'test.html'),
  sep='')) # open file in browser
```

see `?markdownHTMLOptions` for more options. E.g.,

```
markdownToHTML('test.md', 'test.html',
  options='fragment_only')
```

pandoc

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- pandoc is a library and command-line tool for converting between many document formats (e.g., HTML, markdown, pdf, LaTeX, docx; also supports multiple plain text slide formats such as beamer)
- Lots of options
- Often requires thought in order to minimise conversion issues

Example

```
pandoc -s file.html -o file.pdf
```

One-click build

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- For simple documents, click `knit to HTML` in RStudio
- For complex documents use a command-line option:
 - e.g., `makefile`, `Rscript`, etc.
 - combine with `pandoc`, `knitr` options, `markdown` options, text manipulation tools (e.g., `sed`, `awk`, scripting languages) etc. to flexibly produce a variety of documents

Example of LaTeX

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

If time permits, show example of knitr with LaTeX

Final thoughts

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

■ knitr and R Markdown

- It makes reproducible analysis as simple as one click
- Great tool for:
 - quick analyses for self and colleagues
 - doing homework
 - creating teaching resources
 - blog posts, websites, etc.
- Scope to make more complex documents, but at a certain point it may be worth exploring other tools

■ knitr and R LaTeX

- Great for journal articles, theses, books (e.g., citations, cross-references, printed works, equations)
- As your needs get more complex
 - pandoc, makefiles, knitr options, markdown package options, scripts, etc.

Links

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

- knitr: <http://yihui.name/knitr/>
- R Studio: <http://rstudio.org/>
- R Markdown with R Studio: http://rstudio.org/docs/authoring/using_markdown
- My Posts <http://jeromyanglim.blogspot.com.au/search/label/reproducible%20research>

Places to ask questions

- R on StackOverflow:
<http://stackoverflow.com/questions/tagged/r>
- LaTeX: <http://tex.stackexchange.com/>
- knitr: <https://github.com/yihui/knitr/issues>

Thank You

Simple
Reproducible
Analysis with
knitr, R
Markdown,
and RStudio

Jeromy
Anglim

Introduction

Markdown

knitr and R
Markdown

LaTeX

Conclusion

Questions?