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

Outline

Types of documents

- Types
  - Journal articles, books, book chapters, theses
  - Preliminary analyses
  - Online content: web pages, blog posts, forum posts
  - Slide show presentations
  - Consulting reports
- Key Distinctions
  - online versus paper-based
  - document or presentation
  - audience: formal versus informal (e.g., self, collaborators)

What is *reproducible analysis*?

- Reproducibility varies on a continuum
- I operationalise it as:
  - 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 has 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`

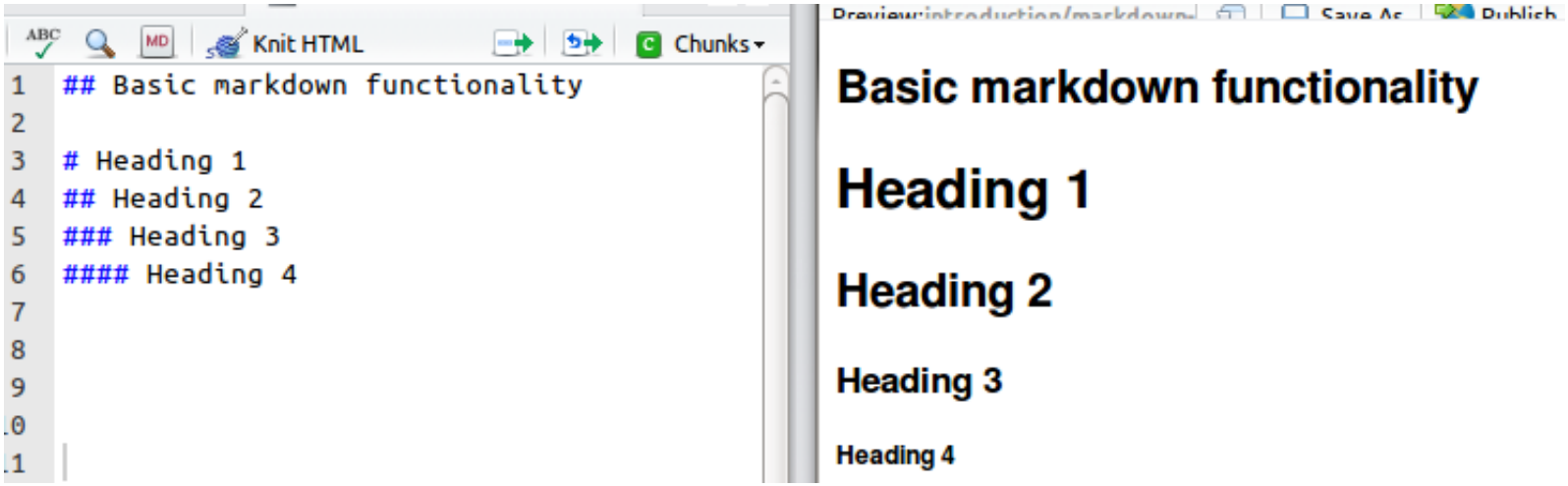
Aims of Reproducible

- 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

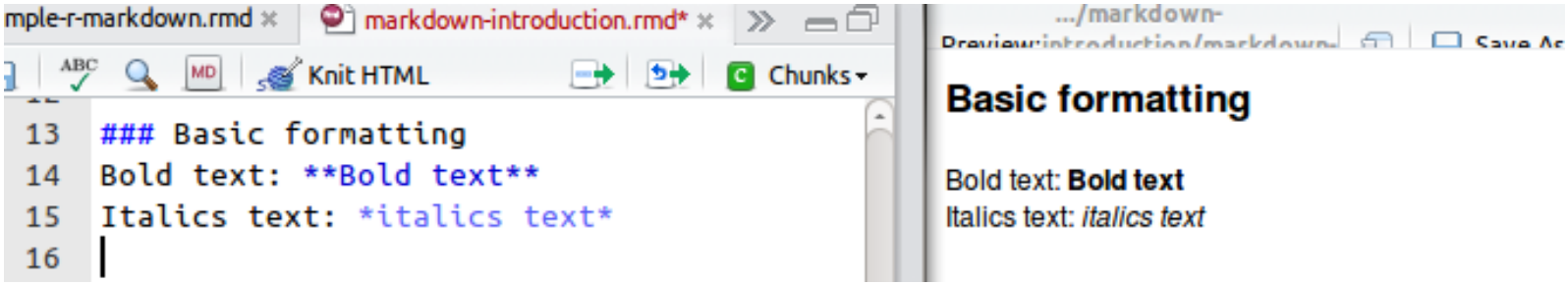
Overview of Markdown

- Ultra simplified and intuitive set of markup
- Limited set of markup
- HTML passed straight through
- Various extensions
- Popular on websites: e.g., StackOverflow, GitHub, Reddit

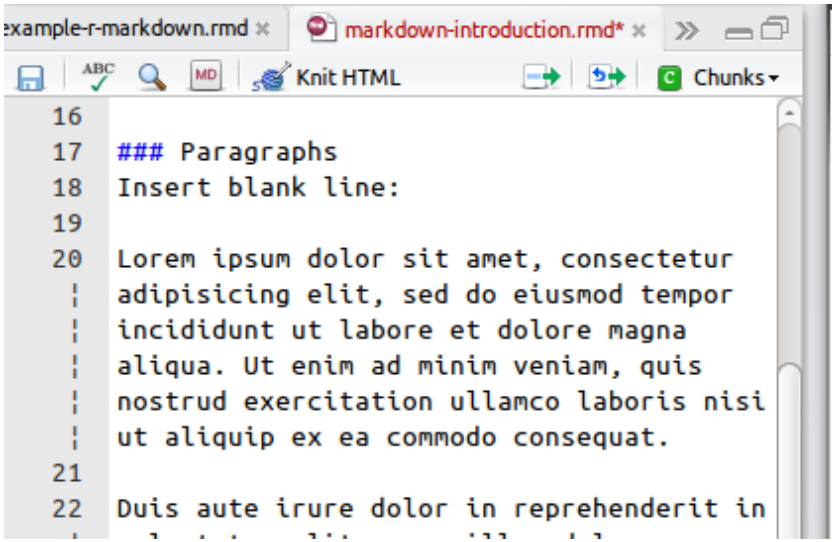
Headings

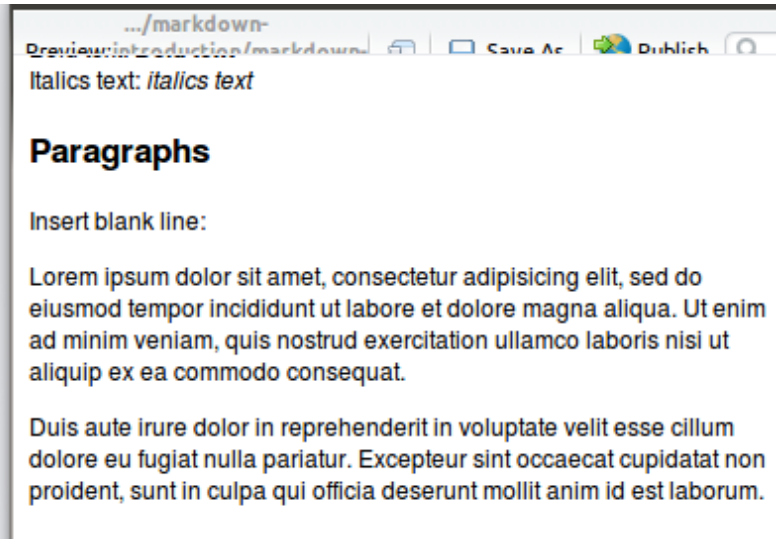

Knit HTML
Chunks
1 ## Basic markdown functionality
2
3 # Heading 1
4 ## Heading 2
5 ### Heading 3
6 #### Heading 4
7
8
9
10
11
Basic markdown functionality
Heading 1
Heading 2
Heading 3
Heading 4

Basic formatting



Paragraphs


example-r-markdown.rmd
markdown-introduction.rmd*
Knit HTML
Chunks
16
17 ### Paragraphs
18 Insert blank line:
19
20 Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.
21
22 Duis aute irure dolor in reprehenderit in


Italics text: italics text
Paragraphs
Insert blank line:
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.
Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Dot points


```
### 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
```

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


```
### Equations
Uses Mathjax to support LaTeX equations.

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

Displayed equations:

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

Equations

Uses Mathjax to support LaTeX equations.

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

Displayed equations:

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

Hyperlinks


```
### Hyperlinks

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

Hyperlinks

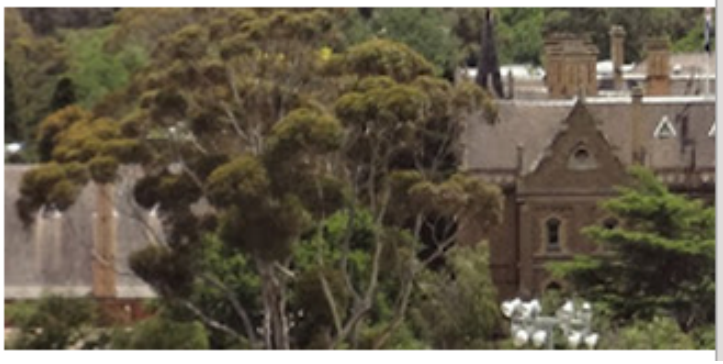
- my RSS feed.
- http://www.r-project.org/

Images


```
### Images

![image description here](figure/buildings.jpg)
```

Images



Code


````
### 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


```
### 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


```
### Tables
Extended github table functionality:

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

Or just write HTML:

<table border="1">
    <tr><td>Cell A1</td>
        <td>Cell B1</td></tr>
    <tr><td>Cell A2</td>
        <td>Cell B2</td></tr>
</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

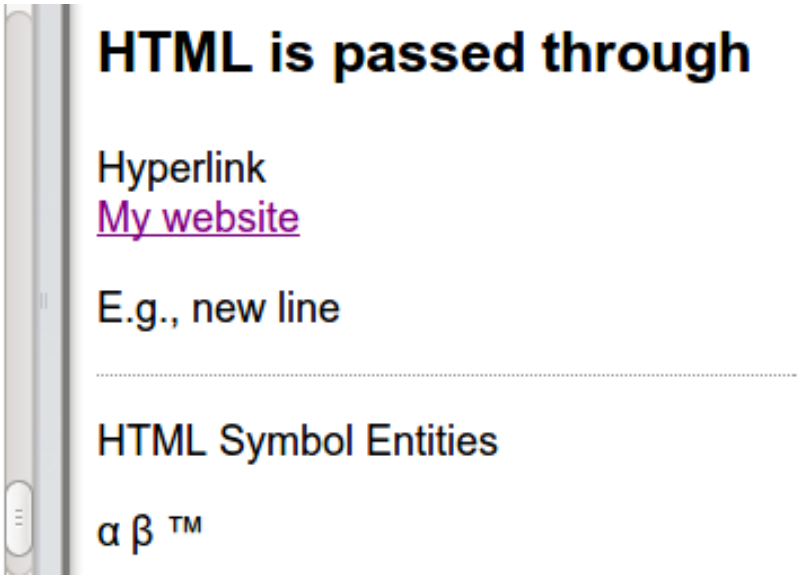

```
### HTML is passed through
Hyperlink
<a href="http://jeromyanglim.blogspot
.com">My website</a>



E.g., new line
<hr />

HTML Symbol Entities

&alpha; &beta; &trade;
```



HTML is passed through

Hyperlink
My website

E.g., new line

HTML Symbol Entities

α β ™

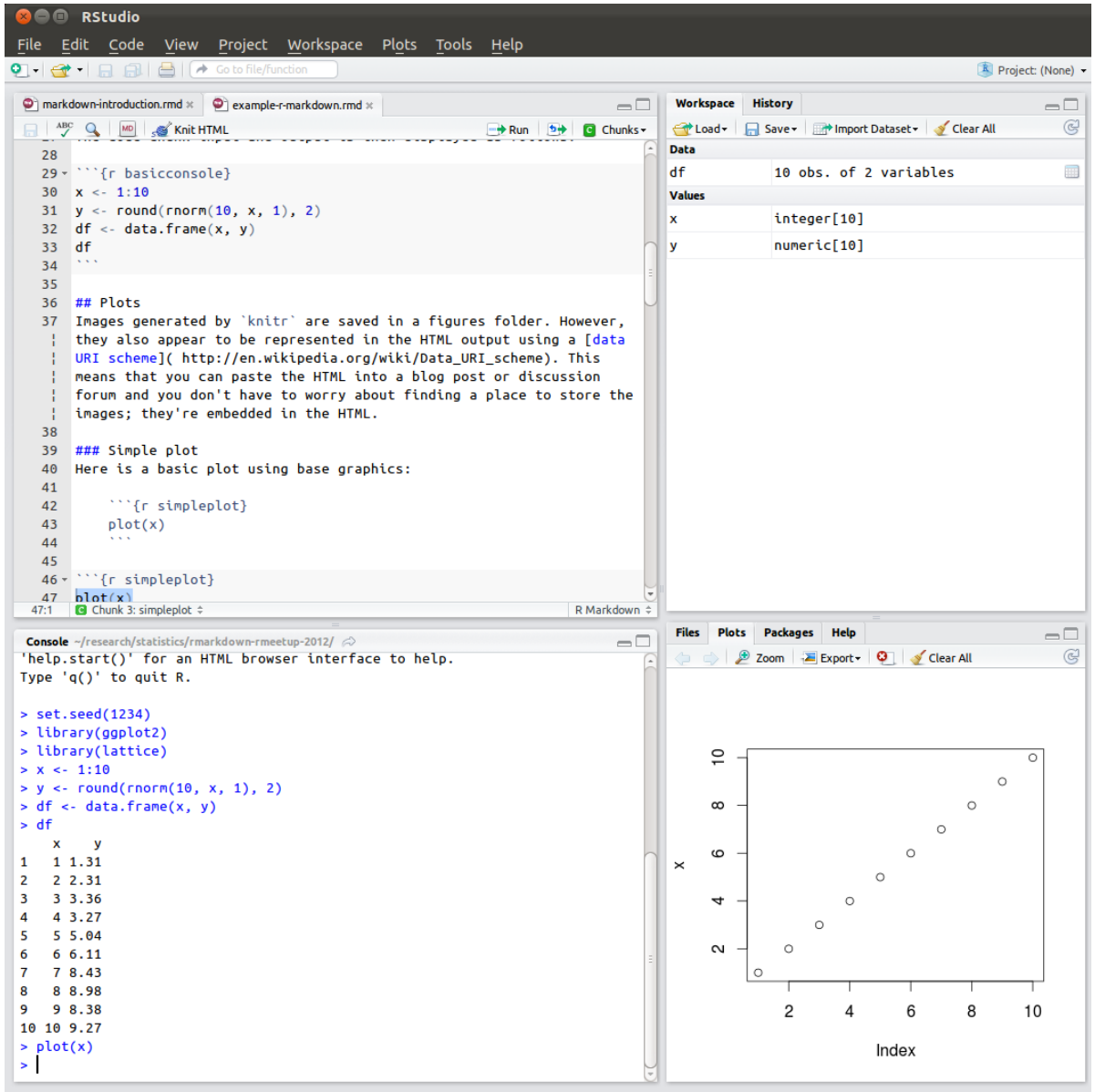
Benefits of knitr

- knitr supports many markups: LaTeX, Markdown, HTML, reStructuredText
- knitr has really nice defaults
- 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`

Rstudio

- 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.

Rstudio screenshot



R Code chunks

Global options:

Installation

- Install Rstudio
- Install knitr `install.packages("knitr")`

Inline R Code

- \texttt{`r 2 + 2` becomes `4` which becomes 4.
- `r I(2+2)`
- Markdown `4` 4 HTML `4` 4

Figures

Tables

- Many options for creating HTML Tables:
  - R packages: `xtable`, `googleVis`, `R2HTML`, `hwriter`
  - markdown extentions: 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

xtable example


```
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

Rstudio

Caching

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`

- Devloped by Jeffrey Horner
- R Package that creates more options for converting Markdown to HTML

Replicating R Studio


```
require(knitr) # required for knitting from rmd to md
require(markdown) # required 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')
```

knitr with LaTeX

- Sexpr
- Code chunk delimiters

Final thoughts

Links

- knitr: `http://yihui.name/knitr/`
- R Studio: `http://rstudio.org/`
- R Markdown with R Studio: `http://rstudio.org/docs/authoring/using_markdown`
- 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`