Reproducible research with R

Bjarki Þór Elvarsson and Einar Hjörleifsson

Marine Research Institute

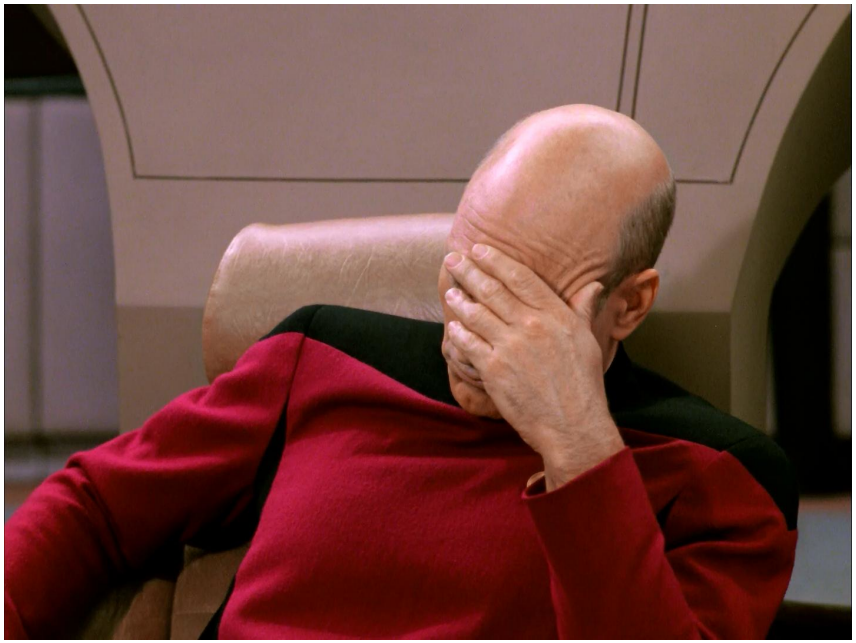
Overview

Reproducible research

- How did you do that?
- What data did you use?
- What scripts were used to do what?

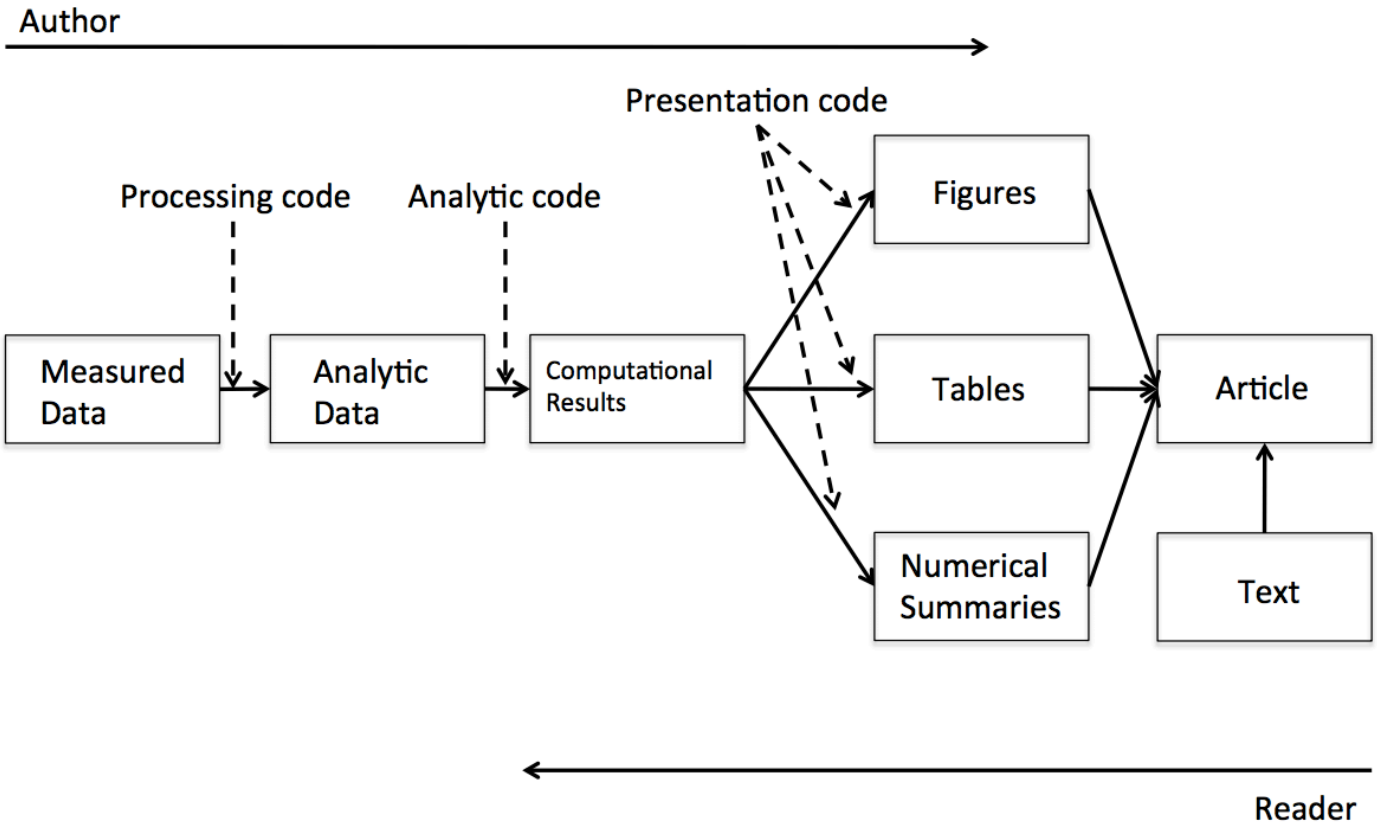
Typical workflow

- Prepare data
- Perform analysis
- Generate report
- Realise that something was wrong in the data or analysis
- Do it all over again



What is reproducible research?

"The final product of research is not only the paper itself, but also the full computation environment used to produce the results in the paper such as the code and data necessary for reproduction of the results and building upon the research." (Xie, 2014).

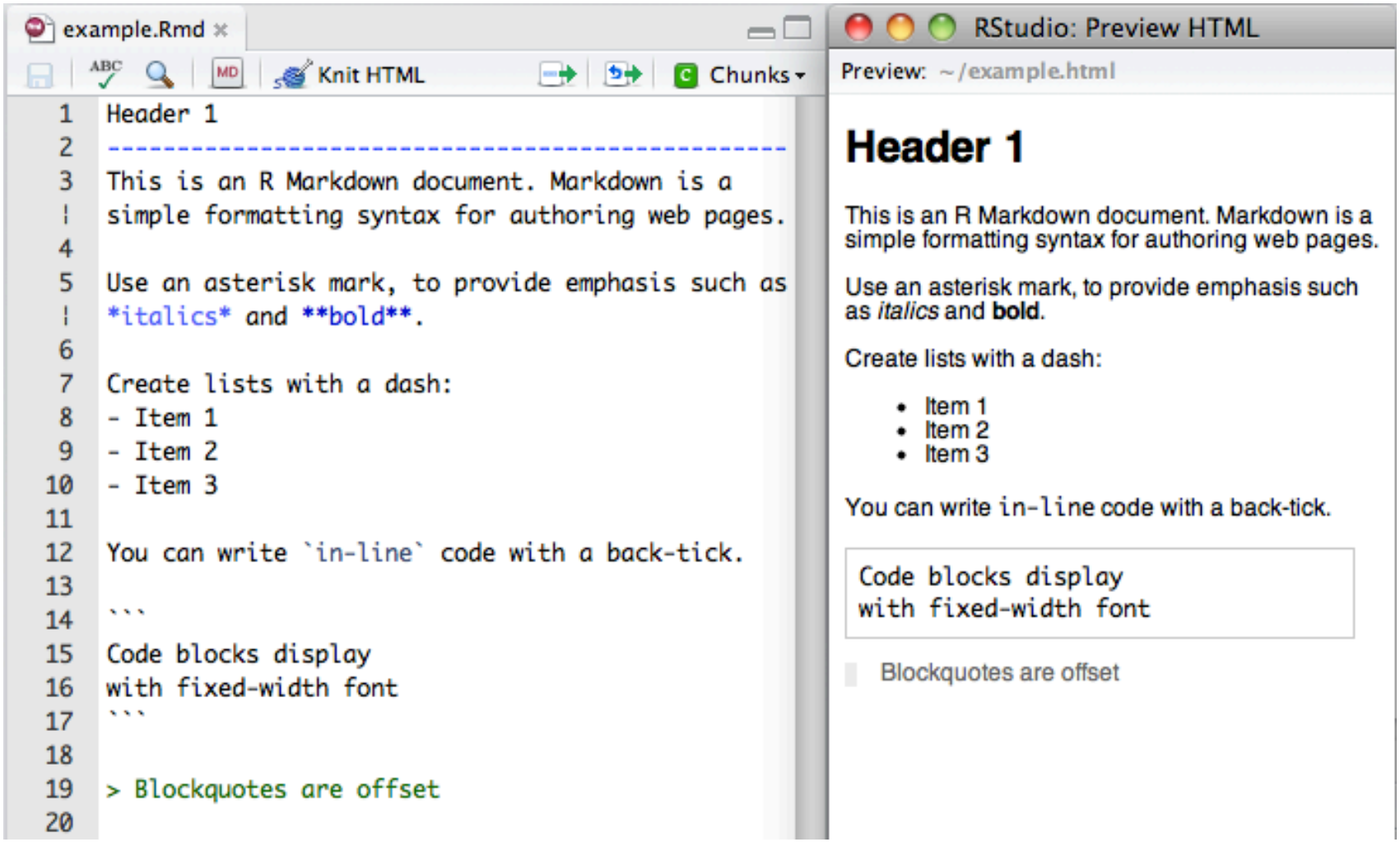


Rmarkdown

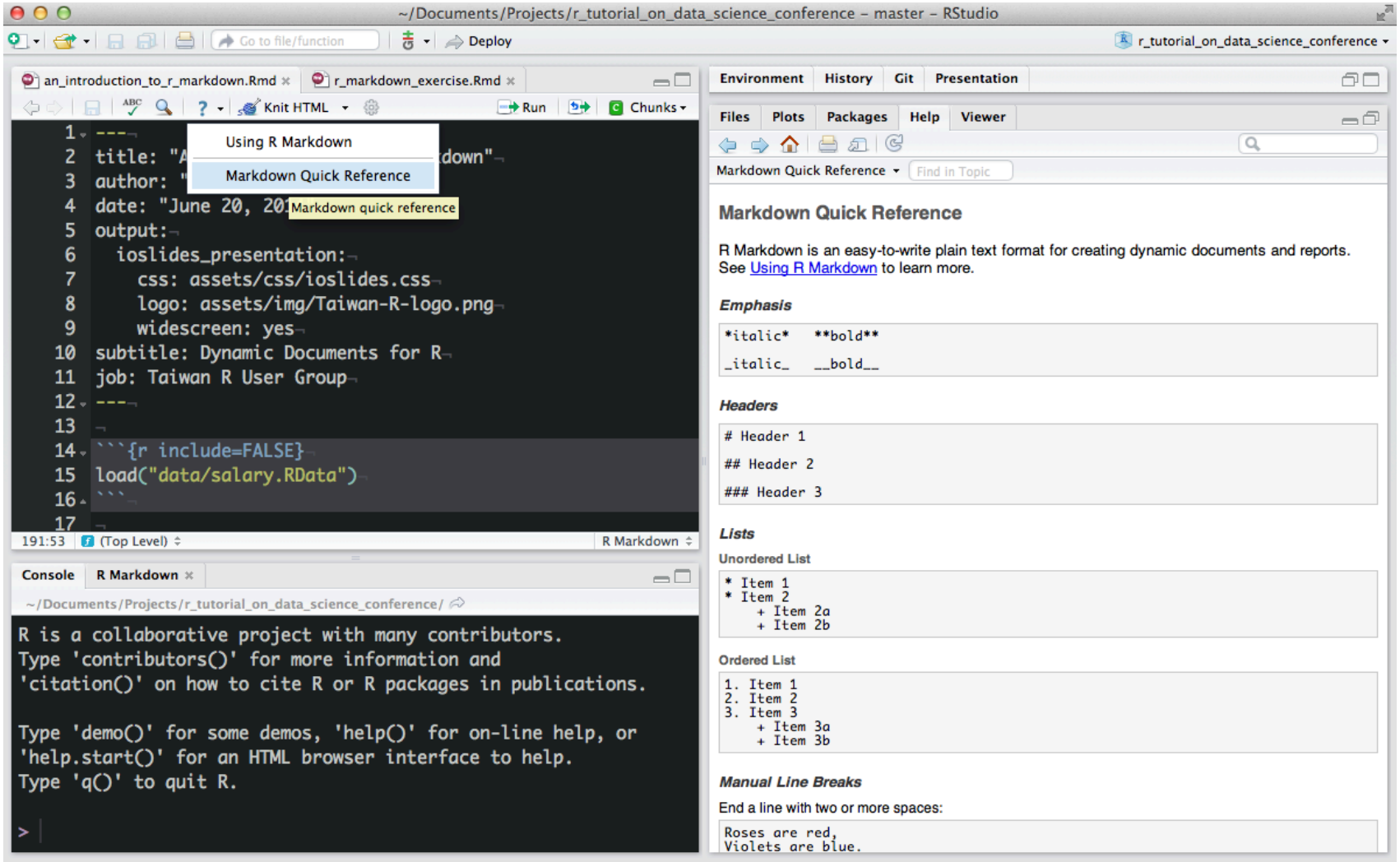
A convenient tool to generate reproducible document.

- Markdown
  - Remove HTML tag for higher readibility.
  - Inline HTML is avaliable.
- R markdown
  - Markdown + embedded R code chunks
  - Rmd -> md -> docx, pdf,html
- Why R Markdown
  - Consolidate your code and document into single file.
  - Easy for version control.

Markdown


example.Rmd
Knit HTML
Chunks
1 Header 1
2 ----------
3 This is an R Markdown document. Markdown is a
simple formatting syntax for authoring web pages.
4
5 Use an asterisk mark, to provide emphasis such as
italics and **bold**.
6
7 Create lists with a dash:
8 - Item 1
9 - Item 2
10 - Item 3
11
12 You can write `in-line` code with a back-tick.
13
14 ```
15 Code blocks display
16 with fixed-width font
17 ```
18
19 > Blockquotes are offset
20
RStudio: Preview HTML
Preview: ~/example.html
Header 1
This is an R Markdown document. Markdown is a simple formatting syntax for authoring web pages.
Use an asterisk mark, to provide emphasis such as italics and bold.
Create lists with a dash:
Item 1
Item 2
Item 3
You can write in-line code with a back-tick.
Code blocks display
with fixed-width font
Blockquotes are offset

Markdown quick refefence



Markdown commands

```
Text formatting
#     1st level header   *a*     italics
##    2nd level header   **b**   bold
###   3rd level header

List:                    Hyperlinks
* item 1                 [Hafro](www.hafro.is)
* item 2
  + subitem
```

Markdown equations

Markdown also supports LaTeX equation so

```
$x^2 + \sum_i a_i y_i^n$
```

becomes

$$x^2 + \sum_i a_i y_i^n$$

Further reading: https://en.wikibooks.org/wiki/LaTeX/Mathematics

Embedding R

chunks.Rmd

Knit HTML | Chunks

```
R Code Chunks
==========

With R Markdown, you can insert R code
chunks including plots:

```{r qplot, fig.width=4, fig.height=3,
message=FALSE}
# quick summary and plot
library(ggplot2)
summary(cars)
qplot(speed, dist, data=cars) +
      geom_smooth()
```
```

RStudio: Preview HTML

Preview: ~/chunks.html | Save As | Publish

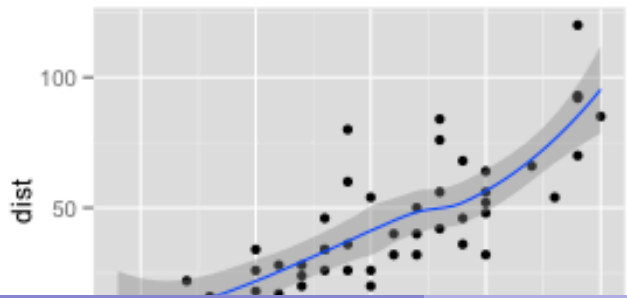
R Code Chunks

With R Markdown, you can insert R code chunks including plots:

```
# quick summary and plot
library(ggplot2)
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2
##  1st Qu.:12.0   1st Qu.: 26
##  Median :15.0   Median : 36
##  Mean   :15.4   Mean   : 43
##  3rd Qu.:19.0   3rd Qu.: 56
##  Max.   :25.0   Max.   :120
```

```
qplot(speed, dist, data = cars) + geom_smooth()
```



Chunks

- Input from R in Markdown is evaluated in `chunks`:

```
```{r}
<insert R code for Markdown>
```
```

- Chunks have a plethora of options available by default:
  - Allows you to display or hide code
  - Similarly display or hide its output
  - Figure dimensions can be set (if created by in the chunk)

R code chunks

````
Here's some code
```{r}
dim(iris)
```
````

Here's some code

```
dim(iris)
```

```
## [1] 150   5
```

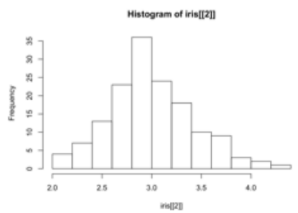
````
Here's some code
```{r echo=FALSE}
dim(iris)
```
````

Here's some code

```
## [1] 150   5
```

````
Here's a plot
```{r echo=FALSE}
hist(iris[[2]])
```
````


Here's a plot




````
Here's some code
```{r eval=FALSE}
dim(iris)
```
````

Here's some code

```
dim(iris)
```

Helpful chunk output options

- eval = TRUE : Evaluate all or part of the current chunk
- echo = TRUE : Show all or part of the source code
- results = 'asis' : Writes raw output from R to the output document without markup. Helpful for creating tables with xtable. markup is the default.
- include = TRUE : Code chunk will be included in output. If you don't want a chunk in the output but still evaluated set this to FALSE

Figure settings

- fig.width controls the figure width (in inches)
- fig.height controls the figure height (in inches)
- fig.cap is the figure caption
- fig.align sets the alignment
- dev allows the user to specify the file type of the plot (png, pdf, bmp, etc..)

Creating tables in RMarkdown

Tables in Markdown are fairly easy:

```
First Header  | Second Header
------------- | -------------
Content Cell  | Content Cell
Content Cell  | Content Ce
```

But one can also use R command to create tables. Note: Set results='asis' to write raw results from R into the output document

```
library(knitr)
kable(tab,caption = 'This is a table')
```

Output options

- Html, both slides and normal webpages
- Word, requires either Word or Libreoffice to display
- Pdf, requires LaTeX:
  - Linux do 'yum install texlive'
  - Windows: install miktex from miktex.org
  - MacOsX: install MacTeX from tug.org/mactex
- Custom output

Other notable features

- Bibtex style citations
- Theming available based on reference word documents, css files and latex templates
- Allows the creation of interactive webpages using shiny
- Support for presentations (beamer, ioslides, slidy)

Class exercise

- Play around with Einar's ggplot2 document