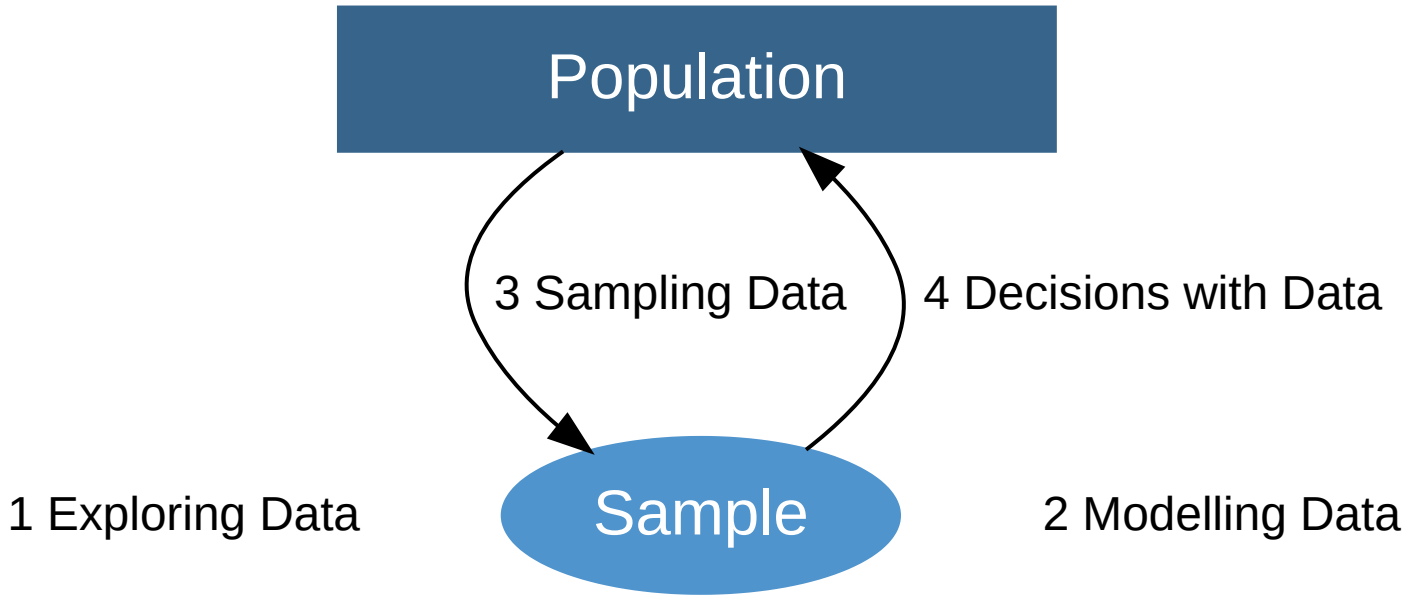
Reproducible Reports

Modelling Data | Normal Model

Unit Overview



Module2 Modelling Data

Normal Model

What is the Normal Curve? How can we use it to model data?

Linear Model

How can we describe the relationship between 2 variables? When is a linear model appropriate?

Reproducible Reports

Reproducible Research

R Markdown

Steps for .Rmd

Handy Tricks

Dangers of Non-reproducible Research

Summary

Reproducible Research

Defining Reproducibility

Getting Started

- Why is ‘reproducibility’ vital in scientific research?
- What does ‘reproducibility’ mean in data science?

Reproducibility Crisis (Surveys from Nature, 2016)

Is there a reproducibility crisis in science?

Reproducibility Crisis (TED-Ed, 2016)

Is there a reproducibility crisis in science? - Matt Anticole

Reproducible Research

- Increasingly, journals are requiring reproducible research, which requires “data sets and software to be made available for verifying published findings and conducting alternative analyses”.
  - A study by Begley and Ellis(2012) found that 47 out of 53 medical research papers focused on cancer research that was irreproducible.
  - A follow up study by Begley (2013) identified “6 flags for suspect work”: studies were not performed by investigators blinded to the experimental versus the control arms, there was a failure to repeat experiments, a lack of positive and negative controls, failure to show all the data, inappropriate use of statistical tests and use of reagents that were not appropriately validated.

New tools for building interaction with data

"Scientific figures are typically rendered as static images. But these are divorced from the underlying data, which prevents readers from exploring them in more detail by, for instance, zooming in on features of interest. For genomicists needing to cram millions of data points into dense visuals a few centimetres big, this can be particularly problematic. The same is true for researchers working with computational algorithms. Scientists often post software on open-source repositories such as GitHub, but getting the code to run properly is easier said than done … Reviewers and other interested parties often require extra software and configuration to make the algorithms work.

Some journals now bridge that gap by supporting interactive figures and code."

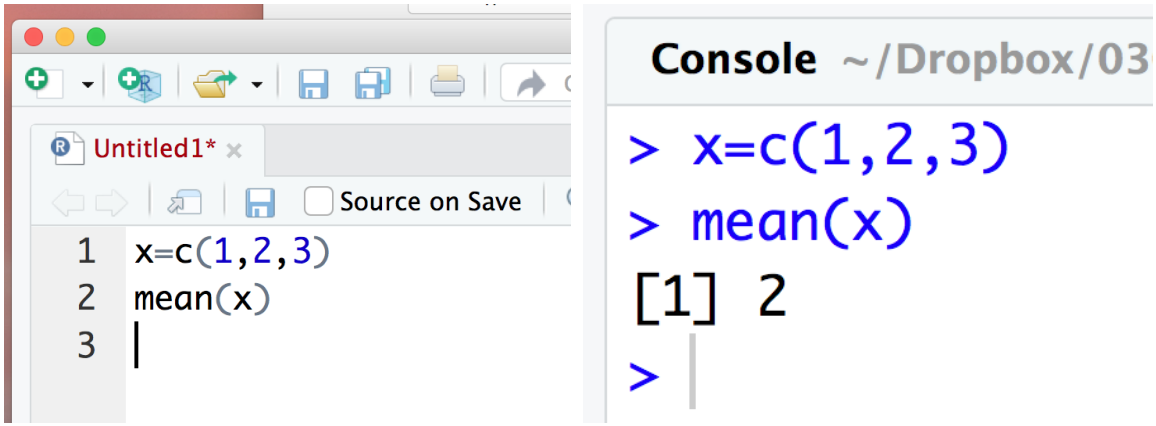
🔗 Nature, Jan 2018

R Markdown

R Script (.R)

- An R script file is a text file which saves your R code and comments `#`.

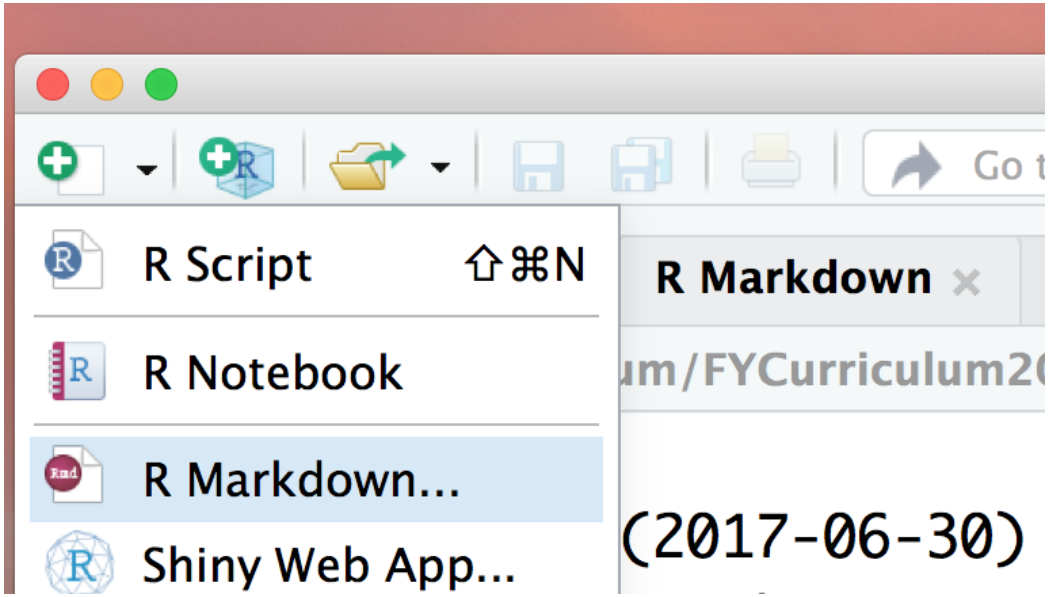
- A collaborator runs the .R script file, and then the code and results appear in the console.



R Markdown (.Rmd)

- R Markdown is an authoring framework for data science which produces dynamic, interactive documents with R.

- It saves and executes code AND produces a high quality report for the collaborator to view and validate the results.



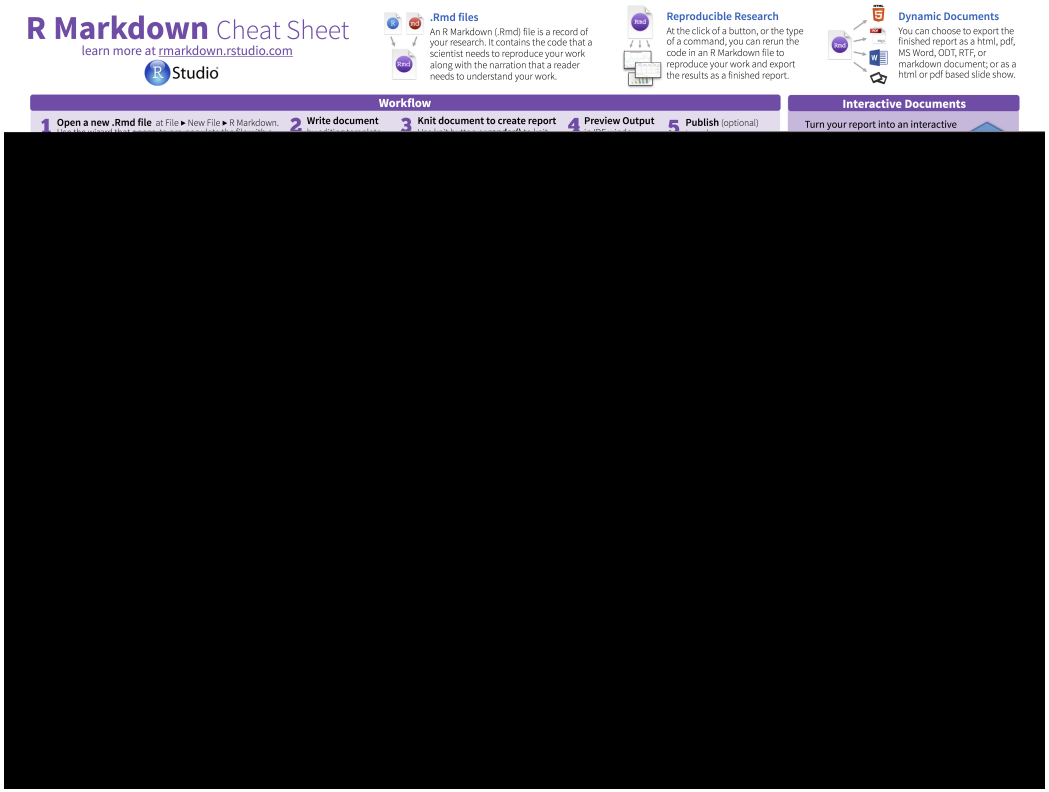

01:12

Inputs and Output Formats

- An R Markdown file combines
  - chunks of text (written in **markdown**)
  - embedded code (eg R, Python, SQL)
  - Latex.
- R Markdown supports dozens of static and dynamic **output** formats including HTML, PDF, MS Word, Beamer, HTML5 slides, Tufte-style handouts, books, dashboards, Shiny apps, scientific articles and websites.

Steps for .Rmd

Cheat Sheet



🔗 Cheatsheet

Step1 (Summary)

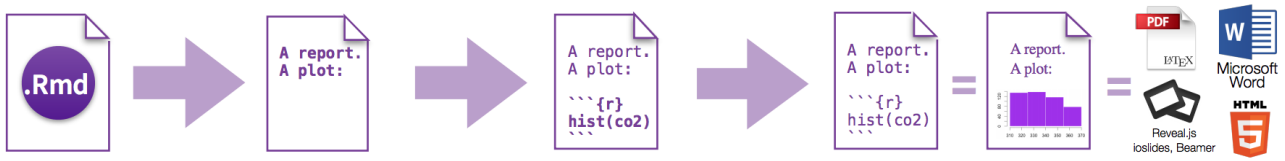
1. Workflow R Markdown is a format for writing reproducible, dynamic reports with R. Use it to embed R code and results into slideshows, pdfs, html documents, Word files and more. To make a report:

i. Open - Open a file that uses the .Rmd extension.

ii. Write - Write content with the easy to use R Markdown syntax

iii. Embed - Embed R code that creates output to include in the report

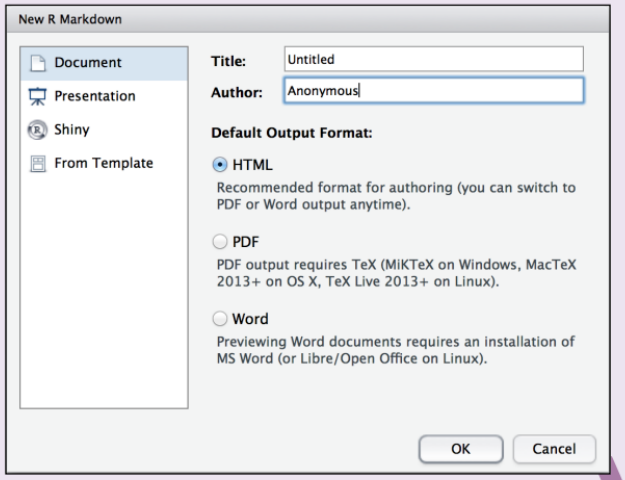
iv. Render - Replace R code with its output and transform the report into a slideshow, pdf, html or ms Word file.



Step2

2. Open File Start by saving a text file with the extension .Rmd, or open an RStudio Rmd template

- In the menu bar, click **File ▸ New File ▸ R Markdown…**
- A window will open. Select the class of output you would like to make with your .Rmd file
- Select the specific type of output to make with the radio buttons (you can change this later)
- Click OK



Step3

3. Markdown Next, write your report in plain text. Use markdown syntax to describe how to format text in the final report.

syntax

```
Plain text
End a line with two spaces to start a new paragraph.
*italics* and _italics_
**bold** and __bold__
superscript^2^
~~strikethrough~~
[link](www.rstudio.com)

# Header 1

## Header 2

### Header 3

#### Header 4

##### Header 5

###### Header 6

endash: --
emdash: ---
ellipsis: ...
inline equation: $A = \pi*r^{2}$
image: ![](path/to/smallorb.png)

horizontal rule (or slide break):

***

> block quote

* unordered list
* item 2
    + sub-item 1
    + sub-item 2

1. ordered list
2. item 2
    + sub-item 1
    + sub-item 2

Table Header  | Second Header
------------- | -------------
Table Cell    | Cell 2
Cell 3        | Cell 4
```

becomes

Plain text
End a line with two spaces to start a new paragraph.
italics and *italics*
bold and **bold**
superscript^2
~~strikethrough~~
link

Header 1
Header 2
Header 3
Header 4
Header 5
Header 6

endash: –
emdash: —
ellipsis: …
inline equation: $A = \pi * r^2$
image: 
horizontal rule (or slide break):

> block quote

- unordered list
- item 2
  - sub-item 1
  - sub-item 2

1. ordered list
2. item 2
   - sub-item 1
   - sub-item 2

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4

Step4

4. Choose Output Write a YAML header that explains what type of document to build from your R Markdown file.

YAML

A YAML header is a set of key: value pairs at the start of your file. Begin and end the header with a line of three dashes (- - -)

```
---
title: "Untitled"
author: "Anonymous"
output: html_document
---

This is the start of my
report. The above is metadata
saved in a YAML header.
```

The RStudio template writes the YAML header for you

The output value determines which type of file R will build from your .Rmd file (in Step 6)

output: html_document	html file (web page)
output: pdf_document	pdf document
output: word_document	Microsoft Word .docx
output: beamer_presentation	beamer slideshow (pdf)
output: ioslides_presentation	ioslides slideshow (html)

Step5

5. Embed Code Use knitr syntax to embed R code into your report. R will run the code and include the results when you render your report.

inline code

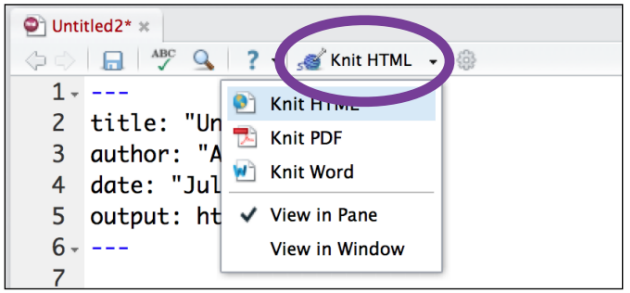
Surround code with back ticks and r.
R replaces inline code with its results.

```
Two plus two equals `r 2 + 2`.
```

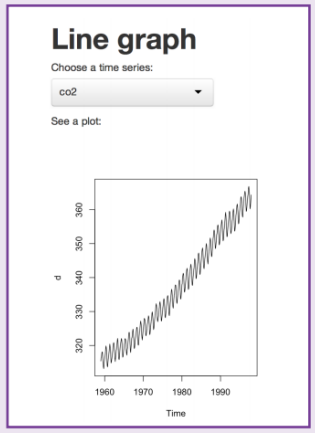
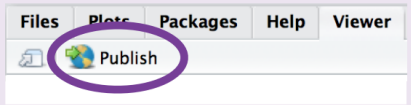
▶ Two plus two equals 4.

code chunks

Start a chunk with ```{r}.
End a chunk with ```

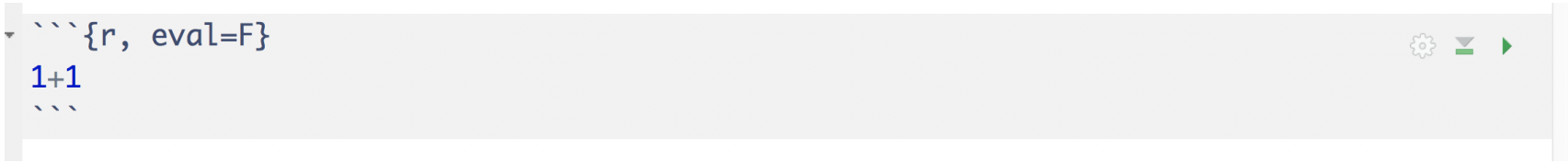
```
Here's some code
```{r}
dim(iris)
```
```

▶

Here's some code

```
dim(iris)
```

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

display options

Use knitr options to style the output of a chunk.
Place options in brackets above the chunk.

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

▶ Here's some code

```
dim(iris)
```

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

▶ Here's some code

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

option	default	effect
eval	TRUE	Whether to evaluate the code and include its results
echo	TRUE	Whether to display code along with its results
warning	TRUE	Whether to display warnings
error	FALSE	Whether to display errors
message	TRUE	Whether to display messages
tidy	FALSE	Whether to reformat code in a tidy way when displaying it
results	"markup"	"markup", "asis", "hold", or "hide"
cache	FALSE	Whether to cache results for future renders
comment	"##"	Comment character to preface results with
fig.width	7	Width in inches for plots created in chunk
fig.height	7	Height in inches for plots created in chunk

For more details visit yihui.name/knitr/

Step6

6. Render Use your .Rmd file as a blueprint to build a finished report.

Render your report in one of two ways

1. Run **rmarkdown::render("<file path>")**
2. Click the **knit HTML** button at the top of the RStudio scripts pane



When you render, R will

- execute each embedded code chunk and insert the results into your report
- build a new version of your report in the output file type
- open a preview of the output file in the viewer pane
- save the output file in your working directory

Step7 (Ext)

7. Interactive Docs Turn your report into an interactive Shiny document in 3 steps

1 Add **runtime: shiny** to the YAML header

```
---
title: "Line graph"
output: html_document
runtime: shiny
---
```

2 In the code chunks, add Shiny **input** functions to embed widgets. Add Shiny **render** functions to embed reactive output

```
---
title: "Line graph"
output: html_document
runtime: shiny
---

Choose a time series:
```{r echo = FALSE}
selectInput("data", "",
  c("co2", "lh"))
```

See a plot:
```{r echo = FALSE}
renderPlot({
  d <- get(input$data)
  plot(d)
})
```
```

3 Render with **rmarkdown::run** or click **Run Document** in RStudio



***** *Note: your report will be a Shiny app, which means you must choose an html output format, like* ***html_document*** *(for an interactive report) or* ***ioslides_presentation*** *(for an interactive slideshow).*

Step8 (Ext)

8. Publish Share your report where users can visit it online

Rpubs.com

Share non-interactive documents on RStudio's free R Markdown publishing site
www.rpubs.com

ShinyApps.io

Host an interactive document on RStudio's server. Free and paid options
www.shinyapps.io

Click the "Publish" button in the RStudio preview window to publish to rpubs.com with one click.



Step9

9. Learn More

Documentation and examples - rmarkdown.rstudio.com
Further Articles - shiny.rstudio.com/articles
- blog.rstudio.com
- @rstudio

Handy Tricks

Annotate your code as you go (best practise)

This ensures:

- understanding for you;
- transferability to someone else, either for collaboration, or to someone who inherits your project.

Customise code chunk output

- Echo = F: Don’t show code

````
```{r, echo=F}
1+1
```
````

```
## [1] 2
```

- Eval = F: Don’t evaluate result

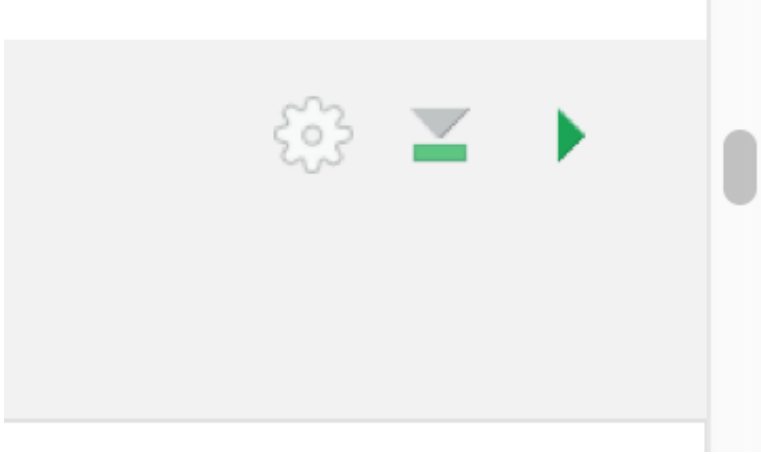



```
1+1
```

Run code chunks

There are 2 options:

- Run result inline.
- Run result in console.



For quick checks, you can run inline. But usually it’s better to run in console, in case later chunks rely on earlier ones.

Put R code in text

Code results can be inserted directly into the text of a .Rmd file.

- Take this line in .Rmd:

```
This is an example `r 1+1`.
```

- It would be rendered as:

This is an example 2.

🔗 https://rmarkdown.rstudio.com/lesson-4.html

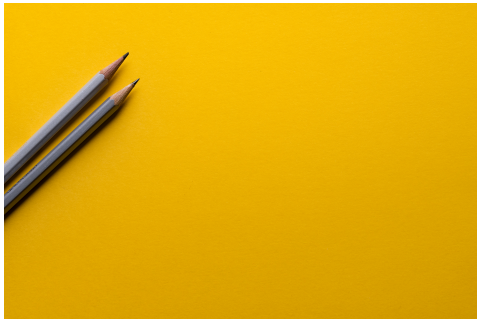
Add hyperlink

Simply `[Name of Link](url)`

Add images

- Make a folder called **figure**, and store **2pencils.jpg** in it.
- The speedy way is: `![Caption](figure/2pencils.jpg)`
- However, to customise the size, we use the following:

```
<div align="center">
<img src="http://www.maths.usyd.edu.au/u/UG/JM/DATA1001/r/2018S1/figure/2pencils.jpg"
width="40%" height="40%"
</div>
```



Or direct from the url.

```
<div align="center">
<img src="http://www.maths.usyd.edu.au/u/UG/JM/DATA1001/r/2018S1/figure/2pencils.jpg"
width="40%" height="40%"
</d
```

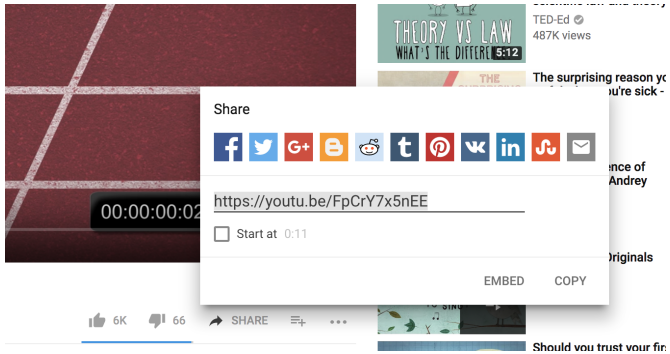
Open Link (Command+Click)

<img src="http://www.maths.usyd.edu.au/u/UG/JM/DATA1001/r/2018S1/figure/2pencils.jpg"
width="40%" height="40%"

Add videos

```
<div align="center">
<iframe width="560" height="315" src="https://www.youtube.com/embed/FpCrY7x5nEE"
frameborder="0" allow="autoplay; encrypted-media" allowfullscreen></iframe>
</div>
```

You can get this embeding code direct from YouTube.



Adding a table

```
|Heading1|Heading2|

|------|------|

|Entry1|Entry2|
```

Heading1	Heading2
Entry1	Entry2

Dangers of Non-reproducible Research

What can go wrong?

Without reproducible research:

- Data versions can change (eg people edit an Excel file without documenting what has changed and why);
- Graphical summaries can change (eg people can photoshop images without keeping record of what changed and why)

Reproducible research is about being responsible with possible human errors, or worse, detecting intentionally changed results.

 Why might a researcher intentionally change results?

Summary

R Markdown is a great way to produce reproducible reports, enabling more transparent reproducible research.

Key words

reproducibility, reproducibiity crisis, collaborator, r script, r markdown, code chunks, inline, console, hyperlink, non-reproducible