

Reproducible Reporting

Session 2
Patrick Mathias
July 16, 2020

July 16 2020	Session	Instructor
1:00 pm - 1:30 pm	Instructor Introductions, Introduction to technology	Amrom Obstfeld
1:30 pm - 2:15 pm	Introduction to R and RStudio	Joe Rudolf
2:30 pm - 3:15 pm	Reproducible Reporting	Patrick Mathias
3:30 pm - 5:00 pm	Data Visualization	Stephan Kadauke
July 17 2020		
1:00 pm - 2:30 pm	Data Transformation	Amrom Obstfeld
2:45 pm - 4:15 pm	Statistical Analysis	Dan Herman
4:30 pm - 5:00 pm	Advanced Reporting	Patrick Mathias

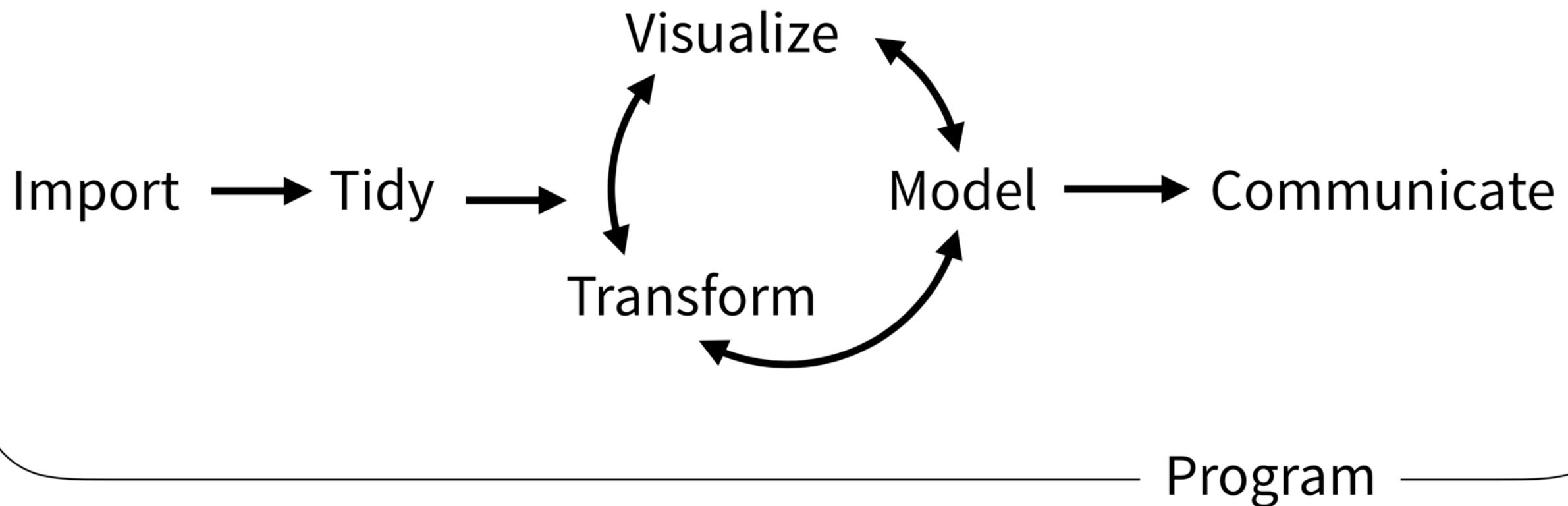
Goals

1. Understand why reproducible reporting is important
2. Learn to work within R Markdown for reproducible reports

Objectives

1. Create an R Markdown document and generate different types of output files
2. Practice modifying each component of a R Markdown file

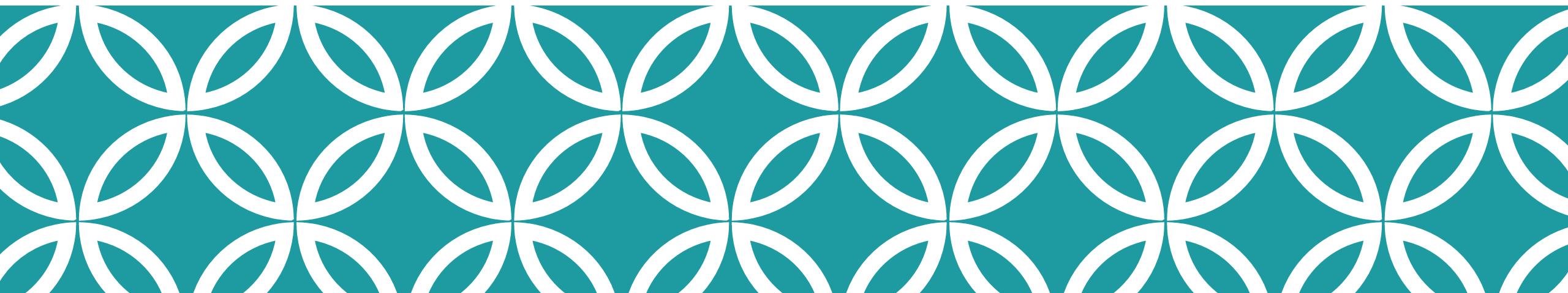
Typical Data Science Pipeline



Your Turn #1

1. List 3 benefits of doing data analysis using spreadsheet software like Microsoft Excel
2. List 3 drawbacks/problems with doing data analysis using spreadsheet software like Microsoft Excel





Why is reproducibility important?

Replication vs Reproduction

- ❖ Replication: other people collect new data
 - Scientific gold standard
 - Difficult and time-consuming

- ❖ Reproduction: other people analyze the same data
 - Does not by itself validate the analysis ...
 - Has been proposed as a minimal standard

Case

37 y/o M informatician with PMH of email overload disorder

Request from informatics staff:

“Please provide detailed data from your 2 year old analysis of total departmental effort spent performing test cancellations for a SBAR calling out the need to invest effort in duplicate checking rules for the ongoing EHR implementation project”

Analysis is in an Excel file but original raw data is nowhere to be found

Consider the above scenario, but with someone else performing the original analysis

Would it be less work to start from scratch and rewrite the analysis?

Point-and-Click Is Not Reproducible

- Interactive tools do not record user actions
- Manual documentation is error-prone
- Manual analyses cannot be repeated on new data sets or shared with collaborators



Computer code can precisely document each step of the analysis

Why YOU Should Do Data Analysis Reproducibly

“Can we redo the analysis with this month’s data?”

“Why do the data in Table 1 not seem to agree with Figure 2?”

“Why did I decide to omit these six samples?”



**YOUR CLOSEST COLLABORATOR IS YOU FROM 6 MONTHS AGO
(BUT YOU DON’T ANSWER E-MAILS)**

Using R for Reproducibility

Programming in R (or another language) allows one to reproduce analysis steps exactly or perform same analysis on new data

Better practice is to create documentation about analysis to accompany and explain code

Best practice is include documentation and code in one place



Using R Markdown to Support Reproducibility

The screenshot shows an RStudio interface with an R Markdown file open. The file contains the following code:

```
1 ---  
2 title: "Summarization and statistics in R"  
3 output: html_notebook  
4 editor_options:  
5   chunk_output_type: inline  
6 ---  
7  
8 ## Setup  
9  
10 ````{r setup}  
11 library(tidyverse)  
12 library(readxl)  
13  
14 orders <- read_excel("data/orders_data_set.xlsx")  
15 ````  
16  
17 ## Summarize  
18  
19 ````{r}  
20 orders %>%  
21   select(order_id, patient_id) %>%  
22   head(4) %>%  
23   summarize(order_count = n(),  
24             pt_count    = n_distinct(patient_id))  
25 ````  
26  
27 ## Your Turn 1  
28  
29 Add onto the code in the above chunk to calculate:  
30  
31 1) Mean count of orders per patient
```

Annotations with blue rounded rectangles highlight specific sections of the code:

- A rectangle surrounds the header section (lines 1-6).
- A rectangle surrounds the first code chunk (lines 18-25).
- A rectangle surrounds the text section (lines 27-31).

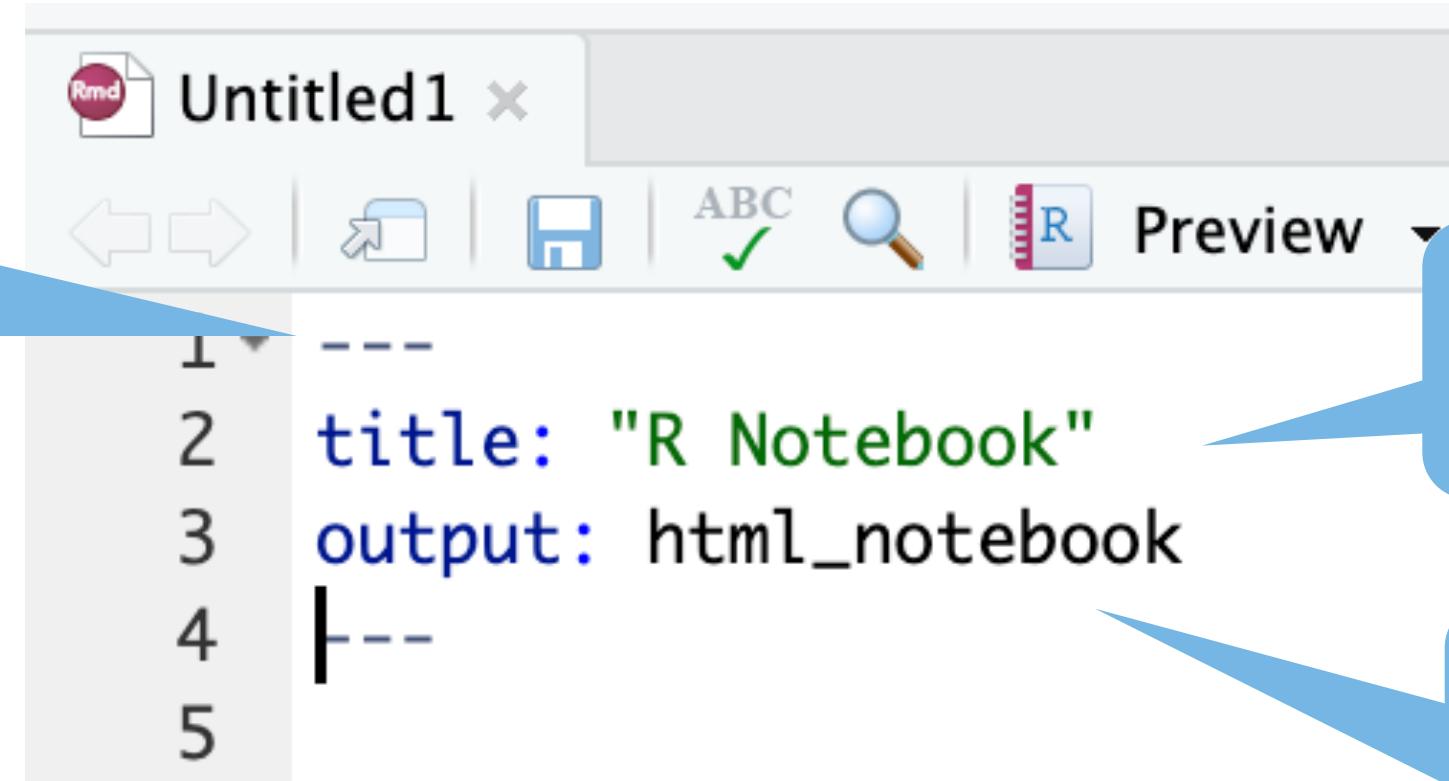
Header

Code chunk

Text

Header

Starts and ends
with 3 dashes



A screenshot of an R Notebook interface titled "Untitled1". The notebook contains five code cells:

```
1 ---  
2 title: "R Notebook"  
3 output: html_notebook  
4 ---  
5
```

Title in quotes

Output format

Text

5

6 This is an [R Markdown](<http://rmarkdown.rstudio.com>) Notebook. When you execute
code within the notebook, the results appear beneath the code.

7

8 Try executing this chunk by clicking the *Run* button within the chunk or by
placing your cursor inside it and pressing *Cmd+Shift+Enter*.

9

1 asterisk for *italics* (*italics*)

2 asterisks for **bold** (**bold**)

Hyphens (-bullet 1) for bullet points

Include hyperlinks
with [text](link)
format

1 asterisk for
italics (*italics*)

R Markdown

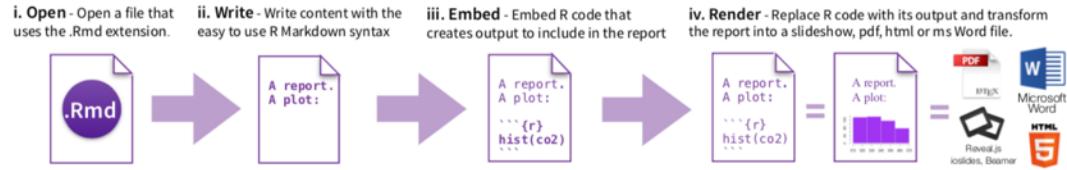
Cheat Sheet

learn more at rmarkdown.rstudio.com

rmarkdown 0.2.50 Updated: 8/14



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:



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

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:   
  
horizontal rule (or slide break):
```

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: R

output: beamer_presentation..... beamer slideshow (pdf)
output: ioslides_presentation..... ioslides slideshow (html)

RStudio® is a trademark of RStudio, Inc. • CC BY RStudio • info@rstudio.com • 844-448-1212 • rstudio.com

syntax

```
Plain text  
End a line with two spaces to start a new paragraph.  
*italics* and _italics_  
**bold** and __bold__
```

> 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

- 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
```

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

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

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

Your Turn #2

1. Open up an R Notebook per guidance on next few slides.
2. Change the title of your R Notebook to “My First R Notebook” by modifying the header.
3. Add your name as the author by adding another line to the header:
author: “Your Name”
3. Add a second level heading (##) at the end of the notebook called “My Calculation”



Step 1

The screenshot shows the RStudio interface. On the left, a file creation dialog is open, listing various file types: R Script, R Notebook, R Markdown..., Shiny Web App..., Plumber API..., Text File, C++ File, Python Script, SQL Script, Stan File, D3 Script, R Sweave, R HTML, R Presentation, and R Documentation. The 'R Notebook' option is highlighted with a blue arrow labeled 'Step 1'. In the center, a code editor window displays the R Foundation license text. A blue arrow labeled 'Step 2' points from the license text towards the right panel. The right panel contains the Environment, Data, and Files panes. The 'Files' pane shows a directory structure with files like .Rhists, .Rprof, 02 - R, 03 - t, 04 - S, 05 - V, 06 - E, data, and project.

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

R Script R Notebook R Markdown... Shiny Web App... Plumber API...

Text File C++ File Python Script SQL Script Stan File D3 Script

R Sweave R HTML R Presentation R Documentation

Environment Global Enviro Data test_catalog

04-26> -- "Planting of a Tree"
R Foundation for Statistical Computing
nux-gnu (64-bit)

Step 2

comes with ABSOLUTELY NO WARRANTY.
istribute it under certain conditions.
ice() for distribution details.

roject with many contributors.
or more information and
cite R or R packages in publications.

demos, 'help()' for on-line help, or
HTML browser interface to help.

New Folder Cloud ..

.Rhists .Rprof 02 - R 03 - t 04 - S 05 - V 06 - E data project

File Edit Code View Plots Session Build Debug Profile Tools Help

+ | Go to file/function | Addins

Console Terminal x Jobs x

/cloud/project/ ↵

```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help,
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

> |

You may see this screen – click yes

Install Required Packages

? Create R Notebook requires an updated version of the rprojroot package.

Do you want to install this package now?

Yes No

Environment History Connections

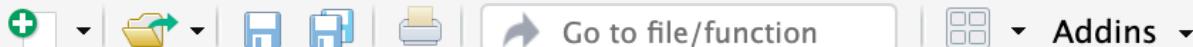
Import Dataset Global Environment

Data test_catalog 559 obs. of 2 variables

Pages Help Viewer

Upload Delete Rename More

		Size	Modified
	02 – Report.Rmd	0 B	Jul 4,
	03 – transform.Rmd	69 B	Jul 4,
	04 – Stats.Rmd	2.4 KB	Jul 4,
	05 – Visualize.Rmd	4.9 KB	Jul 4,
	06 – Exploratory Data Analysis.Rmd	3.6 KB	Jul 4,
	data	2.7 KB	Jul 4,
	project.Rproj	9.9 KB	Jul 4,
		205 B	Jul 4,



Rmd Untitled1 x



```
1 ---  
2 title: "R Notebook"  
3 output: html_notebook  
4 ---  
5  
6 This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute  
code within the notebook, the results appear beneath the code.  
7  
8 Try executing this chunk by clicking the *Run* button within the chunk or by  
placing your cursor inside it and pressing *Cmd+Shift+Enter*.  
9  
10```{r}  
11 plot(cars)  
12```  
13  
14 Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by  
pressing *Cmd+Option+I*.  
15  
16 When you save the notebook, an HTML file containing the code and output will be  
saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview  
the HTML file).  
17  
18 The preview shows you a rendered HTML copy of the contents of the editor.  
4:1 # R Notebook
```

Your Turn #2

1. Open up an R Notebook per guidance on next few slides.
2. Change the title of your R Notebook to “My First R Notebook” by modifying the header.
3. Add your name as the author by adding another line to the header:
Author: “Your Name”
3. Add a second level heading (##) at the end of the notebook called “My Calculation”



Code chunks

Open/close with
3 backticks

Language

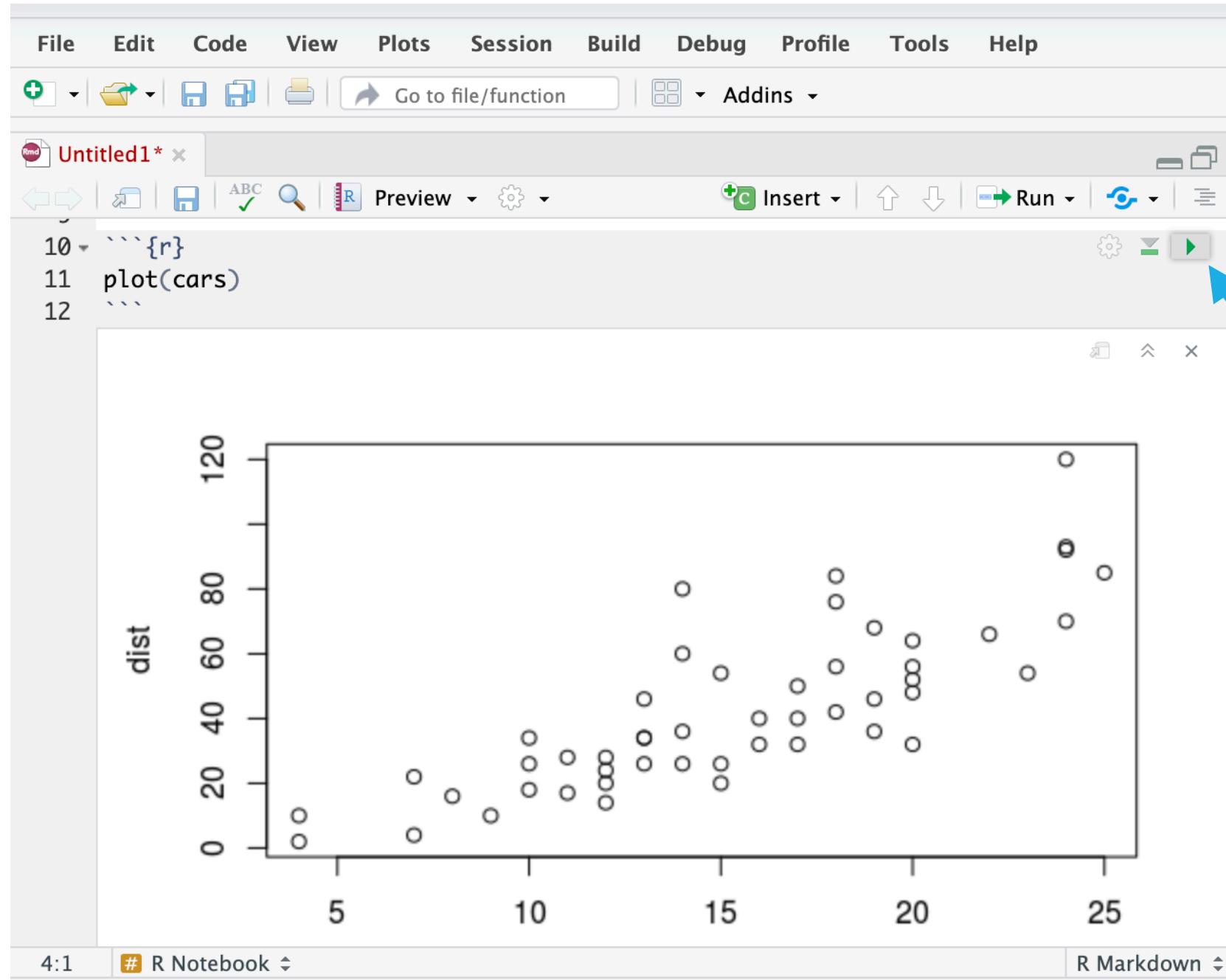
Chunk
name

Run chunk

Code in body of
chunk

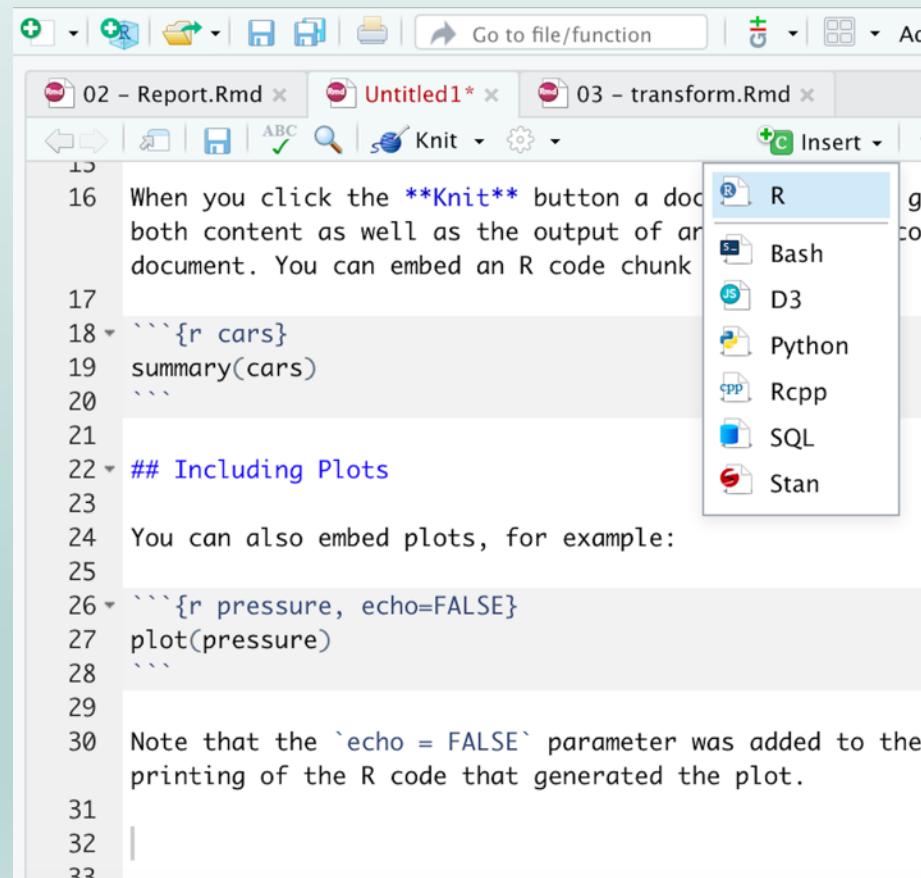
```
17  
18 - ``{r cars}  
19 summary(cars)  
20 ...  
21
```





Run code chunk

Your Turn #3



The screenshot shows the RStudio interface. The code editor contains the following R code:

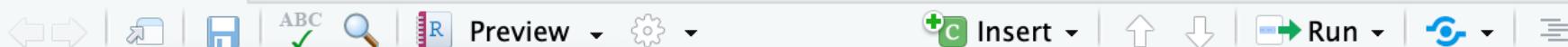
```
16 When you click the **Knit** button a doc  
both content as well as the output of an  
document. You can embed an R code chunk  
17  
18 ```{r cars}  
summary(cars)  
```  
19
20 ## Including Plots
21
22 You can also embed plots, for example:
23
24
25 ```{r pressure, echo=FALSE}
plot(pressure)
```  
26  
27 Note that the `echo = FALSE` parameter was added to the  
printing of the R code that generated the plot.  
28  
29  
30
```

An 'Insert' menu is open on the right side of the code editor, showing options for R, Bash, D3, Python, Rcpp, SQL, and Stan.

1. Under your new “My Calculation” heading, insert a code chunk into white space using Insert button on top right of code window
2. Add the following to your new code chunk:
`mean(c(10, 20, 30))`
3. Execute code chunk by pressing Run button on top right of code chunk



Rmd Untitled1*



10 ``

11 pl ``

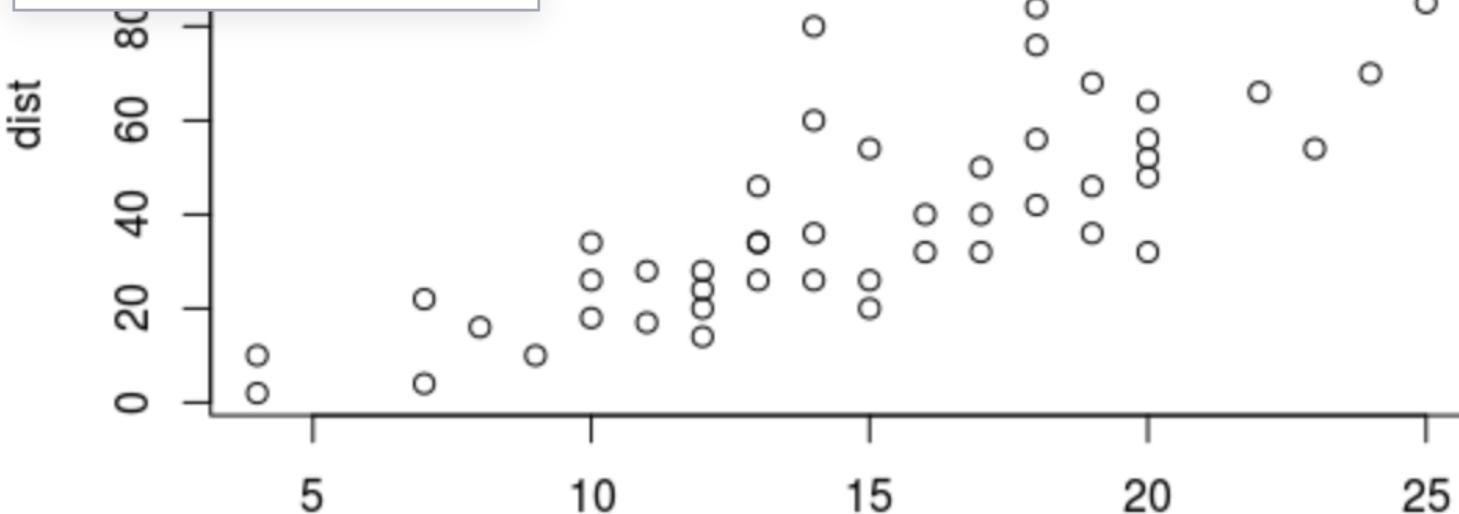
12

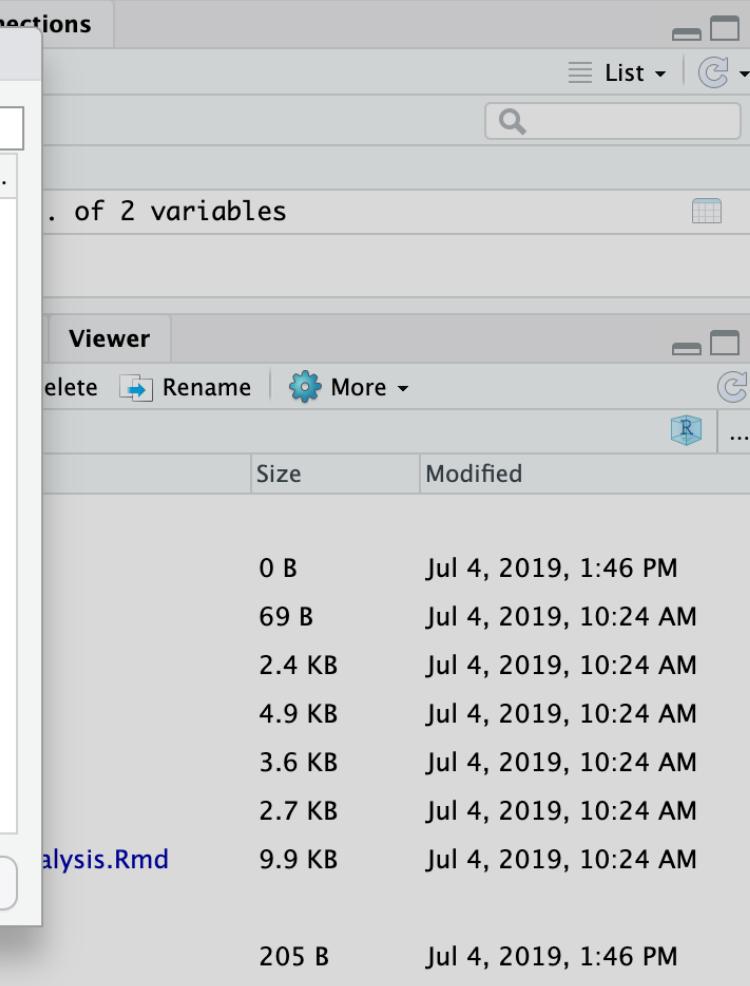
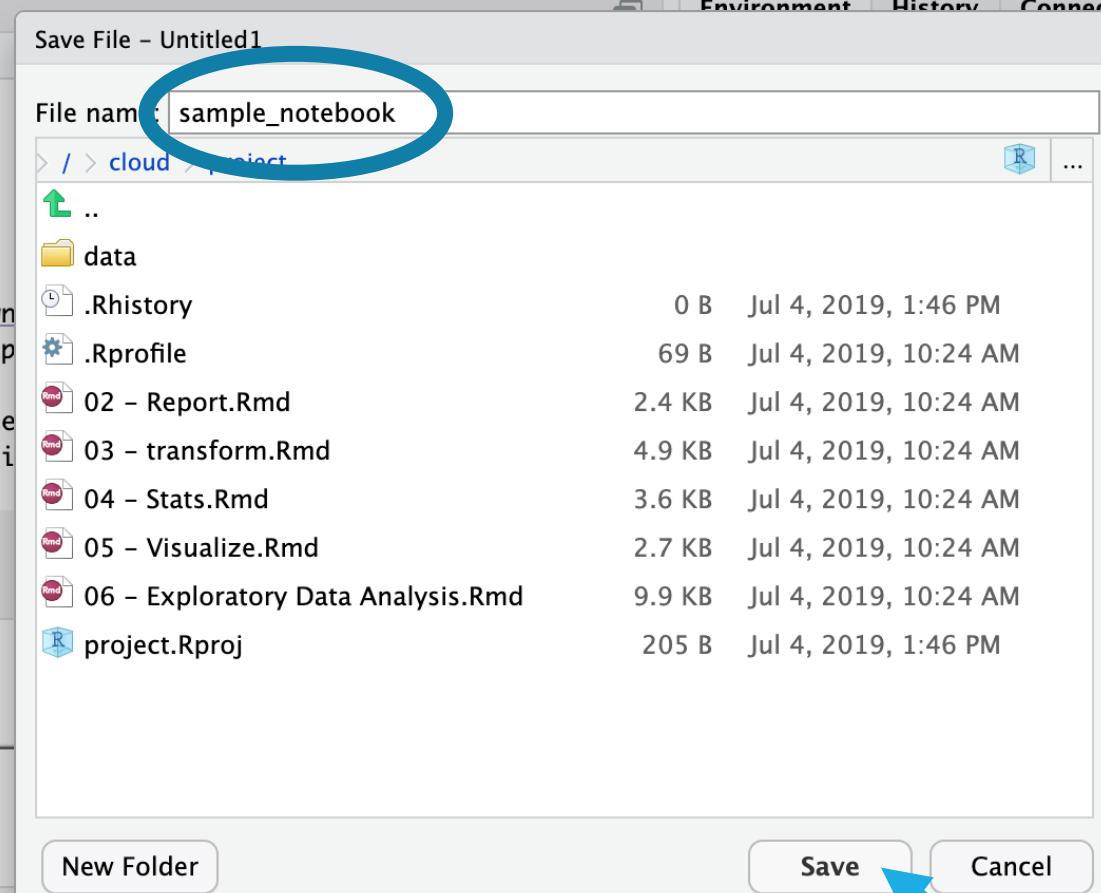
- Preview Notebook
- Knit to HTML
- Knit to PDF
- Knit to Word

Knit with Parameters...

Knit Directory ▾

Clear Knitr Cache...



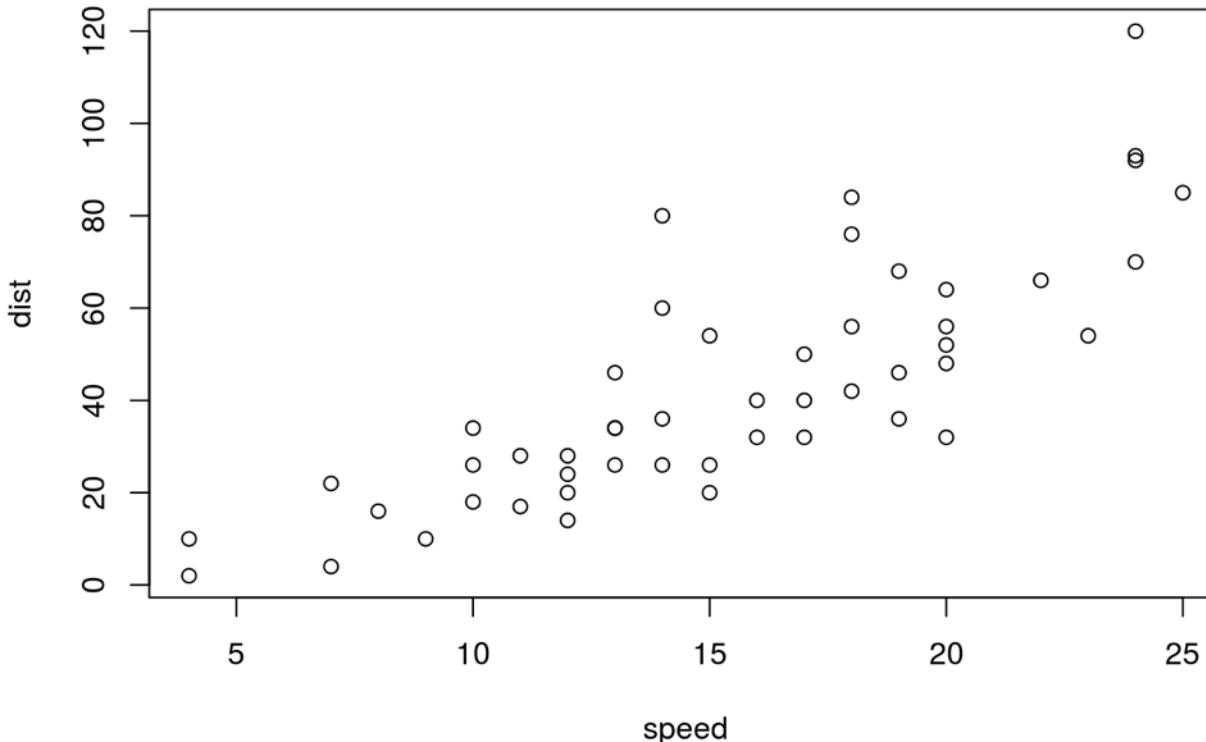


R Notebook

This is an [R Markdown](#) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

```
plot(cars)
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.



What Else?

Why name Code Chunks?

Jump between
chunks

The screenshot shows an RStudio interface with a code editor and a terminal. The code editor contains R code with numbered lines. A tooltip is displayed over line 24, which starts with 'Yo'. The tooltip content is 'Sample Markdown' with a list of items: 'Chunk 1: setup', 'R Markdown', 'Chunk 2: cars', 'Including Plots', and 'Chunk 3: pressure'. The 'Chunk 2: cars' item is highlighted with a blue background. The terminal below shows the R version and a directory path.

```
16 When you click the **Knit** button a document includes both content as well as the output of within the document. You can embed an R code c
17
18 ````{r cars}
19 summary(cars)
````

21
22 ## Including Plots
23
24 Yo Sample Markdown
25 Chunk 1: setup
26 ````{r cars}
27 pl R Markdown
28 ````{r cars}
29 Including Plots
30 No Chunk 3: pressure
```

# Sample Markdown

Console Terminal × Jobs ×

/cloud/project/ ↗

R version 3.5.2 (2018-12-20) -- "Faasheil Taloa"

# “Setup” Chunk & Chunk Options

chunk name  
(optional)

“chunk option”

don't show code in rendered document

```
6 - ```{r setup, include=FALSE}
7 library(tidyverse)
8 library(lubridate)
9 ...````
```

for dealing  
with dates

# Chunk options

| 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               |

# Keyboard Shortcuts

Insert a code chunk into white space within your open R Markdown document using:

- Windows: CTRL+ALT+i
- Mac: COMMAND+OPTION+I

Execute using shortcuts:

- Windows: CTRL+SHIFT+ENTER
- Mac: COMMAND+SHIFT+ENTER

# Multiple Output Formats are Available

Pandoc universal document converter can create multiple document types:

- html
- pdf
- docx

Can also create presentations and dashboards

- Including Powerpoint (most recent version of RStudio)

# Creating a pdf report

R & RStudio require additional packages to create a nicely formatted pdf report

Behind the scenes, R will use a markup language called LaTeX to turn your markdown into the pdf

Install the `tinytex` package with the following commands:

- `install.packages('tinytex')`
- `tinytex::install_tinytex()`

# Goals

1. Understand why reproducible reporting is important
2. Learn to work within R Markdown for reproducible reports

# Objectives

1. Create an R Markdown document and generate different types of output files
2. Practice modifying each component of a R Markdown file