

Introduction to R and Reproducible Reporting

Joseph Rudolf
AACC 2021

September 26, 2021	Session	Instructor
8:30 am – 8:45 am	Course Introduction	Patrick Mathias
8:45 am – 9:30 am	Intro to R and Reproducible Reporting	Joseph Rudolf
9:45 am – 10:30 am	Coding Basics and Importing Data	Joseph Rudolf
10:45 am – 11:30 am	Data Visualization	Patrick Mathias
LUNCH		
12:30 pm – 1:30 pm	Data Transformation	Patrick Mathias
1:45 pm – 2:45 pm	Grouping and Summarizing Data	Joseph Rudolf
3:00 pm – 3:30 pm	Dashboard Demo and Course Wrap Up	Patrick Mathias

Lesson Goals

1. Get oriented to the R programming environment
2. Appreciate the importance of reproducible data analysis

Lesson Objectives

1. Define R, RStudio and R Markdown
2. Log in and navigate RStudio Cloud
3. Discuss the risks of unreproducible data analysis
4. Create a R Markdown document



R

Programming
language for
data analysis



RStudio

Interactive
development
environment (IDE)

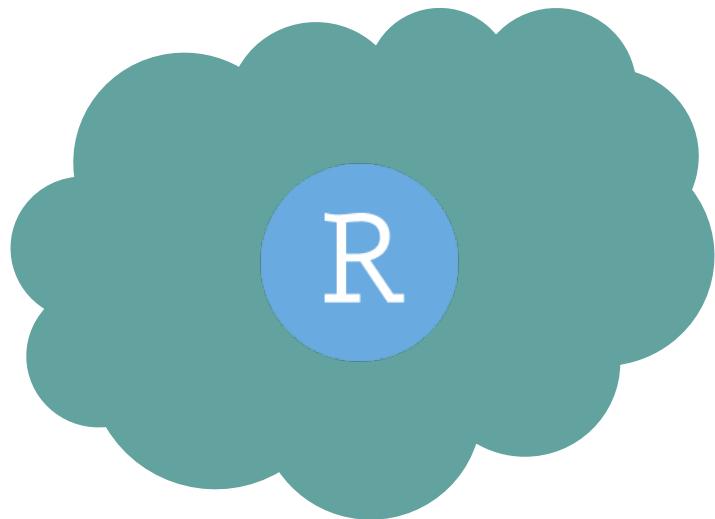


R Markdown

Computational
document format

Getting Started with RStudio

RStudio: On the Web and In Your Home



RStudio Cloud
Hosted on a server
(in the cloud)



RStudio Desktop
Installed locally on
your computer

Note: Use RStudio Cloud only for this course. Do not upload protected health information to the cloud!

Your Turn

Navigate to: <https://bit.ly/AACC-2021-R-Intro>

Enter your log in credentials

Join Space

Make a copy of the Core Exercises for yourself





Log In

Don't have an account?
Sign Up

Email

Continue

Forgot your password?

or

 Log In with Google

 Log In with GitHub

Join Space?

Joining a space gives you access to it and to its contents.

Once you join, admins will be able to see your email address.

Would you like to join this space?



Join Space

Cancel

The screenshot shows the RStudio Cloud interface with a sidebar on the left and a main content area on the right.

Left Sidebar:

- Spaces
- Your Workspace
- AACC 2021 Introduction to R (highlighted with a blue arrow labeled 1)
- AACC Introduction to R Copy
- New Space

Main Content Area:

AACC 2021 Introduction to R
Patrick Mathias

Projects Members About

All Projects

List All projects Sort By name

AACC-2021-Introduction-to-R
PM Patrick Mathias R RStudio Project Created Sep 19, 2021 12:22 PM

New Project

Copy Project Overlay:

Copy Project

Make your own copy of AACC-2021-Introduction-to-R?

OK

Three numbered arrows indicate the steps:

- An arrow points to the "AACC 2021 Introduction to R" project in the sidebar.
- An arrow points to the "Copy" button in the top right corner of the project card.
- An arrow points to the "OK" button in the "Copy Project" overlay.

Spaces

Your Workspace

AACC 2021 Introduction to R
Patrick MathiasAACC Introduction to R Cours
Patrick Mathias

New Space

Learn

Guide

What's New

Primers

Cheat Sheets

Help

Current System Status

RStudio Community

Info

Plans & Pricing

Terms and Conditions

File Edit Code View Plots Session Build Debug Profile Tools Help

           Go to file/function  Addins

R 4.1.0

Console Terminal  Jobs R 4.1.0 · /cloud/project/ 

```
R version 4.1.0 (2021-05-18) -- "Camp Pontanezen"  
Copyright (C) 2021 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, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

> |

Environment History Connections Git Tutorial

 Import Dataset  123 MiB R  Global Environment

Environment is empty

Files Plots Packages Help Viewer

 New Folder  Upload  Delete  Rename  More Cloud > project

Name	Size	Modified
..		
.gitignore	621 B	Sep 19, 2021, 12:22 PM
.Rhistory	0 B	Sep 19, 2021, 12:22 PM
coursepack		
exercises		
presentations		
project.Rproj	205 B	Sep 25, 2021, 12:55 PM
README.md	5.5 KB	Sep 19, 2021, 10:55 PM
resources.md	3.6 KB	Sep 19, 2021, 12:22 PM
solutions		
template.pptx	8.2 MB	Sep 19, 2021, 12:22 PM

Your Turn

Navigate to: <https://bit.ly/AACC-2021-R-Intro>

Enter your log in credentials

Join Space

Make a copy of the Core Exercises for yourself



EDITOR

The screenshot shows the RStudio Editor interface. A file named "Untitled1.Rmd" is open, containing R Markdown code. The code includes YAML front matter, a heading, and a code block. The "R Markdown" tab is selected at the bottom. Below the editor is the R Console, which displays the standard R startup message and information about the version and platform.

```
1 ---  
2 title: ''  
3 output:  
4  
5 ---  
6  
7 # Heading  
8  
9  
10 ``{r}  
11  
12 Code  
13  
14 ``  
15  
16  
8:1 # Heading
```

R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 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, or
'help.start()' for an HTML browser interface to help.

ENVIRONMENT

The screenshot shows the RStudio Environment pane. It displays the "Global Environment" tab, which shows that the environment is currently empty. The pane also includes tabs for Environment, History, Connections, and Tutorial, and features like Import Dataset and Global Environment search.

Environment is empty

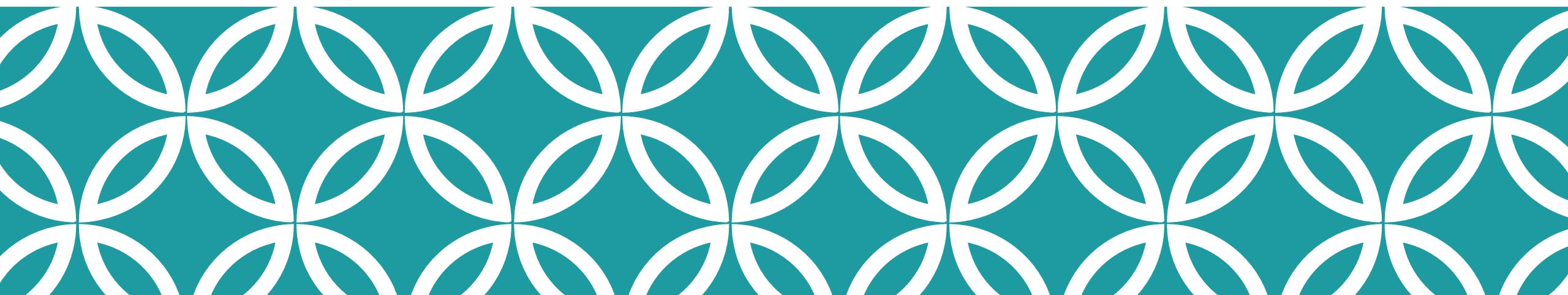
MISC

The screenshot shows the RStudio Files pane. It displays the "Home" directory structure, which contains three subfolders: "exercises", "R", and "solutions". The Files tab is selected at the top.

Name	Size	Modified
exercises		
R		
solutions		



Reproducible Data Analysis and R Markdown



The Duke Cancer Scandal

- ❖ Chemo sensitivity from microarrays
- ❖ Errors first, then cover-up
- ❖ Clinical trials based on flawed models
- ❖ Papers retracted, lawsuits settled



Duke

"1881_at"

"31321_at"

"31725_s_at"

"32307_r_at"

...

MD Anderson

"1882_g_at"

"31322_at"

"31726_at"

"32308_r_at"

Off-by-one indexing error

“Common problems are simple...

Off-by-one indexing error

Sensitive / resistant label reversal

Confounding in experimental design

Inclusion of data from non-reported sources

Wrong figure shown

... and simple problems are common.”

Point-and-click is not reproducible



Computer code can precisely document each step of the analysis

Why YOU should analyze your data 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?”

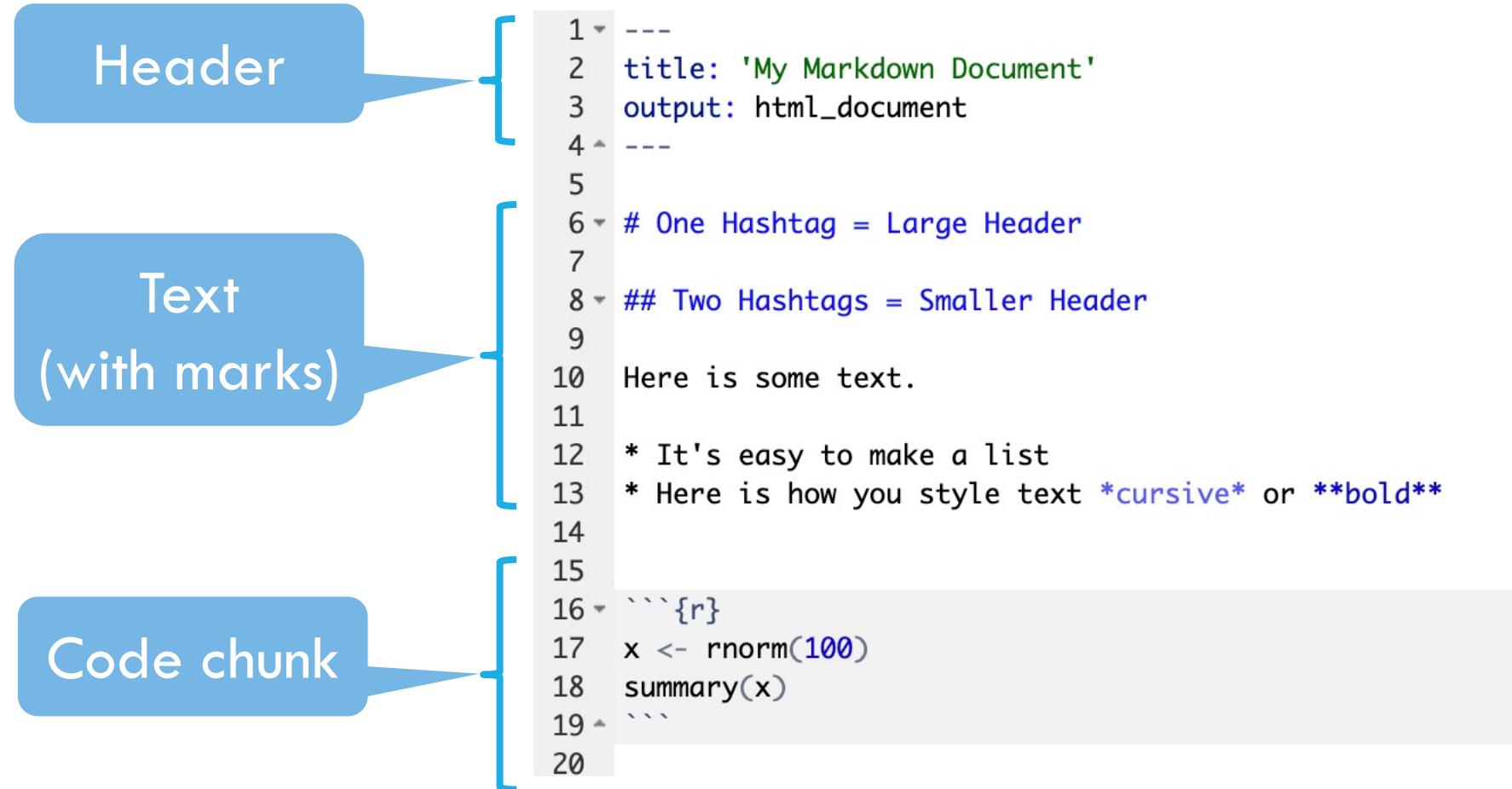
“Why did I decide to omit these six samples from my analysis?”



YOUR CLOSEST COLLABORATOR IS YOU FROM 6 MONTHS AGO



Anatomy of an R Markdown Document



```

1 ---  

2 title: 'My Markdown Document'  

3 output: html_document  

4 ---  

5  

6 # One Hashtag = Large Header  

7  

8 ## Two Hashtags = Smaller Header  

9  

10 Here is some text.  

11  

12 * It's easy to make a list  

13 * Here is how you style text *cursive* or **bold**  

14  

15  

16 ```{r}  

17 x <- rnorm(100)  

18 summary(x)  

19 ````  

20  

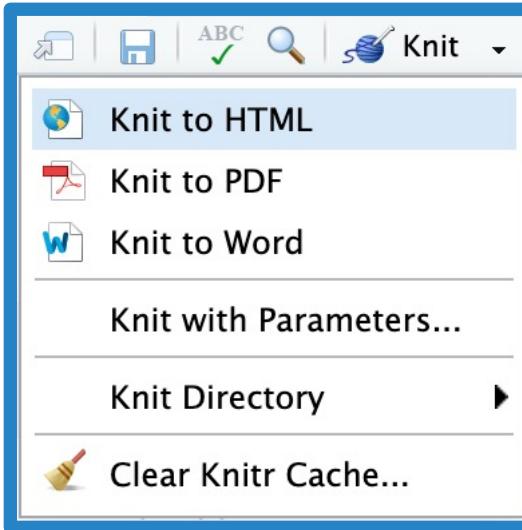
21 ## Including Plots|  

22  

23 ```{r, echo=FALSE}  

24 hist(x)  

25 ````
```



My Markdown Document

One Hashtag = Large Header

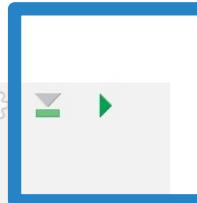
Two Hashtags = Smaller Header

Here is some text.

- It's easy to make a list
- Here is how you style text *cursive* or **bold**

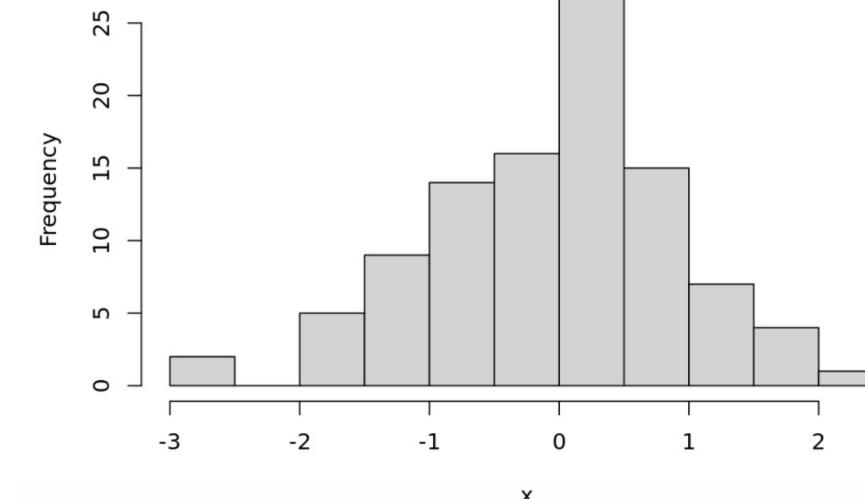
```
x <- rnorm(100)
summary(x)
```

```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
## -2.99204 -0.64726  0.14853 -0.02832  0.58218  2.07410
```



Including Plots

Histogram of x



```
1 --
2 title: 'My Markdown Document'
3 output: html_document
4 ---
```

```
5
6 # One Hashtag = Large Header
7
8 ## Two Hashtags = Smaller Header
9
10 Here is some text.
11
12 * It's easy to make a list
13 * Here is how you style text *cursive* or **bold**
14
```

```
15
16 ``{r}
17 x <- rnorm(100)
18 summary(x)
19 ````
```



My Markdown Document

One Hashtag = Large Header

Two Hashtags = Smaller Header

Here is some text.

- It's easy to make a list
- Here is how you style text *cursive* or **bold**

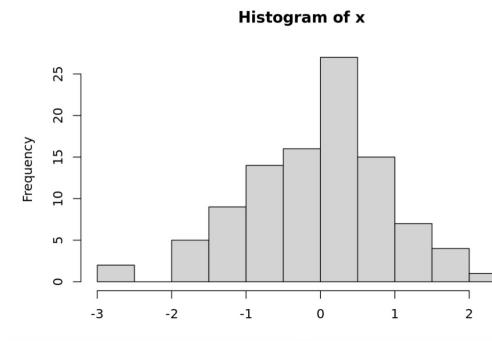
```
x <- rnorm(100)
summary(x)
```

```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## -2.99204 -0.64726  0.14853 -0.02832  0.58218  2.07410
```

```
20
21 ## Including Plots|
22
23 ``{r, echo=FALSE}
24 hist(x)
25 ````
```



Including Plots



Your Turn #2

Open a sample R Markdown document (File -> New File -> R Markdown).

Review the format of the document: header, text, code chunks

Execute the individual code chunks by selecting the Run Current Chunk arrow.

Knit the document to HTML (Preview or Knit Button -> Knit to HTML). You may be prompted to save your R Markdown first. In this case select a name for your document and click save. Review the knitted document.



Recap



Programming
Language



IDE

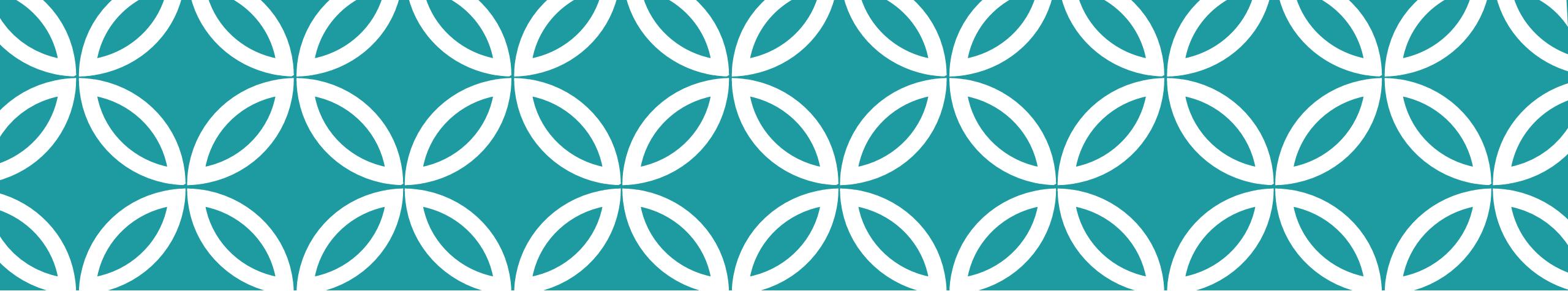


Document
Format

The **RStudio Panes** include editor (writing code), environment (interact with objects), misc (interact with files), and console (enter individual commands).

Reproducible Data Analysis is a best practice for working with clinical and research data.

R Markdown provides us with an electronic lab notebook to mix executable data analysis code elements with annotation for effective reproducibility.



What else?

rmarkdown :: CHEAT SHEET

What is rmarkdown?

.Rmd files • Develop your code and ideas side-by-side in a single document. Run code as individual chunks or as an entire document.

Dynamic Documents • Knit together plots, tables, and results with narrative text. Render to a variety of formats like HTML, PDF, MS Word, or MS Powerpoint.

Reproducible Research • Upload, link to, or attach your report to share. Anyone can read or run your code to reproduce your work.

Workflow

- 1 Open a new .Rmd file in the RStudio IDE by going to File > New File > R Markdown.
- 2 Embed code in chunks. Run code by line, by chunk, or all at once.
- 3 Write text and add tables, figures, images, and citations. Format with Markdown syntax or the RStudio Visual Markdown Editor.
- 4 Set output format(s) and options in the YAML header. Customize themes or add parameters to execute or add interactivity with Shiny.
- 5 Save and render the whole document. Knit periodically to preview your work as you write.
- 6 Share your work!

Embed Code with knitr

CODE CHUNKS

Surround code chunks with `{{r}}` and `{{` or use the Insert Code Chunk button. Add a chunk label and/or chunk options inside the curly braces after {{r}}.

```
```{r chunk-label, include=FALSE}
summary(mtcars)
```
```

SET GLOBAL OPTIONS

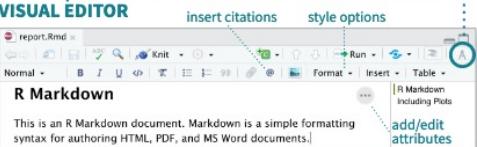
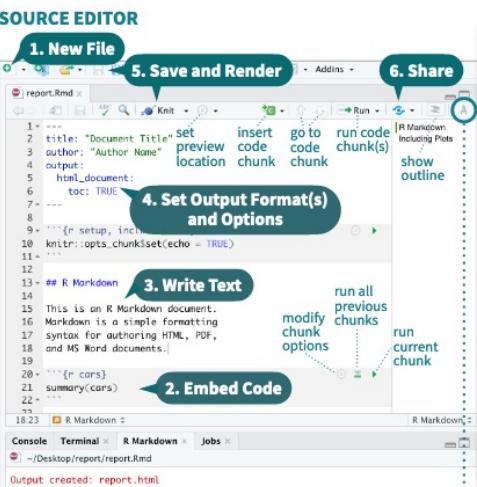
Set options for the entire document in the first chunk.

```
```{r include=FALSE}
knitr::opts_chunk$set(message = FALSE)
```
```

INLINE CODE

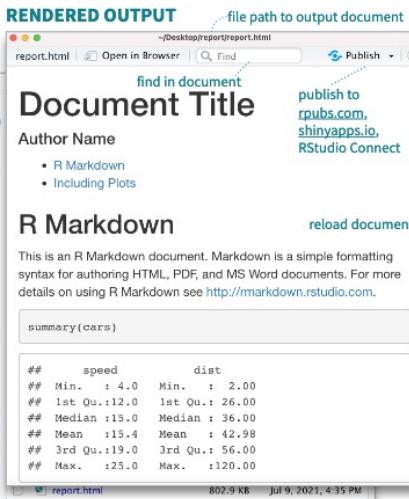
Insert `{{r <code>}}` into text sections. Code is evaluated at render and results appear as text.

"Built with `{{r getRversion()}}`" -> "Built with 4.1.0"



| OPTION | DEFAULT EFFECTS |
|------------------------|---|
| echo | TRUE display code in output document |
| error | FALSE (display error messages in doc) |
| eval | TRUE run code in chunk |
| include | TRUE include chunk in doc after running |
| message | TRUE display code messages in document |
| warning | TRUE display code warnings in document |
| results | "markup" "hide" (don't display results)
"hold" (put all results below all code) |
| fig.align | "default" "left", "right", or "center" |
| fig.alt | NULL alt text for a figure |
| fig.cap | figure caption as a character string |
| fig.path | "figure/" prefix for generating figure file paths |
| fig.width & fig.height | 7 plot dimensions in inches |
| out.width | rescales output width, e.g. "75%", "300px" |
| collapse | FALSE collapse all sources & output into a single block |
| comment | "##" prefix for each line of results |
| child | NULL files(s) to knit and then include |
| purl | TRUE include or exclude a code chunk when extracting source code with knitr::purl() |

See more options and defaults by running `str(knitr::opts_chunk$get())`



Insert Citations

Create citations from a bibliography file, a Zotero library, or from DOI references.

BUILD YOUR BIBLIOGRAPHY

- Add BibTeX or CSL bibliographies to the YAML header.


```
---
title: "My Document"
bibliography: references.bib
link-citations: TRUE
```
- If Zotero is installed locally, your main library will automatically be available.
- Add citations by DOI by searching "from DOI" in the **Insert Citation** dialog.

INSERT CITATIONS

- Access the **Insert Citations** dialog in the Visual Editor by clicking the @ symbol in the toolbar or by clicking **Insert > Citation**.
- Add citations with markdown syntax by typing `[@cite]` or `@cite`.

Insert Tables

Output data frames as tables using `kable(data, caption)`.

```
```{r}
data <- faithful[1:4,]
knitr::kable(data,
 caption = "Table with kable")
```
```

Other table packages include `flextable`, `gt`, and `kableExtra`.

Write with Markdown

The syntax on the left renders as the output on the right.

Plain text.

End a line with two spaces to start a new paragraph. Also end with a backslash\ to make a new line.

italics and **bold**

superscript²/subscript₂

~~strikethrough~~

escaped: `_`

endash: ---, emdash: —

Header 1

Header 2

Header 6

- unordered list
 - item 2
 - item 2a (indent 1 tab)
 - item 2b
- ordered list
 - item 2
 - item 2a (indent 1 tab)
 - item 2b

<http://www.rstudio.com/>
This is a link.

This is another link.



Caption.

verbatim code

multiple lines of verbatim code

block quotes

equation: $\sum_{i=1}^n i = \frac{n(n+1)}{2}$

equation block:

$$E = mc^2$$

horizontal rule:

| Right | Left | Default | Center |
|-------|------|---------|--------|
| 12 | 12 | 12 | 12 |
| 123 | 123 | 123 | 123 |
| 1 | 1 | 1 | 1 |

HTML Tables

Results {table}

Plots text

text

Tables

more text



HTML

