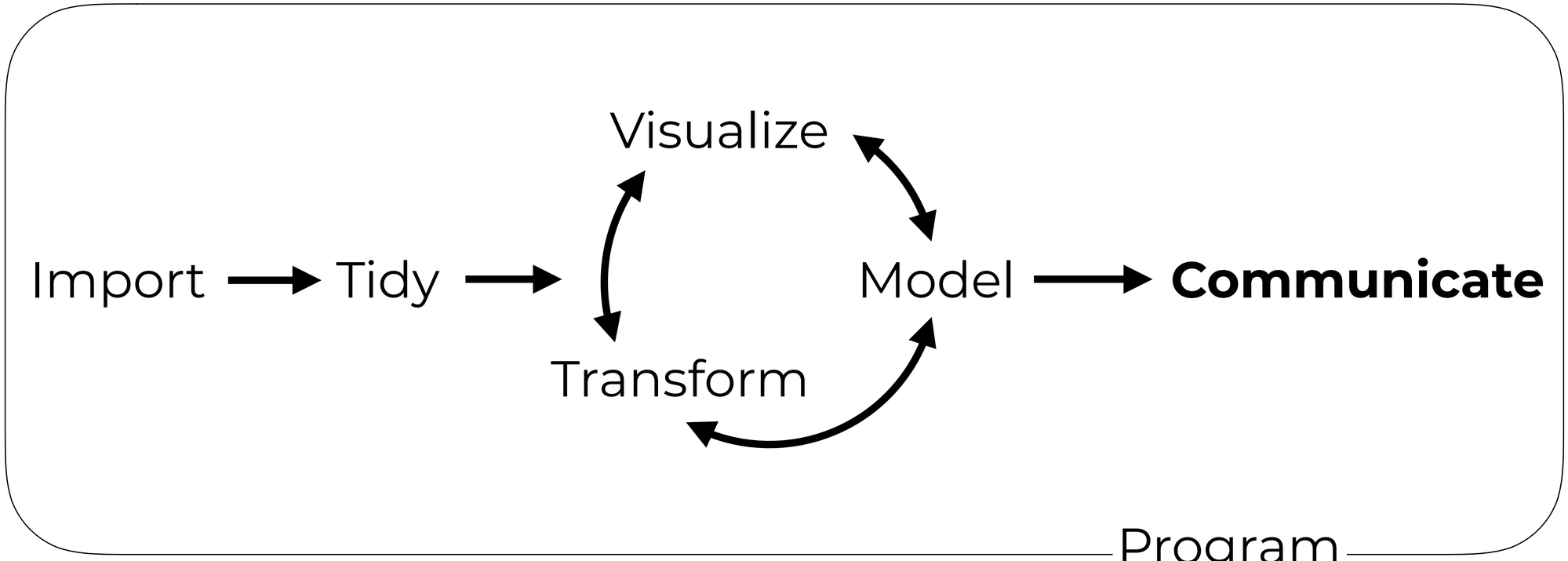
Communicating Results

Jake Thompson
wjakethompson.com
@wjakethompson

ATLAS
Accessible Teaching, Learning, and Assessment Systems

CC BY



Adapted from Master the Tidyverse, CC BY RStudio

Consider

- What is the definition of a reproducible data analysis?
- What tools can you use to make a data analysis reproducible?

Reproducibility

- Are tables and figures reproducible from the code and data?
- Does the code do what you think it does?
- Is it clear **why** things were done?
- Can the code be used for other data?

From *Steps to Reproducible Research*, Karl Broman

In R4DS

R Markdown

ATLAS

Markdown

- Lightweight markup language with plain text formatting syntax
- Designed to be converted to HTML (and other formats, e.g. PDF)

```
This is the repo for *"Using R and the tidyverse for Data Science"*
given at the 2019 Summer Statistical Institute in May 2019.

## Description

This workshop is designed for those who want to learn how to use R to
analyze data. The material is based on Hadley Wickham and Garrett
Grolemund's [*R for Data Science*](https://r4ds.had.co.nz/index.html).
We'll talk about how to conduct a complete data analysis from data
import to final reporting in R using a suite of packages known as the
[**tidyverse**](https://tidyverse.org). The two goals of this workshop
are: 1) learn how to use R to answer questions about our data; and 2)
write code that is human readable and reproducible. We will also talk
about how to share our code and analyses with others.

## Software Requirements

In order to install the workshop materials (e.g., slides, exercises),
you'll need **a laptop that can access the internet** (wifi will be
available). Before the workshop, install the following:

1.  A recent version of R (~3.6.0), which is available for free at
    [cran.r-project.org](http://www.cran.r-project.org)
2.  A recent version of RStudio IDE (~1.2.1335), available for free at
    [www.rstudio.com/download](http://www.rstudio.com/download)
3.  The set of relevant R packages, which you can install by connecting
    to the internet, opening RStudio, and running the code in `setup.R`.

```

README.md

This is the repo for *"Using R and the tidyverse for Data Science"* given at the 2019 Summer Statistical Institute in May 2019.

Description

This workshop is designed for those who want to learn how to use R to analyze data. The material is based on Hadley Wickham and Garrett Grolemund's *R for Data Science*. We'll talk about how to conduct a complete data analysis from data import to final reporting in R using a suite of packages known as the **tidyverse**. The two goals of this workshop are: 1) learn how to use R to answer questions about our data; and 2) write code that is human readable and reproducible. We will also talk about how to share our code and analyses with others.

Software Requirements

In order to install the workshop materials (e.g., slides, exercises), you'll need **a laptop that can access the internet** (wifi will be available). Before the workshop, install the following:

1. A recent version of R (~3.6.0), which is available for free at cran.r-project.org
2. A recent version of RStudio IDE (~1.2.1335), available for free at www.rstudio.com/download
3. The set of relevant R packages, which you can install by connecting to the internet, opening RStudio, and running the code in `setup.R`.

More detailed instructions can be found on the workshop website. Don't forget to bring your power cord!

R Markdown

- Markdown + R
- Text + R code (in chunks) gets converted to text + R code + R output
- 3 types of content:
  - YAML header surrounded by ---
  - R code chunks surrounded by ```
  - Text with markdown formatting

Markdown Formatting

Text and Headers

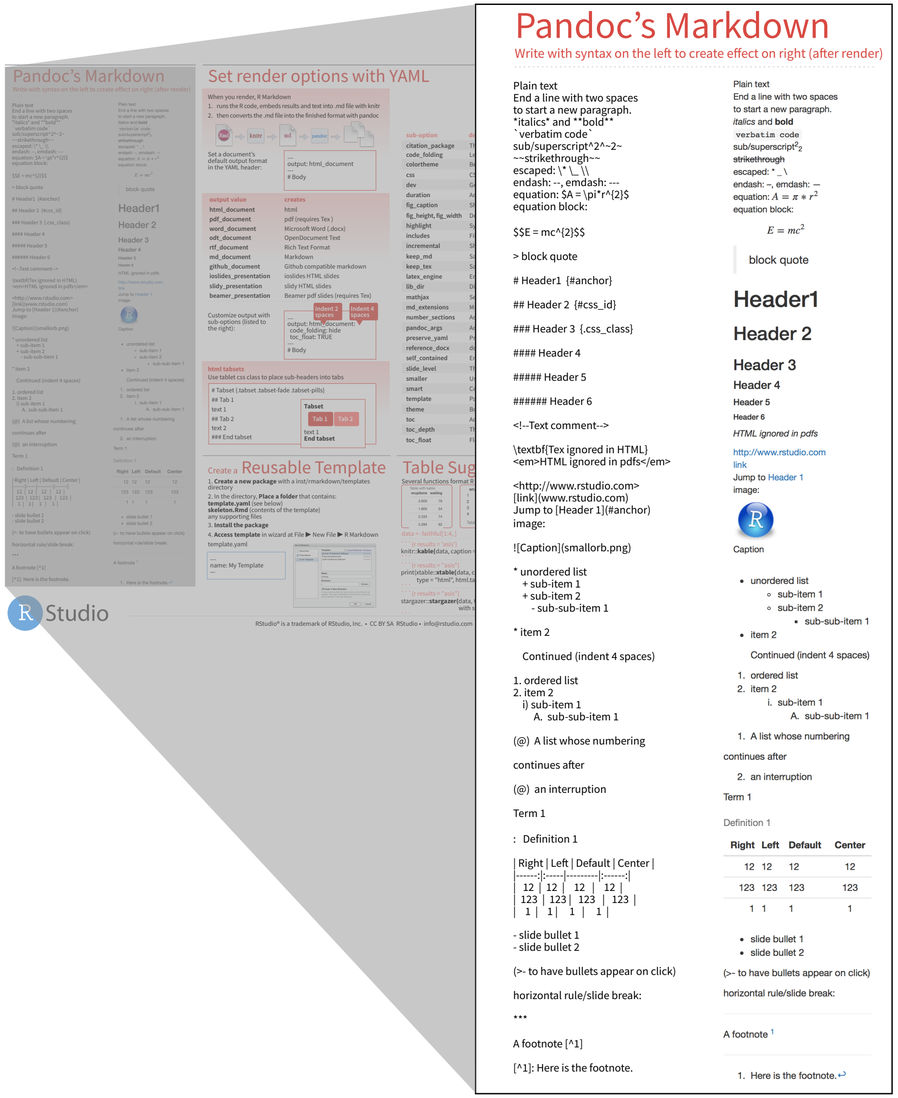
- Text can be `plain`, `*italic*`, `**bold**`, or `` `code` ``
- Headers use #

```
# Header 1
## Header 2
### Header 3
```

Links

- Plain web address, or custom phrase
  - `https://tidy-ds.wjakethompson.com`
  - `[Workshop Website](https://tidy-ds.wjakethompson.com)`
- Images from the web or local
  - `![RStudio logo](https://www.rstudio.com/wp-content/uploads/2018/10/RStudio-Logo-Flat.png)`
  - `![RStudio logo](resources/rstudio-logo.png)`

Alt text



Pandoc's Markdown

Write with syntax on the left to create effect on right (after render)

```
Plain text
End a line with two spaces
to start a new paragraph.
*italics* and **bold**
`verbatim code`
sub/superscript^2^~2~
~~strikethrough~~
escaped: \* \_ \\
endash: --, emdash: ---
equation: $A = \pi*r^{2}$
equation block:

$$E = mc^{2}$$

> block quote

# Header1  {#anchor}

## Header 2  {#css_id}

### Header 3  {.css_class}

#### Header 4

##### Header 5

###### Header 6

<!--Text comment-->

\textbf{Tex ignored in HTML}
<em>HTML ignored in pdfs</em>

<http://www.rstudio.com>
[link](www.rstudio.com)
Jump to [Header 1](#anchor)
image:

![Caption](smallorb.png)

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

* item 2

  Continued (indent 4 spaces)

1. ordered list
2. item 2
   i) sub-item 1
      A. sub-sub-item 1

(@)  A list whose numbering

continues after

(@)  an interruption

Term 1

:  Definition 1

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

- slide bullet 1
- slide bullet 2

(>- to have bullets appear on click)

horizontal rule/slide break:

***

A footnote [^1]

[^1]: Here is the footnote.
```

Plain text
End a line with two spaces
to start a new paragraph.
italics and **bold**
`verbatim code`
sub/superscript^2~2~
~~strikethrough~~
escaped: * _ \
endash: –, emdash: —
equation: $A = \pi * r^2$
equation block:

$$E = mc^2$$

> block quote

Header1

Header 2

Header 3

Header 4

Header 5

Header 6

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

  Continued (indent 4 spaces)

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

(@) A list whose numbering

continues after

(@) an interruption

Term 1

Definition 1

Right	Left	Default	Center
12	12	12	12
123	123	123	123
1	1	1	1

- slide bullet 1
- slide bullet 2

(>- to have bullets appear on click)

horizontal rule/slide break:

A footnote [1]

1. Here is the footnote. ↩

R Markdown Reference Guide

Learn more about R Markdown at rmarkdown.rstudio.com
Learn more about Interactive Docs at shiny.rstudio.com/articles

Contents:
1. **Markdown Syntax**
2. Knitr chunk options
3. Pandoc options

Syntax	Becomes
`Plain text`	Plain text
`End a line with two spaces to start a new paragraph.`	End a line with two spaces to start a new paragraph.
`*italics* and _italics_`	*italics* and *italics*
`**bold** and __bold__`	**bold** and **bold**
`superscript^2^`	superscript^2
`~~strikethrough~~`	~~strikethrough~~
`[link](www.rstudio.com)`	link
`# Header 1`	Header 1
`## Header 2`	Header 2
`### Header 3`	Header 3
`#### Header 4`	Header 4
`##### Header 5`	Header 5
`###### Header 6`	Header 6
`endash: --`	endash: –
`emdash: ---`	emdash: —
`ellipsis: ...`	ellipsis: …
`inline equation: $A = \pi*r^{2}$`	inline equation: $A = \pi * r^2$
`image: ![](path/to/smallorb.png)`	image:
`horizontal rule (or slide break):`	horizontal rule (or slide break):
`***`	
`> block quote`	> block quote

`* unordered list`
`* item 2`
`  + sub-item 1`
`  + sub-item 2`
	• unordered list
	• item 2
	  ◦ sub-item 1
	  ◦ sub-item 2

`1. ordered list`
`2. 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`
	Table Header / Second Header
	Table Cell / Cell 2
	Cell 3 / Cell 4

RStudio Updated 10/30/2014 © 2014 RStudio, Inc. CC BY RStudio.

1

RStudio:

Help → Markdown Quick Reference

Your Turn 1

- Open the **tidyds-2019.Rproj** project
- File > New File > R Markdown...
- Adjust the title, author, ...
- Add an image or equation
- Save your document as **11-Communicate/my-first-rmd.Rmd**
- Compile document to HTML

05 : 00

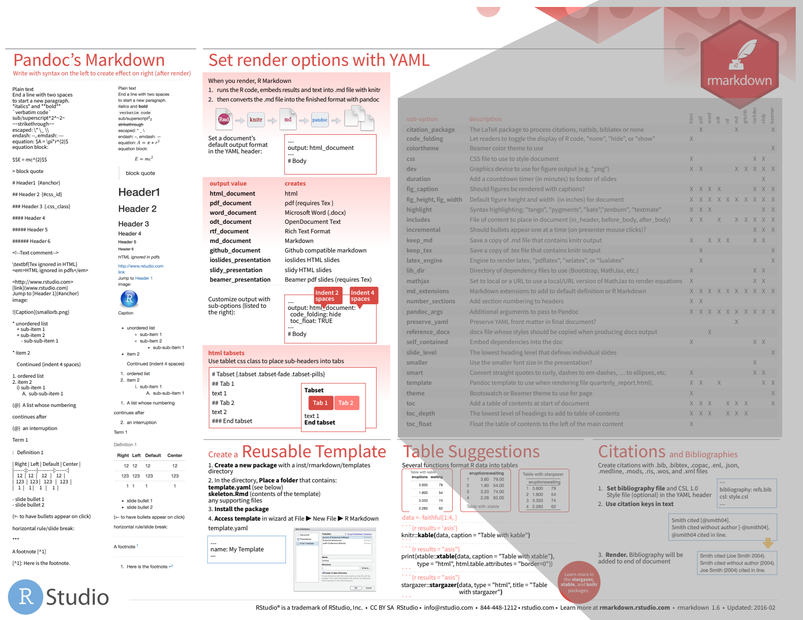
YAML

Customize Output

```
---
title: "Communicate - Solution"
author: "Jake Thompson"
date: "`r Sys.Date()`"
output: html_document
---
```

Customize Output

```
---
title: "Communicate - Solution"
author: "Jake Thompson"
date: "`r Sys.Date()`"
output:
  html_document:
    number_sections: true
    toc: true
---
```



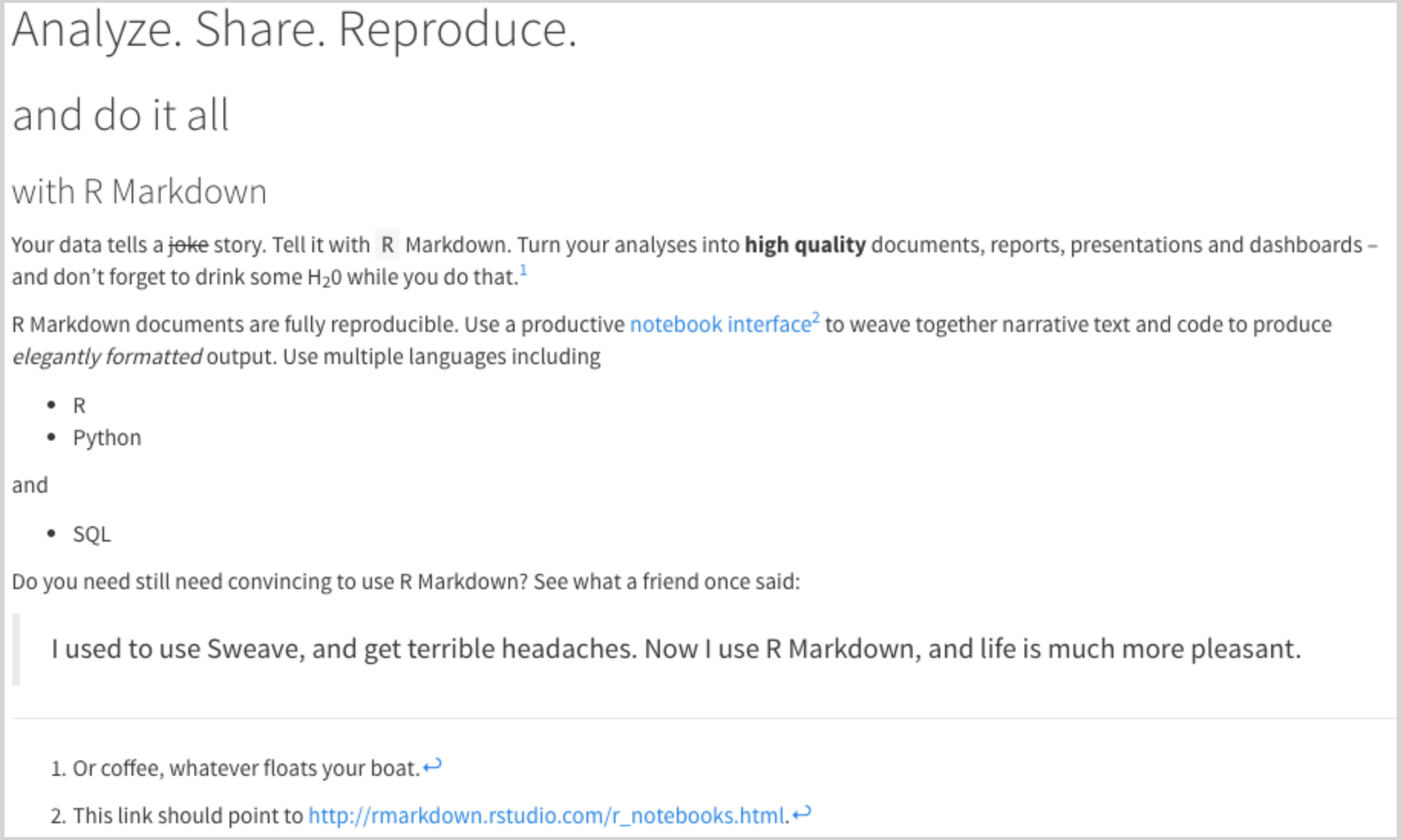
sub-option	description	html	pdf	word	odt	rtf	md	gituhb	ioslides	slidy	beamer
citation_package	The LaTeX package to process citations, natbib, biblatex or none		X				X				X
code_folding	Let readers to toggle the display of R code, "none", "hide", or "show"	X									
colortheme	Beamer color theme to use										X
css	CSS file to use to style document	X							X	X	
dev	Graphics device to use for figure output (e.g. "png")	X	X				X	X	X	X	X
duration	Add a countdown timer (in minutes) to footer of slides									X	
fig_caption	Should figures be rendered with captions?	X	X	X	X				X	X	X
fig_height, fig_width	Default figure height and width (in inches) for document	X	X	X	X	X	X	X	X	X	X
highlight	Syntax highlighting: "tango", "pygments", "kate","zenburn", "textmate"	X	X	X					X	X	
includes	File of content to place in document (in_header, before_body, after_body)	X	X		X		X	X	X	X	X
incremental	Should bullets appear one at a time (on presenter mouse clicks)?								X	X	X
keep_md	Save a copy of .md file that contains knitr output	X		X	X	X			X	X	
keep_tex	Save a copy of .tex file that contains knitr output		X								X
latex_engine	Engine to render latex, "pdflatex", "xelatex", or "lualatex"		X								X
lib_dir	Directory of dependency files to use (Bootstrap, MathJax, etc.)	X							X	X	
mathjax	Set to local or a URL to use a local/URL version of MathJax to render equations	X							X	X	
md_extensions	Markdown extensions to add to default definition or R Markdown	X	X	X	X	X	X	X	X	X	X
number_sections	Add section numbering to headers	X	X						X		
pandoc_args	Additional arguments to pass to Pandoc	X	X	X	X	X	X	X	X	X	X
preserve_yaml	Preserve YAML front matter in final document?						X				
reference_docx	docx file whose styles should be copied when producing docx output			X							
self_contained	Embed dependencies into the doc	X							X	X	
slide_level	The lowest heading level that defines individual slides										X
smaller	Use the smaller font size in the presentation?								X		
smart	Convert straight quotes to curly, dashes to em-dashes, … to ellipses, etc.	X							X	X	
template	Pandoc template to use when rendering file quarterly_report.html).	X	X		X					X	X
theme	Bootswatch or Beamer theme to use for page	X									X
toc	Add a table of contents at start of document	X	X	X		X	X	X			X
toc_depth	The lowest level of headings to add to table of contents	X	X	X		X	X	X			
toc_float	Float the table of contents to the left of the main content	X									

Your Turn 2

- Edit the YAML for **11-Communicate/my-first-rmd.Rmd**
- Add a table of contents
- Change the theme to **darkly**

```
---
...
output:
  html_document:
    toc: true
    theme: "darkly"
---
```

Your Turn 3



Analyze. Share. Reproduce.

and do it all

with R Markdown

Your data tells a ~~joke~~ story. Tell it with `R` Markdown. Turn your analyses into **high quality** documents, reports, presentations and dashboards – and don't forget to drink some H_20 while you do that.[1]

R Markdown documents are fully reproducible. Use a productive notebook interface[2] to weave together narrative text and code to produce *elegantly formatted* output. Use multiple languages including

- R
- Python

and

- SQL

Do you need still need convincing to use R Markdown? See what a friend once said:

> I used to use Sweave, and get terrible headaches. Now I use R Markdown, and life is much more pleasant.

1. Or coffee, whatever floats your boat. ↩
2. This link should point to http://rmarkdown.rstudio.com/r_notebooks.html. ↩

- Open **11-Communicate/analyze-share-repro.Rmd**
- Add styling to the text to create the look on the left.

10 : 00

Putting the R in R Markdown

Run R code inside Markdown

- Code chunks
  - Most powerful
  - Control over computations
  - Handles plots and other non-text output
- Inline code
  - Simple, but only handles text
  - Character strings (including Markdown formatting)
  - Numbers (but not vectors)

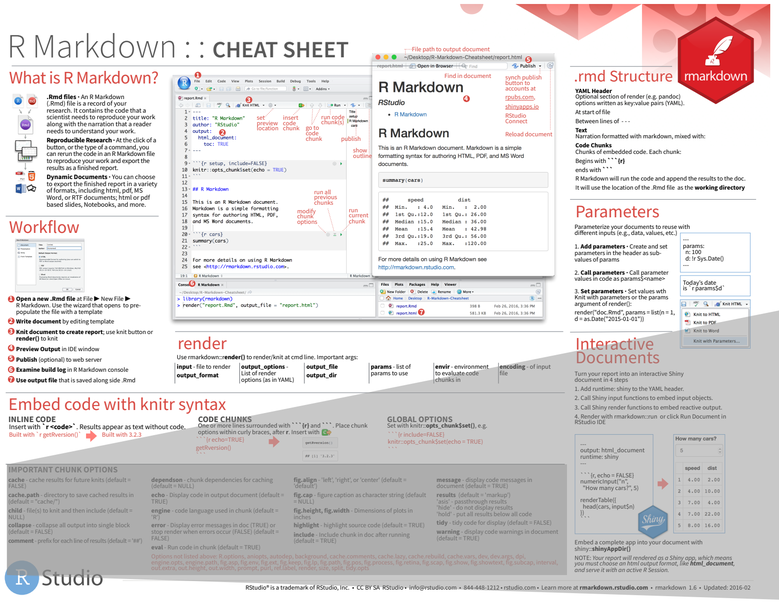
Code chunks

- "Fenced" regions of code
- When report is rendered:
  - Each code chunk is run
  - Output placed directly beneath the code chunk
- Insert new chunks with:
  - *Add Chunk* button in toolbar
  - Keyboard shortcut:
    - Mac: **Cmd** + **option** + **I**
    - Windows: **Ctrl** + **Alt** + **I**

Chunk options

Control how the chunk is displayed

- Should the code display? Or just the output?
- Include errors and warnings?
- How should figures be rendered?



IMPORTANT CHUNK OPTIONS

cache - cache results for future knits (default = FALSE)

cache.path - directory to save cached results in (default = "cache/")

child - file(s) to knit and then include (default = NULL)

collapse - collapse all output into single block (default = FALSE)

comment - prefix for each line of results (default = '##')

dependson - chunk dependencies for caching (default = NULL)

echo - Display code in output document (default = TRUE)

engine - code language used in chunk (default = 'R')

error - Display error messages in doc (TRUE) or stop render when errors occur (FALSE) (default = FALSE)

eval - Run code in chunk (default = TRUE)

fig.align - 'left', 'right', or 'center' (default = 'default')

fig.cap - figure caption as character string (default = NULL)

fig.height, fig.width - Dimensions of plots in inches

highlight - highlight source code (default = TRUE)

include - Include chunk in doc after running (default = TRUE)

message - display code messages in document (default = TRUE)

results (default = 'markup')
'asis' - passthrough results
'hide' - do not display results
'hold' - put all results below all code

tidy - tidy code for display (default = FALSE)

warning - display code warnings in document (default = TRUE)

Options not listed above: R.options, aniopts, autodep, background, cache.comments, cache.lazy, cache.rebuild, cache.vars, dev, dev.args, dpi, engine.opts, engine.path, fig.asp, fig.env, fig.ext, fig.keep, fig.lp, fig.path, fig.pos, fig.process, fig.retina, fig.scap, fig.show, fig.showtext, fig.subcap, interval, out.extra, out.height, out.width, prompt, purl, ref.label, render, size, split, tidy.opts

Some normal text.

```{r make-plot, echo = FALSE, fig.align = "c"}
ggplot(mtcars, aes(mpg, disp)) +
  geom_point()
```

Some more plain text.

````
Some normal text.

```{r make-plot, echo = FALSE, fig.align = "c"}
ggplot(mtcars, aes(mpg, disp)) +
  geom_point()
```

te plain text.
````

Language engine

````
Some normal text.

```{r make-plot, echo = FALSE, fig.align = "c"}
ggplot(mtcars, aes(mpg, disp)) +
  geom_point()
```

Some normal text.
````

Language engine

Chunk label

````
Some normal text.

```{r make-plot, echo = FALSE, fig.align = "c"}
ggplot(mtcars, aes(mpg, disp)) +
  geom_point()
```

Some normal text.
````

Language engine

Chunk label

Chunk options

Inline code

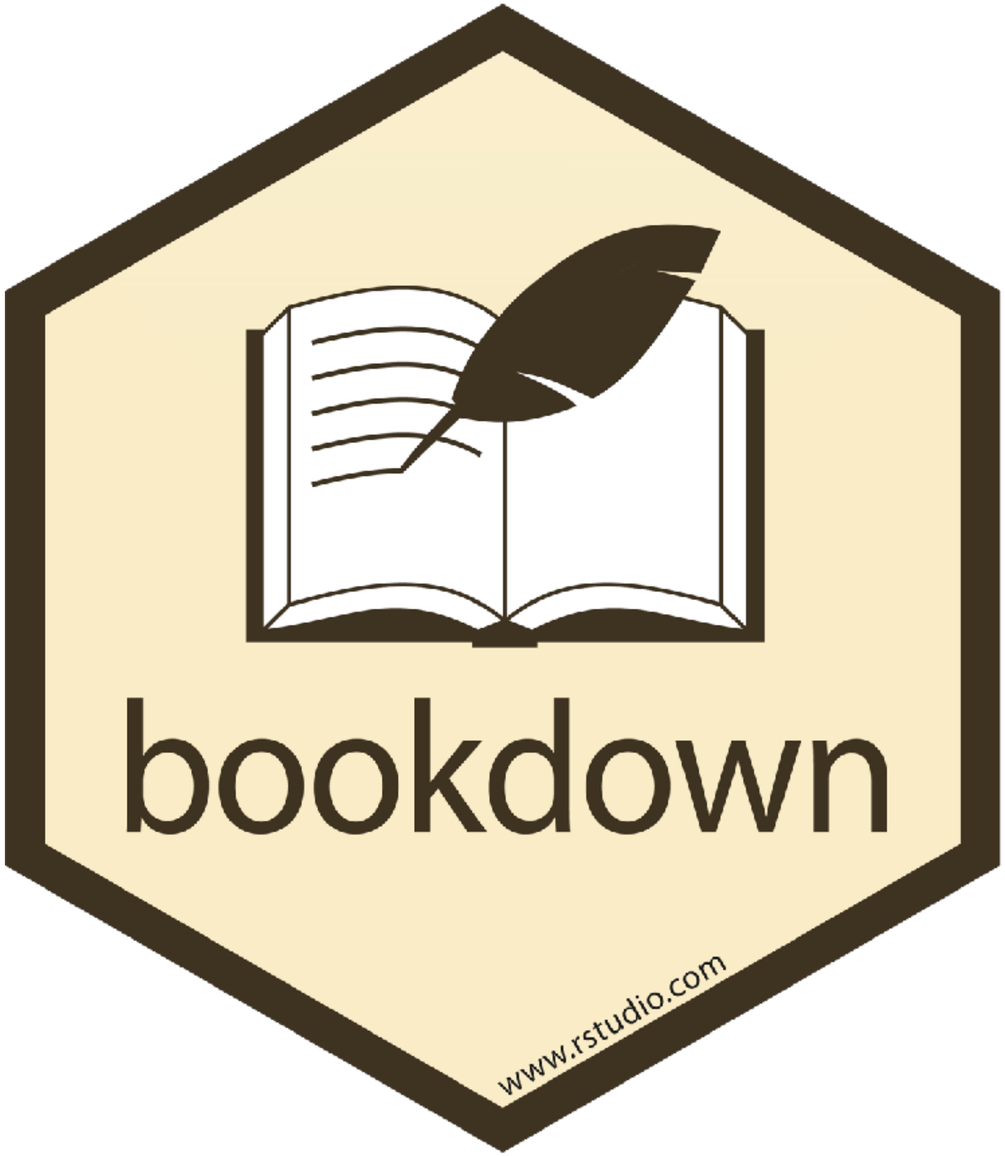
- Can be anywhere in your document... except a chunk
- Fenced with a single back tick `` `r contents` ``
- Uses current workspace

Your Turn 4

- Open **11-Communicate/chunk-basics.Rmd**
- Follow the numbered directions to add and modify code chunks and inline code.

05 : 00

Writing Reports with R Markdown


bookdown
www.rstudio.com

```
---
...
output:
  html_document:
    toc: true
    theme: "darkly"
---
```

```
---
...
output:
  bookdown::html_document2:
    toc: true
    theme: "darkly"
---
```

Case Study 2 Report

Your Turn 5

- Create a new R Markdown document.
- Title it "*Examing the Effect of Gender on Graduate Program Admission*" and add yourself as the author
- Make the date update automatically to today's date
- Change the output to **bookdown::html_document2**
- Save the file as **11-Communicate/case-study-2-report/case-study-2-report.Rmd**

Your Turn 6

Setup global options in the **setup** chunk

- Load the packages that we used in Case Study 2 + **knitr**, **hrbrthemes**, and **colorblindr**
- Set global chunk options inside **knitr::opts_chunk$set()**
  - Don't show code
  - Don't show **messages**, **warnings**, or **errors**
  - Save figure to **figures/**
  - Make all figures **8** inches wide with an aspect ratio of **0.618**
  - Align all pictures in the **center** and use **90%** of the output width

05 : 00

````
```{r setup, include = FALSE}
library(tidyverse)
library(broom)
library(rsample)
library(tidydscompanion)
library(knitr)
library(hrbrthemes)
library(colorblindr)

knitr::opts_chunk$set(
  echo = FALSE,
  message = FALSE,
  warning = FALSE,
  error = FALSE,
  fig.path = "figures/",
  fig.retina = 3,
  fig.width = 8,
  fig.asp = 0.618,
  fig.align = "center",
  out.width = "90%"
)
```
````

Your Turn 7

- Remove the boilerplate text from the document
- Add an introduction that describes the purpose of the report.

05 : 00

Citations and References

Citations are added to a **.bib** file

bib/references.bib

```
@book{gre,
  author = {{Educational Testing Service}},
  year = {2012},
  title = {The official guide to the {GRE} revised general test},
  edition = {2nd},
  address = {New York, NY},
  publisher = {McGraw-Hill},
  isbn = {9780071791236}
}
```

Citation key

```
@book{gre,
  author = {{Educational Testing Service}},
  year = {2012},
  title = {The official guide to the {GRE} revised general test},
  edition = {2nd},
  address = {New York, NY},
  publisher = {McGraw-Hill},
  isbn = {9780071791236}
}
```

syntax	renders as
`Blah blah [@gre].`	Blah blah (Educational Testing Service, 2012).
`ETS say blah [-@gre].`	ETS says blah (2012).
`@gre says blah.`	Educational Testing Service (2012) says blah.

More on citations here:
https://rmarkdown.rstudio.com/authoring_bibliographies_and_citations.html#citations

Example bib entries:
https://www.verbosus.com/bibtex-style-examples.html

```
---
...
bibliography: ["bib/references.bib"]
biblio-style: apalike2
csl: csl/apa.csl
link-citations: true
...
---
```

List of .bib files containing references

Bibliography style

Citation style guide

Citations include hyperlinks

Your Turn 8

- Update the YAML to include **bib/references.bib** as a bibliography
- Set **biblio-style** to "apalike2" and **csl** to "csl/apa.csl"
- Make citations have a hyper link

Where do the references appear?

Add a "References" heading.

Tables

```
kable(x, digits, col.names = NA, align, caption, format.args, ...)
```

- Data frame or matrix
- For numbers, how many digits to round to
- Column names for the table
- Column alignment
- Table caption
- Formatting options
- Additional arguments

```
kable(head(admission), digits = 1, align = "c",
      col.names = c("Decision", "Verbal", "Quant", "Writing", "GPA", "Gender"),
      caption = "A table of data", format.args = list(big.mark = ","),
      booktabs = TRUE)
```

Table 1: A table of data

Decision	Verbal	Quant	Writing	GPA	Gender
0	142	151	2.5	3.1	Female
0	148	140	2.5	3.4	Male
1	156	147	2.5	3.5	Male
0	154	152	4.0	3.3	Male
1	160	164	4.0	2.9	Male
1	154	164	3.5	3.1	Male

Look at those results in Table \@ref(tab:some-table)!

```{r some-table}
kable(head(admission), caption = "Here is a table")
```

It really is incredible!

Your Turn 9

- Add a "Data" section that describes the data and provides summary statistics
- Include a table of the median score on academic measures by gender

05 : 00

Equations

```
Let define the Log Loss in equation below.


$$
\text{LogLoss} = -\frac{1}{n}\sum_{i=1}^n[...]
$$
```

Equations

```
Let define the Log Loss in equation \@ref(eq:logloss).

\begin{equation}
  \text{LogLoss} = -\frac{1}{n}\sum_{i=1}^n[...]
  (\#eq:logloss)
\end{equation}
```

Cite your software!

```
```{r write-packages, include = FALSE}
if (!file.exists("bib/packages.bib")) file.create("bib/packages.bib")
if (!file.exists("bib/knit.bib")) file.create("bib/knit.bib")
suppressWarnings(
  knitr::write_bib(c("rmarkdown", "bookdown"), "bib/knit.bib")
)
suppressWarnings(
  knitr::write_bib(c(.packages()), "bib/packages.bib")
)
```
```

Then cite the package with @R-pkg.

Your Turn 10

- Add a methods section that describes what we did
- Include the equation for the Log Loss

$$\text{LogLoss} = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

- Cite the relevant R packages we used

05 : 00

Figures

Include code for figures in their own chunk.

````
```{r figure, fig.cap = "Here is a figure!"}
ggplot() +
  ...
```
````

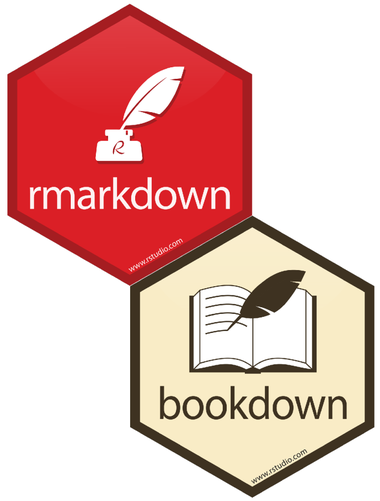
Add a **fig.cap** so the figure can be referenced with **\@ref(fig:figure)**

Your Turn 11

- Add a results section that include the analysis we did in Case Study 2
- Include each figure in its own chunk
- Reference the figures in the text

10 : 00

Communicating Results



ATLAS
Accessible Teaching, Learning, and Assessment Systems

wjakethompson.com wjakethompson@ku.edu @wjakethompson @wjakethompson