Reproducible Research using

Go here and log in (free):

https://rstudio.cloud/project/233945

https://rstudio.cloud/project/233945

@StatGarrett

https://rstudio.cloud/project/233945

Pop Quiz

What does **IMRAD** refer to? Poll your neighbors.

Standard structure
for a scientific paper.

https://rstudio.cloud/project/233945

What does **IMRAD** stand for? Poll your neighbors.

Introduction
What hypothesis was tested and why?

Methods
How was the study done?

Results
What answer was discovered?

And Discussion
What does the answer imply?

https://rstudio.cloud/project/233945

Which words do you associate with **math**?

hypotheses	**axioms**
messy	**logical**
best guess	**certain**
discover	**prove**

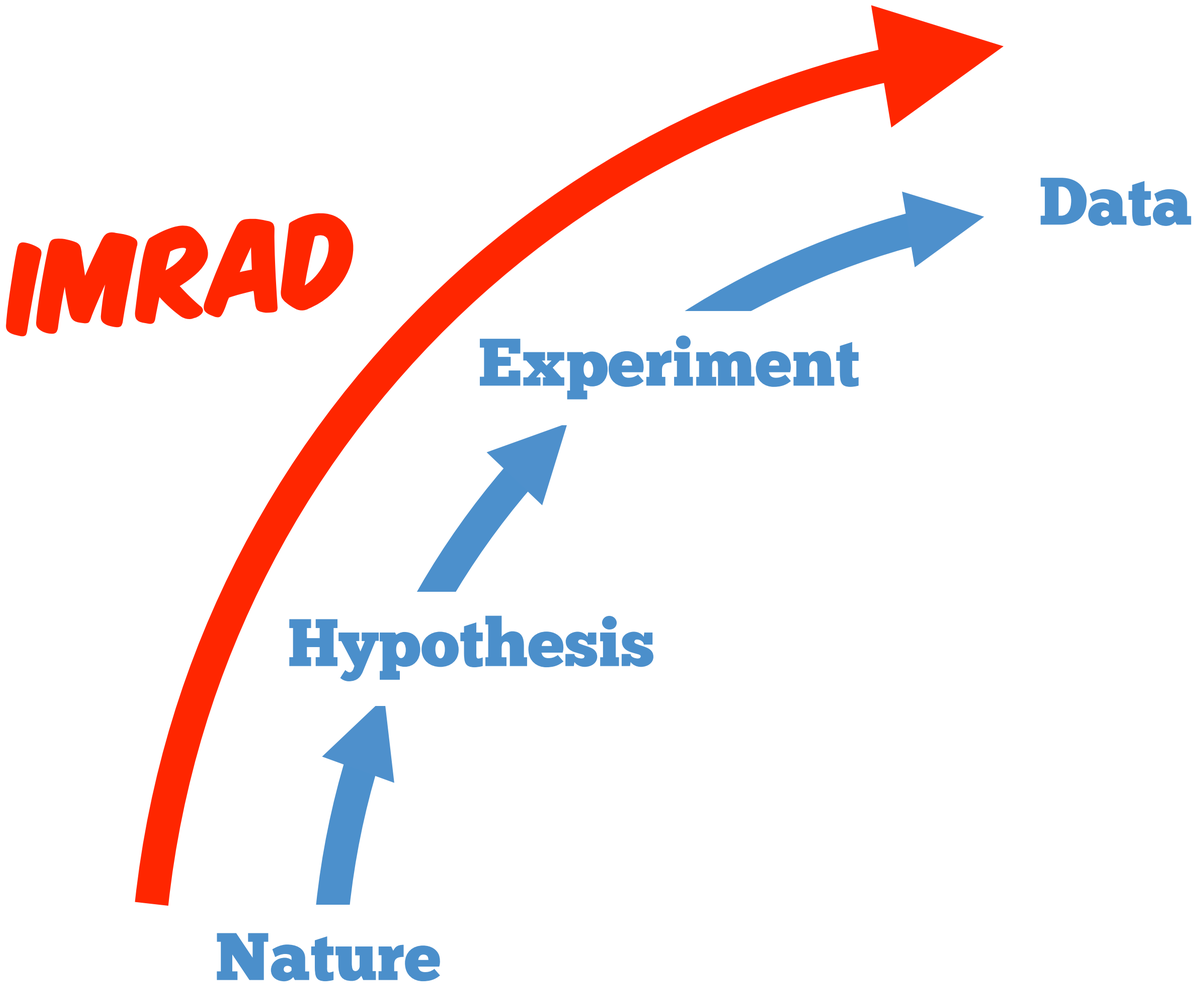
https://rstudio.cloud/project/233945

Which words do you associate with **Science**?

hypotheses	**axioms**
messy	**logical**
best guess	**certain**
discover	**prove**


CREATE MAPS,
NOT PROOFS
IMRAD

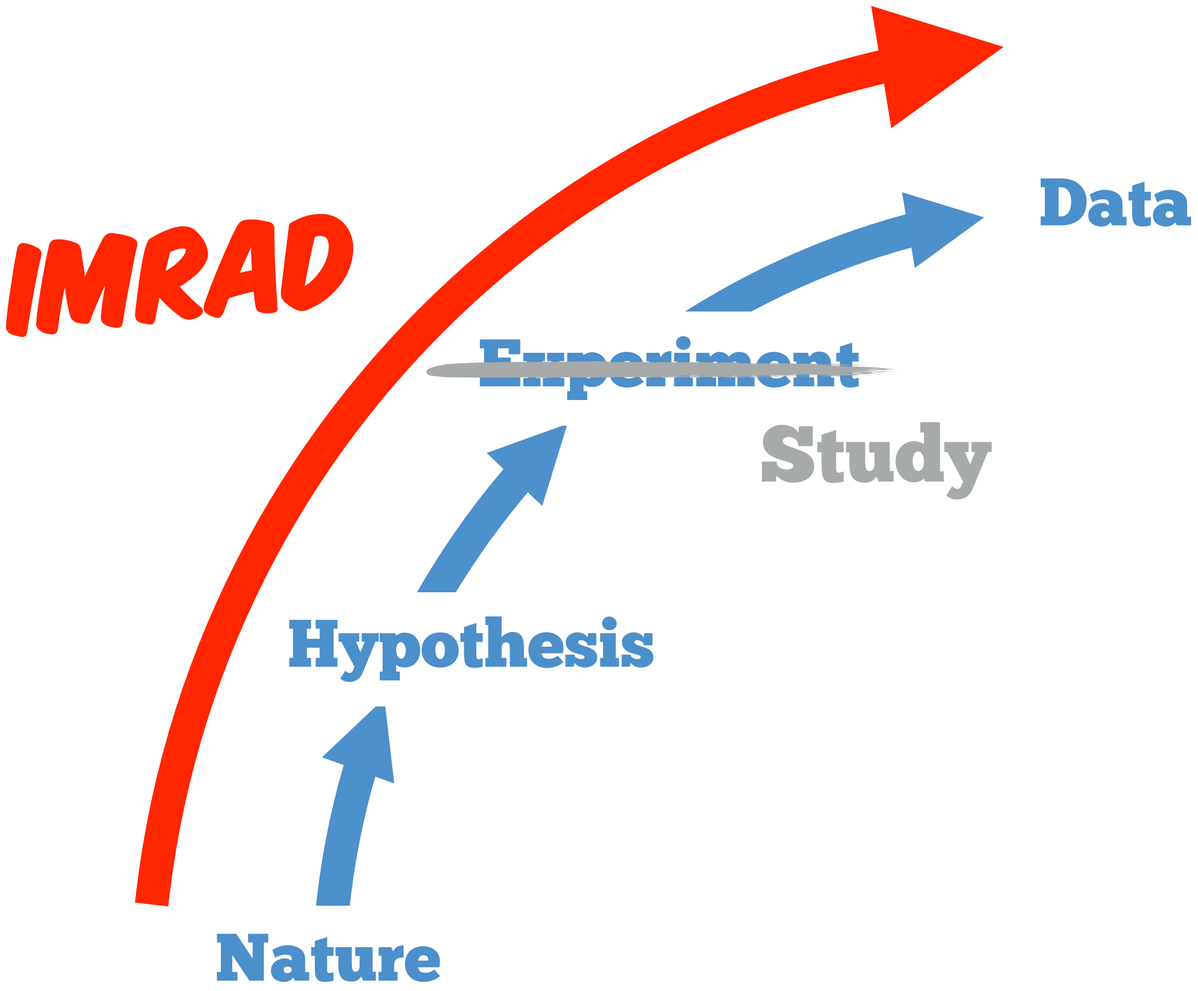
https://rstudio.cloud/project/233945



But where's the math?

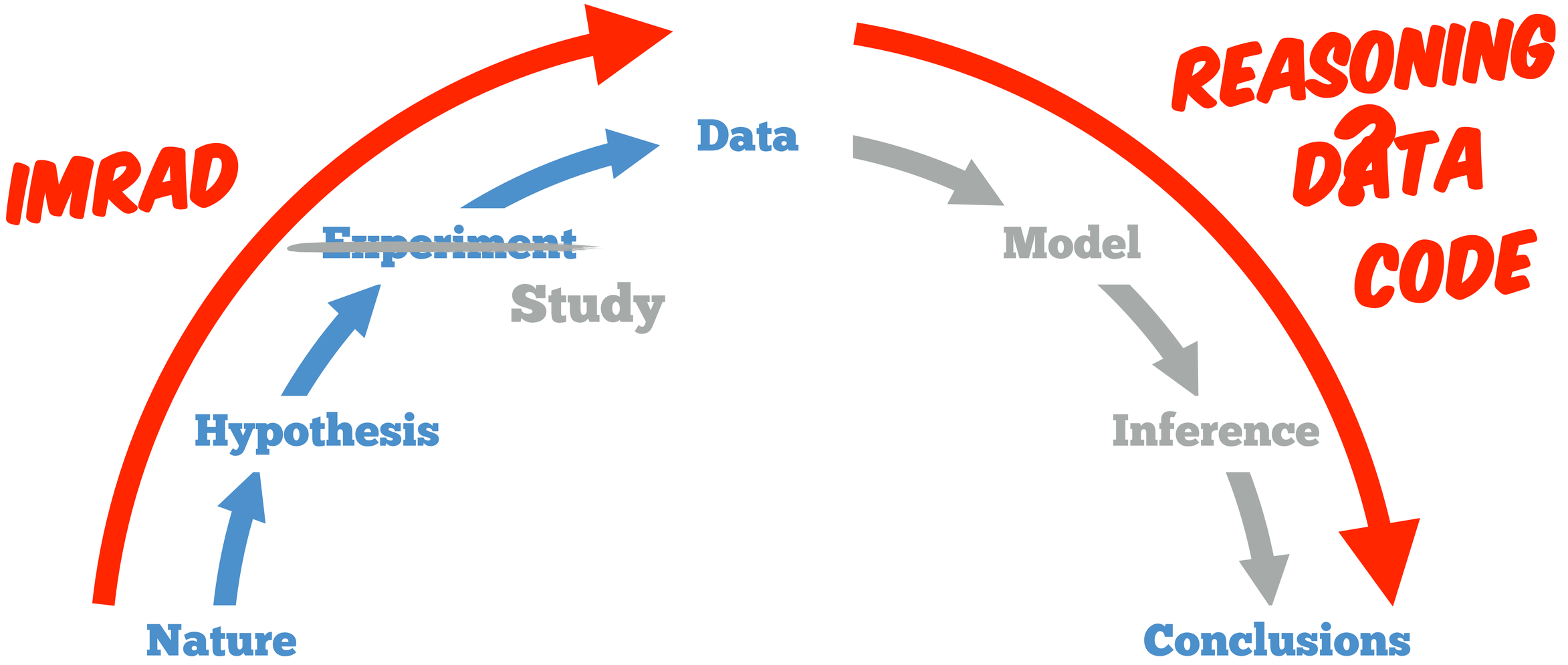
Adapted from Levy DG, Norris D. Methods and Guidelines for Integrity in Multivariate Analysis of Real World (Observational) Data. Unpublished Manuscript.

https://rstudio.cloud/project/233945



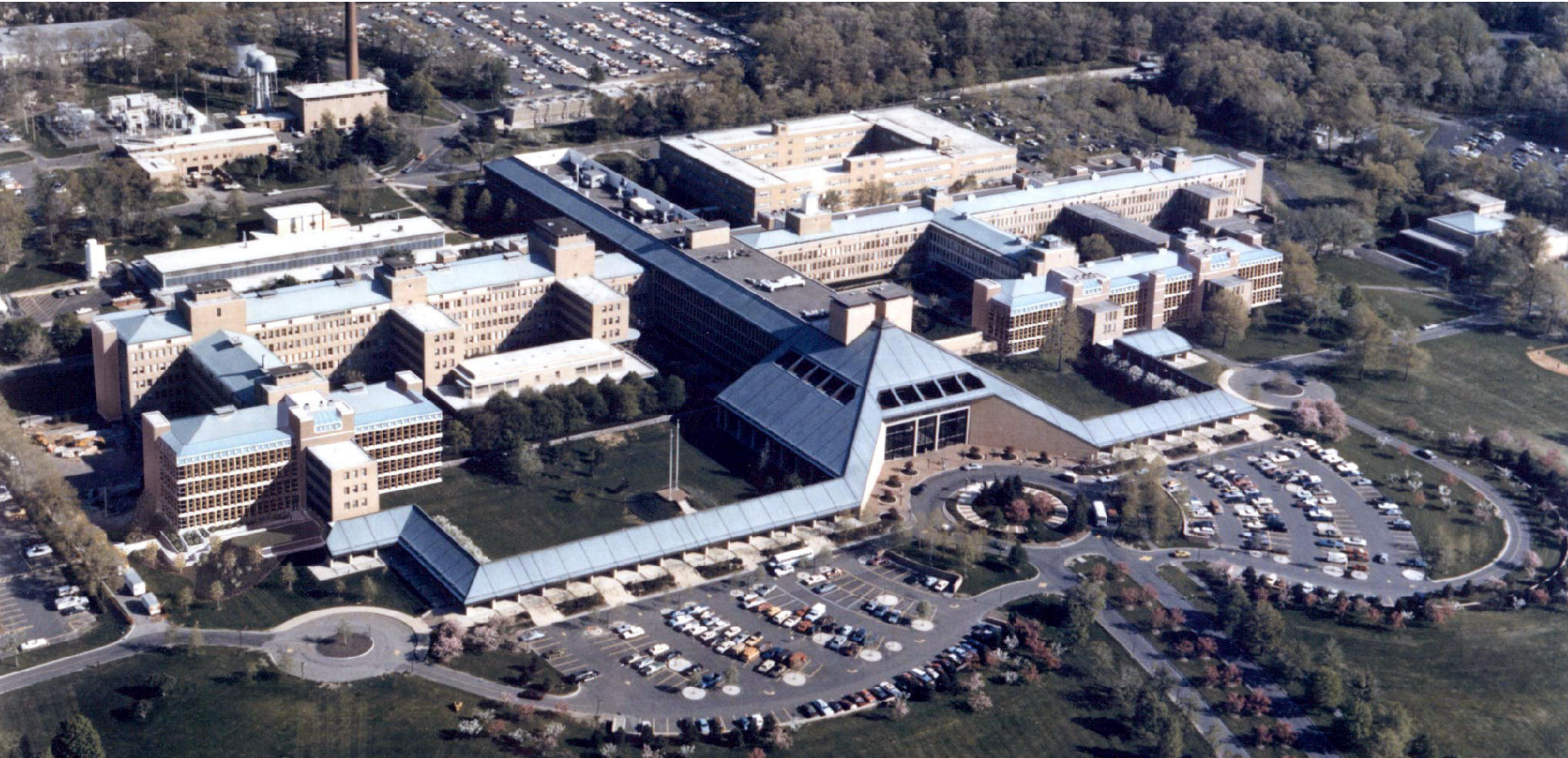
Adapted from Levy DG, Norris D. Methods and Guidelines for Integrity in Multivariate Analysis of Real World (Observational) Data. Unpublished Manuscript.

https://rstudio.cloud/project/233945



Adapted from Levy DG, Norris D. Methods and Guidelines for Integrity in Multivariate Analysis of Real World (Observational) Data. Unpublished Manuscript.

But which code language?

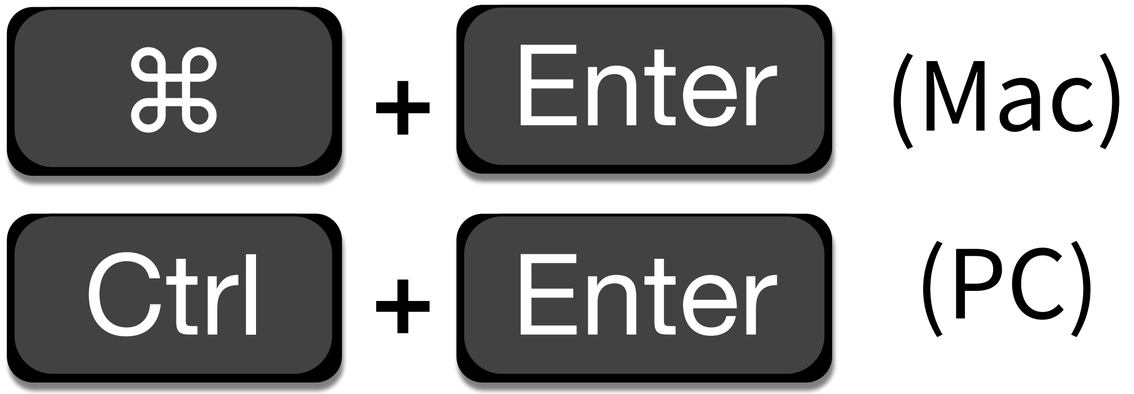


https://rstudio.cloud/project/233945

1. R and RStudio
2. R Markdown
3. Projects
4. Git and Github

R and RStudio

To run highlighted code:



Your Turn

Open **01-script.R**. Read the content and run it.

Determine: what does this code do?

Think 02:00

Pair 01:00

Your Turn

Open **01-script.R**. Read the content and run it.

Determine: what does this code do?

Share your conclusions with your neighbor.
Do they agree?

Think 

Pair 

Your Turn

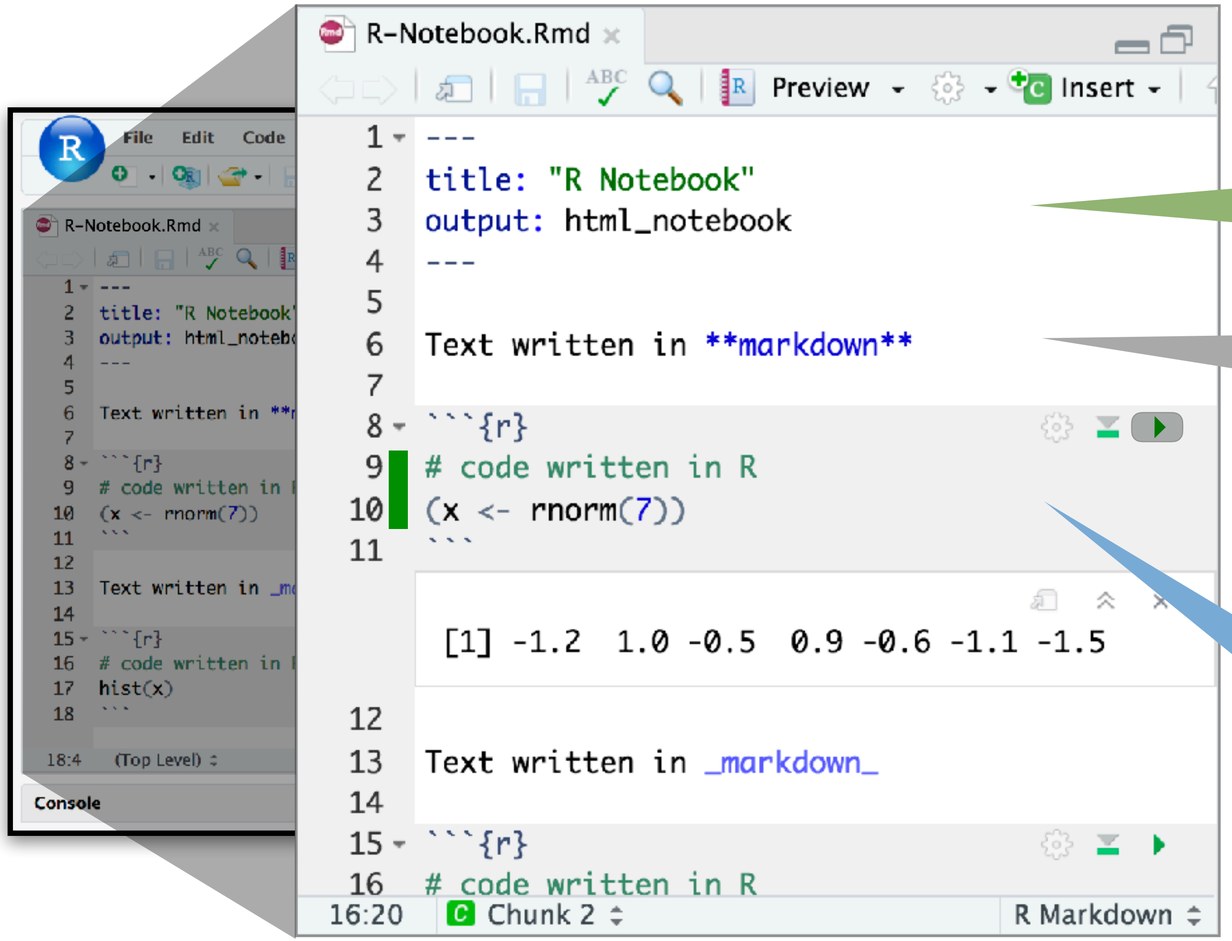
Open **02-narrative.Rmd**, which narrates the analysis in 01-script.R.

Were you right?

R Markdown

R Markdown

Plain text file with 3 types of content:



A YAML header surrounded by — — —

Text in markdown

Code chunks surrounded by ```

Markdown

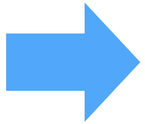
Web sites that use markdown

* **GitHub** www.github.com
* **StackOverflow** www.stackoverflow.com
* **Reddit** www.reddit.com
* many more

Headers

Use # to create headers.
Multiple #’s create lower level headers.

```
# Header 1
## Header 2
### Header 3
#### Header 4
##### Header 5
###### Header 6
```



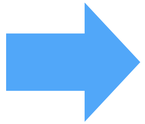
Header 1
Header 2
Header 3
Header 4
Header 5
Header 6

Text

Add two spaces at the end of a line to start a new line

Text is rendered as plain text. Surround text with _, **, or ` to format it.

```
Text
_italics_
**bold**
`code`
```



Text

italics

bold

`code`

Lists

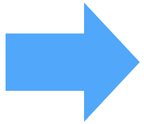
Use asterisks to make bullet points.
Use numbers to make numbered lists.

```
Bullets

* bullet 1
* bullet 2

Numbered list

1. item 1
2. item 2
```



Bullets

- bullet 1
- bullet 2

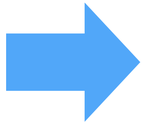
Numbered list

1. item 1
2. item 2

Hyperlinks

Use brackets to denote a link.
Place the URL in parentheses.

```
This is a
[link](www.git.com).
```



Images

Use a link preceded by an ! to insert an image.

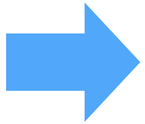
The link text should be a URL (if the image is hosted online), or a file path (if the image is saved as a file)

```
![](images/1.png)
The RStudio logo.
```

Equations

Write equations with latex math commands and surround them in $'s.

```
According to Einstein, $E=mc^{2}$
```



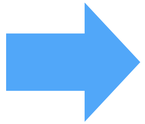
According to Einstein, $E = mc^2$

Equation blocks

Use two $’s to make centered equation blocks.

```
According to Einstein,

$$E=mc^{2}$$
```



According to Einstein,

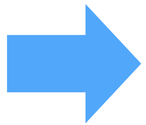
$$E = mc^2$$

Embed code

Inline code

Place code in a sentence with `` `r <code>` ``. R Markdown will replace the code with its results.

```
Today is
`r Sys.Date()`.
```



Today is 2015-04-16.

Code chunks

Insert a chunk of R code with

````
```{r}
# some code
```
````

When you render the report, R Markdown will run the code and include its results. R Markdown will also remove the ```` ```{r} ```` and ```` ``` ````.

Code chunks

Insert a chunk of R code with

```
```{r}
# some code
```
```

Your Turn

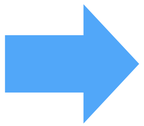
Replace every TODO in **02-narrative.Rmd**, with code from **01-script.R**. Then knit your document.

````
```{r}
# some code
```
````

chunk options

By default, R Markdown includes both the code and its results

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



Here's some code

```
dim(iris)
```

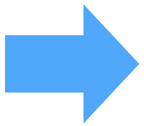
```
## [1] 150   5
```

echo

Add options in the brackets after r.

echo = FALSE hides the code.

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



Here's some code

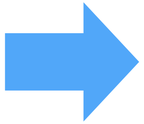
```
## [1] 150   5
```

Very useful for plots

eval

eval = FALSE prevents the code from being run. As a result, no results will be displayed with the code.

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



Here's some code

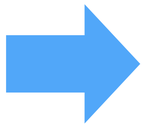
```
dim(iris)
```

include

include = FALSE runs the code, but prevents both the code and the results from appearing (e.g. to setup).

````
Here's some code

```{r include=FALSE}
dim(iris)
```
````

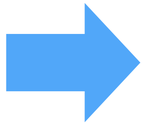


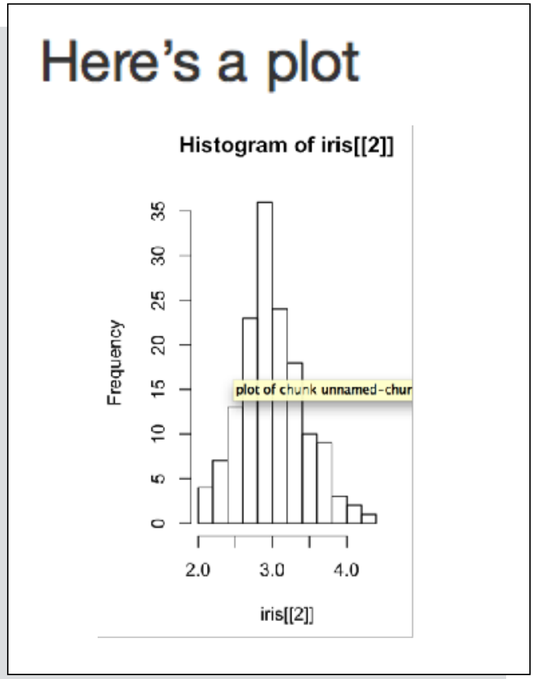
Here's some code

fig.height, fig.width

Specify the dimension of plots (in inches) with fig.width and fig.height. Separate multiple arguments with commas.

````
Here's a plot
```{r echo=FALSE, fig.width=3, fig.height=5}
hist(iris[[2]])
```
````



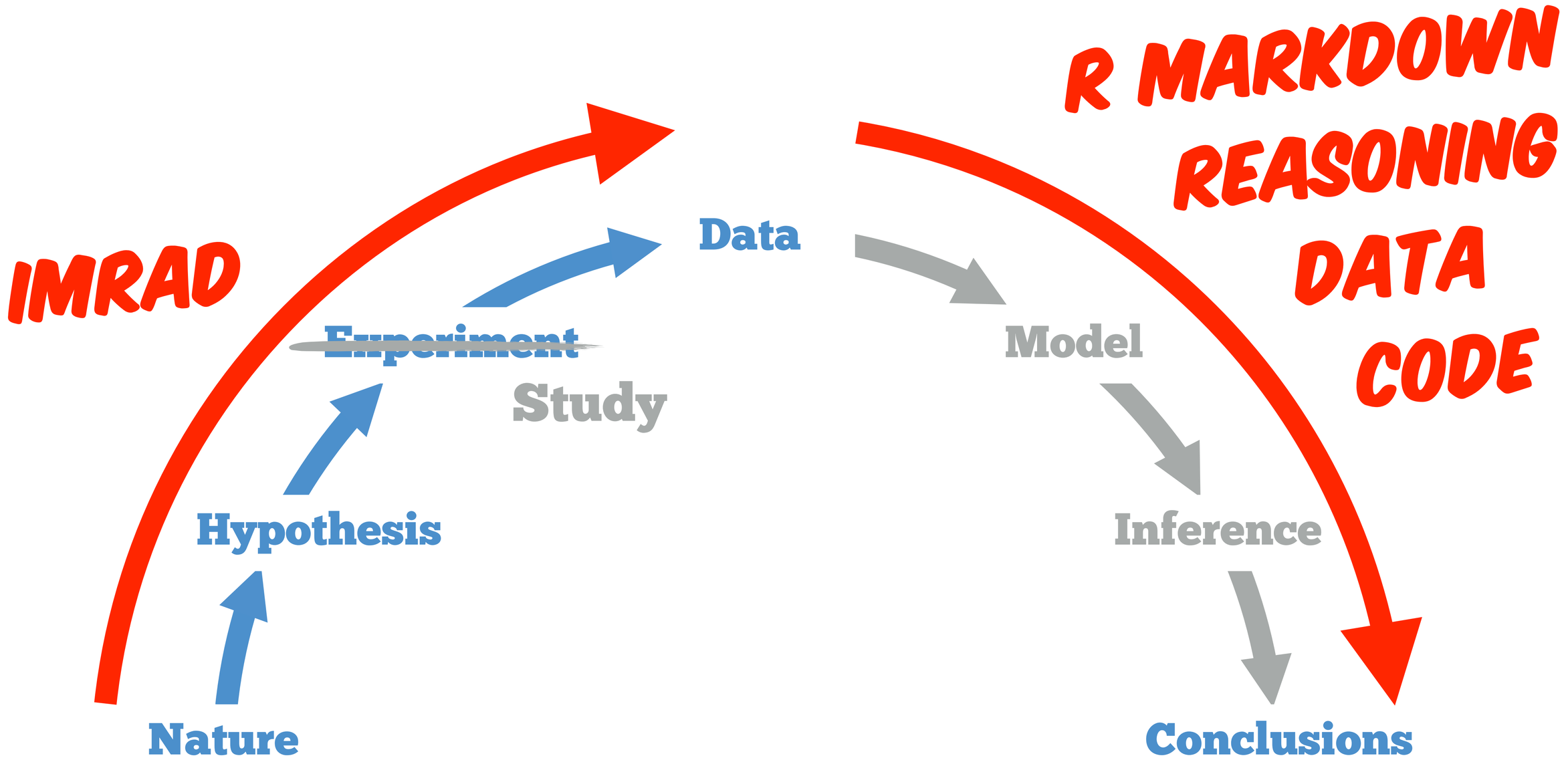


Your Turn

Add **include = FALSE** or **echo = FALSE** to each chunk as appropriate to suppress the code.

Set the height of the figure to 2 inches.

Then re-knit your document.



Adapted from Levy DG, Norris D. Methods and Guidelines for Integrity in Multivariate Analysis of Real World (Observational) Data. Unpublished Manuscript.

Parameters

Your Turn

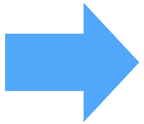
Open **03-report.Rmd**. Then use Knit with Parameters to render the document.

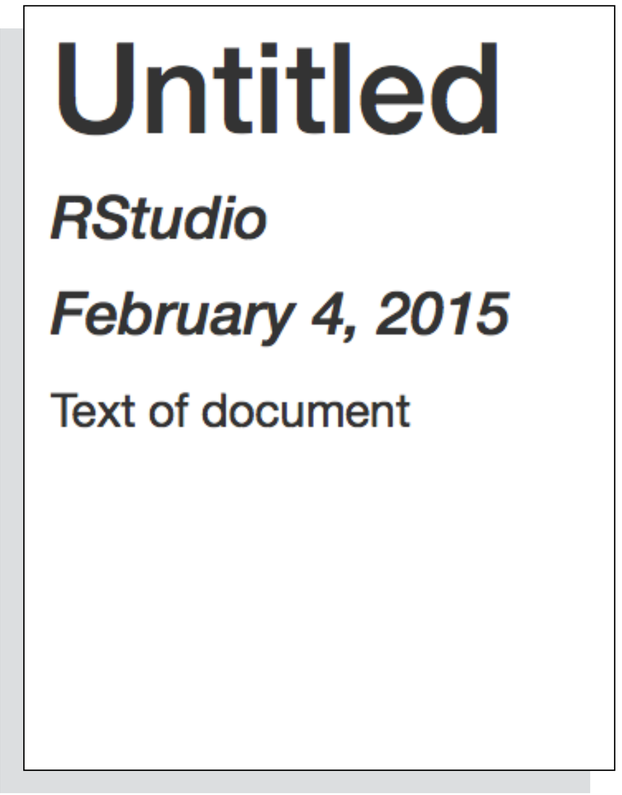
What happens if you use a different parameter?

YAML

A section of key:value pairs
separated by dashed lines ———

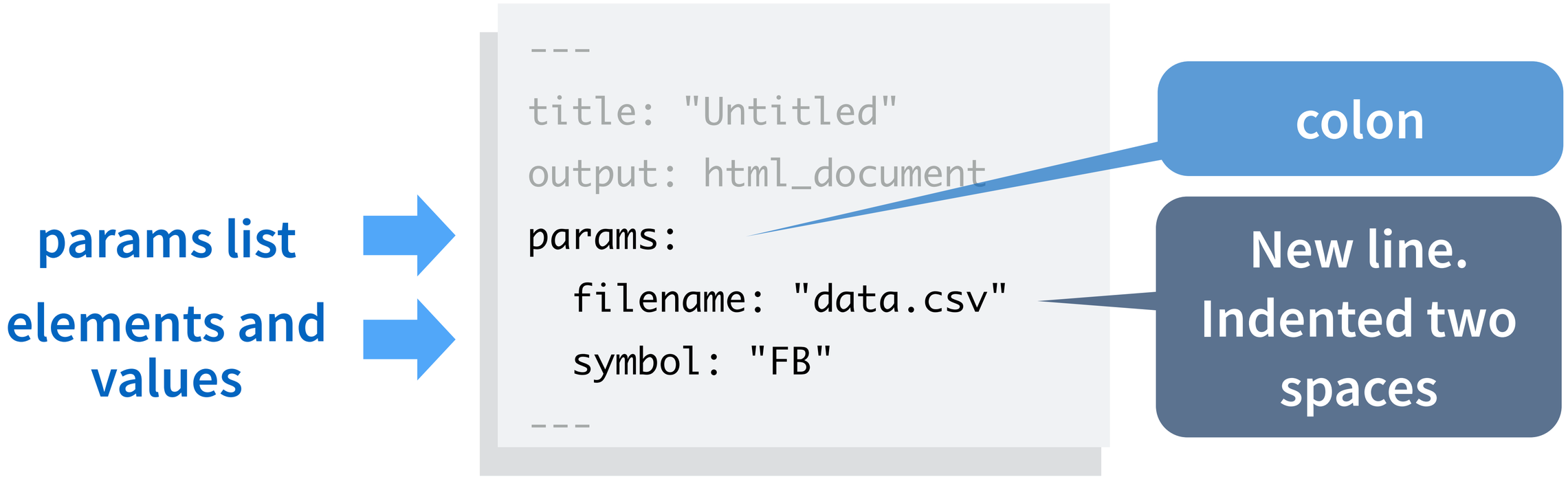
```
---
title: "Untitled"
author: "RStudio"
date: "February 4, 2015"
output: html_document
---
Text of document
```





Parameters

A list of values that you can call in R code chunks

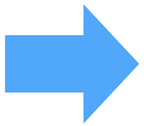


Using Parameters

Call parameter values as elements of the params list, **`params$num`**

````
---
params:
  num: 42
---
The value of the
parameter is
`r params$num`, e.g.

```{r}
params$num
```
````



The value of the parameter is 42, e.g.

```
params$num
```

```
## [1] 42
```

render()

Render at the command line with default YAML options

```
library(rmarkdown)
render("03-reports.Rmd")
```

Render at the command line, set parameters.

```
render("03-reports.Rmd",
       params = list(data = "london.csv"))
```

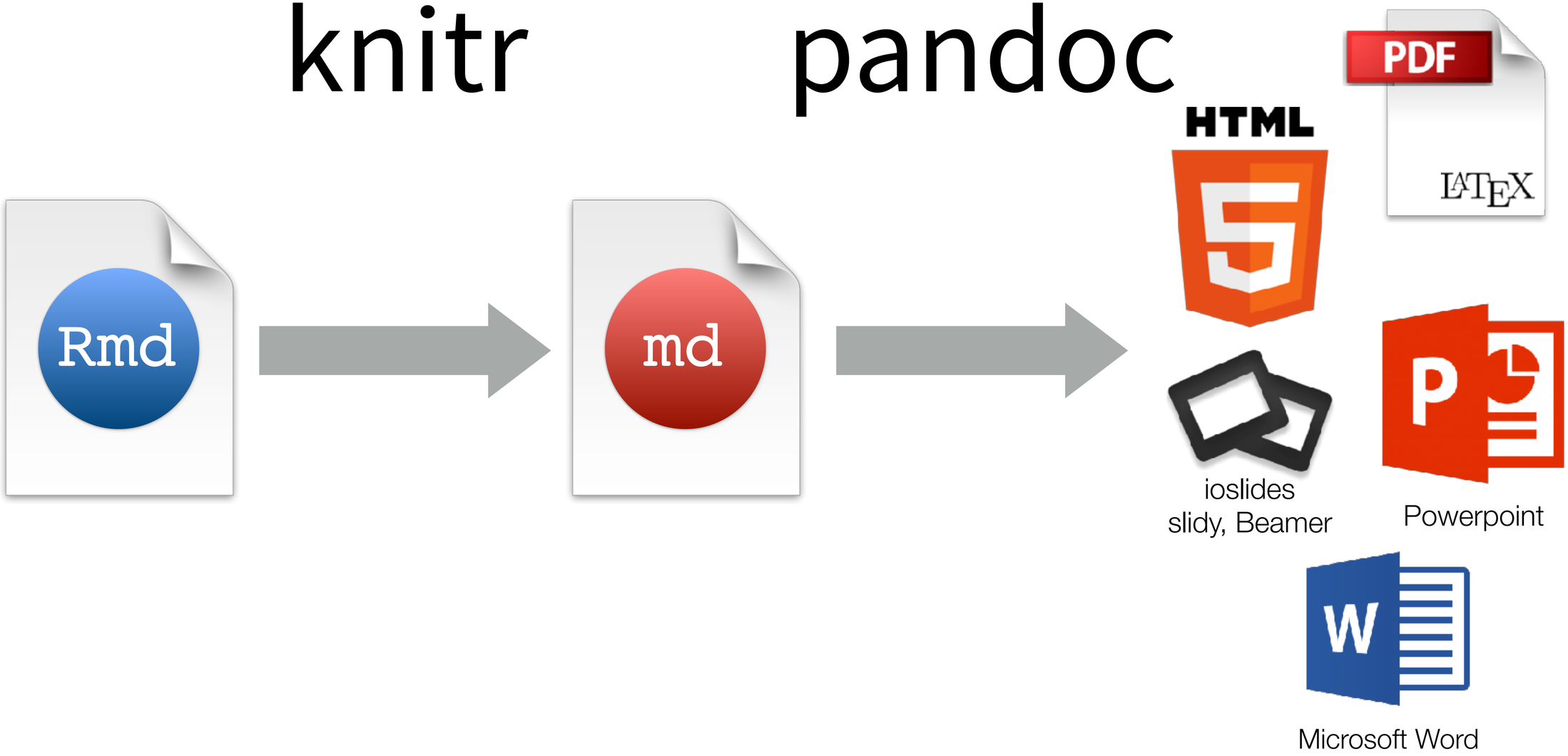
render() + for

```
datasets <- c("dublin.csv", "london.csv")

for (name in datasets) {
  render("03-reports.Rmd",
         params = list(data = name))
}
```

How it works

knitr pandoc



Logistics

1 Knitr runs the document in a fresh R session, which means you need to load the libraries that the document uses *in the document*

Logistics

1. Knitr runs the document in a fresh R session, which means you need to load the libraries that the document uses *in the document*

2. Objects made in one code chunk will be available to code in later code chunks.

KNITR IS MULTI-LINGUAL!

- [x] SAS
- [x] PYTHON
- [x] MORE

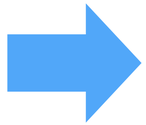
engine

python

To embed non R code, change the chunk label from r to the language to use.

````
Some python code,

```{python}
x = 'hello, python world!'
print(x)
print(x.split(' '))
```
````



Some python code:

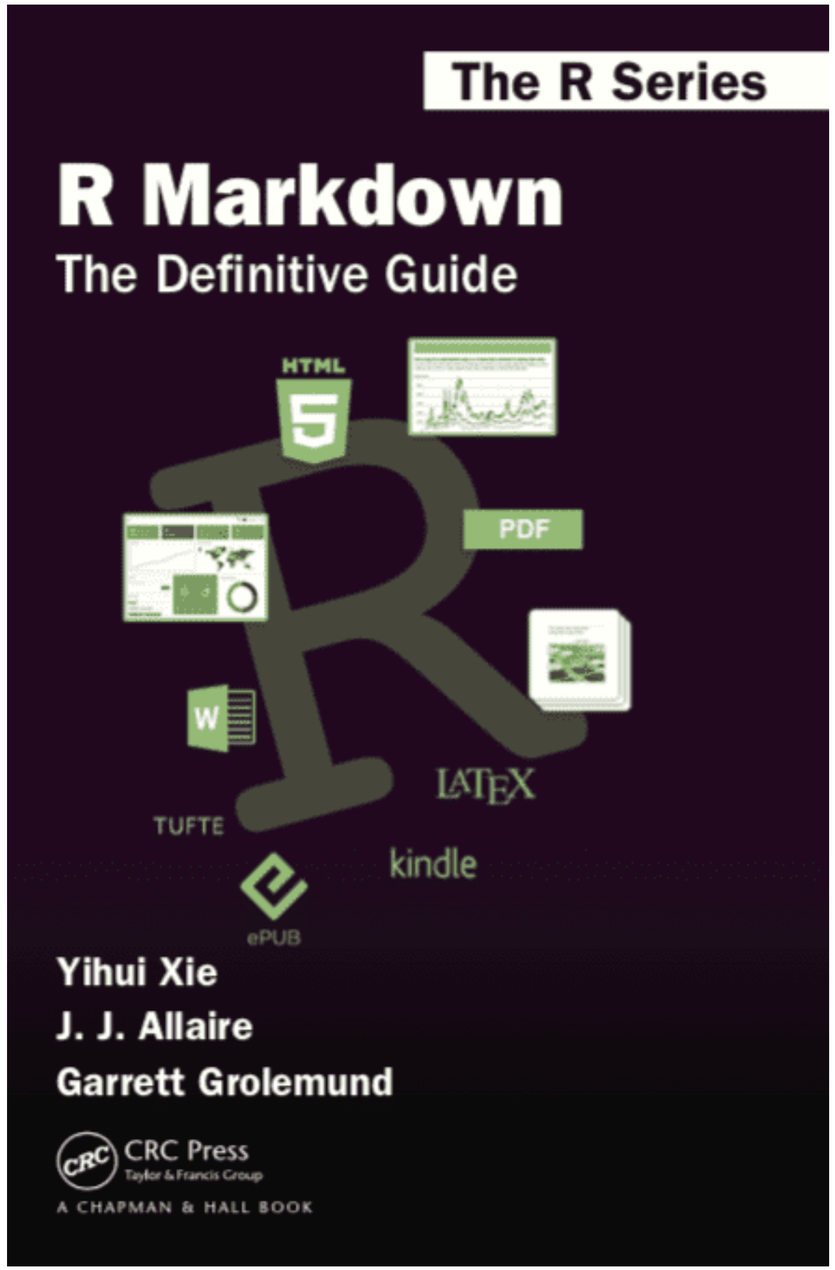
```
x = 'hello, python world!'
print(x)
print(x.split(' '))
```

```
## hello, python world!
## ['hello,', 'python', 'world!']
```

If the language's interpreter is not on your system's PATH variable, use **engine.path** to tell knitr where to find the interpreter

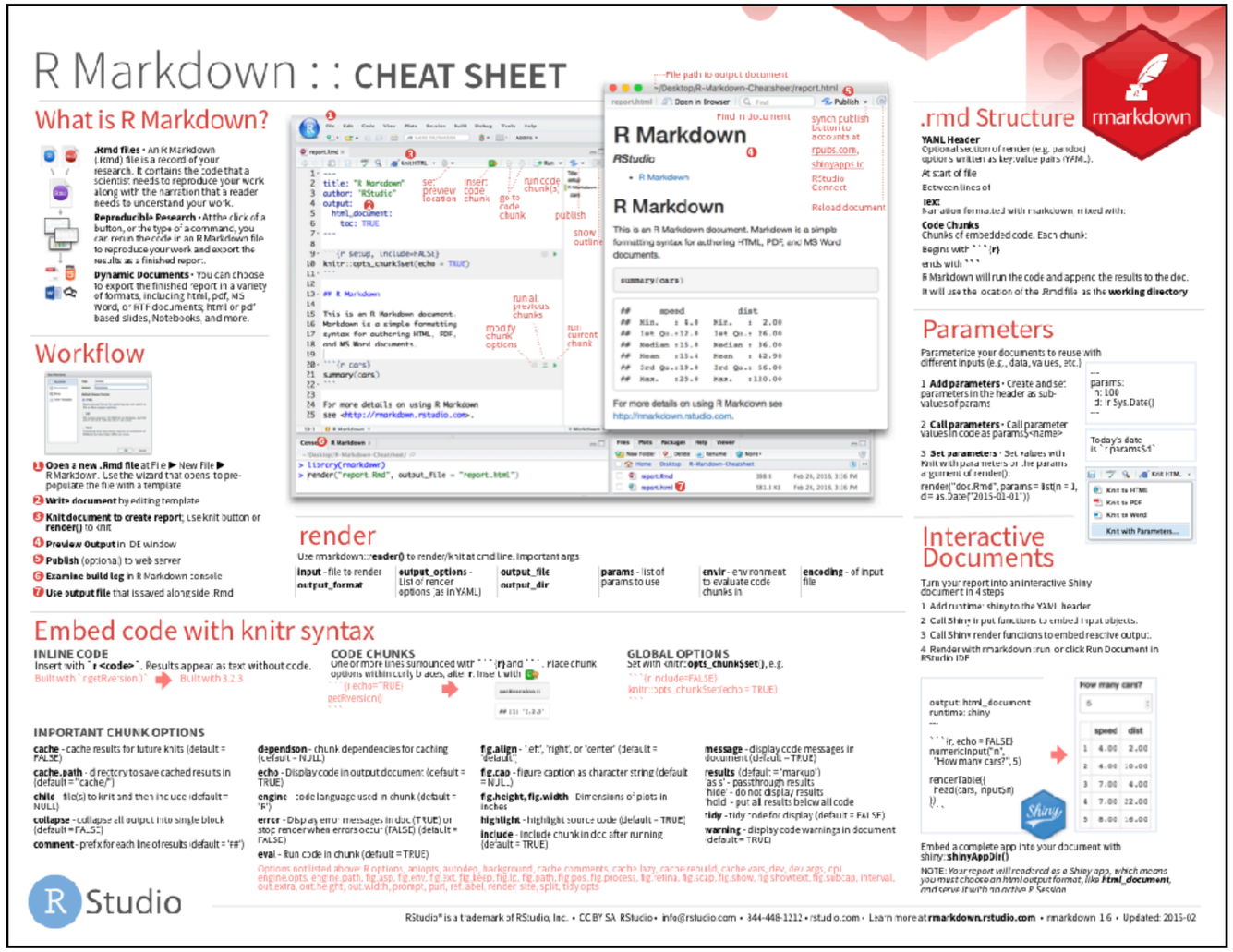
````
```{sas, engine.path = "C:/Program Files/SASHome/9.4/sas.exe"}
Some SAS code


```
````



bookdown.org/yihui/rmarkdown/

ONLINE, FREE

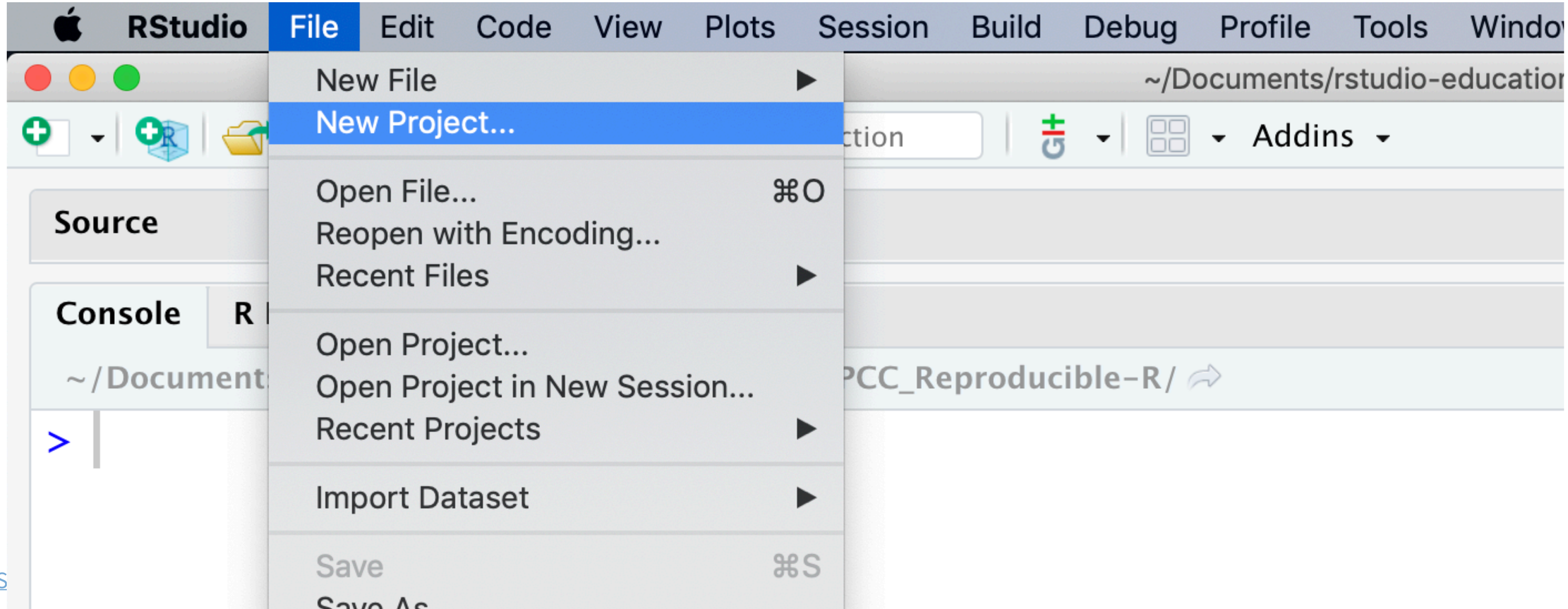


rstudio.com/cheatsheets

Projects

How to keep track of your files?

Put them in a directory and make it a **PROJECT**



Git and Github

What is version control?

Why should you use it?


git

History

1st Commit → 2nd Commit → 3rd Commit → 4th Commit → 5th Commit

```
"name","year","time","lat","long"
"Allison",1995,1995-06-03 00:00:00,17.4,-84.3
"Allison",1995,1995-06-03 06:00:00,18.3,-84.9
"Allison",1995,1995-06-03 12:00:00,19.3,-85.7
"Allison",1995,1995-06-03 18:00:00,20.6,-85.8
"Allison",1995,1995-06-04 00:00:00,22,-86
"Allison",1995,1995-06-04 06:00:00,23.3,-86.3
"Allison",1995,1995-06-04 12:00:00,24.7,-86.2
"Allison",1995,1995-06-04 18:00:00,26.2,-86.2
"Allison",1995,1995-06-05 00:00:00,27.6,-86.1
"Allison",1995,1995-06-05 06:00:00,28.5,-85.6
"Allison",1995,1995-06-05 12:00:00,29.6,-84.7
"Allison",1995,1995-06-05 18:00:00,30.7,-83.8
"Allison",1995,1995-06-06 00:00:00,31.8,-82.8
"Allison",1995,1995-06-06 06:00:00,32.7,-81.5
"Allison",1995,1995-06-06 12:00:00,33.6,-80
"Allison",1995,1995-06-06 18:00:00,34.5,-78.1
"Allison",1995,1995-06-07 00:00:00,35.6,-75.9
"Allison",1995,1995-06-07 06:00:00,37.1,-73.6
"Allison",1995,1995-06-07 12:00:00,38.5,-71
"Allison",1995,1995-06-07 18:00:00,39.8,-69.2
"Allison",1995,1995-06-08 00:00:00,41,-67.7
"Allison",1995,1995-06-08 06:00:00,42.4,-66
"Allison",1995,1995-06-08 12:00:00,43.8,-63.7
```

```
# 0-Clean.R

library(dplyr)
library(lubridate)

storms <- read.csv("storms.csv")

storms <- storms %>%
  mutate(time = ymd_h(paste(year, month, day,
    hour))) %>%
  select(name, year, time, lat, long,
pressure, wind, type)

write.csv(storms, file = "storms.csv",
  row.names = FALSE)
```

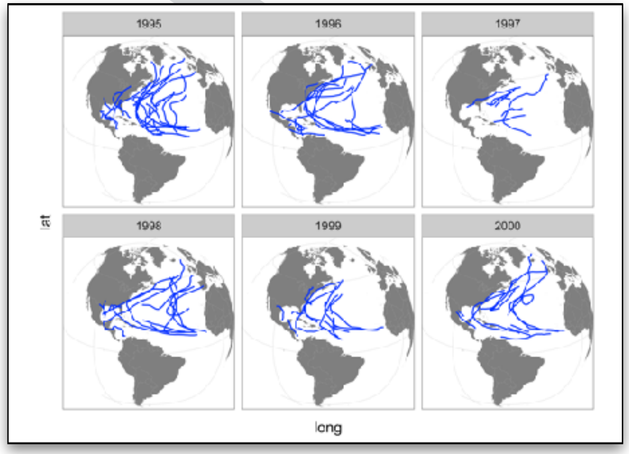
```
# 1-Plot.R

library(ggplot2)
library(dplyr)

map <- map_data("world") %>%
  filter(region != "USSR")

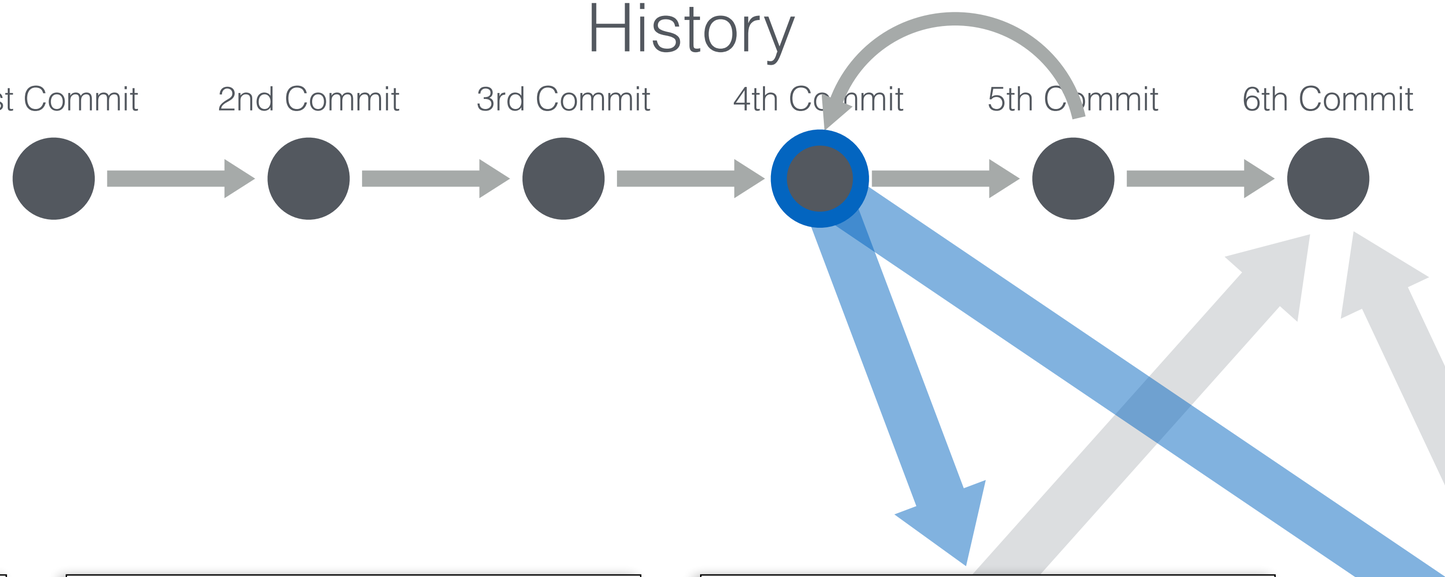
ggplot(storms, aes(x = long, y = lat)) +
  geom_polygon(aes(group = group),
    fill = "grey50", data = map) +
  geom_path(aes(group = name),
    color = "blue") +
  facet_wrap(~ year) +
  theme_bw() +
  coord_map(projection = "ortho",
    orientation = c(21, -60, 0))

ggsave("storms.png", width = 7, height = 5)
```



Project

History




```
"name","year","time","lat","long"
"Allison",1995,1995-06-03 00:00:00,17.4,-84.3
"Allison",1995,1995-06-03 06:00:00,18.3,-84.9
"Allison",1995,1995-06-03 12:00:00,19.3,-85.7
"Allison",1995,1995-06-03 18:00:00,20.6,-85.8
"Allison",1995,1995-06-04 00:00:00,22,-86
"Allison",1995,1995-06-04 06:00:00,23.3,-86.3
"Allison",1995,1995-06-04 12:00:00,24.7,-86.2
"Allison",1995,1995-06-04 18:00:00,26.2,-86.2
"Allison",1995,1995-06-05 00:00:00,27.6,-86.1
"Allison",1995,1995-06-05 06:00:00,28.5,-85.6
"Allison",1995,1995-06-05 12:00:00,29.6,-84.7
"Allison",1995,1995-06-05 18:00:00,30.7,-83.8
"Allison",1995,1995-06-06 00:00:00,31.8,-82.8
"Allison",1995,1995-06-06 06:00:00,32.7,-81.5
```

```
# 0-Clean.R

library(dplyr)
library(lubridate)

storms <- read.csv("storms.csv")

storms <- storms %>%
  mutate(time = ymd_h(paste(year, month, day,
    hour))) %>%
  select(name, year, time, lat, long,
pressure, wind, type)

write.csv(storms, file = "storms.csv",
  row.names = FALSE)
```

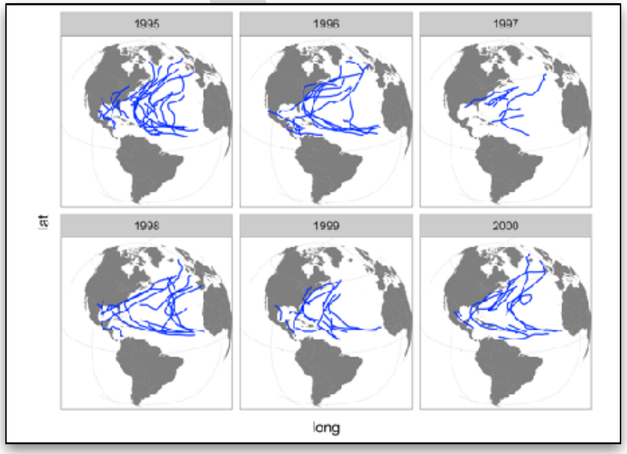
```
# 1-Plot.R

library(ggplot2)
library(dplyr)

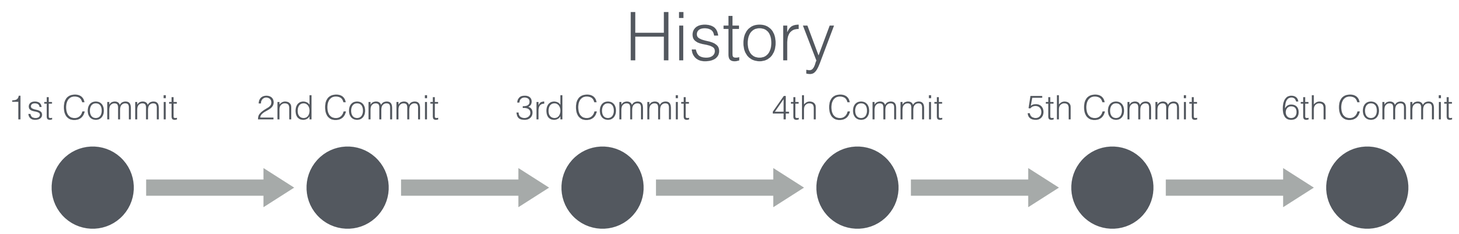
map <- map_data("world") %>%
  filter(region != "USSR")

ggplot(storms, aes(x = long, y = lat)) +
  geom_polygon(aes(group = group),
    fill = "grey50", data = map) +
  geom_path(aes(group = name),
    color = "blue") +
  facet_wrap(~ year) +
  theme_bw() +
  coord_map(projection = "ortho",
    orientation = c(21, -60, 0))

ggsave("storms.png", width = 7, height = 5)
```

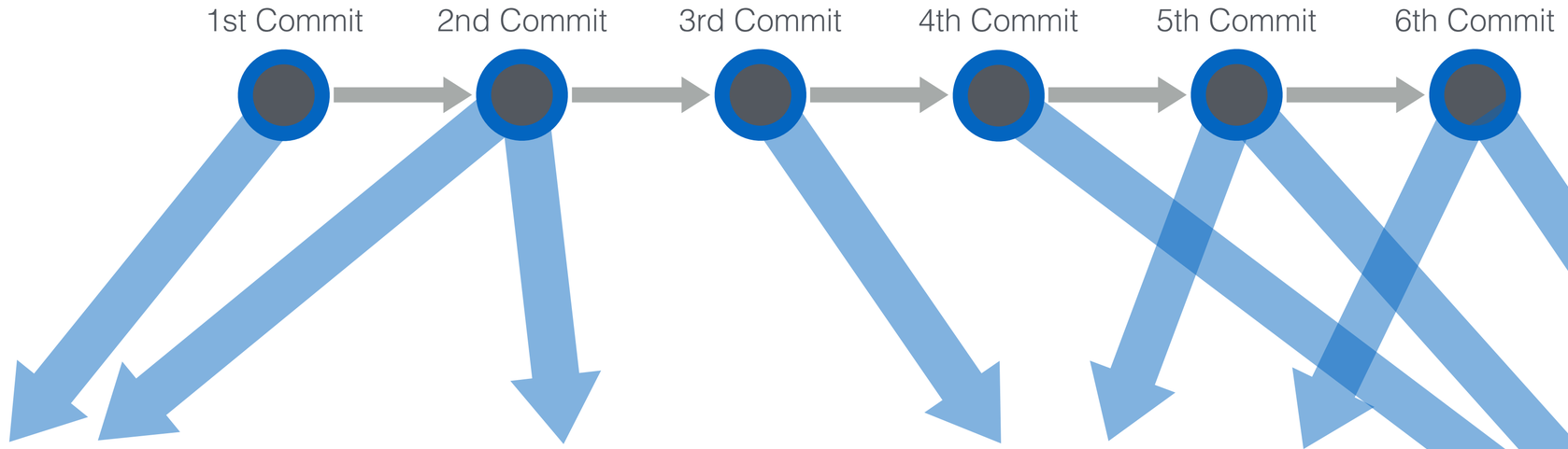


Project


History
1st Commit
2nd Commit
3rd Commit
4th Commit
5th Commit
6th Commit

.git

History




```
"name","year","time","lat","long"
"Allison",1995,1995-06-03 00:00:00,17.4,-84.3
"Allison",1995,1995-06-03 06:00:00,18.3,-84.9
"Allison",1995,1995-06-03 12:00:00,19.3,-85.7
"Allison",1995,1995-06-03 18:00:00,20.6,-85.8
"Allison",1995,1995-06-04 00:00:00,22,-86
"Allison",1995,1995-06-04 06:00:00,23.3,-86.3
"Allison",1995,1995-06-04 12:00:00,24.7,-86.2
"Allison",1995,1995-06-04 18:00:00,26.2,-86.2
"Allison",1995,1995-06-05 00:00:00,27.6,-86.1
"Allison",1995,1995-06-05 06:00:00,28.5,-85.6
"Allison",1995,1995-06-05 12:00:00,29.6,-84.7
"Allison",1995,1995-06-05 18:00:00,30.7,-83.8
"Allison",1995,1995-06-06 00:00:00,31.8,-82.8
"Allison",1995,1995-06-06 06:00:00,32.7,-81.5
"Allison",1995,1995-06-06 12:00:00,33.6,-80
"Allison",1995,1995-06-06 18:00:00,34.5,-78.1
"Allison",1995,1995-06-07 00:00:00,35.6,-75.9
"Allison",1995,1995-06-07 06:00:00,37.1,-73.6
"Allison",1995,1995-06-07 12:00:00,38.5,-71
"Allison",1995,1995-06-07 18:00:00,39.8,-69.2
"Allison",1995,1995-06-08 00:00:00,41,-67.7
"Allison",1995,1995-06-08 06:00:00,42.4,-66
"Allison",1995,1995-06-08 12:00:00,43.8,-63.7
```

```r
# 0-Clean.R

library(dplyr)
library(lubridate)

storms <- read.csv("storms.csv")

storms <- storms %>%
  mutate(time = ymd_h(paste(year, month, day,
    hour))) %>%
  select(name, year, time, lat, long,
pressure, wind, type)

write.csv(storms, file = "storms.csv",
  row.names = FALSE)
```

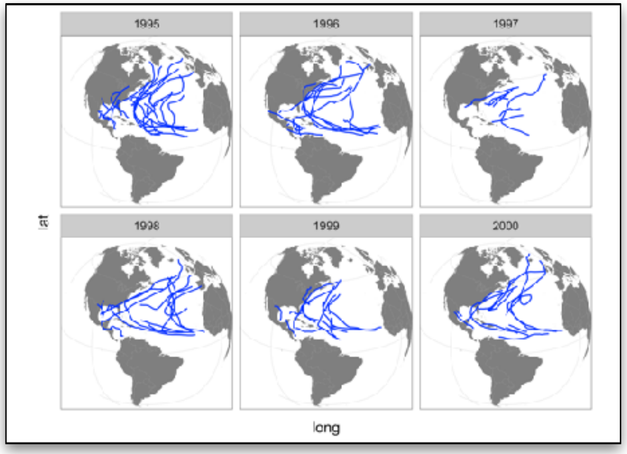
```r
# 1-Plot.R

library(ggplot2)
library(dplyr)

map <- map_data("world") %>%
  filter(region != "USSR")

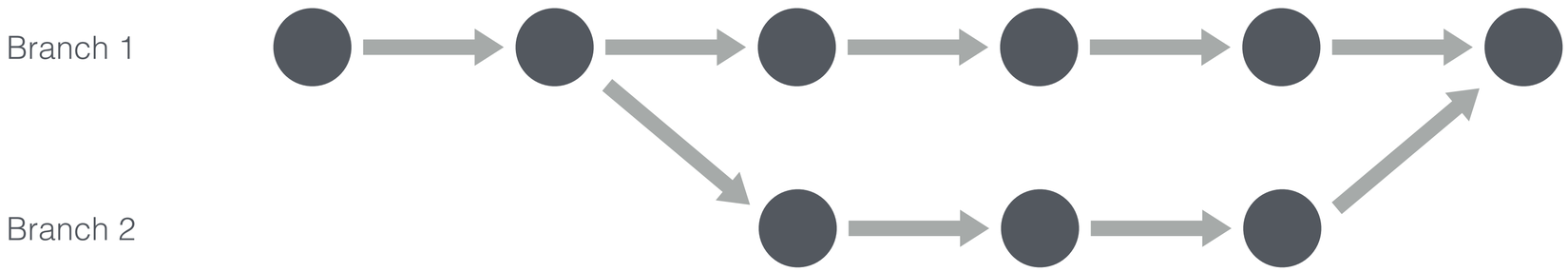
ggplot(storms, aes(x = long, y = lat)) +
  geom_polygon(aes(group = group),
    fill = "grey50", data = map) +
  geom_path(aes(group = name),
    color = "blue") +
  facet_wrap(~ year) +
  theme_bw() +
  coord_map(projection = "ortho",
    orientation = c(21, -60, 0))

ggsave("storms.png", width = 7, height = 5)
```



Project

History



"Official" Version

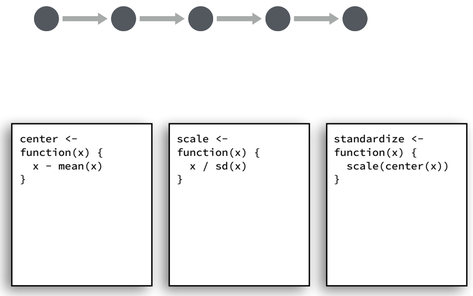
implied by commit history

Real Life Version

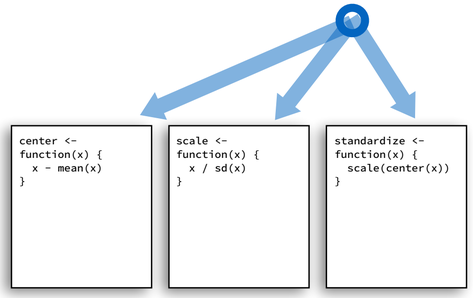
uncommitted changes
in blue

GitHub

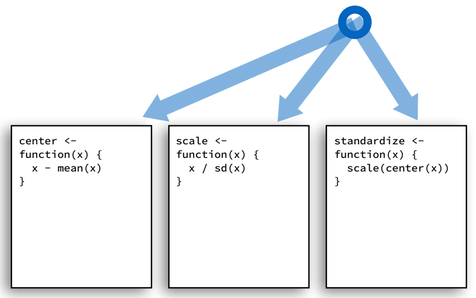
www.github.com



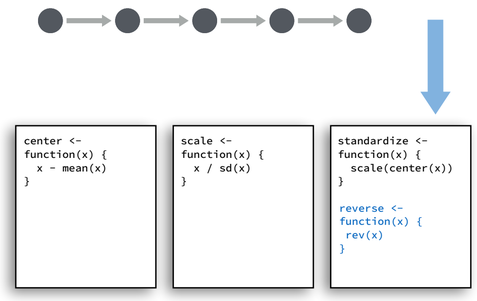
You



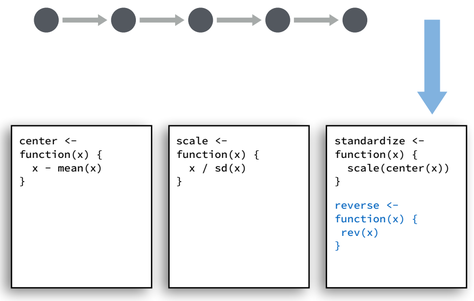
Collaborator 1



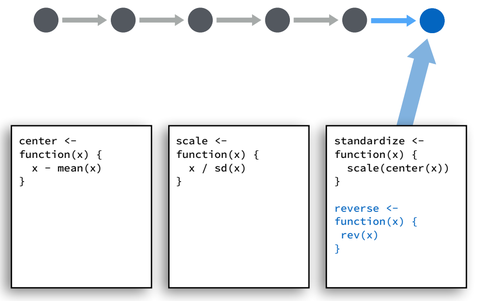
Collaborator 2



You



Collaborator 1



Collaborator 2

3. Github Version

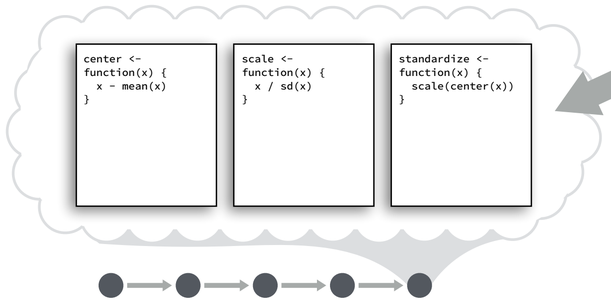
available to public

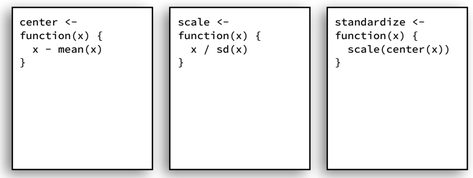
2."Official" Version

implied by commit history

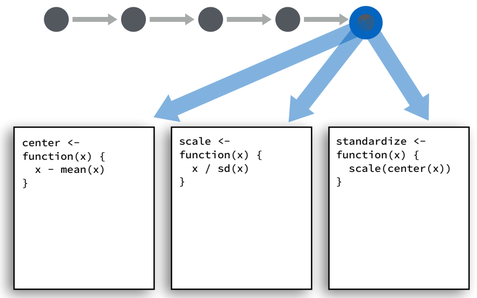
1.Real Life Version

in your working directory





You



3. Github Version

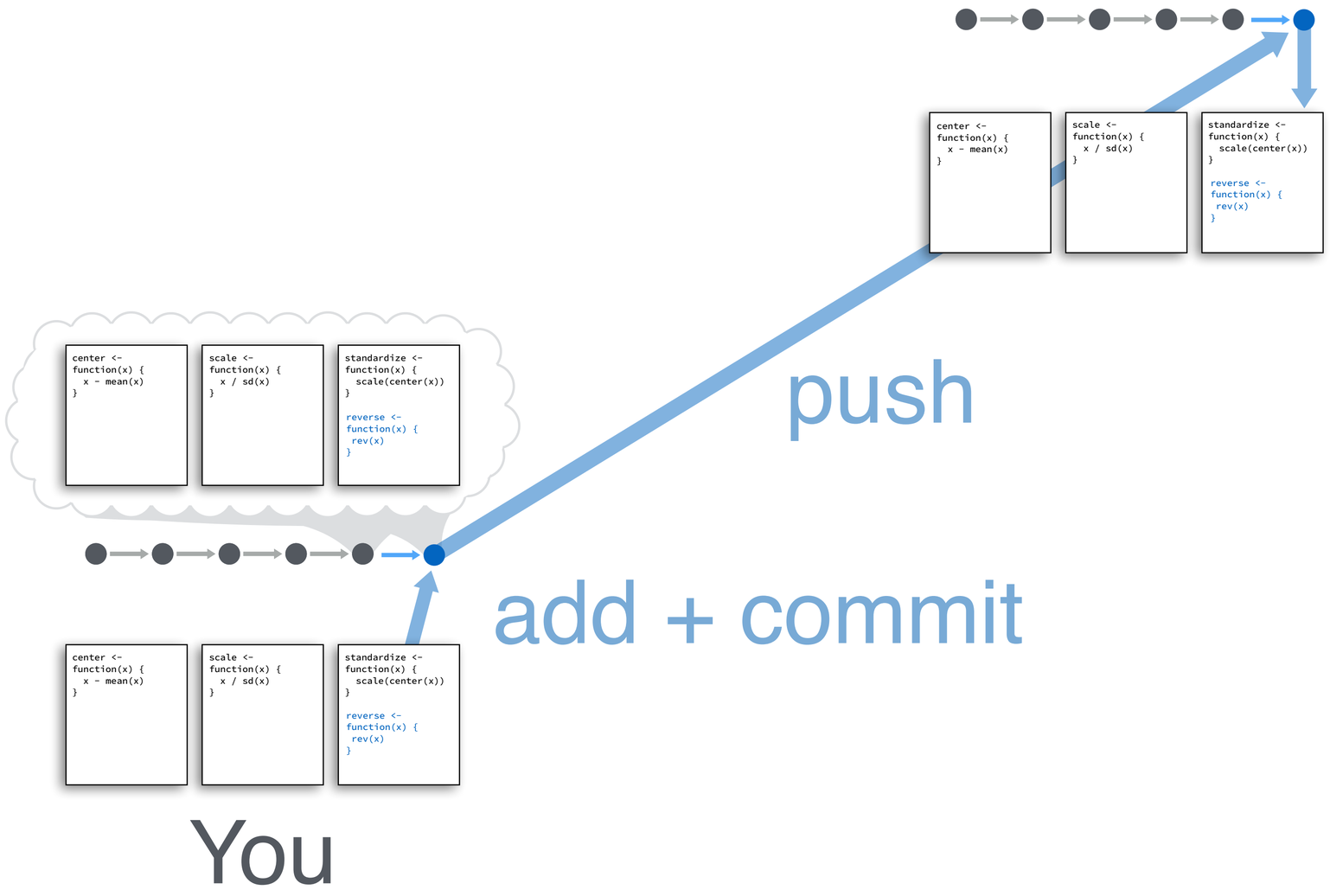
available to public

2."Official" Version

implied by commit history

1.Real Life Version

in your working directory



3. Github Version

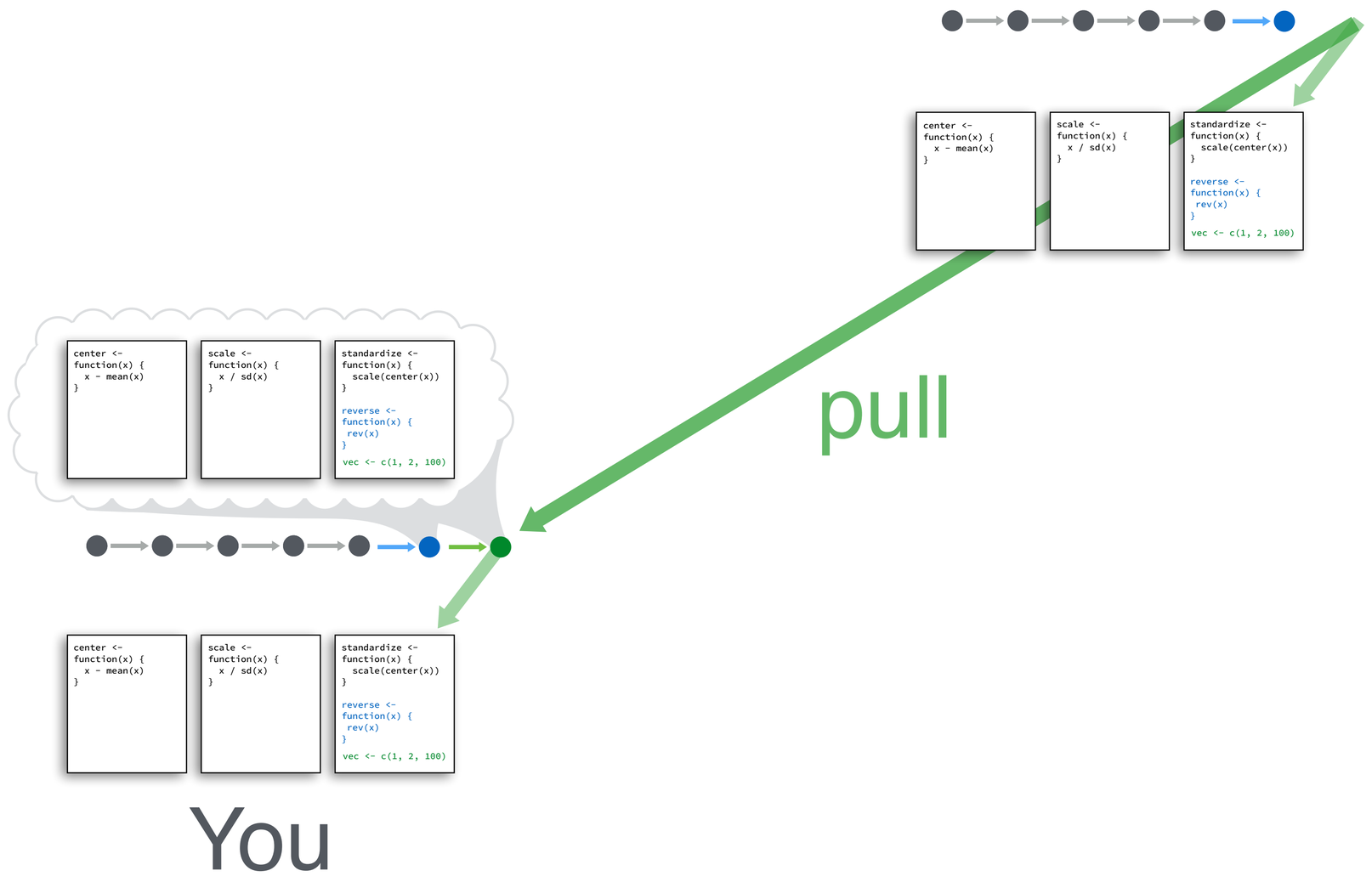
available to public

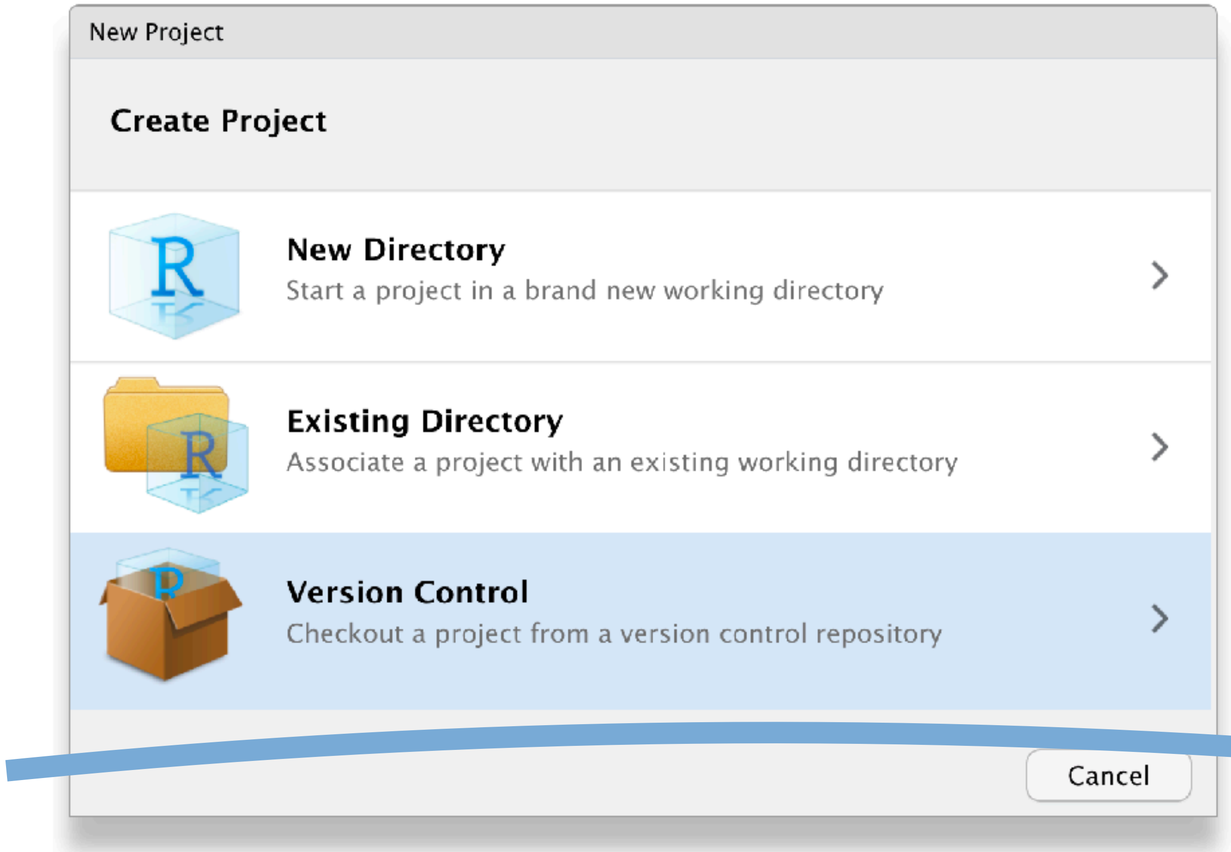
2."Official" Version

implied by commit history

1.Real Life Version

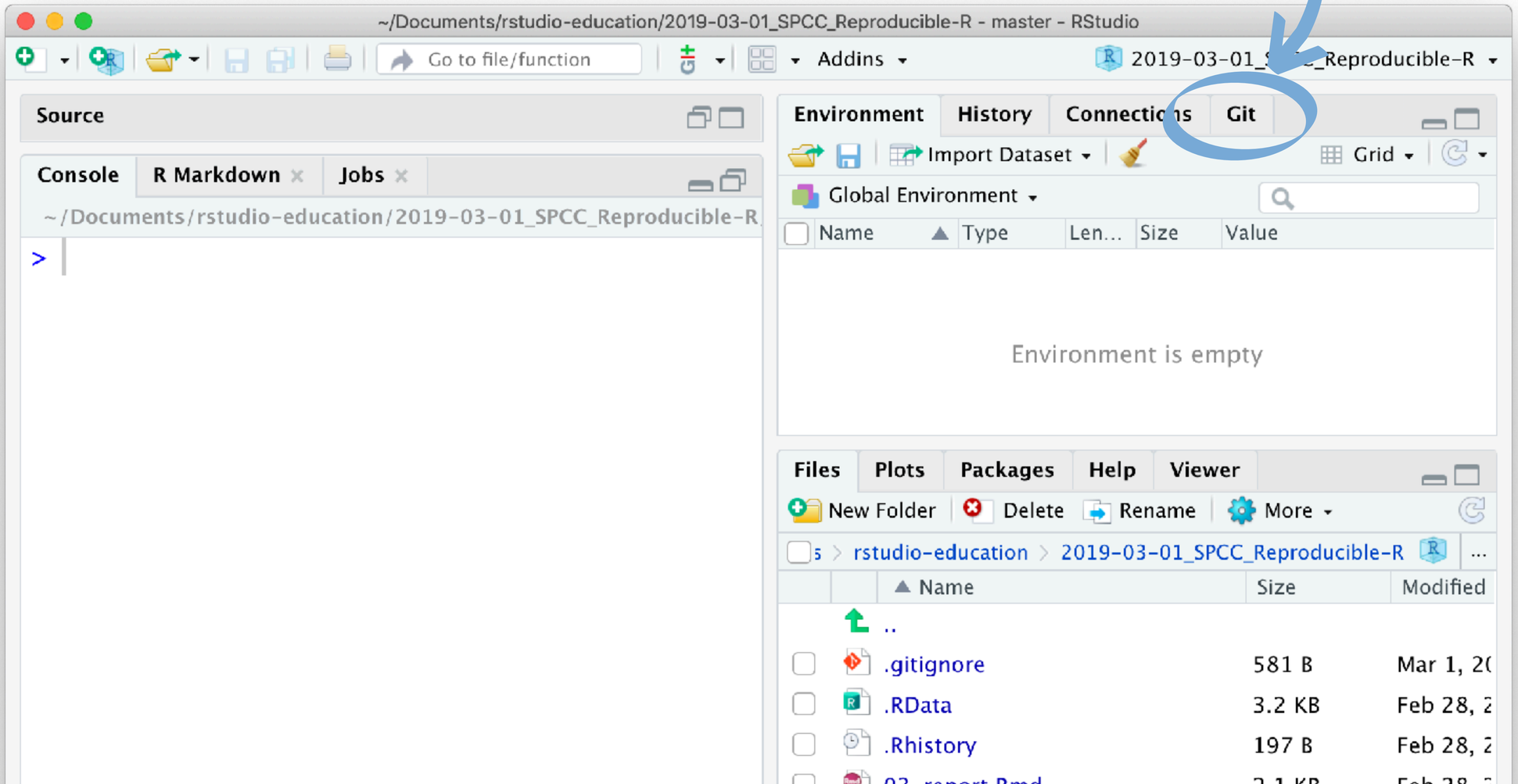
in your working directory

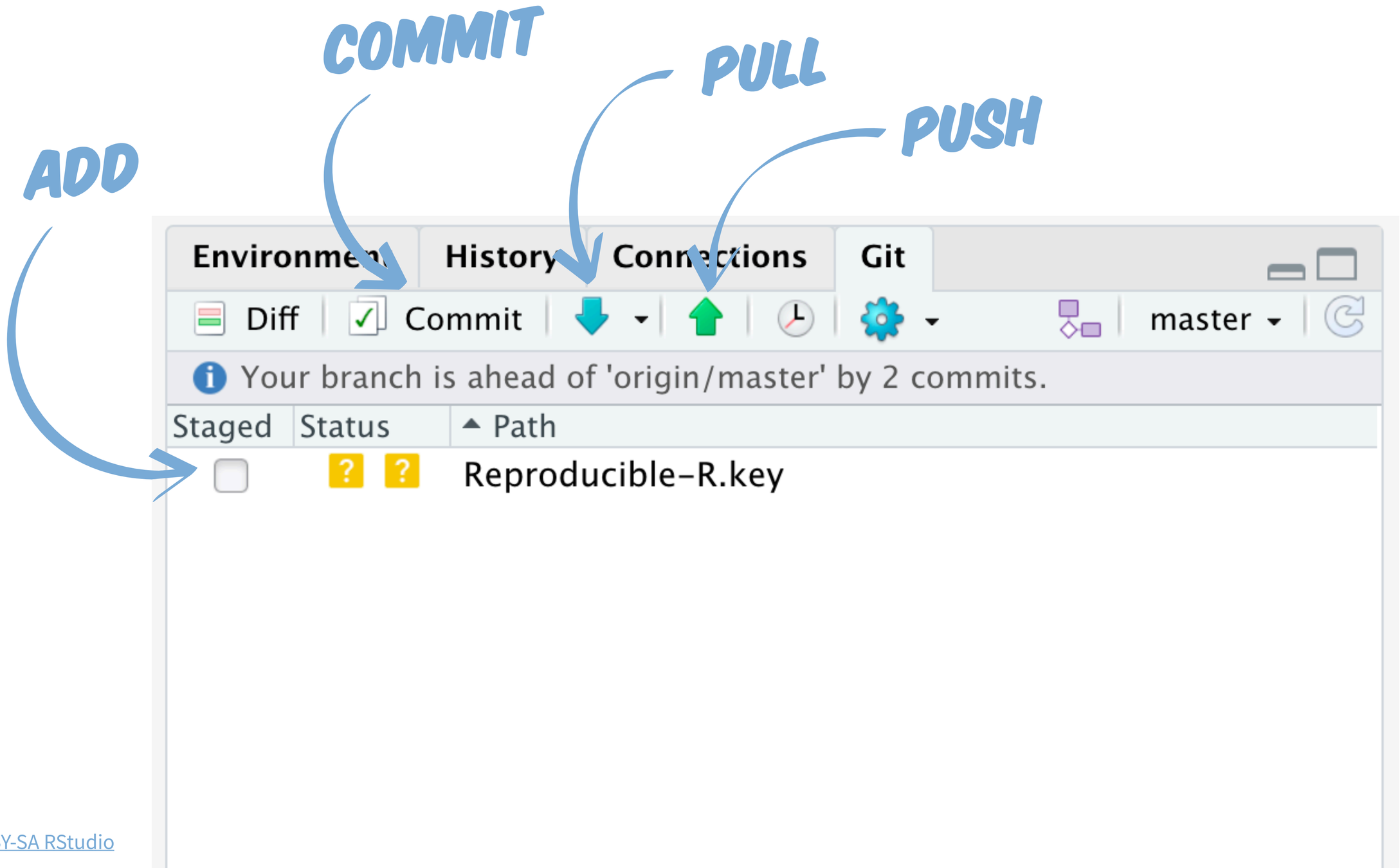




DOWNLOAD A REPOSITORY AS A PROJECT

A GIT PANE





R Markdown Takeaway


Mythos
PREMIUM QUALITY HELLENIC BEER
PREMIUM
QUALITY
Mythos
HELLENIC BEER
750mL?

Dublin - Google Sheets

Garrett

Secure https://docs.google.com/spreadsheets/d/1OrHXGZJIyE_xyrbYs4YBEIF...

Dublin

File Edit View Insert Format Data

SHARE

100% $ % .0 .00 123

fx bottle

	A	B	C	D	E
1	bottle	amount			
2	1	748.7			
3	2	746.4			
4	3	750.1			
5	4	751.9			
6	5	747.6			
7	6	748.8			
8	7	748.6			
9	8	752.2			
10	9	749.9			
11	10	745.9			
12	11	747.1			
13	12	748.3			
14	13	750.4			
15	14	749.6			
16	15	746			
17	16	750.5			
18	17	750.5			
19		750.1			
20		743.9			
21		750.6			

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

rmarkdown-demo.Rmd*

Knit | Insert | Run

```
1  ---
2  title: 'Re: Bottle Machine'
3  output: html_document
4  ---
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
```

14:1 (Top Level) R Markdown

Files | Plots | Packages | Help | Viewer

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

rmarkdown-demo.Rmd*

Knit | Insert | Run

```
1  ---
2  title: 'Re: Bottle Machine'
3  output: html_document
4  ---
5
6  ```{r setup, include=FALSE, message=FALSE}
7  library(googlesheets)
8  library(tidyverse)
9
10 key <- "1OrHXGZJIyE_xyrbYs4YBElFUtmESYZE7n1rK5HuoR4w"
11 samples <- gs_read(gs_key(key), lookup = FALSE)
12 obs <- mean(samples$amount)
13 ```
14
15 ```{r echo = FALSE}
16 samples$amount
17 ```
18
19 ```{r echo = FALSE}
20 sim <- map_dbl(1:1000,
21                ~mean(rnorm(50, mean = 750, sd = 3)))
22 delta <- abs(750 - obs)
23 pvalue <- mean(abs(750 - sim) >= delta)
24 ```
25
26 ```{r echo = FALSE, fig.height = 2}
27 ggplot() +
```

14:1 (Top Level) R Markdown

Files | Plots | Packages | Help | Viewer

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

rmarkdown-demo.Rmd*

Knit | Insert | Run

````
```{r echo = FALSE, fig.height = 2}
ggplot() +
  geom_histogram(aes(x = sim), bins = 30) +
  geom_rect(aes(xmin = -Inf, xmax = 750 - delta,
                ymin = -Inf, ymax = Inf),
            fill = "#4197D9", alpha = 0.2) +
  geom_rect(aes(xmin = 750 + delta, xmax = Inf,
                ymin = -Inf, ymax = Inf),
            fill = "#4197D9", alpha = 0.2) +
  geom_vline(aes(xintercept = obs),
             color = "#4197D9", size = 3)
```
````

14:1 (Top Level) R Markdown

Files | Plots | Packages | Help | Viewer

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

rmarkdown-demo.Rmd*

Knit | Insert | Run

```
10  key <- "1OrHXGZJIyE_xyrbYs4YBElFUtmESYZE7n1rK5HuoR4w"
11  samples <- gs_read(gs_key(key), lookup = FALSE)
12  obs <- mean(samples$amount)
13  ```
14
15  Does our bottle machine fill each bottle with 750 mL
    of beer (on average)?
16
17
18
19
20
21  ```{r echo = FALSE}
```

Does our bottle machine fill each bottle with 750 mL of beer (on average)?

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

rmarkdown-demo.Rmd*

Knit | Insert | Run

```
10  key <- "1OrHXGZJIyE_xyrbYs4YBElFUtmESYZE7n1rK5HuoR4w"
11  samples <- gs_read(gs_key(key), lookup = FALSE)
12  obs <- mean(samples$amount)
13  ```
14
15  ## Data
    Beer bottle machine fill each bottle with 750 mL
    of beer (on average)?
16  I selected 50 bottles at random from the Dublin
17  factory, which contained the following amounts of
18  beer (mL):
19
20
21  ```{r echo = FALSE}
```

Data

I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

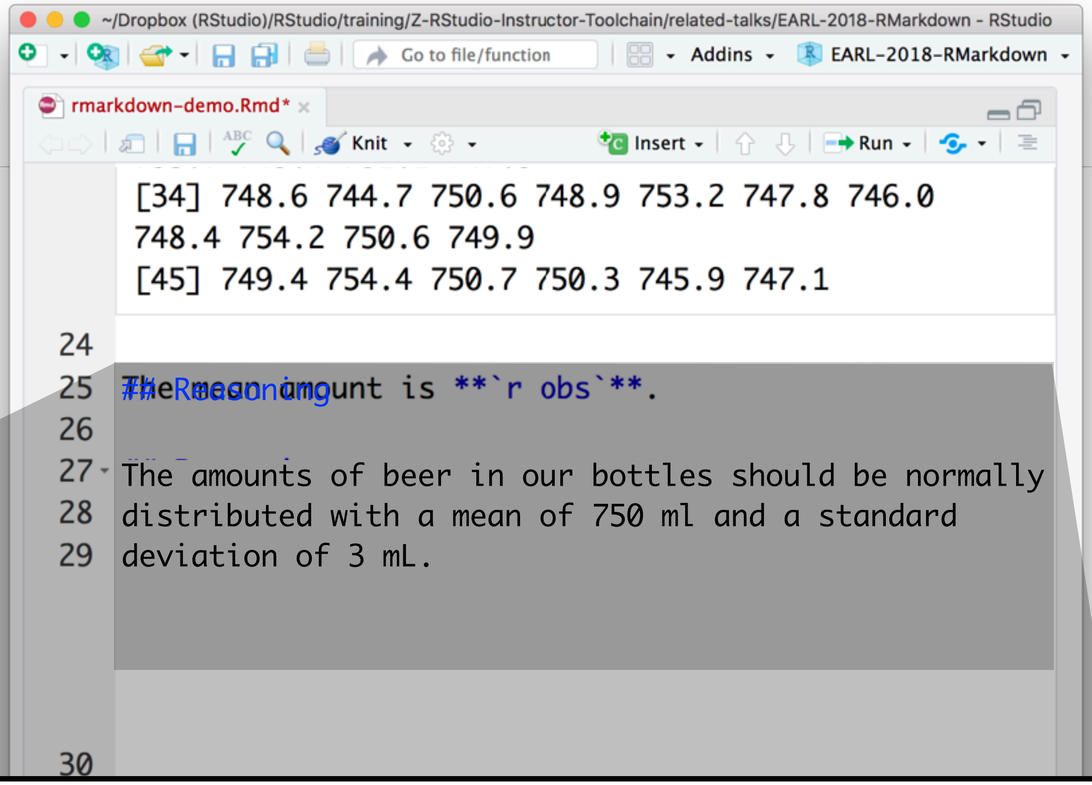
rmarkdown-demo.Rmd*

Knit | Insert | Run

```
13
14
15  Does our bottle machine fill each bottle with 750 mL
    of beer (on average)?
16
17  ## Data
18
19  I selected 50 bottles at random from the Dublin
    factory, which contained the following amounts of
    beer (mL):
20
21  ```{r echo = FALSE}
22  samples$amount
23  ```
```

The mean amount is **`r obs`**.

The mean amount is **`r obs`**.



Reasoning

The amounts of beer in our bottles should be normally distributed with a mean of 750 ml and a standard deviation of 3 mL.

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

rmarkdown-demo.Rmd*

Knit | Insert | Run

```
[45] 749.4 754.4 750.7 750.3 745.9 747.1

24
25 The mean amount is **`r obs`**.
26
27 ## Reasoning
28
29 The amounts of beer in our bottles should be normally
   distributed with a mean of 750 ml and a standard
   deviation of 3 mL. Let's use simulation to calculate
   the sample means of 1000 samples of size 50 generated
   from such a distribution.
30
31 ```{r echo = FALSE}
32 sim <- map_dbl(1:1000,
```

Reasoning

The amounts of beer in our bottles should be normally distributed with a mean of 750 ml and a standard deviation of 3 mL. Let's use simulation to calculate the sample means of 1000 samples of size 50 generated from such a distribution.

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

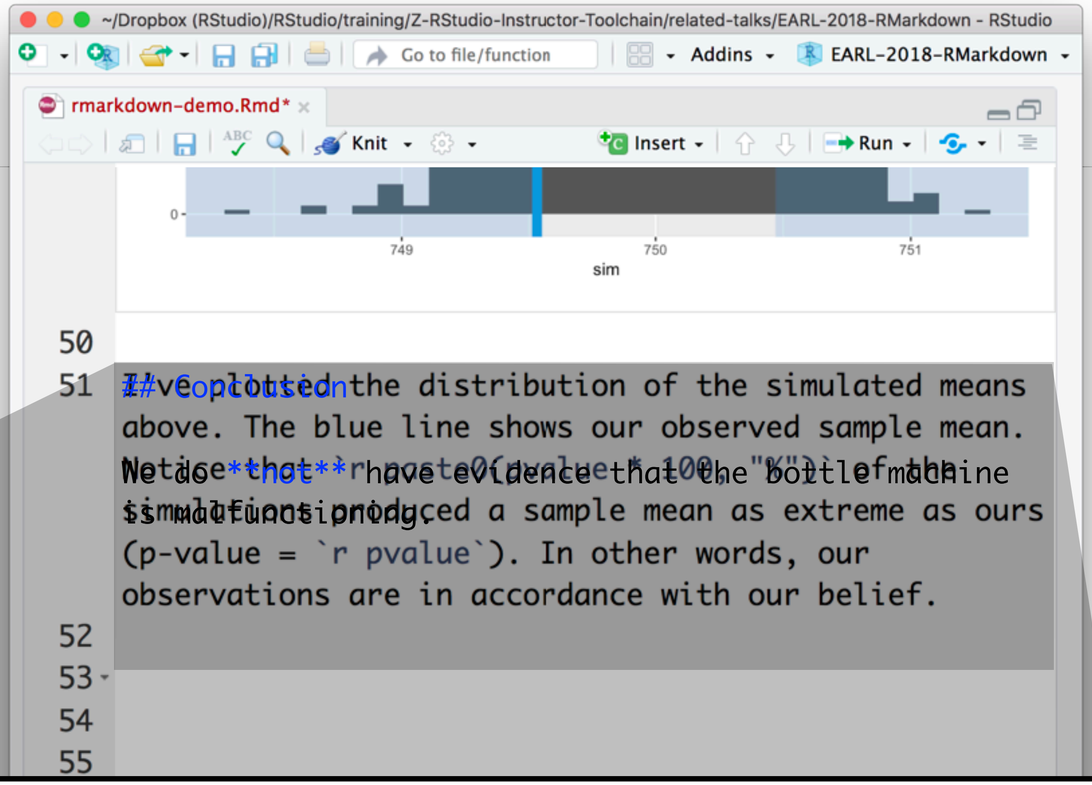
rmarkdown-demo.Rmd*

Knit | Insert | Run

```
[45] 749.4 754.4 750.7 750.3 745.9 747.1

24
25 The mean amount is **`r obs`**.
26
27 ## Results
28 I've plotted the distribution of the simulated means
29 above. The blue line shows our observed sample mean.
   Notice that `r paste0(pvalue * 100, "%")` of the
   simulations produced a sample mean as extreme as ours
   (p-value = `r pvalue`). In other words, our
   observations are in accordance with our belief.
30
31 ```{r echo = FALSE}
32 sim <- map_dbl(1:1000,
```

I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that `r paste0(pvalue * 100, "%")` of the simulations produced a sample mean as extreme as ours (p-value = `r pvalue`). In other words, our observations are in accordance with our belief.



Conclusion

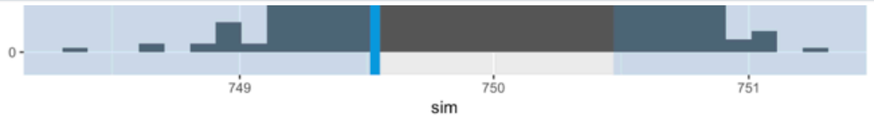
We do **not** have evidence that the bottle machine is malfunctioning.

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

rmarkdown-demo.Rmd*

Knit | Insert | Run




```
50
51  I've plotted the distribution of the simulated means
    above. The blue line shows our observed sample mean.
    Notice that `r paste0(pvalue * 100, "%")` of the
    simulations produced a sample mean as extreme as ours
    (p-value = `r pvalue`). In other words, our
    observations are in accordance with our belief.
52
53  ## Conclusion
54
55  We do **not** have evidence that the bottle machine
    is malfunctioning.
56
```

14:1 (Top Level) R Markdown

Files | Plots | Packages | Help | Viewer

RStudio: View PDF

Page: 1 of 1 Automatic Zoom

Re: Bottle Machine

Does our bottle machine fill each bottle with 750 mL of beer (on average)?

Data

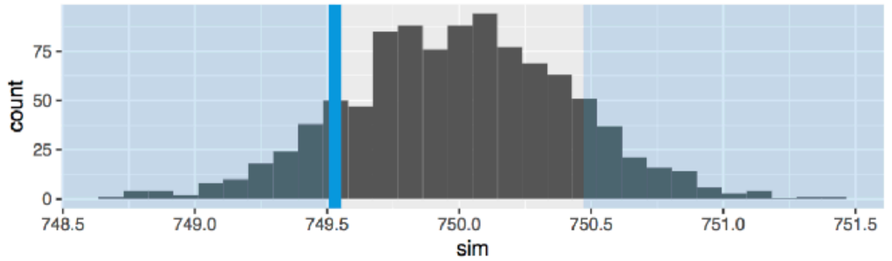
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 748.6 752.2 749.9 745.9 747.1
## [12] 748.3 750.4 749.6 746.0 750.5 750.5 750.1 743.9 750.6 758.0 746.3
## [23] 752.5 751.6 748.1 751.8 744.6 751.0 751.7 753.7 749.4 752.1 747.9
## [34] 748.6 744.7 750.6 748.9 753.2 747.8 746.0 748.4 754.2 750.6 749.9
## [45] 749.4 754.4 750.7 750.3 745.9 747.1
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should be normally distributed with a mean of 750 mL and a standard deviation of 3 mL. Let's use simulation to calculate the sample means of 1000 samples of size 50 generated from such a distribution.



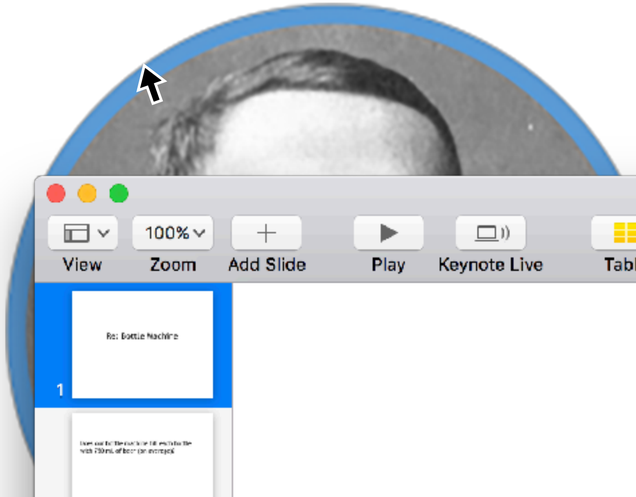
I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that 26.2% of the simulations produced a sample mean as extreme as ours (p-value = 0.262). In other words, our observations are in accordance with our belief.

Conclusion

We do **not** have evidence that the bottle machine is malfunctioning.

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Files Plots Packages Help Viewer



~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

rmarkdown-demo.Rmd*

Knit | Insert | Run

rmarkdown-demo

View | Zoom | Add Slide | Play | Keynote Live | Table | Chart | Text | Shape | Media | Comment | Front | Back | Forward | Backward | Format | Animate | Document

1 2 3 4 5 6 7

Re: Bottle Machine

Slide Layout

Title Slide

Change Master

Appearance

- [x] Title
- [x] Body
- [ ] Slide Number

Background

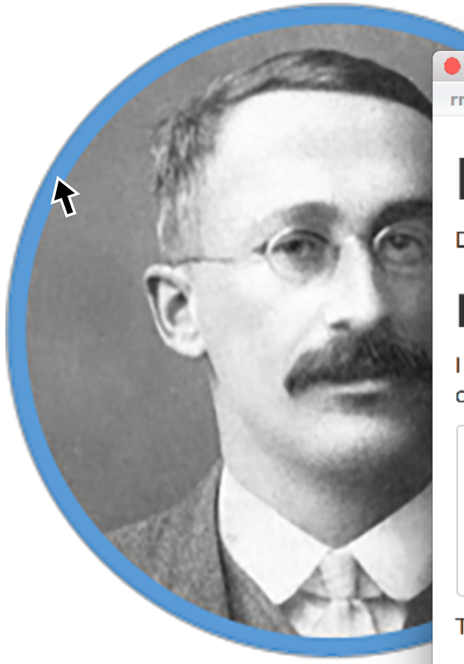
Color Fill

Edit Master Slide

```
21 ```{r echo = FALSE}
22 samples$amount
23 ```
```

14:1 (Top Level) R Markdown

Files | Plots | Packages | Help | Viewer



~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMark...

rmarkdown-demo.html | Open in Browser | Find | Publish

Re: Bottle Machine

Does our bottle machine fill each bottle with 750 mL of beer (on average)?

Data

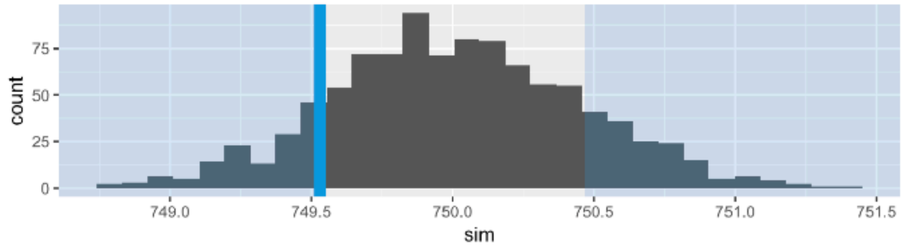
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 748.6 752.2 749.9 745.9 747.1
## [12] 748.3 750.4 749.6 746.0 750.5 750.5 750.1 743.9 750.6 758.0 746.3
## [23] 752.5 751.6 748.1 751.8 744.6 751.0 751.7 753.7 749.4 752.1 747.9
## [34] 748.6 744.7 750.6 748.9 753.2 747.8 746.0 748.4 754.2 750.6 749.9
## [45] 749.4 754.4 750.7 750.3 745.9 747.1
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should be normally distributed with a mean of 750 mL and a standard deviation of 3 mL. Let's use simulation to calculate the sample means of 1000 samples of size 50 generated from such a distribution.



I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that 28.3% of the simulations produced a sample mean as extreme as ours (p-value = 0.283). In other words, our observations are in accordance with our belief.

Conclusion

We do **not** have evidence that the bottle machine is malfunctioning.

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Addins | EARL-2018-RMarkdown

Run

sage=FALSE}

FUtmESYZE7n1rK5HuoR4w"
lookup = FALSE)

ch bottle with 750 mL

from the Dublin
llowing amounts of

R Markdown

RStudio Connect

Garrett

Secure | https://connect.rstudioservices.com/connect/#/apps/149/access/105

Content / rmarkdown-demo

WG bill

Re: Bottle Machine

Does our bottle machine fill each bottle with 750 mL of beer (on average)?

Data

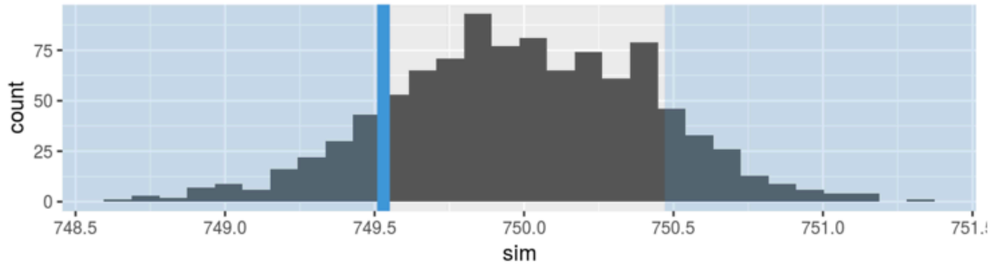
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 748.6 752.2 749.9 745.9 747.1
## [12] 748.3 750.4 749.6 746.0 750.5 750.5 750.1 743.9 750.6 758.0 746.3
## [23] 752.5 751.6 748.1 751.8 744.6 751.0 751.7 753.7 749.4 752.1 747.9
## [34] 748.6 744.7 750.6 748.9 753.2 747.8 746.0 748.4 754.2 750.6 749.9
## [45] 749.4 754.4 750.7 750.3 745.9 747.1
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should be normally distributed with a mean of 750 mL and a standard deviation of 3 mL. Let's use simulation to calculate the sample means of 1000 samples of size 50 generated from such a distribution.



I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that 27.4% of the simulations produced a sample mean as extreme as ours (p-value = 0.274). In other words, our observations are in accordance with our belief.

Conclusion

We do **not** have evidence that the bottle machine is malfunctioning.

Info | Access | Runtime | Schedule | Tags | Vars | Logs

Who can view this document

You

Who can change this document

WG William Gosset bill

Add collaborator

Who runs this document on the server

The default user

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.

Your custom url:

https://connect.rstudioservices.com/garrett/rmarkdown-demo

RStudio Connect

Garrett

Secure | https://connect.rstudioservices.com/connect/#/apps/149/access/105

Content / rmarkdown-demo

WG bill

Your email request has been submitted.

Does our bottle machine fill each bottle with 750 mL of beer

Data

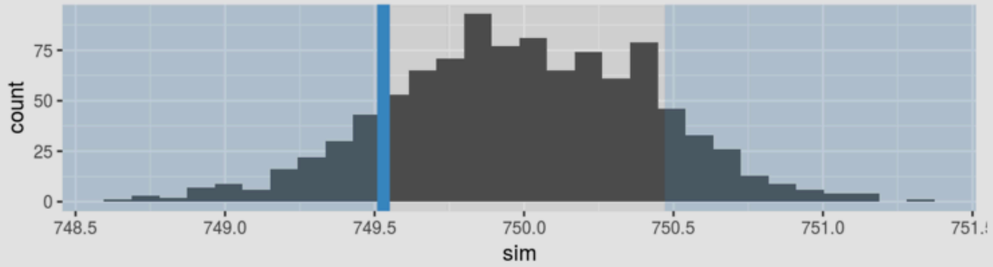
I selected 50 bottles at random from the Dublin factory, whic

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 748
## [12] 748.3 750.4 749.6 746.0 750.5 750.5 750
## [23] 752.5 751.6 748.1 751.8 744.6 751.0 751
## [34] 748.6 744.7 750.6 748.9 753.2 747.8 746
## [45] 749.4 754.4 750.7 750.3 745.9 747.1
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should be normally distributed with a mean of 750 mL and a standard deviation of 3 mL. Let's use simulation to calculate the sample means of 1000 samples of size 50 generated from such a distribution.



I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that 27.4% of the simulations produced a sample mean as extreme as ours (p-value = 0.274). In other words, our observations are in accordance with our belief.

Conclusion

We do **not** have evidence that the bottle machine is malfunctioning.

Email the report to:

- just me
- collaborators
- collaborators & viewers

Send

Info | Access | Runtime | Schedule | Tags | Vars | Logs

Who can view this document

Collaborators & you

Who can change this document

WG William Gosset bill

VP VP Virginia virginia

Add collaborator

Who runs this document on the server

The default user

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.

Your custom url:

https://connect.rstudioservices.com/garrett/rmarkdown-demo

One more thing

Reproducibility is an opportunity

RStudio Connect

Garrett

Secure | https://connect.rstudioservices.com/connect/#/apps/149/access/105

Content / rmarkdown-demo

WG bill

Re: Bottle Machine

Does our bottle machine fill each bottle with 750 mL of beer (on average)?

Data

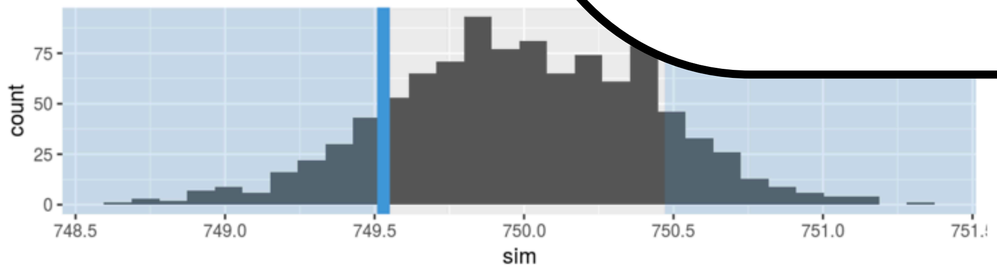
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (ml):

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 748.6 75
## [12] 748.3 750.4 749.6 746.0 750.5 750.5 750.1
## [23] 752.5 751.6 748.1 751.8 744.6 751.0 751
## [34] 748.6 744.7 750.6 748.9 753.2 747.8 74
## [45] 749.4 754.4 750.7 750.3 745.9 747.1
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should be normally distr
simulation to calculate the sample means of 1000 samples o



I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that 27.4% of e
simulations produced a sample mean as extreme as ours (p-value = 0.274). In other words, our observations are in accordance wi
our belief.

Conclusion

We do **not** have evidence that the bottle machine is malfunctioning.

Info | Access | Runtime | Schedule | Tags | Vars | Logs

Who can view this document

Collaborators & you

can change this document

William Gosset bill

VP Virginia virginia

dd collaborator

Your custom url:

https://connect.rstudioservices.com/garrett/rma wn-demo

Can you do this ***every*** week?

Schedule saved successfully.

Info Access Runtime Schedule Tags Vars Logs

Schedule output for default

Start date & time

Fri Sep 07 2018 12:39:20 GMT-0500

Set to Now

Schedule type

Weekly

Run every 1 week.

Run every...

- Sunday
- Monday
- Tuesday
- Wednesday
- Thursday
- Friday
- Saturday

Publish output after it is generated

Send email after update

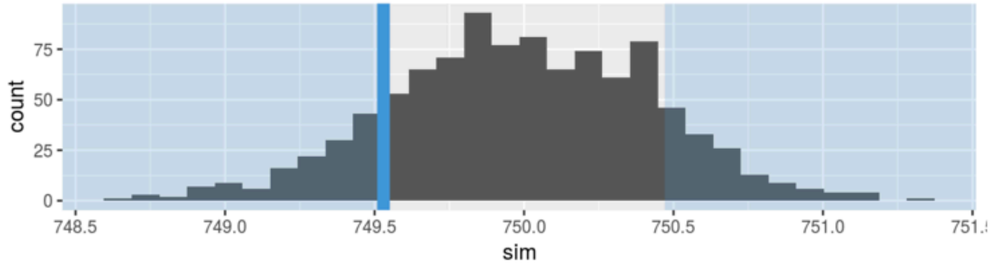
ed the following amounts of beer (mL):

```
##  [1] 752.2 749.9 745.9 747.1
##  [5] 743.9 750.6 758.0 746.3
## [23] 753.7 749.4 752.1 747.9
## [34] 756.0 748.4 754.2 750.6 749.9
## [45] 749.
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should be normally distributed with a mean of 750 mL and a standard deviation of 3 mL. Let's use simulation to calculate the sample means of 1000 samples of size 50 generated from such a distribution.



I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that 27.4% of the simulations produced a sample mean as extreme as ours (p-value = 0.274). In other words, our observations are in accordance with our belief.

Conclusion

We do **not** have evidence that the bottle machine is malfunctioning.

RStudio Connect

Garrett

Secure | https://connect.rstudioservices.com/connect/#/apps/149/access/105

Content / rmarkdown-demo

WG bill

Re: Bottle Machine

Does our bottle machine fill each bottle with 750 mL of beer (on average)?

Data

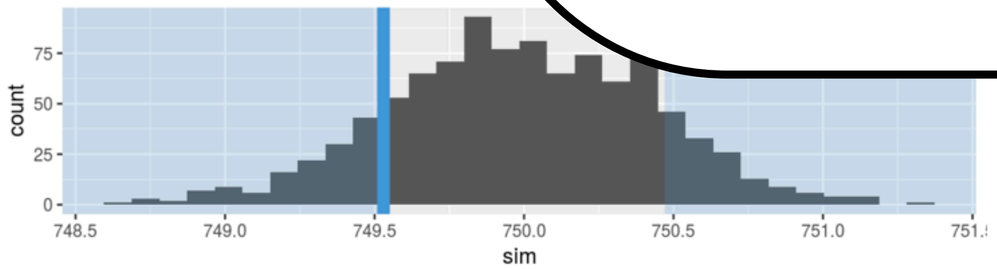
I selected 50 bottles at random from the Dublin factory, which co

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 7
## [12] 748.3 750.4 749.6 746.0 750.5 750.5
## [23] 752.5 751.6 748.1 751.8 744.6 751.
## [34] 748.6 744.7 750.6 748.9 753.2 747
## [45] 749.4 754.4 750.7 750.3 745.9 747
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should be normally d
simulation to calculate the sample means of 1000 sampl



I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that 27.4% of e
simulations produced a sample mean as extreme as ours (p-value = 0.274). In other words, our observations are in accordance wi
our belief.

Conclusion

We do **not** have evidence that the bottle machine is malfunctioning.

Info | Access | Runtime | Schedule | Tags | Vars | Logs

Schedule output for default

Start date & time

Fri Sep 07 2018 12:39:20 GMT-0500

Set to Now

le type

dy

every

eek

un every...

Sunday

Monday

Send email after update

Can you do this for the ***London*** factory?

~/Dropbox (RStudio)/RStudio/training/Z-RStudio-Instructor-Toolchain/related-talks/EARL-2018-RMarkdown - RStudio

Go to file/function | Addins | EARL-2018-RMarkdown

rmarkdown-demo.Rmd*

Knit | Insert | Run

```
1  ---
2  title: 'Re: Bottle Machine'
3  output: html_document
4  params:
5    factory:
6      value: Dublin
7      choices:
8        - Dublin
9        - London
10
11 ```{r setup, include=FALSE, message=FALSE}
12 library(googlesheets)
13 library(tidyverse)
14
15 key <- "1OrHXGZJIyE_xyrbYs4YBElFUtmESYZE7n1rK5HuoR4w"
16 samples <- gs_read(gs_key(key), lookup = FALSE)
17 obs <- mean(samples$amount)
18 ```
```

14:1 (Top Level) R Markdown

Files | Plots | Packages | Help | Viewer

```
params:
  factory:
    value: Dublin
    choices:
      - Dublin
      - London
```

RStudio Connect

Garrett

Secure https://connect.rstudioservices.com/connect/#/apps/150/access/106

Content / rmarkdown-demo

WG bill

Re: Bottle Machine

Does our bottle machine fill each bottle with 750 mL of beer (on average)?

Data

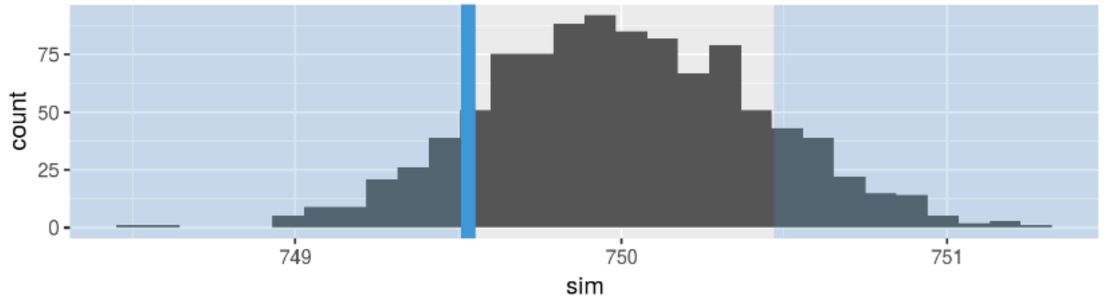
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 748.6 752.2 749.9 745.9 747.1
## [12] 748.3 750.4 749.6 746.0 750.5 750.5 750.1 743.9 750.6 758.0 746.3
## [23] 752.5 751.6 748.1 751.8 744.6 751.0 751.7 753.7 749.4 752.1 747.9
## [34] 748.6 744.7 750.6 748.9 753.2 747.8 746.0 748.4 754.2 750.6 749.9
## [45] 749.4 754.4 750.7 750.3 745.9 747.1
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should be normally distributed with a mean of 750 mL and a standard deviation of 3 mL. Let's use simulation to calculate the sample means of 1000 samples of size 50 generated from such a distribution.



I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that 26% of the simulations produced a sample mean as extreme as ours (p-value =

INPUT

Info | Access | Runtime | Schedule | Tags | Vars | Logs

Who can view this document

You

Who can change this document

WG William Gosset bill

Add collaborator

Who runs this document on the server

The default user

Administrators can create a custom "vanity" url to access this document. Your custom URL will be appended to your domain to form the complete path to your document.

Your custom url:

https://connect.rstudioservices.com/garrett/rmarkdown-demo

RStudio Connect

Garrett

Secure https://connect.rstudioservices.com/connect/#/apps/150/access/106

Content / rmarkdown-demo

WG bill

Re: Bottle Machine

Does our bottle machine fill each bottle with 750 mL of beer (on average)?

Data

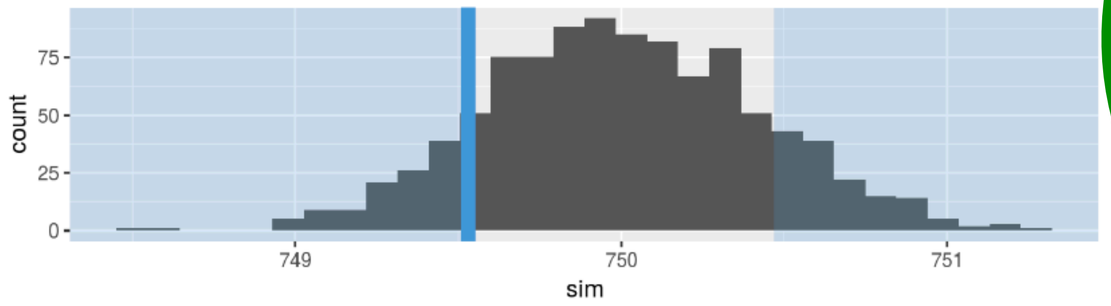
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 748.6 752.2 749.9 745.9 747.1
## [12] 748.3 750.4 749.6 746.0 750.5 750.5 750.1 743.9 750.6 758.0 746.3
## [23] 752.5 751.6 748.1 751.8 744.6 751.0 751.7 753.7 749.4 752.1 747.9
## [34] 748.6 744.7 750.6 748.9 753.2 747.8 746.0 748.4 754.2 750.6 749.9
## [45] 749.4 754.4 750.7 750.3 745.9 747.1
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should be normally distributed with a mean of 750 mL and a standard deviation of 3 mL. Let's use simulation to calculate the sample means of 1000 samples of size 50 generated from such a distribution.



I've plotted the distribution of the simulated means above. The blue line shows our observed sample mean. Notice that 26% of the simulations produced a sample mean as extreme as ours (p-value =

INPUT

Info Access Runtime Schedule Tags Vars Logs

Who can view this document

You

Who can change this document

WG William Gosset bill

Add collaborator

Who runs this document on the server

https://connect.rstudioservices.com/garrett/rmarkdown-dem

RStudio Connect

Garrett

Secure https://connect.rstudioservices.com/connect/#/apps/150/access/106

Content / rmarkdown-demo

bill

Re: Bottle Mac

Does our bottle machine fill each bottle wi

Data

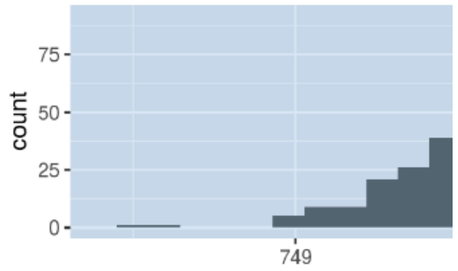
I selected 50 bottles at random from the D
(mL):

```
##  [1] 748.7 746.4 750.1 751.
## [12] 748.3 750.4 749.6 746.
## [23] 752.5 751.6 748.1 751.
## [34] 748.6 744.7 750.6 748.
## [45] 749.4 754.4 750.7 750.
```

The mean amount is **749.53**.

Reasoning

The amounts of beer in our bottles should
standard deviation of 3 mL. Let's use simu
50 generated from such a distribution.



I've plotted the distribution of the simulate
mean. Notice that 26% of the simulations

Re: Bottle Machine

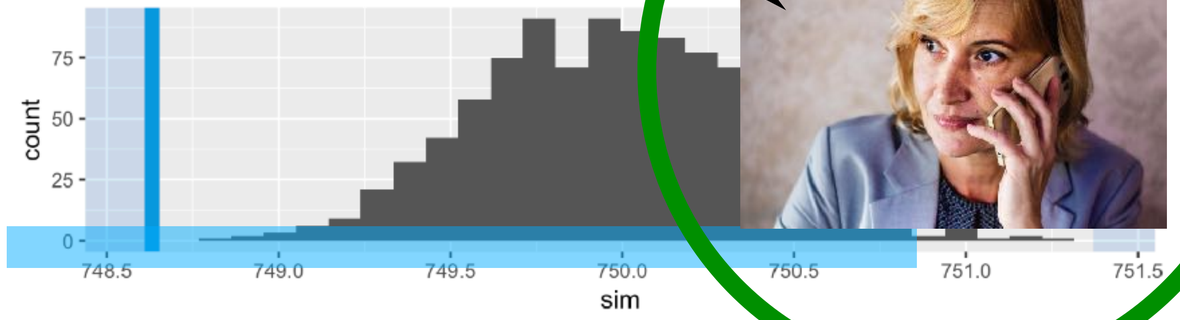
Does our bottle machine fill each bottle with 750 mL of beer (on average)?

Data

ined the following amounts of beer (mL):

```
7.7 747.4 751.6
.1 745.5 752.0
.5 750.7 748.9
2 751.4 749.5
```

with a mean of 750 mL and a standard devi ion
h a
distribution.



I've plotted the distribution of the simulated means above. The blue line show ple mean. Notice that 0% of the simulations produced a sample mean as extreme as ours (p-value = 0). In other words, our

I like it!

Reproducibility is an opportunity

schedule

parameterize

automate

What should we do the leftover beer?

Thank You