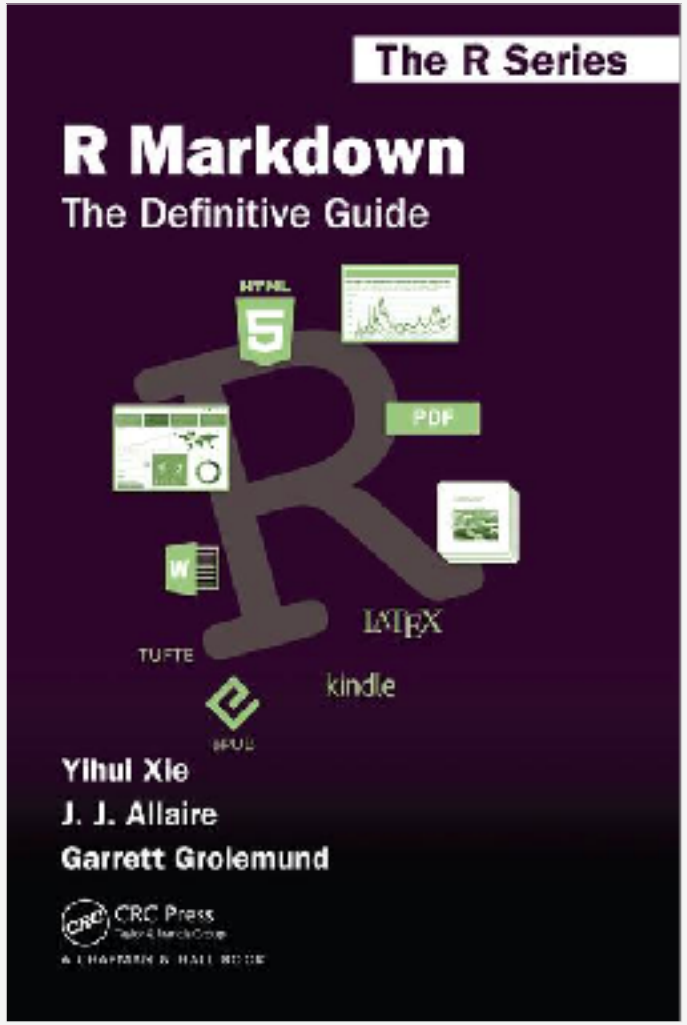
R Markdown:

The Bigger Picture

January 17, 2019

Garrett Grolemund, RStudio @StatGarrett

http://**bookdown.org/yihui/rmarkdown**



Why Do So Many Studies Fail to Replicate?

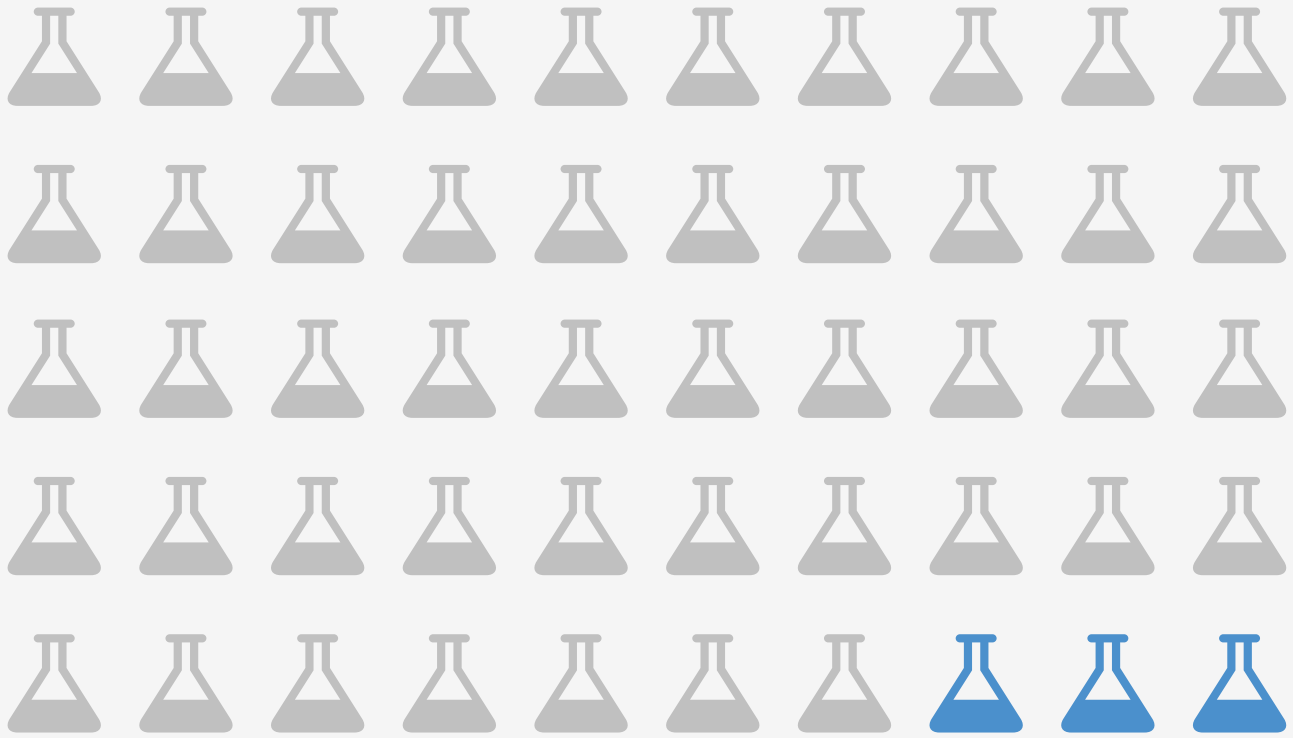
- The New York Times, May 2016

Scientists' Elusive Goal: Reproducing Study Results

- The Wall Street Journal, Dec 2011

Challenges in irreproducible research

- Nature, July 2018



Amgen 2012: could reproduce only **6** of **53** landmark results

75%–90%

of preclinical results cannot be reproduced (estimate)

Begley CG, Ioannidis JP (2015) Reproducibility in Science: improving the standard for basic and preclinical research.
Circ Res 116: 116–126. pmid:25552691

$28 billion

£ 22B € 24B

Costs to biomedical industry per year, US

Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165. https://doi.org/10.1371/journal.pbio.1002165

Economics

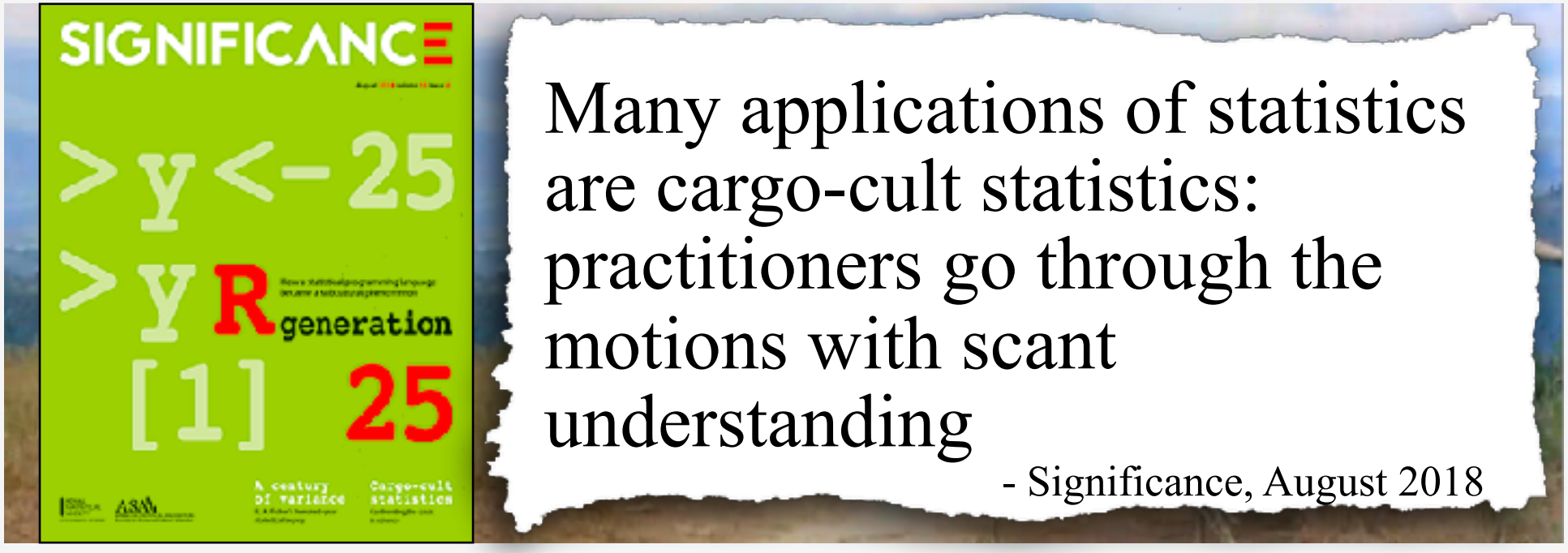
6 of **18**

Nature & Science

13 of **21**

Credibility


SIGNIFICANCE
>y<-25
>y
R generation
[1] 25
A century of variance
Cargo-cult statistics



Many applications of statistics are cargo-cult statistics: practitioners go through the motions with scant understanding

- Significance, August 2018

In the case considered in Ex. 13·1, we have therefore

$$\frac{18!\,12!\,17!\,13!}{30!}\left\{\frac{1}{12!\,3!\,10!\,15!}+\frac{1}{2!\,11!\,16!}+\frac{1}{11!\,12!\,17!}\right\}$$

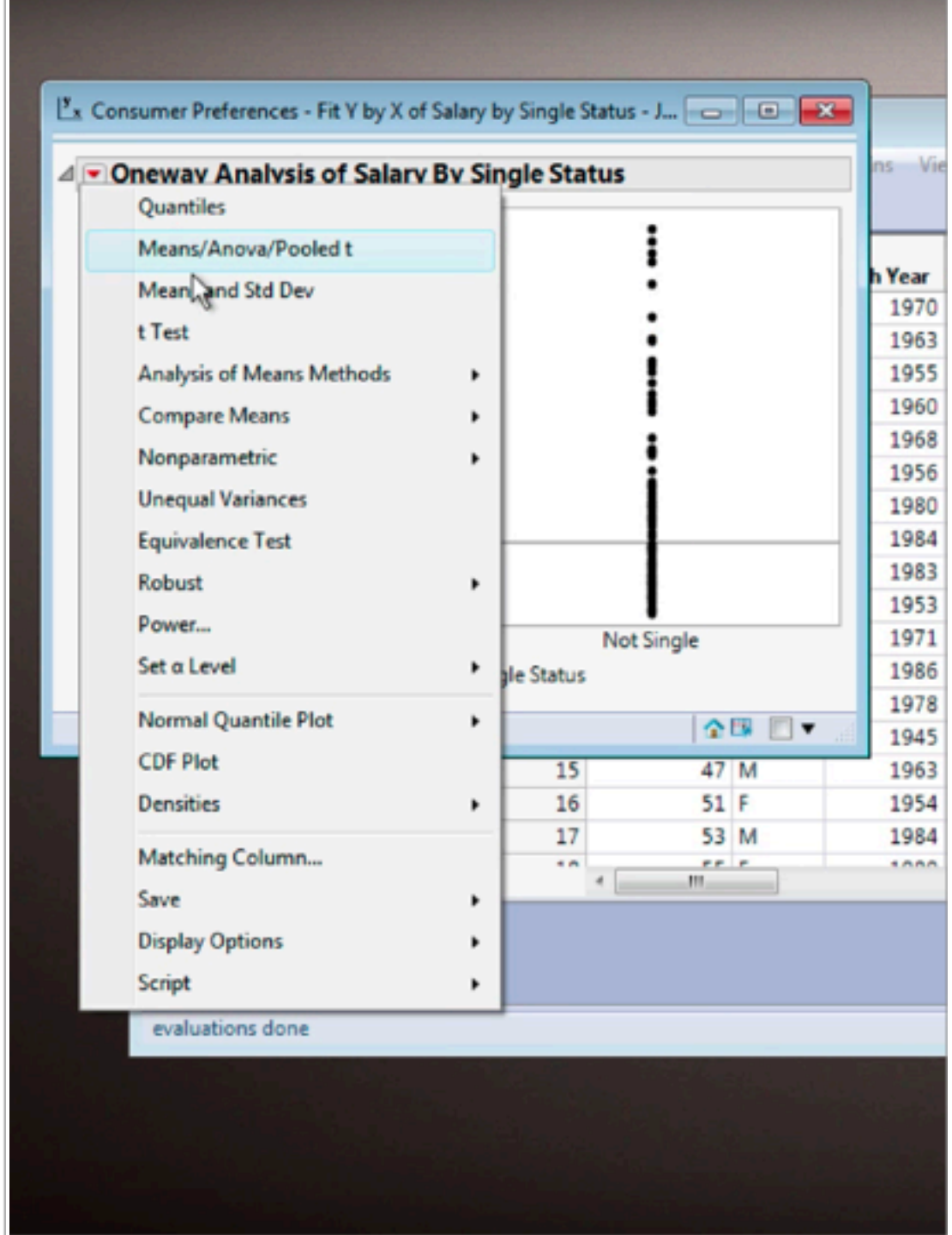
for the probabilities of the set of frequencies observed, and the two possible more extreme sets of frequencies which might have been observed. Without any assumption or approximation, therefore, the table observed may be judged significantly to contradict the hypothesis of proportionality if

$$\frac{18!\,13!}{30!}(2992+102+1)$$

is a small quantity. This amounts to 619/1330665, or about 1 in 2150, showing that if the hypothesis of proportionality were true, observations of the kind recorded would be highly exceptional.

21·03. In its primary purpose of the comparison of a series of observed frequencies with those expected on the hypothesis to be tested, the χ^2 test is an approximate one, though validly applicable in an immense range of important cases. For other tests where the observations are measurements, instead of frequencies, it provides exact tests of significance. Of these the two most important are:—

(i.) its use to test whether a sample from a normal distribution confirms or contradicts the variance which this distribution is expected on theoretical grounds to have, and

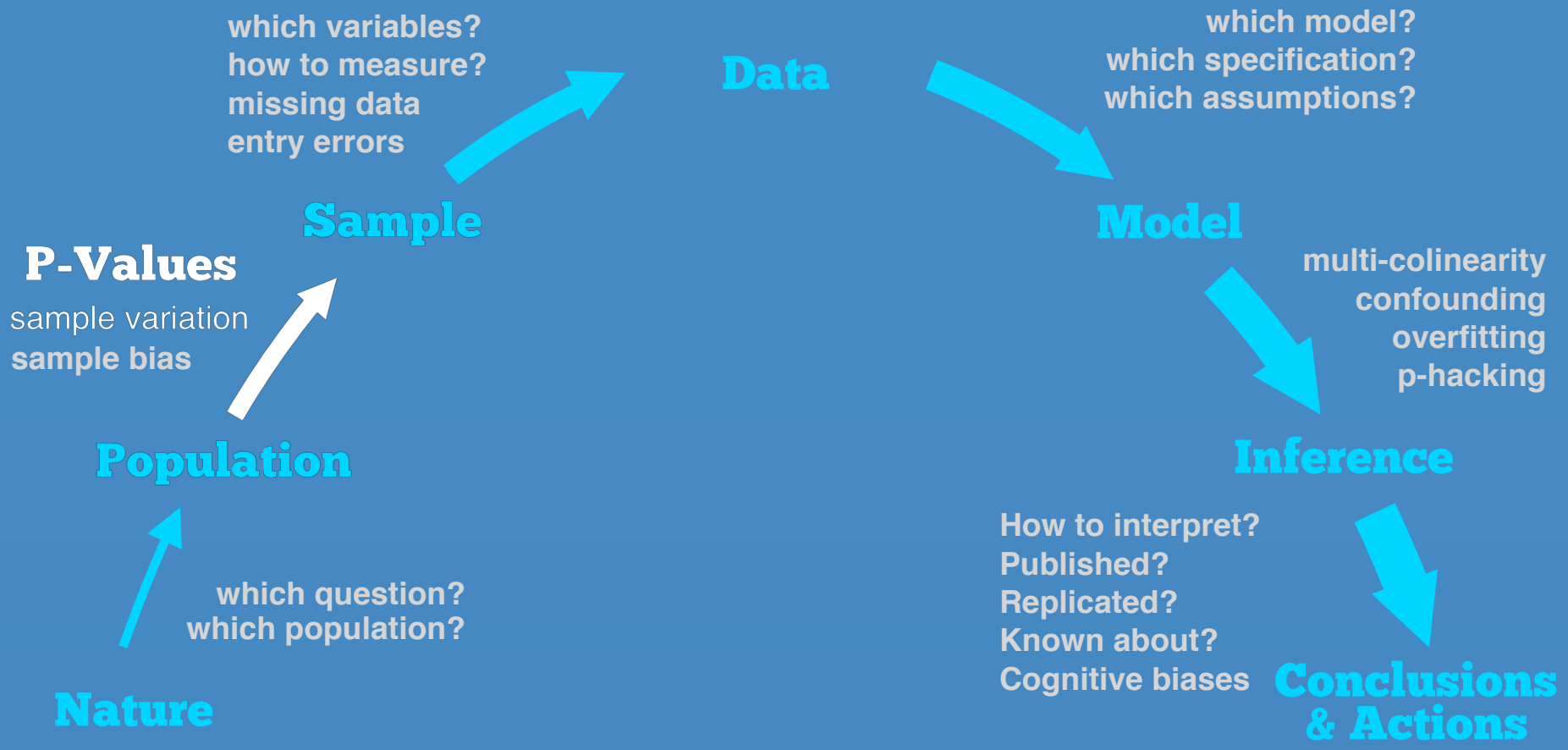


P-Values

sample variation

Sample

Population



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

where $q = 1 - p$. The probability that in a sample of $(c+d)$ trials it will occur c times is

$$\frac{(c+d)!}{c!\,d!} p^c q^d.$$

So that the probability of the observed frequencies a, b, c, and d in a 2×2 table is the product

$$\frac{(a+b)!\,(c+d)!}{a!\,b!\,c!\,d!} p^{a+c} q^{b+d},$$

and this in general must be unknown if p is unknown. The unknown factor involving p and q will, however, be the same for all tables having the same marginal frequencies $a+c$, $b+d$, $a+b$, $c+d$, so that among possible sets of observations having the same marginal frequencies, the probabilities are in proportion to

$$\frac{1}{a!\,b!\,c!\,d!},$$

whatever may be the value of p, or, in other words, for all populations in which the four frequencies are in proportion.

Now the sum of the quantities $1/a!\,b!\,c!\,d!$ for all samples having the same margins is found to be

$$\frac{n!}{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!},$$

where $n = a+b+c+d$; so that, given the marginal frequencies, the probability of any observed set of entries is

$$\frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{n!} \cdot \frac{1}{a!\,b!\,c!\,d!}.$$

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

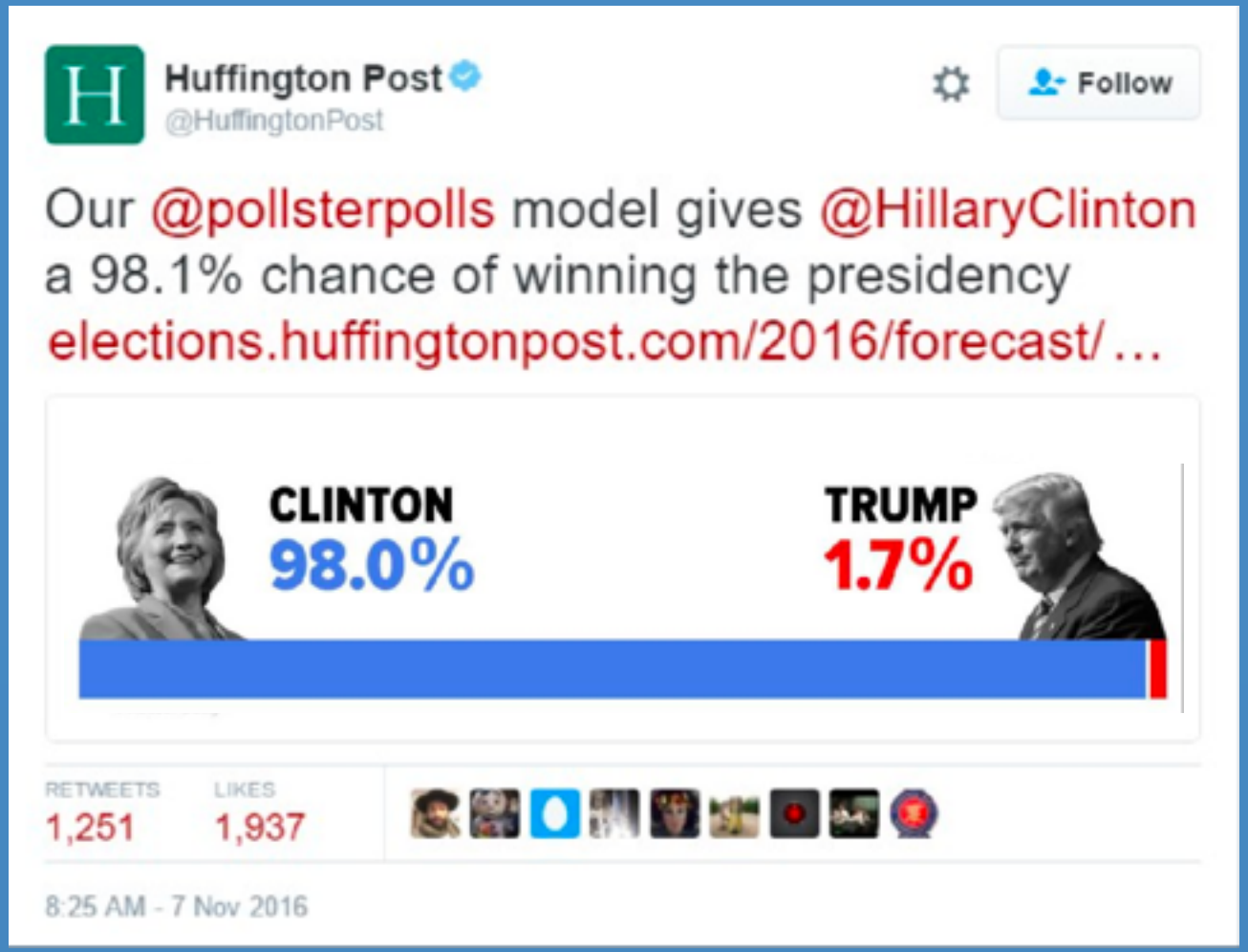
hypotheses	**axioms**
messy	**logical**
best guess	**certain**
discover	**prove**

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

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

We created a cargo cult
by confusing science
with math.

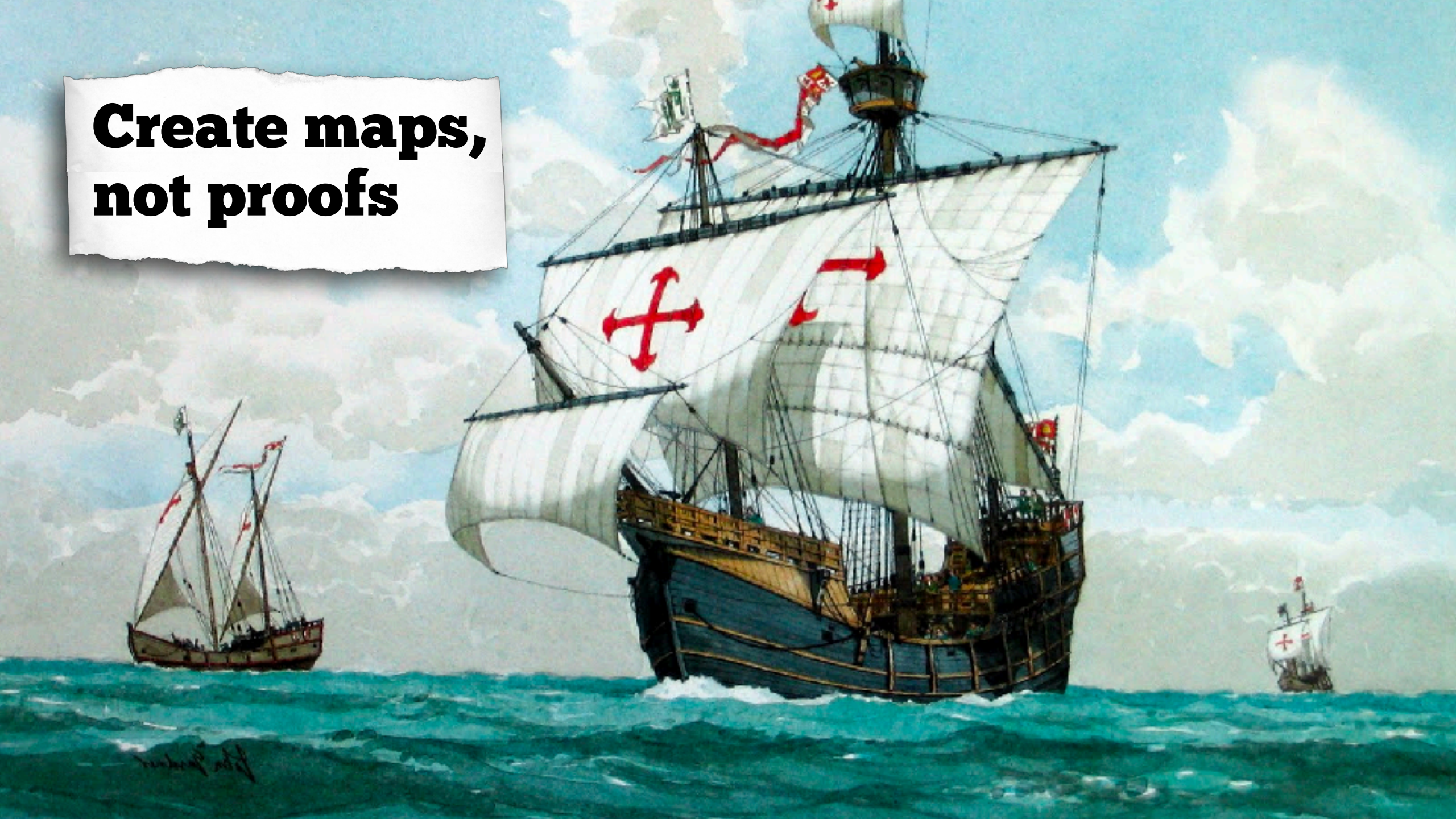
Now we must undo it.

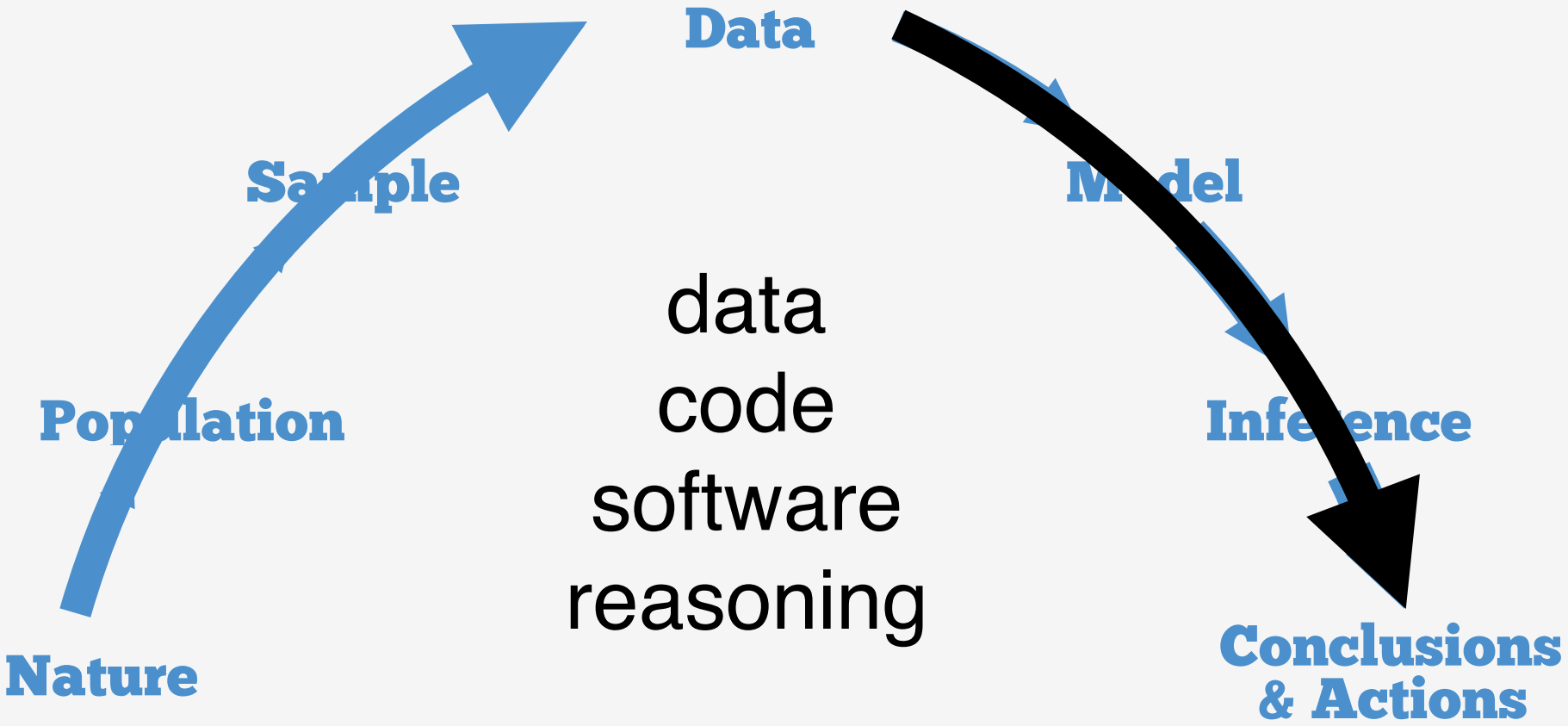

Huffington Post
@HuffingtonPost
Follow
Our @pollsterpolls model gives @HillaryClinton a 98.1% chance of winning the presidency
elections.huffingtonpost.com/2016/forecast/ …
CLINTON
98.0%
TRUMP
1.7%
RETWEETS
1,251
LIKES
1,937
8:25 AM - 7 Nov 2016

10 hours later…

rstudio::conf
AUSTIN

Create maps, not proofs





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


R Markdown is an authoring format for data science


Mythos
PREMIUM QUALITY
Mythos
HELLENIC BEER
750mL?

Dublin - Google Sheets

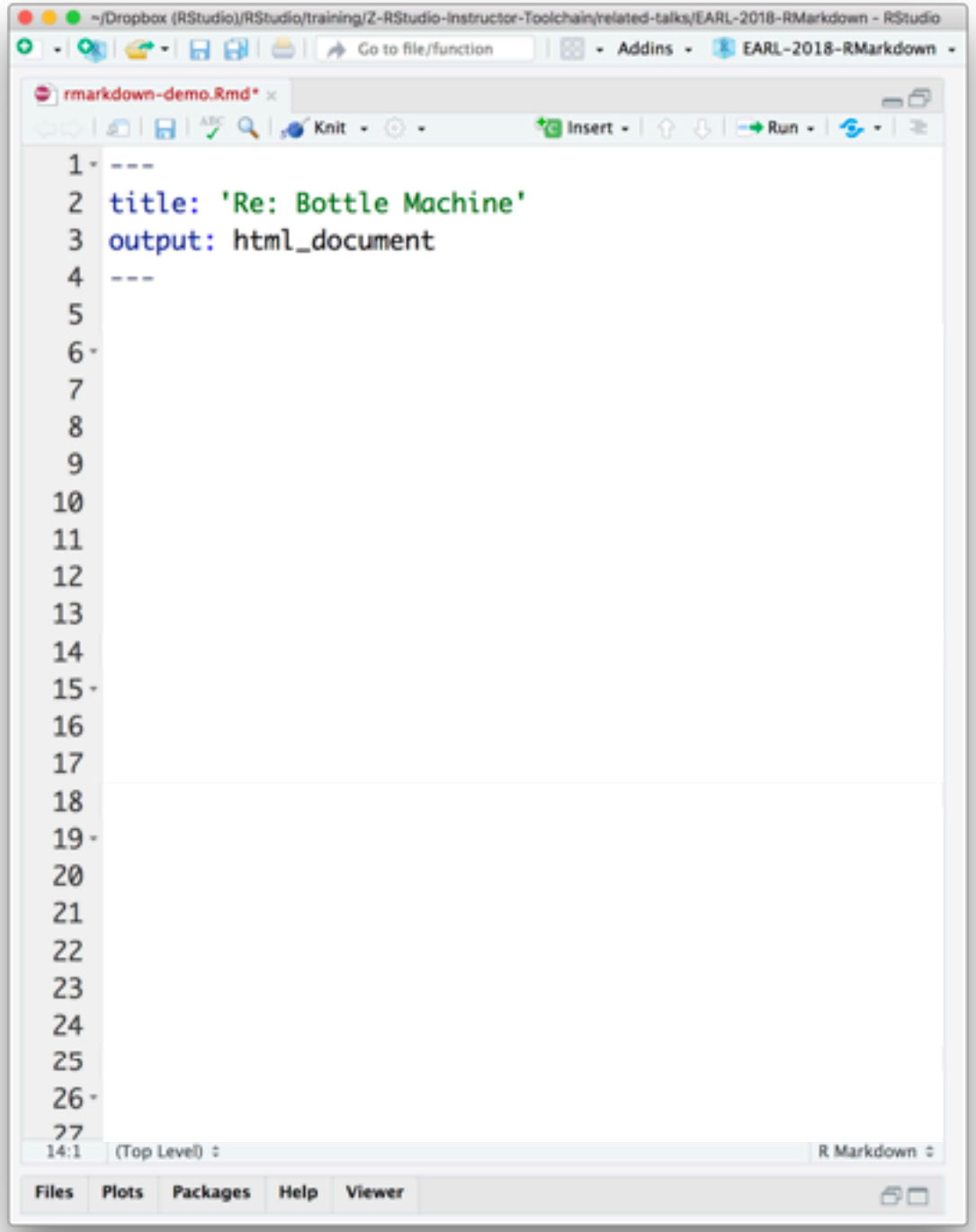
Dublin

File Edit View Insert Format Data

SHARE

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	748			
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 ---
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
   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

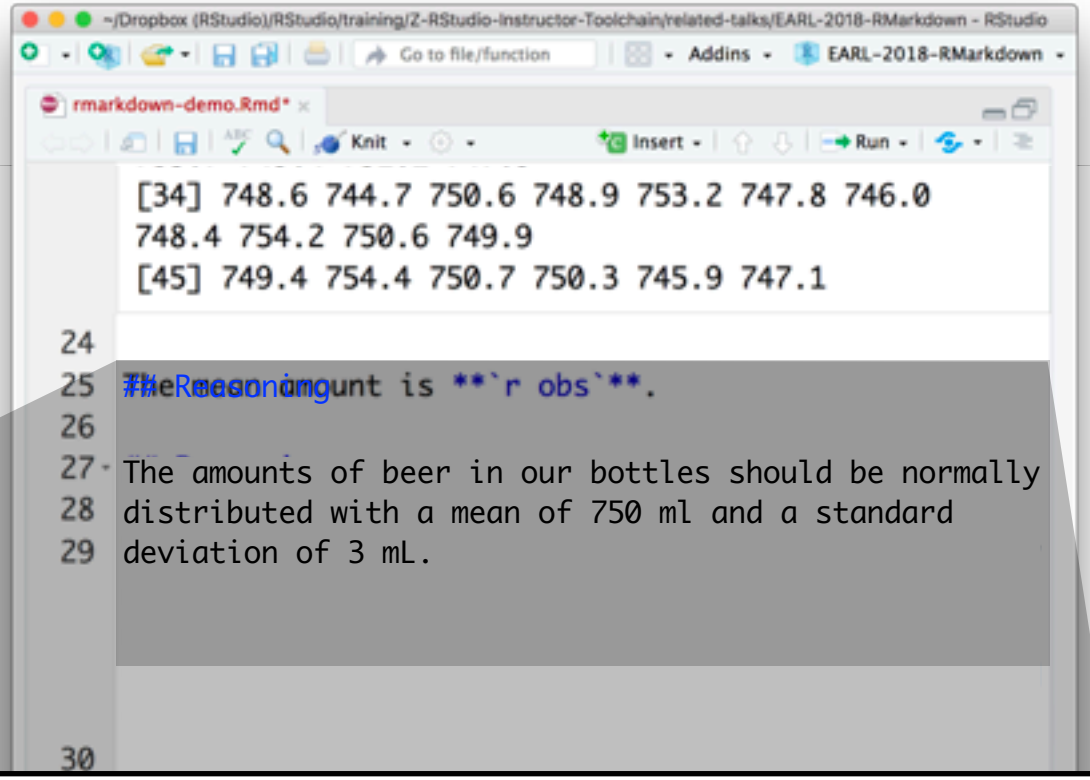
rmarkdown-demo.Rmd*

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  ```
```

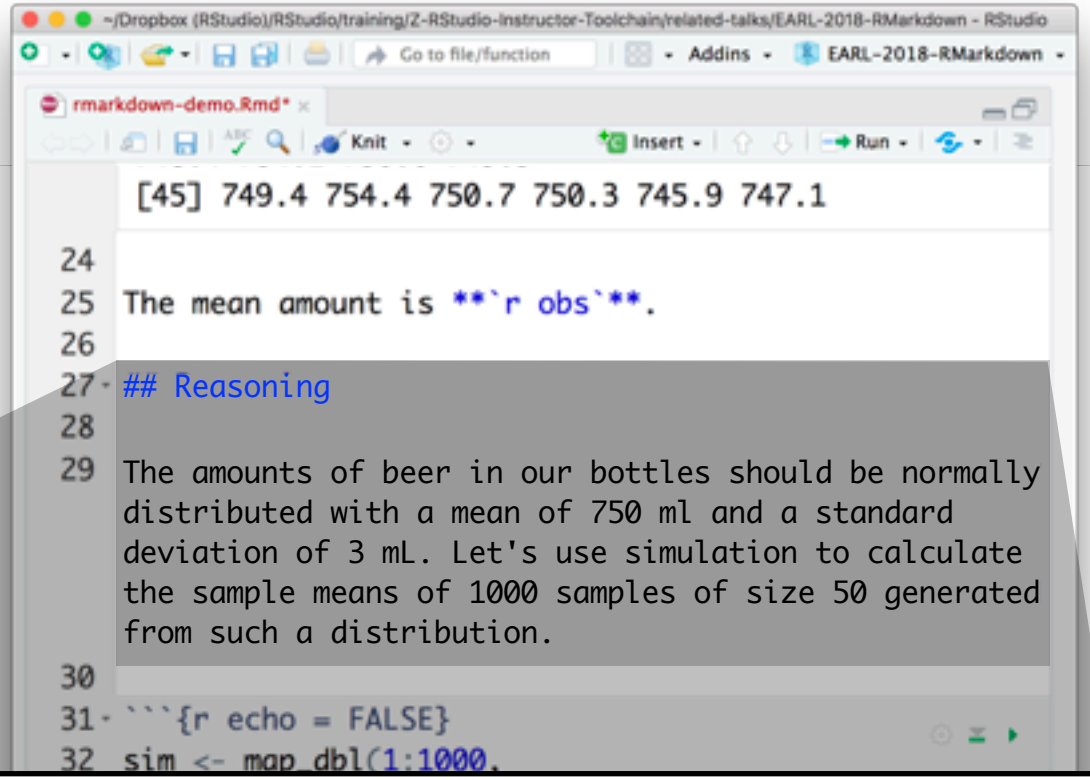
The mean amount is **`r obs`**.

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.



Reasoning

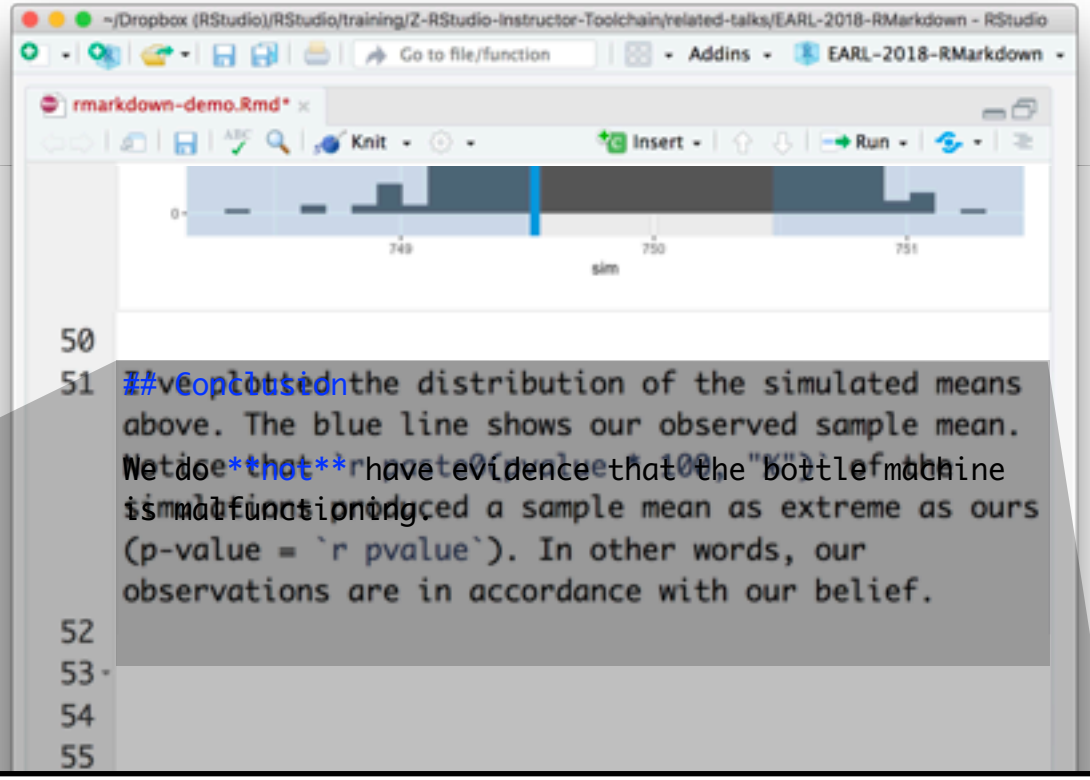
The amounts of beer in our bottles should be normally distributed with a mean of 750 ml and a standard deviation of 3 mL. Let's use simulation to calculate the sample means of 1000 samples of size 50 generated from such a distribution.

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

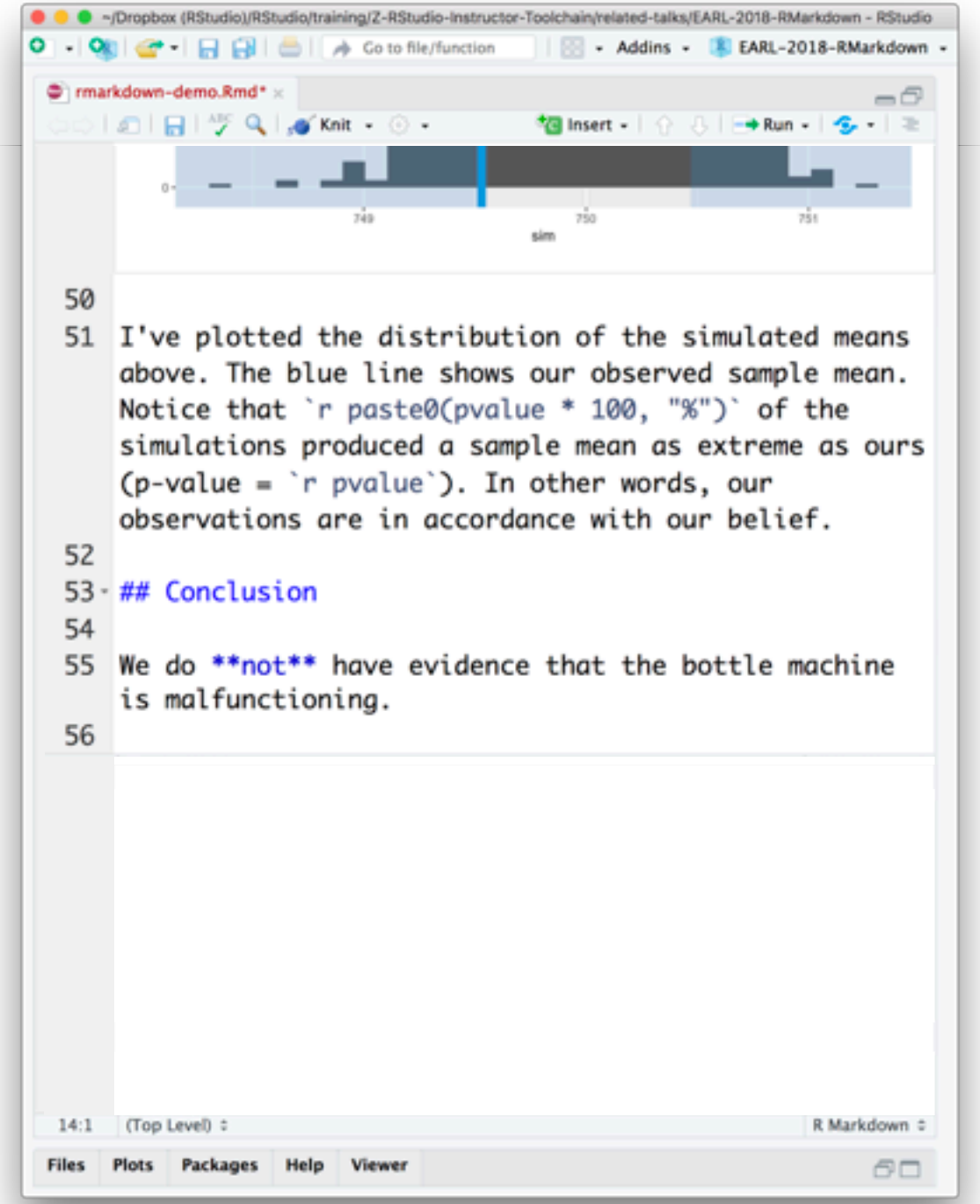
rmarkdown-demo.Rmd*

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

24
25 The mean amount is **`r obs`**.
26
27 I've plotted the distribution of the simulated means
28 above. The blue line shows our observed sample mean.
29 Notice that of the
   simulations produced a sample mean as extreme as ours
   (p-value = ). 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
0
749
750
751
sim
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

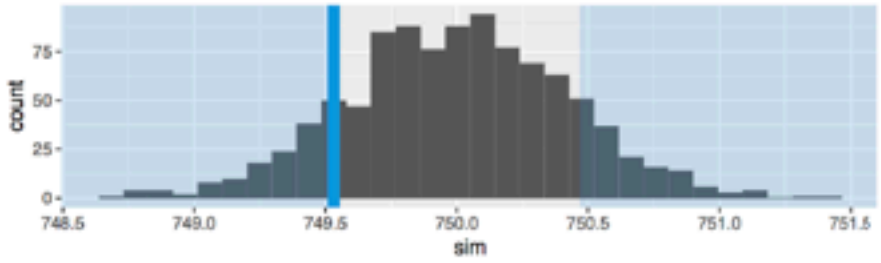
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [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

Re: Bottle Machine

Slide Layout

Title Slide

Change Master

Appearance

Title

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

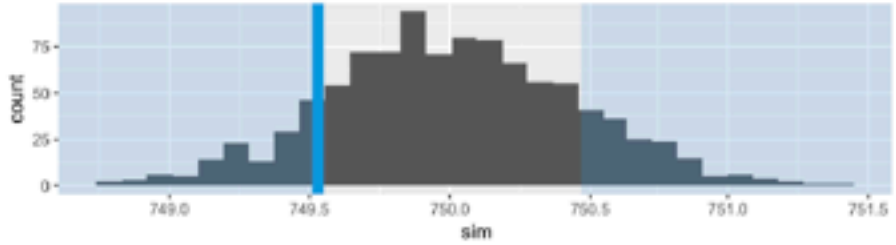
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [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

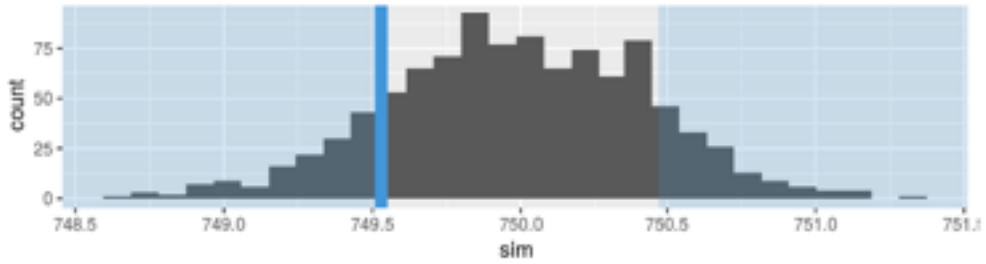
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 748.6 752.2 749.9 745.9 747.1
## [12] 748.3 750.4 749.6 746.0 750.5 750.5 750.1 747.9 750.6 758.0 746.3
## [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

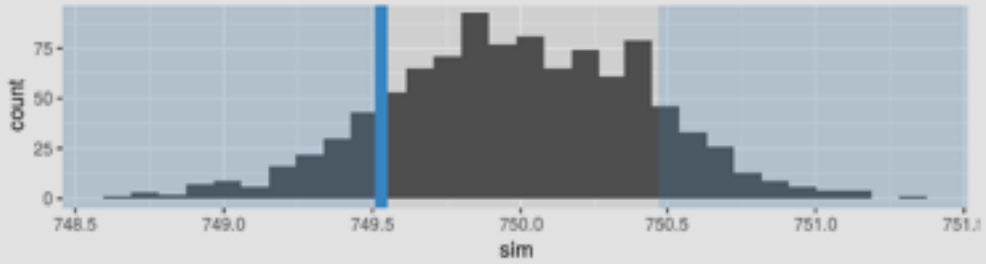
I selected 50 bottles at random from the Dublin factory, whic

```
##  [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

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

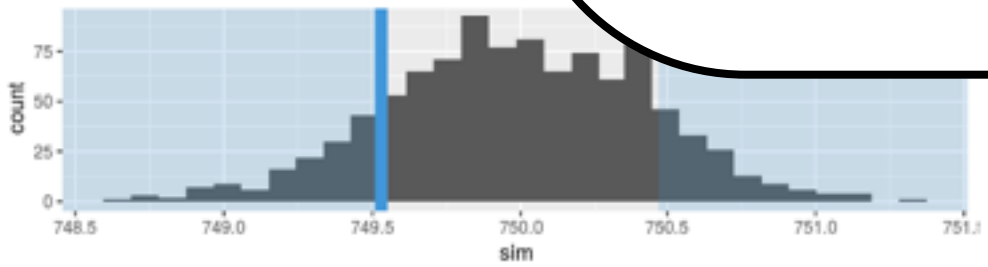
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [1] 748.7 746.4 750.1 751.9 747.6 748.8 748.6 7
## [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 dist
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
simulations produced a sample mean as extreme as ours (p-value = 0.274). In other words, our observations are in accordance w
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

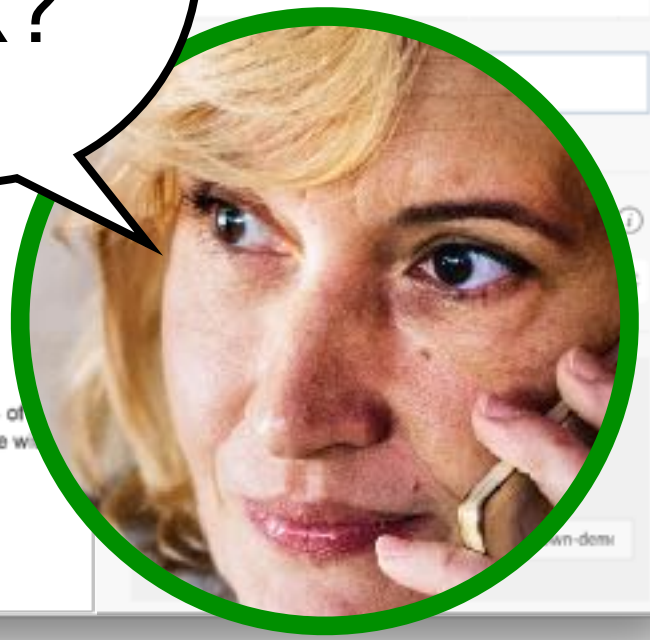
Collaborators & you

an change this document

William Gosset bill

VP Virginia virginia

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



Schedule saved successfully.

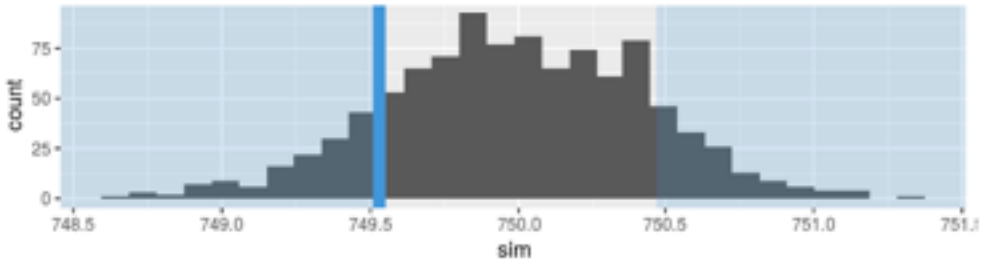
...hed the following amounts of beer (mL):

```
##  [1] 752.2 749.9 745.9 747.1
## [13] 741.9 750.6 758.0 746.3
## [25] 751.7 749.4 752.1 747.9
## [34] 746.0 748.4 754.2 750.6 749.9
## [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.

Garrett

WG bill

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

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

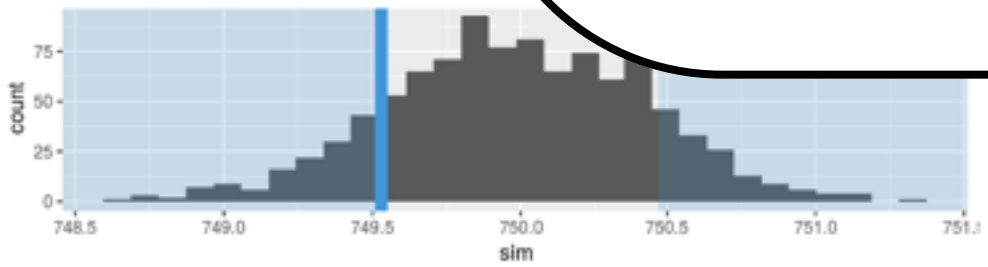
I selected 50 bottles at random from the Dublin factory, which c

```
##  [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
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
simulations produced a sample mean as extreme as ours (p-value = 0.274). In other words, our observations are in accordance w
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

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

Set to Now

le type

ly

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  ```{r setup, include=FALSE, message=FALSE}
7  library(googlesheets)
8  library(tidyverse)
9        - London
10 key <- "1OrHXGZJIyE_xyrbYs4YBElFUtmESYZE7n1rK5HuoR4w"
11 samples <- gs_read(gs_key(key), lookup = FALSE)
12 obs <- mean(samples$amount)
13 ```
```

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/160/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

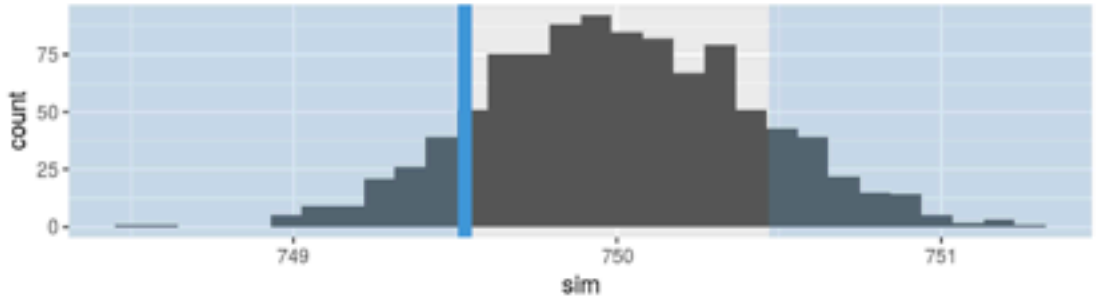
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [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/160/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

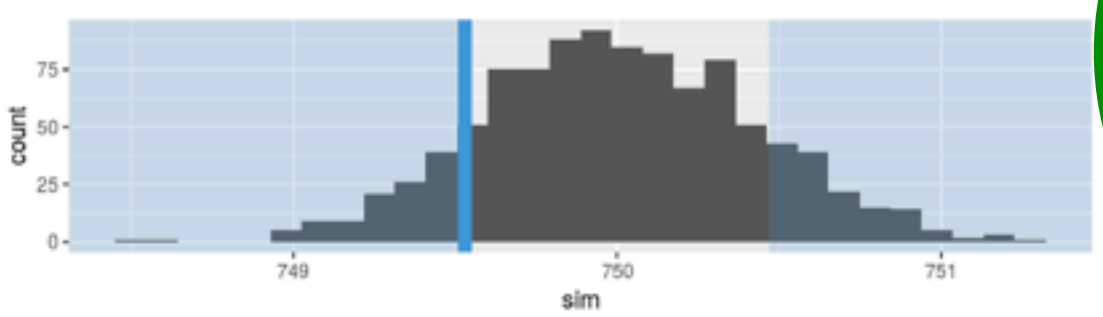
I selected 50 bottles at random from the Dublin factory, which contained the following amounts of beer (mL):

```
##  [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

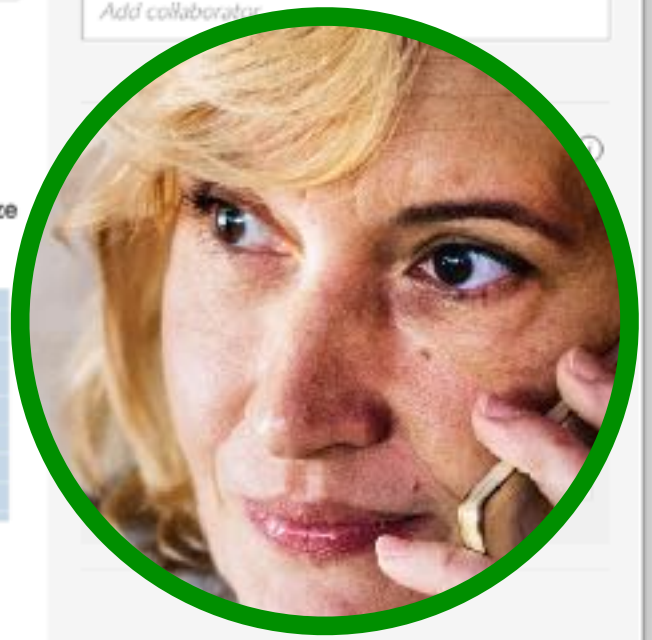
Who can view this document

You

Who can change this document

WG William Gosset bill

Add collaborator



RStudio Connect

Garrett

Secure https://connect.rstudioservices.com/connect/#/apps/160/access/106

Content / rmarkdown-demo

WG bill

Re: Bottle Mac

Does our bottle machine fill each bottle wi

Data

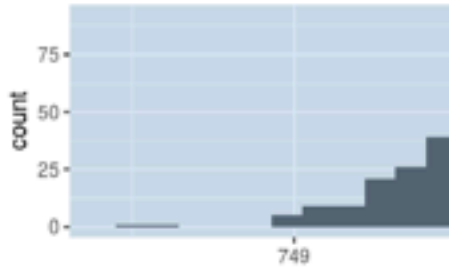
I selected 50 bottles at random from the D
(mL):

```
##  [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

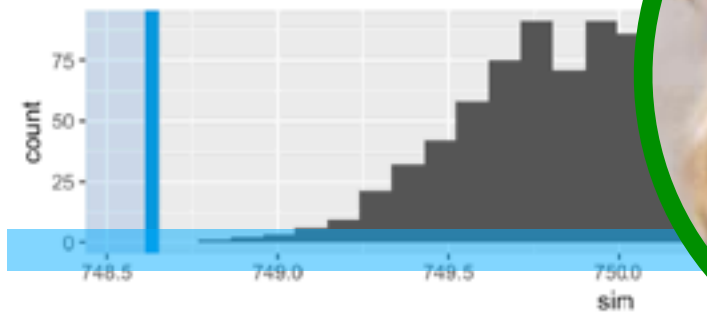
Does our bottle machine fill each bottle with 750 mL of beer (on average)?

Data

ned 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
```

The
of 3 mL. Let
distribution.



I've plotted the distribution of the simulated means above. The blue line sho
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?

rstudio::conf

Photo by Wil Stewart on Unsplash

Thank You