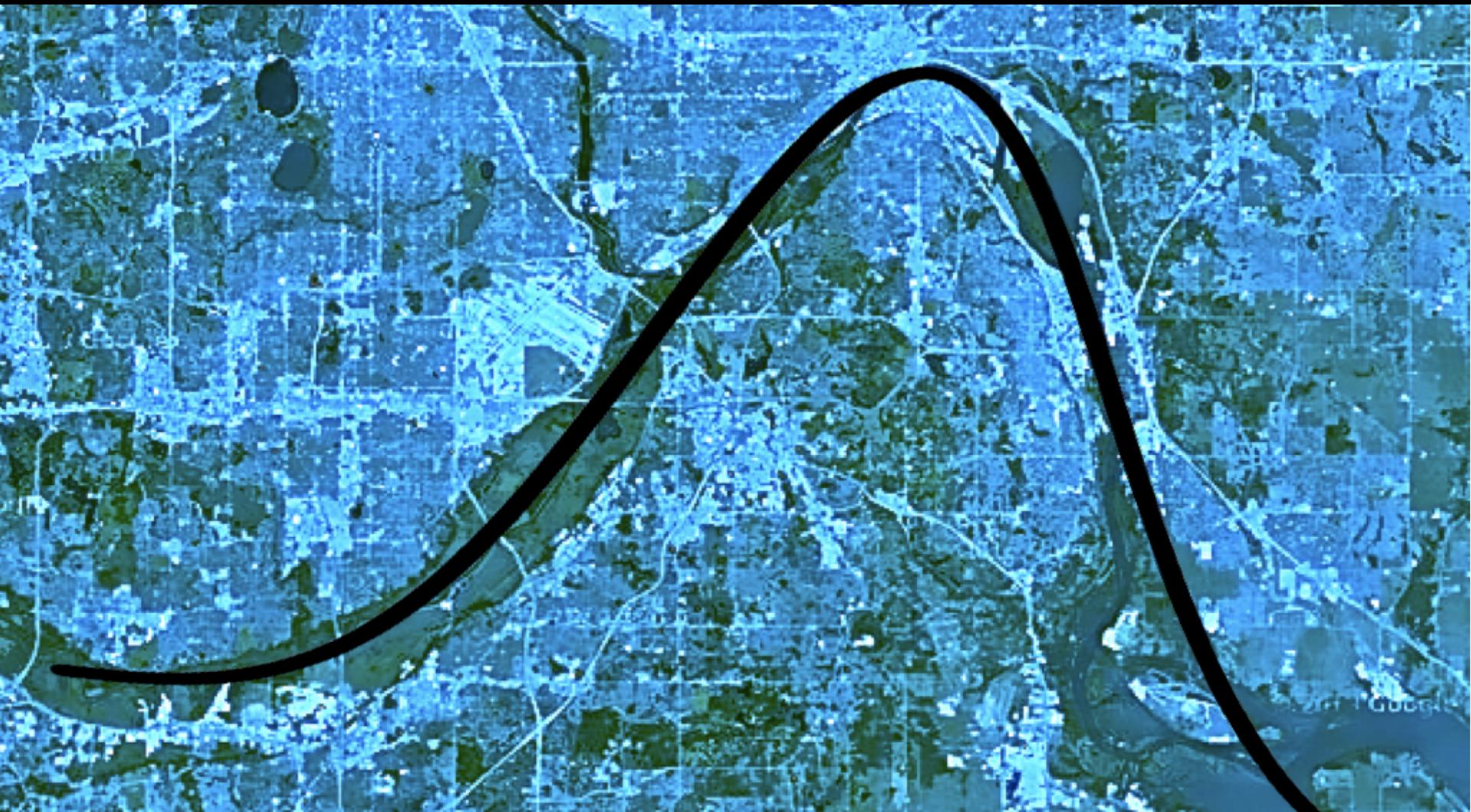


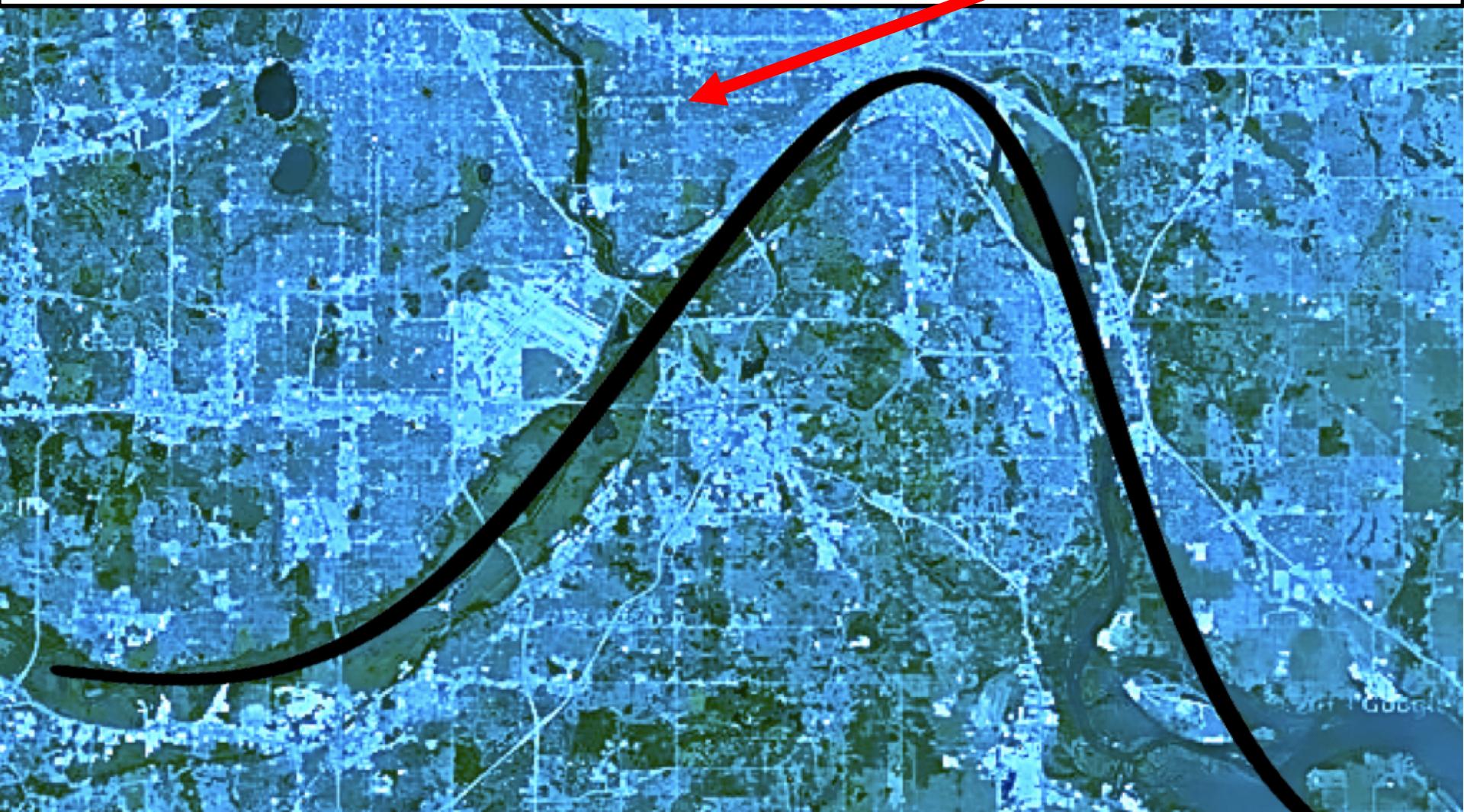
# Authoring & Utilizing Open Source, Reproducible Statistics/Data Science Textbooks

Alicia Johnson, Macalester College



# Authoring & Utilizing Open Source, Reproducible Statistics/Data Science Textbooks

Alicia Johnson, Macalester College



# 2013 office recreation



# introductory



[see definition of introductory](#)

*adj* preliminary, first

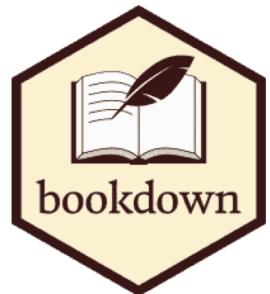
## More words related to **introductory**

### **elementary**

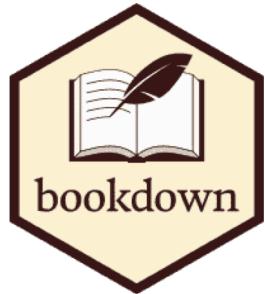
adjective. simple, basic

ABCs	easy	introductory	primitive	substratal
abecedarian	elemental	meat and potatoes	primo	uncomplex
basal	essential	original	rudimentary	uncomplicated
beginning	facile	plain	simplest	underlying
child's play	foundational	prefatory	simplex	
clear	fundamental	preliminary	simplified	
duck soup	initial	primary	straightforward	

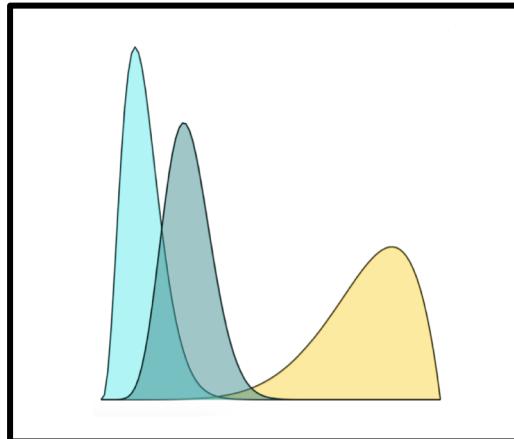
# BUILD



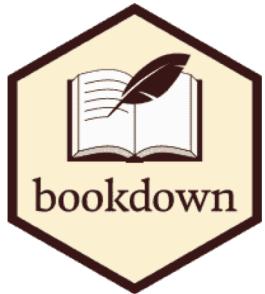
# BUILD



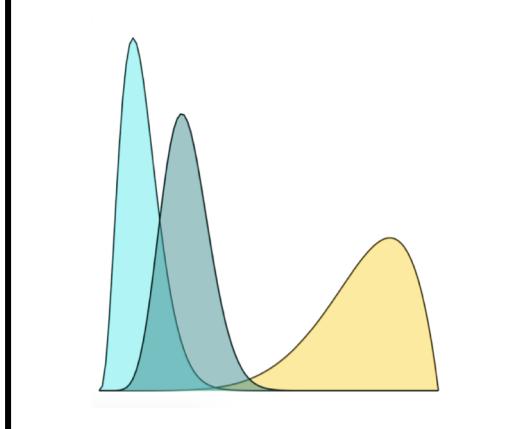
# DEVELOP



# BUILD



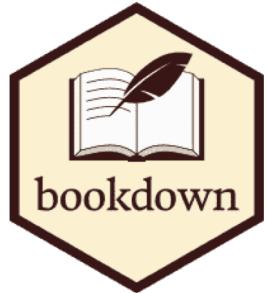
# DEVELOP



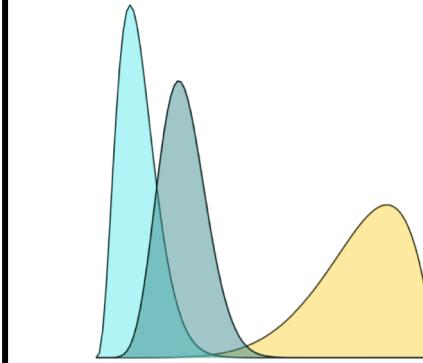
# DISTRIBUTE



# BUILD



# DEVELOP



# DISTRIBUTE



**reflect current best practices in statistical & modern data science workflow**

The R Series

# bookdown

Authoring Books and Technical  
Documents with R Markdown



Yihui Xie



CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

The R Series

# bookdown

Authoring Books and Technical  
Documents with R Markdown

R Markdown



Yihui Xie



CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

R Markdown

# A raw R Markdown (Rmd) file

A recent [fivethirtyeight.com](http://fivethirtyeight.com) article asked “What do men think it means to be a man?” \*\*Q: How much have you heard about the #MeToo movement?\*\*

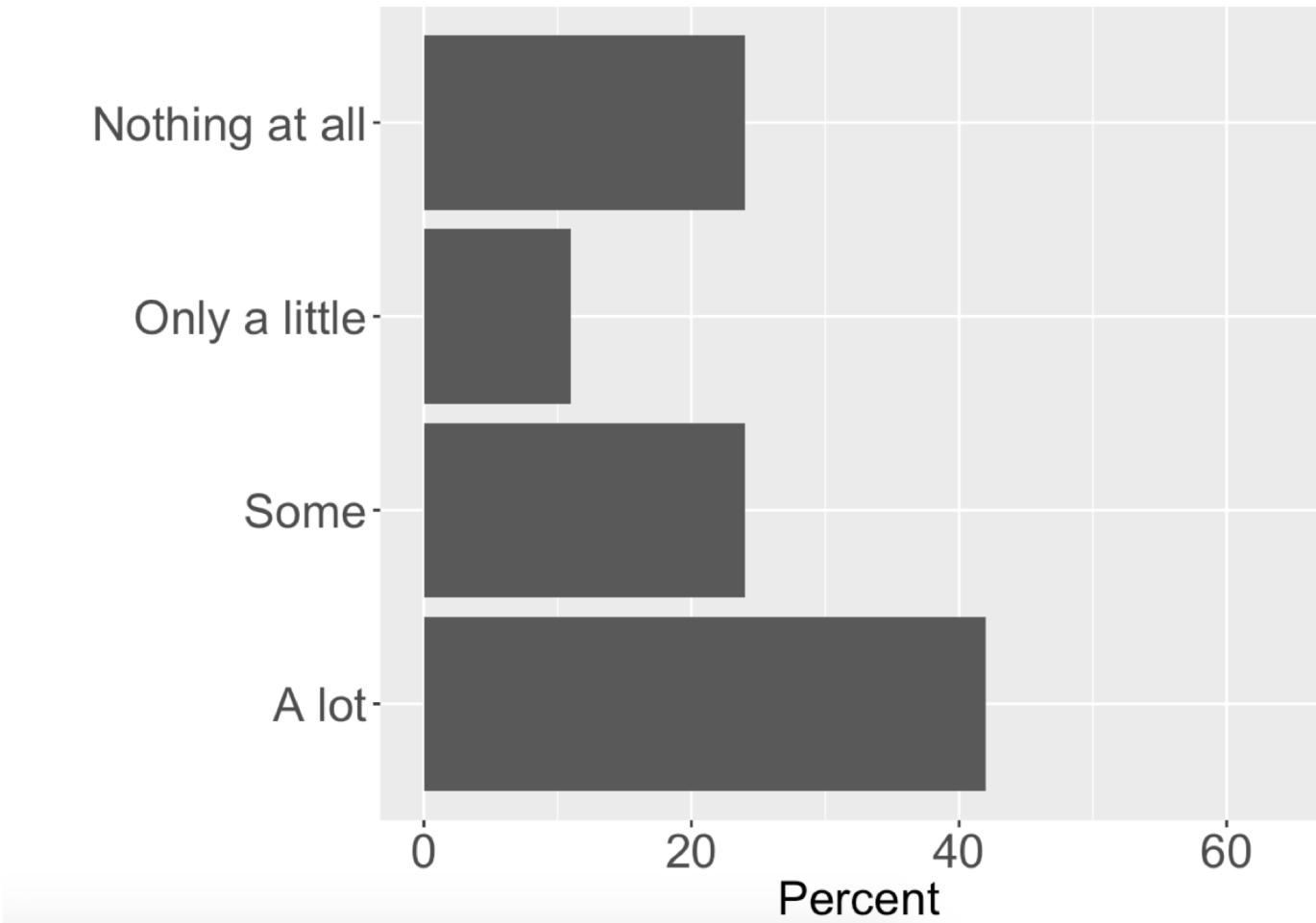
```
```{r echo = FALSE}
# Load data
survey <- read.csv("https://raw.github....masculinity-survey.csv")

# Plot survey results
ggplot(survey, aes(x = Response, y = Percent)) +
  geom_col() +
  coord_flip()
...```

```

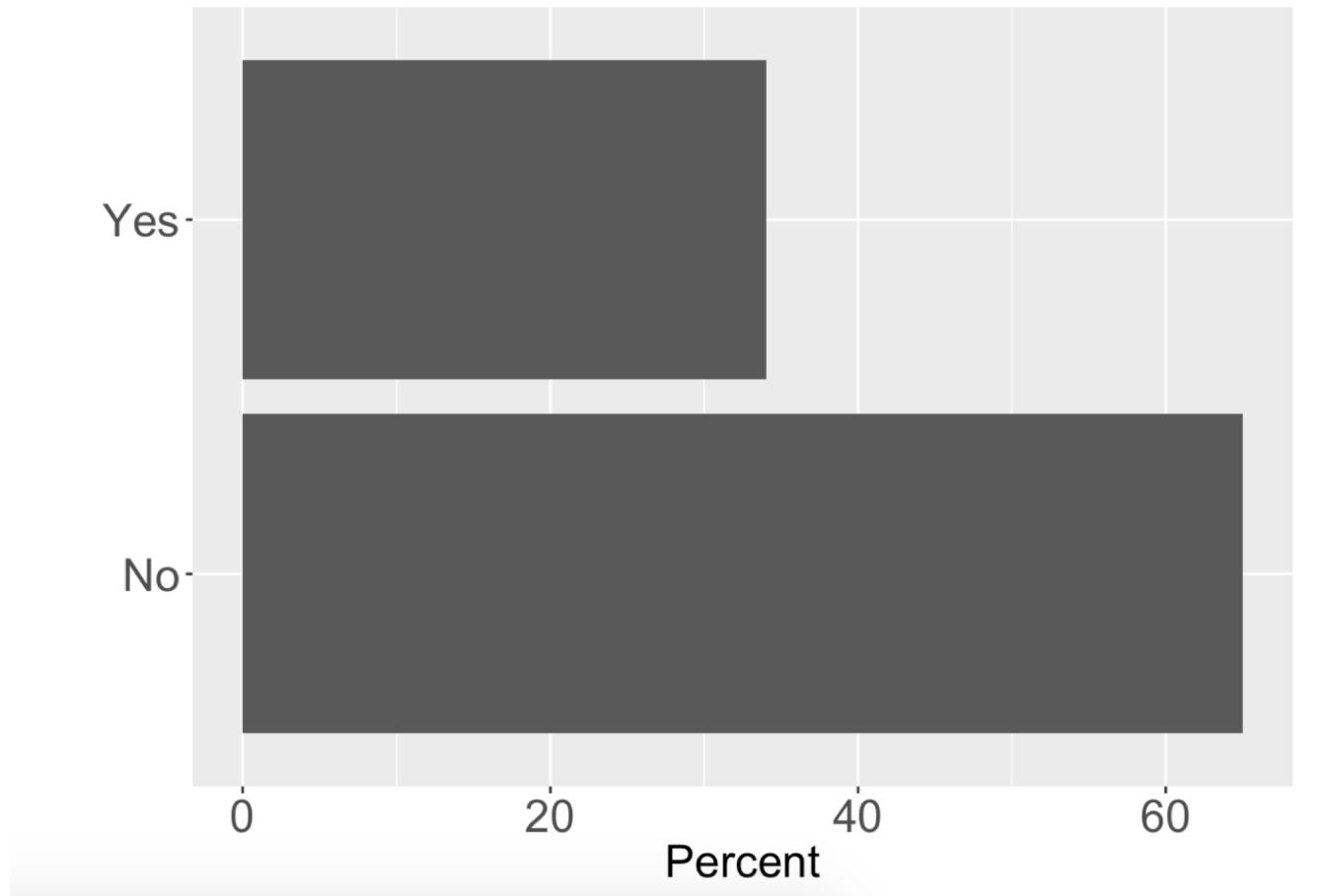
# A compiled R Markdown (Rmd) file

A recent [fivethirtyeight.com](https://fivethirtyeight.com/article/what-do-men-think-it-means-to-be-a-man/) article asked [What do men think it means to be a man?](#) **Q: How much have you heard about the #MeToo movement?**



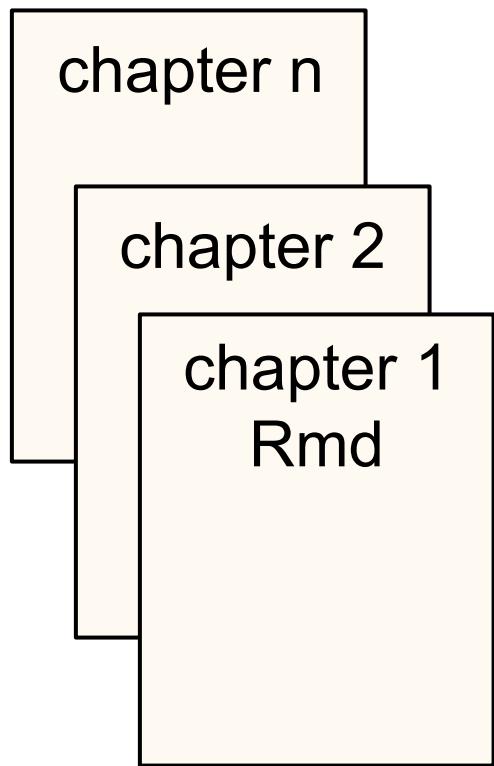
# A compiled R Markdown (Rmd) file

A recent [fivethirtyeight.com](#) article asked **What do men think it means to be a man? Q: Do you think about your behavior at work differently in the wake of #MeToo?**

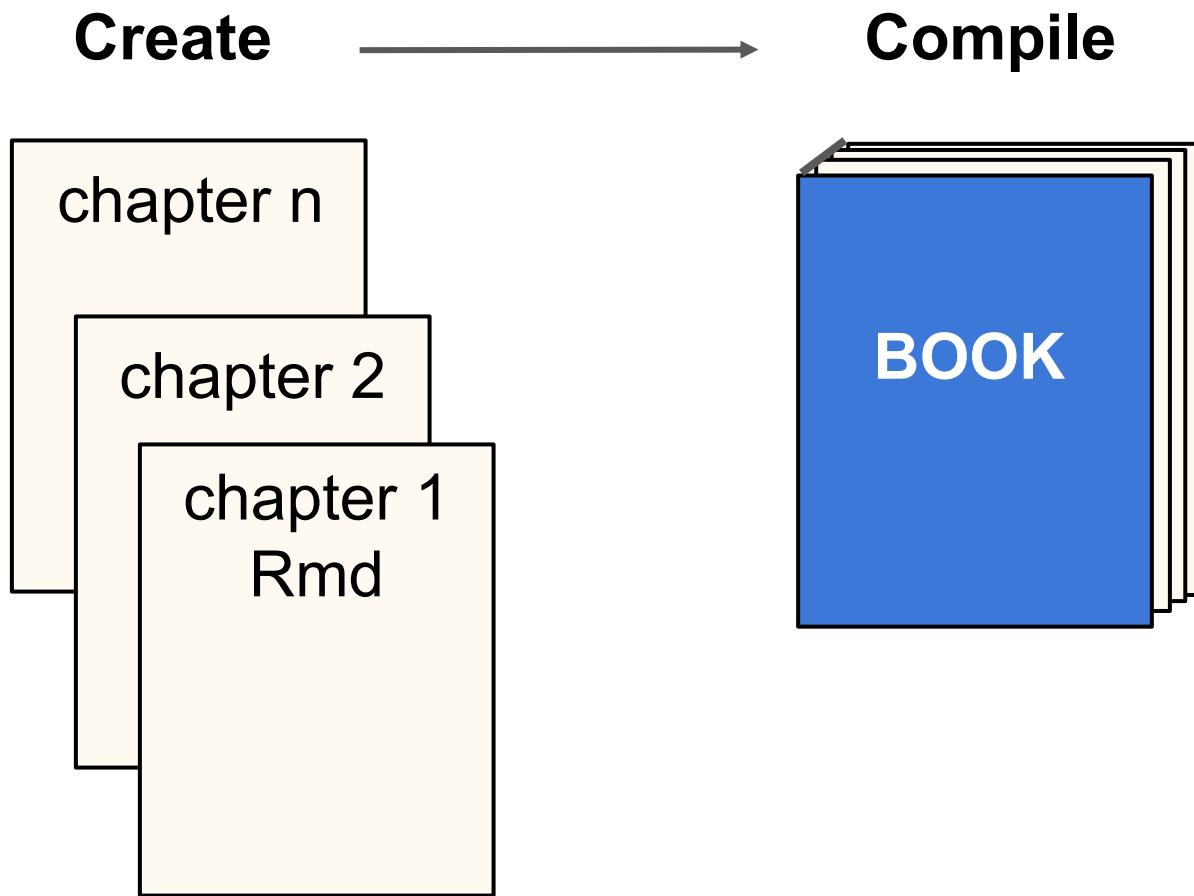


# A common bookdown workflow

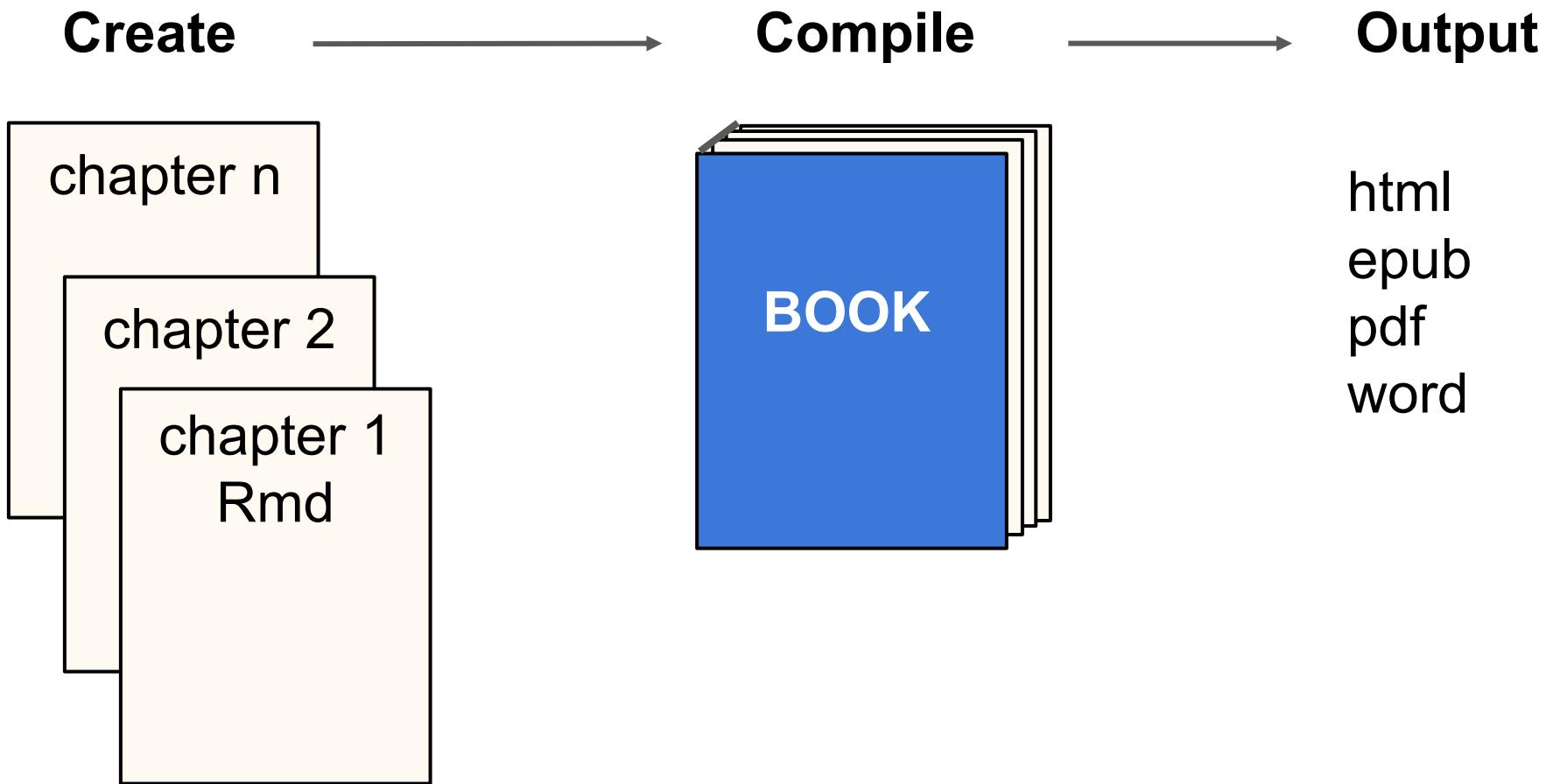
## Create



# A common bookdown workflow



# A common bookdown workflow



**1 Hello there**

**I Foundations**

**2 Bayesian Statistics?!**

**3 Modeling Foundations**

**4 Thinking Like a Bayesian**

**II Build, simulate, & analyze**

**5 The Beta-Binomial Bayesian Model**

**6 Posterior Simulation**

**7 Posterior Inference**

**III A family of Normal models**

**IV Generalizing the Normal model**

**V Hierarchical Bayesian models**

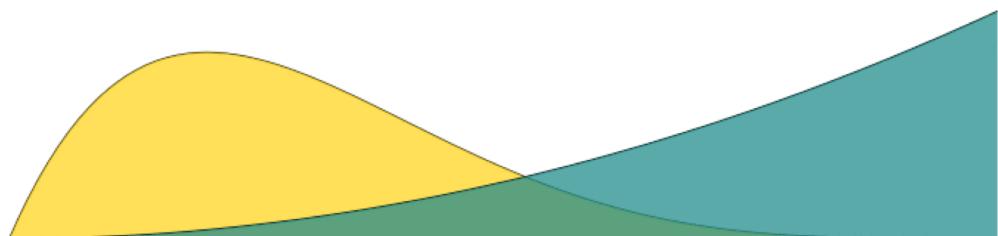
**VI Optional: MCMC algorithms**

**VII Appendices**

**8 References**

# **Project Bayes: an interactive introduction to Bayesian modeling with R**

***Alicia A. Johnson***



# DEVELOP

reflect modern best practices in statistics education

# DEVELOP

reflect modern best practices in statistics education

- emphasize conceptual understanding!
- promote data-driven inquiry!
- incorporate technology (eg: R)!

# DEVELOP

reflect modern best practices in statistics education

- emphasize conceptual understanding!
- promote data-driven inquiry!
- incorporate technology (eg: R)!
- **foster active learning!**

# Learn by doing: R

## 6.2 Markov chains via RJAGS

Let's examine these 3 steps in the context of our Bayesian election model. In Step 1, we *define* the structure of the likelihood & prior pieces of this model. Recall that the likelihood structure is Binomial and the prior structure is Beta:

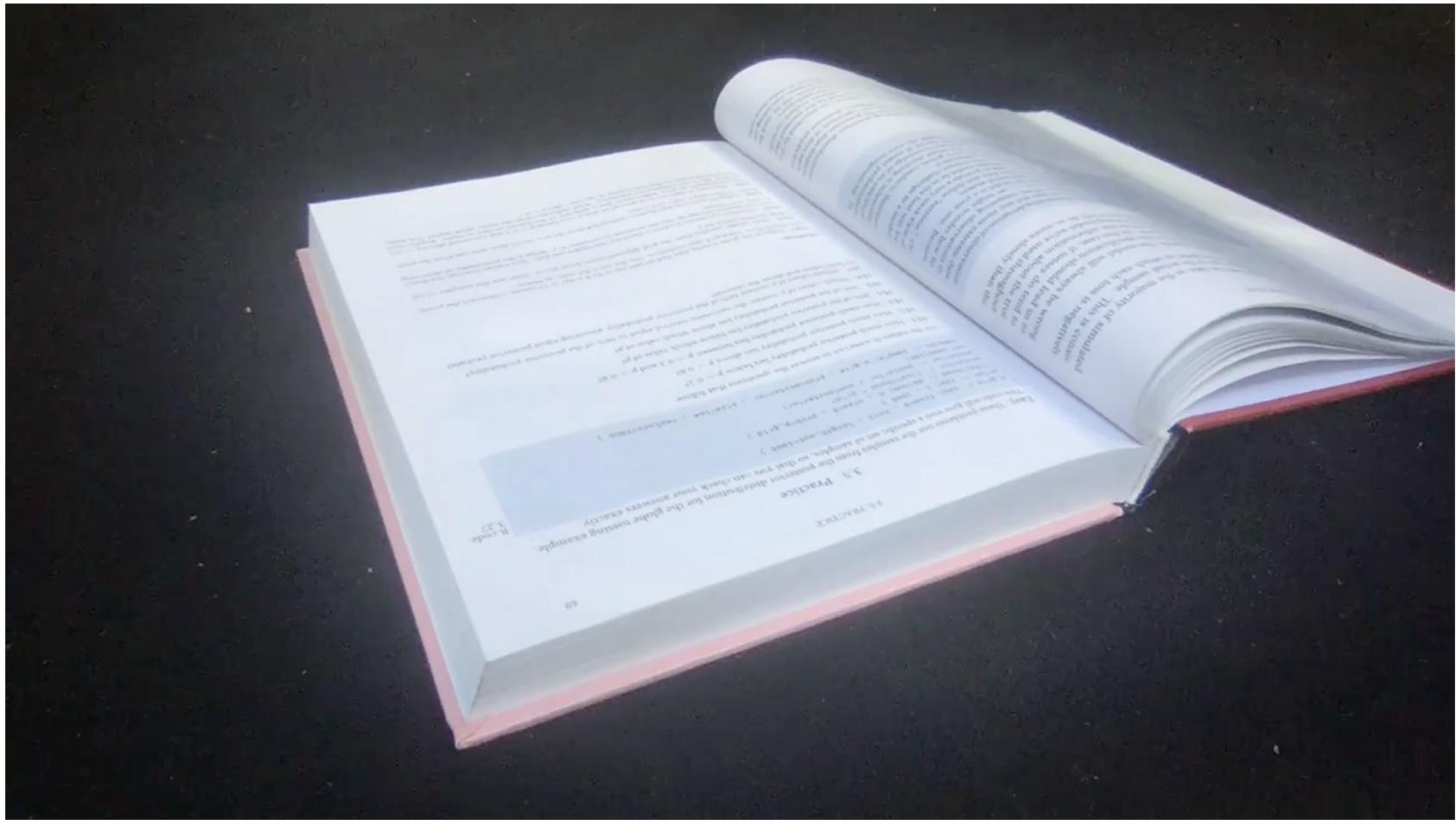
$$\begin{aligned} X|\theta &\sim \text{Bin}(n, \theta) \\ \theta &\sim \text{Beta}(\alpha, \beta) \end{aligned}$$

Translating this mathematical notation into `rjags` syntax, we define this model structure in `R` as a *model string* stored as `vote_model`:

```
# DEFINE the model
vote_model <- "model{
  # Likelihood model for X
  X ~ dbin(theta, n)

  # Prior model for theta
  theta ~ dbeta(alpha, beta)
}"
```

# Try to learn by doing...



# Really learn by doing: DataCamp Light

script.R R Console

```
1 # Add 2 and 3
2 |           +
3
4 # Calculate the square root of 16
5
```

**Solution**

**Run**



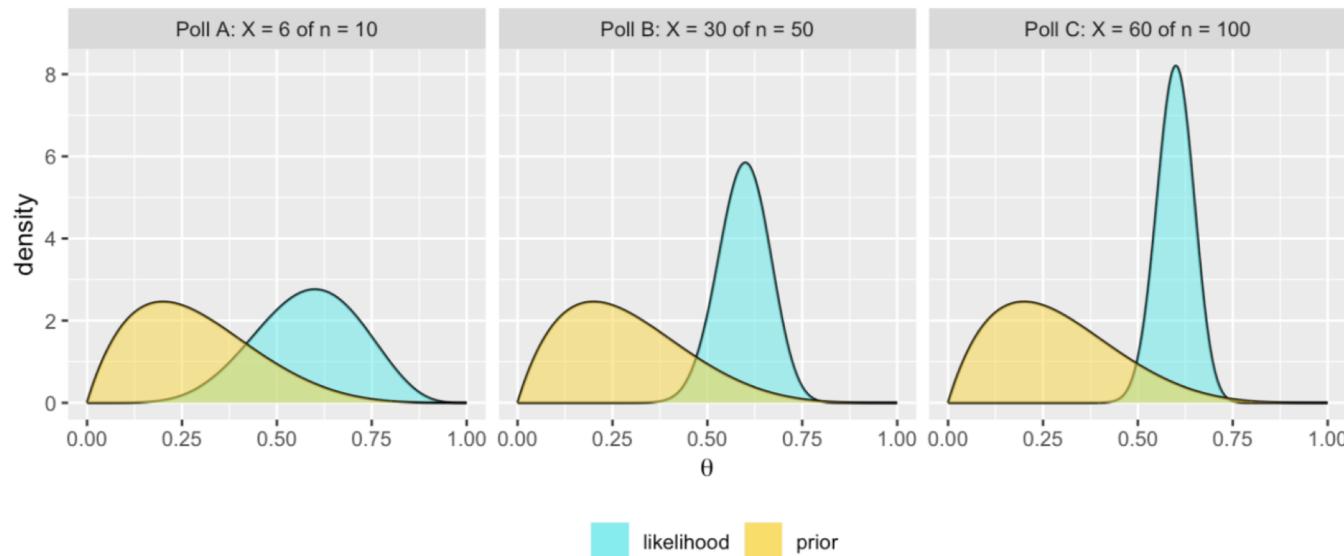
Setting Up Workspace

Powered by DataCamp 

# Stop and think: code folding

## Gut check 2

Suppose 3 different analysts start with the same Beta(2,5) prior for your election support  $\theta$ , but have access to 3 different polls. In each poll, 60% of the polled voters support you ( $X/n = 0.6$ ). However, the polling sample sizes  $n$ , thus the analyst's likelihood functions differ. In turn, each analyst will have a different posterior model of  $\theta$ . Which analyst's posterior do you anticipate will look the most like the likelihood, ie. will be "swayed" the most by the polling results? Explain.

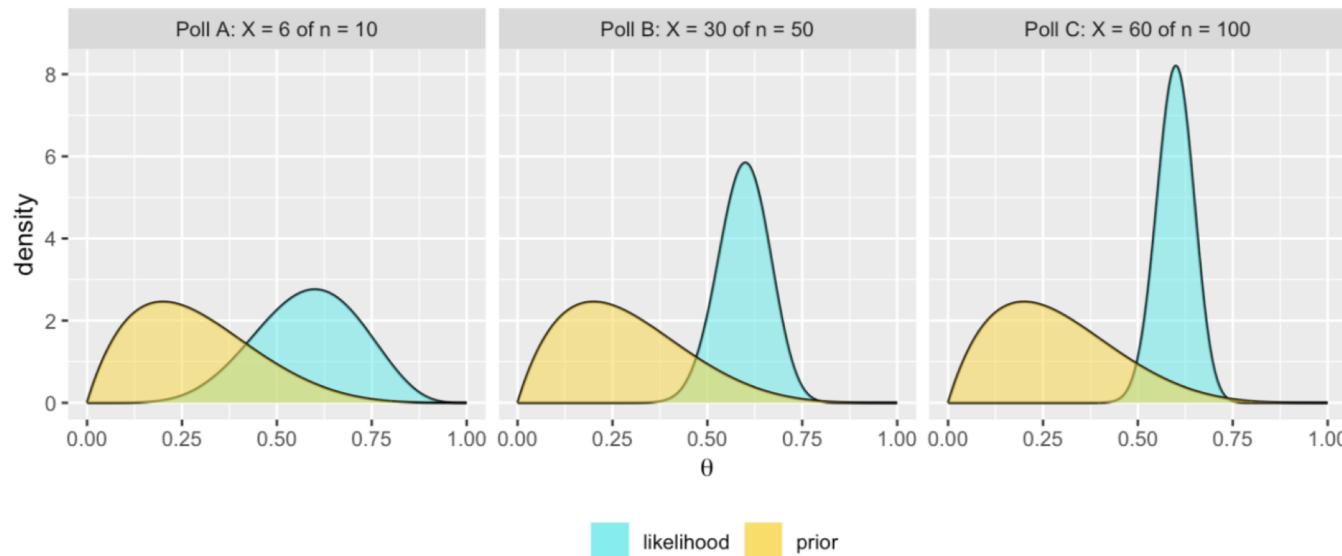


## Answer

# Stop and think: code folding

## Gut check 2

Suppose 3 different analysts start with the same Beta(2,5) prior for your election support  $\theta$ , but have access to 3 different polls. In each poll, 60% of the polled voters support you ( $X/n = 0.6$ ). However, the polling sample sizes  $n$ , thus the analyst's likelihood functions differ. In turn, each analyst will have a different posterior model of  $\theta$ . Which analyst's posterior do you anticipate will look the most like the likelihood, ie. will be "swayed" the most by the polling results? Explain.



Answer

show

# Stop and think: code folding

As expected, the posterior is a balance of the prior and likelihood - in this case, it's slightly “closer” to the prior than to the likelihood. We'll soon discuss why this is intuitive. For now, notice that the posterior also looks like it could be modeled by a Beta with updated shape parameters (the current section title is also a dead giveaway). This is indeed the case! In fact, conditioned on the polling data  $X = 30$ , the updated or posterior model of  $\theta$  is Beta(75,75):

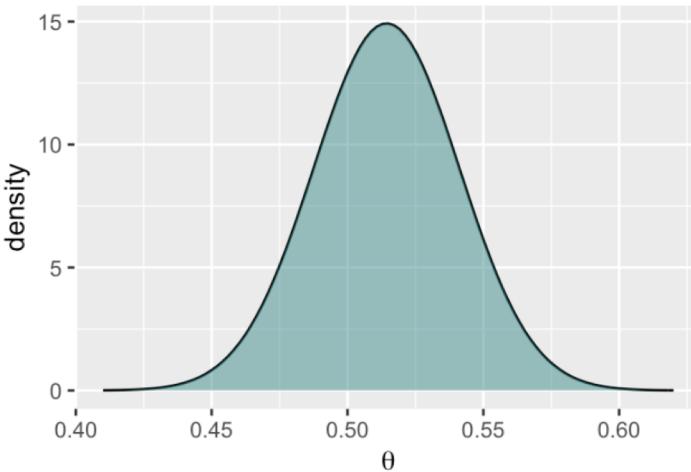
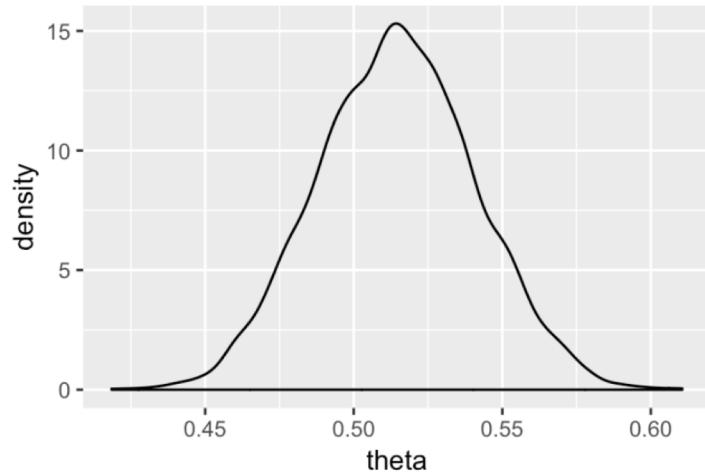
$$\theta | (X = 30) \sim Beta(75, 75)$$

**Extra: proof of Beta posterior**

show

# Stop and think: code folding

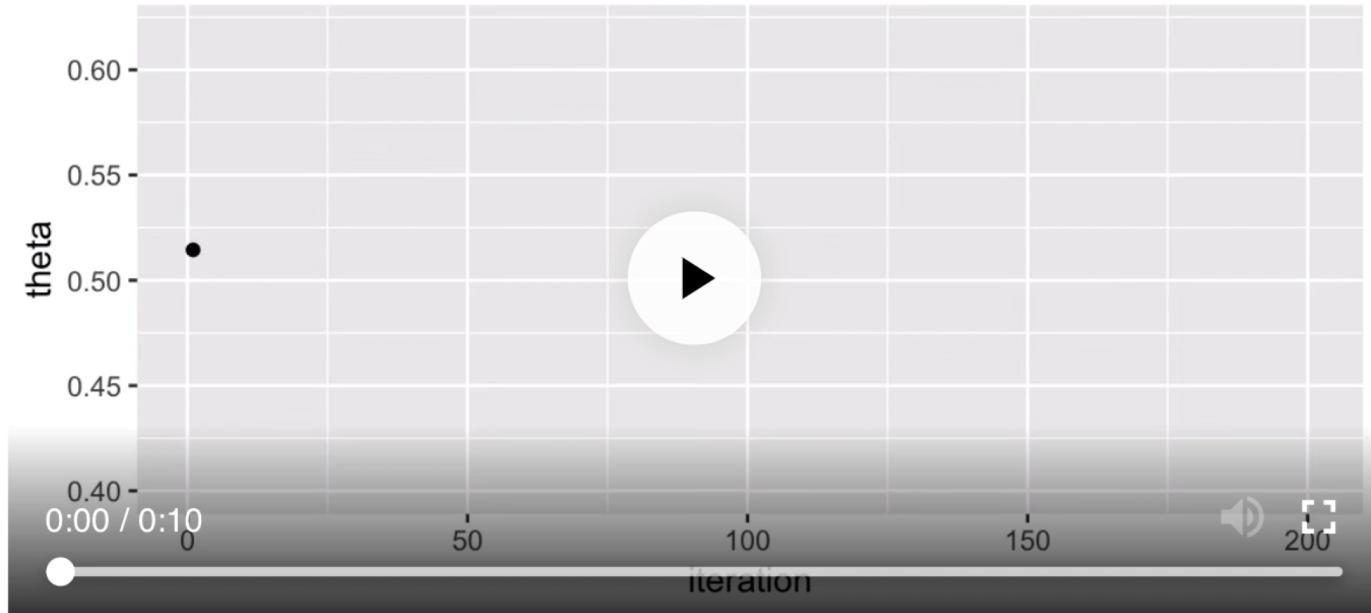
The hard part is over. Once we have the Markov chain simulation results in hand, it's quite intuitive to use them to conduct posterior inference. Simply put, we can estimate any feature of the  $\text{Beta}(180, 170)$  posterior model (right) by the corresponding feature of the Markov chain distribution (left):



**Plot code**

# Experience: video & animation

After 200 iterations, the Markov chain has started to explore new territory, traversing a slightly wider range of values between 0.45 and 0.58:

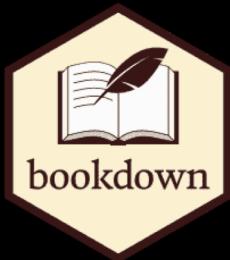


# DISTRIBUTE

reflect modern data science workflow

# DISTRIBUTE

reflect modern data science workflow



+ **GitHub** →

reproducible  
open-source  
agile

# Non-reproducible

My Drive > ComputerDocuments > Statistics > Research > GraduateSchool > UMN > Thesis ▾

Name ↑	Owner	Last opened by me	File size
 acf1.eps	me	Jan 19, 2017	4 KB
 acf1.jpg	me	Jan 19, 2017	14 KB
 acf1.pdf	me	Jan 19, 2017	3 KB
 acf1.ps	me	Jan 19, 2017	4 KB
 Background.tex	me	Jan 19, 2017	42 KB
 CLT.tex	me	Jan 19, 2017	11 KB
 d1d2.eps	me	Jan 19, 2017	4 KB
 d1d2.pdf	me	Jan 19, 2017	4 KB
 drift1.eps	me	Jan 19, 2017	7 KB
 drift1.pdf	me	Jan 19, 2017	7 KB

# Reproducible

A recent [fivethirtyeight.com](https://fivethirtyeight.com) article asked “What do men think it means to be a man?” \*\*Q: How much have you heard about the #MeToo movement?\*\*

```
```{r echo = FALSE}
# Load data
survey <- read.csv("https://raw.githubusercontent.com/fivethirtyeight/data/master/masculinity-survey.csv")

# Plot survey results
ggplot(survey, aes(x = Response, y = Percent)) +
  geom_col() +
  coord_flip()
...```

```

# Distributing open-source materials on GitHub

ajohns24 / project\_bayes Private

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Branch: master project\_bayes / docs /

Create new file Upload files Find file History

ajohns24 try again Latest commit ab885d8 37 seconds ago

..

File	Commit Message	Time
bayes_book_files/figure-html	try again	36 seconds ago
images	all current chapters	6 minutes ago
libs	all current chapters	6 minutes ago
beta-binomial.html	try again	36 seconds ago
hello-there.html	try again	36 seconds ago
index.html	all current chapters	6 minutes ago
introduction.html	try again	36 seconds ago
modeling-foundations.html	try again	36 seconds ago
posterior-inference.html	try again	36 seconds ago
posterior-simulation.html	try again	36 seconds ago

# Open-source: open access



# Open-source: open access, customizable

ajohns24 / project\_bayes Private X

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Branch: master project\_bayes / docs / Create new file Upload files Find file History

ajohns24 try again Latest commit ab885d8 37 seconds ago

..

File	Status	Time
bayes_book_files/figure-html	try again	36 seconds ago
images	all current chapters	6 minutes ago
libs	all current chapters	6 minutes ago
beta-binomial.html	try again	36 seconds ago
hello-there.html	try again	36 seconds ago
index.html	all current chapters	6 minutes ago
introduction.html	try again	36 seconds ago
modeling-foundations.html	try again	36 seconds ago
posterior-inference.html	try again	36 seconds ago
posterior-simulation.html	try again	36 seconds ago

# Open-source: open access, customizable

ajohns24 / project\_bayes Private 

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Branch: master  project\_bayes / docs / Create new file Upload files Find file History

ajohns24 try again Latest commit ab885d8 37 seconds ago

..

File	Status	Time
bayes_book_files/figure-html	try again	36 seconds ago
images	all current chapters	6 minutes ago
libs	all current chapters	6 minutes ago
beta-binomial.html	try again	36 seconds ago
hello-there.html	try again	36 seconds ago
index.html	all current chapters	6 minutes ago
introduction.html	try again	36 seconds ago
modeling-foundations.html	try again	36 seconds ago
posterior-inference.html	try again	36 seconds ago
posterior-simulation.html	try again	36 seconds ago

# Open-source: open access, customizable

ajohns24 / project\_bayes Private

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Branch: master [project\\_bayes / docs /](#)

Create new file Upload files Find file History

ajohns24 try again Latest commit ab885d8 37 seconds ago

..

File	Status	Time
bayes_book_files/figure-html	try again	36 seconds ago
images	all current chapters	6 minutes ago
libs	all current chapters	6 minutes ago
beta-binomial.html	try again	36 seconds ago
hello-there.html	try again	36 seconds ago
index.html	all current chapters	6 minutes ago
introduction.html	try again	36 seconds ago
modeling-foundations.html	try again	36 seconds ago
posterior-inference.html	try again	36 seconds ago
posterior-simulation.html	try again	36 seconds ago

# Open-source: open access, customizable

ajohns24 / project\_bayes Private

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Branch: master [project\\_bayes / docs /](#)

Create new file Upload files Find file History

ajohns24 try again Latest commit ab885d8 37 seconds ago

..

bayes\_book\_files/figure-html

images

libs

beta-binomial.html

hello-there.html

index.html

introduction.html

modeling-foundations.html

posterior-inference.html

posterior-simulation.html

Add tailored example

Commit Message	Time Ago
try again	36 seconds ago
all current chapters	6 minutes ago
all current chapters	6 minutes ago
try again	36 seconds ago
try again	36 seconds ago
all current chapters	6 minutes ago
try again	36 seconds ago
try again	36 seconds ago
try again	36 seconds ago
try again	36 seconds ago
try again	36 seconds ago

# Open-source: open access, customizable, collaborative

The screenshot shows a GitHub repository page. At the top, the repository name is 'mac-comp112 / 112-course-development' with a 'Private' button. To the right are 'Unwatch' (with a count of 4), 'Star' (with a count of 0), and 'Fork' (with a count of 0) buttons. Below this is a navigation bar with links for 'Code', 'Issues 0', 'Pull requests 0', 'Projects 0', 'Wiki', 'Insights', and 'Settings'. A red arrow points from the top right towards the 'Contributors' section.

No description, website, or topics provided.

Add topics

Edit

65 commits 2 branches 0 releases 5 contributors

Branch: gh-pages New pull request Create new file Upload files Find file Clone or download

ajohns24	publish: final directions	...	Latest commit 1caf410 on Dec 11, 2017
CompMath112book_files/figure-html	publish: final directions		7 months ago
images	publish: updated 494 text analysis typos		9 months ago
libs	publish: update checkpoint2		8 months ago
CompMath112book.epub	publish: final directions		7 months ago
activities-acquiring-data.html	publish: final directions		7 months ago
activities-classification-prediction.html	publish: final directions		7 months ago
activities-eda.html	publish: final directions		7 months ago

# Open-source: open access, customizable, collaborative

[Home](#)

**1** Syllabus

**2** Activities: Introductory

**3** Activities: Visualization

**4** Activities: Wrangling

**5** Activities: EDA

**6** Activities: Classification & Prediction

**7** Activities: Acquiring Data

**8** Additional Topics

**9** Homework

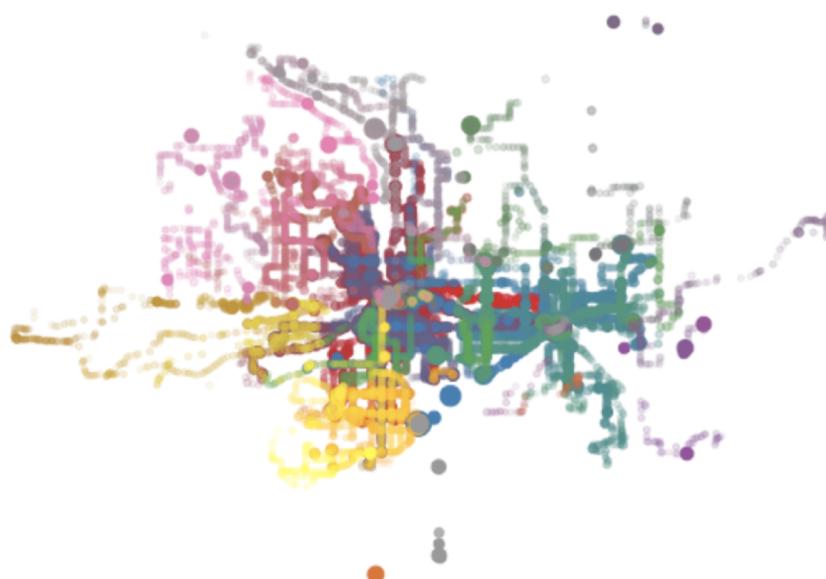
**10** RStudio Cheat Sheets

**11** Final Project

## COMP/MATH 112: Introduction to Data Science

*Fall 2017*

[Home](#)



Addona, Johnson, Kaplan, Sen, Shuman

<https://mac-comp112.github.io/112-course-development/>

# versions

Screenshot of a GitHub repository page for 'ajohns24 / project\_bayes'. The page shows a list of files and their recent activity. The files listed include bayes\_book\_files/figure-html, images, libs, beta-binomial.html, hello-there.html, index.html, introduction.html, modeling-foundations.html, posterior-inference.html, and posterior-simulation.html. Most files show activity from 'try again' or 'all current chapters' with timestamps ranging from 36 seconds ago to 6 minutes ago.

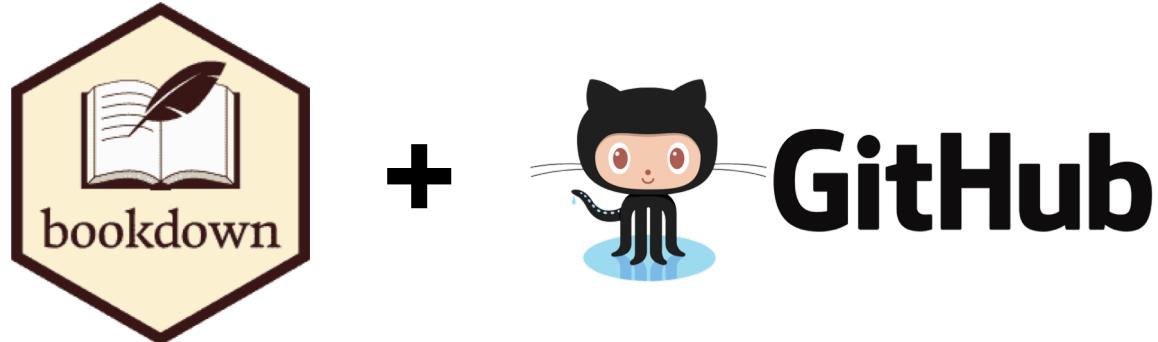
File	Activity	Timestamp
bayes_book_files/figure-html	try again	36 seconds ago
images	all current chapters	6 minutes ago
libs	all current chapters	6 minutes ago
beta-binomial.html	try again	36 seconds ago
hello-there.html	try again	36 seconds ago
index.html	all current chapters	6 minutes ago
introduction.html	try again	36 seconds ago
modeling-foundations.html	try again	36 seconds ago
posterior-inference.html	try again	36 seconds ago
posterior-simulation.html	try again	36 seconds ago

# not editions



shutterstock.com • 193735652

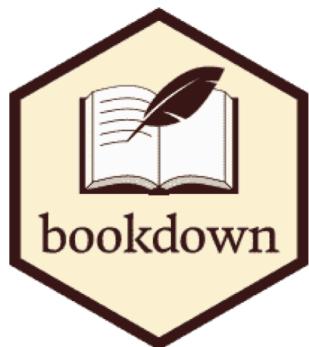
# Putting this all together...



- align with current best practices in statistics education;
- encourage readers to learn by doing;
- reflect the modern data science workflow; and
- look great.

# ACTION ITEM:

Convert a single project (eg: book, notes, manual, diary) to

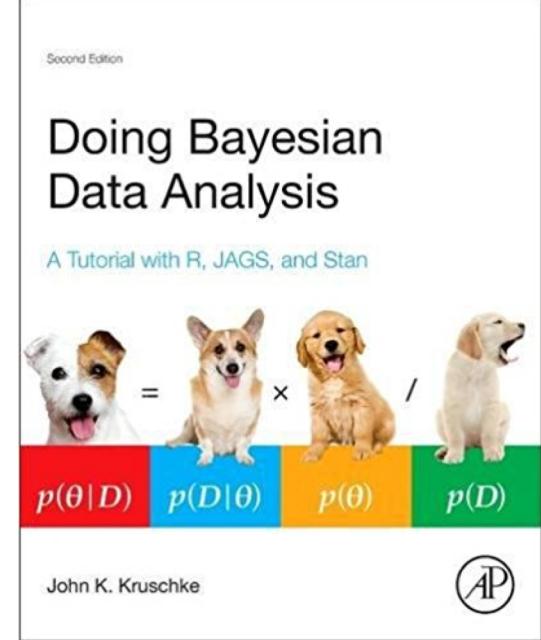
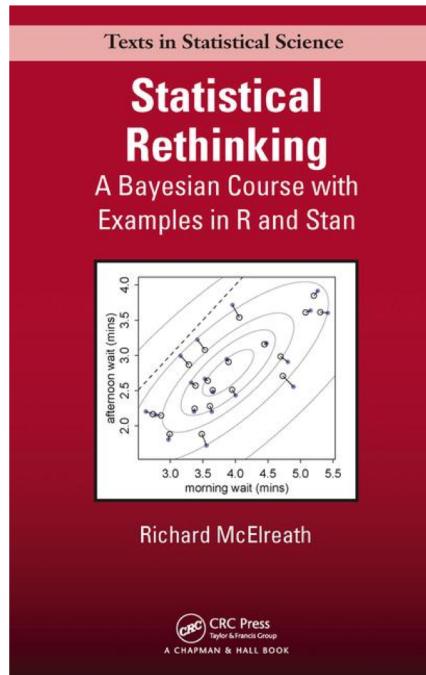
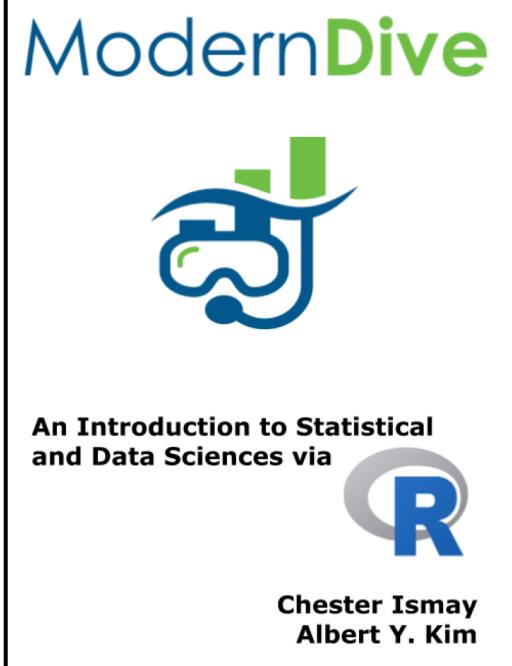


+



**GitHub**

# THANK YOU to some inspirations:



# THANK YOU for listening

## Contact:

- [ajohns24@macalester.edu](mailto:ajohns24@macalester.edu)
- aliciajohnson78.weebly.com
- github: ajohns24

## Upcoming projects:

- Rough bookdown sample:  
[https://ajohns24.github.io/project\\_bayes/](https://ajohns24.github.io/project_bayes/)
- DataCamp:  
Bayesian Modeling with RJAGS