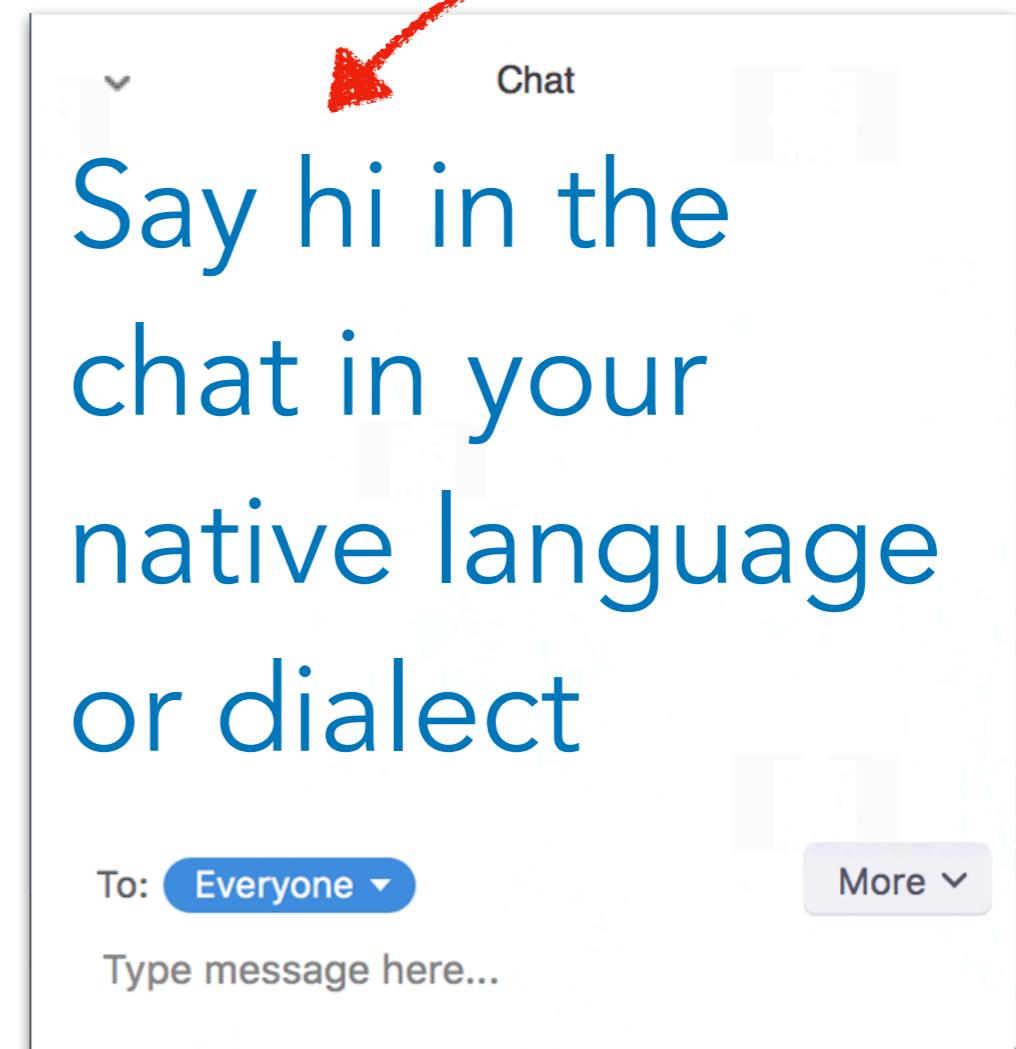
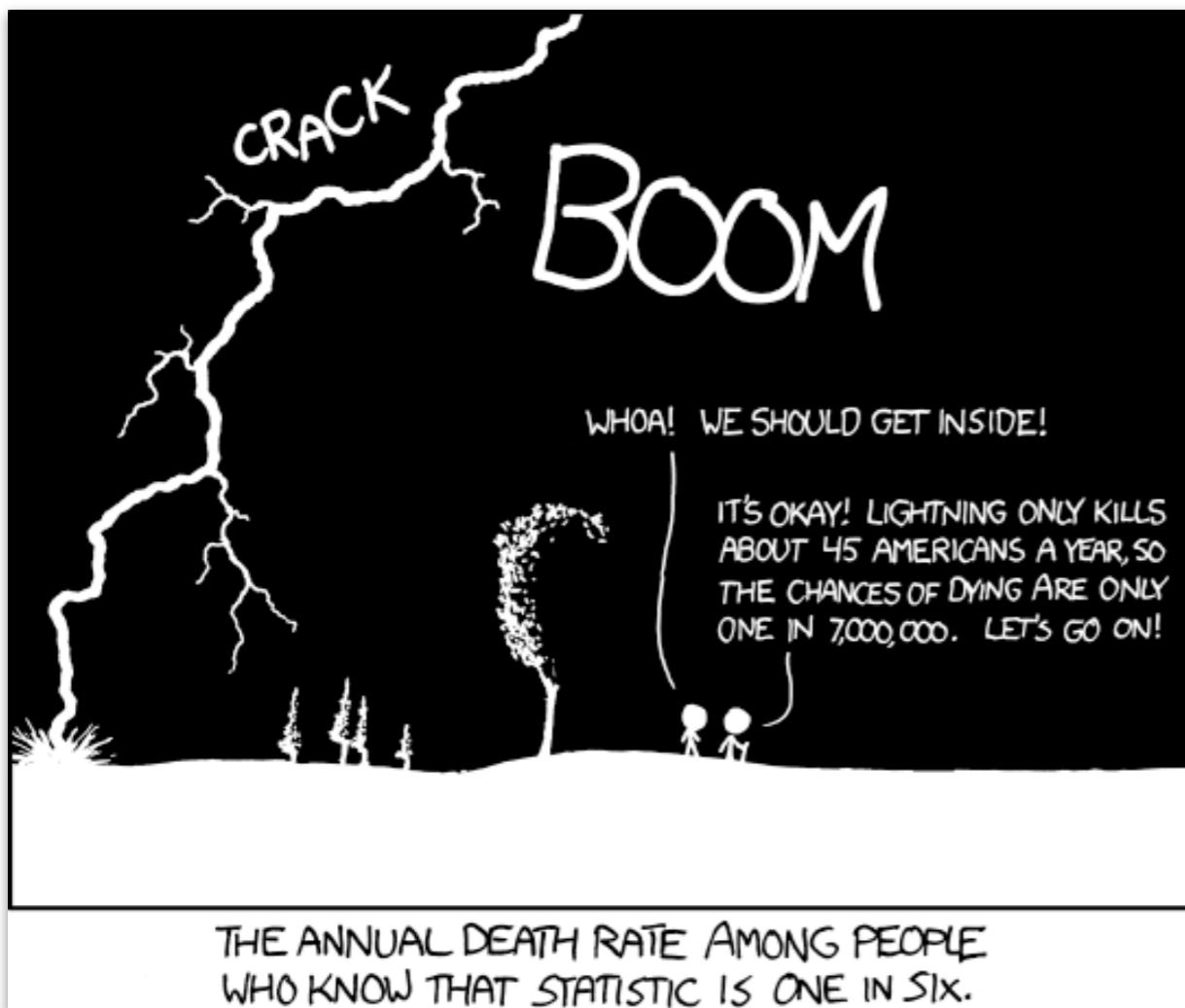


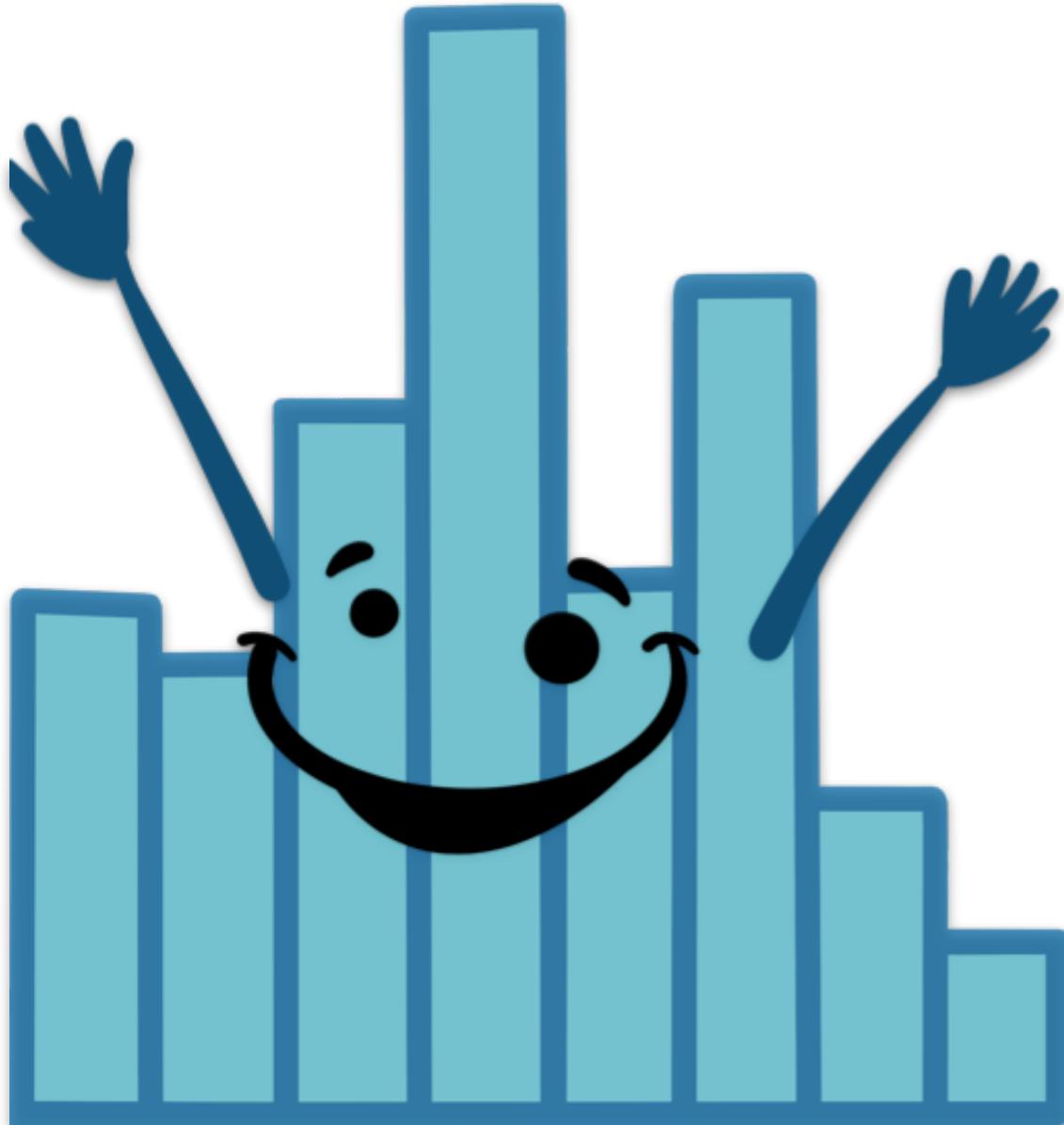
# Introduction

do this!



COLLABORATIVE PLAYLIST  
**psych252**  
<https://tinyurl.com/psych252spotify22>

01/04/2022



many of the  
illustrations I use  
in class are by

<https://github.com/allisonhorst/stats-illustrations>



**Allison Horst**  
allisonhorst

in the right

zoom session?



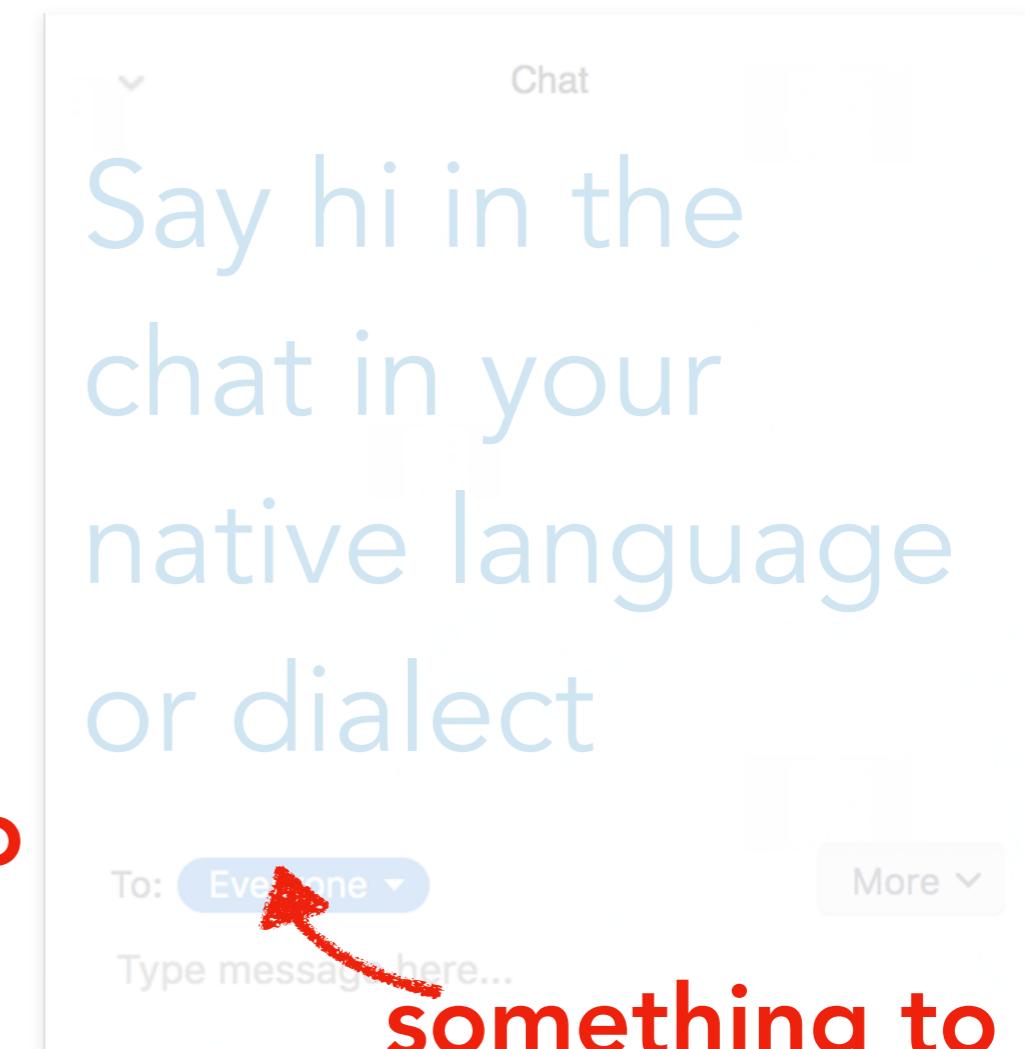
something to  
laugh about



something to  
listen to

# Introduction

soup of the day



01/04/2022

# Outline

- Introduction
- Getting the Zoomies
- **What** will we learn?
- **How** will we learn?
- Tools we will use in class
- Some general thoughts
- Feedback

# Introduction

# Introduction to the class

master the tools that empower  
you to do your research



develop statistical literacy

# Introduction to the team

## Team

**Tobi Gerstenberg**



**Ari Beller**



**Sarah Wu**



**Chengxu Zhuang**



Role

Instructor

Teaching assistant

Teaching assistant

Teaching assistant

Pronouns

he/him

he/they

she/her

he/him

Email (@stanford.edu)

gerstenberg

abeller

sarahawu

chengxuz

# Breakout rooms

## Tasks:

1. Who are you?

google  
slide intros



<https://tinyurl.com/psych252intros22>



2. What did you do during the winter break?

3. What are you hoping to get out of this course?

**Size:** ~3 people

**Time:** 10 minutes

**Report:** A few students may report back on Task 3.

# Introduction

The screenshot shows a survey interface titled "Psych 252: Introductory survey". The top navigation bar includes a back arrow, the title "Psych 252: Introductory", a file icon, a star icon, and various settings and send buttons. Below the title, there are tabs for "QUESTIONS" and "RESPONSES", with "QUESTIONS" being the active tab.

**Psych 252: Introductory survey**

This survey is meant to obtain some background information about the students in Psych 252, as well as to obtain some different types of data that we will use to analyze in the class. Please answer all questions as honestly as possible.

Some of the data collected in this survey may be made available to the class for use in analysis exercises. However, the data will be made anonymous before anything is released by removing any potentially identifying information, so that it will not be possible to determine which data came from which person in the group.

Please use your Stanford email address in order to get credit for completing the survey.

**Email address\***

Valid email address

This form is collecting email addresses. [Change settings](#)

**What year of graduate school are you in?\***

1. Undergraduate
2. 1-2
3. 3-4
4. 5+

**What department are you in?\***

Psychology

Education

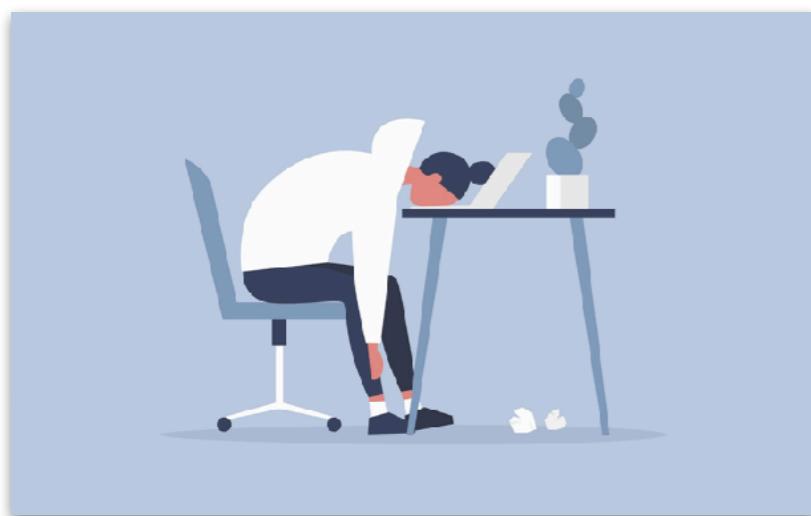
<https://tinyurl.com/psych252survey22>

A small, white, shaggy dog is captured in mid-stride, running towards the camera through a lush green grassy field. The dog's tongue is slightly out, suggesting it might be panting or excited. The background is a soft-focus green, indicating a natural outdoor setting.

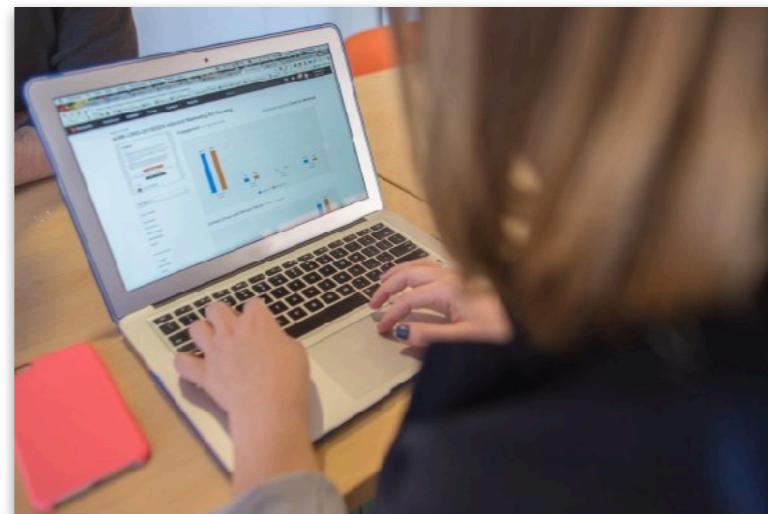
Getting the Zoomies

# Challenges of online classes

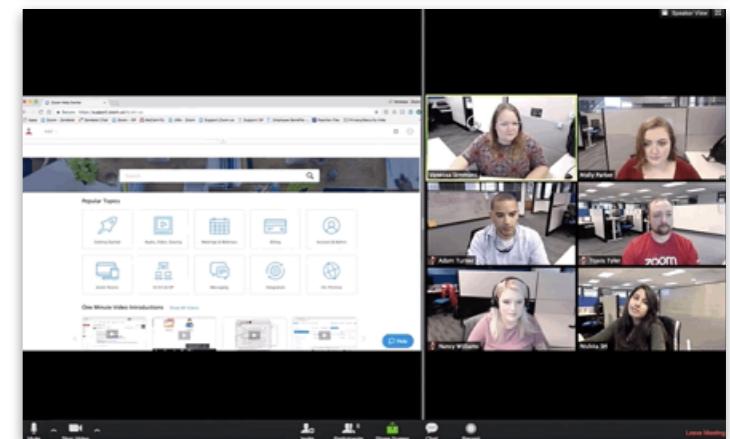
zoom fatigue



can't look your friend over the shoulder



different perspective



difficult to create community



difficult to read the room



How was your meeting experience?



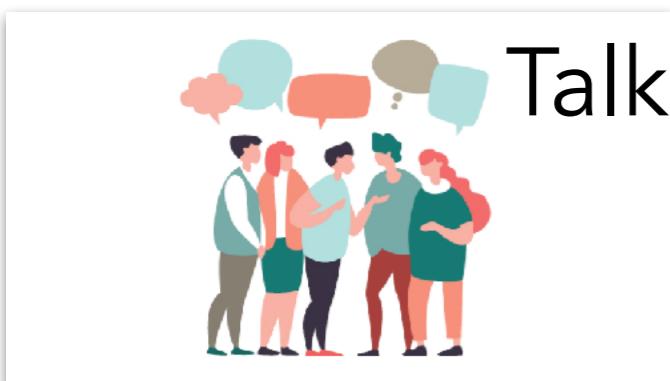
Good



Bad

# Advantages of online classes

Multiple ways of participation



Meeting new people



Recording lectures



# Zoom norms

Video on (if possible)



# Zoom norms

In the main room



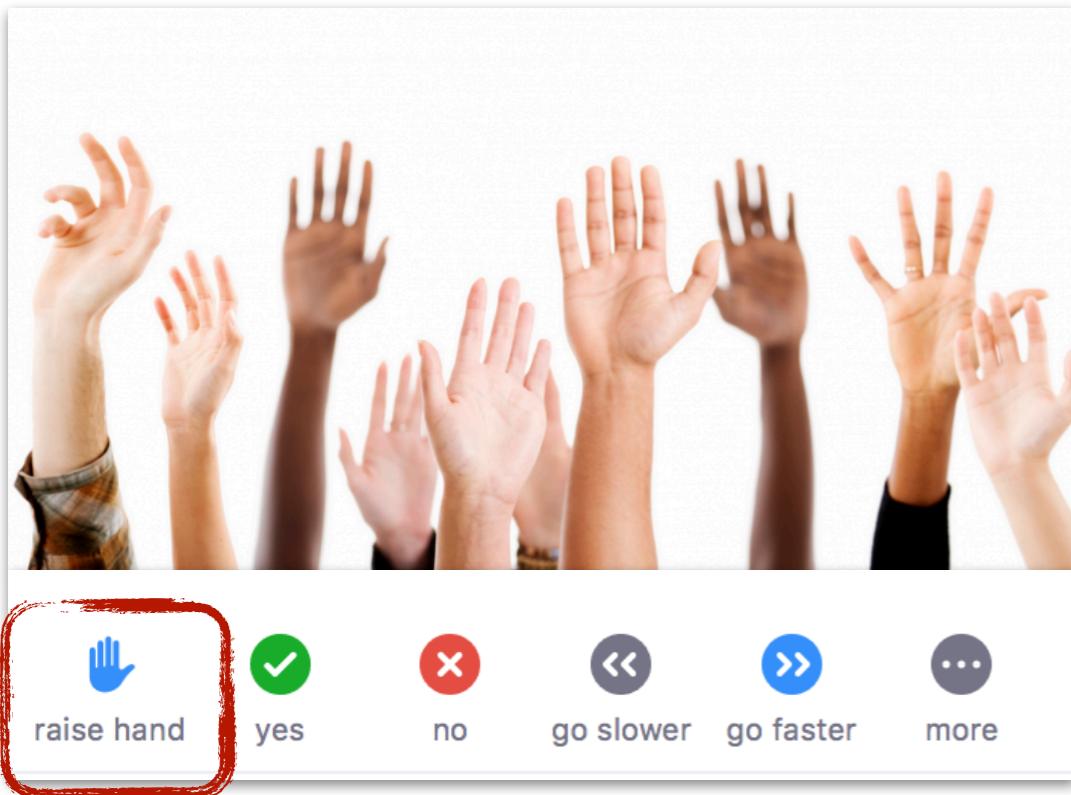
In breakout rooms



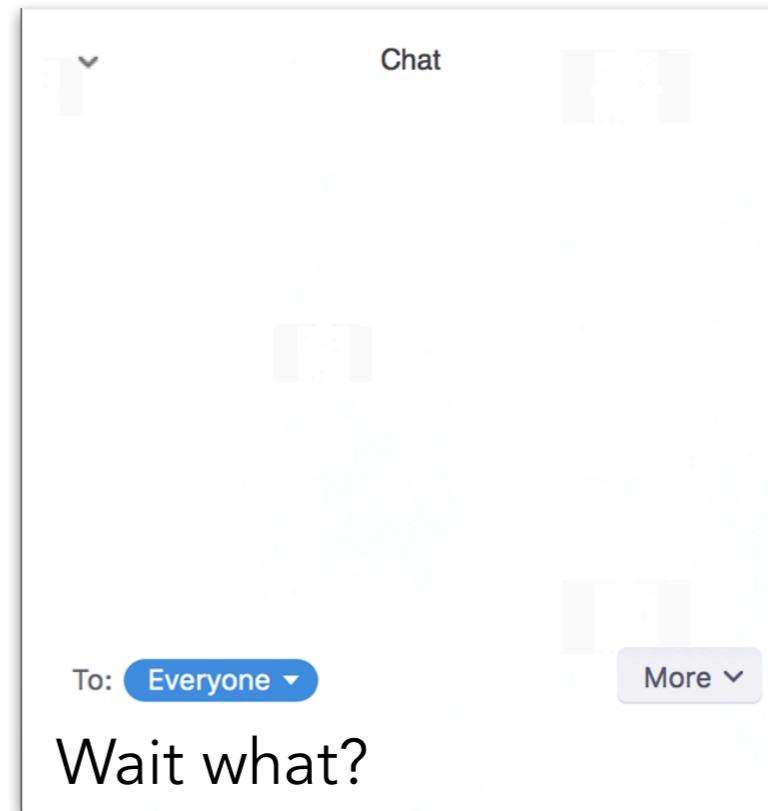
# Zoom norms

## Participating

Raise hand



Ask in the chat



- chat will be monitored by TAs
- no need to monitor the chat (can be distracting)
- recording of public chat will be available after class
- private chat TAs for technical help

# **What will we learn?**

# Teaching philosophy

found in here



Teaching Effectively with Zoom  
A practical guide to engage your students and help them learn  
By Dan Levy

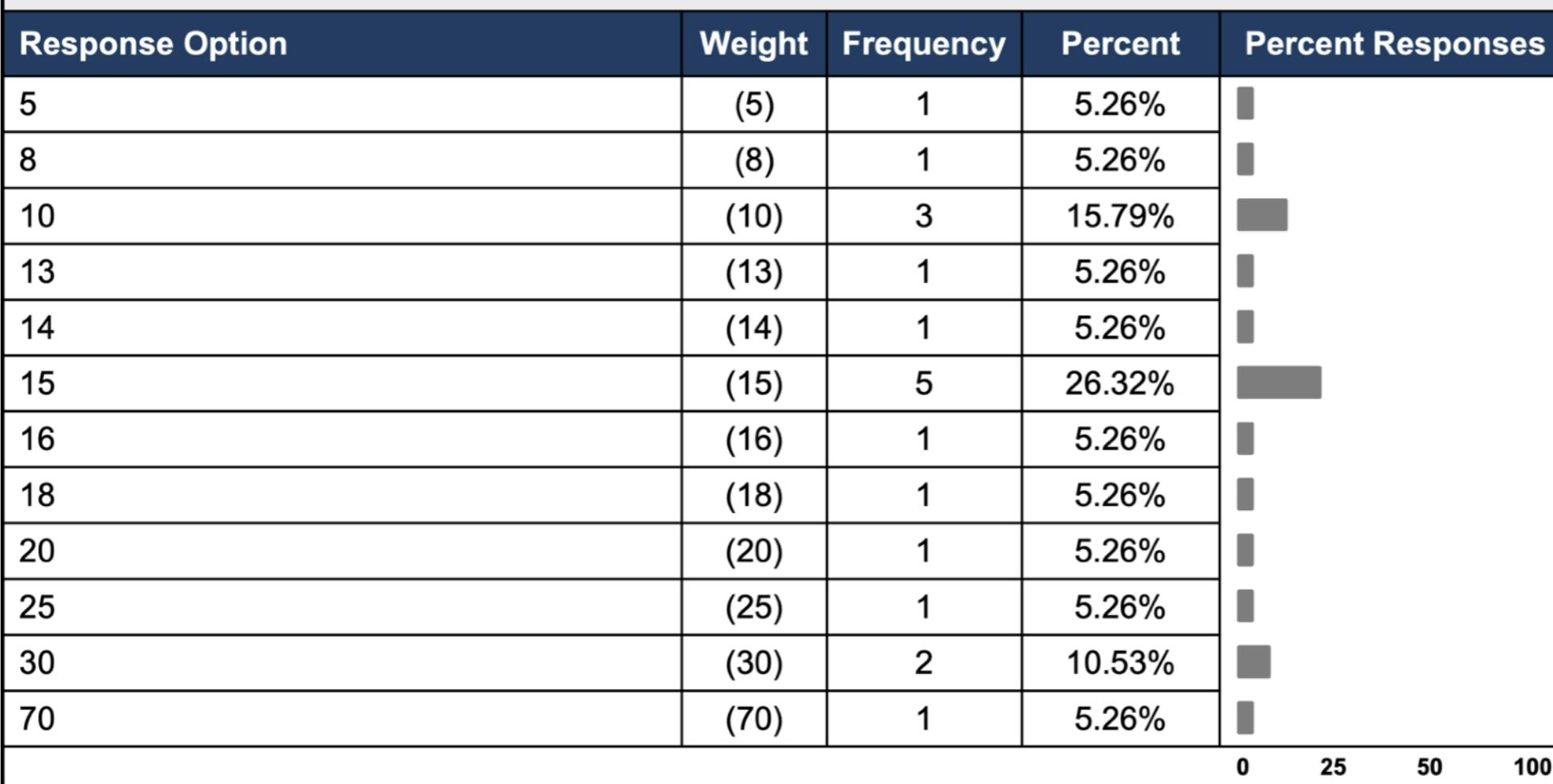
**"The one who does the work does the learning." (Doyle, 2008)**

we will try our best to make you do the work!

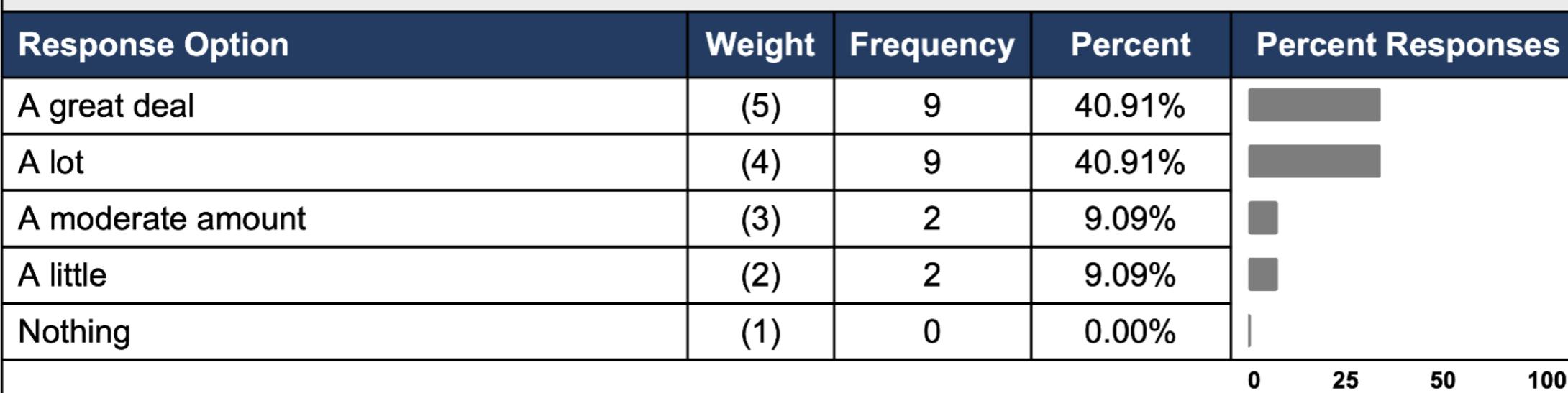
Doyle, T. (2008). Helping students learn in a learner-centered environment: A guide to facilitating learning in higher education. Stylus Publishing, LLC.

# Doing the work takes time ...

## 6 - How many hours per week on average did you spend on this course (including class meetings)?



## 3 - How much did you learn from this course?



... but it's  
worth it

# Learning goals

You will learn how to **use R** to ...

- read, wrangle, simulate and analyze data
- make publication-ready plots

# What will we learn?

## **Weeks 1-3** 1. Use R!

- Data visualization
- Data manipulation/wrangling
- Understand key statistical concepts
  - Simulation, manipulation, visualization

## 2. Build models

- Weeks 4-7**
- Formulate hypotheses as statistical models

- Weeks 8-9**
- Bayesian statistics

## 3. Report results

- all the time  
(& Week 10)**
- Reproducible research

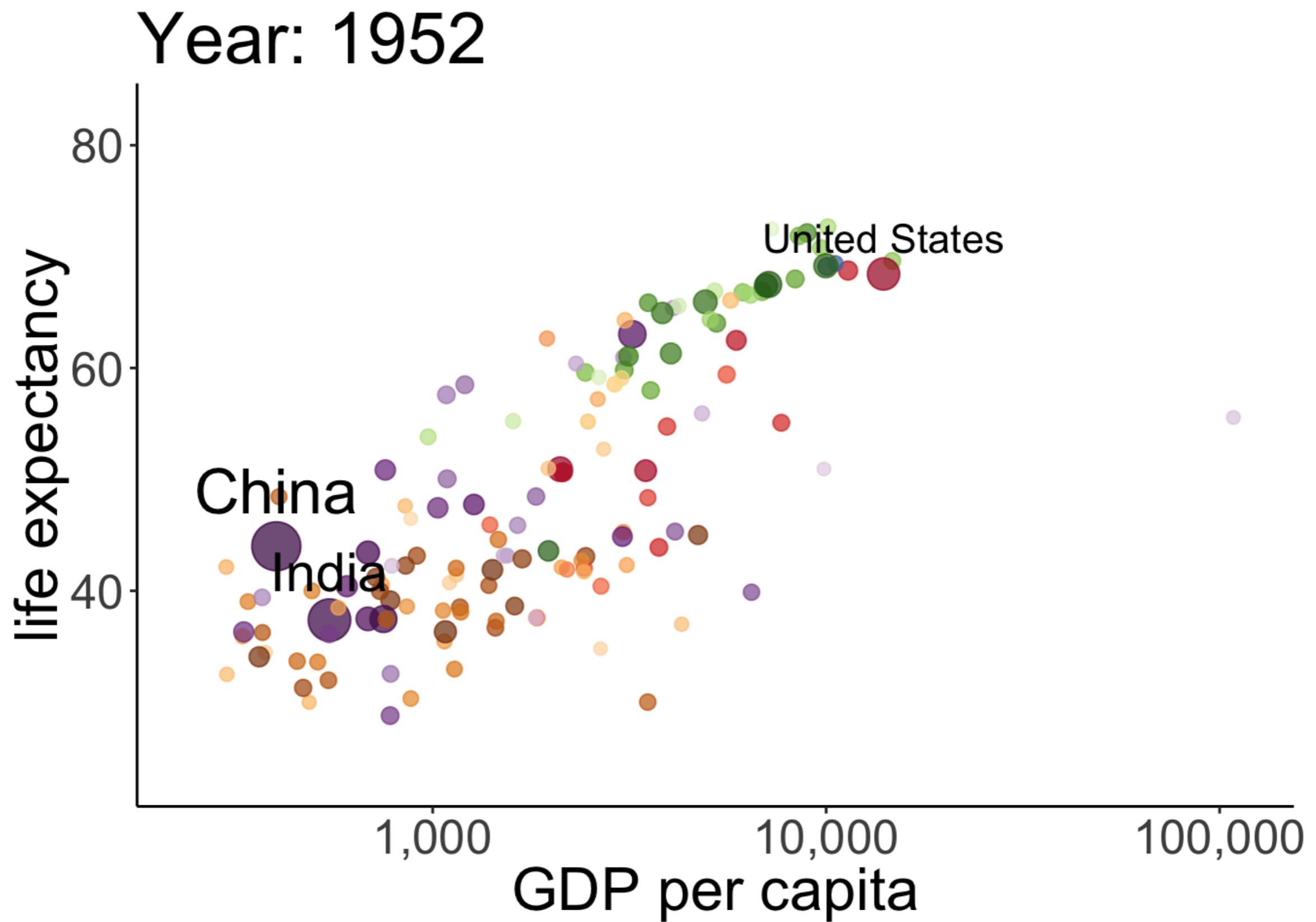
1. Use R!

# 4 Reasons for why ?

1. Visualization
2. Data manipulation
3. Statistical modeling
4. Communicating results

# Why R ?

## 1. Visualization



# Why R? 1. Visualization

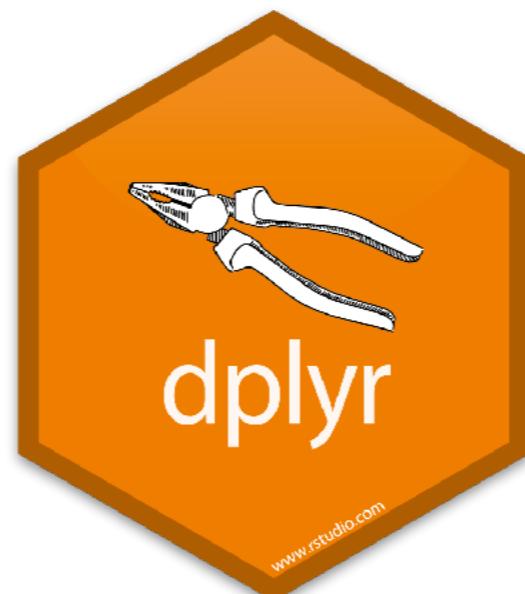
```
1 ggplot(gapminder, mapping = aes(x = gdpPercap, y = lifeExp, size = pop, color = country)) +
2   geom_point(alpha = 0.7, show.legend = FALSE) +
3   geom_text(data = gapminder %>% filter(country %in% c("United States", "China", "India")),
4             mapping = aes(label = country),
5             color = "black",
6             vjust = -0.75,
7             show.legend = FALSE) +
8   scale_colour_manual(values = country_colors) +
9   scale_size(range = c(2, 12)) +
10  scale_x_log10() +
11  labs(title = "Year: {frame_time}", x = "GDP per capita", y = "life expectancy") +
12  transition_time(year) +
13  ease_aes("linear")
```



# 2. Data manipulation

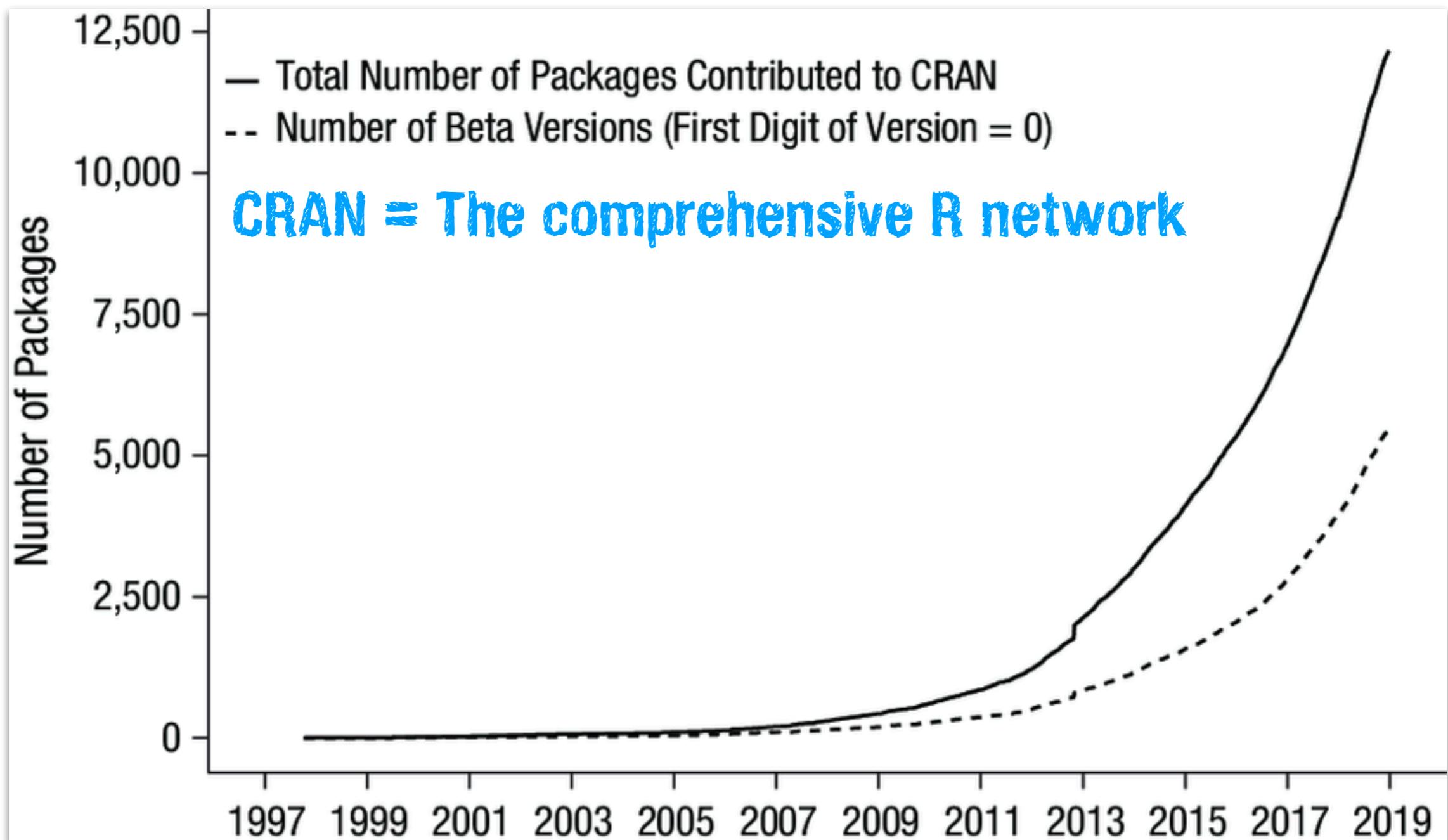
A1	StartDate;EndDate;Status;IPAddress;Progress;Duration (in seconds);Finished;RecordedData;ResponseID;RecipientLastName;RecipientFirstName;ExternalReference;LocationLatitude;LocationLongitude
1	StartDate;EndDate;Status;IPAddress;Progress;Duration (in seconds);Finished;RecordedData;ResponseID;RecipientLastName;RecipientFirstName;ExternalReference;LocationLatitude;LocationLongitude
2	StartDate;EndDate;ReasonsType;IPAddress;Progress;Duration (in seconds);Finished;RecordedDate;ResponseID;RecipientLastName;RecipientFirstName;RecipientEmail;ExternalDataReference;LocationLatitude;LocationLongitude
3	
4	
5	
6	Liam says to isn't it?"
7	
8	
9	Who is Liam talking to? (case sensitive)";"Are you a bot?
10	Mie says to E isn't it?"
11	
12	
13	Who is Mia talking to? (case sensitive)";"Are you a bot?
14	William says isn't it?"
15	
16	
17	Who is William talking to? (case sensitive)";"Are you a bot?
18	Amelia says isn't it?"
19	
20	
21	Who is Amelia talking to? (case sensitive)";"Are you a bot?
22	Furlyn says t isn't it?"
23	
24	
25	Who is Evelyn talking to? (case sensitive)";Did Chad vote for or against the pol?;How many votes are required for the policy to pass?;Did the policy pass?;How morally bad is it to ban electric scooters from driving?;1. Age - Selected Choice;"Please provide the following demographic information to complete the study.
26	1. Age - Selected Choice"2. Gender"3. Race - Selected Choice"3. Race - Text"4. Ethnicity"5. One or more of your responses were incorrect. Please re-read the instructions and try again"PL_31 - Block R
27	2020-11-17 09:31:41;2020-11-17 10:23:17;IP Address:171.66.161.111;73095;False;2020-11-24 10:21:03;R_0Tx4nCgG9vBPH;anonymous;EN;Emma;;For the pol;2;Yes;52;3;86;87;1;
28	2020-11-17 09:31:41;2020-11-17 10:23:17;IP Address:171.66.161.111;73095;False;2020-11-24 10:21:03;R_0Tx4nCgG9vBPH;anonymous;EN;Emma;;For the pol;2;Yes;52;3;86;87;1;
29	2020-11-17 08:36:54;2020-11-19 06:42:56;IP Address:134.76.136.123;12.165542;False;2020-11-26 06:42:38;R_0dpUiaUvcYOkTL;anonymous;EN;Emma;;For the policy;2;Yes;100;26;55;22;54;
30	2020-12-04 1 then their w the number they could have a significant ethnic or religious minority with strong views that oppose the majority. Hence none of the answers were terribly surprising. "My age in years
31	2020-12-14 07:22:31;2020-12-14 07:23:23;Survey_Preview;100;51;True;2020-12-14 07:23:23;R_cGAxLivoSnl6055...;2.4149017333994375;7 07659912109375;preview;EN;James;Against the policy;1;Yes

thanks qualtrics :(



Making Data Wrangling Suck Less

# Why R? 3. Statistical modeling



This is what R was developed to do.  
(and it continues to develop!!)

# Why R? 4. Communicating results

```
01-intro.Rmd x
ABC 🔎 Knit ⚙️
1 # Introduction {#intro}
2
3 You can label chapter and section titles using `#{label}` after them,
do not manually label them, there will be automatic labels anyway, e.
4
5 Figures and tables with captions will be placed in `figure` and `tabl
6
7
8 ```{r nice-fig, fig.cap='Here is a nice figure!', out.width='80%', fi
9 par(mar = c(4, 4, .1, .1))
10 plot(pressure, type = 'b', pch = 19)
11 ```
12
13 Reference a figure by its code chunk label with the `fig:` prefix, e.
can reference tables generated from `knitr::kable()`, e.g., see Table
14
15 ```{r nice-tab, tidy=FALSE}
16 knitr::kable(
17   head(iris, 20), caption = 'Here is a nice table!',
18   booktabs = TRUE
19 )
20 ```
21
22 You can write citations, too. For example, we are using the **bookdown**
was built on top of R Markdown and **knitr** [@xie2015].
23
24 ```{r stats-help, fig.cap='Stats cheatsheet',fig.align='center',echo=
25 knitr::include_graphics('figures/cheatsheets/stats-help.jpg')
26 ```
27
28 See figure \@ref(fig:stats-help)
29
30
31 ```{r klippy, echo=FALSE, include=TRUE}
32 klippy::klippy()
33 ...
34
```

## Chapter 1 Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 1. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter ??.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Figure 1.1: Here is a nice figure!

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 1.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 1.1.

R Markdown

# Why R ? 4. Communicating results

The screenshot shows the homepage of the R for Data Science website. The left sidebar contains a table of contents for the book, organized into four main sections: I Explore, II Wrangle, III Program, and IV Model. The main content area features the book's title "R for Data Science" in large bold letters, followed by the authors' names, Garrett Grolemund and Hadley Wickham. Below this is a "Welcome" section with a detailed description of what the book teaches. To the right of the welcome text is an image of the book cover, which features a green parrot (kākāpō) and the title "R for Data Science". At the bottom of the page, there is a note about the website's license and a link to the book's source code.

Welcome

R for Data Science

Garrett Grolemund  
Hadley Wickham

**Welcome**

This is the website for “R for Data Science”. This book will teach you how to do data science with R: You’ll learn how to get your data into R, get it into the most useful structure, transform it, visualise it and model it. In this book, you will find a practicum of skills for data science. Just as a chemist learns how to clean test tubes and stock a lab, you’ll learn how to clean data and draw plots—and many other things besides. These are the skills that allow data science to happen, and here you will find the best practices for doing each of these things with R. You’ll learn how to use the grammar of graphics, literate programming, and reproducible research to save time. You’ll also learn how to manage cognitive resources to facilitate discoveries when wrangling, visualising, and exploring data.

This website is (and will always be) **free to use**, and is licensed under the [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License](#). If you’d like a **physical copy** of the book, you can order it from [amazon](#); it was published by O’Reilly in January 2017. If you’d like to **give back** please make a donation to [Kākāpō Recovery](#): the kākāpō (which appears on the cover of R4DS) is a critically endangered native NZ parrot; there are only 148 left.

The book is written in [RMarkdown](#) with [bookdown](#). It is automatically rebuilt from [source](#) by [travis](#). R4DS is a collaborative effort and many people have contributed fixes and improvements via pull request.

O'REILLY

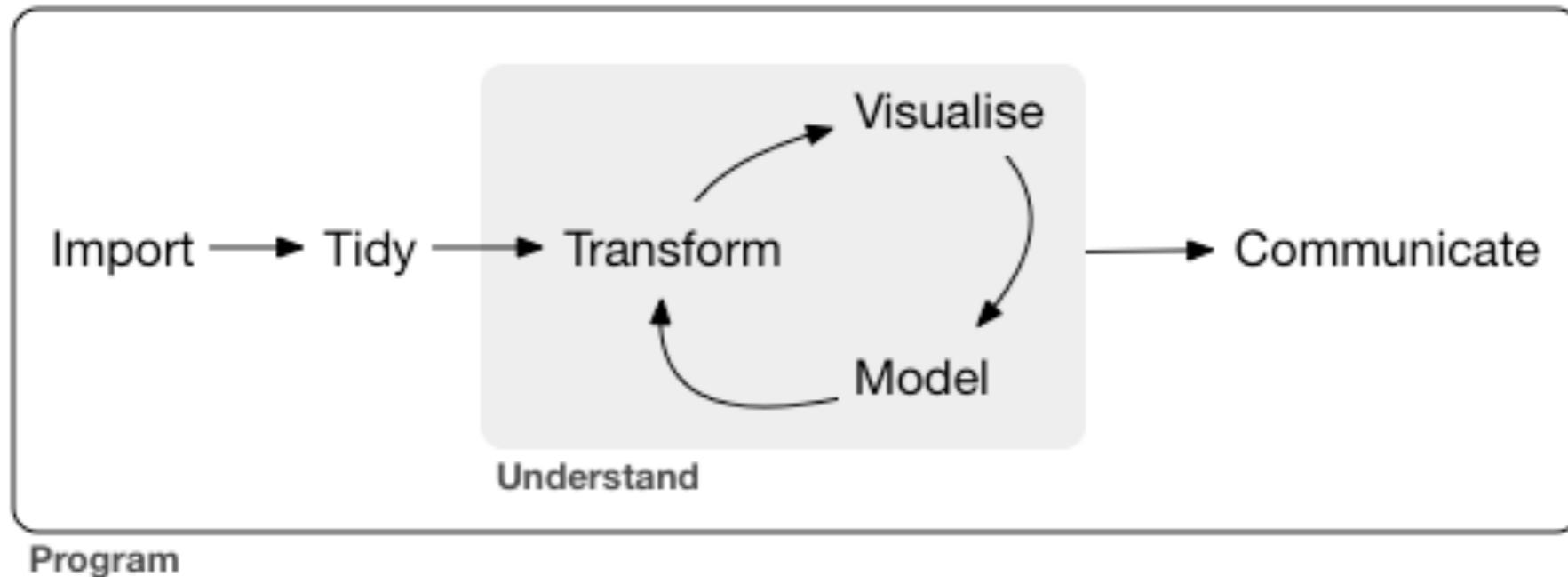
R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham & Garrett Grolemund

<https://r4ds.had.co.nz/>

# Why R? Tidyverse

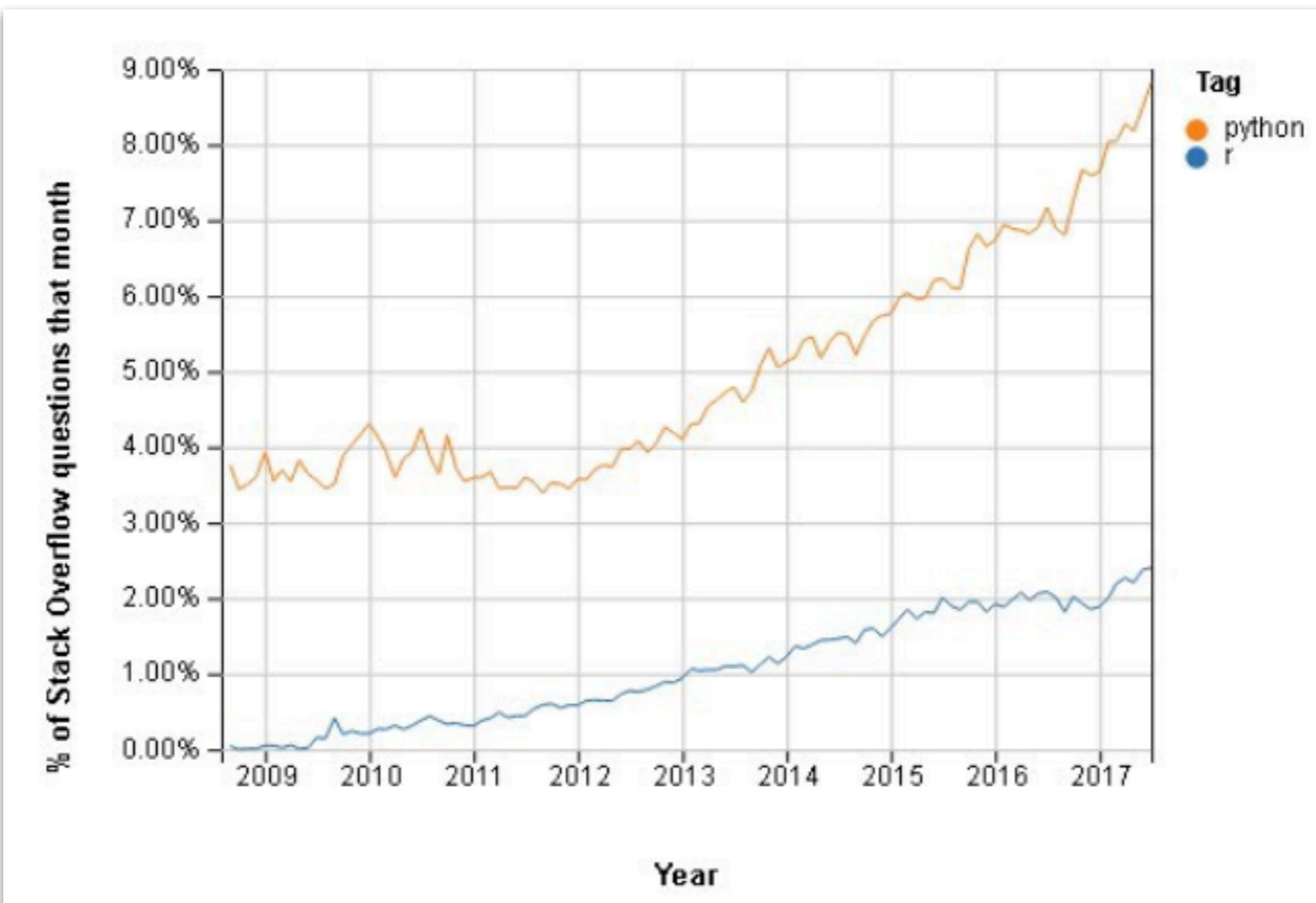


- *Import*: get data into R
- *Tidy*: clean and format the data
- *Transform*: select variables, create new ones, group and summarize
- *Visualize*: "look" at the data in different ways
- *Model*: answer questions about the data
- *Communicate*: write reproducible research reports

# Why ? Also ...

- many of us here at Stanford use it
- R has a nice online community. **#rstats**
- RStudio is a great IDE (Integrated Development Environment)!

# But what about python ™ ?



*Image Source: <https://dzone.com/articles/r-or-python-data-scientists-delight>*

# But what about python™ ?

- You should learn python, too! :)
- Python is great for machine learning (e.g. deep learning).
- You can learn about all of these things in **Psych 253!**

**Psych 253**

Basics / Course Schedule & Syllabus / Software Tools

[f](#) [e](#)

**High-Dimensional Methods for Behavioral and Neural Data**  
Spring 2018, Stanford University

Introduction to high-dimensional data analysis and machine learning methods for use in the behavioral and neurosciences, including: supervised methods such as SVMs, linear and nonlinear regression and classifiers, and regularization techniques; statistical methods such as bootstrapping, signal detection, factor analysis, and reliability theory; metrics for model/data comparison such as representational similarity analysis; and unsupervised methods such as clustering. Students will learn how to both use existing statistical data analysis packages (such as [scikit-learn](#)) as well to build, optimize, and estimate their own custom models using the [Tensorflow](#) optimization framework.

**Time:** Tue./Thu. 12:00p - 1:20p

**Location:** Science Teaching and Learning Center (STLC), Rm 105

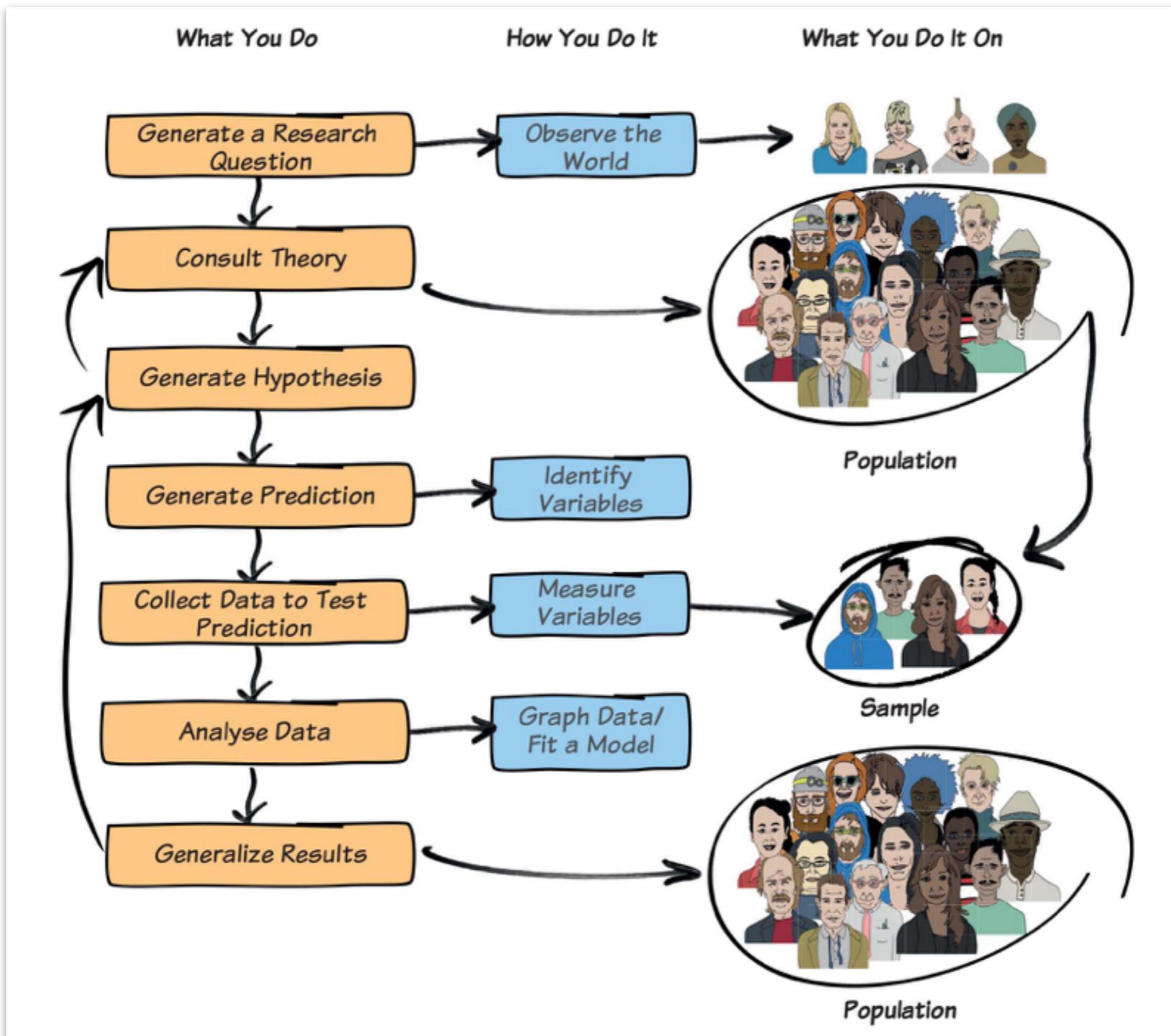
**Staff:** Daniel Yamins (x@stanford.edu where x=yamins) & Chengxu Zhuang (x=chengxuz)

- visualization
- data manipulation
- statistical modeling
- reporting



## 2. Build models

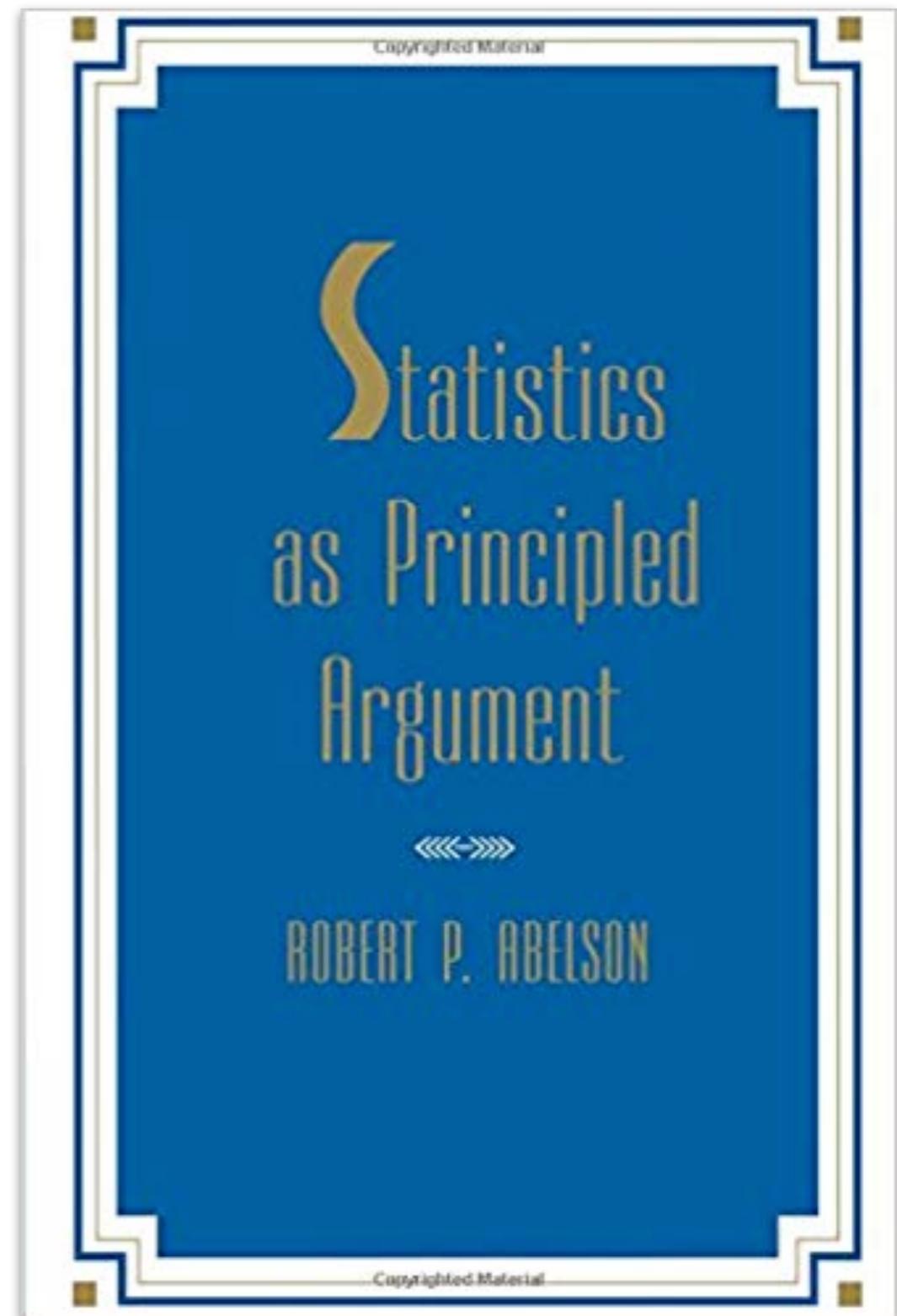
# The research process



# Statistics as argument

Researchers use empirical observations  
(data) for **making arguments about  
research questions.**

Statistics is part of a narrative and  
requires **figures**, context, rhetoric,  
prose, ...



Abelson, R. P. (1995). *Statistics as principled argument*. New York: Psychology Press.



# helps us with building models

instead of

$$\begin{aligned}
 \sum_{\mu,\nu,\lambda} \hat{M}_{\mu\nu\lambda}^{(3)} F_{\mu\nu\lambda} &= \hat{M}_{000}^{(3)} F_{000} + 3 \sum_i \hat{M}_{00(i)}^{(3)} F_{00(i)} + 3 \sum_i \hat{M}_{0ii}^{(3)} F_{0ii} + 3 \sum_i \hat{M}_{0(i)i}^{(3)} F_{0(i)i} \\
 &\quad + 6 \sum_{i < j} \hat{M}_{0(ii)(jj)}^{(3)} F_{0(ii)(jj)} + 3 \sum_{i < j} \hat{M}_{0(ii)(ij)}^{(3)} F_{0(ii)(ij)} + 3 \sum_{i < j} \hat{M}_{ii(jj)}^{(3)} F_{ii(jj)} + 6 \sum_{i < j} \hat{M}_{ij(ji)}^{(3)} F_{ij(ji)} + \dots \\
 &= S^{(0)} F_{000} - \frac{2S^{(1)}}{d} \frac{\sigma_1^2}{\sigma_0^2} \left( 2 \sum_i F_{00(ii)} - \sum_i F_{0ii} \right) - \frac{2S^{(2)}}{d^2(d+2)} \frac{\sigma_1^4}{\sigma_0^4} \\
 &\quad \times \left[ 3(d-1) \sum_i F_{0(ii)(ii)} - 6 \sum_{i < j} F_{0(ii)(jj)} + \frac{3d}{2} \sum_{i < j} F_{0(ii)(ij)} + (d+2) \sum_{i < j} F_{ii(jj)} - (d+2) \sum_{i < j} F_{ij(ji)} \right] \\
 &\quad + \frac{3S_2^{(2)}}{d(d+2)} \frac{\sigma_1^4}{\sigma_0^4} \left( 3 \sum_i F_{0(ii)(ii)} + 2 \sum_{i < j} F_{0(ii)(jj)} + \sum_{i < j} F_{0(ii)(ij)} \right) + \dots. \tag{85}
 \end{aligned}$$

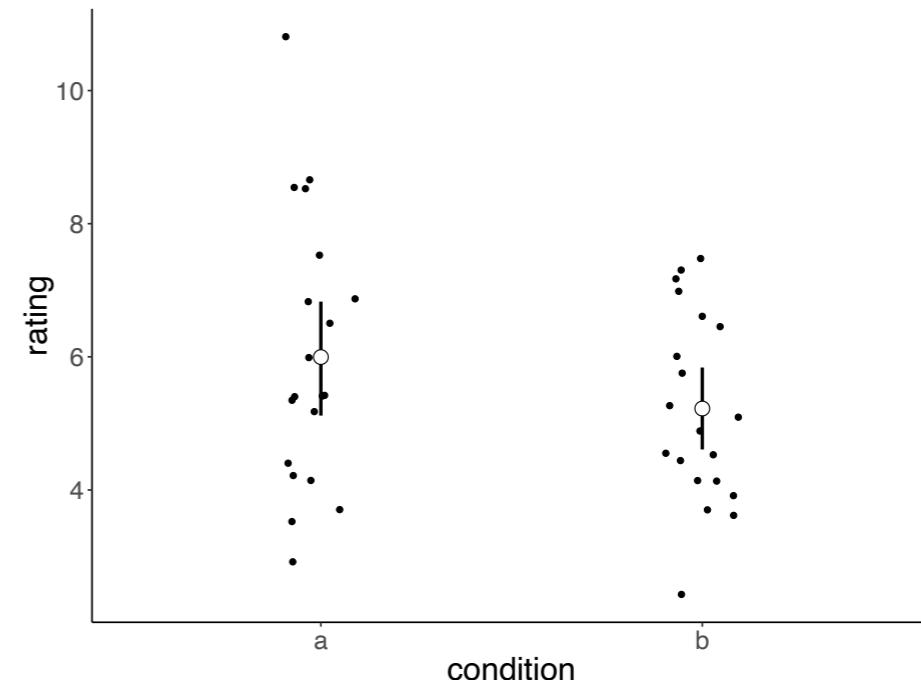
(google search "complicated statistical equation")

we'll do

```

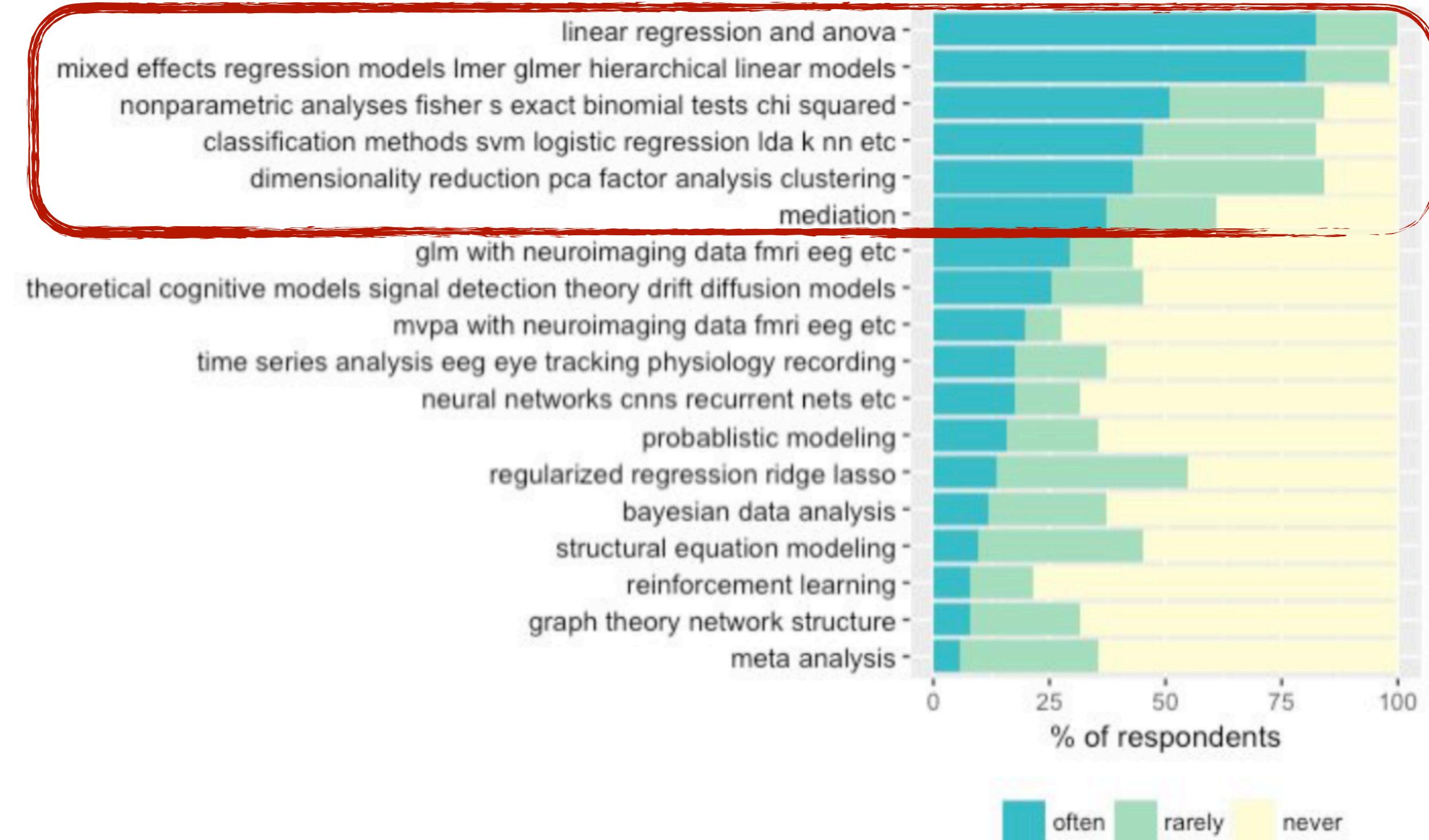
1 set.seed(0)
2 df.data = tibble(
3   a = rnorm(20, mean = 6, sd = 2),
4   b = rnorm(20, mean = 5, sd = 2)
5 ) %>%
6 gather("condition", "rating")
7
8 df.data %>%
9 group_by(condition) %>%
10 summarize(rating.mean = mean(rating),
11           rating.sd = sd(rating)) %>%
12 kable()
13
14 # calculate the difference between conditions
15 difference.actual = df.data %>%
16 group_by(condition) %>%
17 summarize(rating.mean = mean(rating)) %>%
18 pull(rating.mean) %>%
19 diff() %>%
20 -

```

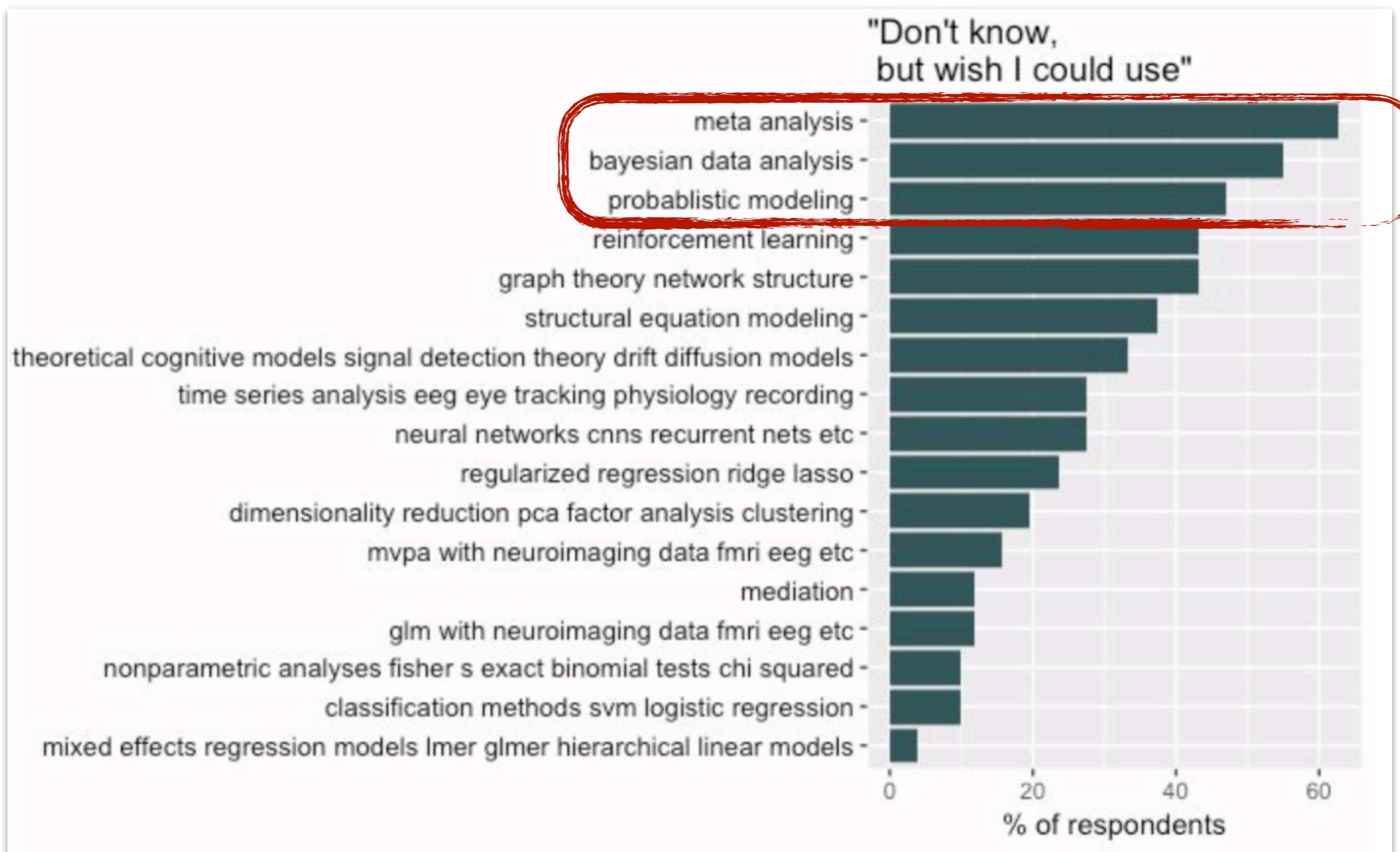


# Psychology grad student survey

“How often do you **use** the following skills?”



# Psychology grad student survey



### 3. Report results

# Replication crisis

Google search results for "replication crisis in psychology".

Search filters: Images, SafeSearch.

Top results:

- No Evidence for a Replicability Crisis** (Harvard Project)
- Economics (n=18) vs Psychology (n=100)** (Camerer et al., 2016; Open Science Collaboration, 2015)
- Journal Replication Rates** (Overall 36%)
- Effect size reduction and independent replication p-value**
- The Replication Crisis in Psychology | NOBA** (Edward Diener and Robert Biswas-Diener)

Other results include:

- How to Fix Psychology's Replication Crisis** (Chronicle of Higher Education)
- Tradition of controlled experiments** (Nature, Atlantic, Science)
- Replication Crisis in Science** (Simply Statistics)
- Only 36% of studies replicated** (Science)
- psychological science** (Science)
- Replication Crisis Overblown** (The Cut)
- Health News Digest** (bfm.my)
- 6 Principles of Open Science** (Noba)
- How will stereotype threat get us** (Progress Focused)
- Replication | TOK Topics** (TOKTopics)
- psychological science** (Science)
- Replication Crisis** (Cambridge)

Study	Field	Replication rate	Effect size quotient	N (samples)	Preregistered	Target selection
Open Science Collaboration (OSCI), 2015	Psychology	36%	0.49	100	No	Quasirandom
Klein, Ratliff, & Vianello, 2014	Psychology	77%	1.26	36	Yes	Nonrandom
Special Issue contributions to <i>Social Psychology</i> , vol. 45(3), 2014	Psychology	85%	—	13	Yes	Nonrandom
Camerer et al., 2016	Economics	61%	0.66	18	No	Nonrandom

# R Markdown

## R Markdown :: CHEAT SHEET

### What is R Markdown?

**.Rmd files** - An R Markdown (.Rmd) file is a record of your research. It contains the code that a scientist needs to reproduce your work along with the narration that a reader needs to understand your work.

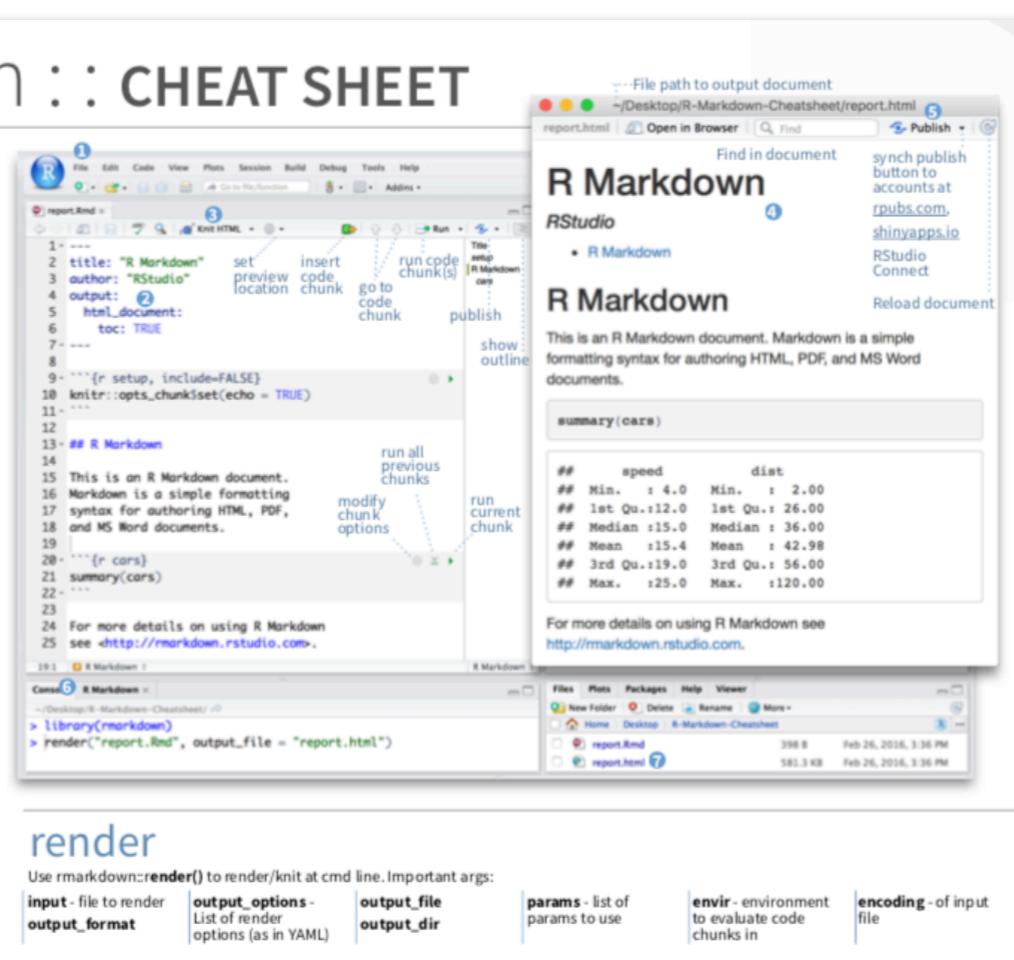
**Reproducible Research** - At the click of a button, or the type of a command, you can rerun the code in an R Markdown file to reproduce your work and export the results as a finished report.

**Dynamic Documents** - You can choose to export the finished report in a variety of formats, including html, pdf, MS Word, or RTF documents; html or pdf based slides, Notebooks, and more.

### Workflow



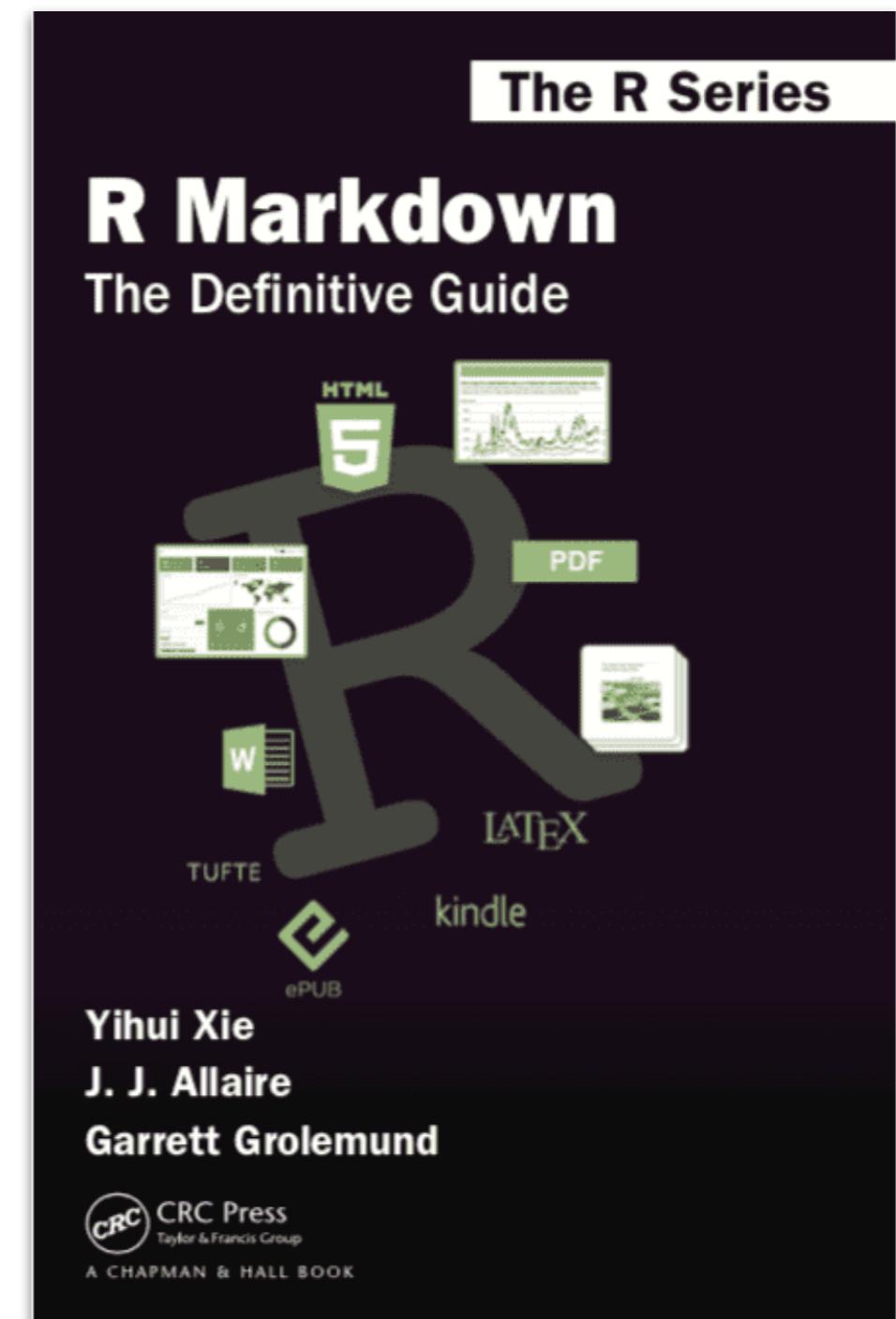
- ① Open a new .Rmd file at File ▶ New File ▶ R Markdown. Use the wizard that opens to pre-populate the file with a template
- ② Write document by editing template
- ③ Knit document to create report; use knit button or render() to knit
- ④ Preview Output in IDE window
- ⑤ Publish (optional) to web server
- ⑥ Examine build log in R Markdown console
- ⑦ Use output file that is saved along side.Rmd



The screenshot shows the RStudio interface with the following components:

- Top Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Tools, Help.
- Left Sidebar:** Projects, Home, Recent, Files, R Markdown, R Notebook, R Data Frame, R Shiny App, R Markdown Document, R Notebook Document, R Data Frame Document, R Shiny App Document.
- Code Editor:** Shows an R Markdown file (report.Rmd) with code and comments. A tooltip explains various buttons: set preview location, insert code chunk, run code chunk(s), publish, show outline, modify chunk options, run all previous chunks, and run current chunk.
- Preview Area:** Displays the rendered HTML output of the R Markdown document, showing a summary of the cars dataset.
- Console:** Shows the command `library(rmarkdown)` and `render("report.Rmd", output\_file = "report.html")` being run.
- File Explorer:** Shows the file structure with `report.Rmd` and `report.html` files.

<https://github.com/rstudio/cheatsheets/raw/master/rmarkdown-2.0.pdf>



**The R Series**

# R Markdown

## The Definitive Guide

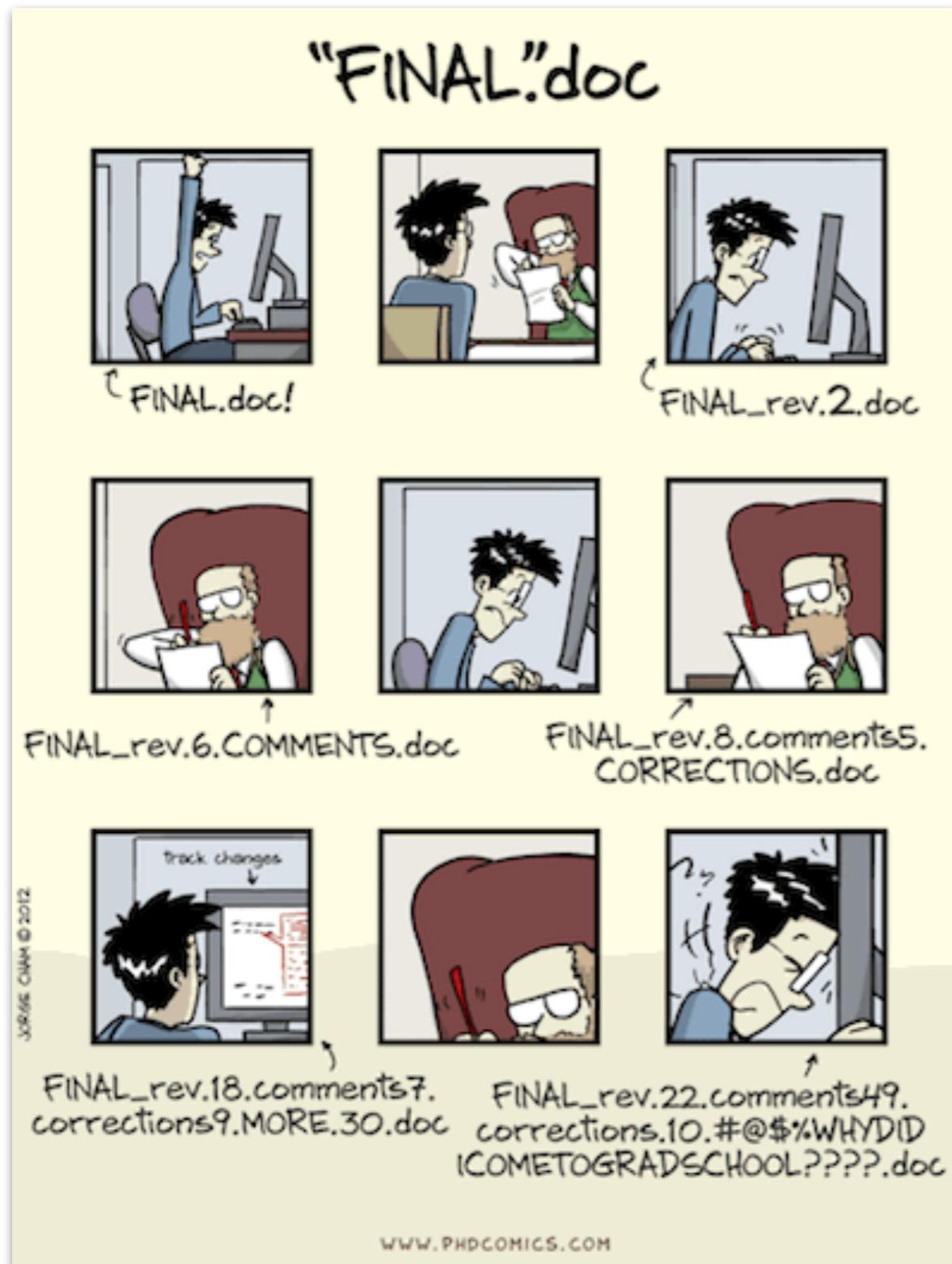
A large stylized R logo is in the center, surrounded by icons for different output formats: HTML 5, PDF, TUFTE, LATEX, kindle, and ePUB.

**Yihui Xie**  
**J. J. Allaire**  
**Garrett Grolemund**

CRC Press  
Taylor & Francis Group  
A CHAPMAN & HALL BOOK

<https://bookdown.org/yihui/rmarkdown/>

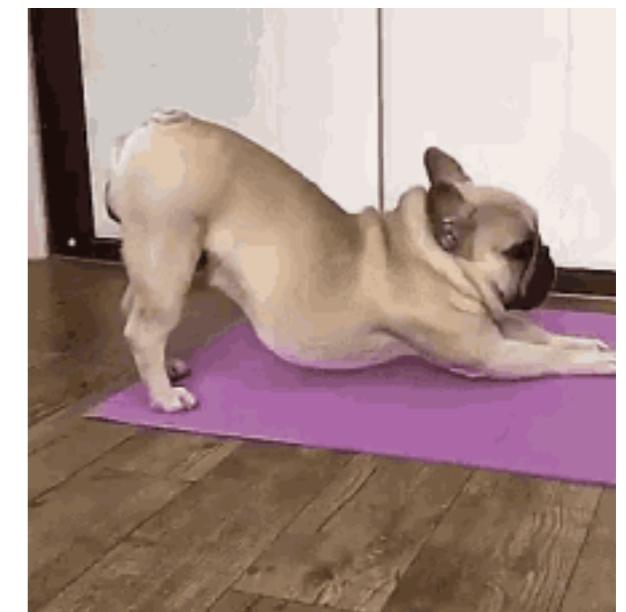
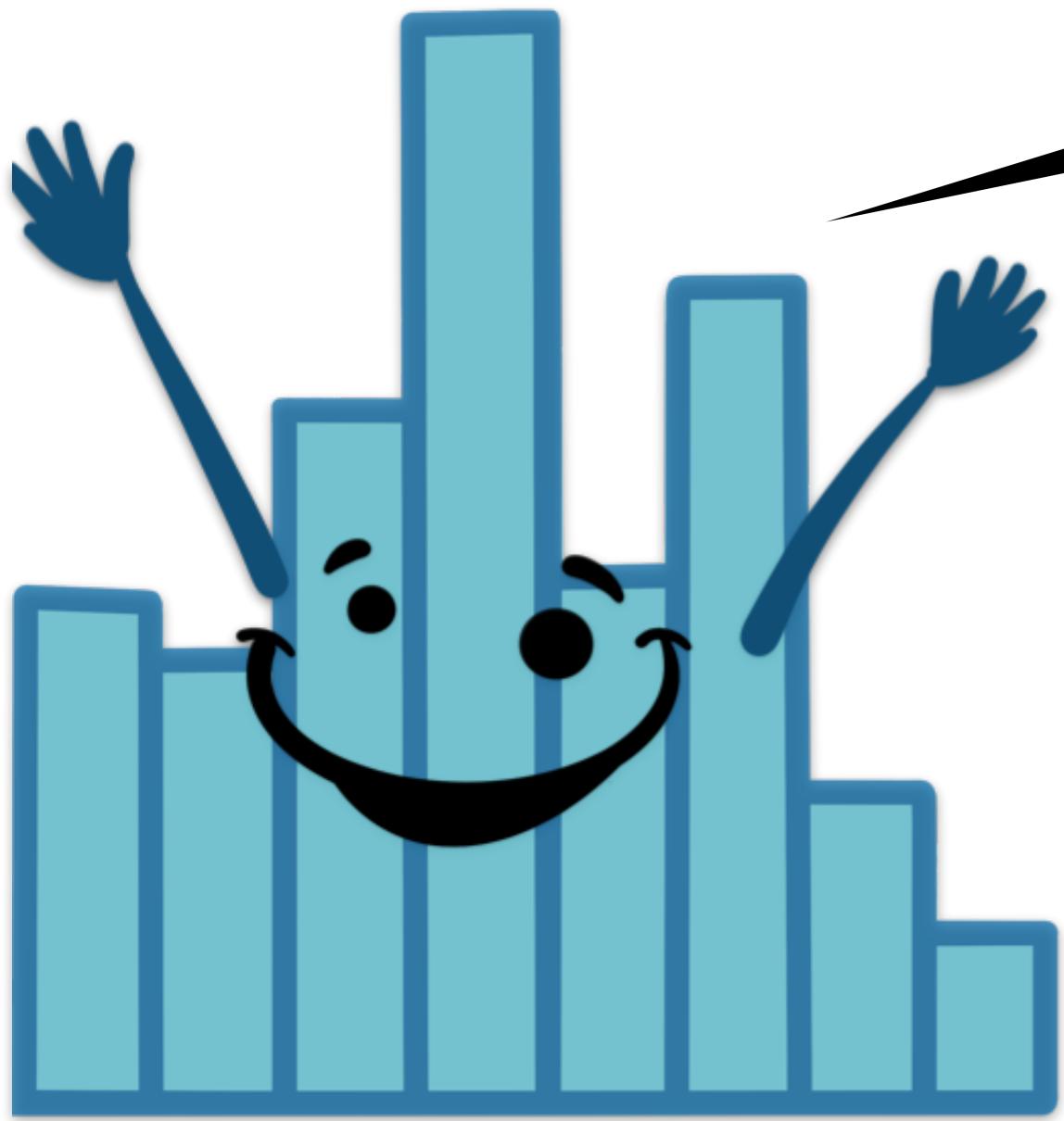
# R Markdown



- if we use a word document ... :
  - figures change
  - results change
  - copy and paste is error prone
- in R Markdown ... :
  - figures and statistics are updated
  - no need for copy and paste
  - everything in one place
  - even better with version control (e.g. via github)

01:00

stretch break!



# **How will we learn?**

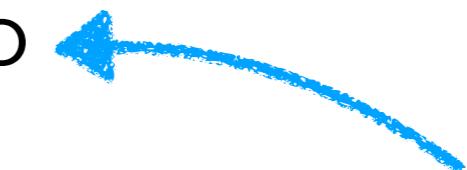
# How will we learn?

- Lectures
- Sections
- Homework
- Midterm exam
- Final project
- Grading
- Office hours

# How will we learn?

- **Lectures**

- Monday, Wednesday, Friday **9:45-11:15am**
- get familiar with R and RStudio
- visualization
- data manipulation
- simulation
- learn about and discuss statistical methods



**if you know R very well  
already, you could skip  
the visualization and data  
wrangling classes**

# How will we learn?

- **Sections**

- section times are **Tuesday 5:30-6:30pm** and **Thursday 2:45-3:45pm** (via zoom for the first two weeks - links are on Canvas)
- **R tutorial** in week 1 ← **recommended if you have little R experience**
- work on the homework assignments together
- ask the TAs questions
- attending sections is **optional** (but **recommended**)

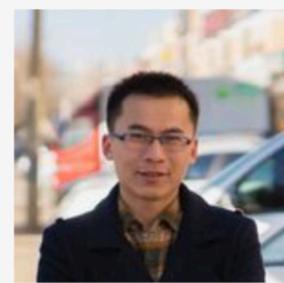
Ari Beller



Sarah Wu



Chengxu Zhuang



# How will we learn?

- Homework

encouraged!

- one assignment per week (7 in total)
- you can work in groups
- clearly indicate who you worked with on your submission
- each group member should write their own code and written responses
- download and submit via Canvas
- you'll submit an RMarkdown file rendered as a pdf
- homework will be available **after class on Fridays**, and is due **Thursday 8pm** the week after

# How will we learn?

- **Homework**
  - late submission policy:
    - you have 5 slip days in total
    - if you submit a homework within 24h of the deadline, this costs 1 slip day (or 2 slip days if it's submitted within 48h, ...)
    - once you have used up all your slip days, a late homework will receive a 0 score

# How will we learn?

- **Midterm exam**
  - like a homework assignment, but:
    - you have to work on it **on your own**
    - will be made available on Friday (2/4) and is due on **Thursday (2/10) at 8pm**
  - you can use any resources you like
  - will be submitted to Canvas as a RMarkdown file

# How will we learn?

- **Final project**
  - you can work in groups of up to 3 members
  - everyone in the group will receive the same grade
  - the expectations for the project scale with the size of the group

# How will we learn?

- **Final project**
  - structure:
    1. short written project proposal
    2. oral presentation
    3. written report

# How will we learn?

- **Final project**

1. short written project proposal

- due Thursday 2/17
- we'll provide an RMarkdown template
- and some examples from prior years

# How will we learn?

- **Final project**

- 2. oral presentation

- during exam week
    - short slideshow of the project
    - groups present together

# How will we learn?

- Final project

- 3. written report

- due **Friday 3/18**
    - ~ 2000 words per group member
    - answer an interesting research question
    - demonstrate what you've learned in class:
      - data wrangling
      - visualization
      - statistical modeling
      - reporting
    - you'll be using github to publish your final project online
    - we'll make some examples available

# How will we learn?

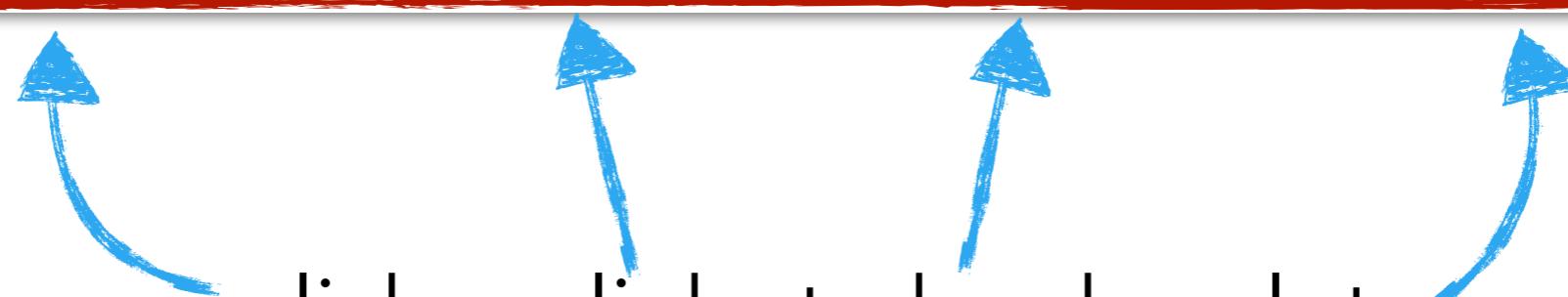
- **Grading**
  - Homework: 40%
  - Midterm: 20%
  - Final project: 40%
    - Proposal: 5%
    - Presentation: 10%
    - Report: 25%

# How will we learn?

- Office hours

## Team

	Tobi Gerstenberg	Ari Beller	Sarah Wu	Chengxu Zhuang
Role	Instructor	Teaching assistant	Teaching assistant	Teaching assistant
Pronouns	he/him	he/they	she/her	he/him
Email (@stanford.edu)	gerstenberg	abeller	sarahawu	chengxuz
Office hours	Monday 1:00-2:00pm	Friday 11:30am-12:30pm	Tuesday 3:00-4:00pm	Tuesday 10:00-11:00am



click on links to book a slot  
(on the course website)

# Tools we will use in class

- Canvas
- Course website
- PollEverywhere
- Datacamp
- Ed Discussion
- Slack
- Github
- Free online books

# Tools we will use in class

- **Canvas**

- **We** will:
  - send announcements
  - upload:
    - slides
    - class notes / code files
    - homework assignments
- **You** will:
  - read announcements :)
  - upload completed homework as RMarkdown files

# Tools we will use in class

- Course website
  - general information
  - schedule with links to materials (will keep being updated)
  - <https://psych252.github.io/>

PSYCH 252: STATISTICAL METHODS

Home Schedule Getting ready Information Book

This course offers an introduction to advanced topics in statistics with the focus of understanding data in the behavioral and social sciences. It is a practical course in which learning statistical concepts and building models in R go hand in hand. The course is organized into three parts: In the first part, we will learn how to visualize, wrangle, and simulate data in R. In the second part, we will cover topics in frequentist statistics (such as multiple regression, logistic regression, and mixed effects models) using the general linear model as an organizing framework. We will learn how to compare models using simulation methods such as bootstrapping and cross-validation. In the third part, we will focus on Bayesian data analysis as an alternative framework for answering statistical questions.

Requirement: [Psych 10](#), [Stats 60](#), or equivalent.

Team

Tobi Gerstenberg	Ari Beller	Sarah Wu	Chengxu Zhuang	
Role	Instructor	Teaching assistant	Teaching assistant	Teaching assistant
Pronouns	he/him	he/they	she/her	he/him
Email (@stanford.edu)	gerstenberg	abeller	sarahawu	chengxuz

# Tools we will use in class

- **PollEverywhere**
  - quick polls in class
  - feedback at the end of class
  - address: [pollev.com/psych252](http://pollev.com/psych252)

# How are you feeling today?



Demonstrate that data are in fact anonymous

# Tools we will use in class

- **DataCamp**

- use your **stanford.edu** address to sign up!
- if you haven't already, sign up here:  
**<https://tinyurl.com/psych252datacamp22>**

**Wednesday, January 13th: Visualization I**

**Content:**

- Get familiar with the RStudio interface.
- Take a look at some suboptimal plots, and think about how to make them better.
- Understand the general philosophy behind ggplot2 – a grammar of graphics.
- Understand the mapping from data to geoms in ggplot2.
- Create informative figures using grouping and facets.

**Resources:**

- [Cheatsheet ggplot2](#)

**Datacamp:**

- [ggplot \(intro\)](#)
- [Reporting](#)
- [visualization best practices](#)

**Reading:**

- [Course notes: Visualization 1](#)
- [Data visualization \(#1\)](#)
- [Data visualization \(#3\)](#)



The DataCamp website homepage features a dark blue header with the DataCamp logo and a search bar. Below the header, a large banner reads "THE SMARTEST WAY TO Learn Data Science Online". A subtext explains that the skills needed for success are changing and that DataCamp helps learners apply data science today. A prominent yellow button says "Start Learning For Free". At the bottom, there are icons for Python, R, SQL, Spark, Git, Shell, and Spreadsheets.

# Tools we will use in class

- **DataCamp:** mixed feelings about it

"I cannot emphasize how much I appreciated the Datacamp access; that was probably crucial to my success in this class (on top of all of the resources and materials) and I'm happy we get to keep using it in the next 3 months."

TECH

## A Multimillion-Dollar Startup Hid A Sexual Harassment Incident By Its CEO — Then A Community of Outsiders Dragged It Into the Light

"Sexual misconduct happens everywhere. But DataCamp was dealing with a community with abnormally high standards and support for each other."

 **Davey Alba**  
BuzzFeed News Reporter

Posted on May 13, 2019, at 9:30 a.m. ET

HOME > ENTERPRISE

### The CEO of DataCamp is leaving his role 'indefinitely' after a sexual misconduct allegation from a former employee came to light

Paige Leskin Apr 25, 2019, 11:46 AM   

 DataCamp Team  
Apr 14th, 2019 

#### A note to our community

Update effective October 16, 2019: DataCamp commissioned an independent, third-party review of the events of October 2017, led by Anurima Bhargava, a former Department of Justice official in the Obama administration, and Pamela Cookson from Working IDEAL. Their report also examines the company's response in the months that followed, and the overall climate and culture at the company. You can read the report in its entirety [here](#). Key findings from the report are also summarized on the DataCamp blog [here](#).

## Don't use DataCamp

A timeline of transgressions & community response

 Daniel McNichol Jul 5, 2020 · 18 min read  

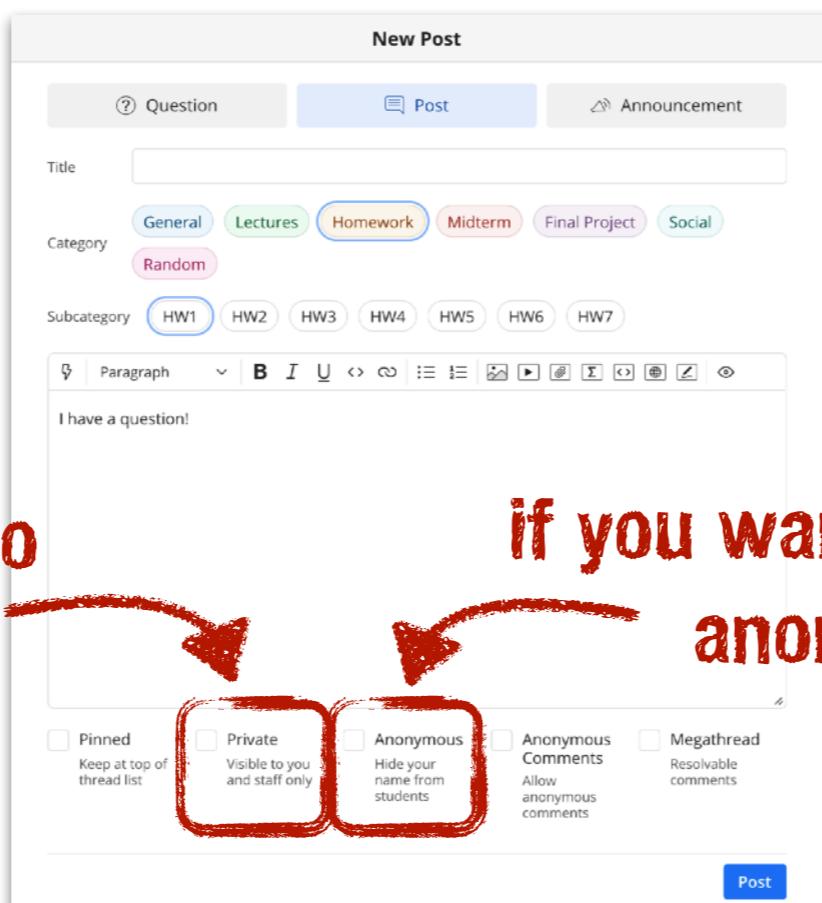
- <https://www.buzzfeednews.com/article/daveyalba/datacamp-sexual-harassment-metoo-tech-startup>
- <https://www.businessinsider.com/datacamp-ceo-jonathan-cornelissen-leave-sexual-misconduct-allegation-2019-4>
- <https://www.datacamp.com/community/blog/note-to-our-community>
- <https://dnlmc.medium.com/dont-use-datacamp-ef04adcf1b7f>

# Tools we will use in class

- **Ed Discussion**

- post your own questions and answer those of your colleagues
- sign up here:

**<https://tinyurl.com/psych252discussion22>**



**if you only want to  
ask the teaching  
team**

**if you want to be  
anonymous**

# Tools we will use in class

- **Slack**

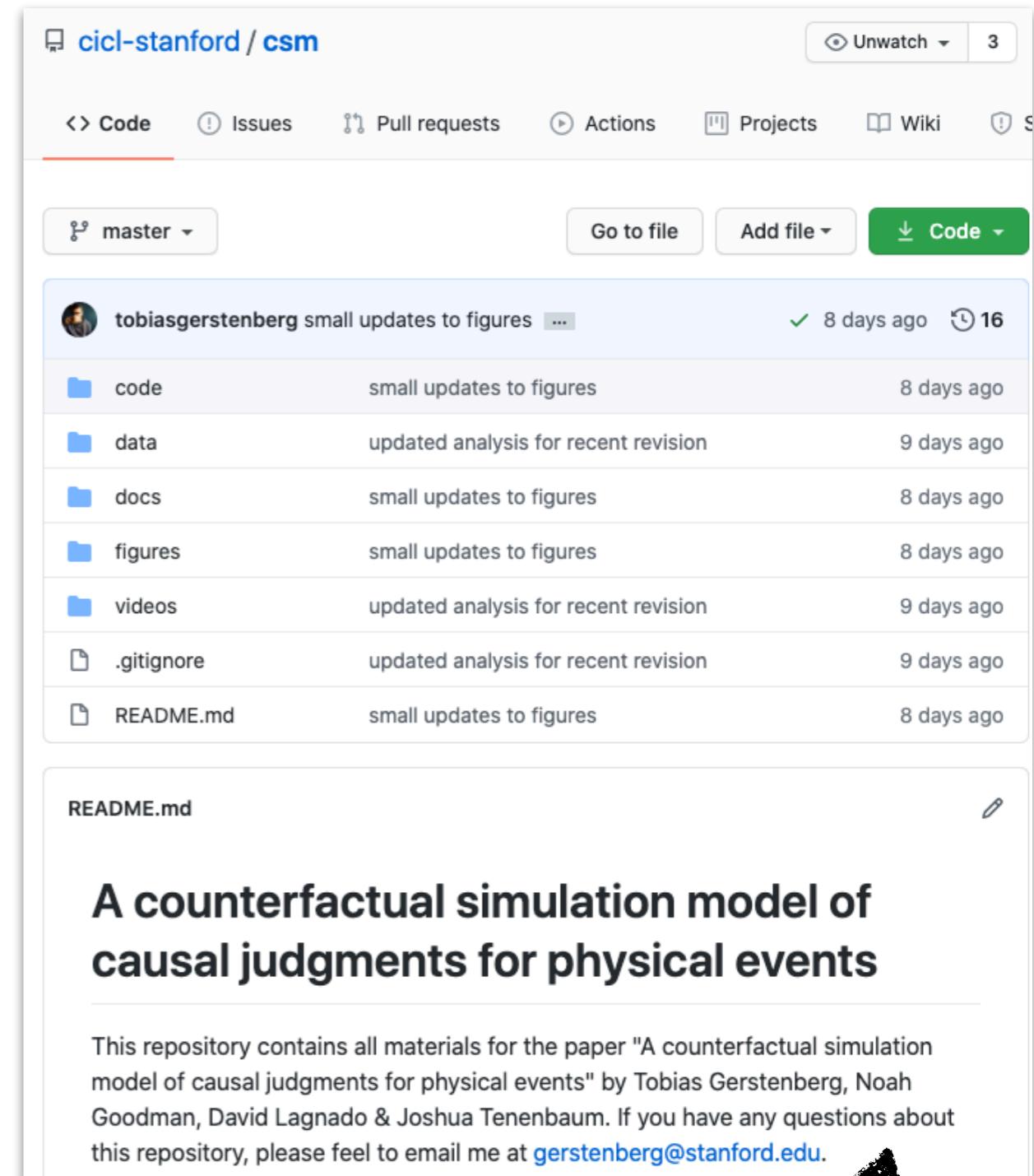
- just a space for you to build community
- the teaching team won't monitor what's happening on slack
- if you haven't received an invite yet, stick around at the end of class and/or send us an email



# Tools we will use in class

- **Github**

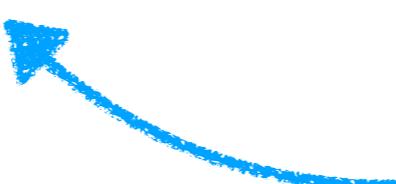
- we will use github for the final project
- it's an essential collaboration tool and critical for contributing to open science



**you want one of these for each paper you publish!**

# Tools we will use in class

- **Free online books**
  - we won't use a text book in class
  - statistics and data science are developing fast and I didn't find a single book that fits the bill
  - **but:** many great, free books online
  - **and:** I will point out suggested readings as we go along
  - **also:** I'll update our course notes as we go along



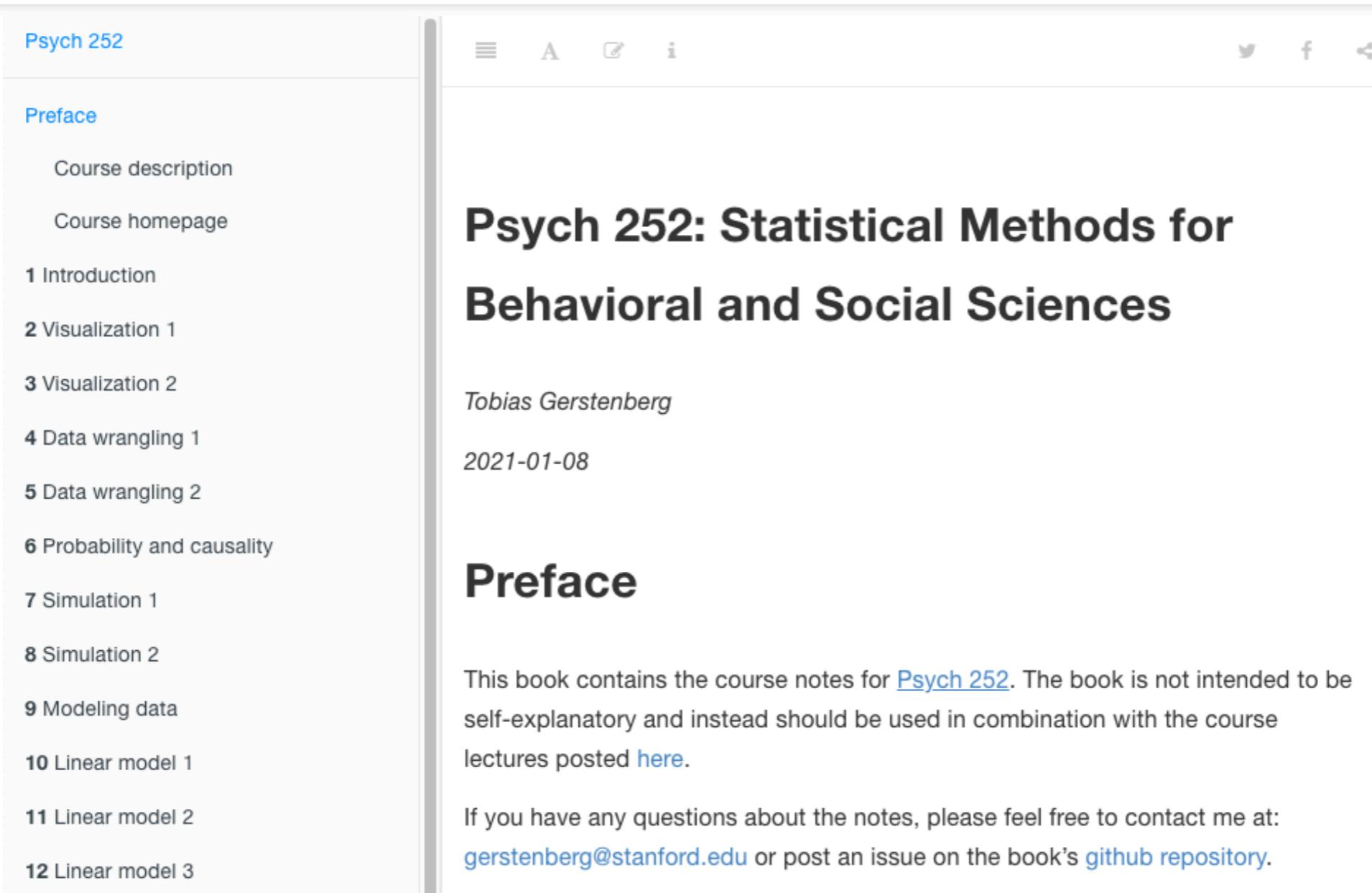
**feedback always welcome!**

# Tools we will use in class

- Free online books

## PSYCH 252: STATISTICAL METHODS

Home Schedule Getting ready Information **Book**



The screenshot shows a website for "Psych 252: Statistical Methods for Behavioral and Social Sciences". The left sidebar contains a navigation menu with links to "Preface", "Course description", "Course homepage", and chapters 1 through 12. The main content area displays the title "Psych 252: Statistical Methods for Behavioral and Social Sciences" by Tobias Gerstenberg, published on 2021-01-08. The "Preface" section is visible, stating that the book contains course notes for Psych 252 and should be used in combination with lectures posted elsewhere. It also provides contact information for questions.

Psych 252

Preface

Course description

Course homepage

1 Introduction

2 Visualization 1

3 Visualization 2

4 Data wrangling 1

5 Data wrangling 2

6 Probability and causality

7 Simulation 1

8 Simulation 2

9 Modeling data

10 Linear model 1

11 Linear model 2

12 Linear model 3

# Psych 252: Statistical Methods for Behavioral and Social Sciences

Tobias Gerstenberg

2021-01-08

## Preface

This book contains the course notes for [Psych 252](#). The book is not intended to be self-explanatory and instead should be used in combination with the course lectures posted [here](#).

If you have any questions about the notes, please feel free to contact me at: [gerstenberg@stanford.edu](mailto:gerstenberg@stanford.edu) or post an issue on the book's [github repository](#).

<https://github.com/psych252/psych252book>

## Ed discussion Canvas

I have a question about the homework.

## Datacamp

I would like to learn more about a topic.

## PollEverywhere

I have an idea how to make this better!

## Slack

I would like to submit my homework on time.

I want to connect with others in my class.



## Course website

I don't remember the schedule.

## Office hours

I'd like to discuss something in person.

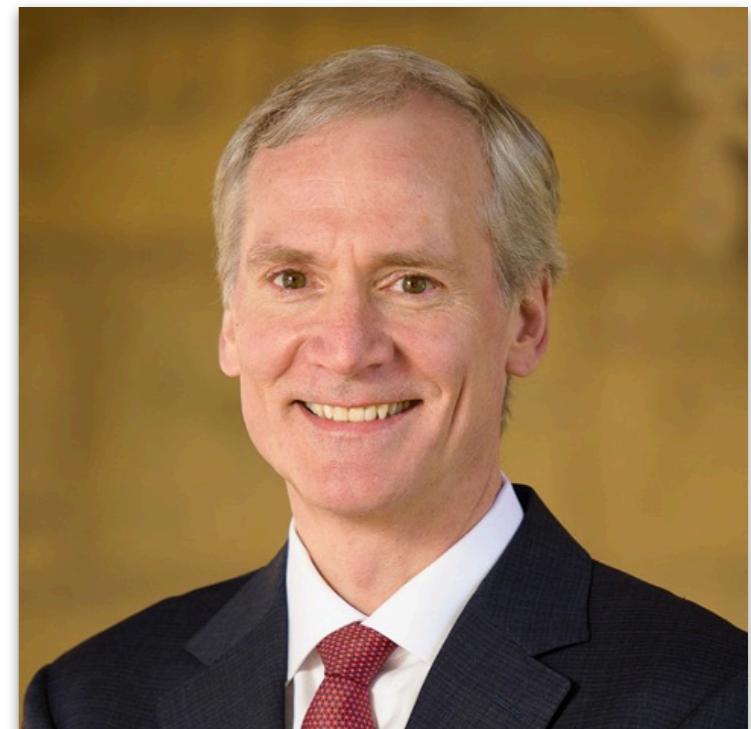
## Anonymous feedback form

Something bothers me ...

# **Some general thoughts**

# Vision for this class

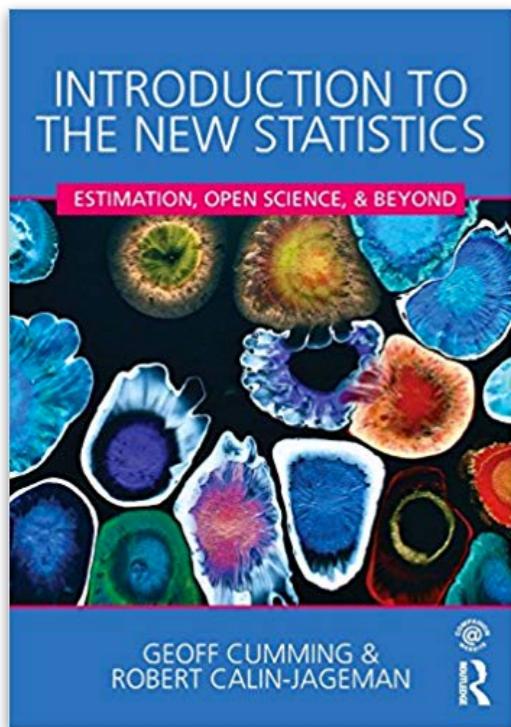
In “[A Vision for Stanford](#)”, university president Marc Tessier-Lavigne states that Stanford wants to be



“an inspired, inclusive and collaborative community of diverse scholars, students and staff, where all are supported and empowered to thrive.”

**Let's try our best together in this class to make this happen!**

# Fear of statistics



**Change a Fixed Mindset to a Growth Mindset**

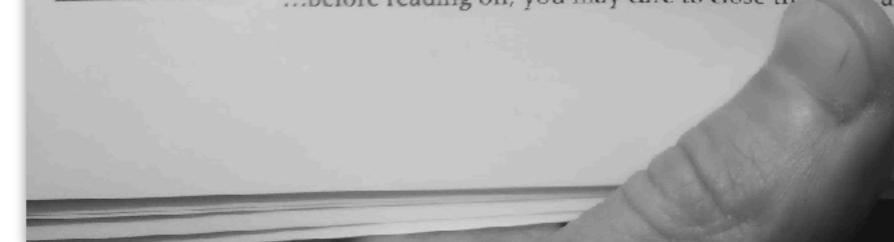
A further key idea is the distinction between a *fixed mindset* and a *growth mindset*. Carol Dweck and colleagues have demonstrated that helping students adopt a growth mindset can be a highly effective way to help them learn better and achieve more. Here's how Dweck describes the two mindsets:

Fixed mindset:  
The belief that my capabilities are more or less fixed, whatever I do.

Growth mindset:  
The belief that effort, persistence, and using good techniques can help me learn more successfully and become more capable.

In a fixed mindset students believe their basic abilities, their intelligence, their talents, are just fixed traits. They have a certain amount and that's that.... In a growth mindset students understand that their talents and abilities can be developed through effort, good teaching and persistence. They don't necessarily think everyone's the same or anyone can be Einstein, but they believe everyone can get smarter if they work at it. (Carol Dweck, [tiny.cc/dwecktalk](http://tiny.cc/dwecktalk))

I've mentioned three important ideas about learning.  
...before reading on, you may care to close the book and practice retrieval...



my thumb

Carol Dweck

Try to adopt a growth mindset!

## fixed mindset:

students believe their basic abilities, their intelligence, their talents, are just fixed traits.

## growth mindset:

students understand that their talents and abilities can be developed through effort, good teaching and persistence

# Fear of statistics

- I don't want to make a mistake ...
- What if my analysis script has bugs?
- Best response: openness!
- We will use **github** for the final projects.

# Stats is important but ...

- it's not the only thing that matters for being a good scientist
- generating research ideas matters
- writing matters
- presenting matters
- ...
- (nobody is very good at all of these things)

# **Feedback**

# What would you like to say about this course to a student who is considering taking it in the future?

- "It will be a lot of work, but it will teach you a lot!"
- "I would absolutely recommend taking this class. I came into this class with random/haphazard stats knowledge (ex. STATS60) and no coding experience, and I left as pretty much a convert to R and model comparison approaches."
- "It can sometimes feel like a ton of work but I advise that **you use office hours at least once**. Even if you think you don't need it. It will help you during the midterm (I struggled a bit with this) and the problem sets."
- "This is a great course to get more comfortable running your own stats and to secure proficiency in R. There is a ton of course material, though, and sometimes that felt really overwhelming. It was **hard to balance actually understanding the stats with bringing my coding skills up** to where they need to be to do this course."
- "Don't take it unless you are an R wizard before the start of the class. It is impossible to get the R skills in 2 weeks. And the class gets REALLY high level just before the midterm (and it just gets worse...)."

# Feedback

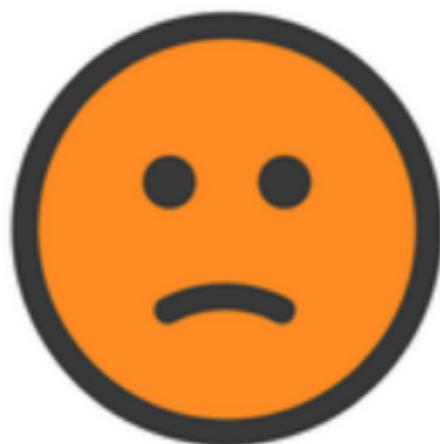
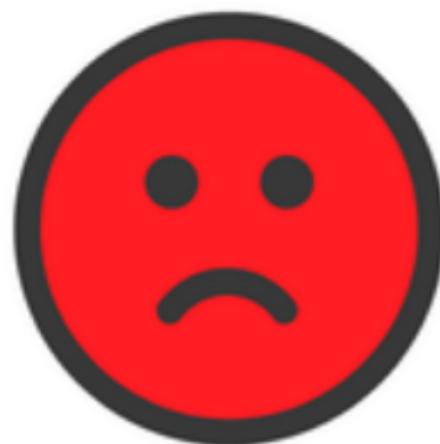
Help us help you!

- ask questions in class and stick around after class
- meet us during office hours
- post questions on Ed discussion (you can post anonymously)
- send us an email (but use Ed discussion for questions about homework or class!)
- use anonymous feedback form (link at the bottom of the course website)  
<https://tinyurl.com/psych252feedback22>

# How was the pace of today's class?

much    a little    just    a little    much  
too        too        right      too        too  
slow      slow                                    fast      fast

# How happy were you with today's class overall?



**What did you like about today's class? What could be improved next time?**

# Thank you to ...

Alexandra Chouldechova

Ben Baumer

Benoit Monin

David Lagnado

Ewart Thomas

Henrik Singmann

Julian Jara-Ettinger

Justin Gardner

Kevin Smith

Maarten Speekenbrink

Matthew Kay

Matthew Salganik

Mika Braginsky

Mike Frank

Mine Çetinkaya-Rundel

Patrick Mair

Peter Cushner Mohanty

Richard McElreath

Russ Poldrack

Stephen Dewitt

Tom Hardwicke

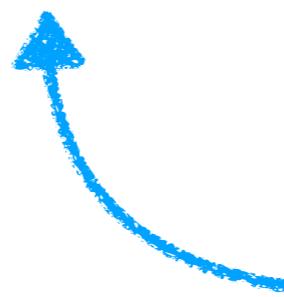
Tristan Mahr

Psych 252 teaching teams

# Thanks!

see you on **Wednesday**

make sure to have  
**R** and **RStudio** up to date



post on Ed discussion if  
you experience any  
problems