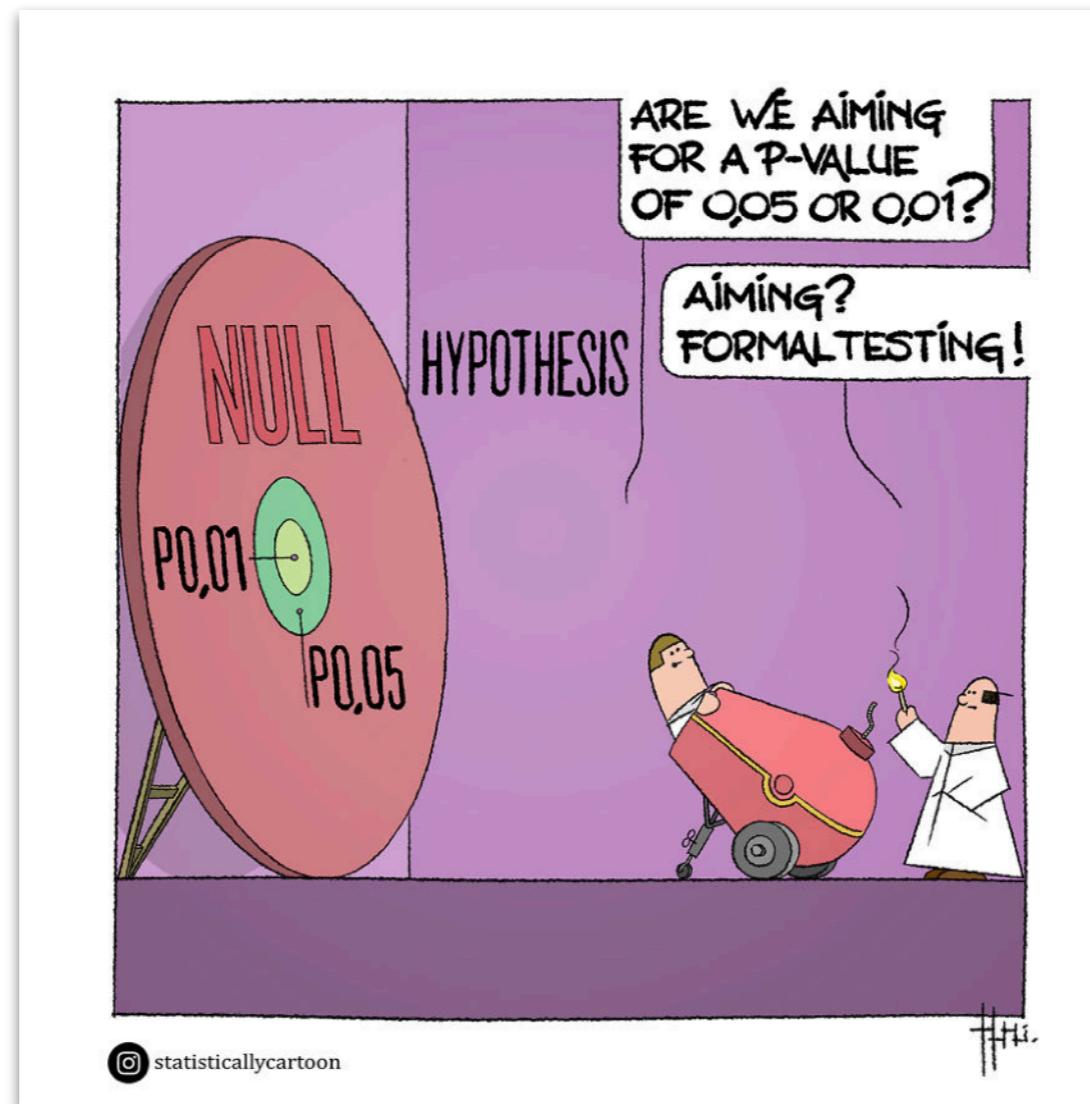


# Introduction



COLLABORATIVE PLAYLIST

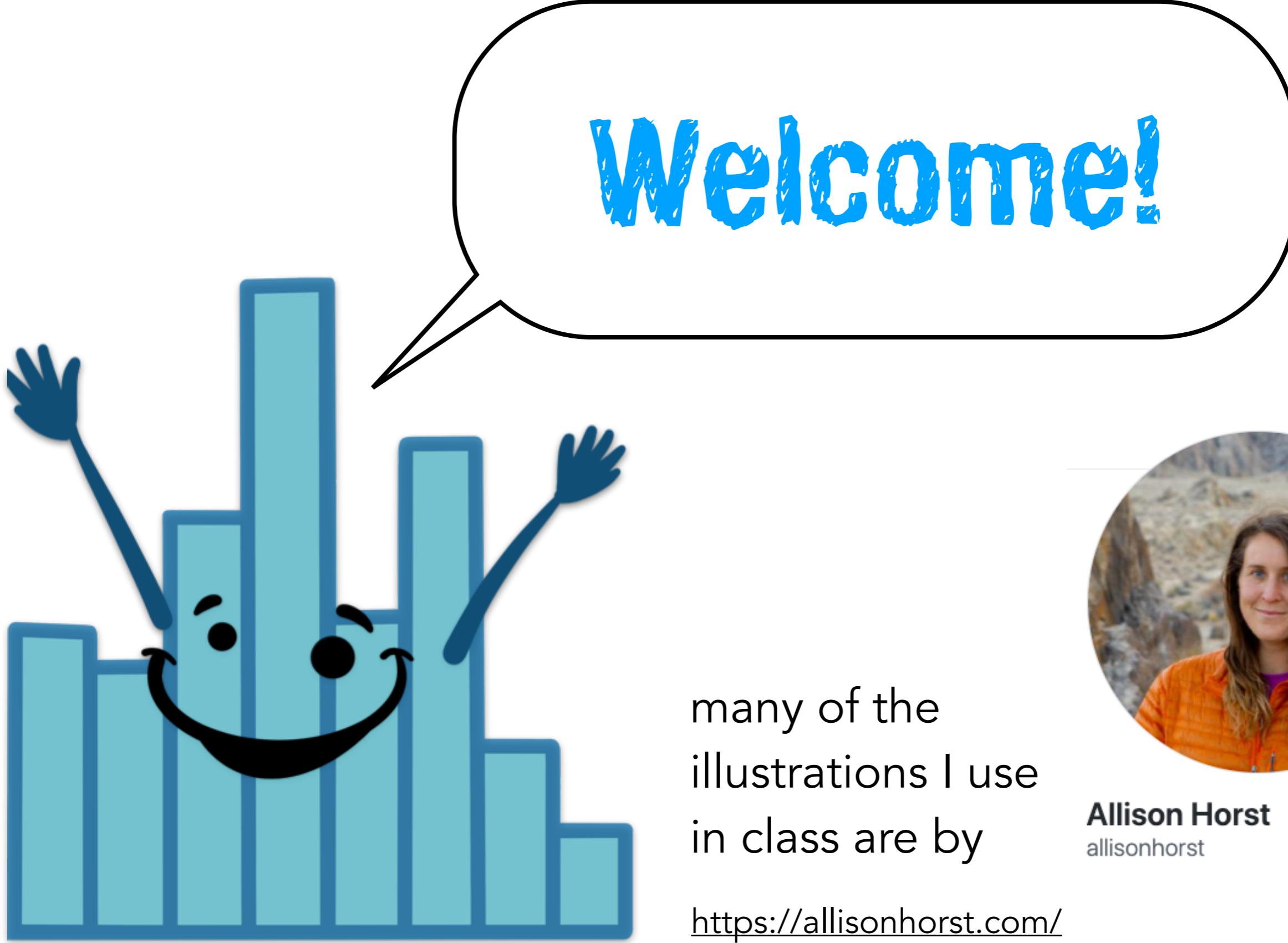
**psych252**

<https://tinyurl.com/psych252spotify24>

PLAY ...

add songs that get you  
into stats mood 

01/08/2024



# Welcome!

many of the  
illustrations I use  
in class are by

<https://allisonhorst.com/>



**Allison Horst**

allisonhorst

# Outline

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

# Introduction

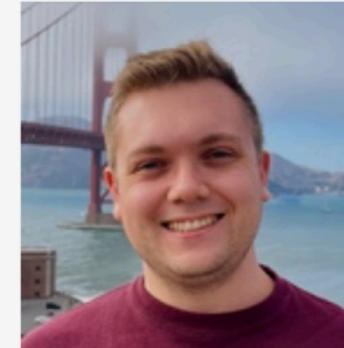
# Introduction to the class

master the tools that empower  
you to do your research



develop statistical literacy

# Introduction to the team

	<b>Tobi Gerstenberg</b>	<b>Ari Beller</b>	<b>Beth Rispoli</b>	<b>Satchel Grant</b>	<b>Shawn Schwartz</b>
					
Role	Instructor	Teaching assistant	Teaching assistant	Teaching assistant	Teaching assistant
Pronouns	he/him	they/them	she/her	he/him	he/him

# Tobi Gerstenberg

**Pronouns:** he/him

**Department:** Psychology

**Interests:**

- running the [Causality in Cognition Lab](#)
- computational modeling

**Fun facts:**

- I like surfing and R



# Ari Beller

**Pronouns:** they/them

**Department:** Psychology

**Interests:**

- computational models for cognition
- intuitive physics
- emotion dynamics

**Fun facts:**

- I live in a co-op



W my niece Lyla :-)

# Beth Rispoli

**Pronouns:** she/her

**Department:** Psychology

## Interests:

- perceptual processing
- cortical organization
- neural architecture

## Fun facts:

- I like adventuring with friends and with corgis



# Satchel Grant

**Pronouns:** he/him

**Department:** Psychology

**Interests:**

- Connectionist modeling
- Mathematical Cognition
- Neural Network Interpretability

**Fun facts:**

- I like mechanical bulls



# Shawn Schwartz

**Pronouns:** he/him

**Department:** Psychology

**Interests:**

- neural readouts of episodic memory
- modeling attention lapsing states
- neurobiology of cognitive aging

**Fun facts:**

- I like to play ultimate frisbee



What did you do during the winter break?

What are you hoping to get out of this class?

**(5 minutes)**

05:00

# 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, a gear icon, a "SEND" button, and a user profile icon with a purple "T". Below the title, there are two tabs: "QUESTIONS" (which is selected) and "RESPONSES". The main content area starts with the survey title and a brief introduction: "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." It also states that data will be made anonymous before release. A note at the bottom asks users to use their Stanford email address for credit. The first question is "Email address \*", with a placeholder "Valid email address" and a link to "Change settings". The second question is "What year of graduate school are you in? \*", with options: 1. Undergraduate, 2. 1-2, 3. 3-4, and 4. 5+. The third question is "What department are you in? \*", with options: Psychology (radio button) and Education (radio button). A help icon is located in the bottom right corner of the form area.

← Psych 252: Introductory ■ ☆ 🔍 ⚙️ SEND ⋮ T

QUESTIONS RESPONSES

## 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/psych252survey24>

# **What will we learn?**



# Learning goals

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

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

# Learning goals

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

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

Understand the philosophy behind null **hypothesis significance testing (NHST)** and **Bayesian statistics** through ...

- running computer simulations and visualizing the results

Formulate **research questions as statistical models** and ...

- determine which models work best for different situations

**Communicate** what you have learned about your data ...

- in short presentations in class, showcasing your visualization and analysis
- in written reports

Contribute to open and **reproducible science** through ...

- adopting good coding practices
- sharing your data and research reports online

# 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

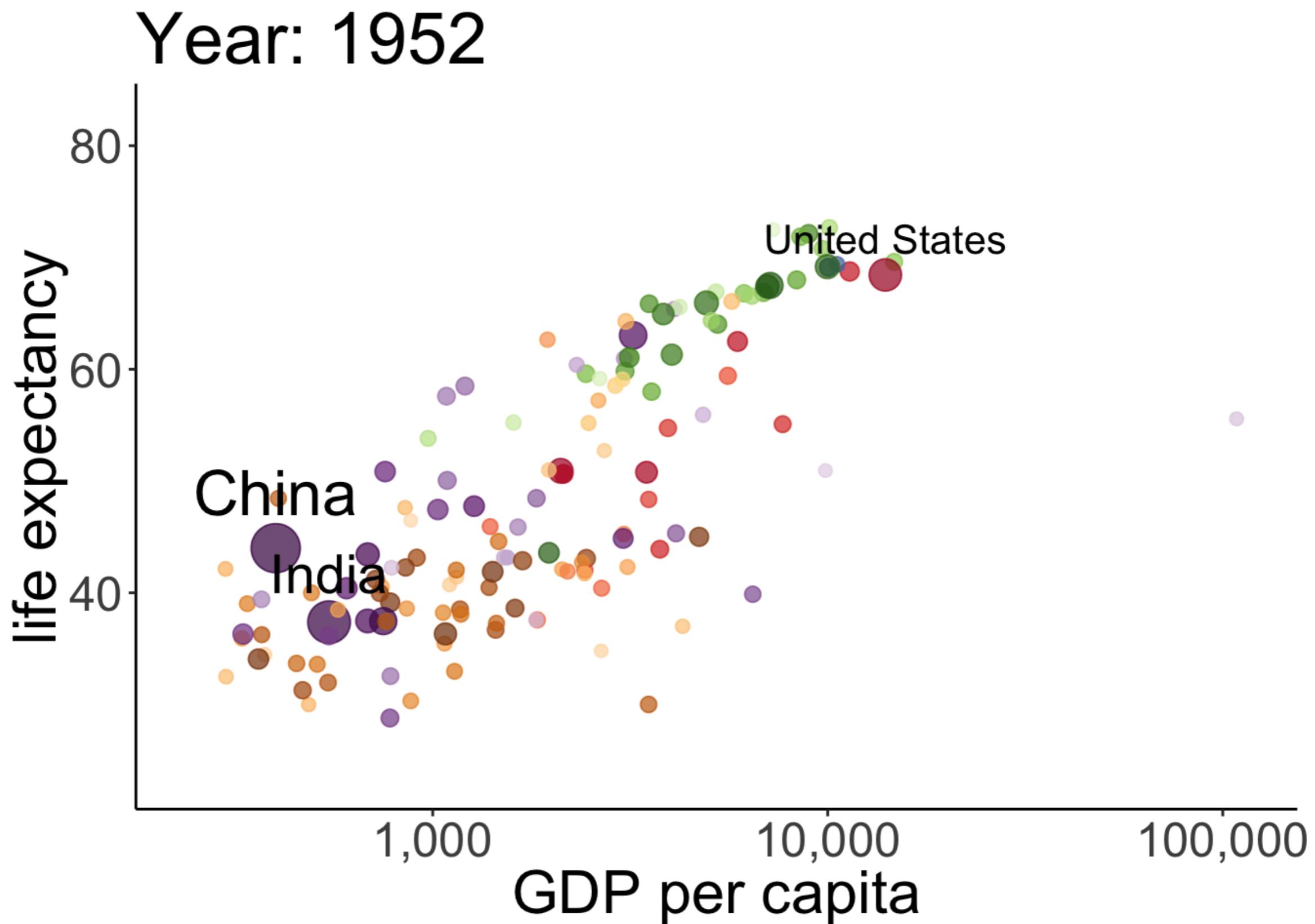
- all the time  
(& Week 10)**
- Reproducible research

1. Use R!

# 4 Reasons for why ?

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

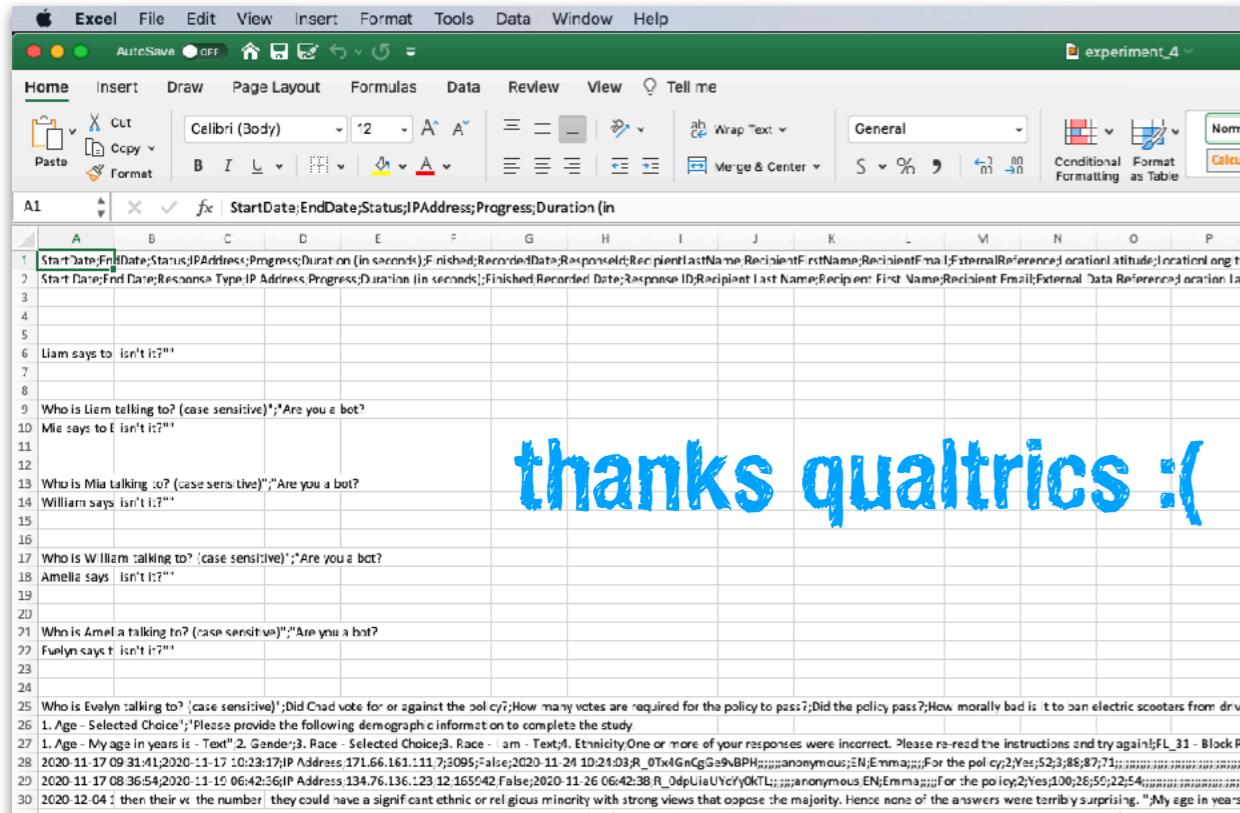
# Why R ? 1. Visualization



# Why R? 1. Visualization

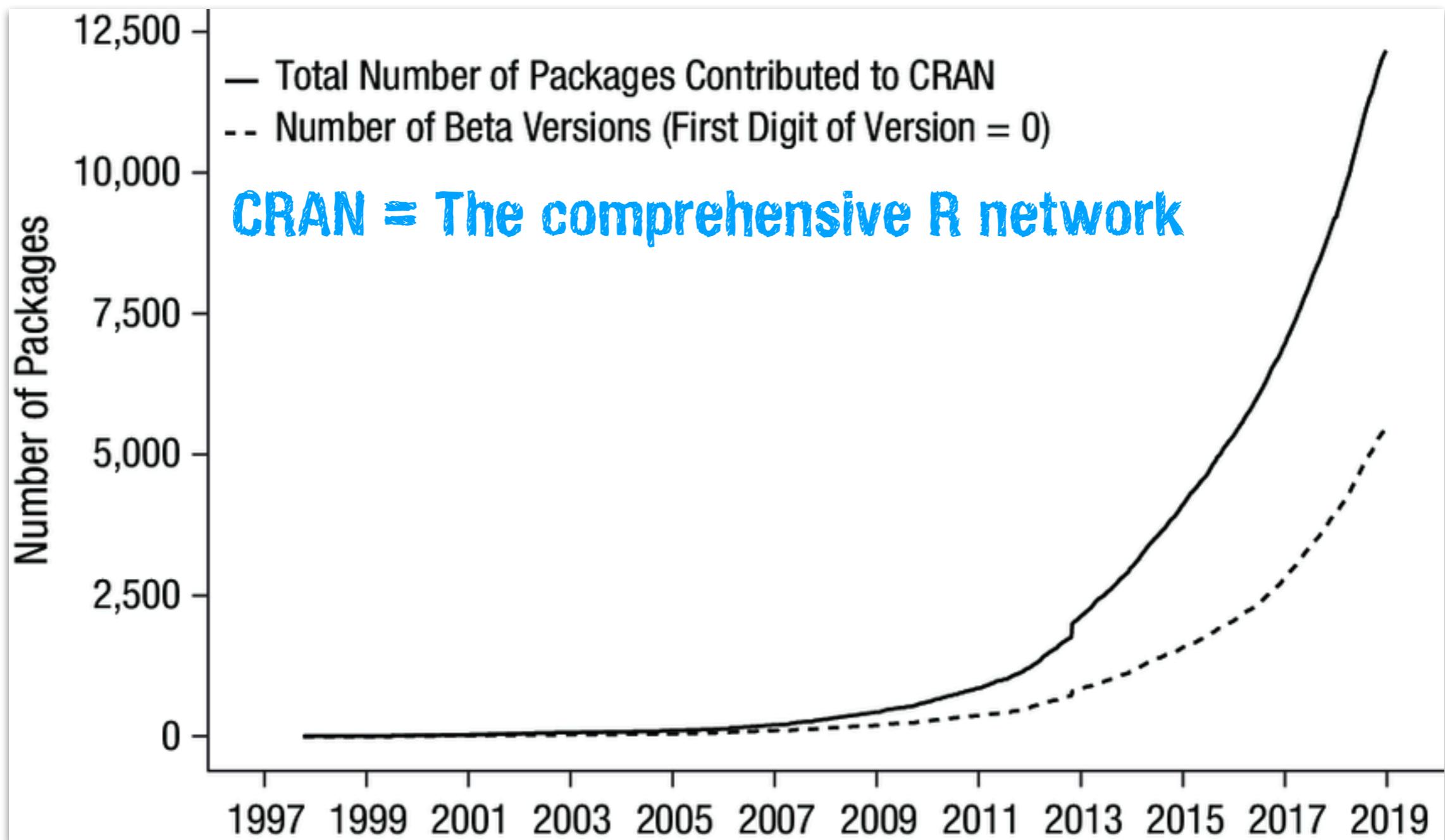
```
1 ggplot(gapminder, mapping = aes(x = gdpPercap, y = lifeExp, size = pop, color = country)) +
2   geom_point(alpha = 0.7, show.legend = FALSE) +
3   geom_text(data = gapminder %>% filter(country %in% c("United States", "China", "India")),
4             mapping = aes(label = country),
5             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")
```

# Why R? 2. Data manipulation



# 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)
```

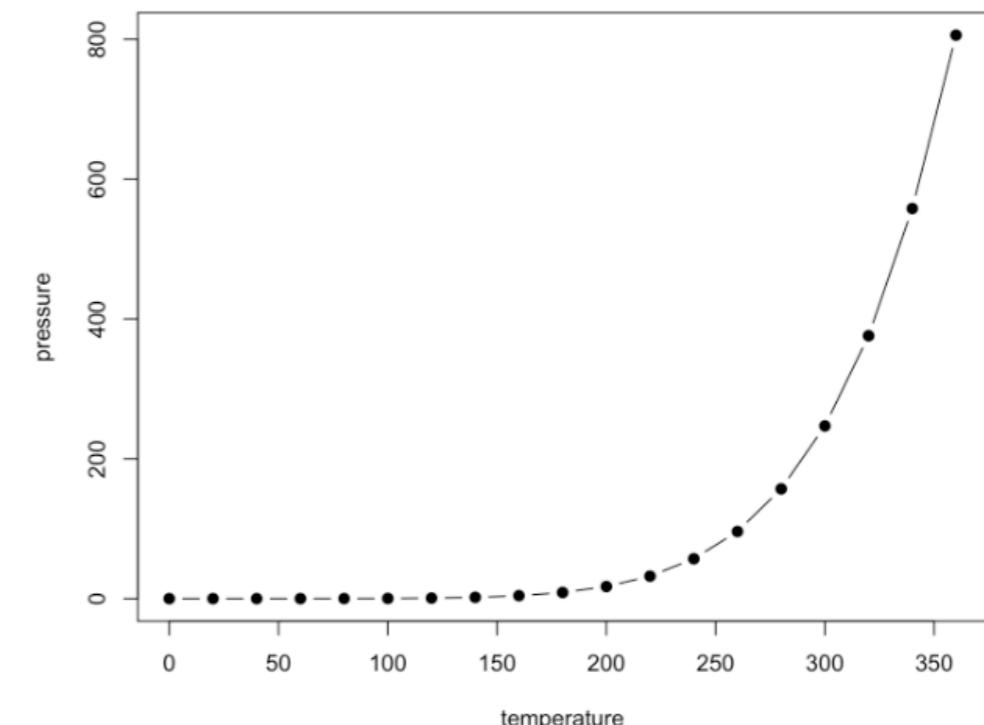


Figure 1.1: Here is a nice figure!

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 title **R for Data Science**, authors **Garrett Grolemund** and **Hadley Wickham**, and a large heading **Welcome**. Below the heading is a detailed description of what the book teaches about data science with R. To the right of the text is a thumbnail image of the book cover, which features a green parrot (Kākāpō) and the title **R for Data Science**.

**R for Data Science**

*Garrett Grolemund*  
*Hadley Wickham*

## Welcome

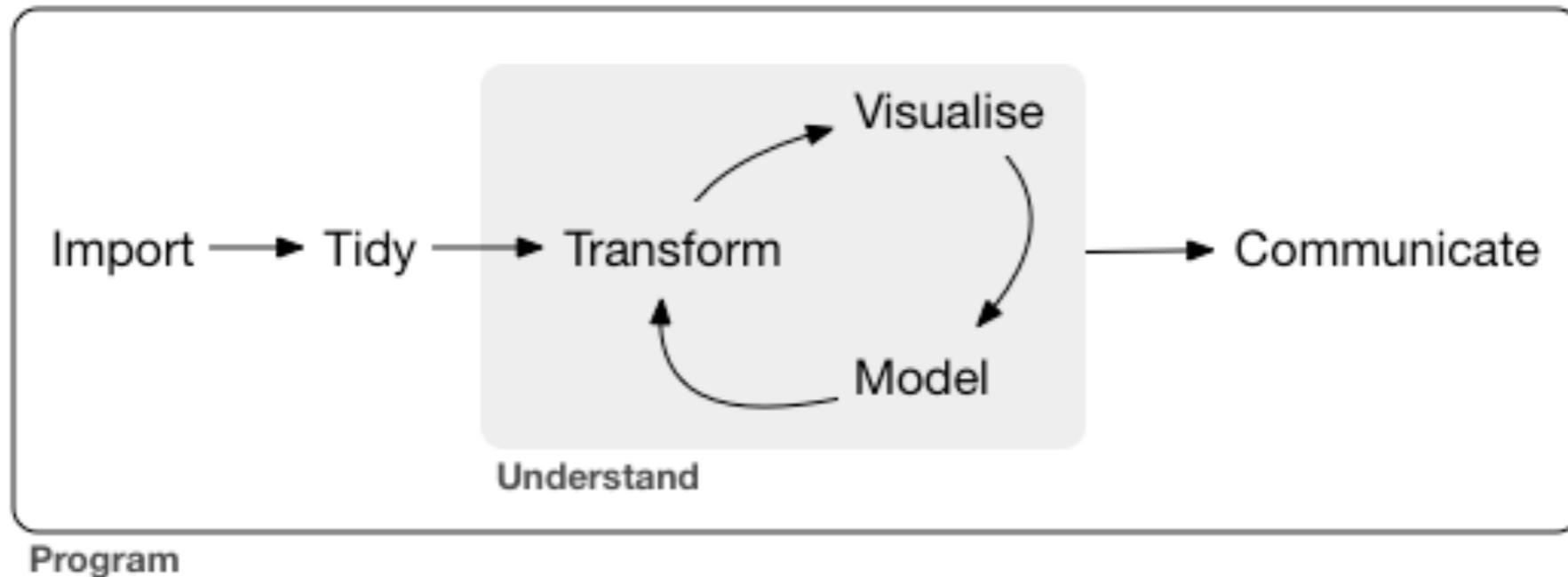
This is the website for “**R for Data Science**”. This book will teach you how to do data science with R: You’ll learn how to get your data into R, get it into the most useful structure, transform it, visualise it and model it. In this book, you will find a practicum of skills for data science. Just 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.

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

# Why R? Tidyverse

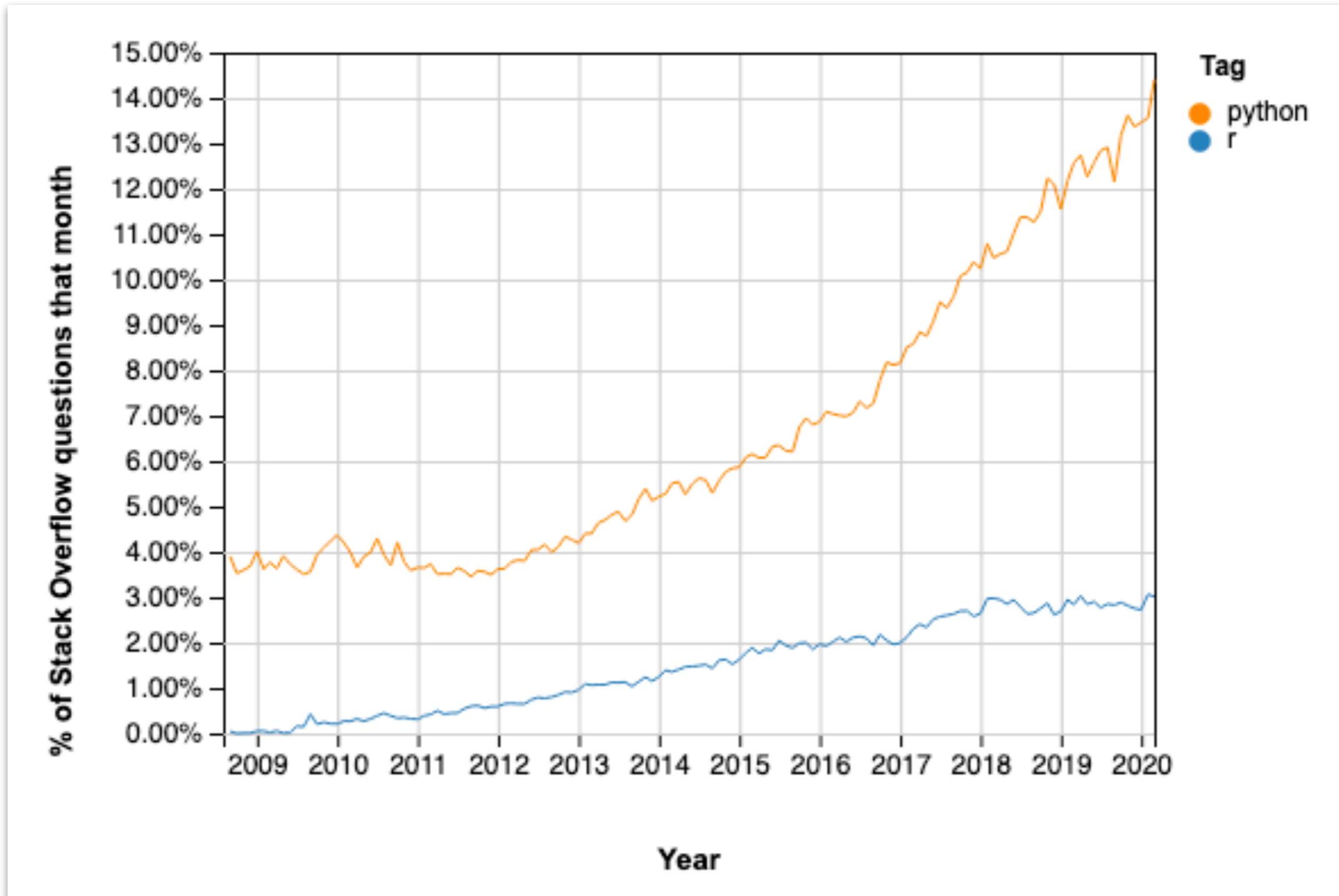


- *Import*: get data into R
- *Tidy*: clean and format the data
- *Transform*: select variables, create new ones, group and summarize
- *Visualize*: "look" at the data in different ways
- *Model*: answer questions about the data
- *Communicate*: write reproducible research reports

# 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 ™ ?



# But what about python™ ?

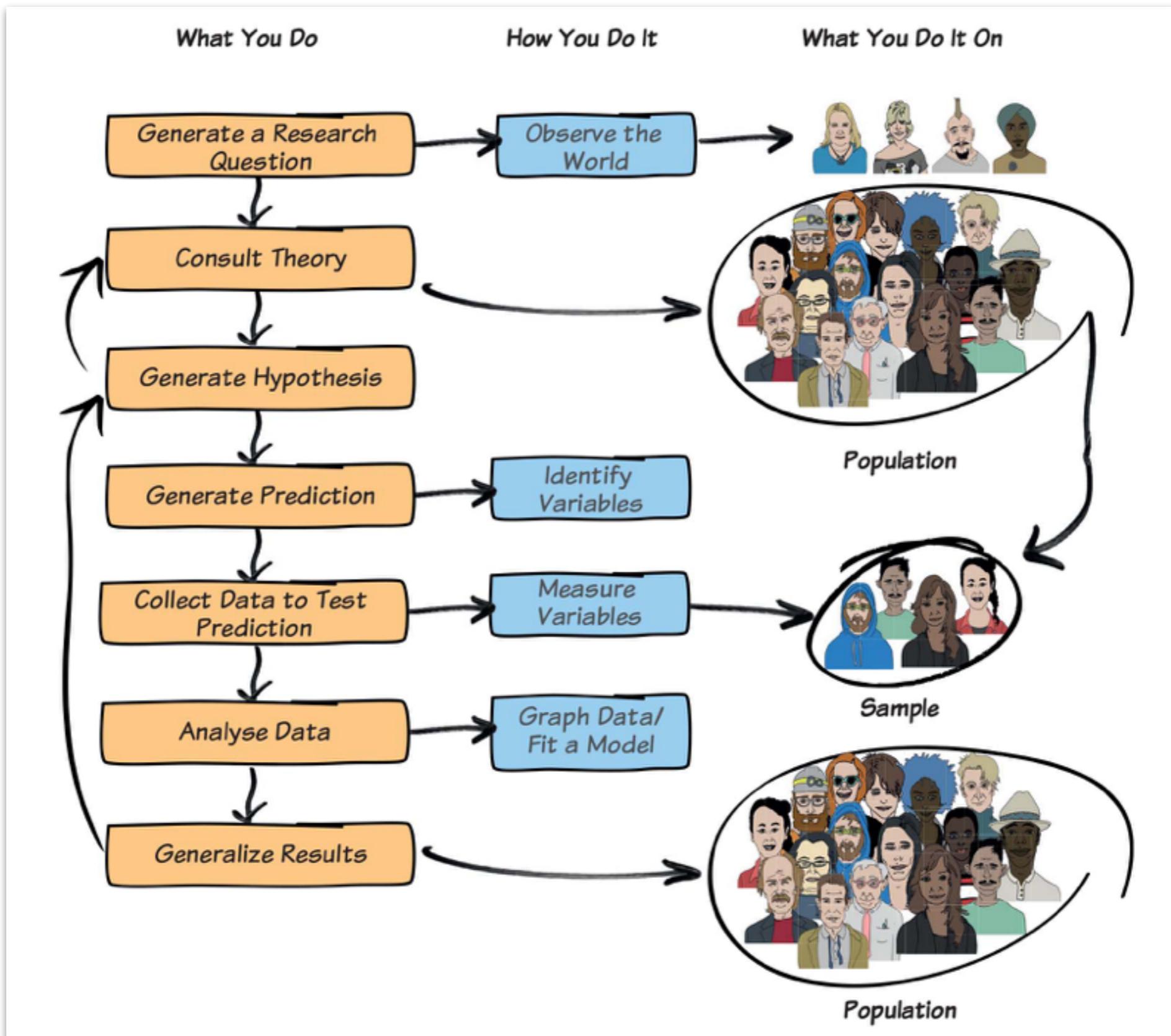
- You should learn python, too! :)
- Python is great for machine learning (e.g. deep learning).

- visualization
- data manipulation
- statistical modeling
- reporting



## 2. Build models

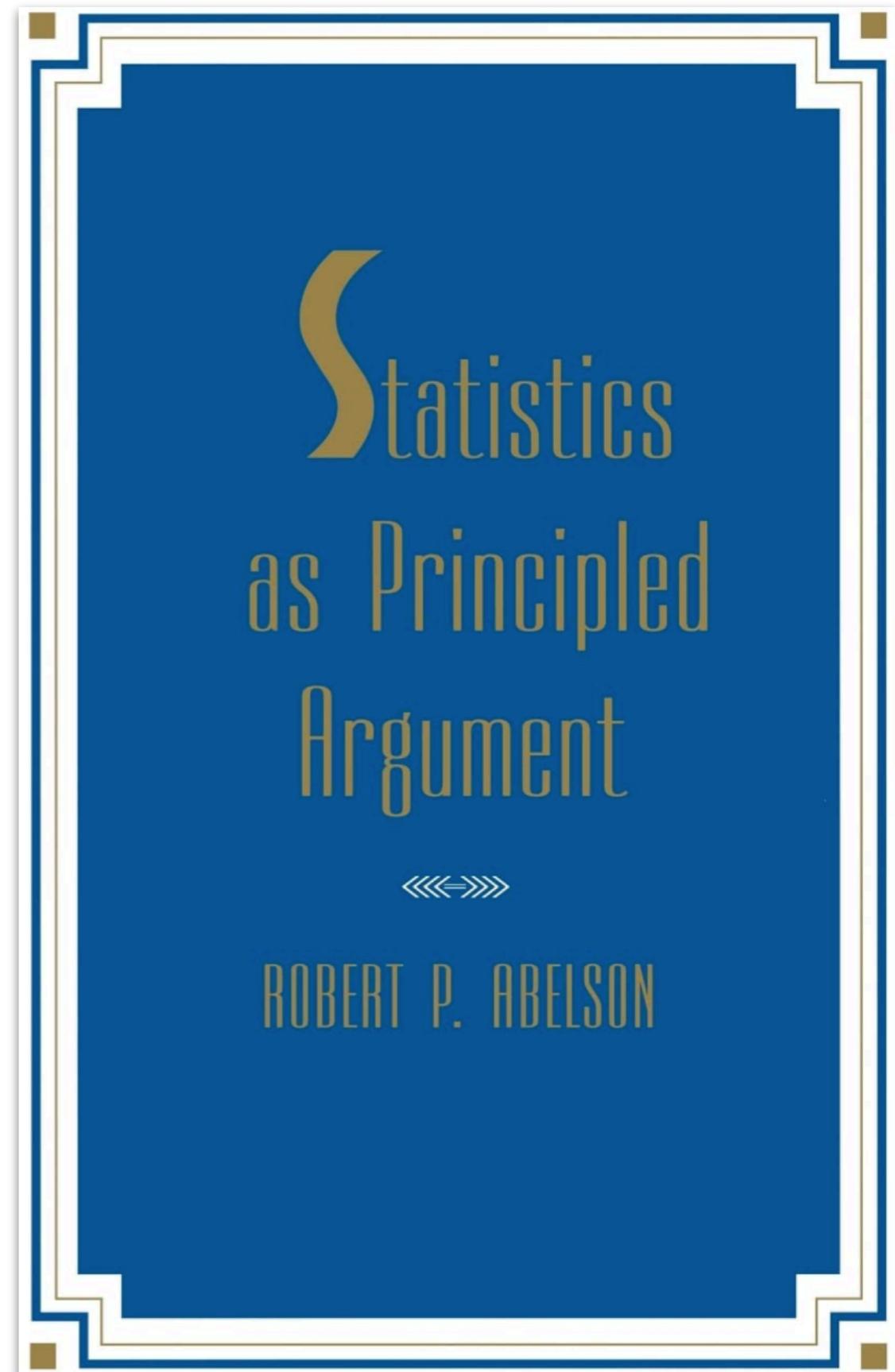
# The research process



# 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

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

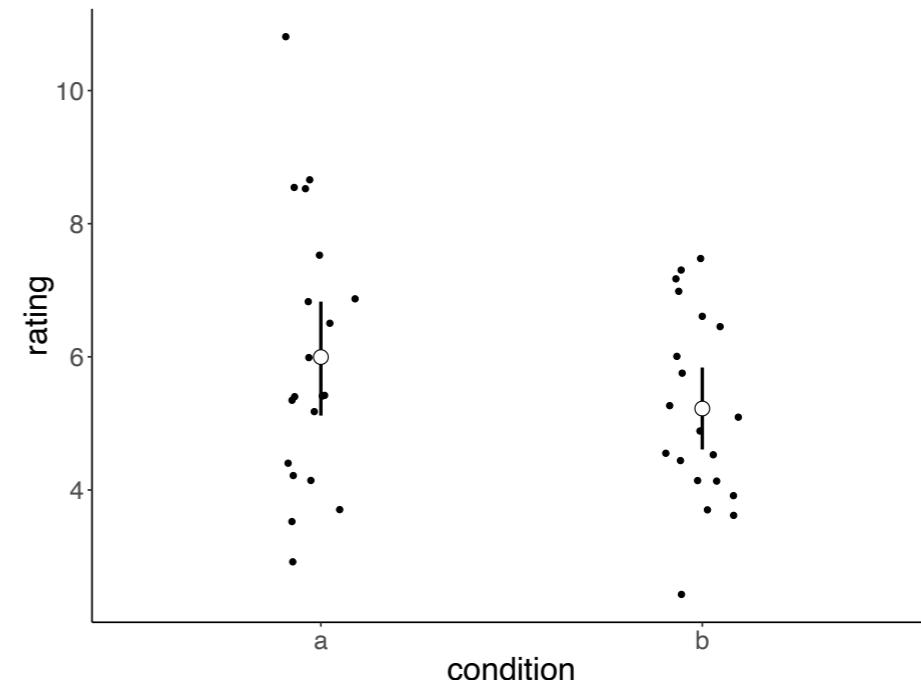
(google search "complicated statistical equation")

we'll do

```

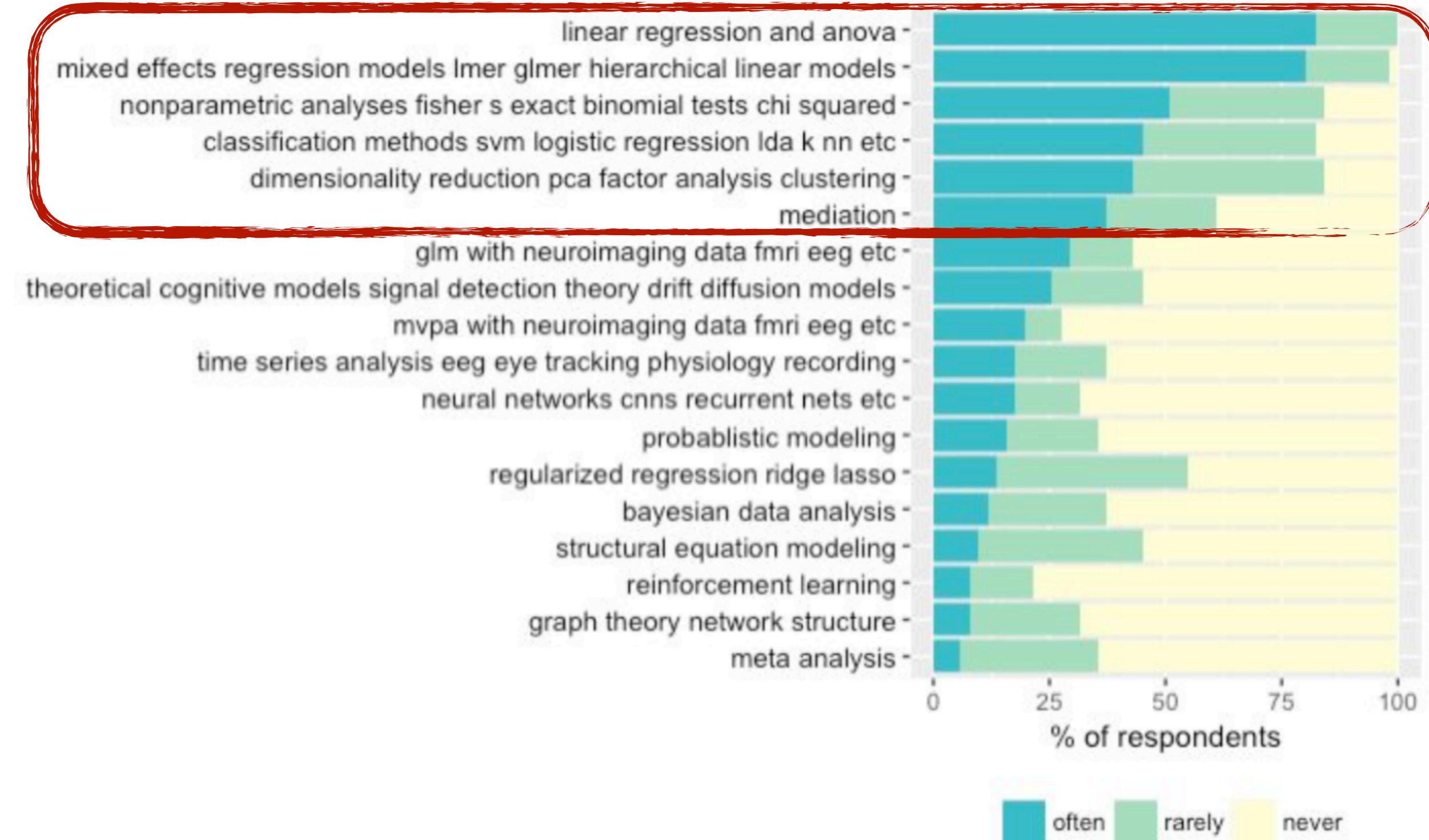
1 set.seed(0)
2 df.data = tibble(
3   a = rnorm(20, mean = 6, sd = 2),
4   b = rnorm(20, mean = 5, sd = 2)
5 ) %>%
6 gather("condition", "rating")
7
8 df.data %>%
9 group_by(condition) %>%
10 summarize(rating.mean = mean(rating),
11           rating.sd = sd(rating)) %>%
12 kable()
13
14 # calculate the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

```

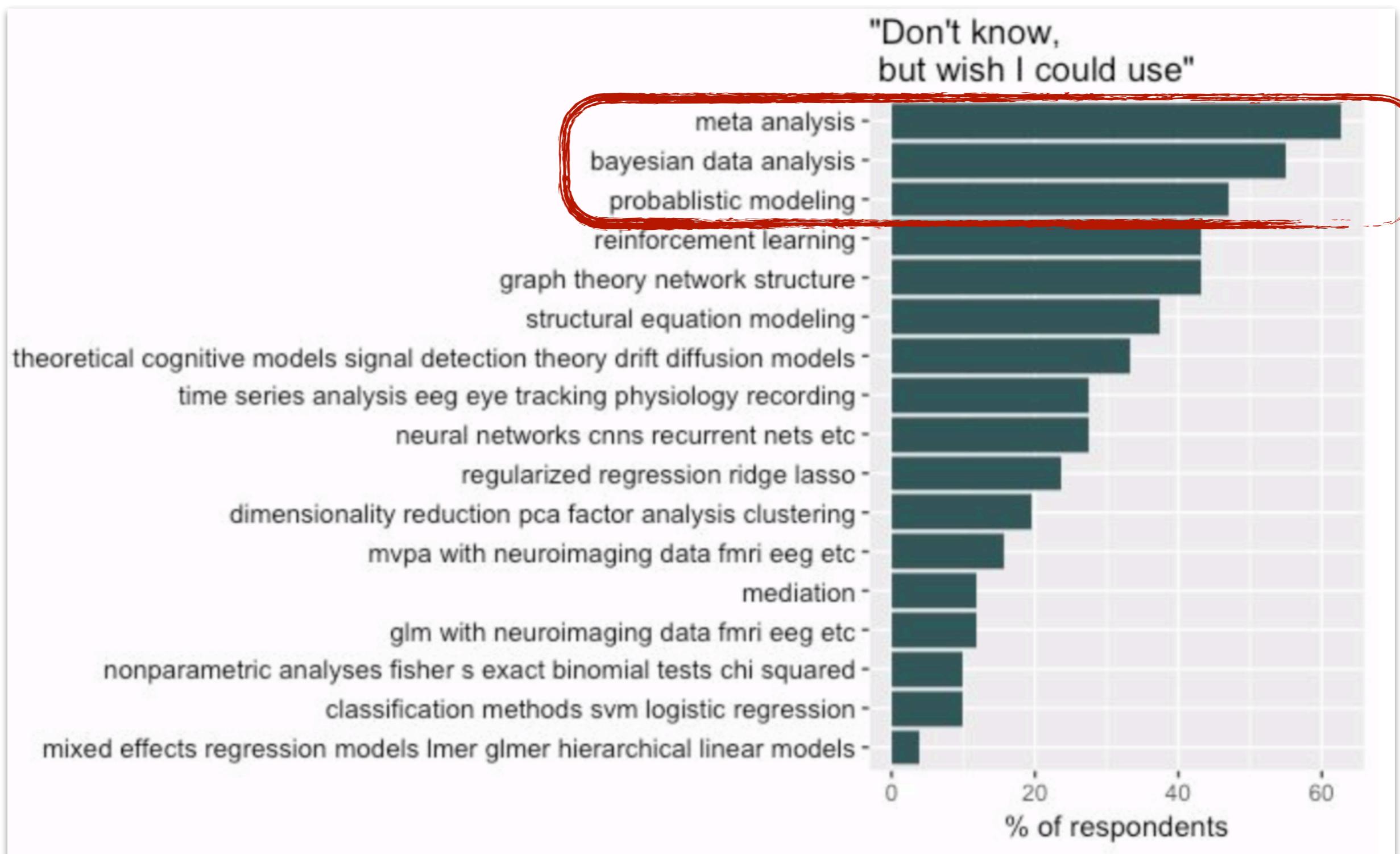


# Psychology grad student survey

“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:

- Science magazine cover: "Many Psychology Findings Not as Strong as We Thought, Study Says" (New York Times)
- Economics (n=18) vs Psychology (n=100) replication rates: 36% overall
- Journal replication rates: Journal of Personality and Social Psychology: Social (23%), Journal of Experimental Psychology: Learning, Memory, and Cognition (48%), Psychological Science, social articles (29%), Psychological Science, cognitive articles (53%)
- Effect size reduction and independent replication p-value funnel plots.
- Cartoon: "REPLICATION CRISIS".
- NOBA Project: "The Replication Crisis in Psychology" by Edward Diener and Robert Biswas-Diener.

Other results include:

- No Evidence for a Replicability Crisis ... projects.iq.harvard.edu
- replication crisis in Psychology real ... theneuroeconomist.com
- The Replication Crisis in Psychology | Noba nobaproject.com
- Are internal replications the solution ... brainsidea.wordpress.com
- What is Replication Crisis? | Popular ... popsci.com
- Teaching High School Psychology ... teachinghighschoolpsychology.blogspot.com
- How to fix psychology's repl... researchgate.net
- Replication Crisis in Science ... simplystatistics.org
- The Reproducibility Crisis in ... slideshare.net
- The Reproducibility Crisis in ... slideshare.net
- psychological science ... science.sciencemag.org
- Replication Crisis Overblown ... thecut.com
- Health News Digest ... bfm.my
- Are conceptual replications part of the ... pigee.wordpress.com
- The Replication Crisis in Psychology | Noba nobaproject.com
- How will stereotype threat get ... progressfocused.com
- Replication | TOK Topics toktopics.com
- psychological science ... science.sciencemag.org
- Replication Crisis ... cambridge.org

Diagram illustrating the 6 Principles of Open Science:

- Open Data
- Open Source
- Open Access
- Open Methodology
- Open Peer Review
- Open Educational Resources

Figure showing mean items solved (adjusted by SAT) for Black Subjects and White Subjects under RACE PRIME and NO RACE PRIME conditions.

Figure illustrating the Reproducibility Problem: Large-scale replication studies like the one performed by Rutter and Niles have begun to reveal that many of the published results in psychology are not replicable, making it difficult to confirm.

Table of replication studies:

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, and 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

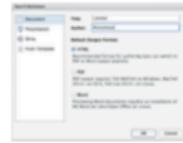
### What is R Markdown?

**.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, Go to R Functions, Publish.
- Left Sidebar:** R Markdown (report.Rmd), Knit HTML, set preview location, insert code chunk, run code chunk(s), go to code chunk, publish, show outline.
- Middle Area:** R Markdown document content:

```
1 ---  
2 title: "R Markdown"  
3 author: "RStudio"  
4 output: 2  
5   html_document:  
6   toc: TRUE  
7 ---  
8   
9 {r setup, include=FALSE}  
10 knitr::opts_chunk$set(echo = TRUE)  
11   
12 # R Markdown  
13   
14 This is an R Markdown document.  
15 Markdown is a simple formatting  
16 syntax for authoring HTML, PDF,  
17 and MS Word documents.  
18   
19 {r cars}  
20 summary(cars)  
21   
22   
23 For more details on using R Markdown  
24 see <http://rmarkdown.rstudio.com>.
```
- Right Panel:** Find in document, synch publish button to accounts at rpubs.com, shinyapps.io, RStudio Connect, Reload document.
- Bottom Area:** R Markdown document content:

```
summary(cars)
```

	speed	dist
## Min.	4.0	Min. : 2.00
## 1st Qu.	12.0	1st Qu.: 26.00
## Median	15.0	Median : 36.00
## Mean	15.4	Mean : 42.98
## 3rd Qu.	19.0	3rd Qu.: 56.00
## Max.	25.0	Max. :120.00

For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
- Console:** R Markdown, report.Rmd, report.html
- File Explorer:** Files, Plots, Packages, Help, Viewer, report.Rmd, report.html

**The R Series**

# R Markdown

## The Definitive Guide

A comprehensive guide to R Markdown, the open-source framework for reproducible research and dynamic documents. This definitive guide covers everything you need to know to get started with R Markdown, from basic syntax to advanced features like R Markdown for data science and R Markdown for presentation.

The book is available in several formats:

- HTML**: A green icon with a large number 5.
- PDF**: A green icon with a small chart.
- TUFTÉ**: A green icon with a world map.
- LATEX**: A green icon with a document.
- ePUB**: A green icon with a book.
- kindle**: A green icon with a tablet.

**Authors:** Yihui Xie, J. J. Allaire, Garrett Grolemund

**Publisher:** CRC Press, Taylor & Francis Group

A CHAPMAN & HALL BOOK

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

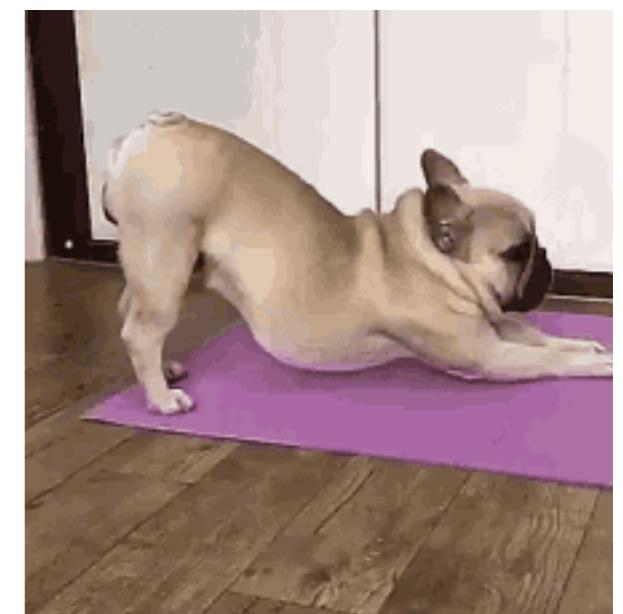
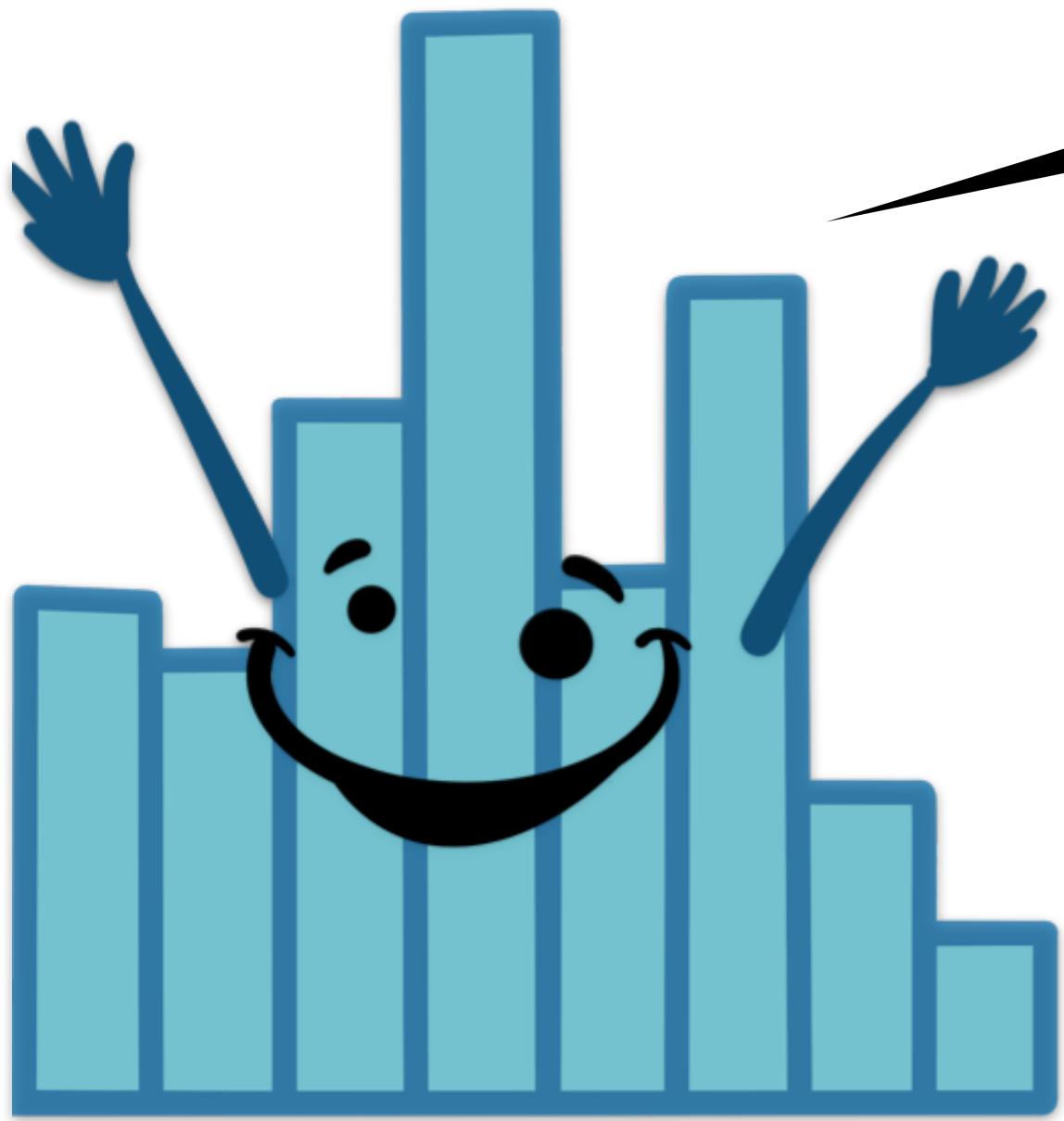
# R Markdown



- 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!





Ray ✅

@ray4tesla

Subscribe

...

80% of Covid patients put on ventilators died. This is probably most shocking revelation unknown to the public.



2:38 PM · Oct 31, 2023 · 10.1M Views

2.2K

2.5K

4.5K

565





## Post



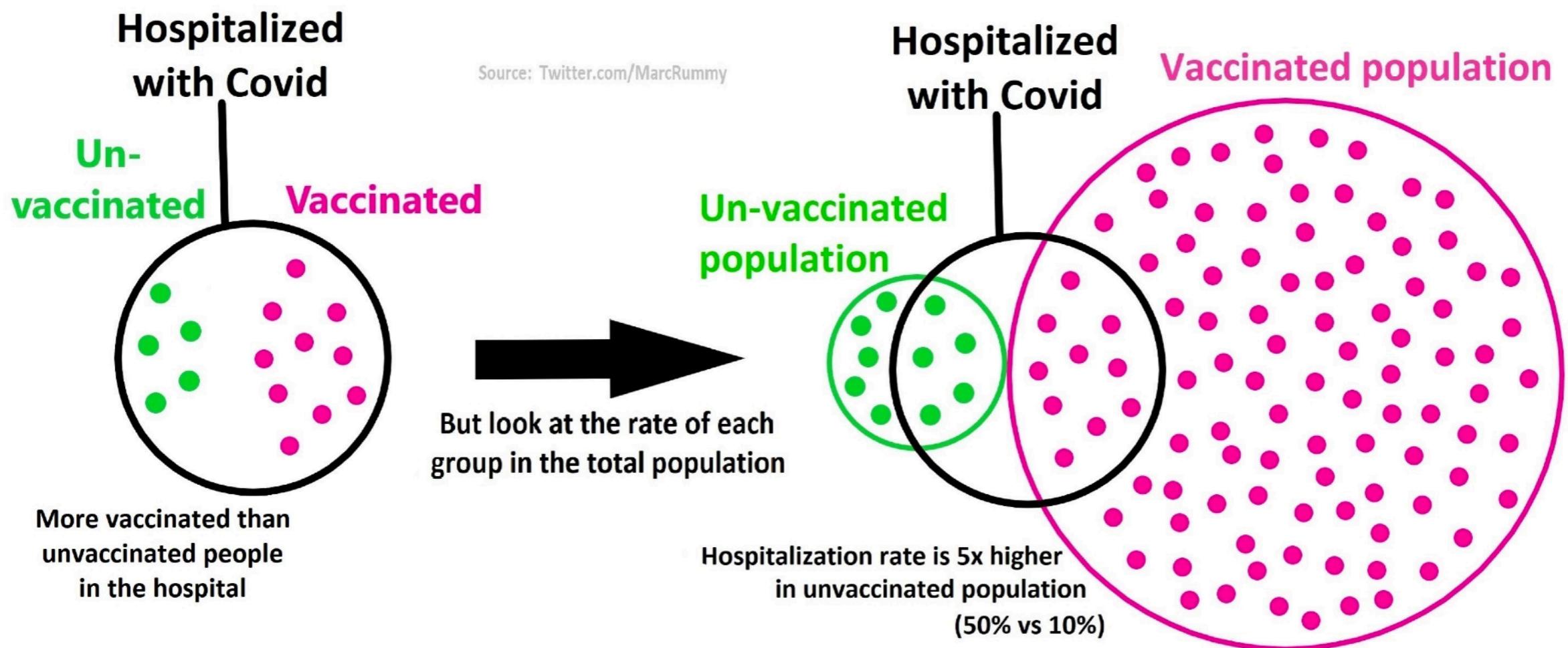
**Daniël Lakens**

@lakens

...

This is why people need basic training in logic (in this case about the positive predictive value). After selecting patients on 'needing to go on a ventilator' many will die, whatever you do.

More vaccinated people are  
hospitalized with COVID than  
unvaccinated people!

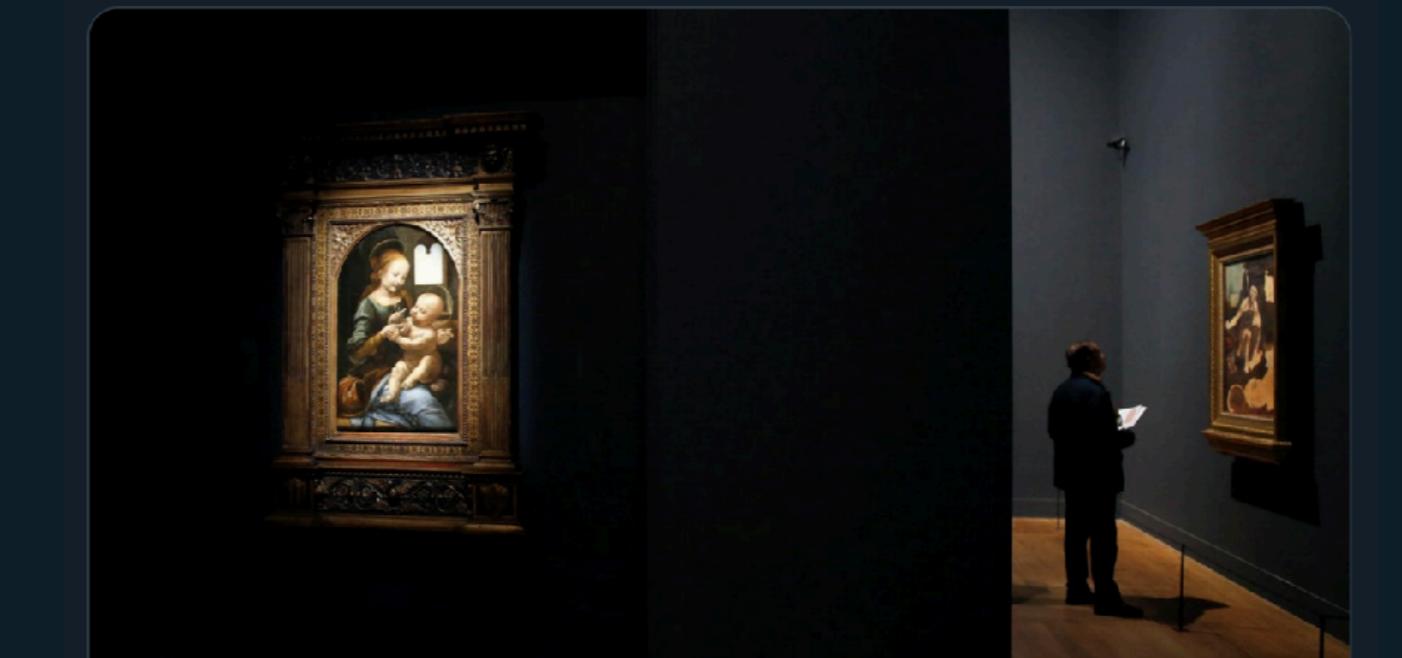


Note: The ratios presented are made to illustrate the concept of the base rate fallacy when the vaccination rate is high



NYT Health  
@NYTHealth

Want to live longer? Try going to the opera. Researchers in Britain have found that people who reported going to a museum or concert even once a year lived longer than those who didn't.



#### Another Benefit to Going to Museums? You May Live Longer

Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.

[nytimes.com](https://nytimes.com)

9:19 AM · Dec 22, 2019 · SocialFlow

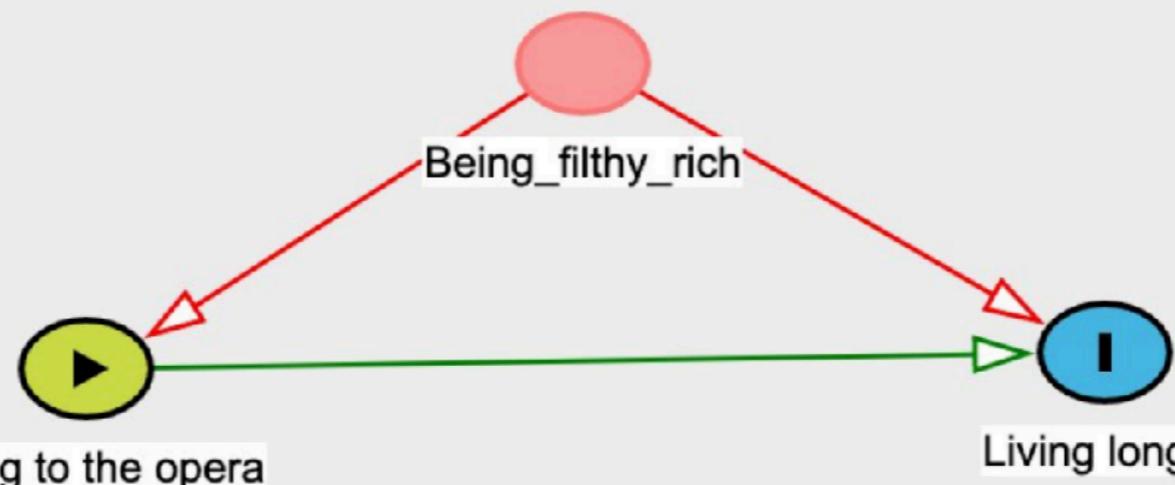
336 Retweets 1.3K Likes





Andrew Heiss  
@andrewheiss

ooh ooh i can draw the dag for this one!



NYT Health @NYTHealth · Dec 22, 2019

Want to live longer? Try going to the opera. Researchers in Britain have found that people who reported going to a museum or concert even once a year lived longer than those who didn't. [nyti.ms/2Q9AmZV](http://nyti.ms/2Q9AmZV)

2:47 PM · Dec 22, 2019 · Twitter Web App

View Tweet activity

837 Retweets 3.9K Likes

# **How will we learn?**

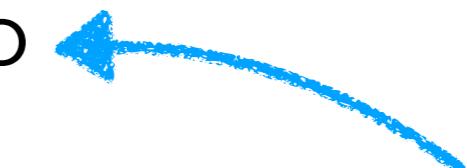
# How will we learn?

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

# How will we learn?

- **Lectures**

- Monday, Wednesday, Friday **10:30-11:50am**
- get familiar with R and RStudio
- visualization
- data manipulation
- simulation
- learn statistical methods



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

# How will we learn?

Ari Beller



Beth Rispoli



Satchel Grant



Shawn Schwartz



- **Sections**

- section times are **Monday 3:30-4:20pm** and **Friday 12:30-1:20pm**
- **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**)

# How will we learn?

- Office hours

**Tobi Gerstenberg**



Role	Instructor
Pronouns	he/him
Email (@stanford.edu)	gerstenberg
Office hours	Monday 1-2pm

<http://tinyurl.com/psych252office24>

# 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 (2/9) and is due on **Thursday (2/15) at 8pm**
    - no class the day before the deadline so you have more time
  - 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 2/22
- 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 3/22**
    - ~ 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%
  - Bonus:
    - Ed discussion: 2%

# Tools we will use in class

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

# 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

**PSYCH 252: STATISTICAL METHODS**

Home Schedule 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.

**2024 Team**

Tobi Gerstenberg	Ari Beller	Beth Rispoli	Satchel Grant	Shawn Schwartz
Role	Instructor	Teaching assistant	Teaching assistant	Teaching assistant
Pronouns	he/him	they/them	she/her	he/him
Email (@stanford.edu)	gerstenberg	abeller	brispoli	grantsrb
Office hours	Monday 1-2pm			

<https://psych252.github.io/>

# 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/psych252datacamp24>**

**Visualization 1**

**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 \(#3\)](#)

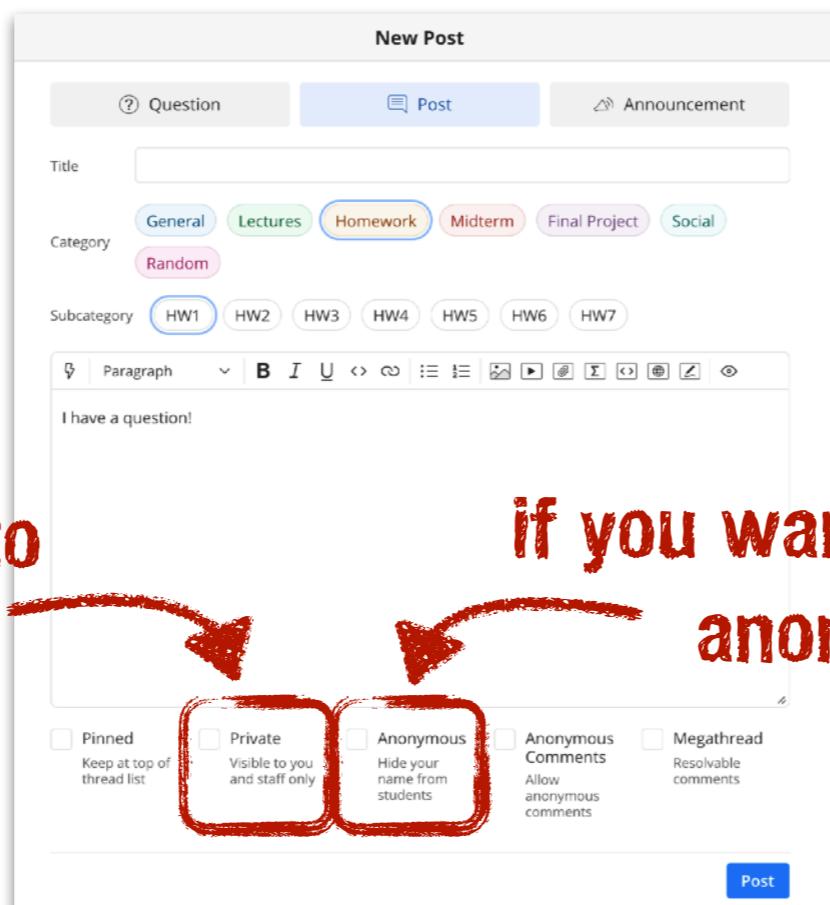


The DataCamp website homepage features a dark blue header with the DataCamp logo and a search bar. Below the header, a large banner reads "THE SMARTEST WAY TO Learn Data Science Online". A text block explains that people and businesses need to learn data science to succeed. A yellow "Start Learning For Free" button is prominently displayed. At the bottom, icons represent various programming languages and tools: Python, R, SQL, Spark, Git, Shell, and Spreadsheets.

# Tools we will use in class

- **Ed Discussion**

- post your own questions and answer those of your colleagues



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

**if you want to be  
anonymous**

# Tools we will use in class

- **Slack**

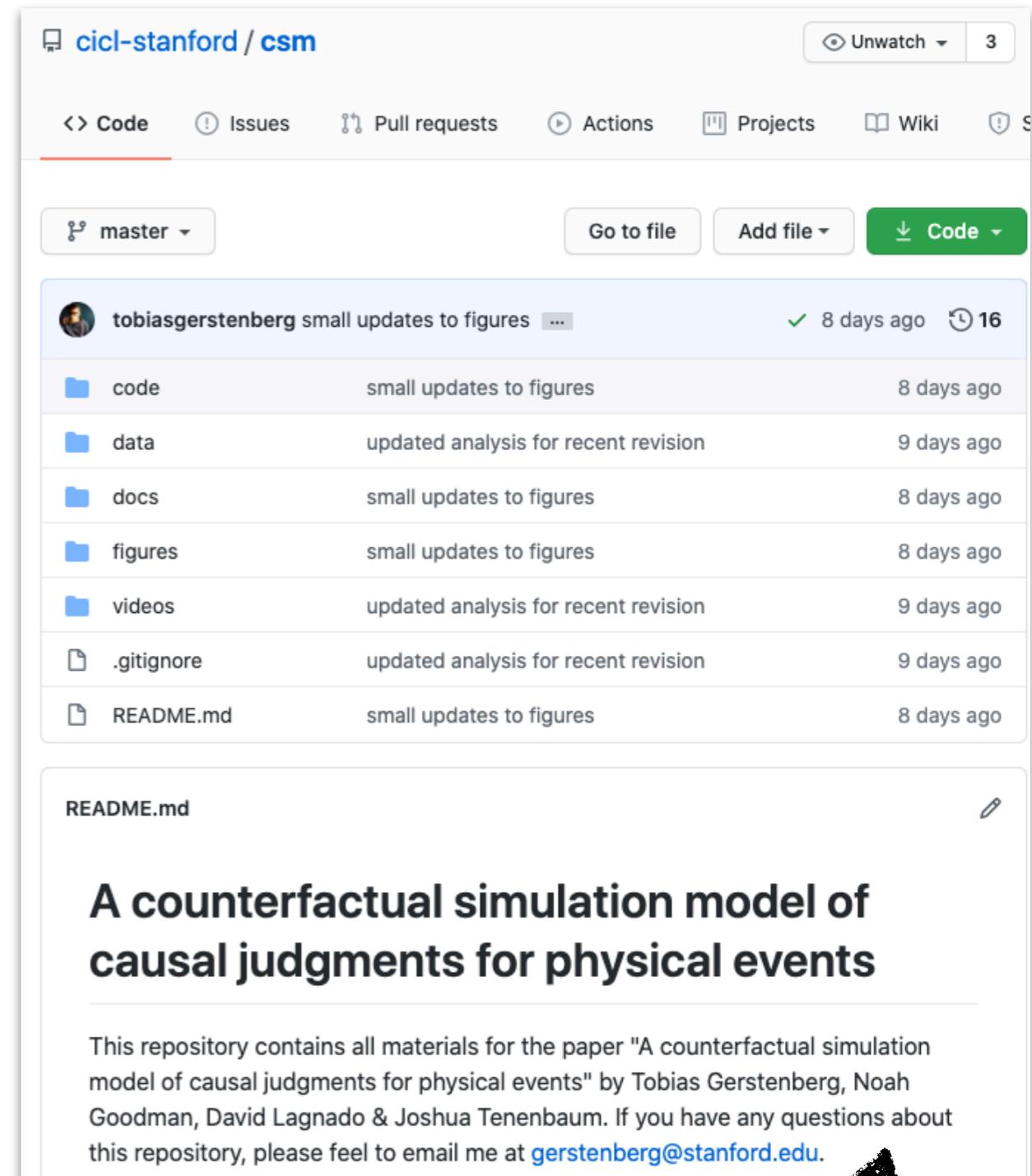
- just a space for you to build community
- the teaching team won't monitor what's happening on slack
- if you haven't received an invite yet, stick around at the end of class and/or send us an email



# Tools we will use in class

- **Github**

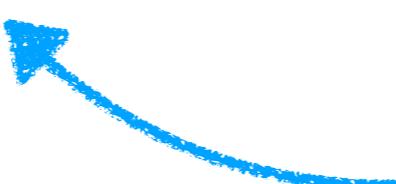
- we will use github for the final project
- it's an essential collaboration tool and critical for contributing to open science



**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**

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>

# Tools we will use in class

- **chatGPT**
  - try it out to see whether you find it helpful
  - we'll make sure that the homework and midterm are GPT-proof

## 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.



with help from

## Course website

I don't remember the schedule.

## Section & Office hours

I'd like to discuss something in person.

## Anonymous feedback form

Something bothers me ...

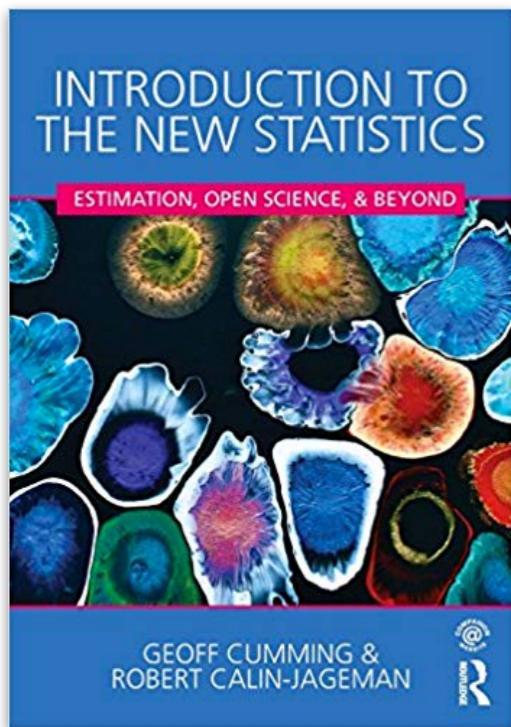
# **Some general thoughts**

# Vision for this class

In “A Vision for Stanford”, it says 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**

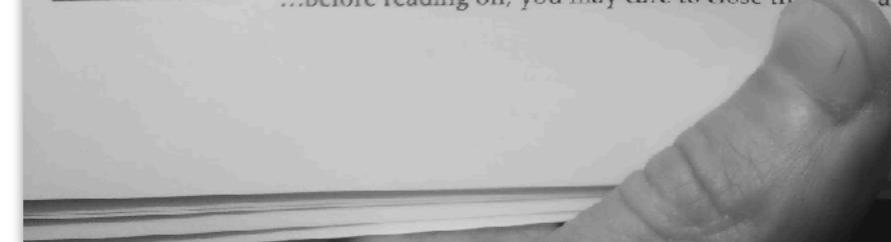
A further key idea is the distinction between a *fixed mindset* and a *growth mindset*. Carol Dweck and colleagues have demonstrated that helping students adopt a growth mindset can be a highly effective way to help them learn better and achieve more. Here's how Dweck describes the two mindsets:

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 ...

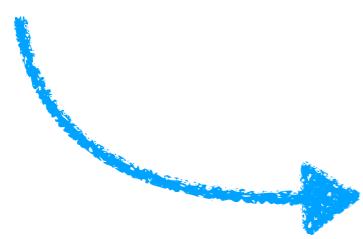
- it's not the only thing that matters for being a good scientist
- generating research ideas matters
- writing matters
- presenting matters
- ...
- (nobody is very good at all of these things)

# **Feedback**

# Teaching philosophy



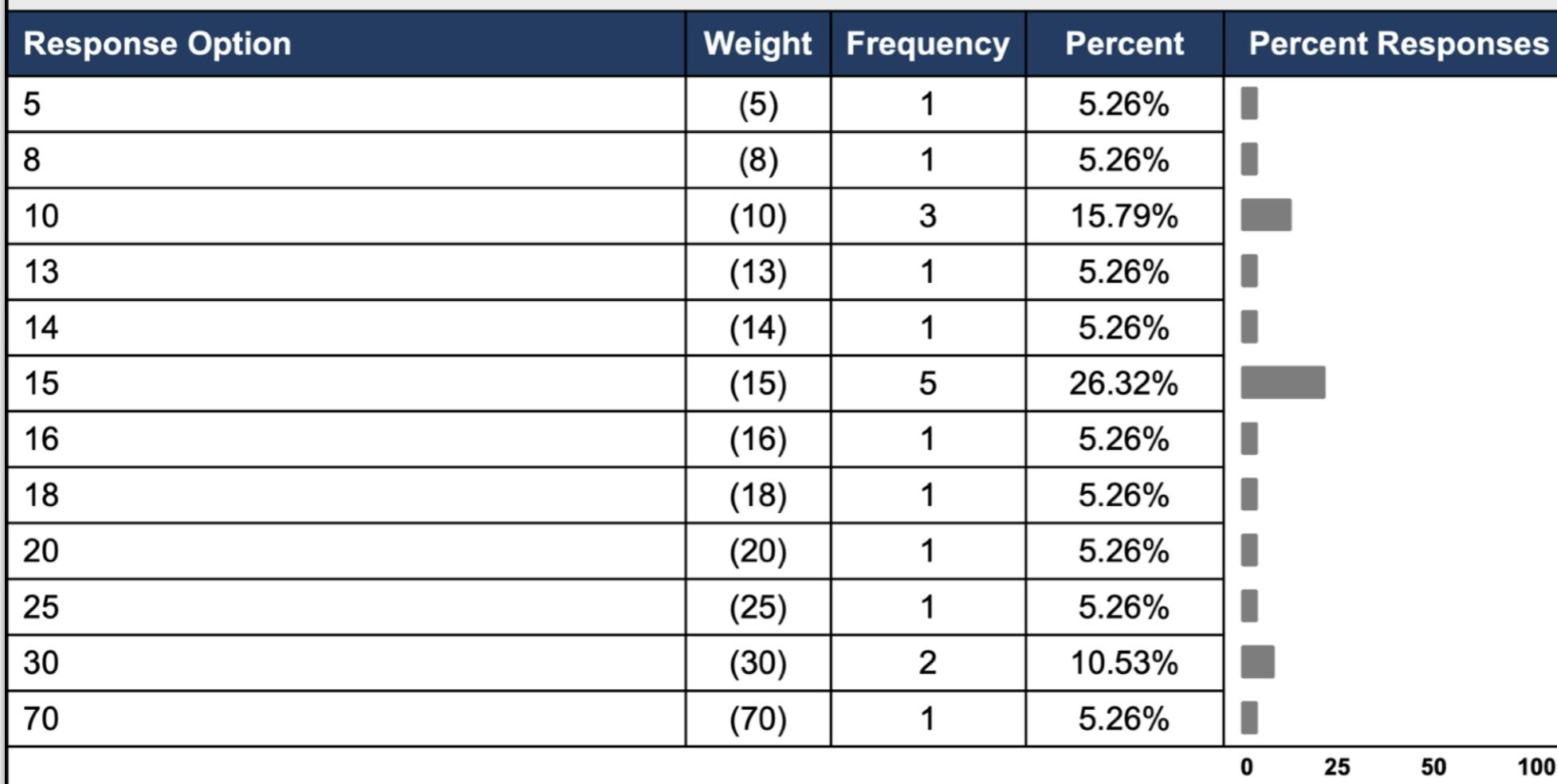
**"The one who does the work does the learning." (Doyle, 2008)**

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

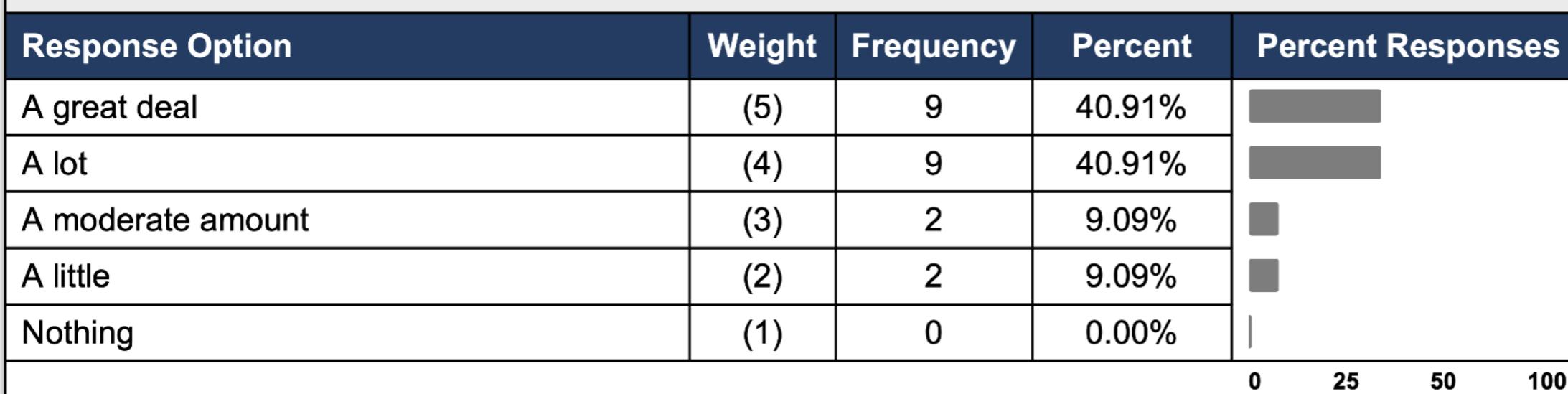
Doyle, T. (2008). Helping students learn in a learner-centered environment: A guide to facilitating learning in higher education. Stylus Publishing, LLC.

# 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

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

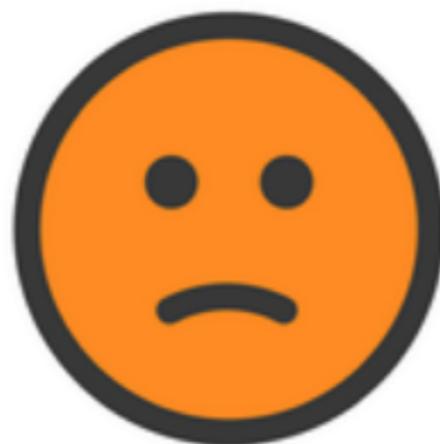
Help us help you!

- ask questions in class and stick around after class
- come to sections
- 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/psych252feedback24>

# 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?**

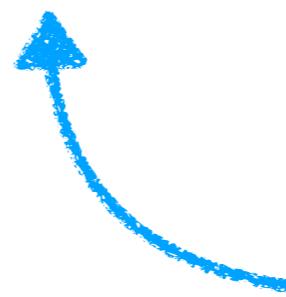
# Thank you to ...

Alexandra Chouldechova	Mika Braginsky
Allison Horst	Mike Frank
Andrew Heiss	Mine Çetinkaya-Rundel
Ben Baumer	Nick C. Huntington-Klein
Benoit Monin	Nilam Ram
Bodo Winter	Patrick Mair
David Lagnado	Paul-Christian Bürkner
Ewart Thomas	Peter Cushner Mohanty
Henrik Singmann	Richard McElreath
Julian Jara-Ettinger	Russ Poldrack
Justin Gardner	Stephen Dewitt
Kevin Smith	Solomon Kurz
Lisa DeBruine	Tom Hardwicke
Maarten Speekenbrink	Tristan Mahr
Matthew Kay	
Matthew Salganik	

# Thanks!

see you on **Wednesday**

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



post on Ed discussion if  
you experience any  
problems