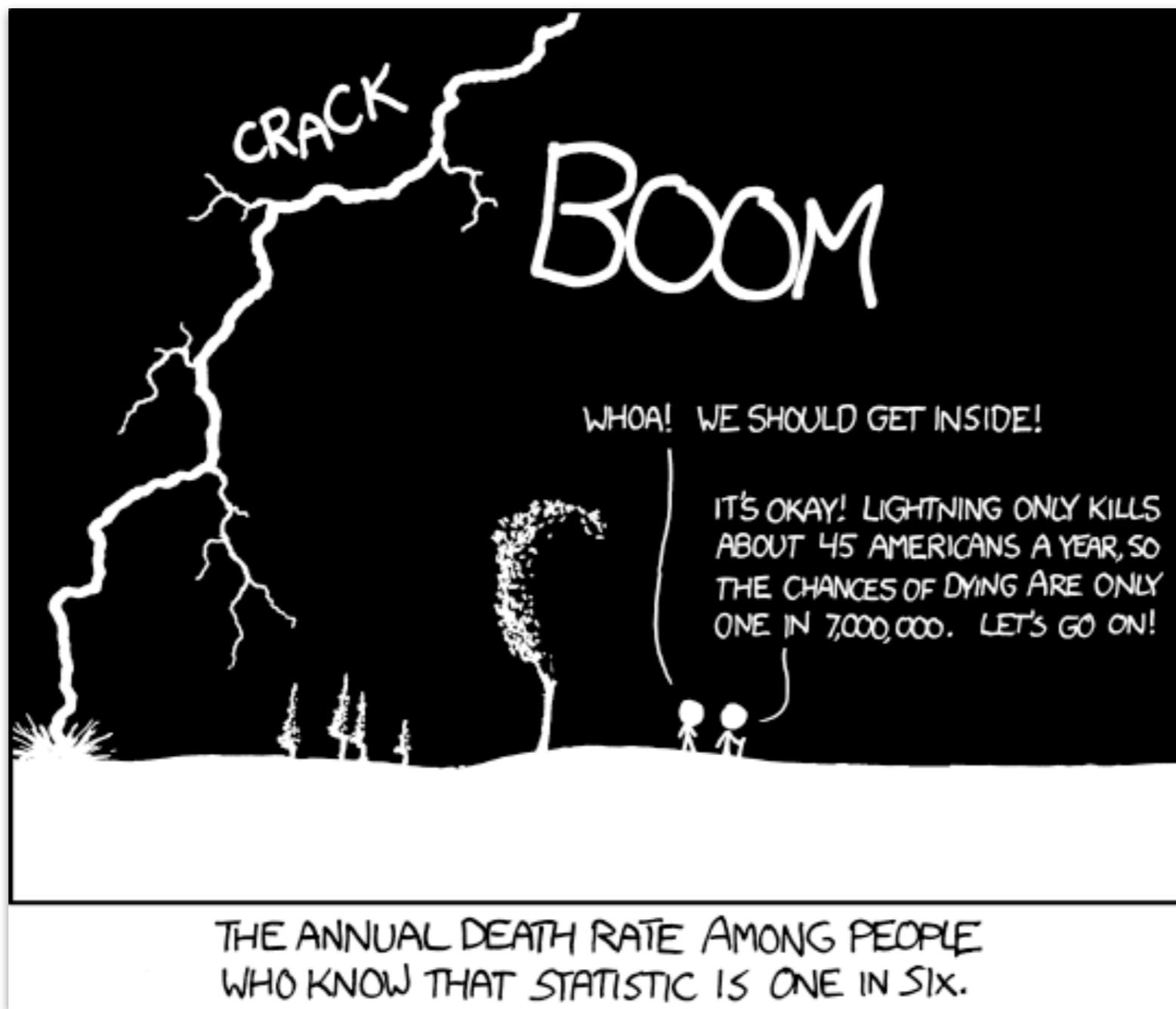


Introduction



xkcd

01/07/2019

Outline

- Introduction
- **What** will we learn?
- **How** will we learn?
- Some general thoughts
- Feedback

Introduction

Introduction

1. What did you do during the winter break?
2. What are you hoping to get out of this course?



Introduction

The screenshot shows a Google Forms survey titled "Psych 252: Introductory survey". The interface includes a green header bar with the title, a toolbar with various icons, and a sidebar on the right with additional tools. The main content area contains the survey introduction, an email address field, a question about graduate school year, and a question about department.

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+

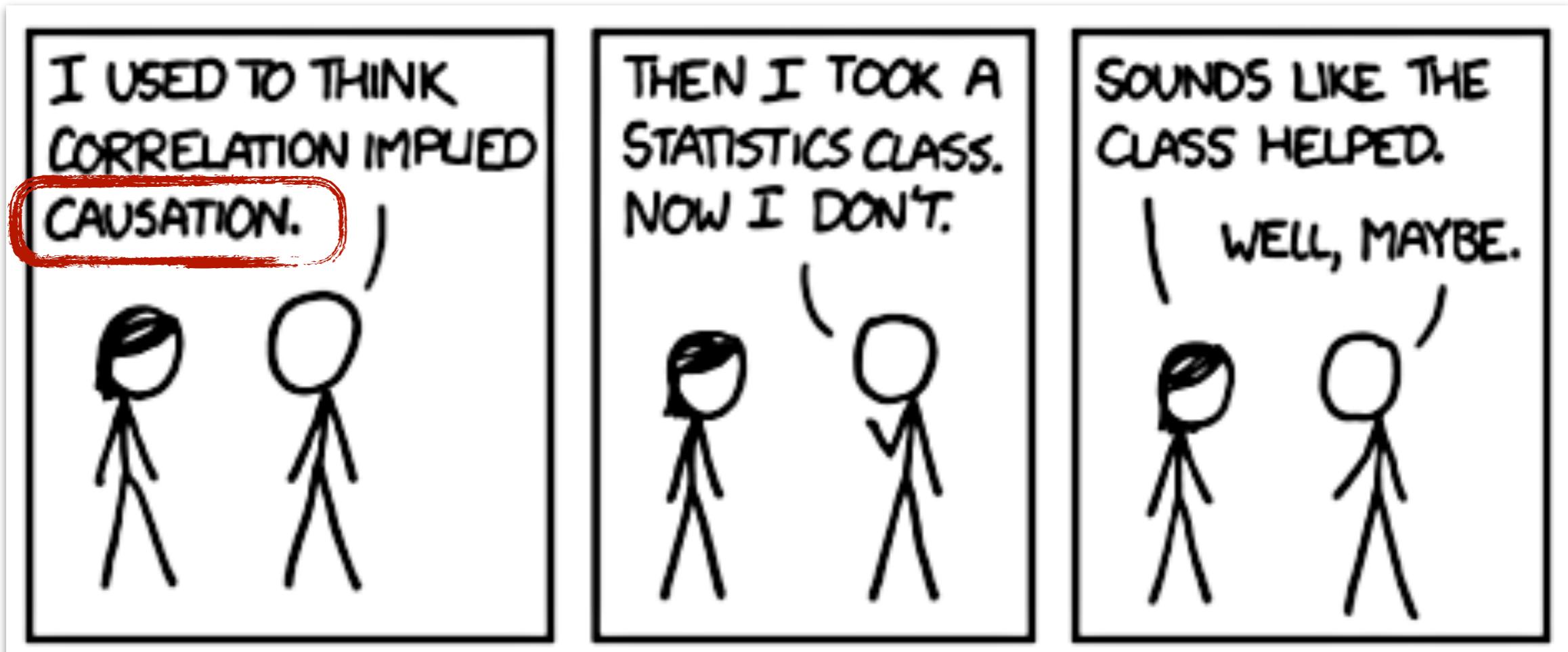
What department are you in?*

Psychology

Education

(announcement soon via Canvas)

Tobias Gerstenberg (he/him/his)



xkcd

Tobias Gerstenberg (he/him/his)

Tobi Gerstenberg

- I am leading the



(<http://cicl.stanford.edu/>).

- The lab studies the role of causality in our understanding of the world, and of each other.

- How does the mind learn to represent the causal structure of the world?
- What is the relationship between causal thinking and counterfactual simulation?
- How do we hold others responsible for the outcomes of their actions?

Role:	Instructor
Email:	gerstenberg@stanford.edu
Office:	302
Office hours:	Monday 2-3pm

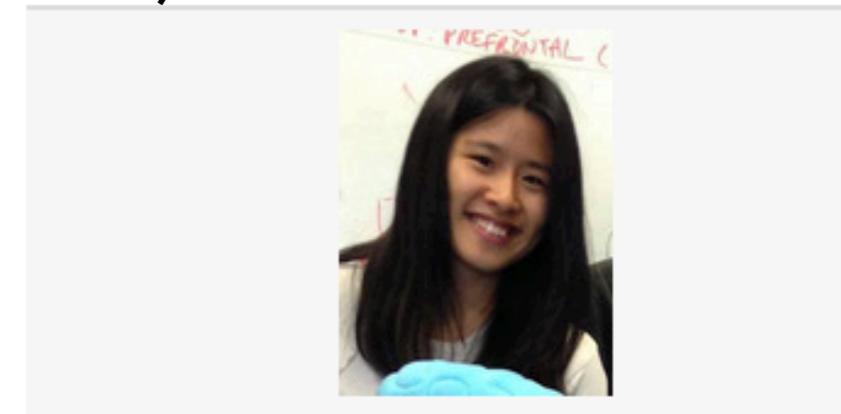
We formalize people's mental models as **computational models**.

We use a combination of large-scale **online experiments**, **interactive experiments** in the lab, and **eye-tracking experiments**.

Shao-Fang (Pam) Wang (she/her/hers)

[Shao-Fang \(Pam\) Wang](#)

- 4th year PhD student working with Dr. Anthony Wagner.
- I am interested in how interactions between the cortex and hippocampus contribute to cortical plasticity during learning. I use lab- and web-based behavioral tests, computational modeling, EEG, and fMRI.
- B.S. in Biology from National Taiwan University.

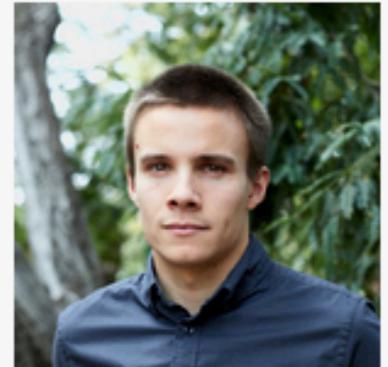


Role:	Teaching assistant
Email:	shaofang@stanford.edu
Office:	409
Office hours:	Wednesday 1-2pm
Section:	Wednesday 2:30-3:20pm in 160-326

Andrew Lampinen (he/him/his)

[Andrew Lampinen](#)

- I am interested in knowledge transfer, meta-learning, abstraction, and flexibility.
- I explore these issues from perspectives ranging from educational psychology to theoretical machine learning.
- I rely on statistically-based cognitive models like deep neural networks, as well as statistical techniques for analyzing and comparing the behavior of computational models and human subjects.



Role:	Teaching assistant
Email:	lampinen@stanford.edu
Office:	316
Office hours:	Friday 12:30-1:30pm
Section:	Friday 1:30-2:20pm in 160-314

Mona Rosenke (she/her/hers)



will give a quick intro this Wednesday

What will we learn?

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

- Formulate hypotheses as statistical models
- Bayesian statistics

3. Report results

- Reproducible research

**all the time
(& Week 10)**

Use R!

Why ?

Visualization

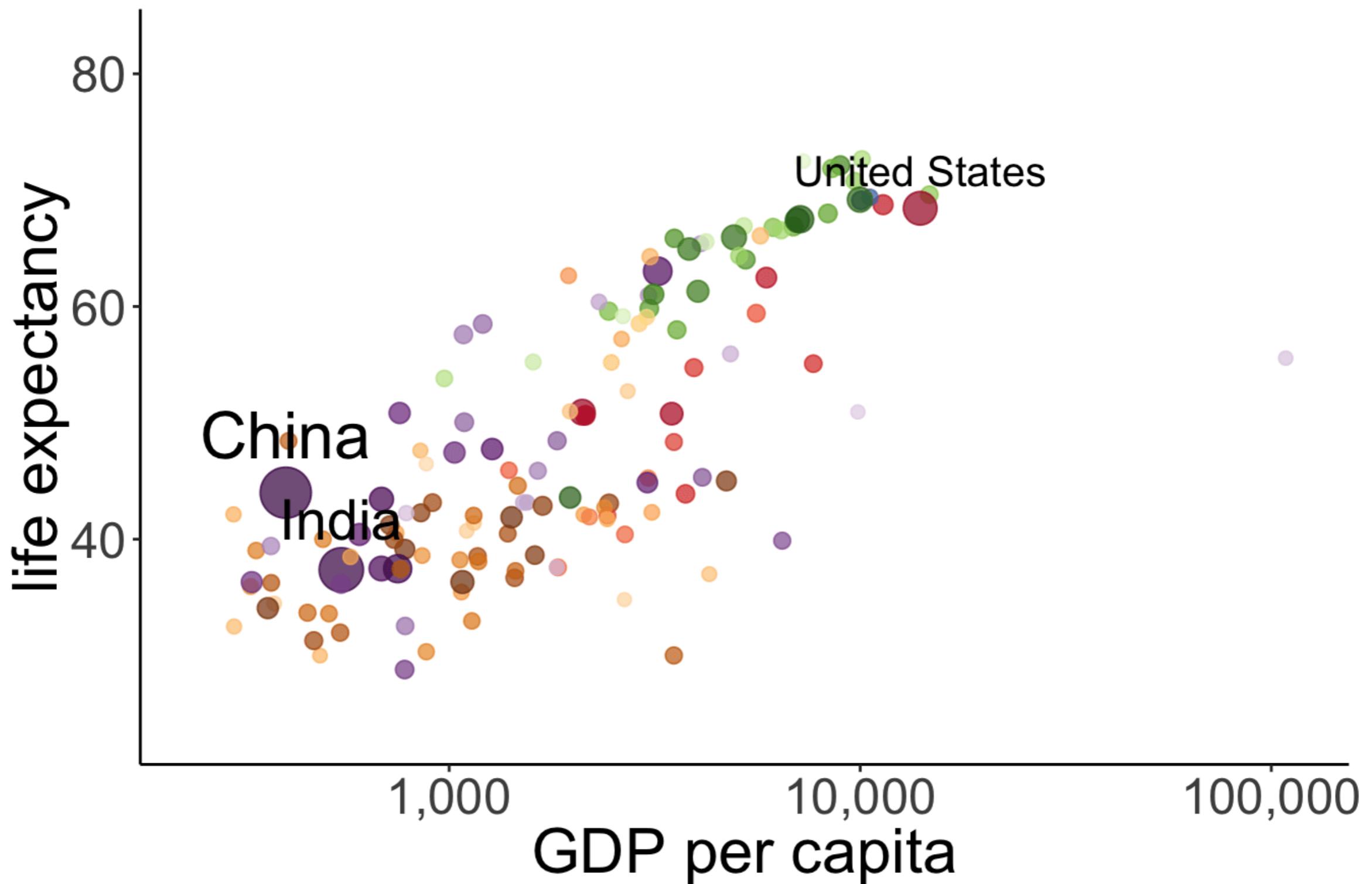
Data manipulation

Statistical modeling

Communicating results

Why ? Visualization

Year: 1952



Why R ? 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? Data manipulation

Data Transformation with dplyr :: CHEAT SHEET



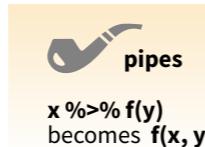
dplyr functions work with pipes and expect **tidy data**. In tidy data:



Each **variable** is in its own **column**

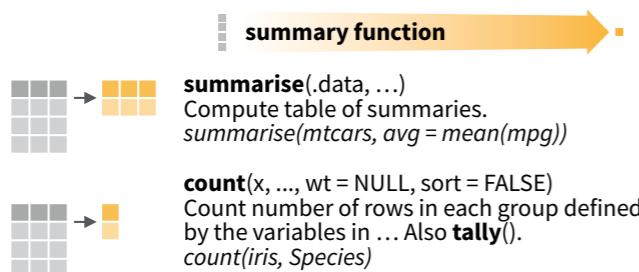


Each **observation**, or **case**, is in its own **row**



Summarise Cases

These apply **summary functions** to columns to create a new table of summary statistics. Summary functions take vectors as input and return one value (see back).

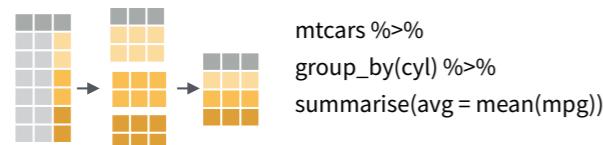


VARIATIONS

- summarise_all()** - Apply funs to every column.
- summarise_at()** - Apply funs to specific columns.
- summarise_if()** - Apply funs to all cols of one type.

Group Cases

Use **group_by()** to create a "grouped" copy of a table. dplyr functions will manipulate each "group" separately and then combine the results.



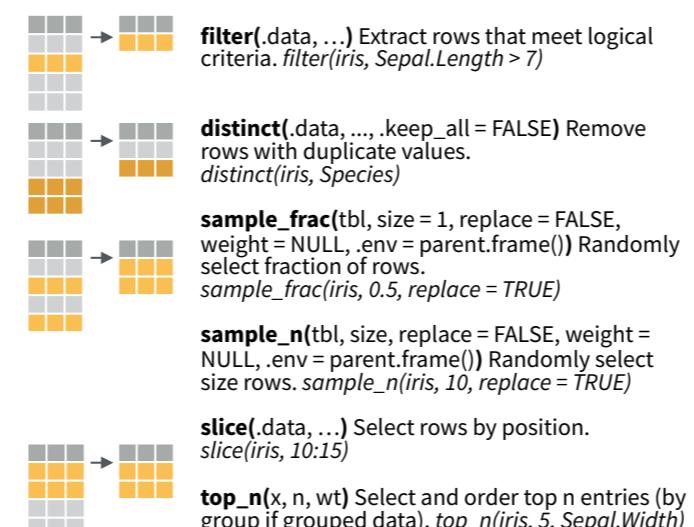
group_by(.data, ..., add = FALSE)
Returns copy of table grouped by ...
`g_iris <- group_by(iris, Species)`

ungroup(x, ...)
Returns ungrouped copy of table.
`ungroup(g_iris)`

Manipulate Cases

EXTRACT CASES

Row functions return a subset of rows as a new table.



Logical and boolean operators to use with filter()

<	<=	is.na()	%in%		xor()
>	>=	!is.na()	!	&	

See **?base:::logic** and **?Comparison** for help.

ARRANGE CASES

arrange(.data, ...) Order rows by values of a column or columns (low to high), use with **desc()** to order from high to low.
`arrange(mtcars, mpg)`
`arrange(mtcars, desc(mpg))`

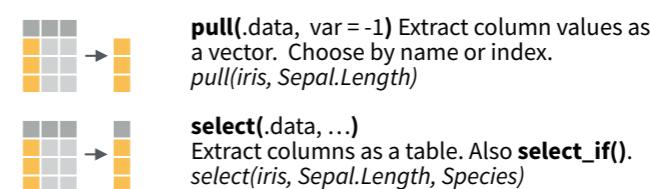
ADD CASES

add_row(.data, ..., .before = NULL, .after = NULL) Add one or more rows to a table.
`add_row(faithful, eruptions = 1, waiting = 1)`

Manipulate Variables

EXTRACT VARIABLES

Column functions return a set of columns as a new vector or table.

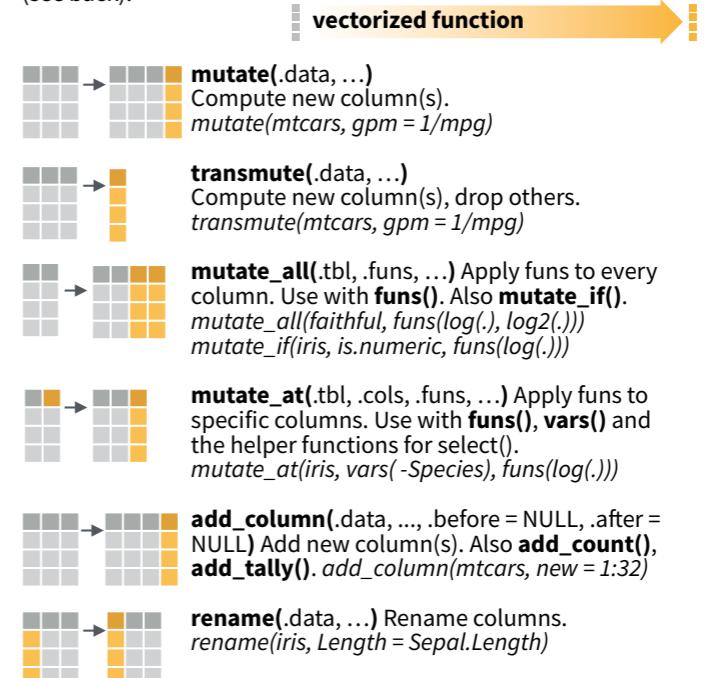


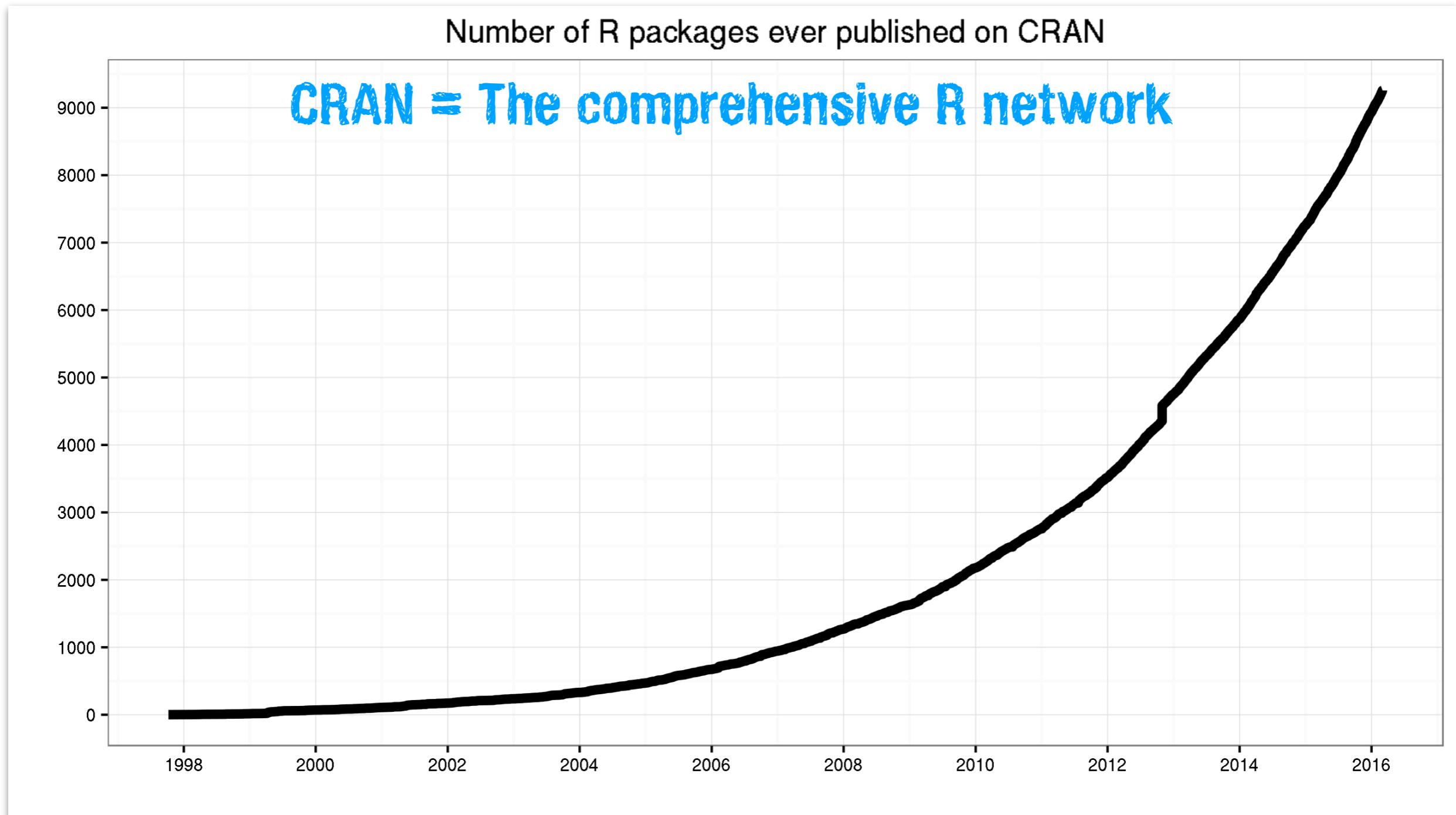
Use these helpers with **select ()**,
e.g. `select(iris, starts_with("Sepal"))`

contains(match)	num_range(prefix, range)	:, e.g. <code>mpg:cyl</code>
ends_with(match)	one_of(...)	-, e.g. <code>-Species</code>
matches(match)	starts_with(match)	

MAKE NEW VARIABLES

These apply **vectorized functions** to columns. Vectorized funs take vectors as input and return vectors of the same length as output (see back).





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

Why R ? 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)
```

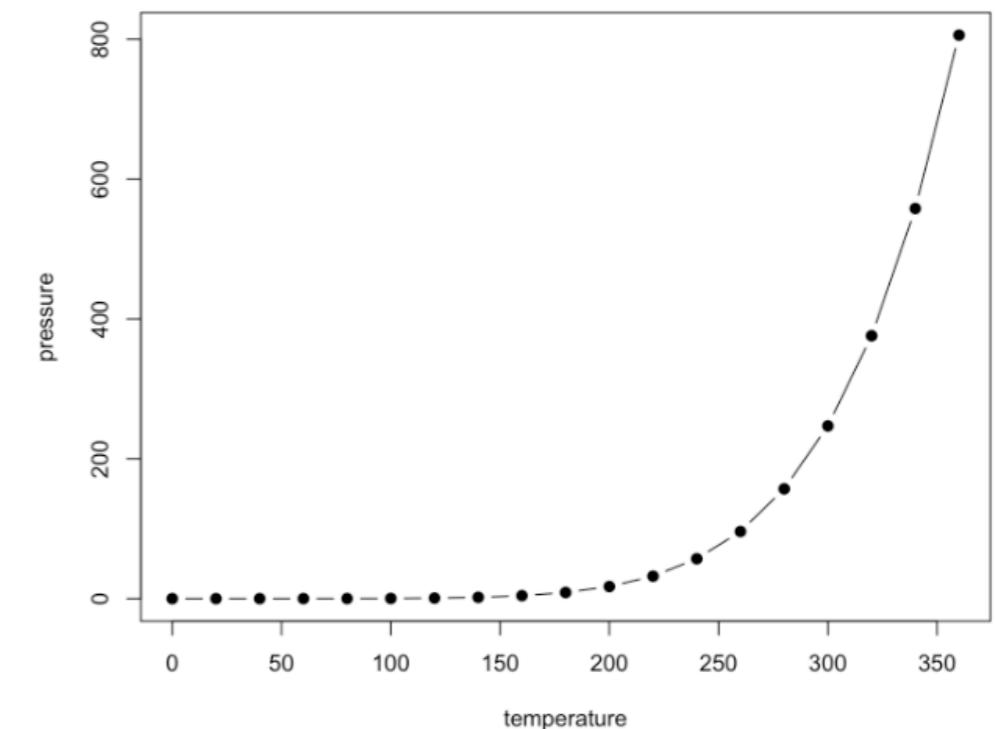


Figure 1.1: Here is a nice figure!

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 ? 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

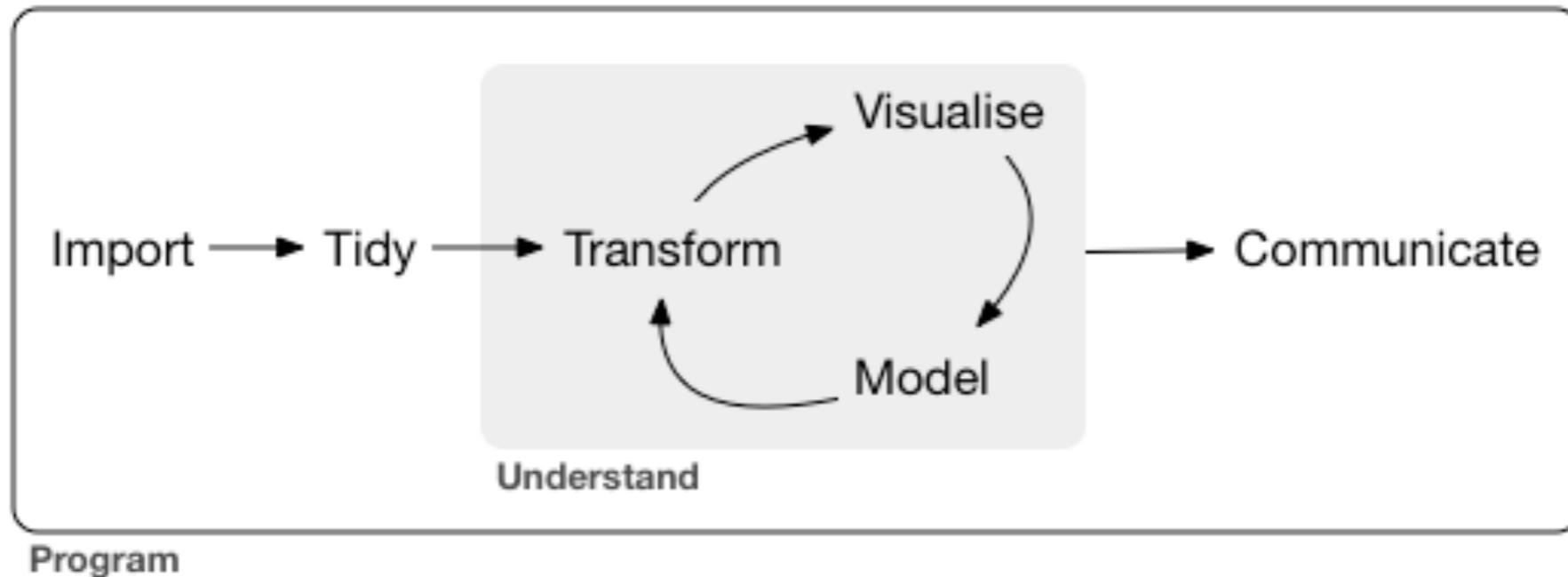
R for Data Science

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)!
- We use R to better understand statistics

Why R? Also ...

instead of

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

(google search "complicated statistical equation")

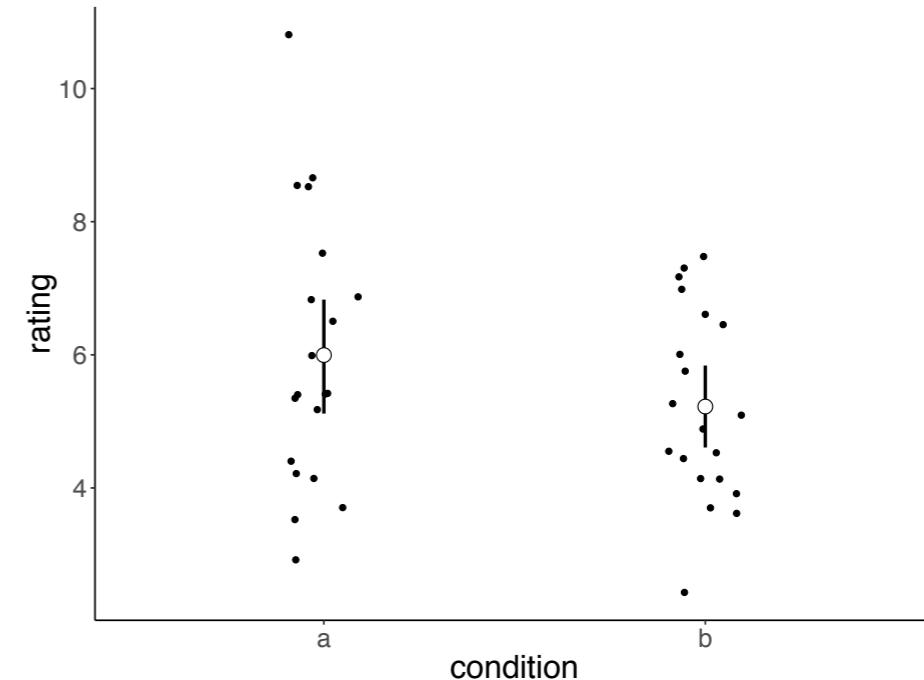
df	Level of Significance for Directional Test (t_{crit})					
	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.025$	$\alpha=0.01$	$\alpha=0.005$	$\alpha=0.0005$
Level of Significance for Non-directional Test (t_{crit})						
1	3.0780	6.3140	12.7100	31.8200	63.6600	636.6000
2	1.8860	2.9200	4.3030	6.9650	9.9250	31.6000
3	1.6380	2.3530	3.1820	4.5410	5.8410	12.9200
4	1.5330	2.1320	2.7760	3.7470	4.6040	8.6100
5	1.4760	2.0150	2.5710	3.3650	4.0320	6.8690
6	1.4400	1.9430	2.4470	3.1430	3.7070	5.9590
7	1.4150	1.8950	2.3650	2.9980	3.4990	5.4080
8	1.3970	1.8600	2.3060	2.8960	3.3550	5.0410
9	1.3830	1.8330	2.2620	2.8210	3.2500	4.7810
10	1.3720	1.8120	2.2280	2.7640	3.1690	4.5870
11	1.3630	1.7960	2.2010	2.7180	3.1060	4.4370
12	1.3560	1.7820	2.1790	2.6810	3.0550	4.3180
13	1.3500	1.7710	2.1600	2.6500	3.0120	4.2210
14	1.3450	1.7610	2.1450	2.6240	2.9770	4.1400
15	1.3410	1.7530	2.1310	2.6020	2.9470	4.0730

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  -

```



But what about python ™ ?

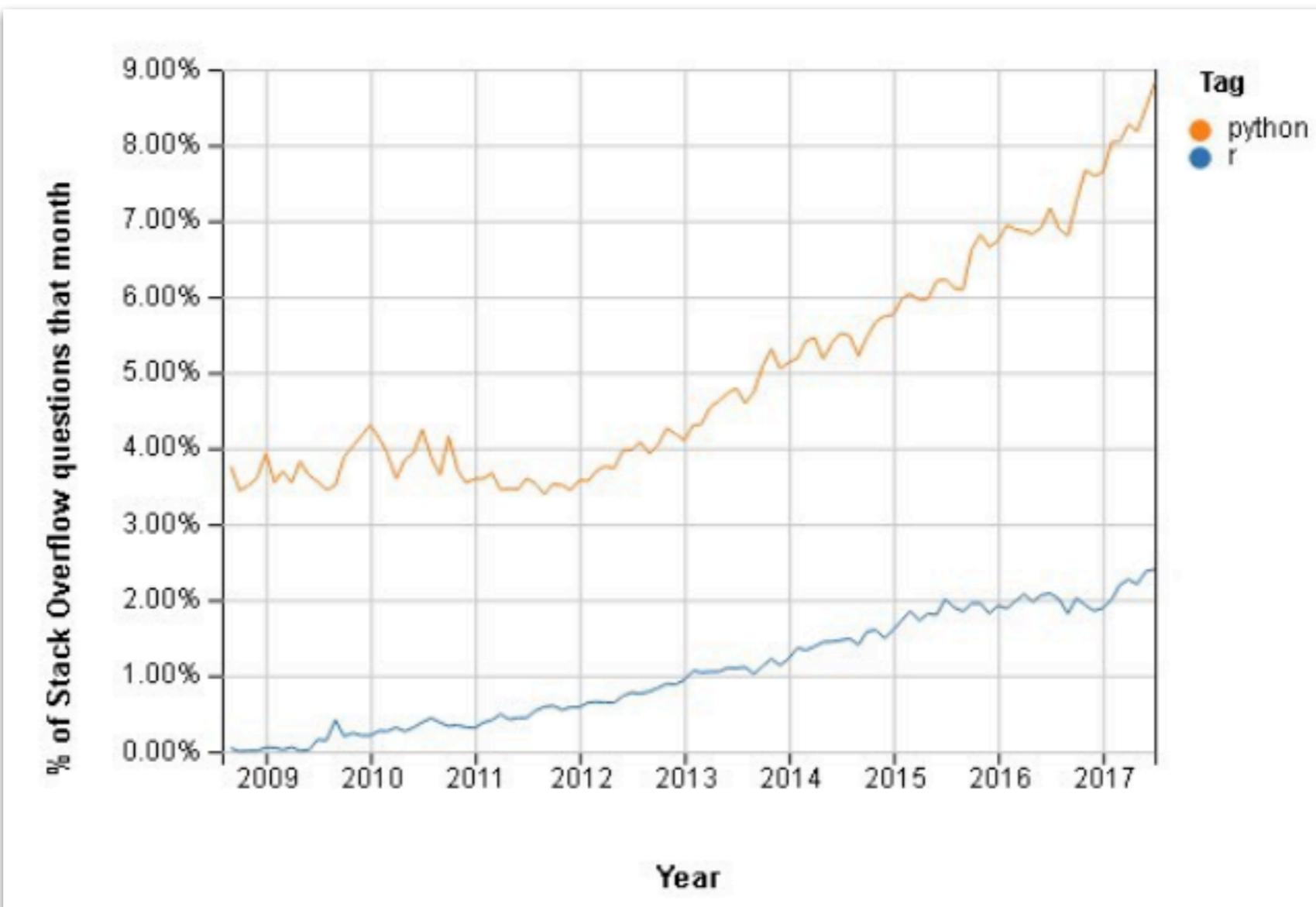
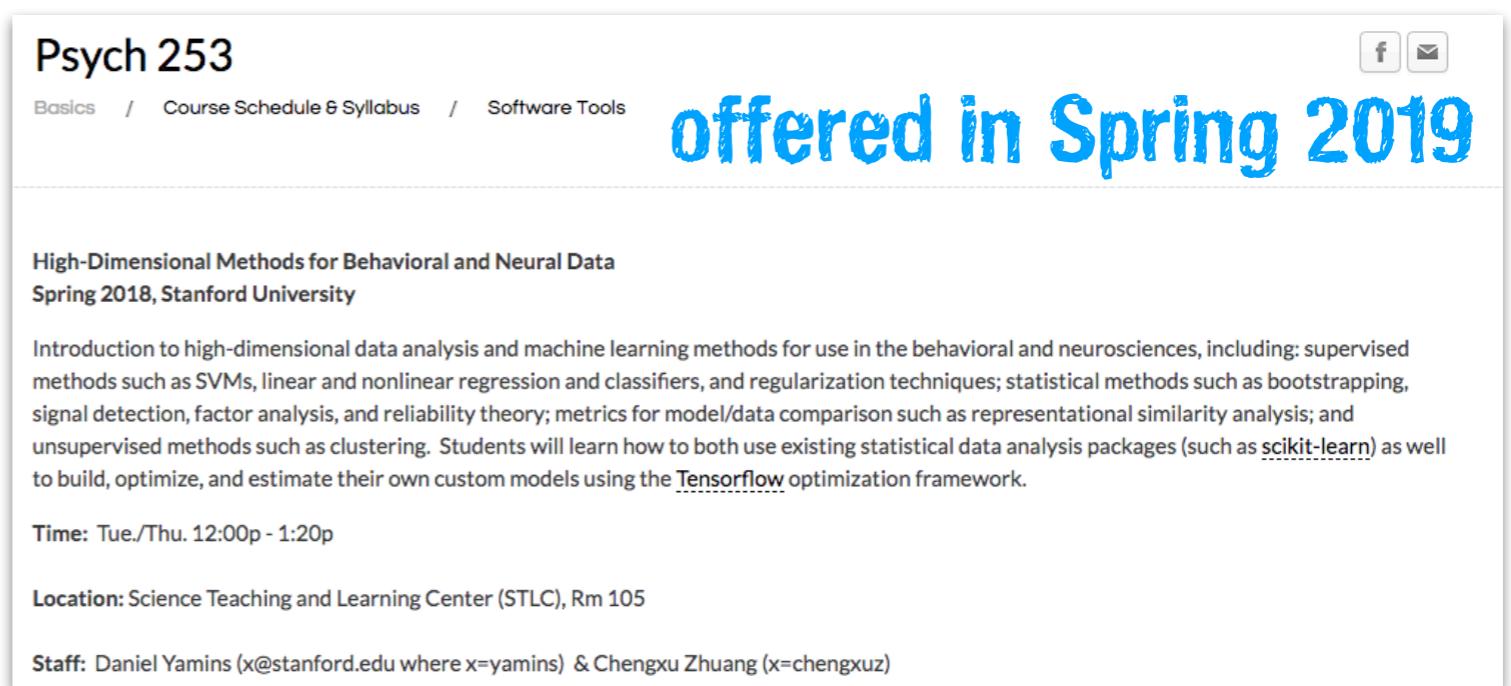


Image Source: <https://dzone.com/articles/r-or-python-data-scientists-delight>

But what about python™ ?

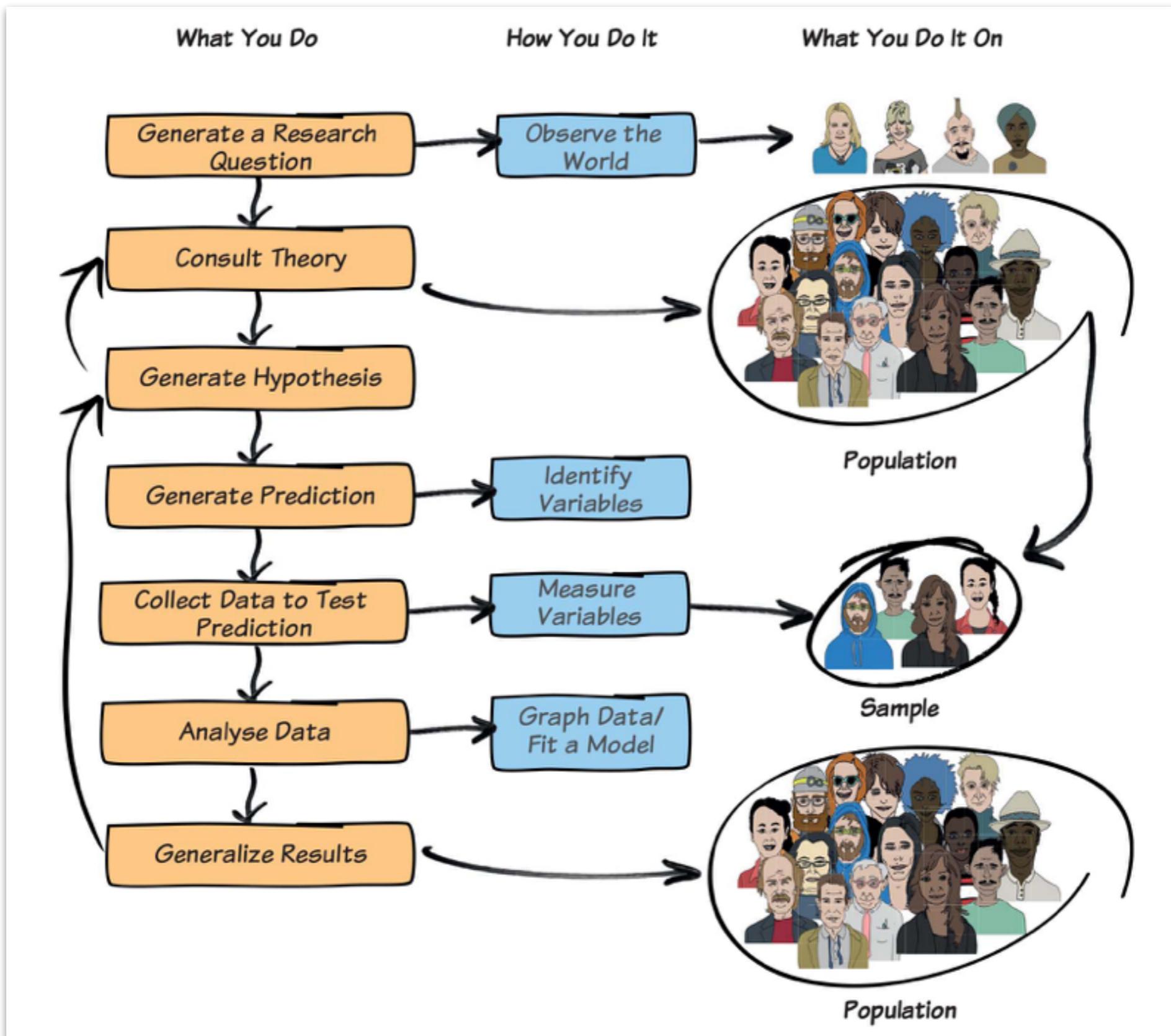
- You should learn python, too! :)
- Python is the more general purpose programming language.
- Python is great for machine learning (e.g. neural networks and deep learning).
- You can learn about all of these things in **Psych 253!**



The screenshot shows the homepage of the Psych 253 course website. At the top, it says "Psych 253" and has links for "Basics", "Course Schedule & Syllabus", and "Software Tools". On the right, there are social media icons for Facebook and Email. A large blue banner across the top reads "offered in Spring 2019". Below the banner, the course title is listed as "High-Dimensional Methods for Behavioral and Neural Data" and "Spring 2018, Stanford University". The course description is: "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." Below the description, it says "Time: Tue./Thu. 12:00p - 1:20p", "Location: Science Teaching and Learning Center (STLC), Rm 105", and "Staff: Daniel Yamins (x@stanford.edu where x=yamins) & Chengxu Zhuang (x=chengxuz)".

Build models

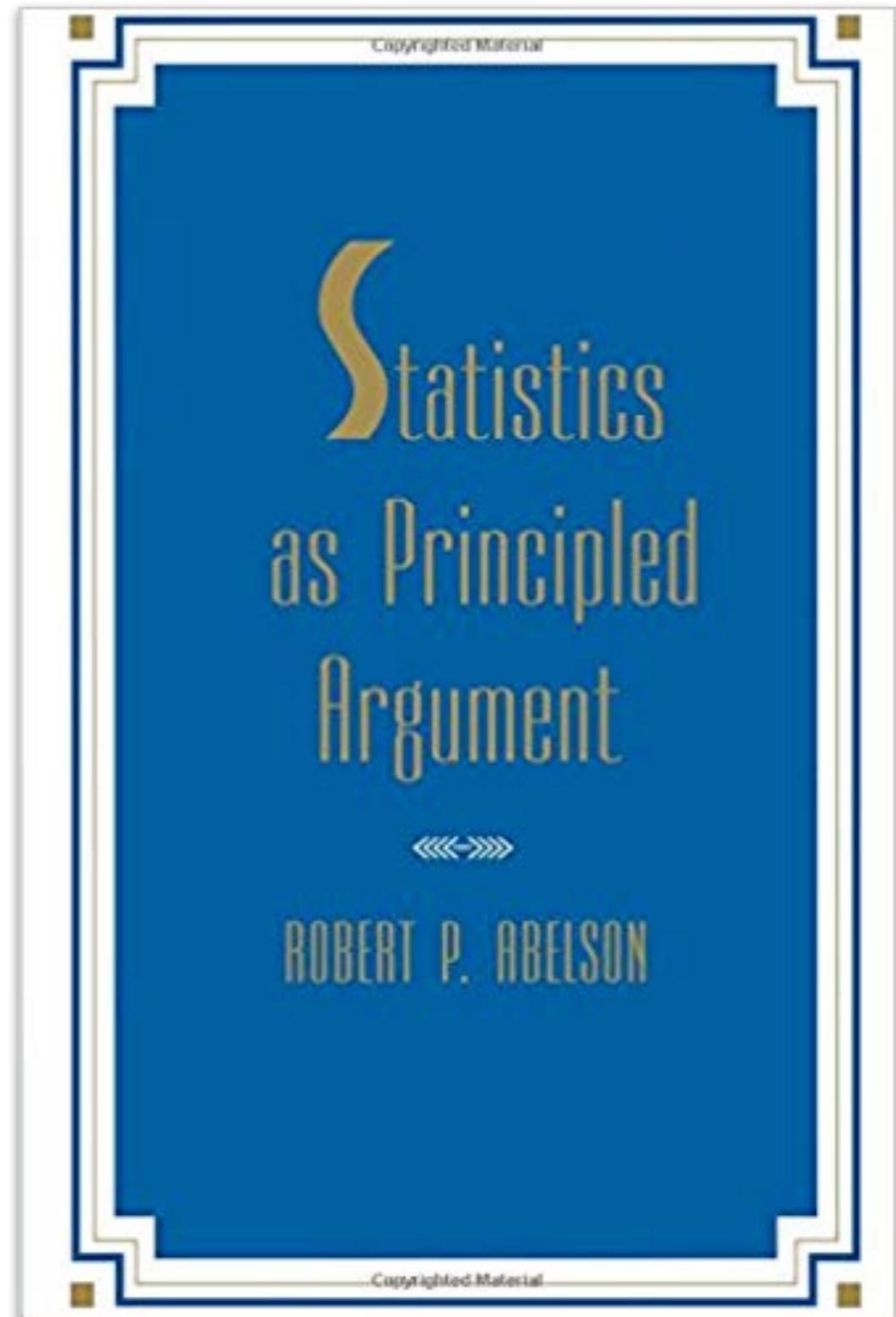
The research process



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, ...

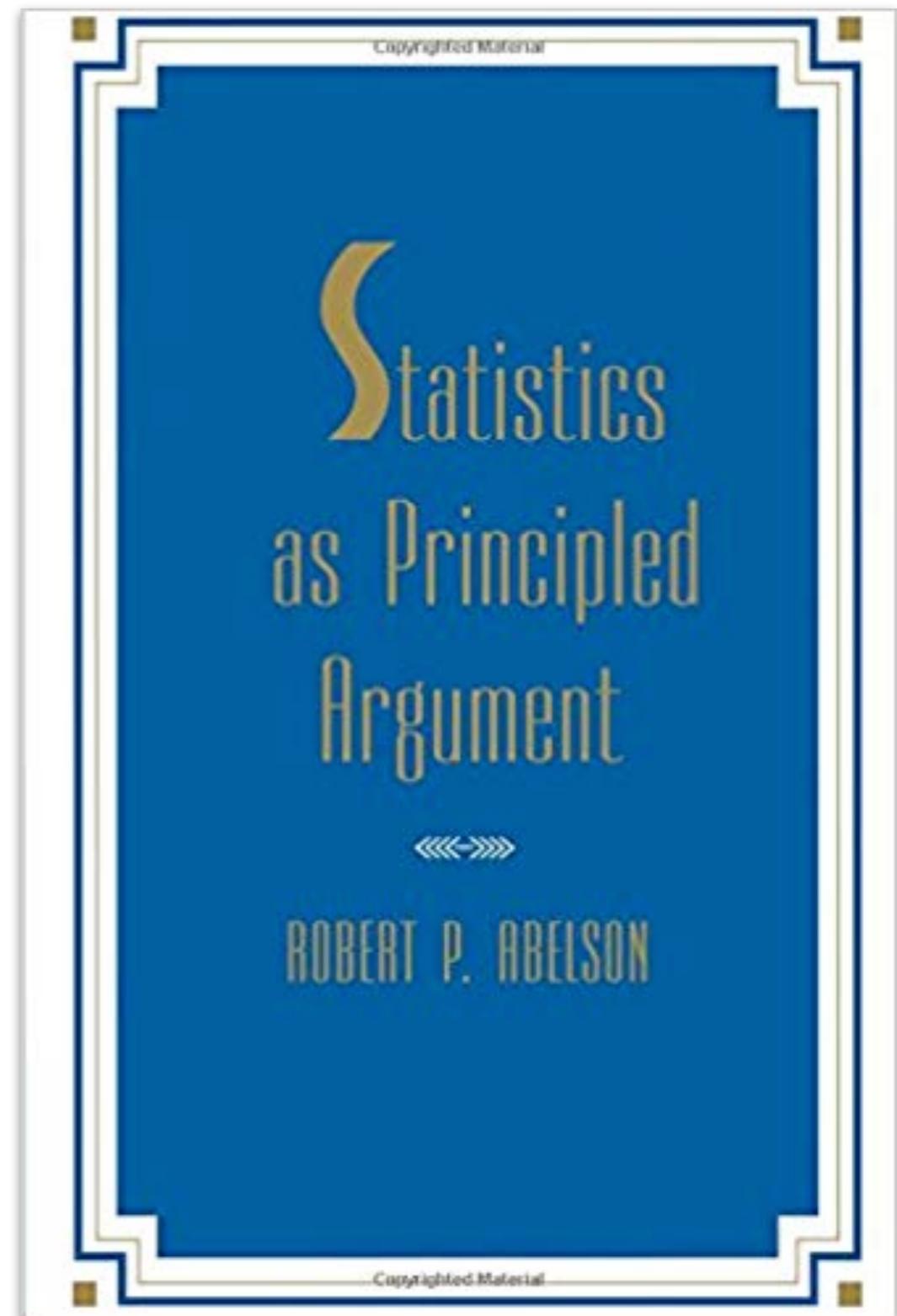


Abelson, R. P. (1995). *Statistics as principled argument*. New York: Psychology Press.

Statistics as argument

5 criteria make statistical arguments persuasive, or **MAGIC**:

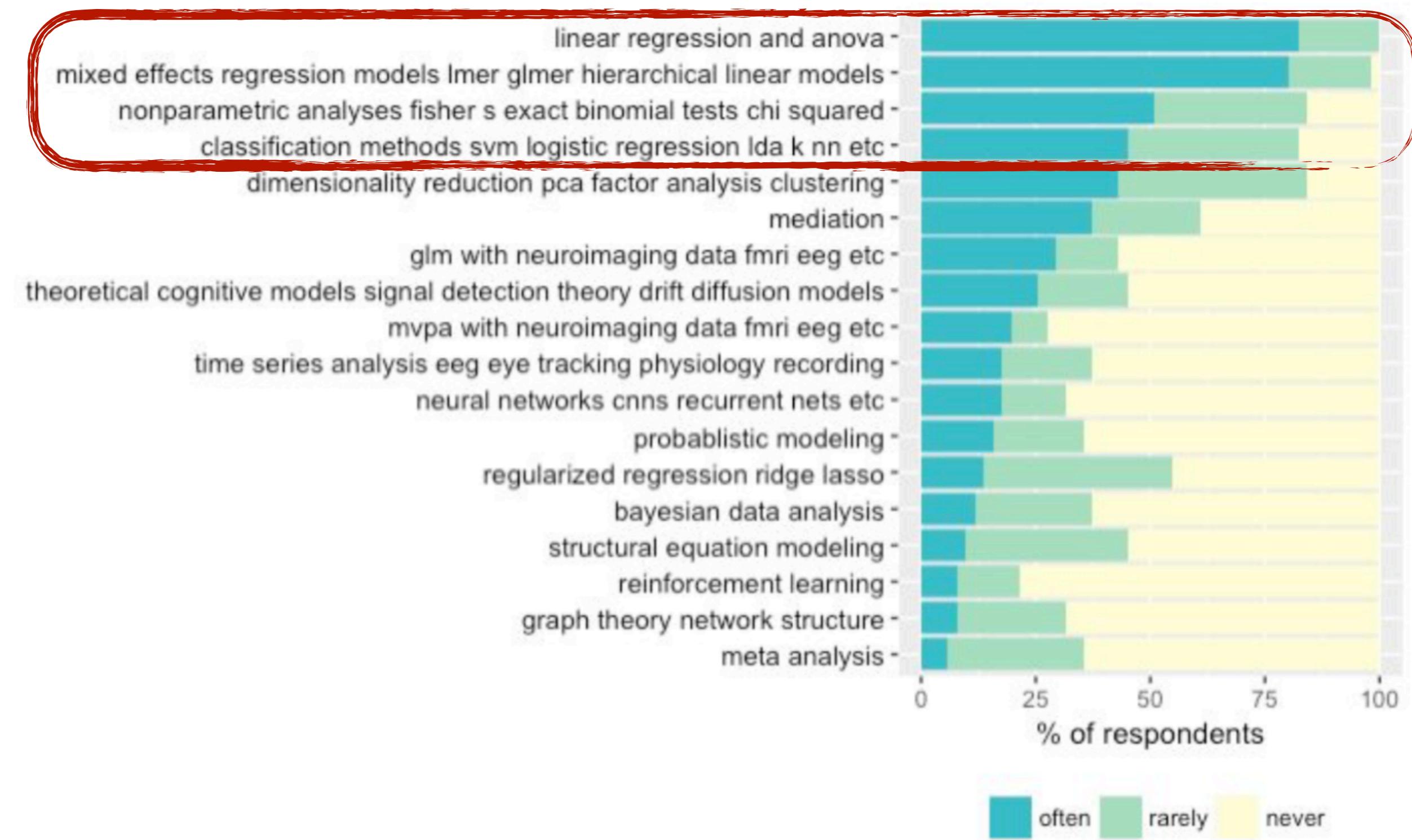
- **Magnitude**: Larger effects are more persuasive
- **Articulation**: Understandability of results
- **Generality**: How much does the effect depend on contextual factors
- **Interestingness**: Change what people believe about an important issue
- **Credibility**: Believability of a research claim, methodological soundness, and theoretical coherence



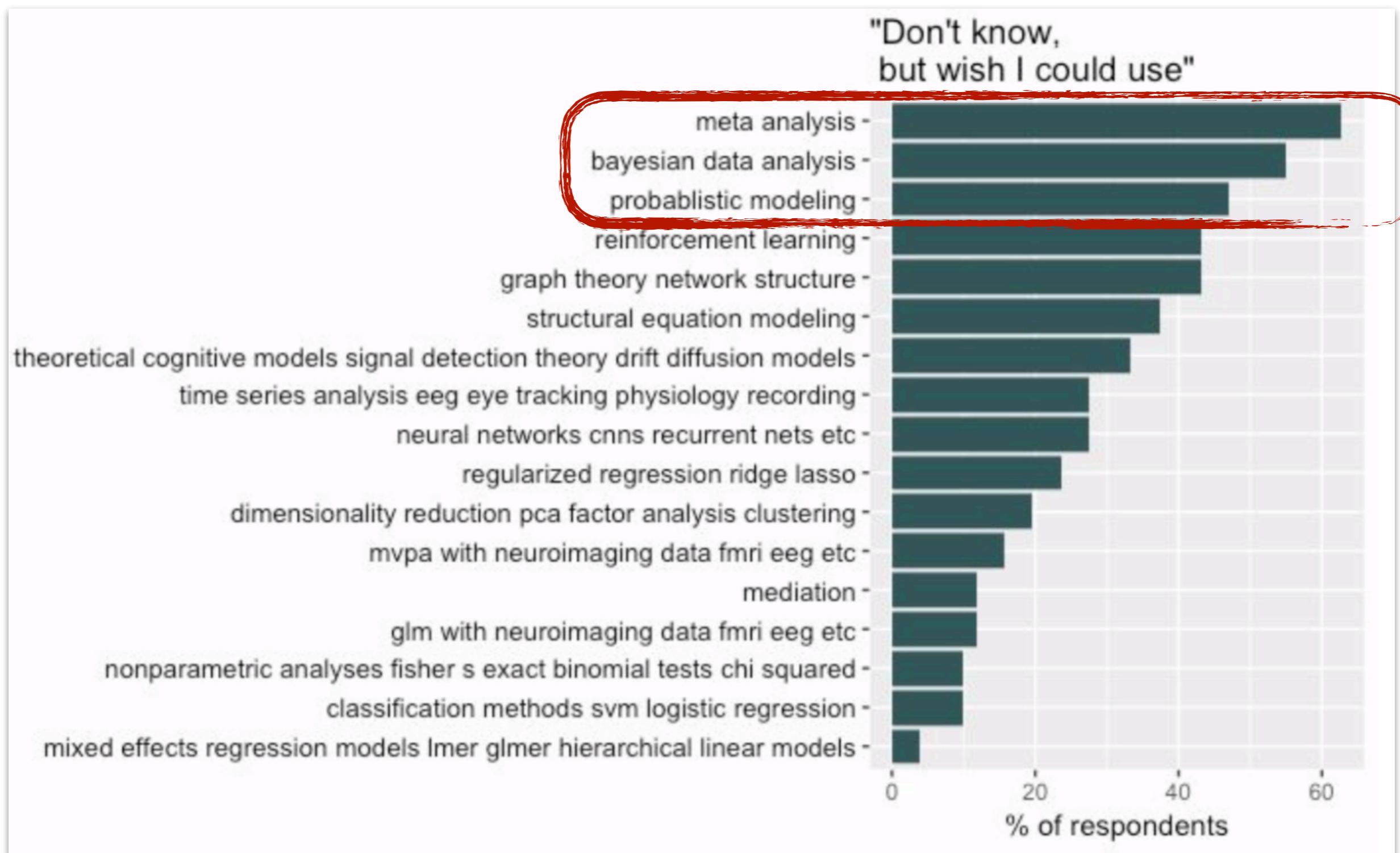
Abelson, R. P. (1995). *Statistics as principled argument*. New York: Psychology Press.

Psychology grad student survey

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



Psychology grad student survey



Report results

Replication crisis

Google replication crisis in psychology

All News Images Videos Shopping More Settings Tools SafeSearch ▾

precis reproducibility crisis scientific estimating science social science psych nature studies recommendations social psychology psychological science research randall monroe stapel

Economics (n=18) Psychology (n=100)
Camerer et al., 2016 Open Science Collaboration, 2015

Failed to replicate Meta analytic effect p<0.05 Replication p<0.05

Journal % Findings Replicated
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
Overall 36

Effect size reduction Independent replication p-value

REPLICATION CRISIS

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

THE CHRONICLE OF HIGHER EDUCATION How to Fix Psychology's Replication Crisis

Tradition of controlled experiments No Yes
No Epidemiology Clinical Sciences CRISIS!
Yes Strong basic theory Astrology Particle physics

nature How Reliable Psychology S Massive International Project Raises Questions about the Validity of Psychology Research

Science Only 36% of studies replicated!!

Replication Effect Size P-value Non-Significant Replication Power

psychological science ... science.sciencemag.org

Replication Crisis Overblown ... thecut.com

Health News Digest ... bfm.my

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

6 Principles of Open Science

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

Lower integrity Higher integrity
Weak science Strongest science
Exact replications Conceptual replications

mean items solved (adjusted by SAT)

RACE PRIME NO RACE PRIME

■ BLACK SUBJECTS □ WHITE SUBJECTS

PSYCHOLOGY'S REPRODUCIBILITY PROBLEM

LARGE SCALE REPPLICATION 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.

A B

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	8%	—	13	Yes	Nonrandom
Camerer et al., 2016	Economics	61%	0.66	18	No	Nonrandom

Project organization

- **Use R Projects**

- makes sure that the working directory is set correctly
- allows you to work on multiple projects at the same time

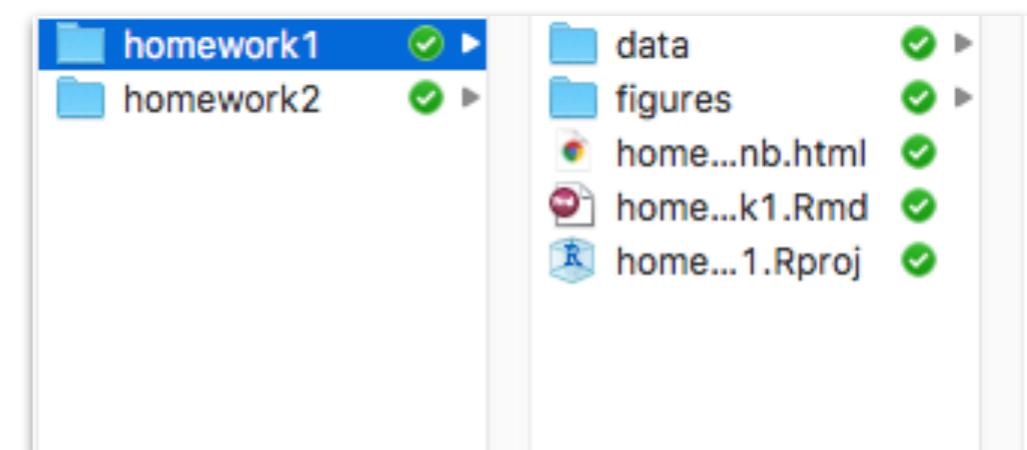
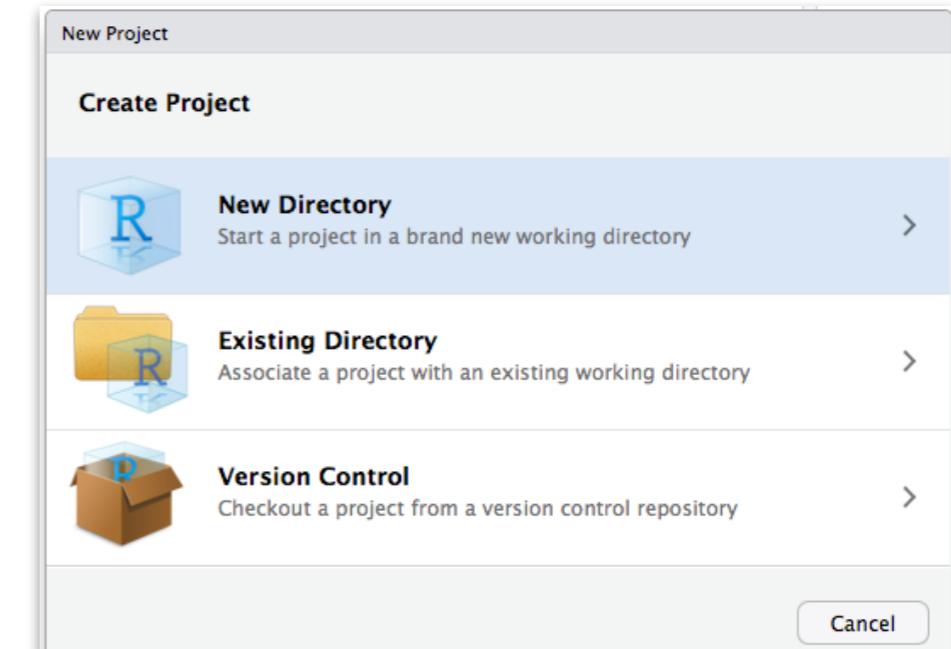
- **Organize** the folder structure
- **Use relative paths**



```
1 read_csv("data/dataset.csv")
```



```
1 read_csv("~/tobi/Documents/work/homework/data/  
dataset.csv")
```



R Markdown

R Markdown :: CHEAT SHEET

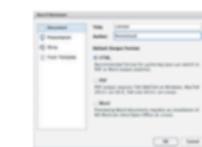
What is R Markdown?

.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.

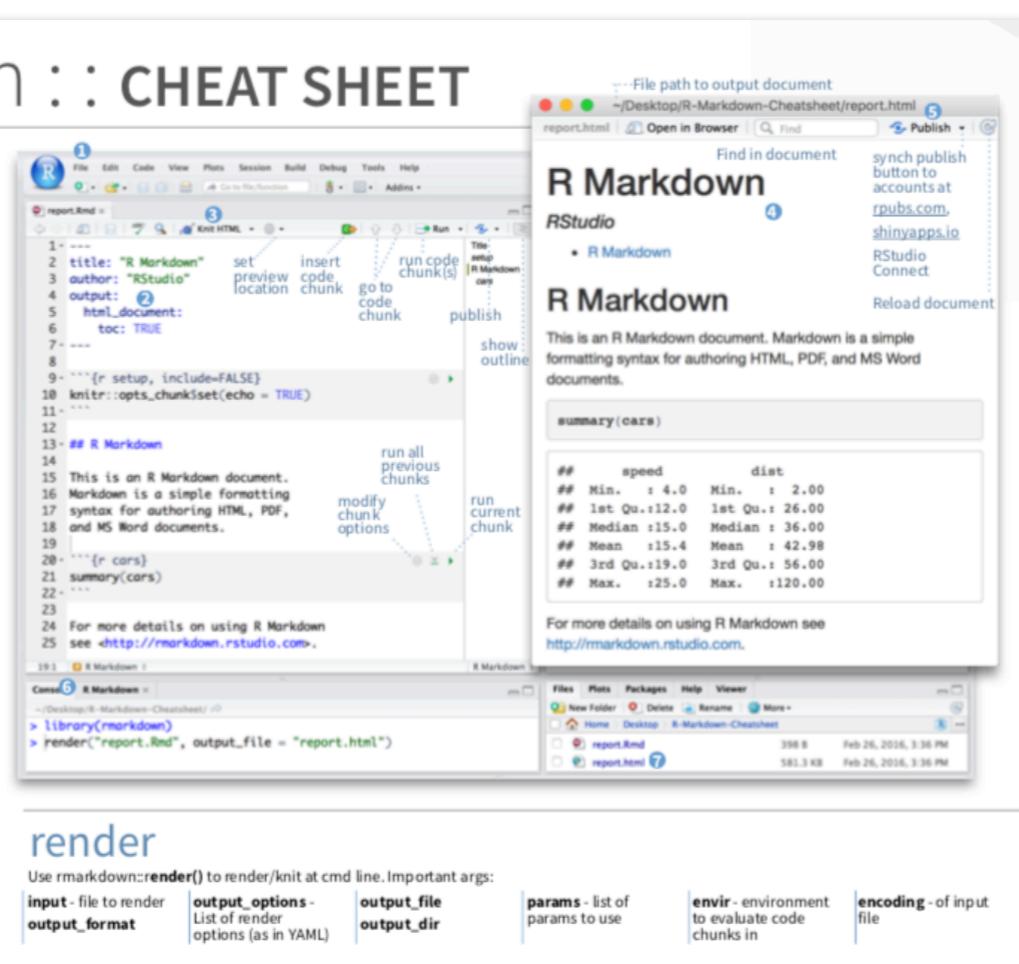
Reproducible Research - At the click of a button, or the type of a command, you can rerun the code in an R Markdown file to reproduce your work and export the results as a finished report.

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

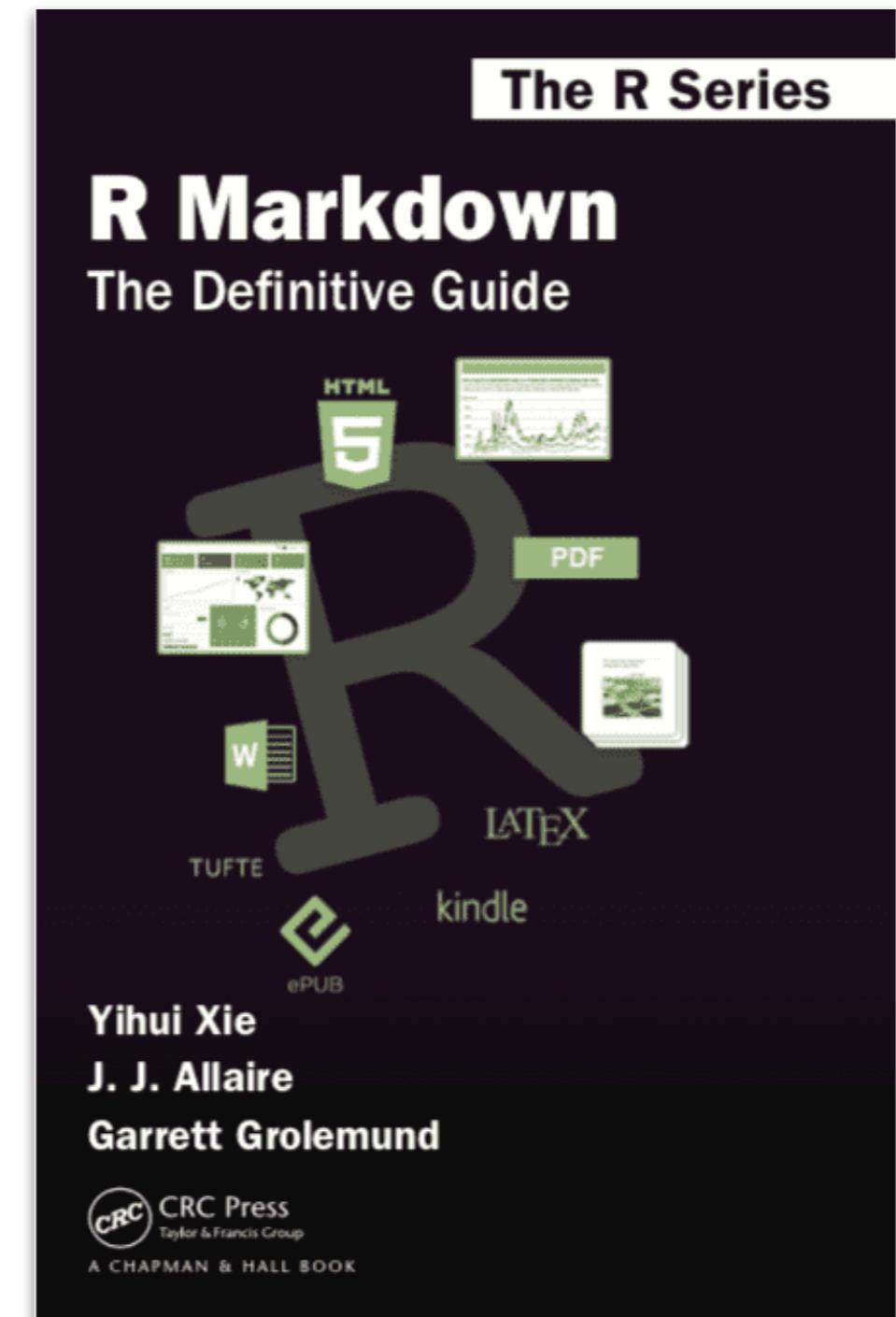


The screenshot shows the RStudio interface with several windows open:

- File Path to Output Document:** ~/Desktop/R-Markdown-Cheatsheet/report.html
- R Markdown Editor:** Shows an Rmd file with code chunks and their execution status (e.g., "run current chunk").
- R Markdown Preview:** Shows the rendered HTML output of the Rmd file.
- Console:** Shows the command: library(rmarkdown); render("report.Rmd", output_file = "report.html")
- File Explorer:** Shows the directory structure with report.Rmd and report.html files.
- Help Window:** Shows the "render" function documentation.

render
Use rmarkdown::render() to render/knit at cmd line. Important args:
input - file to render **output_options** - List of render options (as in YAML) **output_file** **output_dir**
params - list of params to use **envir** - environment to evaluate code chunks in **encoding** - of input file

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



The R Series

R Markdown

The Definitive Guide

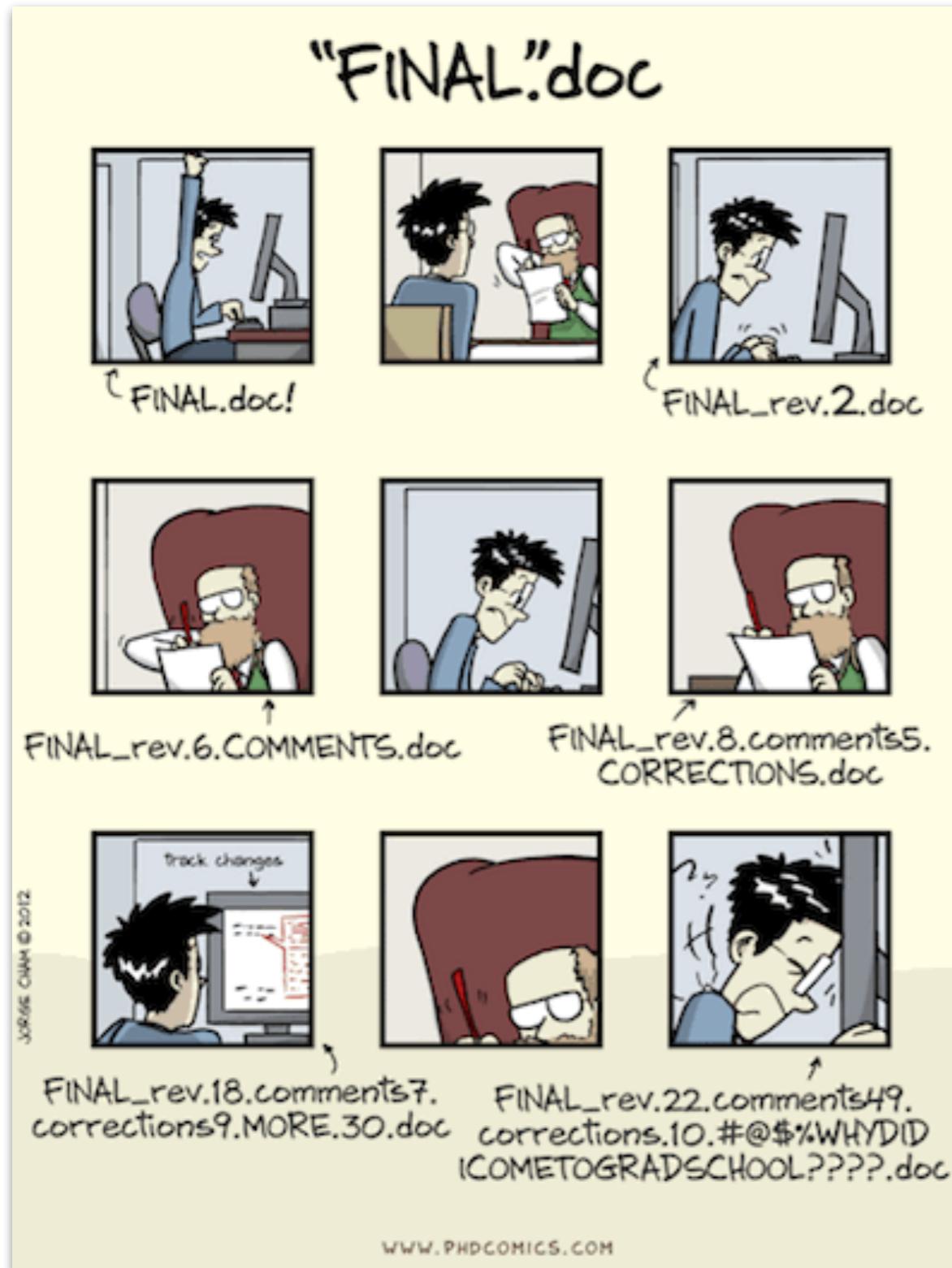
Yihui Xie
J. J. Allaire
Garrett Grolemund

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

HTML 5
PDF
TUFTE
LATEX
kindle
ePUB

<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)

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 @ 10:30am
 - get familiar with R and RStudio
 - visualization
 - data manipulation
 - simulation
 - introduce and discuss statistical methods

How will we learn?

- **Sections**

Andrew Lampinen



Shao-Fang (Pam) Wang



Role:

Teaching assistant

Teaching assistant

Email:

lampinen@stanford.edu

shaofang@stanford.edu

Office:

316

409

Office hours: Friday 12:30-1:30pm

Wednesday 1-2pm

Section:

Friday 1:30-2:20pm

Wednesday 2:30-3:20pm

in 160-314

in 160-326

How will we learn?

- **Sections**
 - work through and expand on the materials introduced in class
 - hands on exercises in smaller groups
 - learning by doing!

How will we learn?

- **Homework**
 - one assignment per week
 - you can work in groups (indicate who you worked with)
 - download and submit via Canvas
 - homework will be available online after class on Wednesdays, and is due **Tuesdays 8pm** the week after

How will we learn?

- **Midterm exam**
 - like a homework assignment, but:
 - you have to work on it **on your own**
 - you'll have **24h** to complete the exam
 - you can use any resources you like

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**
 - substantive report based on an interesting data set
 - demonstrate what's been learned in class: visualization, data manipulation, statistical modeling, reporting of results
 - publish the project on github (make sure it's reproducible)
 - more information soon ...

How will we learn?

- **Grading**
 - Homework: 40% (the homework with the lowest score doesn't count)
 - Midterm: 20%
 - Final project: 40%
 - Proposal: 5%
 - Presentation: 10%
 - Report: 25%
 - Extra credit
 - Piazza: 2%
 - Datacamp: 2%

How will we learn?

- **Office hours**

	Tobi Gerstenberg	Andrew Lampinen	Shao-Fang (Pam) Wang	Mona Rosenke
				
Role:	Instructor	Teaching assistant	Teaching assistant	Teaching assistant
Email:	gerstenberg@stanford.edu	lampinen@stanford.edu	shaofang@stanford.edu	rosenke@stanford.edu
Office:	302	316	409	424
Office hours:	Monday 2-3pm	Friday 12:30-1:30pm	Wednesday 1-2pm	Tuesday 4:30-5:30pm
Section:		Friday 1:30-2:20pm	Wednesday 2:30-3:20pm	

Tools we will use in class

- Canvas
- Course homepage
- RStudio
- PollEverywhere
- Datacamp
- Piazza
- Free online books

Tools we will use in class

- **Canvas**

- **We** will:
 - send announcements
 - upload:
 - slides
 - code
 - assignments
- **You** will:
 - read announcements :)
 - upload homework

Tools we will use in class

- Course homepage
 - general information
 - schedule with links to materials (will keep being updated)
 - <https://psych252.github.io/>
 - you can access it via canvas

PSYCH 252: STATISTICAL METHODS

Home Getting ready Schedule Information

This course offers an introduction to advanced topics in statistics with the focus of understanding data in the behavioral and social sciences. We will cover a range of methods such as regression, mixed effects models, and generalized linear models. You will learn how these methods work, as well as how to implement them using the statistical computing environment R. In addition to these more traditional methods for analyzing data, we will also discuss simulation methods (e.g. Monte Carlo, bootstrapping), and Bayesian statistics.

Team

Tobi Gerstenberg	Andrew Lampinen	Shao-Fang (Pam) Wang	Mona Rosenke
Role: Instructor	Teaching assistant	Teaching assistant	Teaching assistant
Email: gerstenberg@stanford.edu	lampinen@stanford.edu	shaofang@stanford.edu	rosenke@stanford.edu
Office: 302	316	409	424
Office hours: Monday 2-3pm	Friday 12:30-1:30pm	Wednesday 1-2pm	Tuesday 4:30-5:30pm
Section:	Friday 1:30-2:20pm	Wednesday 2:30-3:20pm	

Where and when?

Lectures: The class meets Monday, Wednesday, and Friday 10:30-11:50 in room 203 in the History Corner (building 200).

Tools we will use in class

- **RStudio**

- write code
- make beautiful plots
- understand statistical concepts
- implement statistical models
- write homework assignments using R Markdown
- contribute to reproducible research

Tools we will use in class

- **PollEverywhere**
 - quick polls in class
 - feedback at the end of class
 - address: pollev.com/psych252

How are you feeling today?



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/psych252datacamp
 - get extra credit for taking courses

Schedule				
Week	Day	Date	Topic	
Content	Reading	Resources	Datacamp	
1	M	7-Jan	Introduction	<ul style="list-style-type: none">• Course introduction
				<ul style="list-style-type: none">• Cheatsheet R Studio• Cheatsheet R Markdown 1• Cheatsheet R Markdown 2• R Markdown for class reports
				<ul style="list-style-type: none">• Introduction to R• RStudio IDE 1• RStudio IDE 2• RMarkdown
				<ul style="list-style-type: none">• ggplot part 1• ggplot part 2• Reporting
	W	9-Jan	Visualization I	<ul style="list-style-type: none">• Best practices• Introduction to RStudio• Introduction to <code>library(ggplot2)</code>• Reporting results using Rmarkdown
				<ul style="list-style-type: none">• Data visualization (#1)• Data visualization (#3)• Cheatsheet ggplot2



Tools we will use in class

- **Piazza**

- post your questions and answer your colleagues' questions (for extra credit, yay!)
- form groups for the final project
- sign up here:
<https://tinyurl.com/psych252piazza>
- access code: **sunnystats**

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 fit the bill
 - **but:** many great, free books online
 - **and:** I will point out suggested readings as we go along

Some general thoughts

Vision for this class

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

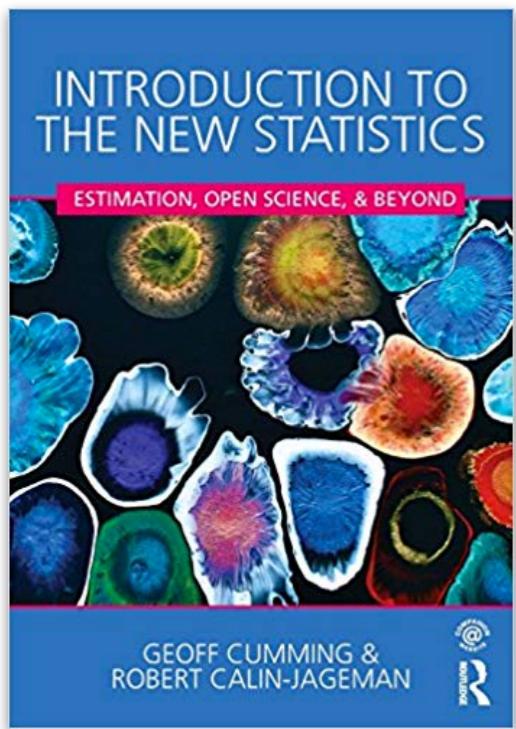
“an inspired, inclusive and collaborative community of diverse scholars, students and staff, where all are supported and empowered to thrive.”

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

Fear of statistics

- 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.

Fear of statistics



Change a Fixed Mindset to a Growth Mindset

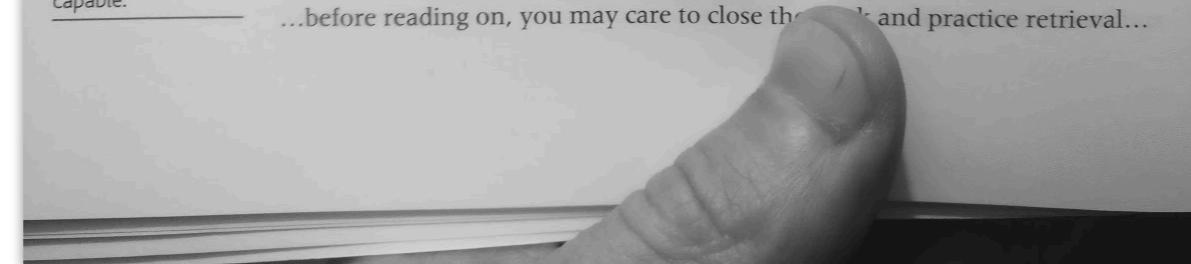
A further key idea is the distinction between a *fixed mindset* and a *growth mindset*. Carol Dweck and colleagues have demonstrated that helping students adopt a growth mindset can be a highly effective way to help them learn better and achieve more. Here's how Dweck describes the two mindsets:

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)

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

A grayscale photograph of a person's hand pointing their index finger towards the text on the page.

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

Feedback

Feedback

- ask questions in class and/or come to us after class
- post questions on Piazza (you can post anonymously)
- send us an email (but use Piazza for questions about homework or class!)
- use anonymous feedback form (link at the bottom of the course website)
<https://tinyurl.com/psych252feedback>
- come to office hours

How was the pace of today's class?

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

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

Datacamp

David Lagnado

Ewart Thomas

Henrik Singmann

Julian Jara-Ettinger

Kevin Smith

Maarten Speekenbrink

Matthew Salganik

Mika Braginsky

Mike Frank

Mine Çetinkaya-Rundel

Patrick Mair

Richard McElreath

Russ Poldrack

Stephen Dewitt

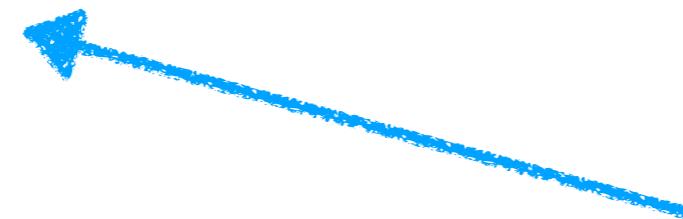
Tom Hardwicke

Thanks!

see you on **Wednesday**

bring a laptop

(with **R** and **RStudio** installed and up to date)



**post on Piazza if you
experience any problems**