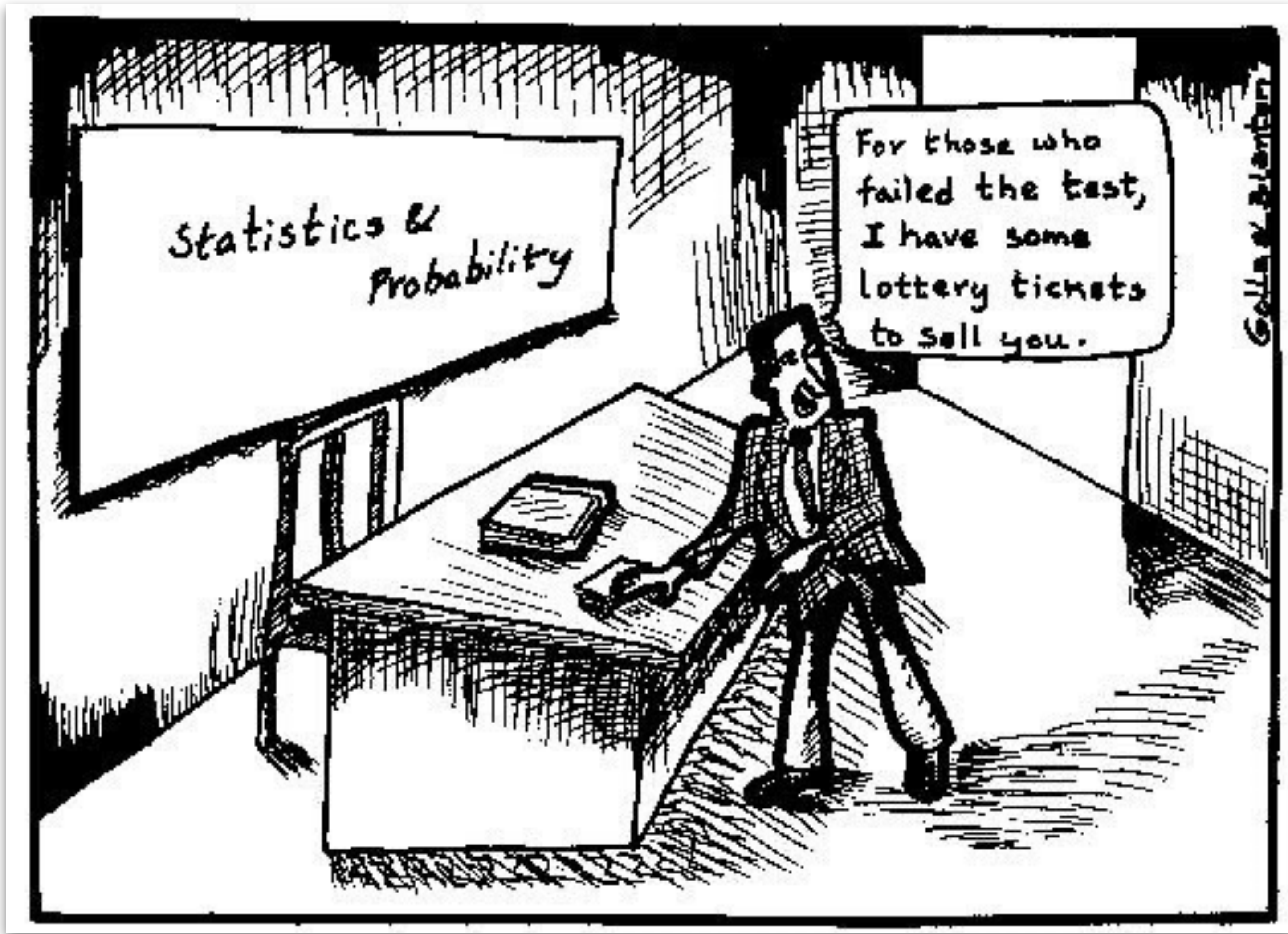
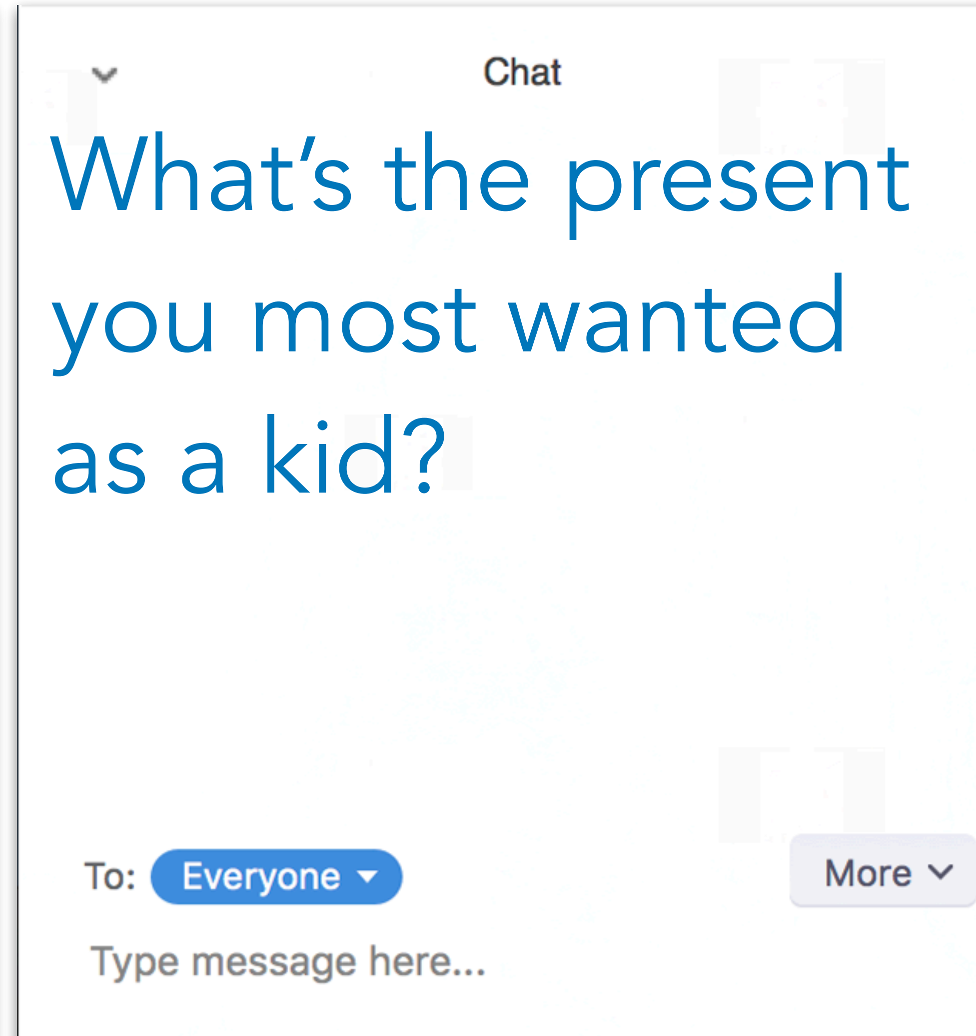
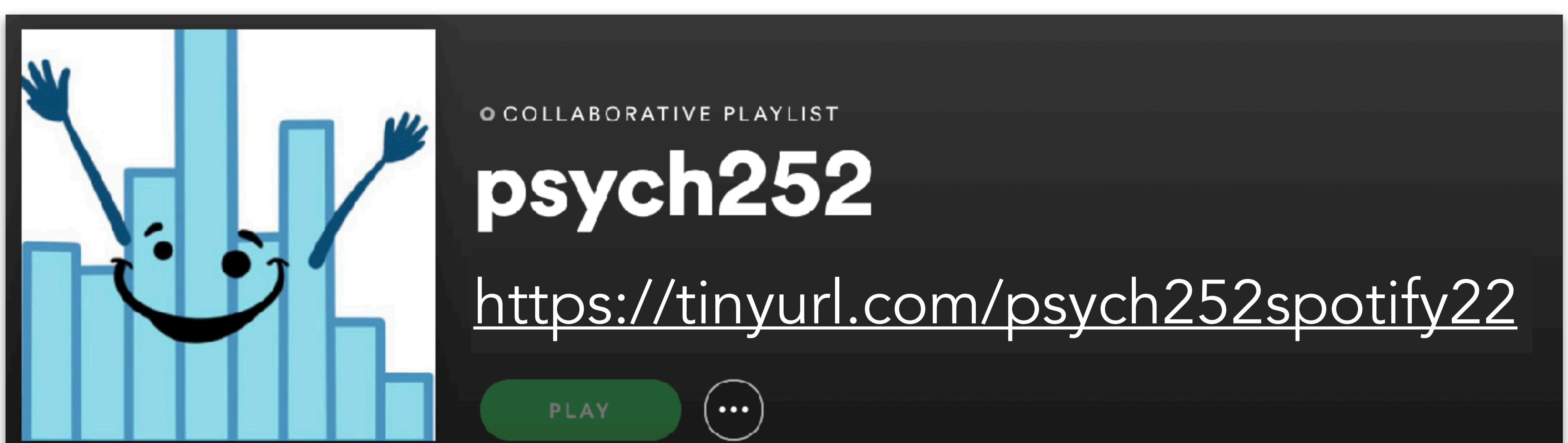
PSYCH 252: Statistical Methods for Behavioral and Social Sciences

Probability





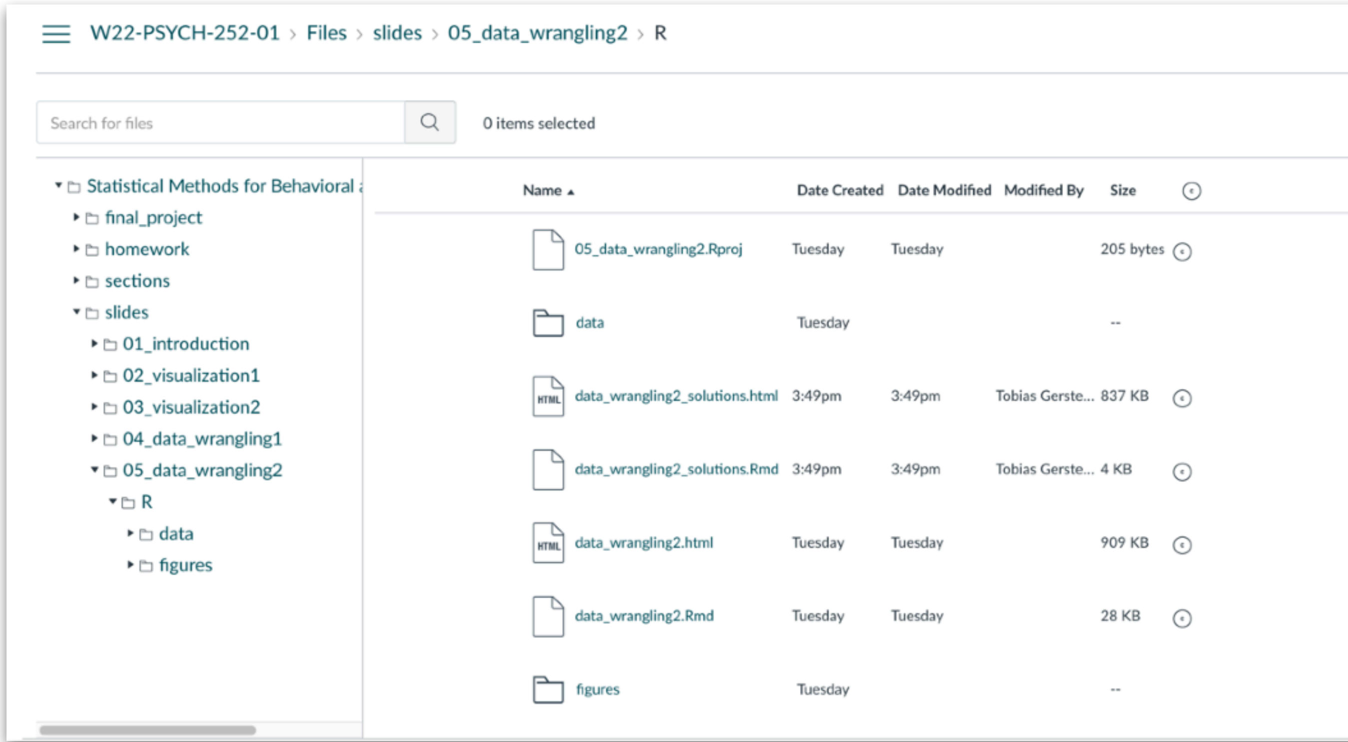


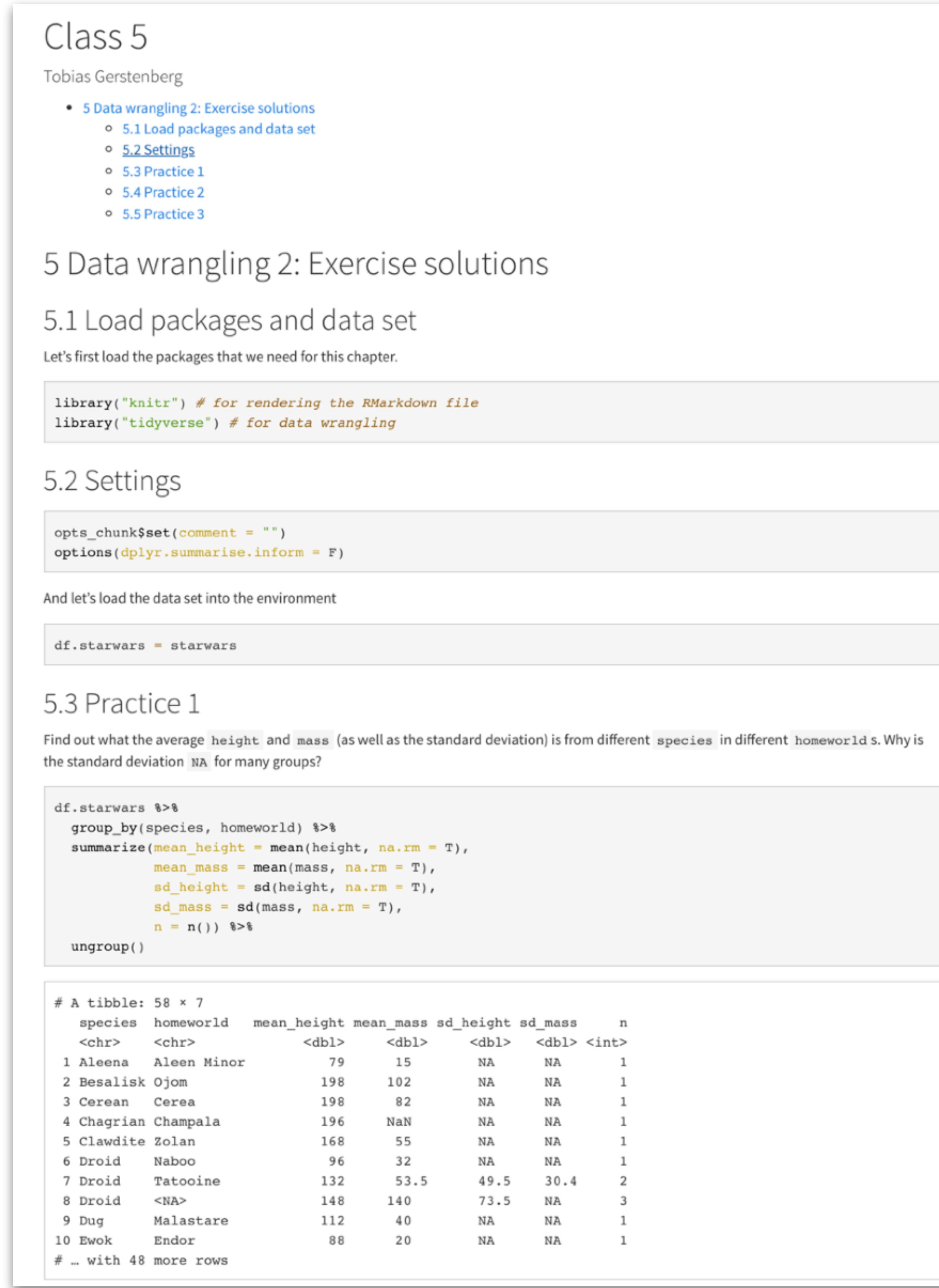
We're listening to "Video" by "India.Arie" submitted by j.av1997

01/14/2022

Logistics

Solutions to practice problems uploaded





Class 5

Tobias Gerstenberg

- 5 Data wrangling 2: Exercise solutions
  - 5.1 Load packages and data set
  - 5.2 Settings
  - 5.3 Practice 1
  - 5.4 Practice 2
  - 5.5 Practice 3

5 Data wrangling 2: Exercise solutions

5.1 Load packages and data set

Let's first load the packages that we need for this chapter.

```
library("knitr") # for rendering the RMarkdown file
library("tidyverse") # for data wrangling
```

5.2 Settings

```
opts_chunk$set(comment = "")
options(dplyr.summarise.inform = F)
```

And let's load the data set into the environment

```
df.starwars = starwars
```

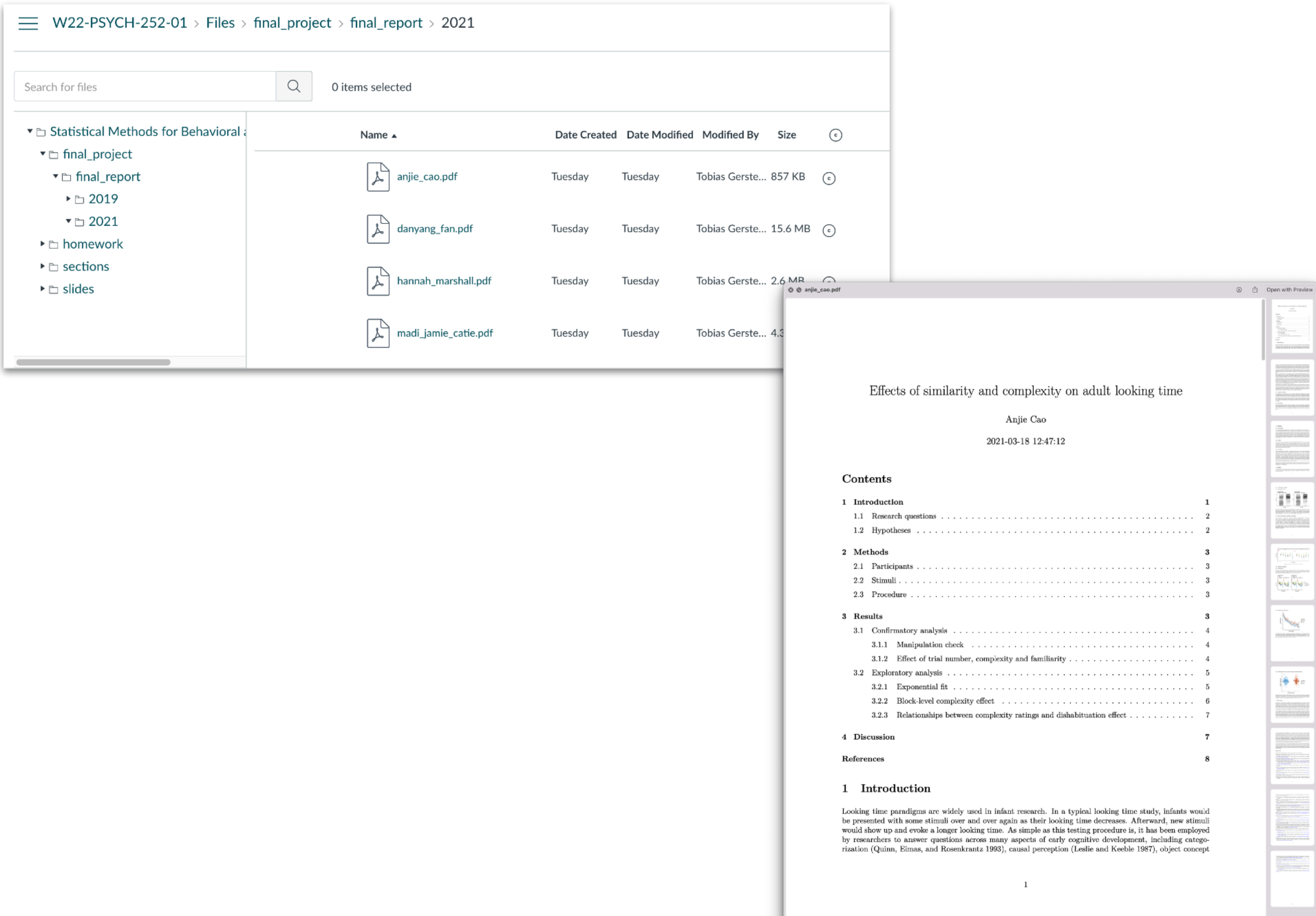
5.3 Practice 1

Find out what the average `height` and `mass` (as well as the standard deviation) is from different `species` in different `homeworld`s. Why is the standard deviation `NA` for many groups?

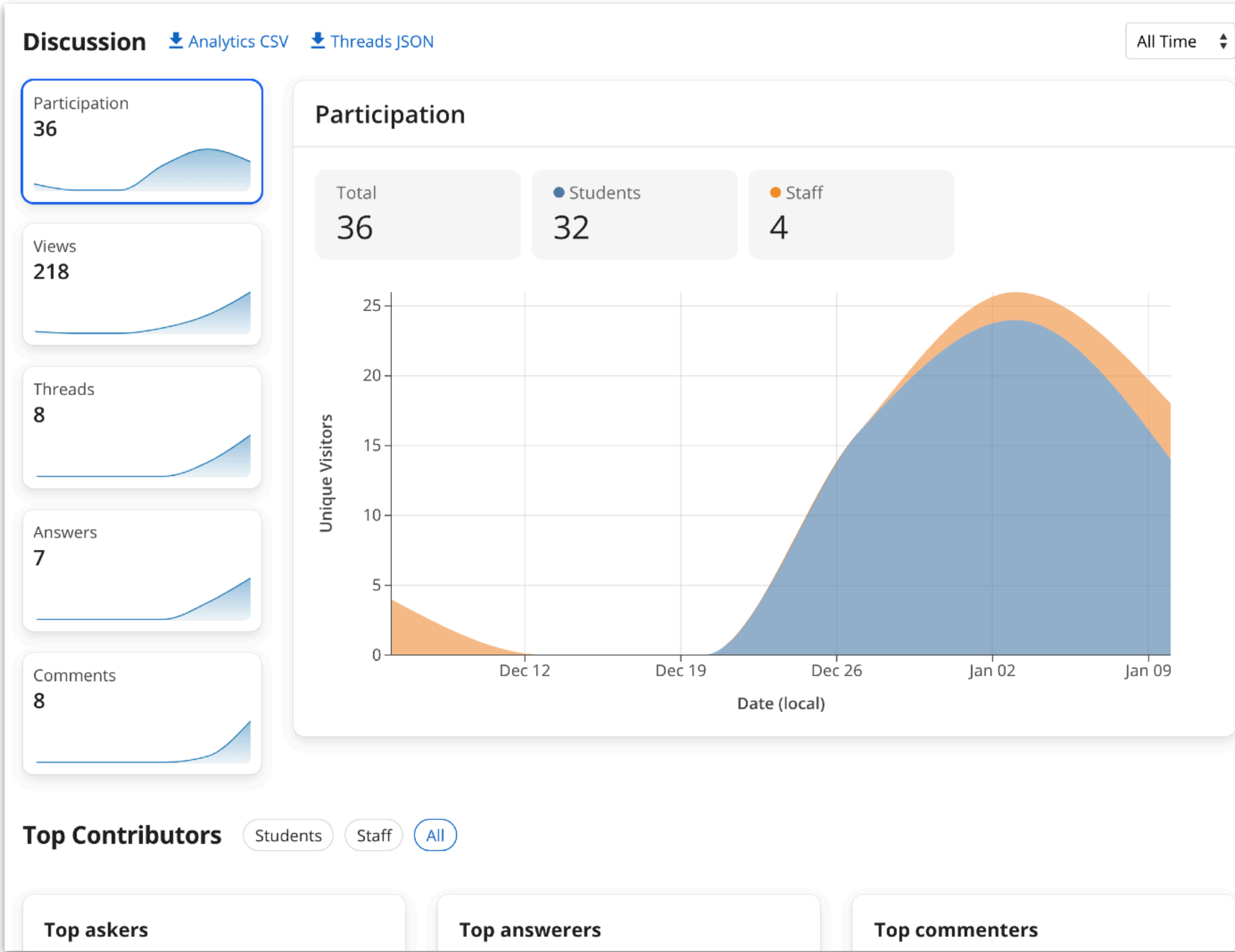
```
df.starwars %>%
  group_by(species, homeworld) %>%
  summarize(mean_height = mean(height, na.rm = T),
            mean_mass = mean(mass, na.rm = T),
            sd_height = sd(height, na.rm = T),
            sd_mass = sd(mass, na.rm = T),
            n = n()) %>%
  ungroup()
```

```
# A tibble: 58 × 7
   species  homeworld   mean_height mean_mass sd_height sd_mass     n
   <chr>    <chr>             <dbl>     <dbl>     <dbl>   <dbl> <int>
 1 Aleena   Aleen Minor          79        15        NA      NA     1
 2 Besalisk Ojom                198       102        NA      NA     1
 3 Cerean   Cerea               198        82        NA      NA     1
 4 Chagrian Champala            196       NaN        NA      NA     1
 5 Clawdite Zolan               168        55        NA      NA     1
 6 Droid    Naboo                96        32        NA      NA     1
 7 Droid    Tatooine            132        53.5      49.5    30.4   2
 8 Droid    <NA>                148       140        73.5    NA     3
 9 Dug      Malastare           112        40        NA      NA     1
10 Ewok     Endor                88        20        NA      NA     1
# … with 48 more rows
```

Final project



Ed discussion bonus points

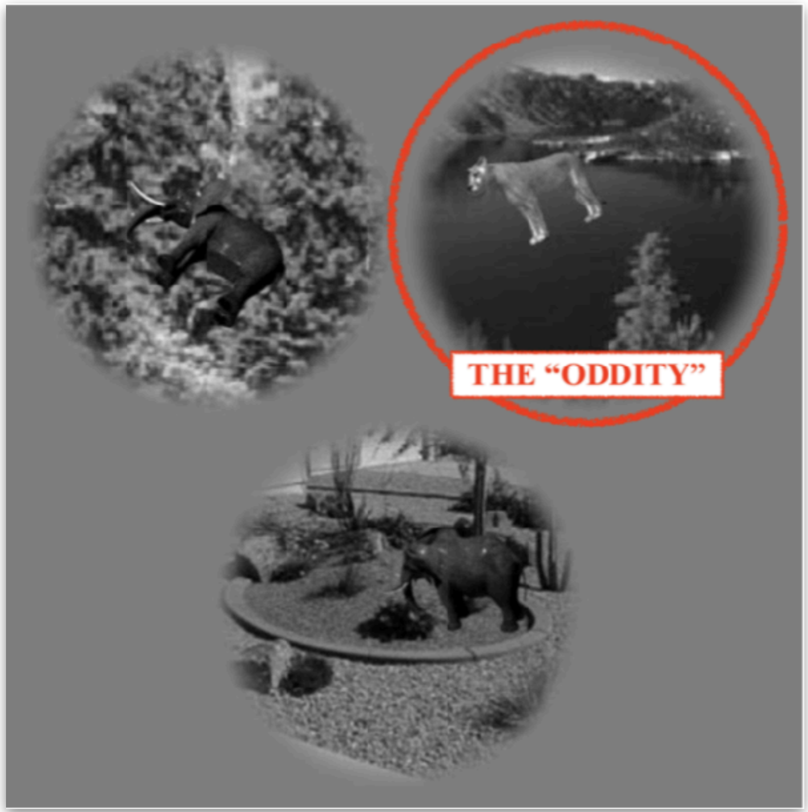


up to 2% bonus points to the final grade for active participation on Ed discussion

Plan for next week

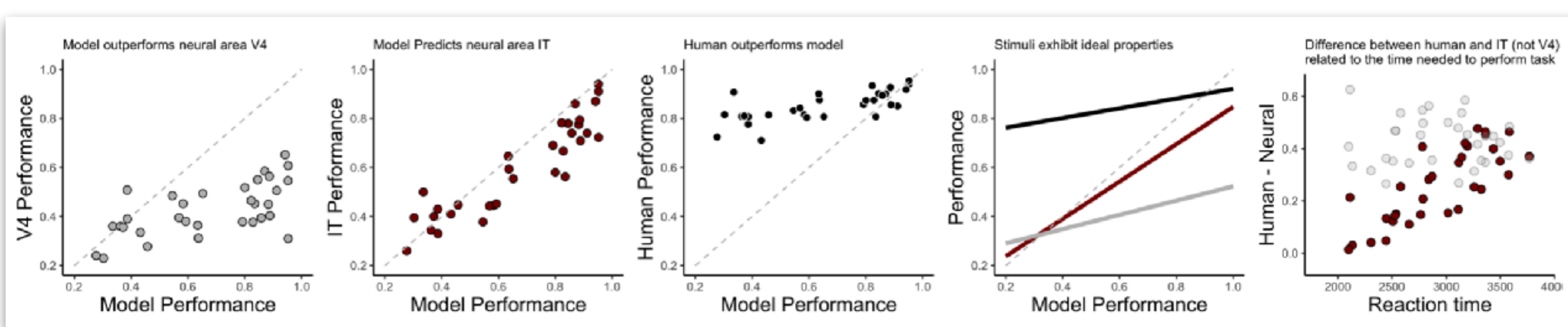
- no class on Monday (Martin Luther King, Jr. Day)
- class on Wednesday will be offered in Hybrid format
- sections next week will be offered both in person and via zoom (one TA each)

Homework 2



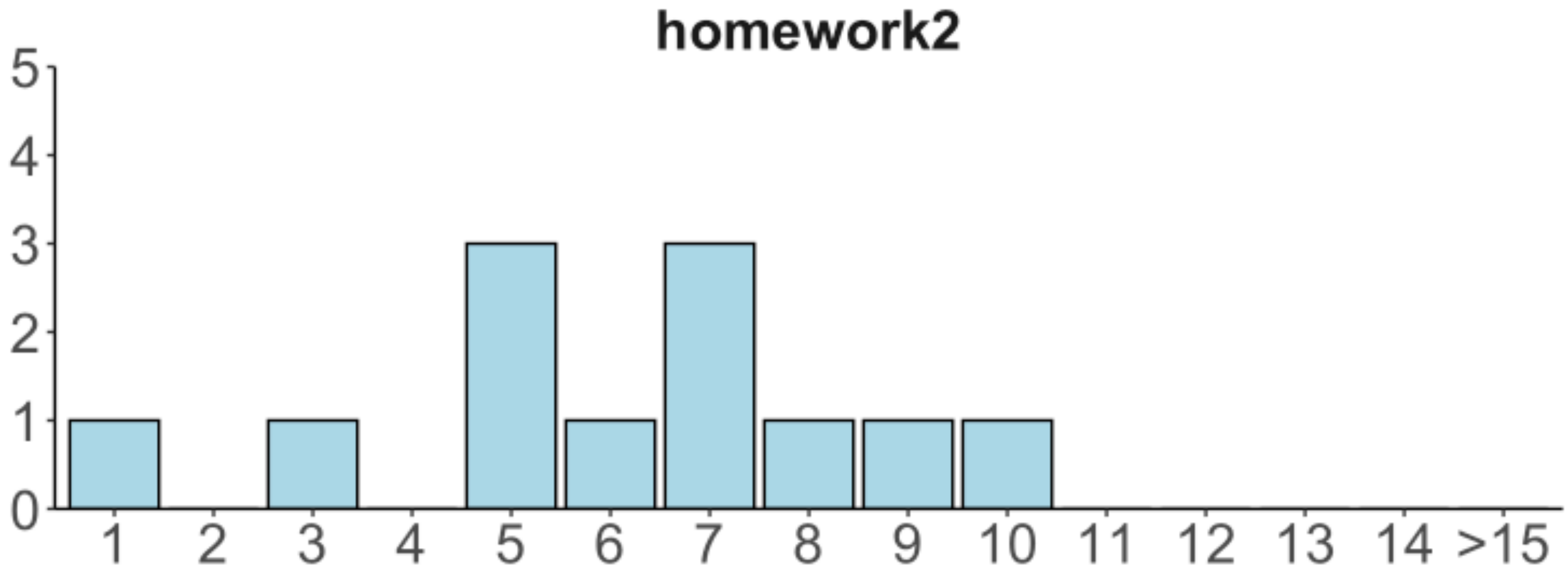
cool experiment

messy dataset



neat plot!

Homework 2



how long it took people last year

Homework 2

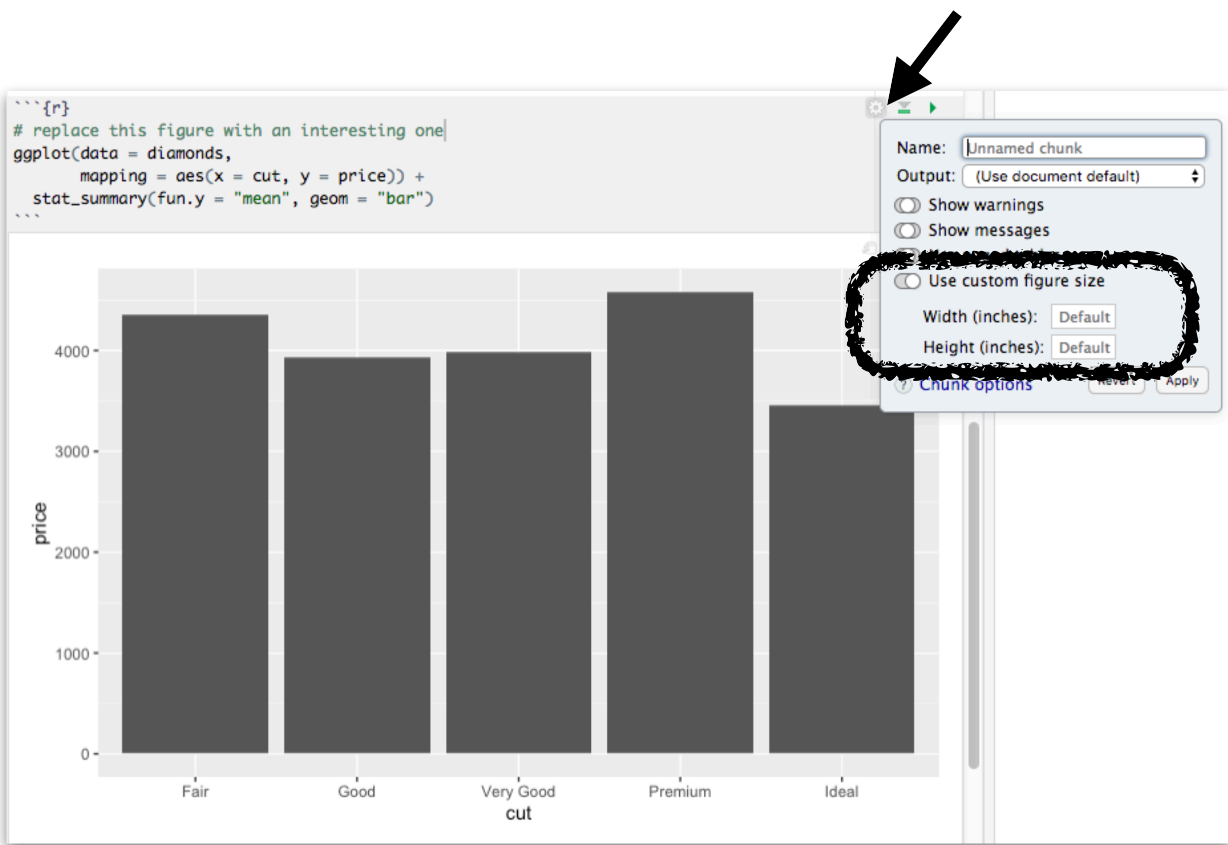
- Due **Thursday 20th, at 8pm**
- Don't wait until the very last moment to knit your RMarkdown file into a pdf. It may not compile and debugging takes time ...
- You can upload earlier versions of your homework on Canvas and still update until the deadline.
- Get and give help via Ed Discussion!

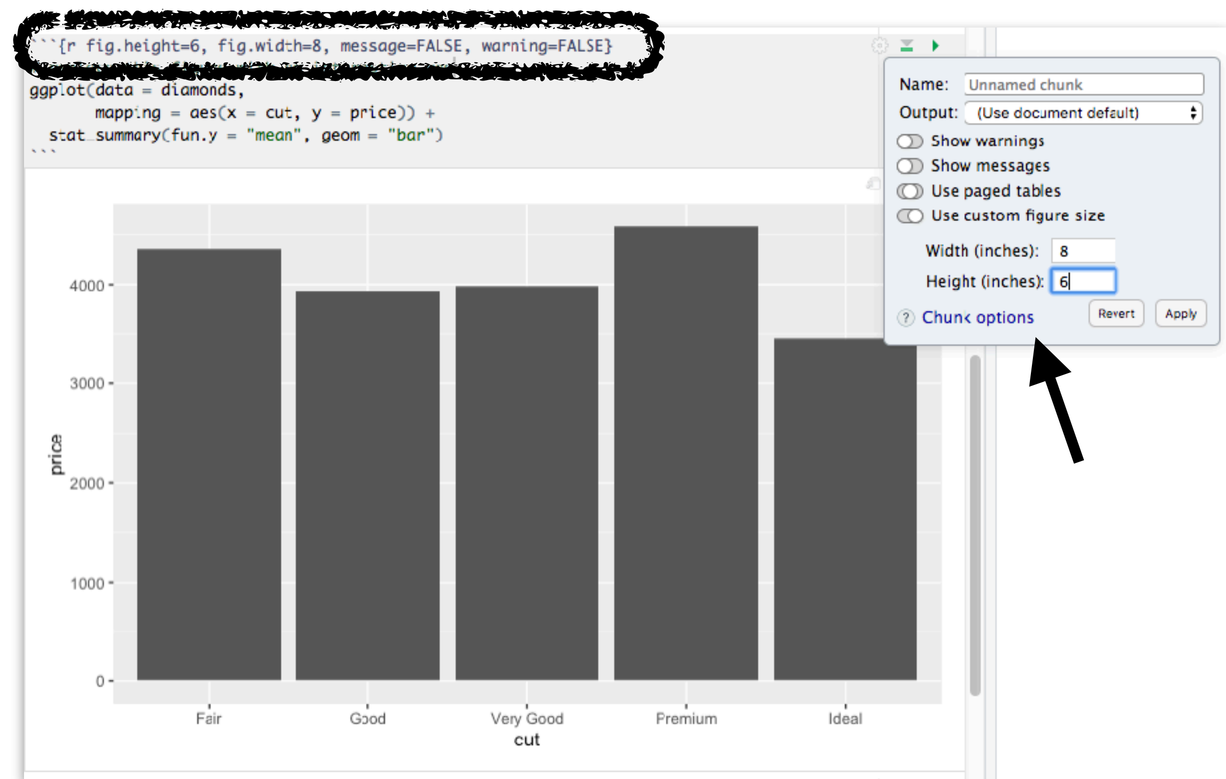
Homework 2

- **We encourage you to work in groups!**
- It's more fun
- You learn more (learning through explaining)

Homework 2

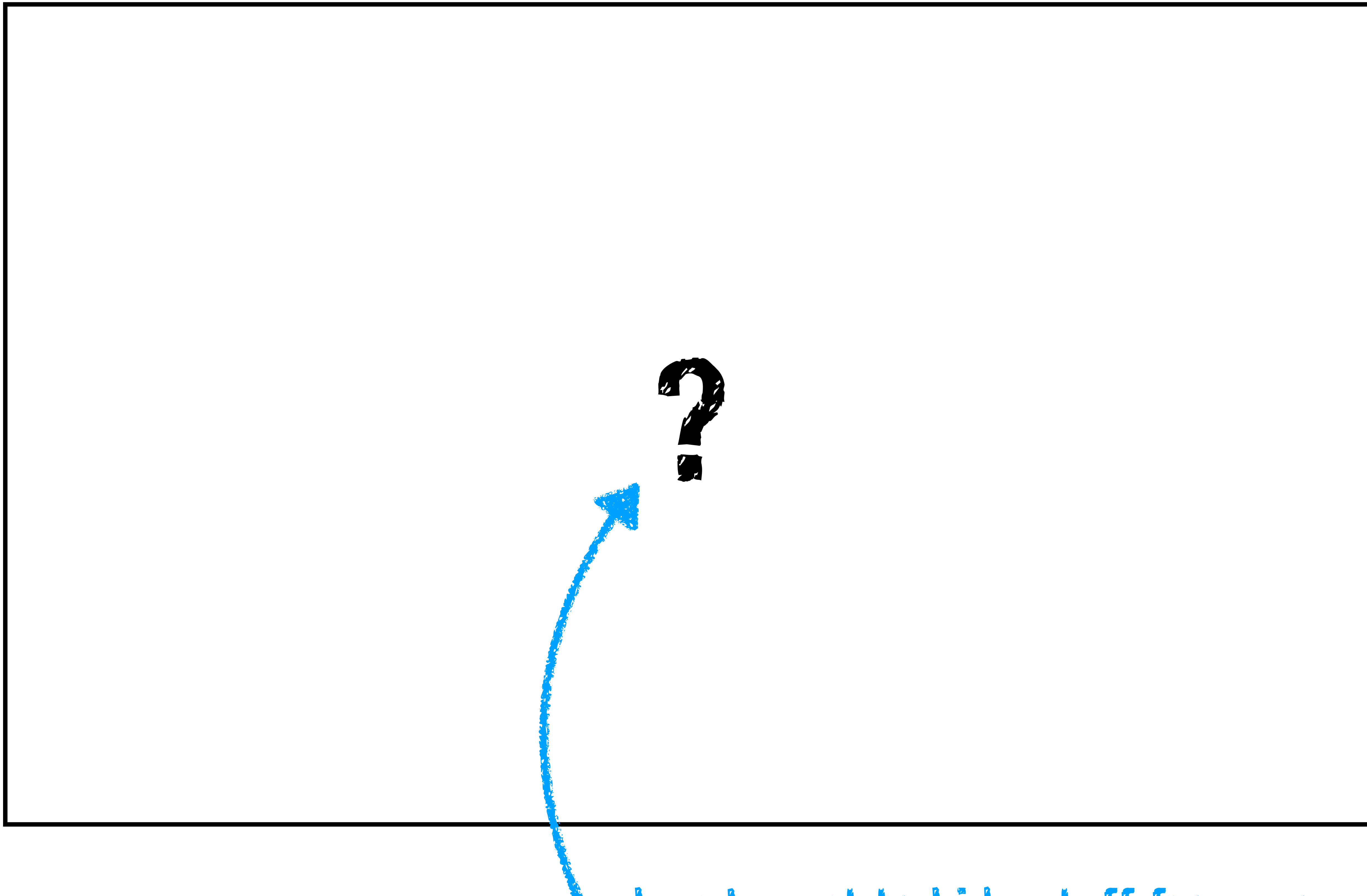
- You can adjust the figure size in your output by using the code chunk options





https://yihui.org/knitr/options/

It's a mystery

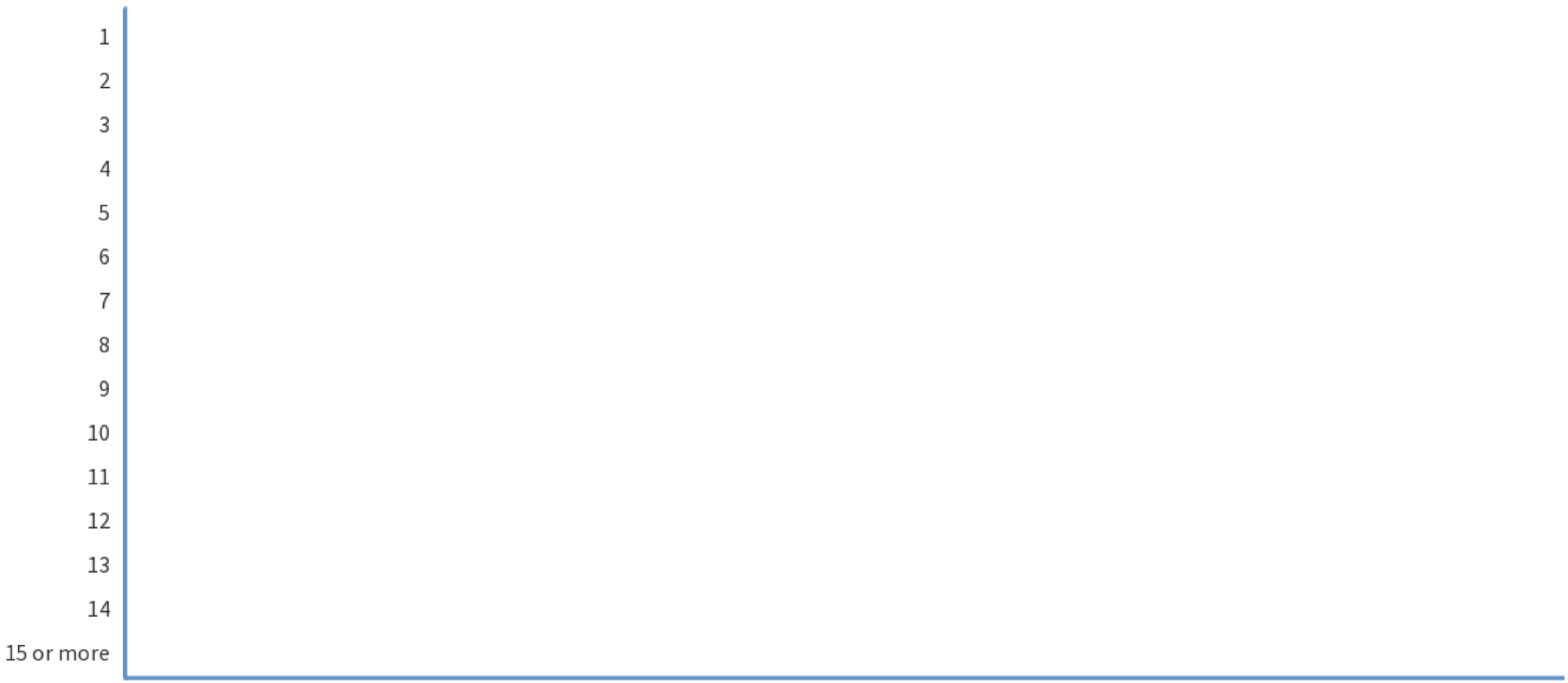


when I want to hide stuff from you

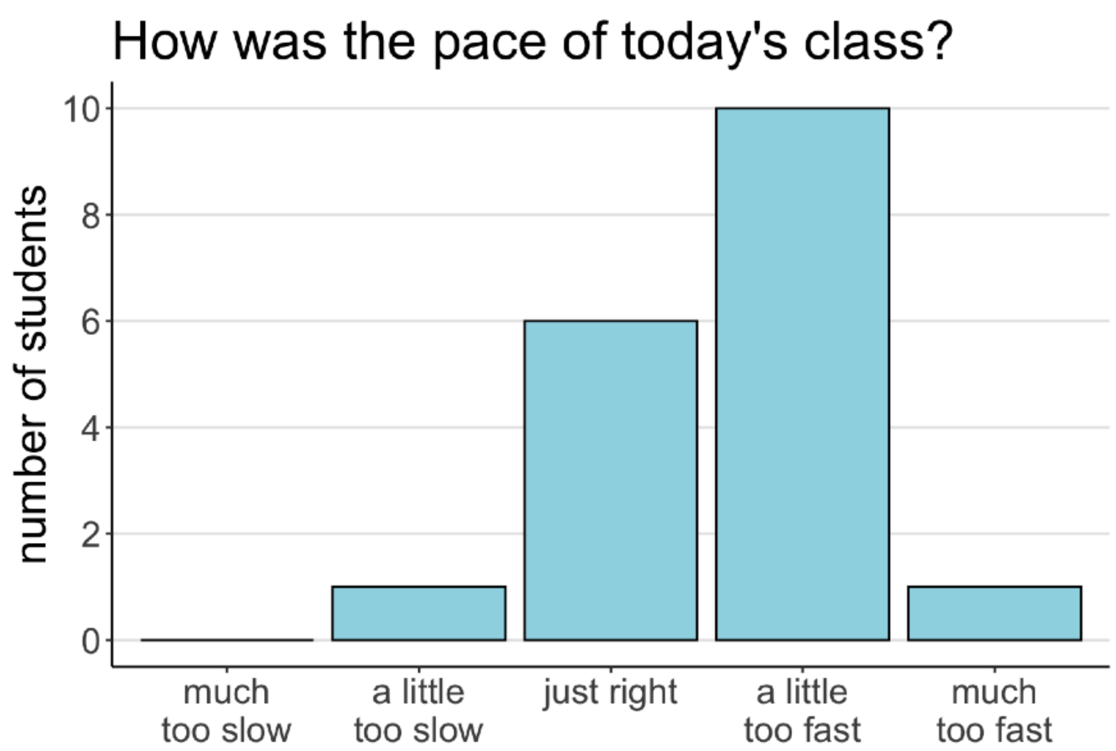
Your feedback

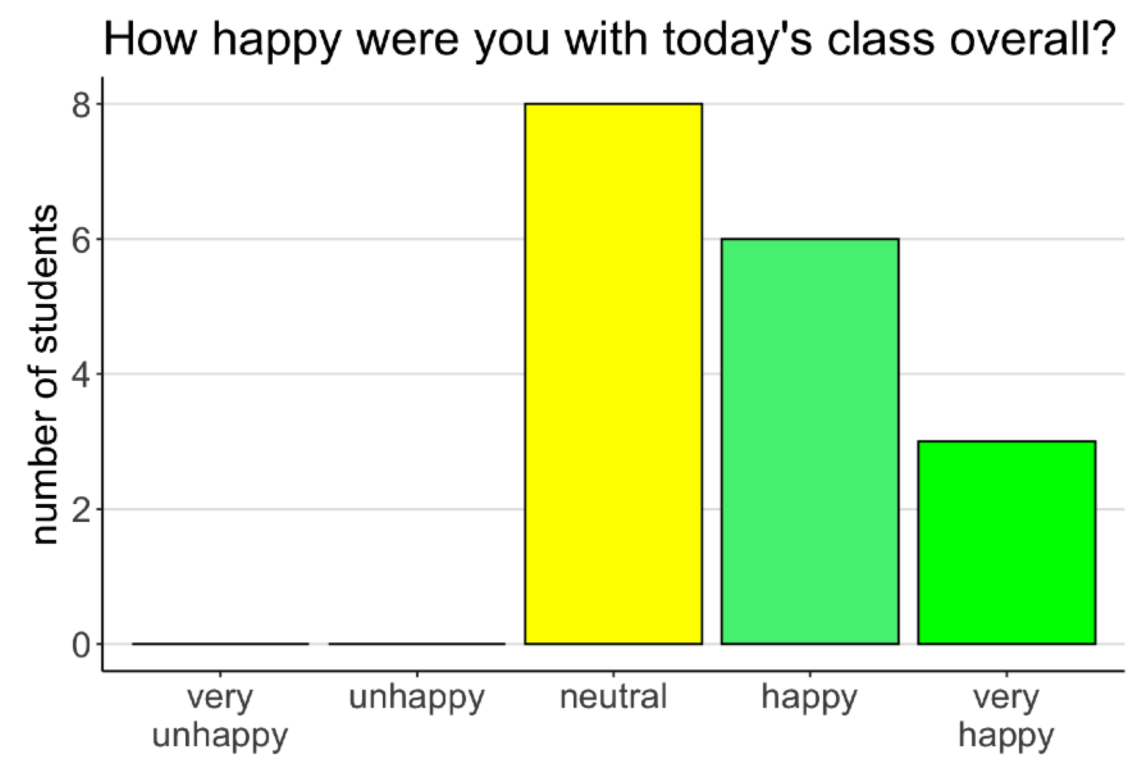
When poll is active, respond at **pollev.com/psych252**

How many hours did it take you to complete Homework 1?



Your feedback





wish we had spent more time on the more complex stuff later in class like pivoting, etc. that stuff is generally a lot less intuitive than simple stuff like adding new columns. I mean, hopefully we won't have to waste our time tidying data and we can learn actually statistics in this class so it won't matter… but ¯_(ツ)_/¯

Really would like another wrangling class just focusing on making data tidy/combining dataframes.

breakout rooms where we work on code chunks are great!

I finally got the hang of a practice problem and was able to contribute and reproduce the code on my own!! Woohoo!!!!

Most things are improved by chinchillas. If an opportunity arises, they would be much appreciated.

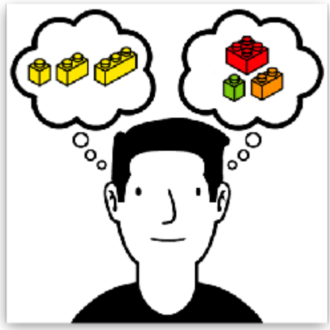
I'm nervous about how much I actually have learned and worried that I'm in worse shape than I think I am. I understand the material in the class but not sure how well I can replicate the code from scratch, or come up with what I need to do myself.

Outline

- Introduction to probability / Recap
  - Motivation
  - Counting possibilities
  - Clue guide to probability
  - Understanding Bayes' Rule
  - Getting Bayes' right matters!
  - Building a Bayesis

Motivation

What does statistics have to do with probability?



Theory

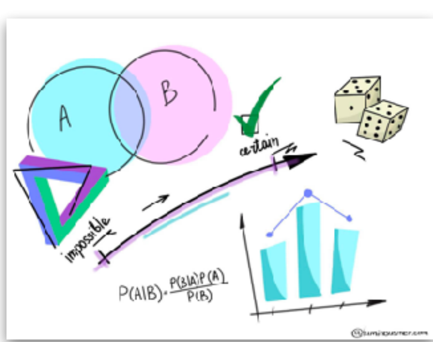
Our goal is to develop theories. In psychology, theories of how the mind works.

Prediction

Our theories need to make testable/falsifiable predictions.

Uncertainty

Because the domains that we are interested in are fundamentally uncertain (e.g. we want to say something about people generally but can only test a sample), we formulate and test these predictions using statistical models.

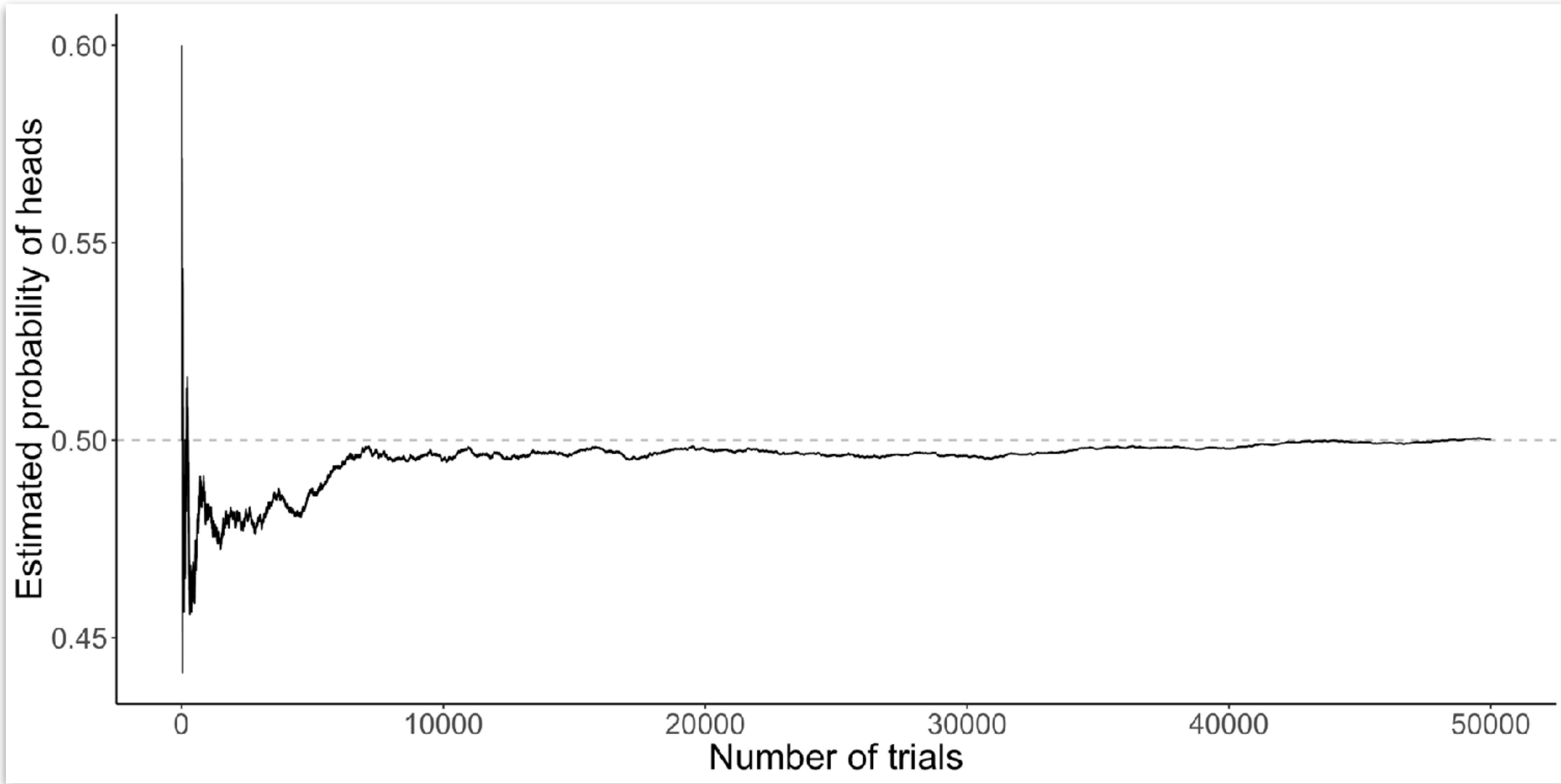


Probability

Probability theory is the formal language for dealing with uncertainty.

Frequentist interpretation

Probabilities = **long-range frequencies**



law of large numbers = empirical probability will approximate the true probability as the sample size increases

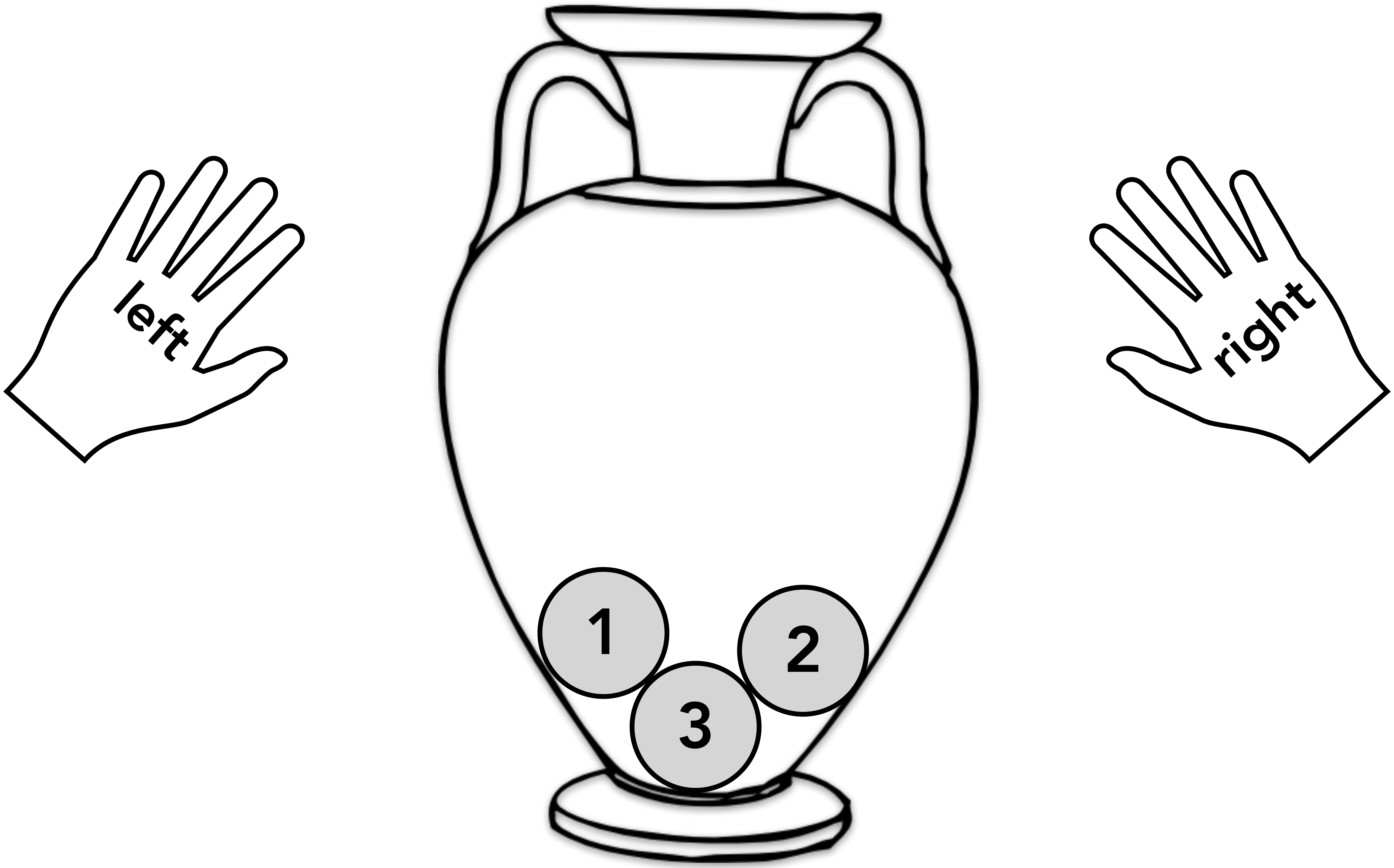
Subjective/Bayesian interpretation

Probabilities = **subjective degrees of belief**

- applies to events which may only happen once
- **"What's the probability that humans will land on Mars someday?"**
- probabilities are not a property of the world, but of a person's beliefs about the world
- at the heart of Bayesian data analysis

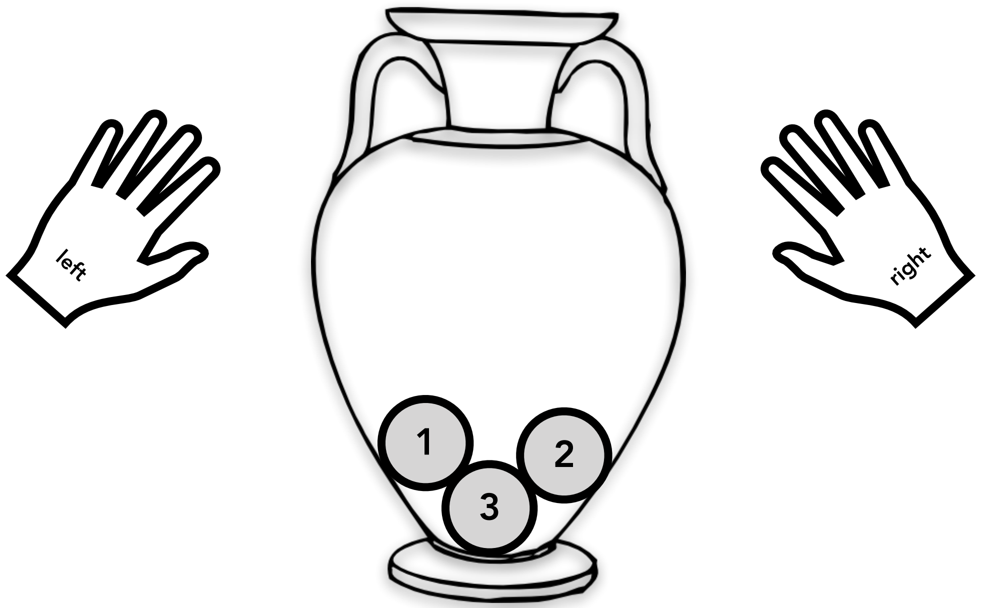
Counting possibilities

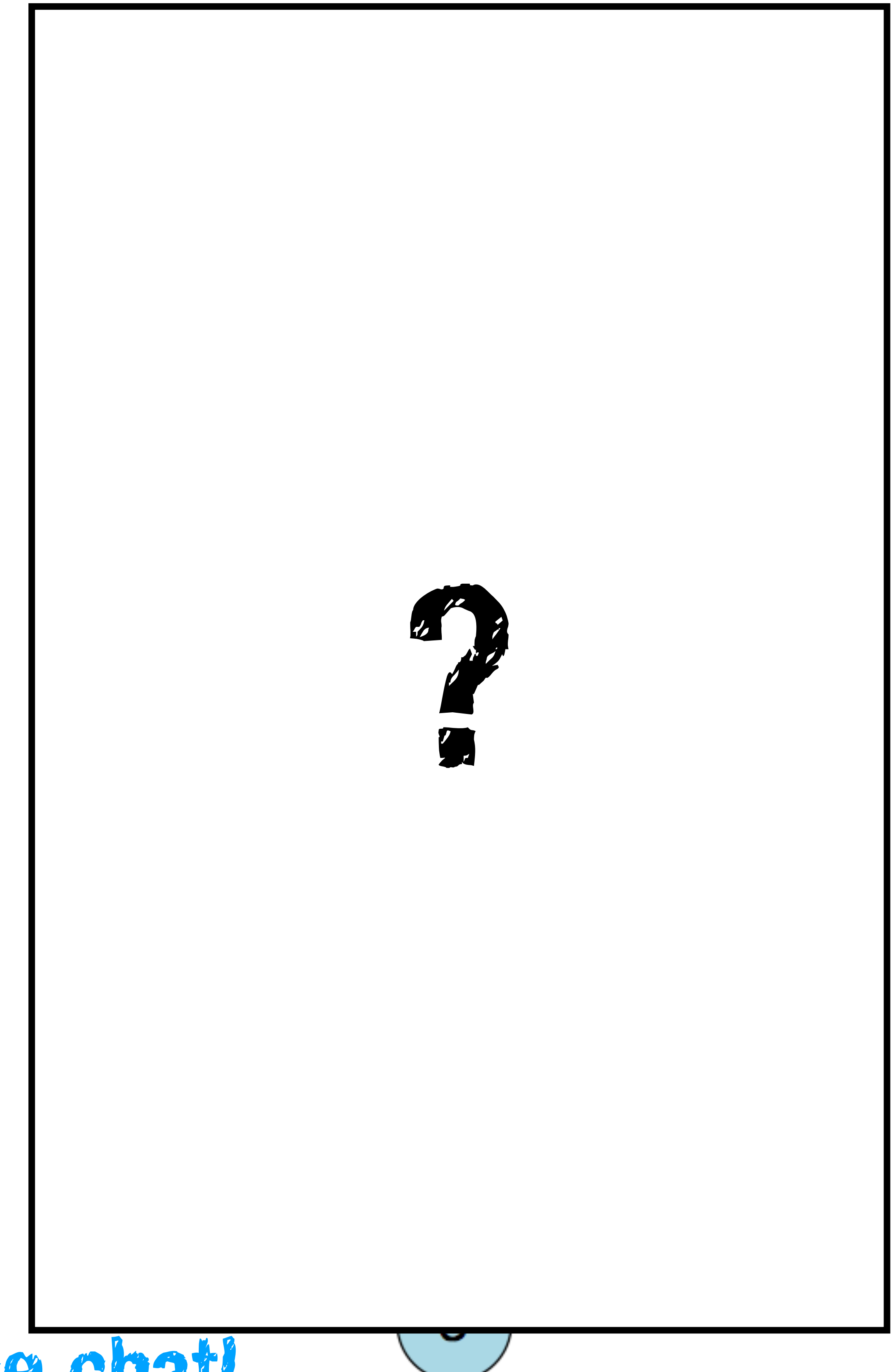
no stats class without urns!



Sampling **with** replacement

$$p(\text{left} = 1, \text{right} = 2) = ?$$



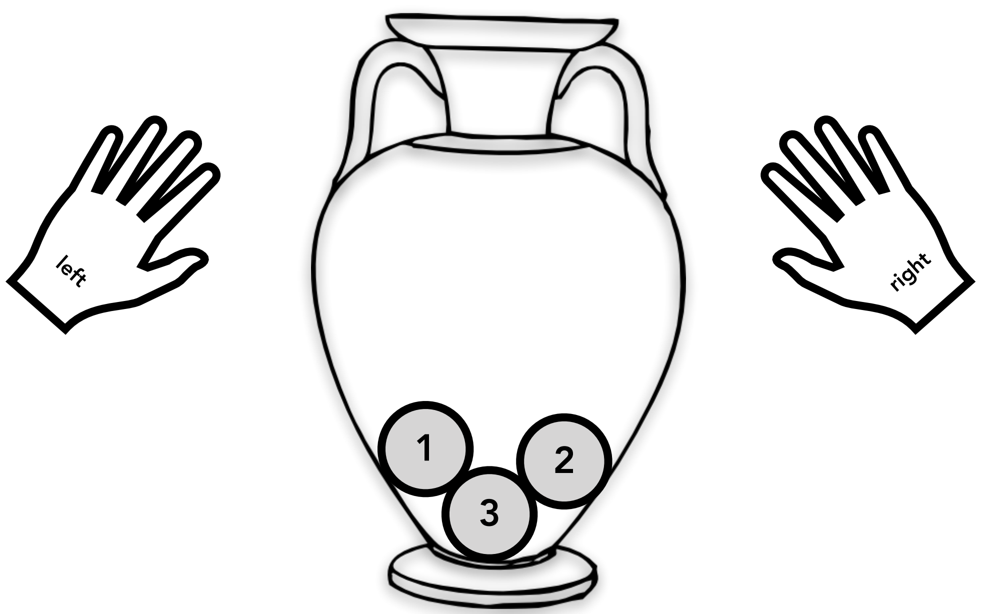


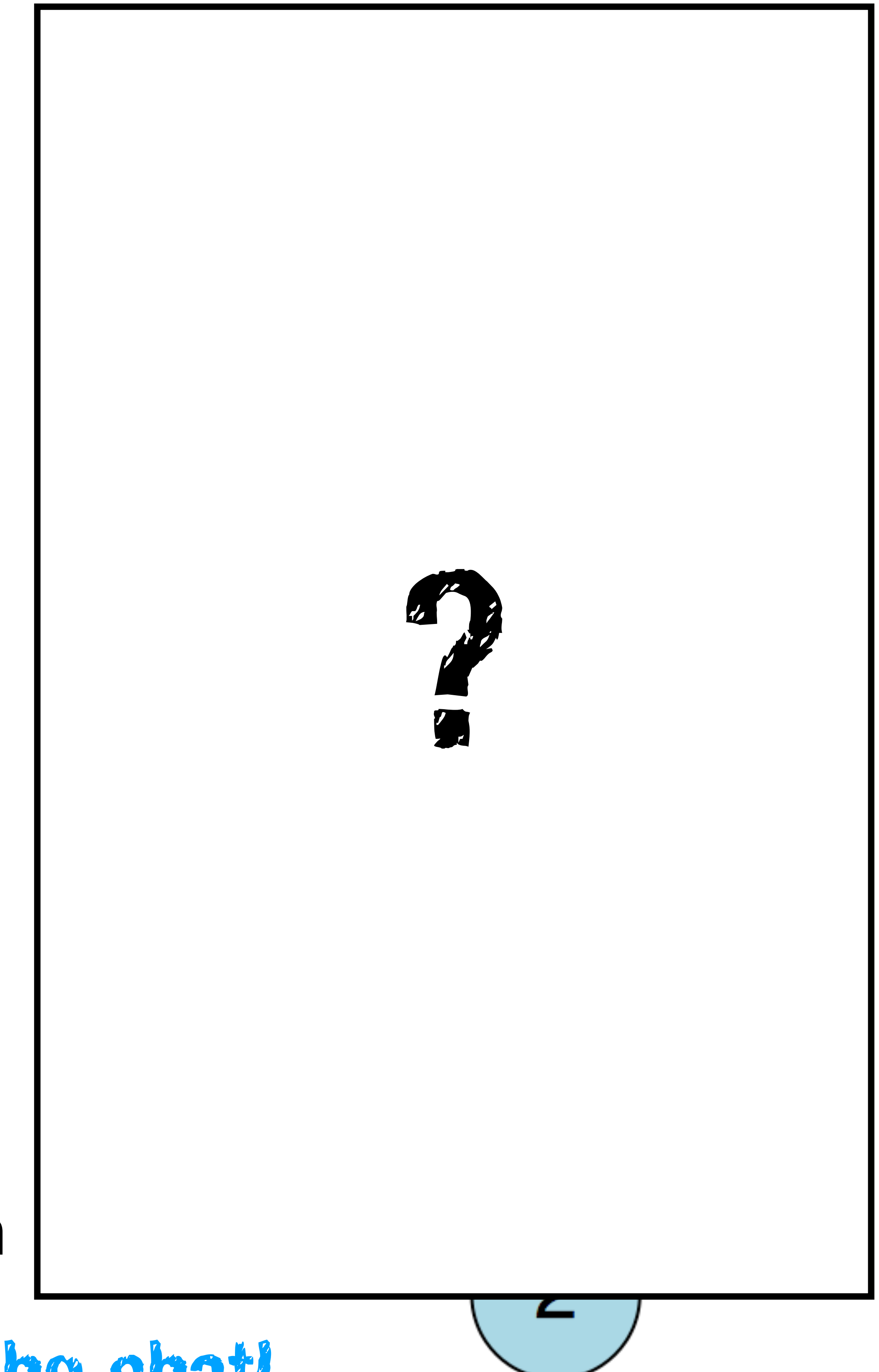
What is the probability that I first draw the 1 with my left hand, and then, after putting the 1 back into the urn again, draw the 2 with my right hand?

write in the chat!

Sampling **without** replacement

$$p(\text{left} = 1, \text{right} = 2) = ?$$



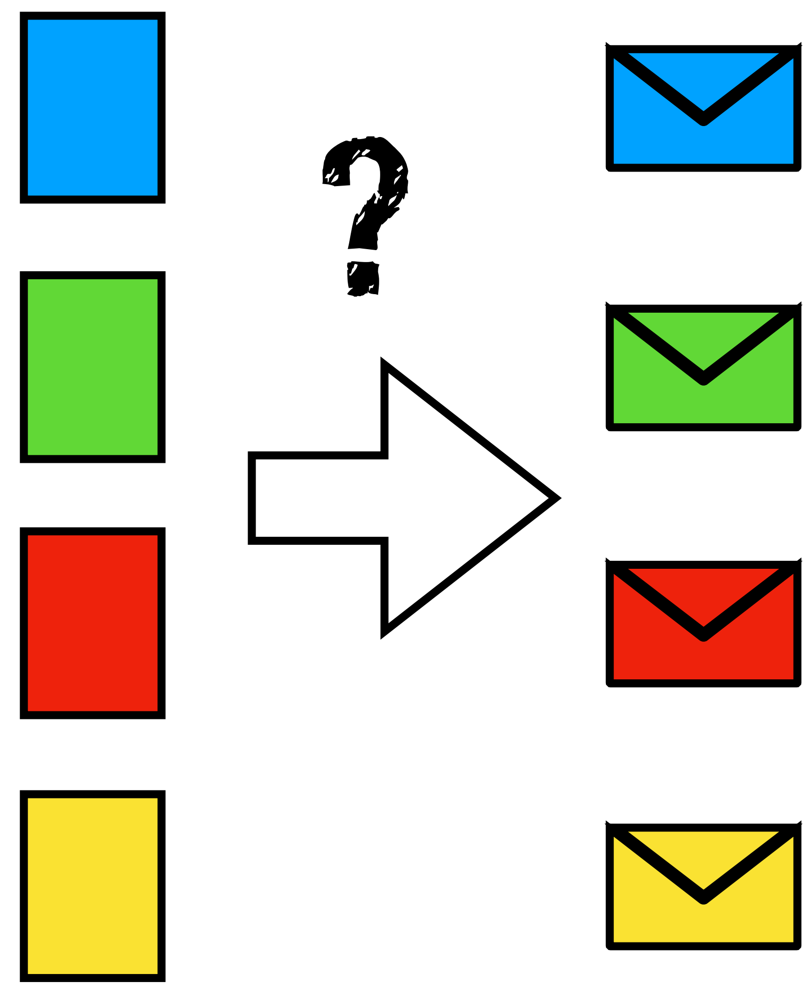


What is the probability that I first draw the 1 with my left hand, and then, without putting the 1 back into the urn, draw the 2 with my right hand?

write in the chat!

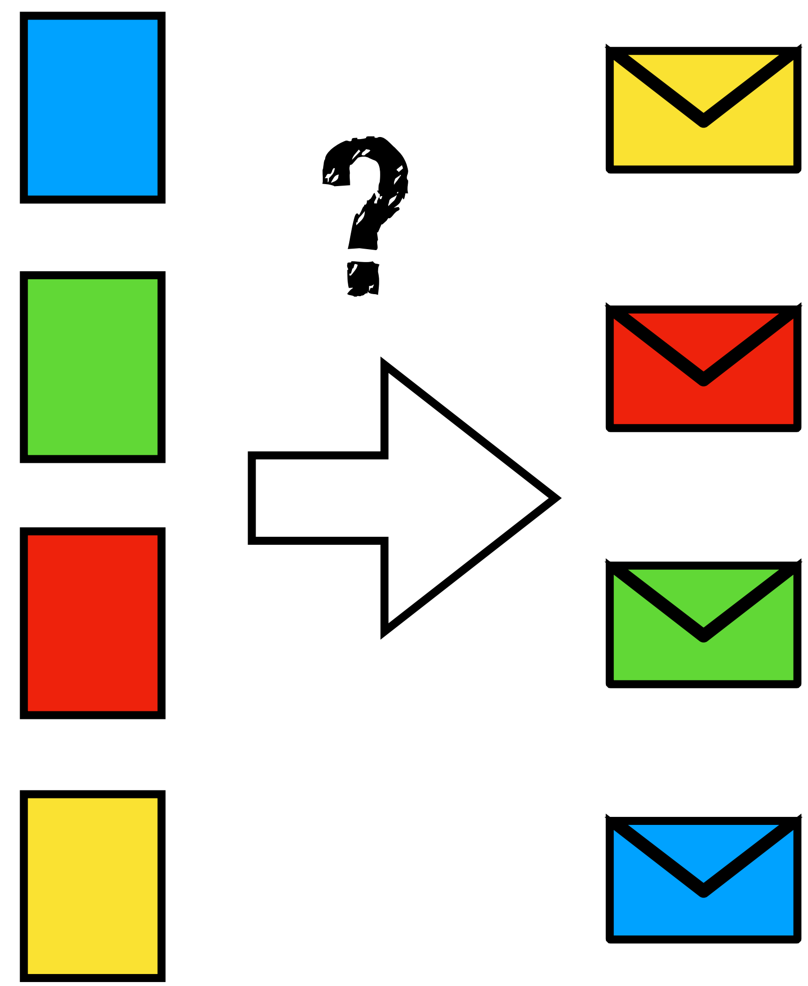
Random secretary

A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, each in a different envelope, what is the probability that exactly three letters will go into the right envelope?



Random secretary

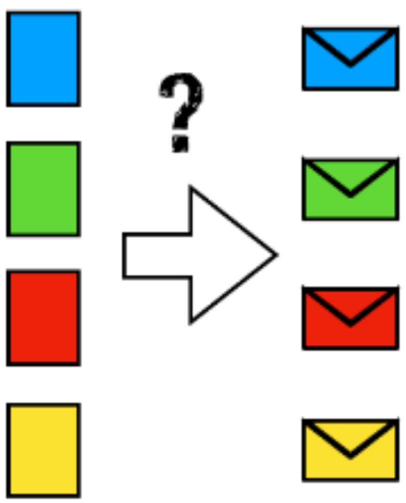
A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, each in a different envelope, what is the probability that exactly three letters will go into the right envelope?



What is the probability that exactly three letters will go into the right envelope?

Random secretary

A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, each in a different envelope, what is the probability that exactly three letters will go into the right envelope?



Random secretary

?

Random secretary

?

Naive definition of probability

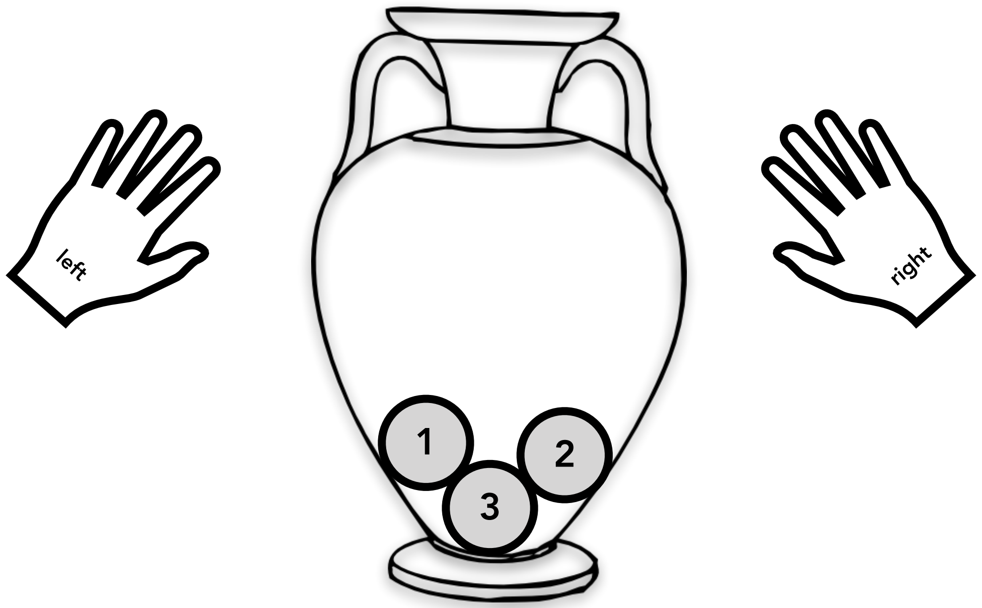
$$P_{\text{naive}}(A) = \frac{\text{number of outcomes favorable to A}}{\text{number of outcomes}}$$

if all outcomes are equally likely!

Definitions

Experiment: Activity that produces or observes an outcome.

Drawing 2 marbles from the urn with replacement, and noting the order.



Sample Space: Set of possible outcomes for an experiment.

$\Omega = \{(1, 1), (1, 2), ..., (2,1), ..., (3, 3)\}$

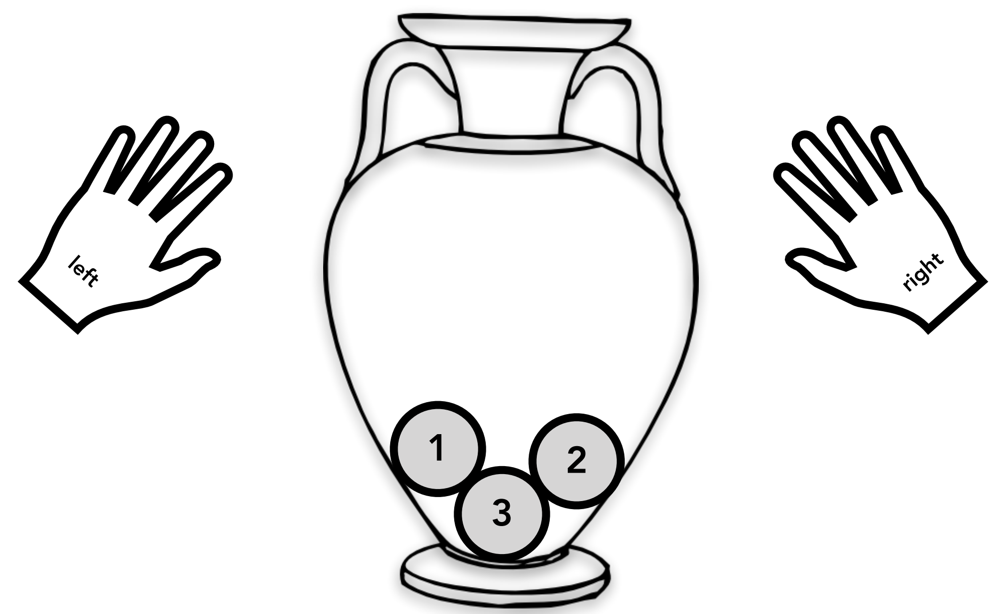
Event: Subset of the sample space. (1, 1)

Definitions

If $P(X_i)$ is the probability of event X_i

1. Probability cannot be negative.

$$P(X_i) \geq 0$$



2. Total probability of all outcomes in the sample space is 1.

$$\sum_{i=1}^{N} P(X_i) = P(X_1) + P(X_2) + \ldots + P(X_N) = 1$$

Clue guide to probability

Clue guide to probability

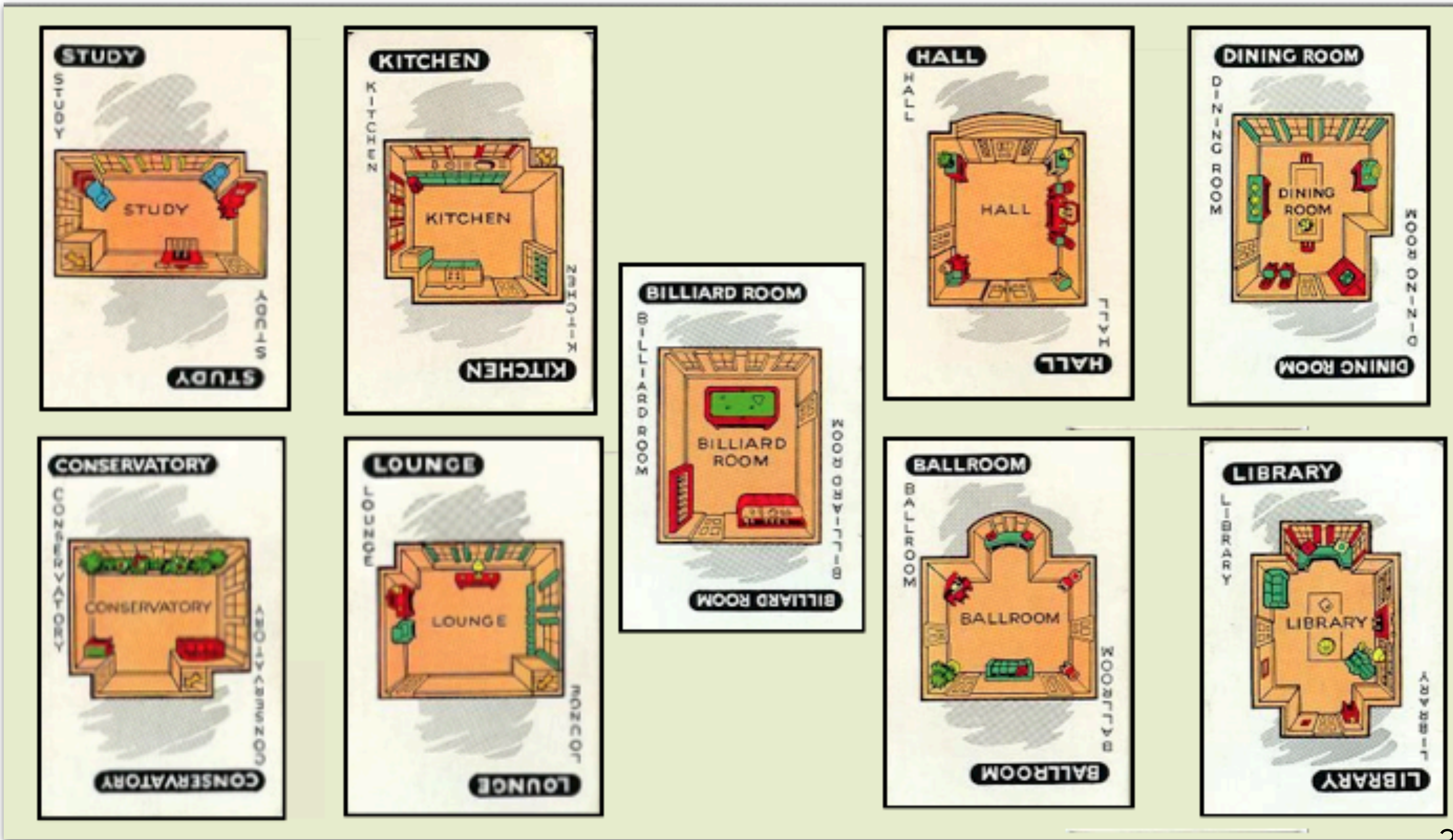
Who killed Mr Boddy?

Clue guide to probability

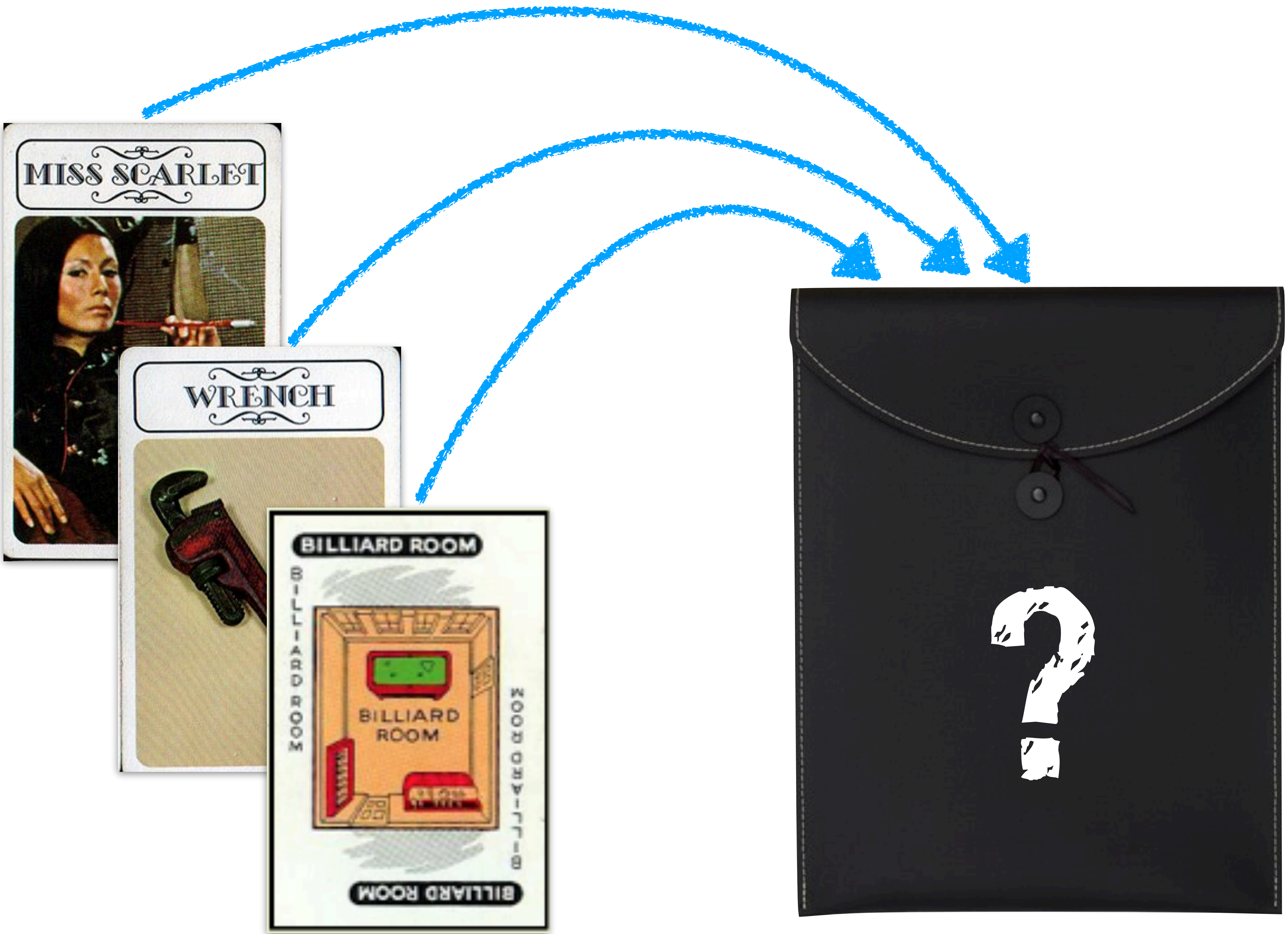
Who killed Mr Boddy, **with what?**

Clue guide to probability

Who killed Mr Boddy, with what, **and where?**



Clue guide to probability



Clue guide to probability

```r
who = c("ms_scarlet", "col_mustard", "mrs_white",
        "mr_green", "mrs_peacock", "prof_plum")
what = c("candlestick", "knife", "lead_pipe",
         "revolver", "rope", "wrench")
where = c("study", "kitchen", "conservatory",
          "lounge", "billiard_room", "hall",
          "dining_room", "ballroom", "library")

df.clue = expand_grid(who = who,
                      what = what,
                      where = where)
```

all combinations

Ω

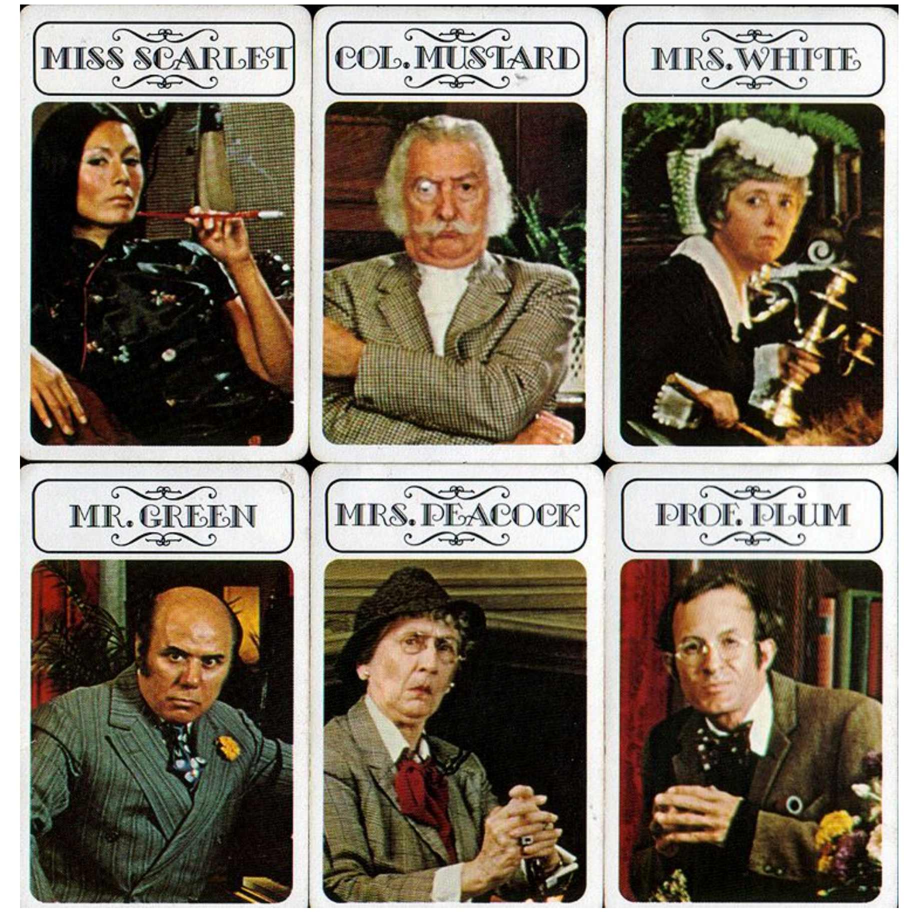
who	what	where
ms_scarlet	candlestick	study
ms_scarlet	candlestick	kitchen
ms_scarlet	candlestick	conservatory
ms_scarlet	candlestick	lounge

⋮

`nrow(df.clue)` = 324

Clue guide to probability

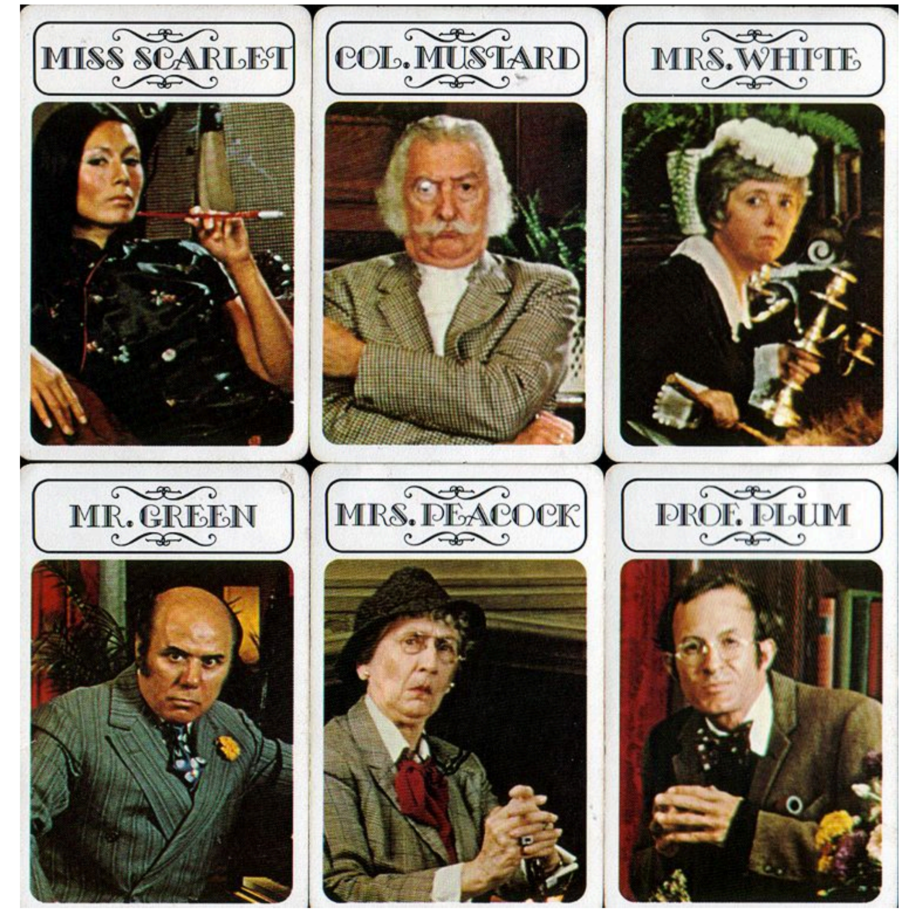
Who?



- 6 suspects
- mutually exclusive and exhaustive
- $p(\text{who} = \text{one of the six}) = 1$
- each equally likely a priori
- $p(\text{who} = \text{Prof. Plum}) = \frac{1}{6}$

Clue guide to probability

Who?

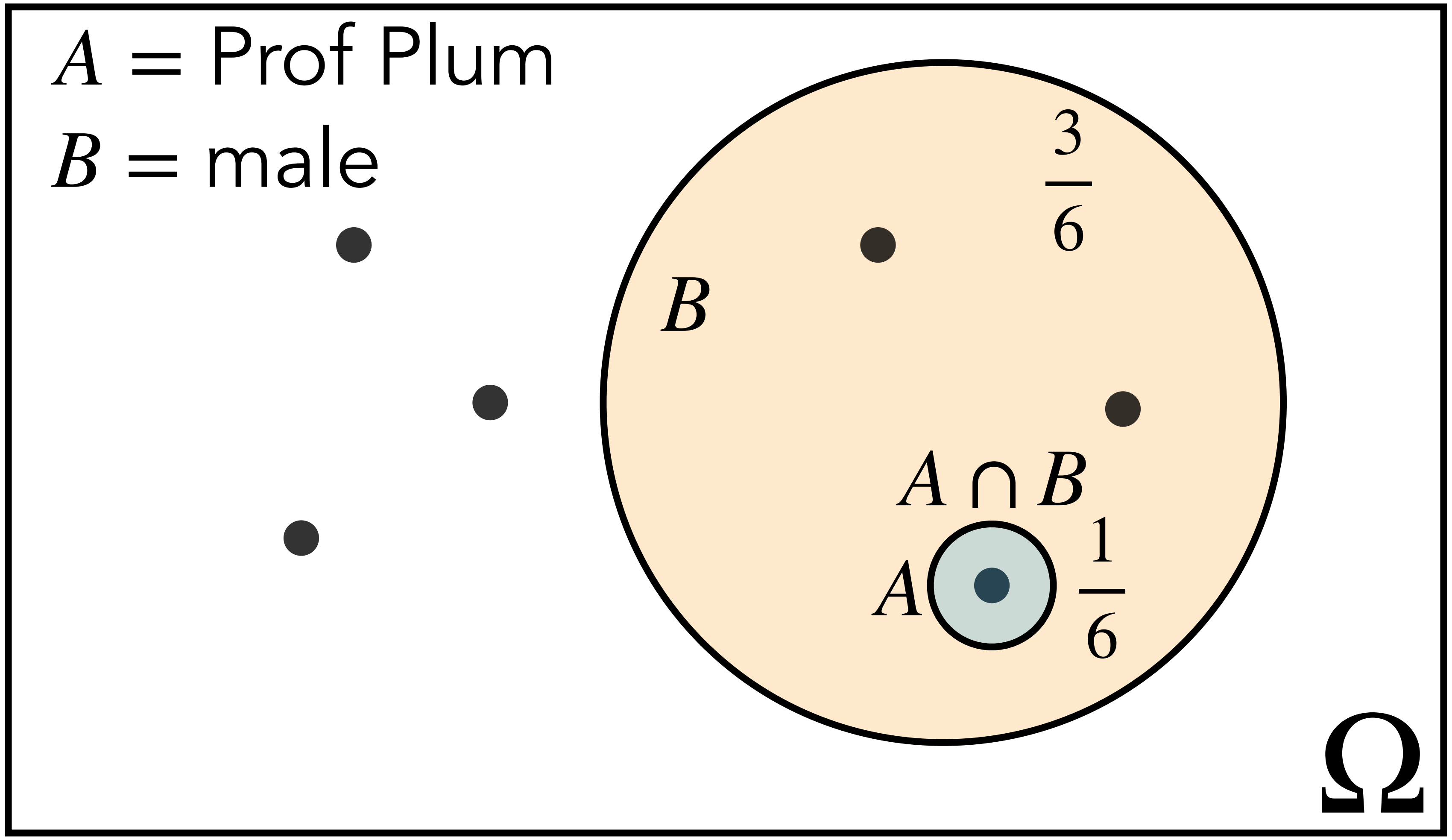


- *conditional probability*:
- $p(A \mid B)$ (probability of A given B)
- **Definition**: $p(A \mid B) = \dfrac{p(A, B)}{p(B)}$

Probability that it was Prof Plum, given that the murderer was male?

$$p(\text{Prof. Plum} \mid \text{male}) = ?$$

Clue guide to probability



Probability that it was Prof Plum, given that the murderer was male?

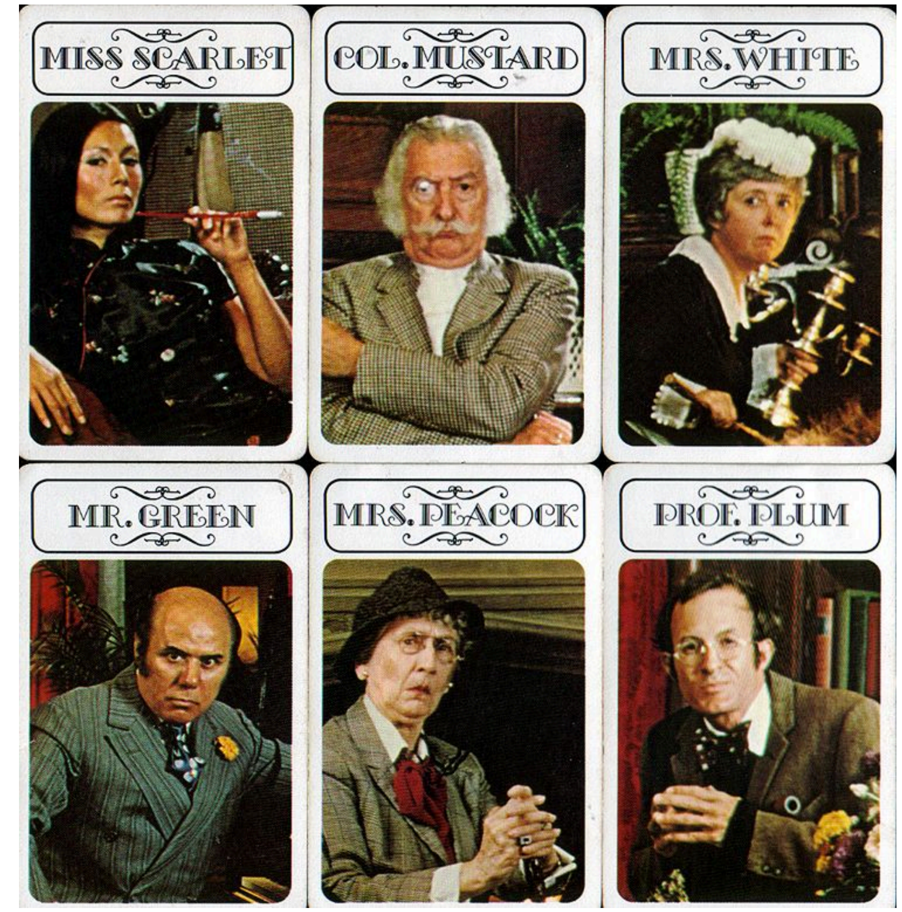
$$\text{Definition:} \quad p(A \mid B) = \frac{p(A,B)}{p(B)} = \frac{1}{3}$$

$$p(A) = \frac{1}{6} \qquad p(A,B) = \frac{1}{6} \qquad p(B) = \frac{3}{6}$$

Clue guide to probability

- *conditional probability*:
- $p(A \mid B)$ (probability of A given B)
- **Definition**: $p(A \mid B) = \dfrac{p(A, B)}{p(B)}$
- $p(\text{Prof. Plum} \mid \text{male}) = \dfrac{1/6}{1/2} = 1/3$

Who?

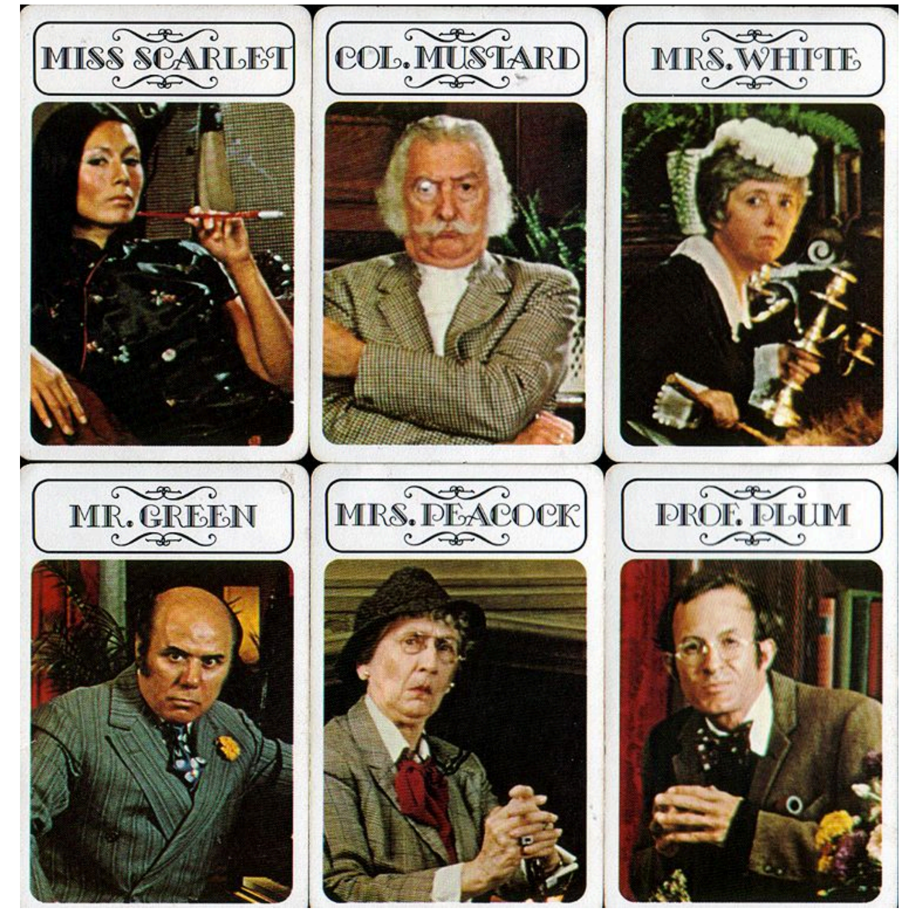


who	gender
col_mustard	male
mr_green	male
prof_plum	male
ms_scarlet	female
mrs_white	female
mrs_peacock	female

```
df.suspects = df.clue %>%
  distinct(who) %>%
  mutate(gender = ifelse(
    test = who %in% c("ms_scarlet",
                      "mrs_white",
                      "mrs_peacock"),
    yes = "female",
    no = "male"))
```

Clue guide to probability

Who?



- *conditional probability*:
- $p(A \mid B)$ (probability of A given B)
- **Definition**: $p(A \mid B) = \dfrac{p(A, B)}{p(B)}$
- $p(\text{Prof. Plum} \mid \text{male}) = \dfrac{1/6}{1/2} = 1/3$

who	gender
col_mustard	male
mr_green	male
prof_plum	male
ms_scarlet	female
mrs_white	female
mrs_peacock	female

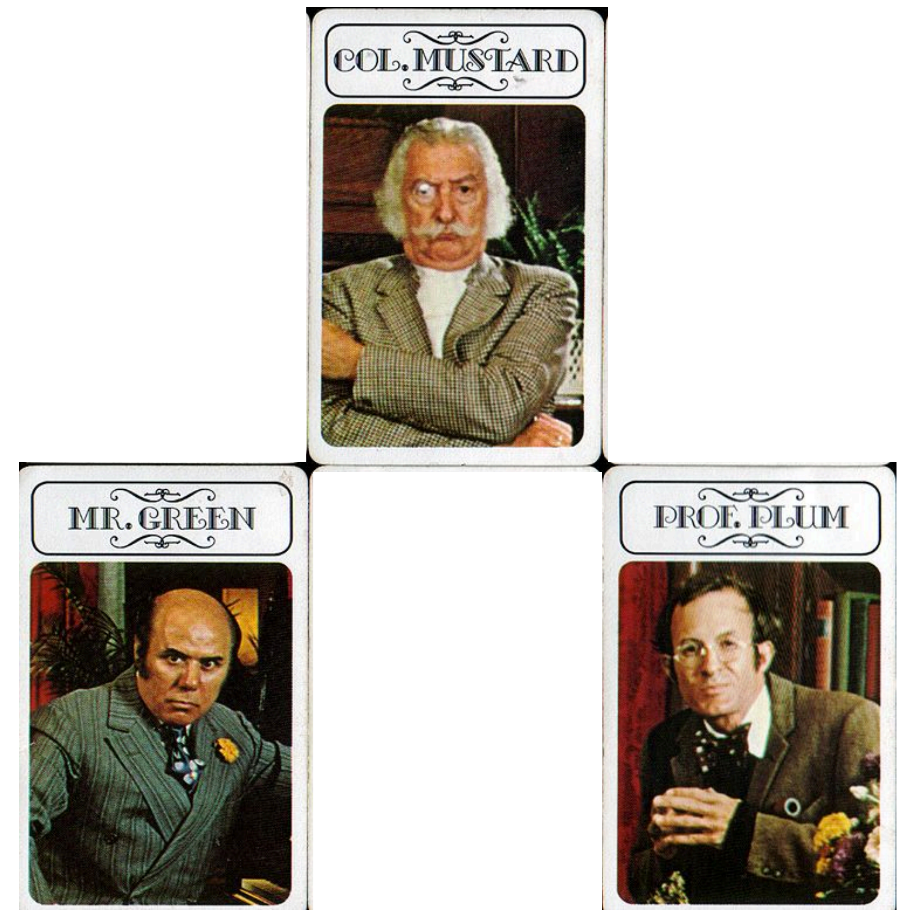
```
df.suspects %>%
  summarize(p_prof_plum_given_male =
              sum(gender == "male" &
                    who == "prof_plum") /
              sum(gender == "male"))
```

use naive definition of probability

Clue guide to probability

- *conditional probability*:
- $p(A \mid B)$ (probability of A given B)
- **Definition**: $p(A \mid B) = \dfrac{p(A, B)}{p(B)}$
- $p(\text{Prof. Plum} \mid \text{male}) = \dfrac{1/6}{1/2} = 1/3$

Who?



who	gender
col_mustard	male
mr_green	male
prof_plum	male

```
df.suspects %>%
  filter(gender == "male") %>%
  summarize(p_prof_plum_given_male =
              sum(who == "prof_plum") /
              n())
```

Clue guide to probability

- *independence*:
- A and B are independent if
- **Definition**: $p(A\,|\,B) = p(A)$
- (probability of A does not change if you know B)

- $p(\text{Prof Plum}\,|\,\text{candle stick}) = p(\text{Prof Plum})$
- each card (who and what) is drawn from a separate pack of cards

Who?

What?

Clue guide to probability

Who?

- *joint probability*:
- if A and B are independent then
- **Definition**: $p(A, B) = p(A) \cdot p(B)$
- $p(\text{Prof Plum, candle stick}) = p(\text{Prof Plum}) \cdot p(\text{candle stick}) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$

What?

Clue guide to probability

- *dependence*:
- **Definition**: $p(A \mid B) \neq p(A)$
- **Definition**: $p(A, B) = p(A) \cdot p(B \mid A)$

- if women were more likely than men to use the revolver then
- $p(\text{Mrs. White} \mid \text{Revolver}) > p(\text{Mrs. White})$

Who?

What?

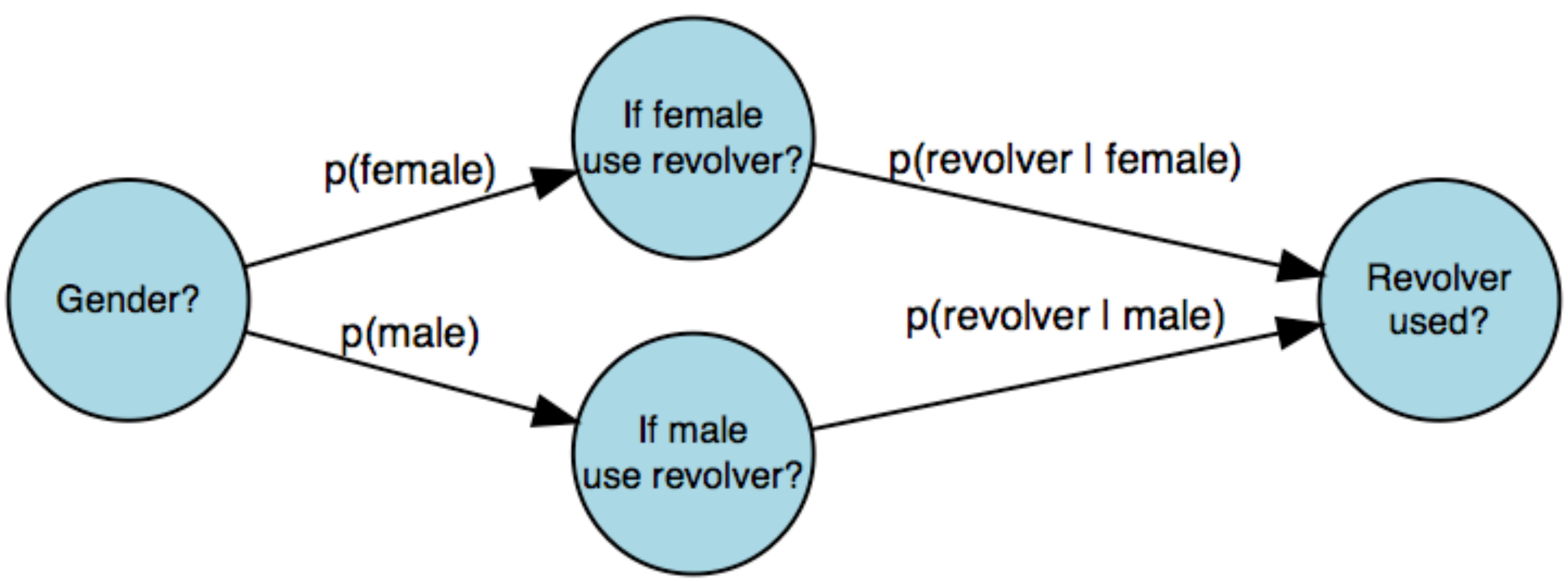
Clue guide to probability

- *law of total probability*
- **Definition**:

$$p(A) = p(A \mid B) \cdot p(B) + p(A \mid \neg B) \cdot p(\neg B)$$

$$p(A) = \sum_{i=1}^{n} p(A \mid B_i) \cdot p(B_i)$$

$$p(\text{what} = \text{Revolver}) = ?$$



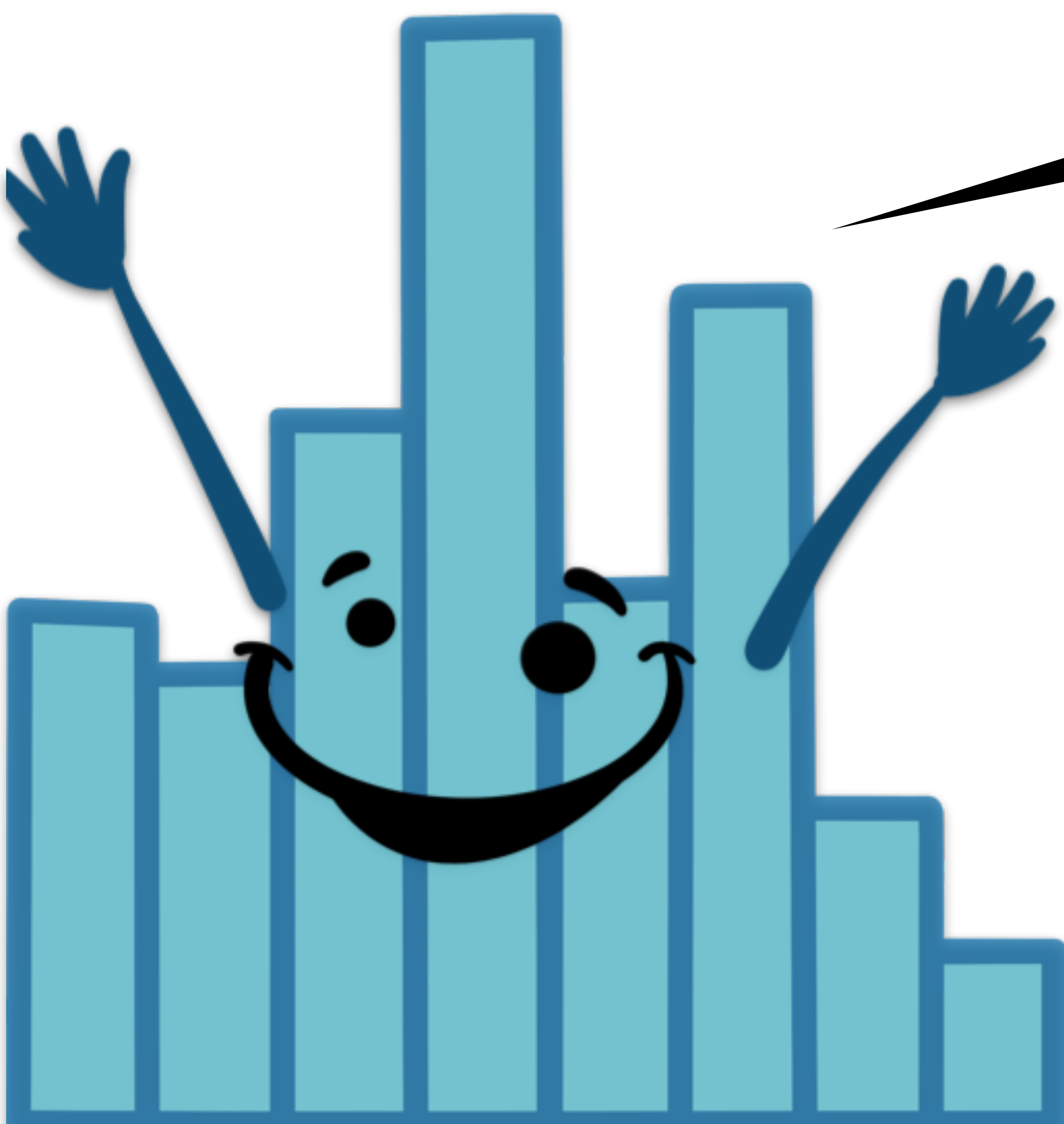
Who?

What?

We're listening to "The Bird"
by "Anderson .Paak"
submitted by Ari Beller

02:00

stretch break!



Understanding Bayes' rule

Clue guide to probability

Bayes Theorem in a few steps

Clue guide to probability

- *Bayes' theorem (derivation)*

$$p(B \mid A) = \frac{p(A,B)}{p(A)} \qquad p(A \mid B) = \frac{p(A,B)}{p(B)}$$

$$p(A,B) = \boxed{p(B \mid A) \cdot p(A) = p(A \mid B) \cdot p(B)}$$

$$p(B \mid A) = \frac{p(A \mid B) \cdot p(B)}{p(A)}$$

Clue guide to probability

$$p(B \mid A) = \frac{p(A \mid B) \cdot p(B)}{p(A)}$$

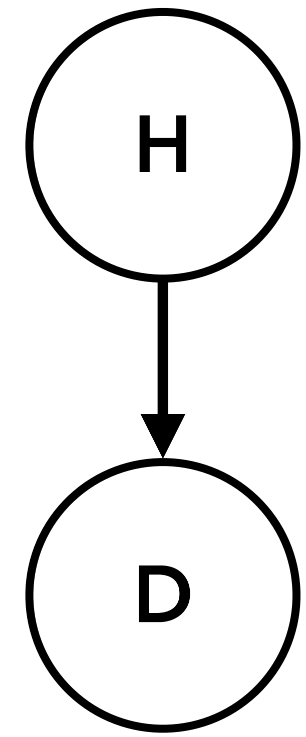
posterior $\quad$ likelihood $\quad$ prior

$$p(H \mid D) = \frac{\overbrace{p(D \mid H)}^{\text{likelihood}} \cdot \overbrace{p(H)}^{\text{prior}}}{p(D)}$$

subjective probability interpretation

H = Hypothesis

D = Data



formal framework for learning from data

updating our prior belief $p(H)$, to a posterior belief $p(H \mid D)$ given some data

Clue guide to probability

posterior $\quad$ likelihood $\quad$ prior

$$p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{p(D)}$$

probability of the data?!

$H =$ Hypothesis

$D =$ Data

law of total probability:

$$p(D) = \sum_{i=1}^{n} p(D \mid H_i) \cdot p(H_i)$$

take into account all the different ways in which the data could have come about

Clue guide to probability

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is "95% accurate"; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that $\mathbf{P(T|D) = 0.95}$ and $\mathbf{P(\neg T|\neg D) = 0.95}$. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

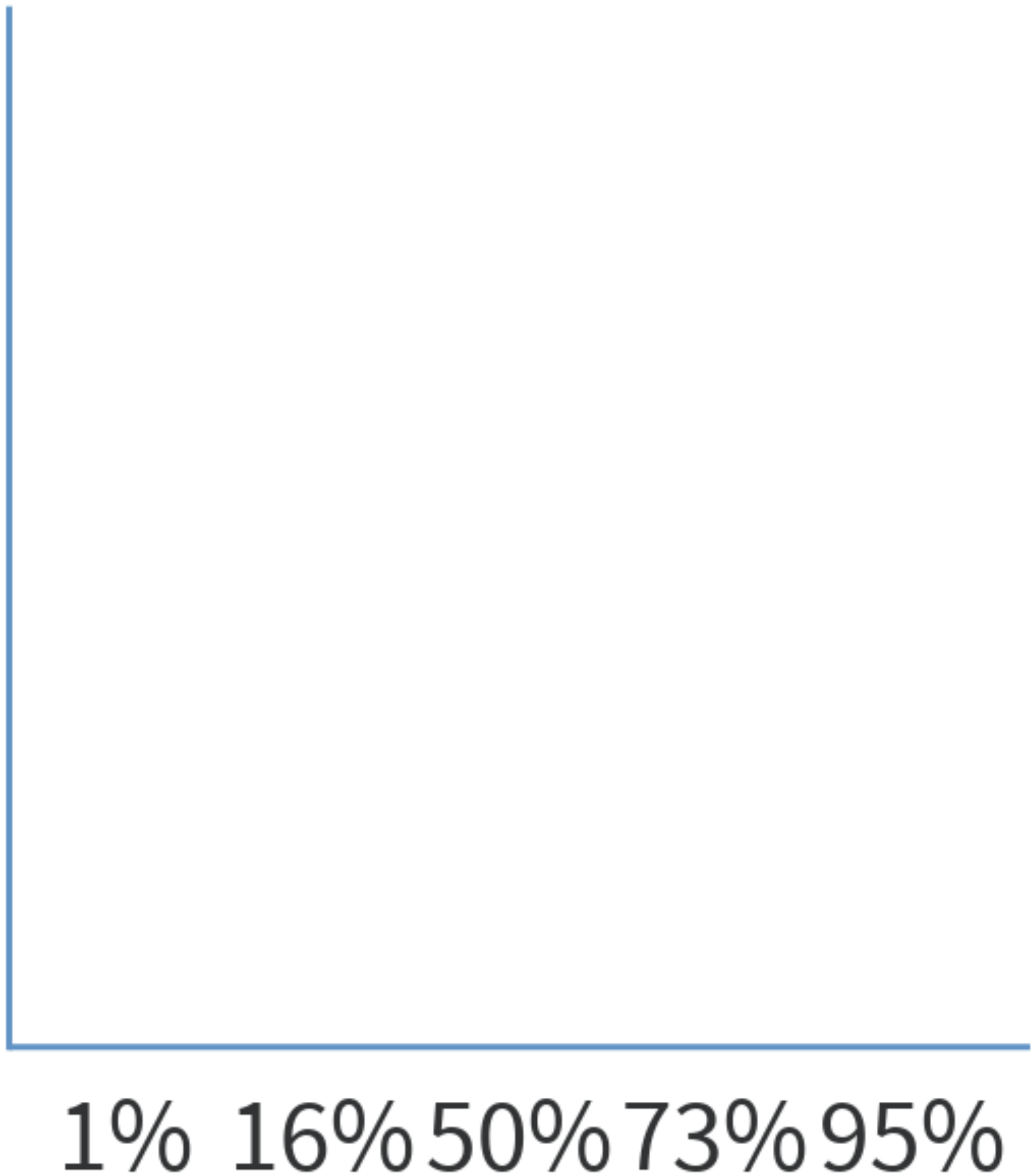
Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

What's the probability that Fred has conditionitis?

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is "95% accurate"; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that $\mathbf{P(T|D) = 0.95}$ and $\mathbf{P(\neg T|\neg D) = 0.95}$. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.



Breakout rooms

Tasks: A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is “95% accurate”; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that **$P(T|D) = 0.95$** and **$P(\neg T|\neg D) = 0.95$**. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

Size: ~3 people

Time: 5 minutes

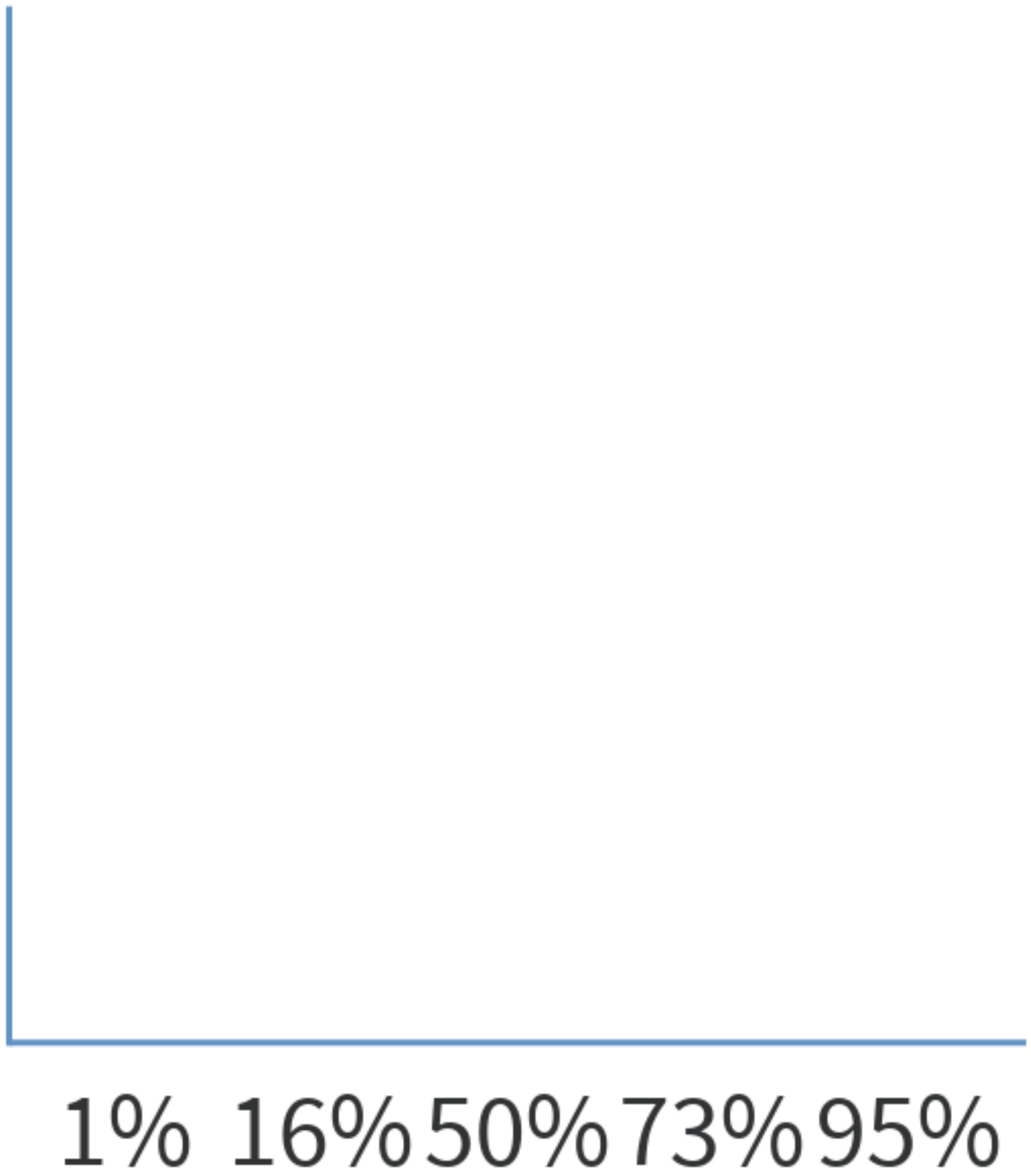
Report: We will vote again.

What's the probability that Fred has conditionitis?

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is "95% accurate"; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that $\mathbf{P(T|D) = 0.95}$ and $\mathbf{P(\neg T|\neg D) = 0.95}$. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.



Clue guide to probability

what we know

$$P(D) = 0.01$$

$$P(T \mid D) = 0.95$$

$$P(T \mid \neg D) = 0.05$$

what we want to know

$$P(D \mid T) = ?$$

$$p(D \mid T) = \frac{p(T \mid D) \cdot p(D)}{p(T)}$$ Bayes' rule

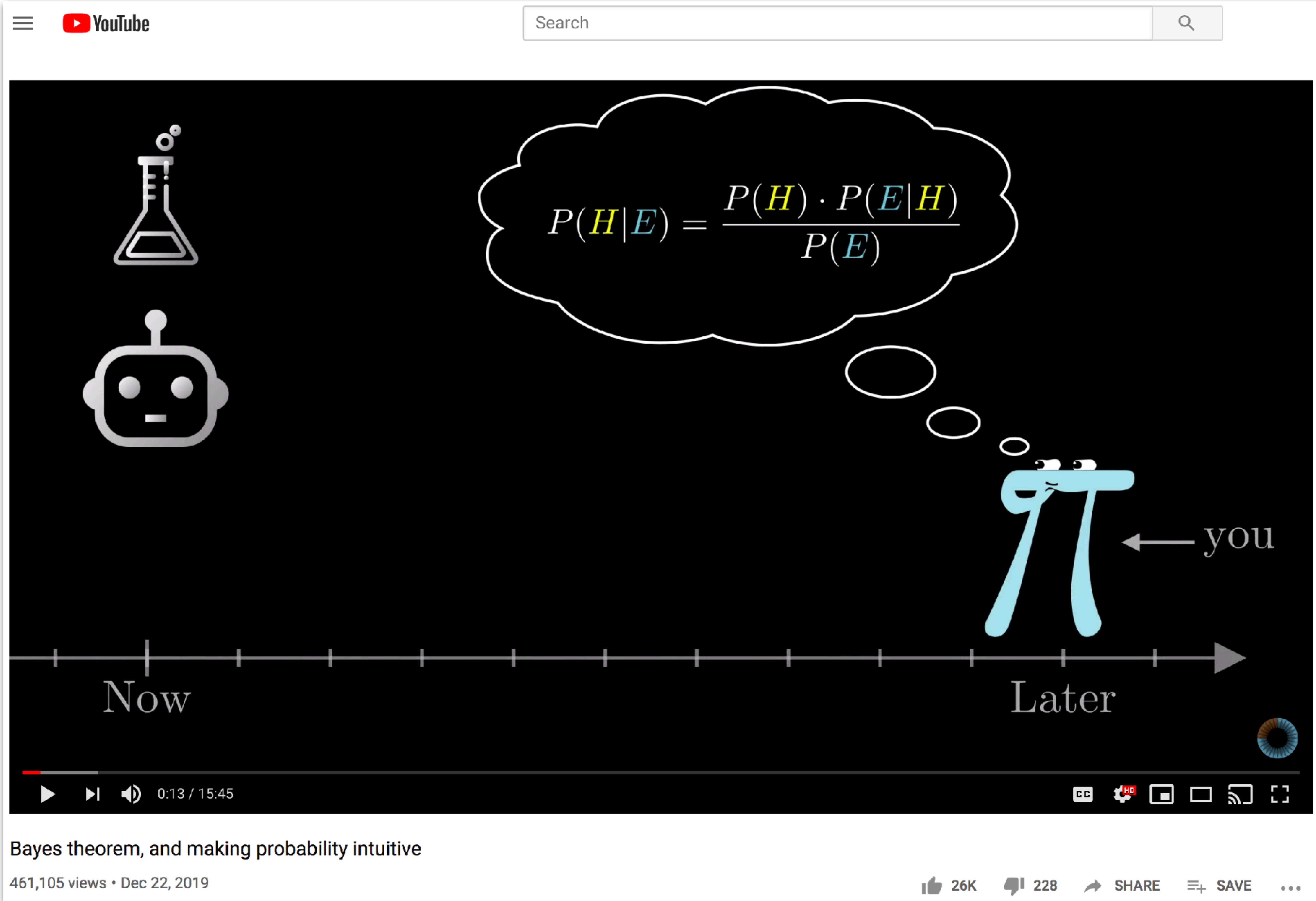
$$p(D \mid T) = \frac{p(T \mid D) \cdot p(D)}{p(T \mid D) \cdot p(D) + p(T \mid \neg D) \cdot p(\neg D)}$$

law of total probability

$$\boxed{?}$$

Clue guide to probability

?



https://www.youtube.com/watch?v=HZGCoVF3YvM&feature=youtu.be

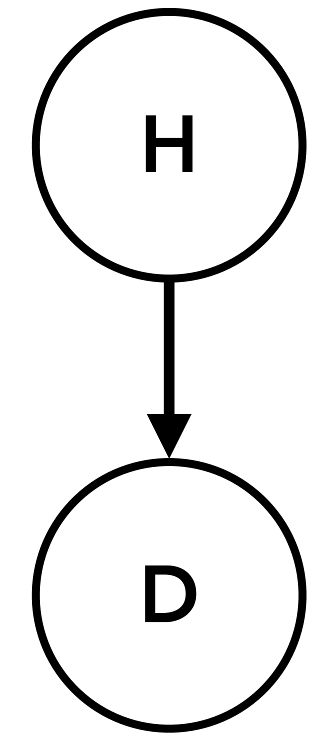
Bayes' theorem in three panels

$$\overset{\text{posterior}}{p(H \mid D)} = \frac{\overset{\text{likelihood}}{p(D \mid H)} \cdot \overset{\text{prior}}{p(H)}}{p(D)}$$

subjective probability interpretation

H = Hypothesis

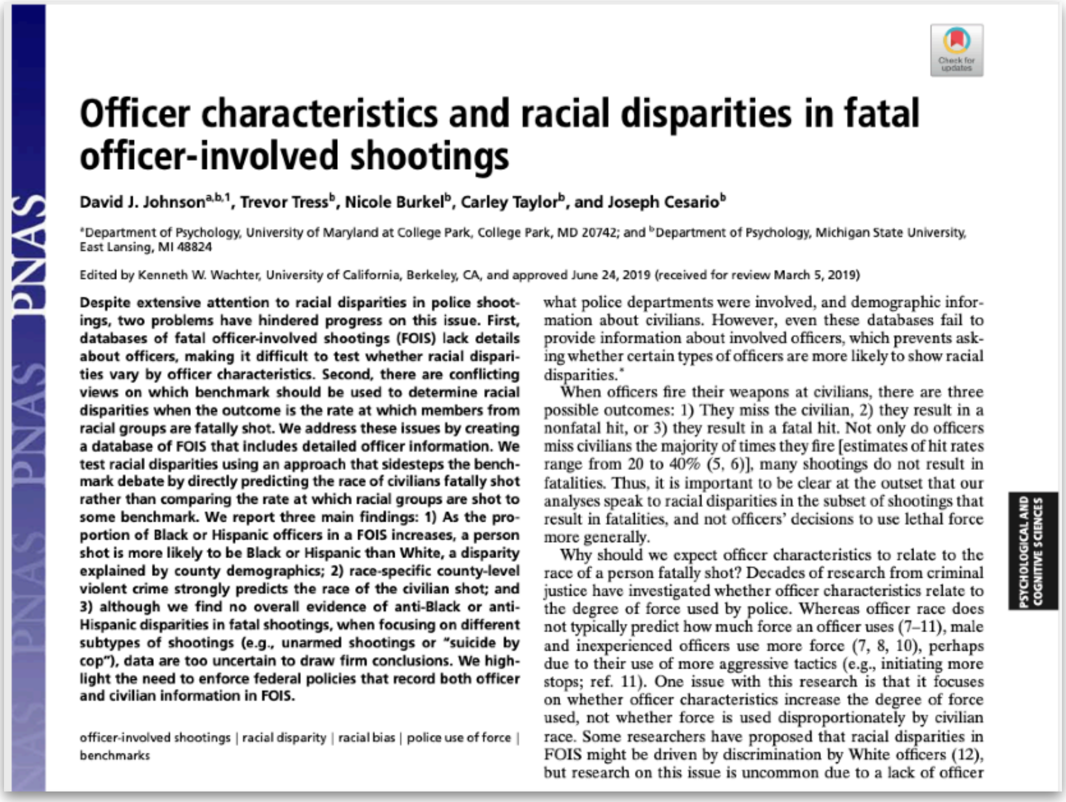
D = Data



https://www.tjmahr.com/bayes-theorem-in-three-panels/

Getting Bayes' right matters

Getting Bayes right matters!



PNAS

Check for updates

Officer characteristics and racial disparities in fatal officer-involved shootings

David J. Johnson[a,b,1], Trevor Tress[b], Nicole Burkel[b], Carley Taylor[b], and Joseph Cesario[b]

[a]Department of Psychology, University of Maryland at College Park, College Park, MD 20742; and [b]Department of Psychology, Michigan State University, East Lansing, MI 48824

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved June 24, 2019 (received for review March 5, 2019)

Despite extensive attention to racial disparities in police shootings, two problems have hindered progress on this issue. First, databases of fatal officer-involved shootings (FOIS) lack details about officers, making it difficult to test whether racial disparities vary by officer characteristics. Second, there are conflicting views on which benchmark should be used to determine racial disparities when the outcome is the rate at which members from racial groups are fatally shot. We address these issues by creating a database of FOIS that includes detailed officer information. We test racial disparities using an approach that sidesteps the benchmark debate by directly predicting the race of civilians fatally shot rather than comparing the rate at which racial groups are shot to some benchmark. We report three main findings: 1) As the proportion of Black or Hispanic officers in a FOIS increases, a person shot is more likely to be Black or Hispanic than White, a disparity explained by county demographics; 2) race-specific county-level violent crime strongly predicts the race of the civilian shot; and 3) although we find no overall evidence of anti-Black or anti-Hispanic disparities in fatal shootings, when focusing on different subtypes of shootings (e.g., unarmed shootings or "suicide by cop"), data are too uncertain to draw firm conclusions. We highlight the need to enforce federal policies that record both officer and civilian information in FOIS.

officer-involved shootings | racial disparity | racial bias | police use of force | benchmarks

what police departments were involved, and demographic information about civilians. However, even these databases fail to provide information about involved officers, which prevents asking whether certain types of officers are more likely to show racial disparities.*

When officers fire their weapons at civilians, there are three possible outcomes: 1) They miss the civilian, 2) they result in a nonfatal hit, or 3) they result in a fatal hit. Not only do officers miss civilians the majority of times they fire [estimates of hit rates range from 20 to 40% (5, 6)], many shootings do not result in fatalities. Thus, it is important to be clear at the outset that our analyses speak to racial disparities in the subset of shootings that result in fatalities, and not officers' decisions to use lethal force more generally.

Why should we expect officer characteristics to relate to the race of a person fatally shot? Decades of research from criminal justice have investigated whether officer characteristics relate to the degree of force used by police. Whereas officer race does not typically predict how much force an officer uses (7–11), male and inexperienced officers use more force (7, 8, 10), perhaps due to their use of more aggressive tactics (e.g., initiating more stops; ref. 11). One issue with this research is that it focuses on whether officer characteristics increase the degree of force used, not whether force is used disproportionately by civilian race. Some researchers have proposed that racial disparities in FOIS might be driven by discrimination by White officers (12), but research on this issue is uncommon due to a lack of officer

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Claim:

"White officers are not more likely to shoot minority civilians than non-White officers"

$$\Pr(\text{shot}|\text{minority civilian}, \text{white officer}, X) - \Pr(\text{shot}|\text{minority civilian}, \text{minority officer}, X) \leq 0,$$

[1]

What the statistic says:

"whether a person fatally shot was more likely to be Black (or Hispanic) than White"

Original claim:

Requires Bayes' rule

$$\begin{aligned}
&\Pr(\text{shot}|\text{minority civilian}, \text{white officer}, X) \\
&\quad - \Pr(\text{shot}|\text{minority civilian}, \text{minority officer}, X) \\
&= \frac{\Pr(\text{min. civ.}\,|\text{shot}, \text{white off.}, X) \times \Pr(\text{shot}|\text{white off.}, X)}{\Pr(\text{minority civilian}|\text{white officer}, X)} \\
&\quad - \frac{\Pr(\text{min. civ.}\,|\text{shot}, \text{min. off.}, X) \times \Pr(\text{shot}|\text{min. off.}, X)}{\Pr(\text{minority civilian}|\text{minority officer}, X)}.
\end{aligned}$$

[2]

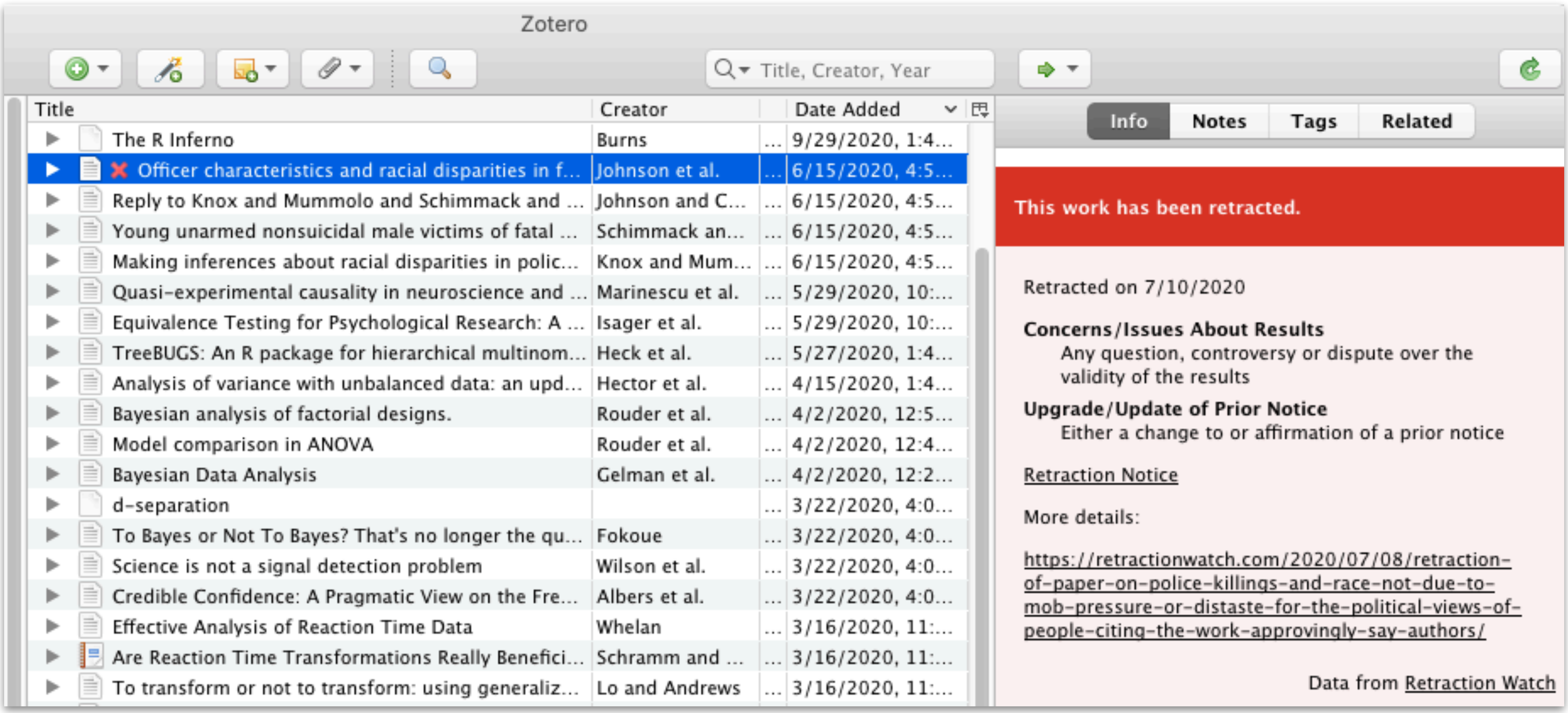
authors didn't have the relevant data to support their claim!

paper was retracted

Johnson, D. J., Tress, T., Burkel, N., Taylor, C., & Cesario, J. (2019). Officer characteristics and racial disparities in fatal officer-involved shootings. *Proceedings of the National Academy of Sciences*, 116(32), 15877–15882.

Knox, D., & Mummolo, J. (2020). Making inferences about racial disparities in police violence. *Proceedings of the National Academy of Sciences*, 117(3), 1261–1262.

Getting Bayes right matters!



Tip: Use Zotero as a reference manager!

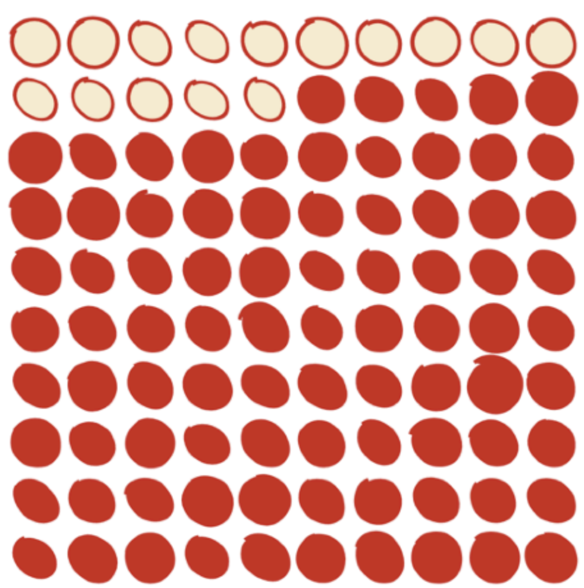
Getting Bayes right matters!

The New York Times

TheUpshot

When They Warn of Rare Disorders, These Prenatal Tests Are Usually Wrong

Some of the tests look for missing snippets of chromosomes. For every 15 times they correctly find a problem ...



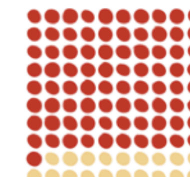
... they are wrong 85 times

For These Five Tests, Positive Results Are Often Wrong

As prenatal tests have expanded to more rare conditions, a larger share of their positive results are incorrect. Some of the worst-performing tests look for microdeletions, which are small missing snippets of chromosomes.

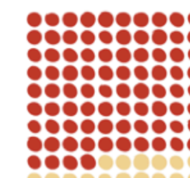
Chance positive results are wrong

DiGeorge syndrome
Affects 1 in 4,000 births
Can cause heart defects and delayed language acquisition. (May appear on lab reports as "22q.")



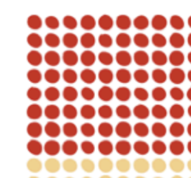
81% wrong

1p36 deletion
1 in 5,000 births
Can cause seizures, low muscle tone and intellectual disability.



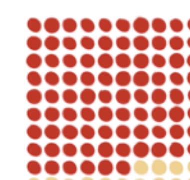
84% wrong

Cri-du-chat syndrome
1 in 15,000 births
Can cause difficulty walking and delayed speech development.



80% wrong

Wolf-Hirschhorn syndrome
1 in 20,000 births
Can cause seizures, growth delays and intellectual disability.



86% wrong

Prader-Willi and Angelman syndromes
1 in 20,000 births
Can cause seizures and an inability to control food consumption.

93% wrong

https://www.nytimes.com/2022/01/01/upshot/pregnancy-birth-genetic-testing.html

Getting Bayes right matters!

sensitivity: $p(T \mid D) = 0.999$

specificity: $p(\neg T \mid \neg D) = 0.999$

prior: $p(D) = 0.0001$

T = positive test result

$\neg T$ = negative test result

D = disease

$\neg D$ = no disease

data: T (positive test result)

81% wrong

posterior: $$p(D \mid T) = \frac{p(T \mid D) \cdot p(D)}{p(T \mid D) \cdot p(D) + p(T \mid \neg D) \cdot p(\neg D)}$$

$$= \frac{0.999 \cdot 0.0001}{0.999 \cdot 0.0001 + 0.001 \cdot 0.9999} \approx 0.09$$

https://www.nytimes.com/2022/01/01/upshot/pregnancy-birth-genetic-testing.html

Getting Bayes right matters!

Most people who are in the hospital being treated for Covid are vaccinated.

Getting Bayes right matters!

likelihood: $p(H \mid V) = 0.2$

$p(H \mid \neg V) = 0.5$

prior: $p(V) = 0.8$

$H =$ hospitalized

$\neg H =$ not hospitalized

$V =$ vaccinated

$\neg V =$ no vaccinated

data: H (the person is in the hospital)

posterior: $$p(V \mid H) = \frac{p(H \mid V) \cdot p(V)}{p(H \mid V) \cdot p(V) + p(H \mid \neg V) \cdot p(\neg V)}$$

$$= \frac{0.2 \cdot 0.8}{0.2 \cdot 0.8 + 0.5 \cdot 0.2} \approx 0.62$$

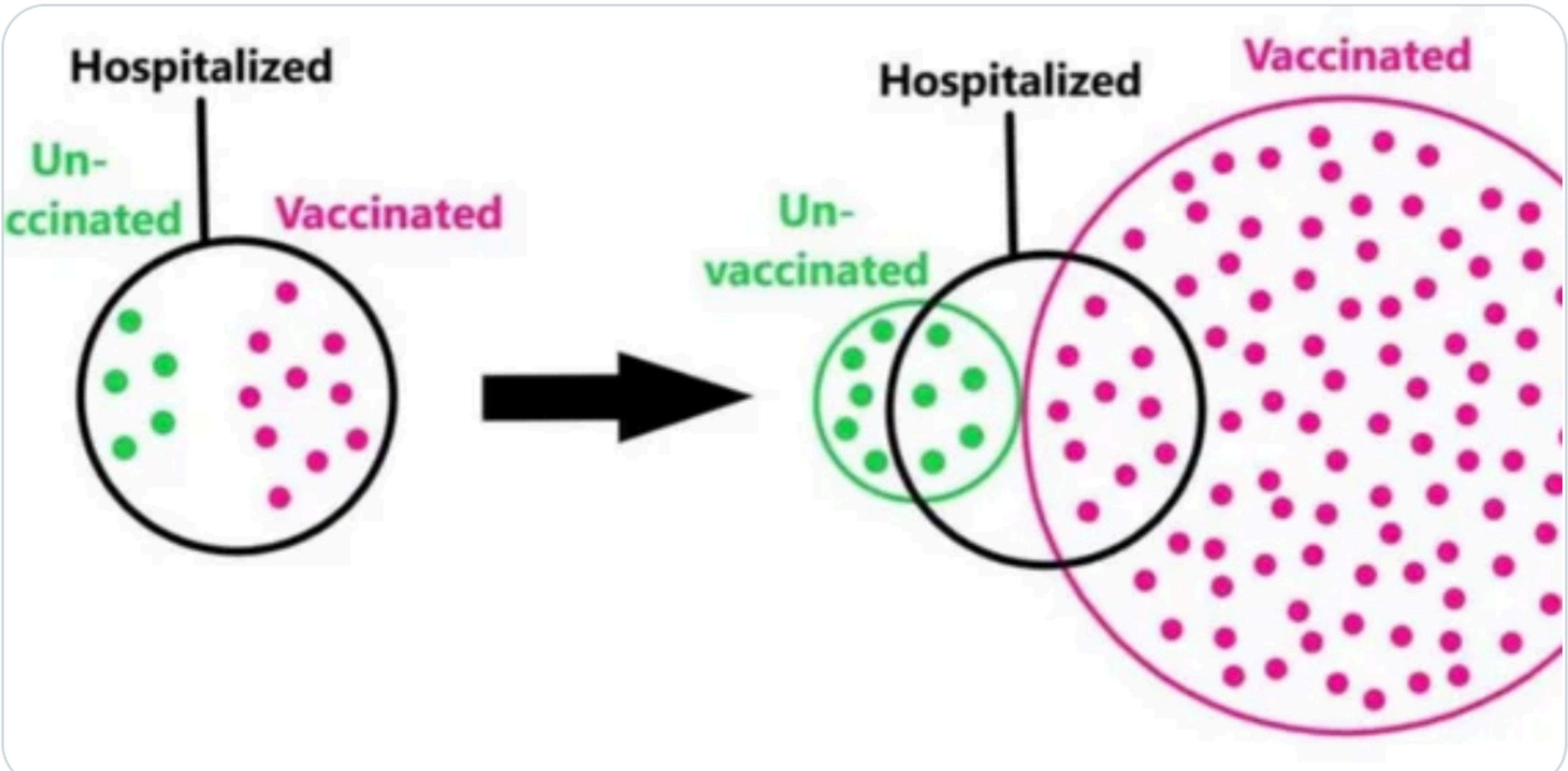
62% of the hospitalized people are vaccinated

https://www.nytimes.com/2022/01/01/upshot/pregnancy-birth-genetic-testing.html

Bayes' rule matters

Nick Brown
@sTeamTraen

Stolen from Reddit. May be of some use.



1:22 PM · Nov 20, 2021 · Twitter Web App

565 Retweets **54** Quote Tweets **2,752** Likes

Building a Bayesis

Rolling the dice

Four sided

Six sided

both dice are equally likely to be picked

p() = p() = 0.5

both dice are equal sided (uniform probability over the different numbers)

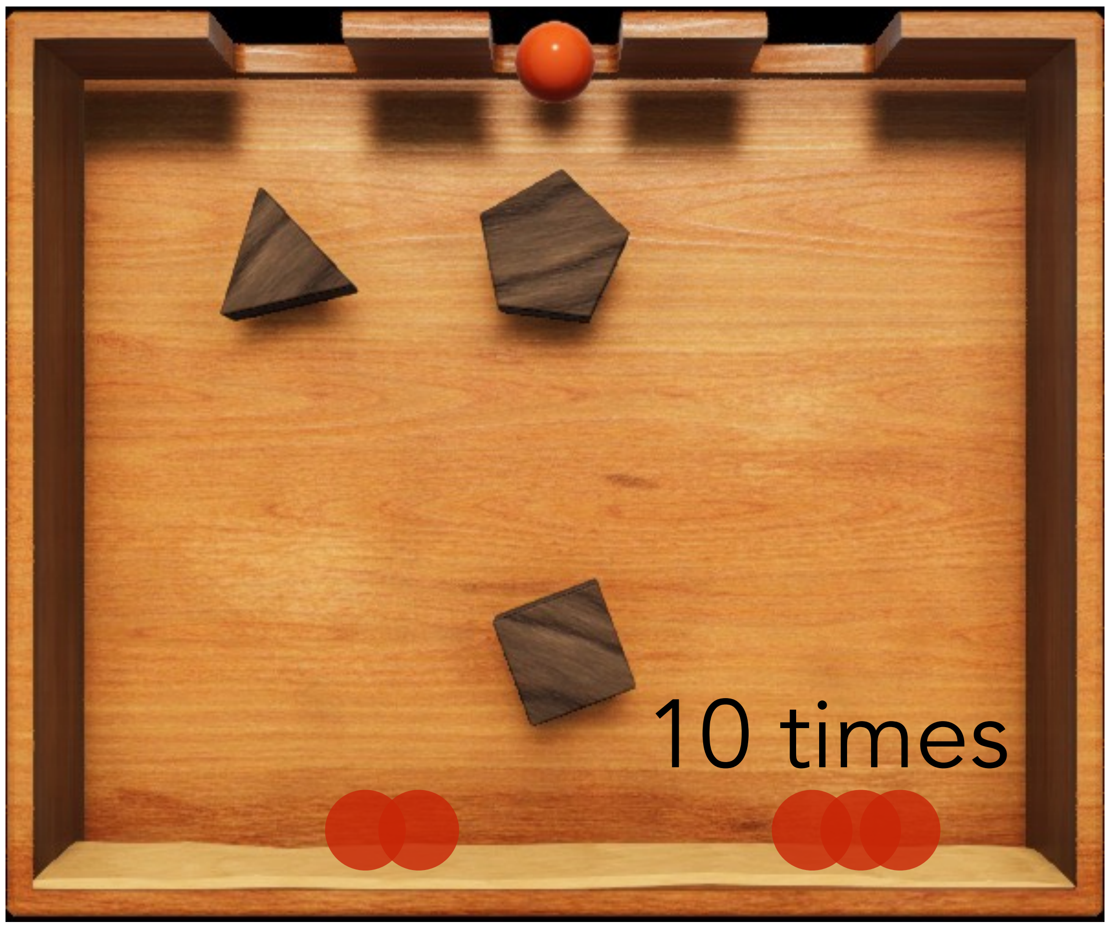
Physical reasoning

Gerstenberg, T., Siegel, M. H., & Tenenbaum, J. B. (2021). What happened? Reconstructing the past from vision and sound. PsyArXiv. https://psyarxiv.com/tfjdk

Physical reasoning

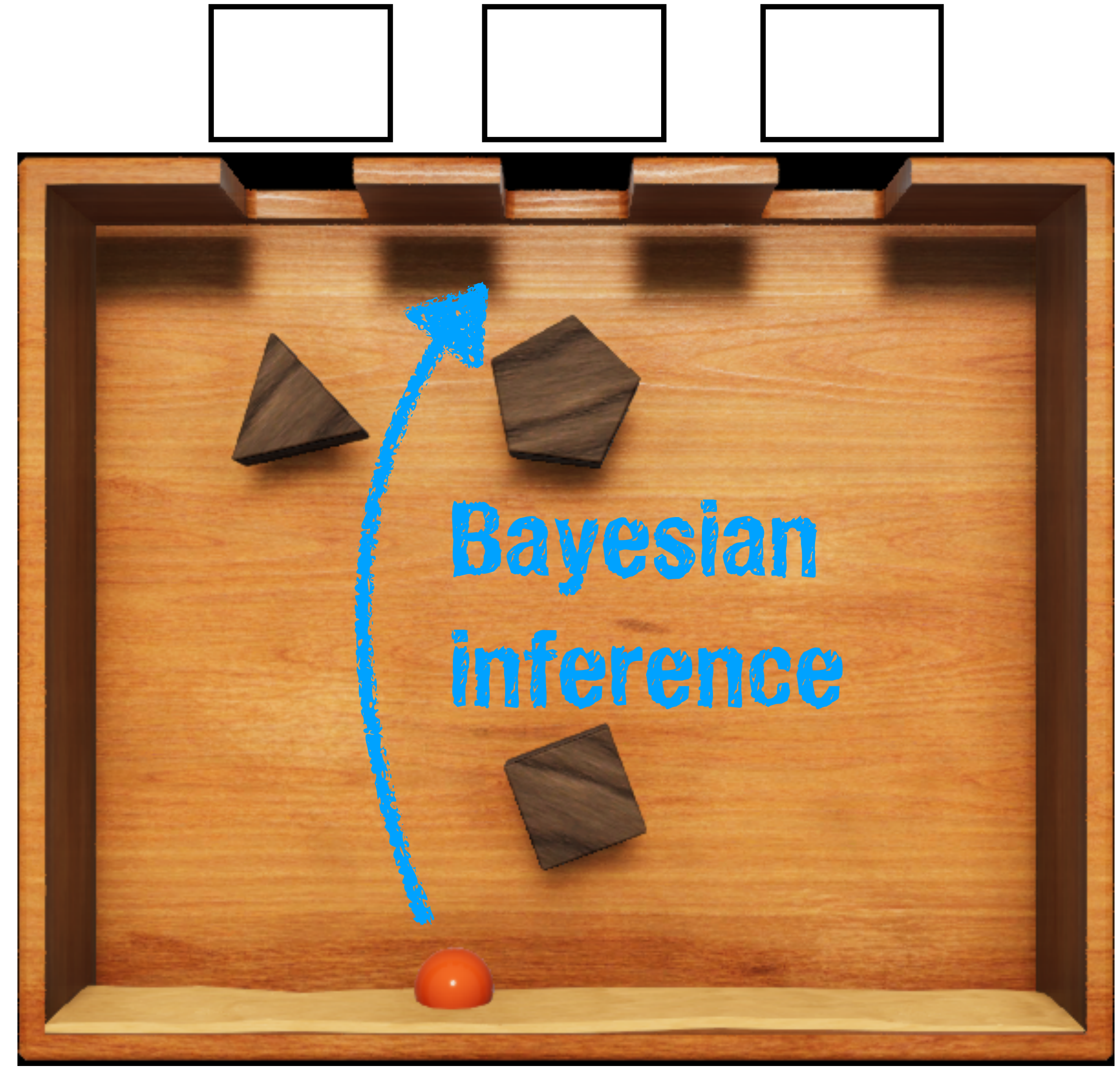
Gerstenberg, T., Siegel, M. H., & Tenenbaum, J. B. (2021). What happened? Reconstructing the past from vision and sound. PsyArXiv. https://psyarxiv.com/tfjdk

Prediction



Where will the ball land?

Inference



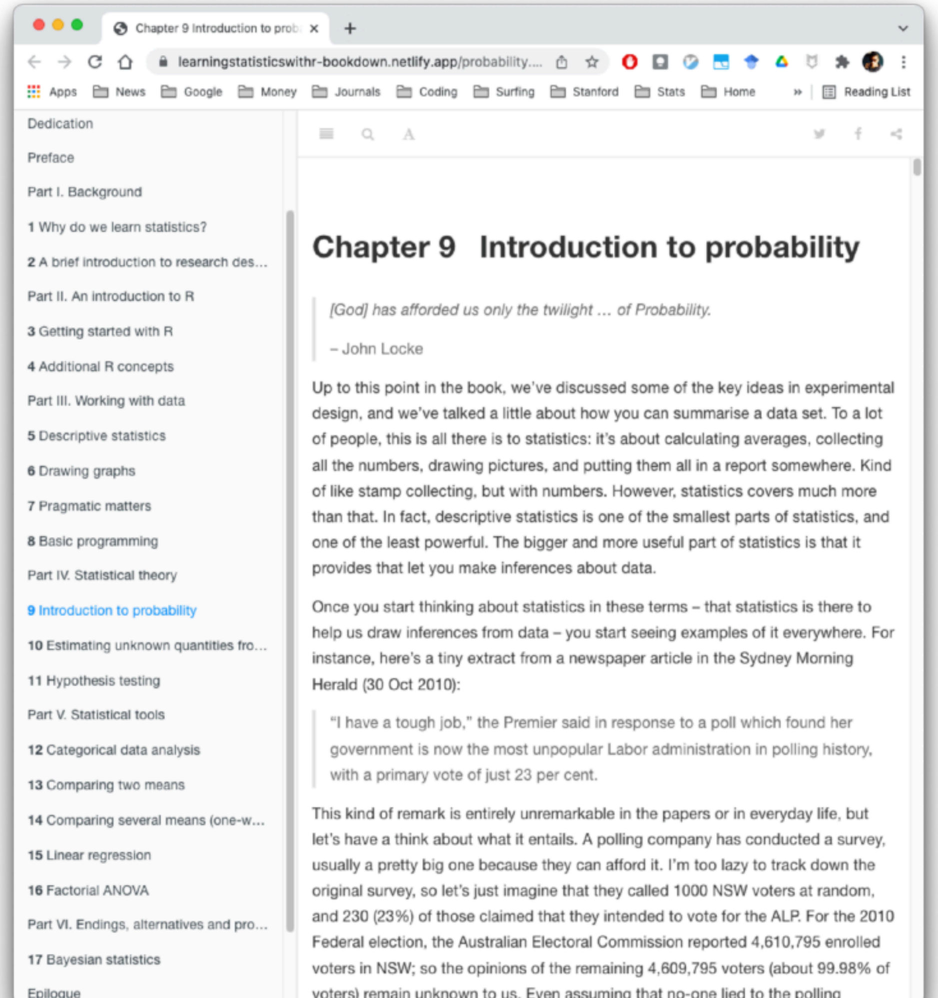
In which hole was the ball dropped?

Gerstenberg, T., Siegel, M. H., & Tenenbaum, J. B. (2021). What happened? Reconstructing the past from vision and sound. PsyArXiv. https://psyarxiv.com/tfjdk

Outline

- Introduction to probability / Recap
  - Motivation
  - Counting possibilities
  - Clue guide to probability
  - Understanding Bayes' Rule
  - Getting Bayes' right matters!
  - Building a Bayesis

I want more!



Chapter 9 Introduction to probability

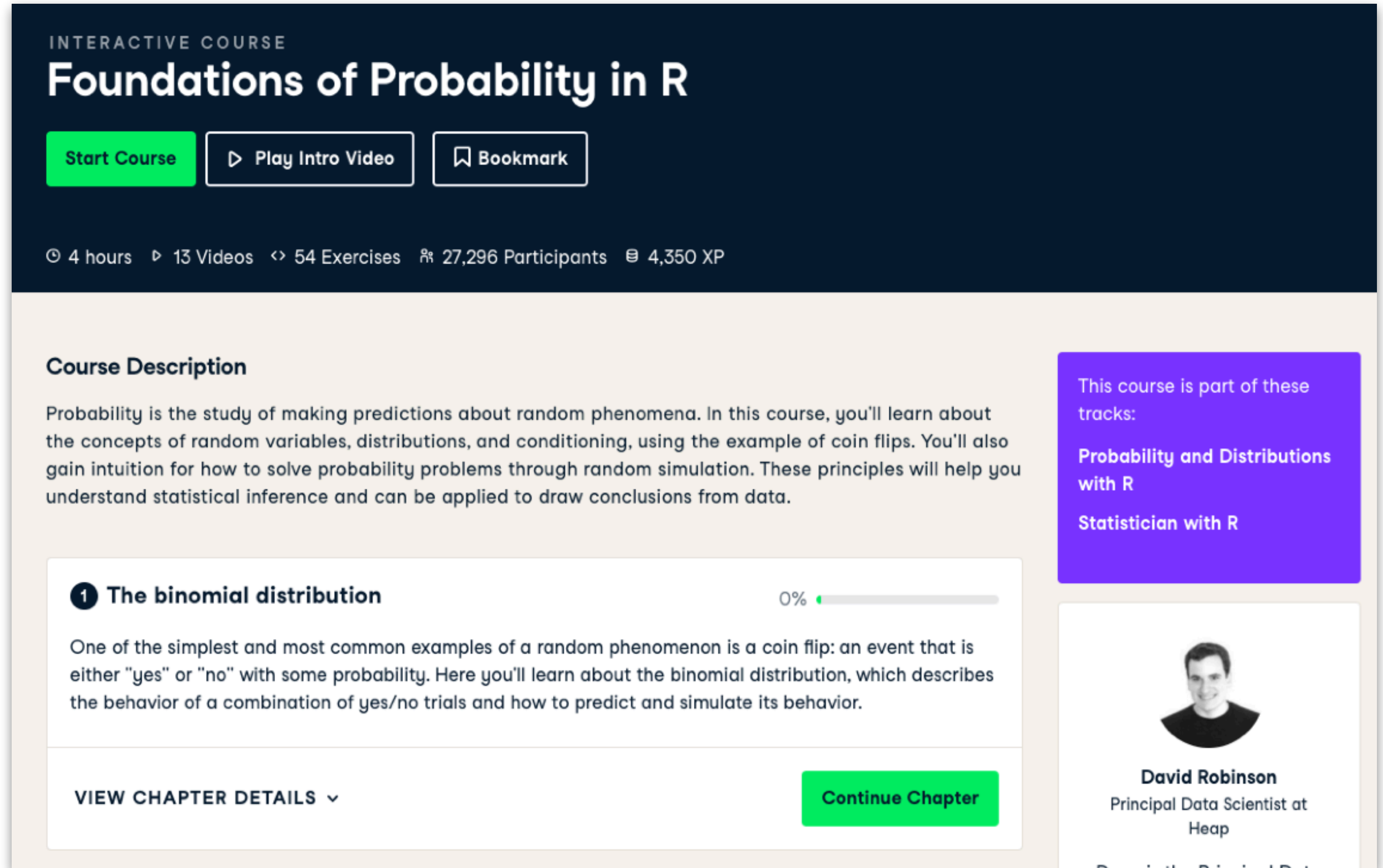
> *[God] has afforded us only the twilight … of Probability.*
>
> – John Locke

Up to this point in the book, we've discussed some of the key ideas in experimental design, and we've talked a little about how you can summarise a data set. To a lot of people, this is all there is to statistics: it's about calculating averages, collecting all the numbers, drawing pictures, and putting them all in a report somewhere. Kind of like stamp collecting, but with numbers. However, statistics covers much more than that. In fact, descriptive statistics is one of the smallest parts of statistics, and one of the least powerful. The bigger and more useful part of statistics is that it provides that let you make inferences about data.

Once you start thinking about statistics in these terms – that statistics is there to help us draw inferences from data – you start seeing examples of it everywhere. For instance, here's a tiny extract from a newspaper article in the Sydney Morning Herald (30 Oct 2010):

> "I have a tough job," the Premier said in response to a poll which found her government is now the most unpopular Labor administration in polling history, with a primary vote of just 23 per cent.

This kind of remark is entirely unremarkable in the papers or in everyday life, but let's have a think about what it entails. A polling company has conducted a survey, usually a pretty big one because they can afford it. I'm too lazy to track down the original survey, so let's just imagine that they called 1000 NSW voters at random, and 230 (23%) of those claimed that they intended to vote for the ALP. For the 2010 Federal election, the Australian Electoral Commission reported 4,610,795 enrolled voters in NSW; so the opinions of the remaining 4,609,795 voters (about 99.98% of voters) remain unknown to us. Even assuming that no-one lied to the polling



INTERACTIVE COURSE

Foundations of Probability in R

Start Course | Play Intro Video | Bookmark

4 hours · 13 Videos · 54 Exercises · 27,296 Participants · 4,350 XP

Course Description

Probability is the study of making predictions about random phenomena. In this course, you'll learn about the concepts of random variables, distributions, and conditioning, using the example of coin flips. You'll also gain intuition for how to solve probability problems through random simulation. These principles will help you understand statistical inference and can be applied to draw conclusions from data.

This course is part of these tracks:

Probability and Distributions with R

Statistician with R

1 The binomial distribution 0%

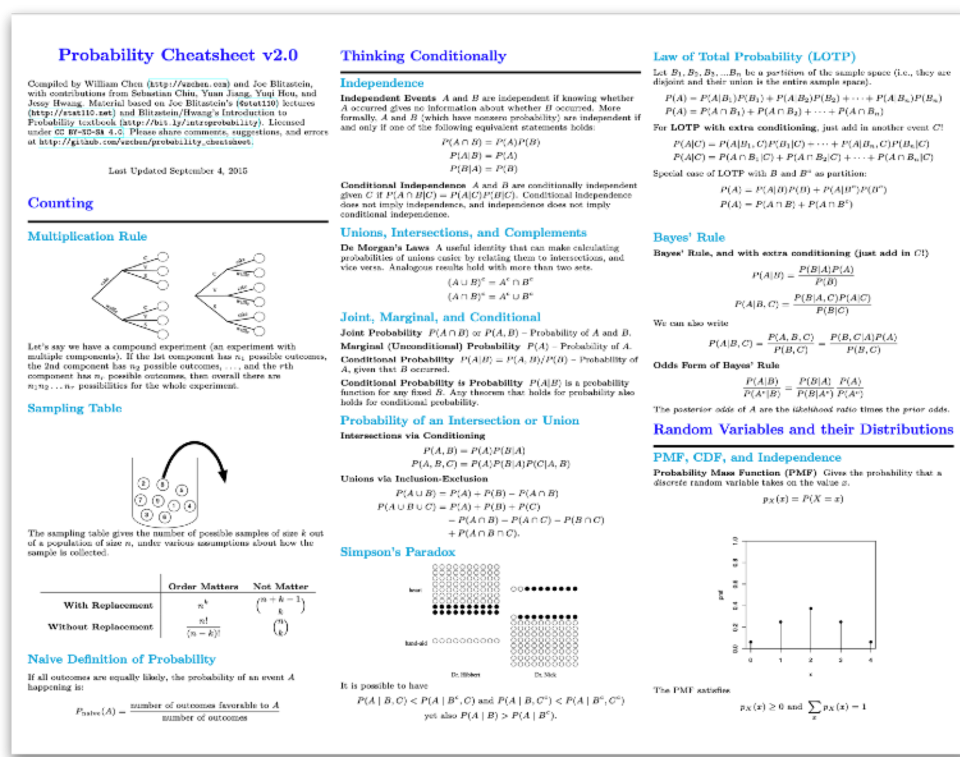
One of the simplest and most common examples of a random phenomenon is a coin flip: an event that is either "yes" or "no" with some probability. Here you'll learn about the binomial distribution, which describes the behavior of a combination of yes/no trials and how to predict and simulate its behavior.

VIEW CHAPTER DETAILS | Continue Chapter

David Robinson

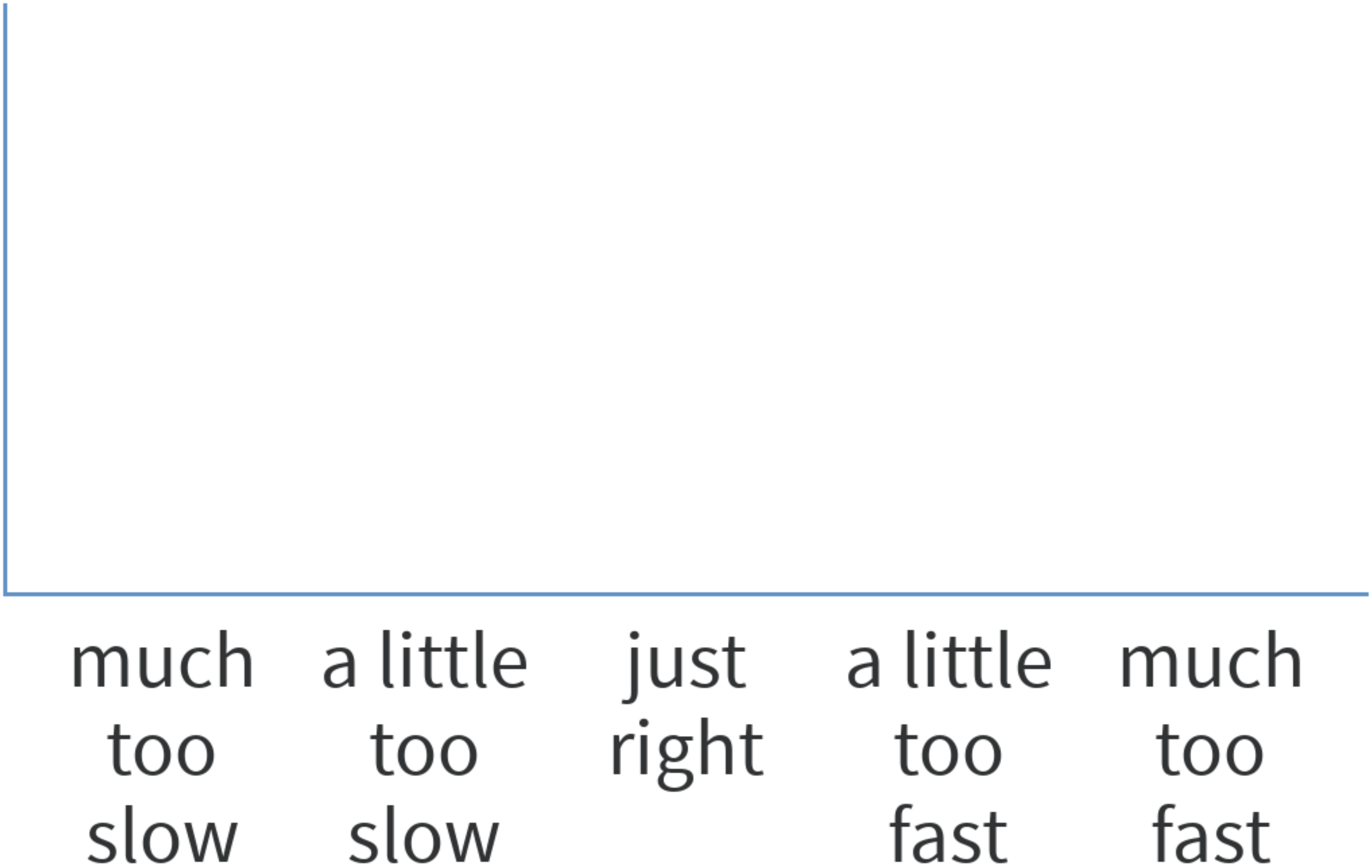
Principal Data Scientist at Heap

in the `figures/` folder for these materials on canvas



Feedback

How was the pace of today's class?



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!

Have a nice Martin Luther King, Jr. day!

The time is always right to do what is right.

- Martin Luther King, Jr.