Experimental Design and Scientific Rigor

An introduction to good science

Ernesto Salcedo, PhD

Experimental Design and Scientific Rigor

talk overview

- Terminology
  - Scientific Method
- Experimental Design
- Science in Crisis?
- Rigor and Reproducibility
- Paper Assignment

Fun with Science Terminology

Science

Fun with Terminology

- From Latin “scientia” for “Knowing, expertise”
- A Systemic and Logical Approach to Discovery
    - to understand how things work
- Also, the Body of Knowledge based on demonstrable and reproducible data

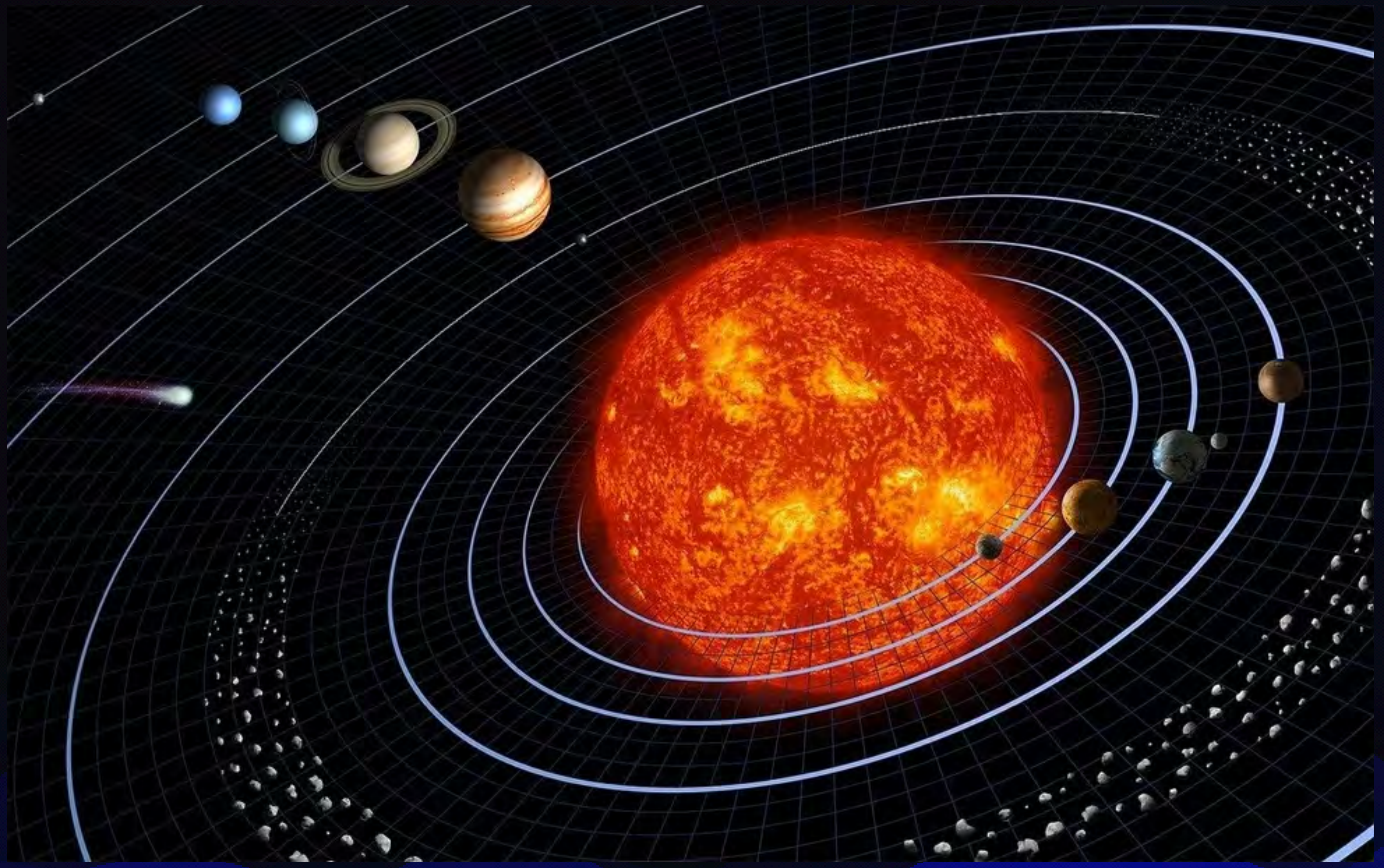
Hypothesis vs Theory

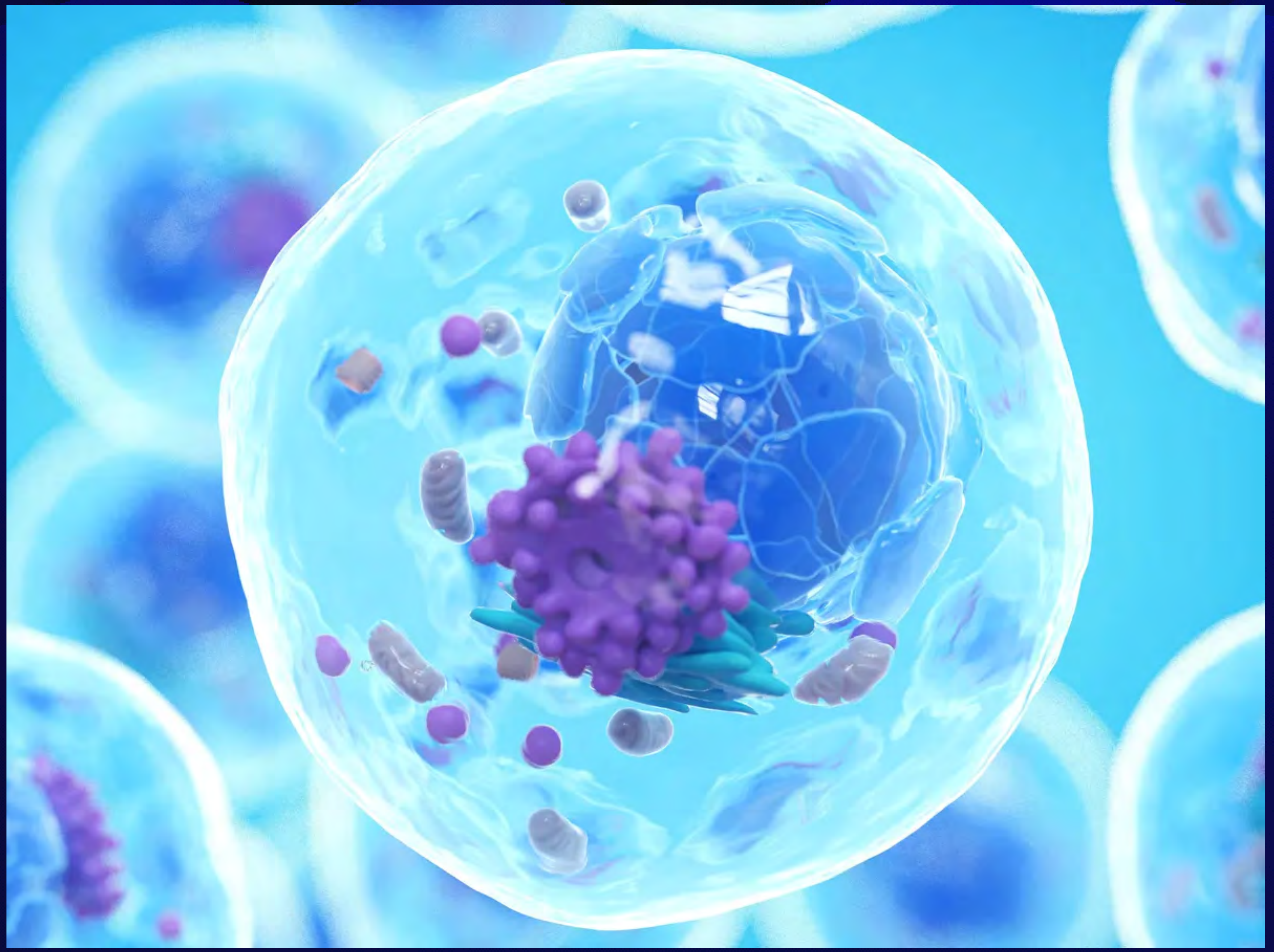
Fun with Terminology

- Which one means a proposed explanation (or guess) of a (Biological) Phenomenon?
- Which one means a well-supported understanding of a (Biological) Phenomenon?

What is a Theory?
Fun with Scientific Terminology

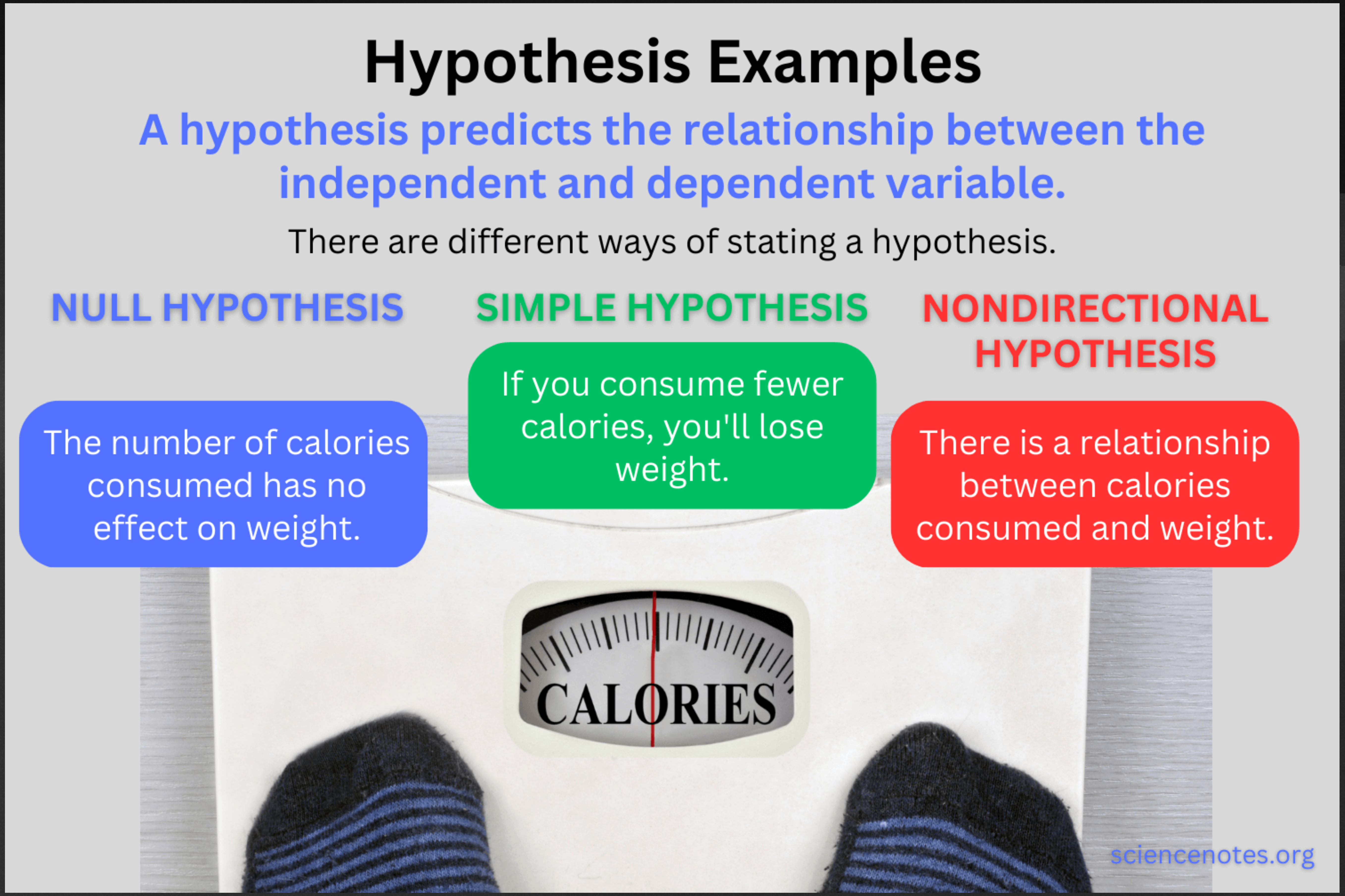
- A comprehensive explanation of some aspect of nature that is supported by a vast body of evidence
- A Theory is WELL Supported; it is not a guess or speculation
- Properties of a theory
  - Has been tested numerous times
  - Found to explain previous observations
  - Makes accurate predictions about future observations.
- Examples
  - heliocentric theory – Earth revolves around the Sun
  - Cell theory – living things are made up of cell





What is a Hypothesis?
Fun with Scientific Terminology

- An educated guess about how things work
- It is an attempt to answer your question with an explanation that can be tested.
- A good hypothesis allows you to then make a prediction:
"If [I do this], then [this] will happen."



Scientific Method

Scientific Method

a series of procedural steps

1. OBSERVATION

2. HYPOTHESIS

3. EXPERIMENT

4. CONCLUSIONS

Scientific Method
a series of procedural steps
Experimentation
Conclusion
Hypothesis
Observation
Exp Working?
yes
no
trouble shoot
Test with an Experiment
check all your steps
Analyze Data
Draw Conclusions
Results align with hypothesis
Results do NOT align with hypothesis
Report Findings
Construct Hypothesis
Background research
Ask a Question
Has anyone asked this question or similar before? What did they find?
Use your Experimental data as the background research
Hmm, Ask another Question

Experimental design

How to design an experiment

Variables

Experimental Design

- A V**ariable** is anything that can change or be changed
- You run an experiment when you want to determine causality
  - i.e. To determine if one Variable causes a change in another Variable
- Consider a car race:
  - What can change in a race?

Variables

Experimental Design

- What can change in a race?
  - type of car, color, speed, distance, traffic weather, mood of driver, age of driver, etc
  - These are called variables
- In an experiment, should we vary all of our variables?
  - You don’t want to vary everything: too complicated
  - You want to **control** some of your variables: make sure that they don’t change

Example

Experimental Design

- **Hypothesis**: The color of the car affects how fast you drive
- Psychological Experiment:
  - Participants are assigned to four conditions: a cyan, blue, red, and green Porsche 911.
  - Each Participant races around an empty racetrack 50 times
- What changes? What stays the same? What's measured?

Types of Variables

Control vs Independent vs Dependent

- **Control Variables**: anything that is held constant or limited in a research study.
- **Independent Variable**: the variable you purposefully modify
- **Dependent variable**: the variable that you measure

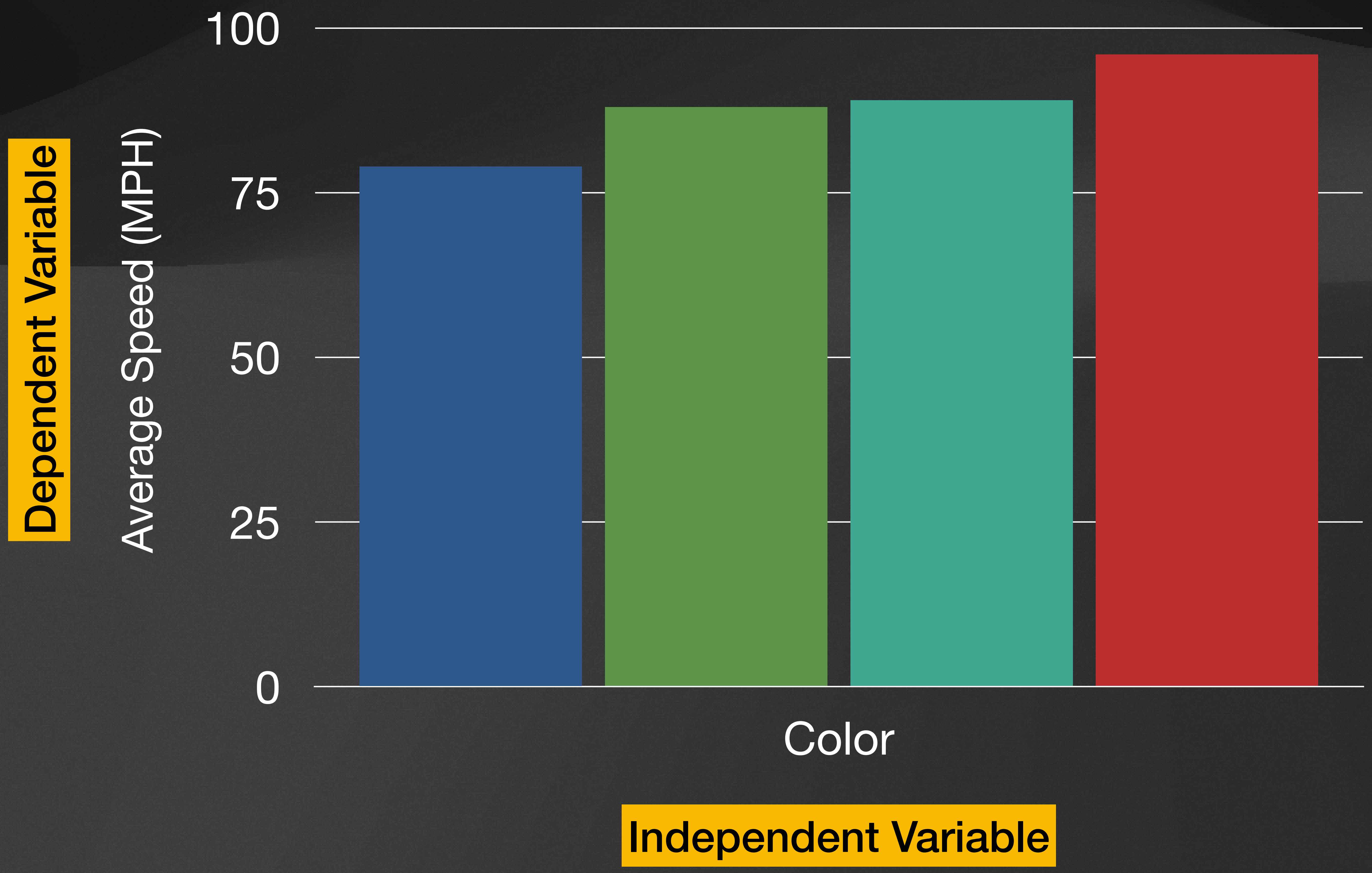
Variable	Value	category
type of car	Porsche 911	Control
distance	50 times around track	Control
weather	Sunny day	*mostly* Control
traffic	no one else on track	Control
color	cyan, blue, red, or green	independent variable
mood of driver	Unknown	assumed control/ignored
age of driver	18-34	Age-group Control
speed	measured for each participant	dependent variable




Types of Variables

Plotting your Results

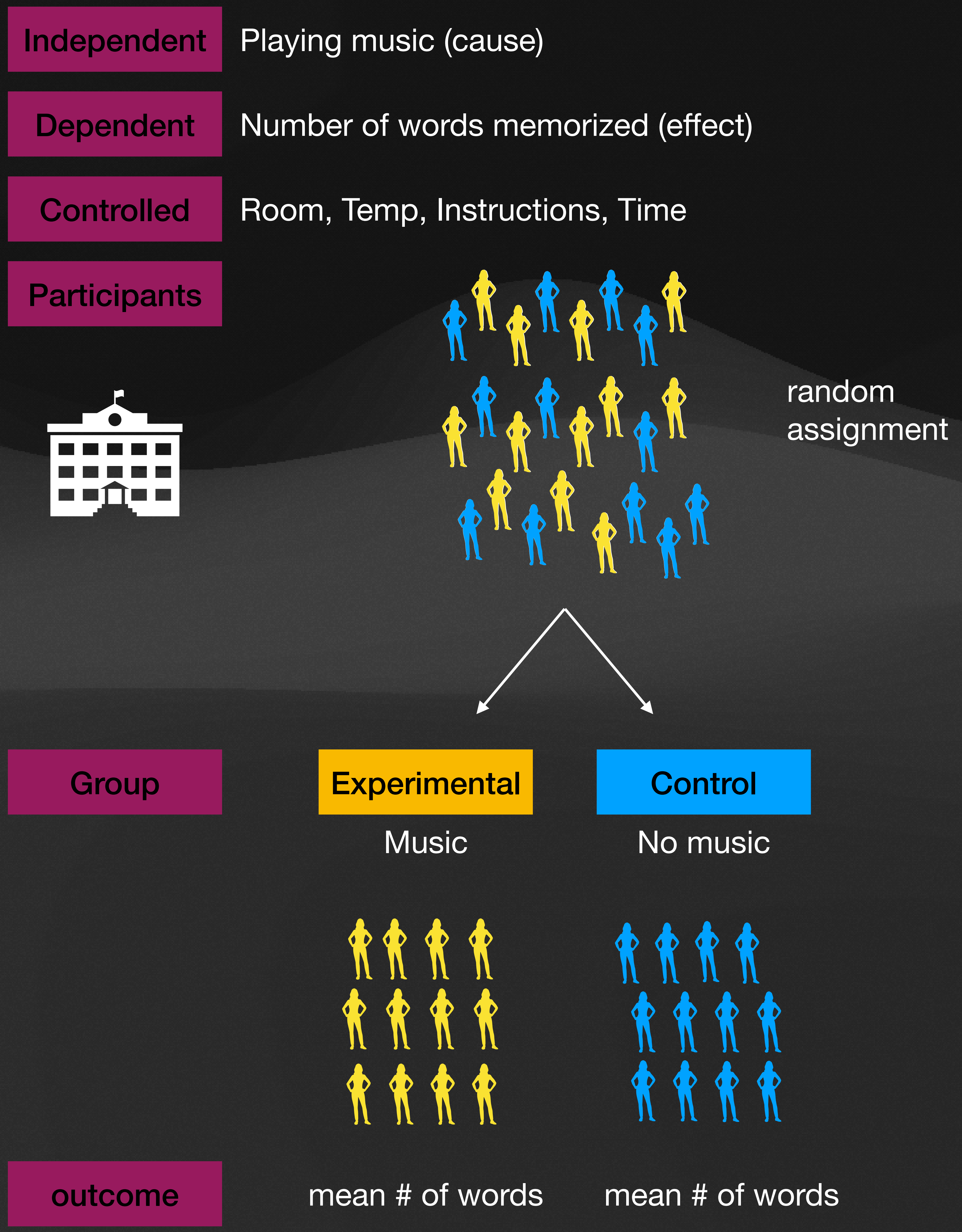
- **Independent Variable**: x-axis (cause)
- **Dependent variable**: y-axis (effect)



Experimental Groups Example

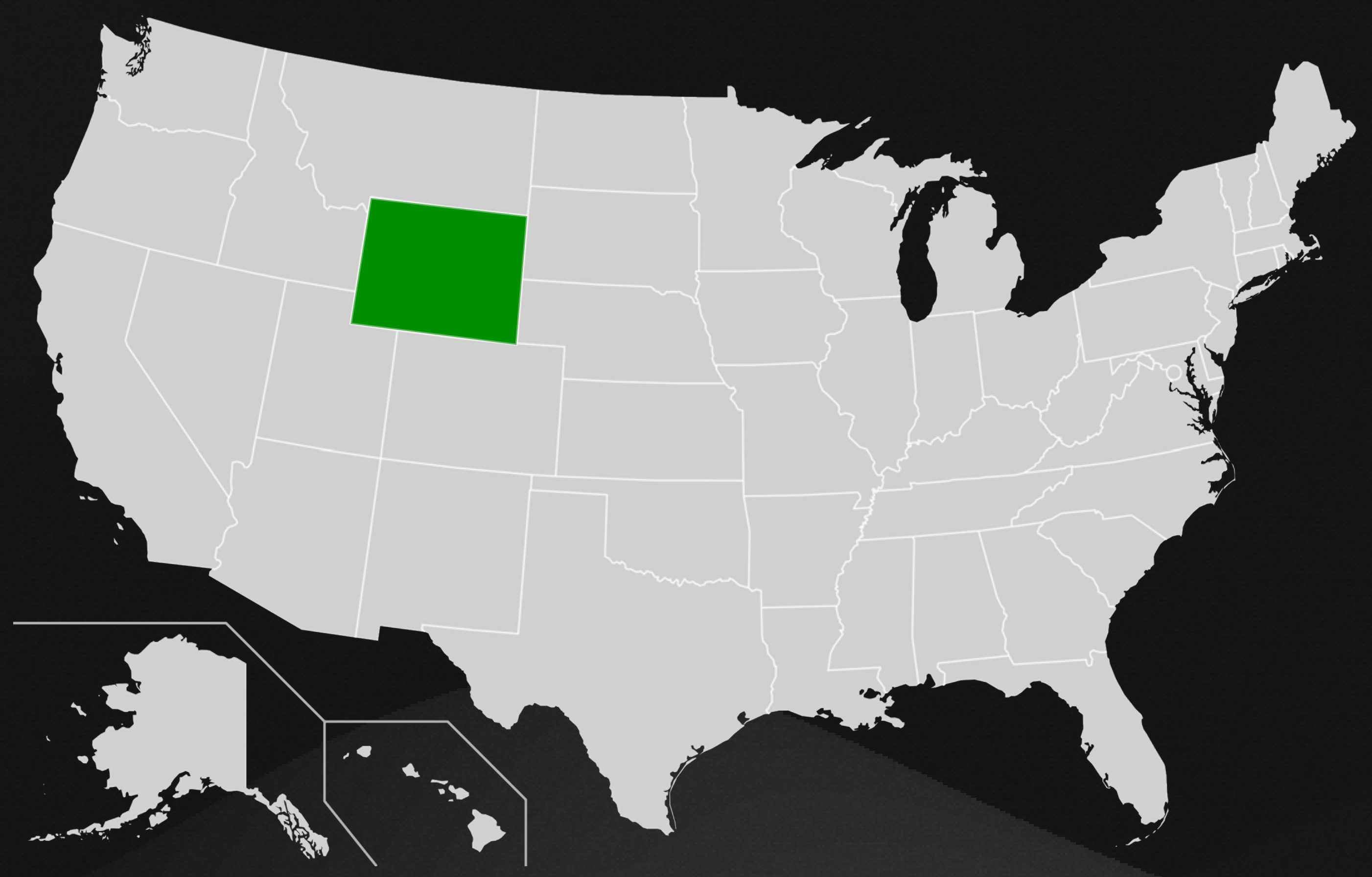
Experimental Design

- **Question**: Does listening to music improve learning?
- **Hypothesis**: Listening to classical music improves the ability to memorize 40 random words in 5 minutes
- **Variables**
  - Independent Variable?
  - Dependent variable?
  - Controlled variables?
- **Participants**: 24 students in a class

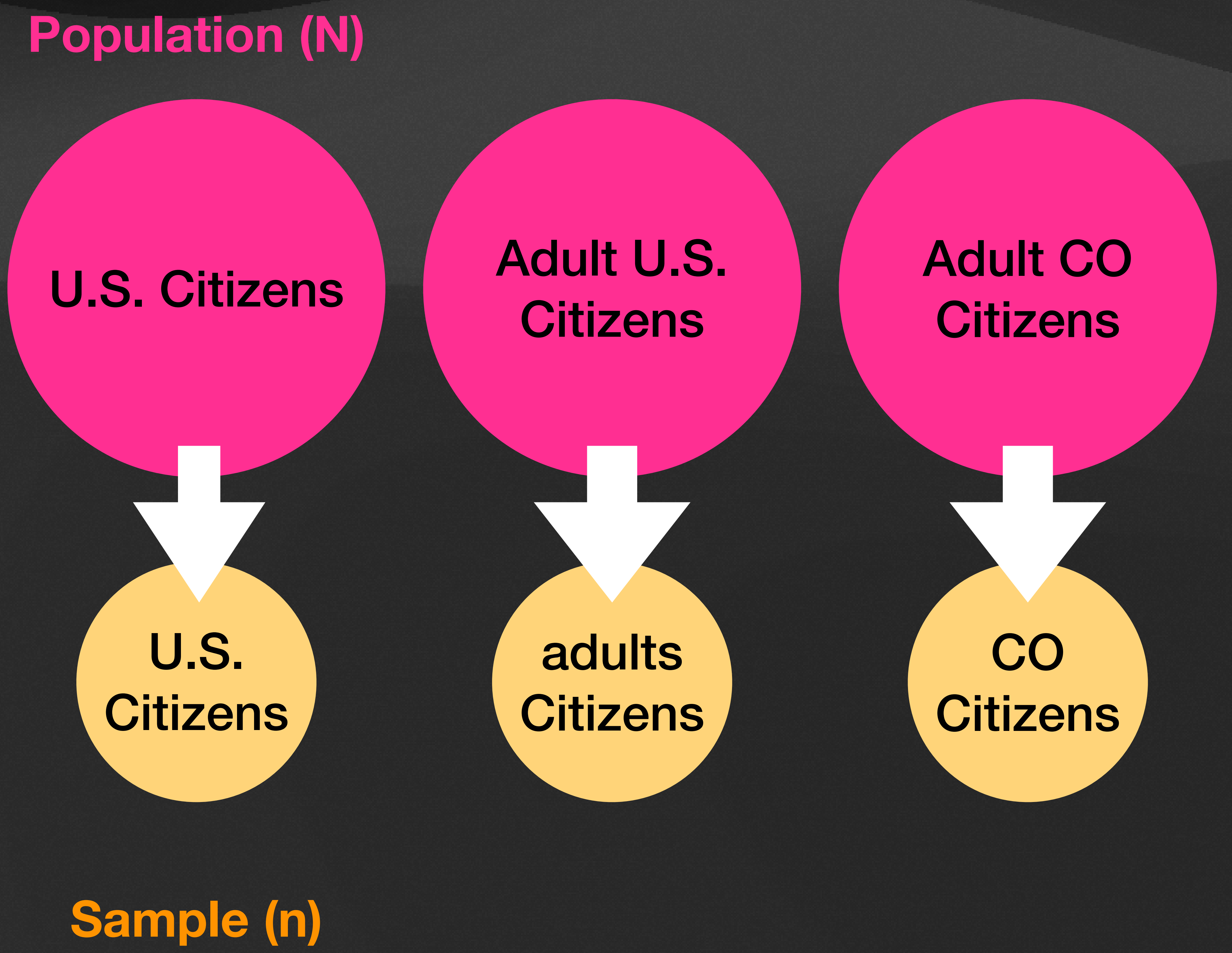


Sampling

Important Terms

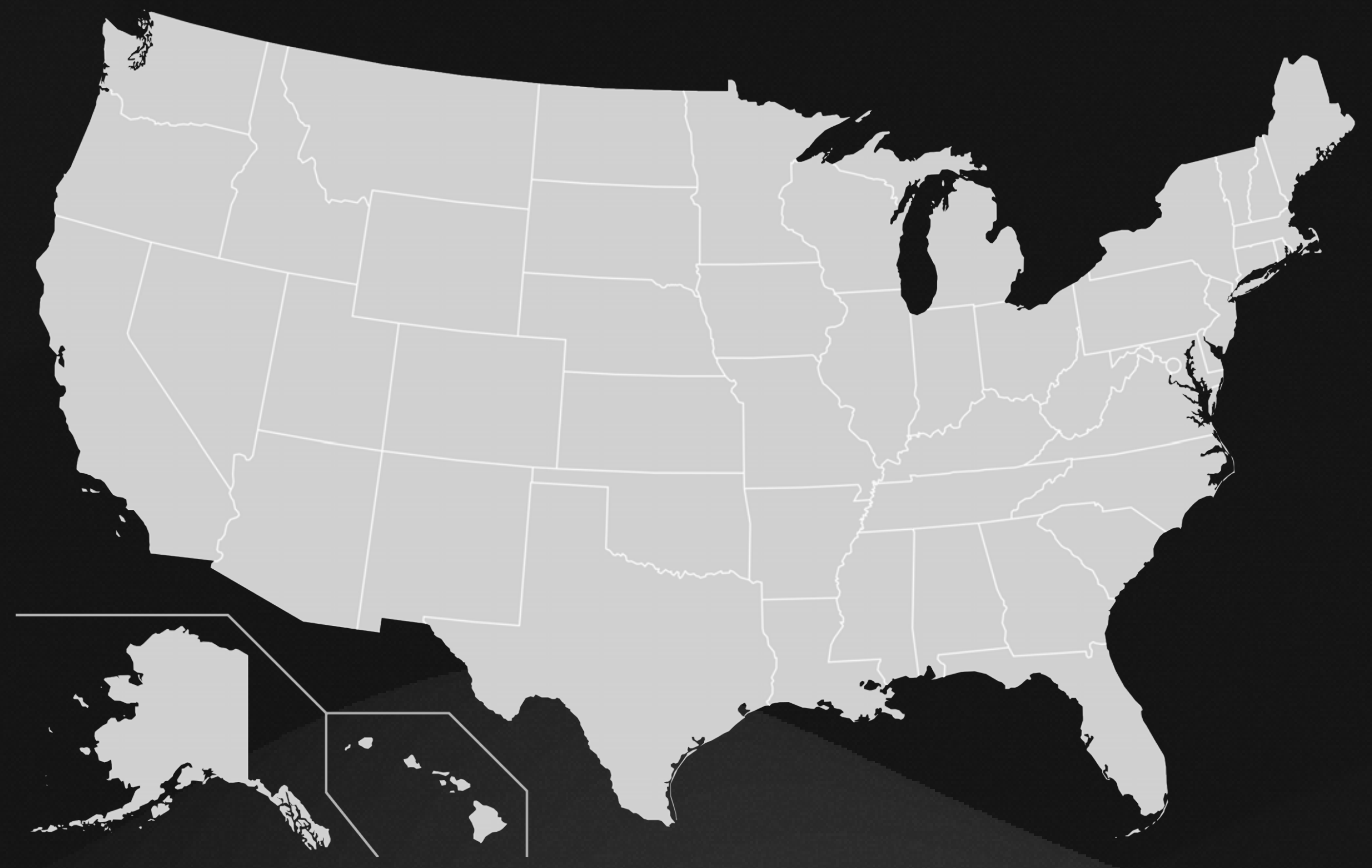


- **Population (N)** - the entirety of something
  - whatever it is that you want to know about
  - Can be defined however you want
- **Sample (n)** - a subset of the Population
  - **Non-probability** Sampling
    - Just sample what's available, convenience
  - **Probability Sampling**
    - Random selection
    - Stratified selection

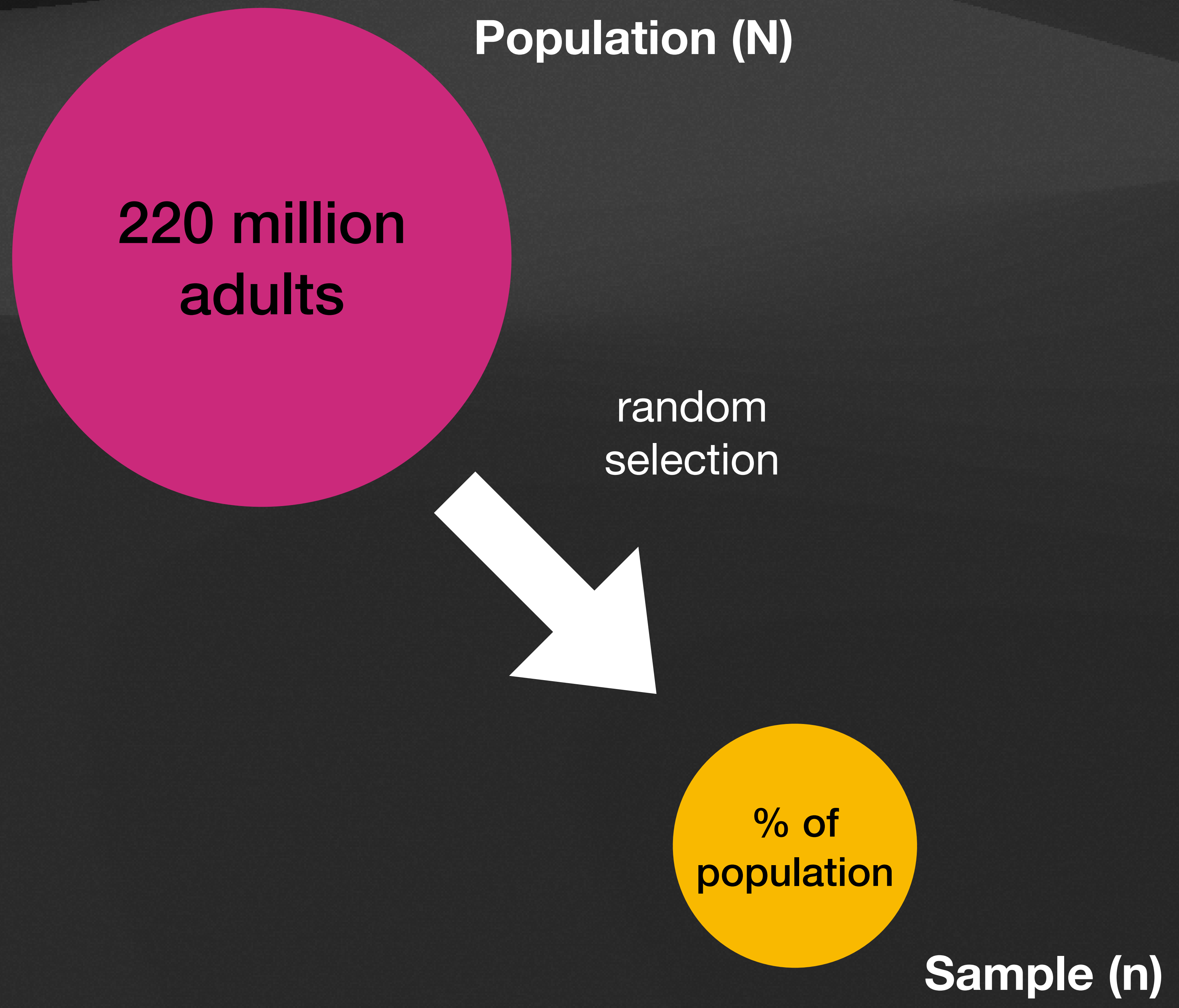


Random Sampling

What % of Americans support Marijuana Legalization?

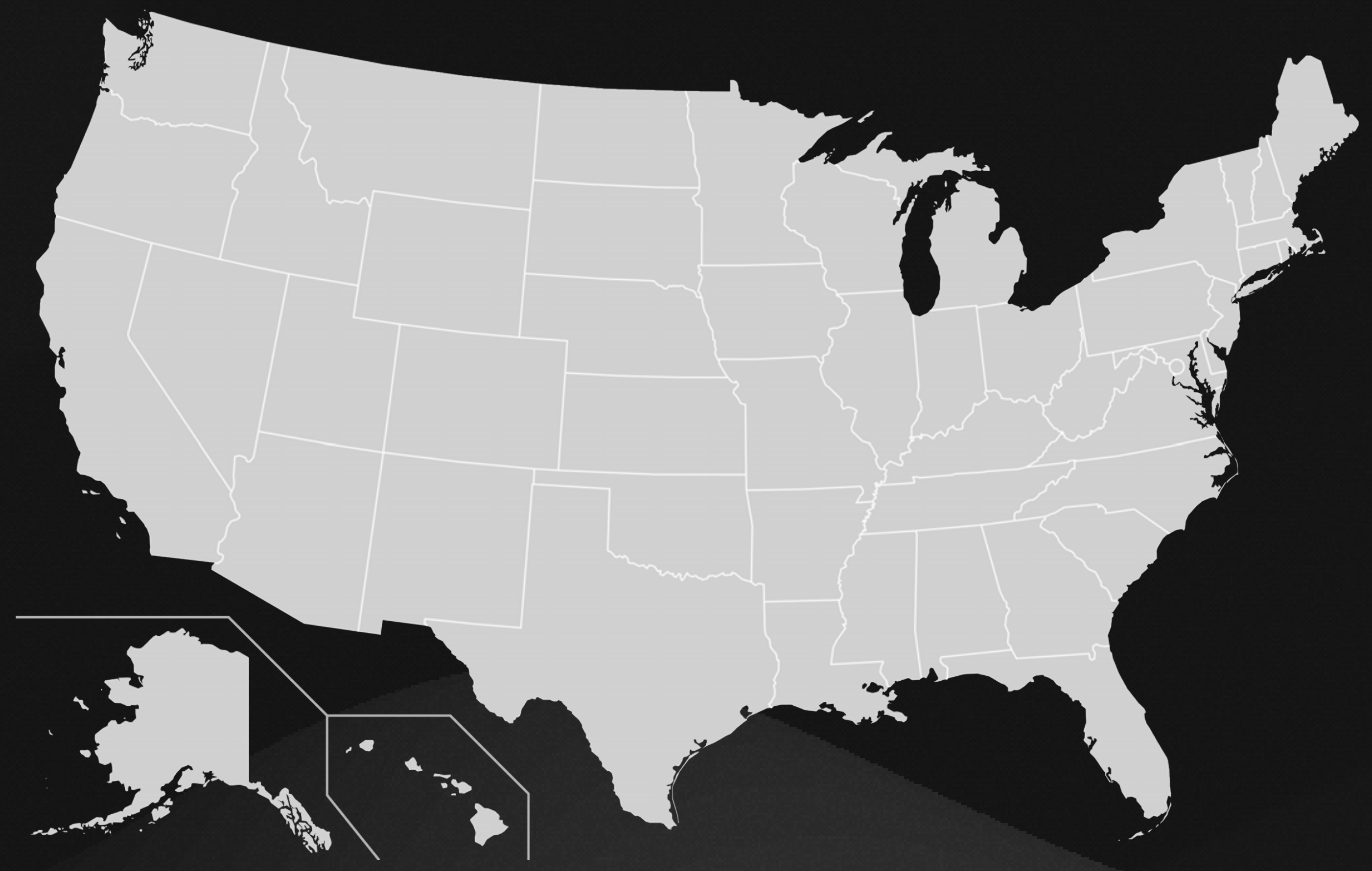


- Population?
  - Everyone over 18
  - they can give consent and answer our questions
- Sample?
  - **Random Selection** (like picking names out of a hat, or the lottery)
  - n = 2,000 people

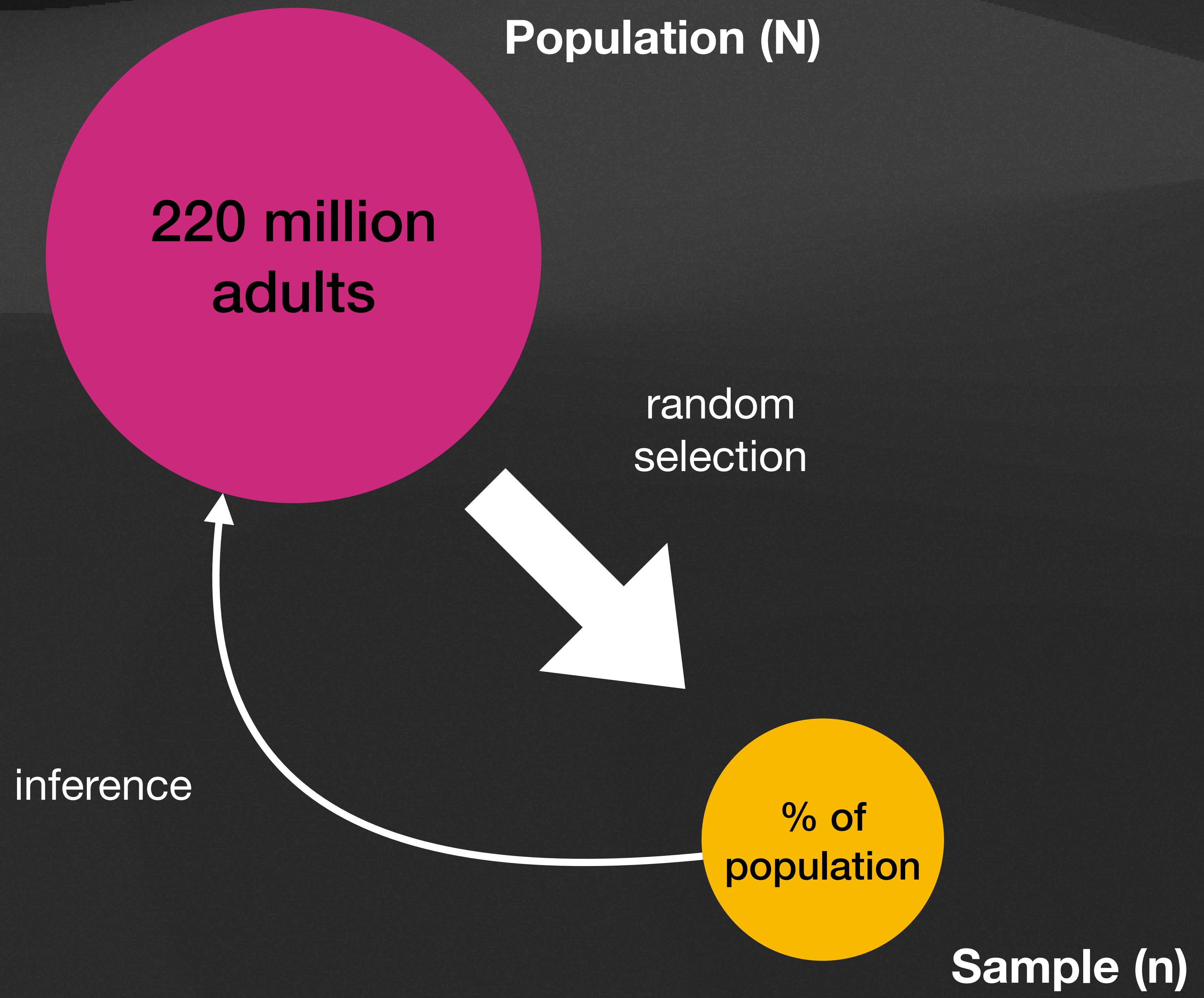


Validity

What % of Americans support Marijuana Legalization?



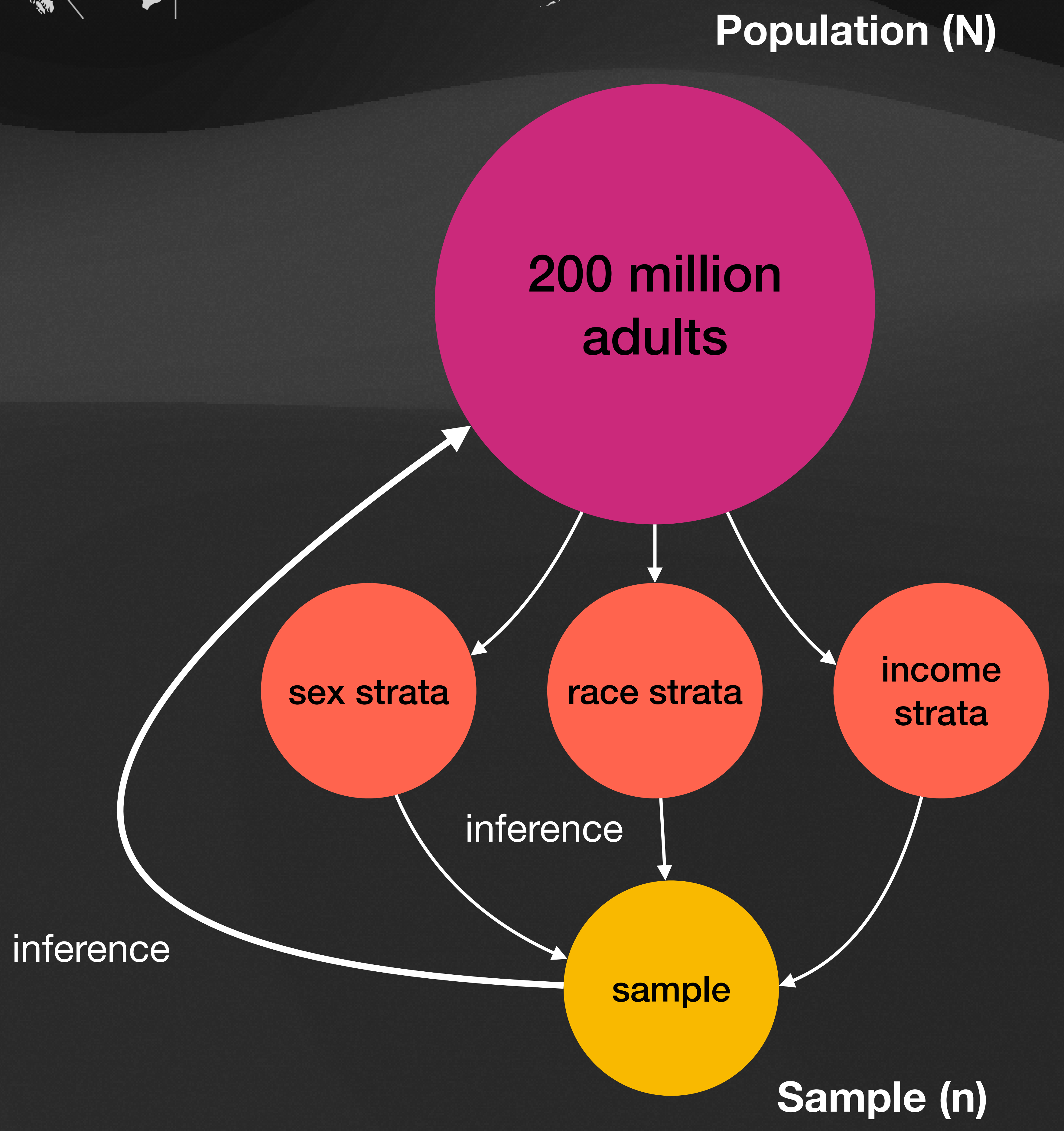
- The Key to a good sample is that it is **Representative** of the entire Population
  - If your Population is 50% women, your sample should be 50% Women
- High External Validity
  - Representative Sample - you can make good inferences about the population
- Low External Validity
  - **Sample bias**: sometimes random numbers can be random
  - Not representative, not good for inferences



Stratified Sampling

Experimental design

- Better Method for Larger Populations
  - helps Avoid sampling bias
  - Ensure all groups are represented
- Strata means layer
  - subgroups of a population
- Divy up your population into different strata
  - age, sex, income, race
    - e.g. Number of Men and Women should be equal
- Sample is generated by randomly choose participants from each strata



Data Wrangling

Managing Data

Measures of Central tendencies

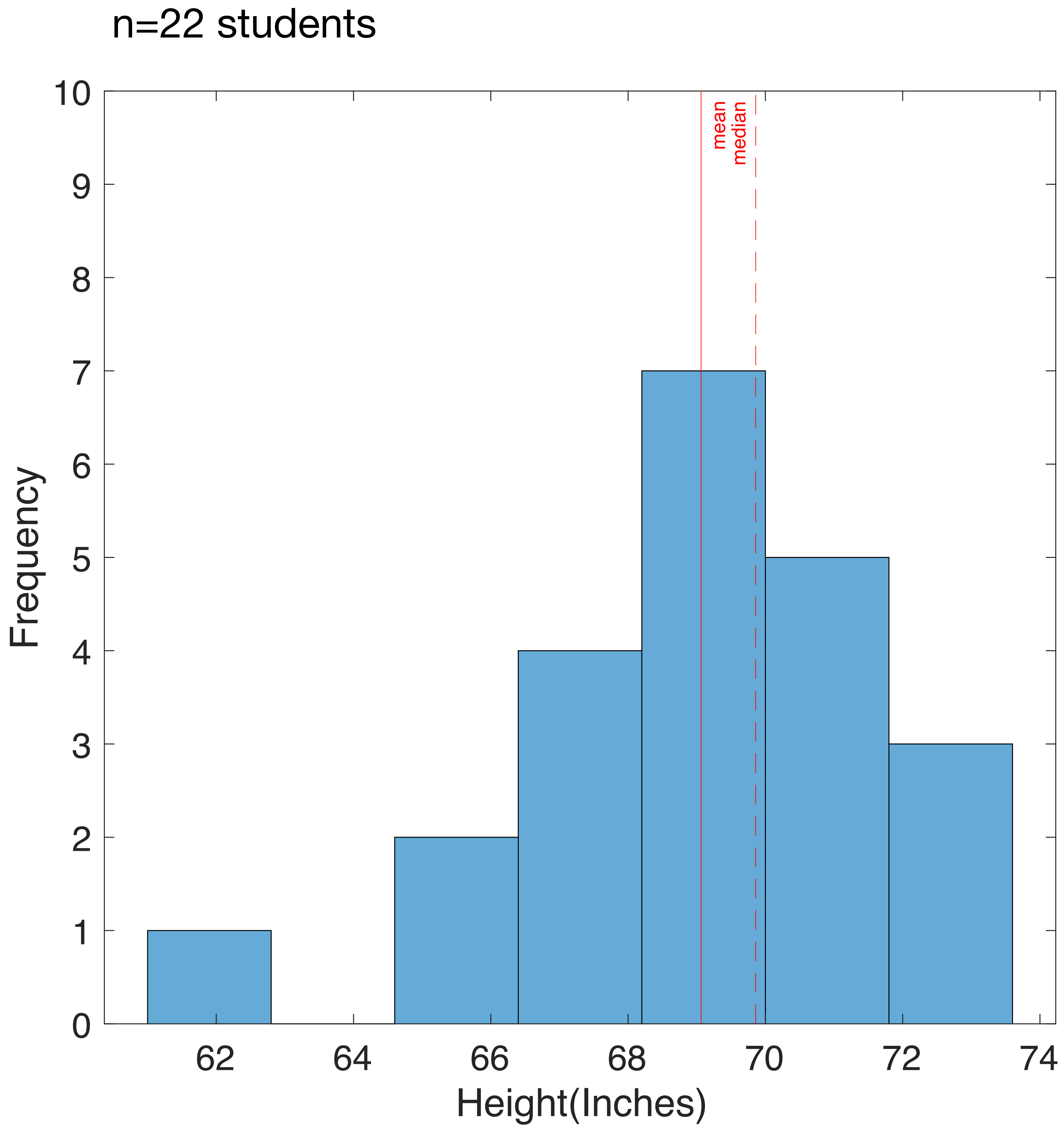
- What do you do once you measure something?
    - like Heights
- Measures of central tendency
    - mean
    - median
    - standard deviation
    - range

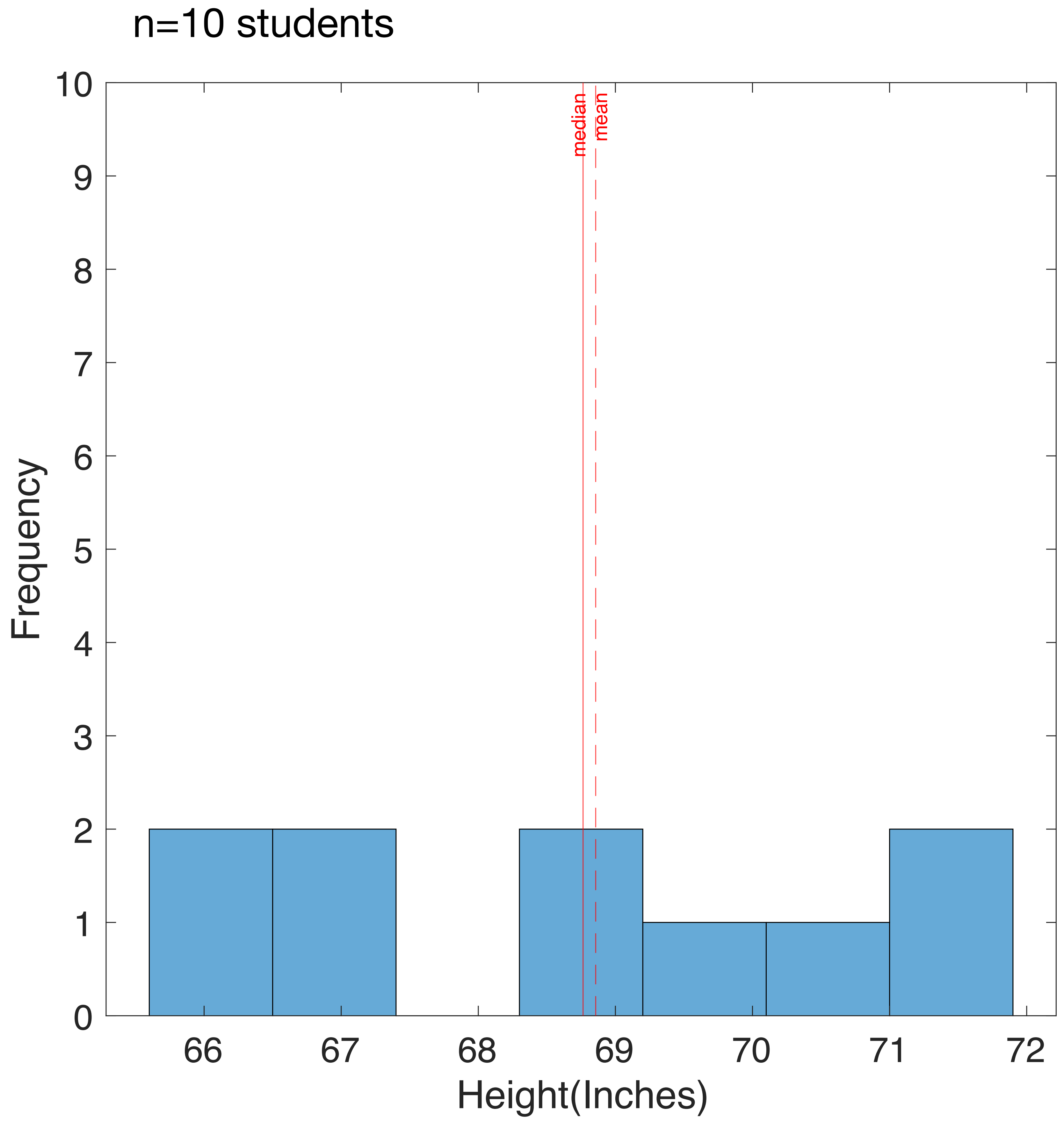
student id	height (in)
1	70.9
2	73.2
3	70.3
4	71.7
5	70.9
6	68.3
7	69.8
8	67.1
9	71.3
10	66.2
11	66.4
12	67.1
13	61.7
14	72.7
15	69.9
16	67.2
17	72.5
18	64.8
19	68.8
20	68.5
21	69.9
22	69.9

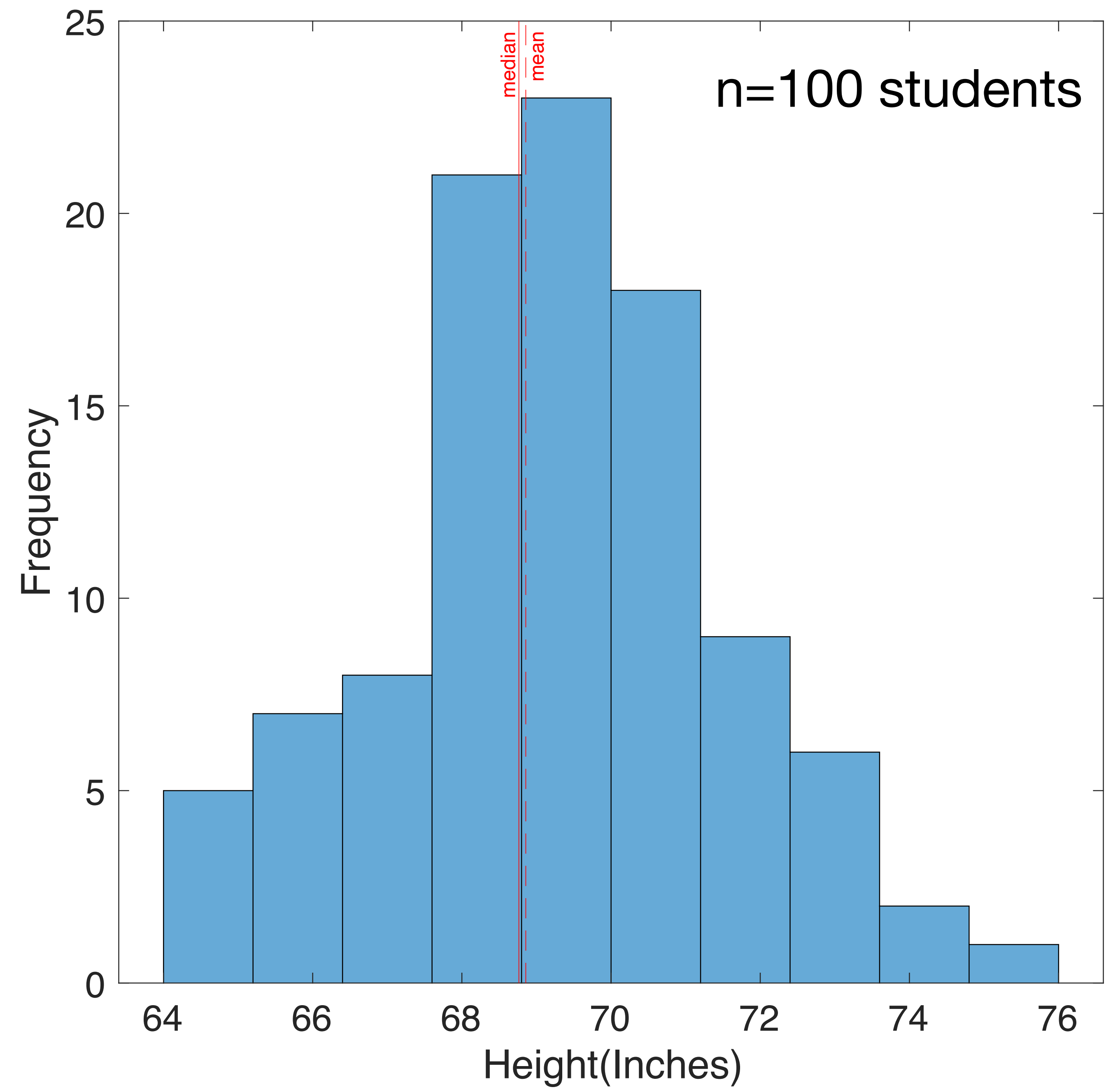
mean = 69.42”
std = 2.2”

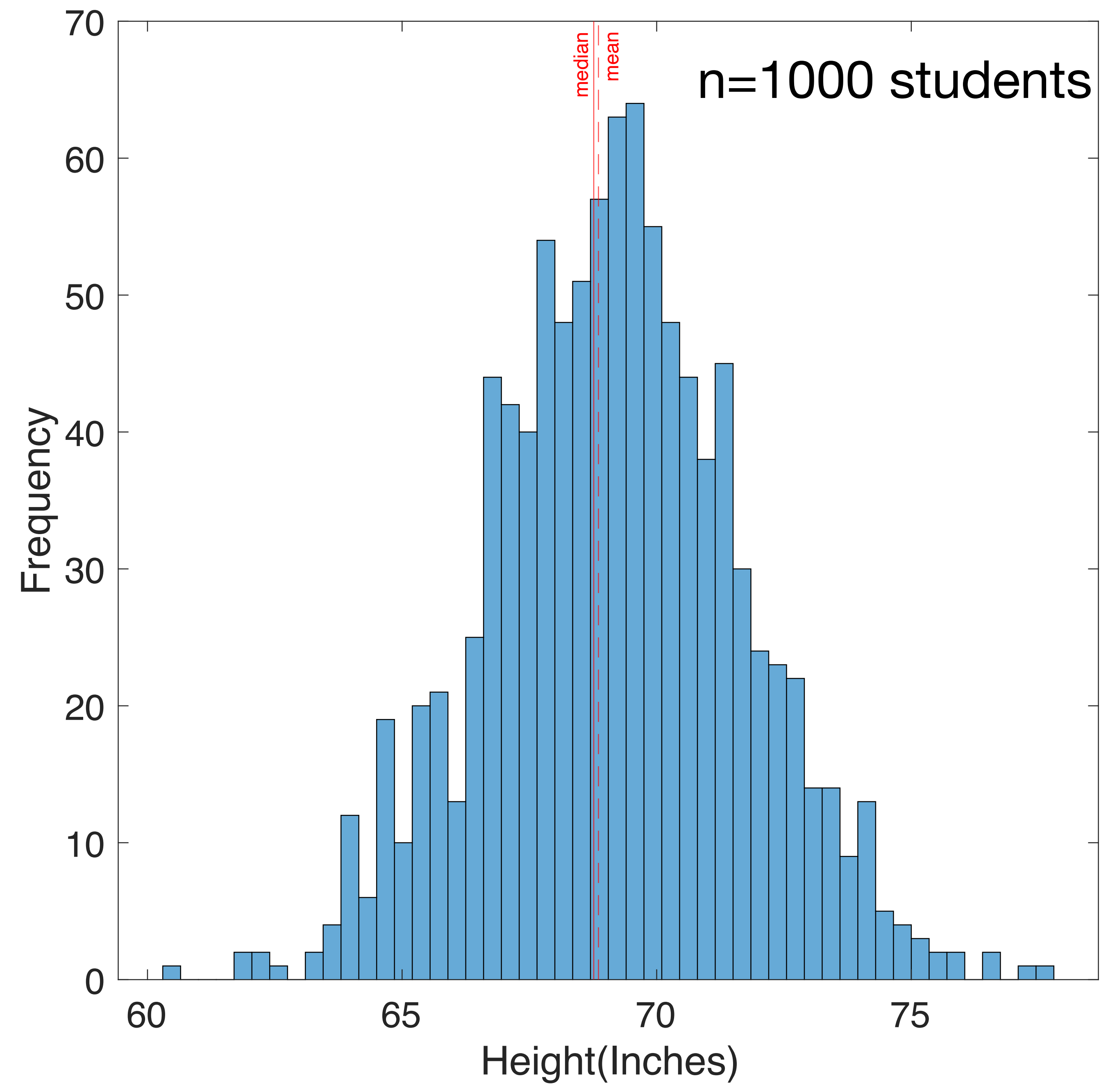
Managing Data

Sample Size affects data analysis





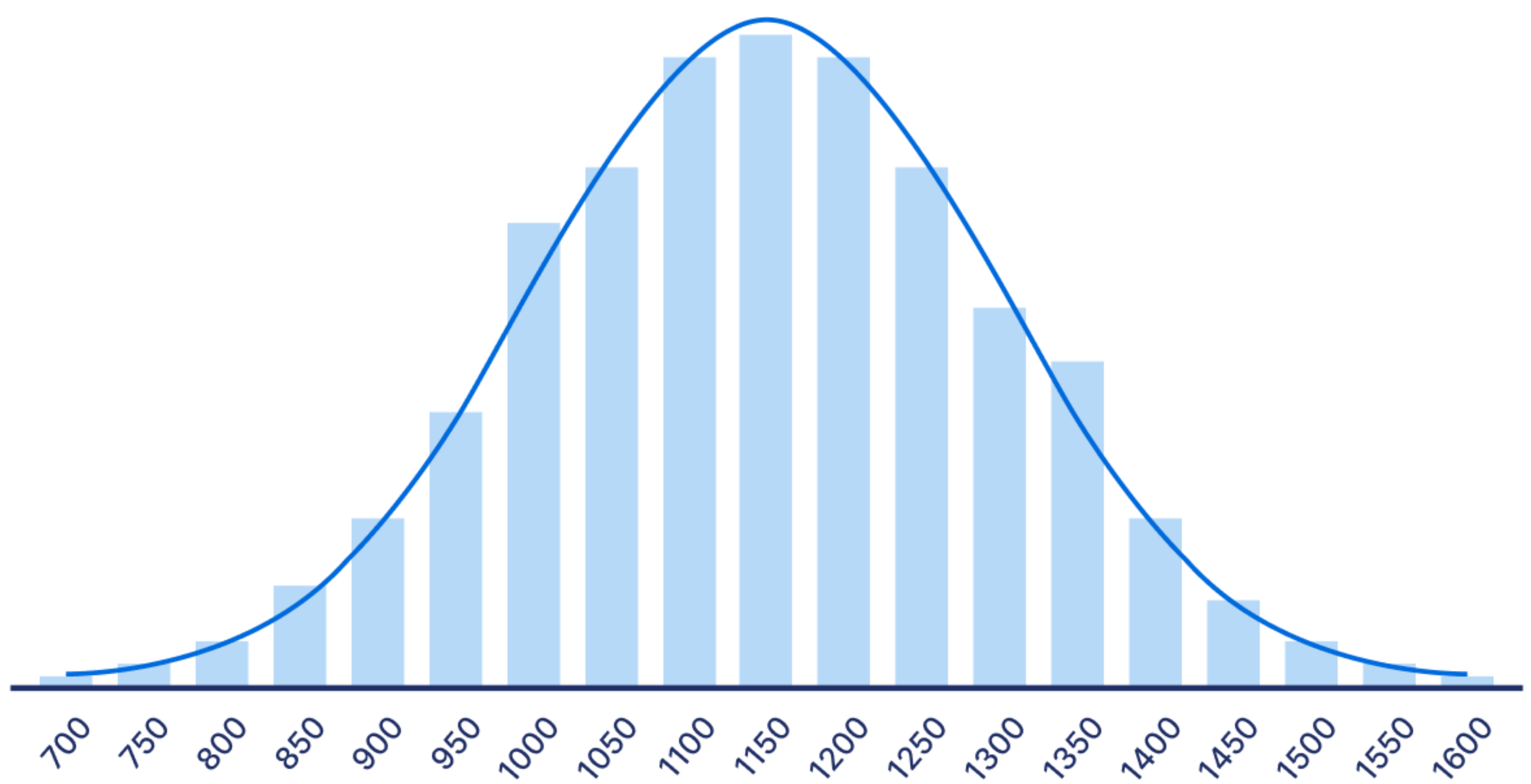




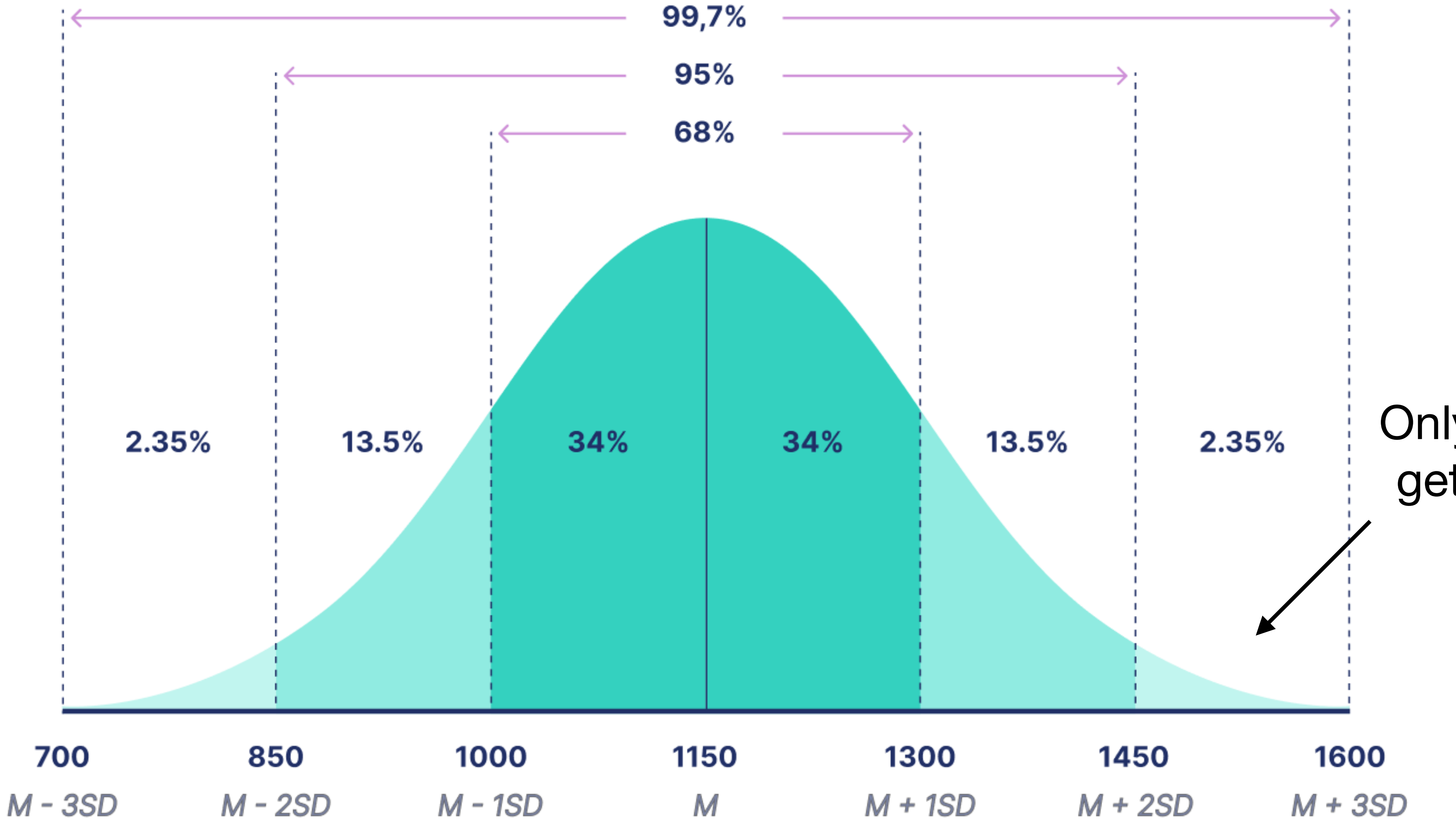
Managing Data

Normal Curve

Normal curve fitted to SAT score data



Using the empirical rule in a normal distribution



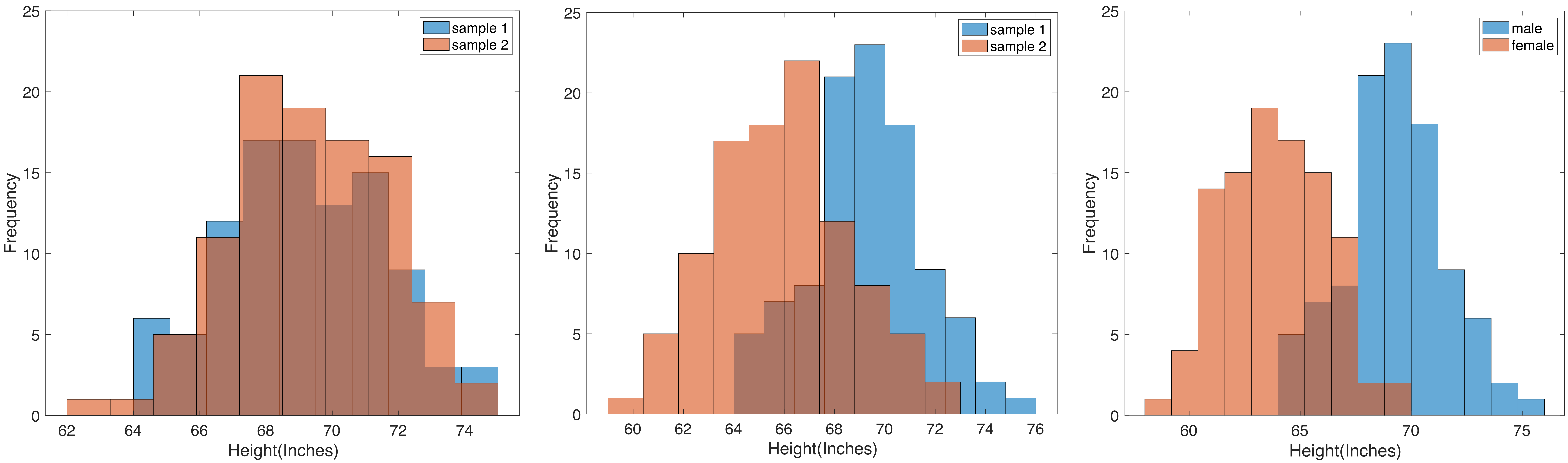
Only 2% of people get a score in this range

Using Normal Curves allows us to assign a probability to each measured value.

Multiple samples

Are they different?

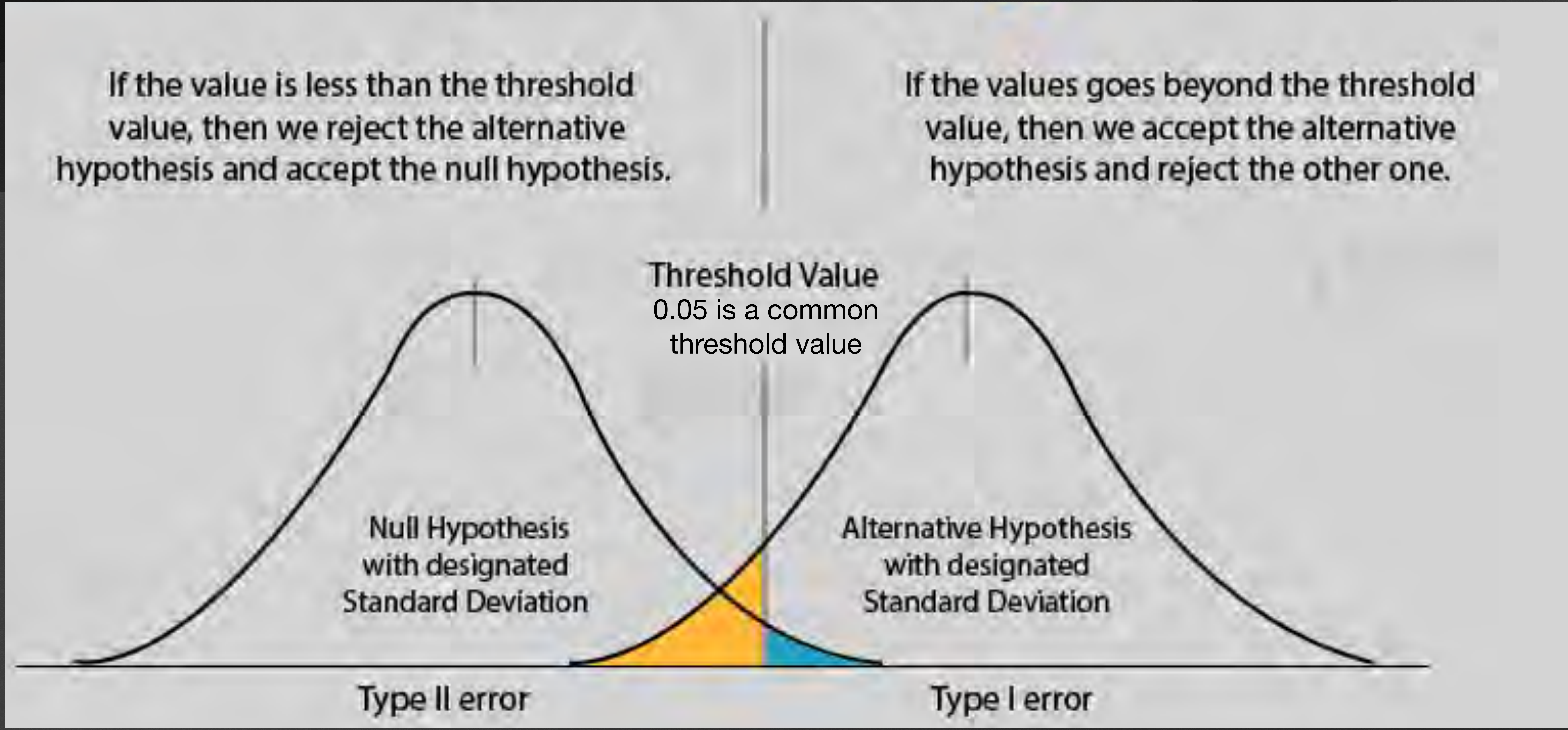
- With Random sampling, you will often get slightly different means and data distributions just by chance
- So, how can you tell if they are the same or different populations?



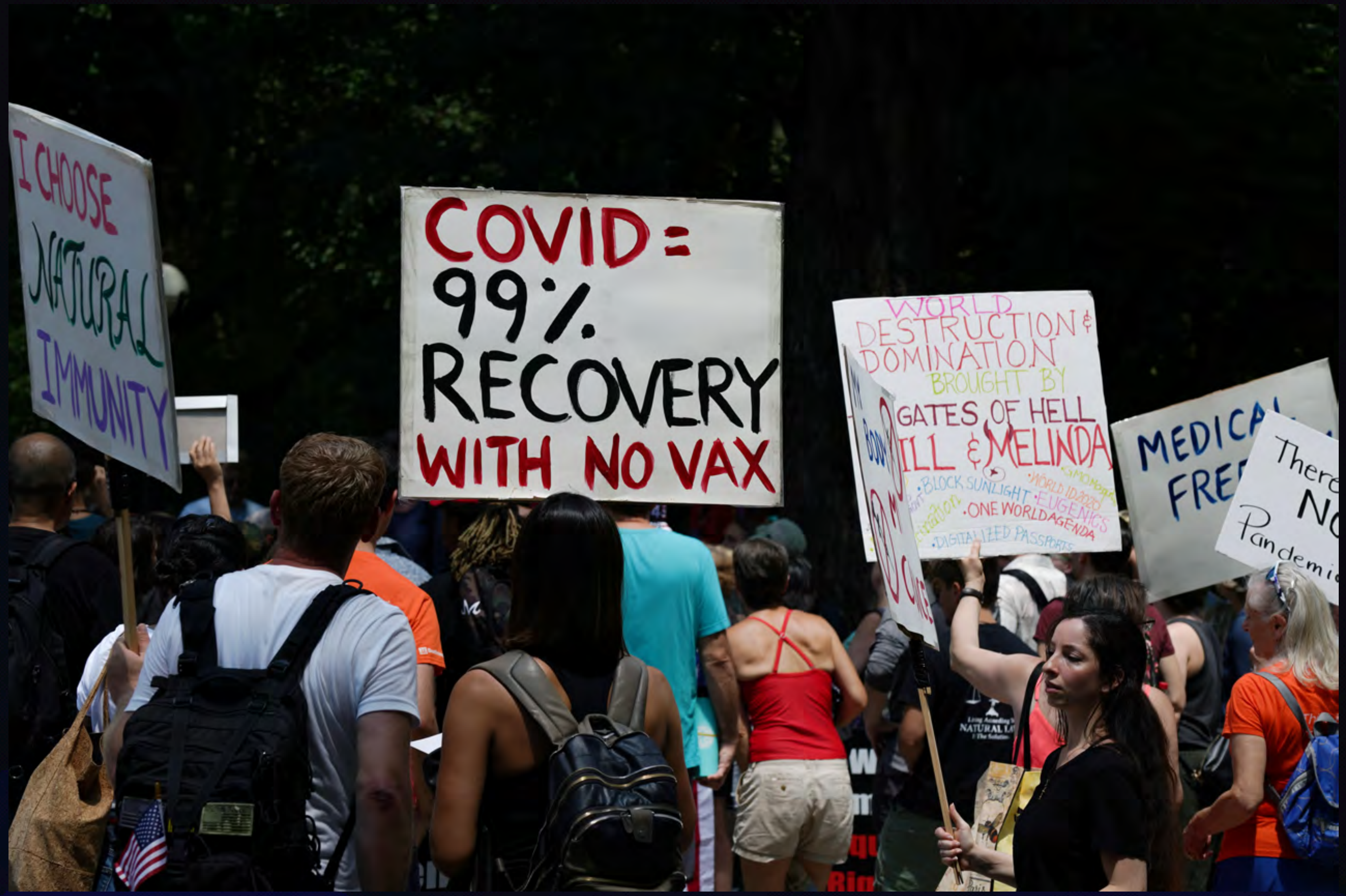
Hypothesis Testing

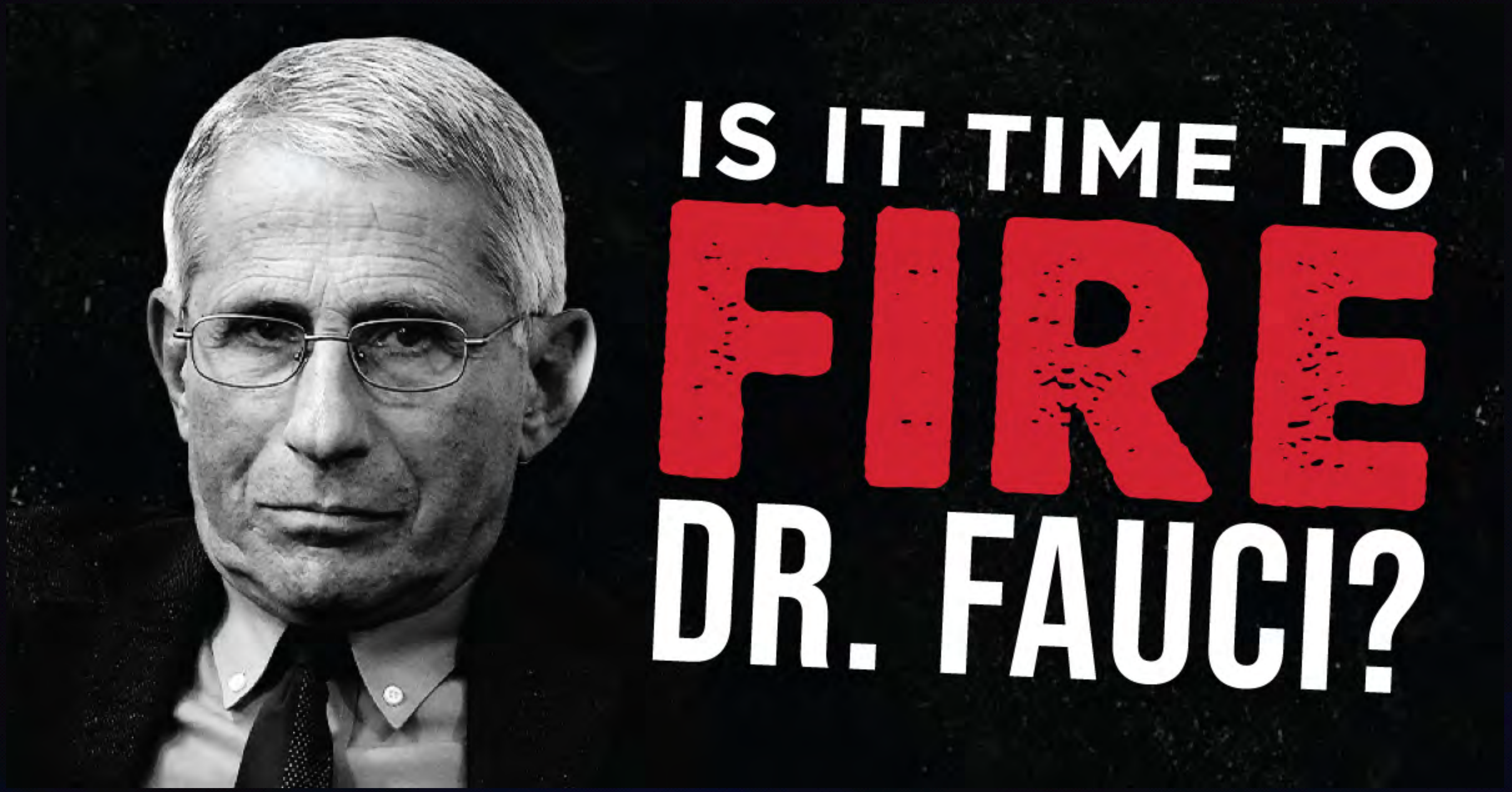
Common stats test to determine whether two samples are the same or different

- Null Hypothesis: the means are the same
- Alternative Hypothesis: the means are different



5% Significance means 1/20 chance - not impossible





Science in Crisis?

"A lie can travel halfway around the world while the truth is still putting on its shoes."

- Mark Twain

"A Lie runs until it is overtaken by the truth"

- Cuban Wisdom

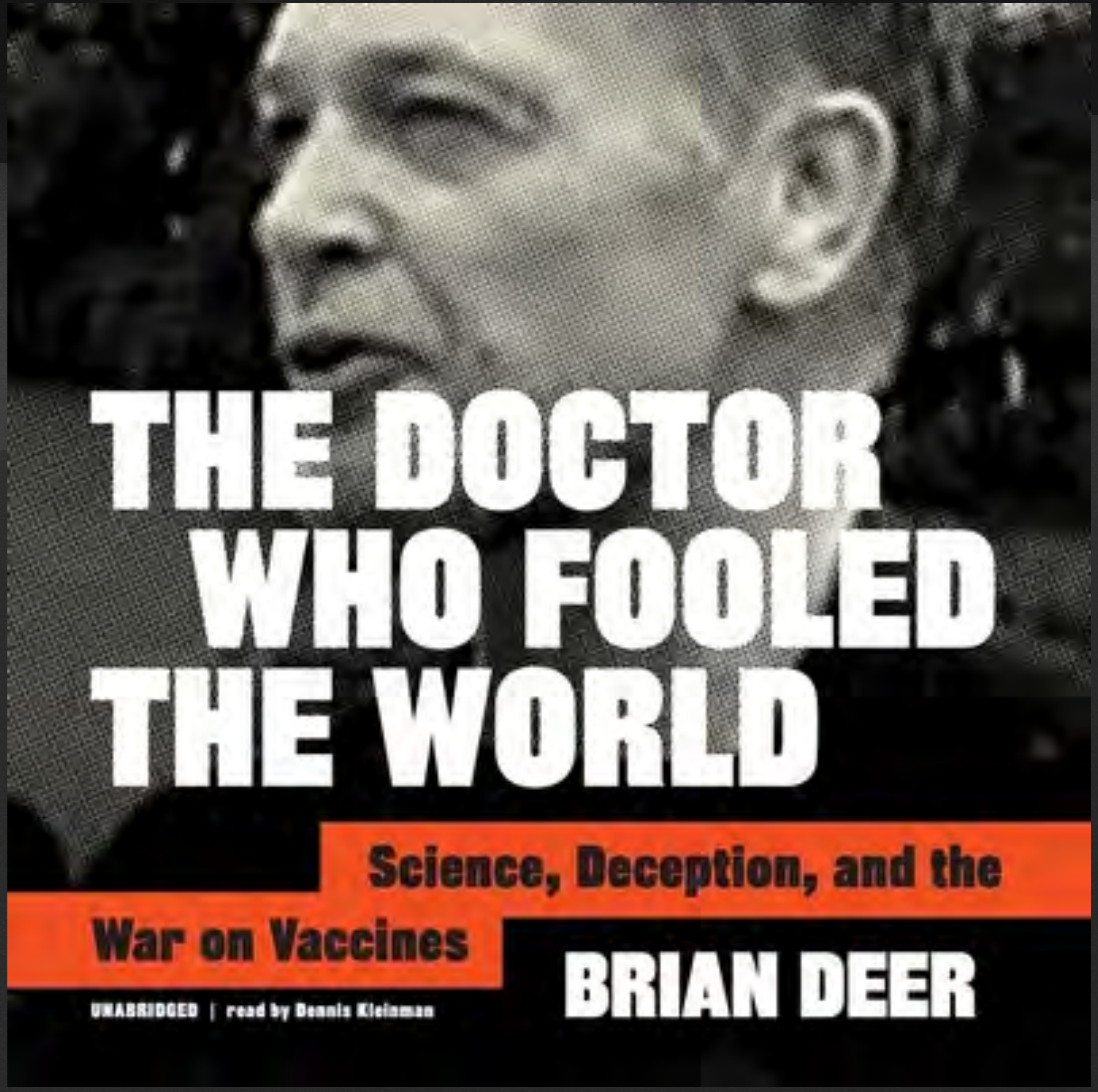
"It takes way more energy to refute bullshit than it does to produce it."

- Brandolini's Law

History of Vaccine Hesitancy

Science in Crisis?

- Anti-vax movement galvanized by 1998 scientific study
    - Andrew Wakefield, former British physician
        - Paper: MMR vaccine causes autism
    - Shoddy Paper was retracted a few years after publication
    - Wakefield found guilty of profession misconduct and his medical license was revoked
    - 20 years of research have thoroughly debunked the claims of the paper
- And yet…
    - Wakefield remains a fairly prominent Anti-vaxxer
    - Vaccine Hesitancy has metastasized to all vaccines, including COVID



The trouble with Ivermectin

Science not always great in the very short term

- Ivermectin is an anti-parasitic drug most commonly used in veterinary medicine to prevent and treat heartworm and acariasis.
- Early during the pandemic, Ivermectin was touted as powerful tool to combat COVID-19
  - Several Early studies showed improvement after taking ivermectin
    - Perhaps random sampling error?
    - Powerful dewormer — maybe effective in countries where people have parasites/worms?
    - Desperate need for any treatment?
  - Meta-analysis since then has shown to not be effective in treating COVID-19
    - still a valuable drug to treat parasites
    - FDA: “Do not use Ivermectin for treatment of COVID-19”
- Similar issues with Mask Usage
  - Early studies showed conflicted results - not even sure how COVID spread, etc
  - also suffered from becoming wrapped up in a Political Identity

Anthony Fauci, MD

Director of the National Institute of Allergy and Infectious Diseases (NIAID)

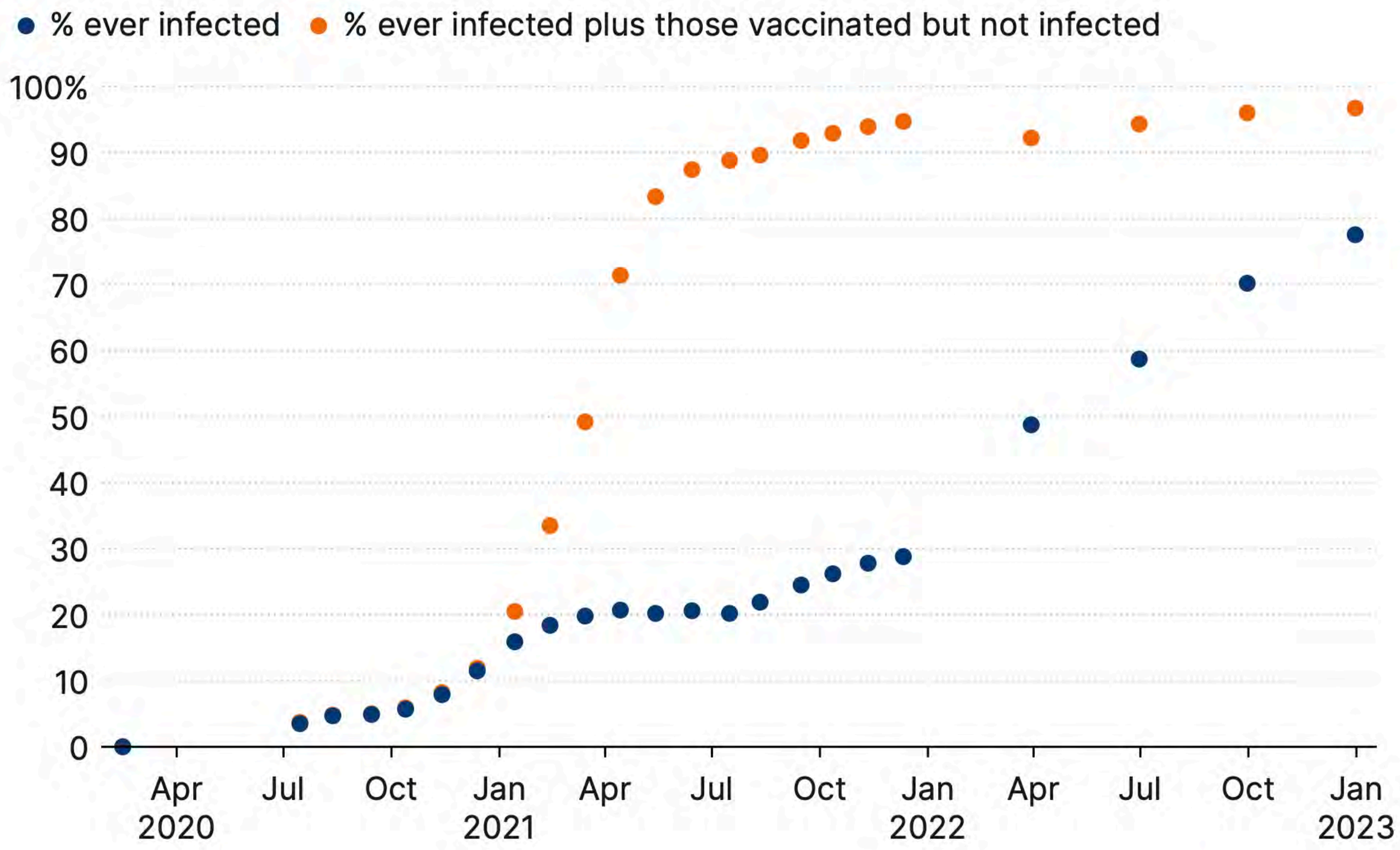
- Has served the public health sector for more than 50 years
- Key figure in battling the HIV/AIDS epidemic in the 1980s
- Awarded the Presidential Medal of Freedom in 2008 by George W Bush
- Member of the White House Coronavirus Task Force
  - became embroiled in controversy because his advice contradicted the Trump Administration (politics)
  - His advice has since borne out to be mostly true

The impact of vaccines and behavior on US cumulative deaths from COVID-19

Brookings Institute Mar 2024 (CU mathematician)

- Created a mathematical model of the Pandemic
  - Behavioral Changes: social distancing, masking, etc
  - Rate of Getting Vaccinated
- Found Vaccines and Behavioral changes inextricably linked
  - Eventually everyone would have been infected, no matter what
  - Without Behavioral Changes, everyone would have been infected faster
  - Without Vaccination, COVID 4X more likely to kill
- Social Distancing and Masking saved 800,000 lives
  - delayed infection rates
  - But had HUGE economic and social costs
- Vaccine hesitancy and the slowdown in their deployment “Cost an additional 273,000 preventable deaths” from the Delta and Omicron variants

Prevalence of COVID in U.S. Population



Source: Jones et al. (2021)

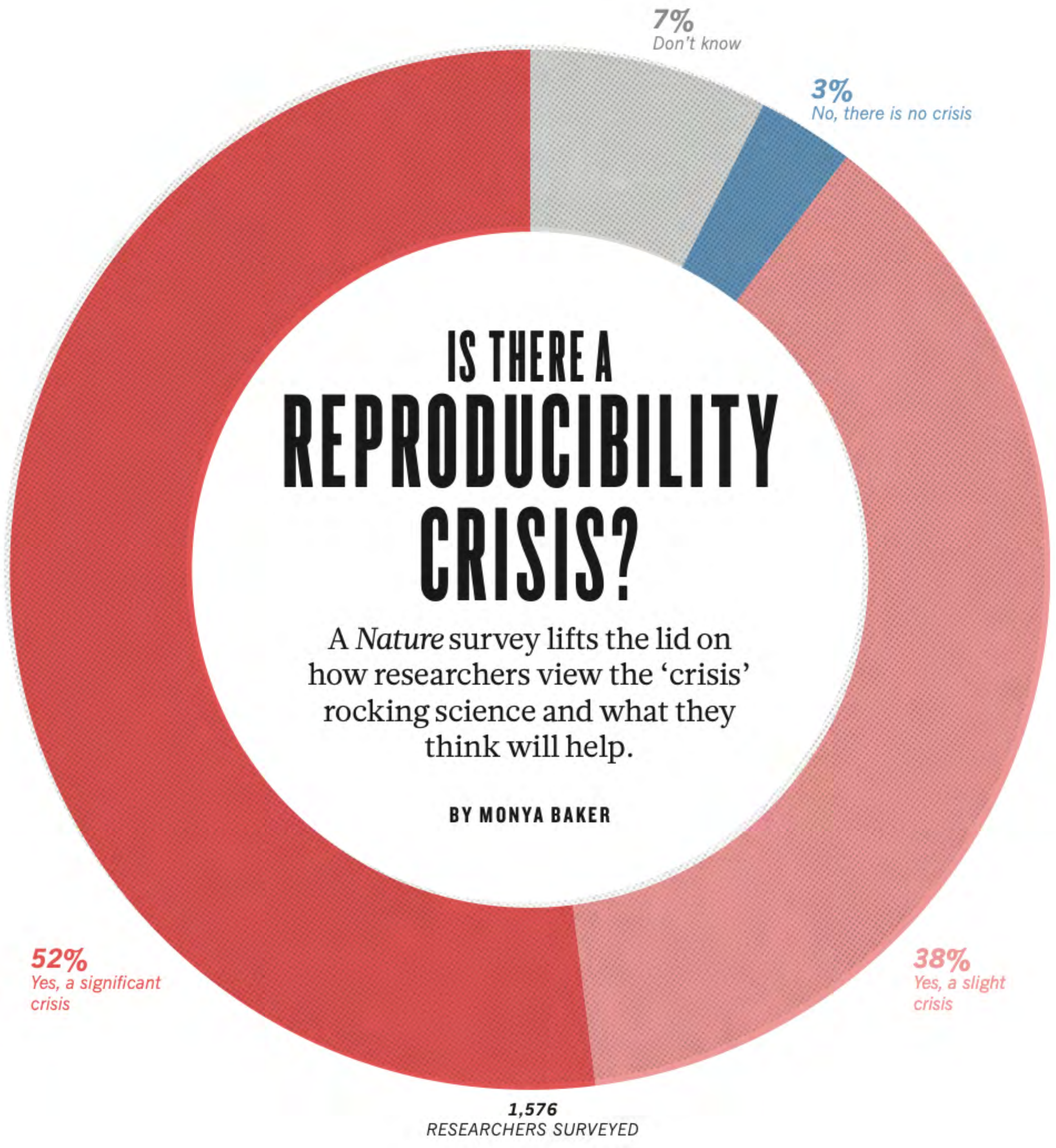
BROOKINGS

Note: Blue dots show infection levels in population; orange dots show combined seroprevalence resulting from infection and vaccination.

Reproducibility Crisis

Uh oh, we’re not getting the same results…

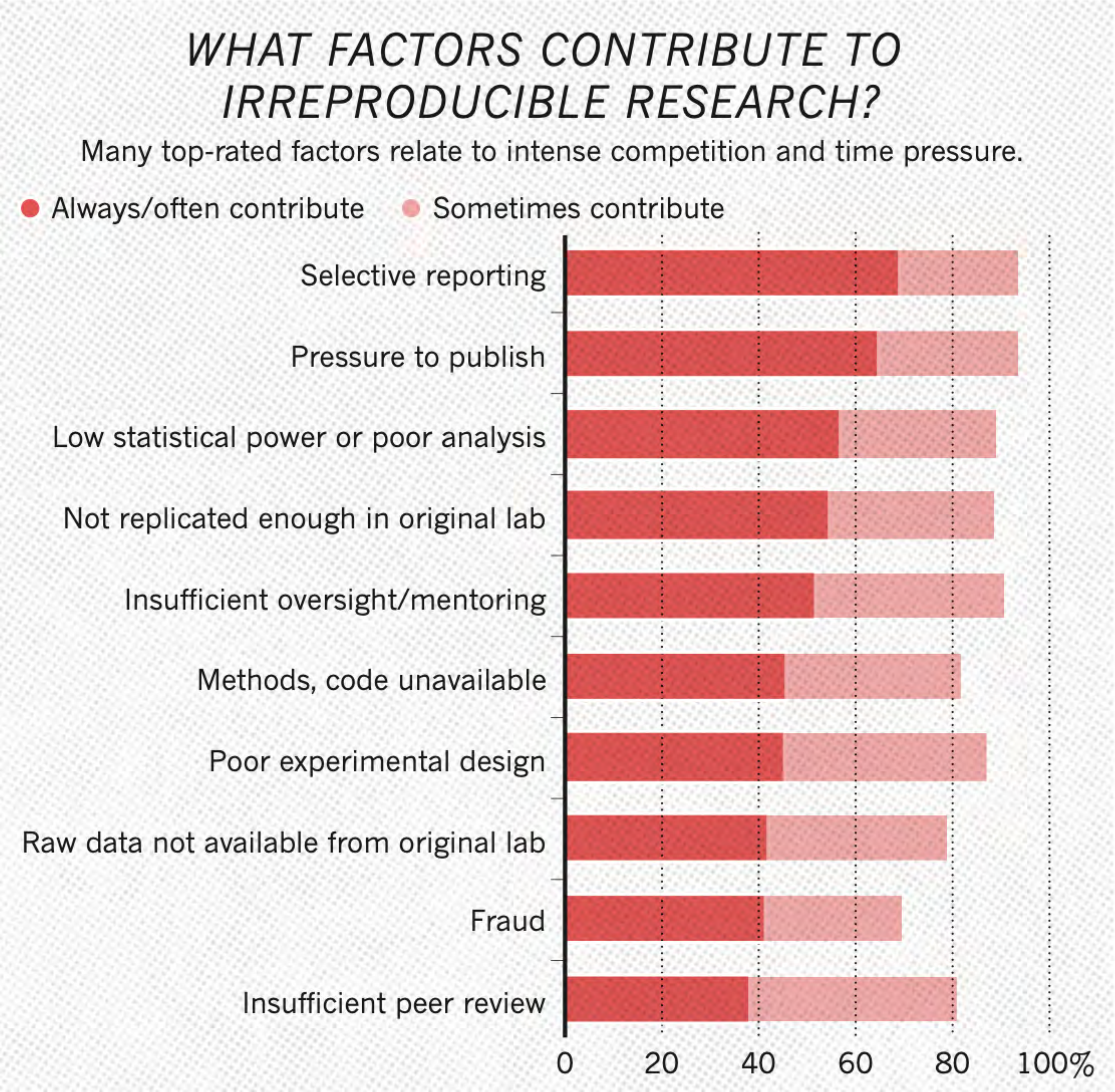
- Recent studies have been unable to recreate results from the past
- Nature Survey
  - More than 70% of researchers have tried and failed to reproduce another scientist's experiments
  - More than 50% have failed to reproduce their own experiments



Potential Causes

Reproducibility Crisis

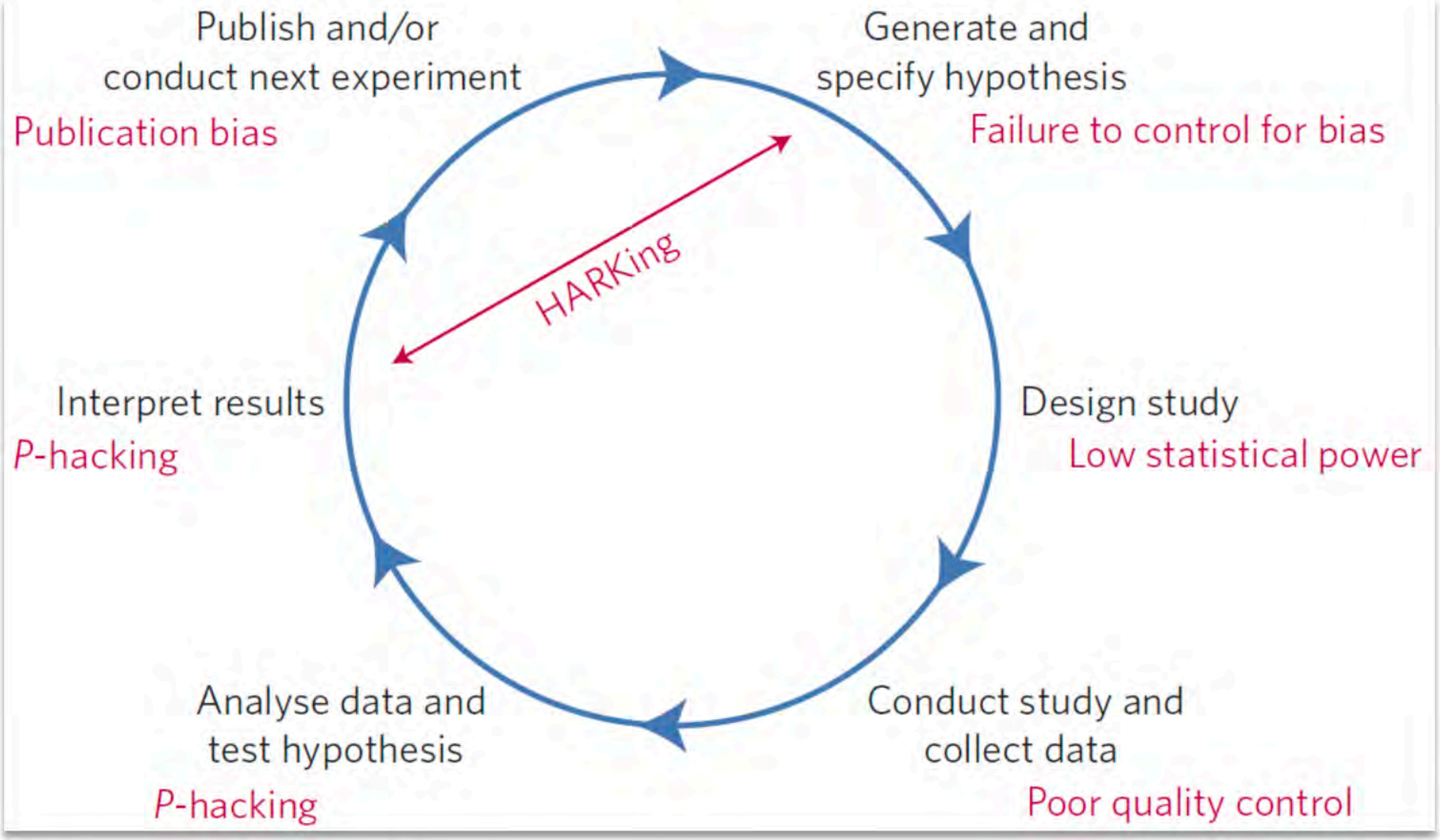
- Reporting Issues
    - Selective reporting: only positive (sexy) results are published
    - Pressure to publish
- Stats problems
    - low stats power (small sample size)
    - poor analysis
- Procedural problems
    - Documentation issues
    - Poor experimental design
    - Data not available
- Not enough time spent on designing good research



Problematic Scientific Approaches
Reproducibility Crisis

- Sample Size too small
  - Variation may simply be from random effects
  - Use a Power Analysis to fix
- P-Hacking
  - If you measure enough things, something will be found to statistically significant
  - p <0.05: means a 1 in 20 probablity
- HARKing
  - “hypothesizing after the results are known”- Creating a new Hypothesis after you get the results
  - Hypothesis now Tailored to the Sample instead previous results
    - increases likelihood that results are not reproducible

Feedforward Cycle of Low Research Quality



Modified from Munafò *et al.*, 2017 *Nature Human Beh* 1: 0021

Rigor and Reproducibility

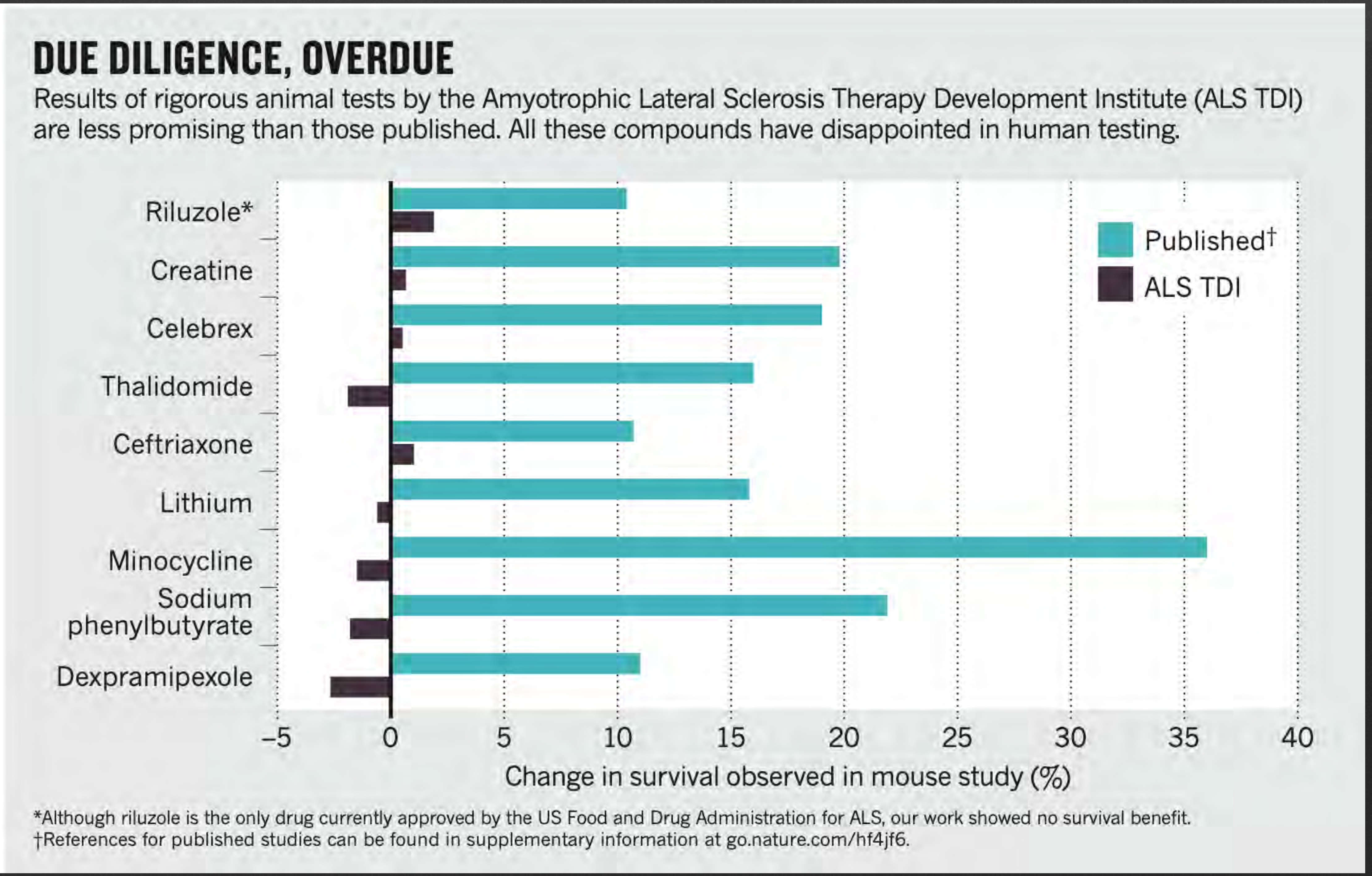
So, what can we do?

Borrowed from San Diego

Call for Transparency

Better reporting of methods

- In 2007, Clinical Trial for a very promising drug for ALS failed
  - NIH went back to review prior mouse studies the motivated the clinical trial
  - Found many procedural errors
    - No Blind testing, studies weren't properly randomized, sex of mice wasn't reported
- When the researchers repeated the experiments with better methods, they failed to reproduce the promising results



Call for Transparency

NIH perspective in Nature

PERSPECTIVE

doi:10.1038/nature11556

A call for transparent reporting to optimize the predictive value of preclinical research

https://www.nature.com/articles/nature11556

- **Randomization**
  - Animals should be assigned randomly to the various experimental groups, and the method of randomization reported.
  - Data should be collected and processed randomly or appropriately blocked.
- **Blinding**
  - Allocation concealment: the investigator should be unaware of the group to which the next animal taken from a cage will be allocated.
  - Blinded conduct of the experiment: animal caretakers and investigators conducting the experiments should be blinded to the allocation sequence.
  - Blinded assessment of outcome: investigators assessing, measuring or quantifying experimental outcomes should be blinded to the intervention.
- **Sample-size estimation**
  - An appropriate sample size should be computed when the study is being designed and the statistical method of computation reported.
  - Statistical methods that take into account multiple evaluations of the data should be used when an interim evaluation is carried out.
- **Data handling**
  - Rules for stopping data collection should be defined in advance.
  - Criteria for inclusion and exclusion of data should be established prospectively.
  - How outliers will be defined and handled should be decided when the experiment is being designed, and any data removed before analysis should be reported.
  - The primary end point should be prospectively selected. If multiple end points are to be assessed, then appropriate statistical corrections should be applied.
  - Investigators should report on data missing because of attrition or exclusion.
  - Pseudo replicate issues need to be considered during study design and analysis.
  - Investigators should report how often a particular experiment was performed and whether results were substantiated by repetition under a range of conditions.

Rigor and Reproducibility in Grants

improving experimental design

- All NIH grants must have section on how the experiments will be properly performed
- Rigor and Reproducibility training is required for all trainees (hence this lecture)
- Your role:
  - be a good trainee
  - Keep good notes — don't delete anything
  - Always ask questions if you don't understand something
  - be consistent in your work
  - Keep Rigor and reproducibility in mind — if something doesn't make sense, ask