Elementary Statistics – Welcome!

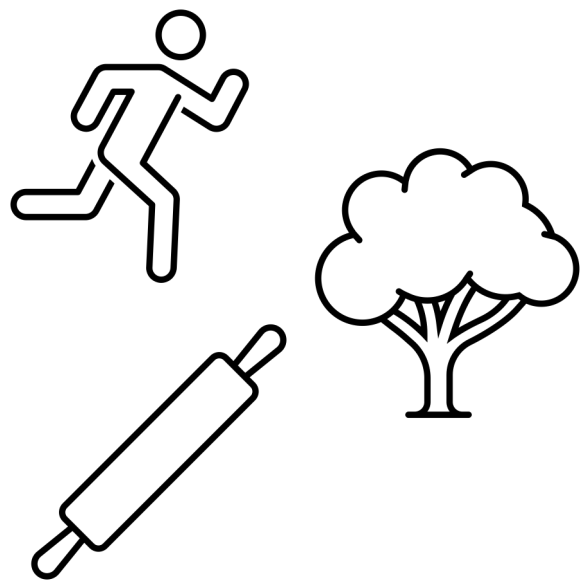
Dr. Ab Mosca (they/them)

Slides based off slides courtesy of Kaitlyn Cook (https://www.smith.edu/people/kaitlyn-cook)

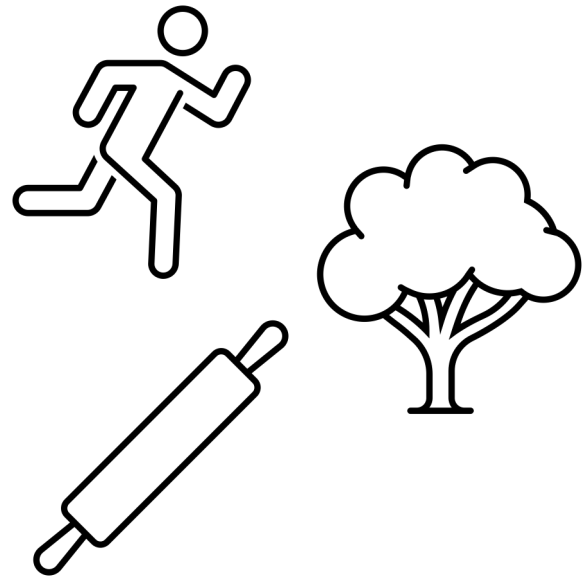
Plan for Today

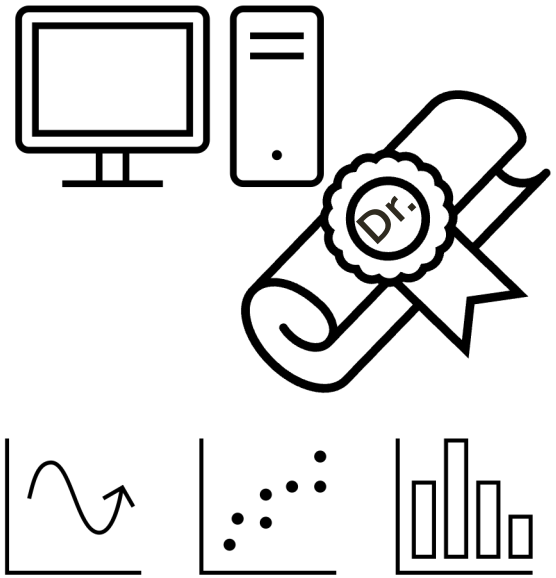
- Who am I?
- Who are you?
- What will we do in this class?
- What are statistics?
- Data vocab

Who Am I?

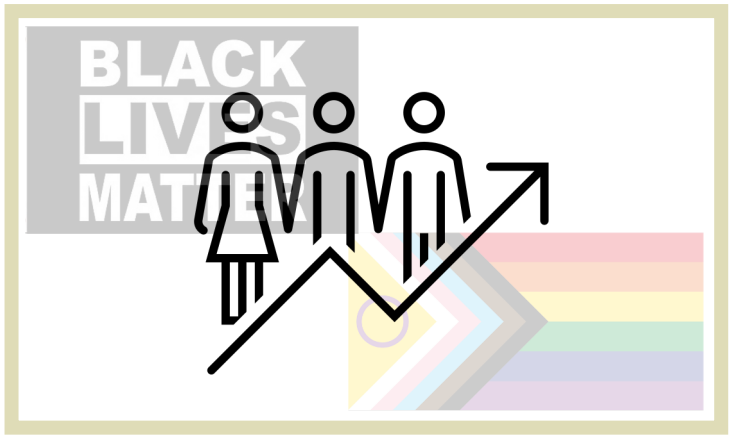


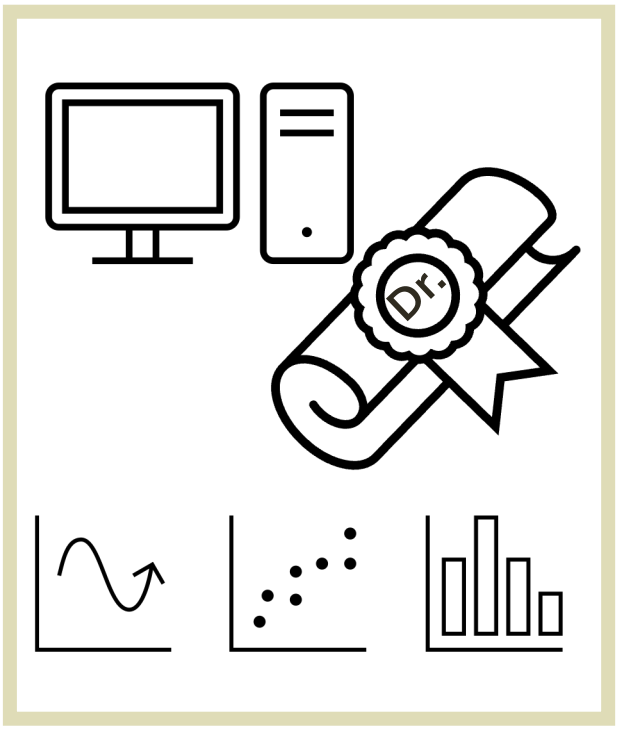
Who Am I?





Who Am I?





Who Are You?

- Form groups of 3
- Introduce yourselves (name, pronouns)
- Share:
  - A highlight of your winter break
- Find 1 thing that your entire group has in common (favorite color? hometown? left-handed? Be creative!)
- After about 5 minutes we will go around, introduce ourselves, and share what each group has in common

Who Are You?

- Form **new groups** of 3 (move around!)
- Introduce yourselves (name, pronouns)
- Share:
  - Would you rather only eat sweet foods for the rest of your life OR only eat savory foods for the rest of your life?
- After about 5 minutes we will go around, introduce ourselves, and share our would you rather answers

Who Are You?

- Form **new new groups** of 3 (move around!)
- Introduce yourselves (name, pronouns)
- Share:
  - Would you rather have a self driving car OR a private jet?
- After about 5 minutes we will go around, introduce ourselves, and share our would you rather answers

What You Will Learn & Logistics

What Is This Class?

- An introduction to statistics
- You will learn to…
  - formulate well-defined research questions;
  - explore data using visualizations and summary statistics, and interpret visualizations and summary statistics of a given study;
  - select and conduct an appropriate analysis for a given research question; and
  - effectively communicate statistical ideas and results, both verbally and in writing

Important Info

- Course website (**write this down**!): https://amosca01.github.io/MATH108-S24/

- Office Hours
  - Wilson Hall 325
    - Wednesday 09:30 - 11:00
    - Thursday 14:30 - 16:30
    - By Appointment

Important Info

- Textbook: *Introduction to Modern Statistics, 1st Edition*
  - See course website for instructions
- Assignments:
  - Turn in on Gradescope – Demo! (https://help.gradescope.com/article/ccbpppziu9-student-submit-work)
- Due Dates: As listed on course schedule.
  - 24hr grace period; no late submissions
  - Lowest homework dropped
  - **See syllabus for revise and resubmit policy**

Important Info

Assignments

- Homework
  - Pair assignments
  - Graded largely on effort
- Quizzes (on PLATO)
  - Individual assignments
  - Can re-take as many times as wanted before deadline
- In-class Activities
  - Graded on effort
- Final Project
  - Small group
  - Graded on creativity and correctness

Important Info

- I’m here to help you succeed
- Please come to office hours or reach out if you need any additional support

Now the good stuff

What is statistics?

On the next slide is an image accompanied by a statement. Take 10 minutes and. . .

→ Think critically about the image, the statement, and their relationship to one another. What do you *notice*? What do you *wonder*?

→ Generate and write down ~10 *open-ended* questions that you have about the statement.

What is statistics?

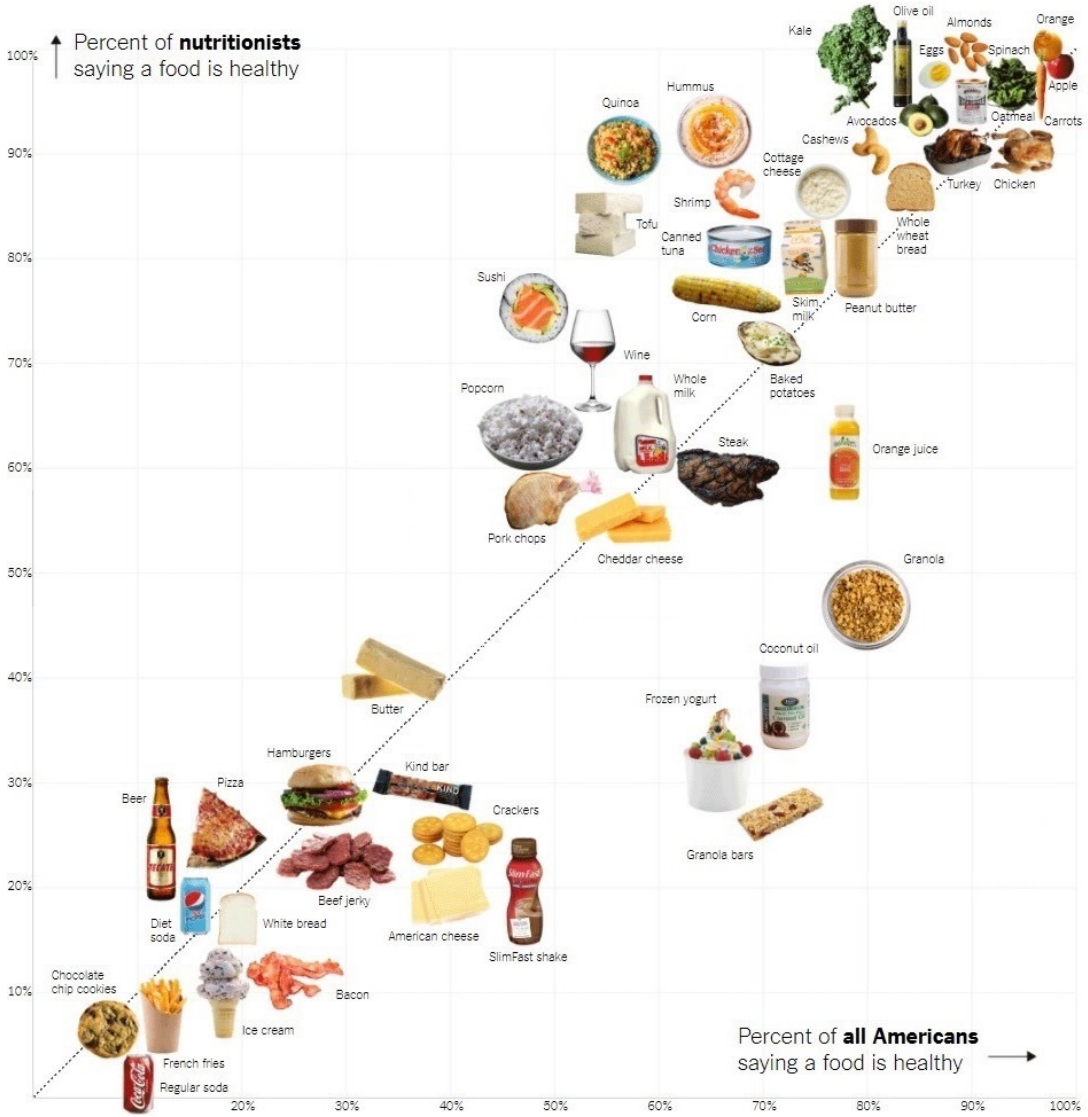


Image courtesy of The Upshot – The New York Times

Americans have a good understanding of which foods are healthy.

What is statistics?

"Statistics: The Art and Science of Learning from Data."
– Alan Agresti and Christine A. Franklin

Distinguished Professor Emeritus of Statistics, University of Florida

Senior Lecture Emeritus in Statistics, University of Georgia

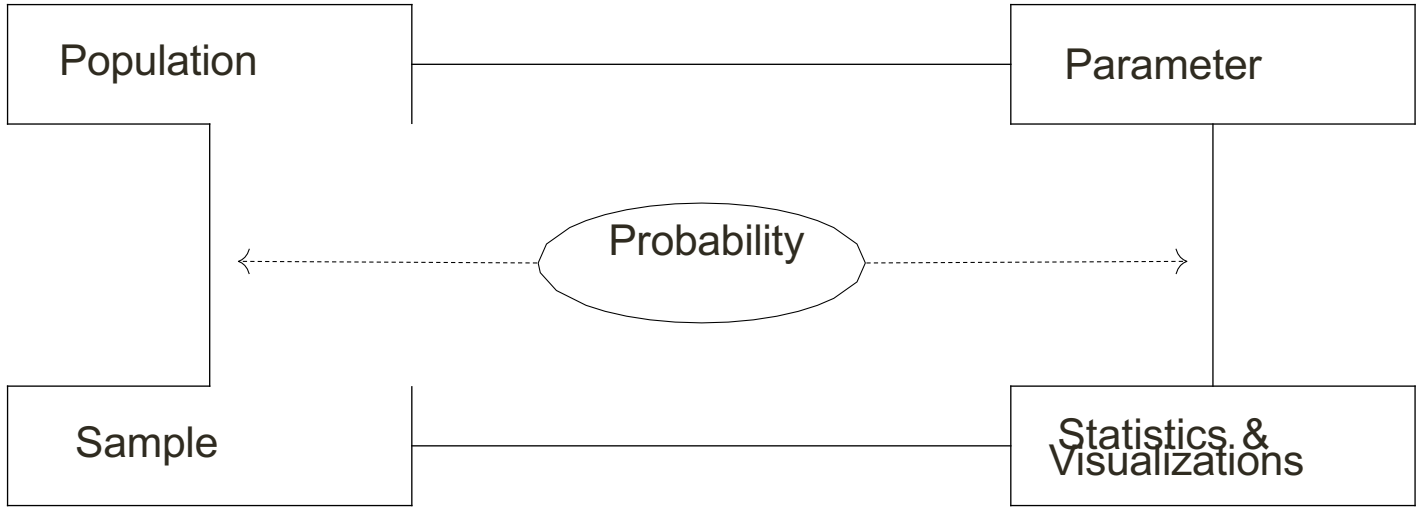
What is statistics?

"Statistics is using data and knowledge about randomness to condense, communicate, and contextualize information and provide insight into the setting from which the data came." – Jo Hardin

Professor of Mathematics and Statistics, Pomona College

Course Overview

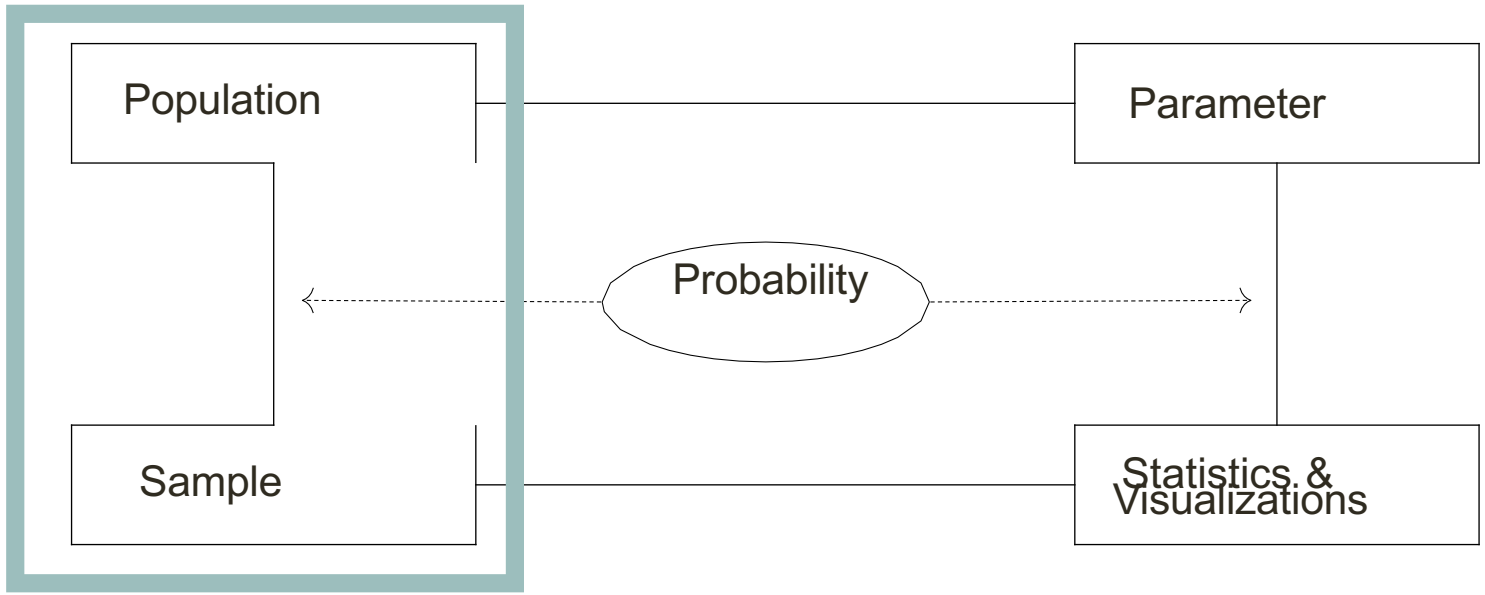
Given a statistical question. . .



- Population: the target group about which we wish to make claims or predictions
- Parameter: numerical summary of the population
- Sample: the data that we have at hand
- Statistic: numerical summary of the sample

Course Overview

Given a statistical question. . .

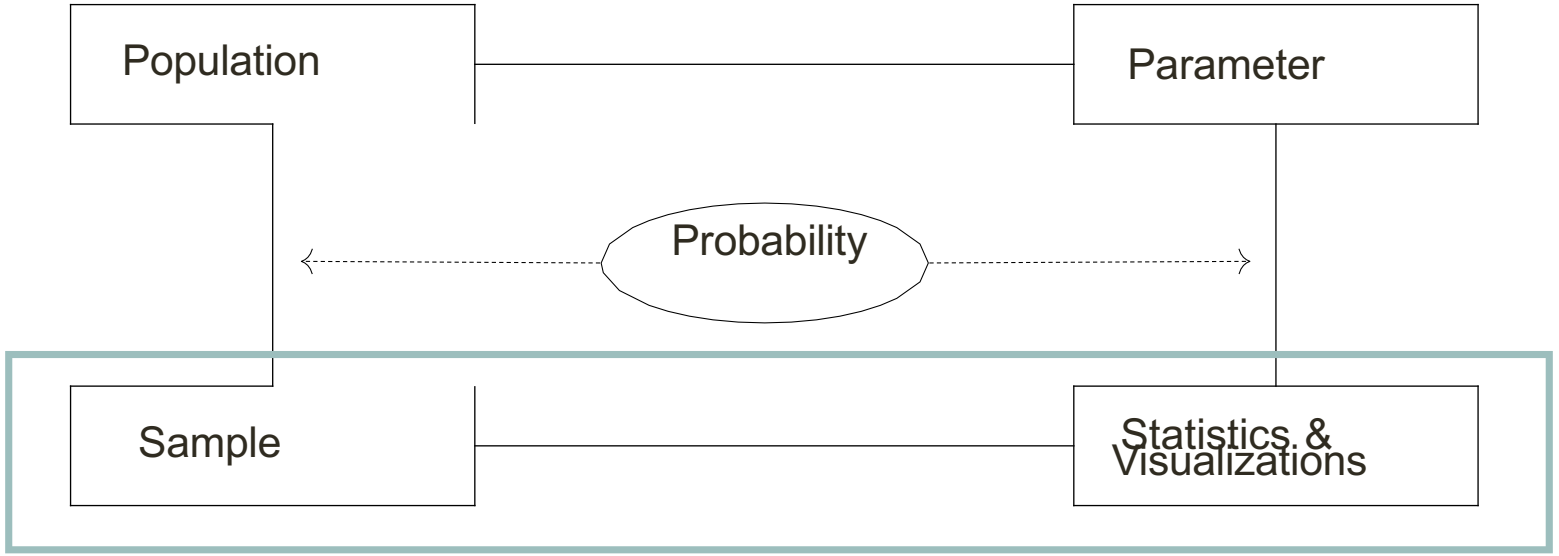


Where do our data come from?

→ Weeks 1 and 2

Course Overview

Given a statistical question. . .

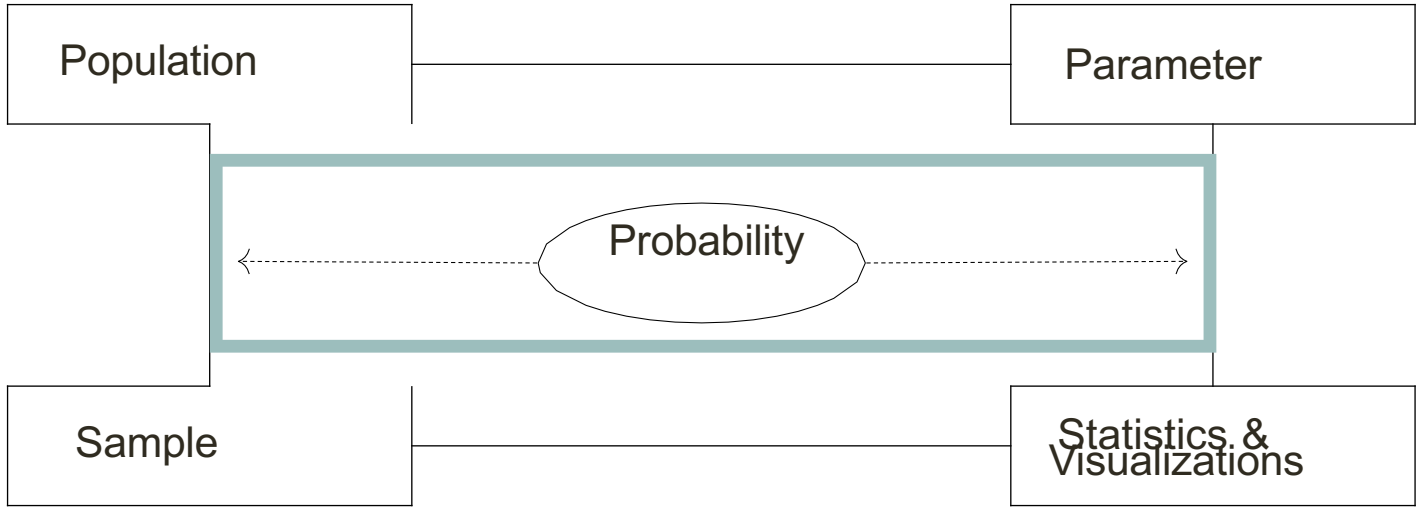


How do we summarize and make sense of all this data (in a way that informs our research question)?

→ Weeks 3, 4, and 5

Course Overview

Given a statistical question. . .

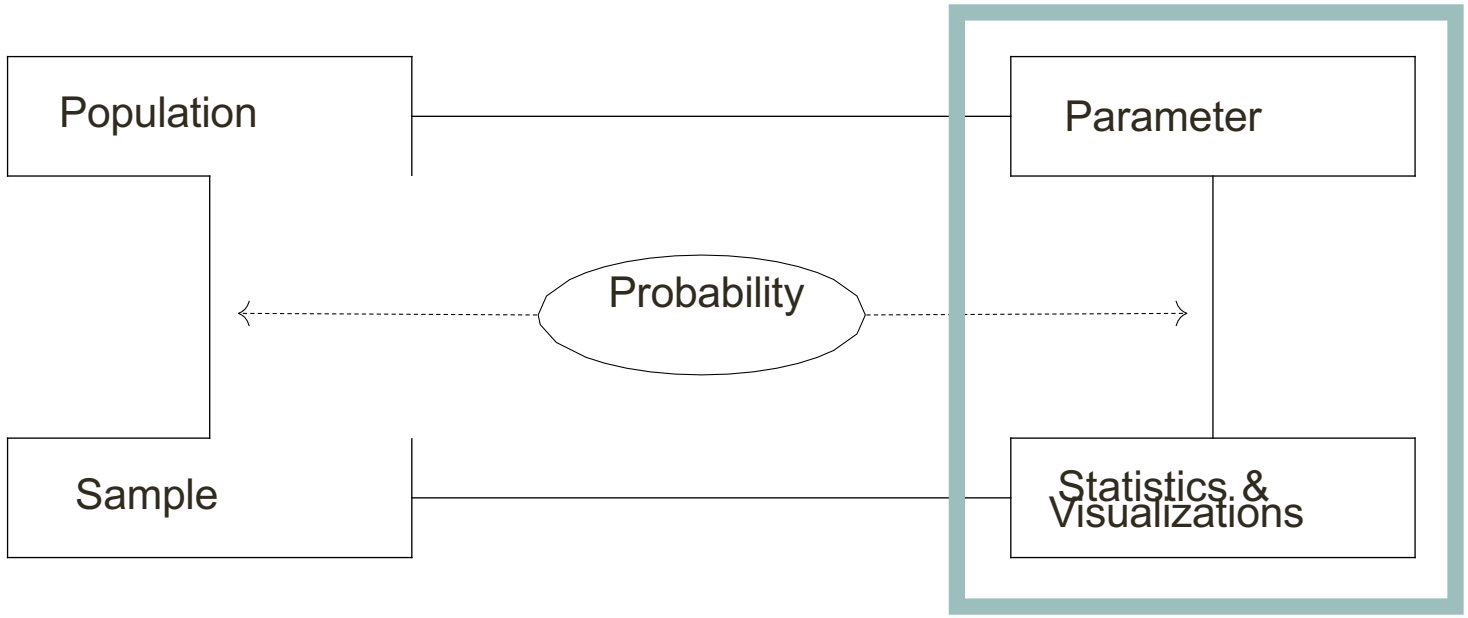


How can we use ideas from mathematics to relate our sample (and sample statistic) back to the population (and parameter of interest)?

→ Weeks 6, 7, and 9

Course Overview

Given a statistical question. . .



How do we use these summaries to draw rigorous and reproducible conclusions about the population?

→ Weeks 9–14

Data: Big Picture

Although our statistical questions are framed in terms of populations and parameters, what we have at our disposal to *answer* those questions is often only a sample

Data: Big Picture

Although our statistical questions are framed in terms of populations and parameters, what we have at our disposal to *answer* those questions is often only a sample

Population		Sample (a.k.a Data)
all likely US voters	→	2000 individuals in a snap political poll
all French paintings from the 1800s	→	French paintings from the 1800s in the Louvre
humanity	→	43,448 individuals in a COVID vaccine trial

Data: Big Picture

Although our statistical questions are framed in terms of populations and parameters, what we have at our disposal to *answer* those questions is often only a sample

Population		Sample (a.k.a Data)
all likely US voters	→	2000 individuals in a snap political poll
all French paintings from the 1800s	→	French paintings from the 1800s in the Louvre
humanity	→	43,448 individuals in a COVID vaccine trial

What does our data look like?

Where does our data come from?

Data: Structure

Observational unit: the person or thing on which measurements are taken

- Often represented by a single row in our dataset
- Ex: a likely US voter, a single French painting from the 1800s, an individual in a clinical trial

	episode	season	episode_num	title	apple_frame	aurora_borealis	barn	beach	boat	bridge
1	S01E01	1	1	A WALK IN THE WOODS	0	0	0	0	0	0
2	S01E02	1	2	MT. MCKINLEY	0	0	0	0	0	0
3	S01E03	1	3	EBONY SUNSET	0	0	0	0	0	0
4	S01E04	1	4	WINTER MIST	0	0	0	0	0	0
5	S01E05	1	5	QUIET STREAM	0	0	0	0	0	0
6	S01E06	1	6	WINTER MOON	0	0	0	0	0	0
7	S01E07	1	7	AUTUMN MOUNTAINS	0	0	0	0	0	0

Data: Structure

Observational unit: the person or thing on which measurements are taken

- Often represented by a single row in our dataset
- Ex: a likely US voter, a single French painting from the 1800s, an individual in a clinical trial

Variable: the characteristic being measured

- Often represented by a column in our dataset
- Ex: preferred political candidate, whether the painting depicts nature as its subject, COVID infection status after 4 weeks

	episode	season	episode_num	title	apple_frame	aurora_borealis	barn	beach	boat	bridge
1	S01E01	1	1	A WALK IN THE WOODS	0	0	0	0	0	0
2	S01E02	1	2	MT. MCKINLEY	0	0	0	0	0	0
3	S01E03	1	3	EBONY SUNSET	0	0	0	0	0	0
4	S01E04	1	4	WINTER MIST	0	0	0	0	0	0
5	S01E05	1	5	QUIET STREAM	0	0	0	0	0	0
6	S01E06	1	6	WINTER MOON	0	0	0	0	0	0
7	S01E07	1	7	AUTUMN MOUNTAINS	0	0	0	0	0	0

Data: Variable Taxonomy

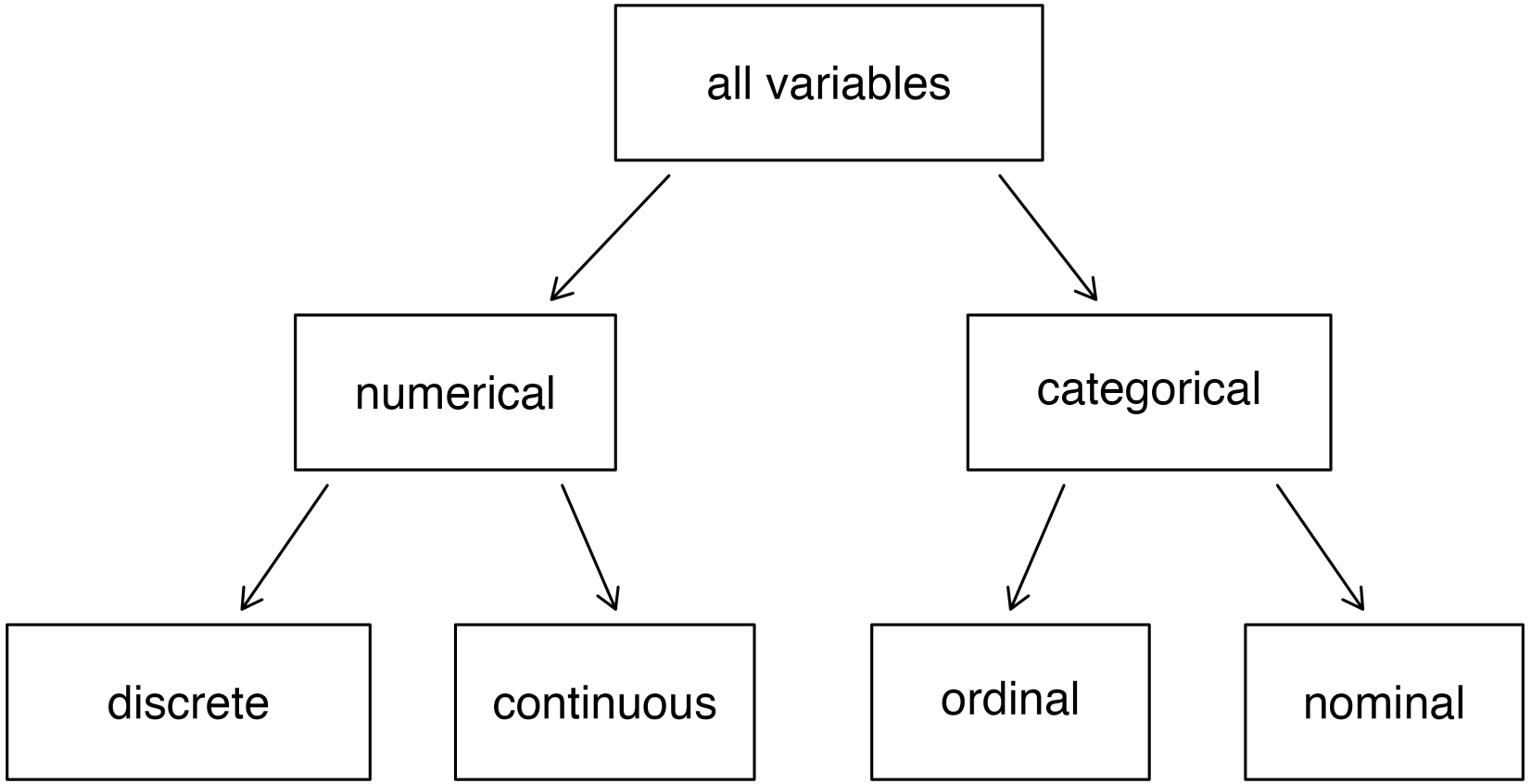
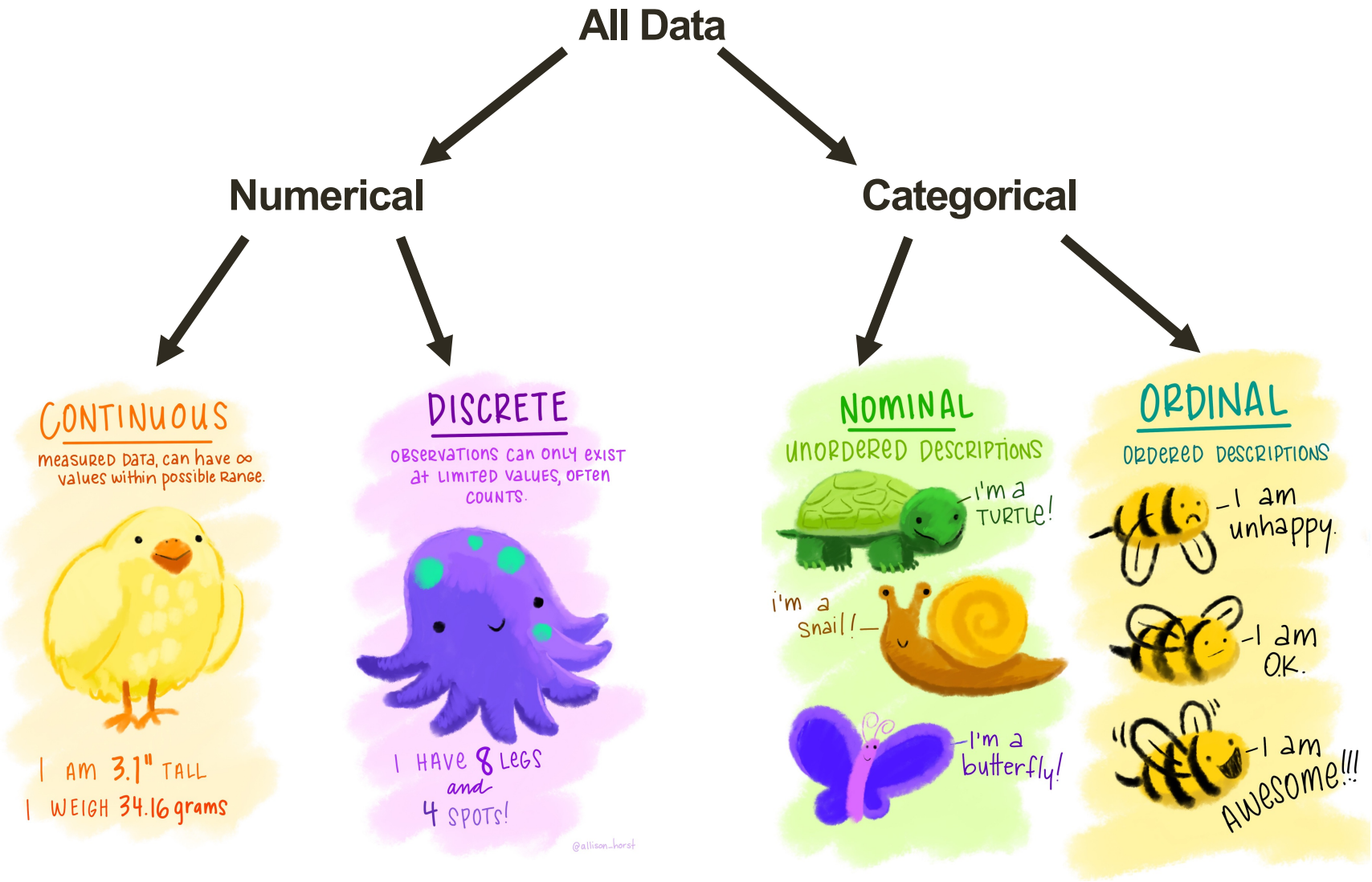


Fig 1.1 from IMS Ch.1: Breakdown of variables into their respective types

Data: Variable Taxonomy



Exercise

Data Collection:

Form a group of three and take turns answering the following questions:

What is your name?

What is your class year?

What is your hometown?

Do you have siblings (1 = Yes, 0 = No)?

What is the furthest away from Westfield (in miles) that you were over winter break?

While one group member answers, the other two will write down the answers. Then open the data collection Google spreadsheet (http://bit.ly/48ceWSa) and start entering the data!

Exercise

Identifying Variable Types:

Take a couple minutes on your own to answer the following:

- What is the observational unit in our dataset?
- For each variable in the spreadsheet, is it discrete, continuous, ordinal, or nominal?

Exercise

Identifying Variable Types:

Take a couple minutes on your own to answer the following:

- What is the observational unit in our dataset?
- For each variable in the spreadsheet, is it discrete, continuous, ordinal, or nominal?

Discuss your answers with your neighbor!

Exercise

What type of variable is "Class Year"?

- Discrete
- Continuous
- Ordinal
- Nominal

Exercise

What type of variable is "Class Year"?

- Discrete
- Continuous
- Ordinal
- Nominal

Exercise

What type of variable is “Siblings”?

- Discrete
- Continuous
- Ordinal
- Nominal

Exercise

What type of variable is “Siblings”?

- Discrete
- Continuous
- Ordinal
- Nominal

Exercise

What type of variable is "Distance from Westfield over Break" (Miles-Away)?

- Discrete
- Continuous
- Ordinal
- Nominal

Exercise

What type of variable is "Distance from Westfield over Break" (Miles-Away)?

- Discrete
- Continuous
- Ordinal
- Nominal