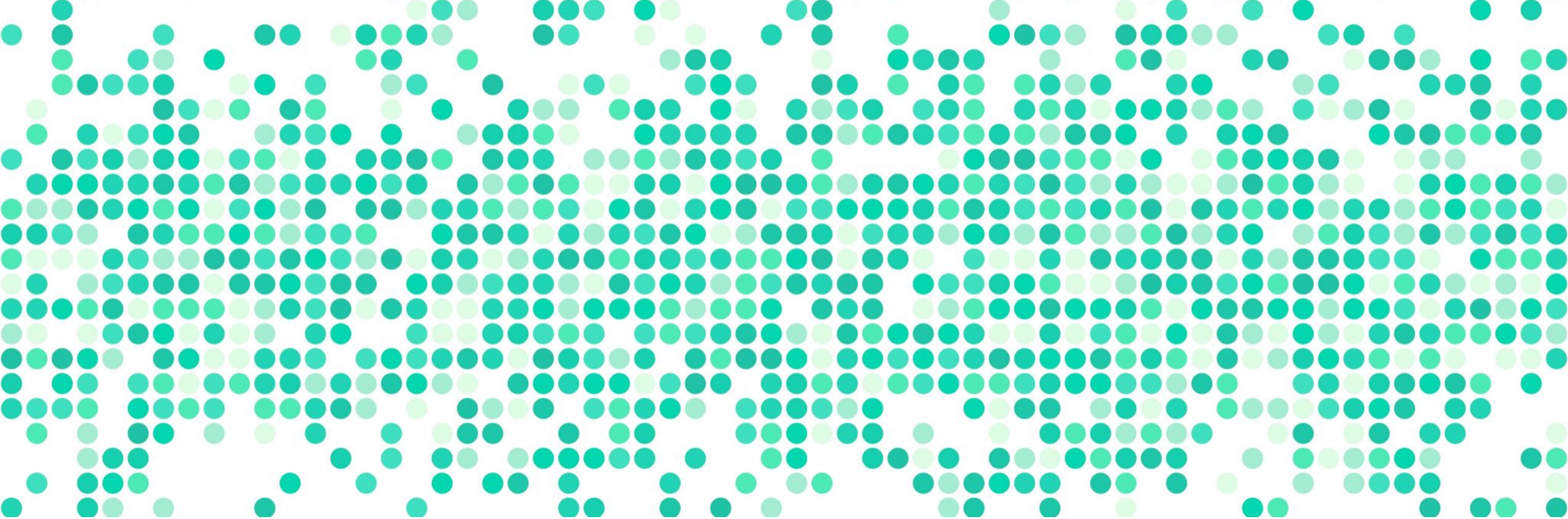
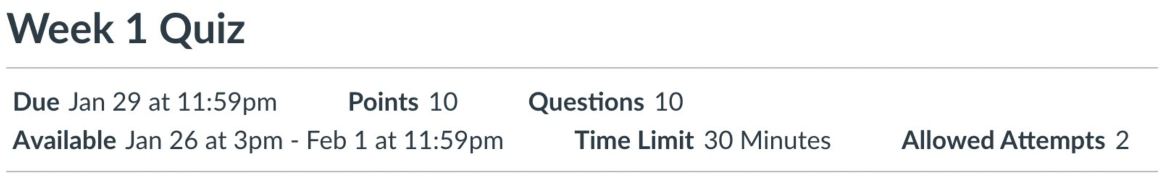


DATA ANALYSIS

Week 1: Statistical Thinking / What are data?

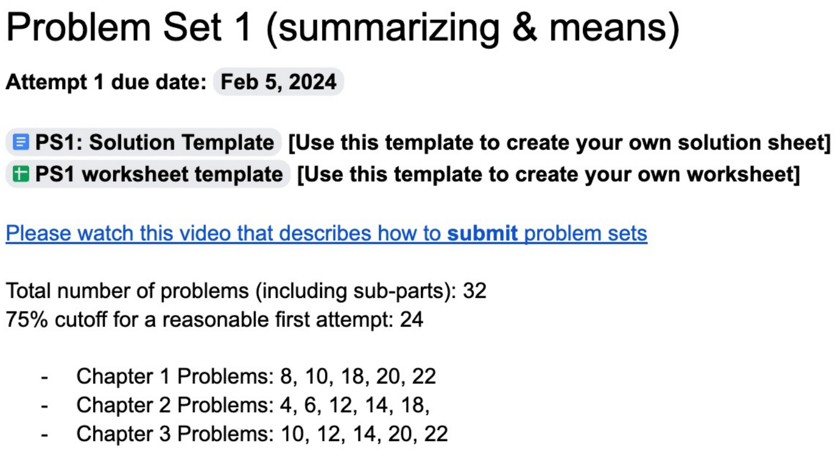
logistics



Week 1 Quiz

Due Jan 29 at 11:59pm **Points** 10 **Questions** 10
Available Jan 26 at 3pm - Feb 1 at 11:59pm **Time Limit** 30 Minutes **Allowed Attempts** 2

- revised quiz policy
    - quizzes will now remain open from Fridays at 3 pm
    - now due on Monday but they will be available until Thursday, 11.59 pm to incorporate flex days (3 max)
- pre-class survey
    - fill out by end of this week, you can still get extra credit!
    - and learn about your attitudes towards statistics!
- problem set submission video updated
- AI policy: use at your own risk!



Problem Set 1 (summarizing & means)

Attempt 1 due date: Feb 5, 2024

PS1: Solution Template [Use this template to create your own solution sheet]
PS1 worksheet template [Use this template to create your own worksheet]

Please watch this video that describes how to **submit** problem sets

Total number of problems (including sub-parts): 32
75% cutoff for a reasonable first attempt: 24

- Chapter 1 Problems: 8, 10, 18, 20, 22
- Chapter 2 Problems: 4, 6, 12, 14, 18,
- Chapter 3 Problems: 10, 12, 14, 20, 22

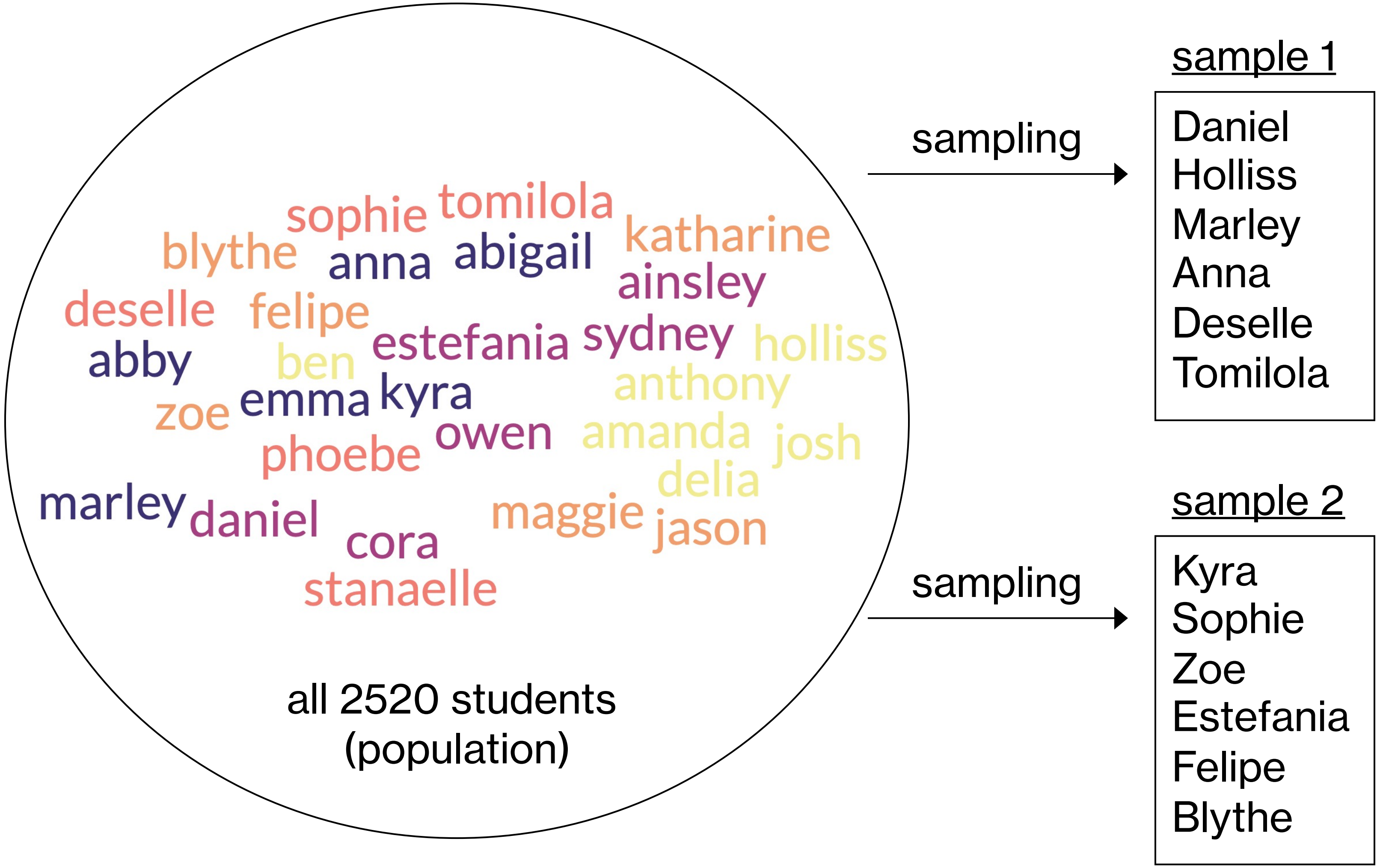
today's agenda

- introduce statistical thinking
- define population / sample / data
- discuss scales of measurement / reliability

what is statistical thinking?

- understanding the complex world in simple terms
  - summarization + uncertainty
- different from other forms of thinking, e.g., human intuition, heuristics, etc.
- three key uses: describe (the world), decide (something), predict (something)
- key concepts:
  - learning from data: we let the data guide us
  - aggregation: we “summarize” raw data
  - uncertainty: we assess how well our raw data maps on to the summarization
  - sampling: we acknowledge that our data are samples from a population

populations and samples



sophie tomilola
blythe anna abigail katharine
ainsley
deselle felipe
abby ben estefania sydney holliss
anthony
zoe emma kyra
owen amanda josh
phoebe
delia
marley daniel maggie jason
cora
stanaelle

all 2520 students
(population)

sampling

sampling

sample 1

Daniel
Holliss
Marley
Anna
Deselle
Tomilola

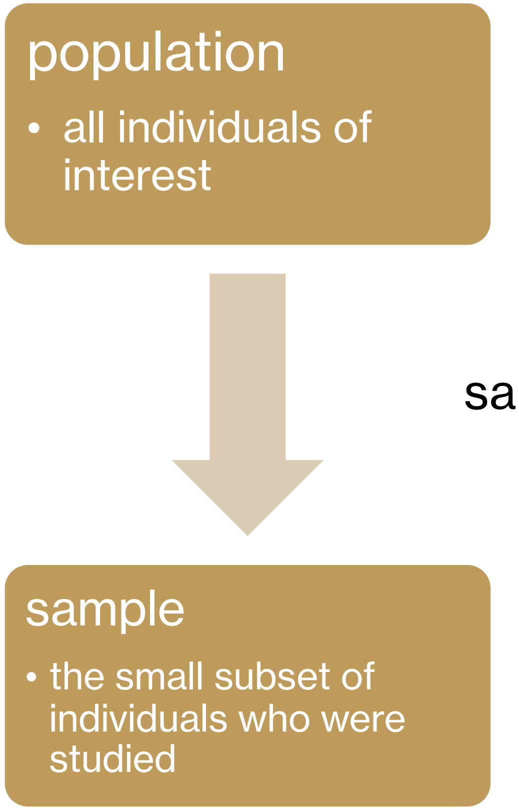
sample 2

Kyra
Sophie
Zoe
Estefania
Felipe
Blythe

samples should be

- representative
- generalizable

populations and samples



population

- all individuals of interest

sample

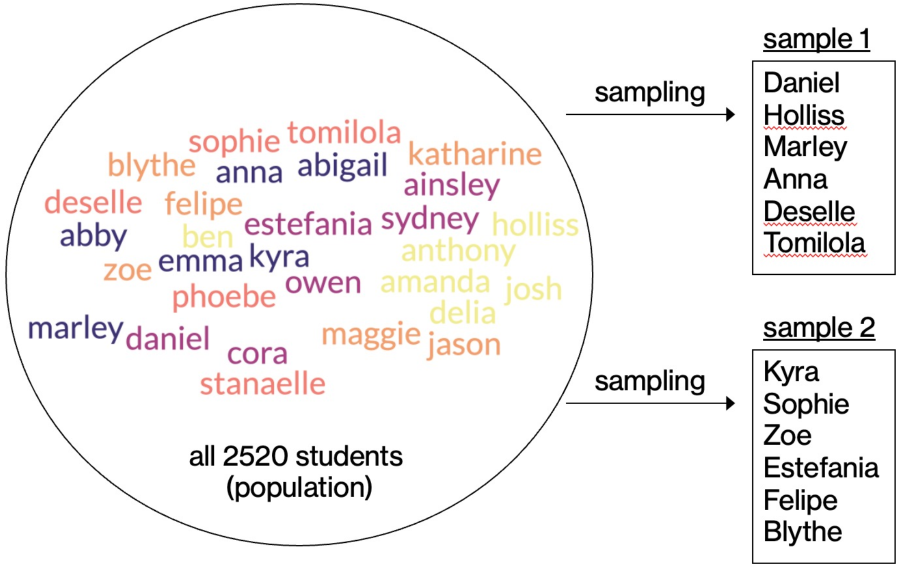
- the small subset of individuals who were studied

results from samples are generalized to populations

samples come from populations

parameters, statistics, sampling error

- parameter: something that describes a population
- statistic: something that describes a sample
- sampling error: the discrepancy between the sample statistic and the true population parameter it is estimating
- to reduce sampling error:
  - use a sufficiently large sample
  - use random selection: selecting individuals from the population at random for your sample to create an unbiased sample



all 2520 students
(population)

sampling

sampling

sample 1

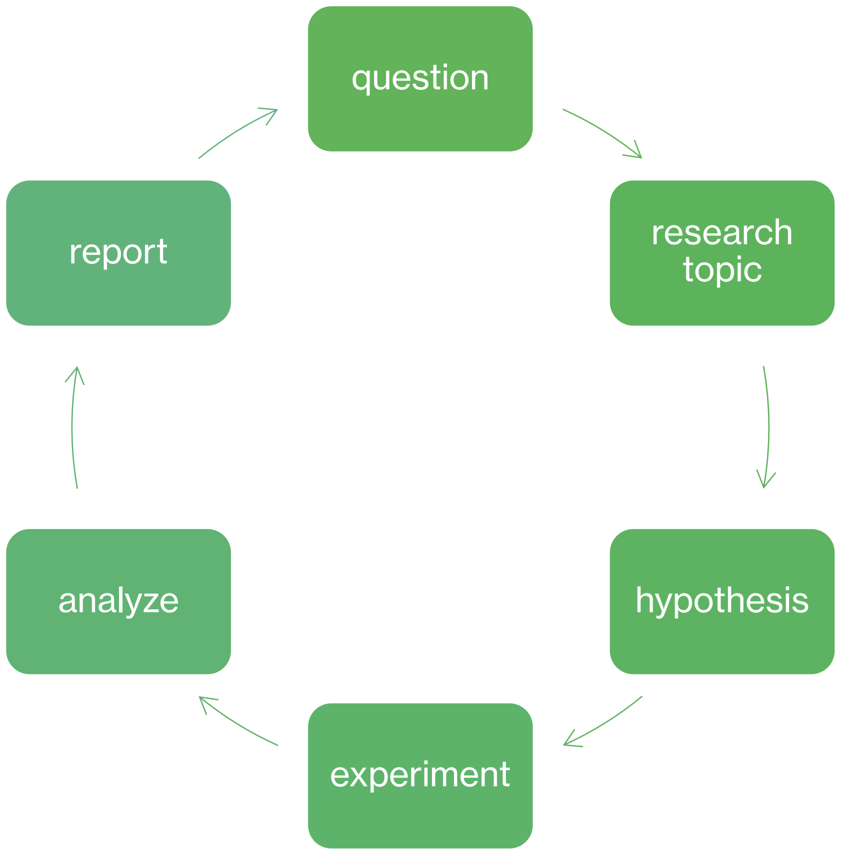
Daniel
Holliss
Marley
Anna
Deselle
Tomilola

sample 2

Kyra
Sophie
Zoe
Estefania
Felipe
Blythe

the scientific method

- the scientific method is a method for acquiring knowledge by making predictions, carrying out experiments to test those predictions, and making inferences based on the observed outcomes
- variables and constants
  - variable: a characteristic that changes across conditions
  - constant: a characteristic that is fixed across conditions
- to make inferences, we manipulate a variable of interest, and observe the effect on an outcome variable, holding all other variables constant



samples in research

experimental research

- test a manipulation to establish a cause-and-effect relationship between two variables

non-experimental research

- quasi-experimental research
  - no actual manipulation, groups/variables defined due to natural variations
- descriptive research
  - single or collection of variables are observed and summarized
- correlational research
  - at least two variables are observed to determine a relationship

research terminology

- independent variable (what is being manipulated?)
    - levels denote the types of “conditions” that a participant could be assigned to
- dependent variable (what is being measured?)
- design type (within- or between-subjects/participants)
    - were all participants exposed to all levels of the independent variable?
- key ideas for controlling other extraneous variables:
    - random assignment
    - matching/holding constant
    - control conditions

activity (think-pair-share)

- a research scenario will be presented
- think about your answers
- pair up and discuss your answers
- share out

scenario #1

- A researcher is testing the effect of alcohol on memory performance. He gives one group of subjects a bottle of vodka, and another a nonalcoholic substance that tastes like vodka. Each group then learns a list of words, and attempts to recall them. Number of words correctly recalled for each group is recorded
  - what kind of study is it (experimental / non-experimental)?
  - independent and dependent variables?
  - design type (within- or between-participant)?
  - what would the data look like?

scenario #2

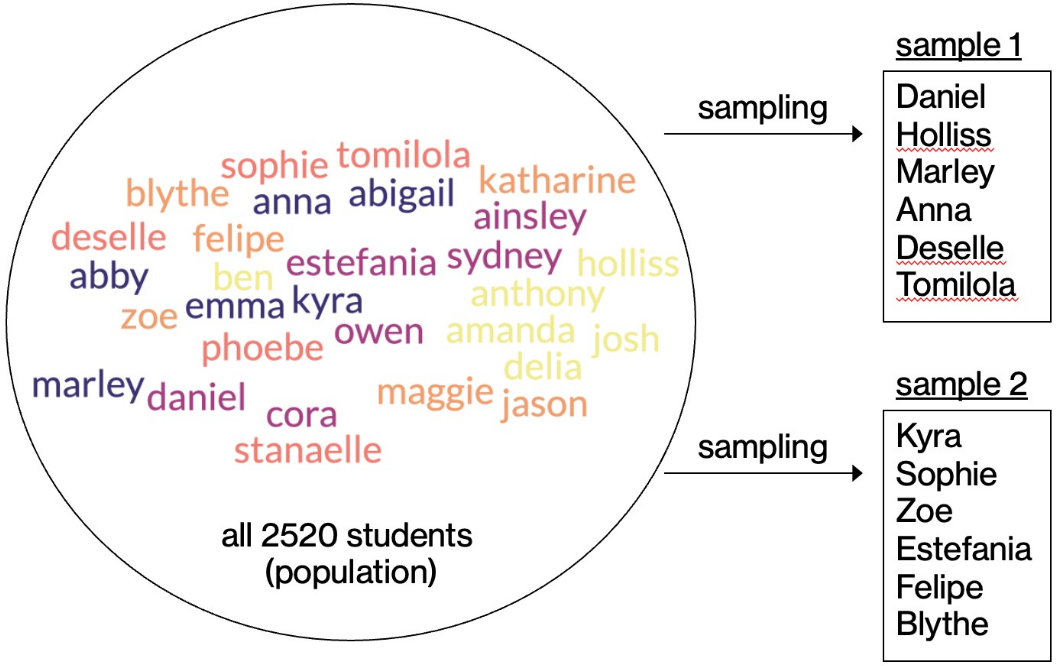
- A social psychologist is interested in gender differences in math performance. She randomly selects men and women from Bowdoin and has them solve a series of equations. Number of equations correctly solved for each participant is recorded.
    - what kind of study is it (experimental / non-experimental)?
    - independent and dependent variables?
    - design type (within- or between-participant)?
    - what would the data look like?

scenario #3

- A clinical psychologist is interested in the effectiveness of a new anti-depression drug. He collects depression scores from a group of individuals diagnosed with depression at time 1. All individuals then take the drug, and are measured again a month later at time 2.
  - what kind of study is it (experimental / non-experimental)?
  - independent and dependent variables?
  - design type (within- or between-participant)?
  - what would the data look like?

from samples to data

- samples provide us with information
- data ***are*** measurements or observations obtained from a sample
    - a dataset is a collection of measurements or observations
    - a datum is a single measurement or observation



all 2520 students
(population)

sampling

sampling

sample 1

Daniel
Holliss
Marley
Anna
Deselle
Tomilola

sample 2

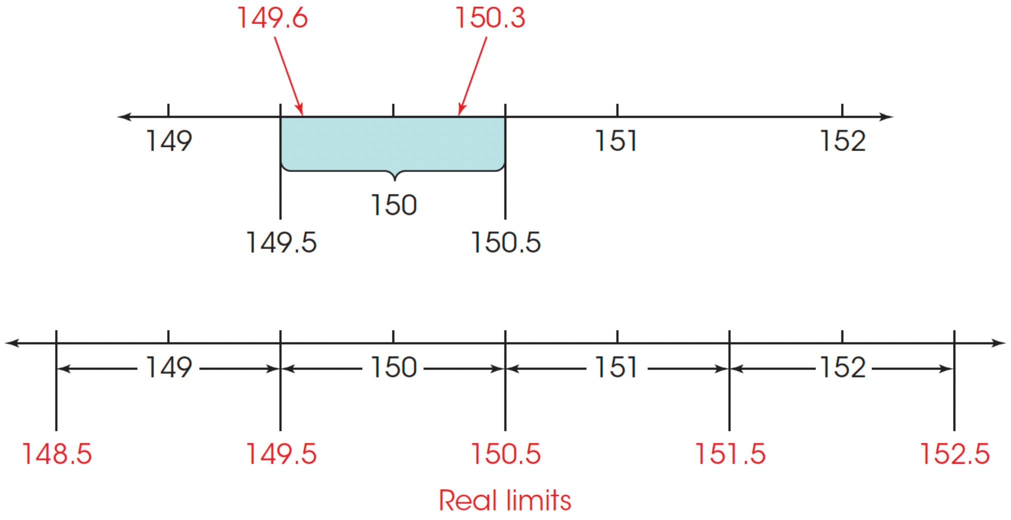
Kyra
Sophie
Zoe
Estefania
Felipe
Blythe

scales of measurement

- data can be measured in several ways:
  - qualitative (put things into categories) vs. quantitative (assign numbers) data
  - discrete: separate, indivisible values. no values can exist between two neighboring values; integer scales
  - continuous: an infinite number of possible values fall between any two observed values. hypothetically divisible into an infinite number of fractional parts.
- how data are measured determines:
  - what kinds of mathematical operations can be applied
  - what kind of statistical computations can be computed

real limits for continuous data

- only applies to continuous data
- the real limit separates two adjacent scores, and is located halfway between the scores
  - each score has an upper real limit (UL) and a lower real limit (LL)
  - lower limit for 150 is 149.5; upper limit is 150.5



scales of measurement

NOIR	each value has a unique meaning	a value has a sense of quantity, some values are larger, some are smaller	units along the scale of measurement are equal to one another	the scale has a true meaningful zero point
	identity	**magnitude**	**equal intervals**	**absolute zero**
nominal				
ordinal				
interval				
ratio				

scales of measurement

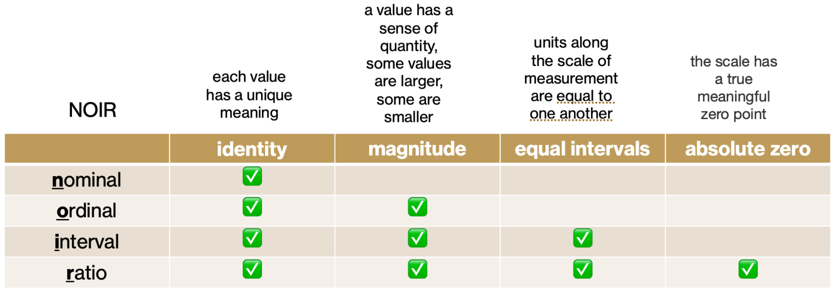
NOIR	each value has a unique meaning	a value has a sense of quantity, some values are larger, some are smaller	units along the scale of measurement are equal to one another	the scale has a true meaningful zero point
	identity	**magnitude**	**equal intervals**	**absolute zero**
nominal	✅			
ordinal	✅	✅		
interval	✅	✅	✅	
ratio	✅	✅	✅	✅

NOIR	each value has a unique meaning	a value has a sense of quantity, some values are larger, some are smaller	units along the scale of measurement are equal to one another	the scale has a true meaningful zero point
	identity	magnitude	equal intervals	absolute zero
nominal	✅			
ordinal	✅	✅		
interval	✅	✅	✅	
ratio	✅	✅	✅	✅

activity

- assign a data type to each variable (NOIR) and whether it is discrete / continuous

variable	NOIR	discrete/continuous
numbers on basketball jerseys		
sizes of Starbucks orders		
weight		
calendar years		
IQ scores		



NOIR	each value has a unique meaning	a value has a sense of quantity, some values are larger, some are smaller	units along the scale of measurement are equal to one another	the scale has a true meaningful zero point
	identity	**magnitude**	**equal intervals**	**absolute zero**
nominal	✓			
ordinal	✓	✓		
interval	✓	✓	✓	
ratio	✓	✓	✓	✓

activity

- assign a data type to each variable (NOIR) and whether it is discrete / continuous

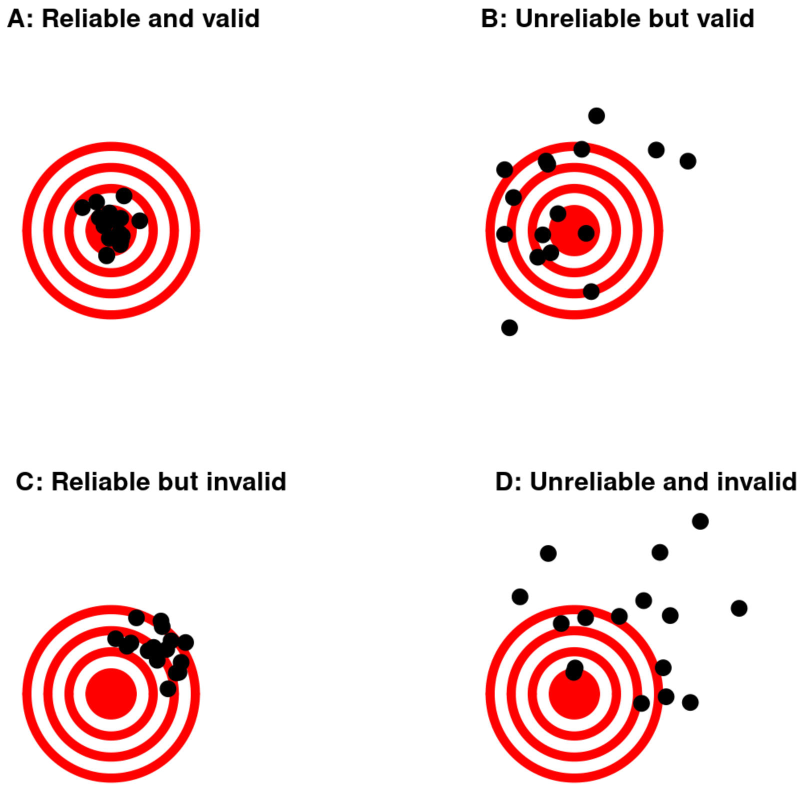
variable	NOIR	discrete/continuous
numbers on basketball jerseys	nominal	discrete
sizes of Starbucks orders	ordinal	discrete
weight	ratio	continuous
calendar years	interval	continuous
IQ scores	interval	continuous

data in scientific abstracts

- table groups
- go to the abstract document and read over the abstract
- make note of (you will need to make a copy to edit the document):
  - independent variable(s) and data type(s)
  - dependent variable(s) and data type(s)
- predicted graph of results?
- key takeaway?

reliability and validity

- reliability: consistency of measurements
  - test-retest reliability
  - inter-rater reliability
- validity: are we measuring what we think we are measuring?
  - ***face*** validity: reality check, does it make sense?
  - ***construct*** validity: is it related to other measurements in a logical manner? convergent vs. divergent validity
  - ***predictive*** validity: can it predict future data?

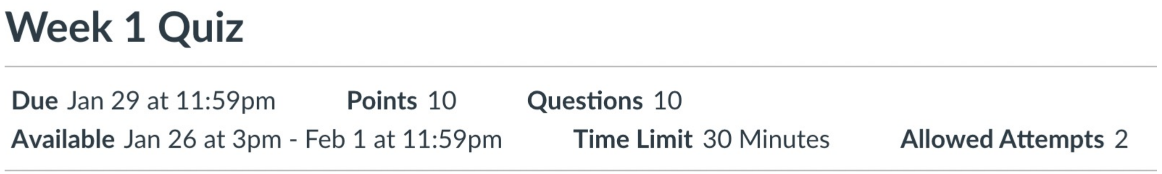


big takeaways from today

- jot down the key takeaways from today **<u>without</u>** looking at the slides/notes someplace you can revisit
- retrieval practice + elaborative encoding

next time

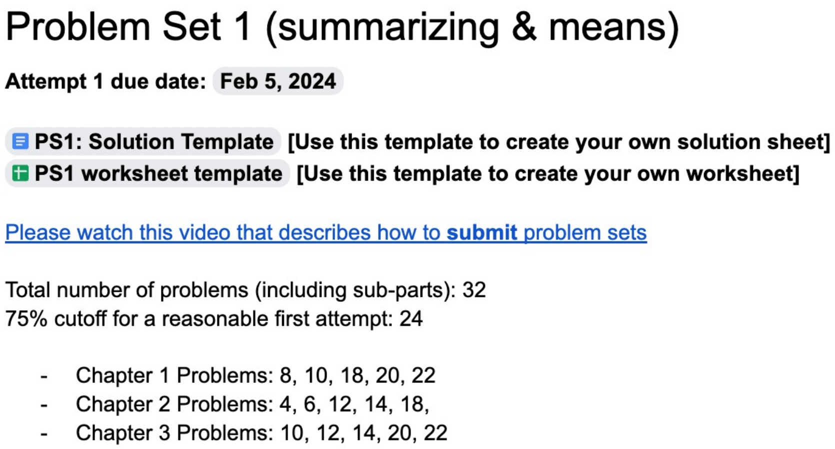
- **before** class
  - *try*: week 1 quiz
  - *apply*: problem set #1 (chapter 1 problems)
  - *apply*: optional meme / discussion board post
  - *prep*: Chapter 2/3 from textbook + videos
- **during** class
  - why/how do we summarize data?
  - how do we “explain” data?



Week 1 Quiz

Due Jan 29 at 11:59pm **Points** 10 **Questions** 10

Available Jan 26 at 3pm - Feb 1 at 11:59pm **Time Limit** 30 Minutes **Allowed Attempts** 2



Problem Set 1 (summarizing & means)

Attempt 1 due date: Feb 5, 2024

PS1: Solution Template [Use this template to create your own solution sheet]
PS1 worksheet template [Use this template to create your own worksheet]

Please watch this video that describes how to **submit** problem sets

Total number of problems (including sub-parts): 32
75% cutoff for a reasonable first attempt: 24

- Chapter 1 Problems: 8, 10, 18, 20, 22
- Chapter 2 Problems: 4, 6, 12, 14, 18,
- Chapter 3 Problems: 10, 12, 14, 20, 22