

SWK: 8408 Statistics I for Social Work

Gia Elise Barboza-Salerno, MA, MS, JD, PhD
barboza-salerno.1@osu.edu

The Ohio State University
Colleges of Public Health & Social Work

Autumn 2023
[Class Website](#)
Week 2

Table of Contents

1. Review from last class
2. Overview & Terminology
 - Overview of scientific method
 - Epistemology & Standpoint Epistemology
 - Observational Biases and Re-framing
 - Variables & Measurement
 - Sampling & Measurement
3. Application & Lab

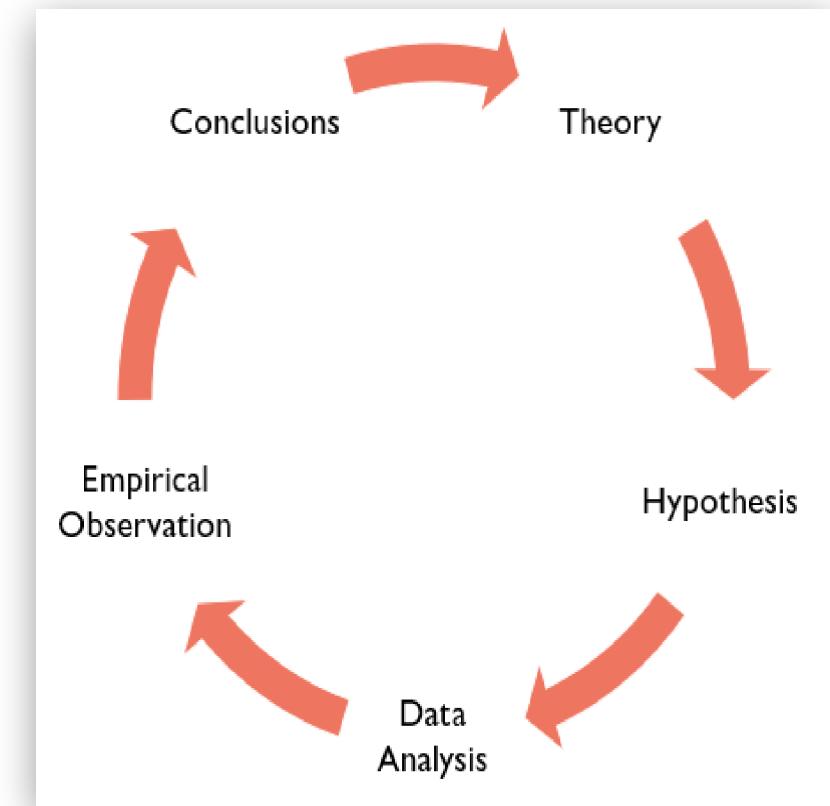
Motivation: Why Study

1. To be able to effectively conduct (empirical) research
2. To be an informed “consumer”
3. To further develop critical and analytic thinking skills



Introduction

1. Characteristics: explain, predict, describe
2. At all stages, there are rules that establish reliability and validity
 - Sampling, measurement, data analysis and interpretation



Starting Point: Observation

Individuals' knowledge about social, political and economic life is “minimal.”

“Esse est percipi” – To be is to be perceived which is to say the so-called objective world is all in the mind

Our only knowledge of this world is what comes to us through our senses

There are serious sense data interpretation problems, as illustrated by my joke

Human “intuition” is based on our understanding of the world...

Factors that affect observation and "intuition"

- selective perception
- cognitive dissonance
- framing

Examples

- [Optical illusions show how we see | Beau Lotto - YouTube](#)
- [Movie Perception Test - YouTube](#)

A brief overview : Observation

Epistemology: The study of what we think we know and how we come to know it.

Neither do the ignorant love wisdom or desire to become wise; for this is the grievous thing about ignorance, that those who are neither good ... nor sensible think they are good enough, and do not desire that which they do not think they are lacking. Plato, Symposium 203E-204A

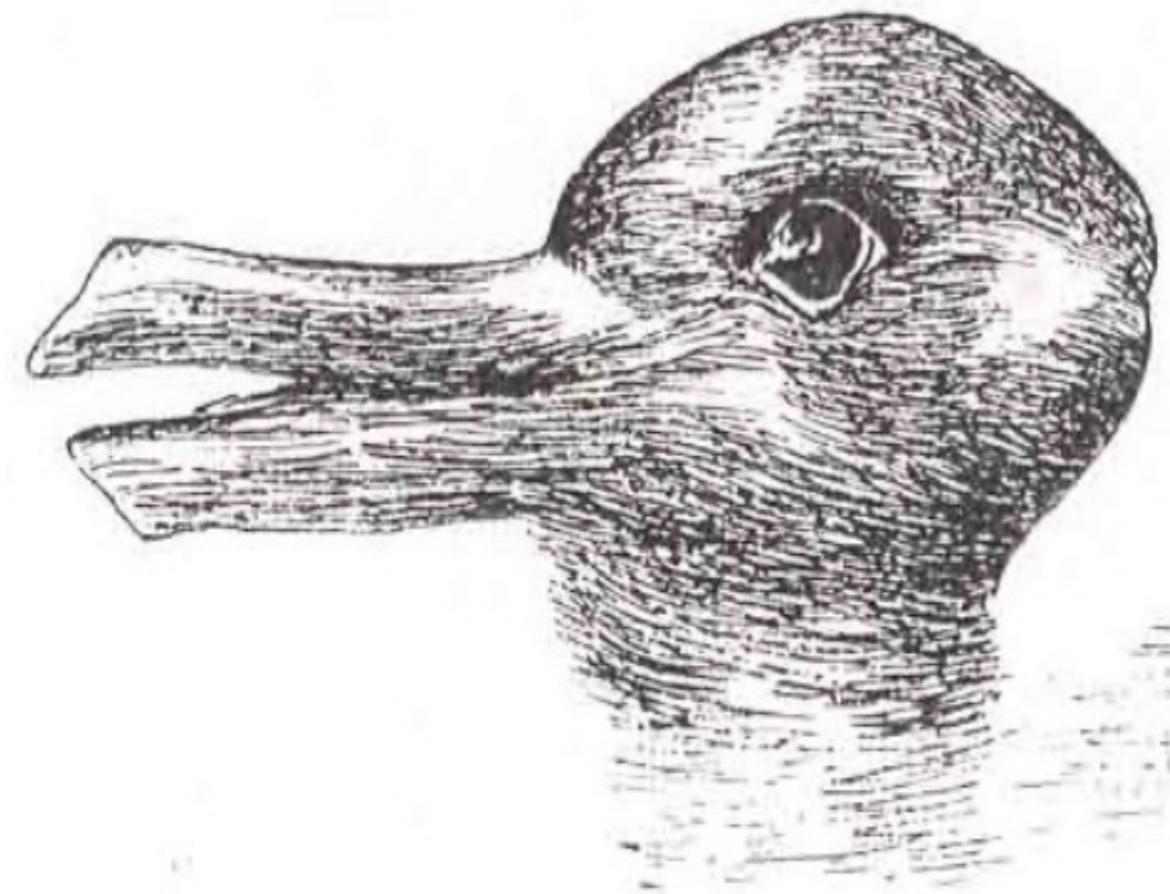
- What does this say about the relationship between ignorance and knowledge?
- Stop, pause, reflect.... always about the statistical analyses and interpretation you do!



**How do you
know what you
know?**

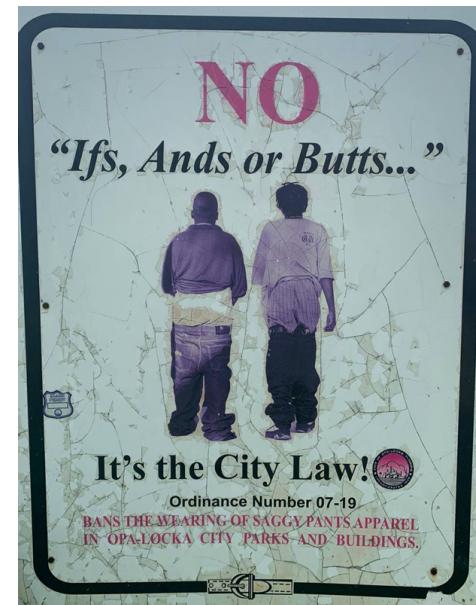
What do you see?

**Do you see what
you expect to
see?**



Example: Standpoint

In support of the law, Commissioner Alvin Burke argued that the law was not intended "to target our young black men" ...but rather to "uplift our young black men."



Alabama Official: God Told Me to Ban Saggy Pants

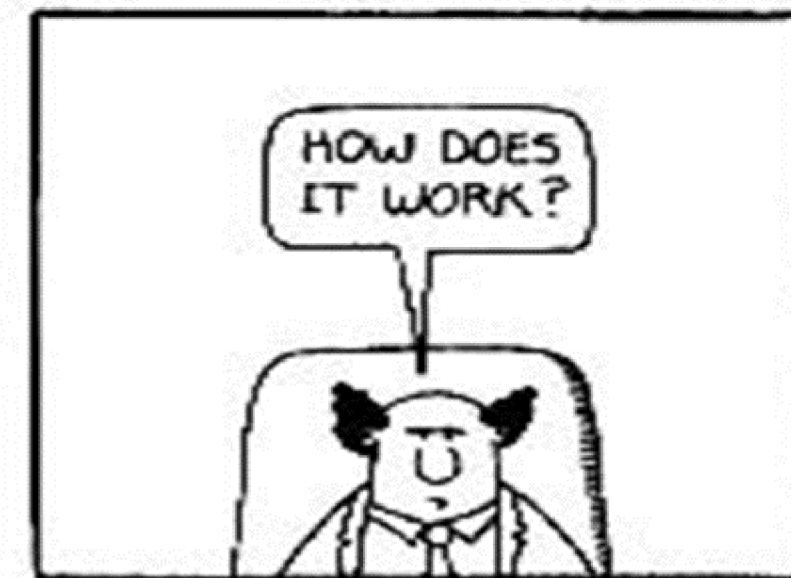
"I prayed about this. I know that God would not go around with pants down."

If I toss a fair coin, then:

- What is the probability of getting heads?
- Probability of two heads in a row?
- Probability of five heads in a row?
- What is the probability of getting tails after four heads?

The idea that the past influences the future forms the basis of the fallacy.

Cognitive Dissonance



Framing

Framing occurs when in the course of describing an issue or event, a speaker's emphasis on a subset of potentially relevant considerations causes individuals to focus on these considerations when constructing their opinions.

Decision frame: The same information, expressed in different forms, can lead to different judgments

The COVID-19 problem

Imagine that the US is preparing for the outbreak of COVID-19, which is expected to kill 600 people. Two alternative programs to combat COVID have been proposed.

Which do you choose?

- A. If Program A is adopted 200 people will be saved
- B. If Program B is adopted, there is a $\frac{1}{3}$ probability that 600 people will be saved, and $\frac{2}{3}$ probability nobody will
- C. If Program C is adopted 400 people will die
- D. If Program D is adopted there is a $\frac{1}{3}$ probability that nobody will die, and a $\frac{2}{3}$ probability that 600 people will die

More life or death examples



Rate an Area

Home

About

Contact

Log in using Facebook

About Ghetto Tracker

Traveling to a new city? Will you be visiting a safe part of town?

Ghetto Tracker can help you determine which parts of town are safe and which parts are not.

It's simple

We use a rating system that allows locals and people familiar with area rate which parts of town are safe and which ones are ghetto, or unsafe.

Who can benefit from the Ghetto Tracker?

People who can benefit from this site include, but are not limited to, students, traveling

professionals, people on vacation, anyone moving to a new city, and it can even help people become more familiar with their own city.

More life or death examples

Drug users are constructed as felons or people who need help

"Where strong social constructions exist, science that goes against the grain is usually ignored, whereas science that reinforces stereotypes is used as a rationale but changes few minds" (p. 109).



Thugs or 'very good' people?



Problem with using statistics as frame

- High school dropout obtained a GED
- Record of legal involvement
- Was a parent as a teenager
- Carries a weapon for protection
- A rebel
- Doesn't trust authority

Congresswoman?

Lauren Boebert

US House of Representatives



Policy Significance of Interpretation and Data Frames

The "policy implications" part of your manuscript is of upmost importance

- *Example:* Robin Shellow's life conviction for manslaughter (Qual)
- *Example:* 2015 lawsuit brought by students against CUSD (Quant)

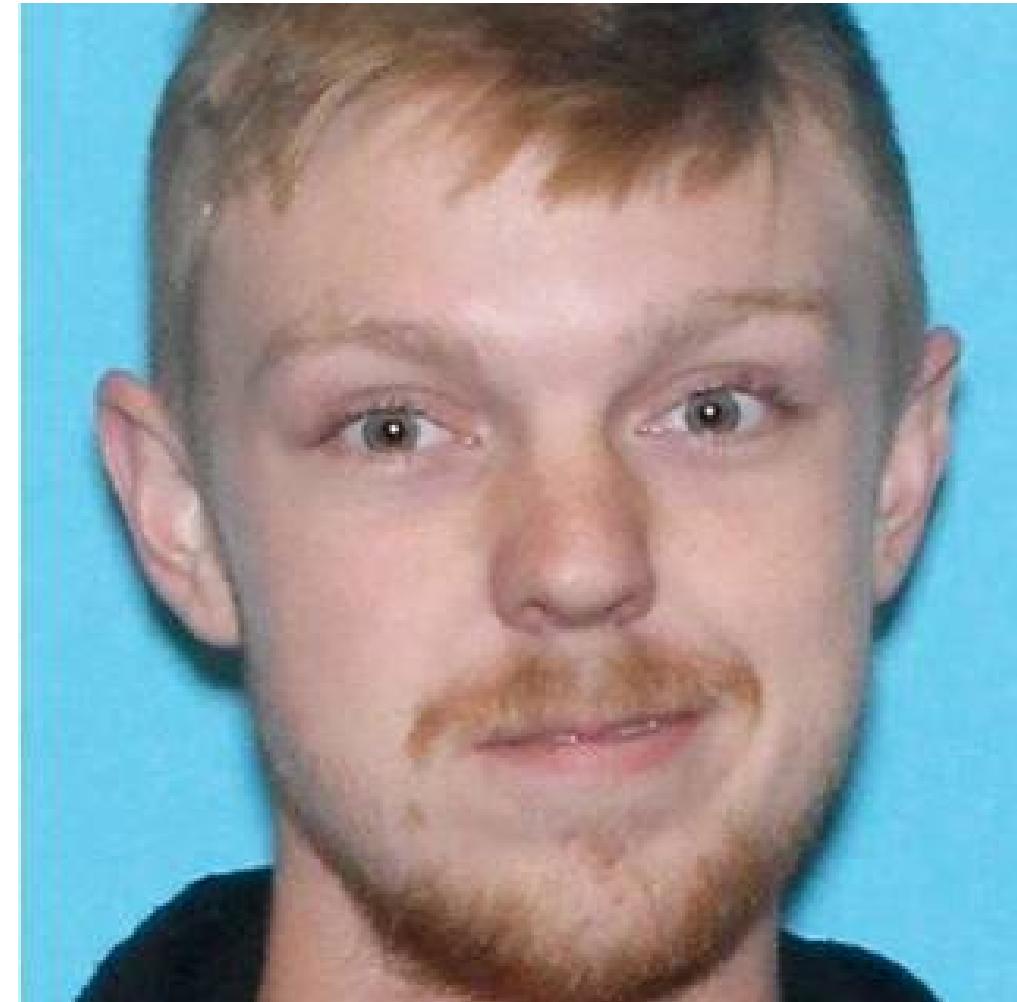
Urban Psychosis rejected as defense

At 17, Felicia Morgan viciously killed someone after a fight over a coat. Her lawyer argued that Morgan's traumatic childhood in a violent inner-city home and neighborhood created in her the urban counterpart of the post-traumatic stress disorders that affected some Vietnam veterans. A psychologist testified that because of her PTSD, Morgan did not know right from wrong



Suburban Neurosis

At age 16, Couch was intoxicated after a wild night of partying. He took his father's pickup truck and killed four people. His lawyer's argued he had a condition called "affluenza" because he was raised in an environment of considerable wealth and privilege. Couch was too affluent to know right from wrong.



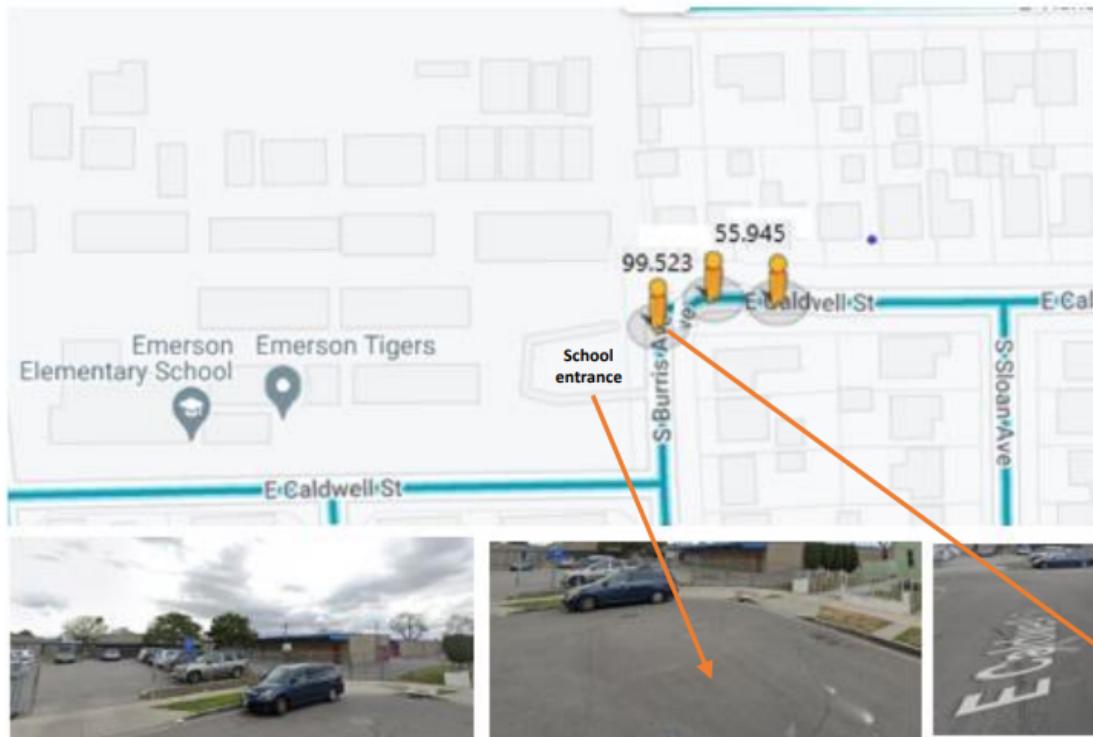


Figure 4. Driveway of the Emerson Elementary School, along a cul-de-sac, and the driveway in relation to the closest gun-related homicide violence (starting at 99ft) for the walking street network.

Network Constrained Spatial Dependence of Crimes Near Schools

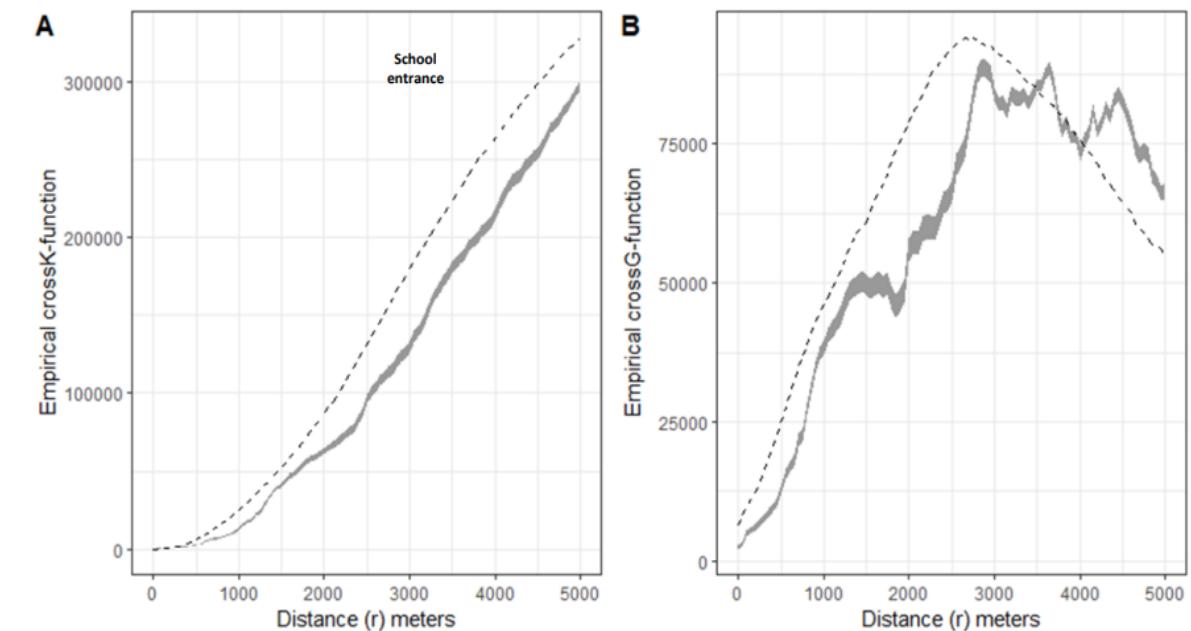


Fig 5. Bivariate Network (Cross-K) function of gun violence in proximity to schools.

What is Childhood Adversity?



2 Replies

MO ❤️

They live each other



Slim and trim version

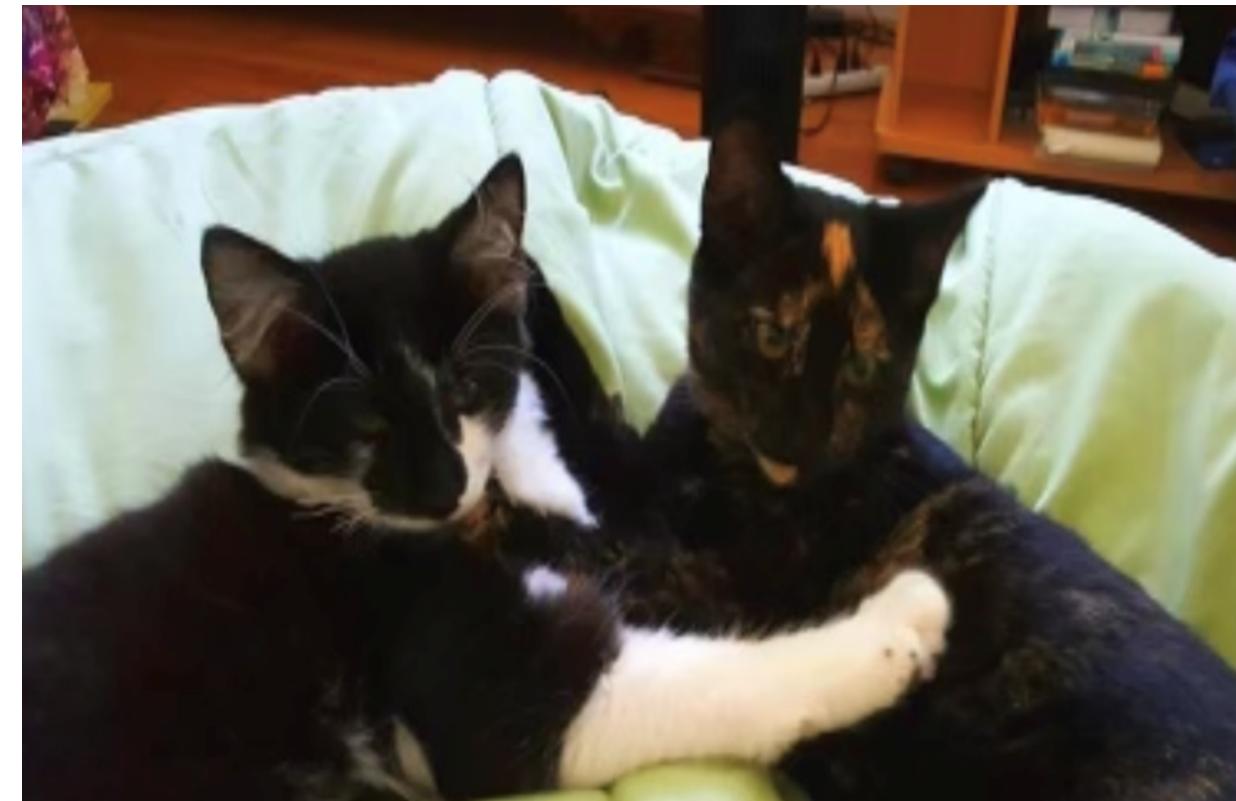


Oji

not anymore lol they got a divorce

You can get divorced and still love each other

In fact if you don't then you never loved at all



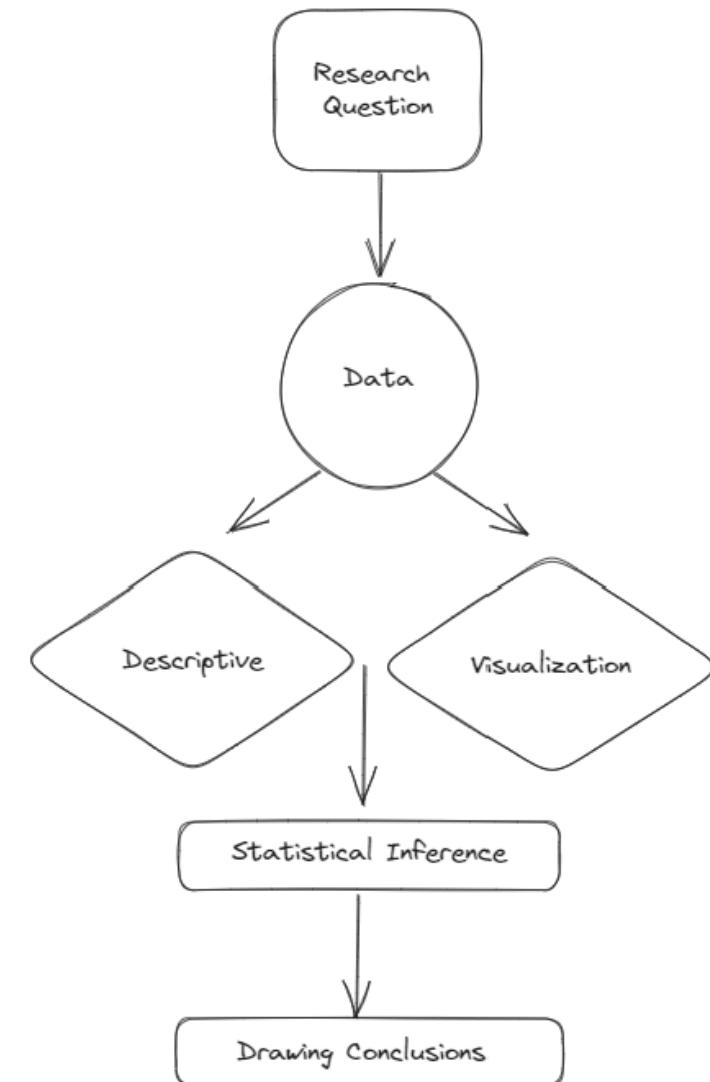
Introduction

Key Question: how do images become embedded in public consciousness?

- Short answer: our analyses, data and statistics
- The strategic combination of images, words, research, data, media operates on human intuition
- Numbers, just like words, or even more so, change the way an issue is framed

The research process

- Formulate a research question
- Find data
- Know the data
- Select a statistical analyses based on the variables you have
- Carefully draw conclusions that consider principles of DEI



Terminology: Concepts

1. Socially constructed abstractions that need clarification
 - Example: what is SES?
 - Gives us a specific working definition so that readers will understand the concept
 - Abstract concepts need to be defined with precision
2. Operational definition (how will we measure SES?)
 - Spells out precisely how the concept will be measured
 - i.e. it is specific and unambiguous

Terminology: Measurement

	Conceptualization	Conceptual Definition	Operational Definition	Real World Measure
Idea	An abstract idea or concept of something	Describes what is meant by the concept	Variables that measure the concept as defined	the actual questions we ask
Example	Traumatic experiences	Adversity in childhood	Ex. Living in a household with someone who abuses substances	Before age 18 have you lived with someone who abused substances or prescription pills

Conceptualization



Conceptual definition



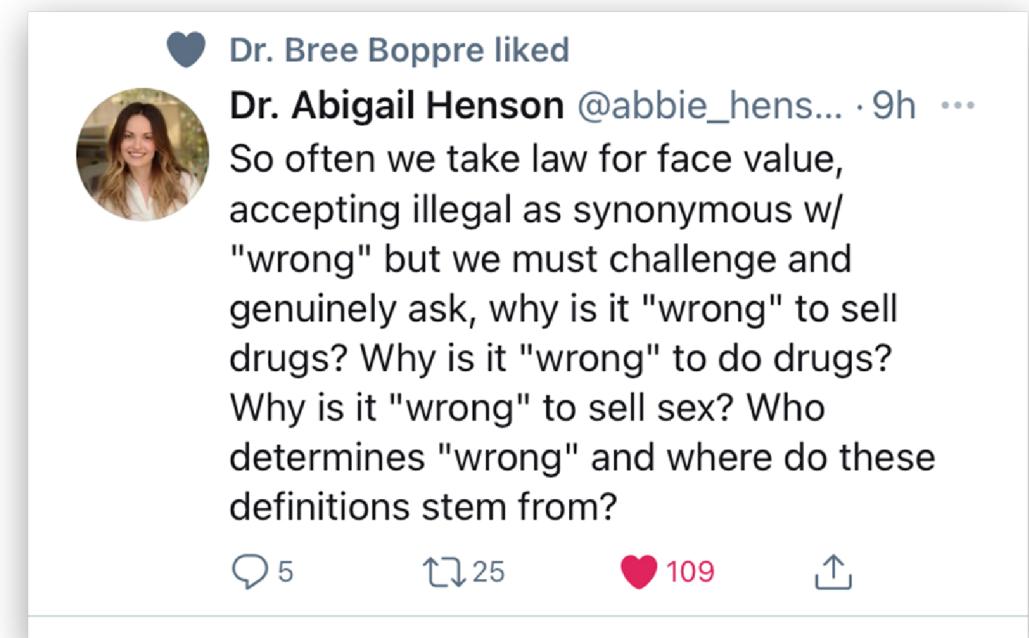
Operational definition



Real world measures

Terminology: Who decides?

- Any given concept has multiple definitions and measures (confuses our understanding)
- This depends on who decides
- Example: Sexual violence



Dr. Bree Boppre liked

Dr. Abigail Henson @abbie_hens... · 9h ...

So often we take law for face value, accepting illegal as synonymous w/ "wrong" but we must challenge and genuinely ask, why is it "wrong" to sell drugs? Why is it "wrong" to do drugs? Why is it "wrong" to sell sex? Who determines "wrong" and where do these definitions stem from?

5 25 109

No one data source measures the full range of violence

TABLE 5.1
Studies Measuring Sexual Violence

Study	Unit of analysis	Target population	Type of sexual violence covered	Best account for	Incidence or prevalence
General population surveys					
ICARIS-2	Individuals (females)	All females in the United States age 18 years and older	Forcible sex; unwanted sexual activity	Females in the United States	2.5% annual; 10.6% lifetime
DAFR/IR	Individuals (females)	All females in the United States age 18 years and older	Forcible rape; drug- and alcohol-facilitated rape	Females in the United States	18% lifetime
Criminal justice					
UCR	Aggregate; reporting agency	All law enforcement agencies; 98% reporting	Limited number reported and recorded crimes	Commercial and business victims	92,455 past year or 60/100,000
NCVS	Victimization, individuals and households	Individuals in households	Household and personal crimes	Household and personal crimes not reported to police	176,450 past year or 1.4/1,000
College studies					
ACHA-NCHA	Students in classrooms	All students attending college or university	Crimes on college campuses	Female college students at participating colleges	1,153,680 lifetime or 11.5%; 501,600 past year, 5%
DAFR/IR	College students	U.S. college students	Crimes on college campuses	Female college students in U.S. colleges	20,200,000 lifetime, 16%; 1,100,000, 0.6% past year <i>(continues)</i>

Conceptualizing and defining 'Elder Abuse'

Defining Body	Definition
Baker (1975)	Granny battering -- physical assault or abuse
American Medical Association	"actions or the omission of actions that result in harm or threatened harm to the health or welfare of the elderly."
Eastman (1988)	the physical, emotional or psychological abuse of an older person by a formal or informal <u>carer</u> ; the abuse is repeated and is the violation of a person's human and civil rights by a person or persons who have the power over the life of a dependent."
The Action on Elder Abuse (1995)	"Elder abuse is a single or repeated act or lack of appropriate action occurring within any relationship where there is an expectation of trust which causes harm or distress to an older person."
World Health Organization (2002)	"a single, or repeated act, or lack of appropriate action, occurring within any relationship where there is an expectation of trust which causes harm or distress to an older person."
Patient Protection and Affordable Care Act 2010	The knowing infliction of physical or psychological harm or the knowing deprivation of goods or services that are necessary to meet essential needs or to avoid physical or psychological harm.

Dangerous Labels

Subject	United States						
	Total		White alone, not Hispanic or Latino		Black or African American		Hispanic or Latino origin (of any race)
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate
Population 15 to 19 years	22,083,463	+/-8,671	12,537,287	+/-4,538	3,436,889	+/-6,578	4,398,876
SCHOOL ENROLLMENT							
Enrolled in school	19,202,446	+/-21,894	11,129,399	+/-13,926	2,913,370	+/-6,863	3,611,862
Public	87.4%	+/-0.1	84.6%	+/-0.1	91.8%	+/-0.1	92.7%
Private	12.6%	+/-0.1	15.4%	+/-0.1	8.2%	+/-0.1	7.3%
Not enrolled in school	2,881,017	+/-23,167	1,407,888	+/-13,838	523,519	+/-5,985	787,014
MARITAL STATUS AND FERTILITY							
Male	11,342,438	+/-6,068	6,439,396	+/-2,967	1,749,416	+/-5,054	2,282,702
Ever married	1.0%	+/-0.1	0.8%	+/-0.1	0.9%	+/-0.1	1.6%
Female	10,741,025	+/-6,463	6,097,891	+/-3,049	1,687,473	+/-4,207	2,116,174
Ever married	1.9%	+/-0.1	1.6%	+/-0.1	1.1%	+/-0.1	3.5%
Female with a birth in the past 12 months	2.7%	+/-0.1	1.7%	+/-0.1	4.3%	+/-0.1	4.5%
HOUSEHOLD TYPE							
Population 15 to 19 years in households	20,199,416	+/-11,594	11,298,069	+/-9,437	3,138,154	+/-6,859	4,221,918
In married-couple family households	61.7%	+/-0.3	69.6%	+/-0.3	34.4%	+/-0.3	59.0%
In male householder, no wife present, family households	7.8%	+/-0.1	7.3%	+/-0.1	7.8%	+/-0.1	9.3%
In female householder, no husband present, family households	26.9%	+/-0.2	18.9%	+/-0.2	55.1%	+/-0.3	29.0%
In nonfamily households	3.6%	+/-0.1	4.3%	+/-0.1	2.7%	+/-0.1	2.7%

1 of 4

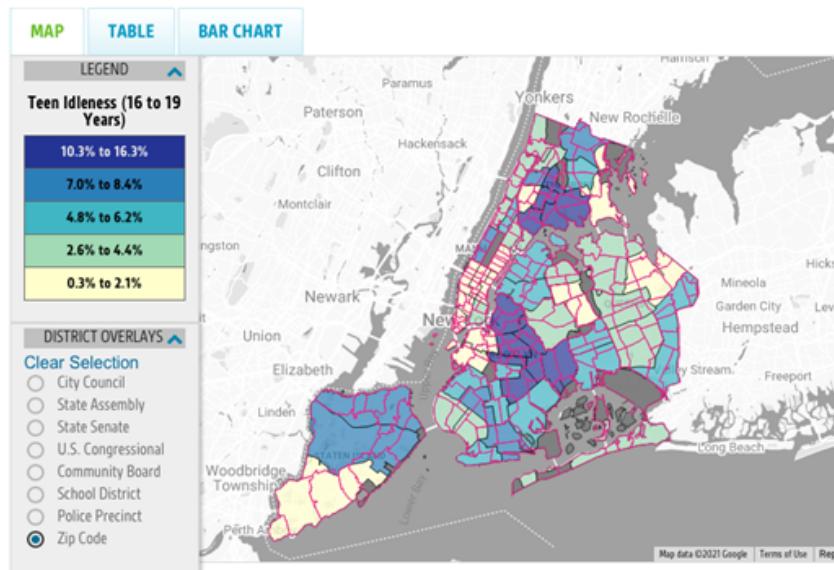
09/16/2013

Subject	United States						
	Total		White alone, not Hispanic or Latino		Black or African American		Hispanic or Latino origin (of any race)
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate
Population 16 to 19 years	17,791,435	+/-14,515	10,117,024	+/-9,668	2,775,825	+/-7,617	3,522,231
IDLENESS							
Not enrolled in school and not in the labor force	5.4%	+/-0.1	4.0%	+/-0.1	8.3%	+/-0.1	7.8%
LABOR FORCE STATUS							
In the labor force	40.5%	+/-0.1	44.2%	+/-0.1	34.9%	+/-0.2	37.9%

And, idleness linked spatially to certain areas

Teen Idleness (16 to 19 Years)

Community District; Percent; 2019 (3-yr. avg.)



New York City

Teen Idleness (16 to 19 Years)

5.8%

IN 2019



Change since 2018 (3-yr. avg.)

New York City

5.8%

-0.0%



Source: U.S. Census Bureau, American Community Survey combined 1-Year Estimates, Public Use Microdata Sample File (2007-2015 3-year averages); retrieved from <https://data.census.gov/>.

Definitions: Teens 16 to 19 years who are not in school and not in the labor force. This excludes teens who serve in the armed forces. Percents refer to the percent of all teenagers 16 to 19 years.

Notes: Due to small sample sizes in smaller geographies, data are presented in 3-year averages; 2019 = 2017-2019 3-year average.

In 2018, CCC revised the methodology for calculating Teen Idleness and updated all previous years data (2007-2015) using this new method. Data for Teen Idleness are used in our annual publication, the [Child and Family Well-being Report](#) (explore the data under the [Rank Communities by Risk](#) tool). Because of the aforementioned methodological change, data for Teen Idleness in the Child and Family Well-being Report prior to 2016 may not match the values represented here.

Trends in Black Male Joblessness: An Employment Crisis Ignored

During the 2001 recession and the jobless recovery of 2002, when wage and salary employment in the U.S. economy actually declined, the idleness rate of 20-64 year old men increased by nearly two percentage points, rising to 13.4 percent by 2002. Idleness rates increased to some degree among males in each of the four race-ethnic groups, but the rise in idleness was greatest among Black men. By 2002, one of every four Black men in the U.S. was idle all year long. This idleness rate was twice as high as that of White and Hispanic males. It is interesting to note that the Hispanic male idleness rate barely changed during this two year period despite a continued surge in Hispanic immigration. As Janny Scott noted in the February 2004 New York Times article on high Black male joblessness in New York City, one of the forces believed to be influencing the steep decline in Black male job opportunities in New York City is “increased competition with immigrants for low skill jobs”.¹³

MCAS Multiple-Choice Results Interpretive Guide

Grade 10 Math MCAS raw scores and percent of answers correct by Achievement Level: A “Deep Investment”?

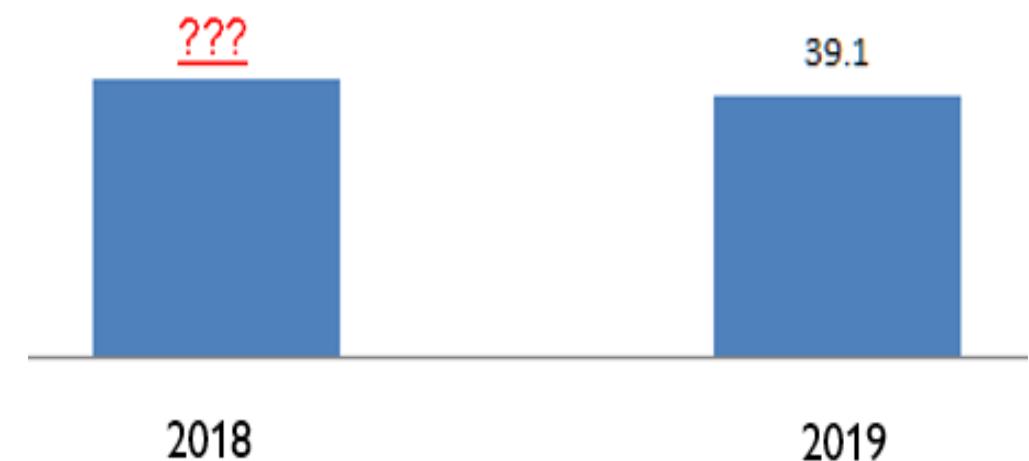
Total Raw Score on MC Items	Likely Achievement Level, Based on MC	Percent Correctly Answered (Gia's calculation)
0-5	Low Failing (200-208)	0 – 15.625%
6-10	High Failing (210-218)	18.75 – 31.25%
11-13	Low Needs Improvement (220 – 228)	34.375 – 40.625%
14-16	High Needs Improvement (230 – 238)	43.75 – 50%
17-19	Low Proficient (240 – 248)	53.125 – 59.375%
20-22	High Proficient (250 – 258)	62.5 – 68.75%
23-29	Low Advanced (260 – 268)	71.875 – 90.625%
30-32	High Advanced (270 – 280)	93.75 – 100%

More trickery

A newspaper reported that the teen birthrate fell to a record low, dropping 6% from 2018 to 39.1 births per 1000 in 2019. It went on to state “The decline in teen births is really quite amazing.”

Solve for ???

hint: $\frac{(T_1 - T_2)}{|T_2|} \times 100$



Statistical Terminology

1. **Unit of analysis:** a single entity (person or object) whose characteristics are of interest
2. **Population of units:** the complete collection of units about which information is sought
3. **Population:** a set of all measurements corresponding to each unit in the entire collection of units about which information is sought
4. **Sample:** a subset of measurements selected from the population of interest

Textbooks

- We use parameter(s) to describe the population of interest
- We use statistic(s) to describe the sample with respect to the population

Characteristic	Population parameter	Sample statistic
mean	μ	\bar{X}
variance	σ^2	s^2
std.dev.	σ	s

Variables

A variable is a characteristic of a unit that may vary for different observations

- There are two main types of variables
 1. Qualitative (aka categorical; discrete); and
 2. Quantitative (aka continuous; numerical)

Levels of measurement

Quantitative variables that can theoretically take on any value in a certain range

Qualitative variables that can take on only a finite number of values and are placed into categories

Levels of measurement: Qualitative variables

1. Qualitative variables have labels or names used to identify an attribute of a unit
2. Qualitative data use either the nominal or ordinal scale of measurement

Levels of measurement: Qualitative variables

Examples

Qualitative variable: has labels or names used to identify an attribute of a unit.

- Nominal: order does not matter
 - numbers assigned to attributes **do not matter**
 - Gender, SSN, race/ethnicity, marital status
- Ordinal: order does matter
 - rankings -- distance separating the attributes has **no meaning**
 - position in race, education

Levels of measurement: Quantitative variables

Examples

Quantitative variable: has numeric values that indicate how much or how many of something.

Quantitative data uses either the interval or ratio scale

Examples

Interval: difference of quantities that are meaningful but ratios of quantities that cannot be compared; interval measures have no meaningful “0”

e.g. temperature

- 50 degrees is less than 70 degrees but 0 does not mean absence of heat
- 80 degrees is not twice as hot as 40 degrees

Examples

Ratio: ratios of quantities are meaningful and can be compared

- Have all structural features above and 0 is meaningful
- 0 represents the absence of something
- Example. Age: 0 years old, 0 years on the job, 0 amount of money, if I'm 40 and you're 20 I'm twice your age, etc.

Levels of measurement: Example

Grade	Major	gpa	credit hours
Sophomore	Psych	3.8	30
Fresh	Philosophy	4.0	12
Senior	Statistics	2.0	130
Senior	Social Work	2.8	99

Questions

What is the unit of analysis?

How many units in the data?

How many variables in the dataset?

What data type is each variable?

Measurement as scoring

- Measurement involves making observations and assigning scores or numbers to those observations
- Every variable should have two important qualities:
 - Exhaustive – You should be able to classify every observation in terms of one of the attributes composing the variable
 - Mutually exclusive – You must be able to classify every observation in terms of one and only one attribute
 - Example: Gender; Employment status; sentence type

Measurement as scoring

Create an exhaustive and mutually exclusive measure for:

- Income
- Sexual and Gender minority

Criteria for Measurement Quality: Reliability

Whether a particular measurement technique, repeatedly applied to the same object, would yield the same result each time

Problem – Even if the same result is retrieved, it may be incorrect every time

- Think of this as getting the wrong result every time
- Reliability does not insure accuracy
- Observer's subjectivity might come into play
- Example: My scale says I weigh 120 lbs every day but I weigh 150

Criteria for Measurement Quality: Validity

The extent to which an empirical measure adequately reflects the meaning of the concept under consideration

Problem: Are you really measuring what you say you are measuring?

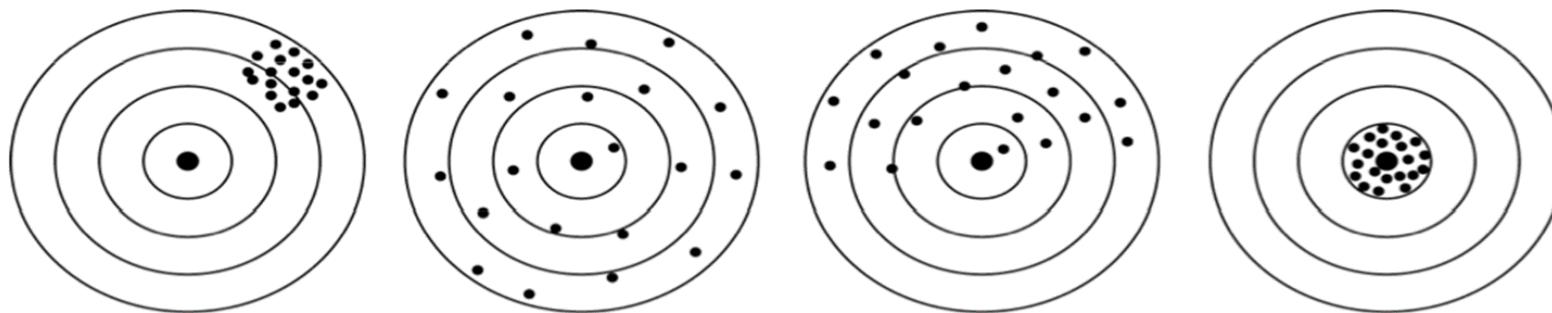
- Think of this as, are are getting the right result every time, but you are not measuring what you thought
- Example: My scale says 150 lbs every day, but it is actually stating my IQ lol

Demonstrating validity is more difficult than demonstrating reliability

Criteria for Measurement Quality: Typical demonstration

In the image below, suppose the bulls-eye of the target represents the true underlying risk of disease in a population and the holes in the target represent multiple objective measurements of the risk. In other words, the whole target is a part of the risk for the population.

- when are we measuring true risk? (valid)
- when are we measuring the risk consistently? (reliable)



Types of Studies

Depending on how a study was conducted, we have the following types of studies:

Observational study: a study in which the investigator observes a variable of interest of an existing sample in order to draw conclusions

Experimental Study: a study in which the investigator examines how a response variable behaves when the researcher manipulates one or more factors in order to determine the effect of those factors on the response.

Examples

A researcher wants to know if smoking during pregnancy leads to children with lower IQ scores. She looks at 200 pregnant women and records smoking status along with the subsequent IQ score (measured a few years after birth)

A scientist tries his weight loss drug on a group of monkeys with identical diets. 40 monkeys are randomly assigned to either get the drug or not get the drug (20 in each group). The weight gained or lost was recorded for each monkey.

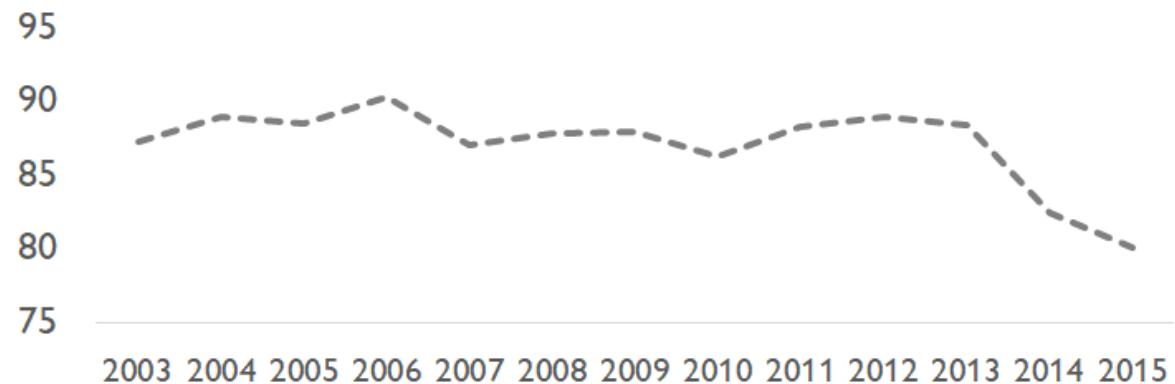
Dataset Types

1. Cross-sectional
2. Longitudinal
3. Spatial
4. Spatiotemporal

Putting it all

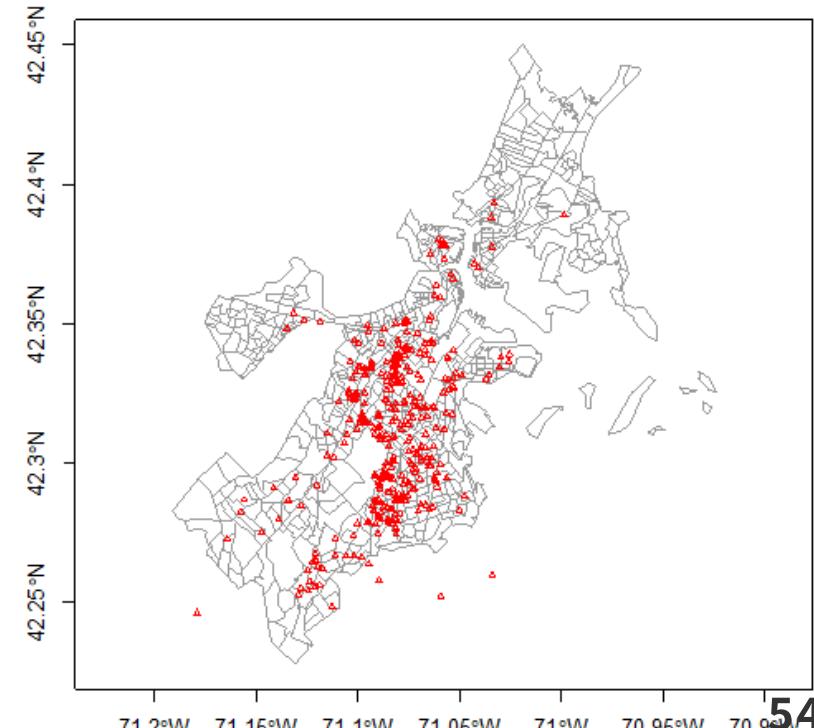
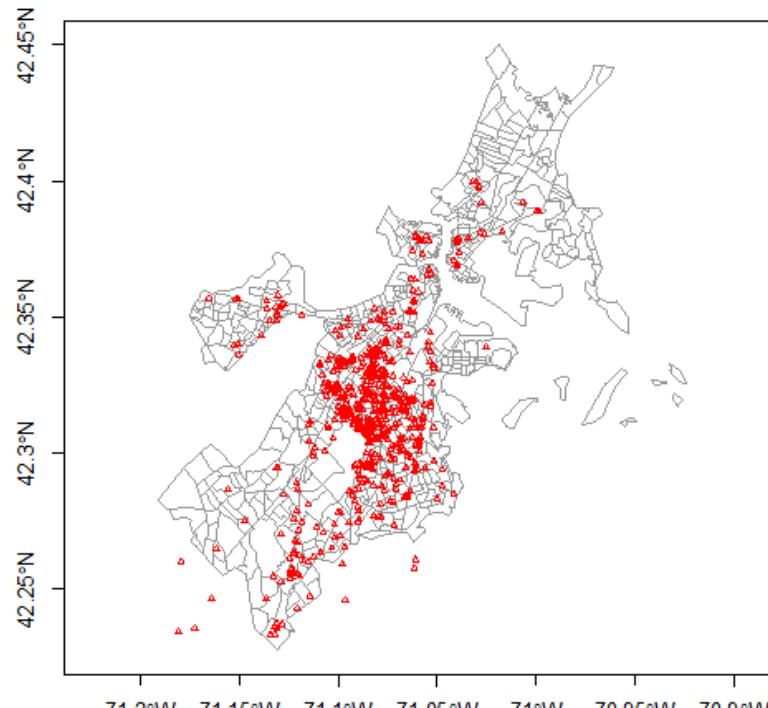
Innocent New Yorker's Stopped and Frisked (2003 - 2015)

- The Stop, Question and Frisk Data | NYC Open Data (cityofnewyork.us)



Investigatory Police Stops

Suspect Wearing a Hoody or Cap



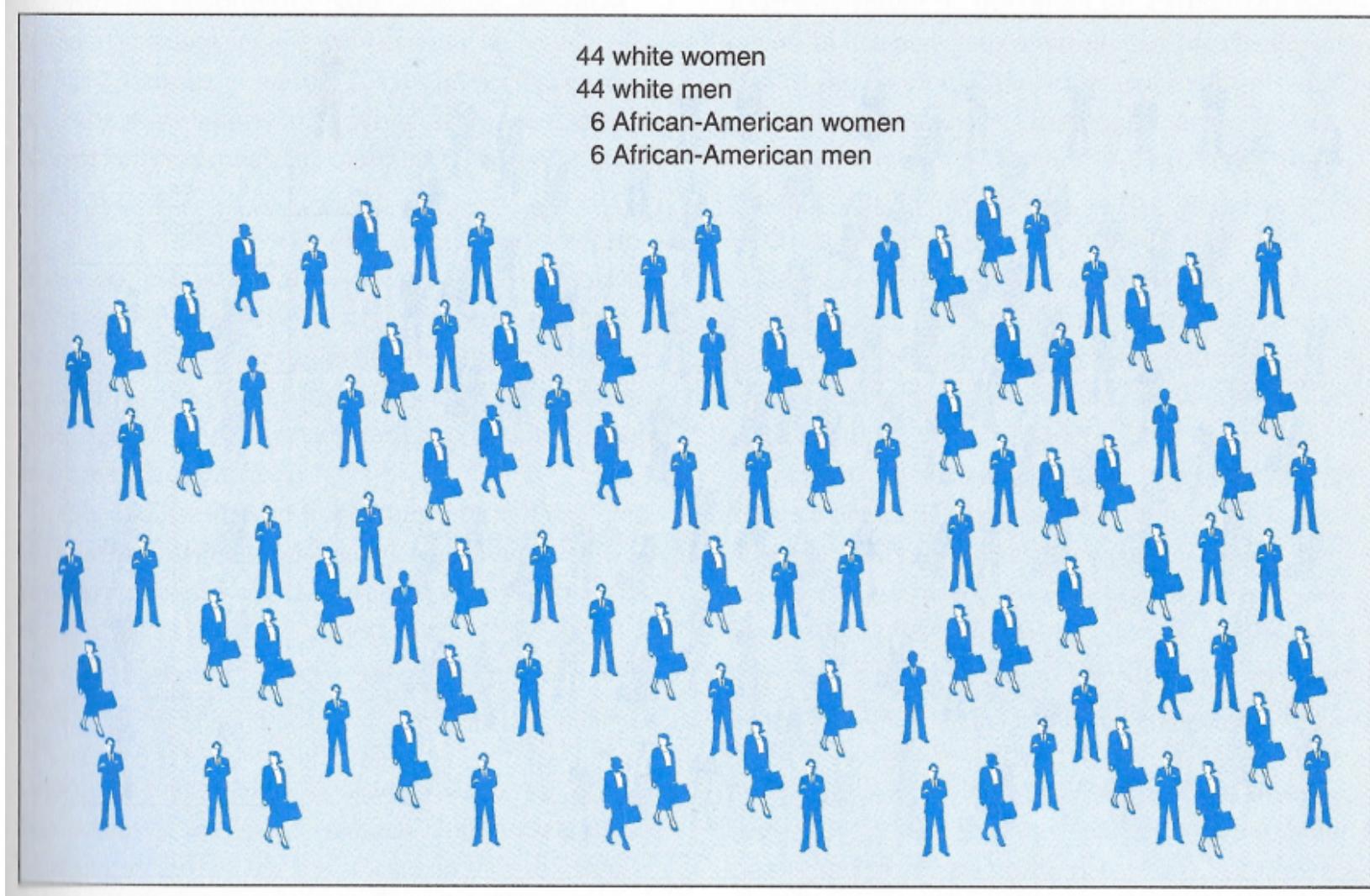
Collecting Data

Polls	Clinton	Trump	Margin
YouGov/Economist	45	41	Clinton +4
IBD/TIPP	41	43	Trump +2
Insights West	45	41	Clinton +4
Bloomberg/Selzer	46	43	Clinton +3
Lucid/The Times-Picayune	45	40	Clinton +5
Fox News	48	44	Clinton +4

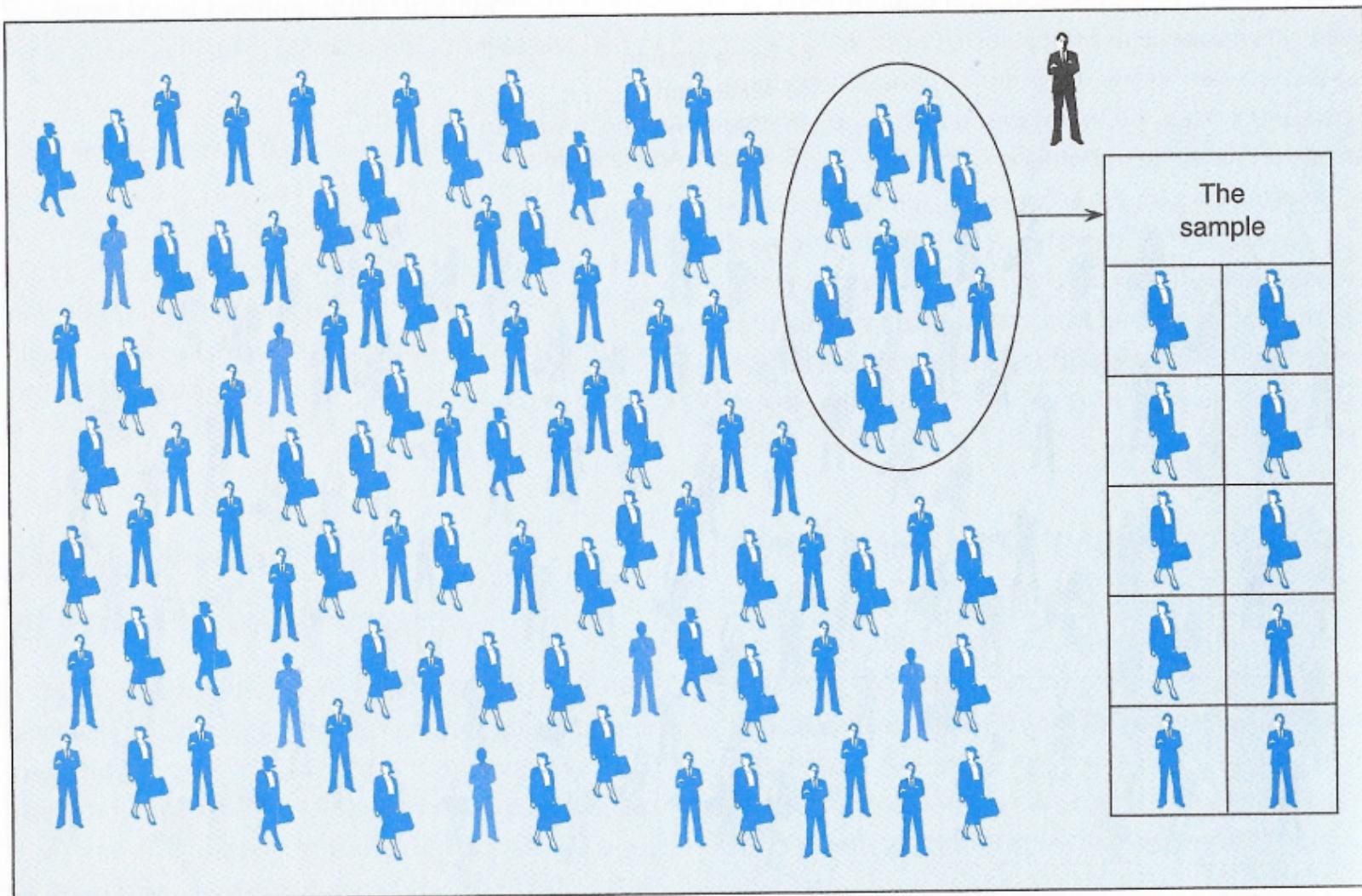
Ensuring Representativeness

- Samples should be representative
 - The sample has the same characteristics as the population
 - Note: Samples need only be representative in regards to characteristics relevant to the study
- EPSEM: equal probability of selection --> ensures representativeness
 - If every case in the **population** has the same chance of being selected, the sample is highly likely to be representative

A population



Not random



Sampling techniques

Samples must be representative of the population!

Statistical sampling is the procedure by which we select a subset from a population of interest that should be *representative* of that population. The most common sampling methods are:

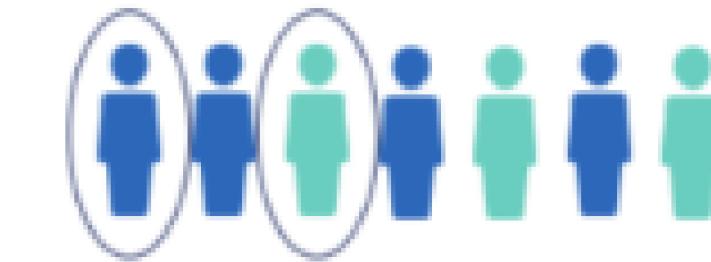
- Simple random sample
- Stratified sample
- Cluster sample
- Systematic sample
- Convenience sample

The sampling distribution is the single most important concept in statistics! (next week)

Simple random sample

Simple random sampling (SRS):
a sample selected such that
each element in the population
has the same probability of
being selected

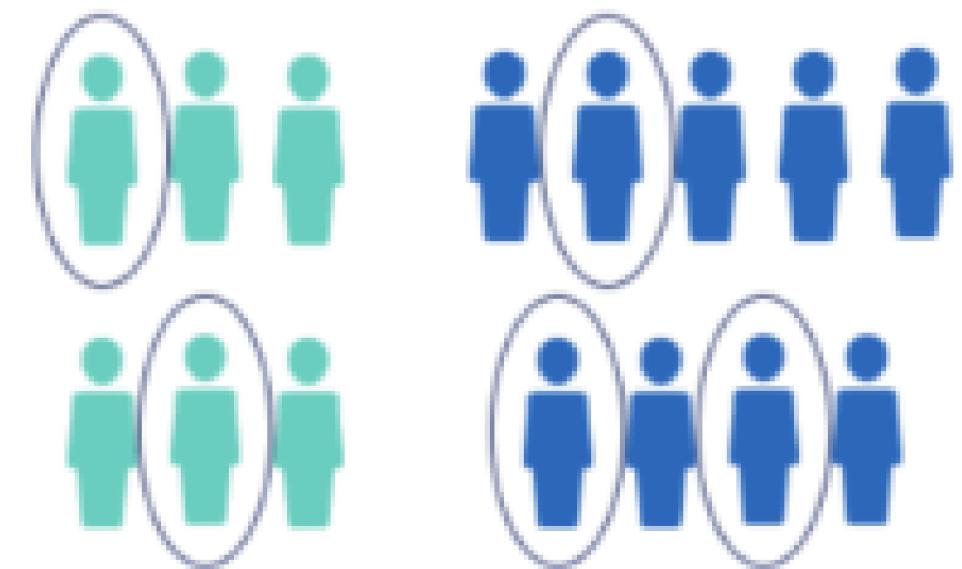
Simple random sample



Stratified sample

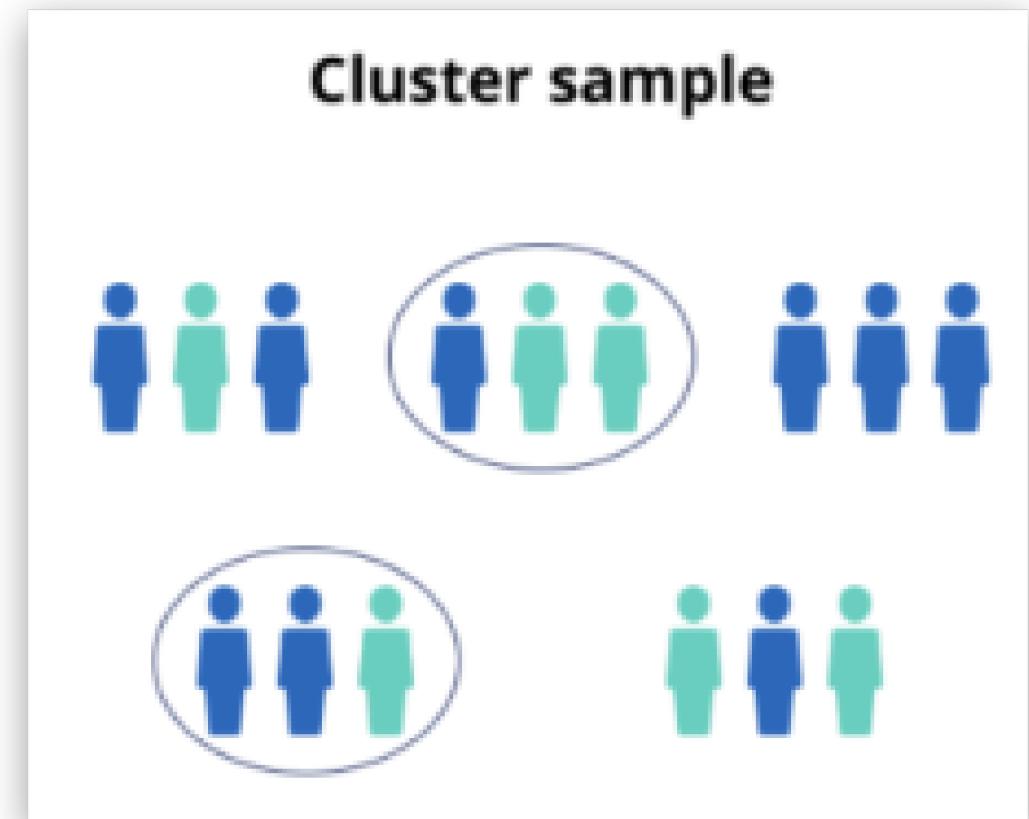
Stratified sample: elements in the population are first divided into groups and a simple random sample is then taken from each group

Stratified sample



Cluster sample

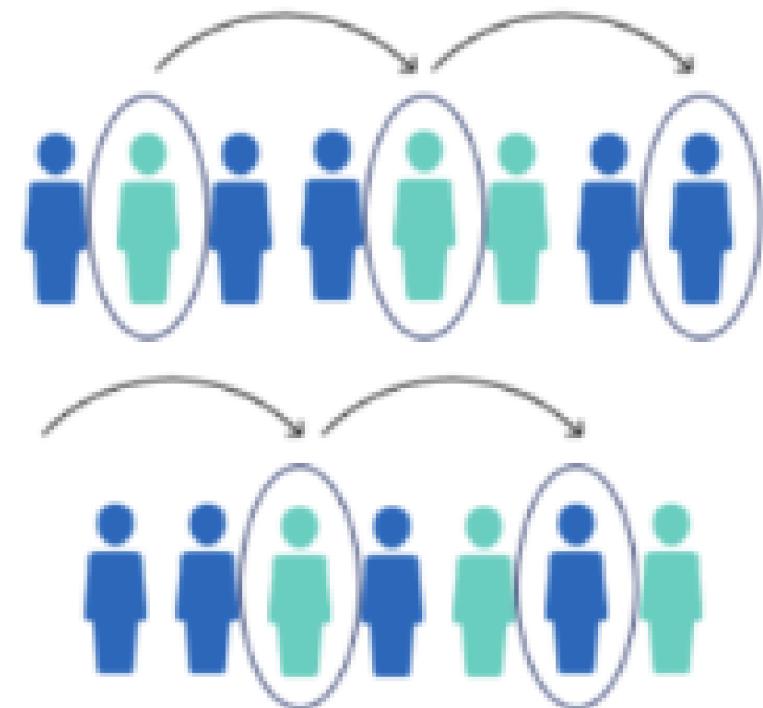
Cluster sampling: the elements in the population are first divided into separate groups called clusters and then a simple random sample of the clusters is taken that all elements in a selected cluster are part of a sample



Systematic sample

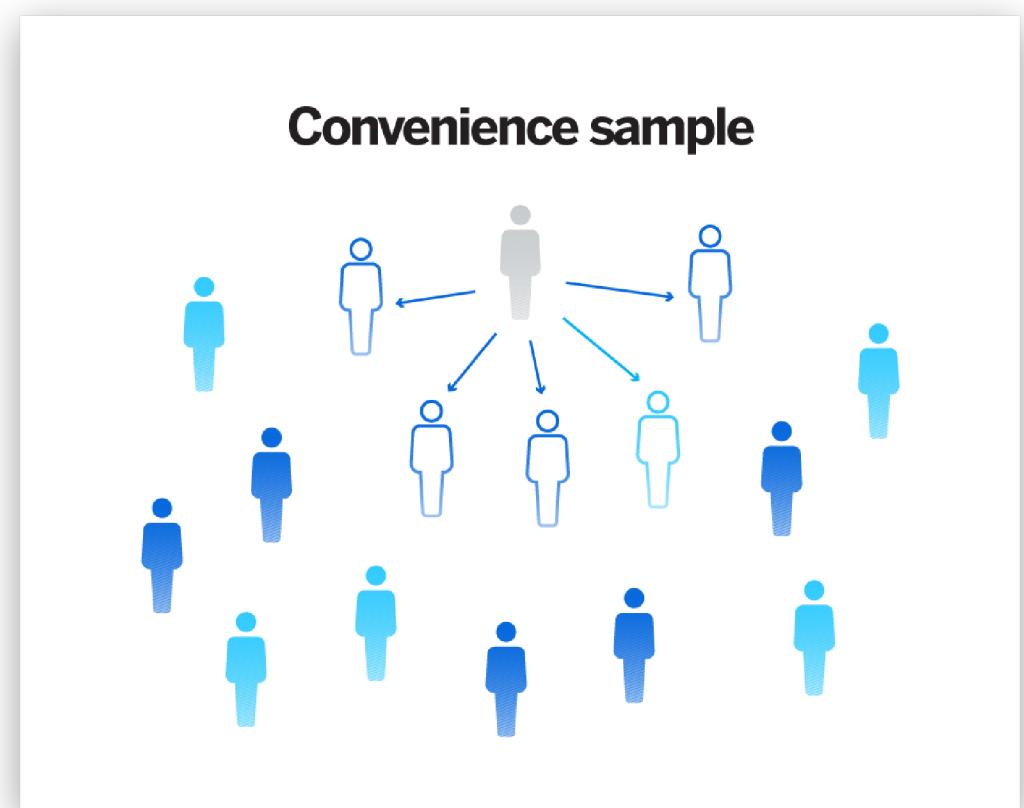
Systematic sampling: randomly select one of the first k elements from the population and then every k^{th} element thereafter is picked

Systematic sample



Convenience sample

Convenience sampling:
elements selected from the
population on the basis of
convenience

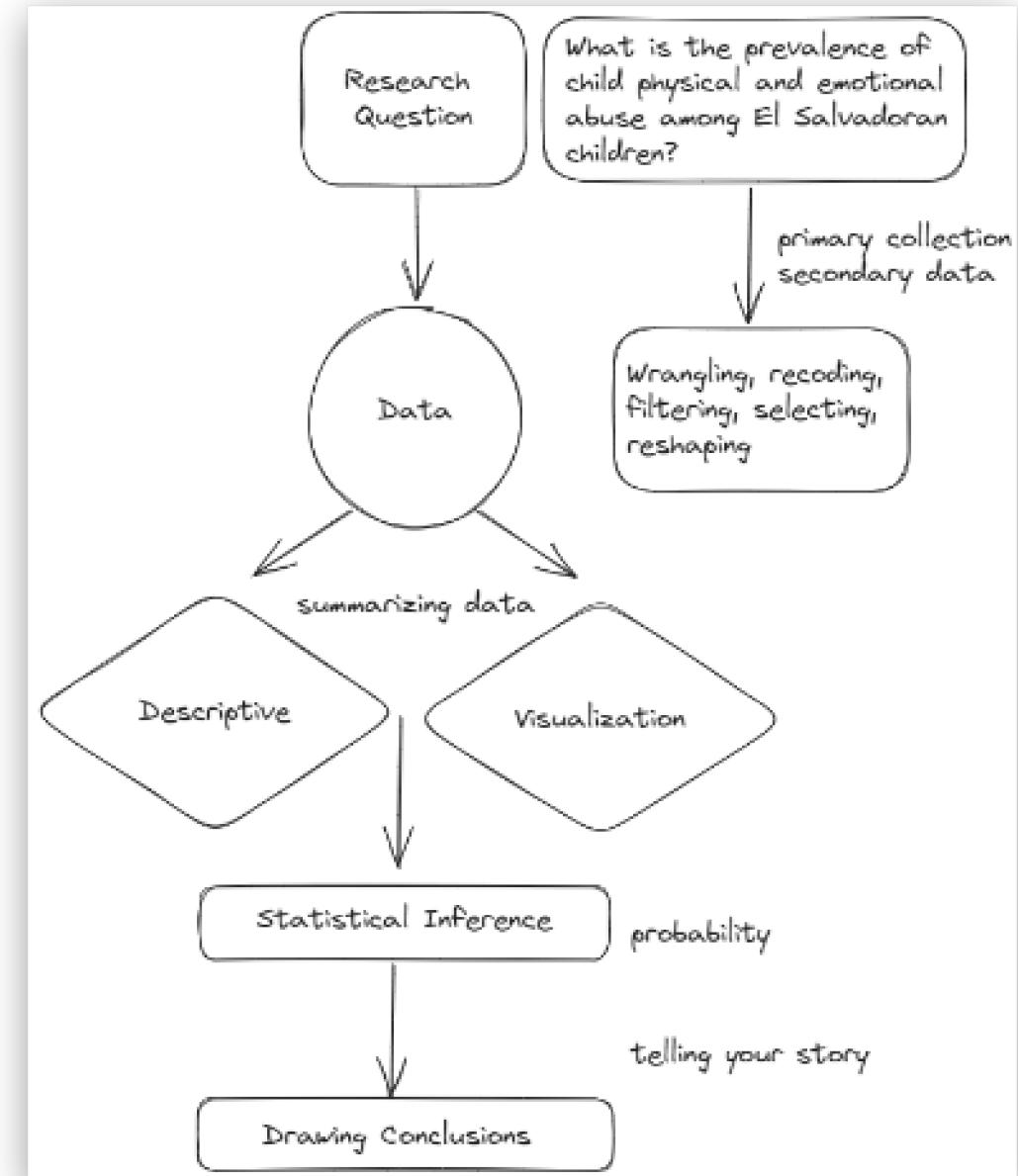


Lab 2

1. What is the prevalence of child physical and sexual abuse among El Salvadoran children?
 2. Is there a relationship between abuse and internalizing symptoms?
- Background & Importance
 - Data
 - Summarizing data set
 - Statistical analyses
 - Interpretation

Overview

1. Importance of finding data
 - Violence against children
2. Look at codebook and identify key variables for El Salvador
3. Download data as SAS file (must request data)
 - I uploaded the data [here](#)
4. Open SAS file in SPSS
5. Locate and clean variables/make chart on next slide using R integration



To Do

1. Look at codebook and find variables on child physical and emotional abuse
2. Rename variables, run analysis, compare to codebook
3. Identify issue, select cases, compare to codebook
4. Identify issue, etc.
5. Make R plot in SPSS

