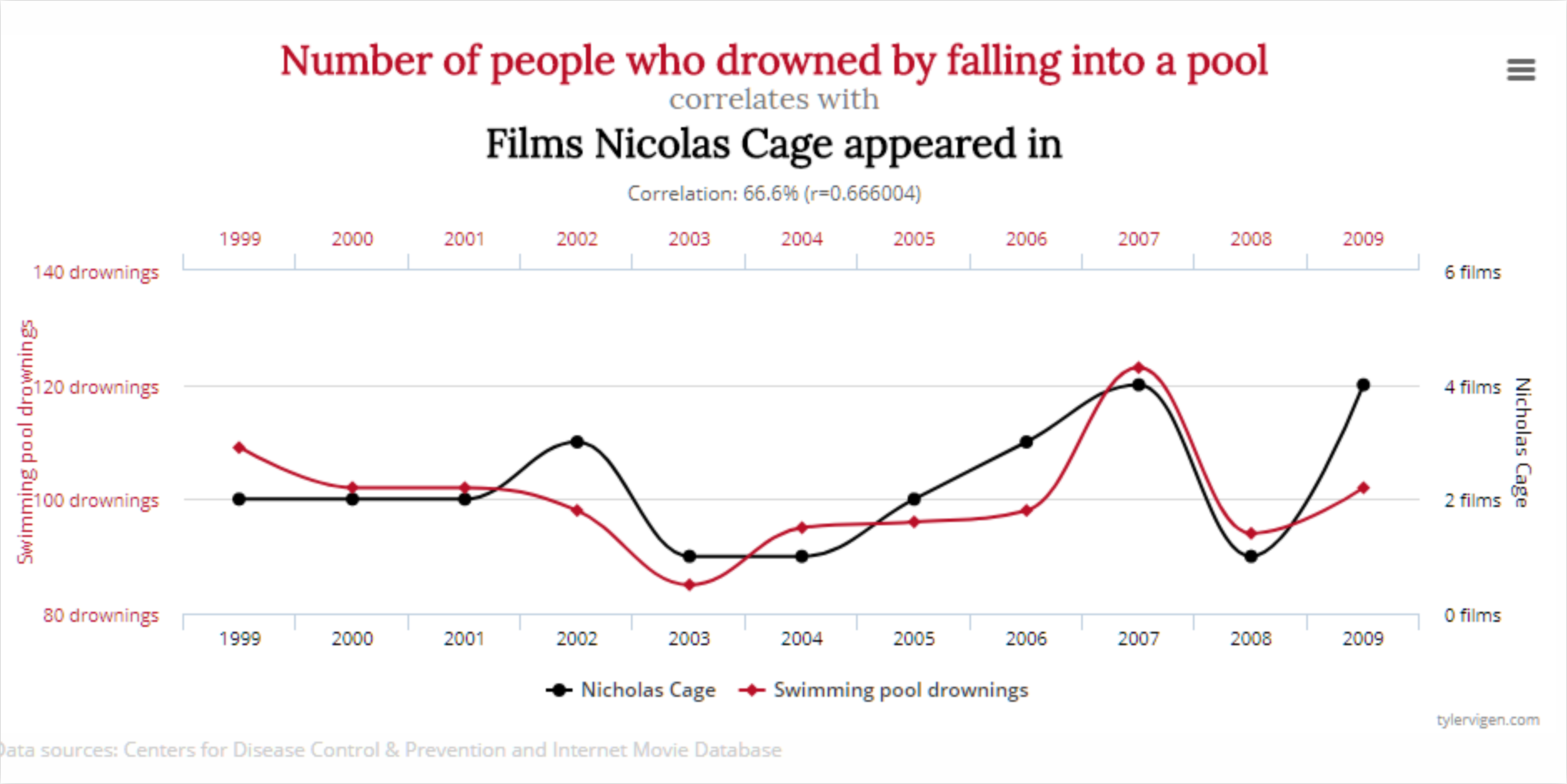
Lecture 1:
Experiments and causal inference

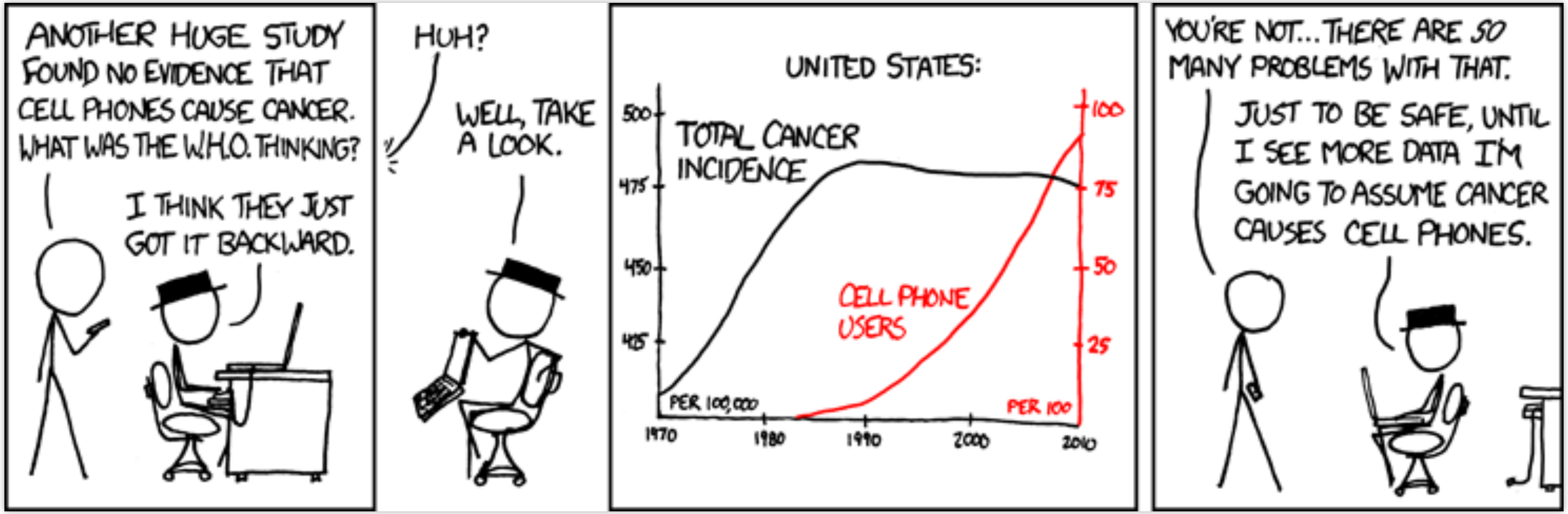
Summer Institute in Computational Social Science @ CU Boulder
August 17th, 2018

University of Colorado Boulder
SICSS

Correlation and causation



Causal mechanisms in observational data

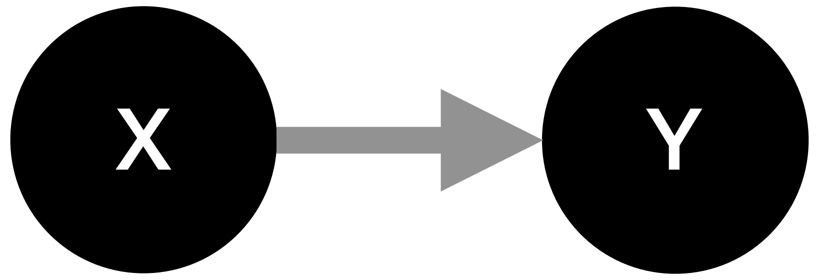


Association or Causation (Hill 1965)

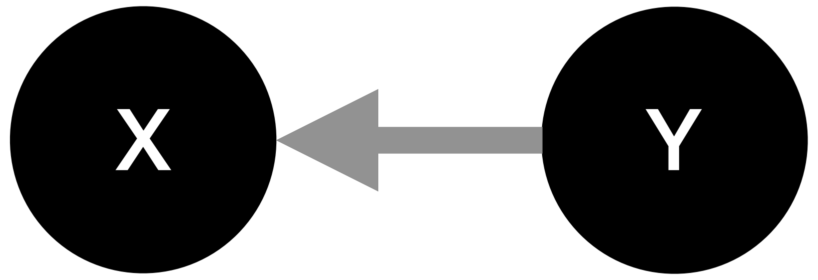
1. **Strength**: How big is effect you are measuring? (Large effects imply causality, but so can small effects)
2. **Consistency**: Can the effect be replicated? (Causal effects should be reproducible)
3. **Specificity**: Can association be pinpointed? (No other mechanisms should plausibly explain)
4. **Temporality**: Do the causes come before effects? (No time traveling)
5. **Gradient**: Do stronger/weaker treatments cause greater/lesser effects? (Same as covariation)
6. **Plausibility**: Does the causal mechanism itself make sense? (How could Cage films cause drowning?)
7. **Coherence**: Is the causal mechanism compatible with other evidence?
8. **Experiment**: Can experiments reproduce the effect?
9. **Analogy**: Is the causal mechanism similar to other established mechanisms?

Alternative causal relationships

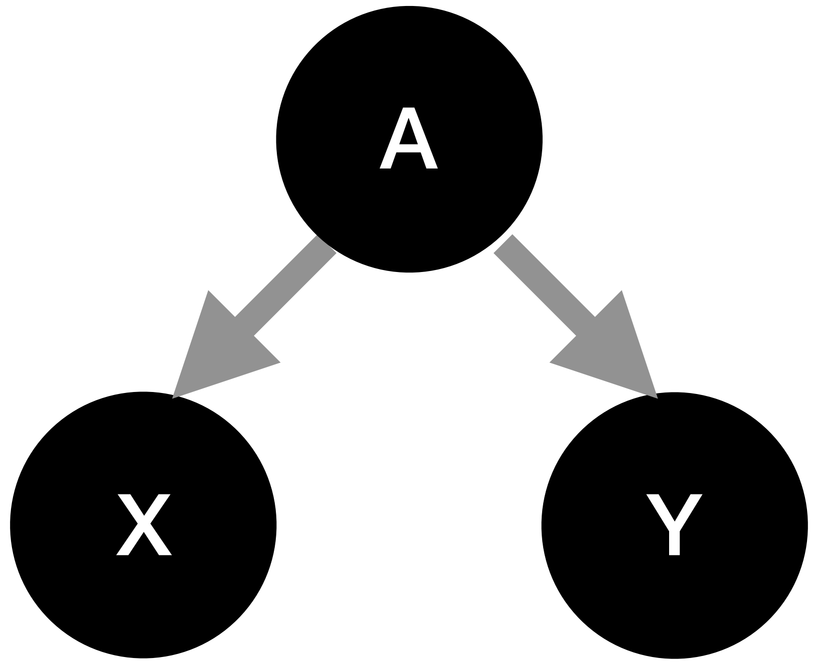
You observe a correlation between X and Y:



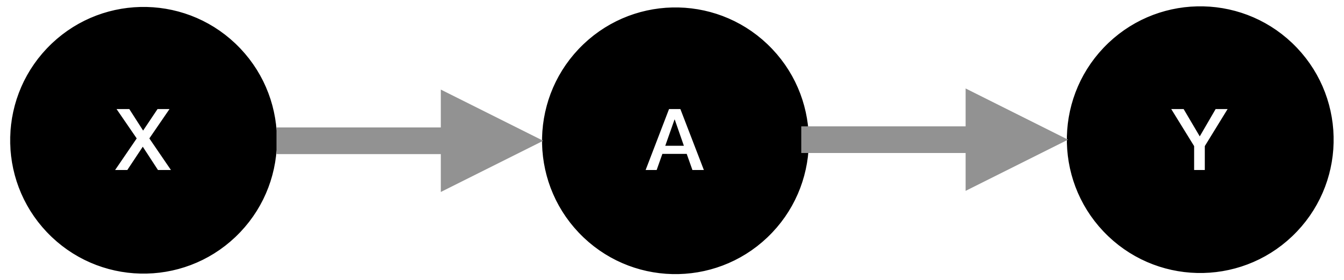
Simple Causation
X causes Y



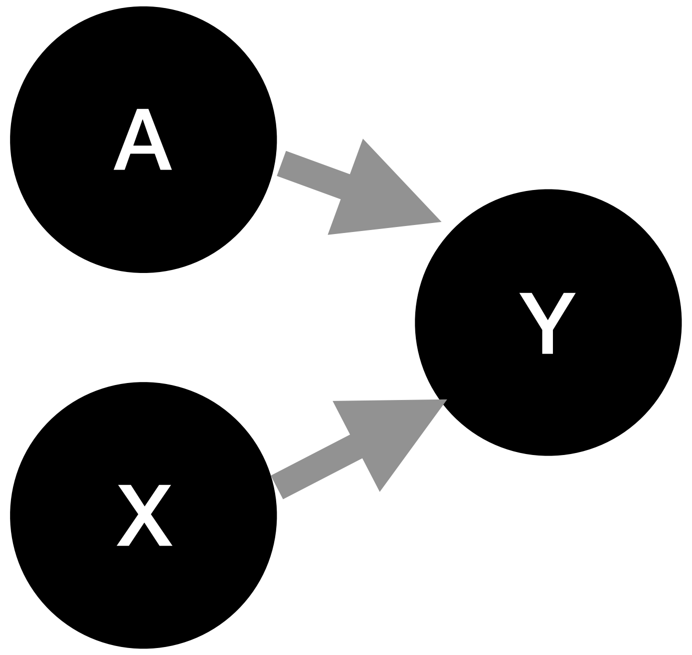
Reverse Causation
Y causes X



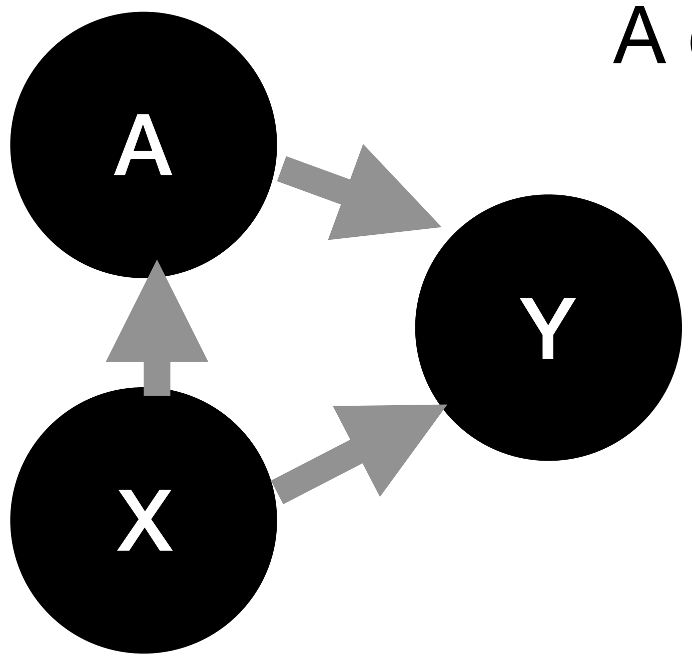
Confounding
A causes X and Y



Mediation
X causes A causes Y

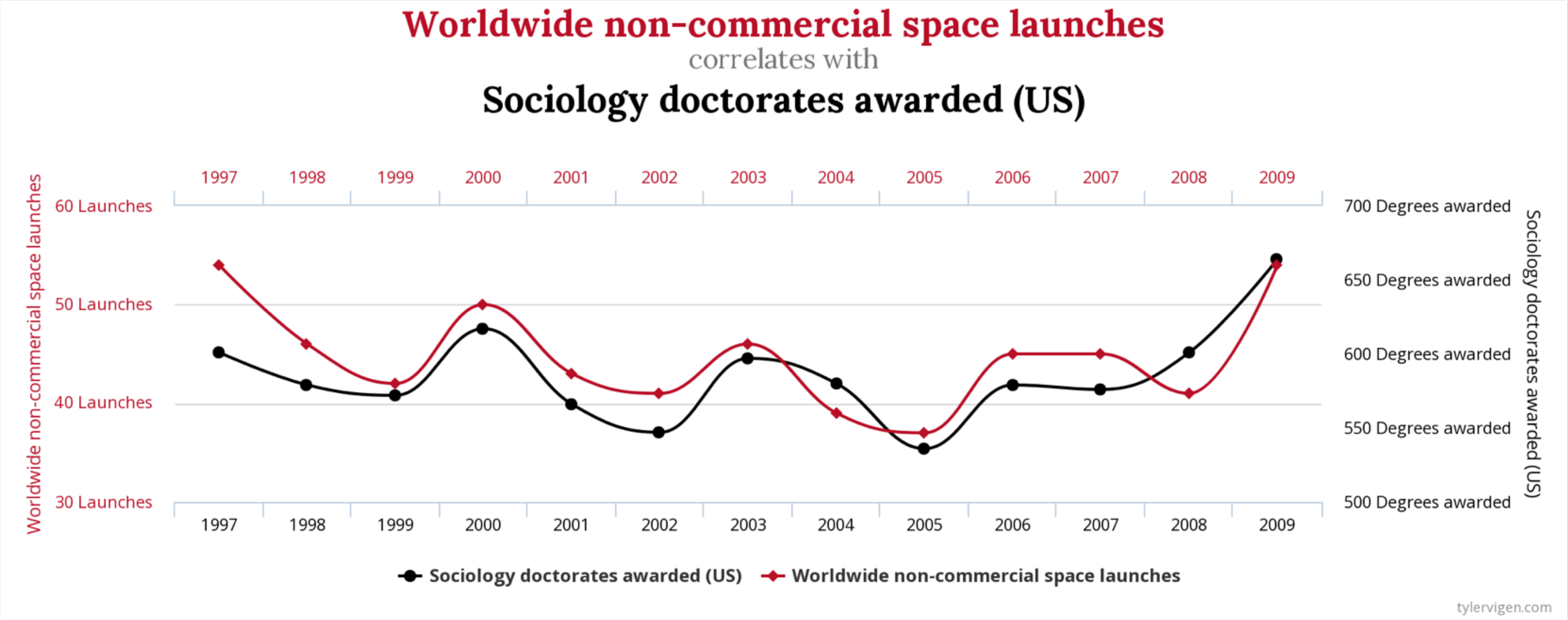


Multiple causation
X and A causes Y



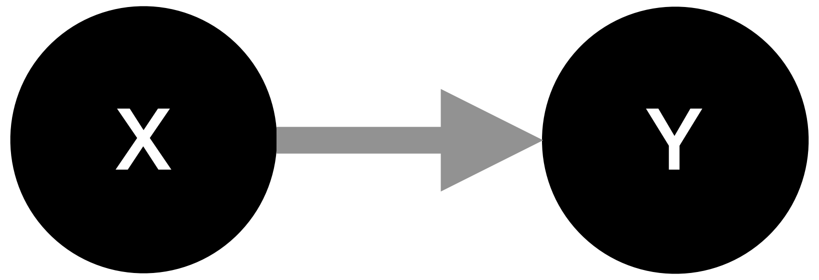
Interaction
X causes A and Y, A causes Y

Causal mechanisms

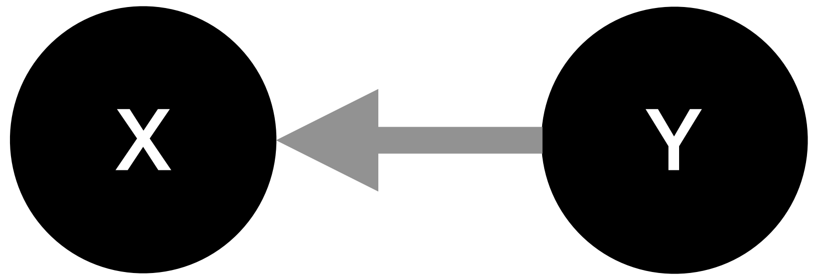


Alternative causal relationships

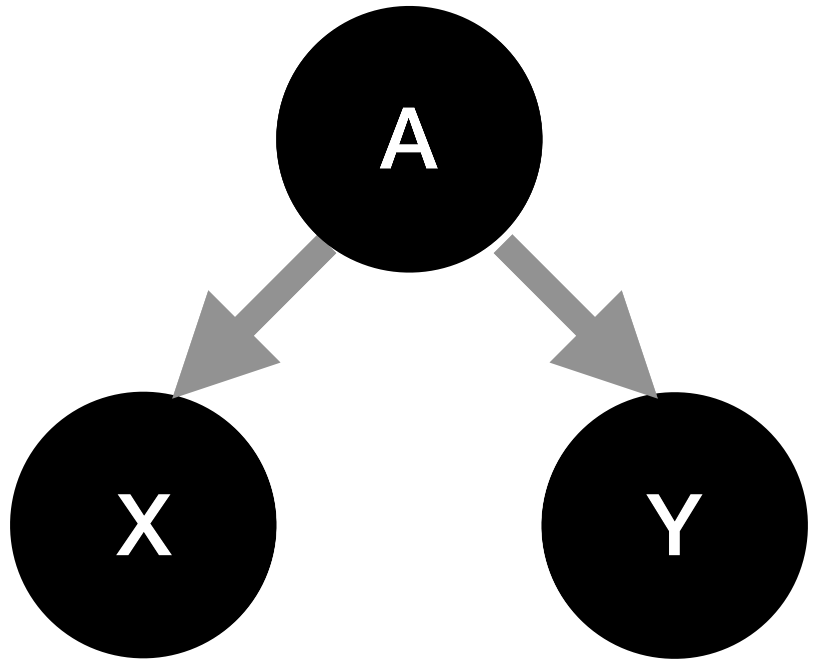
You observe a correlation between space launches and sociology PhDs:



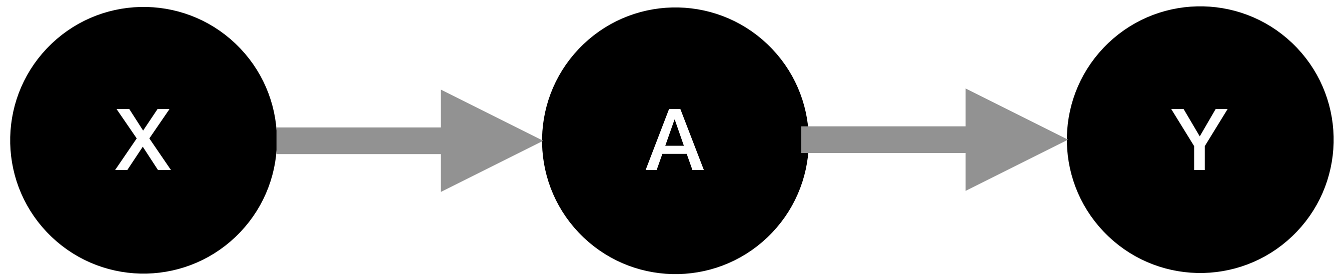
Simple Causation
PhDs cause launches



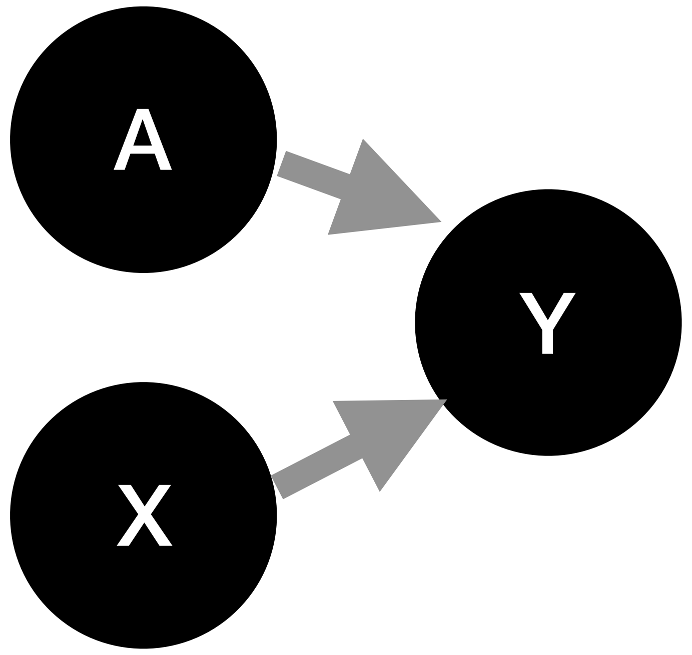
Reverse Causation
Launches cause PhDs



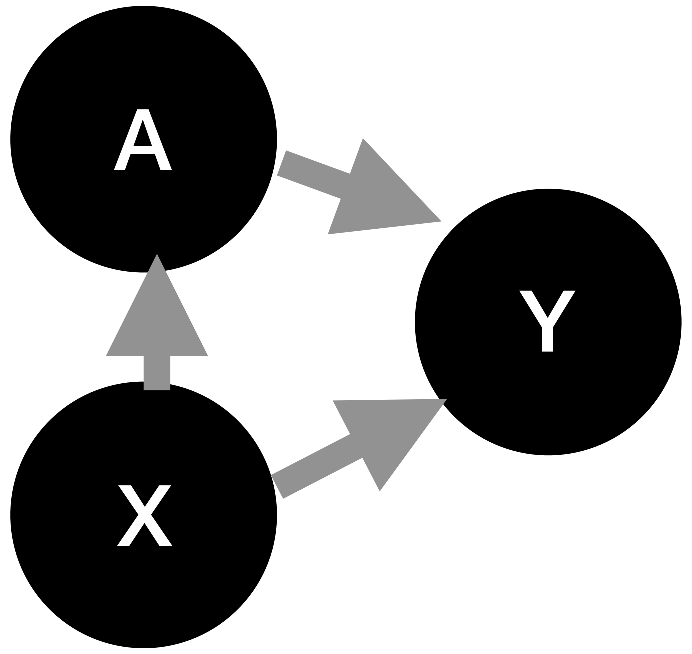
Confounding
Funding causes PhDs and launches



Mediation
PhDs cause researchers cause launches



Multiple causation
PhDs and mineral prices cause launches



Interaction
???

Methods for answering causal questions

Validity

- **Randomization**
    - A/B tests
    - Multi-armed bandits
- **Natural experiments**
    - Regression discontinuity
    - Instrumental variables
- **Conditioning**
    - Stratification, Matching
    - Propensity scores

Ease of use

Adapted from: http://www.github.com/amit-sharma/causal-inference-tutorial

Moving beyond simple experiments

- **Validity**: internal, construct, and external validity
- **Heterogeneous treatment effects**: lots of people and lots of influences

Example causal research questions

- Has cannabis legalization…
  - reduced opioid deaths?
  - reduced border seizures?
  - reduced harder substance abuse?
  - increased DUIs?
  - increased crime around dispensaries?
  - reduced racial bias in policing?
  - generated more jobs?

Internal validity

- There are many factors outside of a researcher's control that can change behavior → how to rule them out?
- Causal requirements
  - *Temporal precedence*: cause has to happen before effect
  - *Covariation*: if more treatment, then more outcome; if less treatment, then less outcome
  - *Alternative explanations:* single group threats, multiple group threats, and social interaction threats
- Experimental design and causal inference methods
  - Blocked, factorial, etc.
  - Matching, differencing, discontinuities, instrumental variables, *etc*.

Threats to internal validity - single group

- RQ: “Does cannabis legalization reduce opioid overdoses?”
- Design: Measure per-capita opioid overdoses in before and after 2014 legalization

- **History**: Legalization didn’t cause reduction in overdoses, an unrelated 2014 mental health program did
- **Maturation**: Risky behavior goes down as people age, overdoses went down because population got older
- **Testing**: Asking about opioid consumption before 2014 caused people to consume/overdose
- **Instrumentation**: Method for measuring opioid overdoses changed between 2010 and 2018
- **Mortality**: People reporting in 2010 left Colorado before 2014, so measuring different people afterwards
- **Regression**: Opioid overdoses were unusually high around 2010, they could only come down

Threats to internal validity – multiple group

- A similar control group that doesn't receive treatment is ideal, but has its own challenges:
  - Treatment and control groups must be comparable → selection threats cause groups to not be comparable
- Design #2: Measure opioid overdoses before and after 2014 legalization by comparing to opioid overdoses in a similar state that did not legalize (Colorado vs. Michigan)

- **History**: Population in CO reacts to Obama's 2012 re-election differently than MI, CO has less abuse
- **Maturation**: Population in CO matures faster than MI, MI has greater risk taking behavior around opioids
- **Testing**: Pre-2014 surveys caused MI people to be more likely to start abusing opioids
- **Instrumentation**: Method for measuring overdoses differs between CO and MI
- **Mortality**: At-risk people in CO are more likely to move than MI, more of them to drop out of statistics
- **Regression**: MI had unusually high rates that had to come down, regardless of treatment/control

Threats to internal validity – social interaction

- People do not exist in isolation, their interactions with others can interfere with the experiment
- Design #2: Measure opioid overdoses before and after 2014 legalization by comparing to opioid overdoses in a similar state that did not legalize (Colorado vs. Michigan)

- **Diffusion/Imitation of Treatment**: Comparison state sees CO effects, MI cannabis prohibition becomes lax
- **Compensatory Rivalry**: Comparison state sees CO effects, MI starts new program to reduce overdoses
- **Resentful Demoralization**: Comparison state sees CO effects, MI increases prohibition policies
- **Compensatory Equalization**: Federal agency sees CO effects, increases support to MI

What's different about digital experiments?

- Similarities between classic lab and digital experiments
  - Recruiting participants, randomization, treatment & control, measurement
- Differences
  - Fully digital experiments have (close to) zero marginal cost
  - Constraints on size are not cost or logistics, but increasingly ethics

Four strategies for digital experiments

	Cost	Control	Realism	Ethics
Partner with the powerful	low	medium	high	complex
Use existing systems	low	low	high	complex
Build an experiment	medium	high	medium	simple
Build a product	high	high	high	simple

Partner with the powerful

- Work with Facebook, Amazon, Google, *etc*.
- **Cost**: Personal relationships: internships, sabbaticals, collaborations
- **Control**: Wizard of Oz (lots of control) but don't kill the golden goose (limitations on what managers will agree to: partnership, not extraction)
- **Realism**: Large-scale social systems engaged in meaningful behavior
- **Ethics**: Governmentality mindsets likely prevails over participatory design

Use existing systems

- Deploy field experiments: Wikipedia, Twitter, Mechanical Turk, Reddit, *etc*.
- **Cost**: Creating accounts, interacting through APIs
- **Control**: Multiple validity threats from noisy social system
- **Realism**: Large scale social systems engaged in meaningful behavior
- **Ethics**: Fidelity *vs*. informed consent? Disclosure and debriefing important

Build an experiment

- Salganik's Music Lab, Centola's health communities, *etc*.
- **Cost**: Building out own technical — but temporary — infrastructure
- **Control**: Precisely control recruitment, treatments, instrumentation, *etc.*
- **Realism**: Engaging enough to sustain motivation, but some contrivances
- **Ethics**: Informed consent baked into recruitment, few competing incentives

Build a product

- MovieLens, FoldIt, *etc*.
- **Cost**: Building and sustaining own social and technical infrastructure
- **Control**: Wizard of Oz and only ecological constraints
- **Realism**: More users, more research, better product
- **Ethics**: Informed consent baked into recruitment, few competing incentives

Replace, Refine, Reduce

- **Replace** experiments with less invasive methods (EDA, natural experiments, *etc*.)

- **Refine** treatments to make them less harmful (boost, don't block)

- **Reduce** number of participants (lower risk of harm, more volunteers)