In-person session 4

February 6, 2025

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

Plan for today

Logic models and evaluation

More regression things

Measuring outcomes

DAGs

Exam 1 details

Final project details

Extra bonus things

Logic models and evaluation

Do people really have the
job title "program evaluator"?

How much does this evaluation stuff cost?

Can you do scaled-down versions
of these evaluations?

Isn't it best to always have an articulated theory?

Should implicit theory and articulated theory be the same thing in most cases?

What if a program exists already and doesn't have a logic model?

Why would a program aim for final outcomes that can't be measured?

What should you do if you find that your theory of change (or logic model in general) is wrong in the middle of the program?
Is it ethical to stop or readjust?

How does regression relate to impact evaluation?

More regression things

Do we care about the actual coefficients or just whether or not they're significant?

How does significance relate to causation?

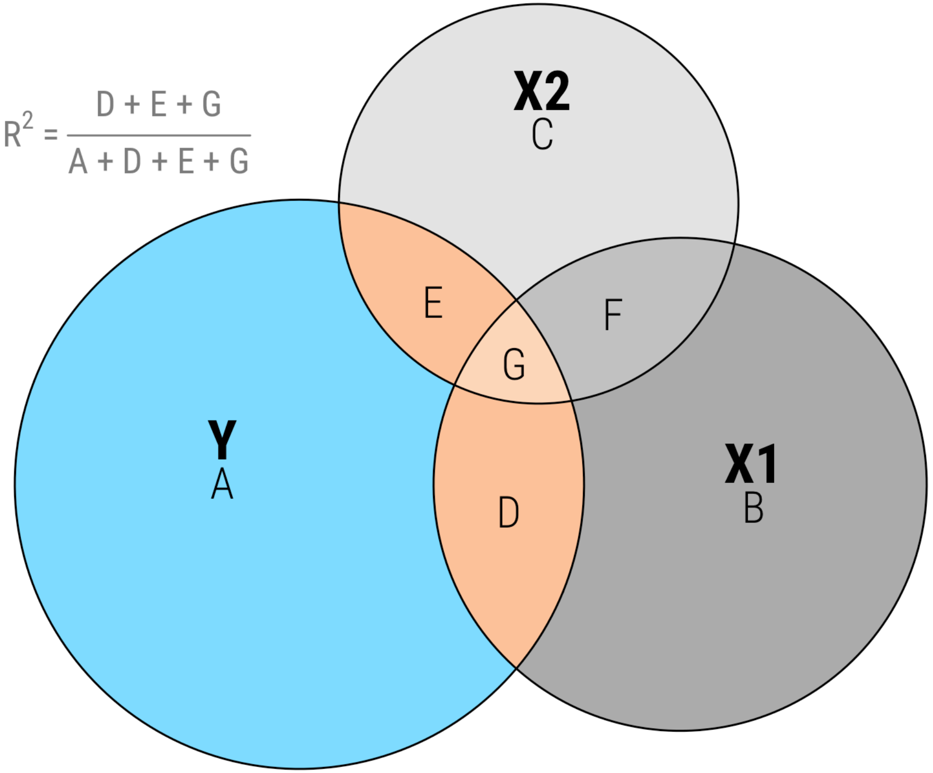
If we can't use statistics to assert causation how are we going to use this information in program evaluation?

What counts as a "good" R^2?

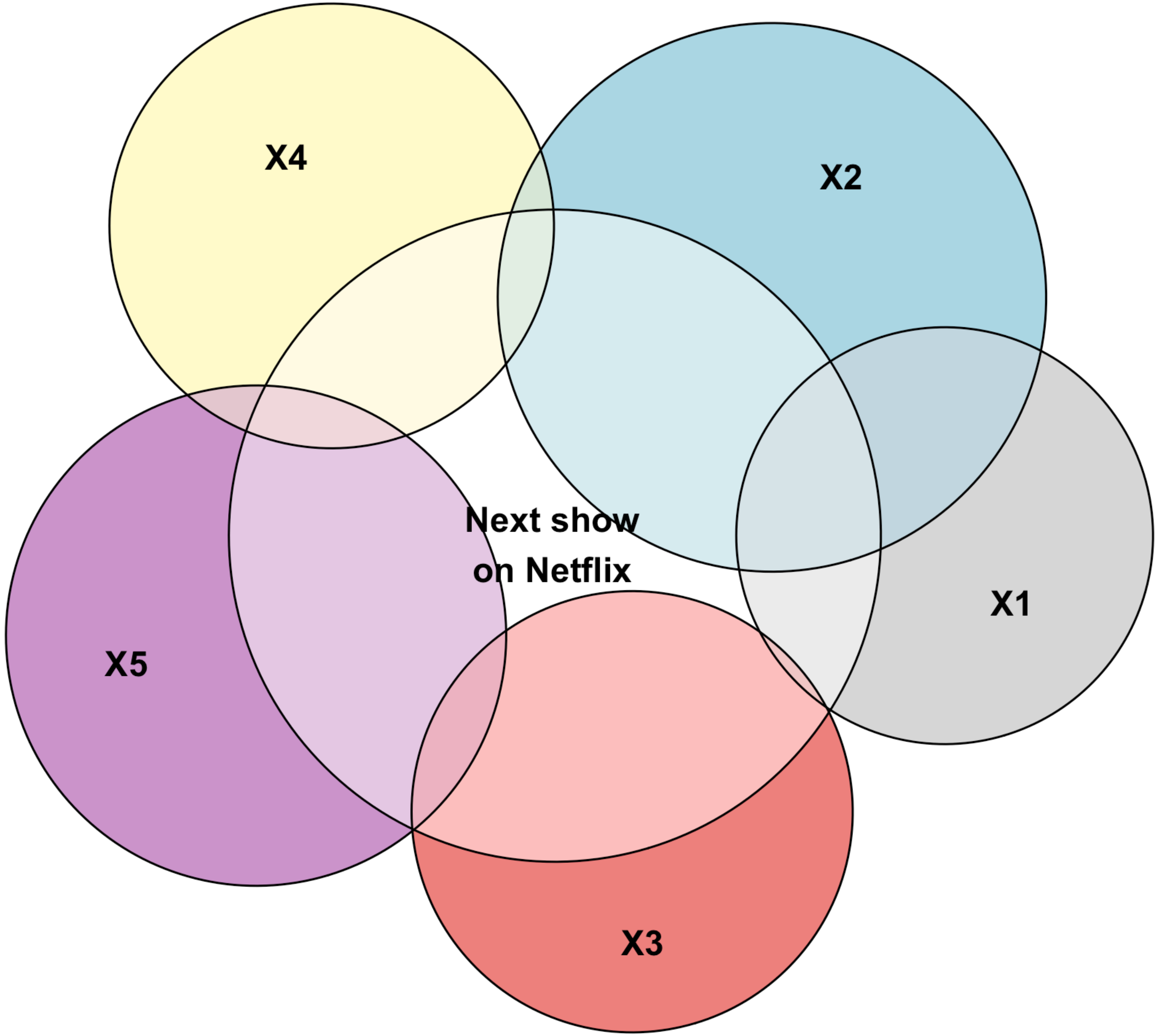
R^2 represented as an Euler diagram

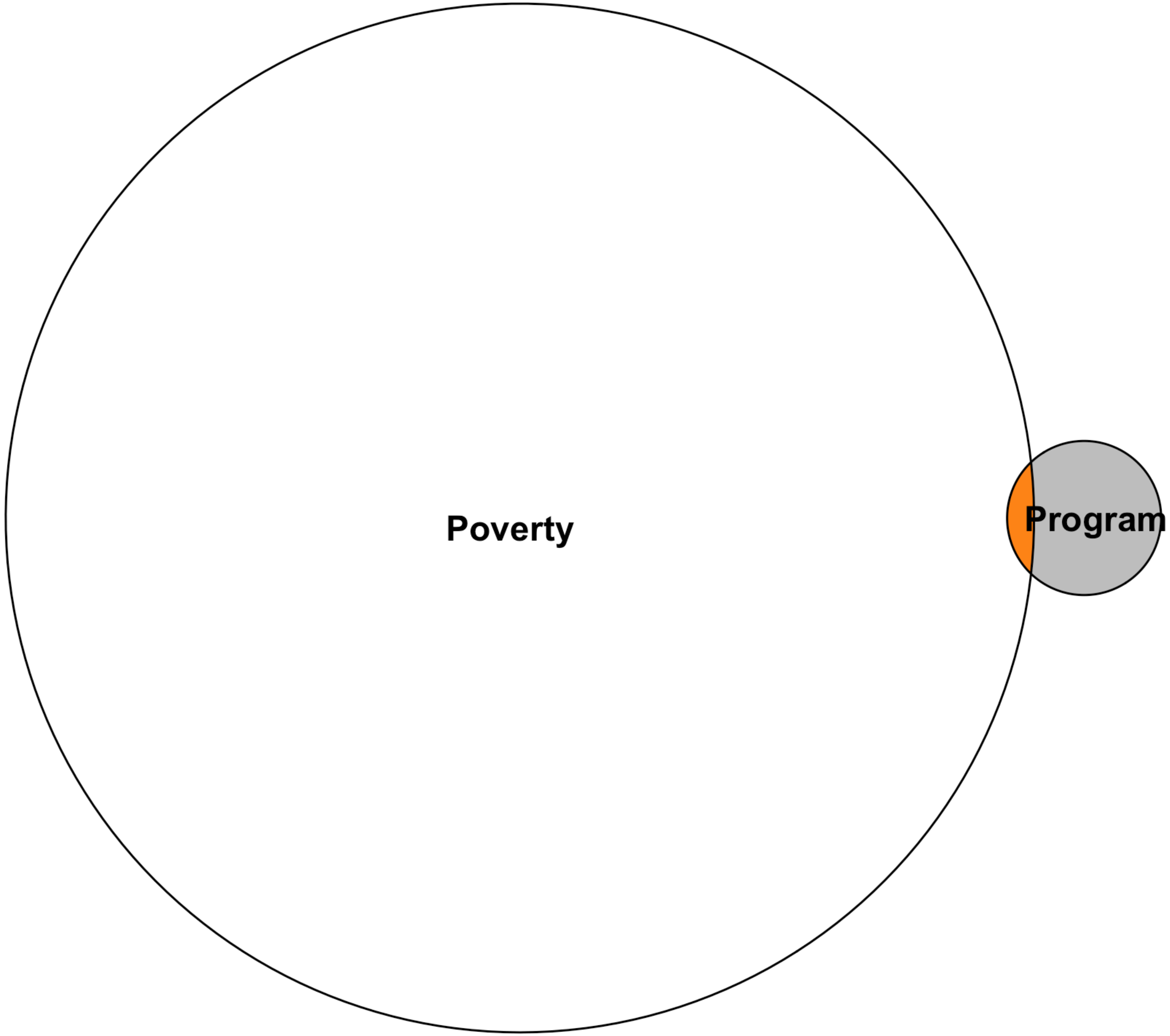
Orange area (D + E + G) shows the total variance in outcome Y that is jointly explained by X1 and X2

$$R^2 = \frac{D + E + G}{A + D + E + G}$$



Circles sized according to each variable's sum of squares; size of overlapping areas is not 100% correct due to limitations in available geometric space

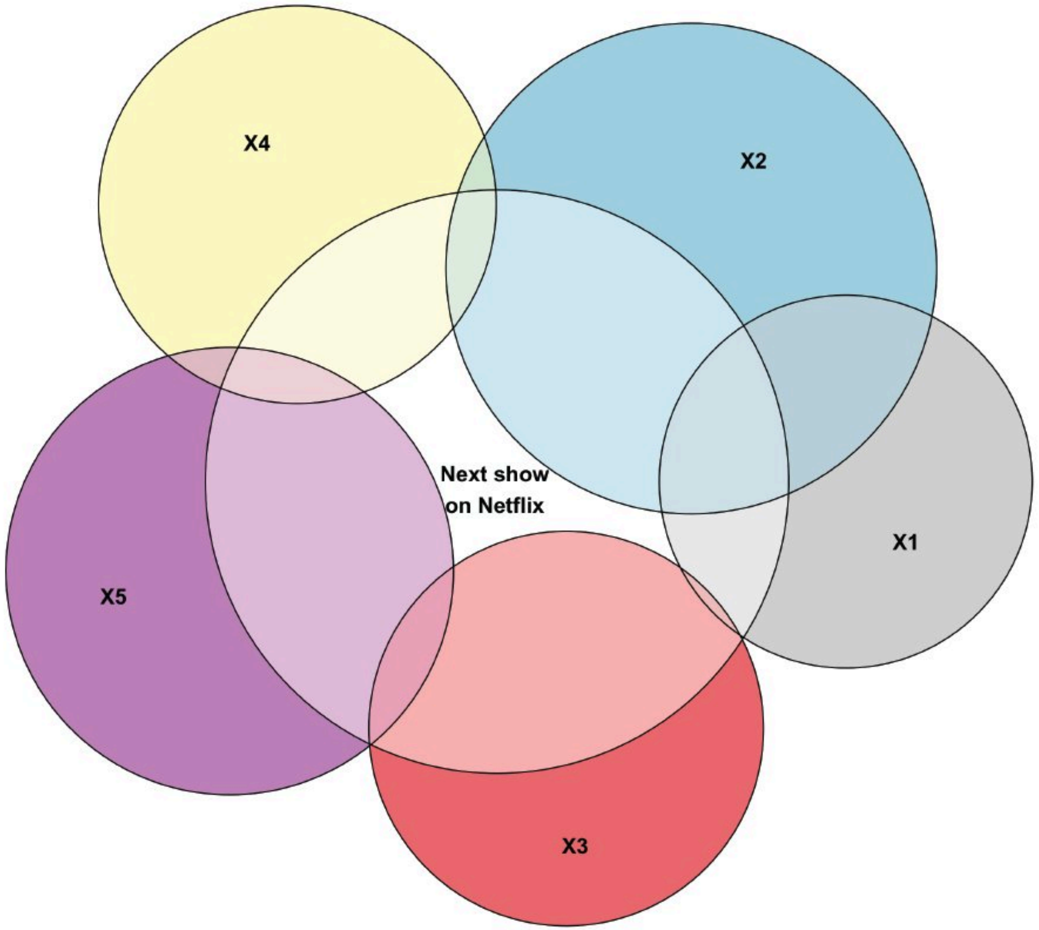

X4
X2
Next show
on Netflix
X1
X5
X3


Poverty
Program

Regression focused on prediction

Focus is on Y

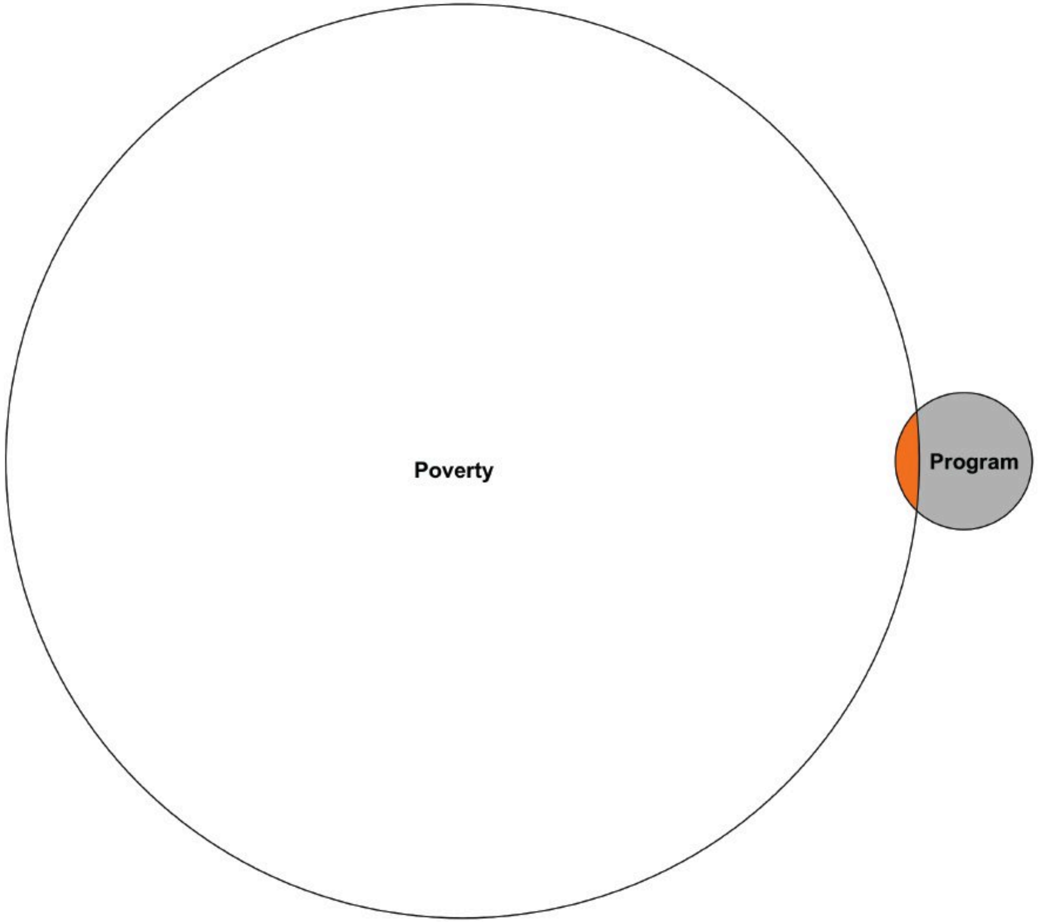
Minimize unexplained variation in the outcome



Regression focused on estimation

Focus is on a single X

Get that little sliver as accurate as possible



Side-by-side regression tables

	(1)	(2)	(3)	(4)
(Intercept)	362.307	-5780.831***	-5736.897***	-5433.534***
	(283.345)	(305.815)	(307.959)	(286.558)
bill_length_mm	87.415***		6.047	-5.201
	(6.402)		(5.180)	(4.860)
flipper_length_mm		49.686***	48.145***	48.209***
		(1.518)	(2.011)	(1.841)
sexmale				358.631***
				(41.572)
Num.Obs.	342	342	342	333
R2	0.354	0.759	0.760	0.807
R2 Adj.	0.352	0.758	0.759	0.805
AIC	5400.0	5062.9	5063.5	4863.3
BIC	5411.5	5074.4	5078.8	4882.4
Log.Lik.	-2696.987	-2528.427	-2527.741	-2426.664
F	186.443	1070.745	536.626	457.118
RMSE	643.54	393.12	392.34	353.66

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

See full documentation and examples for `modelsummary()` here

Make nicer tables with {tinytable}

Measuring outcomes

The paradox of evaluation

Evaluation is good, but expensive

"Evaluation thinking"

Too much evaluation is bad

Taming programs

Outcomes and programs

Outcome variable

Thing you're measuring

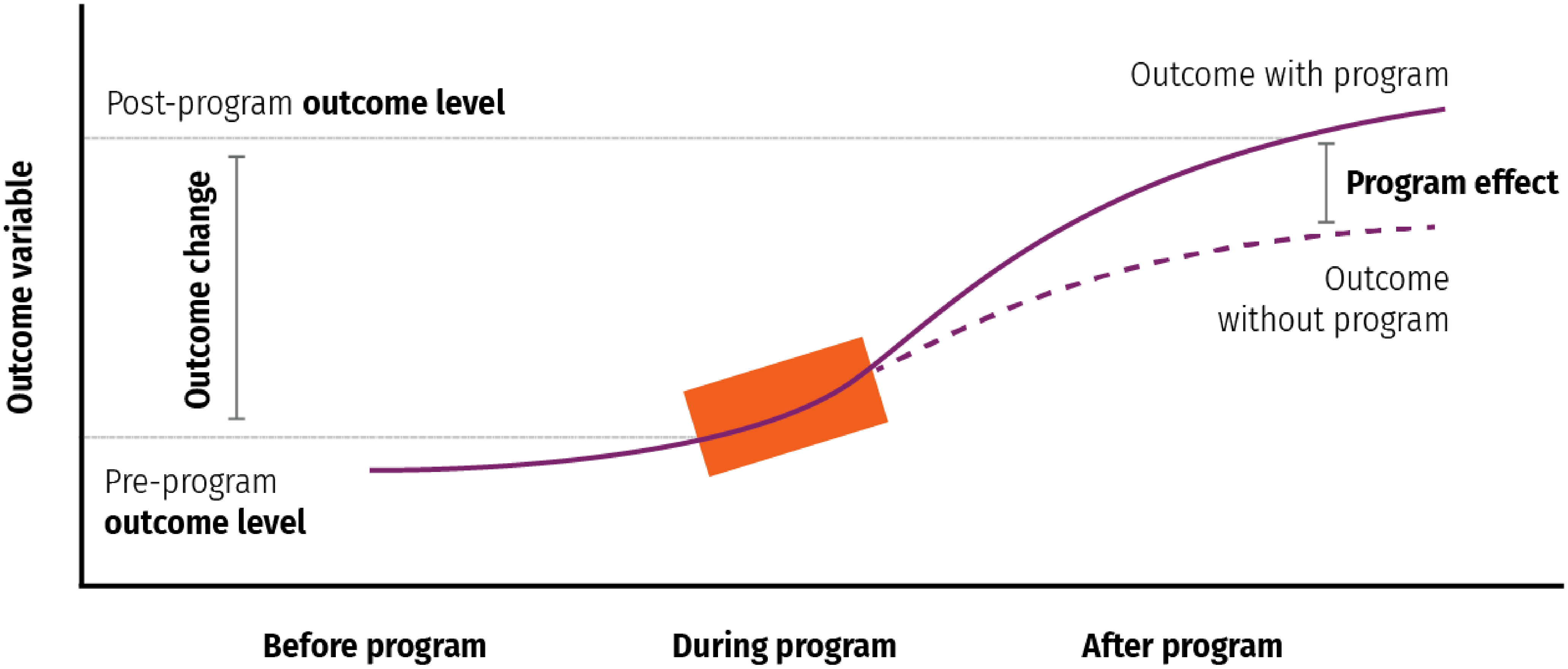
Outcome change

∆ in thing you're measuring over time

Program effect

∆ in thing you're measuring over time *because of* the program

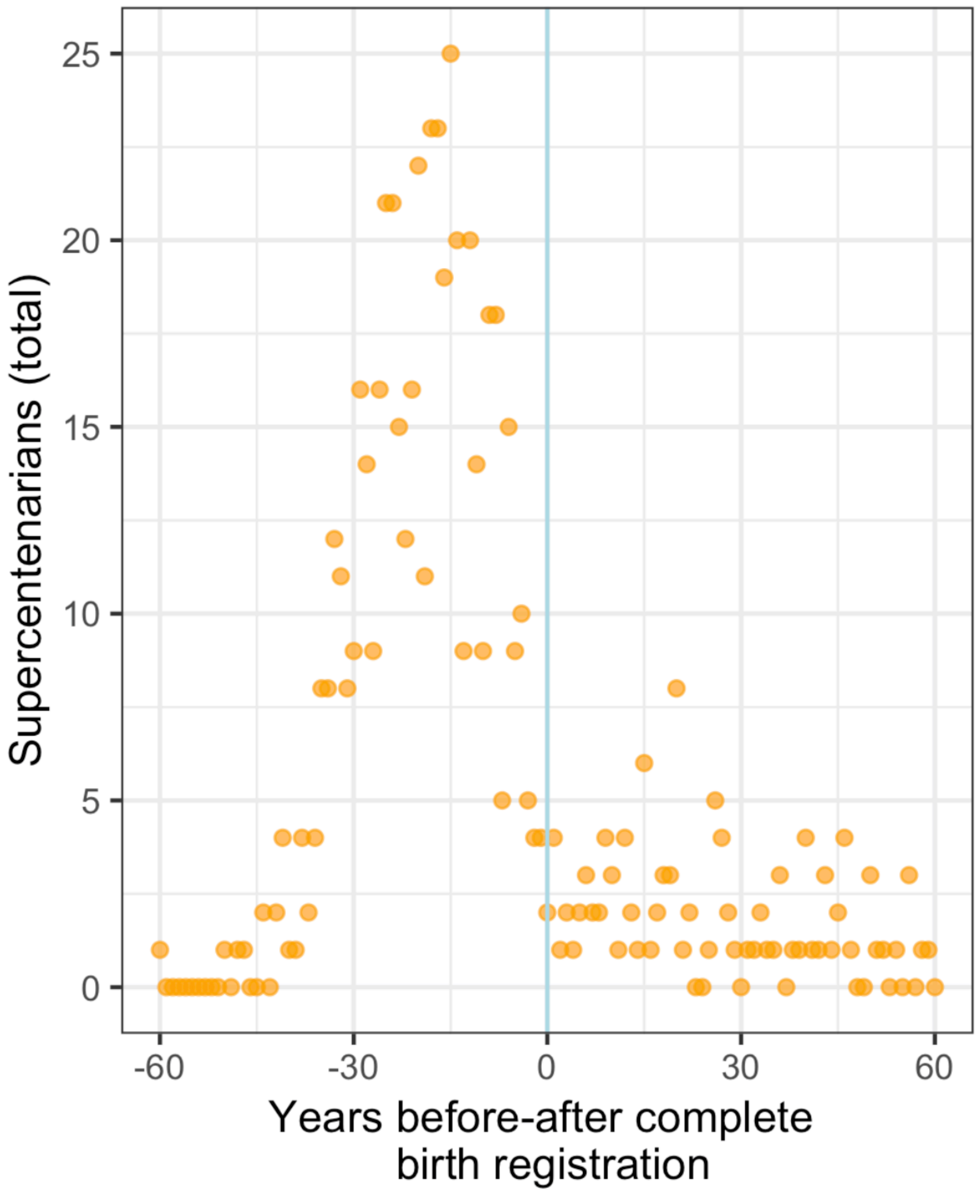
Outcomes and programs



Abstraction

DAGs

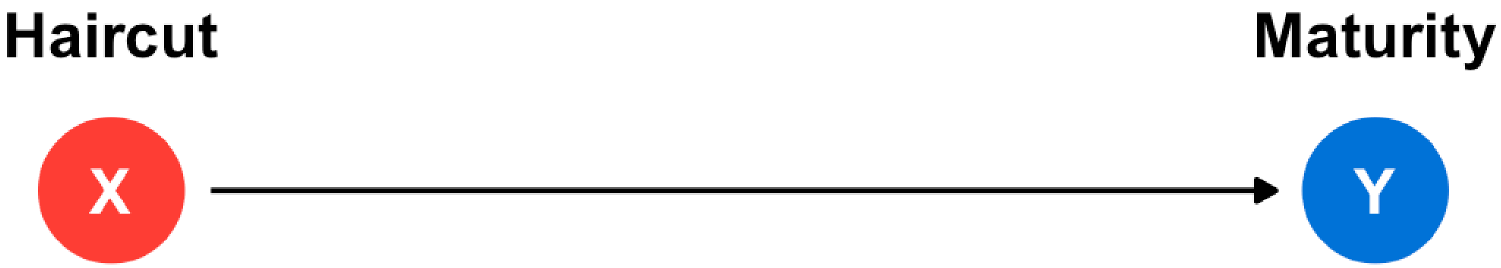
Causal thinking is necessary—
even for descriptive work!



Supercentenarians (total)
25
20
15
10
5
0
-60
-30
0
30
60
Years before-after complete
birth registration

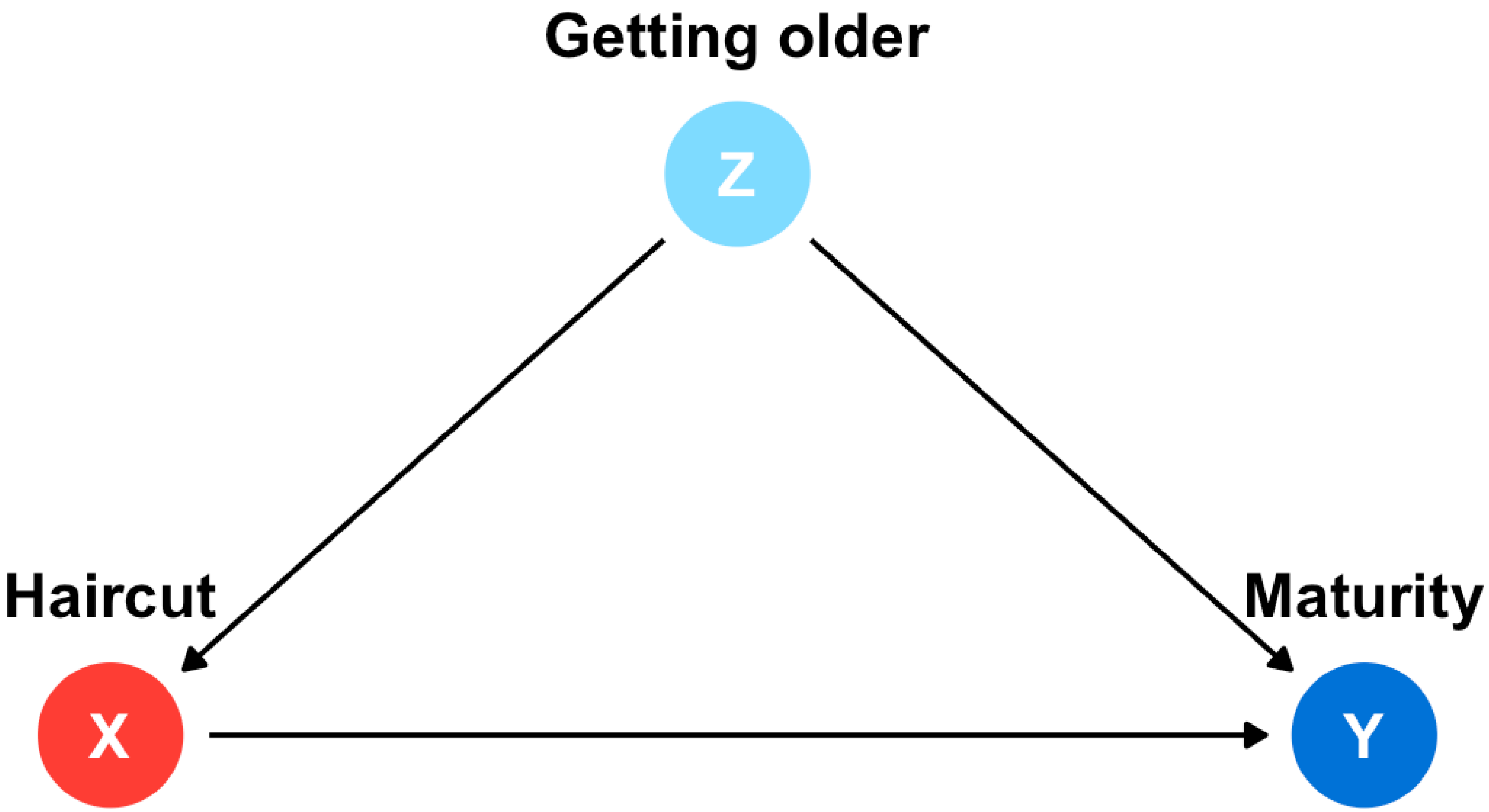
"Every time I get a haircut, I become more mature!"

"Every time I get a haircut, I become more mature!"

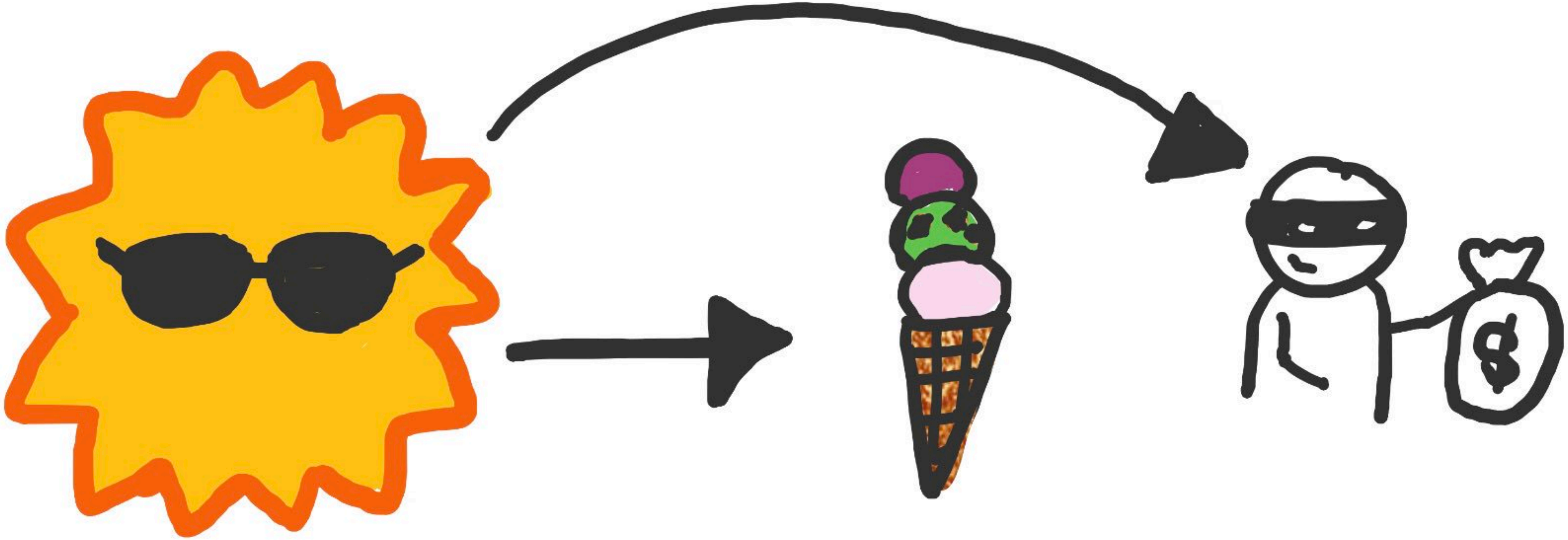


$$E[\text{Maturity} \mid \text{do}(\text{Get haircut})]$$

Getting older opens a backdoor path



Ice cream causes crime

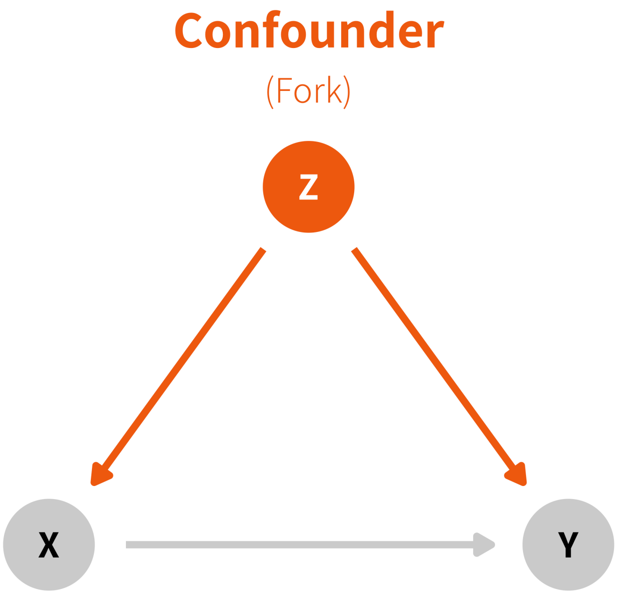


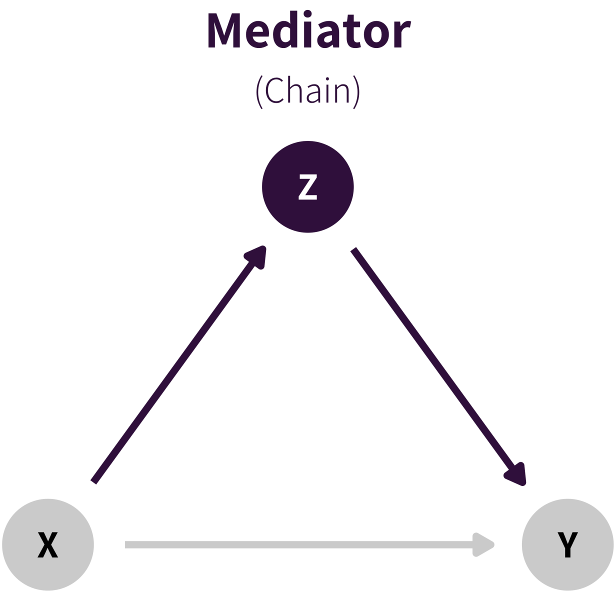
Summer weather opens a backdoor path

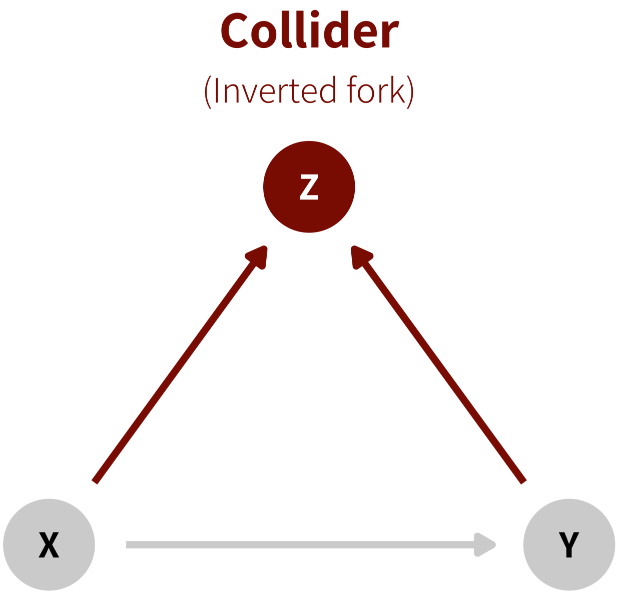
But what does that mean, "opening a backdoor path"?

How does statistical association get passed through paths?

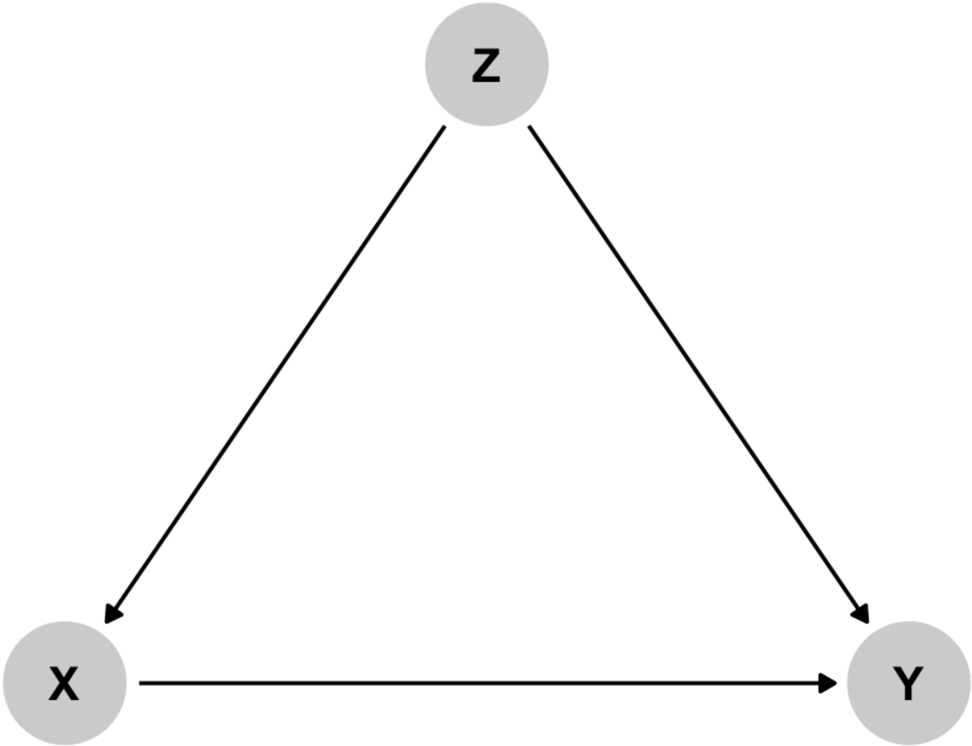
How do I know which of these is which?



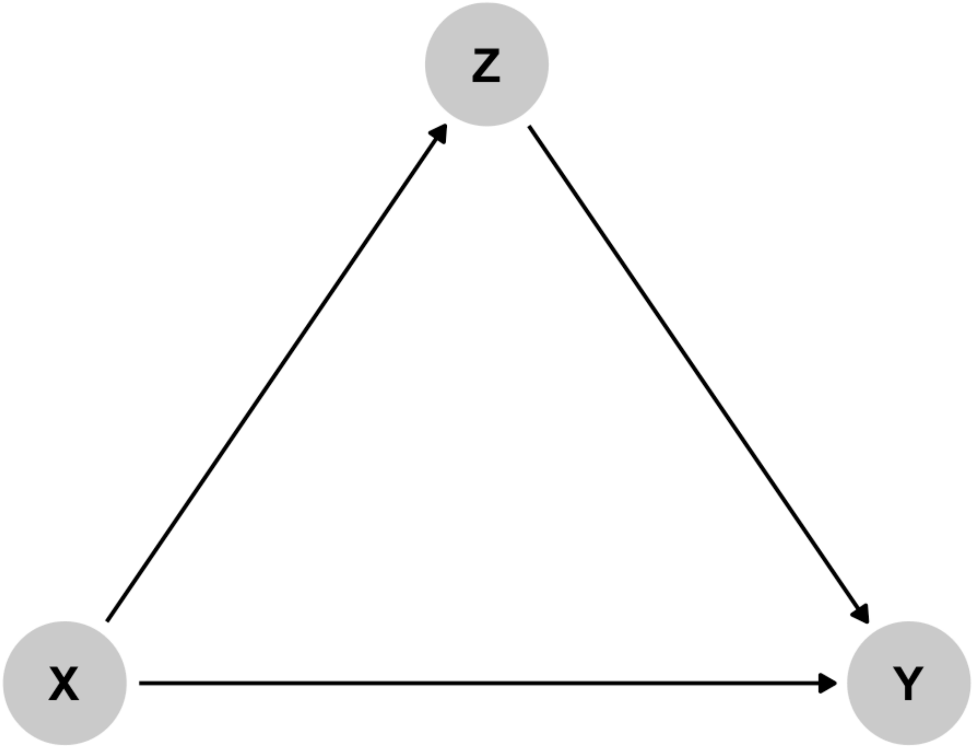




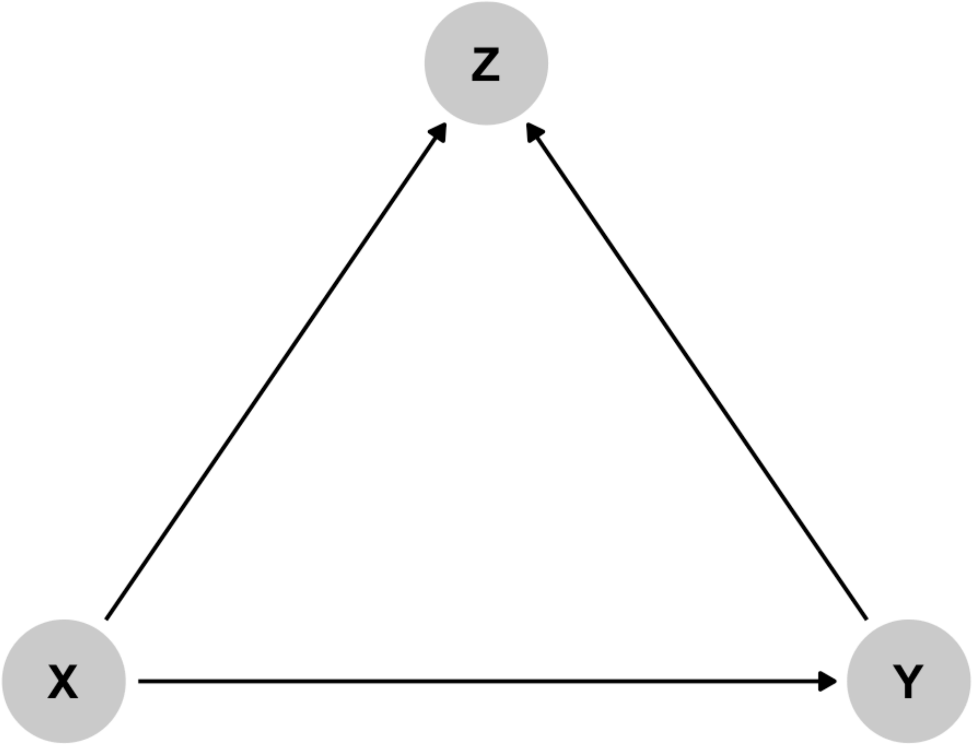

LUTRON


Z
X
Y

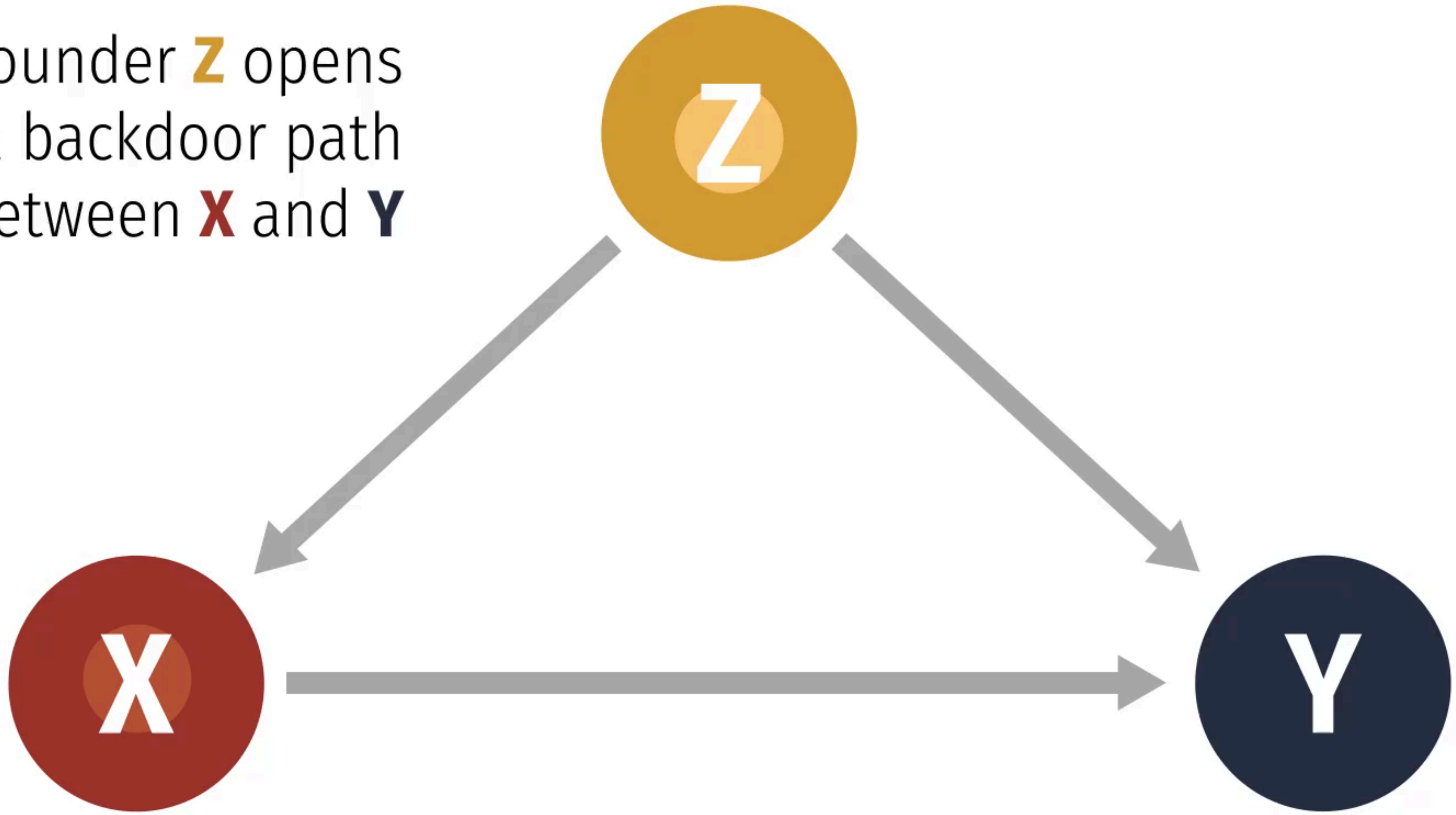

LUTRON


Z
X
Y


LUTRON

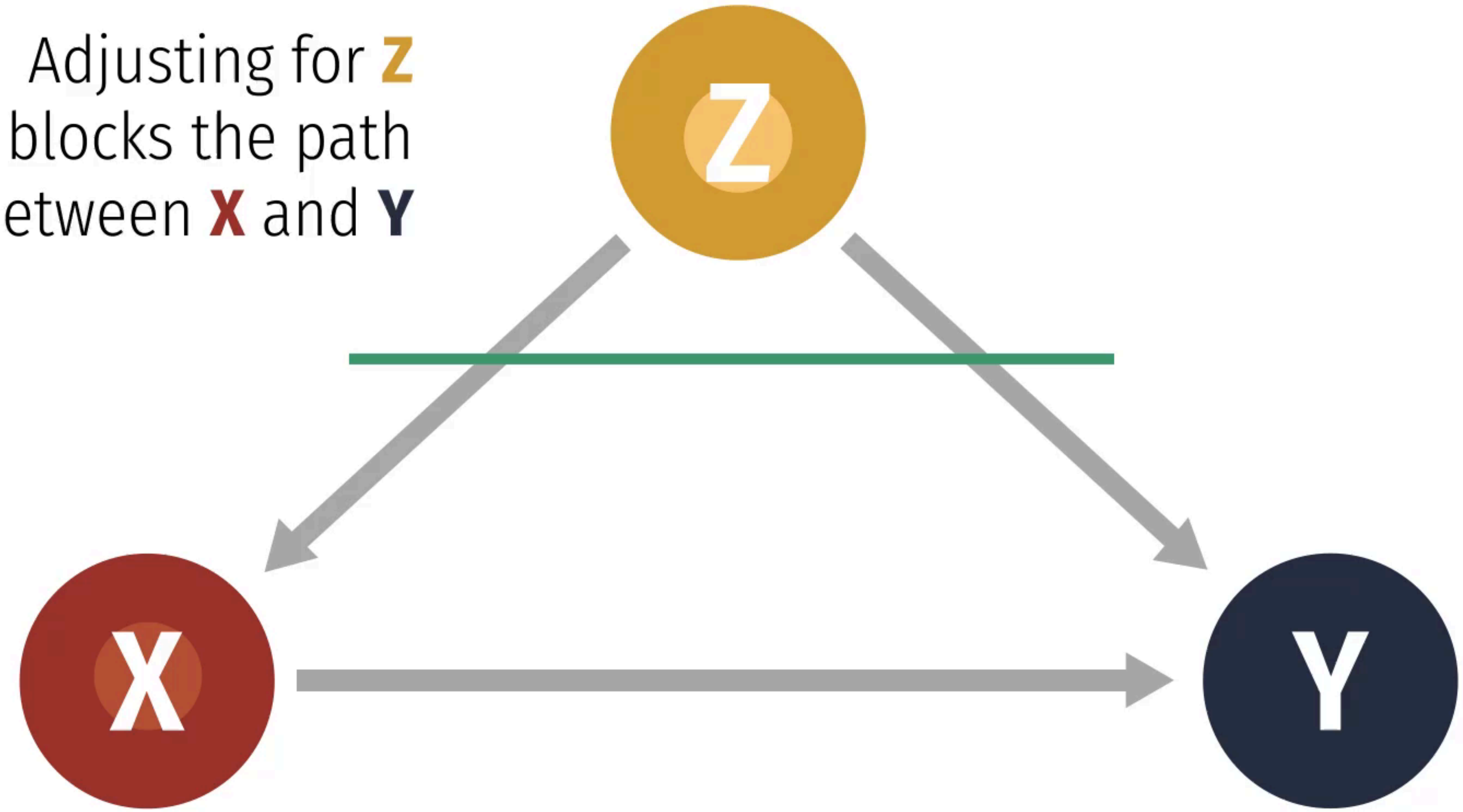

Z
X
Y

Confounder Z opens a backdoor path between X and Y



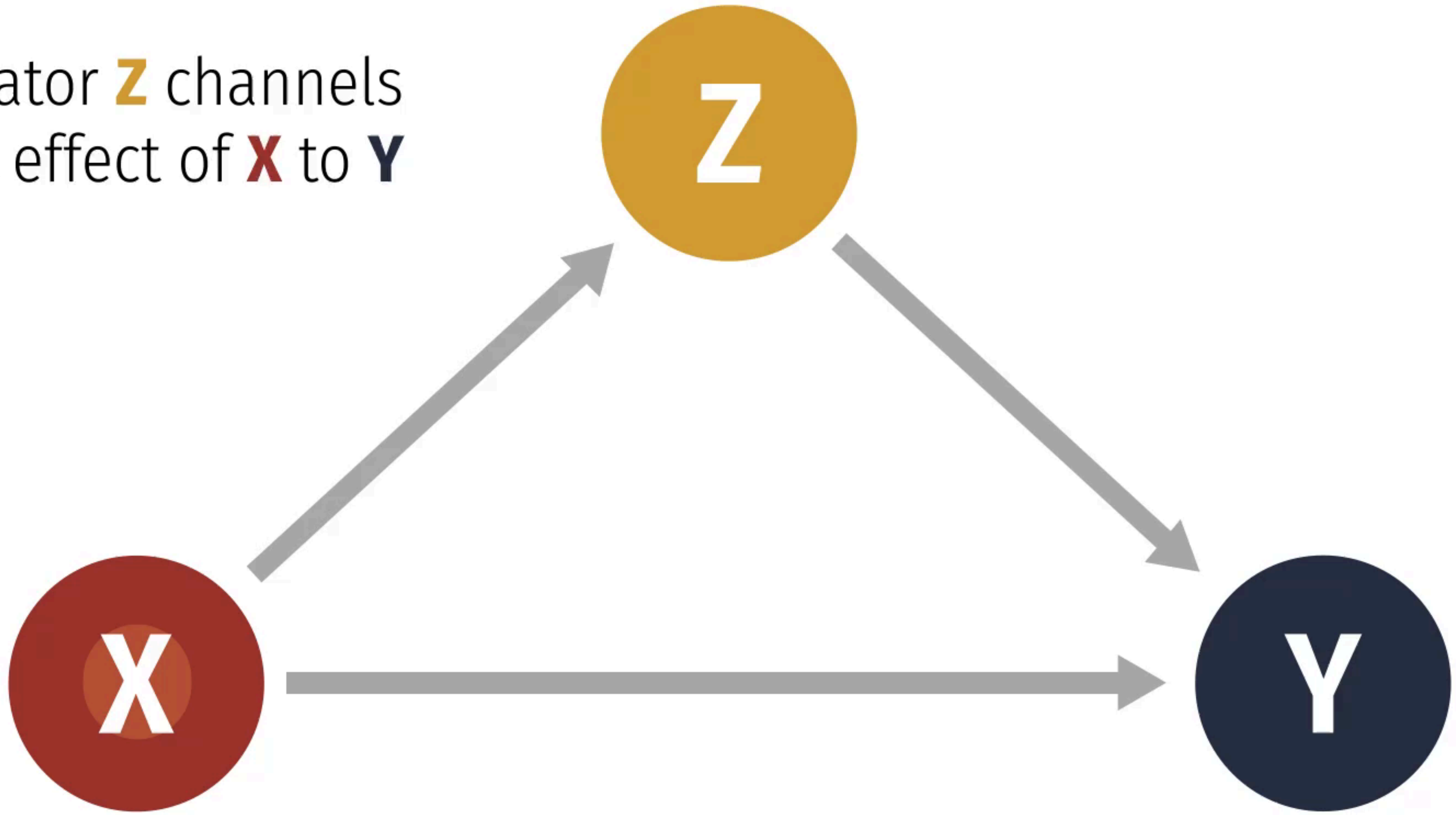
Y ~

Adjusting for Z blocks the path between X and Y



Y ~

Mediator Z channels indirect effect of X to Y



Y ~

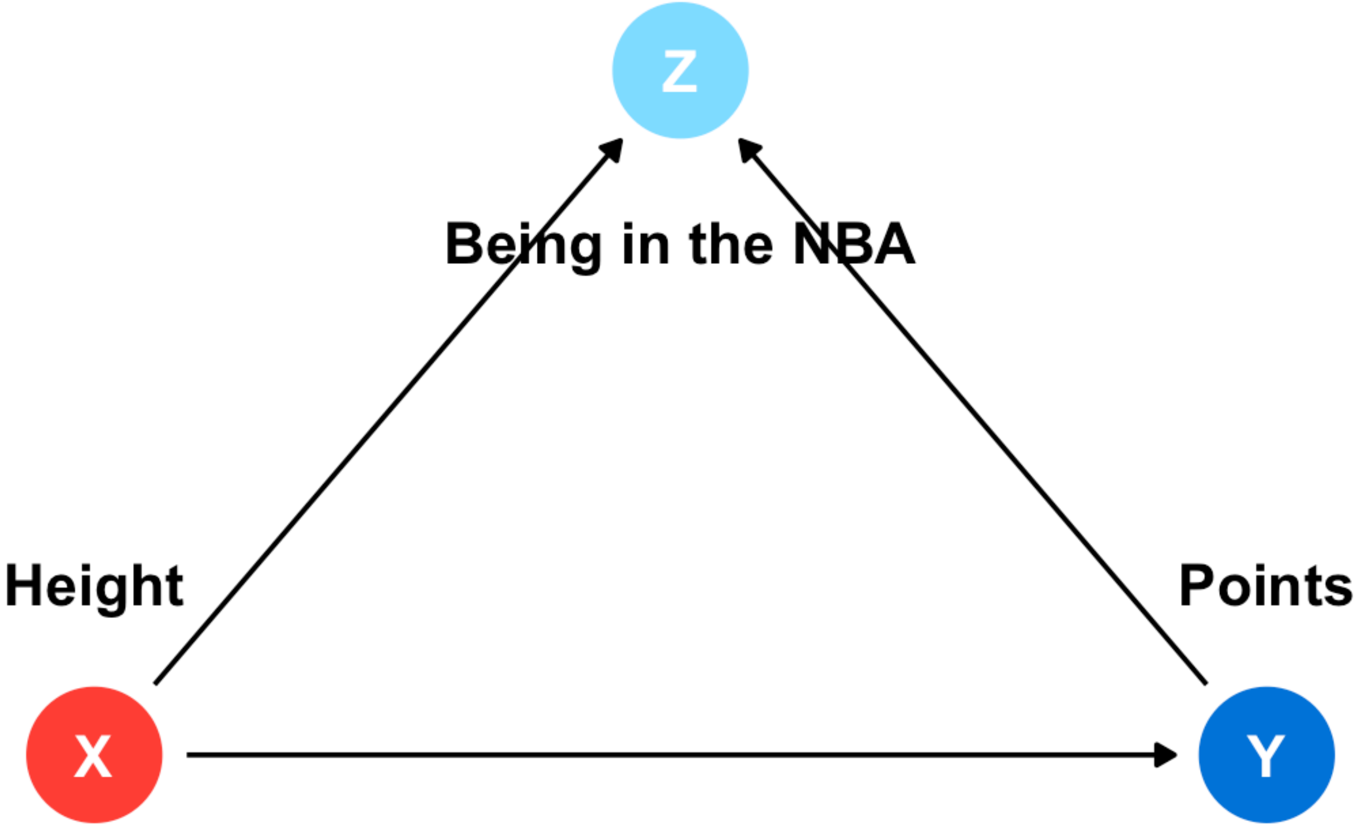
d-separation

Except for the one arrow between X and Y, no statistical association can flow between X and Y

This is *identification*— all alternative stories are ruled out and the relationship is isolated

How exactly do colliders mess up your results?

It looks like you can still get the effect of X on Y

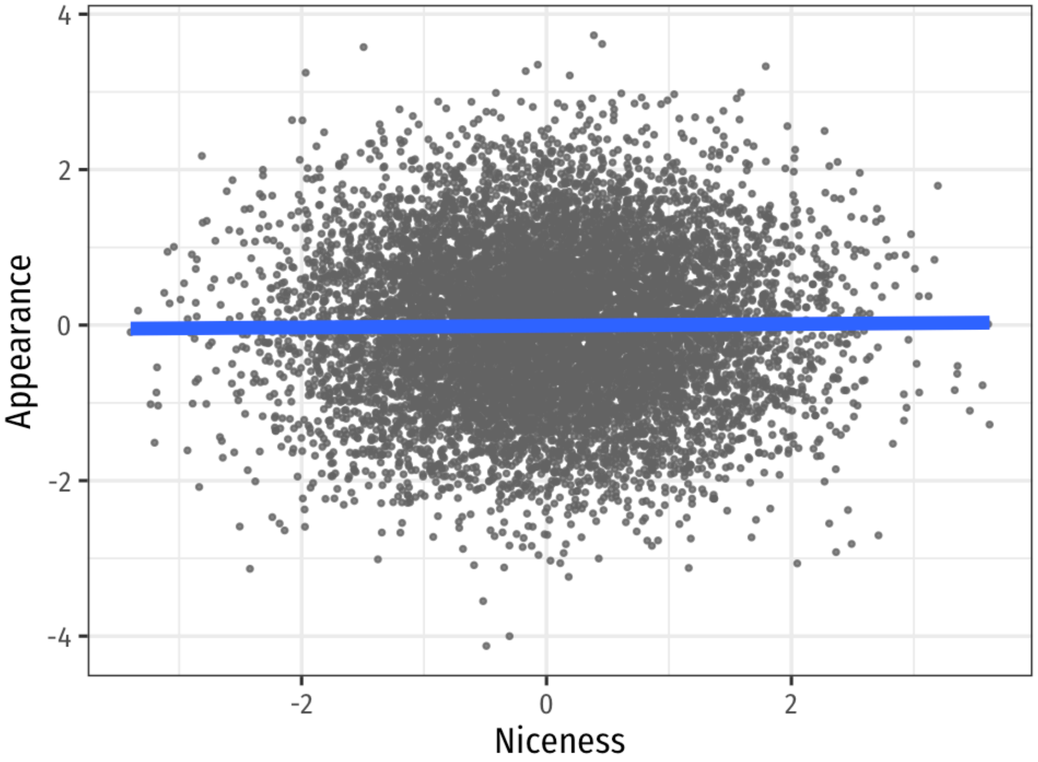

Z
Being in the NBA
Height
Points
X
Y

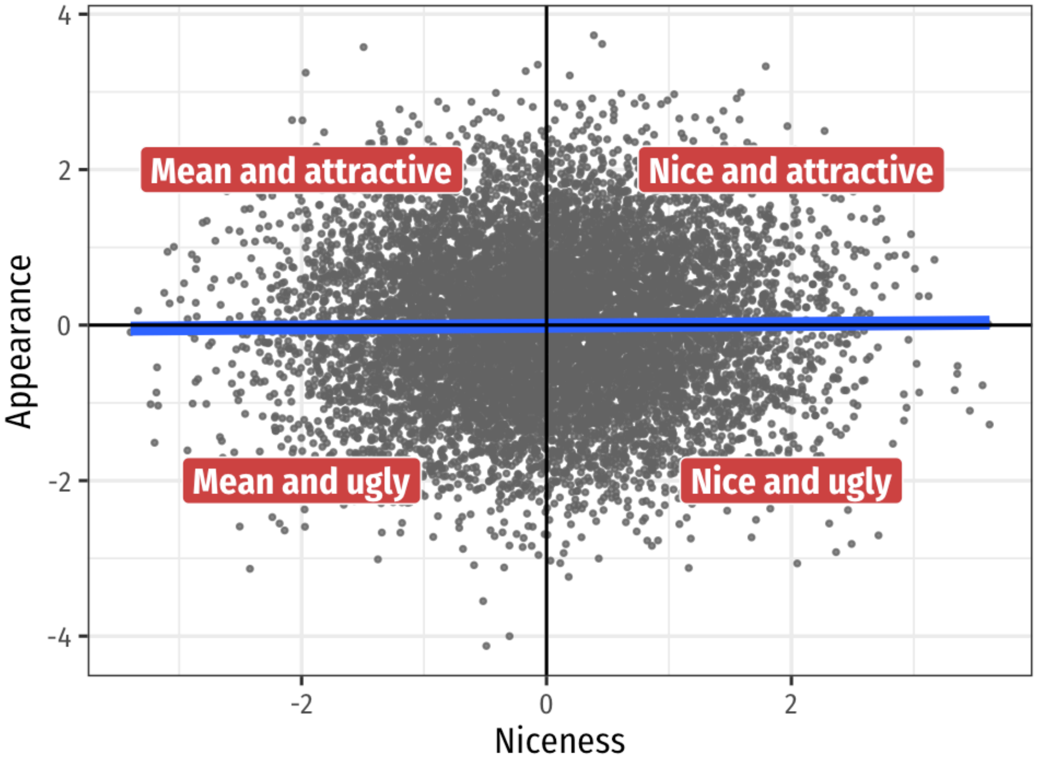
Sept. 10, 2021, 3:58 p.m. ET
By Davey Alba

Facebook sent flawed data to misinformation researchers.

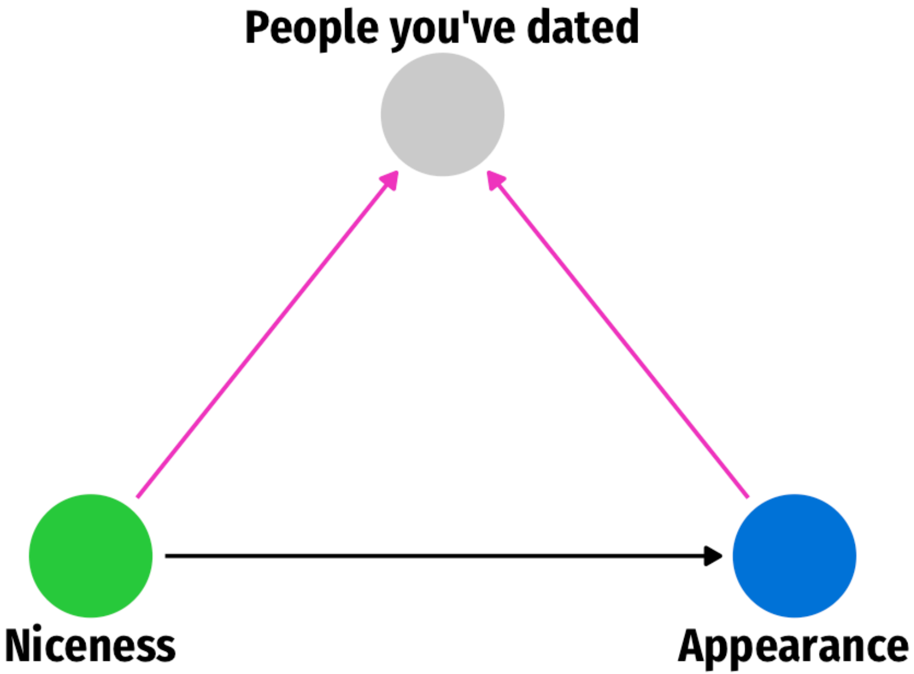
Mark Zuckerberg, chief executive of Facebook, testifying in Washington in 2018. Tom Brenner/The New York Times

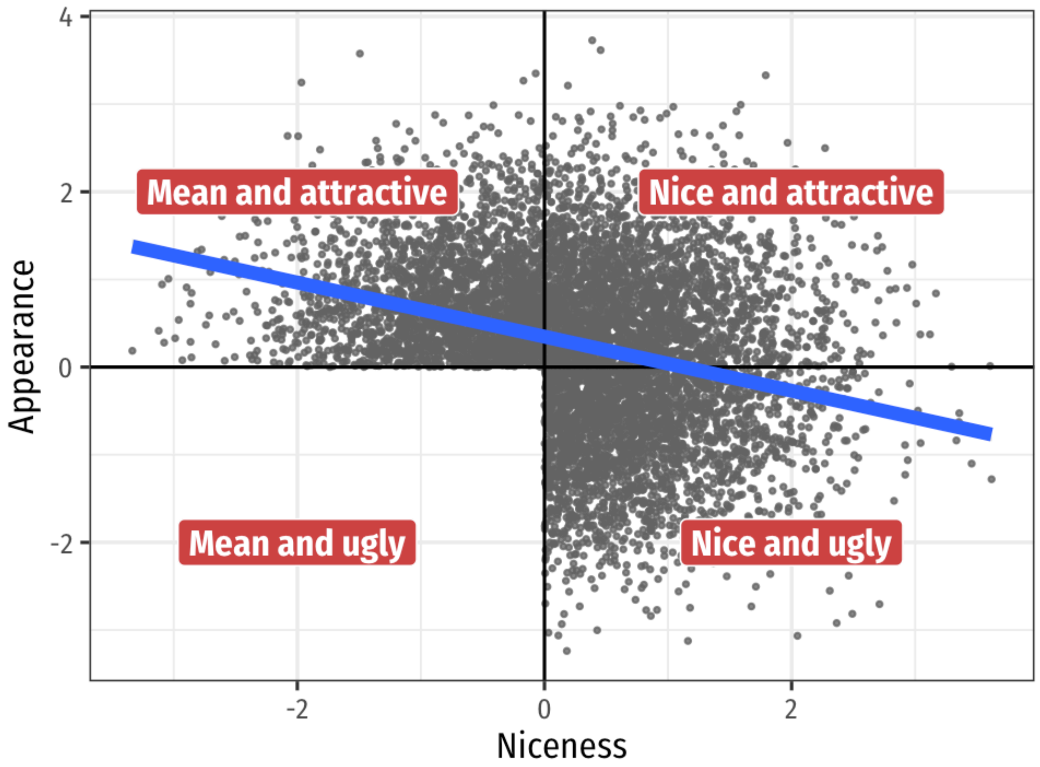
Does niceness improve appearance?





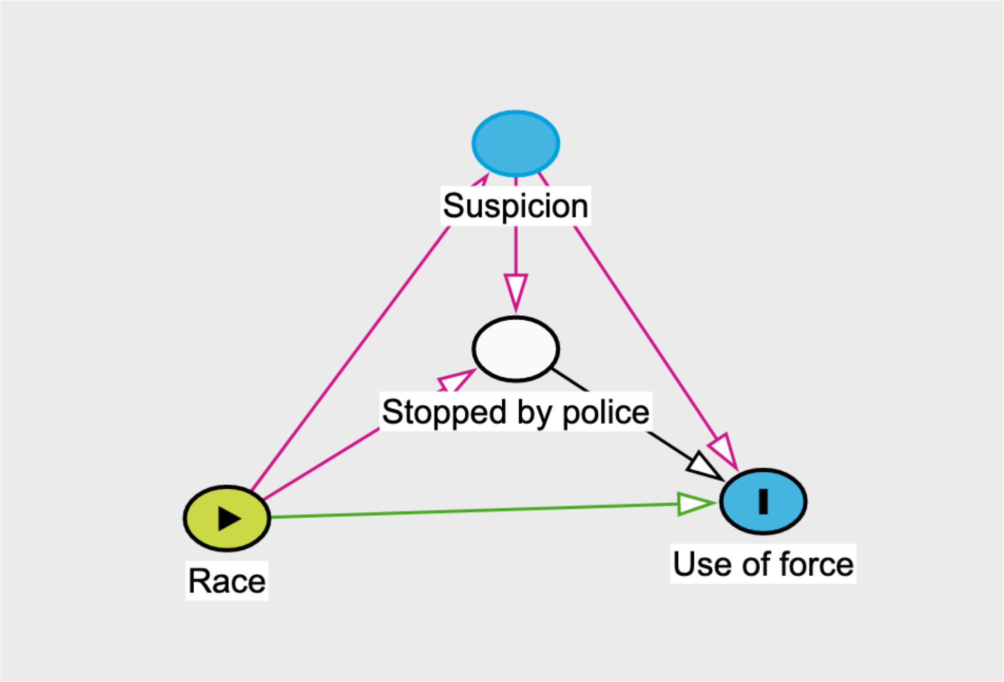
Collider distorts the true effect!





Effect of race on police use of force using administrative data

Effect of race on police use of force using administrative data



American Political Science Review, Page 1 of 19
doi:10.1017/S0003055420000039

Administrative Records Mask Racially Biased Policing

DEAN KNOX *Princeton University*
WILL LOWE *Hertie School of Governance*
JONATHAN MUMMOLO *Princeton University*

Researchers often lack the necessary data to credibly estimate racial discrimination in policing. In particular, police administrative records lack information on civilians police observe but do not investigate. In this article, we show that if police racially discriminate when choosing whom to investigate, analyses using administrative records to estimate racial discrimination in police behavior are statistically biased, and many quantities of interest are unidentified—even among investigated individuals—absent strong and untestable assumptions. Using principal stratification in a causal mediation framework, we derive the exact form of the statistical bias that results from traditional estimation. We develop a bias-correction procedure and nonparametric sharp bounds for race effects, replicate published findings, and show the traditional estimator can severely underestimate levels of racially biased policing or mask discrimination entirely. We conclude by outlining a general and feasible design for future studies that is robust to this inferential snare.

Concern over racial bias in policing, and the public availability of large administrative data sets documenting police–civilian interactions, have prompted a raft of studies attempting to quantify the effect of civilian race on law enforcement behavior. These studies consider a range of outcomes including ticketing, stop duration, searches, and the use of force (e.g., Antonovics and Knight 2009; Fryer 2019; Ridgeway 2006; Nix et al. 2017). Most research in this area attempts to adjust for omitted variables that may correlate with suspect race and the outcome of interest. In contrast, this study addresses a more fundamental problem that remains even if the vexing issue of omitted variable bias is solved: the inevitable statistical bias that results from studying racial discrimination using records that are themselves the product of racial discrimination (Angrist and Pischke 2008; Elwert and Winship 2014; Rosenbaum 1984). We show that when there is any biased absent additional data and/or strong and untestable assumptions.

This study makes several contributions. We clarify the causal estimands of interest in the study of racially discriminatory policing—quantities that many studies appear to be targeting, but are rarely made explicit—and show that the conventional approach fails to recover any known causal quantity in reasonable settings. Next, we highlight implicit and highly implausible assumptions in prior work and derive the statistical bias when they are violated. We proceed to develop informative nonparametric sharp bounds for the range of possible race effects, apply these in a reanalysis and extension of a prominent article on police use of force (Fryer 2019), and present bias-corrected results that suggest this and similar studies drastically underestimate the level of racial bias in police–civilian interactions. Finally, we outline strategies for future data collection and re-