DAGs and the Causal Revolution

Types of data

Experimental

You have control over which units get treatment

Observational

You don't have control over which units get treatment

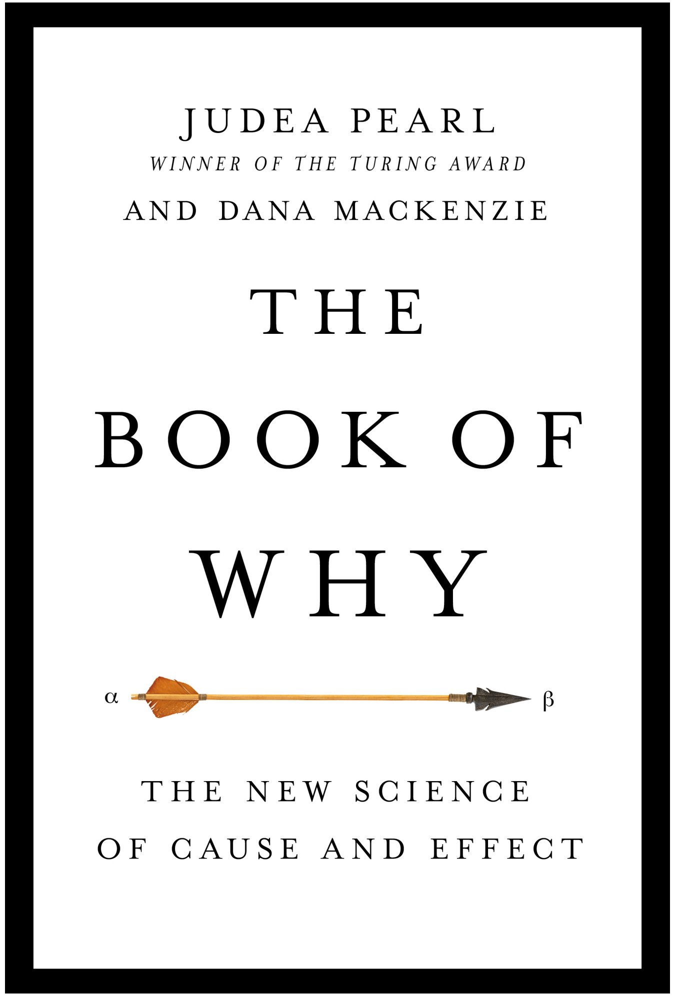
Which kind lets you prove causation?

Causation with observational data

Can you prove causation with observational data?

Why is it so controversial to use observational data?

The causal revolution



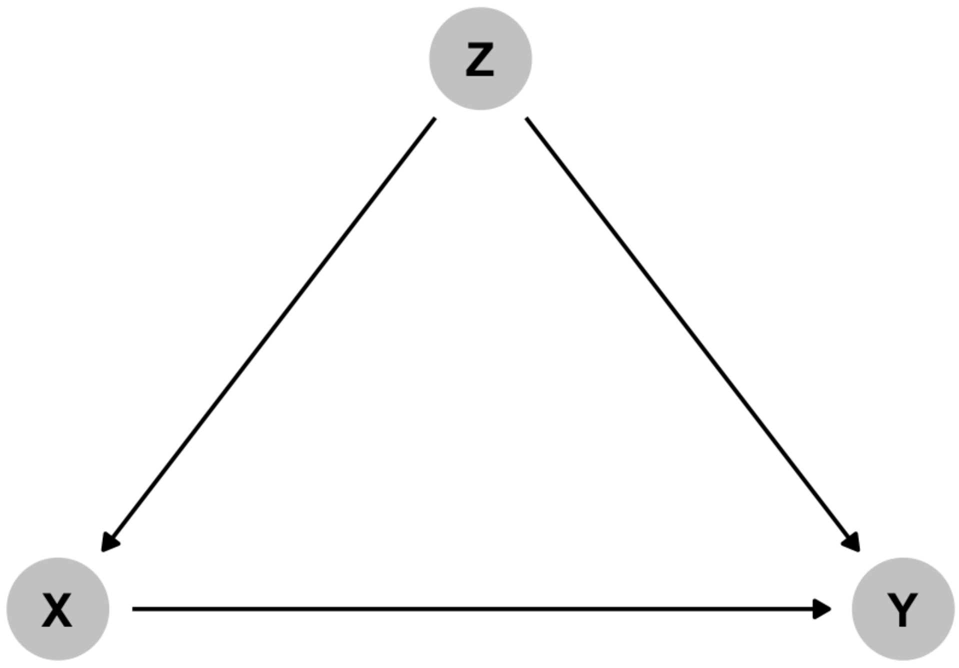
Causal diagrams

Directed acyclic graphs (DAGs)

Graphical model of the process that generates the data

Maps your philosophical model

Fancy math ("*do*-calculus") tells you what to control for to isolate and identify causation



How to draw a DAG

What is the causal effect of an additional year of education on earnings?

Step 1: List variables

Step 2: Simplify

Step 3: Connect arrows

Step 4: Use logic and math to determine which nodes and arrows to measure

1. List variables

Education (treatment) $\longrightarrow$ Earnings (outcome)

Location

Ability

Demographics

Socioeonomic status

Year of birth

Compulsory schooling laws

Job connections

2. Simplify

Education (treatment) $\longrightarrow$ Earnings (outcome)

Location

Ability

Demographics

Socioeonomic status

Year of birth

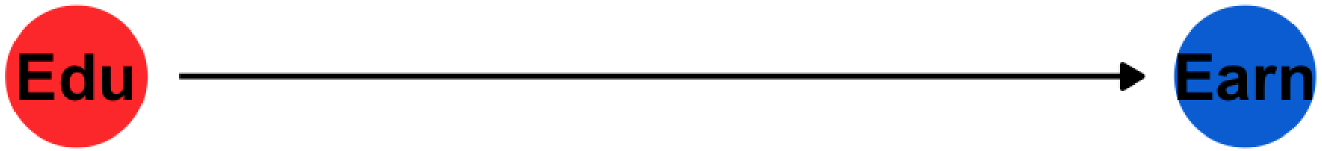
Compulsory schooling laws

Job connections

Background

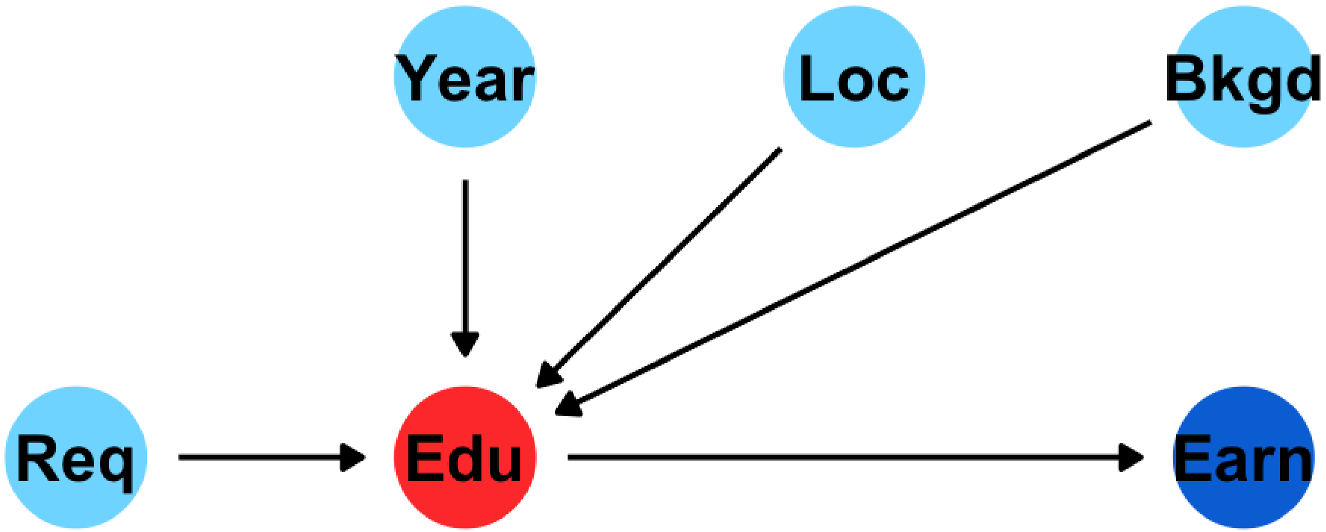
3. Draw arrows

Education causes earnings



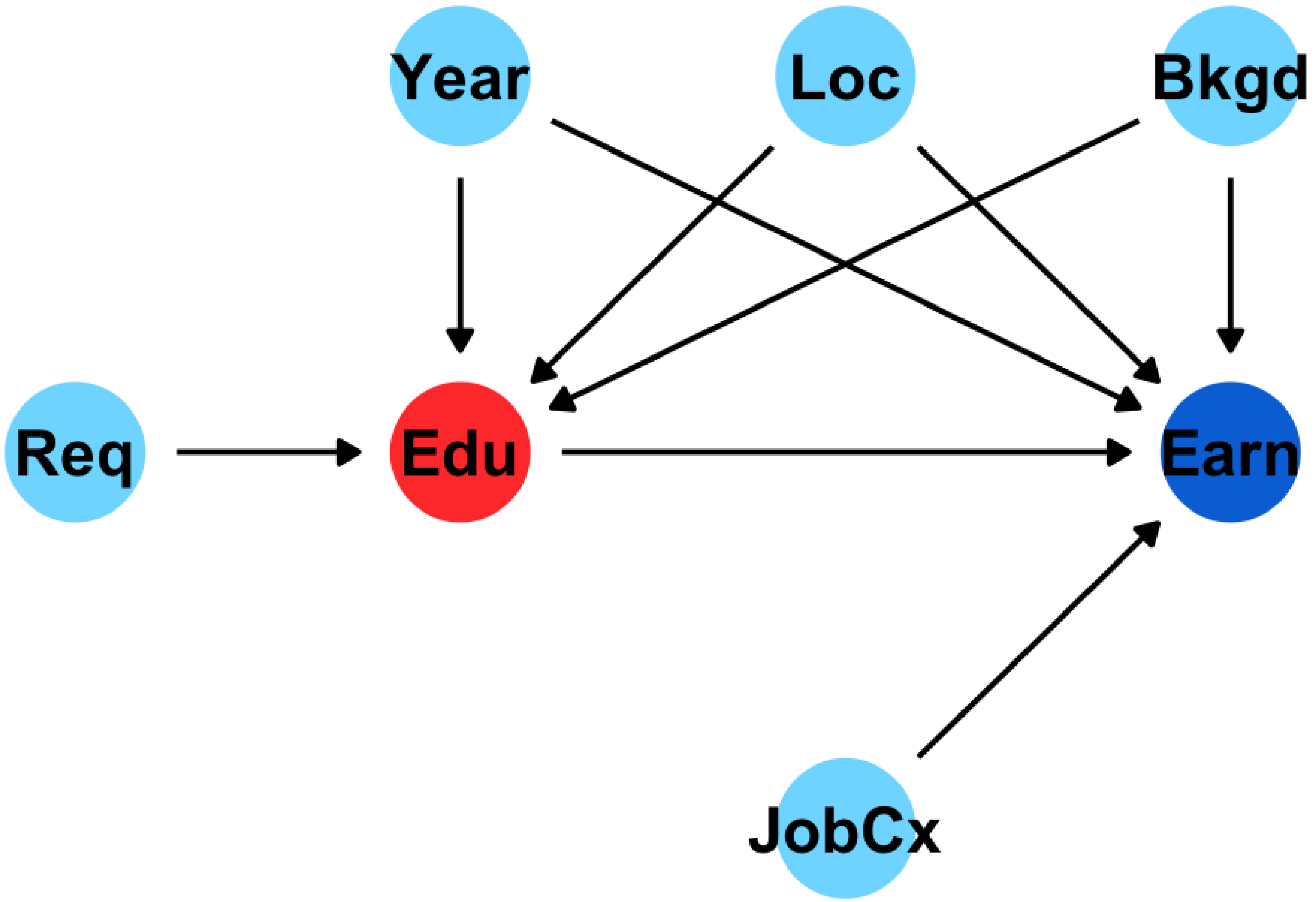
3. Draw arrows

Background, year of birth, location, job connections, and school requirements all cause education



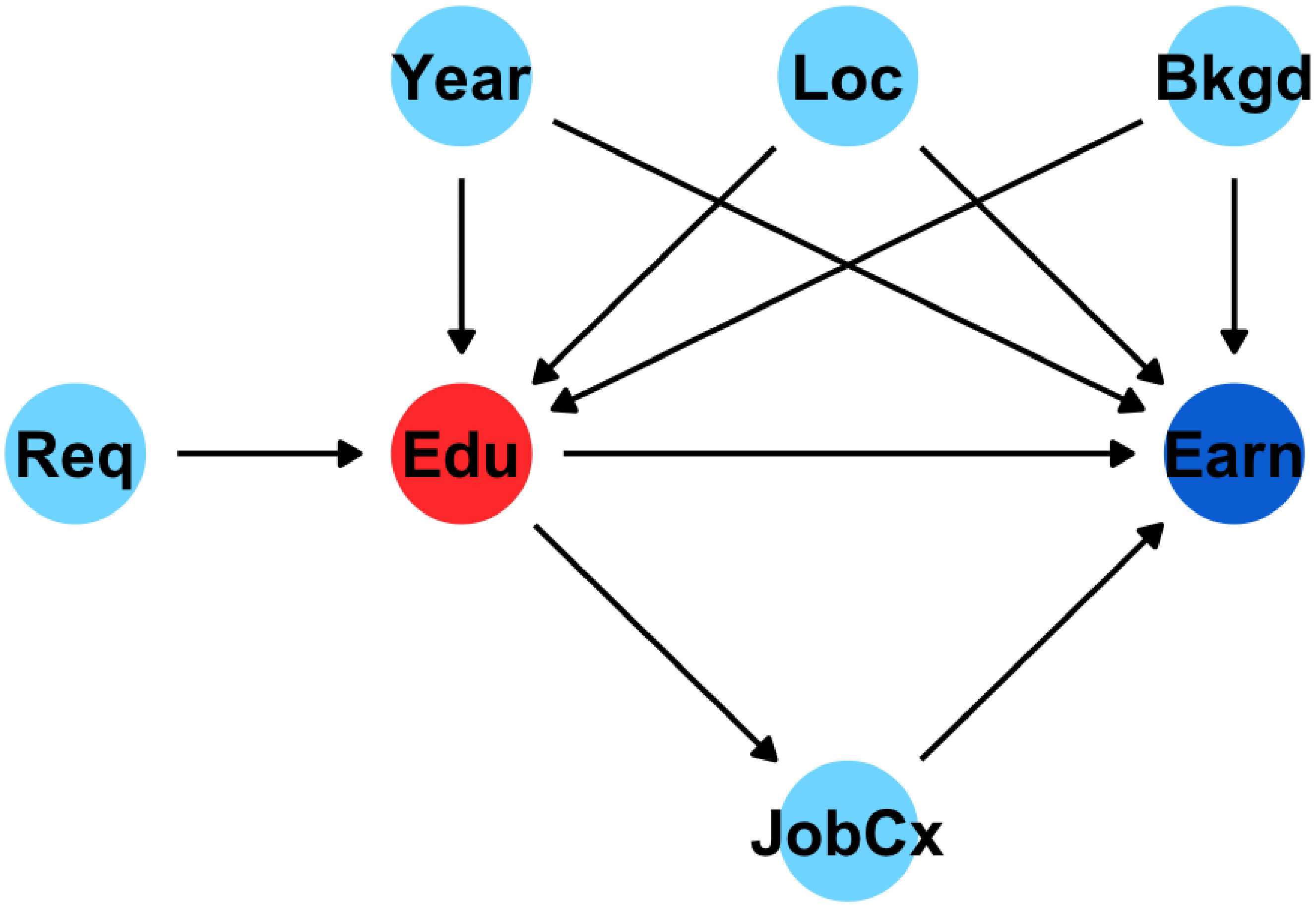
3. Draw arrows

Background, year of birth, and location all cause earnings too



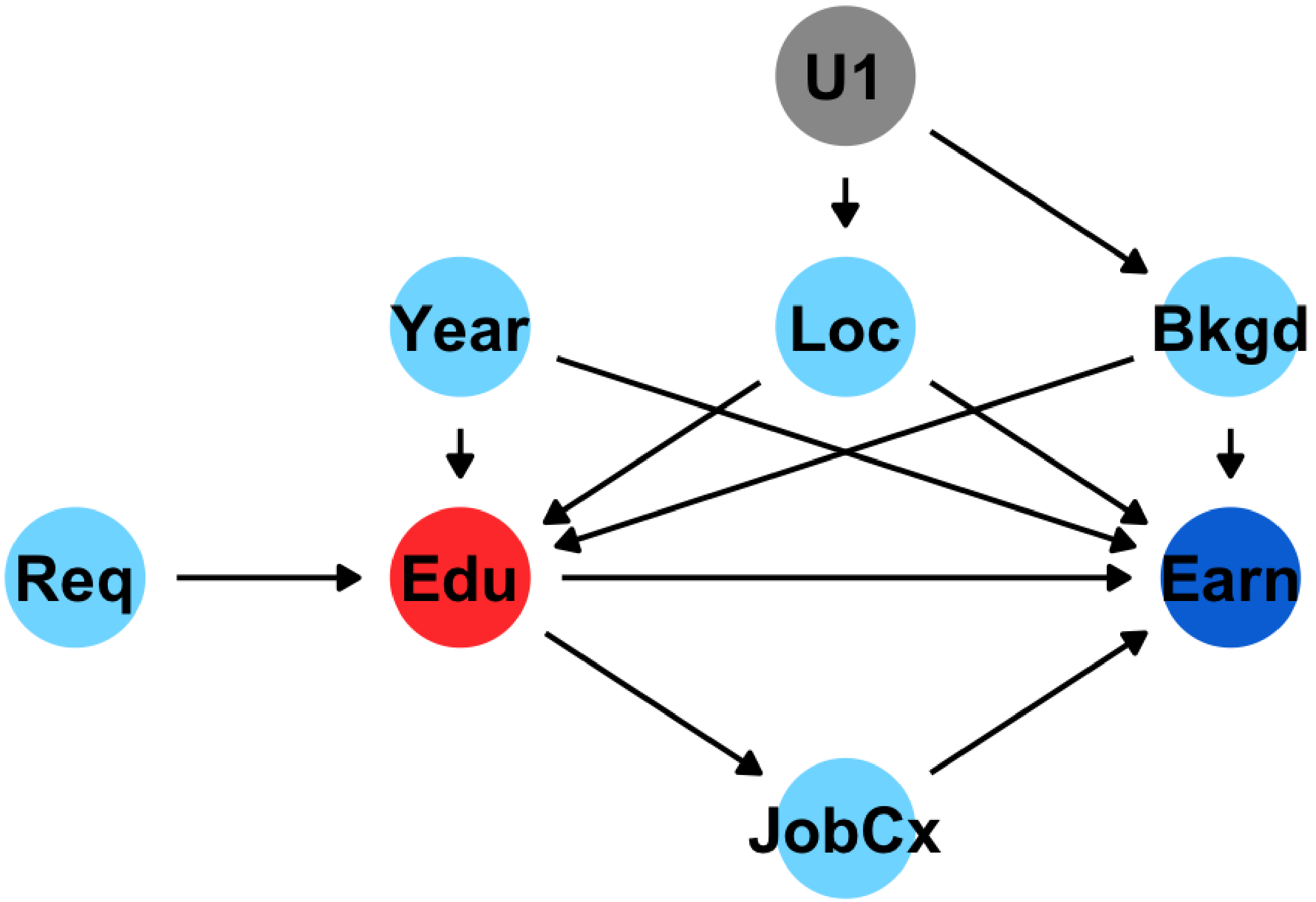
3. Draw arrows

Education causes job earnings



3. Draw arrows

Location and background are probably related, but neither causes the other. Something unobservable (U1) does that.



Your turn #1

Does a longer night's sleep extend your lifespan?

Step 1: List variables

Step 2: Simplify

Step 3: Connect arrows

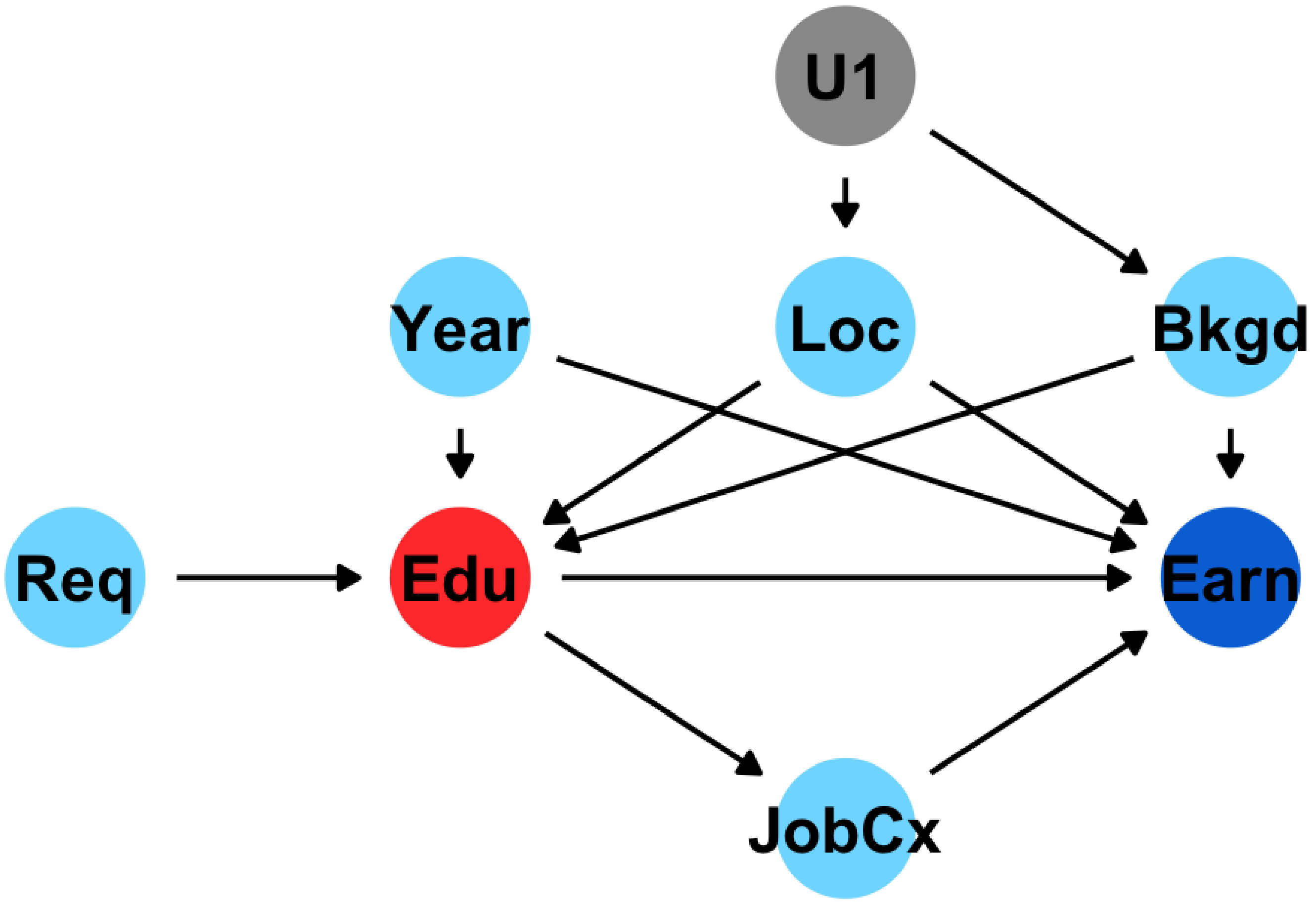
Use dagitty.net

05:00

Causal identification

All these nodes are related; there's correlation between them all

We care about Edu → Earn, but what do we do about all the other nodes?



Causal identification

A causal effect is *identified* if the association between treatment and outcome is propertly stripped and isolated

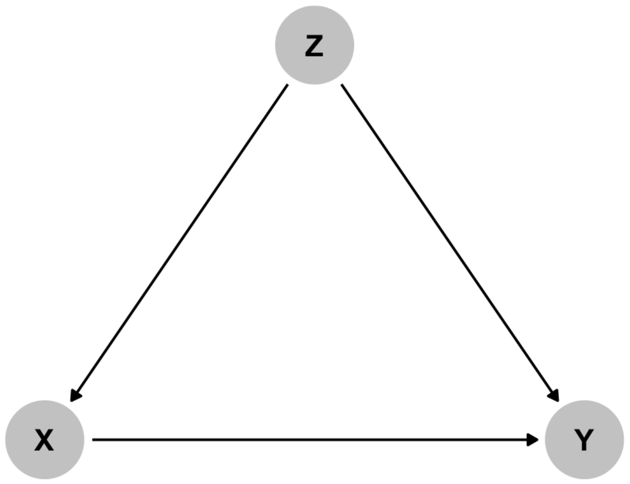
Paths and associations

Arrows in a DAG transmit associations

You can redirect and control those paths by "adjusting" or "conditioning"

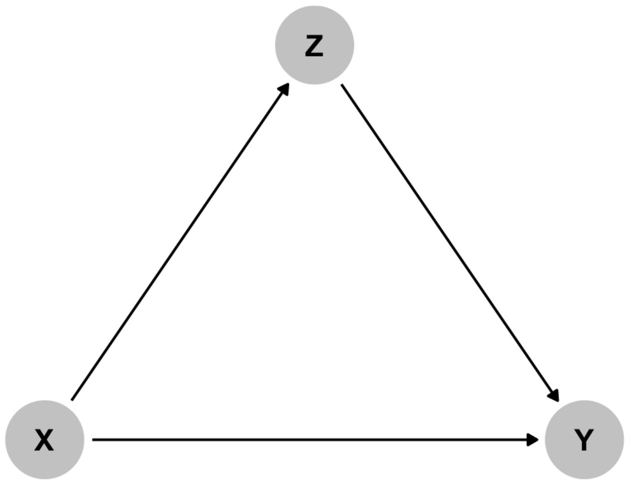
Three types of associations

Confounding



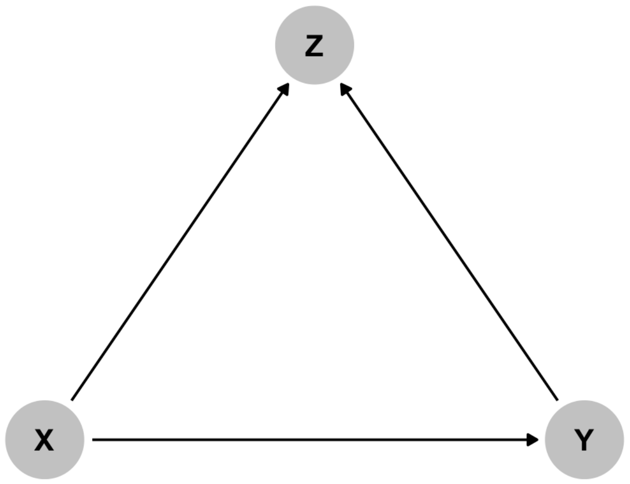
Common cause

Causation



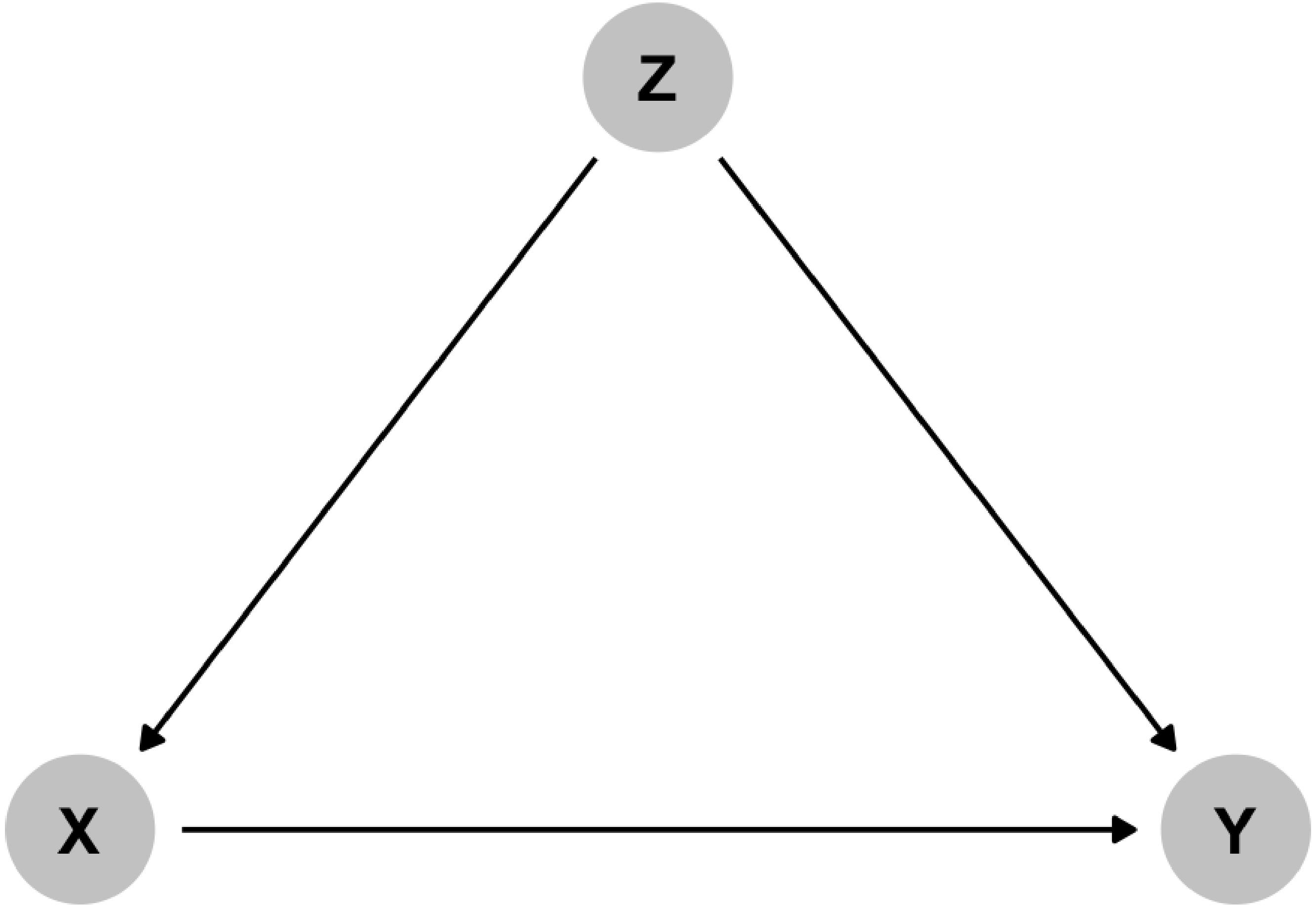
Mediation

Collision



Selection / endogeneity

Confounding

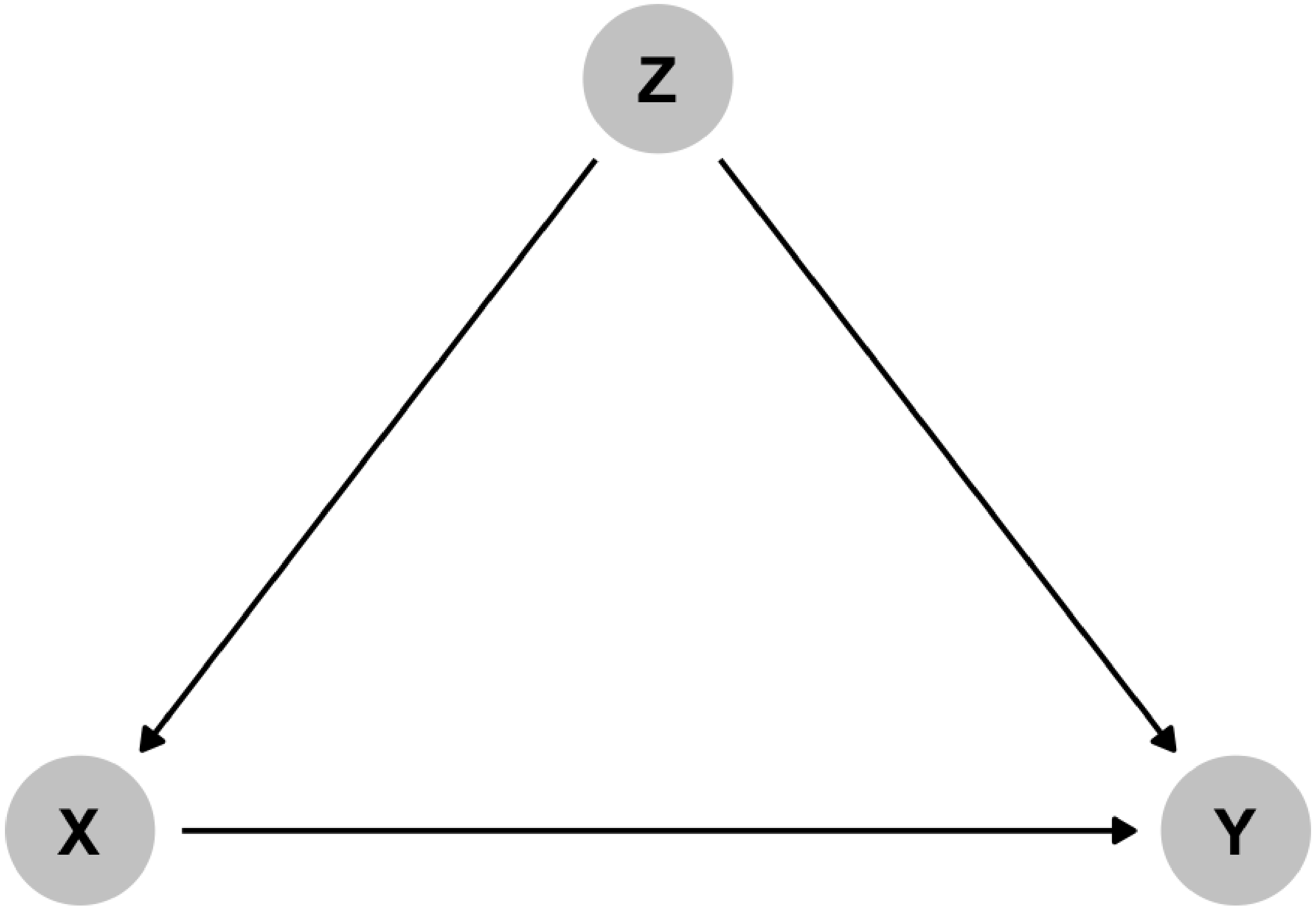


X causes Y

But Z causes both X and Y

Z *confounds* the X $\rightarrow$ Y association

Paths



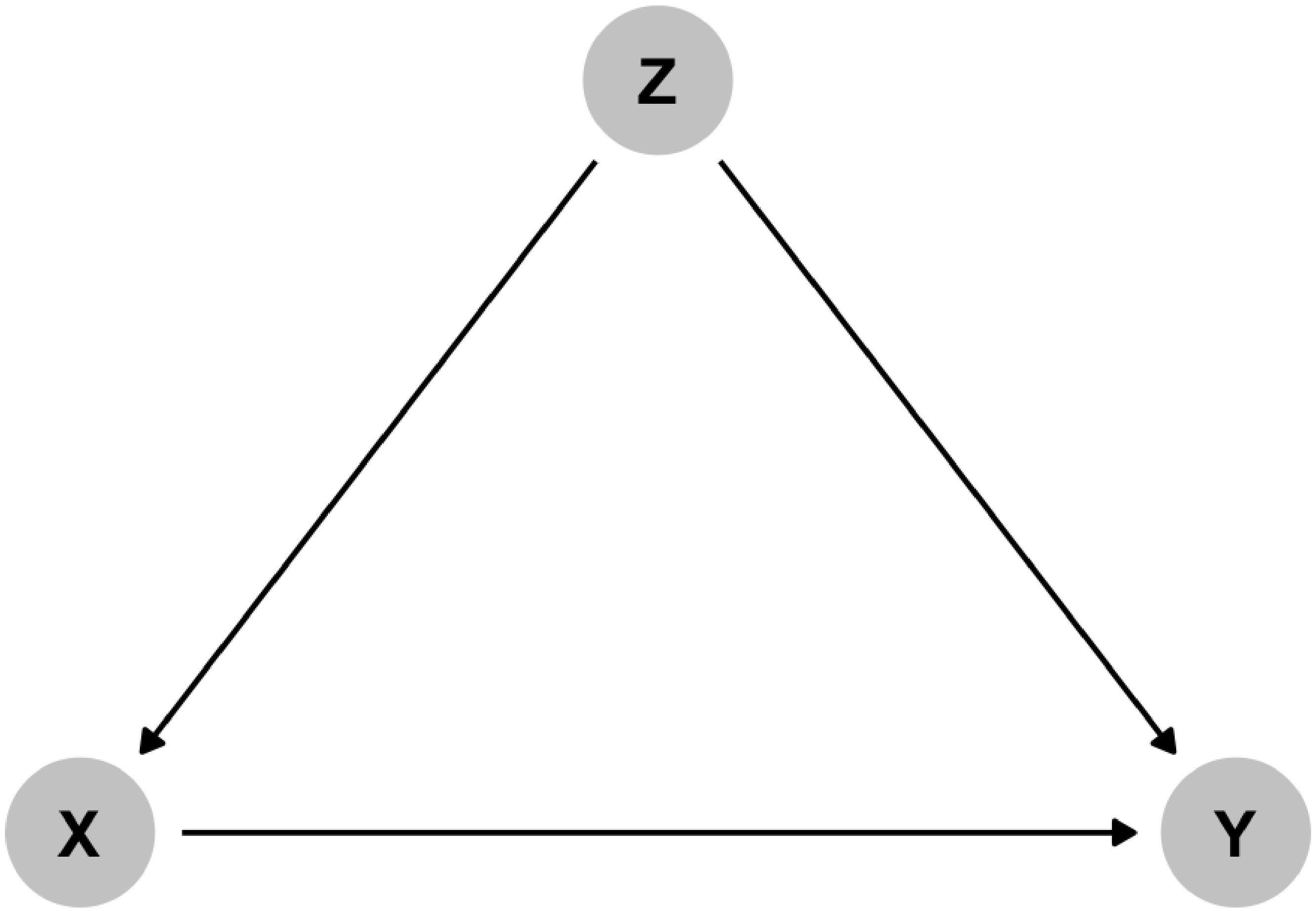
Paths between X and Y?

X → Y

X ← Z → Y

Z is a *backdoor*

d-connection

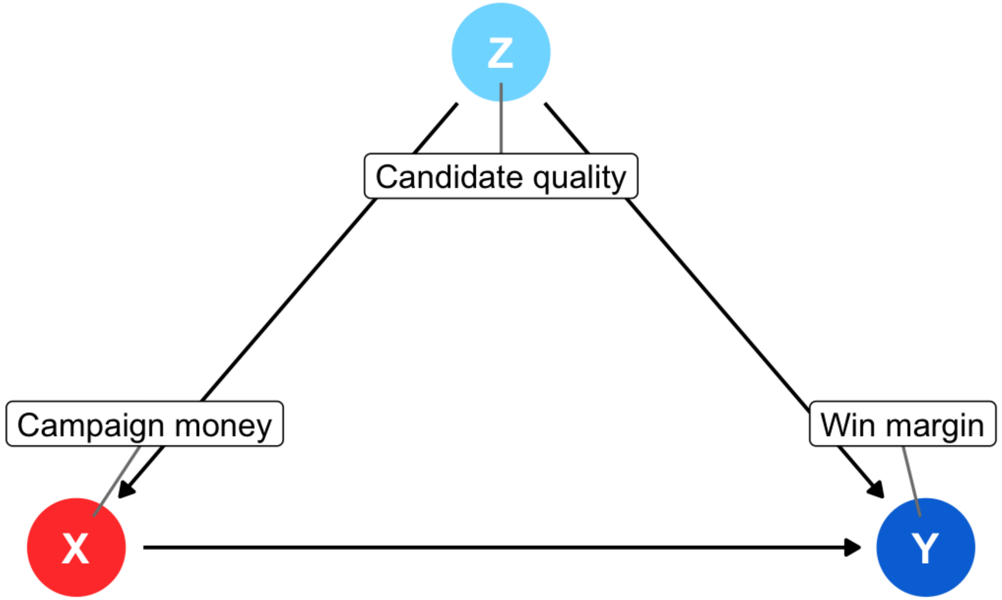


X and Y are "*d*-connected" because associations can pass through Z

The relationship between X and Y is not identified / isolated

Effect of money on elections

What are the paths between money and win margin?

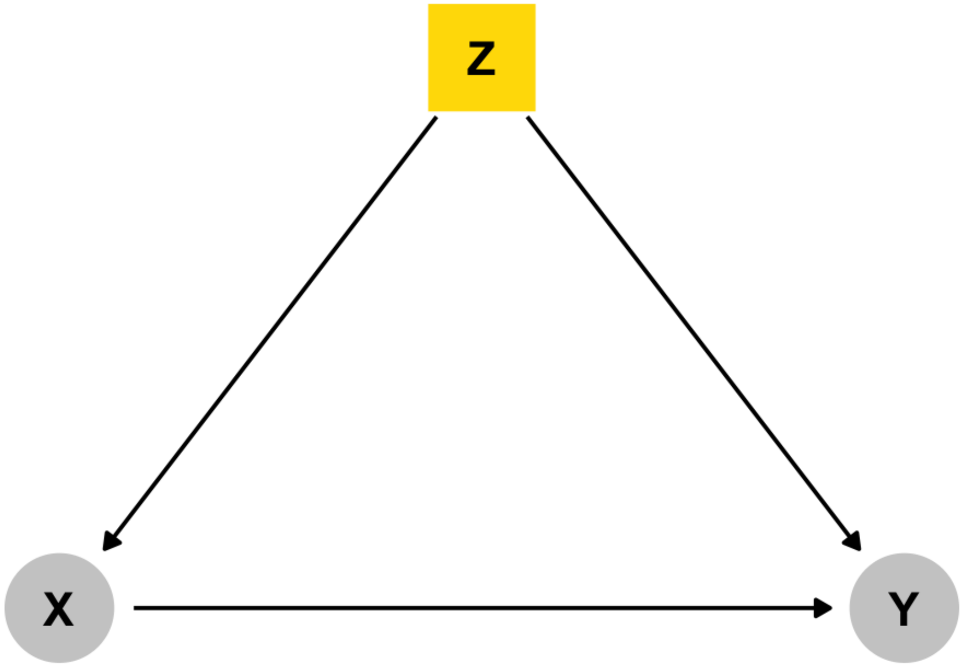


Money → Margin

Money ← Quality → Margin

Quality is a *backdoor*

Closing doors



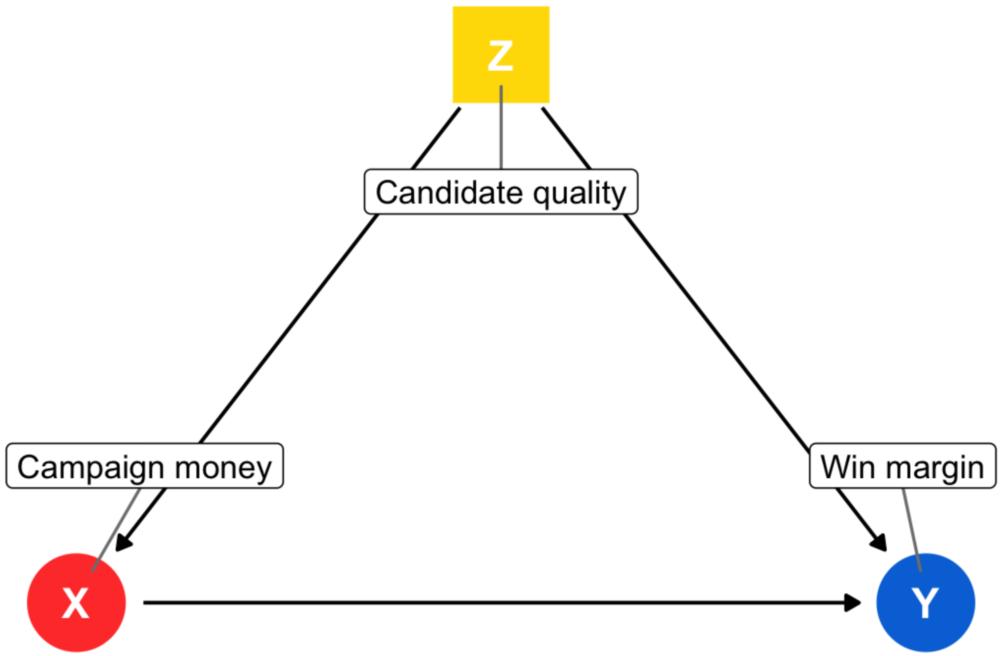
Close the backdoor by adjusting for Z

Closing doors

Find the part of campaign money that is explained by quality, subtract it out. This is the residual part of money.

Find the part of win margin that is explained by quality, subtract it out. This is the residual part of win margin.

Find the relationship between the residual part of money and residual part of win margin.
This is the causal effect.

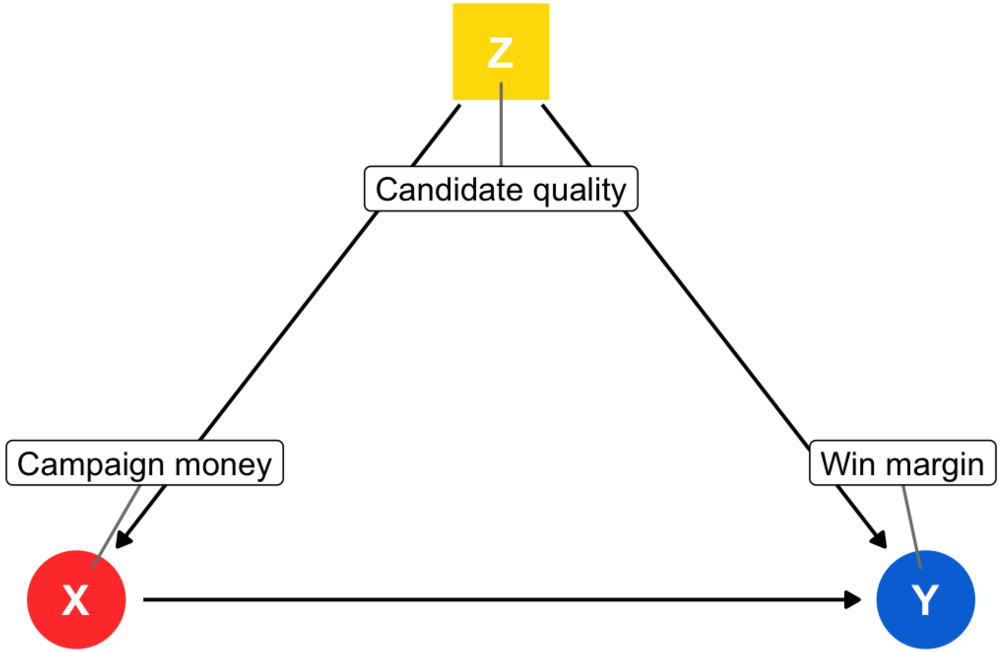


Closing doors

Compare candidates as if they had the same quality

Remove differences that are predicted by quality

Hold quality constant



How to adjust

Include term in regression

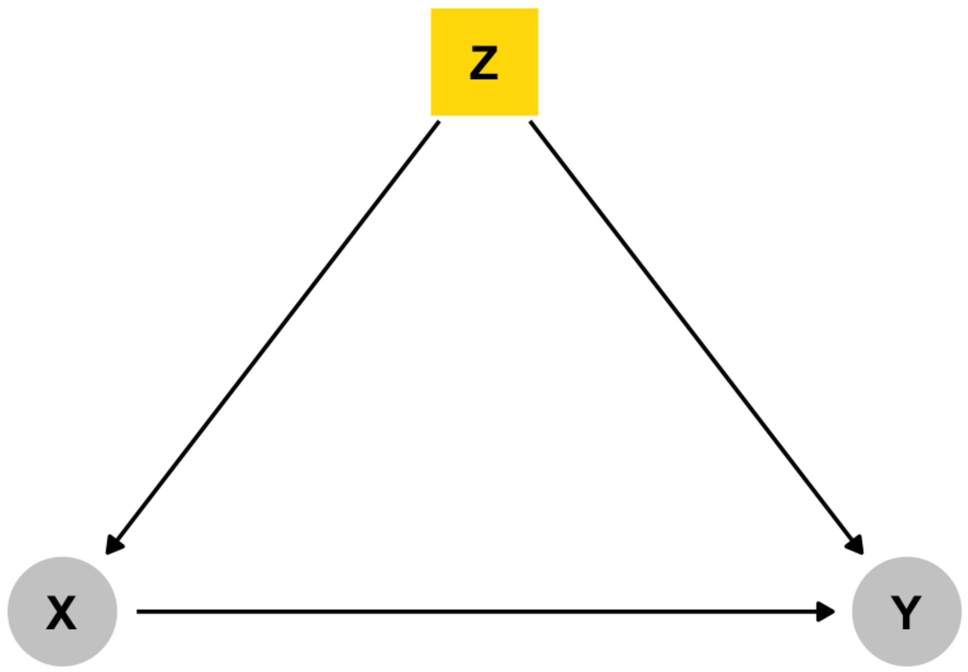
$$\begin{aligned}\text{Win margin} = &\beta_0 + \beta_1 \text{Campaign money} + \\ &\beta_2 \text{Candidate quality} + \varepsilon\end{aligned}$$

Matching

Stratifying

Inverse probability weighting

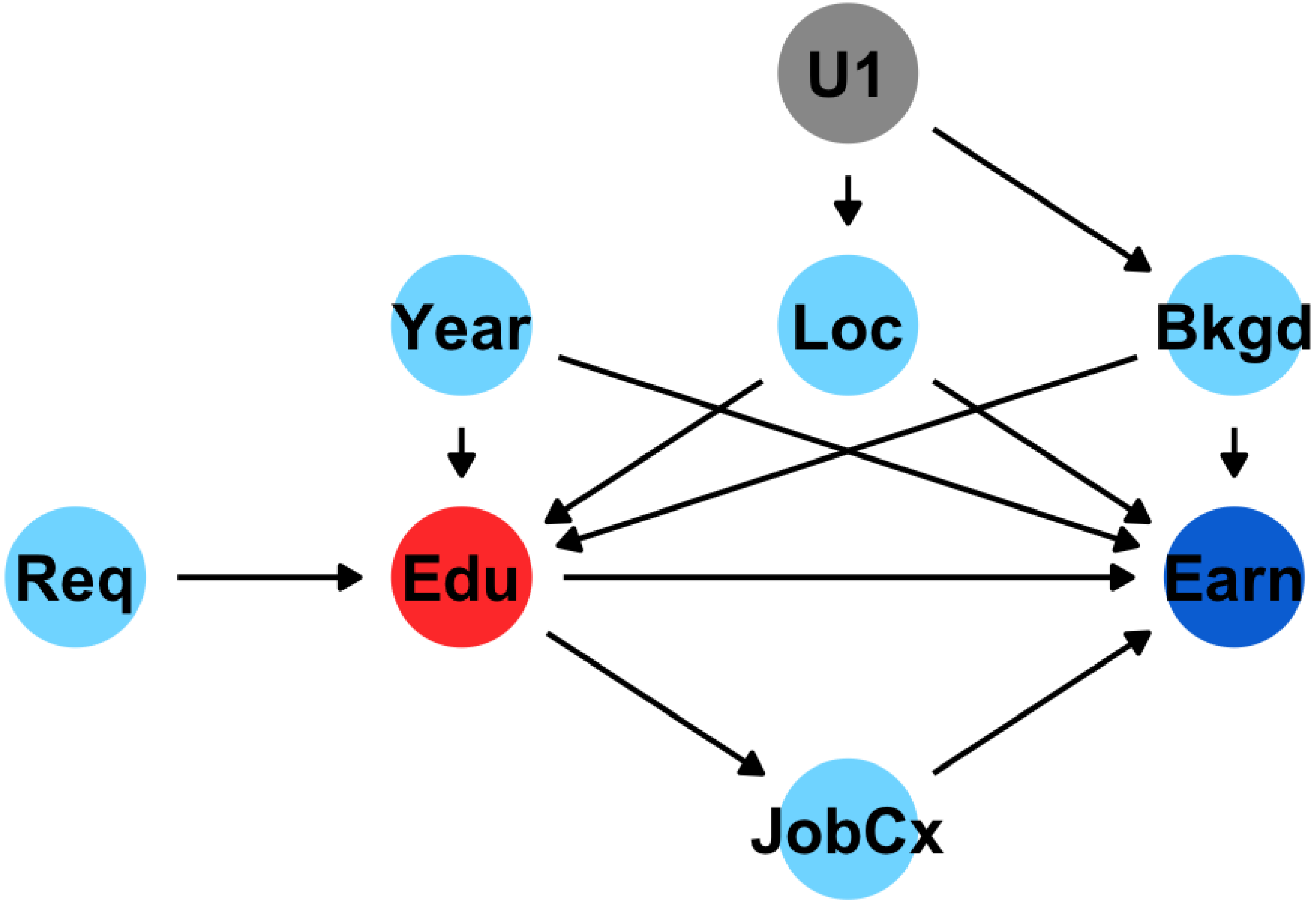
d-separation



If we control for Z, X and Y are now "*d*-separated" and the association is isolated!

Closing backdoors

Block all backdoor paths to identify the main pathway you care about



All paths

Education → Earnings

Education → Job connections → Earnings

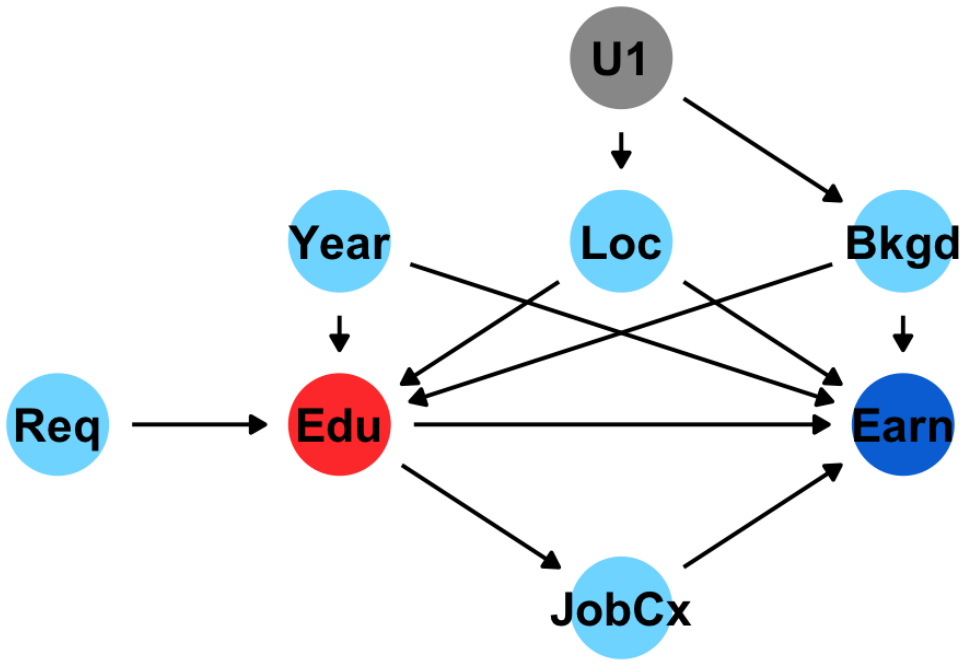
Education ← Background → Earnings

Education ← Background ← U1 → Location → Earnings

Education ← Location → Earnings

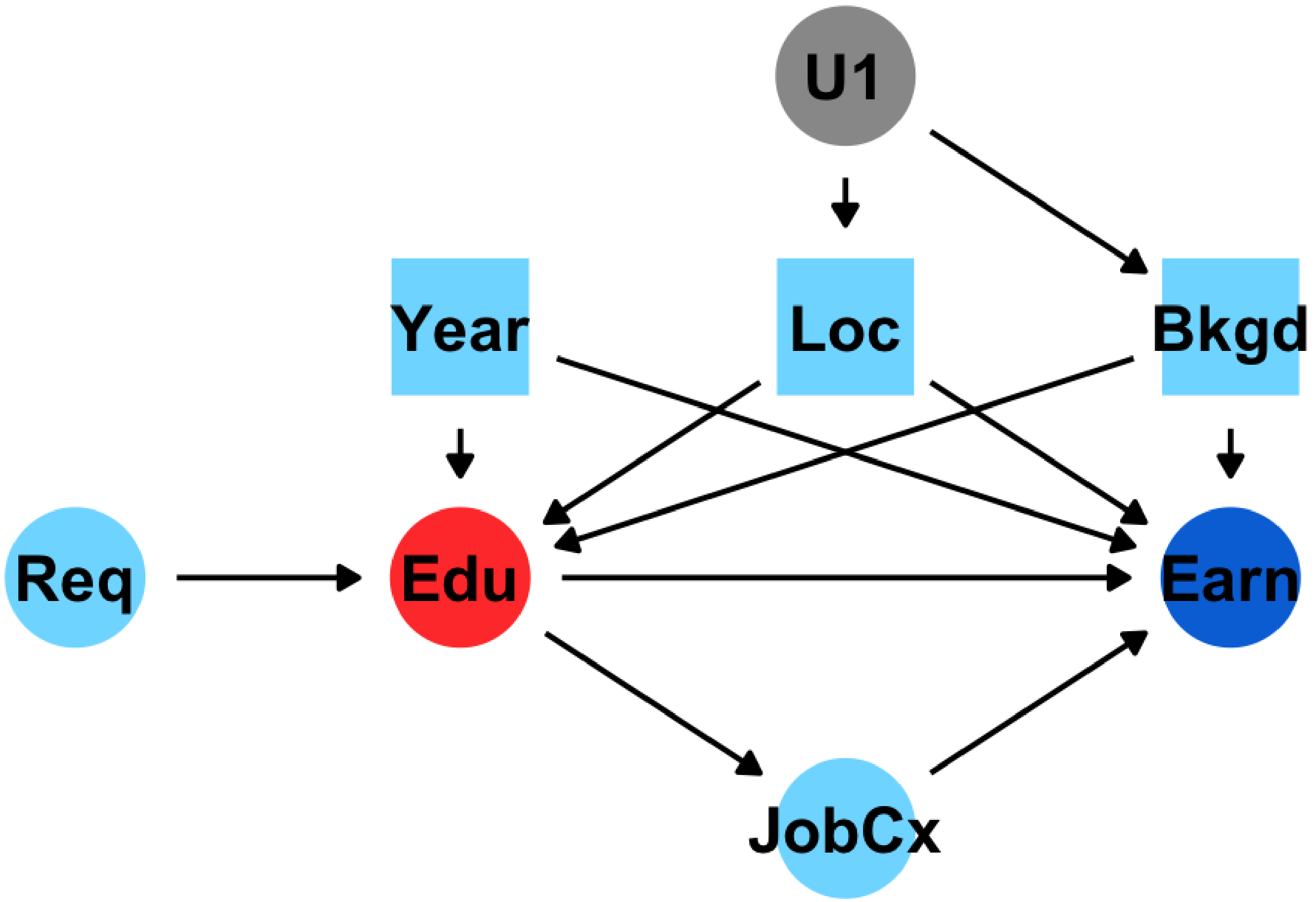
Education ← Location ← U1 → Background → Earnings

Education ← Year → Earnings



All paths

Adjust for **Location**, **Background** and **Year** to isolate the **Education** → **Earnings** causal effect



Let the computer do this!

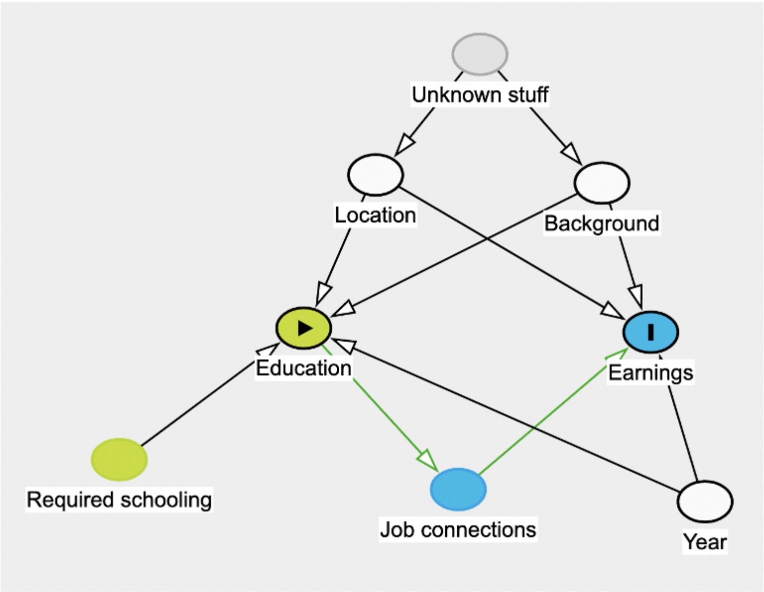
dagitty.net

How do you know if this is right?

You can test the implications of the model to see if they're right in your data

$$X \perp Y \mid Z$$

X is independent of Y, given Z



Testable implications

The model implies the following conditional independences:

- Education ⊥ Earnings | Background, Job connections, Location, Year
- Required schooling ⊥ Job connections | Education
- Required schooling ⊥ Year
- Required schooling ⊥ Earnings | Background, Job connections, Location, Year
- Required schooling ⊥ Earnings | Background, Education, Location, Year
- Required schooling ⊥ Background
- Required schooling ⊥ Location
- Job connections ⊥ Year | Education
- Job connections ⊥ Background | Education
- Job connections ⊥ Location | Education
- Year ⊥ Background
- Year ⊥ Location

Your turn #2

Go to andhs.co/nyt and skim the article

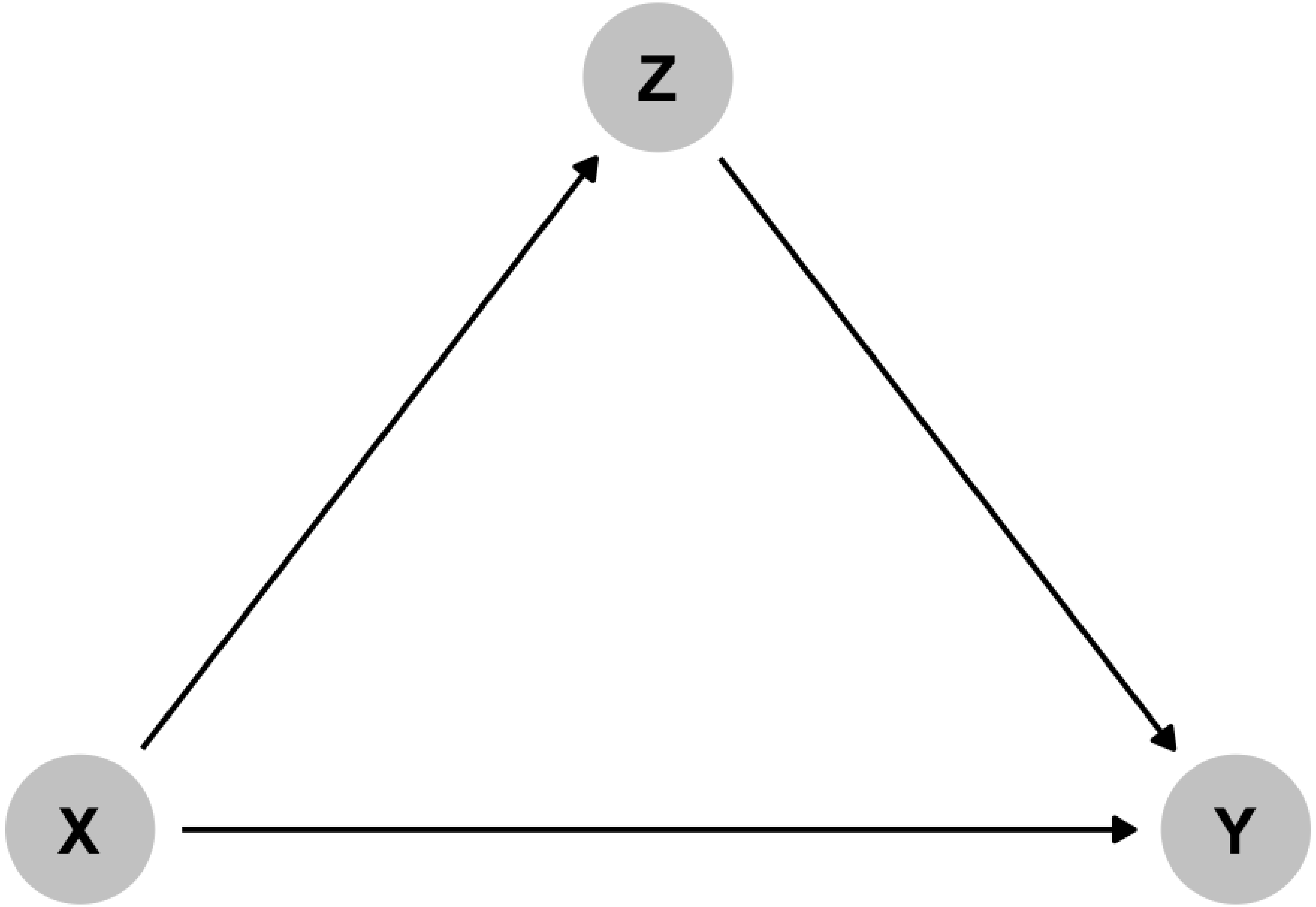
Pick one of the causal claims in the article

Draw a DAG for that causal claim

Determine what needs to be adjusted to identify the effect

06:00

Causation

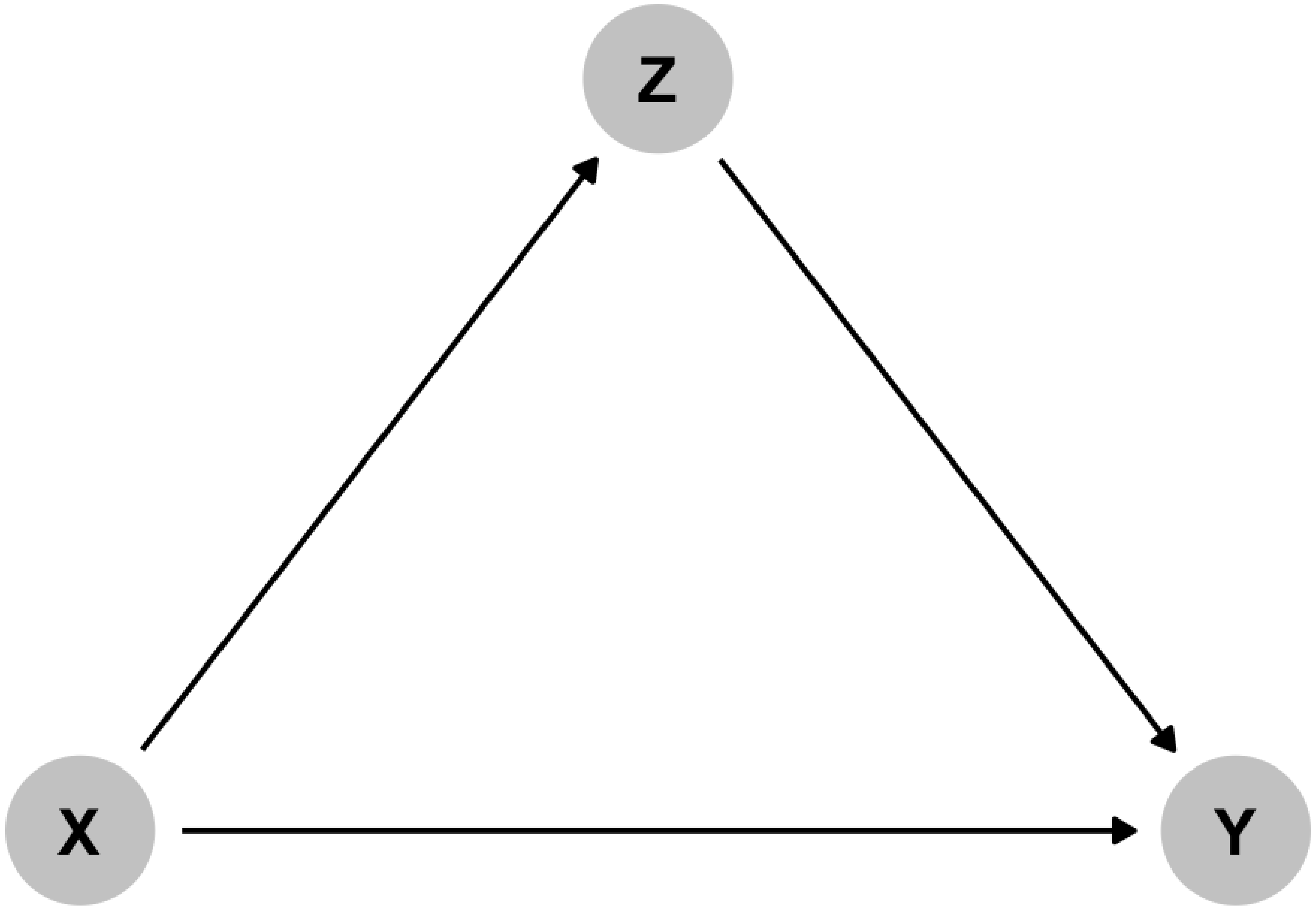


X causes Y

X causes Z which causes Y

Should you control for Z?

Causation

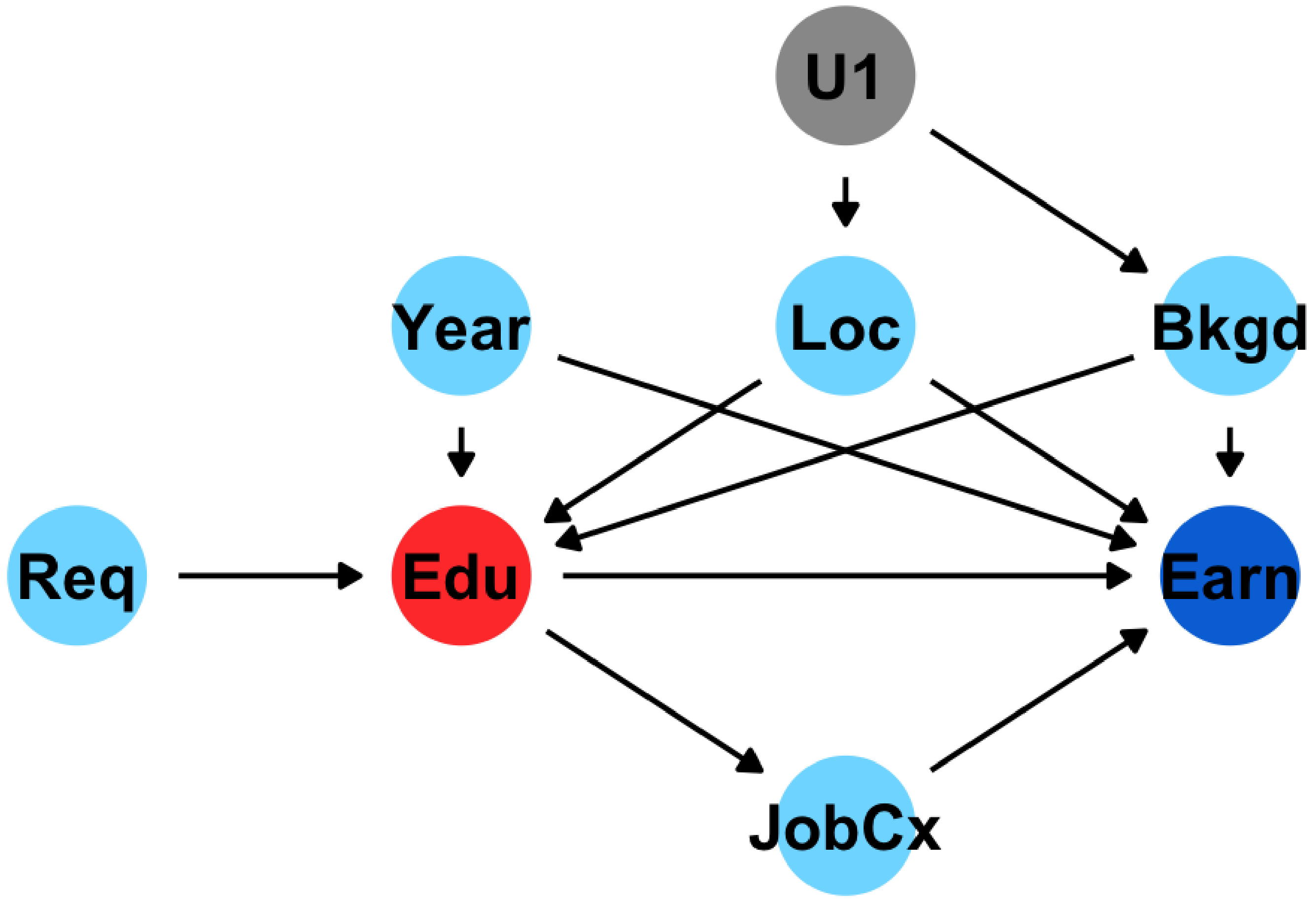


Should you control for Z?

No!

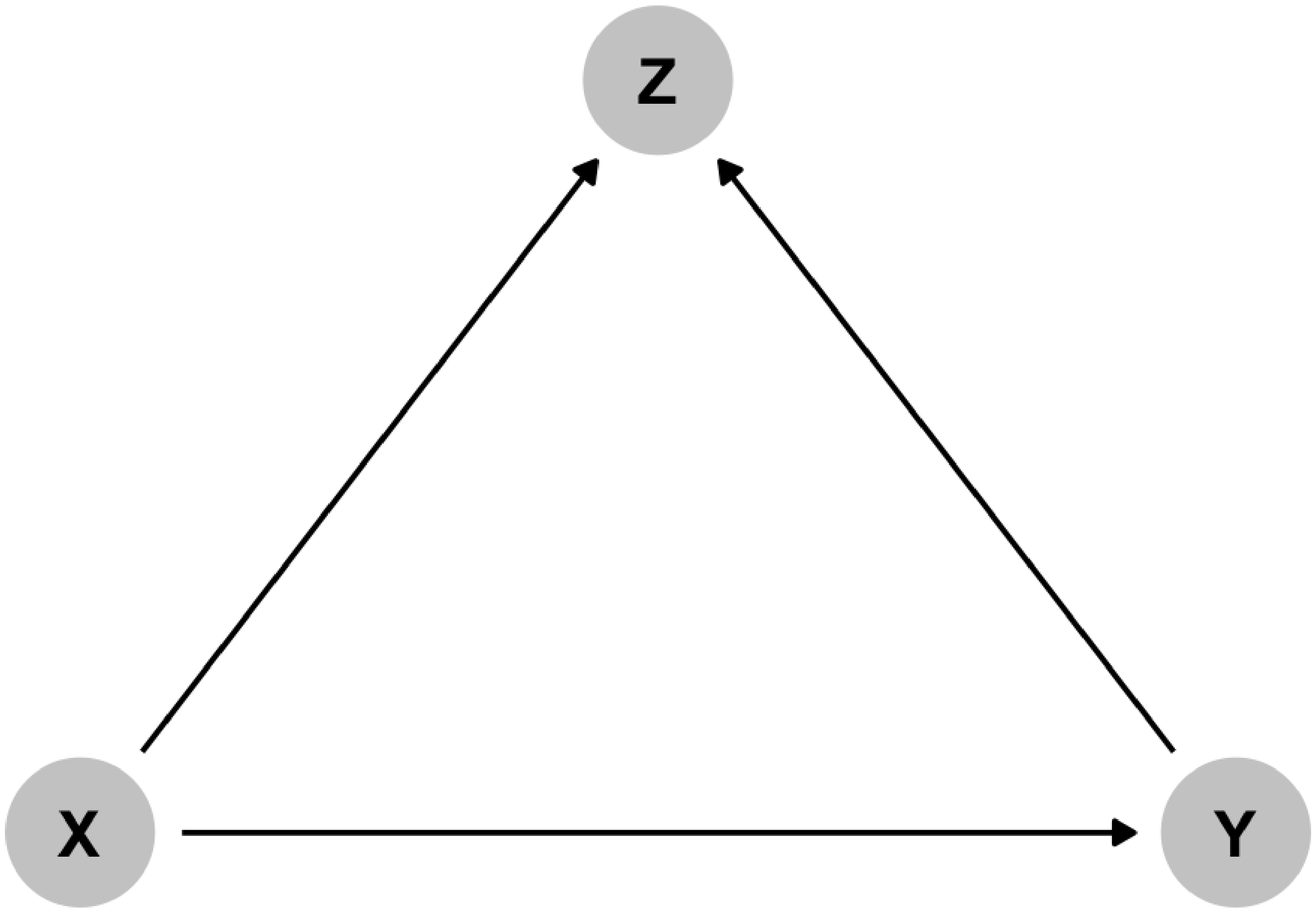
Overcontrolling

Causation and overcontrolling



Should you control for job connections?

Colliders



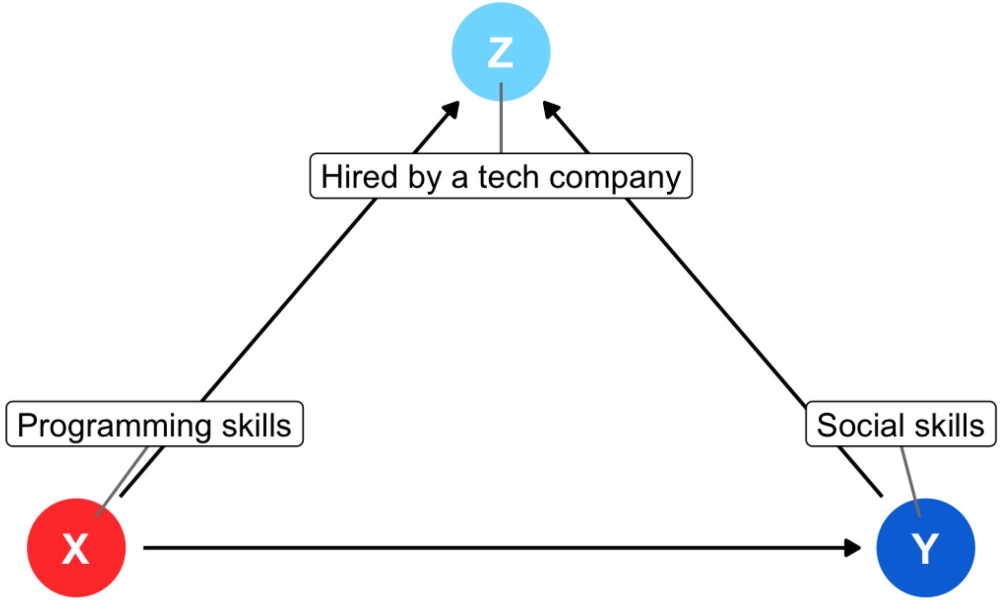
X causes Z

Y causes Z

Should you control for Z?

Programming and social skills

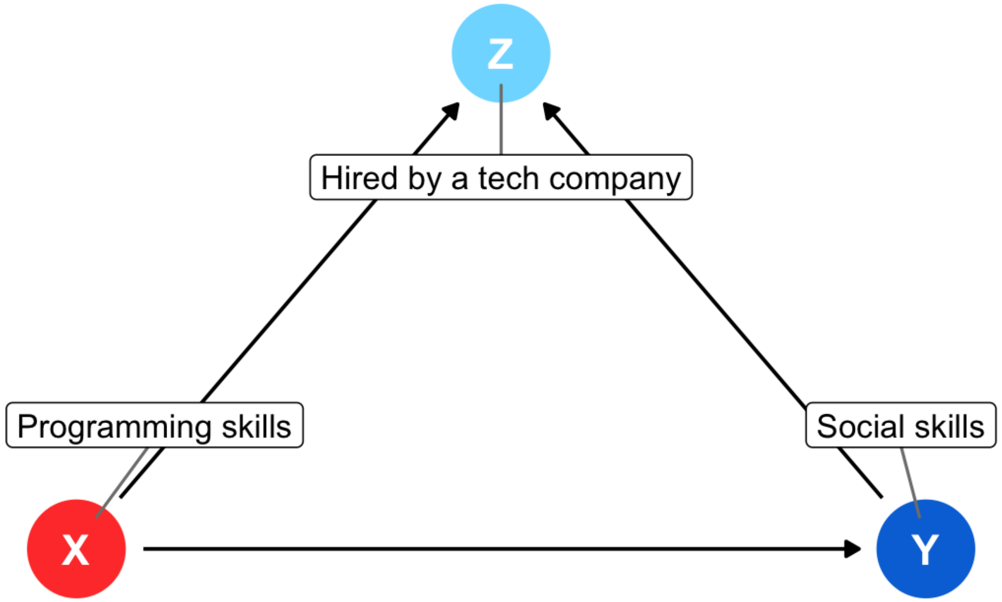
Do programming skills reduce social skills?



**You go to a tech company and conduct a survey. You find a negative relationship!
Is it real?**

Programming and social skills

Do programming skills reduce social skills?

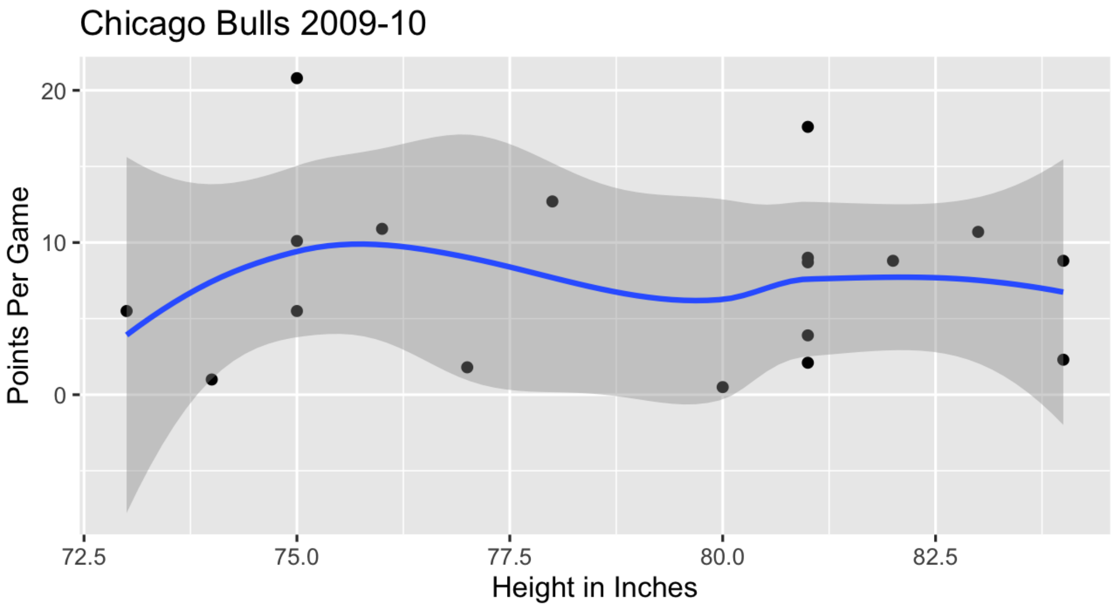


No! Hired by a tech company is a collider and we controlled for it.

This inadvertently connected the two.

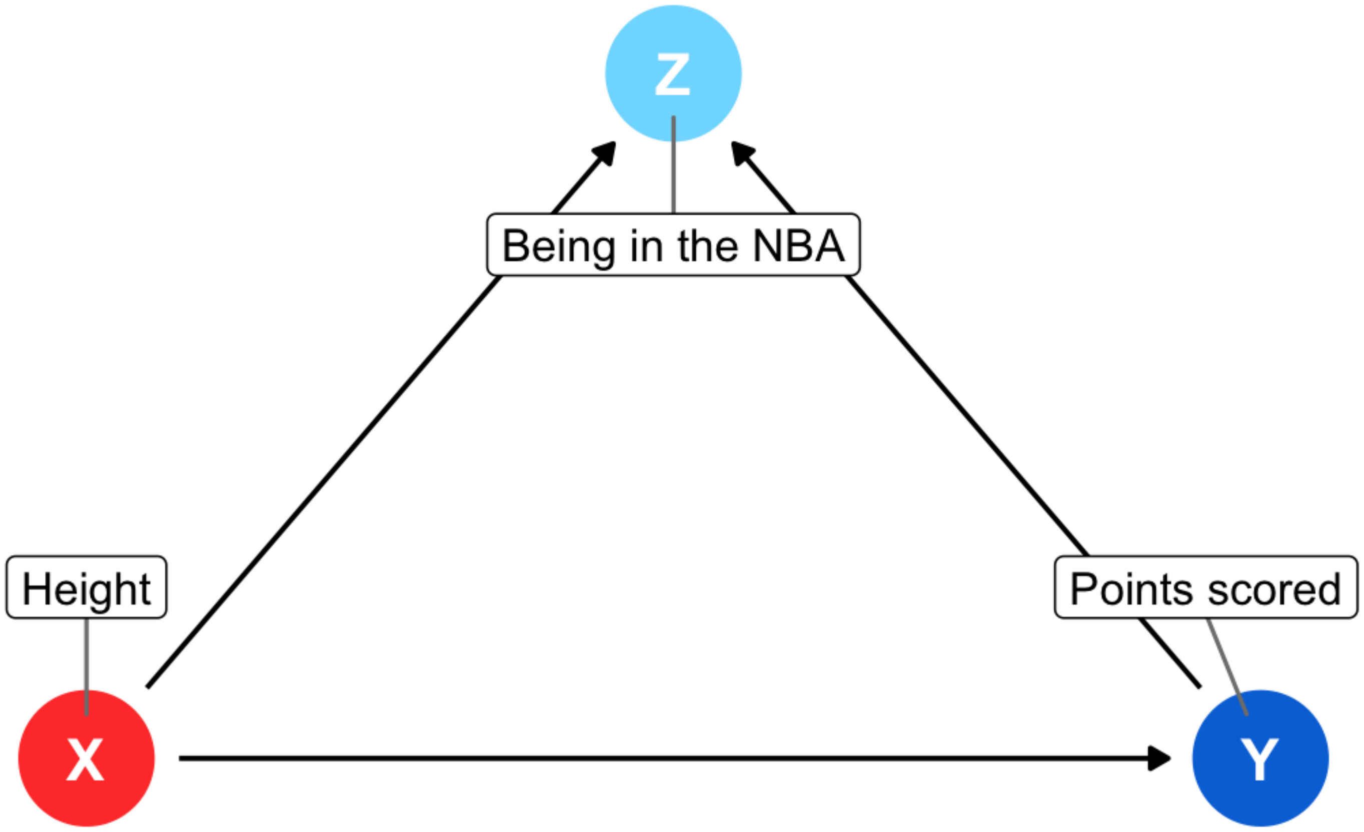
Colliders can create fake causal effects

Colliders can hide real causal effects



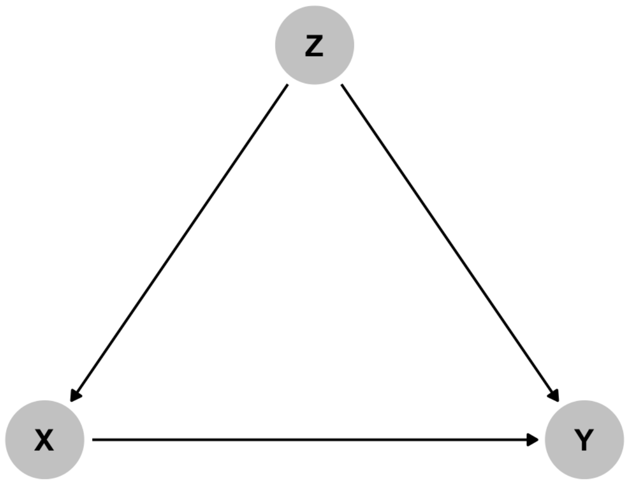
Height is unrelated to basketball skill… among NBA players

Colliders and selection bias



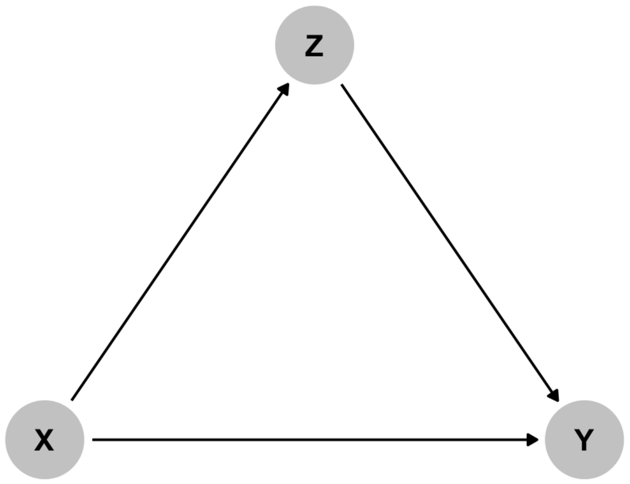
Three types of associations

Confounding



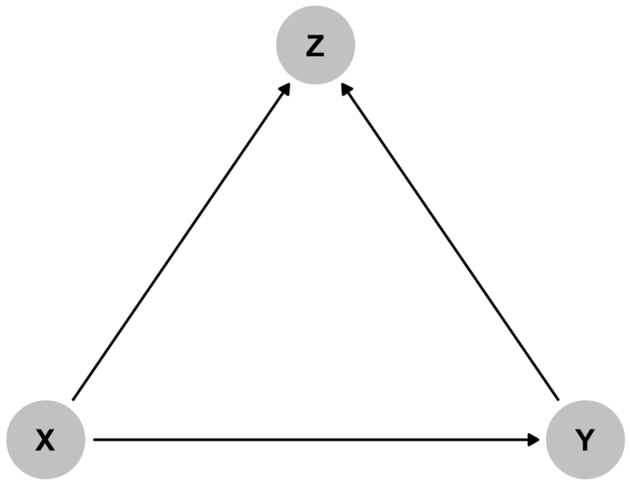
Common cause

Causation



Mediation

Collision



Selection / endogeneity

Next up

How to analyze RCTs