

# Measurement and DAGs

## Session 4

PMAP 8521: Program evaluation  
Andrew Young School of Policy Studies

# Plan for today

Abstraction, stretching,  
and validity

Causal models

Paths, doors, and adjustment

# Abstraction, stretching, and validity

# Indicators

## Inputs, activities, and outputs

Generally directly measurable

# of citations mailed,  
% increase in grades, etc.

## Outcomes

Harder to measure directly

Loftier and more abstract

Commitment to school,  
reduced risk factors

# How do you measure abstract outcomes?

**Move up the ladder of abstraction.**

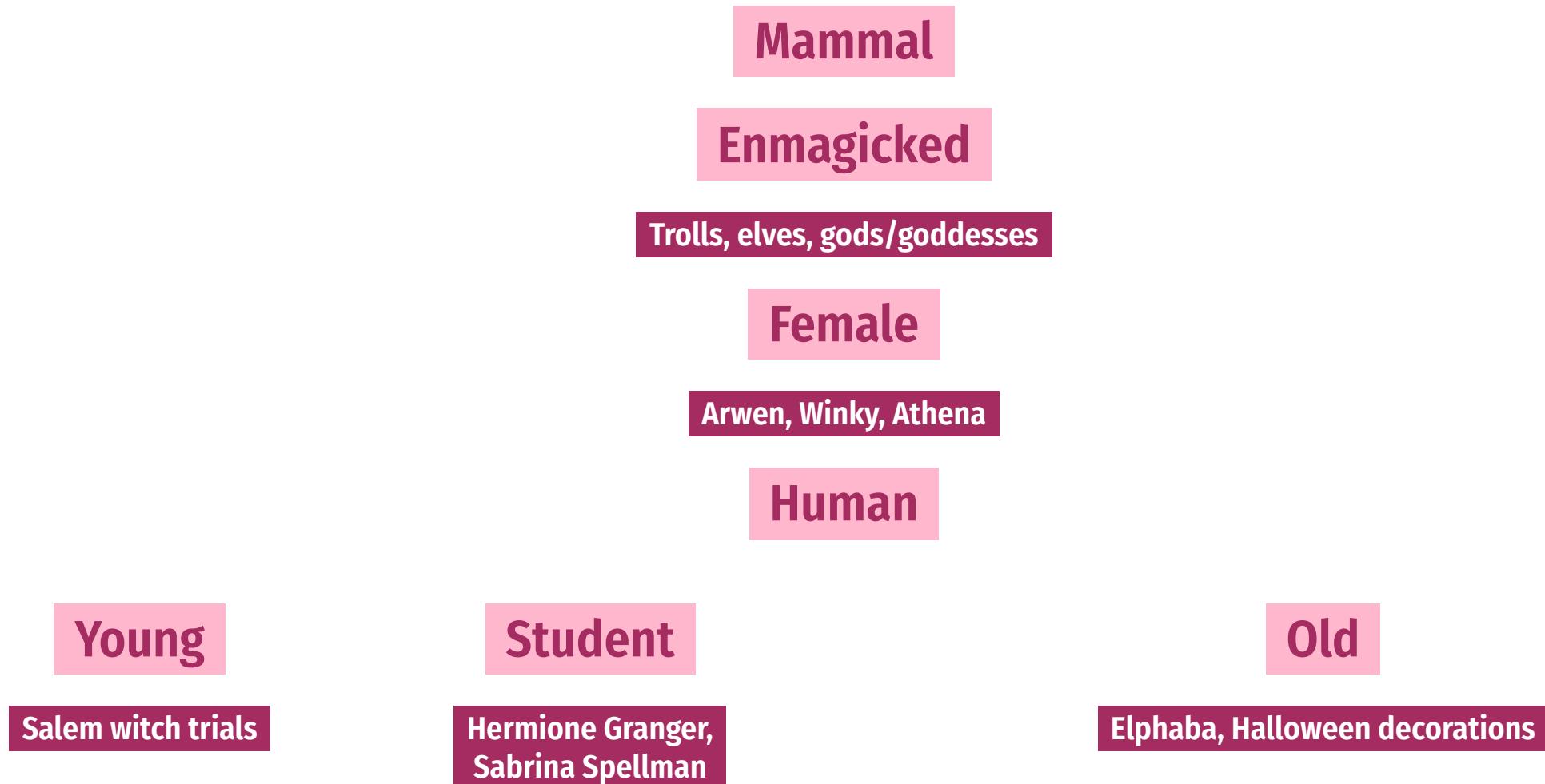




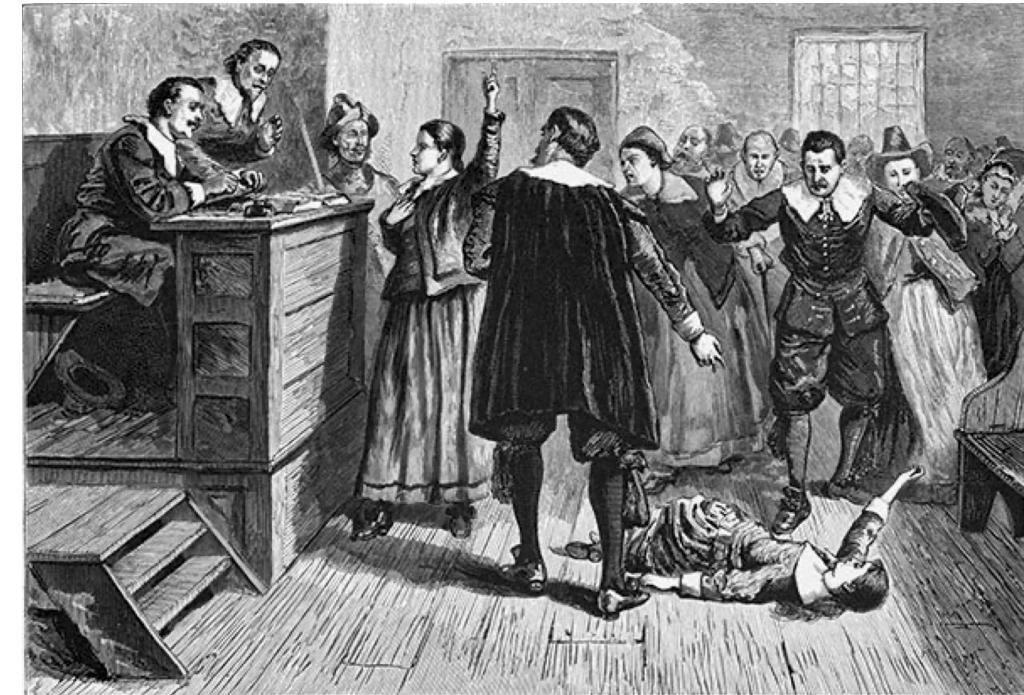
# Conceptual stretching



# Ladder of abstraction for witches



# Connection to theory



# Outcomes and programs

Outcome variable

Thing you're measuring

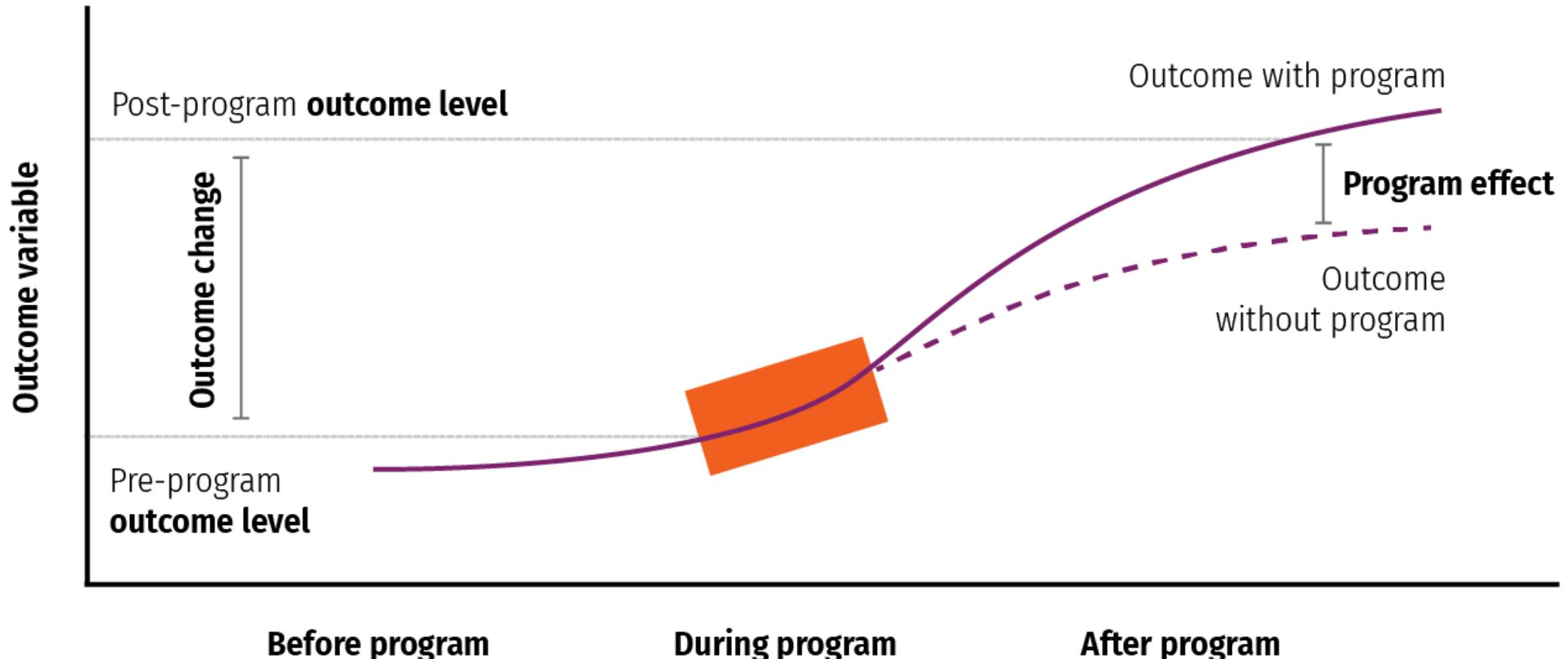
Outcome change

$\Delta$  in thing you're measuring over time

Program effect

$\Delta$  in thing you're measuring over time *because of* the program

# Outcomes and programs



# Connecting measurement to programs

**Measurable definition of program effect**

**Ideal measurement**

**Feasible measurement**

**Connection to real world**

# Causal models

# 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?**



Laura Hatfield  
@laura\_tastic

Wow: this comment from fresh page proofs.

Guess all of us researching causal inference in observational data need to find new jobs?

**Commented [DT1]:** Causal language (including use of terms such as effect, efficacy, and predictor) should be used only for randomized clinical trials. For all other study designs, methods and results should be described in terms of association or, if appropriate tests were used, correlation, and should avoid cause-and-effect wording. We have eliminated causal language from the manuscript.



Seva  
@SevaUT

normal person: this rain is making us wet

me, RCT genius: whoa there! First, take twenty walks and randomly apply the rain treatment



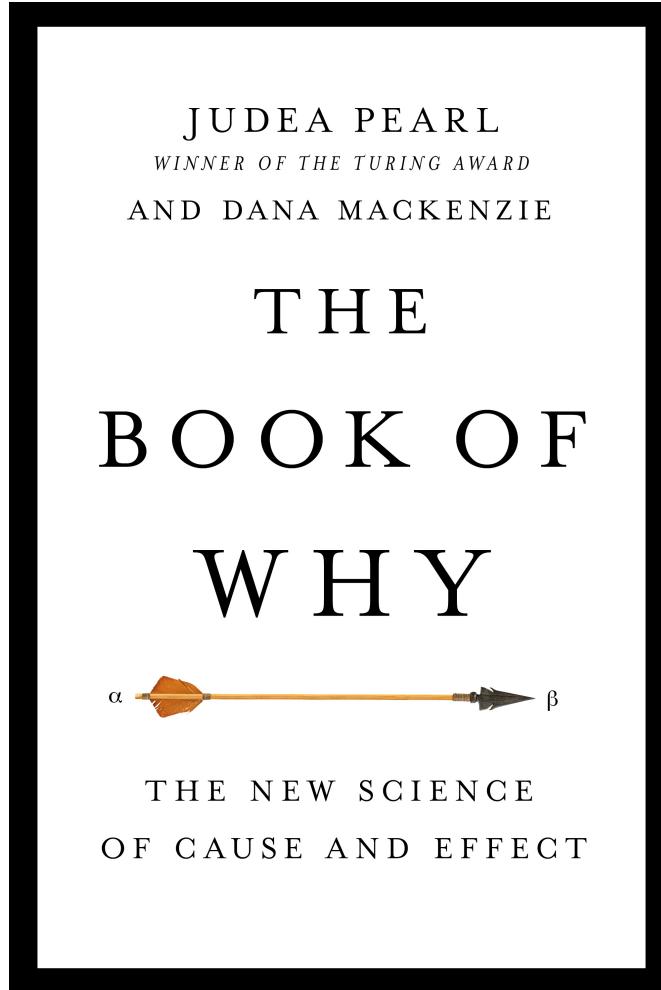
Laura Hatfield @laura\_tastic · Jan 16

Wow: this comment from fresh page proofs.

Guess all of us researching causal inference in observational data need to find new jobs?

**Commented [DT1]:** Causal language (including use of terms such as effect, efficacy, and predictor) should be used only for randomized clinical trials. For all other study designs, methods and results should be described in terms of association or, if appropriate tests were used, correlation, and should avoid cause-and-effect wording. We have

# The causal revolution



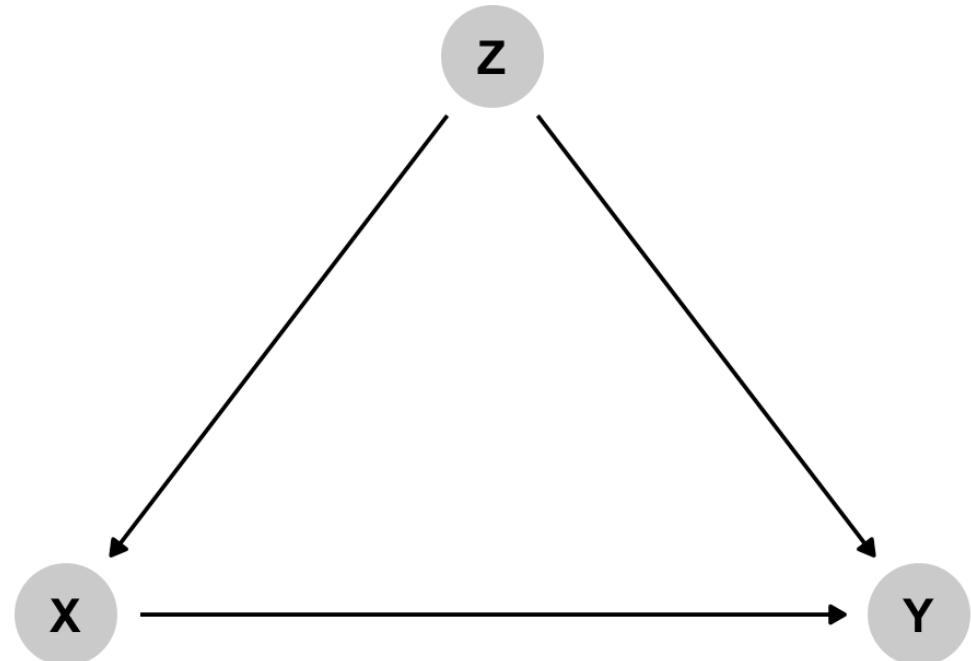
# Causal diagrams

## Directed acyclic graphs (DAGs)

**Directed:** Each node has an arrow that points to another node

**Acyclic:** You can't cycle back to a node (and arrows only have one direction)

**Graph:** It's... um... a graph



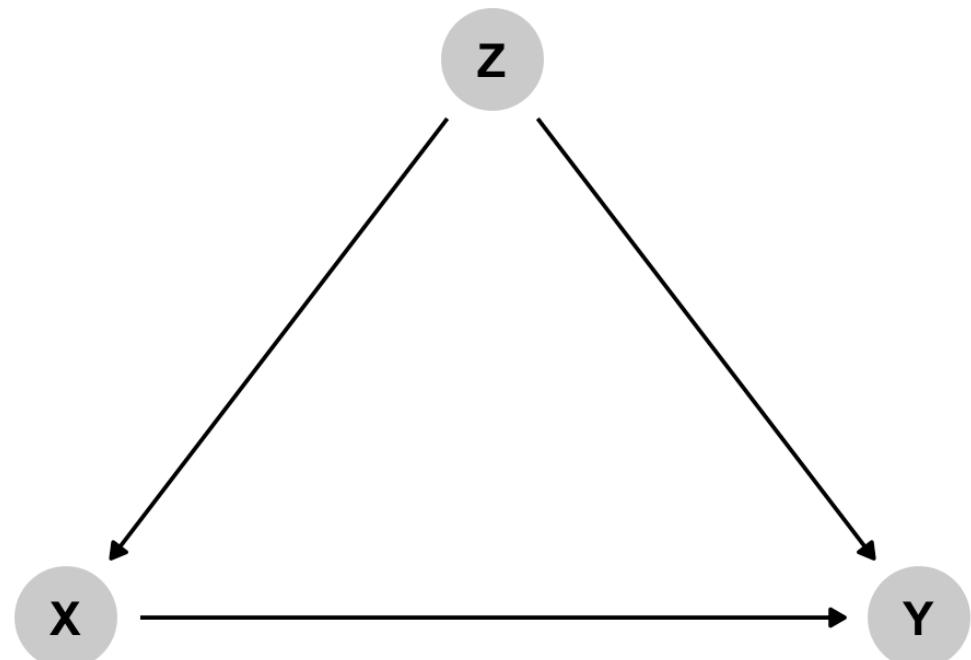
# Causal diagrams

## 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



# Acyclicalness

What if there's something that really is cyclical?

Wealth → Power → Wealth

This isn't acyclic!  
Wealth ↔ Power

Split the node into different time periods

Wealth<sub>t - 1</sub> → Power<sub>t</sub> → Wealth<sub>t</sub>

# 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) → Earnings (outcome)**

**Location**

**Ability**

**Demographics**

**Socioeconomic status**

**Year of birth**

**Compulsory schooling laws**

**Job connections**

## 2. Simplify

**Education (treatment) → Earnings (outcome)**

**Location**

**Ability**

**Demographics**

**Socioeconomic status**

**Year of birth**

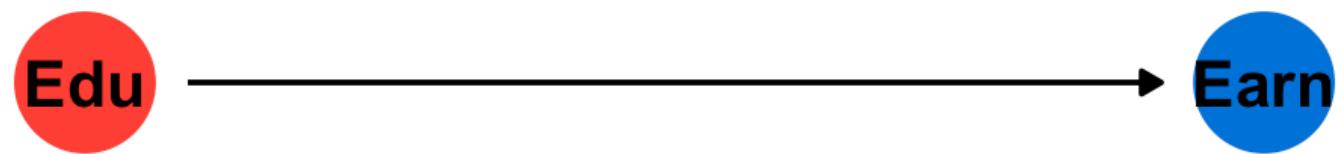
**Compulsory schooling laws**

**Job connections**

**Background**

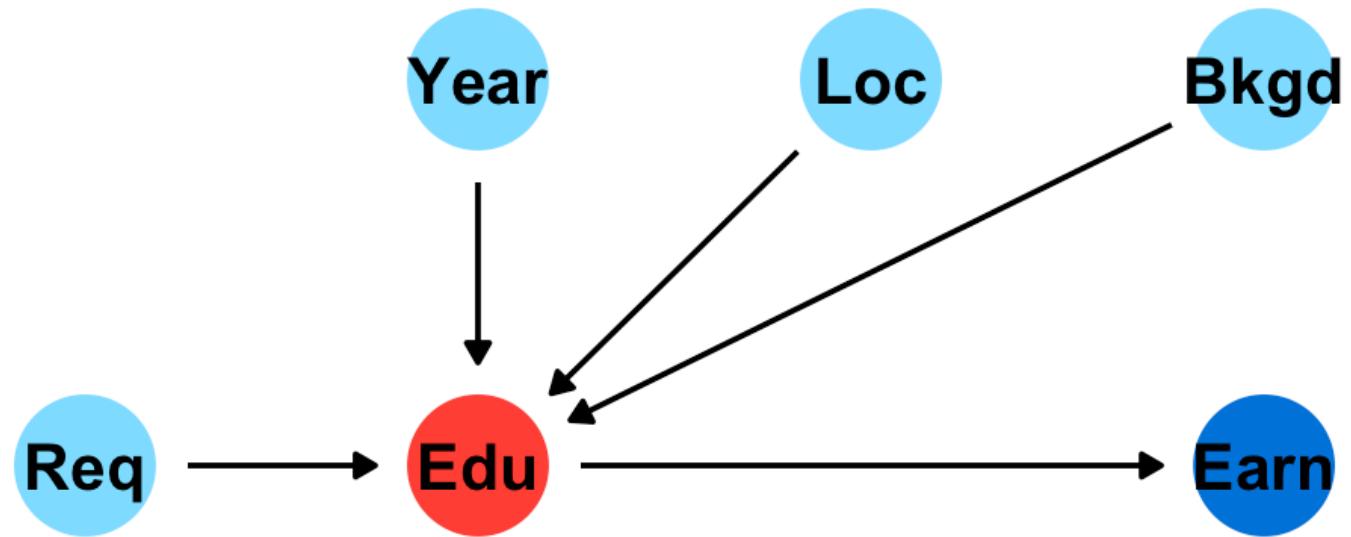
### 3. Draw arrows

Education causes  
earnings



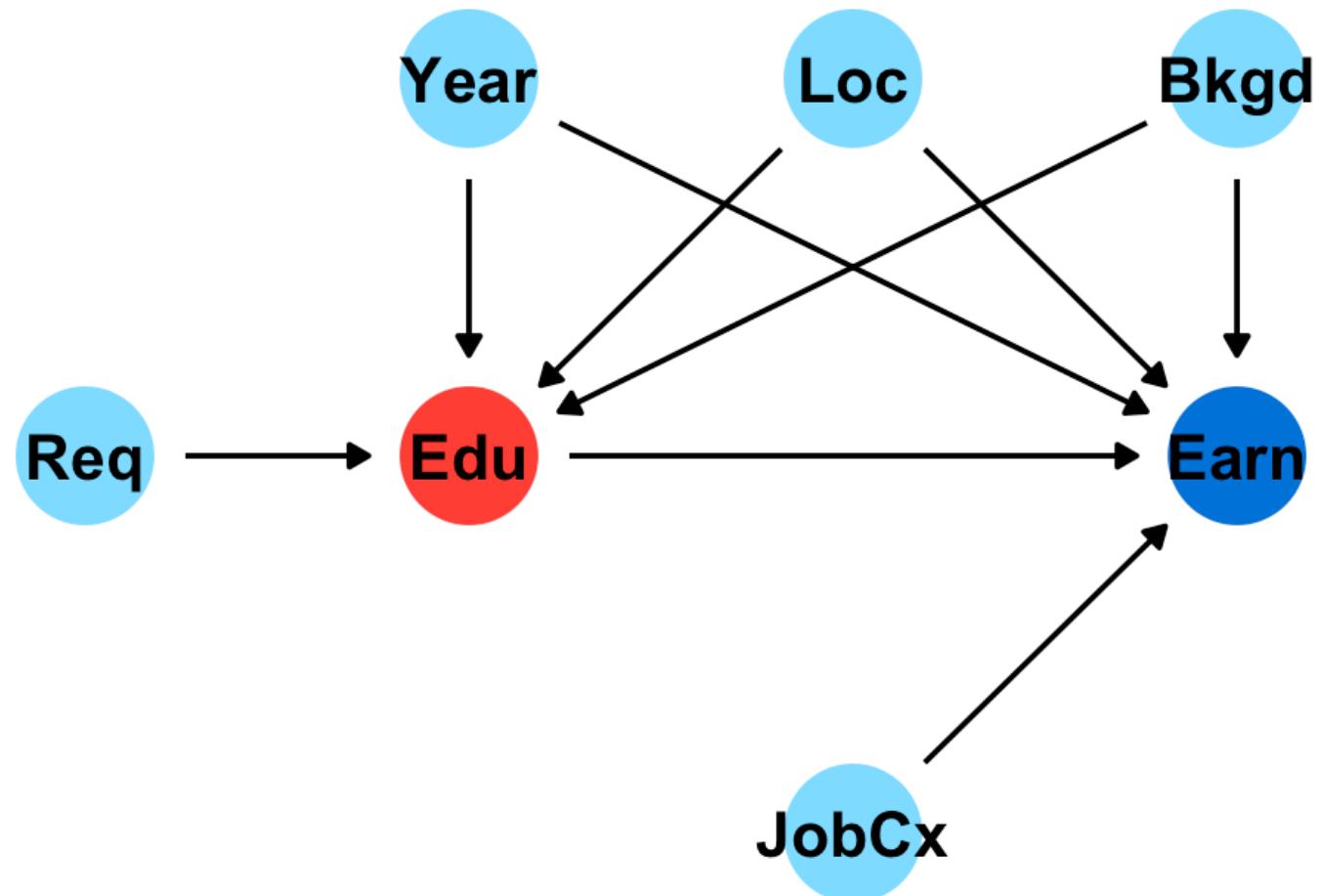
### 3. Draw arrows

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



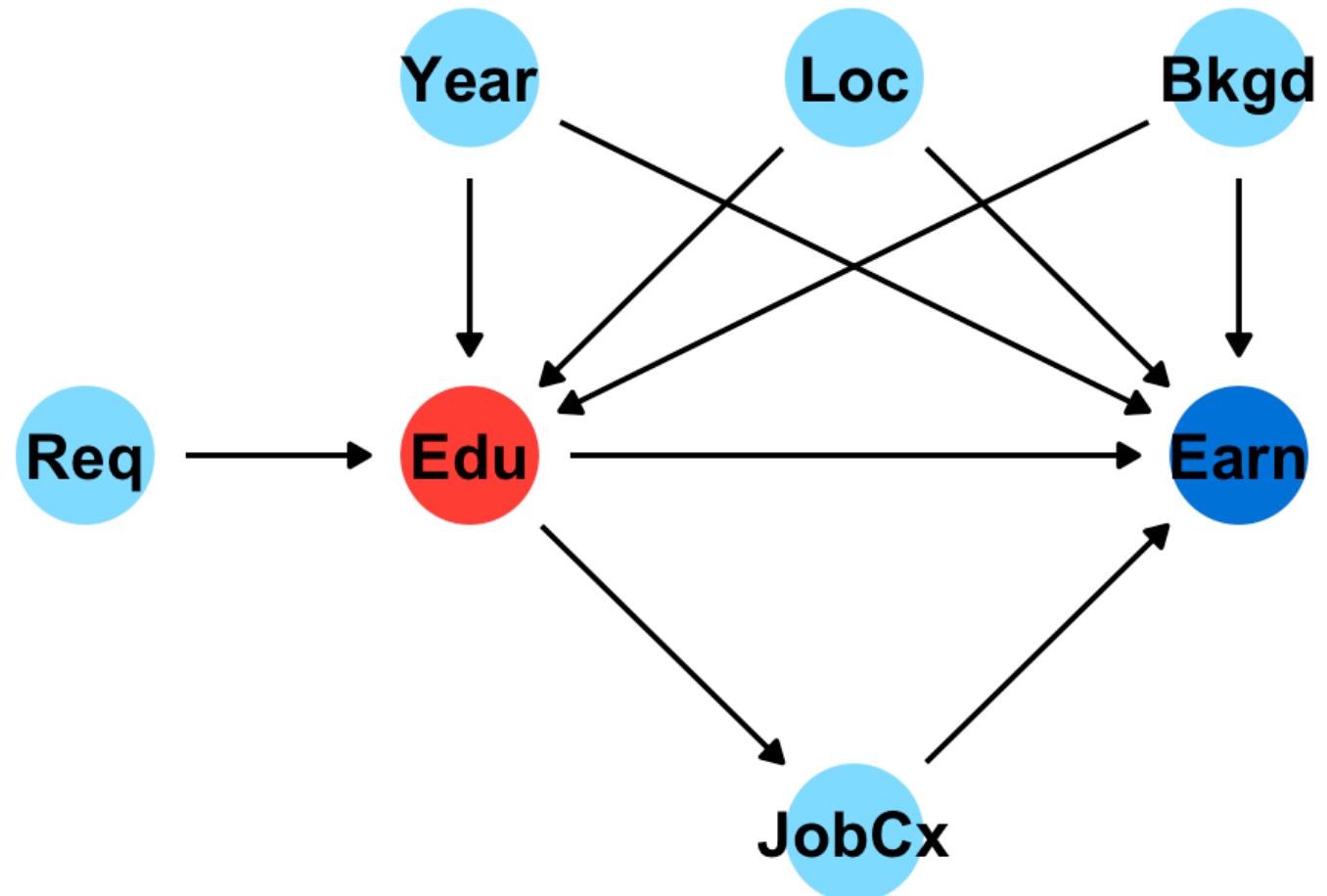
### 3. Draw arrows

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



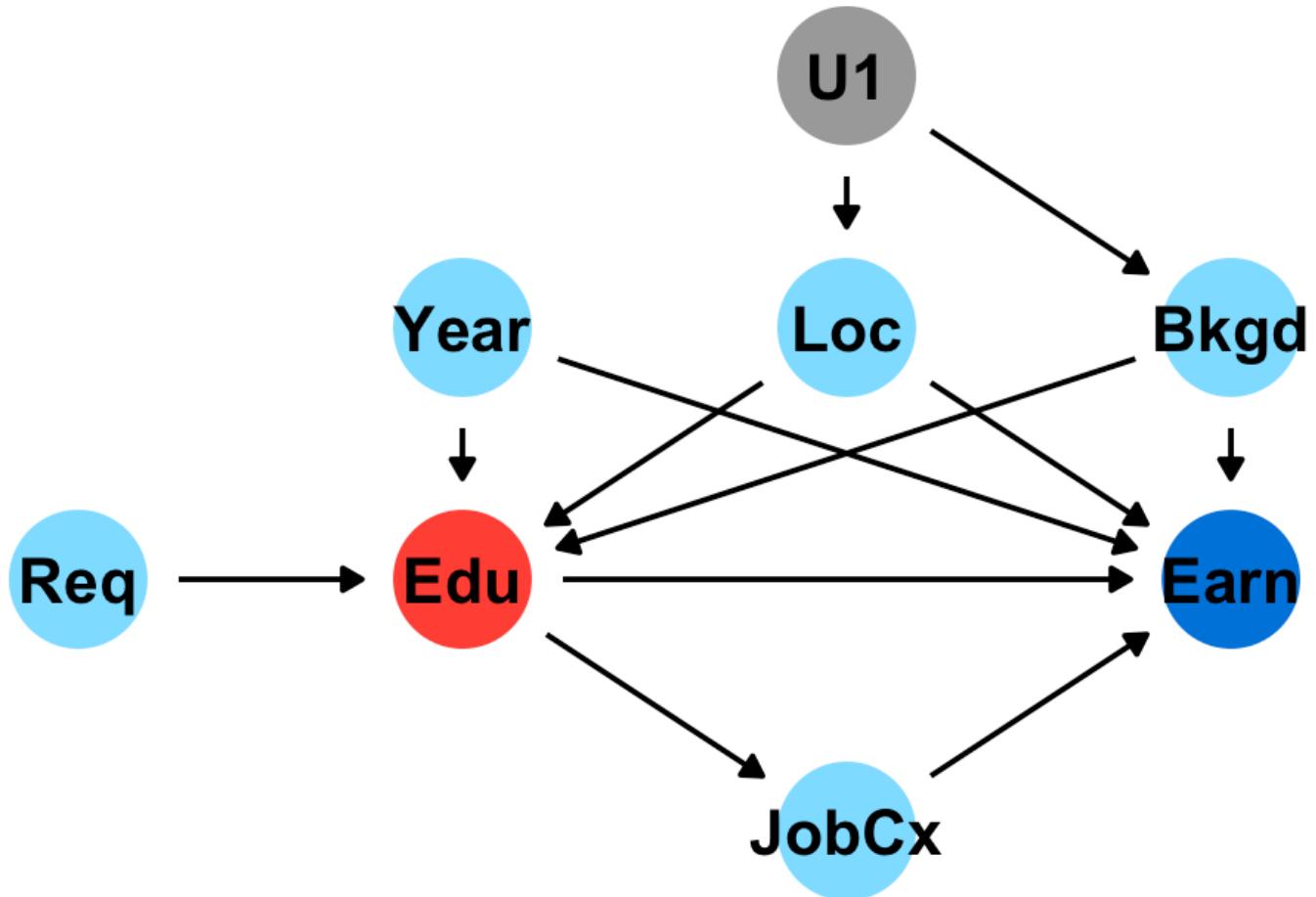
### 3. Draw arrows

Education causes  
job earnings



### 3. Draw arrows

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



# Let the computer do this!

[dagitty.net](http://dagitty.net)

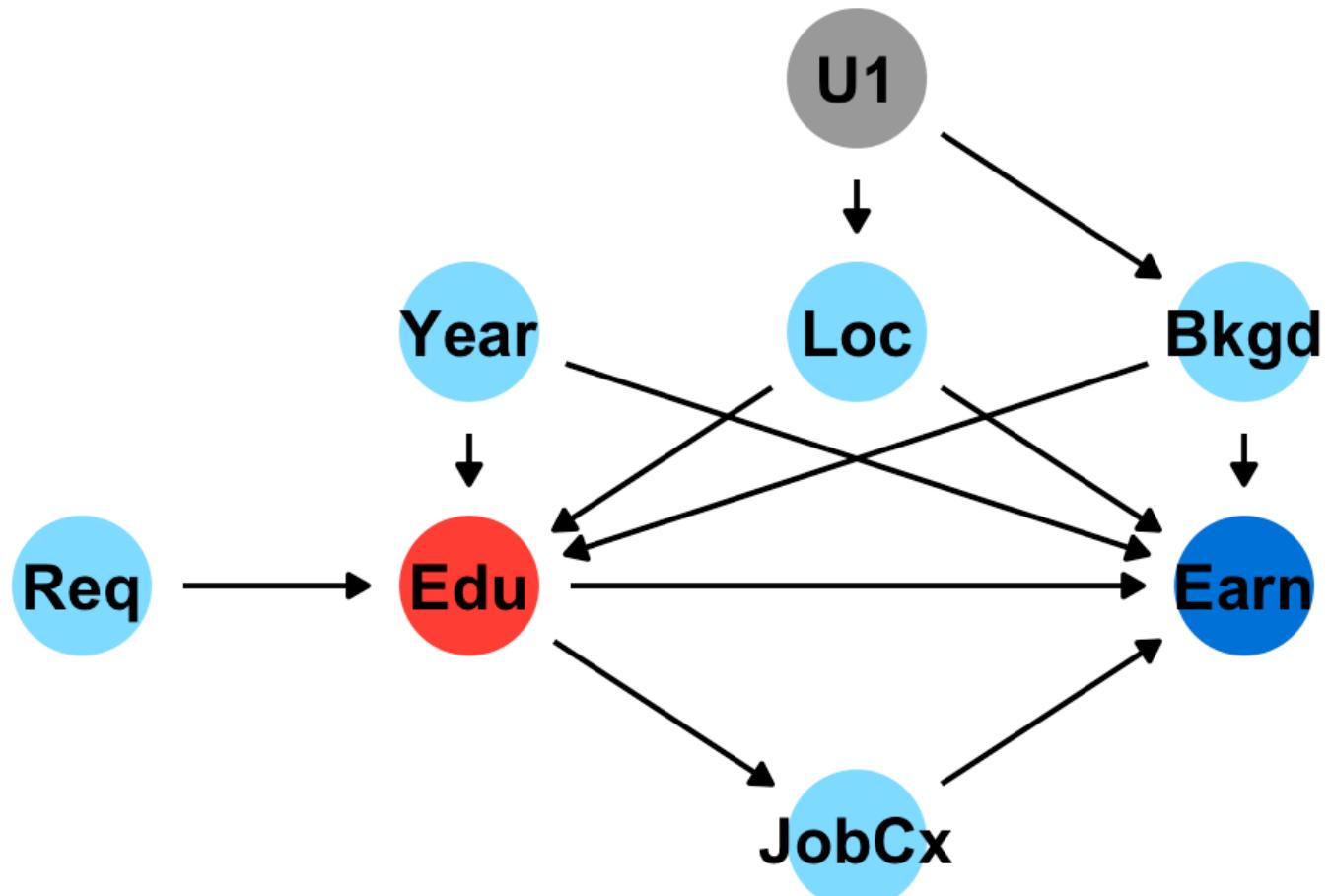
**ggdag** package in R

# Paths, doors, and adjustment

# 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 properly stripped and isolated

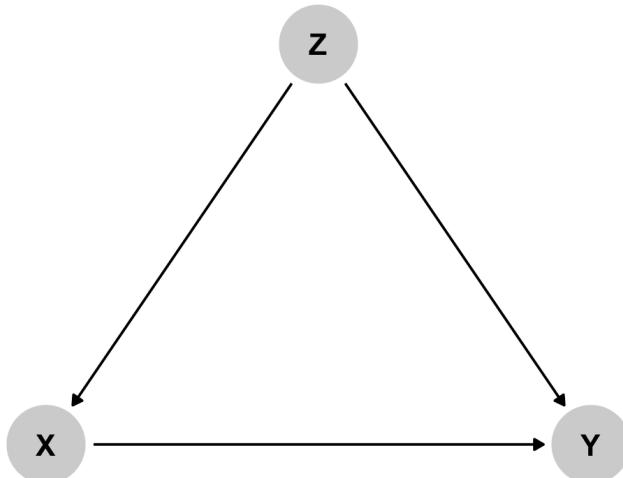
# Paths and associations

Arrows in a DAG transmit associations

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

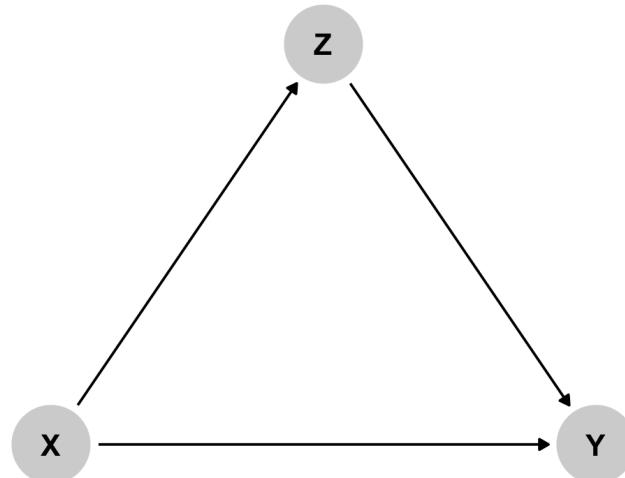
# Three types of associations

## Confounding



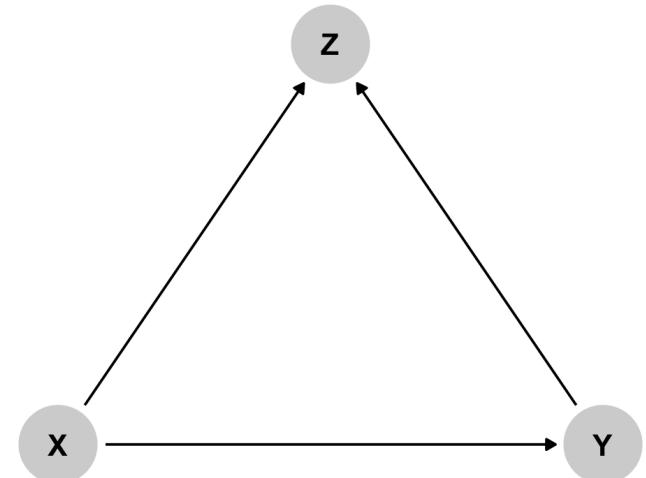
Common cause

## Causation



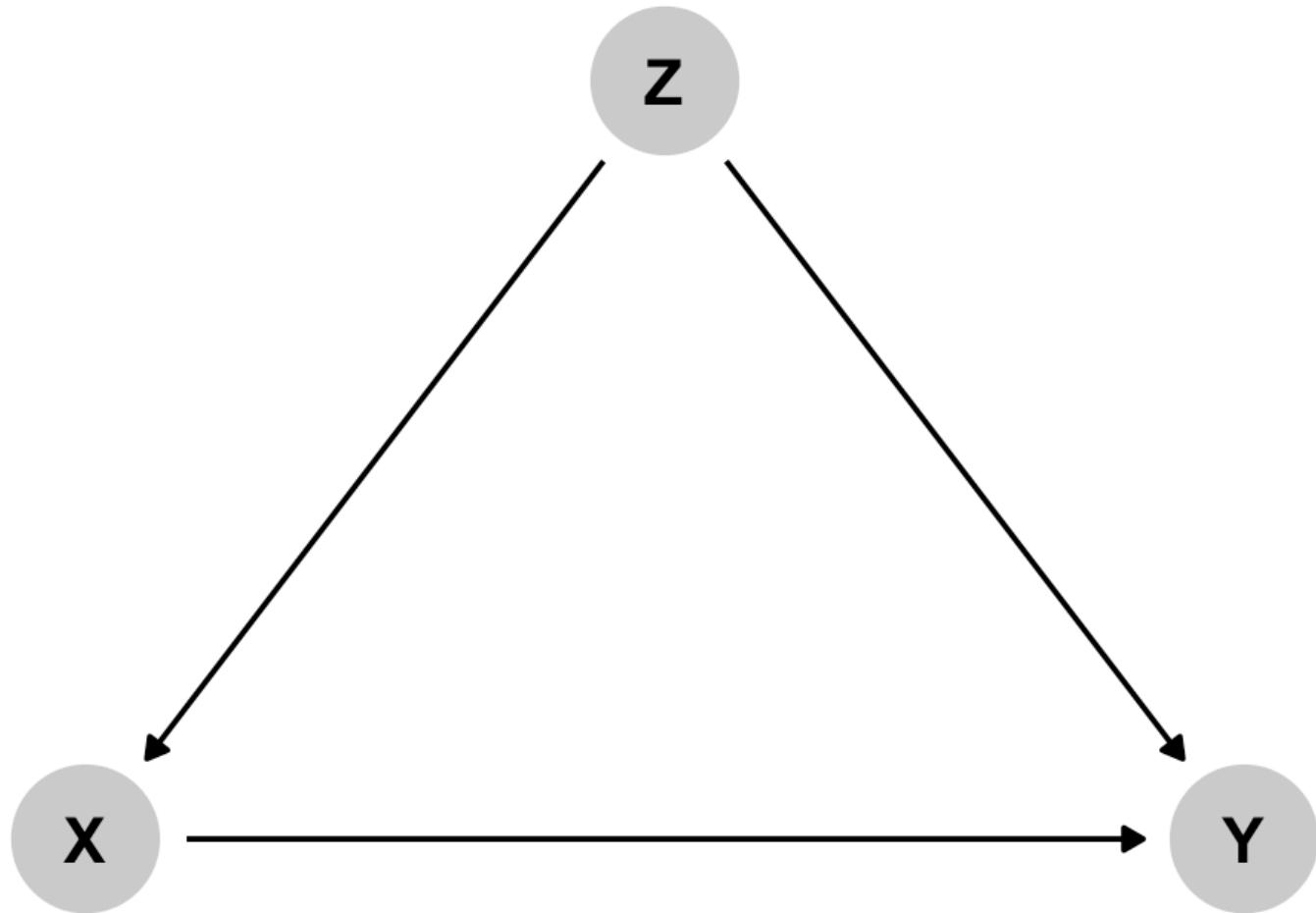
Mediation

## Collision



Selection /  
endogeneity

# Confounding

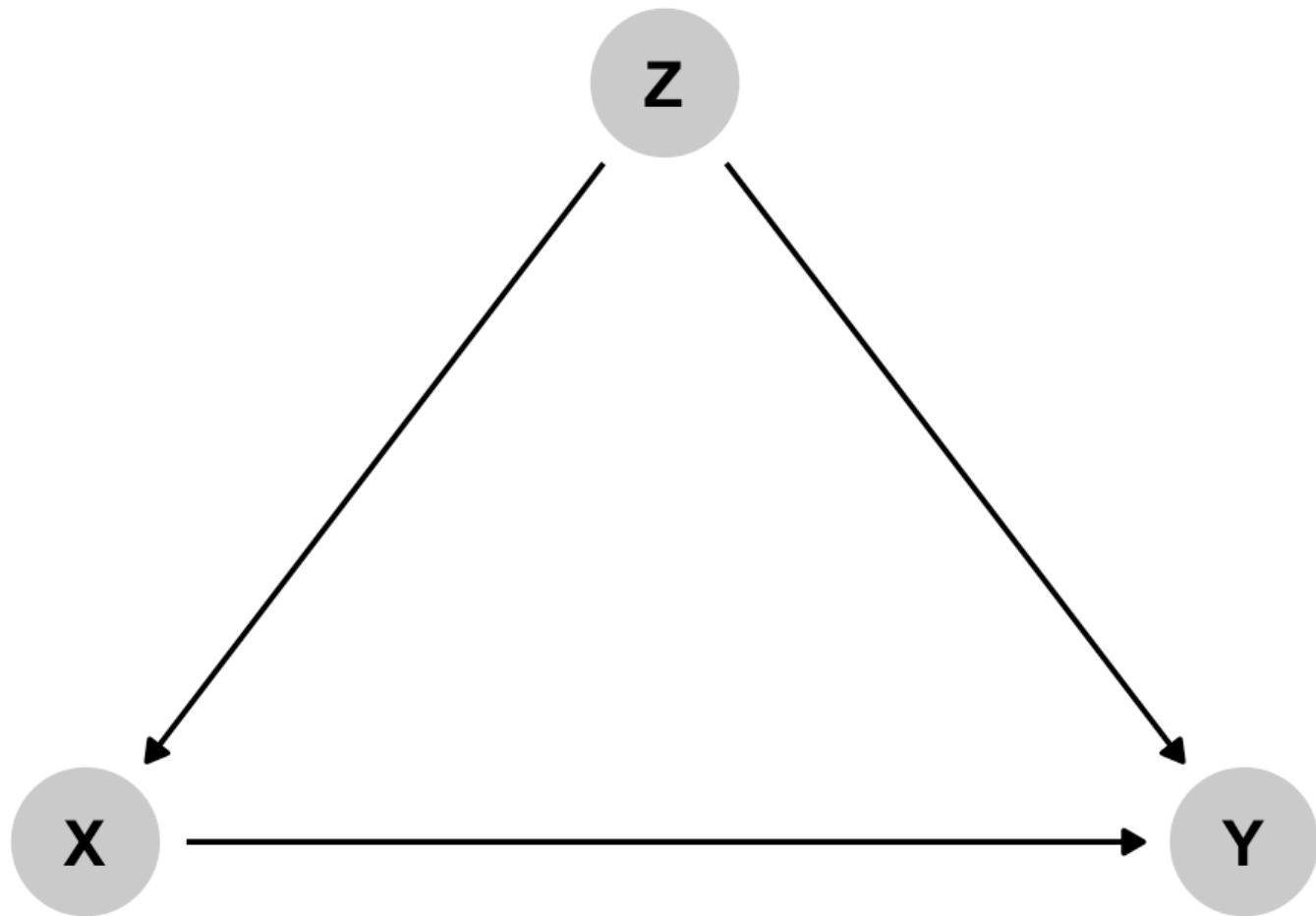


**X causes Y**

**But Z causes  
both X and Y**

**Z *confounds*  
the X → Y  
association**

# Paths



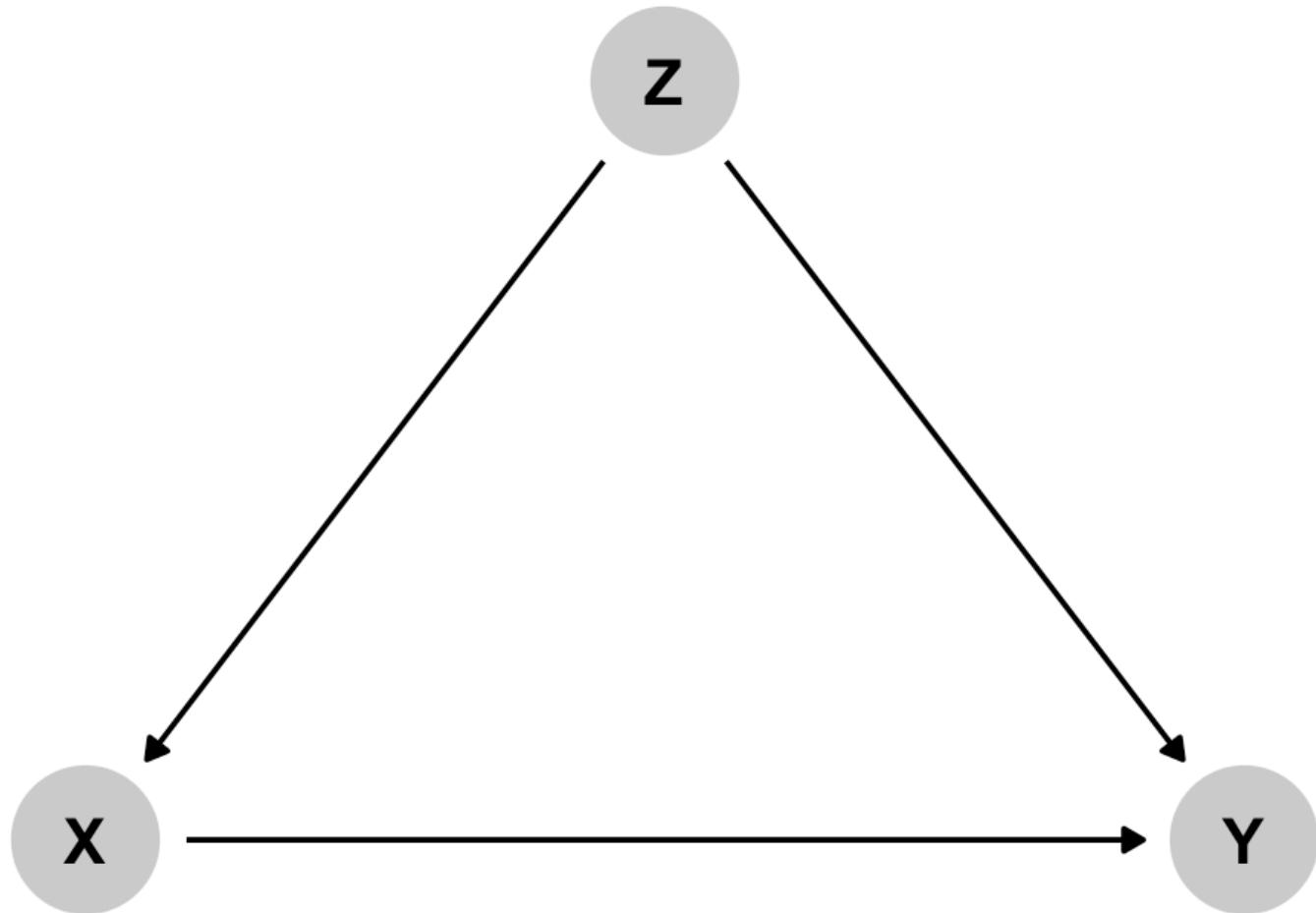
Paths between  
X and Y?

$X \rightarrow Y$

$X \leftarrow Z \rightarrow Y$

Z is a  
*backdoor*

# *d*-connection

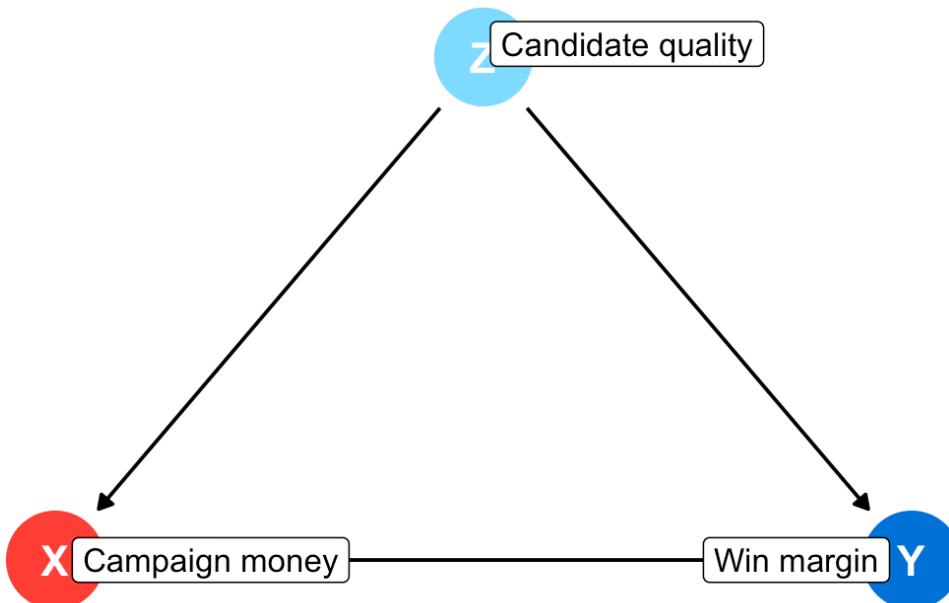


**X and Y are  
"d-connected"  
because  
associations can  
pass through Z**

**The relationship  
between X and Y is  
not identified /  
isolated**

# Effect of money on elections

What are the paths  
between money and win margin?

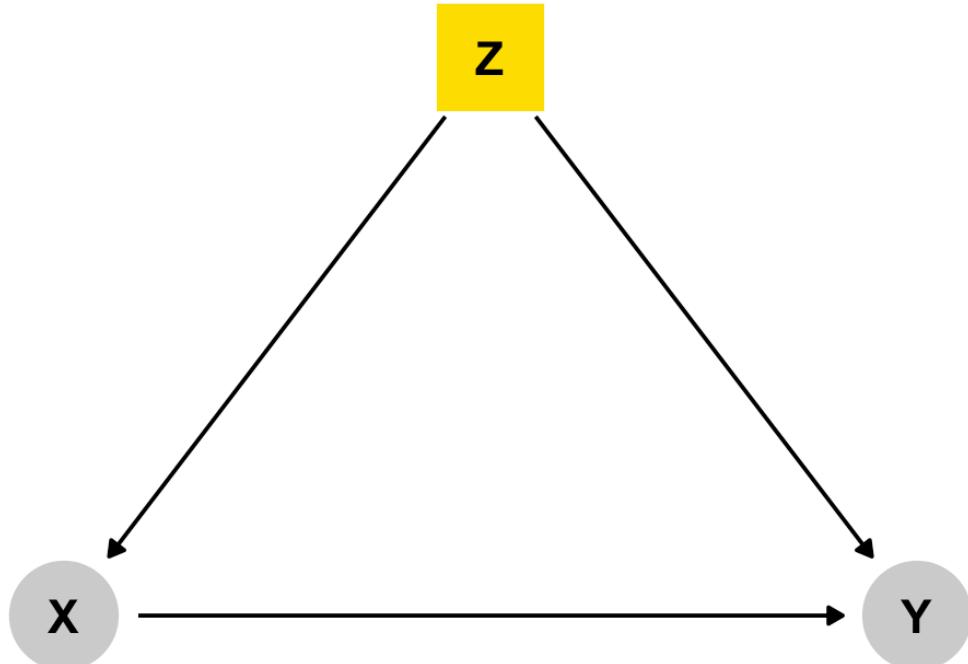


Money → Margin

Money ← Quality → Margin

Quality is a *backdoor*

# Closing doors



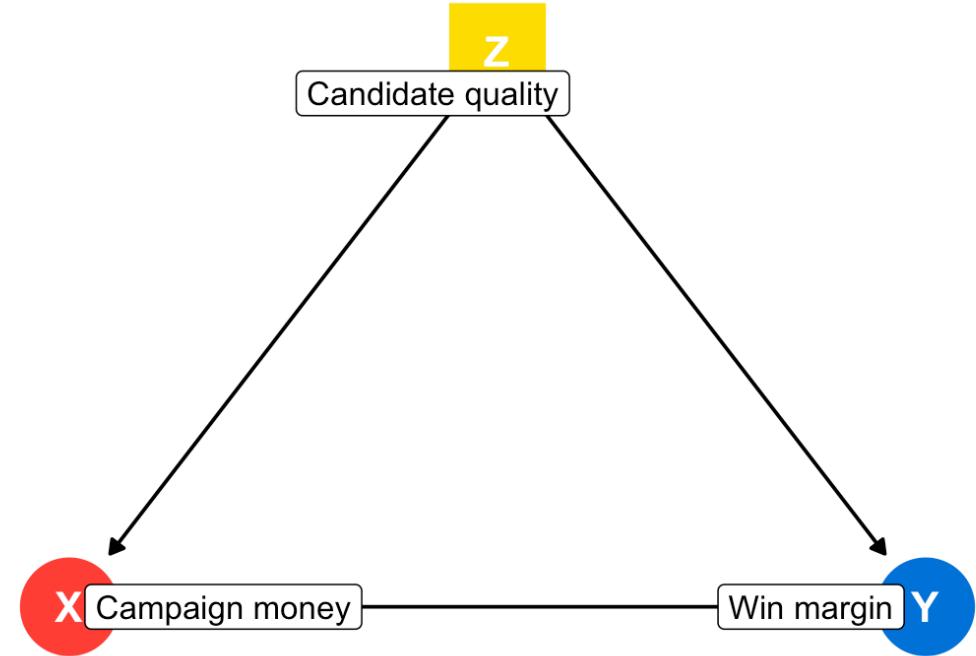
Close the backdoor  
by adjusting for **Z**

# Closing doors

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

Find the part of win margin that is explained by quality, remove it. 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.

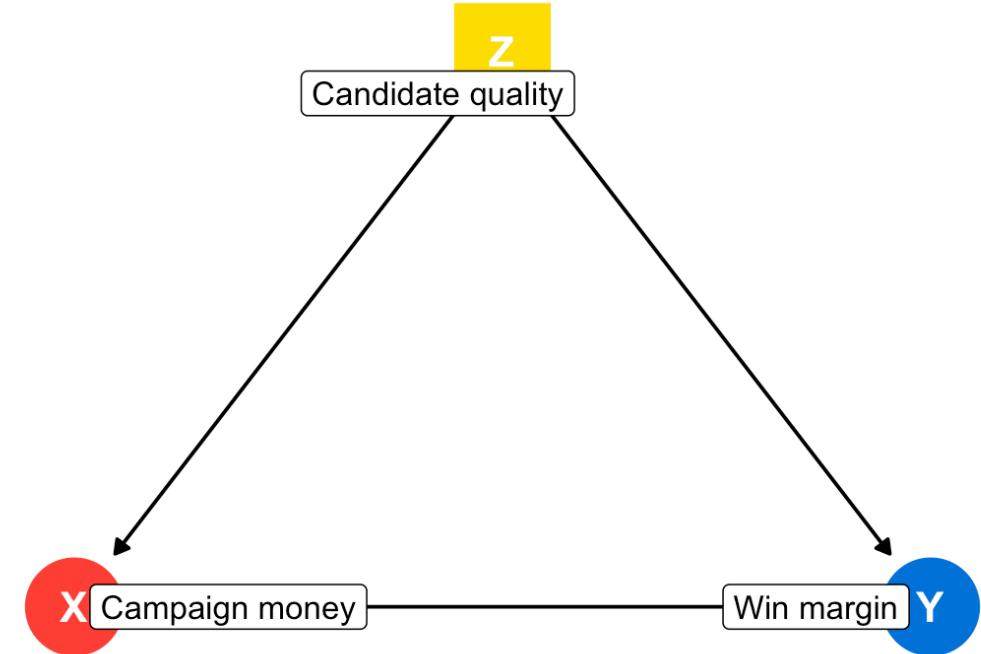


# Closing doors

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

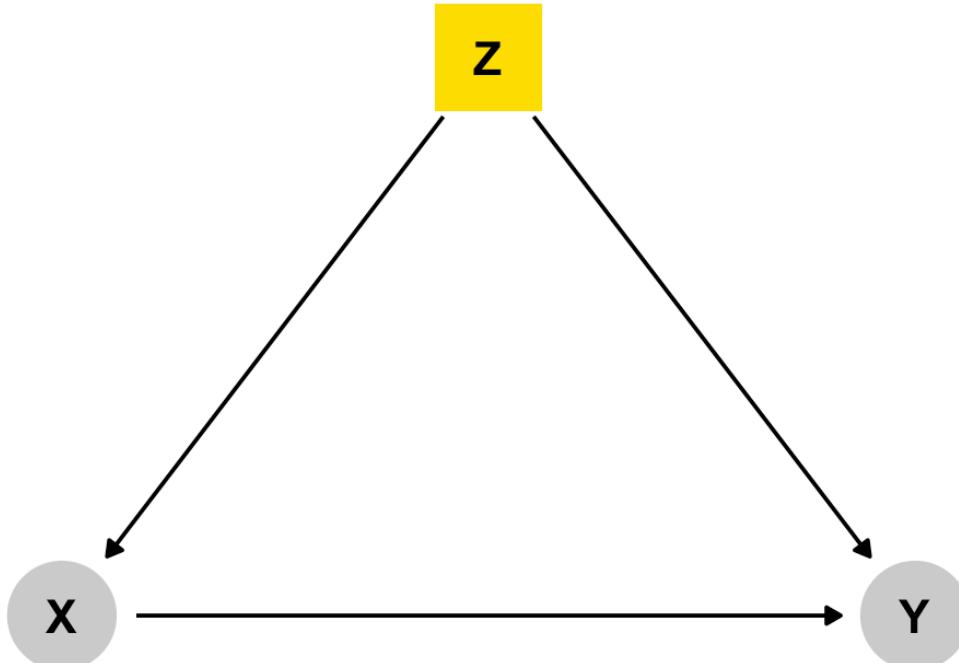
Win margin =  $\beta_0 + \beta_1$  Campaign money +  
 $\beta_2$  Candidate quality +  $\varepsilon$

Matching

Stratifying

Inverse probability weighting

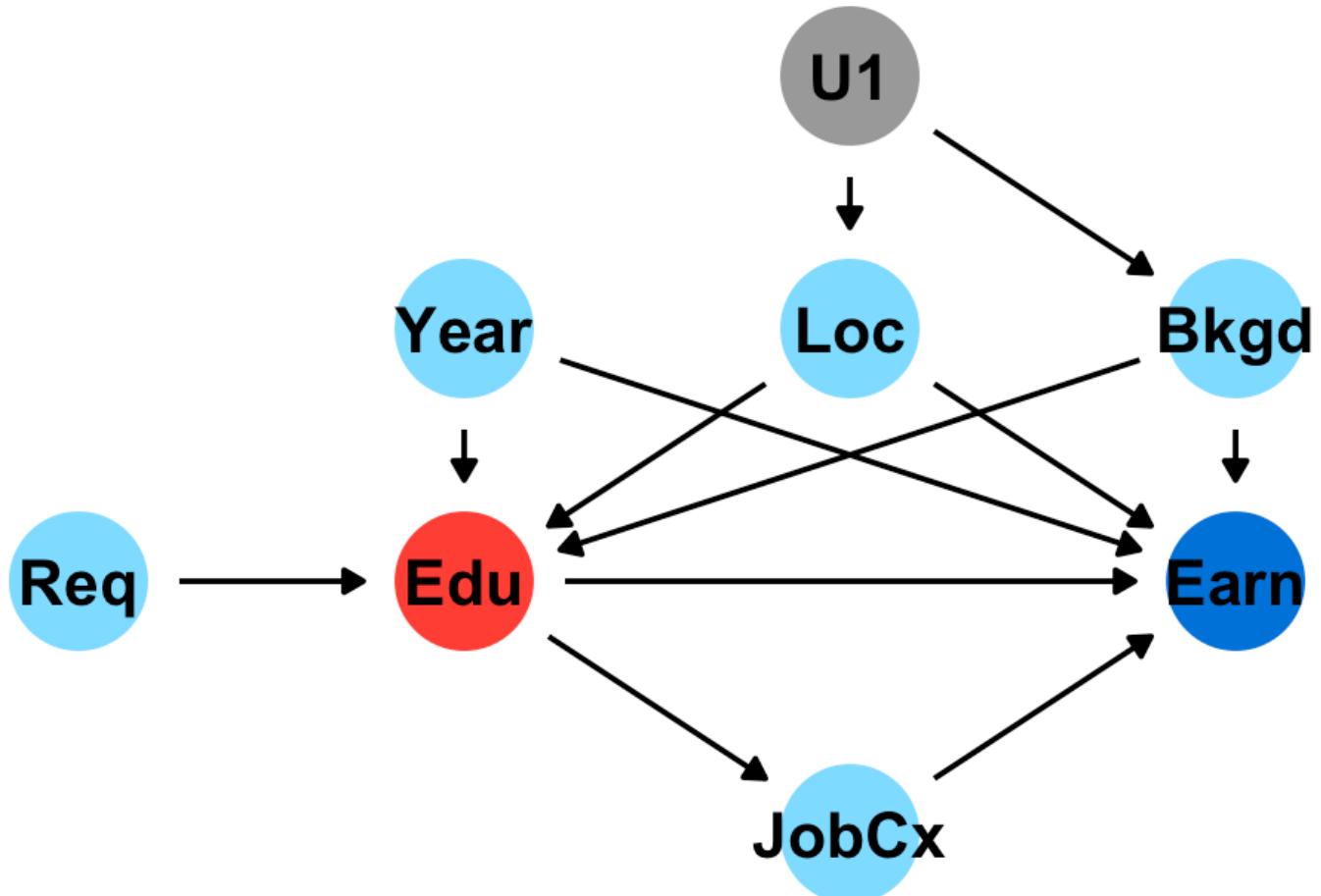
# *d*-separation



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

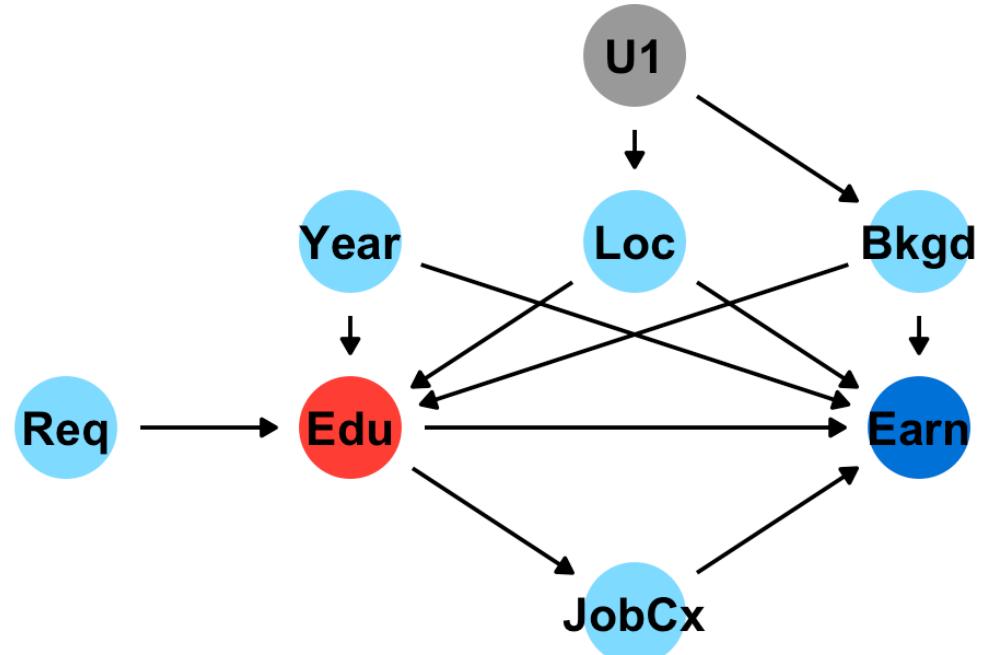
Education ← Background → Earnings

Education ← Background ← U1 → Location → Earnings

Education ← Location → Earnings

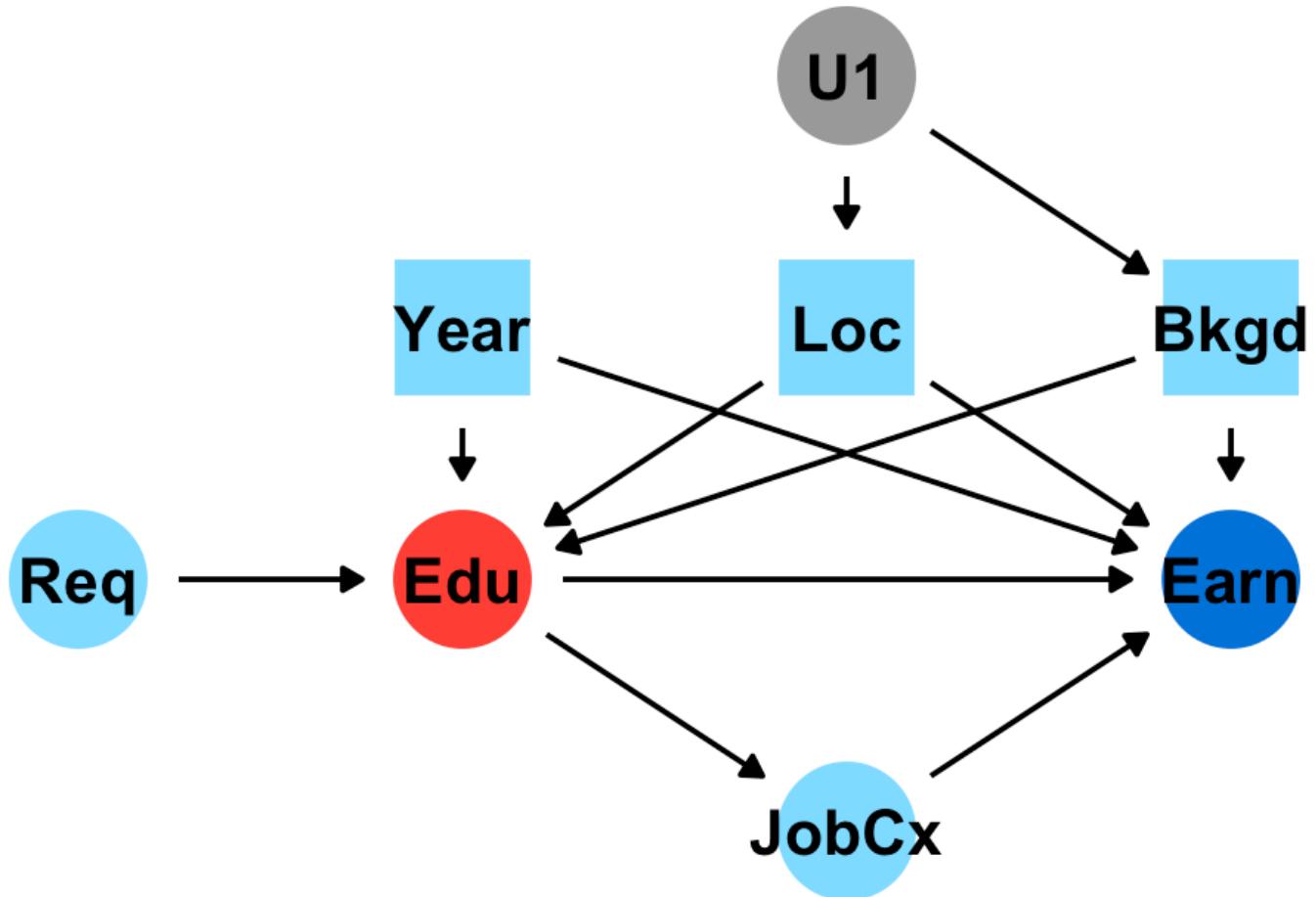
Education ← Location ← U1 → Background → Earnings

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](http://dagitty.net)

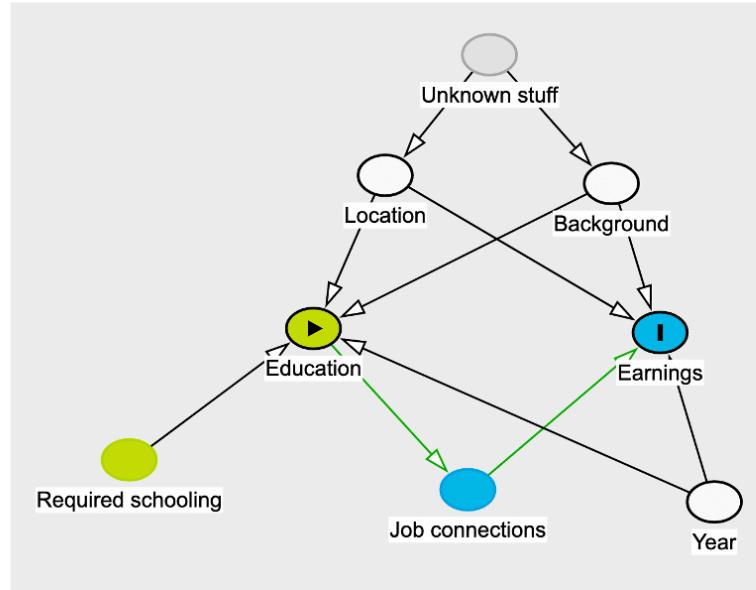
## The **ggdag** and **dagitty** packages in R

# 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

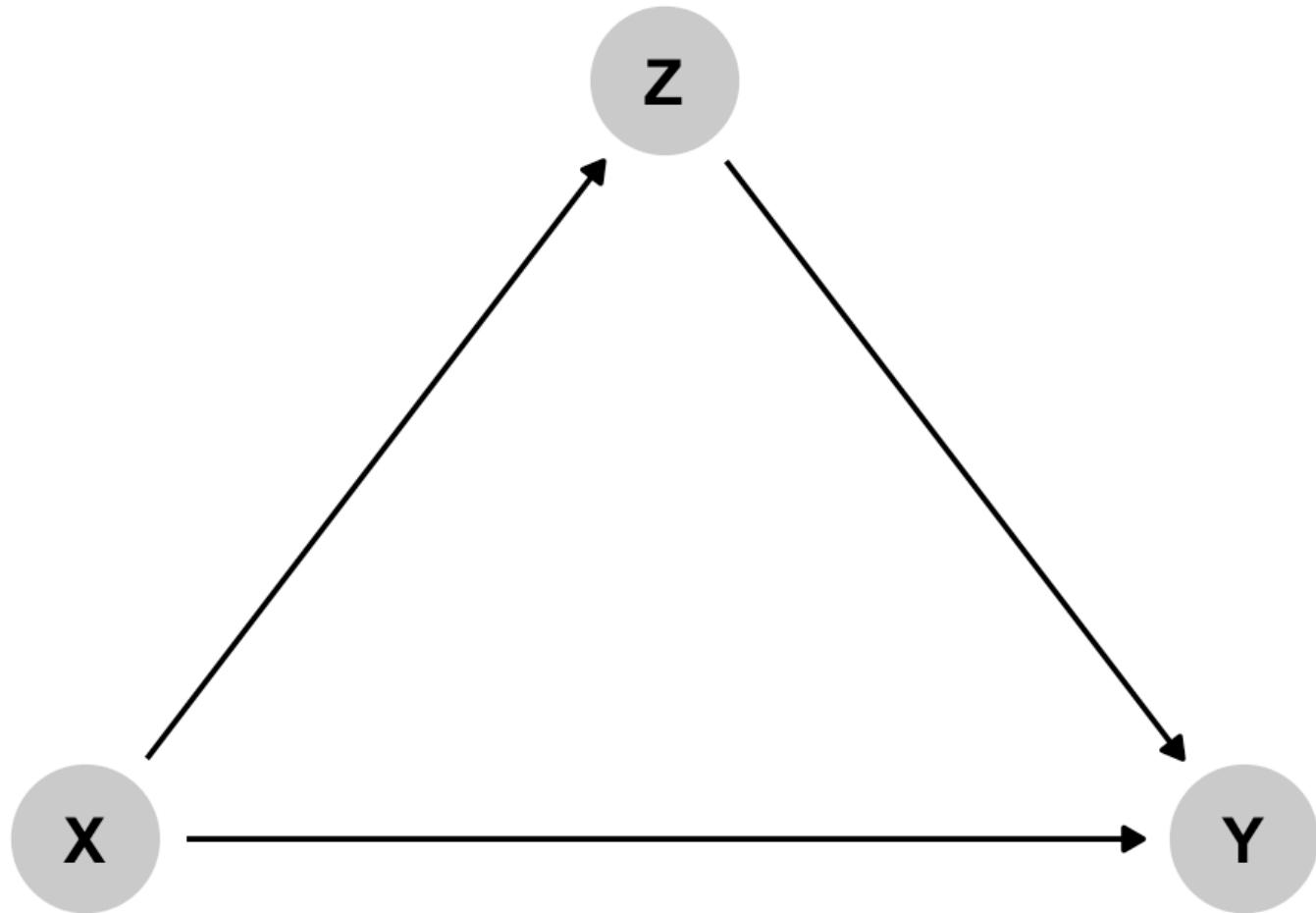


## Testable implications

The model implies the following conditional independences:

- Education  $\perp$  Earnings | Background, Job connections, Location, Year
- Required schooling  $\perp$  Job connections | Education
- Required schooling  $\perp$  Year
- Required schooling  $\perp$  Earnings | Background, Job connections, Location, Year
- Required schooling  $\perp$  Earnings | Background, Education, Location, Year
- Required schooling  $\perp$  Background
- Required schooling  $\perp$  Location
- Job connections  $\perp$  Year | Education
- Job connections  $\perp$  Background | Education
- Job connections  $\perp$  Location | Education
- Year  $\perp$  Background
- Year  $\perp$  Location

# Causation

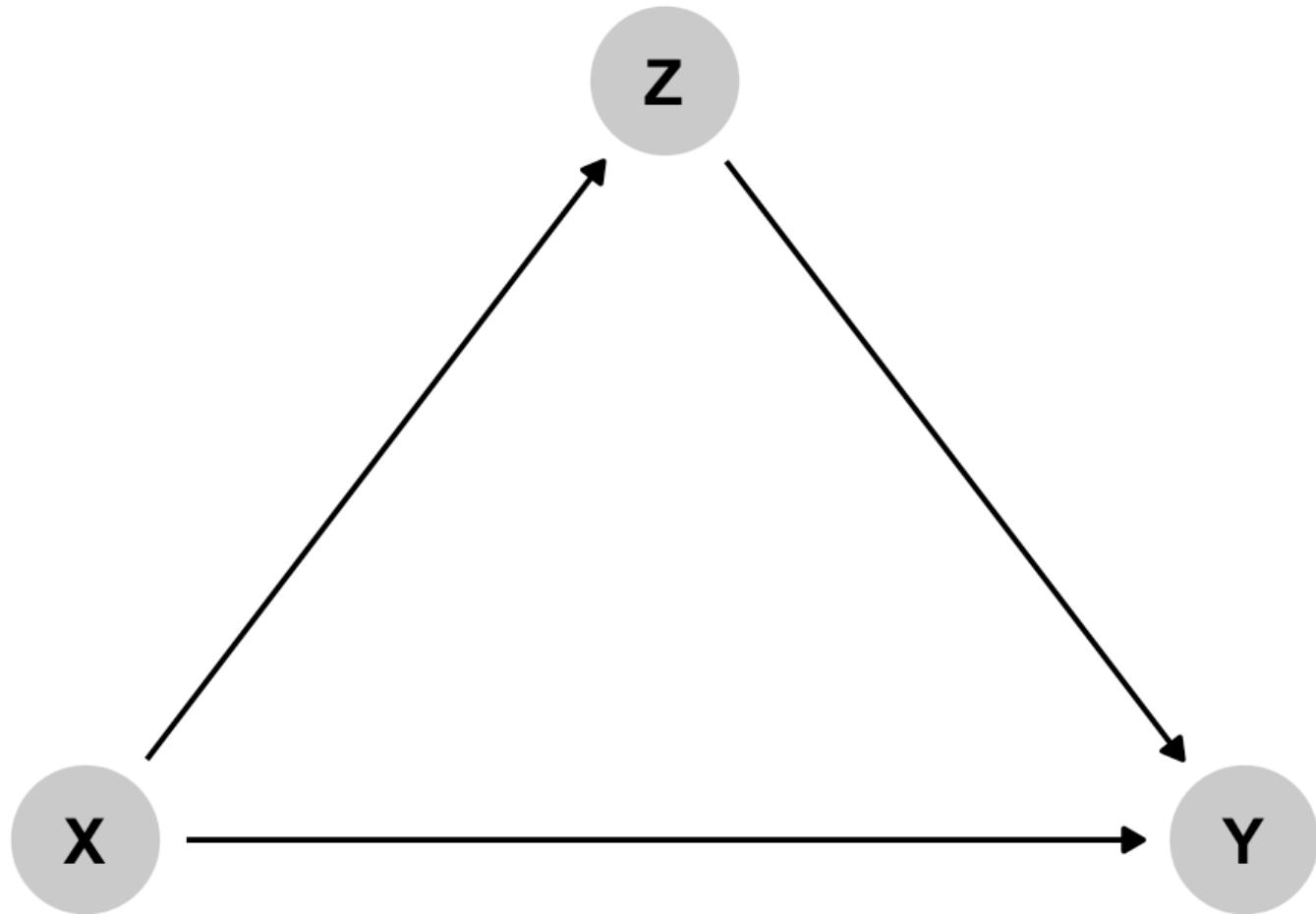


**X causes Y**

**X causes  
Z which  
causes Y**

**Should you  
control for Z?**

# Causation

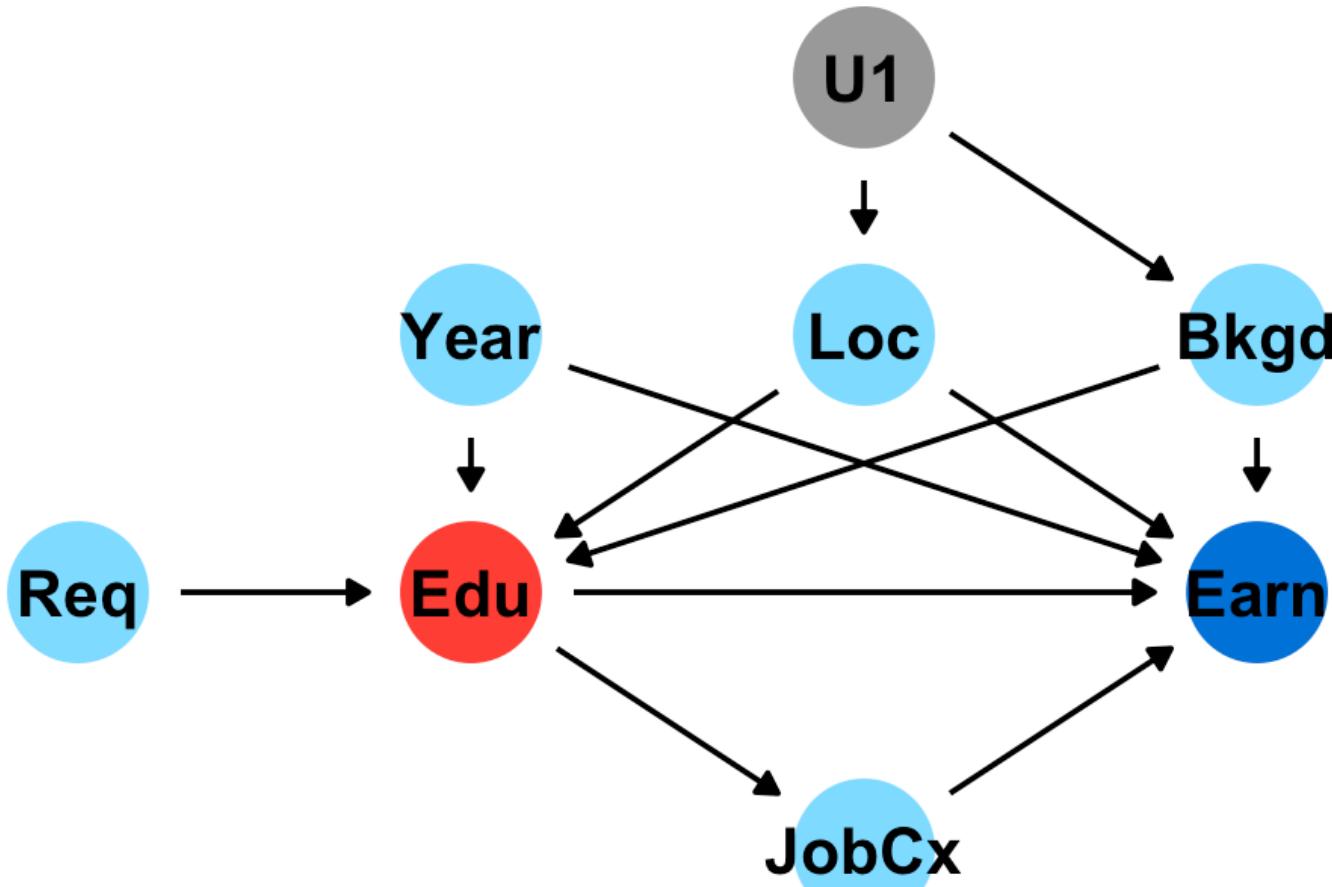


Should you  
control for Z?

No!

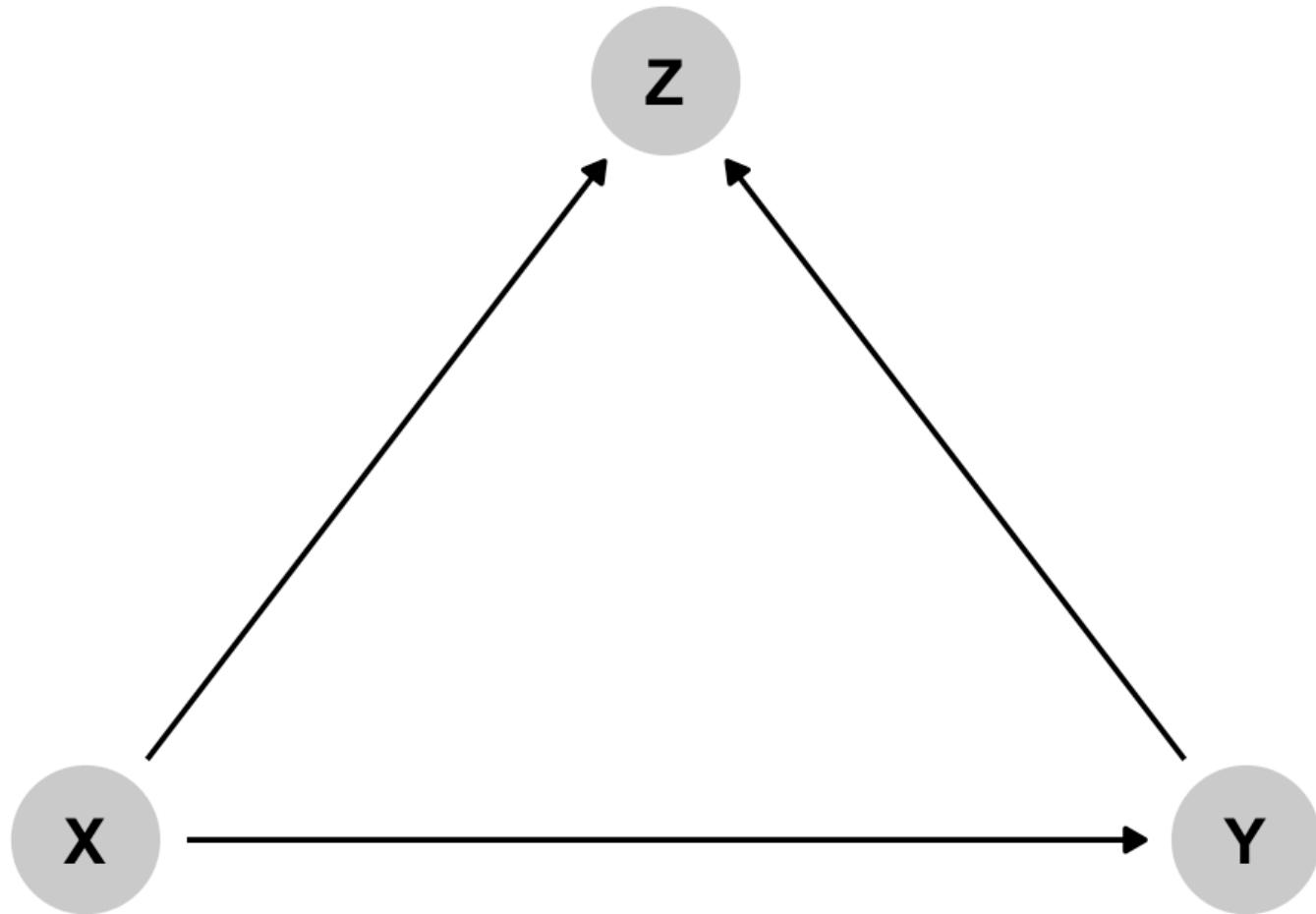
Overcontrolling

# Causation and overcontrolling



Should you control  
for job  
connections?

# Colliders



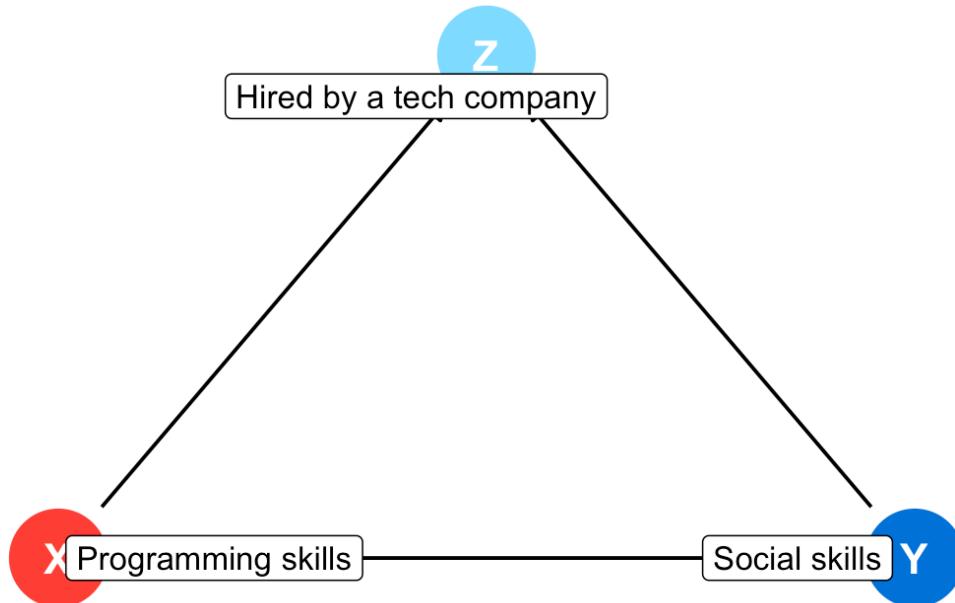
**X causes Z**

**Y causes Z**

**Should you  
control for Z?**

# Programming and social skills

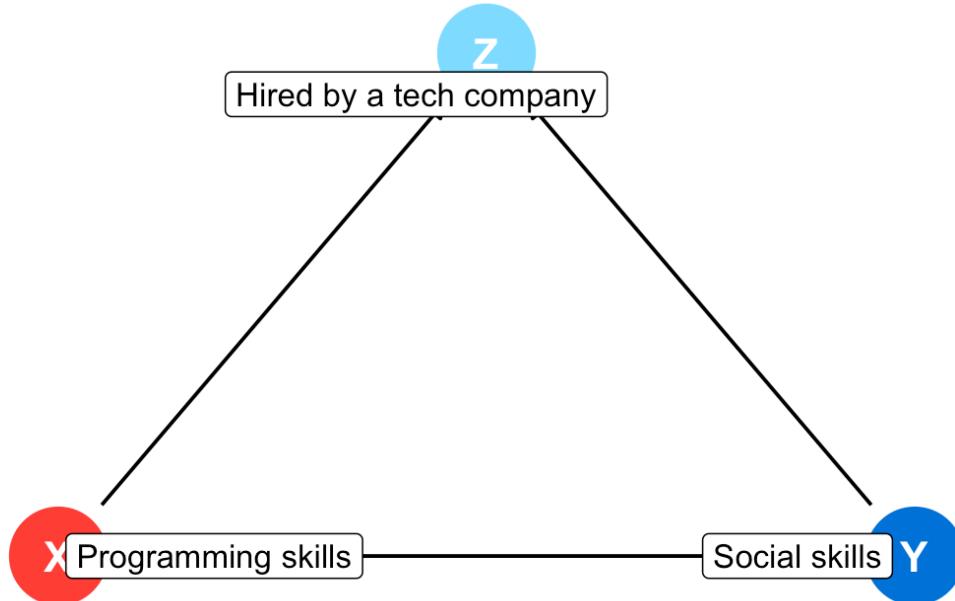
## Do programming skills reduce social skills?



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?

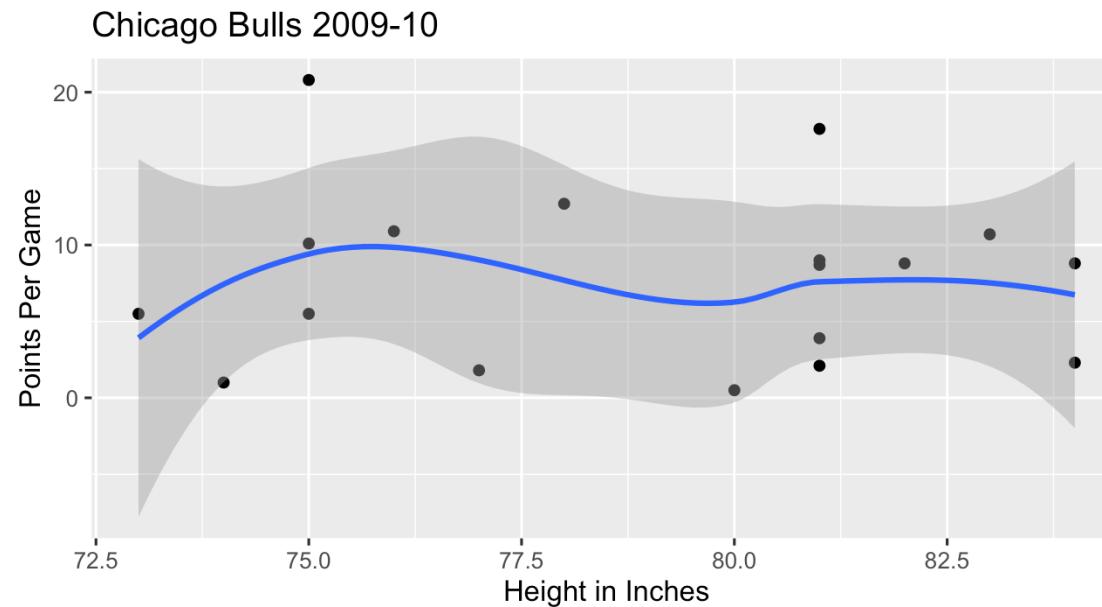


**No! Hired by a tech company is a collider and we controlled for it.**

**This inadvertently connected the two.**

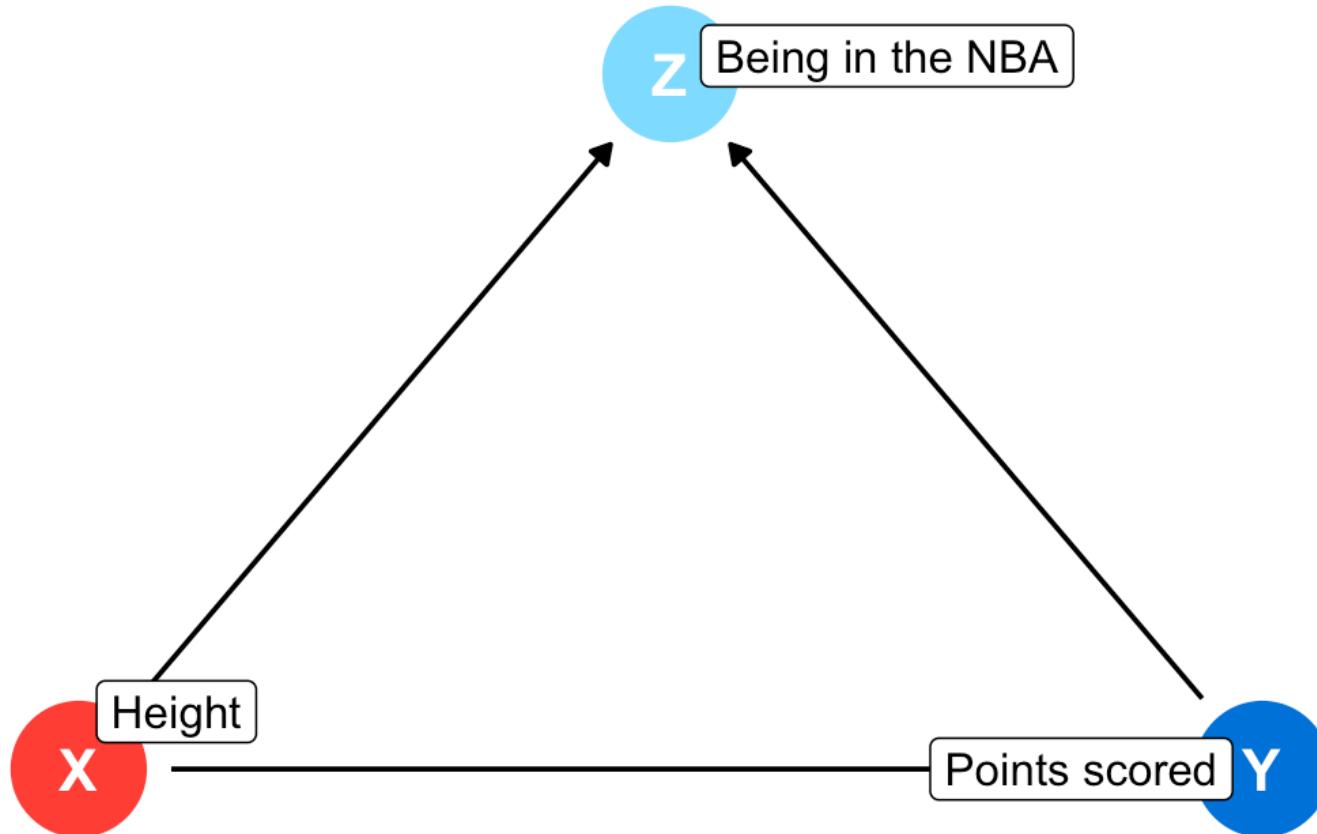
Colliders can create fake causal effects

Colliders can hide real causal effects



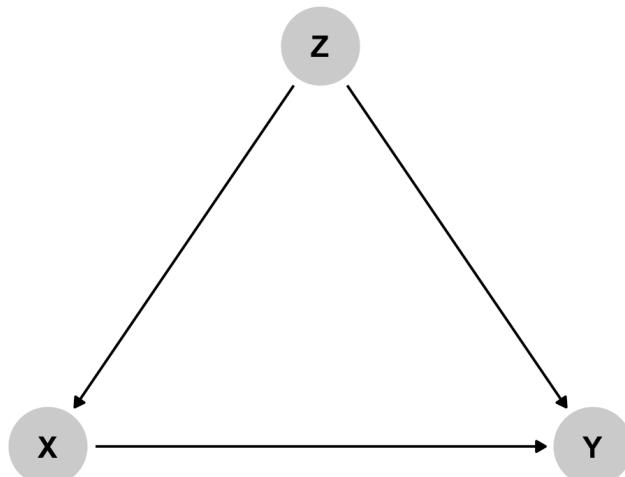
Height is unrelated to basketball skill... among NBA players

# Colliders and selection bias



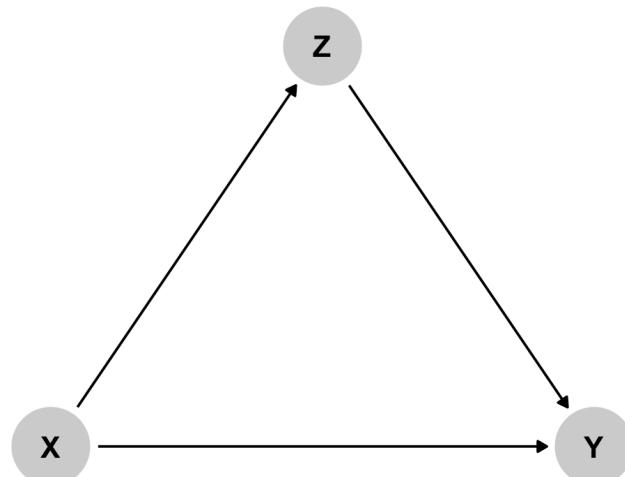
# Three types of associations

## Confounding



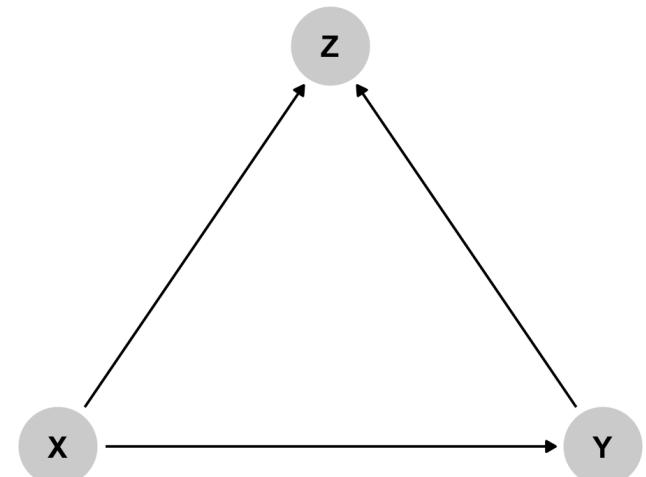
Common cause

## Causation



Mediation

## Collision



Selection /  
endogeneity