Measurement and DAGs

February 5, 2020

PMAP 8521: Program Evaluation for Public Service
Andrew Young School of Policy Studies
Spring 2020

Fill out your reading report on iCollege!

Plan for today

Abstraction, stretching, and validity

Causal models

Equations, paths, doors, and adjustment

Abstraction, stretching, and validity

Indicators

Inputs, activities, & outputs	Outcomes
Generally directly measurable	Harder to directly measure
# of citations mailed, % increase in grades, etc.	Commitment to school, reduced risk factors






Conceptual stretching

Ladder of abstraction for witches

Mammal

Trolls, elves, gods/goddesses

Enmagicked

Arwen, Winky, Athena

Female

Human

Salem witch trials

Young

Old

Elphaba

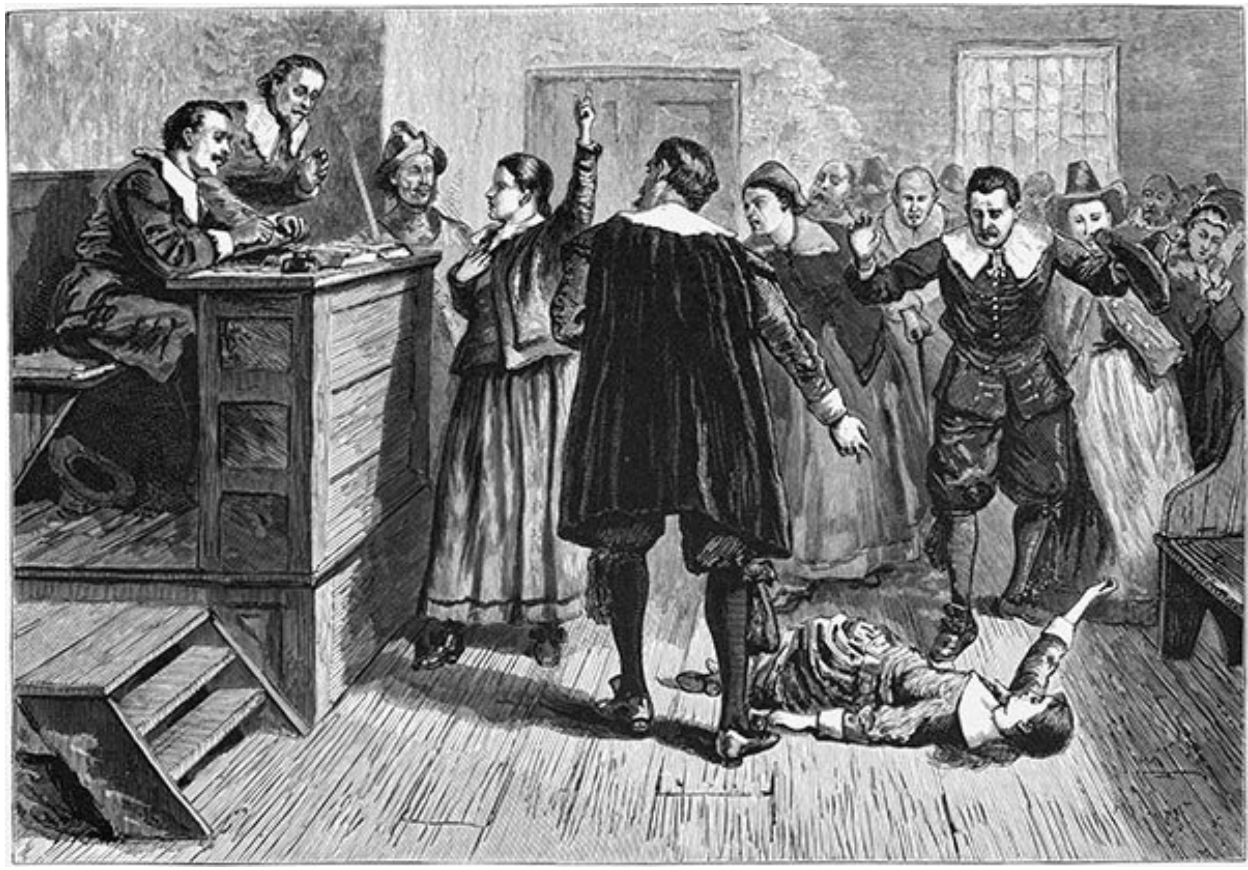
Hermione Granger

Student

Halloween decorations

Sabrina Spellman

Connection to theory



Practice

Choose an outcome

List all the possible attributes of that outcome

Build a ladder of abstraction with all the attributes

Determine which level is sufficient for showing an effect

Juvenile delinquency

School performance

Poverty

Outcomes and programs

Outcome variable

Thing you're measuring

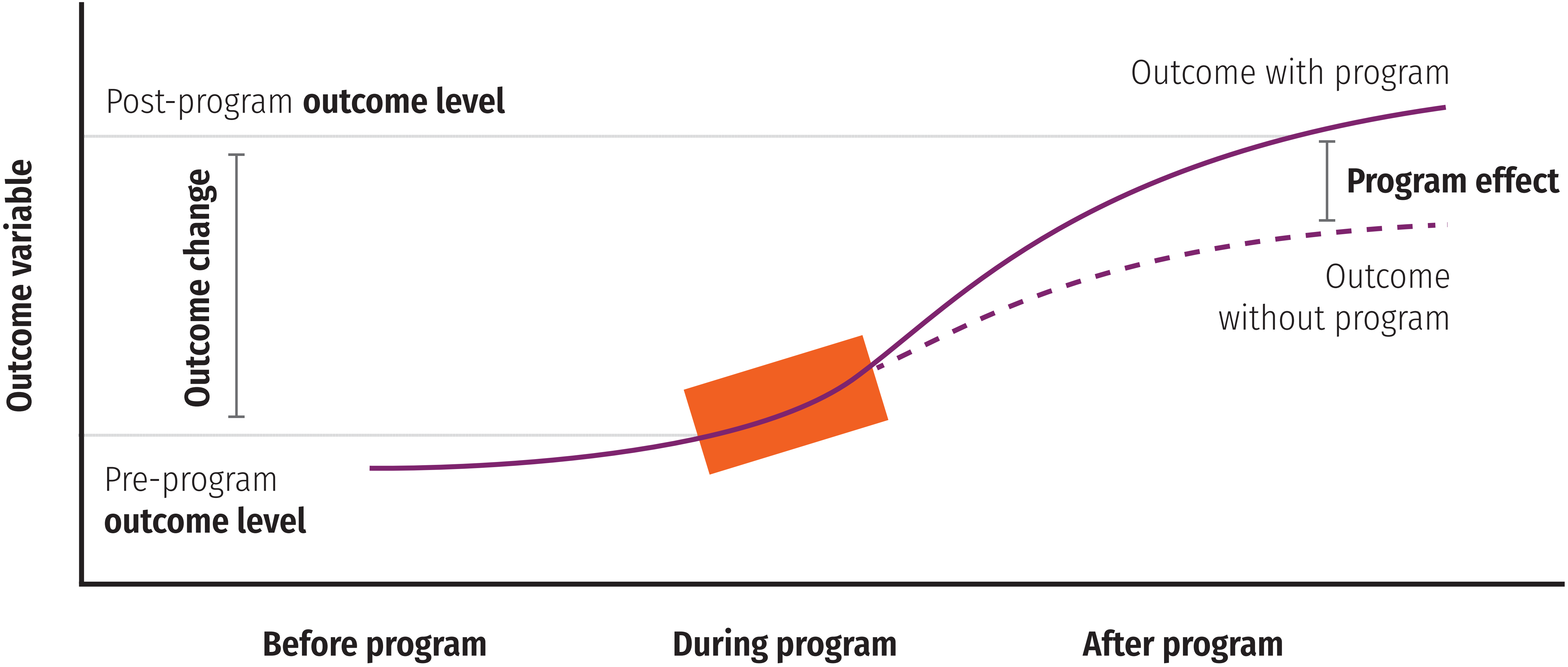
Outcome change

∆ in thing you're measuring over time

Program effect

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

Outcomes and programs



Connecting measurement to programs

Measurable definition of program effect

Ideal measurement

Feasible measurement

Connection to real world

Juvenile delinquency

School performance

Poverty

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

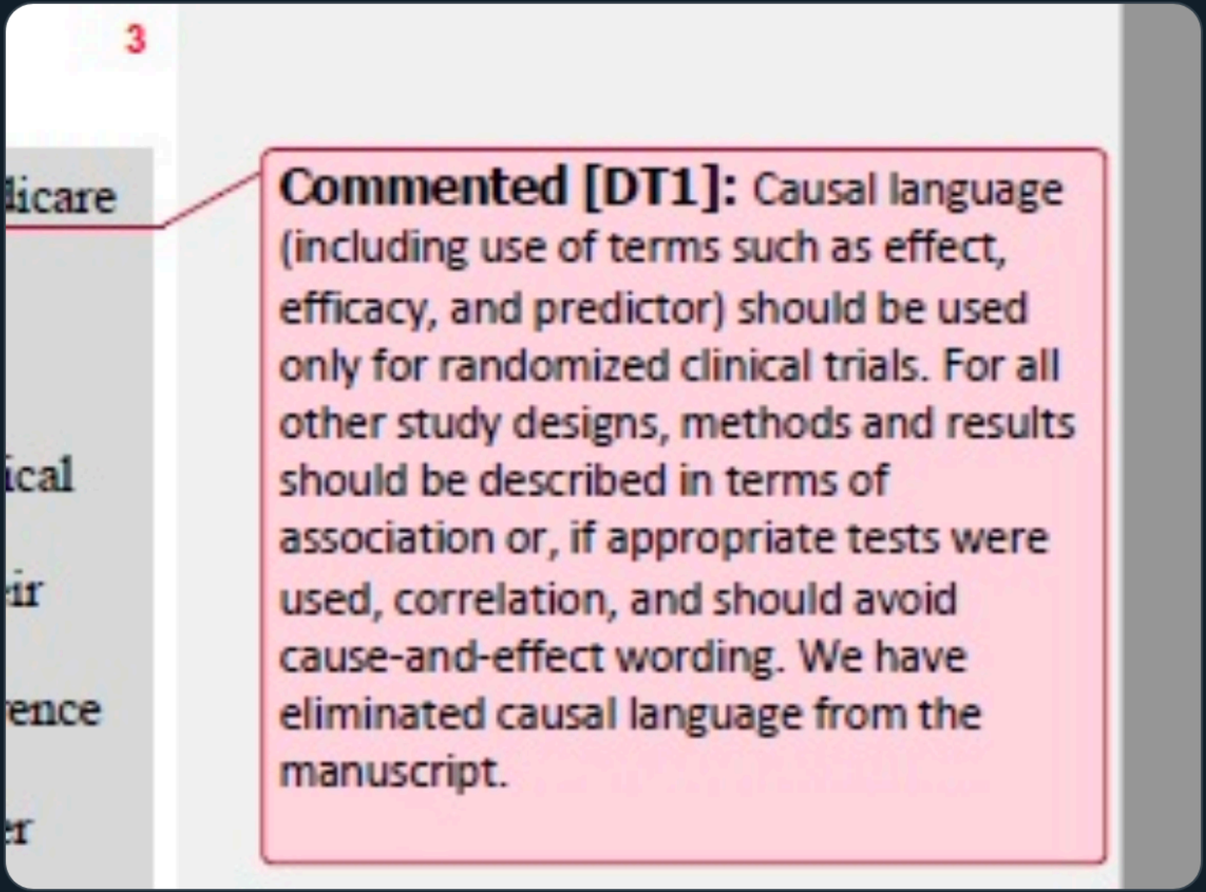
Can you prove causation with observational data?

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?


3

licare

ical

ir

ence

r

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.

5:02 PM · Jan 16, 2020 · TweetDeck

Seva
@SevaUT

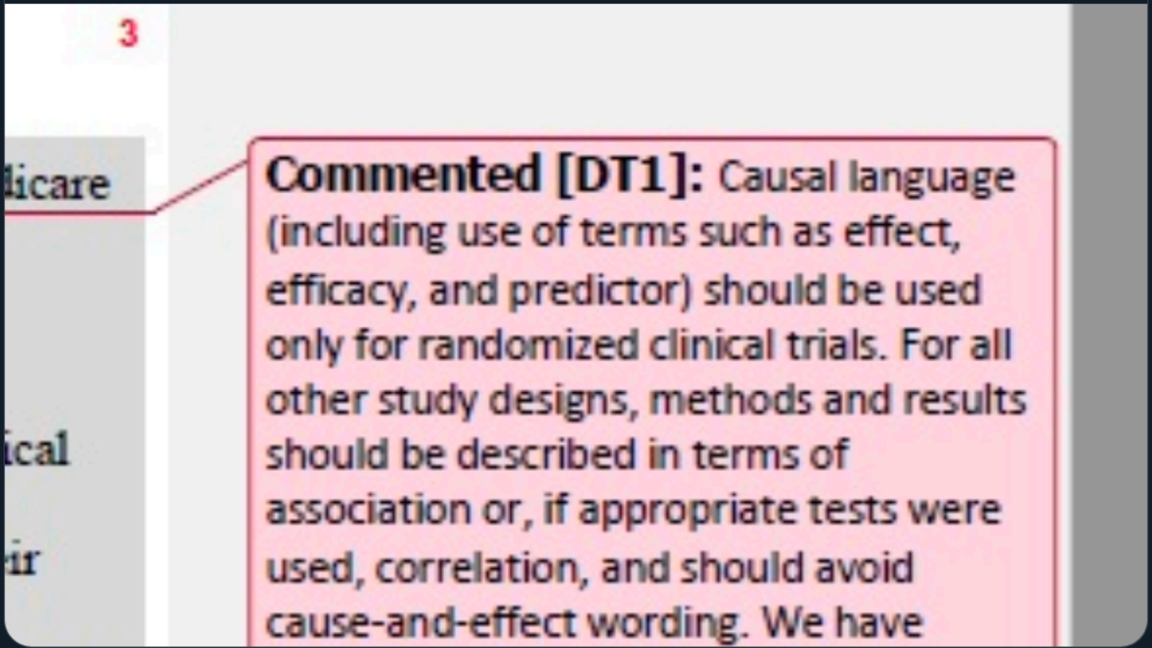
normal person: this rain is making us wet

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?


3

licare

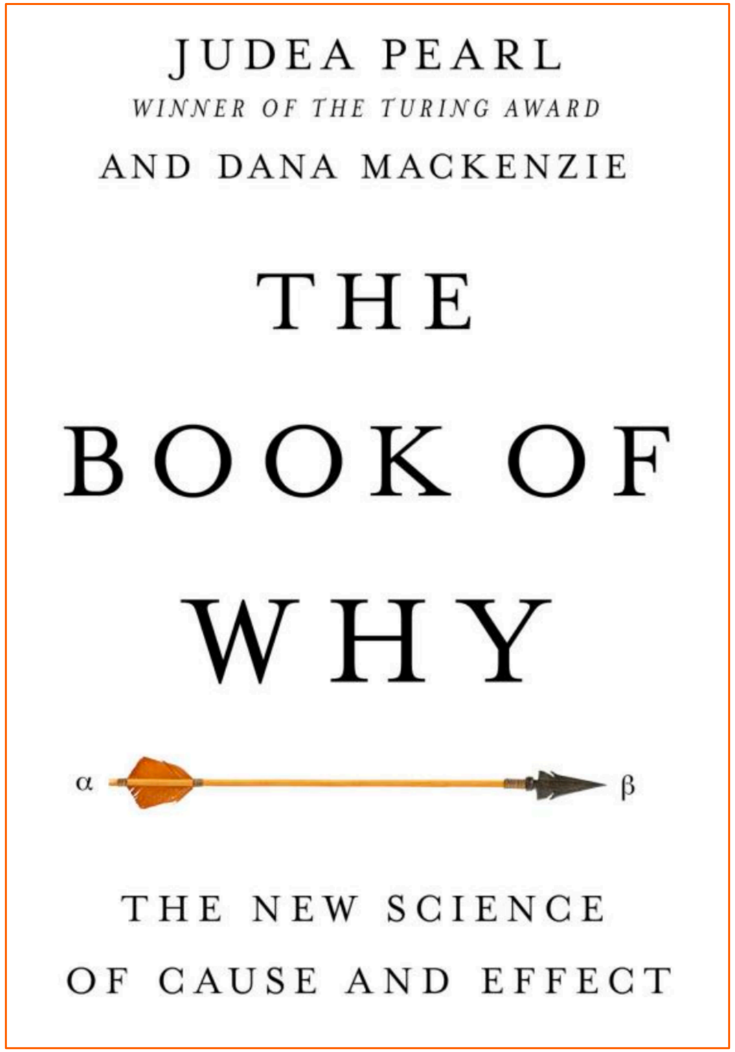
ical

ir

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

5:55 PM · Jan 17, 2020 · Twitter Web Client

The causal revolution



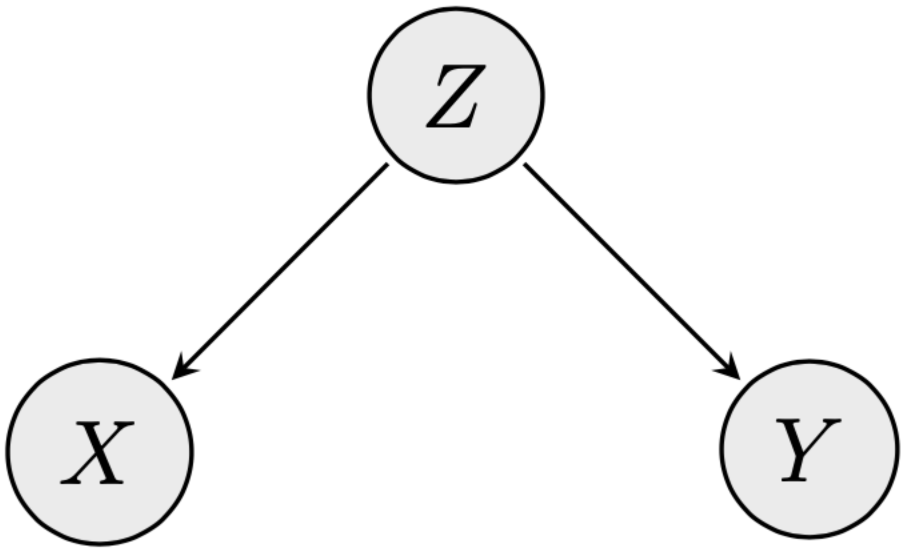
Causal diagrams

Directed acyclic graphs (DAGs)

Graphical model of the process that generates the data

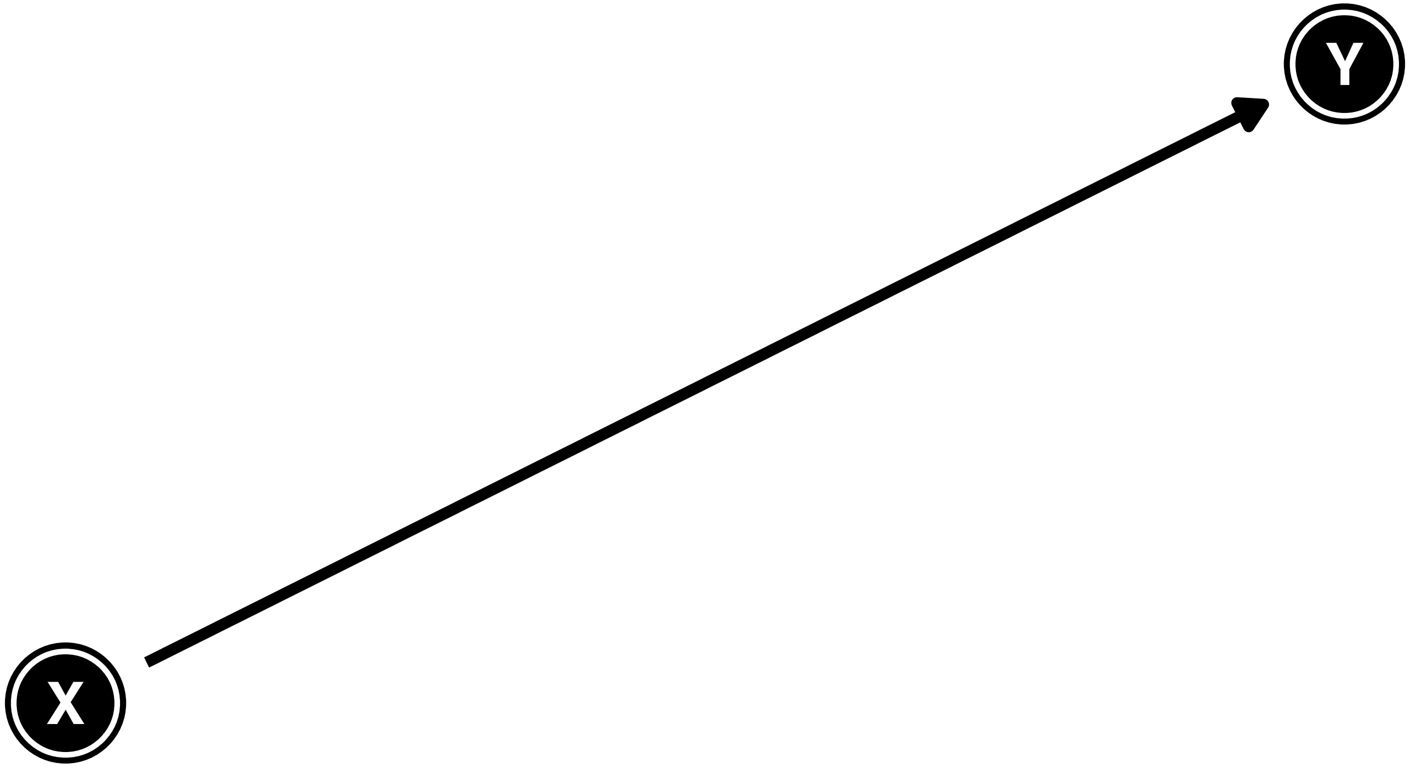
Maps your philosophical model

Fancy math ("do-calculus") tells you what to control for to find causation



DAGs

Directed acyclic graphs encode our understanding of the causal model (or philosophy)



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)

List anything that's relevant

Things that cause or are caused by treatment, especially if they're related to both treatment and outcome

You don't have to actually observe or measure them all

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

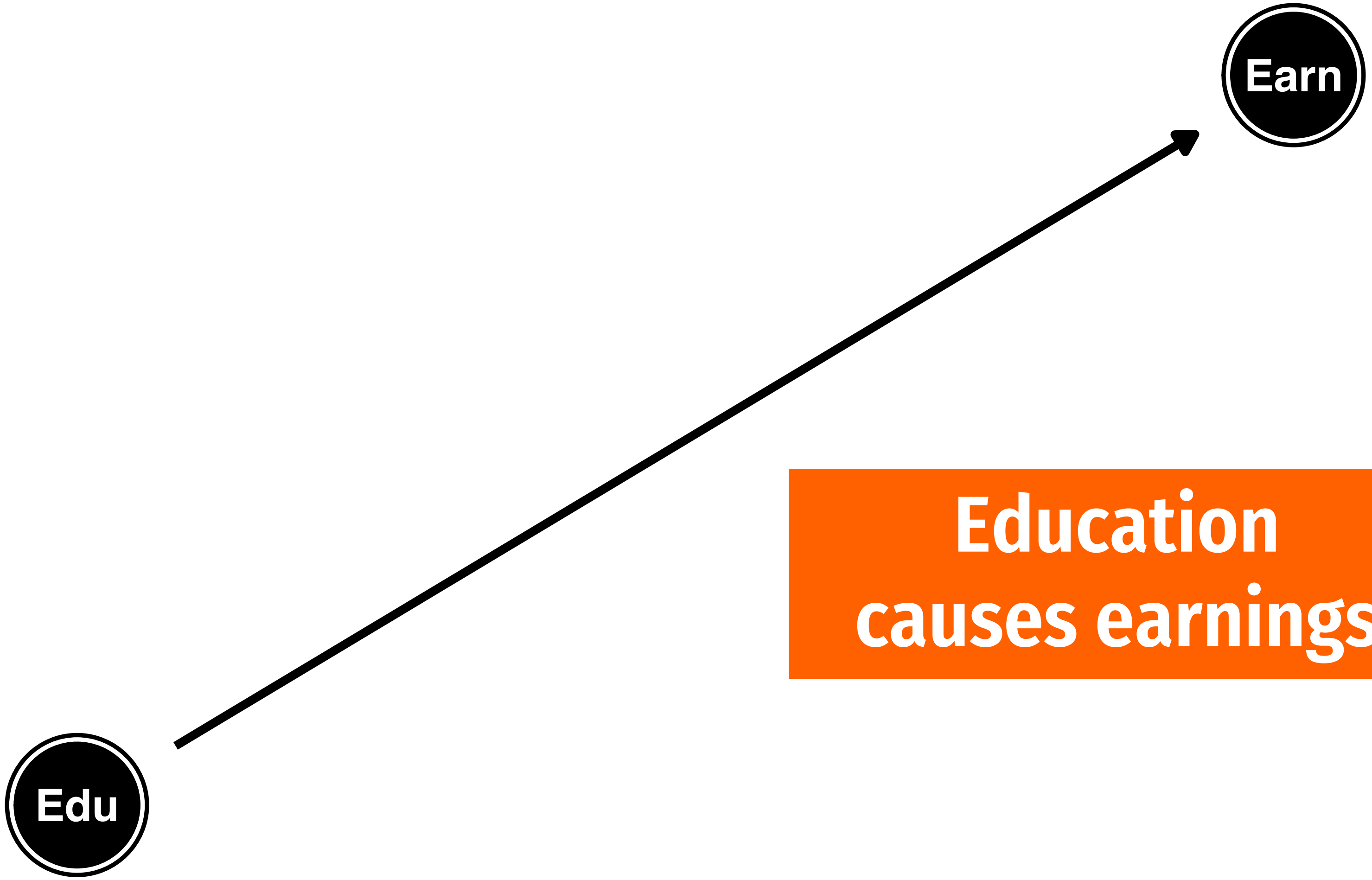
Year of birth

Compulsory schooling laws

Job connections

Background

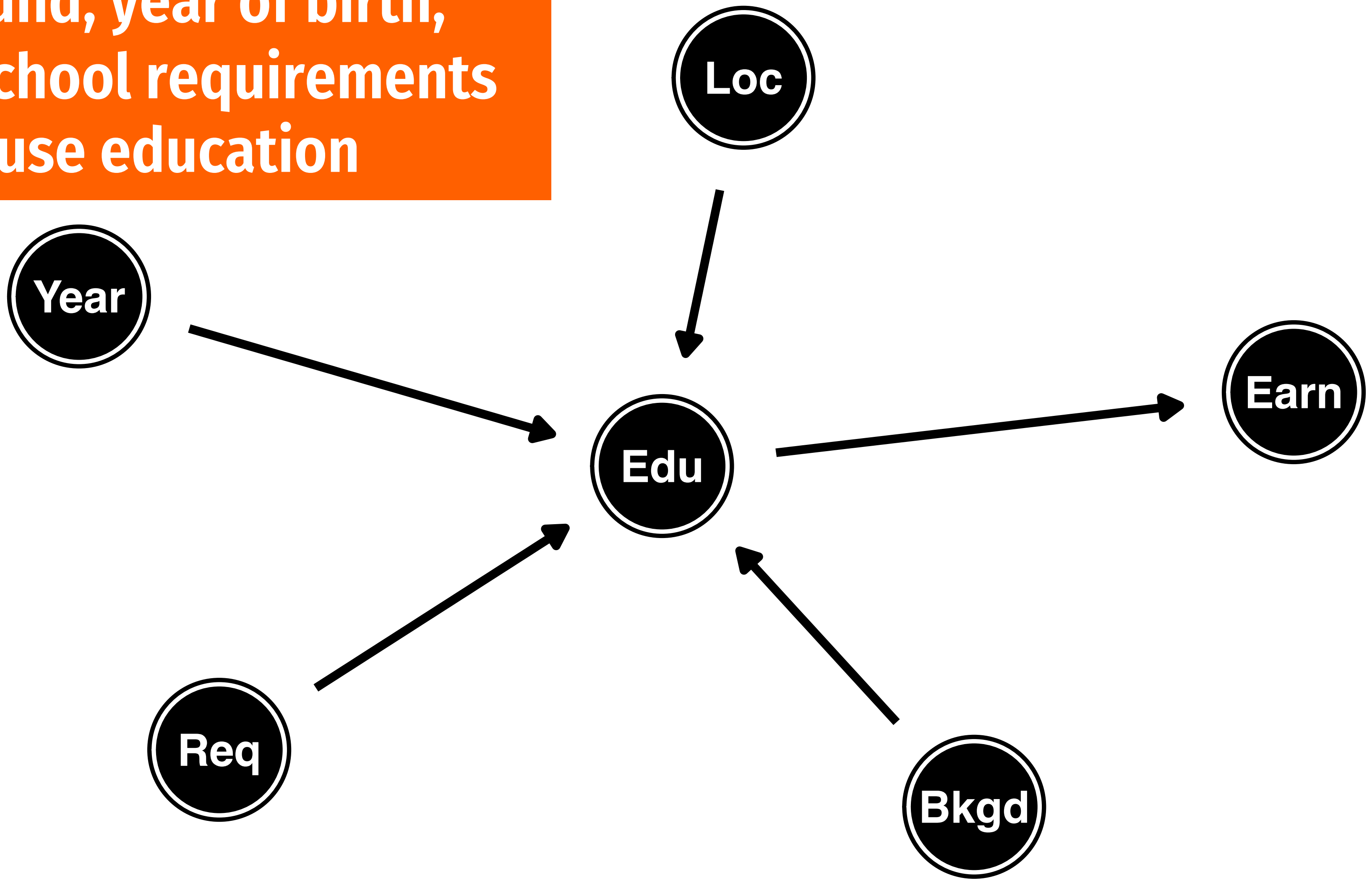
3. Draw arrows



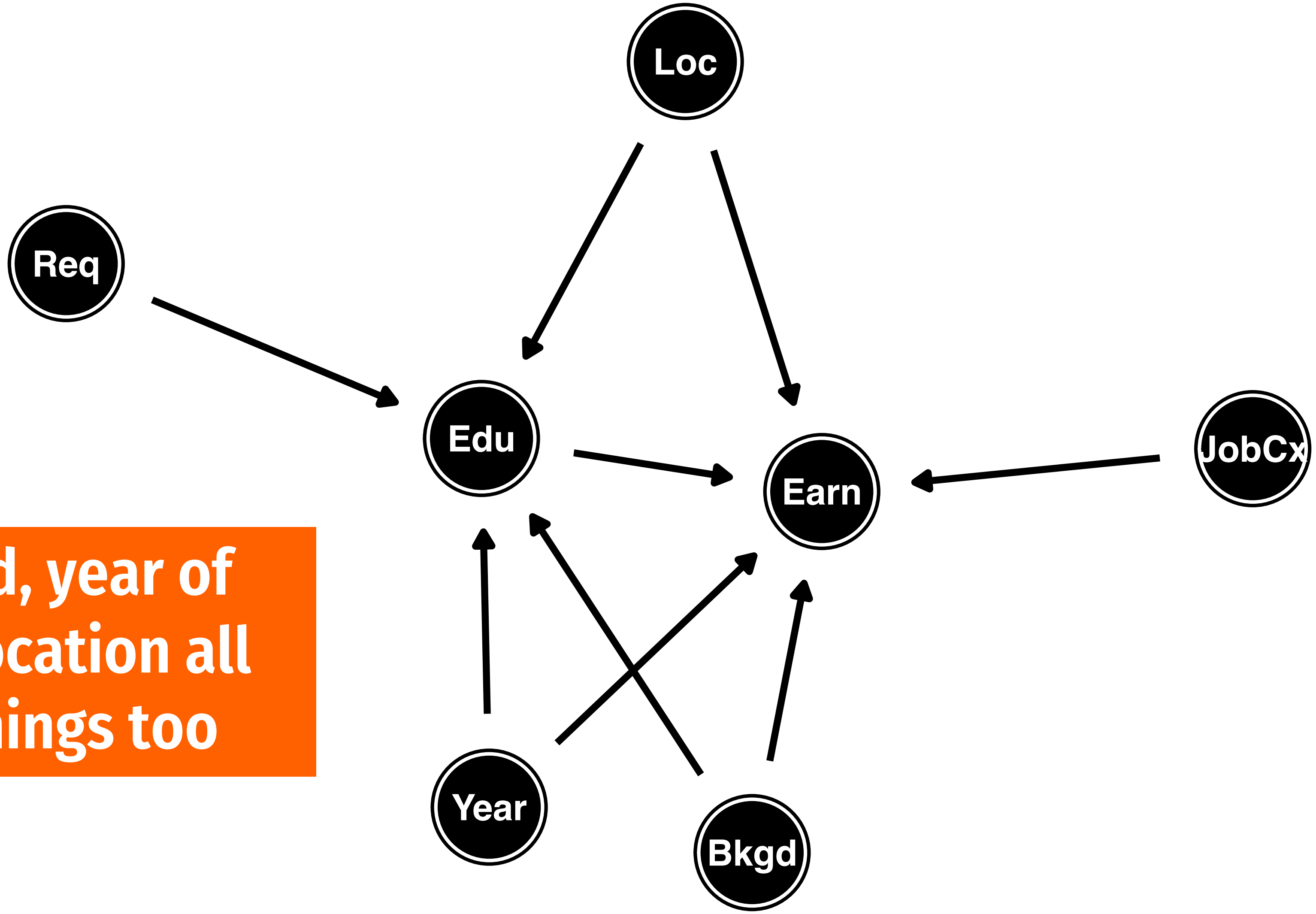
Education causes earnings

3. Draw arrows

Background, year of birth, location, school requirements all cause education

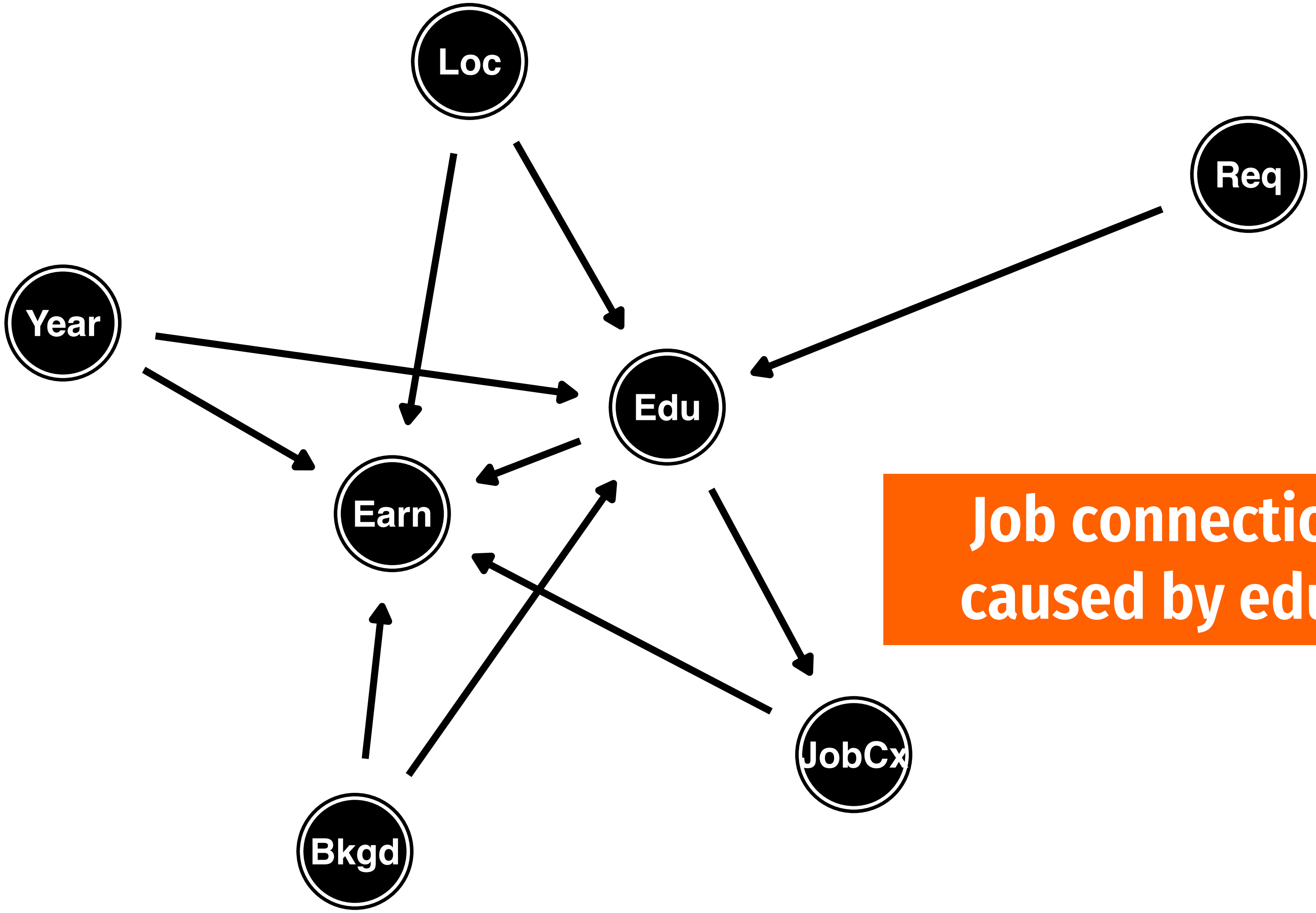


3. Draw arrows



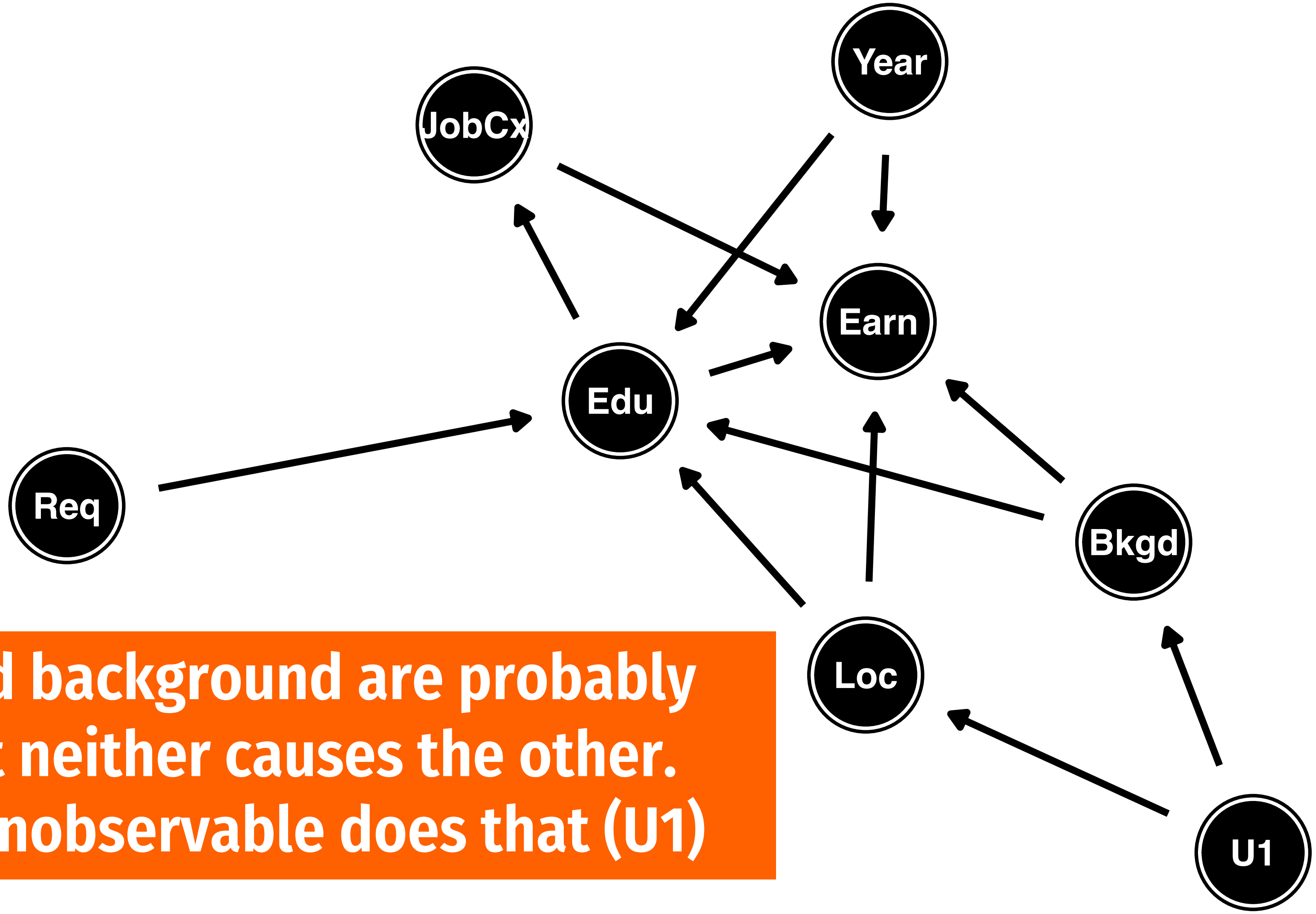
Background, year of birth, and location all effect earnings too

3. Draw arrows



Job connections are caused by education

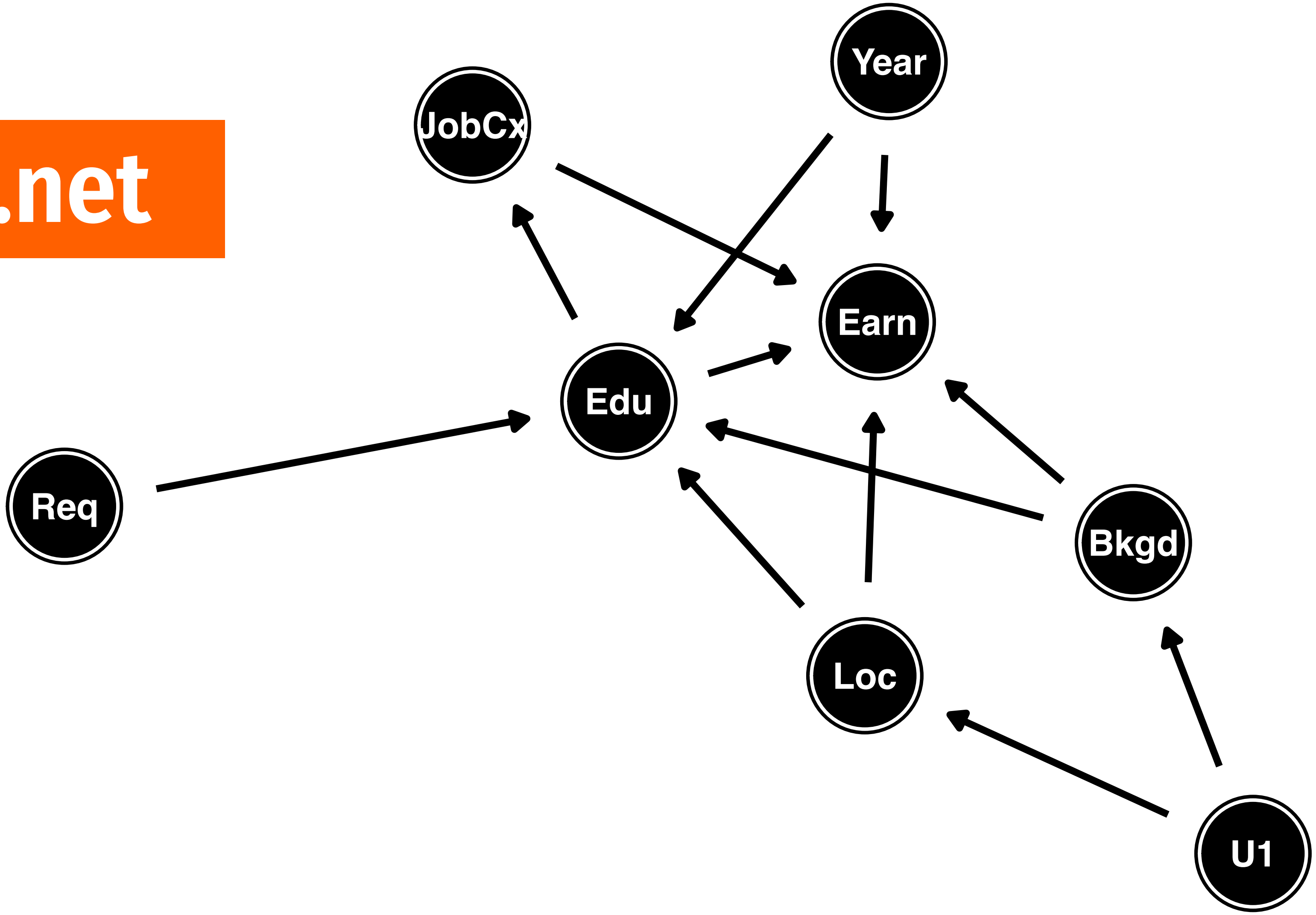
3. Draw arrows



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

Let the computer do this!

dagitty.net



Your turn

Does a longer night's sleep extend your lifespan?

Step 1: List variables

Step 2: Simplify

Step 3: Connect arrows

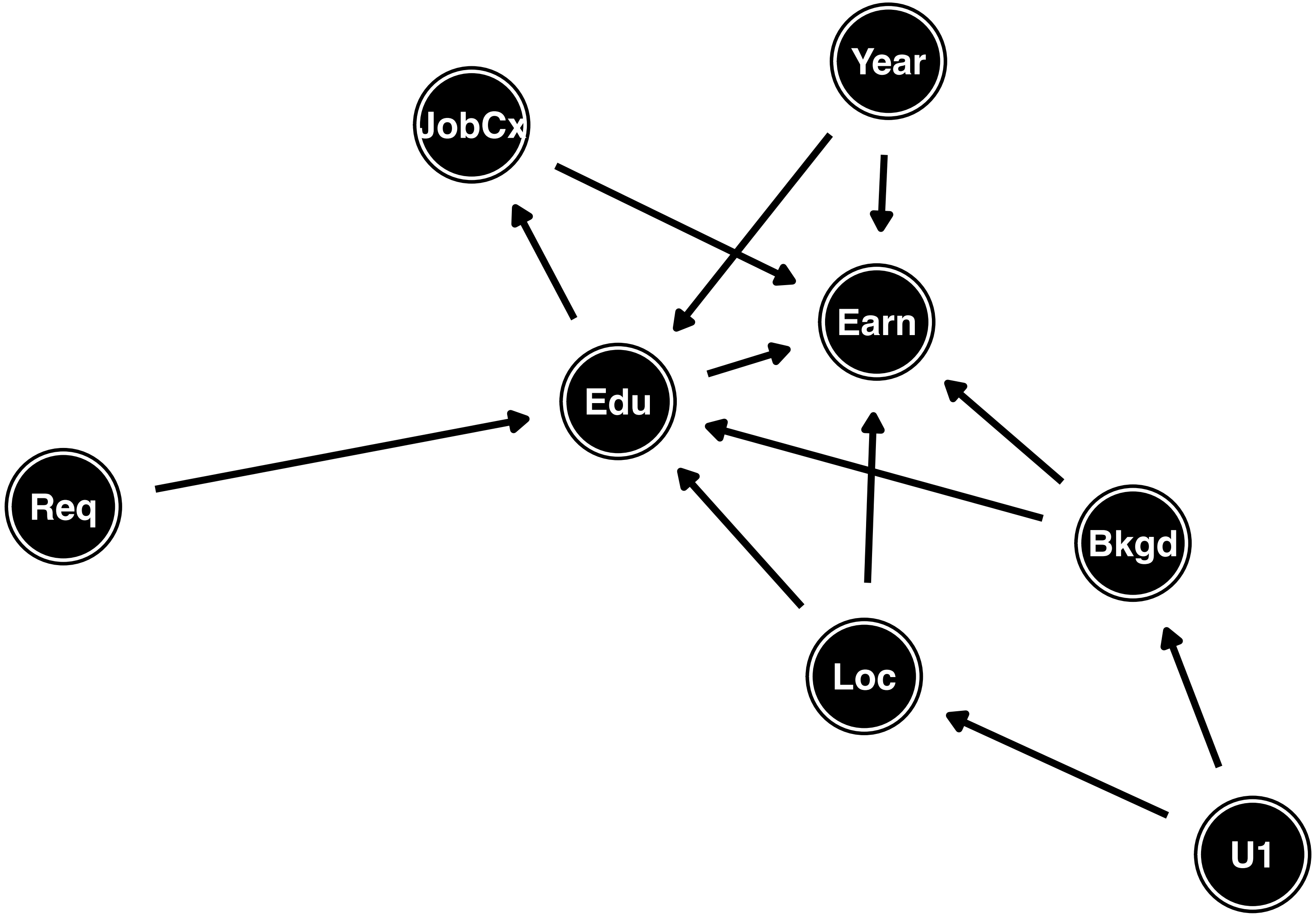
Use dagitty.net

Equations, 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 with all the other nodes?



Causal identification

A causal effect is "identified" if the association between treatment and outcome is properly stripped and isolated

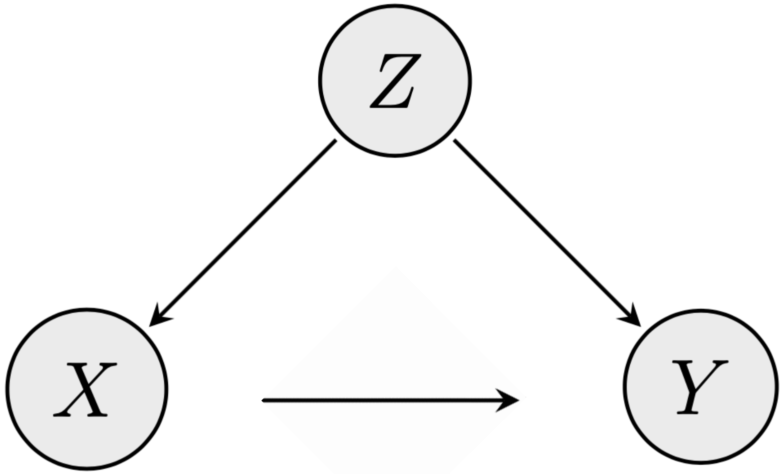
Paths and associations

Arrows in a DAG transmit associations

You can redirect and control those paths by “adjusting” or “conditioning”

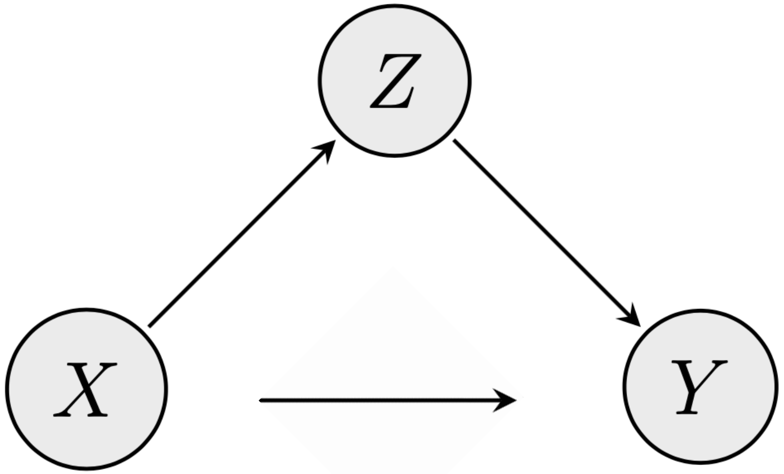
Three types of associations

Confounding



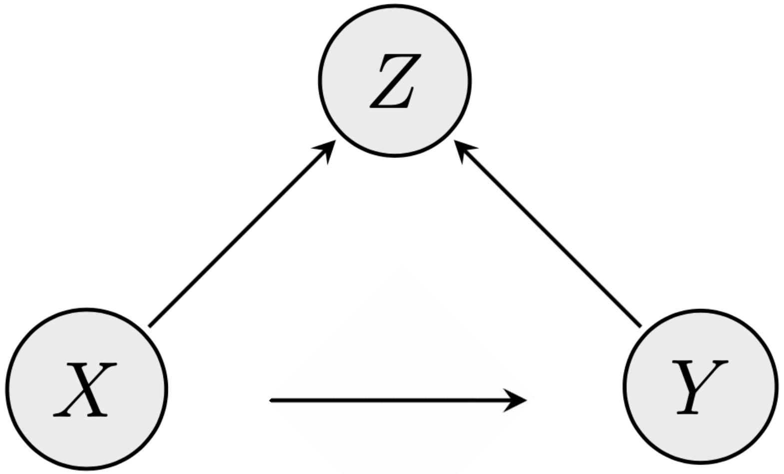
Common cause

Causation



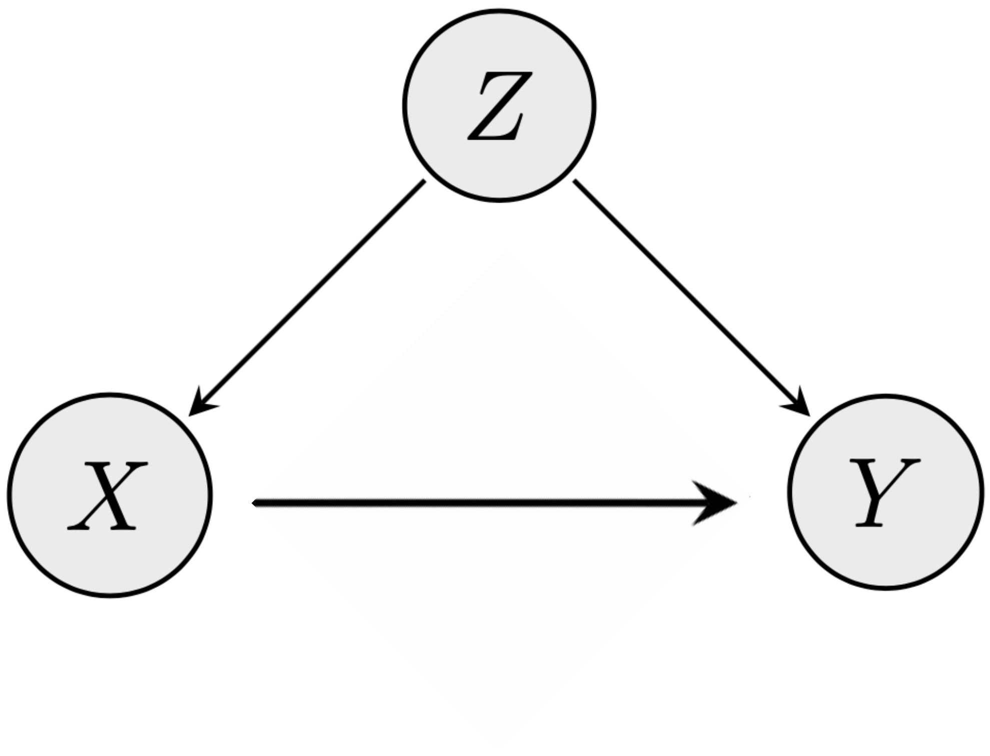
Mediation

Collision



Selection / Endogeneity

Confounding

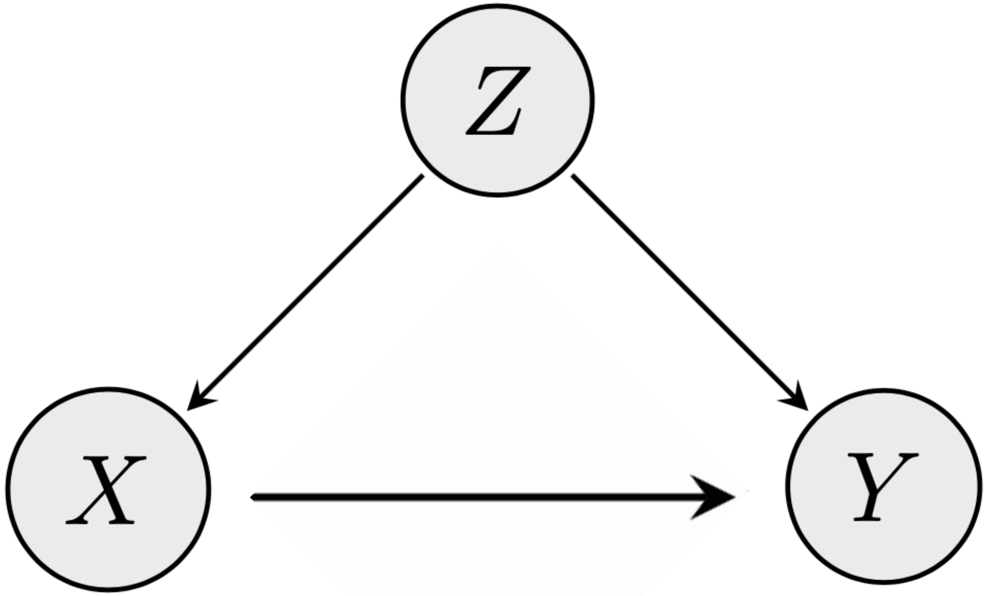


X causes Y

But Z causes both X and Y

Z confounds X → Y association

Paths

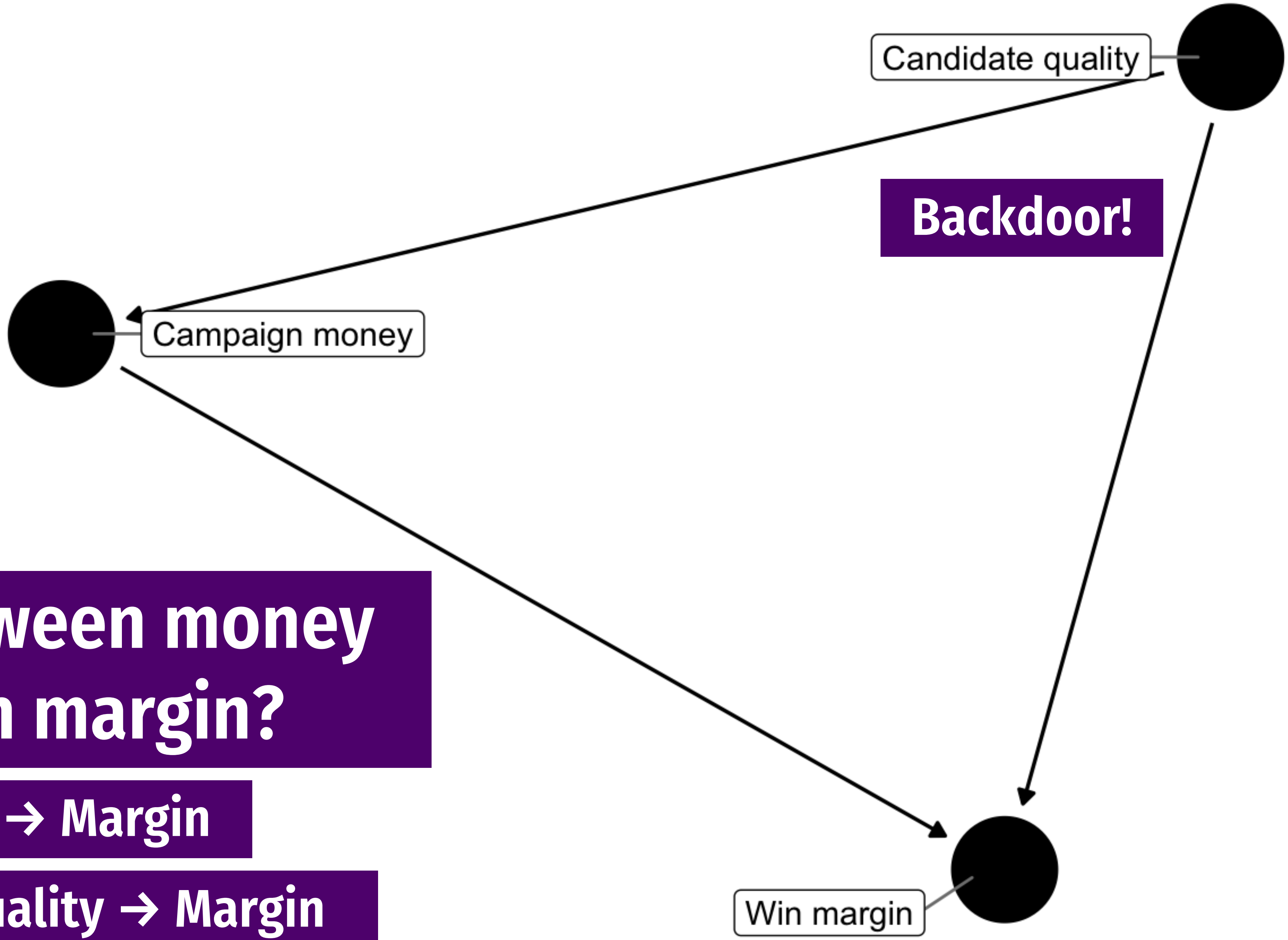


Paths between X and Y?

X → Y

X ← Z → Y

Z is a backdoor

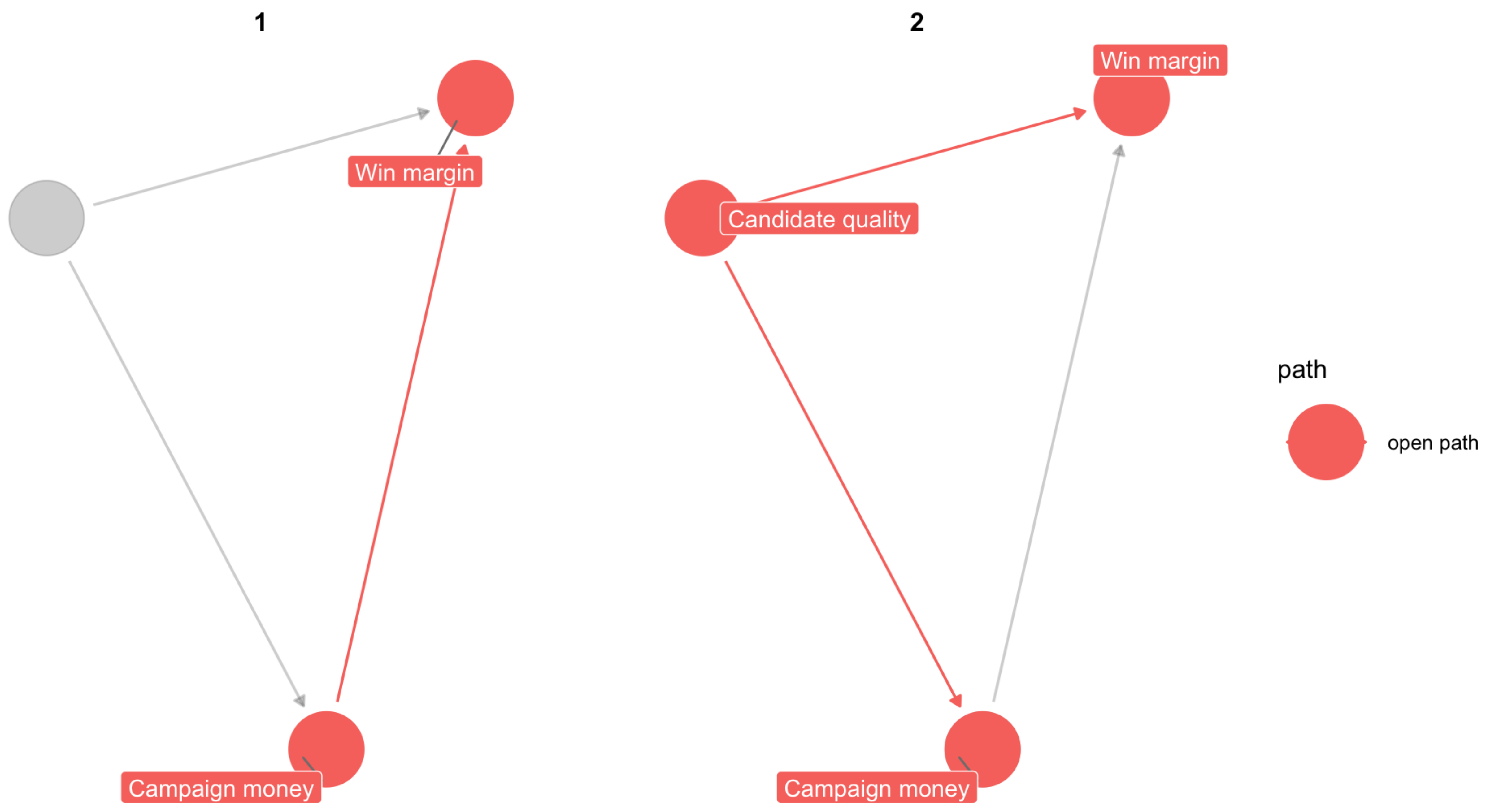


Backdoor!

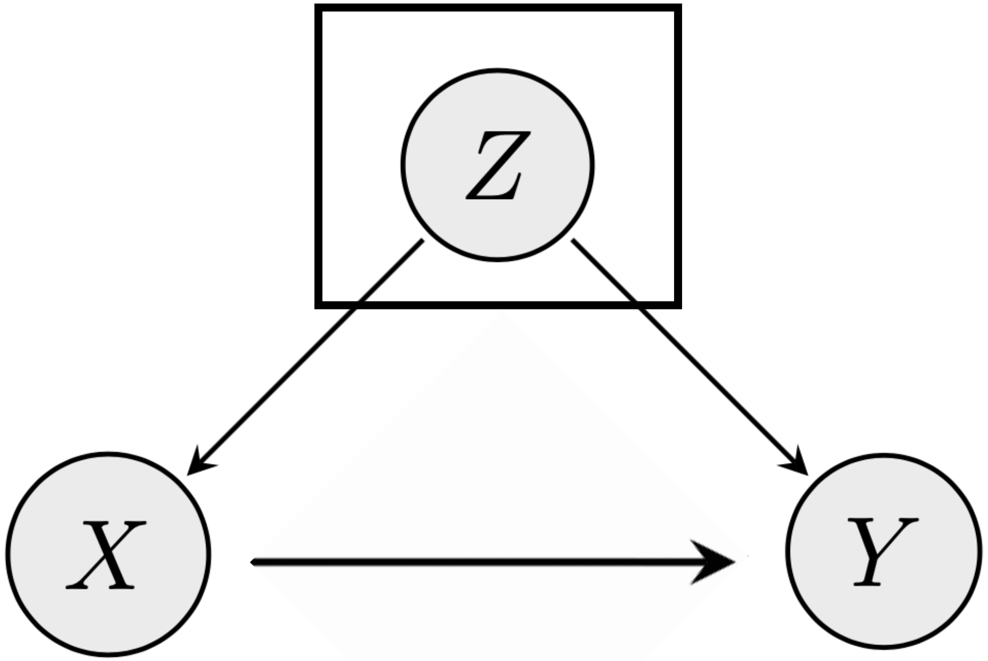
Paths between money and win margin?

Money → Margin

Money ← Quality → Margin


1
Win margin
Campaign money
2
Win margin
Candidate quality
Campaign money
path
open path

Closing doors

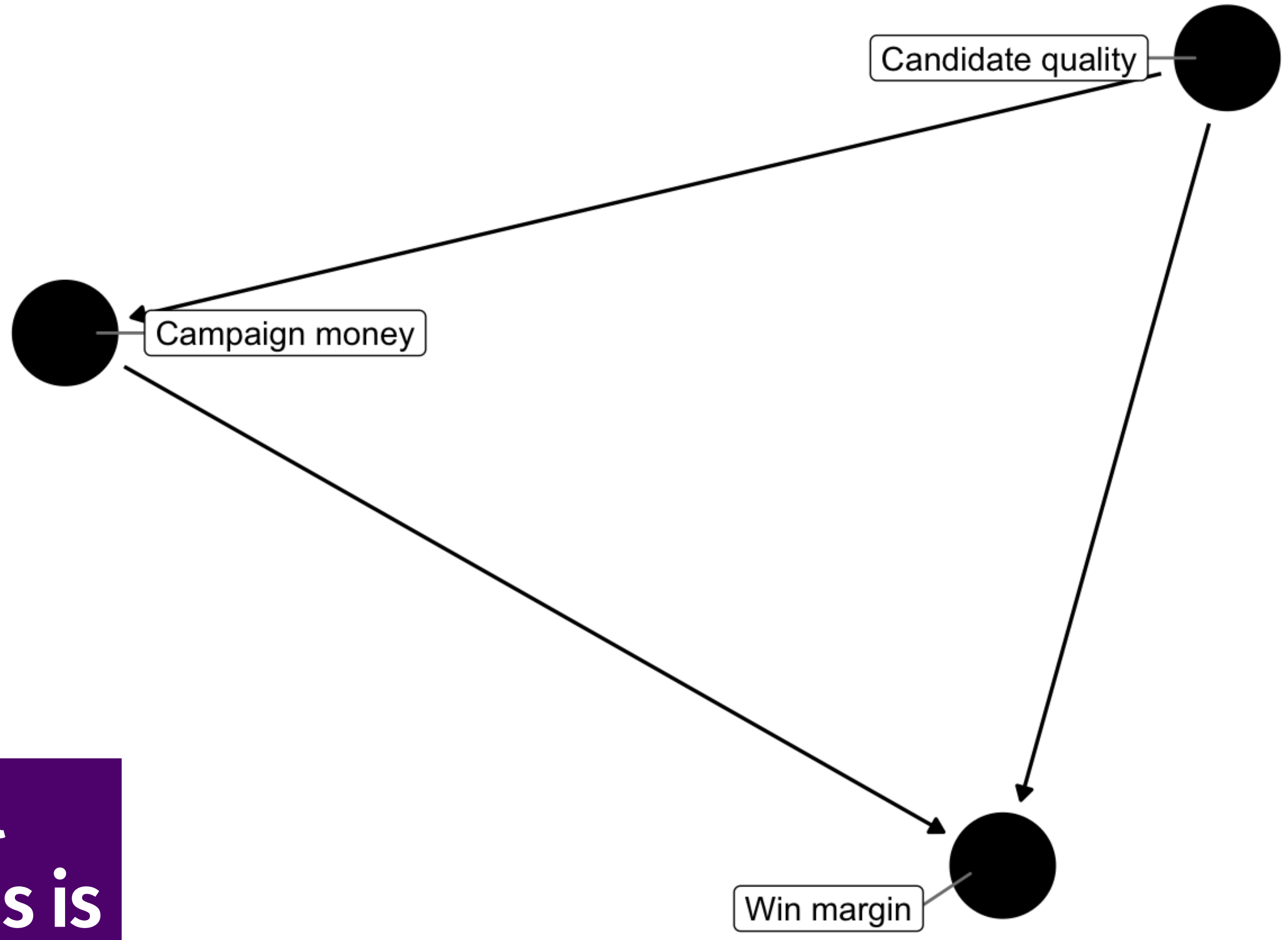


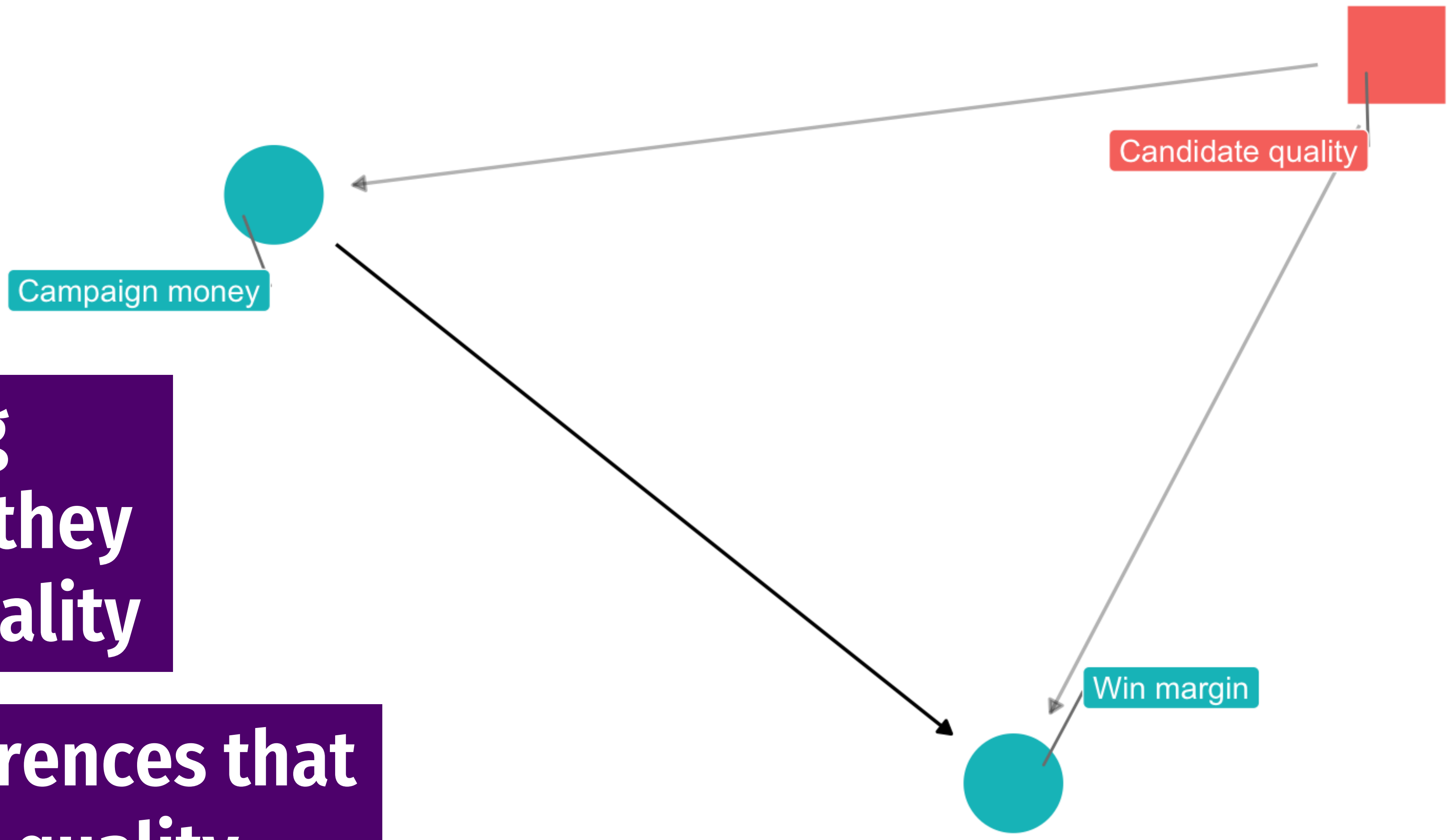
Close the backdoor by adjusting for Z

Find what part of X (campaign money) is explained by Q (quality), subtract it out. This creates the residual part of X.

Find what part of Y (the win margin) is explained by Q (quality), subtract it out. This creates the residual part of Y.

Find relationship between residual part of X and residual part of Y. This is the causal effect.





We're comparing candidates as if they had the same quality

We remove differences that are predicted by quality

Holding quality constant

How to adjust?

Include term in regression

$$\text{Win margin} = \beta_0 + \beta_1 \text{Campaign money} + \beta_2 \text{Candidate quality} + \epsilon$$

$$\text{Win margin} = \alpha + \beta \text{ Campaign money} + \gamma \text{ Candidate quality} + \epsilon$$

Matching **Do-calculus** **Inverse probability weighting**