Causality

Francisco Villamil

Research Design for Social Sciences
MA Computational Social Science, UC3M
Fall 2023

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

About prediction

- What is prediction?

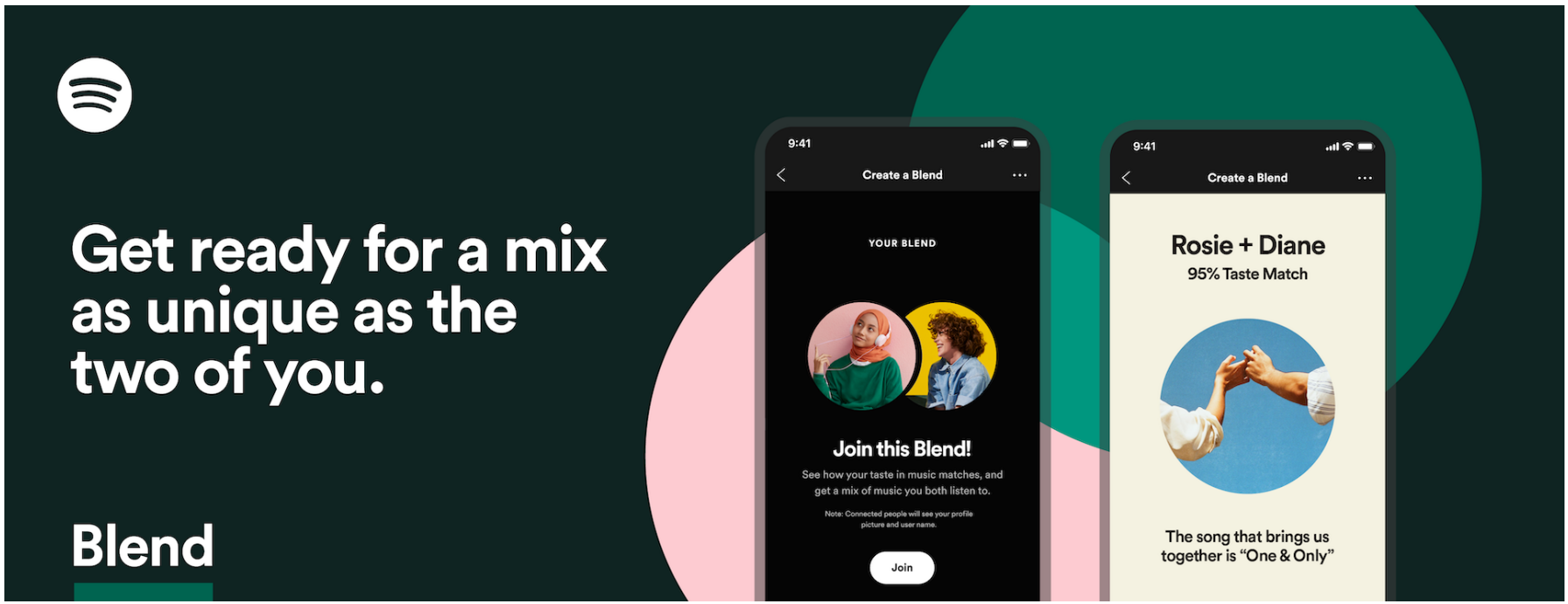
About prediction

- What is prediction?

The two concepts of prediction:

- Predicting another variable
- Predicting the future (or out of sample prediction)

About prediction



About prediction

TECH

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Kashmir Hill Former Staff

Welcome to The Not-So Private Parts where technology & privacy collide

Follow

Feb 16, 2012, 11:02am EST

About prediction

ECONOMICS

Predicting poverty and wealth from mobile phone metadata

Joshua Blumenstock,[1*] Gabriel Cadamuro,[2] Robert On[3]

Accurate and timely estimates of population characteristics are a critical input to social and economic research and policy. In industrialized economies, novel sources of data are enabling new approaches to demographic profiling, but in developing countries, fewer sources of big data exist. We show that an individual's past history of mobile phone use can be used to infer his or her socioeconomic status. Furthermore, we demonstrate that the predicted attributes of millions of individuals can, in turn, accurately reconstruct the distribution of wealth of an entire nation or to infer the asset distribution of microregions composed of just a few households. In resource-constrained environments where censuses and household surveys are rare, this approach creates an option for gathering localized and timely information at a fraction of the cost of traditional methods.

About prediction

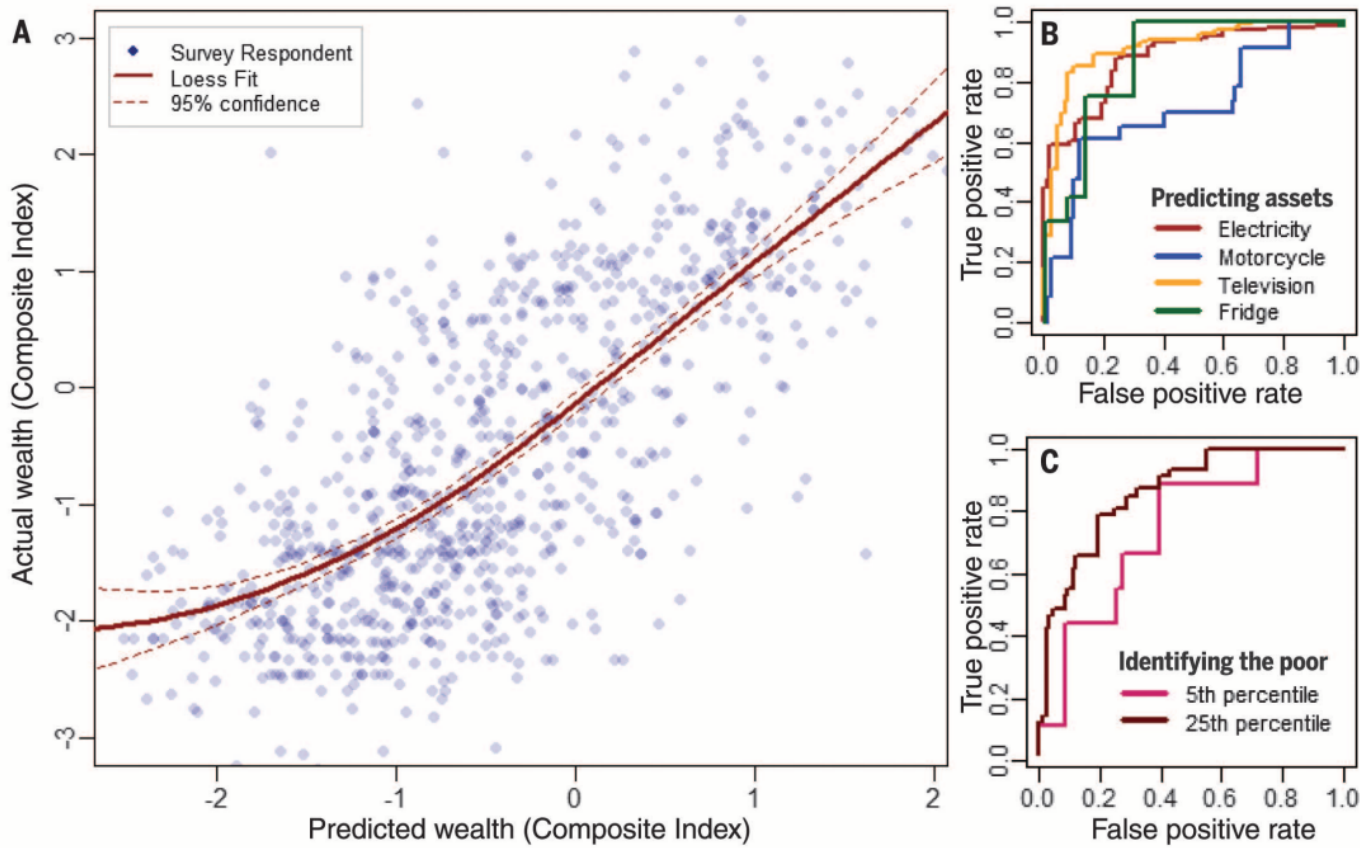


Fig. 1. Predicting survey responses with phone data. (**A**) Relation between actual wealth (as reported in a phone survey) and predicted wealth (as inferred from mobile phone data) for each of the 856 survey respondents. (**B**) Receiver operating characteristic (ROC) curve showing the model's ability to predict whether the respondent owns several different assets. AUC values for electricity, motorcycle, television, and fridge, respectively, are as follows: 0.85, 0.67, 0.84, and 0.88. (**C**) ROC curve illustrates the model's ability to correctly identify the poorest individuals. The poor are defined as those in the 5th percentile (AUC = 0.72) and the 25th percentile (AUC = 0.81) of the composite wealth index distribution.

About prediction

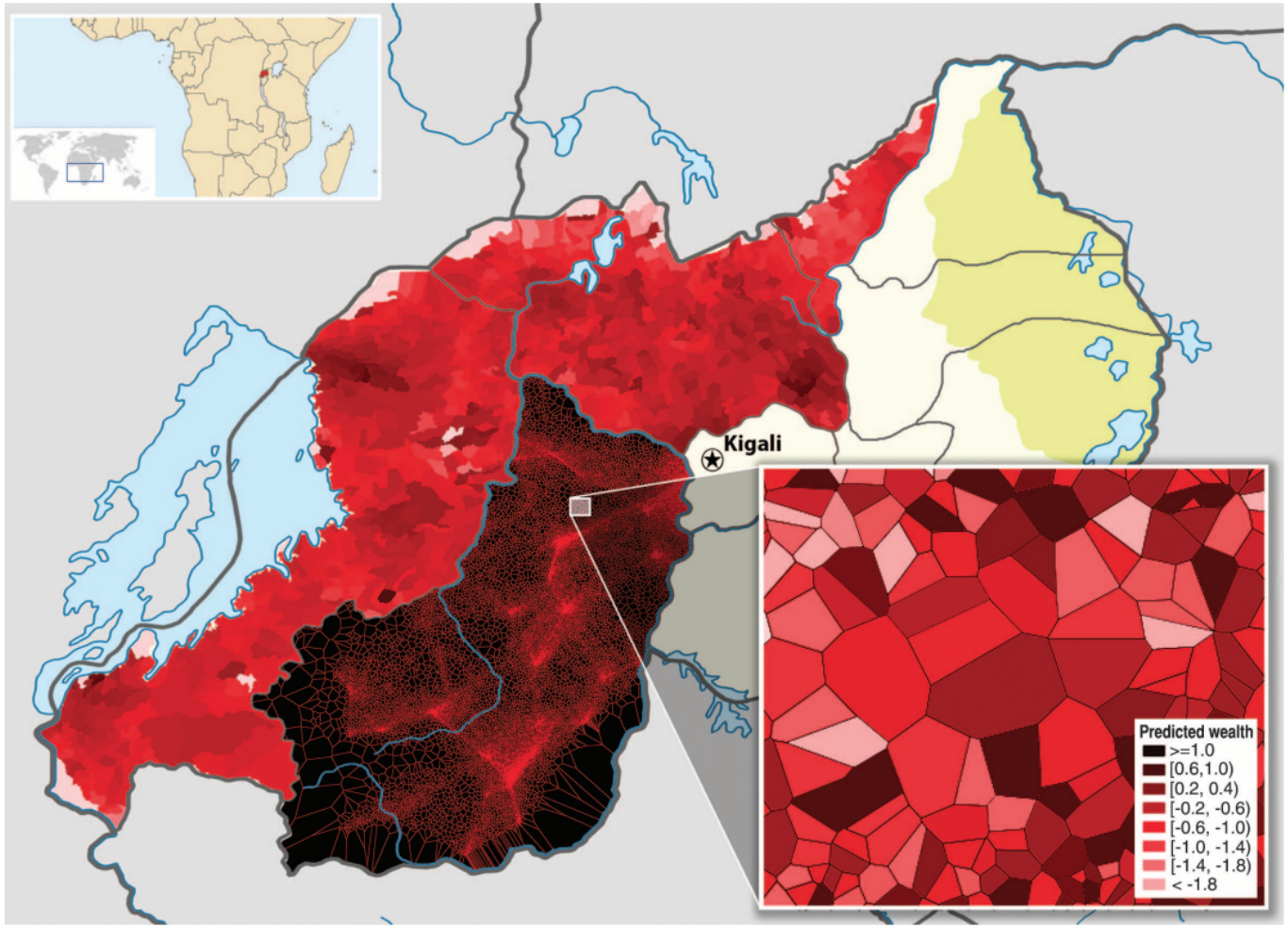


Fig. 2. Construction of high-resolution maps of poverty and wealth from call records. Information derived from the call records of 1.5 million subscribers is overlaid on a map of Rwanda. The northern and western provinces are divided into cells (the smallest administrative unit of the country), and the cell is shaded according to the average (predicted) wealth of all mobile subscribers in that cell. The southern province is overlaid with a Voronoi division that uses geographic identifiers in the call data to segment the region into several hundred thousand small partitions. (**Bottom right inset**) Enlargement of a 1-km^2 region near Kiyonza, with Voronoi cells shaded by the predicted wealth of small groups (5 to 15 subscribers) who live in each region.

About prediction

- Causality and prediction, is it the same?

About prediction

- Causality and prediction, is it the same?
- The importance of **counterfactuals** (more on this later)

Background: explanatory questions and data

- When we are dealing with *explanation*, we want to use data to get closer to the *data generating process*
- This is the causal process that generates the outcomes that we are measuring (data)

Background: explanatory questions and data

- When we are dealing with *explanation*, we want to use data to get closer to the *data generating process*
- This is the causal process that generates the outcomes that we are measuring (data)
- Example:
  - → What is the process generating the data that Spotify receives about your music tastes (i.e. song choice)?

Background: explanatory questions and data

- When we are dealing with *explanation*, we want to use data to get closer to the *data generating process*
- This is the causal process that generates the outcomes that we are measuring (data)
- Example:
  - → What is the process generating the data that Spotify receives about your music tastes (i.e. song choice)?
  - → When we try to predict it the way Spotify does, are we using data to uncover the data generating process?

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Explaining relationships

- Key thing: we want to know whether X actually *causes* Y

Explaining relationships

- Key thing: we want to know whether X actually *causes* Y
  - $\rightarrow$ I.e., we want to do *causal inference*

Explaining relationships

- Key thing: we want to know whether X actually *causes* Y
  - $\rightarrow$ I.e., we want to do *causal inference*
- (**Note** that this does not mean that X is the only cause of Y, but that **changing** X **alters** Y)

Explaining relationships

- How could we observe causal relationships? Repeating history

Explaining relationships

- How could we observe causal relationships? Repeating history
- The 'fundamental problem of causal inference' is that we cannot

Explaining relationships

- How could we observe causal relationships? Repeating history
- The 'fundamental problem of causal inference' is that we cannot
  - $\rightarrow$ In other words, that for every unit of observation, we can only observe **either** $Y(X = 0)$ **or** $Y(X = 1)$

Explaining relationships

- How could we observe causal relationships? Repeating history
- The 'fundamental problem of causal inference' is that we cannot
  - → In other words, that for every unit of observation, we can only observe **either** $Y(X = 0)$ **or** $Y(X = 1)$
- If we observe $Y(X = 1)$, causal inference essentially means trying to find as good an approximation to $Y(X = 0)$ as we can find

Explaining relationships

- How could we observe causal relationships? Repeating history
- The 'fundamental problem of causal inference' is that we cannot
  - → In other words, that for every unit of observation, we can only observe **either** $Y(X = 0)$ **or** $Y(X = 1)$
- If we observe $Y(X = 1)$, causal inference essentially means trying to find as good an approximation to $Y(X = 0)$ as we can find
  - → i.e., we want to find something that is valid as a *counterfactual*

Potential outcomes framework

- Also called Neyman–Rubin causal model

Potential outcomes framework

- Also called Neyman–Rubin causal model
- An effect is the difference between the *actual world* and an *alternative reality* (counterfactual)

Potential outcomes framework

- Also called Neyman–Rubin causal model
- An effect is the difference between the *actual world* and an *alternative reality* (counterfactual)
  - $\rightarrow$ Causal effect of X, is $E(Y|X = 1)$ - $E(Y|X = 0)$

Potential outcomes framework

What is the effect of smoking on life expectancy?

Potential outcomes framework

- Let's take Gary, a man who smokes, doesn't exercise, but is vegetarian. We can wait and see how long he lives:

$$E(LExp|S = 1, G = Male, E = 0, V = 1)$$

Gary

Potential outcomes framework

- Let's take Gary, a man who smokes, doesn't exercise, but is vegetarian. We can wait and see how long he lives:

  $E(LExp|S = 1, G = Male, E = 0, V = 1)$

- The Q is, **what is the causal effect of smoking** on Gary?

Gary

Potential outcomes framework

- Let's take Gary, a man who smokes, doesn't exercise, but is vegetarian. We can wait and see how long he lives:

  $E(LExp|S = 1, G = Male, E = 0, V = 1)$

- The Q is, **what is the causal effect of smoking** on Gary?
- To know that, we need to estimate the life expectancy of an alternative Gary that is *exactly* the same except for the smoking:

  $E(LExp|S = 0, G = Male, E = 0, V = 1)$

Gary

Potential outcomes framework

- Let's take Gary, a man who smokes, doesn't exercise, but is vegetarian. We can wait and see how long he lives:

  $E(LExp|S = 1, G = Male, E = 0, V = 1)$

- The Q is, **what is the causal effect of smoking** on Gary?
- To know that, we need to estimate the life expectancy of an alternative Gary that is *exactly* the same except for the smoking:

  $E(LExp|S = 0, G = Male, E = 0, V = 1)$

- The problem is that the alternative Gary is unobservable: we have a **missing data problem**

Gary

Potential outcomes framework

- That would be for Gary. What about the 'general' effect of smoking?

Potential outcomes framework

- That would be for Gary. What about the 'general' effect of smoking?
- Well, we need to find an alternative for every person (smoking and non-smoking), and just calculate the difference between the alternative and the reality:

  $E[LifeExp_i^1] - E[LifeExp_i^0]$

  which would be the **Average Treatment Effect** (or **ATE**)

Potential outcomes framework

- That would be for Gary. What about the 'general' effect of smoking?
- Well, we need to find an alternative for every person (smoking and non-smoking), and just calculate the difference between the alternative and the reality:

  $E[LifeExp_i^1] - E[LifeExp_i^0]$

  which would be the **Average Treatment Effect** (or **ATE**)
- Problem is we have missing data: we don't have $E[LifeExp_i^1]$ for non-smokers, and we don't have $E[LifeExp_i^0]$ for smokers

Potential outcomes framework

- Same goes with other quantities of interest we'll see:

Potential outcomes framework

- Same goes with other quantities of interest we'll see:
- **ATT**, or **average treatment effect on the treated**:

  $E[Y_i^1|D_i = 1] - E[Y_i^0|D_i = 1]$

  (effect of smoking among smokers)

Potential outcomes framework

- Same goes with other quantities of interest we'll see:
- **ATT**, or **average treatment effect on the treated**:

  $E[Y_i^1|D_i = 1] - E[Y_i^0|D_i = 1]$

  (effect of smoking among smokers)
- Or the **ATC** (or ATU), or **average treatment effect on the untreated**:

  $E[Y_i^1|D_i = 0] - E[Y_i^0|D_i = 0]$

  (effect of smoking among non-smokers)

Potential outcomes framework

- Same goes with other quantities of interest we'll see:
- **ATT**, or **average treatment effect on the treated**:

  $E[Y_i^1|D_i = 1] - E[Y_i^0|D_i = 1]$

  (effect of smoking among smokers)
- Or the **ATC** (or ATU), or **average treatment effect on the untreated**:

  $E[Y_i^1|D_i = 0] - E[Y_i^0|D_i = 0]$

  (effect of smoking among non-smokers)

- When is $ATT \neq ATC$?

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Estimating causal effects

- **So how do we solve this missing data problem?**

Estimating causal effects

- **So how do we solve this missing data problem?**
- 'No causation without manipulation'

Estimating causal effects

- **So how do we solve this missing data problem?**

- 'No causation without manipulation'
- Model initially developed for *experimental data*: randomized controlled trials are the gold-standard in approximating the alternative reality (counterfactual)

Issues in experimental designs

Experiments also have their problems, e.g.:

- Non-perfect randomization (esp. if block-r)
- SUTVA (or stable unit treatment value assumption)
- Attrition
- Treatment compliance (estimating the intent-to-treat, or ITT)
- External validity
- etc

Randomization issues

- Obviously, the basic of any experiment is that **treatment assignment is random**
- It's not frequent, but could happen that this randomization is not well done
- Also it might not let us detect the effect, and having statistical issues, especially when using *block* randomization, or *unit* vs. *cluster* randomization

- Also could be an issue when doing *block randomization*

SUTVA

- SUTVA stands for **Stable Unit Treatment Value Assumption**, and it is a key assumption in experimental designs
  - $\rightarrow$ It is basically that the outcome in one unit is **not** affected by treatment assignment in other units
- Diffusion effects among subjects?
- (This problem is also discussed in causal inference with observational data)

Attrition

- **Attrition** is just the case when 'participants leave the study'
- More generally, when some of the units in the experiment do not complete it
- The key question is, to what extent is this biasing the results?

Treatment compliance

- Are all units assigned to treatment really exposed to it?
- In clinical trials, e.g. do they take the pill or spit it?
- How would this look like in an experiment when you pay (treated) individuals to watch TV or use Facebook?
- Concept of **intention-to-treat** (ITT) analyses and the **complier average causal effect** or **local average treatment effect** (LATE)

External validity

- To what extend can we **generalize the results of an experiment**?
- This is a more general issue that we will also discuss with observational-data studies, but perhaps very relevant for experiments because of the setting it usually takes place
- Example: media exposure studies
  - → Treatment validity?
  - → Outcome validity of survey hypothetical questions? (behavioral outcomes)

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

So how to we approximate Y^0?

- Experiments are fine, but often not possible

So how to we approximate Y^0?

- Experiments are fine, but often not possible
- How do we do this with **observational data**?,
  - $\rightarrow$ We need to 'build' a counterfactual

So how to we approximate Y^0?

- Experiments are fine, but often not possible
- How do we do this with **observational data**?,
  - $\rightarrow$ We need to 'build' a counterfactual
- Basic idea: we come up with a strategy where the only variation we analyze is (according to us) *due* to the independent variable (cause) we are interested in
  - $\rightarrow$ **Read it again:** it is actually the same idea as in the experimental method, where we use randomization to achieve that

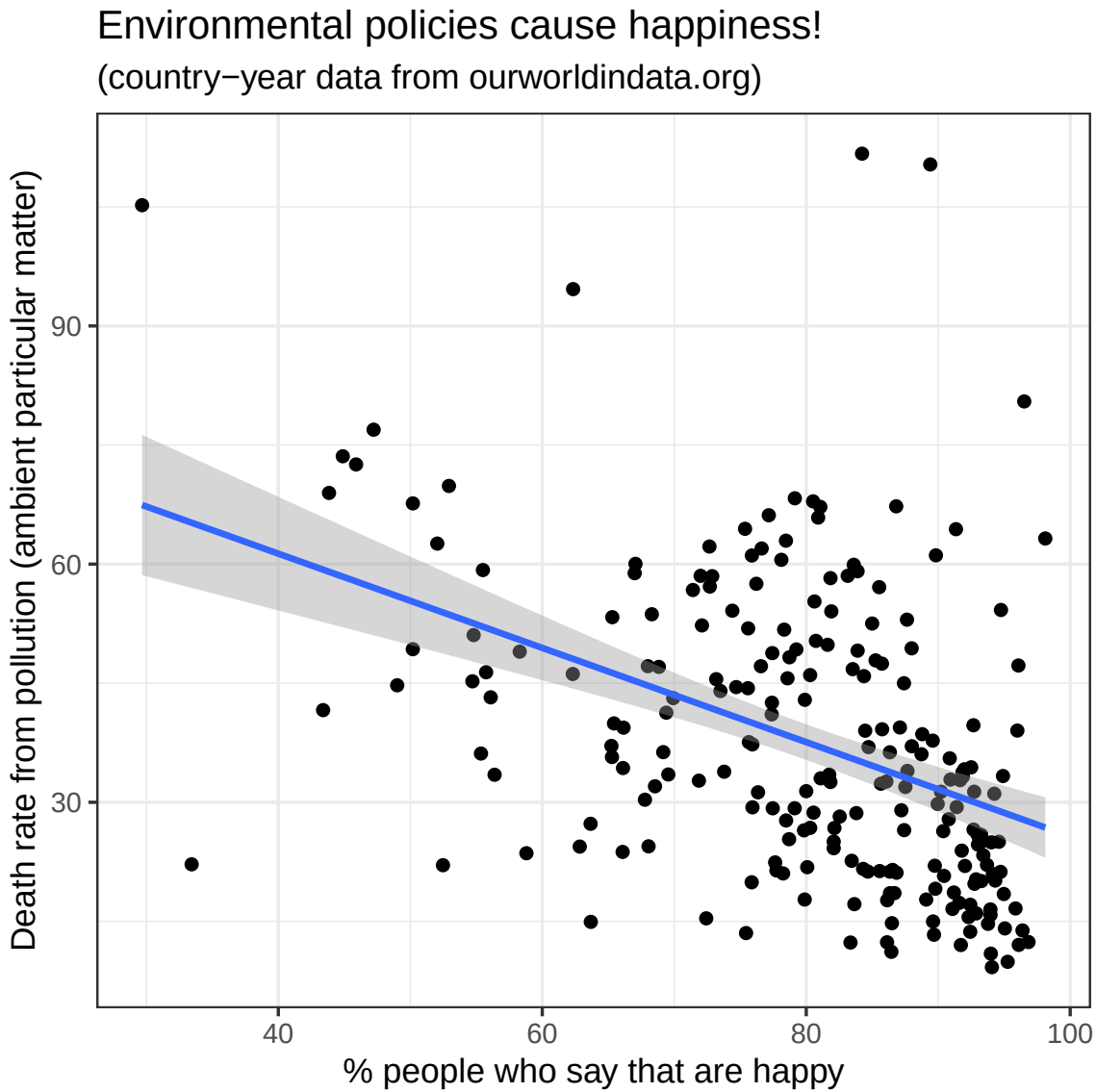
So how to we approximate Y^0?

- Experiments are fine, but often not possible
- How do we do this with **observational data**?,
  - → We need to 'build' a counterfactual
- Basic idea: we come up with a strategy where the only variation we analyze is (according to us) *due* to the independent variable (cause) we are interested in
  - → **Read it again:** it is actually the same idea as in the experimental method, where we use randomization to achieve that
- But in order to do that, we need to be clear about the **causal model** that is causing Y, so we know what we need to control for
  - → And we're gonna use causal diagrams for that

Example

- Let's say we want to know whether a cleaner environment makes people happier

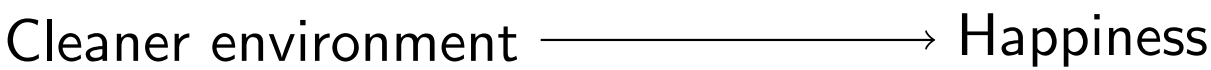
Example


Environmental policies cause happiness!
(country−year data from ourworldindata.org)
Death rate from pollution (ambient particular matter)
90
60
30
40
60
80
100
% people who say that are happy

Example

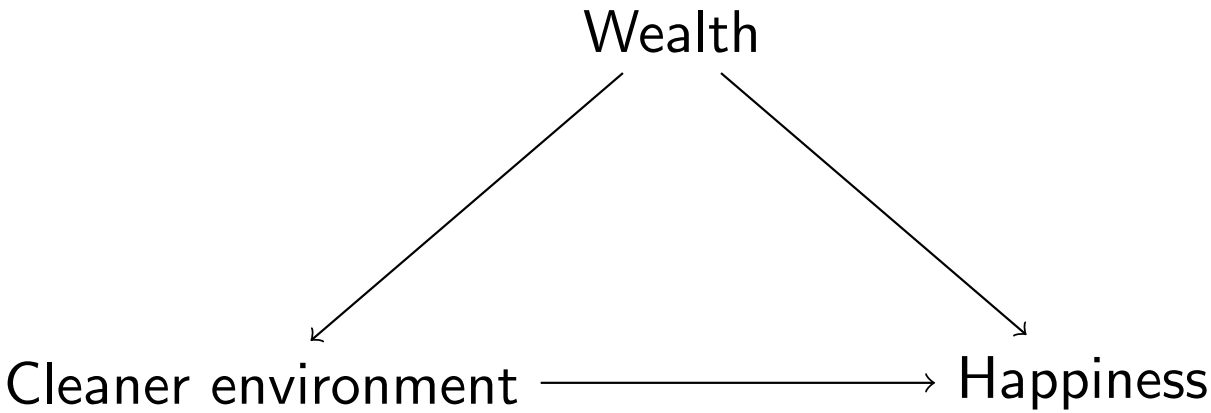
- Remember that out problem (the ‘fundamental problem of causal inference‘ etc) is that we can observe e.g. Pakistan, where the level of pollution (measured as death date) is 46, and 58% of the people say they’re happy
- But **we cannot observe** how many people say they are happy in an **alternative Pakistan** where the pollution death date is 15
- So to approximate this, we’ll build a causal model to know what we should be controlling for

Our causal model



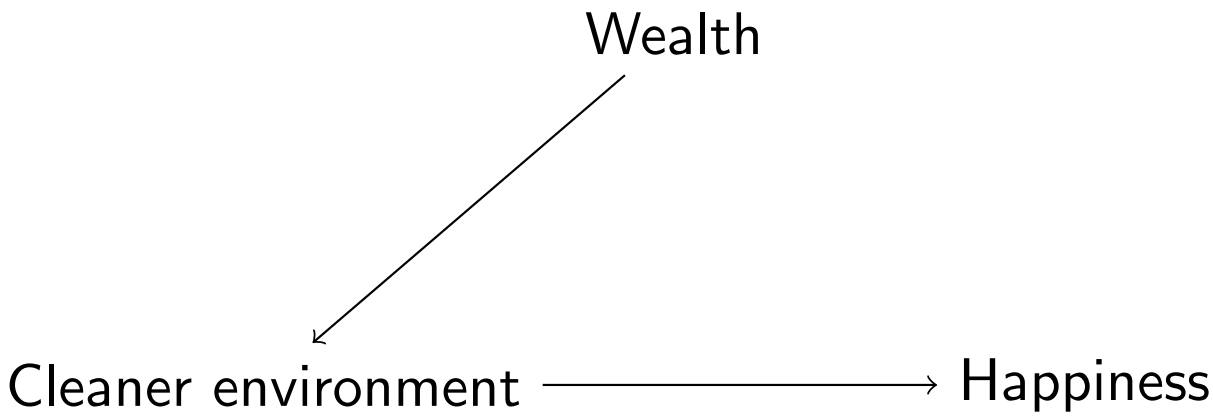
- This is our initial causal model: having a cleaner environment makes people happier (because they like looking into a blue sky without smog), and that's it. We do not have to control for anything nor do anything else.

Our causal model



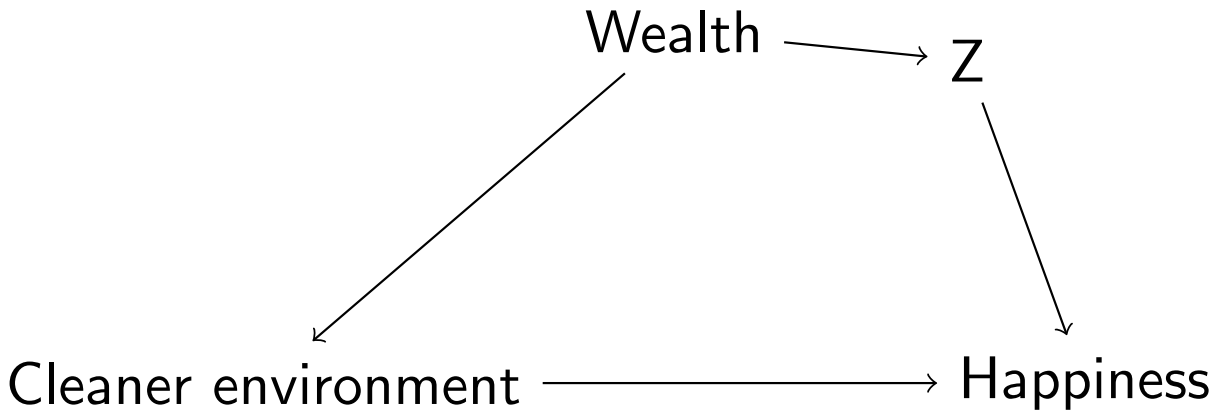
- Wait, but maybe it's about money, isn't it? Actually, wealthier countries tend to have cleaner environments and, at the same time, money causes happiness. **We need to control for wealth.**

Our causal model



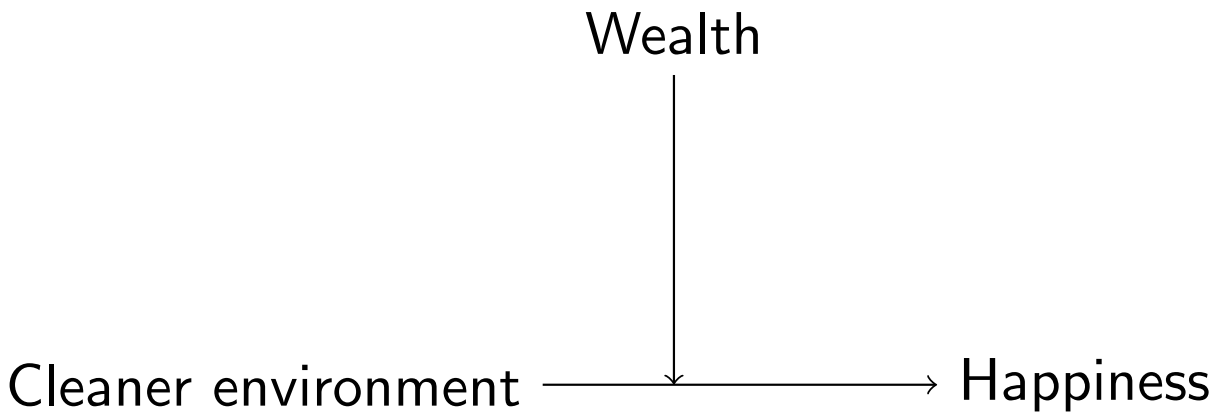
- Or perhaps is not that money increases happiness *per se*, but that it does so through other **mediators**: wealth allows countries to focus on environment, which increases happiness. **Again, no need to control.** As long as this is the **only** mediator.

Our causal model



- We are happy with that model, but we're still missing something. Say we believe that money does not have any direct causal effect, but it does causes some other things (labour conditions, cultural offer, ... let's call them Z) and these, in turn, have an effect on happiness. **We need to control for wealth and all Z.**

Our causal model



- (Another thing would be if money **moderates** the relationship between environmental policies and happiness: spending resources to take care of our environment makes you happier only if you have enough money – this is an special case, we could talk about heterogenous effects)

Basics of causal inference

- So to come up with an strategy, we need to understand what's going on in terms of the data generating process

Basics of causal inference

- So to come up with an strategy, we need to understand what's going on in terms of the data generating process
  - → This applies from the most basic strategy (add controls) to the more complicated ones (e.g. evaluating DiD or RDD)

Basics of causal inference

- So to come up with an strategy, we need to understand what's going on in terms of the data generating process
  - $\rightarrow$ This applies from the most basic strategy (add controls) to the more complicated ones (e.g. evaluating DiD or RDD)
- Once we have that, we can **identify** an effect (in other words: isolating the causal variation from other sources of variation we are not interested in)

Causal models, mechanisms, and DAGs

- We will use Directed Acyclic Graphs (DAG) (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

Causal models, mechanisms, and DAGs

- We will use Directed Acyclic Graphs (DAG) (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

  A few things:

Causal models, mechanisms, and DAGs

- We will use Directed Acyclic Graphs (DAG) (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

A few things:

- Only one-directional causality (*acyclic*)
  - → if you have feedback cycles, write multiple nodes for t_1, t_2

Causal models, mechanisms, and DAGs

- We will use Directed Acyclic Graphs (DAG) (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

A few things:

- Only one-directional causality (*acyclic*)
  - → if you have feedback cycles, write multiple nodes for t_1, t_2
- Sometimes: solid lines → observed, dashed → unobserved (U)

Causal models, mechanisms, and DAGs

- We will use Directed Acyclic Graphs (DAG) (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

A few things:

- Only one-directional causality (*acyclic*)
  - → if you have feedback cycles, write multiple nodes for t_1, t_2
- Sometimes: solid lines → observed, dashed → unobserved (U)
- Treatment usually written as D (and Y the outcome)

Causal models, mechanisms, and DAGs

- We will use Directed Acyclic Graphs (DAG) (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

A few things:

- Only one-directional causality (*acyclic*)
  - → if you have feedback cycles, write multiple nodes for t_1, t_2
- Sometimes: solid lines → observed, dashed → unobserved (U)
- Treatment usually written as D (and Y the outcome)
- Combine variables (usually B for background, or U for unknown)

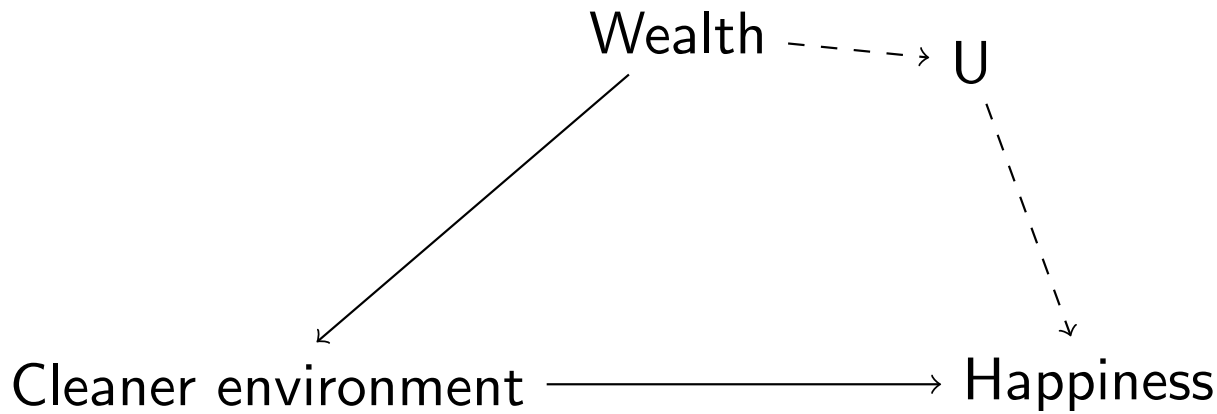
Causal models, mechanisms, and DAGs

- We will use Directed Acyclic Graphs (DAG) (causal diagrams), a graph where we link **variables** (nodes) with **causal effects** (arrows)

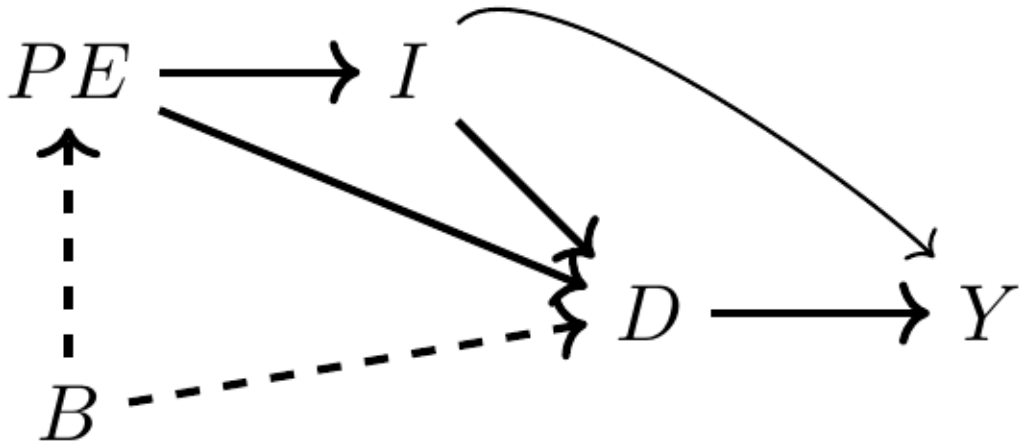
A few things:

- Only one-directional causality (*acyclic*)
  - → if you have feedback cycles, write multiple nodes for t_1, t_2
- Sometimes: solid lines → observed, dashed → unobserved (U)
- Treatment usually written as D (and Y the outcome)
- Combine variables (usually B for background, or U for unknown)
- No arrow means *no effect*, explicitly

This is a DAG



This is another DAG



- Y = earnings (outcome)
- D = college education (treatment)
- PE = parental education
- I = family income
- B = unobserved background factors (intelligence, abilities, home, etc)

from https://mixtape.scunning.com/03-directed_acyclical_graphs

Causal models, mechanisms, and DAGs

We use DAGs for mainly two things related to causal inference:

- Drawing up the **mechanism** that explains the outcome
- Come up with the strategy we need to **identify** the causal effect

Causal models, mechanisms, and DAGs

We use DAGs for mainly two things related to causal inference:

- Drawing up the **mechanism** that explains the outcome
- Come up with the strategy we need to **identify** the causal effect
  - → The difference between the mechanism and the causal model is that not all intermediate steps are relevant for causal inference, even though they do work as an additional check

Mediation and moderation

- We usually find more than one variable present in a mechanism
- Two typical variables: mediator and moderator
- Mediation: a third variable explains the causal relationship between two variables (e.g. flu infection $>$ immune reaction $>$ fever)
- Moderation: a third variable changes the effect of one variable on another (e.g. how age changes the immune reaction)

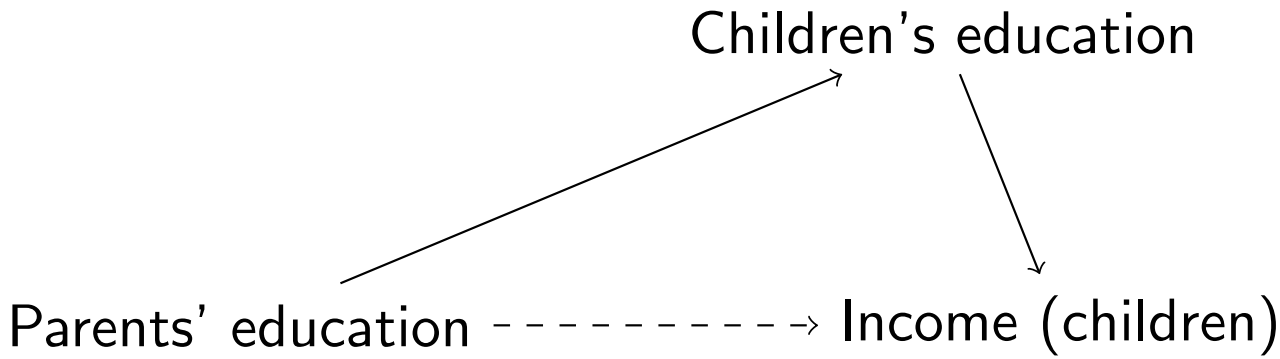
Example: income inequalities

Parents' education $\longrightarrow$ Income (children)

- Say we want to explain income inequality, and we find that people whose parents went to university earn, on average, more. This would be the basic causal model.

(*Note:* in this case I use solid lines for direct effects and dashed lines for indirect effects, kind of)

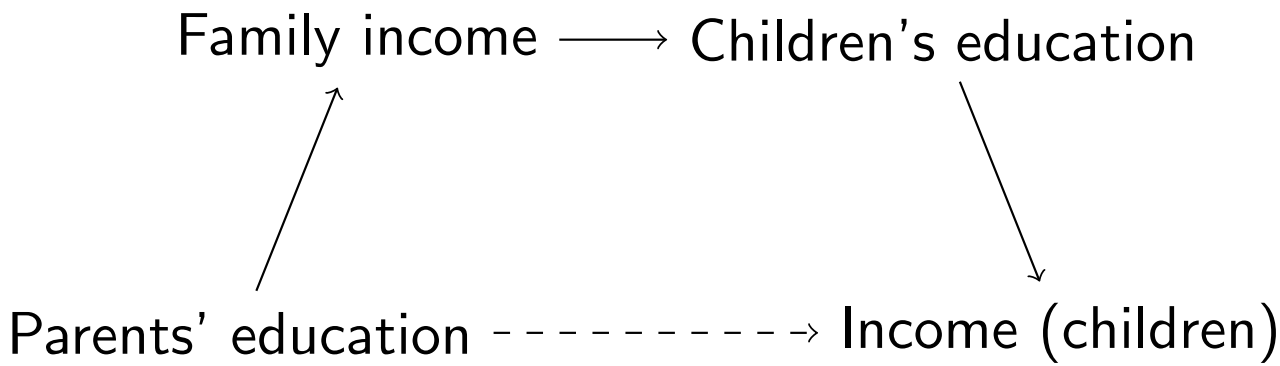
Example: income inequalities



- But *why* it is so? Someone comes and says: "It's because parents with higher education are more likely to send their children to university and help them get through."

(*Note:* in this case I use solid lines for direct effects and dashed lines for indirect effects, kind of)

Example: income inequalities



- And then someone comes and says: "It's not only that, it's money. Parents with higher education are richer and are able to send their kids to private schools and universities."

(*Note:* in this case I use solid lines for direct effects and dashed lines for indirect effects, kind of)

DAGs and mechanisms

- We can draw the process we're trying to study

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
  - → What had to happen for kids with university-degree parents to get richer?

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
  - → What had to happen for kids with university-degree parents to get richer?
  - → In other words, what mechanisms were in play behind what we see in the data?

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
  - → What had to happen for kids with university-degree parents to get richer?
  - → In other words, what mechanisms were in play behind what we see in the data?
- This matters when choosing our empirical strategy:

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
  - → What had to happen for kids with university-degree parents to get richer?
  - → In other words, what mechanisms were in play behind what we see in the data?
- This matters when choosing our empirical strategy:
  - → Main identification strategy

DAGs and mechanisms

- We can draw the process we're trying to study
- Main idea: focus on the *data generating process*
  - → What had to happen for kids with university-degree parents to get richer?
  - → In other words, what mechanisms were in play behind what we see in the data?
- This matters when choosing our empirical strategy:
  - → Main identification strategy
  - → Additional checks or implications (testing the mechanism, heterogenous effects, etc)

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Front doors and back doors

- We are interested in *identifying* the effect of pollution on happiness: that's out **front door**

Front doors and back doors

- We are interested in *identifying* the effect of pollution on happiness: that's out **front door**
- To do so, you have to make sure that the 'water flows' through this front door, and not through other 'pipe', or back door

Front doors and back doors

- We are interested in *identifying* the effect of pollution on happiness: that's out **front door**
- To do so, you have to make sure that the 'water flows' through this front door, and not through other 'pipe', or back door

- `Pollution -> Happiness` is our front door
- `Pollution <- Wealth -> Happiness` is a back door

Front doors and back doors

- We are interested in *identifying* the effect of pollution on happiness: that's out **front door**
- To do so, you have to make sure that the 'water flows' through this front door, and not through other 'pipe', or back door

- `Pollution -> Happiness` is our front door
- `Pollution <- Wealth -> Happiness` is a back door
- How do you close it? Just controlling for `Wealth`

Front doors and back doors

- There are esentially two ways to do causal inference:

1. Close all back doors and leave only the front door open

   That's where DAGs help to identify these variables

2. Using some other method where only the front door is opened

   (Finding and analysing exogenous variation)

Off topic: Controlling

- How does alcohol consumption affect health?
- Imagine we take data from a group of people:

```
df = data.frame(
  # In this group of people, one-third are rich
  rich = rbinom(500, 1, 0.3)) %>%
  # Rich people have 3x more money to buy whiskey
  mutate(whiskey = 3*rich + runif(500, 0, 4)) %>%
  # Health risk is worse if you drink more whiskey, but
    rich people have better health overall
  mutate(risk = -2*rich + .3*whiskey + rnorm(500, 2))

```

Off topic: Controlling

```
cor(df$whiskey, df$risk)
[1] -0.1150553
```

Off topic: Controlling

- Controlling for `rich` look at the variation **not** explained by `rich`
- i.e., take the group prediction out (mean of whiskey/risk for rich or non-rich)

```
df = df %>%
  group_by(rich) %>%
  mutate(whiskey_resid = whiskey - mean(whiskey),
         risk_resid = risk - mean(risk)) %>%
  ungroup()
```

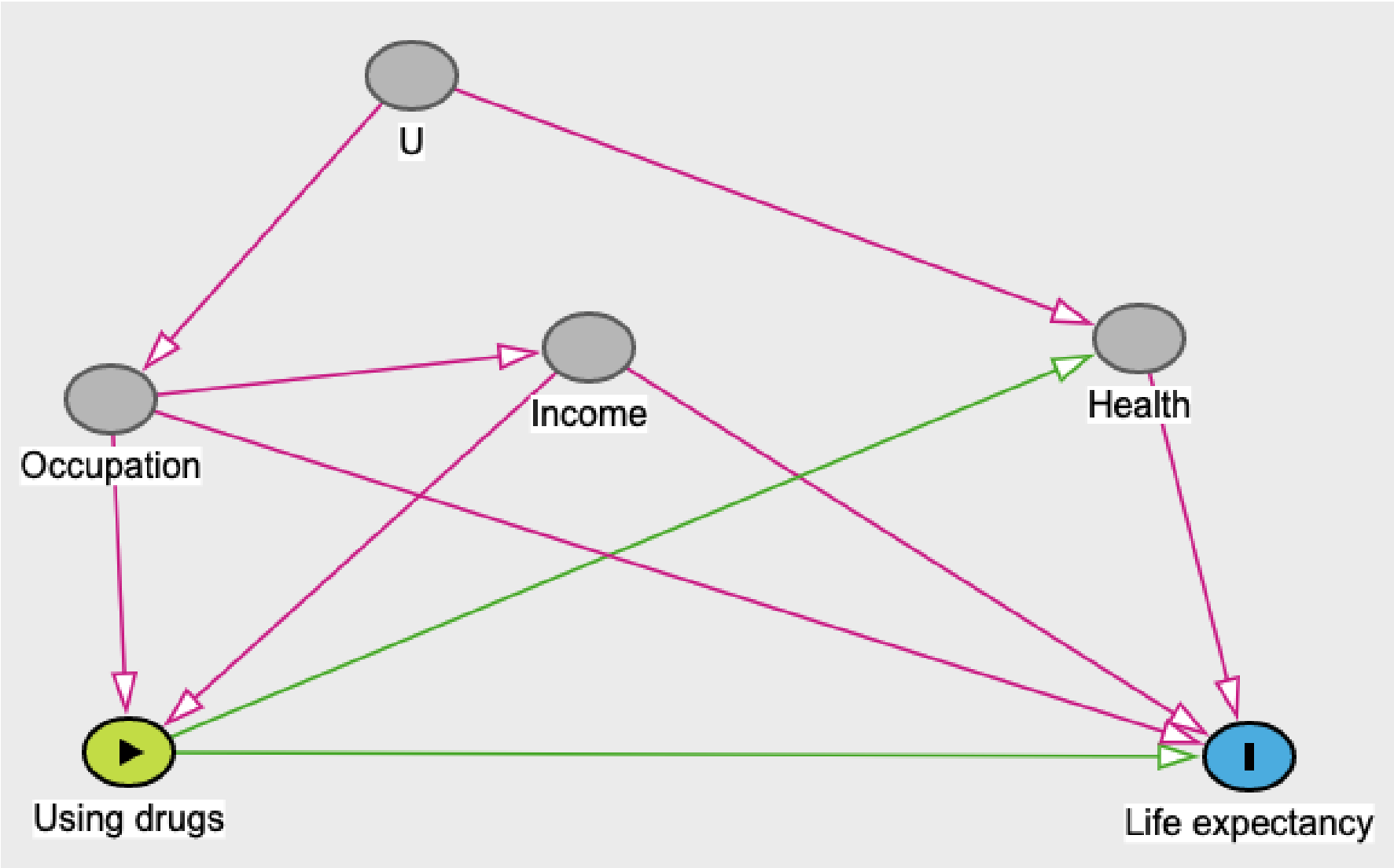
Off topic: Controlling

- The *true* model we created:

  `risk = -2 * rich + .3 * whiskey + error`

```
cor(df$whiskey_resid, df$risk_resid)
[1] 0.3242735
```

Front doors and back doors



Front doors and back doors

- `Drugs > LifeExp`
- `Drugs > Health > LifeExp`
- `Drugs < Income > LifeExp`
- `Drugs < Occup > LifeExp`
- `Drugs < Occup > Income > LifeExp`
- `Drugs < Occup < U > Income > LifeExp`
- `Drugs < Occup < U > Health > LifeExp`

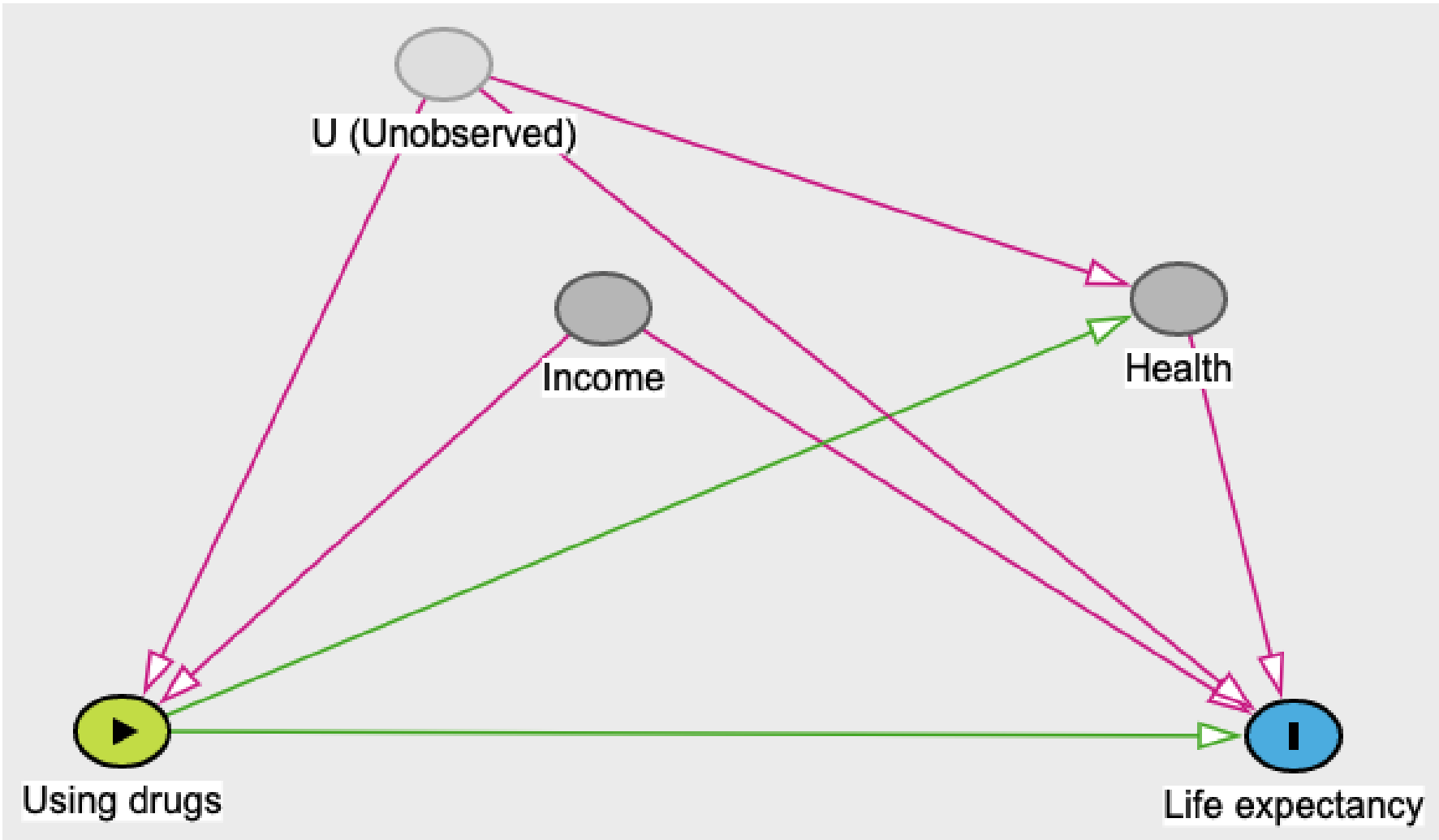
Front doors and back doors

- Drugs > LifeExp
- Drugs > Health > LifeExp
- Drugs < Income > LifeExp
- Drugs < Occup > LifeExp
- Drugs < Occup > Income > LifeExp
- Drugs < Occup < U > Income > LifeExp
- Drugs < Occup < U > Health > LifeExp

Front doors and back doors

- We just need to control for one of the variables in the path of a back door to close that path
- In this example, it would be enough to control for `income` and `occupation`
- This is the **back door criterion**

Front doors and back doors



Front doors and back doors

- Drugs > LifeExp
- Drugs > Health > LifeExp
- Drugs < Income > LifeExp
- Drugs < U > Income > LifeExp
- Drugs < U > Health > LifeExp
- Drugs < U > LifeExp

Front doors and back doors

- **What if we control for health?**

Front doors and back doors

- **What if we control for health?**
- We would be blocking part of the causal 'water flow' from drugs to life expectancy

Front doors and back doors

- **What if we control for health?**
- We would be blocking part of the causal 'water flow' from drugs to life expectancy
- That's part of the mechanism: imagine that drugs has a direct effect, e.g. higher probability of dying on an accident, and an indirect effect through its effect on health

Front doors and back doors

- **What if we control for health?**
- We would be blocking part of the causal 'water flow' from drugs to life expectancy
- That's part of the mechanism: imagine that drugs has a direct effect, e.g. higher probability of dying on an accident, and an indirect effect through its effect on health
- (Unless you want to calculate the *direct effect*)

Front doors and back doors

- **What if we control for health?**
- We would be blocking part of the causal 'water flow' from drugs to life expectancy
- That's part of the mechanism: imagine that drugs has a direct effect, e.g. higher probability of dying on an accident, and an indirect effect through its effect on health
- (Unless you want to calculate the *direct effect*)

Front doors and back doors

- **What if we control for health?**
- We would be blocking part of the causal 'water flow' from drugs to life expectancy
- That's part of the mechanism: imagine that drugs has a direct effect, e.g. higher probability of dying on an accident, and an indirect effect through its effect on health
- (Unless you want to calculate the *direct effect*)

- So we would have to control for **U**, which would close all other paths

Front doors and back doors

- `Drugs > LifeExp`
- `Drugs > Health > LifeExp`
- `Drugs < Income > LifeExp`
- `Drugs < U > Income > LifeExp`
- `Drugs < U > Health > LifeExp`
- `Drugs < U > LifeExp` **(!)**

- Problem? So?

Roadmap

Intro to explanation

Potential outcomes framework

Experiments

Causal models and diagrams

Back doors and front doors

Usual suspects

Usual suspects

- Confounding
- Reverse causality
- Bidirectional causation
- Selection bias
- Collider bias
- Post-treatment bias

Confounding

- Typical example: as the number of pirates in the oceans decreased, global mean temperature increased. Does it mean the disappearance of pirates is causing global warming?

Confounding

- Typical example: as the number of pirates in the oceans decreased, global mean temperature increased. Does it mean the disappearance of pirates is causing global warming?
- No, both are caused by the industrial revolution or technological development

Confounding

- Typical example: as the number of pirates in the oceans decreased, global mean temperature increased. Does it mean the disappearance of pirates is causing global warming?
- No, both are caused by the industrial revolution or technological development
- Months when people eat more ice-creams, also more people drown in the beach. Ice-creams causing drownings?

Reverse causality

- Many examples where correlations we think imply a particular causal effect might be explained by its reverse: Violent videogames making teenagers violence? Maybe violent teeneagers prefer those games. Drug use causes psychological problems? Maybe psychological problems can also cause drug use. (In many cases it's also two-way causality)

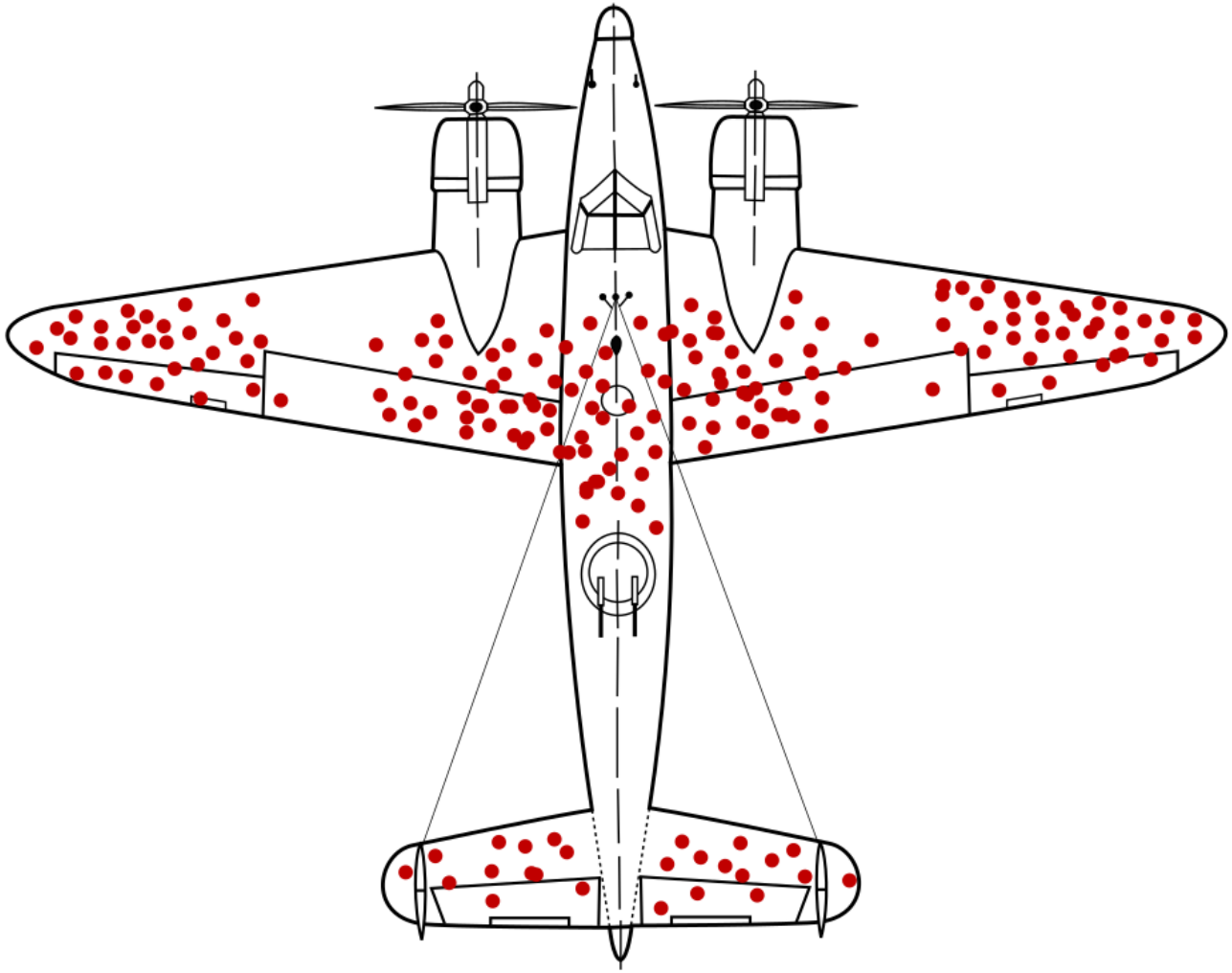
Reverse causality

- Many examples where correlations we think imply a particular causal effect might be explained by its reverse: Violent videogames making teenagers violence? Maybe violent teeneagers prefer those games. Drug use causes psychological problems? Maybe psychological problems can also cause drug use. (In many cases it's also two-way causality)
- "Hospitals make people sick:" if you collect data on illness development, you might find that people fare worse if they go to the hospital. Obviously, it's a case of reverse causality: being sick causes going to the hospital.

Bidirectional causation

- (endogenous cycles, $\neq$ reverse causality)

- Political values and voting: they way you think makes you vote in a particular way, but the way you vote can also affect the way you think (group influence, cognitive processes, etc)
- Can be closely related to selection bias: imagine we go to Madrid Rio and we measure if people doing exercises are more likely to be overweight than those lying around
- We probably don't find any result. Does it mean exercise does not decrease overweight? No, it's probably bidirectional causation: overweight makes people more likely to exercise, and exercise reduces overweight



Selection bias

- Our observations are not representative

Selection bias

- Our observations are not representative
- Famous example from World War II airplanes

Selection bias

- Our observations are not representative
- Famous example from World War II airplanes
- Many examples: advice from successful CEOs, ex-heroin addicts more likely to do sports, etc

Selection bias

- Our observations are not representative
- Famous example from World War II airplanes
- Many examples: advice from successful CEOs, ex-heroin addicts more likely to do sports, etc
- Why?

Selection bias

- Our observations are not representative
- Famous example from World War II airplanes
- Many examples: advice from successful CEOs, ex-heroin addicts more likely to do sports, etc
- Why?
  - $\rightarrow$ Sampling

Selection bias

- Our observations are not representative
- Famous example from World War II airplanes
- Many examples: advice from successful CEOs, ex-heroin addicts more likely to do sports, etc
- Why?
  - → Sampling
  - → Attrition ($\approx$ survivorship bias)

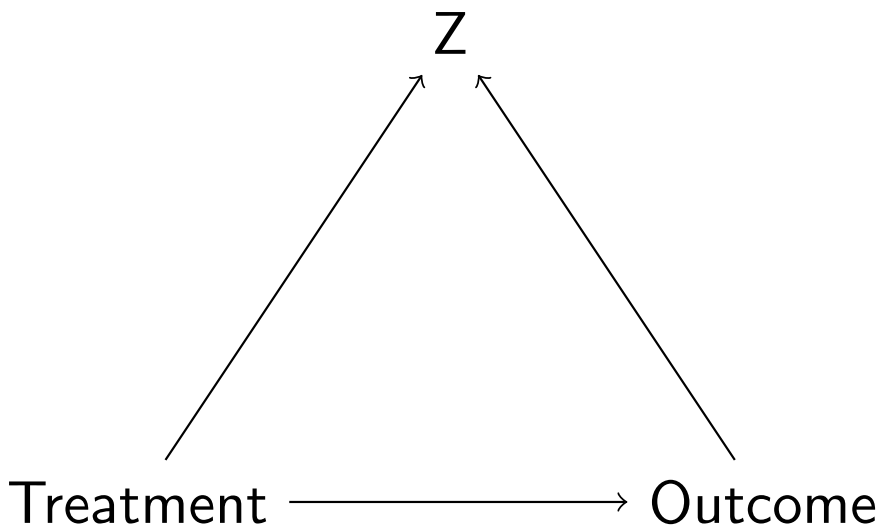
Selection bias

- Our observations are not representative
- Famous example from World War II airplanes
- Many examples: advice from successful CEOs, ex-heroin addicts more likely to do sports, etc
- Why?
  - → Sampling
  - → Attrition ($\approx$ survivorship bias)
  - → etc

Selection bias in causal inference

- Selection bias in statistics: sampling issue
- Quite different in causality: we're dealing with **selection into treatment**
- Remember example from HIV treatments studies

Collider bias



Collider bias

Collider bias

- Are good looking people jerks?
- We have 1,000 people, with **randomly** distributed beauty and niceness

```
df = data.frame(
  niceness = rnorm(1000, mean = 5, sd = 1.5),
  beauty = rnorm(1000, mean = 5, sd = 1.5))

```

Collider bias

- No correlation

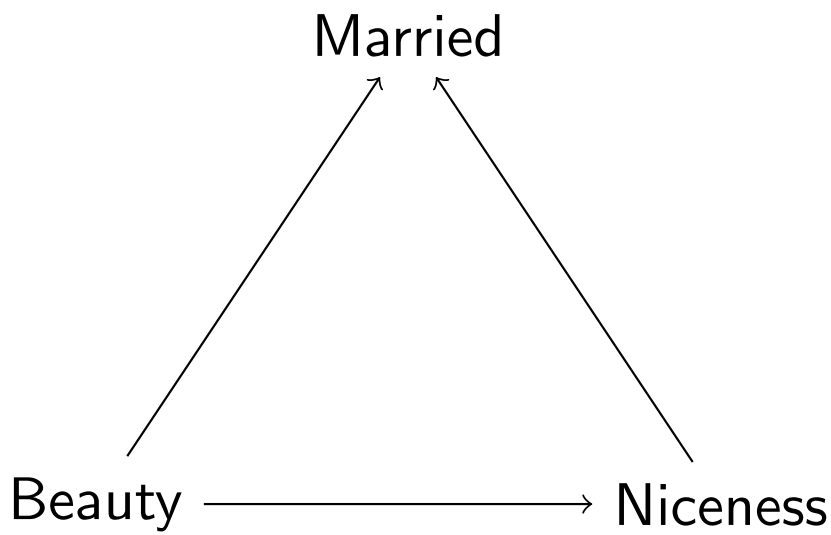
```
cor(df$niceness, df$beauty)
[1] 0.02876464

```

Collider bias

	Model 1
(Intercept)	4.857***
	(0.155)
beauty	0.027
	(0.030)
Num.Obs.	1000
R2	0.001

Collider bias

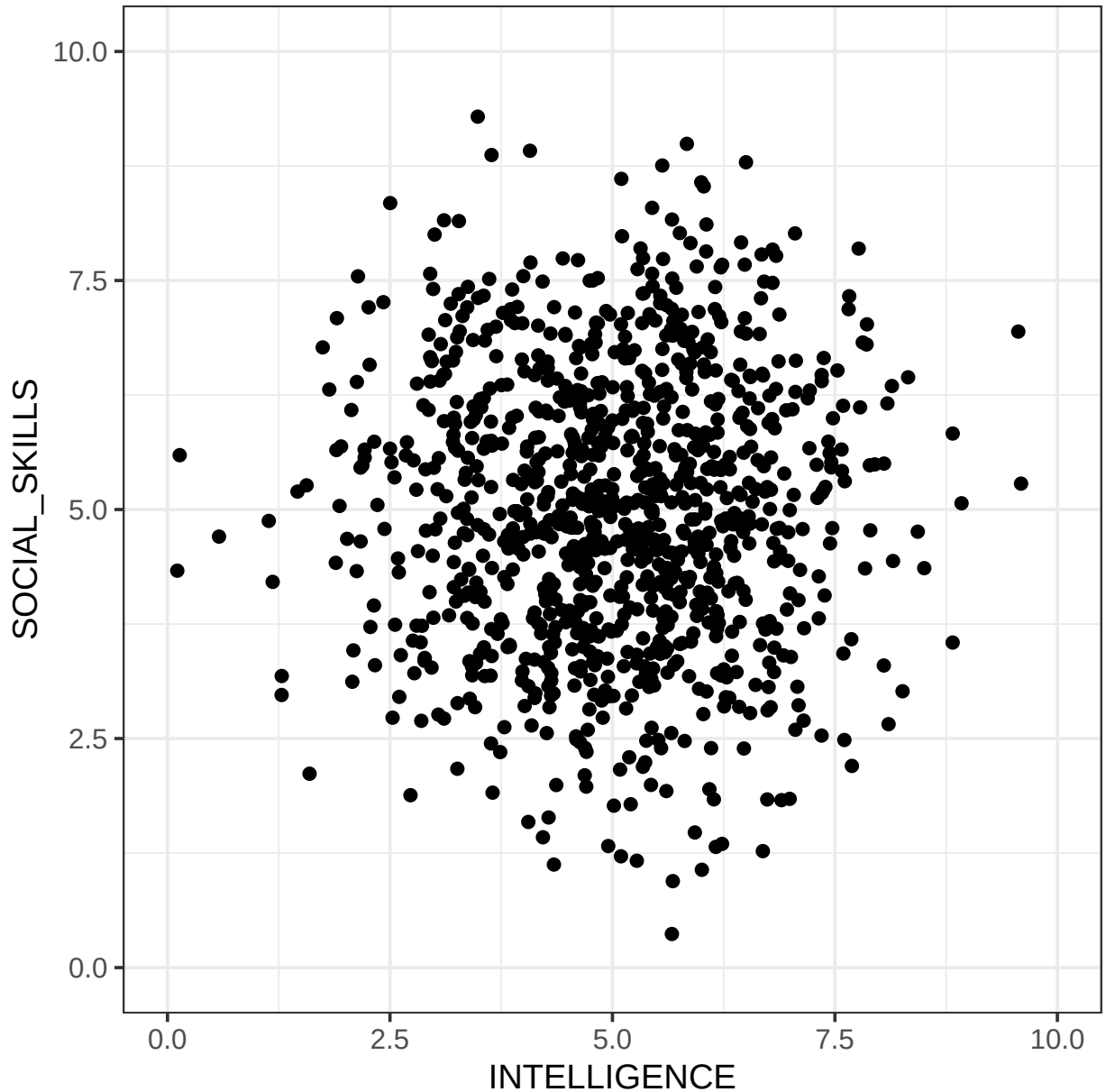
- Now imagine that we have another variable, the probability of being married, which is we will say is caused by both niceness and beauty:



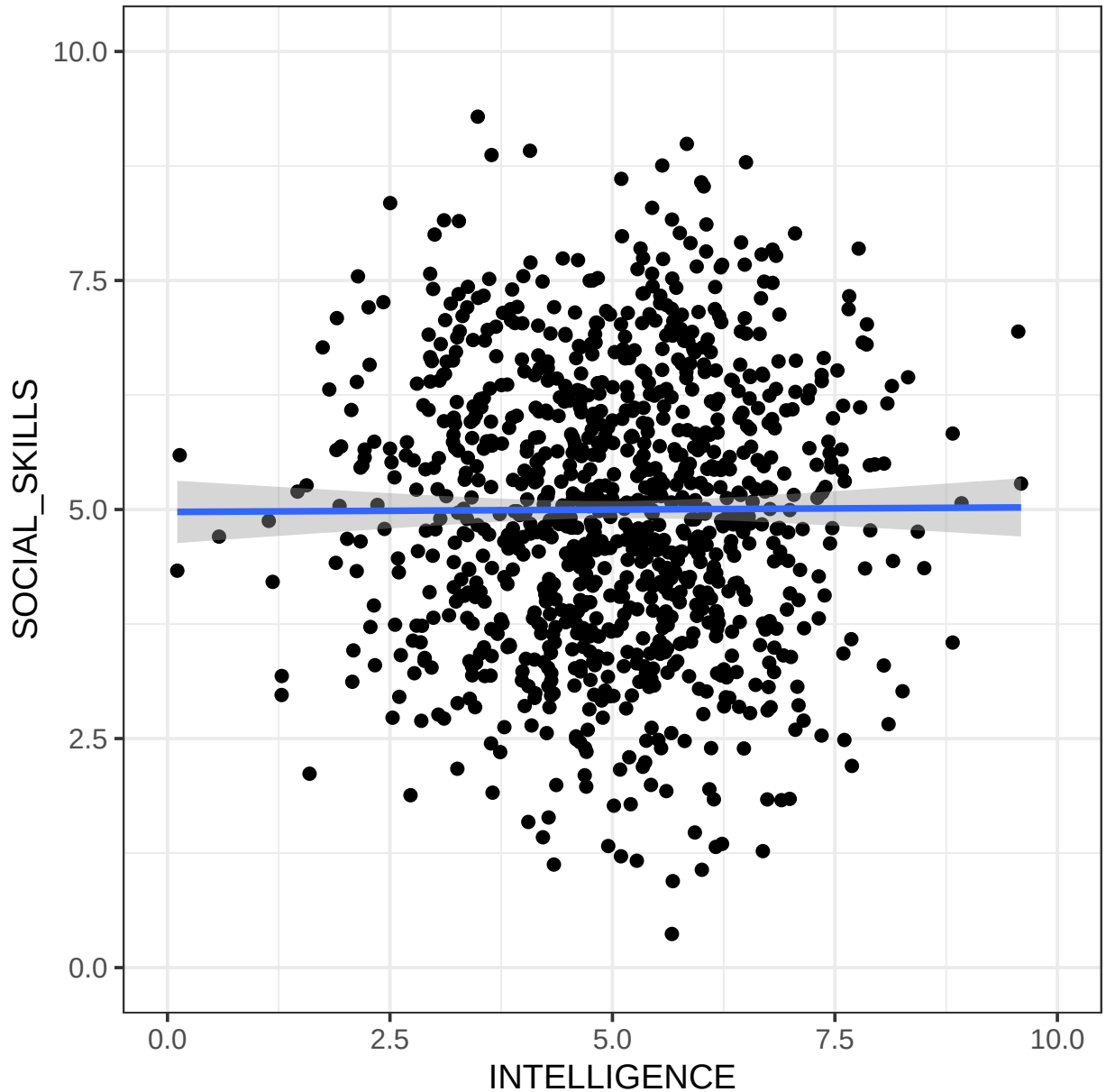
Collider bias

	Model 1	Model 2
(Intercept)	4.857***	0.000***
	(0.155)	(0.000)
beauty	0.027	−2.000***
	(0.030)	(0.000)
married		28.671***
		(0.000)

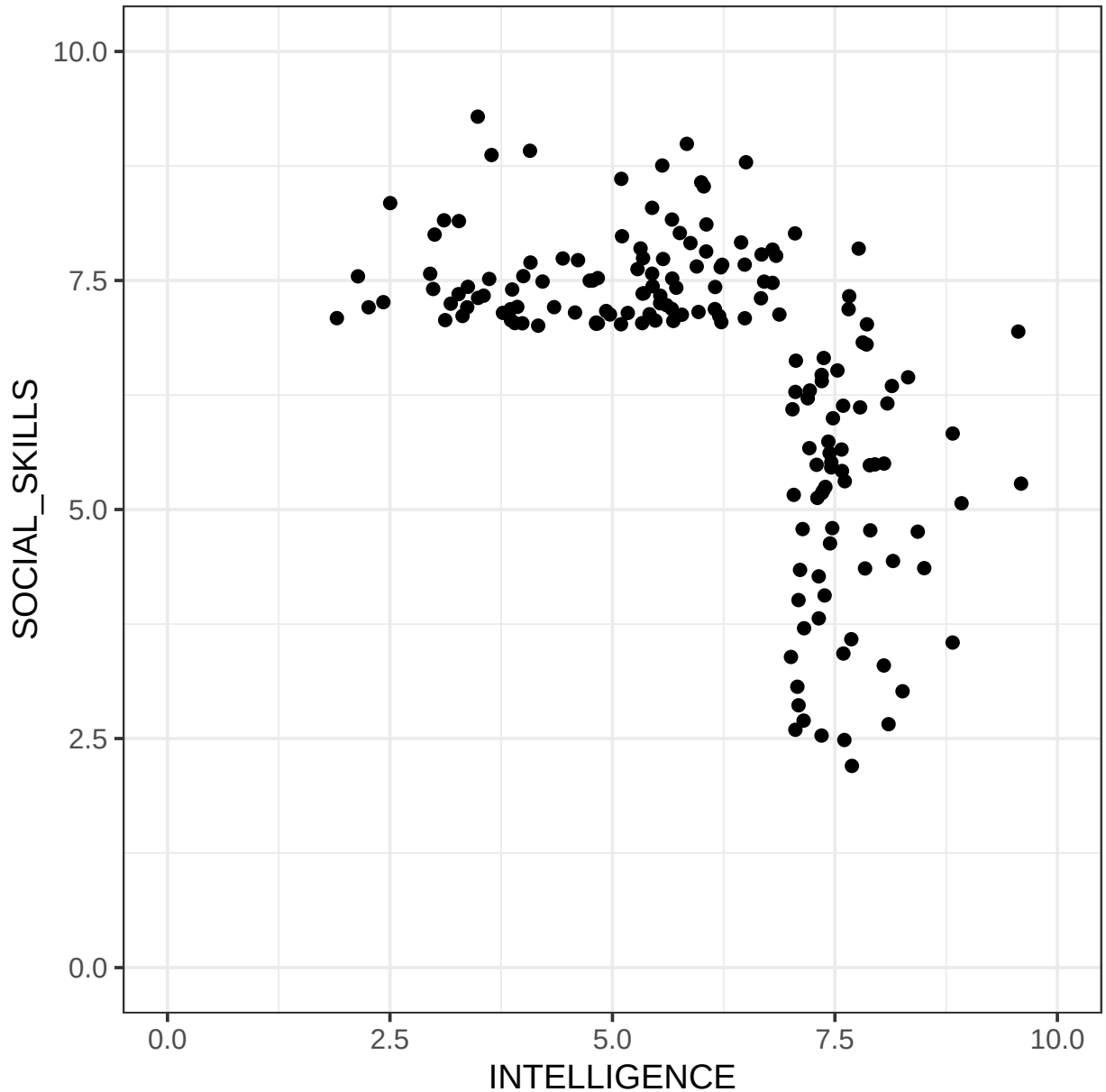
Collider bias



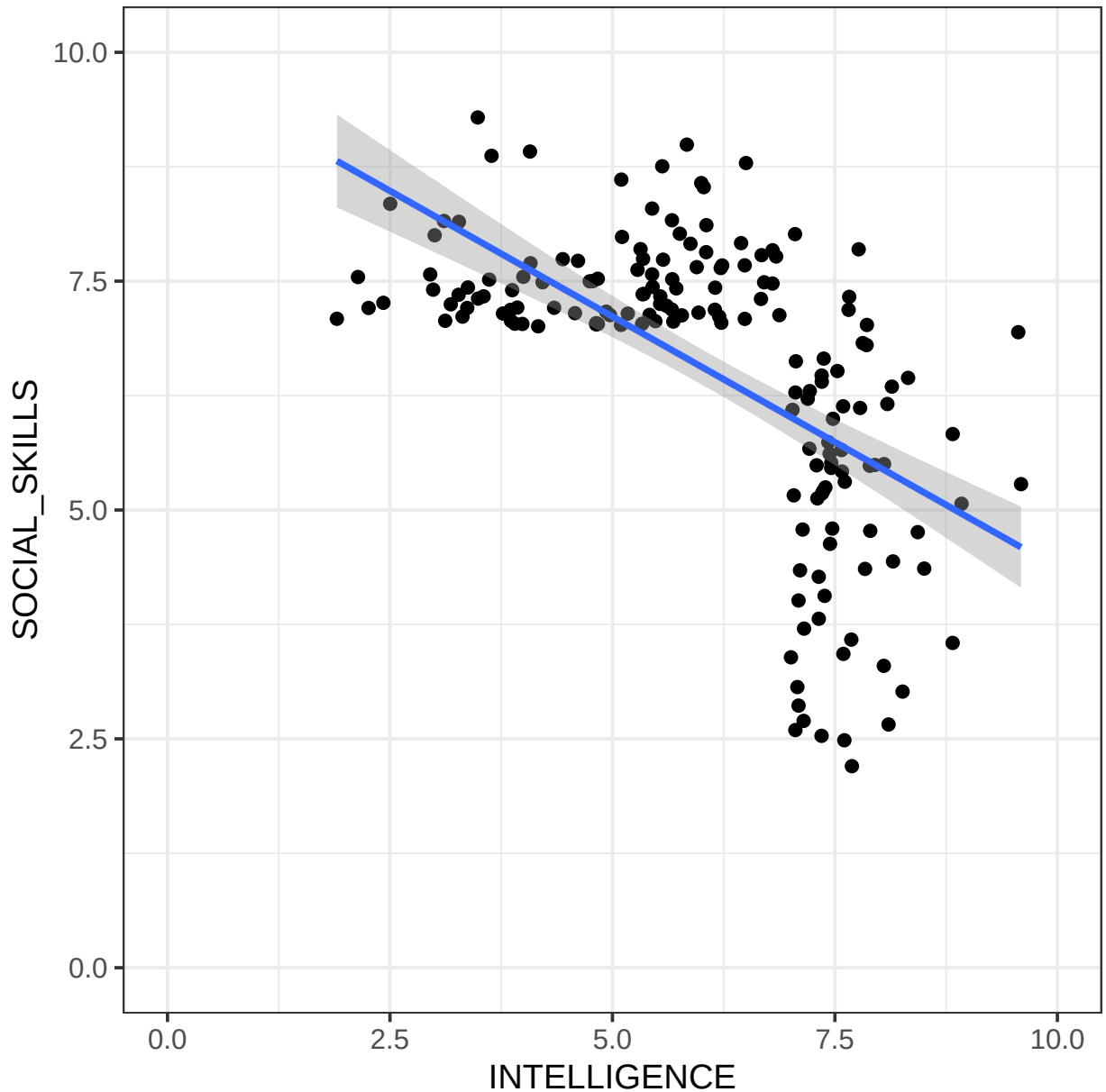
Collider bias



Collider bias



Collider bias

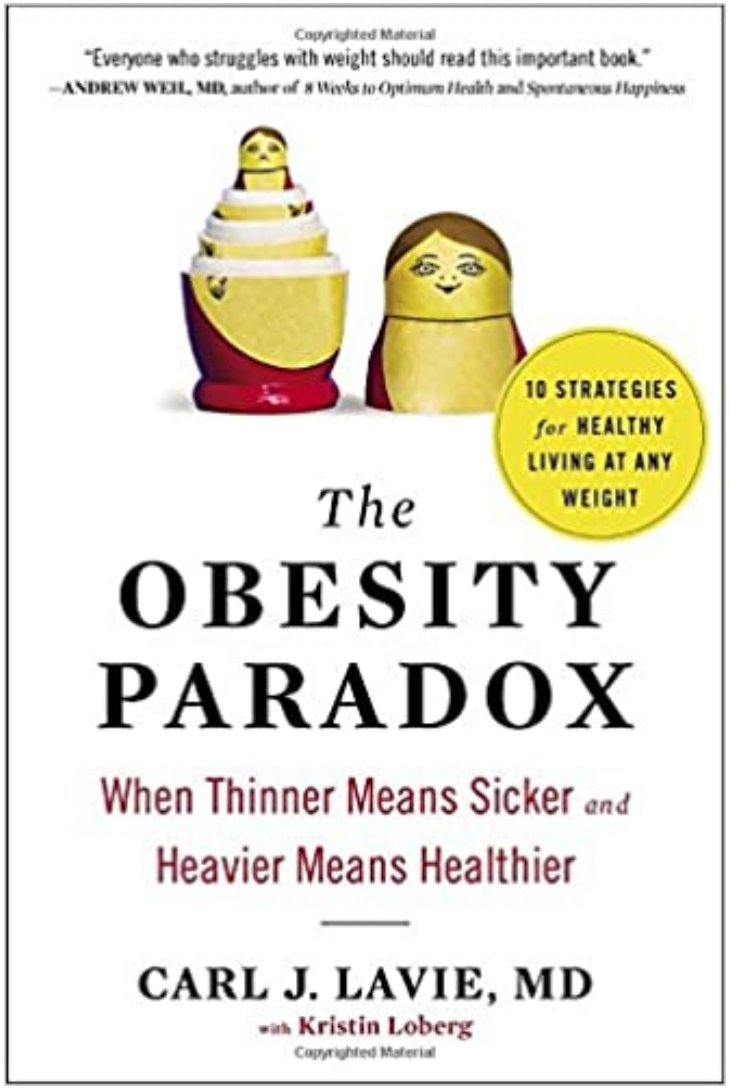


Collider bias

- A collider bias **opens** a path when you control for the variable

Collider bias

- Another example in life sciences where we can only use observational data
- Obesity reduces mortality among older people or patients with some chronic diseases (?)
- Collider bias? Y = health, X = environment/genetics



Collider bias

- Animated: https://nickchk.com/causalgraphs.html

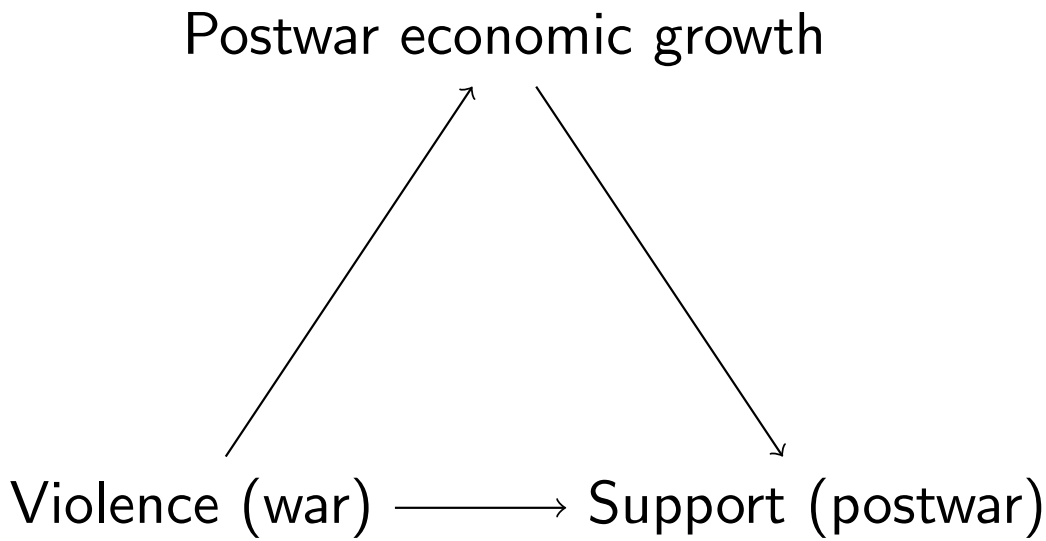
Post-treatment bias (collider again)

- We want to know whether suffering violence during a civil wars makes people more or less likely to support certain authorities decades after the war

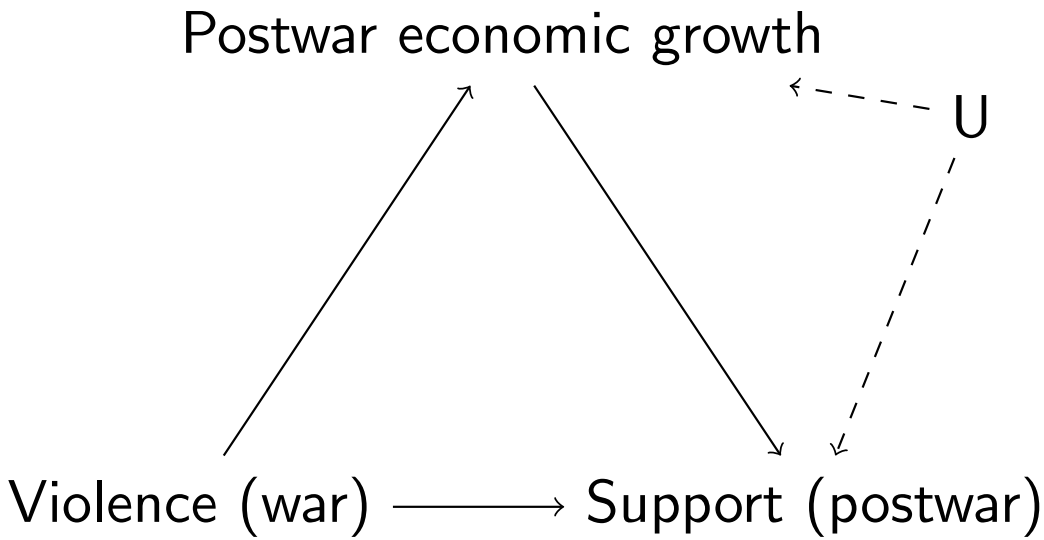
Post-treatment bias (collider again)

- We want to know whether suffering violence during a civil wars makes people more or less likely to support certain authorities decades after the war
- And we say: well, the country develop economically after the war, so maybe it makes sense to control for local increase in GDPpc, because it will also affect support

Post-treatment bias (collider again)



Post-treatment bias (collider again)



Recap: what should not be controlled for

1. Front-door paths
   - → Blocking some of the effect through a mediator variable
   - → (There are almost always mediator variables, so you could potentially just eliminate all the effect you're trying to identify)

2. Collider bias
   - → Opens a new, uncontrolled-for path
   - → Sometimes you might be inadvertently controlling for a collider because of *selection* issues
   - → Extra care with post-treatment bias