Causality and experiments

Francisco Villamil

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

Missing data in DAG framework

Missing completely at random (MCAR) Missing at random (MAR) Missing not at random (MNAR)

Roadmap

Prediction and explanation

Explanation

Explanatory questions and data

- Let's we want to know what's going on between two things (the process or mechanism between two variables)
- Data always comes from somewhere
- So when we look at or analyze data, we try to uncover the *data generating process*
  - → There could be several mechanisms generating the same data (variable), or they might vary over time
- (A model is actually our simplified guess of that process)

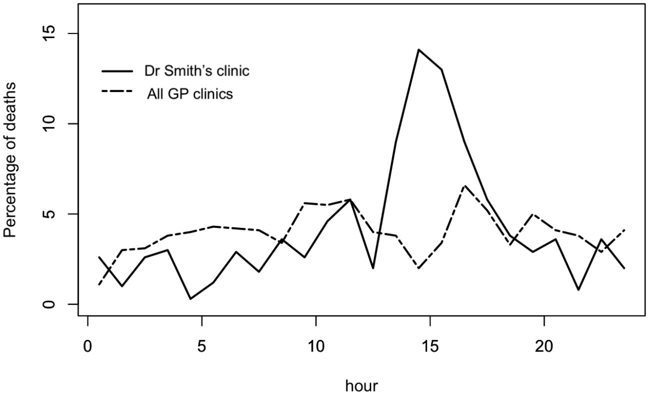
Data generating process

- How does the data look like, and what's the data generating process in...

- Flipping a coin?
- Choosing Mahou instead of another beer brand?
- Going on Erasmus and getting a job?

- It's kind of like the *causal model* that creates the outcomes, but thinking about data

Example



- Different generating processes? If so, what?

About prediction

- What is prediction?

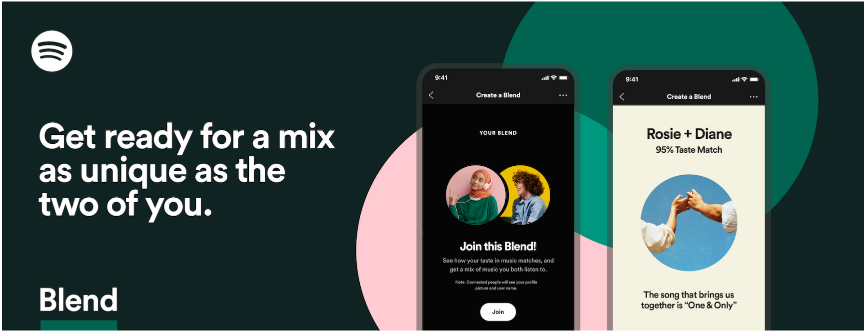
About prediction

- 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

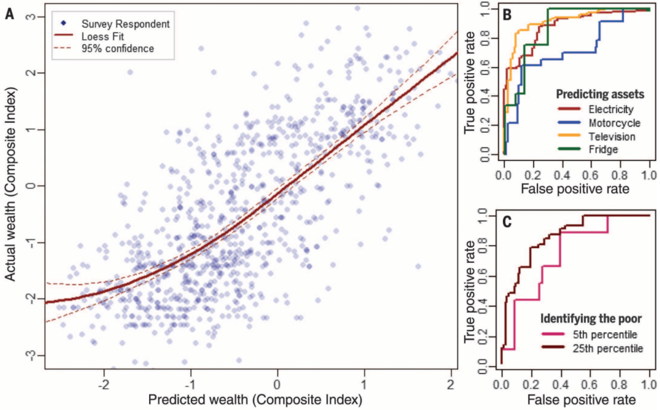


Fig. 1. Predicting survey responses with phone data. (**A**) Relation between actual wealth (as reported in a phone survey) and predicted wealth (as inferred from mobile phone data) for each of the 856 survey respondents. (**B**) Receiver operating characteristic (ROC) curve showing the model's ability 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

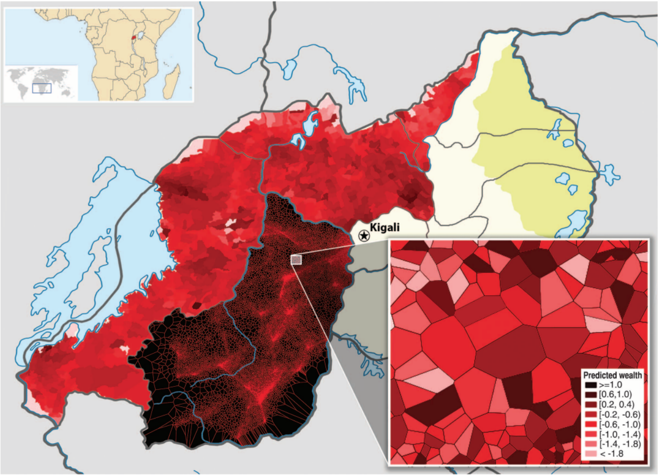


Fig. 2. Construction of high-resolution maps of poverty and wealth from call records. Information derived from the call records of 1.5 million subscribers is overlaid on a map of Rwanda. The northern and western provinces are divided into cells (the smallest administrative unit of the country), and the 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)

Roadmap

Prediction and explanation

Explanation

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
  - → I.e., we want to do *causal inference*

Explaining relationships

- Key thing: we want to know whether X actually *causes* Y
  - → 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
  - → 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
  - $\rightarrow$ 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**, that is, we will always 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)
- (ATT can be different from ATC when these two groups of units differ in characteristics, and kind of matters as well when estimating them)

Potential outcomes framework

- 'No causation without manipulation'
- Model initially developed for *experimental data*

Potential outcomes framework

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

Experimental design

Experiments also have their problems, e.g.:

Experimental design

Experiments also have their problems, e.g.:

- Non-perfect randomization (esp. if block-r)

Experimental design

Experiments also have their problems, e.g.:

- Non-perfect randomization (esp. if block-r)
- SUTVA (or stable unit treatment value assumption)

Experimental design

Experiments also have their problems, e.g.:

- Non-perfect randomization (esp. if block-r)
- SUTVA (or stable unit treatment value assumption)
- Attrition

Experimental design

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)

Experimental design

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)
- etc

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
- Using **observational data**, we need to 'build' a counterfactual. How?

So how to we approximate Y^0?

- Experiments are fine, but often not possible
- Using **observational data**, we need to ‘build’ a counterfactual. How?
- Basic idea: we build a causal model that explains the data generating process, and use that model to detect and correct biases in our causal inference

So how to we approximate Y^0?

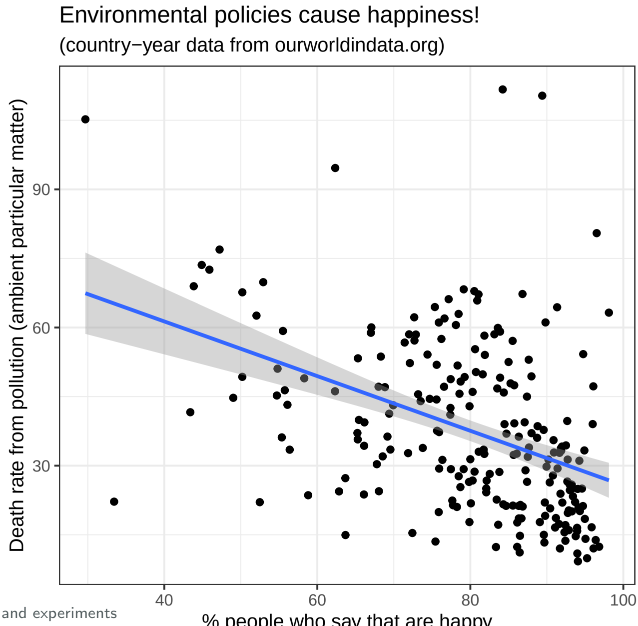
- Experiments are fine, but often not possible
- Using **observational data**, we need to 'build' a counterfactual. How?
- Basic idea: we build a causal model that explains the data generating process, and use that model to detect and correct biases in our causal inference

  (e.g., to know what we should control for)

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

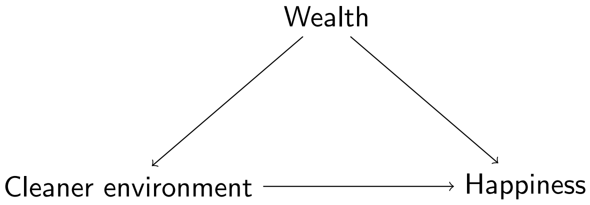
- Remember that out problem (the ‘fundamental problem of causal inference‘ etc) is that we can observe e.g. Pakistan, where the level of pollution (measured as death date) is 46, and 58% of the people say they’re happy
- But **we cannot observe** how many people say they are happy in an **alternative 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

Cleaner environment $\longrightarrow$ Happiness

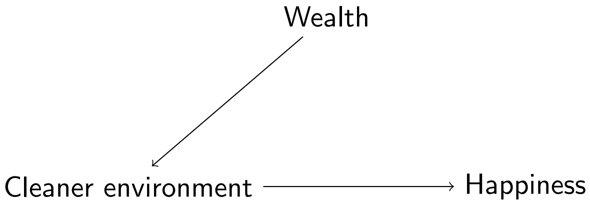
- This is our initial causal model: having a cleaner environment makes people happier (because they like looking into a blue sky without smog), and that's it. We do not have to control for anything nor do anything else.

Our causal model



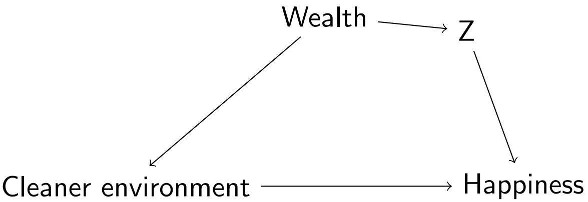
- Wait, but maybe it's about money, isn't it? Actually, wealthier countries tend to have cleaner environments and, at the same time, money causes happiness. **We need to control for wealth.**

Our causal model



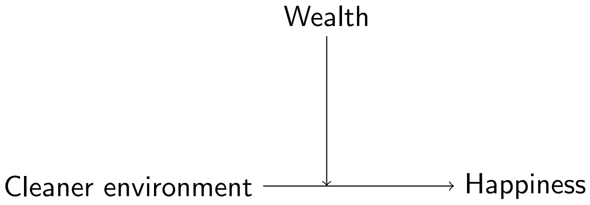
- Or perhaps is not that money increases happiness *per se*, but that it does so through other **mediators**: wealth allows countries to focus on environment, which increases happiness. **Again, no need to control.** As long as this is the **only** mediator.

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
  - $\rightarrow$ This applies from the most basic strategy (add controls) to the more complicated ones (e.g. evaluating DiD or RDD)

Basics of causal inference

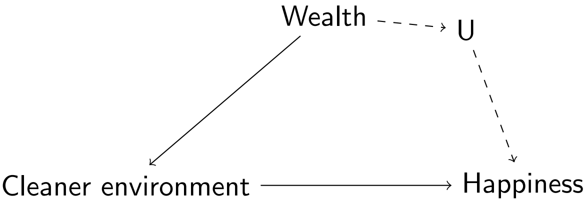
- So to come up with an strategy, we need to understand what's going on in terms of the data generating process
    - → This applies from the most basic strategy (add controls) to the more complicated ones (e.g. evaluating DiD or RDD)
- Once we have that, we can **identify** an effect (in other words: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

Causal models, mechanisms, and DAGs

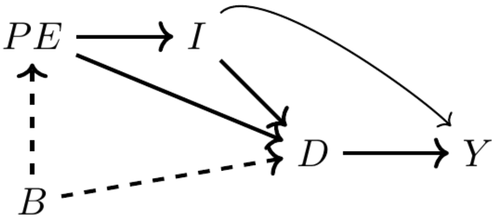
- We will use Directed Acyclic Graphs (DAG), or causal diagrams
- These are basically a graph where we link **variables** (nodes) with **causal effects** (arrows)

  Couple things:
- Only one-directional causality (*acyclic*)
  - $\rightarrow$ if you have feedback cycles, you'd have to 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 will use DAGs or similar, 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

Mechanisms

- A mechanism is a type of causal explanation that specifies **why** and **how** something happened
- Not exactly the same as a causal model, but very interrelated (not all intermediate steps are relevant for causal inference, but they do work as an additional check)

- Example: we know the flu gives us fever, but why?
  - → Correlation: flu infection and fever go together
  - → Causation: flu infection *causes* fever
  - → Mechanism: the immune system detects the infection and reacts by increasing the body temperature

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)

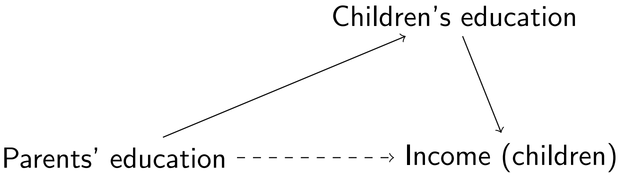
Example: income inequalities

$$\text{Parents' education} \longrightarrow \text{Income (children)}$$

- 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

$$
\begin{tikzcd}
\text{Family income} \arrow[r] & \text{Children's education} \arrow[d] \\
\text{Parents' education} \arrow[u] \arrow[r, dashed] & \text{Income (children)}
\end{tikzcd}
$$

- 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)