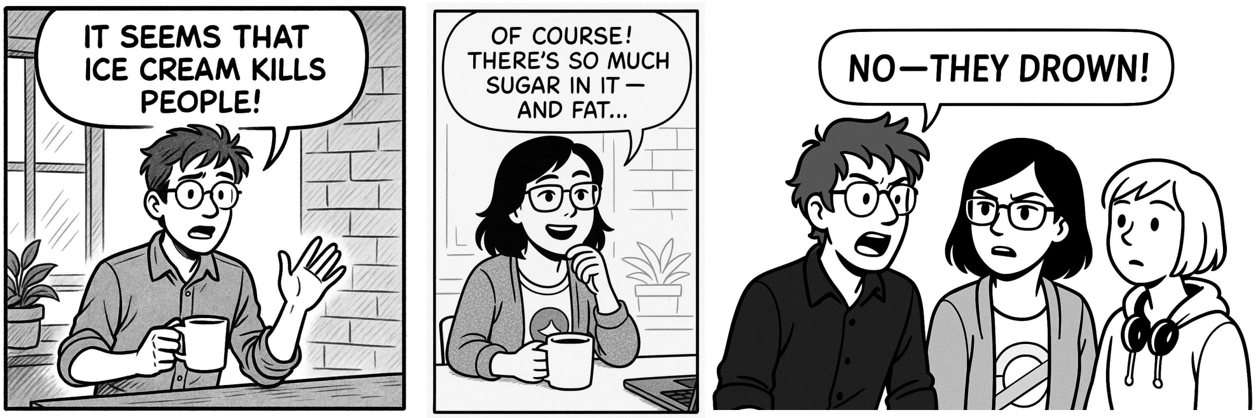
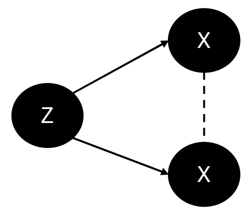


- A **fork** is a causal structure where **a single variable causes two others**:
- Z → X
- Z → Y
- **Z is a common cause** (a **confounder**) of both **X** and **Y**
- The path between **X and Y** is **opened** by **Z**, even though there's **no direct causal link** between X and Y



It looks like consuming ice cream causes drowning (X → Y), but really:

- **Hot day (Z)** causes both
- So there's a **fork structure**
- If we **don't adjust for hot day**, we might wrongly blame the ice cream for the drowning.

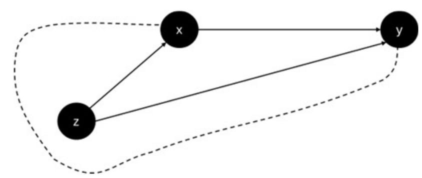
- Run icecream-kills.py and study the visualisations.
- Can you calculate Bayes (X|Y) or (Y|X)?
- Could you predict number of drownings by looking at the ice cream sales on a given day?
    - Fill in the blanks:
        - *Ice cream sales and drownings correlate / do not correlate.*
        - *... causes higher numbers of ice sales*
        - *... causes greater number of drownings*
        - *Ice sales are not causing ...*
        - *Drownings are not causing ...*

Backdoor paths (Confounders)

A **backdoor path** is a **non-causal path** from a treatment (or exposure) variable **X** to an outcome variable **Y** that can **create spurious associations** — it "sneaks in through the back door" and messes with your causal conclusions.

It brings in confounding — common causes of X and Y. You need to control for variables that block these paths to isolate the true causal effect.

- Exercising has an effect on health (more exercise [x] → better health [y]?).
- Age has an effect on health (older [z] → unhealthier [y]?).
- Age has an effect on exercise (older [z] → less exercise [y]?)

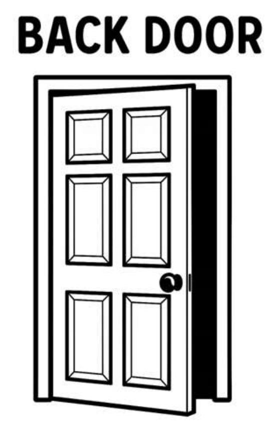


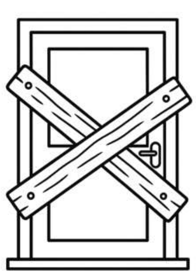
Controlling via regression (Blocking the backdoor path):

1. We use **linear regression** to **remove the part of Exercise and Health that can be predicted by Age**.

   ```
   model_x = LinearRegression().fit(df[['Age']], df['Exercise’])
   exercise_resid = df['Exercise'] - model_x.predict(df[['Age']])
   ```

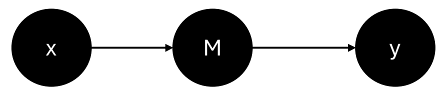
1. This gives us the part of Exercise that is not explained by Age. In other words: “what’s left of Exercise after removing Age’s effect.”
2. We do the same for Health
3. The correlation between these residuals tells us how **Health varies with Exercise, after removing the part that varies with Age**.





- Run the code in backdoor.py.
- Look at the visualizations:
- Can you see, how age “confounds” the results from exercise on health?
- What is different from the ice-cream – drownings example?
- What is the effect of blocking the path?
- Look in the code, to better understand how controlling works.

- A **mediator** is a node on the causal path between two nodes.
- If **X** causes **M**, and **M** causes **Y**, then **M is a mediator**.
- The treatment exerts some or all of its influence on the outcome through that mediator.
- Decomposing studies, how much of the effect is mediated.



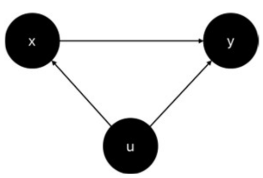
- Run the code in mediator.py.
- Look at the visualizations:
- What is the effect of coffee on alertness?
- What is the effect of alertness on productivity?
- Compare the direct effect of coffee on alertness and alertness on productivity with the indirect effect of coffee on productivity.
- Look in the code, to better understand the concept of mediator.

You're studying whether **eating jelly beans (X)** leads to **reduced focus during study sessions (Y)** for **university students**.

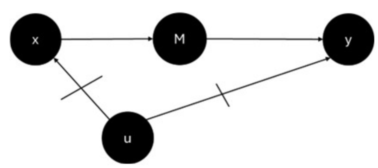
There's an **unobserved confounder (U)**: the **student's lifestyle** (e.g. sleep schedule, stress level, diet quality) — which affects both:

- How many jelly beans they eat, and
- How well they can focus.

But we **can observe a mediator**: the **sugar level in their bloodstream (M)**.



1. unobserved confounder (u) affects x and y.

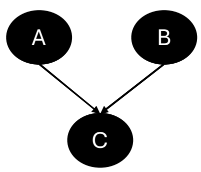


2. Mediator (m) is not affected (directly) by u, backdoor x via u to y is of no concern.

- Run the code in frontdoor.py.
- Look at the visualizations:
- Can you see, how life style (c.f. far right plot) “confounds” the results from jelly beans on study focus?
- What is the effect of sugar level on study focus?
- What is the effect of jelly beans on sugar level?
- Fill in the blanks:

  While we cannot tell the effect of jelly beans on study focus, as other factors ..., we can confirm, that ... has a negative effect on study focus. Insofar as jelly beans have a direct impact on ..., we advise ... to improve study focus.
- Look in the code, to better understand how controlling works in this case.

A **collider** is a variable **C** in a causal graph such that there are two (or more) variables, say **A** and **B**, that both have **directed edges into C.**



Colliders are **critical** in determining whether a path in a causal graph is **blocked or open** (which affects whether two variables are statistically independent).

- **Unconditioned**, a path that includes a collider is **blocked** — it **does not allow information to flow** between the variables.
- **Conditioning on a collider** (or on its descendants) **opens the path**, creating a **spurious association** between the upstream variables.

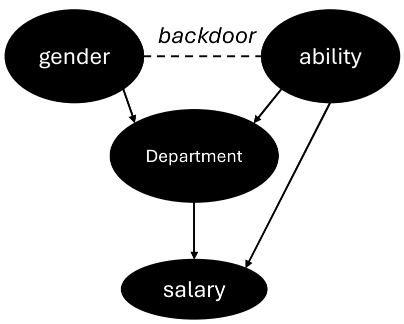
- Run collider.py.
- Look at the visualization.
- Describe and explain: Literary Quality and Curriculum Score.
- Describe and explain: Moral Message and Curriculum.
- Confirm or deny:

  There is (no) correlation between Moral Message and Literary Quality.
- Look in the code. How is the curriculum score calculated?

When we adjust a curriculum score and pick books based on it, we condition on the curriculum score.

- Confirm or deny only for those books, that are picked for the curriculum:

  There is (no) correlation between Moral Message and Literary Qualtity



A firm says, they do not discriminate between men or women, they both get similar wages for the same kind of jobs simply based on their abilities.

- Run collider2.py.
- Look at ALL the visualizations (there are two windows). What would you say?
- The firm conditions on departments ("for the same kind of jobs"). Why is that a problem?