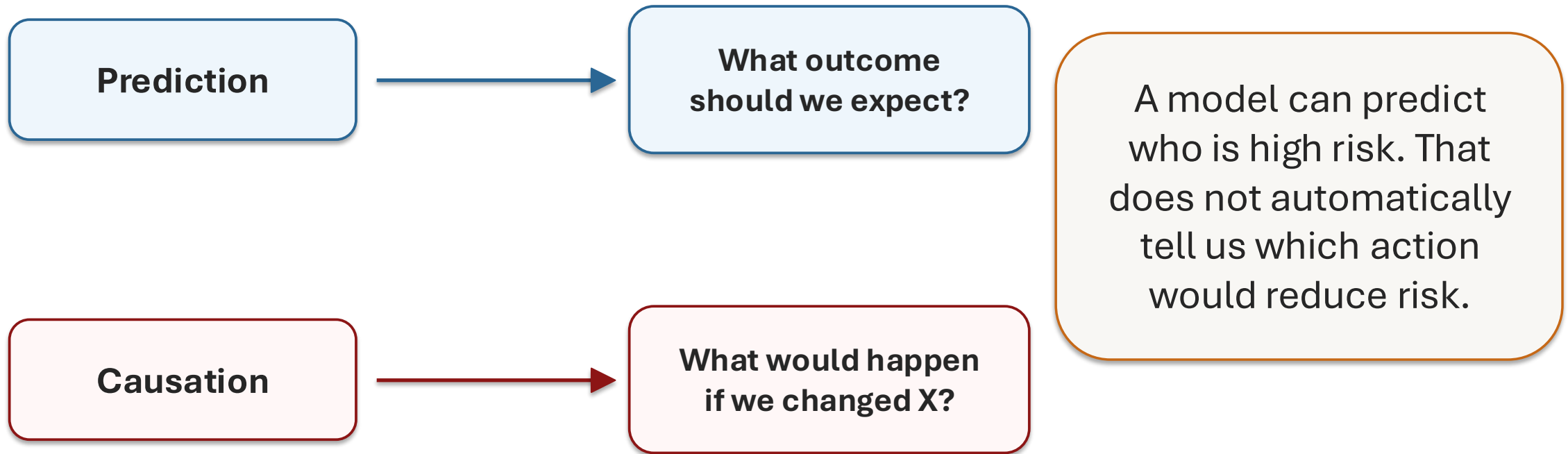


Causal Inference I

Prediction Is Not Causation

Correlation, counterfactuals, and the first language of causal graphs

From prediction to intervention



Two kinds of questions

Prediction question	Causal question
Which students are likely to earn more?	Would attending Stanford increase earnings?
Which users are likely to churn?	Would a retention email reduce churn?
Which hospitals have high readmission rates?	Would a discharge protocol reduce readmissions?
Which patients are high risk?	Would a treatment improve outcomes?

Prediction helps identify risk. Causal inference helps choose actions.

The minimal vocabulary

Term	Meaning	Example
Unit	The object we compare	student, user, patient, school
Treatment / exposure	The thing we might change	tutoring, email, Stanford attendance
Outcome	The result we care about	score, churn, earnings
Comparison group	The substitute for the missing counterfactual	similar students without tutoring

DISCUSSION

- 1. Which users are likely to churn?**
- 2. Would a discount reduce churn?**
- 3. Which colleges have high-earning graduates?**
- 4. Would attending Stanford increase earnings?**

Lecture 18 map

1

Three suspicious correlations

2

Why each causal story can fail

3

Potential outcomes: missing half the data

4

Stanford premium: real association, hard causality

5

DAGs: first graphical language

A correlation can be true and still not causal

Correlation question: Do X and Y move together?

Causal question: If we intervened to change X, would Y change?

Thinking: Some intuitions why these two are different?

Three suspicious correlations

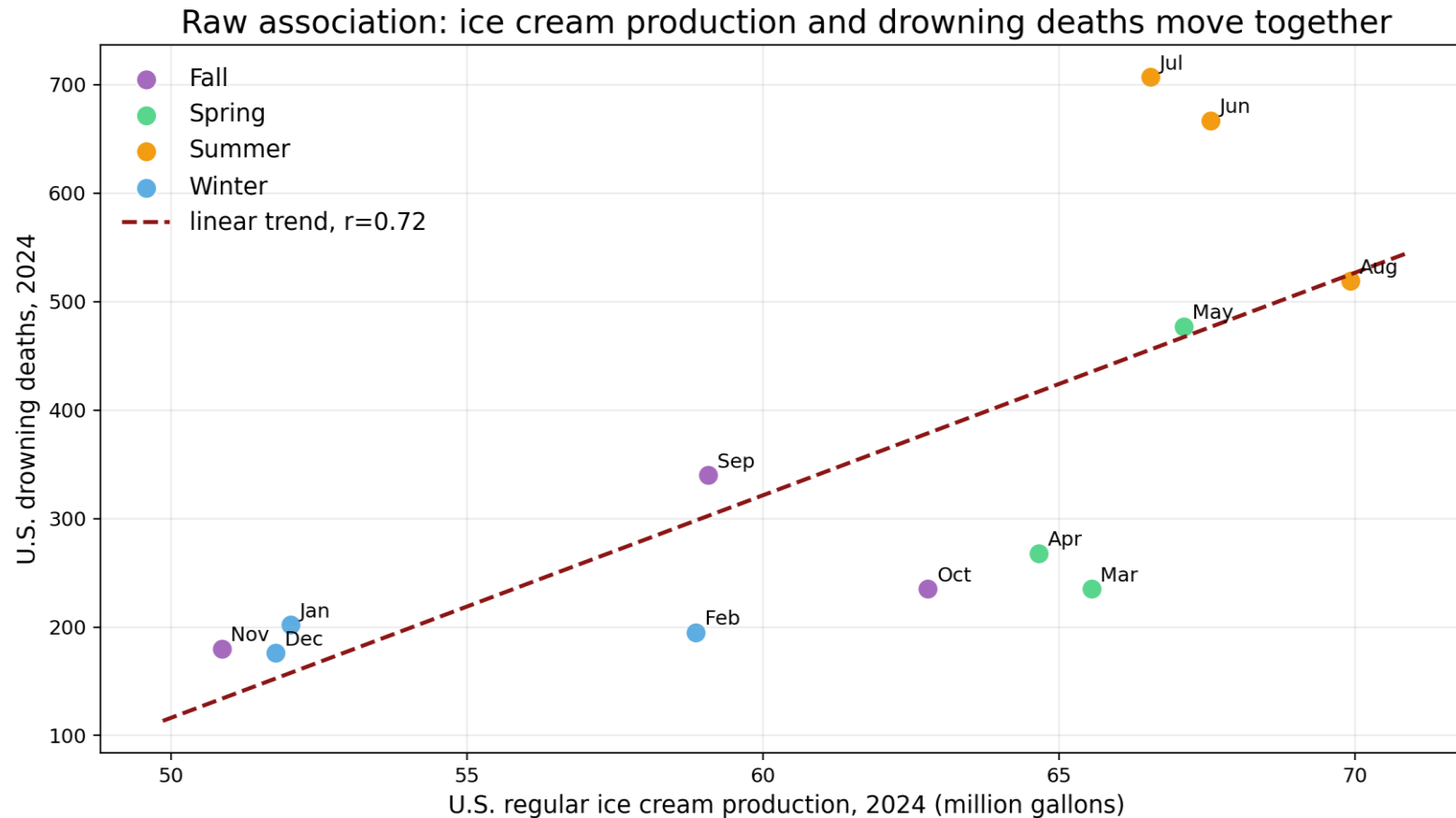
Three correlations to question

Observed relationship	Naive causal question
Ice cream production and drowning deaths move together	Does ice cream increase drowning?
Places with more crime also have more police activity	Does police activity increase crime?
Selected groups sometimes have worse outcomes than non-selected groups	Did the treatment/selection make outcomes worse?

Case 1: confounding

A third variable can create a real association.

Case 1: ice cream and drowning



Ice cream production is higher in months with more drowning deaths.

Does ice cream cause drowning?

From a correlation perspective, the pattern is real.

Notation:

Role	In this example
Exposure X	Ice cream production
Outcome Y	Drowning deaths
Other factors Z	?

What else changes between winter and summer besides ice cream production?

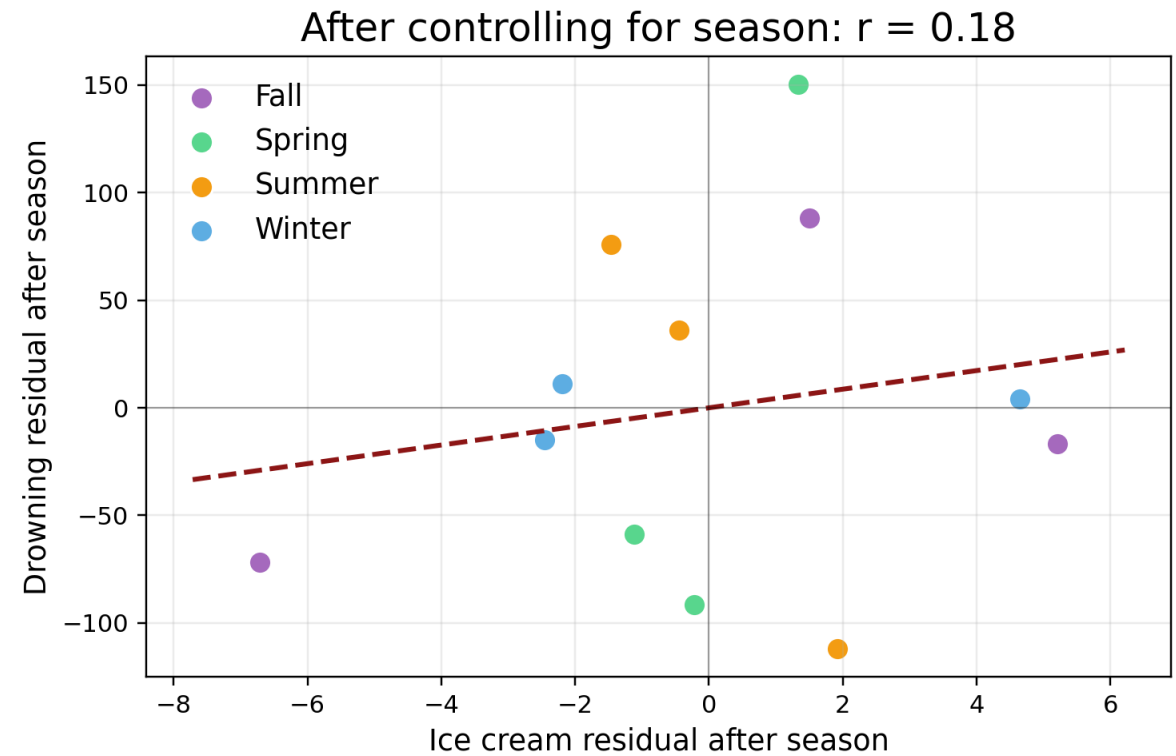
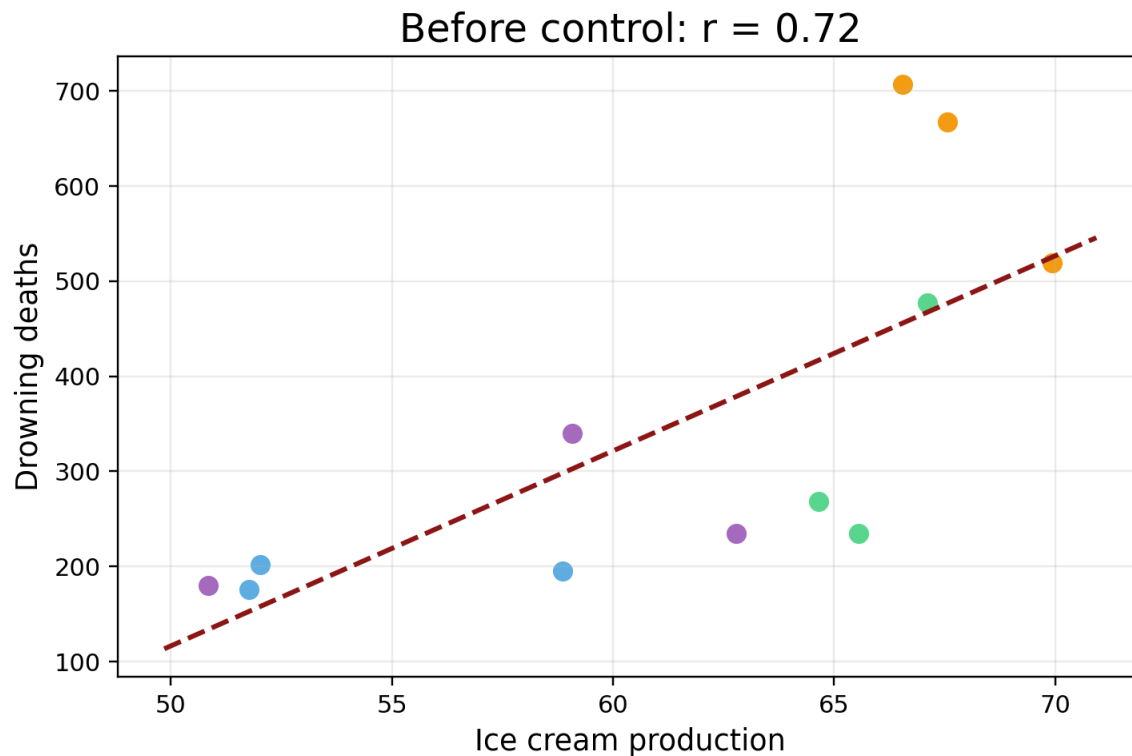
Notation:

Role	In this example
Exposure X	Ice cream production
Outcome Y	Drowning deaths
Other factors Z	Season / temperature / swimming exposure

What else changes between winter and summer besides ice cream production?

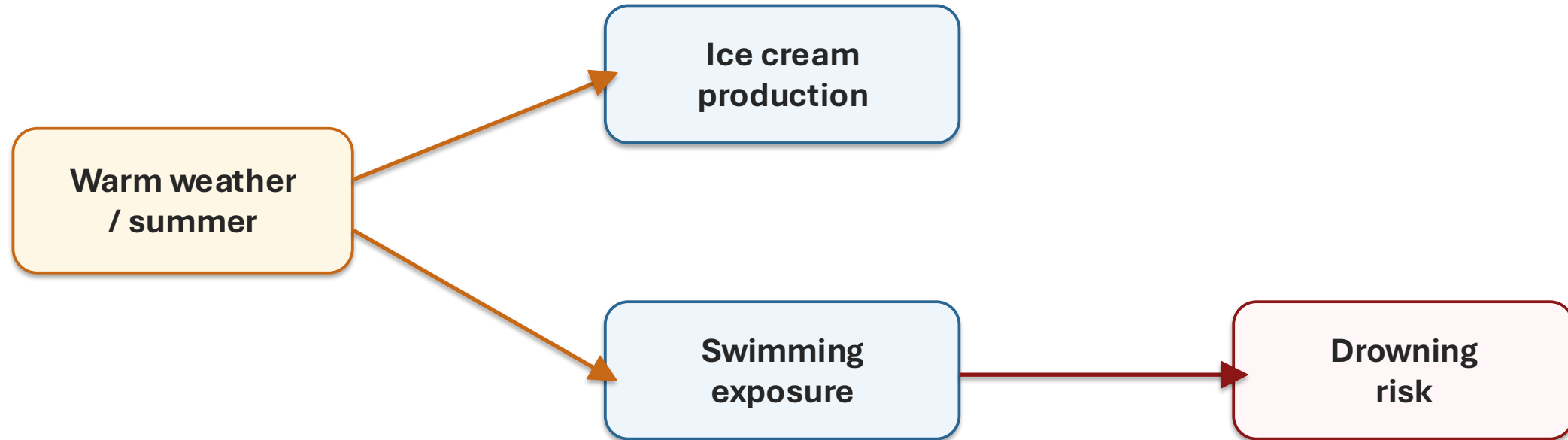
Control for season: the relationship shrinks

Controlling for season removes much of the raw correlation



After removing season-level differences, the correlation drops sharply. The original pattern was mostly about season/temperature, not ice cream itself.

Confounding: a common cause creates the association



A **confounder** is a common cause of the exposure and the outcome. It creates an association that is not the causal effect of X on Y.

Confounding is everywhere

Observed association	Possible confounder	Why it matters
Bigger shoe size → better reading	Age	Older children have bigger feet and read better.
Coffee drinking → heart disease	Smoking, lifestyle	Coffee drinkers may differ in other risk factors.
Selective college → higher earnings	Family background / preparation	Students differ before college begins.

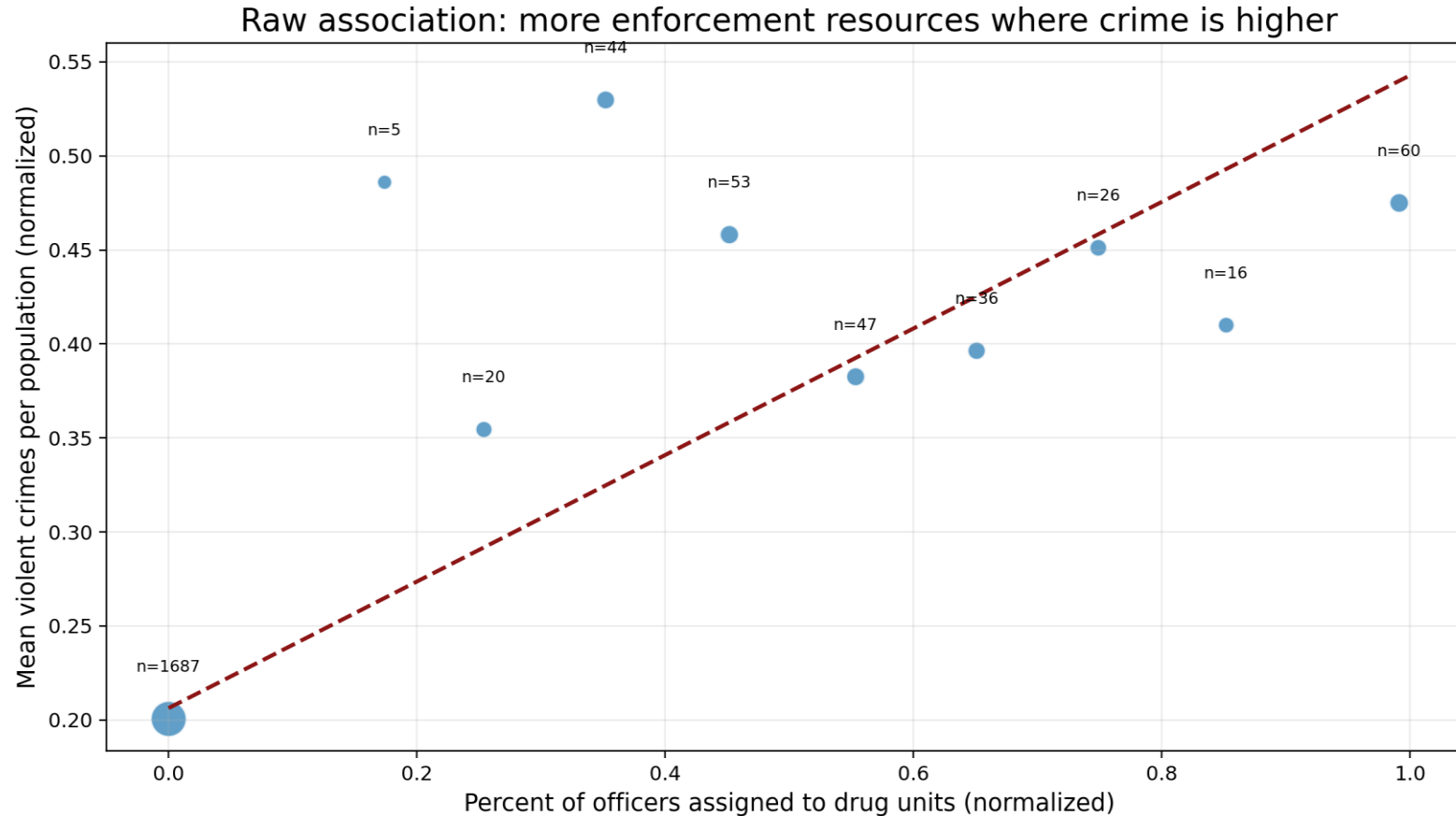
Question for any correlation: what common cause might affect both sides?

Case 2: reverse causality

The direction of the arrow may be the opposite of the naive story.

Case 2: crime and enforcement activity

The positive association does not settle the causal direction.



Communities with more crime also report more enforcement activity.

Does enforcement activity increase crime?

Correlation alone cannot decide which arrow is correct.

Reverse causality: the same correlation can fit different arrows

Naive causal reading



Alternative causal reading



The correlation can be correct in both stories.

Causal inference asks what would happen under an intervention, not just which variables move together.

Reverse causality: more examples

Observed relationship	Naive claim	Alternative direction
More medicine, worse health	Medicine makes people sick	Sicker people take more medicine
More consultants, poor firm performance	Consultants hurt firms	Struggling firms hire consultants
More hospital care, higher mortality	Hospitals harm patients	Sicker patients receive more care

How do we handle reverse causality?

Use timing: cause must precede effect.

Use interventions: what happens if we change X directly?

Use research designs (next lecture): randomization or external shocks can create a direction.

Use DAGs (more on this latter): write down which arrows we believe before interpreting a coefficient.

The data pattern alone is not enough; we need a causal story plus a credible comparison.

Case 3: selection and Simpson's paradox

The comparison groups may have been formed in systematically different ways.

Selection bias is about group formation

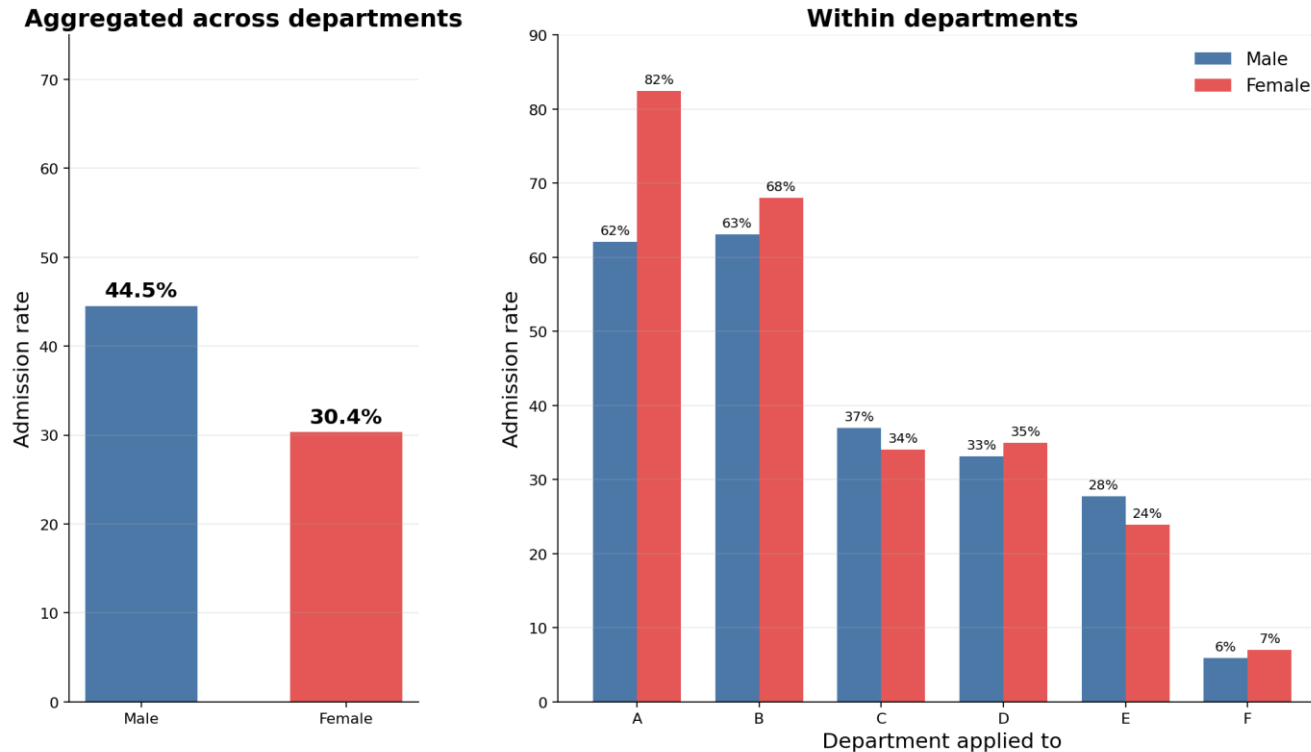
The treated group is often not a random group.

People may select into treatment, or be selected into treatment, because they are already higher-risk, higher-need, or different on background variables.

Example	Who selects into treatment?
Tutoring	Students who are already struggling
Retention emails	Users predicted to be likely to churn
Insurance	People expecting higher medical costs

Simpson's paradox: aggregation can reverse the story

Selection / aggregation can reverse the apparent pattern: UC Berkeley admissions, 1973



The overall pattern partly reflects different application patterns across departments with different admission rates. Source: Bickel, Hammel & O'Connell (1975), UC Berkeley graduate admissions data.

If the overall admission rates differ sharply, does the aggregate comparison alone explain the mechanism?

The aggregate combines groups with different baseline admission rates.

Selection example: tutoring looks harmful overall

Group	Average score
Received tutoring	76
No tutoring	84

Naive causal question: Does tutoring lower scores by 8 points?

But who was most likely to receive tutoring in the first place?

Within comparable groups, the story can reverse

Student group	Tutored	Not tutored	Within-group difference
Struggling students	72	68	+4
Strong students	92	88	+4

Overall, tutored students looked worse because they were more likely to be struggling students. Within each preparation group, tutoring helped.

So... What happened?

Student group	Tutored	Not tutored	Proportion
Struggling students	x	$1-x$	100%
Strong students	y	$1-y$	100%

$$\text{Tutored group: } 72x + 92y = 78$$

$$\text{No tutored group: } 68(1-x) + 88(1-y) = 84$$

$$(x, y) = (0.8, 0.2)$$

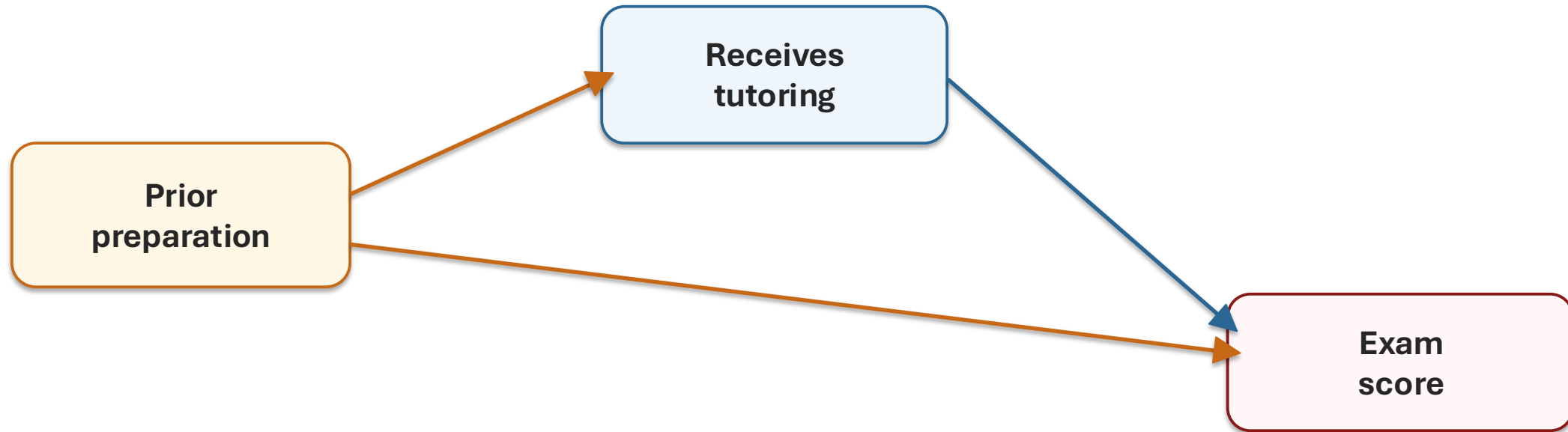
So... What happened?

Student group	Tutored	Not tutored	Proportion
Struggling students	0.8	0.2	100%
Strong students	0.2	0.8	100%

Simple aggregation can be misleading!

Struggling students tend to join the tutoring, which makes tutoring harmful.

Selection bias as a DAG



Selection bias is often a form of confounding. The emphasis is not just “third variable,” but how the treatment group was formed before the comparison.

Three failure modes: summary

Failure mode	Core issue	Question to ask
Confounding	A third variable causes both X and Y	What common cause could explain this?
Reverse causality	The arrow may run from Y to X	Which direction is plausible?
Selection bias	Groups differ before comparison	How were units selected into groups?

Potential outcomes

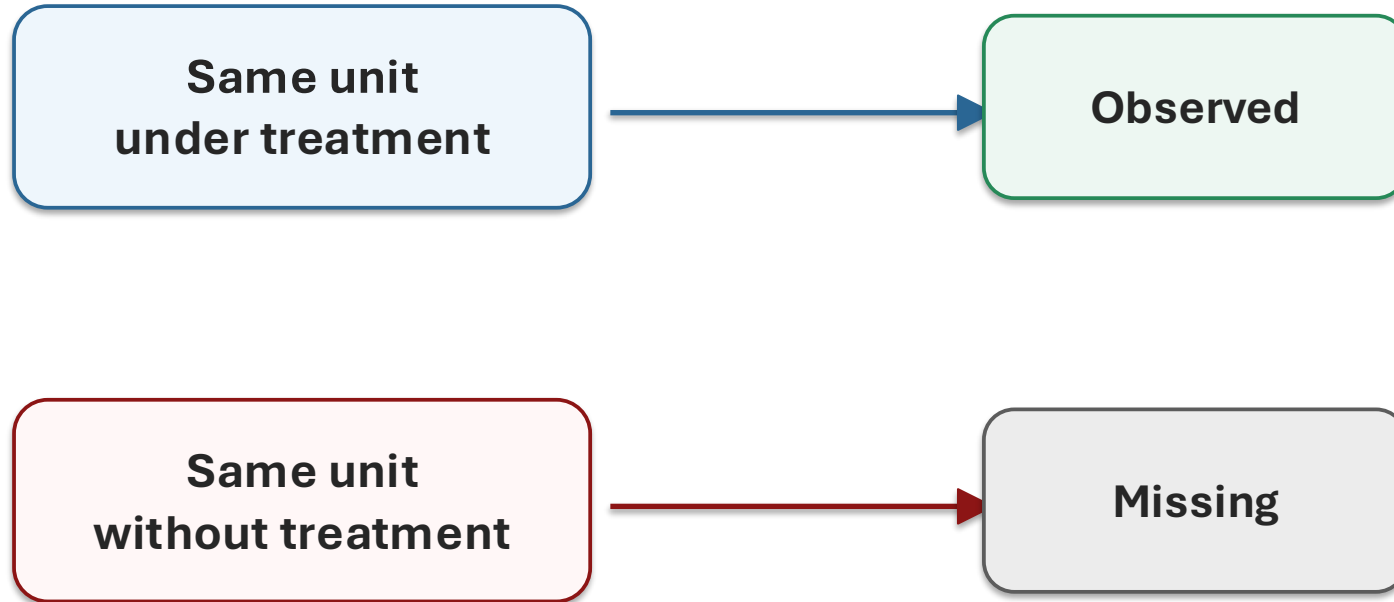
Causal inference is hard because half the data are missing.

Potential outcomes: the missing half of the data

Student	Score if tutored $Y(1)$	Score if not tutored $Y(0)$	Observed status
Maya	82	?	Tutored
Diego	?	75	Not tutored
Aisha	91	?	Tutored
Noah	?	70	Not tutored

Individual causal effect = $Y(1) - Y(0)$

The fundamental problem of causal inference



For each unit, we observe one potential outcome and do not observe the other.

Every causal method tries to construct a credible substitute for the missing outcome.

Raw difference is not automatically a causal effect

Observed difference = causal effect + pre-existing group differences

Question	What we need
Are treated and untreated groups different?	A descriptive comparison
Would treatment change the outcome?	A counterfactual comparison
Are the groups comparable before treatment?	A causal design or assumptions

Four ways to approximate the missing outcome

Strategy	Basic idea	Preview
Randomized experiment	Create comparable groups by random assignment	Lecture 19
Natural experiment	Find external as-if-random variation	Lecture 19
Difference-in-differences	Use a comparison group to subtract background trends	Lecture 19
DAG + adjustment	Use assumptions to compare similar units	Starts today; deepens next lecture

The Stanford premium

The Stanford premium: the question

Suppose Stanford costs substantially more than a state school.
Stanford graduates have high earnings.

However, is the earnings difference caused by Stanford?

Prediction version

Which schools have high-earning graduates?

Uses observed patterns across schools.

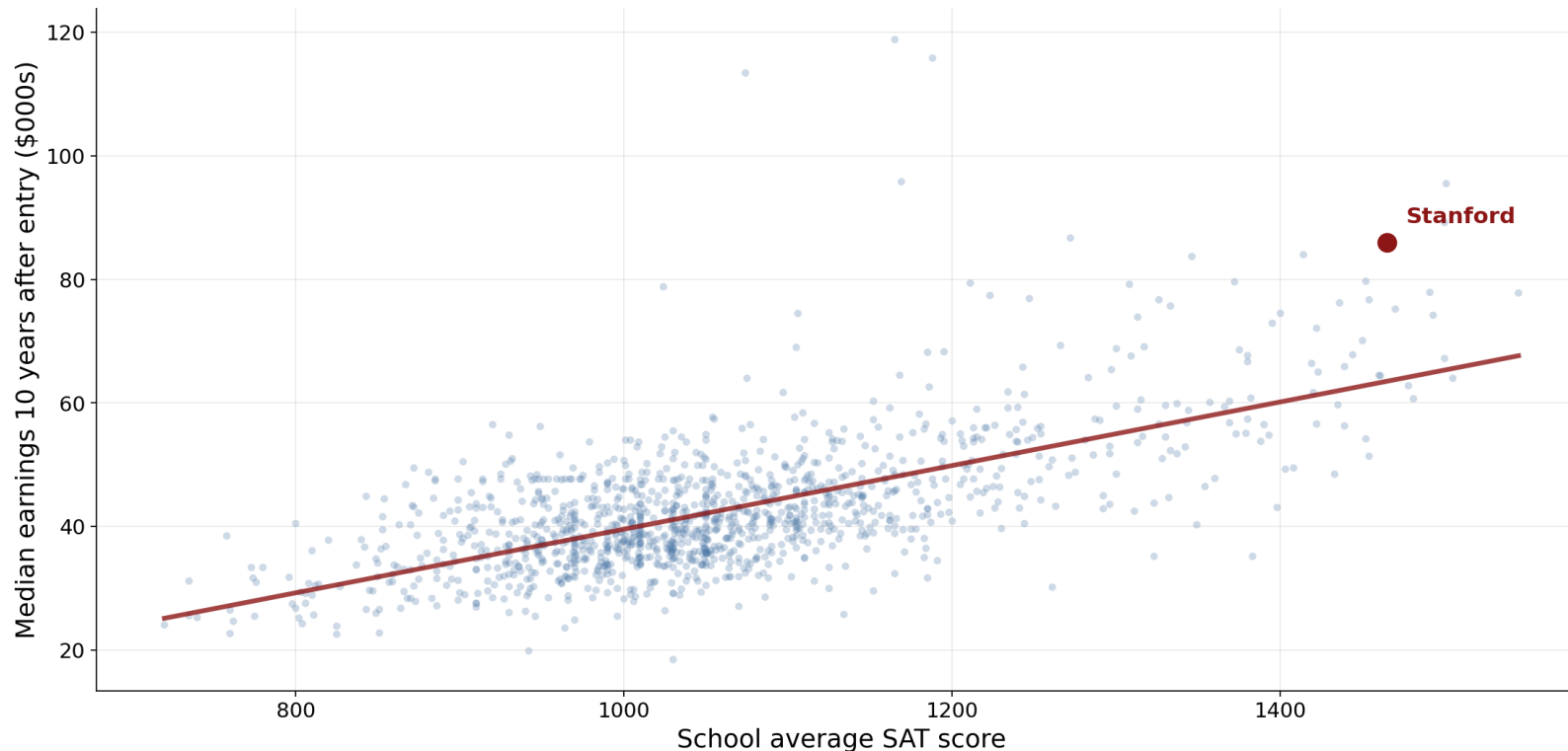
Causal version

Would this student earn more if they attended Stanford?

Requires a missing counterfactual for the same student.

Stanford premium: the association is real

College Scorecard: selective schools have higher-earning graduates ($r = 0.65$)



Source: U.S. Department of Education College Scorecard, course scorecard.csv. Each point is a school.

From a prediction perspective, the pattern is informative.

But the plot compares different schools with different students.

The graph does not show what the same student would have earned elsewhere.

The Stanford counterfactual

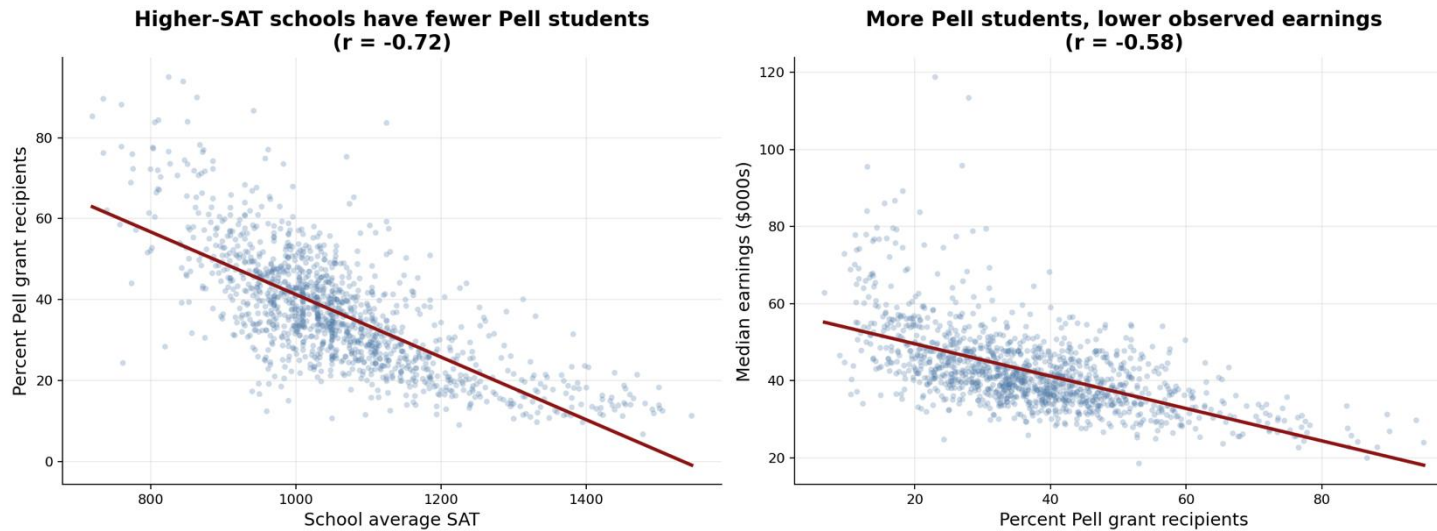
Student	Earnings if attends Stanford $Y(1)$	Earnings if attends state school $Y(0)$	Observed?
Student A	?	\$80K	State school
Student B	\$130K	?	Stanford
Student C	\$105K	?	Stanford

The causal question compares the **same student** under two possible college choices.

The observed graph compares **different students** at different schools.

Why the counterfactual is hard

A plausible confounding story in the Stanford premium example



Source: U.S. Department of Education College Scorecard, course scorecard.csv. Pell share is a rough proxy for socioeconomic composition.

Selective colleges do not enroll a random sample of students.

Students differ before college: preparation, family resources, geography, networks, intended majors.

Any comparison of earnings mixes school effects with pre-existing differences.

DAGs: first graphical language

What is a DAG?

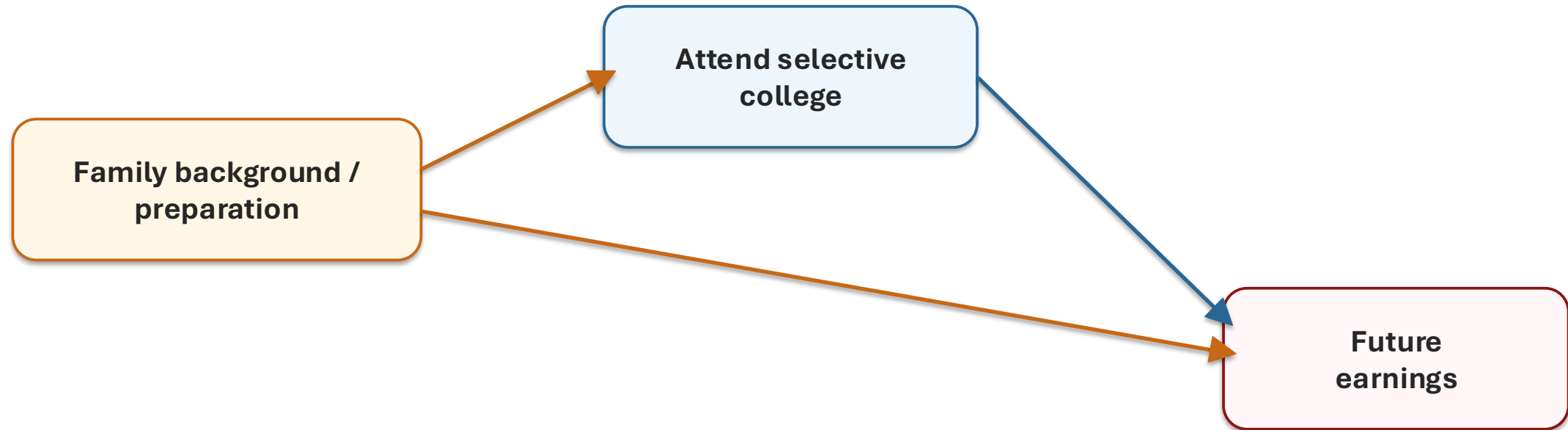
Term	Meaning
Directed	Arrows have direction: A causes B.
Acyclic	No causal loops inside the graph.
Graph	Variables are nodes; arrows are assumed direct causal links.

A DAG is a causal story drawn with arrows. It is a statement of assumptions, not something the data automatically gives us.

How to draw a DAG: a practical recipe

Step	What to do
1. Start with the causal question	Identify treatment X and outcome Y.
2. List variables that occur before treatment	These can confound the comparison.
3. Add arrows for causal influence	Do not draw arrows just because variables are correlated.
4. Mark variables after treatment	These may be mediators or colliders.
5. Ask what comparison the DAG implies	Which groups are comparable?

DAG example: Stanford and earnings



The DAG separates the causal path we care about from the noncausal path through background differences.

What a DAG can and cannot do

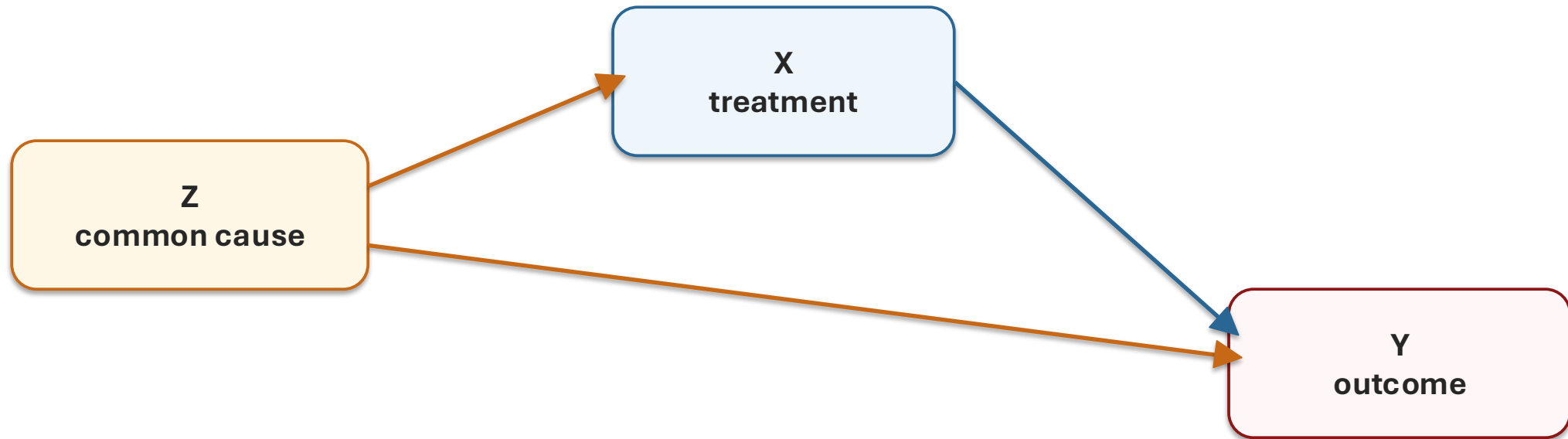
DAGs can help us...	DAGs cannot...
Make assumptions explicit.	Prove the assumptions are true.
Separate causal paths from spurious paths.	Automatically estimate effects.
Discuss what to adjust for.	Save a bad research design.
Communicate the comparison we want.	Replace domain knowledge.

Three DAG motifs to recognize

Motif	Shape	First intuition
Confounder	$X \leftarrow Z \rightarrow Y$	Z makes X and Y move together.
Mediator	$X \rightarrow M \rightarrow Y$	M is part of how X affects Y.
Collider	$X \rightarrow C \leftarrow Y$	Conditioning on C can create a fake association.

Today: recognize the patterns. Lecture 19: use them to decide what to control for.

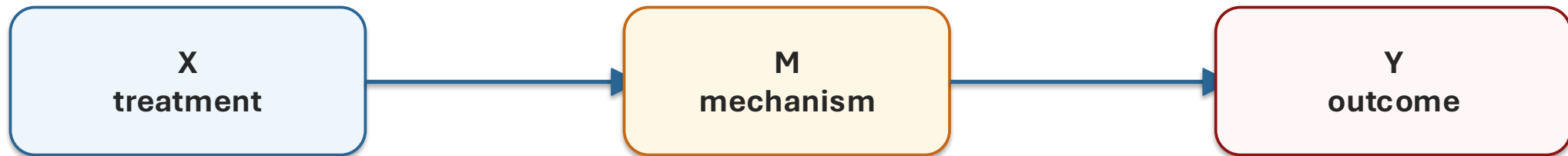
Motif 1: confounder



A confounder Z is a common cause for both treatment X and outcome Y.

Example: preparation affects both tutoring and exam scores. This is the pattern behind many misleading treatment comparisons.

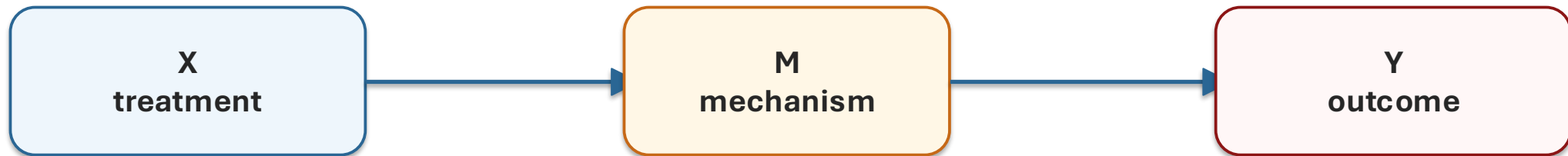
Motif 2: mediator



A mediator X captures how the treatment X effect outcome Y.

Example: college may affect alumni network, which may affect earnings. A mediator is part of the causal pathway, not a pre-treatment common cause.

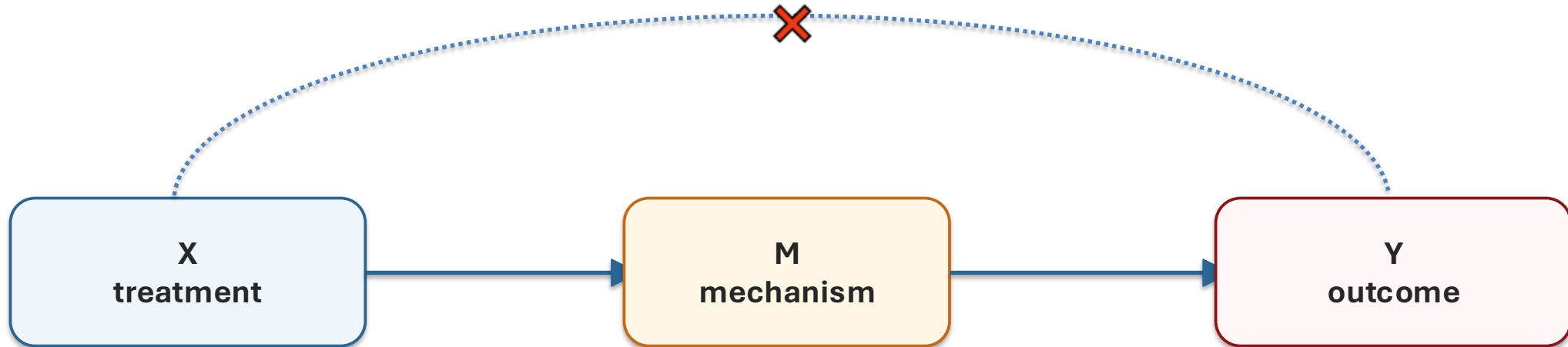
Motif 2: mediator



A mediator X captures how the treatment X effects outcome Y .

Example: college may affect alumni network, which may affect earnings. A mediator is part of the causal pathway, not a pre-treatment common cause.

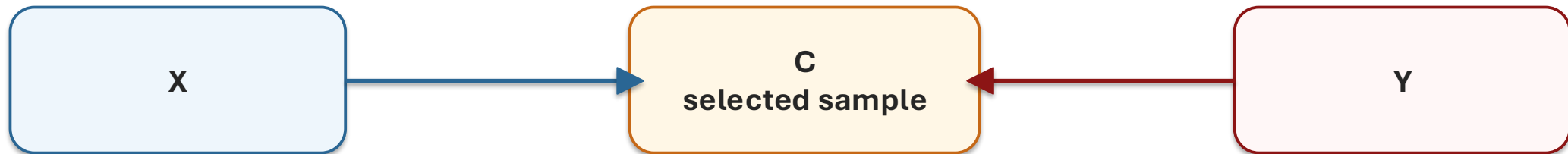
Motif 2: mediator



A mediator X captures how the treatment X affects outcome Y.

Example: college may affect alumni network, which may affect earnings. A mediator is part of the causal pathway, not a pre-treatment common cause.

Motif 3: collider



A collider measures the common effect of X and Y.

Example: if we only look at people hired by a top firm, talent and connections may become related inside that selected group even if they were unrelated in the full population.

More about these in next lecture

Which variables should we control for?

Why can controlling for a mediator change the causal question?

Why can conditioning on a collider create bias?

How do DAGs connect to RCTs, natural experiments, and difference-in-differences?

Preview: how can we make causal comparisons credible?

Strategy	What it tries to do
Randomized experiment / A-B test	Create comparable groups by random assignment.
Natural experiment	Use external variation that is plausibly as-if random.
Difference-in-differences	Use a comparison group to remove background trends.
DAGs + adjustment	Make assumptions explicit and compare similar units.

Lecture 18 takeaways

Prediction and causation answer different questions.

Correlation can be real but causally misleading.

Three common failures: confounding, reverse causality, and selection/aggregation.

The fundamental problem: for each unit, one potential outcome is missing.

DAGs are a compact language for causal assumptions.

Causal inference is about constructing a credible comparison to the missing counterfactual.

More thinkings:

Give one example where a correlation is real but the causal interpretation is wrong.

Explain the “missing half of the data” in one sentence.

Draw a simple DAG for the tutoring example.

What would make the Stanford premium question easier to answer causally?