Causal origins of statistical associations

Causal origins of statistical associations

There are two reasons A and Y can be associated

Causal origins of statistical associations

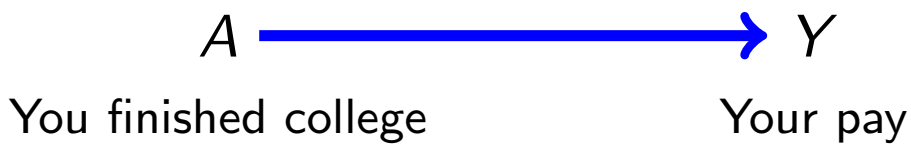
There are two reasons A and Y can be associated

- A **causal path**: $A \to Y$

Causal origins of statistical associations

There are two reasons A and Y can be associated

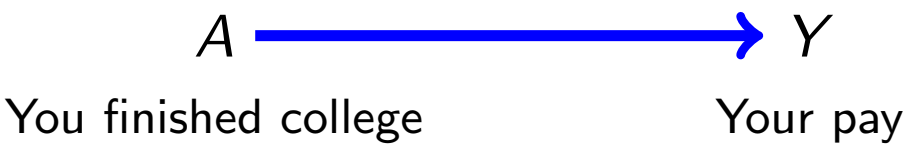
- A **causal path**: $A \rightarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

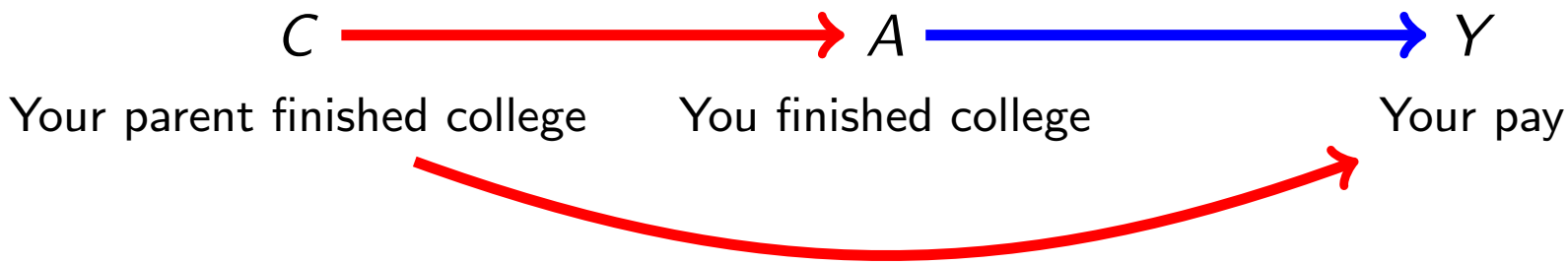
- A **causal path**: $A \to Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \to Y$
  - or blocked colliders $A \to \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

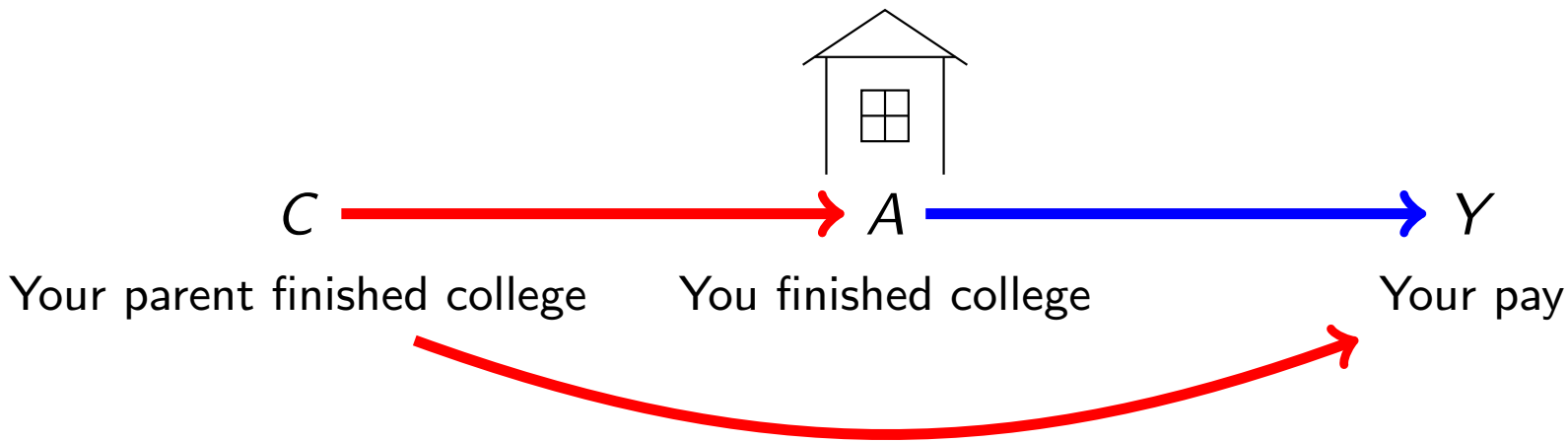
- A **causal path**: $A \rightarrow Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \rightarrow Y$
  - or blocked colliders $A \rightarrow \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

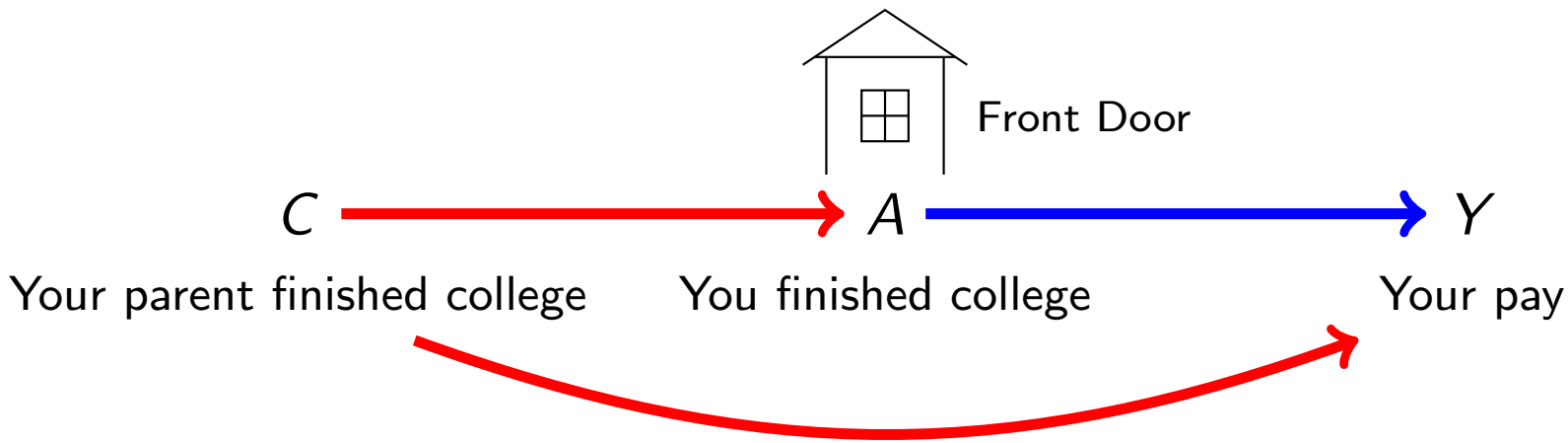
- A **causal path**: $A \rightarrow Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \rightarrow Y$
  - or blocked colliders $A \rightarrow \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

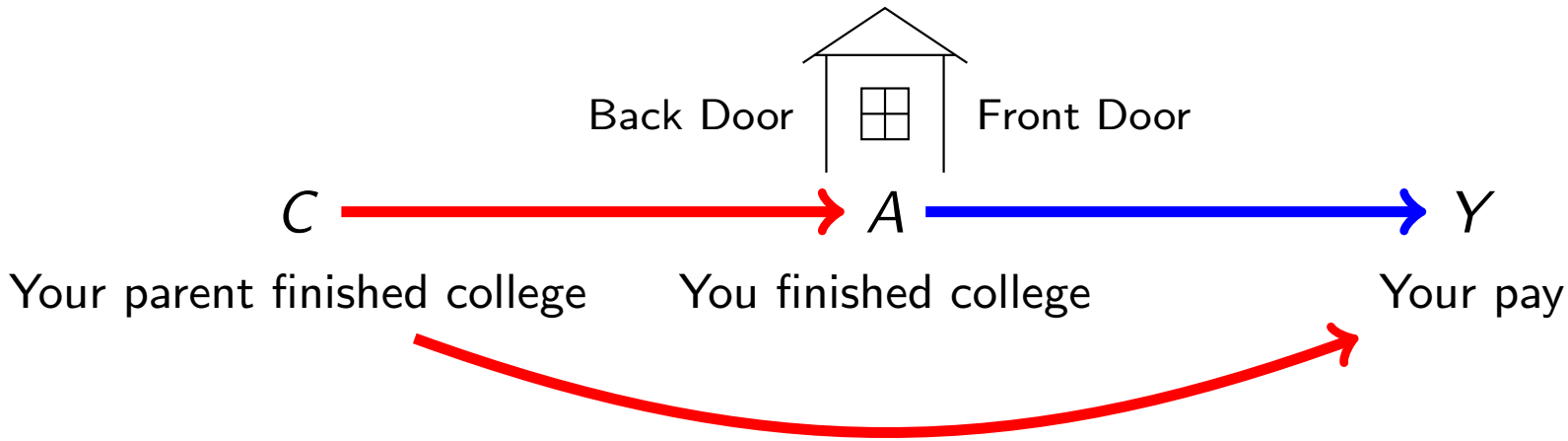
- A **causal path**: $A \rightarrow Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \rightarrow Y$
  - or blocked colliders $A \rightarrow \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

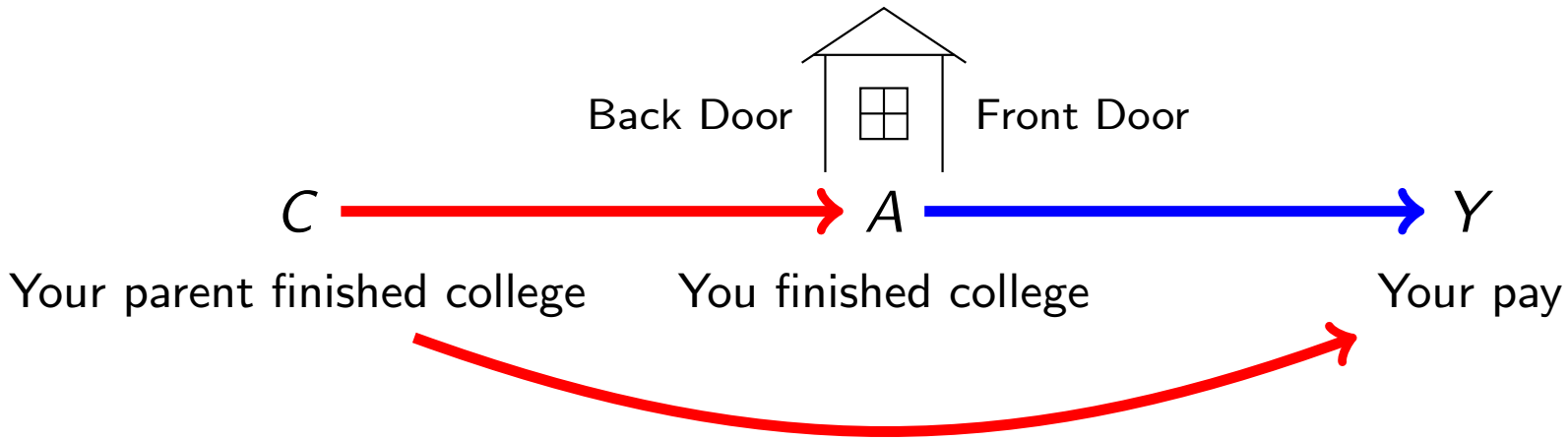
- A **causal path**: $A \rightarrow Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \rightarrow Y$
  - or blocked colliders $A \rightarrow \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

- A **causal path**: $A \rightarrow Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \rightarrow Y$
  - or blocked colliders $A \rightarrow \boxed{C} \leftarrow Y$

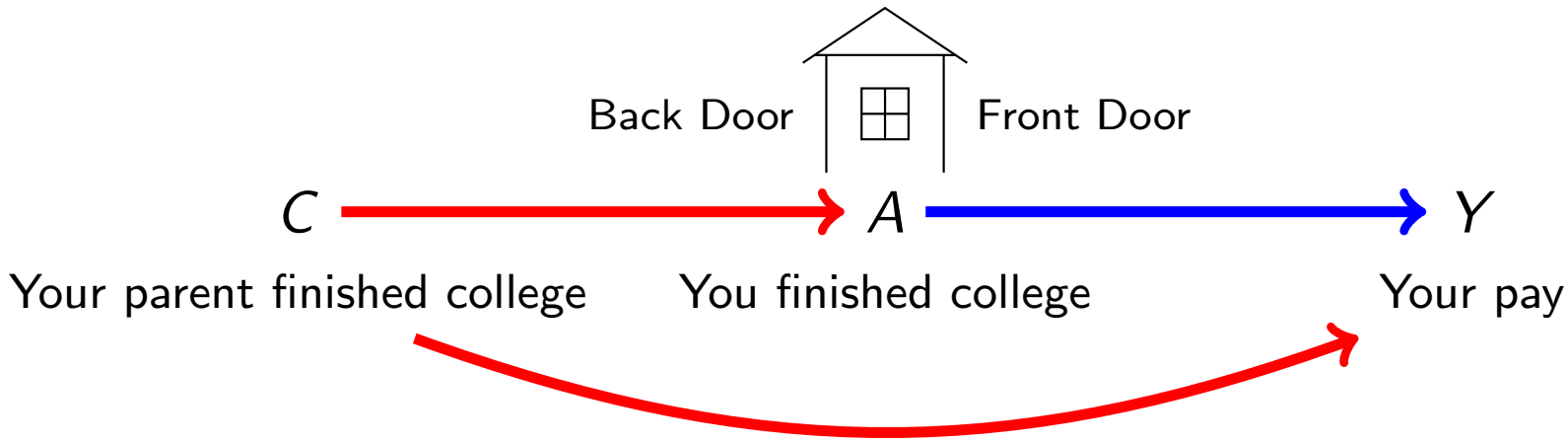


To block this backdoor path, condition on C (a **confounder**)

Causal origins of statistical associations

There are two reasons A and Y can be associated

- A **causal path**: $A \rightarrow Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \rightarrow Y$
  - or blocked colliders $A \rightarrow \boxed{C} \leftarrow Y$



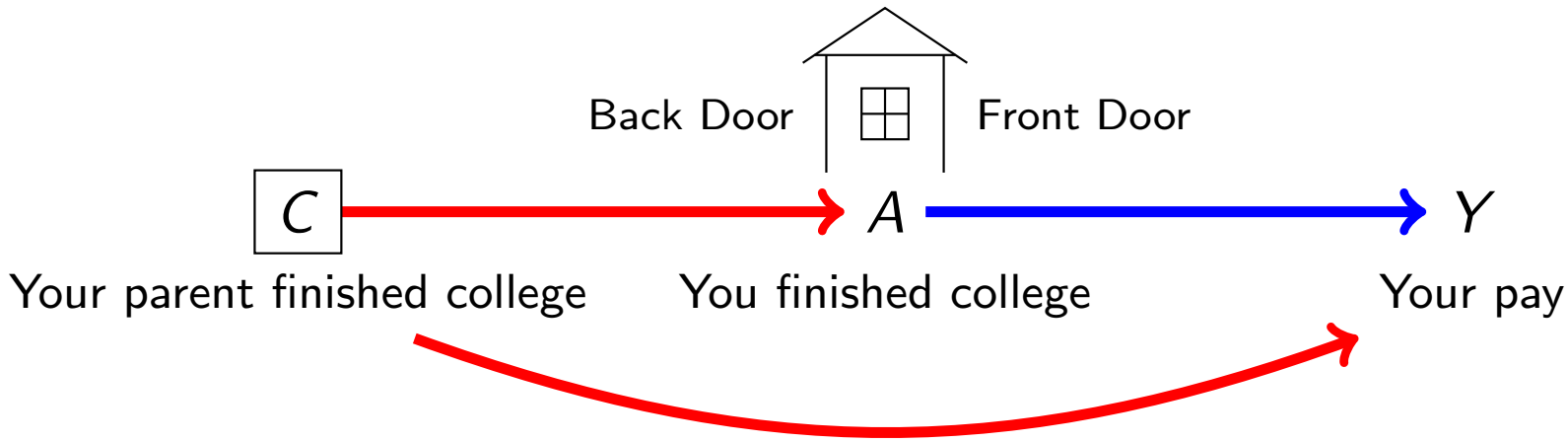
To block this backdoor path, condition on C (a **confounder**)

- Analyze within subgroups defined by C

Causal origins of statistical associations

There are two reasons A and Y can be associated

- A **causal path**: $A \rightarrow Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \rightarrow Y$
  - or blocked colliders $A \rightarrow \boxed{C} \leftarrow Y$



To block this backdoor path, condition on C (a **confounder**)

- Analyze within subgroups defined by C

Colliders[1]

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

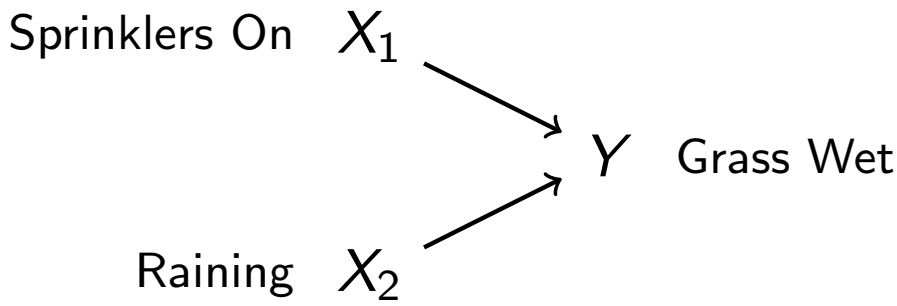
Colliders[1]

Suppose I have sprinklers on a timer.

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

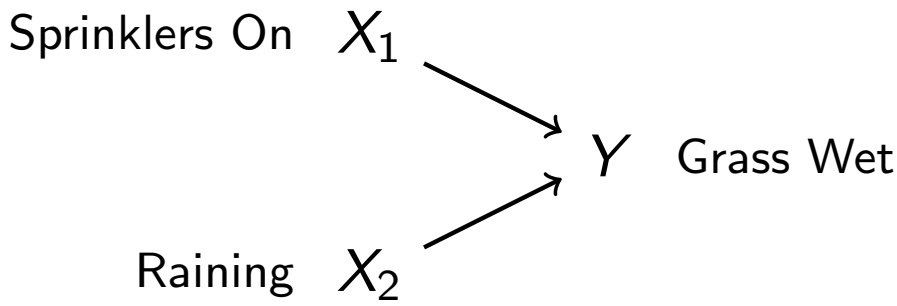
Suppose I have sprinklers on a timer.



[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

Suppose I have sprinklers on a timer.

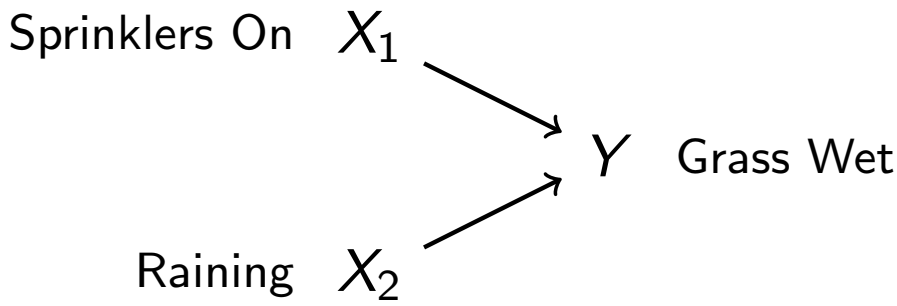


We say Y is a **collider** along the path $X_1 \rightarrow Y \leftarrow X_2$

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

Suppose I have sprinklers on a timer.



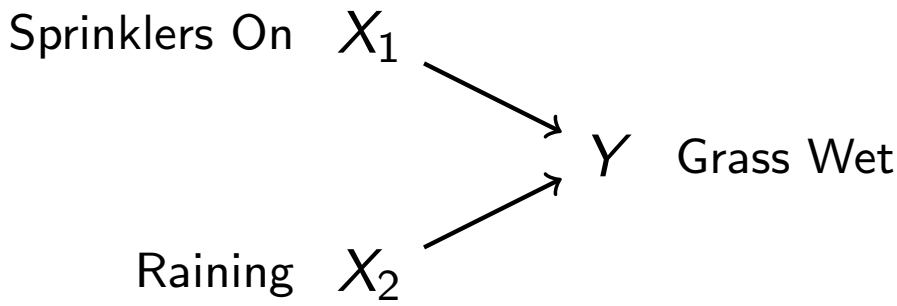
We say Y is a **collider** along the path $X_1 \rightarrow Y \leftarrow X_2$

- The collider blocks the path

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

Suppose I have sprinklers on a timer.



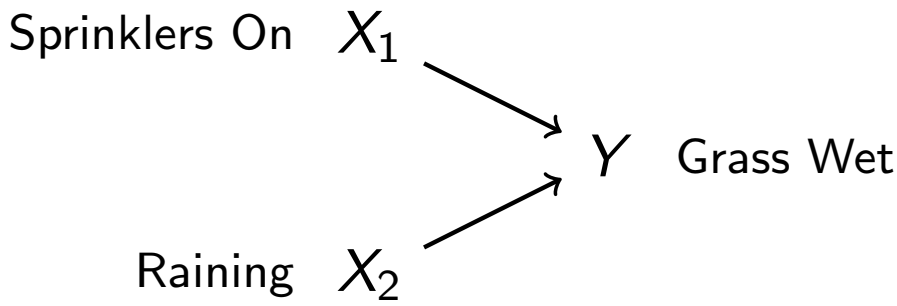
We say Y is a **collider** along the path $X_1 \rightarrow Y \leftarrow X_2$

- The collider blocks the path
- X_1 is independent of X_2
  - (Sprinklers On) is uninformative about (Raining)

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

Suppose I have sprinklers on a timer.



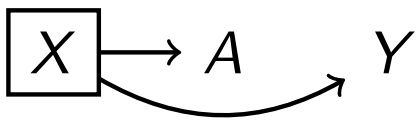
We say Y is a **collider** along the path $X_1 \to Y \leftarrow X_2$

- The collider blocks the path
- X_1 is independent of X_2
  - (Sprinklers On) is uninformative about (Raining)
- Conditioning on Y opens the path
  - If the grass is wet (conditional on $Y = 1$), then either (Sprinklers On) or (Raining)

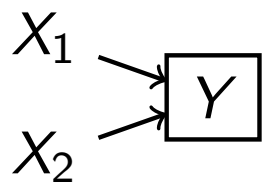
[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Ancestors vs. Colliders

Conditioning on an ancestor **closes** an open path

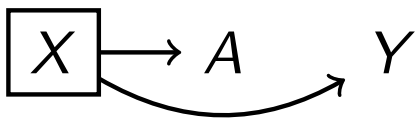


Conditioning on an collider **opens** a closed path



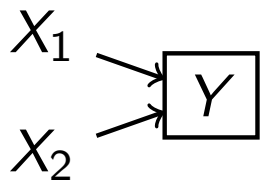
Ancestors vs. Colliders

Conditioning on an ancestor **closes** an open path



Example
— X is your parent's education
— A is your education
— Y is your pay

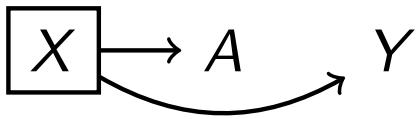
Conditioning on an collider **opens** a closed path



Example
— X_1 is sprinklers on
— X_2 is rain
— Y is wet grass

Ancestors vs. Colliders

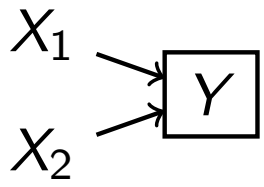
Conditioning on an ancestor **closes** an open path



Example
— X is your parent's education
— A is your education
— Y is your pay

In the population,
A and Y are **related**

Conditioning on an collider **opens** a closed path

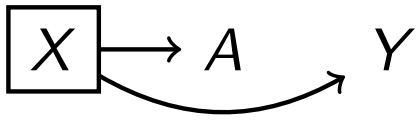


Example
— X_1 is sprinklers on
— X_2 is rain
— Y is wet grass

In the population,
X_1 and X_2 are **independent**

Ancestors vs. Colliders

Conditioning on an ancestor **closes** an open path

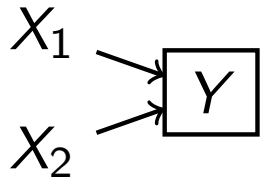


Example
— X is your parent's education
— A is your education
— Y is your pay

In the population,
A and Y are **related**

Within strata of X,
A and Y are **independent**

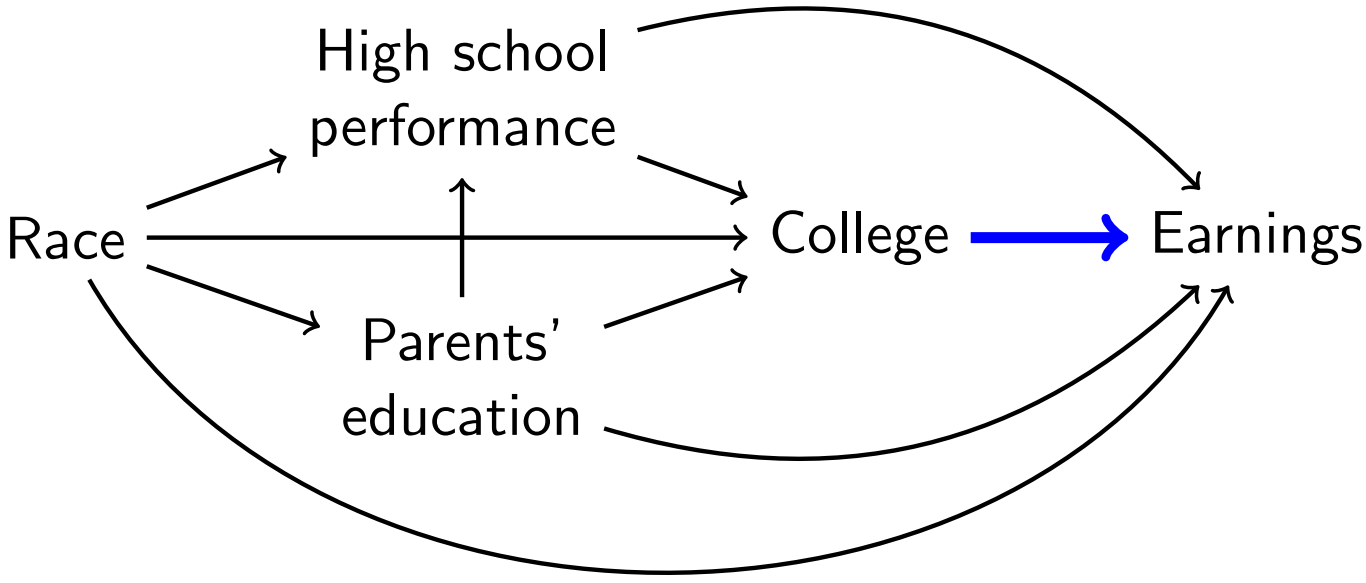
Conditioning on an collider **opens** a closed path



Example
— X_1 is sprinklers on
— X_2 is rain
— Y is wet grass

In the population,
X_1 and X_2 are **independent**

Within strata of Y,
X_1 and X_2 are **related**


High school
performance
Race
College
Earnings
Parents'
education

How to find adjustment variables to identify causal effects

Goal:

Block all backdoor paths so treatment A and outcome Y are associated only by the causal path

How to find adjustment variables to identify causal effects

Goal:
Block all backdoor paths so treatment A and outcome Y are associated only by the causal path

Backdoor path: Any sequence of edges $A \leftarrow \texttt{nodes} \rightarrow Y$

Blocked if it contains an adjusted variable along a fork

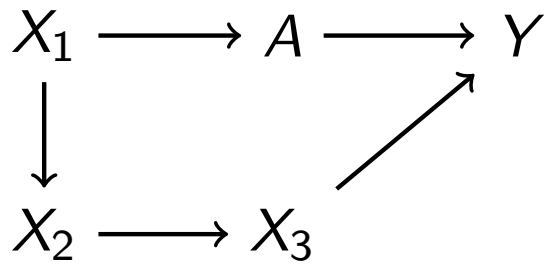
$$A \leftarrow \boxed{C} \rightarrow Y$$
$$A \leftarrow \boxed{C} \leftarrow \cdots \rightarrow Y$$
$$A \leftarrow \cdots \rightarrow \boxed{C} \rightarrow Y$$

Blocked if it contains an unadjusted collider

$$A \rightarrow C \leftarrow Y$$

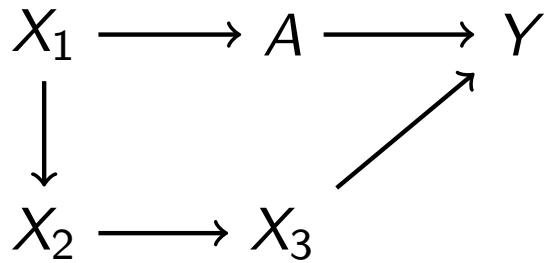
Exercise 1

Find adjustment sets that identify the effect of A on Y



Exercise 1

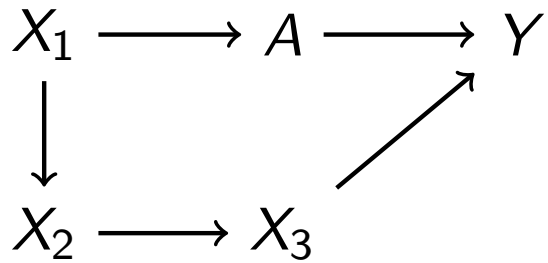
Find adjustment sets that identify the effect of A on Y



We can block the backdoor path in several ways:

Exercise 1

Find adjustment sets that identify the effect of A on Y

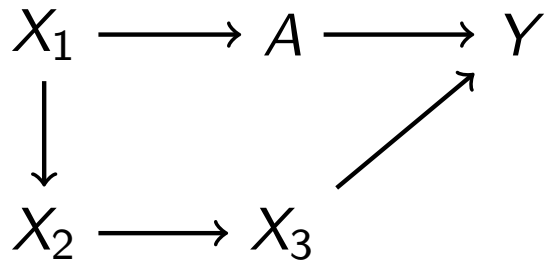


We can block the backdoor path in several ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$

Exercise 1

Find adjustment sets that identify the effect of A on Y

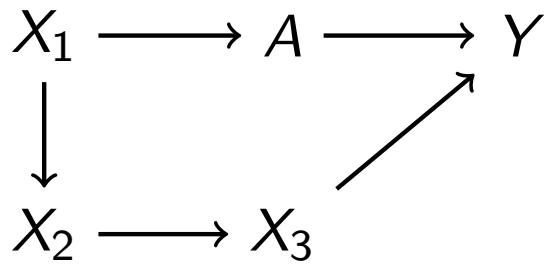


We can block the backdoor path in several ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$
- Condition on X_2: $A \leftarrow X_1 \rightarrow \boxed{X_2} \rightarrow X_3 \rightarrow Y$

Exercise 1

Find adjustment sets that identify the effect of A on Y

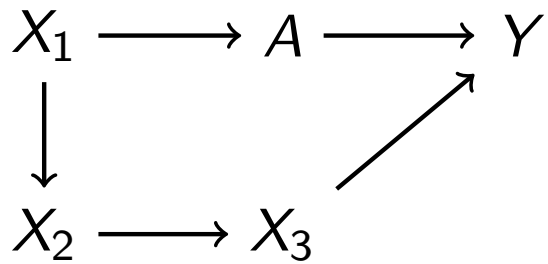


We can block the backdoor path in several ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$
- Condition on X_2: $A \leftarrow X_1 \rightarrow \boxed{X_2} \rightarrow X_3 \rightarrow Y$
- Condition on X_3: $A \leftarrow X_1 \rightarrow X_2 \rightarrow \boxed{X_3} \rightarrow Y$

Exercise 1

Find adjustment sets that identify the effect of A on Y

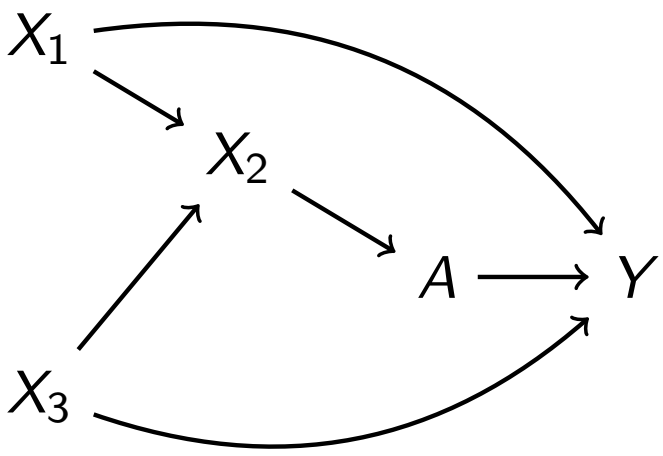


We can block the backdoor path in several ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$
- Condition on X_2: $A \leftarrow X_1 \rightarrow \boxed{X_2} \rightarrow X_3 \rightarrow Y$
- Condition on X_3: $A \leftarrow X_1 \rightarrow X_2 \rightarrow \boxed{X_3} \rightarrow Y$
- Any combination of the above

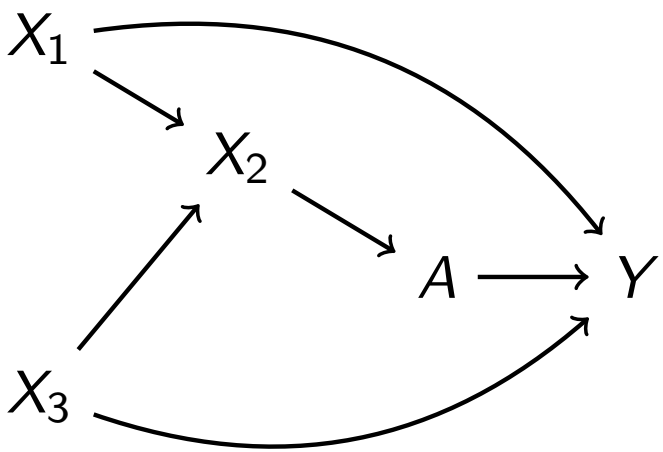
Exercise 2

Find 3 sufficient adjustment sets to identify $A \rightarrow Y$



Exercise 2

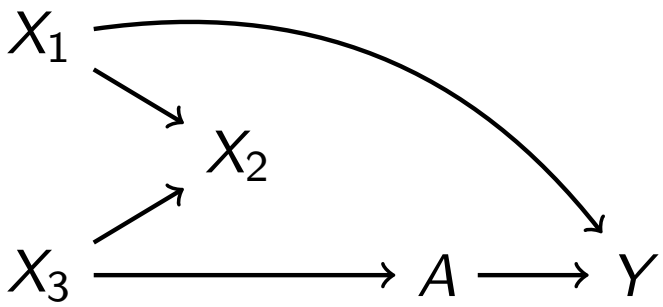
Find 3 sufficient adjustment sets to identify $A \to Y$



Answer: $\{X_2\}, \{X_1, X_3\}, \{X_1, X_2, X_3\}$

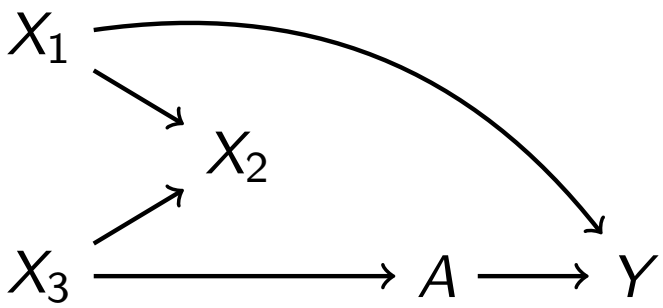
Exercise 3

What is the smallest adjustment set that identifies $A \to Y$?



Exercise 3

What is the smallest adjustment set that identifies $A \to Y$?



Answer: The empty set! Don't condition on anything.
The collider X_2 already blocks the path.

Learning goals for today

By the end of class, you will be able to

- Formalize causal assumptions in Directed Acyclic Graphs (DAGs)
- Use DAGs to find a sufficient adjustment set of variables within which a statistical association is causal