Lab 2: DAGs

Olivier Bergeron-Bouutin

January 26th, 2021

Introductory meme

When your code is a mess but it somehow still works.

Why use DAGs?

A short history

DAGs are *not* widely used, but increasingly popular

A short history

DAGs are *not* widely used, but increasingly popular

- You'll essentially never see them in papers, though it could arguably be useful

A short history

DAGs are *not* widely used, but increasingly popular

- You'll essentially never see them in papers, though it could arguably be useful

Most closely associated with Judea Pearl

A short history

DAGs are *not* widely used, but increasingly popular

- You'll essentially never see them in papers, though it could arguably be useful

Most closely associated with Judea Pearl

Especially useful for causal inference with observational data

A short history

DAGs are *not* widely used, but increasingly popular

- You'll essentially never see them in papers, though it could arguably be useful

Most closely associated with Judea Pearl
Especially useful for causal inference with observational data
Several reasons for that:

- A compact way to represent causal relationships

A short history

DAGs are *not* widely used, but increasingly popular

- You'll essentially never see them in papers, though it could arguably be useful

Most closely associated with Judea Pearl

Especially useful for causal inference with observational data

Several reasons for that:

- A compact way to represent causal relationships
- They clarify causal ordering

A short history

DAGs are *not* widely used, but increasingly popular

- You'll essentially never see them in papers, though it could arguably be useful

Most closely associated with Judea Pearl

Especially useful for causal inference with observational data

Several reasons for that:

- A compact way to represent causal relationships
- They clarify causal ordering
- Help clarify research designs/estimators, e.g. instrumental variables

A short history

DAGs are *not* widely used, but increasingly popular

- You'll essentially never see them in papers, though it could arguably be useful

Most closely associated with Judea Pearl

Especially useful for causal inference with observational data

Several reasons for that:

- A compact way to represent causal relationships
- They clarify causal ordering
- Help clarify research designs/estimators, e.g. instrumental variables
- They force you to think about the assumptions you're making

A short history

DAGs are *not* widely used, but increasingly popular

- You'll essentially never see them in papers, though it could arguably be useful

Most closely associated with Judea Pearl

Especially useful for causal inference with observational data

Several reasons for that:

- A compact way to represent causal relationships
- They clarify causal ordering
- Help clarify research designs/estimators, e.g. instrumental variables
- They force you to think about the assumptions you're making

The famous 3rd variable?

Think about the famous problem of “third variables”

The famous 3rd variable?

Think about the famous problem of “third variables”

- How would you describe this problem?

The famous 3rd variable?

Think about the famous problem of “third variables”

- How would you describe this problem?
- You’ll see there are different types

The famous 3rd variable?

Think about the famous problem of “third variables”

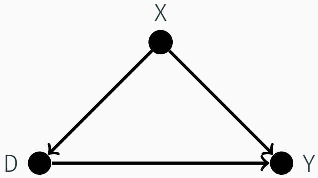
- How would you describe this problem?
- You’ll see there are different types
- And your empirical strategy depends on the type

Anatomy of a DAG

DAGs use a set of **nodes** and directed **edges**

Anatomy of a DAG

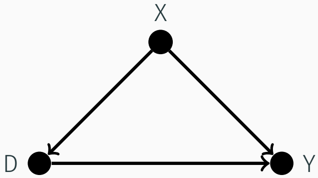
DAGs use a set of **nodes** and directed **edges**



- The nodes  represent random variables

Anatomy of a DAG

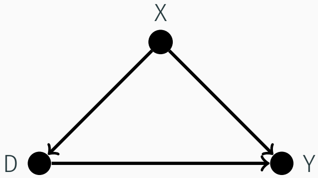
DAGs use a set of **nodes** and directed **edges**



- The nodes ● represent random variables
- The directed edges represent potential direct causal effects

Anatomy of a DAG

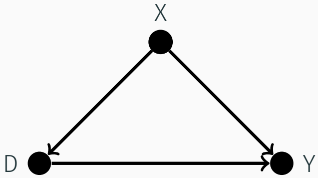
DAGs use a set of **nodes** and directed **edges**



- The nodes ● represent random variables
- The directed edges represent potential direct causal effects
- Here, we have some treatment D that affects some outcome Y

Anatomy of a DAG

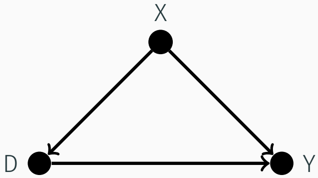
DAGs use a set of **nodes** and directed **edges**



- The nodes ● represent random variables
- The directed edges represent potential direct causal effects
- Here, we have some treatment D that affects some outcome Y
- We also have a random variable X that affects both D and Y

Anatomy of a DAG

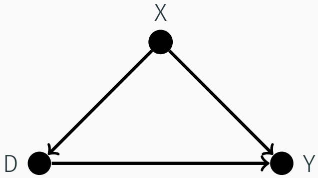
DAGs use a set of **nodes** and directed **edges**



- The nodes  represent random variables
- The directed edges represent potential direct causal effects
- Here, we have some treatment D that affects some outcome Y
- We also have a random variable X that affects both D and Y
- We have a **backdoor path** between D and Y

Anatomy of a DAG

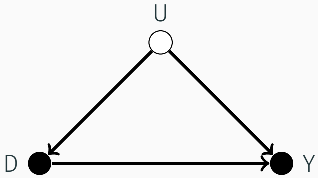
DAGs use a set of **nodes** and directed **edges**



- The nodes represent random variables
- The directed edges represent potential direct causal effects
- Here, we have some treatment D that affects some outcome Y
- We also have a random variable X that affects both D and Y
- We have a **backdoor path** between D and Y
- What does that mean? What do we need to do?

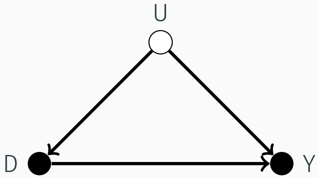
Types of third variables

Confounder



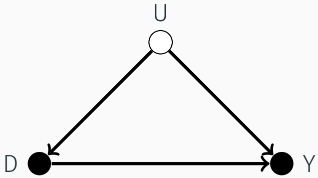
- Here our node $\bigcirc$ is white and called U
- It's unobserved...

Confounder



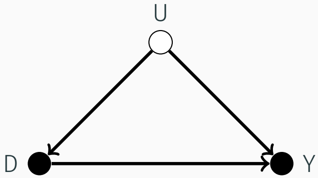
- Here our node ◯ is white and called U
- It's unobserved...
- either because the data has not been collected

Confounder



- Here our node ◯ is white and called U
- It's unobserved…
- either because the data has not been collected
- or because it's fundamentally unobservable (examples?)

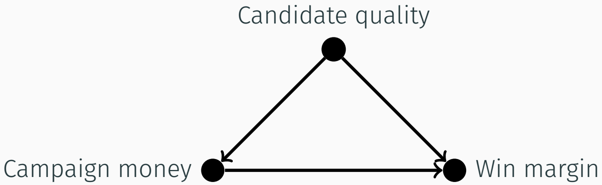
Confounder



- Here our node ◯ is white and called U
- It's unobserved…
- either because the data has not been collected
- or because it's fundamentally unobservable (examples?)
- Is the causal effect of D on Y identified?

Simple example

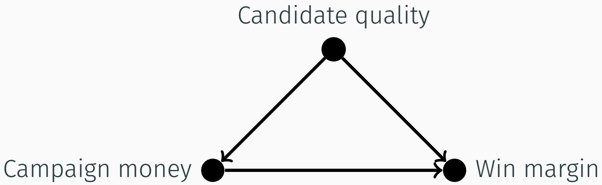
From Andrew Heiss:



There's an open backdoor path: Money ←Quality →Margin

Simple example

From Andrew Heiss:

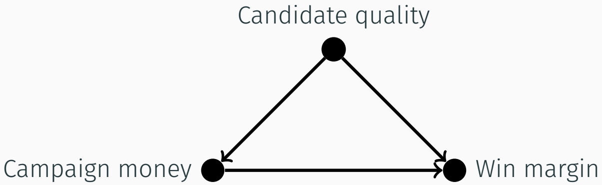


There's an open backdoor path: Money ←Quality →Margin

- To retrive the causal effect Money →Margin, I need to adjust/condition/control for Candidate Quality

Simple example

From Andrew Heiss:

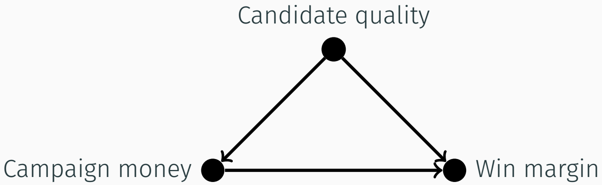


There's an open backdoor path: Money ←Quality →Margin

- To retrive the causal effect Money →Margin, I need to adjust/condition/control for Candidate Quality
- In other words: to block the backdoor path

Simple example

From Andrew Heiss:

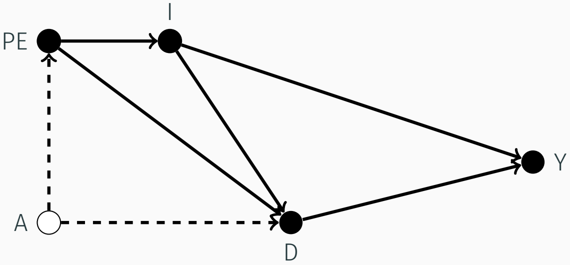


There's an open backdoor path: Money ←Quality →Margin

- To retrive the causal effect Money →Margin, I need to adjust/condition/control for Candidate Quality
- In other words: to block the backdoor path
- Matching, regression...

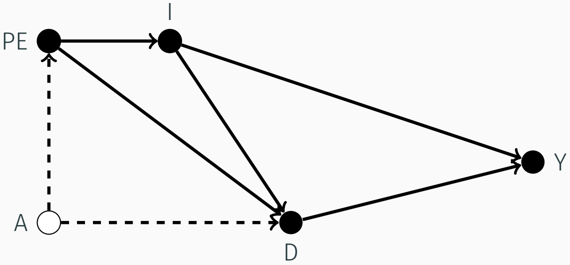
A more complex example

Drawn from the *Causal Inference Mixtape*:



A more complex example

Drawn from the *Causal Inference Mixtape*:



D: education

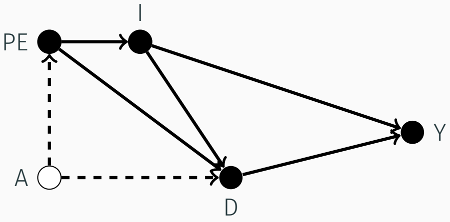
Y: wages

I: Parental income

PE: Parental education

A: (unobserved) ability

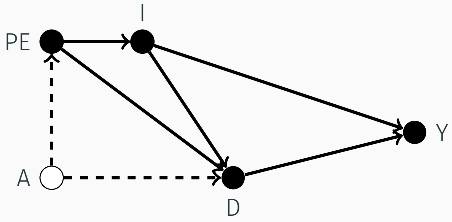
Paths in this DAG



We need to list all direct and indirect (backdoor) paths between D and Y

$D \rightarrow Y$ (causal effect of education on wages)

Paths in this DAG

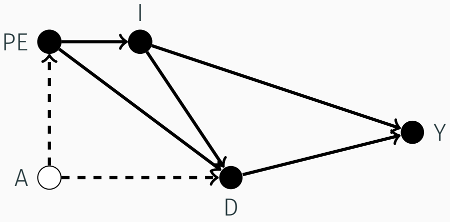


We need to list all direct and indirect (backdoor) paths between D and Y

$D \rightarrow Y$ (causal effect of education on wages)

$D \leftarrow I \rightarrow Y$ (backdoor path 1)

Paths in this DAG



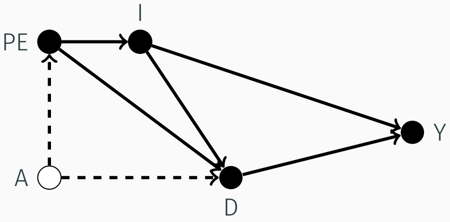
We need to list all direct and indirect (backdoor) paths between D and Y

$D \rightarrow Y$ (causal effect of education on wages)

$D \leftarrow I \rightarrow Y$ (backdoor path 1)

$D \leftarrow PE \rightarrow I \rightarrow Y$ (backdoor path 2)

Paths in this DAG



We need to list all direct and indirect (backdoor) paths between D and Y

$D \rightarrow Y$ (causal effect of education on wages)

$D \leftarrow I \rightarrow Y$ (backdoor path 1)

$D \leftarrow PE \rightarrow I \rightarrow Y$ (backdoor path 2)

$D \leftarrow A \rightarrow PE \rightarrow I \rightarrow Y$ (backdoor path 3)

Closing backdoor paths

Our goal is to close all backdoor paths

Closing backdoor paths

Our goal is to close all backdoor paths

- Such that we're left with the only path we care about: the direct causal effect of D on Y

Closing backdoor paths

Our goal is to close all backdoor paths

- Such that we're left with the only path we care about: the direct causal effect of D on Y
- If we succeed, we have met the *backdoor criterion*

Closing backdoor paths

Our goal is to close all backdoor paths

- Such that we're left with the only path we care about: the direct causal effect of D on Y
- If we succeed, we have met the *backdoor criterion*

In the example above, if we condition on I, we've met the backdoor criterion:

$D \rightarrow Y$ (causal effect of education on wages)

Closing backdoor paths

Our goal is to close all backdoor paths

- Such that we're left with the only path we care about: the direct causal effect of D on Y
- If we succeed, we have met the *backdoor criterion*

In the example above, if we condition on I, we've met the backdoor criterion:

$D \rightarrow Y$ (causal effect of education on wages)
$D \leftarrow I \rightarrow Y$ (backdoor path 1)

Closing backdoor paths

Our goal is to close all backdoor paths

- Such that we're left with the only path we care about: the direct causal effect of D on Y
- If we succeed, we have met the *backdoor criterion*

In the example above, if we condition on I, we've met the backdoor criterion:

$D \rightarrow Y$ (causal effect of education on wages)
$D \leftarrow I \rightarrow Y$ (backdoor path 1)
$D \leftarrow PE \rightarrow I \rightarrow Y$ (backdoor path 2)

Closing backdoor paths

Our goal is to close all backdoor paths

- Such that we're left with the only path we care about: the direct causal effect of D on Y
- If we succeed, we have met the *backdoor criterion*

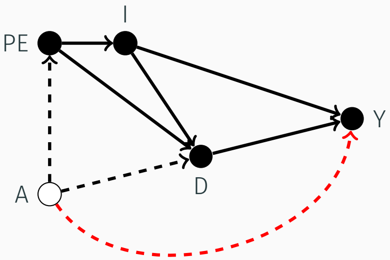
In the example above, if we condition on I, we've met the backdoor criterion:

$D \rightarrow Y$ (causal effect of education on wages)
$D \leftarrow I \rightarrow Y$ (backdoor path 1)
$D \leftarrow PE \rightarrow I \rightarrow Y$ (backdoor path 2)
$D \leftarrow A \rightarrow PE \rightarrow I \rightarrow Y$ (backdoor path 3)

We can identify the causal effect of interest!

One more complication

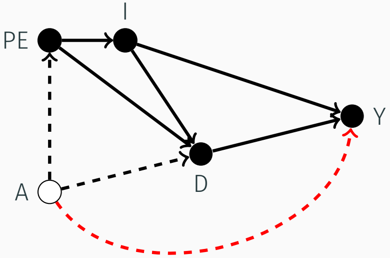
But what if ability has a direct effect on wages? And not only through education?



I now have a backdoor path $D \leftarrow A \rightarrow Y$

One more complication

But what if ability has a direct effect on wages? And not only through education?

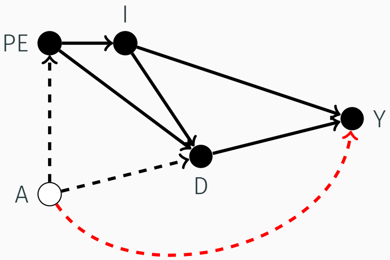


I now have a backdoor path $D \leftarrow A \rightarrow Y$

- The solution is to condition on A...But it's unobserved!

One more complication

But what if ability has a direct effect on wages? And not only through education?

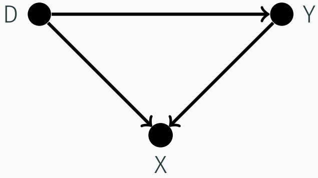


I now have a backdoor path $D \leftarrow A \rightarrow Y$

- The solution is to condition on A...But it's unobserved!
- Any solutions?

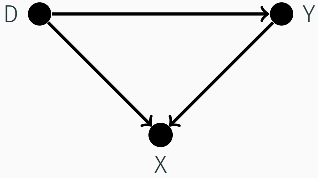
2nd type: Colliders

We've seen confounders, but we also need to be careful with colliders:



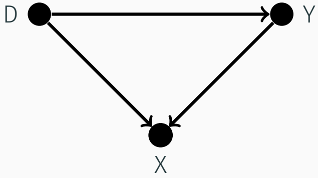
2nd type: Colliders

We've seen confounders, but we also need to be careful with colliders:



We still have a backdoor path: $D \rightarrow X \leftarrow Y$

We've seen confounders, but we also need to be careful with colliders:

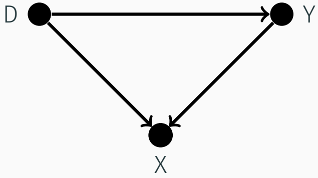


We still have a backdoor path: $D \rightarrow X \leftarrow Y$

- But the arrows are pointing *into* X rather than outward

2nd type: Colliders

We've seen confounders, but we also need to be careful with colliders:

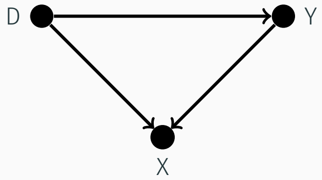


We still have a backdoor path: $D \to X \leftarrow Y$

- But the arrows are pointing *into* X rather than outward
- Collider: the causal effects of D and Y are colliding at X

2nd type: Colliders

We've seen confounders, but we also need to be careful with colliders:

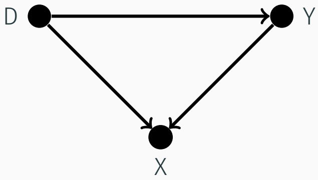


We still have a backdoor path: $D \rightarrow X \leftarrow Y$

- But the arrows are pointing *into* X rather than outward
- Collider: the causal effects of D and Y are colliding at X
- Colliders, if left alone, close the backdoor path

2nd type: Colliders

We've seen confounders, but we also need to be careful with colliders:

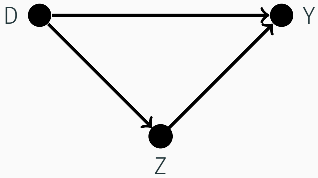


We still have a backdoor path: $D \rightarrow X \leftarrow Y$

- But the arrows are pointing *into* X rather than outward
- Collider: the causal effects of D and Y are colliding at X
- Colliders, if left alone, close the backdoor path
- But if you condition on the collider, **you open the backdoor path**

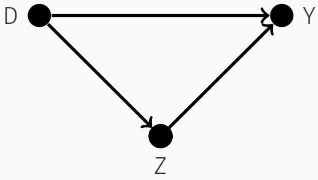
Causation between D and Z

What's going on here? Should I condition on Z?



Causation between D and Z

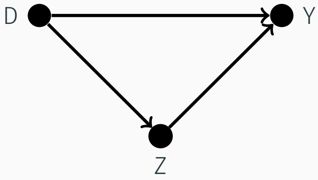
What's going on here? Should I condition on Z?



- Z is a mediator: D causes Z, which causes Y

Causation between D and Z

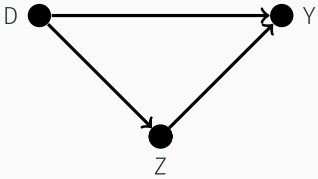
What's going on here? Should I condition on Z?



- Z is a mediator: D causes Z, which causes Y
- Both causal paths can be interesting!

Causation between D and Z

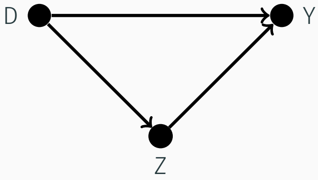
What's going on here? Should I condition on Z?



- Z is a mediator: D causes Z, which causes Y
- Both causal paths can be interesting!
- But again, you do **not** want to control for a mediator

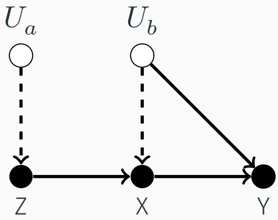
Causation between D and Z

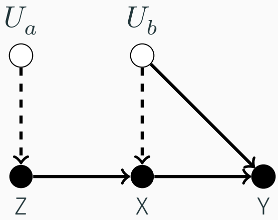
What's going on here? Should I condition on Z?



- Z is a mediator: D causes Z, which causes Y
- Both causal paths can be interesting!
- But again, you do **not** want to control for a mediator
- Advanced topics: Mediation analysis for estimation of direct effect vs total effect

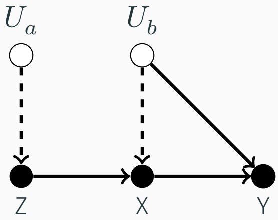
Some exercises



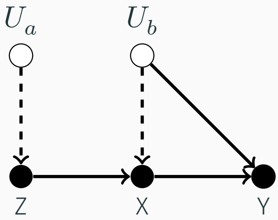


1. List all nodes

Some exercises

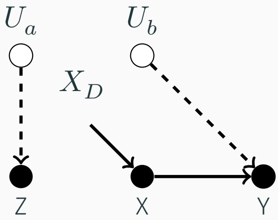


1. List all nodes
2. List all directed edges

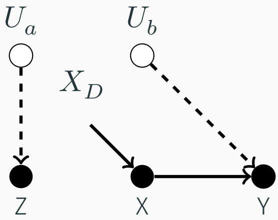


1. List all nodes
2. List all directed edges
3. Is the relationship between X and Y identified? Why?

Some exercises



Some exercises



1. What is X_D?
2. Is the relationship between X and Y identified?

How do I draw these things?

- *ggdag* in R

How do I draw these things?

- *ggdag* in R
- This website

How do I draw these things?

- *ggdag* in R
- This website
- tikz in Latex:

```
\begin{center}
  \begin{tikzpicture}[scale=0.8]
    \node[shape = circle, fill = black, label=west:D] (D) at (0, 0) {};
    \node[shape = circle, fill = black, label=east:Y] (Y) at (4, 0) {};
    \node[shape = circle, fill = white, draw = black, label=north:U] (X) at (2, 2) {};

    \path[->] (D) edge[draw=black,very thick] node {} (Y);
    \path[->] (X) edge[draw=black,very thick] node {} (D);
    \path[->] (X) edge[draw=black,very thick] node {} (Y);

  \end{tikzpicture}
\end{center}
```