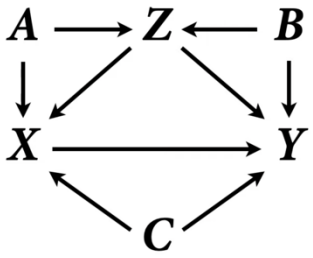
DAG STUFF

1. Complex DAG.

Identify paths connecting X to Y

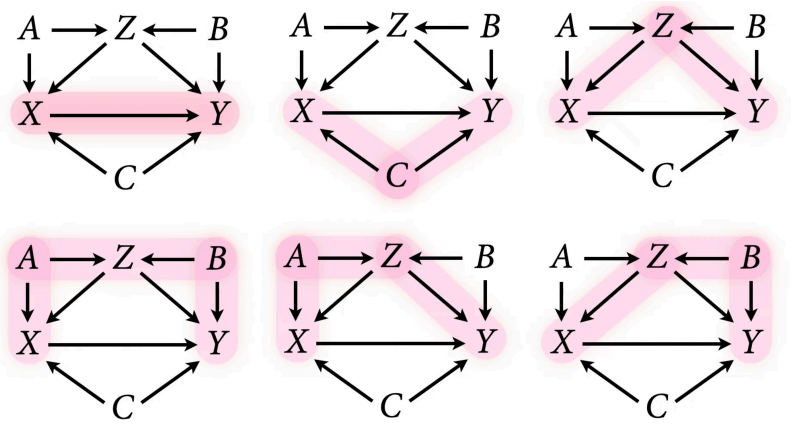
Identify which ones have arrows connecting X to Y where the arrow goes "into" x

Close the backdoors appropriately.



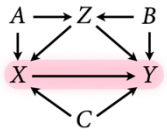
What variables should you include in your model?

Explanation: First, find all paths connecting X to Y



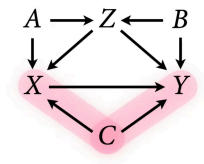
Identify the "minimal adjustment set:"

1. Leave the path from X to Y open.



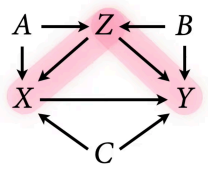
Causal path, open

2. C is a fork. Close the fork by adding “C” to the model



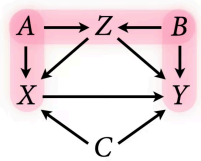
Backdoor path, open
Close with *C*

3. Z is a fork as well. Add “Z” to the model.



Backdoor path, open
Close with *Z*

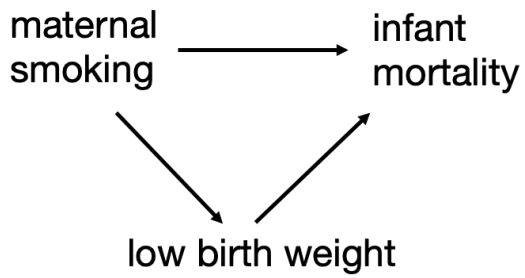
4. However, adding Z to the model creates a problem because it’s also a collider between A and B and opens up the path from X to Y through A and B. A and B are forks as well, so adding either one to the model will block this path.



Backdoor path, opened by *Z*
A or *B* to close

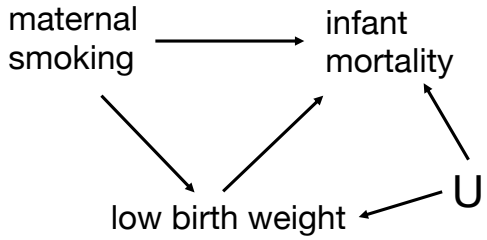
Resulting Model:
Y ~ X + C + Z + B or Y ~ X + C + Z + A

2. Maternal smoking is positively associated with infant mortality; however among babies born with a low birth weight (LBW), the relationship is reversed.
Consider this DAG and explain this paradoxical result:



Explanation: This "Birth Weight paradox" was the source of many controversies in epidemiology. One simple explanation is that LBW was a pipe, i.e., a consequence of smoking that also affected mortality rates.

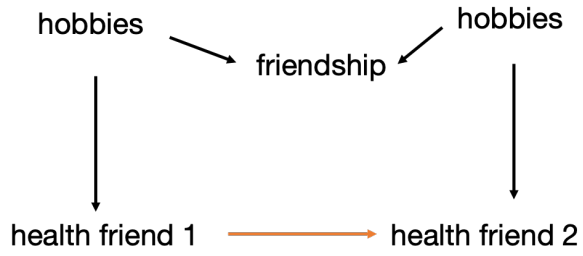
However, there may be other factors at play. It's possible that there are other, unmeasured processes (U) that generate low birth weight, e.g., birth defects, exposure to other drugs, malnutrition, etc.



By "stratifying" or only looking at babies within the LBW group, we have made it into a collider i.e., a variable that is a shared consequence of multiple causes (smoking and U).

Thinking through it: Imagine that the most common causes of LBW in babies are birth defects, malnutrition, and maternal smoking. By limiting the sample to just LBW babies, we are selecting for babies that have either birth defects, malnutrition, and/or tobacco exposure. Thus, this means that any baby in this group that *wasn't* exposed to tobacco must either have a birth defect was malnourished. If these two exposures (birth defects or malnourishment) have higher effects on infant mortality, it will make the babies in the smoking group look "ok" by comparison. By comparing our sample to babies who had exposure to smoking vs. babies who had exposure to other things associated with infant mortality, we will be biasing our estimation of the effect of smoking on infant mortality.

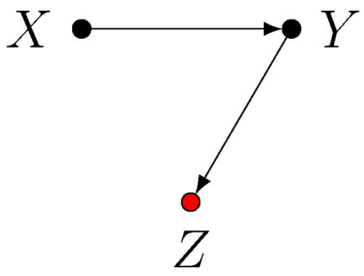
3. We're want to study whether friends influence each other's health
What happens if we limit our sample to people we know are friends?



Explanation: We want to study the effects of Friend 1 on Friend 2. However, there are other ways their health could be related. People's health is affected by their hobbies, e.g., people who do a lot of exercise may have better health, while people who drink a lot of alcohol may have worse health. Hobbies also may affect who becomes friends. Friendship therefore becomes a **collider.** By only looking at people who are already friends, we have narrowed our sample and effectively included "Friendship" in our model. This inclusion opens a backdoor path through hobbies -friendship-hobbies. This DAG shows an example of "M-Bias," a phenomenon that makes studying causal inference in social networks challenging.

4. In the 1800s records from men in the army in the US and Britain show that even as childhood nutrition rose, average height fell.

Explain this paradox:



Explanation: In this DAG, X represents childhood nutrition, Y is height, and Z is whether the individual was enlisted in the army. You can imagine that height influences whether a person enlists and/or is accepted into the army. Z therefore is an outcome of Y, and conditioning on it produces "case-control bias." By limiting our sample to just men who ended up in the army, we have inadvertently included Z in our model, leading to counterintuitive effects of nutrition on height.

Read more:
Cinelli, Forney, Pearl 2021 A Crash Course in Good and Bad Controls