Get unlimited access
Open in app

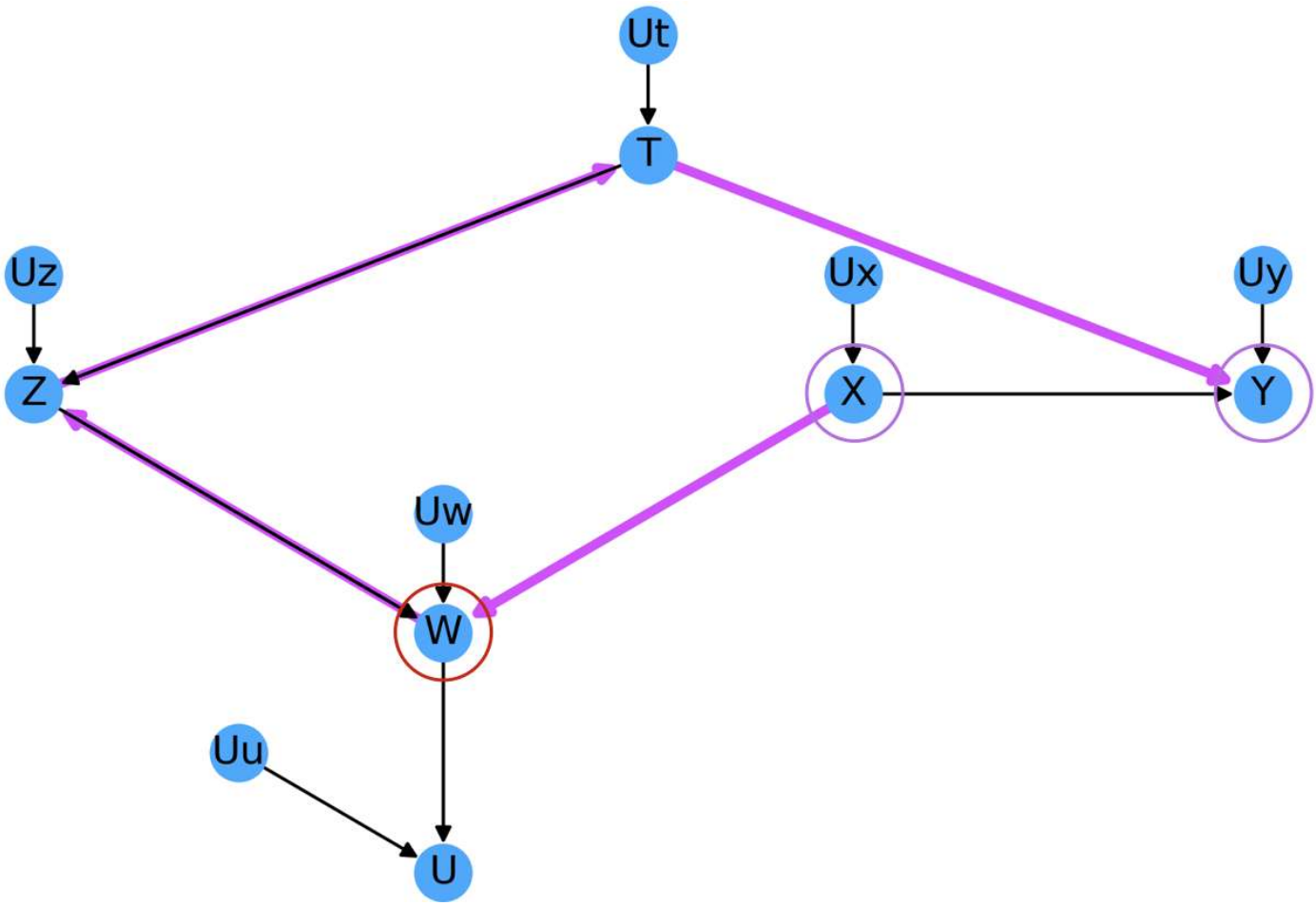
Published in Data For Science

You have 1 free member-only story left this month. Upgrade for unlimited access.

Bruno Gonçalves Follow

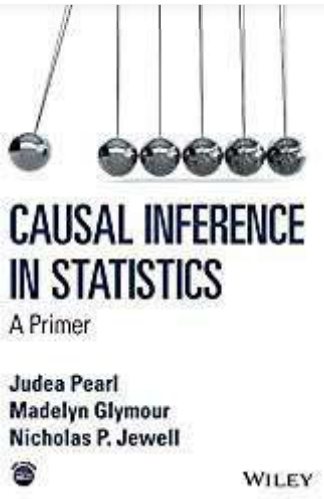
Nov 8, 2020 · 5 min read ★ · Listen

Save



CAUSAL INFERENCE

Backdoor Criterion



Amazon Affiliate Link

You can find the previous post here and all the we relevant Python code in the companion GitHub Repository:

DataForScience/Causality

How do causes lead to effects? Can you associate the cause leading to the observed effect? Big Data opens the doors for...

github.com

While I will do my best to introduce the content in a clear and accessible way, I highly recommend that you get the book yourself and follow along. So, without further ado, let's get started!

3.3 — Backdoor Criterion

One of the main goals of causal analysis is to understand how one variable causally influences another. In particular, and for practical reasons, we are interested in understanding under what conditions we can use observational data to compute causal

be present. The easiest way to do this is to make sure all non-causal paths between X and Y are blocked off. We can easily identify the variables we need to condition on by applying the so called '*backdoor criterion*' which is defined as:

> Backdoor Criterion — Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the backdoor criterion relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.

This definition is easy to understand intuitively: to understand the direct effect of X on Y we simply must make sure to keep all direct paths intact while blocking off any and all spurious paths.

If the backdoor criterion is satisfied, then the causal effect of X on Y is given by:

$$P\left(Y|\text{do}\left(X\right)\right) = \sum_{z} P\left(Y|X,Z\right)P\left(Z\right)$$

Which you'll recognize as a variant of the adjustment formula where the parents of X have been replace by Z. This implies that the parents of X naturally satisfy the backdoor criterion although in practice we are often interested in finding some other set of variables we can use.

Let's consider the DAG in Fig. 3.6:

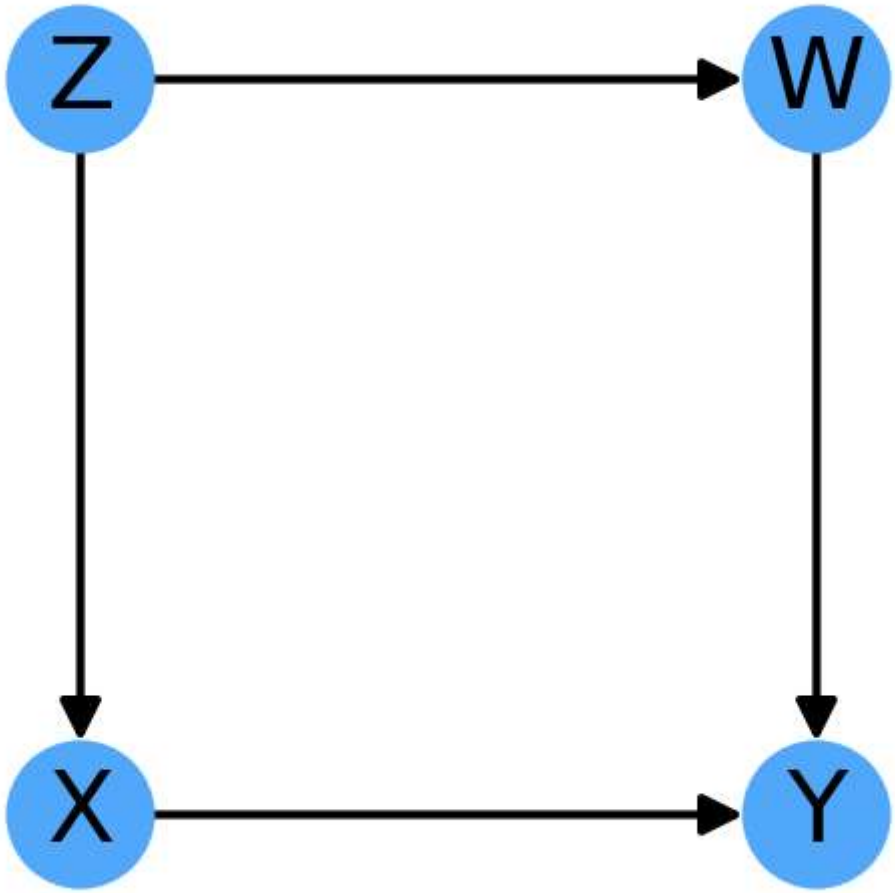


Fig 3.6

From this figure, it is clear that if we are interested in isolating the effect of X on Y, we can simply condition on Z (the parent of X). However, if for some reason we our dataset doesn't include information about Z we can also condition on W to obtain the same effect.

For a more complex example, consider Fig. 2.8, where we wish to measure the effect of X and Y. We already saw that there are no unblocked paths between X and Y as they are all blocked by the collider at W.

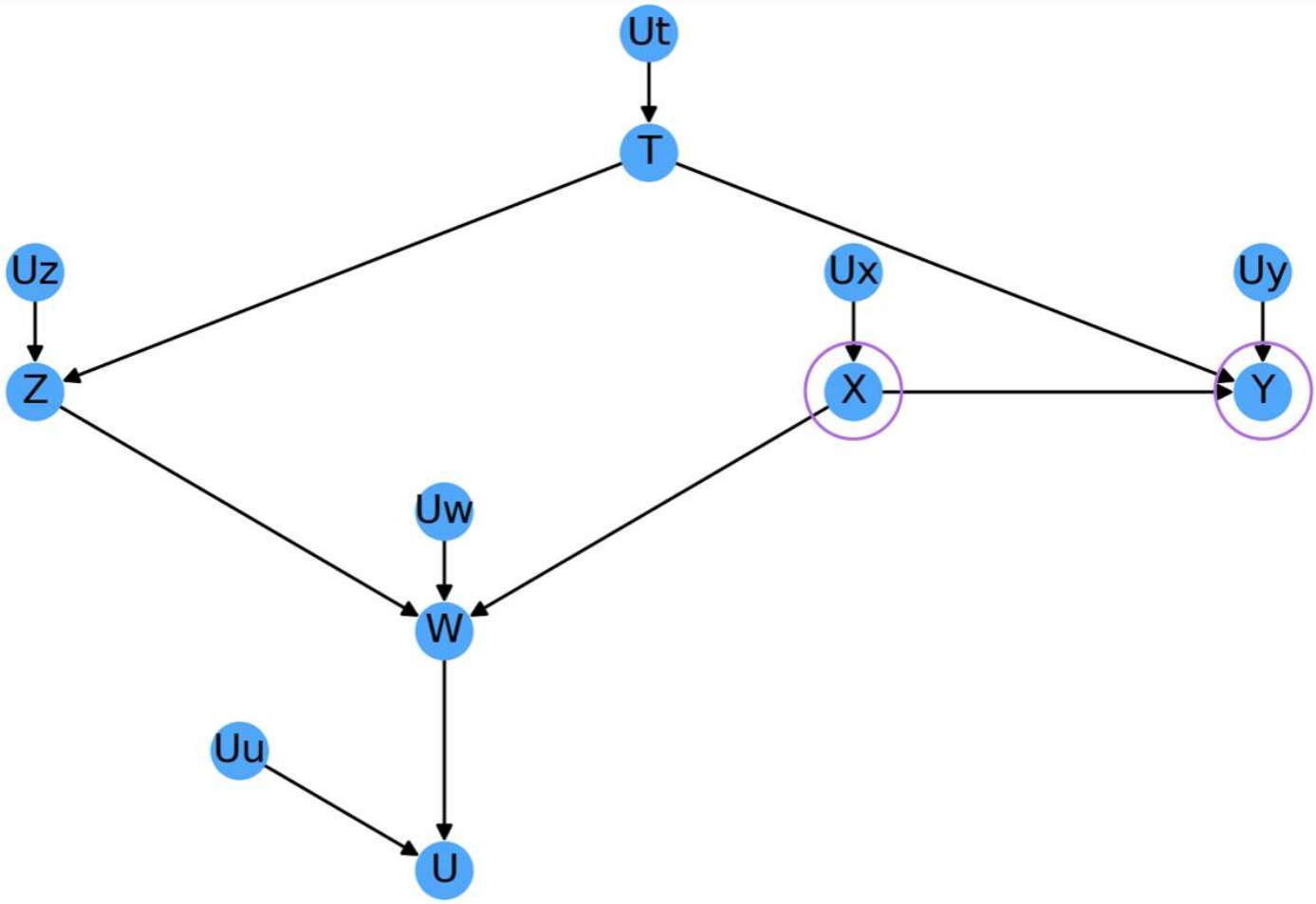
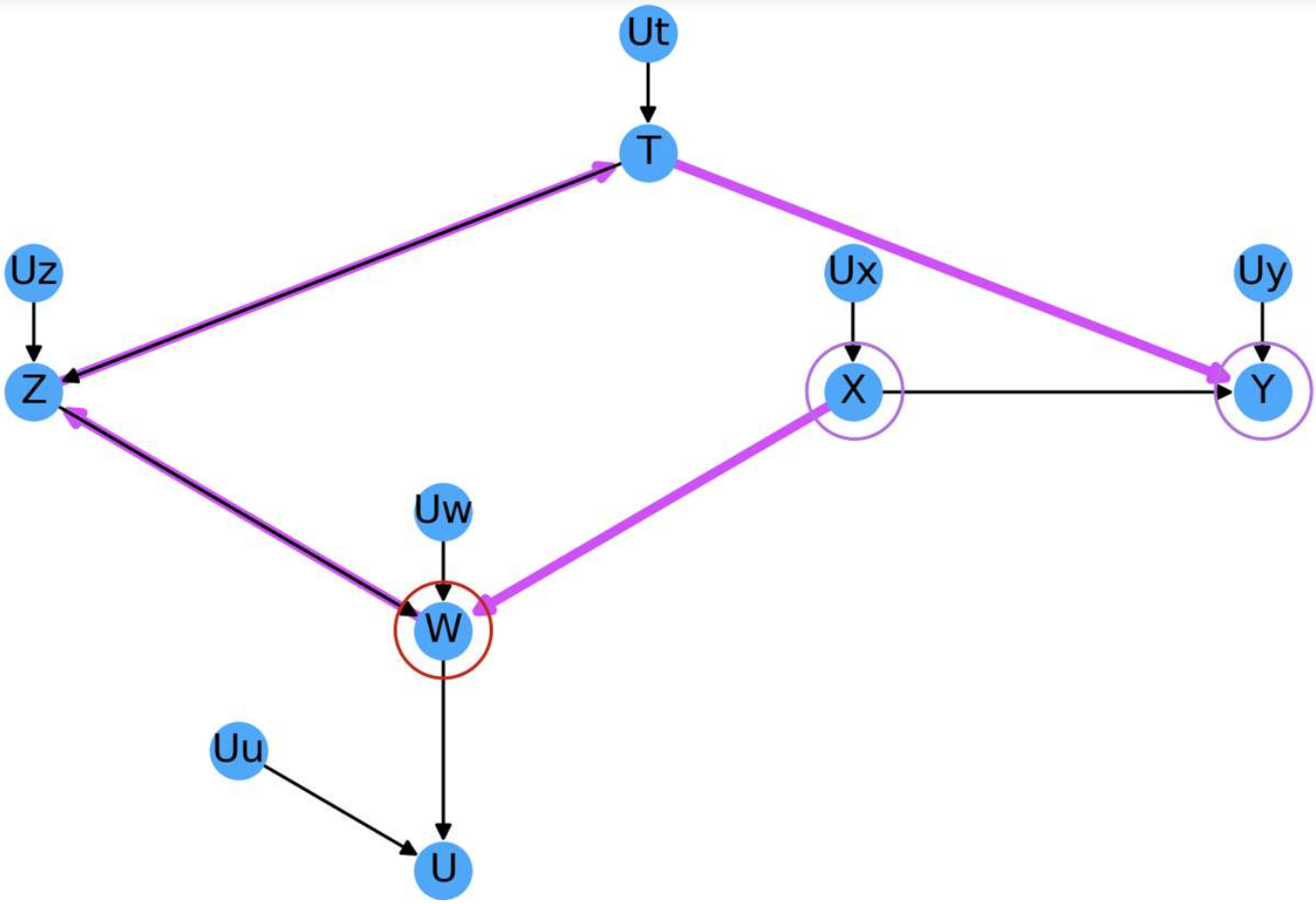


Fig 2.8

Now let's consider a slightly different question: We want to measure how tthe effect of X on Y depends on the observed values of W. To do this we need to condition on W which in turn opens up a new path:



Backdoor path between X and Y

Naturally, we can block this new path by conditioning on any one of the nodes that lie along it, say T. In this case, our expression then becomes:

$$P\left(Y|\mathrm{do}\left(X\right),W\right)=\sum_{t}P\left(Y|X,W,T\right)P\left(T|W\right)$$

Where we sum over all value of T in order to eliminate any dependence on it. This is known as the *W-specific causal effect*.

To firm up our understanding of the backdoor criterion, let us consider the more complex case of Fig. 3.8:

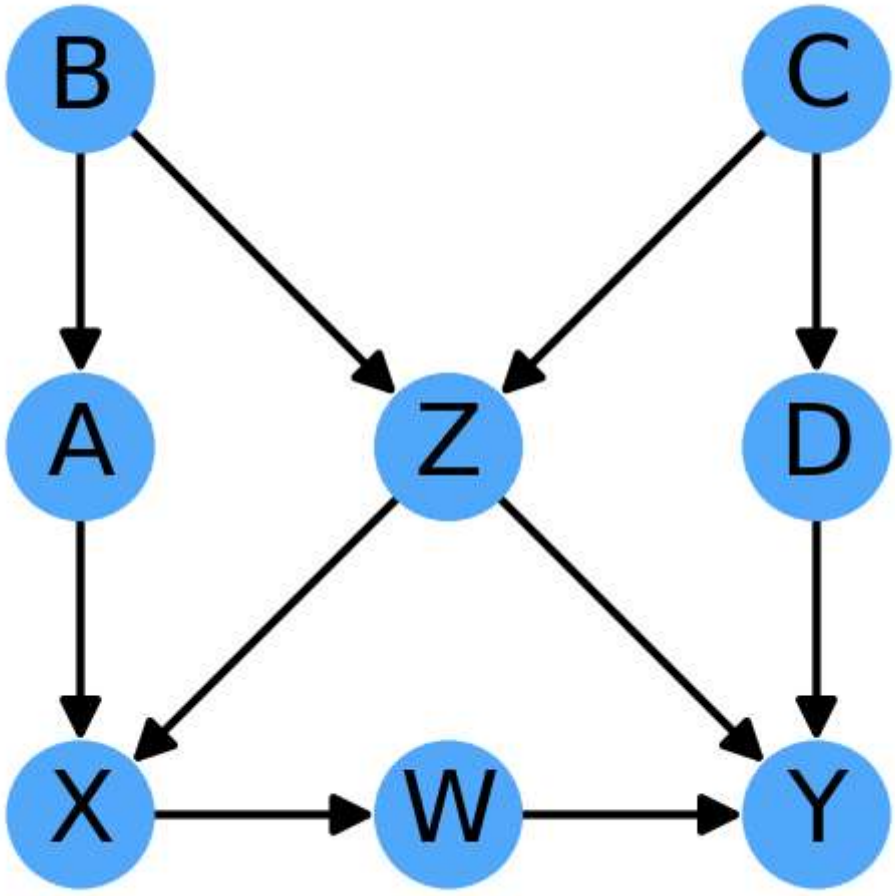
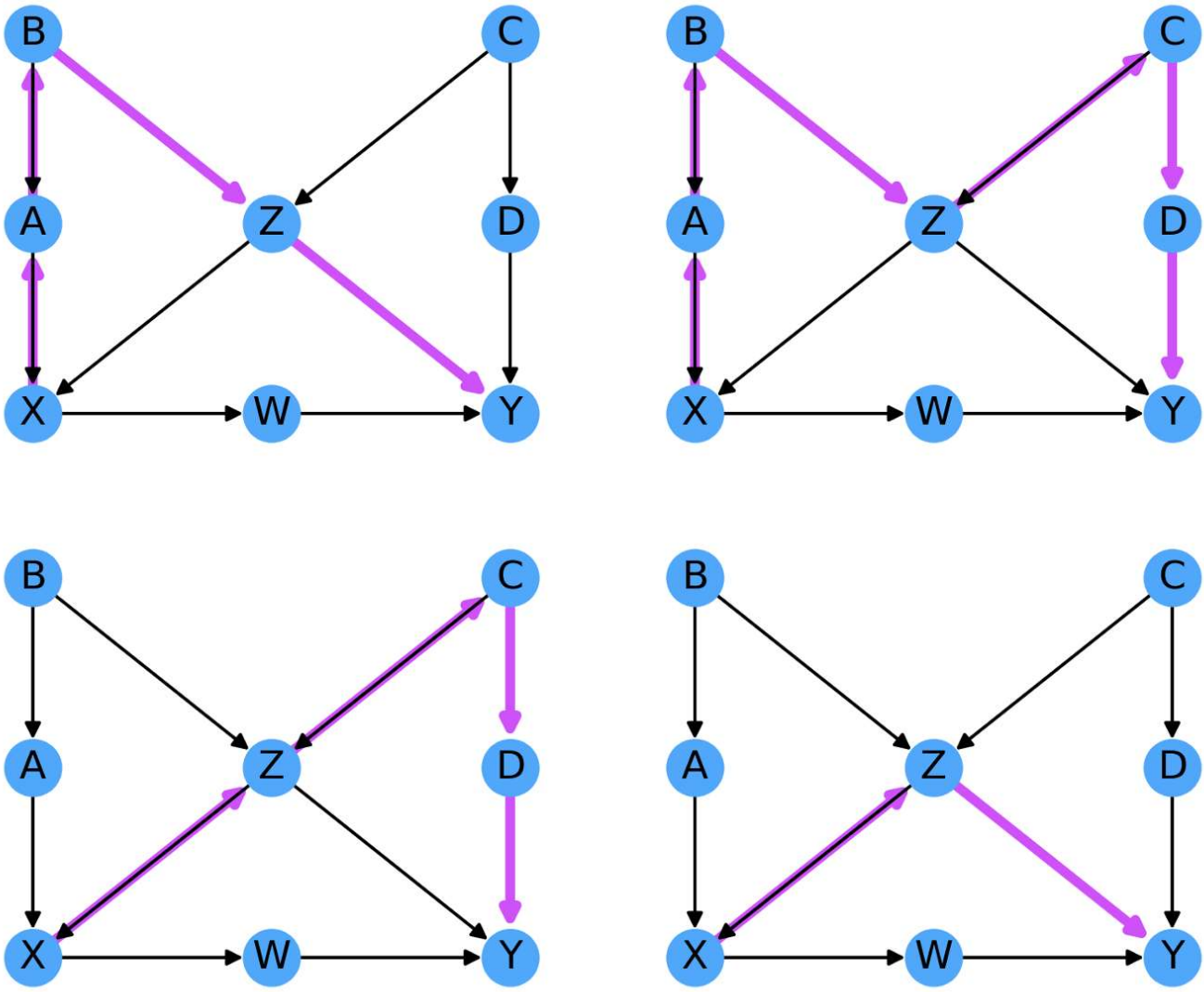


Fig 3.8

In order to determine the effect of X on Y, we start by identifying all the non-directed paths from X to Y:



Non-causal paths between X and Y

71 1

From the image above, it's easy to [illegible] present in all paths, so we should condition on it. However, since Z is a collider, we must also condition on one of its parents (or their descendants), giving us a choice of one of (A, B, C, or D). Z in addition to any combinations of these 4 nodes will fulfill the back-door criteria. So we could condition on (Z, A), (Z, B), (Z, A, C), (Z, A, B, C, D), etc.

Congratulations on working with us through another technical Causality post. We are

As always, you can find all the notebooks of this series in the GitHub repository:

DataForScience/Causality

How do causes lead to effects? Can you associate the cause leading to the observed effect? Big Data opens the doors for...

github.com

The next post on the series is already available:

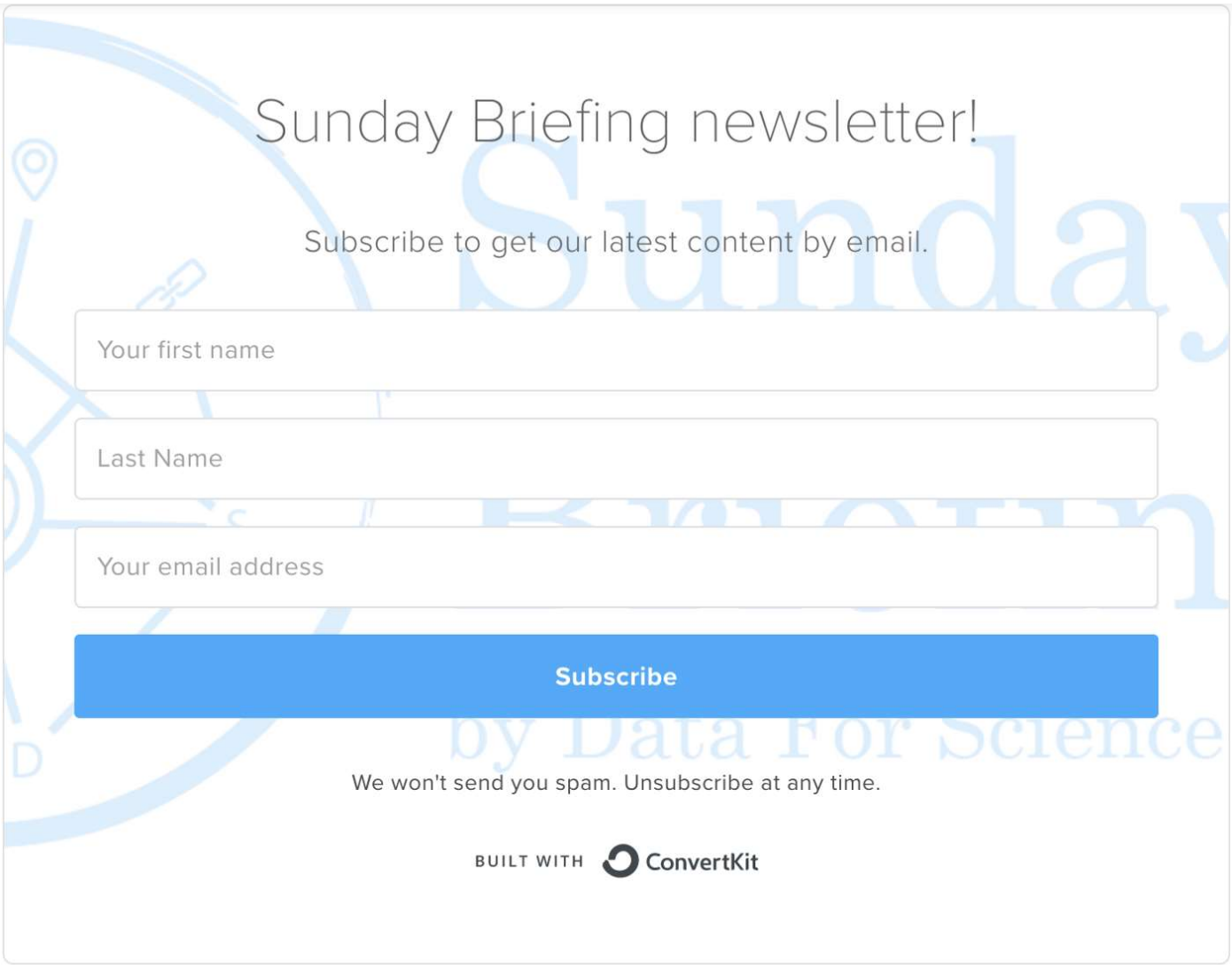
Causal Inference — Part XII — Front-door Criterion

This is the twelveth post on the series we work our way through “Causal Inference In Statistics” a nice Primer...

medium.com

And if you would like to be notified when the next post comes out, you can subscribe to the The Sunday Briefing newsletter:

Get unlimited access Open in app



Get an email whenever Bruno publishes.

Subscribe

Emails will be sent to burkanjobs@gmail.com.
Not you?

Get unlimited access
Open in app