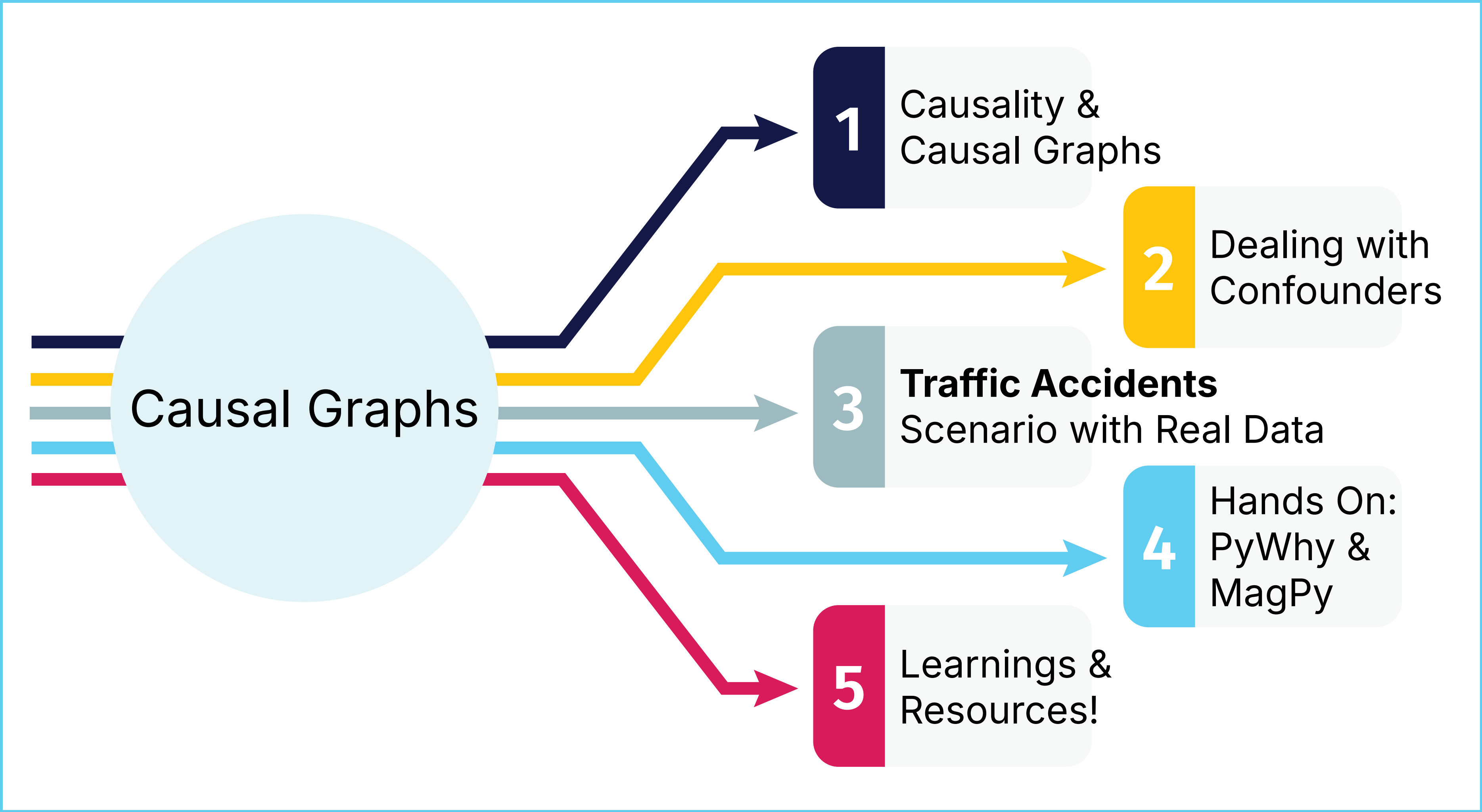
Causal Graphs in Practice:
Navigating the Arrow of Why

Amy Hodler & Michelle Yi
January 2025

data day texas

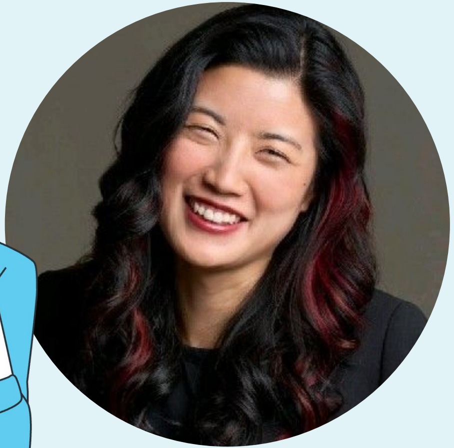

Causal Graphs
1 Causality & Causal Graphs
2 Dealing with Confounders
3 **Traffic Accidents** Scenario with Real Data
4 Hands On: PyWhy & MagPy
5 Learnings & Resources!

Michelle Yi

Applied AI, Artist, Speaker
Board Member, Women in Data
Michelle@generationship.ai

Amy Hodler

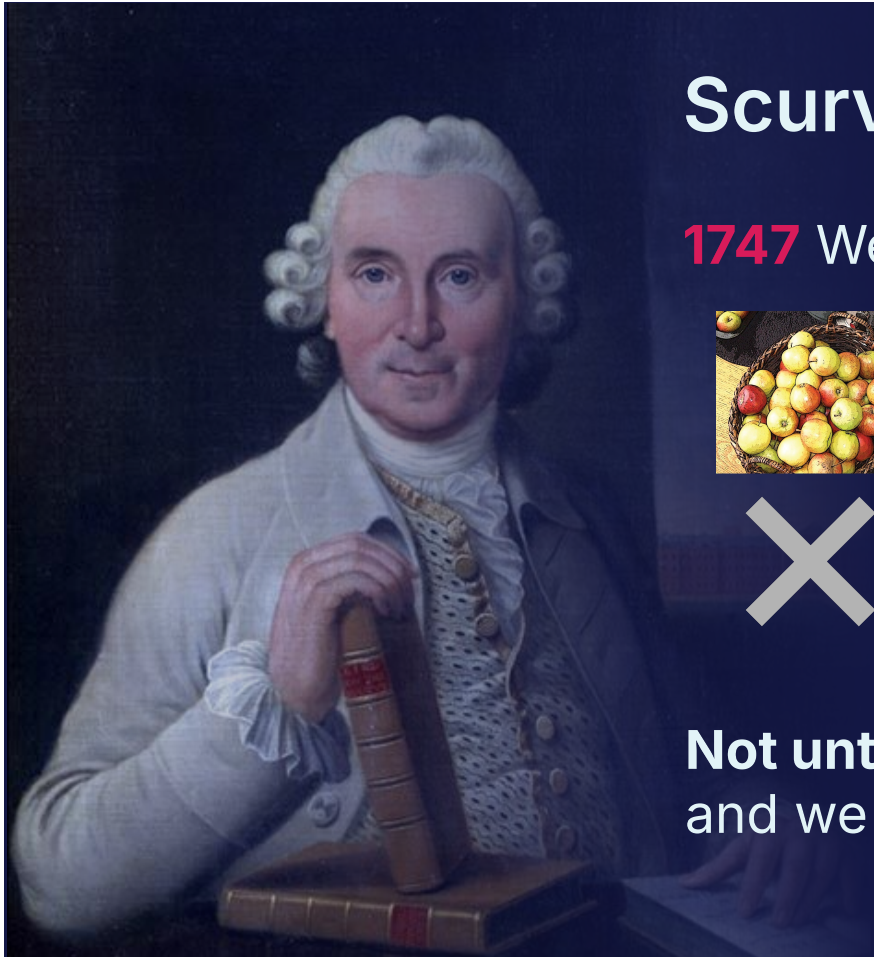
Graph Advisor, Author, Speaker
Founder of GraphGeeks
Amy.Hodler@GraphGeeks.org



WOMEN IN DATA™

GraphGeeks

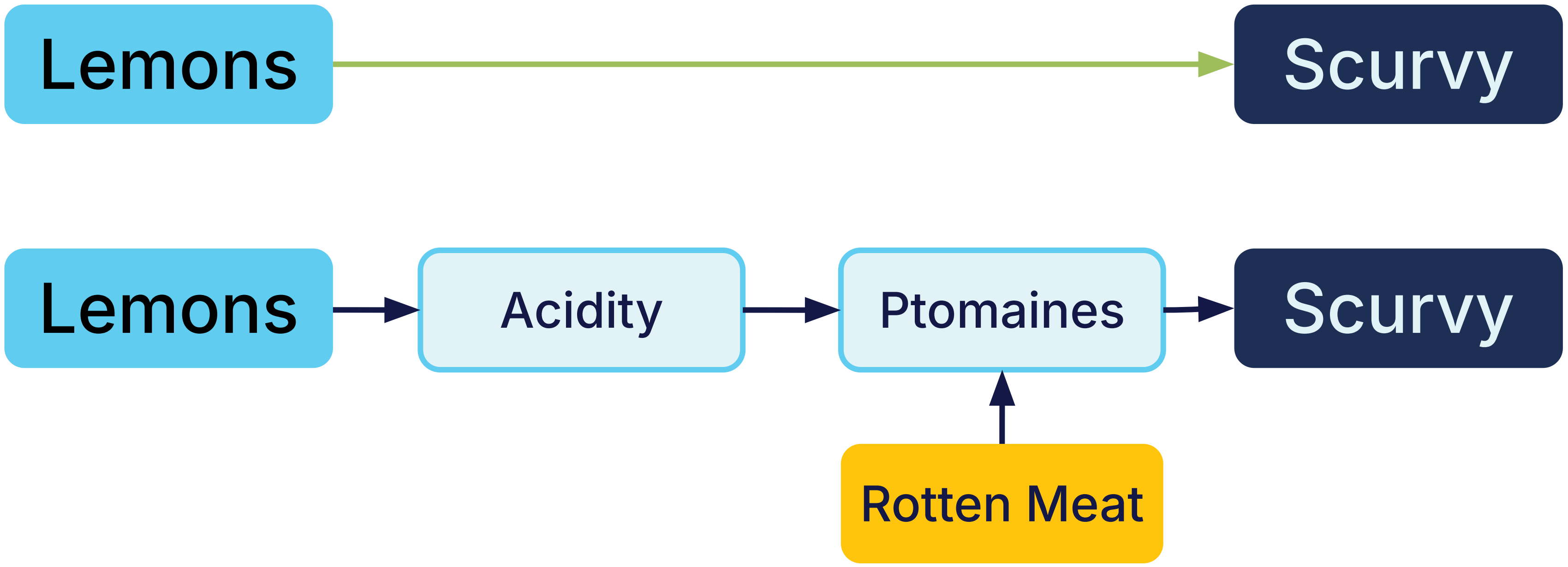
Scurvy Killed Millions of People



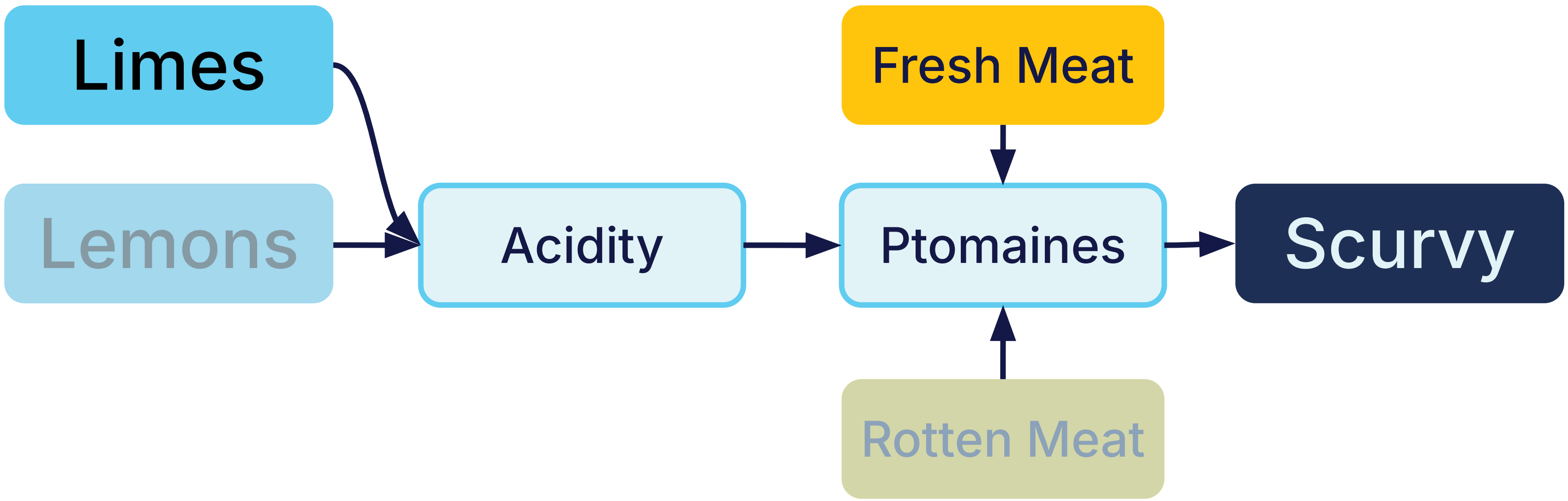
James Lind

1747 We knew lemons prevented scurvy

Not until 1932 was the cause understood
and we eradicated it as a common disease



The leading theory on why lemons cured scurvy was that it neutralized some harmful effects of rotting meat



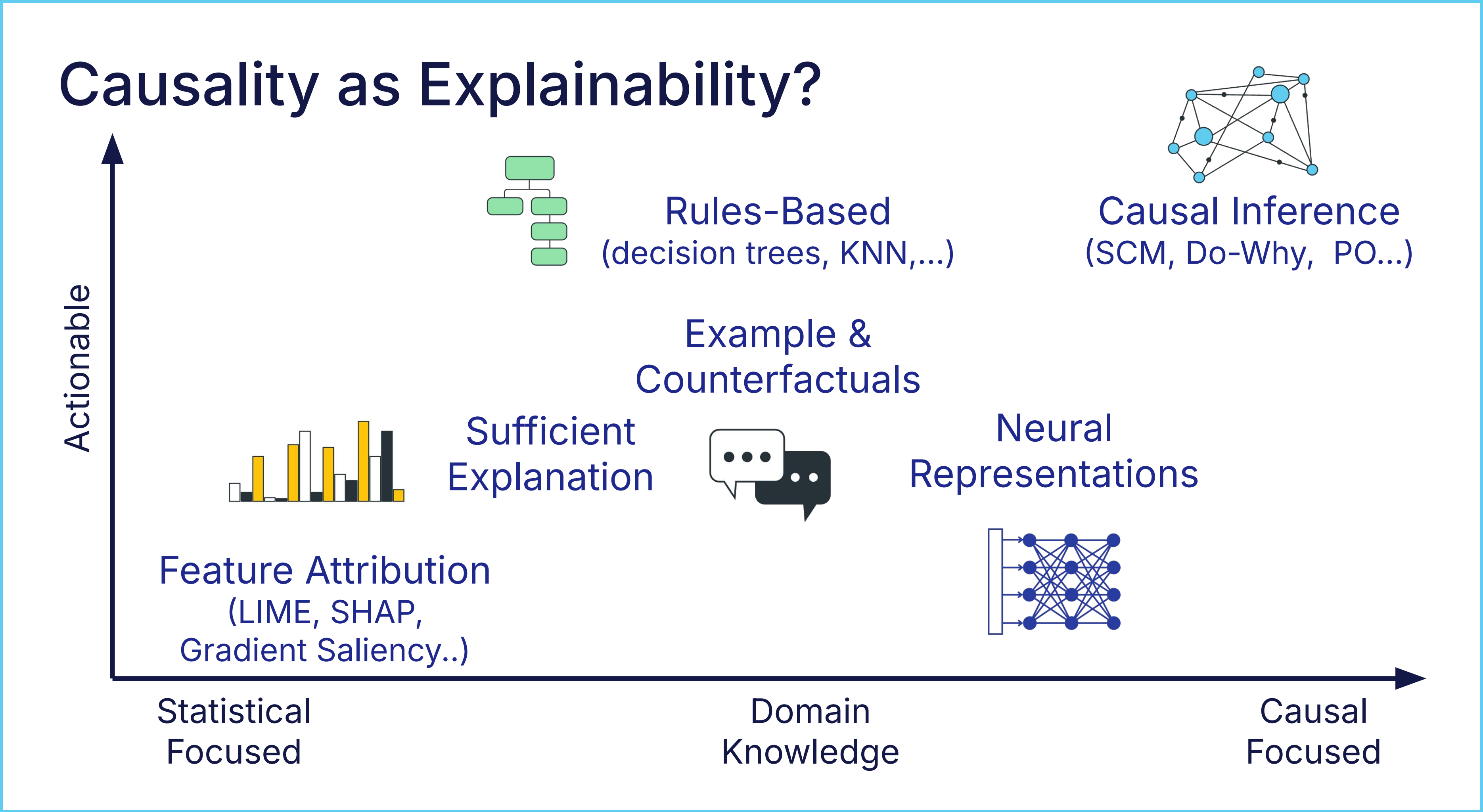
If acidity is the mediating effect, then we can substitute lemons for limes.

Or we can add fresh sources of meat to avoid the ptomaines in the first place.

"What will we be slapping our foreheads about sixty years from now, wondering how we missed something so obvious?"

Maciej Cegłowski
[Scott and Scurvy]

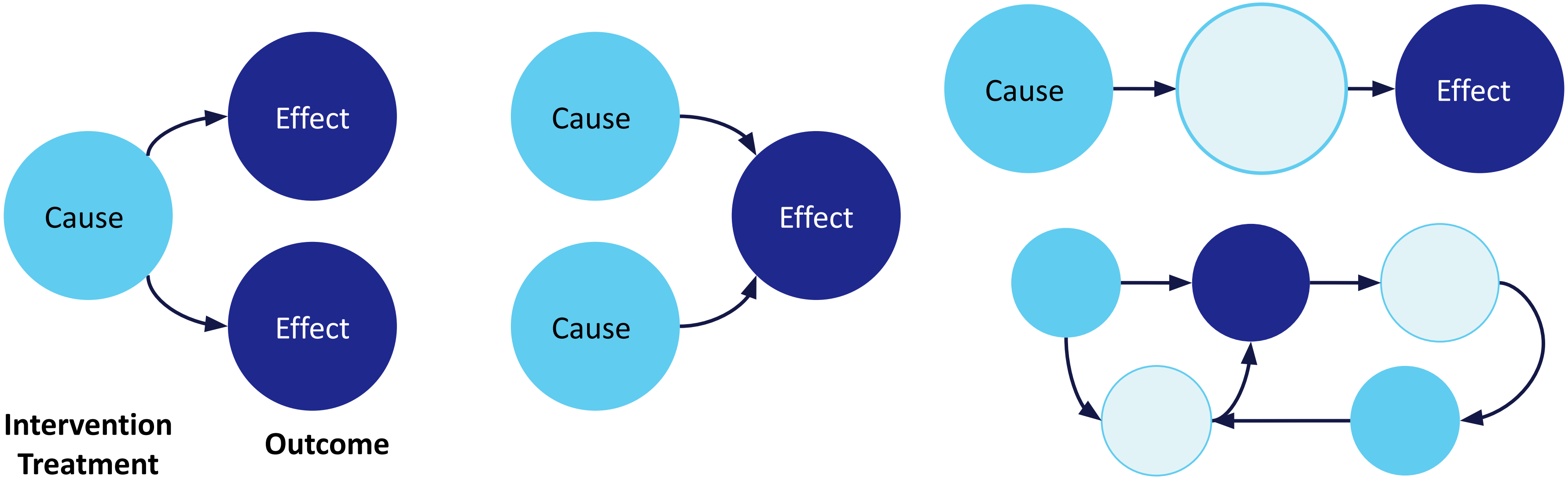
The Arrow of Why


Causality as Explainability?
Actionable
Rules-Based
(decision trees, KNN,...)
Causal Inference
(SCM, Do-Why, PO...)
Example &
Counterfactuals
Sufficient
Explanation
Neural
Representations
Feature Attribution
(LIME, SHAP,
Gradient Saliency..)
Statistical
Focused
Domain
Knowledge
Causal
Focused

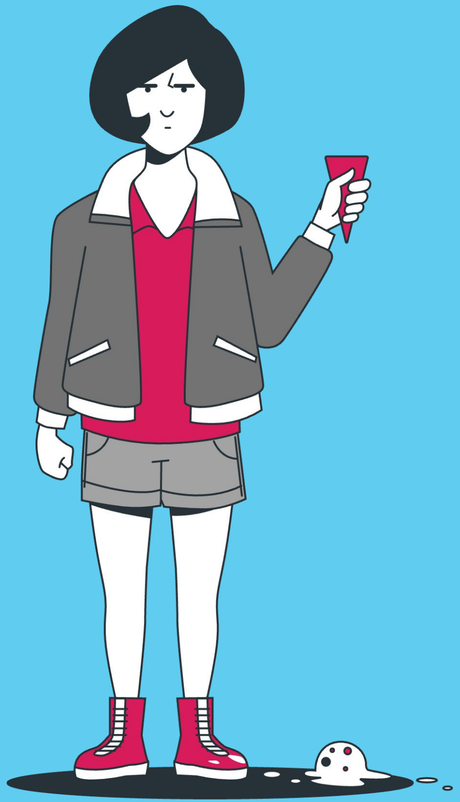
Causality Goes Beyond Explainability

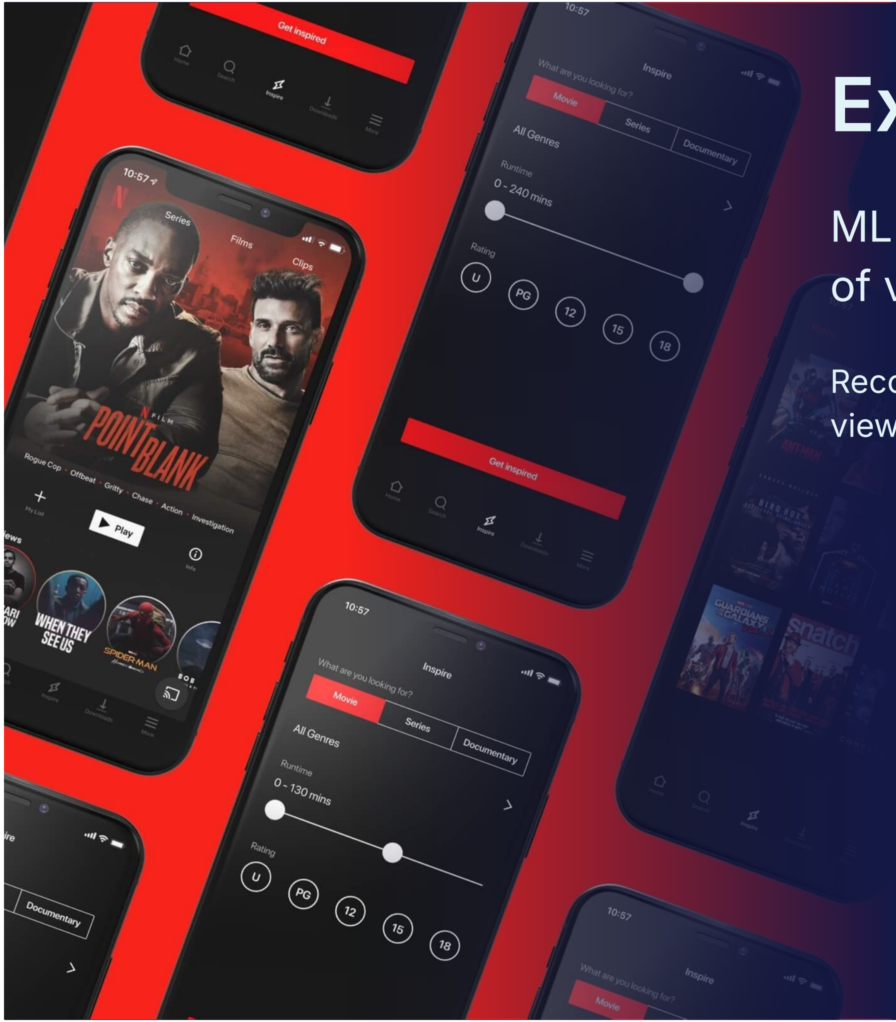
Why does this outcome happen?

- Finds how a change in one variable causes a change in another
- Causality is inferred using the difference between outcomes



No, You Can’t Just Use Plain Machine Learning



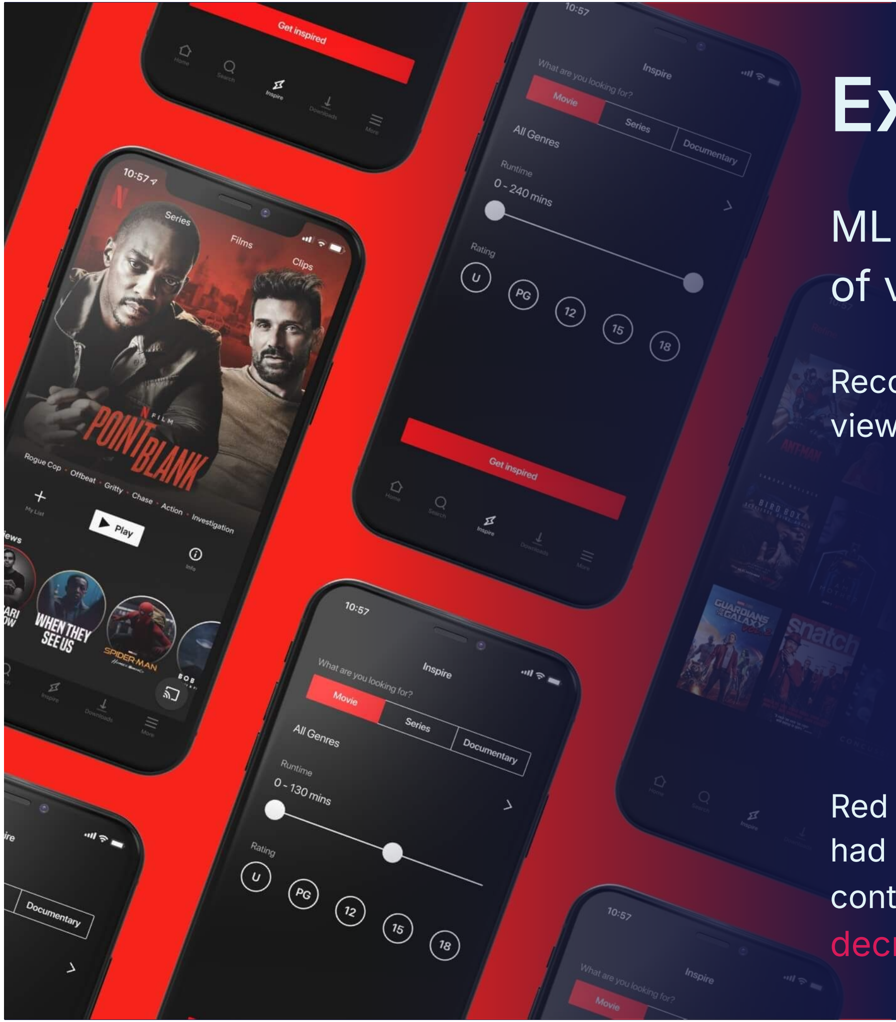


Example of ML #Fail

ML identified Red thumbnails as a top predictor of viewership

Recommended making every thumbnail red. Focused on predicting viewership rather than increasing viewership

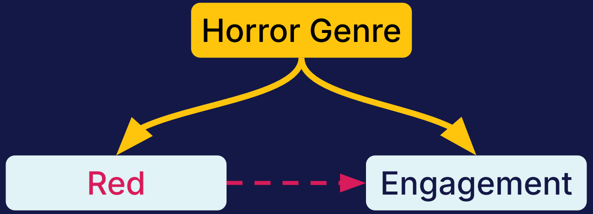
Red → Engagement



Example of ML #Fail

ML identified Red thumbnails as a top predictor of viewership

Recommended making every thumbnail red. Focused on predicting viewership rather than increasing viewership



Red thumbnails were more likely to appear in horror movies, which had naturally higher engagement. By applying causal AI in the context of movie genres, we learn that red images actually decrease engagement overall.

Plain ML is Great for Correlations

Observational data tells us about joint distributions

$$p(y \mid x)$$ probably of Y given X

x Anisha buys a padded backpack

y Anisha buys a laptop

There's a high correlation between Anisha buying a laptop and a fancy padded backpack

But Plain ML is Awful at Causality

Correlations tell us nothing about directionality or interventions

$p(y \mid do(x))$ probably of Y given we **do** X

x Anisha buys a padded backpack

y Anisha buys a laptop

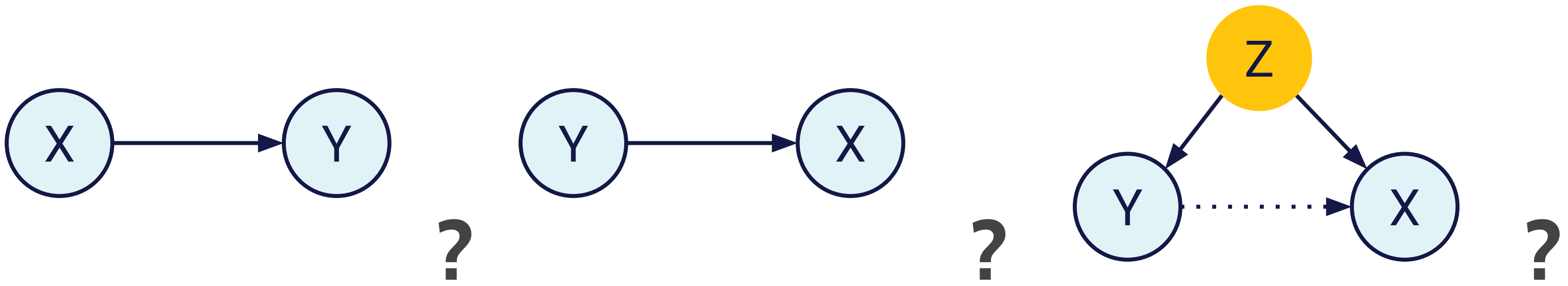
Give her a backpack (intervention) and it's very unlikely she'll buy a laptop

$$p(y \mid do(x)) \sim 0$$

Give her a laptop (intervention) and she'll likely want to buy a backpack to keep it safe

$$p(x \mid do(y)) \sim \text{high}$$

Causal Structures Are Inductive Biases



These structures are not present in typical ML but are absolutely necessary to make the right decisions!

Causal Inference Is The Next Big Wave

Databricks Industry survey from 400 AI experts

Organisations have realised that the majority of use-cases require causal AI - and the language of causal AI is becoming commonplace in many industries, from media to the financial sector. Causal AI had the highest score from respondents on *'Not using it but planning to in next 12 mths'.* Link

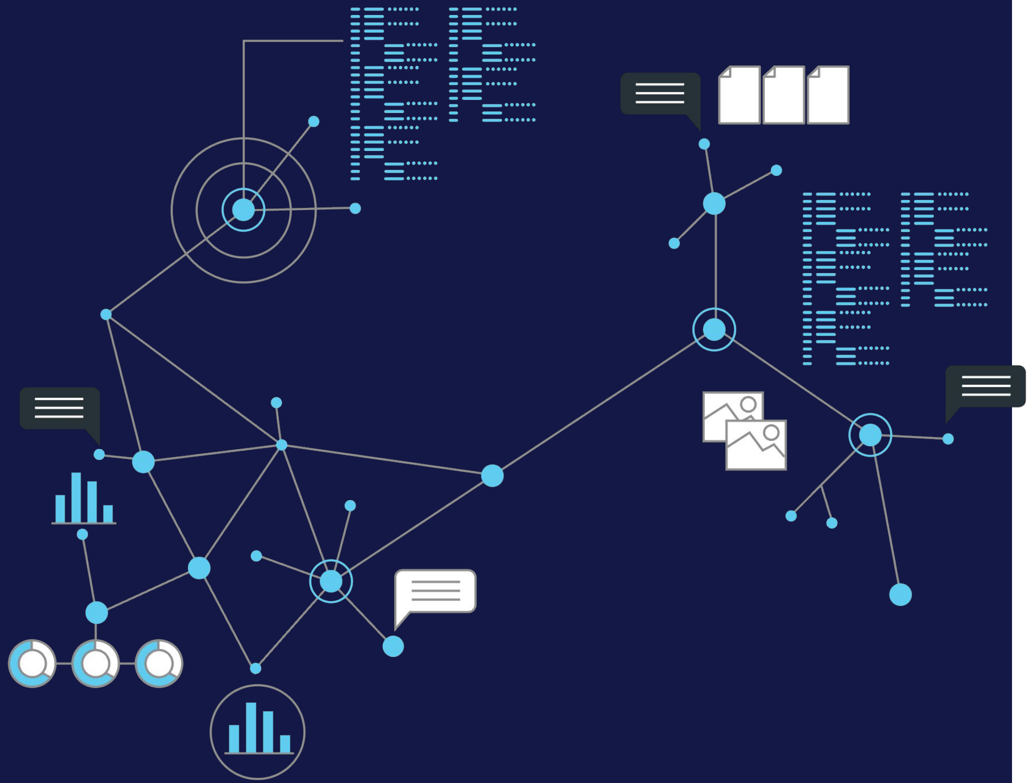
BigTech Investment in Causal Inference

Many BigTechs have heavily invested in Causal Inference approaches, publishing research and open source packages to drive commercial adoption and for hiring.

Regulatory Agenda: Explainable & Bias-Free AI

Clarity is still emerging across international regulators, but AI for high-impact use cases will require oversight and explainability at its core: strongly suited to rapid adoption

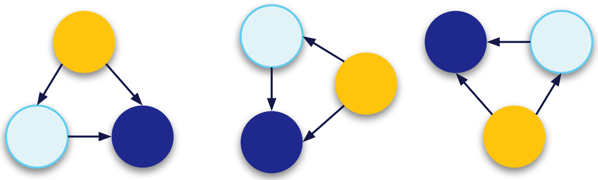
A Better Model



Breakthrough and Turning Point

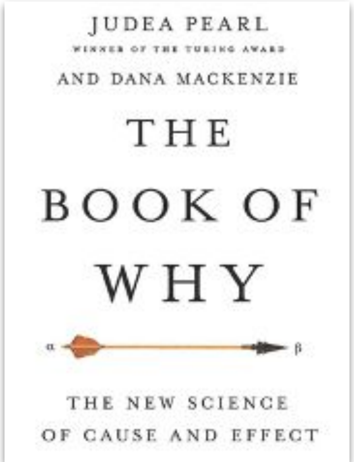
Accessible arguments on causation

Using graphs to model relationships changed the way we think about interventional questions



Creation of a *do*-calculus provided a mathematical way to calculate causal strength of *"doing"* something

Megapost summary:
https://engineeringideas.substack.com/p/megapost-about-causality-the-summary



$P(Y|X)$

to

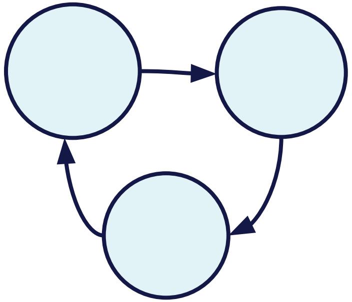
$P(Y|do(X))$

Causal Graphs as a Unifying Model

Directed Acyclic Graphs (DAG) ONLY

No feedback loops: no intervention can cause itself!

Arrow direction illustrates causal influence



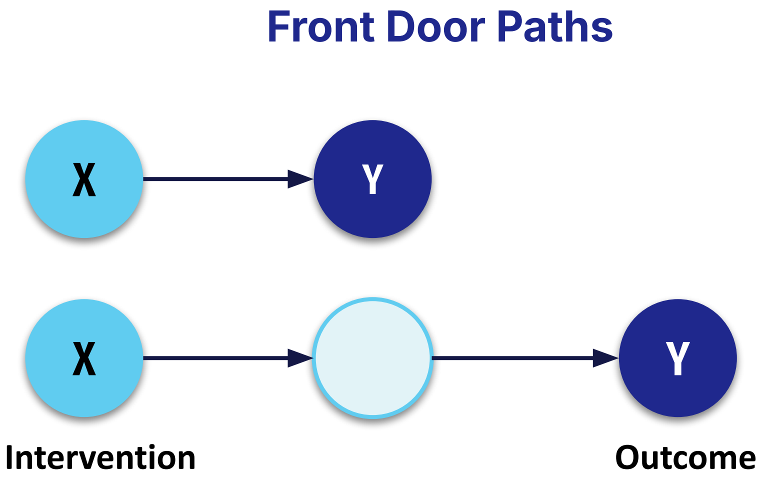
Cycles NOT Allowed

Causal Graphs as a Unifying Model

Directed Acyclic Graphs (DAG) ONLY

No feedback loops: no intervention can cause itself!

Arrow direction illustrates causal influence



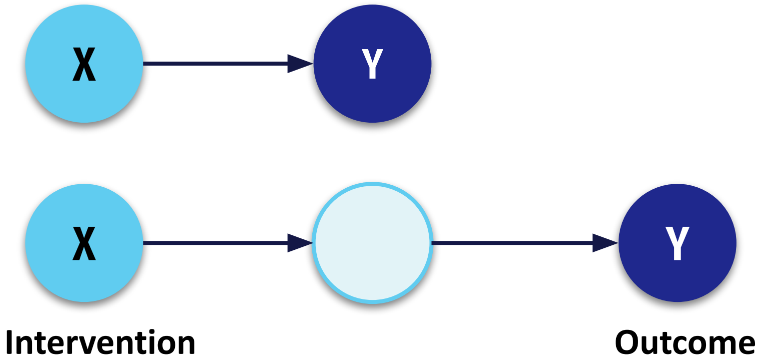
Causal Graphs as a Unifying Model

Directed Acyclic Graphs (DAG) ONLY

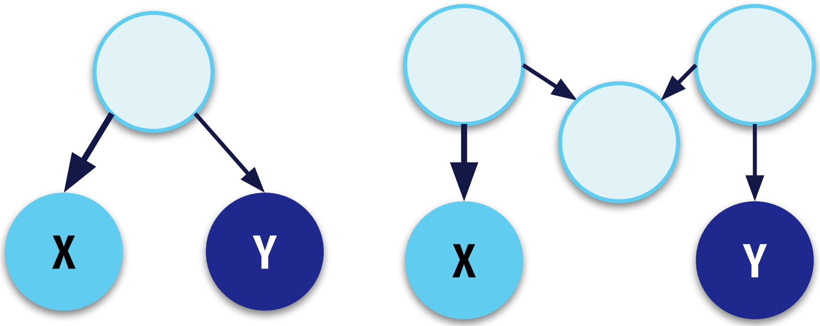
No feedback loops: no intervention can cause itself!

Arrow direction illustrates causal influence

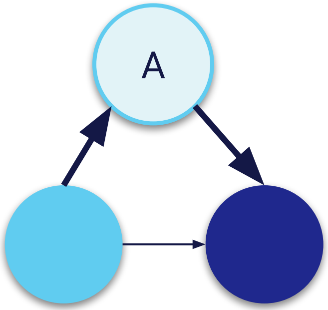
Front Door Paths



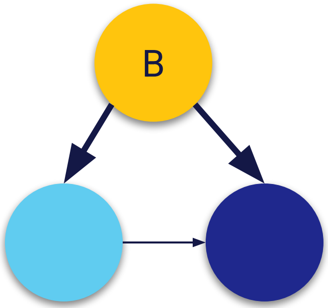
Back Door Paths



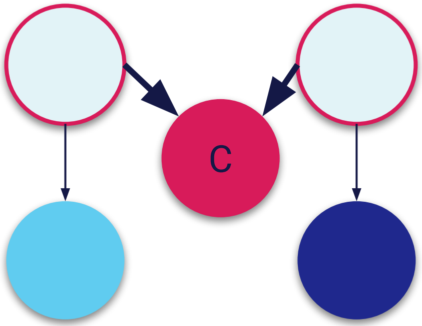
Basic Structures of Causal Graphs



A mediator

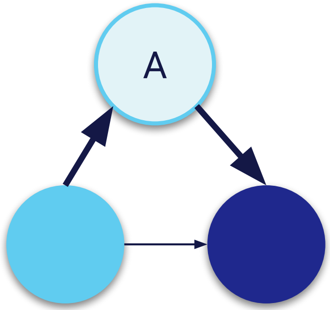


B confounder

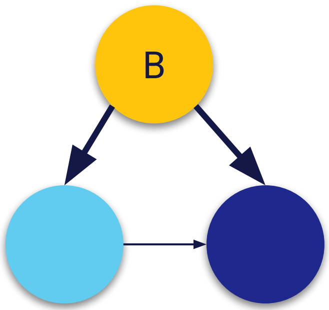


C collider

Basic Structures of Causal Graphs

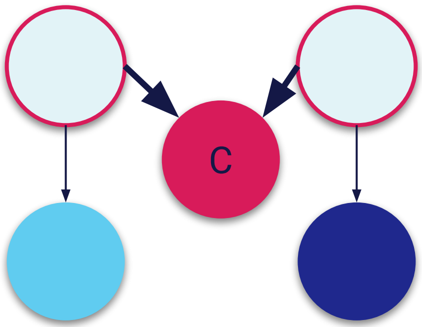


A mediator



B confounder

Control For
if pointing to
treatment & outcome

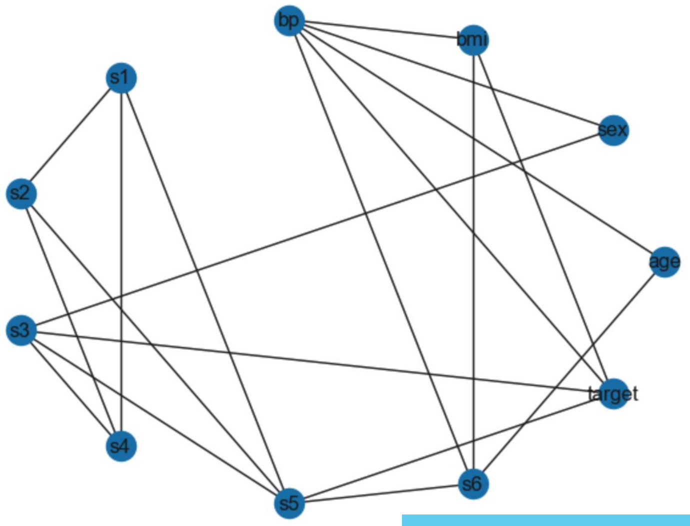


C collider

Do Not Control
As it introduces correlations
that require further controls

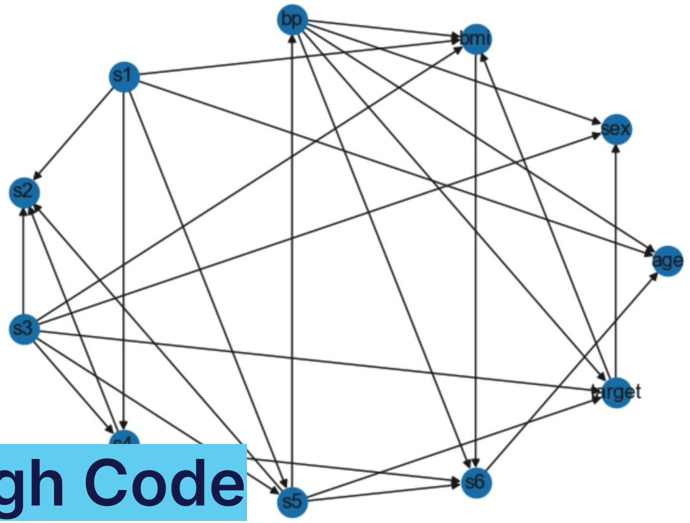
Don’t Worry - There Are Tools To Help!

```python
import networkx as nx

g = nx.from_pandas_adjacency(skeleton)
nx.draw_circular(g, with_labels=True)
plt.show()
```




```python
# discovered graph with domain expertise
astar = AStarSearch(df_std, super_graph=skeleton, include_graph=include_graph)
astar.run_scoring(parallel=False, func=bic_score_node)
mat = astar.search()

g = nx.from_pandas_adjacency(mat, create_using=nx.DiGraph)
nx.draw_circular(g, with_labels=True)
plt.title("Discovered graph with domain expertise")
plt.show()
```

Discovered graph with no domain expertise



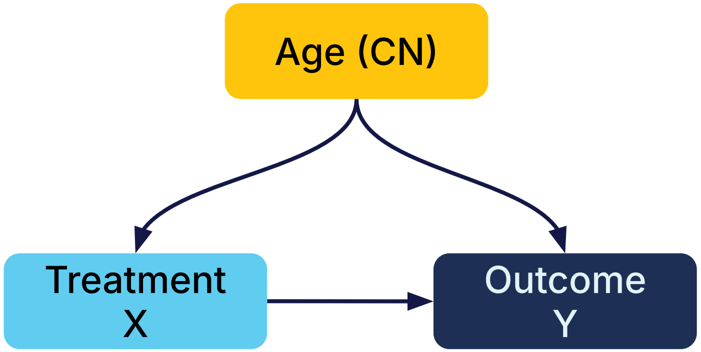
Graphs Through Code

Logic Behind the Tools

Confounders - Back Door Adjustments

Control Node (CN) 1) Must Not be a child of the treatment your estimating and 2) Must block the path between treatment and outcome.

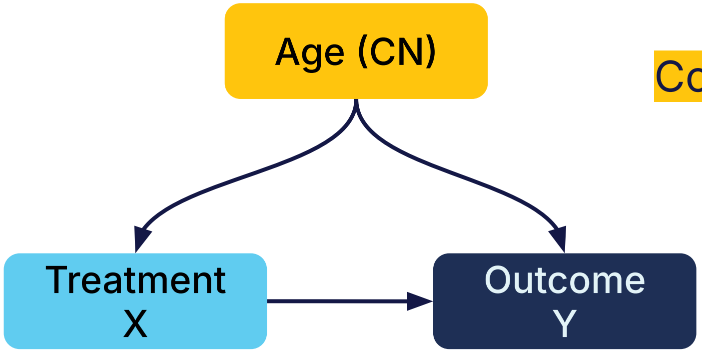
Observed Data



Confounders - Back Door do-calculus

Interventions represented by the operator, **do(x)**, which "erases" the function while keeping the rest of the model unchanged

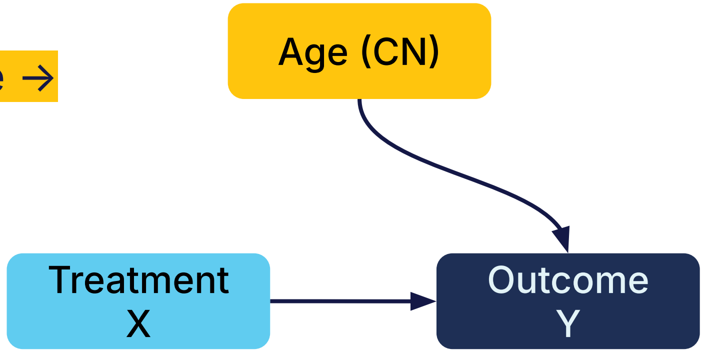
Observed Data



$P[Y|X]$

Control for Age →
(hold steady)

Intervention Analysis



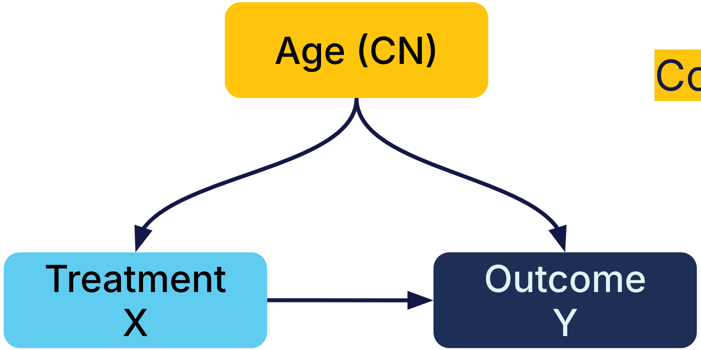
$P[Y|\mathbf{do(X)}]$

Confounders - Back Door do-calculus

$$P(Y|do(X)) - P(Y)$$

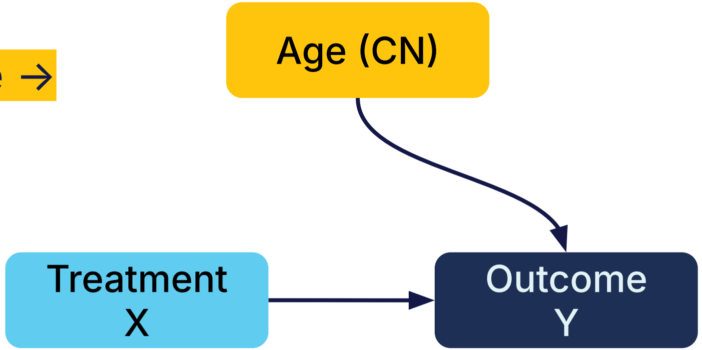
Replaces probability formulas with constant variables to simulate a change in a controlled manner and "closes" the backdoor.

Observed Data



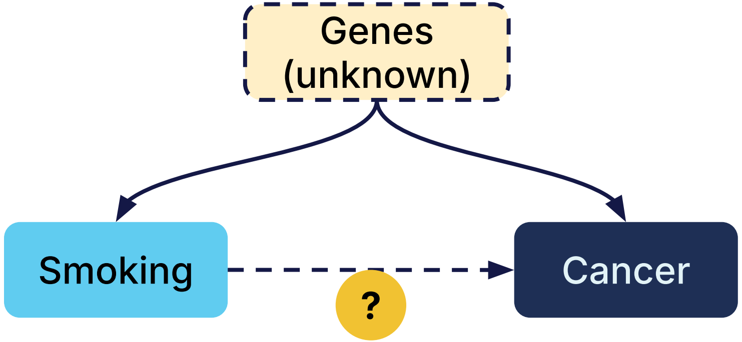
Control for Age →
(hold steady)

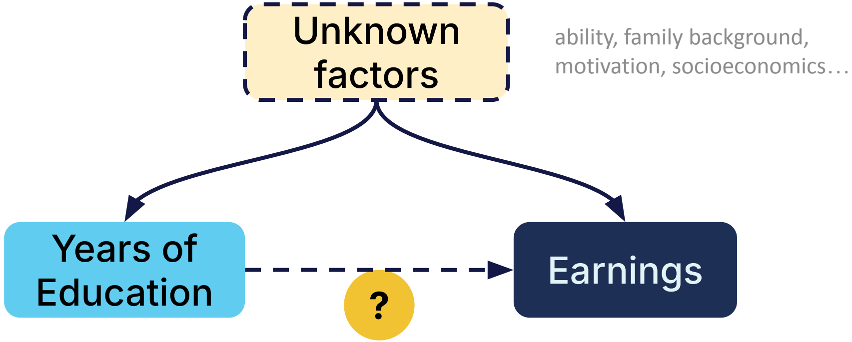
Intervention Analysis



Confounders - What if they're unknown?

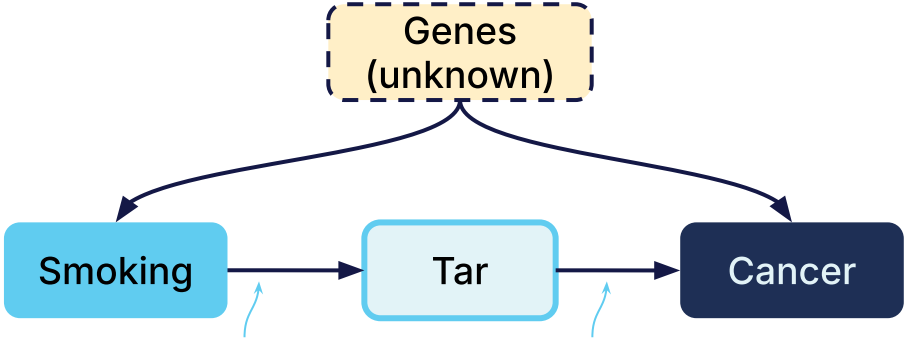
In most cases we actually know that there could be *latent confounders* - but we don't know how to quantify them!





Latent Confounders -
We Use Mediators and Front Door Adjustment

Even if we have latent confounders, we can estimate the impact of Smoking on Cancer by identifying mediators (**Tar**) that are independent of the confounding factors.



If we can establish that these arrows exist, independently of genes, then we can estimate the impact of smoking on cancer as simply the product of individual effects:

$$P[Tar|do(Smoke)] * P[Cancer|do(Tar)]$$

Mediators tell us about the *"intermediate mechanisms"* through which a potential cause is related to a potential effect.

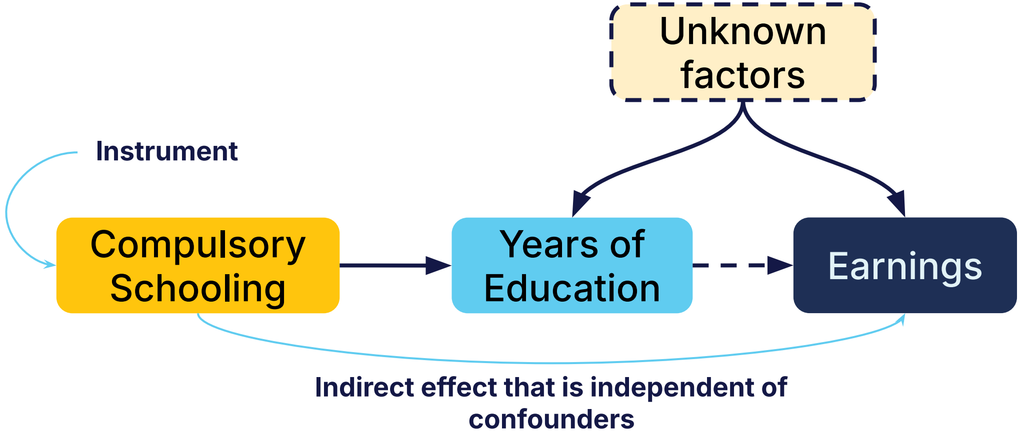
We state that a mediator is **shielded** whenever it is independent of any potential confounder that affects both the cause and the effect.

Instrumental Variables

202

We may not know the mechanism, but we may have "natural experiments" happening in the data that moc the treatment independently of confounders.

Maybe Hide? I'm a litt



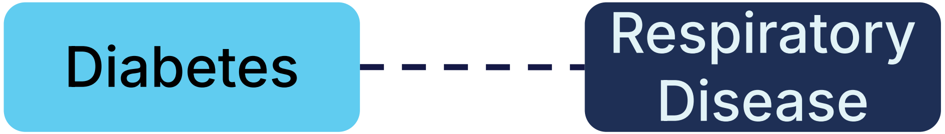
Instruments are variable treatment, but do not sh (confounder) with the outcome.

In this scenario, imagine that schooling laws were changed in certain states, but confounding factors are the same (socioeconomics, etc).

Then comparing different states allow us to remove the impact of the exogenous factors.

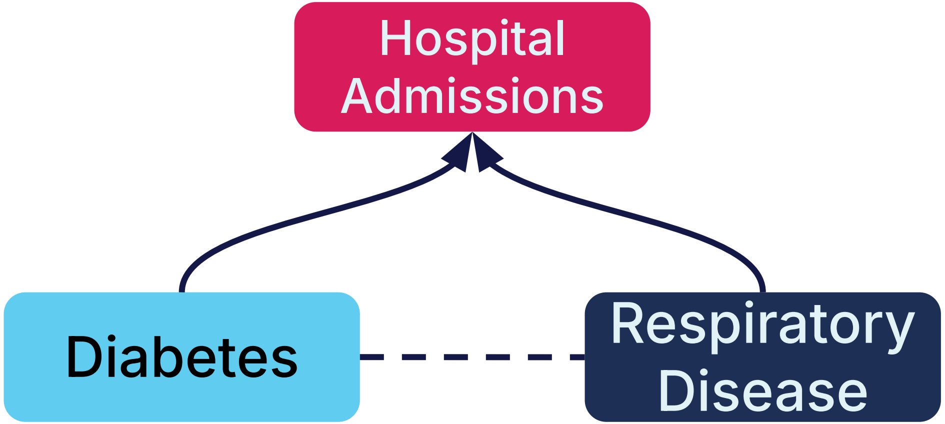
Collider - Should We Get the Flu?

Hospital admissions showed a negative correlation between diabetes and respiratory illness



Collider - Should We Get the Flu?

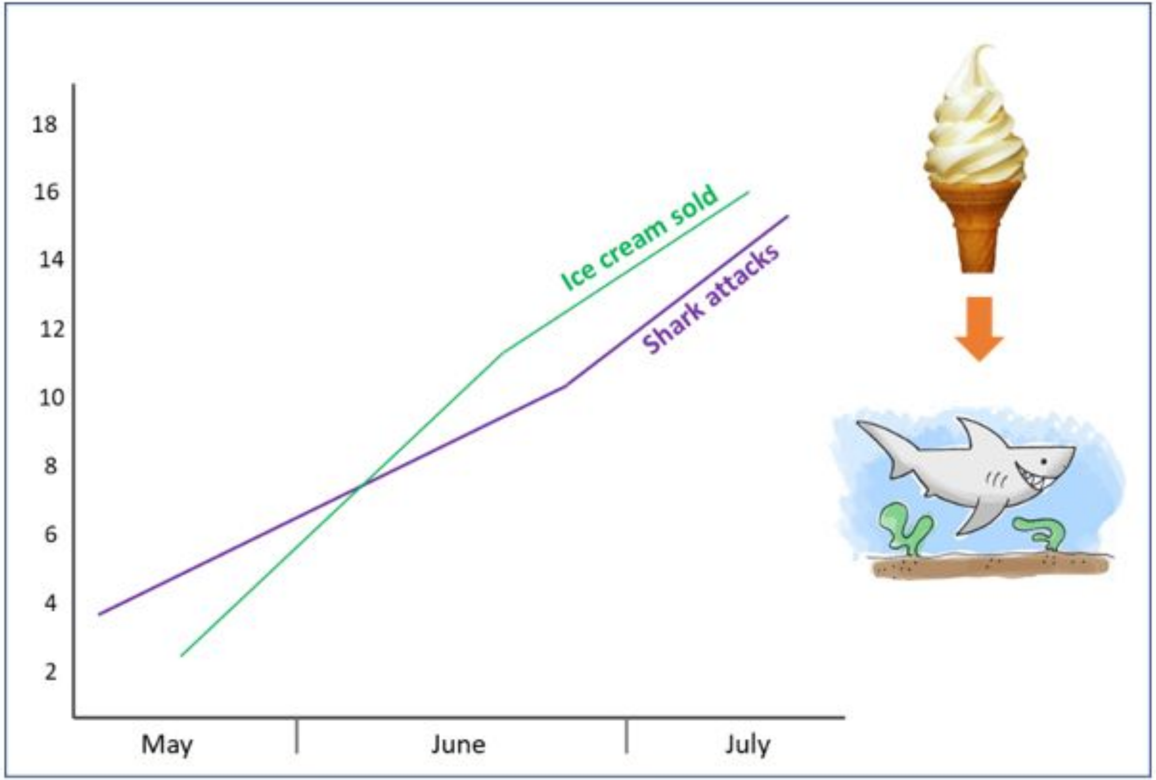
Hospital admissions showed a negative correlation between diabetes and respiratory illness



Conditioning on a collider, the hospital admissions, created a correlation between both risk factors, even though there's a priori no causal relationship between them.

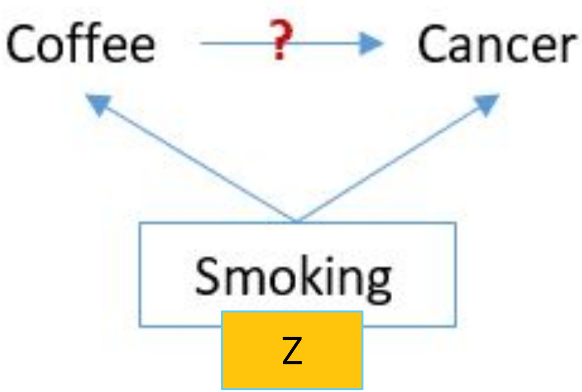
So, if you're running an ML analysis you should not add hospital admissions to your model

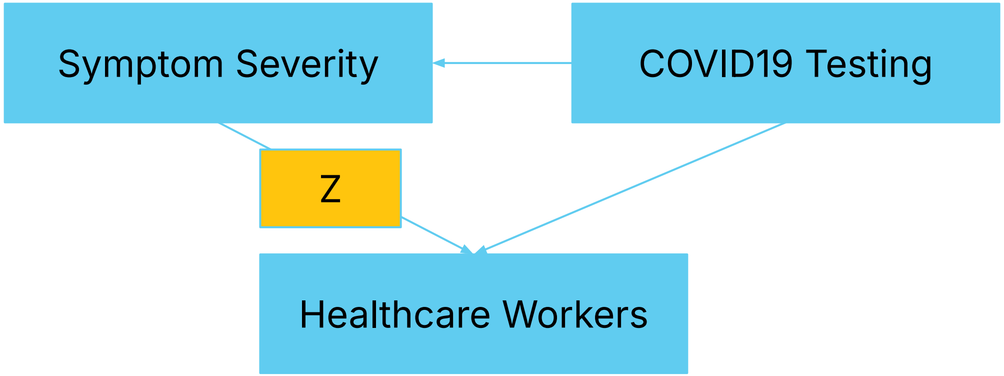
QUIZ TIME



QUIZ TIME

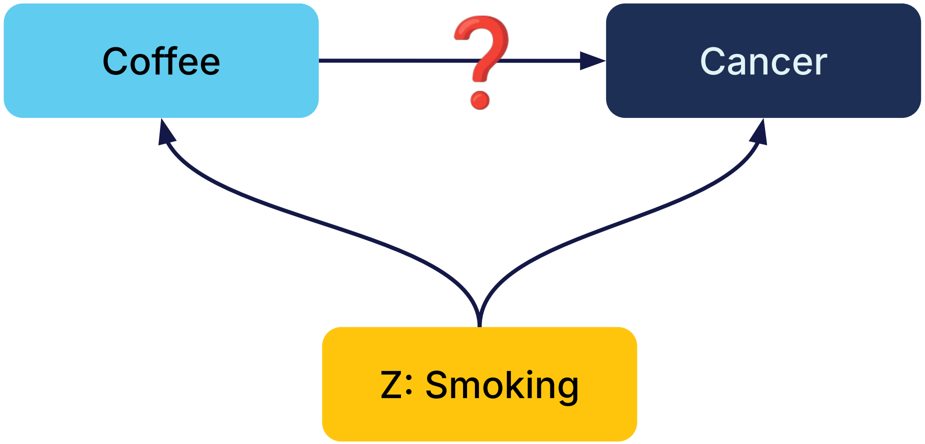
Is Z a confounder, collider and mediator?





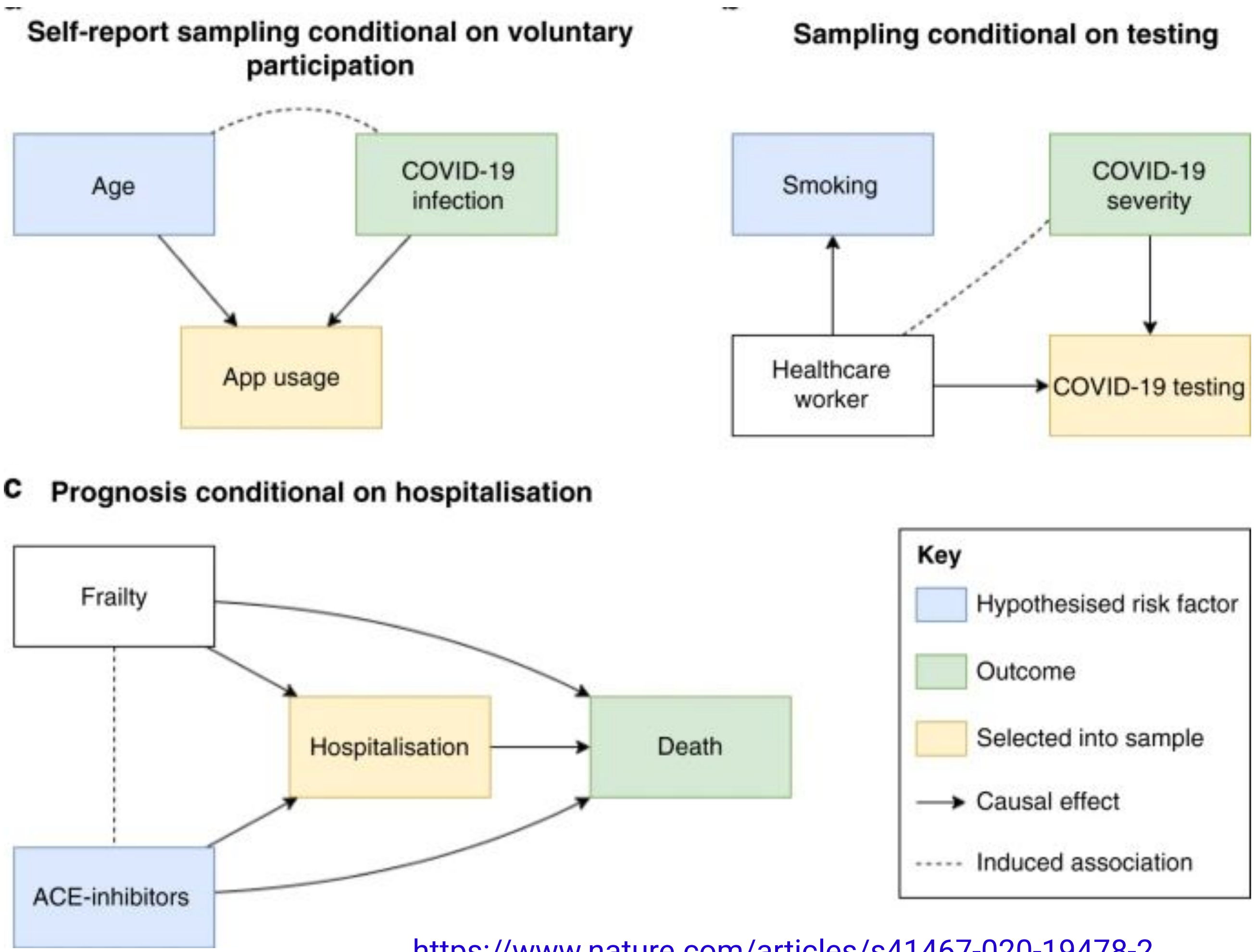
QUIZ TIME

Is Z a confounder, collider or mediator?



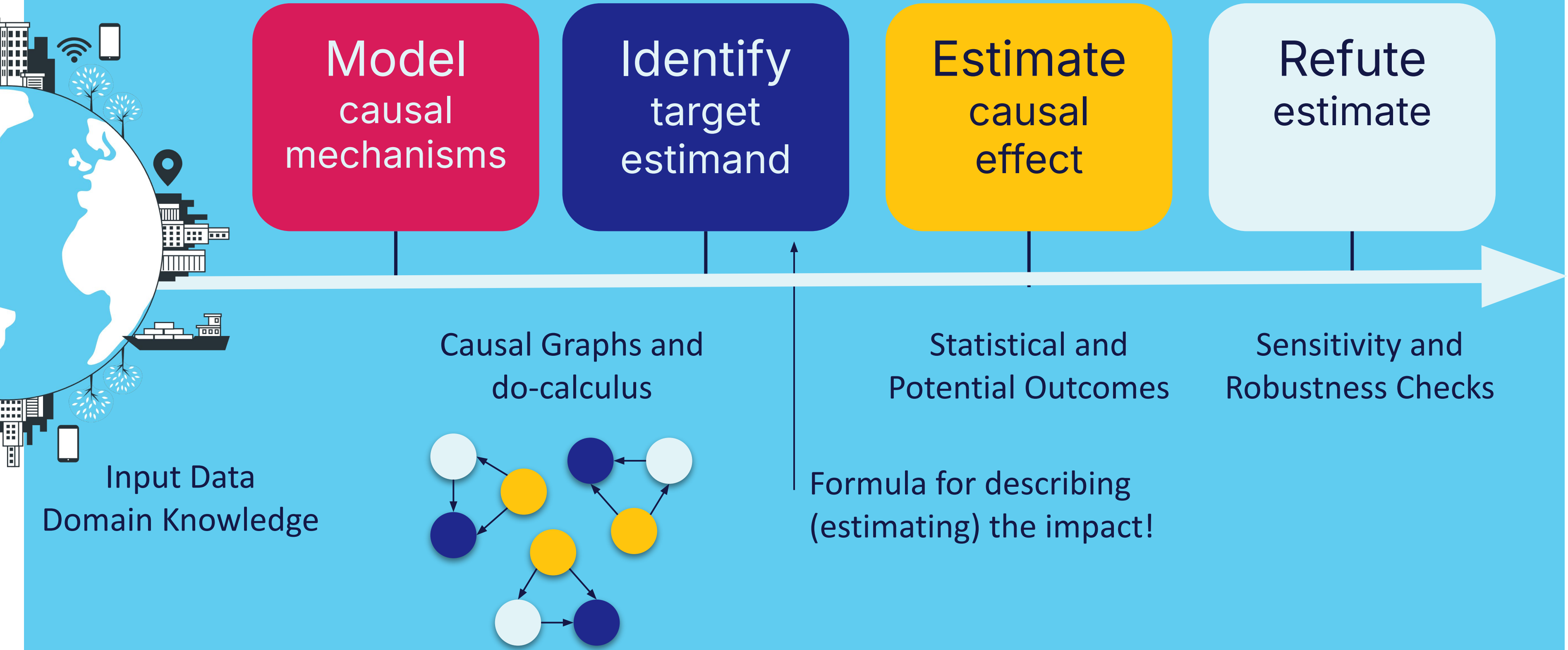
QUIZ TIME

Are the Yellow Boxes confounders, colliders or mediators?



https://www.nature.com/articles/s41467-020-19478-2

DoWhy Process for Causal Inference

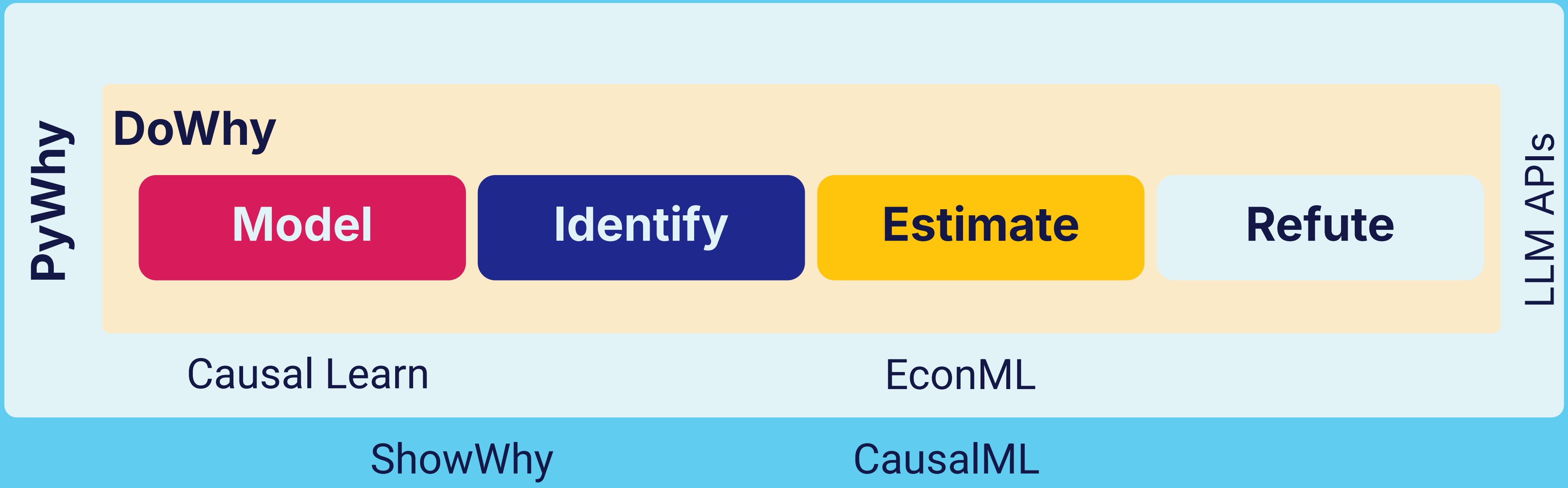


DoWhy, PyWhy, MagPy, and Other Tools

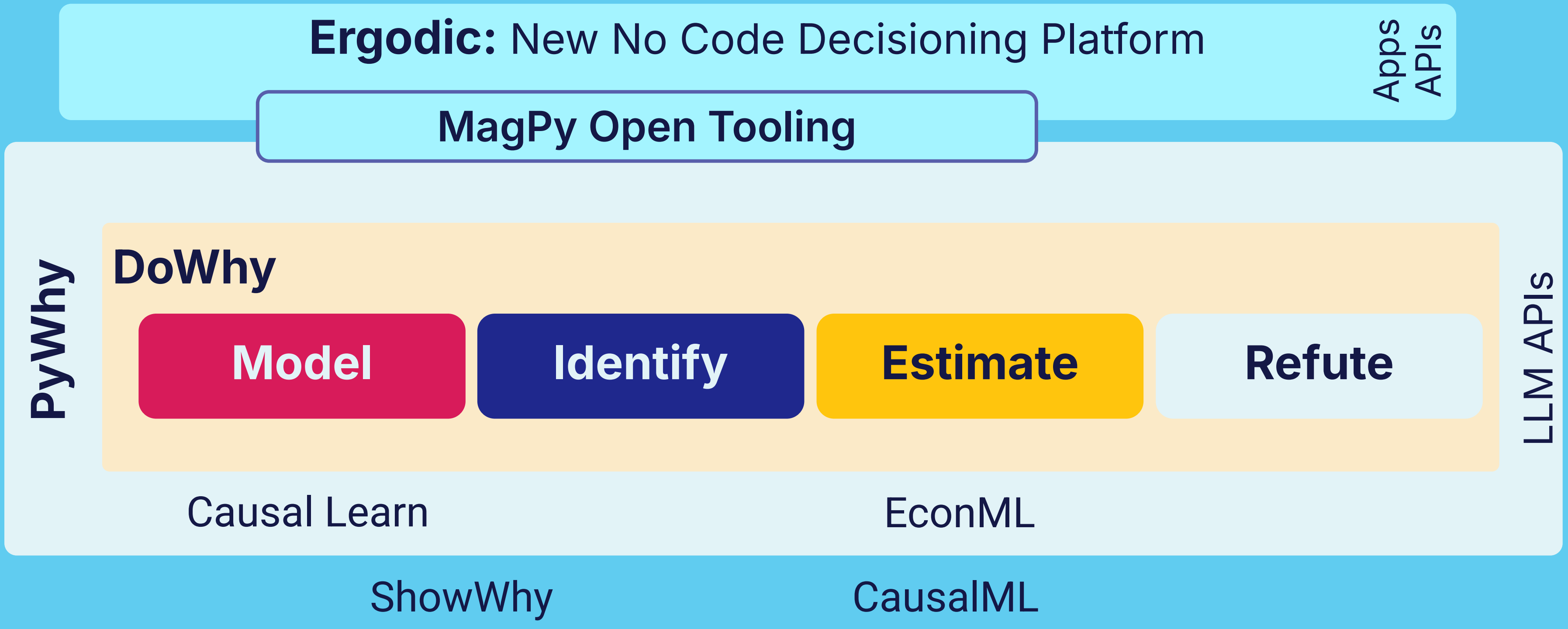
DoWhy

Model | Identify | Estimate | Refute

DoWhy, PyWhy, MagPy, and Other Tools



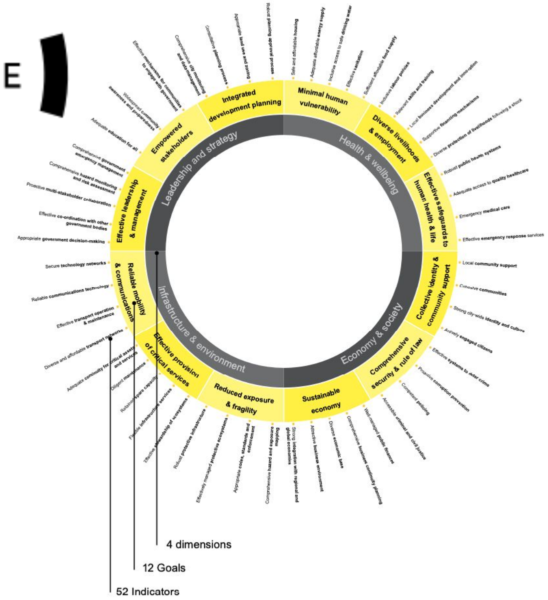
DoWhy, PyWhy, MagPy, and Other Tools



Can We Make Streets Safer?

From City Resilience to Traffic Accidents

City Resilience Index - Practical Framework
www.cityresilienceindex.org



Pared down big questions

- Subcomponents: Leadership, Infrastructure, **Health**, Economy
- Used a credible measurement framework
- Evaluated various data sources
- Map available data to measurable goals
  San Francisco Open City Data www.sf.gov/data/

Reducing the impact of traffic accidents as a **desirable, measurable, actionable** (hopefully) goal

The San Francisco Standard

Crazy driving, road deaths are rampant. SF candidates vow to crack down

By Gabe Greschler

Published Oct. 16, 2024 • 5:42pm

What would have the biggest impact to traffic safety in San Francisco?

Is there a clear intervention that would reduce accident severity?

What policy changes might help save lives?

Cleaned Accident Data

Vehicle Group

Road Conditions

Time of Day

Collision Type

Weather Conditions

Day of Week

Proceeding Action of 1

Light Conditions

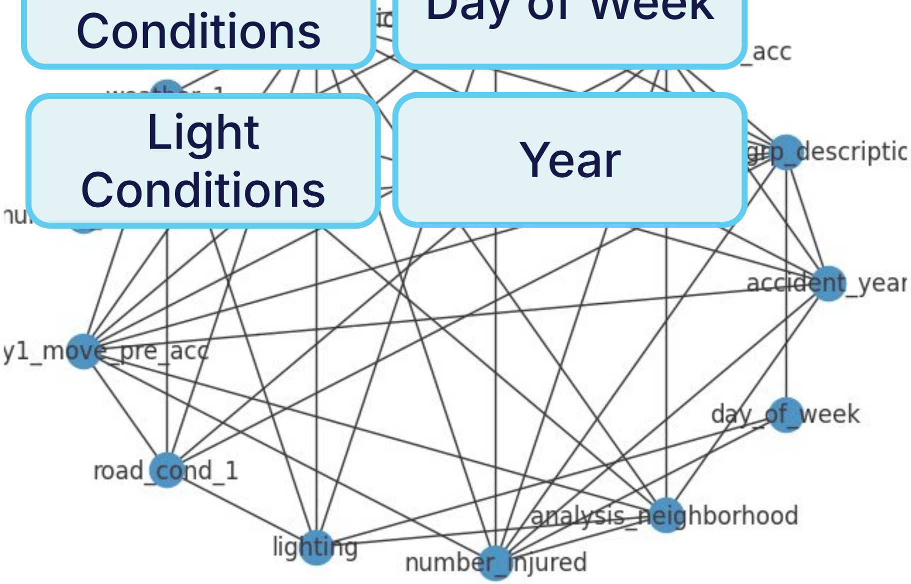
Year

Proceeding Action of 2

Fatalities

Collision(injury) Severity

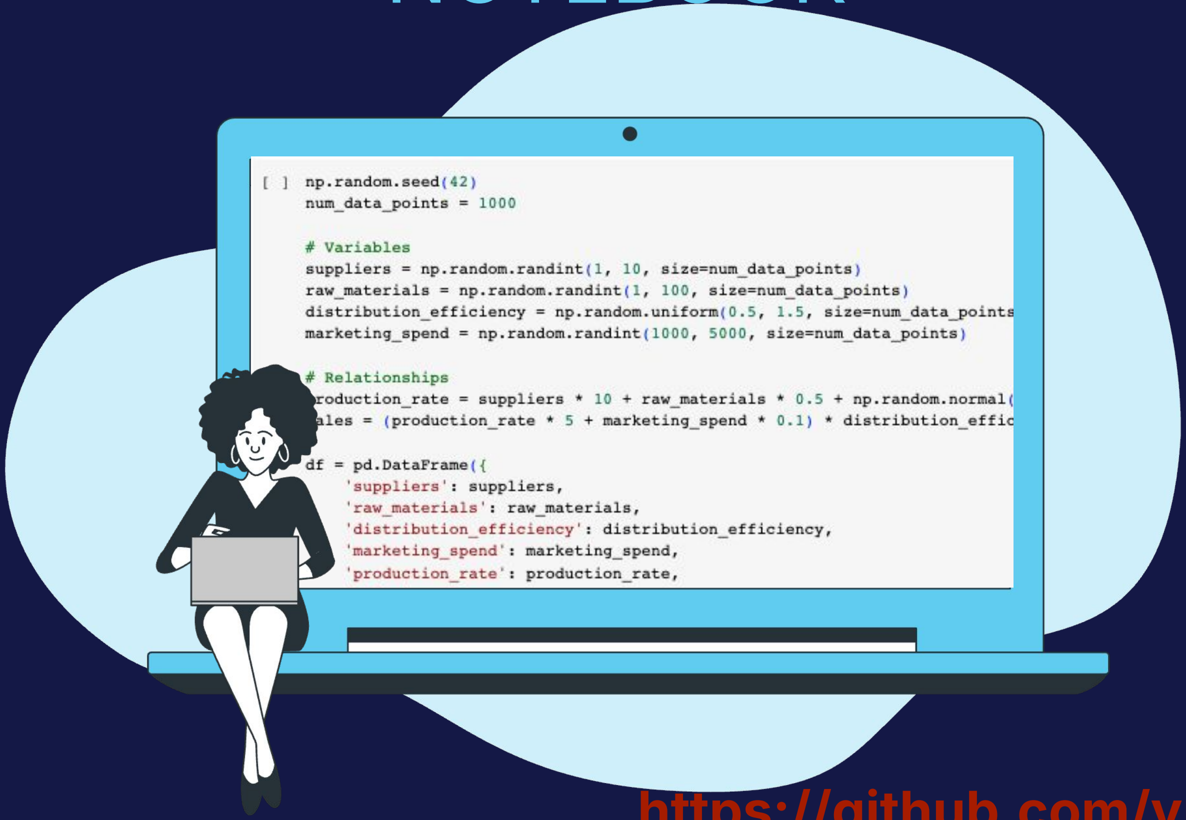
Injured



Our Approach

Model	Identify	Estimate	Refute
Model causal relationships as a graph comparing different methods Consider/update for confounders	Evaluate different data types to measure the effect of variables on injury severity and fatalities	Find the causal strength of interventions on injury severity and fatalities	Test robustness of findings by adding randomness, checking new vs estimated effect and the p values
PyWhy: GCM MagPy: PC Algo: skeleton (A* not for categorical)	MagPy: PC Algo: Mixed Data Oracle	PyWhy: Causal Arrow MagPy: Random Forest	PyWhy

NOTEBOOK



https://github.com/yulleyi/ddt2025

Causal Insights: Reducing Accident Severity

Lighting is a major influence on severity of injuries and fatalities

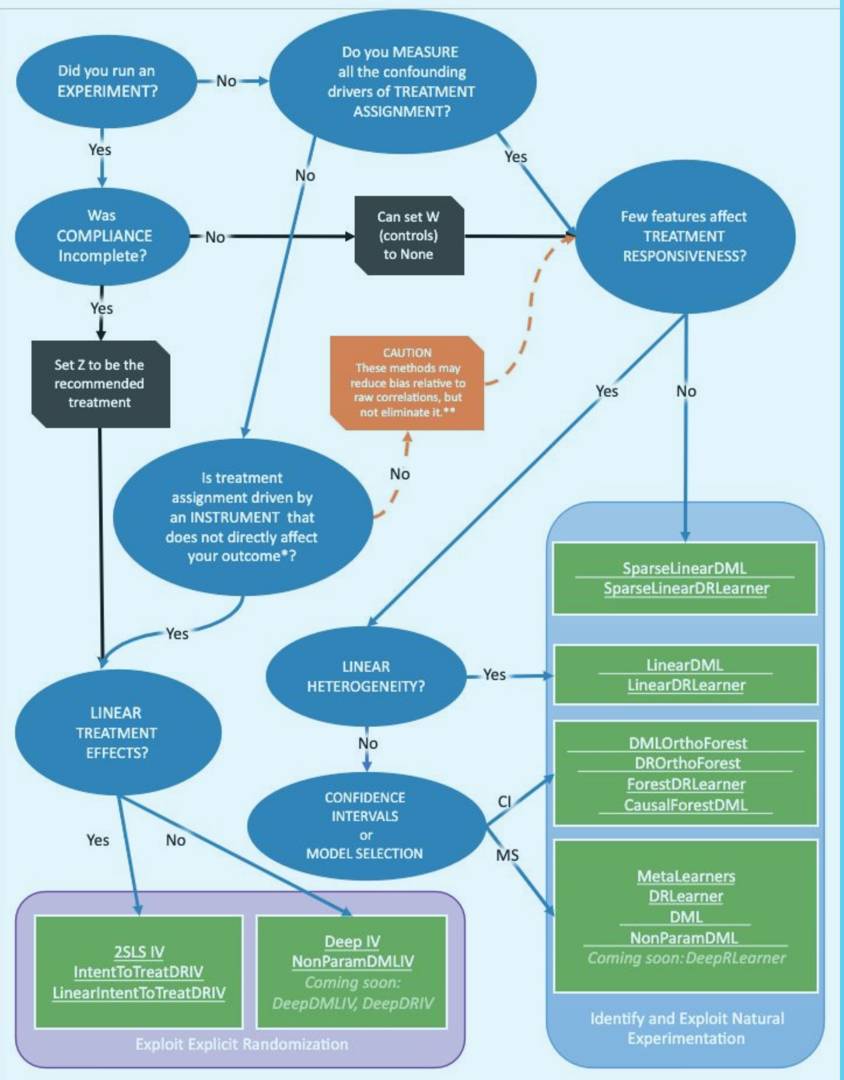
A simulation showed the potential to save over 700 lives if lighting issues were eliminated (5% reduction)

- Consider the additional injuries, services, and economic impacts!

Biggest reduction in high severity accidents is car-only accidents vs pedestrians (U-Turns & Driving Off the Road)

Lots of opportunity for further analysis, connecting more data sources and evaluating change over time

What We Learned



Try Not To Get Overwhelmed

Start Simple and Add Layers

DoWhy

- Graph Causal Model (GCM) - New
- Find relationship strengths
- Simple models (linear regression)
- Limited to averages, not a lot of tuning

MagPy

- Causal discovery & modeling
- Testing of the causal graph
- Estimate interventions
- Mixed variable types

EconML

- Play with other models (🌲🌲, DML, +)
- More & deeper controls, cross validation
- No robustness testing → So fit the log back in DoWhy

Don't Forget

Include Domain Experts!
Confirm assumptions on causal links, missing data, confounders

Questions Matter A Lot
Changes data used & causal graph
You will iterate!

Missing Data
Beware of spurious correlations and deal with confounders

Categorical Data is Cool but...
More difficult/slower to evaluate
Mixed variables easier in MagPy

Causal Graphs Take Time
Use tools like MagPy, CausalGraph, Casual Learn

One Outcome at a Time
Can't observe the treatment and not having a treatment in one pass

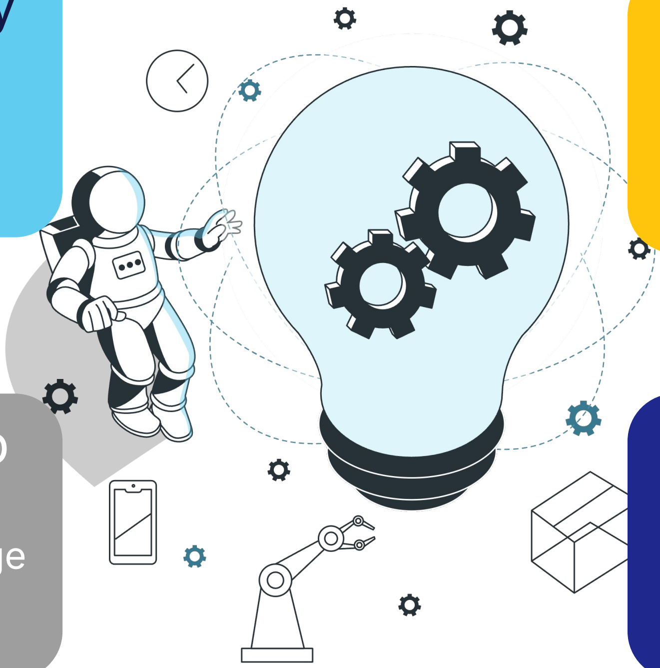
Causal Inference is Required for Better Decisions

Answering Why

Distinguish cause and effect
Find biggest influencers
Inherent explainability

Actionable Results

What changes outcomes?
Estimate impact of an action
Supports decision intelligence



Human-in-the-Loop

Closest to human inference
Integrates domain knowledge
Emphasize human action

Detecting Problems

Uncover poor associations
Is more data needed?
Highlight biased features

Smart Teams Are Adopting Causal Approaches

"Causality is very important for the next steps of progress of machine learning," said Yoshua Bengio, a Turing Award-wining scientist known for his work in deep learning, in an interview with *IEEE Spectrum* in

The complex math of counterfactuals could help Spotify pick your next favorite song

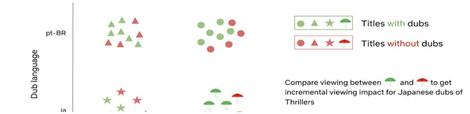
A new kind of machine-learning model is set to improve automated decision making in finance, health care, ad targeting, and more.

Data / ML

Using Causal Inference to Improve the Uber User Experience

Netflix Technology Blog in Netflix TechBlog
May 21, 2022 · 6 min read

But we can match *similar* titles, some of which are with dubs while others are without.



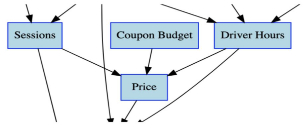
A Survey of Causal Inference Applications at Netflix

Salesforce CausalAI Library: A Fast and Scalable Framework for Causal Analysis of Time Series and Tabular Data

airbnb | Microsoft | NETFLIX | Uber | Spotify | amazon advertising | salesforce | Walmart | shopify | ...

bp's Causal Inference Symposium: Discussing the next frontier in AI

DJ Rich in Lyft Engineering
Jun 28, 2022 · 8 min read



Causal Forecasting at Lyft (Part 1)

Microsoft | Research

Project Causica:

Project Causica: Decision Optimization with Causal ML

Shopify Engineering

How to Use Quasi-experiments and Counterfactuals to Build Great Products

but they often come with the caveat that *correlation isn't causation*. At Shopify, we believe that understanding **causality** is the key to unlocking maximum business value. We aim to identify insights that actually indicate *why we see things in the*

Walmart Global Tech

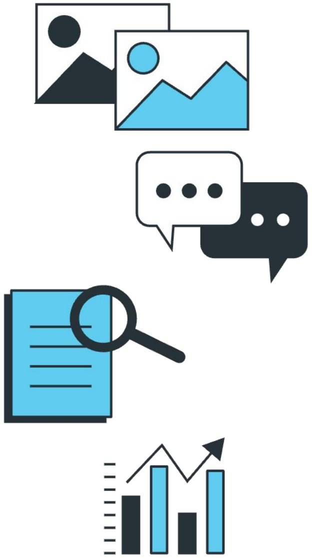
Walmart's scientific approach to evaluating marketing campaign effectiveness is centered on the 'causal inference' method, which determines if a change in one variable causes a change in another. It differentiates between correlation (two events occurring together) and causation (one event causing another). For instance, if a business has an ad campaign followed by increased sales, causal inference methods help establish if the ad caused the sales increase.

Resources

- Finding good data is a challenge but DoWhy has some curated
  - Synthetic Healthcare Data Example - Synthea
  - SF Open City Data www.sf.gov/data/

Slides and Notebook - https://github.com/yulleyi/ddt2025

- PyWhy pywhy.org/
- MagPy github.com/ergodic-ai/magpy/
- Become an early user of Ergodic! - ergodic.ai/
- Helpful causal inference guide tinyurl.com/5e4auenh
- PyWhy Discord tinyurl.com/365d7e37
- GraphGeeks Discord tinyurl.com/hrjanc3p



GraphGeeks

Start Asking *Why*

Special Thanks 🙏

Andre Franca
Patrick Blöbaum
Jason Grafft

Michelle@Generationship.ai
Amy.Hodler@GraphGeeks.org

Available for Training

PARTY! 6:30 PM in the Bar

GraphGeeks