GRAPHS

The Next Frontier of GenAI Explainability

Michelle Yi & Amy Hodler
Fall 2023

Michelle Yi

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

Amy Hodler

Graph Evangelist, Author, Speaker
Founder of Graph Geeks
Amy@GraphGeeks.org

Today

- **Background**: Terminology, Landscape
- **Hands On**: Hypothesis & Tools, Collab Notebook
- **Tips**: Learnings, Resources
- **Questions**: What did we miss?

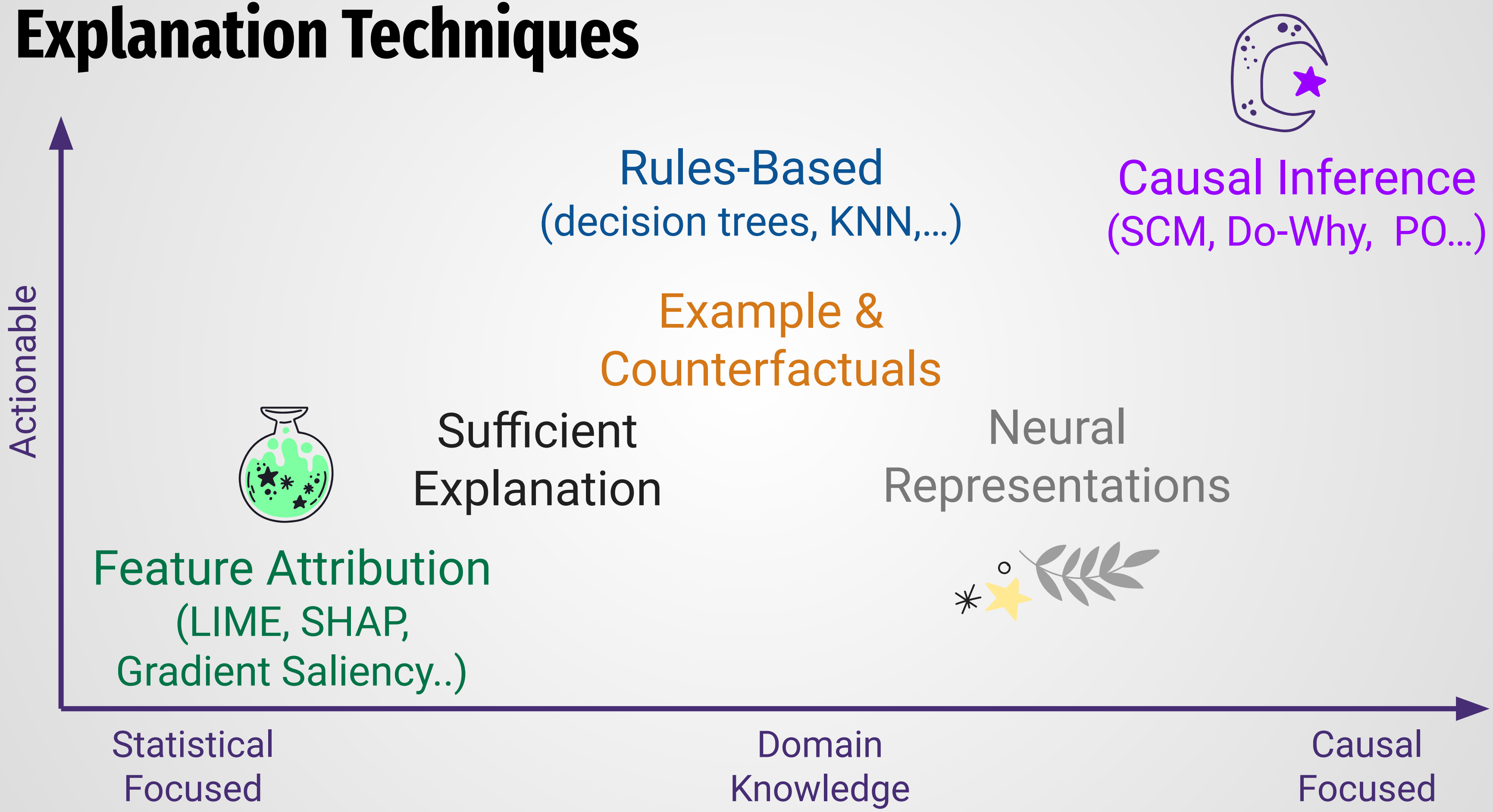

Quick
Review

Explainable AI (XAI) Today

Why is this model giving me this prediction?

- Uncovers strong co-occurrences in data
- Good at finding feasible actions that would change a prediction
- Quick diagnosis of problems in data
- Not great with data drift
- It doesn’t actually explain anything

Explanation Techniques

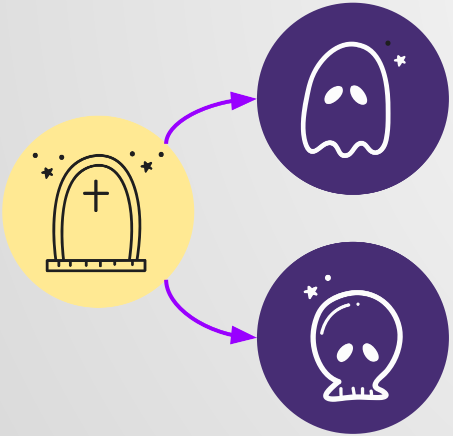


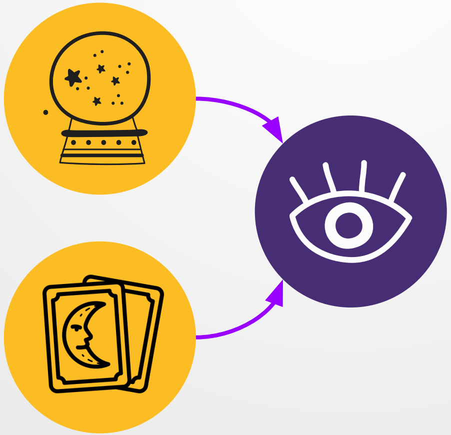
Causality: Another Layer of Explainability

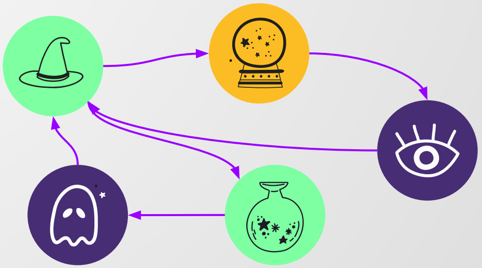
Why does this outcome happen?

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

Different Types







We Need Causal Inference

Answering Why

What changes outcomes?
Overall effect of intervention?
Inherent explainability

Better Results

Improve accuracy
More actionable results
Supports decision intelligence

Detecting Problems

Uncover poor associations
Is more data needed?
Highlight biased features

Human-in-the-Loop

Closest to human inference
Integrates domain knowledge
Emphasize human action

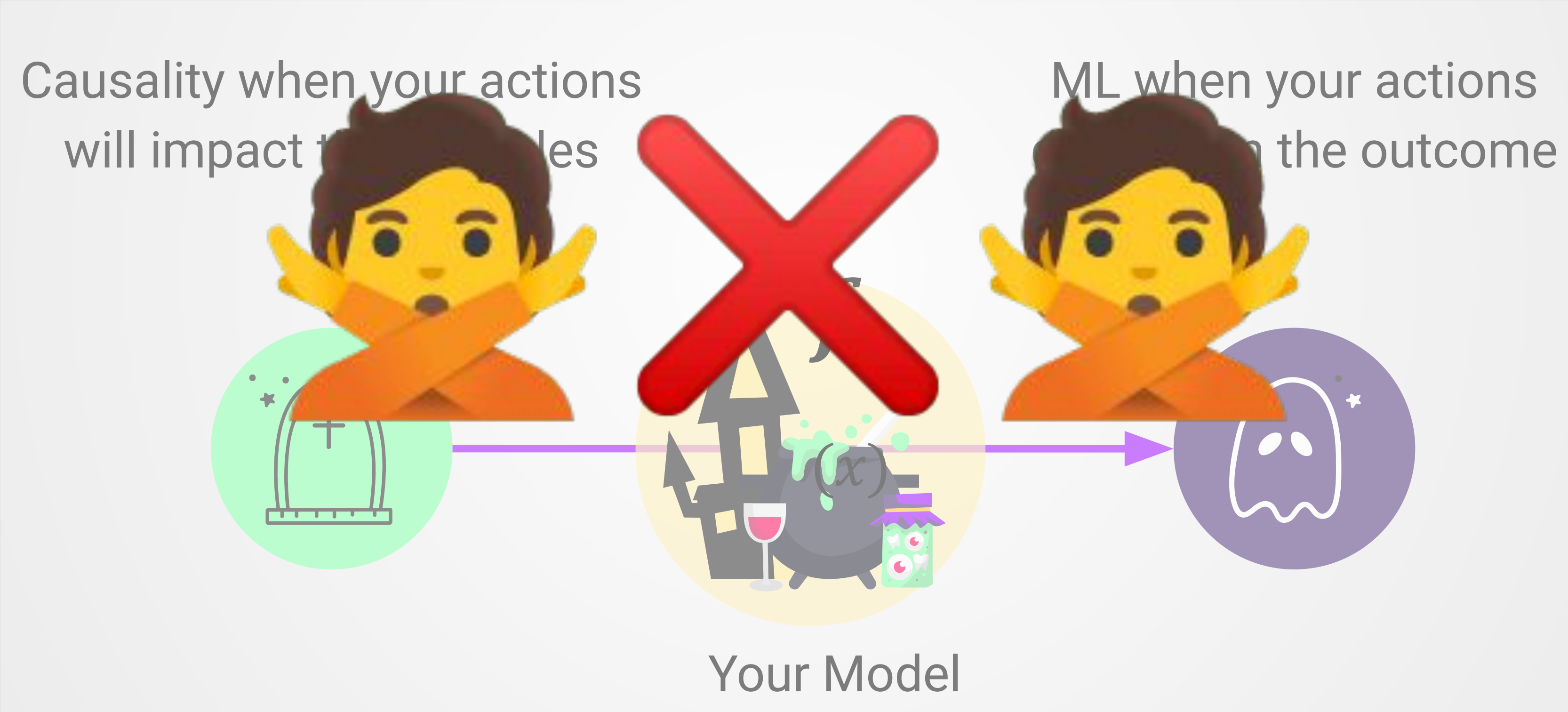
When To Use Causality vs Machine Learning?

Causality when your actions will impact the variables

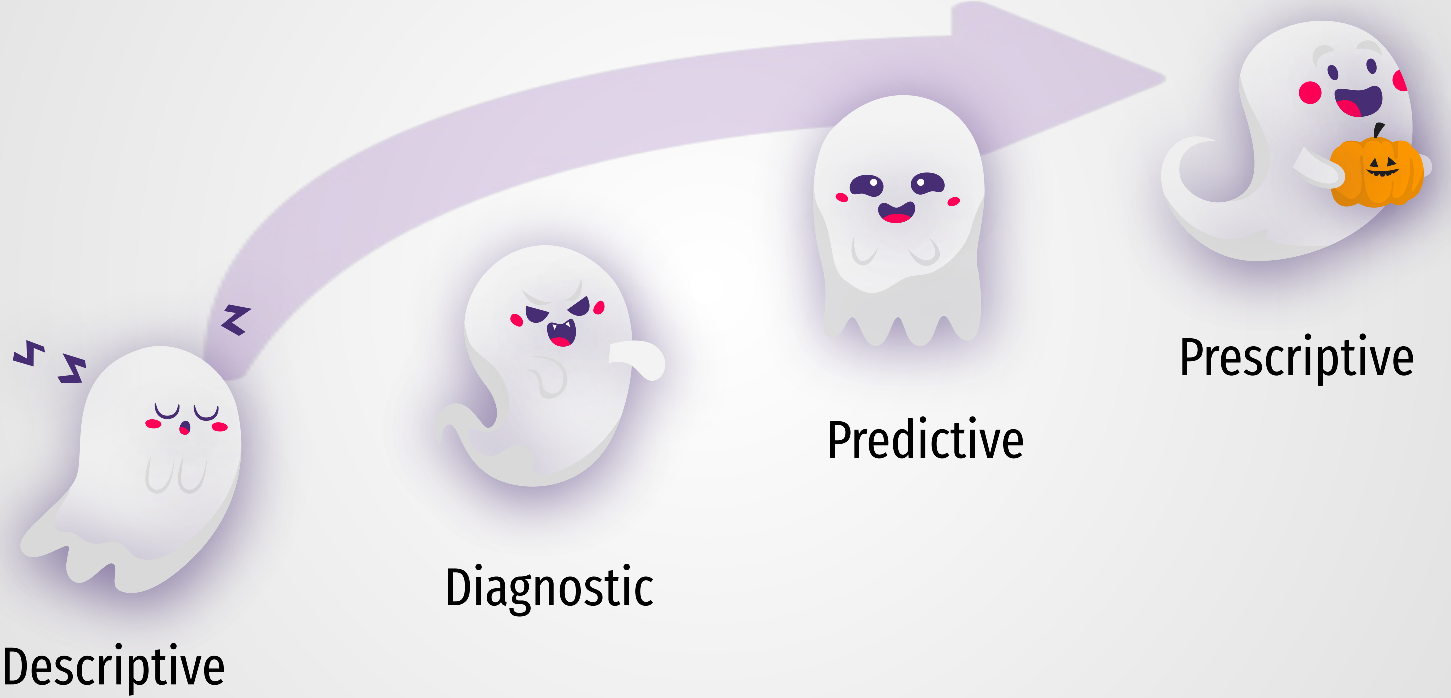
ML when your actions depend on the outcome

Your Model

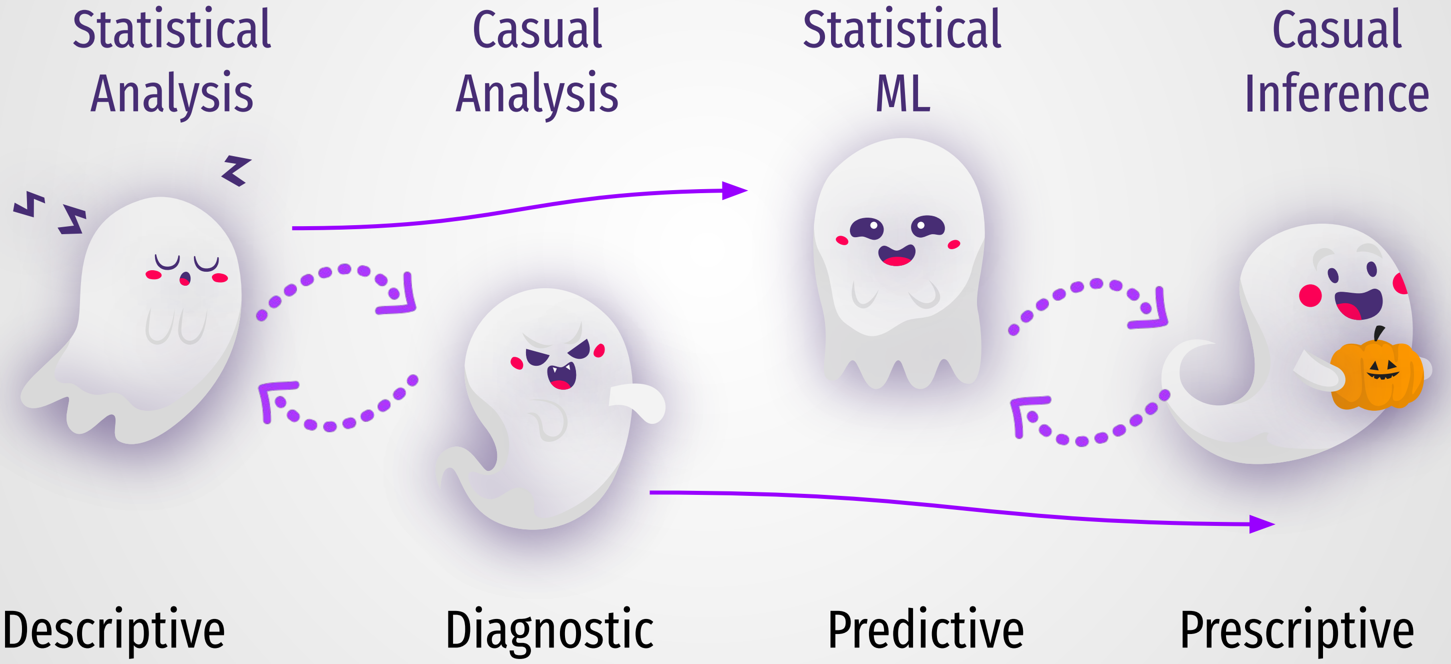
When To Use Causality vs Machine Learning? NOT



Is Causality a Progression?



Causal + Statistical Approaches Are Complementary

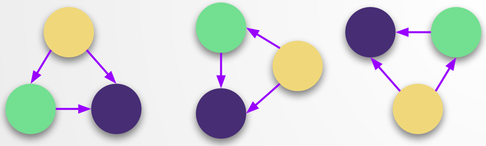


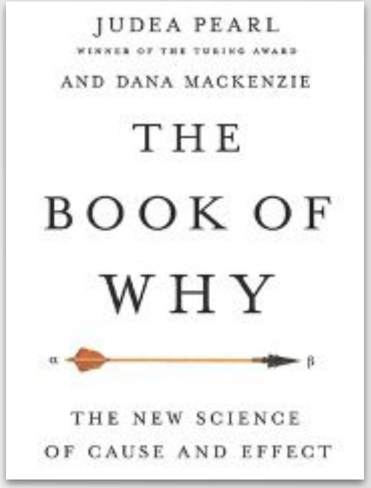

Putting the Ingredients Together

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

$P(Y|X)$

to

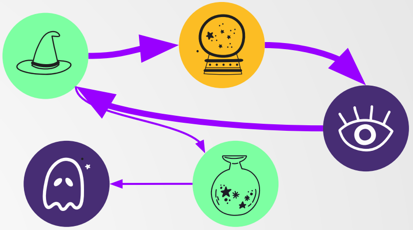
$P(Y|do(X))$

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

Causal Graphs as a Unifying Model

Directed Acyclic Graphs (**DAG**) ONLY

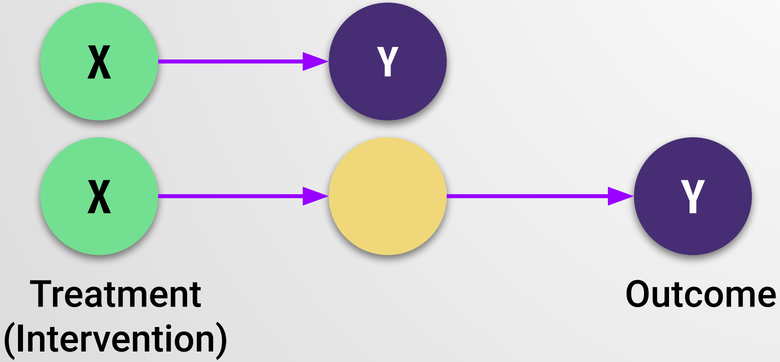
No feedback loops: no intervention can cause itself!



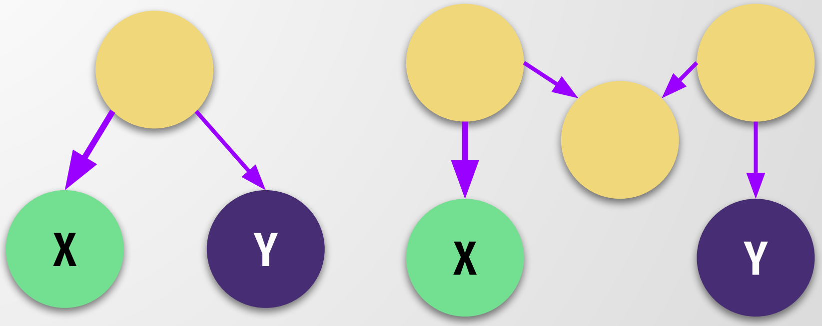
Cycles NOT Allowed

Arrow direction illustrates causal influence

Front Door Paths

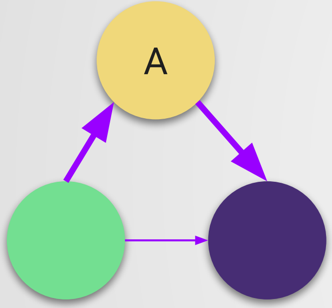


Back Door Paths

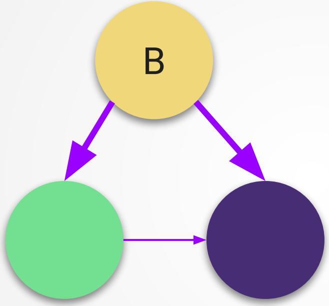


Causal Graphs as a Unifying Model

Basic structures



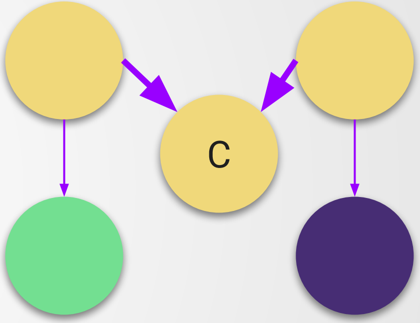
A mediator



B confounder

Often Controlled

if pointing to
treatment & outcome



C collider

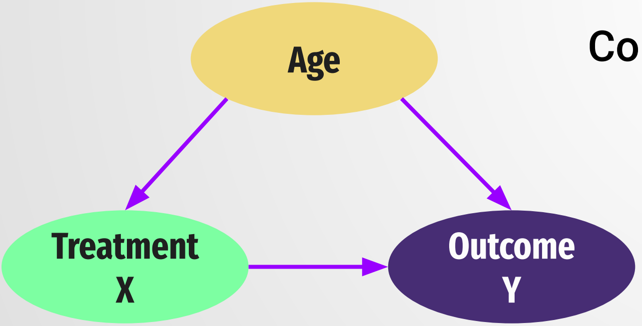
Avoid Controlling

As it introduces correlations
that require further controls

Confounders - Back Door Adjustments

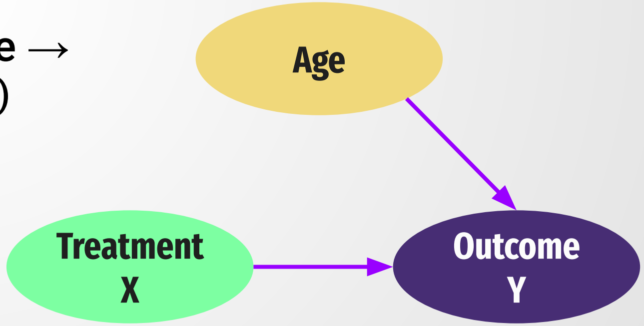
Control Node (CN) Criteria: 1) CN must not be a child of the treatment your estimating and 2) CN must block the path between treatment and outcome.

Observed Data



Control on Age →
(hold steady)

For Intervention Analysis



we've "closed" the backdoor

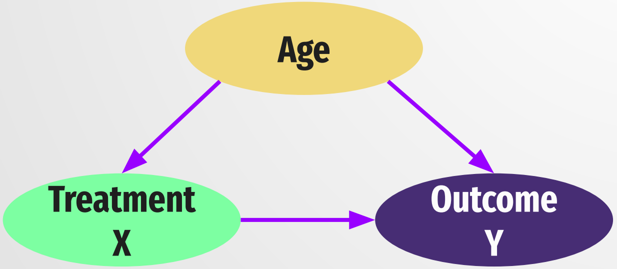
do-calculus

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

Replaces probability formulas with constant variables to simulate the change in a controlled manner

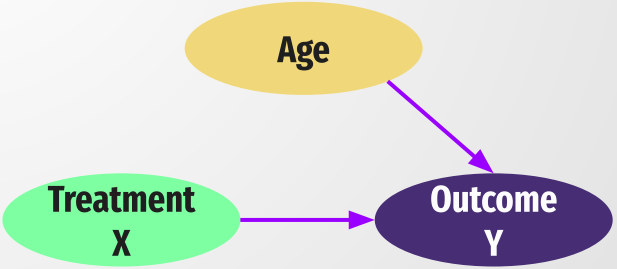
Interventions & counterfactuals are represented by the operator, **do(x)**, which “erases” the function while keeping the rest of the model unchanged

Observed Data



$P[Y|X]$

For Intervention Analysis

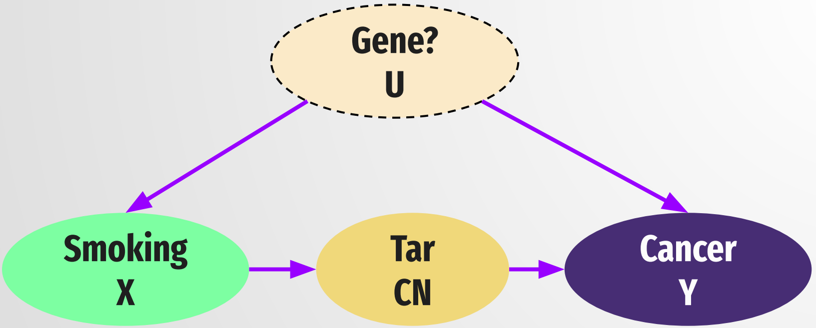


$P[Y|do(X)]$

Front Door Adjustments (if you can’t do a back door)

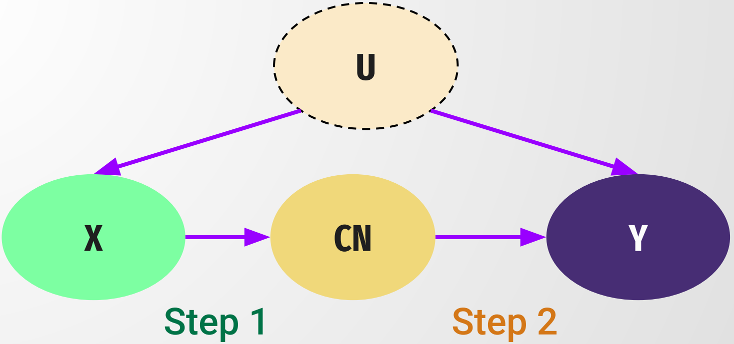
Control Node (CN) Criteria: 1) CN must intercept all directed paths from treatment to outcome 2) no unblocked backdoor path from treatment to CN and 3) all backdoor paths from CN to outcome are blacked by the treatment.

Observed & Unknown Data



Shielded Mediator

For Intervention Analysis



Step 3 - Sum over Mediator and chain together

$$\sum_z P[Z|do(X)]\ P[Y|do(Z)]$$

DoWhy for Causal Inference

Formula for describing (estimating) the impact!

Model causal mechanisms

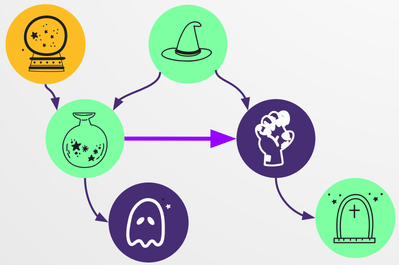
Identify target estimand

Estimate causal effect

Refute estimate

Input Data
Domain Knowledge

Causal Graphs and do-calculus



Statistical and Potential Outcomes

Sensitivity and Robustness Checks

LLM API

LLM API

LLM API

PyWhy – Brings It All Together

Model	Identify	Estimate	Refute
DoWhy	DoWhy	DoWhy (basic)	DoWhy
Casual Learn (discovery & investigation)		EconML (deeper)	

ShowWhy 😢😢😢

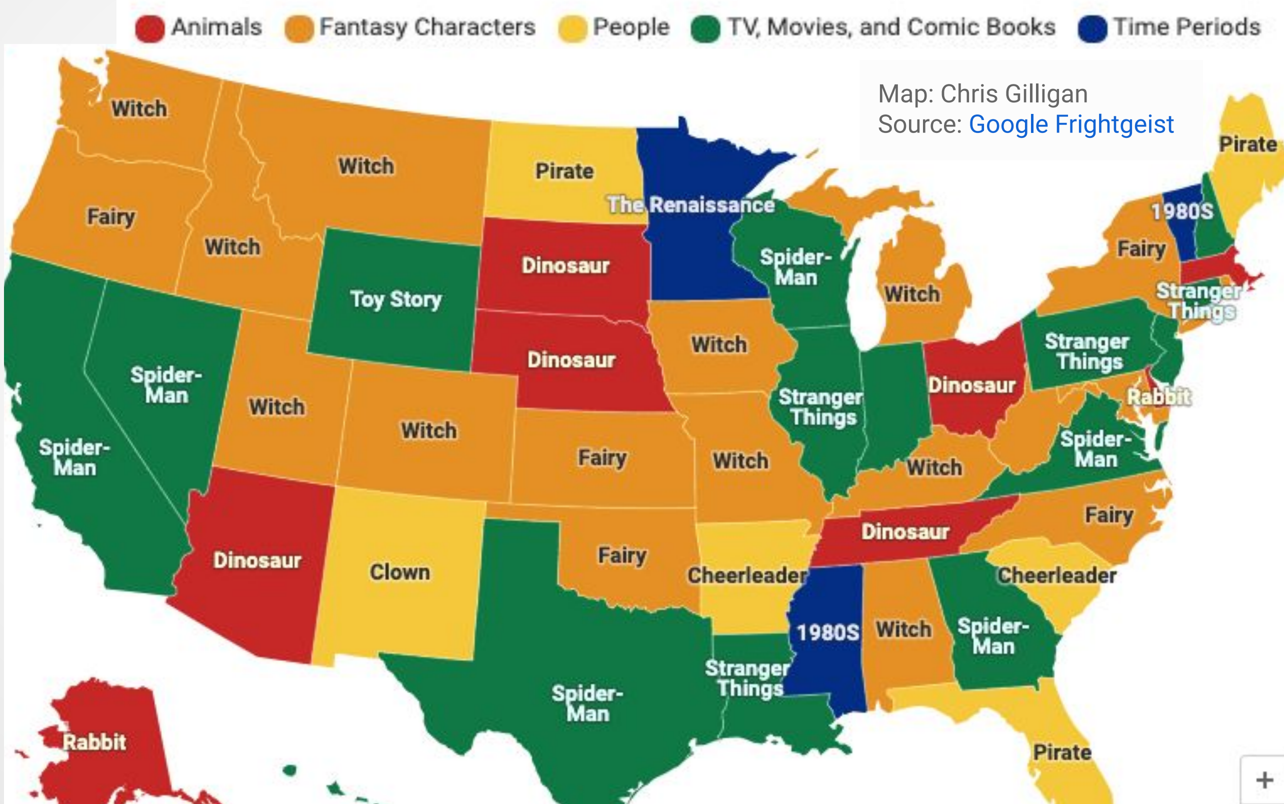
CausalML (deeper)


Today’s Scenario

Halloween Costume Vendor Wants to Increase Sales

We have lots of data and experience…and even more opinions!

Top Costume Searches in 2022



We're Data Scientists: We Can Help Settle the Argument

What if we changed where we focus our resources this year?

What changes have the most impact to sales?

Model

Costume Supplier (branded/not)

Costume variety

Production rate

Material (quality)

Pricing

Sales (output)

Identify

Use ATE (ave. treatment effect) to measure the effect of these variables on sales.

SME input → Updates

Estimate

Find the causal strength of interventions on sales using linear regression.

Then use EconML for Causal Forest

Refute

Test robustness of findings by adding randomness, checking new vs estimated effect and the p values.

Present Results!

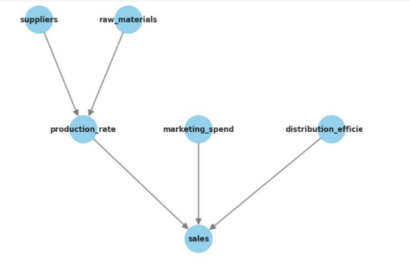
Let's DO This!

```
[ ] np.random.seed(42)
    num_data_points = 1000

    # Variables
    suppliers = np.random.randint(1,
    raw_materials = np.random.randint
    distribution_efficiency = np.ran
    marketing_spend = np.random.rand

    # Relationships
    production_rate = suppliers * 10
    sales = (production_rate * 5 + ma

    df = pd.DataFrame({
        'suppliers': suppliers,
        'raw_materials': raw_materia
        'distribution_efficiency': distribution_efficiency,
        'marketing_spend': marketing_spend,
        'production_rate': production_rate,
        'sales': sales
    })
```



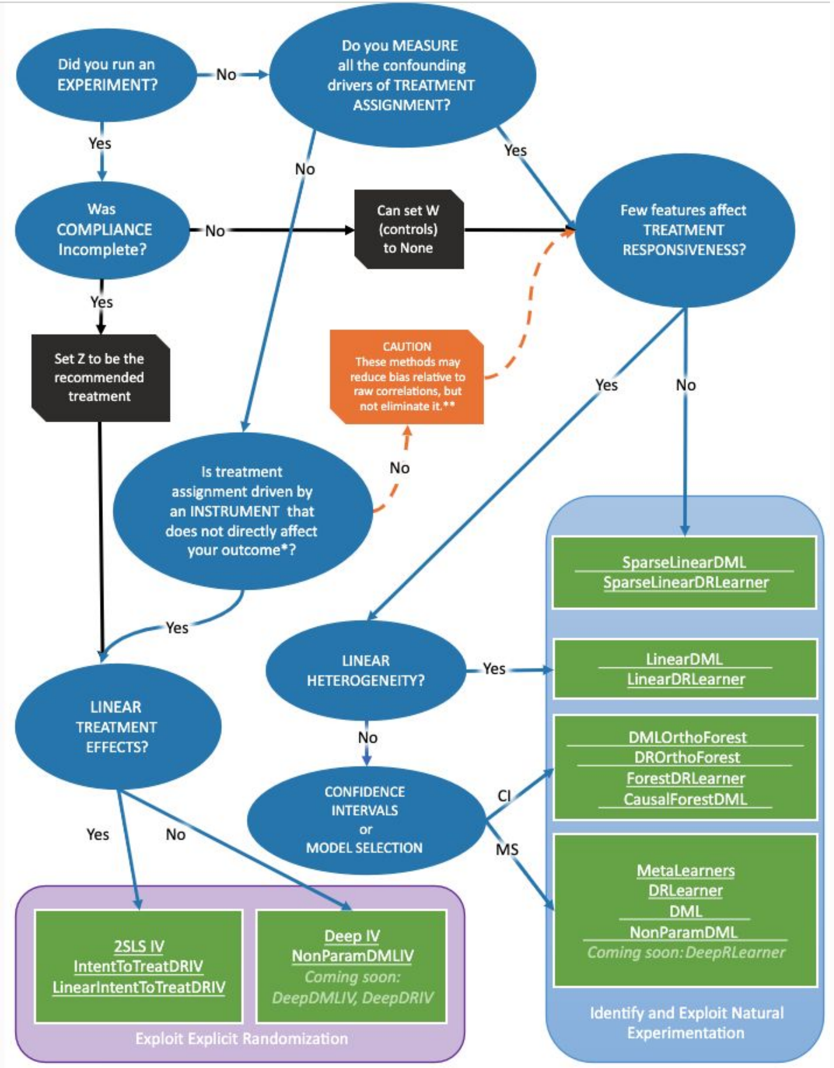
github.com/yulleyi/causal-graph-pywhy

copy or download the notebook


What We Learned

Um...Yeah, This Isn't Helpful

(maybe for more advanced?)



Start Simple and Add Ingredients!

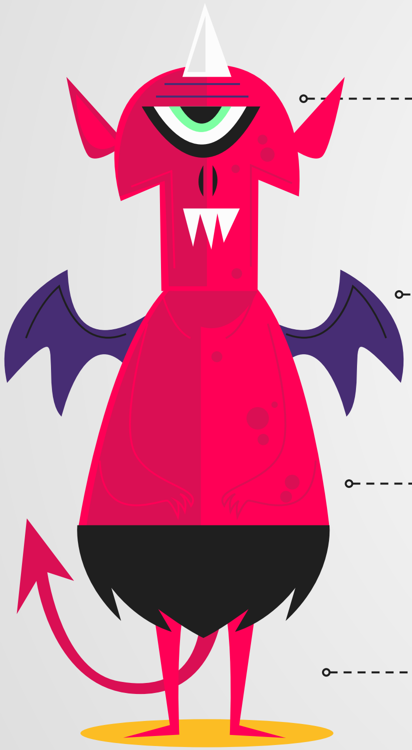
DoWhy

- Find the strength of different relationships
- Simple models (linear regression)
- Limited to averages, not a lot of tuning

EconML

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

Causality Limitations / Considerations



Missing Data
Missing confounders or other data left out can cause spurious correlations

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

Domain Expertise
Need to confirm your causal links, possible missing data, and confounder assumptions

Graph Model
No auto-magical creation, use your SME!
Causal Learn can help

Final Tips and Resources

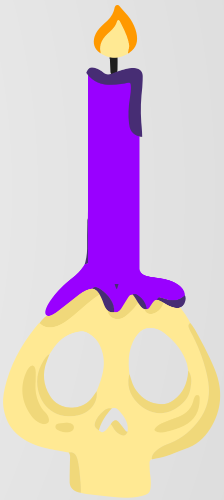
- Don't be intimated, the math and implementations are done for you!
- Finding good data is a challenge but DoWhy has some curated data
- There are multiple ways to do things, just try it out

Slides and Notebook - **github.com/yulleyi/causal-graph-pywhy**

Causal Resources

- PyWhy pywhy.org/ (with links to DoWhy and EconML)
- Another causal inference guide tinyurl.com/5e4auenh
- Discord community is very helpful! tinyurl.com/365d7e37

New Graph Community: GraphGeeks.org & Discord tinyurl.com/hrjanc3p



Thank You

Patrick Blöbaum

Jason Grafft

Questions?

Michelle@generationship.ai

Amy@GraphGeeks.org

Amit Sharma, Emre Kiciman. DoWhy: An End-to-End Library for Causal Inference. 2020. https://arxiv.org/abs/2011.04216

Patrick Blöbaum, Peter Götz, Kailash Budhathoki, Atalanti A. Mastakouri, Dominik Janzing. DoWhy-GCM: An extension of DoWhy for causal inference in graphical causal models. 2022. https://arxiv.org/abs/2206.06821

Real Use Cases??

tinyurl.com/48dk6bmm