UNIVERSITAT DE BARCELONA

CAUSAL INFERENCE AND MACHINE LEARNING

DataScience Lab

Dr. Jordi Vitrià,
Full Professor,
Departament de Matemàtiques i Informàtica,
Universitat de Barcelona.

Roger Pros,
PhD student
in Causal Machine Learning at the Universitat de Barcelona & Data Scientist at Zenital.

Enrique Mora,
PhD student in Causality and Reinforcement Learning at the Universitat de Barcelona & Tech Lead at Nestlé Global IT Innovation.

About the course

The relationship between causality and artificial intelligence can be seen from two points of view: how causality can help solve some of the current problems of AI and how causal inference can leverage machine learning techniques. In this course we will review the two points of view with special emphasis on examples and practical cases.

01 Jordi
Introduction
Observational and Interventional Distributions. Causal Thinking.

02 Roger
Potential Outcomes
Fundamental Problem of Causal Inference

03 Jordi
Causal Graphs
Do Calculus

11:45-12:10 Coffee Break

04 Roger
Estimand-based Estimation
Metalearners

05 Jordi
Estimand-agnostic Estimation
Counterfactuals

14:00-15:30 Lunch

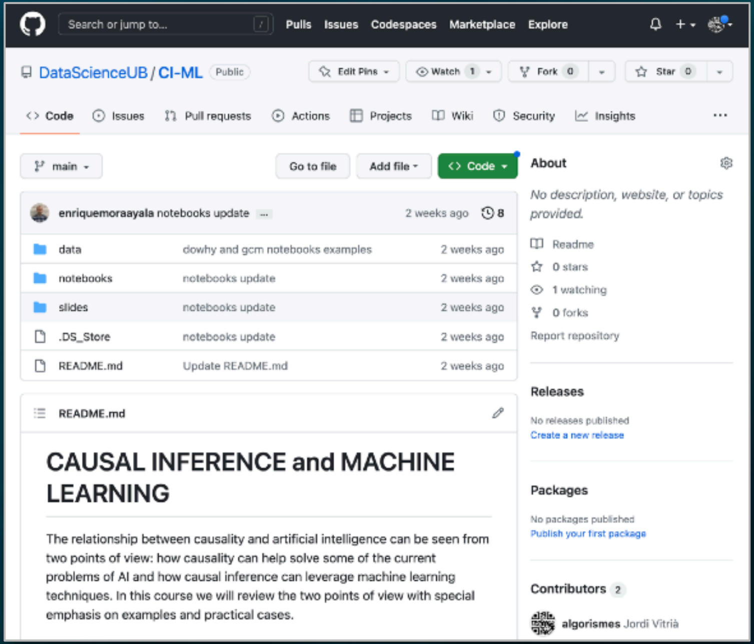
06 Jordi & Enrique
Causal Machine Learning
Supervised and Reinforcement Learning

07 Enrique
Practical Causal Inference
Exercises

17:30 End

About the course

The relationship between causality and artificial intelligence can be seen from two points of view: how causality can help solve some of the current problems of AI and how causal inference can leverage machine learning techniques. In this course we will review the two points of view with special emphasis on examples and practical cases.



https://github.com/DataScienceUB/CI-ML

Introduction

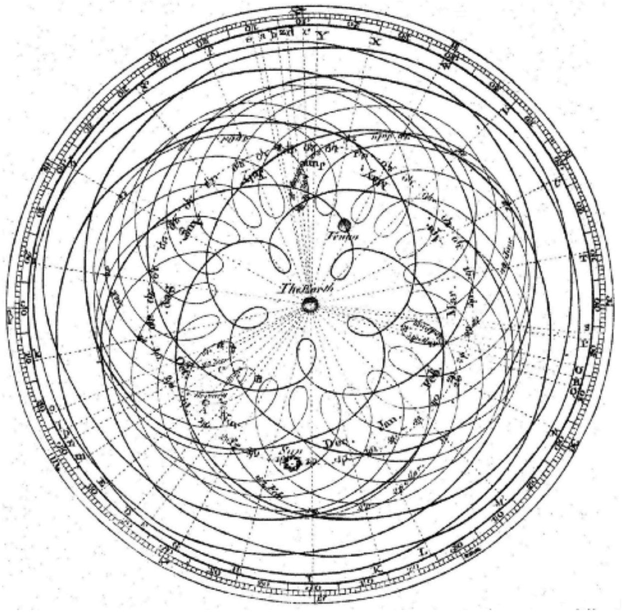
Observational and Interventional Distributions.
Causal Thinking.

Jordi Vitrià
jordi.vitria@ub.edu

UNIVERSITAT DE BARCELONA

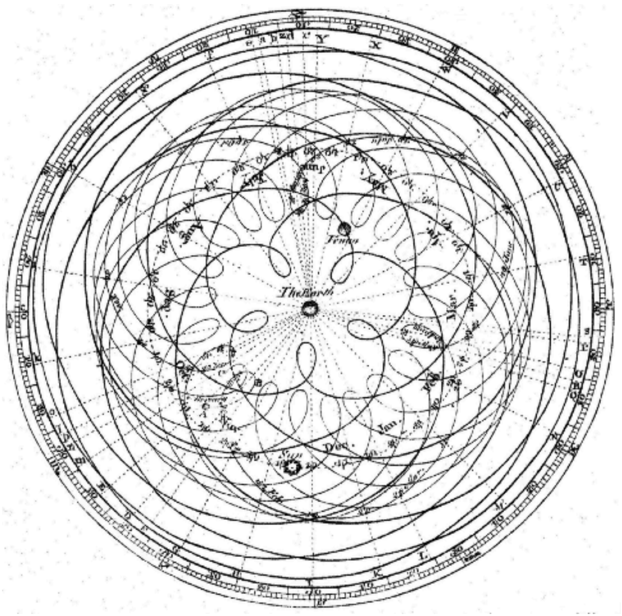
ML Mindset = Gathering data + Building a model

Predicting observations vs **predicting interventions**

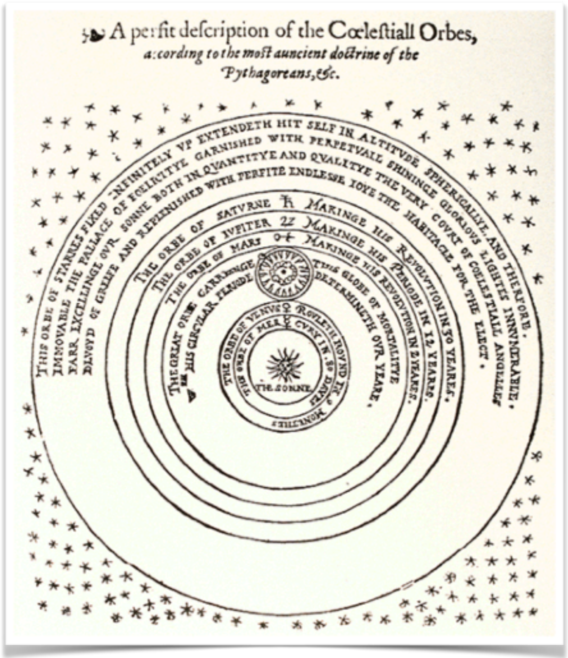


Ptolemaic model (circular orbits, geocentric)
100 AD

Predicting observations vs predicting interventions



Ptolemaic model (circular orbits, geocentric)
100 AD



Copernican model (heliocentric, harmonious = fewer causes)
1543 AD

Predicting observations vs predicting interventions

Ptolomeus and Copernicus build models with **high predictive power.** (Statistical/ML Mindset)

But they both were "false"!
(Causal/Scientific Mindset)

Predicting observations vs predicting interventions

- Predictions were not false in the "predictive" (Statistical/ML) sense, but in the "**interventional**" (scientific/causal) sense.

- What about aliens destroying (**intervening**) a planet instantly?

Predicting observations vs predicting interventions

Statistical inference and machine learning models are designed to predict observations (observational data) in stable environments.

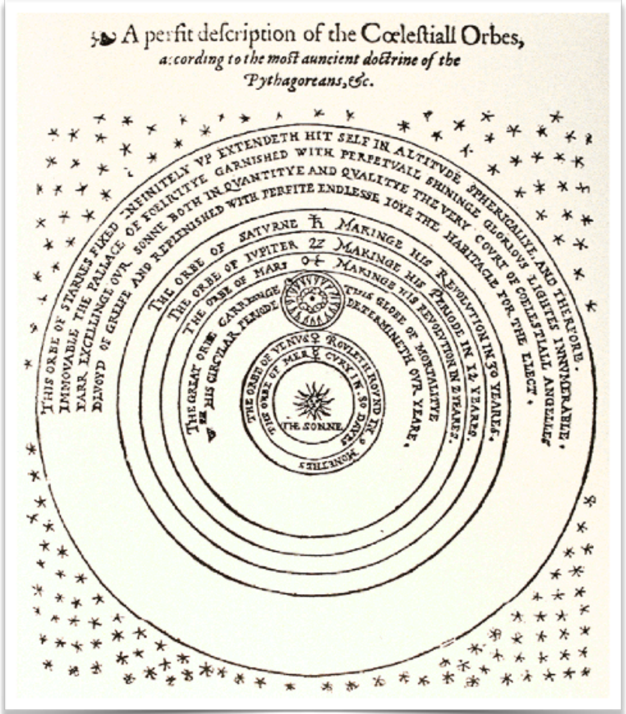
They are based on analyzing data to answer **associative questions**.

Predicting observations vs predicting interventions

What can I say about Y given that I have observed X?

What can I say about X given that I have observed Y?

Predicting observations vs **predicting interventions**



Modern ML models are very good at discovering and using associative structures in (X, Y) for predicting the value of Y in pure observational settings.

But predictive models can be accurate without being “correct” in interventional settings.

AI sucks…
or how CI can help ML

AI sucks...

Bogdan Kulynych @hiddenmarkov · 19 h
"Number of elevators in the building where the client *currently lives*"
1 1 8

Use of spurious correlations.

Bogdan Kulynych
@hiddenmarkov

Looking through a popular public dataset for risk scoring in home credit from a real home credit company, and some of the features included there are absolutely wild:
Tradueix el tuit

3:28 p. m. · 12 d'oct. de 2021 · Twitter Web App

14 Retuits 3 Tuits amb cita 43 Agradaments

Tuita una resposta Respon

Bogdan Kulynych @hiddenmarkov · 19 h
En resposta a @hiddenmarkov
"Who was accompanying client when he was applying for the loan?" (Unaccompanied, spouse, partner, group of people)
1 6

Bogdan Kulynych @hiddenmarkov · 19 h
"On which day of the week did the client apply for the loan?"
1 7

Bogdan Kulynych @hiddenmarkov · 19 h
"Approximately at what hour did the client apply for the loan?"
1 6

Bogdan Kulynych @hiddenmarkov · 19 h
"Number of elevators in the building where the client *currently lives*"
1 1 8

Bogdan Kulynych @hiddenmarkov · 19 h
"Number of enquiries to Credit Bureau about the client one hour/day/month before application"
2 5

Bogdan Kulynych @hiddenmarkov · 19 h
... I don't doubt these might be "correlated" with defaulting on loans, but someone at these agencies must realize that using these features for individual-level decisions will result in absurd outcomes?

AI sucks...

(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98

(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94

Use of spurious correlations.

AI sucks…

→ We want to minimize the **Empirical Risk**.

→ We want to maximize **robustness** against changes in data distribution.

→ We want to maximize **robustness** against adversarial attacks.

→ We want to be able of **explaining** my predictions to different stakeholders.

→ We want to measure and mitigate harmful biases (**discrimination**).

→ We want to use predictions to support a **decision** that may influence the outcome they aim to predict (performative predictions).

→ Etc.

AI sucks…

→ We want to minimize the **Empirical Risk**.

→ We want to maximize **robustness** against changes in data distribution.

→ We want to maximize **robustness** against adversarial attacks.

→ We want to be able of **explaining** my predictions to different stakeholders.

→ We want to measure and mitigate harmful biases (**discrimination**).

→ We want to use predictions to support a **decision** that may influence the outcome they aim to predict (performative predictions).

All these considerations involve causal thinking.

Causal Data Science

Causal Data Science

OBSERVATIONAL DATASET
(passive observation of the world)

	Sex	Race	Height	Income	Marital Status	Years of Educ.	Liberal-ness
R1001	M	1	70	50	1	12	1.73
R1002	M	2	72	100	2	20	4.53
R1003	F	1	55	250	1	16	2.99
R1004	M	2	65	20	2	16	1.13
R1005	F	1	60	10	3	12	3.81
R1006	M	1	68	30	1	9	4.76
R1007	F	5	66	25	2	21	2.01
R1008	F	4	61	43	1	18	1.27
R1009	M	1	69	67	1	12	3.25

Let's consider some different features in this dataset, (X, Y, Z).

Which can of questions can we answer from this dataset?

Causal Data Science

		X		Y		Z	
	Sex	Race	Height	Income	Marital Status	Years of Educ.	Liberal-ness
R1001	M	1	70	50	1	12	1.73
R1002	M	2	72	100	2	20	4.53
R1003	F	1	55	250	1	16	2.99
R1004	M	2	65	20	2	16	1.13
R1005	F	1	60	10	3	12	3.81
R1006	M	1	68	30	1	9	4.76
R1007	F	5	66	25	2	21	2.01
R1008	F	4	61	43	1	18	1.27
R1009	M	1	69	67	1	12	3.25

Causal Data Science

Q1: Which is the *expected income* Y that would have been observed if an individual had $X = x$ and $Z = z$?

Association (or prediction) is using data to map some features of the world (the inputs) to other features of the world (the outputs) based on the observed $p(X, Y, Z)$. For example, $\mathbb{E}(Y \mid X, Z)$.

All we need to do prediction is a dataset sampled from $p(X, Y, Z)$ and some inference tools (statistical inference & machine learning).

Mapping observed inputs to observed outputs is a **natural candidate for automated data analysis** because this task only requires 1) a large dataset with inputs and outputs, 2) an algorithm that establishes a mapping between inputs and outputs, and 3) a metric to assess the performance of the mapping, often based on a gold standard.

Causal Data Science

Causal effect of Race on Income

Q2: Estimate the **mean income** Y that would have been observed if all individuals had $(X = 1)$ vs. if they had $(X \neq 1)$.

Causal Inference is using data to **predict certain features of the world if the world had been different.** We cannot get these data by passive observation of the world! The world was different!

Answers to causal questions cannot be derived exclusively from $p(X, Y, Z)$. Answering a causal question (yes, sometimes is possible!) typically requires a combination of data, analytics, and **expert causal knowledge**.

Causal Data Science

	Sex	Race	Height	Income	Marital Status	Years of Educ.	Liberal-ness
R1001	M	1	70	50	1	12	1.73
R1002	M	2	72	100	2	20	4.53
R1003	F	1	55	250	1	16	2.99
R1004	M	2	65	20	2	16	1.13
R1005	F	1	60	10	3	12	3.81
R1006	M	1	68	30	1	9	4.76
R1007	F	5	66	25	2	21	2.01
R1008	F	4	61	43	1	18	1.27
R1009	M	1	69	67	1	12	3.25

$$\mathbb{E}[Y \mid do(X = 1)] - E[Y \mid do(X \neq 1)] \, ?$$

Causal effect of Race on Income

Causal Data Science

Causal effect of Race on Income

Intervention

Intervention

The answer to

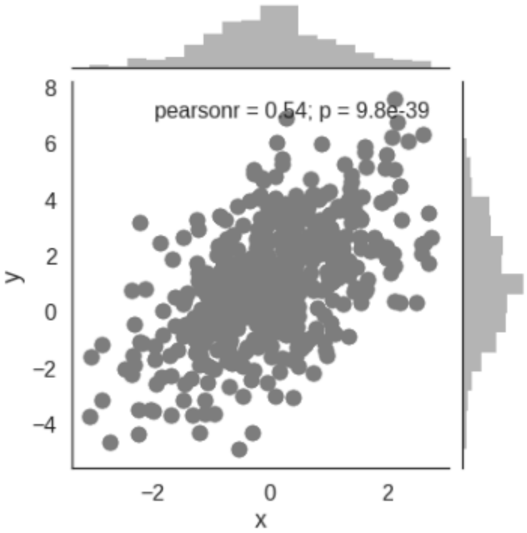
$$\mathbb{E}[Y \mid do(X = 1)] - E[Y \mid do(X \neq 1)]$$

is not

$$\mathbb{E}[Y \mid X = 1] - E[Y \mid X \neq 1]$$

Causal Data Science

In order to understand what is $p(Y | do(X = x))$, let's suppose I have observed $p(X, Y)$.



This is all we need to compute $p(Y | X)$. We can give an answer to any <u>associational question</u>.

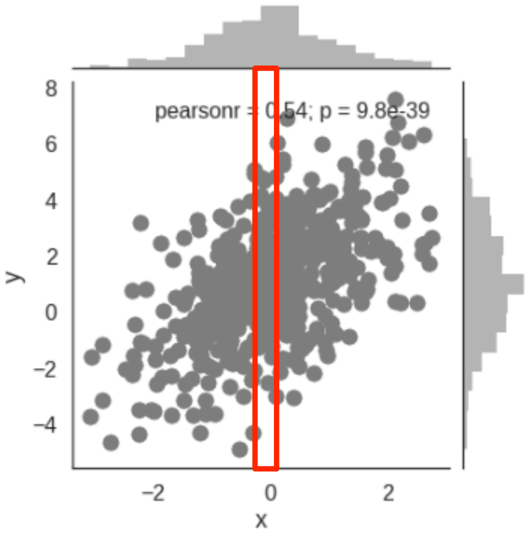
For example:

- What is the expected value of Y if we observe $X = 0$, $\mathbb{E}(Y | X = 0)$? (**Regression**)
- What is the expected MAX/MIN/MEDIAN value of Y if we observe $X = 0$? (**Quantile regression**)
- Etc.

Causal Data Science

- What is the expected value of Y if we observe $X = 0$, $\mathbb{E}(Y \mid X = 0)$? (**Regression**)

$$p(Y \mid X = 0)$$



https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/

Causal Data Science

OBS $p(Y|X)$

INT $p(Y|do(X = x))$

OBS and INT are not generally the same!
Let's consider three generative models corresponding to the same $p(X, Y)$

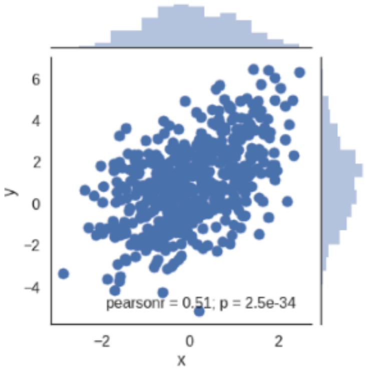
Causal Data Science

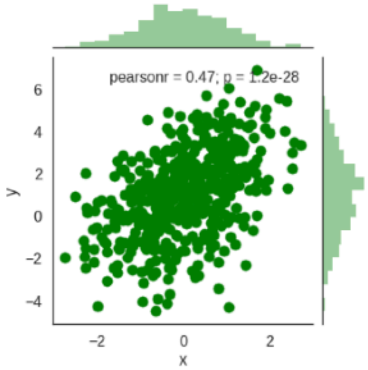
Generative Models

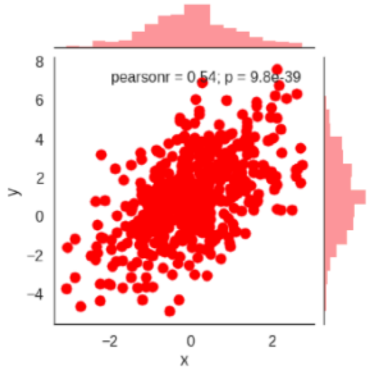
```
x = randn()
y = x + 1 + sqrt(3)*randn()
```

```
y = 1 + 2*randn()
x = (y-1)/4 + sqrt(3)*randn()/2
```

```
z = randn()
y = z + 1 + sqrt(3)*randn()
x = z
```







Based on the joint distribution the three scripts are indistinguishable.

https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/

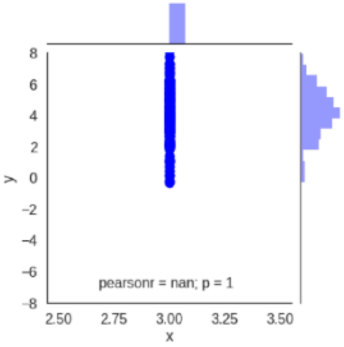
Causal Data Science

Intervention $p(Y \mid do(X = 3))$

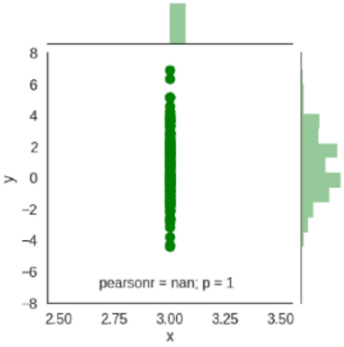
```
x = randn()
x = 3
y = x + 1 + sqrt(3)*randn()
x = 3
```

```
y = 1 + 2*randn()
x = 3
x = (y-1)/4 + sqrt(3)*randn()/2
x = 3
```

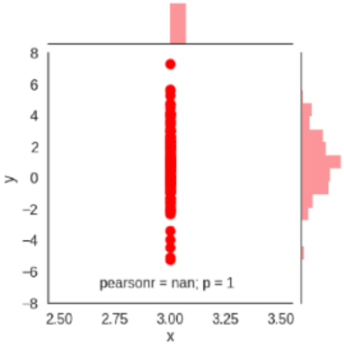
```
z = randn()
x = 3
x = z
x = 3
y = z + 1 + sqrt(3)*randn()
x = 3
```



$p(Y \mid do(X = 3))$



$p(Y \mid do(X = 3))$



$p(Y \mid do(X = 3))$

The joint distribution of data $p(X, Y, Z)$ alone is insufficient to predict behavior under interventions.

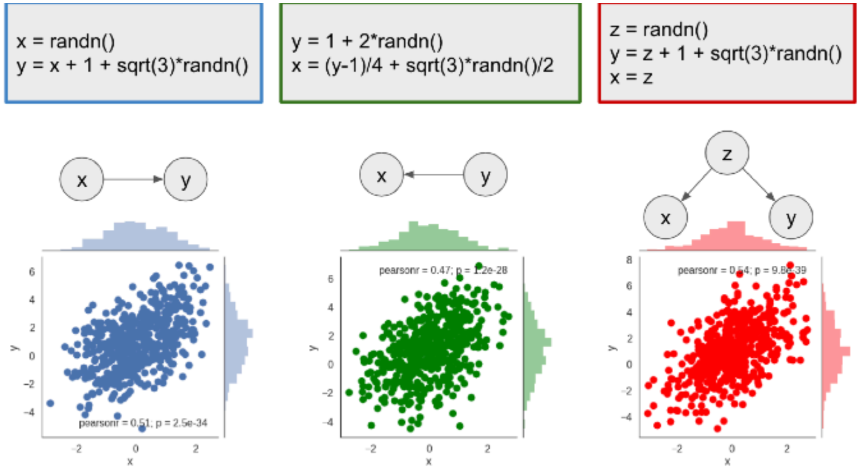
https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/

Causal Data Science

Generative Models

An intervention can be understood as a **modification of the generative model of the data, producing a different probability distribution**

$$p(do(X = 3), Y, Z)$$



Directed Acyclic Graphs (DAG).

No assumptions about the exact form of the functional relationships are needed. The only requirement is that causal relationships are **acyclic**.

https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/

Causal Data Science

Generative Models

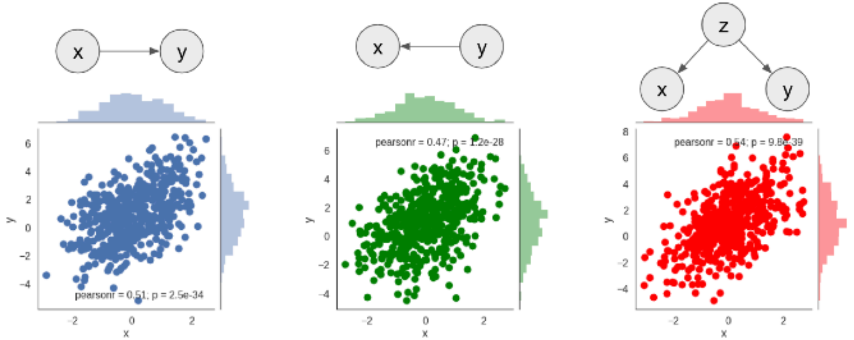
An intervention can be understood as a **modification of the generative model of the data, producing a different probability distribution**

$$p(do(X = 3), Y, Z)$$

```
x = randn()
y = x + 1 + sqrt(3)*randn()
```

```
y = 1 + 2*randn()
x = (y-1)/4 + sqrt(3)*randn()/2
```

```
z = randn()
y = z + 1 + sqrt(3)*randn()
x = z
```



X is the cause of Y

Y is the cause of X

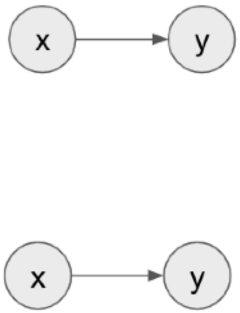
X and Y are not causally related (but they are associated!)

https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/

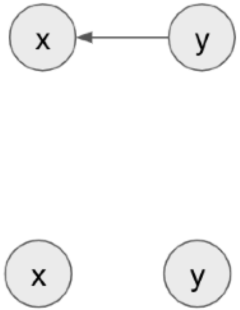
Causal Data Science

What is an intervention?

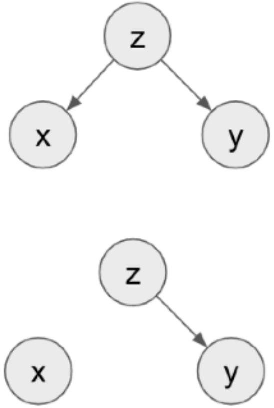
Graphically, to **simulate the effect of an intervention**, you **mutilate** the graph by removing all edges that point into the variable on which the intervention is applied, in this case x.



$$P(Y \mid do(X = x)) = P(Y \mid X = x)$$



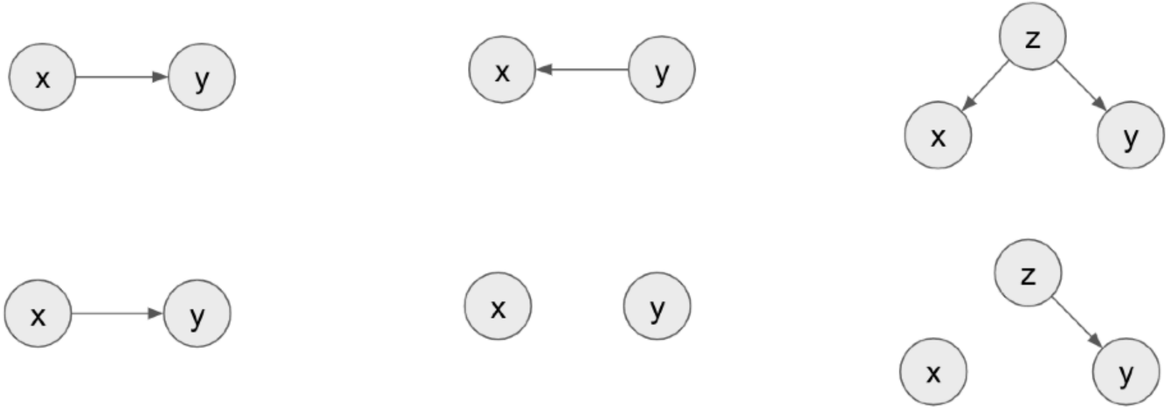
$$P(Y \mid do(X = x)) = P(Y)$$



$$P(Y \mid do(X = x)) = P(Y)$$

Causal Data Science

What is an intervention?

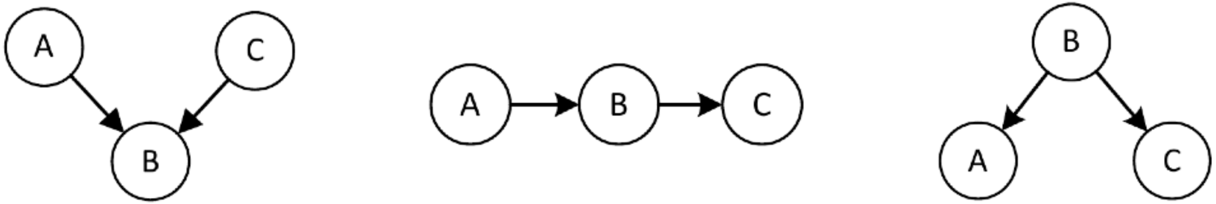


$P(Y \mid do(X = x)) = P(Y \mid X = x)$ $\quad$ $P(Y \mid do(X = x)) = P(Y)$ $\quad$ $P(Y \mid do(X = x)) = P(Y)$

Just by looking at the causal diagram, we are now **able to predict** how the scripts are going to behave under the intervention $X = 3$.

Causal Data Science

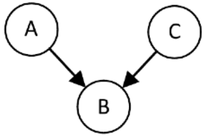
- The primary language for modeling causal mechanisms and expressing our assumptions is the **language of causal graphs**.



- Causal graphs encode our domain knowledge about the causal mechanisms underlying a system or phenomenon under study.
- Causal graphs are assumed to be **acyclic**. This is why they are called **DAGs (Directed Acyclic Graphs)**.

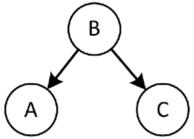
Causal Data Science

- Fundamentally, a causal graph describes a **non-parametric data-generating process** over its nodes.
- By specifying independence and dependence between the nodes, the graph constrains relationship between generated variables corresponding to those nodes.



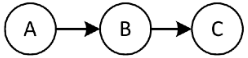
B is a **collider** for A and C
A and B create an **inverted fork** to B

A and C are independent



B is a **confounder**
B creates a **fork** to A and C

A and C are not independent.
A and C are independent conditional on B



B forms a **chain** from A to C

A and C are conditionally independent given B

Causality Theory

A DAG provides enough extra-data information
(in terms of conditional independences)

to answer many causal queries,

even with the data generating process hidden.

Causal Thinking

Basic Concepts: Causal Effect

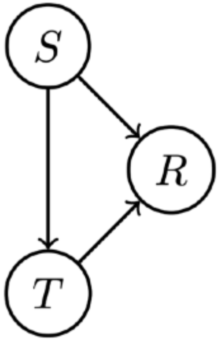
- The **average treatment/causal effect** (ATE) of T on R:

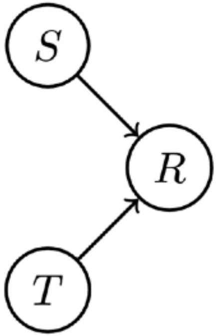
$$\mathbb{E}[R \mid do(T=1)] - E[R \mid do(T=0)]$$

- The **conditional average treatment/causal effect** (CATE):

$$\mathbb{E}[R \mid do(T=1), S] - E[R \mid do(T=0), S]$$

Symptoms, Treatment, Recovery

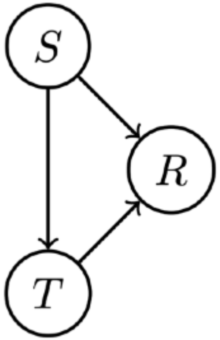




Intervened graph
$do(T=t)$

Basic Concepts: Counterfactual

- **Counterfactuals**: hypothetical result that an intervention may have on an individual for whom we have already observed a different *factual* outcome.
    - Counterfactuals allow us to mix factual information with alternative scenarios.
    - Explainability and Fairness applications.
- Given a certain patient with symptoms s, who was not given a treatment and didn't recover, **would they have recovered had we given them the treatment?**
- The **individual treatment/causal effect** (ITE):



$$\mathbb{E}[R_i \mid do(T_i = 1)] - E[R_i \mid do(T_i = 0)]$$

Causal Thinking Process

1. Asking a causal/counterfactual query (ATE, CATE, ITE,...)
2. Gathering knowledge from experts
3. Building a DAG
4. Identifying the causal query
5. Gathering data.
6. Computing and estimand/building a SCM
7. Answering the causal/counterfactual query

Asking a causal query

- Let's start with this setting:
  - $\mathbb{E}[S \mid do(G = male)] - \mathbb{E}[S \mid do(G = female)]$: **ATE** of gender (binary) on salary.
  - $\mathbb{E}[S_i^* \mid do(G_i^* = male), G_i = female, S_i = s]$: given a particular woman i with salary s, **counterfactual** salary when male.

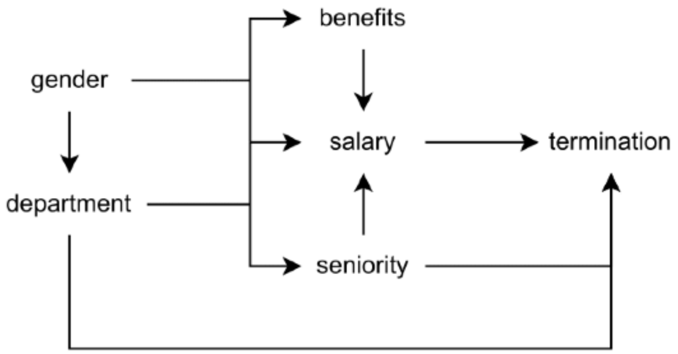
Salary Dataset

- Variables:
  - **G**ender.
  - **D**epartment.
  - **B**enefits.
  - **Se**niority.
  - **S**alary.
  - **T**ermination.

Gathering knowledge and data

Salary Dataset

- We need to find the corresponding causal graph.
  - Causal Discovery algorithms.
  - Domain Experts.
  - Experiments.



Causal Graph

Identifiying a causal query

There are <u>two ways</u> to measure the **causal relationship** between two variables, S and G:

1. The easiest way is an **intervention** in the real world: You **randomly** force G to have different values and you measure S.

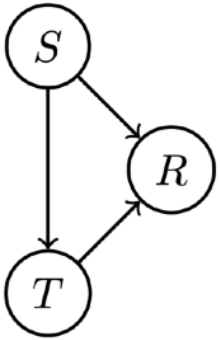
   This is what we do in Randomized Clinical Trial (RCT) or in an A/B Test.

   This is not always feasible (because of **practical**, **ethical** or **economical** reasons)

Identifiying a causal query

2. If the query is **identifiable** we can compute an estimand.

   For example, in this case, do-calculus allows us to massage $p(S, R, T)$ until we can express $p(R \mid do(T))$ in terms of various marginals, conditionals and expectations under $p(S, R, T)$.

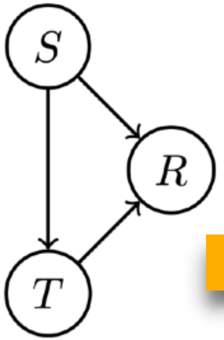


$$\Pr(R \mid do(T)) = \sum_{s} \Pr(R \mid T, S) \Pr(S)$$

Identifiying a causal query

2. If the query is **identifiable** we can compute an estimand.

For example, in this case, do-calculus allows us to massage $p(S, R, T)$ until we can express $p(R \mid do(T))$ in terms of various marginals, conditionals and expectations under $p(S, R, T)$.



$$\Pr(R \mid do(T)) = \sum_s \Pr(R \mid T, S) \Pr(S)$$

It only depends on the causal graph!

Identifying the causal query

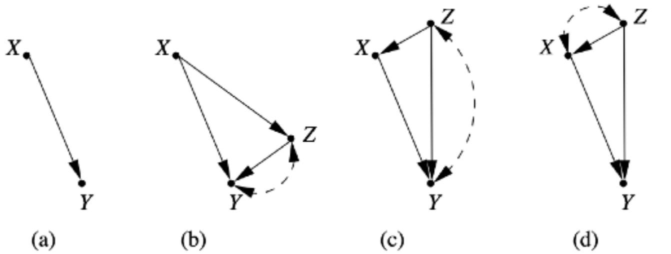
- Causal query $\mathcal{Q}$, e.g., $\mathcal{Q} := \mathbb{E}[S \mid do(G = male)] - \mathbb{E}[S \mid do(G = female)]$.
  - It contains interventional terms, so we can't use the dataset directly.
  - **Identification**: transform every interventional term into an expression using only observational terms $\Rightarrow$ **estimand**.
  - There are **automated algorithms** that do this work for us.

Identifying the causal query

Salary Dataset

Given a causal query for a certain DAG, we say it is **identifiable** if we can derive an statistical estimand (**only using observational terms**) for this query using the rules of **do-calculus.**

The ***do*-calculus** is an axiomatic system for replacing probability formulas containing the *do* operator with ordinary conditional probabilities. It consists of three axiom schemas that provide **graphical criteria** for when certain substitutions may be made.

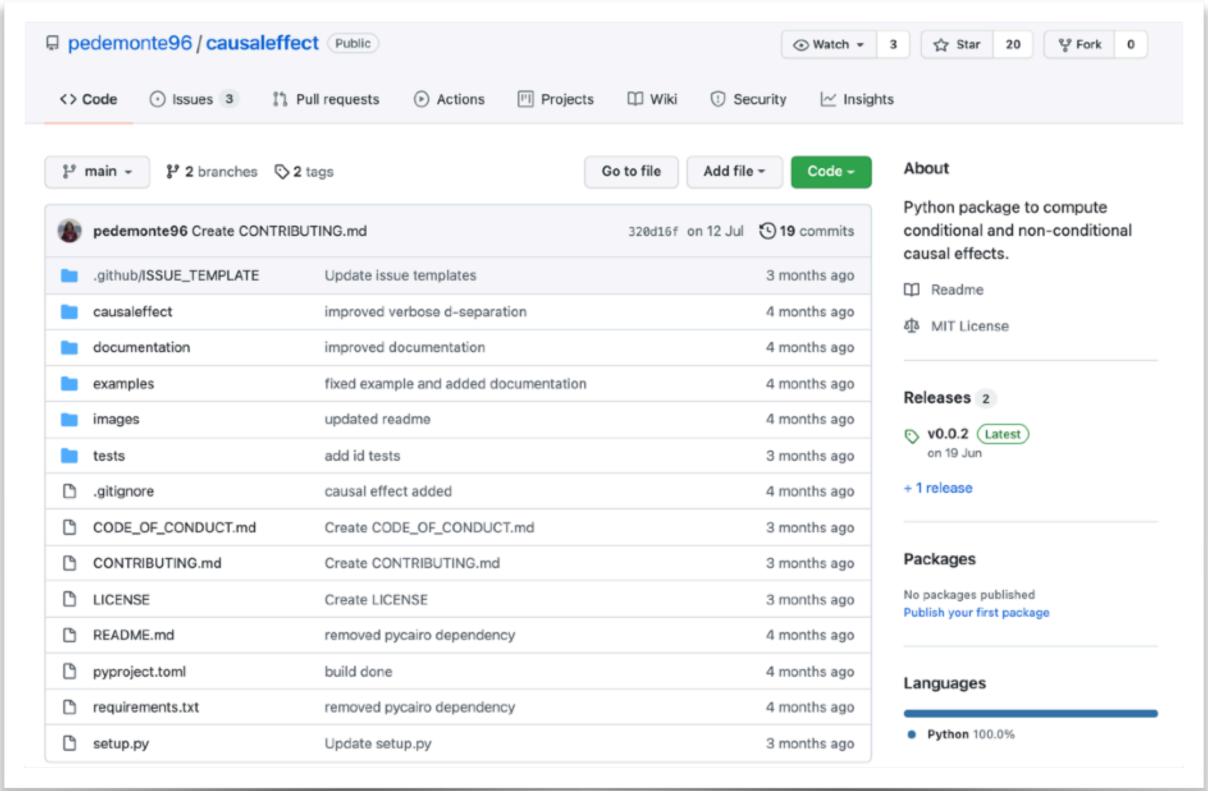


Causal graphs where $P(y|do(\mathbf{x}))$ is identifiable

Source: Complete Identication Methods for Causal Inference, PhD Thesis, University of California. I.Shpitser

Dashed lines correspond to **unobserved confounders**, associations produced by unobserved variables.

Identifying the causal query



Identifying the causal query

```python
import causaleffect

G = causaleffect.createGraph(['X<->Y', 'Z->Y', 'X->Z', 'W->X', 'W->Z'])
causaleffect.plotGraph(G)
```

```python
P = causaleffect.ID({'Y'}, {'X'}, G)
P.printLatex()
```

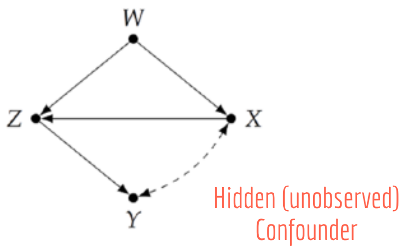
The code above computes the causal effect, and returns a string encoding the distribution in LaTeX notation:

```
'\sum_{w, z}P(w)P(z|w, x)\left(\sum_{x}P(x|w)P(y|w, x, z)\right)'
```

This string, in LaTeX, is

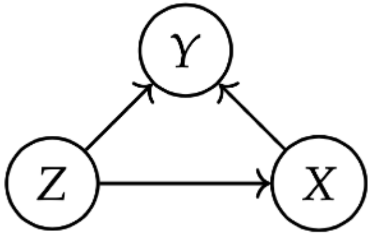
$$\sum_{w,z} P(w)P(z|w,x)\left(\sum_{x} P(x|w)P(y|w,x,z)\right)$$

$$p(X, W, Z, Y)$$



Building a causal model

- Once we have an estimand, we need to build **models** to use it for our estimation.
  - $\mathcal{Q} := \mathbb{E}[Y \mid do(X = x)] = \mathbb{E}_Z[\mathbb{E}_{Y|x,Z}[Y]]$.
  - We need to model the $f(x, Z) := \mathbb{E}_{Y|x,Z}[Y]$ term.
    - If Y is binary, with a ML classifier.
    - If Y is continuous, with a ML regressor.



Answering the query

→ Now that we have our ML model, we can follow the **estimand** formula:

- → $\mathcal{Q} := \mathbb{E}[Y \mid do(X = x)] = \mathbb{E}_Z[\mathbb{E}_{Y|x,Z}[Y]].$

- → The $\mathbb{E}_Z$ expectation can be estimated by averaging dataset samples:

$$\mathcal{Q} = \mathbb{E}_Z[\mathbb{E}_{Y|x,Z}[Y]] \approx \frac{1}{n} \sum_{i=1..N} f(x, z_i).$$

Answering the query

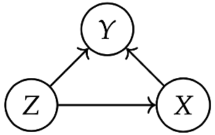
→ The **estimand-based** approach:

1. Derive an **estimand** for our graph & query.
2. Train ML **models** for the terms we need to compute.
3. Follow the estimand formula to obtain an **estimation**.

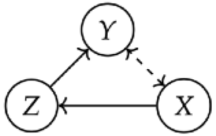
Answering the query

→ However, depending on the graph, even the same query results in different estimands.

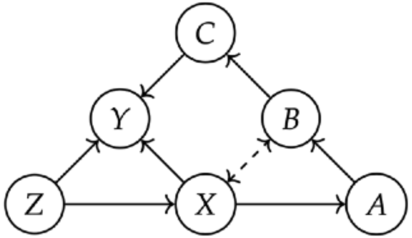
→ $P(Y \mid do(X = x))$:



$\mathcal{Q} = \mathbb{E}_Z\big[P(Y \mid x, Z)\big]$



$\mathcal{Q} = \mathbb{E}_{Z|x}\big[\mathbb{E}_X\big[P(Y \mid X, Z)\big]\big]$



$\mathcal{Q} = \mathbb{E}_{Z,C}\Big[\frac{P(Y \mid x, Z, C)}{P(Z, C)} \cdot \mathbb{E}_{A|x}\big[\mathbb{E}_X\big[P(Z, C \mid X, A)\big]\big]\Big]$

Example

From "Causal Inference in AI Education: A Primer"

Example 3.1. ***AdBot*** *Consider an online advertising agent attempting to maximizing clickthroughs, with* $X \in \{0,1\}$ *representing two ads,* $Y \in \{0,1\}$ *whether or not it was clicked upon, and* $Z \in \{0,1\}$ *the sex of the viewer. A marketing team collects the following data on purchases following ads shown to focus groups to be used by AdBot:*

	Ad 0	**Ad 1**
Male	108/120 (90%)	340/400 (85%)
Female	266/380 (70%)	65/100 (65%)
Total	374/500 (75%)	405/500 (81%)

Table 1. Clickthroughs in the AdBot setting striated by the ad shown to participants in a focus group, and the sex of the viewer.

If the sex of a viewer is not know, which ad is the best choice?

Example

From "Causal Inference in AI Education: A Primer"

Example 3.1. ***AdBot*** *Consider an online advertising agent attempting to maximizing clickthroughs, with $X \in \{0,1\}$ representing two ads, $Y \in \{0,1\}$ whether or not it was clicked upon, and $Z \in \{0,1\}$ the sex of the viewer. A marketing team collects the following data on purchases following ads shown to focus groups to be used by AdBot:*

	Ad 0	Ad 1
Male	108/120 (90%)	340/400 (85%)
Female	266/380 (70%)	65/100 (65%)
Total	374/500 (75%)	405/500 (81%)

Table 1. Clickthroughs in the AdBot setting striated by the ad shown to participants in a focus group, and the sex of the viewer.

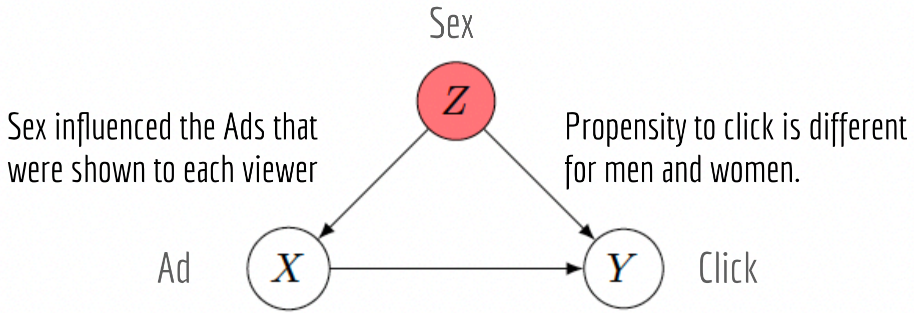
If the sex of a viewer is not know, which ad is the best choice?

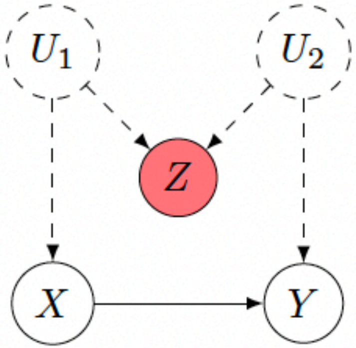
Example

From "Causal Inference in AI Education: A Primer"

If the sex of a viewer is not know, which ad is the best choice?

These are two different causal stories:

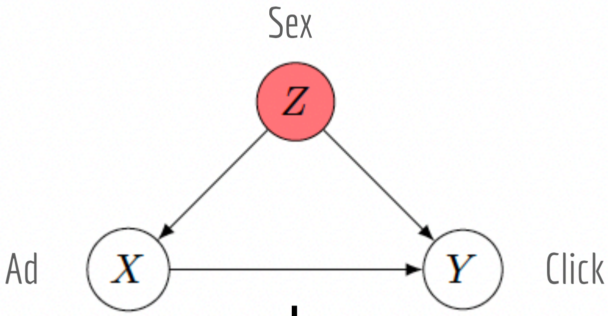




Relevant question: $p(Y \mid \mathrm{do}(X_0)) > p(Y \mid \mathrm{do}(X_1))?$

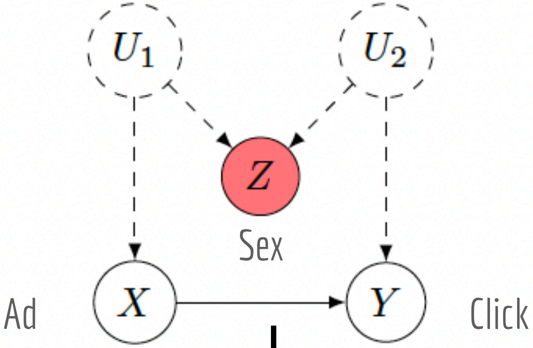
Example

From "Causal Inference in AI Education: A Primer"




```
G = causaleffect.createGraph(['X->Y', 'Z->Y', 'Z->X'])
P = causaleffect.ID({'Y'}, {'X'}, G)
```

$$p(Y \mid \mathrm{do}(X)) = \sum_z P(Y \mid X, Z) P(Z)$$




```
G = causaleffect.createGraph(['U1<->X', 'U1<->Z', 'U2<->Z', 'U2<->Y','X->Y'])
P = causaleffect.ID({'Y'}, {'X'}, G)
P.printLatex()
```

$$p(Y \mid \mathrm{do}(X)) = P(Y \mid X)$$

Example

From "Causal Inference in AI Education: A Primer"

Example 3.1. ***AdBot*** *Consider an online advertising agent attempting to maximizing clickthroughs, with* $X \in \{0,1\}$ *representing two ads,* $Y \in \{0,1\}$ *whether or not it was clicked upon, and* $Z \in \{0,1\}$ *the sex of the viewer. A marketing team collects the following data on purchases following ads shown to focus groups to be used by AdBot:*

	Ad 0	Ad 1
Male	108/120 (90%)	340/400 (85%)
Female	266/380 (70%)	65/100 (65%)
Total	374/500 (75%)	405/500 (81%)

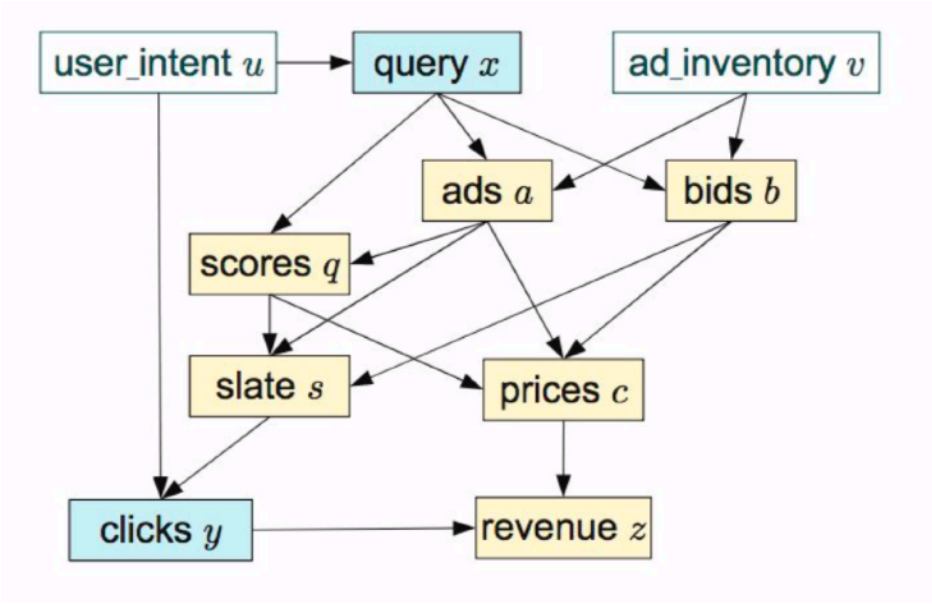
If (a) is our explanation of the data, then AdBot should display Ad0.

$$p(Y \mid \mathrm{do}(X)) = \sum_z P(Y \mid X, Z)P(Z)$$

If (b) is our explanation of the data, then AdBot should display Ad1.

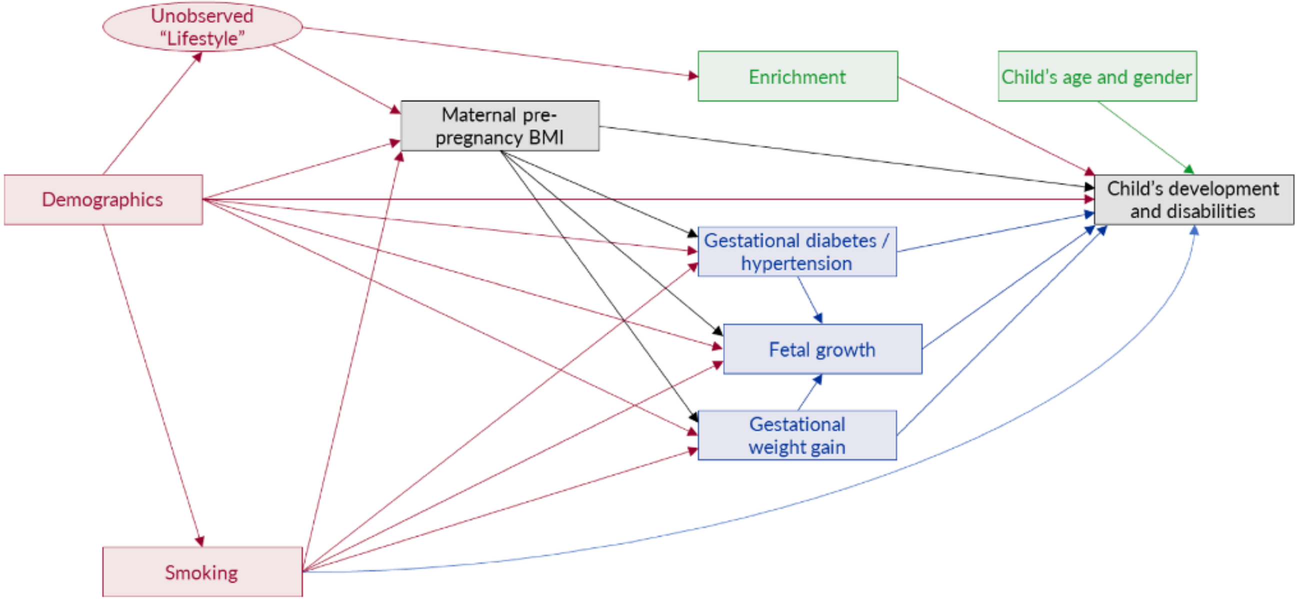
$$p(Y \mid \mathrm{do}(X)) = P(Y \mid X)$$

Example



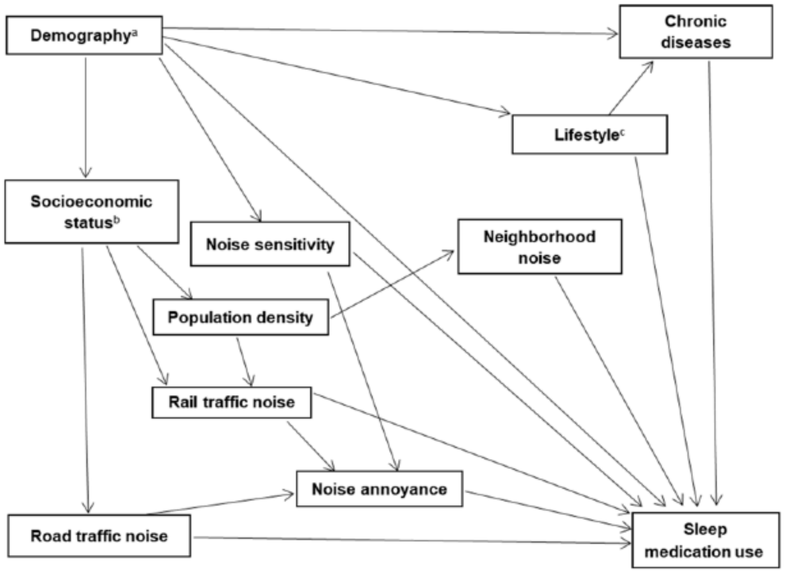
Bottou, Léon, et al. "Counterfactual reasoning and learning systems: the example of computational advertising." *The Journal of Machine Learning Research* 14.1 (2013): 3207-3260.

Example

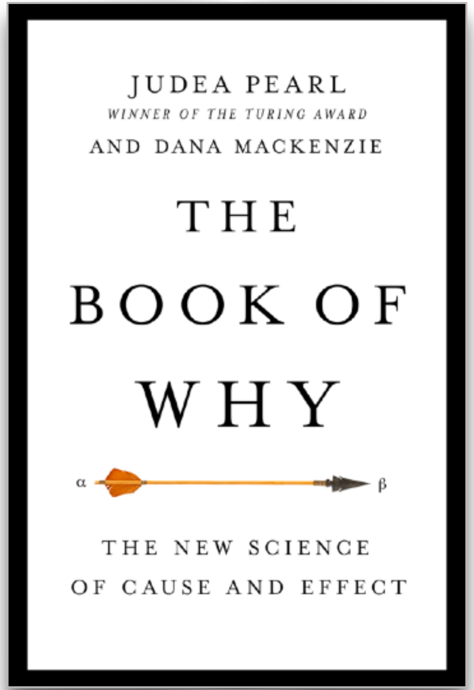


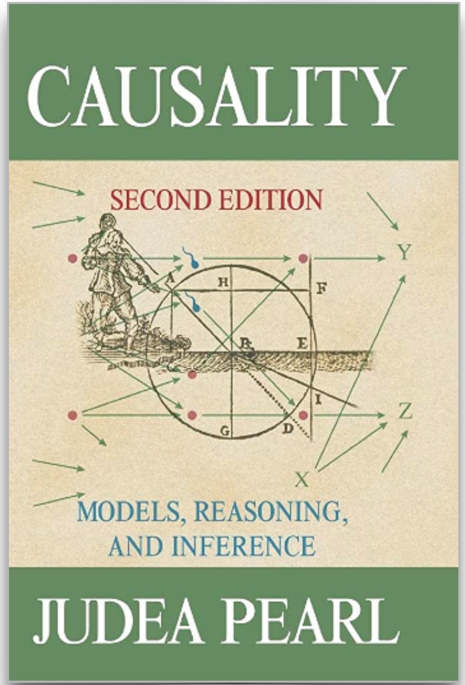
ADAPTED FROM: Hinkle SN, Sharma AJ, Kim SY, Schieve LA. Maternal prepregnancy weight status and associations with children's development and disabilities at kindergarten. *Int J Obes (Lond)*. 2013;37(10):1344-51. DOI: 10.1038/ijo.2013.128 (Figure 1). Freely available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4407562

Example


Demography[a]
Chronic diseases
Lifestyle[c]
Socioeconomic status[b]
Noise sensitivity
Neighborhood noise
Population density
Rail traffic noise
Noise annoyance
Road traffic noise
Sleep medication use

References


JUDEA PEARL
WINNER OF THE TURING AWARD
AND DANA MACKENZIE
THE
BOOK OF
WHY
α
β
THE NEW SCIENCE
OF CAUSE AND EFFECT


CAUSALITY
SECOND EDITION
MODELS, REASONING,
AND INFERENCE
JUDEA PEARL

References

- Glymour, Madelyn, Judea Pearl, and Nicholas P. Jewell. Causal inference in statistics: A primer. John Wiley & Sons, 2016.
- Imbens, G. W., & Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.
- Hünermund, Paul, and Elias Bareinboim. "Causal inference and data fusion in econometrics." The Econometrics Journal, forthcoming. Also in *arXiv preprint arXiv:1912.09104* (2019).
- Neal, Brady. "Introduction to causal inference from a machine learning perspective." Course Lecture Notes (draft) (2020).
- Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. Elements of causal inference: foundations and learning algorithms. The MIT Press, 2017.
- Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.
- Kaddour, J., Lynch, A., Liu, Q., Kusner, M. J., & Silva, R. (2022). Causal machine learning: A survey and open problems. arXiv preprint arXiv:2206.15475.
- Mitchell, Shira, et al. "Algorithmic fairness: Choices, assumptions, and definitions." Annual Review of Statistics and Its Application 8 (2021): 141-163.
- Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. Proceedings of the IEEE, 109(5), 612-634.
- Schölkopf, B. (2022). Causality for machine learning. In Probabilistic and Causal Inference: The Works of Judea Pearl (pp. 765-804).
- Hünermund, Paul and Kaminski, Jermain and Schmitt, Carla, Causal Machine Learning and Business Decision Making (February 19, 2022). Available at SSRN: https://ssrn.com/abstract=3867326 or http://dx.doi.org/10.2139/ssrn.3867326.

Which causal inference book you should read

https://www.bradyneal.com/which-causal-inference-book