The Causal Revolution

Jungle Program - Machine Learning & Data Science Community Meetup

Amric Trudel - OCTO Technology

June 29th, 2021

There is a better way

Outline

There is a better way

01

OCTO Technology

There is a better way

OCTO in Numbers

IT Consulting & Outstanding Software

1 INTERNATIONAL SUBSIDIARIES

4 OFFICES IN FRANCE

HEAD-QUARTERS

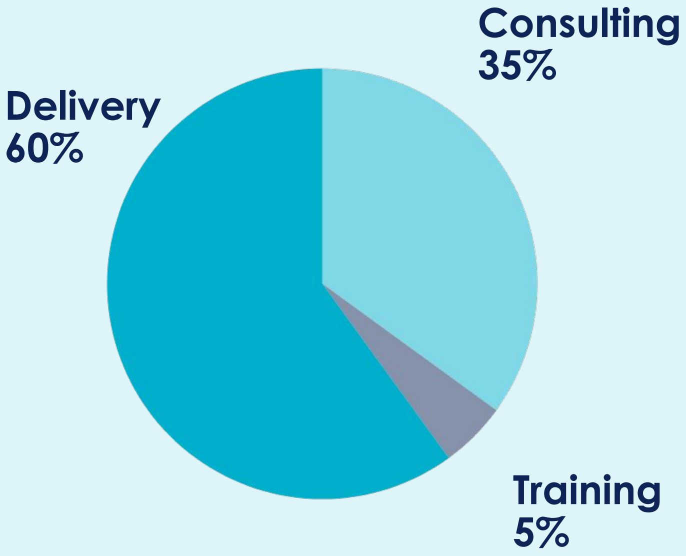
PARIS

740 EMPLOYEES

with over 630 Consultants, Experts, and Coaches

all well-versed in technology, methodology, and sensitive to your professional challenges

Breakdown by Activity



There is a better way

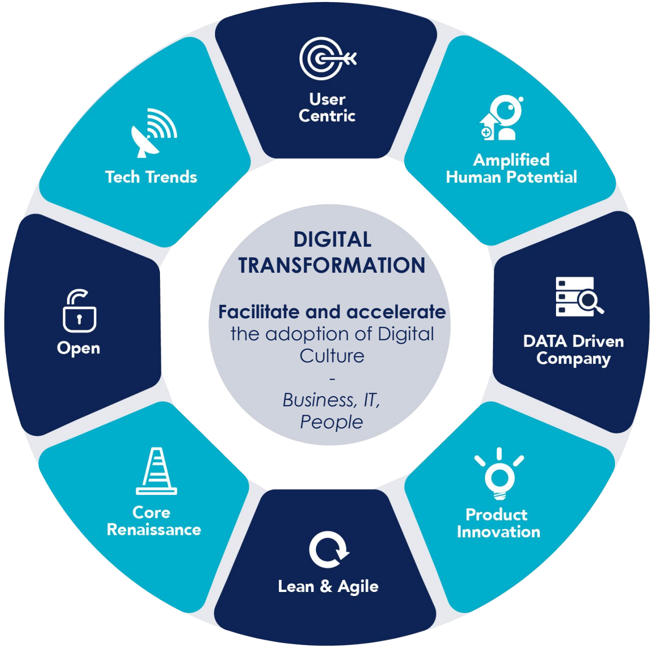
OCTO - Digital Transformation Accelerator

USER CENTRIC
Rely on relevant technologies to **build** User Experience that can answer any need : ATAWAD "*any time*, *anywhere*, *any device*".

TECH TRENDS
Provide expertise on the potential of technology to boost your business models and economic assets.

OPEN
Open your information system and **secure** access to your services to share, monetize, and benefit from the key drivers of Open Innovation.

CORE RENAISSANCE
Redesign the core of your information systems to meet digital's high standards, reduce TTM, and refocus on essential features.



AMPLIFIED HUMAN POTENTIAL
Share the codes of digital by establishing collaborative practices and boost potential and expertise within your teams.

DATA-DRIVEN COMPANY
Build data architecture that can drive your activities and make the most of opportunities provided by Big Data, Data Science, and Machine Learning.

PRODUCT INNOVATION
Structure and **Ensure** effective deployment of the Product Innovation approach (Design Thinking and Lean startup) to firmly plant it in your organization and processes (portfolio).

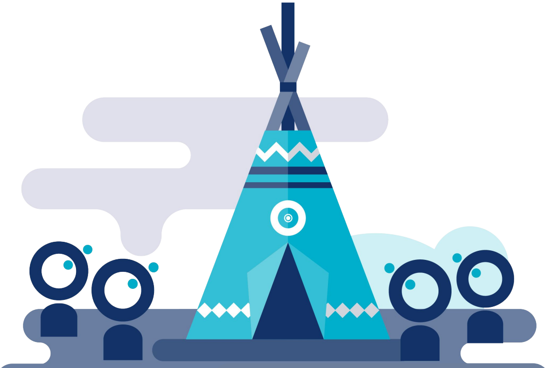
LEAN & AGILE
Coach your teams to take ownership of the best practices on the market: autonomous and multidisciplinary teams, short cycles, Test & Learn culture, industrialization of software procedures, DevOps, Software Craftsmanship.

There is a better way

Our Ecosystem

We provide support to all industries

- INDUSTRIAL COMPANIES
- TRAVEL
- DEFENSE AEROSPACE
- ENERGY UTILITIES
- BANKING
- INSURANCE
- RETAIL
- CONSUMER GOODS
- SERVICE PROVIDERS
- PUBLIC SECTOR
- HEALTHCARE
- ENTERTAINMENT
- INTERNET
- MEDIA
- SERVICES



600 consultants are ready and willing to help you find solutions. Organized into tribes, they all share expertise and knowledge to provide the best!

- SCALING AGILE
- APPLICATION SECURITY
- DATA VISUALIZATION EXPERIENCE
- SOFTWARE CRAFTSMANSHIP
- NEW DATA ARCHITECTURE
- INTELLIGENCE PLATFORM
- BIG DATA ANALYTICS
- AGILE & COMPLEX DELIVERY
- ARCHI EXCHANGES & INTEGRATION
- CONNECTED EXPERIENCE
- VIRTUAL IMMERSION AND BOT EXPERIENCE
- PRODUCT MANAGEMENT & DESIGN
- LIQUID SAP ANALYTICS
- CULTURE HACKING
- INFRASTRUCTURES & OPERATIONS
- DATA MANAGEMENT EXCELLENCE
- APPLIED INTELLIGENCE DATA & ADVISORY
- ENTERPRISE PERFORMANCE, PLANNING INTELLIGENCE COMMUNITY
- UX USER EXPERIENCE
- WEB FRONT-END
- BLOCKCHAIN
- MOBILITY
- WOA & API
- DELIVERY AGILE
- MANAGEMENT OF INFORMATION SYSTEMS AND TECHNOLOGIES

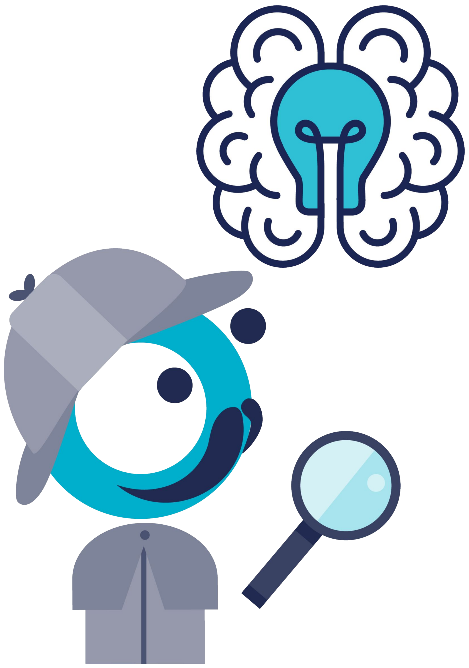
There is a better way

02

Causation

Why Study Causation?

Humans Think With Causation



There is a better way

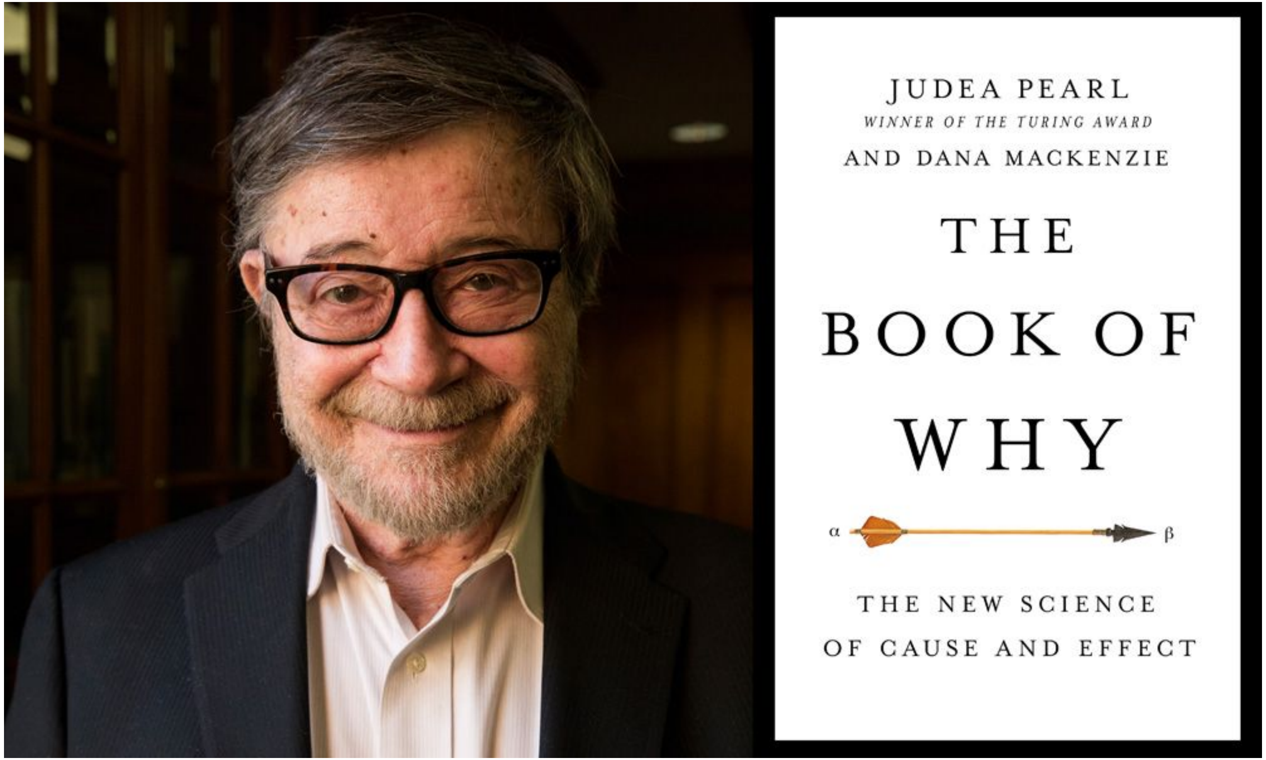
A Few Questions That Entail Causation

1. Why are some neighborhoods poorer than others?
2. Was it the new tax regulation or our advertising campaign that made our sales go up?
3. What is the efficacy of the AstraZeneca vaccine against coronavirus infection?
4. Should the European Medicines Agency approve the Sputnik V vaccine?
5. I'm about to switch careers to become a Data Scientist. Should I?

There is a better way

Judea Pearl

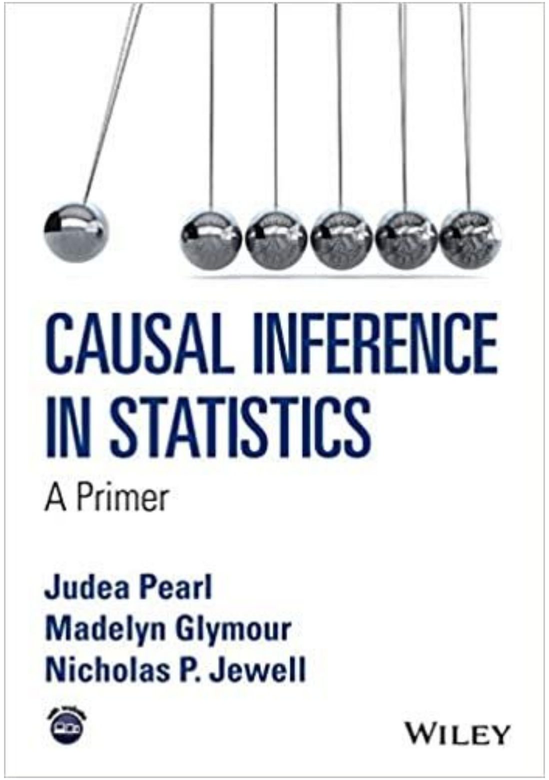
pioneer of the field of Causal Inference



There is a better way

Judea Pearl

To get started with causal calculus



[Series of Medium posts](#)
by Bruno Gonçalves

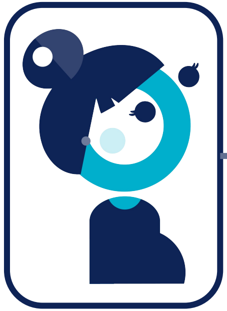
There is a better way

How Can We Find a Cause and its Effect?

Does a drug...

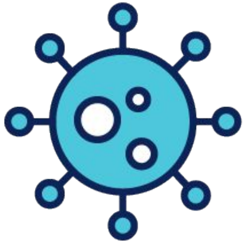
have an effect on...

disease recovery?



Does pregnancy...

have an effect on...

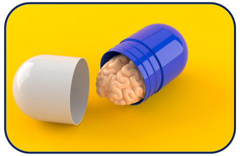


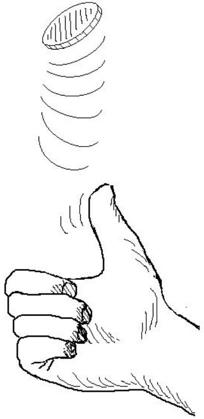
vulnerability to a virus?

How Can We Find a Cause and its Effect?

Research Methods

- **Randomized Controlled Trial** *(RCT)*
  - The **Gold standard** in causal research
  - Individuals are **randomly assigned** to a treatment condition that corresponds to a possible value of the **controlled variable**.
  - ***Randomness*** guarantees that we find a causal relationship.
- **Observational Study**
  - Data is simply collected **without intervention** on the studied environment.
  - Causation can't be distinguished from **correlation**





There is a better way

03

The Tools of Science and Statistics

There is a better way

The Tools of Science

Mathematical equations cannot express causal relationships

- **B**: Barometer reading
- **P**: Atmospheric Pressure
- **k**: constant of proportionality

$$B = kP$$

There is a better way

The Tools of Science

$$B = kP$$

$$P = \frac{B}{k}$$

$$k = \frac{B}{P}$$

$$B - kP = 0$$

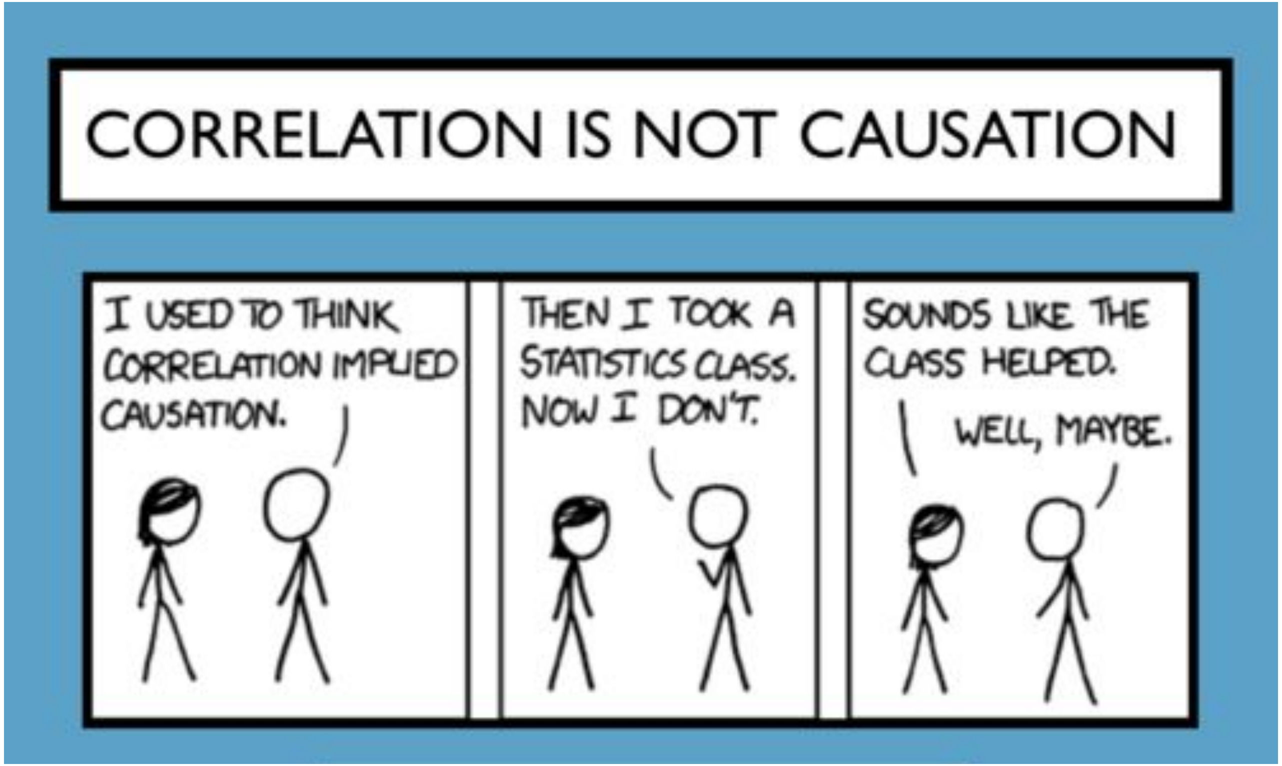
There is a better way

The Tools of Science

Probability

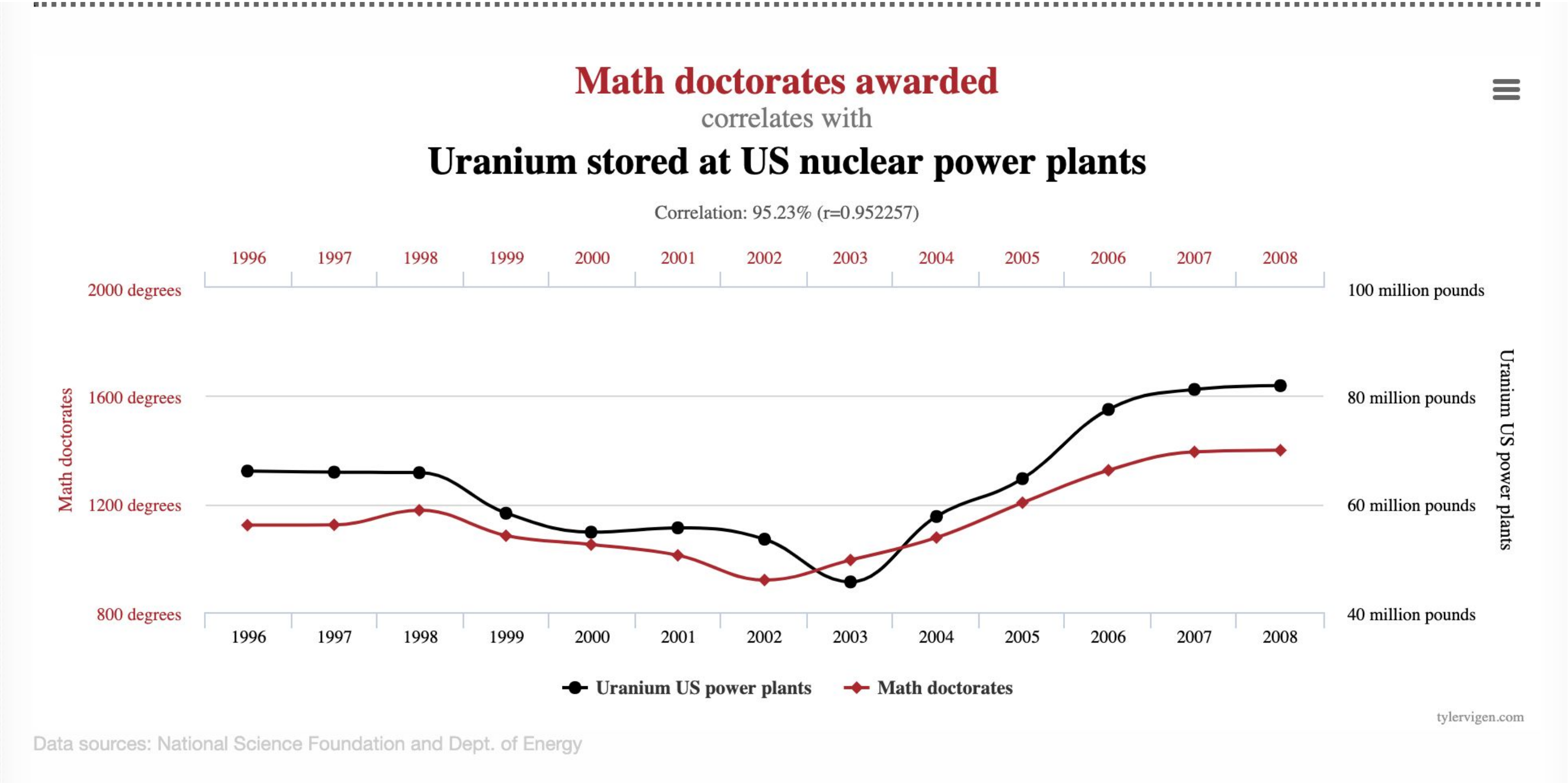
$$P(S|P = 101kPa)$$

Correlation ⇏ Causation



There is a better way

Spurious Correlations



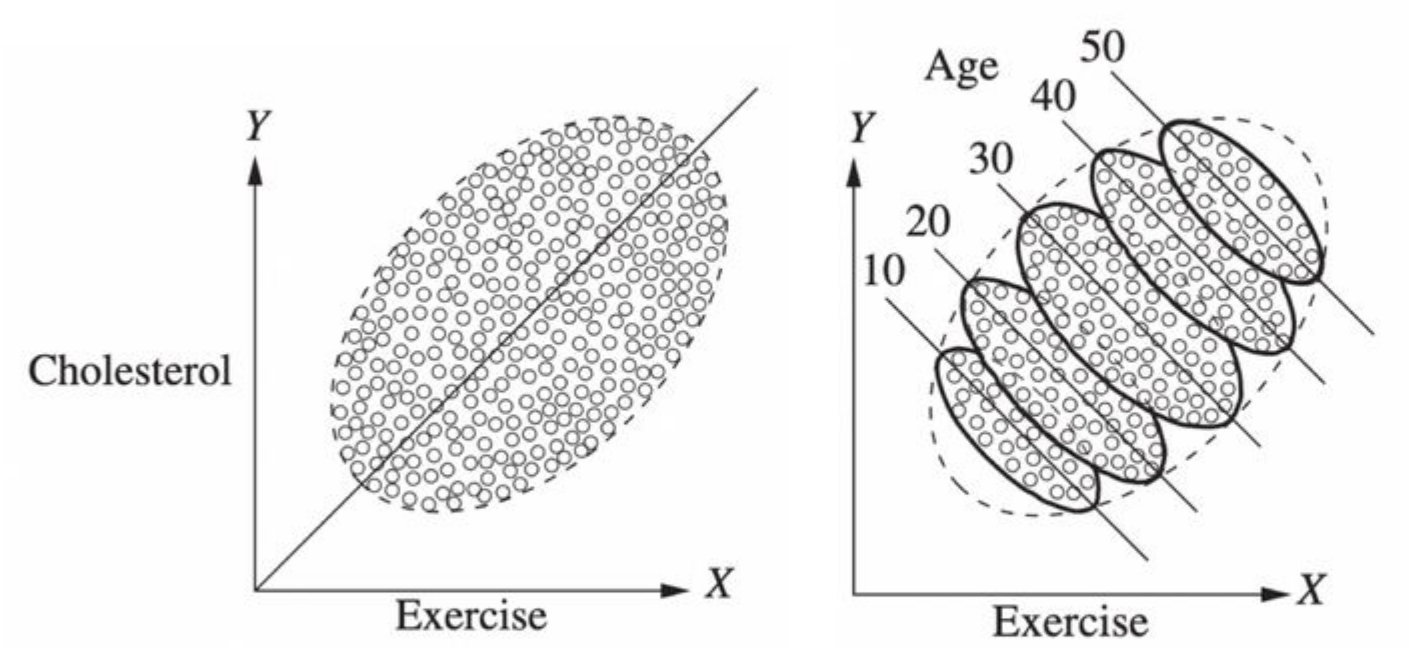
https://www.tylervigen.com/spurious-correlations

04

Simpson’s Paradox

There is a better way

Simpson’s Paradox



There is a better way

Simpson’s Paradox: Fictitious Example

Recovery Rate of a Drug

Drug taking (it’s a choice)

Recovery

Simpson's Paradox: Fictitious Data as an Example

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

There is a better way

Counter-intuitive Mathematics

$$\frac{a}{b} + \frac{c}{d} \geq \frac{e}{f} + \frac{g}{h}$$

$$\frac{a+c}{b+d} \not\geq \frac{e+g}{f+h}$$

There is a better way

Statistics, an “Objective” Field

- Modern statistics’ objective is to summarize data.
- Contingency tables alone cannot tell the story of the mechanism that generated the data.
- So far, causation has been handled by the scientists’ **intuition.**
- **New tools** are needed to manipulate causal links easily.

There is a better way

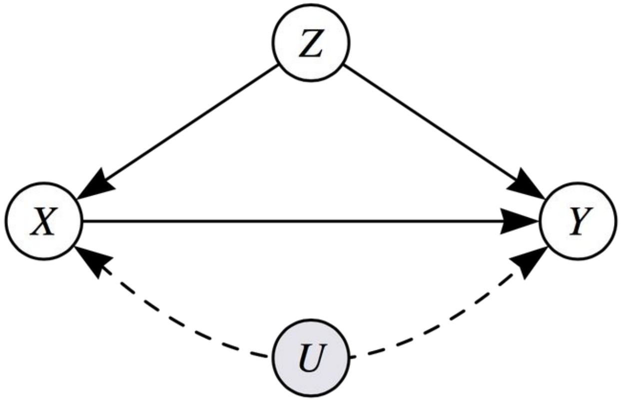
05

Towards a Causal Model

Causal Calculus

Two Components

Causal Diagram



To express what we know

Symbolic Language

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

To express what we want to know

There is a better way

Causal Calculus

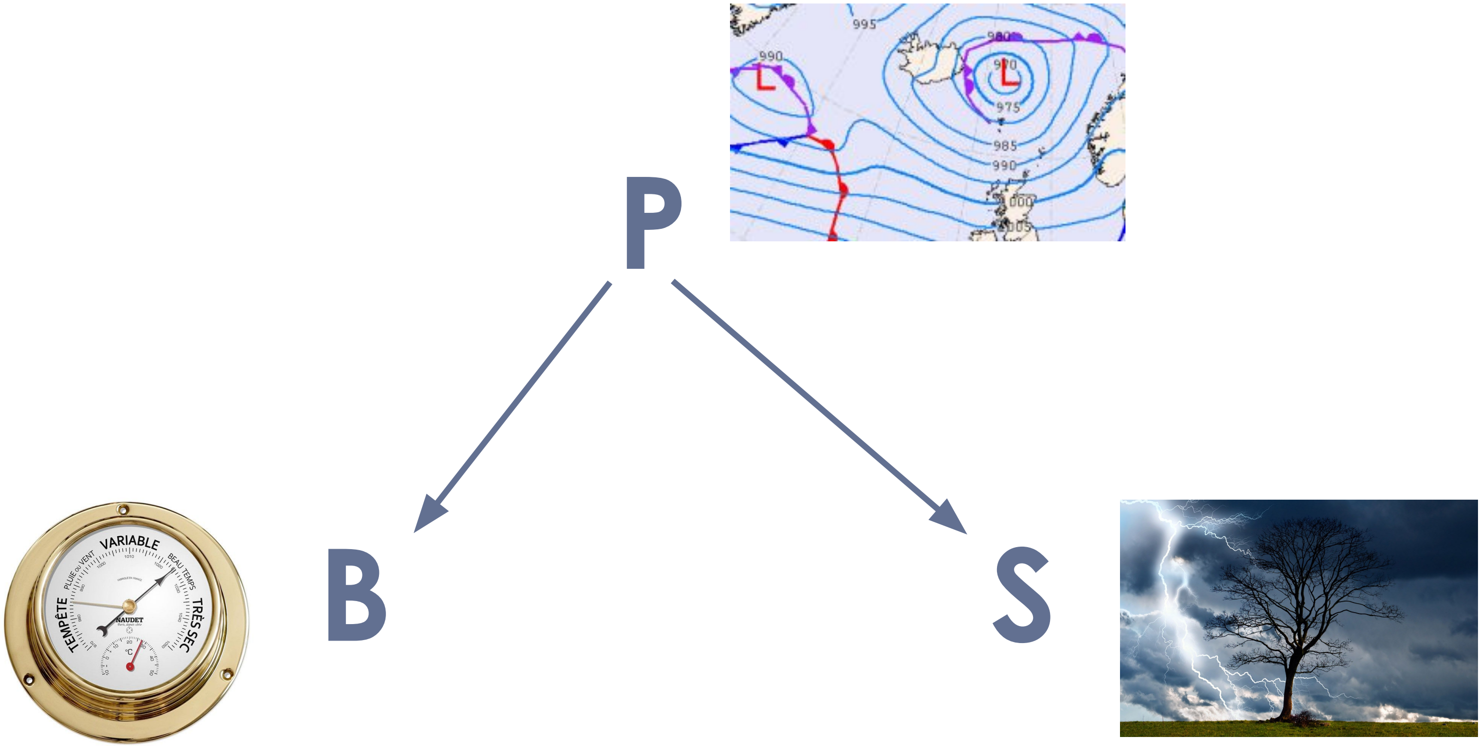
The **do()** operator

$$P(S \mid B) \quad \text{vs.} \quad P(S \mid do(B))$$

There is a better way

Causal Calculus

Causal Relationships Matter



There is a better way

Back to our Drug Example

We want to know the effect of the drug on recovery.

do(X=1)

do(X=0)

Drug taking
(X)

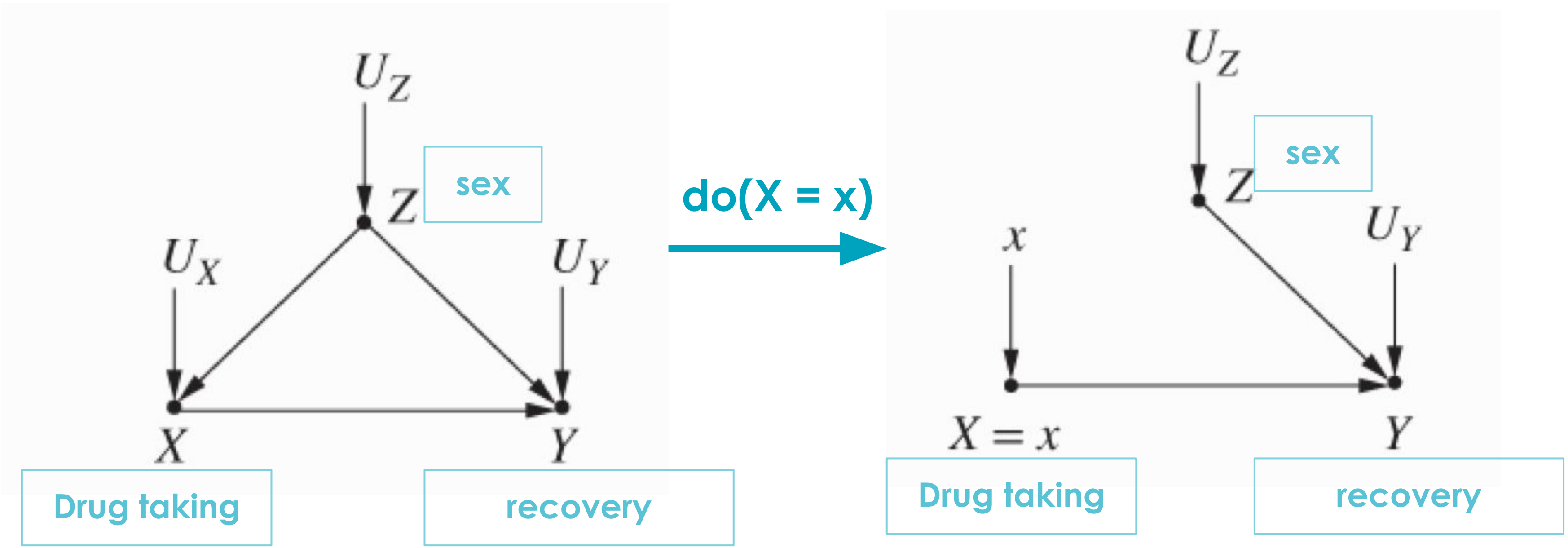
Recovery
(Y)

Calulate the Causal Effect

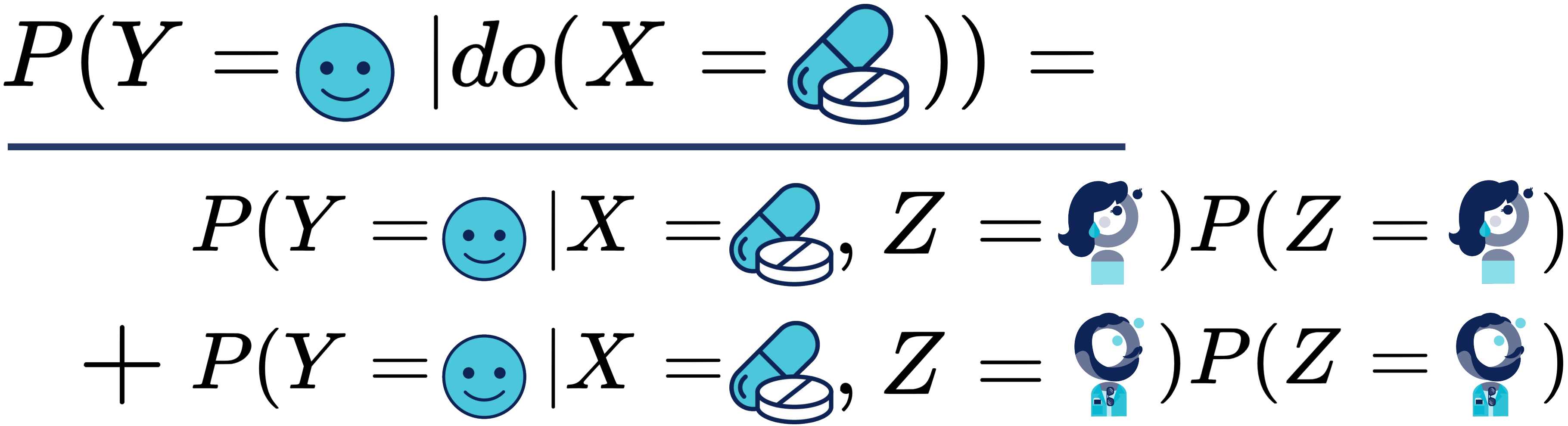
$$ACE = P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0))$$

ACE = Average Causal Effect

Do-calculus


U_Z
Z
sex
U_X
U_Y
X
Y
Drug taking
recovery
do(X = x)
U_Z
Z
sex
U_Y
x
$X = x$
Y
Drug taking
recovery

The Adjustment Formula



The Adjustment Formula

$$P(Y = y | do(X = x)) = \sum_z P(Y = y | X = x, Z = z) P(Z = z)$$

There is a better way

Back to the Original Data

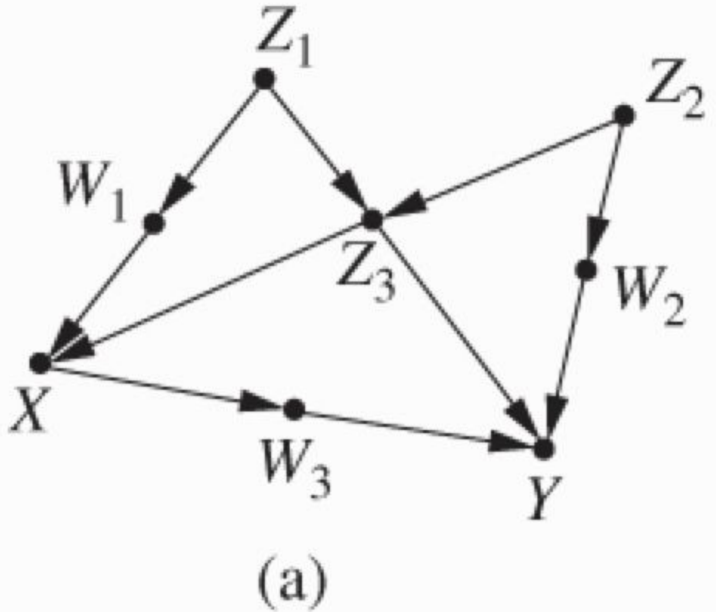
	Drug	No drug
Men (51%)	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women (49%)	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

83%

78%

There is a better way

Graphs Can Get More Complex



There is a better way

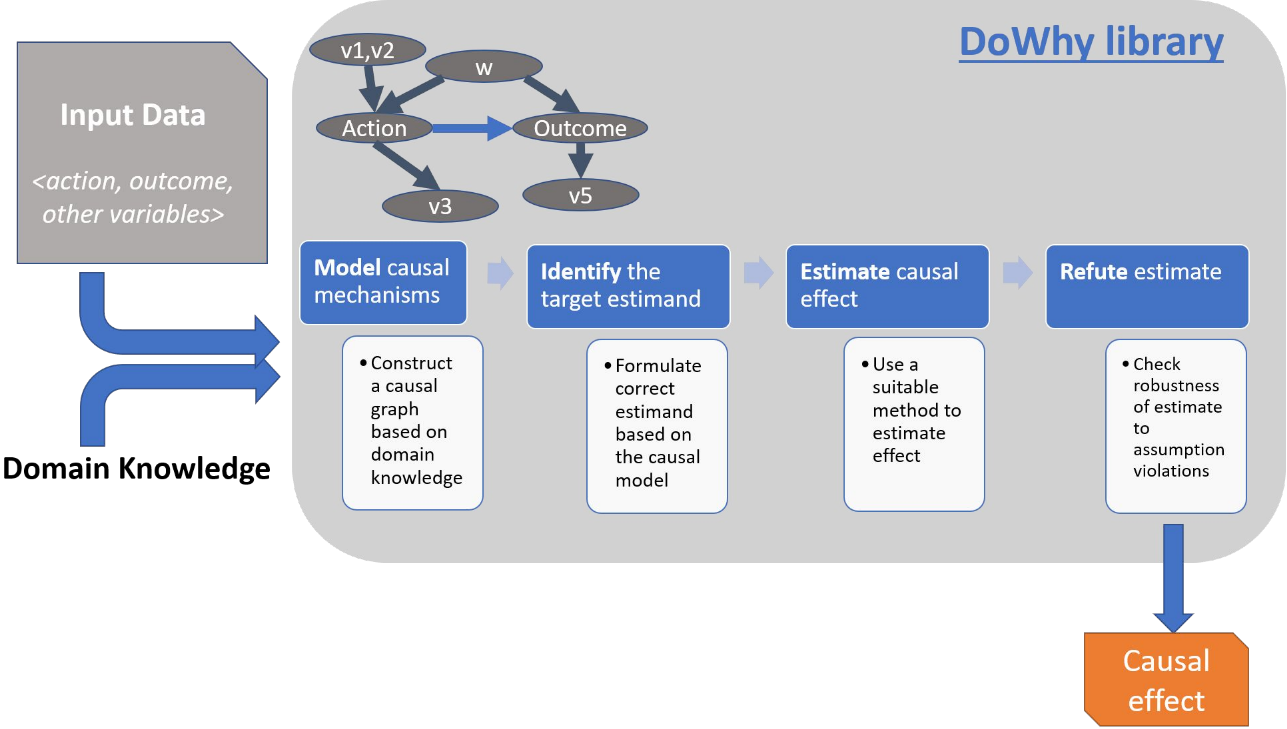
There is a better way

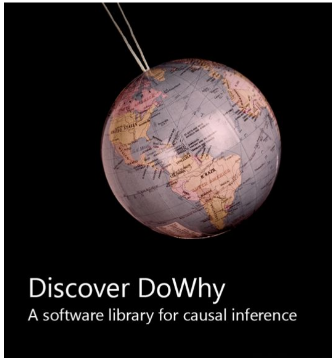
06

Inference Engines

The DoWhy Library

by Microsoft



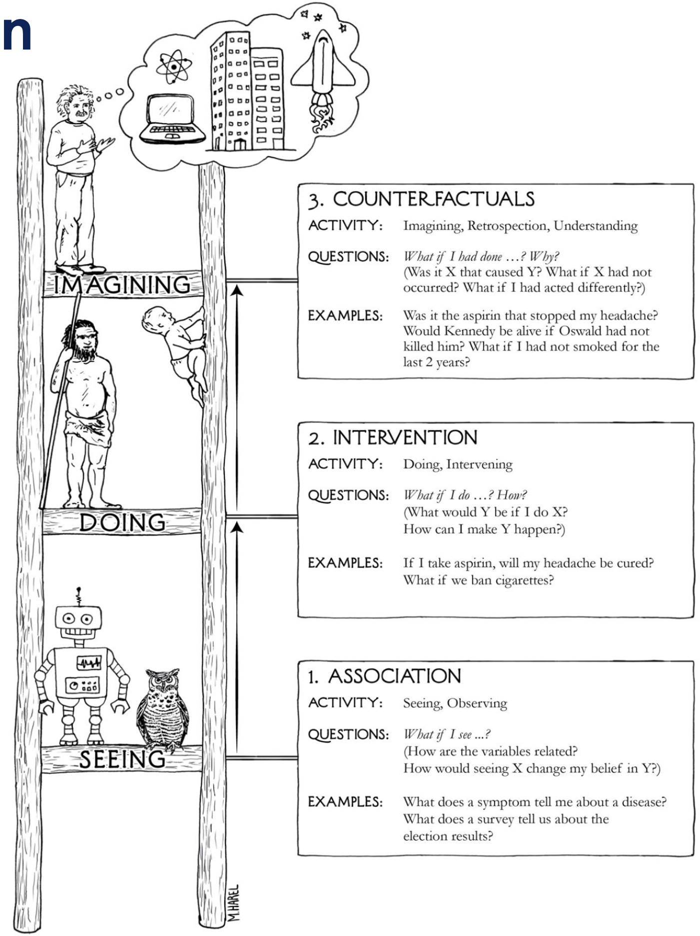


There is a better way

07

Conclusion

The Ladder of Causation



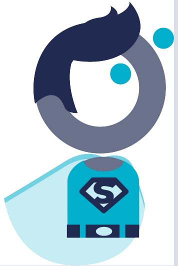
There is a better way

The Potential of Causal Inference

How could we use it in our Data Science projects?

Strengths

- Precise modelling of **causal links**
- Libraries can handle **complex** interactions.
- Transparent and **interpretable** model.
- Leverages **domain expertise.**



Difficulties

- MUST leverage **domain expertise.**
- Forces the consultant to **look beyond the data** themselves.
- Causal diagrams can be **tedious** to build.
- **Any other ideas?**

There is a better way

There is a Better Way