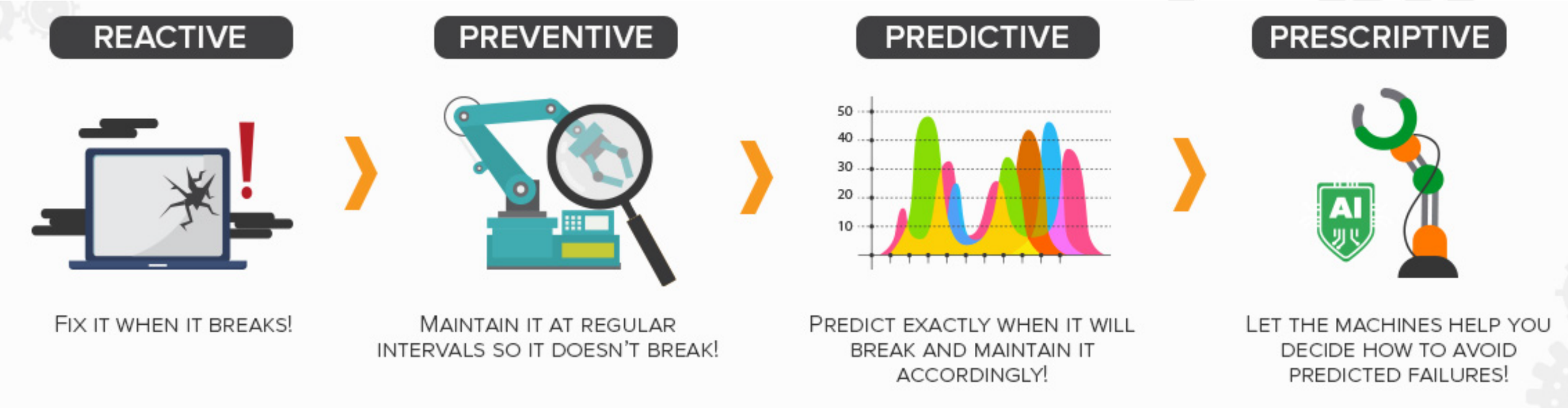
Causal inference and discovery

Tom Heskes
Radboud University Nijmegen

PrimaVera colloquium

June 24, 2020

Evolution of maintenance strategies



from https://limblecmms.com/blog/predictive-maintenance/

Smoking and lung cancer

Results

- clear correlation
- strong risk factor for lung cancer

Source: OECD Health Data 2010.

Chocolate consumption and Nobel prizes

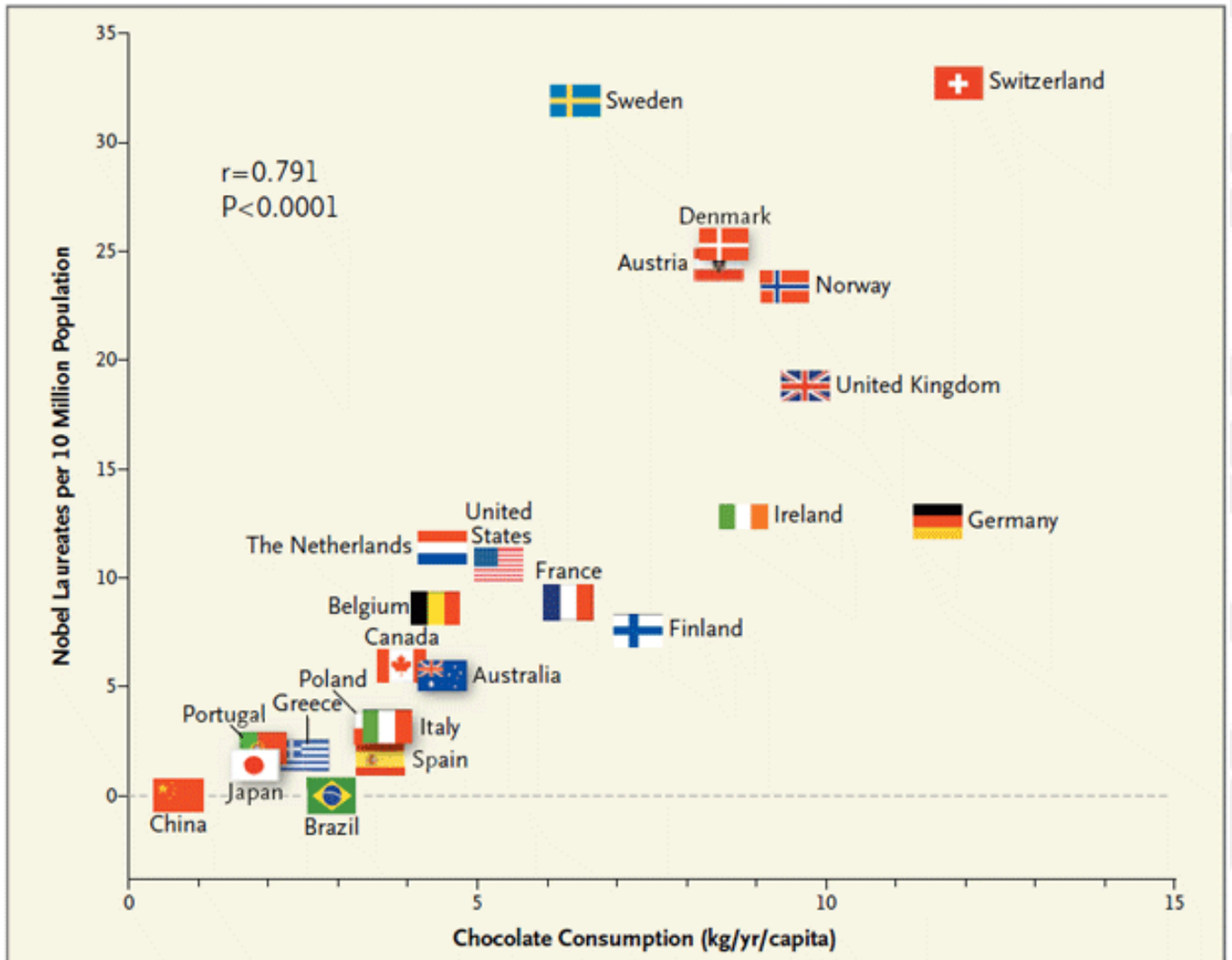


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

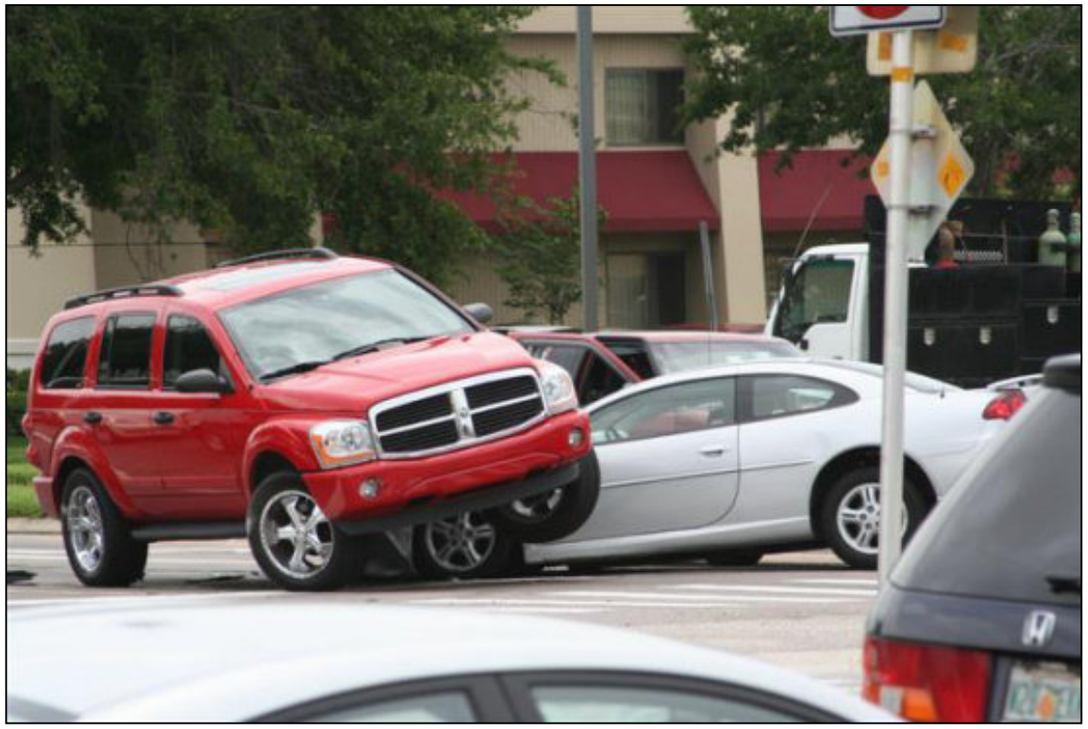
NEW ENGLAND JOURNAL OF MEDICINE

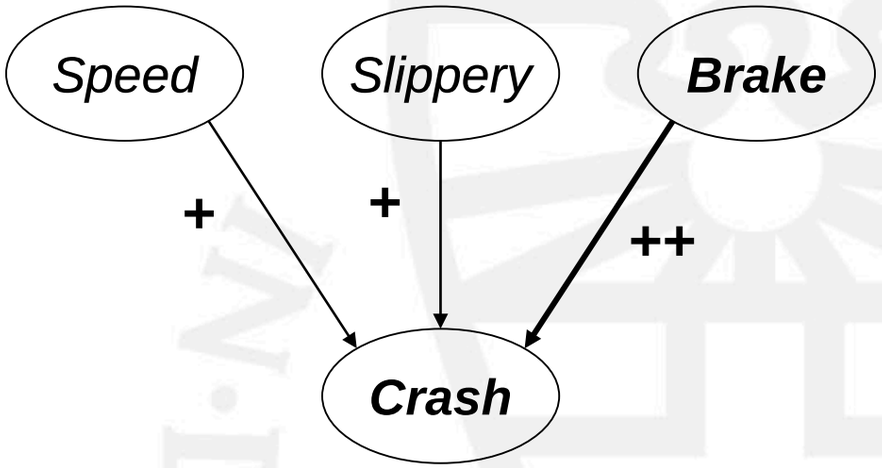
Results

- even stronger correlation!
- good predictor of chance on Nobel prize...

Messerli, "Chocolate Consumption, Cognitive Function, and Nobel Laureates", New England Journal of Medicine, 2012

Accident hot spots



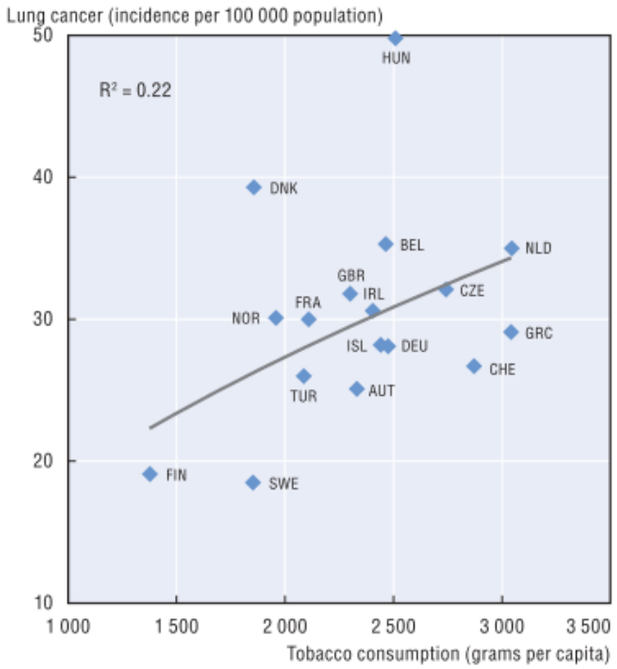


regression coefficients

Results

- strong positive correlation between *Braking heavily* and *Car Crash*?

From observation to action



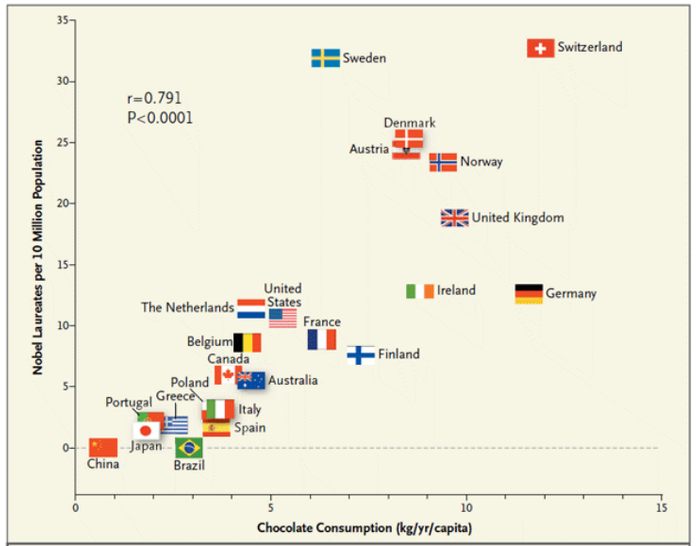
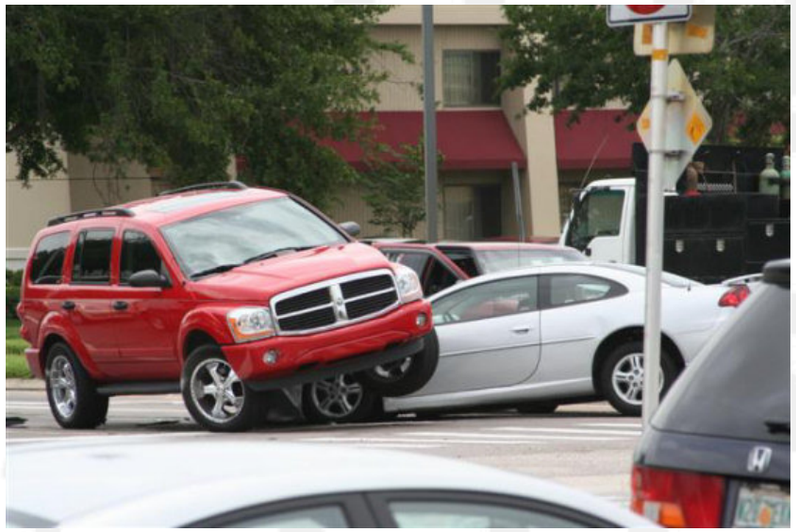


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

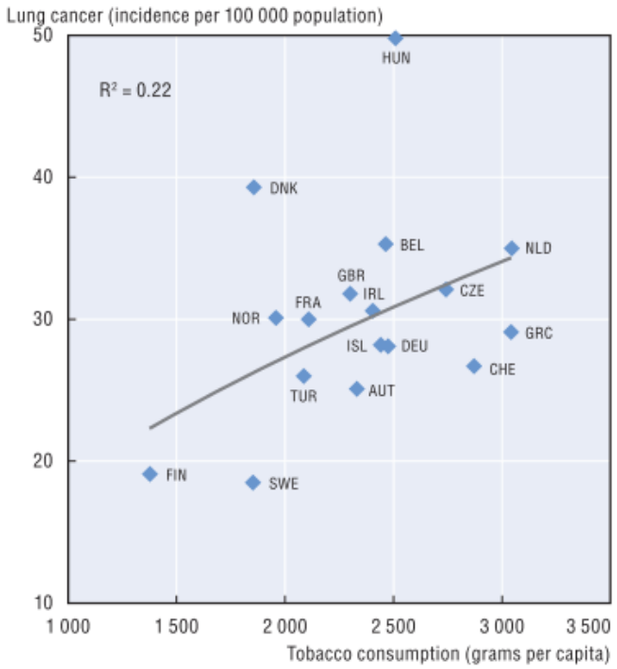
NEW ENGLAND JOURNAL OF MEDICINE



- correlations describe the world as we see it
- causal relations predict how the world will change when we intervene

$\Rightarrow$ main goal of causal inference and causal discovery

Challenge: recognize causal pathways from data



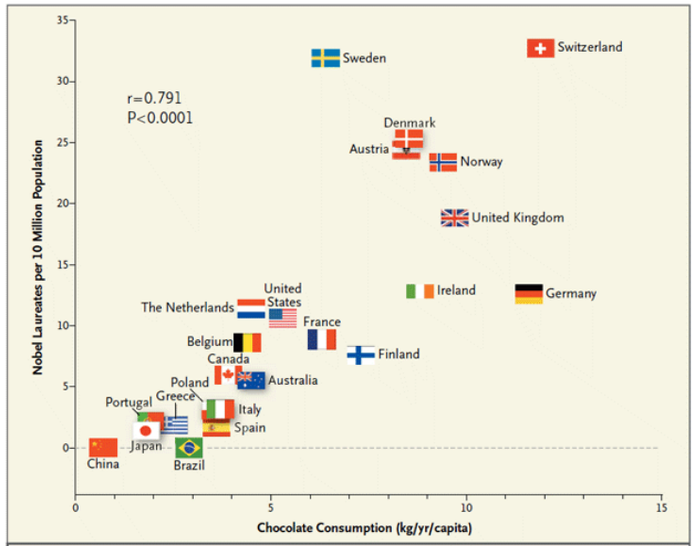
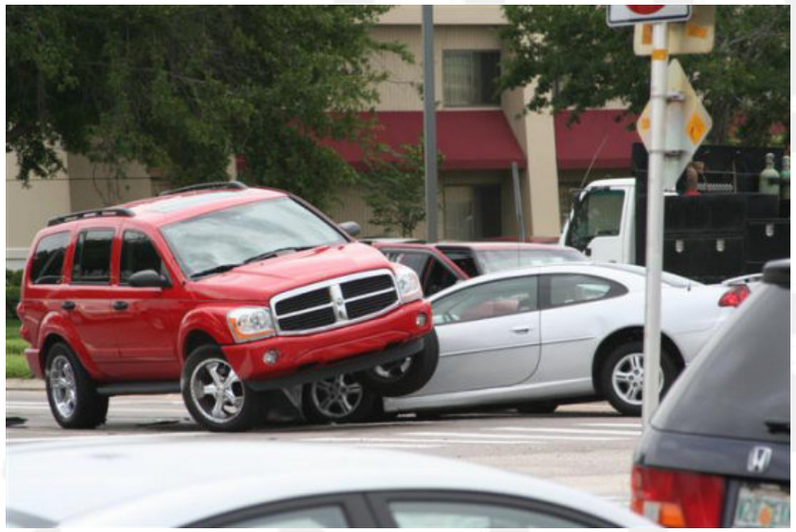
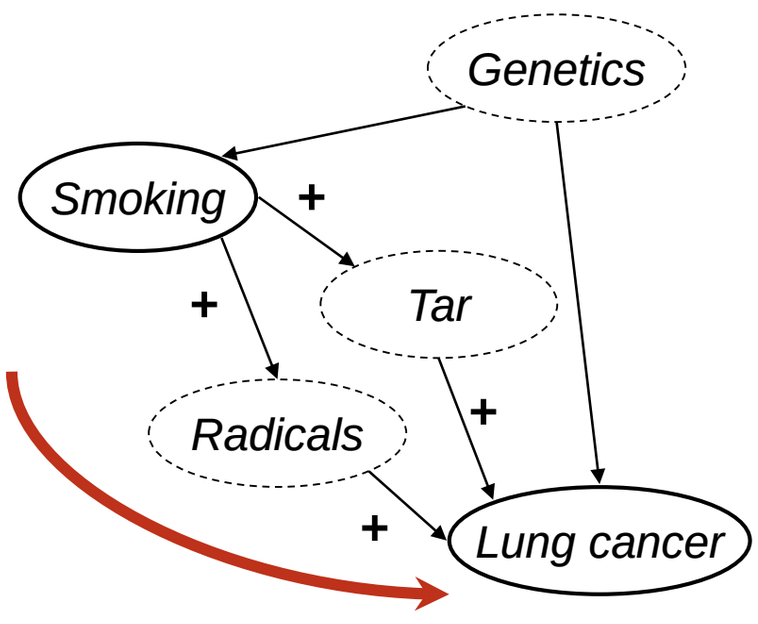
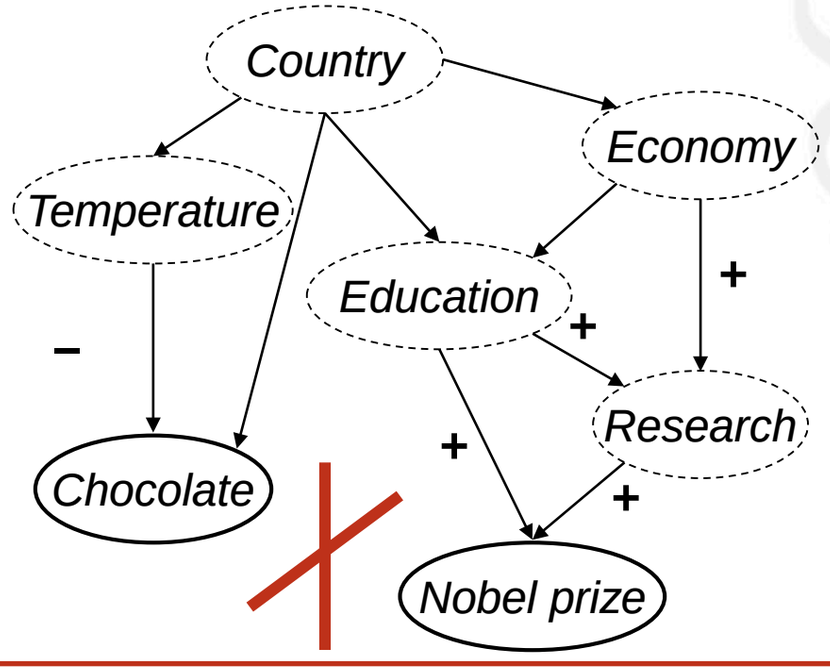


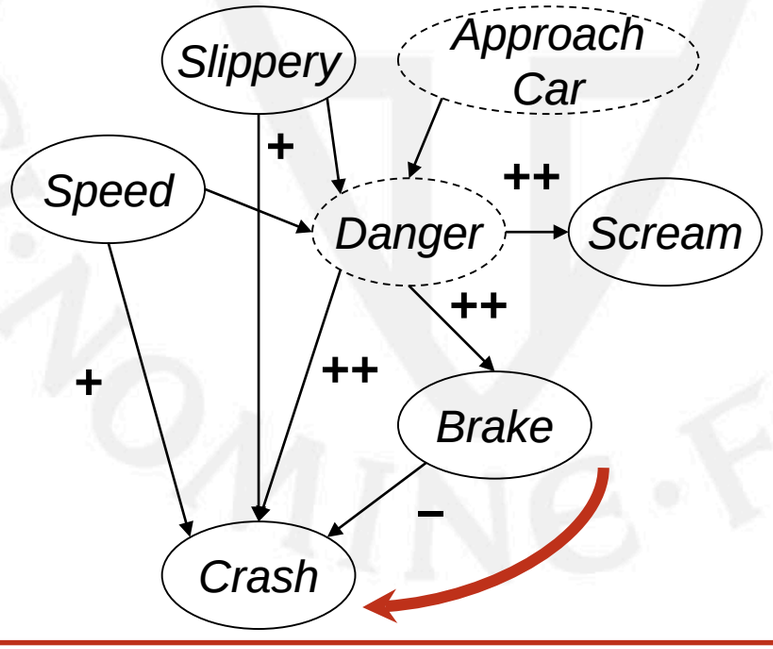
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

NEW ENGLAND JOURNAL OF MEDICINE

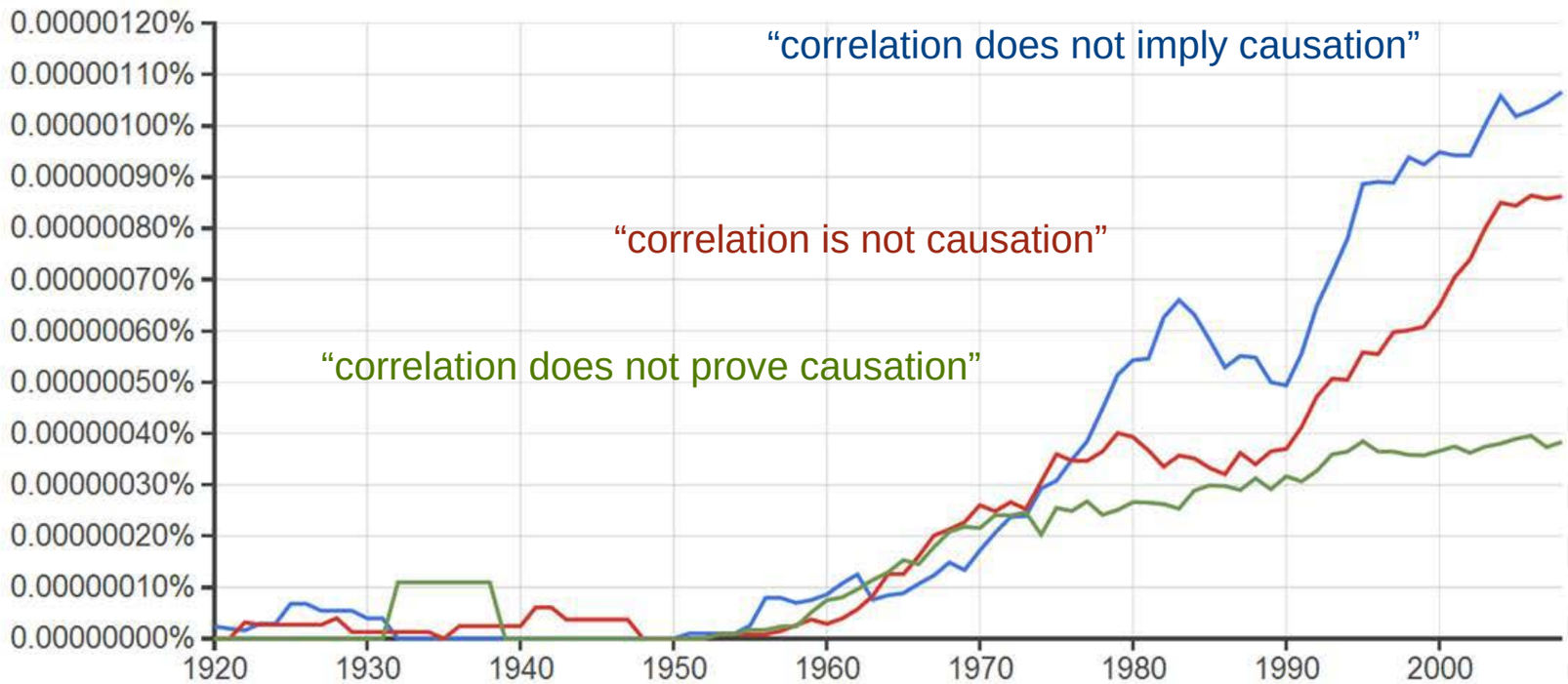


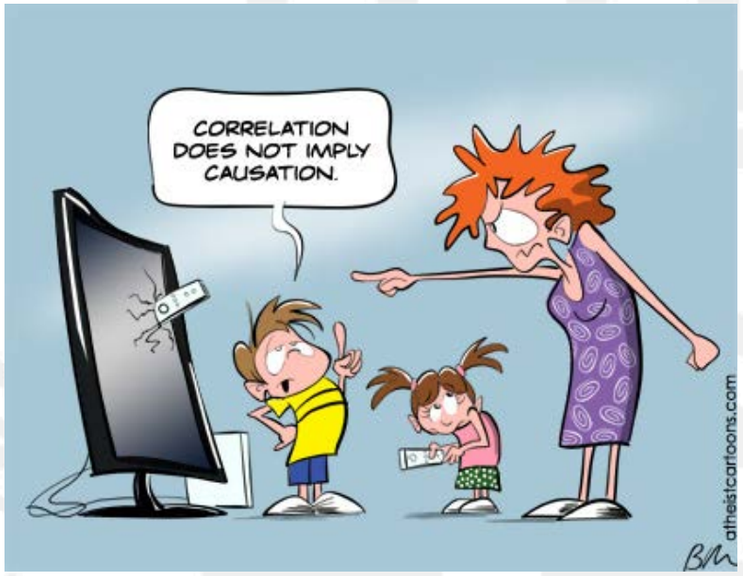






A popular saying





Why do people love to say that correlation does not imply causation?

Daniel Engber: “The internet blowhard’s favorite phrase”
http://www.slate.com/articles/health_and_science/science/2012/10/correlation_does_not_imply_causation_how_the_internet_fell_in_love_with_a_stats_class_clich_.html

Big data and causality

[...] society will need to shed some of its obsession for causality in exchange for simple correlations: not knowing *why* but only *what*. This overturns centuries of established practices and challenges our most basic understanding of how to make decisions and comprehend reality.

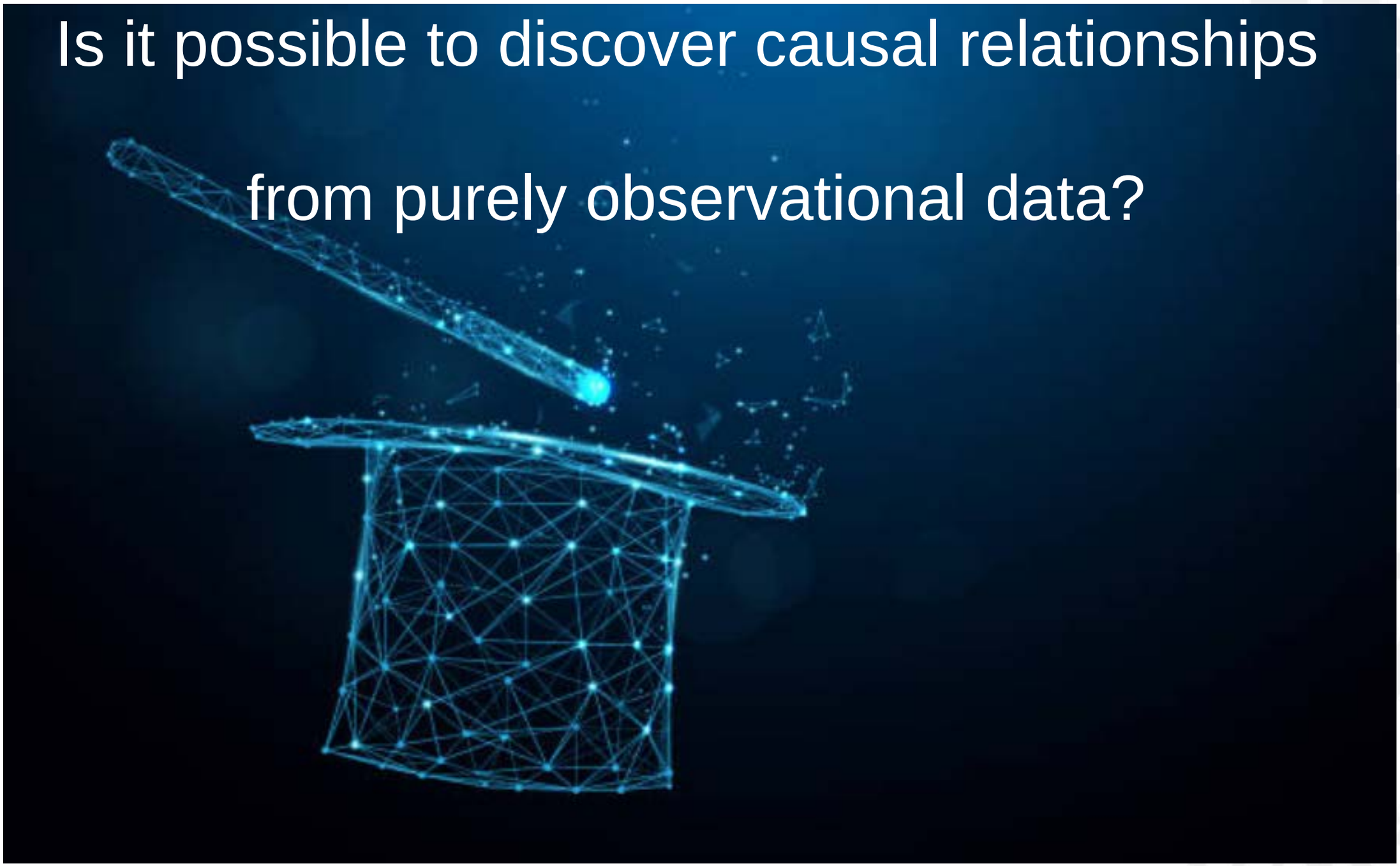
Mayer-Schönberger & Cukier

Logical reasoning

correlation does not imply causation

thus

it is impossible to discover causal relationships from purely observational data


Is it possible to discover causal relationships
from purely observational data?

Logical fallacy

correlation does not imply causation

thus

it is impossible to discover causal relationships from purely observational data

In fact

a single, simple correlation does not imply causation

yet

it is possible to discover causal relationships from purely observational data
(which of course requires some assumptions, as any statistical approach)

Statistical inference

- Given a set of input-output samples and background assumptions, and input observations **X**, predict the outcome $Y(\mathbf{X})$

- E.g., train a deep neural network to detect cracks in images of bridges (Zaharah and Nils' paper)
- Fitting crack depths (Tiedo)
- Prediction of failures with 'black box' techniques (Geert-Jan)
- Learn decision (fault) trees (Marielle)

Causal inference

- Given a set of input-output samples, possibly (but not necessarily) gathered under different manipulations, background assumptions, and a manipulation M and observations $\mathbf{X}$, predict the outcome $Y(\mathbf{X}, \text{do}(M))$

Typically decomposed into two steps:

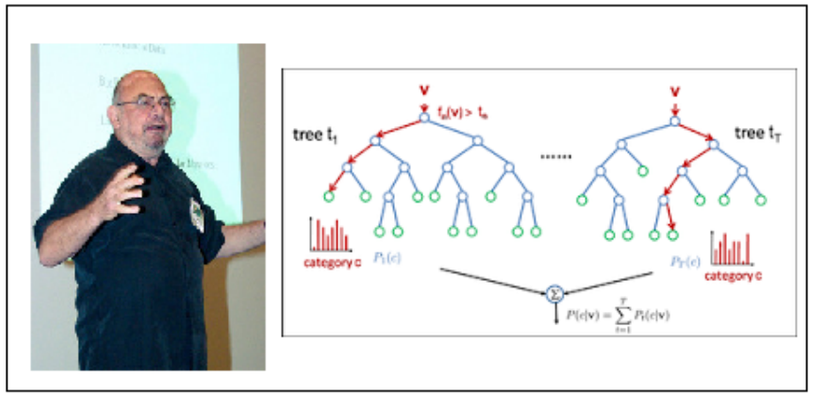
1. Given a set of samples, possibly (but not necessarily) gathered under different manipulations and background assumptions, construct a (set of) causal model(s) $\mathbf{C}$

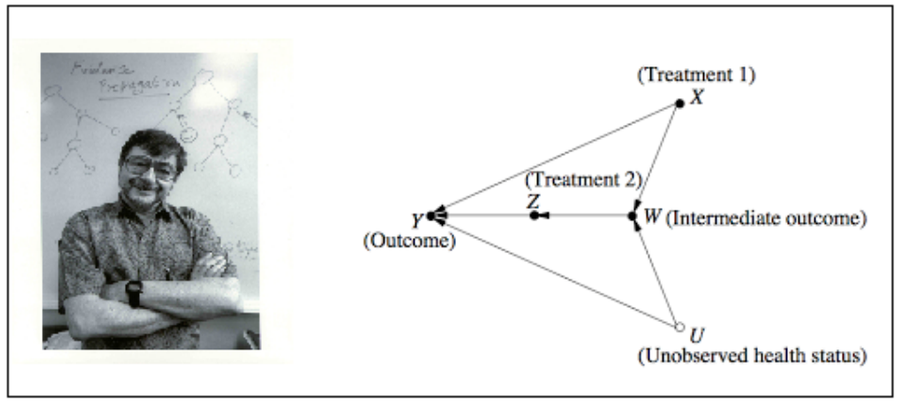
2. Given a (set of) causal model(s) $\mathbf{C}$, manipulation M and observations $\mathbf{X}$, predict the outcome $Y(\mathbf{X}, \text{do}(M))$

Outline

- Motivation
- Causal inference (known causal structure)
- Causal discovery (unknown causal structure)

Causal forests



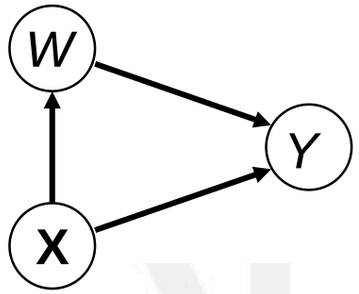


Estimation and Inference of Heterogeneous Treatment Effects using Random Forests*

Stefan Wager
Department of Statistics
Stanford University
swager@stanford.edu

Susan Athey
Graduate School of Business
Stanford University
athey@stanford.edu

November 22, 2016



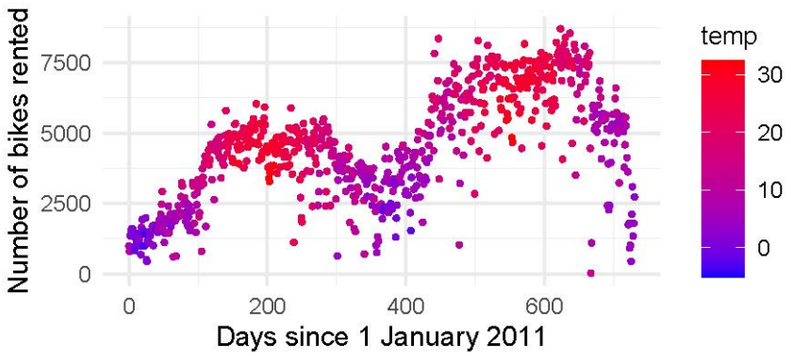
$$\hat{\tau}(x) = \frac{1}{|\{i : W_i = 1,\, X_i \in L\}|} \sum_{\{i:W_i=1,\, X_i \in L\}} Y_i - \frac{1}{|\{i : W_i = 0,\, X_i \in L\}|} \sum_{\{i:W_i=0,\, X_i \in L\}} Y_i.$$

Counterfactual Fairness

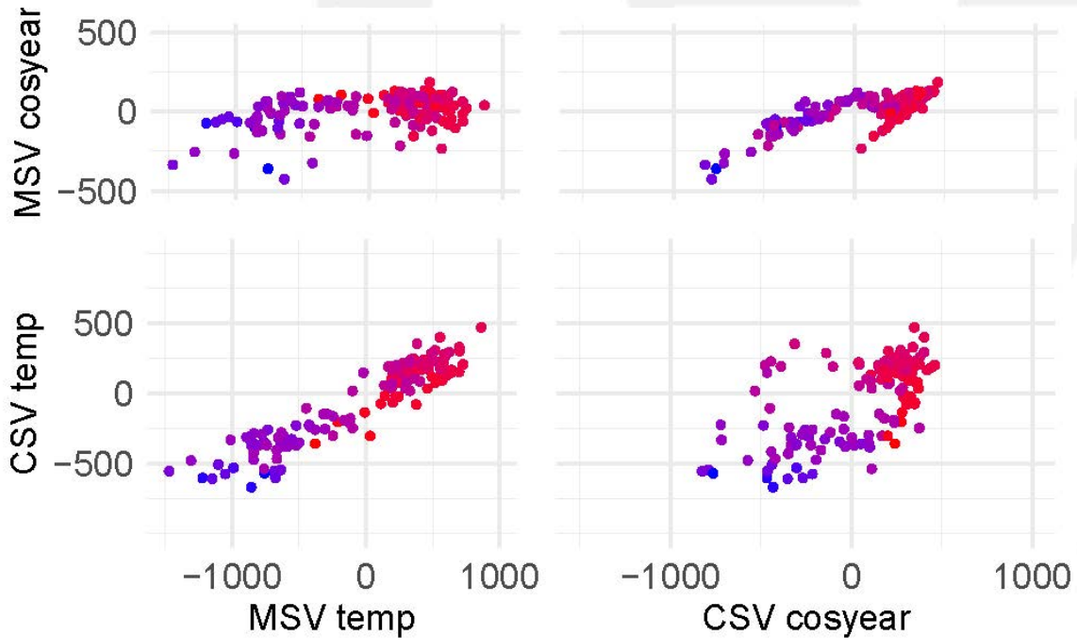
Kusner et al., NIPS 2017

Causal Shapley values

- Shapley values: *the* state-of-the-art approach for explaining individual predictions of complex ‘black-box’ machine learning models
- Standard Shapley values (MSV) assume independence of model inputs
- Causal Shapley values (CSV) apply causal inference on the model inputs to estimate the total effect each input has on the model’s prediction



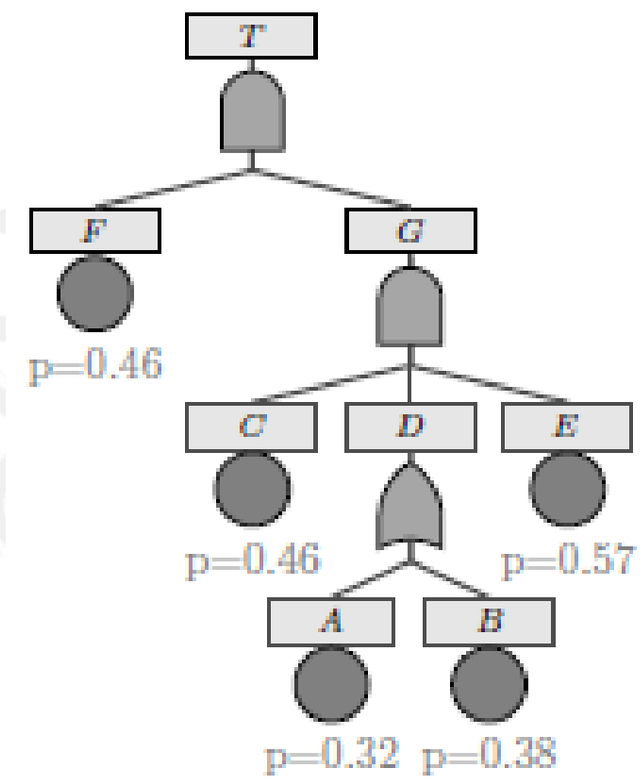
Bike rental data set: MSV’s attribute model predictions predominantly to weather, CSV’s properly represent the effect of season.



LIFT

Nauta, Bucur, Stoelinga, QEST 2018

- Learning fault trees from observational data
- Aims to find a causal chain of events leading to global system failure
- Underlying assumption that all events are causally related
- Iteratively searches for the strongest direct associations using Mantel-Haenszel test



Outline

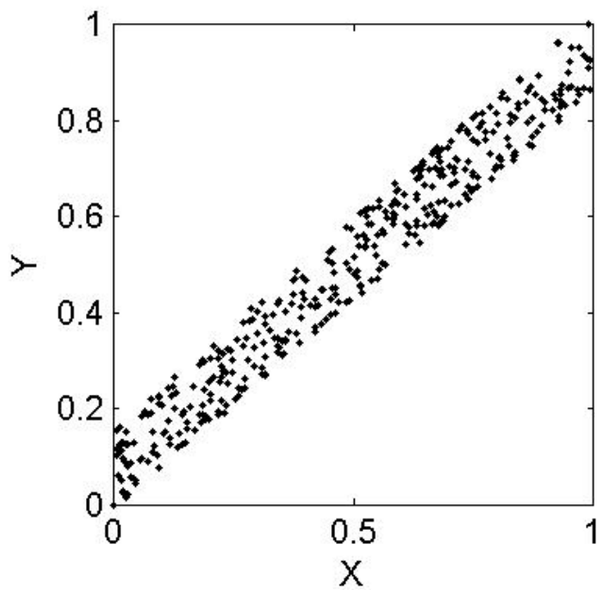
- Motivation
- Causal inference (known causal structure)
- Causal discovery (unknown causal structure)

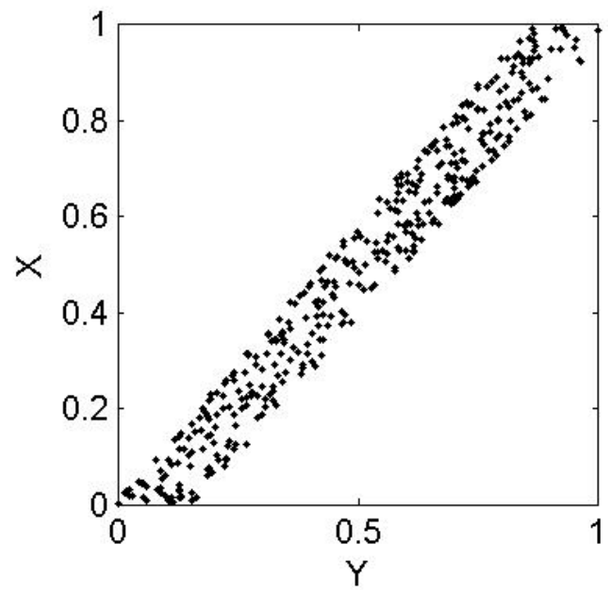
Causal discovery

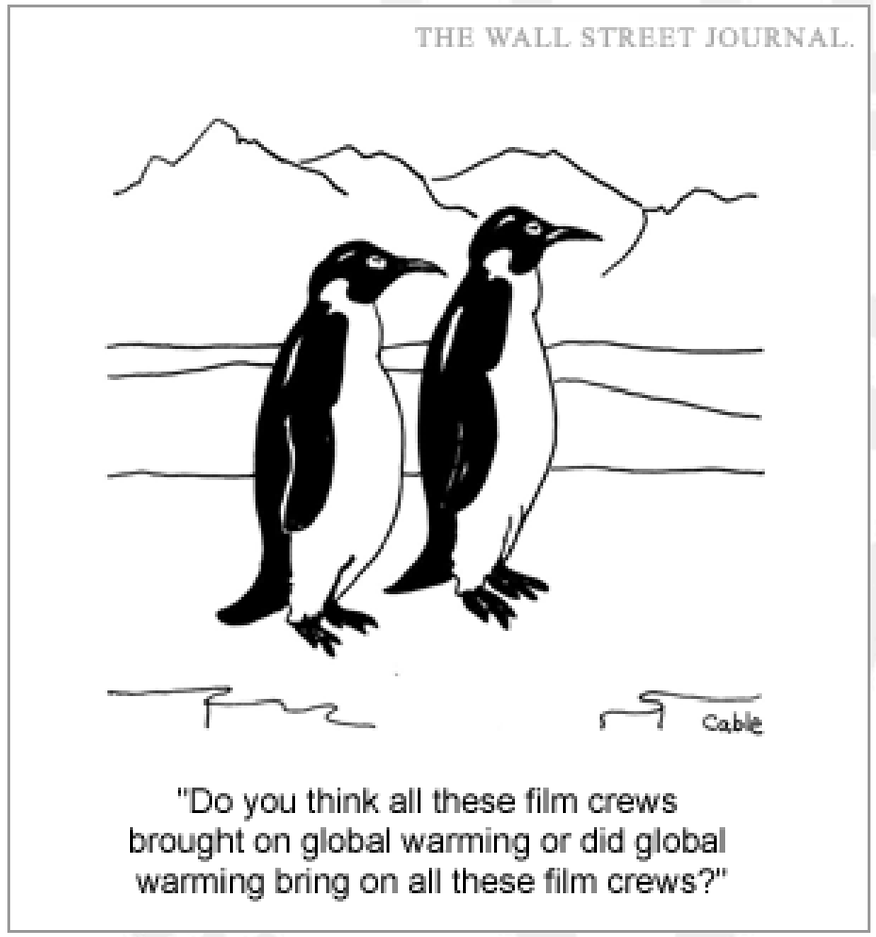
- Focuses on the first step of causal inference: inferring the structure of causal models from data
- Typically from purely observational data
- Mainly used for exploratory data analysis
- Three main types:

1. Pairwise, using higher-order statistics
2. Multiple variables, using conditional (in)dependencies
3. Using temporal data, e.g., Granger causality, transfer entropy

Causal pairs

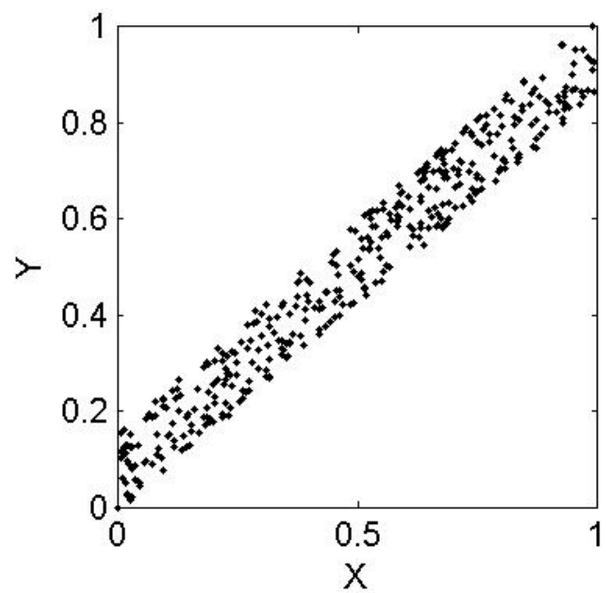


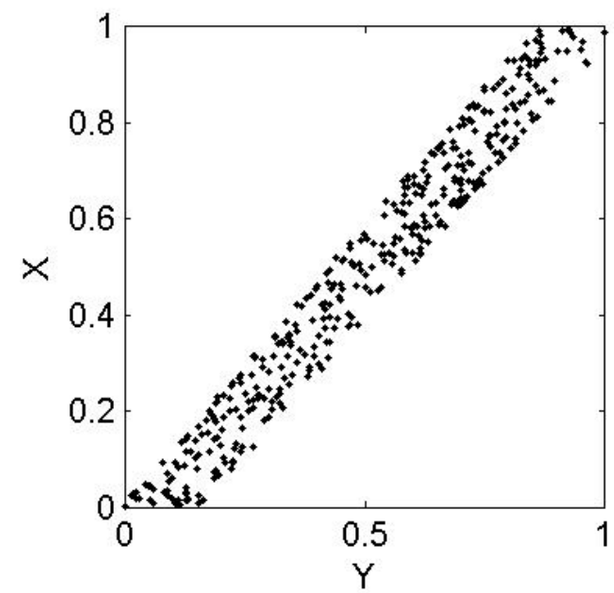




does X cause Y or does Y cause X?

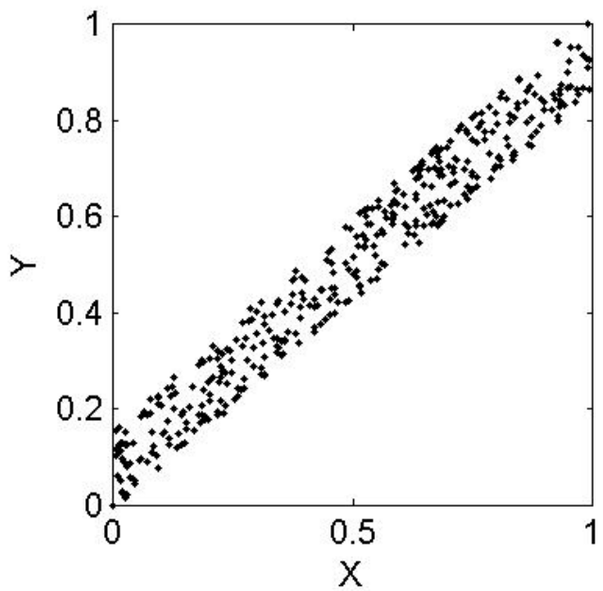
Which one is more likely?

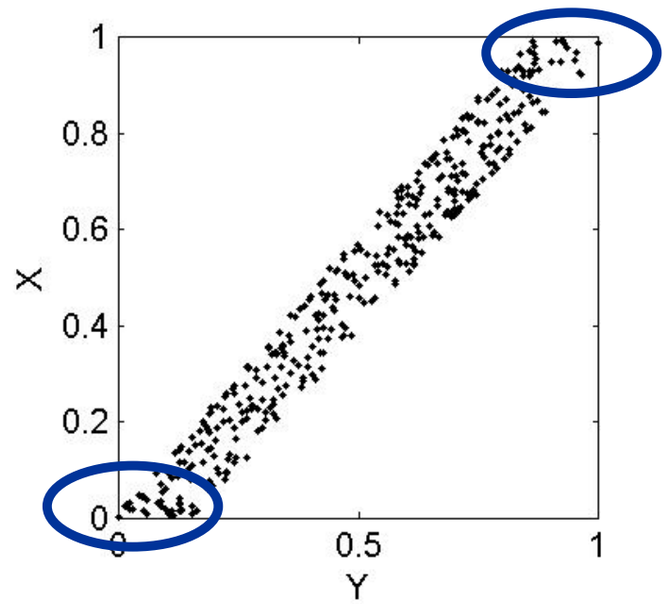




X causes Y or Y causes X

Causal direction

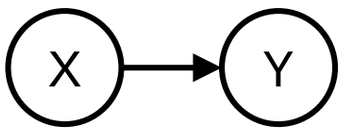




Ockham chooses a razor

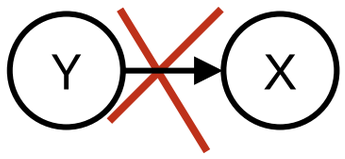
easy to explain as

$Y = f(X) + \text{noise}$

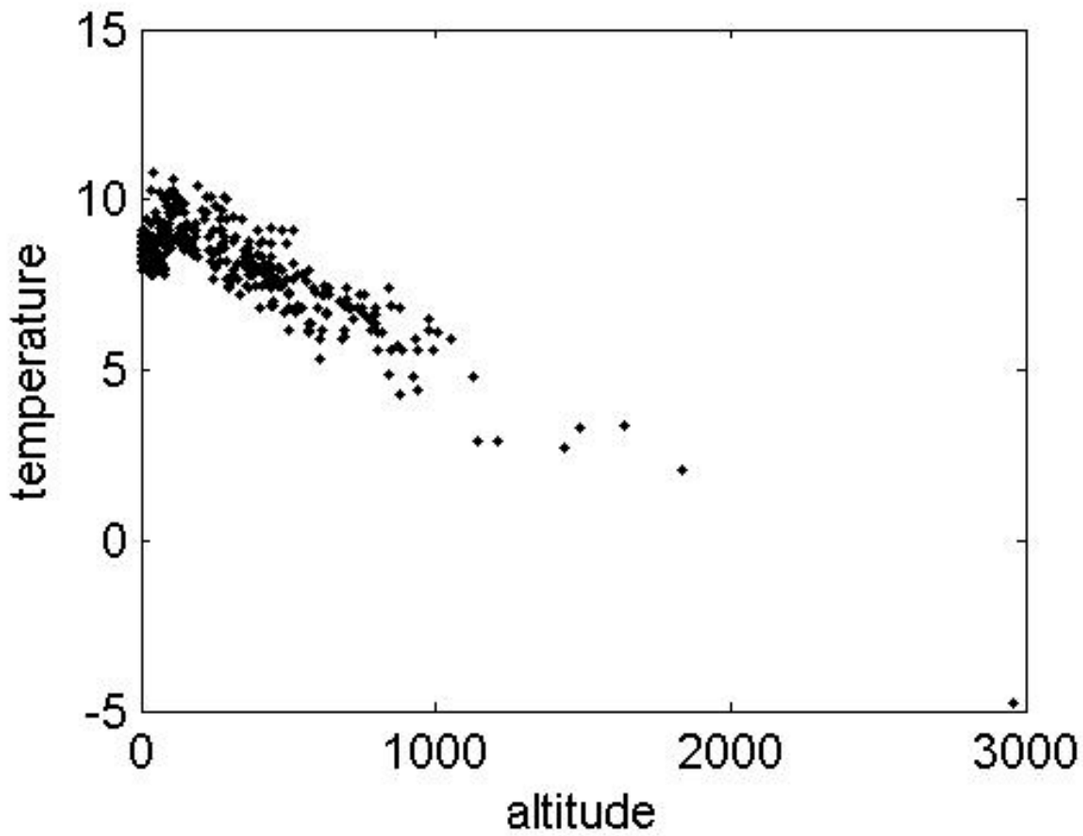


difficult to explain as

$X = g(Y) + \text{noise}$

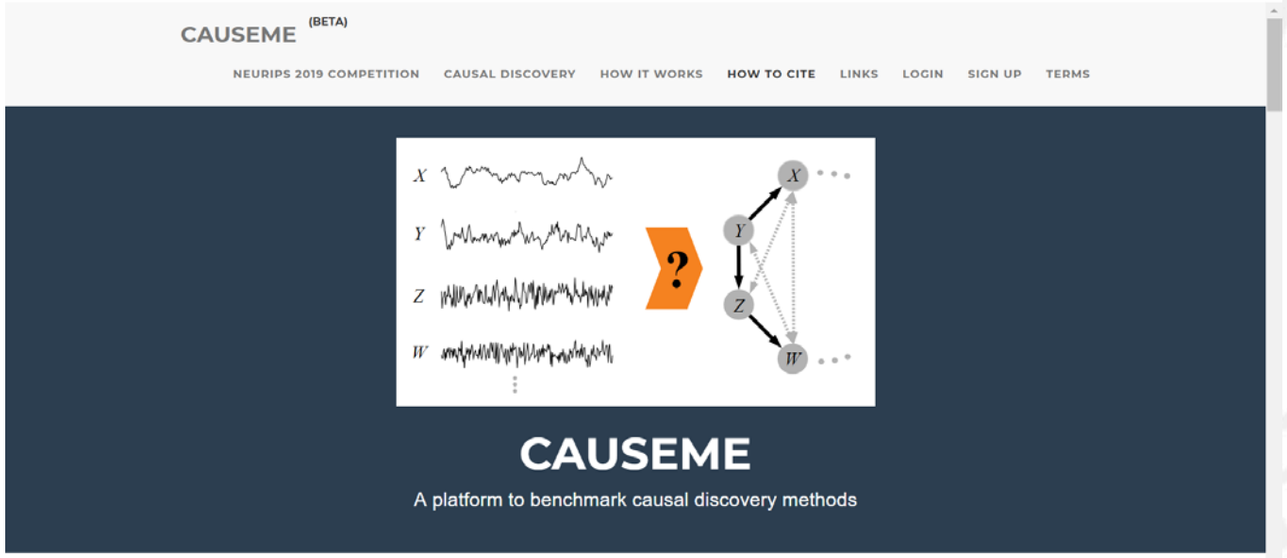


Real-world cause-effect pairs



X: altitude of weather station

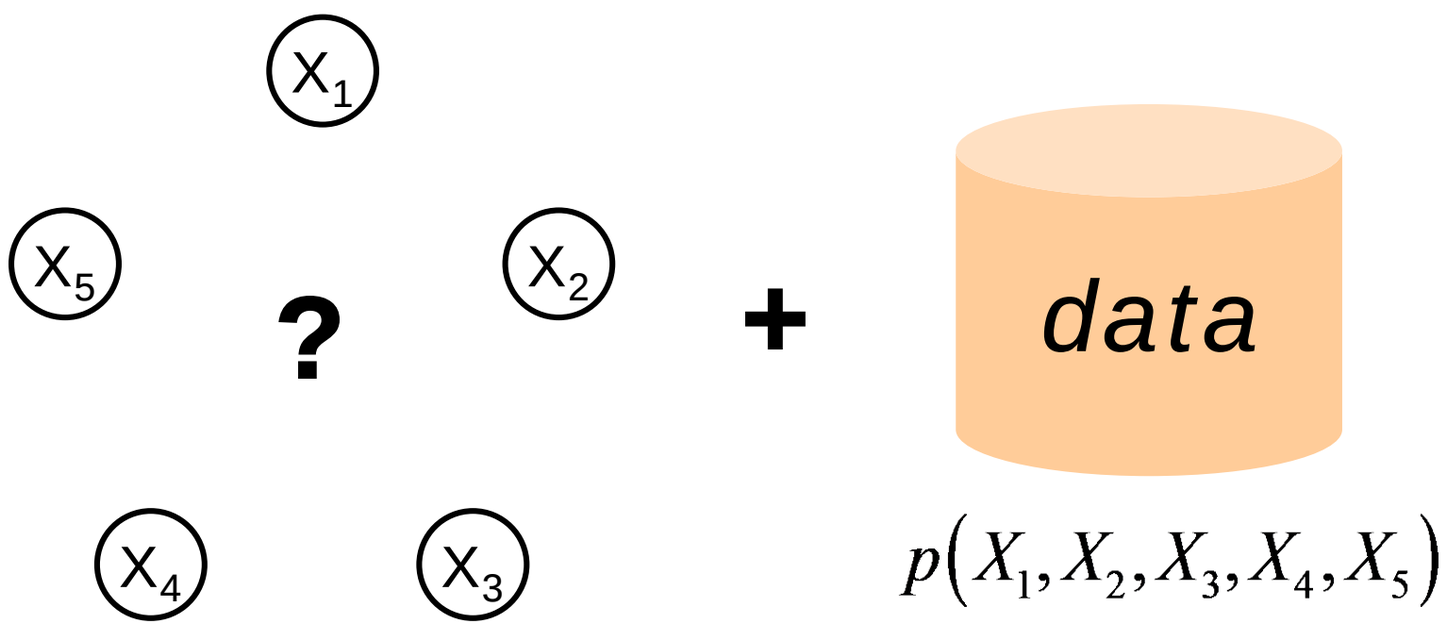
Y: temperature (average over 1961-1990)



http://webdav.tuebingen.mpg.de/cause-effect/
http://www.kaggle.com/c/cause-effect-pairs
https://causeme.uv.es

More than two variables

- Given observed data from some distribution $p(X_1,..,X_d)$
- Some reasonable assumptions,
- Can we still predict $p(X_j \mid do(X_i = \xi))$?



$$= p\left(X_j \middle| do\left(X_i = \xi\right)\right)?$$

From causal graph to (in)dependencies and back

- Given a causal graph, we can read off all conditional (in)dependencies
- For causal discovery we need to invert this and reason in the opposite direction:

Given an observed set of conditional (in)dependencies, e.g., derived from a set of data, what can we say about the underlying causal graph?

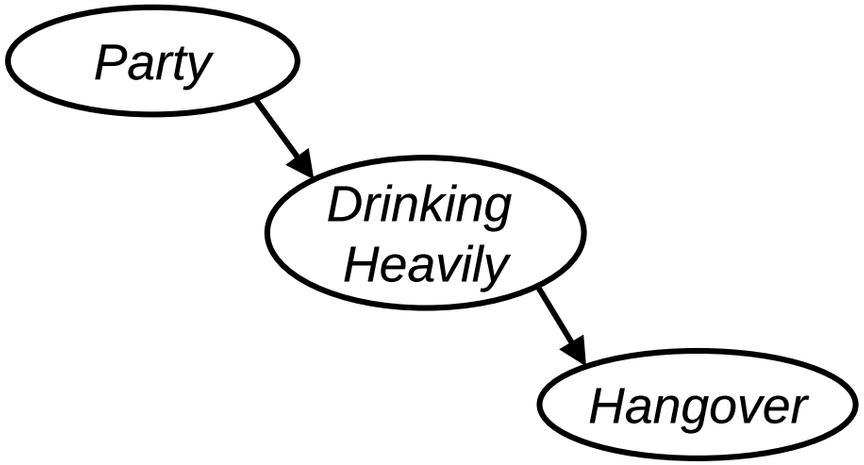
Key connection: two rules

1. $X \perp\!\!\!\perp Y \,|\, [Z] \quad : \quad (Z \Rightarrow X) \vee (Z \Rightarrow Y)$

square brackets denote 'minimal'

"is a cause of"

"if variable Z makes variables X and Y independent, then Z must have a causal relation to X and/or Y"



Minimal conditional independence

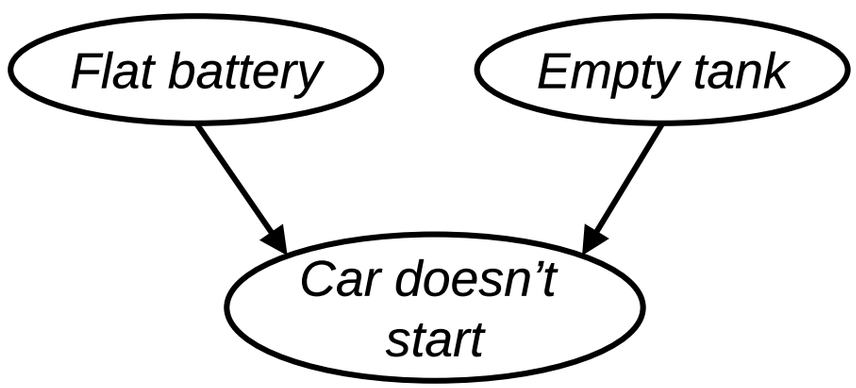
Reasoning:
every possible DAG in which variables *X* and *Y* are dependent when we do not condition on *Z*, yet become independent when we do condition on *Z*, has a (possibly indirect) directed path from *X* to *Z* and/or from *Y* to *Z*

Key connection: two rules

1. $X \perp\!\!\!\perp Y \mid [Z] \quad : \quad (Z \Rightarrow X) \vee (Z \Rightarrow Y)$
2. $X \not\perp\!\!\!\perp Y \mid [Z] \quad : \quad (Z \not\Rightarrow X) \wedge (Z \not\Rightarrow Y)$

"is NOT a cause of"

"if variable Z makes variables X and Y dependent, then Z cannot have a causal relation to X and/or Y"



Reasoning:
a *DAG* in which variables *X* and *Y* are independent when we do not condition on *Z*, yet become dependent when we do condition on *Z*, cannot have a directed path from *Z* to *X*, nor from *Z* to *Y*

Minimal conditional dependence ('v-structure')

Logical Causal Inference (LoCI)

1. $X \perp\!\!\!\perp Y \mid [Z] \quad : \quad (Z \Rightarrow X) \vee (Z \Rightarrow Y)$
2. $X \not\perp\!\!\!\perp Y \mid [Z] \quad : \quad (Z \not\Rightarrow X) \wedge (Z \not\Rightarrow Y)$
3. [something slightly more complicated, needed for completeness]

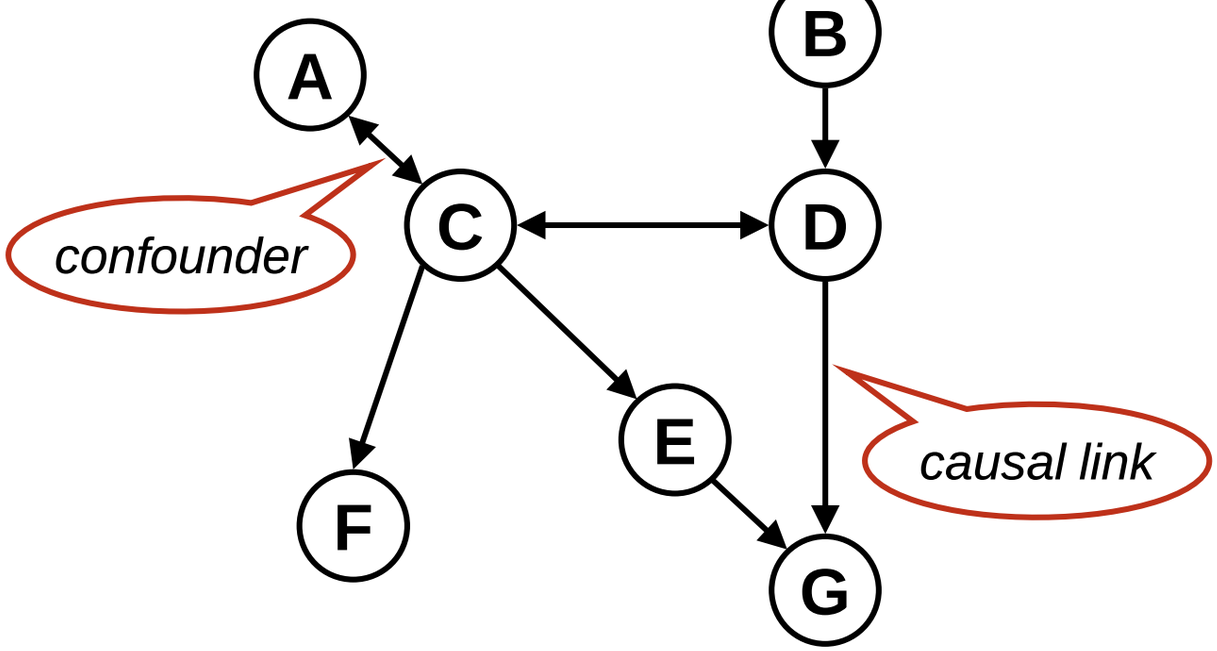
+ subsequent logical deduction on standard causal properties

- transitivity $\quad (X \Rightarrow Y) \wedge (Y \Rightarrow Z) \quad : \quad (X \Rightarrow Z)$
- acyclicity $\quad (X \Rightarrow Y) \quad : \quad (Y \not\Rightarrow X)$

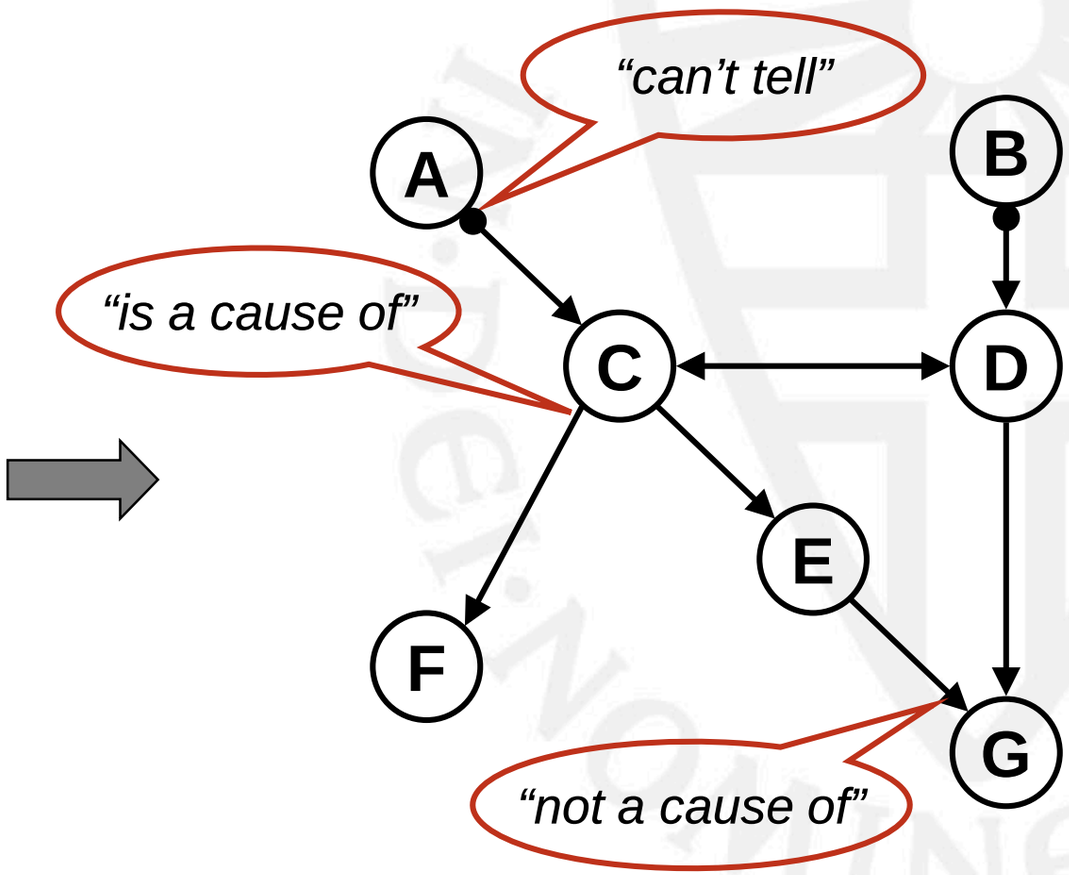
Theorem: "LoCI rules are sound and complete for causal discovery in the presence of latent confounders and selection bias." [Claassen & Heskes, 2011]

Oracle

- Given an “infinite” amount of data from the unknown causal structure on the left
- Constraint-based causal discovery algorithms return the inferred (set of) causal model(s) on the right



true underlying causal structure



inferred causal model

ADHD

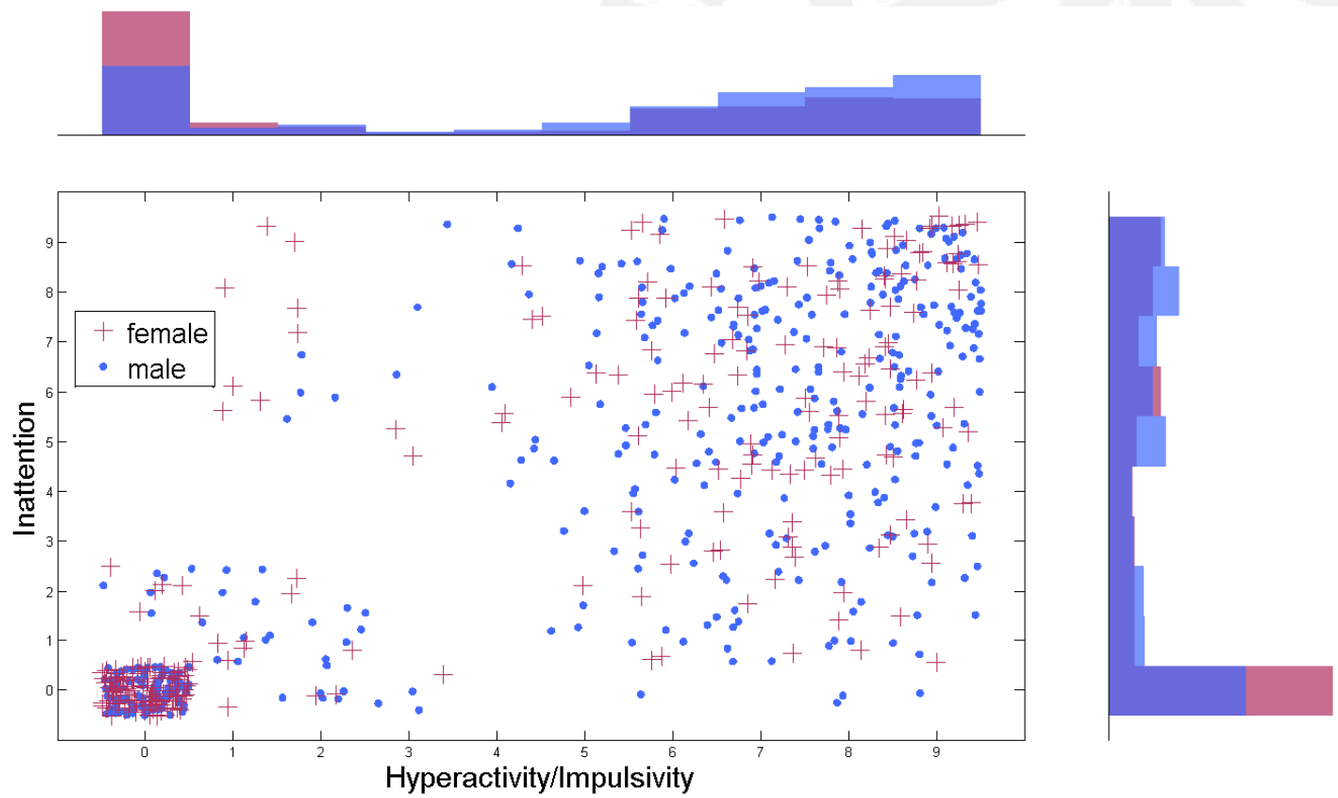
E. Sokolova et al., Statistical evidence suggests that inattention drives hyperactivity/impulsivity in Attention Deficit-Hyperactivity Disorder, PLOS ONE, 2016

- ADHD - Attention Deficit Hyperactivity Disorder
- Two types of symptoms:
  - Hyperactivity / Impulsivity (HI)
  - Inattention / concentration problems (In)
- M-H test indicates that

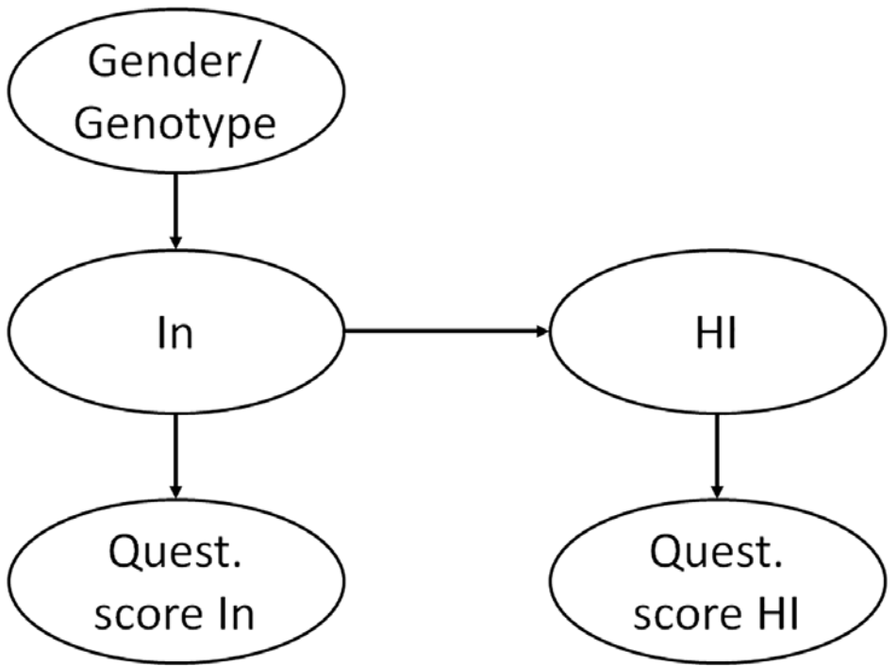
  Gender $\perp\!\!\!\perp$ HI | In

  suggesting the causal link

  In $\Rightarrow$ HI



Causal model for ADHD



[redacted]

Voorpagina
Net binnen
Algemeen
Eindexamens
Achtergronden
Economie
Sport
Tech
Entertainment
Overig
Wetenschap
Opmerkelijk
Lifestyle
Gezondheid
Eten en Drinken
Reizen
Dieren
NUcheckt
Auto
Podcasts

NU.nl > Overig > Lifestyle > Gezondheid

Foto: ANP

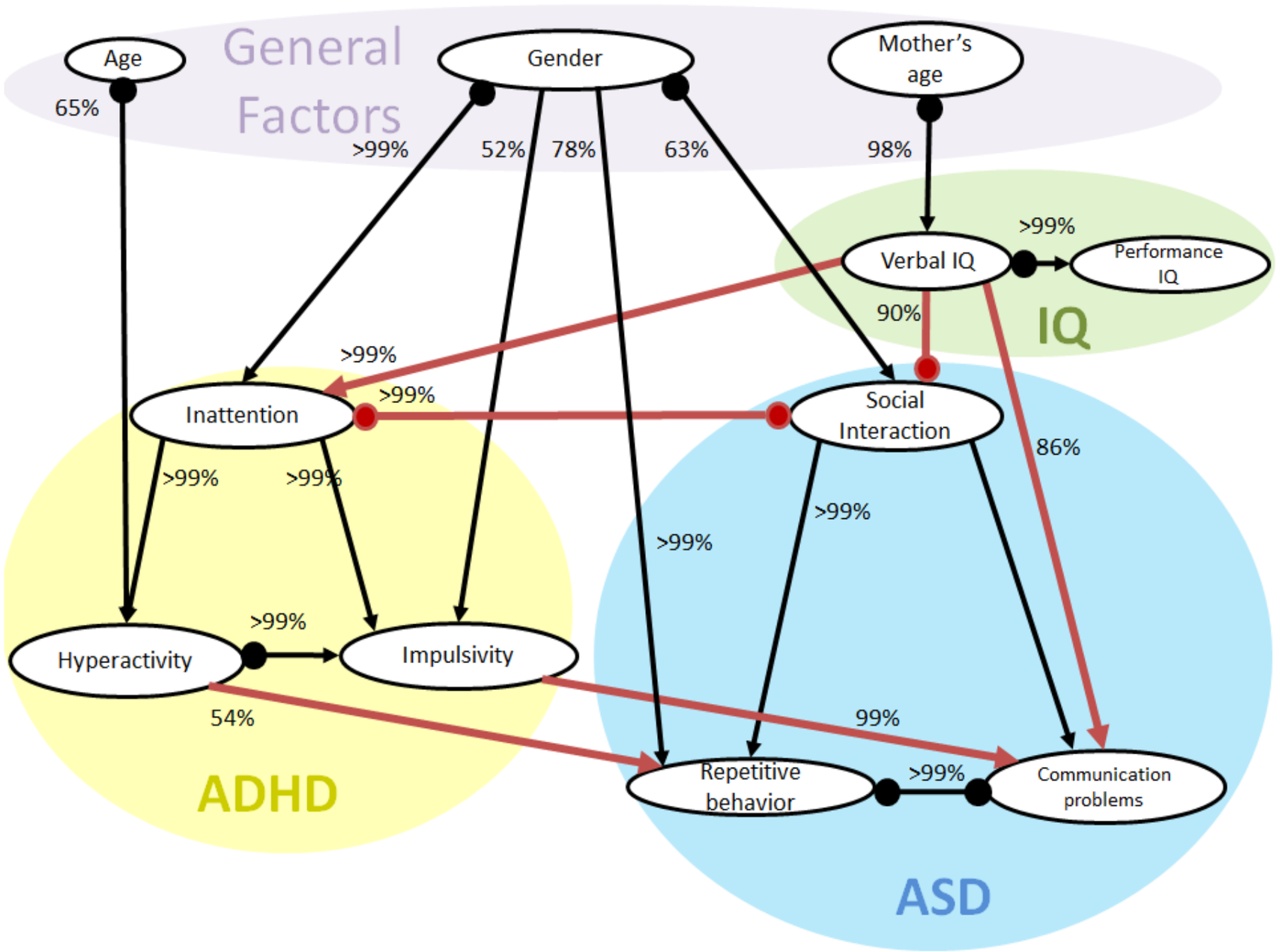
'Focus van ADHD moet liggen op aandachtstekort'

Gepubliceerd: 29 november 2016 13:55

Uit onderzoek van het Radboudumc is gebleken dat de aanpak van ADHD zich beter op het aandachtstekort dan op de hyperactiviteit kan richten.

ADHD is een combinatie van je moeilijk kunnen focussen en druk en impulsief

Comorbidity between autism and ADHD



E. Sokolova et al., Journal of Autism and Developmental Disorders, 2017

Time-series data

- ASML case
- 15000 system variables, spanning time scales from nanoseconds to days or 14 orders of magnitude
- Transfer entropy to identify causal effects between pairs of variables (nodes)
- Graph analysis to find the most influential nodes
- Essential for root cause analysis
- But beware of the rooster that crows before the sun rises…

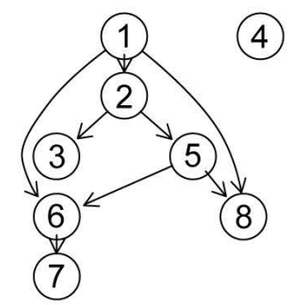
From philosophy to math to engineering

Philosophy **Math** **Engineering**

"How can we infer causal relations from observations?"

$$p(\mathrm{X} = \mathrm{x}) = \prod_i p(X_i = x_i | \mathrm{PA_i} = \mathrm{pa_i})$$

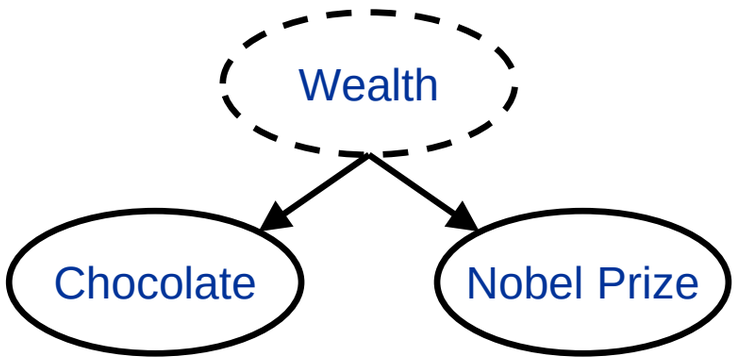
```
R> require("pcalg")
R> data("gmG")
R> suffStat <- list(C = cor(gmG$x), n = nrow(gmG$x))
R> pc.fit <- skeleton(suffStat, indepTest = gaussCItest,
                      p = ncol(gmG$x), alpha = 0.01)
R> par(mfrow = c(1,2))
R> plot(gmG$g, main = "")
R> plot(pc.fit, main = "")
```

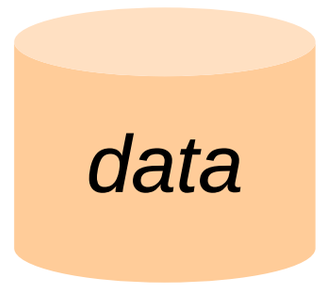


picture courtesy: Luke Muelhauser

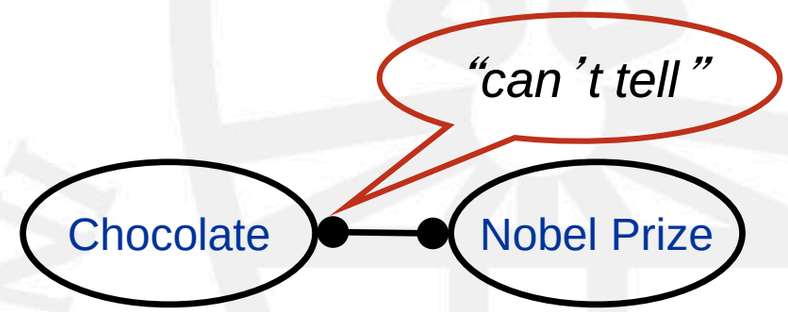
Take-home message

unknown underlying causal model





inferred causal model



Correlation does not imply causation

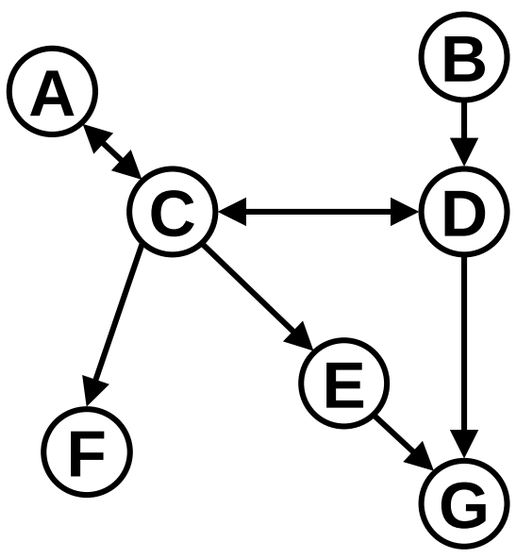
BORING!

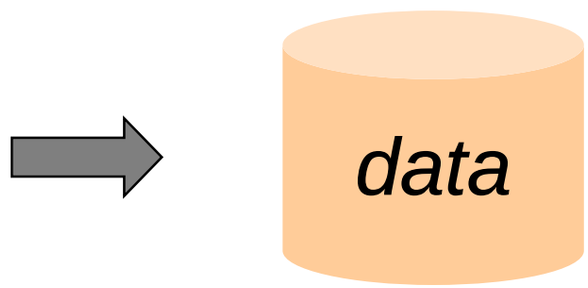
just a pair of variables

just a single symmetric number summarizing their dependence

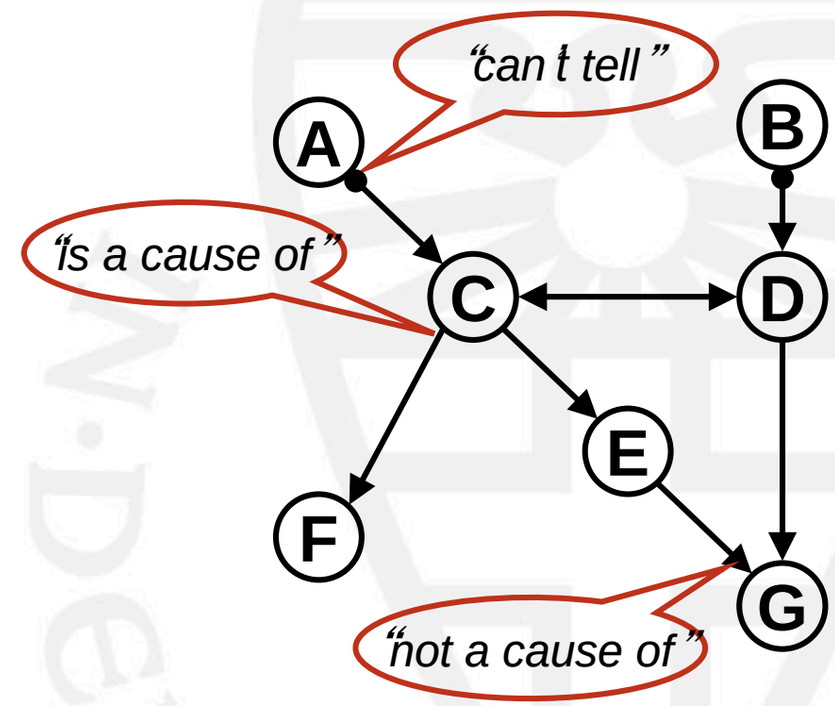
Take-home message

unknown underlying causal model





inferred causal model



Many thanks to:
Tom Claassen, Joris Mooij, Elena Sokolova, Perry Groot, Ridho Rahmadi, Gabriel Bucur, Ruifei Cui, Gido Schoenmacker, Errol Zalmijn et al.

challenging multi-disciplinary research
exciting opportunities

predictive → prescriptive maintenance?