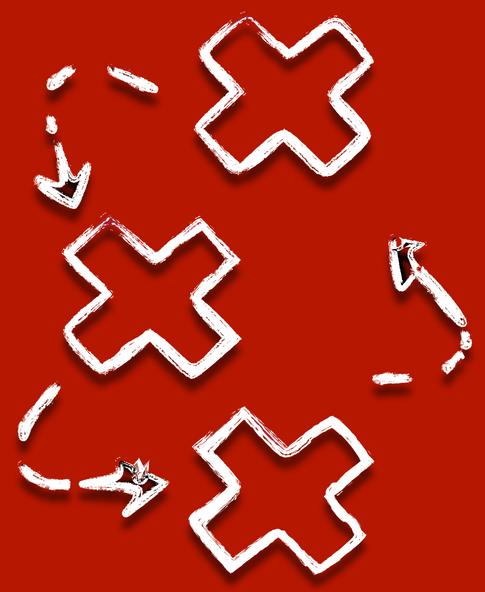


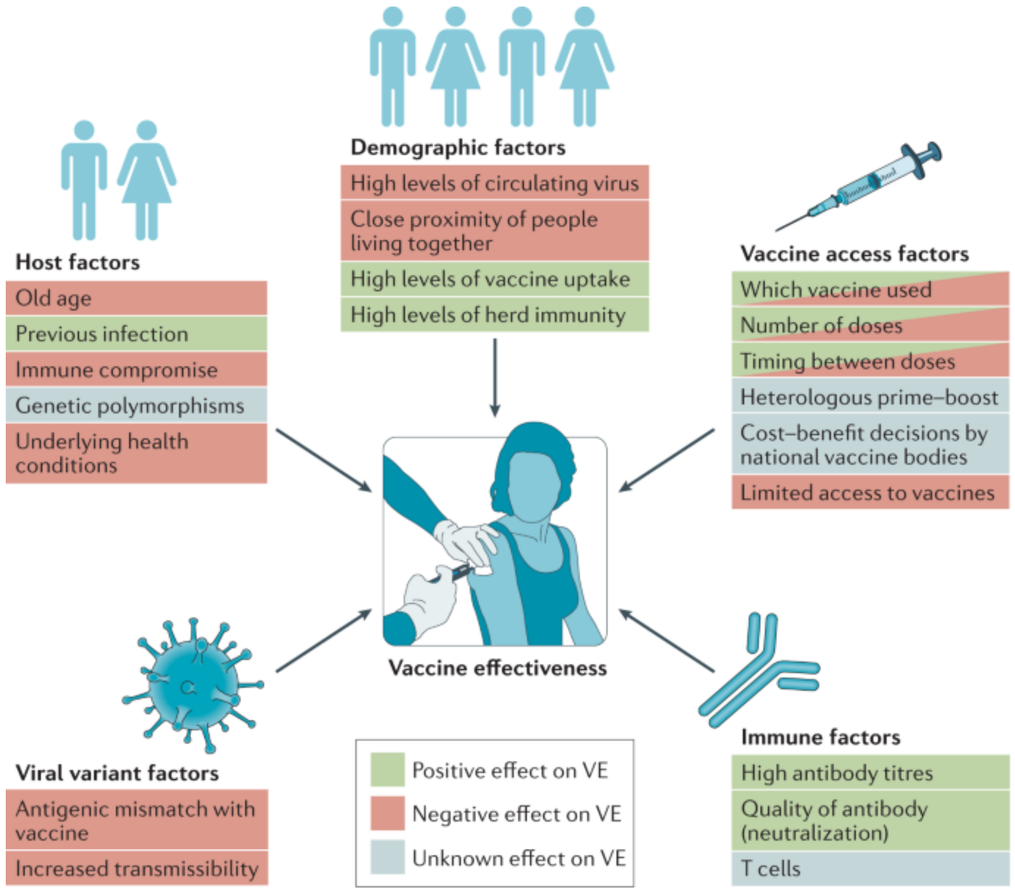
Causal Data Science

Lecture 1.2: Introduction

Lecturer: Sara Magliacane

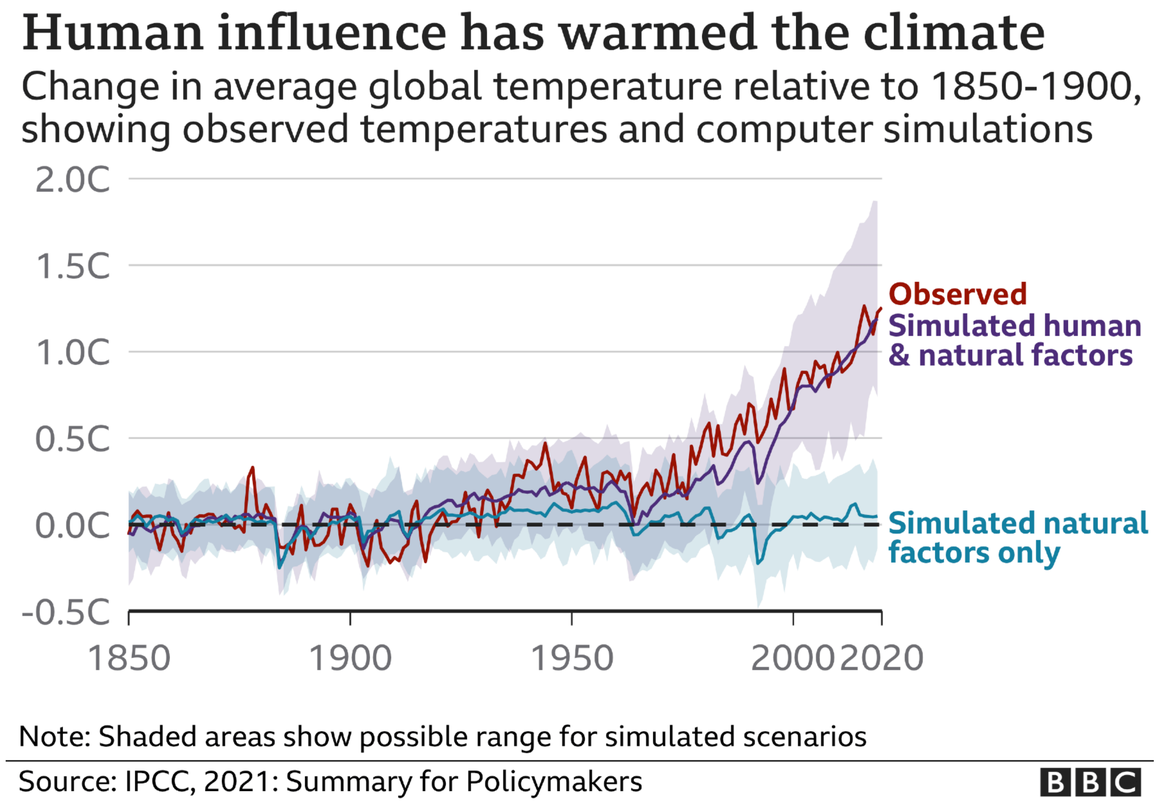
UvA - Spring 2024

Causal questions are ubiquitous: healthcare



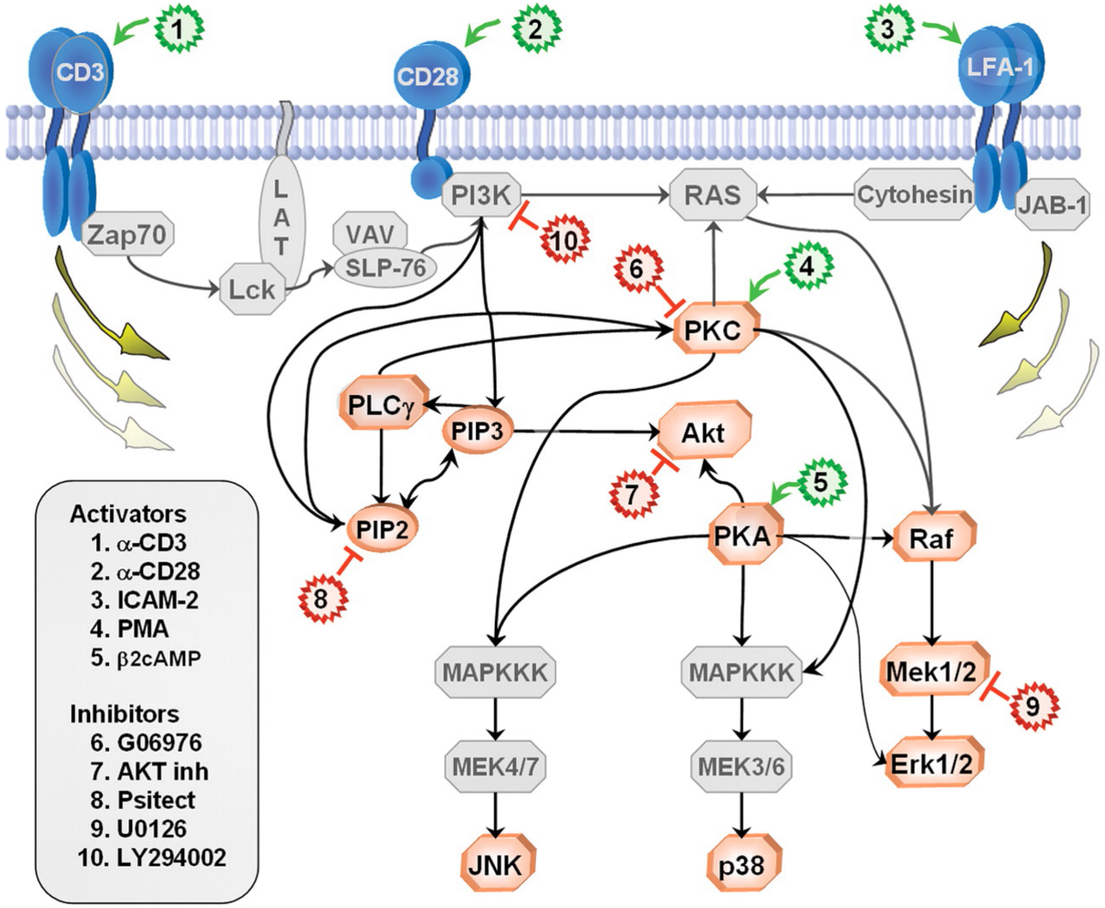
https://www.nature.com/articles/s41577-021-00592-1

Causal questions are ubiquitous: climate change



https://www.bbc.com/news/science-environment-58600723

Causal questions are ubiquitous: biology



https://www.science.org/doi/abs/10.1126/science.1105809

Correlation or causation? Does chocolate cause Nobel prizes?

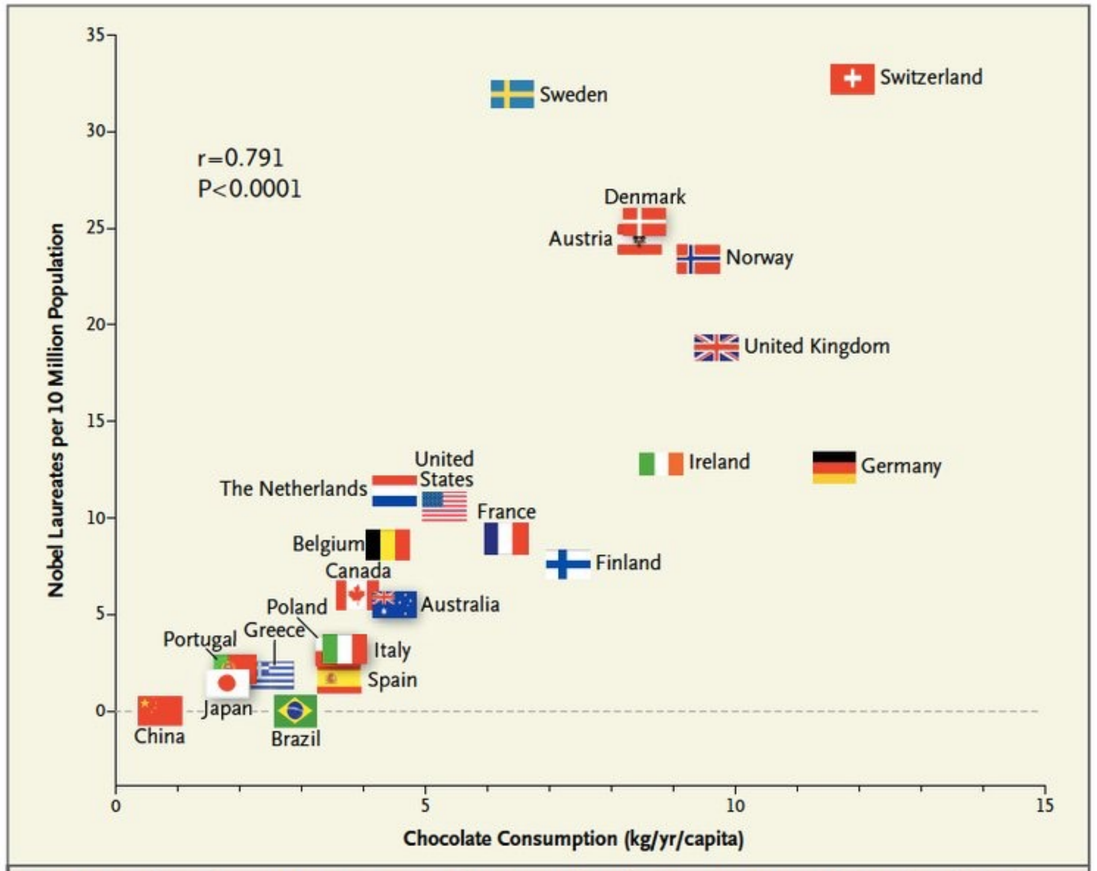


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Correlation or causation? Can we predict a new data point?

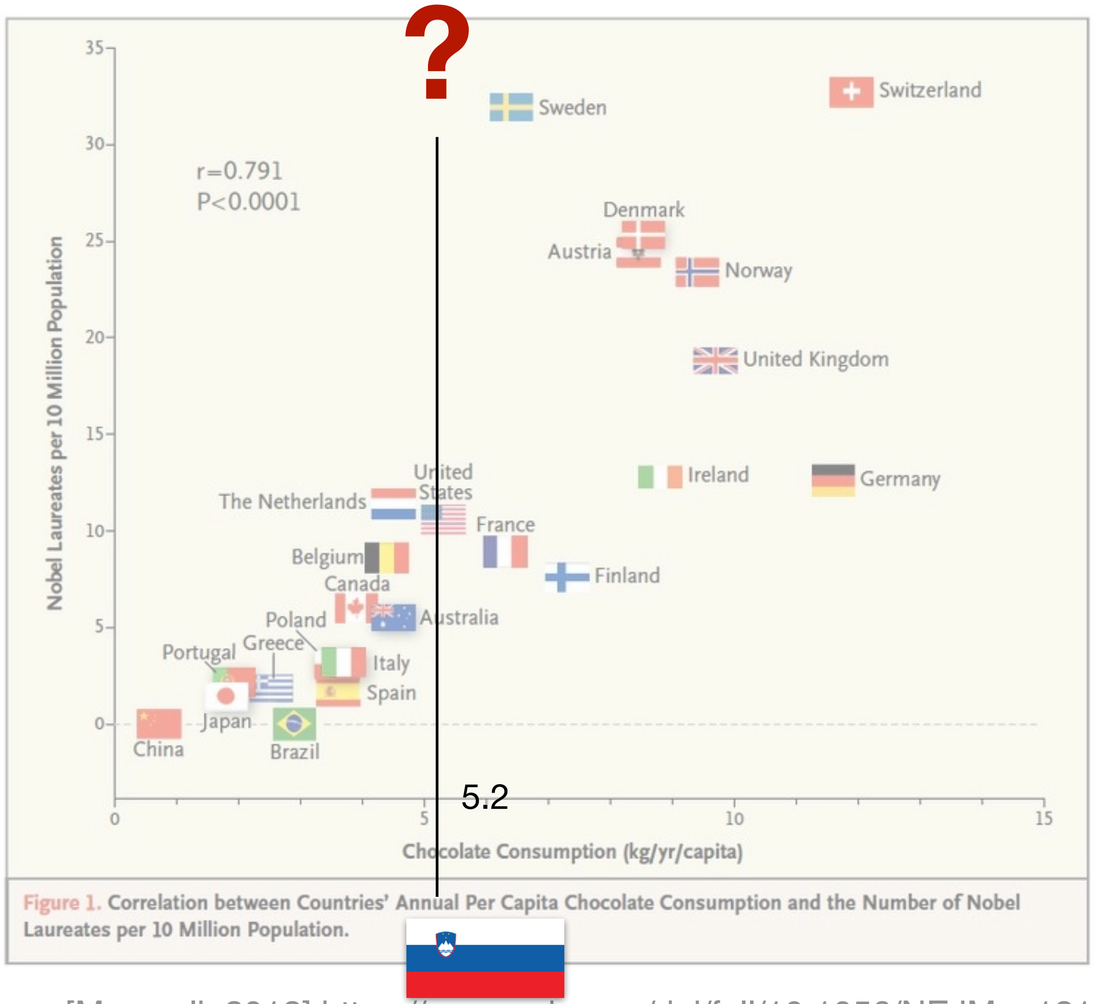


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Correlation or causation? Can we predict a new data point?

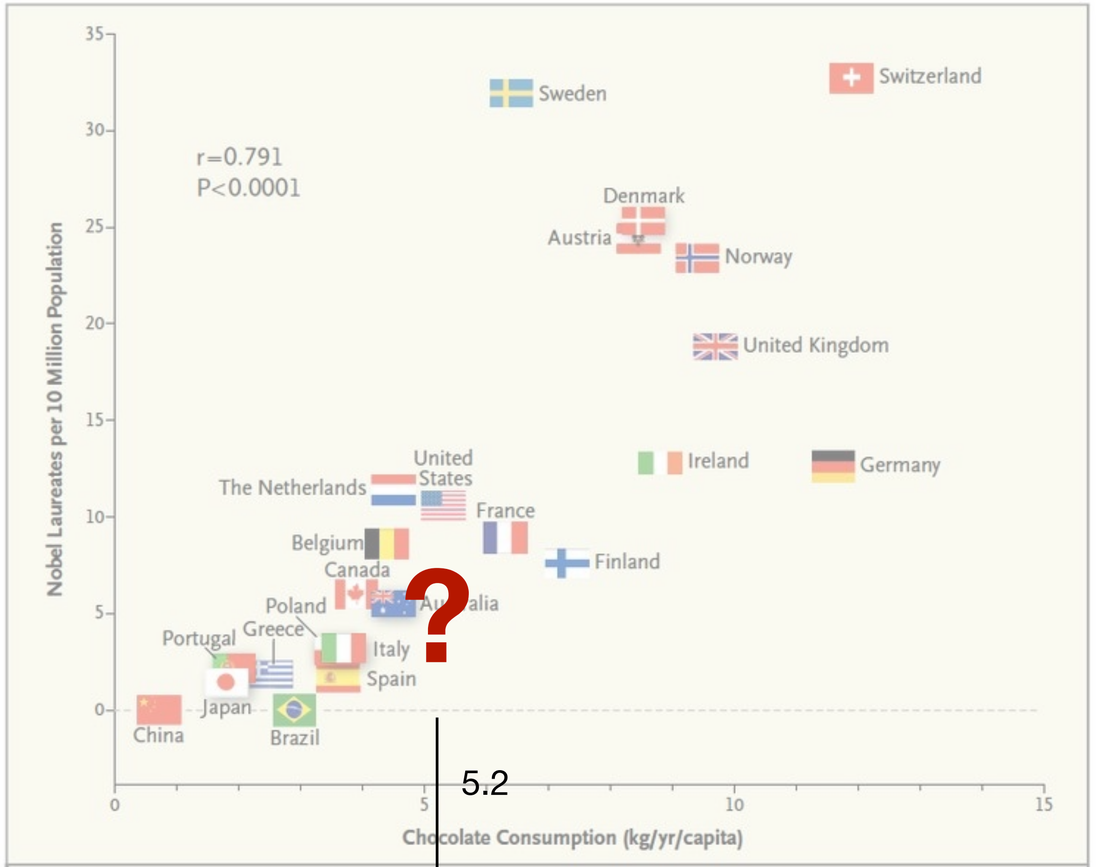


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Correlation or causation? Can we predict a new data point?

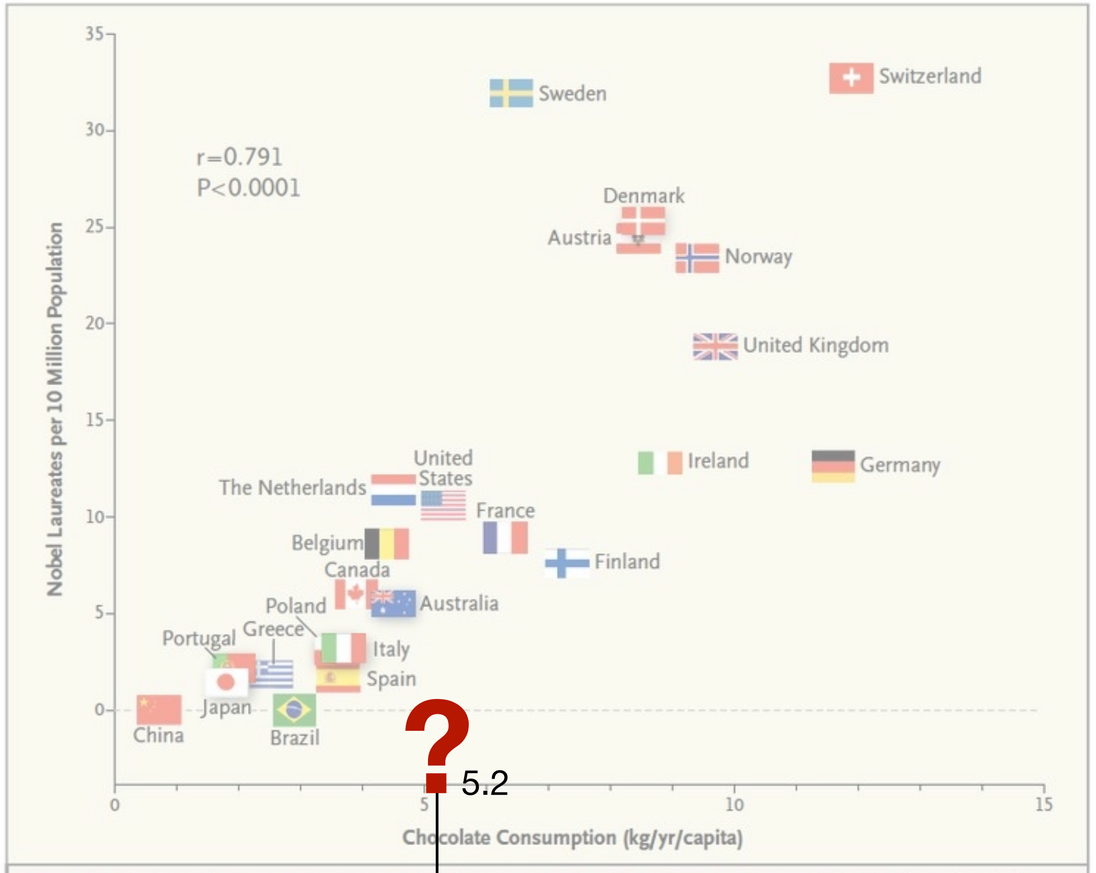
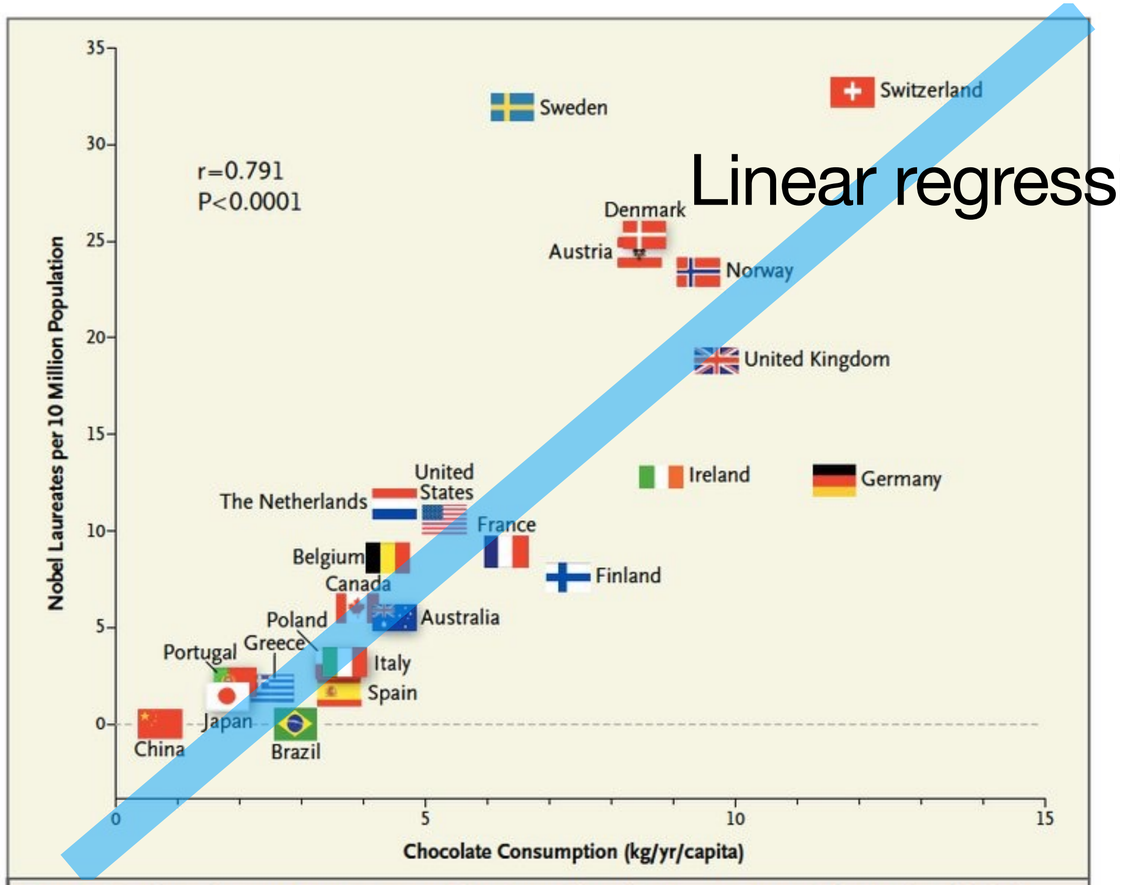


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Correlation or causation? Can we predict a new data point?



Linear regression

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

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Correlation or causation? Can we predict a new data point?

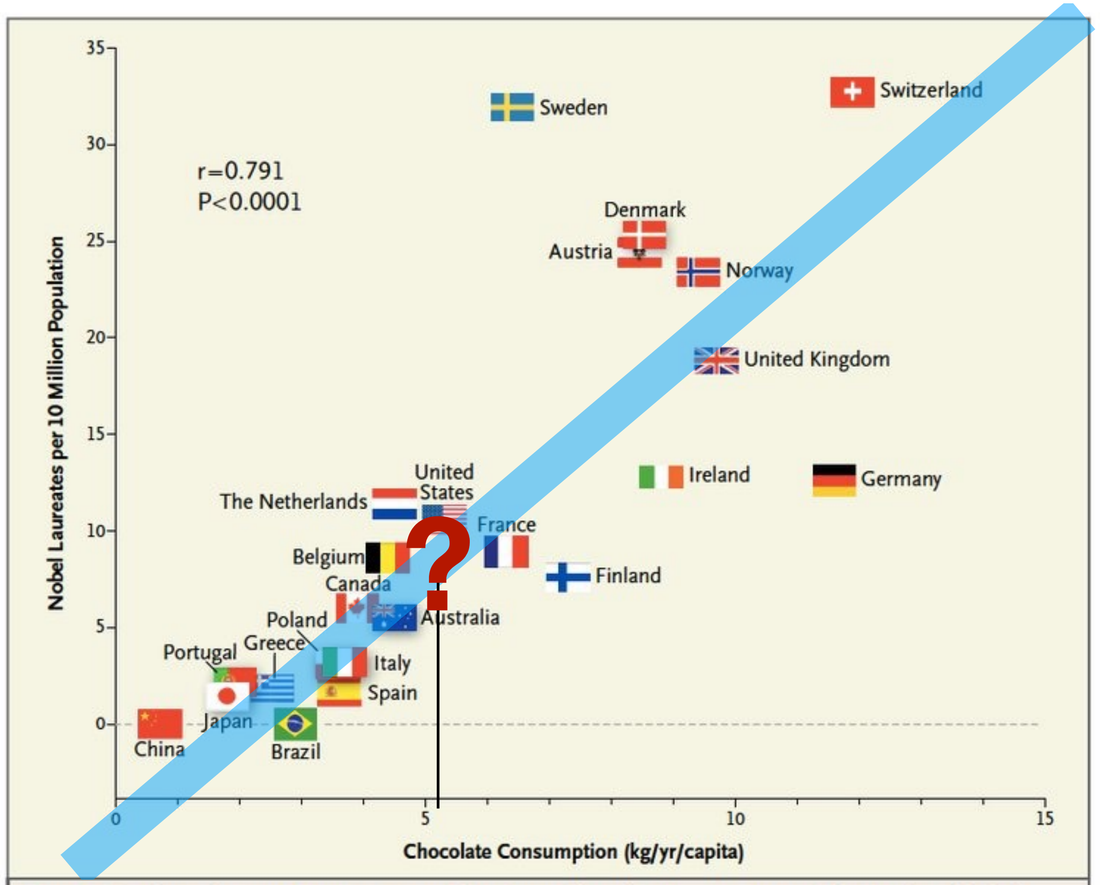
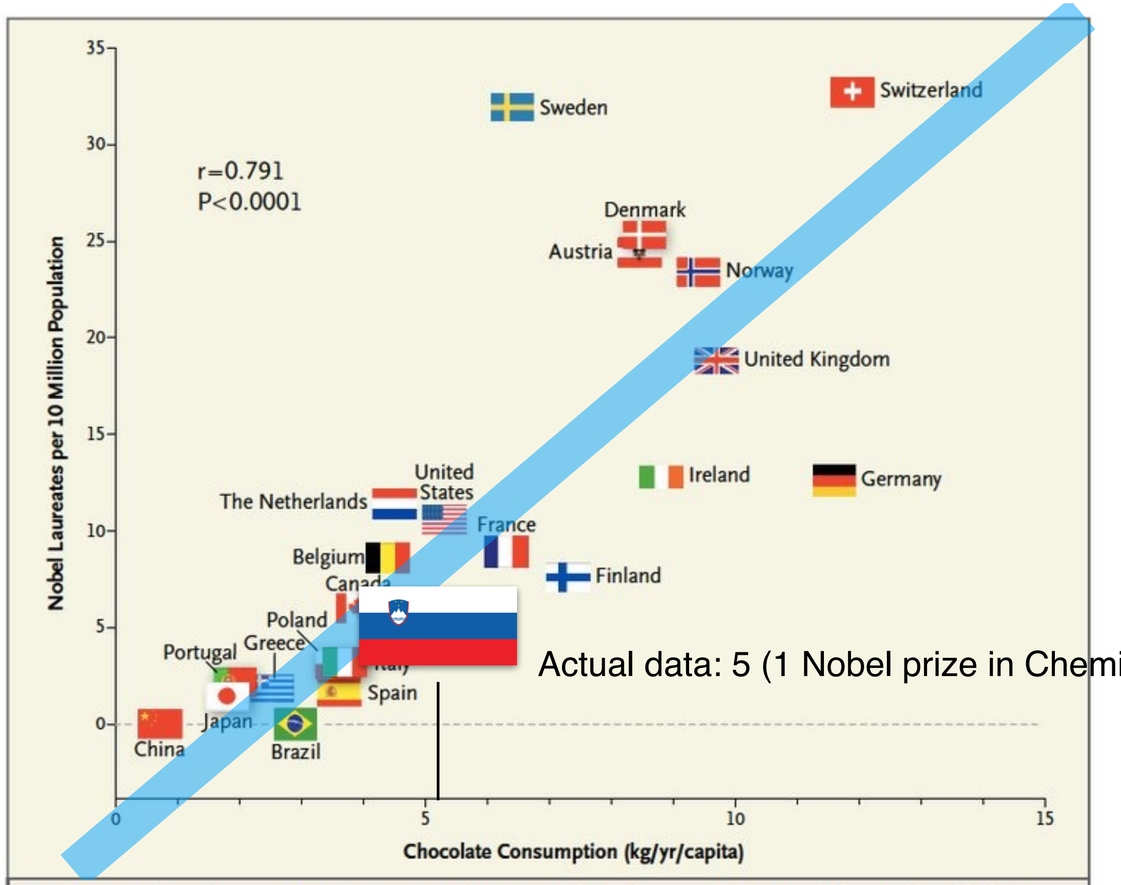


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Correlation or causation? Can we predict a new data point?



Actual data: 5 (1 Nobel prize in Chemistry in 1928, but Slovenia only has 2M citizens)

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

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Correlation or causation? Can we predict a new data point?

Prediction works!!!

Why shall we care about causality?

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

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Helping the Dutch government with decision making

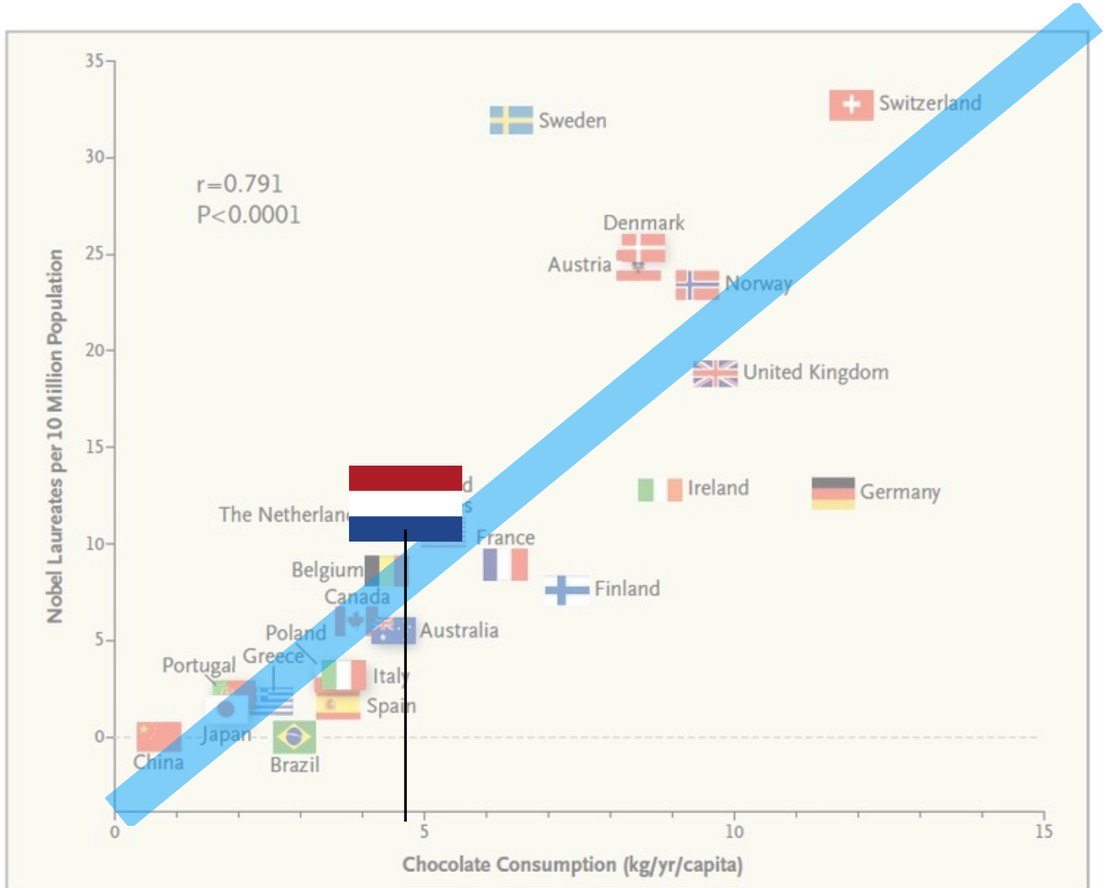


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Helping the Dutch government with decision making

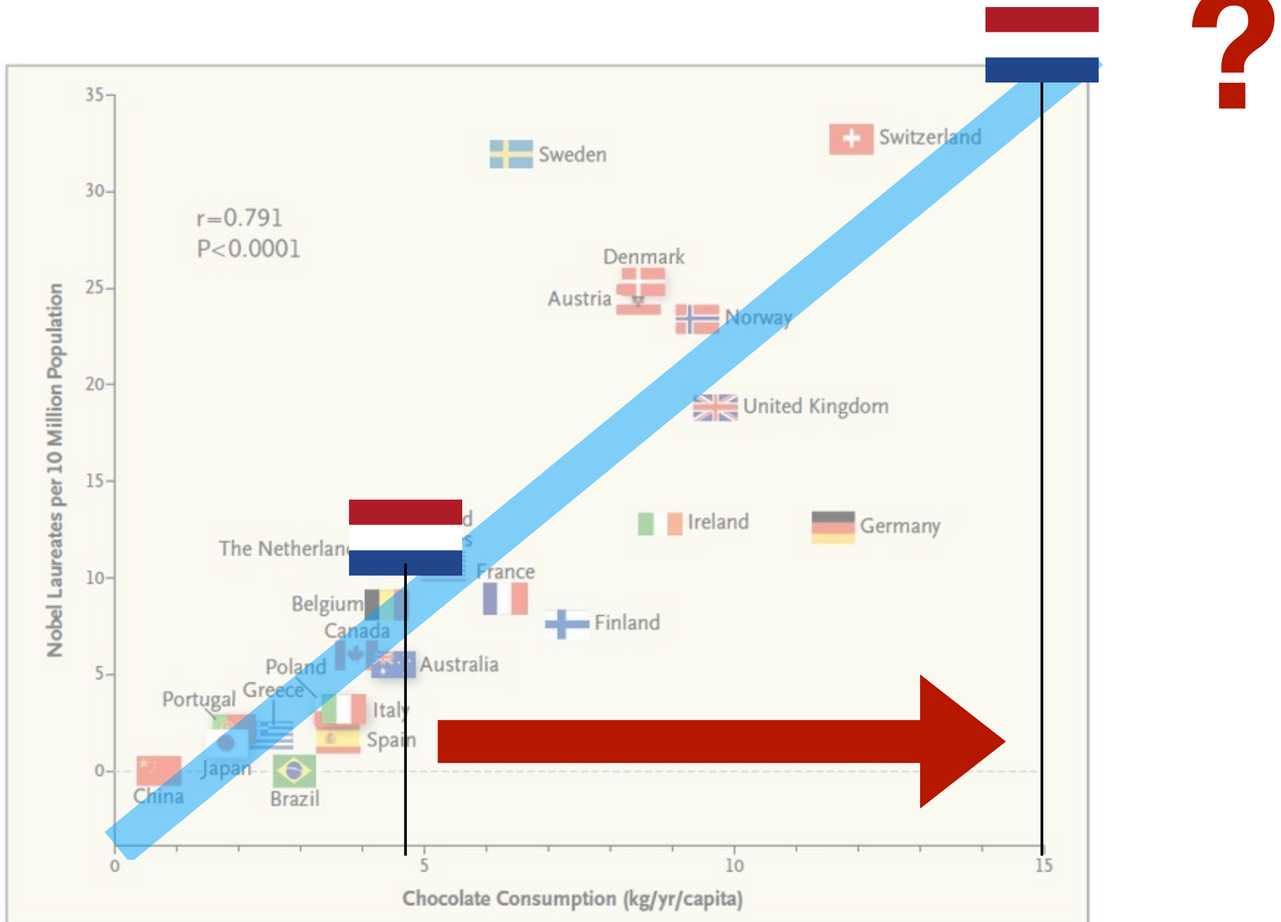


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Helping the Dutch government with decision making

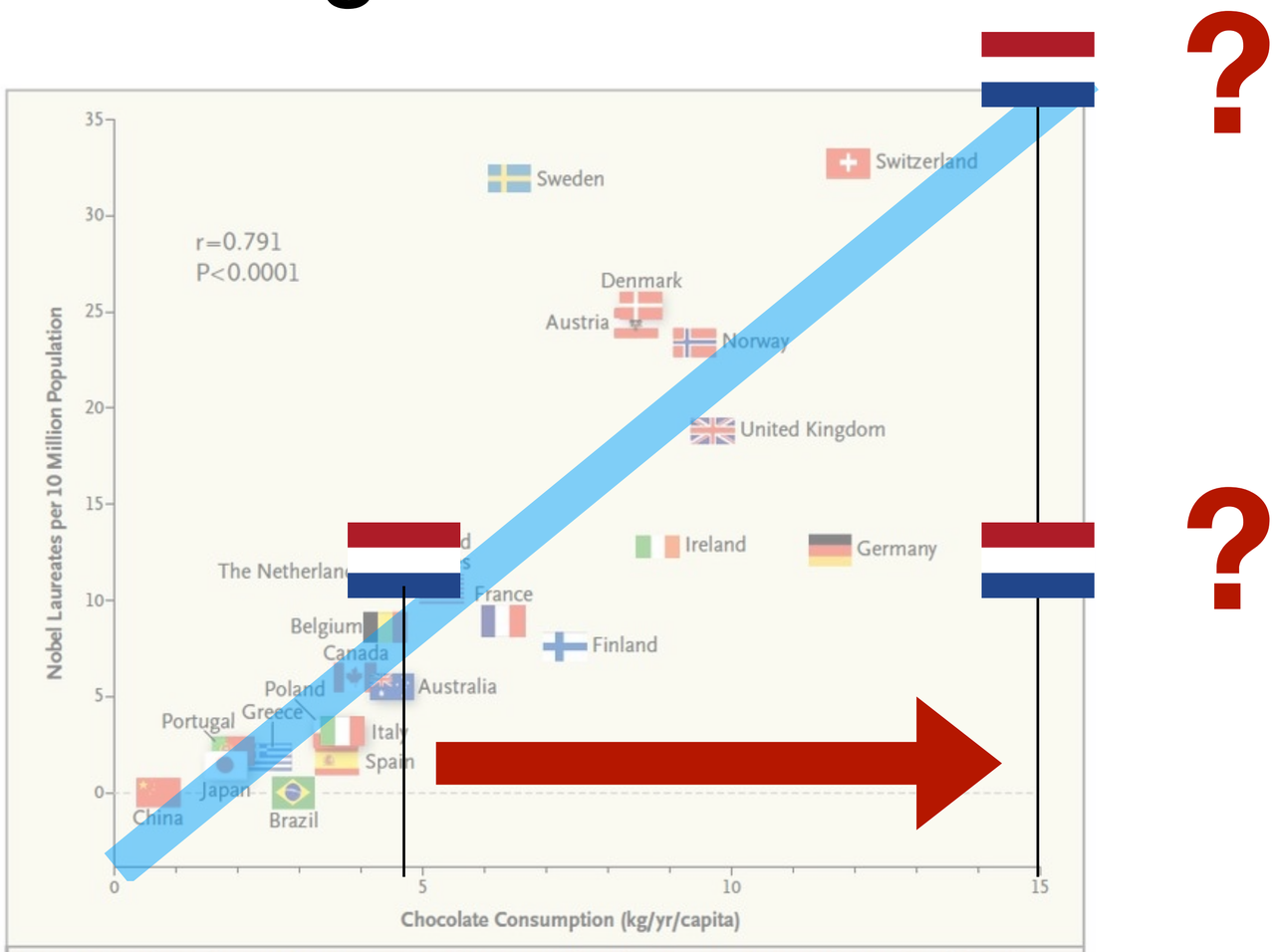


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Helping the Dutch government with decision making

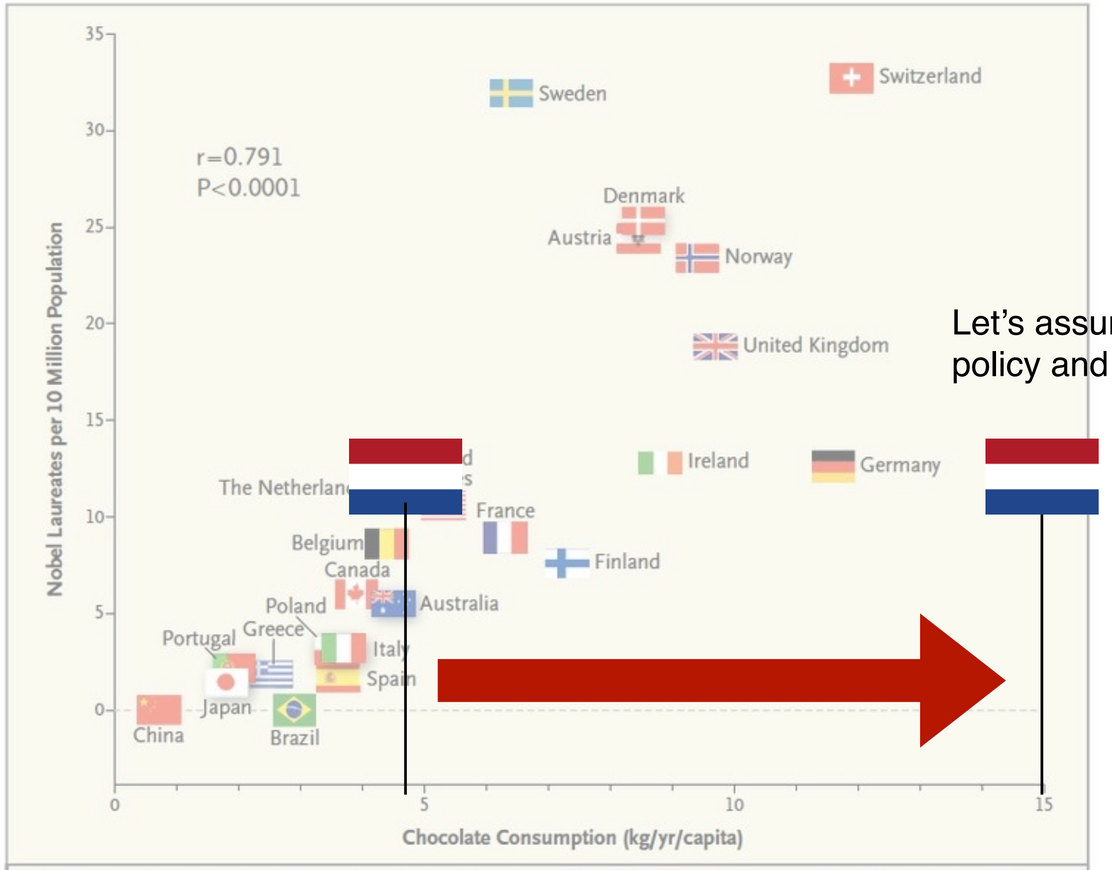


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Let's assume we implement the "Eat more chocolate" policy and this happens

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Correlation or causation? Direction? Alternative explanations?

Can you think of alternative explanations of the correlation?

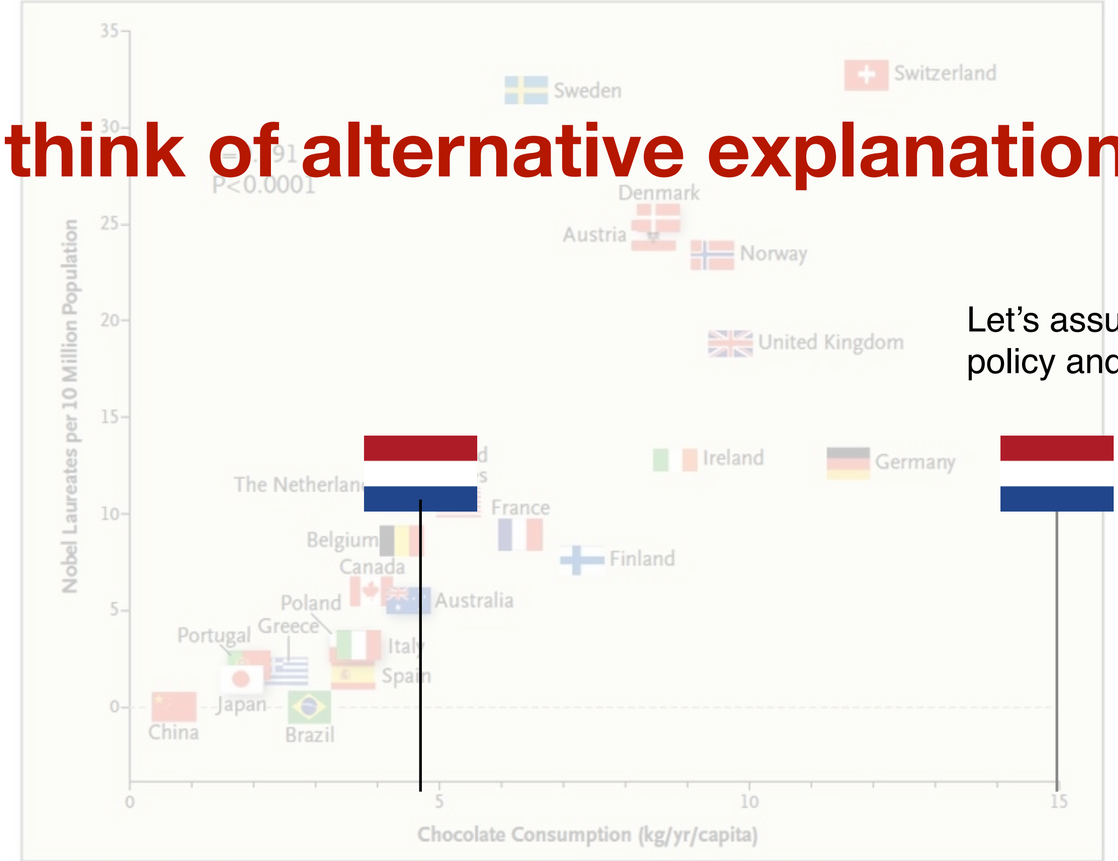


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Let's assume we implement the "Eat more chocolate" policy and this happens

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Canvas quiz: “[Optional] Does chocolate cause Nobel prizes?”

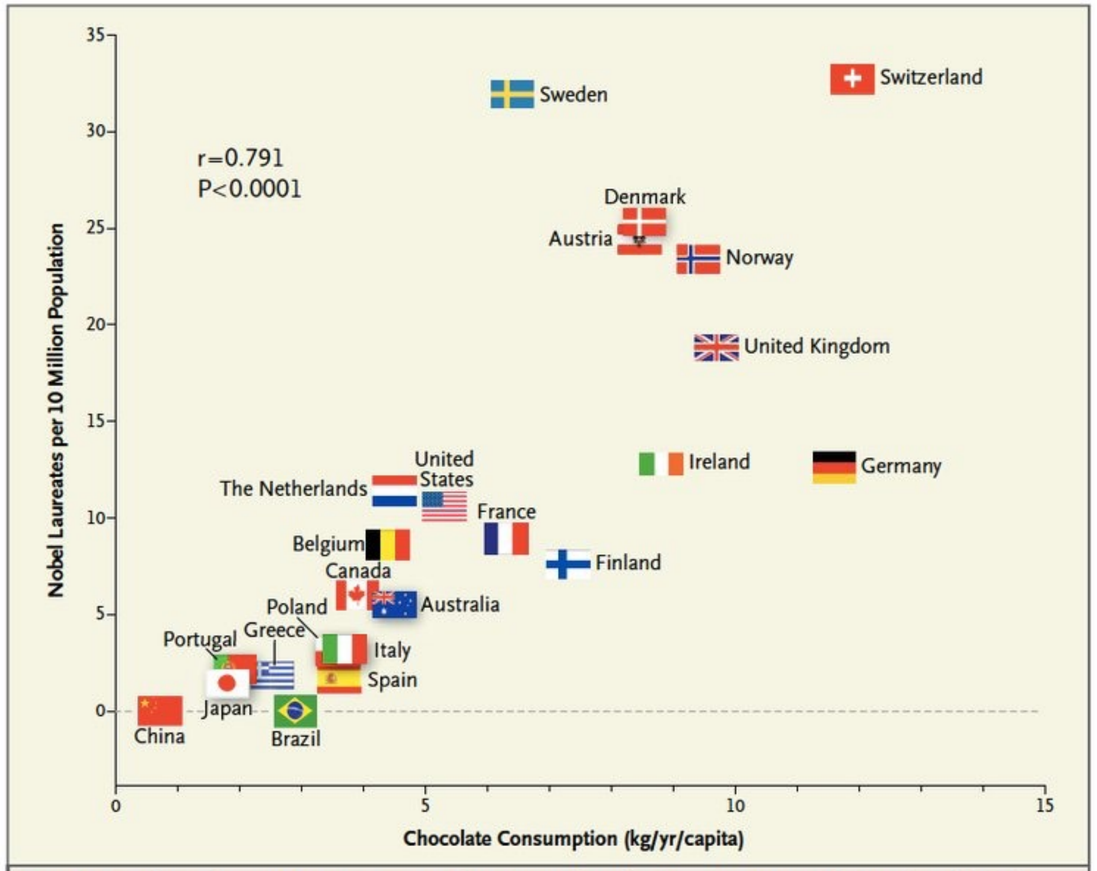
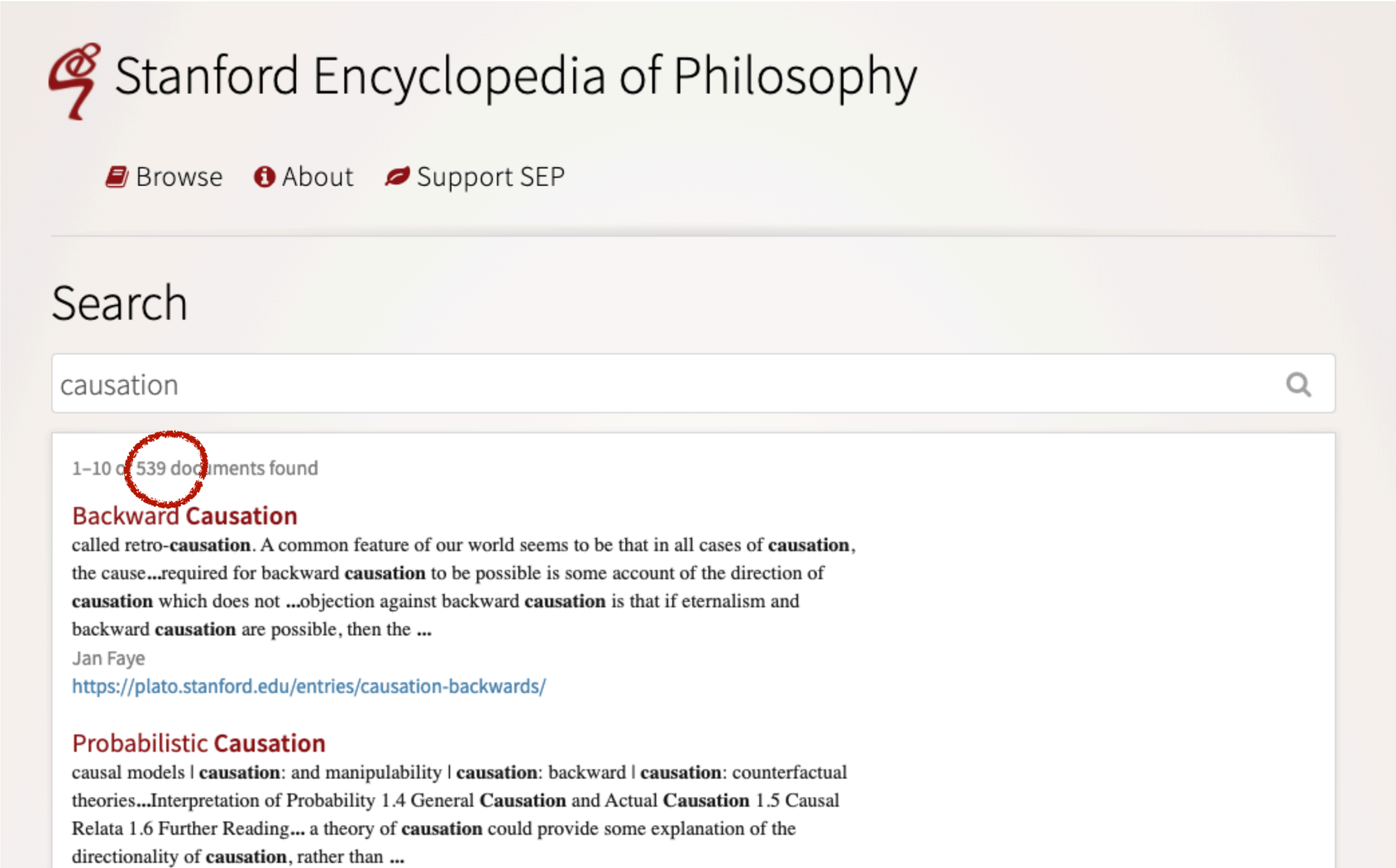


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

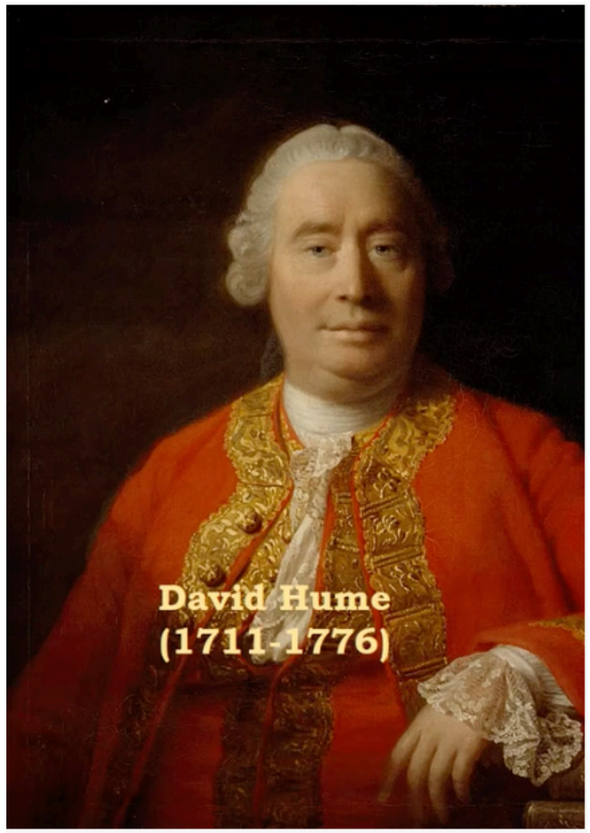
[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

How can we define causality, any ideas?

How can we define causality, any ideas?



Humean regularity theory



David Hume defines a cause:

> [a]n object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of priority and contiguity to those objects, that resemble the latter. (Hume 1739: book I, part III, section XIV [1978: 169])

- A cause precedes its effect
- A cause is spatiotemporally close to the effect
- All similar causes are followed by similar effects

(https://plato.stanford.edu/entries/causation-regularity)

Humean regularity theory

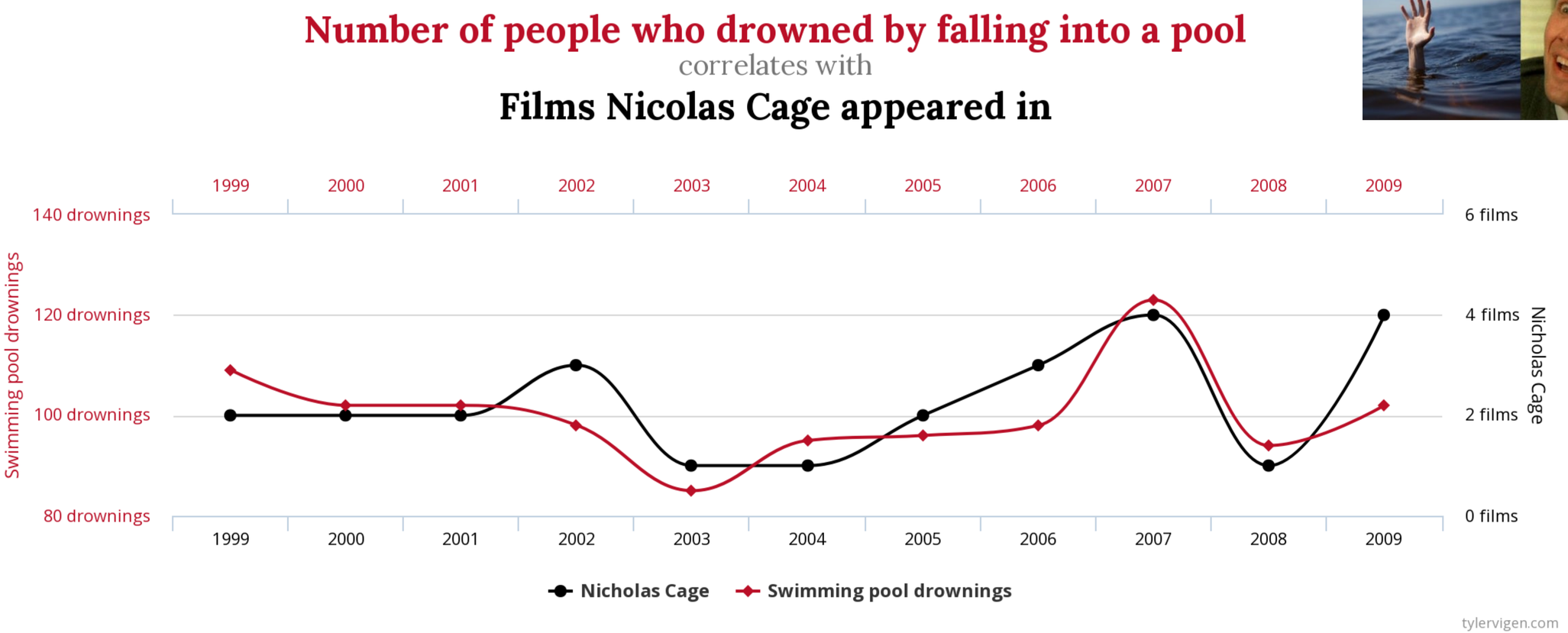
David Hume defines a cause:

> [a]n object precedent and contiguous to another, and w
> former are plac'd in like relations of priority and conti
> resemble the latter. (Hume 1739: book I, part III, secti

- A cause precedes its effect
- A cause is spatiotemporally close to the effect
- All similar causes are followed by similar effects

(https://plato.stanford.edu/entries/causation-regularity)

Other examples of regularity theories



http://tylervigen.com/spurious-correlations

Causal Hierarchy [Pearl 2009, 2018]

https://mobile.twitter.com/yudapearl

Causal Hierarchy [Pearl 2009, 2018]

Most ML

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?

Causal Hierarchy [Pearl 2009, 2018]

Most ML

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?

Causal Hierarchy [Pearl 2009, 2018]

Most ML

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Causal Hierarchy [Pearl 2009, 2018]

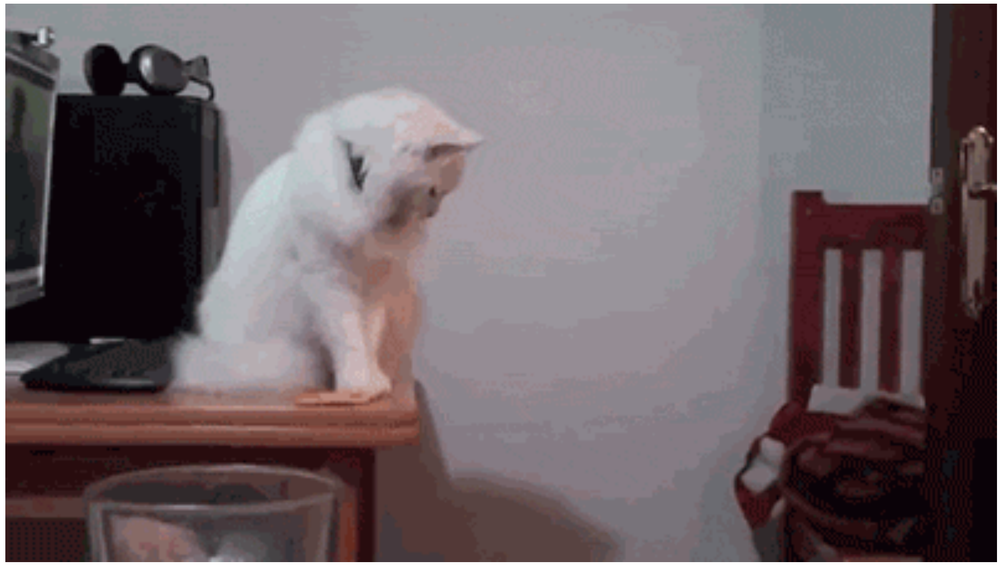
Most ML

Causality

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

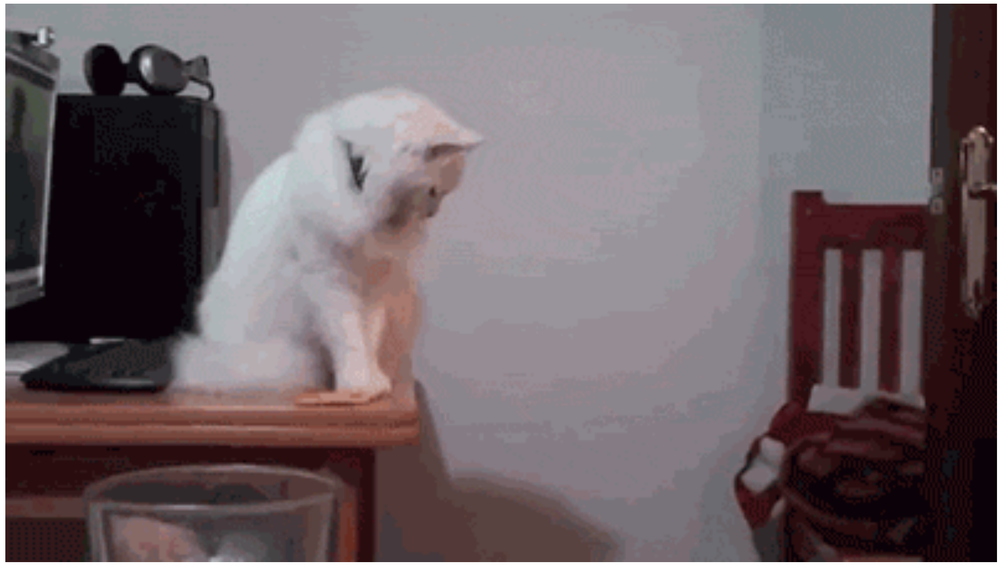
We know what causality is not, but what is it?

We know what causality is not, but what is it?



The classical approach to causality is based on experimentation.

We know what causality is not, but what is it?



The classical approach to causality is based on experimentation.

Can we use a definition of causality based on **manipulation** (interventions)?

A working definition of causality in machine learning

Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

A working definition of causality in machine learning

Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

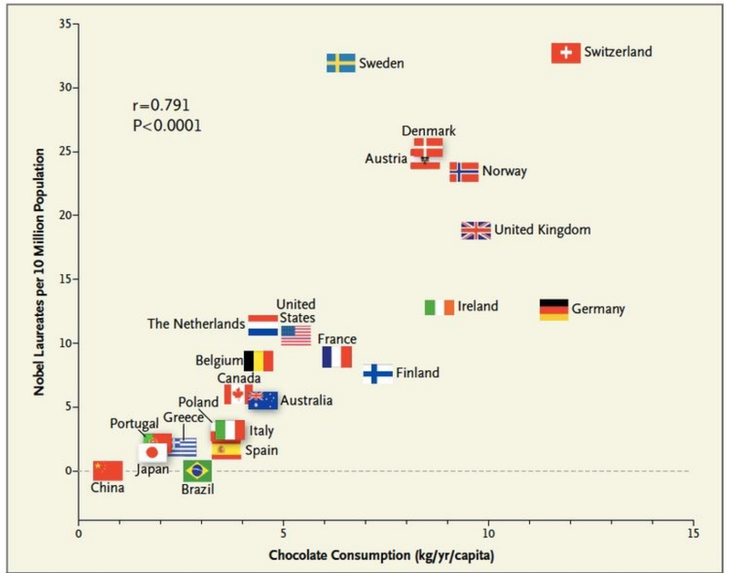


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

A working definition of causality in machine learning

Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

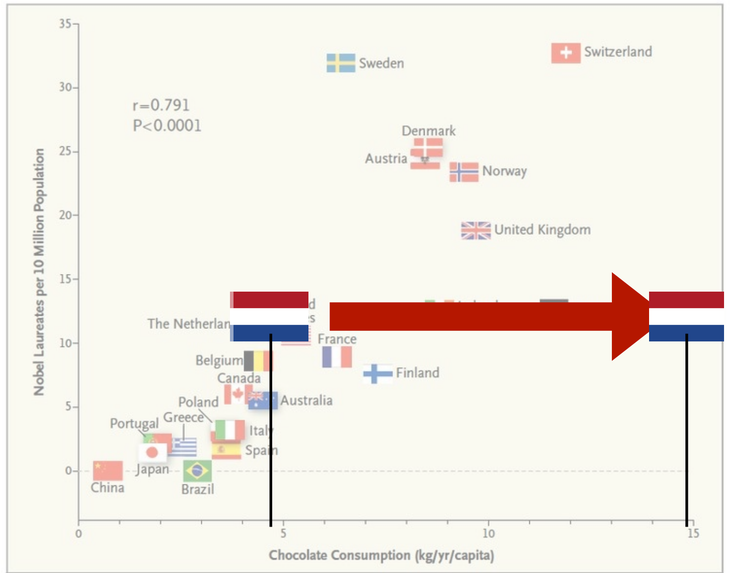


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

NL eats more chocolate => nothing changes

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

A working definition of causality in machine learning

Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

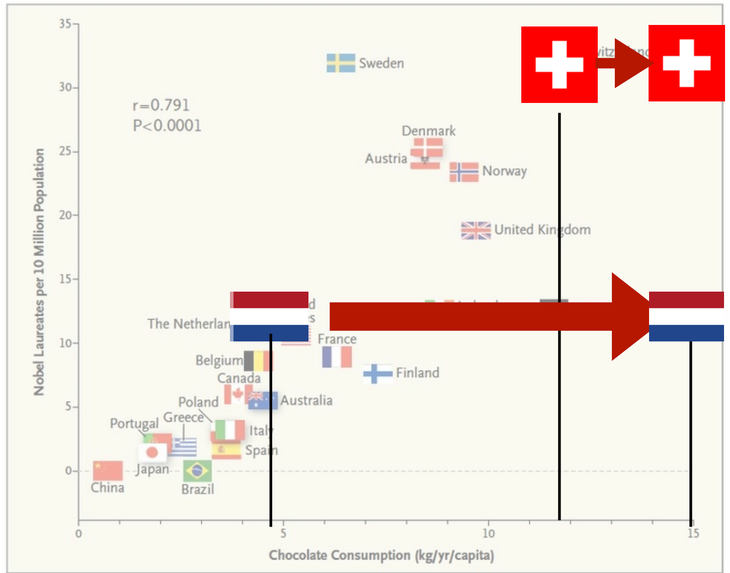


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

NL eats more chocolate => nothing changes

… and similarly for other countries (and other values)

Chocolate does not cause Nobel prizes

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

A working definition of causality in machine learning

Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

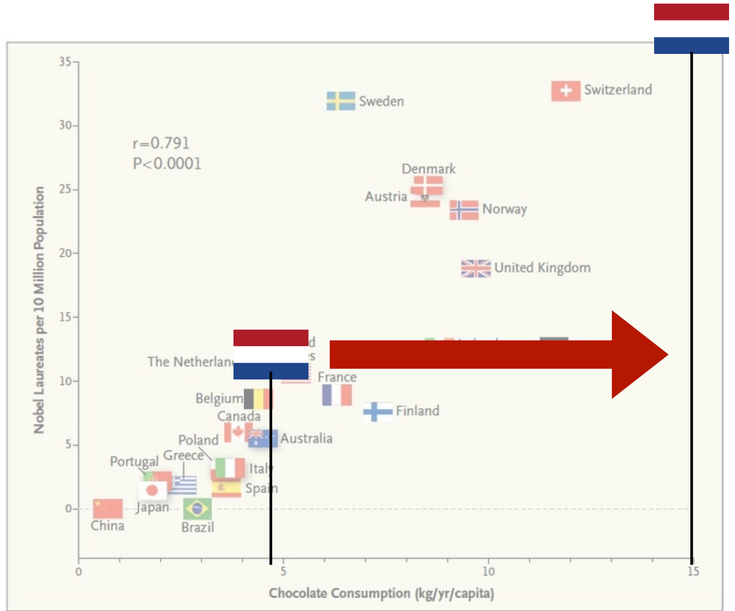


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

In a hypothetical universe:

NL eats more chocolate => more Nobel prizes

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

A working definition of causality in machine learning

Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

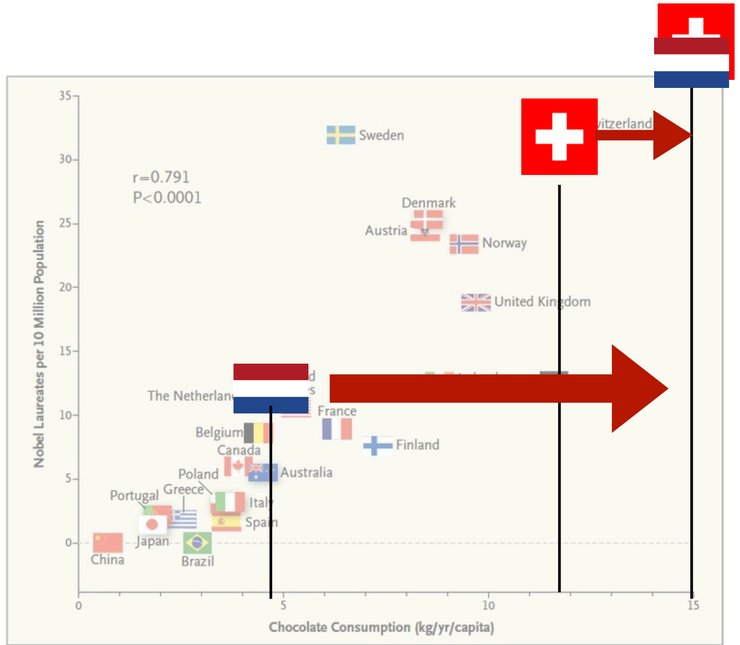


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

In a hypothetical universe:

NL eats more chocolate => more Nobel prizes

CH eats more chocolate => more Nobel prizes

… and similarly for (some) other countries

Chocolate causes Nobel prizes

Based on experimental data

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

A working definition of causality in machine learning

Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

Intervention

A working definition of causality in machine learning

Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

Intervention

Challenge: estimate the causal effect of an intervention, when we do not have (all possible) interventional data **(e.g. observational data)**

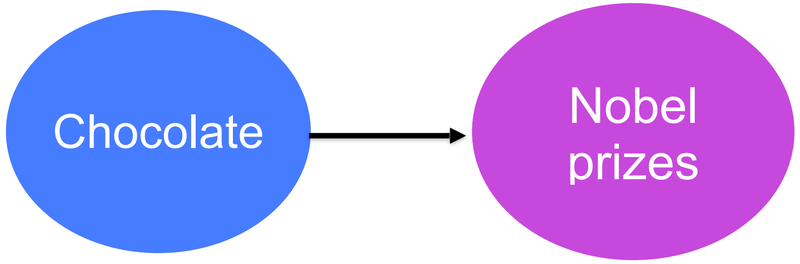
A working definition of causality in machine learning

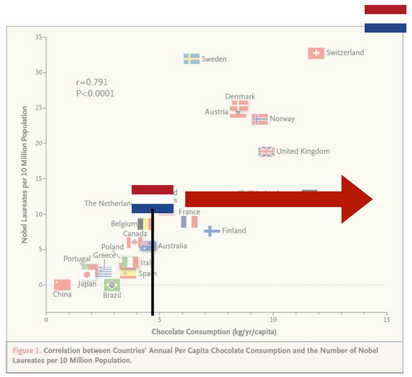
Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

Intervention

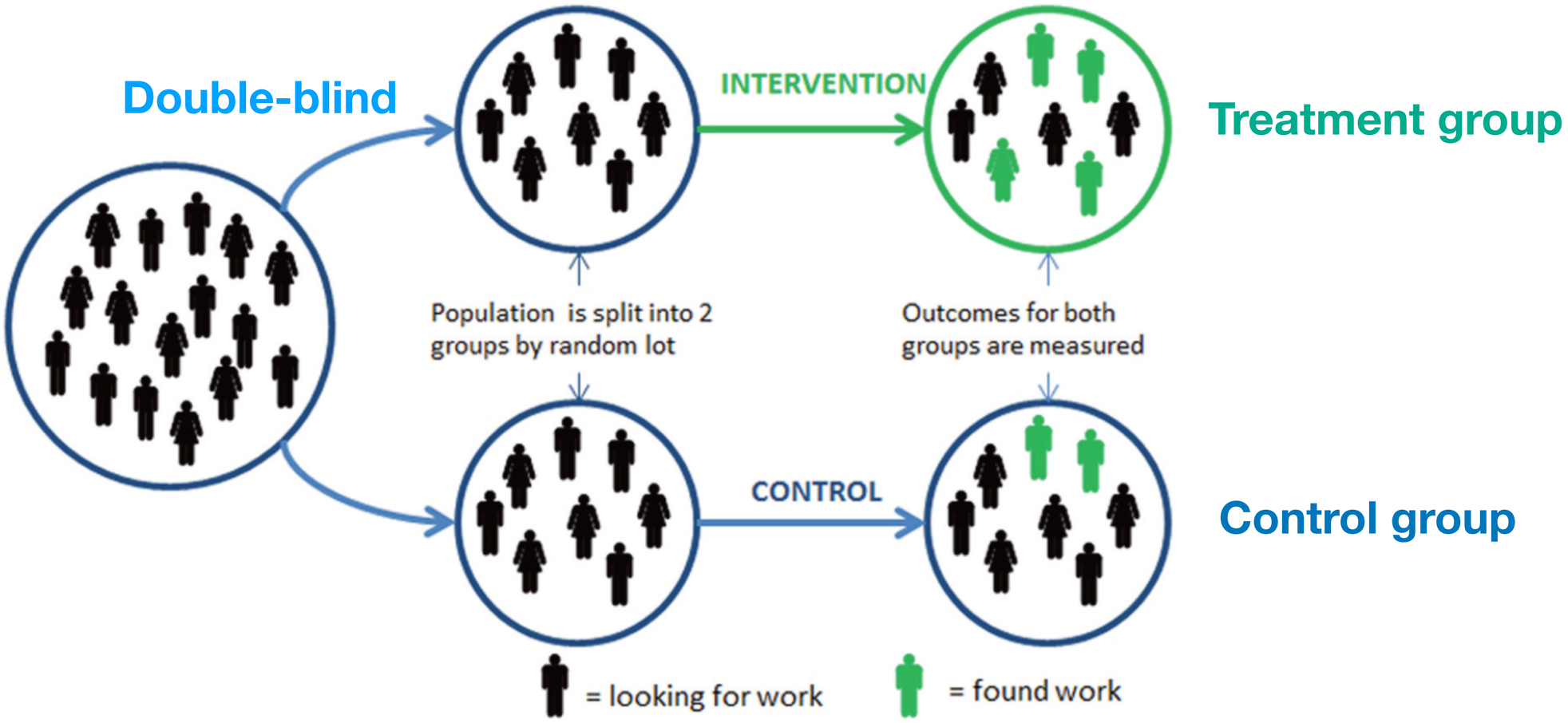
Challenge: estimate the causal effect of an intervention, when we do not have (all possible) interventional data **(e.g. observational data)**

Representation: We can represent causal relations in **causal graphs:** nodes are random variables, edges causal relations





Gold standard for experiments: Randomized Controlled Trials



https://www.researchgate.net/publication/256031307_Test_Learn_Adapt_Developing_Public_Policy_with_Randomised_Controlled_Trials

Randomised controlled trials (RCT): spot the difference?

Figure A:

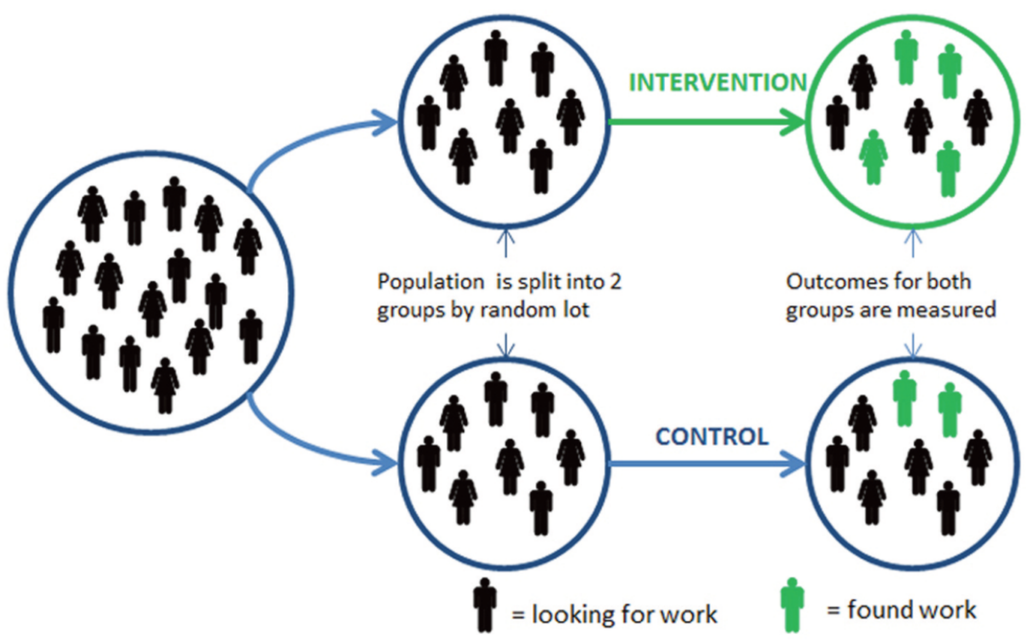
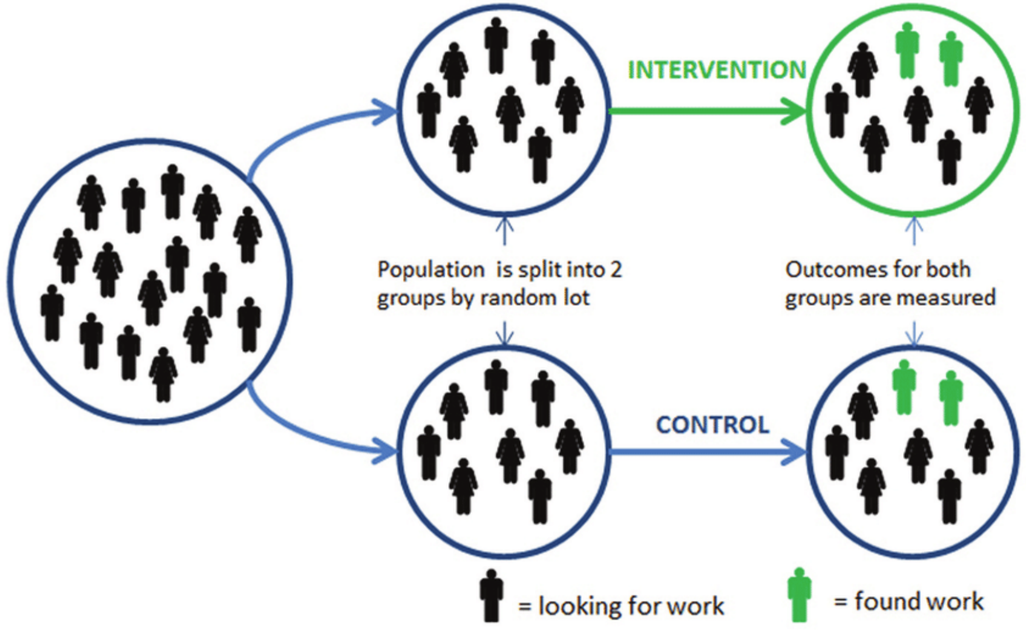


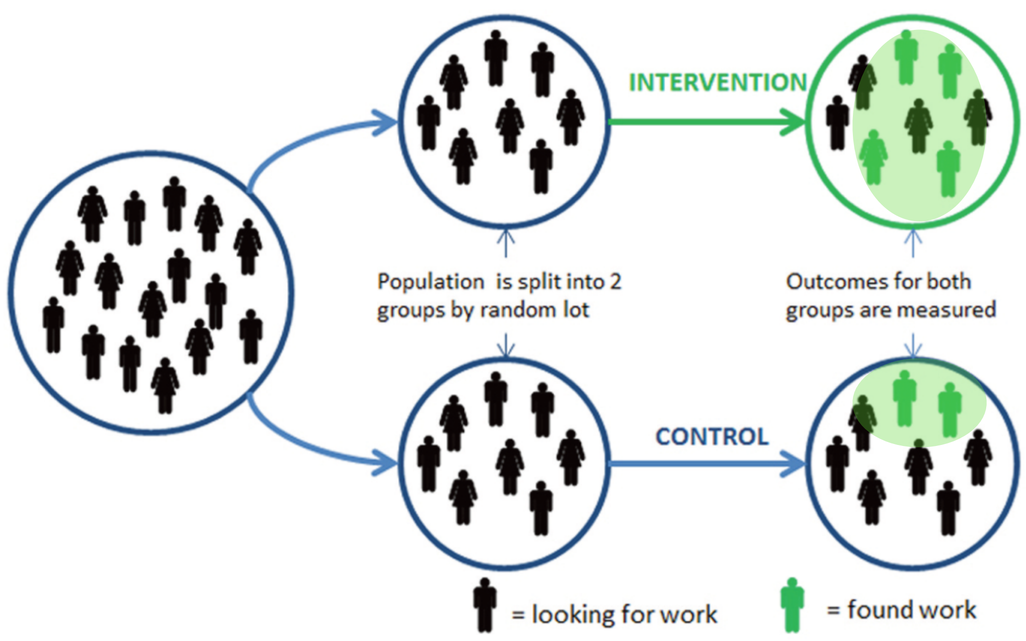
Figure B:



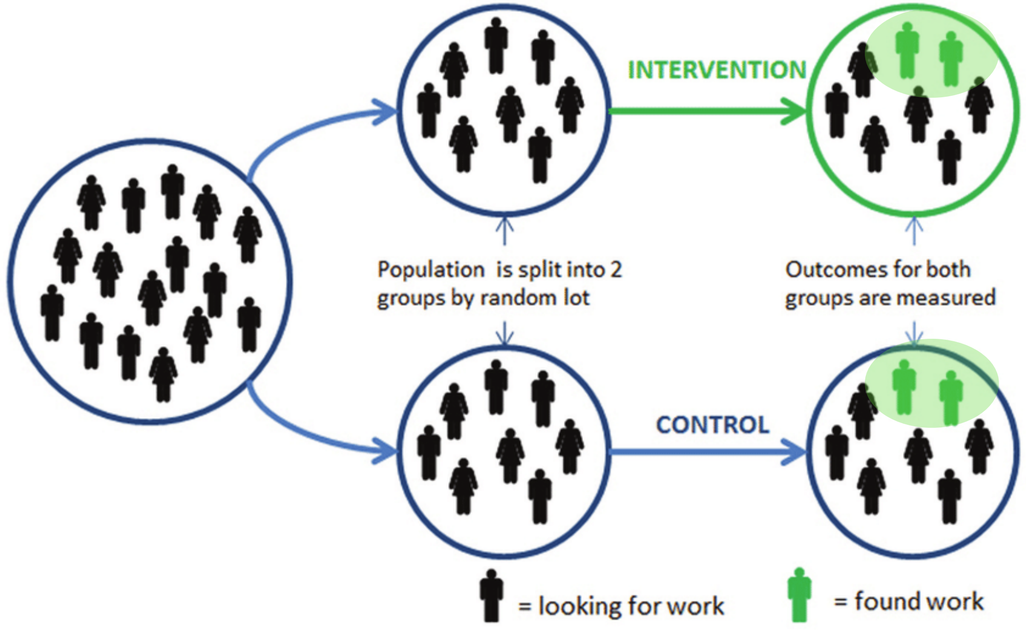
https://www.researchgate.net/publication/256031307_Test_Learn_Adapt_Developing_Public_Policy_with_Randomised_Controlled_Trials

Randomised controlled trials (RCT): spot the difference?

Positive effect



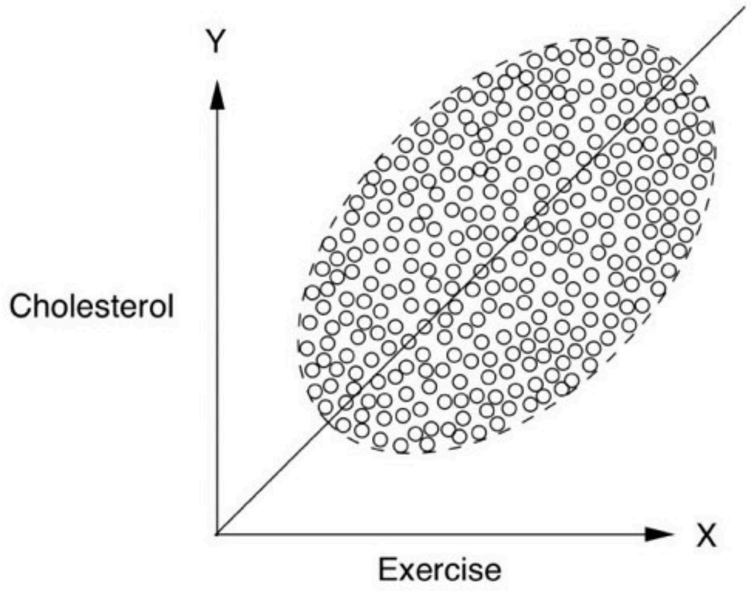
No effect



https://www.researchgate.net/publication/256031307_Test_Learn_Adapt_Developing_Public_Policy_with_Randomised_Controlled_Trials

Observational data: when we do not have RCTs

Let's assume we have **observational data** (e.g. data collected by hospitals)



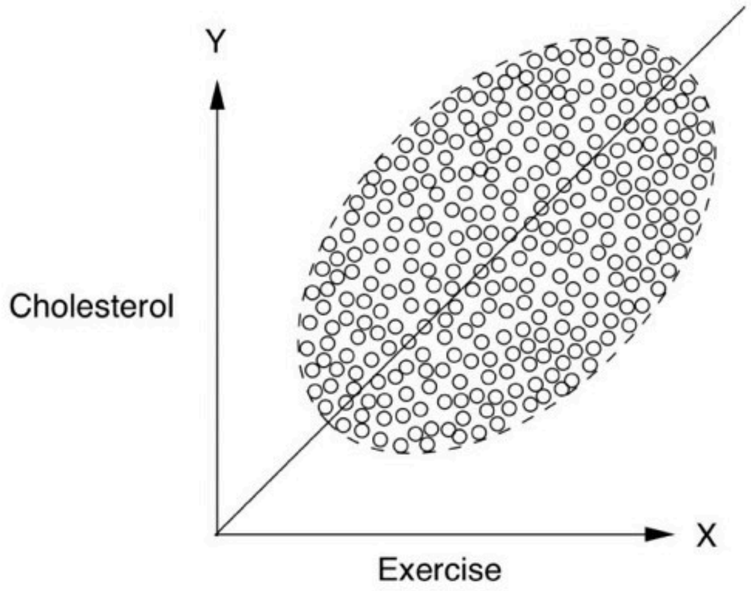
We don't know if there is a policy (and in case, which one) of how people decide to exercise.

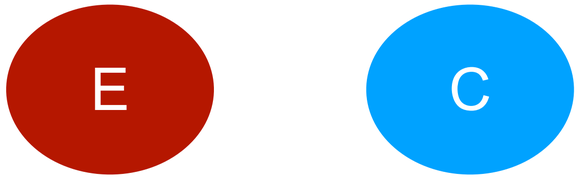
We know they were not randomly split in two similar groups and randomly assigned exercise.

From the Book of Why [Pearl 2018]

What if we don’t have an RCT?

Let’s assume we have **observational data** (e.g. data collected by hospitals)

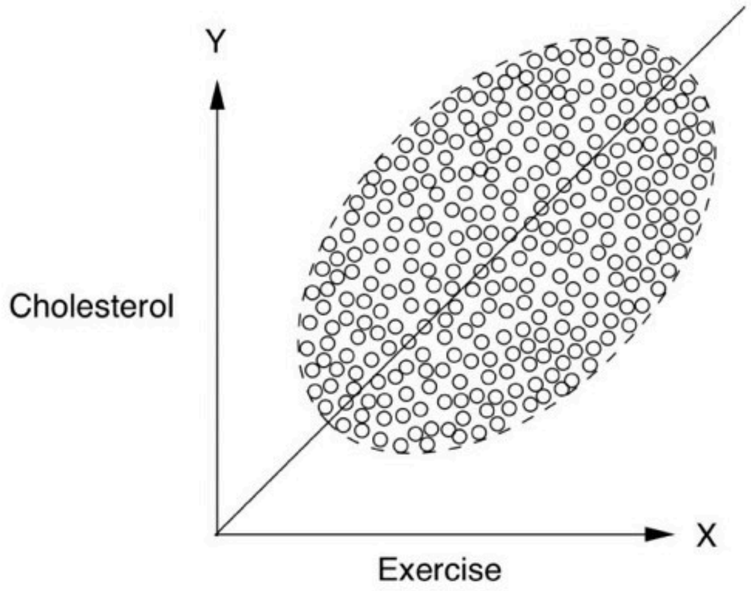


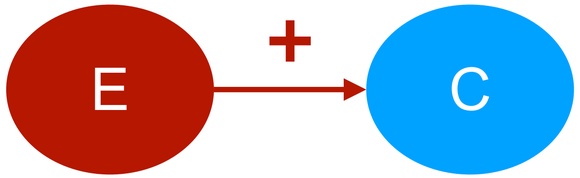


From the Book of Why [Pearl 2018]

What if we don’t have an RCT?

Let’s assume we have **observational data** (e.g. data collected by hospitals)



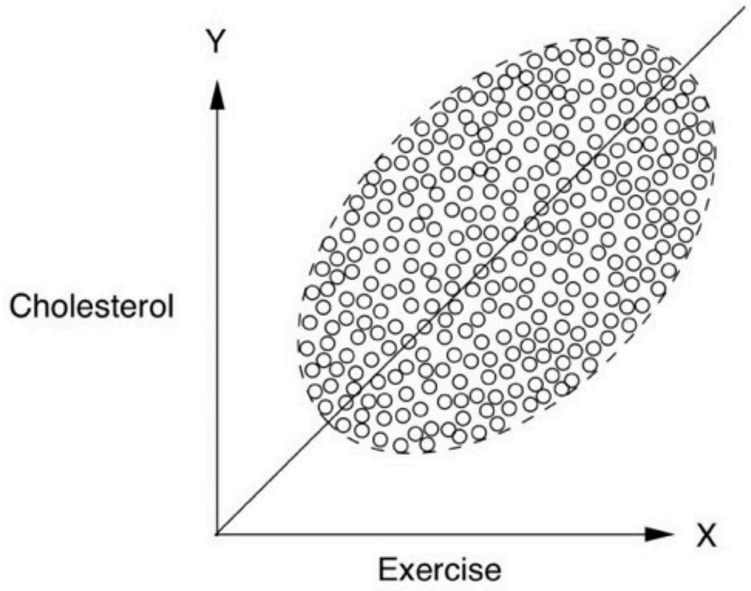


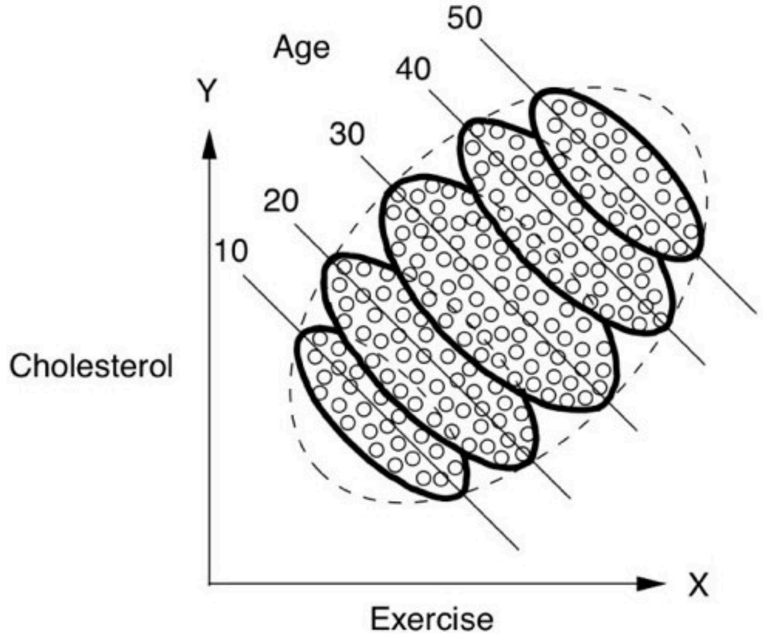
Exercise increases cholesterol

From the Book of Why [Pearl 2018]

What if we don’t have an RCT? Careful!

Let’s assume we have **observational data** (e.g. data collected by hospitals)

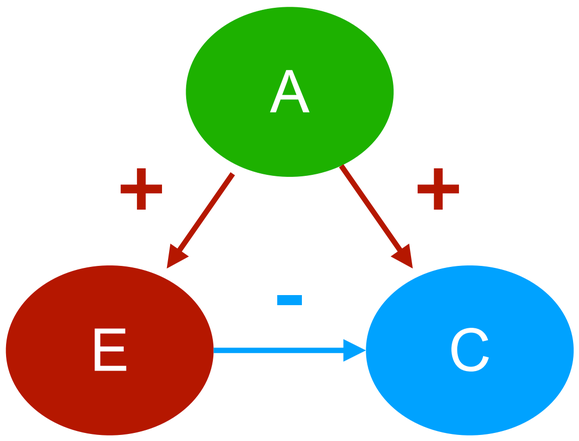




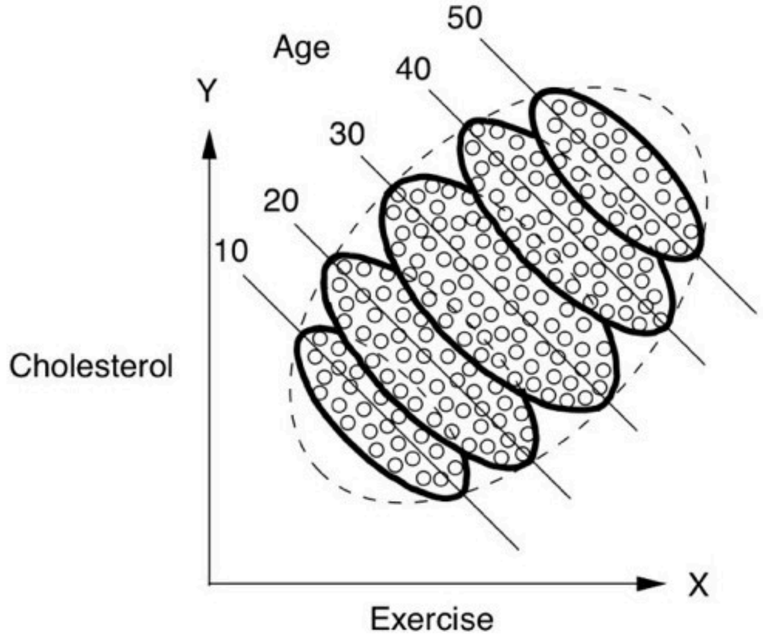
From the Book of Why [Pearl 2018]

What if we don’t have an RCT? Opposite conclusion

Let’s assume we have **observational data** (e.g. data collected by hospitals)



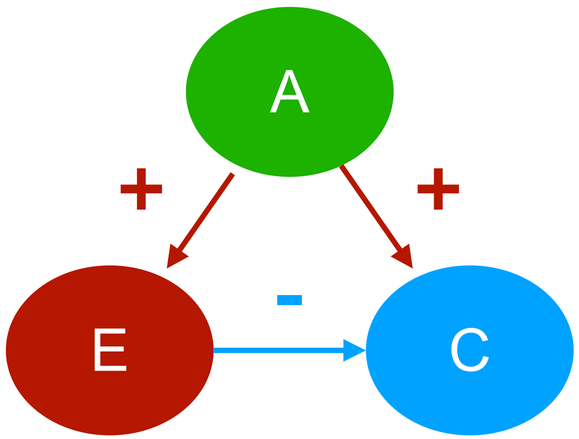
Exercise decreases cholesterol

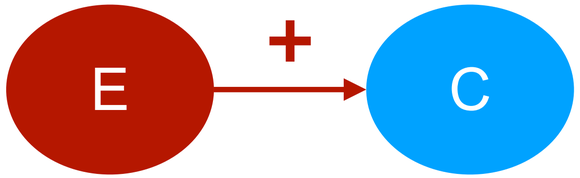


From the Book of Why [Pearl 2018]

What if we don’t have an RCT? Opposite conclusion

Let’s assume we have **observational data** (e.g. data collected by hospitals)

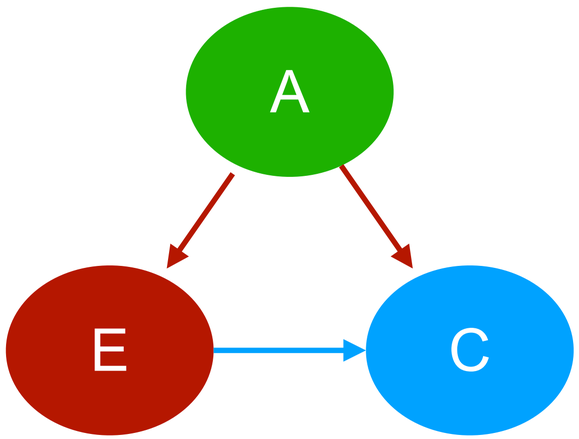


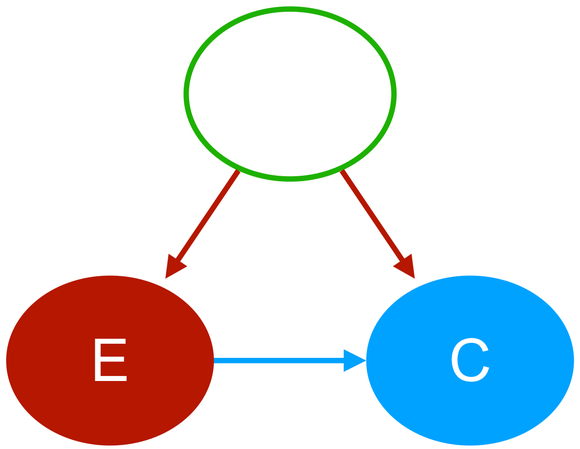


From the Book of Why [Pearl 2018]

What if we don't have an RCT? Opposite conclusion

Let's assume we have **observational data** (e.g. data collected by hospitals)





Controlling for confounding

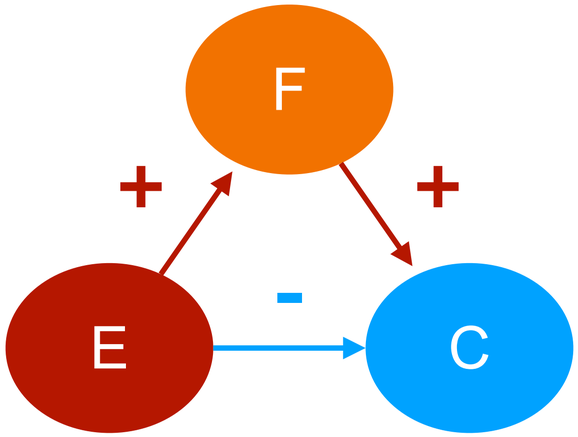
It confounds the effect

A is a confounder (a common cause of the treatment and outcome)

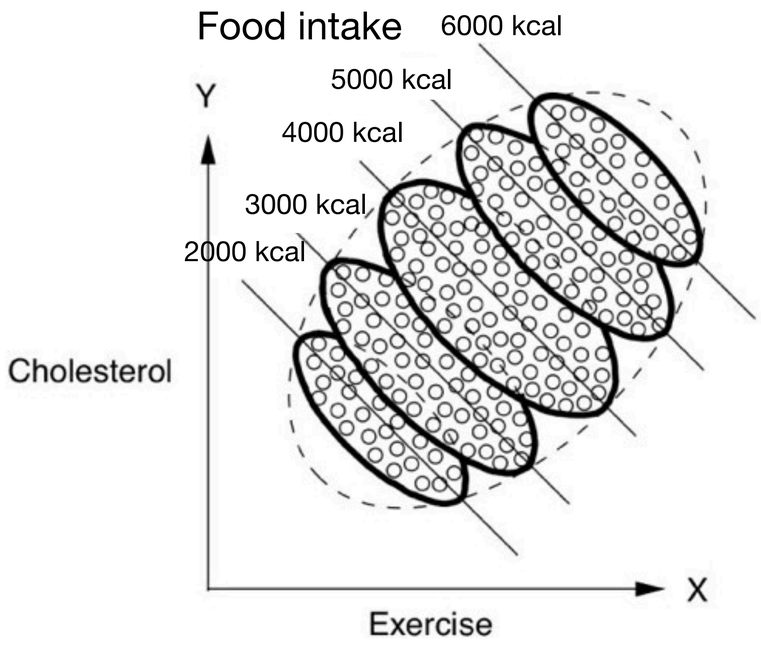
From the Book of Why [Pearl 2018]

Shall we always just control on everything?

In alternative universe:



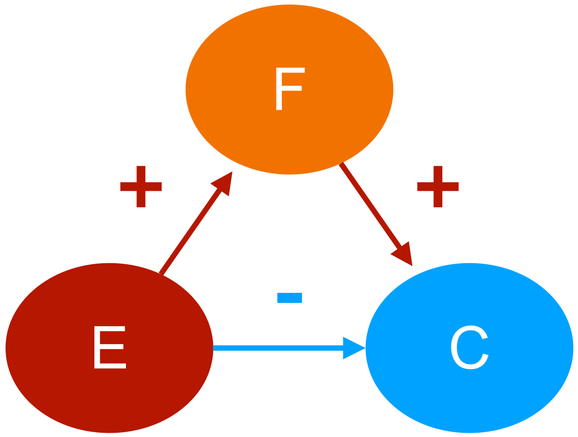
Exercise **increases** cholesterol

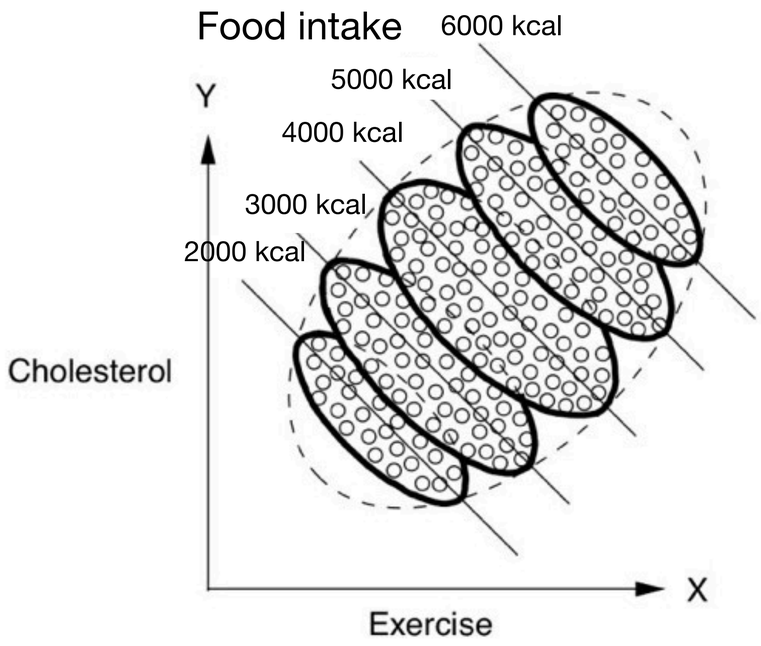


Edited based on the Book of Why [Pearl 2018]

Shall we always just control on everything?

In alternative universe:





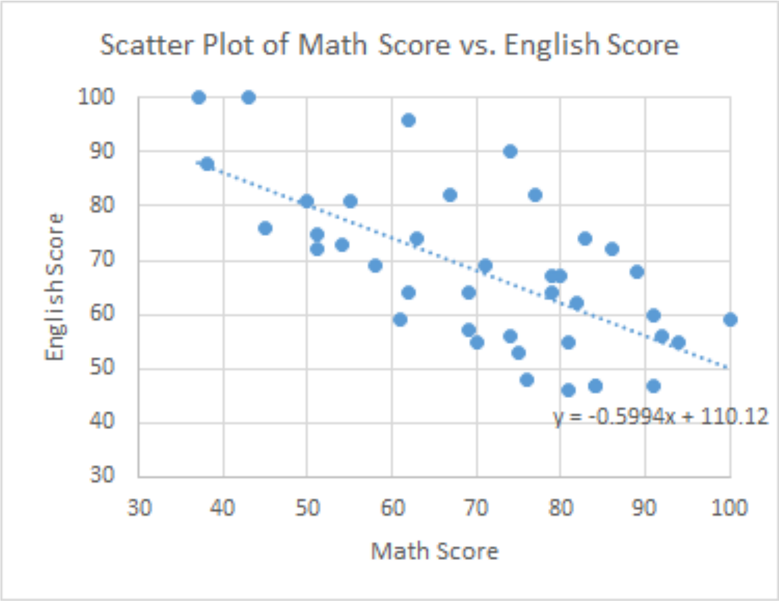
Exercise **increases** cholesterol

E has both a direct and indirect (through F) effect on C

We will see later in the course on which **covariates** we should **adjust**.

Edited based on the Book of Why [Pearl 2018]

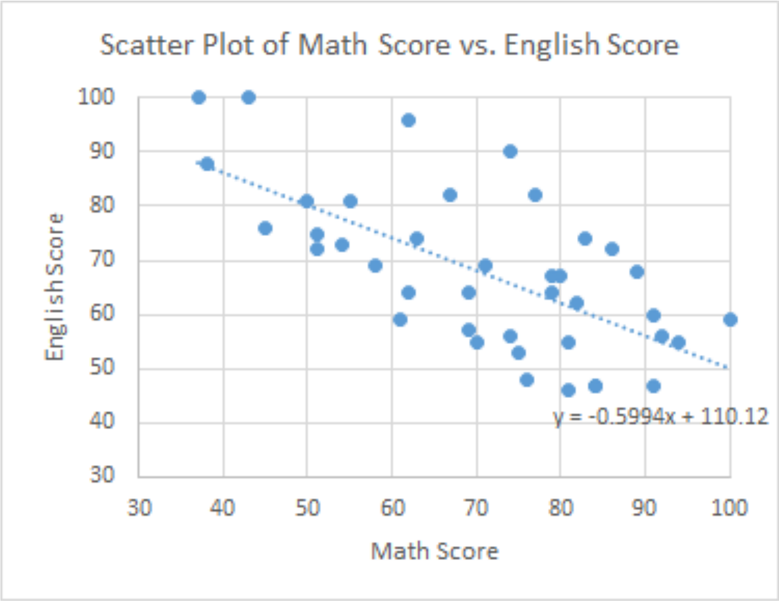
Another example of bias: Selection bias

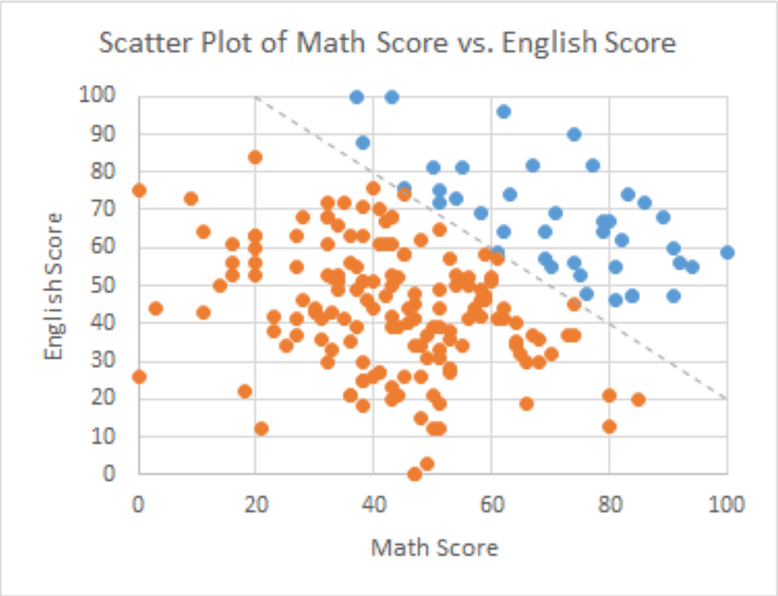


College admissions: English score is negatively correlated with Math Score

https://medium.com/analytics-vidhya/simpsons-paradox-when-you-derive-a-wrong-insight-from-your-analysis-ee488b346427

Another example of bias: Selection bias

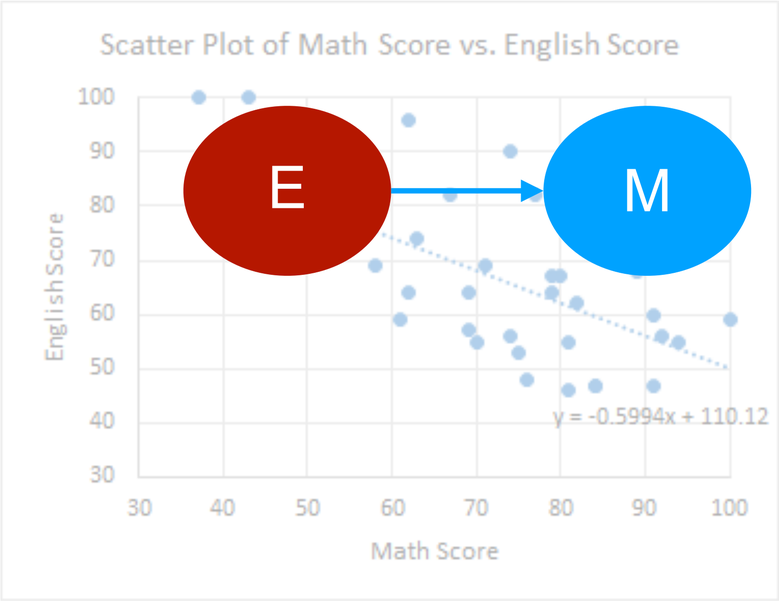


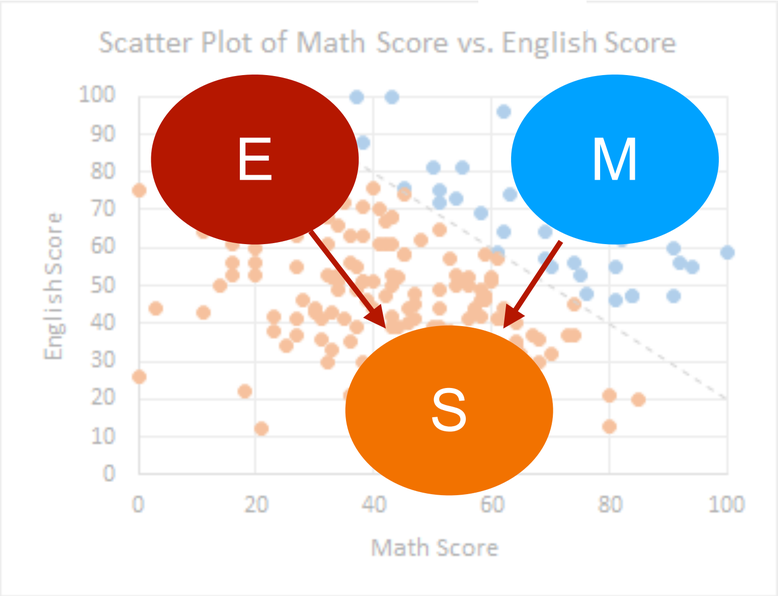


Selection rule: English score + Math Score > 120

https://medium.com/analytics-vidhya/simpsons-paradox-when-you-derive-a-wrong-insight-from-your-analysis-ee488b346427

Another example of bias: Selection bias



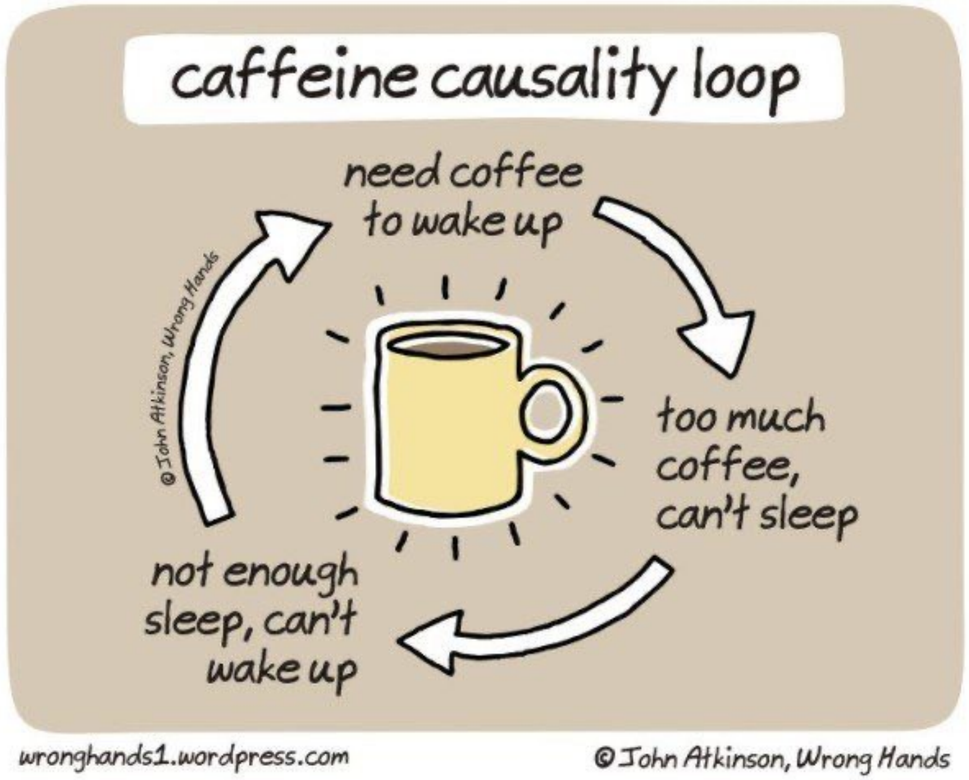


Selection rule: English score + Math Score > 120

https://medium.com/analytics-vidhya/simpsons-paradox-when-you-derive-a-wrong-insight-from-your-analysis-ee488b346427

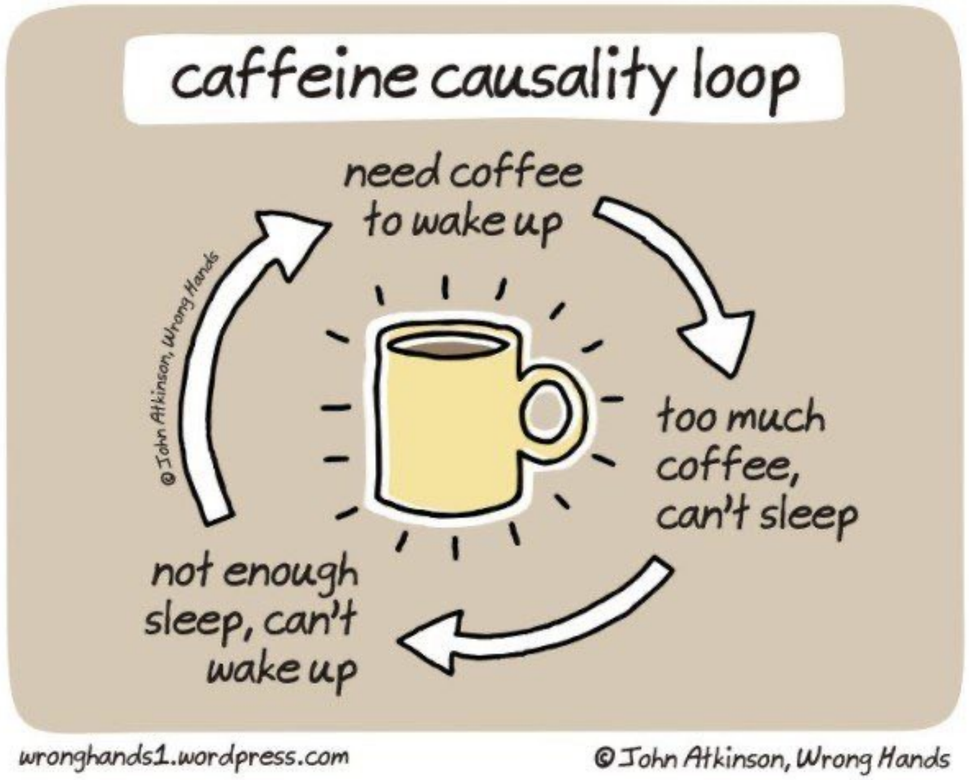
Cycles can also introduce spurious relations

Definition: If X causes Y and Y causes X (even indirectly), X and Y are in a **cycle.**

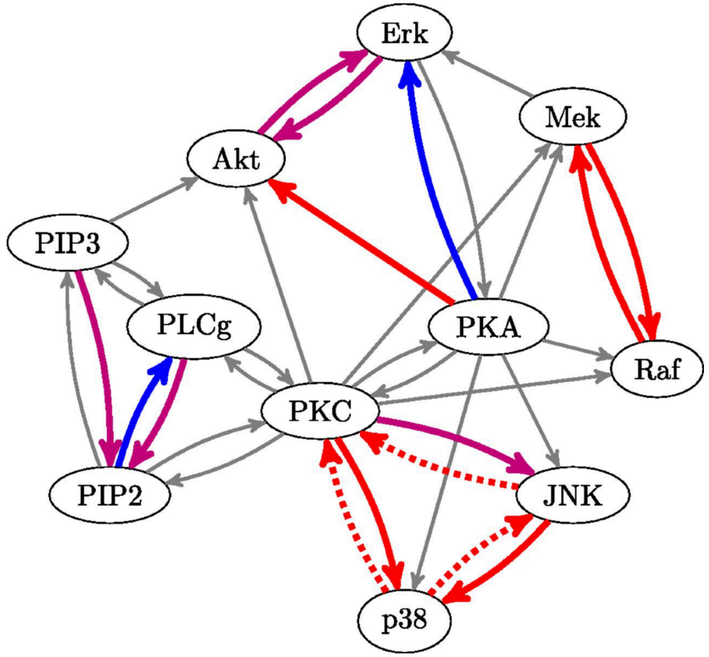


Cycles can also introduce spurious relations

Definition: If X causes Y and Y causes X (even indirectly), X and Y are in a **cycle.**



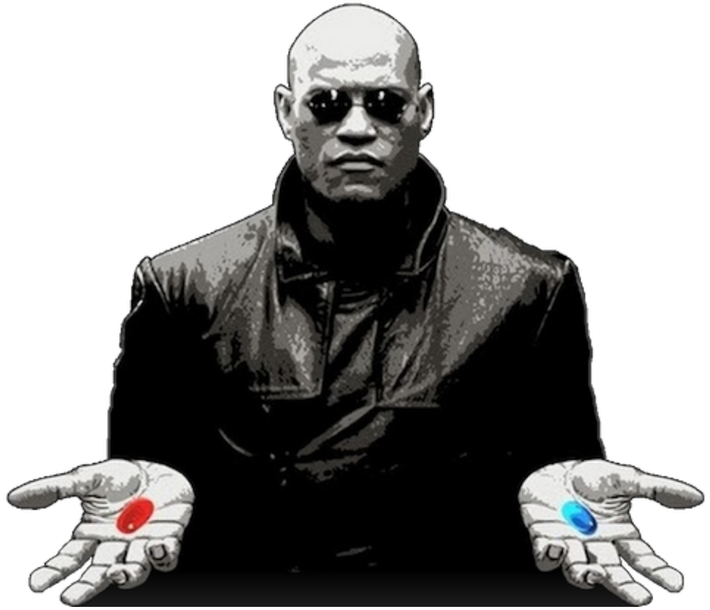
Can introduce spurious correlations



https://www.pnas.org/content/113/27/7361

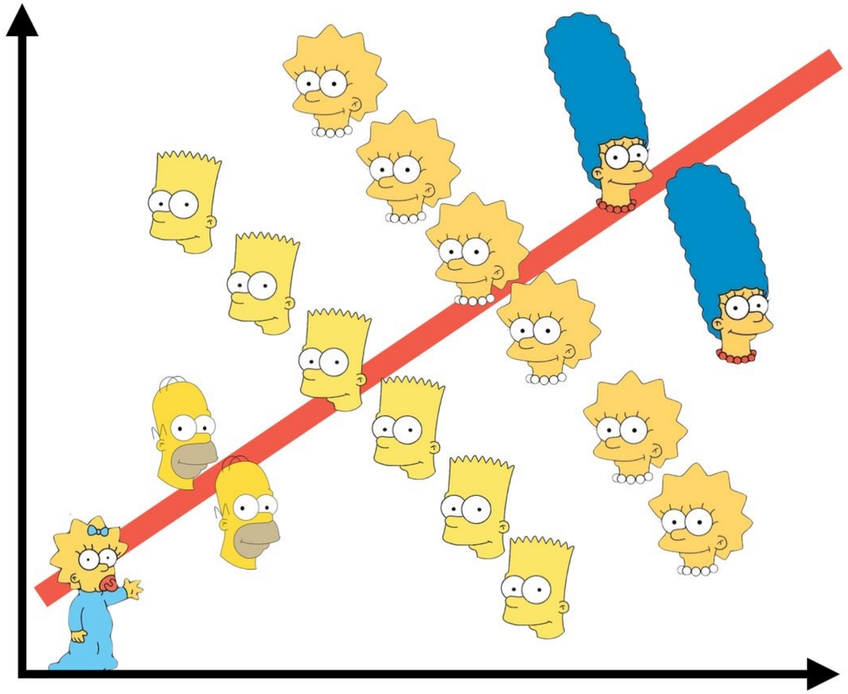
Canvas quiz: Treatment A or treatment B?

- In a hypothetical future there is a new disease and you are in charge of deciding which of the two new treatments, treatment A or treatment B is the most effective.



https://www.npmjs.com/package/blue-pill

Simpson’s paradox



What if we don't know the graph? Causal discovery

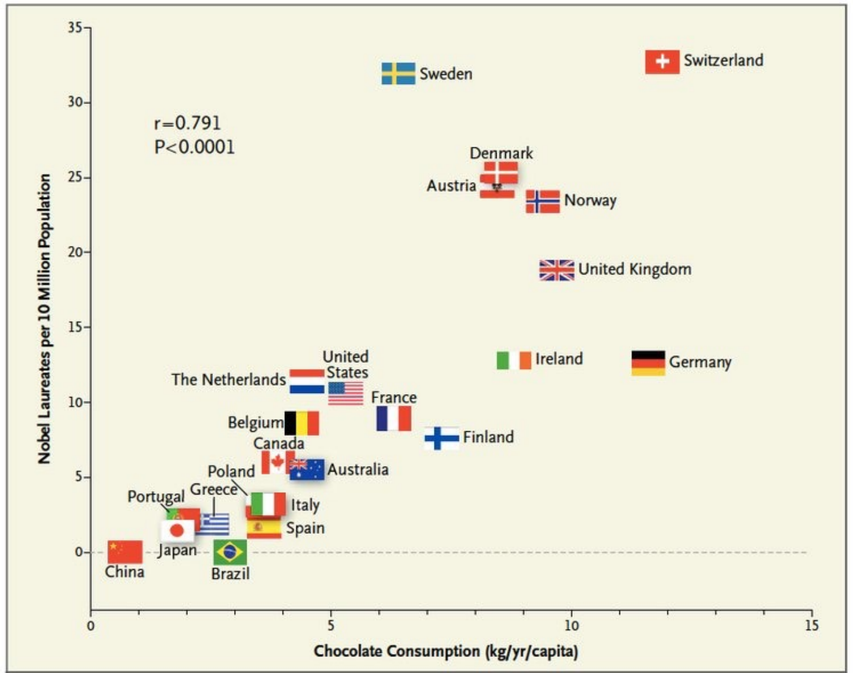


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Reichenbach's principle of common cause: A correlation between X and Y implies that

- X causes Y, or
- Y causes X, or
- There exists a common cause between X and Y

(or any combination of the above)

((we ignore selection bias))

What if we don't know the graph? Causal discovery

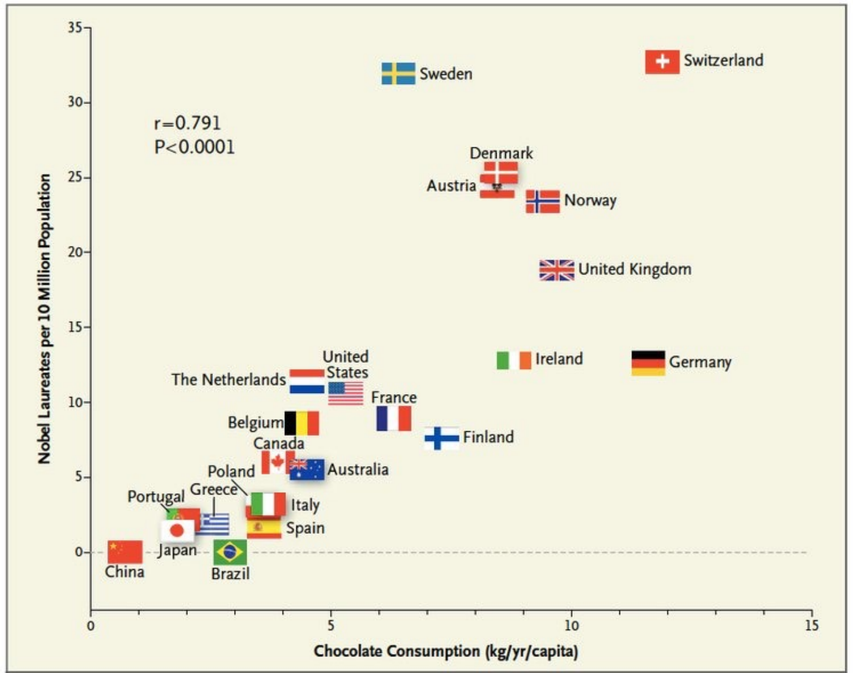
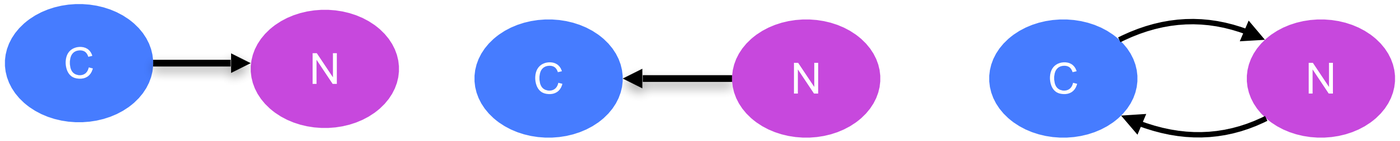


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



Reichenbach's principle of common cause:

A correlation between X and Y implies that X causes Y, Y causes X, or there exists a common cause between X and Y (or any combination)

What if we don't know the graph? Causal discovery

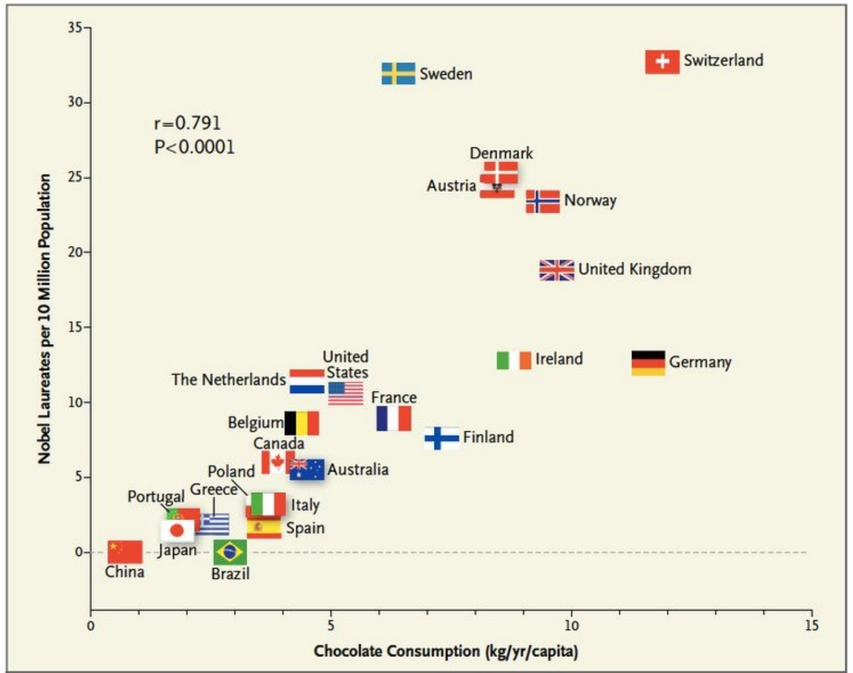
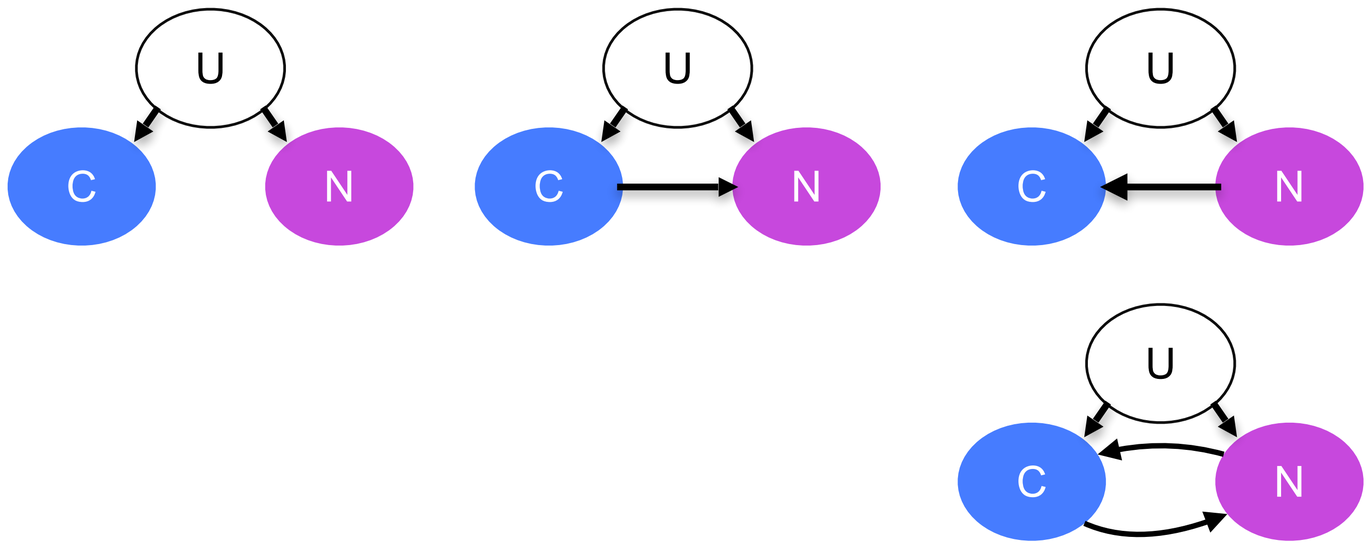


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



Reichenbach's principle of common cause:

A correlation between X and Y implies that X causes Y, Y causes X, or there exists a common cause between X and Y (or any combination)

Another example of confounding

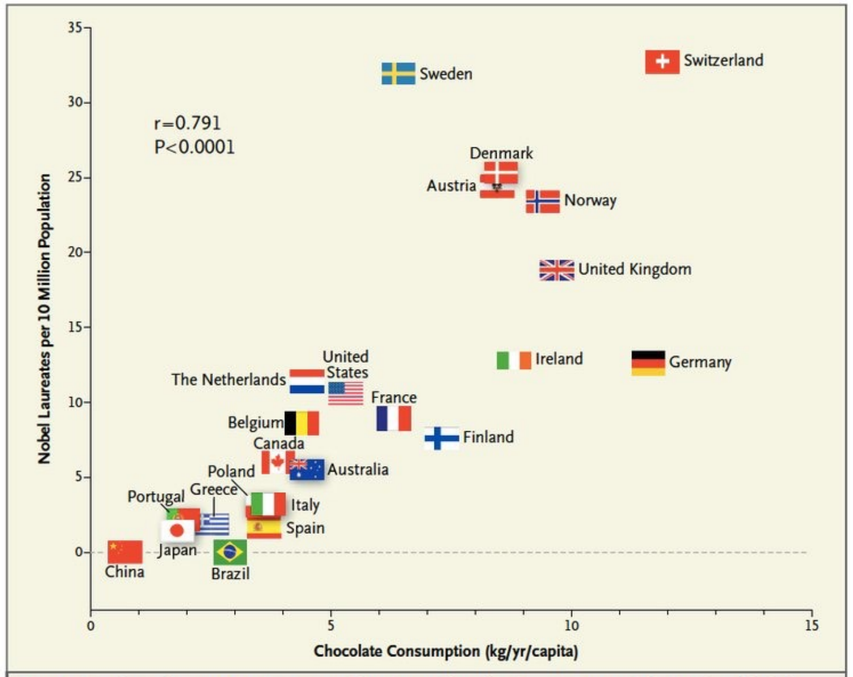
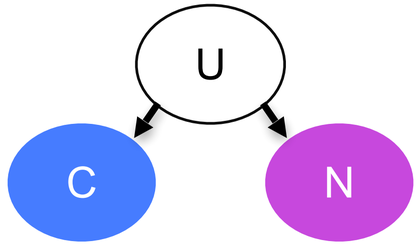


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



Another example of confounding

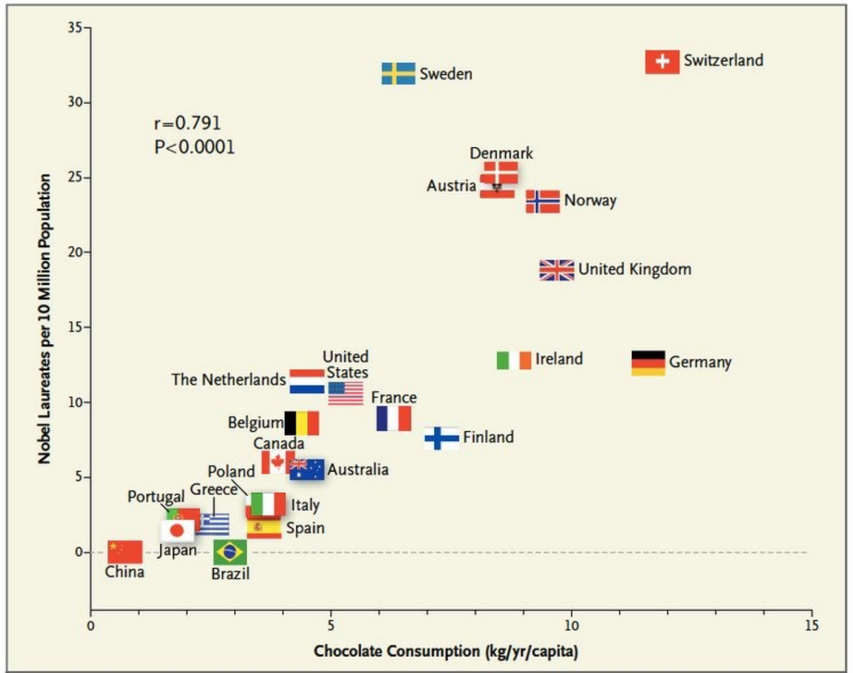
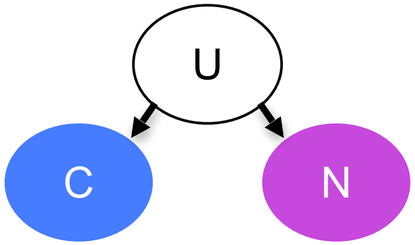
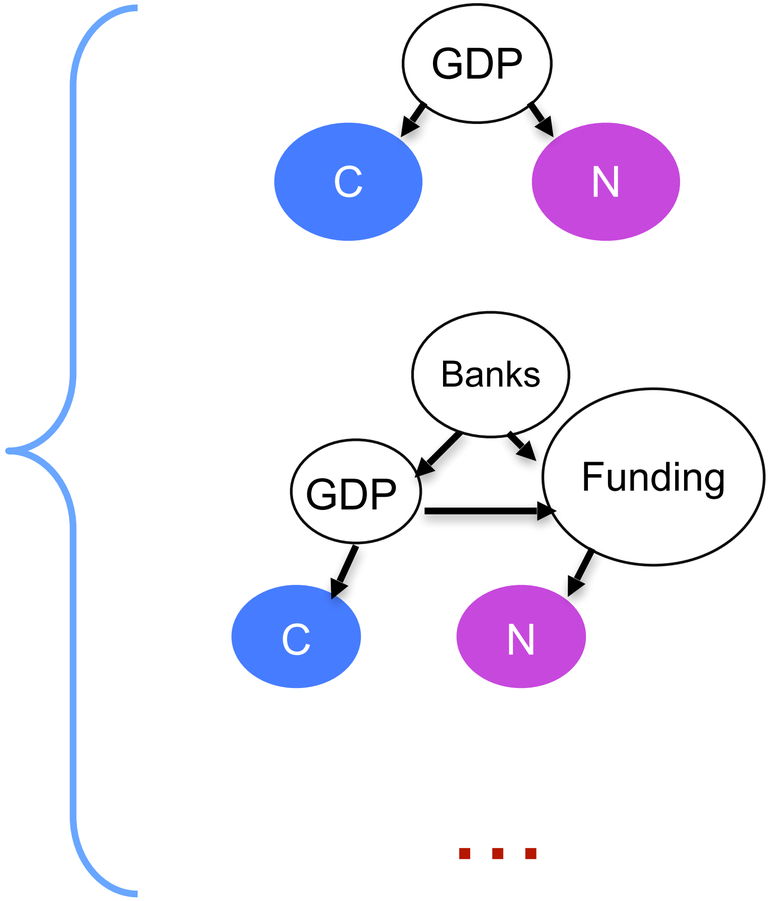


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.





Statistically indistinguishable from only these data

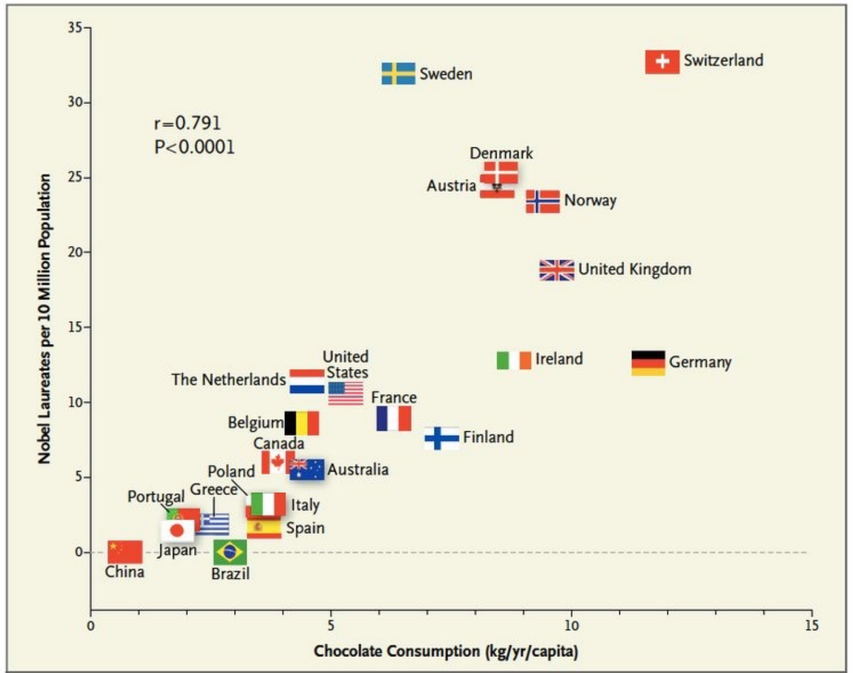
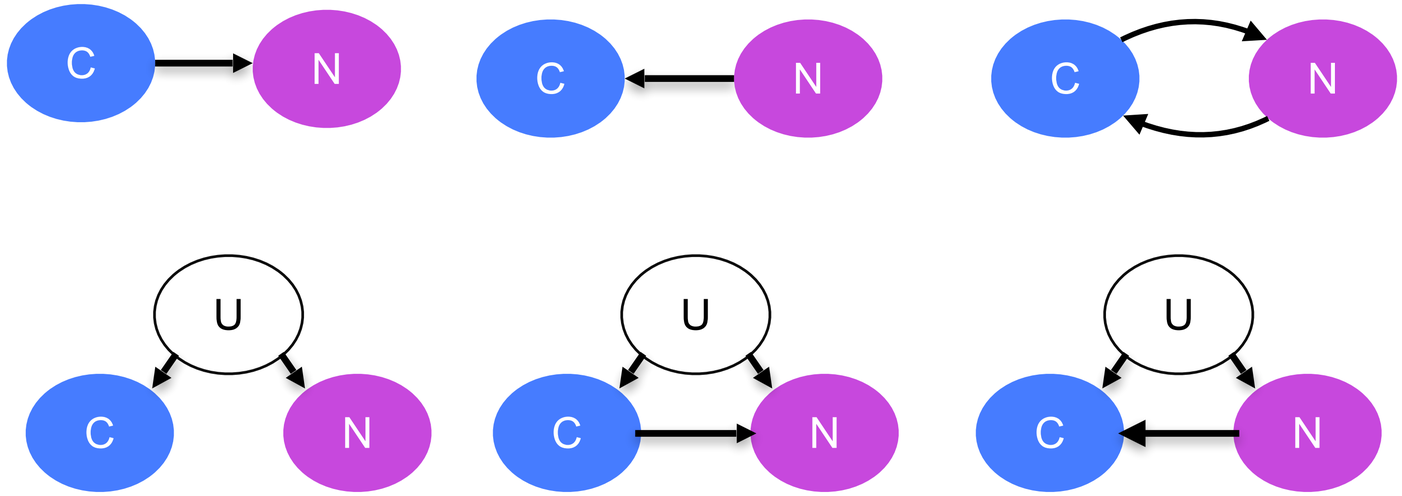
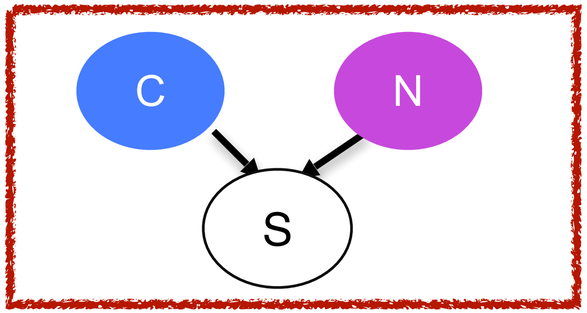
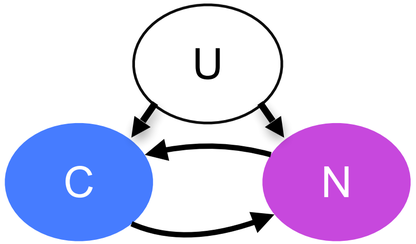


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



Or even selection bias





Statistically indistinguishable from only these data

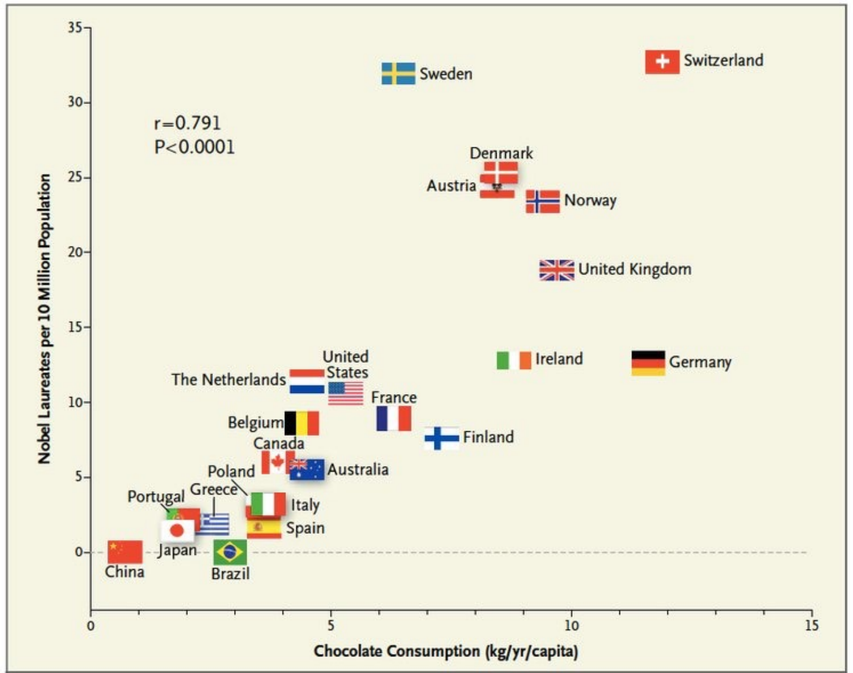
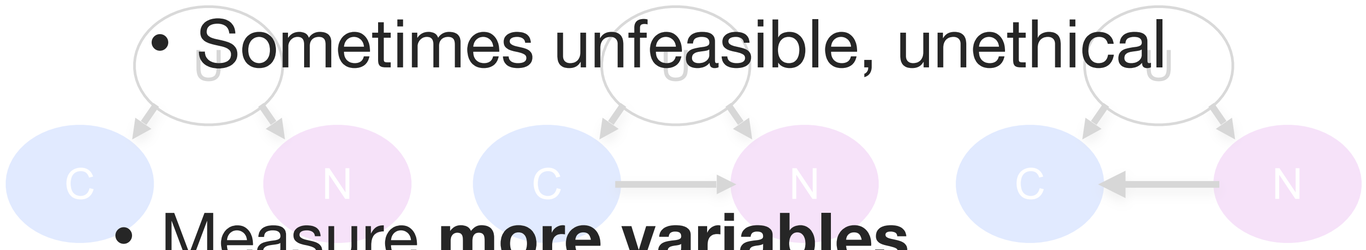


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

- Add some **experimental** data
  - Sometimes unfeasible, unethical

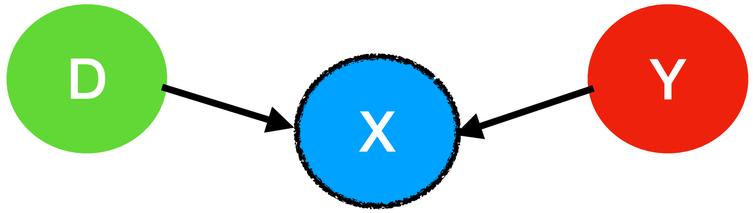


- Measure **more variables**

Or even selection bias

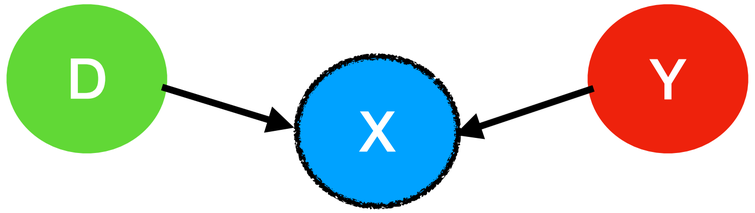
A more general method for causal discovery

- Given a causal graph, **d-separation** is a graphical criterion that (under standard conditions) allows us to read **conditional independences**



A more general method for causal discovery

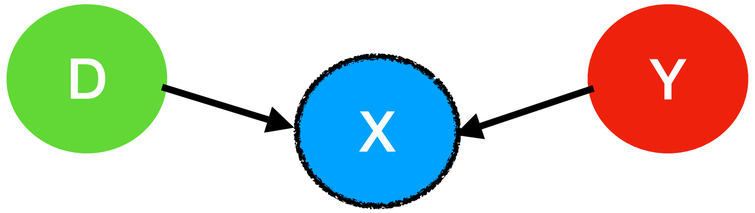
- Given a causal graph, **d-separation** is a graphical criterion that (under standard conditions) allows us to read **conditional independences**



Graph: $Y \not\perp\!\!\!\perp_d D$ $\quad Y \perp\!\!\!\perp_d D \mid X$ $\quad Y \perp\!\!\!\perp_d D$ $\quad Y \not\perp\!\!\!\perp_d D \mid X$

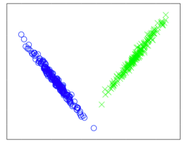
A more general method for causal discovery

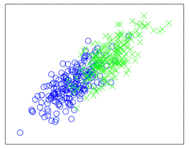
- Given a causal graph, **d-separation** is a graphical criterion that (under standard conditions) allows us to read **conditional independences**

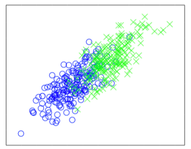


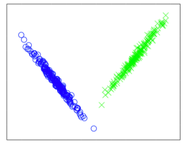
Graph: $Y \not\perp\!\!\!\perp_d D$ $\quad Y \perp\!\!\!\perp_d D \mid X$ $\quad Y \perp\!\!\!\perp_d D \mid X$ $\quad Y \not\perp\!\!\!\perp_d D \mid X$

Data: $Y \not\perp\!\!\!\perp D$ $\quad Y \perp\!\!\!\perp D \mid X$ $\quad Y \perp\!\!\!\perp D$ $\quad Y \not\perp\!\!\!\perp D \mid X$



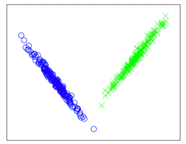


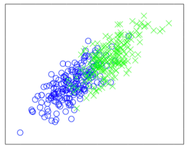


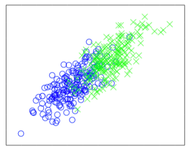


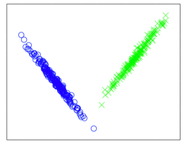
A more general method for causal discovery

- Given a causal graph, **d-separation** is a graphical criterion that (under standard conditions) allows us to read **conditional independences**



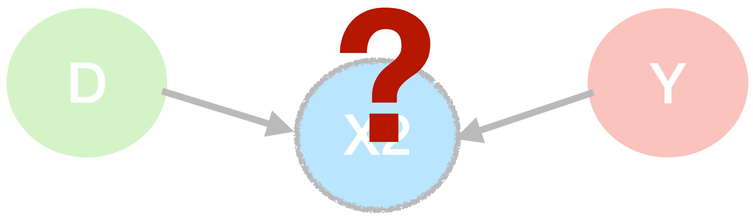




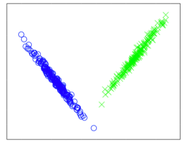


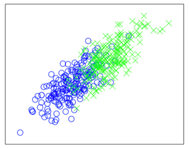
A more general method for causal discovery

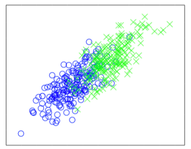
- Given a causal graph, **d-separation** is a graphical criterion that (under standard conditions) allows us to read **conditional independences**

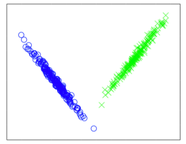


Data: $Y \not\perp\!\!\!\perp D$ $\quad Y \perp\!\!\!\perp D \mid X_1$ $\quad Y \perp\!\!\!\perp D$ $\quad Y \not\perp\!\!\!\perp D \mid X_2$



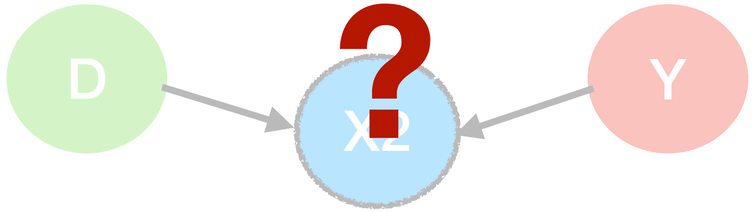






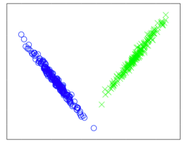
A more general method for causal discovery

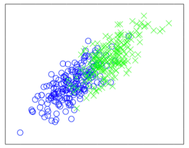
- Given a causal graph, **d-separation** is a graphical criterion that (under standard conditions) allows us to read **conditional independences**

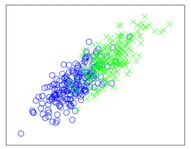


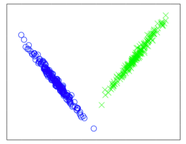
Graph: $Y \not\perp\!\!\!\perp_d D$ $\quad$ $Y \perp\!\!\!\perp_d D \mid X_1$ $\quad$ $Y \perp\!\!\!\perp_d D$ $\quad$ $Y \not\perp\!\!\!\perp_d D \mid X_2$

Data: $Y \not\perp\!\!\!\perp D$ $\quad$ $Y \perp\!\!\!\perp D \mid X_1$ $\quad$ $Y \perp\!\!\!\perp D$ $\quad$ $Y \not\perp\!\!\!\perp D \mid X_2$



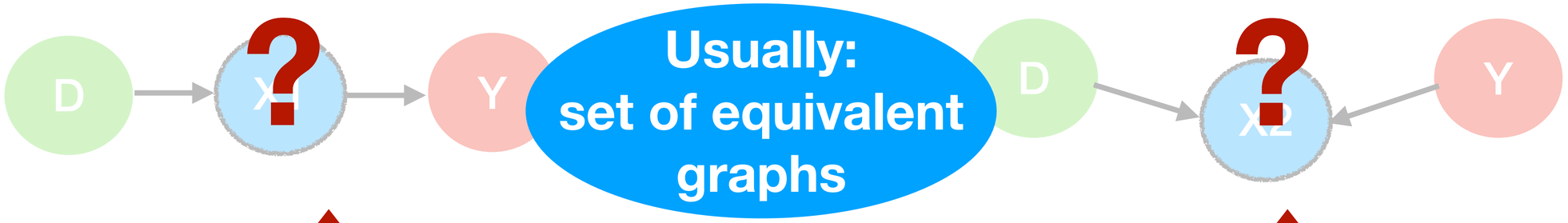






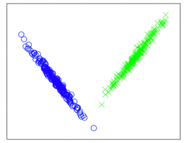
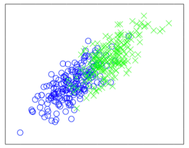
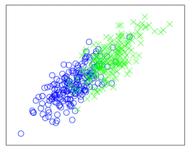
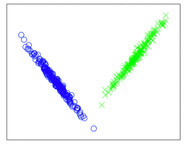
A more general method for causal discovery

- Given a causal graph, **d-separation** is a graphical criterion that (under standard conditions) allows us to read **conditional independences**



Graph: $Y \not\perp\!\!\!\perp_d D$ $\quad Y \perp\!\!\!\perp_d D \mid X_1$ $\quad Y \perp\!\!\!\perp_d D$ $\quad Y \not\perp\!\!\!\perp_d D \mid X_2$

Data: $Y \not\perp\!\!\!\perp D$ $\quad Y \perp\!\!\!\perp D \mid X_1$ $\quad Y \perp\!\!\!\perp D$ $\quad Y \not\perp\!\!\!\perp D \mid X_2$

Causal discovery example: instrumental variables

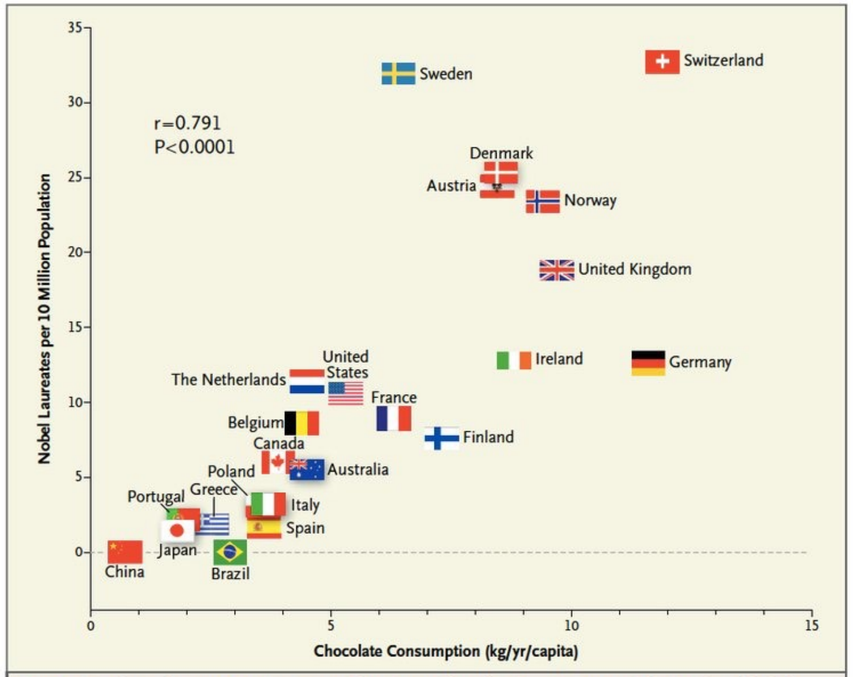
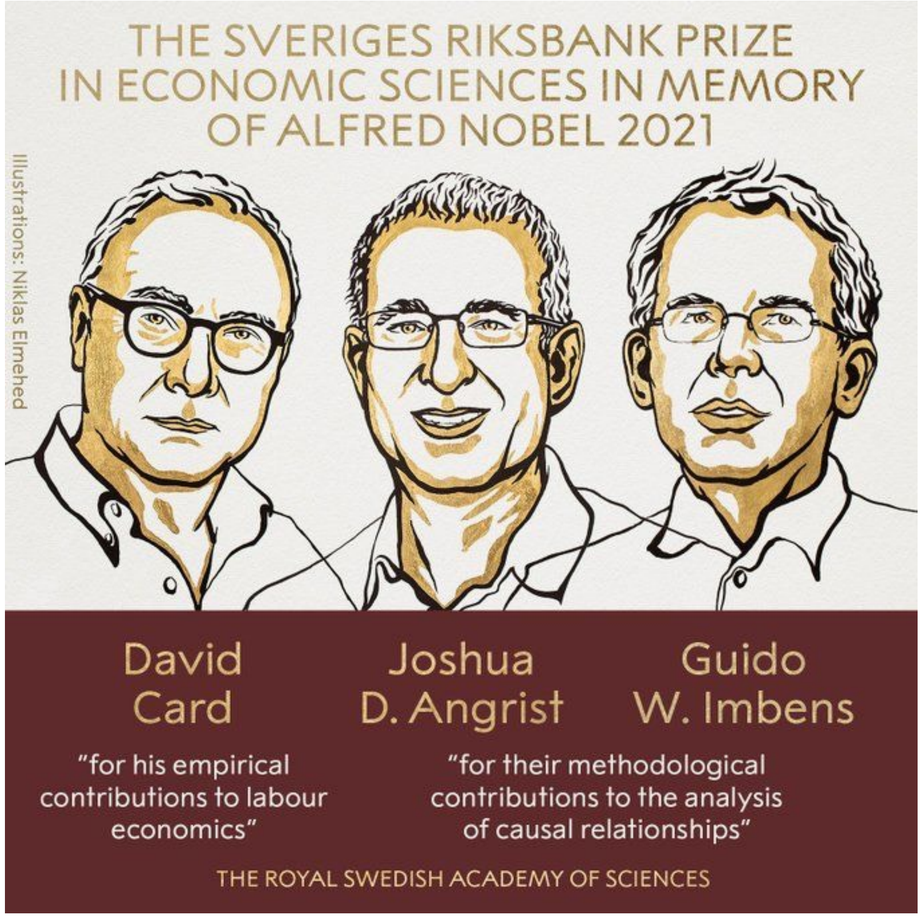


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



Causal discovery example: instrumental variables

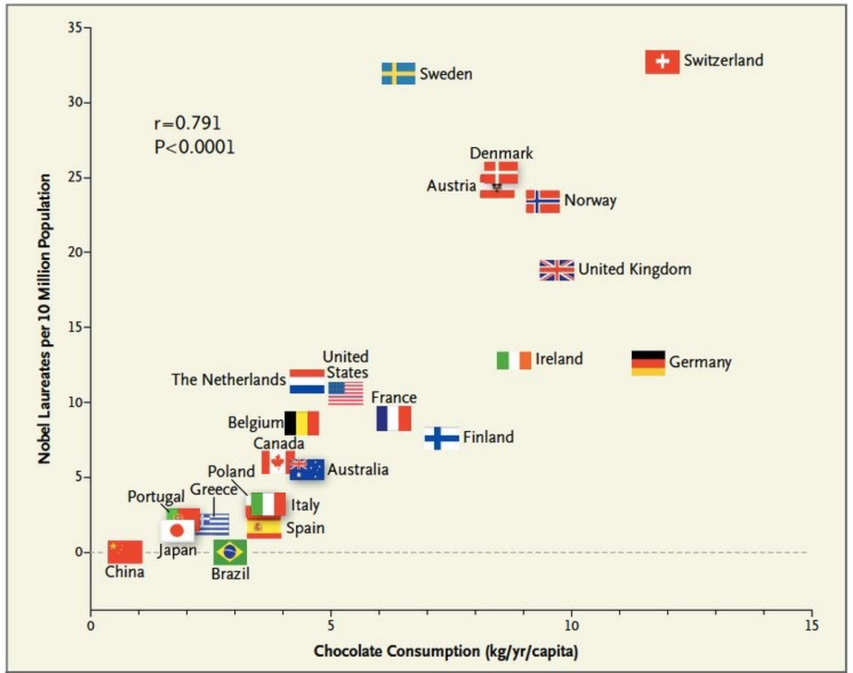
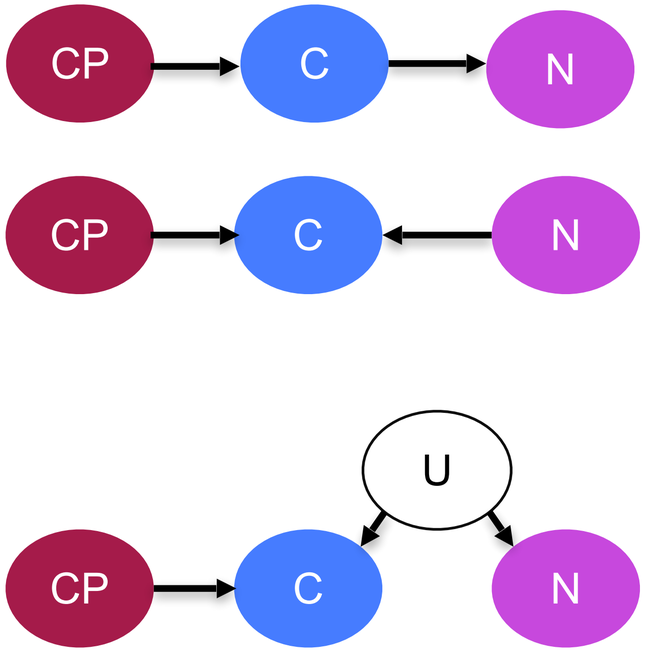


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



...

Chocolate price does not cause Nobel Prizes directly

Causal discovery example: instrumental variables

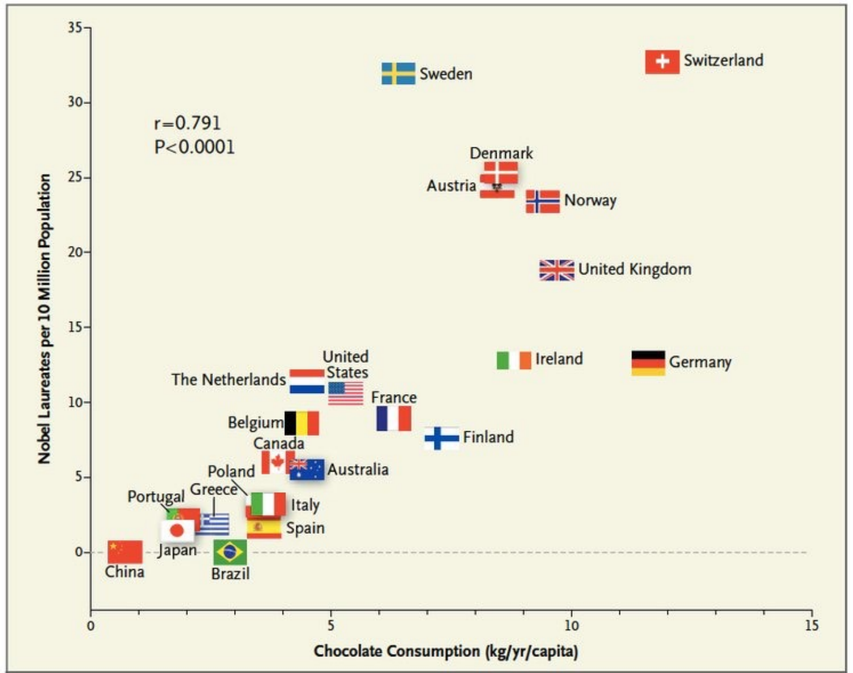
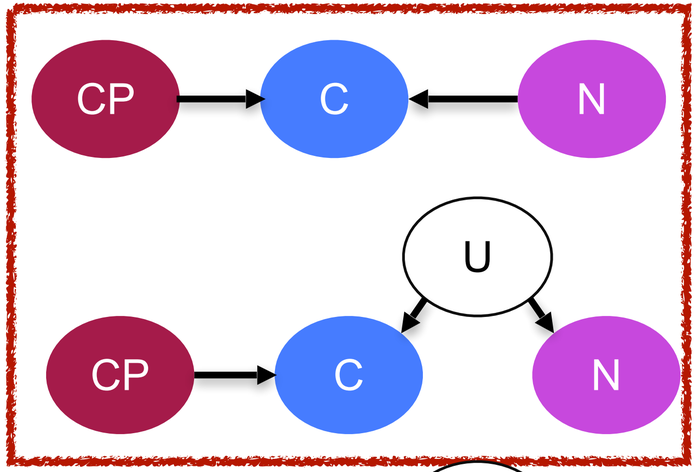
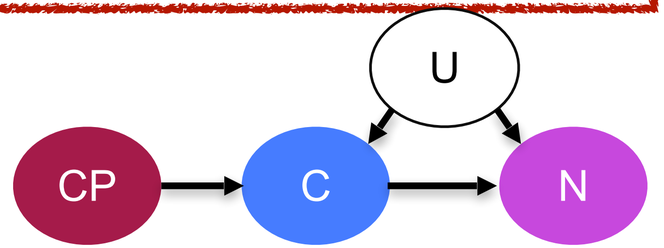
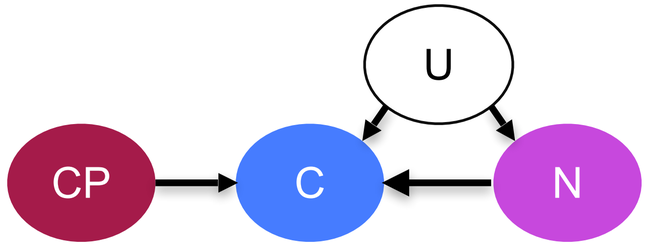


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



If CP is independent of N
(no cycles, no selection bias)





Causal discovery example: beyond instrumental variables

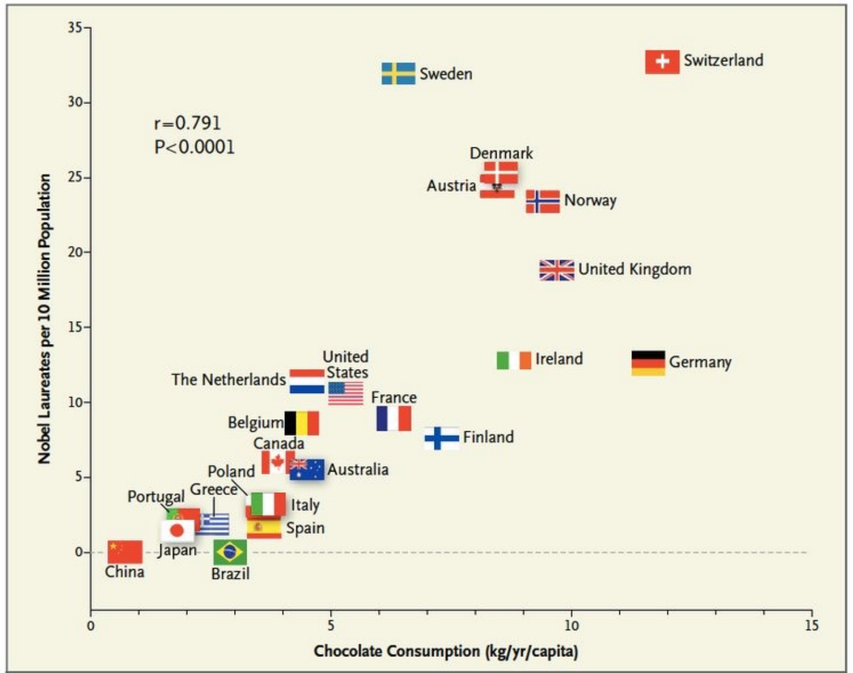
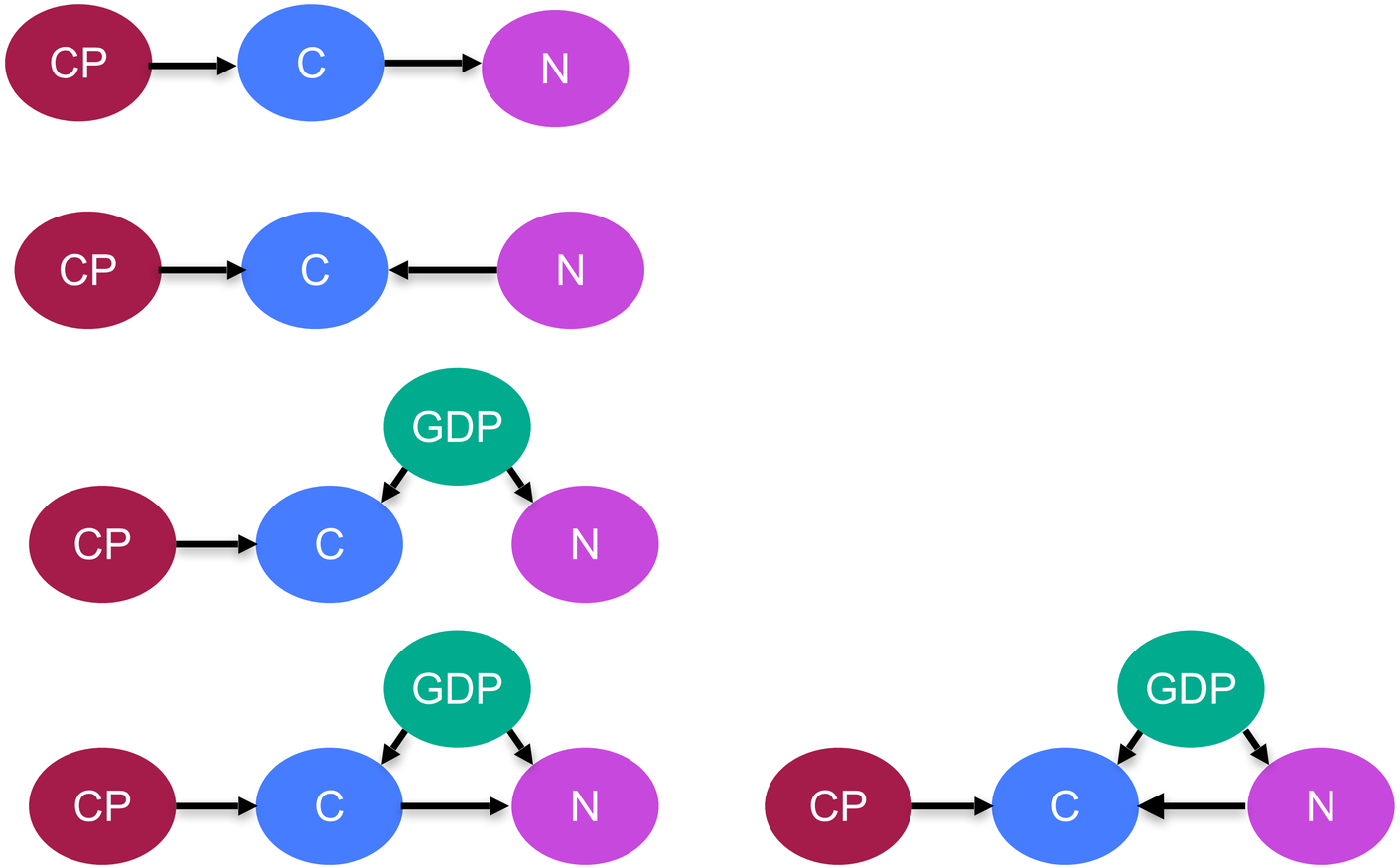


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



...

Causal discovery example: beyond instrumental variables

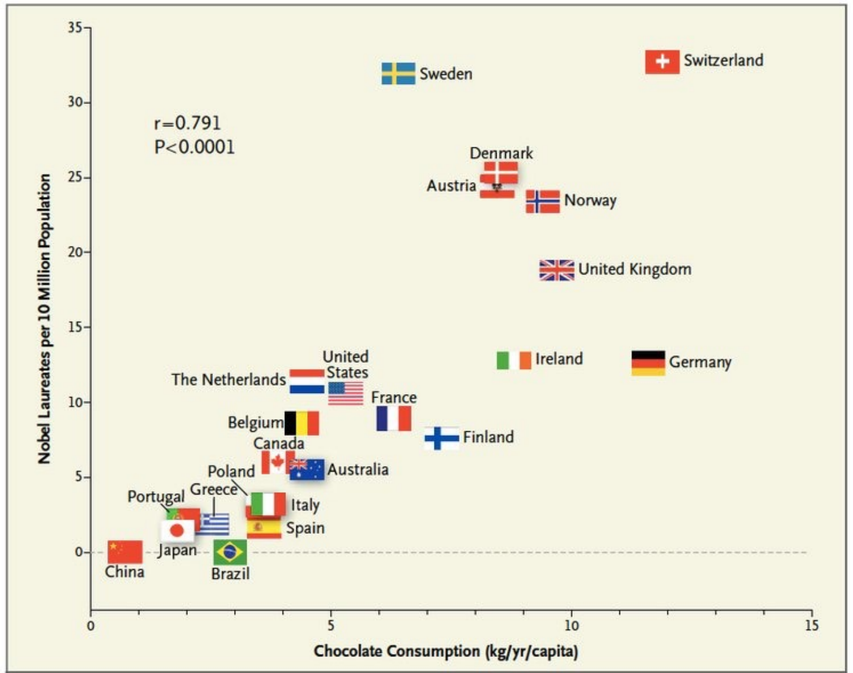
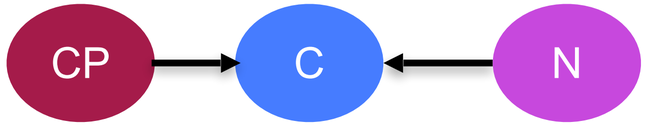
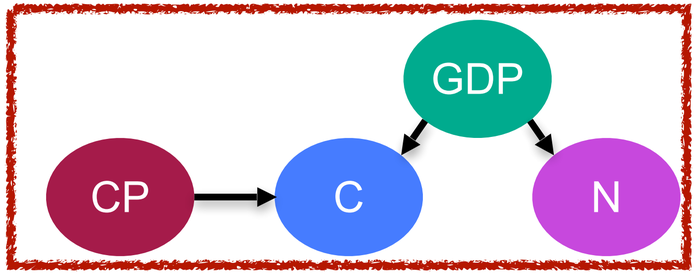


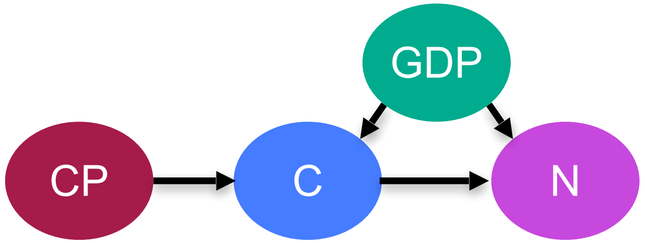
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

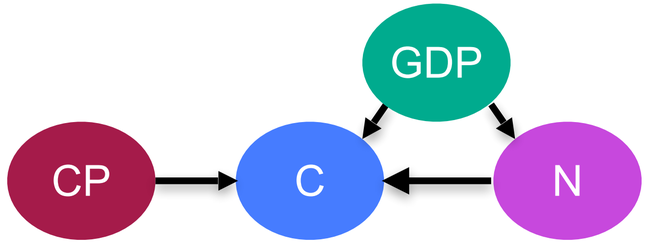




If CP is independent of N

If C is independent of N given GDP





...

Causal discovery example: beyond instrumental variables

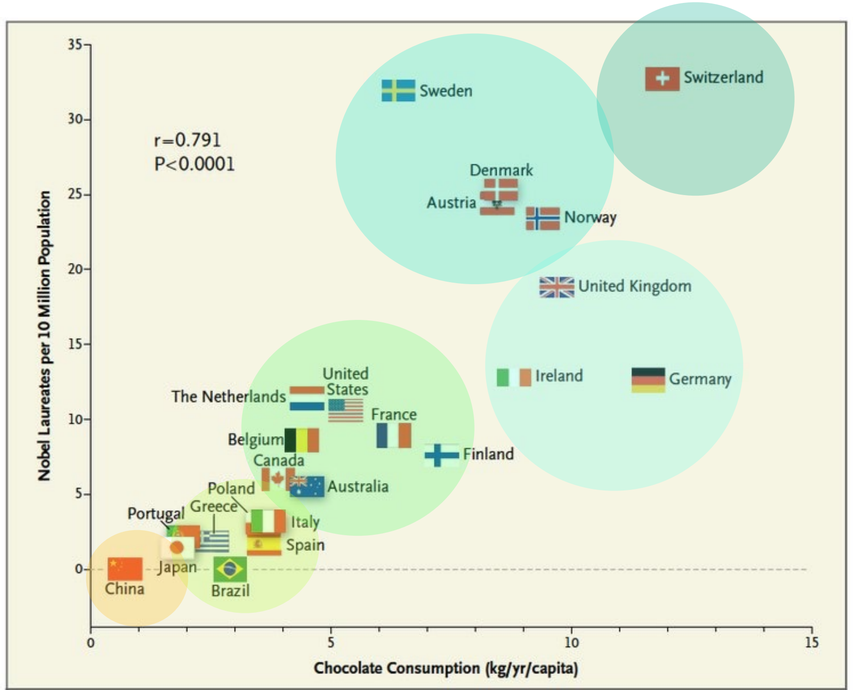
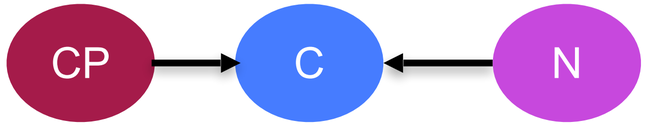
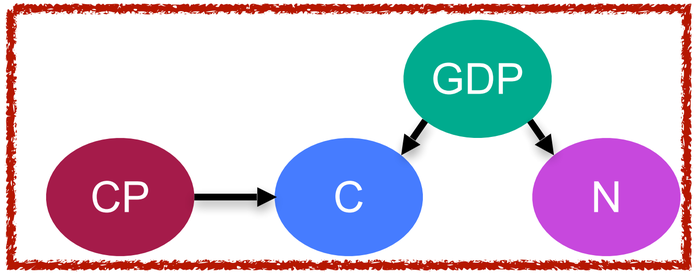


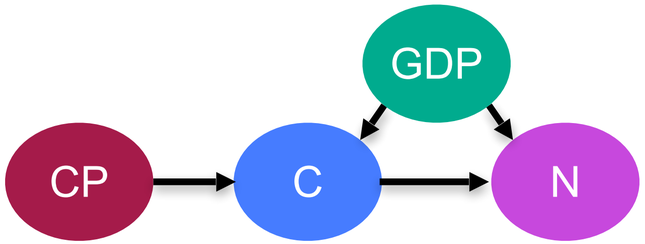
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

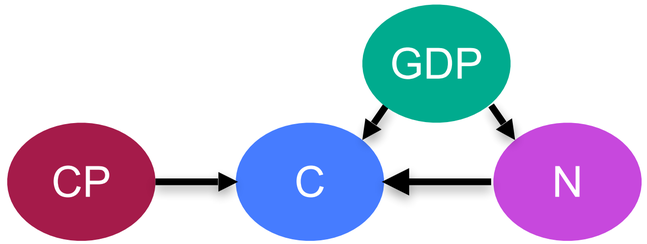




If CP is independent of N

If C is independent of N given GDP





...

Causal discovery example: beyond instrumental variables

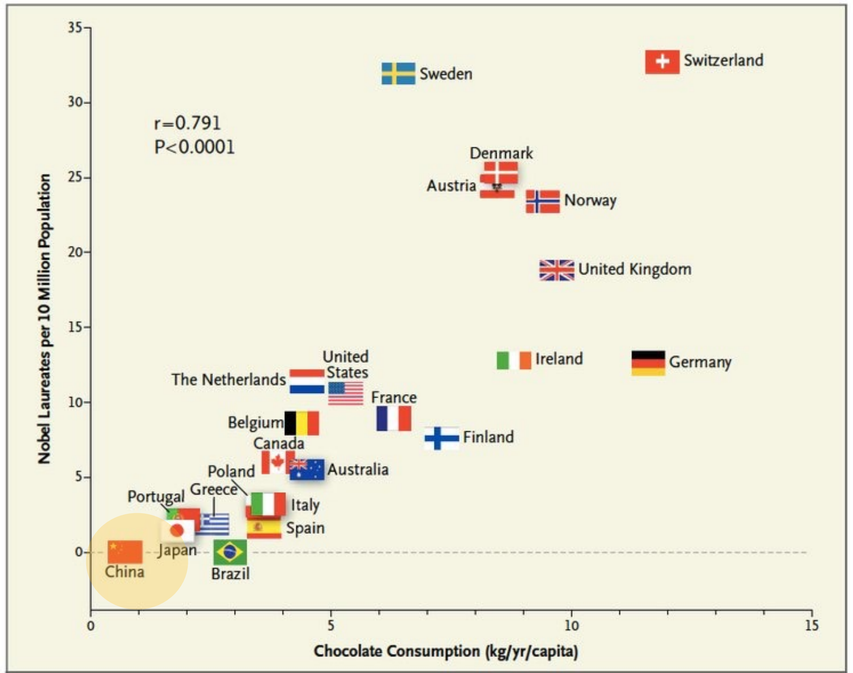
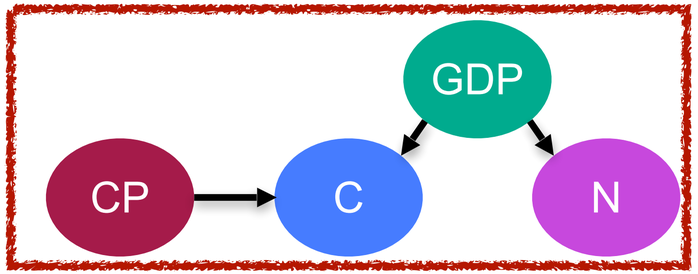


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



If CP is independent of N
If C is independent of N given GDP

Under assumptions we will discuss later in the course, we can sometimes find the causal graph even from observational data.

Causal discovery (causal structure learning)

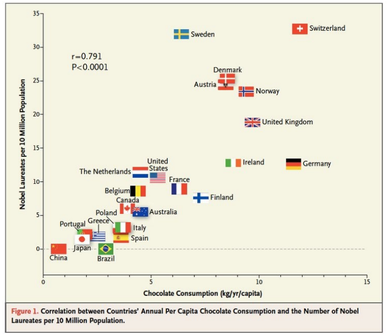


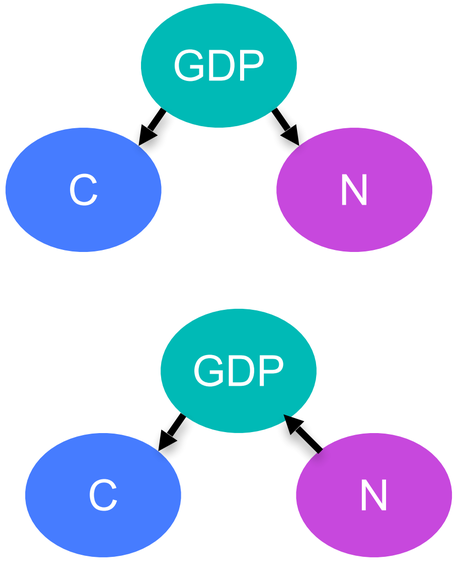
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

C	N	GDP
4.5	5	33k
12	30	86k
10	20	46k
....	...	...

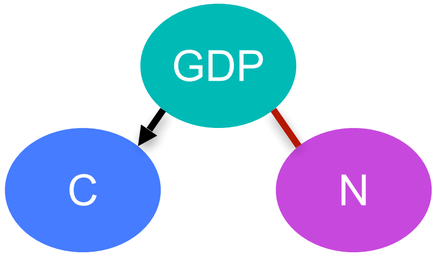
Observational data

$C \nrightarrow GDP$

[Optional] Background knowledge



Sets of graphs that fit the data and background knowledge



Summary graph

Causal discovery simplified overview

Constraint-based causal discovery

- Conditional independence tests
- Observational data
- Output: MEC
- SGS, PC, FCI

Score-based causal discovery

- Penalised likelihood
- Observational data
- Output: MEC
- GES, MMHC

Restricted models

- Nonlinear additive noise, Linear Non-Gaussianity
- Observational data
- Output: DAG
- RESIT, LINGAM

Interventional causal discovery / causal invariance

- Observational and Interventional data
- Output: parents of Y, I-MEC
- ICP, GIES, JCI

Tentative course outline - topics

1	Introduction
2	Probability recap
3	Graphical models, d-separation
4	Causal graphs, Interventions, SCMs
5	Covariate adjustment: backdoor criterion
6	Covariate Frontdoor criterion, Instrumental variables
7	Counterfactuals, potential outcomes, estimating causal effects 1
8	Estimating causal effects 2 (matching, IPW)
9	Constraint based structure learning
10	Score based structure learning, restricted models
11	Do-calculus, transportability, Joint Causal Inference
12	Causality-inspired ML, recap of the course

Background on causal graphs

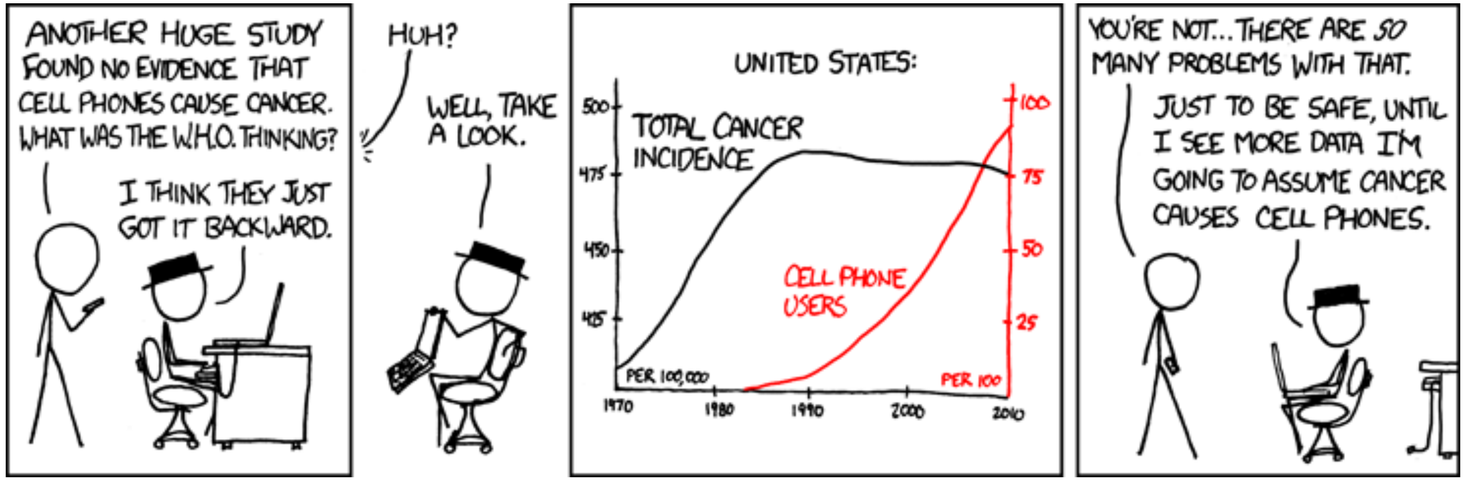
We know the causal graph, how do we estimate causal effects?

What happens if the graph is unknown?

Cutting edge research

More details here: https://canvas.uva.nl/courses/39307/pages/course-schedule-tentative

Questions?



https://xkcd.com/925/