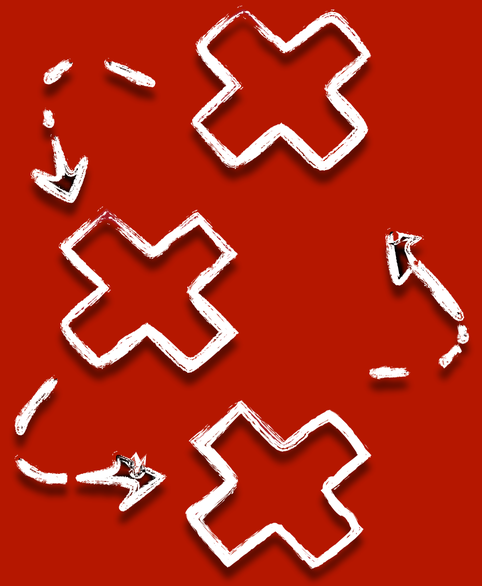


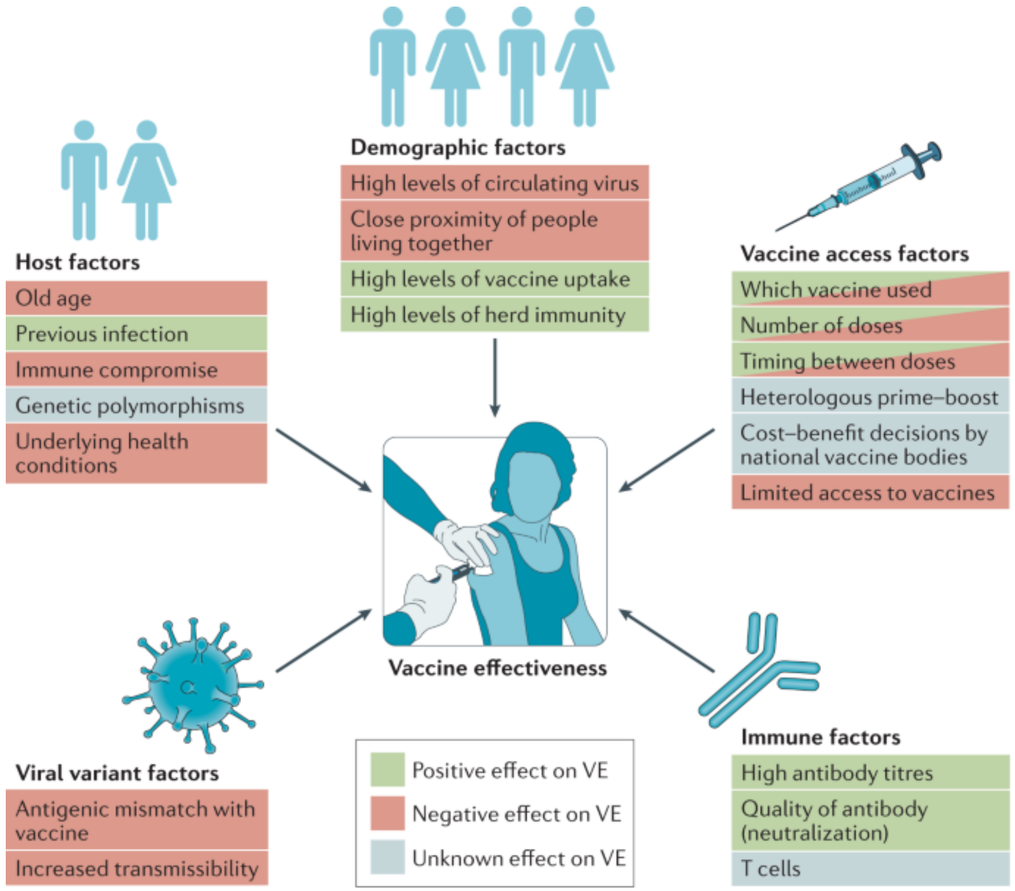
Deep learning 2: Causality & DL

1.1: Introduction

Lecturer: Sara Magliacane

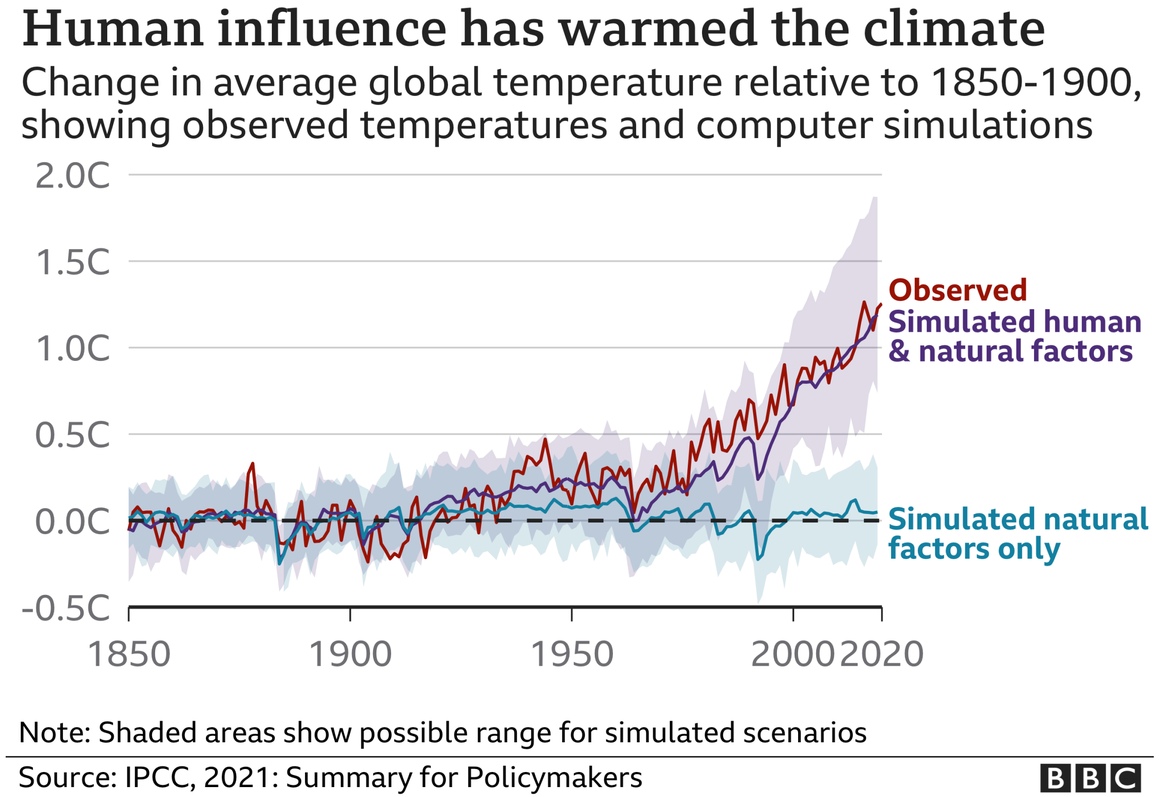
UvA - Spring 2022

Causal questions are ubiquitous: healthcare



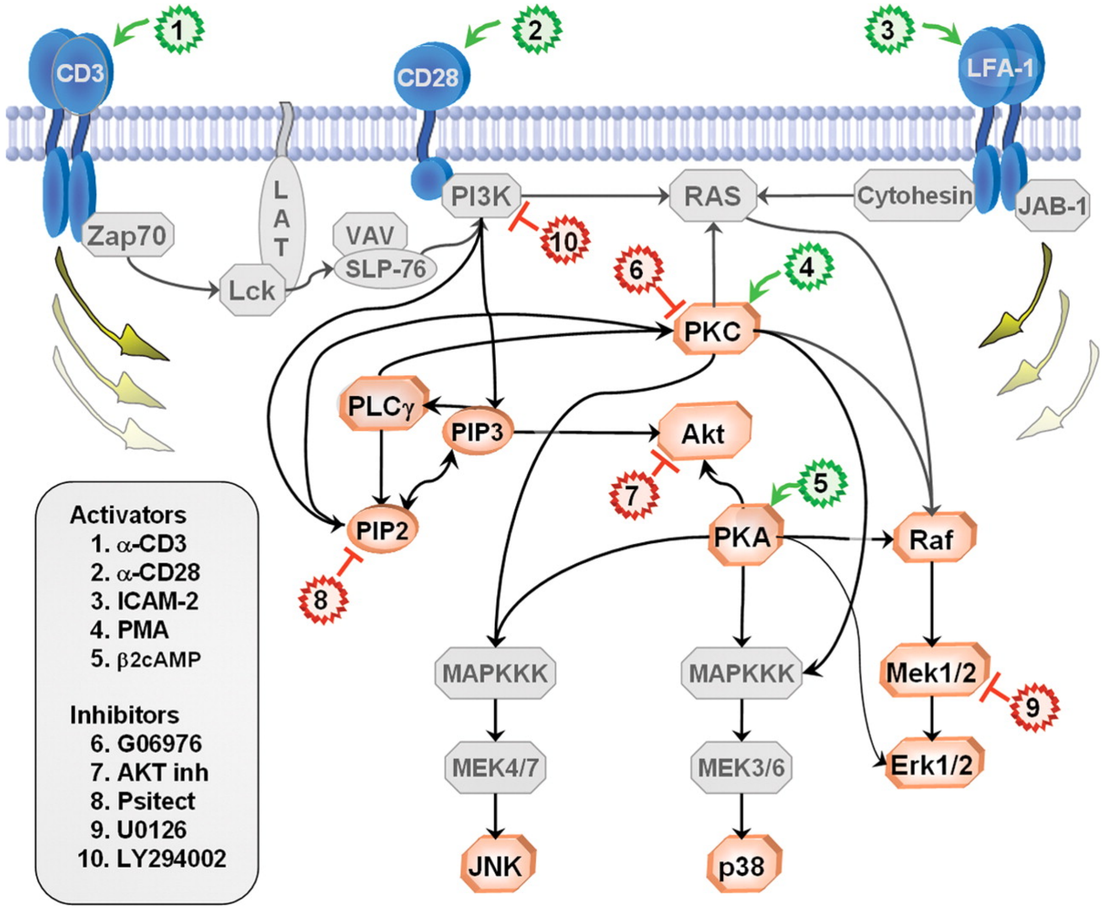
https://www.nature.com/articles/s41577-021-00592-1

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? Direction? Alternative explanations?

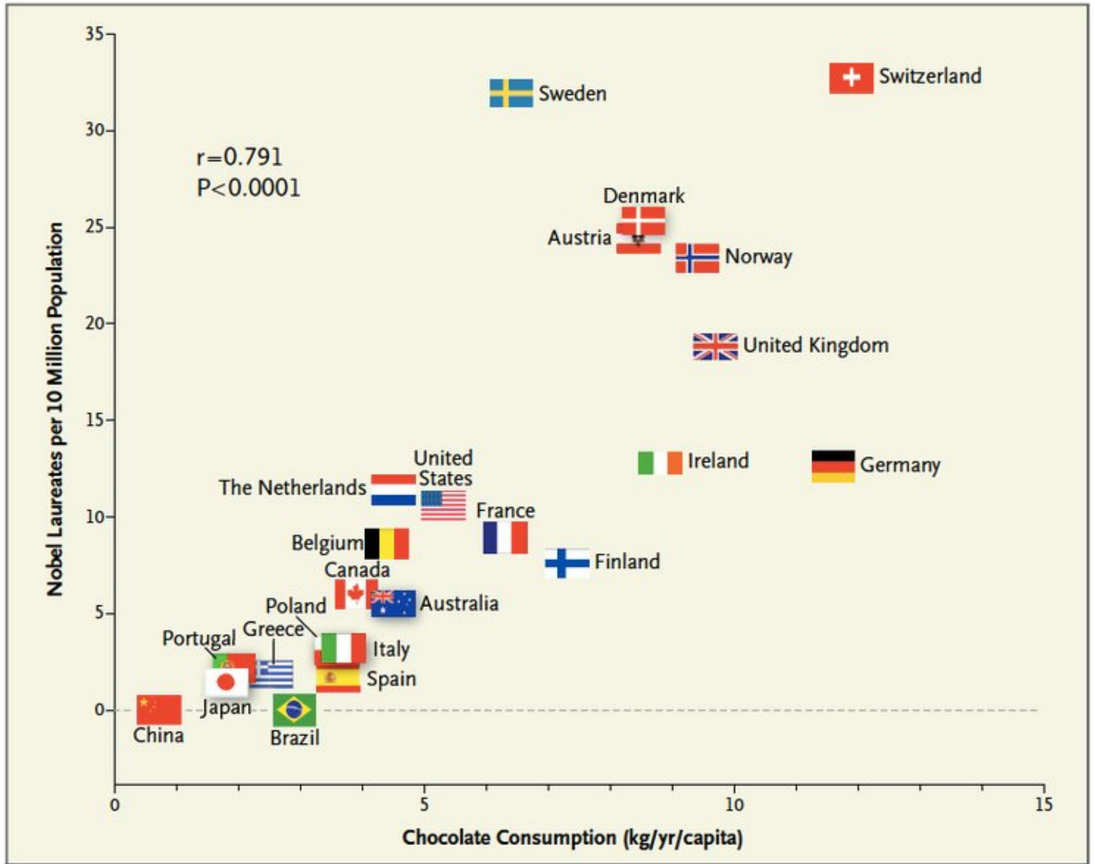


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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

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

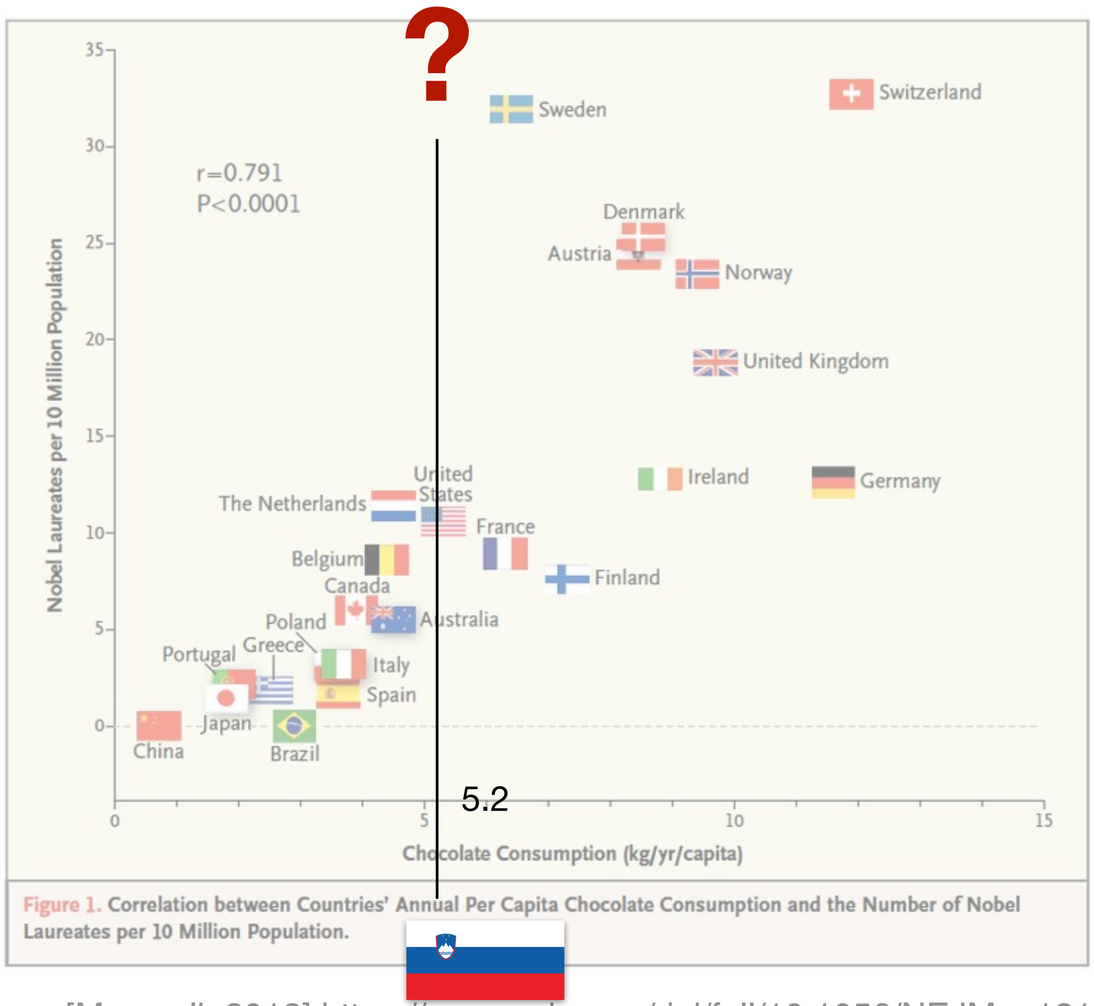


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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

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

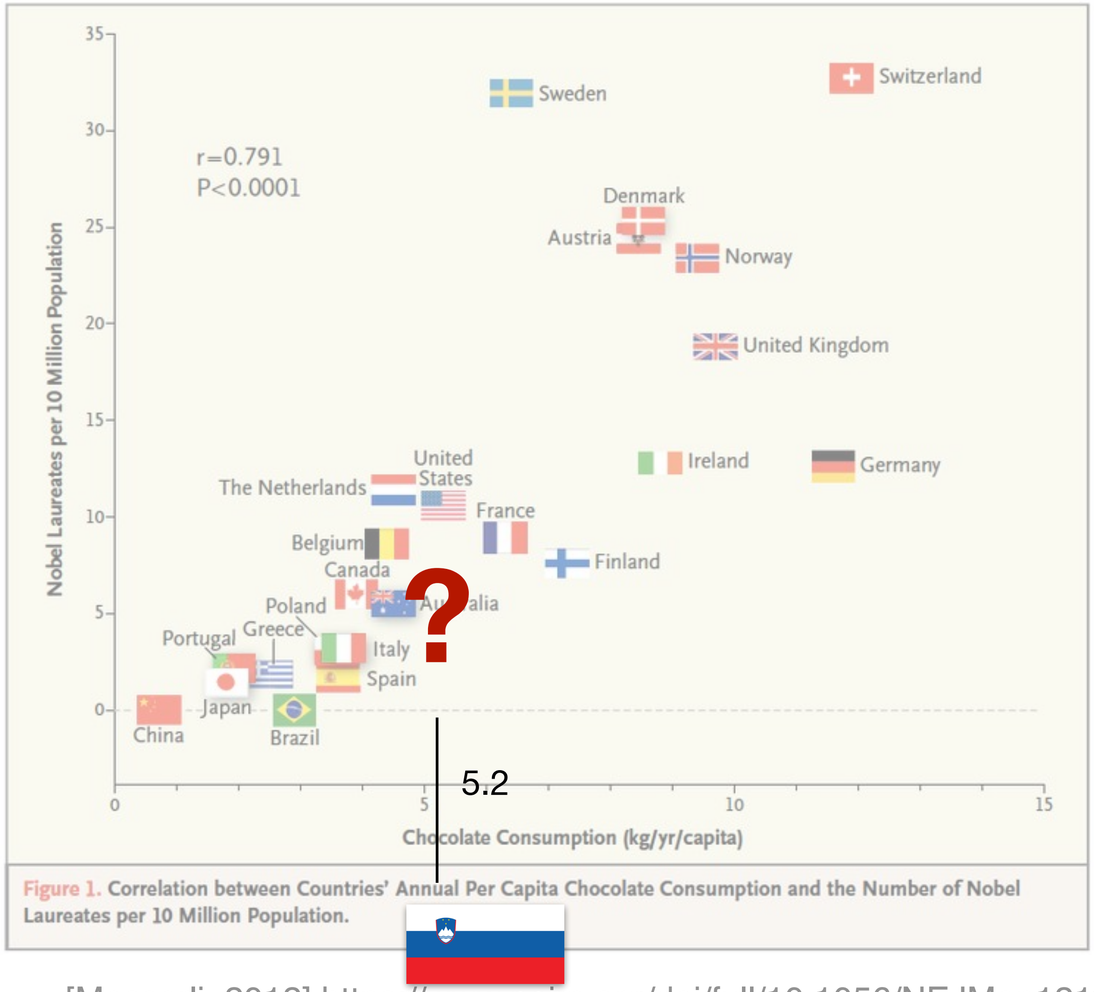


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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

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

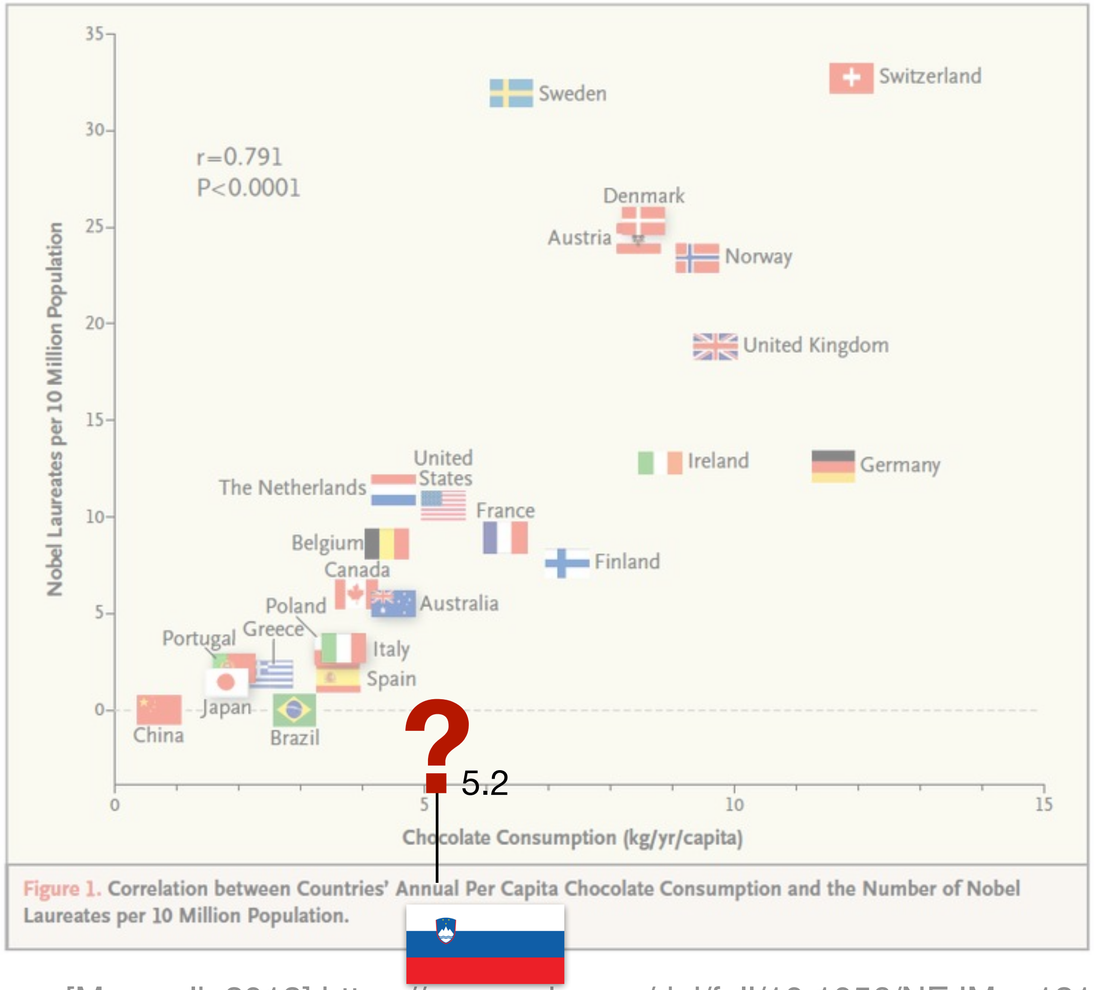
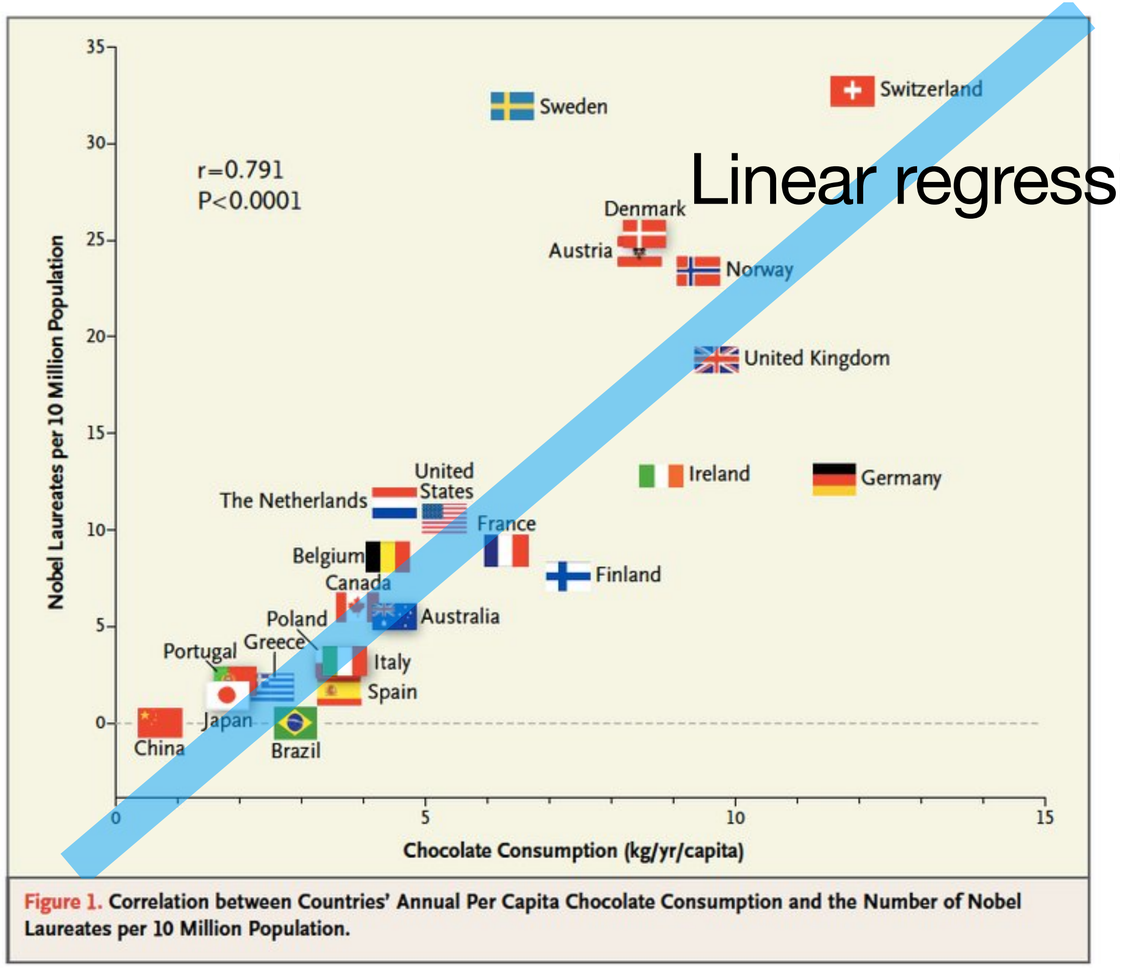


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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

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



Linear regression, NNs, …

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?

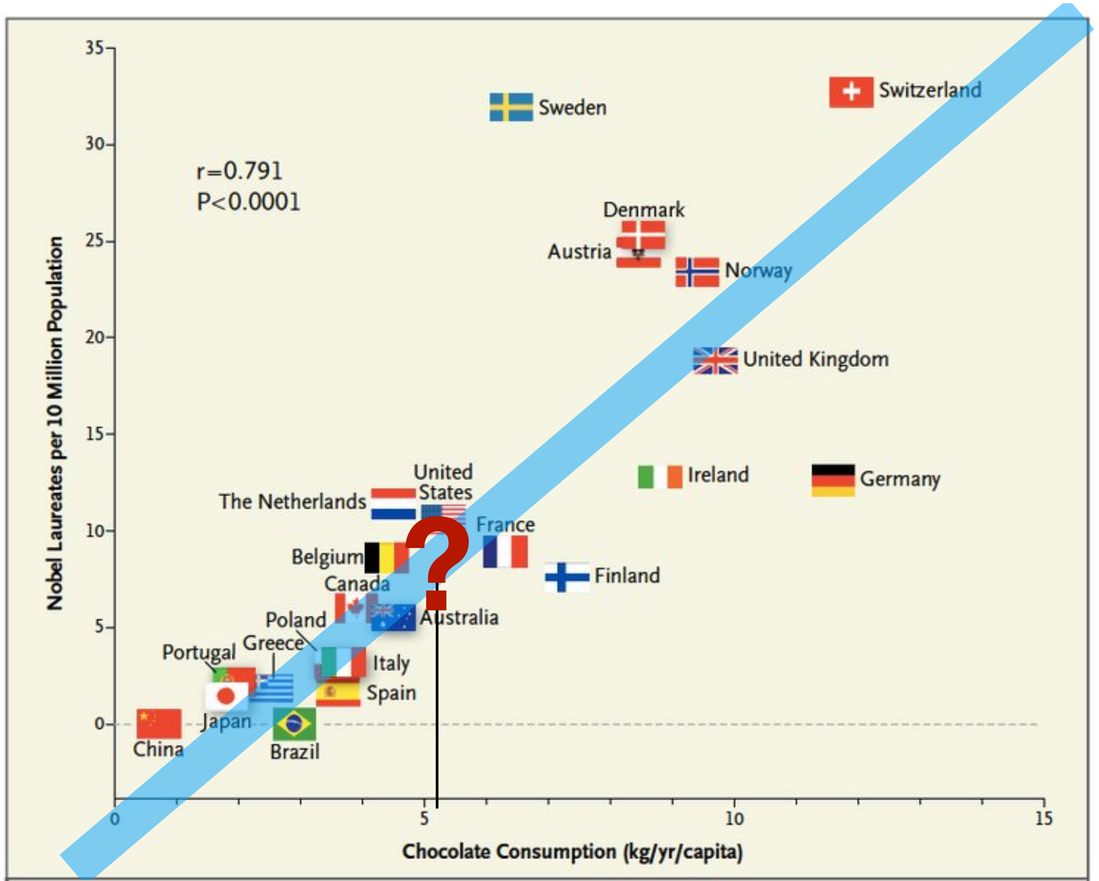


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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

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

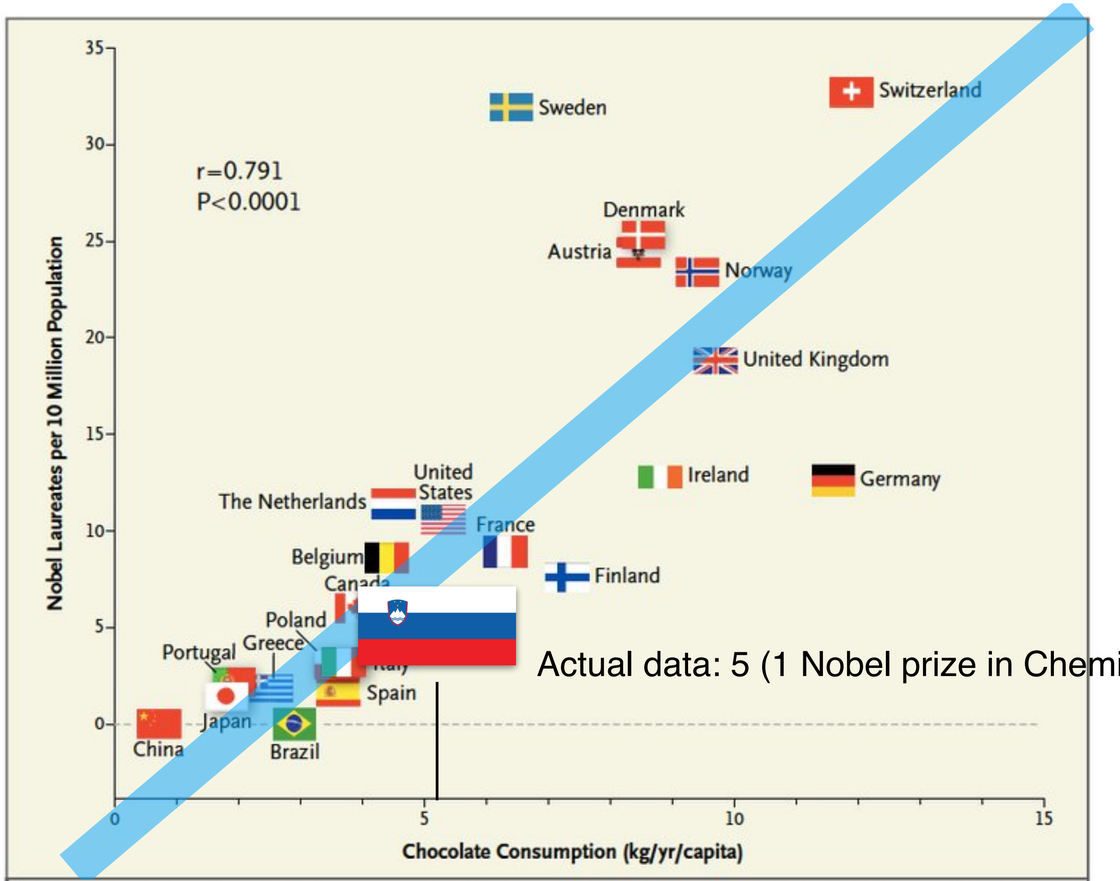


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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

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

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

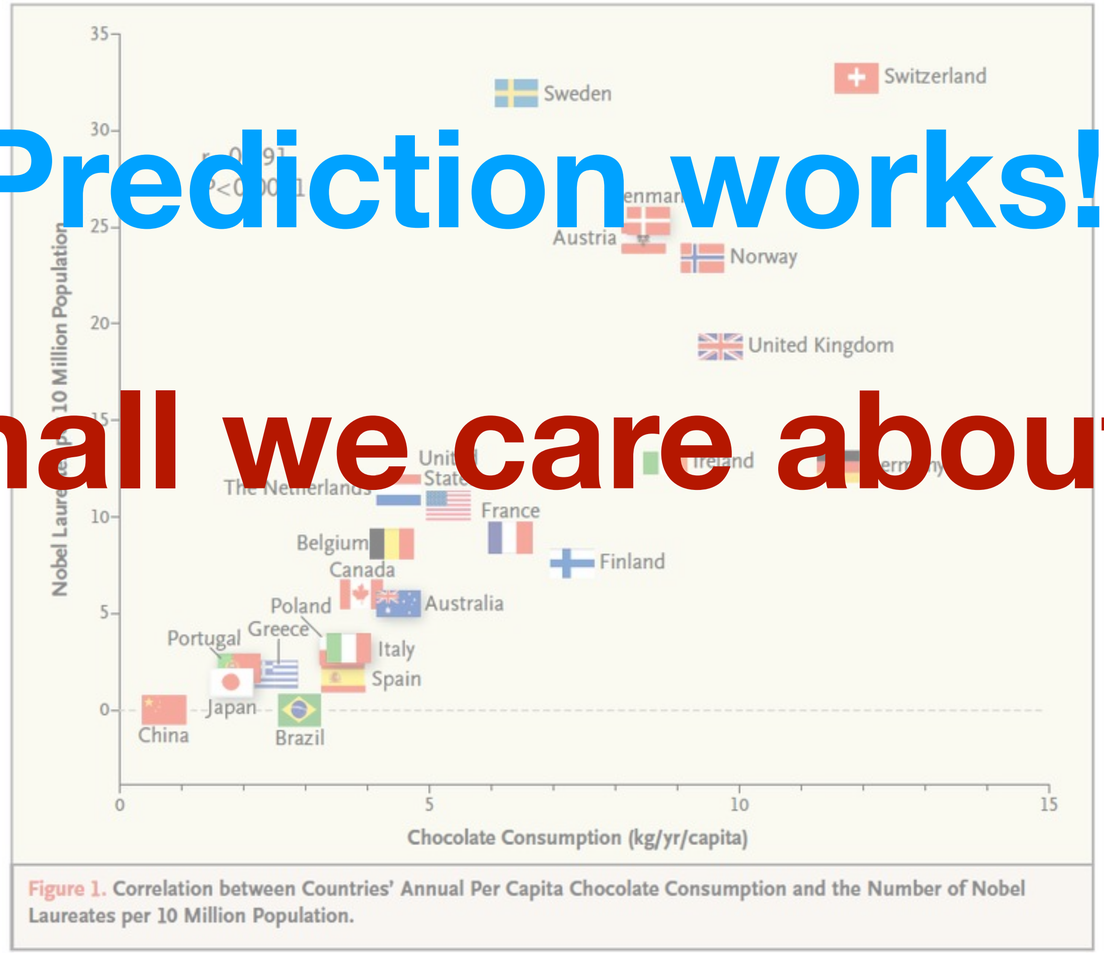


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Prediction works!!!

Why shall we care about causality?

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

Helping the Dutch government with decision making

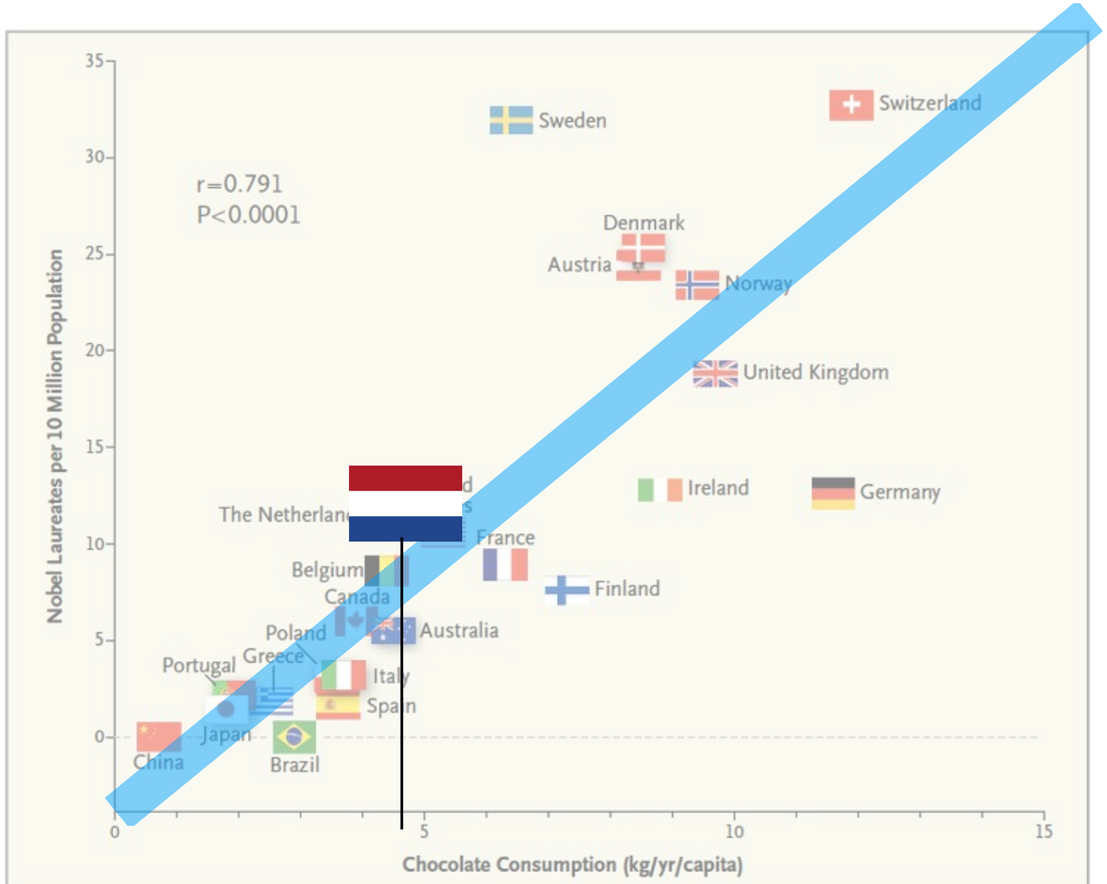


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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

Helping the Dutch government with decision making

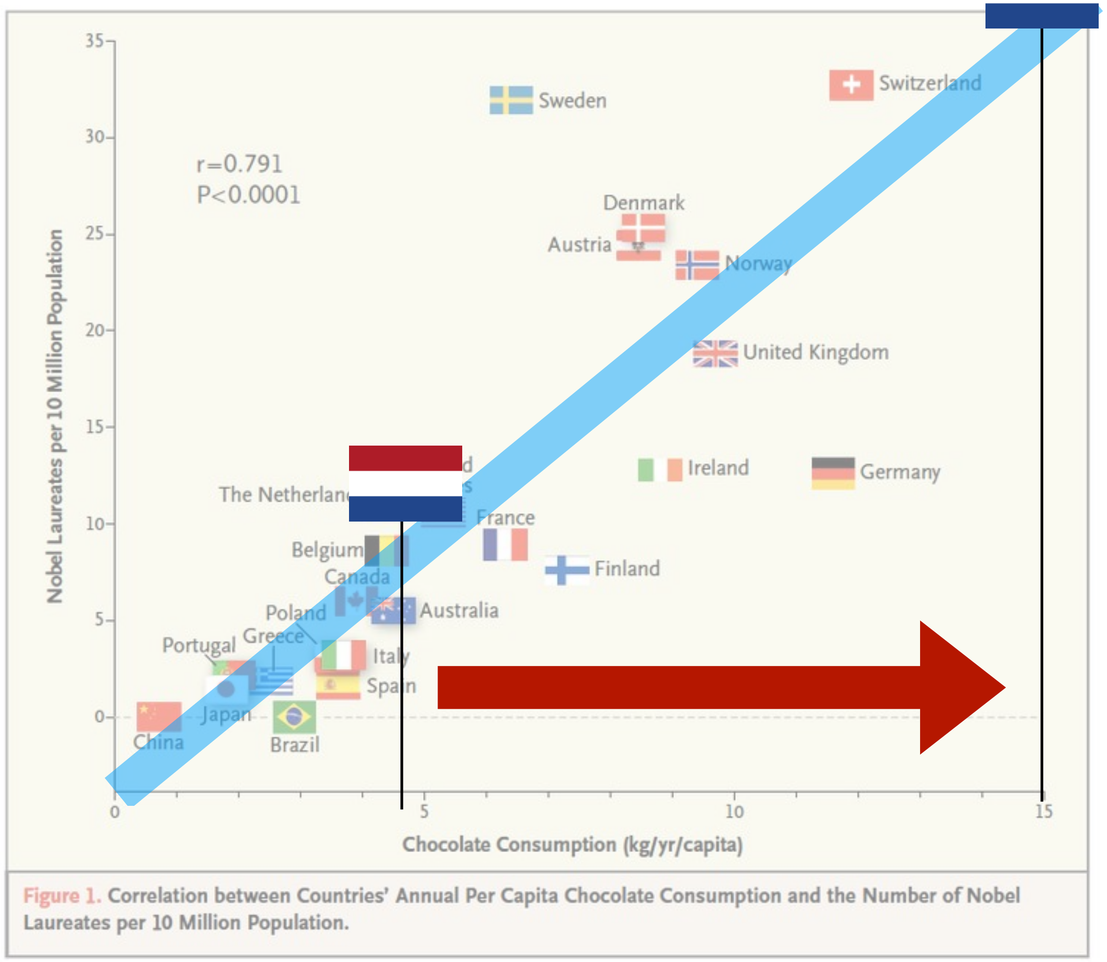


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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

Helping the Dutch government with decision making

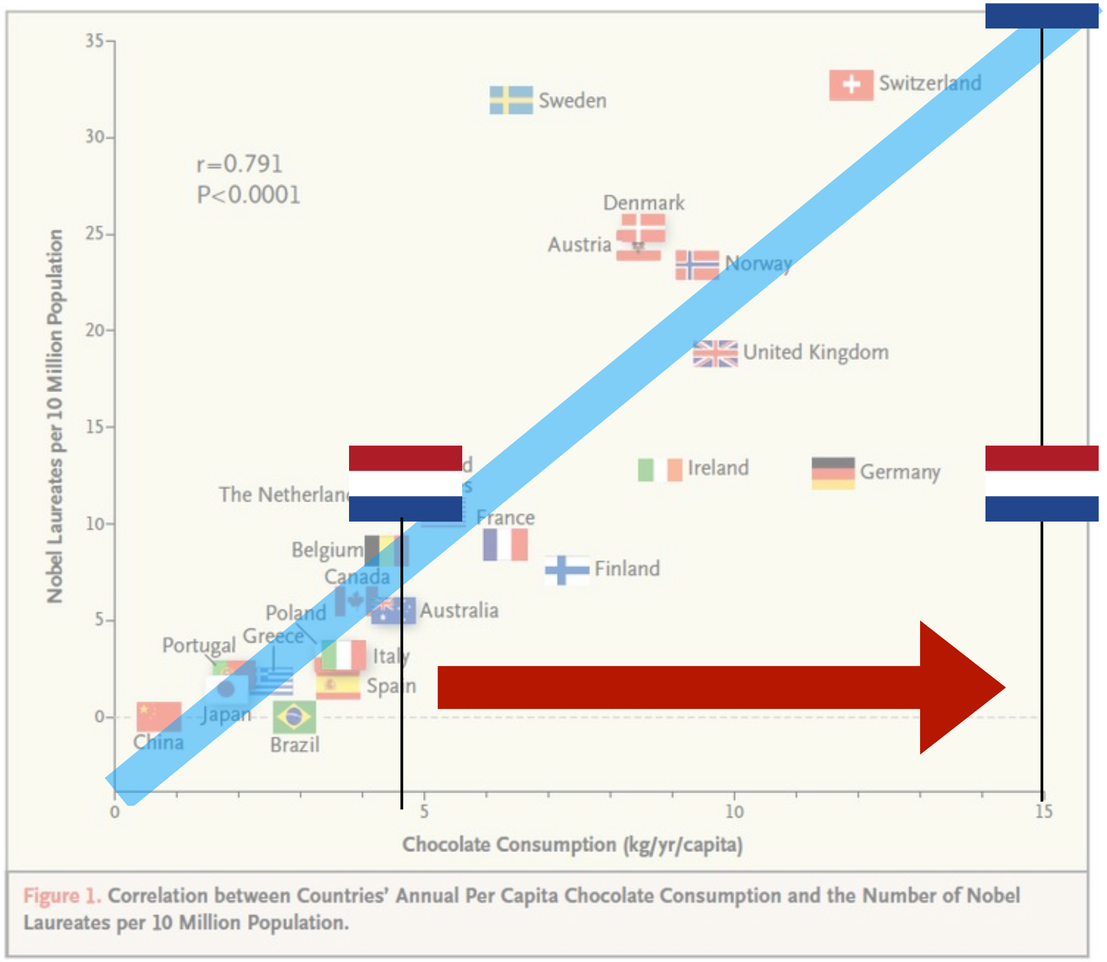


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

?

?

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

Helping the Dutch government with decision making

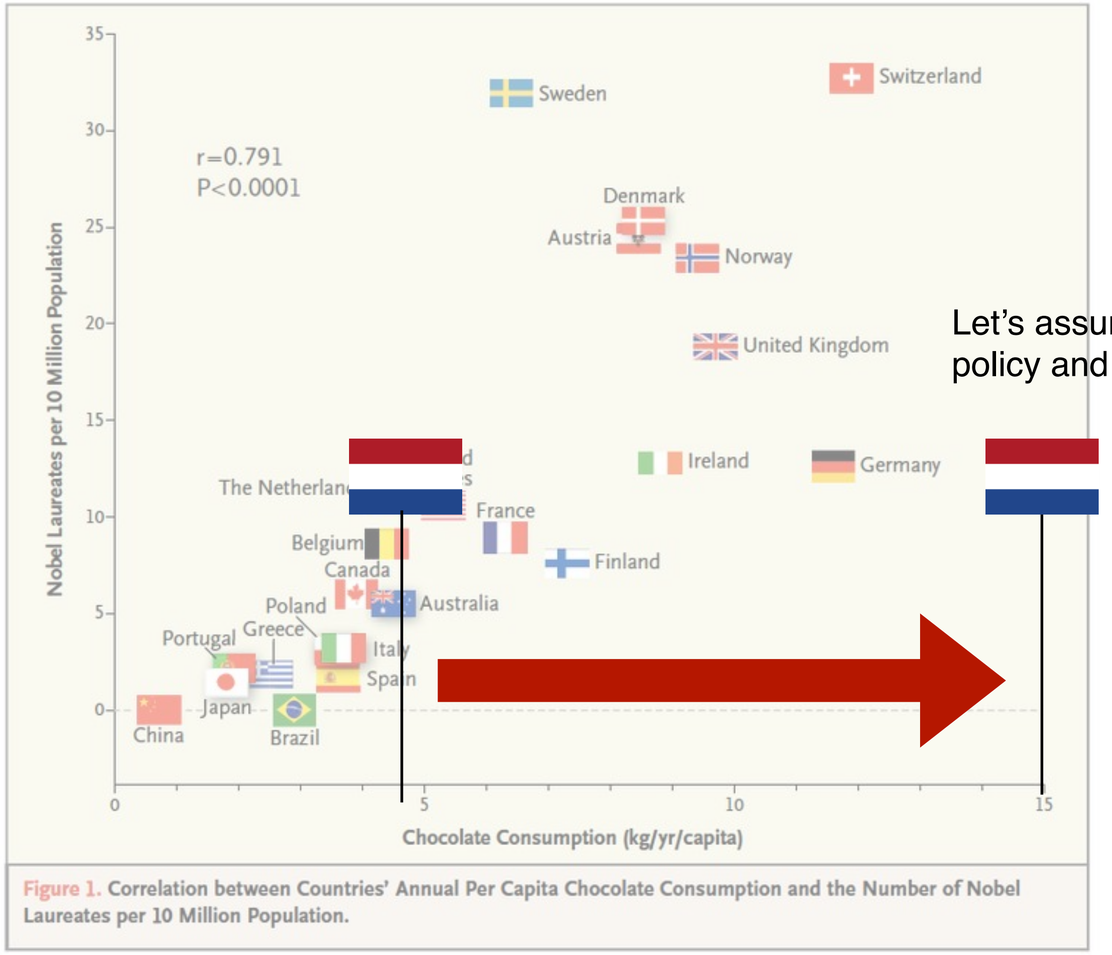


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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?

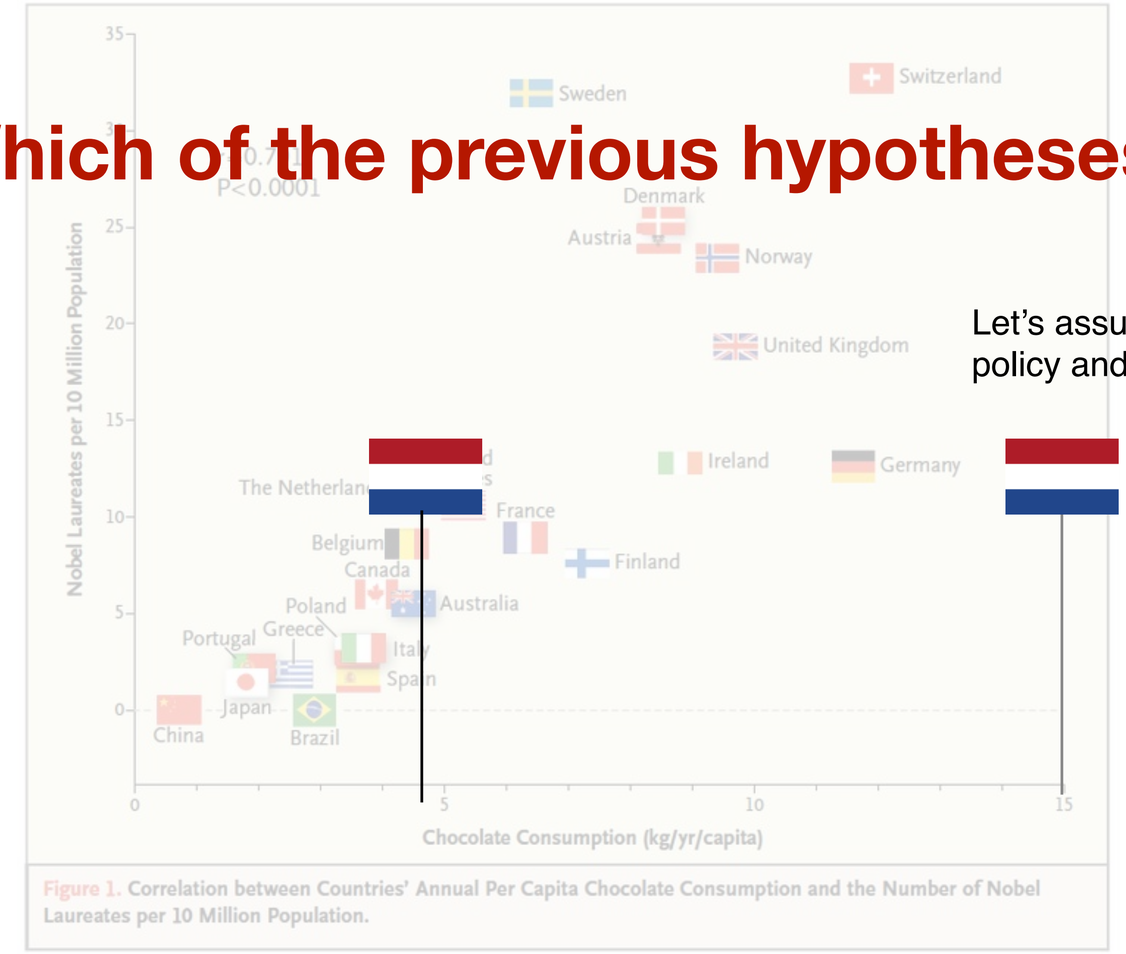


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Which of the previous hypotheses is still valid?

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 - Optional quiz on Canvas

[Optional] Correlation or causation?

Instructions

Given only the data plotted in this graph, which of the following hypotheses is plausible (without using any domain knowledge?)

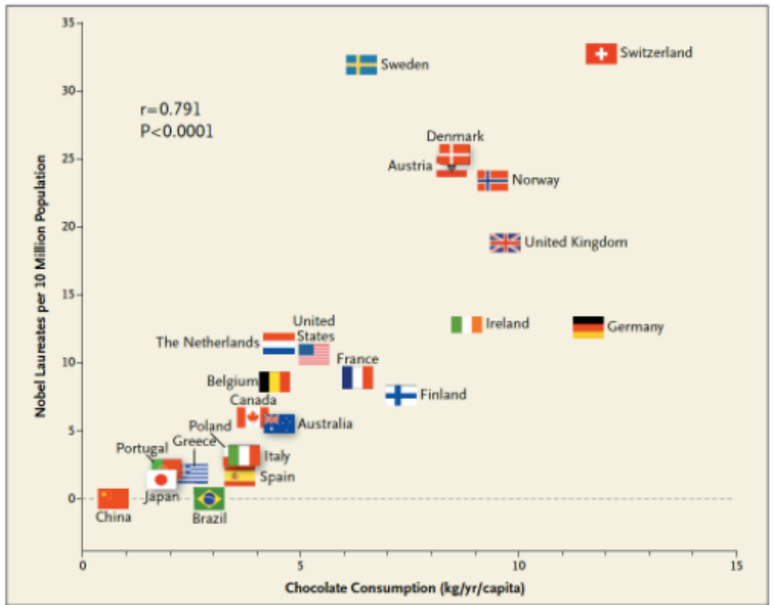
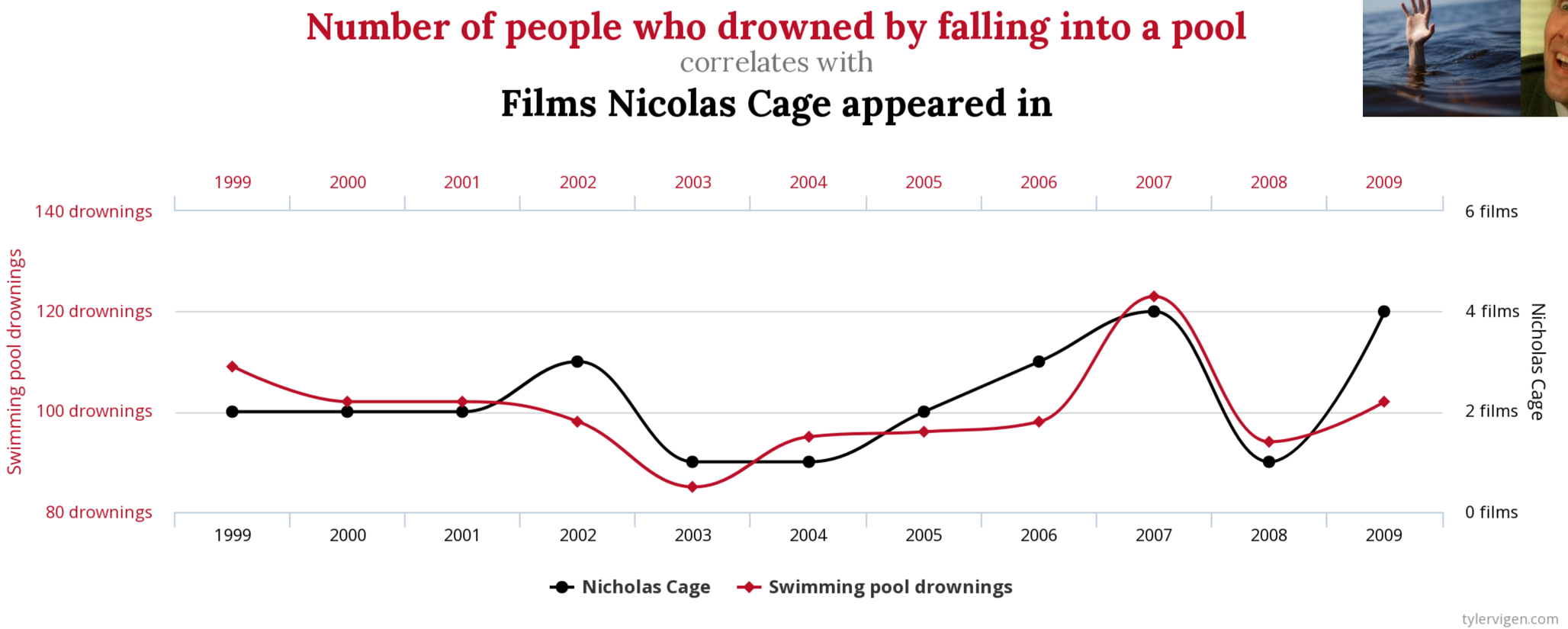


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Correlation or causation? Direction? Alternative explanations?



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?

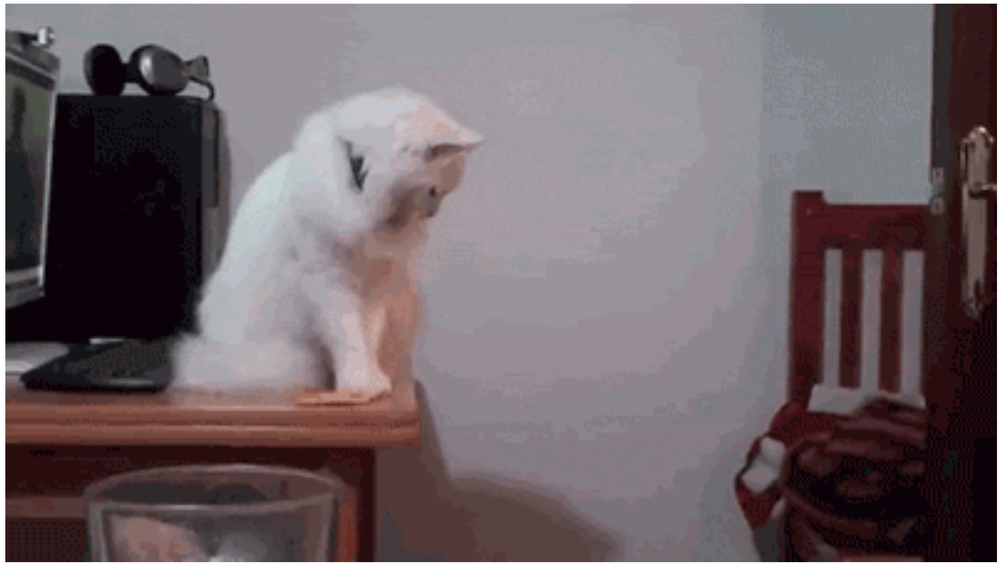
Causal Hierarchy [Pearl 2009, 2018]

Most ML

Causality

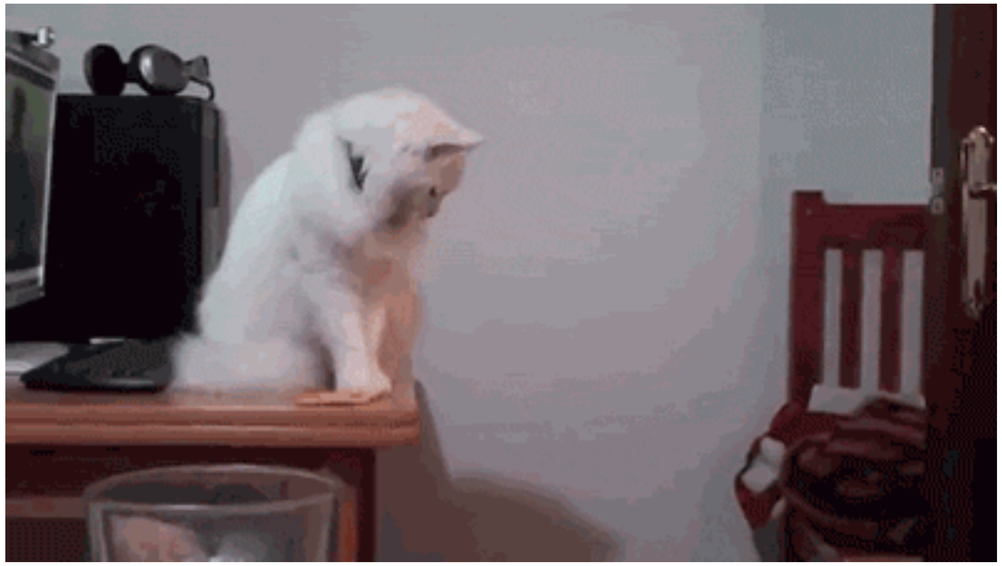
Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if?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



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

Intuitive definition: A variable A causes another variable B, if changing/ intervening upon variable A , changes (the distribution of) B

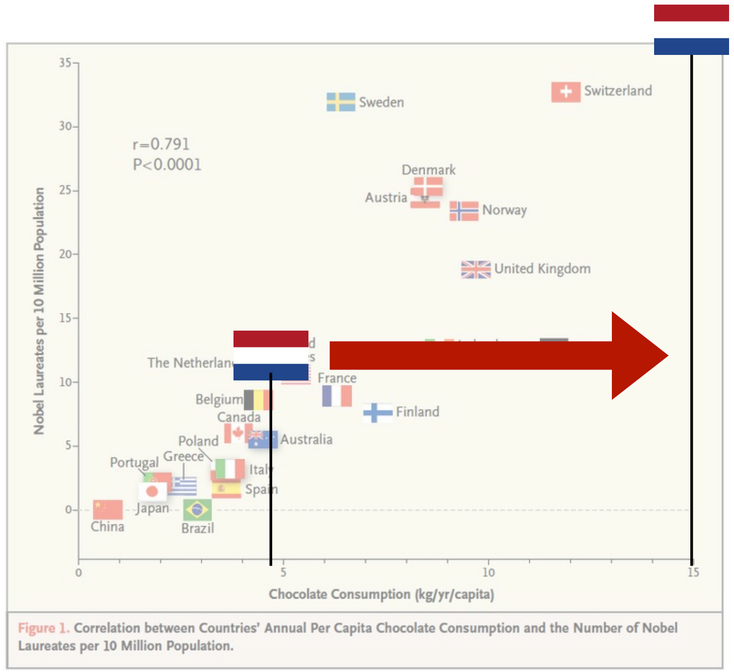


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

In an alternative universe:

NL eats more chocolate => more Nobel

A working definition

Intuitive definition: A variable A causes another variable B, if changing/ intervening upon variable A , changes (the distribution of) B

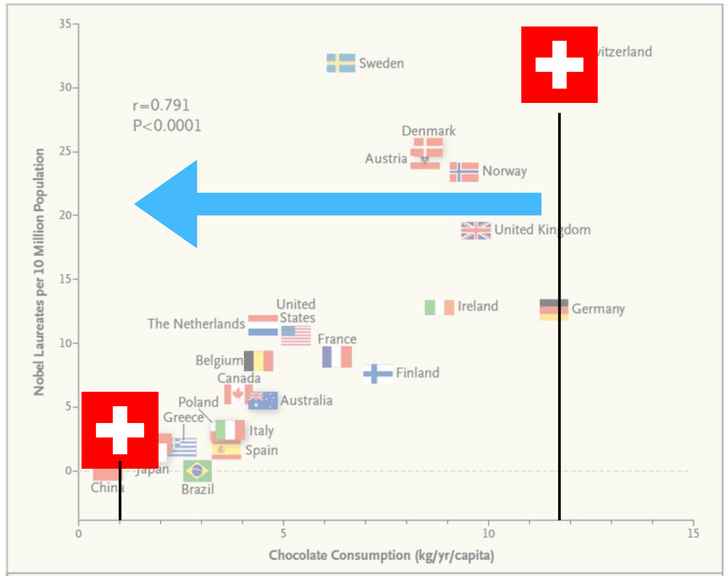


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

In an alternative universe:

CH eats less chocolate => less Nobel

A working definition

Intuitive definition: A variable A causes another variable B, if changing/ intervening upon variable A , changes (the distribution of) B

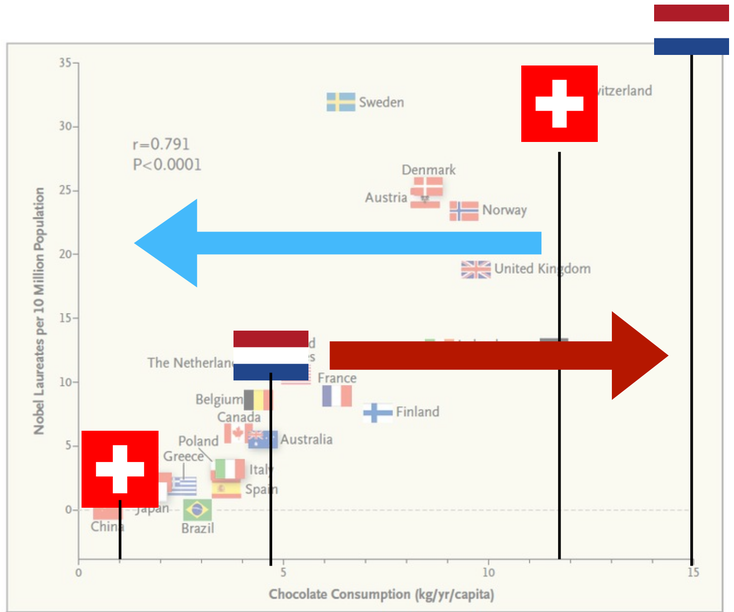


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

In an alternative universe:

NL eats more chocolate => more Nobel

CH eats less chocolate => less Nobel

… and similarly for other countries

A working definition

Intuitive definition: A variable A causes another variable B, if changing/ intervening upon variable A , changes (the distribution of) B

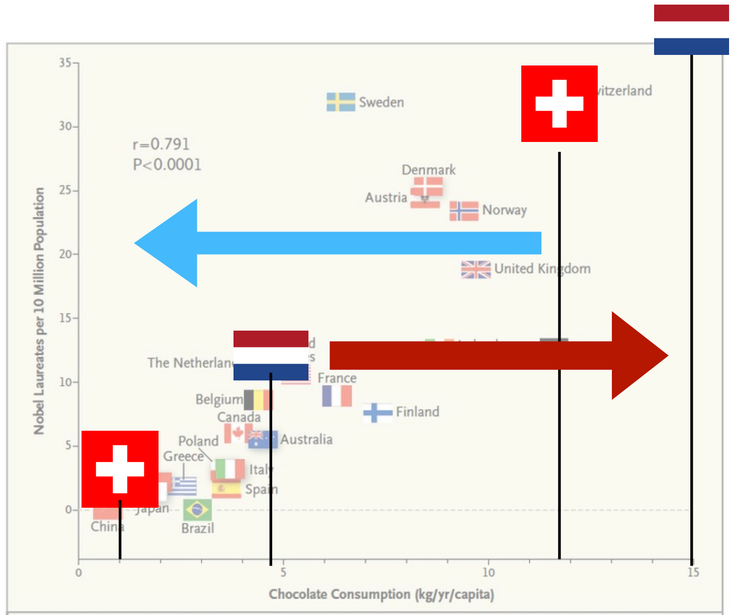


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

In an alternative universe:

NL eats more chocolate => more Nobel

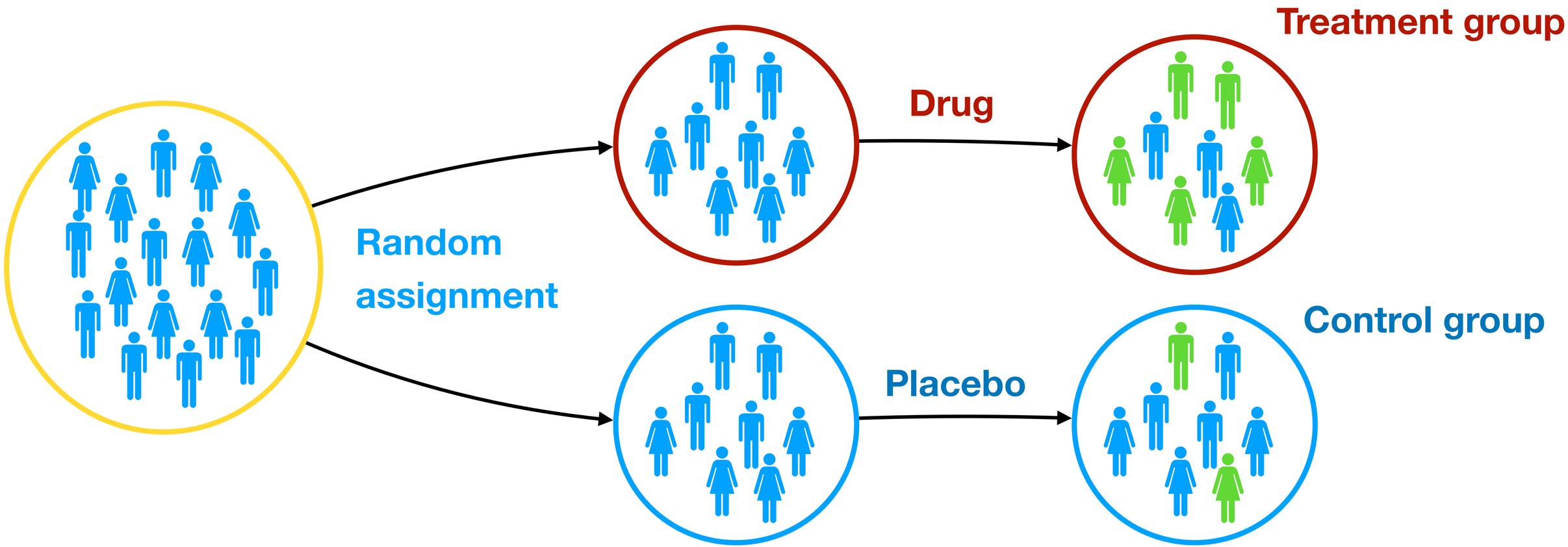
CH eats less chocolate => less Nobel

… and similarly for other countries

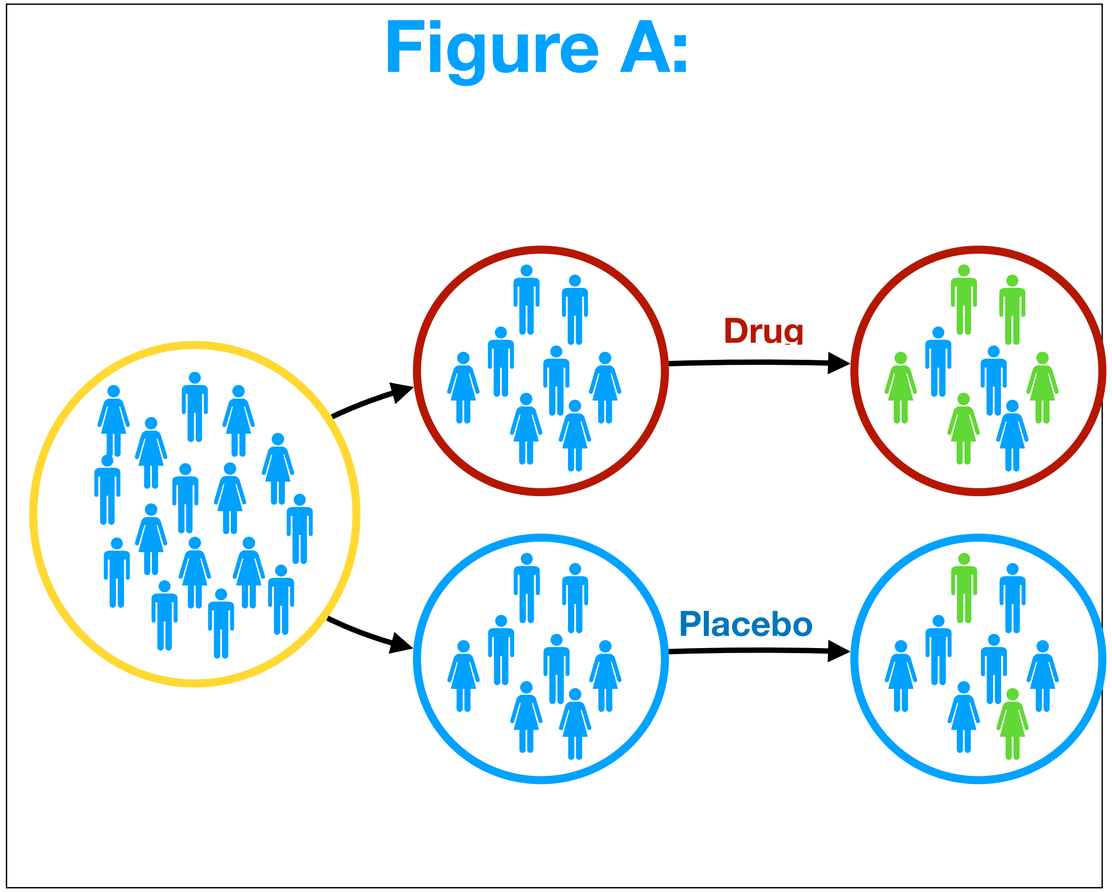
Chocolate => Nobel prizes

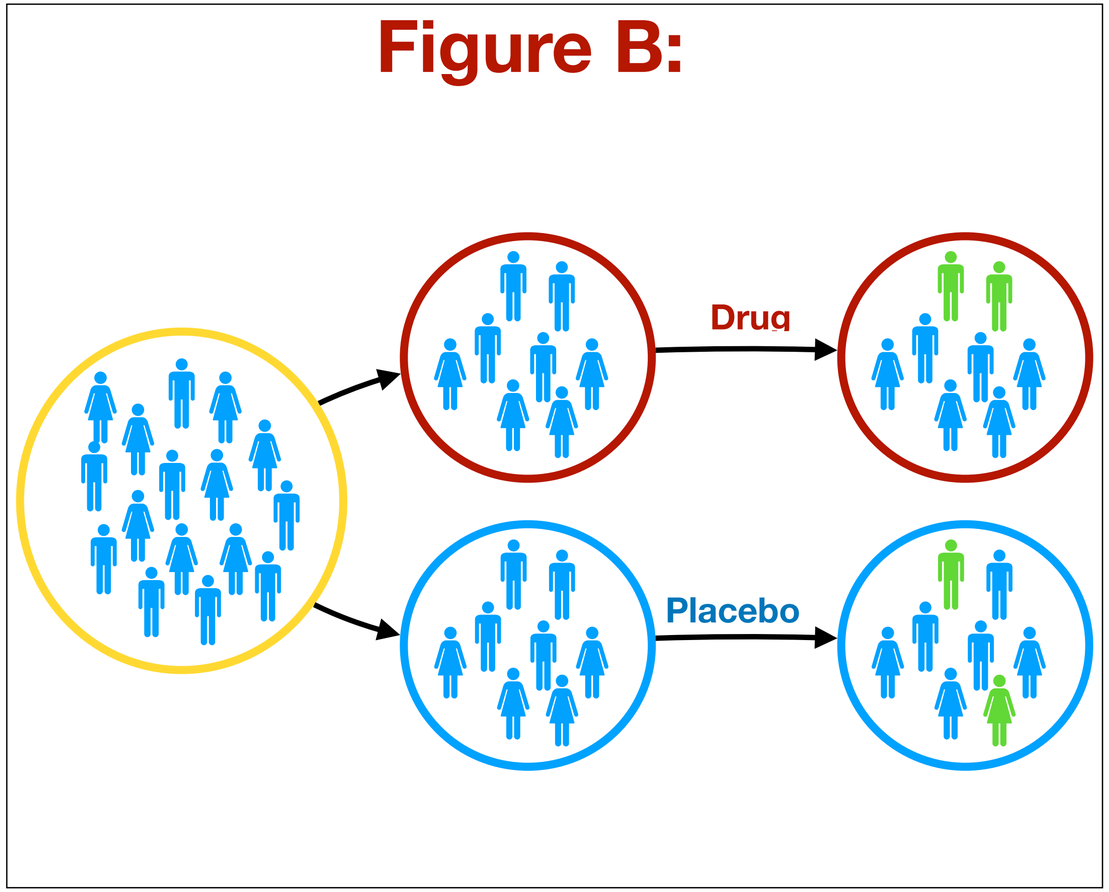
Based on experimental data

Gold standard of experiments: Randomized Controlled Trials (RCTs)

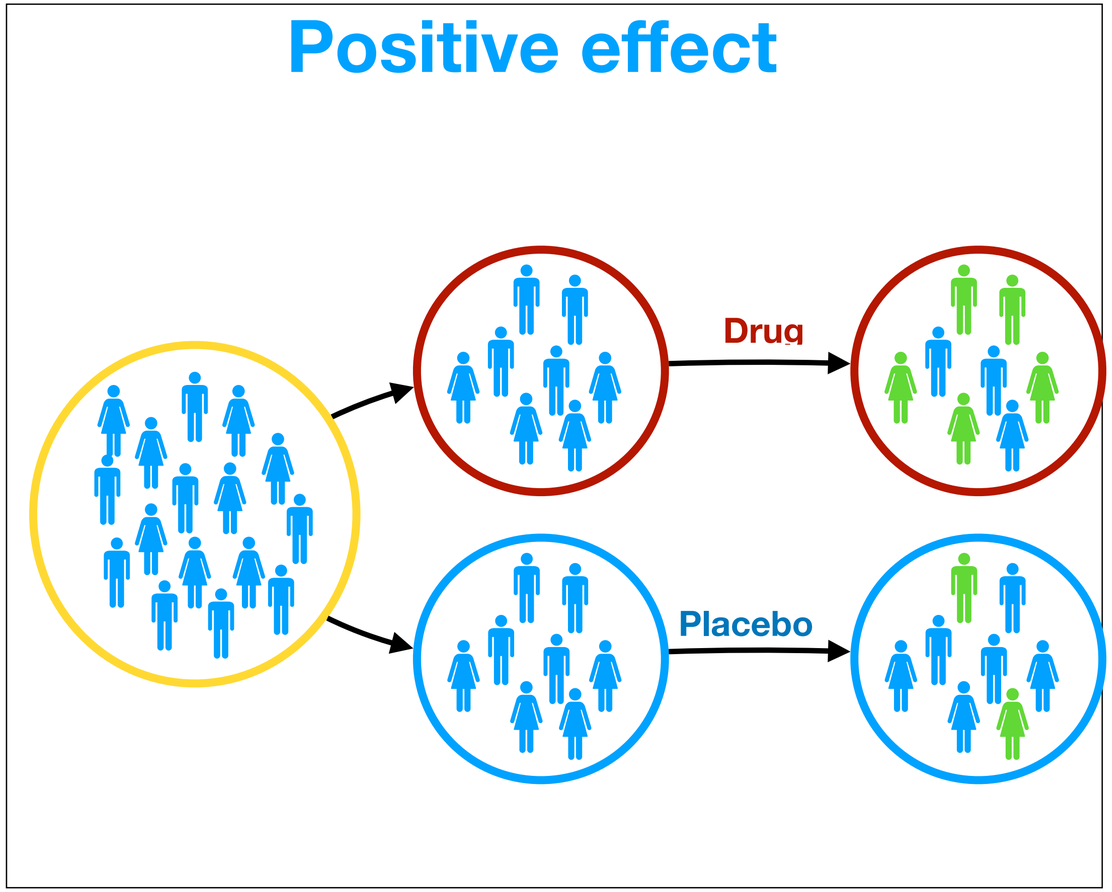


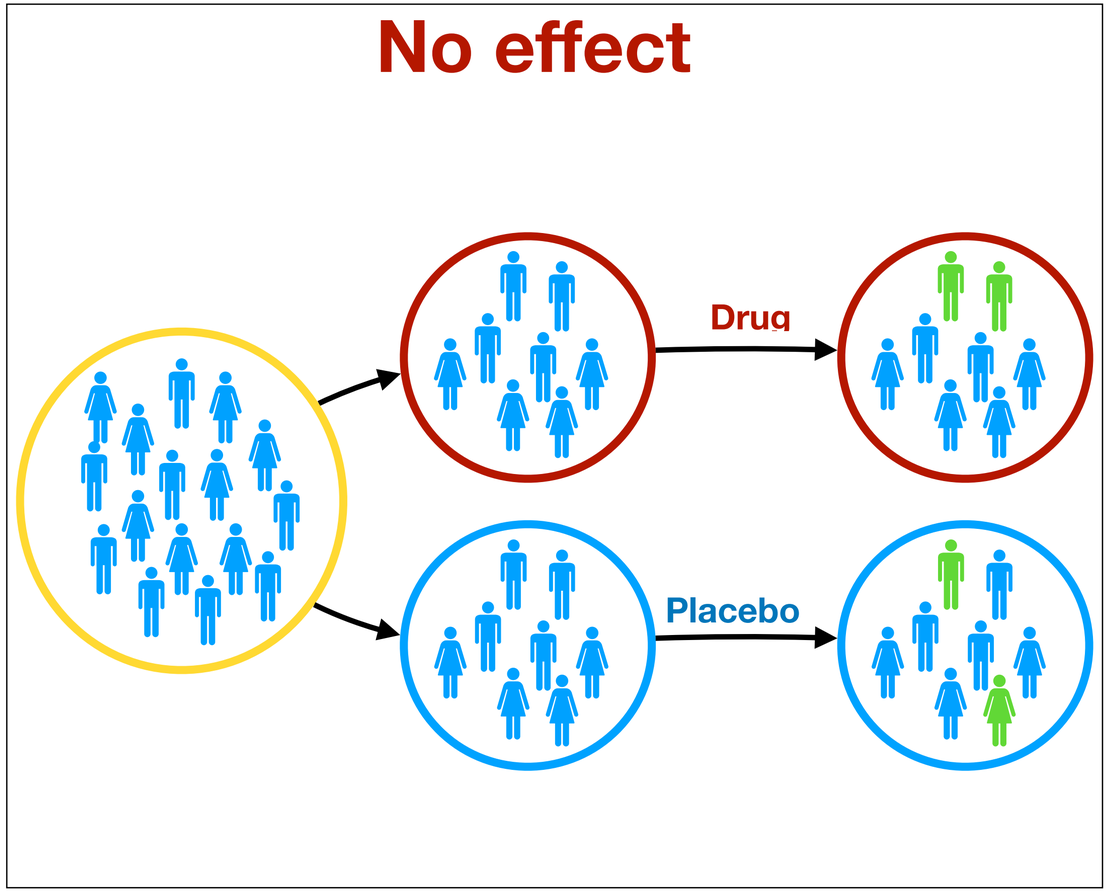
RCT: Spot the difference?





RCT: Spot the difference?





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

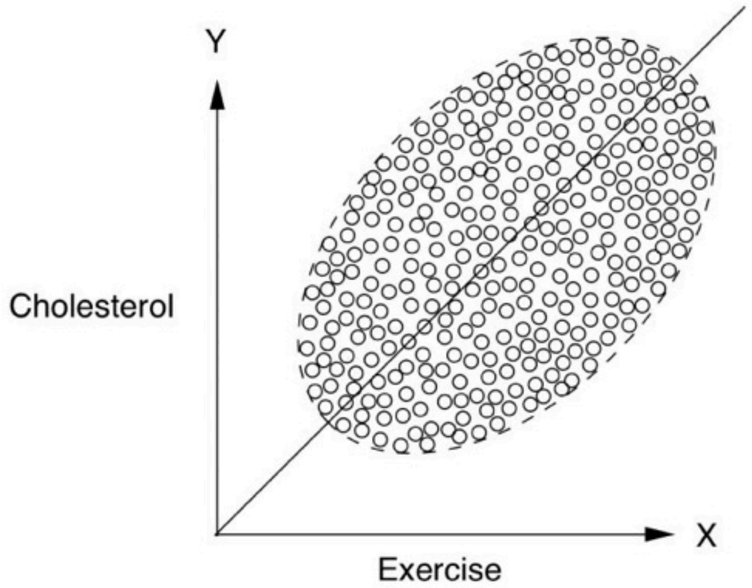
Experimental/interventional vs observational data

- **Experimental data:** We perform an experiment by changing variable A **(intervention)** and check the effects on the other variables
  - For example, we can encourage one country to eat more chocolate
  - Or to test if a drug works, we set up an RCT

- Sometimes these experiments are **expensive**, **unfeasible** or **unethical** (e.g. consider forcing people to smoke to see what effects it has)
  - In this case, we use **observational (non-experimental)** data

What if we don't have an RCT?

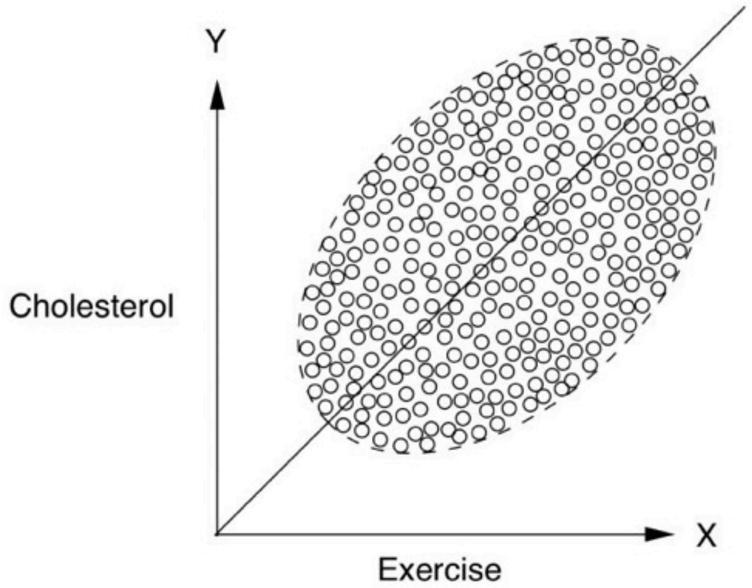
Let's assume we have **observational data** (e.g. data collected by hospitals)

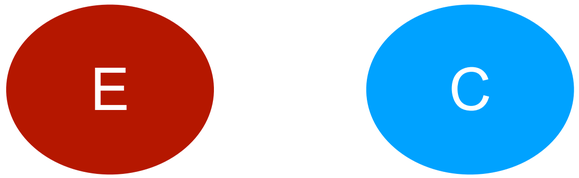


From the Book of Why [Pearl 2018]

What if we don't have an RCT?

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

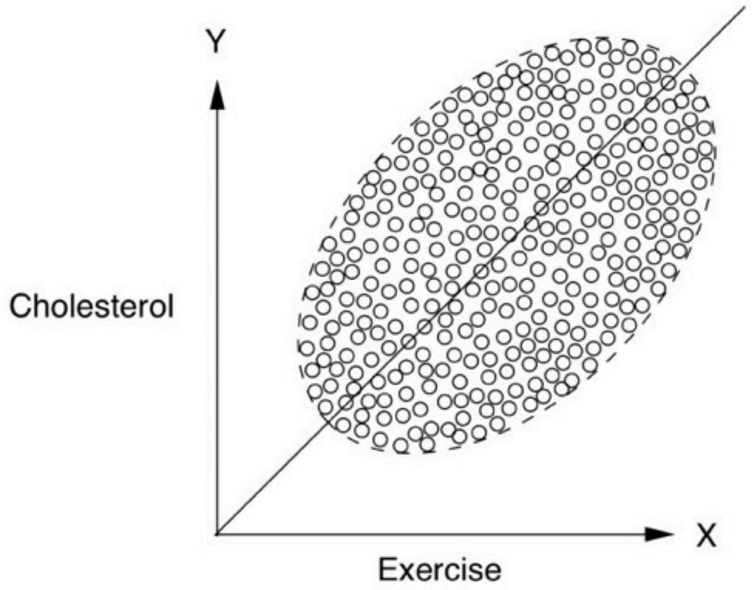


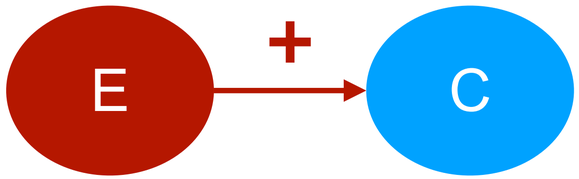


From the Book of Why [Pearl 2018]

What if we don't have an RCT?

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



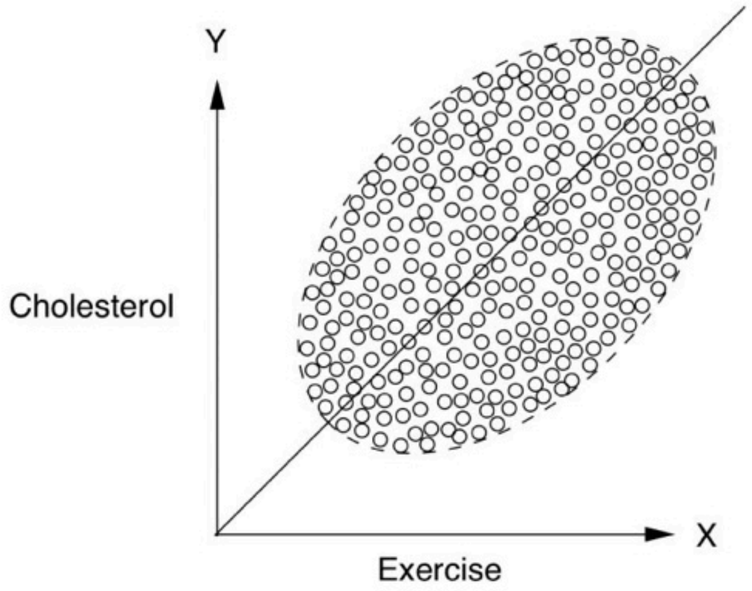


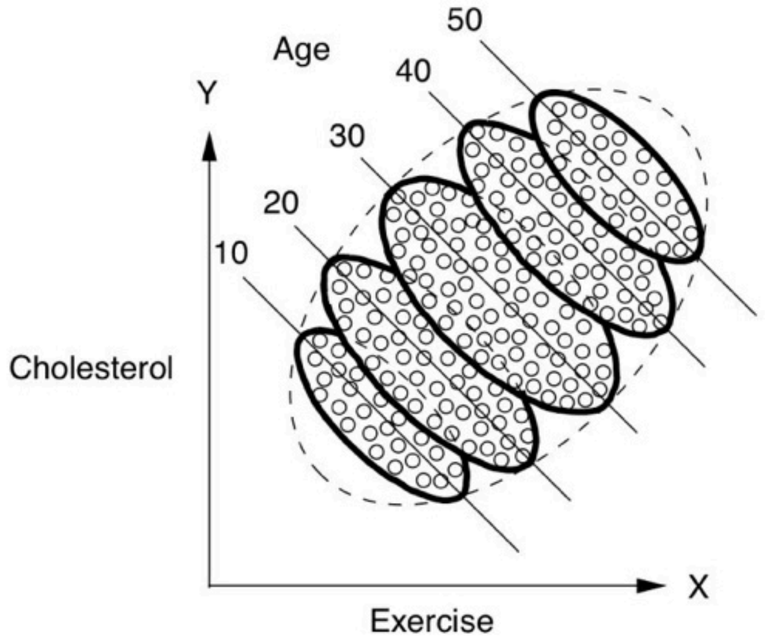
Exercise increases cholesterol

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)

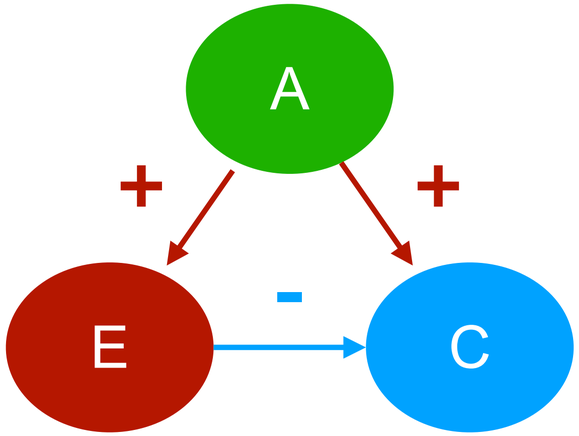




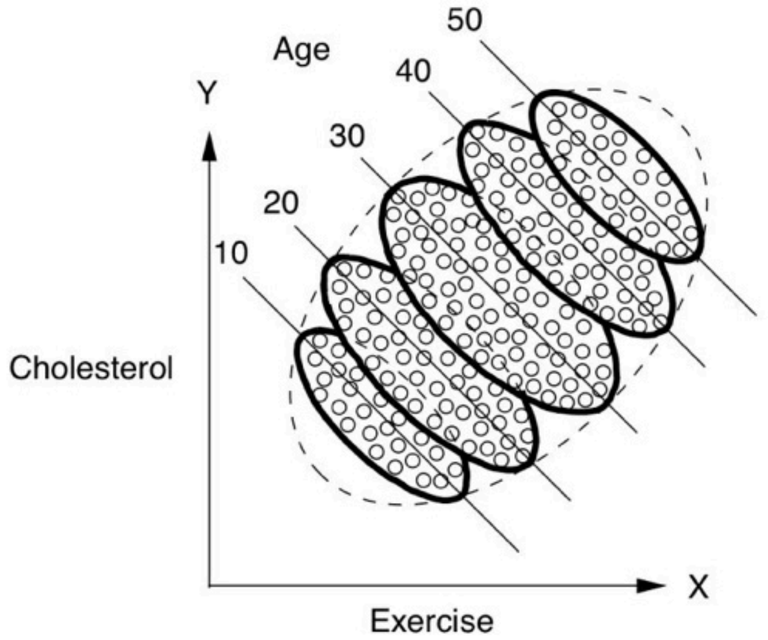
From the Book of Why [Pearl 2018]

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

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



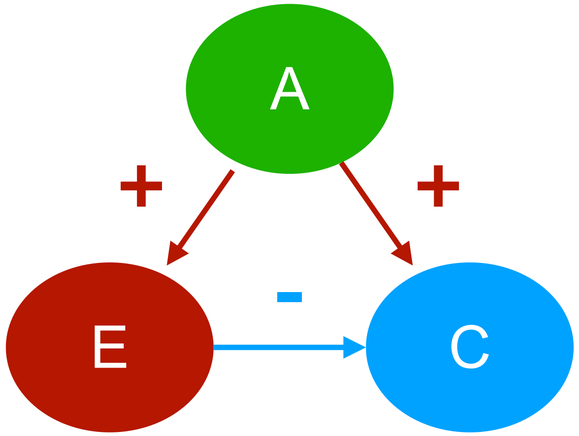
Exercise decreases cholesterol

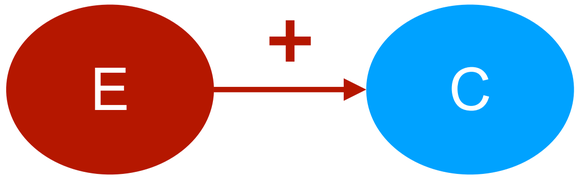


From the Book of Why [Pearl 2018]

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

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

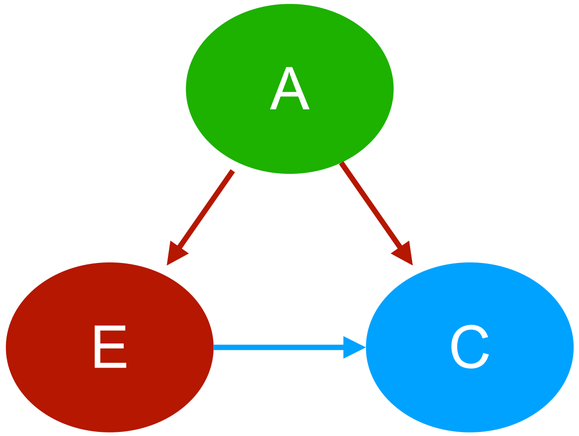




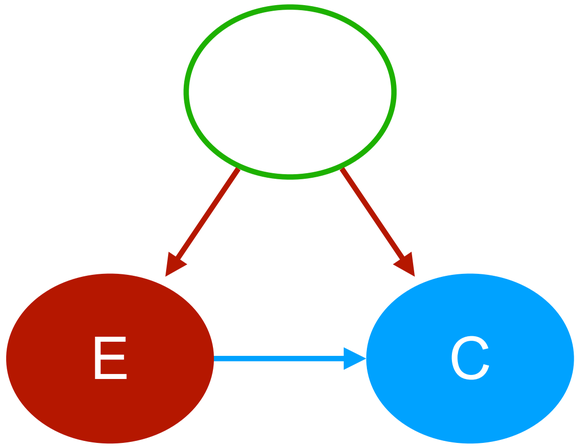
From the Book of Why [Pearl 2018]

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



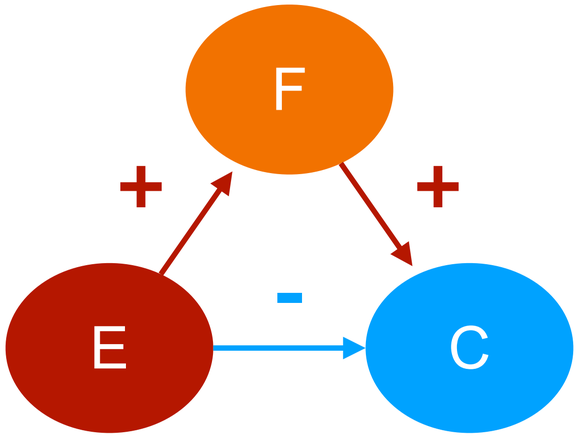
It confounds the effect

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

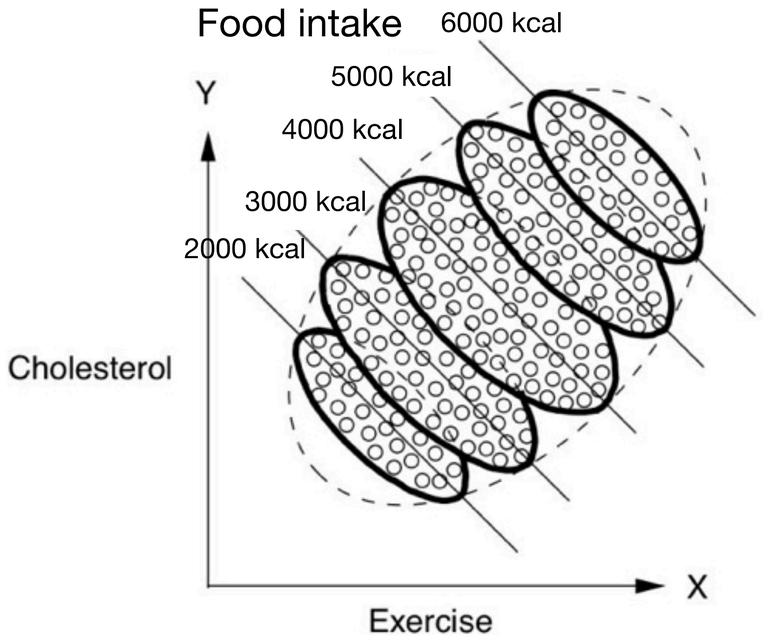
From the Book of Why [Pearl 2018]

Shall we always just control on everything?

In alternative universe:



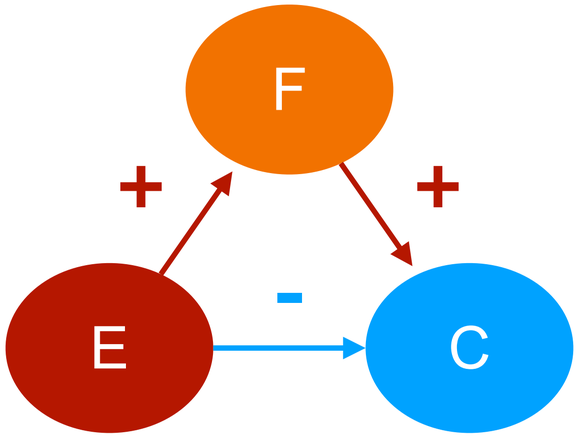
Exercise **increases** cholesterol



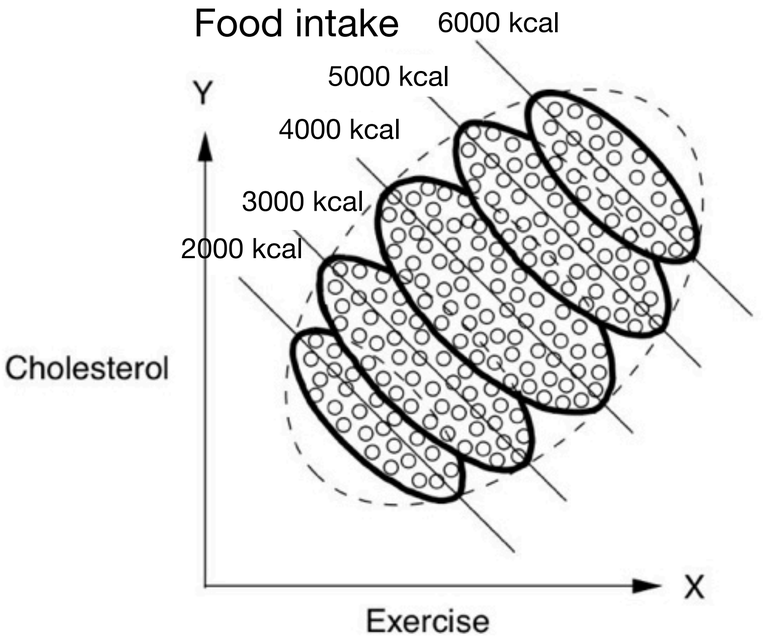
Edited based on the Book of Why [Pearl 2018]

Shall we always just control on everything?

In alternative universe:



Exercise **increases** cholesterol



If we know the true causal graph, we can select **covariates** to **adjust** for.

Edited based on the Book of Why [Pearl 2018]

Simpson’s paradox: Optional quiz on Canvas

- 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