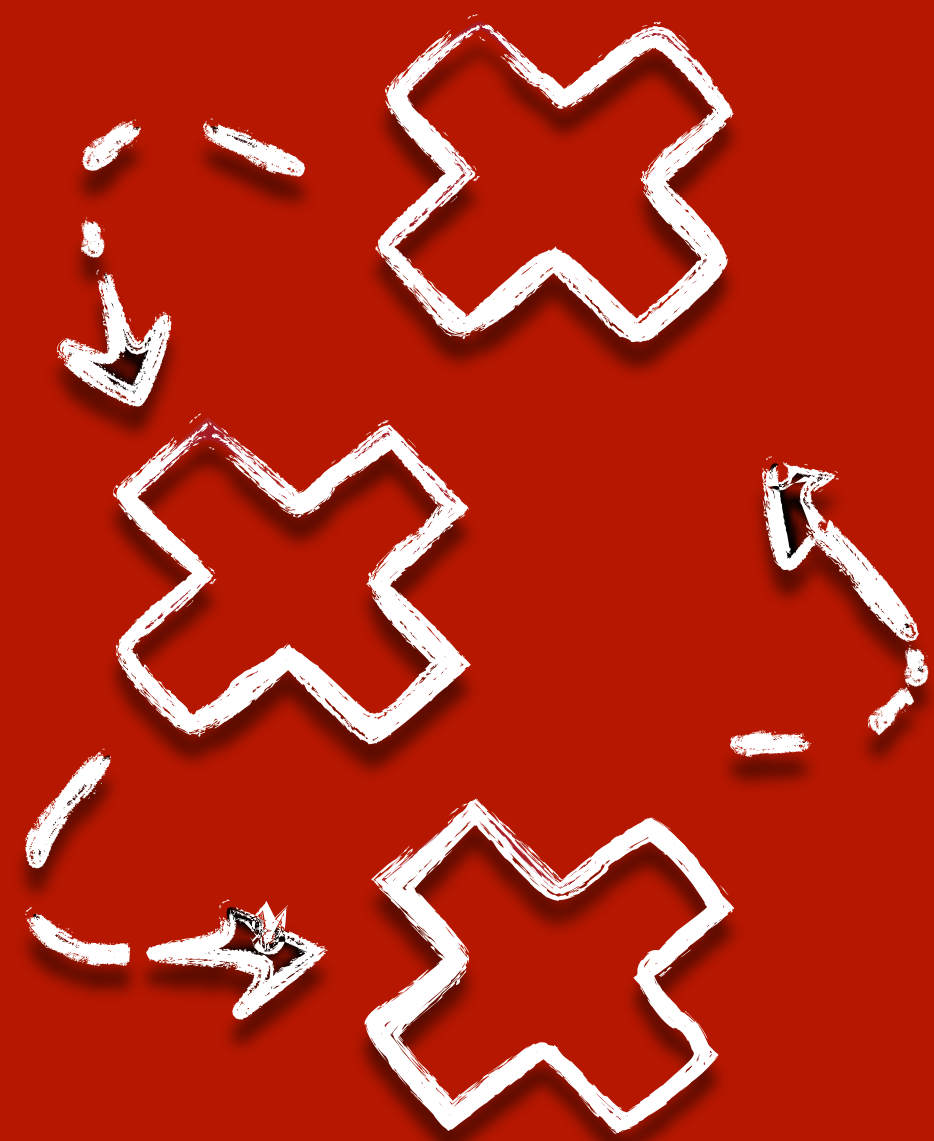


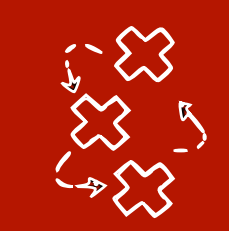


Seminar: Advanced topics in causality and causal ML research

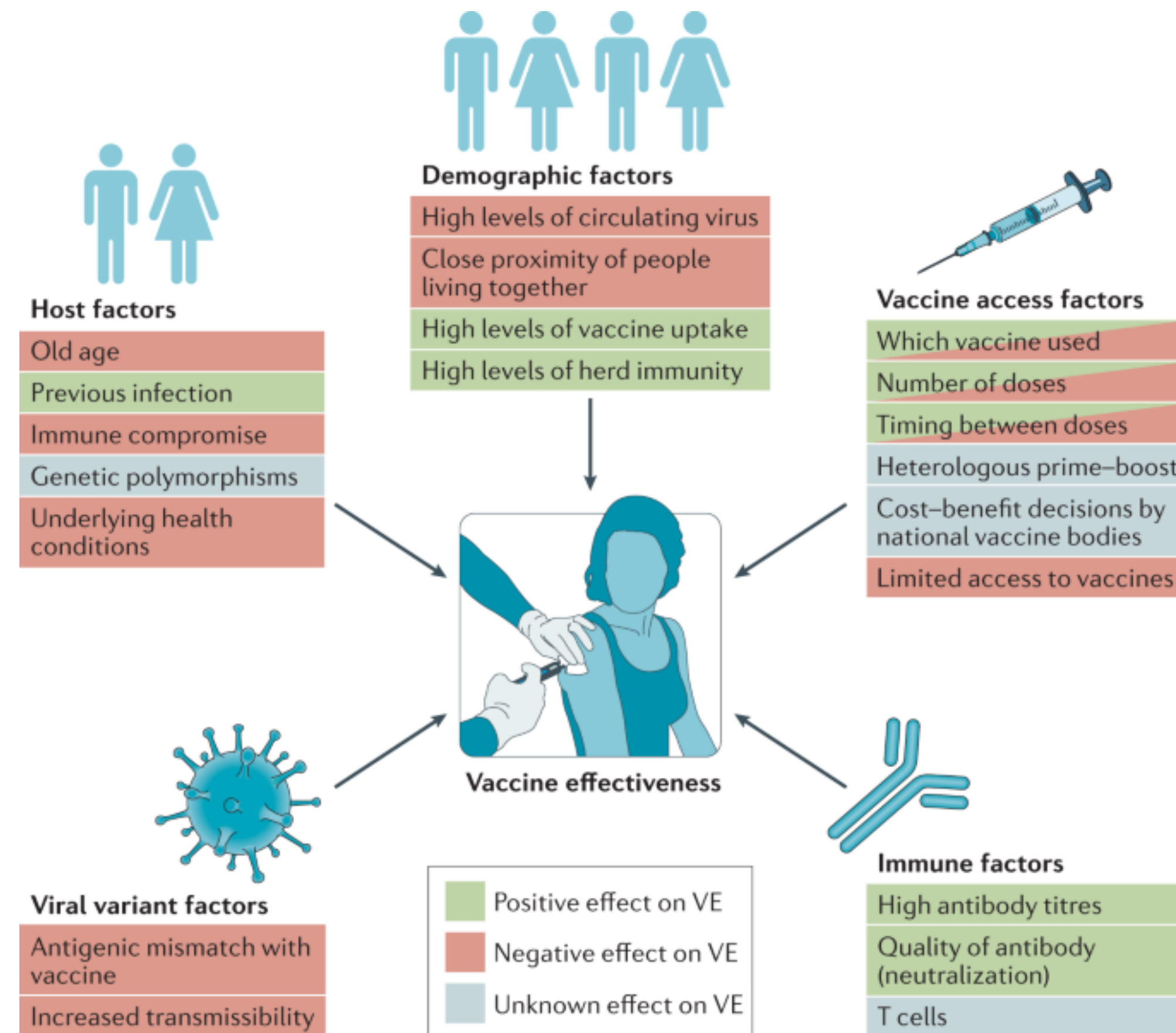
Lecture 1.2: Introduction

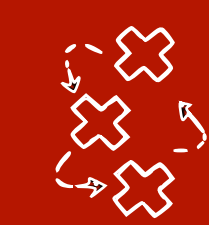
Lecturer: Sara Magliacane

UdS - Summer 2026



Causal questions are ubiquitous: healthcare

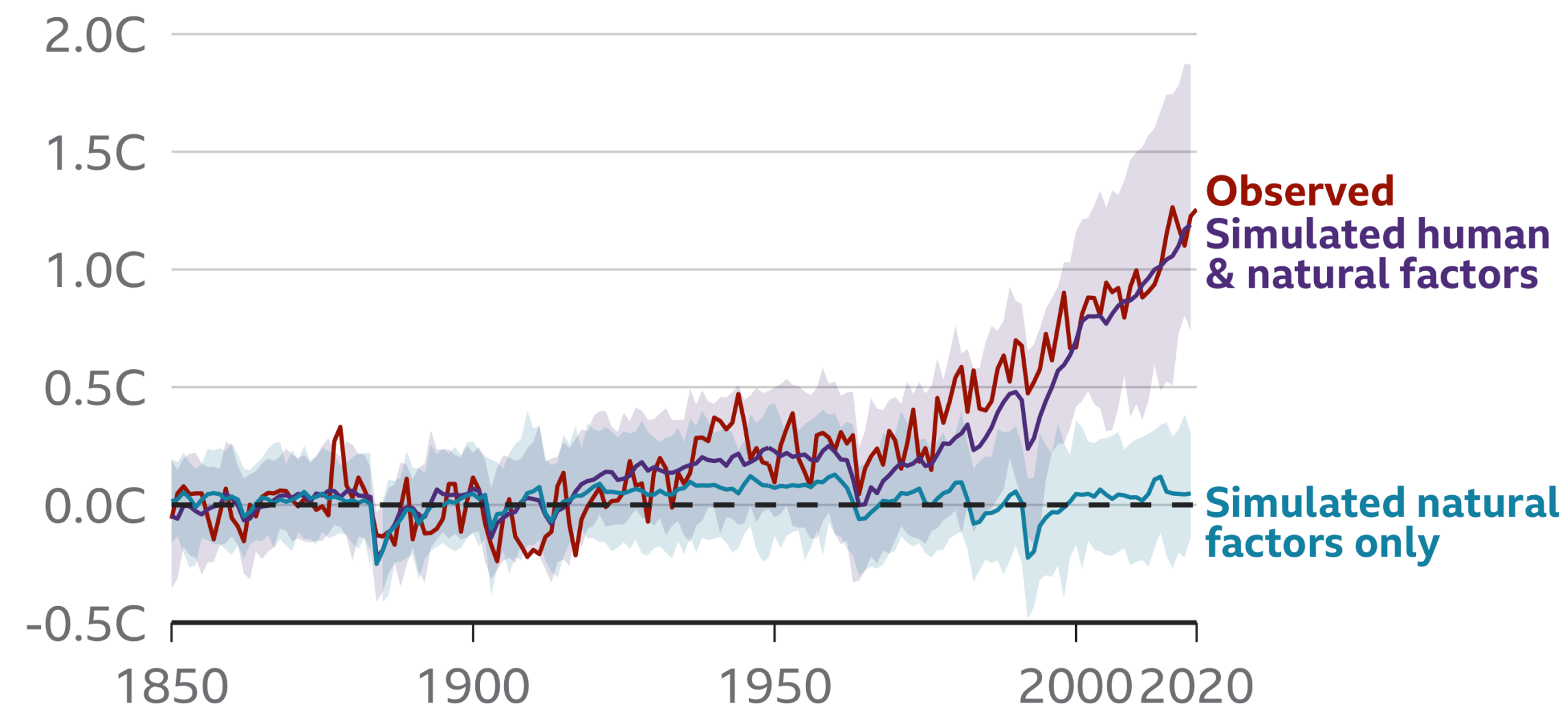




Causal questions are ubiquitous: **climate change**

Human influence has warmed the climate

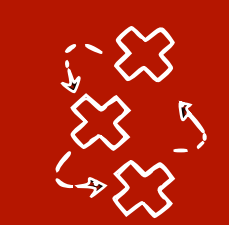
Change in average global temperature relative to 1850-1900, showing observed temperatures and computer simulations



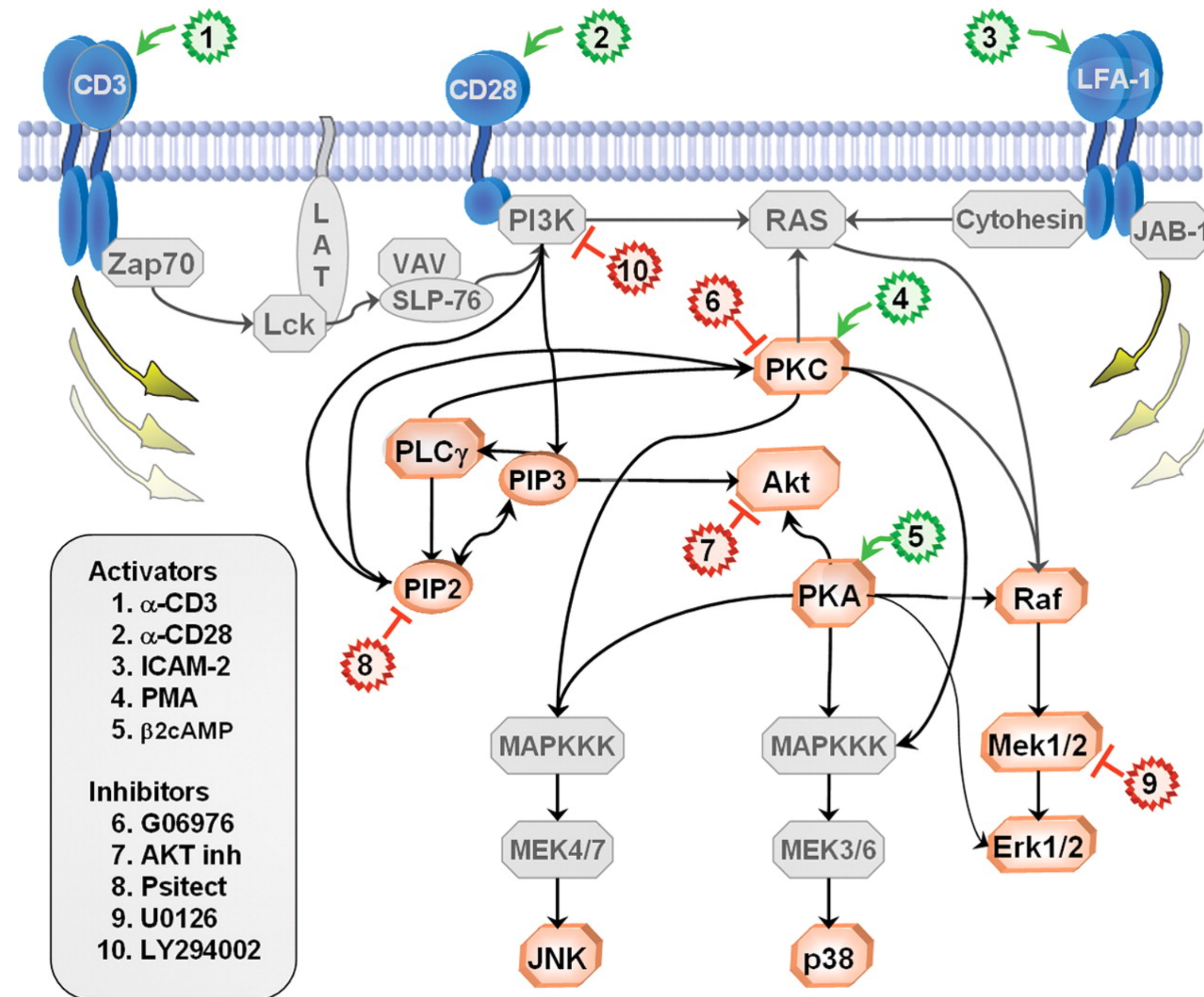
Note: Shaded areas show possible range for simulated scenarios

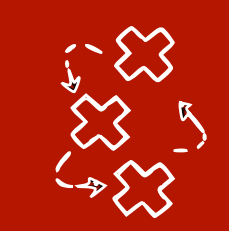
Source: IPCC, 2021: Summary for Policymakers



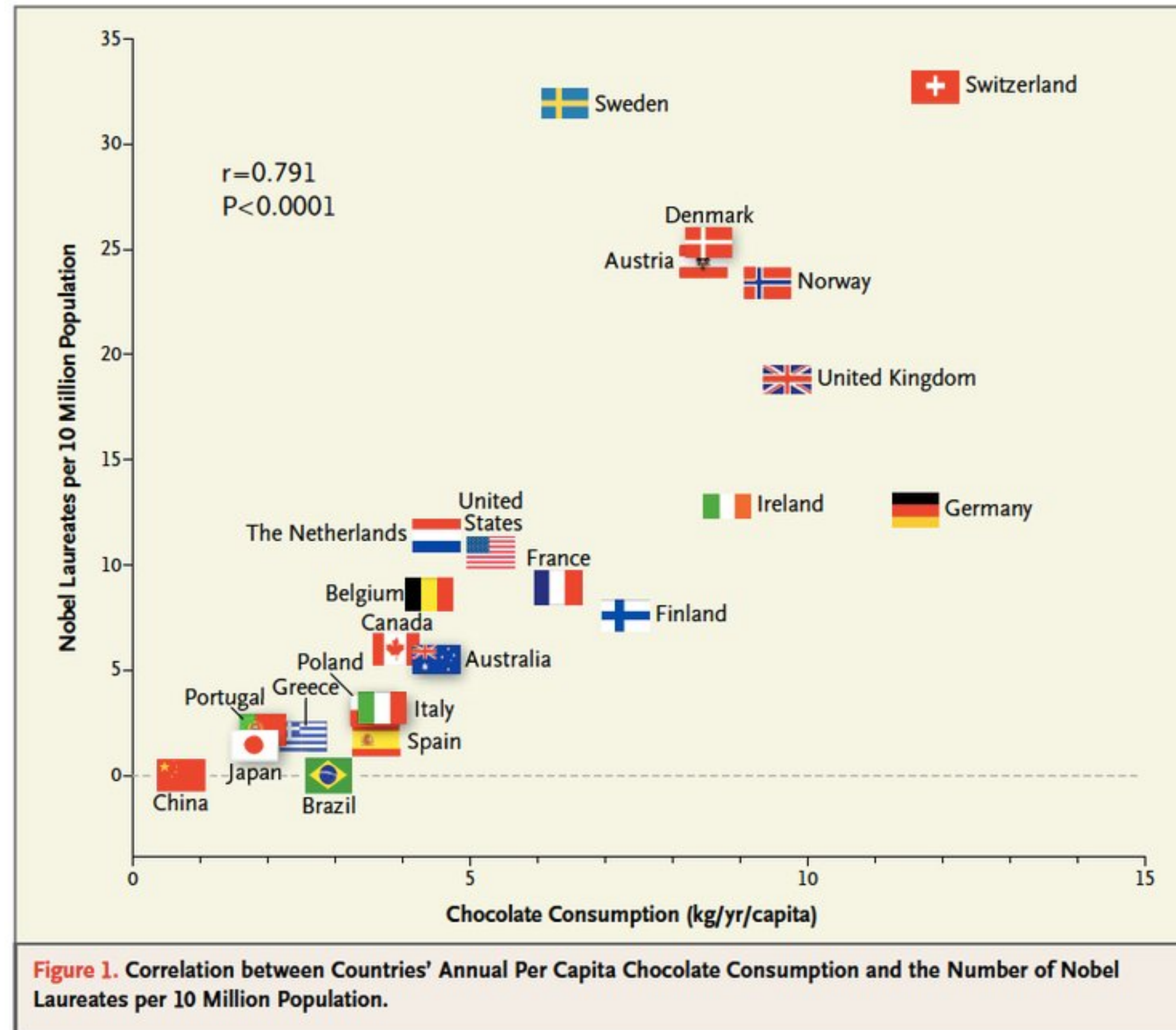


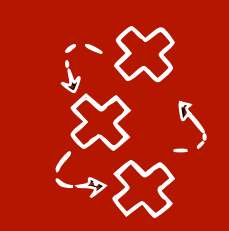
Causal questions are ubiquitous: **biology**



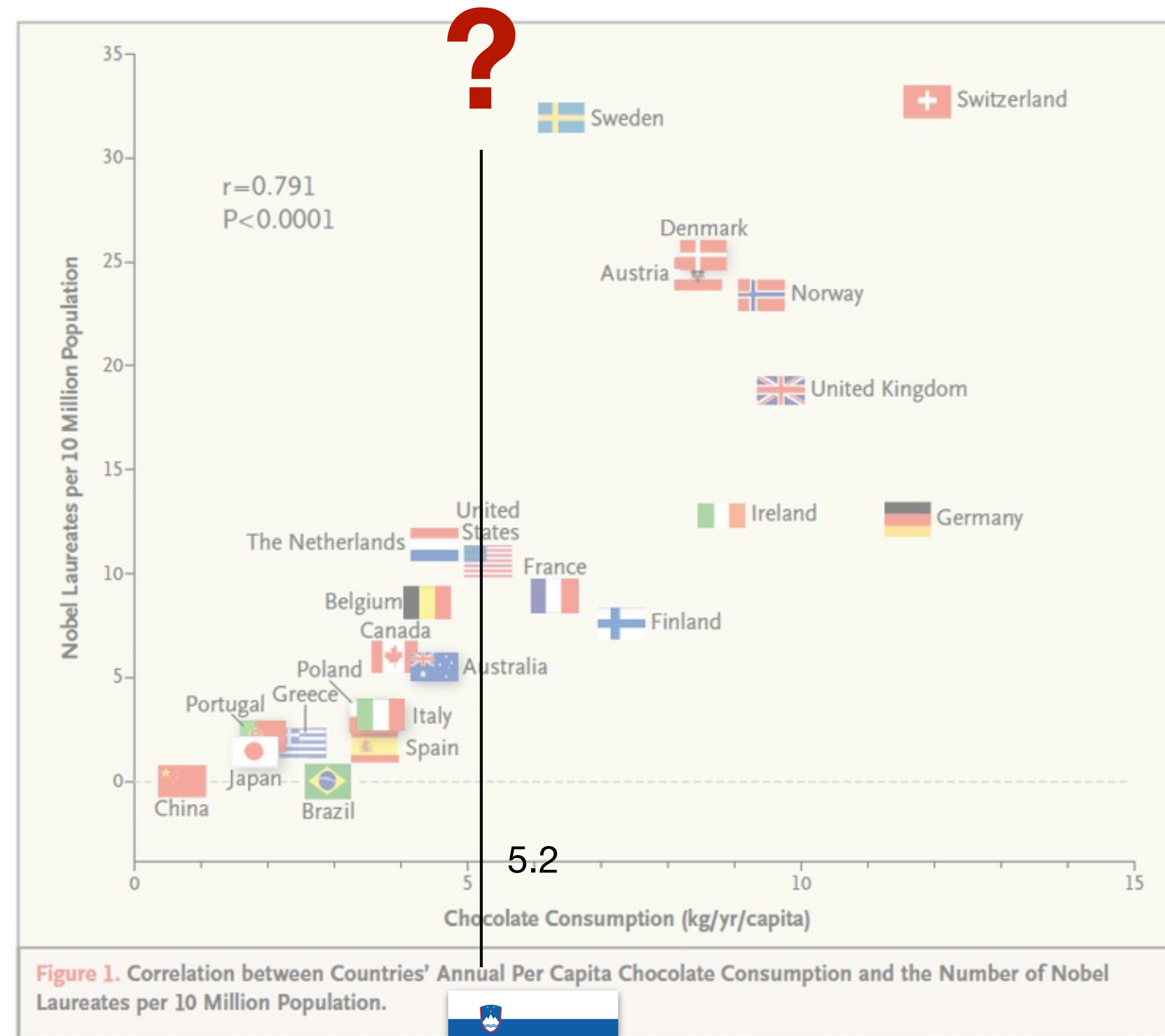


Correlation or causation? Does chocolate cause Nobel prizes?

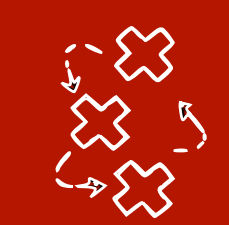




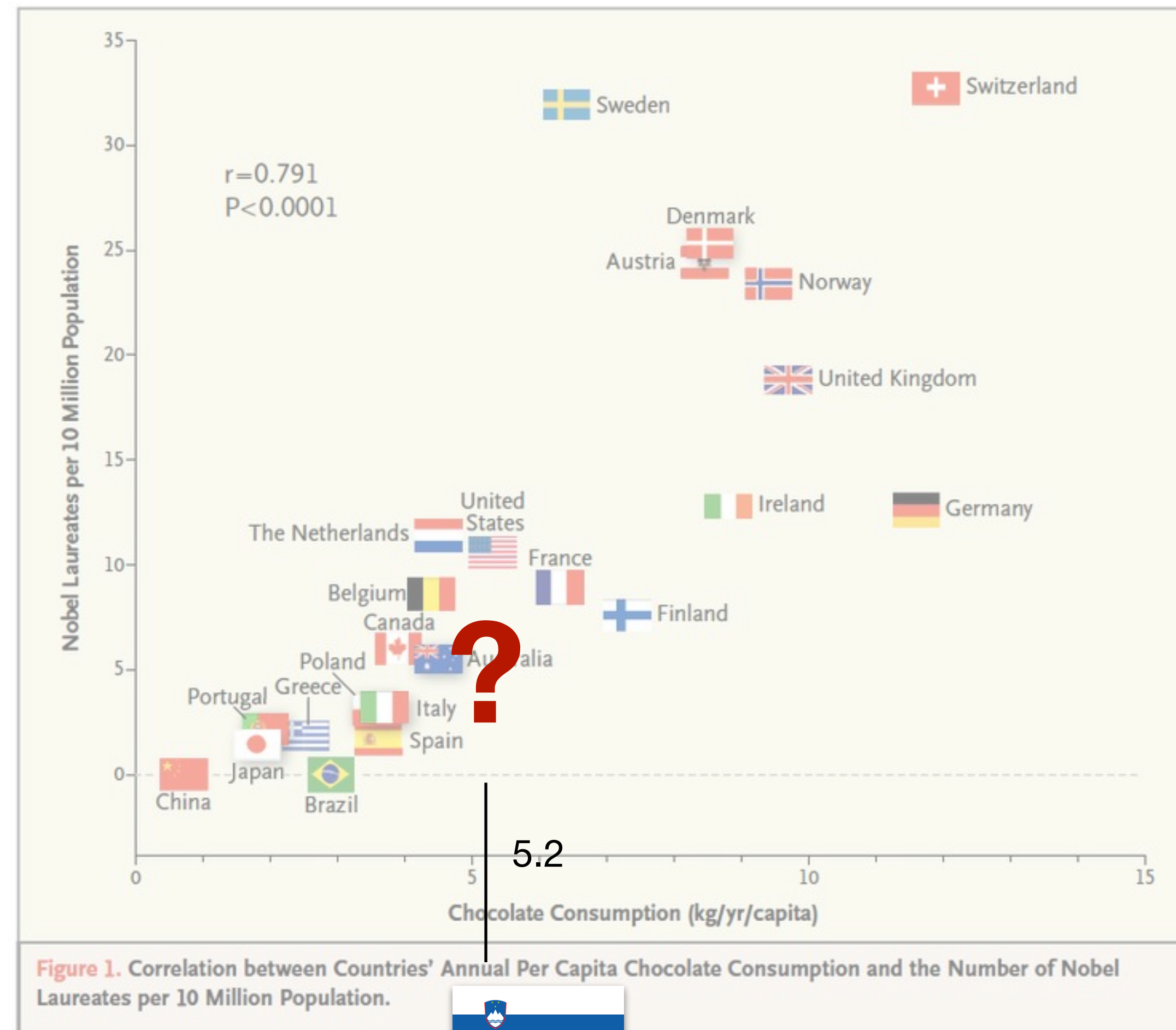
Correlation or causation? Can we predict a new data point?



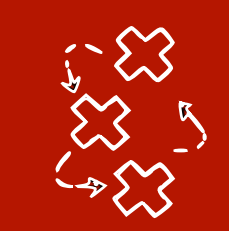
[Messerli, 2012] <https://www.nejm.org/doi/full/10.1056/NEJMon1211064>



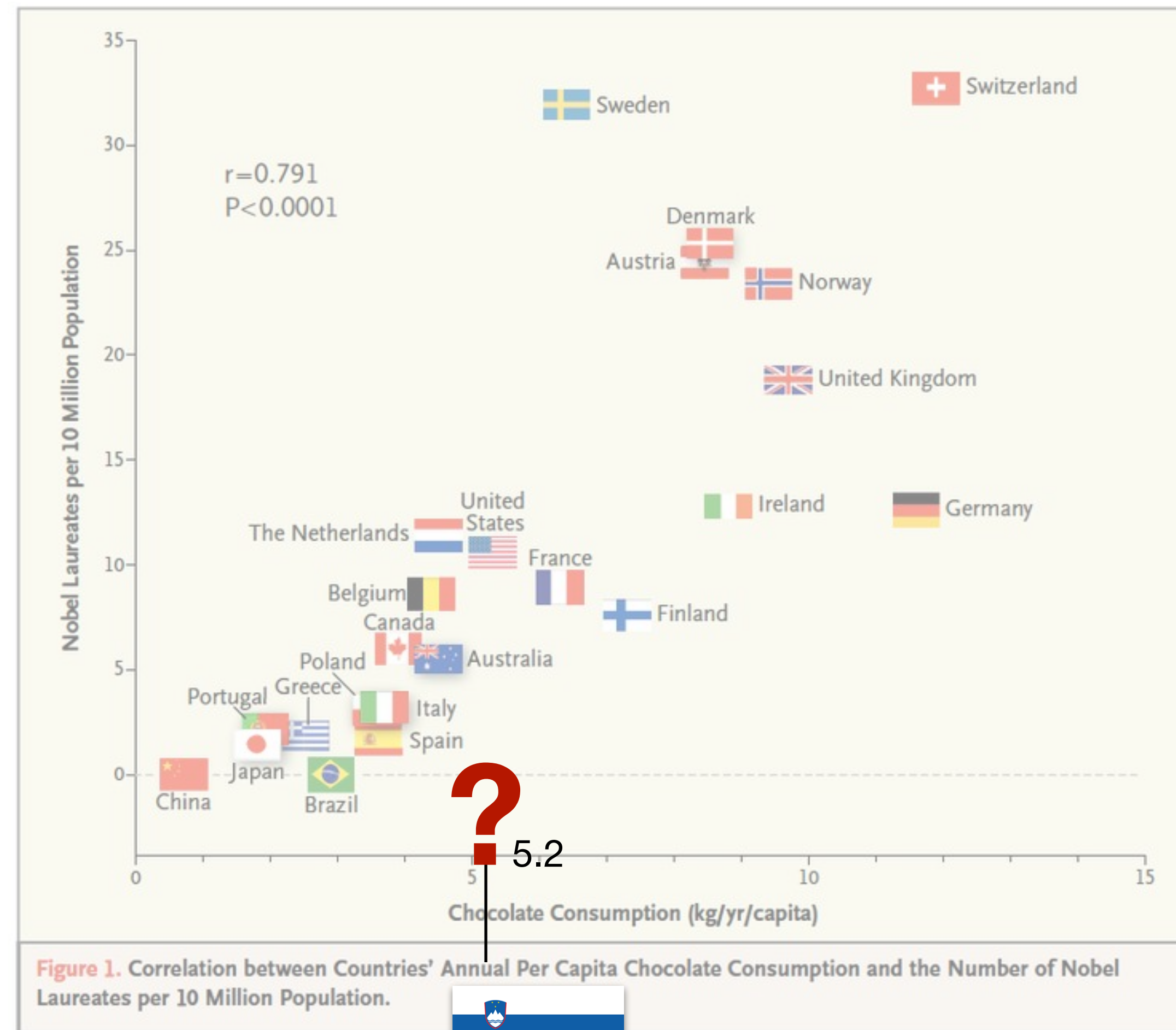
Correlation or causation? Can we predict a new data point?



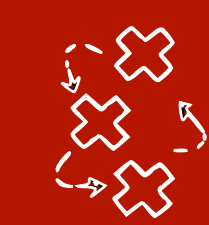
[Messerli, 2012] <https://www.nejm.org/doi/full/10.1056/NEJMon1211064>



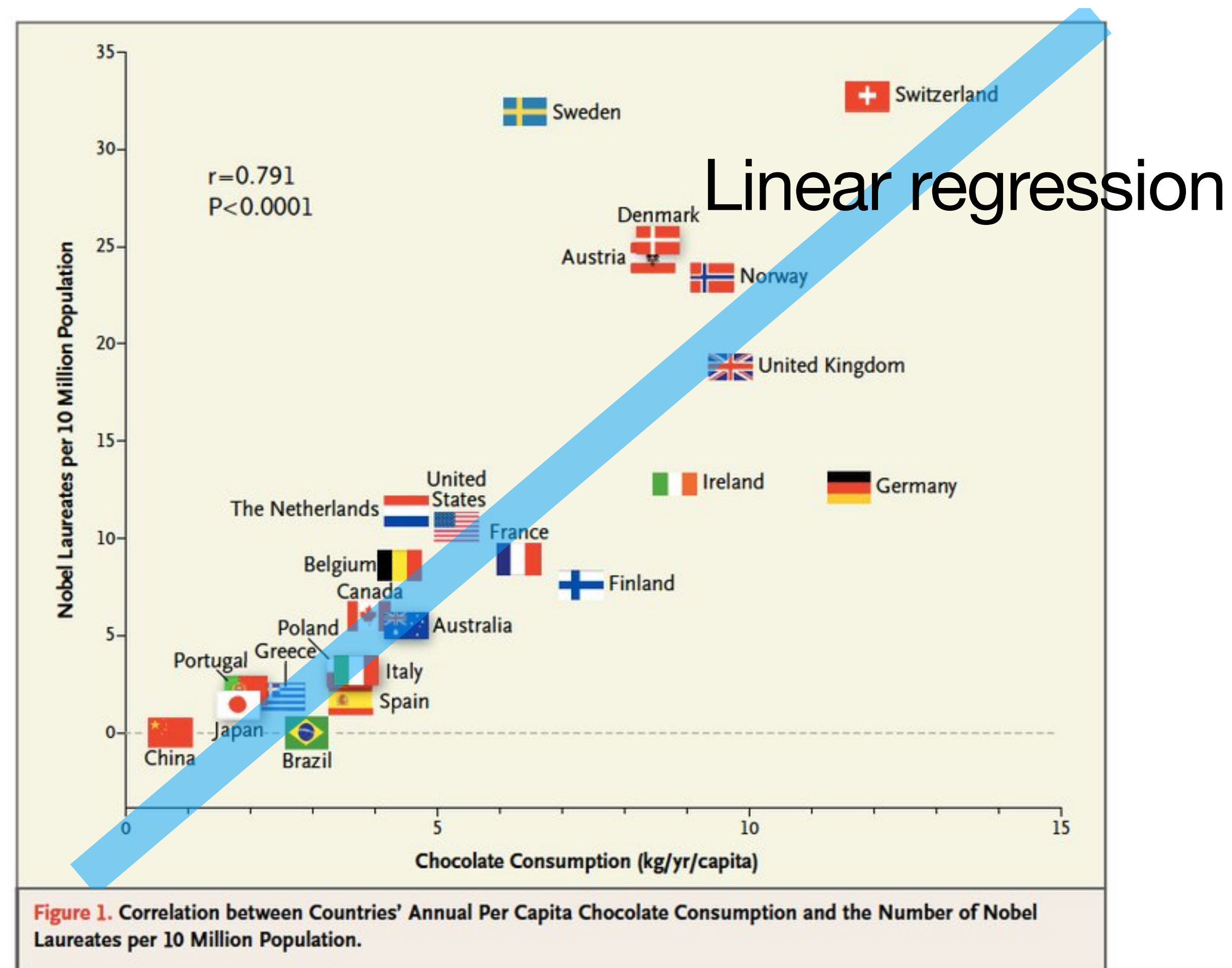
Correlation or causation? Can we predict a new data point?

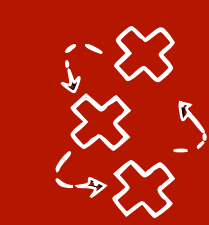


[Messerli, 2012] <https://www.nejm.org/doi/full/10.1056/NEJMon1211064>

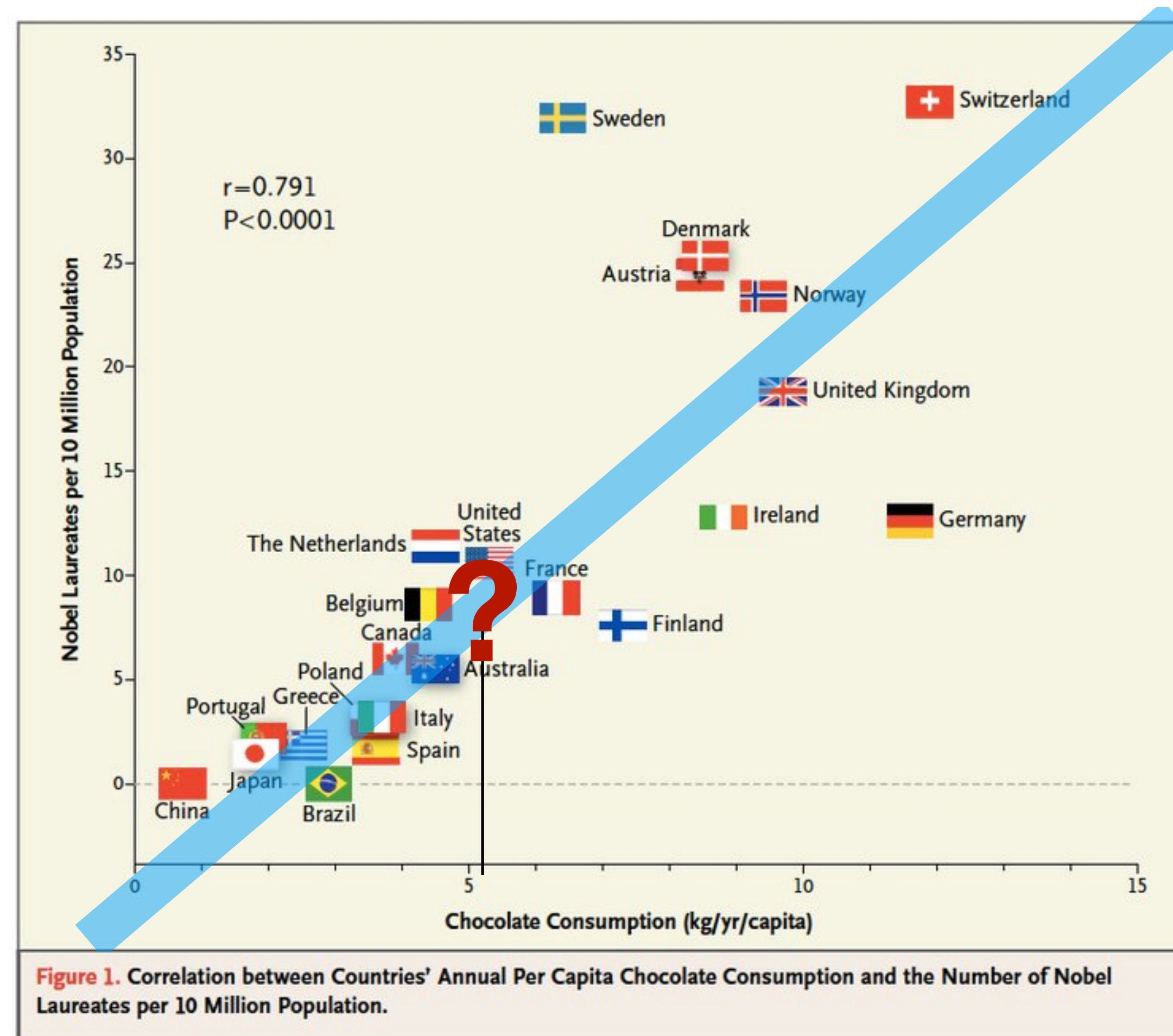


Correlation or causation? Can we predict a new data point?

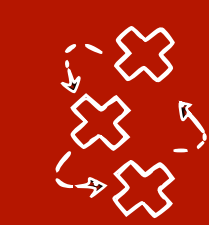




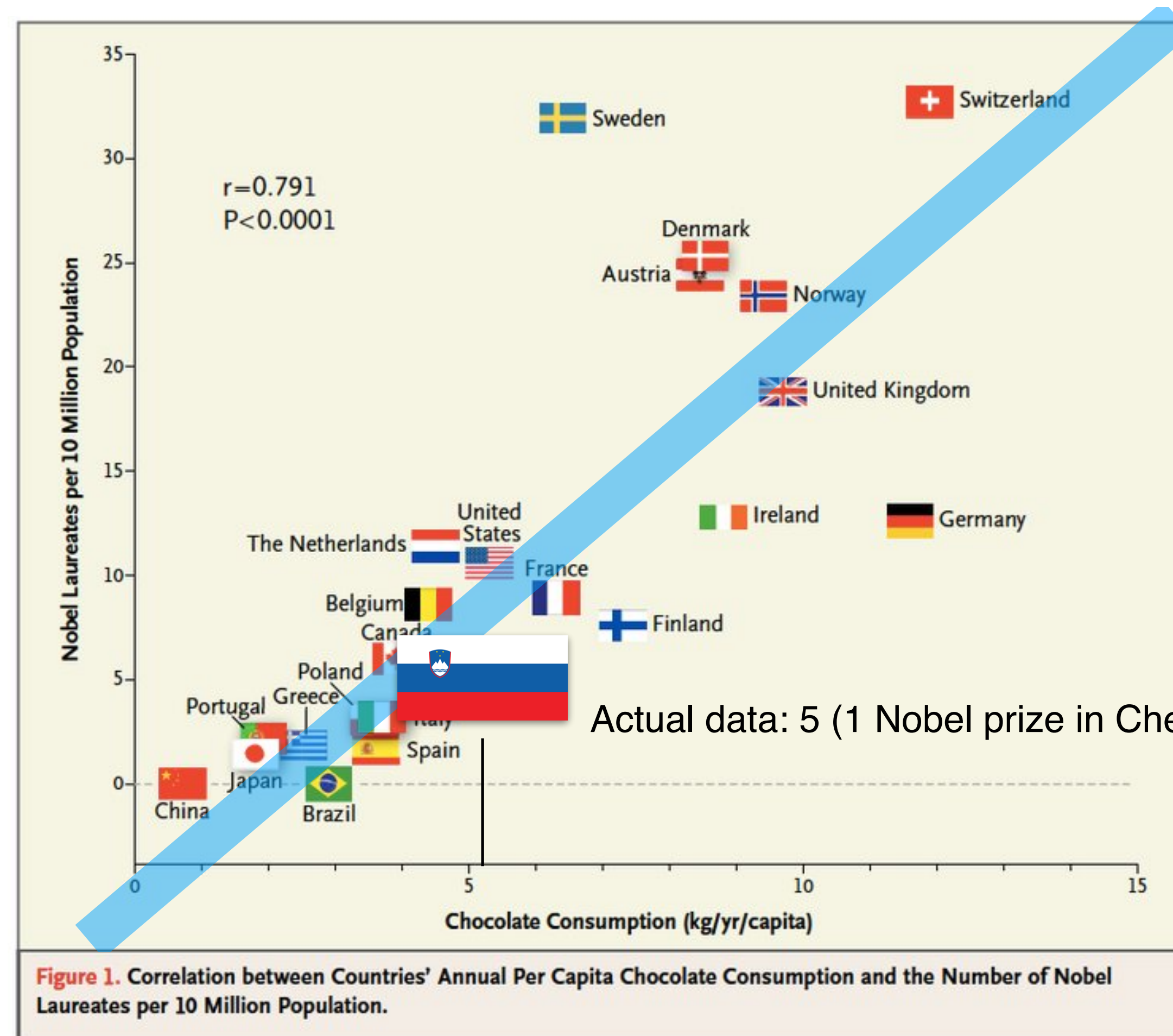
Correlation or causation? Can we predict a new data point?

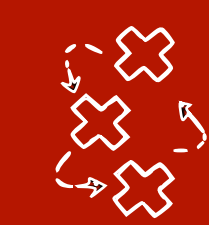


[Messerli, 2012] <https://www.nejm.org/doi/full/10.1056/NEJMon1211064>



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



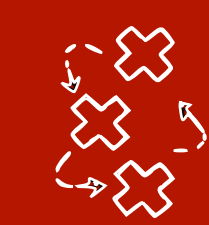


Correlation or causation? Can we predict a new data point?

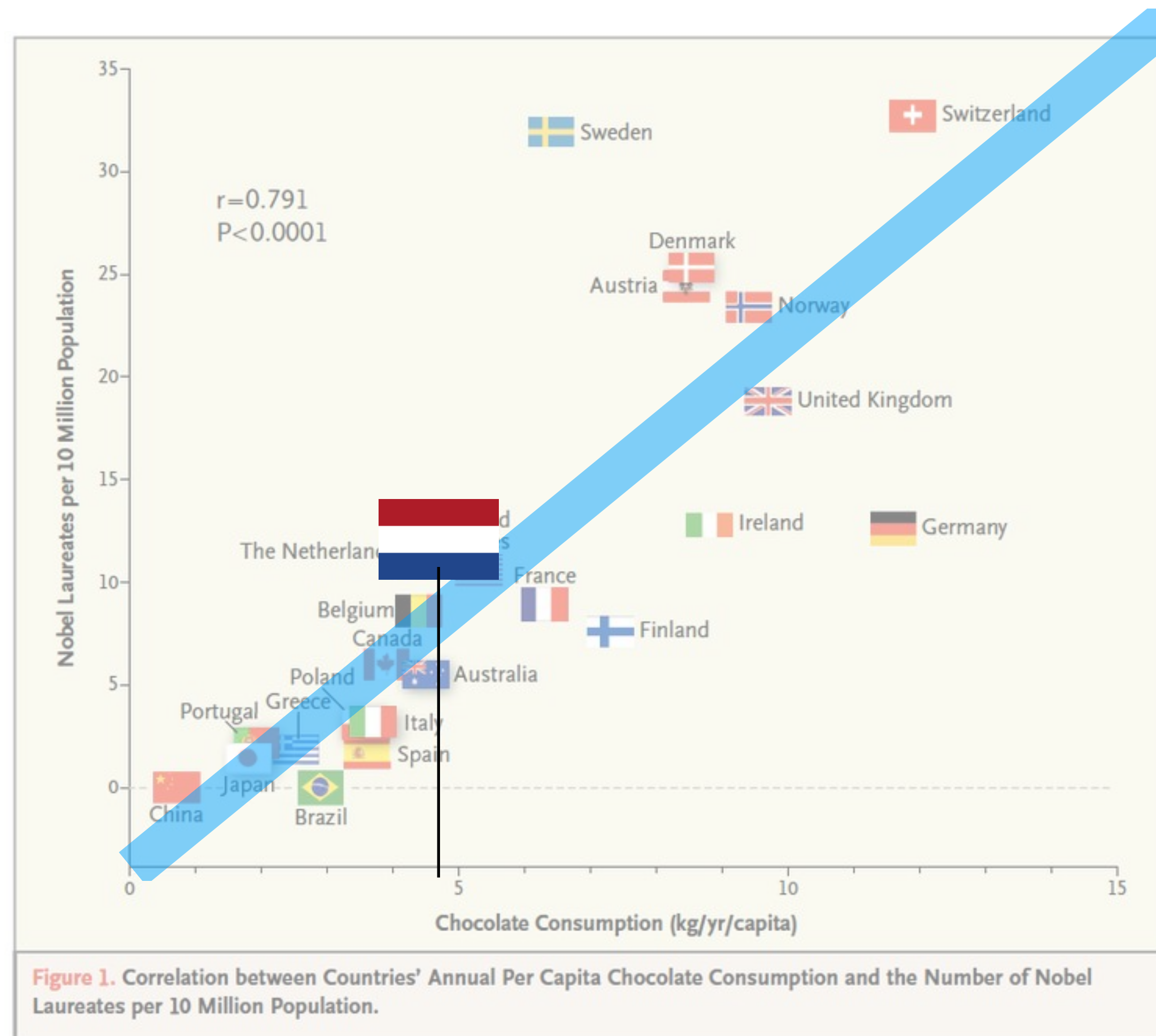
Prediction works!!!

Why shall we care about causality?

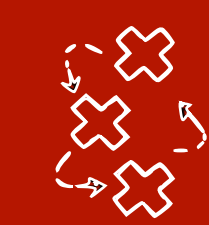




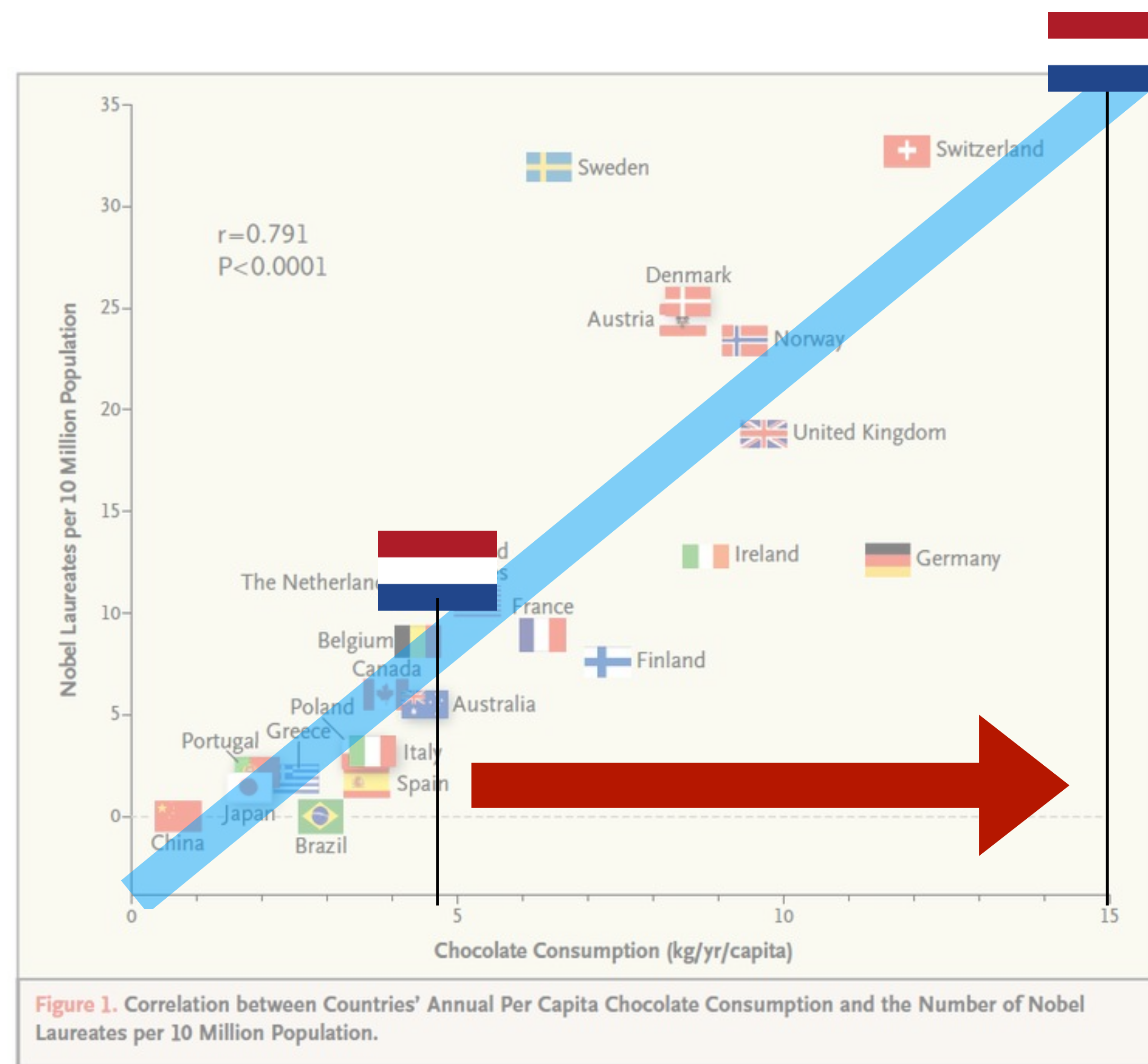
Helping the Dutch government with decision making



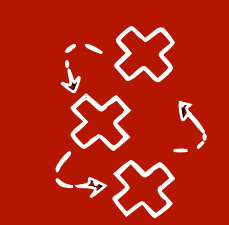
[Messerli, 2012] <https://www.nejm.org/doi/full/10.1056/NEJMon1211064>



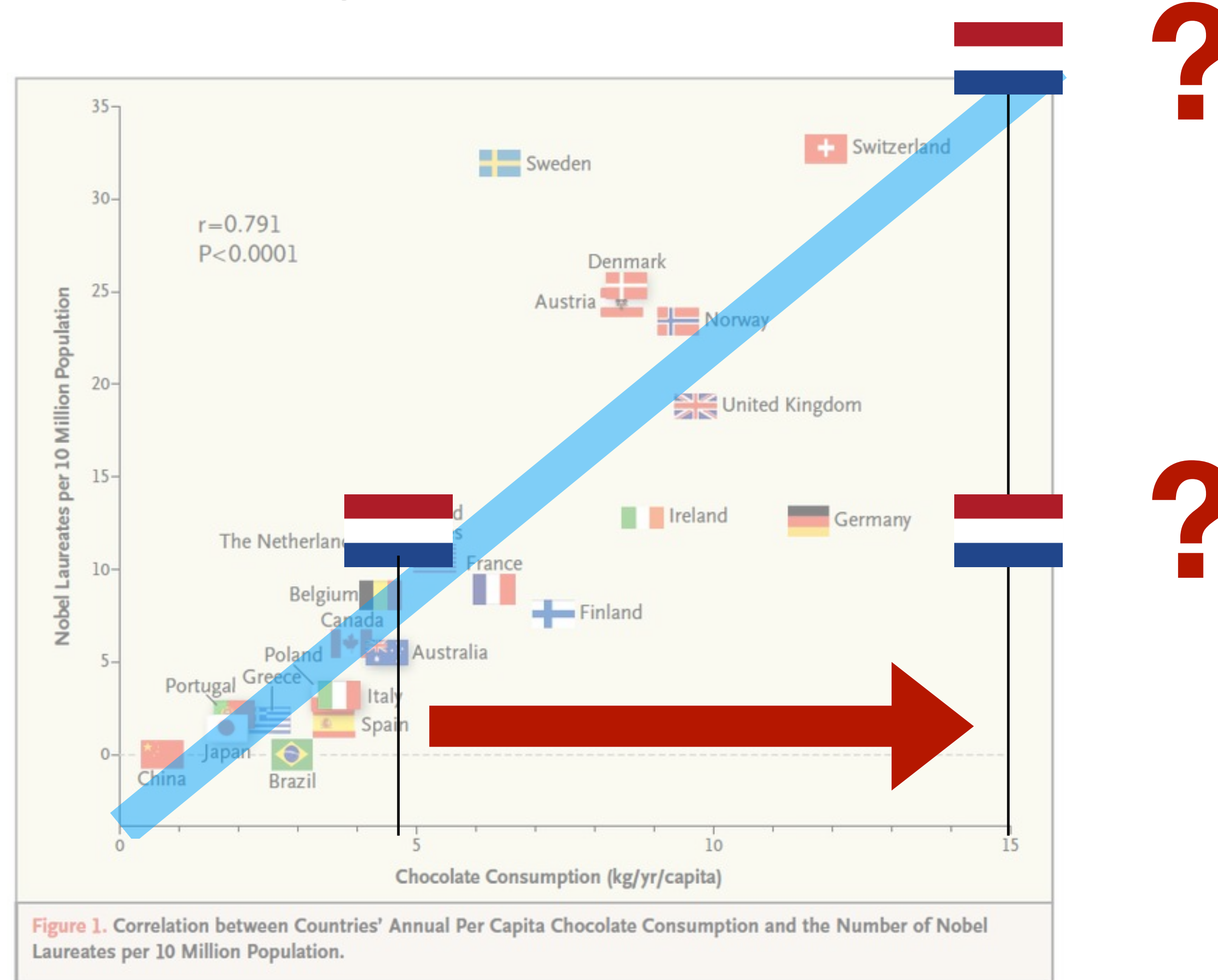
Helping the Dutch government with decision making



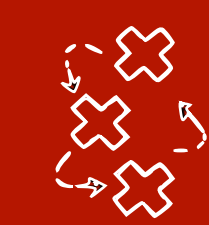
[Messerli, 2012] <https://www.nejm.org/doi/full/10.1056/NEJMon1211064>



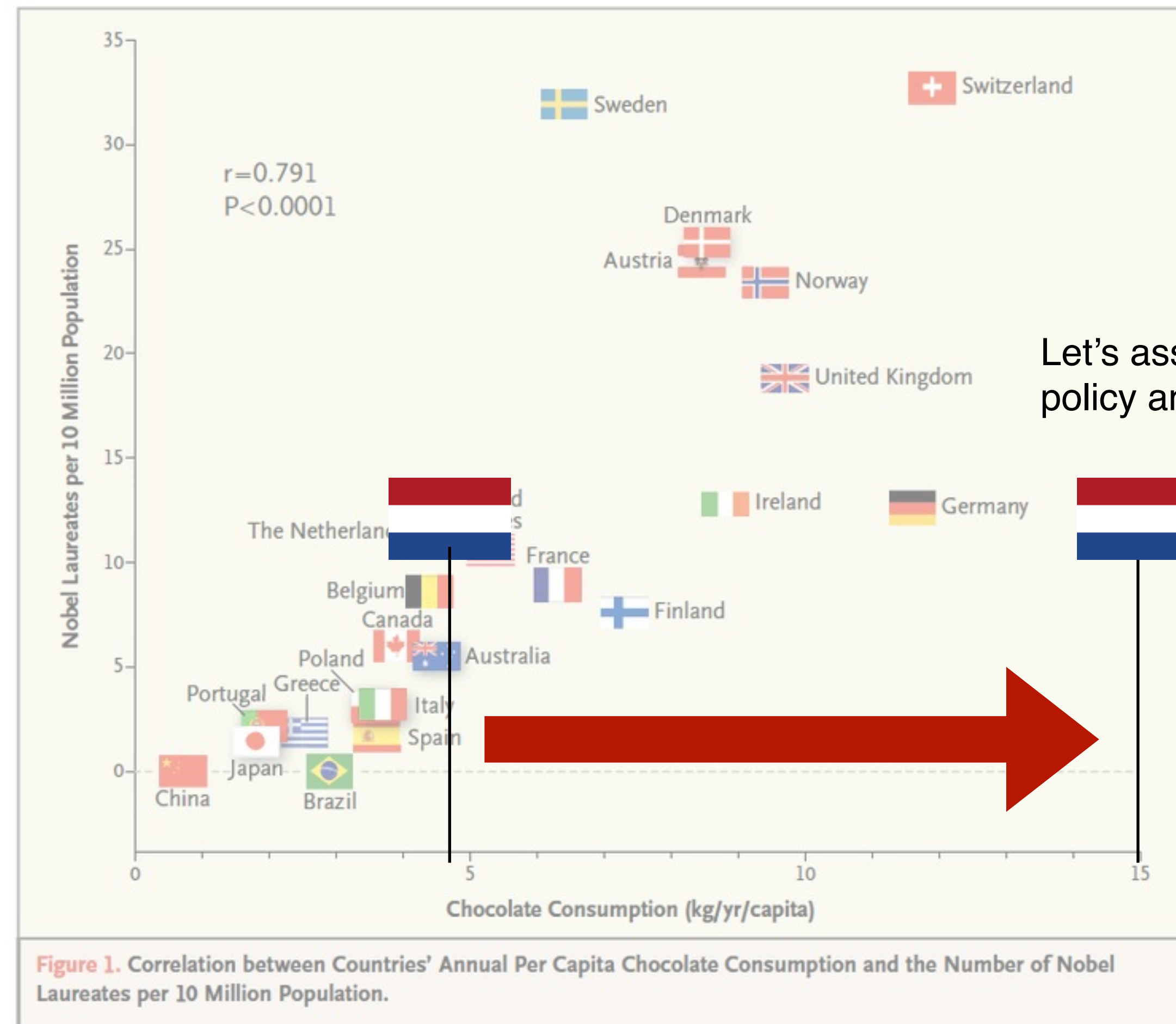
Helping the Dutch government with decision making



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

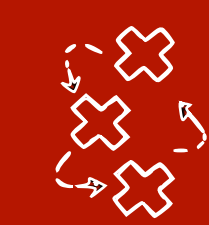


Helping the Dutch government with decision making



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

=> Chocolate probably does not cause Nobel prizes ... how to formalize it?



Many ways to define causality

The screenshot shows the Stanford Encyclopedia of Philosophy website. At the top, there is a navigation bar with links for 'Browse', 'About', and 'Support SEP'. Below this is a search bar containing the word 'causation'. The search results show '1-10 of 539 documents found'. The first result is titled 'Backward Causation' and is authored by Jan Faye. The snippet of the article describes retro-causation and mentions objections to backward causation. The second result is titled 'Probabilistic Causation' and includes a list of related topics such as causal models, manipulation, and various theories of causation.

Stanford Encyclopedia of Philosophy

[Browse](#) [About](#) [Support SEP](#)

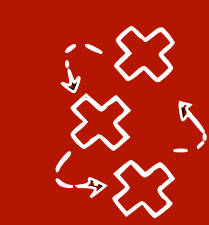
Search

causation

1-10 of 539 documents found

Backward Causation
called retro-causation. A common feature of our world seems to be that in all cases of **causation**, the cause...required for backward **causation** to be possible is some account of the direction of **causation** which does not ...objection against backward **causation** is that if eternalism and backward **causation** are possible, then the ...
Jan Faye
<https://plato.stanford.edu/entries/causation-backwards/>

Probabilistic Causation
causal models | **causation**: and manipulability | **causation**: backward | **causation**: counterfactual theories...Interpretation of Probability 1.4 General **Causation** and Actual **Causation** 1.5 Causal Relata 1.6 Further Reading... a theory of **causation** could provide some explanation of the directionality of **causation**, rather than ...



Humean regularity theory

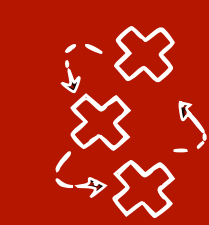


David Hume defines a cause:

[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



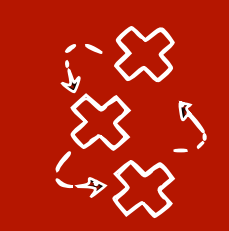
David Hume defines a cause:

[a]n object precedent and contiguous to another, and where the former are plac'd in like relations of priority and contiguity to resemble the latter. (Hume 1739: book I, part III, section IV)

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

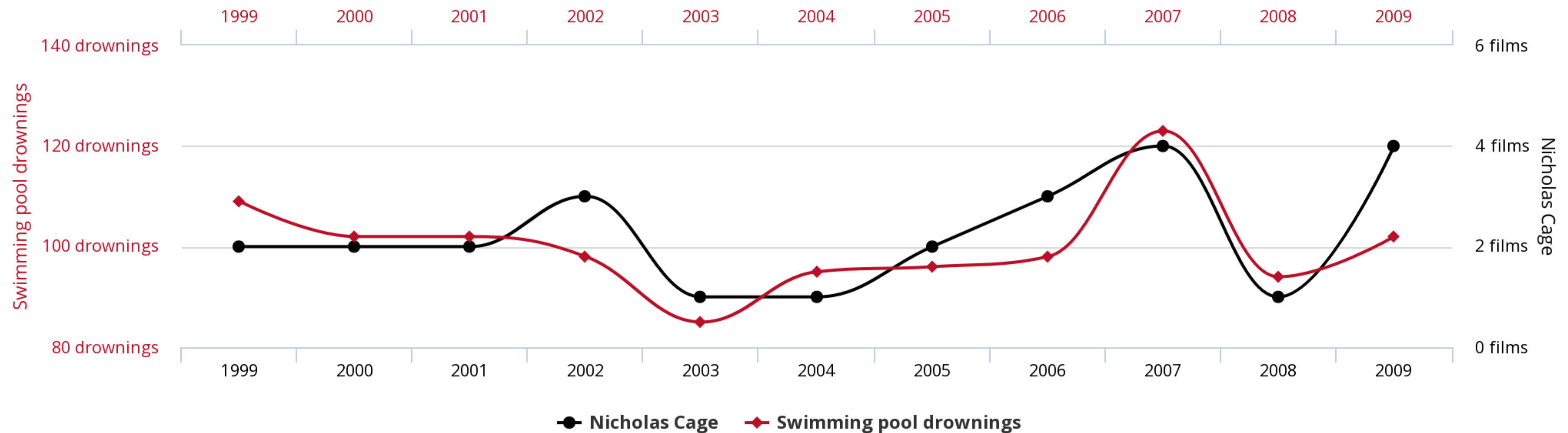


(<https://plato.stanford.edu/entries/causation-regularity>)



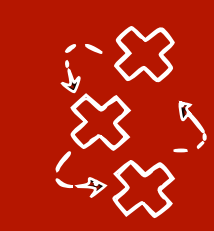
Other examples of regularity theories

Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in



tylervigen.com

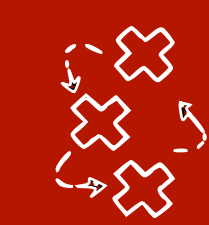
<http://tylervigen.com/spurious-correlations>



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



The classical approach to causality is based on experimentation.

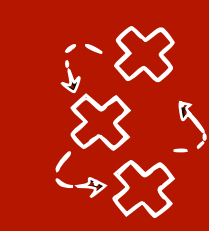


We know what causality is not, but what is it?



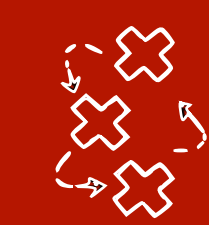
The classical approach to causality is based on experimentation.

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



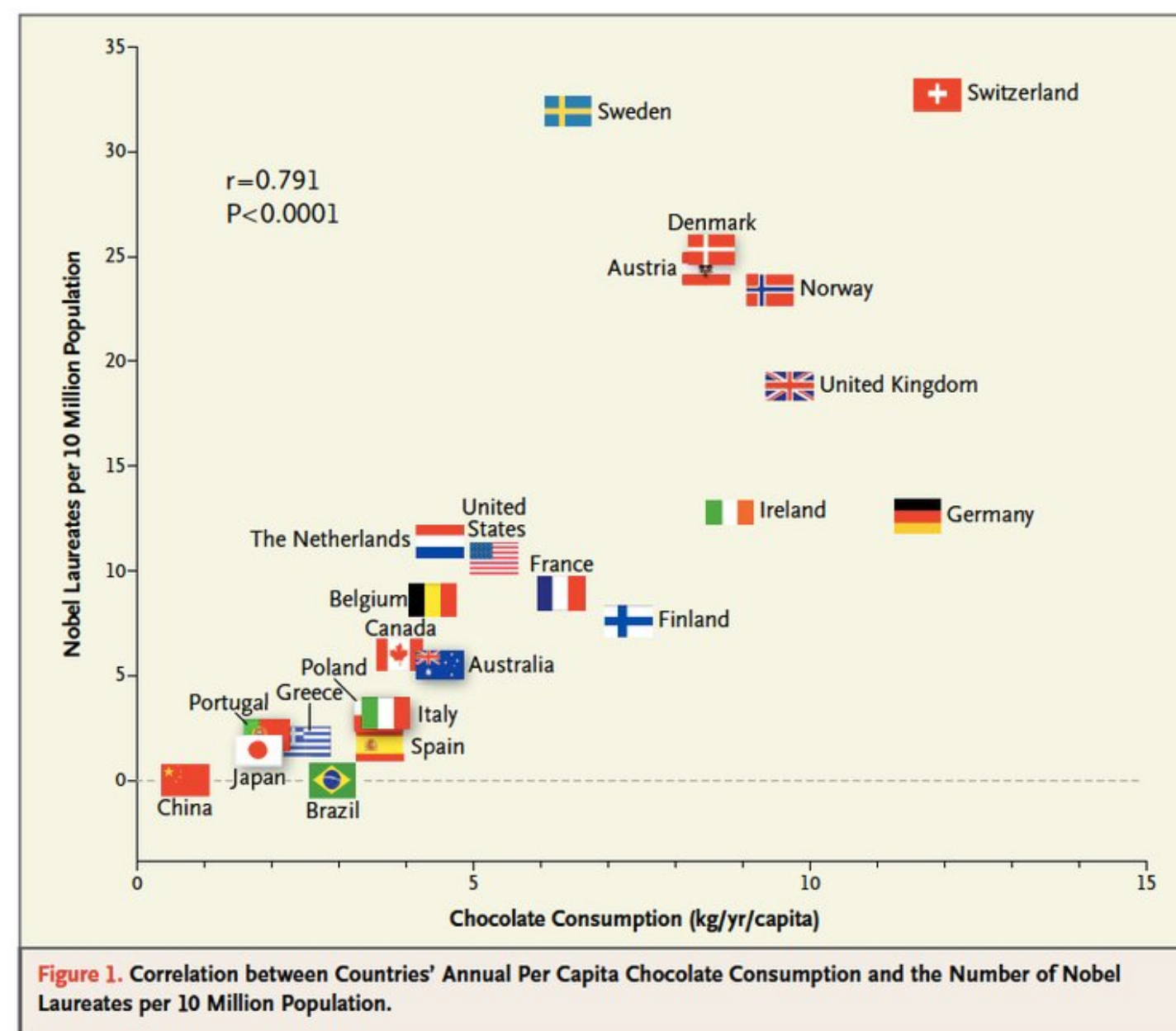
A working definition of causality in machine learning

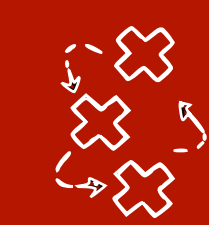
Informal definition: A variable X causes another variable Y , if changing (the distribution of) X , e.g. by fixing its value, changes (the distribution of) Y



A working definition of causality in machine learning

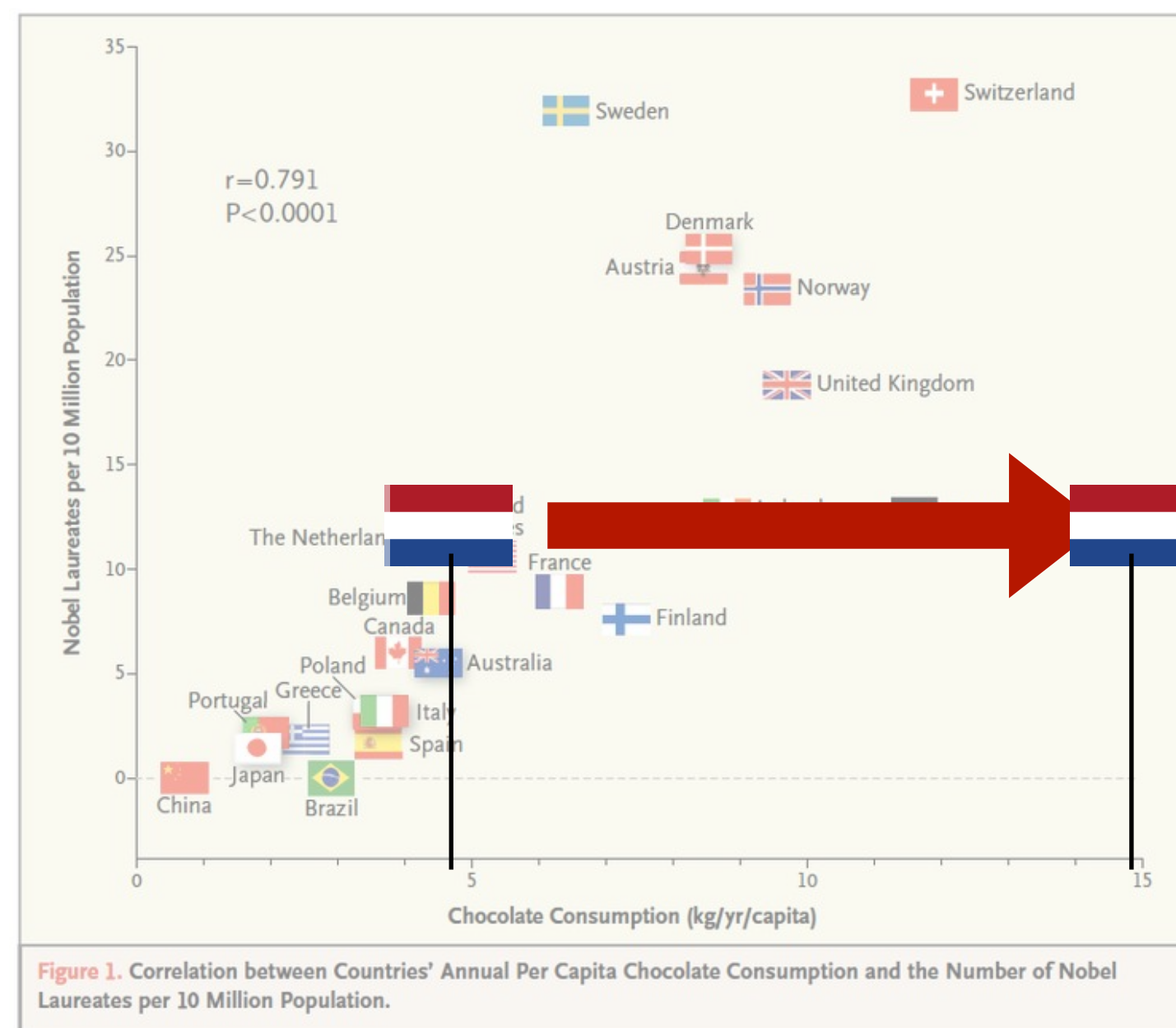
Informal definition: A variable X causes another variable Y , if changing (the distribution of) X , e.g. by fixing its value, changes (the distribution of) Y



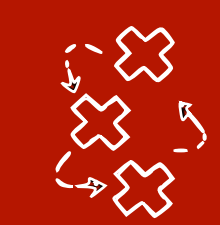


A working definition of causality in machine learning

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

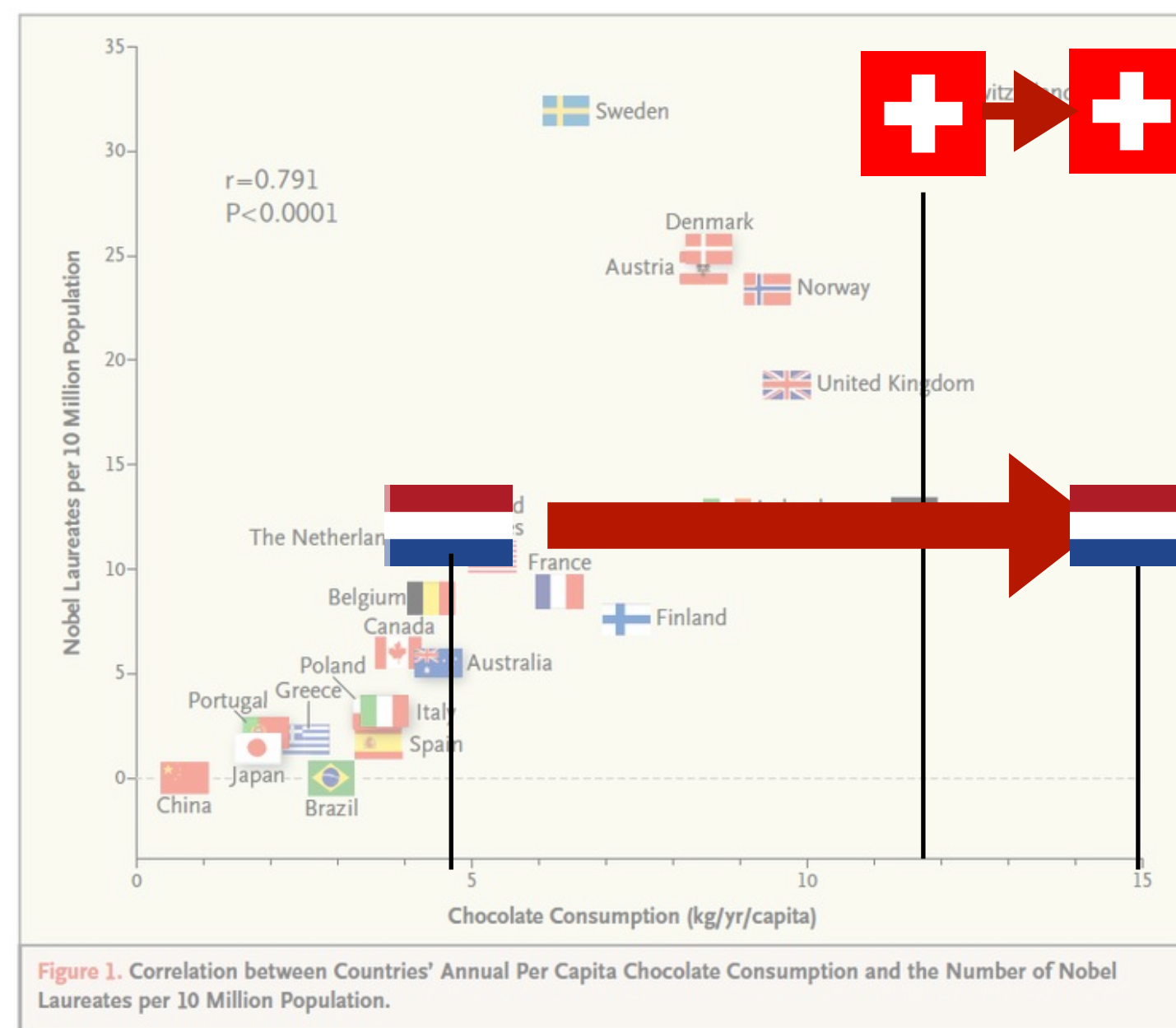


NL eats more chocolate => nothing changes



A working definition of causality in machine learning

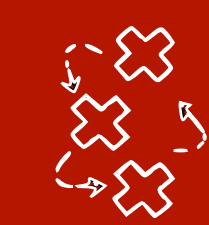
Informal definition: A variable X causes another variable Y , if changing (the distribution of) X , e.g. by fixing its value, changes (the distribution of) Y



NL eats more chocolate => nothing changes

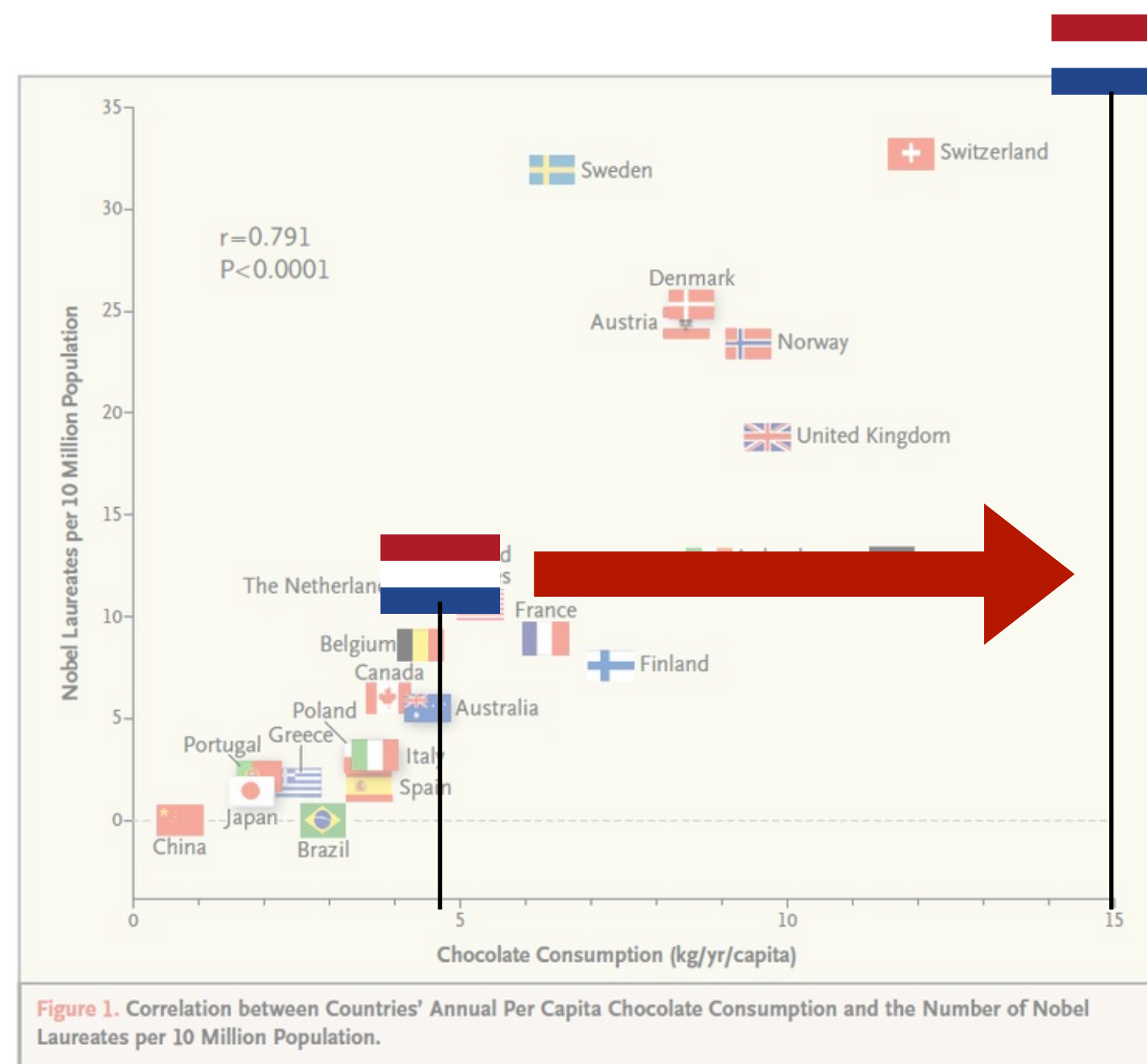
... and similarly for other countries (and other values)

Chocolate does not cause Nobel prizes



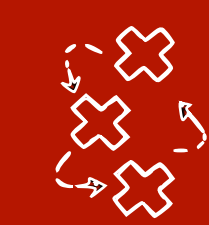
A working definition of causality in machine learning

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



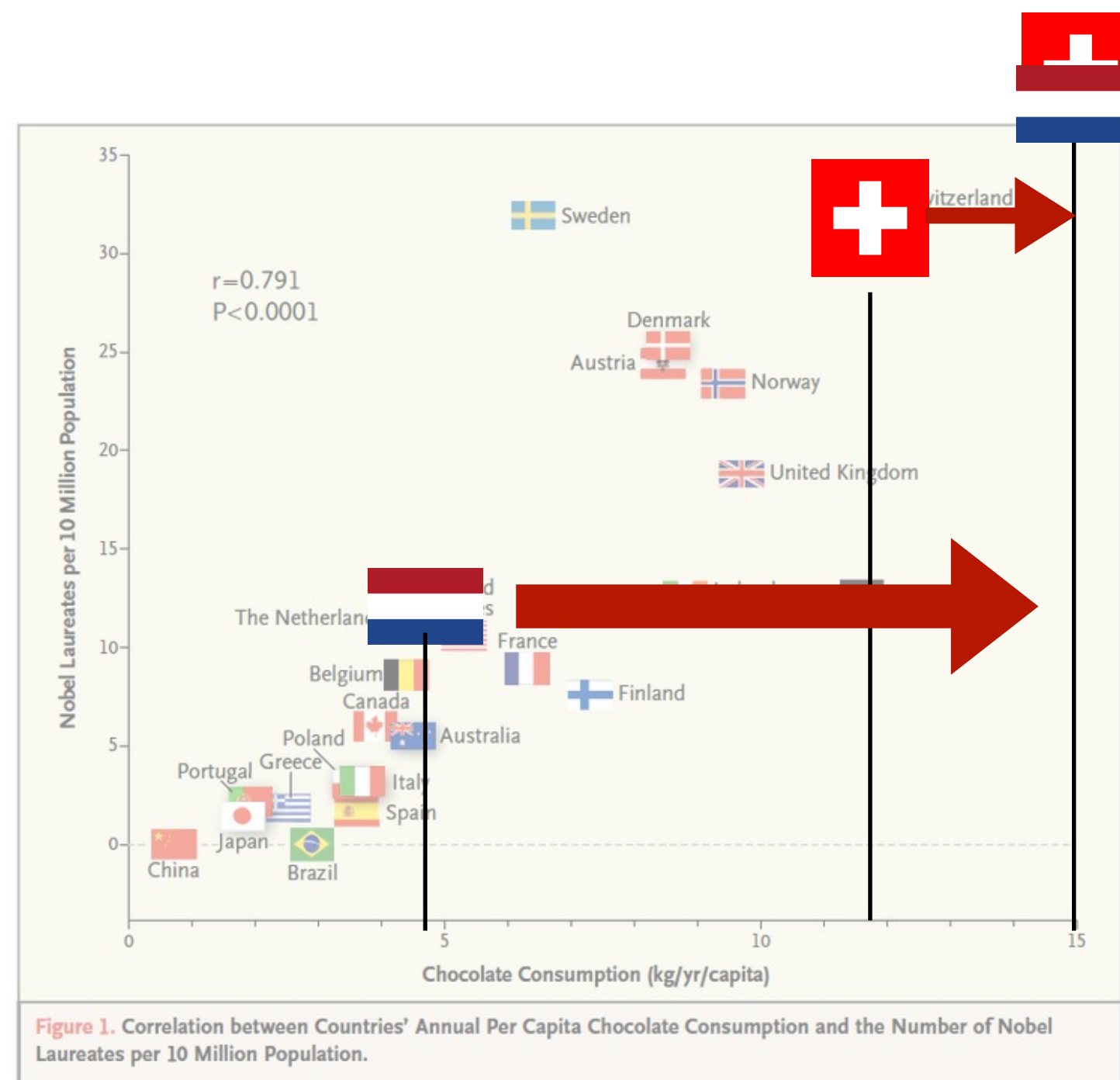
In a hypothetical universe:

NL eats more chocolate => more Nobel prizes



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



In a hypothetical universe:

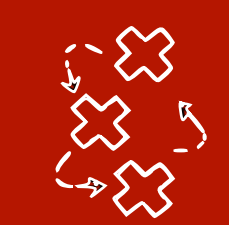
NL eats more chocolate => more Nobel prizes

CH eats more chocolate => more Nobel prizes

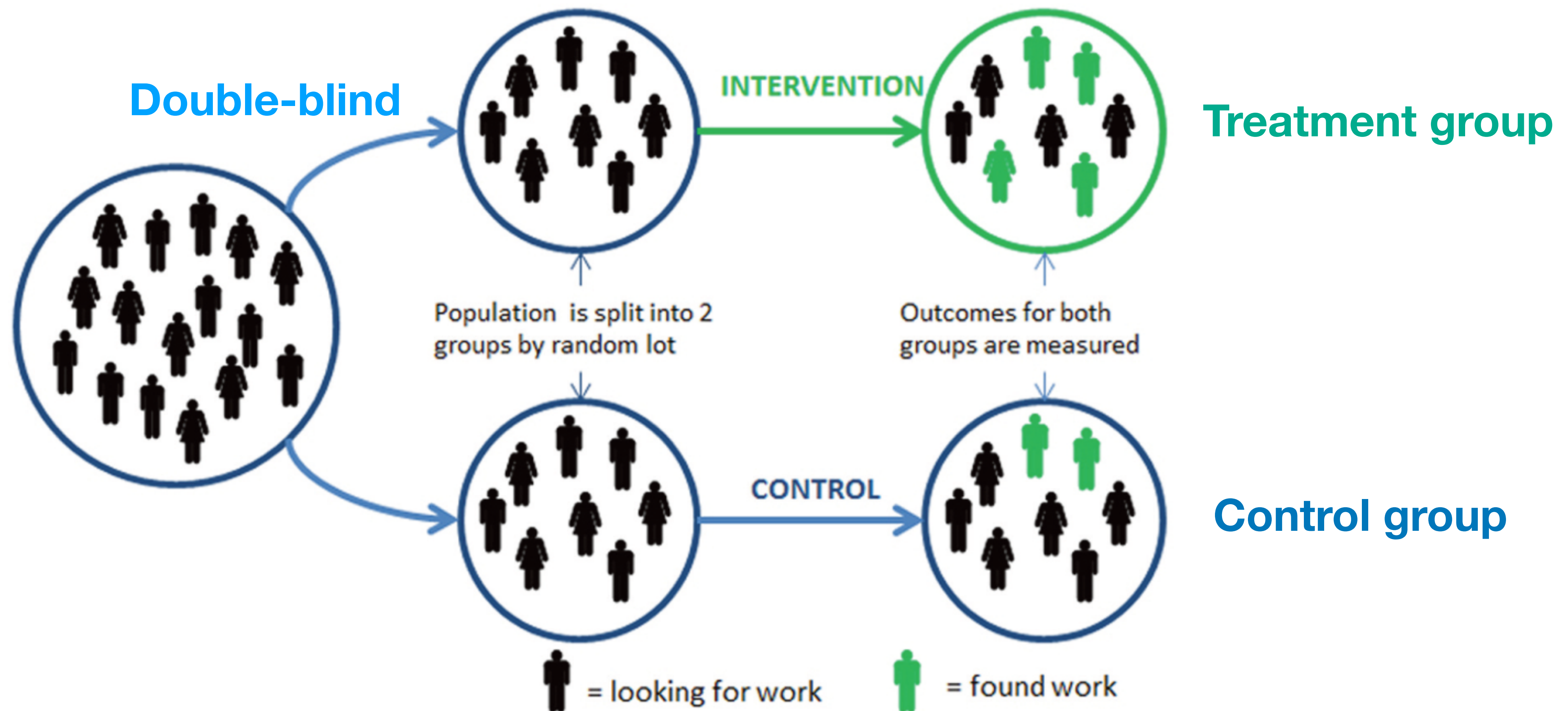
... and similarly for (some) other countries

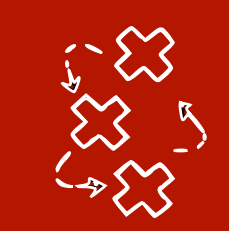
Chocolate causes Nobel prizes

Based on experimental data



Gold standard for experiments: Randomized Controlled Trials





Randomised controlled trials (RCT): spot the difference?

Figure A:

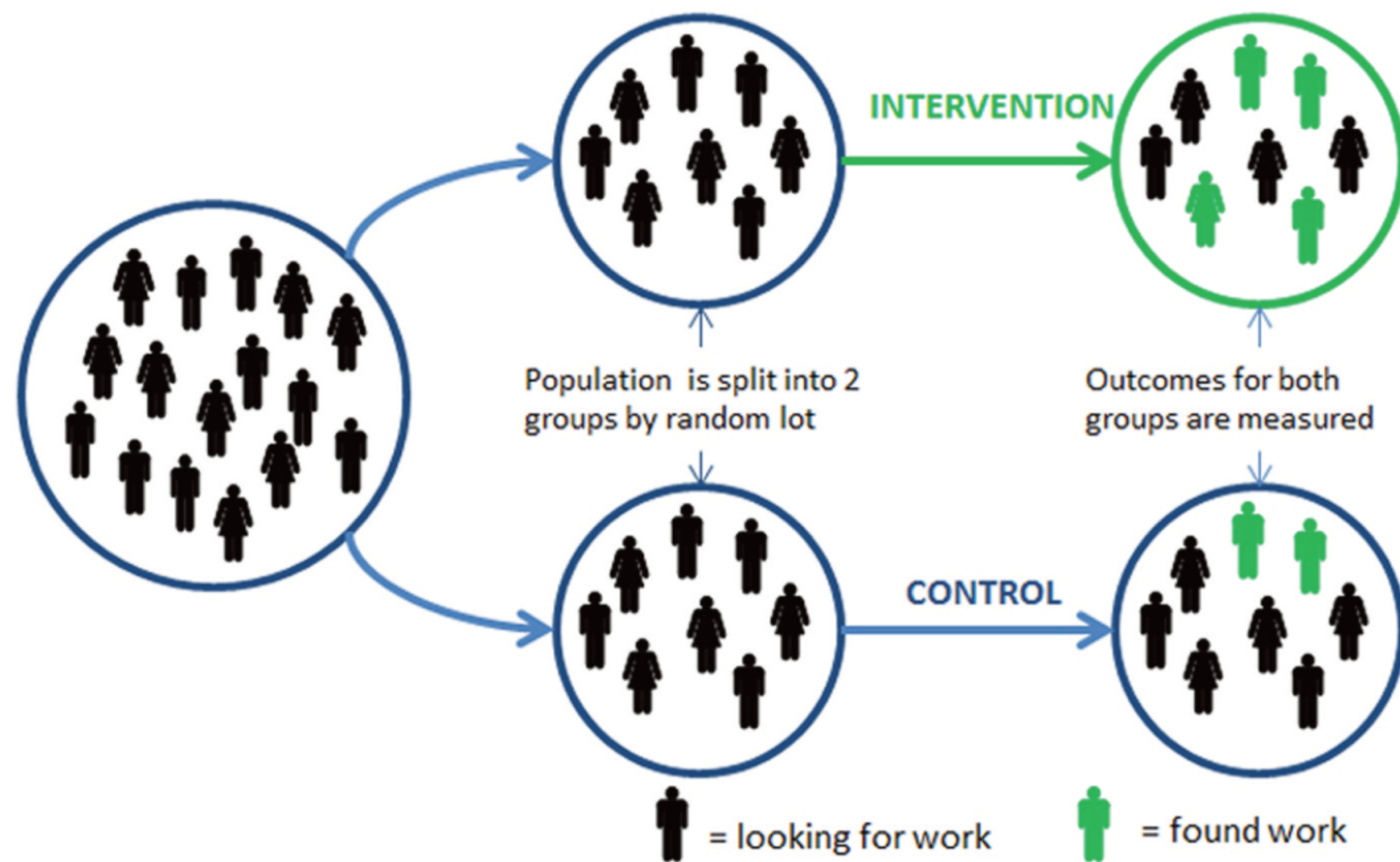
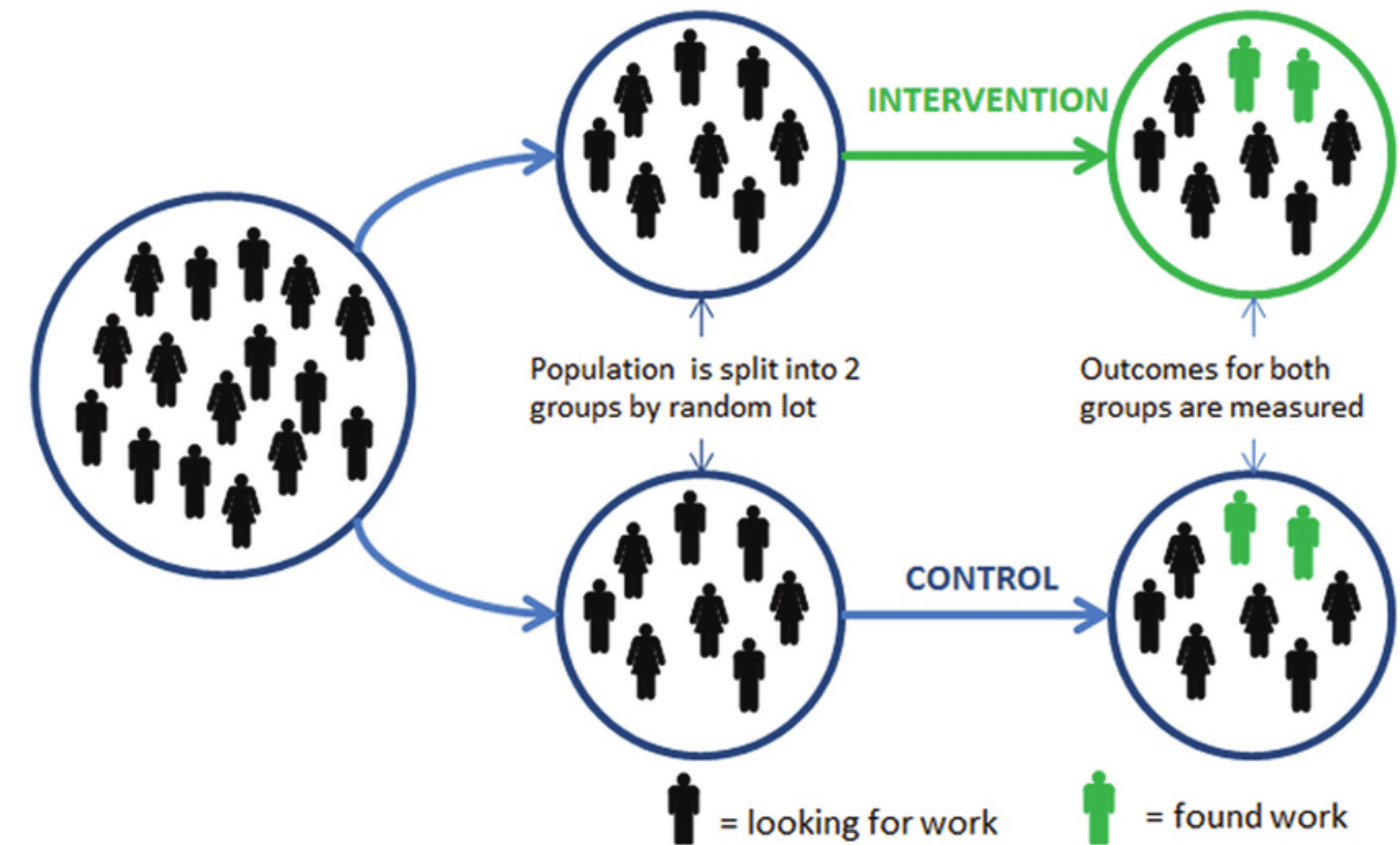
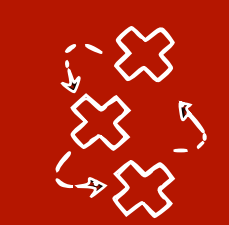
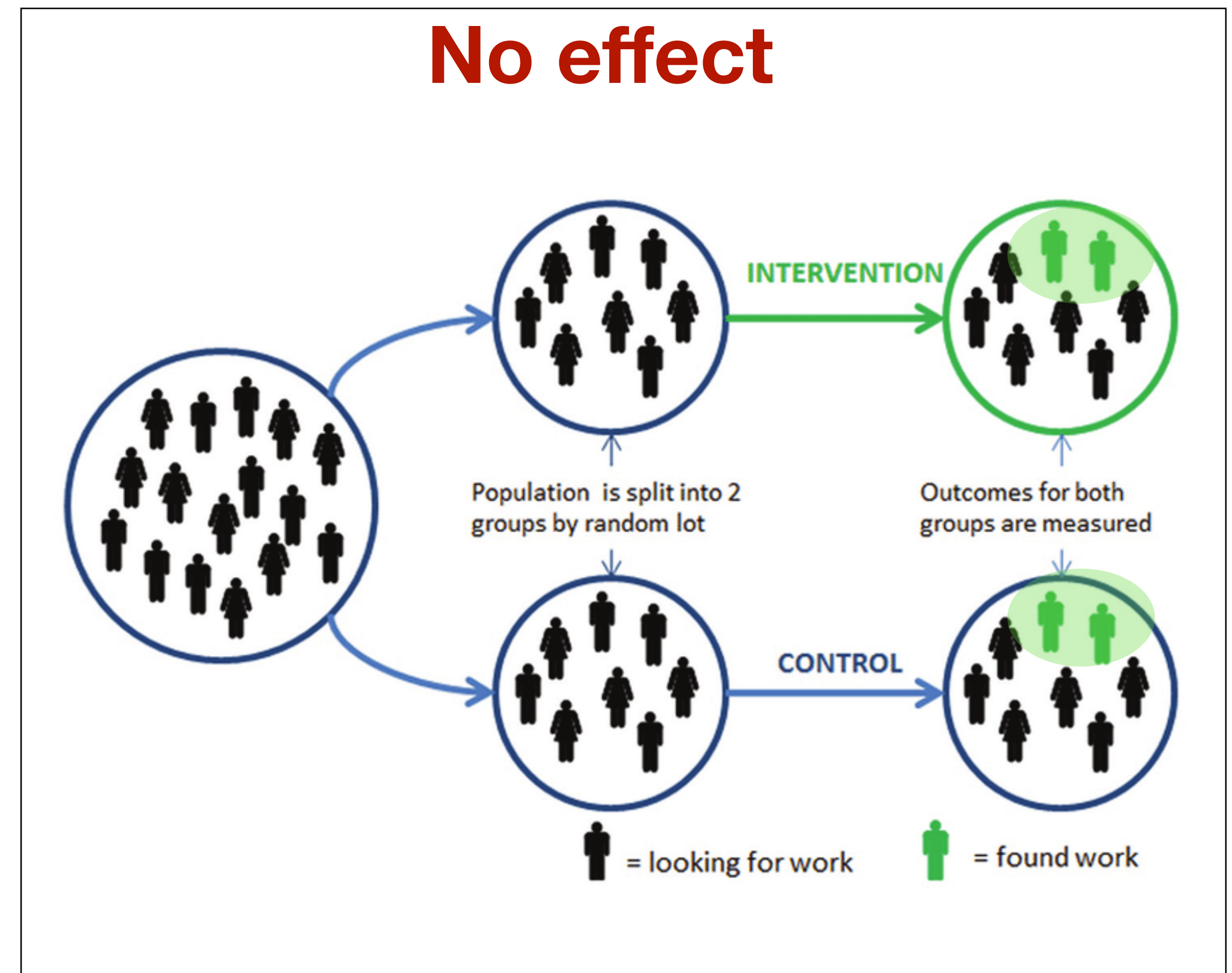
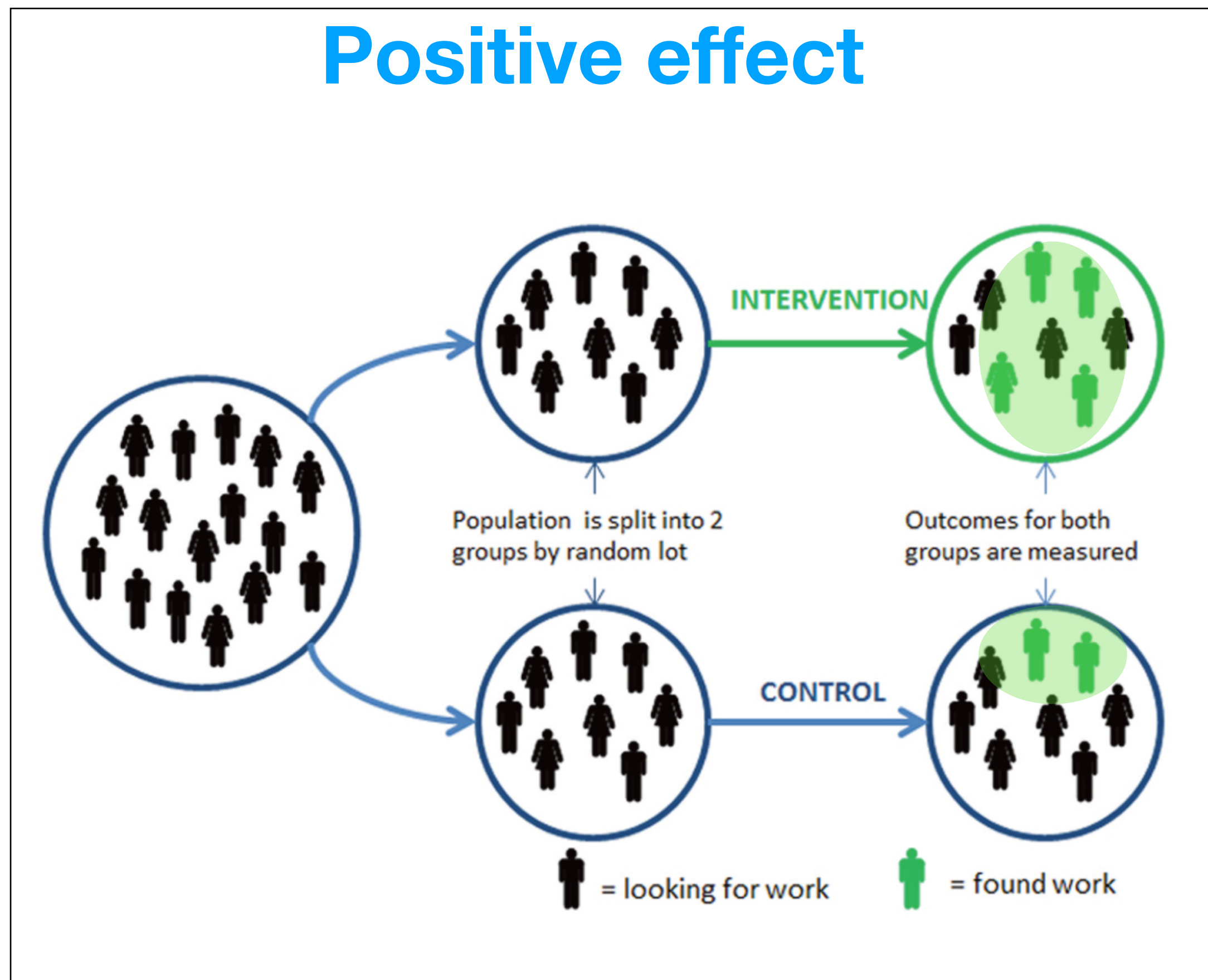


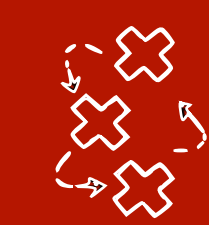
Figure B:





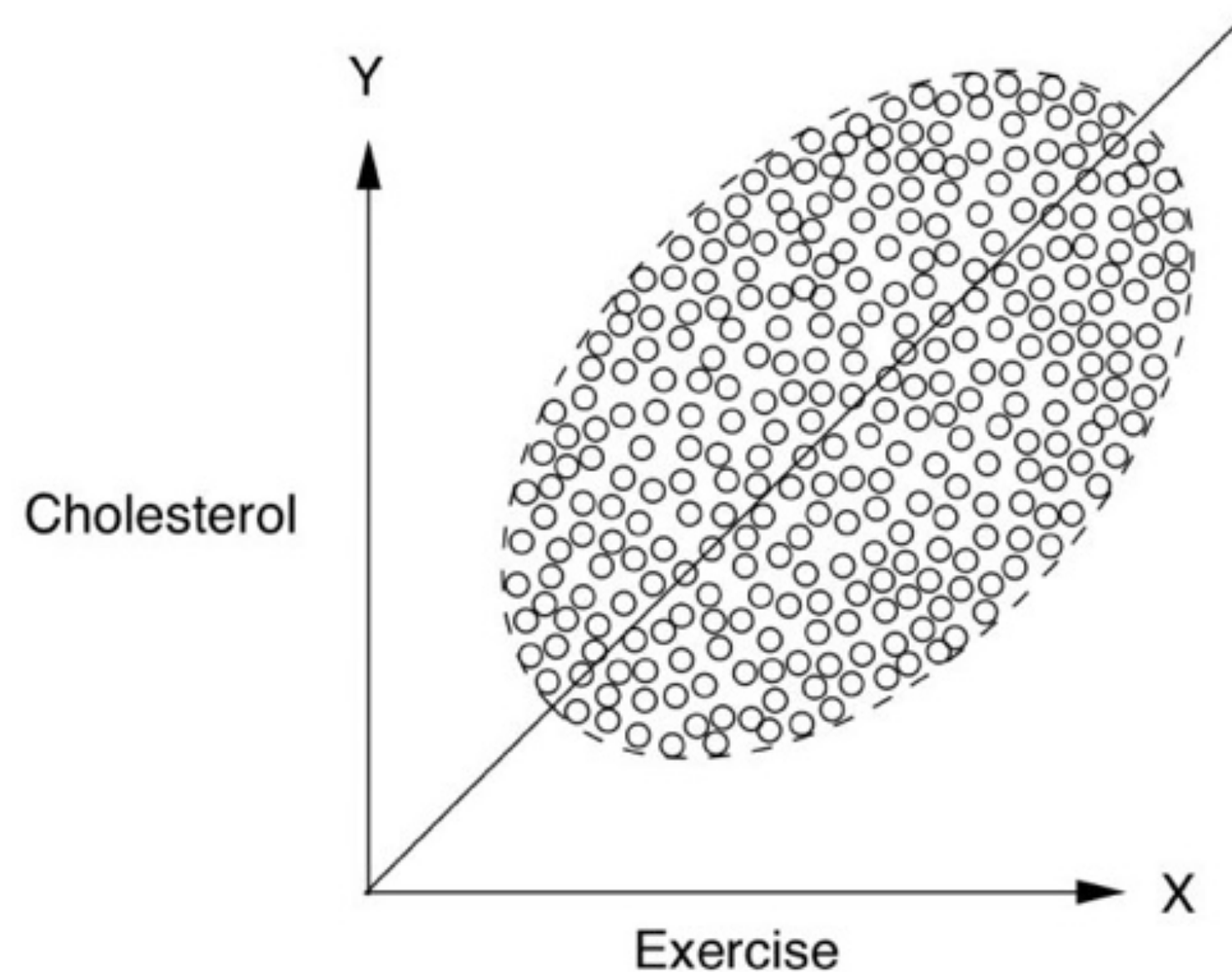
Randomised controlled trials (RCT): spot the difference?





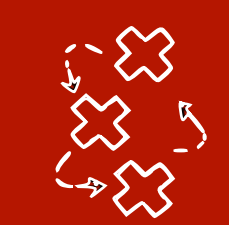
Observational data: when we do not have RCTs

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



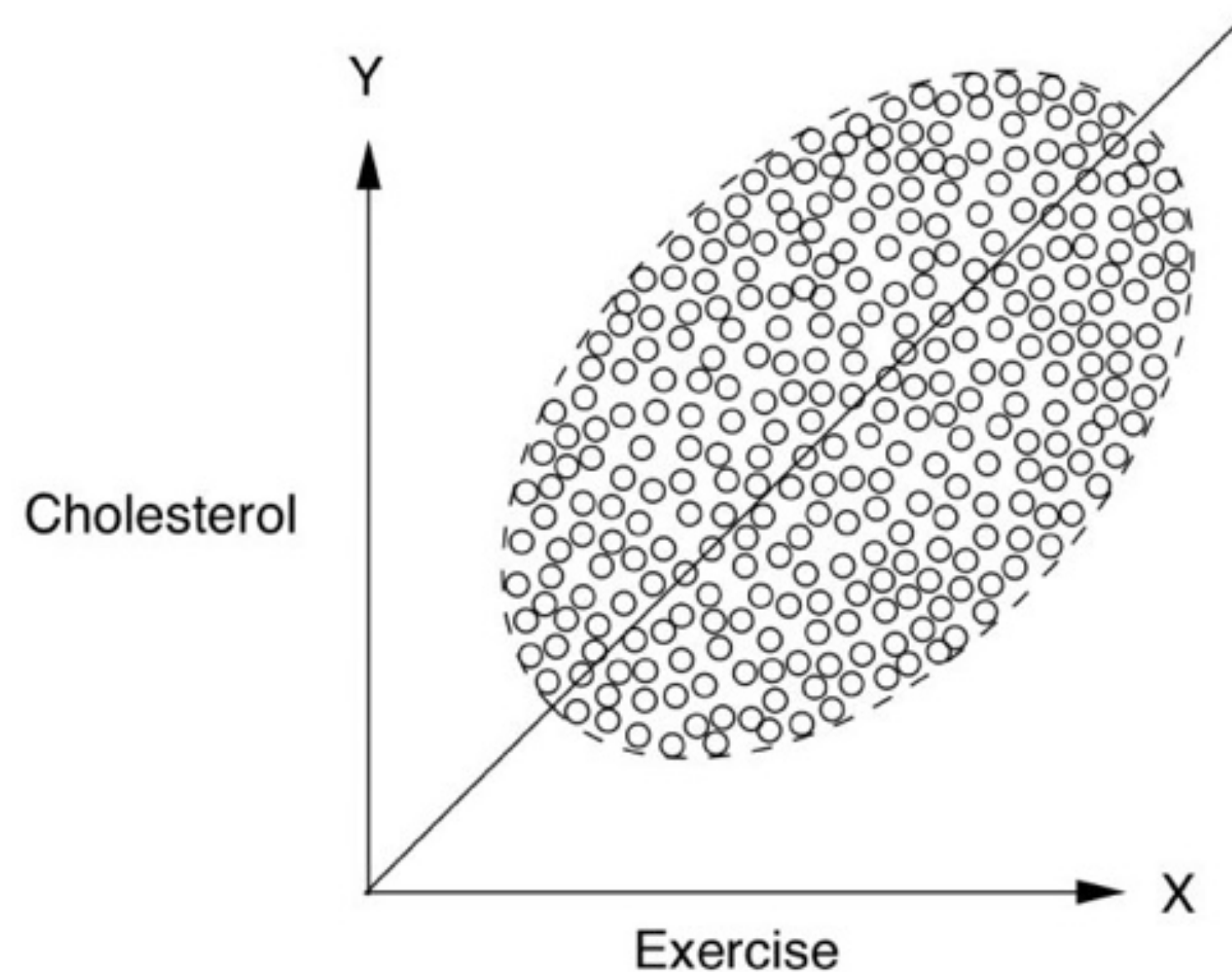
We don't know if there is a policy (and in case, which one) of how people decide to exercise.

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

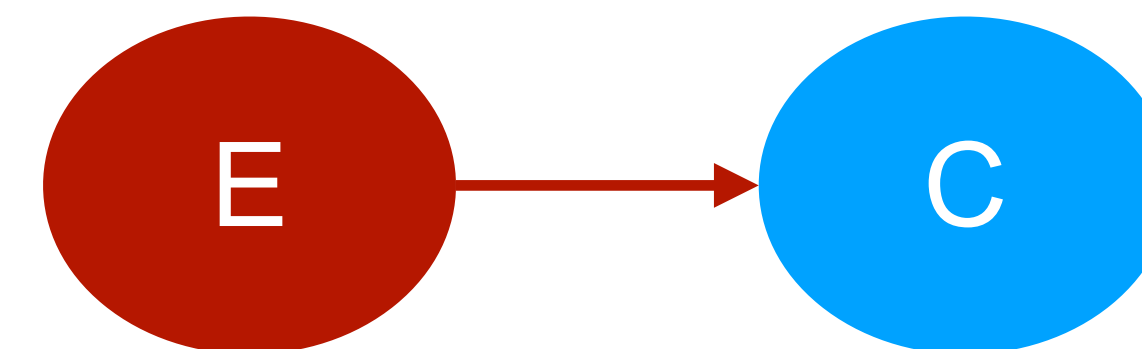


What if we don't have an RCT?

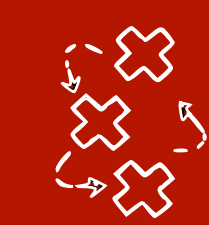
Let's assume we have **observational data** (e.g. data collected by hospitals)



Initial hypothesis:

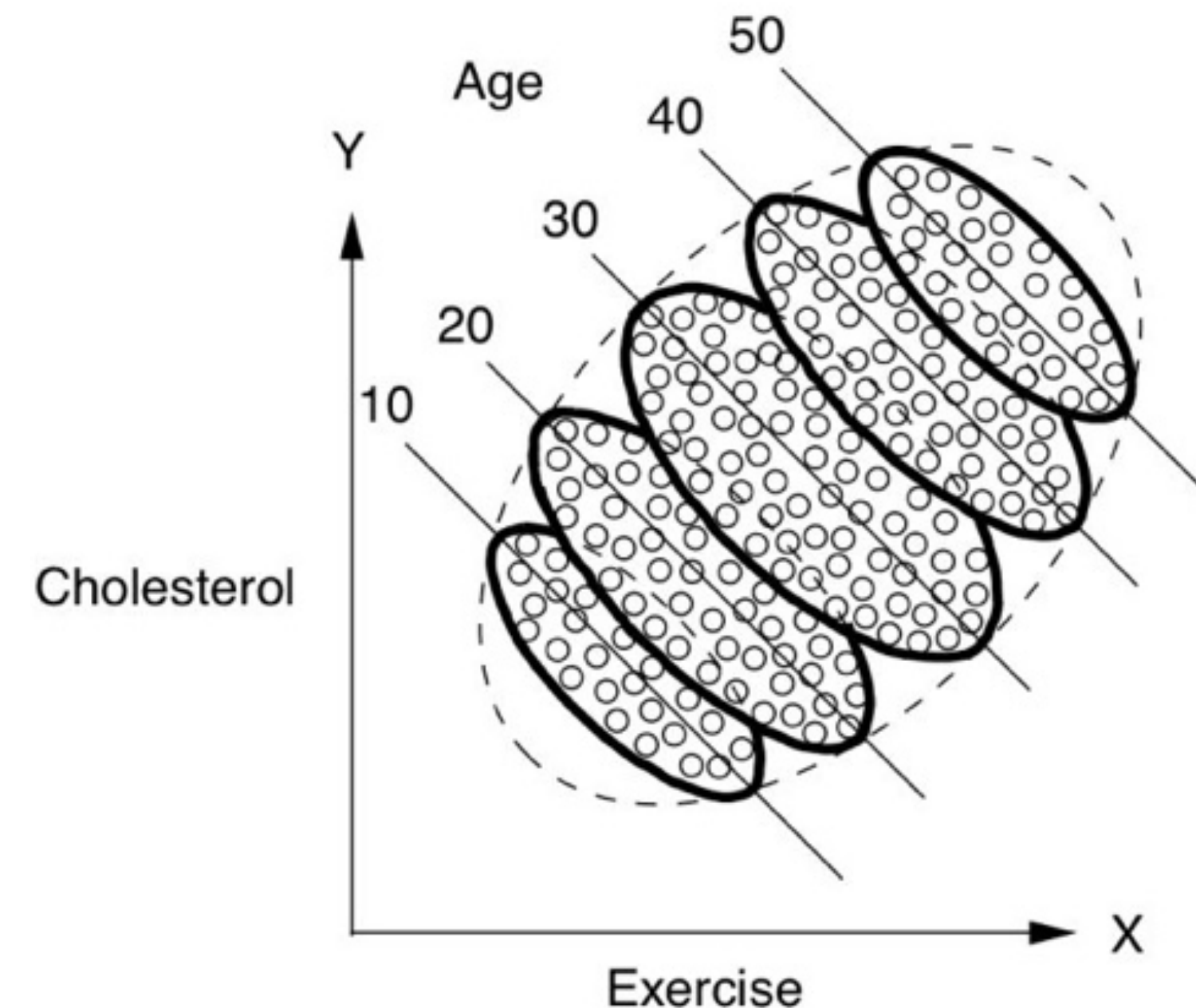
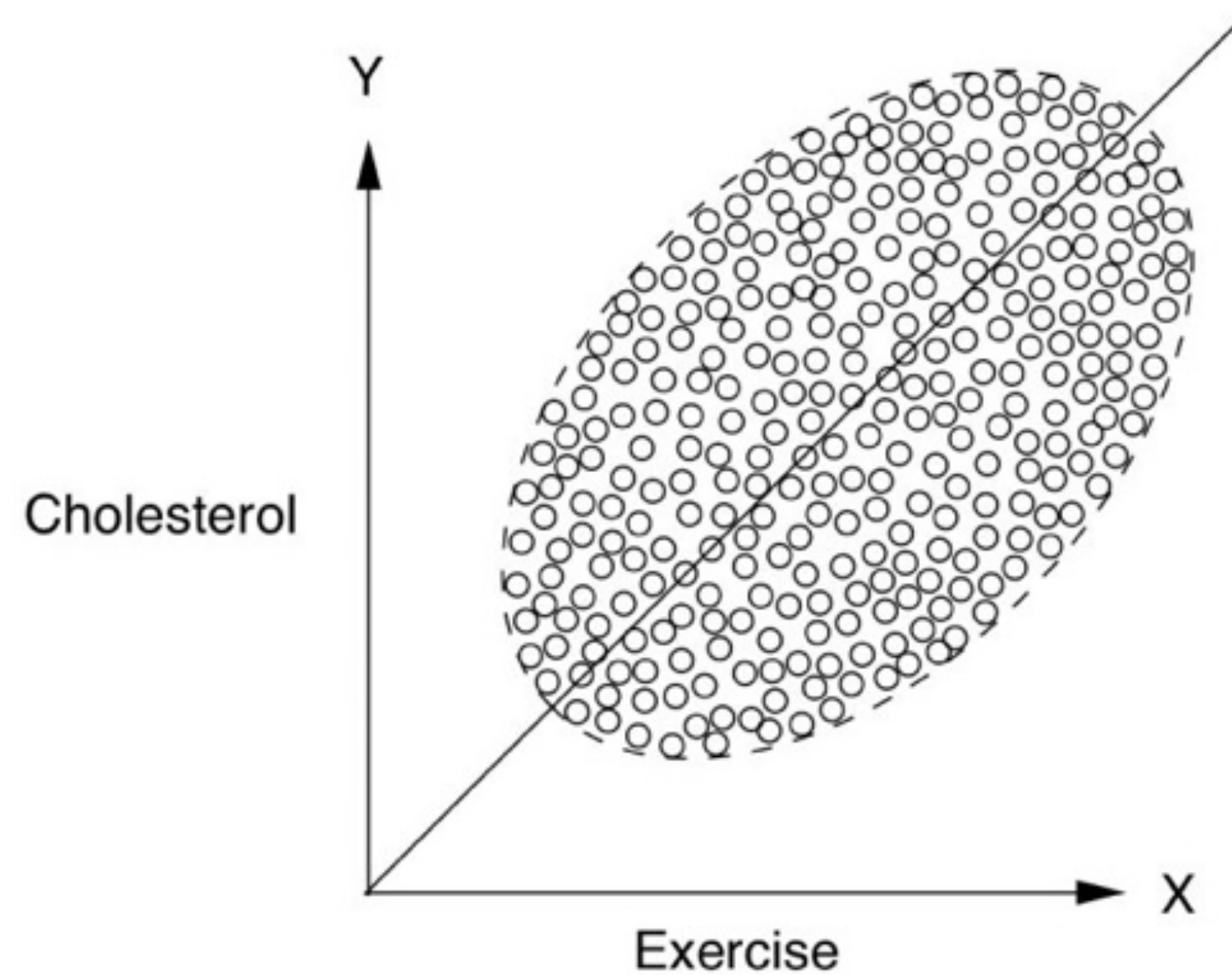


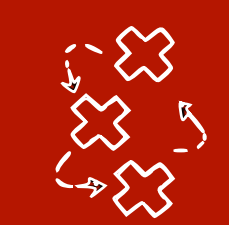
Exercise increases cholesterol



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

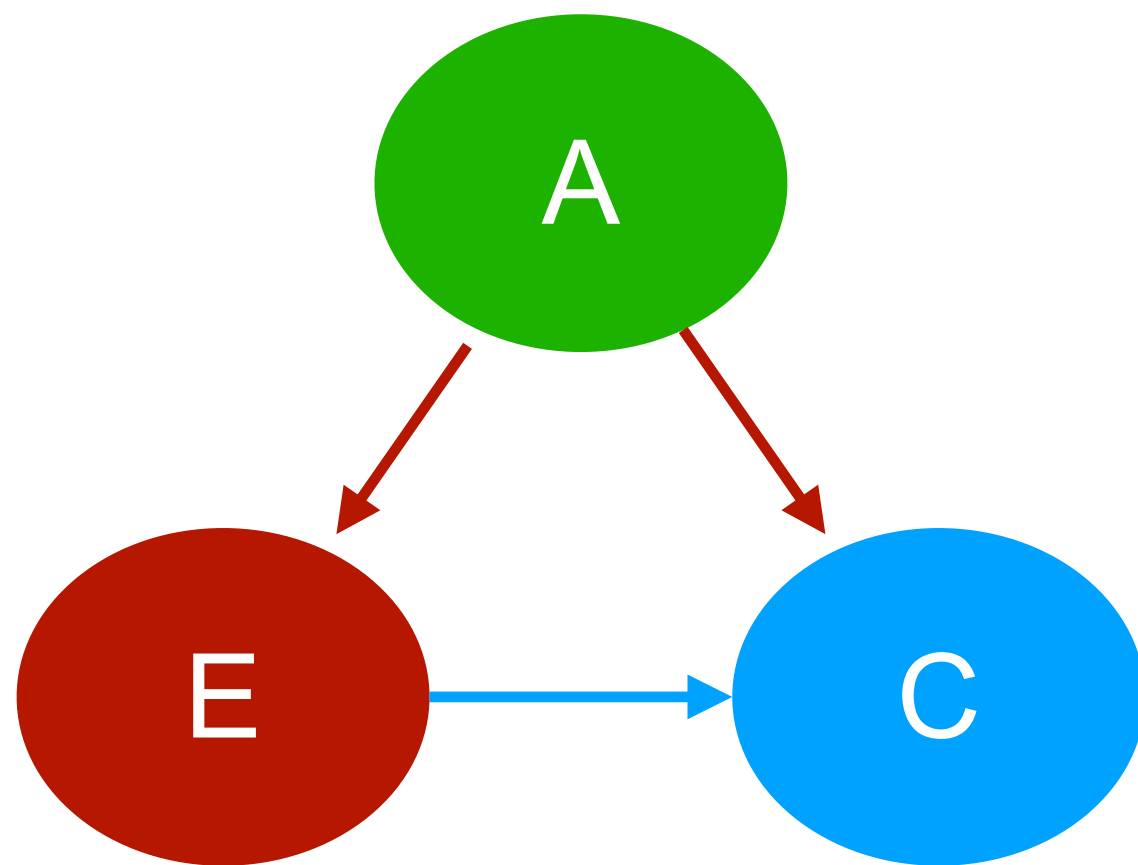
Let's assume we have **observational data** (e.g. data collected by hospitals)



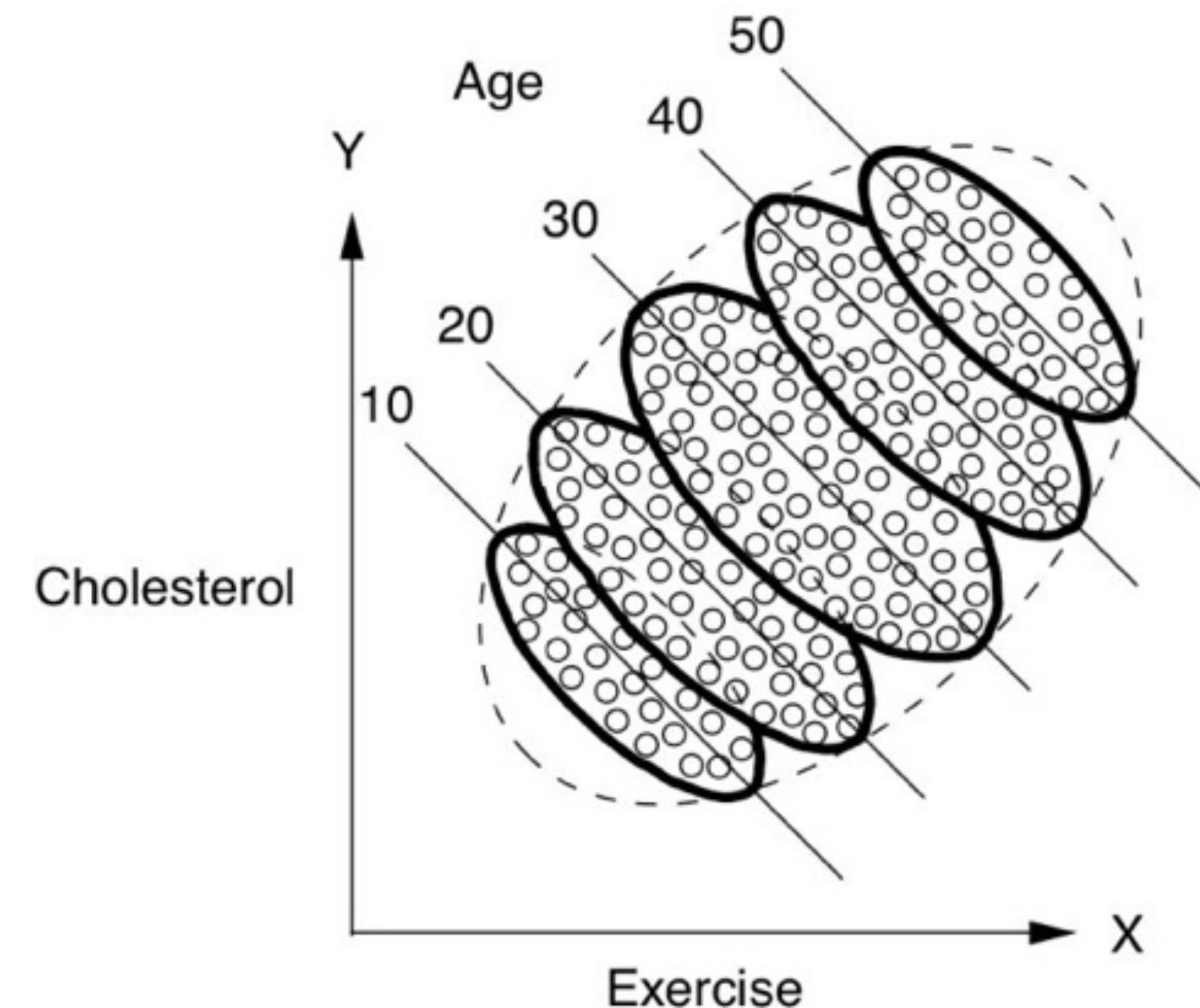


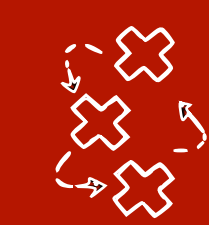
What if we don't have an RCT? Opposite conclusion

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



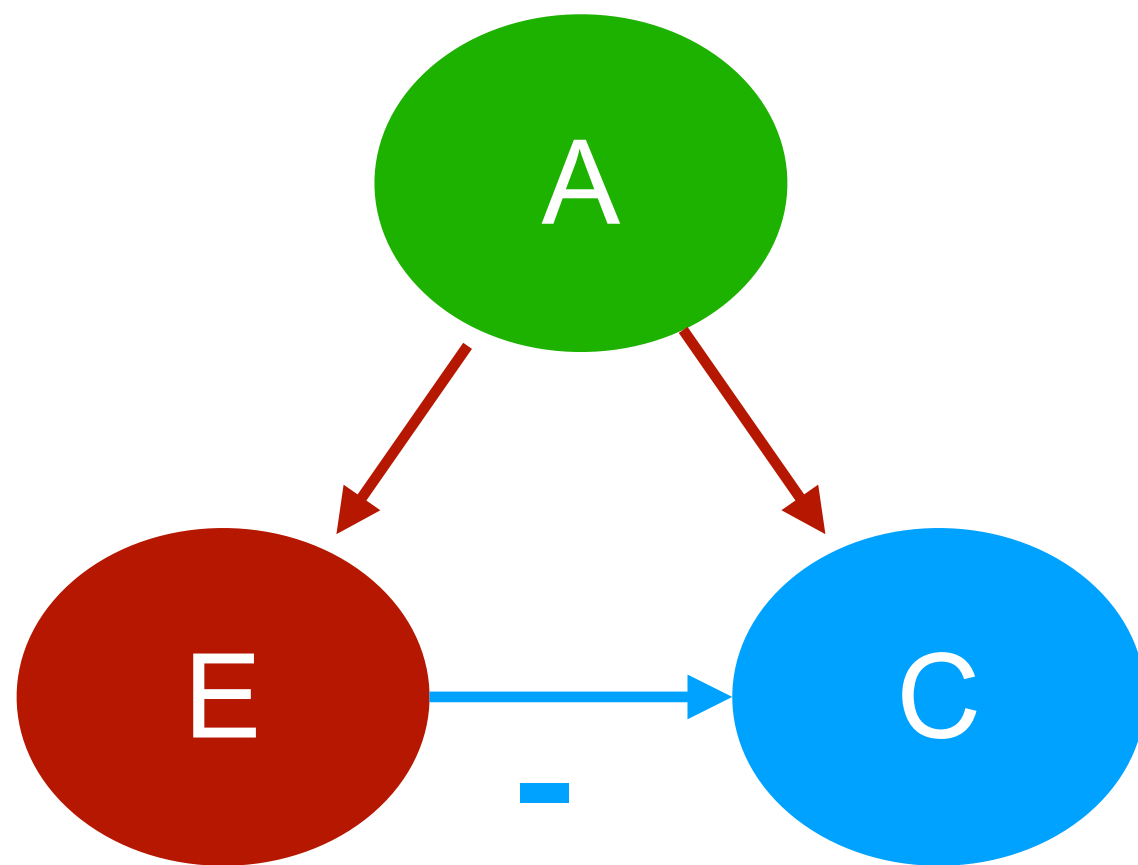
Exercise decreases cholesterol



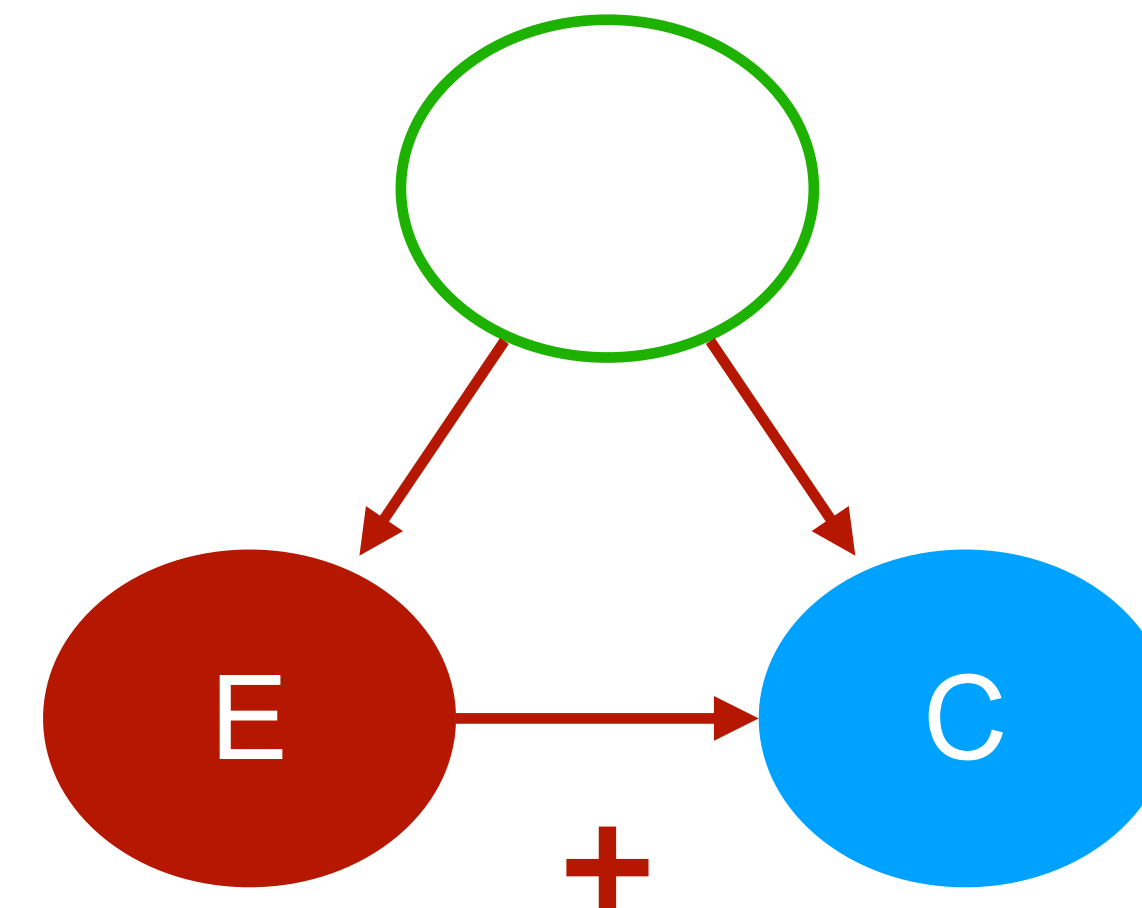


What if we don't have an RCT? Opposite conclusion

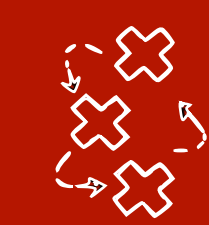
A is a confounder (a common cause of the treatment and outcome)



**Conditioning on A =
Controlling for confounding**

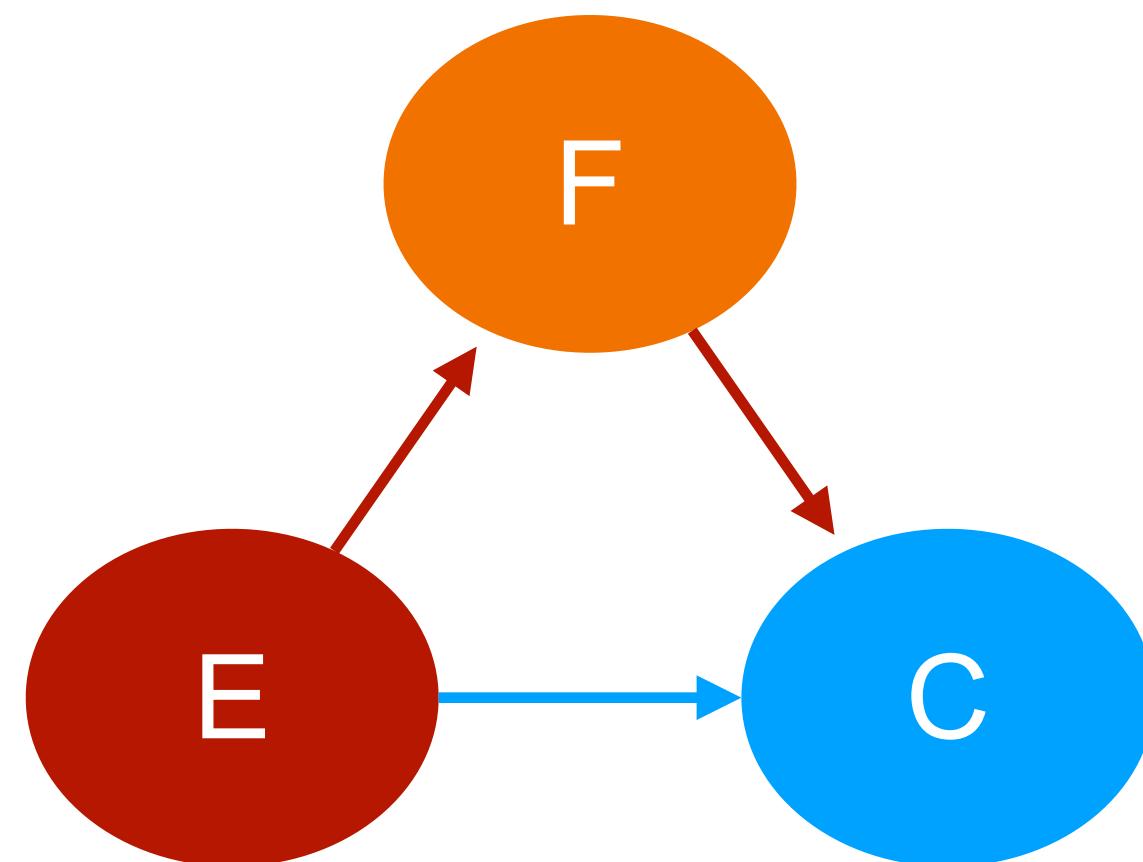


It confounds (biases) the effect, if we do not consider it



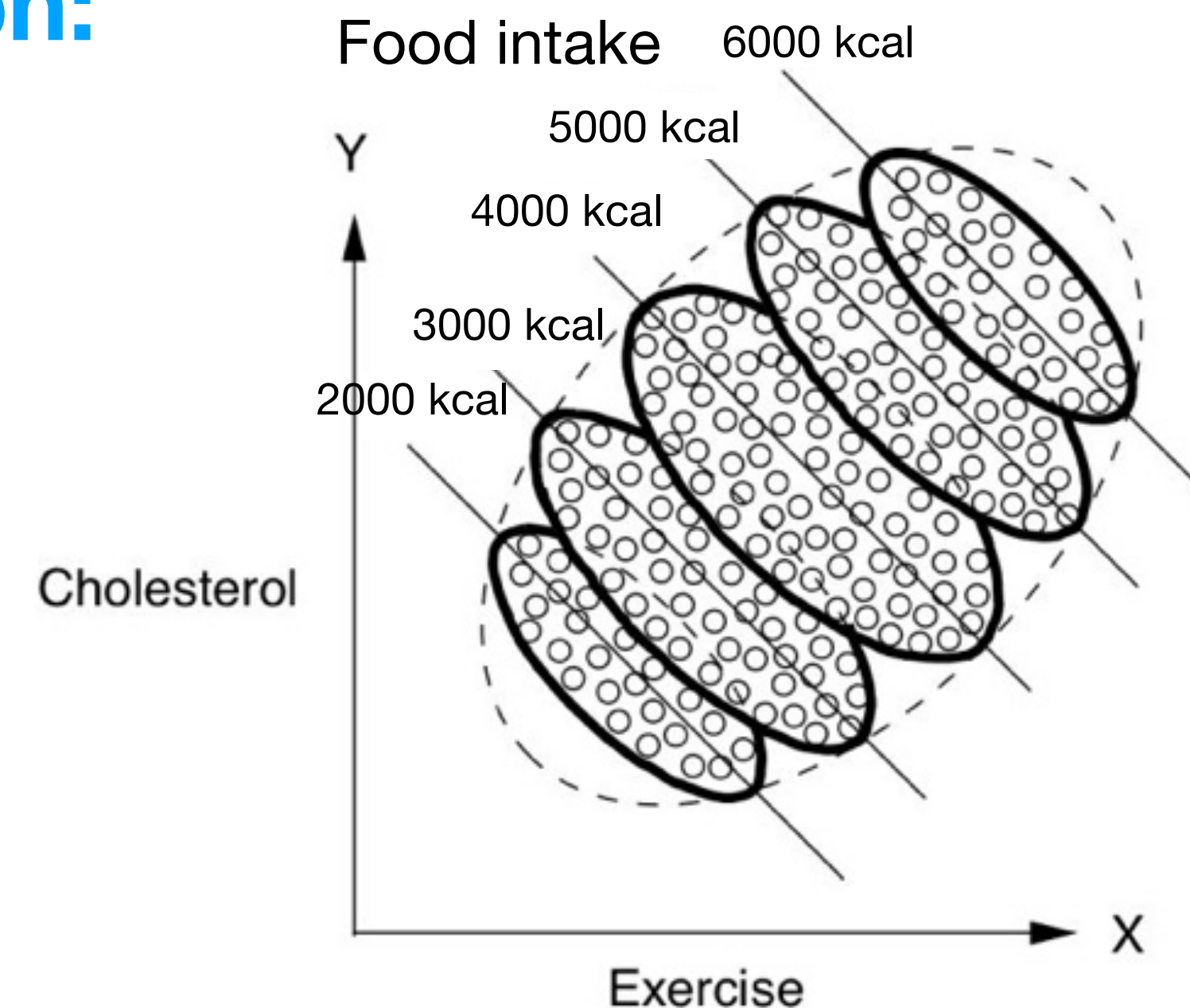
Shall we always just control on everything?

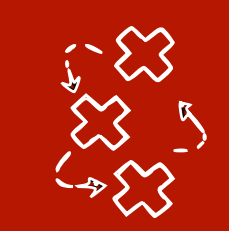
In alternative universe with this true graph:



Exercise **increases** cholesterol

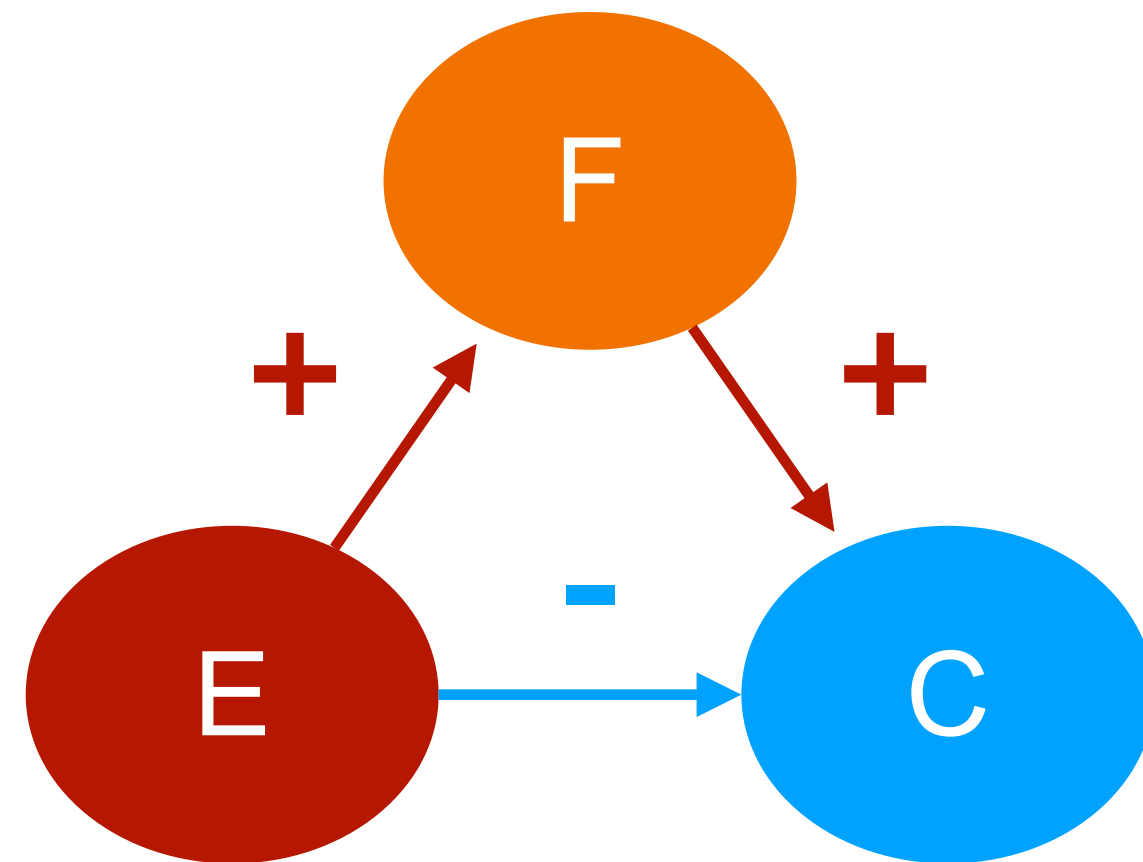
E has both a direct and indirect (through F) effect on C, conditioning on C removes the indirect effect





Shall we always just control on everything?

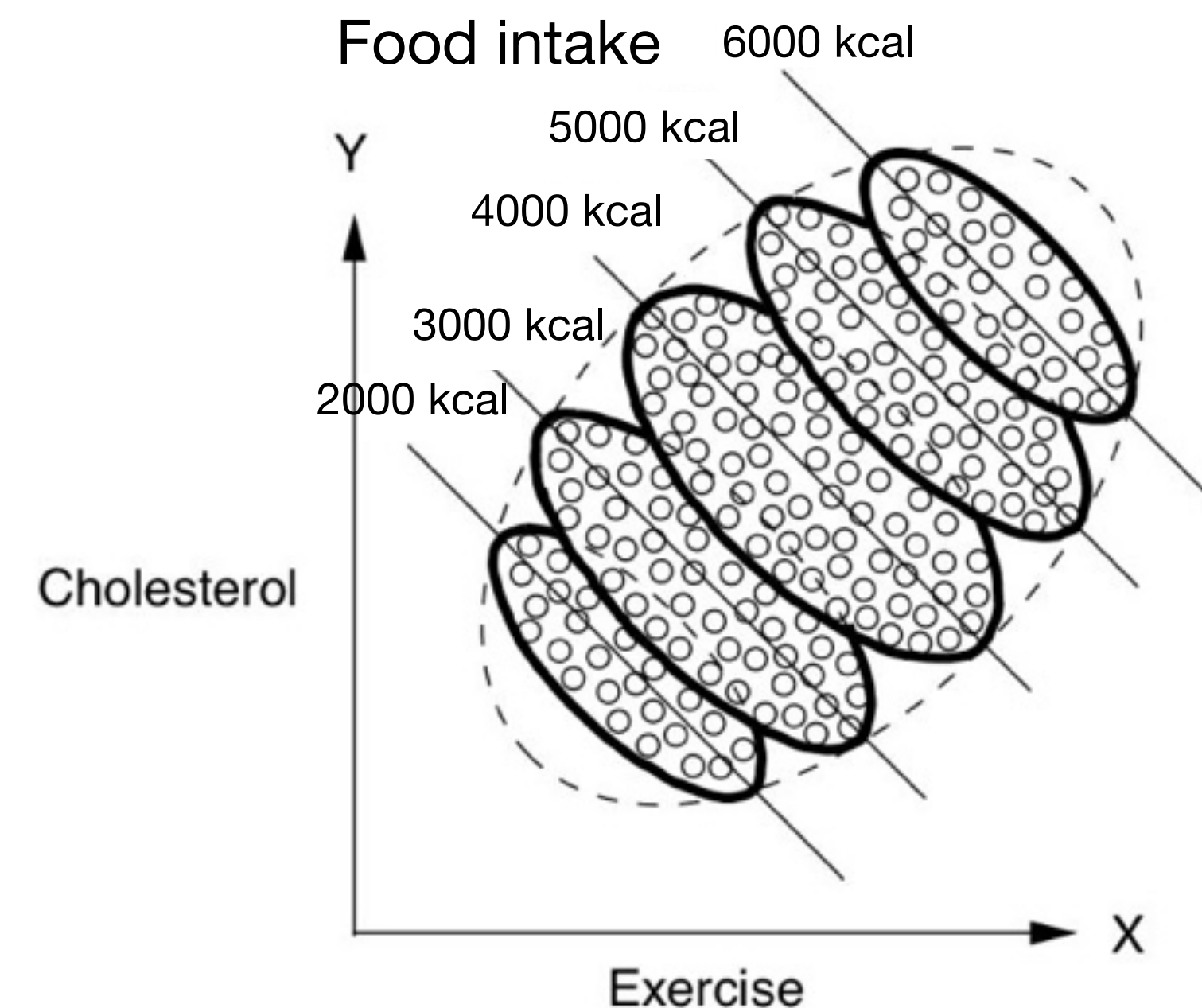
In alternative universe:

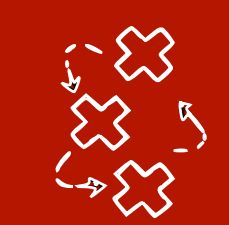


Exercise **increases** cholesterol

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

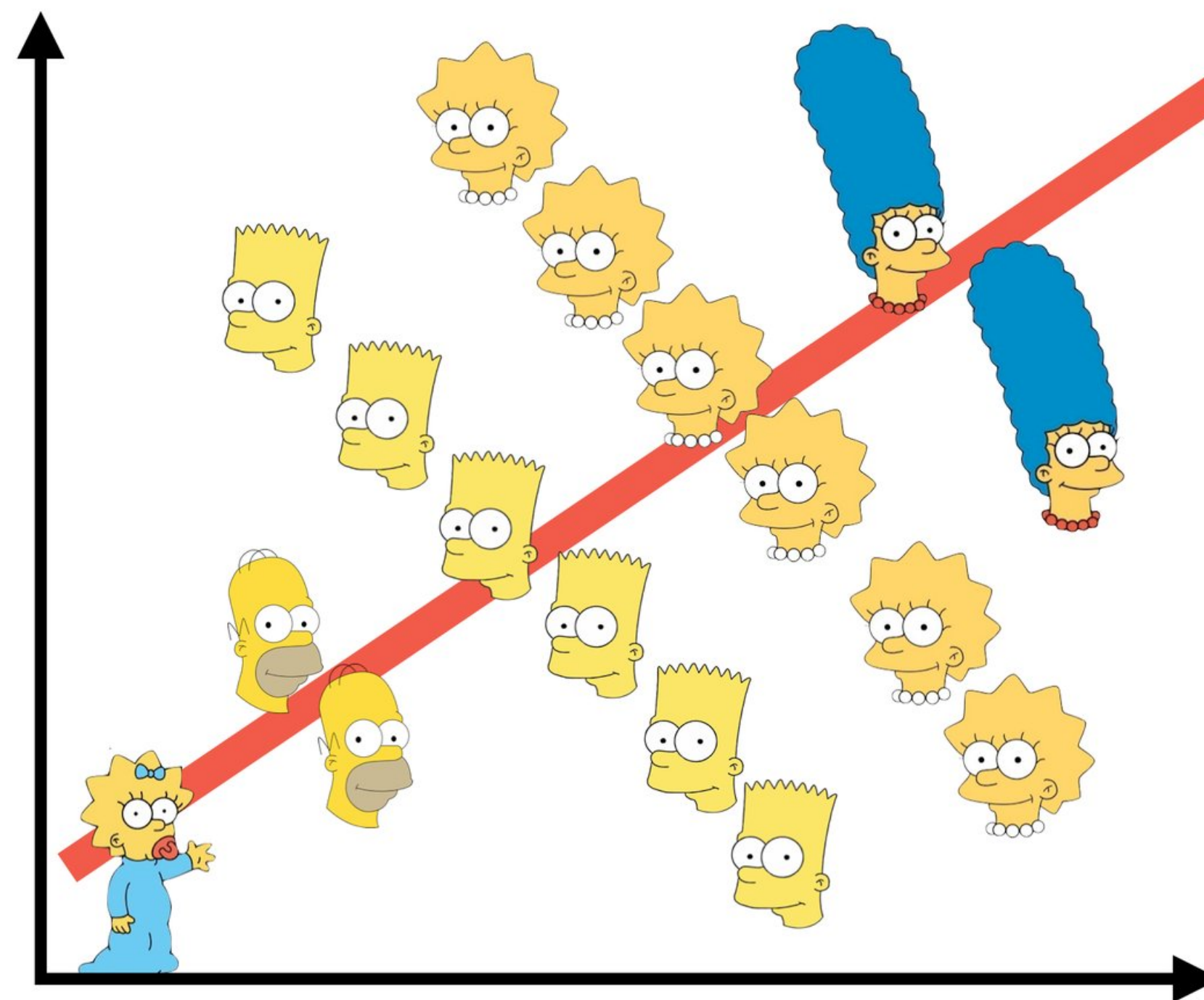
We will see later in the course on which **covariates** we should **adjust**.

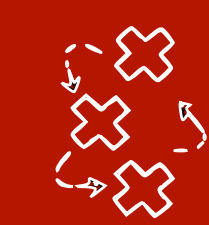




Simpson's paradox (not really these Simpsons)

In order to know what to control/adjust for so we get the correct causal effect, we need to know the underlying causal graph

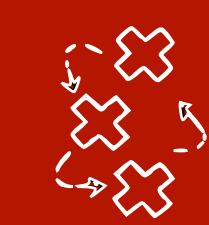




A working definition of causality in machine learning

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

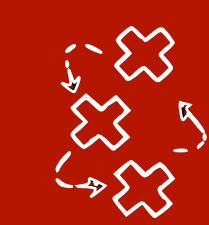


A working definition of causality in machine learning

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. we have observational data)**



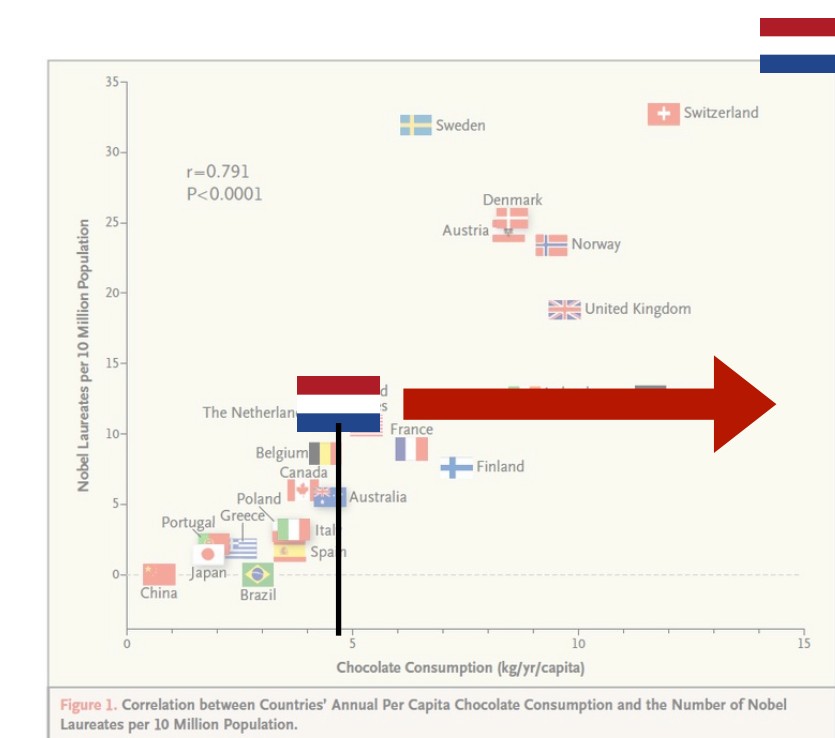
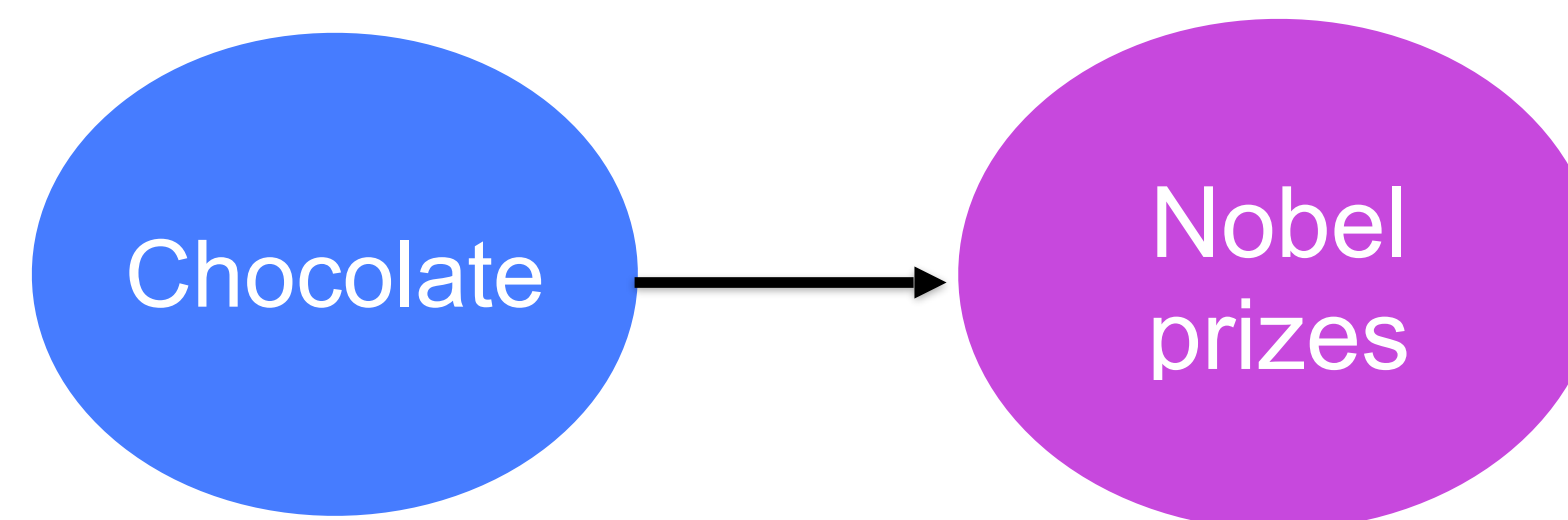
A working definition of causality in machine learning

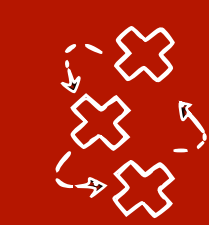
Informal definition: A variable X causes another variable Y , if **changing (the distribution of) X** , e.g. by fixing its value, changes (the distribution of) Y

Intervention

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

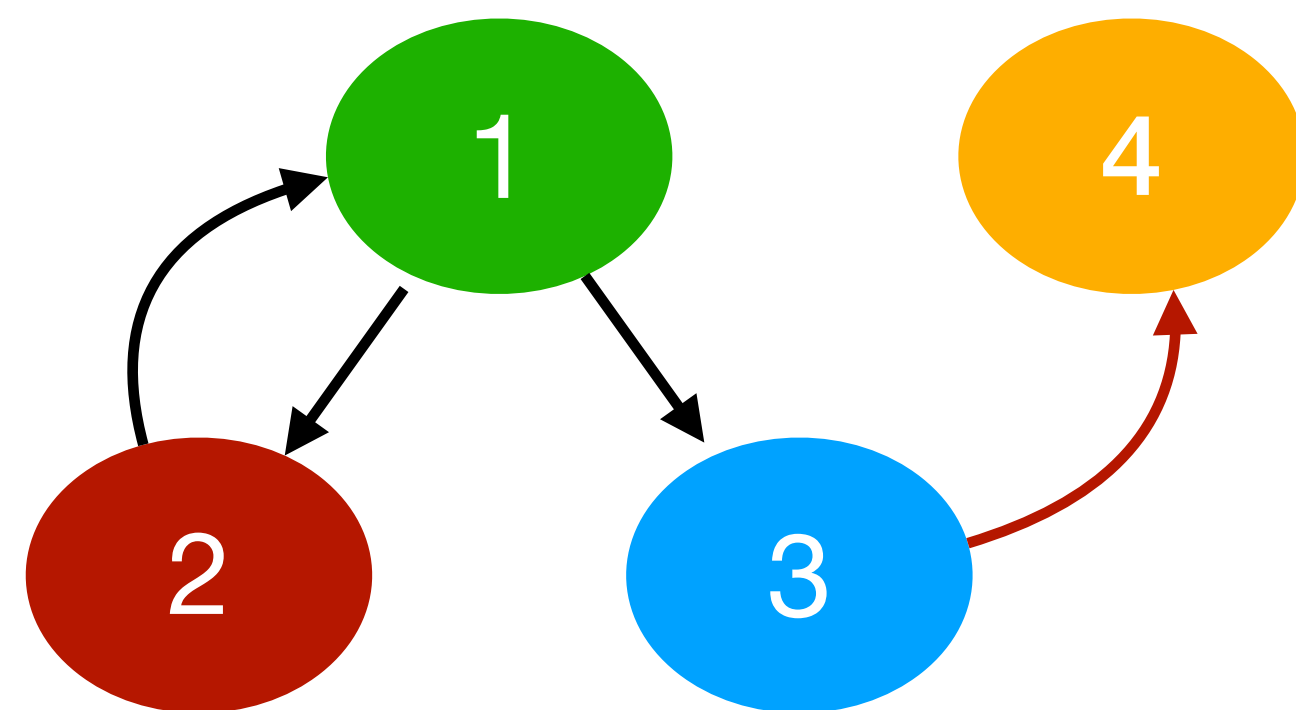
Representation: We can represent causal relations in **causal graphs:** nodes are random variables, edges causal relations



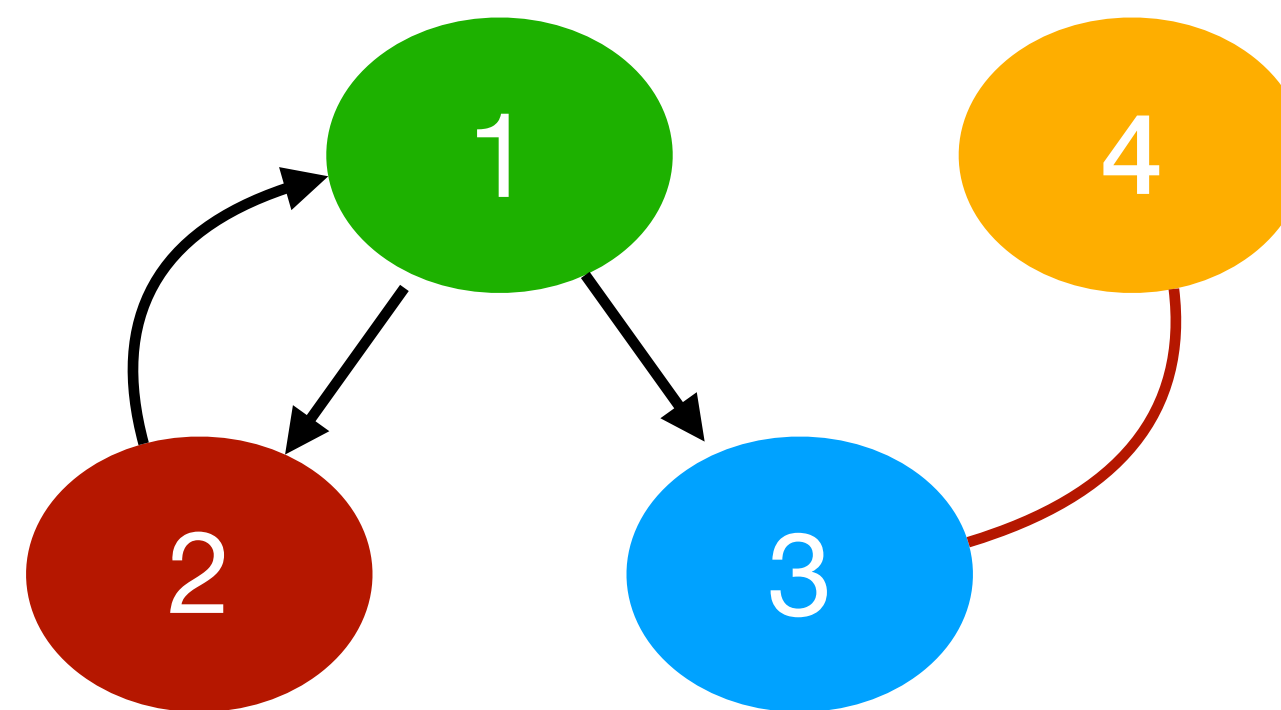


Basic types of graphs in causality

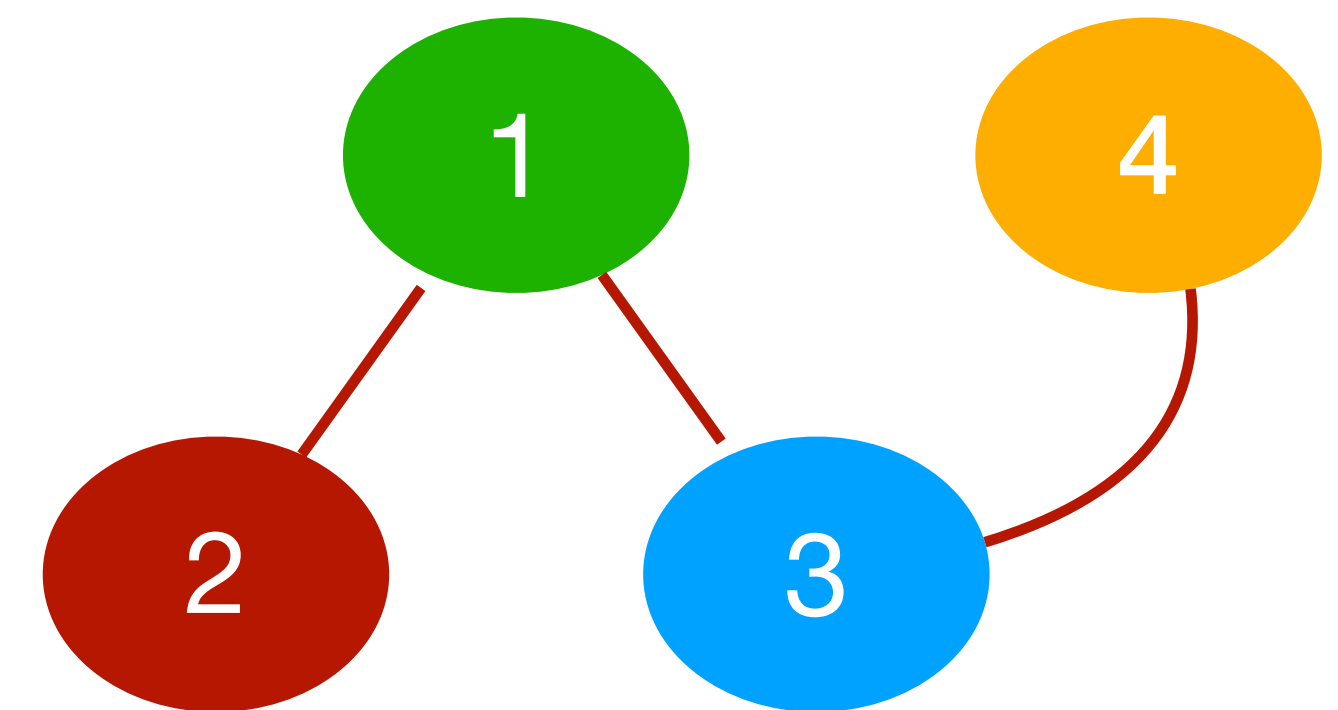
- A graph G is a tuple $G = (\mathbf{V}, \mathbf{E})$:
 - $\mathbf{V}$ is the set of **nodes** (vertices)
 - $\mathbf{E}$ is the set of **edges** between two nodes, i.e. $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$
 - If all edges are **directed** $\rightarrow$ then the graph is **directed**



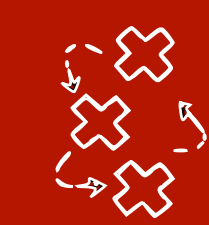
Directed graph



Mixed graph



Undirected graph



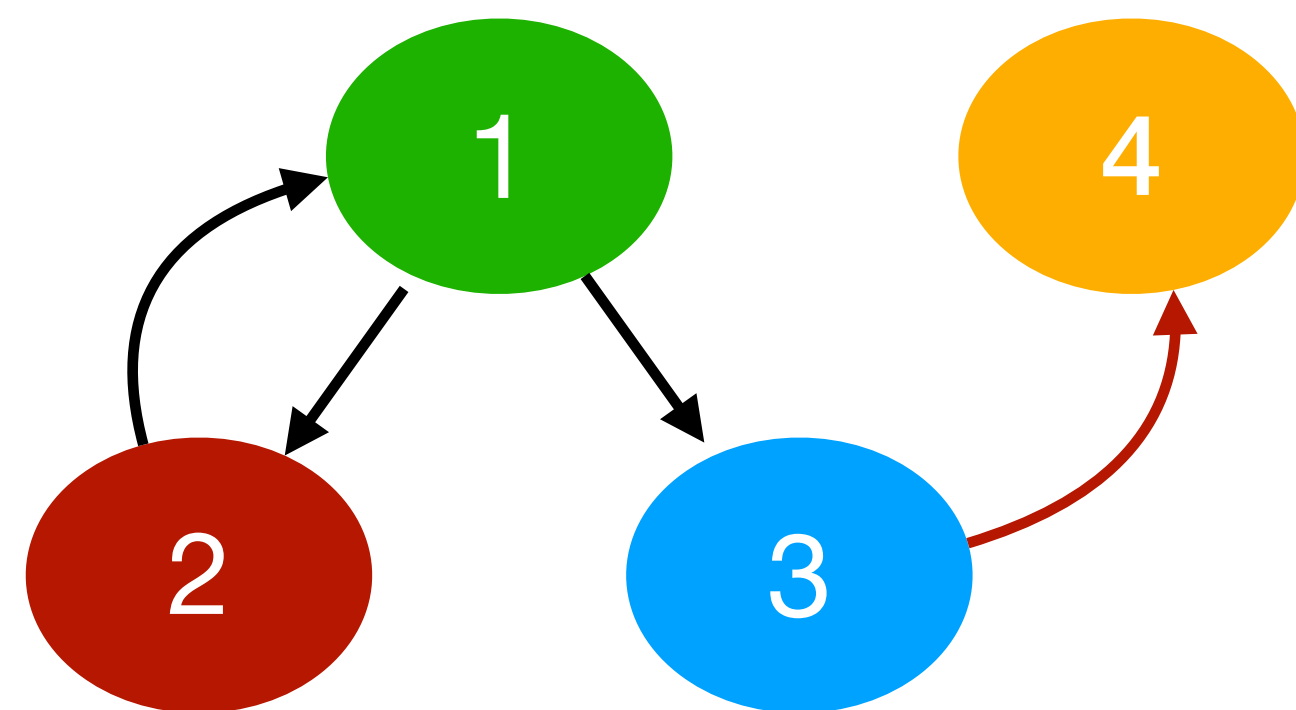
Basic types of graphs in causality

- A graph G is a tuple $G = (\mathbf{V}, \mathbf{E})$:

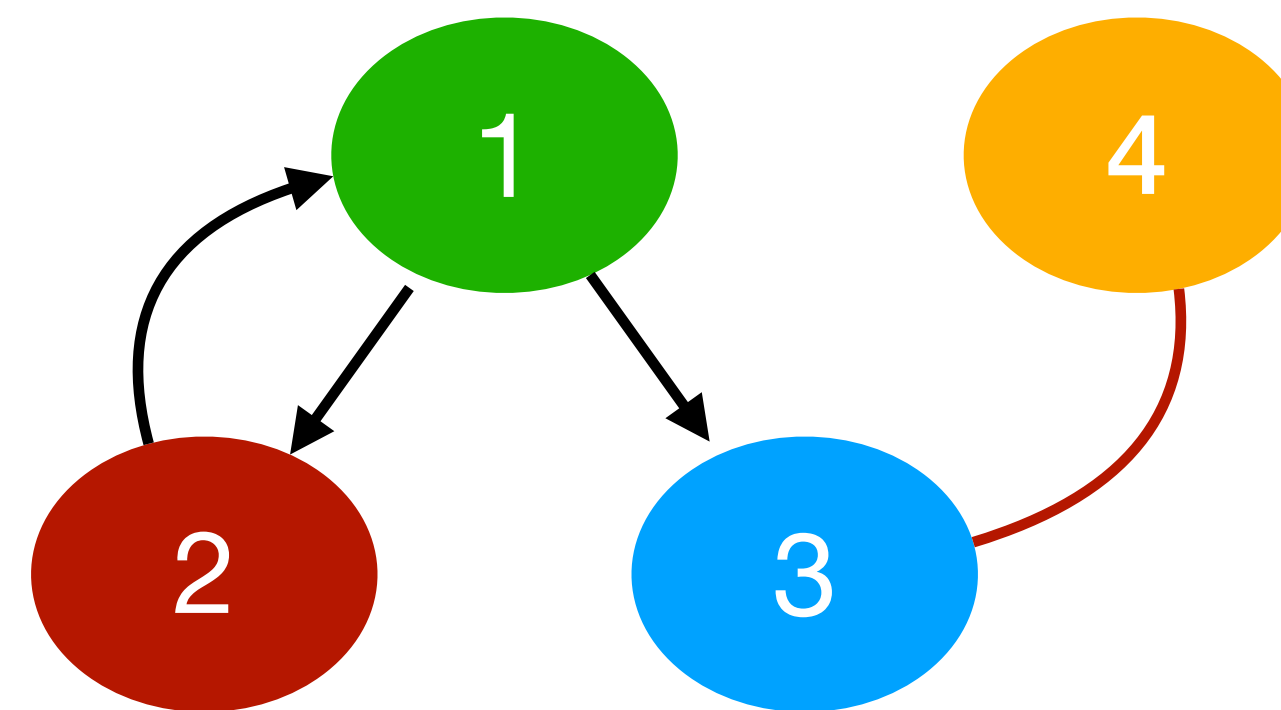
- $\mathbf{V}$ (vertices)

We assume the ground truth graph is directed (and often acyclic, a DAG)

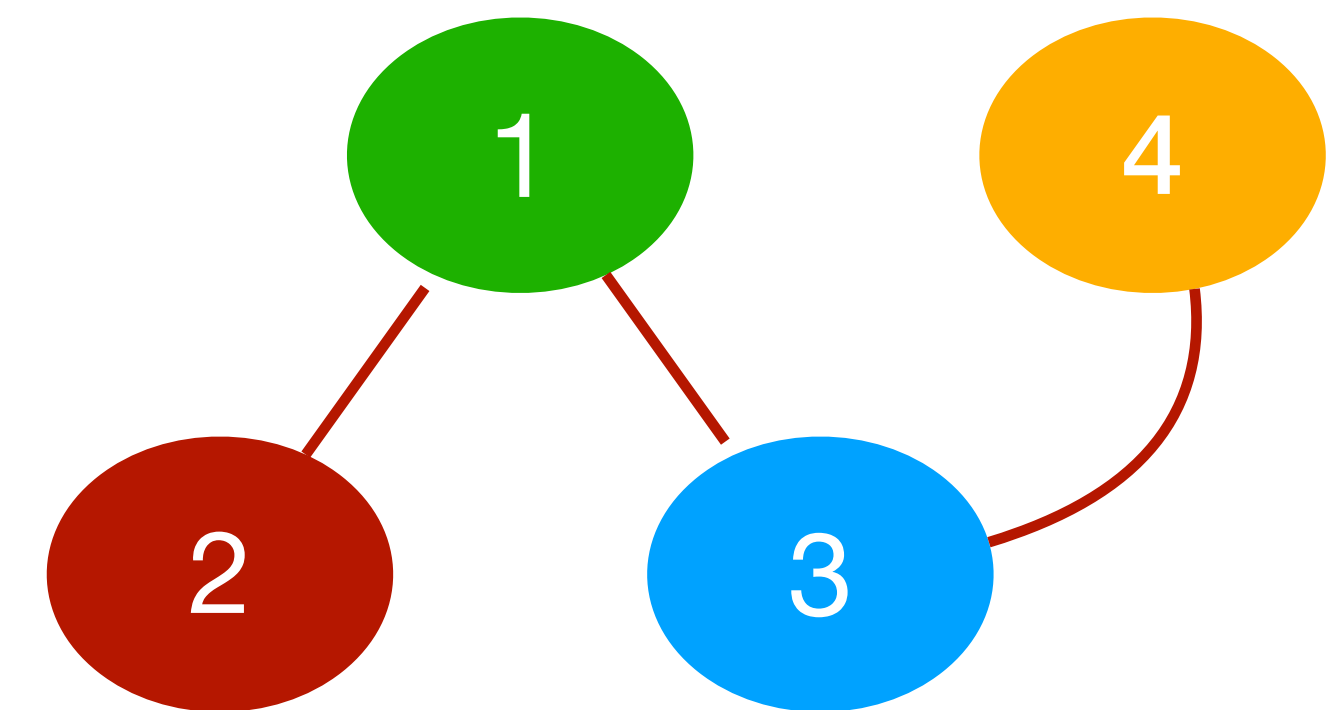
These two types are useful when we are progressively learning causal relations from data or to represent sets of graphs



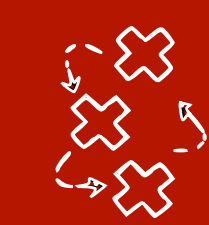
Directed graph



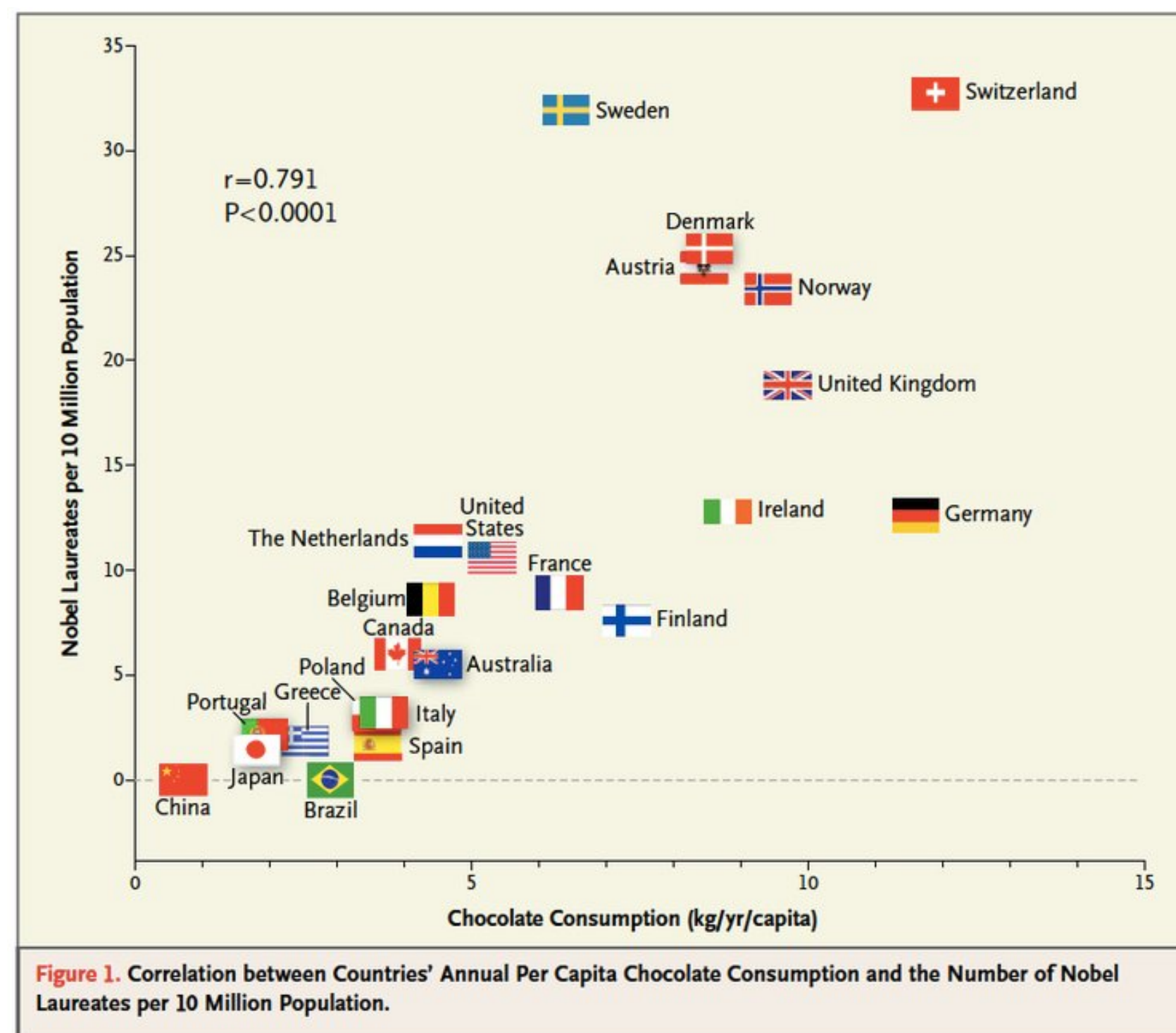
Mixed graph



Undirected graph



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



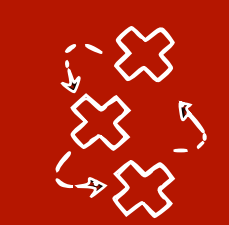
Reichenbach's principle of common cause:

A correlation between X and Y implies that

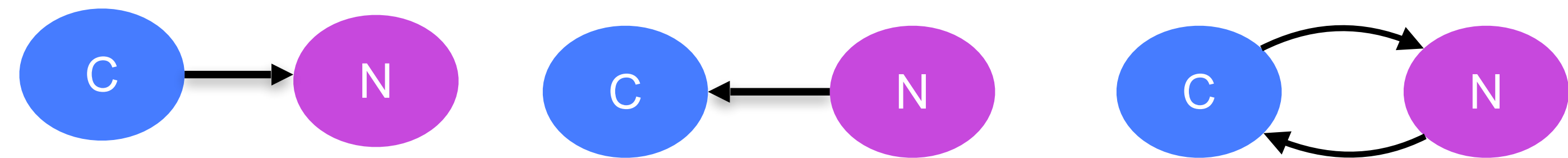
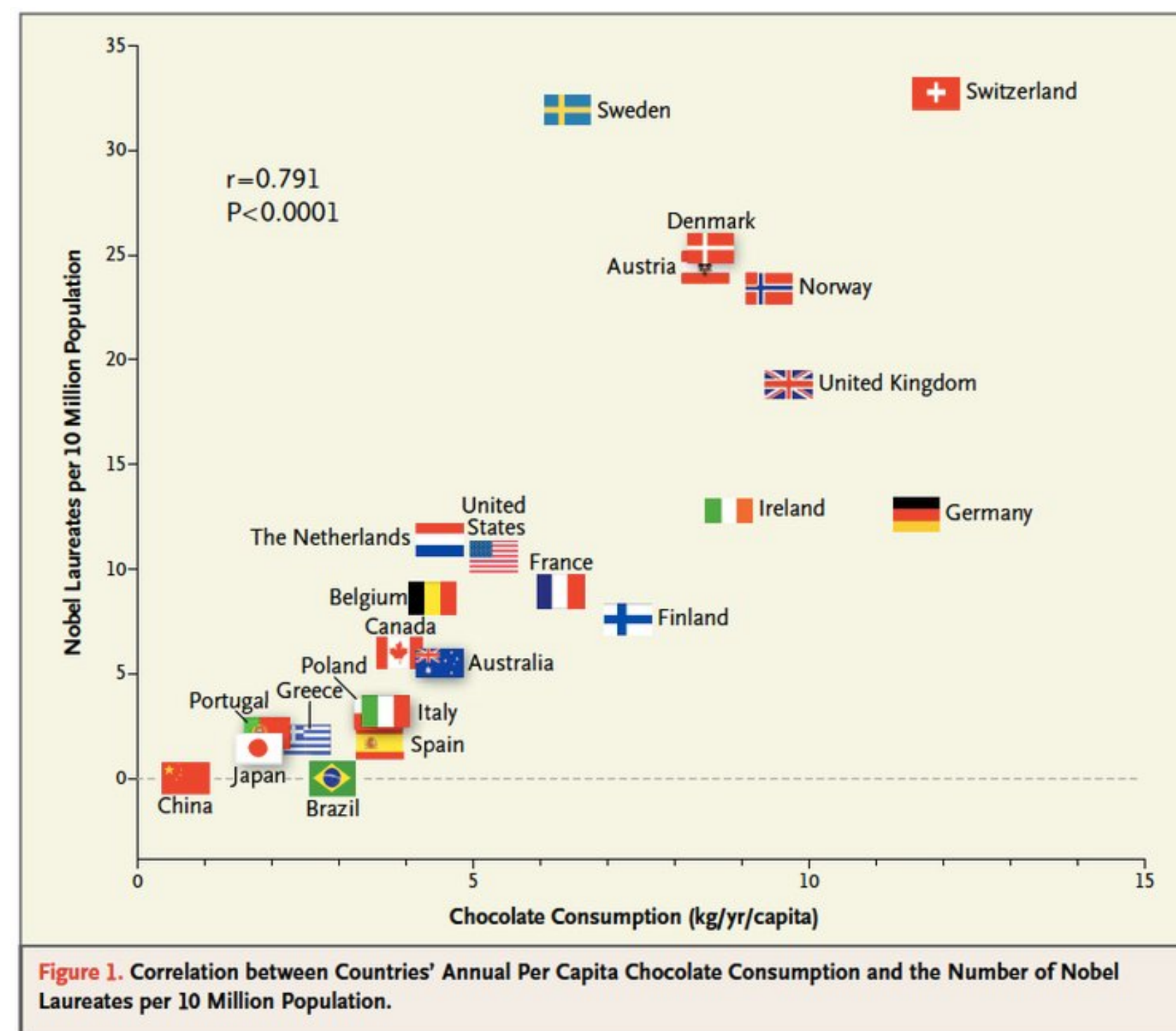
- X causes Y, or
- Y causes X, or
- There exists a common cause between X and Y

(or any combination of the above)

((for now, we ignore selection bias))

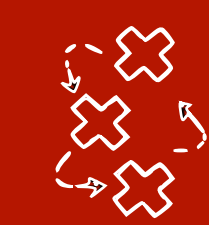


What if we don't know the graph? Causal discovery

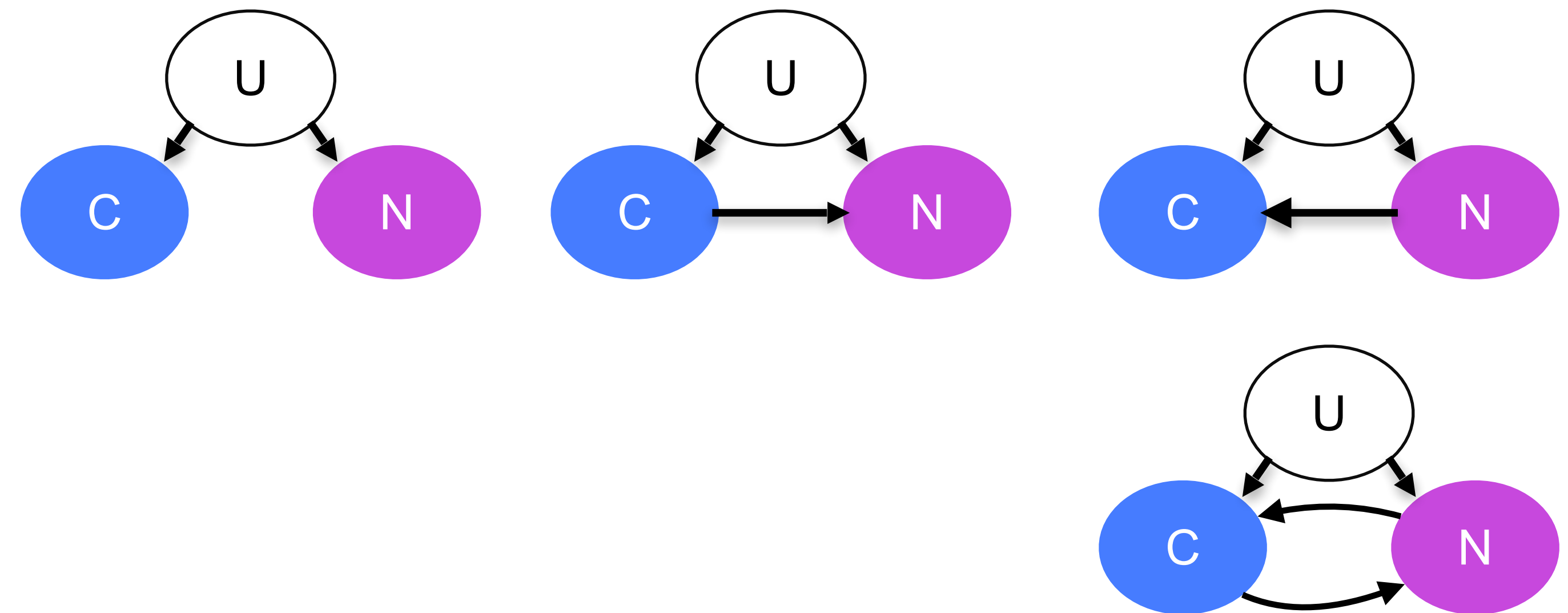
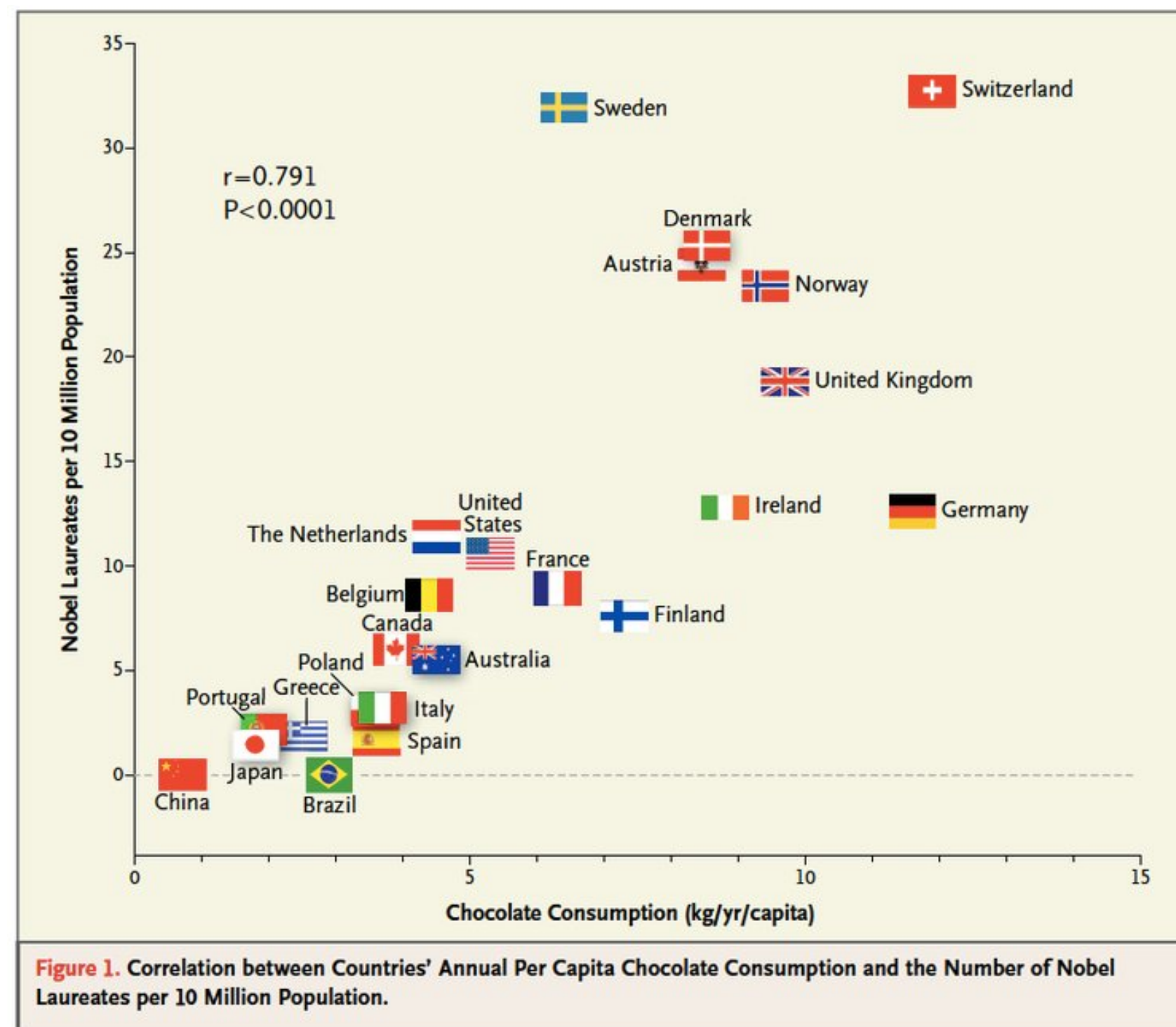


Reichenbach's principle of common cause:

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)

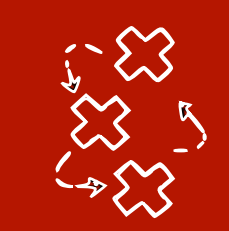


What if we don't know the graph? Causal discovery

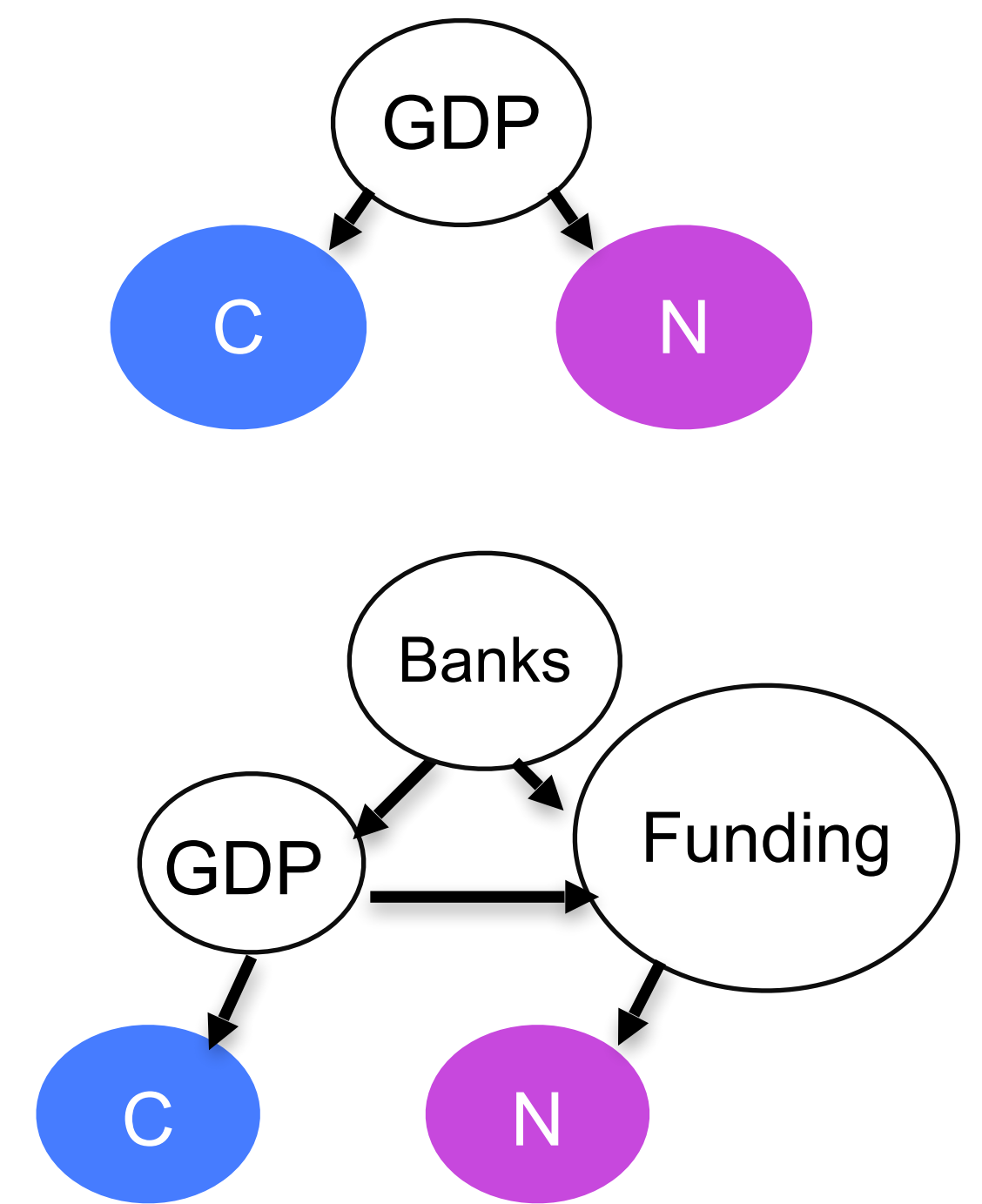
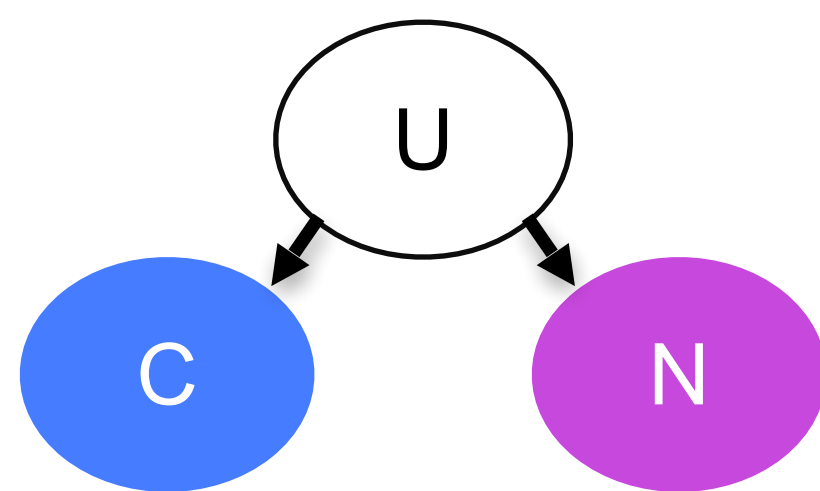
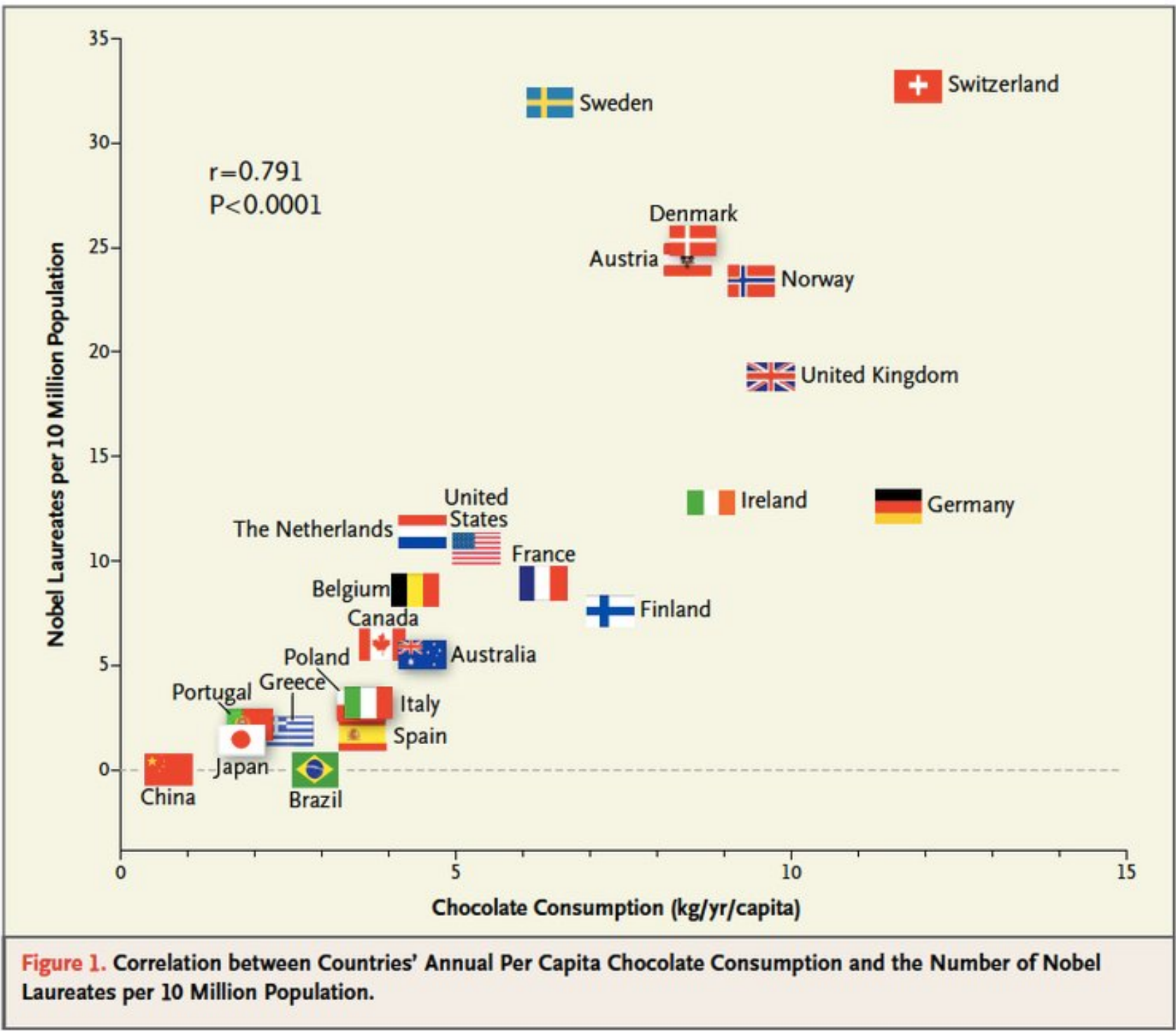


Reichenbach's principle of common cause:

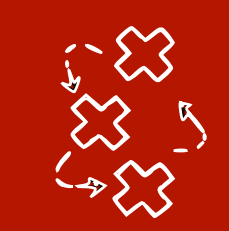
A correlation between X and Y implies that X causes Y, Y causes X, or there exists a common cause between X and Y (or any combination)



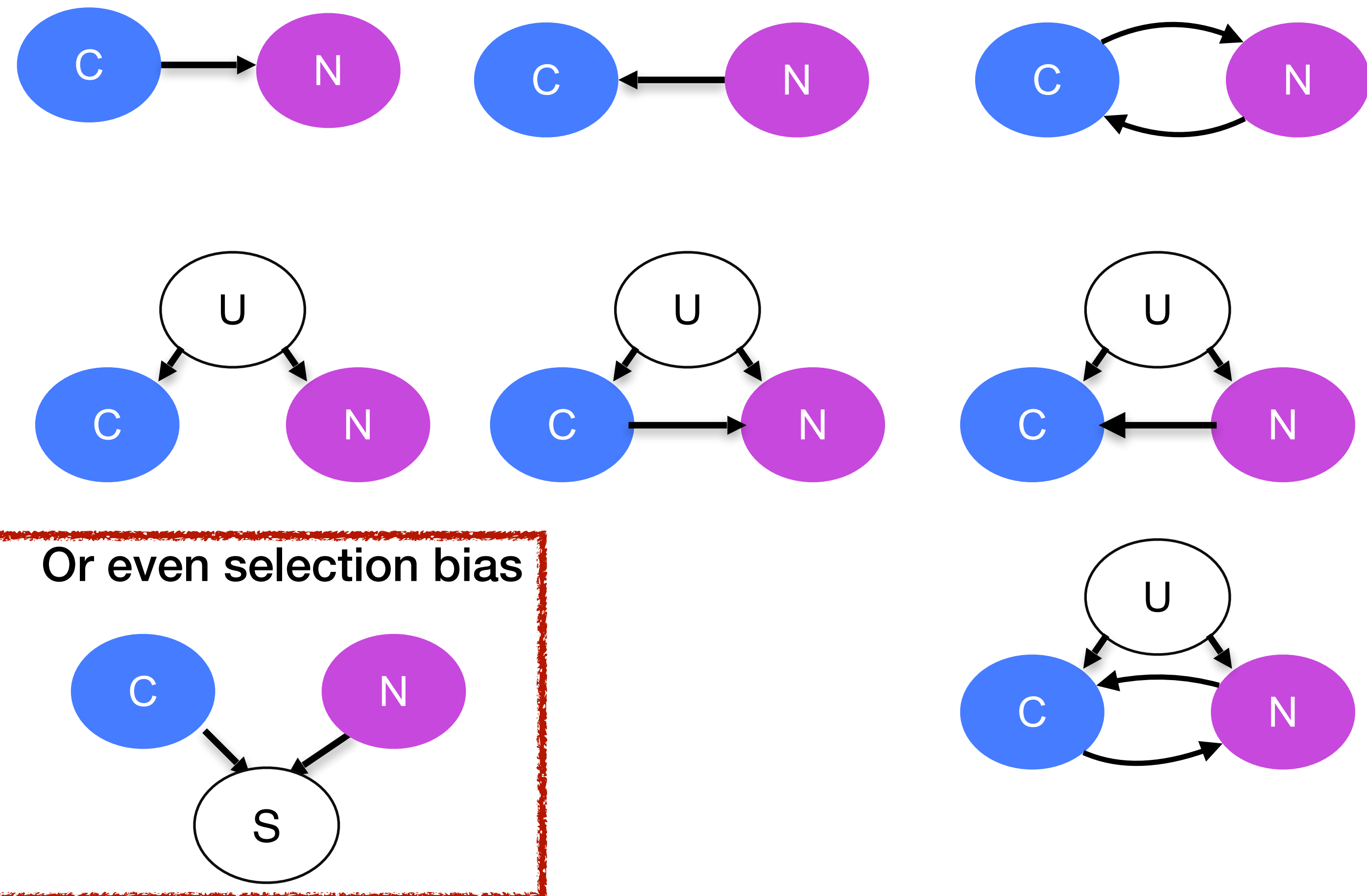
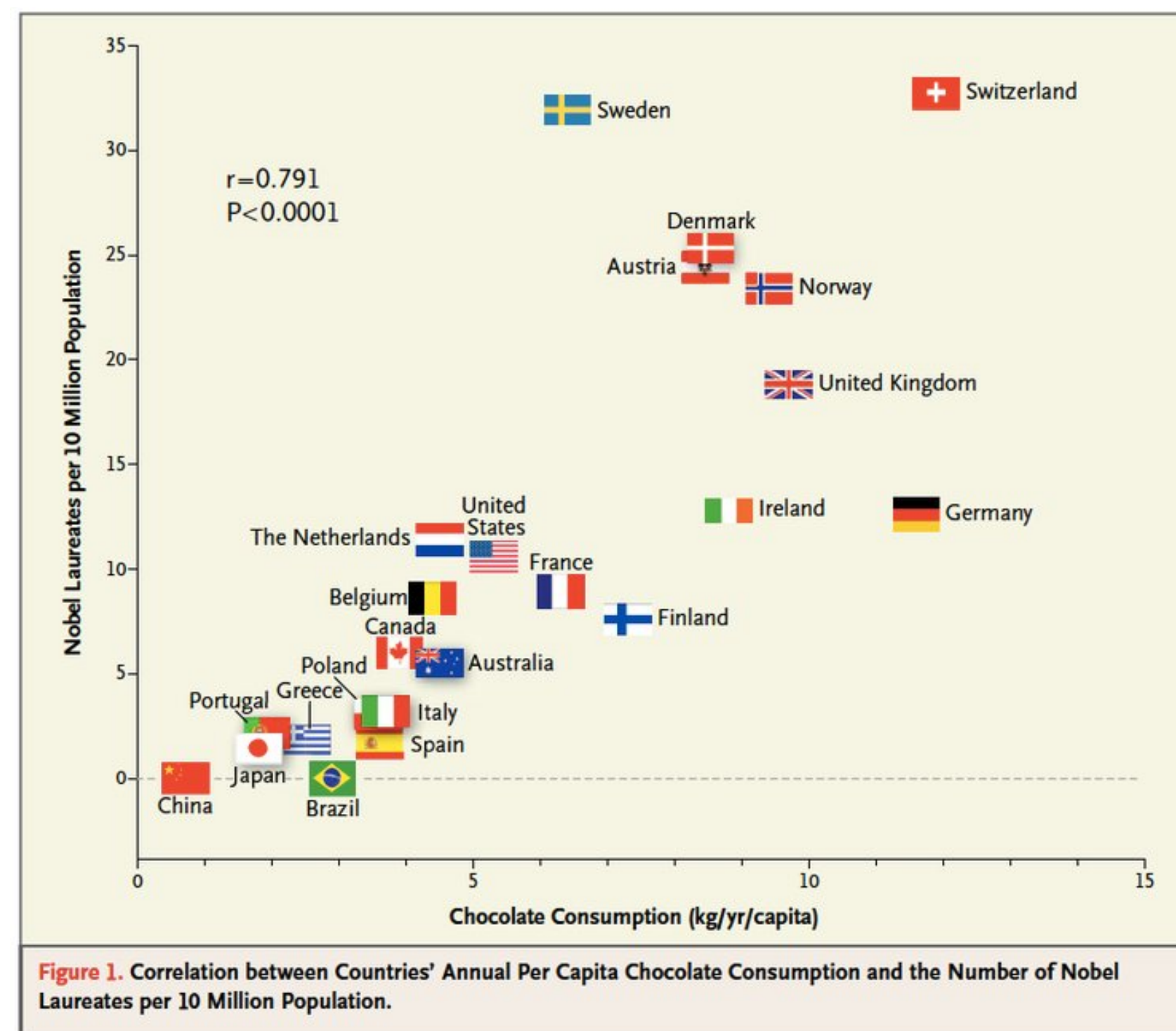
An example of confounding

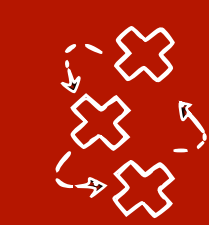


...

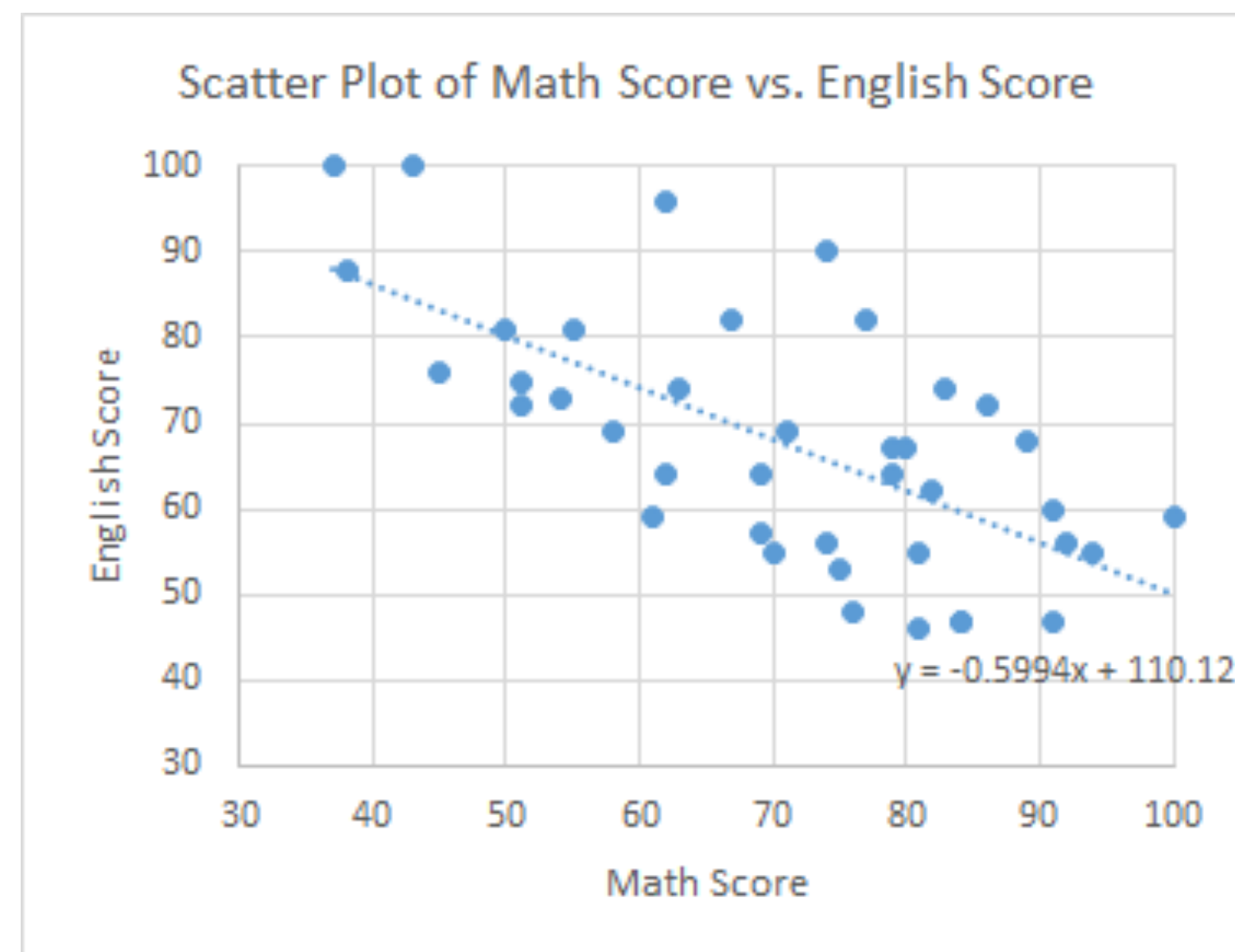


Statistically indistinguishable from only these data

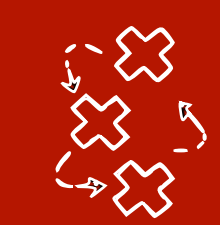




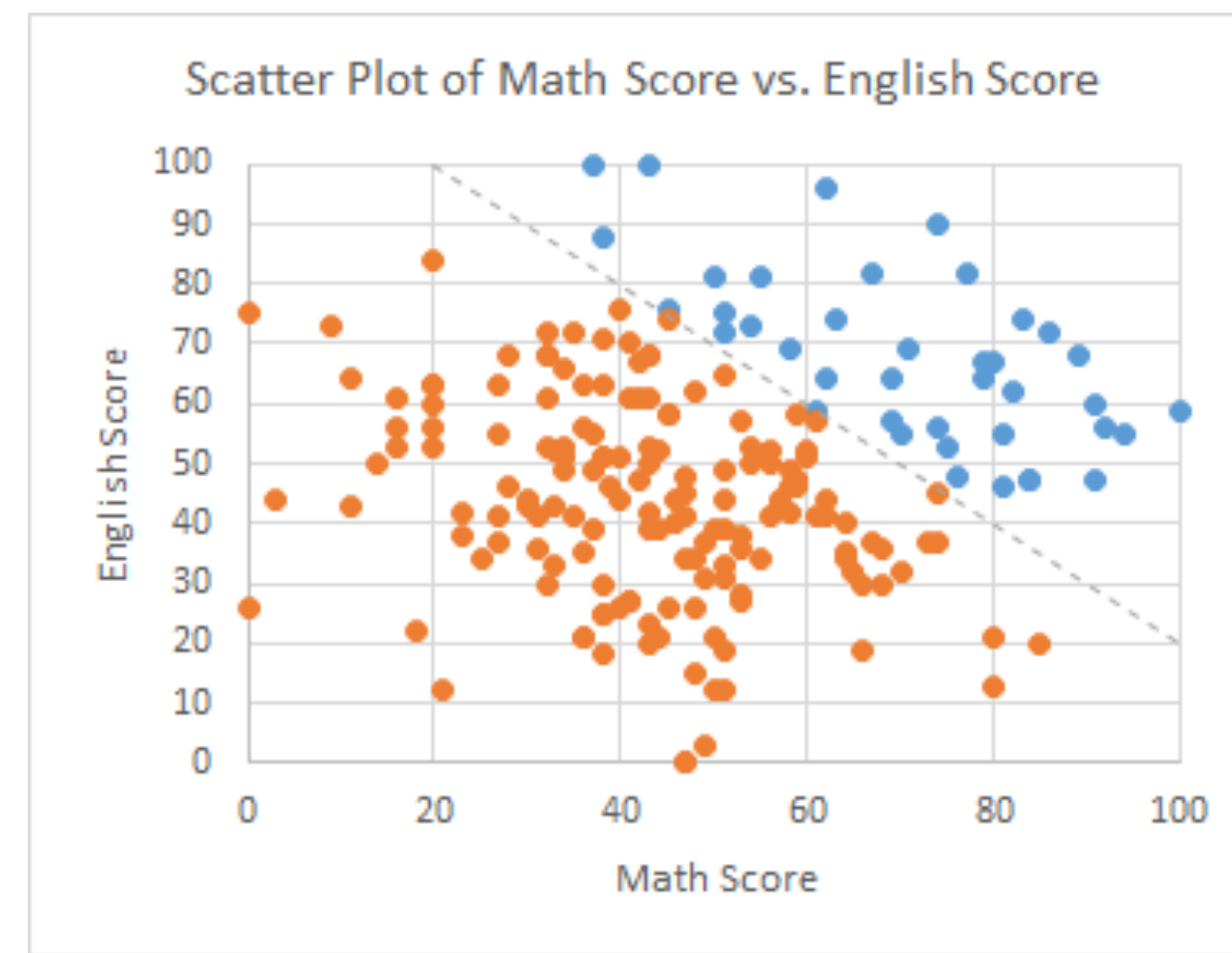
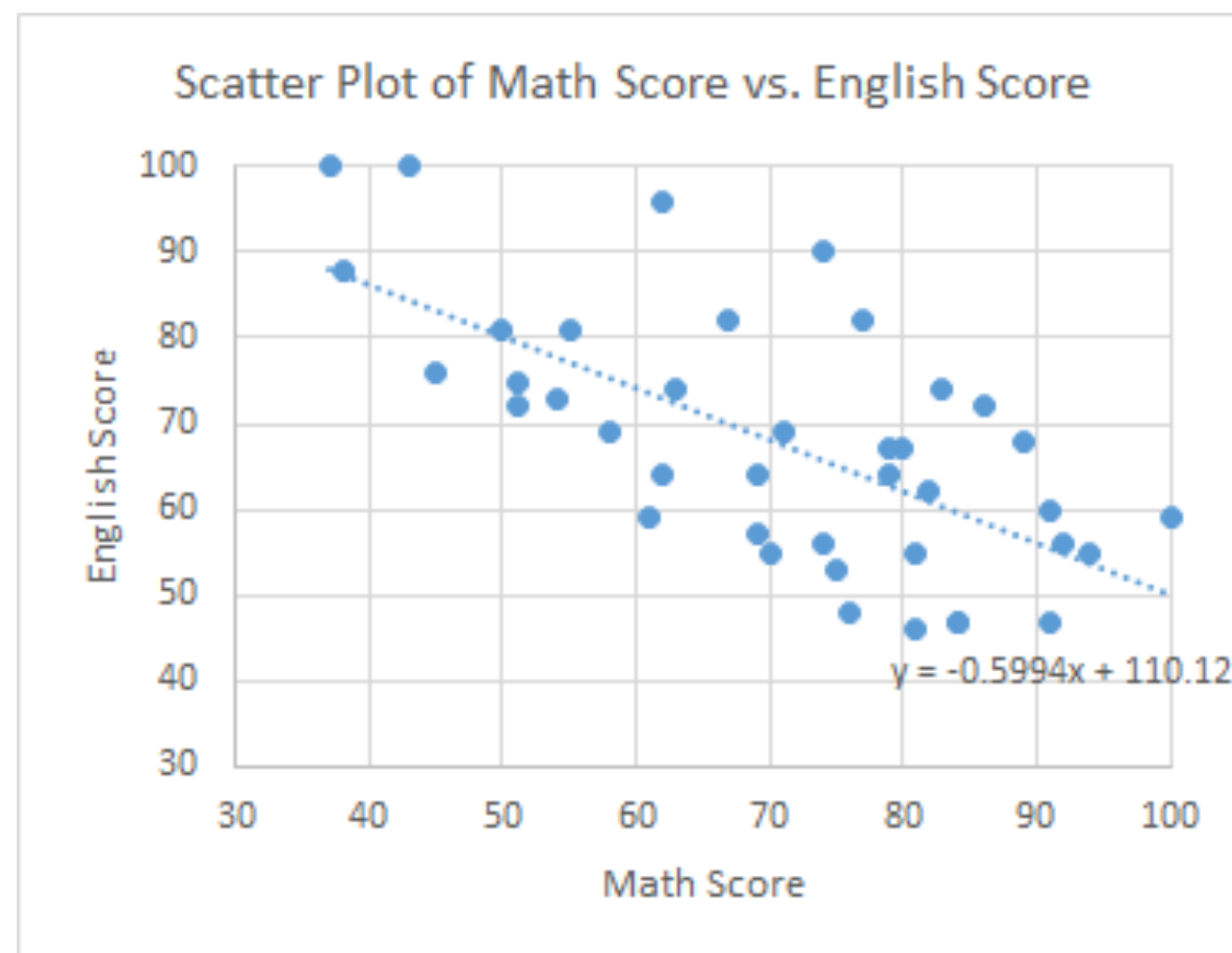
Another example of bias: Selection bias



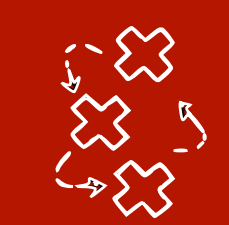
College admissions: English score is negatively correlated with Math Score



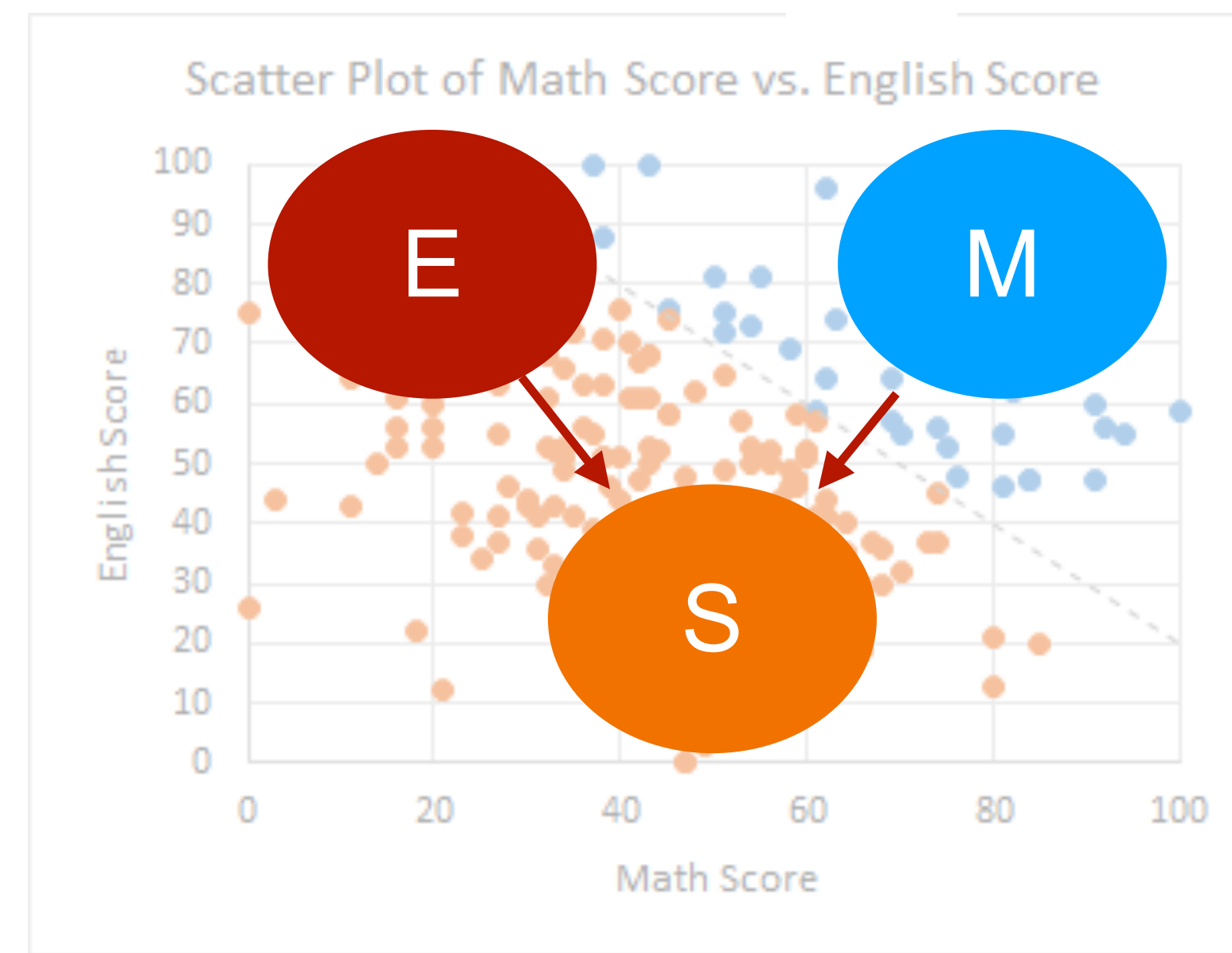
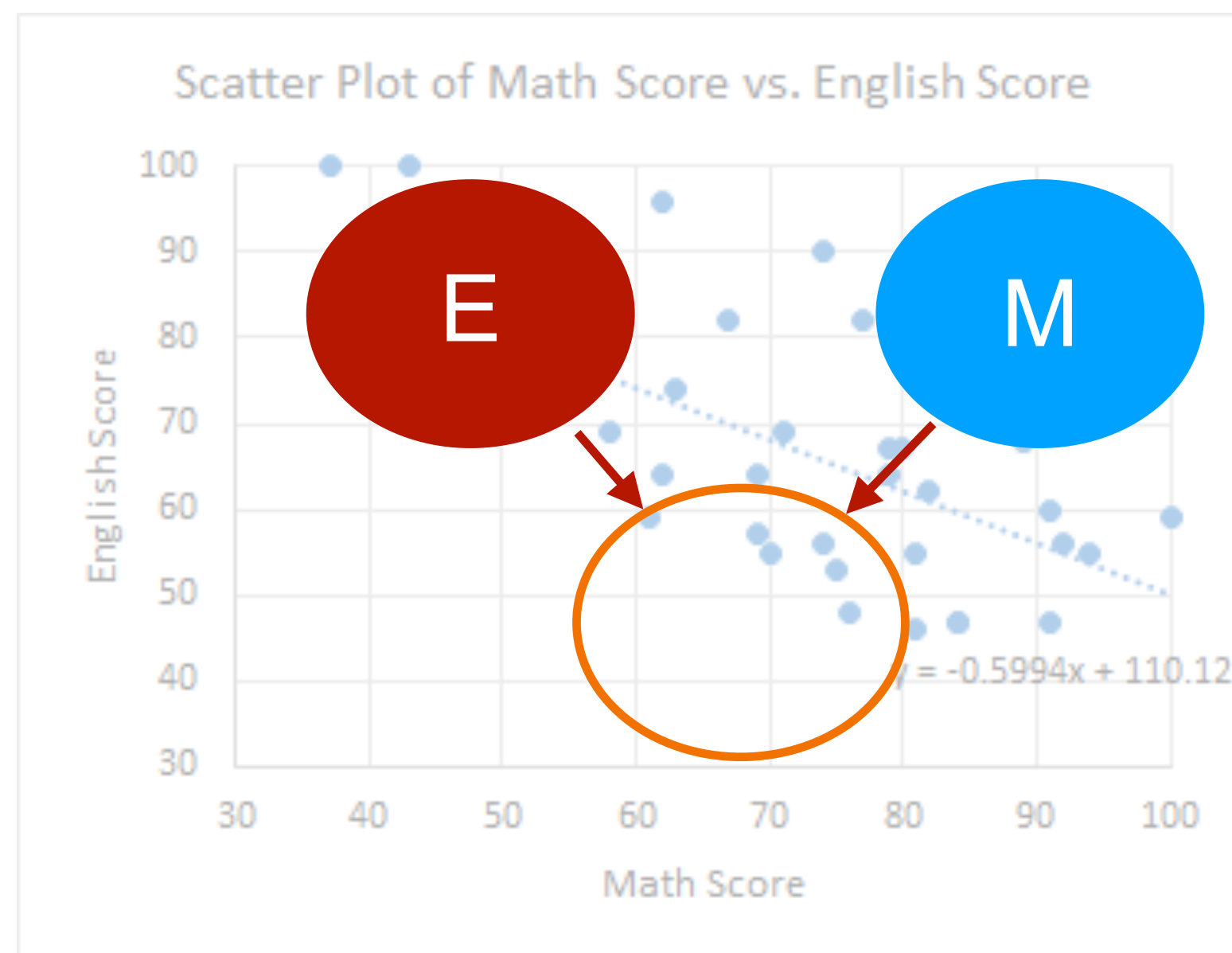
Another example of bias: Selection bias



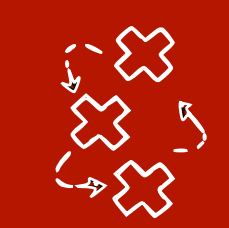
Selection rule: English score + Math Score > 120



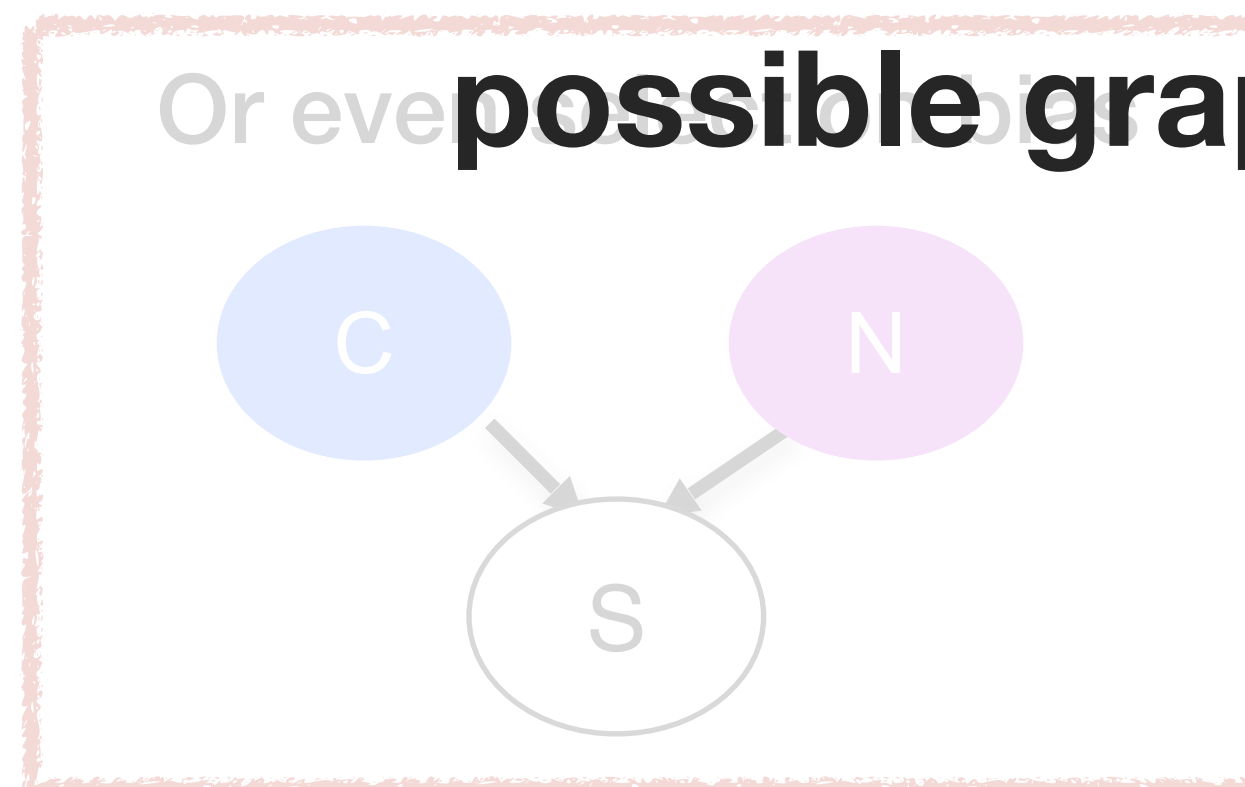
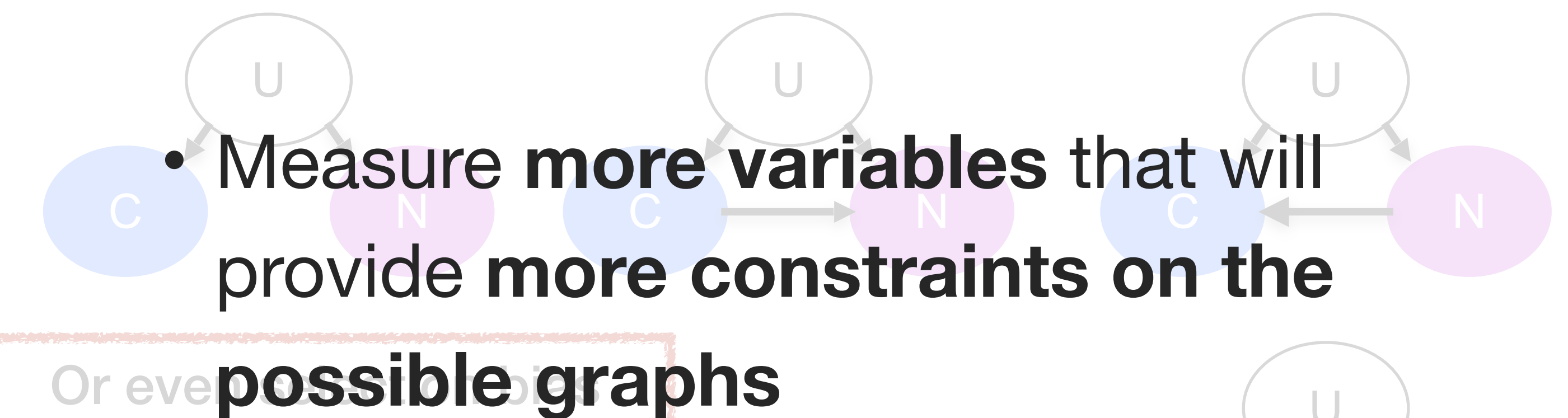
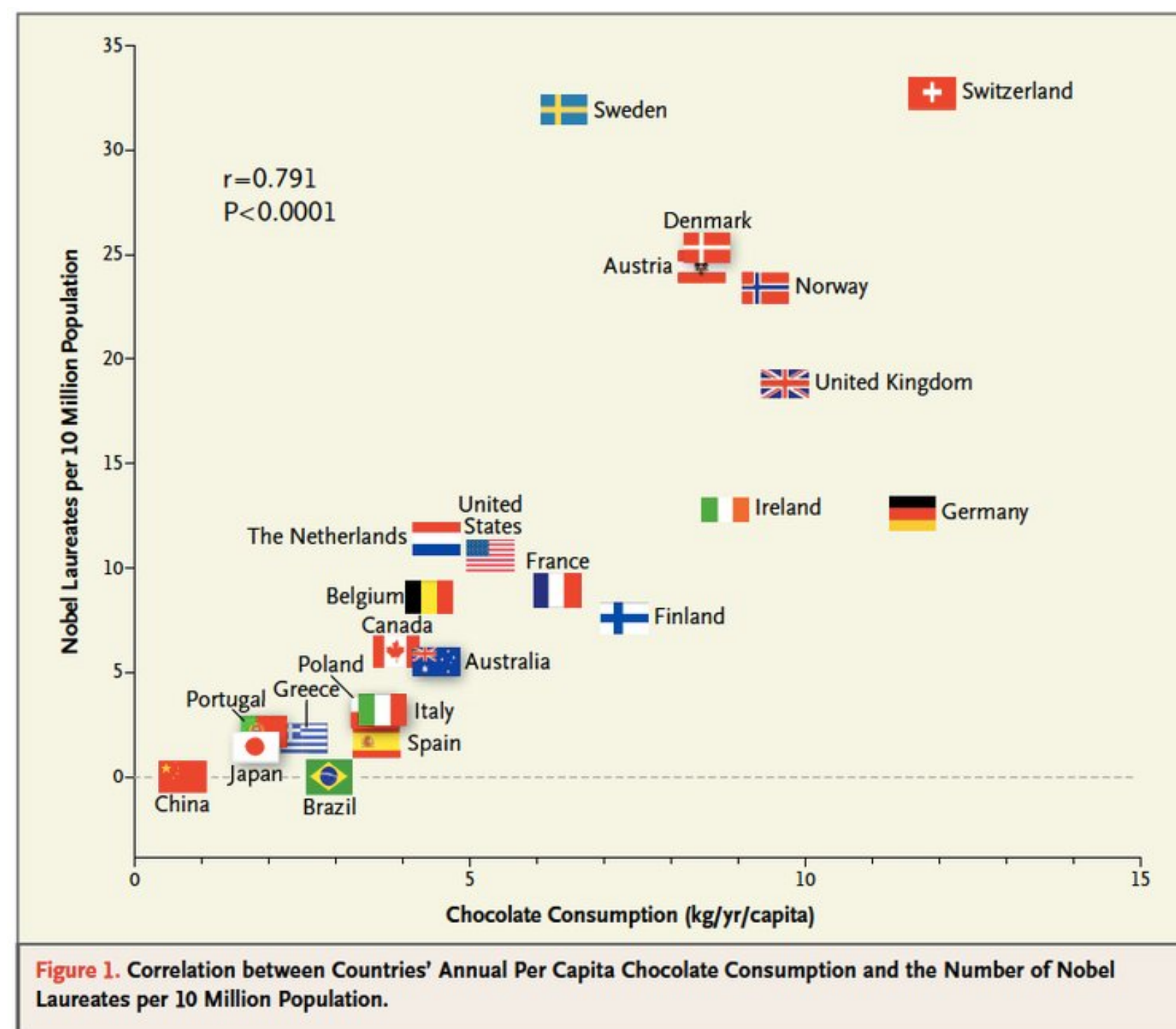
Another example of bias: Selection bias

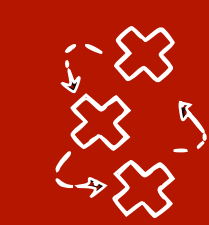


Selection rule: English score + Math Score $>$ 120



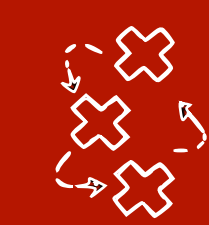
Statistically indistinguishable from only these data





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**

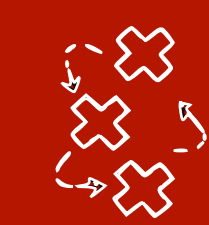


Recap: Independent random variables

- **Definition:** Two random variables X and Y are **independent** iff:

$$\forall x, y: P(X = x, Y = y) = P(X = x)P(Y = y)$$

- We then write $X \perp\!\!\!\perp Y$, otherwise we write $X \not\perp\!\!\!\perp Y$ (**dependent**)



Recap: Independent random variables

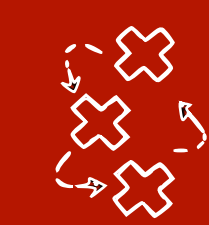
- **Definition:** Two random variables X and Y are **independent** iff:

$$\forall x, y: P(X = x, Y = y) = P(X = x)P(Y = y)$$

- We then write $X \perp\!\!\!\perp Y$, otherwise we write $X \not\perp\!\!\!\perp Y$ (**dependent**)
- Equivalently $\forall x, y$:

$$P(X = x \mid Y = y) = P(X = x) \text{ and } P(Y = y \mid X = x) = P(Y = y)$$

- Intuitively, knowing the value of Y **will not tell us anything** about the distribution of X (and viceversa, it's symmetric)



Recap: Independent random variables

- **Definition:** Two random variables X and Y are **independent** iff:

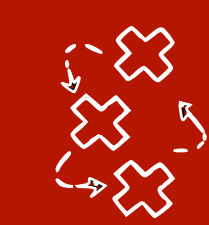
$$\forall x, y: P(X = x, Y = y) = P(X = x)P(Y = y)$$

Correlation is a special type of dependence that assumes a linear relation between X and Y

- We then write $X \perp Y$
- Equivalently $\forall x, y:$

$$P(X = x | Y = y) = P(X = x) \text{ and } P(Y = y | X = x) = P(Y = y)$$

- Intuitively, knowing the value of Y **will not tell us anything** about the distribution of X (and viceversa, it's symmetric)

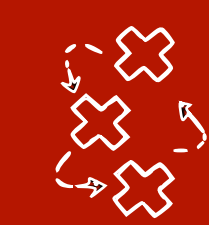


Recap: Independent random variables - example

- Consider throwing an **unbiased coin** two times $X_1, X_2 \in \{H, T\}$:



- Knowing that $X_1 = H$ will not change $P(X_2 = H | X_1 = H) = P(X_2 = H) = 0.5$,
so $X_1 \perp\!\!\!\perp X_2$

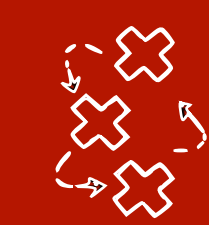


Recap: Independent random variables - example

- Consider throwing an **unbiased coin** two times $X_1, X_2 \in \{H, T\}$:



- Knowing that $X_1 = H$ will not change $P(X_2 = H | X_1 = H) = P(X_2 = H) = 0.5$
- On the other hand, if you have a smoke alarm $A \in \{0, 1\}$ that accurately detects smoke $S \in \{0, 1\}$, then $P(A = 1 | S = 1) \gg P(A = 1 | S = 0)$
- This means also that $P(A = 1 | S = 1) \neq P(A = 1) \implies A \not\perp S$

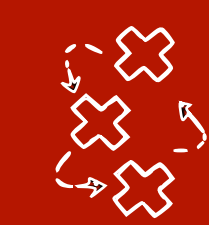


Recap: Conditional independence - definition

- X is independent of Y **conditioned/given** Z iff

$$\forall x, y, z: P(X = x | Y = y, Z = z) = P(X = x | Z = z) \quad (\text{for } P(Z = z) > 0)$$

- We then write $X \perp\!\!\!\perp Y | Z$, otherwise $X \not\perp\!\!\!\perp Y | Z$
- Intuitively this means that **Y does not add any information** to predict X that isn't already offered by Z
- Z can be a **set of variables**, e.g. $X \perp\!\!\!\perp Y | \{Z_1, Z_2\}$

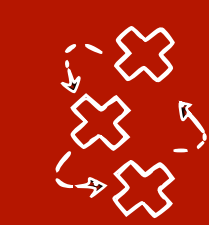


Recap: Conditional independence - example

- Alice and Bob throw each an **unbiased coin** $A, B \in \{H, T\}$:



- $P(B = H | A = H) = P(B = H) = 0.5$, so $A \perp\!\!\!\perp B$ and Alice cannot guess Bob's result better than random

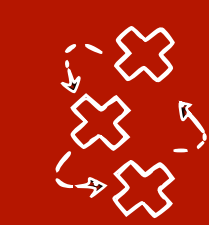


Recap: Conditional independence - example

- Alice and Bob throw each an **unbiased coin** $A, B \in \{H, T\}$:

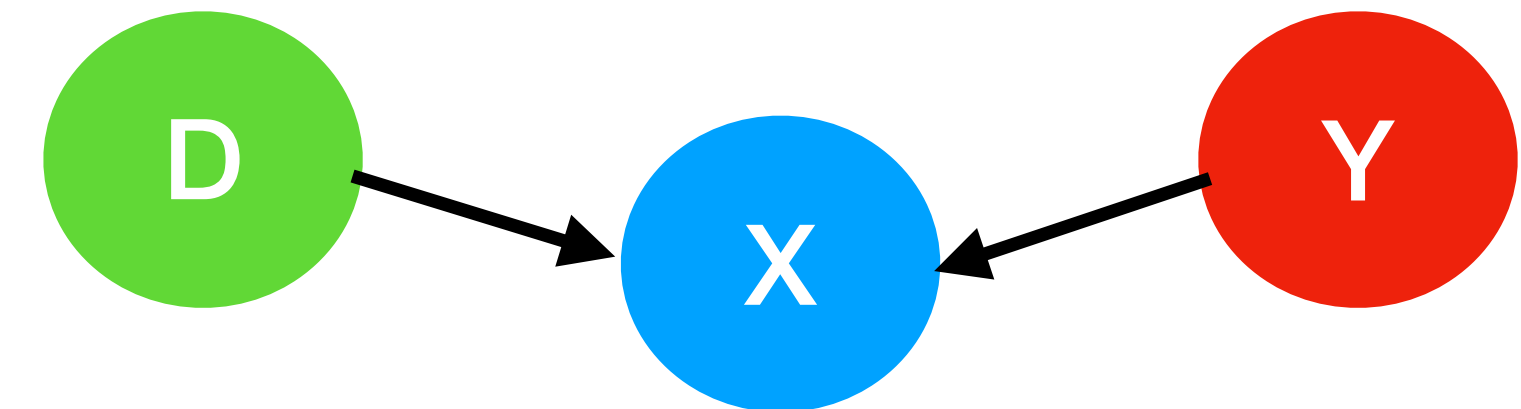
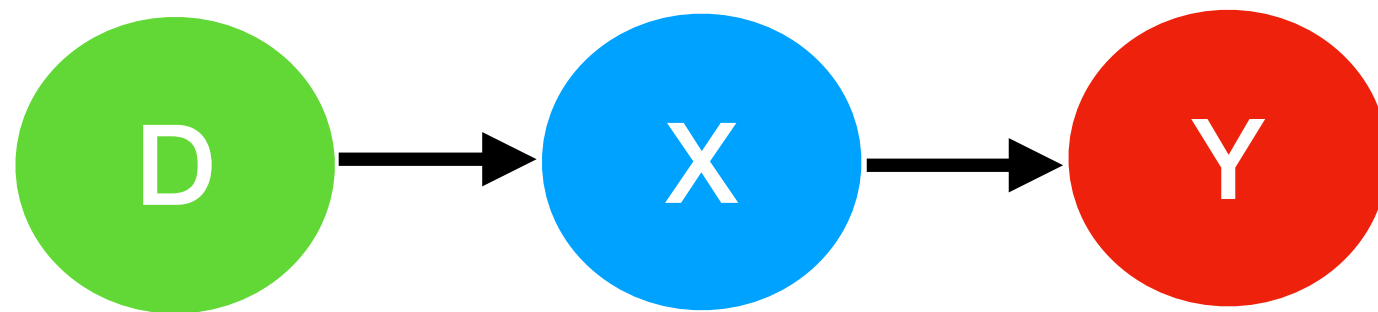


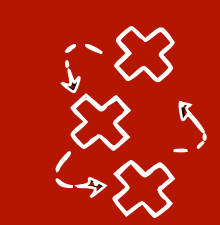
- $P(B = H | A = H) = P(B = H) = 0.5$, so $A \perp\!\!\!\perp B$ and Alice cannot guess Bob's result better than random
- Carlo observes both and tells them if they are different $C \in \{eq, diff\}$,
 $P(B = H | A = H, C = diff) = 0 \neq P(B = H | A = H) \implies A \not\perp\!\!\!\perp B | C$



A more general method for causal discovery

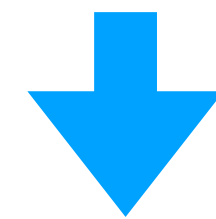
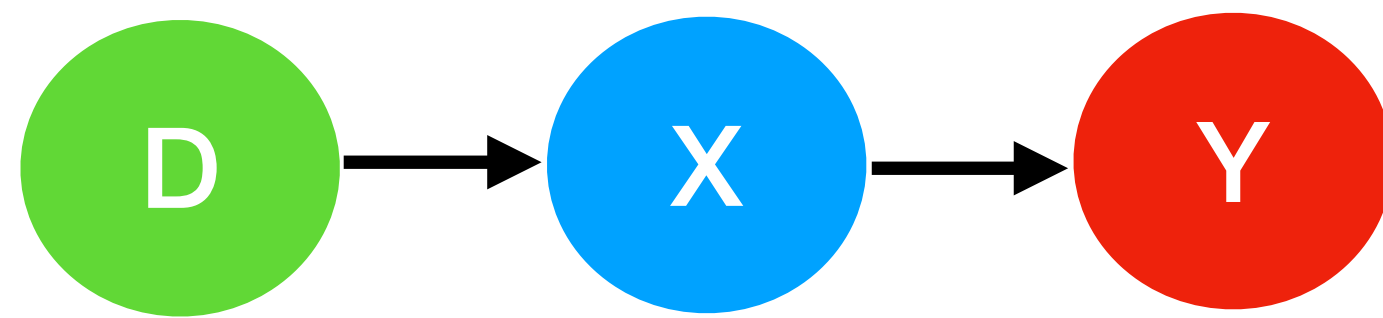
- Given a causal graph, **d-separation** is a graphical criterion that (under standard conditions) allows us to read **conditional independences**





A more general method for causal discovery

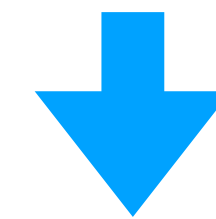
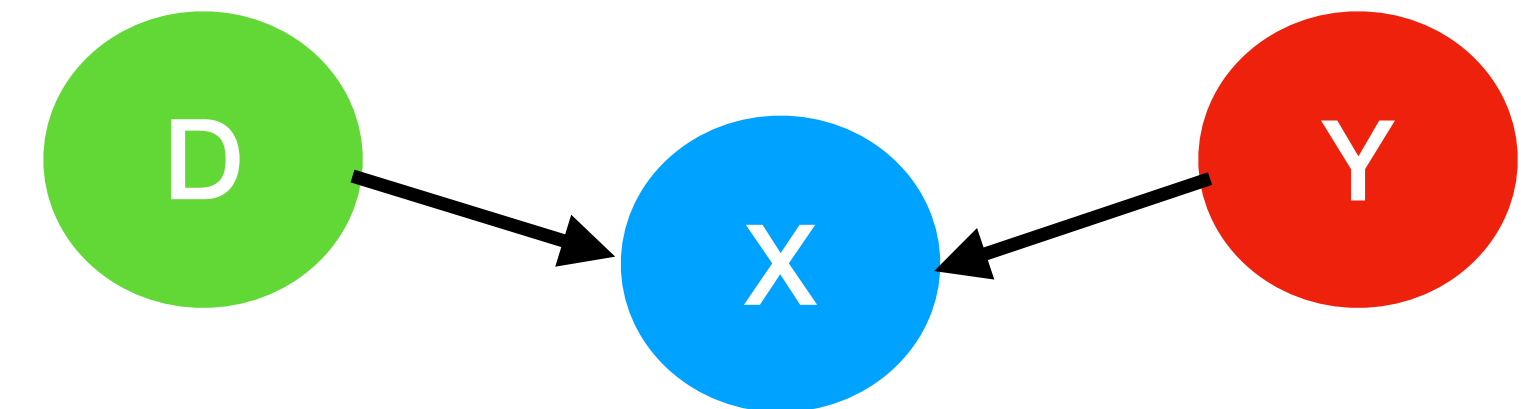
- Given a causal graph, **d-separation** is a graphical criterion that (under standard conditions) allows us to read **conditional independences**



Graph:

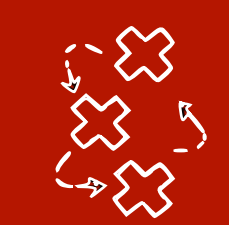
$Y \not\perp_d D$

$Y \perp_d D | X$



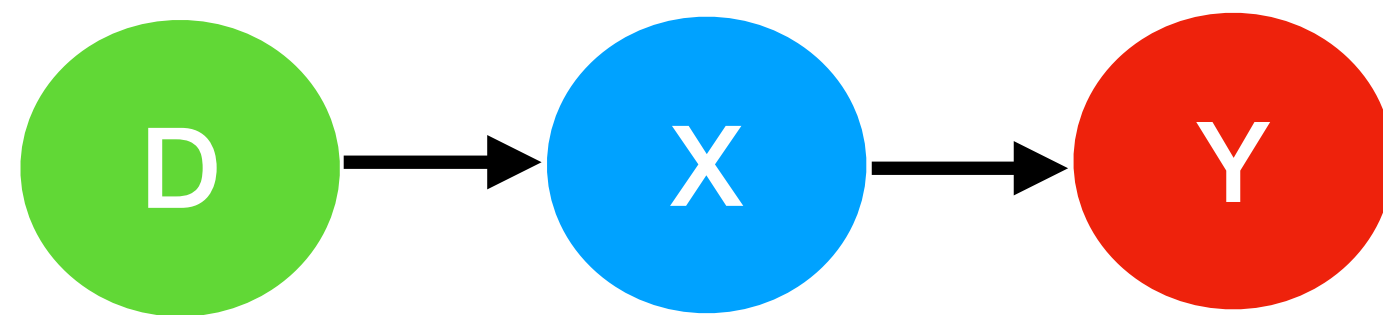
$Y \perp_d D$

$Y \not\perp_d D | X$



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:

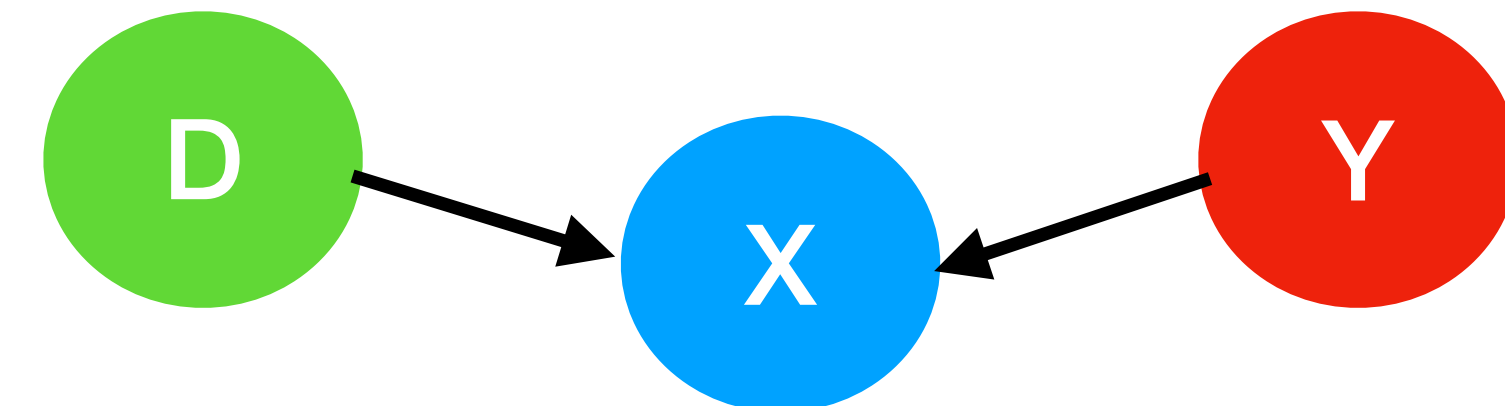
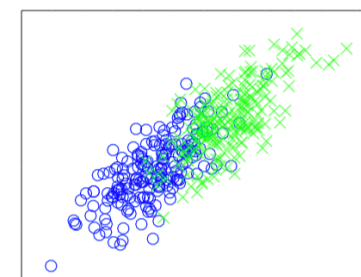
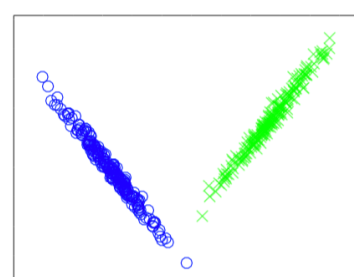
$$Y \not\perp_d D$$

$$Y \perp_d D | X$$

Data:

$$Y \not\perp D$$

$$Y \perp D | X$$

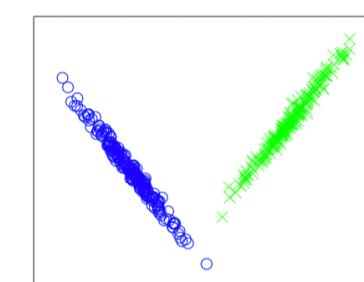
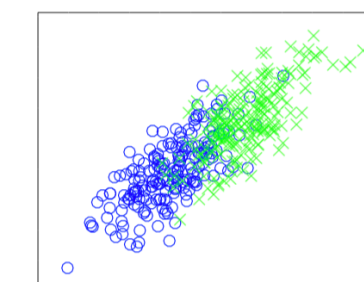


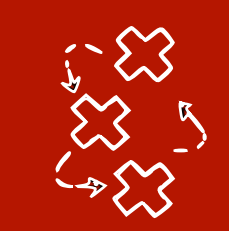
$$Y \perp_d D$$

$$Y \not\perp_d D | X$$

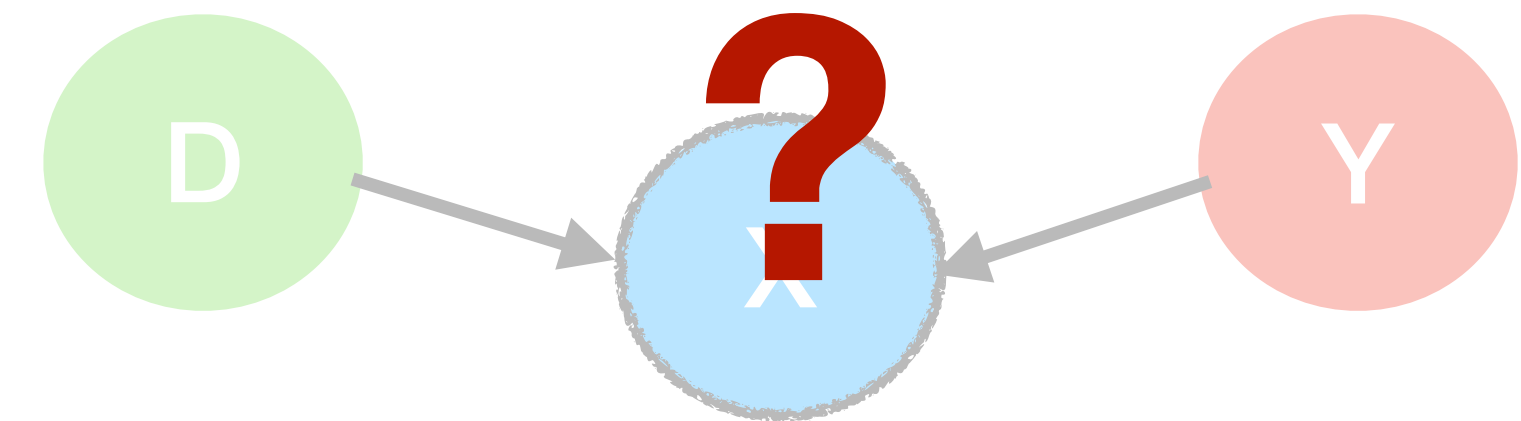
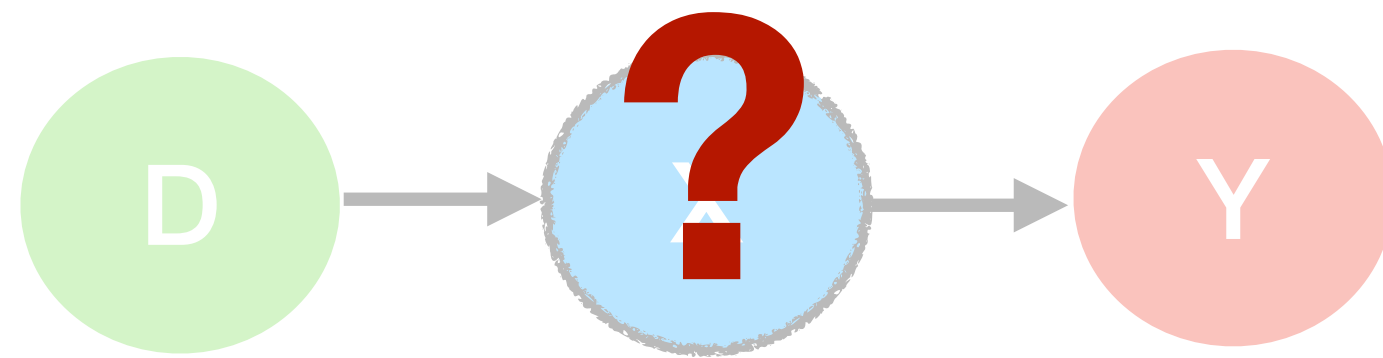
$$Y \perp D$$

$$Y \not\perp D | X$$



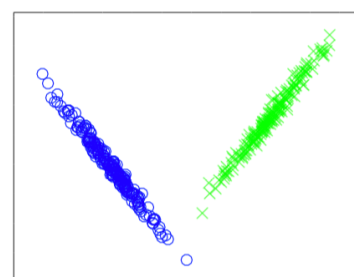


A more general method for causal discovery

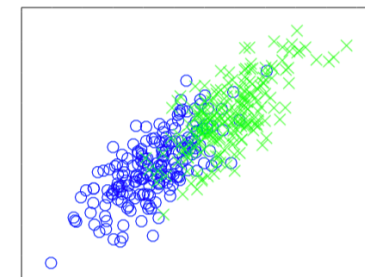


Data:

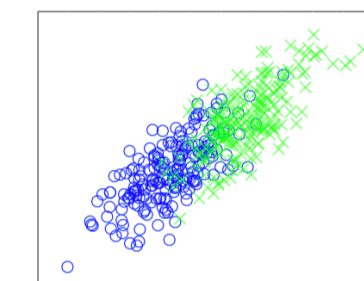
$Y \not\perp\!\!\!\perp D$



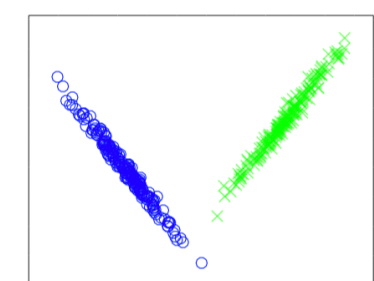
$Y \perp\!\!\!\perp D | X$

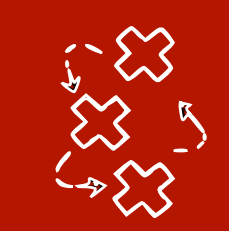


$Y \perp\!\!\!\perp D$

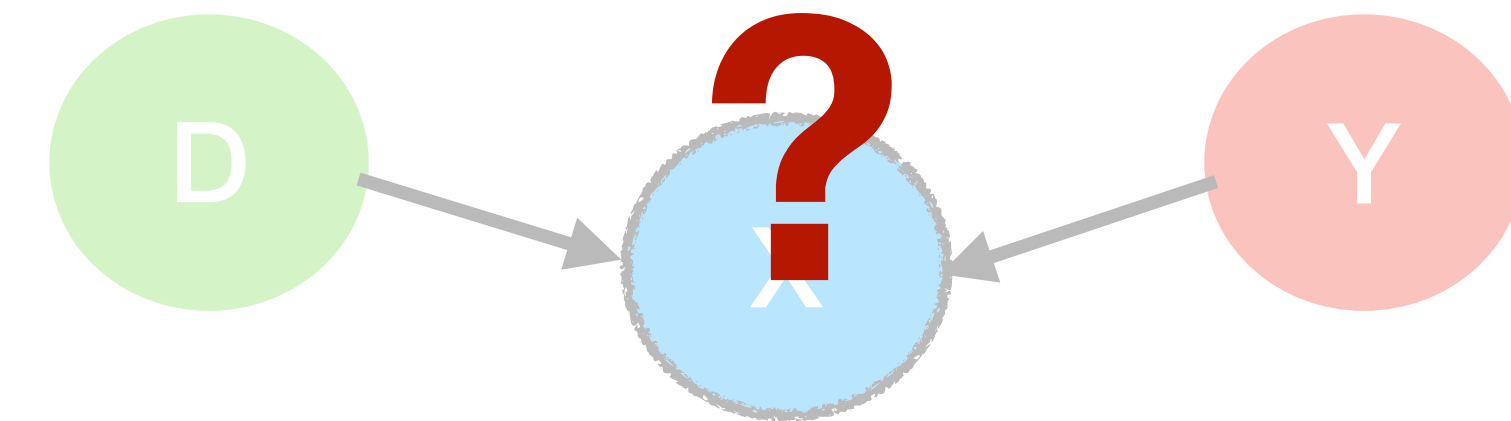
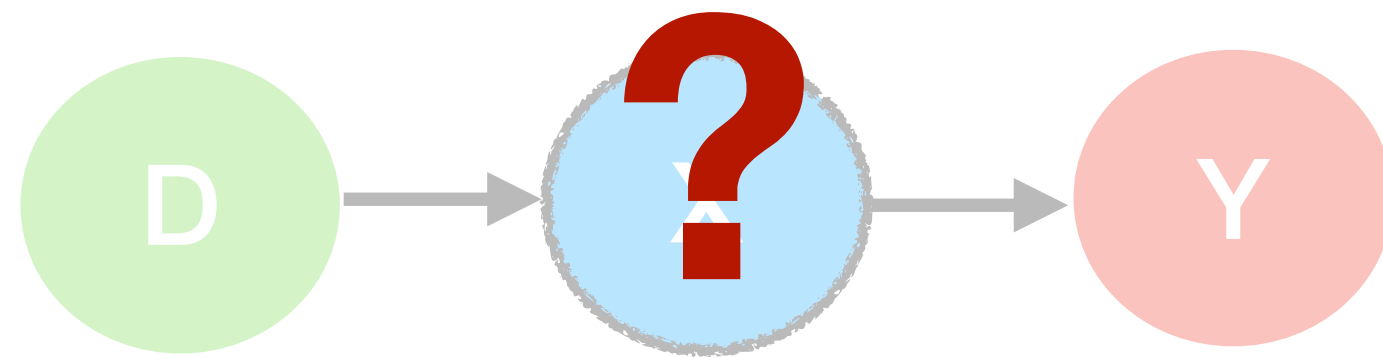


$Y \not\perp\!\!\!\perp D | X$





A more general method for causal discovery



Graph:

$$Y \not\perp_d D$$

$$Y \perp_d D | X$$

$$Y \perp_d D$$

$$Y \not\perp_d D | X$$

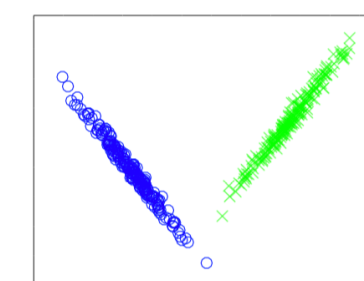
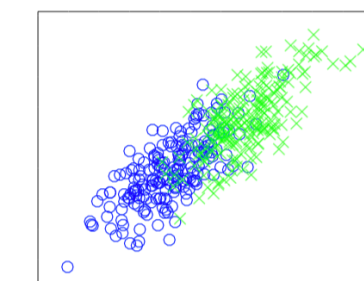
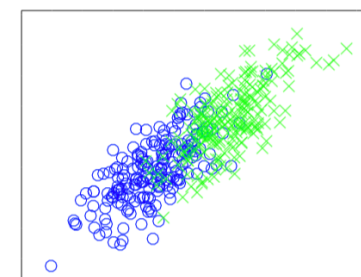
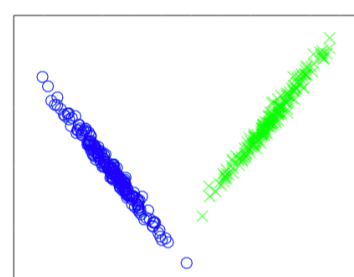
Data:

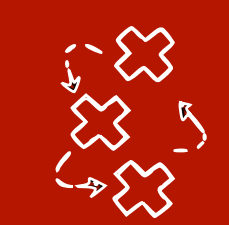
$$Y \not\perp D$$

$$Y \perp D | X$$

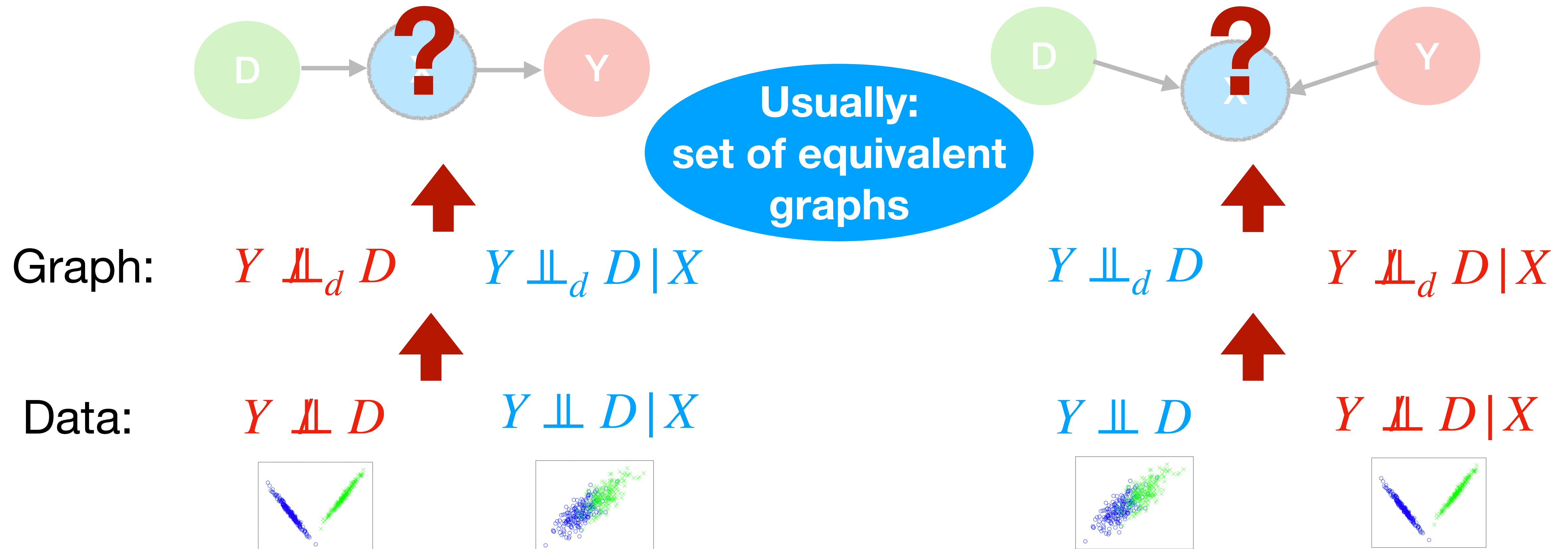
$$Y \perp D$$

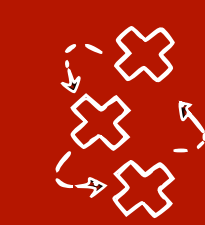
$$Y \not\perp D | X$$



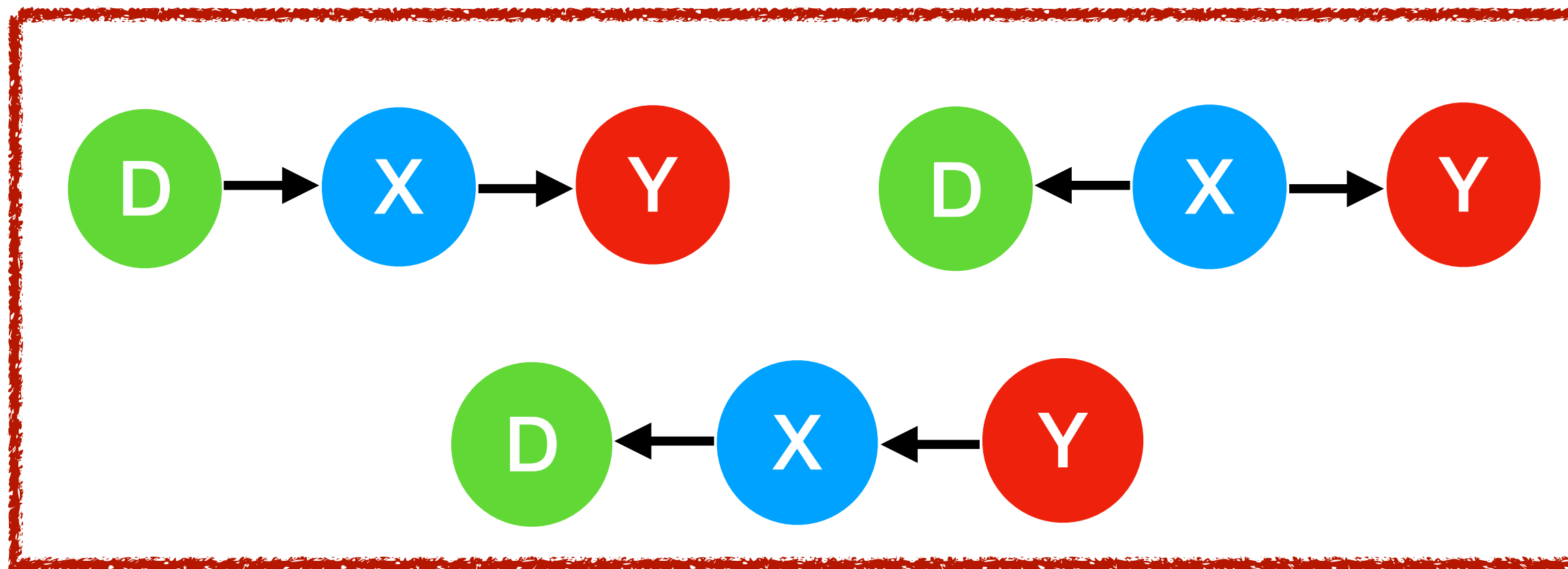


A more general method for causal discovery





Example: assuming no selection bias or latent confounding



Graph:

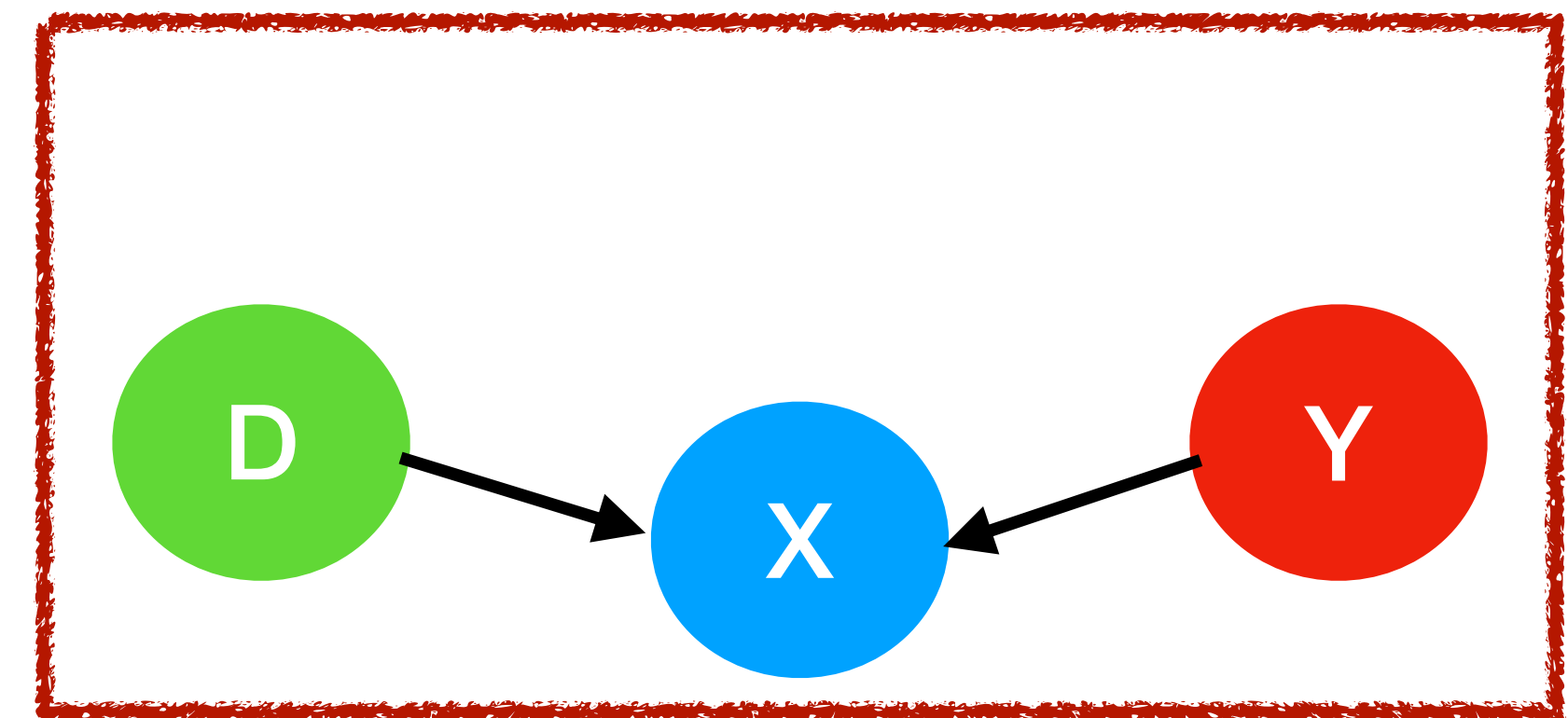
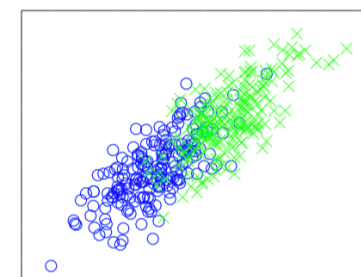
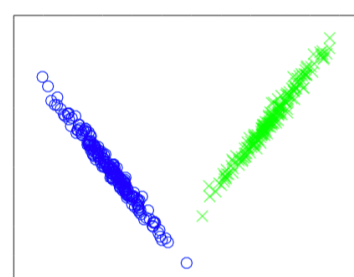
$$Y \perp\!\!\!\perp_d D$$

$$Y \perp\!\!\!\perp_d D | X$$

Data:

$$Y \perp\!\!\!\perp D$$

$$Y \perp\!\!\!\perp D | X$$

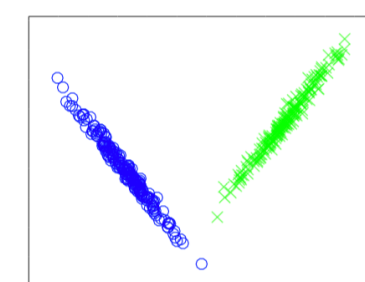
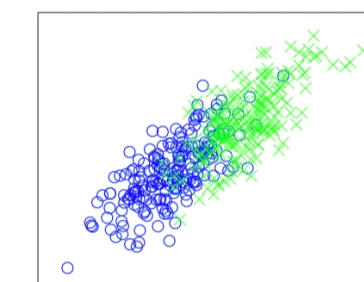


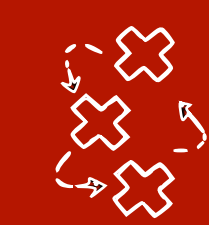
$$Y \perp\!\!\!\perp_d D$$

$$Y \perp\!\!\!\perp_d D | X$$

$$Y \perp\!\!\!\perp D$$

$$Y \perp\!\!\!\perp D | X$$

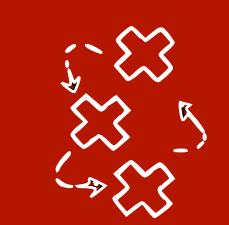




Cheatsheet: d-separation - Algorithm

- A **path** π between **node i and node j** is a sequence of **distinct nodes** $(i, \dots, j)$ such that each two **consecutive nodes** are **adjacent (regardless of orientation)**
- A **directed path** π between **node i and node j** is a path where **all edges point towards j**, i.e. $i \rightarrow \dots \rightarrow j$





Cheatsheet: d-separation - Algorithm

- A **path** π between **node i and node j** is a sequence of **distinct nodes** $(i, \dots, j)$ such that each two **consecutive nodes** are **adjacent (regardless of orientation)**
- A **directed path** π between **node i and node j** is a path where **all edges point towards j**, i.e. $i \rightarrow \dots \rightarrow j$



Algorithm 2 Compute if nodes i and j are d-separated by conditioning set A

Require: Nodes $i \neq j \in V$, conditioning set $A \subseteq V \setminus \{i, j\}$

function ISDSEPARATED(i, j, A)

for all paths π between i and j **do**

$\triangleright$ Check all paths between i and j . If i and j are adjacent, this includes the path (i, j) .

if ISPATHBLOCKED(i, j, π, A) == False **then**

return False

$\triangleright$ There is at least a path that is not blocked by A , so they are not d-separated.

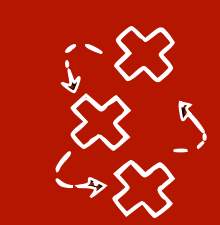
end if

end for

return True

$\triangleright$ All paths between i and j are blocked by A , so they are d-separated.

end function



Cheatsheet: Blocked Path - Algorithm

- A **collider** k on a **path** $\pi = (i, \dots, j)$ is a non-endpoint node ($k \neq i, j$) s.t. the path π contains $\dots \rightarrow k \leftarrow \dots$. Other nodes on π without this pattern are **non-colliders**

Algorithm 1 Compute if path π between nodes i and j is blocked by conditioning set A

Require: Nodes $i \neq j \in V$, path $\pi = (i, \dots, j)$, conditioning set $A \subseteq V \setminus \{i, j\}$

function ISPATHBLOCKED(i, j, π, A)

for all nodes k on π , such that $k \neq i, j$ **do**

if k is a non-collider on π **then**

if $k \in A$ **then**

return True

end if

else

if $\text{Desc}(k) \cap A = \emptyset$ **then**

return True

end if

end if

end for

return False

end function

‣ A non-collider on π is a node on the path without two incoming edges.

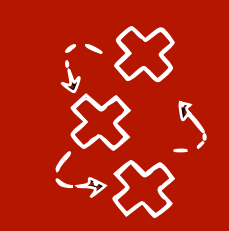
‣ If a non-collider on π is in A , the path is blocked, else continue searching.

‣ Else k is a collider on π , so a node with two incoming edges on π .

‣ Note: the descendants of a node include the node itself.

‣ If a collider on π or its descendants are not in A , the path is blocked, else continue.

‣ We did not find any node that could block this path. This includes the case in which i is adjacent to j on π .

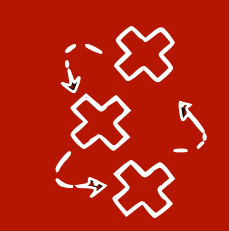


Causal Hierarchy [Pearl 2009, 2018]

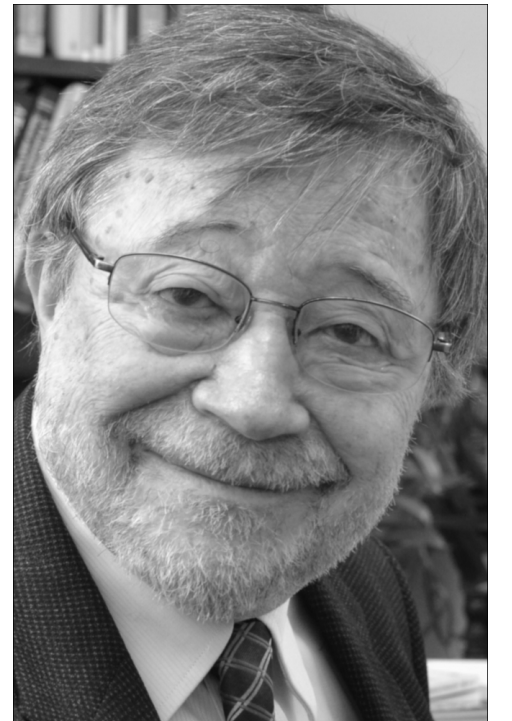


Most ML

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?

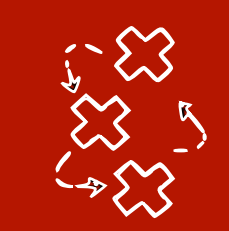


Causal Hierarchy [Pearl 2009, 2018]

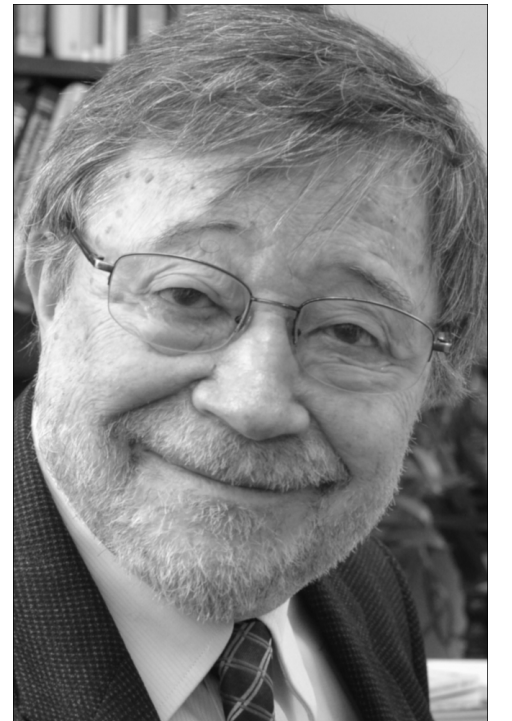


Most ML

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?

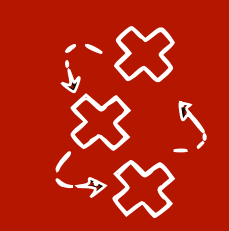


Causal Hierarchy [Pearl 2009, 2018]

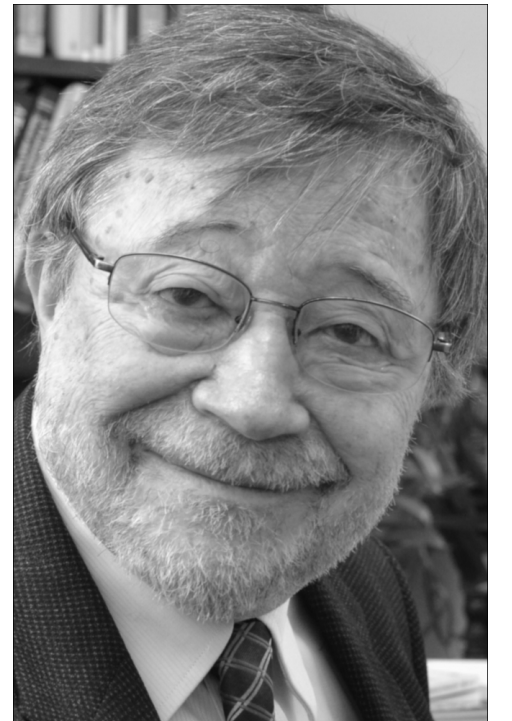


Most ML

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 smok- ing the past 2 years?



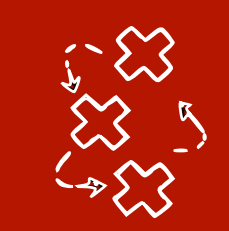
Causal Hierarchy [Pearl 2009, 2018]



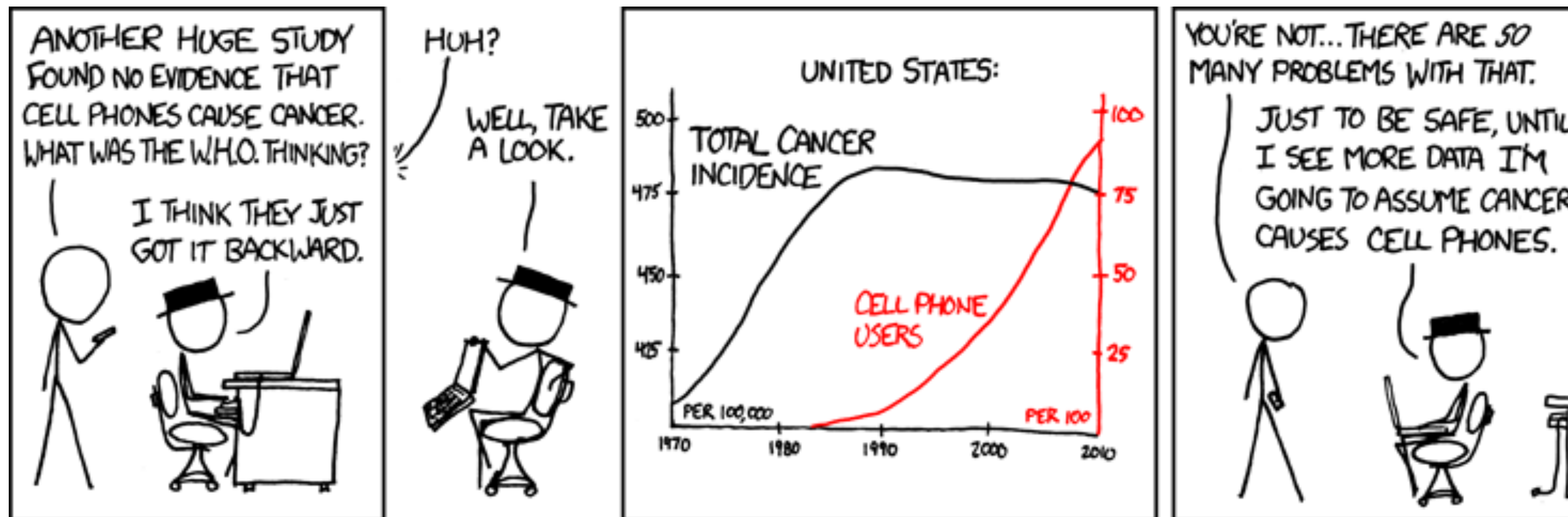
Most ML

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 smok- ing the past 2 years?



Questions?



<https://xkcd.com/925/>