

# Geographic Data Science - Lecture X

## Causal Inference

Dani Arribas-Bel

# Today

- Correlation Vs Causation
- Causal inference
- Why/when causality matters
- Hurdles to causal inference & strategies to overcome them

# Correlation Vs Causation

# Correlation Vs Causation

Two fundamental ways to look at the relationship between two (or more) variables:

## Correlation

Two variables have **co-movement**. If we know the value of one, we know something about the value of the other one.

## Causation

There is a “**cause-effect**” link between the two and, as a result, they display **co-movement**.

# Correlation Vs Causation

- Both are useful, but for different purposes
- Causation *implies* correlation but not the other way around
- It is vital to keep this distinction in mind for meaningful and credible analysis

# Examples

*Sign correlation? Causal link?*

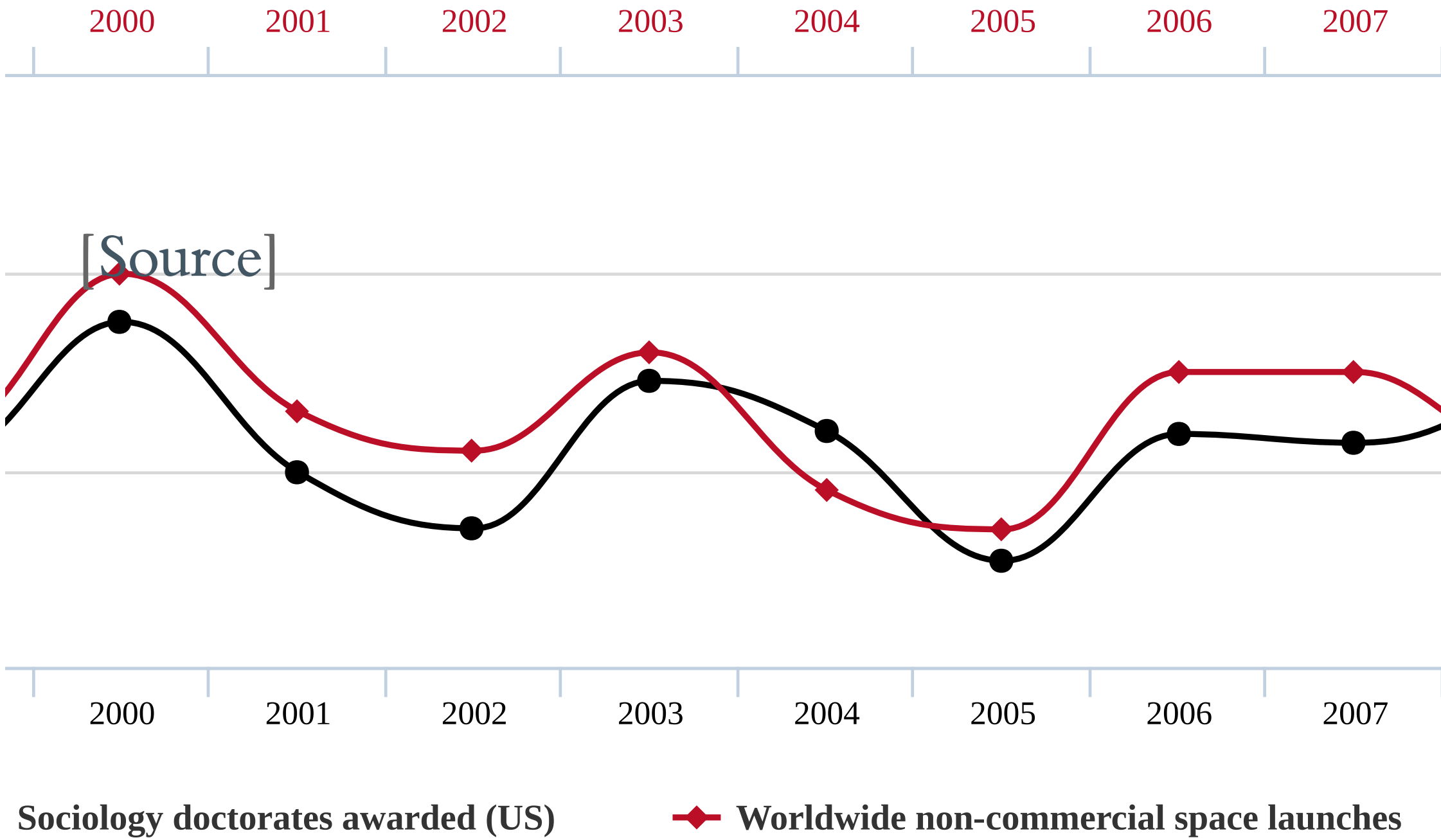
Take a guess (2mins)...

- Temperature and ice-cream consumption →  
**Positive. Positive.**
- Non-commercial space launches & Sociology  
PhDs awarded
- Crime & policing
- IMD in an area Vs its neighbors (Liverpool)

# Worldwide non-commercial space launches

correlates with

# Sociology doctorates awarded (US)



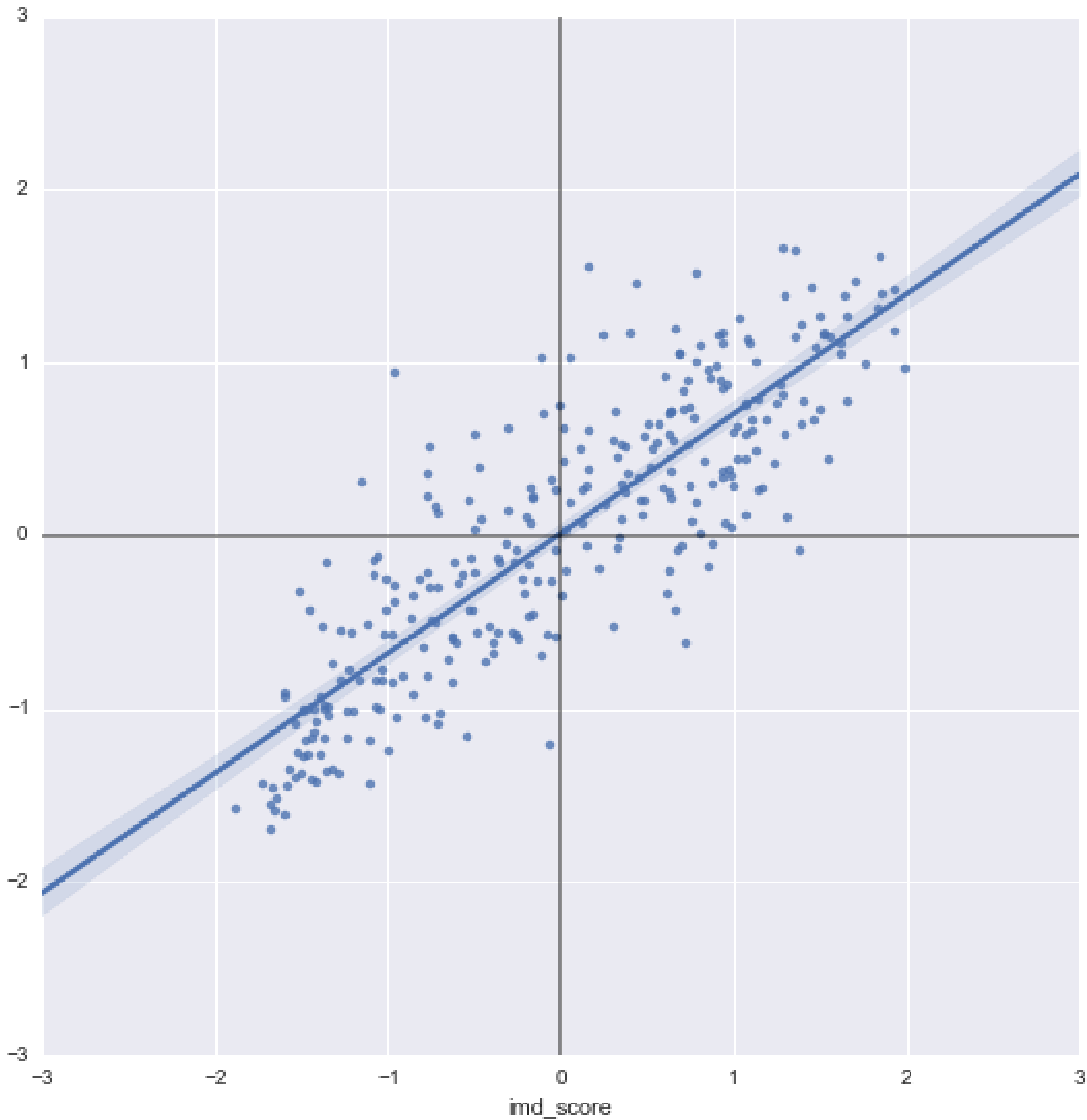
# Examples

*Positive or negative correlation? Causal link?*

Take a guess (2mins)...

- Temperature and ice-cream consumption → **Positive. Positive.**
- Non-commercial space launches & Sociology PhDs awarded → **Positive. None.**
- Crime & policing → **Positive. Negative.**
- IMD in an area Vs its neighbors (Liverpool)





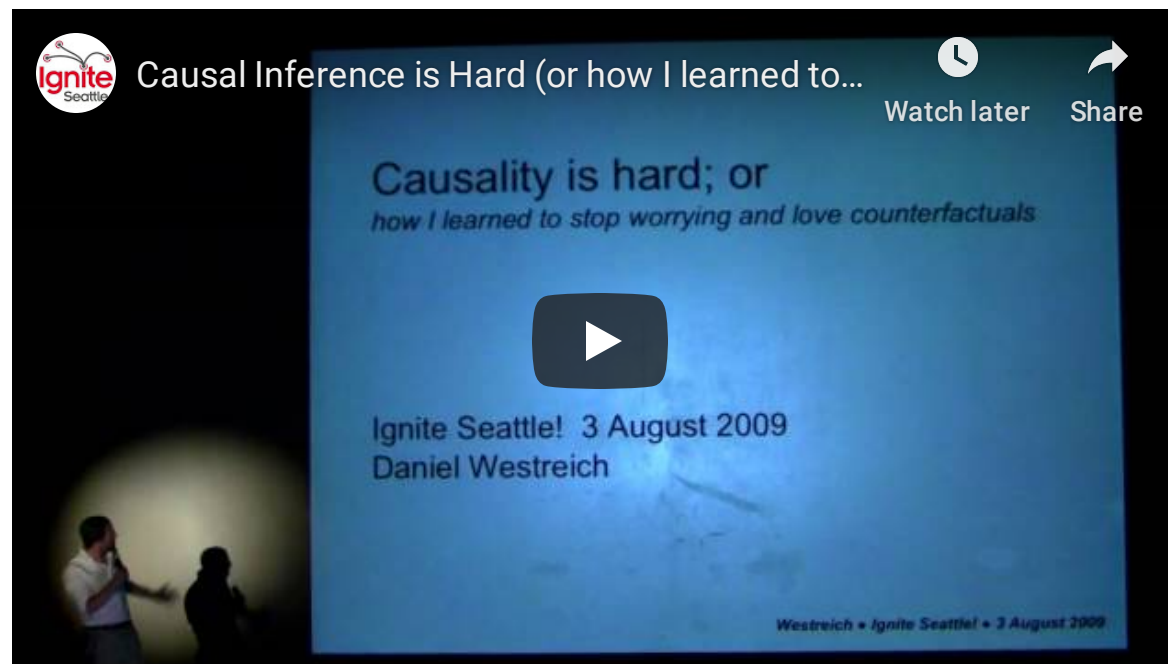
# Examples

*Positive or negative correlation? Causal link?*

Take a guess (2mins)...

- Temperature and ice-cream consumption → **Positive. Positive.**
- Non-commercial space launches & Sociology PhDs awarded → **Positive. None.**
- Crime & policing → **Positive. Negative.**
- IMD in an area Vs its neighbors (Liverpool) → **Positive. ?**

# Causal Inference



[Source]

*Why/When to get Causal?*

# Why

- Most often, we are interested in understanding the **processes** that *generate* the world, not only in observing its outcomes
- Many of these processes are only **indirectly observable** through outcomes
- The only way to link both is through **causal channels**

# When

Essentially when the core interest is to find out if something *causes* something else

- Policy interventions
- Medical trials
- Business decisions (product/feature development...)
- Empirical (Social) Sciences
- ...

# When Not (necessarily)

## Exploratory analysis

Distracting if not enough knowledge about the dataset

## Predictive settings

Interest not in understanding the underlying mechanisms but want to obtain best possible estimates of a variable you do not have by combining others you do have (e.g. Kriging)



# Hurdles to Causal Inference

# Hurdles to causal inference

Causation *implies* Correlation

Correlation *does not imply* Causation

*Why?*

- Reverse causality
- Confounding factors/endogeneity

# Reverse Causality

There *is* a causal link between the two variables but it either runs the opposite direction as we think, or runs in both

E.g. Education and income

# Confounding Factors

Two variables are correlated because they are *both* determined by other, unobserved, variables (factors) that *confound* the effect

E.g. Ice cream and cold beverages consumption

# Strategies

*Is there any way to overcome reverse causality and confounding factors to recover causal effects?*

The key is to get an “*exogenous source of variation*”

# Strategies

## Randomized Control Trials

*Treated Vs control* groups. Probability of treatment is independent of everything else

## Quasi-natural experiments

Like a RCT, but that just “*happen to occur naturally*” (natural disasters, exogenous law changes...)

## Econometric techniques

For the interested reader: space-time regression, instrumental variables, propensity score matching, differences-in-differences, regression discontinuity...

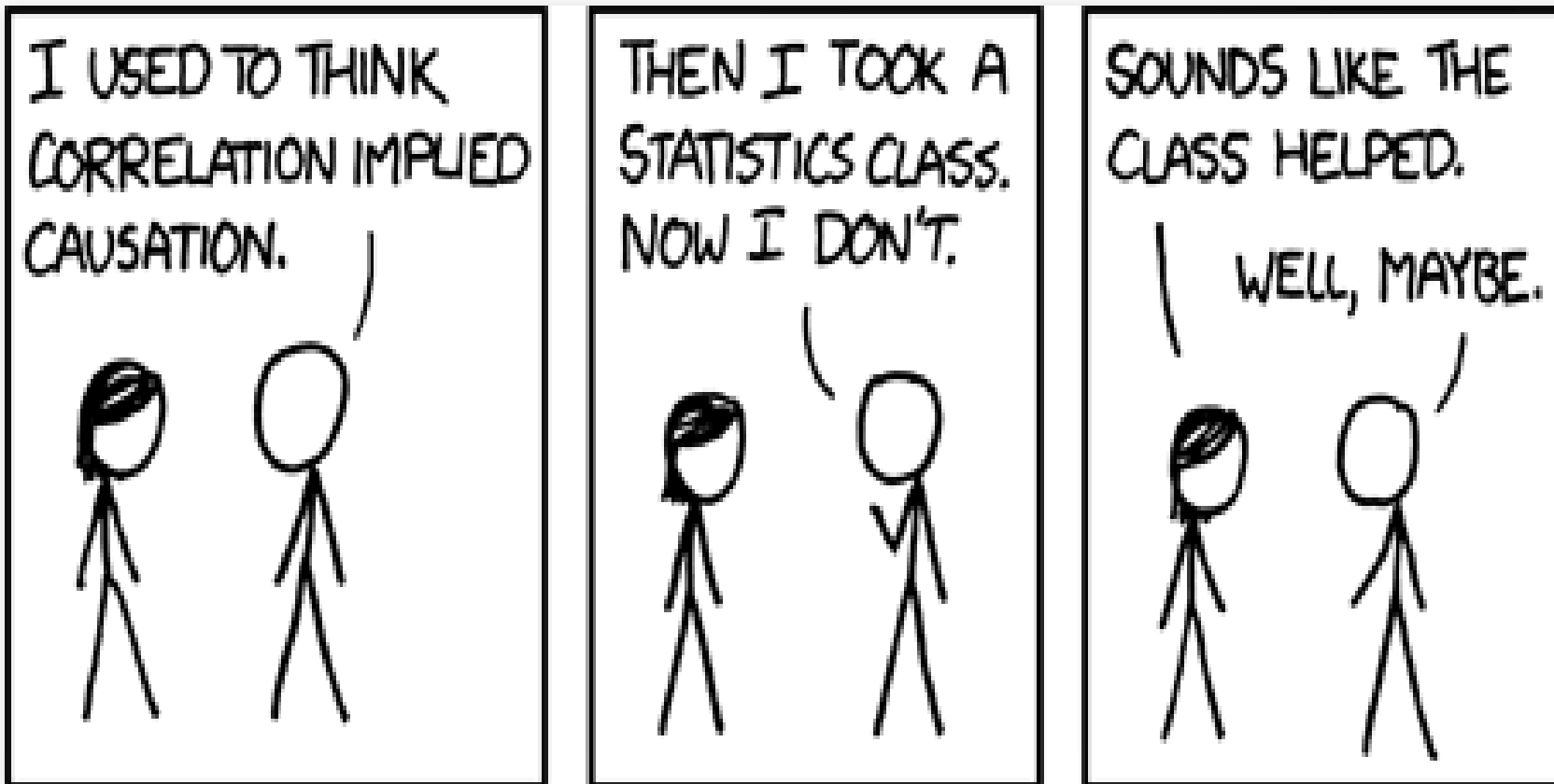
# Correlation *or* Causation?

Establishing causality is much harder than identifying correlation, but sometimes it's needed to move forward!

Correlation *precludes* causation and, in some cases, it is all that is needed.

It is **important** to always draw *conclusions based on analysis*, know what the data can and cannot tell, and stay **honest**.





[Source]

```
<a rel="license" href="http://creativecommons.org/licenses
```