Geographic Data Science - Lecture IX

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

"Association breeds similarity" (sometimes)

Nasir bin Olu Dara Jones (a.k.a. *Nas*)

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?

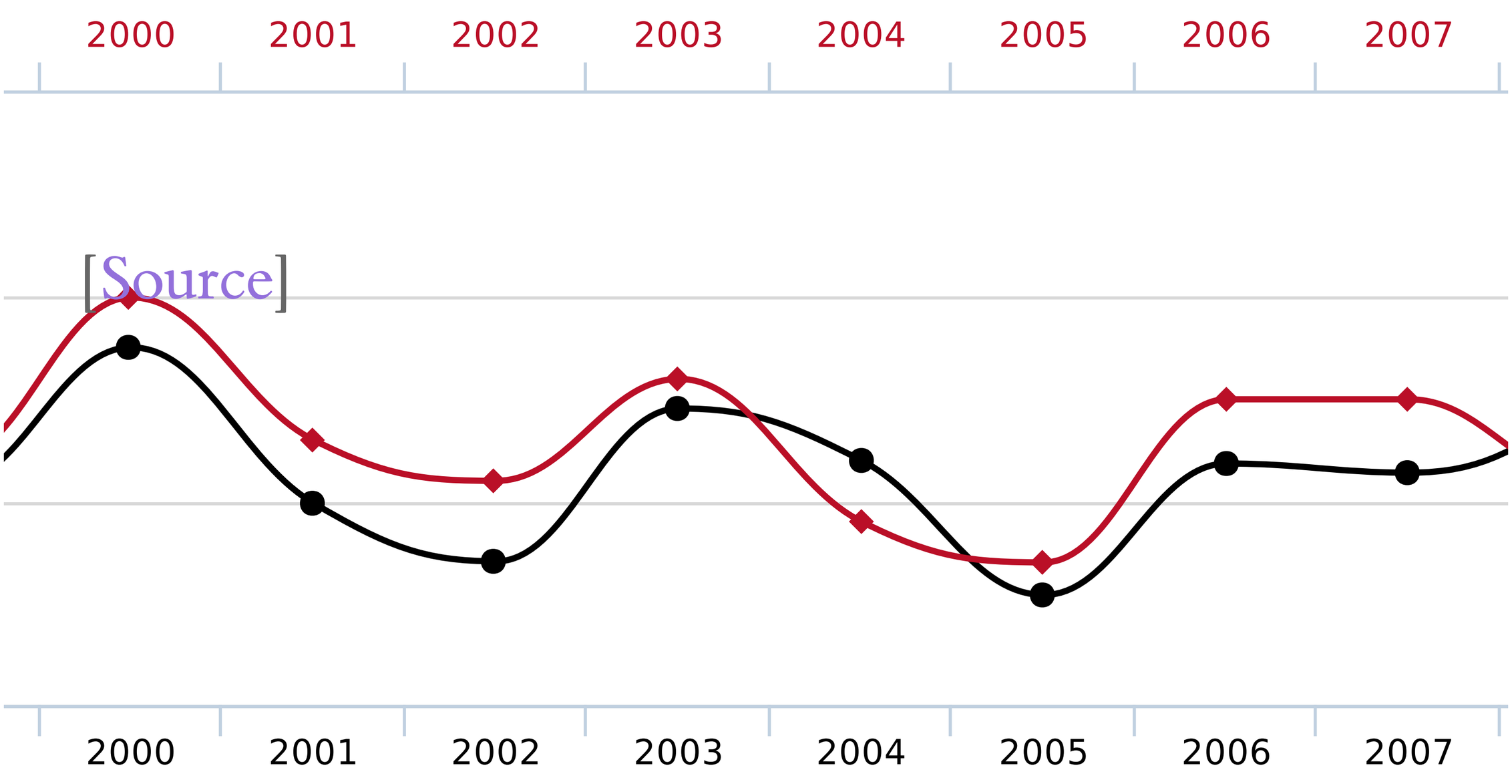
Take a guess (2mins)...

- Temperature and ice-cream consumption → **Positive. Positive.**
- Non-commercial space launches & Sociology PhDs awarded
- Crime & policing
- IMD Moran Plot in Liverpool

ldwide non-commercial space launc

correlates with

Sociology doctorates awarded (US)



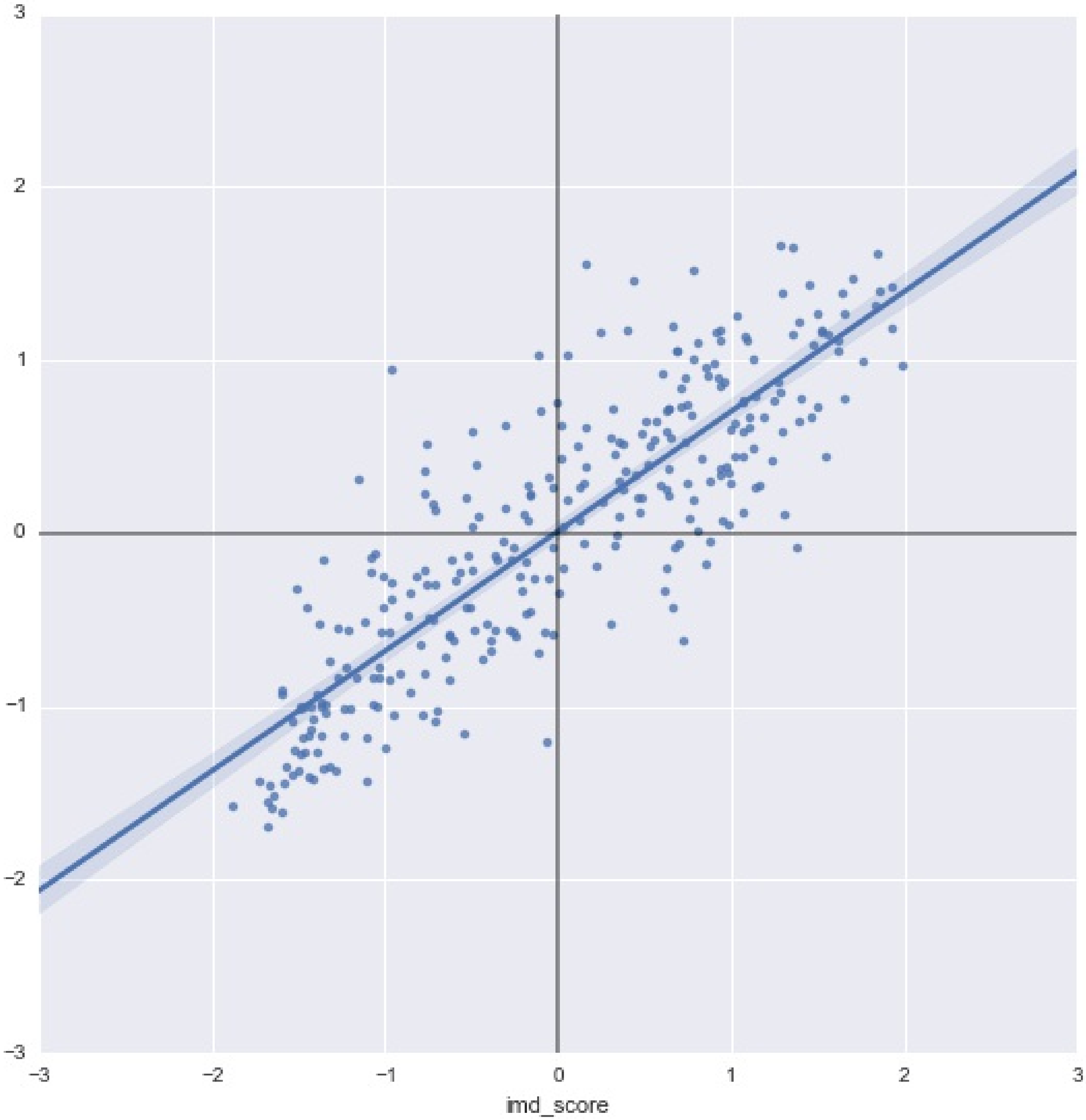
Sociology doctorates awarded (US) Worldwide non-commercial space l

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 Moran Plot in Liverpool


3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
imd_score

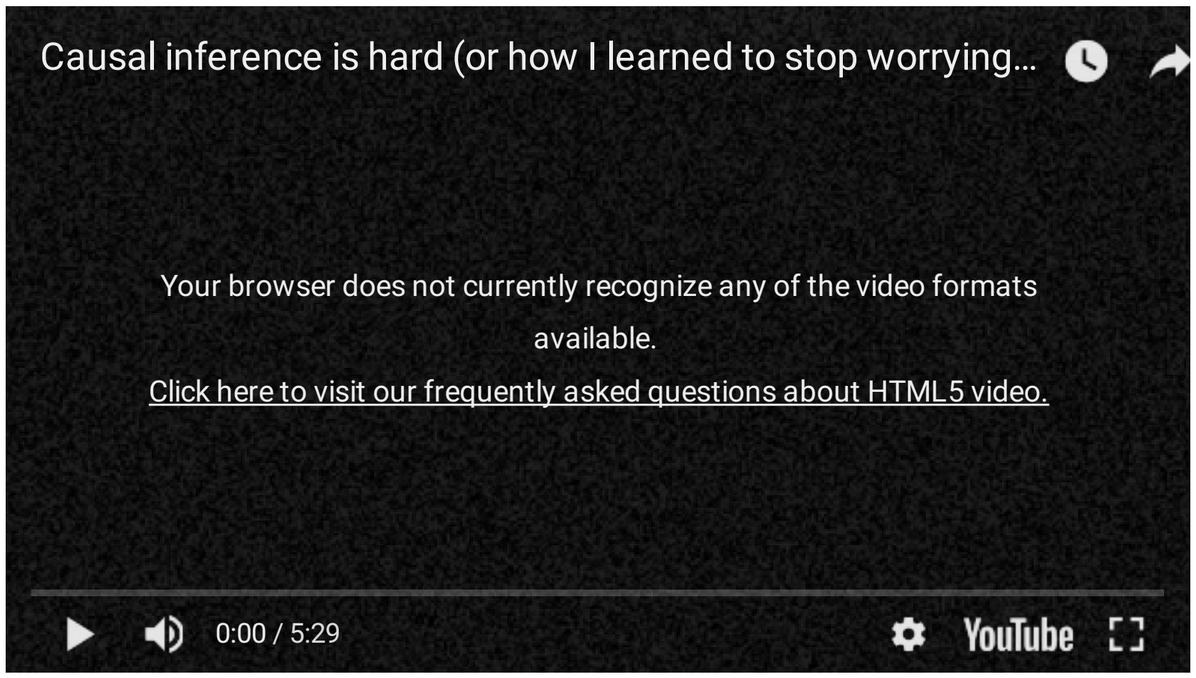
Examples

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 Moran Plot in Liverpool → **Positive. ?**

Causal inference


Causal inference is hard (or how I learned to stop worrying...
Your browser does not currently recognize any of the video formats available.
Click here to visit our frequently asked questions about HTML5 video.
0:00 / 5:29
YouTube

[Source]

Why/When 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

When you are not sure what you are after, inferring causality might be too high of a price to pay to get a sense of the main relationships

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

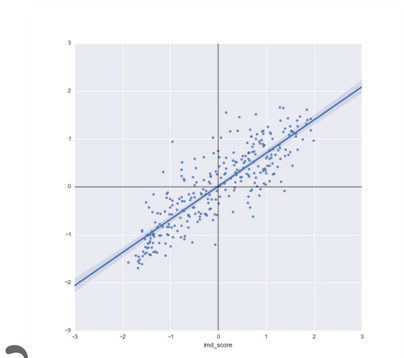
E.g. Population density in a specific point using population density in all available nearby locations

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 oposite 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 and *control* groups

Probability of treatment is independent of everything else

Quasi-natural experiments

Like a RCT, but that just *"happen to occur naturally"* (natural dissasters, exogenous law changes...)

Econometric techniques

For the interested reader: space–time regression, instrumental variables, propensity score matching

So, why correlation at all?

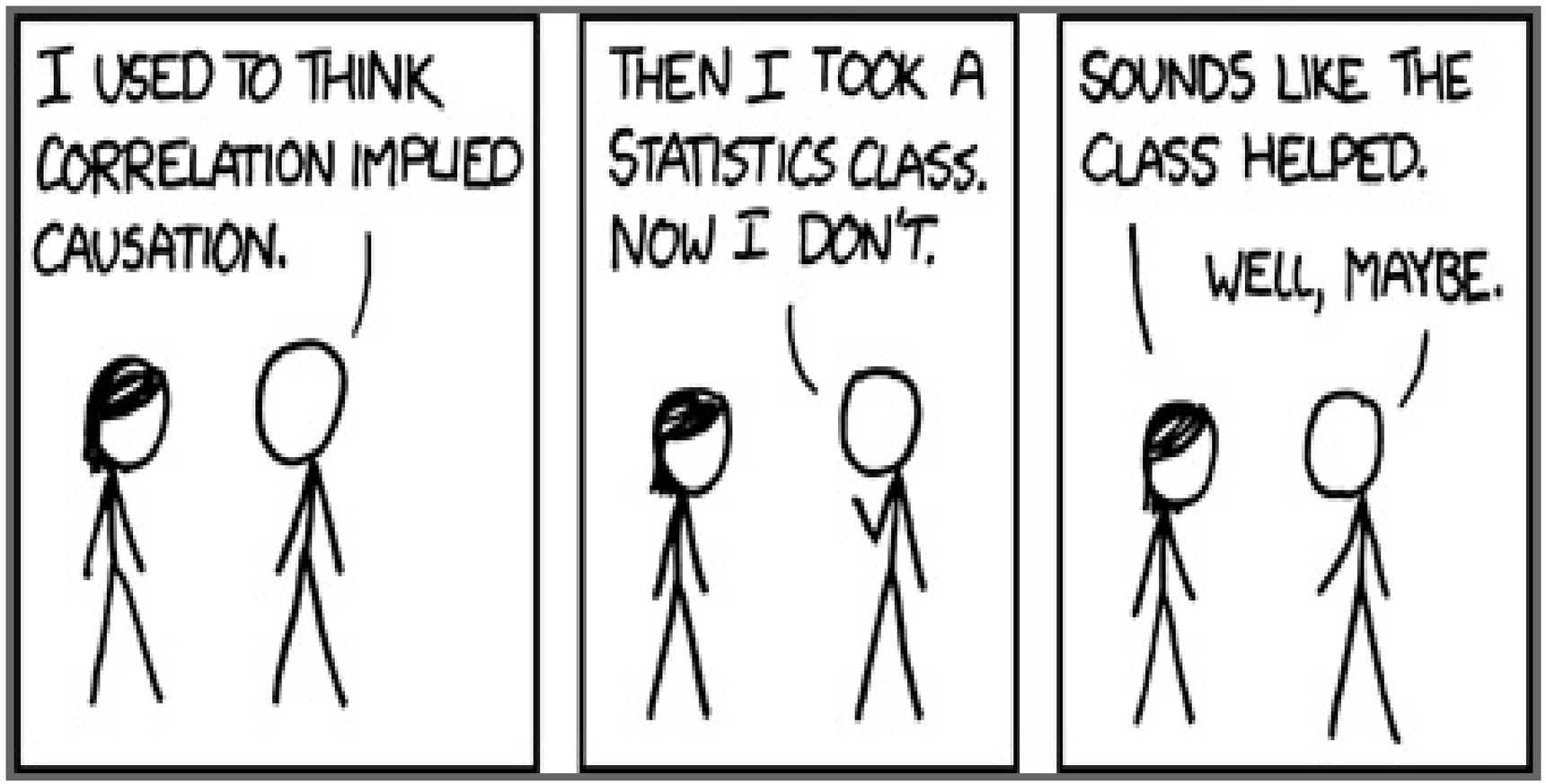
Establishing **causality is much harder** than identifying correlation, and sometimes it is just not possible with a given dataset (e.g. many observational surveys).

... correlation most often *precludes* causation and, depending on the application/analysis, 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**.

Recapitulation

- Correlation does **NOT** imply causation
- Causality implies more than correlation, a direct **effect channel** that is **harder** to identify but might be **worthwhile**
- There are several techniques to identify causality, all usually based on obtaining **exogenous sources of variation**
- You don't always need causality


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

[Source]

Geographic Data Science'16 - Lecture 9 by Dani Arribas-Bel is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.