



INF 110 Discovering Informatics

Cause and Effect



Some content © John DeNero
and Ani Adhikari.

In this lecture:

- I will talk about how *observational studies* work
- What types of conclusions can we make when we ask:
 - *Does A cause B?*

Observational Studies

A study where scientists collect and interpret data but don't actively control an experiment

(which is what they might do in a laboratory environment)

You can probably think of some examples..

Coffee - good.

HEALTH AND SCIENCE

BIOTECH AND PHARMA | HEALTH INSURANCE | HOSPITALS | SCIENCE

Three coffees a day linked to range of health benefits, study says

- A review of more than 200 studies found coffee consumption was "more often associated with benefit than harm"
- Meantime, while three of four cups of coffee was likely to be the optimum number, those who drank as many as seven cups a day still appeared to benefit, the study said
- Health experts warned people should not start drinking coffee, or increasing their intake, for health reasons

Sam Meredith | [@smeredith19](#)

Published 4:15 AM ET Thu, 23 Nov 2017 | Updated 10:43 AM ET Fri, 24 Nov 2017



Chocolate – very good.

Dark chocolate can improve stress, mood, memory and immunity, studies claim

Ashley May

Published 9:05 AM ET Fri, 27 April 2018

USA TODAY



Diana Miller | Getty Images

Dark chocolate

Dark Chocolate Reduces Stress In Humans, According To A New Study

By LUCIA PETERS | Apr 25 2018 |



Processed meat - bad.

Processed meats may increase cancer risk, Tufts University says



The Tufts report suggests reducing or eliminating consumption of hot dogs, corned beef, sausage and other processed meats. (Photo: Salwan Georges, Detroit Free Press)

Susan Selasky, Detroit Free Press Published 6:00 a.m. ET Jan. 16, 2019 | Updated 7:40 a.m. ET Jan. 16, 2019

Share your feedback to help improve our site experience!

MORE STORIES



Travel Channel visits Detroit-area restaurants
Jan. 16, 2019, 6 a.m.



Report: Processed meats may pose cancer risk
Jan. 16, 2019, 7:40 a.m.



Asian-spiced hurricane

Eating even a little bit of processed meat regularly can increase your risk of some deadly cancers

Hilary Brueck Dec. 19, 2018, 9:30 AM



smpics/Getty Images

■ Eating red meat and processed meat has been linked to higher cancer rates.

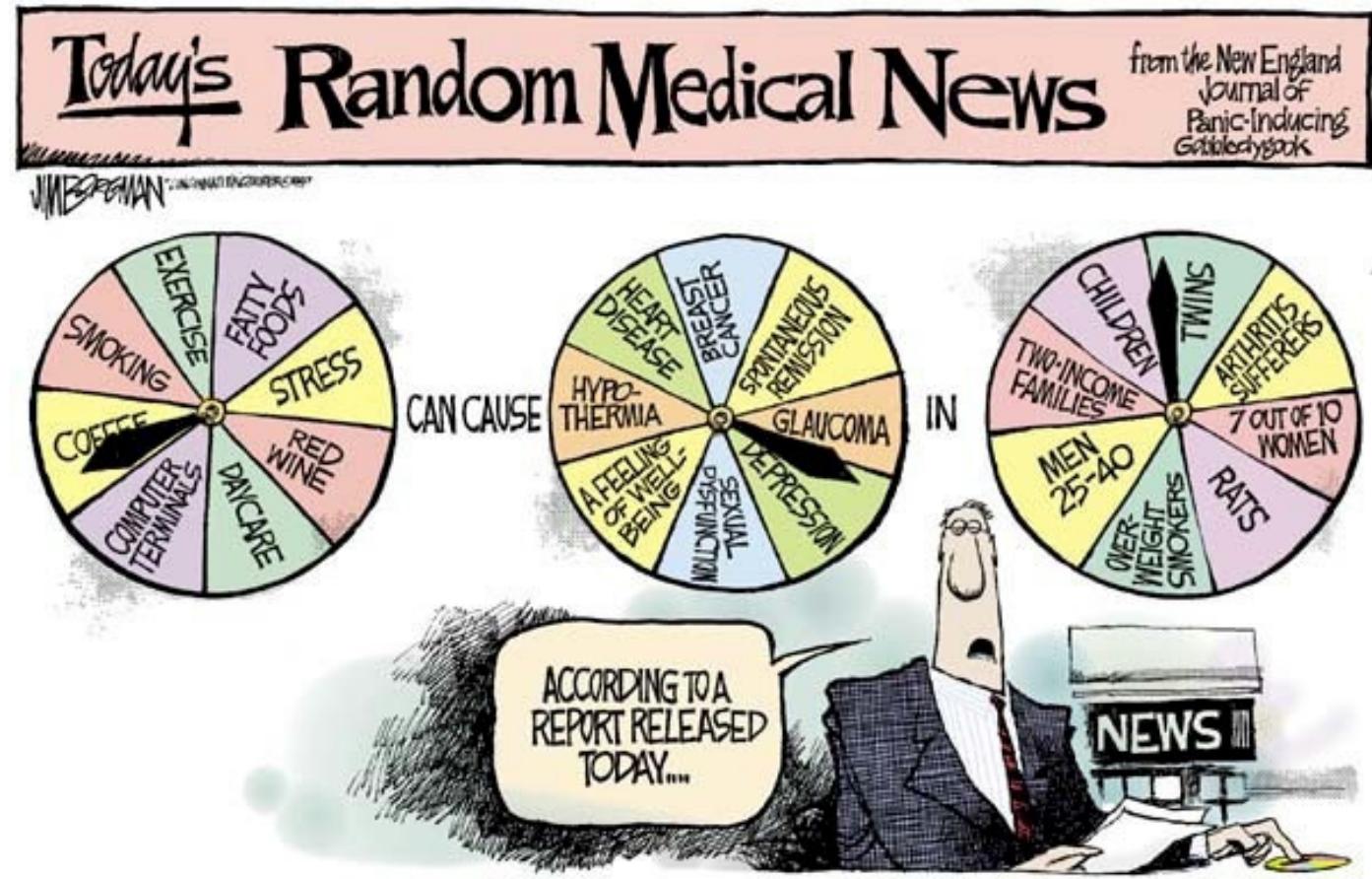
■ But research is increasingly finding that processed meats are much worse for you than other kinds.

■ The average omnivore may be consuming more processed meat than is healthy.



But what does this all mean?

- How can you collect information from a group of people and determine that one habit is related to a particular outcome?
- How can you then apply that information to make better choices?
- This is where ***informatics techniques*** come in.



Cartoon by Jim Borgman, first published by the Cincinnati Inquirer and King Features Syndicate Apr 27, 1997

Components of an Observational Study

- **Population sample** – random selection of individuals from the population you intend to study
Example: 200 American adults
- **Treatment** – the factor under consideration
Example: Processed meat consumption
- **Outcome** – the possible effect
Example: Cancer

Components of an Observational Study

- **Population sample** – random selection of individuals from the population you intend to study

Example: 200 American adults

- **Treatment** – the factor under consideration

Example: Processed meat consumption

- **Outcome** – the possible effect

Example: Cancer

Now you can ask the question:
In my sample, does the treatment lead to the outcome?

Or:

In 200 American adults, does processed meat consumption lead to cancer?

Inferences

Observational studies seek to draw inferences or associations between differences in population properties or behaviors

So really what you want to ask is:

Is there an **association** between processed meat consumption and cancer?

When a report says:

“for processed meat, every 50 grams consumed daily -- about one hot dog -- linked to a 16 percent increased risk of [colorectal] cancer.”

It's because:

The data was based on reported meat consumption and incidence of colorectal cancer.

- both of these things were measured in the sample.

Finding that increased meat consumption tracks cancer incidence points to an **association**.

Causality

- But does eating meat cause colorectal cancer?
- Or does being predisposed to colorectal cancer cause you to eat more meat?
- Or is there another relationship?
- Or is it all happenstance?

1854 Broad Street Cholera Outbreak

Cholera is an infectious disease caused by a bacterium humans come into contact with after expose to contaminated water supplies.

But how did we find this out?

Many theories existed, but none were tested:

1854 Broad Street Cholera Outbreak

Miasma theory claimed that diseases such as cholera and bubonic plague were caused by pollution or a noxious form of "bad air"

Suggested remedies:

- "fly to clene air"
- "a pocket full o'posies"
- "fire off barrels of gunpowder"

Proponents:

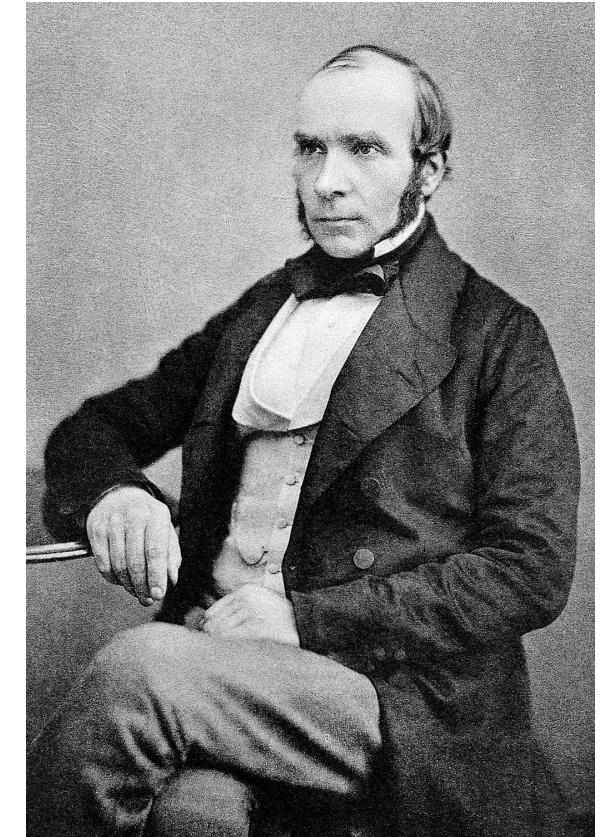
- Florence Nightingale
- Edwin Chadwick, Commissioner of the General Board of Health

Not Everyone Agreed

- John Snow was a skeptic
- He became interested in studying the pattern or spread of the disease

Epidemiology:

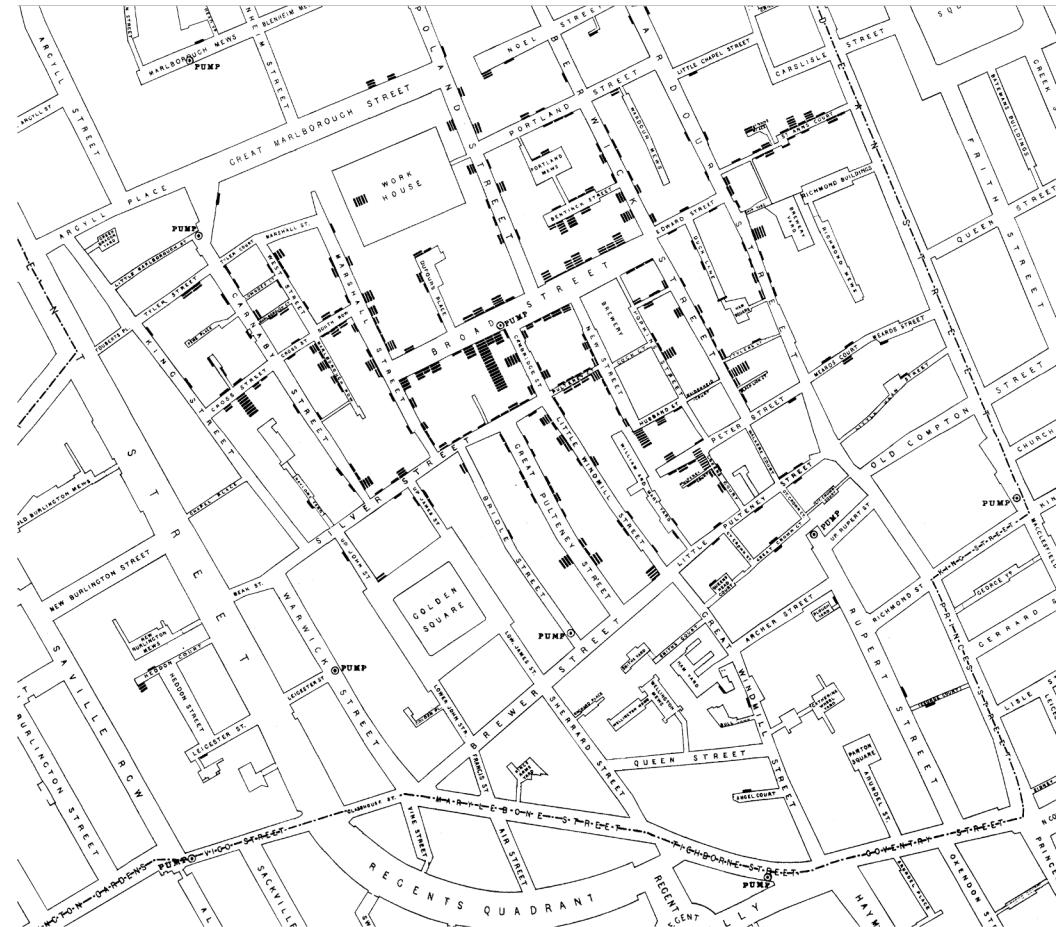
The study of the patterns, distribution and determinants of health outcomes in a population



John Snow

Data Visualization

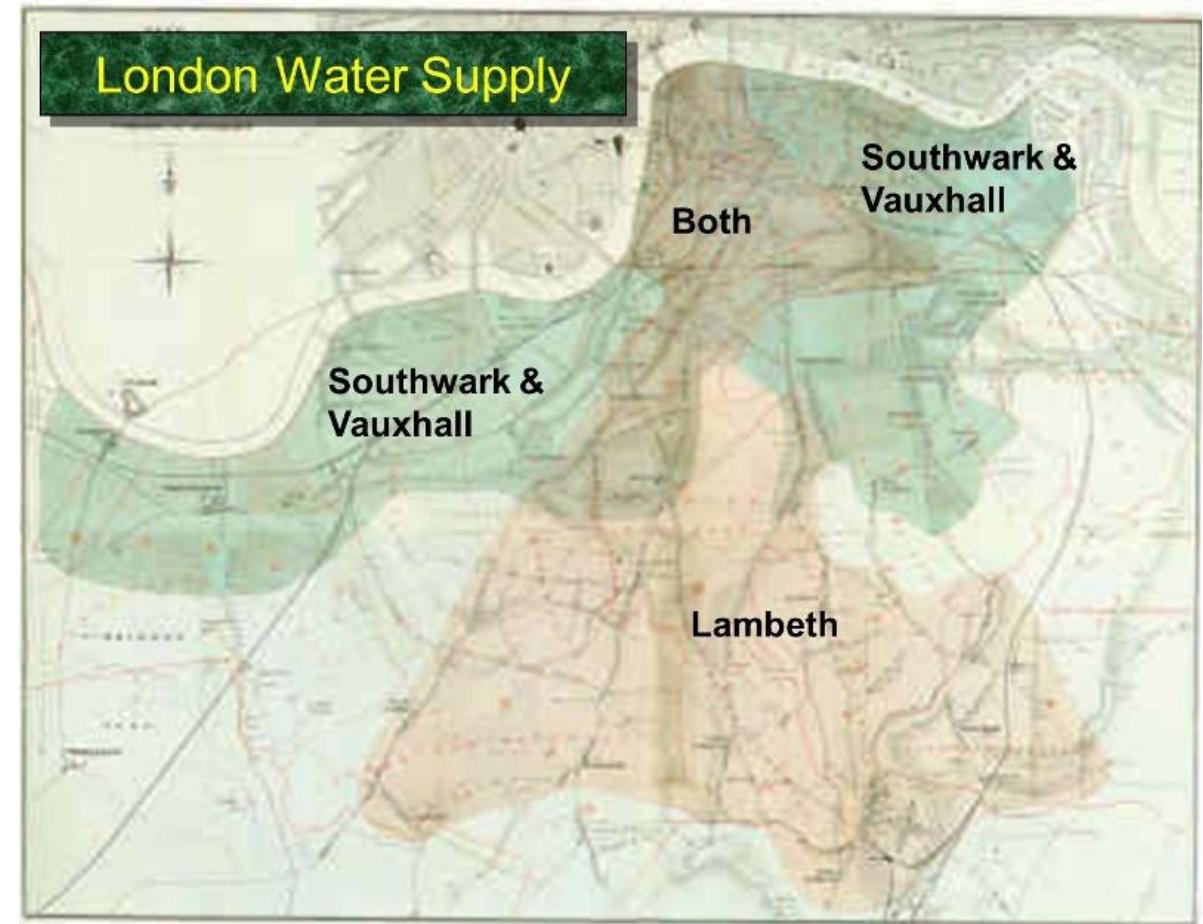
- Snow used a *dot map* to illustrate the cluster of cholera cases around the Broad Street water pump
- He also used statistics to illustrate the connection between the quality of the water source and cholera cases.



Snow's Grand Experiment of 1854

London was getting its water from two places:

- S&V provided water from parts of the Thames close to sewage outlets
- Lambeth provided water further from sewage outlets



Snow's Grand Experiment of 1854

“... there is no difference whatever in the houses or the people receiving the supply of the two Water Companies, or in any of the physical conditions with which they are surrounded ...”

The two groups were similar except for the **treatment**.

In this case, the ***treatment is the different water supplies.***

Snow's Table

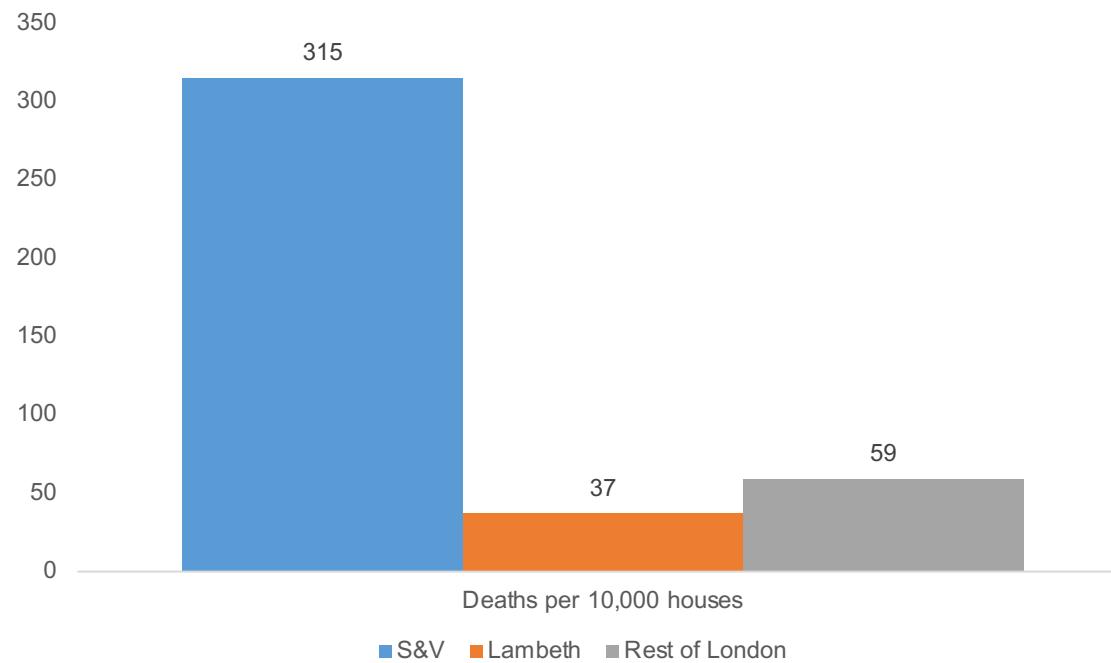
Supply Area	Number of houses	Cholera deaths	Deaths per 10,000 houses
S&V	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

Are the number of cholera deaths simply related to the fact that there are more houses at S&V?

No, not when viewed in terms of deaths per 10,000 houses.

This shows that people living in houses supplied by S&V were more likely to die from cholera.

Charting Snow's Data: Bar Chart



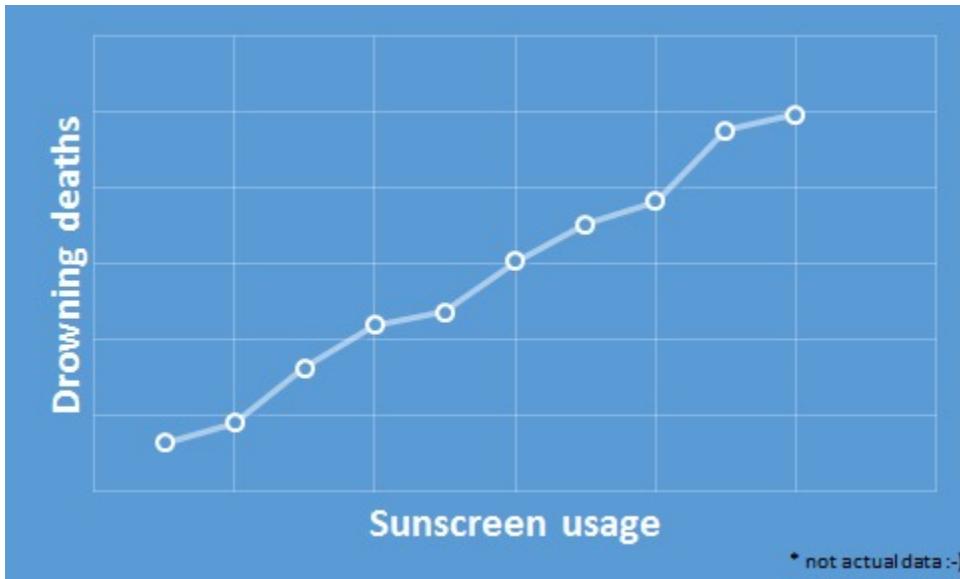
Causality: It's hard to prove

If the treatment and control groups are **similar apart from the treatment**, then differences between the outcomes in the two groups can be ascribed to the treatment.

Confounding Factors

- If the treatment and control groups have systematic differences other than the treatment, then it might be difficult to identify causality.
- Such differences are often present in observational studies.
- When they lead researchers astray, they are called **confounding factors**.

Confounding Correlation Example



- It looks like sunscreen usage is correlated with drowning deaths.
- Does sunscreen usage cause drowning?
- What is a confounding factor here?

[http://www.setheliot.com/blog/2015/05/10/
misleading-correlations/](http://www.setheliot.com/blog/2015/05/10/misleading-correlations/)

Confounding Example: Coffee and Cancer

- Data in the 1960s linked drinking coffee to cancer
- Scientists failed to also note that coffee drinkers also tended to be smokers
- The important link was actually with smoking
- Coffee did not cause the cancer being observed

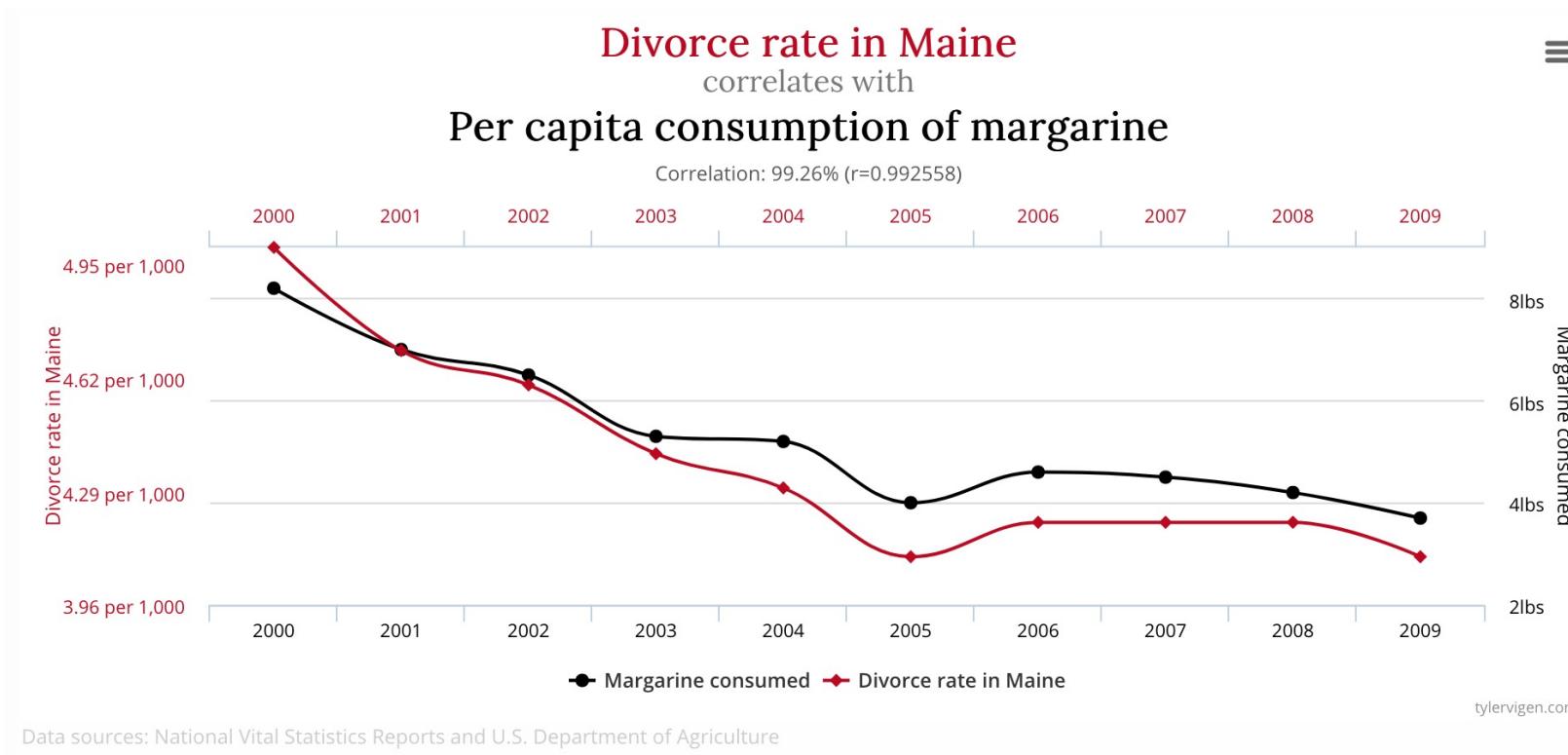


<https://www.youtube.com/watch?v=CxCUHjx7U7Y>

Spurious Correlations

- There's a lot of data in the world (and it's growing)
- Random correlations can emerge by chance
- (especially if you go looking for them)

Spurious Correlation Example

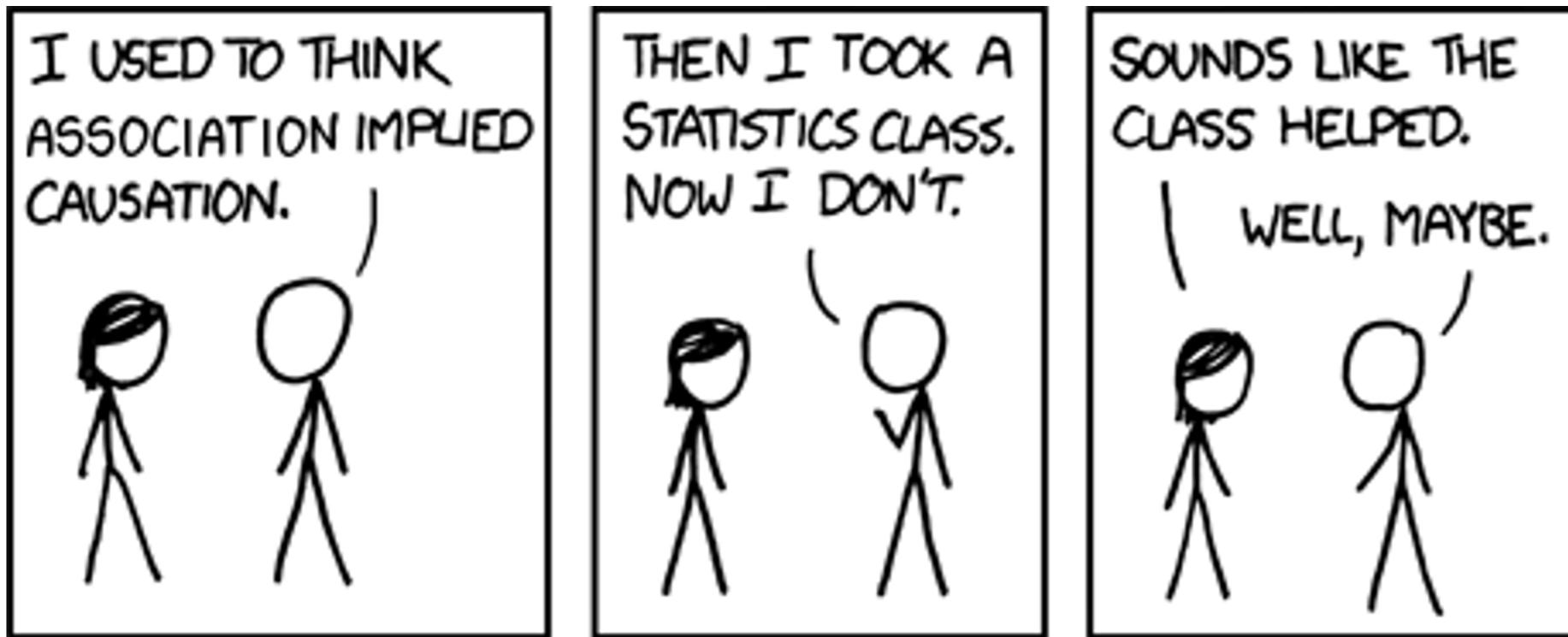


<http://tylervigen.com/spurious-correlations>

Randomization

- If you assign individuals to treatment and control at random, then the two groups are likely to be similar apart from the treatment.
- If you were studying how many people like hamburgers, you wouldn't want to have a large part of your sample coming from people walking out of a McDonald's
- You can account – mathematically – for variability in the assignment.

Association implies causation?



<https://xkcd.com/552/>

end