

YData: An Introduction to Data Science

Lecture 02: Cause and Effect

Elena Khusainova & John Lafferty
Statistics & Data Science, Yale University
Spring 2021

Credit: data8.org



Announcements

- Assignment 0: Something new; to test Anaconda install
- Assignment 1: Also out Friday; due next Thursday at midnight
- Cluster computing environment—accounts avail Feb 5
- Office hours: See Canvas homepage

Questions?

Outline for today

- Demo use of calendar and Jupyter
- Some motivating examples
- Association vs. causation
- John Snow and the Broad Street Pump
- Observational data, confounding, randomization

Demo of Calendar, Anaconda/Jupyter

PHYS ED

The Best Time of Day to Exercise

Men at risk for diabetes had greater blood sugar control and lost more belly fat when they exercised in the afternoon than in the morning.



Getty Images



By Gretchen Reynolds

Jan. 27, 2021

Is it better for our bodies to work out at certain times of day?

A [useful new study of exercise timing and metabolic health](#) suggests that, at least for some people, the answer is a qualified yes. The study, which looked at men at high risk for Type 2 diabetes, found that those who completed afternoon workouts upped their metabolic health far more than those who performed the same exercise earlier in the day. The results add to growing evidence that when we exercise may alter how we benefit from that exercise.

Three coffees a day linked to a range of health benefits

Research based on 200 previous studies worldwide says frequent drinkers less likely to get diabetes, heart disease, dementia and some cancers

Staff and agencies

Wed 22 Nov 2017
19.54 EST



The findings supported other studies showing the health benefits of drinking coffee. Photograph: Wu Hong/EPA

www.theguardian.com/lifeandstyle/2017/nov/23/three-coffees-a-day-linked-to-a-range-of-health-benefits

A Stronger Link?

Pick your NPR Station
There are at least three stations nearby

the salt

NEWSCAST

LIVE RADIO

SHOWS

EATING AND HEALTH

Chocolate, Chocolate, It's Good For Your Heart, Study Finds

LISTEN · 2:07

QUEUE

Download

Transcript

June 19, 2015 · 5:03 AM ET

Heard on Morning Edition



There's a growing body of evidence suggesting that compounds found in cocoa beans, called polyphenols, may help protect against heart disease.
Philippe Huguen/AMF/Bridgeman Images

www.npr.org/sections/thesalt/2015/06/19/415527652/chocolate-chocolate-its-good-for-your-heart-study-finds

Study: Kwok et al. 2015

Observation

- **Individuals, study subjects, participants, units**
 - European adults
- **Treatment**
 - chocolate consumption
- **Outcome**
 - heart disease

The First Question

Is there any relation between chocolate consumption and heart disease?

- Association
 - Any relation
 - Link

Some data:

"Among those in the top tier of chocolate consumption, 12 percent developed or died of cardiovascular disease during the study, compared to 17.4 percent of those who didn't eat chocolate."

[Howard LeWine of the Harvard Health Blog](#)

→ This suggests there may be an association

The next question

Does chocolate consumption lead to a reduction in heart disease?

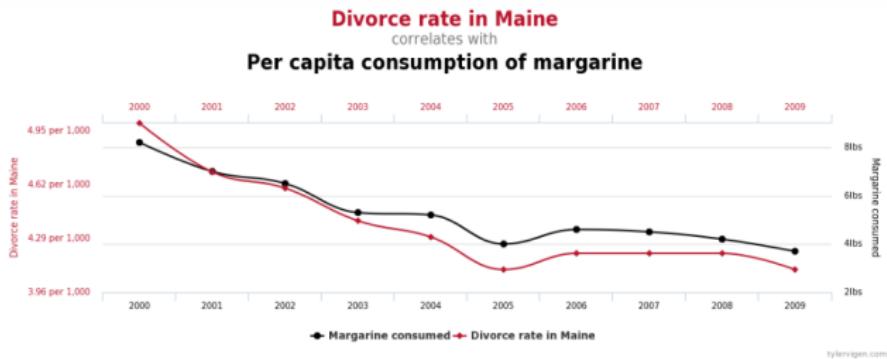
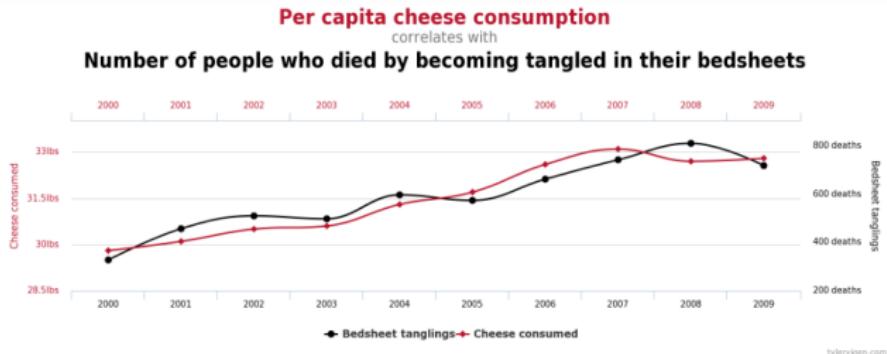
- Causality
→ This question is often harder to answer.

"[The study] doesn't prove a cause-and-effect relationship between chocolate and reduced risk of heart disease and stroke."

→ it is an *observational study*

JoAnn Manson, chief of the Division of Preventive Medicine at Brigham and Women's Hospital in Boston

Association ≠ Causation



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

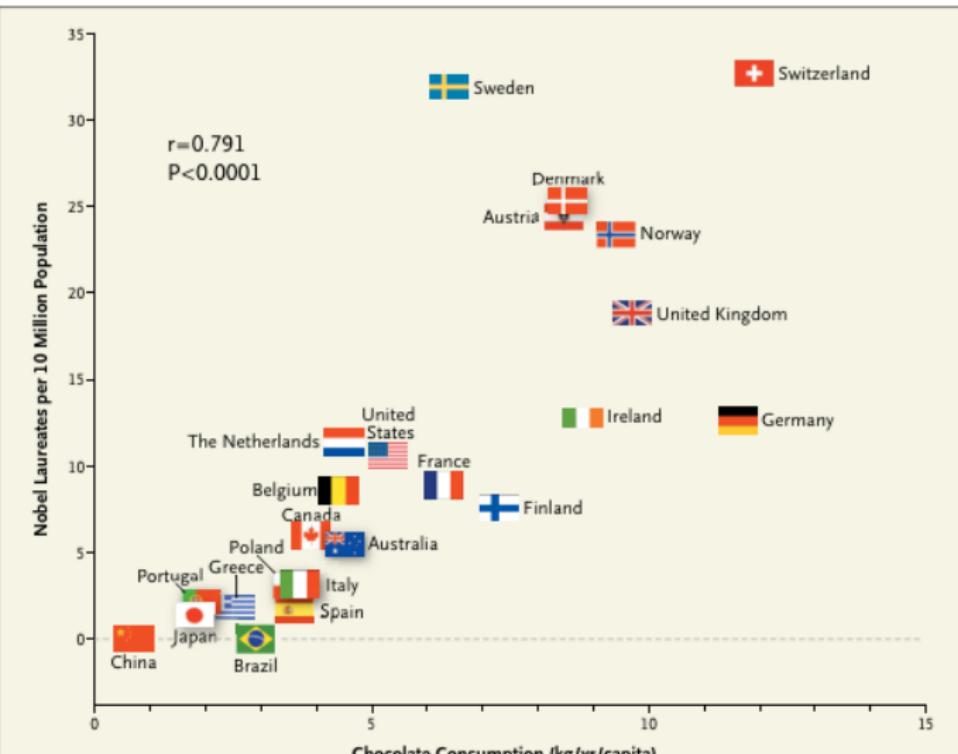
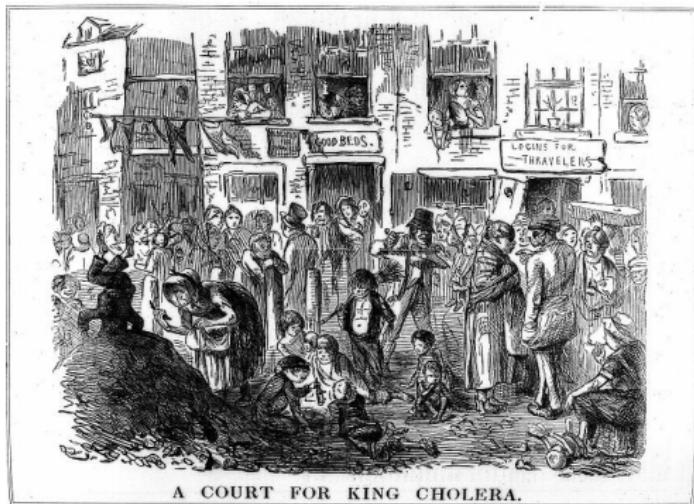


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Messerli 2012 (New England Journal of Medicine)

Association

London, early 1850's



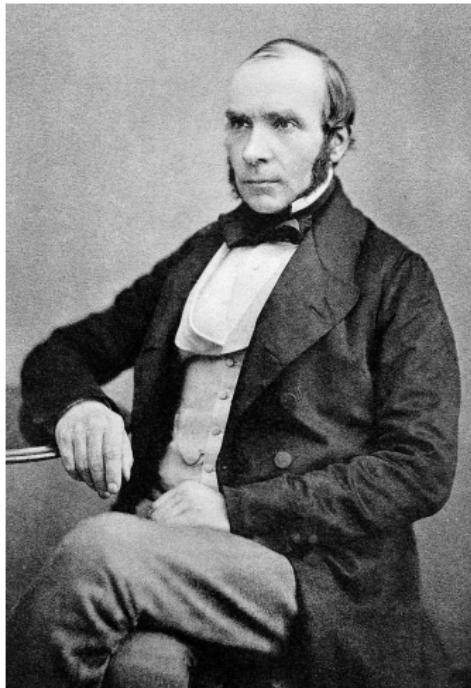
- Cholera reached London in early 1830s
- It was greatly feared in 19th century London as it was often deadly
- Referred to as “King Cholera” in London
- An outbreak in 1849 killed over 14,000 people in London
- Foul-smelling neighborhoods

Illustration by John Leech from Punch (1852).

Miasmas, miasmatism, miasmatists

- Bad smells given off by waste and rotting matter
- Believed to be the main source of disease
- Suggested remedies:
 - “a pocket full o’posies”
 - “fire off barrels of gunpowder”
- Staunch believers:
 - Florence Nightingale
 - Edwin Chadwick, Commissioner of the General Board of Health

John Snow, 1813-1858



John Snow

- Anesthesiologist in London
- Did not believe the miasmas theory
- From examining symptoms, suspected food/drink
- Final suspect: dirty water
- Pioneer of modern epidemiology

50 0 50 100 150 200
Yards

X Pump • Deaths from cholera



john snow london

Back to results

John Snow

4.0 ★★★★☆ · 571 reviews

Pub

Directions

SAVE NEARBY SEND TO YOUR PHONE SHARE

Dark-wood saloon bar serving Yorkshire ales, named after doctor who traced London cholera outbreak.

Cozy · Casual · Groups

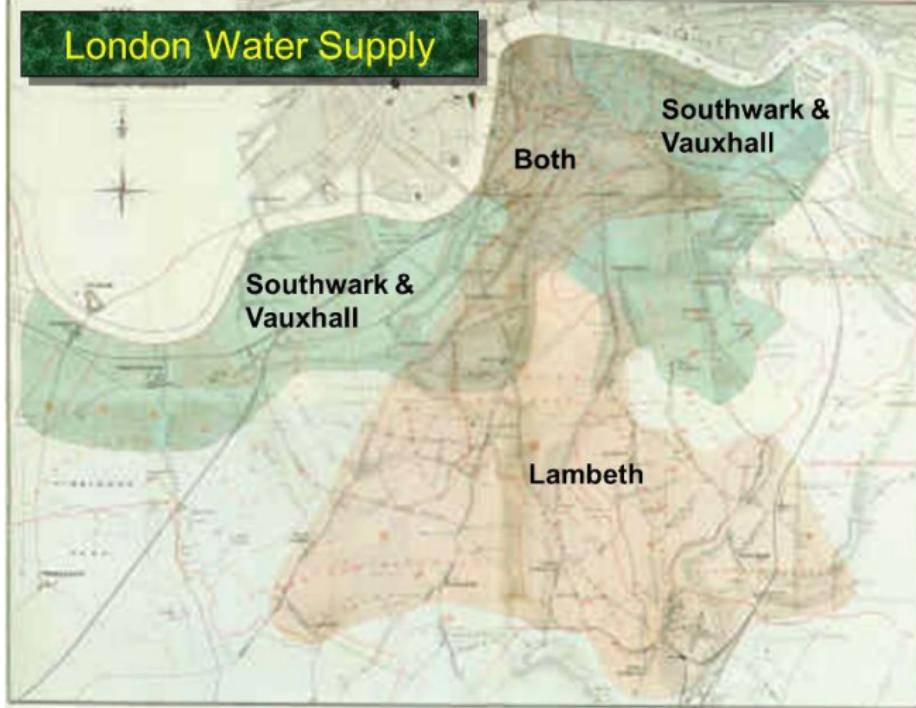
39 Broadwick St, Soho, London W1F 9QJ, UK
+44 20 7437 1344
Open now: 12–11PM

The Google Maps interface shows the location of John Snow pub on Regent Street. The map includes labels for various businesses and landmarks such as The London Palladium, Aqua Spirit, Ping Pong Soho, Reckless Records, Milk & Honey, Cirque le Soir, Red Lion, Old Coffee House, Flat Iron, Bob Bob Ricard, Anthropologie, Nopi, Bill's Soho, Arang, The Glasshouse Stores, and Palace Skateboards. The pub is marked with a red pin and a small description box: "John Snow Intimate pub named after famous doctor". The map also shows the location of the Golden Square area and the Piccadilly Theatre.



Causation

London Water Supply



Lambeth drew water upriver from sewage dump into the River Thames
Southwark & Vauxhall drew from below the sewage dump
Snow focused on intersection ("Both")

Method of Comparison

- **Treatment group**
- **Control group**
→ does not receive the treatment

Compare the outcomes of these two groups
If the results differ, it could suggest an association

Snow's "Grand Experiment"

"... 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.*

Snow's table

	Number of houses.	Deaths from Cholera.	Deaths in each 10,000 houses.
Southwark and Vauxhall Company	40,046	1,263	315
Lambeth Company	26,107	98	37
Rest of London	256,423	1,422	59

Image credit: <https://www.vauxhallandkennington.org.uk/cholera.shtml>

Key to establishing causality

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

Trouble

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**.

Randomize!

- If you assign individuals to treatment and control **at random**, then the two groups are likely to be similar apart from the treatment.
- You can account – mathematically – for variability in the assignment.
- **Randomized Controlled Experiment**
- A “natural experiment” results in data that are a lot like randomized controls

I USED TO THINK
CORRELATION IMPLIED
CAUSATION.

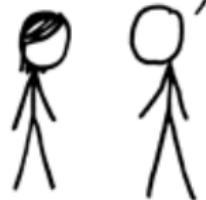


THEN I TOOK A
STATISTICS CLASS.
NOW I DON'T.



SOUNDS LIKE THE
CLASS HELPED.

| WELL, MAYBE.



<https://xkcd.com/552>

Summary

- Causation is different than correlation/assocation
- Be “on the lookout” for studies that suggest causation without sufficient evidence
- Randomized control trials are the gold standard
- Observational data have to be approached with caution
- Confounding variables indicate systematic differences between the control and treatment groups

Readings

Chapter 2 in “Computational and Inferential Thinking”

[https://www.inferentialthinking.com/chapters/02/
causality-and-experiments.html](https://www.inferentialthinking.com/chapters/02/causality-and-experiments.html)

References

- Kwok, C. S., Boekholdt, S. M., Lentjes, M. A., Loke, Y. K., Luben, R. N., Yeong, J. K., Wareham, N. J., Myint, P. K., and Khaw, K.-T. (2015), "Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women," *Heart*, heartjnl–2014.
- Messerli, F. H. (2012), "Chocolate Consumption, Cognitive Function, and Nobel Laureates," *The New England Journal of Medicine*, 367, 16.