

# YData: An Introduction to Data Science

## Lecture 02: Cause and Effect

Jessi Cisewski-Kehe  
Statistics & Data Science, Yale University  
Spring 2020

Credit: [data8.org](https://data8.org)



# Announcements

# Questions

## 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](http://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/AP/ Getty Images

[www.npr.org/sections/thesalt/2015/06/19/415527652/chocolate-chocolate-its-good-for-your-heart-study-finds](http://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?**

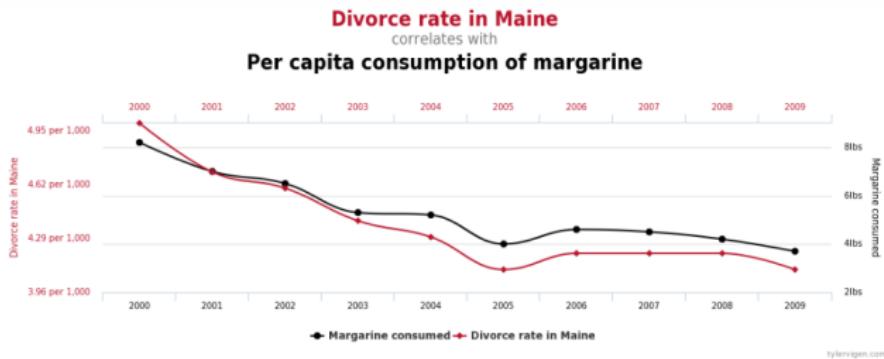
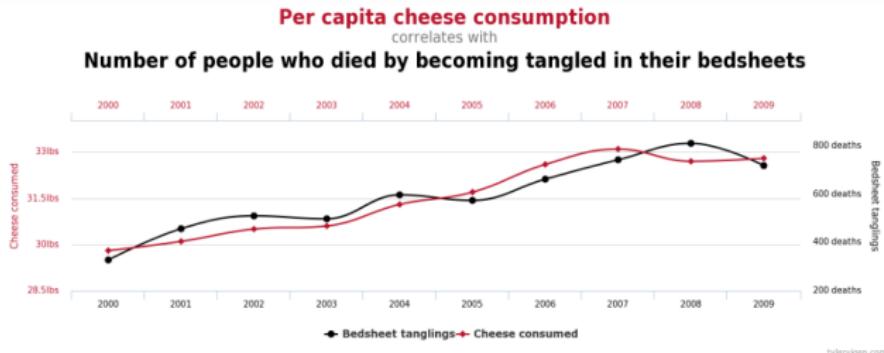
- Causality
  - This question is often harder to answer.

"[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>

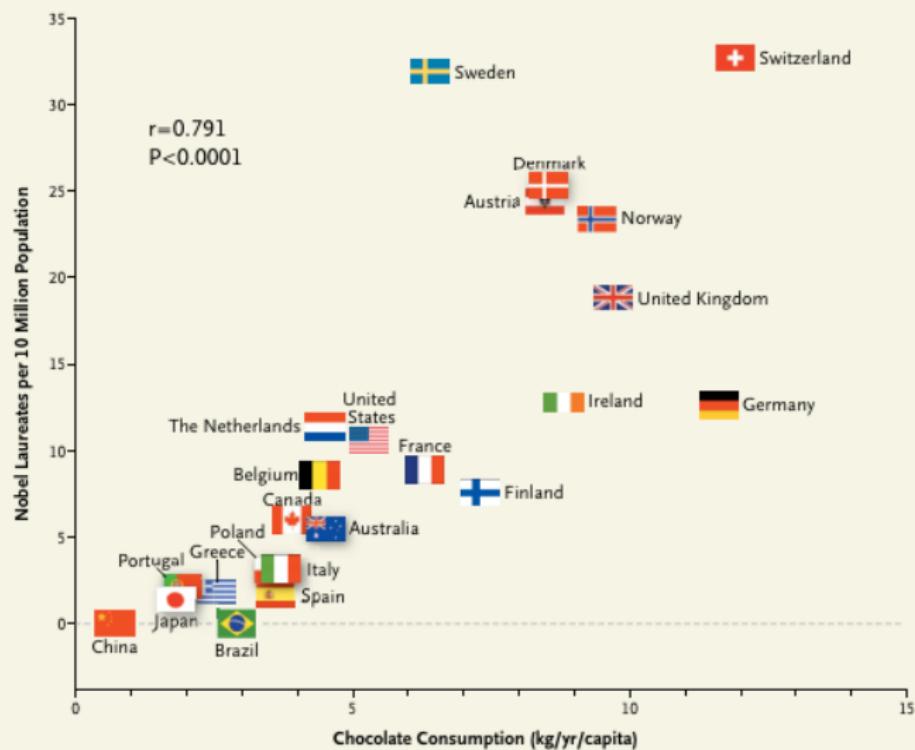


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Messerli 2012 (New England Journal of Medicine)

# Association

# London, early 1850's



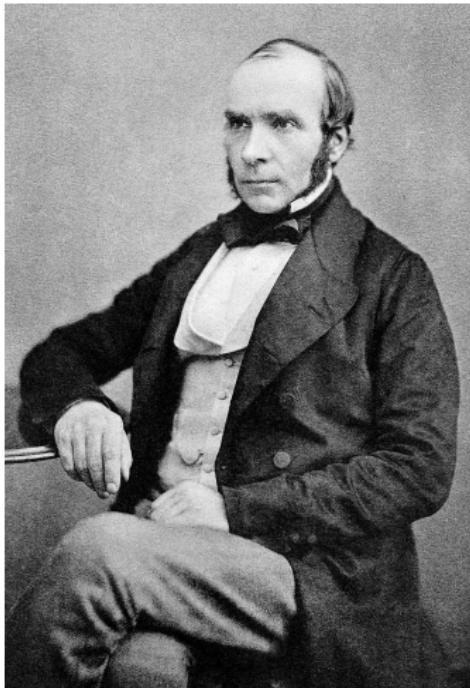
- Cholera reached London in early 1830s
- It was greatly feared in 19th century London as it was often deadly
- Referred to as “King Cholera” in London
- An outbreak in 1849 killed over 14,000 people in London
- Foul-smelling neighborhoods

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:
  - “fly to clene air”
  - “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

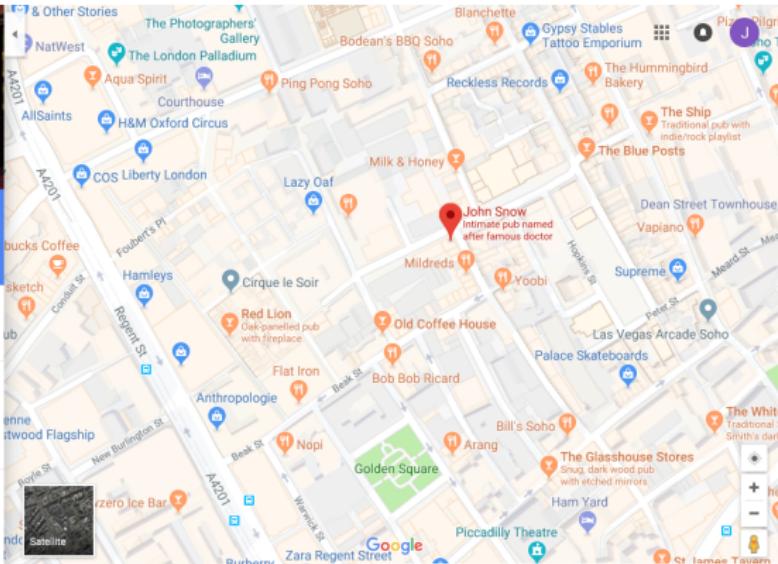
Directions

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

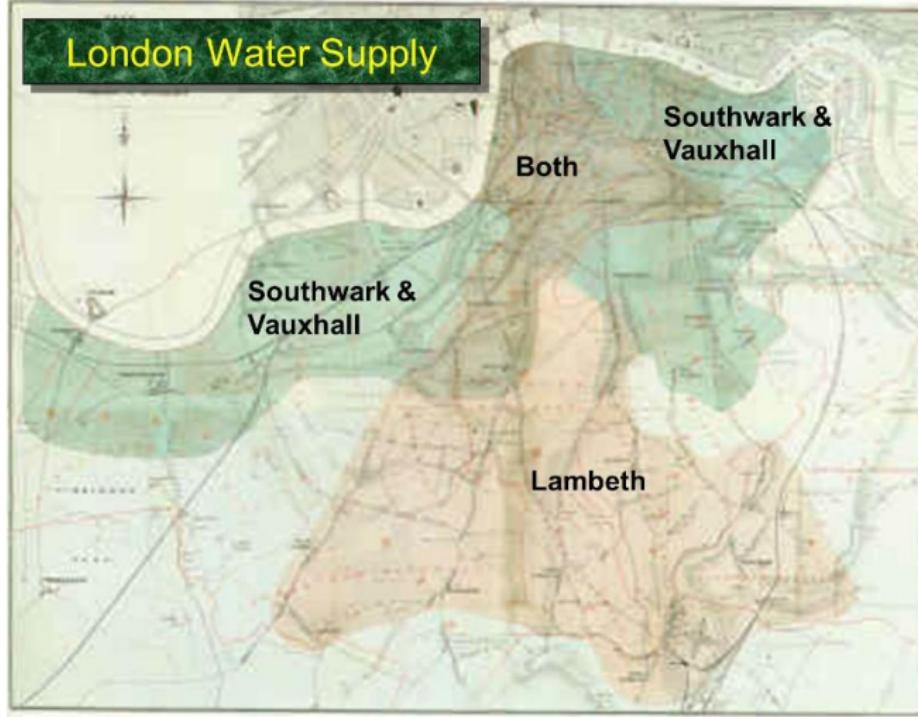
39 Broadwick St, Soho, London W1F 9QJ, UK  
+44 20 7437 1344  
Open now: 12–11PM





# 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.
<b>Southwark and Vauxhall Company</b>	<b>40,046</b>	<b>1,263</b>	<b>315</b>
<b>Lambeth Company . . .</b>	<b>26,107</b>	<b>98</b>	<b>37</b>
<b>Rest of London . . .</b>	<b>256,423</b>	<b>1,422</b>	<b>59</b>

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**

## Careful...

Regardless of what the dictionary says, in probability theory

**Random  $\neq$  Haphazard**

## 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.