YData: An Introduction to Data Science

Lecture 02: Cause and Effect

Jessi Cisewski-Kehe and John Lafferty
Statistics & Data Science, Yale University
Spring 2019

Credit: data8.org

Announcements

Questions

A Link

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

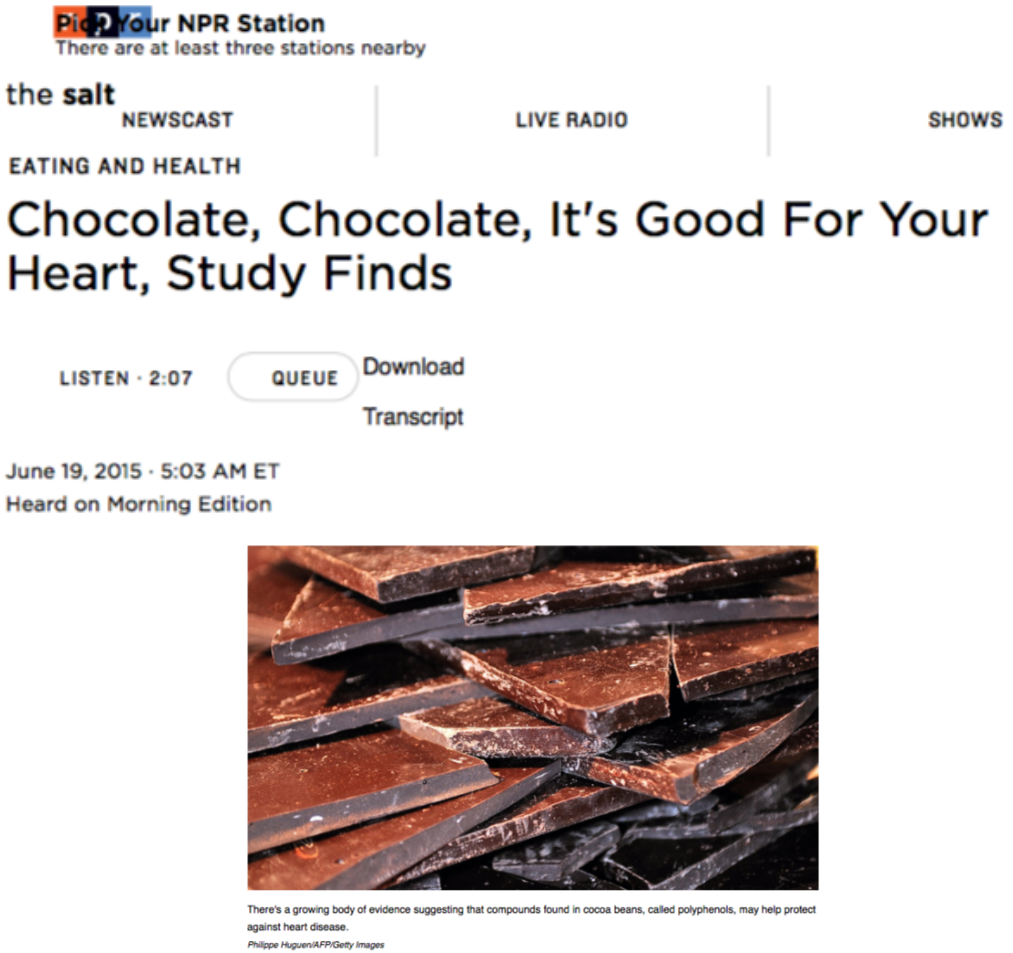
Staff and agencies

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/AFP/Getty 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**

  $\longrightarrow$ European adults

- **Treatment**

  $\longrightarrow$ chocolate consumption

- **Outcome**

  $\longrightarrow$ heart disease

The First Question

Is there any relation between chocolate consumption and heart disease?

- Association
  - $\longrightarrow$ Any relation
  - $\longrightarrow$ Link

An Answer

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

$\longrightarrow$ This suggests there may be an association

The next question

Does chocolate consumption lead to a reduction in heart disease?

- Causality

  $\longrightarrow$ This question is often harder to answer.

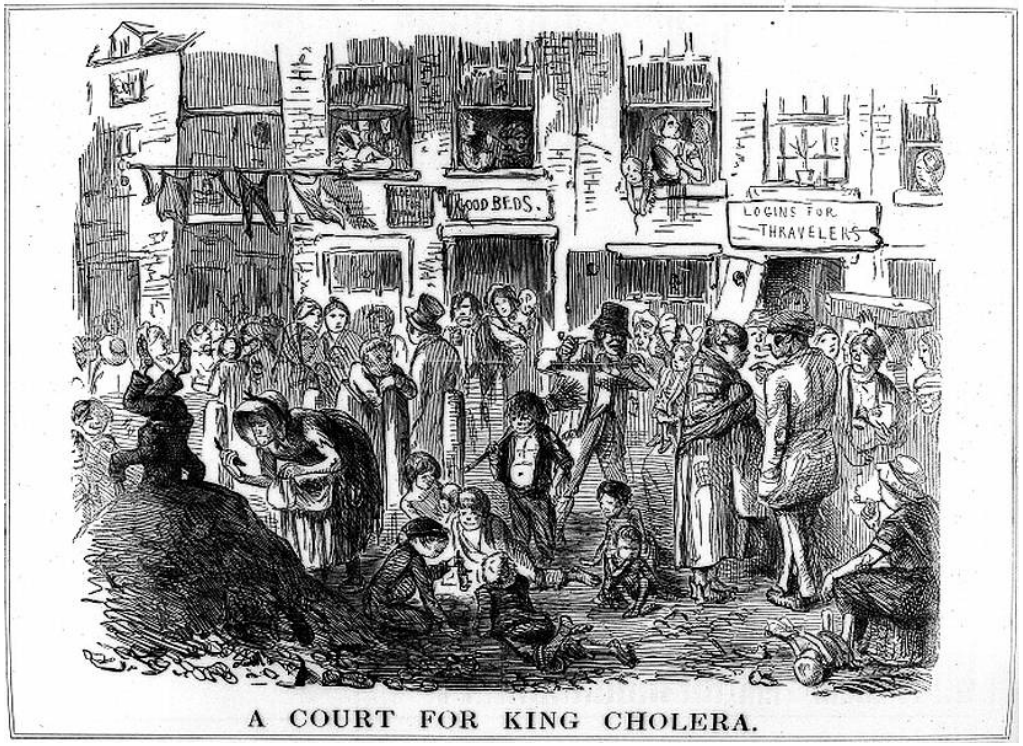
"[The study] doesn't prove a cause-and-effect relationship between chocolate and reduced risk of heart disease and stroke."

$\longrightarrow$ it is an *observational study*

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

Association

London, early 1850's



A COURT FOR KING CHOLERA.

- 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).

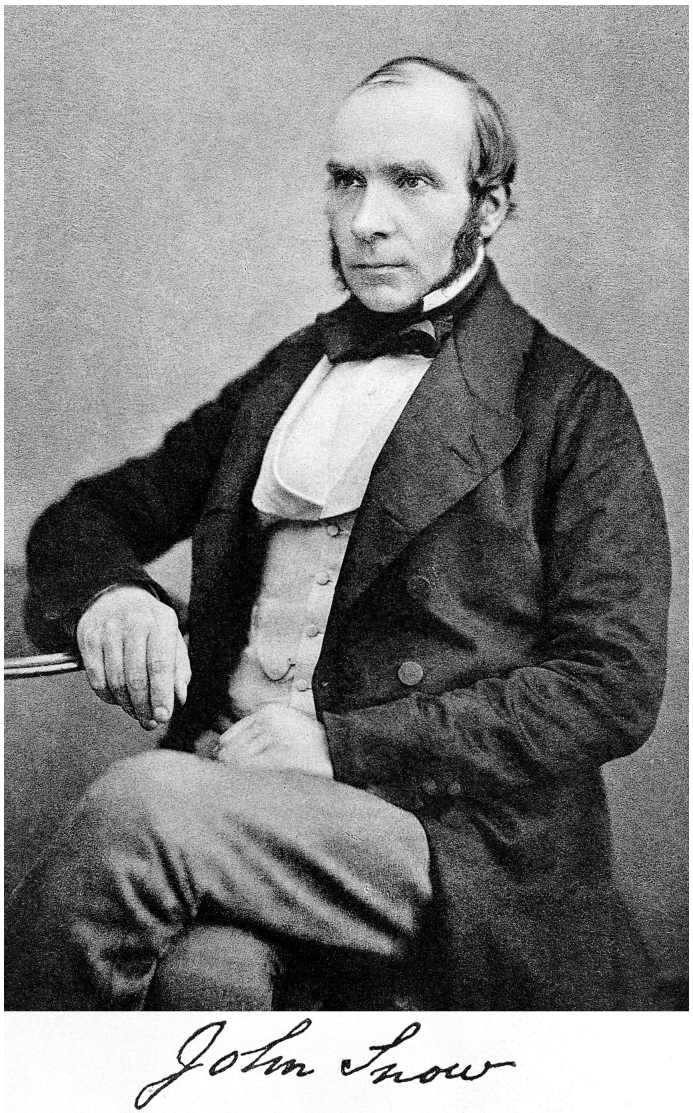
Miasmas, miasmatism, miasmatists

- Bad smells given off by waste and rotting matter
- Believed to be the main source of disease
- Suggested remedies:
  - $\longrightarrow$ "fly to clene air"
  - $\longrightarrow$ "a pocket full o'posies"
  - $\longrightarrow$ "fire off barrels of gunpowder"
- Staunch believers:

  Florence Nightingale

  Edwin Chadwick, Commissioner of the General Board of Health

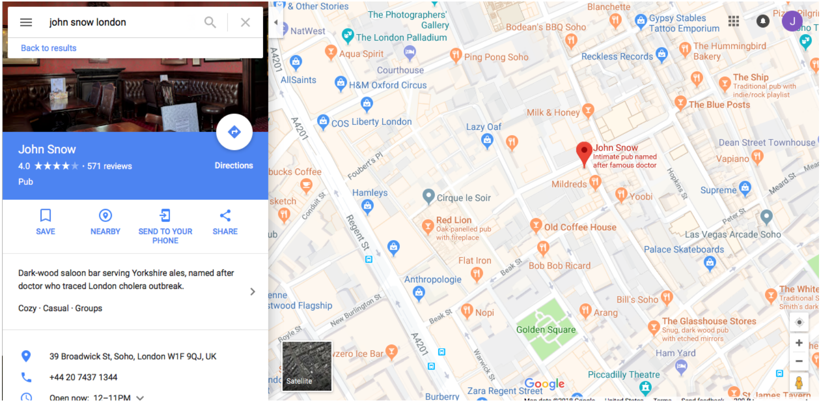
John Snow, 1813-1858



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

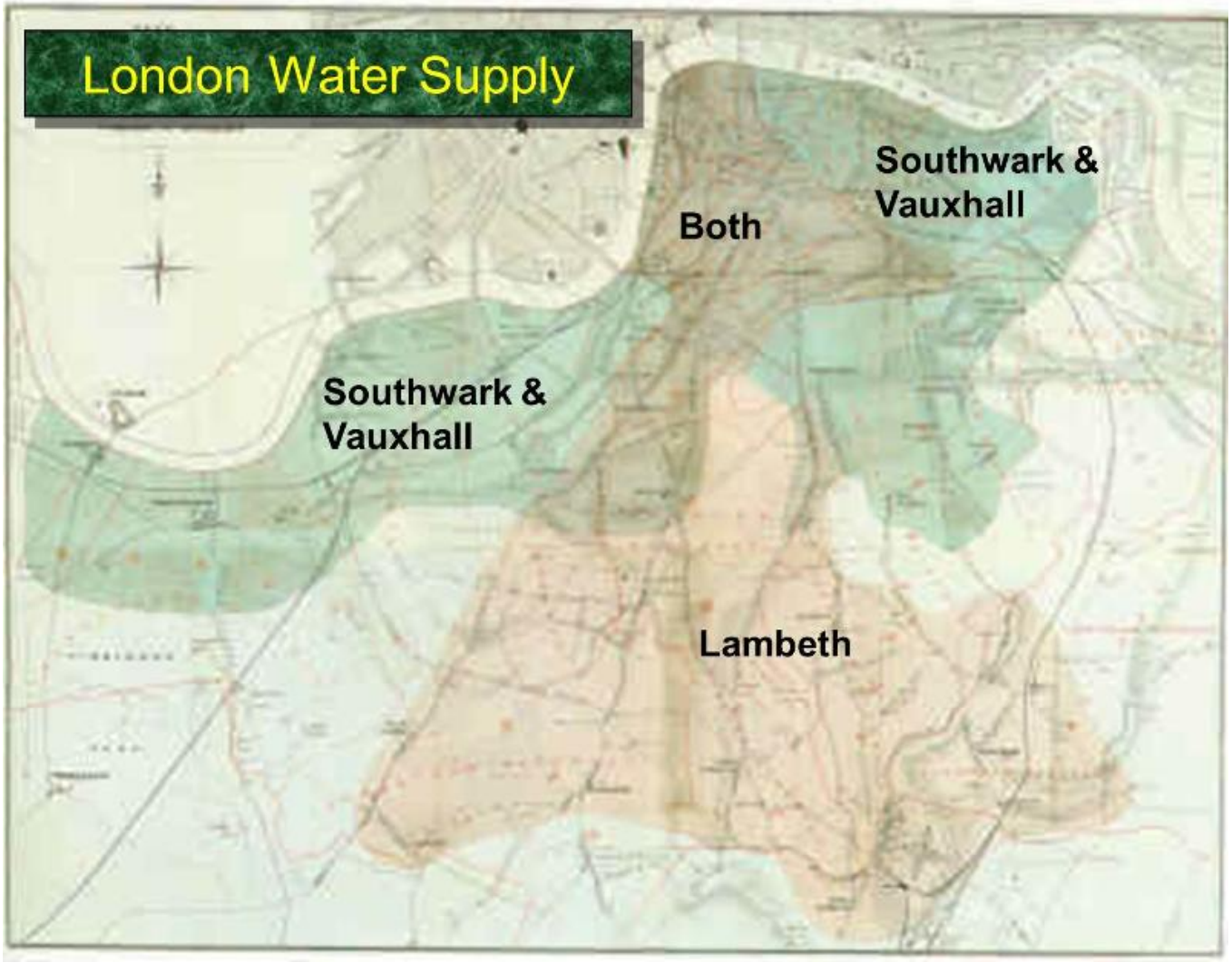
- Pioneer of modern epidemiology


50 0 50 100 150 200 Yards
× Pump • Deaths from cholera
OXFORD STREET
REGENT STREET
GT. MARLBOROUGH ST.
BROAD STREET
WARDOUR STREET
DEAN STREET
CONDUIT STREET
KING STREET
GOLDEN SQUARE
BREWER STREET
SAVILLE ROW
NEW BOND STREET
REGENTS QUADRANT
PICCADILLY


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
John Snow
Intimate pub named after famous doctor
The Photographers' Gallery
The London Palladium
Bodean's BBQ Soho
Blanchette
Gypsy Stables Tattoo Emporium
NatWest
Aqua Spirit
Courthouse
Ping Pong Soho
Reckless Records
The Hummingbird Bakery
AllSaints
H&M Oxford Circus
The Ship
Traditional pub with indie/rock playlist
The Blue Posts
Milk & Honey
COS
Liberty London
Lazy Oaf
Dean Street Townhouse
Vapiano
Foubert's Pl
Hamleys
Cirque le Soir
Mildreds
Yoobi
Supreme
Red Lion
Oak-panelled pub with fireplace
Old Coffee House
Las Vegas Arcade Soho
Palace Skateboards
Regent St
Flat Iron
Bob Bob Ricard
Anthropologie
Beak St
Bill's Soho
The White
Nopi
Arang
Golden Square
The Glasshouse Stores
Snug, dark wood pub with etched mirrors
New Burlington St
Ham Yard
Satellite
Piccadilly Theatre
Zara Regent Street
Burberry
Google


The Red Granite kerbstone
marks the site of the historic
BROAD STREET PUMP
associated with Dr. John Snow's
discovery in 1854
that Cholera is conveyed by water

Causation



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

Method of Comparison

- **Treatment group**
- **Control group**

  $\longrightarrow$ 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**

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.