COMS BC1016
Introduction to Computational Thinking and Data Science

Lecture 2: Cause and Effect

Sept 8, 2025

Course Updates and Reminders

- TA office hours start next week
- Labs start this week!
  - Make sure you are registered for 1017!
- EdStem: Course discussion board (https://edstem.org/us/courses/86257)
  - Post questions you have about labs, homework, office hours, …
  - Do *not* post solutions!

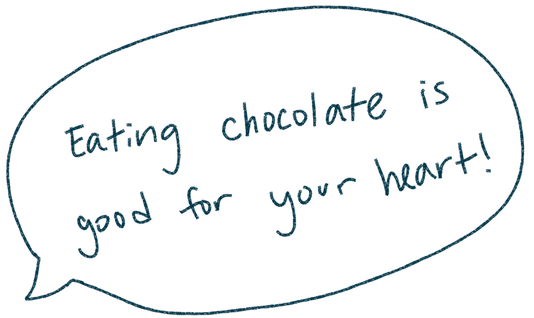
Recall: What is Data Science?

Data science allows us to draw useful conclusions from large and diverse data sets through:

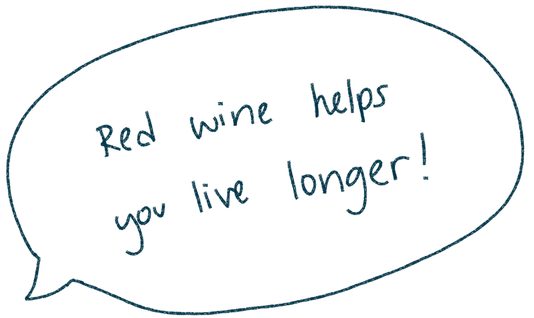
- Exploration: Identifying patterns and trends using data
- Inference: Drawing reliable conclusions using statistics
- Prediction: Making informed guesses about patterns using models

Cause and Effect

A link between chocolate and health



A link between red wine and health



A link between cats and grades

Observational Studies

Definition: A study in which researchers / scientists make conclusions based on *observed data* in which they had no hand in generating

- Individuals: study subjects, participants, units

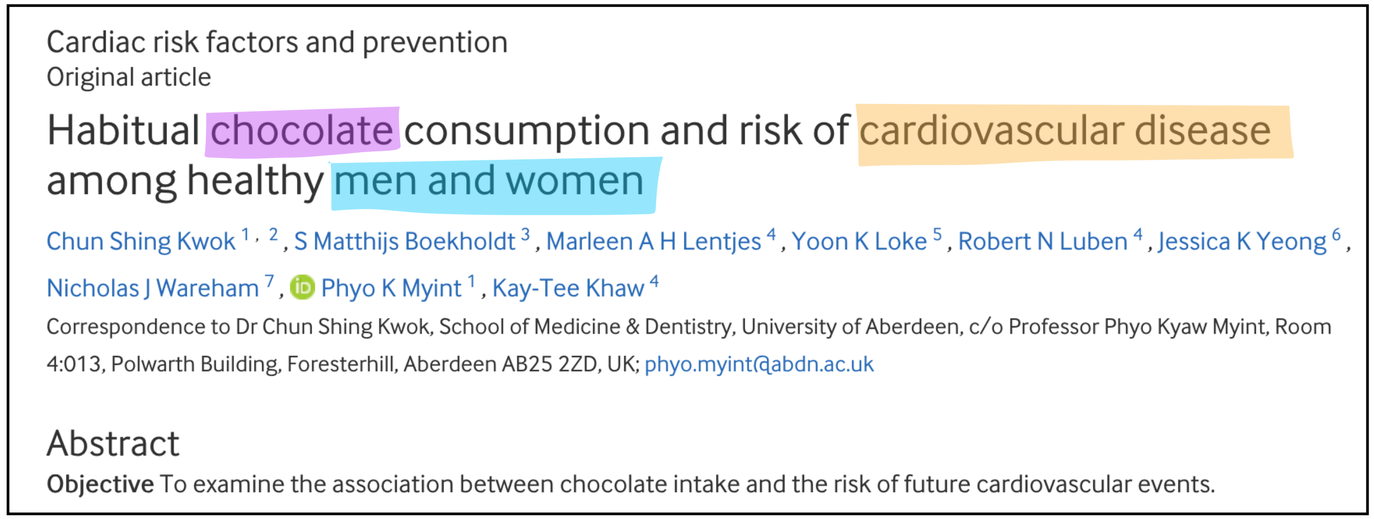
- Treatment: factor of interest

- Outcome: result of treatment

Observational Studies

Definition: A study in which researchers / scientists make conclusions based on *observed data* in which they had no hand in generating

- Individuals: study subjects, participants, units → people
- Treatment: factor of interest → chocolate consumption
- Outcome: result of treatment → heart health



Cardiac risk factors and prevention
Original article

Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok [1, 2], S Matthijs Boekholdt [3], Marleen A H Lentjes [4], Yoon K Loke [5], Robert N Luben [4], Jessica K Yeong [6], Nicholas J Wareham [7], Phyo K Myint [1], Kay-Tee Khaw [4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

Objective To examine the association between chocolate intake and the risk of future cardiovascular events.

Source: https://heart.bmj.com/content/101/16/1279

Association and Causality

Association: Any relation between the treatment and the outcome

Causality: if the treatment causes the outcome to occur

Cardiac risk factors and prevention
Original article

Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok [1, 2], S Matthijs Boekholdt [3], Marleen A H Lentjes [4], Yoon K Loke [5], Robert N Luben [4], Jessica K Yeong [6], Nicholas J Wareham [7], Phyo K Myint [1], Kay-Tee Khaw [4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

Objective To examine the association between chocolate intake and the risk of future cardiovascular events.

Association and Causality

eating chocolate

Association: Any relation between the treatment and the outcome

heart health

Causality: if the treatment causes the outcome to occur

Cardiac risk factors and prevention
Original article

Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok [1, 2], S Matthijs Boekholdt [3], Marleen A H Lentjes [4], Yoon K Loke [5], Robert N Luben [4], Jessica K Yeong [6], Nicholas J Wareham [7], Phyo K Myint [1], Kay-Tee Khaw [4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

Objective To examine the association between chocolate intake and the risk of future cardiovascular events.

Association and Causality

eating chocolate

Association: Any relation between the treatment and the outcome

heart health

Causality: if the treatment causes the outcome to occur

Is there an association between chocolate consumption and heart disease?
According to this study, yes ☺

Cardiac risk factors and prevention
Original article

Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok [1, 2], S Matthijs Boekholdt [3], Marleen A H Lentjes [4], Yoon K Loke [5], Robert N Luben [4], Jessica K Yeong [6], Nicholas J Wareham [7], Phyo K Myint [1], Kay-Tee Khaw [4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

Objective To examine the association between chocolate intake and the risk of future cardiovascular events.

Conclusions Cumulative evidence suggests that higher chocolate intake is associated with a lower risk of future cardiovascular events, although residual confounding cannot be excluded. There does not appear to be any evidence to say that chocolate should be avoided in those who are concerned about cardiovascular risk.

https://doi.org/10.1136/heartjnl-2014-307050

Association and Causality

eating chocolate

Association: Any relation between the treatment and the outcome

heart health

Causality: if the treatment causes the outcome to occur

Is there an association between chocolate consumption and heart disease? According to this study, yes ☺

Does eating chocolate lead to a reduction in heart disease? Well... that's harder to say...

Cardiac risk factors and prevention
Original article

Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok [1, 2], S Matthijs Boekholdt [3], Marleen A H Lentjes [4], Yoon K Loke [5], Robert N Luben [4], Jessica K Yeong [6], Nicholas J Wareham [7], Phyo K Myint [1], Kay-Tee Khaw [4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

Objective To examine the association between chocolate intake and the risk of future cardiovascular events.

Conclusions Cumulative evidence suggests that higher chocolate intake is associated with a lower risk of future cardiovascular events, although residual confounding cannot be excluded. There does not appear to be any evidence to say that chocolate should be avoided in those who are concerned about cardiovascular risk.

https://doi.org/10.1136/heartjnl-2014-307050

Confounding Factors

Definition: In observational studies, underlying difference(s) between two groups other than the treatment (that make it difficult to identify causality)

It's possible that people who like to eat chocolate do something else that offers heart protection, like eat a wide variety of healthful foods. One of the interesting things about this research is that participants in the non-chocolate group had higher average weight, more artery-damaging inflammation, more diabetes, were less physically active and had diets with the least amount of fat compared to chocolate eaters.

By Howard E. LeWine, MD, Chief Medical Editor, Harvard Health Publishing; Editorial Advisory Board Member, Harvard Health Publishing

Source: https://www.health.harvard.edu/blog/sweet-dreams-eating-chocolate-prevents-heart-disease-201506168087

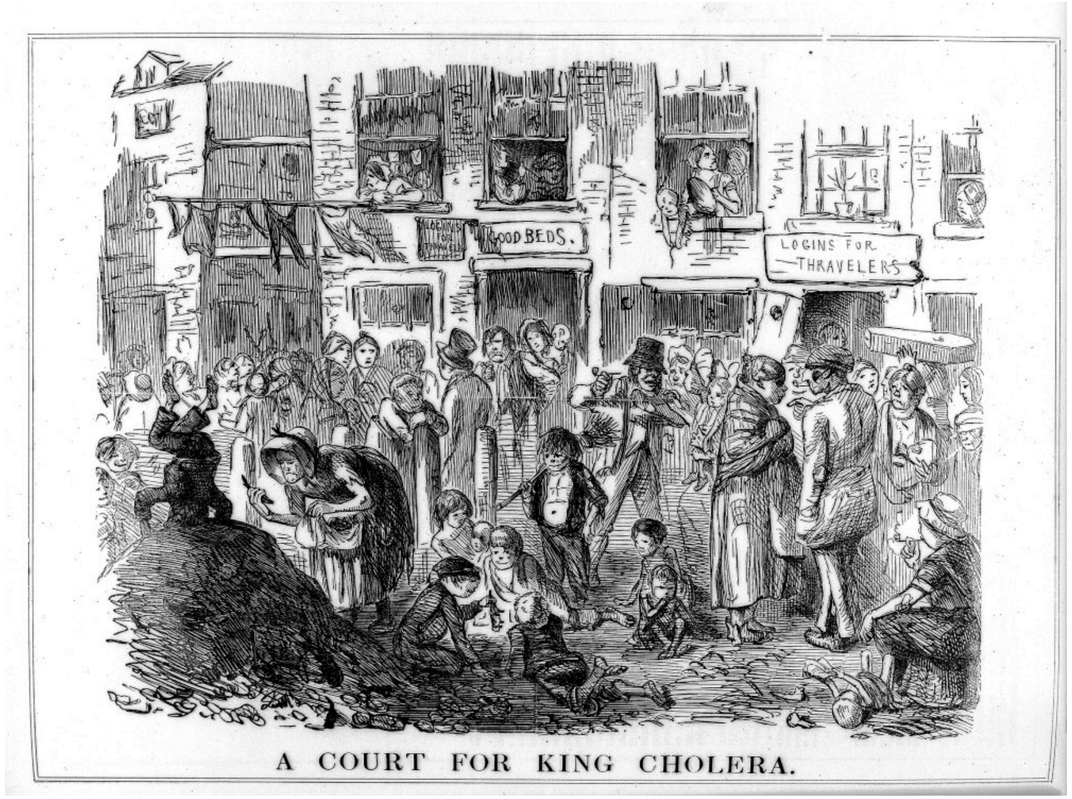
In general, observational studies are not enough to determine causation

Association and Causation

London 1850s: Cholera Epidemic

- Large migration into London in the 1700-1800s, leading to overcrowding

- What's causing cholera to spread?



A COURT FOR KING CHOLERA.

https://www.sciencemuseum.org.uk/objects-and-stories/medicine/cholera-victorian-london

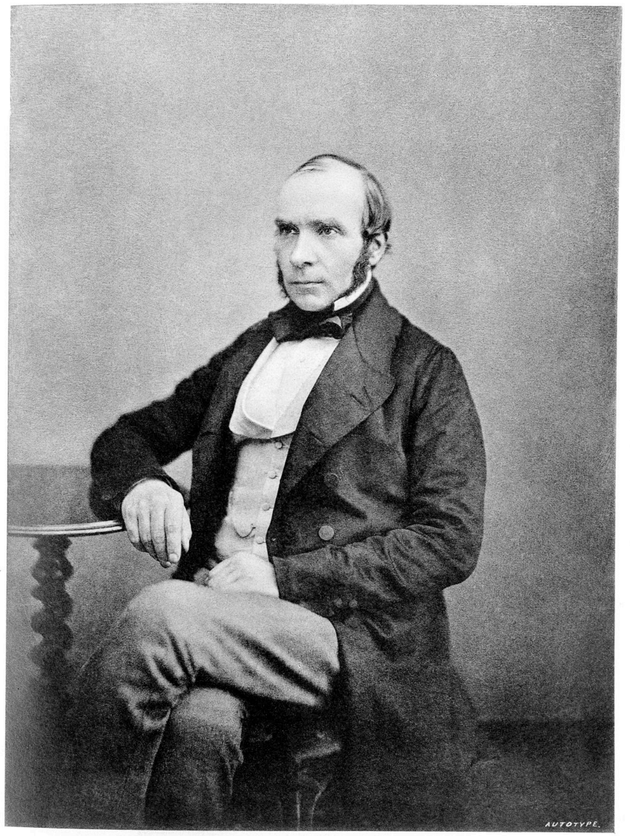
Miasma

- ‘Bad air’ given off by waste and rotting matter
- Believed to the main source of how infectious disease spread
- Potential remedies: “Fly to clean air”, “pocket full of posies”, “fire off barrels of gunpowder”
- Popular medical theory of the time
  - Florence Nightingale (founder of modern nursing)
  - Edwin Chadwick (Commissioner of the Board of Health)

John Snow (1813 - 1858)

- English physician
- Used data and visualizations to understand why cholera was spreading in the way that it was
  - Noticed people could be in close proximity without the same outcomes

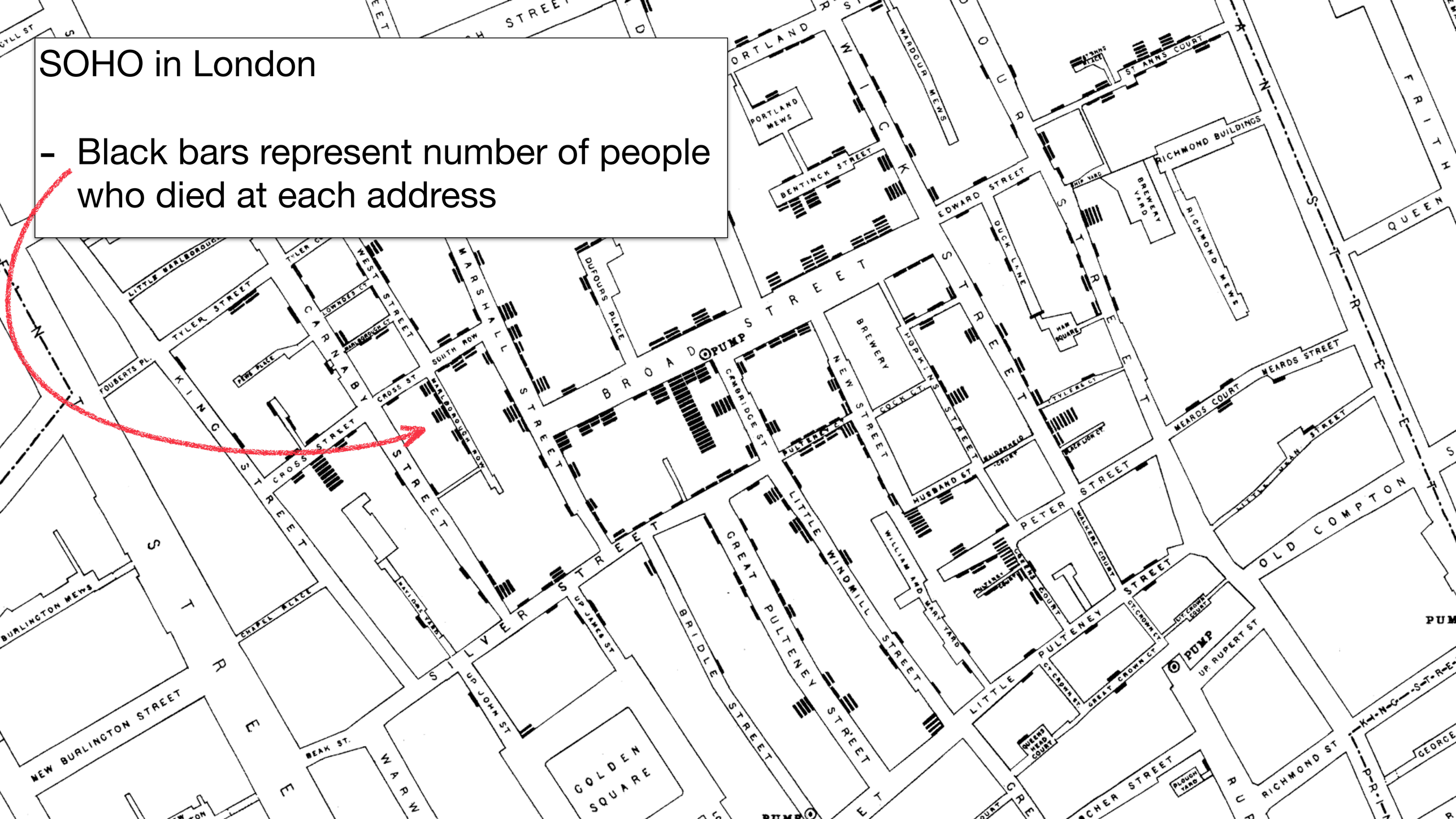
not this dude



John Snow

(Autotype from a Presentation Portrait, 1856, and Autograph facsimile.—B. W. R.)

https://wellcomecollection.org/works/dkk5vx46


SOHO in London
- Black bars represent number of people who died at each address
BROAD STREET
PUMP
MARSHALL STREET
DUFOURS PLACE
CAMBRIDGE ST
NEW STREET
BREWERY
HOPKINS STREET
CROSS STREET
KING STREET
CHAPEL PLACE
SILVER STREET
GREAT PULTENEY STREET
LITTLE PULTENEY STREET
PETER STREET
WARDOUR STREET
RICHMOND BUILDINGS
RICHMOND MEWS
EDWARD STREET
DUCK LANE
BREWERY YARD
HAM SQUARE
HEARDS COURT
HEARDS STREET
OLD COMPTON
QUEEN
PORTLAND MEWS
BENTINCK STREET
ST ANNS COURT
TYLER STREET
LITTLE MARLBOROUGH
MARLBOROUGH PLACE
MARLBOROUGH COURT
SOUTH ROW
FOUBERTS PL.
NEW BURLINGTON STREET
BURLINGTON MEWS
BEAK ST.
GOLDEN SQUARE
BRIDLE STREET
UP JAMES ST
UP JOHN ST
UP RUPERT ST
RICHMOND ST
GEORGE
PLOUGH YARD
QUEENS HEAD COURT
GT CROWN CT
LITTLE WINDMILL STREET
WILLIAM AND MARY YARD
HUSBAND ST
COCK CT
TYLERS CT
BLACK LION CT
WALKERS COURT
CROWN COURT
GT CROWN COURT
PULTENEY ST


GREAT MARLBOROUGH STREET
ARGYLL PLACE
PUMP
GREEN DRAGON YARD
LITTLE MARLBOROUGH ST
TYLER STREET
TYLER COURT
FOUBERTS PL.
KING STREET
CARNABY
CROSS STREET
WORK HOUSE
MARSHALL ST.
MARSHALL STREET
WEST STREET
LOWNDES CT
SOUTH ROW
CROSS ST
MARLBOROUGH ROW
DUFOURS PLACE
BROAD STREET
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
WARDOUR MEWS
ST ANNS COURT
RICHMOND BUILDINGS
RICHMOND MEWS
BREWERY YARD
EDWARD STREET
DUCK LANE
HAM SQUARE
QUEEN
FRITH
CAMBRIDGE ST
NEW STREET
BREWERY
HOPKINS STREET
COCK CT
HUSBAND ST
MAIDENHEAD COURT
TYLER CT
BLACK LION CT
PETER STREET
MEARDS COURT
MEARDS STREET
OLD COMPTON
WALKERS COURT
LITTLE PULTENEY STREET
GREAT CROWN CT
UP. RUPERT ST
QUEENS HEAD COURT
ARCHER STREET
PLOUGH YARD
RICHMOND ST
K-I-N-G-S-T-R-E-E-T
GEORGE
SILVER STREET
UP JAMES ST
UP JOHN ST
GOLDEN SQUARE
BRIDLE STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
WILLIAM AND MARY YARD
BURLINGTON MEWS
NEW BURLINGTON STREET
CHAPEL PLACE
BEAK ST.
STREET


GREAT MARLBOROUGH STREET
ARGYLL PLACE
PUMP
WORK HOUSE
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
WARDOUR MEWS
ST ANNS COURT
RICHMOND BUILDINGS
RICHMOND MEWS
BREWERY YARD
EDWARD STREET
DUCK LANE
QUEEN
FRITH
GREEN DRAGON YARD
LITTLE MARLBOROUGH ST
TYLER COURT
TYLER STREET
MARSHALL ST.
MARSHALL STREET
DUFOURS PLACE
BROAD STREET
PUMP
CARNABY
WEST STREET
LOWNDES CT
SOUTH ROW
CROSS ST
MARLBOROUGH ROW
FOUBERTS PL.
KING STREET
CROSS STREET
CAMBRIDGE ST
NEW STREET
BREWERY
HOPKINS STREET
COCK CT
HUSBAND ST
HAM YARD
TYLERS CT
BLACK LION CT
MEARDS STREET
MEARDS COURT
PETER STREET
OLD COMPTON
SILVER STREET
LITTLE WINDMILL STREET
WILLIAM AND MARY YARD
GREAT PULTENEY STREET
BRIDLE STREET
UP JAMES ST
UP JOHN ST
GOLDEN SQUARE
BEAK ST.
NEW BURLINGTON STREET
BURLINGTON MEWS
CHAPEL PLACE
LITTLE PULTENEY STREET
WALKERS COURT
QUEENS HEAD COURT
GREAT CROWN CT
PUMP
UP. RUPERT ST
RICHMOND ST
ARCHER STREET
PLOUGH YARD
KING STREET
GEORGE
PUMP

BROAD STREET
PUMP
SOUTH ROW
MARLBOROUGH ROW
CAMBRIDGE ST
NEW STREET
BREWERY
HOPKINS STREET
COCK CT.
PULTENEY CT.
HUSBAND ST
LITTLE WINDMILL ST
GREAT PULT
WILLIAM AND MARY YARD
STREET

Water Pumps


PUMP
PUMP
PUMP
PUMP
GREAT MARLBOROUGH STREET
MARLBOROUGH MEWS
ARGYLL PLACE
ARGYLL ST
ARGYLL STREET
BLENHEIM MEWS
POLAND STREET
NOEL STREET
BERWICK STREET
WARDOUR STREET
LITTLE CHAPEL
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
WARDOUR MEWS
WORK HOUSE
ST ANNS COURT
RICHMOND BUILDINGS
RICHMOND MEWS
BREWERY YARD
EDWARD STREET
DUCK LANE
QUEEN
DUFOURS PLACE
BROAD STREET
CAMBRIDGE ST
MARSHALL ST.
MARSHALL STREET
LITTLE MARLBOROUGH ST
TYLER STREET
TYLER COURT
CARNABY STREET
WEST STREET
LOWNDES CT
SOUTH ROW
KING STREET
FOUBERTS PL
CROSS STREET
CROSS ST
MARLBOROUGH ROW
SILVER STREET
NEW STREET
HOPKINS STREET
COCK CT
BREWERY
HUSBAND ST
PULTENEY CT
LITTLE WINDMILL STREET
GREAT PULTENEY
BRIDLE
WILLIAM AND MARY YARD
PETER STREET
WALKERS COURT
MEARDS COURT
MEARDS STREET
OLD COMPTON
PULTENEY
UP RUPERT ST
OLD BURLINGTON MEWS
CHAPEL PLACE
UP JAMES ST
HAM SQUARE
TYLERS CT
BLACK LION CT
GREEN DRAGON YARD


Water Pumps
PUMP
PUMP
PUMP
PUMP
GREAT MARLBOROUGH STREET
ARGYLL PLACE
WORK HOUSE
BROAD STREET
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
WARDOUR MEWS
EDWARD STREET
RICHMOND BUILDINGS
RICHMOND MEWS
MARSHALL ST.
MARSHALL STREET
DUFOURS PLACE
CARNABY STREET
KING STREET
CROSS STREET
SILVER STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
NEW STREET
BREWERY
PETER STREET
OLD COMPTON
MEARDS STREET
MEARDS COURT
LITTLE CHAPEL
NOEL
BLENHEIM MEWS
ST ANNS COURT
TYLER STREET
SOUTH ROW
CHAPEL PLACE
OLD BURLINGTON MEWS
BRIDLE
PULTENEY

Water Pumps

PUMP
PUMP
PUMP
PUMP
GREAT MARLBOROUGH STREET
ARGYLL PLACE
ARGYLL STREET
WORK HOUSE
BROAD STREET
BREWERY
NEW ST
POLAND STREET
BERWICK STREET
WARDOUR STREET
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
NOEL STREET
EDWARD STREET
MARSHALL STREET
MARSHALL ST.
CARNABY STREET
KING STREET
CROSS STREET
SILVER STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
PETER STREET
OLD COMPTON
MEARDS STREET
MEARDS COURT
RICHMOND BUILDINGS
RICHMOND MEWS
ST ANNS COURT
LITTLE CHAPEL
DUFOURS PLACE
CAMBRIDGE ST
BRIDLE
HUSBAND ST
SOUTH ROW
WEST STREET
TYLER STREET
LITTLE MARLBOROUGH ST
BLENHEIM STREET
BLENHEIM MEWS
DUCK LANE
HAM YARD
WILLIAM AND MARY YARD
PULTENEY
QUEEN

Broad Street Pump

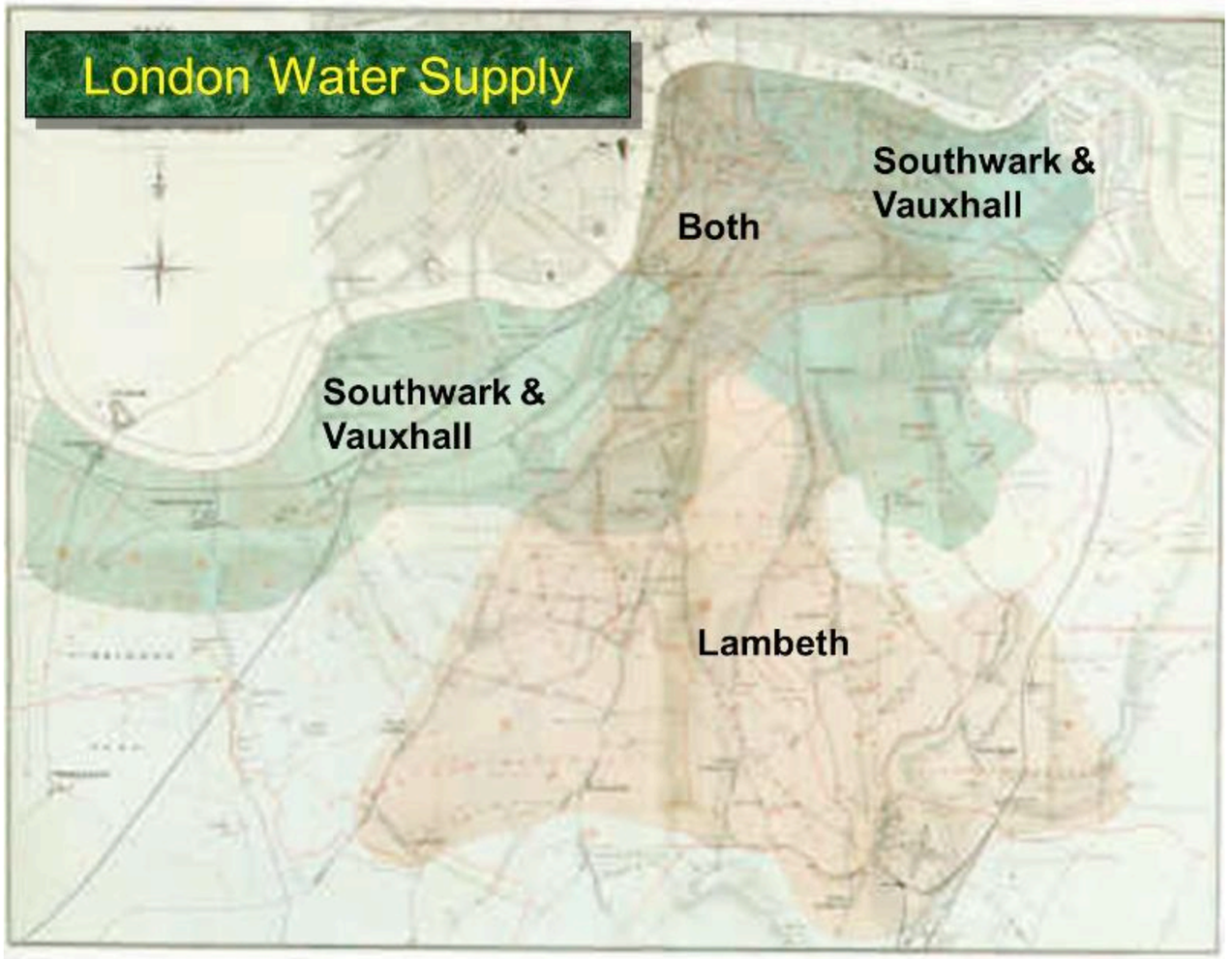
- Snow got pump handle removed
  - It was later discovered a cesspit had been leaking into the well
- Observational study
  - Snow did not control the pump or the people
  - Association was strong

Showing Causation

Snow investigated cases in an area served by two water companies

- Southwark and Vauxhall's intake was in heavily polluted area of the Thames

- Lambeth's water intake was in a less polluted area upstream



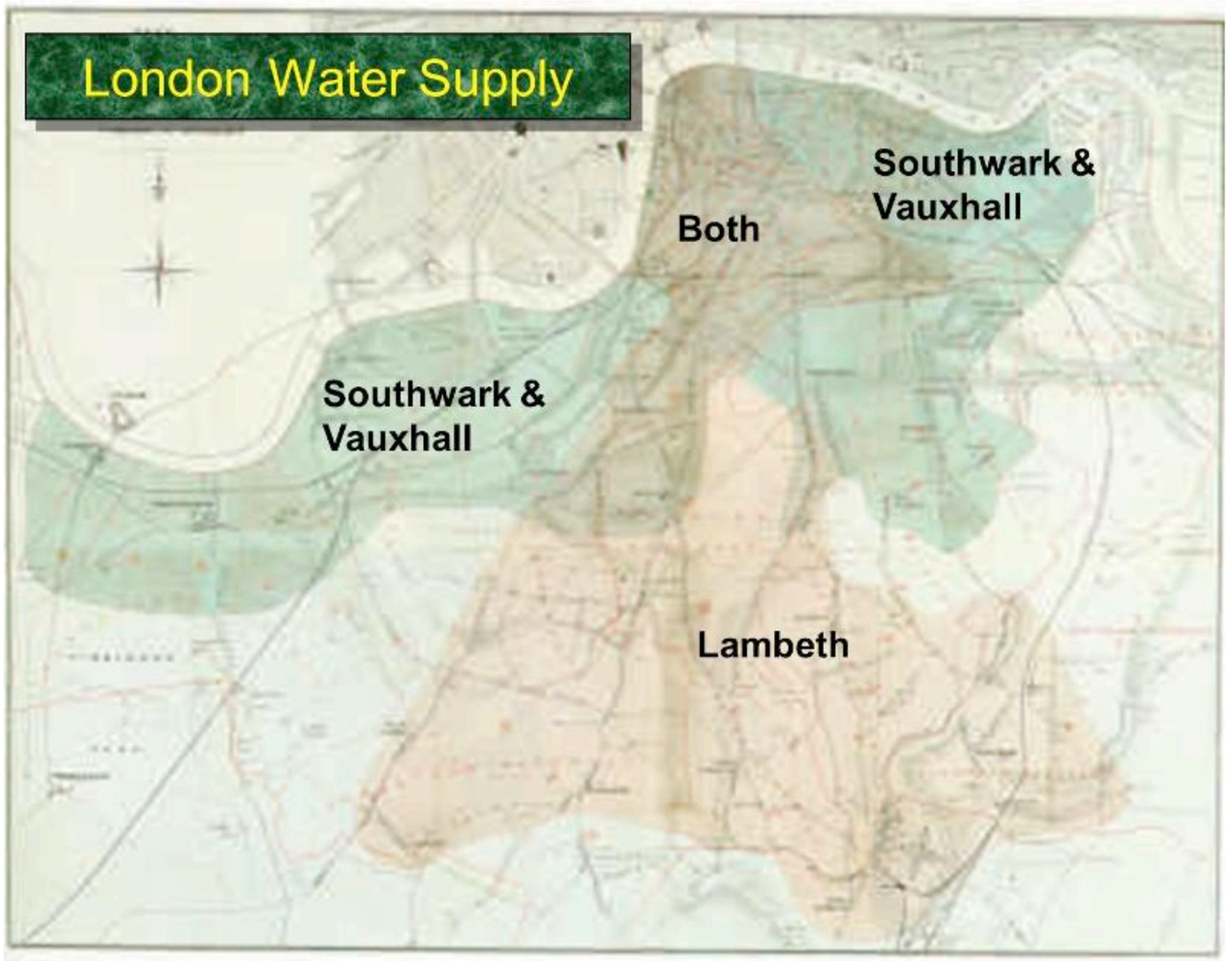
Snow's "Grand Experiment"

Treatment group: S&V

Control group: Lambeth

"… 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…"

Two groups are similar except for the treatment



Snow’s “Grand Experiment”

Deaths From Cholera Epidemic in Districts of London Supplied by Two Water Companies Over 7 Weeks, 1854

Water Supply Company	Number of Houses	Deaths From Cholera	Cholera Deaths per 10,000 Houses
Southwark and Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

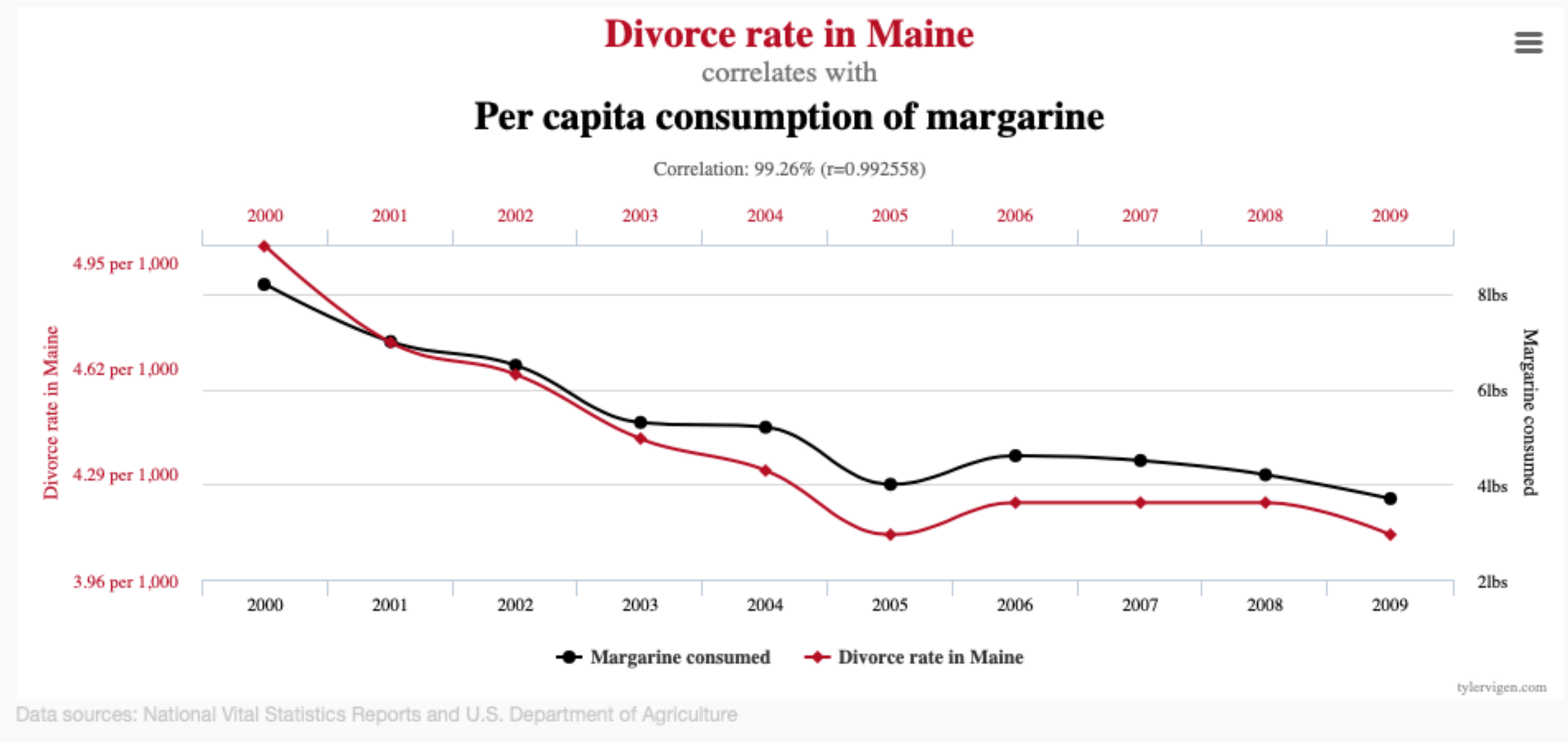
Key to Establishing Causality

- Treatment group: Receives the treatment

- Control group: Does not receive the treatment

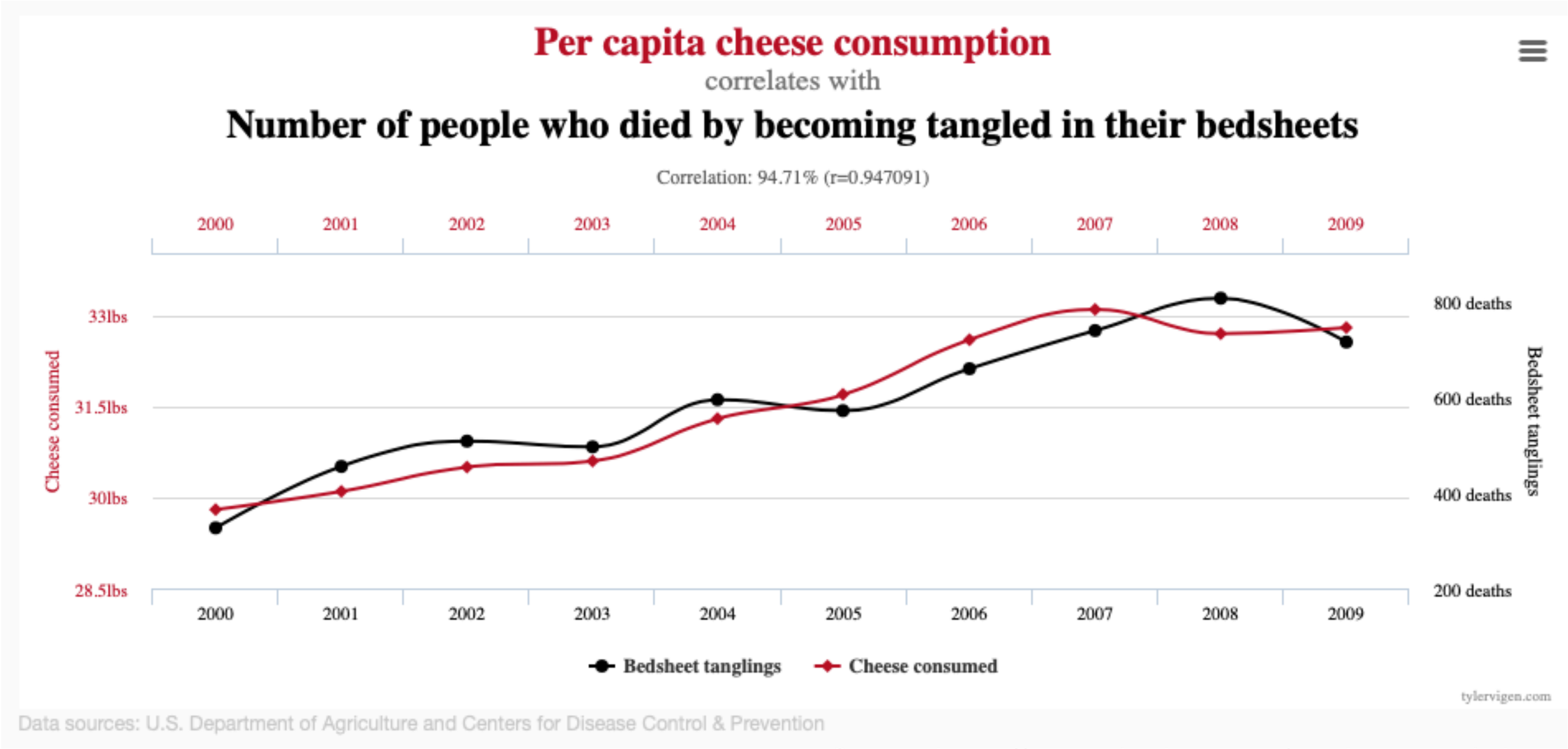
If the treatment and control groups are *similar apart from the treatment*, then differences between the outcomes in the two groups can be ascribed to the treatment

Confounding Factors / Spurious Correlations



Source: https://www.tylervigen.com/spurious-correlations

Confounding Factors / Spurious Correlations



Source: https://www.tylervigen.com/spurious-correlations

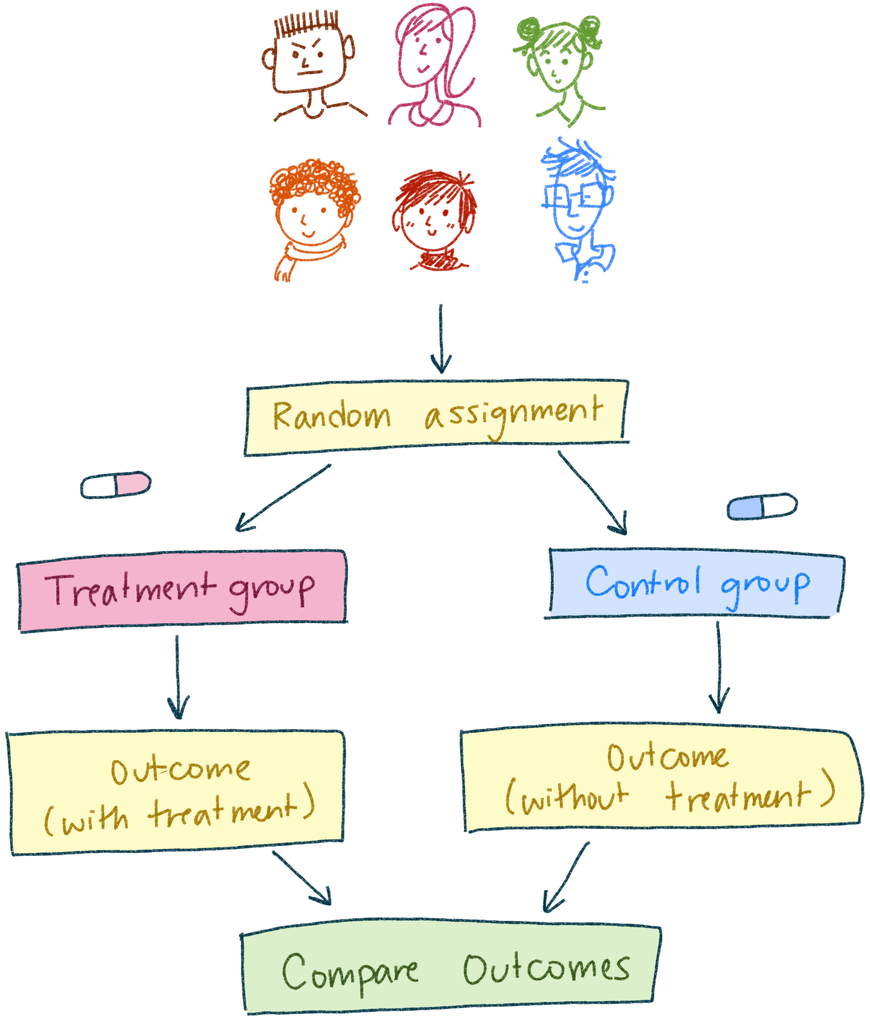
How to determine causation vs random correlation?

Randomization!

- If you assign individuals to treatment and control at random, then the two groups are likely to be similar apart from the treatment.

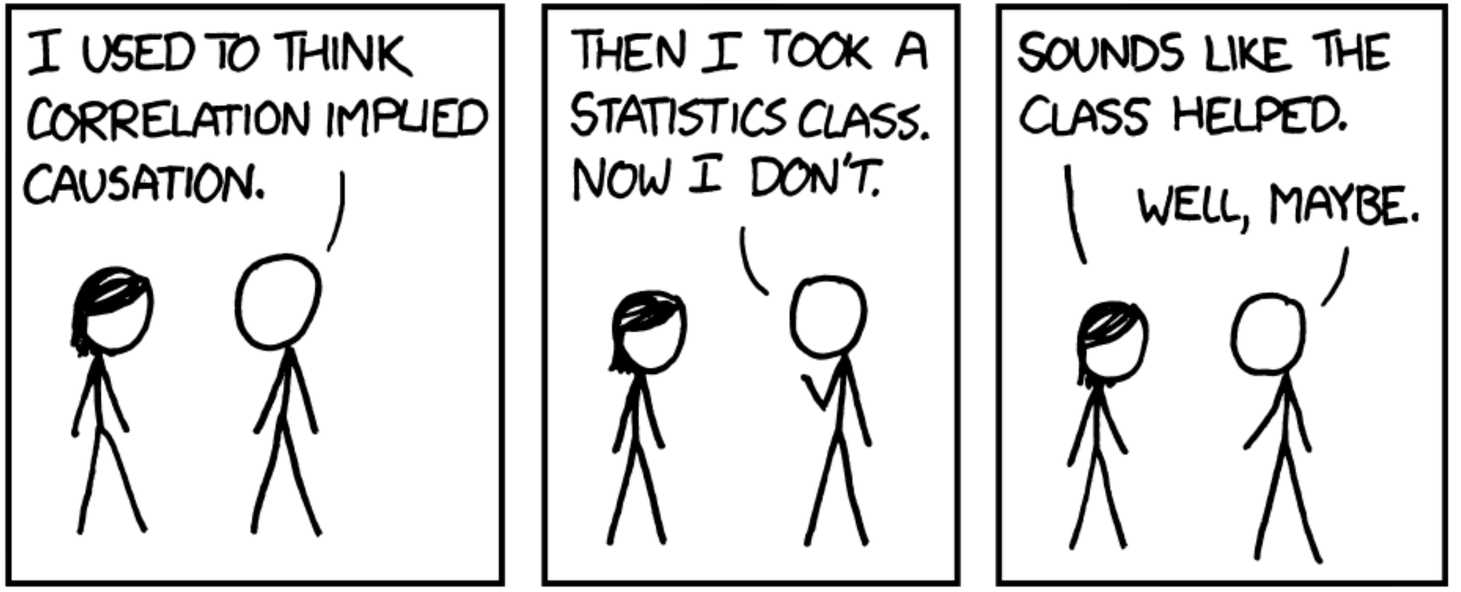
- Known as a Randomized Controlled Experiment or Randomized Controlled Trial (RCT)

- You can account (mathematically) for variability in the assignment



Correlation

|< < Prev Random Next > >|



|< < Prev Random Next > >|

https://xkcd.com/552/

Next Class

- Intro to Python: Expressions and Data Types
- Remember:
  - Labs start this week!!
  - Office hours start next week!