DATA 8
Summer 2018

Lecture 2

Cause and Effect

Contributions by Vinitra Swamy (vinitra@berkeley.edu) and Fahad Kamran (fhdkmrn@berkeley.edu)
Slides created by John DeNero (denero@berkeley.edu) and Ani Adhikari (adhikari@berkeley.edu)

Context: Last Lecture

- **Introduction**
  - Data science: drawing useful conclusions from data using computation
    - Exploration
    - Inference
    - Prediction
  - Course Policies
  - Demo

Cause and Effect

Announcements

Questions

A Link

Coffee

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

34 6

Staff and agencies

Wednesday 22 November 2017 19.54 EST

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

Guardian UK

A Stronger Link?

eating and health

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

JUNE 19, 2015 5:03 AM ET


ALLISON AUBREY

npr.org (report on a study in heart.bmj.com)

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

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 Harvard Health Blog, reported by [npr.org](npr.org)*

- Yes, this points to an association
  (in my opinion)

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.”

- *JoAnn Manson, chief of Preventive Medicine at Brigham and Women’s Hospital, Boston*

Association

London, early 1850's

A COURT FOR KING CHOLERA.

Illustration from *Punch* (1852).

Miasmas, miasmatism, miasmatists

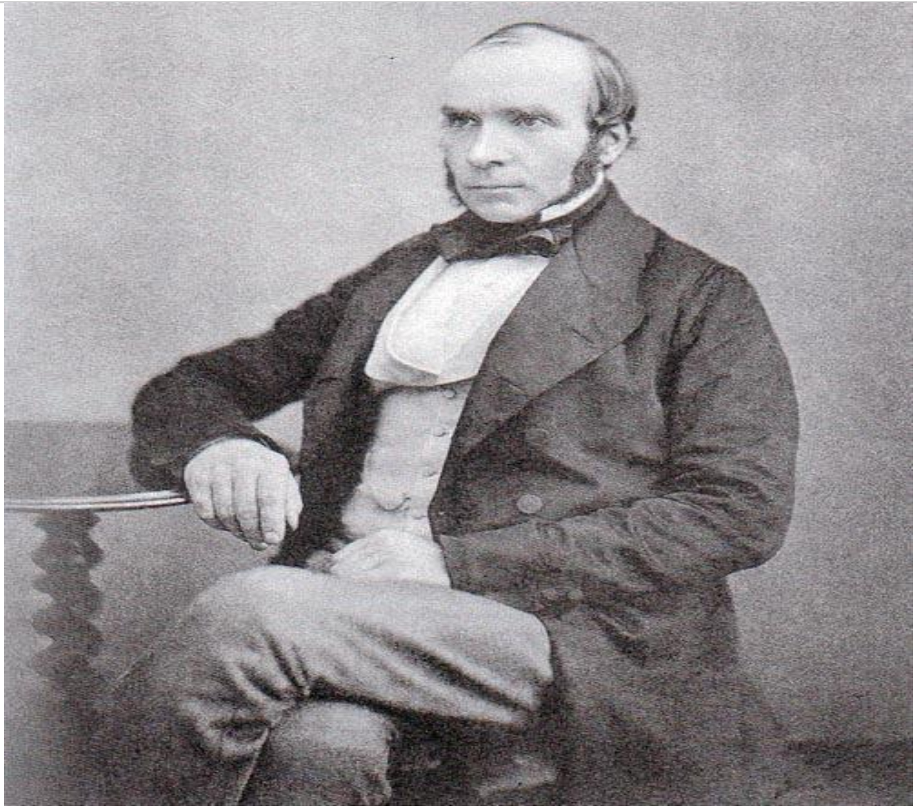
- **Bad smells** given off by waste and rotting matter
- **Believed to be the main source of disease**
- Suggested remedies:
  - "fly to clene air"
  - "a pocket full o'posies"
  - "fire off barrels of gunpowder"

Celebrity Miasmatists

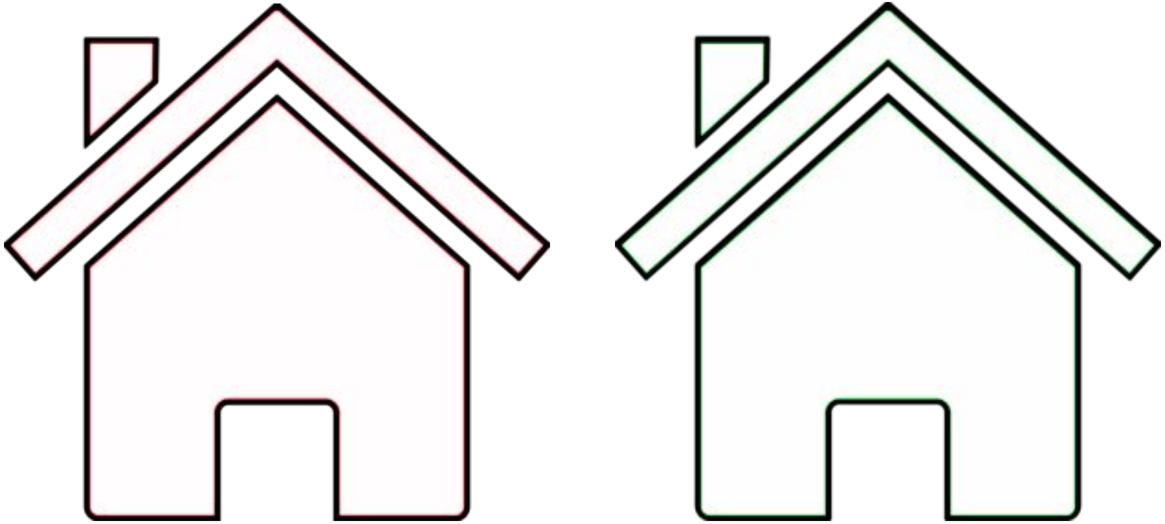
- Florence Nightingale
  - “Lady with the Lamp”
- Edwin Chadwick
  - Commissioner of the General Board of Health
- There was one person who was a little doubtful...

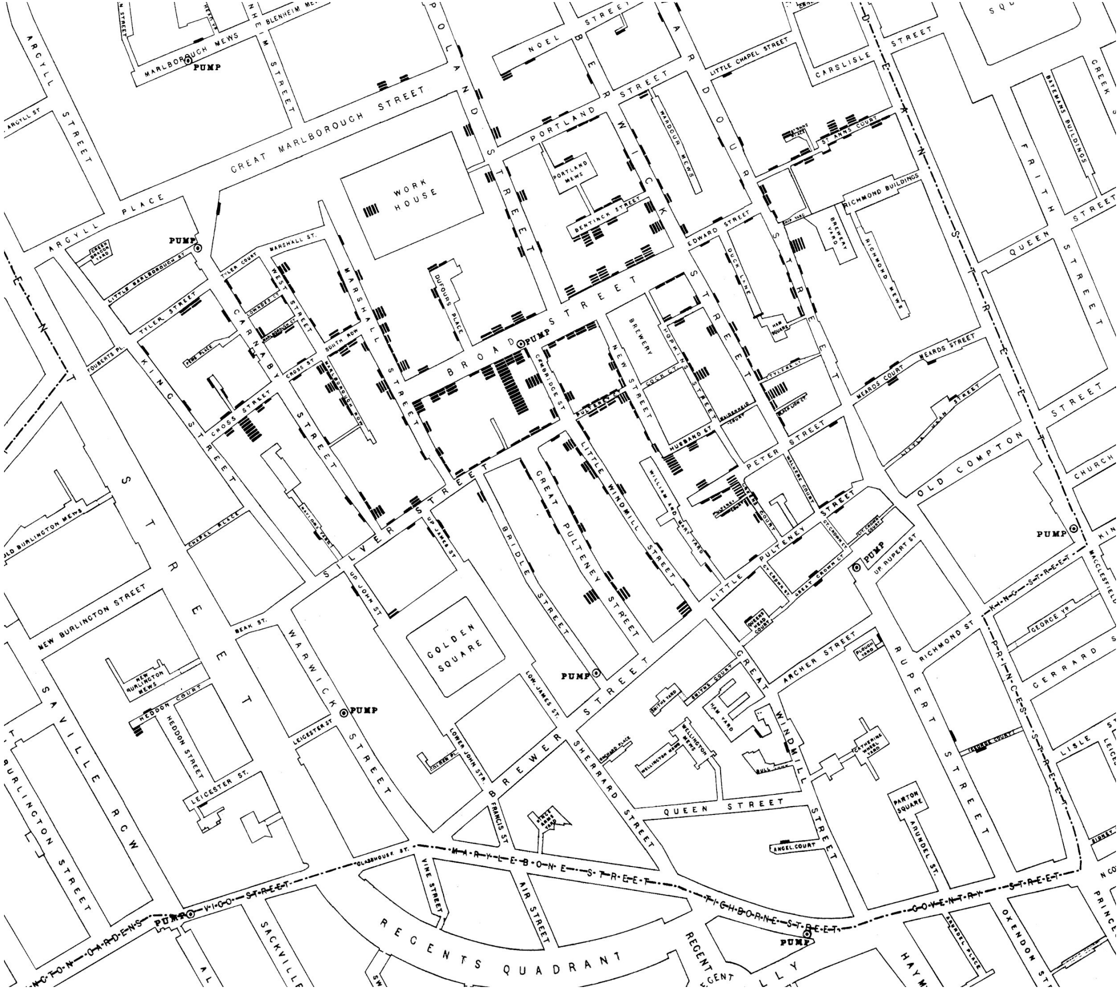
Jon Snow, 281-302

John Snow, 1813-1858

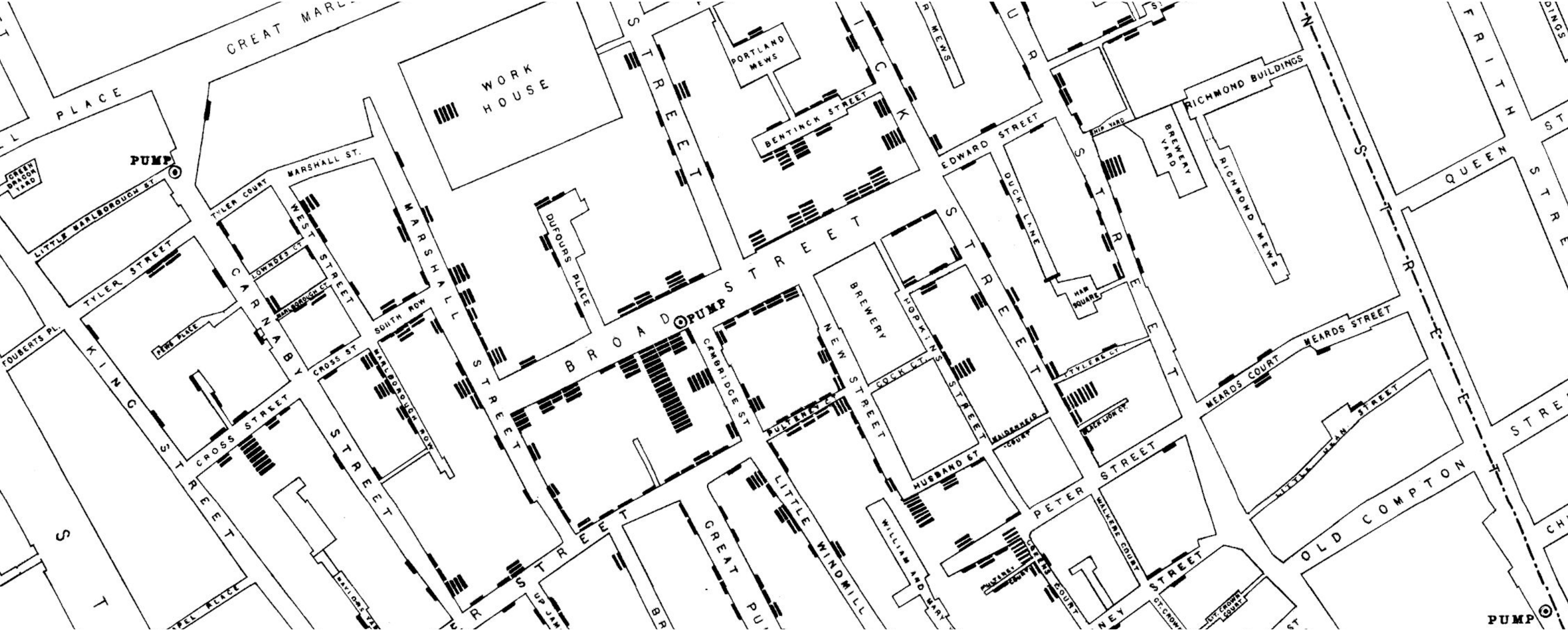


John Snow's Doubt





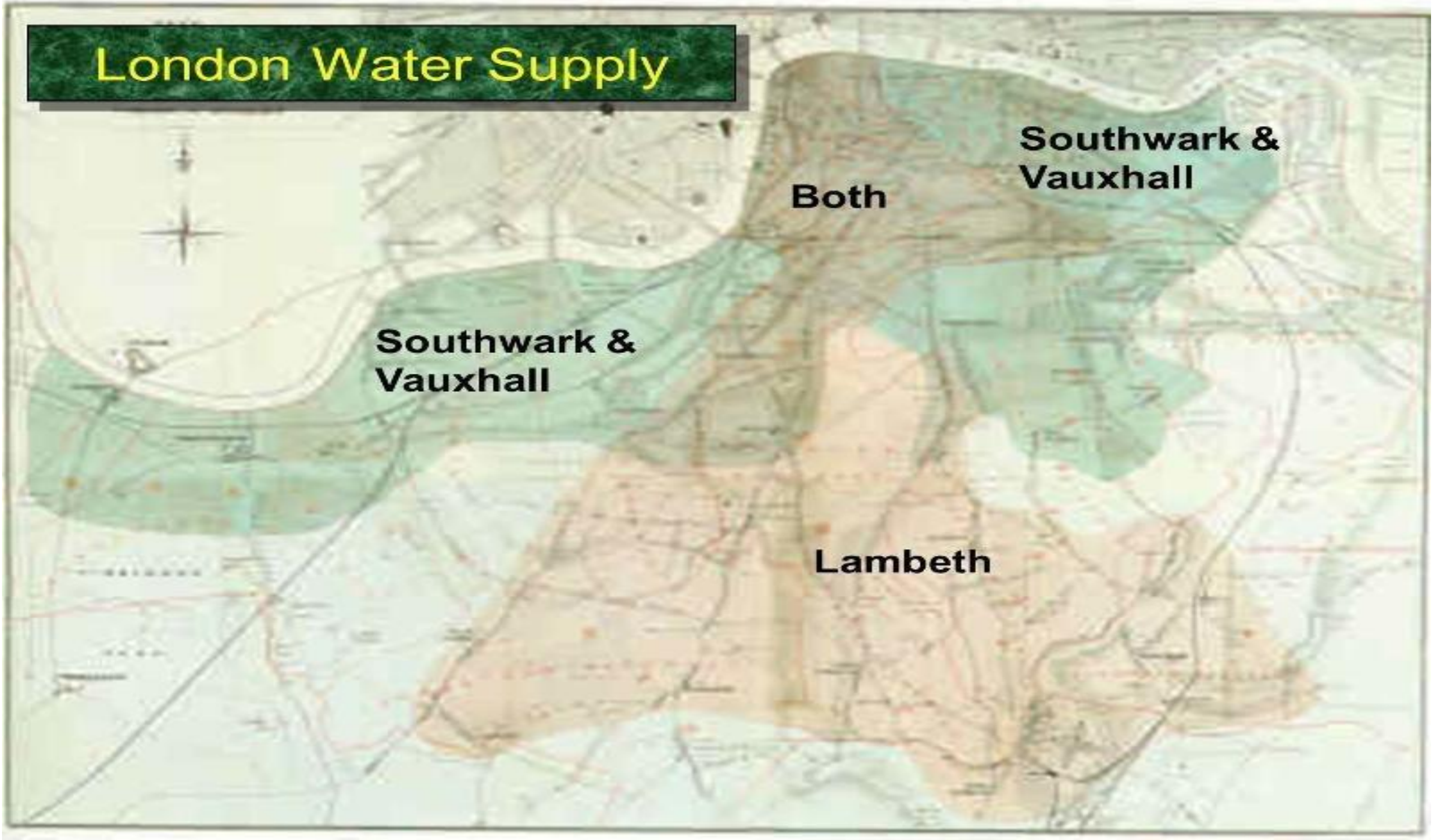
GREAT MARLBOROUGH STREET
MARLBOROUGH MEWS
PUMP
ARGYLL STREET
ARGYLL PLACE
WORK HOUSE
POLAND STREET
NOEL STREET
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
BERWICK STREET
WARDOUR STREET
LITTLE CHAPEL STREET
CARLISLE STREET
ST ANNS COURT
RICHMOND BUILDINGS
RICHMOND MEWS
FRITH STREET
QUEEN STREET
DEAN STREET
EDWARD STREET
DUFOURS PLACE
BROAD STREET
MARSHALL STREET
MARSHALL ST.
TYLER STREET
LITTLE MARLBOROUGH ST.
CARNABY STREET
KING STREET
CROSS STREET
SILVER STREET
CAMBRIDGE ST.
NEW STREET
BREWERY
HOPKINS STREET
HUSBAND ST.
PETER STREET
MEARDS STREET
MEARDS COURT
OLD COMPTON STREET
CHURCH STREET
LITTLE WINDMILL STREET
WILLIAM AND MARY YARD
GREAT PULTENEY STREET
LITTLE PULTENEY STREET
BRIDLE STREET
REGENT STREET
NEW BURLINGTON STREET
NEW BURLINGTON MEWS
OLD BURLINGTON MEWS
HEDDON COURT
HEDDON STREET
LEICESTER ST.
WARWICK STREET
GOLDEN SQUARE
UP. JOHN ST.
LOWER JOHN STR.
BREWER STREET
SHERRARD STREET
GREAT WINDMILL STREET
ARCHER STREET
RUPERT STREET
UP. RUPERT ST.
RICHMOND ST.
GEORGE YD
GERRARD STREET
PRINCES STREET
KING STREET
QUEEN STREET
PARTON SQUARE
ARUNDEL ST.
ANGEL COURT
SAVILLE ROW
BURLINGTON STREET
VIGO STREET
MARYLEBONE STREET
TICHBORNE STREET
COVENTRY STREET
VINE STREET
AIR STREET
REGENTS QUADRANT
SACKVILLE
OXENDON ST
LISLE STREET



PLACE
PUMP
GREEN DRAGON YARD
LITTLE MARLBOROUGH ST
TYLER STREET
FOUBERTS PL.
KING STREET
CARNABY
TYLER COURT
MARSHALL ST.
WEST STREET
LOWNDES CT
CROSS ST
CROSS STREET
SOUTH ROW
MARSHALL STREET
MARLBOROUGH ROW
WORK HOUSE
DUFOURS PLACE
BROAD STREET
PUMP
CAMBRIDGE ST
PORTLAND MEWS
BENTINCK STREET
EDWARD STREET
BREWERY
NEW STREET
HOPKINS STREET
COCK CT.
HUSBAND ST
LITTLE WINDMILL
WILLIAM AND MARY
GREAT
DUCK LANE
TYLERS CT
BLACK LION CT.
PETER STREET
WALKERS COURT
RICHMOND BUILDINGS
RICHMOND MEWS
BREWERY YARD
SHIP YARD
QUEEN
FRITH
MEARDS COURT
MEARDS STREET
OLD COMPTON
PUMP


JOHN SNOW

Causation


London Water Supply
Southwark & Vauxhall
Both
Southwark & Vauxhall
Lambeth

Comparison

- **treatment group**
- **control group**
  - does not receive the treatment

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

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

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 ≠ Haphazard

Jupyter Notebooks

Website