DATA 8

Fall 2016

Lecture 2, August 26

Causality

Slides created by Ani Adhikari and John DeNero

Announcements

- Please join Piazza (piazza.com) for Data 8 Foundations of Data Science [Fall 2016]
- Lab 1 is due at (that is, at or before) 7 pm today.
- Homework 1 is due at 5 p.m. Thursday 9/1. You get a bonus point for turning it in by 5 p.m. Wednesday 8/31.
- GSI and tutor office hours will be announced soon. Watch Piazza and data8.org.

Really?

eating and health

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

JUNE 19, 2015 5:03 AM ET

ALLISON AUBREY

[npr.org](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”

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*

- 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.

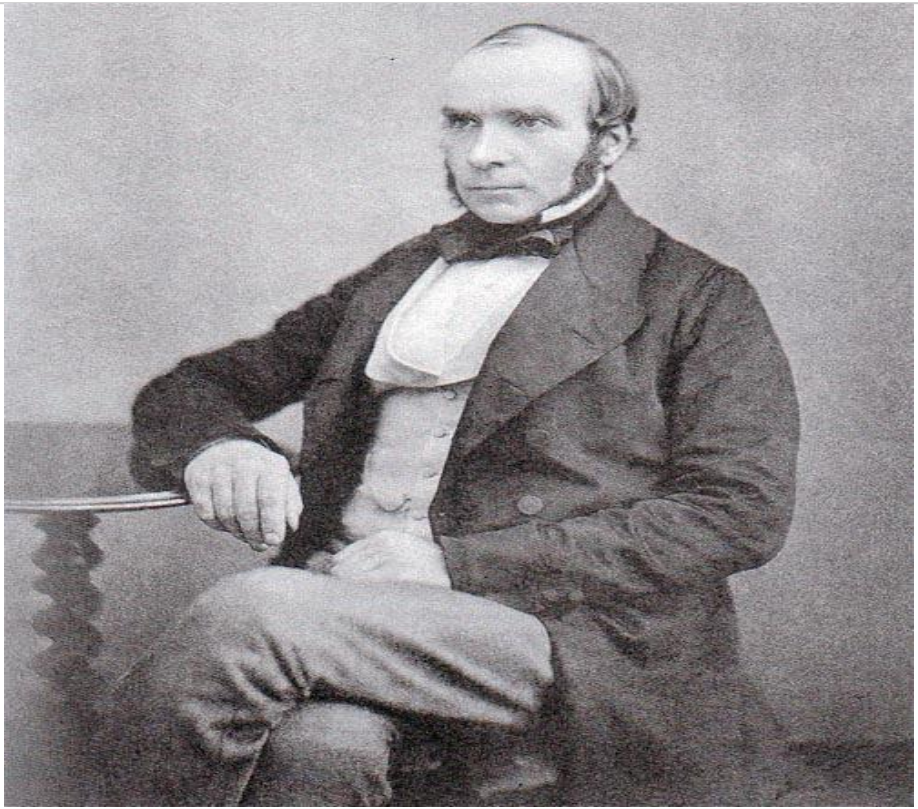
“[The study] doesn’t prove a cause-and-effect relationship between chocolate and reduced risk of heart disease and stroke.”

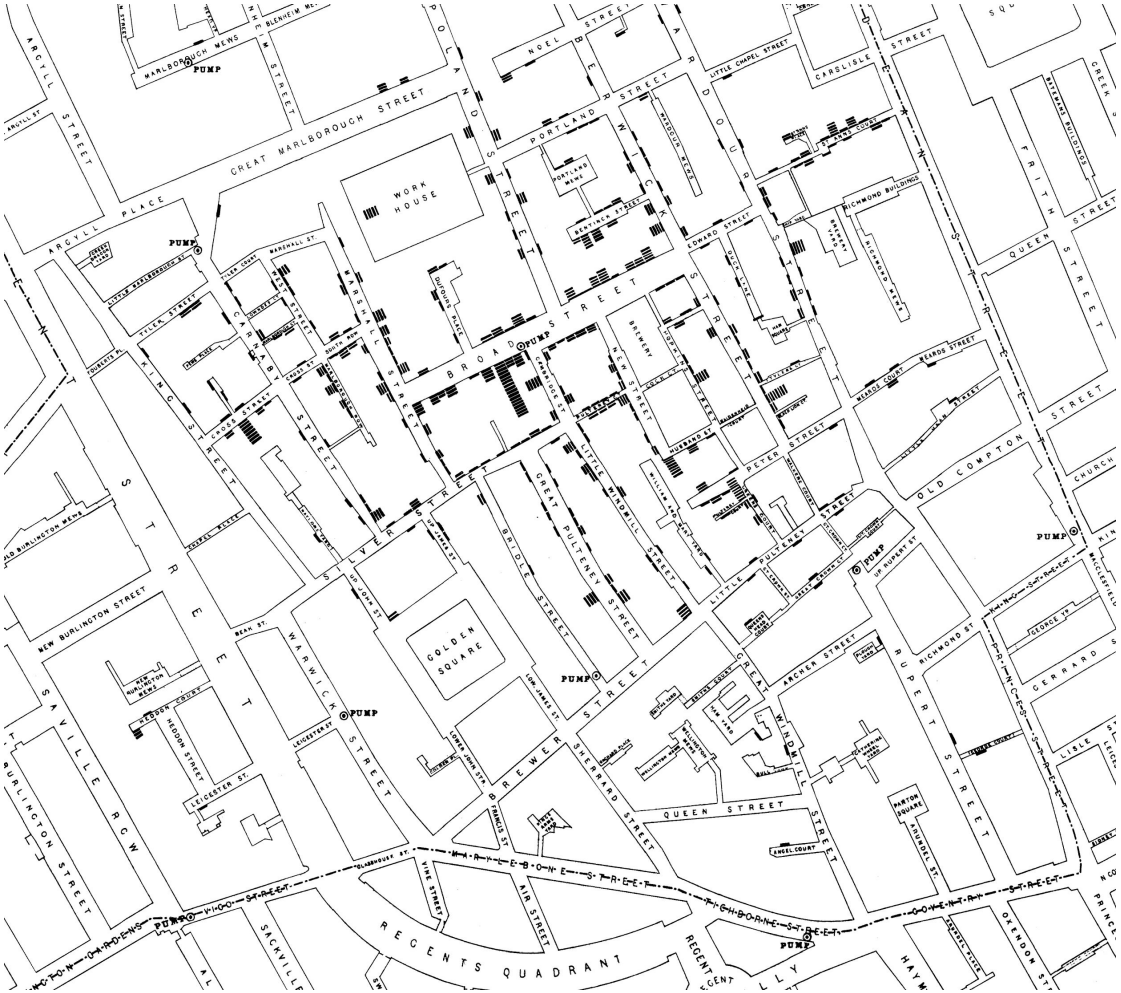
- *JoAnn Manson, chief of Preventive Medicine at Brigham and Women’s Hospital, Boston*

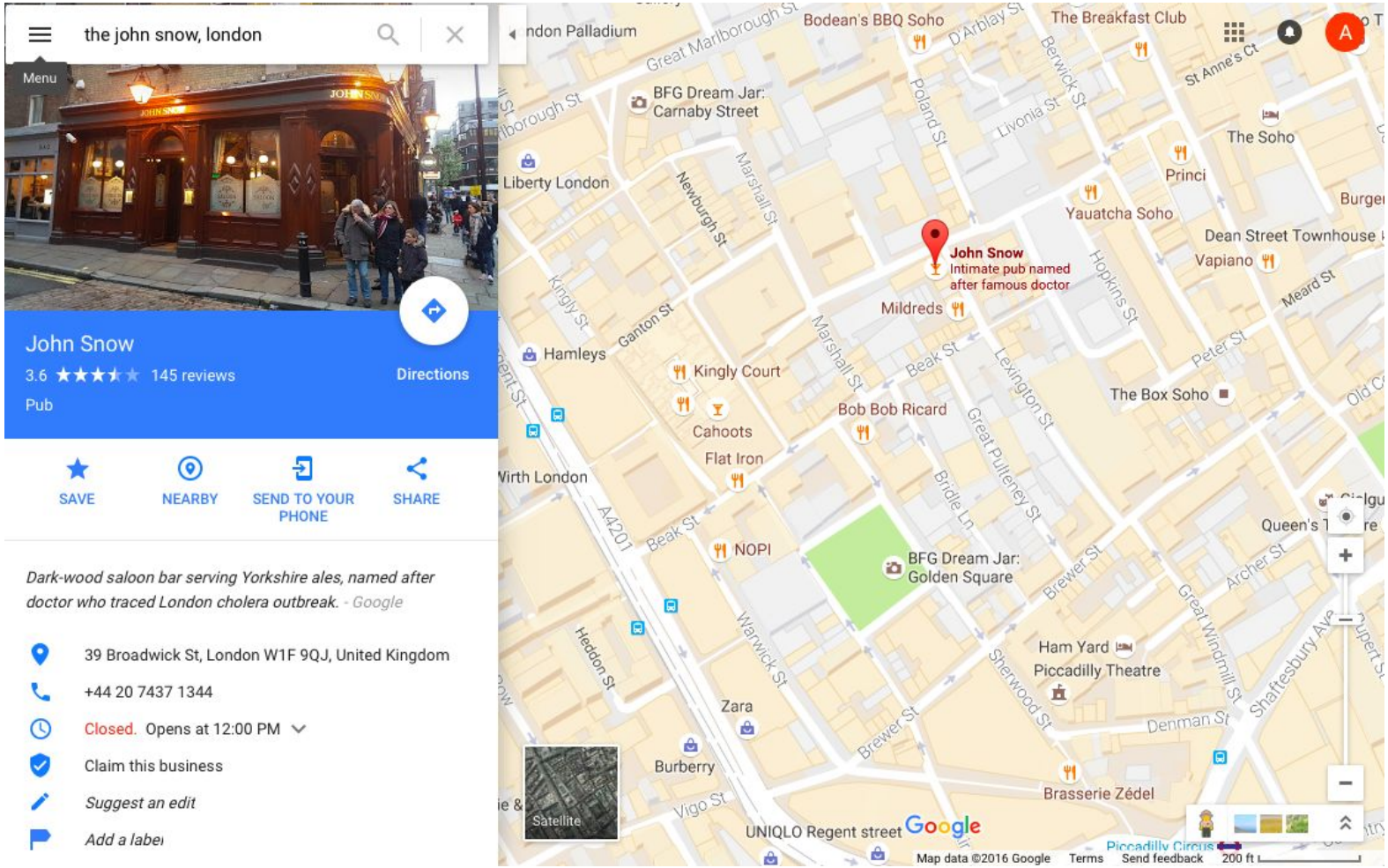
Miasmas, miasmatism, miasmatists

- **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

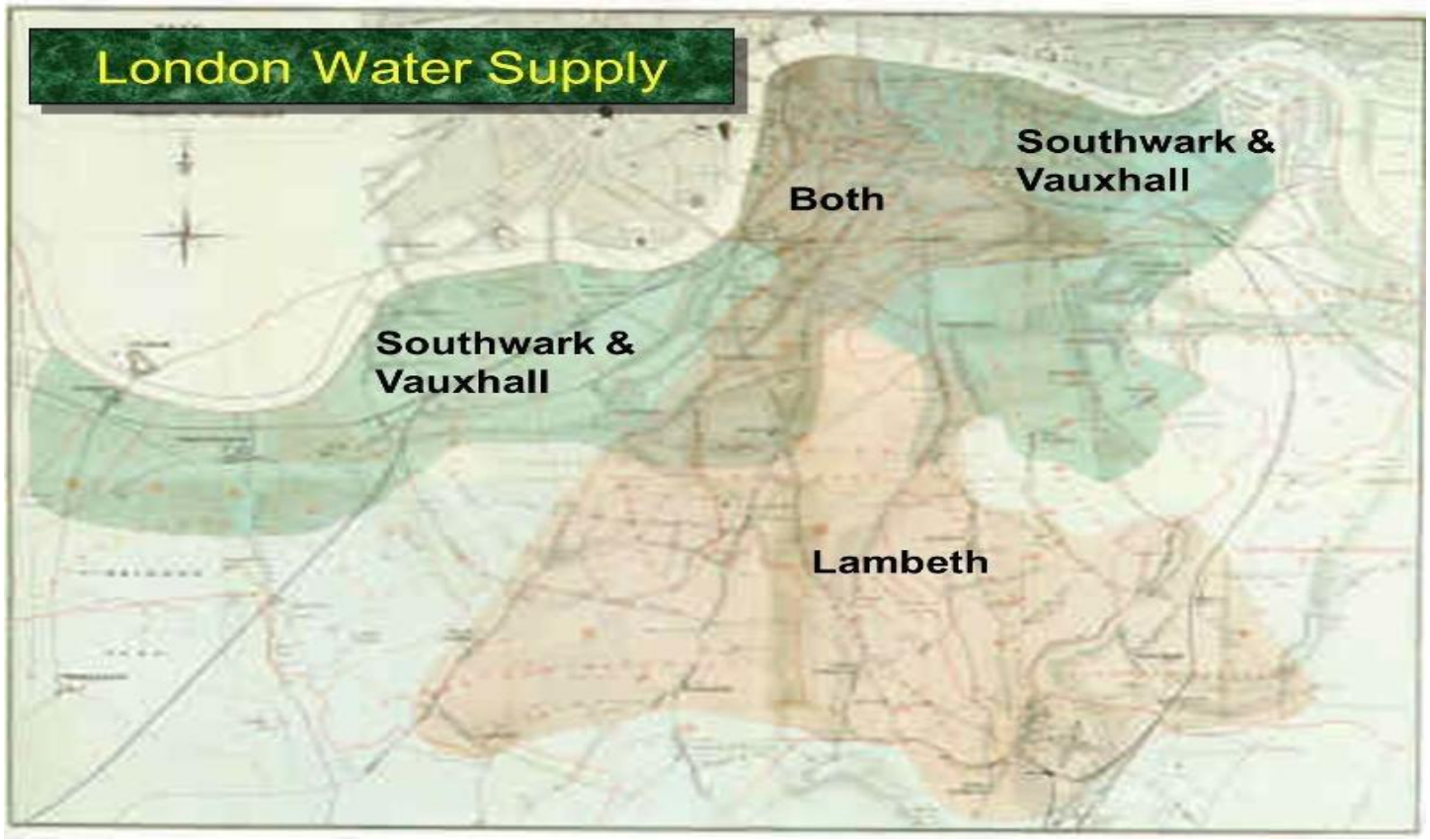
John Snow, 1813-1858




GREAT MARLBOROUGH STREET
ARGYLL PLACE
ARGYLL STREET
MARLBOROUGH MEWS
PUMP
WORK HOUSE
POLAND STREET
PORTLAND MEWS
PORTLAND STREET
NOEL STREET
LITTLE CHAPEL STREET
CARLISLE STREET
RICHMOND BUILDINGS
RICHMOND MEWS
LOWNDES STREET
BROAD STREET
MARSHALL ST.
CHAPLIN STREET
SILVER STREET
KINGLY STREET
BEAK STREET
GOLDEN SQUARE
GREAT PULTENEY STREET
BRIDLE STREET
LITTLE PULTENEY STREET
PETER STREET
OLD COMPTON STREET
FRITH STREET
QUEEN STREET
NEW BURLINGTON STREET
NEW BURLINGTON MEWS
SAVILLE ROW
BURLINGTON STREET
HEDDON COURT
HEDDON STREET
LEICESTER ST.
WARWICK STREET
LOWER JOHN ST.
BREWER STREET
SHERRARD STREET
ARCHER STREET
WINDMILL STREET
RUPERT STREET
RICHMOND ST.
GERRARD ST.
PARTON STREET
ANGEL COURT
WARDOUR STREET
REGENTS QUADRANT
AIR STREET
PIPE STREET
PICCADILLY
REGENT
HAYMARKET


the john snow, london
Menu
John Snow
3.6 145 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. - Google
39 Broadwick St, London W1F 9QJ, United Kingdom
+44 20 7437 1344
Closed. Opens at 12:00 PM
Claim this business
Suggest an edit
Add a label
John Snow
Intimate pub named after famous doctor
Bodean's BBQ Soho
The Breakfast Club
BFG Dream Jar: Carnaby Street
Liberty London
The Soho
Princi
Yauatcha Soho
Dean Street Townhouse
Vapiano
Mildreds
Hamleys
Kingly Court
Cahoots
Flat Iron
Bob Bob Ricard
The Box Soho
NOPI
BFG Dream Jar: Golden Square
Ham Yard
Piccadilly Theatre
Zara
Burberry
Brasserie Zédel
UNIQLO Regent street
Satellite
Map data ©2016 Google
Terms
Send feedback
200 ft


JOHN SNOW
JOHN SNOW


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.

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