



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 whenever you can get it done.
 - 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 (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](https://www.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.”

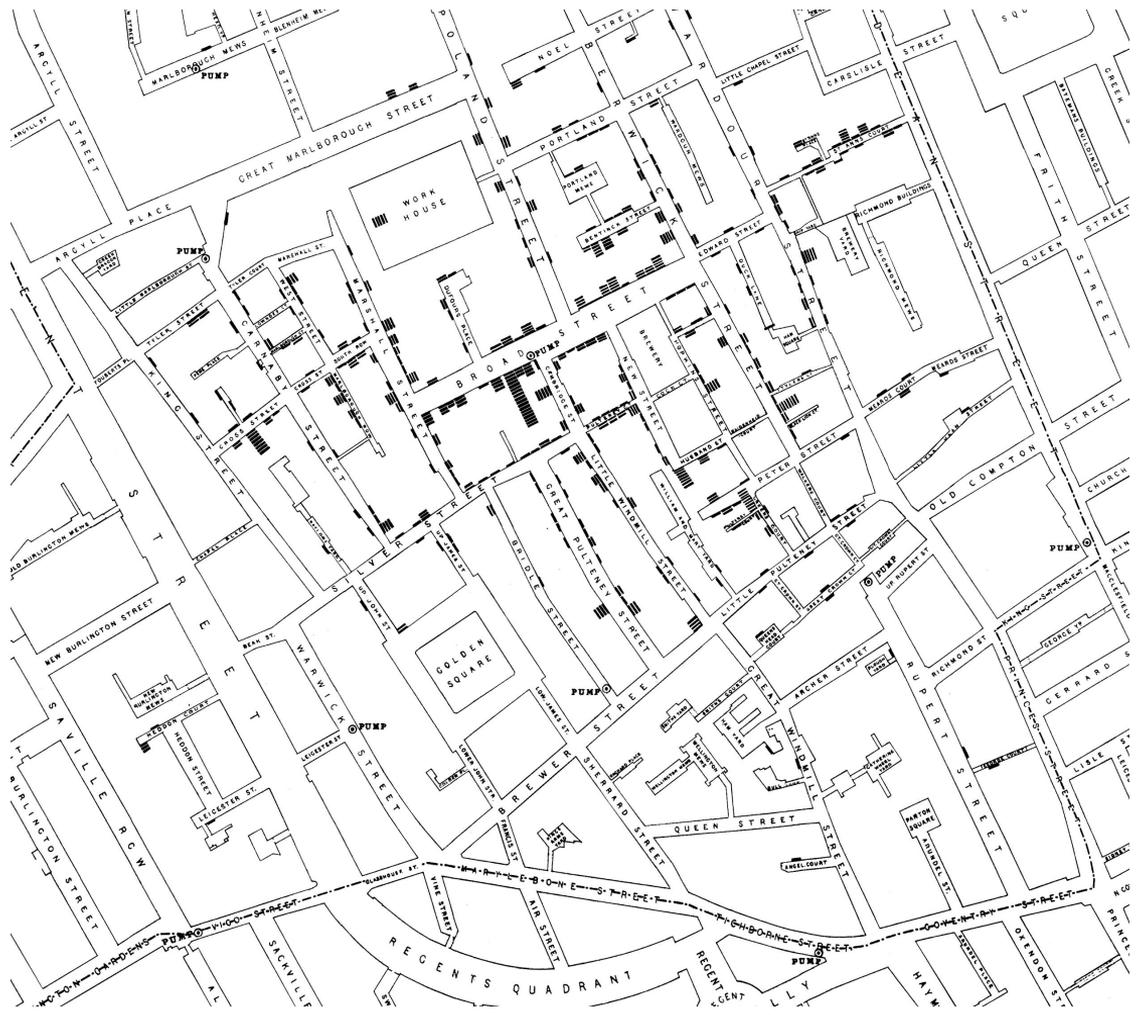
- JoAnn Manson, chief of Preventive Medicine at Brigham and Women’s Hospital, Boston

Miasmas, miasmaticism, miasmaticists

- **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





the john snow, london



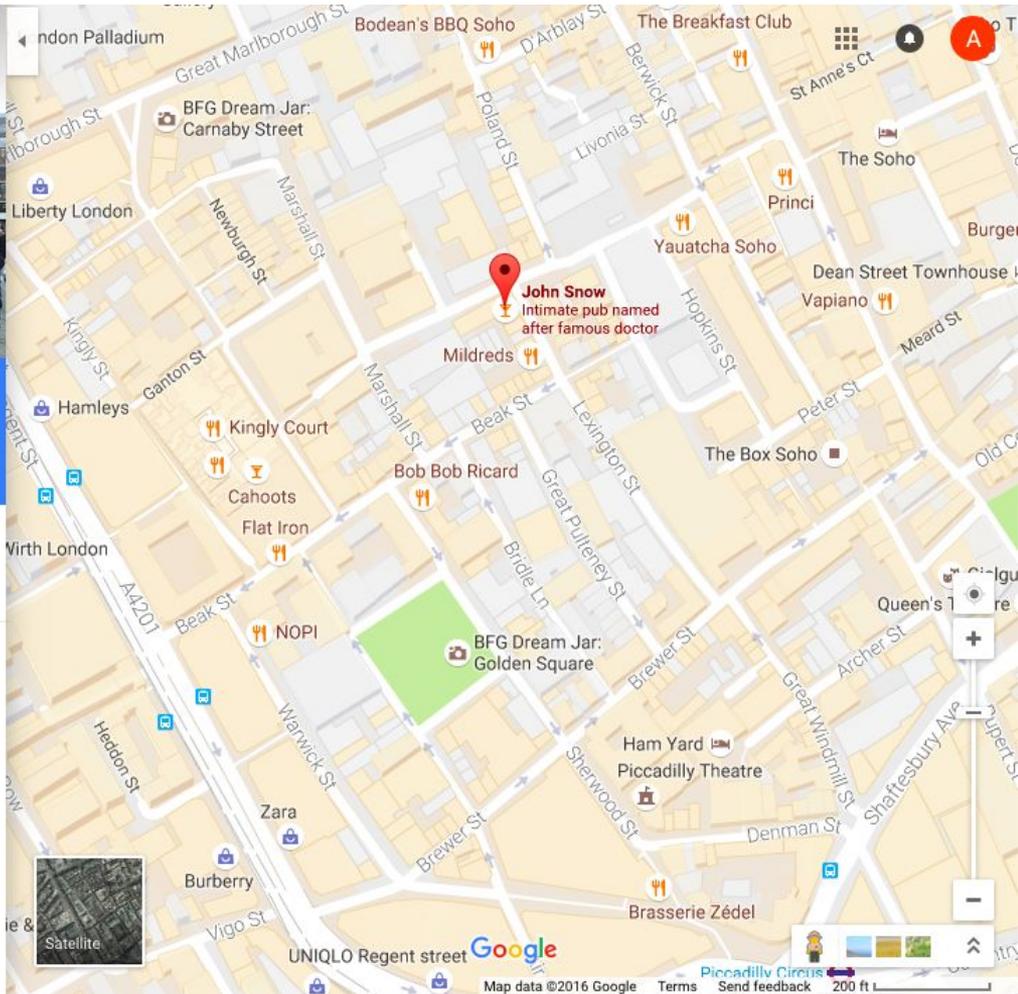
John Snow
3.6 ★★★★★ 145 reviews
Pub

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



London Palladium

Bodean's BBQ Soho

The Breakfast Club

Great Marlborough St

BFG Dream Jar: Carnaby Street

Polard St

Berwick St

St Anne's Ct

Liberty London

Newburgh St

Marshall St

The Soho

Princi

Yauatcha Soho

Dean Street Townhouse

Vapiano

John Snow
Intimate pub named after famous doctor

Mildreds

Kingly Court

Bob Bob Ricard

The Box Soho

Hamleys

Ganton St

Beak St

Lexington St

Peter St

Meard St

Ham Yard

Piccadilly Theatre

Queen's 1

Flat Iron

NOPI

BFG Dream Jar: Golden Square

Brewer St

Archer St

Ham Yard

Piccadilly Theatre

Denman St

Warwick St

Zara

Burberry

Sherwood St

Brasserie Zédel

Ham Yard

Piccadilly Theatre

Denman St

Sherwood St

Brasserie Zédel

UNIQLO Regent street

Google

Map data ©2016 Google

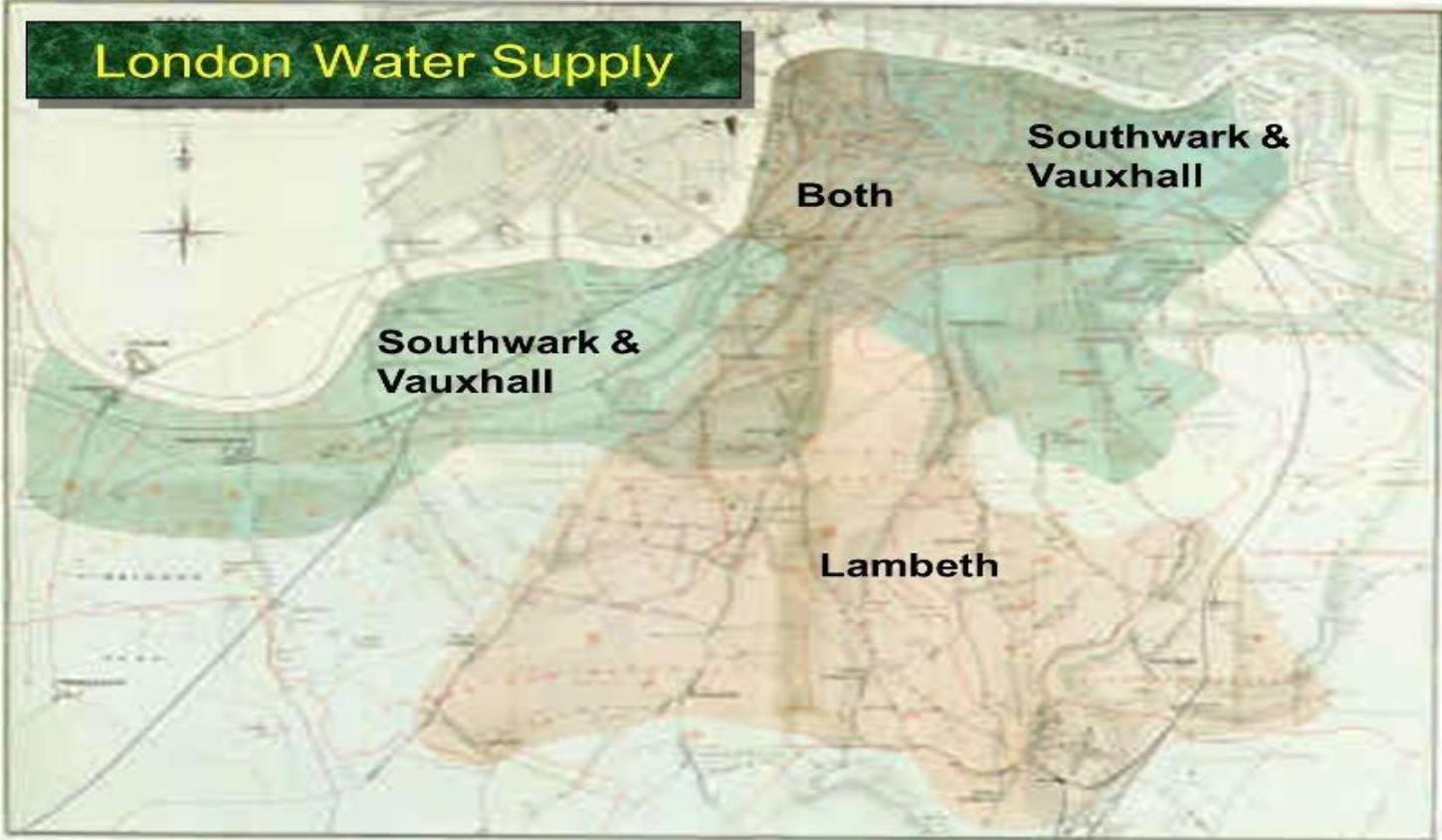
Terms

Send feedback

200 ft



London Water Supply



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 \neq Haphazard
