

CompSci 190: Cause & Effect

Jeff Forbes

August 27, 2018

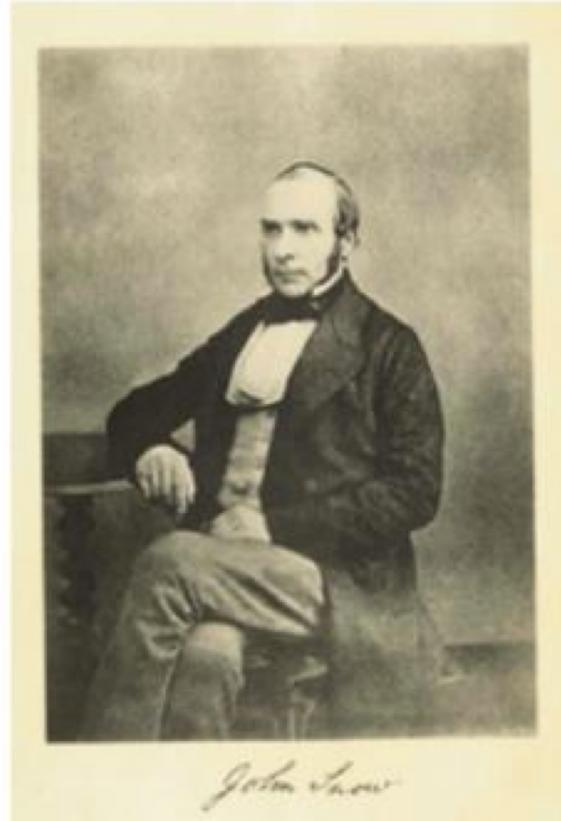
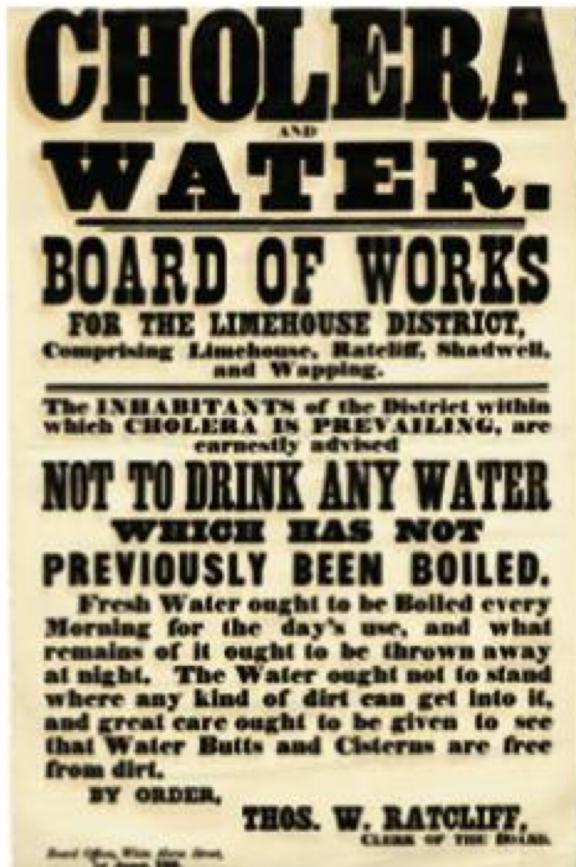
Sit in the first six rows!

Plan For The Day (PFTD)

- Be able to understand how association can lead to uncovering causality
- Be able to use a Jupyter notebook and the testing framework
- Be able to compute values with some basic Python expressions

Causality

Cholera & John Snow?

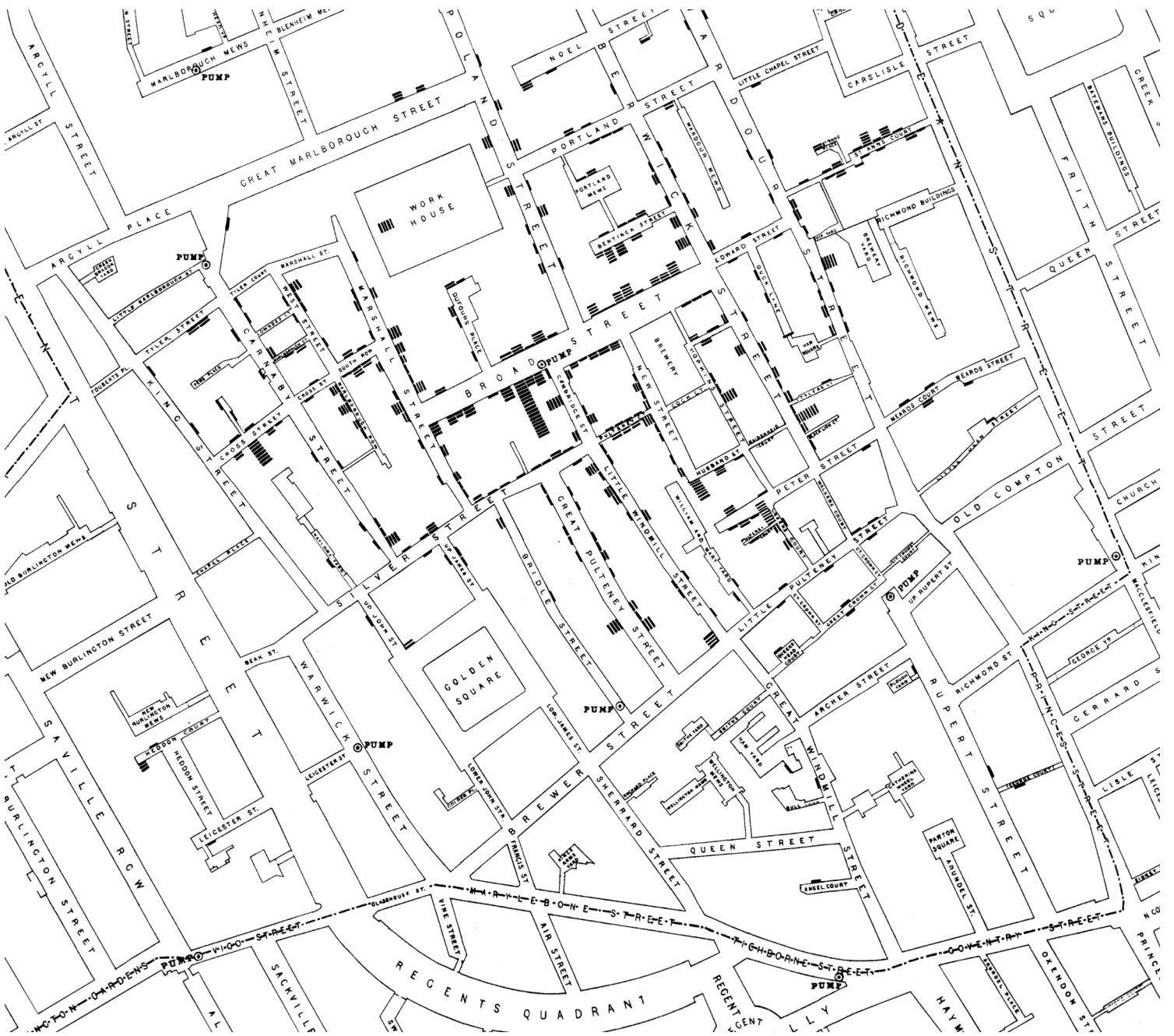


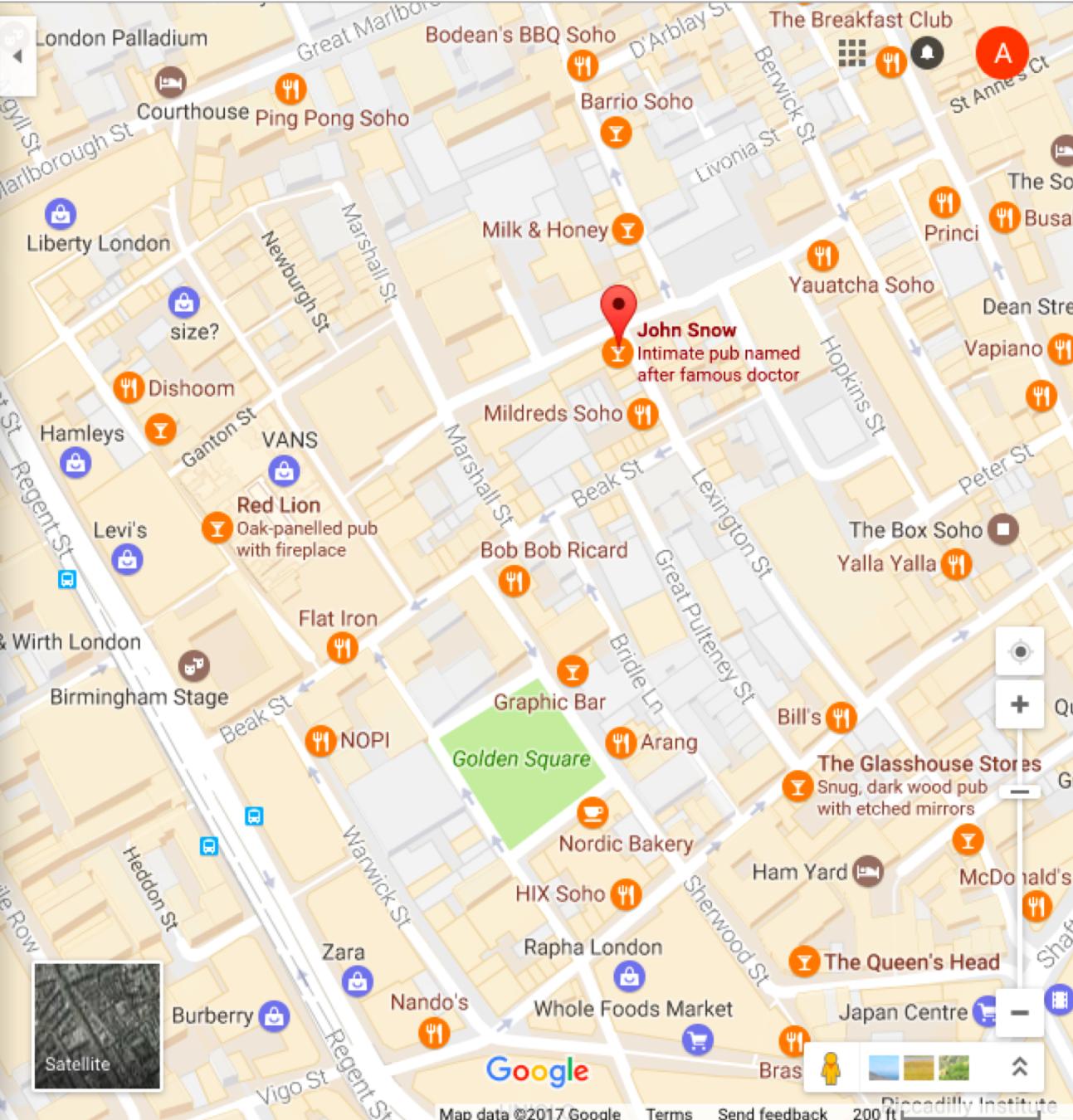
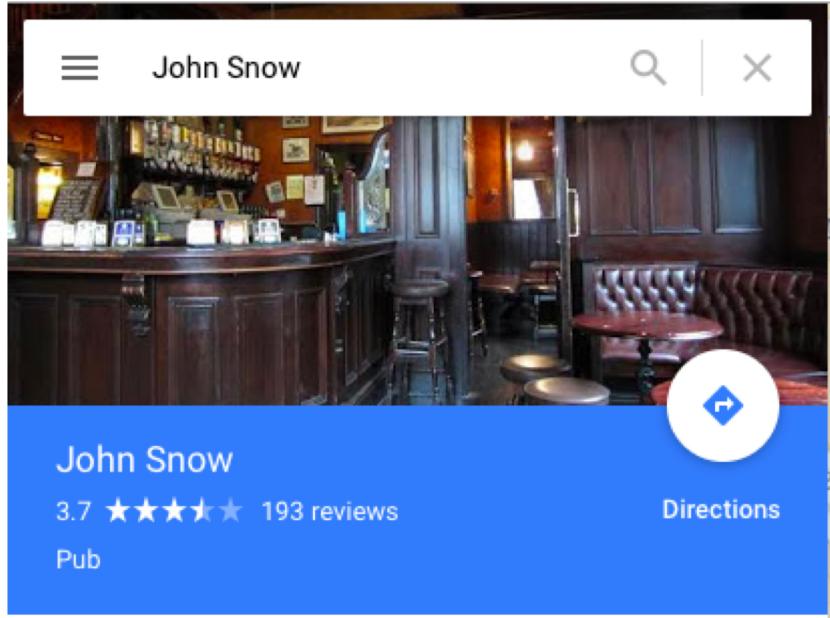
Miasmas, miasmatism, miasmatists

- 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

Observation

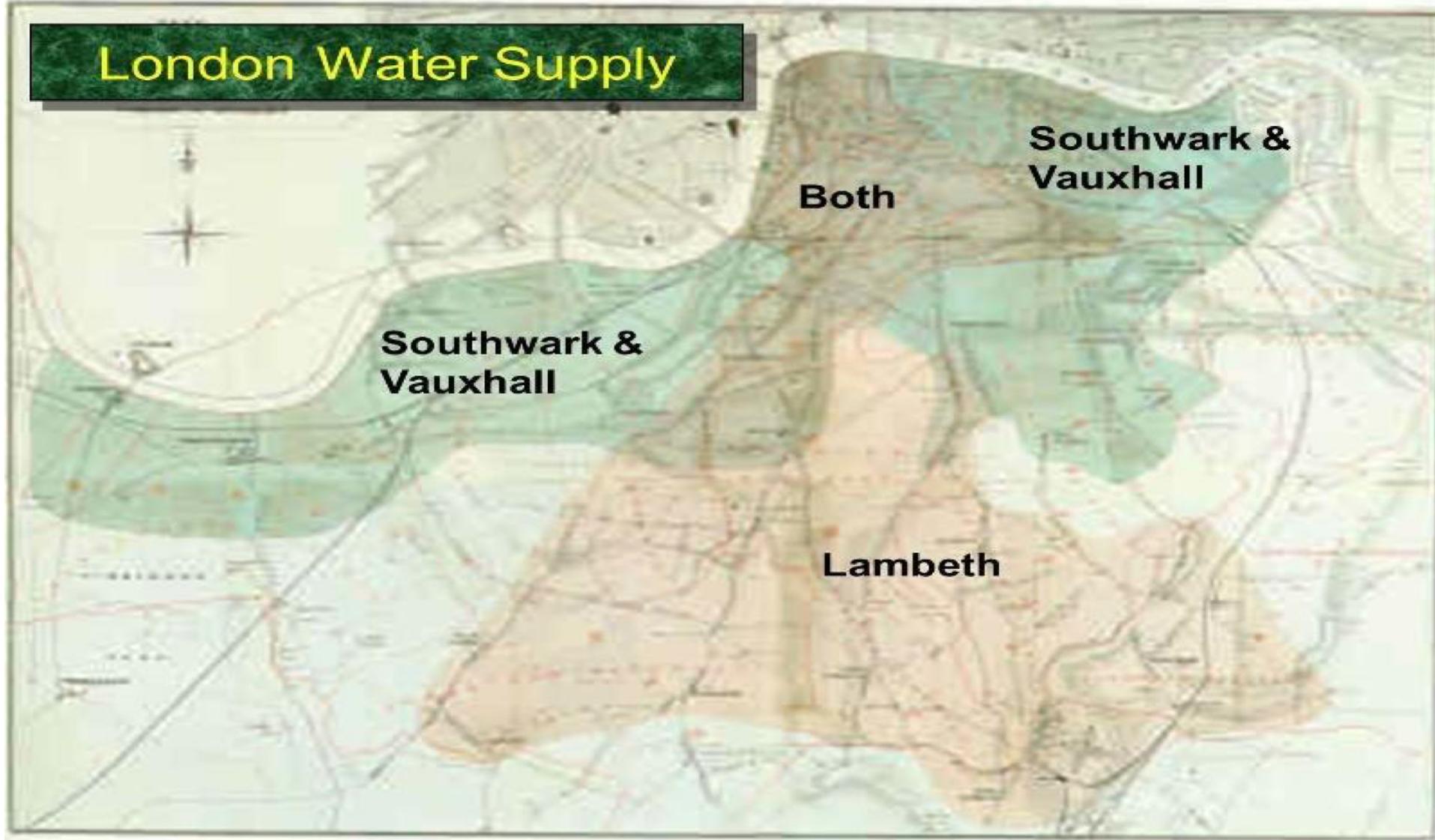
- **individuals**, study subjects, participants, units
 - *People living in 1854 London*
- **treatment**
 - *Drinking water*
- **outcome**
 - *Cholera*
- Is there **any relation** between water consumption and cholera?
 - **association**







London Water Supply



Establishing Causality

- Does drinking tainted water **lead to** cholera?
- Consider water from different water companies
 - **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

What does this table tell us about S&V vs. Lambeth?

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**.
- Solution: If you assign individuals to treatment and control **at random**, then the two groups are likely to be similar apart from the treatment.
 - **Randomized Controlled Experiment**

Lab 0: Notebooks & Expressions

Arithmetic Operators

Operation	Operator	Example	Value
Addition	+	$2 + 3$	5
Subtraction	-	$2 - 3$	-1
Multiplication	*	$2 * 3$	6
Division	/	$7 / 3$	2.66667
Remainder	%	$7 \% 3$	1
Exponentiation	**	$2 ** 0.5$	1.41421

(Lab)

Ints and Floats

- Python has two real number types
 - `int`: an integer of any size
 - `float`: a number with an optional fractional part
 - An `int` never has a decimal point; a `float` always does
 - A `float` might be printed using scientific notation

Limitations of `float` values:

1. They have limited size (but the limit is huge)
2. They have limited precision of 15-16 decimal places
3. After arithmetic, the final few decimal places can be wrong

Arithmetic Question

Rank the results of the following expressions in order from least to greatest

A. $3 * 10 ** 10$

B. $10 * 3 ** 10$

C. $(10 * 3) ** 10$

D. $10 / 3 / 10$

E. $10 / (3 / 10)$

[http://bit.ly/
FoDS-f18-0829-1](http://bit.ly/FoDS-f18-0829-1)

What's next?

- Review Chapter 2 and 3 of *Computational and Inferential Thinking*
- Complete the team-maker survey (Will be published to website on Thursday!)
- Tell a friend
 - There's still space!