

2) Causality and Experiments

Vitor Kamada

December 2019

Tables, Graphics, and Figures from

**Computational and Inferential Thinking:
The Foundations of Data Science**

Adhikari & DeNero (2019): Ch 2. Causality and Experiments

<https://www.inferentialthinking.com/chapters/02/causality-and-experiments.html>

How Cholera is Contracted and Spread?

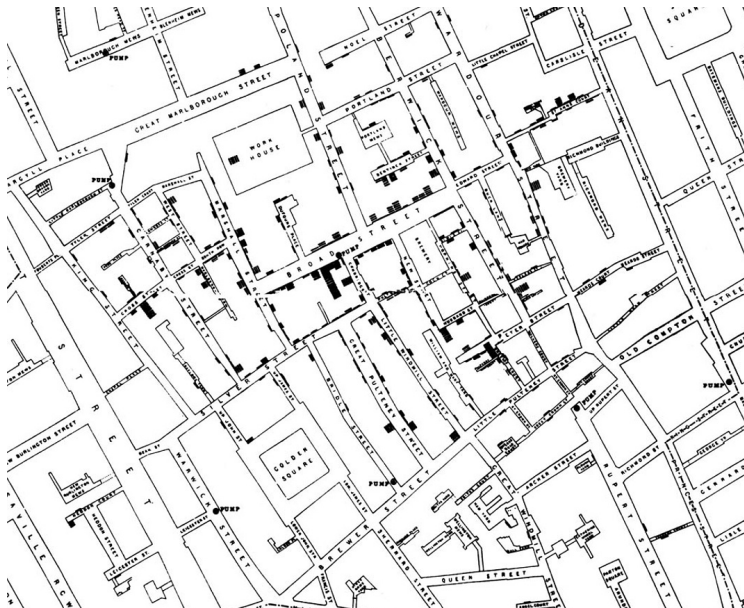
London in the 1850's

“miasmas”: invisible poisonous particles from decaying matter (bad smells)

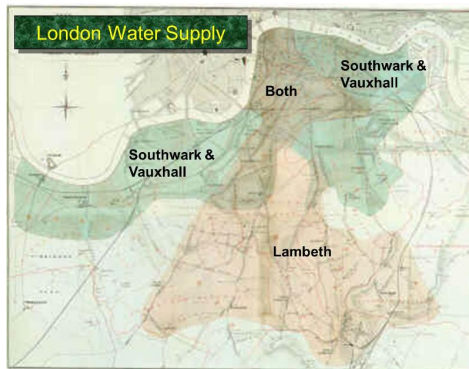
John Snow: Vomiting and Diarrhea

- Deaths clustered around the Broad Street pump (BSP)
- Scattered deaths in houses several blocks away from the BSP: children who drank from the BSP on their way to school

Snow's Original Map



Snow's “Natural Experiment”



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

Coffee and Lung Cancer

Observational Studies in the 1960's:

- coffee drinkers had higher rates of lung cancer than those who did not drink coffee

$\uparrow \text{coffee} \leftrightarrow \uparrow \text{lung cancer}$

Confounding Factor: smoking

PROGRESA increased School Enrollment

Mexican villages in the 1990's

Randomized Controlled Trial (RCT) implies no confounding factors

Money to poor families if their children went to school regularly

	Control	Treatment
Boys	73%	77%
Girls	67%	75%

Banerjee & Duflo (2012). Poor Economics.

Using Stents to Prevent Strokes

Chimowitz MI, Lynn MJ, Derdeyn CP, et al. 2011. **Stenting versus Aggressive Medical Therapy for Intracranial Arterial Stenosis.** *New England Journal of Medicine* 365:993-1003.

	0-30 days		0-365 days	
	stroke	no event	stroke	no event
treatment	33	191	45	179
control	13	214	28	199
Total	46	405	73	378

$$P(\text{Stroke}|\text{treatment}) = \frac{45}{224} = 20\%$$

$$P(\text{Stroke}|\text{control}) = \frac{28}{227} = 12\%$$

US Bureau of Transportation Statistics (2016)

Arrivals to Four Destinations				
		Airline		
Count	Column %	Delta	American	Total
Arrival	Delayed	659 (20%)	1,685 (22%)	2,344
	On Time	2,596 (80%)	5,966 (78%)	8,562
	Total	3,255	7,651	10,906

Simpson's Paradox

On Time %	Delta	American
Boston	80.1%	81.7%
Orlando	80.5%	84.5%
Philadelphia	70.5%	74.3%
San Diego	84.2%	85.4%

Airlines by Destination

Row %	Destination				
Arrival	Boston	Orlando	Philadelphia	San Diego	Total
Delta	1,409	1,168	312	366	3,255
	43%	36%	10%	11%	
American	1,826	970	4,423	432	7,651
	24%	13%	58%	6%	
Total	3,235	2,138	4,735	798	10,906