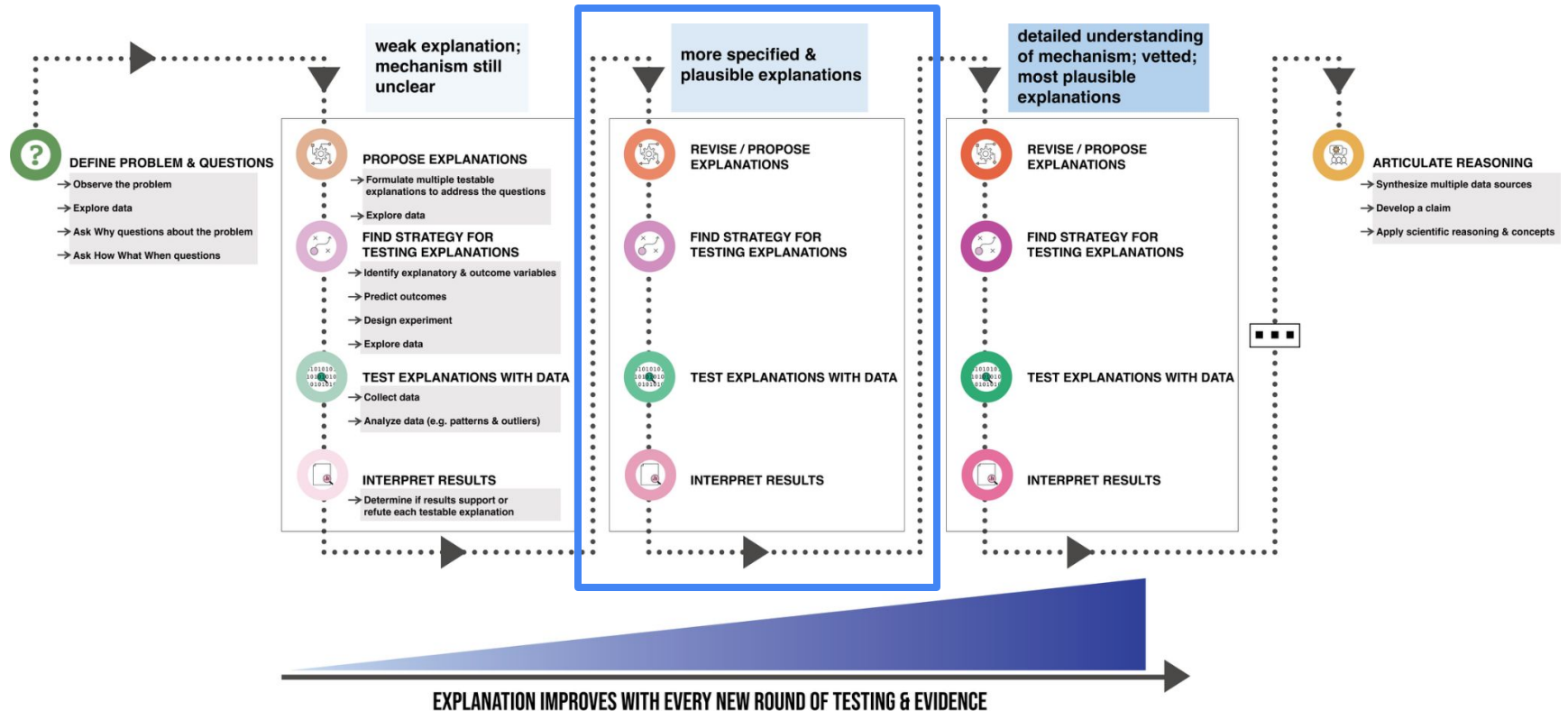


The South London Experiment

Data Science Reasoning Framework





London Water Suppliers

North of the Thames

GREEN	CHELSEA COMPANY	source of supply at C. NEAR THE RED HOUSE
YELLOW	WEST MIDDLESEX C ^o	HAMMERSMITH
PURPLE	GRAND JUNCTION C ^o	KEW BRIDGE
BLUE	HAMPSTEAD COMP ^y	SPRINGS ON THE HILL
PINK	NEW RIVER COMP ^y	RIVERS AMWELL & LEA
BROWN	EAST LONDON COMP ^y	RIVER LEA

South of the Thames.

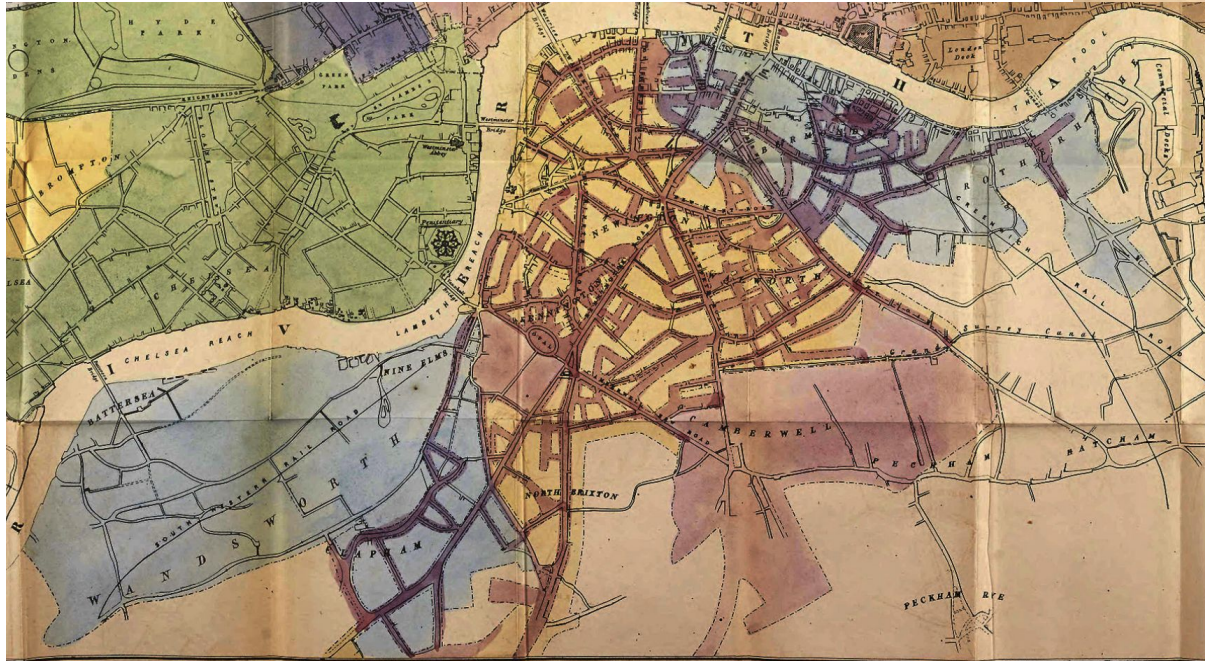
BLUE	SOUTHWARK COMP ^y	BATTERSEA
YELLOW	LAMBETH COMPANY	PEDLAR'S ACRE NEAR WATERLOO BRIDGE
PINK	VAUXHALL COMPANY	VAUXHALL BRIDGE

South London Water Suppliers

South of the Thames.

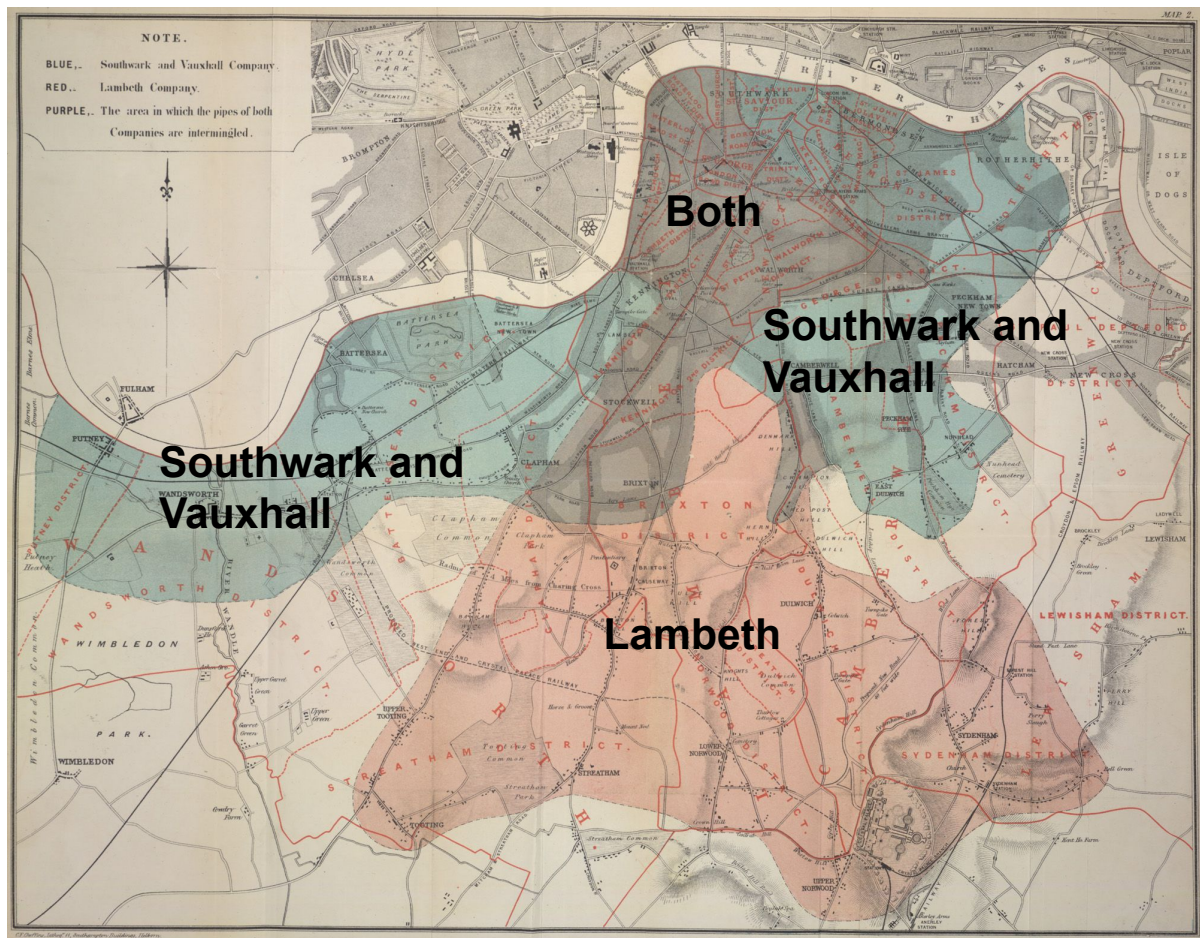
BLUE	SOUTHWARK COMPT	BATTERSEA
YELLOW	LAMBETH COMPANY	PEDLAR'S ACRE NEAR WATERLOO BRIDGE
PINK	VAUXHALL COMPANY	VAUXHALL BRIDGE

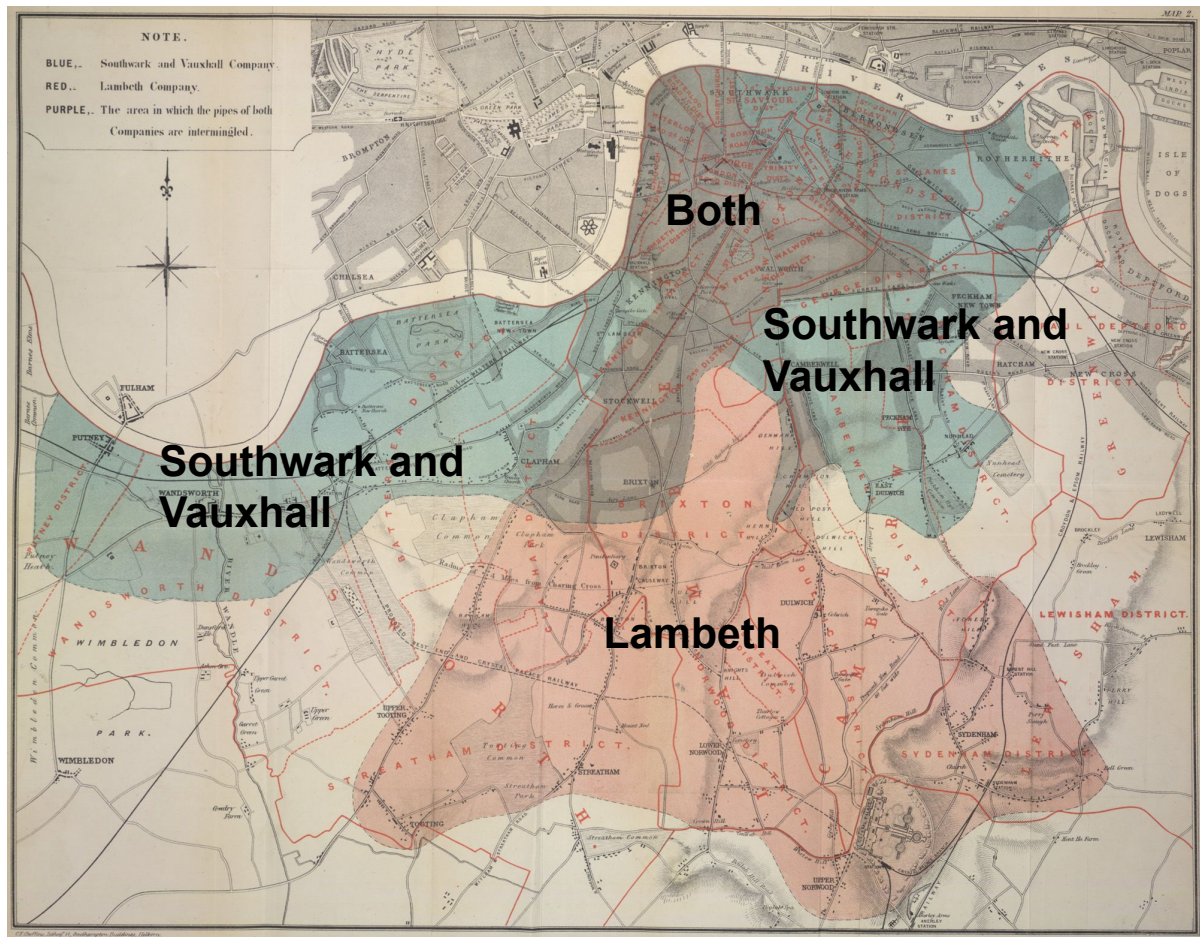
*Note. Where the Blue & Pink are intermingled the pipes of the Southwark & Vauxhall Companies are in competition
Where the Pink & Yellow are intermingled the pipes of the Vauxhall & Lambeth Companies are in competition
In some parts as in the New Kent Road, Dover Road, Newington Causeway, Kennington Road, Gravel Lane, and other places the three Companies are in competition*



Designing Our Data Experiment

Difference in Difference

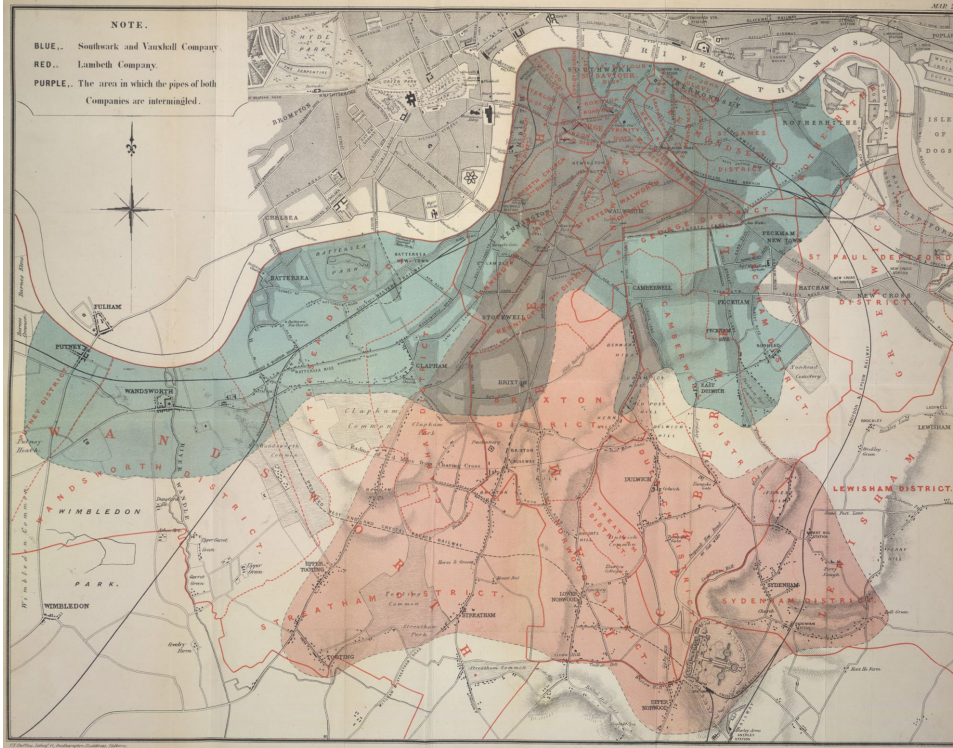




Why South London?

- Cholera impacted this area of London more than others
- 2 water suppliers that both got their water from similar locations on the Thames River in 1849
- The Lambeth company moved its water source upstream to less polluted waters in 1852
- There were cholera outbreaks in 1849 and 1854

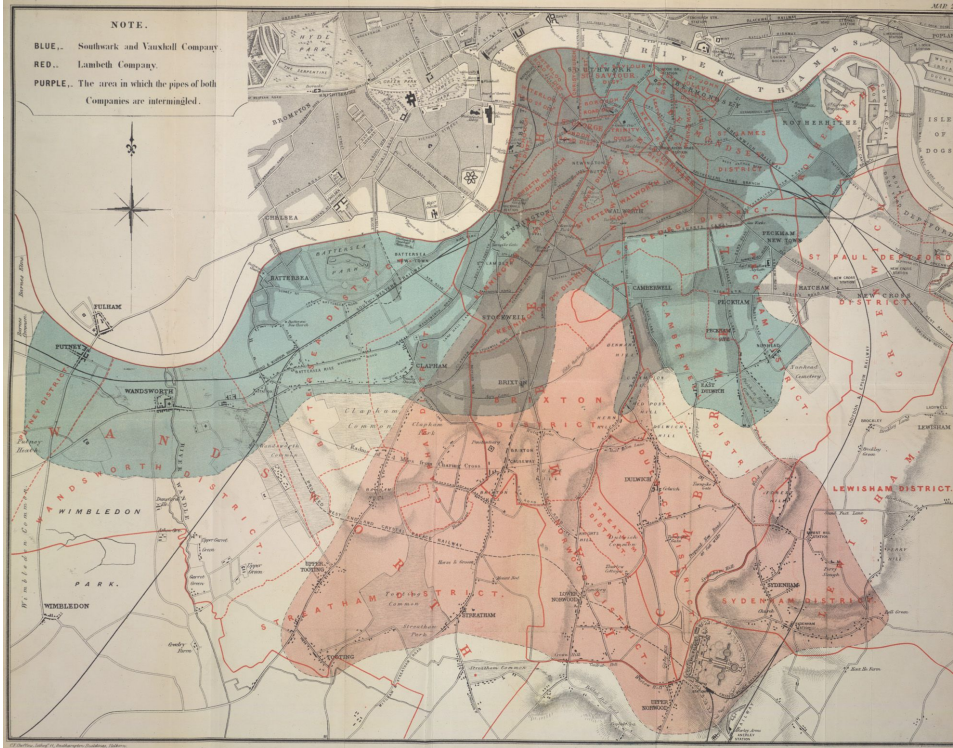
How can we think of our experiment...



South London Experiment

- Treatment group:
- Control group:
- Explanatory variable:
- Outcome variable:

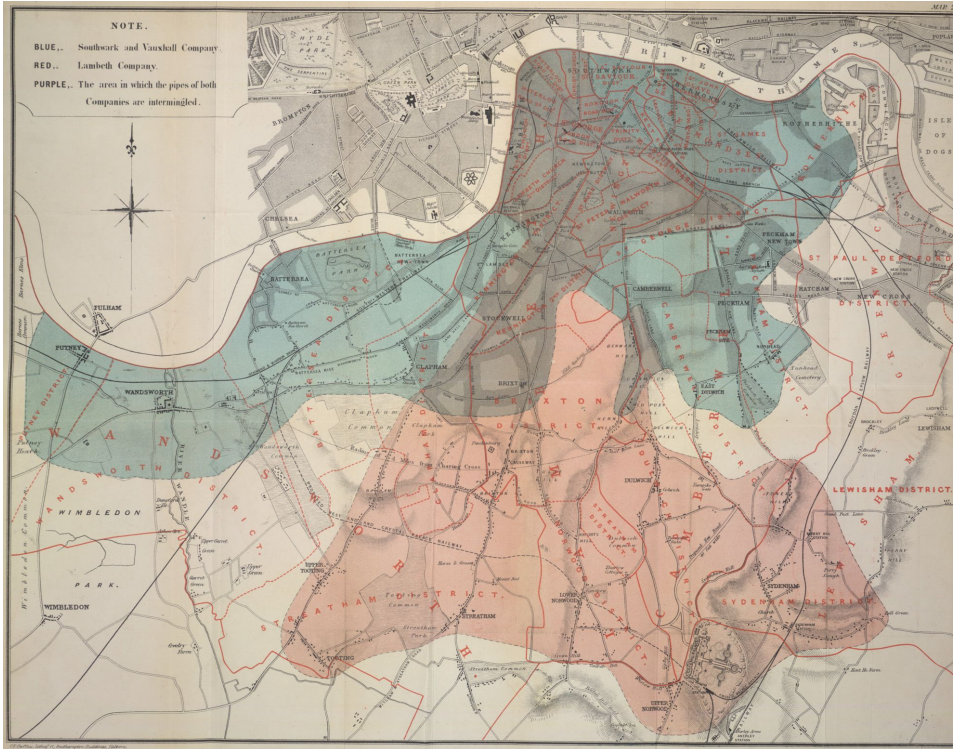
How can we think of our experiment...



South London Experiment

- Treatment group: Lambeth customers
- Control group:
- Explanatory variable:
- Outcome variable:

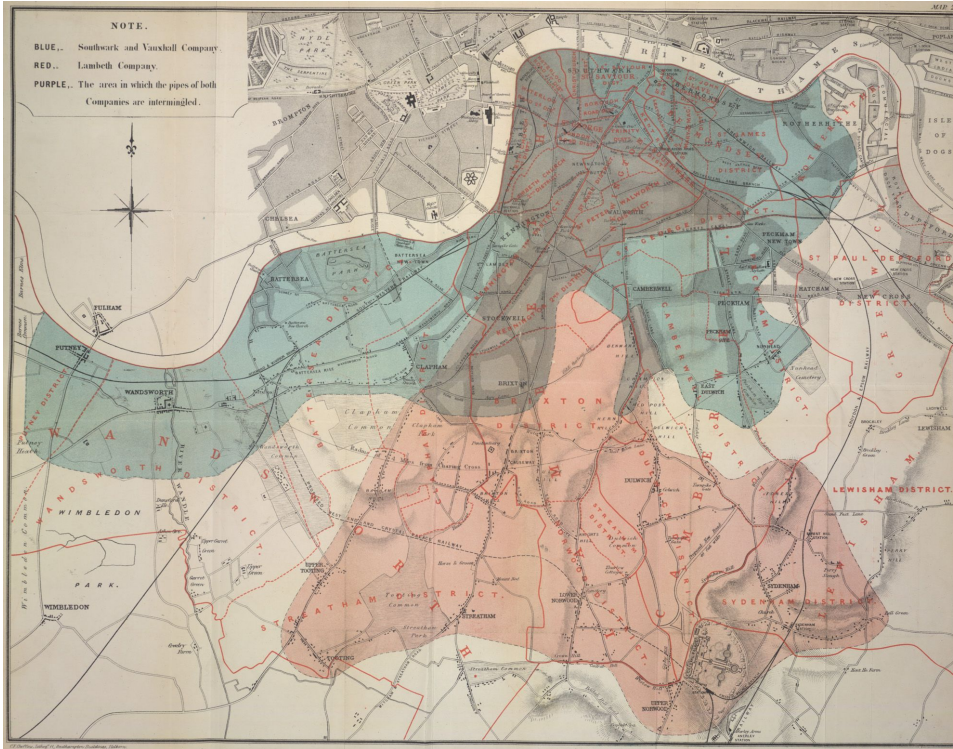
How can we think of our experiment...



South London Experiment

- Treatment group: Lambeth customers
- Control group: Southwark and Vauxhall customers
- Explanatory variable:
- Outcome variable:

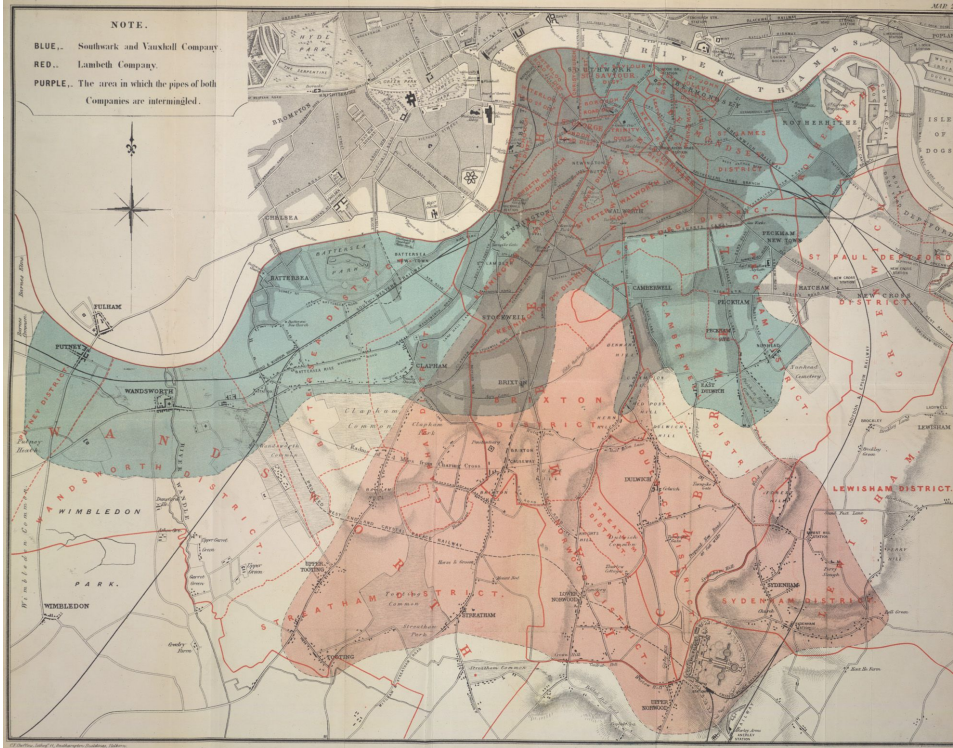
How can we think of our experiment...



South London Experiment

- Treatment group: Lambeth customers
- Control group: Southwark and Vauxhall customers
- Explanatory variable: Water source
- Outcome variable:

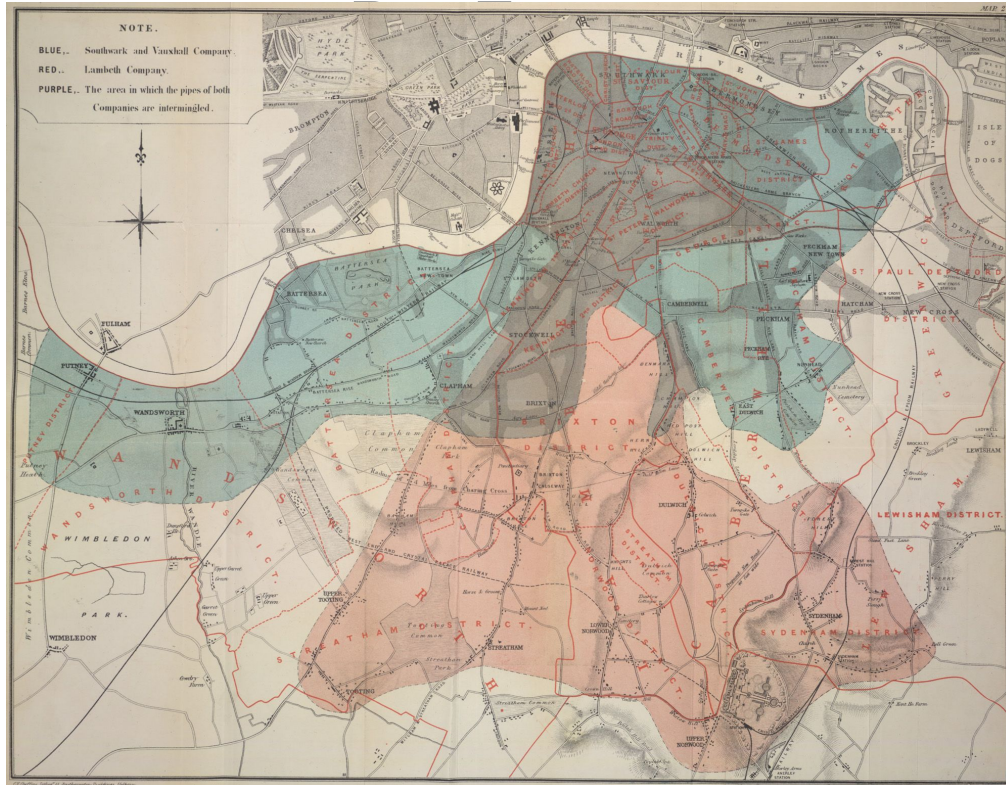
How can we think of our experiment...



South London Experiment

- Treatment group: Lambeth customers
- Control group: Southwark and Vauxhall customers
- Explanatory variable: Water source
- Outcome variable: Deaths from cholera

We can apply a Difference in Difference analysis to the South London Experiment!



South London Experiment

- Treatment group: Lambeth customers
- Control group: Southwark and Vauxhall customers
- Explanatory variable: Water source
- Outcome variable: Deaths from cholera
- **Water source was changed between 1849 and 1854 for Lambeth customers**

A look back at contingency tables


- They allow us to compare the outcome in one group to the outcome of another group, contingent on a condition (the proposed explanation). The condition is the potential explanation.

	Southside	Northside
White Sox		
Cubs		

A look back at contingency tables

- They allow us to compare the outcome in one group to the outcome of another group, contingent on a condition (the proposed explanation). The condition is the potential explanation.

	Southside	Northside
White Sox		
Cubs		



Proposed explanation is that Chicago side you identify with causes you to prefer one baseball team over the other.

Difference in Difference (DiD)

- Focuses on the change over time
- Both groups start out the same and one becomes the treatment at some point during the experiment because of change related to the explanatory variable
- For the bag toss game...

	First half of game	Second half of game
Group A (magnetic cans)		
Group B (normal cans)		

Difference in Difference (DiD)

- Focuses on the change in time
- Both groups start out the same and one becomes the treatment at some point during the experiment because of change related to the explanatory variable
- What would the DiD table look like for the South London Experiment?

	Treatment Group	Control Group
Time A		
Time B		

Difference in Difference (DiD)

- Focuses on the change in time
- Both groups start out the same and one becomes the treatment at some point during the experiment because of change related to the explanatory variable
- What would the DiD table look like for the South London Experiment?

	Mostly Lambeth (Treatment Group)	Mostly S&V (Control Group)
1849		
1854		