190F Foundations of Data Science

Fall 2018

Lecture 2

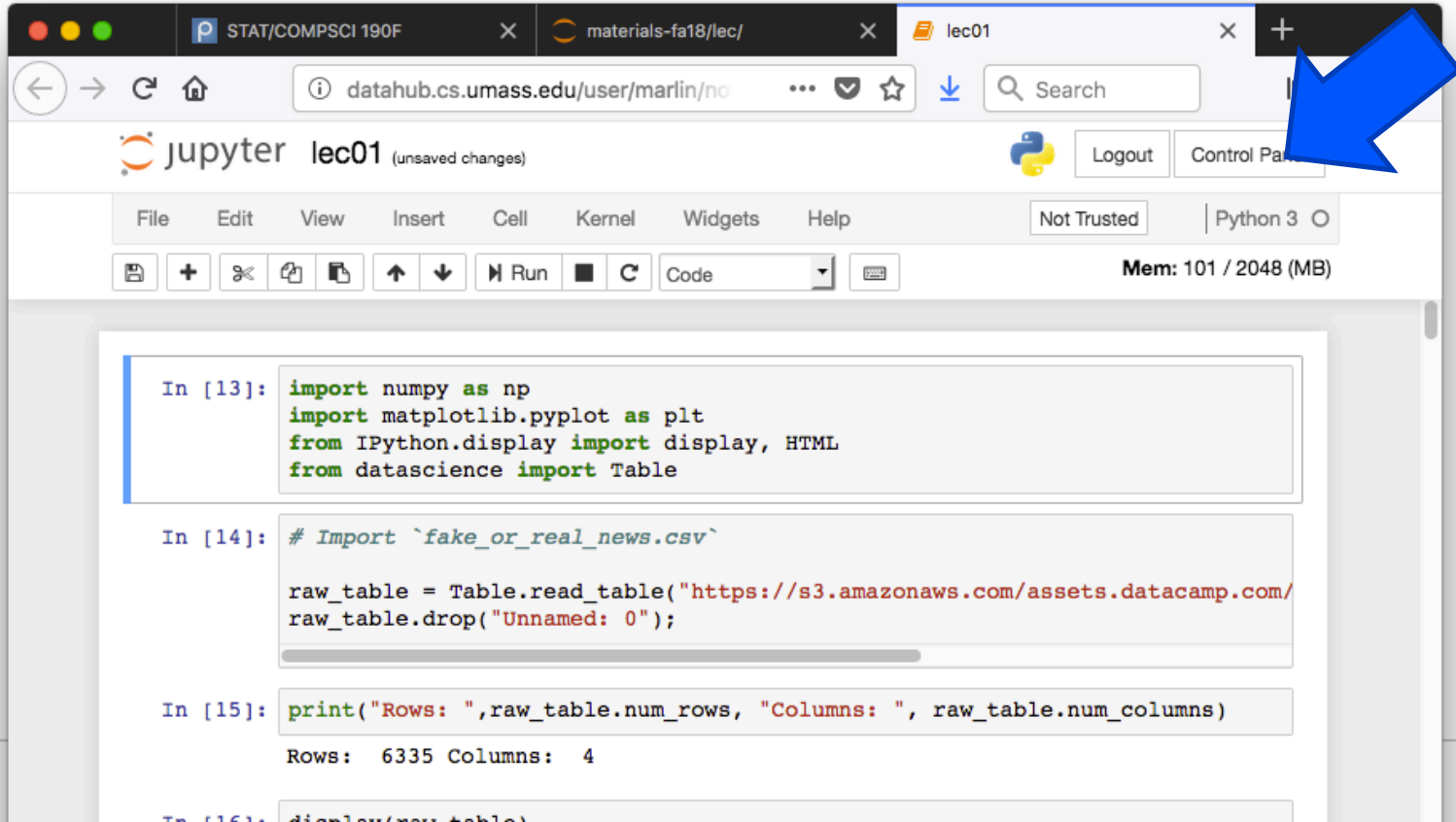
Cause and Effect

Announcements

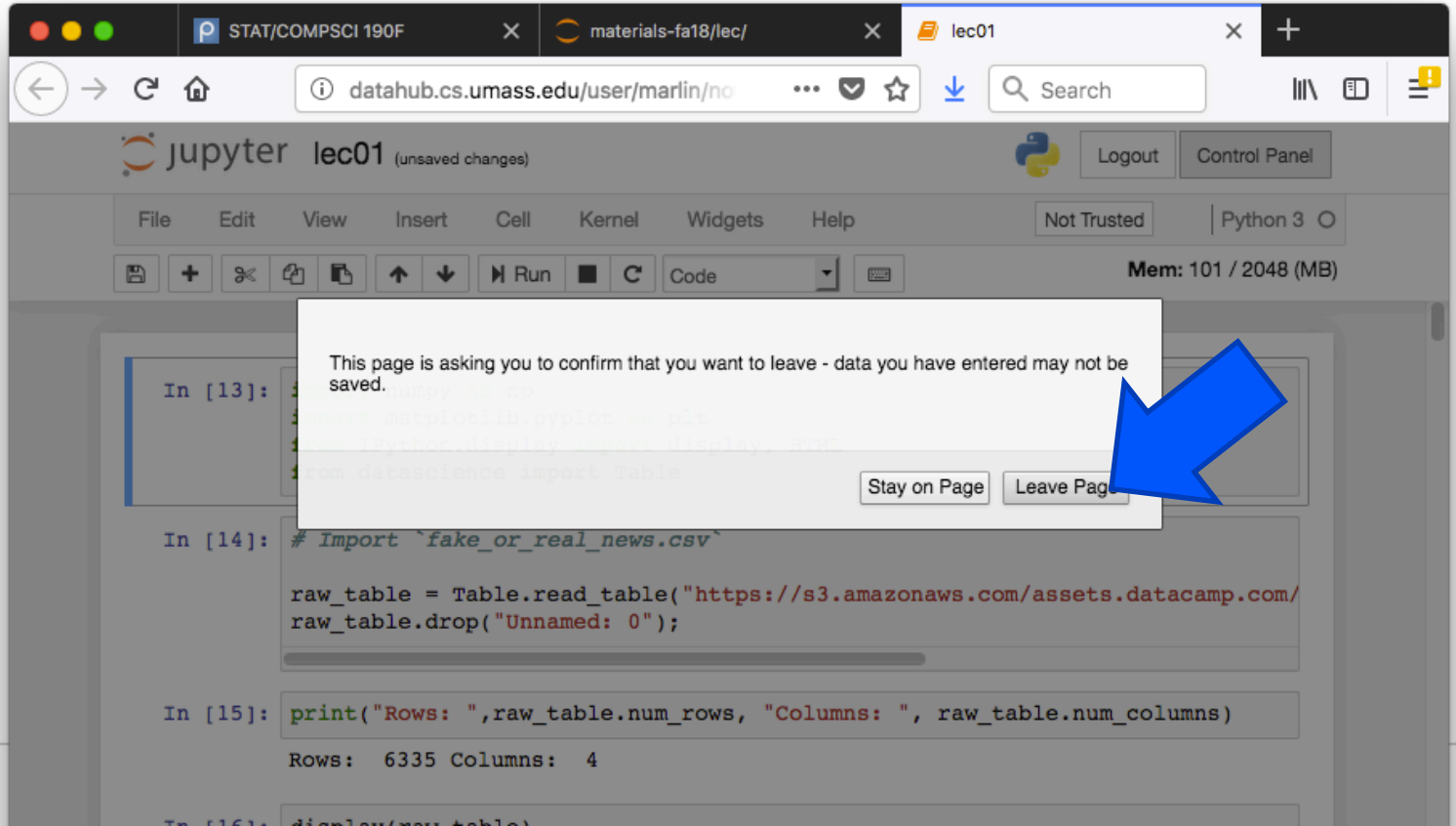
Datahub

- If you have an @umass.edu email address, you should now be able to access the course's data hub.
- Datahub uses UMass Google Apps authentication. Use your @umass.edu email address and Spire password to log in. It takes a minute to start up.
- When you're done working with the Datahub, make sure to shut your datahub server down and then log out.

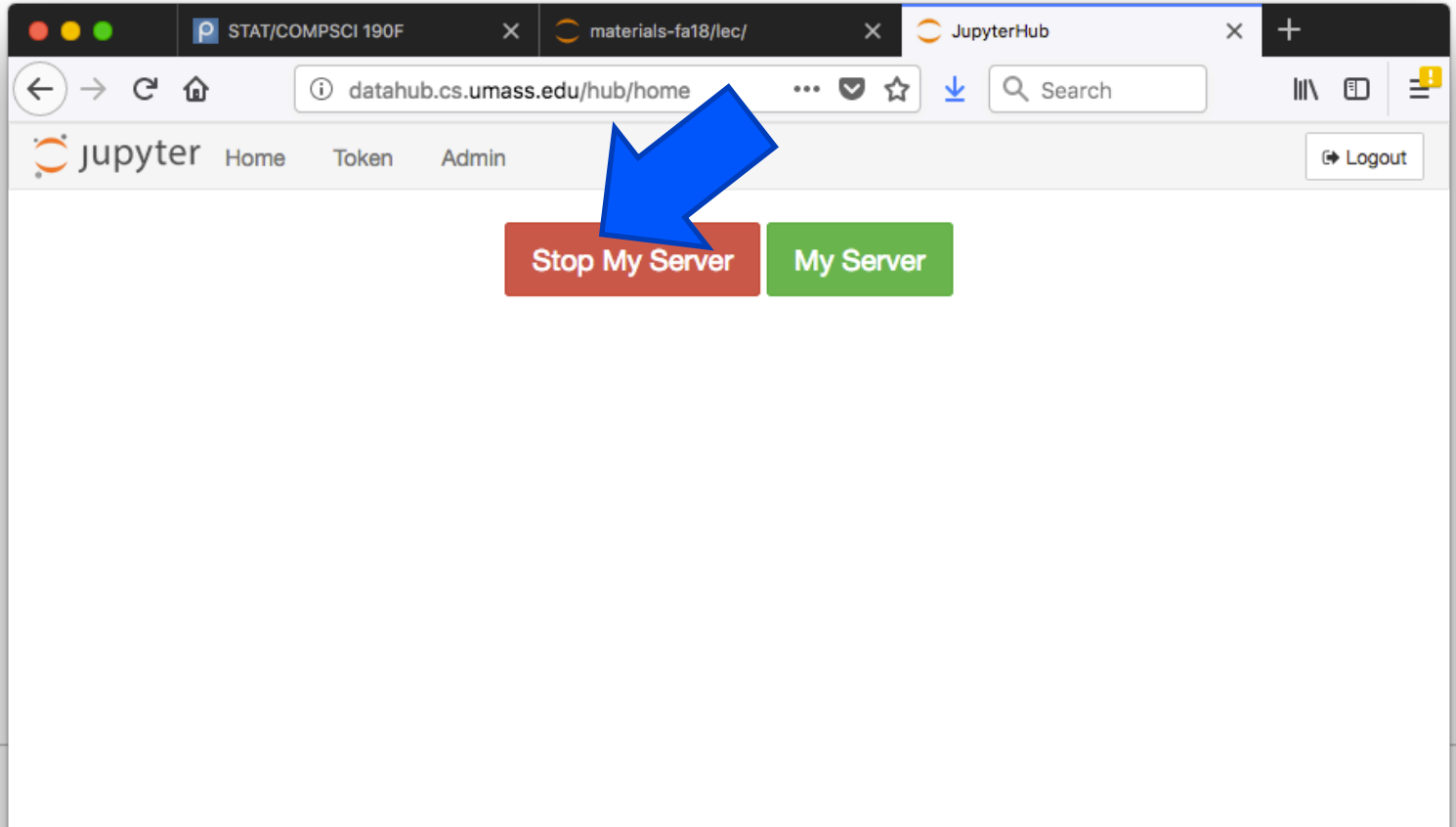
Stopping Your Data Hub



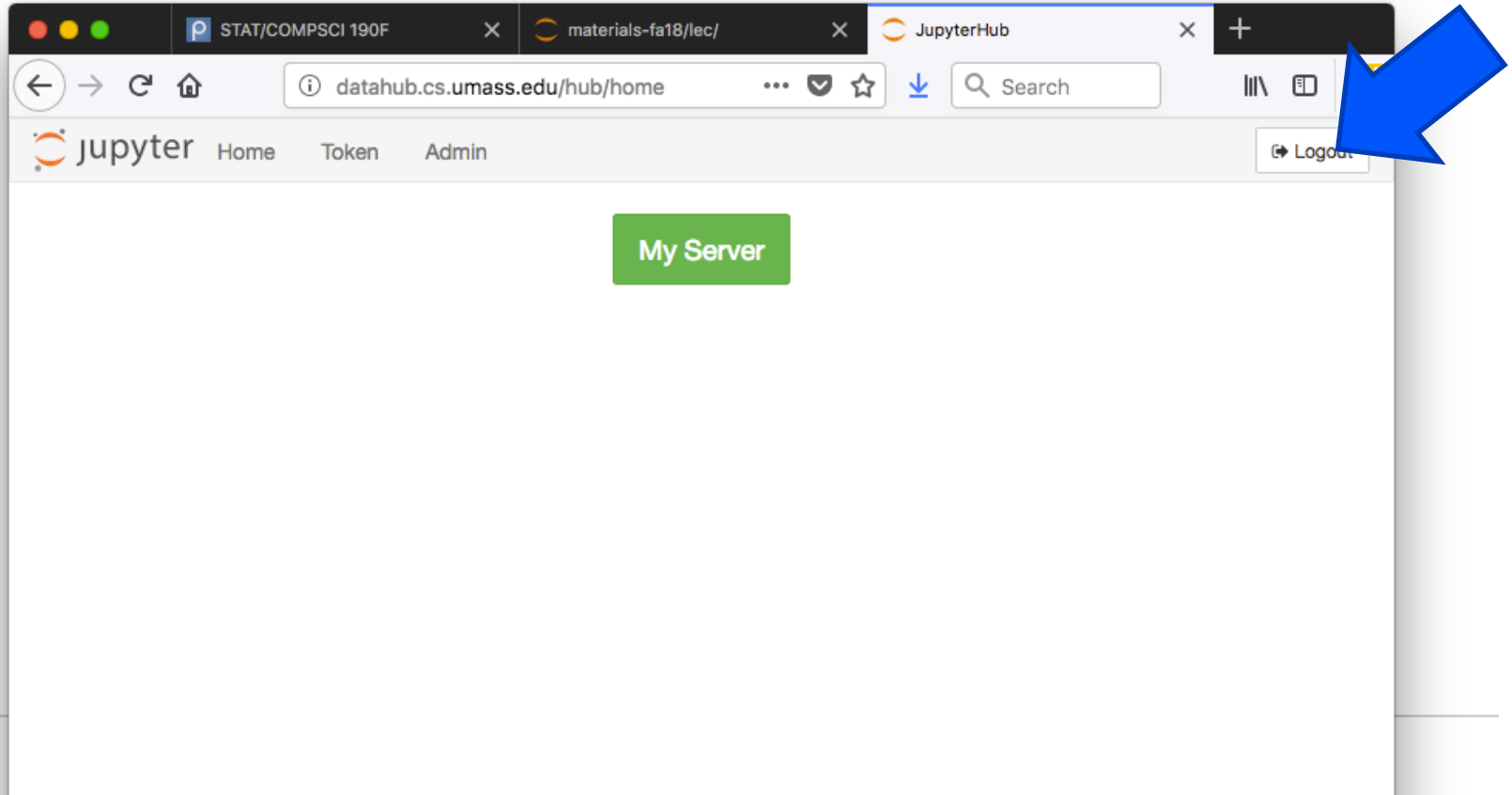
Stopping Your Data Hub



Stopping Your Data Hub



Stopping Your Data Hub



Association vs Causation

Examples

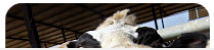
No level of alcohol consumption is healthy, scientists say

Fo

Dairy and meat 'beneficial for heart health and longevity'

Medic

Eating cheese may be associated with a lower risk of death —



and

tha

Busi

Eating chocolate regularly reduces your chance of heart failure according to new study

Huddersfield Examiner • yesterday

Chocolate and Heart Disease: Study

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

June 19, 2015 · 5:03 AM ET
Heard on Morning Edition

ALLISON AUBREY

- **Population** (Individuals, study subjects, participants, units): 20K *European adults followed for 12 years.*
- **Treatment:** *chocolate consumption*
- **Outcome:** *heart disease*

https://www.npr.org/sections/thesalt/2015/06/19/415527652/chocolate-chocolate-its-good-for-your-heart-study-finds

Chocolate and Heart Disease: Association

Question 1: Is there any association (any relationship) between chocolate consumption and heart disease?

- **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.”
- **Answer:** Yes, this points to an **association**

Chocolate and Heart Disease: Causation

Question 2: Does chocolate consumption lead to a reduction in heart disease?

- **This questions asks about 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*

Chocolate and Heart Disease: Alternatives

Question 3: Is the fact that people ate more chocolate the only possible cause for the observed effect of decreased heart disease risk?

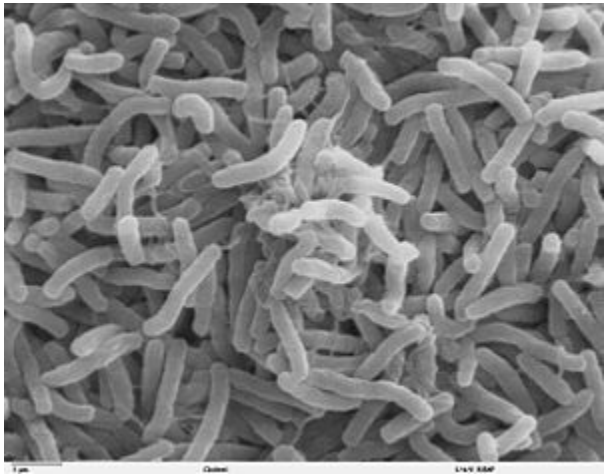
- For example, suppose the people who ate more chocolate tended to live in European countries with better health care?

- What if wealthier people eat more chocolate and can also afford better health care?

Proving Cause and Effect

Broad Street Cholera Outbreak

- The Broad Street cholera outbreak was a severe outbreak of cholera that occurred in 1854 near Broad Street in the Soho district of London, England.
- This outbreak killed 616 people.

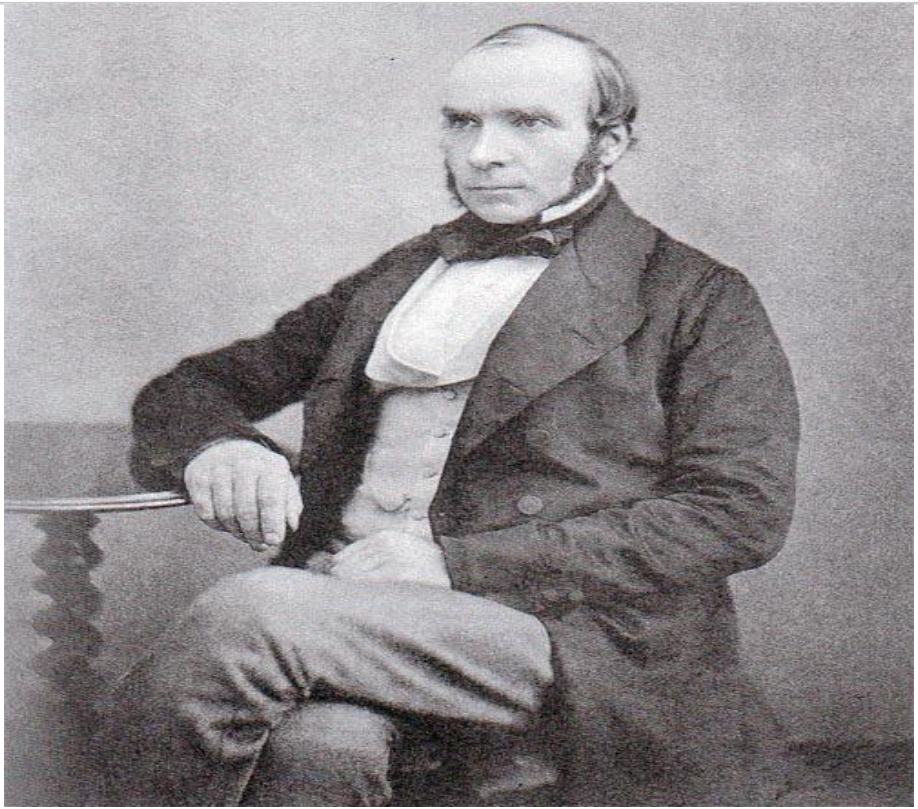


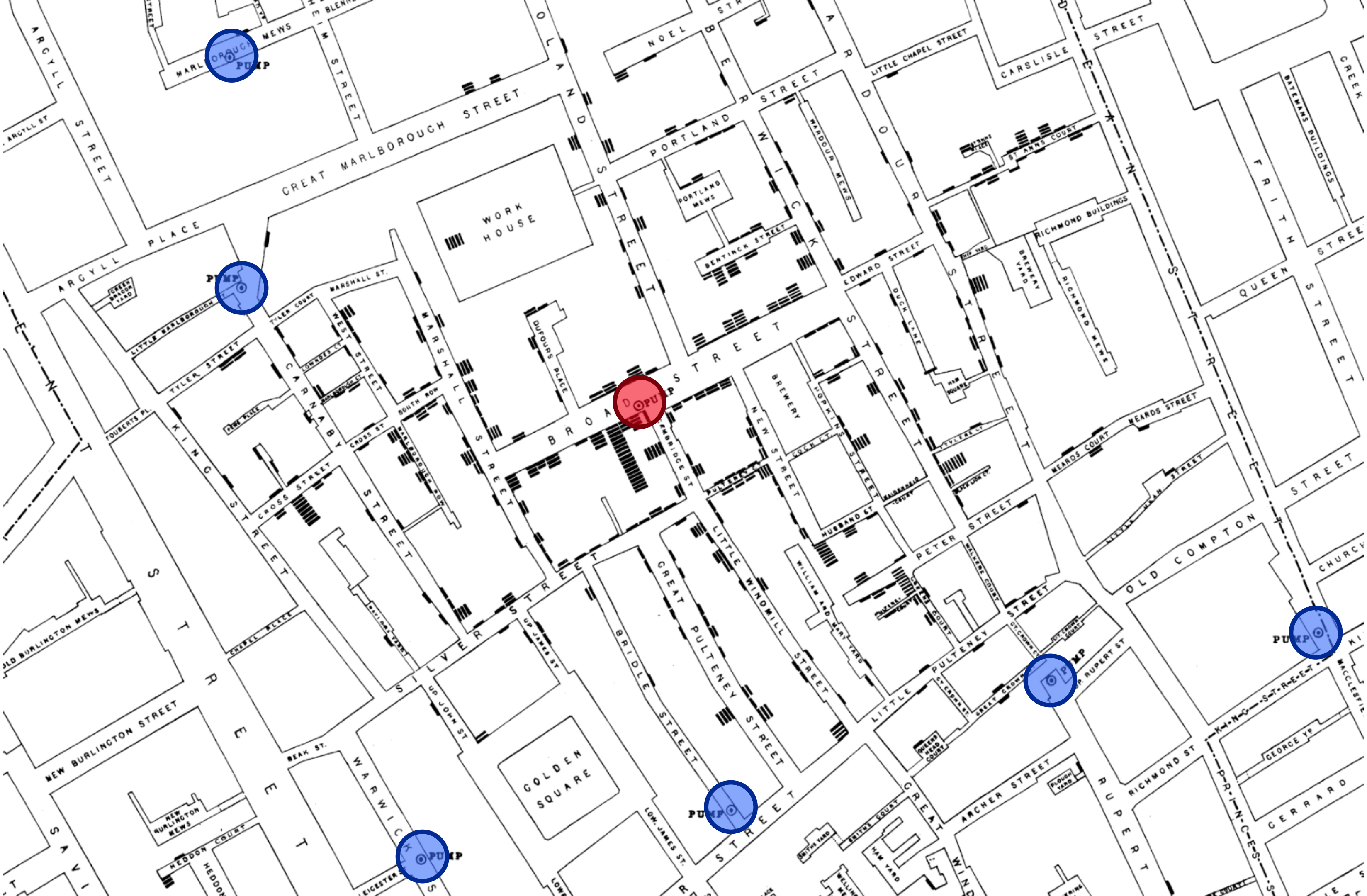
Two Theories of Cholera

- **Miasma theory:** cholera was caused by **particles in the air**, or "miasmata", which arose from decomposing matter or other dirty organic sources.

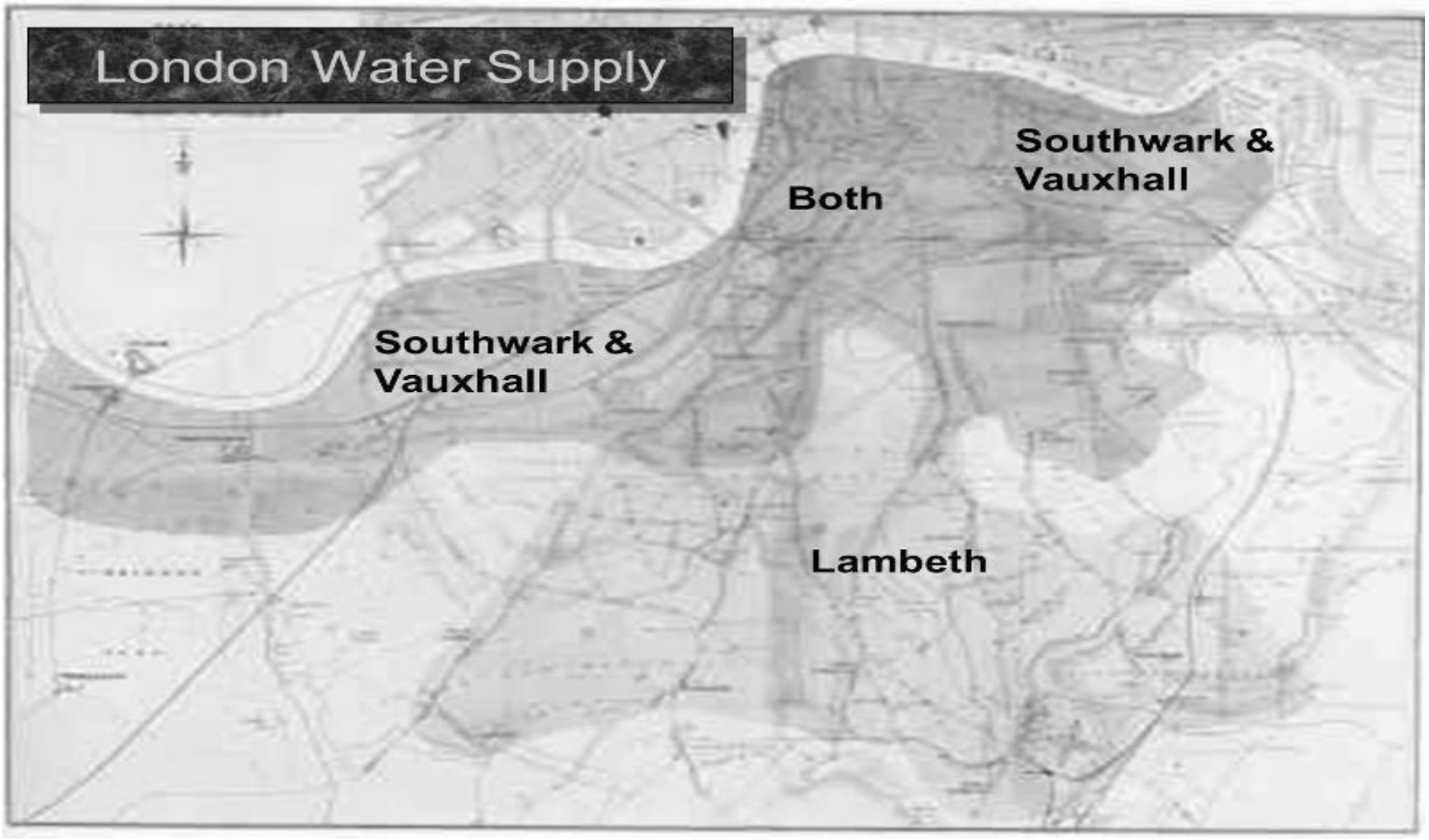
- **Germ theory:** the principal cause of cholera was a germ cell that had not yet been identified, but was **transmitted through food or drink**.

John Snow, 1813-1858




MARLBOROUGH MEWS
PUMP
ARGYLL STREET
GREAT MARLBOROUGH STREET
ARGYLL PLACE
LITTLE MARLBOROUGH
TYLER STREET
MARSHALL ST.
WORK HOUSE
DUFOURS PLACE
PORTLAND STREET
NOEL STREET
PORTLAND MEWS
BENTINCK STREET
BROAD STREET
LITTLE CHAPEL STREET
CARSLISLE STREET
ST ANNS COURT
RICHMOND BUILDINGS
RICHMOND MEWS
EDWARD STREET
QUEEN STREET
FRITH STREET
BREWERY
NEW STREET
MEARDS STREET
MEARDS COURT
PETER STREET
HUSBAND ST
OLD COMPTON STREET
CARNABY STREET
KING STREET
CROSS STREET
SOUTH ROW
MARSHALL STREET
SILVER STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
BRIDLE STREET
LITTLE PULTENEY STREET
GREAT CROWN COURT
RUPERT STREET
RICHMOND ST
GERRARD STREET
ARCHER STREET
GOLDEN SQUARE
NEW BURLINGTON STREET
NEW BURLINGTON MEWS
OLD BURLINGTON MEWS
CHAPEL PLACE
BEAK ST.
WARWICK STREET
UP JOHN ST
UP JAMES ST
LOW. JAMES ST.
HEDDON COURT
FOUBERTS PL.
GREEN DRAGON YARD
CHURCH
GEORGE YD


JOHN SNOW
JOHN SNOW


London Water Supply
Southwark &
Vauxhall
Both
Southwark &
Vauxhall
Lambeth

Comparison

- **Treatment group:** Do receive the treatment
- **Control group:** Do 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

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.

Confounding

- 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.

Randomization and Confounding

- 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 Experiments are the gold standard for establishing cause and effect.**

RCEs vs Observational Studies

- **Question:** If randomized controlled experiments can establish causality while observational studies are subject to confounding, why are so many studies observational?