Lecture 2

Cause and Effect

Contributions by Vinitra Swamy (vinitra@berkeley.edu) and Fahad Kamran (fhdkmrn@berkeley.edu)
Slides created by John DeNero (denero@berkeley.edu) and Ani Adhikari (adhikari@berkeley.edu)
Context: Last Lecture

- **Introduction**
  - Data science: drawing useful conclusions from data using computation
    - Exploration
    - Inference
    - Prediction
  - Course Policies
  - Demo

Cause and Effect
Announcements
Questions
Three coffees a day linked to a range of health benefits

Research based on 200 previous studies worldwide says frequent drinkers less likely to get diabetes, heart disease, dementia and some cancers.

The findings supported other studies showing the health benefits of drinking coffee. Photograph: Wu Hong/EPA
A Stronger Link?

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*
Is there any relation between chocolate consumption and heart disease?

- association
  - any relation
  - link
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

- 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
London, early 1850’s

Illustration from *Punch* (1852).
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”
Celebrity Miasmatists

- Florence Nightingale
  - “Lady with the Lamp”

- Edwin Chadwick
  - Commissioner of the General Board of Health

- There was one person who was a little doubtful...
Jon Snow, 281-302
John Snow, 1813-1858
Causation
London Water Supply

Southwark & Vauxhall

Both

Southwark & Vauxhall

Lambeth
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

<table>
<thead>
<tr>
<th>Supply Area</th>
<th>Number of houses</th>
<th>Cholera deaths</th>
<th>Deaths per 10,000 houses</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;V</td>
<td>40,046</td>
<td>1,263</td>
<td>315</td>
</tr>
<tr>
<td>Lambeth</td>
<td>26,107</td>
<td>98</td>
<td>37</td>
</tr>
<tr>
<td>Rest of London</td>
<td>256,423</td>
<td>1,422</td>
<td>59</td>
</tr>
</tbody>
</table>
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
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
Regardless of what the dictionary says, in probability theory

**Random ≠ Haphazard**
Jupyter Notebooks