

BC COMS 1016: Intro to Computational Thinking & Data Science

Lecture 1 – Course Introduction 10/26/2020

What is Computational Thinking?



Viewpoint | Jeannette M. Wing

Computational Thinking

It represents a universally applicable attitude and skill set everyone, not just computer scientists, would be eager to learn and use.



Computational thinking builds on the power and limits of computing processes, whether they are executed by a human or by a machine. Computational methods and models give us

cisely. Stating the difficulty of a problem accounts for the underlying power of the machine—the computing device that will run the solution. We must consider the machine's instruction set, its resource constraints, and its operating environment.

In solving a problem efficiently, we might further ask whether an approximate solution is good

What is Computational Thinking?



- Reformulating a seemingly difficult problem into one we know how to solve by:
 - reduction, transformation, or simulation
- Thinking at multiple levels of abstraction
- Fundamentals, not rote skills

<https://coms1016.barnard.edu/readings/Wing06-Comp-thinking.pdf>

What is Data Science?



- “*Data science is the study of extracting value from data*” – *Jeannette Wing*

What is Data Science?



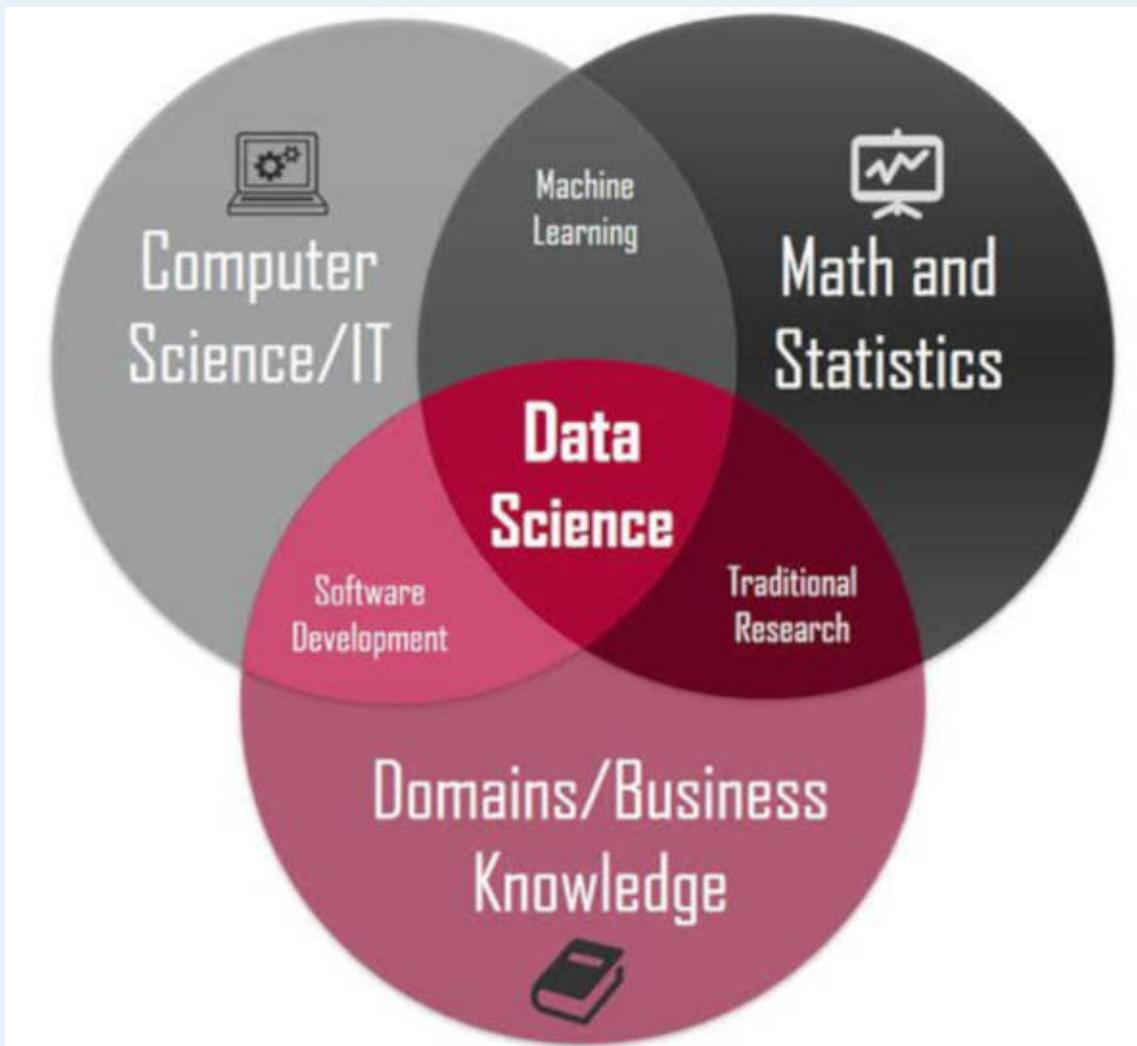
- “*Data science is the study of extracting value from data*” – Jeannette Wing
- Value
 - Requires domain expertise to determine what value is
 - *Value from data* is different based on the domain and the needs

What is Data Science?



- “*Data science is the study of extracting value from data*” – Jeannette Wing
- Extracting
 - emphasizes action on data
 - mining information

Math + Computer Science + Domain Knowledge



Data Science in this course





Data Science in this course

- Exploration
 - Discover patterns in data
 - Articulate insights (visualizations)

Data Science in this course



- Exploration
 - Discover patterns in data
 - Articulate insights (visualizations)

- Inference
 - Make reliable conclusions about the world
 - Statistics is useful

Data Science in this course



- Exploration
 - Discover patterns in data
 - Articulate insights (visualizations)
- Inference
 - Make reliable conclusions about the world
 - Statistics is useful
- Prediction
 - Informed guesses about unseen data



- Exploration **Week 1 - 3**
 - Discover patterns in data
 - Articulate insights (visualizations)
- Inference **Week 3 - 5**
 - Make reliable conclusions about the world
 - Statistics is useful
- Prediction **Week 6-7**
 - Informed guesses about unseen data



- Exploration **Week 1 - 3**
 - Introduction to Python
 - Working with data
- Inference **Week 3 - 5**
 - Probability
 - Statistics
- Prediction **Week 6-7**
 - Machine Learning
 - Regression & Classification



— Logistics —



- Course webpage:
 - <https://coms1016.barnard.edu/>
- Slack:
 - <https://bc-coms-1016-fallb.slack.com/>
- Zoom link:
 - Same for lecture and labs



▼ Channels

- # announcements
- # find-a-partner
- # homeworks
- # in-person-office-h...
- # labs
- # office-hours
- # projects
- # random



Slack - Announcements

A screenshot of a Slack interface showing a list of channels. The channel '# announcements' is highlighted with a red rectangular border. The other channels listed are: find-a-partner, homeworks, in-person-office-h..., labs, office-hours, projects, and random.

- ▼ Channels
- # announcements
- # find-a-partner
- # homeworks
- # in-person-office-h...
- # labs
- # office-hours
- # projects
- # random

- course staff post course wide announcements
- Do not post here
- Encouraged to reply to posts that we create there



- ▼ Channels
 - # announcements
 - # find-a-partner
 - # homeworks
 - # in-person-office-h...
 - # labs
 - # office-hours
 - # projects
 - # random

- Use this channel to find partners
- Different parts of course can be completed in pairs

Slack – Homeworks/Labs/Projects



▼ Channels

announcements

find-a-partner

homeworks

in-person-office-h...

labs

office-hours

projects

random

- Ask questions when working on homework, labs, and projects
- **Do not post solutions**



Slack – Office-Hours

▼ Channels

- # announcements
- # find-a-partner
- # homeworks
- # in-person-office-h...
- # labs
- # office-hours
- # projects
- # random

- Changes to Office Hours will be posted here
- Ask questions about Office Hours posted here



Slack – In-person-office-hours

▼ Channels

announcements

find-a-partner

homeworks

in-person-office-h..

labs

office-hours

projects

random

- Potential in-person office hours



- Live classes
 - Primarily lectures
 - Discussions and exercises about course material
 - Q/A
 - Recorded
- Pre-class readings:
 - Expected to read the assigned reading(s) before class
 - Distributed on course schedule

Assignments

BARNARD COLLEGE OF COLUMBIA UNIVERSITY

Learn By Doing

Assignments



- Labs
- Homework
- Projects
- Midterm
- Final Project



- Complete it by deadline (Wednesdays/Fridays)
- Work in pairs
 - First week paired randomly during lab
- Submit on Gradescope

Homeworks



- Twice a week
 - Released: Monday and Thursday
 - Due: Thursday and Monday
- 12 through out the semester
- Complete individually
- Generous late day policy

Projects



- Similar to HW but a bit longer
- 1 week to complete
- Can be done in pairs
- 2 during the semester

Midterm



- Take-home
- Open notes/open book
- Roughly a week to complete

Final Project



- Open-ended assignment
- Choose from a few datasets
- Develop a question to ask about the dataset
- Deliverables:
 - Analysis proposal
 - Due Friday 12/11
 - Describe the analysis in a report
 - Due Wednesday 12/23

Grading – High Level Overview



- 70% from assignments
- 25% from midterm & quizzes
- 5% participation



Grading – More Details

Participation	5%
Weekly HW	25%
Projects	20%
Midterm + daily quizzes	25%
Final Project	25%



Participation

- During class meetings:
 - Topic discussion
 - Asking questions
- Asynchronous
 - Active on Slack
 - Response questions prior to lectures (daily quizzes)

Assignment Logistics



- Distribution:
 - <https://coms1016.barnard.edu/schedule.html>
- JupyterHub:
 - <http://jupyterhub.coms1016.barnard.edu/>
- Gradescope



Course Staff



Adam Poliak (apoliak@barnard.edu)

- PhD in Computer Science from Johns Hopkins University
- First semester at Barnard
- Research:
 - Natural Language Processing
 - Data Science applied to text data

Course staff



Priya Rajbabu – TA
prajbabu@barnard.edu

- Last semester Master's student in Data Science
- Data Science at Morgan Stanley





Francesca Loiodice

- Barnard CS '22
- Computing Fellow

Lucie le Blanc

- Barnard CS' 21
- Computing Fellow

Susu Rawwagah

- Barnard Political Science '21
- Preceptor



Our job is to help
you succeed!



Office Hours

7 hours of office hours a week

Monday	9pm-10pm
Tuesday	9pm-10pm
Thursday	9pm-10pm
Wednesday	12pm - 1pm
Friday	11am - 12pm
Wednesday	7 - 8pm
Monday	11am - 11:30am
Wednesday	11am - 11:30am



12 hours a week

Tuesday

10am – 12pm
6pm – 8pm

Thursday

10am – 12pm
3pm – 5pm
7pm – 9pm

Friday

1pm – 3pm

A black and white photograph of the exterior of a classical-style building, likely Barnard College. The building features large, fluted Corinthian columns supporting a prominent entablature. The word "BARNARD" is inscribed in capital letters along the top edge of the entablature. The sky is clear and blue.

Course Policies

Collaboration



- Encouraged to discuss problems
- Do not share solutions

Late Days & Dropped Assignments



- 10 Late Days for homeworks and projects
 - Can only use 2 per assignment
- Drop 2 lowest homeworks & 2 lowest labs



—

Learn By Doing

—



Cause & Effect

BARNARD COLLEGE OF COLUMBIA UNIVERSITY



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



EATING AND HEALTH

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

June 19, 2015 · 5:03 AM ET

Heard on [Morning Edition](#)



ALLISON AUBREY



Observation





Observation

- **individuals**, study subjects, participants, units
 - European adults



Observation

- **individuals**, study subjects, participants, units
 - European adults
- **Treatment**
 - *Chocolate through out the day*



- **individuals**, study subjects, participants, units
 - European adults
- **Treatment**
 - *Chocolate through out the day*
- **outcome**
 - *heart disease*



Question 1: Association

- Is there **any relation** between consuming chocolate and heart disease?
 - **association**
 - any relation
 - Three coffees a day **linked** to improve health



Question 1: Association

- Is there **any relation** between consuming chocolate and heart disease?
- **association**
 - any relation
 - Three coffees a day **linked** to improve health



Look at 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



Look at 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

- *Does this point to an association?*



Question 2: Causation

- Does eating chocolate **lead to** reduced heart disease?
 - **Causality**

Question 2: Causation



- Does eating chocolate **lead to** reduced heart disease?
 - **Causality**
- Causality 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



— Association



King Cholera – London 1850's



A COURT FOR KING CHOLERA.

Miasma, miasmatism, miasmatists



- **Bad smells** given off by waste & rotting matter

Miasma, miasmatism, miasmatists



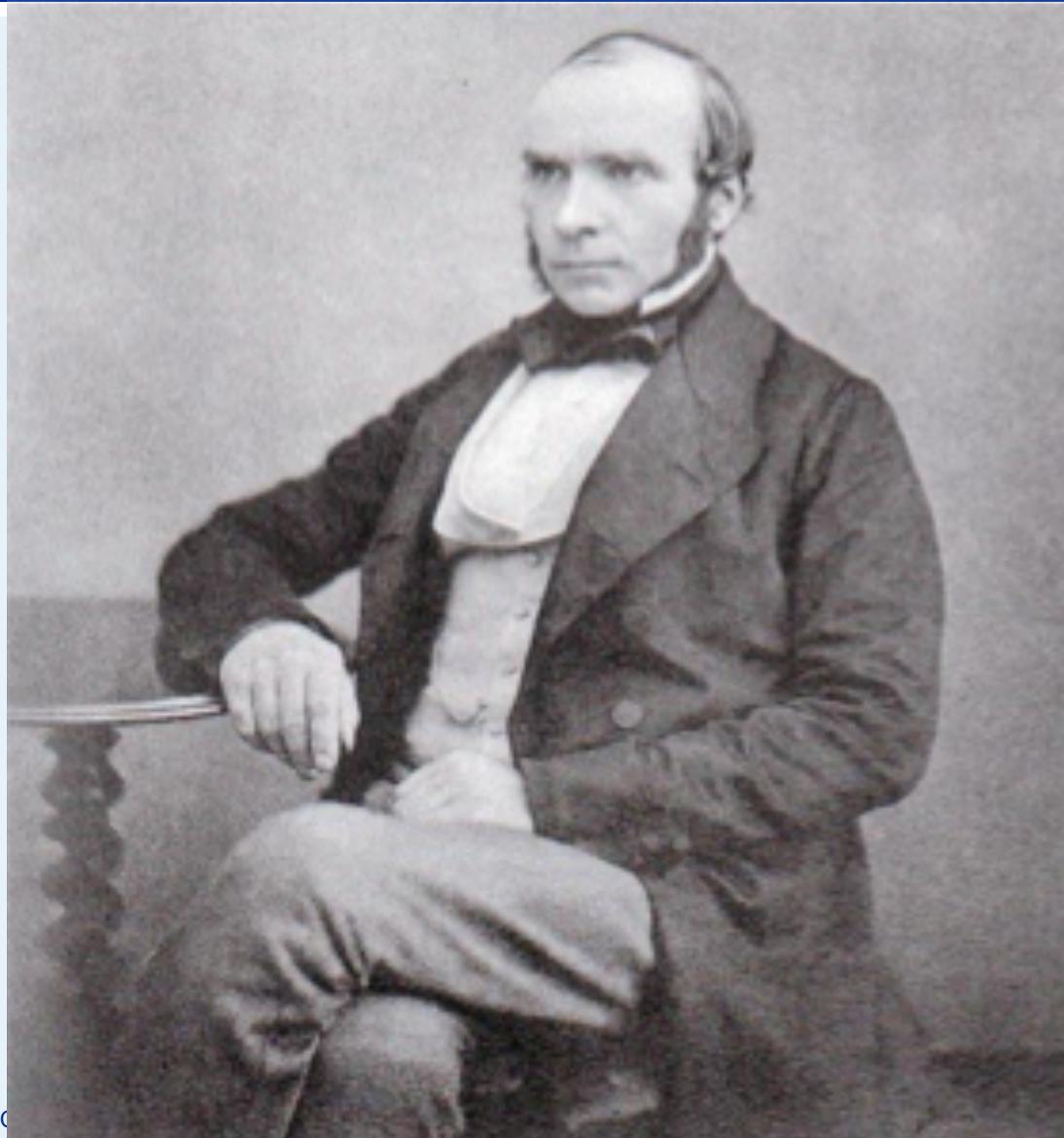
- **Bad smells** given off by waste & rotting matter
- Potential remedies:
 - “fly to clean air”
 - “a pocket full of posies”
 - “fire off barrels of gunpowder”

Miasma, miasmatism, miasmatists



- **Bad smells** given off by waste & rotting matter
- Potential remedies:
 - “fly to clean air”
 - “a pocket full of posies”
 - “fire off barrels of gunpowder”
- Popular miasmatists
 - Florence Nightingale (founder of modern nursing)
 - Edwin Chadwick (Commissioner of the Board of Health)

John Snow, 1813 - 1858





Mapping the disease



Broad Street

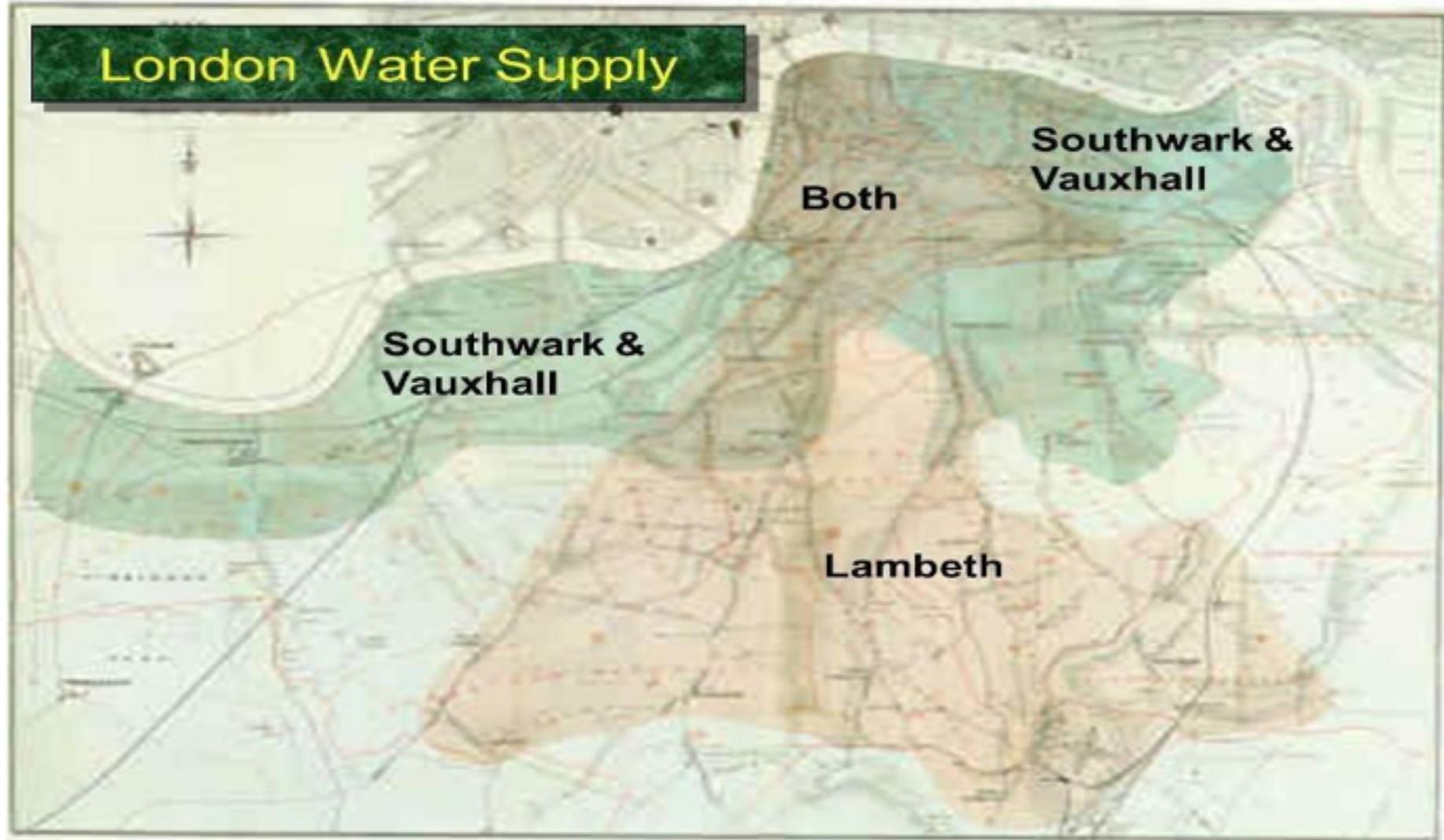




Causation



London Water Supply





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 ...”
- Two groups different **only in the treatment**



Snow's table

Supply Area	Number of houses	Cholera deaths	Deaths per 10,000 houses
S&V	40,046	1,263	315
Lamberth	26,107	98	37
Rest of London	256,423	1,422	59

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

A dark, high-contrast statue of a person in a dynamic pose, possibly a runner or dancer, set against a bright, overexposed background. The statue is rendered in a dark, almost black, color, while the background is a bright, washed-out blue.

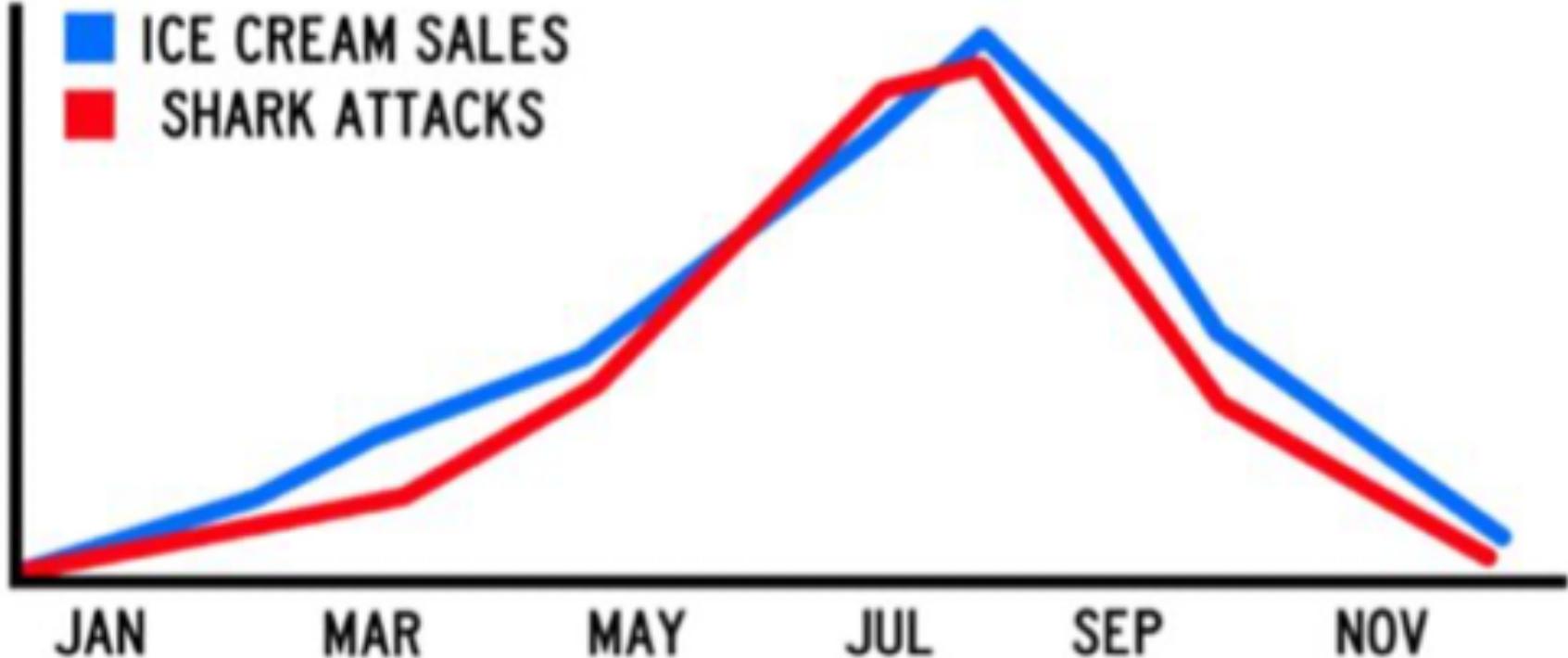
Confounding



- If the treatment and control groups have **systemic differences other than the treatment**, then it might be difficult to identify causality
- Such differences are often present in **observational studies**
- When these differences lead researchers astray, they are called **confounding factors**



Confounding Factor: Example



Randomize!



- If you assign individuals to treatment and control groups **at random**, the two groups are likely to be similar apart from the treatment
- You can account (mathematically) for variability in the assignment
- **Randomized Controlled Experiment**