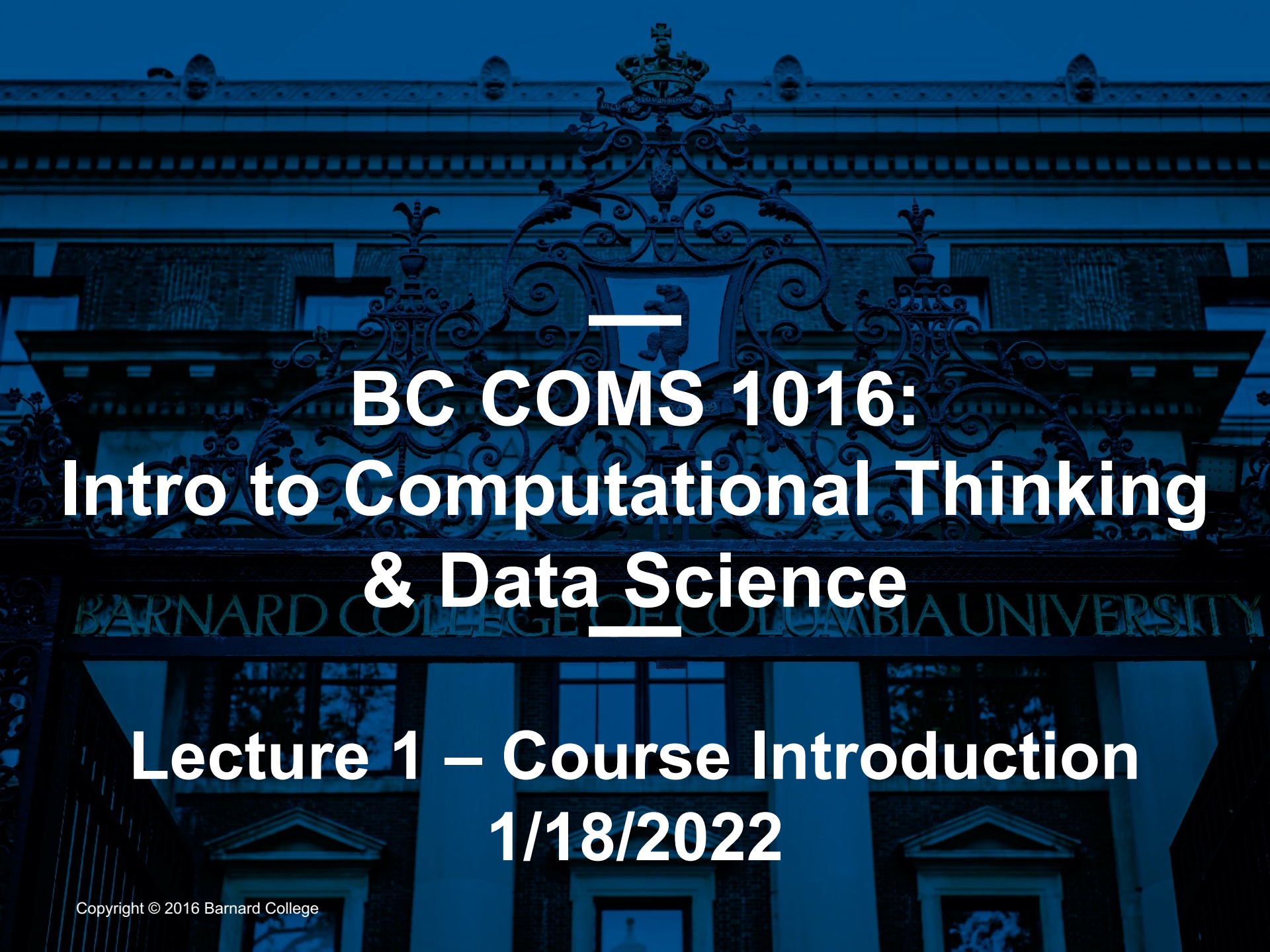


BC COMS 1016: Intro to Computational Thinking & Data Science

Lecture 1 – Course Introduction 1/18/2022

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

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

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

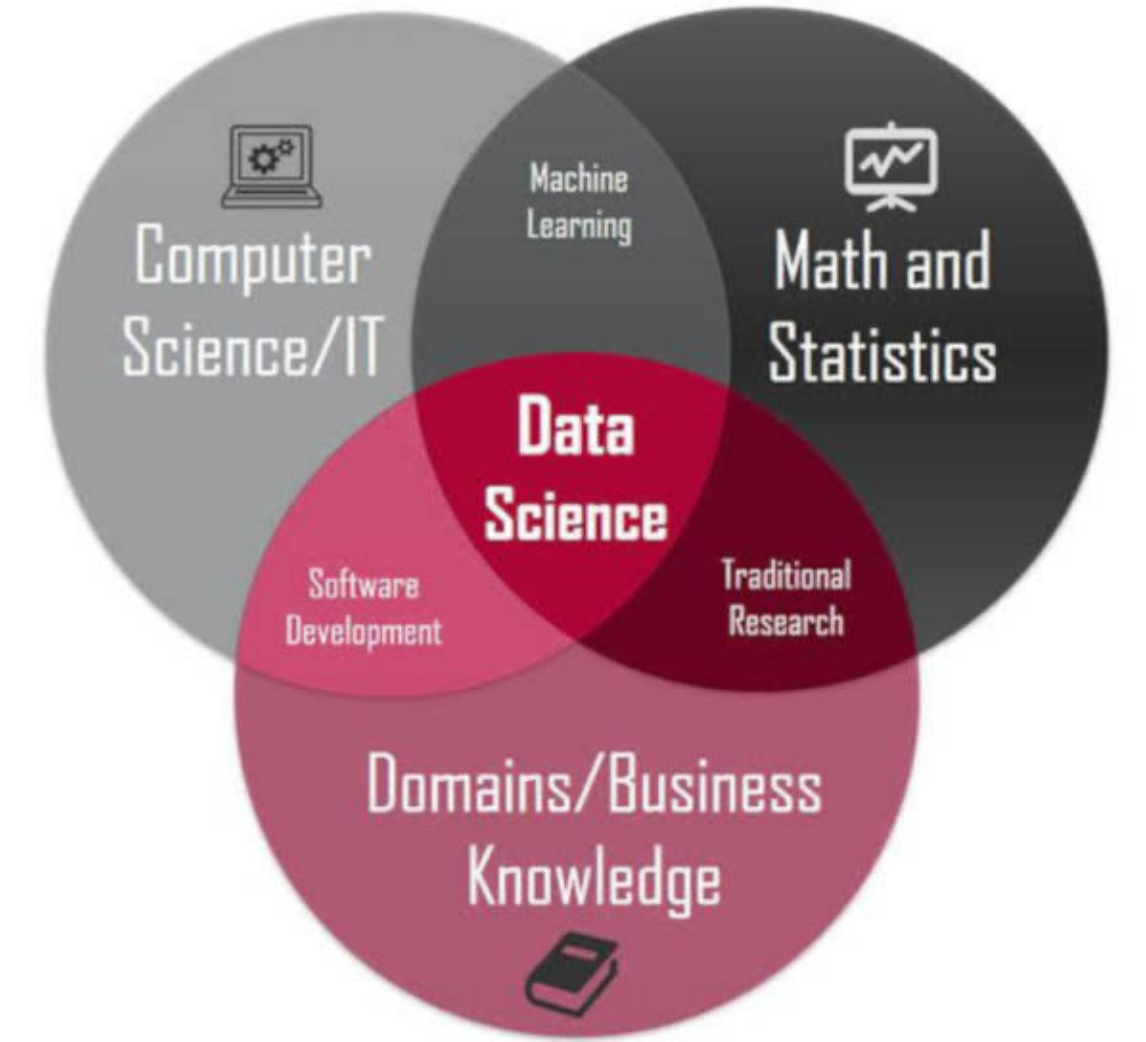
- Value
  - Requires domain expertise to determine what value is
  - *Value from data* is different based on the domain and the needs

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

Course Outline

- Exploration **Week 1 - 5**
  - Discover patterns in data
  - Articulate insights (visualizations)
- Inference **Week 6 - 10**
  - Make reliable conclusions about the world
  - Statistics is useful
- Prediction **Week 11-14**
  - Informed guesses about unseen data

Course Outline

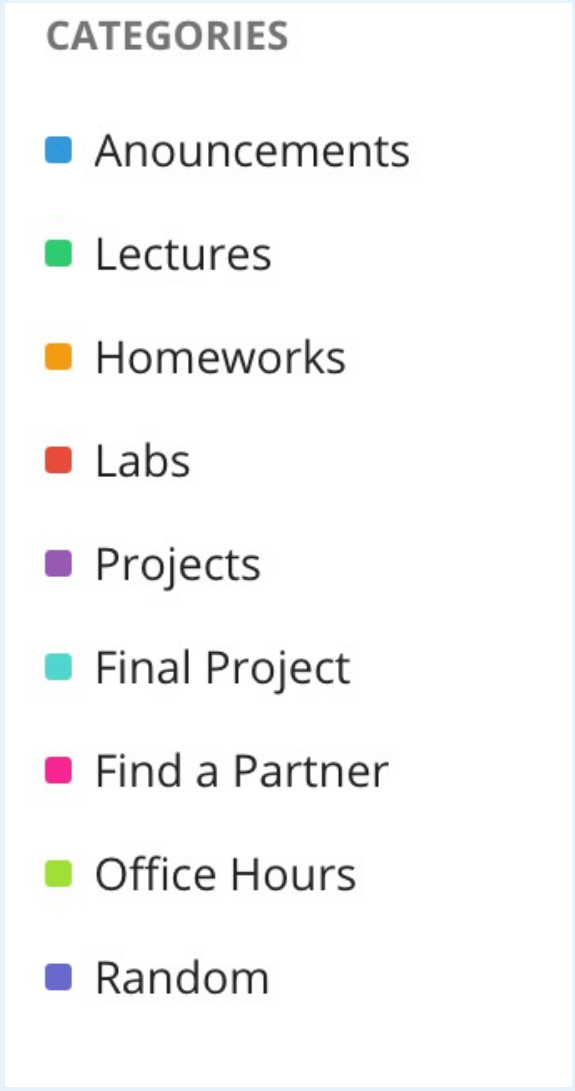
- Exploration **Week 1 - 5**
  - Introduction to Python
  - Working with data
- Inference **Week 6 - 10**
  - Probability
  - Statistics
- Prediction **Week 11-14**
  - Machine Learning
  - Regression & Classification

Logistics

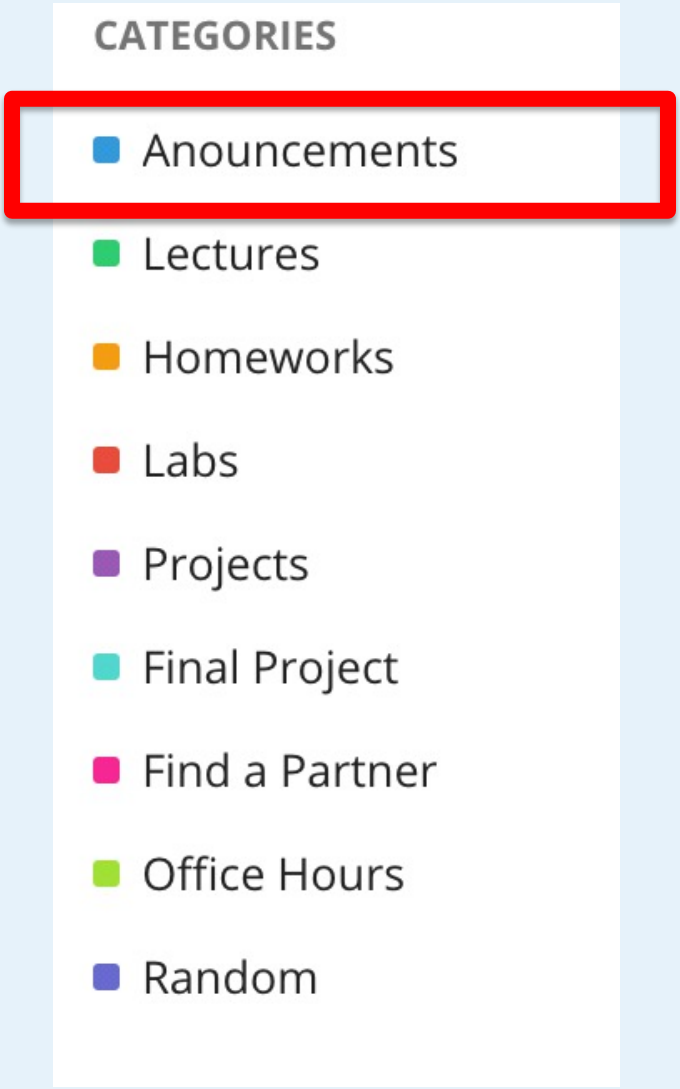
Communication

- Course webpage:
  - https://coms1016.barnard.edu/
- EdStem:
  - https://edstem.org/us/courses/18868
- Zoom link:
  - Same for lecture and labs

EdStem

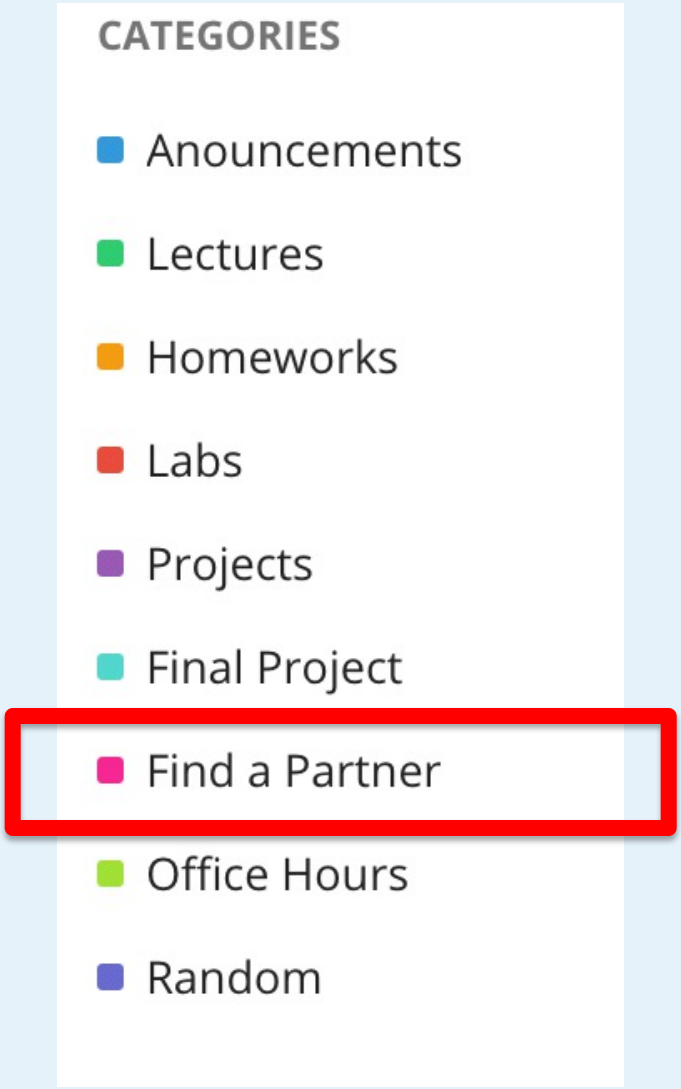


EdStem - Announcements



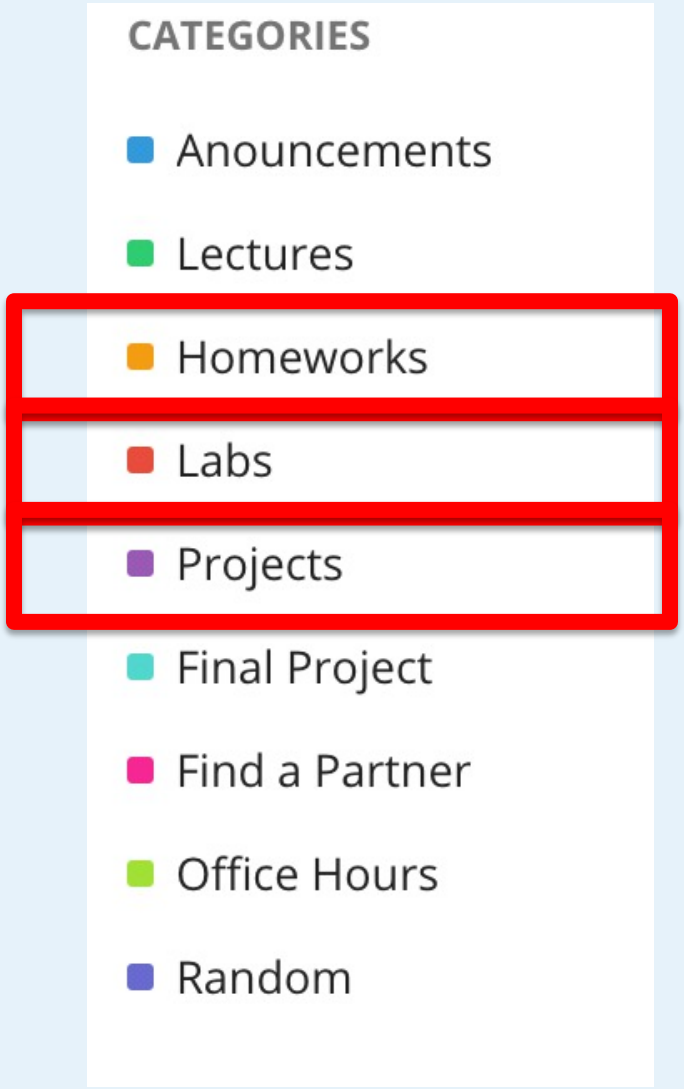
- course staff post course wide announcements
- Do not post here
- Encouraged to reply to posts that we create there

EdStem – Find-a-Partner



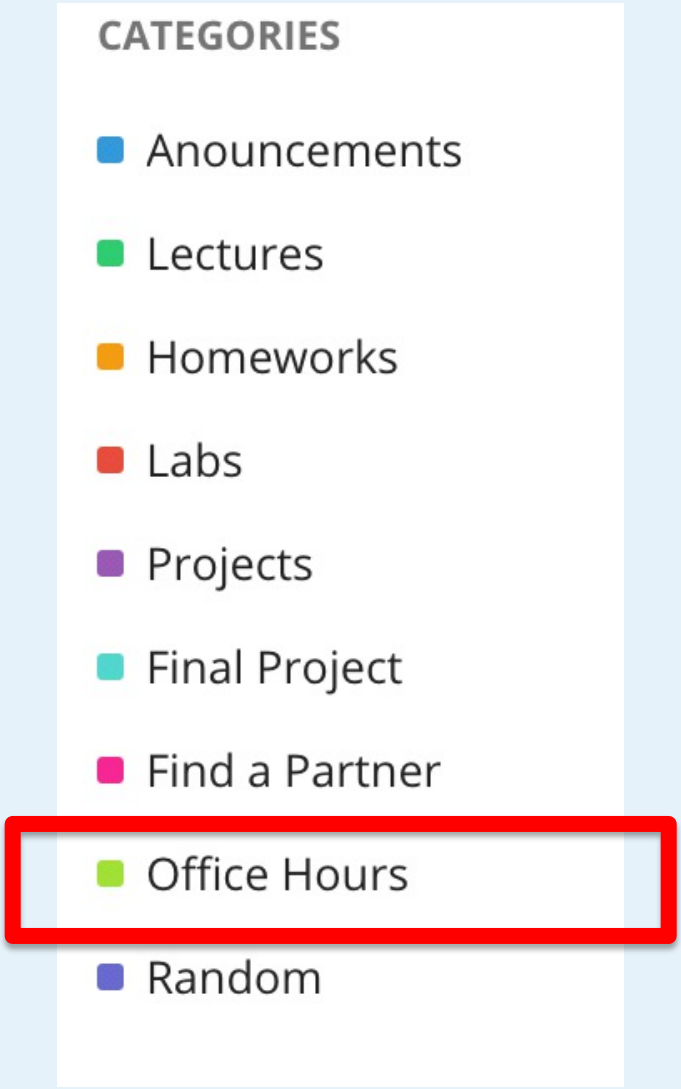
- Use this channel to find partners
- Different parts of course can be completed in pairs

EdStem – Homeworks/Labs/Projects



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

EdStem – Office-Hours

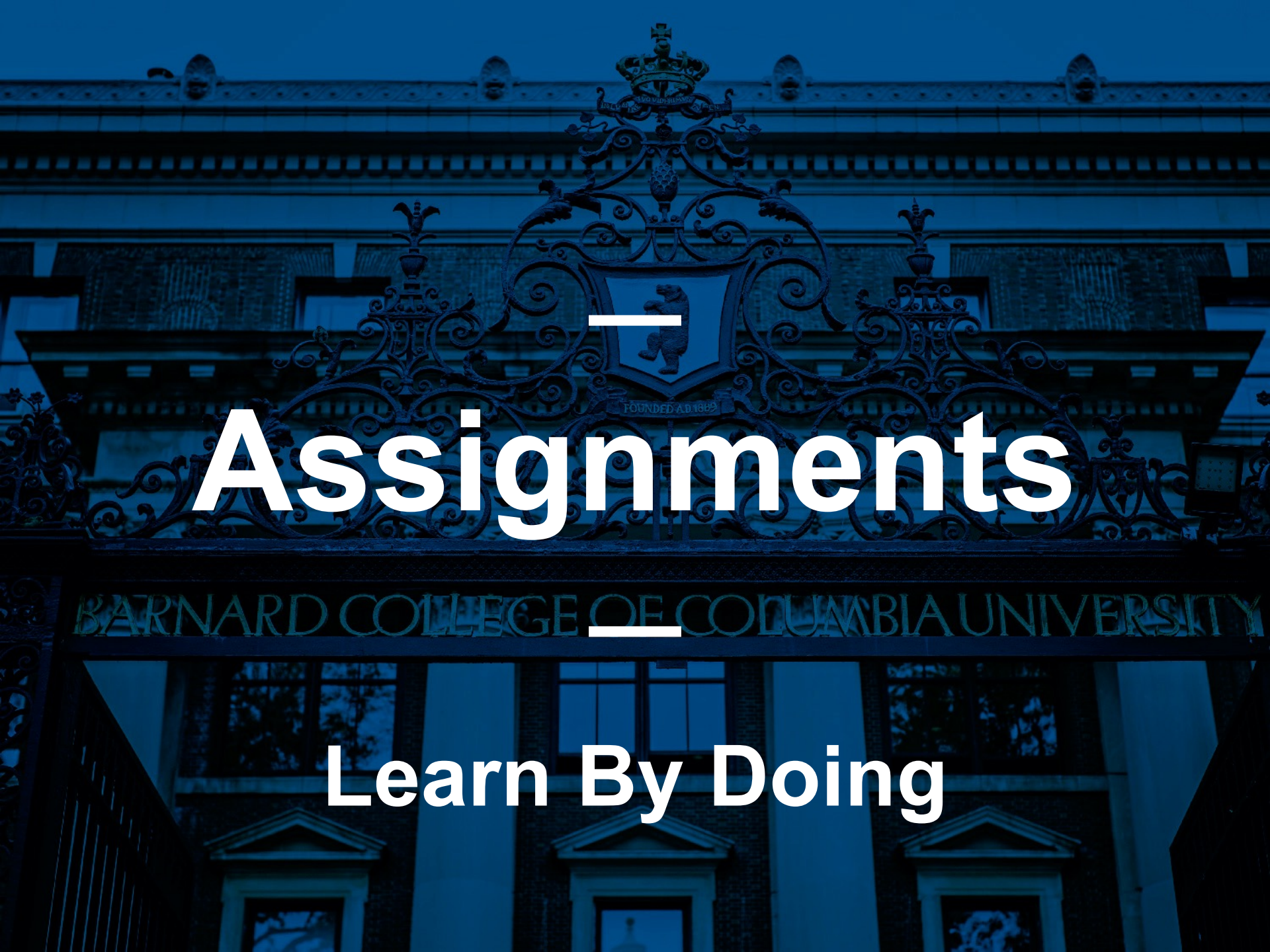


- Changes to Office Hours will be posted here
- Ask questions about Office Hours posted here

Course Meetings: 11:40-12:55 (EST) TR

- 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

Learn By Doing

Assignments

- Labs
- Homework
- Projects
- Midterm
- Final Project

Lab

- Complete it by Monday (11:59pm)
- Work in pairs
  - First week paired randomly during lab
- Submit on Gradescope

Homeworks

- Complete it by Thursday
- 12 through out the semester
- Complete individually
- Generous late day policy

Projects

- Similar to HW but a bit longer
- ~2-3 weeks to complete
- Can be done in pairs
- 3 during the semester

Midterm

- TBD

Final Project

- Open-ended assignment
- Choose from a few datasets
- Develop a question to ask about the dataset
- Deliverables:
  - Analysis proposal
  - Describe the analysis in a report

Grading – More Details

Participation	5%
Weekly HW	25%
Weekly Lab	10%
Projects	20%
Midterm	15%
Final Project	25%

Participation

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

Assignment Logistics

- Distribution:
  - https://coms1016.barnard.edu/schedule.html
- JupyterHub:
  - https://bc-coms-1016-poliak.columbiajupyter2.org/
- Gradescope

Course Staff

Course staff

Adam Poliak (apoliak@barnard.edu)

- PhD in Computer Science from Johns Hopkins University
- 2nd year at Barnard
- Research:
  - Natural Language Processing
  - Data Science applied to text data

Our job is to help you succeed!

Office Hours

- At least 8 a week
- Barnard CS Help Room
- Times will be posted this week

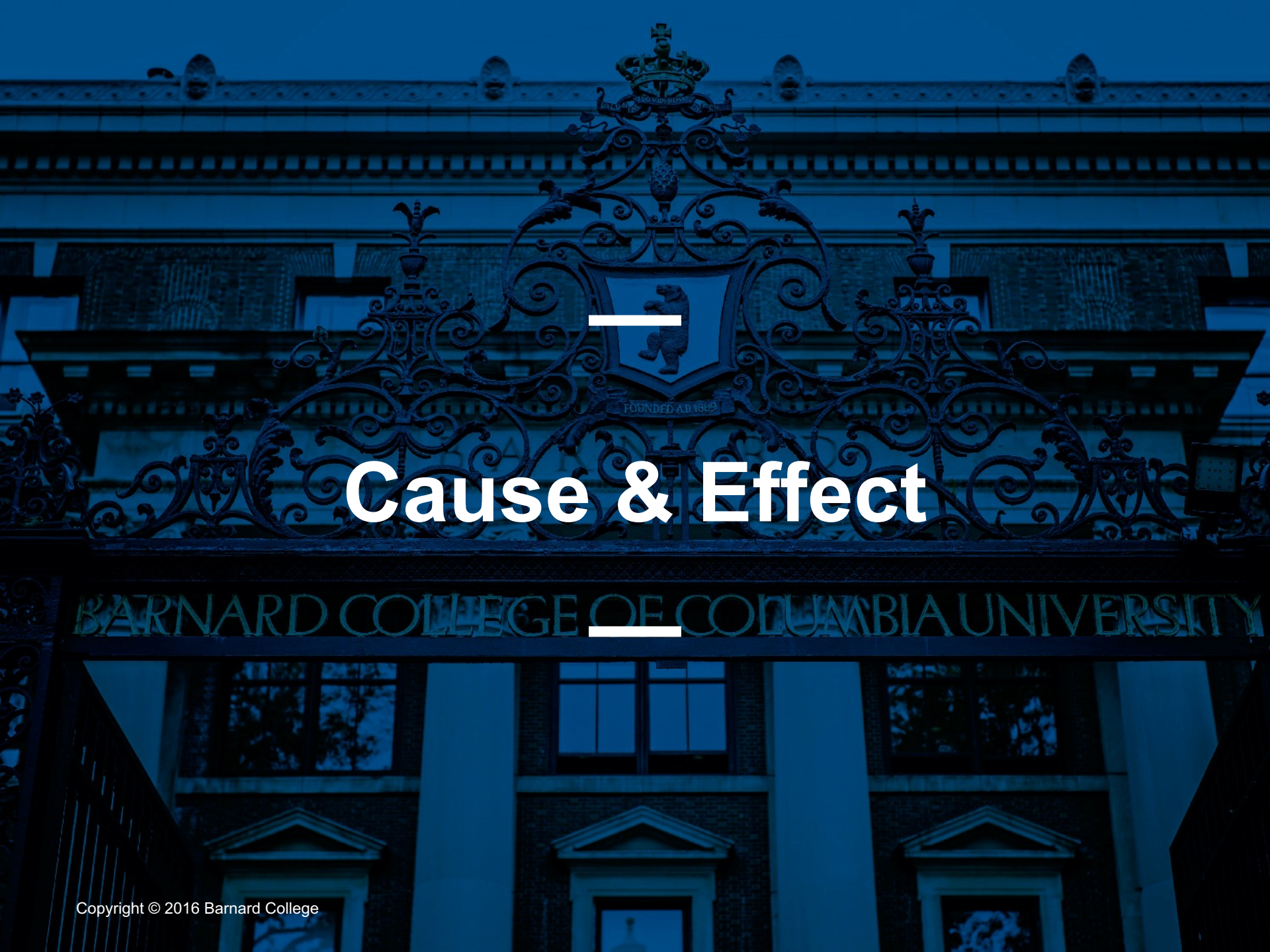
Course Policies

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 1 lowest homework & 1 lowest lab



Cause & Effect

A link (Coffee and Health)

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 link (Chocolate and Health)

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*

Observation

- **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
    - Thee coffees a day **linked** to improve health

Question 1: Association

- Is there **any relation** between consuming chocolate and heart disease?

- **association**
  - any relation
  - Thee coffees a day **linked** to improve health

An Answer

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

An Answer

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

Illustration from *Punch* (1852)

Miasma, miasmatism, miasmatists

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

Miasma, miasmatism, miasmatists

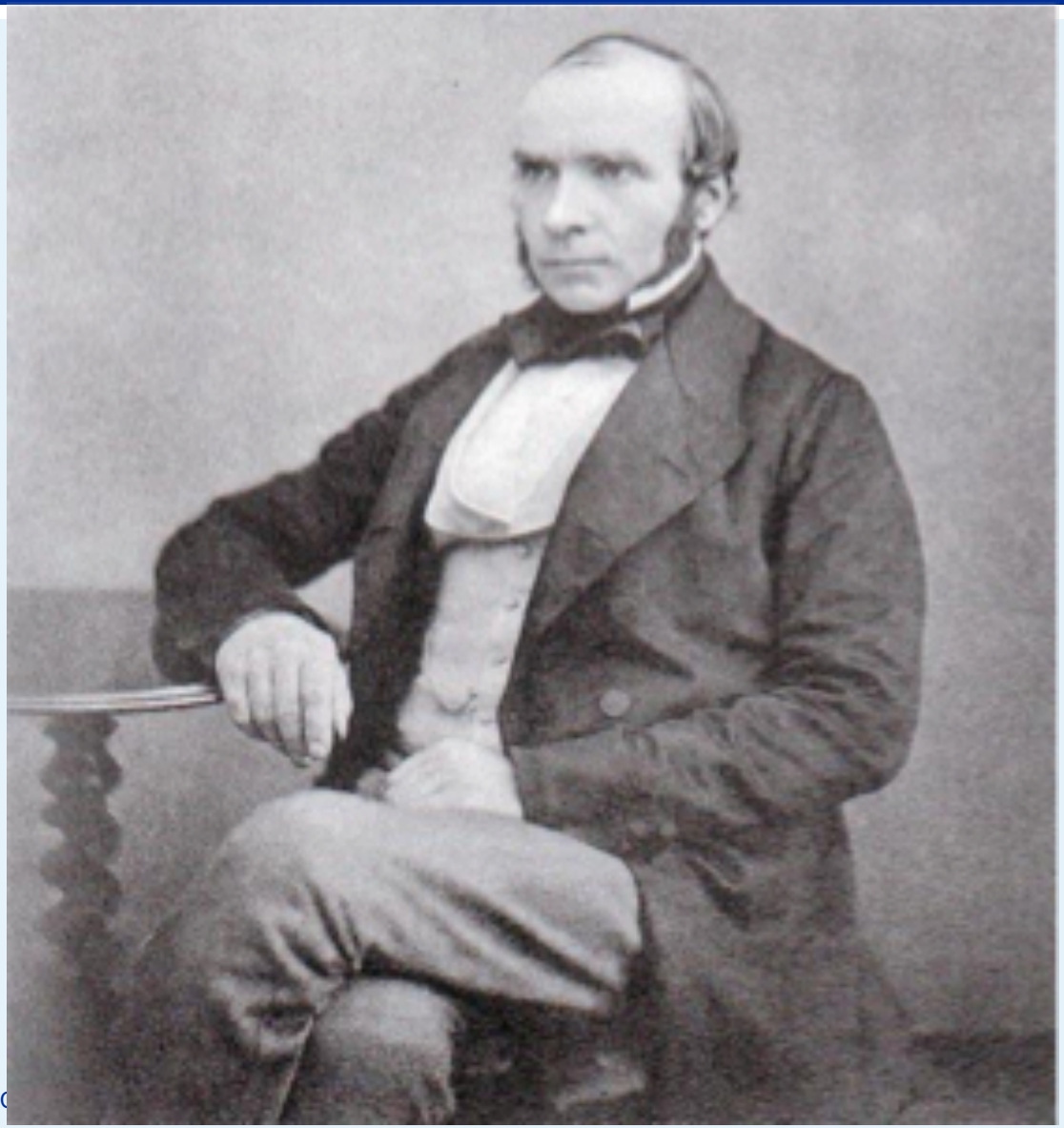
- **Bad smells** given off by waste & rotting matter

- 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)

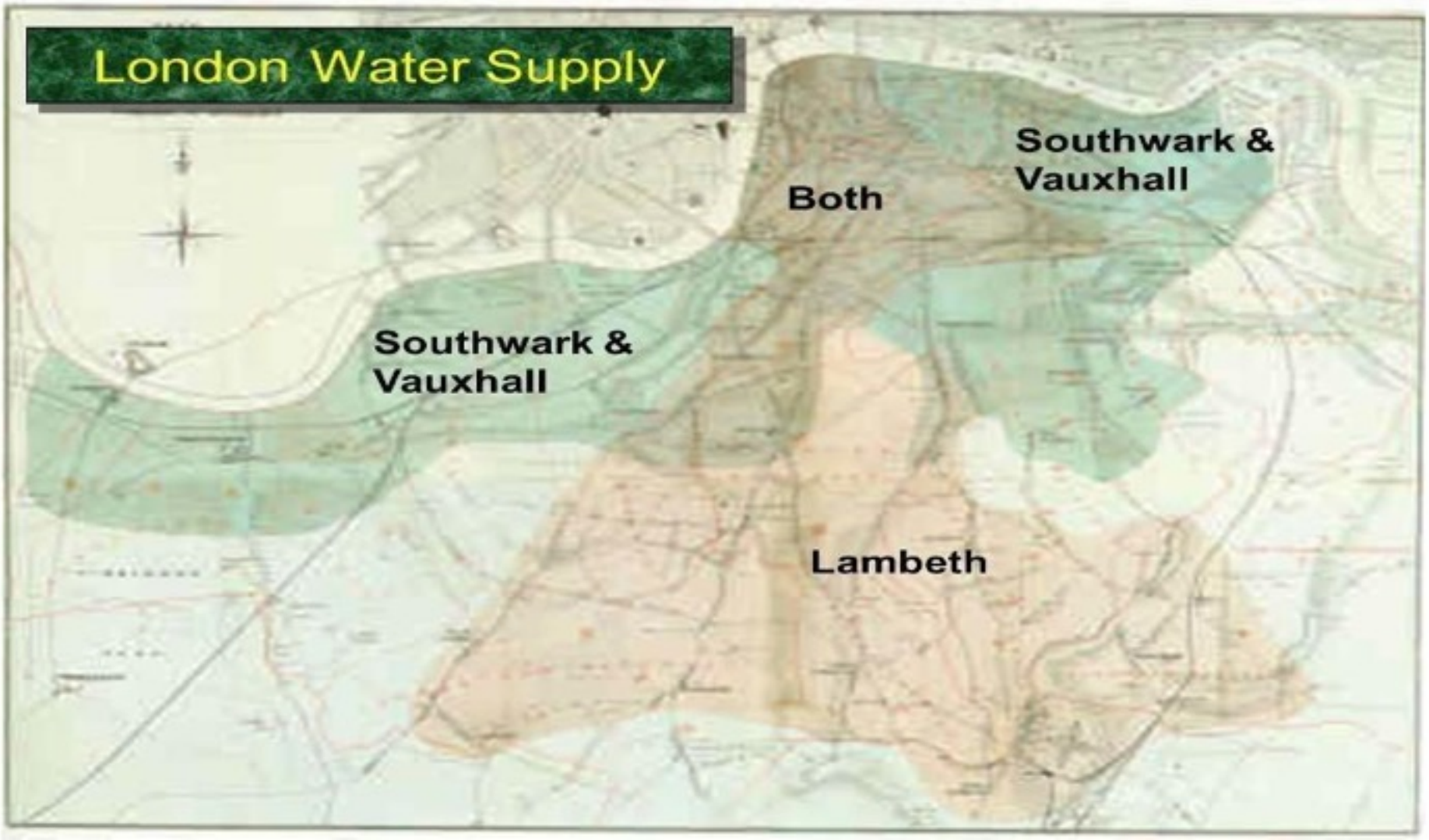
John Snow, 1813 - 1858



Mapping the disease

Broad Street

Causation



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

Confounding

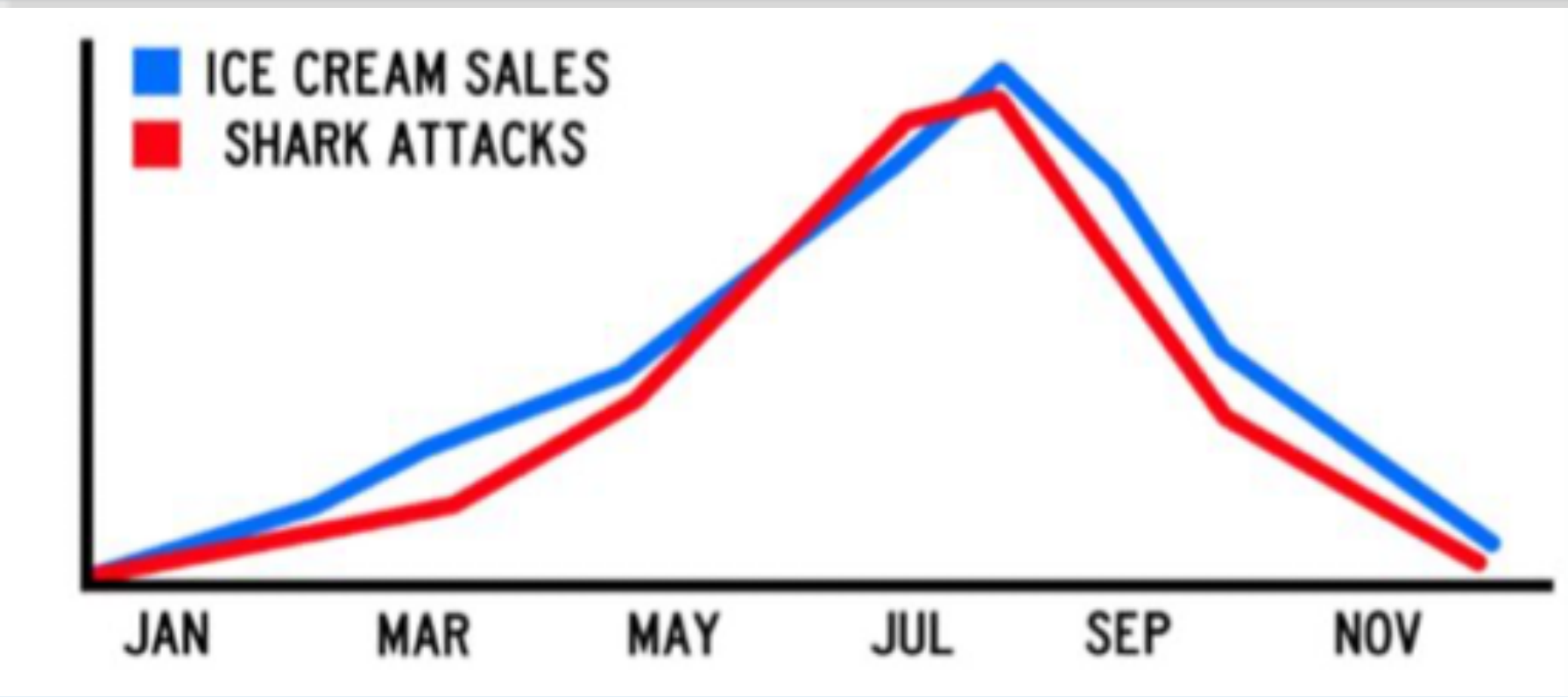
Issue

- 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**