COMS BC1016: Introduction to Computational Thinking and Data Science

Lecture 1: Introduction

Sept. 3, 2025

Copyright © 2026 Barnard College

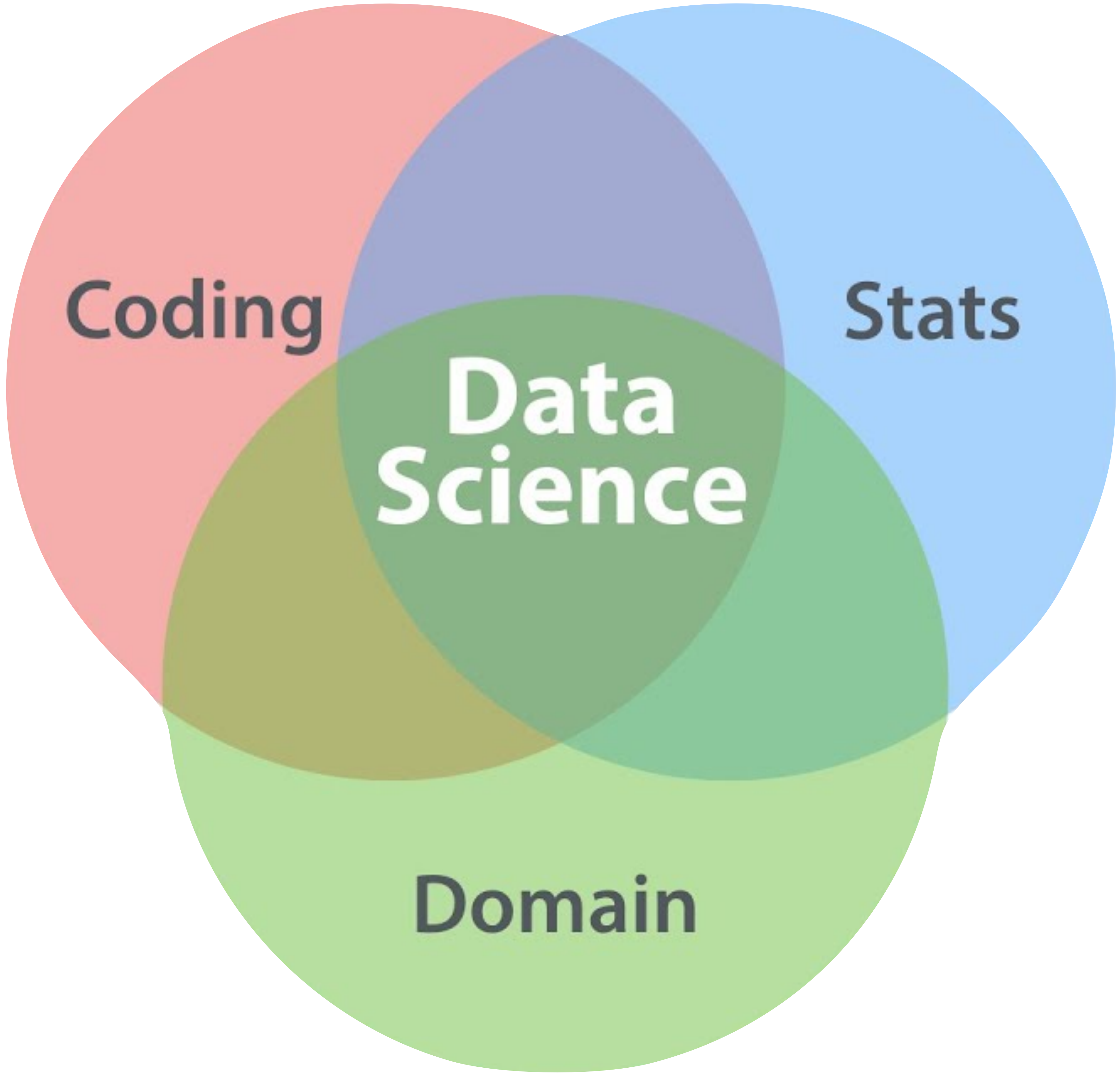
January 21, 2026

Instructor

That's me!

Eysa Lee
eylee@barnard.edu

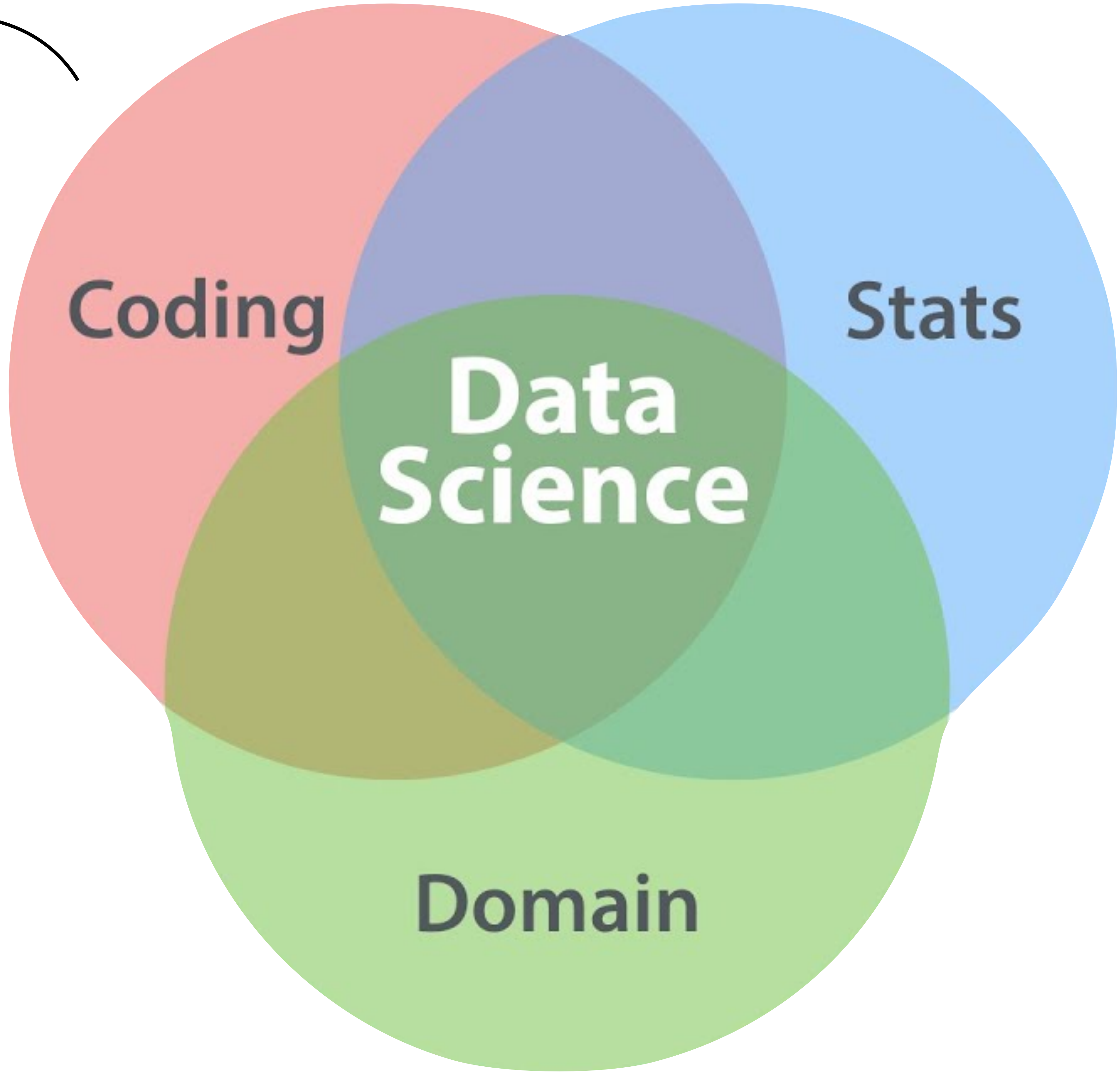
What is Data Science?



What is Data Science?

Prepare &
Analyze Data

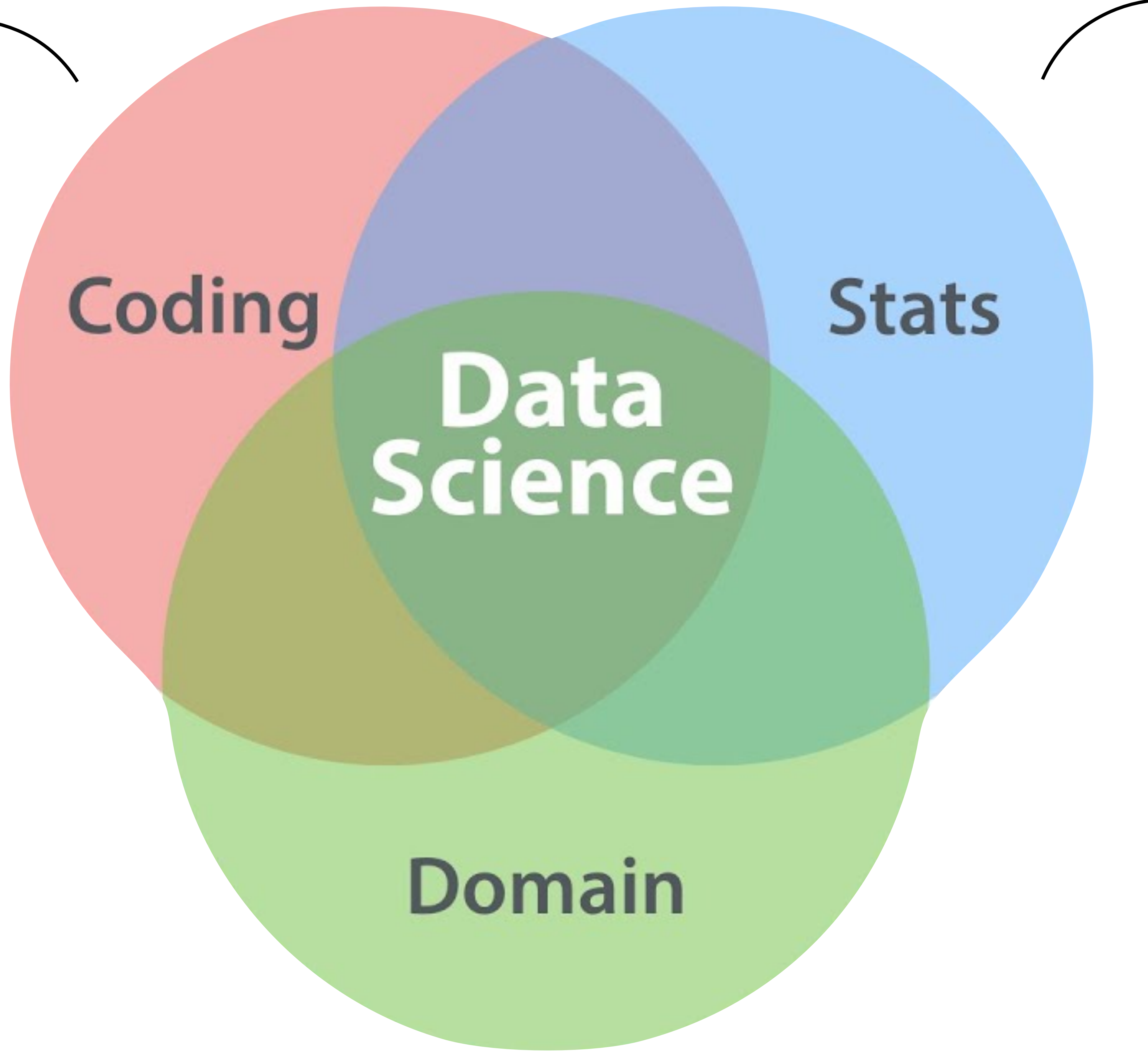
Python &
Computational
Notebooks



What is Data Science?

Prepare &
Analyze Data

Python &
Computational
Notebooks



Probability

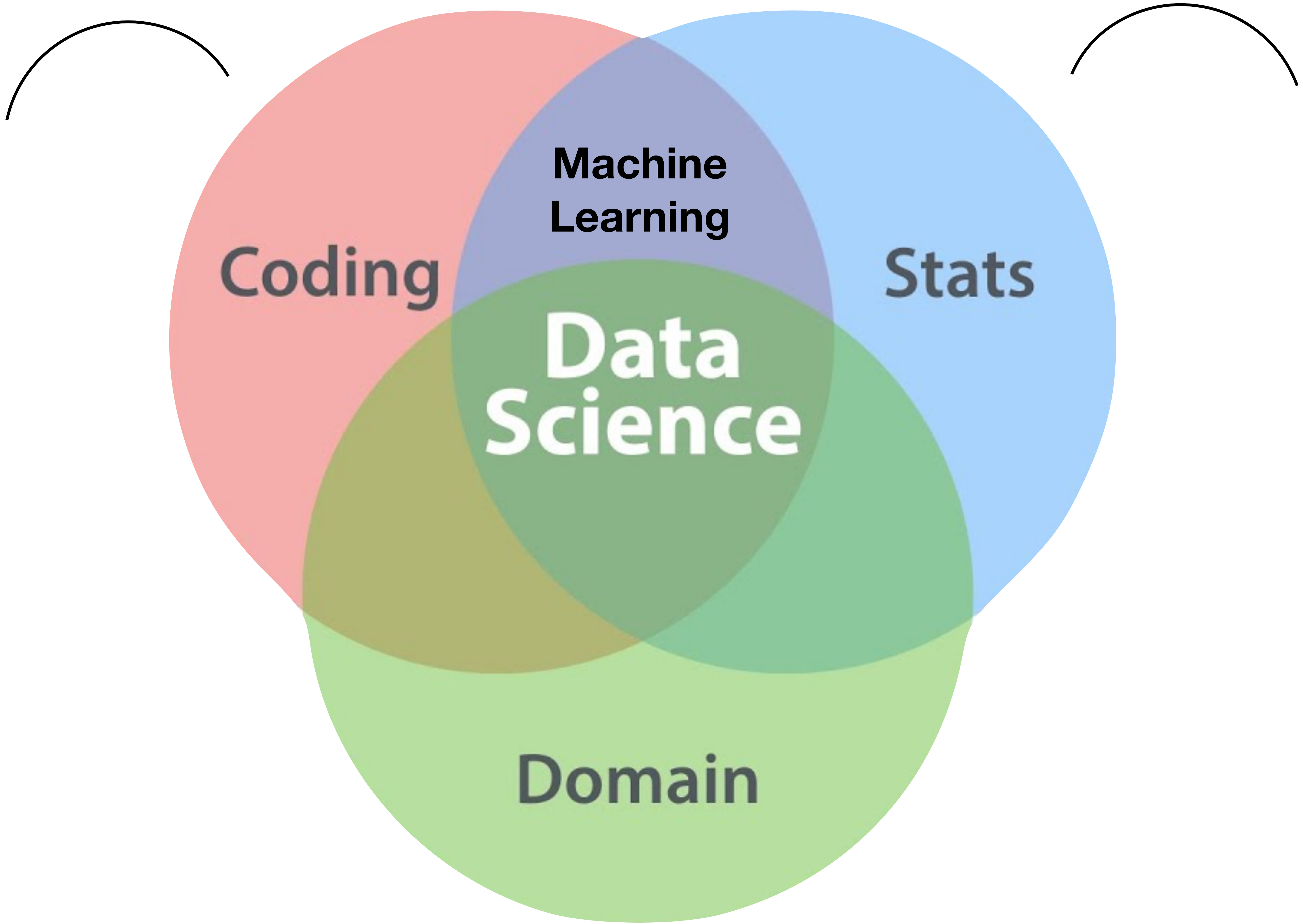
Regression

Confidence

What is Data Science?

Prepare &
Analyze Data

Python &
Computational
Notebooks



Probability

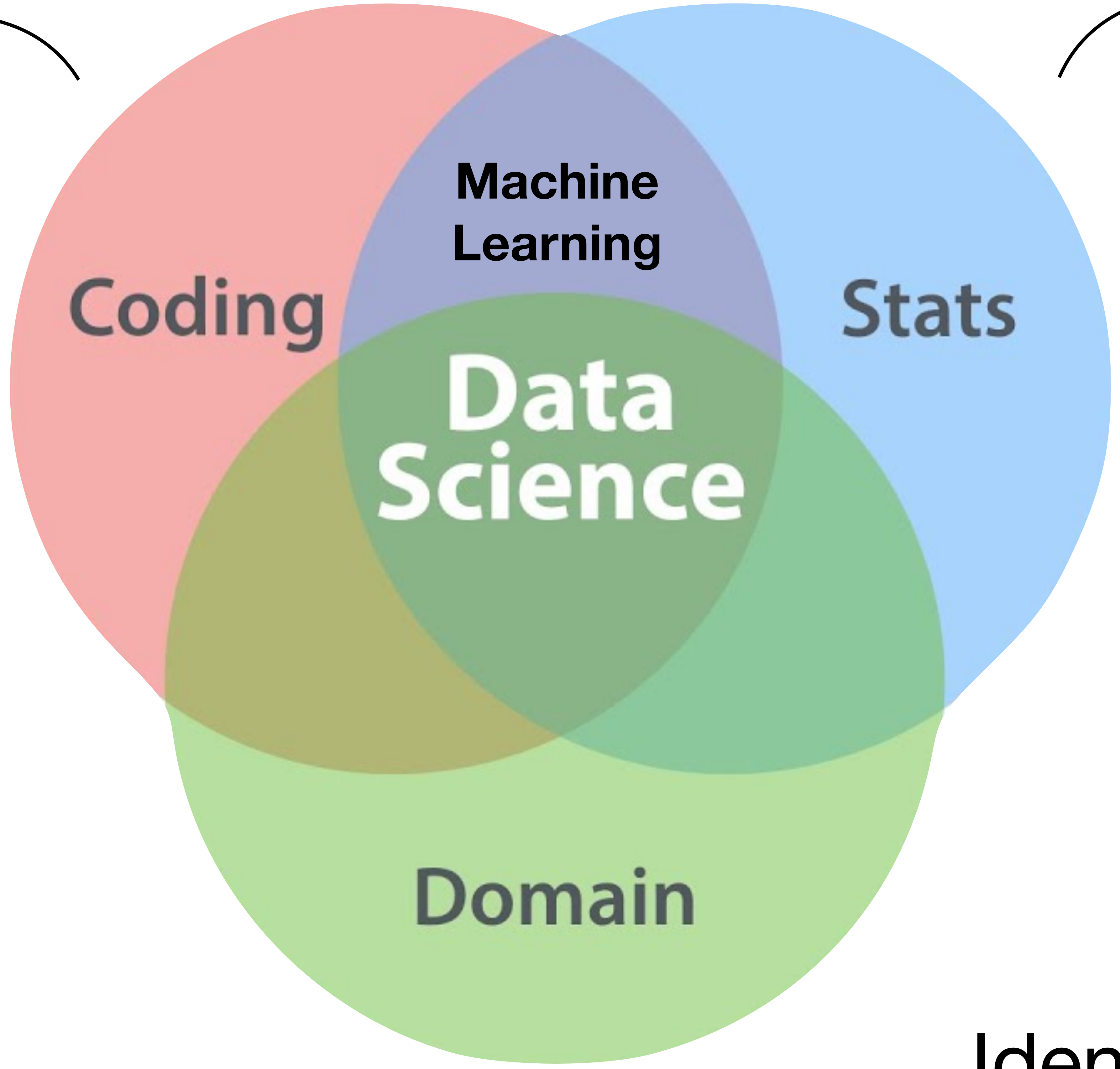
Regression

Confidence

What is Data Science?

Prepare &
Analyze Data

Python &
Computational
Notebooks



Probability

Regression

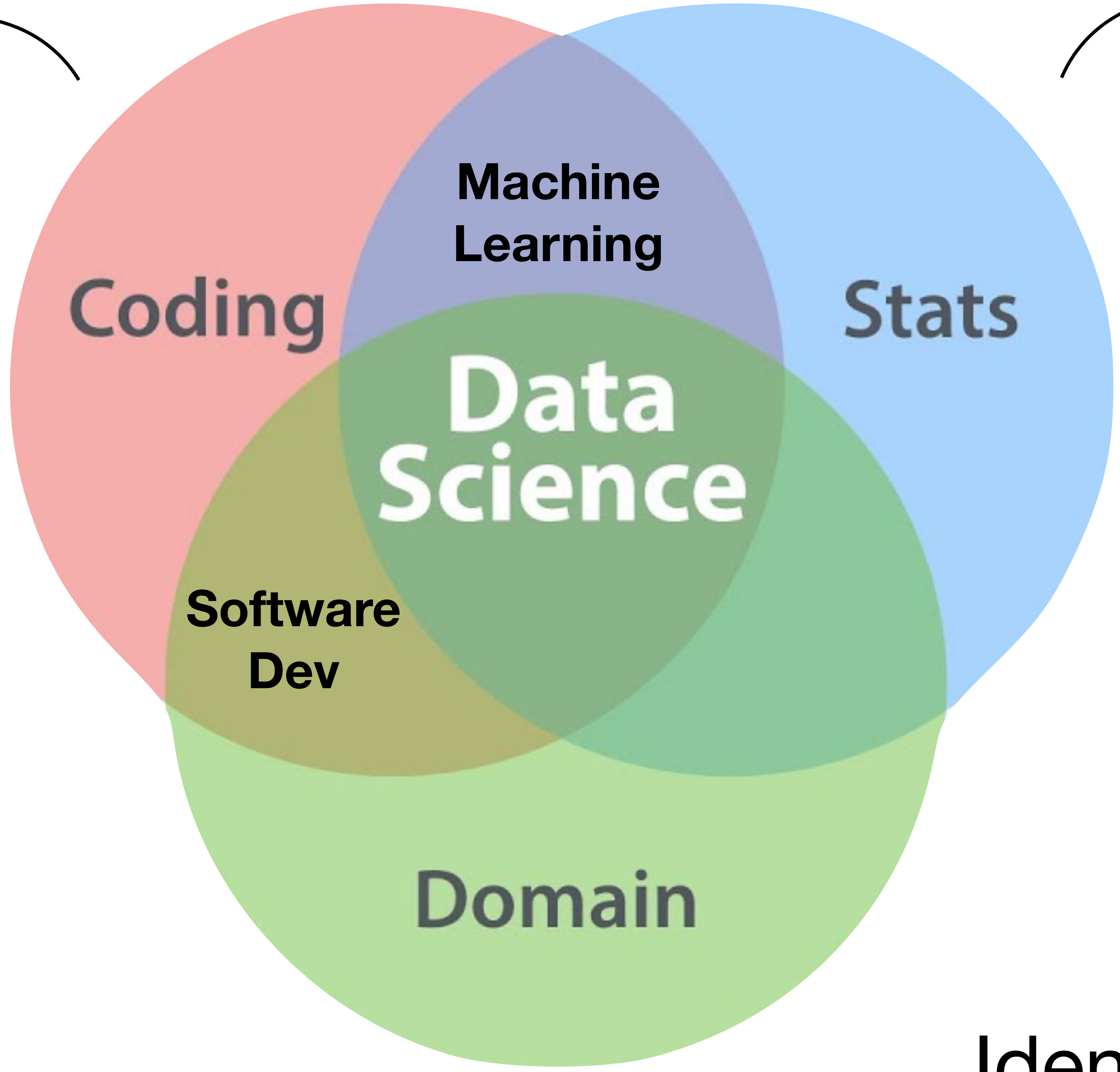
Confidence

Identifying relevant &
meaningful applications

What is Data Science?

Prepare & Analyze Data

Python & Computational Notebooks



Probability

Regression

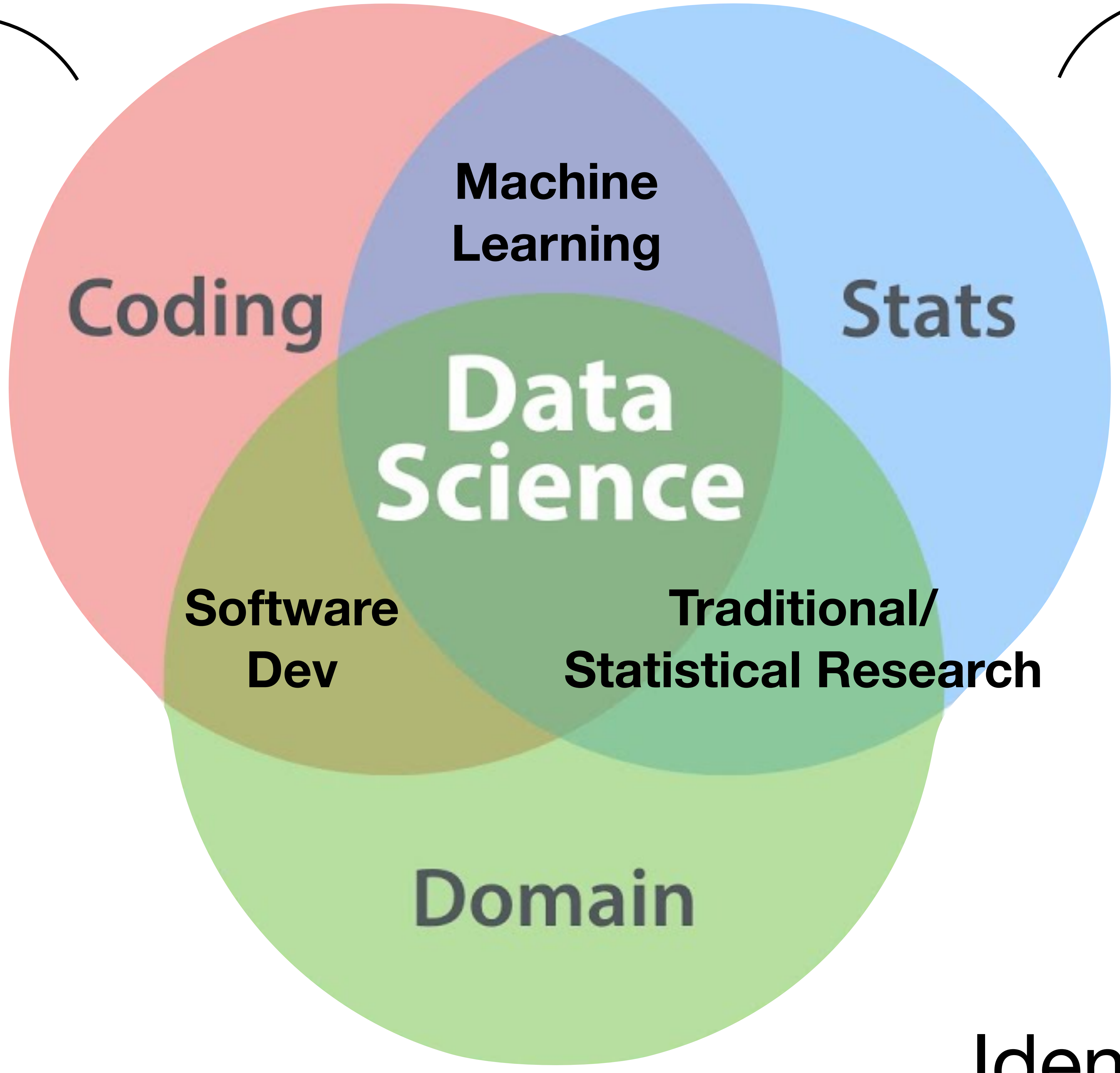
Confidence

Identifying relevant & meaningful applications

What is Data Science?

Prepare &
Analyze Data

Python &
Computational
Notebooks



Probability

Regression

Confidence

Identifying relevant &
meaningful applications

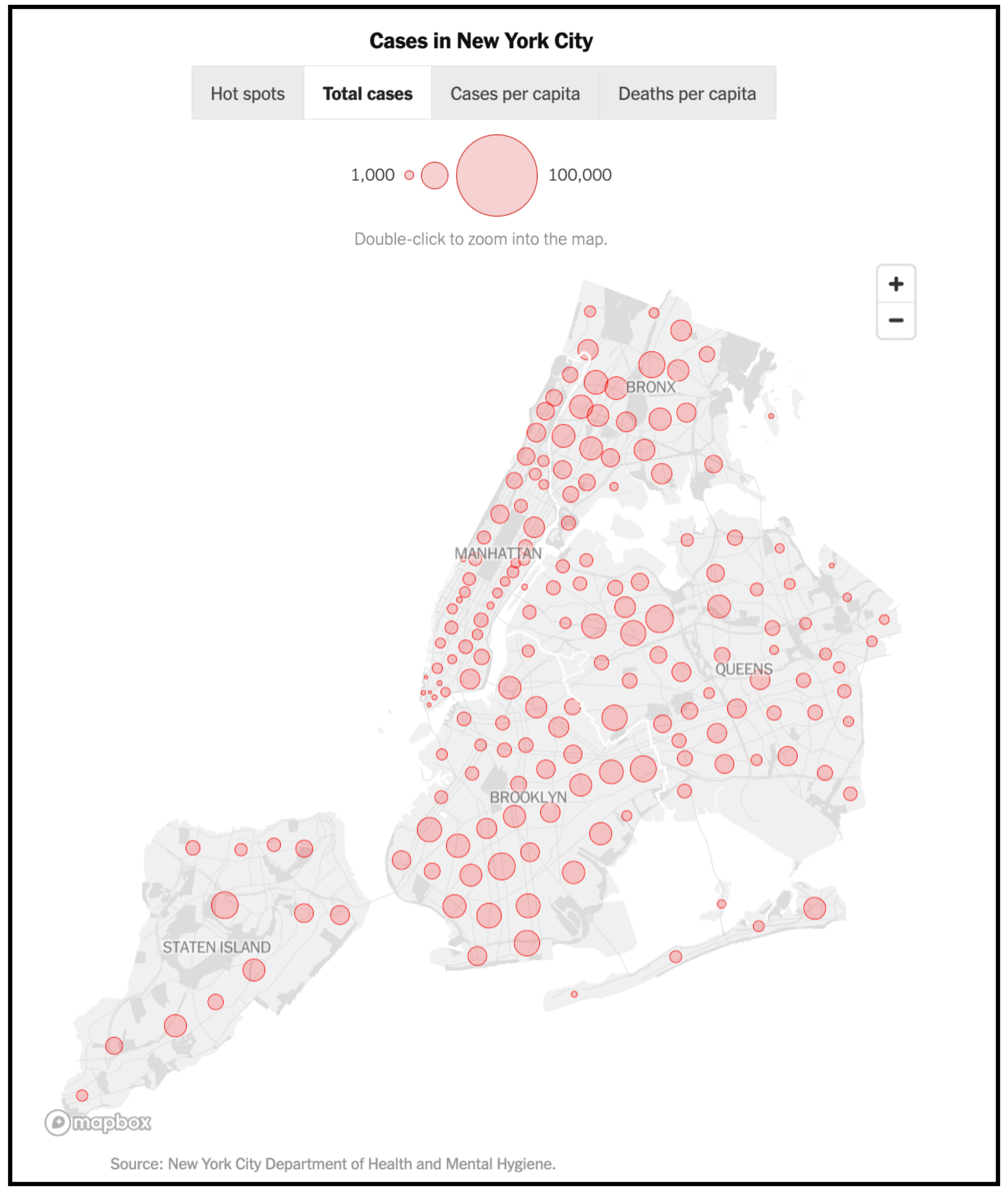
What is Data Science?

Data science is about drawing useful conclusions from large and diverse data sets through…

What is Data Science?

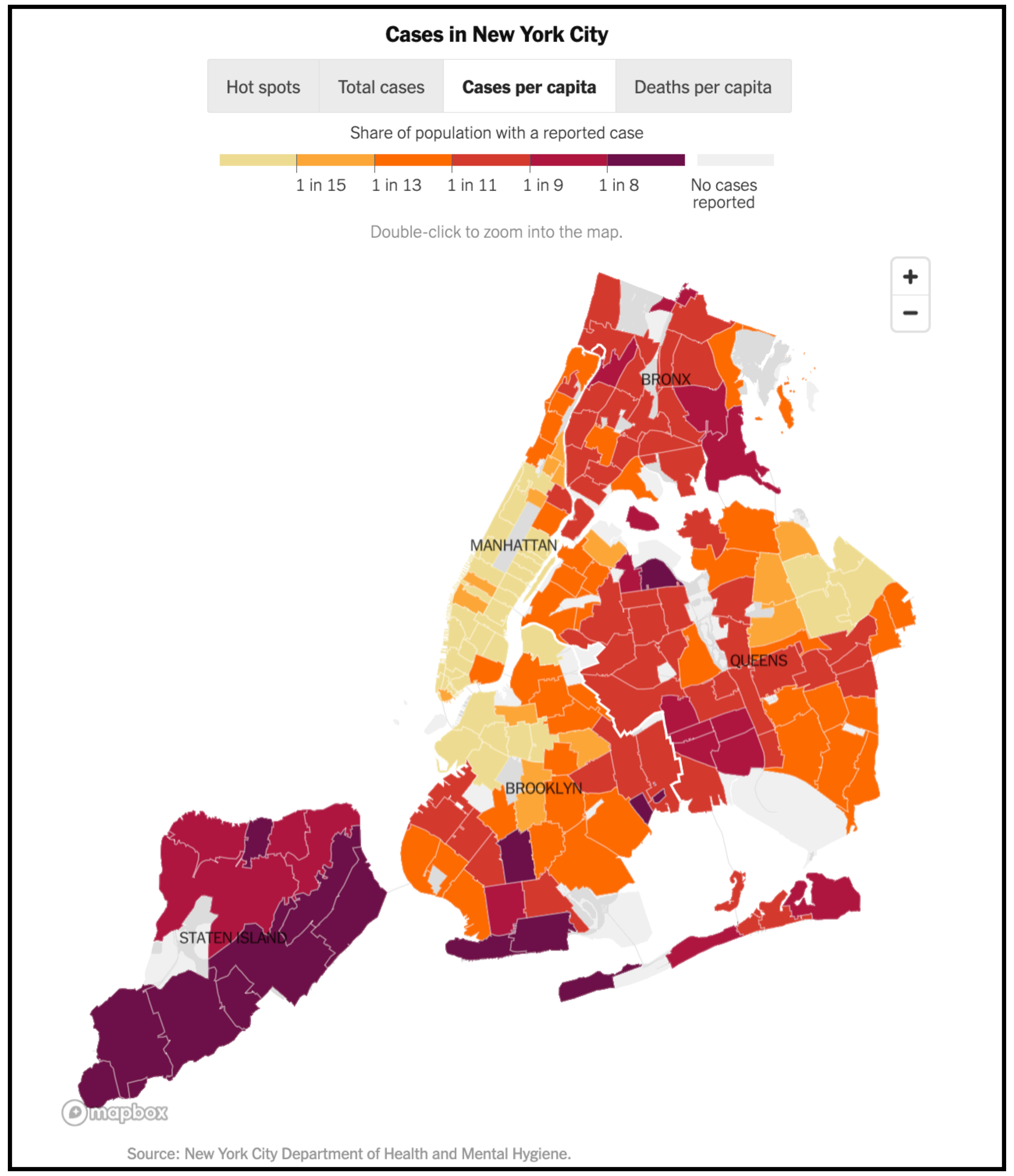
Data science is about drawing useful conclusions from large and diverse data sets through…

- Exploration: Identifying patterns and trends using data (e.g., through visualization)



Cases in New York City

Source: New York City Department of Health and Mental Hygiene.



Cases in New York City

Source: New York City Department of Health and Mental Hygiene.

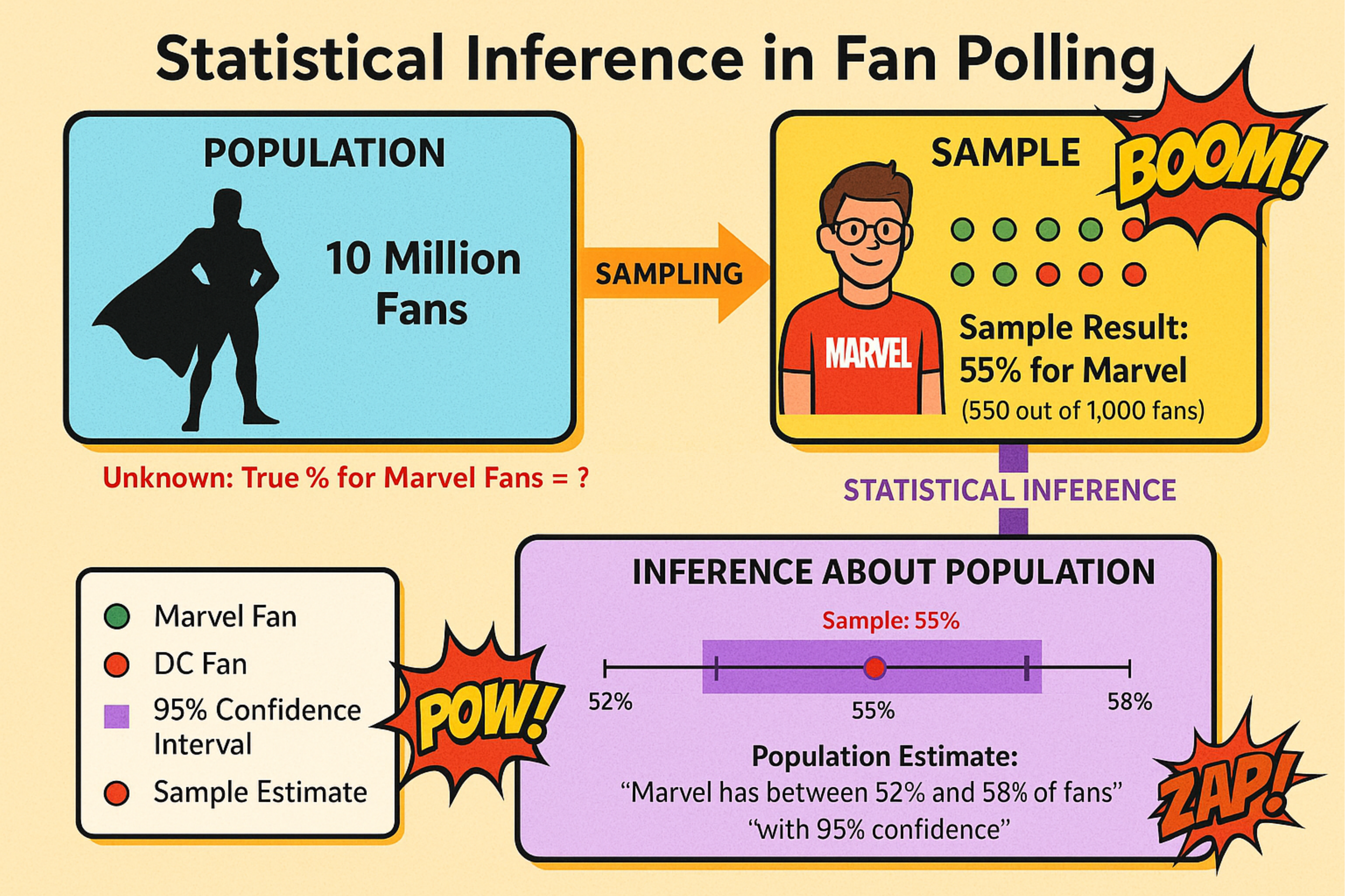
Source: https://www.nytimes.com/interactive/2020/nyregion/new-york-city-coronavirus-cases.html
Data as of May 25, 2021

What is Data Science?

Data science is about drawing useful conclusions from large and diverse data sets through…

- Exploration: Identifying patterns and trends using data (e.g., through visualization)

- Inference: Drawing reliable conclusions using statistics



Source: Murad Megjhani and his AI image generator of choice

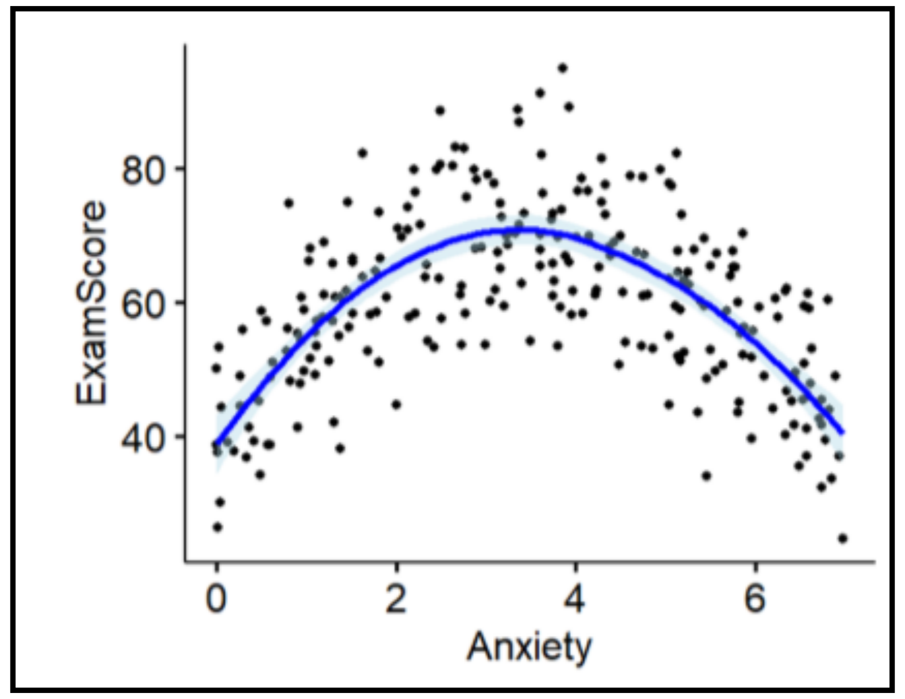
What is Data Science?

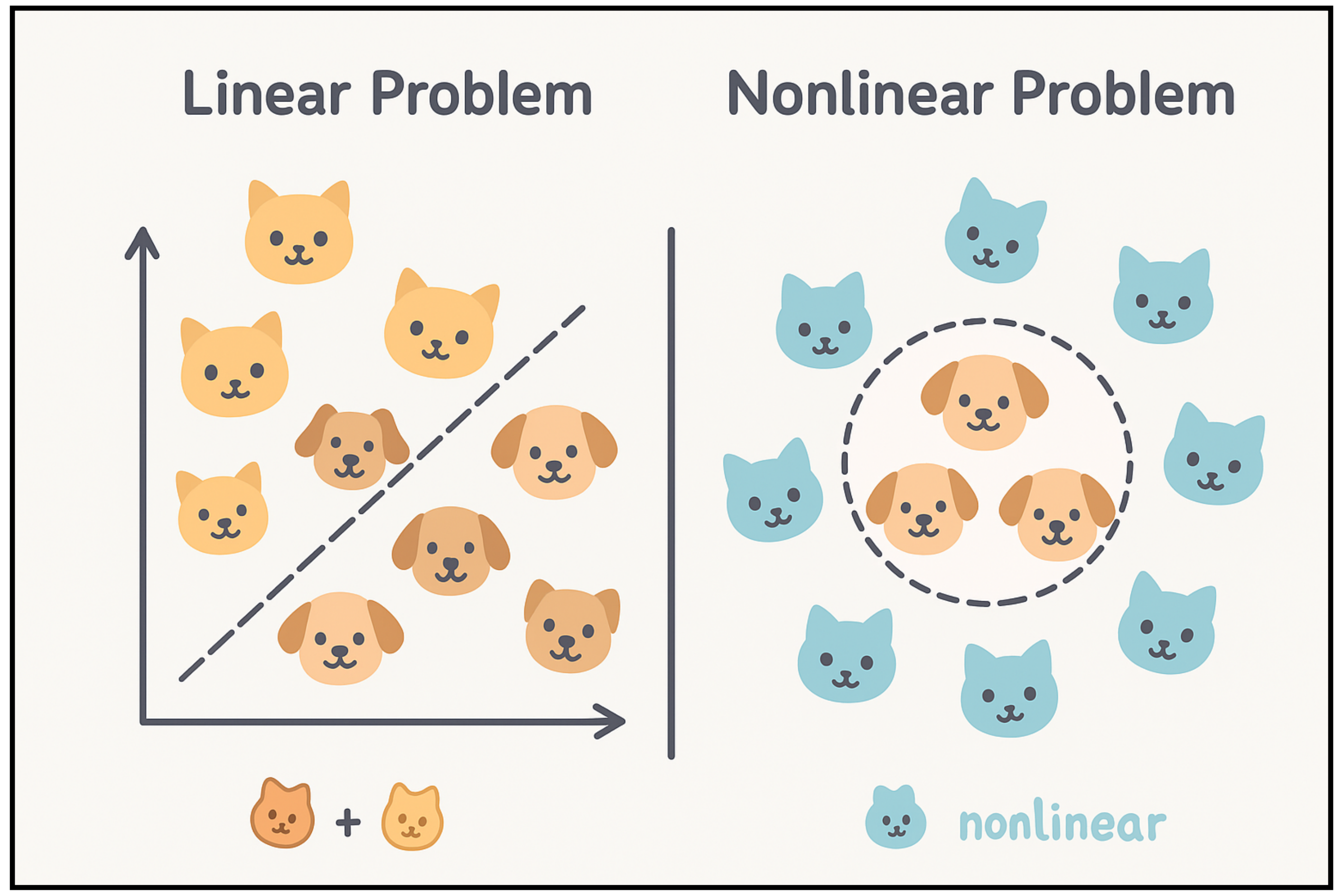
Data science is about drawing useful conclusions from large and diverse data sets through…

- Exploration: Identifying patterns and trends using data (e.g., through visualization)

- Inference: Drawing reliable conclusions using statistics

- Prediction: Making informed guesses about patterns using models





Source: Murad Megjhani and his AI image generator of choice

Course Topics

Course Topics

Programming

Midterm Exam

Statistics

Final Project

Course Topics

Programming

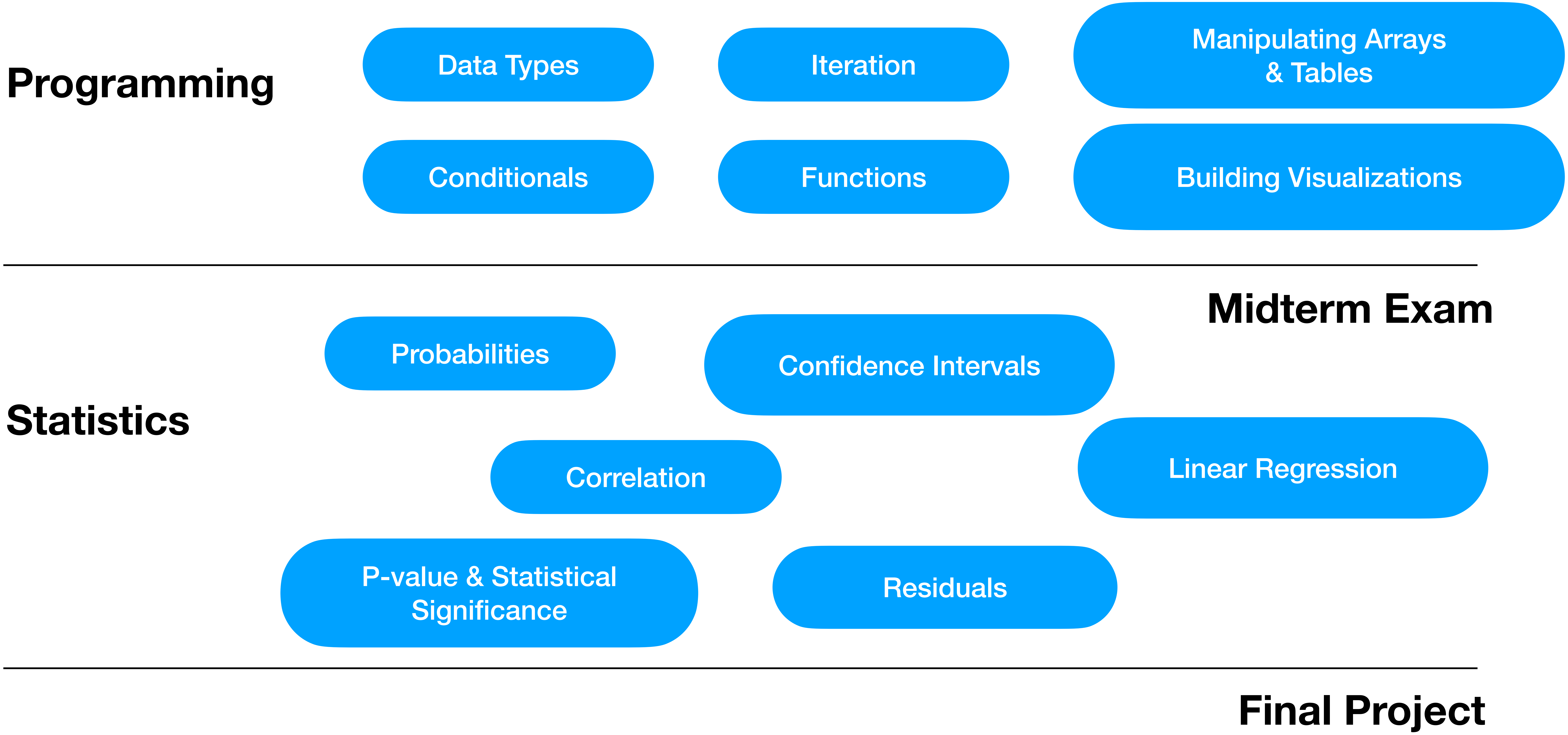
- Data Types
- Iteration
- Manipulating Arrays & Tables
- Conditionals
- Functions
- Building Visualizations

Midterm Exam

Statistics

Final Project

Course Topics



Programming

Data Types

Iteration

Manipulating Arrays & Tables

Conditionals

Functions

Building Visualizations

Midterm Exam

Statistics

Probabilities

Confidence Intervals

Correlation

Linear Regression

P-value & Statistical Significance

Residuals

Final Project

Datasets You’ll Explore

Climate Data

Vaccinations

Restaurant Reviews

Unemployment

Sports Records

Birth Rates

Movie Reviews

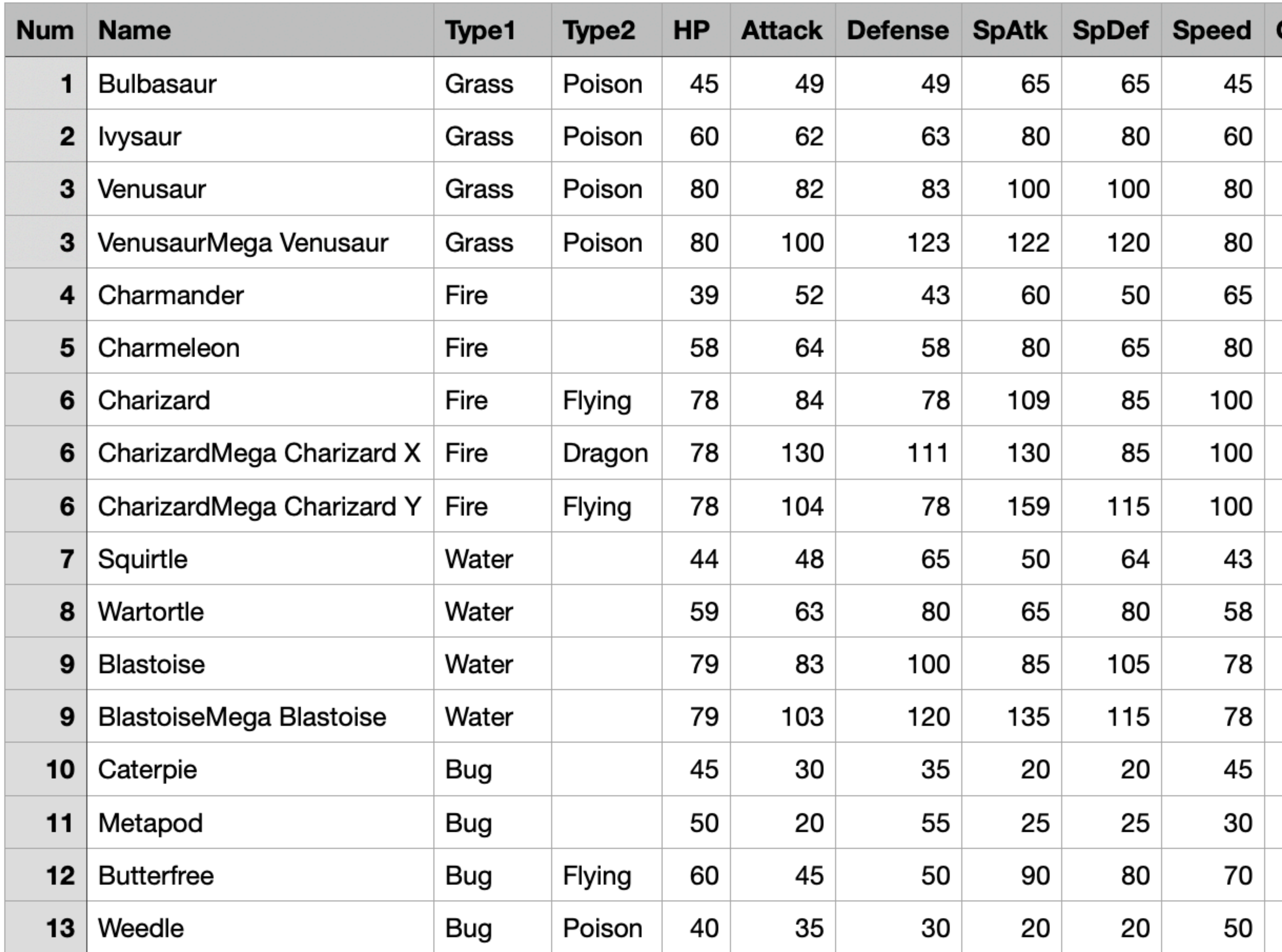
Compensation / Salaries

Happiness Scores

Ride Share Data

What does Data look like?

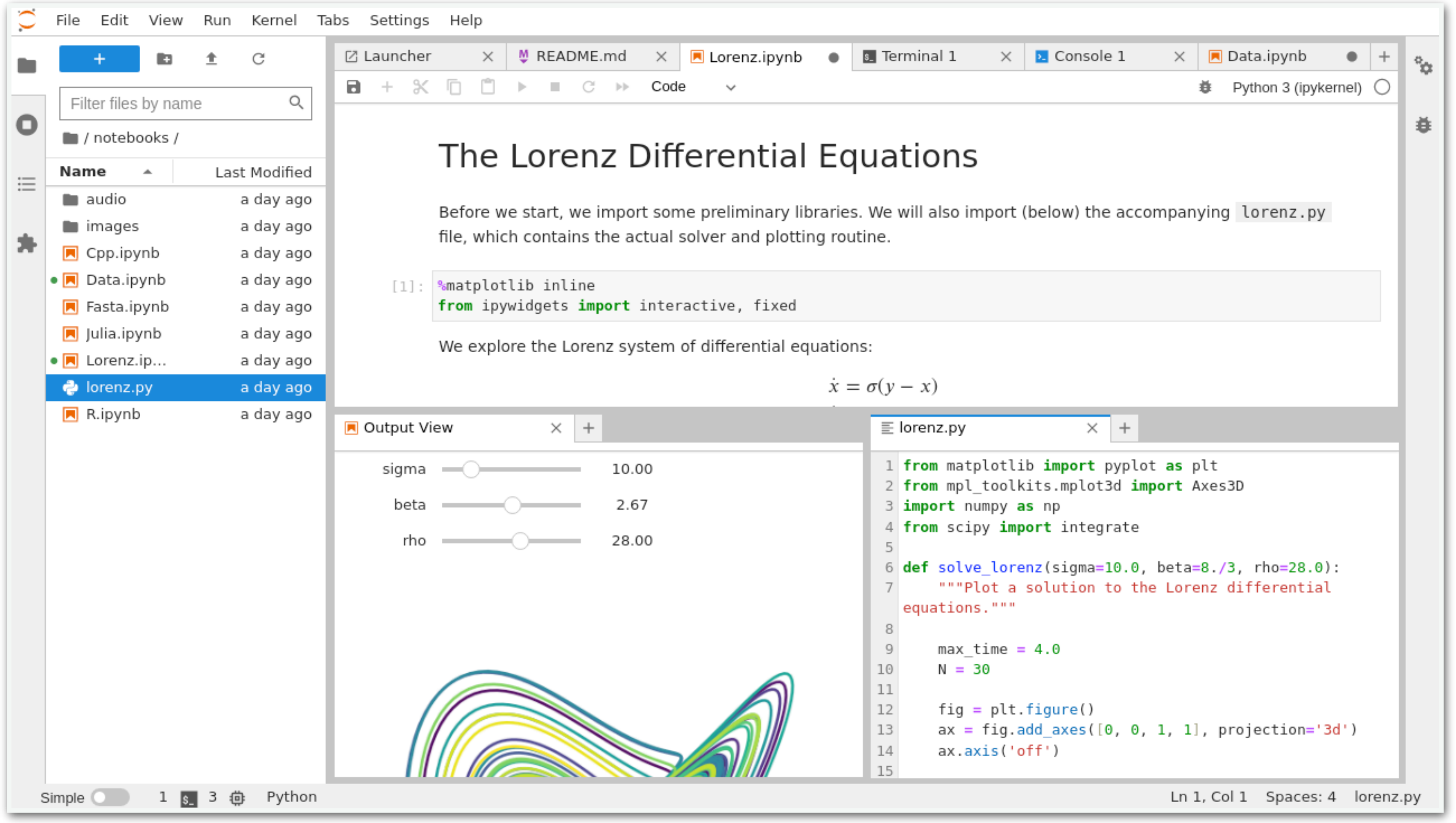
- Tabular data typically in the form of a CSV
- Header row with clear field names
- You will use Jupyter Notebooks to read tabular data and perform analyses on it



Num	Name	Type1	Type2	HP	Attack	Defense	SpAtk	SpDef	Speed
1	Bulbasaur	Grass	Poison	45	49	49	65	65	45
2	Ivysaur	Grass	Poison	60	62	63	80	80	60
3	Venusaur	Grass	Poison	80	82	83	100	100	80
3	VenusaurMega Venusaur	Grass	Poison	80	100	123	122	120	80
4	Charmander	Fire		39	52	43	60	50	65
5	Charmeleon	Fire		58	64	58	80	65	80
6	Charizard	Fire	Flying	78	84	78	109	85	100
6	CharizardMega Charizard X	Fire	Dragon	78	130	111	130	85	100
6	CharizardMega Charizard Y	Fire	Flying	78	104	78	159	115	100
7	Squirtle	Water		44	48	65	50	64	43
8	Wartortle	Water		59	63	80	65	80	58
9	Blastoise	Water		79	83	100	85	105	78
9	BlastoiseMega Blastoise	Water		79	103	120	135	115	78
10	Caterpie	Bug		45	30	35	20	20	45
11	Metapod	Bug		50	20	55	25	25	30
12	Butterfree	Bug	Flying	60	45	50	90	80	70
13	Weedle	Bug	Poison	40	35	30	20	20	50

What are notebooks?

- Jupyter notebooks are environments for creating and sharing computational documents
    - Combination of notes (text and comments), code, data, and figures
- Data science is typically done in Jupyter notebooks using Python
    - Python has a rich developer community & set of libraries made for data science



Name	Last Modified
audio	a day ago
images	a day ago
Cpp.ipynb	a day ago
Data.ipynb	a day ago
Fasta.ipynb	a day ago
Julia.ipynb	a day ago
Lorenz.ip...	a day ago
lorenz.py	a day ago
R.ipynb	a day ago

The Lorenz Differential Equations

Before we start, we import some preliminary libraries. We will also import (below) the accompanying `lorenz.py` file, which contains the actual solver and plotting routine.

```
[1]: %matplotlib inline
     from ipywidgets import interactive, fixed
```

We explore the Lorenz system of differential equations:

$$\dot{x} = \sigma(y - x)$$

sigma	10.00
beta	2.67
rho	28.00

```
from matplotlib import pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np
from scipy import integrate

def solve_lorenz(sigma=10.0, beta=8./3, rho=28.0):
    """Plot a solution to the Lorenz differential equations."""

    max_time = 4.0
    N = 30

    fig = plt.figure()
    ax = fig.add_axes([0, 0, 1, 1], projection='3d')
    ax.axis('off')

```

https://docs.jupyter.org/en/latest/#what-is-a-notebook

What will you learn to do?

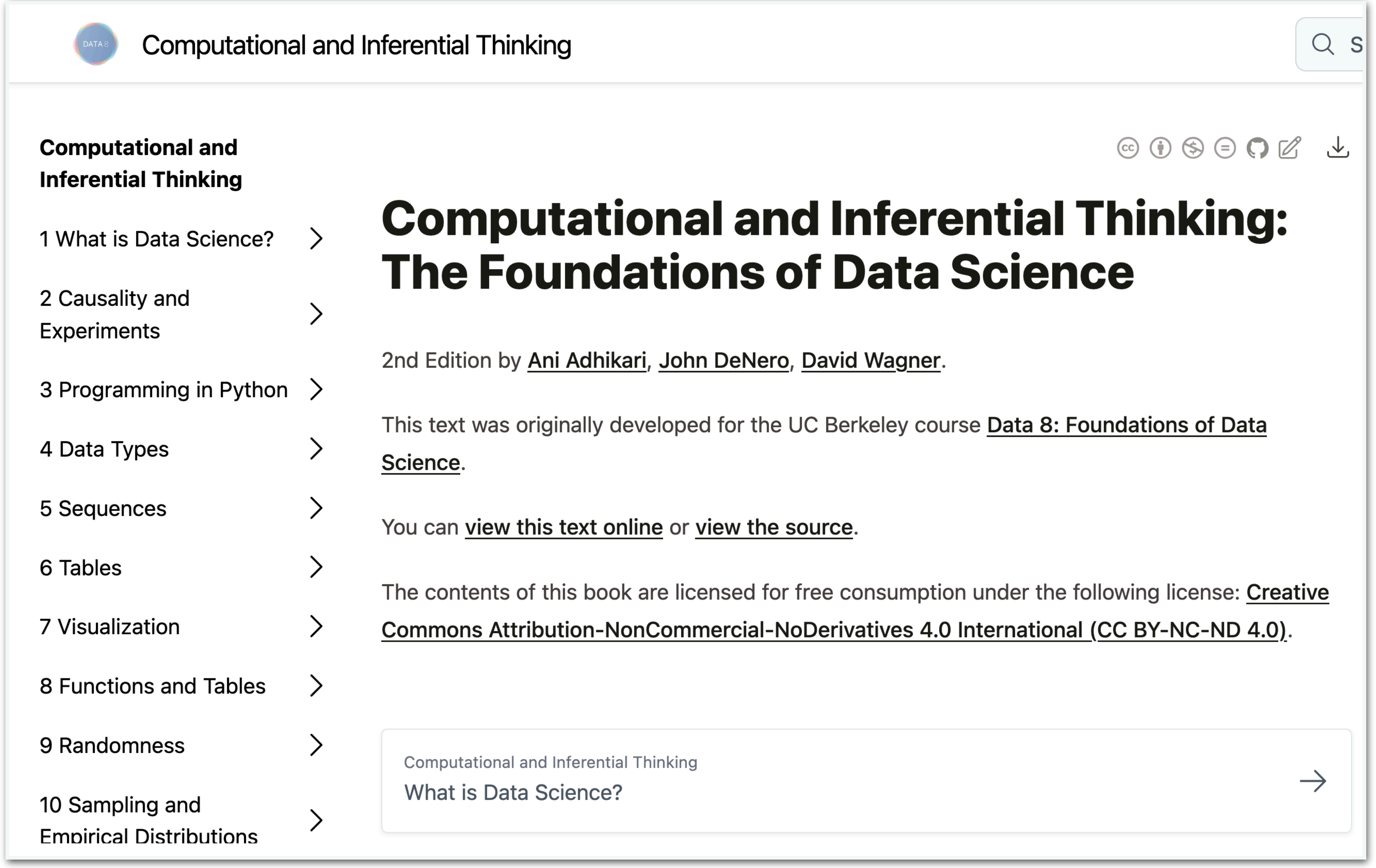
- Take a dataset and explore it with visualizations
- Write Python code to support your experimentation
- Uncover interesting patterns and insights that might help you understand the data in a new way

What is this course?

- Introduction to Computational Thinking and Data Science!
  - There is a lab section (BC 1017) associated with this course
  - You can register for any lab section, but you must be registered for one to take this course!

Course Acknowledgments

- Builds on top of Data 8 (Berkeley Data Science course)
    - Lots of universities build on top of it (UW, NYU, UCSD, McGill, Cornell, etc.)
- Their textbook is great!
    - https://inferentialthinking.com/chapters/intro.html



Computational and Inferential Thinking

Computational and Inferential Thinking: The Foundations of Data Science

2nd Edition by Ani Adhikari, John DeNero, David Wagner.

This text was originally developed for the UC Berkeley course Data 8: Foundations of Data Science.

You can view this text online or view the source.

The contents of this book are licensed for free consumption under the following license: Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0).

Computational and Inferential Thinking
What is Data Science?

TAs, Computing Fellows, & Lab Sections

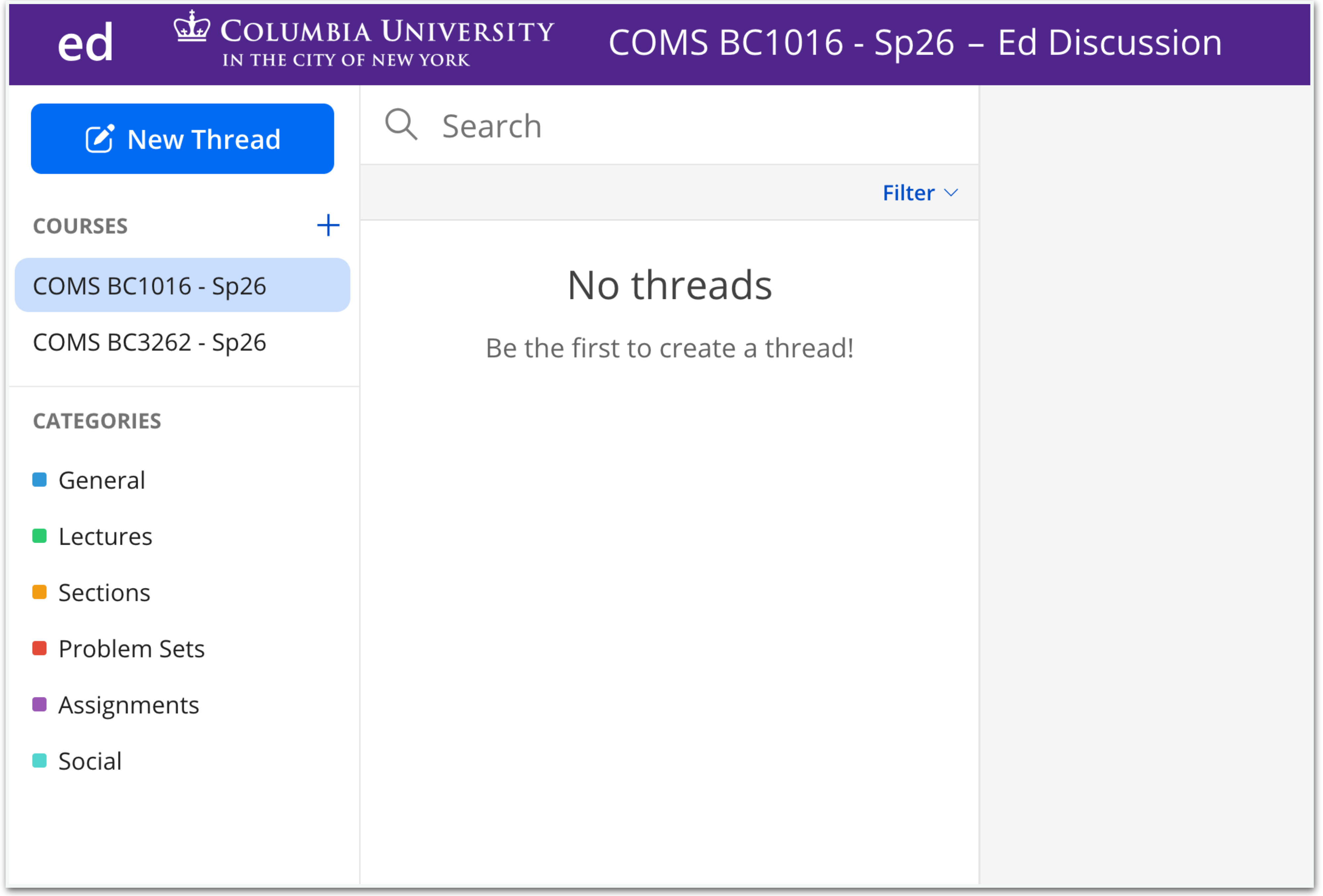
	TA	Computing Fellow
Wednesday **2:10-3:40pm**	Nami Jain nbj2115@columbia.edu	Elena Lukac
Thursday **9:00-10:30am**	Sathya Raman sr4213@columbia.edu	Madeline Gutierrez

Course Office Hours

- Office hours start week of Jan 26
- Professor Lee: Mondays 3:00pm-5:00pm (Milstein 512)
- TAs and Computing Fellows will offer 1.5 hours of OHs each week
    - Times TBA

EdStem

- We also have a class discussion forum: https://edstem.org/us/courses/89029/discussion
    - The EdStem will help us not lose track of your questions
- The TAs, Computing Fellows, and I will be monitoring and answering questions
    - You are also able to post privately to instructors if needed



No threads

Be the first to create a thread!

Note: labs start next week

(no lab this week)

Course Expectations: Assignments

- Weekly Lab Assignments are due **Fridays 11:59pm** via Courseworks
    - Intended to be finished and submitted during lab itself
- (Mostly) weekly Homeworks, due **Wednesdays 11:59pm** through Gradescope (via Courseworks)
    - These may take a little bit more time… Start early
- The lowest lab and lowest homework grade will be dropped

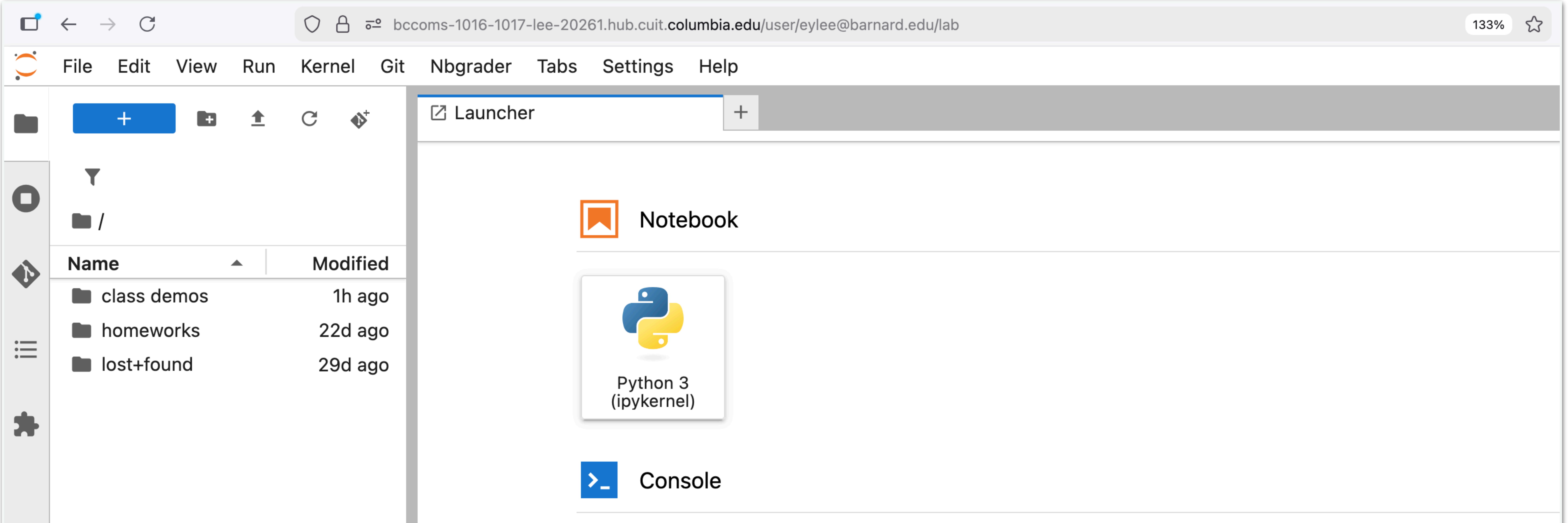
Monday	Tuesday	Wednesday	Thursday	Friday
Lecture Homework Released		Lecture Lab Lab Assignment Released	Lab	
		HW due 11:59pm		Lab due 11:59pm

Course Expectations: Assignments

- All assignments will be completed using cloud-based Jupyter notebooks

- You can access our course Jupyter Hub at: https://bccoms-1016-1017-lee-20261.hub.cuit.columbia.edu/

    - All you need is a web browser, no special software



bccoms-1016-1017-lee-20261.hub.cuit.columbia.edu/user/eylee@barnard.edu/lab

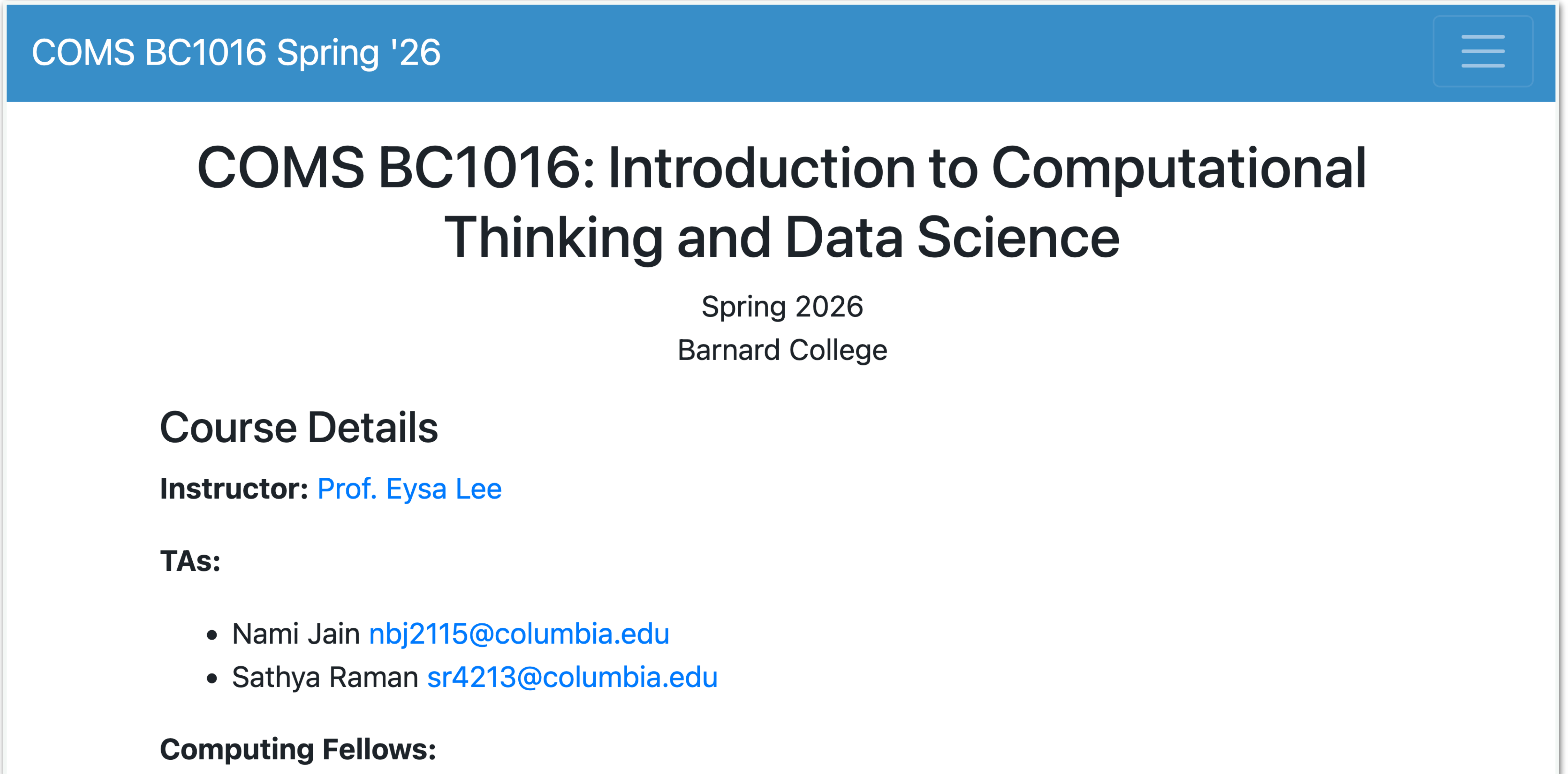
Name	Modified
class demos	1h ago
homeworks	22d ago
lost+found	29d ago

Course Website

https://www.eysalee.com/courses/s26/bc1016.html

- Contains syllabus, schedule (lectures, assignments, labs), and links to handy resources

- Slides and class demos will be uploaded or linked to from here



COMS BC1016 Spring '26

COMS BC1016: Introduction to Computational Thinking and Data Science

Spring 2026
Barnard College

Course Details

Instructor: Prof. Eysa Lee

TAs:

- Nami Jain nbj2115@columbia.edu
- Sathya Raman sr4213@columbia.edu

Computing Fellows:

Course Expectations: Lecture Grading

Your grade will be determined based on the following breakdown:

- 35%: Homework Assignments
- 25%: Midterm Exam
- 40%: Final Project

The lowest homework grade will be dropped.

Course Expectations: Lab Grading

Labs are graded out of 10 points:

- 5 points: Lab Assignments
- 5 points: Attendance

If you are going to be late or are unable to attend, email your lab TA in advance or you will receive 0 points for attendance.

- You are permitted one unexcused absence from lab during the semester.
- The lowest lab grade will be dropped.

Course Expectations: Lab Grading

Labs are graded out of 10 points:

- 5 points: Lab Assignments
- 5 points: Attendance

Assignment Grading:

- Complete and correct lab notebooks receive 5 points. Partially complete lab notebooks receive 3 points.
- Submit notebooks as a PDF via Courseworks.

Course Expectations: Regrade Requests

- TAs will grade all assignments within one week of submission.

- Any regrade requests must be submitted **within 1 week** of your grade being received

  - We will not consider any regrades after this timeframe

- If you request a regrade, we reserve the right to lower your grade if the original grading was found to be too generous.

Course Expectations: Late Policy

- Any late assignment (submitted after the due date) will be docked **10% of the total possible points per late day** for that assignment **up to five days**.
  - Any assignment submitted more than five days after the original due date will receive a no credit.
- This policy does not apply to the final project, which cannot be accepted after the due date except in exceptional circumstances.

Course Expectations: Generative AI Policy

- This course is meant to build your programming skills, so it is not advised to use generative AI tools.
    - We want you to build intuition about how to write code and fix common bugs!
    - Be aware generated code does not always represent best practices and may be verbose (or potentially incorrect!).
- **AI generated code or final report text is not permitted**

Midterm Exam

- Paper exam happens during class **Wednesday, March 11, 2026**
  - This is the week before Spring Recess
- You will be allowed a note sheet to use as a reference during the exam
  - It will be submitted along with your exam
- If you need particular accommodations, please contact CARDS

Final Project

- Groups of 2-3
- We will provide datasets to select from
- You will use the dataset to perform analyses using a combination of visualization and statistical analyses
- Final project report & Jupyter Notebook will be submitted during Finals week
- More info will be released after the midterm

Syllabus Recap

- Info can be found on the course website: https://www.eysalee.com/courses/s26/bc1016.html

- Assignments done through the course JupyterHub: https://bccoms-1016-1017-lee-20261.hub.cuit.columbia.edu/

- **Labs start next week**

    - Email your TA if you will be late or absent

    - Attendance is 50% of your lab grade, 1 excused absence

Let’s think about data

What is Data Science?

Data science allows us to draw useful conclusions from large and diverse data sets through:

- Exploration: Identifying patterns and trends using data

- Inference: Drawing reliable conclusions using statistics

- Prediction: Making informed guesses about patterns using models

Cause and Effect

A link between chocolate and health

A link between red wine and health

Red wine helps you live longer!

A link between cats and grades

Observational Studies

Definition: A study in which researchers / scientists make conclusions based on *observed data* in which they had no hand in generating

- Individuals: study subjects, participants, units

- Treatment: factor of interest

- Outcome: result of treatment

Observational Studies

Definition: A study in which researchers / scientists make conclusions based on *observed data* in which they had no hand in generating

people

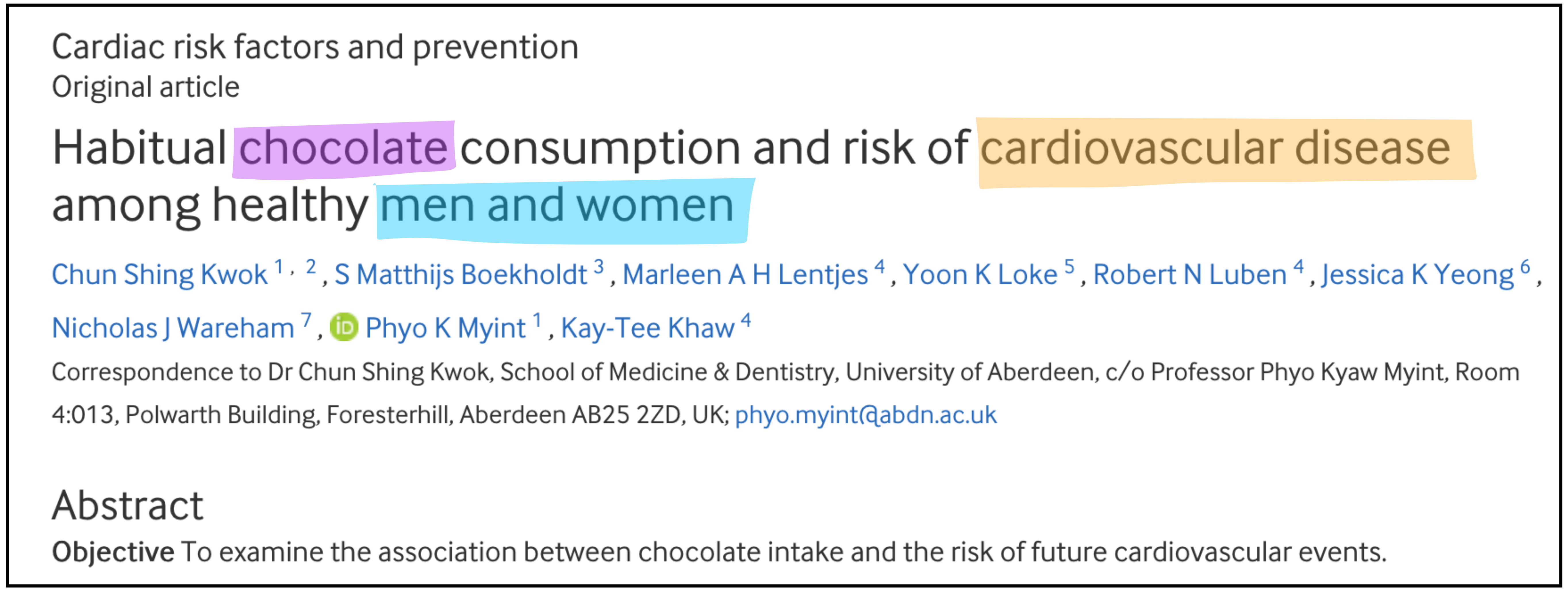
- Individuals: study subjects, participants, units

chocolate consumption

- Treatment: factor of interest

- Outcome: result of treatment

heart health



Cardiac risk factors and prevention
Original article

Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok [1, 2], S Matthijs Boekholdt [3], Marleen A H Lentjes [4], Yoon K Loke [5], Robert N Luben [4], Jessica K Yeong [6], Nicholas J Wareham [7], Phyo K Myint [1], Kay-Tee Khaw [4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

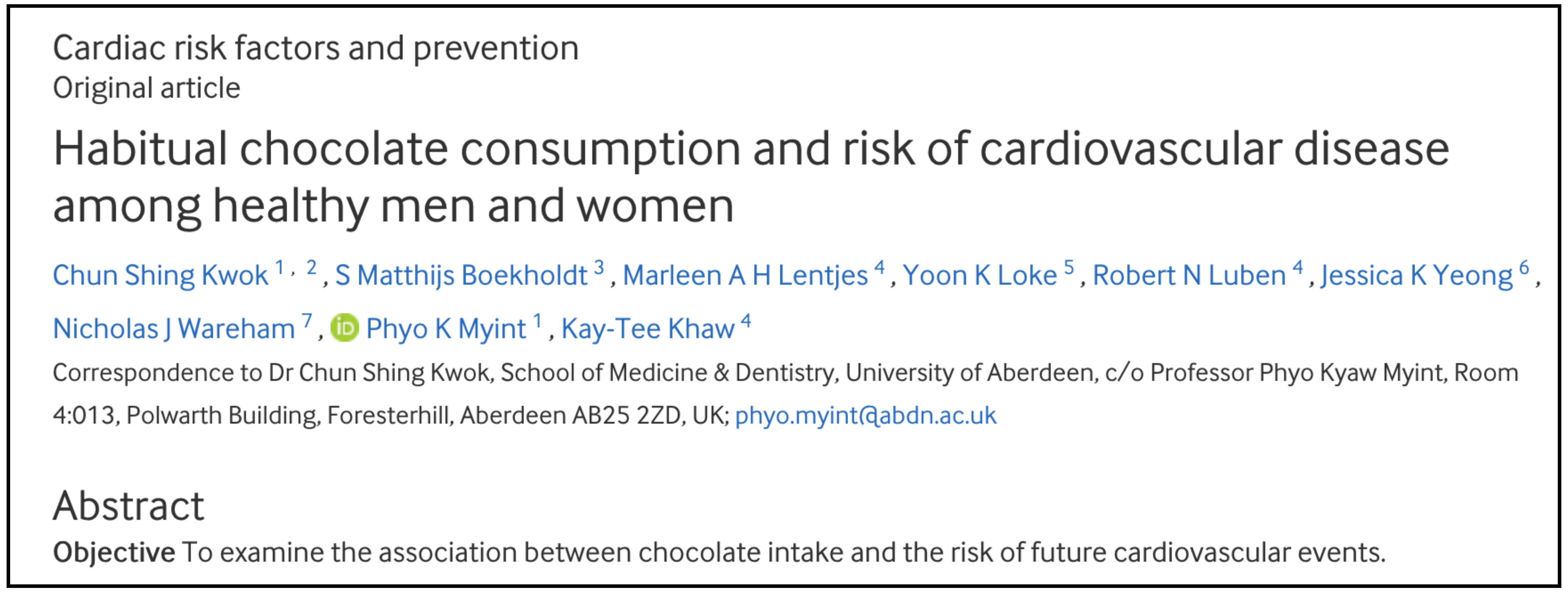
Objective To examine the association between chocolate intake and the risk of future cardiovascular events.

Source: https://heart.bmj.com/content/101/16/1279

Association and Causality

Association: Any relation between the treatment and the outcome

Causality: if the treatment causes the outcome to occur



Cardiac risk factors and prevention
Original article

Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok [1, 2], S Matthijs Boekholdt [3], Marleen A H Lentjes [4], Yoon K Loke [5], Robert N Luben [4], Jessica K Yeong [6], Nicholas J Wareham [7], Phyo K Myint [1], Kay-Tee Khaw [4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

Objective To examine the association between chocolate intake and the risk of future cardiovascular events.

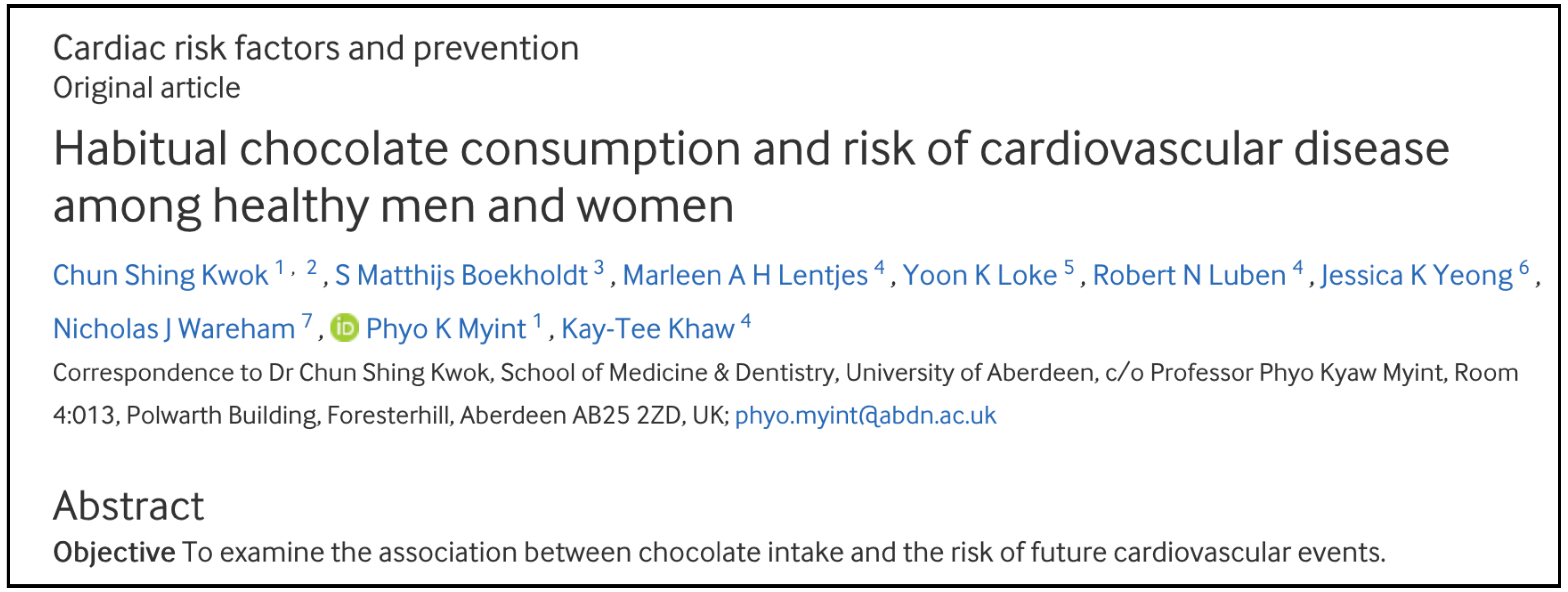
Association and Causality

eating chocolate

Association: Any relation between the treatment and the outcome

heart health

Causality: if the treatment causes the outcome to occur



Cardiac risk factors and prevention
Original article

Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok [1, 2], S Matthijs Boekholdt [3], Marleen A H Lentjes [4], Yoon K Loke [5], Robert N Luben [4], Jessica K Yeong [6], Nicholas J Wareham [7], Phyo K Myint [1], Kay-Tee Khaw [4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

Objective To examine the association between chocolate intake and the risk of future cardiovascular events.

Association and Causality

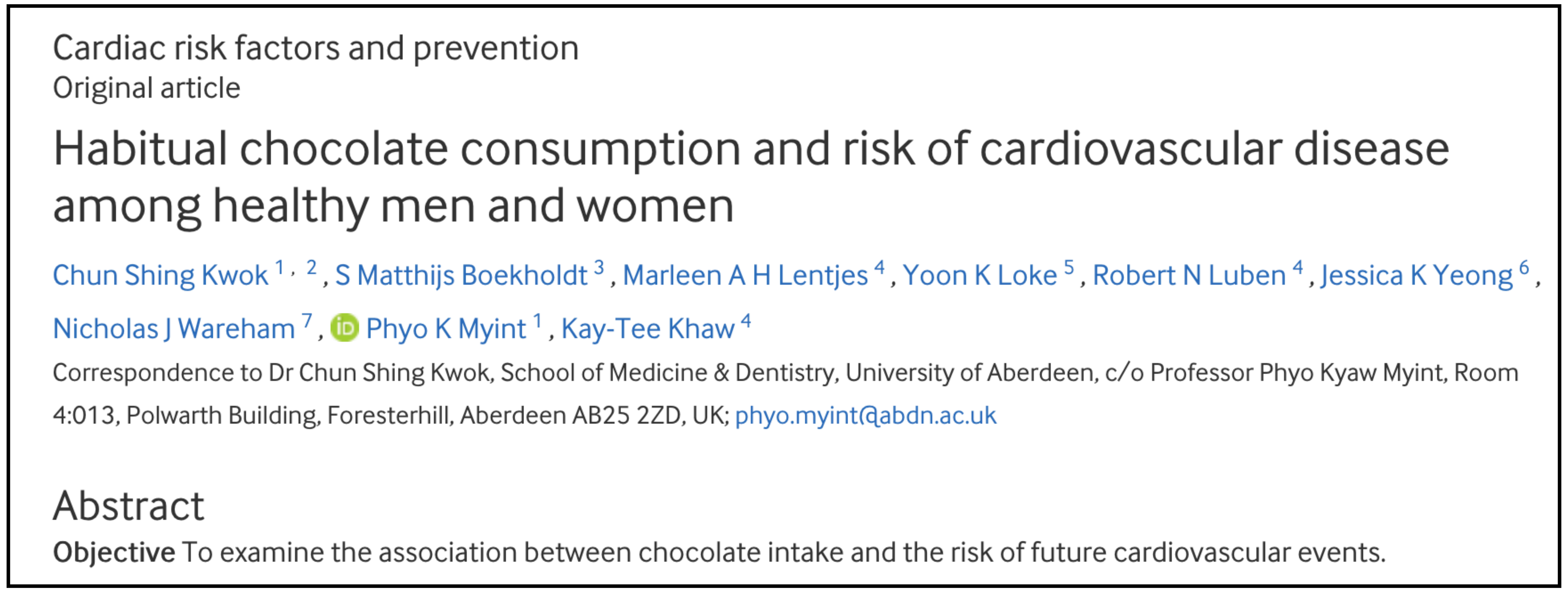
eating chocolate

Association: Any relation between the treatment and the outcome

heart health

Causality: if the treatment causes the outcome to occur

Is there an association between chocolate consumption and heart disease?
According to this study, yes ☺



Cardiac risk factors and prevention
Original article

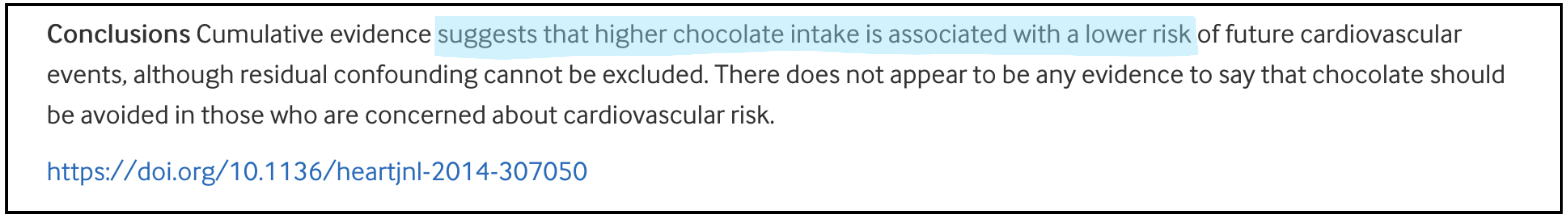
Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok [1, 2], S Matthijs Boekholdt [3], Marleen A H Lentjes [4], Yoon K Loke [5], Robert N Luben [4], Jessica K Yeong [6], Nicholas J Wareham [7], Phyo K Myint [1], Kay-Tee Khaw [4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

Objective To examine the association between chocolate intake and the risk of future cardiovascular events.



Conclusions Cumulative evidence suggests that higher chocolate intake is associated with a lower risk of future cardiovascular events, although residual confounding cannot be excluded. There does not appear to be any evidence to say that chocolate should be avoided in those who are concerned about cardiovascular risk.

https://doi.org/10.1136/heartjnl-2014-307050

Association and Causality

eating chocolate

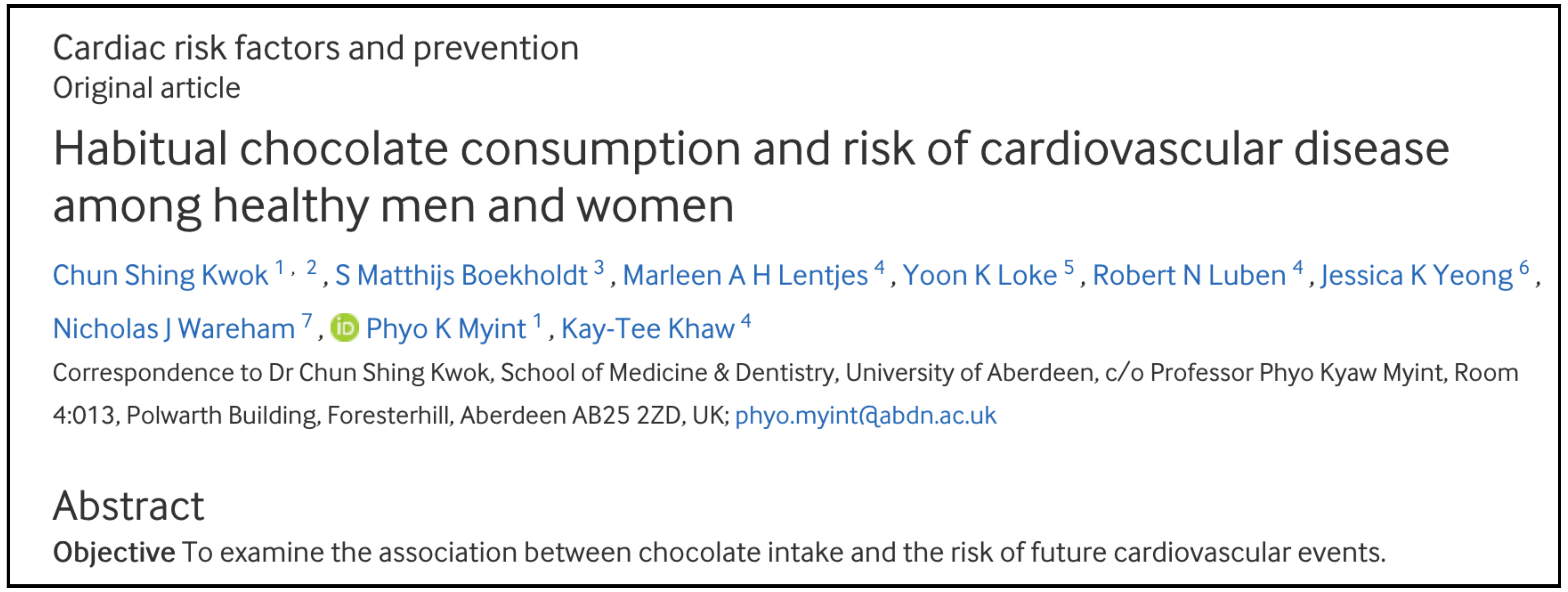
Association: Any relation between the treatment and the outcome

heart health

Causality: if the treatment causes the outcome to occur

Is there an association between chocolate consumption and heart disease? According to this study, yes ☺

Does eating chocolate lead to a reduction in heart disease? Well... that's harder to say...



Cardiac risk factors and prevention
Original article

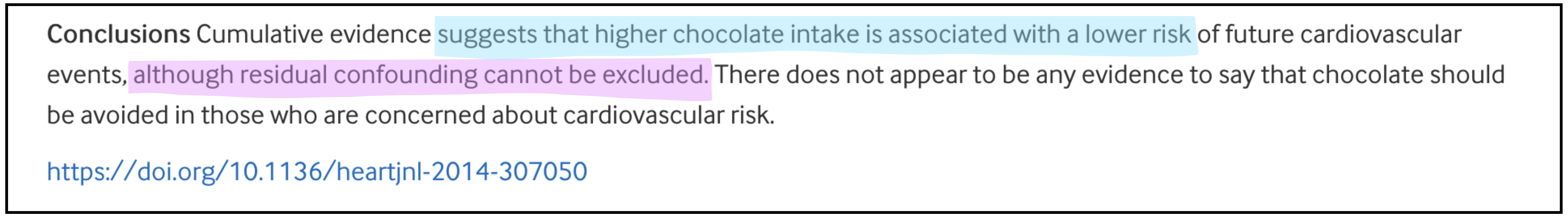
Habitual chocolate consumption and risk of cardiovascular disease among healthy men and women

Chun Shing Kwok[1, 2], S Matthijs Boekholdt[3], Marleen A H Lentjes[4], Yoon K Loke[5], Robert N Luben[4], Jessica K Yeong[6], Nicholas J Wareham[7], Phyo K Myint[1], Kay-Tee Khaw[4]

Correspondence to Dr Chun Shing Kwok, School of Medicine & Dentistry, University of Aberdeen, c/o Professor Phyo Kyaw Myint, Room 4:013, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; phyo.myint@abdn.ac.uk

Abstract

Objective To examine the association between chocolate intake and the risk of future cardiovascular events.

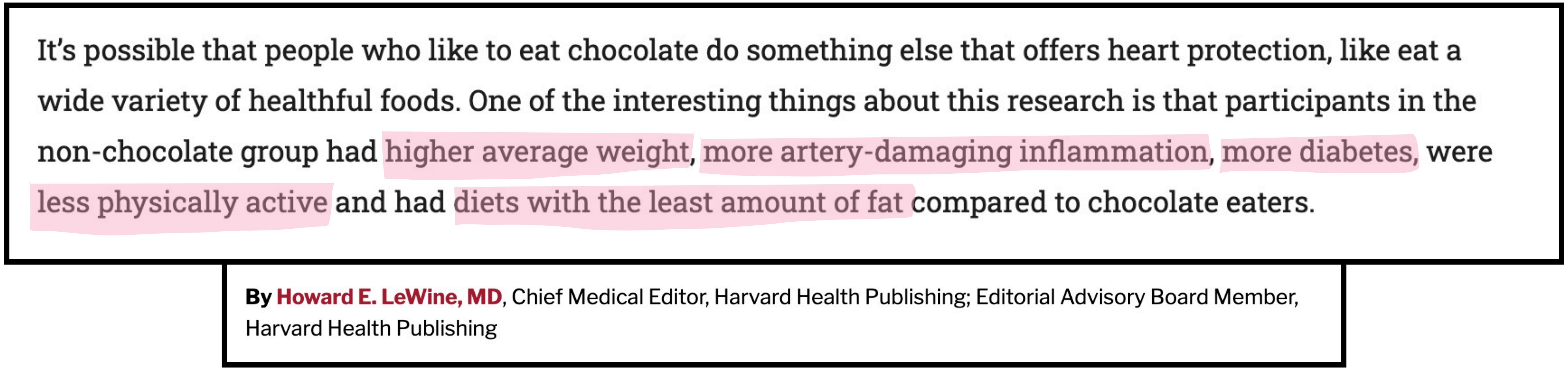


Conclusions Cumulative evidence suggests that higher chocolate intake is associated with a lower risk of future cardiovascular events, although residual confounding cannot be excluded. There does not appear to be any evidence to say that chocolate should be avoided in those who are concerned about cardiovascular risk.

https://doi.org/10.1136/heartjnl-2014-307050

Confounding Factors

Definition: In observational studies, underlying difference(s) between two groups other than the treatment (that make it difficult to identify causality)



It's possible that people who like to eat chocolate do something else that offers heart protection, like eat a wide variety of healthful foods. One of the interesting things about this research is that participants in the non-chocolate group had higher average weight, more artery-damaging inflammation, more diabetes, were less physically active and had diets with the least amount of fat compared to chocolate eaters.

By Howard E. LeWine, MD, Chief Medical Editor, Harvard Health Publishing; Editorial Advisory Board Member, Harvard Health Publishing

Source: https://www.health.harvard.edu/blog/sweet-dreams-eating-chocolate-prevents-heart-disease-201506168087

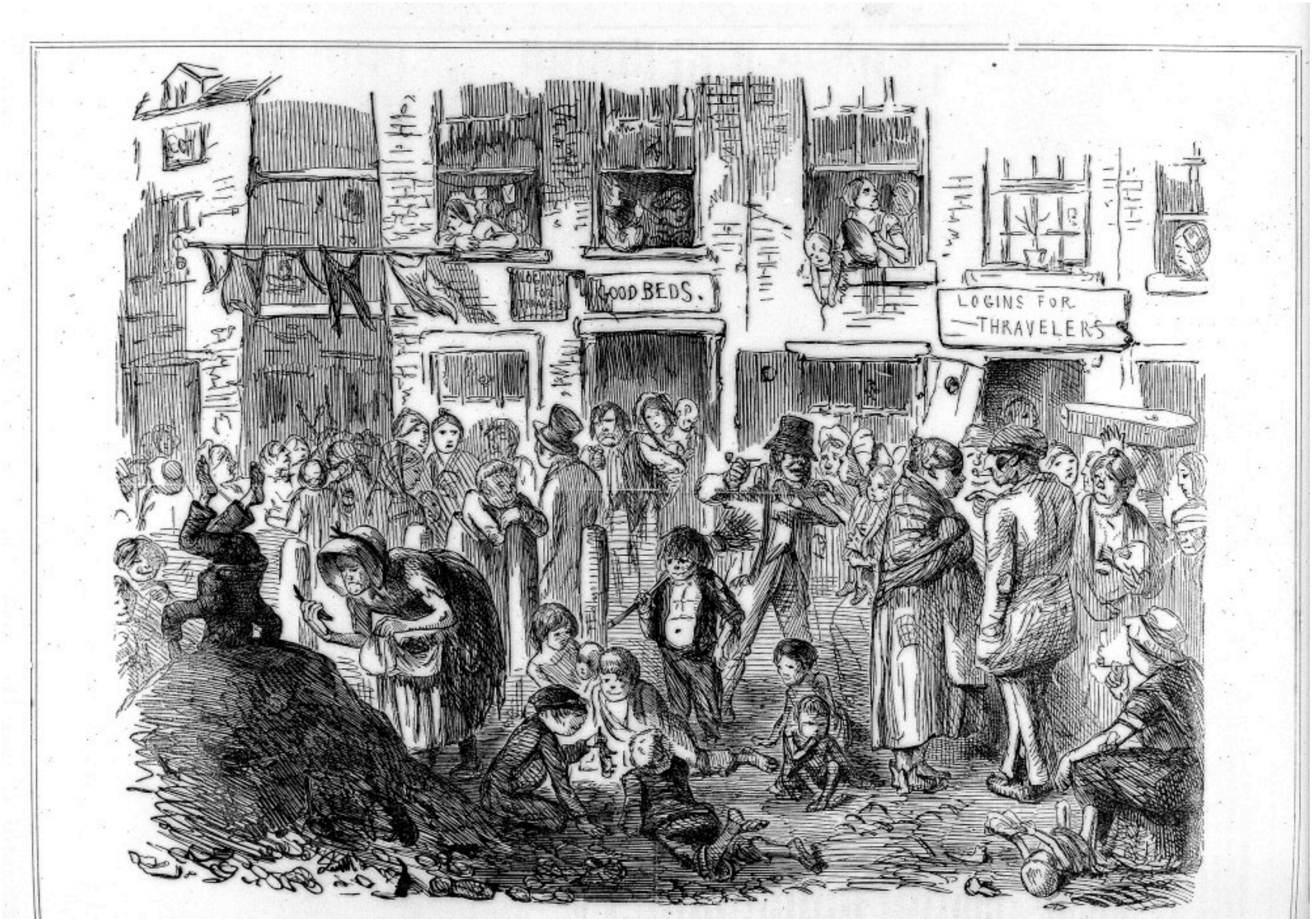
In general, observational studies are not enough to determine causation

Association and Causation

London 1850s: Cholera Epidemic

- Large migration into London in the 1700-1800s, leading to overcrowding

- What's causing cholera to spread?



A COURT FOR KING CHOLERA.

https://www.sciencemuseum.org.uk/objects-and-stories/medicine/cholera-victorian-london

Miasma

- ‘Bad air’ given off by waste and rotting matter

- Believed to the main source of how infectious disease spread

- Potential remedies: “Fly to clean air”, “pocket full of posies”, “fire off barrels of gunpowder”

- Popular medical theory of the time

  - Florence Nightingale (founder of modern nursing)

  - Edwin Chadwick (Commissioner of the Board of Health)

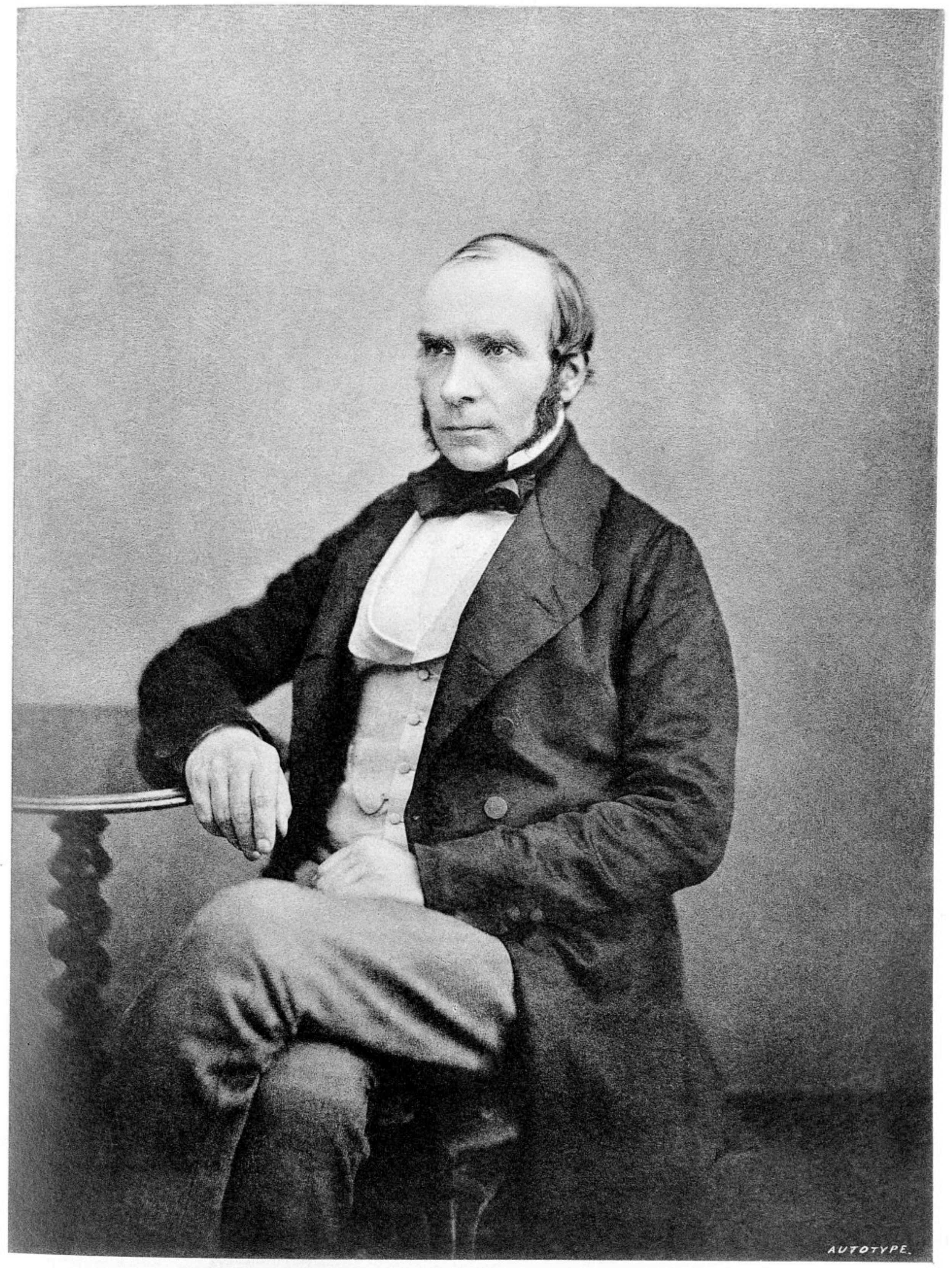
John Snow (1813 - 1858)

- English physician
- Used data and visualizations to understand why cholera was spreading in the way that it was
  - Noticed people could be in close proximity without the same outcomes

Jon Snow

Fictional character

← not this dude



John Snow

(Autotype from a Presentation Portrait, 1856, and Autograph facsimile.—B. W. R.)

https://wellcomecollection.org/works/dkk5vx46

STREET
PORTLAND
PORTLAND MEWS
WARDOUR MEWS
BENTINCK STREET
EDWARD STREET
ST ANNS COURT
RICHMOND BUILDINGS
BREWERY YARD
RICHMOND MEWS
SHIP YARD
DUCK LANE
HAM SQUARE
FRITH
QUEEN
DEAN STREET
LITTLE MARLBOROUGH
TYLER CT
TYLER STREET
LOWNDES CT
MARLBOROUGH CT
WEST STREET
MARSHALL STREET
DUFOURS PLACE
SOUTH ROW
BROAD STREET
PUMP
CARNABY
CROSS ST
MARLBOROUGH ROW
FOUBERTS PL
KING STREET
CROSS STREET
CAMBRIDGE ST
NEW STREET
BREWERY
HOPKINS STREET
COCK CT
PULTENEY CT
HUSBAND ST
MAIDENHEAD COURT
TYLERS CT
BLACK LION CT
MEARDS STREET
MEARDS COURT
OLD COMPTON
PETER STREET
WALKERS COURT
STREET
CT CROWN CT
ST CROWN COURT
PUMP
UP. RUPERT ST
RICHMOND ST
ARCHER STREET
PLOUGH YARD
GREAT CROWN CT
QUEENS HEAD COURT
LITTLE PULTENEY
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
WILLIAM AND MARY YARD
BRIDLE STREET
GOLDEN SQUARE
UP JAMES ST
UP JOHN ST
SILVER STREET
BEAK ST
CHAPEL PLACE
BURLINGTON MEWS
NEW BURLINGTON STREET
K-I-N-G-S-T-R-E-E-T
GEORGE
PUMP

SOHO in London

- Black bars represent number of people who died at each address

GREAT MARLBOROUGH STREET
ARGYLL ST
STREET
ARGYLL PLACE
PUMP
GREEN DRAGON YARD
LITTLE MARLBOROUGH ST
TYLER STREET
TYLER COURT
MARSHALL ST.
WORK HOUSE
PORTLAND
PORTLAND MEWS
BENTINCK STREET
WARDOUR MEWS
D STREET
WICK
OUR
ST ANNS PLACE
ST ANNS COURT
RICHMOND BUILDINGS
BREWERY YARD
RICHMOND MEWS
SHIP YARD
FRITH
QUEEN
DEAN STREET
EDWARD STREET
DUCK LANE
HAM SQUARE
DUFOURS PLACE
WEST STREET
LOWNDES CT
MARLBOROUGH CT
CARNABY
MARSHALL STREET
SOUTH ROW
CROSS ST
MARLBOROUGH ROW
BROAD STREET
PUMP
CAMBRIDGE ST
BREWERY
HOPKINS STREET
NEW STREET
PULTENEY CT
COCK CT
TYLERS CT
BLACK LION CT
MEARDS STREET
MEARDS COURT
LITTLE DEAN STREET
FOUBERTS PL.
KING STREET
PEWS PLACE
CROSS STREET
STREET
MAIDENHEAD COURT
HUSBAND ST
PETER STREET
WALKERS COURT
OLD COMPTON
LITTLE WINDMILL STREET
WILLIAM AND MARY YARD
CHAPEL PLACE
NAYLORS YARD
SILVER STREET
UP JAMES ST
GREAT PULTENEY STREET
BRIDLE STREET
PULTENEY COURT
CHAPEL COURT
LITTLE PULTENEY
CT CROWN CT
LT CROWN COURT
GREAT CROWN CT
PUMP
UP RUPERT ST
STREET
PUMP
BURLINGTON MEWS
NEW BURLINGTON STREET
REGENT STREET
BEAK ST.
WARW
UP JOHN ST
GOLDEN SQUARE
PUMP
QUEENS HEAD COURT
ARCHER STREET
PLOUGH YARD
RUPERT
RICHMOND ST
KING STREET
PRINCES
GEORGE
PUMP

Great Marlborough Street
Argyll Place
Argyll St
Pump
Green Dragon Yard
Little Marlborough St
Tyler Street
Tyler Court
Marshall St.
West Street
Lowndes Ct
Marlborough Ct
Carnaby Street
Marshall Street
Work House
Dufours Place
Poland Street
Portland Street
Portland Mews
Bentinck Street
Berwick Street
Wardour Mews
Wardour Street
St Anns Place
St Anns Court
Richmond Buildings
Richmond Mews
Brewery Yard
Ship Yard
Edward Street
Duck Lane
Ham Square
Frith Street
Queen
Dean Street
Broad Street
Pump
Cambridge St
New Street
Brewery
Hopkins Street
Cock Ct
Husband St
Walker's Court
Tylers Ct
Black Lion Ct
Meards Street
Meards Court
Little Dean Street
Old Compton
Peter Street
Walkers Court
Crown Court
Pulteney Court
Pulteney Court
William and Mary Yard
Little Windmill Street
Great Pulteney Street
Bridle Street
Little Pulteney Street
Great Crown Ct
Ct Crown Ct
Lt Crown Court
Queens Head Court
Pump
Up. Rupert St
Richmond St
Archer Street
Plough Yard
Rupert
King Street
George
Pump
South Row
Cross St
Marlborough Row
Cross Street
King Street
Foubert's Pl.
Pewe Place
Silver Street
Up James St
Up John St
Naylor's Yard
Chapel Place
Golden Square
Beak St.
Warwick
Burlington Mews
New Burlington Street
Regent Street
Great Windmill Street
Pump

BROAD STREET
PUMP
SHALL STREET
URS PLACE
SOUTH ROW
CROSS ST
MARLBOROUGH ROW
STREET
STREET
CAMBRIDGE ST
NEW STREET
BREWERY
HOPKINS STREET
COCK CT.
PULTENEY CT.
STREET
MAIDENHEAD COURT
HUSBAND ST
LITTLE WINDMILL ST
GREAT PULT
WILLIAM AND MARY YARD
PULTENEY COURT
VER STREET
UP JAMES ST
BRI
P

Water Pumps

PUMP
PUMP
PUMP
PUMP
MARLBOROUGH MEWS
BLENHEIM MEWS
BLENHEIM STREET
ARGYLL STREET
ARGYLL ST
ARGYLL PLACE
GREAT MARLBOROUGH STREET
POLAND STREET
NOEL STREET
BERWICK STREET
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
WARDOUR MEWS
WARDOUR STREET
LITTLE CHAPEL
ST ANNS PLACE
ST ANNS COURT
RICHMOND BUILDINGS
RICHMOND MEWS
BREWERY YARD
DEAN STREET
FRITH STREET
QUEEN
WORK HOUSE
GREEN DRAGON YARD
LITTLE MARLBOROUGH ST
TYLER STREET
TYLERS COURT
MARSHALL ST.
MARSHALL STREET
WEST STREET
LOWNDES CT
CARNABY STREET
DUFOURS PLACE
BROAD STREET
CAMBRIDGE ST
EDWARD STREET
DUCK LANE
HAM YARD
HAM SQUARE
BREWERY
HOPKINS STREET
NEW STREET
COCK CT
MEARDS COURT
MEARDS STREET
TYLERS CT
BLACK LION CT
FOUBERTS PL
KING STREET
CROSS STREET
CROSS ST
SOUTH ROW
MARLBOROUGH ROW
PEERS PLACE
PULTENEY CT
HUSBAND ST
MAIDENHEAD COURT
PETER STREET
LITTLE DEAN STREET
OLD COMPTON
SILVER STREET
UP JAMES ST
BRIDLE
GREAT PULTENEY
LITTLE WINDMILL STREET
WILLIAM AND MARY YARD
WALKERS COURT
CROWN CT
PULTENEY
UP RUPERT ST
OLD BURLINGTON MEWS
CHAPEL PLACE
TAYLORS YARD
REGENT STREET

Water Pumps

PUMP
MARLBOROUGH MEWS
BLENHEIM MEWS
BLENHEIM STREET
ARGYLL STREET
ARGYLL ST
ARGYLL PLACE
GREAT MARLBOROUGH STREET
POLAND STREET
NOEL STREET
BERWICK STREET
WARDOUR STREET
LITTLE CHAPEL
PORTLAND STREET
PORTLAND MEWS
WARDOUR MEWS
ST ANNS PLACE
ST ANNS COURT
WORK HOUSE
BENTINCK STREET
RICHMOND BUILDINGS
RICHMOND MEWS
BREWERY YARD
EDWARD STREET
DEAN STREET
DUCK LANE
HAM SQUARE
QUEEN
GREEN DRAGON YARD
PUMP
LITTLE MARLBOROUGH ST
TYLER COURT
MARSHALL ST.
MARSHALL STREET
WEST STREET
LOWNDES CT
TYLER STREET
CARNABY STREET
DUFOURS PLACE
BROAD STREET
PUMP
CAMBRIDGE ST
BREWERY
HOPKINS STREET
NEW STREET
COCK CT
TYLERS CT
MEARDS STREET
MEARDS COURT
FOUBERTS PL
KING STREET
SOUTH ROW
CROSS ST
MARLBOROUGH ROW
CROSS STREET
PULTENEY CT
HUSBAND ST
MAIDENHEAD COURT
BLACK LION CT
PETER STREET
WALKERS COURT
OLD COMPTON
LITTLE WINDMILL STREET
GREAT PULTENEY STREET
BRIDLE
WILLIAM AND MARY YARD
GREENS COURT
PULTENEY STREET
CROWN CT
PUMP
UP RUPERT ST
SILVER STREET
UP JAMES ST
OLD BURLINGTON MEWS
CHAPEL PLACE
TAYLORS YARD
SWALLOW STREET
FRITH

Water Pumps

PUMP
PUMP
PUMP
PUMP
MARLBOROUGH MEWS
GREAT MARLBOROUGH STREET
ARGYLL PLACE
ARGYLL STREET
BLENHEIM MEWS
BLENHEIM STREET
POLAND STREET
NOEL STREET
BERWICK STREET
PORTLAND STREET
PORTLAND MEWS
BENTINCK STREET
WARDOUR STREET
WARDOUR MEWS
LITTLE CHAPEL
ST ANNS COURT
ST ANNS PLACE
RICHMOND BUILDINGS
RICHMOND MEWS
BREWERY YARD
EDWARD STREET
DUCK LANE
HAM YARD
WORK HOUSE
DUFOURS PLACE
MARSHALL ST.
MARSHALL STREET
GREEN DRAGON YARD
LITTLE MARLBOROUGH ST
TYLER COURT
TYLER STREET
CARNABY STREET
WEST STREET
LOWNDES CT
MARLBOROUGH CT
PEMS PLACE
SOUTH ROW
CROSS ST
MARLBOROUGH ROW
BROAD STREET
CAMBRIDGE ST
BREWERY
NEW ST
HOPKINS STREET
COCK CT
PULTENEY COURT
HUSBAND ST
MAIDENHEAD COURT
TYLERS CT
BLACK LION CT
MEARDS COURT
MEARDS STREET
PETER STREET
WALKERS COURT
DEAN STREET
QUEEN STREET
FRITH STREET
OLD COMPTON
FOUBERTS PL
KING STREET
CROSS STREET
SILVER STREET
REGENT STREET
OLD BURLINGTON MEWS
CHAPEL PLACE
TAYLORS YARD
UP JAMES ST
BRIDLE
GREAT PULTENEY
LITTLE WINDMILL STREET
WILLIAM AND MARY YARD
PULTENEY COURT
GREENS COURT
PULTENEY
CROWN CT
UP RUPERT ST
LITTLE CROWN CT

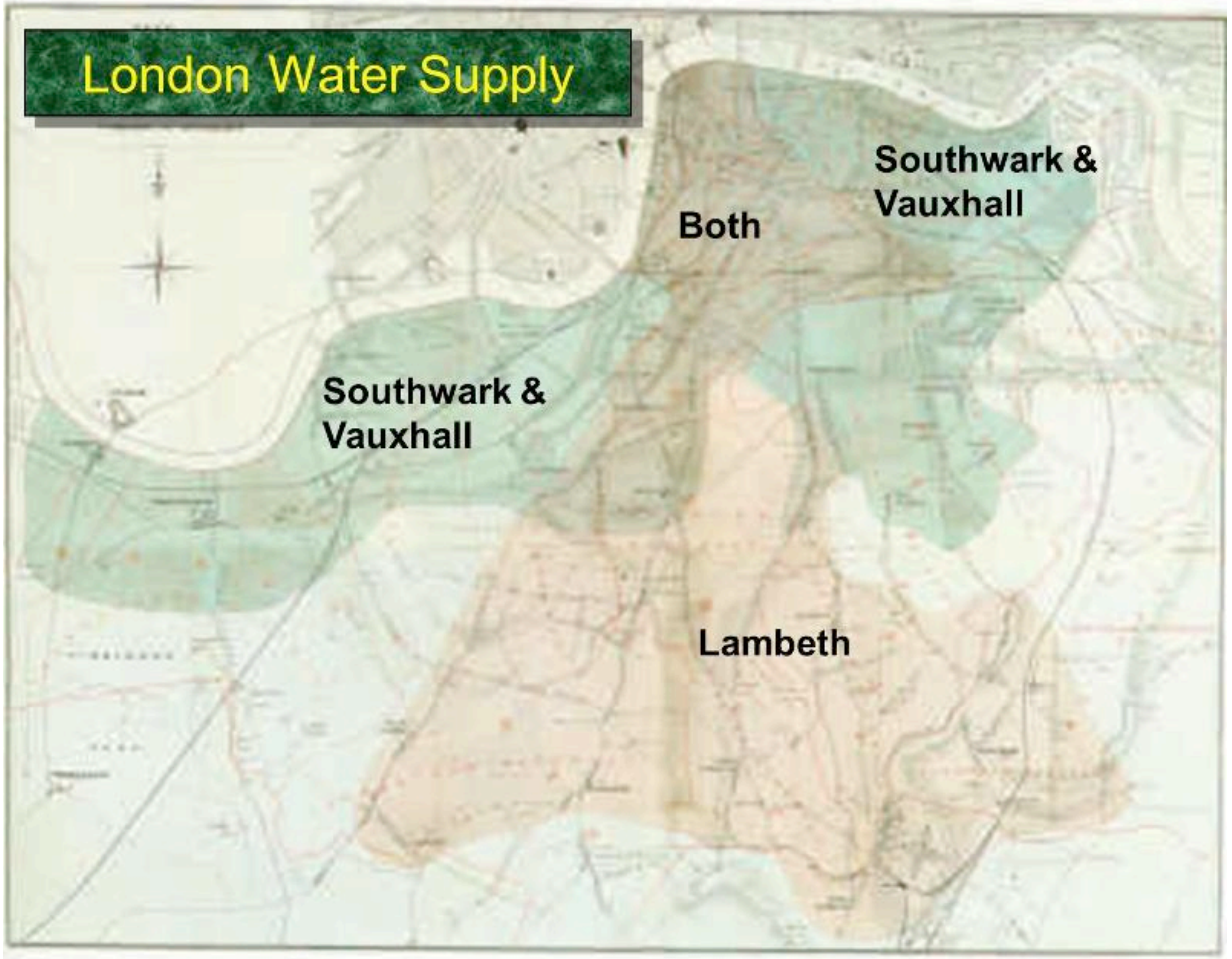
Broad Street Pump

- Snow got pump handle removed
    - It was later discovered a cesspit had been leaking into the well
- Observational study
    - Snow did not control the pump or the people
    - Association was strong

Showing Causation

Snow investigated cases in an area served by two water companies

- Southwark and Vauxhall's intake was in heavily polluted area of the Thames
- Lambeth's water intake was in a less polluted area upstream



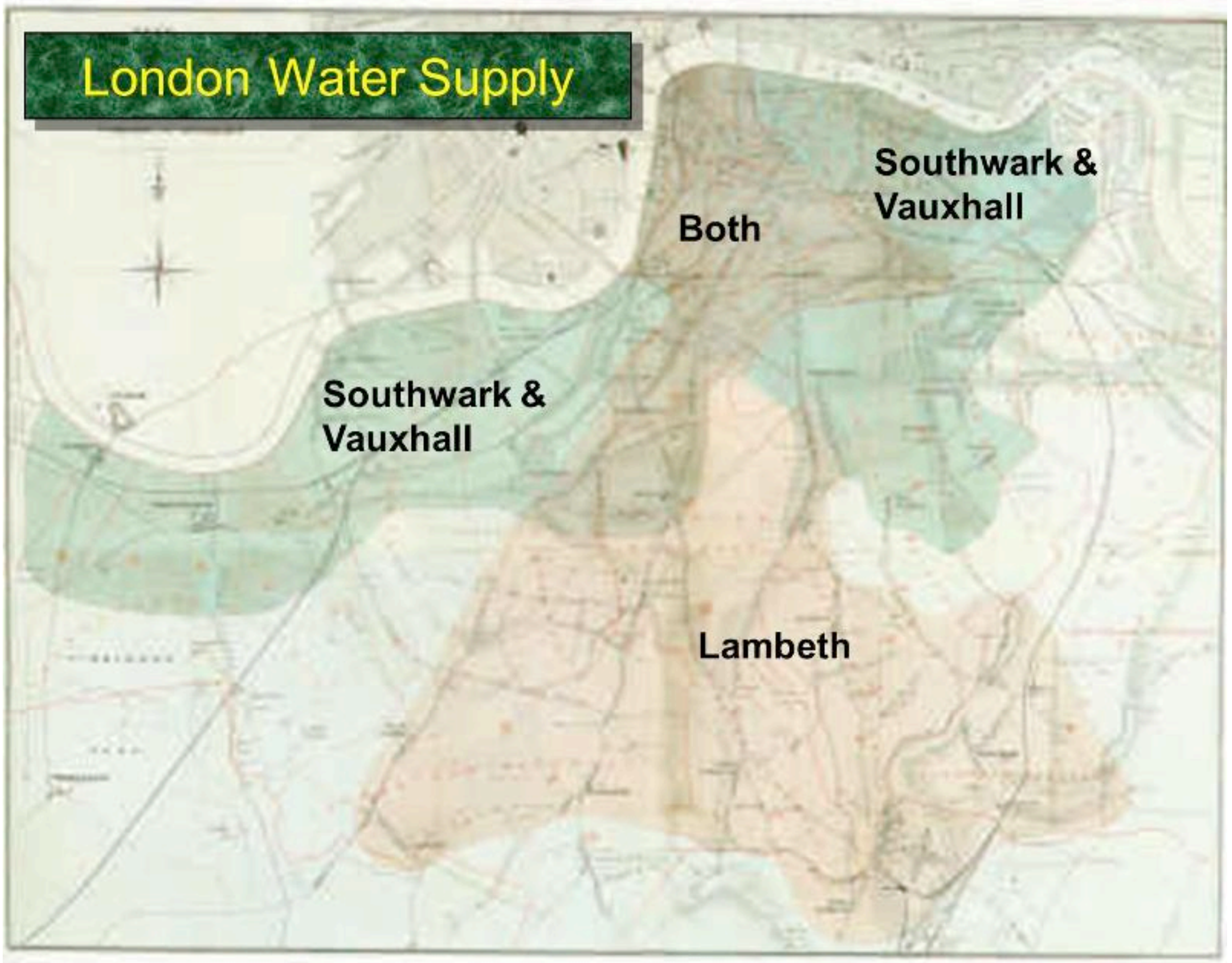
Snow's "Grand Experiment"

Treatment group: S&V

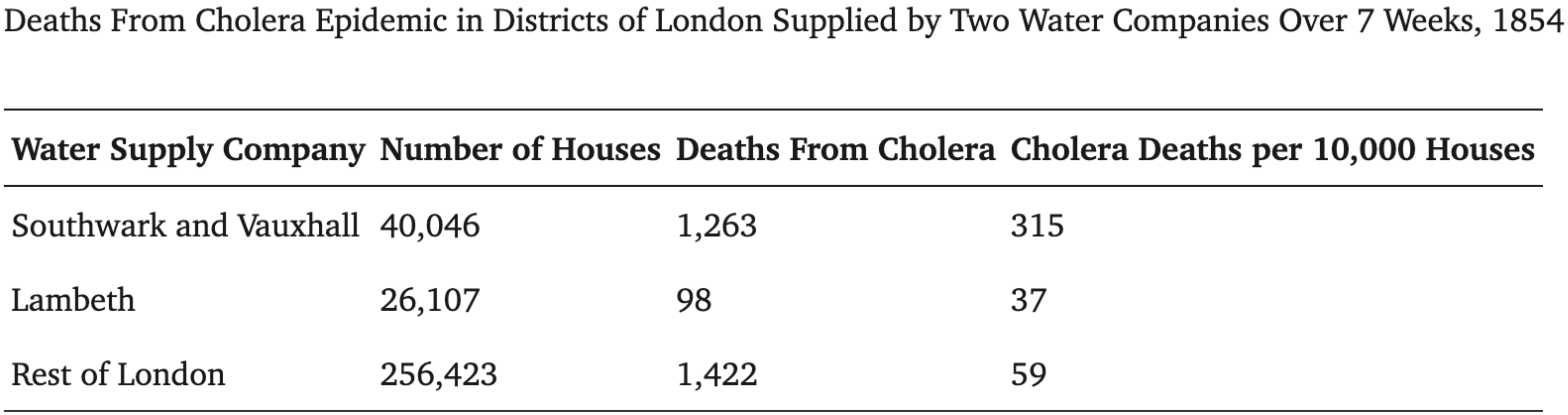
Control group: Lambeth

"… 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 are similar except for the treatment



Snow’s “Grand Experiment”



Deaths From Cholera Epidemic in Districts of London Supplied by Two Water Companies Over 7 Weeks, 1854

Water Supply Company	Number of Houses	Deaths From Cholera	Cholera Deaths per 10,000 Houses
Southwark and Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

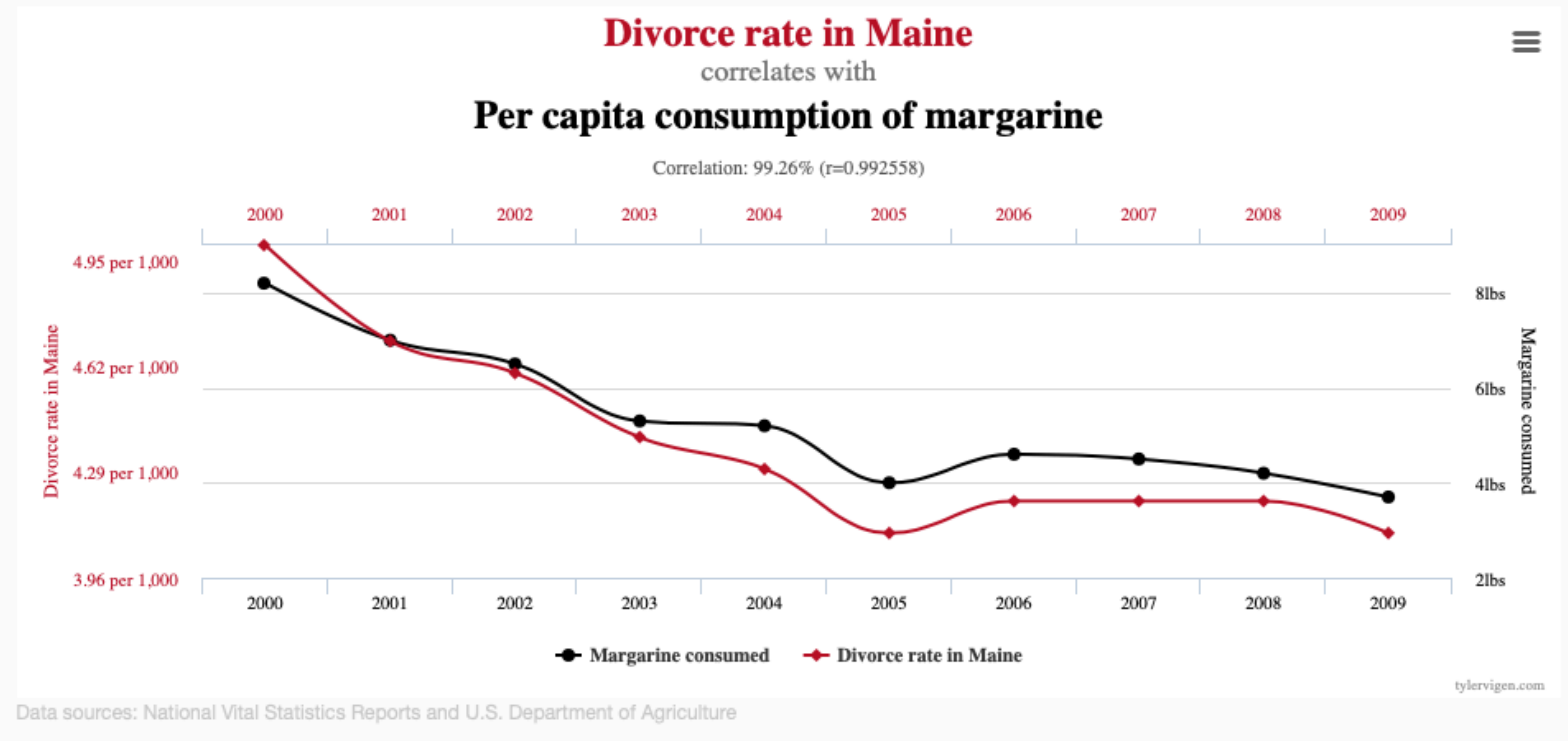
Key to Establishing Causality

- Treatment group: Receives the treatment

- Control group: Does not receive the treatment

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

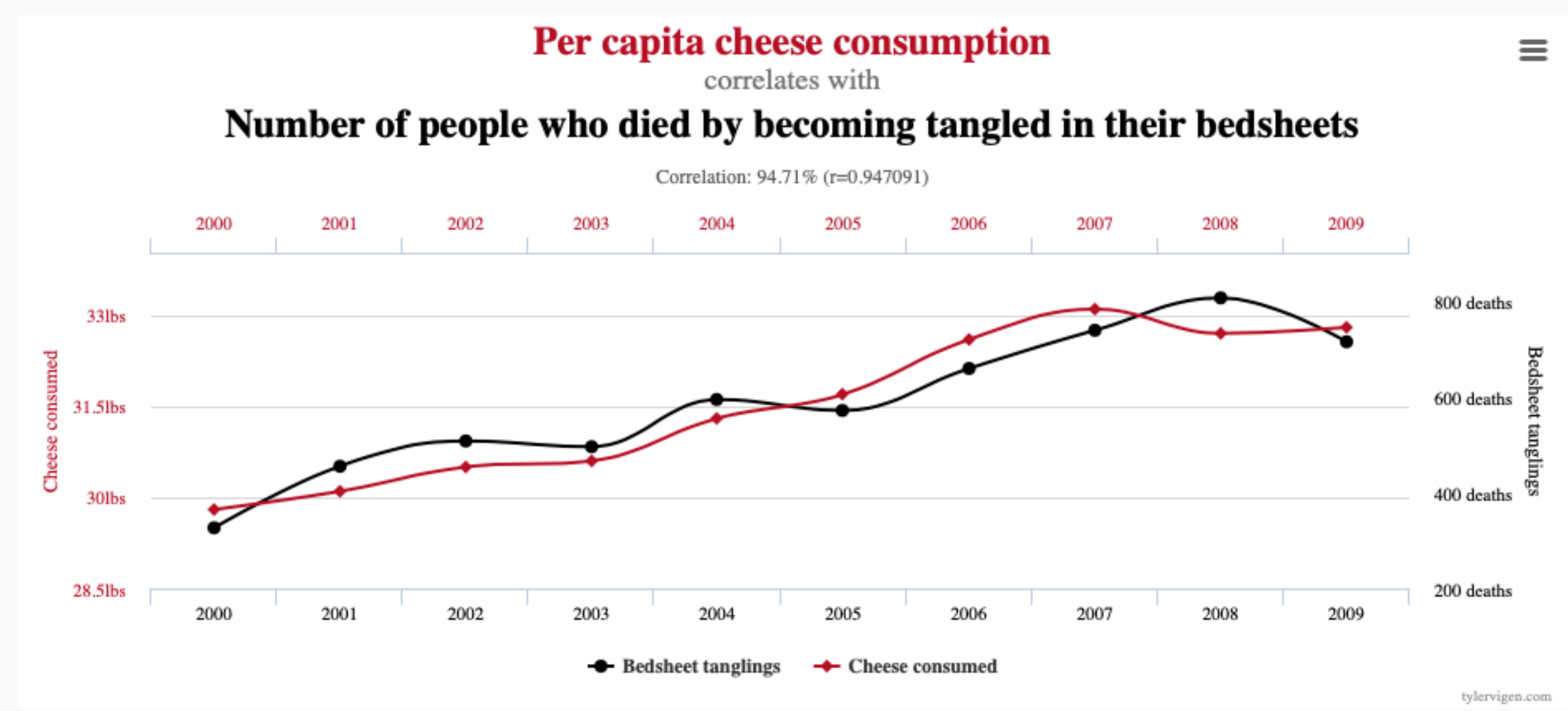
Confounding Factors / Spurious Correlations



Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

Source: https://www.tylervigen.com/spurious-correlations

Confounding Factors / Spurious Correlations



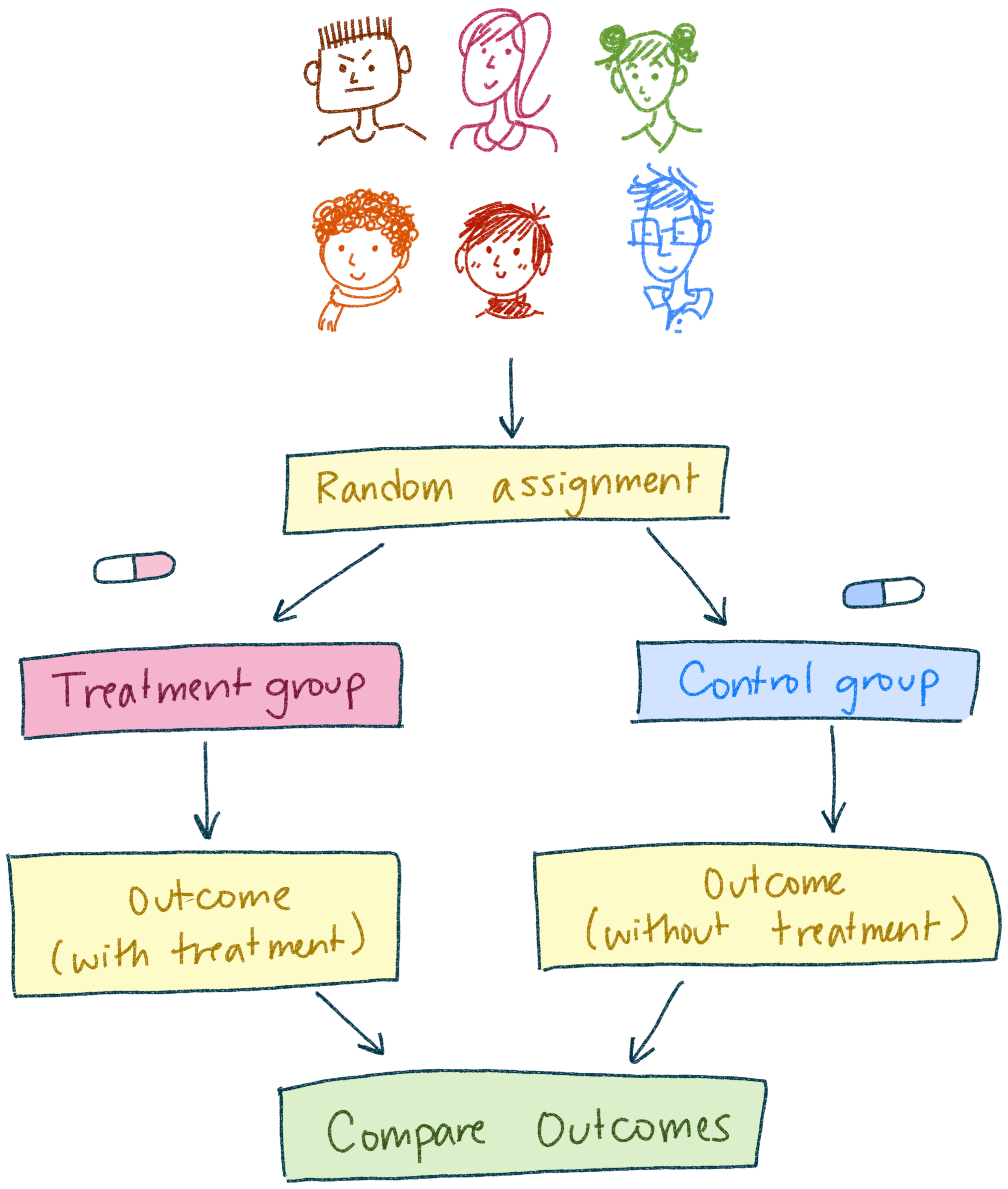
Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention

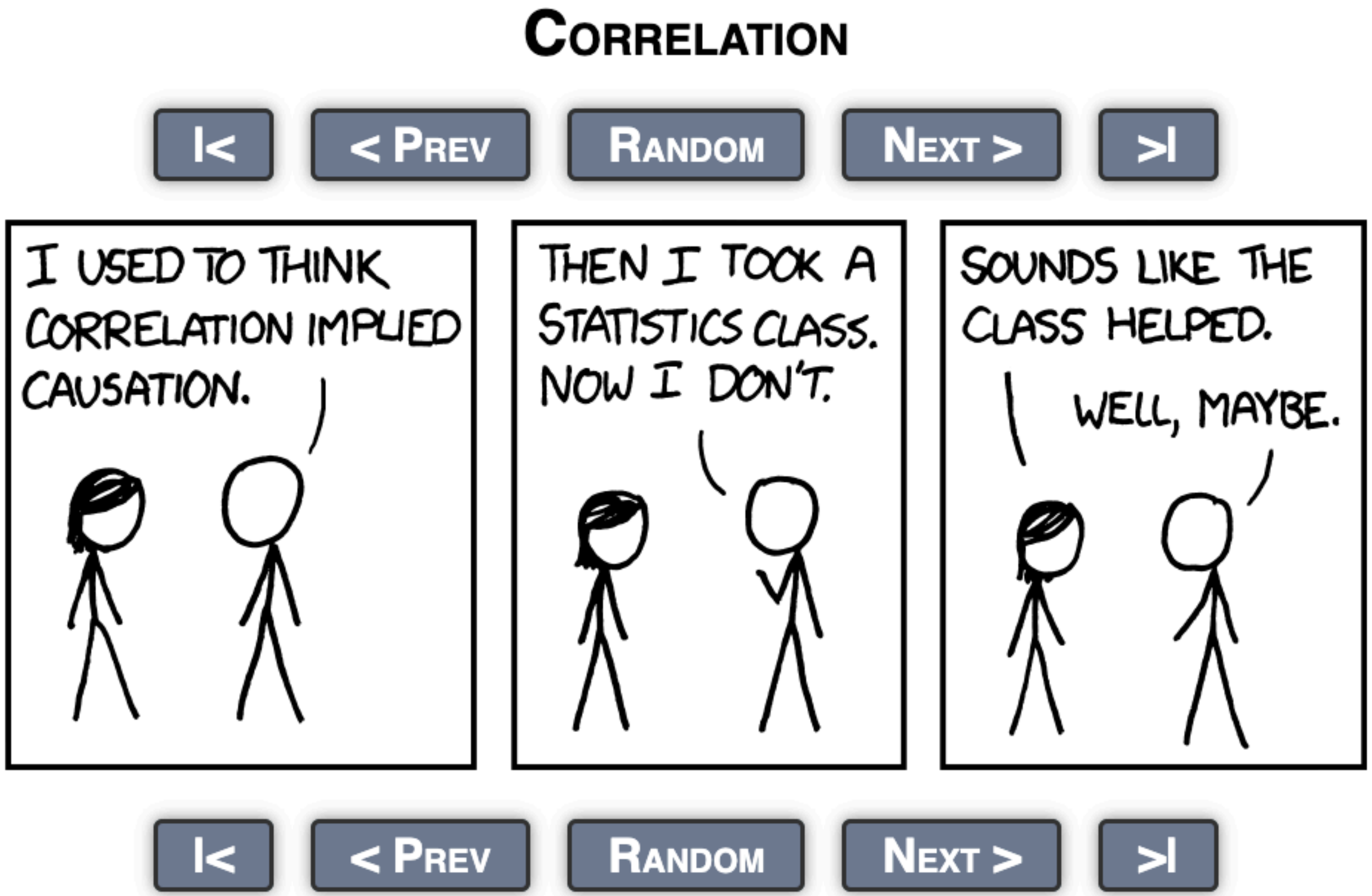
Source: https://www.tylervigen.com/spurious-correlations

How to determine causation vs random correlation?

Randomization!

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




CORRELATION
|<
< PREV
RANDOM
NEXT >
>|
I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.
|<
< PREV
RANDOM
NEXT >
>|

https://xkcd.com/552/

Next Class

- Intro to Python: Expressions and Data Types
- Remember:
    - Labs start next week!!