Lecture 1: Introduction

Sept 3, 2025

Instructor(s)

That's me!

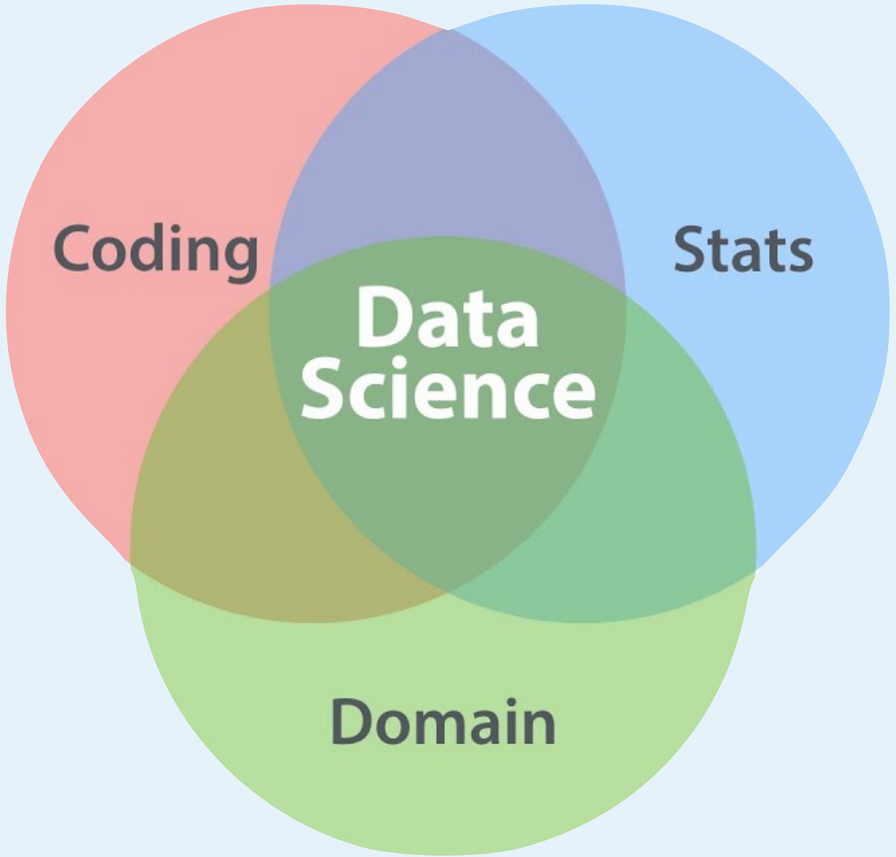
Eysa Lee
eylee@barnard.edu

That's Murad Megjhani, the instructor for the other section of this course!

Lecture Outline

- What is data science?
- Course introduction
  - Topics
  - Teaching Staff
  - Expectations
  - Assignments and Grading
- Demo

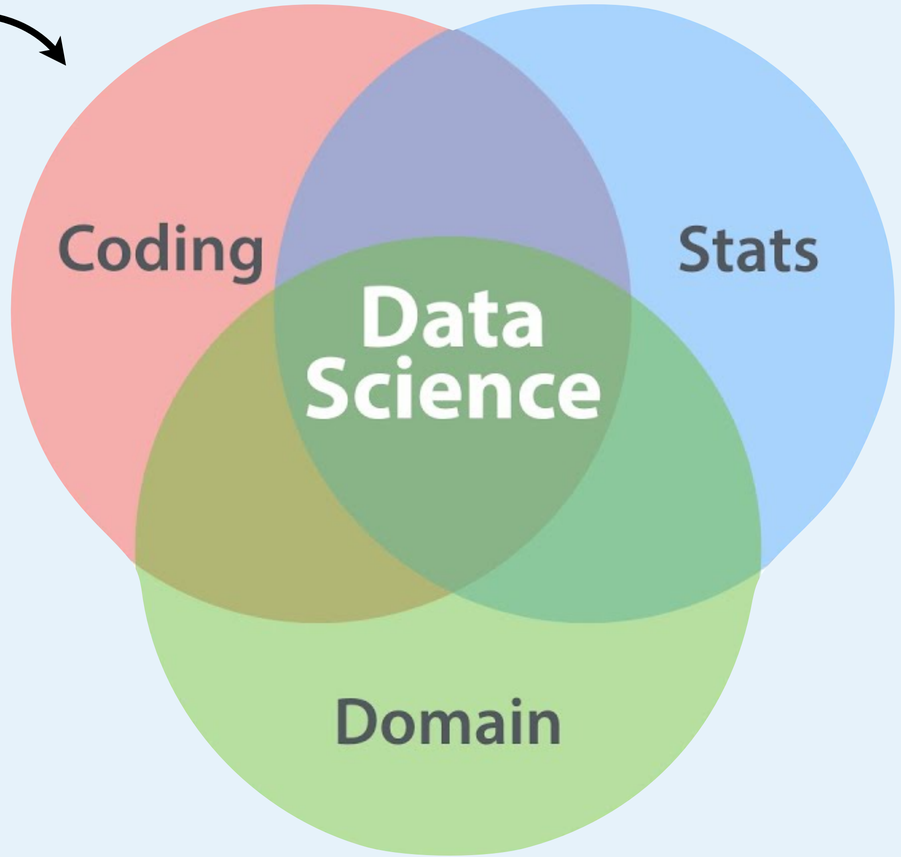
What is Data Science?



What is Data Science?

Prepare &
Analyze Data

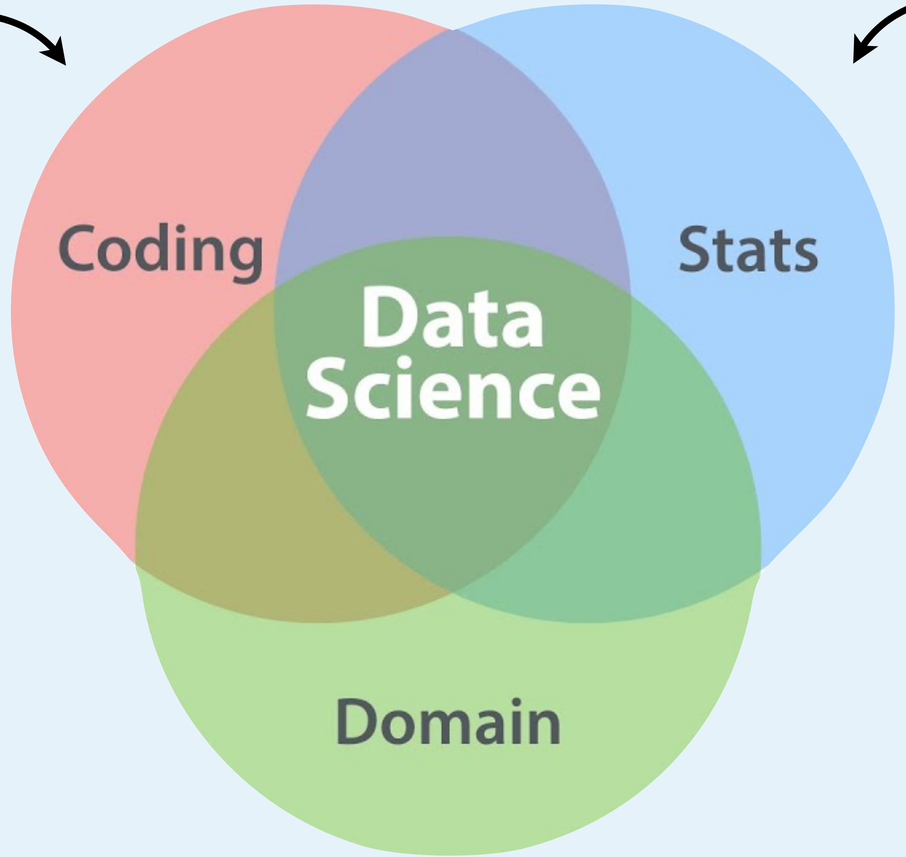
Python &
Computational
Notebooks



What is Data Science?

Prepare &
Analyze Data

Python &
Computational
Notebooks



Probability

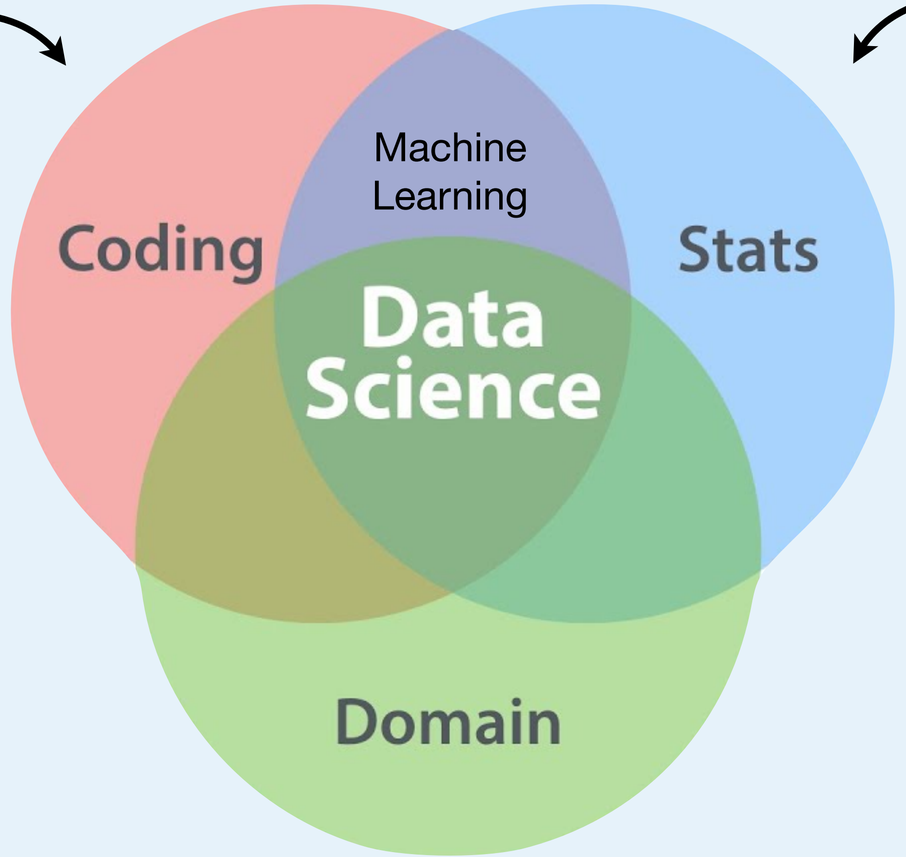
Regression

Confidence

What is Data Science?

Prepare &
Analyze Data

Python &
Computational
Notebooks



Probability

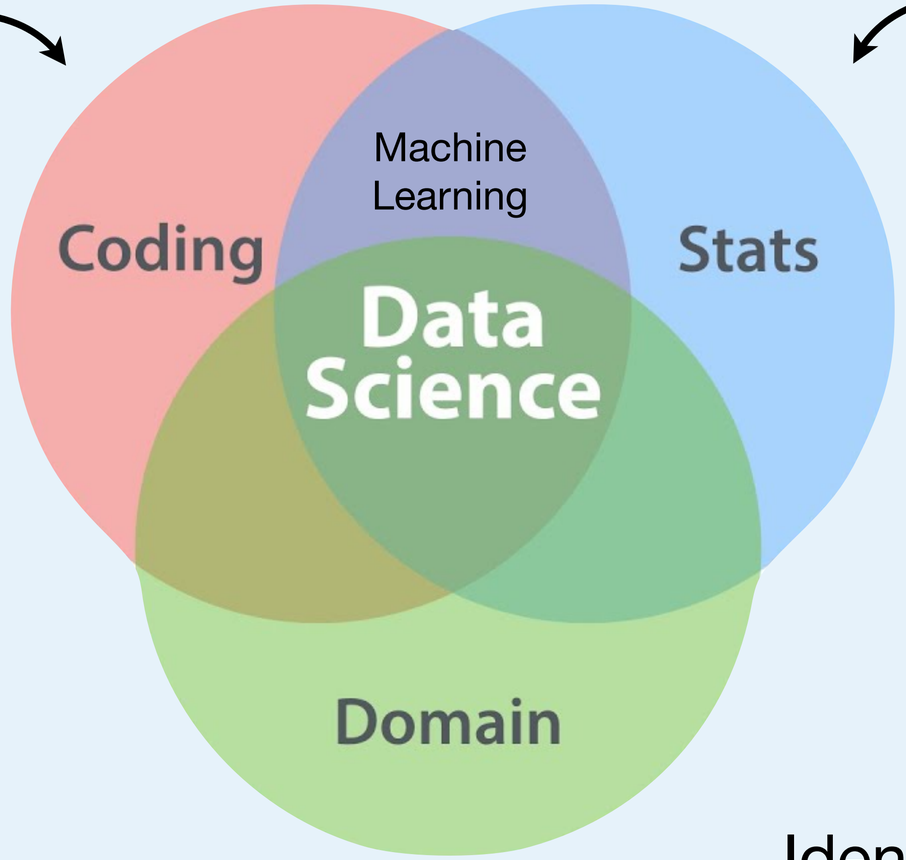
Regression

Confidence

What is Data Science?

Prepare & Analyze Data

Python & Computational Notebooks



Probability

Regression

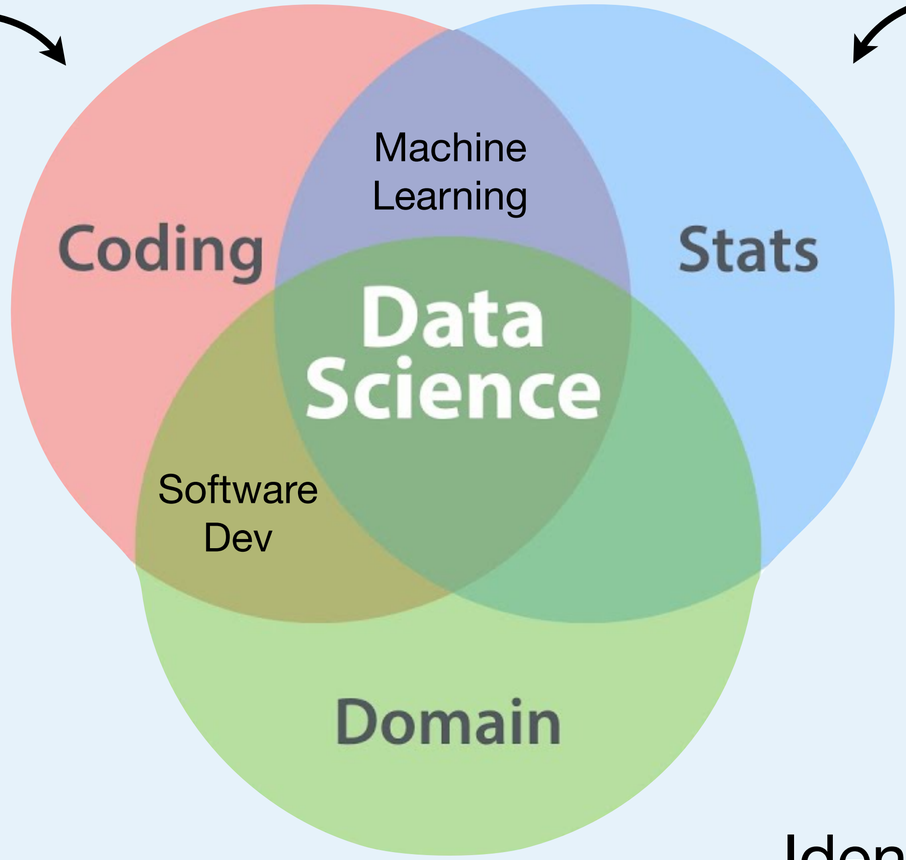
Confidence

Identifying relevant & meaningful applications

What is Data Science?

Prepare & Analyze Data

Python & Computational Notebooks



Probability

Regression

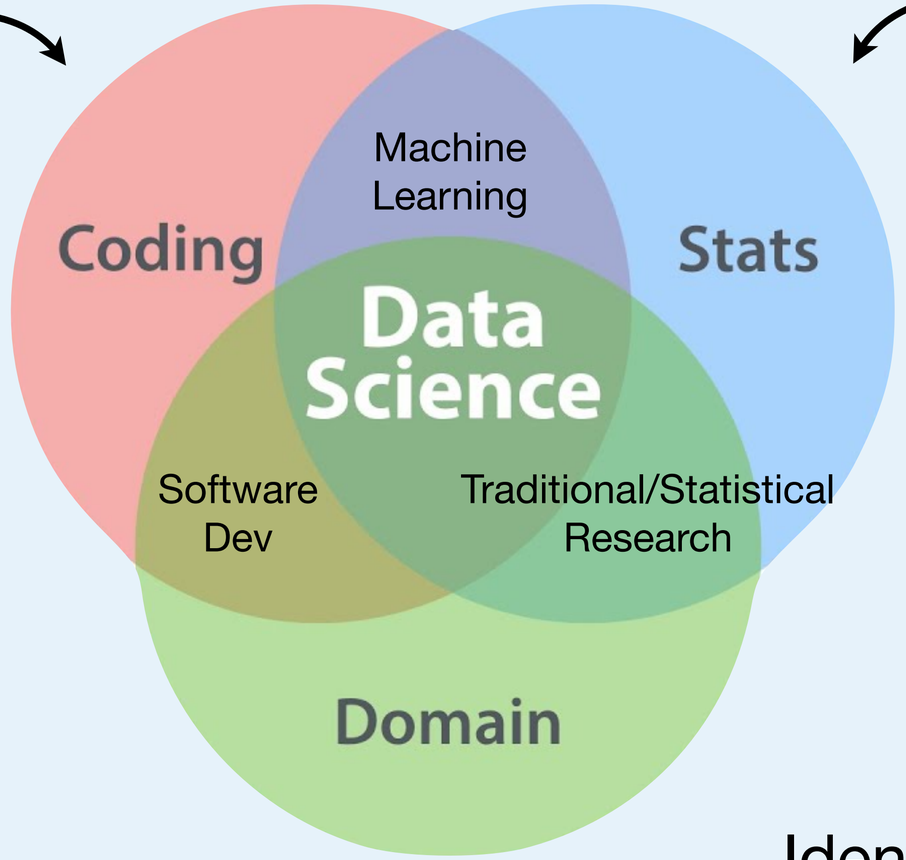
Confidence

Identifying relevant & meaningful applications

What is Data Science?

Prepare &
Analyze Data

Python &
Computational
Notebooks



Probability

Regression

Confidence

Identifying relevant &
meaningful applications

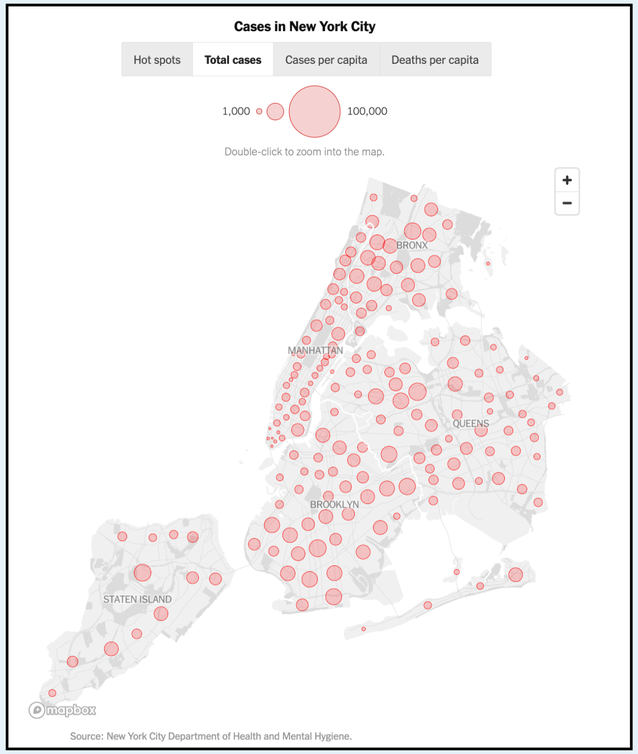
What is Data Science?

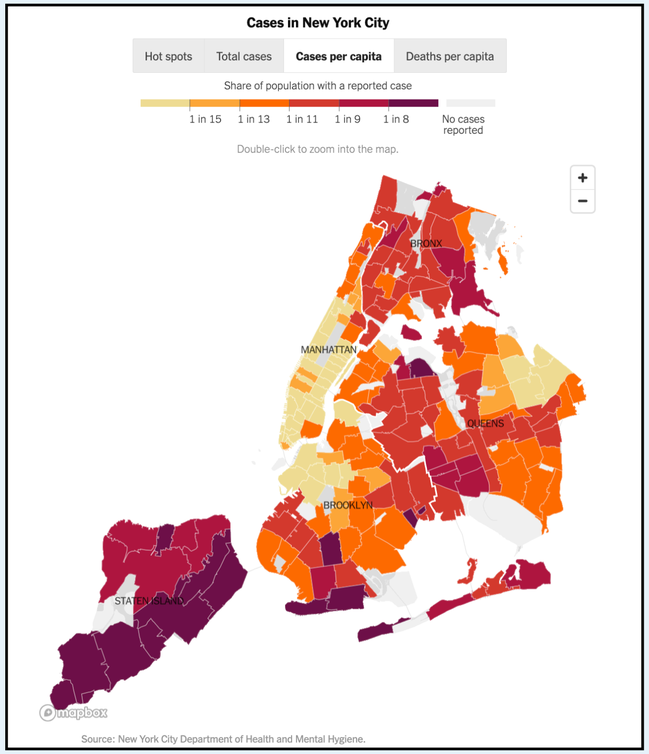
Data science is about drawing useful conclusions from large and diverse data sets through…

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)



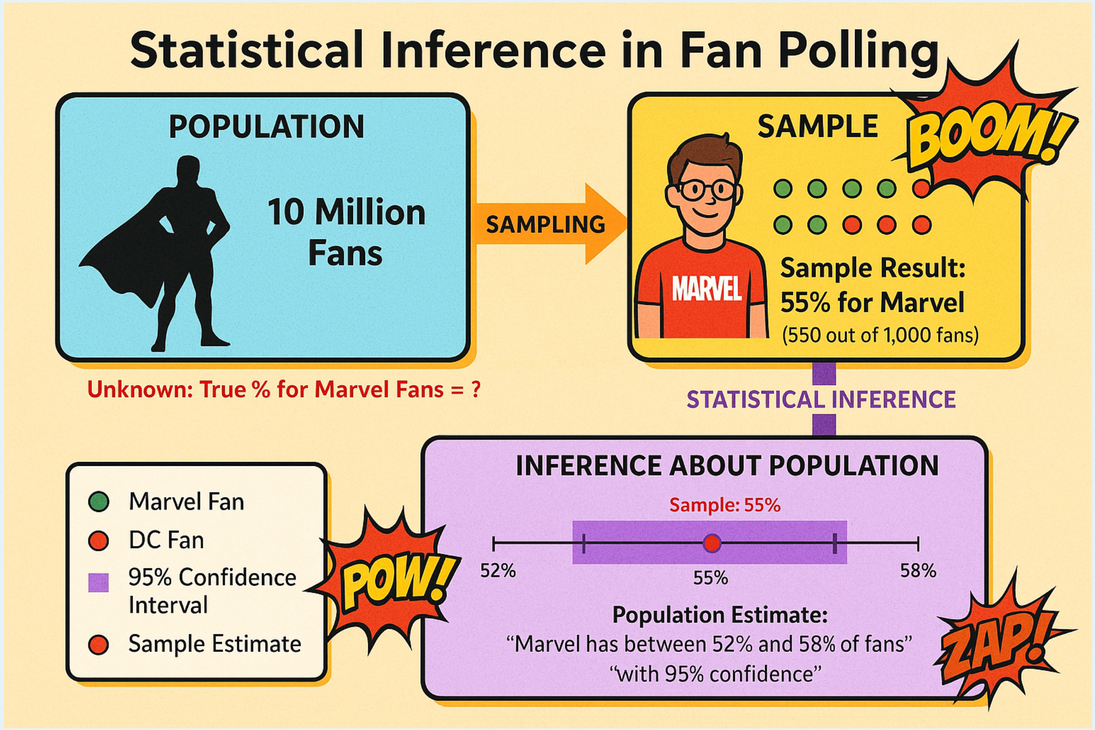


Source: https://www.nytimes.com/interactive/2020/nyregion/new-york-city-coronavirus-cases.html
Data as of May 25, 2021

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

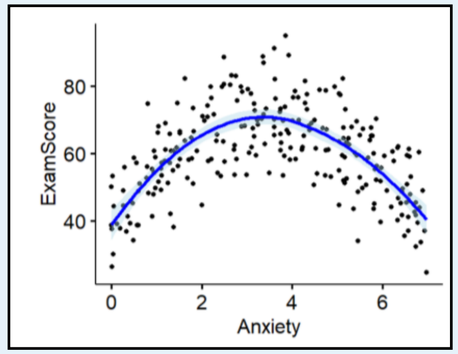


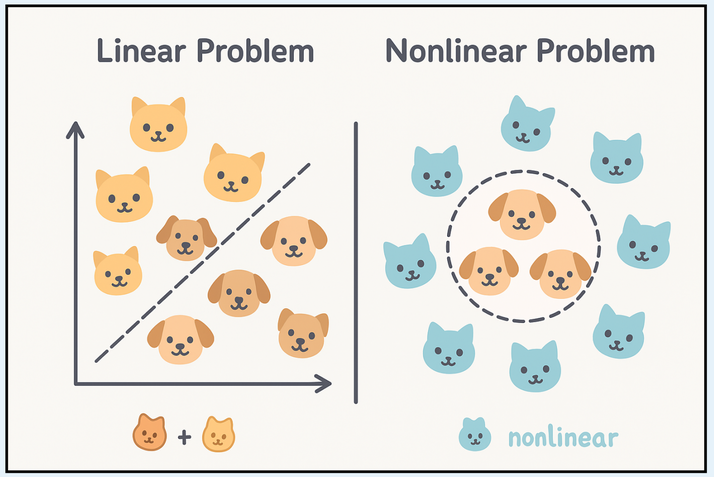
Source: Murad Megjhani and his AI image generator of choice

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

Programming

Midterm Exam

Statistics

Final Project

Programming

Data Types

Iteration

Manipulating Arrays & Tables

Conditionals

Functions

Building Visualizations

Midterm Exam

Statistics

Final Project

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 is this course?

- Introduction to Computational Thinking and Data Science!
  - There is a lab section (BC 1017) associated with this course
  - Labs for both instructors are the same
  - 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.)
  - … including last semester's iteration :)
- Their textbook is great!
  - https://inferentialthinking.com/chapters/intro.html

TAs & Lab Sections

Aditya Vurukala
av3244@columbia.edu

W 2:30-4

Erin Ma
ema@barnard.edu

W 4-5:30

Ken Mah
kmah@barnard.edu

W 5:35-7:05

Amaya Kerjiwal
ark2235@columbia.edu

Th 9:40-11:10

Justin Zeng
jzeng@barnard.edu

Th 11:20-12:50

Computing Fellows

Elena Lukac
epl2135@barnard.edu

Sachi Patel
sdp2173@barnard.edu

Arya Sinha
aps2212@barnard.edu

Riley Stacy
rs4142@barnard.edu

Course Office Hours

- Office hours starting next week
- Professor Lee: Wed 3:00pm-4:00pm (Milstein 512)
- Each TA will offer 1.5 hours of OHs each week
  - Amaya: Tuesdays 2-3:30
  - Erin: Mondays 2:30-4
  - Justin: Wednesdays 4-5:30

Note: labs start *next week* (no lab this week)

Course Expectations: Assignments

- **Lab Assignments** - Due **Fridays 11:59pm** via Courseworks
  - Intended to be finished and submitted during lab itself
- **Homeworks** - Mostly weekly, due **Wednesdays 11:59pm** via Courseworks

The lowest lab and lowest homework grade will be dropped

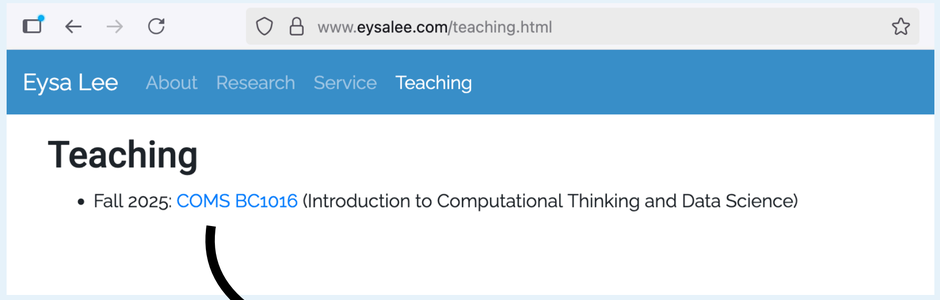
Monday	Tuesday	Wednesday	Thursday	Friday
Lecture Homework Released		Lecture Lab Lab Assignment Released	Lab	
		HW due 11:59pm		Lab due 11:59pm

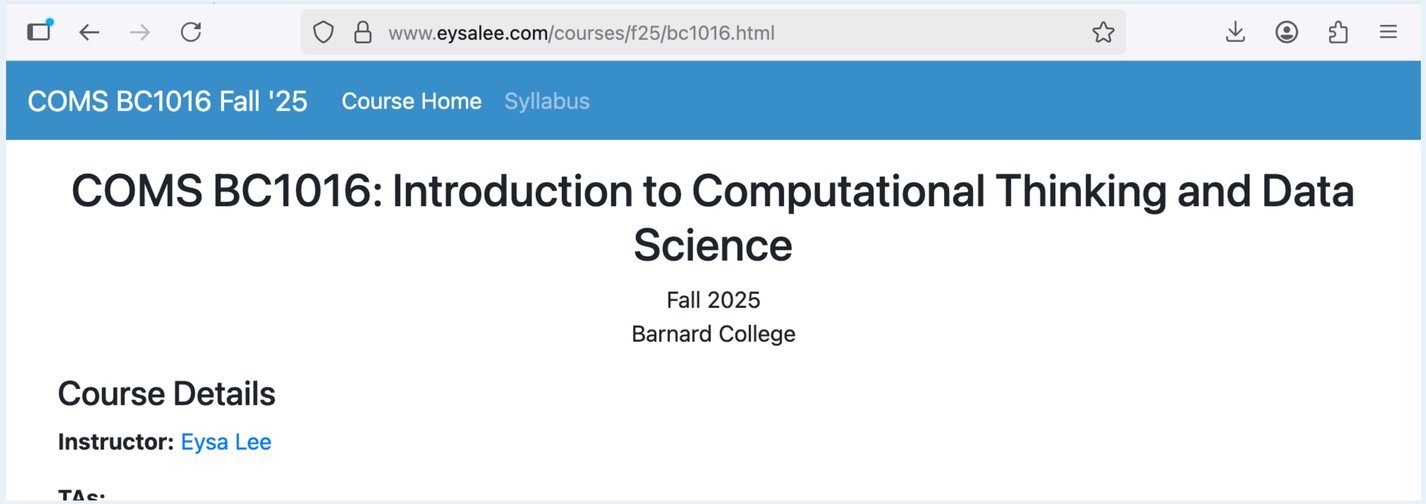
Course Expectations: Assignments

- All assignments will be completed using cloud-based Jupyter notebooks
- You can access our course Jupyter Hub at: **https://bccoms-1016-lee-20253.hub.cuit.columbia.edu** (all you need is a web browser, no special software)

Course Website

- https://www.eysalee.com/courses/f25/bc1016.html





Course Website

- Lecture slides will be posted to the course website
- Class Jupyter demos will be posted after class (end of day Monday / Wednesday)

Course Expectations: 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

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!).
- For your final project, AI generated text is not permitted as part of your written descriptions in your final report.

Midterm Exam

- Paper exam happens during class October 20, 2025
- You will be allowed a 7x5 notecard with notes to use as a reference during the exam

Final Project

- Groups of 2
- 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 TBA

Notebook Demo

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

What did we do?

- Import Data
- Create tables
- Analyze character counts
- Write functions
- Create array and iterate on it
- Build visualizations
- Compare data

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

Next Week

- Cause & Effect
- Tables
- Labs start!