

Data Science

“Computational Statistics Across the Sciences”
CAS PY 191S 1310

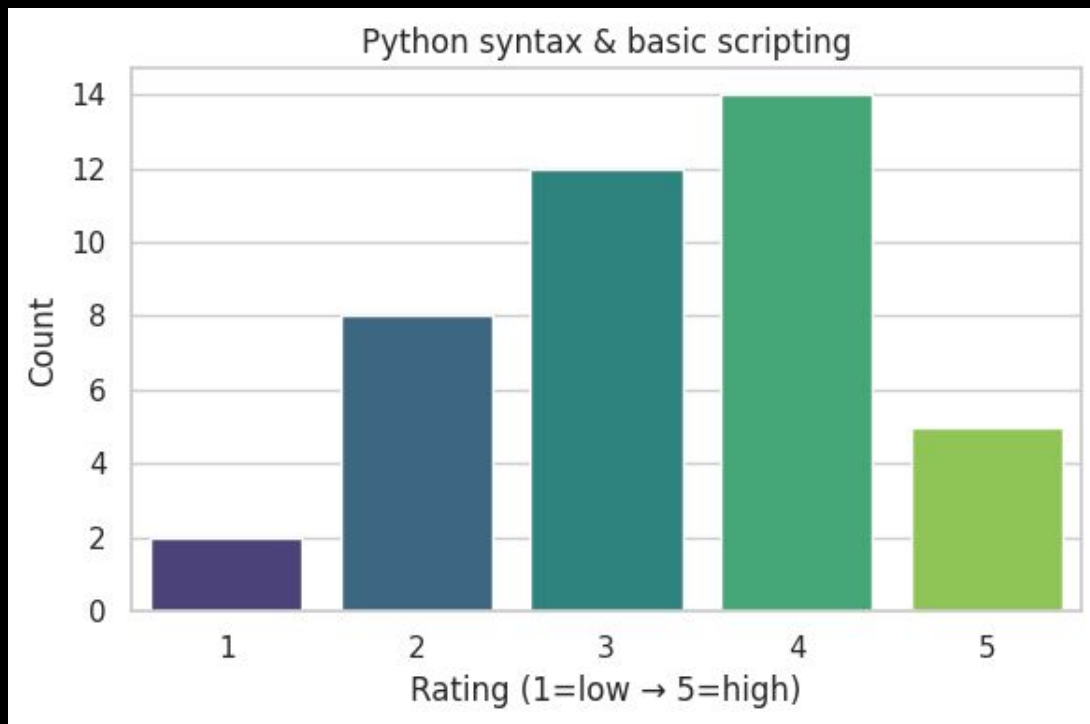
Attendance



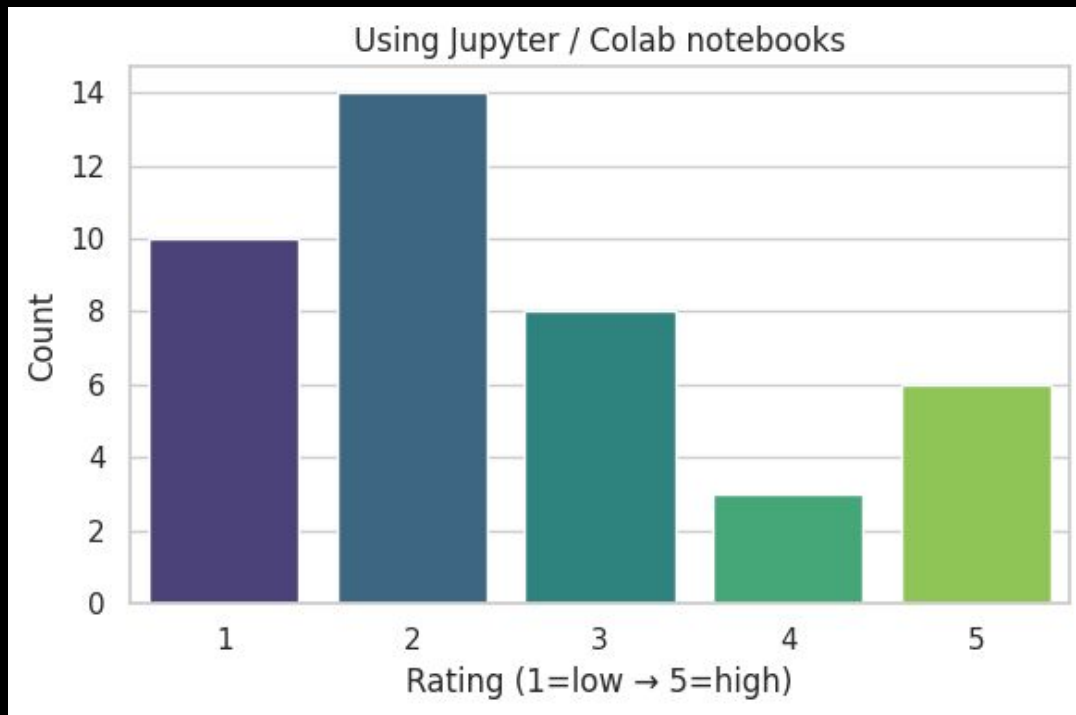
<https://forms.gle/iX5EvN6pLrqi2rej8>

Introductions

A bit about our class



A bit about our class



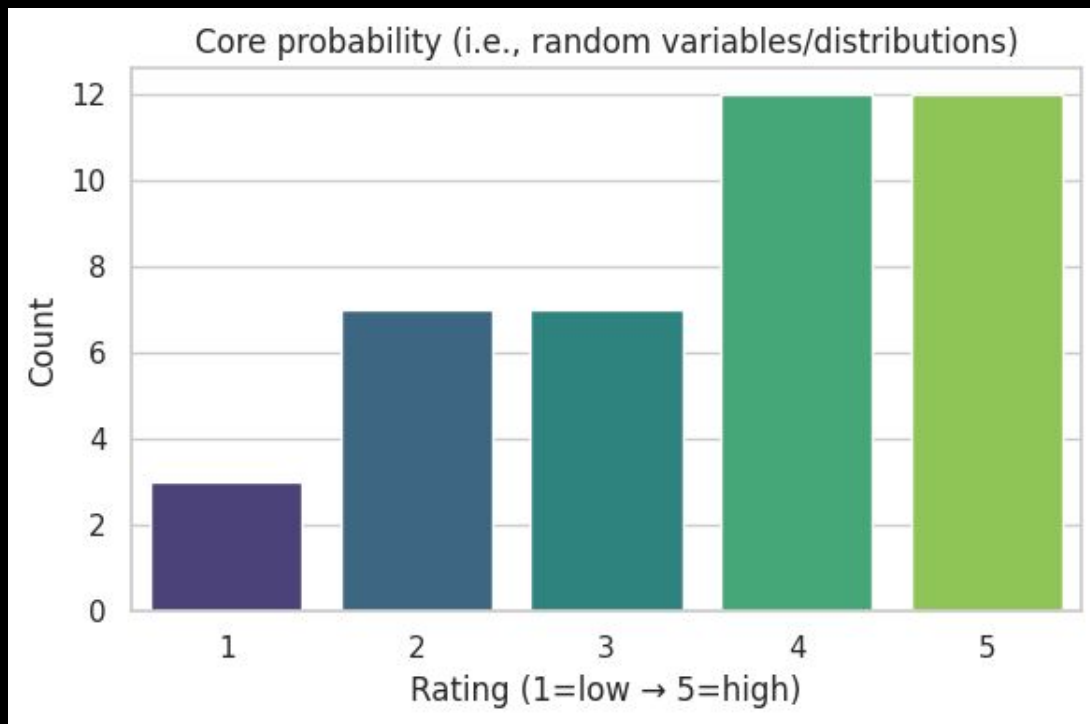
A bit about our class



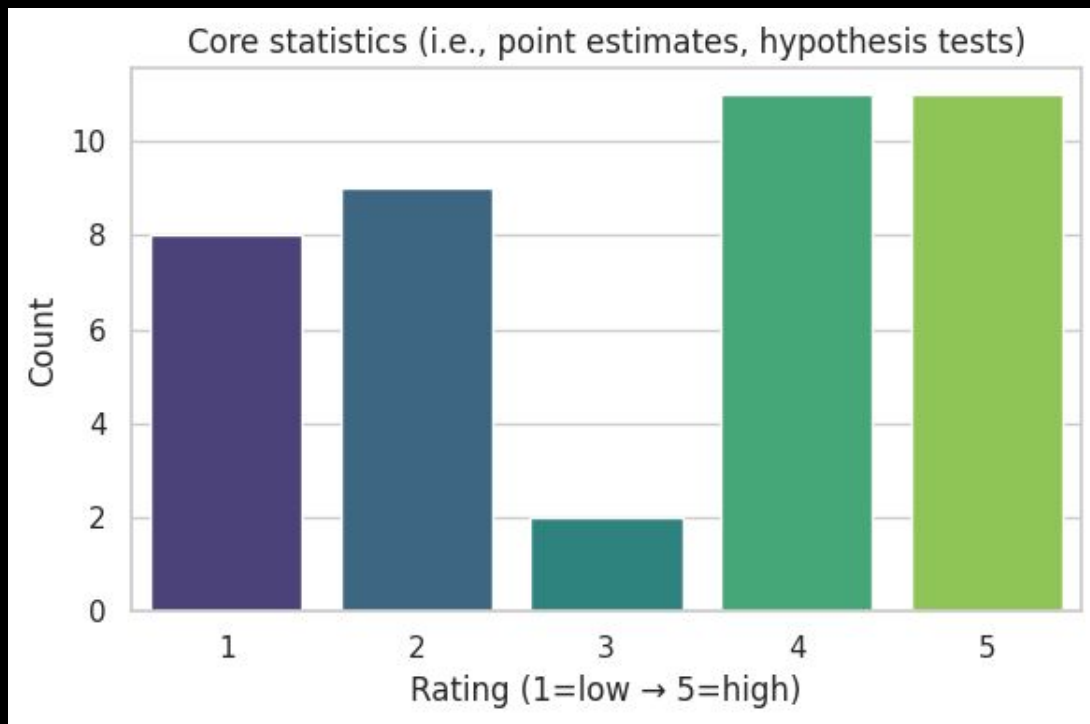
A bit about our class



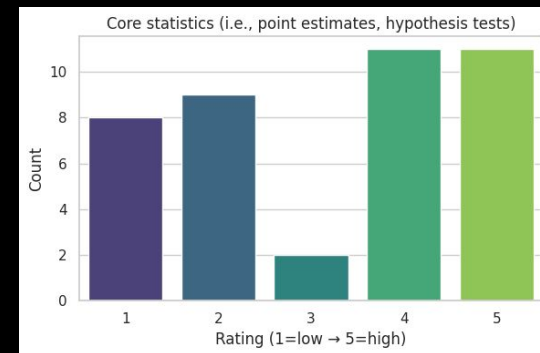
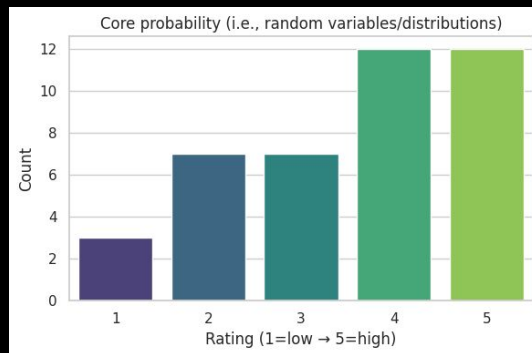
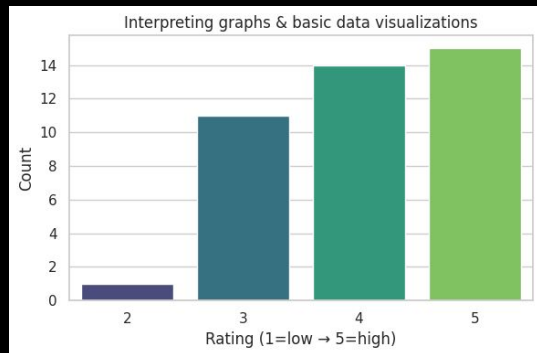
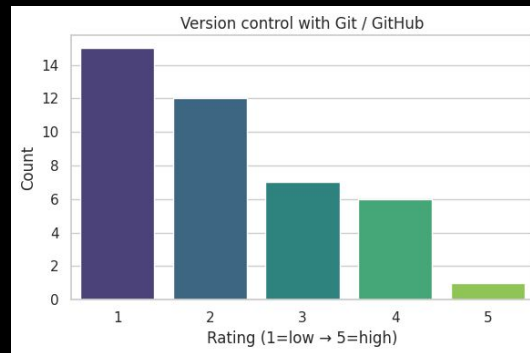
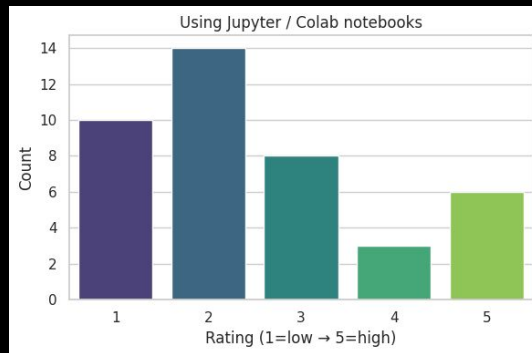
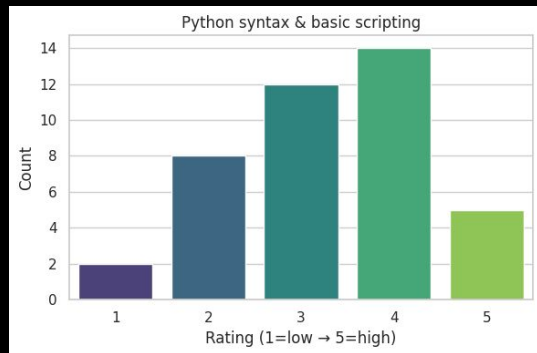
A bit about our class



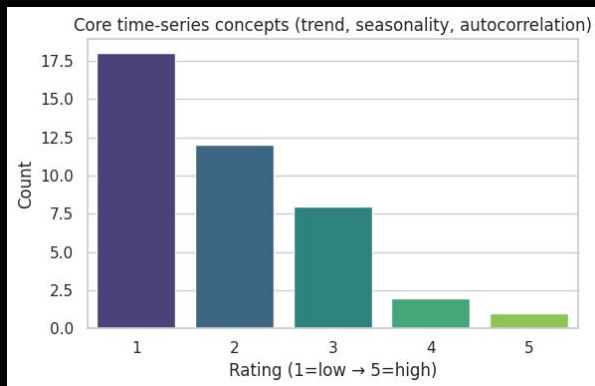
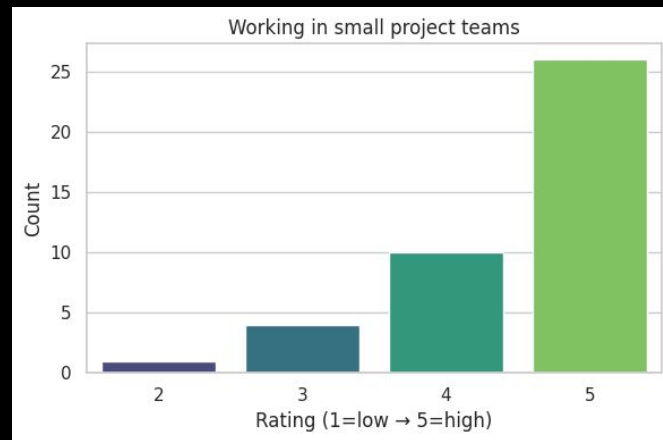
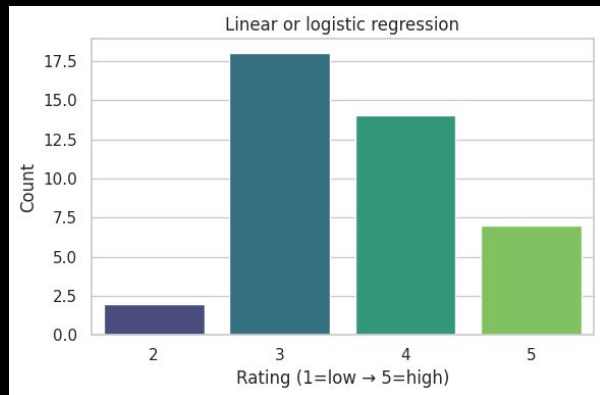
A bit about our class



A bit about our class



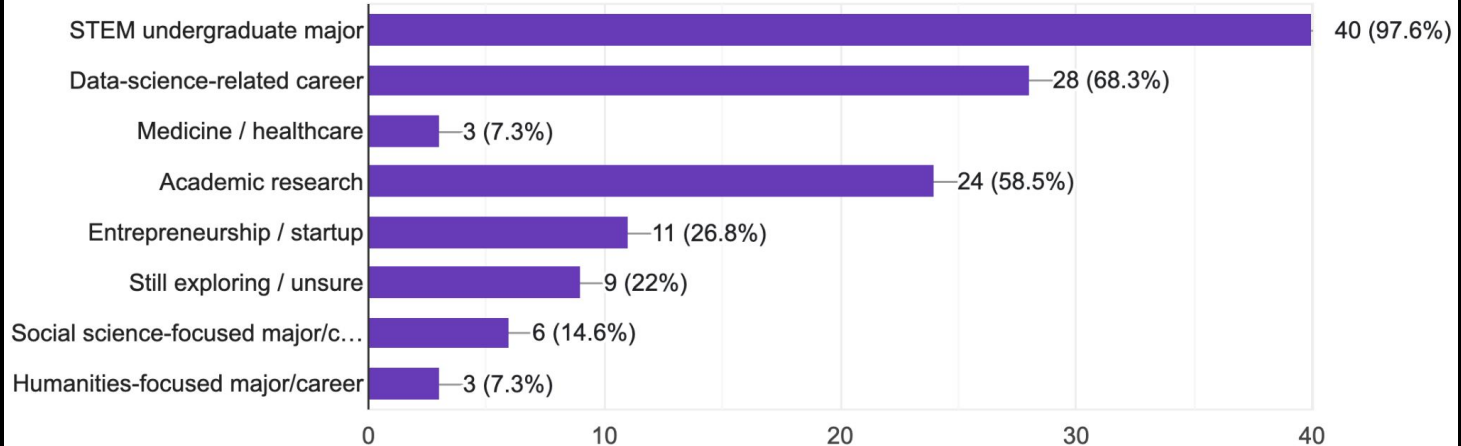
A bit about our class



A bit about our class

What are your long-term goals (choose all that apply).

41 responses

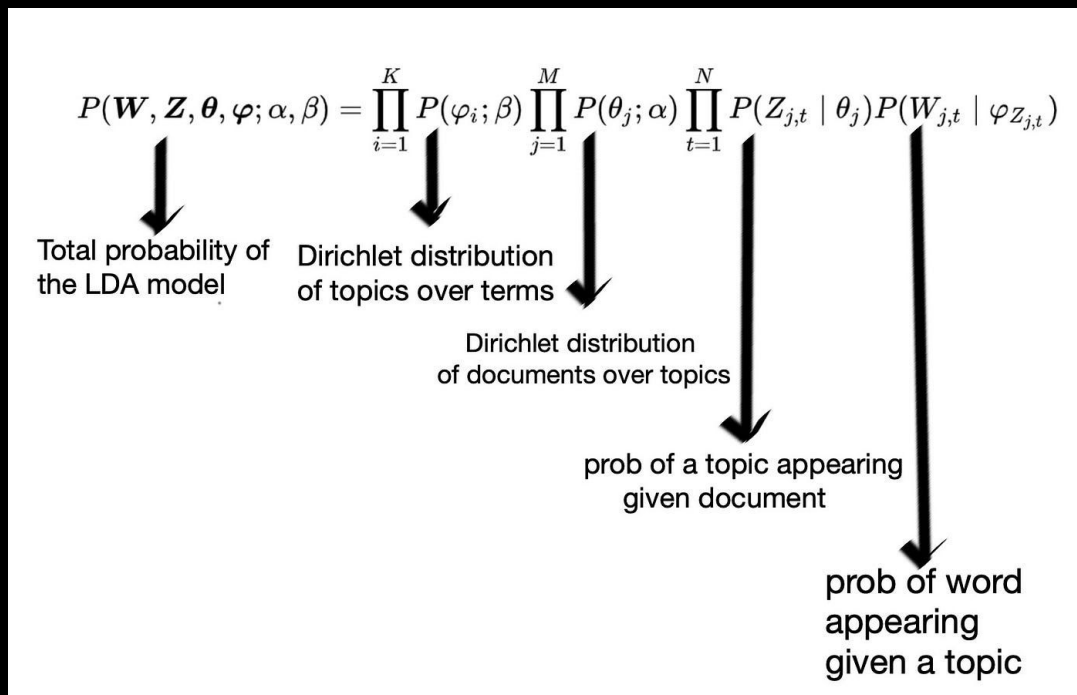


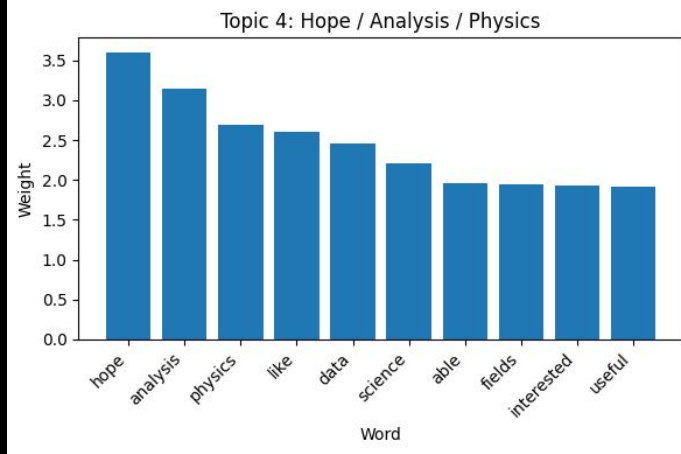
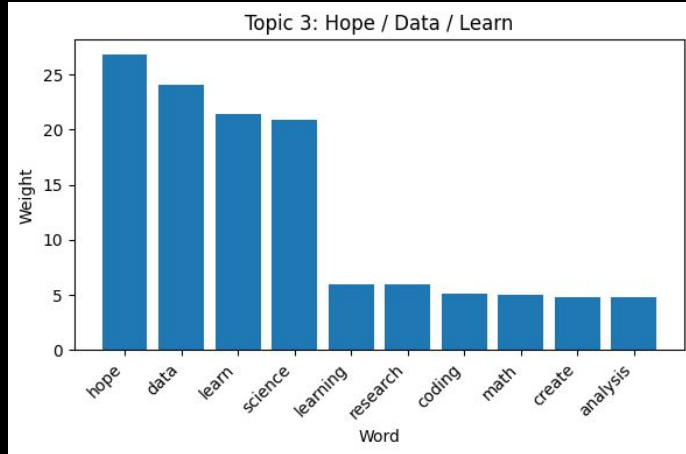
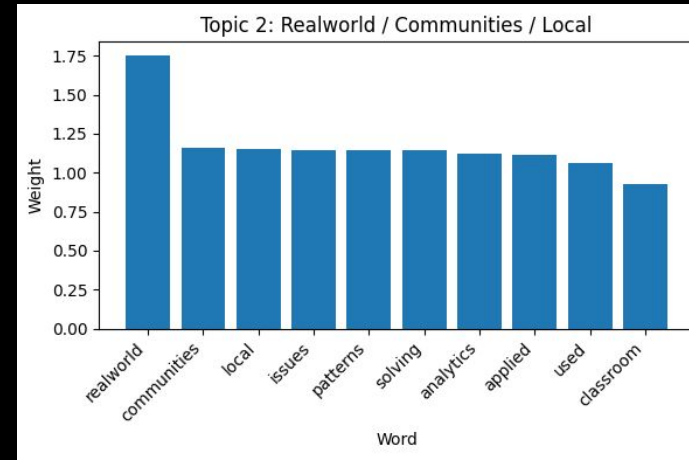
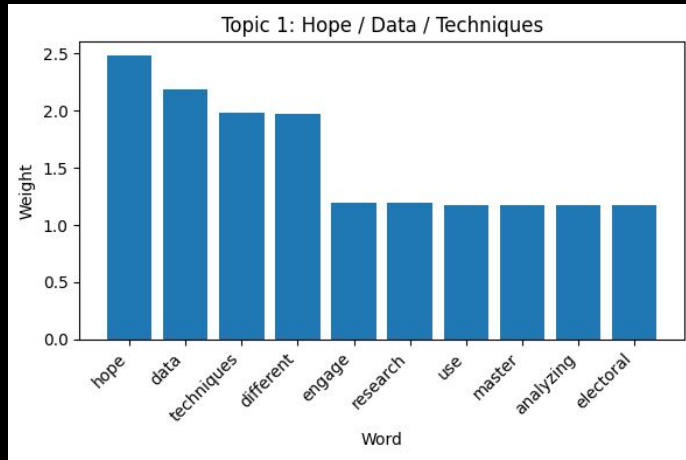
Open response questions:

- 1) In 1–2 sentences, what do you hope to create or learn during this course?
- 2) In 1–2 sentences, what do you hope to get out of the hands on research/computational project for this course?

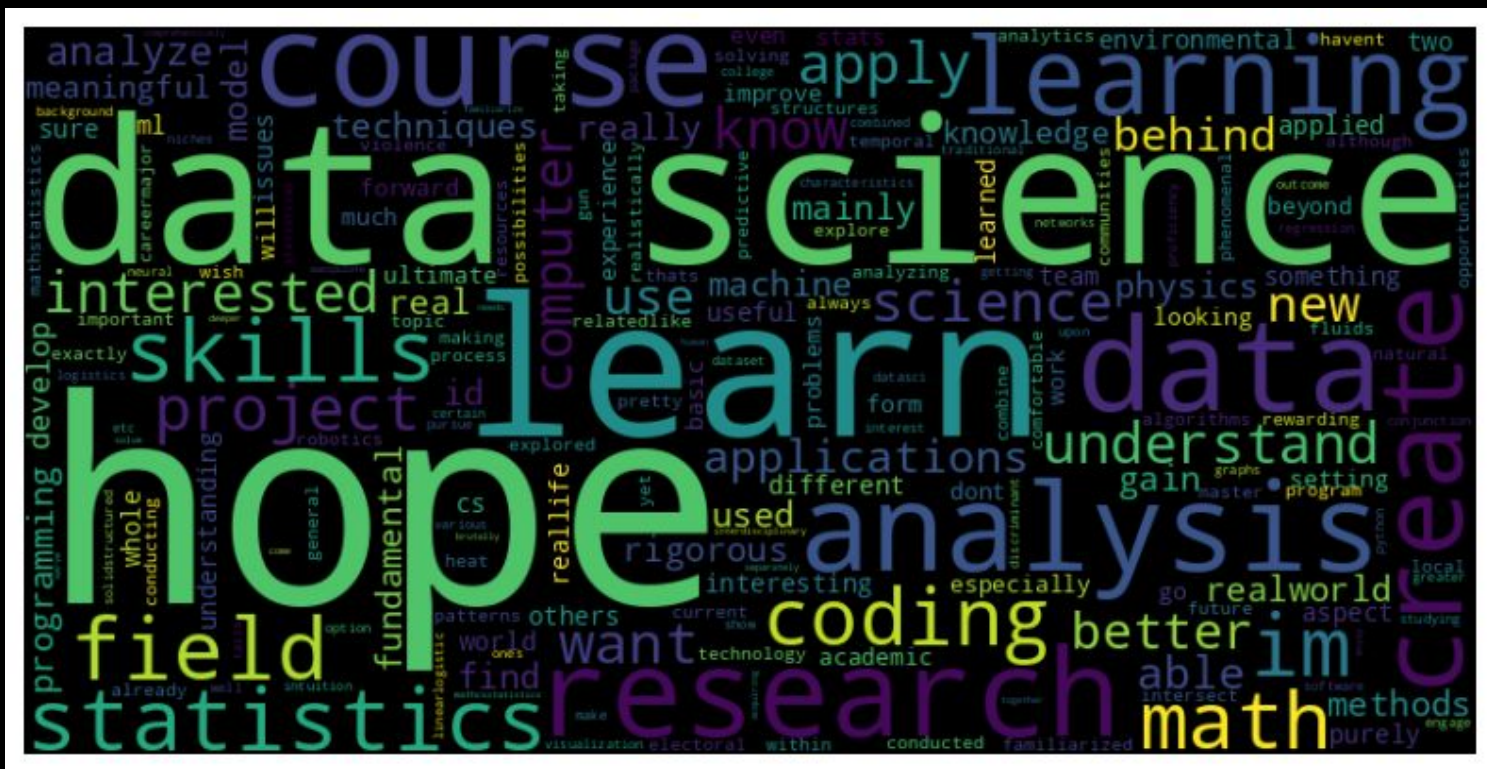
Topic Modeling (Latent Dirichlet Allocation)

- Bayesian probabilistic model used for topic modeling in natural language processing
- It assumes that each document is a mixture of different topics, and each topic is a mixture of words.
- Helpful for identifying core themes





Most commonly used words in 1-2 sentence reflection



Recall our pre class survey: “What are your goals for this summer?”

- Learning to code from scratch?
- Learning data science best practices?
- Learn about a potential college major or career?
- Learn how it feels to be at an R1 university?
- If you’re a residential student, what it feels like to live away from home?
- ...



Learning Objectives for the Course

What are my overall goals for you?

- Learn how to be **kind** and **humble** in a technical environment.
 - How do you give good feedback
 - How do we make diverse educational backgrounds and interests a strength
 - How do we decouple knowing a lot/the right answer from being valuable

What are my overall goals for you?

- Learn how to be **kind** and **humble** in a technical environment.
 - How do you give good feedback
 - How do we make diverse educational backgrounds and interests a strength
 - How do we decouple knowing a lot/the right answer from being valuable
- Learn how to be **rigorous** computational statisticians in a world of generative AI.
 - Core skills in coding, testing, computational reproducibility, statistics, and communication

What are my overall goals for you?

- Learn how to be **kind** and **humble** in a technical environment.
 - How do you give good feedback
 - How do we make diverse educational backgrounds and interests a strength
 - How do we decouple knowing a lot/the right answer from being valuable
- Learn how to be **rigorous** computational statisticians in a world of generative AI.
 - Core skills in coding, testing, computational reproducibility, statistics, and communication
- Learn about the value of **Open Access, Open Source** data science projects

Technical Skills

- Strong conception of the data science life cycle
- Strong Python coding skills
 - Basic data types, object oriented programming
 - Unit tests
 - Version control and project maintency through Github
 - Basic code review
- Basic statistical skills in major areas of the field
 - **Learn how to learn probability and statistics**
- Responsible project communication

Core Team – Two instructors

Patrick Bloniasz



Dr. Eugene Pinsky



Core Team – Four Teaching Fellows (Kevin)



Kevin

Tharunya



Tejovan

Zhengyang

About Me



- Undergraduate: Neuroscience, Digital and Computational Studies
- PhD Candidate in Computational Neuroscience (all but dissertation)

About Me



- Undergraduate: Neuroscience, Digital and Computational Studies
- PhD Candidate in Computational Neuroscience (all but dissertation)
- Previous teaching experience:
 - Adjunct Faculty and Researcher at New York University (Cognitive Science)
 - RISE Neurobiology Dry and Wet Lab Teaching Fellow (2 years)
 - Teaching Fellow at Bowdoin (Biomathematics, Biostatistics, Mathematical Statistics, Intro to DCS; 3 years)

About Me



- Undergraduate: Neuroscience, Digital and Computational Studies
- PhD Candidate in Computational Neuroscience (all but dissertation)
- Previous teaching experience:
 - Adjunct Faculty and Researcher at New York University (Cognitive Science)
 - RISE Neurobiology Dry and Wet Lab Teaching Fellow (2 years)
 - Teaching Fellow at Bowdoin (Biomathematics, Biostatistics, Mathematical Statistics, Intro to DCS; 3 years)
- Research Themes:
 - Stochastic Processes (Time Series Analysis); Point Process/Queuing Theory
 - Statistical Signal Processing
 - Electrical Field Potentials, Electroencephalography
 - Anesthesia (Propofol, Ketamine, Sevoflurane)

How to think of me as an academic resource

Even if we don't overlap in career paths, undergraduate styles of education, or research areas, there are still ways to grow:

- Connect you to researchers who **do** have the experiences you're interested in
- Provide you with guidance on how to enter into a field
- Identify skills required to get onto a particular path

Advice: Talk to your afternoon instructor and teaching fellows

Group Discussion: What are data?

Data are ...

Data are individual facts, statistics, or items of information—usually numeric or categorical—that represent measurements or observations about the world.

Data serve as the **foundational input for analysis, modeling, and decision-making**



Data are ...

Data are individual facts, statistics, or items of information—usually numeric or categorical—that represent measurements or observations about the world.

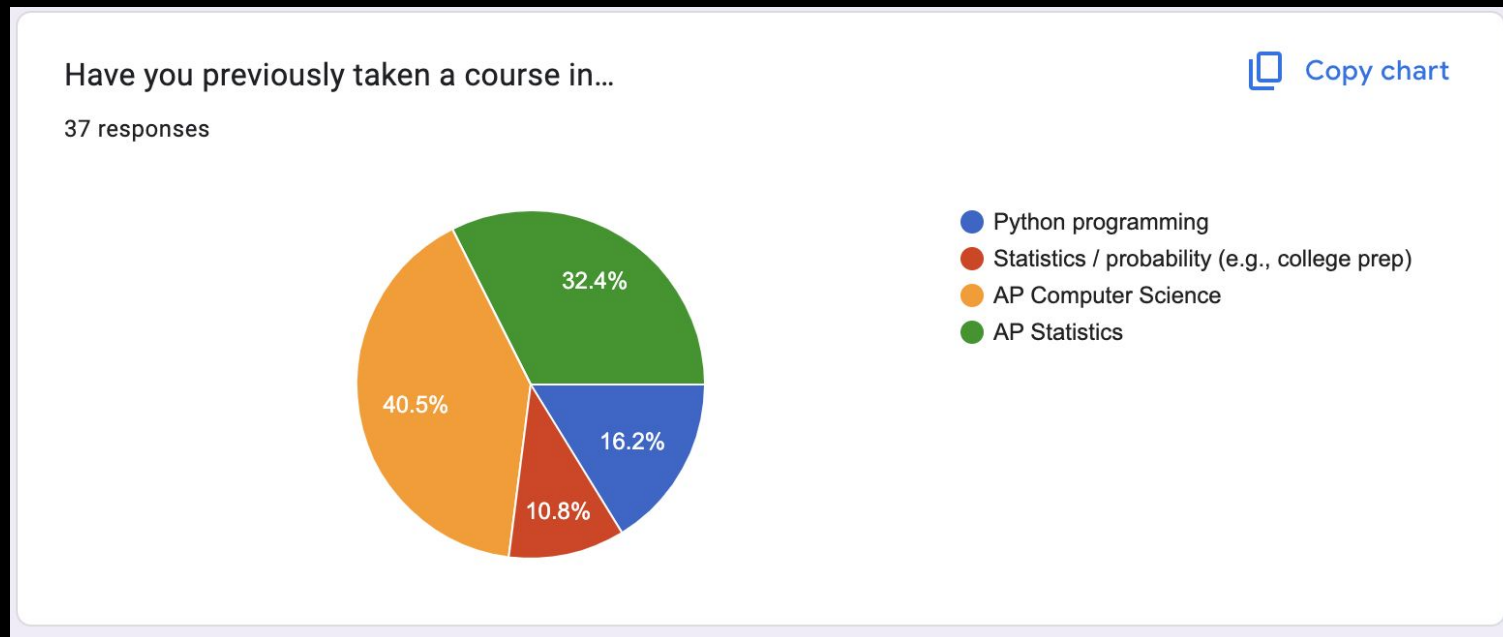
My working definition:

1. Data is any representation of a feature or aspect of a system (e.g., community, human body, physical structure) expressed in abstract terms (e.g., in terms of classes, numbers, vectors etc.) where we have a model of the **generating process**.

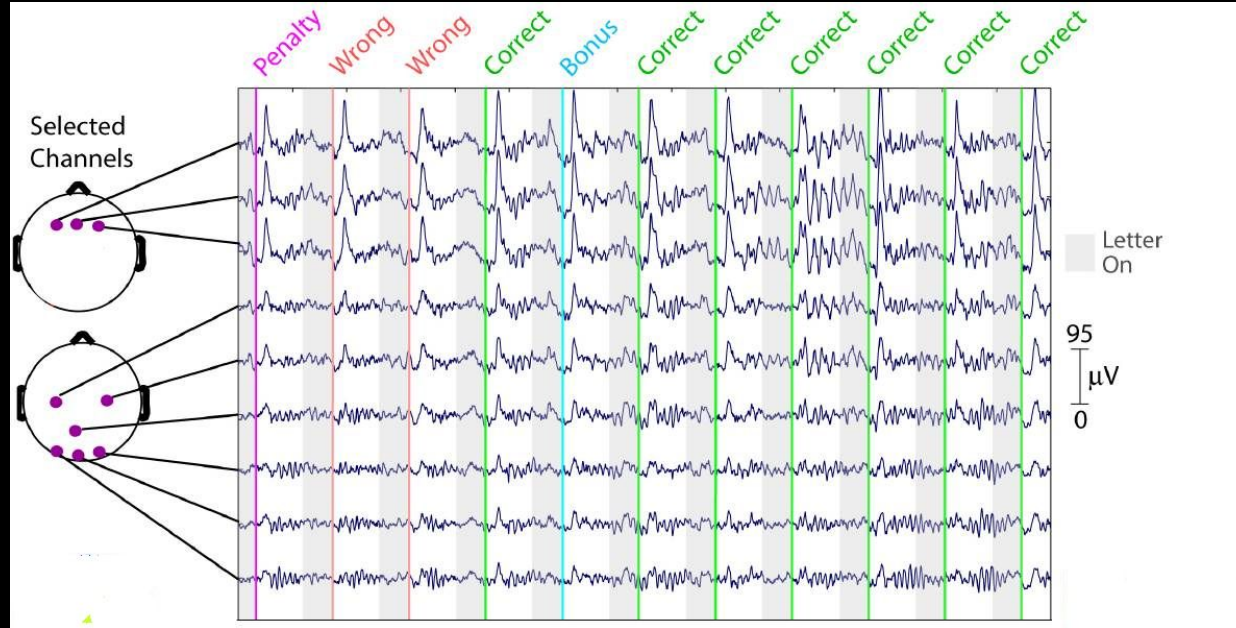
If you don't know any of the following we should be wary of working with our data:

1. The system feature being represented
2. The measurement tool used to generate the data
3. The limitations or omissions of the measurement tool

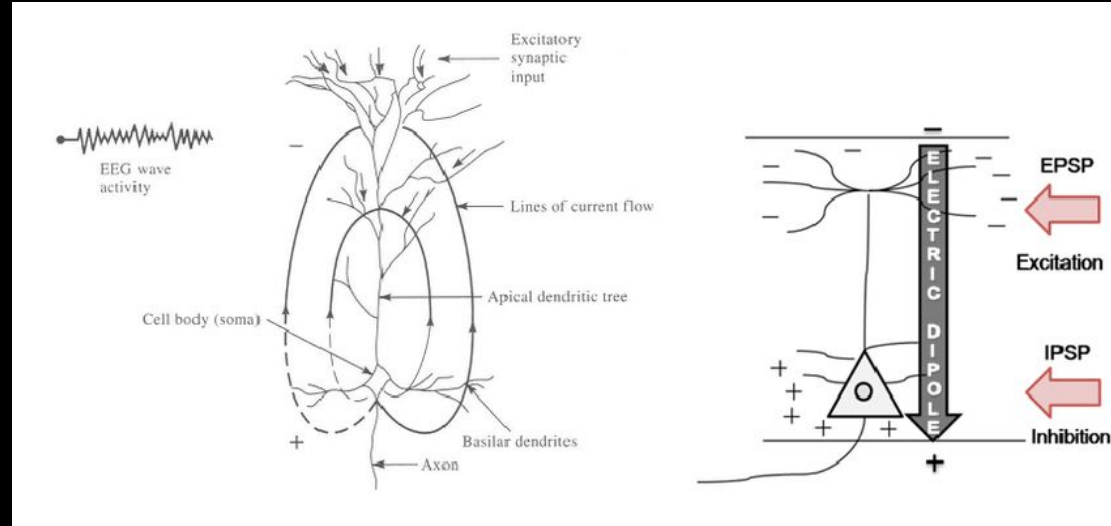
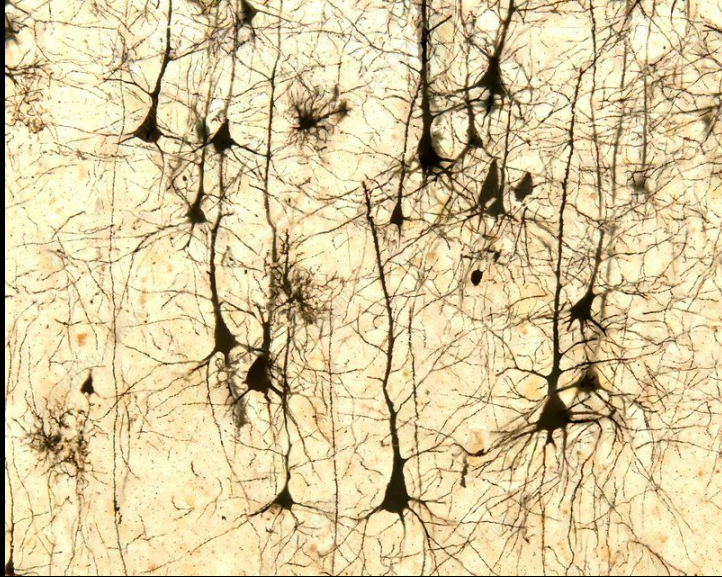
Example: “Have you previously taken a course in...”



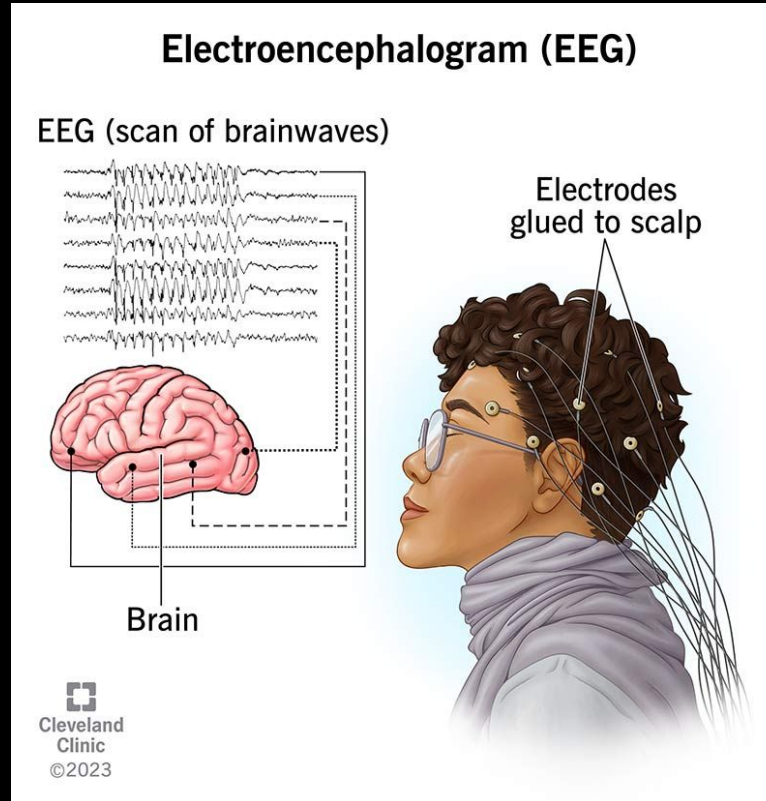
Data Example 1: Brain Data (Signal processing)



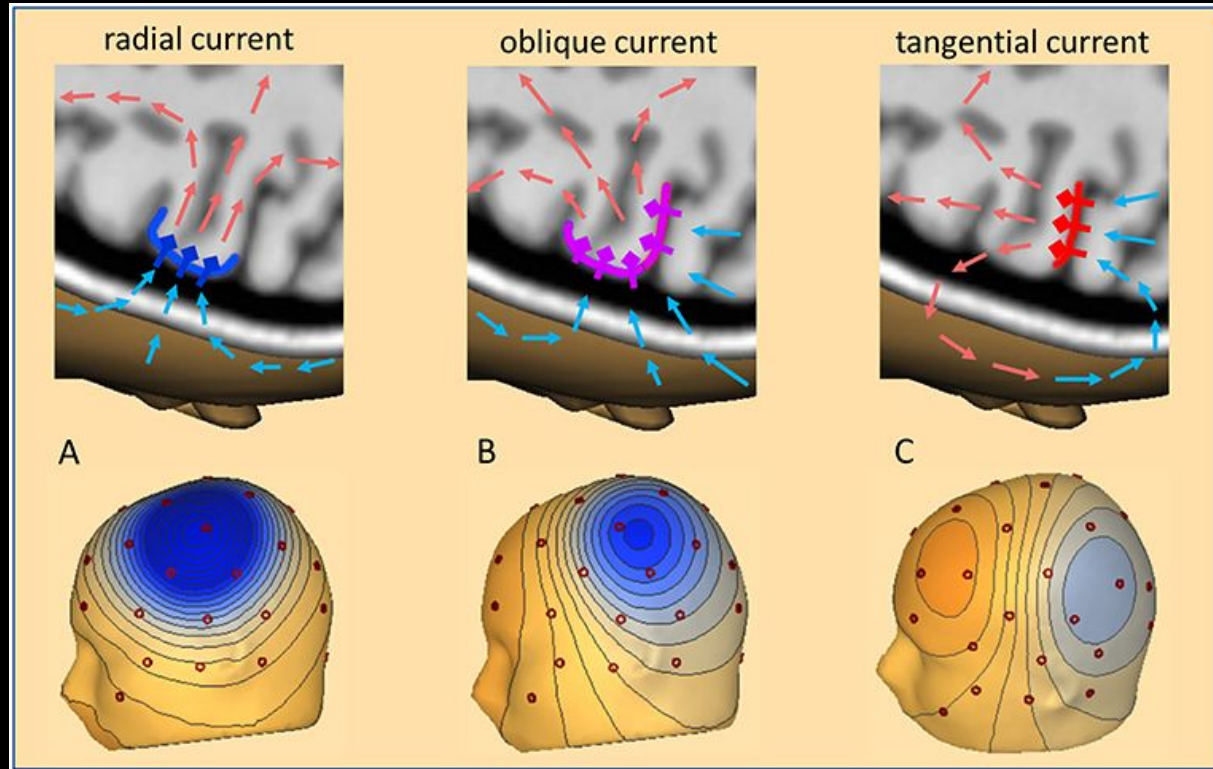
System feature being represented: Cortical neurons in the brain (layer 5)



The measurement tool used to generate the data

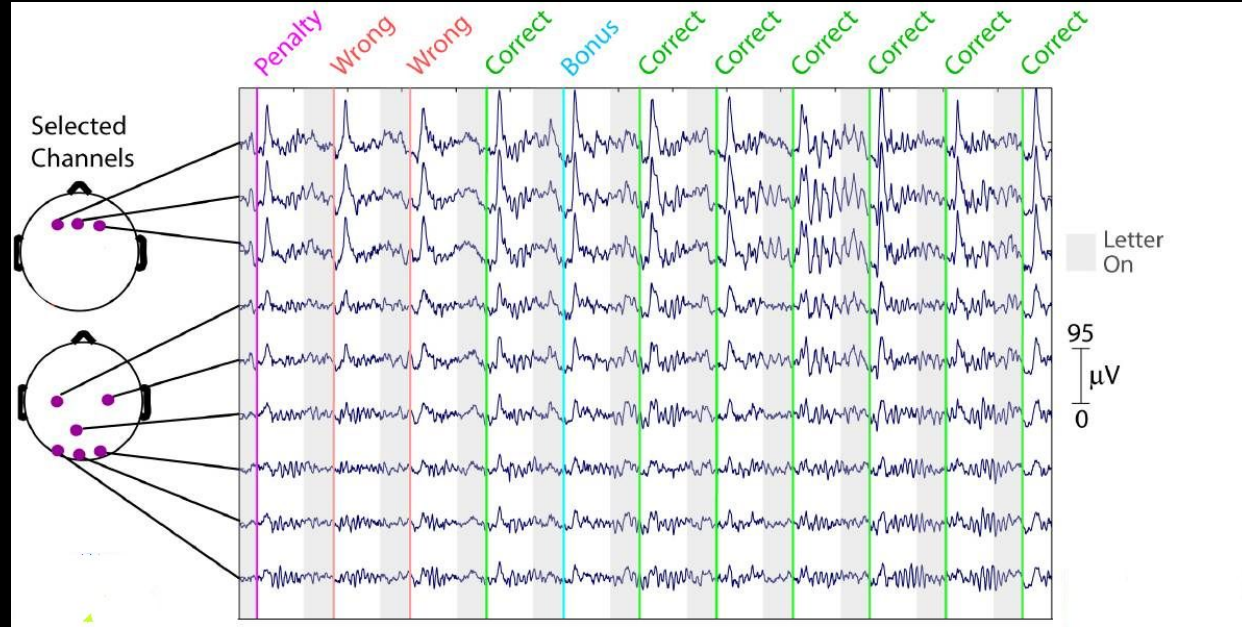


Limitations: Doesn't record the majority of activity



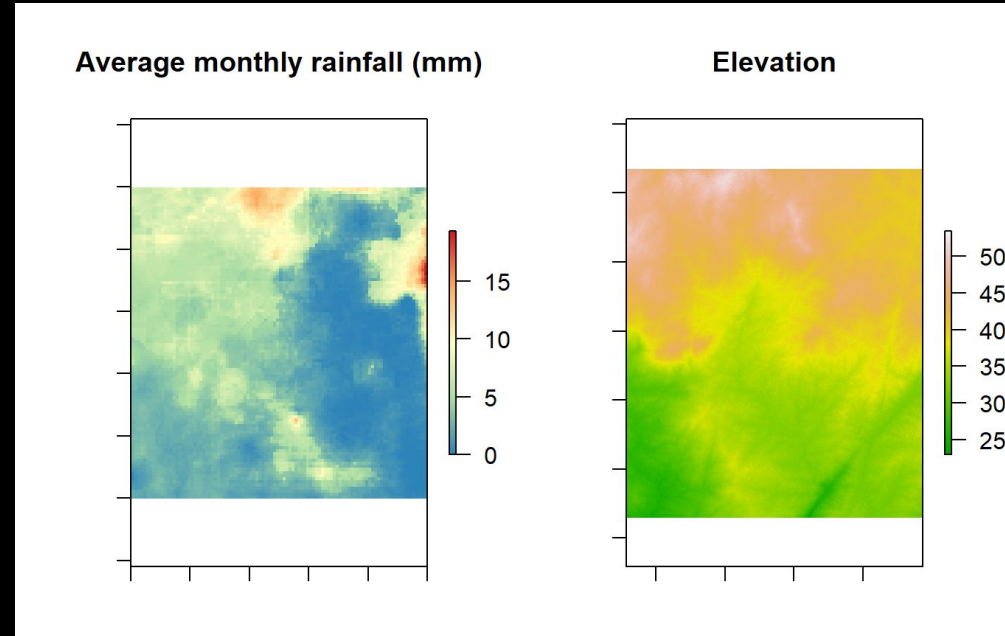
Data Example 1: Brain Data (Signal processing)

1. Cortical neurons in the brain (layer 5)
2. Extracellular electrical recordings (EEG)
3. Ignores all non-electrical signals and electrical signals that don't contribute to “dipoles”



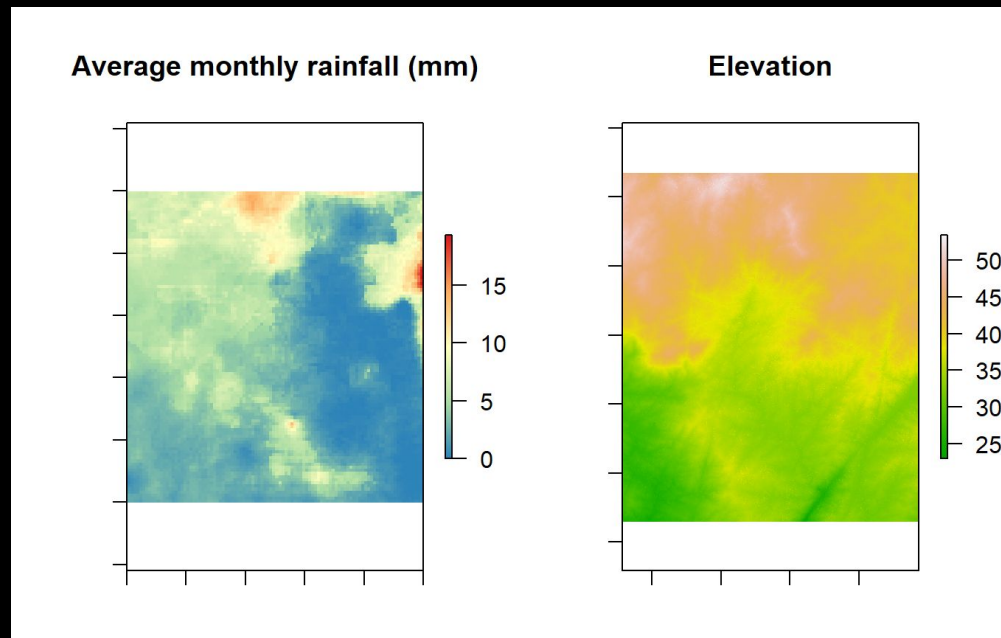
Data Example 2: Environmental Science (Point Processes)

1. The process by which precipitation is deposited in a geographic region, influenced by elevation, topography, and atmospheric circulation.



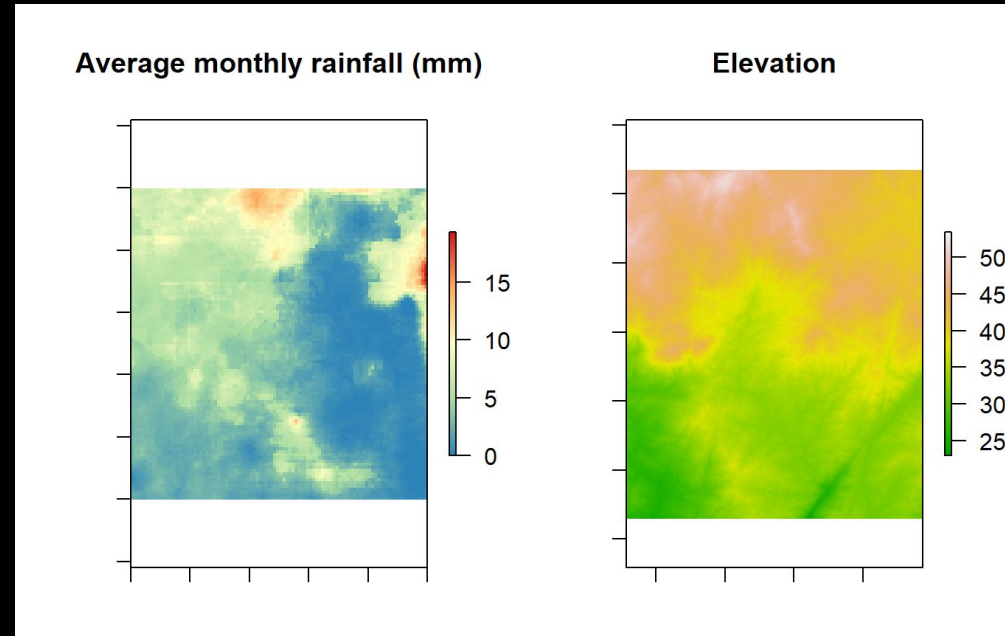
Data Example 2: Environmental Science (Point Processes)

2. Rain gauges at discrete elevation levels that record cumulative precipitation each month.



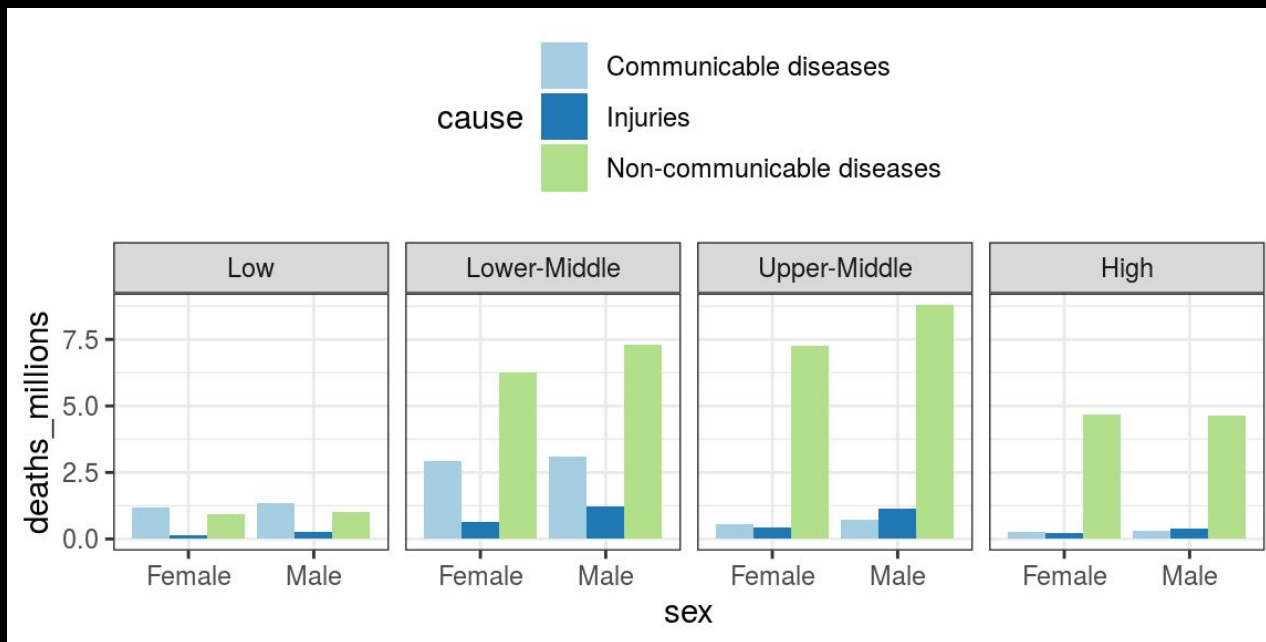
Data Example 2: Environmental Science (Point Processes)

3. Temporal omissions:
missing data during
equipment downtime; no
measurement of
sub-monthly rainfall
intensity.



Many more...

If you can't identify these for any particular data set, it is an area you need to learn more about or you should find a different data set!

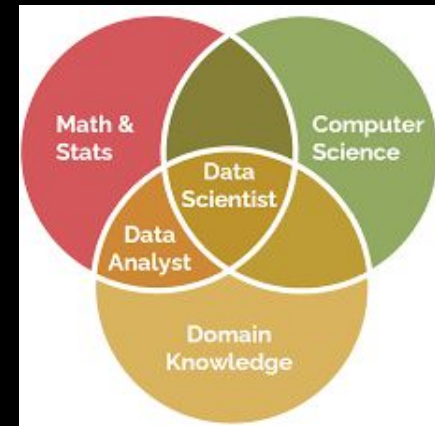


Group Discussion: What is Data Science?

What is Data Science?

Data science is the interdisciplinary field that uses scientific methods, algorithms, and systems to extract knowledge and insights from structured and unstructured data.

Data science transforms raw data into actionable insights—fueling decision-making in business, science, health, and beyond.



Final Projects

You can pick one:

1. Time Series Analysis and/or Prediction – e.g., apply statistical signal processing methods and forecasting techniques to temporal data.

You can pick one:

1. Time Series Analysis and/or Prediction

2. Categorical and Count Data Analysis – e.g., analyze experimental or observational data, comparing conditions and drawing inferences using basic hypothesis testing regimes or regression-based analysis (GLMs) in both or either frequentist and/or Bayesian frameworks.

You can pick one:

1. Time Series Analysis and/or Prediction

2. Categorical and Count Data Analysis

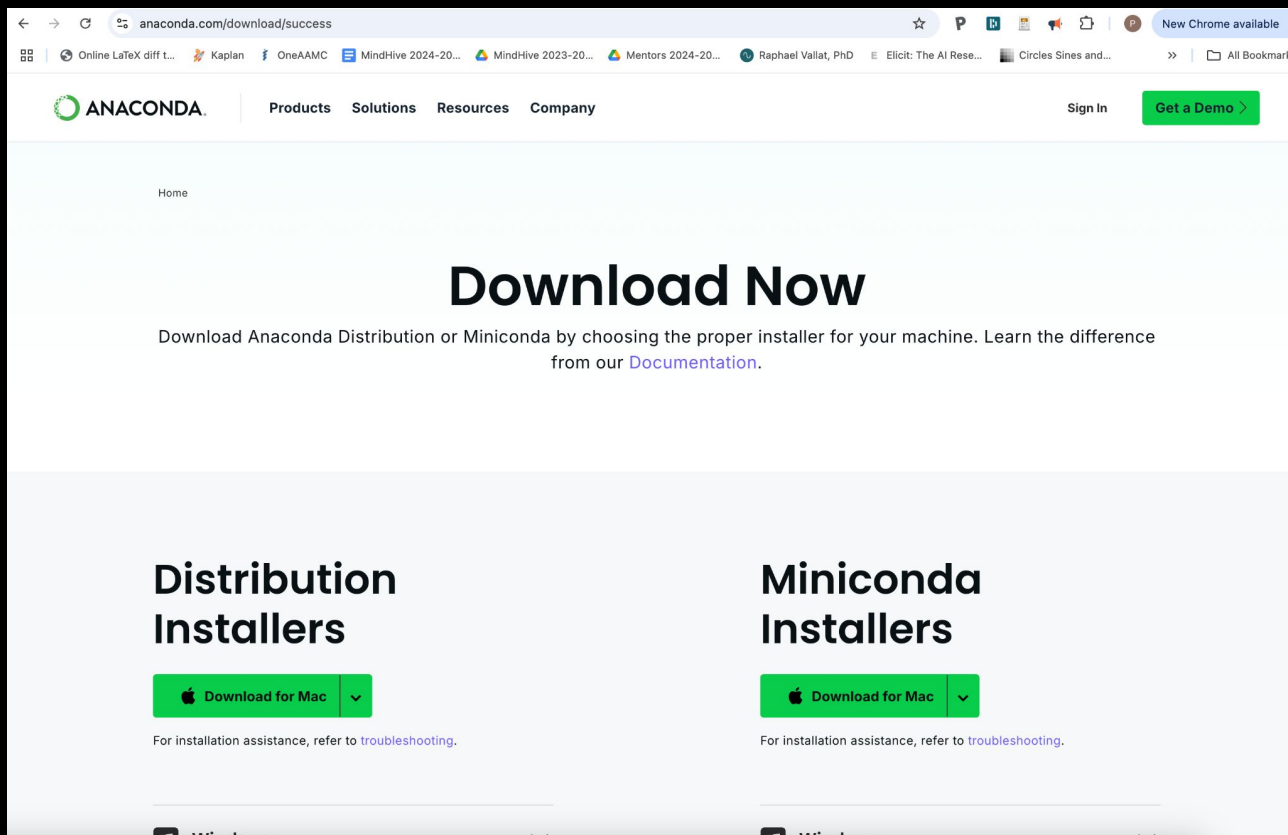
3. Scientific Software Package Development – e.g., implement a statistical method or algorithm in Python as a reusable package, with well-documented code, unit tests, and user tutorials.

You can pick one:

- 1. Time Series Analysis and/or Prediction**
- 2. Categorical and Count Data Analysis**
- 3. Scientific Software Package Development**
- 4. Teaching a Statistical Technique** – e.g., create an interactive lesson in a Jupyter Notebook to explain a concept including narrative, examples, and exercises.

Classroom Tools

Python and Anaconda



The screenshot shows the Anaconda website's download page. The browser's address bar displays 'anaconda.com/download/success'. The website's navigation bar includes the Anaconda logo, links for 'Products', 'Solutions', 'Resources', and 'Company', a 'Sign In' link, and a green 'Get a Demo >' button. The main content area features a large 'Download Now' heading, followed by a paragraph: 'Download Anaconda Distribution or Miniconda by choosing the proper installer for your machine. Learn the difference from our [Documentation](#).' Below this, there are two columns. The left column is titled 'Distribution Installers' and contains a green button with an Apple logo and the text 'Download for Mac' and a dropdown arrow. The right column is titled 'Miniconda Installers' and contains a similar green button. Both columns have a link to 'troubleshooting' for installation assistance. At the bottom of each column, there is a partially visible 'Windows' button.

anaconda.com/download/success

ANACONDA

Products Solutions Resources Company

Sign In Get a Demo >

Home

Download Now

Download Anaconda Distribution or Miniconda by choosing the proper installer for your machine. Learn the difference from our [Documentation](#).

Distribution Installers

Download for Mac

For installation assistance, refer to [troubleshooting](#).

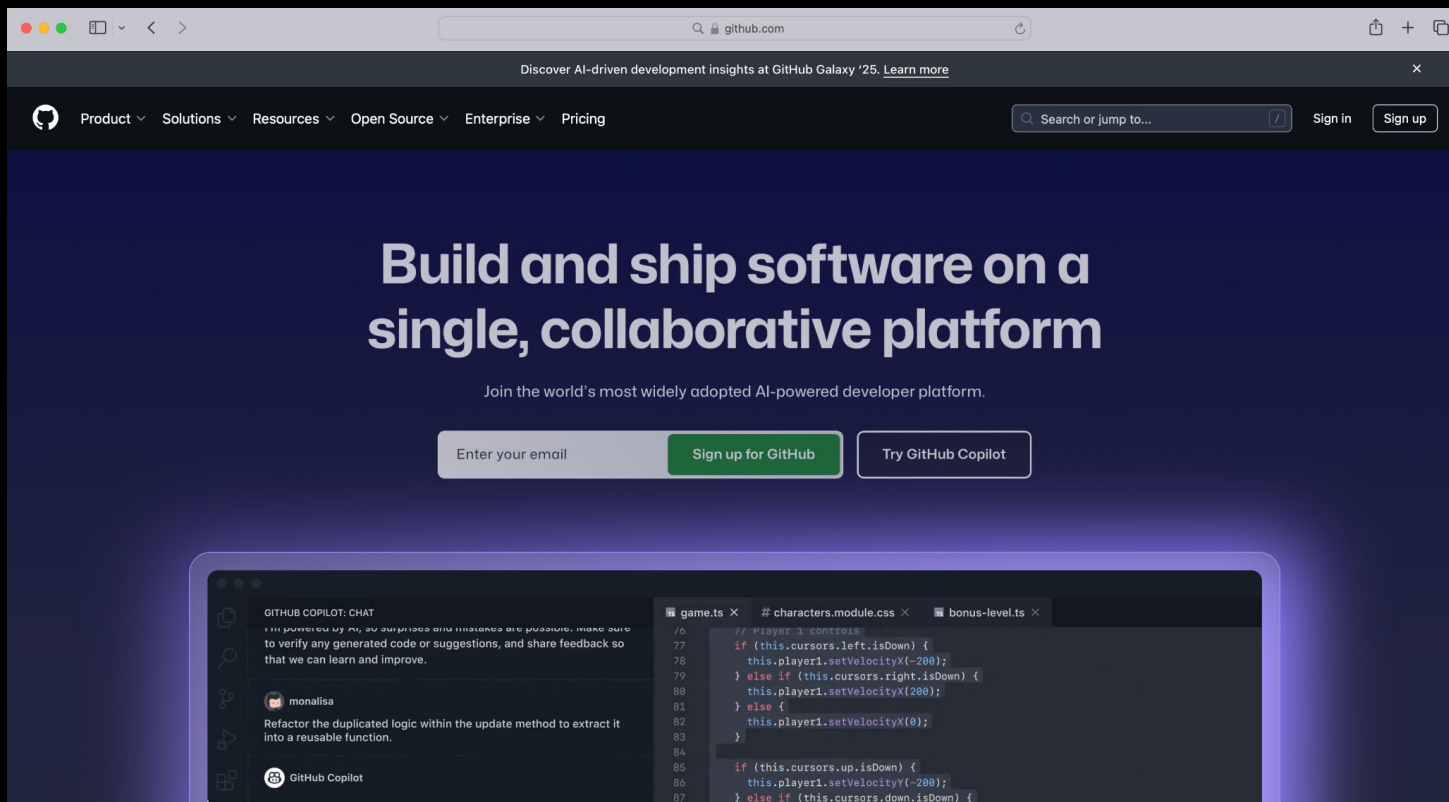
Miniconda Installers

Download for Mac

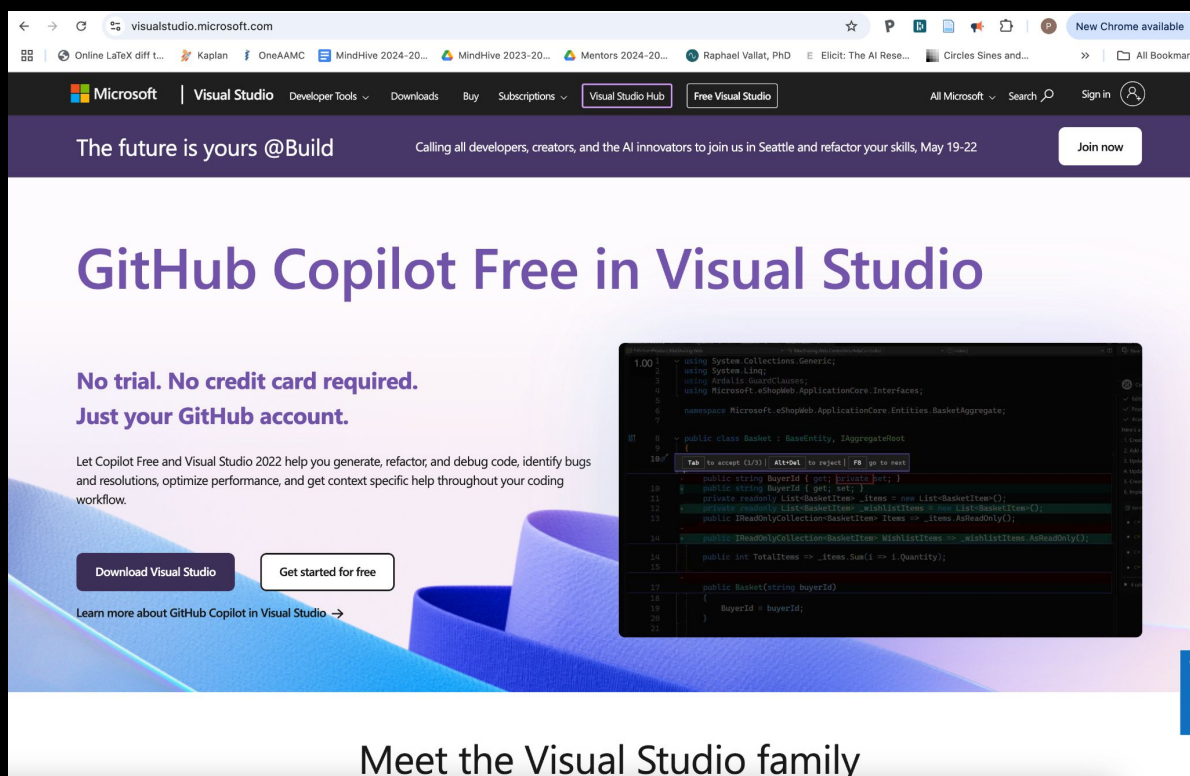
For installation assistance, refer to [troubleshooting](#).

Windows

GitHub (and GitHub classroom)



VSCode (Visual Studio)



The screenshot shows the Visual Studio website homepage. At the top, there's a navigation bar with the Microsoft logo, 'Visual Studio' link, and various developer tools links. Below this is a purple banner with the text 'The future is yours @Build' and a 'Join now' button. The main heading is 'GitHub Copilot Free in Visual Studio'. Below this, it states 'No trial. No credit card required. Just your GitHub account.' and provides a brief description of Copilot's capabilities. There are two buttons: 'Download Visual Studio' and 'Get started for free'. A link 'Learn more about GitHub Copilot in Visual Studio' is also present. On the right side, there's a code editor window showing C# code for a 'Basket' class, with a tooltip suggesting 'to access (x/y)' and 'Alt+Del to reject!'.

visualstudio.microsoft.com

Microsoft | Visual Studio Developer Tools Downloads Buy Subscriptions Visual Studio Hub Free Visual Studio All Microsoft Search Sign in

The future is yours @Build Calling all developers, creators, and the AI innovators to join us in Seattle and refactor your skills, May 19-22 Join now

GitHub Copilot Free in Visual Studio

No trial. No credit card required.
Just your GitHub account.

Let Copilot Free and Visual Studio 2022 help you generate, refactor, and debug code, identify bugs and resolutions, optimize performance, and get context specific help throughout your coding workflow.

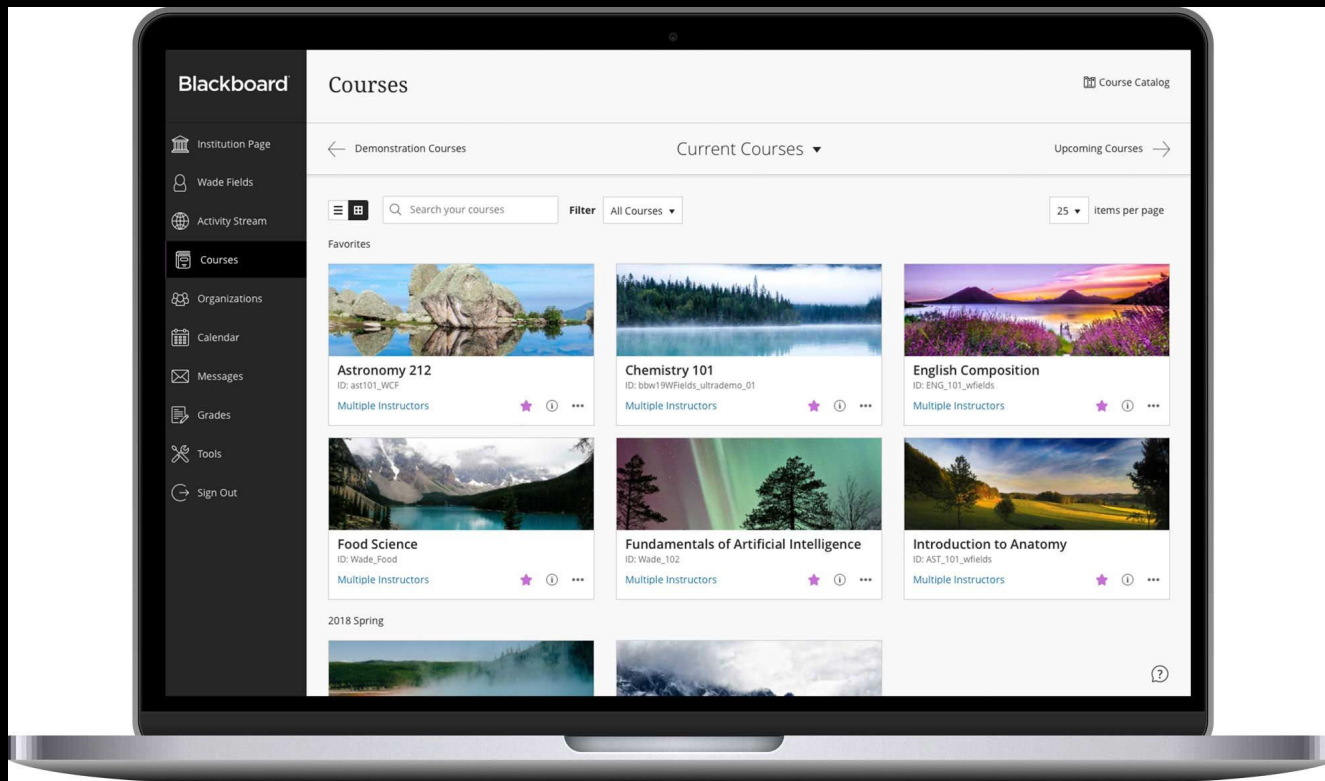
[Download Visual Studio](#) [Get started for free](#)

[Learn more about GitHub Copilot in Visual Studio](#) →

```
1.00: using System.Collections.Generic;
2: using System.Linq;
3: using System.Threading.Tasks;
4: using Microsoft.AspNetCore.Mvc;
5: using Microsoft.AspNetCore.Mvc.Rendering;
6: namespace Microsoft.eShopWeb.ApplicationCore.Entities.BasketAggregate;
7:
8: public class Basket : BaseEntity, IAggregateRoot
9: {
10:     public string BuyerId { get; set; }
11:     public string BuyerId { get; set; }
12:     private readonly List<BasketItem> _items = new List<BasketItem>();
13:     public readonly List<BasketItem> _wishlistItems = new List<BasketItem>();
14:     public IReadOnlyCollection<BasketItem> Items => _items.AsReadOnly();
15:     public IReadOnlyCollection<BasketItem> WishlistItems => _wishlistItems.AsReadOnly();
16:
17:     public int TotalItems => _items.Sum(i => i.Quantity);
18:
19:     public Basket(string BuyerId)
20:     {
21:         BuyerId = BuyerId;
22:     }
23: }
```

Meet the Visual Studio family

BlackBoard



GitHub Classroom (auto-grade your assignments)

The screenshot shows the GitHub Classroom web interface. At the top, the browser address bar displays the URL `classroom.github.com/classrooms/211573695-bu-rise-data-science-2025`. The page header includes the GitHub Classroom logo and the text "GitHub Education". Below the header, the breadcrumb "Classrooms / BU-RISE-Data-Science-2025" is visible. The main section is titled "BU-RISE-Data-Science-2025" with the subtitle "BU-RISE-Data-Science". A navigation bar contains links for "Assignments 1", "Students 1", "TAs and Admins 1", and "Settings". The "Assignments" link is highlighted with a red underline. On the right side of the main section, there is a green button labeled "+ New assignment". Below this, a light blue notification banner states "Get verified to get Codespaces Education benefit." with a "Get verified" button and a close icon.

classroom.github.com/classrooms/211573695-bu-rise-data-science-2025

Classroom

GitHub Education

Classrooms / BU-RISE-Data-Science-2025

BU-RISE-Data-Science-2025

BU-RISE-Data-Science

☆ Assignments 1 👤 Students 1 👤 TAs and Admins 1 ⚙️ Settings

Assignments

+ New assignment

🔔 Get verified to get Codespaces Education benefit. Get verified ×

Zoom (via your BU account, install before tomorrow)

The screenshot shows the Zoom website homepage. At the top, there's a navigation bar with the Zoom logo, links for Products, AI, Solutions, Resources, and Pricing, and buttons for Support, Meet, Sign in, Contact Sales, and Sign Up Free. A banner below the navigation bar says "Don't miss the next Work Transformation Summit | Register here". The main content area features the headline "Find what you need when you need it with AI Companion". Below this, it says "Accomplish more with Zoom Workplace: your AI-first work platform featuring AI Companion 2.0, included at no extra cost.*". There are two buttons: "Plans & pricing" and "Discover Zoom Workplace". To the right, there's a section titled "How can I help you?" with three options: "Prepare notes", "Catch me up", and "Action items". Below this, there's a video thumbnail showing two people working on a laptop. To the right of the video, there's another section titled "Ask AI Companion" with the question "What has the team decided?".

zoom Products AI Solutions Resources Pricing Support Meet Sign in Contact Sales Sign Up Free

Don't miss the next Work Transformation Summit | Register here

Find what you need when you need it with AI Companion

Accomplish more with Zoom Workplace: your AI-first work platform featuring AI Companion 2.0, included at no extra cost.*

Plans & pricing Discover Zoom Workplace

How can I help you?

- Prepare notes
- Catch me up
- Action items

Ask AI Companion

What has the team decided?

Start Introduction to Python Coding Activity 1 (Blackboard and Google Colab)