

Developing Thoughtful Scientists through an Introduction to Data Science Course

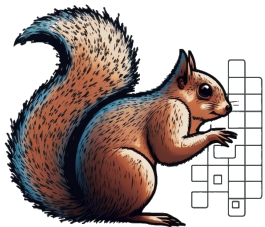
Course: Computational Mathematics, Science, and Engineering (CMSE) 201

Emily Bolger, PhD Candidate

Department of Computational Mathematics, Science, and Engineering (CMSE)

Michigan State University

eCOTS Posters and Beyond 2024



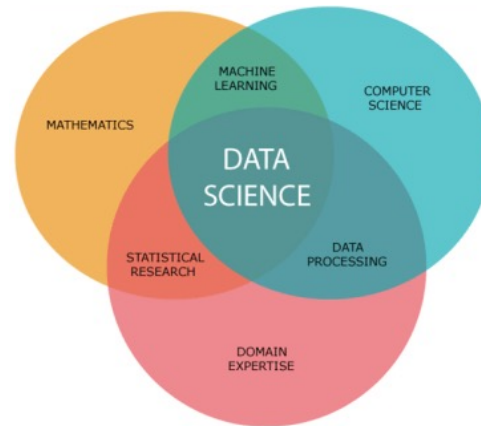
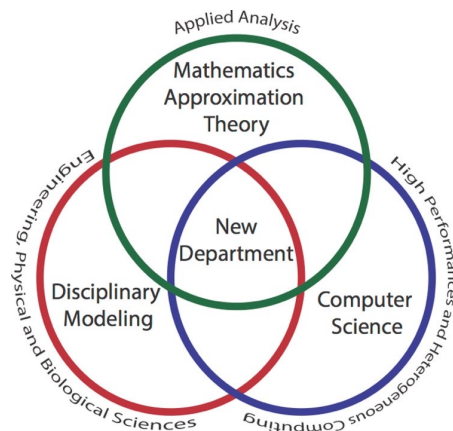
CERL@MSU

COMPUTING EDUCATION RESEARCH LAB



Department of Computational Mathematics, Science, and Engineering

- We are NOT math/cs/stats + data
- Focus on the context and developing data science skills
 - But, intentional about not making the context the challenging part
- Courses are frequently updated
 - Demographic of students and development of DS/ML methods

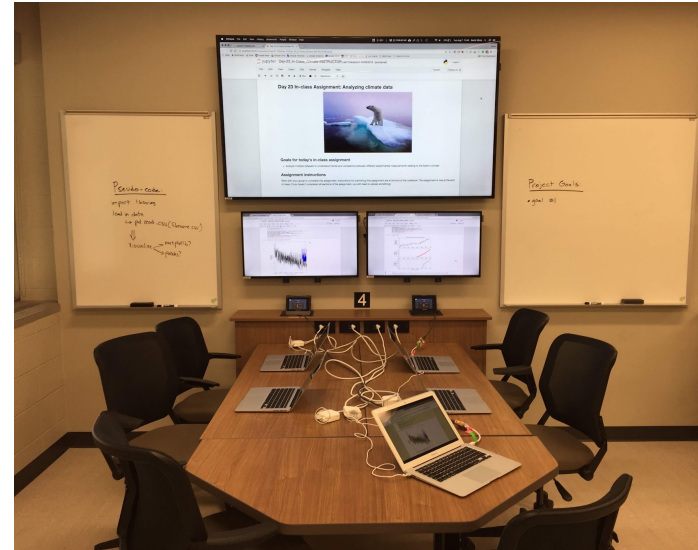


Students have many opportunities to connect with the content and develop a data science identity



CMSE 201

- First course in our undergraduate major
- Instructional Staff:
 - 1 Lead instructor (Faculty)
 - 1 Teaching Assistant (Graduate Student)
 - 1-3 Learning Assistants (Undergraduate Students)
- Flipped Classroom
 - Pre-class, In-Class



Example Classroom Setup

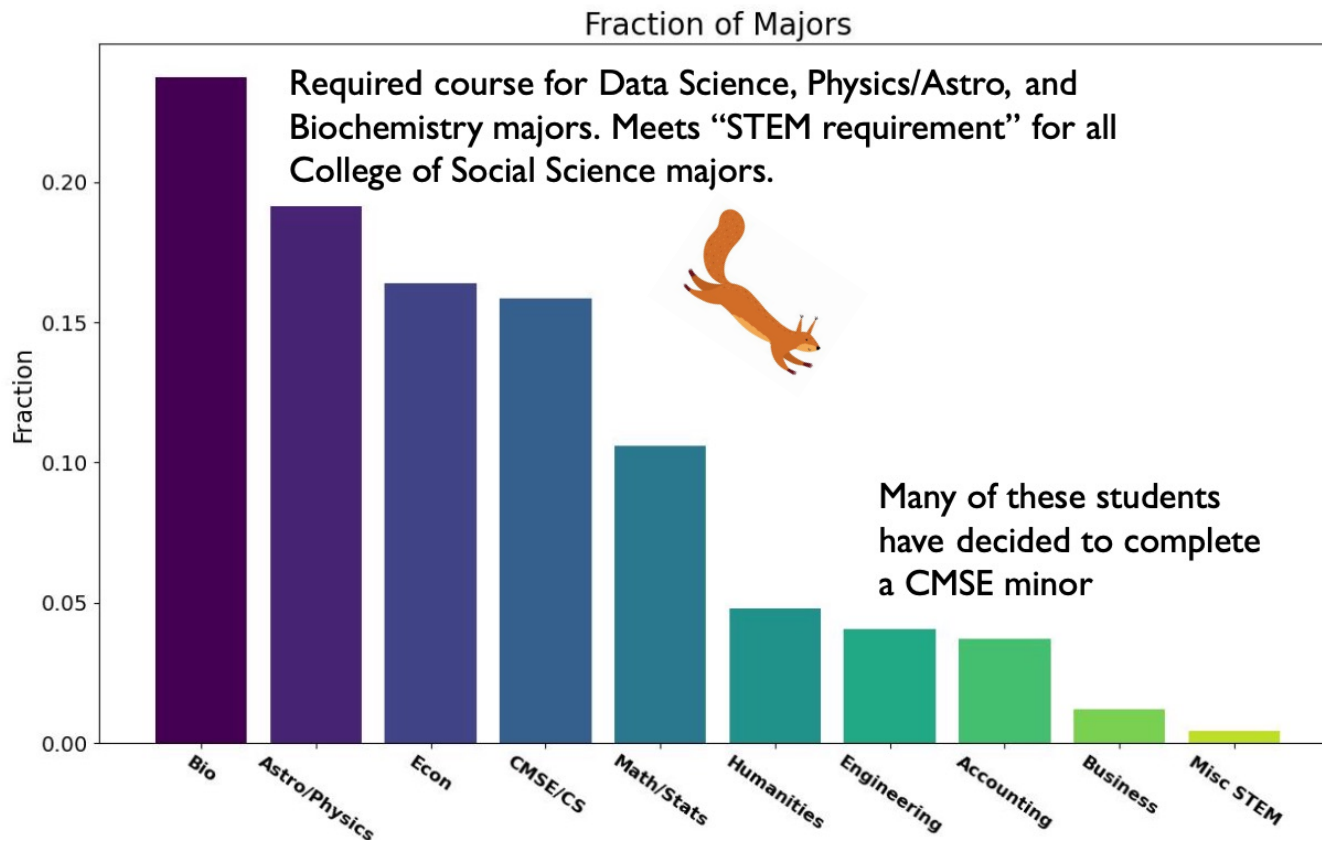
Learning Goals Include:

- Manipulate, Visualize, and Analyze Data
- Write programs to solve common problems in a variety of disciplines
- Present results from scientific computing problem verbally and in writing



CMSE 201 Student Population

Primary Goal: Provide students with computational knowledge to apply skills to their own discipline



Highlighting Cool Assignments from CMSE 201



What Students Have Learned So Far

Day 1:
Algorithmic
Thinking

Day 2:
Order of
Magnitude
Modeling

Day 3:
Lists, Loops, and
Savings

Day 4: Data and Algorithmic Bias

- Introduce bias in a relatable way
 - e.g. supporting sports team, buying product
- Connect to more complex bias
 - Looks like in data and algorithms
- Discussion with classmates
 - Gives them the language to talk about their experience
 - Lots of info to instructional team to minimize harm

2. What is data bias?

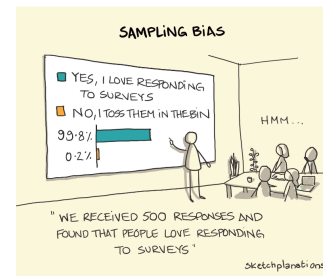


Image from: <https://sketchplanations.com/sampling-bias>

Bias is defined to be "prejudice in favor of or against one thing, person, or group compared with another, usually in a way considered to be unfair." Data bias occurs when parts of a dataset are overemphasized, underemphasized, or are completely nonexistent.

✓ Task:

Read the article at [this link](#) that highlights different types of data bias. While this is discussing machine learning, it is applicable to anyone collecting, using, and visualizing data. Don't be discouraged if you are unclear about machine learning.

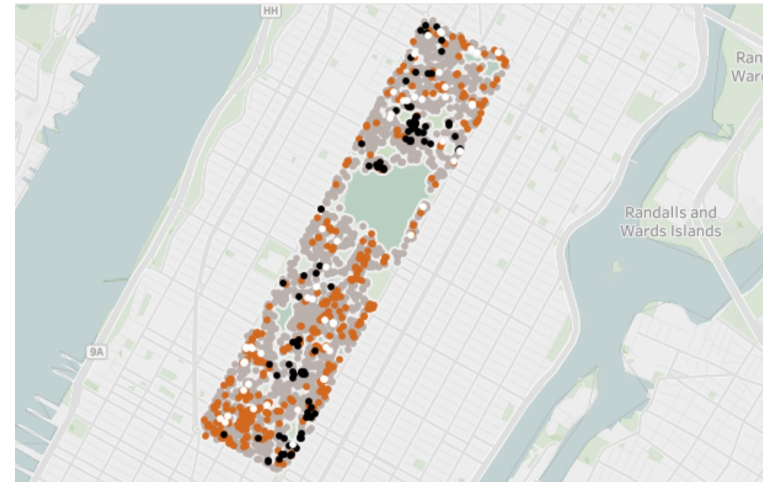
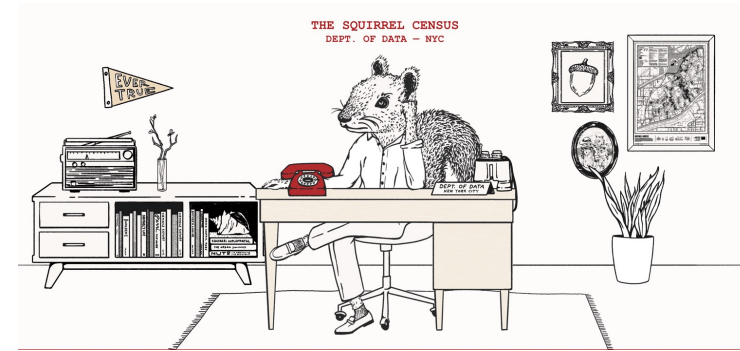
Day 5:
If Statements
& Functions

Nothing is free from bias



Day 6: A Scurry of Squirrels and Functions

- First day with real data
 - [NYC Central Park Squirrel Census](#)
- Real, fun data with context
 - Data is given in lists
 - No reading in/cleaning yet
 - Effort from instructional team



Day 6: A Scurry of Squirrels and Functions

1.4 Writing your Function

In the cells below, you will **write your function with your small group.** The lists you will use for testing your function are in the first cell. All of the lists are given, but you will only need to use the lists relevant to the function you are writing.

```
In [1]: # sample data from 2018 Central Park Squirrel census
squirrel_primary_fur_color = ["Gray", "White", "Black", "Gray"]
squirrel_shift = ["AM", "AM", "PM", "AM"]

In [2]: # put your code here
```

Part 2: Using Functions Inside Other Functions

In this part, you will work with **all** of your group members to use the Function 2 that one of your small groups wrote to write a new function to determine how many squirrels there are with one of the unique fur colors.

2.1 Sharing your work

The first step is to make sure everyone has a copy of both functions. Without making any edits to your original function, copy your group's function and distribute it to the other group (you can use Slack, Teams, Email, etc.). In the cell(s) below, add the function from the other small group.

```
[10]: # Add the additional functions here
```

13:45 - 15:00 **Nifty Assignments at Oregon Ballroom 203** Nifty Assignments
 Chair(s): [Nick Parlante](#) Stanford University, [Julie Zelenski](#) Stanford University, [Dave Reed](#) Creighton University, [Michael Guerzhoy](#) University of Toronto

13:57 12m ★ **Exploring the Scurry of Squirrels in Central Park** CC HYBRID
 Rachel Frisbie Department of Computational Mathematics, Science, and Engineering; Michigan State University, [Devin Silvia](#) Department of Computational Mathematics, Science, and Engineering; Michigan State University, [Marcos Caballero](#) Michigan State University, [Rachel Roca](#) Department of Computational Mathematics, Science, and Engineering; Michigan State University, [Amanda Bowerman](#) Computational Mathematics, Science, and Engineering, [Krithi Sachithanand](#) Computational Mathematics, Science, and Engineering

Day 7-10:
 Matplotlib,
 NumPy,
 Pandas

Day 11: Flint Water Data Analysis

- First full data analysis of [Flint Water Data](#)
 - Includes:
 - Read in Data
 - Data Exploration (quantitative and visuals)
 - Using computation to test data against the EPA limit
 - Repeat for different samples
 - Make Conclusions

6. (Time Permitting) Looking for Other Sources of Analysis

For your project, it will be good to draw upon other resources to compare your results/conclusions to.

Go through the [following article from the New York Times](#) and compare it's conclusions to your own from information in this article, *Would you feel comfortable drinking a glass of water in Flint at the height*

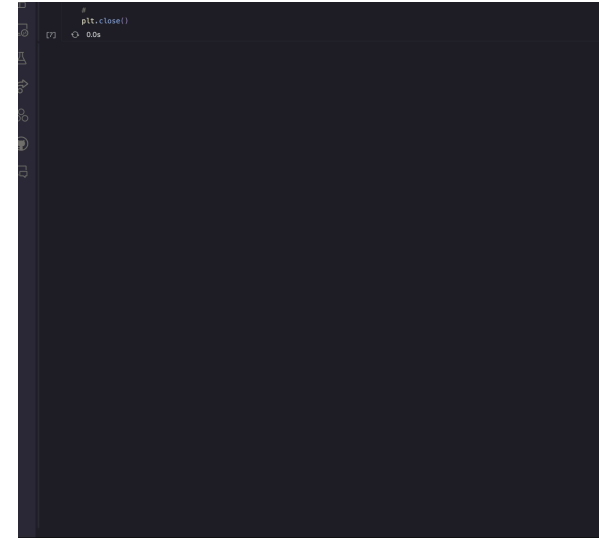
Relevant data; analysis from start to finish

Day 12-21:
File Formats,
Data Viz, Curve
Fitting, ODEs,
2D NumPy
arrays



Day 22/23: Agent Based Models with Forest Fires

- Two day in class assignment
- Heavy scaffolding to focus on code execution and learning goals:
 - 2D indexing skills
 - Manipulating 2D arrays
 - Building and using a simulation
 - Calculating statistics based on the simulation



```
for i in range(board.shape[0]):  
    for j in range(board.shape[1]):
```

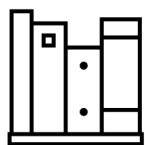
	j=0	j=1	j=2
i=0	(0,0) 2	(0,1) 0	(0,2) 1
i=1	(1,0) 2	(1,1) 1	(1,2) 1
i=2	(2,0) 2	(2,1) 1	(2,2) 0

Day 23-24:
Final Projects

Complex tasks can be accessible to new coders

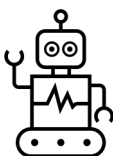


Future Curriculum Development Updates



Data Literacy

- Integrating [assignment developed](#) by myself and Rachel Roca
- Connecting simulation to models used in real hiring processes



Thoughtful AI Use

- Productive uses of ChatGPT
- Scaffolding AI critical thinking
- Teaching basics of what AI is and is not capable of



What to hear more about CERL?



Share your thoughts with me at: bolgerem@msu.edu

