

Software Design for Data Science

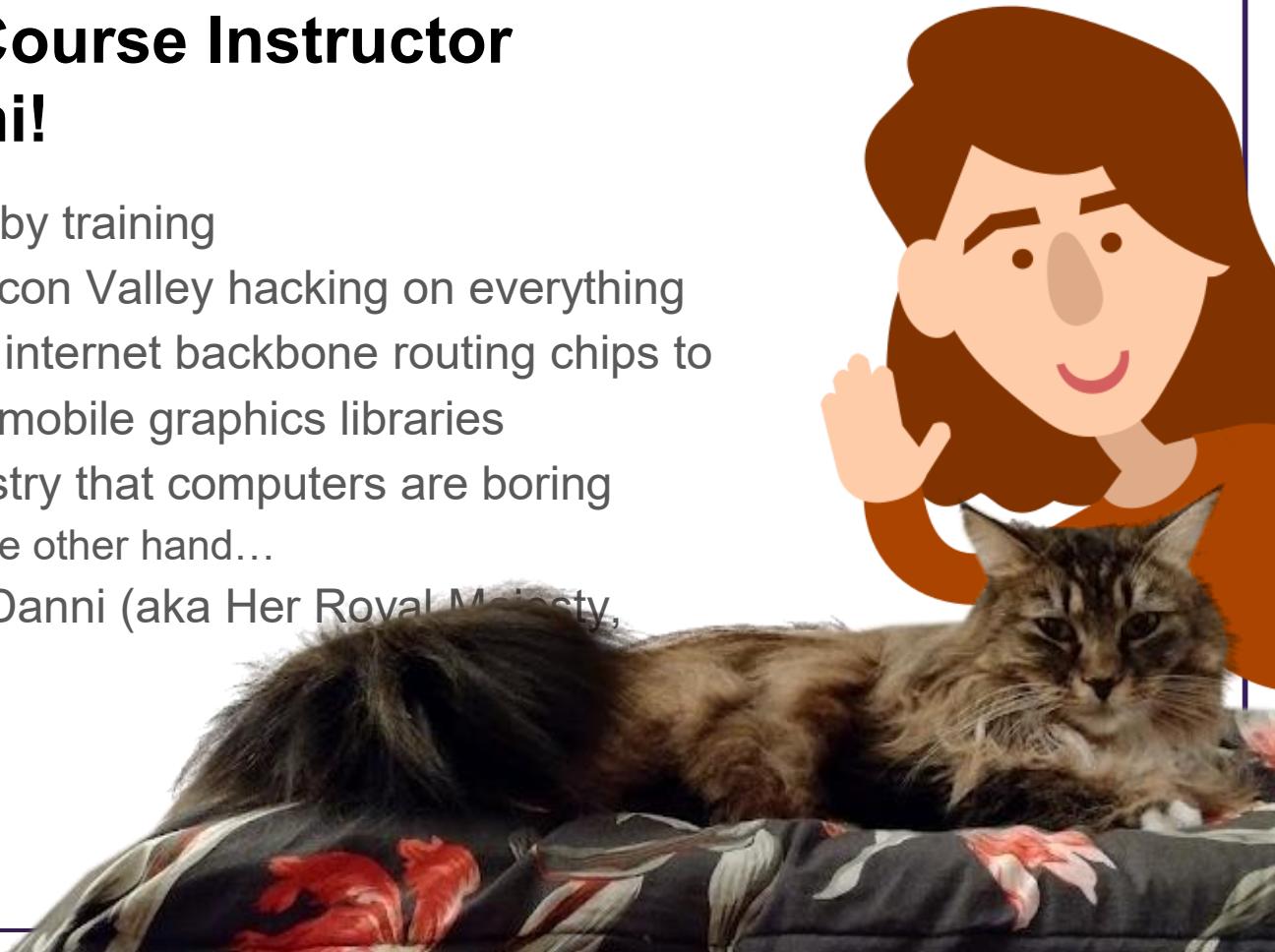
Introduction

*Naomi Altermann
University of Washington
January 6, 2025*



Introductions: Course Instructor Professor Naomi!

- Electrical engineer by training
- Bopped around Silicon Valley hacking on everything from OS kernels to internet backbone routing chips to LIDAR firmware to mobile graphics libraries
- Discovered in industry that computers are boring
 - But *people* on the other hand...
- Proud cat mom to Danni (aka Her Royal Majesty, Queen Baby)

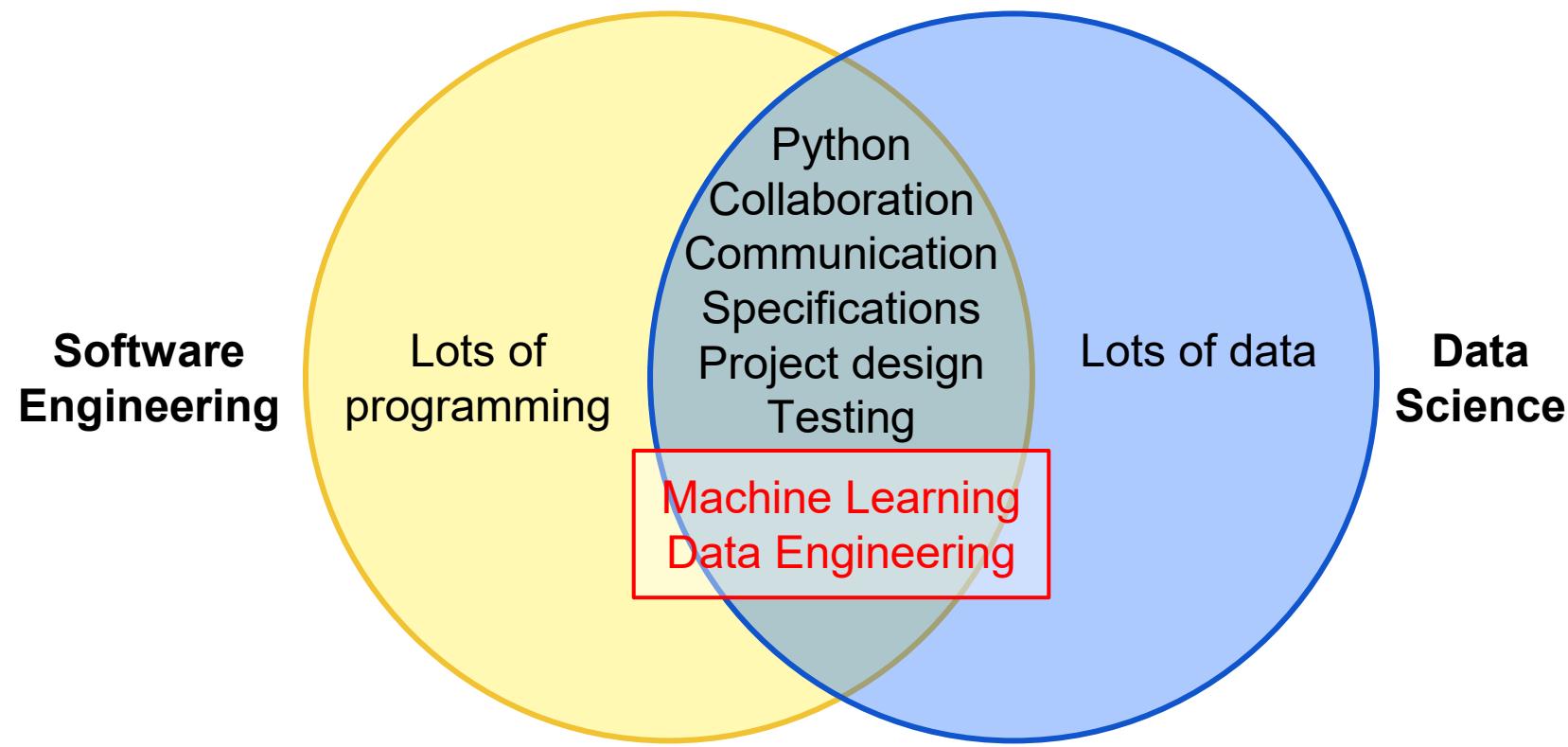


Introductions: Teaching Assistant Riyosha (She/Her)

- 2nd Year MSDS student (2024-26)
- '24 graduate from Ashoka University
 - Majored in Mathematics with minors in CS and Pol Sci
- Most recently researched with the Transport-Human Interaction Network Lab at UW over the summer
- Enjoy hiking, reading and overanalyzing movies with friends (unasked)



Software Engineering vs Data Science



This course

- Practical
- Hands-on
- Useful after graduation
- Focus on HOW to learn,
not just the skills



Course Website

<https://uwdata515.github.io/>

Linked from MyUW

+ Canvas for grades, surveys

+ Ed Discussions

4 hours, with breaks between topics

Course Project

- Collaborative software engineering experience
- Teams of 3 to 4 with 4 being optimal
- Develop project using version control



Course Project

Collaborative software engineering experience

- Design (use cases, component specification)
- Documentation (how to, docstrings)
- Style (PEP8, pylint)
- Coding, testing & milestones
- Standup & code reviews

Course Project Types

1. Answer “Research” Questions
2. Create a Tool
3. Teach Analysis
4. Present Data

Getting Started

Step 1

Students present statements of interest

Step 2

Gather with like-minded students

Step 3

Verify the project idea

Step 4

Size the effort

More on the Data

- At least two non-trivial data sets
- Data need to be combined, joined, merged, etc. to answer the scientific questions
- Have access to the data NOW!

Some Public Data

<http://drugbank.ca>

<https://www.nlm.nih.gov/toxnet/>

<https://data.seattle.gov/browse?tags=traffic&sortBy=relevance&pageSize=20>

<https://divvybikes.com/system-data>

<https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

<https://www.kaggle.com/datasets>

<https://www.seattle.gov/transportation/projects-and-programs/programs/new-mobility-program/scooter-bike-share-data> American Fact Finder Data

[European union data \(World bank\)](#)

[Russian federation data \(World bank\)](#)

[China data \(World bank\)](#)

Project Ideation

Over the first few weeks:

- What areas are you interested in? E.g. social good or a job demo.
- What data are available in that space?
- What tools already exist in that space?
- What type of project is this? (answer research question, create reusable data, create a tool, other?)
- **Give a one slide, 1 minute project idea pitch at the start of class #5!**

Academic Integrity

- **Software development is a highly collaborative endeavor**
- We expect you to collaborate, but your work is your own
- In software, there is rarely one correct solution to any problem
 - Standing around a white board brainstorming is OK
 - Directly copying code someone else in your class is not OK
- The point is for you to learn the concepts and copying answers can defeat that point

But what about AI?

- Lmao what ***about*** AI?? 😂 😂
- Working on getting content together for later in the quarter
- In the meantime though: treat AI as a classmate
 - You might ask a classmate to help explain something to you
 - You might even collaborate with a classmate when OK'ed by the instructor
 - You wouldn't copy a classmate's answers on a solo homework assignment
 - You wouldn't ask a classmate if your answer is right or not

