Your teachers

Me: software engineer
TA: data scientist

You: data scientists

Software Engineering vs Data Science

Software Engineering   Lots of programming

Python
Collaboration
Communication
Specifications
Project design
Testing

Machine Learning
Data Engineering

Lots of data   **Data Science**
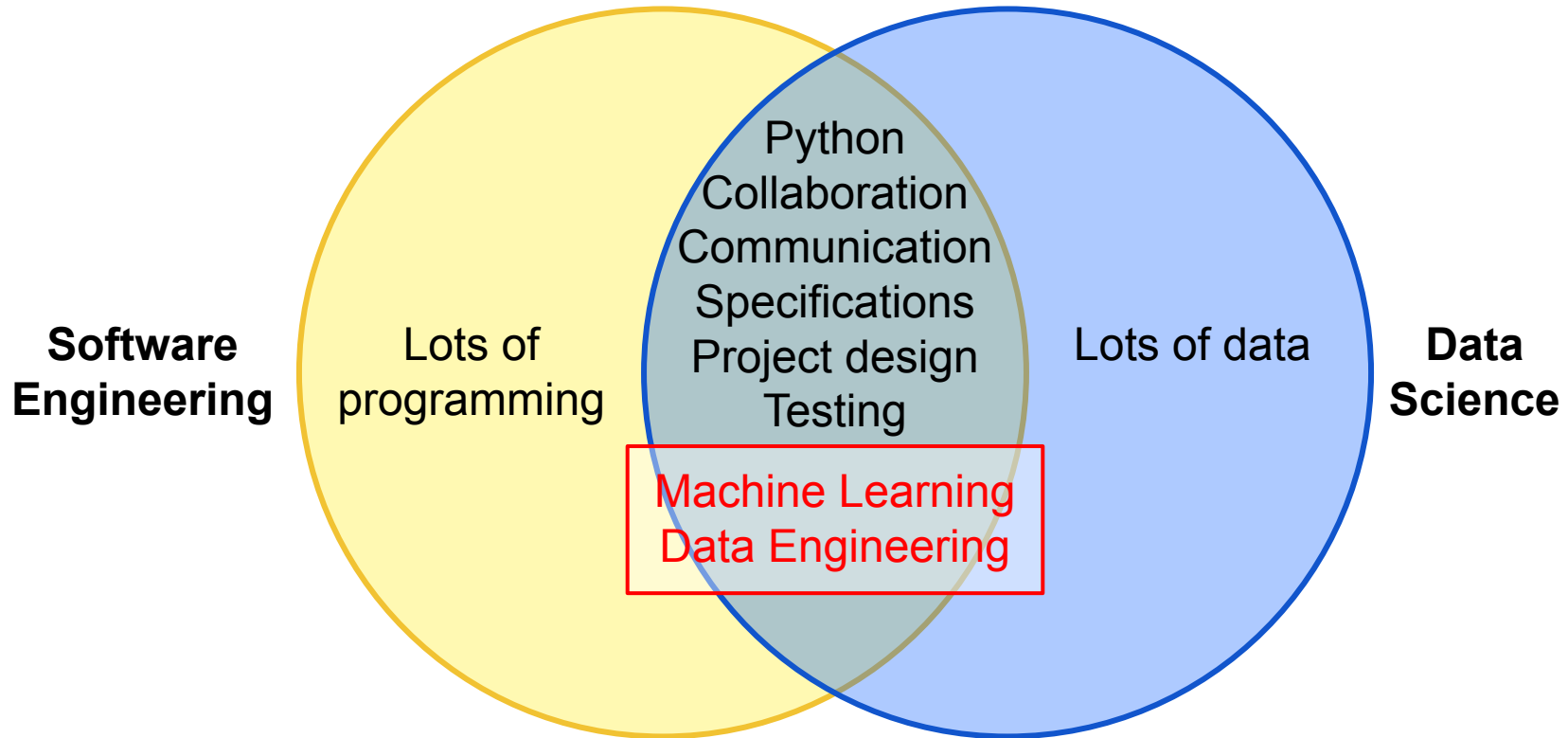
This course

- 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	Step 2	Step 3	Step 4
Students present statements of interest	Gather with like-minded students	Verify the project idea	Size the effort

UNIVERSITY *of* WASHINGTON

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

http://toxnet.nlm.nih.gov

https://data.seattle.gov/Transportation/Traffic-Flow-Counts/7svg-ds5z

https://www.divvybikes.com/data

http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

https://www.kaggle.com

Pronto bike 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