Introduction to Scientific Computing I
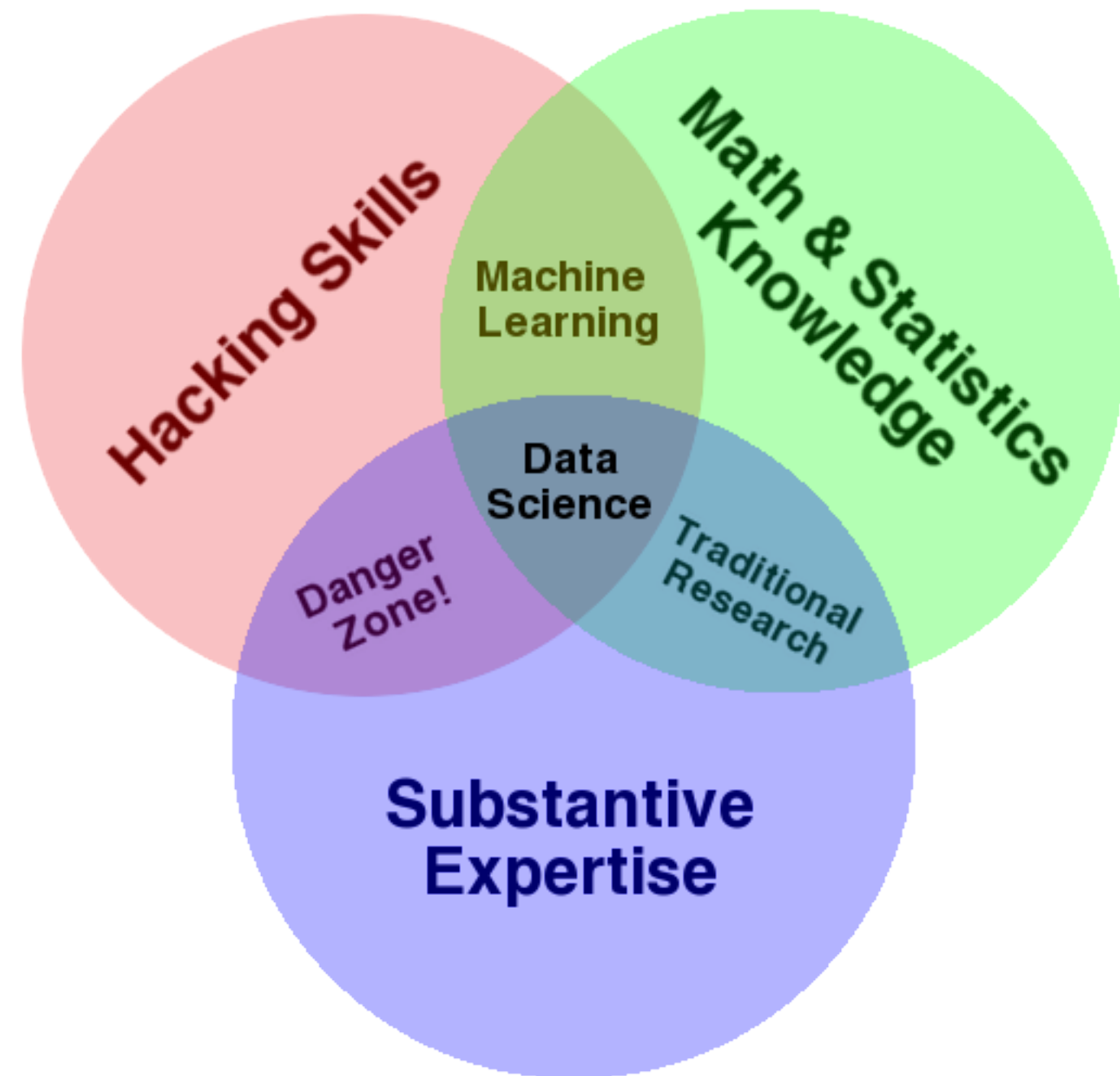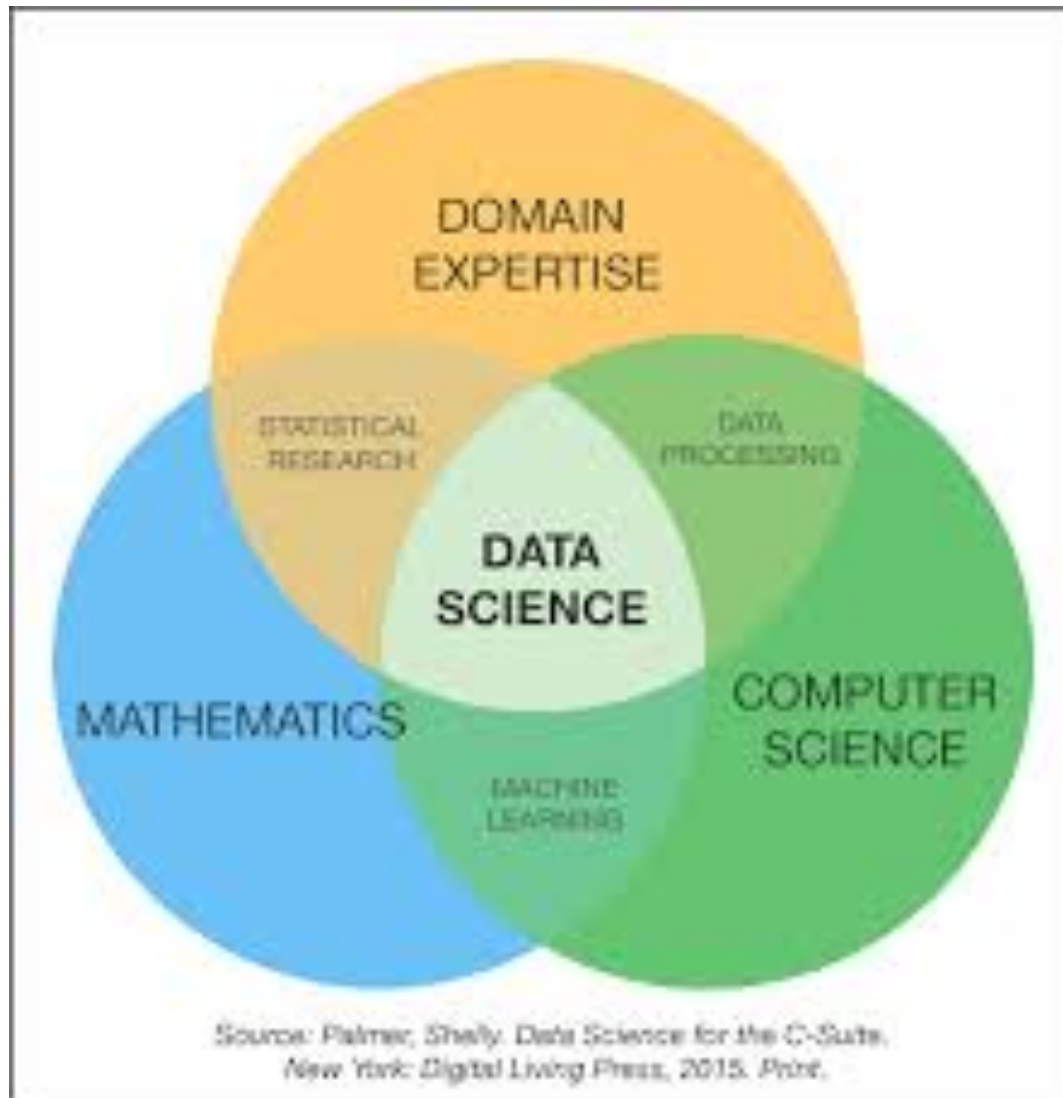
Amir Farbin

Amir Shahmoradi

What is Data Science?

- From *Wikipedia*: Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms…

- There are concepts that unify statistics, data analysis, machine learning, …

- Take techniques from mathematics, statistics, information science, and computer science

- Apply it to a domain:

  - science: biology, physics, psychology, …

  - marketing, business analytics, …

Data Science

Source: Palmer, Shelly. Data Science for the C-Suite.
New York: Digital Living Press, 2015. Print.

Data Science @ UTA

- College of Science is in the process of getting approval for undergraduate BA and BS degrees in Data Science.

- Minor in Data Science possible now.

- Tentative New Courses:

  - DATA 1301: Introduction to Data Science

  - DATA 1401/2: Introduction to Scientific Computing 1 and 2

  - DATA 3401: Data Mining, Management, and Curation

  - DATA 34xx: Statistical Modeling

  - DATA 34xx: Simulation

  - DATA 34xx: Machine Learning

  - DATA 43xx: Data Problems

  - DATA 43xx: Data Capstone Project 1 and 2

  -

Degree Goals

- If you are heading to industry: BS degree → join a data science team.

- If you are heading to grad school: BS degree → enable research.

Logistics

- *Lectures*: Monday/Wednesday/Friday

  - No official Textbook. Unique class. Not following an existing paradigm.

    - Lecture and labs are essential.

    - Welcome to use any python book as a reference.

    - We will learn to use online docs as resource.

  - *Quizzes* (10%): Short and Easy. "Pop": Randomly scheduled. Serves as attendance tool. Quiz will often integrate with the lecture.

    - Drop 2 lowest grades (including being sick, unless previously made arrangements).

- **Lab** (50%) Friday

  - Hands-on: I'll walk through something.

  - **Lab**: (15%) has to be accomplished in during the session (maybe end of day).

  - **Homework** (35%): Remainder is general is "Homework". Due by following lab session.

- **Midterm** (20%) / **Final** (20%): Comprehensive… similar to homework. Take home. 24-48 hours. No makeups.

Logistics (2)

- Homework Policy

  - You can work with others, but do not copy/paste code from another student.

  - Submitted via git (Version Control System).

- Help

  - *Clinic*: We usually run a clinic where students can ask for help… wasn't effective virtually.

    - Will run a poll to determine optimal time for an additional help session.

  - *Office Hours*: I'll generally be available after every classes session.

  - Google group… **https://groups.google.com/forum/#!forum/uta-data-1401-fall-2020**

- *Laptop* (with a physical keyboard).

  - Highly recommended to have a laptop for this class

    - Doesn't matter what OS you run… all you need is a browser.

  - If you don't have a laptop, you can rent one. Details …

- First lab will be NEXT Friday (9/4), will be to get setup.

Style

- Really a Python Programming Class taught from a Data Science perspective.

- Introduce and then reinforce data science concepts through successively more advanced implementations each introducing new programming techniques and culminating in use of advanced libraries.

- For example,

  - Learn about distributions and histogram

  - Example: Construct histograms

    - by hand

    - procedurally

    - write histogramming functions

    - write histogramming object oriented classes

    - use standard libraries that provide histogramming

Course Plan

- Class Intro
- What is Data Science?
- Conceptual Understanding of Computing
  - From Transistor to the iPhone (Hardware)
  - From Machine Language to Python (Programming)
  - From Operating Systems to Apps (Software)
- Fundamentals
  - Git
  - Basic Unix
  - Jupyter
- Python Introduction → Build your first game: Checkers / Tic-Tac-Toe
  - Data types / Syntax
  - Loops / Control Flow
  - Abstraction: Functions
  - Basic I/O
  - Modules / Libraries
  - → Build your first game: Checkers / Tic-Tac-Toe
- Linear Algebra
  - Basic algorithms
  - → Matrix Operations

- Object Oriented Programming (take 1)
  - → Matrix Class
- Probability Theory
  - Understanding Distributions
  - → Histograms and Monte Carlo / Functional Programming
- Object Oriented Programming (take 2)
  - Scripting vs Building Software
  - OO Design Patterns / UML
  - → Gradebook Software
- Data Analysis
  - Introduction to: numpy, Pandas, matplotlib
  - Representing Data
  - Visualization
  - → Basic Data Analysis
- Machine Learning
  - scikit
  - → LHC Data Analysis