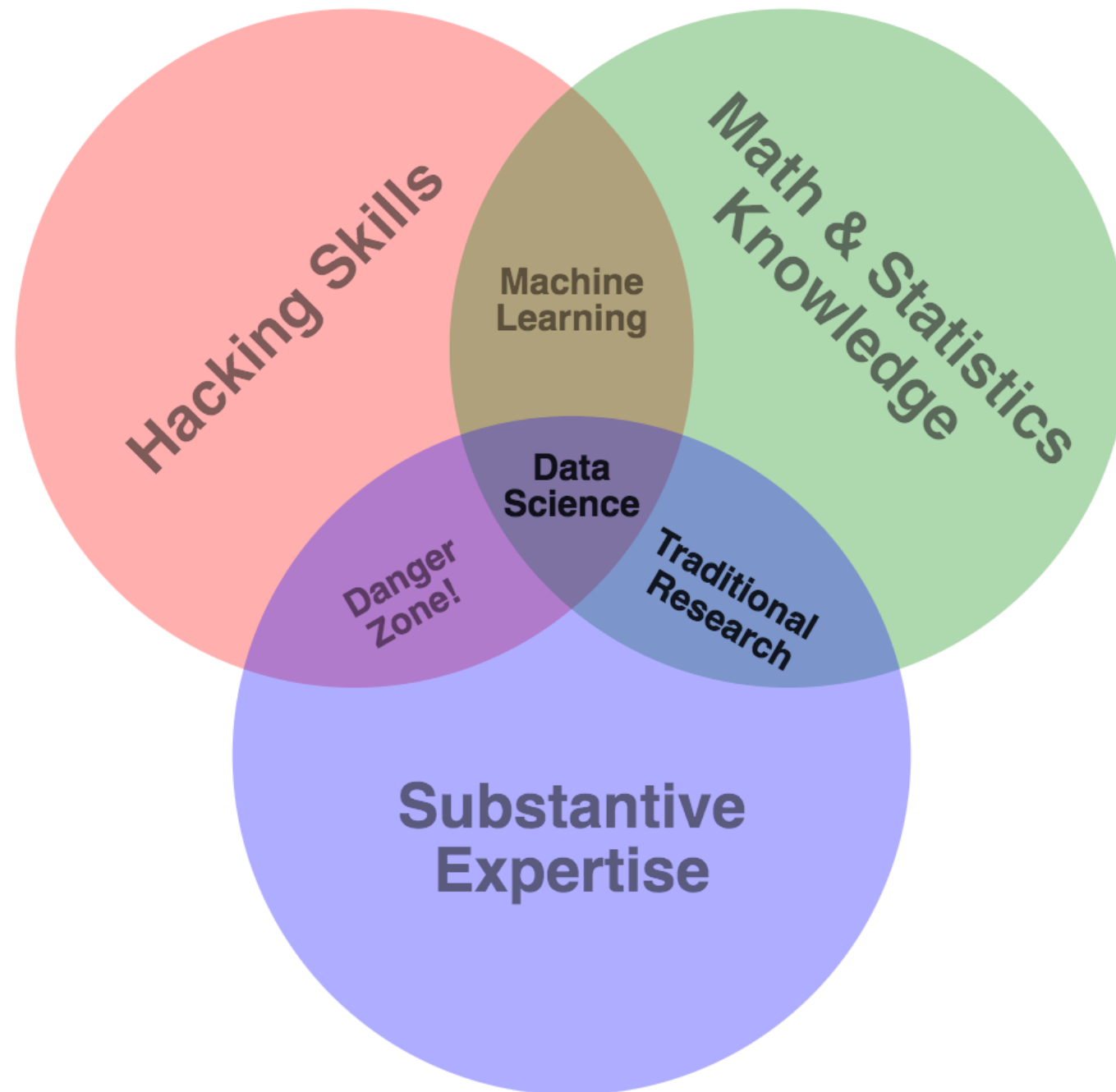


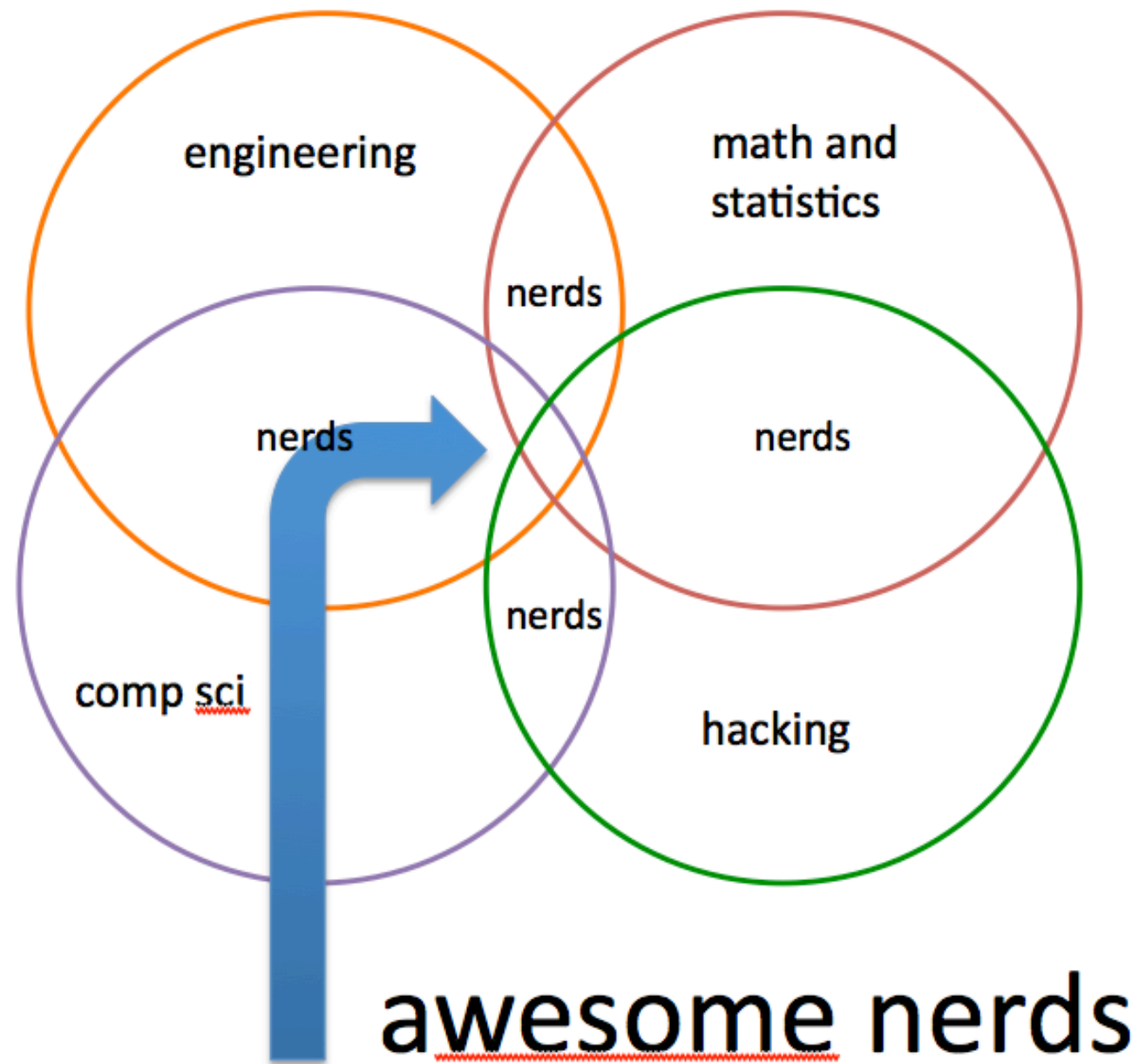
Lecture 1: Half-Day Introduction

COSC 480 Data Science, Spring 2017
Michael Hay

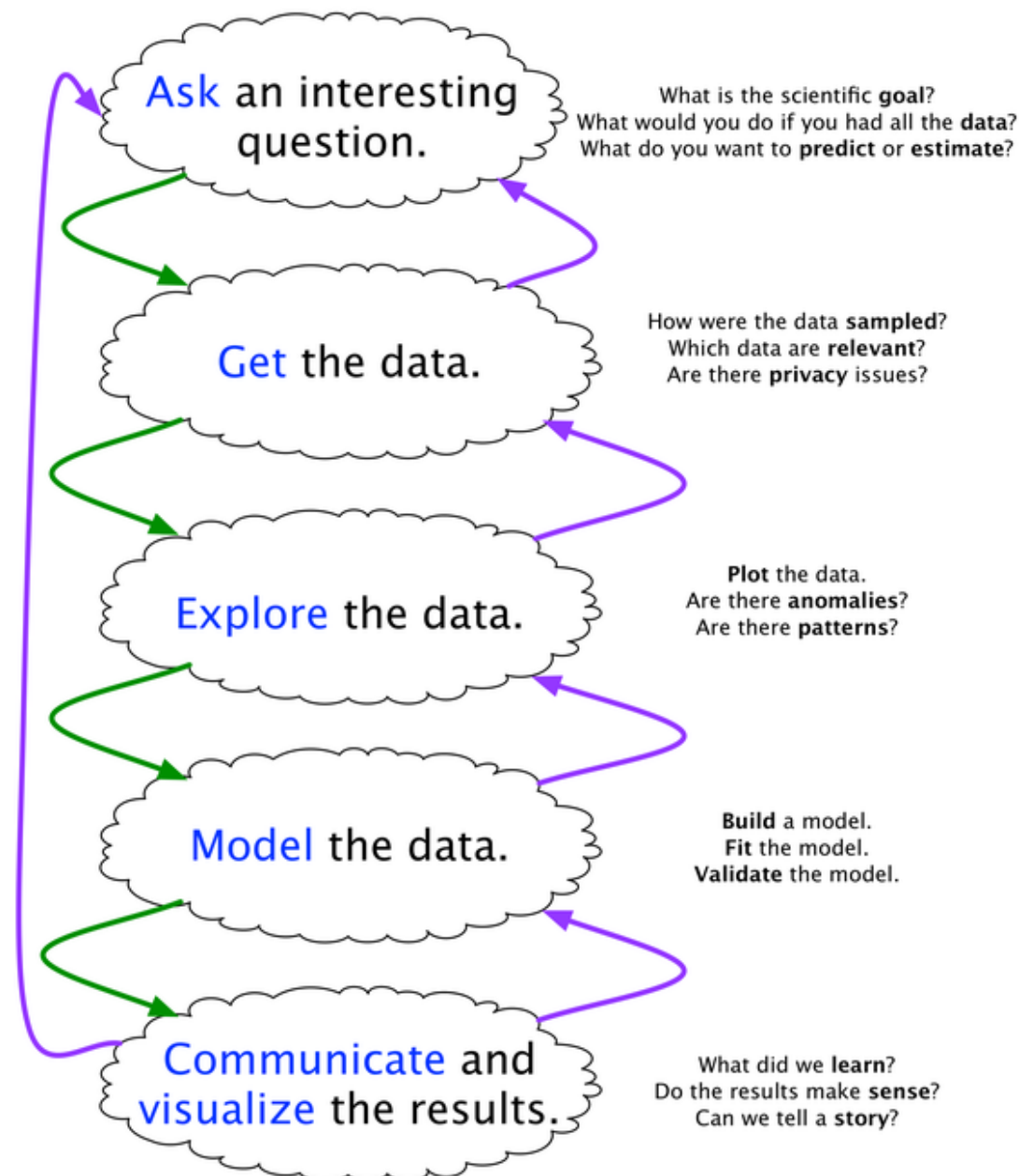
What is data science?



Data scientists?



Process of Data Science



Joe Blitzstein and Hanspeter Pfister, created for the Harvard data science course <http://cs109.org/>.

Goals for course

BREADTH

- Develop *foundational skills* for key steps in process
- *Hands-on experience* to deepen conceptual understanding
- *Exposure* to a suite of useful tools

DEPTH

- Occasional “deep dives” into details
- Course project: apply process to real data!

Course Format

- Reading *before* lecture
- Labs: hands-on, emphasis on conceptual understanding
- Quizzes
- Capstone project (no final exam!*)

* but hold time slot for the last quiz of semester

Course topics

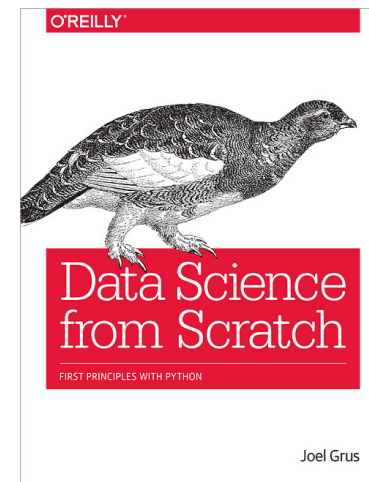
- **Data processing:** acquisition, cleaning, processing, manipulation
- **Computational statistics:** stats, probability, inference
- **Machine learning:** foundational ideas, various methods
- **Visualization:** tools, graphics grammar, principles
- **Additional topics:** different kinds of data (text, networks, streaming); ethical: privacy, fairness

Capstone Project

- Open ended, small teams (2-3)
- Two alternatives
 - *Data-oriented*: follow process outlined earlier for question/dataset of your choosing
 - *Methods-oriented*: deeper study of a specific data science tool, algorithm design/evaluation (e.g., privacy-preserving machine learning)
- Deliverables
 - Milestones throughout semester
 - Presentation (**3-6:30pm 5/3**) and final “report” (Python notebook)

Course resources

- Textbook
- Website (Moodle + github)
- Github classroom for labs
- Piazza



Announcements

- *No lab this week*
- Reading for Wednesday on course schedule

- Looking for another course? ➡ ➡

