Data Science & Computing Across the Curriculum

Jeremiah W. Johnson, Department of Applied Engineering and Sciences, University of New Hampshire
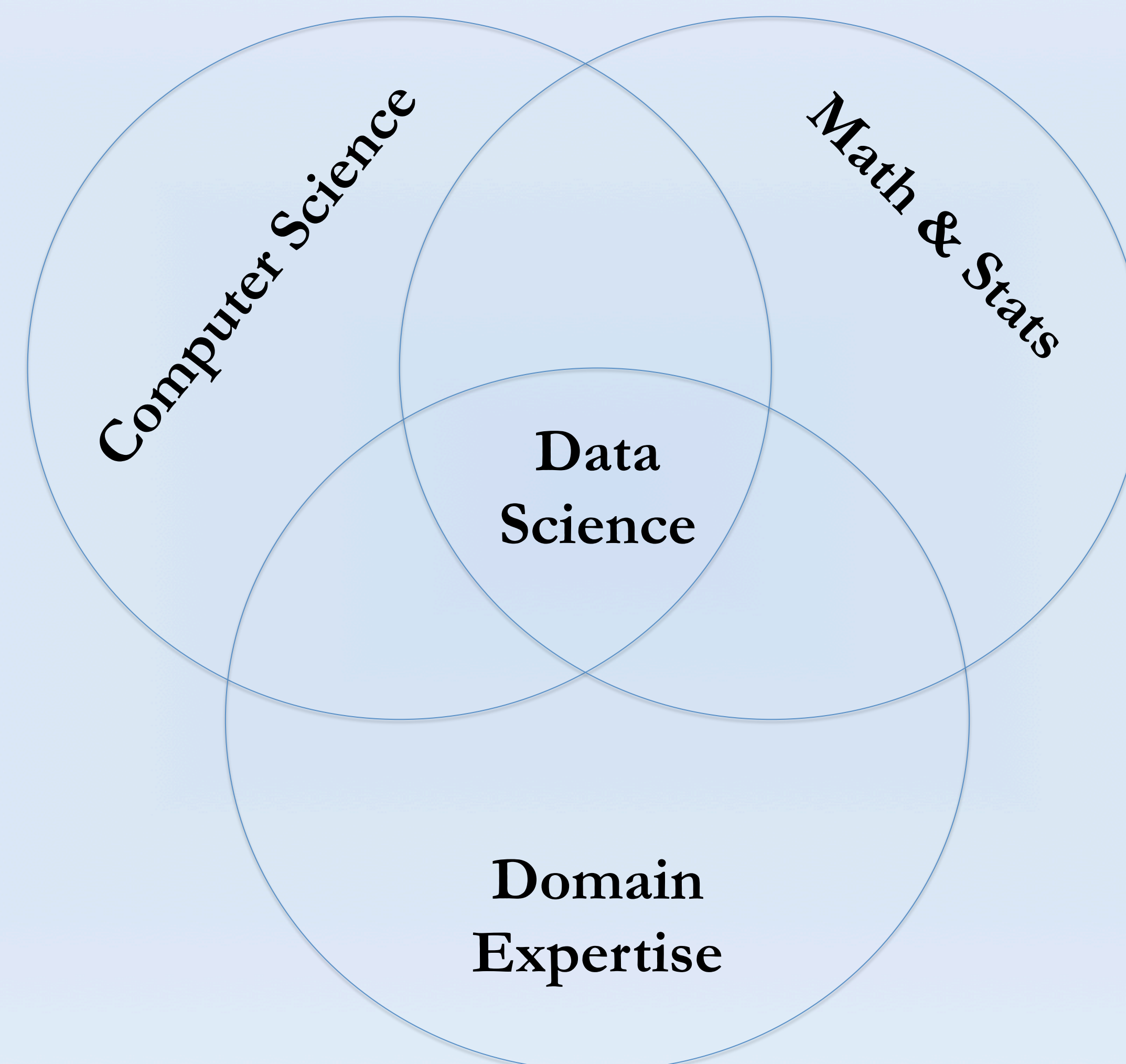
Introduction

Data Science is a multi-disciplinary field, drawing on techniques from computer science and mathematics. Although it is usually considered a graduate-level topic, data science is accessible and interesting to undergraduates[1]. An introductory undergraduate course in data science is a unique opportunity to introduce broadly applicable fundamentals of computer science to a large and diverse audience, and provides an entrée to computing across the curriculum.

Course Overview

- Introduction to Data Science & Analytics first offered at UNH in the fall of 2015
- Intended for first/seecond year students; no prerequisite requirement
- Meets a general education requirement, which brings in a wide range of students with diverse skills and interests
- Course objectives specific to data analysis, not computer science
- Project-based, 'flipped' classroom

Computing Fundamentals

Students in the course learn to program in a high-level scripting language. Students are introduced to

- Common control and program flow structures
- Data types and data structures
- Databases and query languages
- Algorithmic complexity
- Computational thinking and problem-solving

Computer Science · Math & Stats · **Data Science** · **Domain Expertise**

Example Project: NYC Taxis

Goal: Using publicly available data, determine whether faster cab drivers earn more in tips.

Some of the Challenges:

- Large amount of data
- Messy and/or corrupted data

Computing Fundamentals incorporated:

- Data types and structures
- Control structures
- Databases, complex queries
- Computational thinking and problem-solving

Conclusions & Metrics

As a proxy for the effectiveness of the course in promoting computing across the curriculum, we track the conversion rate for nonmajor students into higher-level data science and computing course. We also track the proportion of nonmajors who convert into majors in the computing and data science programs.

- Fall 2015: 25% major, 12.5% higher-level coursework
- Spring 2016: 0% major, 20% higher-level coursework

1. Anderson *et. al.,* 2014. An undergraduate degree in data science: curriculum and a decade of implementation experience. In *Proceedings of the 45th ACM technical symposium on Computer science education* (SIGCSE '14). ACM, New York, NY, USA, 145-150. DOI: http://dx.doi.org/10.1145/2538862.2538936