Non-majors Enrolling in CS Courses

Many reasons for the growth in non-major enrollment, e.g. programming and CS skills highly valued in the workplace.

I focus on data science as a driver for non-major enrollment growth.

Source: Tracy Camp, CRA Presentation 2016
Data- or Computationally-enabled Research is Pervasive

The software contains “ideas that enable biology…” Stories from the Supplement, 2013
A Story of an Undergraduate

What were the drivers behind Asian voting preferences in the 2008 and 2000 elections?

A Response to the SES Model: The Main Drivers Behind Asian Voting Preferences in the 2008 and 2000 U.S. Presidential Elections

Christine Byun
Department of Statistics, Department of Political Science, Columbia University

1. Introduction

SES Model

Of the many models in the field that explain political participation, one of the most widely accepted is the Socioeconomic Status (SES) model. The SES model states that the higher an individual's
Science and engineering bachelor’s degrees earned by underrepresented minority women, by field: 1993–2012

NOTE: Data not available for 1999.
Why are Non-majors Enrolling?

Data Science a new and compelling interest for undergraduates, cuts across domain research areas (climate, energy use, water supply, voter patterns, etc), Data Science uses foundational CS techniques, increasing demand for CS courses such as:

- software design, data structures, building packages and libraries, …
- interpreted languages: python, R, MATLAB, …
- algorithms, machine learning, scalability, …
- database management systems, …
- HPC and cloud computing, networks, …
Master of Information and Data Science, UC Berkeley

Foundation Courses (15 units)
- Research Design and Application for Data and Analysis (3 units)
- Statistics for Data Science (3 units)
- Storing and Retrieving Data (3 units)
- Applied Machine Learning (3 units)
- Data Visualization and Communication (3 units)

Advanced Courses (9 units)
- Experiments and Causal Inference (3 units)
- Behind the Data: Humans and Values (3 units)
- Scaling Up! Really Big Data (3 units)
- Statistical Methods for Discrete Response, Time Series, and Panel Data (3 units)
- Machine Learning at Scale (3 units)
Emerging Computational Science Infrastructure

Dissemination Platforms
ResearchCompendia.org  IPOL  Madagascar
MLOSS.org  thedatahub.org  nanoHUB.org
Open Science Framework  nanoHUB.org  RunMyCode.org

Workflow Tracking and Research Environments
Vistrails  Kepler  CDE  Jupyter
Galaxy  GenePattern  Sumatra  Taverna
Pegasus  Kurator

Embedded Publishing
Verifiable Computational Research  SOLE  knitR  clearScience
Collage Authoring Environment  SHARE  Sweave  Paper of the Future