

The Undergraduate Data Science Degree at Arizona State University

Eric Kostelich
kostelich@asu.edu



ARIZONA STATE UNIVERSITY
SCHOOL OF MATHEMATICAL & STATISTICAL SCIENCES

January 10, 2025

Disclaimer

- The views expressed in this presentation are strictly my own and do not represent those of the ASU administration

My background in data science

- I am one of the principal developers of the Local Ensemble Transform Kalman Filter
- **Goal:** To estimate initial conditions for an operational global weather model with quantified uncertainty
- In operational weather centers, the process must be completed in ~ 10 minutes with upwards of 10^9 obs and 10^{11} model grid points
- Unlike 4DVar, the LETKF does not require a tangent linear model and so is relatively easy to implement
- Now used operationally in Japan, Brazil, and elsewhere

- Fall 2024 enrollment: $\sim 153,000$ students
- This count includes $\sim 73,000$ enrolled in ASU Online
- The School of Mathematical and Statistical Sciences (SoMSS): enrolls $\sim 35,000$ unique students each year in courses at all levels
- Online enrollments are the fastest growing and include partnerships with Starbucks and Uber, among others
- The same ASU faculty develop and deliver online and immersion courses

Some history on the B.S. in Data Science

- An external program review of SoMSS in 2016 provided significant impetus
- Both SoMSS and the College of Liberal Arts and Sciences (CLAS) seek more majors
- Data science seemed like a good fit for SoMSS
- The degree launched in August 2020 with 13 majors and has grown to 1,262 majors as of Fall 2024
- ~ 70 percent of ASU majors are online: data science is in high demand for “upskilling” and “reskilling”

All politics is local

- Many units offer courses in statistics and plausibly “own” data science
- Negotiations at ASU involved all the natural and social science units in CLAS plus the engineering and business schools
- **Compromise:** CLAS got the B.S. in Data Science (with SoMSS as the lead unit) and Computer Science got the Ph.D. in Data Science (with SoMSS as a junior partner)

The Bachelor of Science degree

- **Fundamental prerequisites:**
 - 2 semesters of calculus
 - 2 semesters of programming
 - 1 semester of linear algebra
- 6 additional “core” courses, taught by SoMSS
- Students pick one of 7 possible “tracks,” each of which requires 6 courses
- **Total:** 17 courses (51–55 credit hours), plus general education requirements (120 credits needed for graduation)

QR code for the major map



The core data science courses

- **Data Science and Society** (text: *Ethics for the Information Age*, by Michael J. Quinn)
- **Mathematical Tools for Data Science** (regression, gradient descent)
- **Exploring Data in R and Python**
- **Statistical Modeling and Inference for Data Science** (bias-variance tradeoff, Bayesian and frequentist approaches, causal inference, cross validation)
- **Machine Learning** (neural networks, regularized regression, dimension reduction)
- **Data Science Capstone** (according to track)

The data science tracks (6 courses in 1 track)

- **Behavioral Sciences** (social science statistics, data analysis)
- **Biosciences** (genomics, computational biology, neuroscience)
- **Business analytics** (big data analytics, data warehouses)
- **Computer science** (data structures, algorithms, AI/ML, security)
- **Mathematics** (upper division math, stat, actuarial science)

The data science tracks (6 courses in 1 track), 2

- **Social sciences** (economics, sustainability, statistical designs)
- **Spatial sciences** (geographic information systems, spatial statistics)
- Exact counts of students per track are problematic, but my guess is that most are in math, CS, or biosciences

Curricular design and administrative challenges

- An interdisciplinary focus sounds great in principle—and can reduce infighting among competing academic units
- **Curricular challenge:** Assemble a coherent set of courses from diverse units without a ballooning set of prerequisites
- **Administrative challenge:** Who pays for which course sections? Negotiations can involve several deans and department heads

Staffing issues and scalability concerns

- Instructor recruitment and compensation are a challenge
- It helps to leverage faculty resources in other units (but cost sharing and teaching loads must be negotiated)
- **Clinical professor** roles (imported from health sciences) are one way to provide higher salaries without creating new tenure lines
- I expect 2,000–3,000 majors eventually
- Tools like **nbgrader** are needed to allow instructors to efficiently serve 100+ students per course

Hardware and software considerations

- We use PyTorch for training large language models (LLMs)
- Students can train a small LLM for a homework assignment in ~ 15 minutes on a multicore laptop
- NVIDIA chips can be rented for a few dollars per hour on cloud services (e.g., cr8dl.ai) for more advanced projects

The online learning challenge

- Current 4-year degree enrollment is 8–9 million
- There are ~ 37 million adults in the U.S. who started a college degree but never finished
- Providing opportunities to even 10% of this group is a major challenge
- Purchases by some traditional universities of online schools have not yielded the promised benefits
- **ASU Online:** Because the same faculty develop the courses and teach them across all modalities, administration and quality control are simplified