A Data Science Major: Building Skills and Confidence

Stephanie Rosenthal*
srosenth@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, PA, USA

Tingting (Rachel) Chung*
rachel.chung@mason.wm.edu
College of William & Mary
Williamsburg, VA, USA

ABSTRACT

Data science is a growing field at the intersection of mathematics, computer science, and domain expertise. Like many universities that are building data science degree programs for undergraduates, our small, liberal-arts university saw increasing opportunities in the region and decided to build a data science degree from the ground up, without a pre-existing computer science (CS) department to leverage for courses or culture. We designed and implemented an academically-demanding curriculum that combined mathematics, information systems, and new data science courses, and that also encouraged and supported student success. Each introductory course included active learning design to engage students. To increase retention, all major courses included assignments designed to build skills but also student confidence in their ability to learn challenging technical topics. Outside of the classroom, we created opportunities for professional advancement and developed a technical culture at the university. We will share our approach, course highlights, and lessons learned from building such a curriculum at an institution without a CS department.

CCS CONCEPTS

• **Social and professional topics → Model curricula**; **Computer science education**.

KEYWORDS

Curriculum design; data science major; data science curriculum

ACM Reference Format:
Stephanie Rosenthal and Tingting (Rachel) Chung. 2020. A Data Science Major: Building Skills and Confidence. In *The 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20), March 11–14, 2020, Portland, OR, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3328778.3366791

1 INTRODUCTION

Data science, sometimes also referred to as data analytics, is an emerging, interdisciplinary field to manage, organize, and extract "actionable knowledge directly from data through a process of discovery, or hypothesis formulation and hypothesis testing" [6]. Data scientists leverage technical skills and systems for working with data to solve problems in specific application domains. Good data scientists must have a strong foundation in statistical analysis to make sense of their data, and communication skills to disseminate their results within the context of their organization. Data science is not simply about the mechanics of data processing, but also the legal and ethical considerations of collecting, storing, and analyzing massive amounts of information [5].

Historically, data scientists have entered the workforce through computer science and statistics degree programs and many of them have PhDs [15]. However, in recent years, more universities are adding data science majors at the undergraduate level in order to meet the workforce demand of this growing field. In December 2016, the McKinsey Global Institute published a report indicating that, in the near future, there could be a shortage of up to 250,000 jobs in data analysis despite 7% growth in data scientists entering the workforce, many through these new undergraduate degree programs [19].

With a cursory review of current data science programs across the US and internationally, we notice that data science courses and entire majors do highlight the interdisciplinary nature of data science and/or data analytics at both the undergraduate and graduate levels. These programs focus on the math, science, and computer science (CS) aspects by leveraging pre-existing courses in computer science, statistics, and machine learning. However, this methodology is simply not feasible for the many small universities that do not have pre-existing technical majors to grow from.

On small liberal-arts university campuses without CS like our own, few students arrive with prior interest in computer science or data science disciplines [1, 3]. Students with little interest may find the discipline to be intimidating. They may have come to the institution for other intellectual pursuits. At the same time, data science programs must be rigorous enough for students to be able to compete for jobs against their peers at other universities. Convincing students that data science is a useful discipline and all students have the potential to succeed is quite challenging yet similar to the challenges faced when broadening participation in computer science [12, 20, 22].

Many computer science degree programs, and STEM programs in general, have been proposing ways to attract and retain students for decades. Our approach to attracting and retaining students largely follows Jolly, Campbell, and Perlman's Engagement, Capacity, and Continuity trilogy of characteristics for success in STEM fields [4, 13, 14] that have been applied to many different settings (e.g., [18, 23]). In particular, they argue that in order to be successful, students must be *engaged* to be attracted to the major, they must be shown they have the *capacity* to learn the material, and they must be provided with institutional and programmatic *continuity* and opportunities for advancement in the field.

*Both authors were faculty at Chatham University at the time.

ACM ISBN 978-1-4503-6793-6/20/03…$15.00
https://doi.org/10.1145/3328778.3366791

We applied this trilogy throughout the design of the data science major curriculum, the implementation of the classes, and culturally through advising sessions and interactions with students. We asked students to evaluate the curriculum and their experiences before and after each of their classes. In this paper, we discuss our university learning environment, approach to the data science major, learning objectives and curriculum, student learning outcomes in several classes, and insights and challenges of implementation.

2 UNIVERSITY AND LEARNING CONTEXT

Our University has around 1,100 undergraduate students total across all majors in the fields of Humanities, Arts, Business, Sciences, Social Sciences, Education, Architecture, Sustainability, and Health. Traditionally, our students value very small class sizes and one-on-one attention from professors. Without many technical courses, the University tends to attract students with little to no math or computing interest or experience. However, it is located in the heart of our city's high-tech industry, and is surrounded by an increasing number of high-tech firms seeking talent and several higher-education institutions with strong degree programs in computer and data science. Because of the growth of tech companies and technical positions in the city, there are many opportunities and also competition for jobs in the area.

In 2016, the Business and Science Departments collaborated to propose a new major in data science to bootstrap the University's offerings in applied technical degrees and to capitalize on the growing need for data scientists regionally and globally. The University's administration supported the rapid launch of the major and the decision to house the major within the Business Department. In the Business Department, the data science students have ample opportunities to work with and learn from business students and faculty in order to become business translators who not only understand the data science but can also translate those results to a non-technical audience [19]. At the same time, the intensive focus on computing and mathematics make the data science students' learning experiences quite distinct from that of their peers in different business majors (marketing, accounting, management, etc).

3 CURRICULUM DEVELOPMENT

Without many existing technical courses to leverage, we were able to design the learning objectives and the curriculum from the ground up, based on our students' needs as well as important skills that the industry is looking for in graduates today. In addition to determining courses and topics, we also focused attention on strategies to attract and retain students from our University to the major, given the potential lack of interest in data science and lack of confidence in their math and science skills. In the end, our goals were to achieve academic rigor to ensure our students can compete for jobs regionally and globally, while also helping the students be successful by 1) making the topics approachable for students with no prior computing experience and applicable to students' broad interests and 2) building student confidence in their new skills and demonstrating that they have the intellectual capacity to succeed in the major, and 3) providing students with the opportunities outside the classroom to advance their careers.

3.1 Support Inside and Outside the Classroom

To design the data science major holistically, we chose to draw from Jolly et al.'s Engagement, Capacity, and Continuity trilogy of characteristics for success [13] as it aligned well with our goals. It allowed us to think about the entire culture around the major, including recruitment, retention, graduation and professional relationships.

Engagement. We believe that the placement of the major in Business made it more approachable to students with little to no technical confidence or experience nor prior interest in a technical education. We use introductory business classes as an opportunity attract and engage with students who may have an interest in the applications of data science within business domains such as marketing and accounting and also within other majors such as sustainability that heavily depend on data. We also work to recruit students in our business classes who are very curious, willing to work hard, and willing to learn from mistakes. Once a student enrolled in our courses, we take an active learning approach to encourage continuing engagement with the material within lectures and in homework assignments, as it has been recognized to do in statistics education [16] and computer science education [17].

Capacity. We focus on building self-confidence and pride in their work in order to help retain students who enroll. We structure our classes to support this learning style by providing frequent reassurance, help, and feedback as students learn new skills. Our assignments and subsequent class discussions demonstrate and reinforce that all of the students have the capacity to learn the new material and can succeed. Because data science often requires failures among successes in order to build intuition for data and models, we also encourage a "growth mindset" [7] and perseverance through challenges [11].

Continuity. Because the major is unique at our University, we also needed to build a community around data science outside of the classroom. We created active mailing lists where we can post relevant articles and jobs. We hold resume and interview preparation sessions, and encourage active participation in two new clubs that were started by students with our help. The clubs allow students to interact professionally outside of class, share their experiences, and support each other through different challenges they may face. Finally, we formed a data science discussion series to bring in faculty from other fields to speak about their needs, applications and challenges with data, and discuss internship and job opportunities. All of these efforts build a culture around the major and a reputation that students are well-supported and successful.

3.2 Courses

Based on data science curriculum guidelines (e.g., [2, 5, 6, 8–10]) and resources available at the University, the major was designed to take advantage of existing and revitalized courses in Mathematics and Management Information Systems (MIS) with a handful of newly designed courses on Data Science (Table 1).

Our curricular approach is heavily scaffolded with many courses requiring others as prerequisites. Because our students enter with almost no experience in computer science or programming and many had never considered a technical major, we designed the first several courses to build up their skills and provide many opportunities to practice the skills in and out of class. We teach computer science skills in Introduction to Programming which includes a

Course Name	Topics	Prerequisite(s)
Intro. to Programming	Representation, Abstraction, Functions, File I/O	N/A
Intro. to Data Science	(see Section 4.1)	Intro to Programming
Data Visualization	Psychology of vision, Graphics, Interaction design	Intro to Data Science
Machine Learning and AI	(see Section 4.3)	Intro to Data Science, Calc. III, Lin. Algebra
Data Science Senior Capstone	(see Section 4.4)	(senior status)
Statistics	Introductory Probability and Statistics	N/A
Discrete Mathematics	Logic, Proofs, Computability	N/A
Calculus I	Derivation	N/A
Calculus II	Integration	Calculus I
Calculus III	Multivariate calculus	Calculus II
Probability	Simulation, Markov Models, Probability Distributions	Statistics
Linear Algebra	Matrices and Linear Algebra	Calculus III co-requisite
Information Systems and Operations	Introductory Networks, Hardware, Software, Databases	N/A
Information and Cybersecurity	Vulnerability and Threat Analysis and Mitigation	Info. Systems and Operations
Database Management Systems	(see Section 4.2)	Intro. to Programming
Business Research Methods	Research methods and Analysis	Statistics

Table 1: Data Science Major courses and descriptions divided into three sections - data science core classes (top), mathematics (middle), and business/information systems (bottom). Bold represents new or revitalized courses for the major. Note that the business courses are non-technical introductions and overviews except for Database Management Systems.

culminating project on an application of their choice to keep them engaged (commonly video games and robotics). Similarly, the Introduction to Data Science course teaches data science-specific skills within the data science process and then provides students the opportunity to repeat the process on a data science project of their choice (e.g., sports analytics, accounting audits and fraud, box office ticket sales for movies). These projects create a sense of pride and accomplishment after a semester of hard work, and builds their confidence in their technical capacity.

The mathematics courses build maturity around the concepts of optimization that our culminating core course - Machine Learning and Artificial Intelligence - requires. By the time students take the Machine Learning course they are comfortable reading and understanding the mathematics of optimization and can also algorithmically translate the math into machine learning programs. Our business courses help students use their data science skills to solve real business problems and help them learn how to explain their work to and interact with business professionals. They get hands-on experience in applying all of their skills in their Data Science Capstone project course where they were assigned a project to help analyze data for the Business Capstone students.

Unlike data science majors at technical universities, we are unable to offer engineering courses on cloud-based big data platforms or hardware, as such coursework would require dedicated engineering staff support. We also devote little coursework to vendor-specific platforms or tools, focusing instead on general-purpose programming languages such as Python and MySQL. The design decision was deliberately made to foster academic concepts and skills that are versatile independent of platform environments.

3.3 Learning Objectives

The major's learning objectives reflect the interdisciplinary nature of the profession:

- Create effective mathematic solutions to analytical problems
- Create effective solutions to computing challenges in analytical projects
- Effectively organize and manage datasets for analytical projects
- Critically analyze problems and identify analytical solutions
- Communicate analytical problems, methods, and findings effectively orally, visually, and in writing
- Critically evaluate ethical, privacy and security challenges in data analytics

Each class incorporates at least two of the learning objectives. Technical, analytical and communication skills are all necessary to be successful in the data science field and practiced frequently in projects and presentations. Notably, ethics and privacy are interwoven through discussions and assignments in each of the courses.

4 COURSE HIGHLIGHTS AND EVALUATION

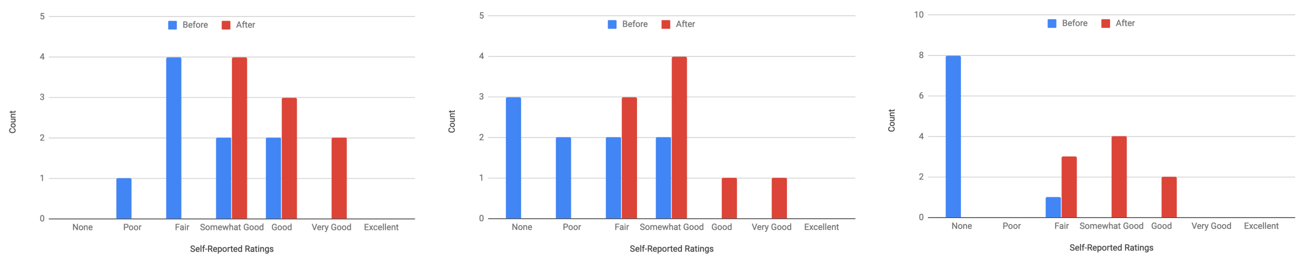
In this section, we highlight several of our required courses including how we incorporate engagement, capacity, and continuity into their structures. Additionally, we report on the student evaluations of the courses and the major.

4.1 Introduction to Data Science

The Introduction to Data Science course serves as the first experience in the field. Our course has been taught two times with eight students each semester.[1] Students in this course have taken programming (prerequisite) but have not directly applied their skills to data science. The course focuses on the data science process: data collection, organization, analysis, machine learning, and visualization/communication of results. Algorithms for downloading, iterating through, parsing, cleaning, summarizing, graphing, and analyzing data are introduced. Differentiation between and intuition behind the broad categories of machine learning techniques give students practical knowledge about how to choose which algorithm to train but not how to implement them algorithmically (yet). Python, Pandas, Jupyter Notebooks, and SciKit-Learn are used.

To make the course approachable and engaging, we split it into two parts. The first half covers all of the steps of the data science process. In the first week, students are introduced to the topic of

[1] While these class sizes seem small, they reflect about 1% of the entire undergraduate student population and 4% of the first year students at the University per year.



(a) Programming **(b) Visualization** **(c) Machine Learning**

Figure 1: End-of-semester student assessments of their knowledge from before and after Introduction to Data Science.

"I think learning the data science process and actually doing it twice was most enjoyable. And, the best is getting a good accuracy at the end of it all."
"Overall I really enjoyed the course, I particularly enjoyed being able to do the data science process twice. Once with the 4 smaller assignments and another with our own projects."
"being able to do the data science process twice helped solidify my knowledge around data science."
"The most challenging part of the class is connecting it all together. It is easier to do each step alone, but as soon as it all has to happen within one project, it seems to become much harder."
"The first 4 assignments. I think that comes from learning many new things and the trial and error involved. But it was good prep for the project, which made it easier."

Table 2: Quotes from students of Introduction to Data Science about what they enjoyed about the course.

research and hypothesis generation and must explore and hypothesize about one of four data sources - weather data, news feeds, twitter, or earthquake reports. Then, as they learn about each step in the process, students are given assignments in the context of their chosen data and hypothesis: downloading the dataset programmatically (1 week), parsing and cleaning the data (1 week), generating new features (2 weeks), and modeling or analysis (2 weeks). Assignments allow the students to apply what they are studying in class with an emphasis on data science process, building intuition, and learning problem solving techniques. They receive ample support in office hours towards the completion of the assignments and are awarded a lot of partial credit for attempts.

In the second half, we engage students with discussions and speakers on topics like data ethics and privacy, and other popular topics that they hear about but have not studied (e.g., self-driving cars, recommender systems). During these discussions, students realize that they already have the basic skills to understand and potentially implement these ideas. Outside of class, students propose a project to apply the data science process a second time. This time, they have confidence in their ability to write the code. They receive less-structured help and focus on independent problem solving. The last week of the semester is spent listening to students' project presentations, and discussing connections and similarities between them. Projects are assessed on their output - understandable code, correct models chosen, correct analysis, appropriate visualizations and presentation. Reinforcing our goal of capacity, students spoke about how accomplished they felt writing several hundred lines of code and in completing the data science process independently.

At the end of each semester, we asked students to self-evaluate their skills on a 7-pt scale in computer science, statistics, machine learning, and visualization and also to provide feedback about what they enjoyed and what was challenging in order to improve the course. The assessment instrument was developed and used in prior data science courses [21]. Nine students responded out of sixteen.

Figure 1a shows the students' perceived skill at programming from the beginning to end of the semester. Students reported not only an increase in programming ability but also in debugging and problem solving which are built up through practice on the assignments and the project. One student wrote "My debugging, problem solving and critical skills have become significantly better. I also know how to reframe my questions when trying to find answers from students, professors or the internet." Similarly, the students reported good skills in visualization after the course compared to having very little experience before the semester began (Figure 1b). Most notably, Figure 1c shows that students had no understanding of machine learning before the semester but rated themselves as having some skills by the end of the semester[2]. Students know how to look at their data, determine which machine learning algorithm is appropriate, and use Python SciKit-Learn to train and test a model.

We also asked students what they enjoyed most about the course. Overwhelmingly, they loved the structure of 4 programming assignments and a project. Some of their responses are in Table 2. We believe that the enjoyment of the assignments is directly related to their challenging yet engaging and confidence-building structure. The students, many of whom were extremely tentative about programming at the start of the semester, are given the opportunity to learn with extensive support to gain confidence. By the time they produce a project, they can predict the potential challenges that they will face and have the tools and support to persevere.

4.2 Business Information Systems Courses

Because much of the world's data are stored and organized in a relational format, we require our students to be proficient in relational database design and queries. The Information Systems and Operations (ISO) course, required for all business and data science majors, makes use of active learning assignments to engage students with no prior interest in technology. To bring relational database

[2]We only expect a broad understanding of machine learning upon course completion.

concepts to life, ISO students are asked to run grocery stores that sell candy. Students play different roles, such as customers who purchase candy using fake money, managers who hire and fire employees, and staff members who update records in a non-relational data table. This exercise exposes students to data integrity issues related to sales transactions, organizational changes, and human resource turnover. By creating chaos in the classroom and in the data table, students discover the importance of thoughtful database design and management for organizations and walk away from the course having a deep understanding of information systems. Those that enjoy the topics are encouraged to take more courses.

The Database Management Systems course is a requirement for MIS, Accounting, and Data Science majors. Introduction to Programming was added as a prerequisite, so students entering the course have foundational skills in coding and debugging. As a result, the course could include more technical content than had been covered previously. The first half of the semester focuses on the rationale for relational database designs and consequences of good versus bad designs. Students practice designing tables, relational structures, and normalization through interesting active learning activities and assignments. In the second half of the course, they work in teams to implement their database designs in MySQL, and develop queries to retrieve datasets to address research or business needs. Upon completion of the course, students are expected to make sound judgments about relational database designs, and to be competent SQL query writers.

This approach of using assignments for conceptual learning and a project to reinforce ideas again allows the students to practice critical skills twice, which reinforces the skills and builds professional confidence while engaging them with topics they care about. Coincidentally, all students from the most recent semester chose to work on a database design and development project for a campus function (e.g., student clubs, campus cafe, campus garden). Therefore, it is clear that the students see the value of relational databases for improving the community they care deeply about. Students in the data science major tend to think more clearly about how data can be organized and accessed after taking this course. Additionally, they understand the business applications of data and can communicate this understanding more clearly to their peers in business.

4.3 Machine Learning and AI

Machine Learning and AI is the final core course of the data science curriculum. It requires mathematical maturity in Calculus III and Linear Algebra, and it requires Introduction to Data Science. We assume students have been programming for at least a year. The course covers many of the basic machine learning algorithms - naive bayes, logistic regression, linear regression, support vector machines, k-nearest neighbors, bayes nets, and decision trees - as well as topics related to properly training and testing models - bias and variance, accuracy, precision, recall, feature reduction, kernels, and ethical bias. It also covers the basics of artificial intelligence - representation, search, heuristics, and constraint satisfaction.

The homework assignments are the same or very similar to those given in machine learning and AI courses at top computer science programs across the US. Students implement algorithms from scratch and assess their properties using given datasets. Successful completion of these assignments once again helps to reinforce that 1) they are capable of doing the same level of work as peers at other universities and 2) they can compete with those peers for the top jobs in data science. While the course is extremely rigorous, the students have the skills to persevere through challenges and ask for help when necessary. They also support each other and can provide each other with conceptual and debugging help.

4.4 Data Science Capstone

Each major at the University is required to have a senior Capstone course. When designing the Data Science Capstone, our goal was to structure the project as if the data science students were supporting business practices within an organization. We decided to have the data science majors support the business capstone projects - pitching their own food trucks. Our team of data science majors analyze Yelp food truck reviews, and the resulting reports are used as supporting evidence in the pitch marketing plans. The data science students are required to create hypotheses, collect and analyze their data, present their findings in a non-technical format to their business peers (harder than it looks after spending several years practicing technical presentations!), and also write a detailed technical report. This format gives the students the opportunity to practice working with non-technical professionals while also demonstrating their skills.

Because these presentations have to happen with enough time for the business students to use the information, the data science students have the opportunity to work on a second smaller project to optimize the food purchases and profits for each food truck. This logistics project is quite challenging for the students and generates very lively, engaging discussions during class and office hours about how to represent the problem so that it could be solved in a reasonable amount of time. While both projects are worthwhile, we will likely pick only one project in the future because the cost for changing projects in the middle of the semester was high.

4.5 Evaluation of the Major

Our first three graduating data science seniors reviewed the major and their senior level courses using the same evaluation survey from Intro. to Data Science. These first students changed major into data science and completed the entire major in two years.

Overall, we saw greatest self-reported improvement in programming, visualization, and machine learning skills in the major. Two students with some prior programming experience reported their programming skills increasing from Good to Excellent, and one who had no prior experience reported Poor to Good. All students reported their visualization skills improving from Fair to Good. Students reported their machine learning skills improving from Poor (2) or None (1) to Good (2) or Very Good (1). The students reported that their skills in Probability/Statistics, Mathematics, Computer Science, and Communication also improved, but only by one scale point on average. Most surprisingly, although the students had taken many math courses, they did not feel that their skills were much better than before - we certainly saw marked improvement.

Based on the major and course evaluations, we believe our goals were met for creating an academically rigorous program that is also approachable, engaging, and supportive. One student summarized the major perfectly:"This major has challenged me in every way possible, which has been extremely valuable for me in my professional and personal life. I feel that I've learned so much about team work, persistence, googling, asking for help, logical thinking, communicating my problem, presenting my problem, presenting my work, learning a bit about everything, how to teach myself concepts and more. I really enjoyed getting a taste of everything and the challenge that each assignment brought. At every turn I wanted to bang my head off a table and probably did but I still saw an opportunity to grow whether in my professional soft skills or in my knowledge." Additionally, our efforts to support the professional advancement of the students was also met. All three students had internships at universities or companies for at least one summer. Upon graduation, one student was hired as a data scientist locally and the two other students are pursuing graduate degrees.

5 INSIGHTS AND LESSONS LEARNED

We summarize our insights into building the major, and highlight challenges and opportunities for improvement, and summarize similarities of our major to other peer institutions.

Engagement through Active Learning. We have found that active learning is critical for attracting and retaining students in our data science major. However, it poses a challenge to support a wide range of projects as class sizes get larger. Our Introduction to Programming class increased from 16 to 40 students in one semester. One way that many schools scale courses is through teaching assistants. However, it was not possible to find many students qualified to support the class because it was new. This is also a problem in our upper-level data science courses where most students graduate before the classes are offered again. As a result, students rely solely on the instructors for office hours, questions and answers, advising and more. Because the faculty are stretched thinner, it is hard to maintain active engagement with all students. We continue to work on ways to support the students as the major grows.

Capacity, Intuition and Bravery. Data science is very intuition-based, and our courses reflect our belief that many examples are needed to gain intuition about data and models in addition to gaining confidence in their own skills. Also reflected in our course designs is the support to allow students to fail to complete an assignment and still learn from that experience. Both the lack of intuition and the inevitable struggle through the learning process tend to make students very anxious, uncomfortable, and even make them doubt their capacity for doing data science which can lead to retention issues. One goal moving forward is to continue building a *brave* community of data scientists who are comfortable taking risks, attempting on hard problems, and participating in hackathons and research to gain additional experience and intuition.

Continuity. To improve retention and encourage students to stay with the major despite setbacks and frustrations, we worked diligently to bootstrap a data science student community that is commonly available at universities where technical programs are well-established. We supported the formation of two student clubs. We facilitated meetings with industry hiring managers, presentations by alumni working in related industries, resume workshops, and company visits. We encouraged students to participate in local coding events and national analytics competitions. We connected students to research internships, and created research internships

	Size	Home Dept	%CS	%MTH	%MIS	%DS
Ours	1105	Business	12%	40%	24%	24%
Peer 1	3477	Business	12%	24%	40%	24%
Peer 2	1837	Math/CS/Stat	25%	33%	17%	25%
Peer 3	1634	CS	28%	36%	0%	36%
Peer 4	2005	CS	25%	17%	0%	33%

Table 3: A comparison of Bachelor of Science Data Science majors from our and four peer small liberal-arts institutions in terms of undergraduate size, and percent of CS, Math, Information Systems (business), and Data Science courses.

on our own to give students opportunities to apply their skills on real-world projects. Even so, students felt like there was a disconnect between classwork and professional work and said they would have appreciated even more corporate interactions in class.

Additionally, despite the extensive support we offered and encouraged the students to offer each other, we still lost several students in their first year because of the intensity of the coursework. Some of the attrition is understandable as we work to maintain academic rigor to make our program competitive with others in the region. However, we suggest developing a strong mentorship program early to help students who may be feeling frustrated with the challenging curriculum.

Data Science at Other Small Schools. Since the inception of our new major, other similar-sized institutions have created Bachelors of Science in Data Science majors. Table 3 outlines several programs and the percentage of their required courses that are in computer science, math, management information systems, and data science[3]. Our major has more math requirements than other schools; we include calculus and linear algebra in addition to statistics. Peer 1 is most similar to our school which houses the major in the business department. However, with fewer math courses and more non-technical MIS courses, their curriculum seems to focus less on the mathematical foundations of data analysis. Peers 2, 3, and 4 have more CS requirements and 3 and 4 include no MIS courses at all. With CS as their home department, these data science programs could draw from existing technical classes, culture, and other resources to launch their data science majors.

To conclude, we created a data science major at a small university with few technical courses and no CS department to leverage. We were able to build an academically rigorous curriculum from the ground up, while making it approachable for students with no prior technical experience. Following Jolly et al's trilogy, we designed courses that 1) valued engagement and active learning with hands-on assignments and projects, 2) built confidence in students' learning capacities, and 3) developed community and advancement opportunities to help students succeed post-graduation. Students evaluated their experiences in the major and reported significant skill improvement in programming, visualization, and machine learning. Looking forward, the data science major and its newly added minor are growing. It has five more majors in the pipeline for graduation in future years, a dozen students interested in the minor with majors in accounting, math, and sustainability, and the MIS degree is also growing with students who decided not to focus as heavily on the mathematics of data science.

[3] 25% of Peer 4's courses are in a depth outside of these areas.

REFERENCES

- [1] Reni A Abraham, John R Slate, D Patrick Saxon, and Wally Barnes. 2014. College-readiness in math: A conceptual analysis of the literature. *Research & Teaching in Developmental Education* 30, 2 (2014), 4.
- [2] Daniel S. Alexander and Bradley Meyer. 2016. Equipping Liberal Arts Students with Skills in Data Analytics. (2016). www.bhef.com/sites/default/files/BHEF_2016_DSA_Liberal_Arts.pdf
- [3] Matthew C Atherton. 2014. Academic preparedness of first-generation college students: Different perspectives. *Journal of College Student Development* 55, 8 (2014), 824–829.
- [4] Patricia B Campbell, Eric J Jolly, and Lesley Perlman. 2007. Introducing the trilogy of success: Examining the role of engagement, capacity and continuity in women's STEM choices. *Women in Engineering ProActive Network* (2007).
- [5] Price Waterhouse Cooper and Business-Higher Education Forum. 2017. Investing in America's data science and analytics talent - the case for action. (2017). www.pwc.com/us/dsa-skills
- [6] Yuri Demchenko, A. Belloum, and Tomasz Wiktorski. 2016. Edison Data Science Framework: Part 1. Data Science Competence Framework (CF-DS) Release 1. (01 2016). https://doi.org/10.5281/zenodo.167585
- [7] Carol S Dweck. 2008. *Mindset: The new psychology of success.* Random House Digital, Inc.
- [8] ACM Data Science Task Force. 2019. Computing Competencies for Undergraduate Data Science Curricula. (2019). http://www.cs.williams.edu/~andrea/DSReportInitialFull.pdf
- [9] Business-Higher Education Forum. 2016. Competency Map for the Data Science and Analytics-Enabled Graduate. (2016). www.bhef.com/news-events/events/webinar-data-science-and-analytics-dsa-enabled-graduate-competency-map
- [10] Geoffrey Fox and Wo Chang (Eds.). 2019. NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements. *NIST Big Data Public Working Group* (2019). https://doi.org/10.6028/NIST.SP.1500-3r2
- [11] Catherine Hill, Christianne Corbett, and Andresse St Rose. 2010. *Why so few? Women in science, technology, engineering, and mathematics.* ERIC.
- [12] Trudy Howles. 2007. Preliminary results of a longitudinal study of computer science student trends, behaviors and preferences. *Journal of Computing Sciences in Colleges* 22, 6 (2007), 18–27.
- [13] E.J. Jolly, P.B. Campbell, and L. Perlman. 2004. Engagement, capacity and continuity: a trilogy for student success. (2004). www.campbell-kibler.com
- [14] E. J. Jolly and P. Campbell. 2014. Ten years of engagement, capacity and continuity: Reflections on a trilogy for student success. (2014). www.campbell-kibler.com/trilogy.pdf
- [15] Sean Kandel, Andreas Paepcke, Joseph M Hellerstein, and Jeffrey Heer. 2012. Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2917–2926.
- [16] Paul H Kvam. 2000. The effect of active learning methods on student retention in engineering statistics. *The American Statistician* 54, 2 (2000), 136–140.
- [17] Jeffrey J McConnell. 1996. Active learning and its use in computer science. *ACM SIGCSE Bulletin* 28, SI (1996), 52–54.
- [18] Susan Staffin Metz. 2007. Attracting the engineers of 2020 today. *Women and minorities in science, technology, engineering, and mathematics: Upping the numbers* (2007), 184–209.
- [19] Michael Chui James Manyika Tamim Saleh Bill Wiseman Nicolaus Henke, Jacques Bughin and Guru Sethupathy. 2016. The Age of Analytics: Competing in a Data-Driven World. (2016).
- [20] Joan Peckham, Lisa L Harlow, David A Stuart, Barbara Silver, Helen Mederer, and Peter D Stephenson. 2007. Broadening participation in computing: issues and challenges. In *ACM SIGCSE Bulletin*, Vol. 39. ACM, 9–13.
- [21] Bina Ramamurthy. 2016. A Practical and Sustainable Model for Learning and Teaching Data Science. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16)*. ACM, New York, NY, USA, 169–174. https://doi.org/10.1145/2839509.2844603
- [22] Eric S. Roberts, Marina Kassianidou, and Lilly Irani. 2002. Encouraging Women in Computer Science. *SIGCSE Bull.* 34, 2 (June 2002), 84–88. https://doi.org/10.1145/543812.543837
- [23] Felesia Stukes, Hang Chen, and Terik Tidwell. 2018. Applying the Engagement, Capacity and Continuity Trilogy for Computing Undergraduates at Johnson C. Smith University. In *2018 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*. IEEE, 1–4.