Rice University

Minor in Data Science

Approved by the Faculty Senate

February 20, 2019

Proposal to the Faculty Senate for an Interdisciplinary Minor in Data Science

Devika Subramanian and Fred Oswald (co-chairs of the Curriculum Committee)
Renata Ramos and Rudy Guerra (School of Engineering representatives)
Members of the Data Science Curriculum Committee

January 30, 2019

Approved by the CUC on January 11, 2019.

1 History and Development of Data Science at Rice

Escalating its vision and commitment to data-driven knowledge-discovery, Rice University has invested roughly $45 million in a university-wide Data Science Initiative (`datascience.rice.edu`). Critically, the goal of this initiative is to develop educational offerings in data science at both the undergraduate and graduate levels. The vision paper defining the Rice Data Science Initiative [3] states that data science is relevant to all Rice students and calls for a data science program that is accessible to a substantial proportion of Rice students. Following this commitment, a Provost-appointed committee issued a report recommending that Rice establish an undergraduate minor in data science that, in addition to more technically oriented courses, includes a required course that focuses on the broader impact of the information age on our understanding of human activity, including discussion of privacy, ethics, decision-making, culture, and truth in the digital age (see [2]). Subsequently, the Provost convened the Data Science Curricular Initiatives Committee, which met once every week during the 2016-2017 academic year (August 2016 to March 2017) to craft an undergraduate data science minor at Rice. The committee gathered information from a wide range of undergraduate and graduate Rice students, existing data science programs, and data science experts tied to substantive and educational/teaching domains; reviewed and discussed detailed operational matters such as the curricular composition of the minor, identifying existing courses upon which to draw, as well as existing gaps in the curriculum. The committee issued a report in May 2017 presenting the design of an innovative curricular structure and process for training a broad spectrum of students in data science [6]. Building on the committee's report, a proposal for the Data Science Minor was submitted to the CUC, and was approved by the CUC on January 11, 2019. This proposal to the Faculty Senate incorporates all the suggestions made by the CUC committee.

Figure 1: Two contemporary views of data science from [5]

2 The Data Science Minor

What is data science, and who is a data scientist? Are data scientists simply statisticians who specialize in data processing, analytics, and computing with data? Are they computer scientists who develop algorithms and systems for handling big data? Furthermore, where do the key topics of data visualization, data wrangling, data privacy and ethics fit in? What is the role of domain expertise in areas such as physics, urban planning or art history, where deep expertise informs data science? The committee considered these questions (and more) when developing curricular goals for the undergraduate data science minor. Data science is a relatively new field, and there are many competing visions for related curricula. The committee surveyed the field and its curricular offerings widely, including the report on undergraduate data science issued by the National Academies [4] in 2018. We found general consensus that teaching data science is inherently interdisciplinary. Statistics, applied mathematics, and computer science are core components, and what varies in emphasis or is often ignored, are the roles and relative importance of domain-relevant applications, data visualization, and data ethics in the data science enterprise.

The Venn diagrams in Figure 1 illustrate two contemporary views of data science [5]. The view on the left does not include interpretation, communication, and data visualization. The view on the right fails to capture the role of domain expertise in data science, focusing instead on tools and algorithms for data analysis. Furthermore, *neither* view emphasizes what our committee viewed as a critical component of data science: understanding issues of data provenance, privacy, and ethics that place data science in the broader context of society, culture, and decision-making.

Our proposed data science minor aims to inculcate the following critical thinking skills and practical capabilities:

- formulate questions in a discipline that can be answered with data;

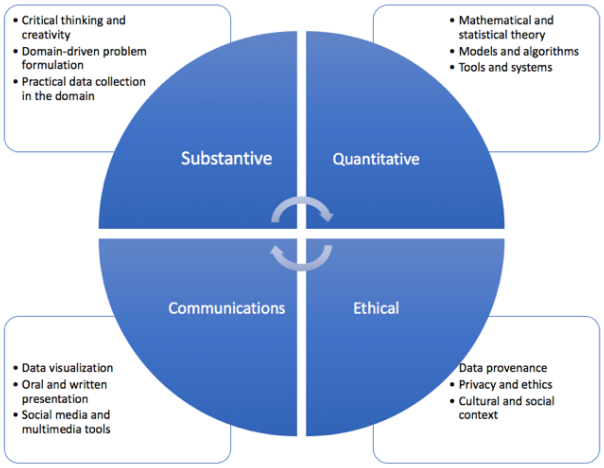


Figure 2: Foundations of the data science minor at Rice

- use tools and algorithms from statistics, applied mathematics, and computer science for analyses;
- visualize, interpret, and explain results cogently, accurately, and persuasively;
- understand the underlying social, political, and ethical contexts that are inevitably tied to data-driven decision-making.

For maximal impact on Rice undergraduates, we have chosen to design a minor that reflects the view of data science expressed in Figure 2. It meets critical criteria of (a) technical heft, (b) intellectual breadth, (c) wide accessibility to students across all disciplines at Rice, and (d) integration with the broader data science research initiative at Rice. Note that the new minor not only places Rice ahead of its peers because of the considerable investment of expertise, time, and resources behind it (e.g., the minor is not merely a combination of existing statistics, mathematics, and computer science courses), it is also unique because the minor does not reflect any one department at the university. Instead, it is to be valued and promoted across all departments and schools as a signature university-level minor.

As shown in Figure 2, the data science minor reflects a domain-grounded view and emphasizes **doing** data science. It is best explained in terms of four core competencies the minor seeks to develop in our students – a set the committee quickly converged on at its very first meeting:

- **Substantive Foundation:** Focusing on high-level critical thinking and creativity to formulate discipline-related questions that can be answered with data. Deep ground-

ing in a specific major to frame sophisticated and significant questions regarding how to identify, gather, analyze, and interpret relevant data.

- **Quantitative Foundation:** Focusing on honing mathematical, statistical, and computational thinking and skills that allow one to translate ideas about modeling and analyses into their actual execution. Thus, students completing the data science minor will be able to **do** data science. They can gather and organize a mass of unstructured data [4] into an analyzable form and execute analyses using state-of-the-art techniques and tools. To this end, we have identified three core quantitative areas: mathematical foundations (probability, statistics, linear algebra, optimization, discrete structures), models (linear, logistic, ridge, and sparse regression; Bayes; generalized linear models, non-linear models), algorithms (unsupervised and supervised machine learning), and tools and systems (statistical computing, programming, databases, big data management).

- **Communications Foundation:** Focusing on heightening data visualization, interpretation, and communication skills to convey results of analyses cogently, accurately, and persuasively to a wide range of audiences (e.g., students, educators, and researchers across disciplines; corporations and non-profits; federal, state, and local policymakers; granting agencies and foundations). An effective data scientist absolutely needs to be well versed in using visualization tools to support the expression of ideas in oral and written form, using different media: slides, posters, and video presentations.

- **Ethics Foundation:** Focusing on improving students' understanding of data provenance, privacy, and ethics in gathering data, as well understanding and knowing how to ensure reproducibility of results, with an awareness and sensitivity to the cultural and social context in which data-driven decisions are inevitably made. News stories from the past several years reinforce the critical need for ethics as the indispensible backbone of data science training.

3 Comparable Minors at Peer Institutions

In what might be viewed as an academic gold rush, US universities have been racing to establish data science degree and certificate programs in the last few years. Our survey of available programs [1] show that most of the programs are at the graduate level (445 masters, 88 certificate and 16 doctoral programs in the US alone), though there are 37 undergraduate major and 5 undergraduate minor programs. The fact that there are so few minor programs may reflect the inherent difficulty of spanning so many distinct interdisciplinary areas of expertise within the typical 6-8 course format of a minor curriculum. The curricula of data science minors at our peer institutions are summarized in Table 1.

School	Housed	Year	#C	#Q	#Cm	#E	Cap	URL
MIT	Statistics	2016	6	5	0	0	1	https://stat.mit.edu/academics/minor-in-statistics/
Georgia Tech	School of Computing	2016	5	4	1	0	0	https://www.cc.gatech.edu/content/minor-computational-data-analysis
Cornell	School of Information Sciences	2016	4	2	1	1	0	http://infosci.cornell.edu/academics/undergraduate/undergraduate-concentrations/concentrations/data-science
Berkeley	Division of Data Sciences	2018?	5	4	0	0	0	https://data.berkeley.edu/education/faqs
UPenn	School of Engineering	2016	6	5	0	1	0	https://catalog.upenn.edu/undergraduate/programs/data-science-minor/

Table 1: A summary table of minors in data science at peer institutions. The table shows where the minors are administratively housed, and the curricular composition in terms of the number of quantitative (#Q), communications (#Cm), ethics (#E), and capstone (Cap) courses (where #C above is the total # of courses). A more detailed description of these programs is in Appendix D.

4 Distinction from Other Minors at Rice

There are no competing minors at Rice that offer a breadth of experience similar to what the data science minor offers. The largest overlap is with the departmental minor in Statistics, offered by the department of Statistics, and the interdisciplinary Minor in Financial Computation and Modeling, offered through a collaboration of the departments of Statistics and Economics. These minors are discussed below.

Minor in Statistics

This minor https://ga.rice.edu/programs-study/departments-programs/engineering/statistics/statistics-minor/ is designed for students from all backgrounds in the university. It has two tracks to accommodate differences in mathematical preparation. Track A is designed for mathematically sophisticated students, and Track B is designed for students seeking a working knowledge of statistics and its applications. The statistics minor requires a minimum of six courses (18 credit hours). The content of the required courses in the statistics minor overlaps with the quantitative core sequence of the data science minor. The statistics minor covers the quantitative core of statistics in far greater depth with 5-6 courses (not including a capstone), whereas the data science minor achieves the coverage in a more applied setting with a tightly coordinated three-course quantitative sequence. The data science minor is unique in terms of its core requirements of communication, ethics, and a capstone experience.

Minor in Financial Computation and Modeling

This minor https://ga.rice.edu/programs-study/departments-programs/engineering/financial-computation-modeling/ focuses on strategies and computational technologies used in the financial industry. It consists of six curated courses from ECON and STAT. It could be viewed as a specialized data science minor with a single focus — financial data analysis. In contrast, the elective structure and capstone experience of the data science minor allows students with any disciplinary focus to acquire broad knowledge and training in data science.

Summary

The data science minor is distinct from other existing minors at Rice. It is a first of its kind, given its broad curricular composition involving multiple departments in multiple schools, and given its mandate to accommodate students from across the university. Its core curricular offerings come from traditional departments as well as a university-level program (PWC). The minor does not duplicate any existing departmental degree programs.

5 Governance

The minor will be initially overseen by the Provost-appointed committee that designed the minor curriculum and the Dean of Engineering-appointed members in-charge of implementation of the minor. The Provost-appointed committee and the Dean of Engineering (DoE) appointed members are listed below.

- David Alexander, Physics and Astronomy
- Rudy Guerra, Statistics (DoE appointment)
- Matthias Heinkenschloss, Computational and Applied Mathematics
- Chris Jermaine, Computer Science
- Luay Nakhleh, Computer Science
- Barbara Ostdiek, Business
- Kirsten Ostherr, English
- Fred Oswald (co-chair), Psychological Sciences
- Renata Ramos, Bioengineering (DoE appointment)
- Devika Subramanian (co-chair), Computer Science
- Marina Vannucci, Statistics
- Ashok Veeraraghavan, Electrical and Computer Engineering
- Jennifer Wilson, Director of the Program for Writing and Communication

A formal faculty steering committee will be appointed by the Dean of Engineering and the Provost after the approval of the minor. This faculty steering committee will meet at least once each semester to review course offerings and enrollments. Members of the steering committee will serve as minor advisors, and sign minor declaration forms, advise students on course sequencing, and approve degree audits for graduation. Future committees will be appointed by the Provost and the Dean of the School of Engineering (with rotations and staggered appointments to retain continuity in decision-making). The composition of the steering committee will reflect the diversity and interdisciplinary nature of the training program and will draw upon faculty in Engineering, Natural Science, Social Sciences, and the Humanities.

The core course offerings and prerequisites of the minor span four departments in the School of Engineering (Computer Science, Electrical and Computer Engineering, Computational and Applied Mathematics and Statistics), the School of Social Sciences, the School

of Humanities, and the Program in Writing and Communication. The Dean of the School of Engineering, the Dean of the School of Social Sciences, the Dean of Humanities, and the Director of the Program in Writing and Communication will be jointly responsible for fulfilling Rice University's ongoing commitment to resources and pertinent staffing required by the minor (we have letters of support from all the Deans). The program will be administered by Chris Jermaine of the Computer Science Department). Presently, we have an Acting Executive Director, Rudy Guerra (Statistics), who will serve in the first year of the program. Changes to the minor's core curriculum or its elective offerings will not be made without appropriate consultation with all participating departments/schools in the University. The Dean of Engineering, with the assistance of the Executive Director and steering committee, will be responsible for adjudicating any issues associated with such changes and for communicating them, as appropriate, to the Faculty Senate.

6 Requirements

To earn a minor in data science, students are required to take a minimum of 9 courses (27-29 credit hours, depending on course selection). They must complete a minimum of 6 courses (18 credit hours) taken at the 300 level or above. These courses include 5 courses (15-16 credit hours based on course selection) to satisfy core requirements and a capstone project (3 credit hours). The core courses are: a sequence of three quantitative courses (DSCI 301/STAT 315/STAT 310, DSCI 302/COMP 330 and DSCI 303), a three-credit course on communication and visualization (DSCI 304), a three-credit course on ethics (DSCI 305). The capstone project course is a three-credit course (DSCI 400). Students must also complete 3 courses (9-10 credit hours) to satisfy prerequisite requirements detailed below.

Prerequisites

- Mathematics: Two terms of calculus are required, for a total of 6 credits. This prerequisite can be satisfied with (MATH 101/MATH 105/MATH 211) and (MATH 102/MATH 106/MATH 212). Students may substitute a higher level MATH course for the MATH 101/Math 105 and MATH 102/MATH 106 prerequisites. Linear algebra is highly recommended via CAAM 334, 335 or MATH 355, which can be taken concurrently with DSCI 301.
- Programming: One term of programming is required, for a total of 4 credits. This prerequisite can be satisfied with COMP 140 or CAAM 210. Python proficiency is assumed in DSCI 302. Students will not be able to register for DSCI 302 unless they have COMP 140 in their academic history, or in exceptional circumstances by permission of the DSCI 302 instructor.

The courses listed above satisfy the requirements for this minor. In certain instances, courses not on this official list may be substituted upon approval of the minor's academic

advisor, or where applicable, the Program Director. (Course substitutions must be formally applied and entered into Degree Works by the minor's Official Certifier). Students and their academic advisors should identify and clearly document the courses to be taken. In certain situations, the DSCI Official Certifier may approve various and specific course substitutions. For example, COMP540 can be substituted for DSCI 303.

Proposed General Announcement Text (currently under review with the Registrar)

Please see attached document.

7 Resources for the minor

We anticipate about 100-120 undergraduate minors at any given time. This estimate is based on a student poll conducted by Rudy Guerra in STAT 315/DSCI 301, a class of 150 students he taught in Fall 2018. This demand will require ongoing resource commitments from central administration, the Deans of Engineering, Social Sciences, Humanities, the Program in Writing and Communication, as well as the relevant departments listed below. We have appraised all the relevant deans of the scale of the commitments needed and have obtained support letters from them.

- DSCI 301/STAT 315: Instructor is Rudy Guerra, Department of Statistics, course offered Fall and Spring.
- DSCI 302: Instructor is Risa Myers, Computer Science, course offered Spring only.
- DSCI 303: Instructor is Reinhard Henkel, ECE course offered Fall only.
- DSCI 304: Instructor is Anneli Joplin, Program in Writing and Communication, course offered Fall and Spring.
- DSCI 305: Instructor is Elizabeth Roberto, Sociology, course offered Spring only.
- DSCI 400: Lead instructor is Genevera Allen, Statistics, course offered Fall and Spring.

We note here that the instructors named above will teach the inaugural versions of these courses and that the Deans of the respective schools, working with School of Engineering, will ensure continuity in course offerings and consistency with the course currricula laid out in this document. Support letters from the Deans make the instructional commitment.

8 New Course Information

Quantitative Core Sequence: DSCI 301, DSCI 302, DSCI 303

Although a serviceable data science quantitative core curriculum could potentially be crafted by selecting existing courses from statistics, computer science, and computational and applied mathematics, the prerequisite chains that arise with these courses, make the data science minor completely inaccessible to students outside of these three majors. Further, these disciplinary courses go deep into topics that are somewhat tangential to the data science enterprise and lack the cross-disciplinary cohesion needed to make a standalone data science minor. By contrast, the set of three quantitative courses described below consists of a carefully curated list of topics from statistics, computer science, and applied mathematics that are essential to build data acumen in our students and to give them the ability to **do** data science.

DSCI 301/STAT 315: Probability and Statistics for data science

Catalog description

An introduction to mathematical statistics and computation for applications to data science. Topics include probability, random variables, expectation, sampling distributions, estimation, confidence intervals, hypothesis testing and regression, and basic linear algebra. A weekly lab will cover the statistical package R and data projects. DSCI 301 is a foundational entry to statistics for students interested in data science (4 credits). Prerequisites: MATH 102. Recommended: MATH 212.

Work and Assessment

DSC 301 is a calculus-based introduction to probability and statistics for data science. Basic mathematical theory of core topics including randomness, descriptive statistics, random variables, probability distributions, inference (estimation and testing), regression and basic linear algebra. Concurrent data projects using R will be used in both standard (comparing two populations with a *t-test*) and applied contexts (regression variable selection associated with analyzing genes in a microarray study; conditional probability and Bayesian inference in drug testing).

- Weekly assignments on mathematical statistics (2-3 hours per week). Assessment: undergraduate grading under a typical homework rubric.
- Biweekly data projects supported by weekly lab meetings (3-4 hours per week). Assessment: Students submit statistical reports to be graded under a typical rubric.
- Midterm and final exams. Assessment: in-class exams covering mathematical theory, statistical applications and R. Assessment: typical exam rubric.

Student learning outcomes

- Understand the role of mathematics in statistics.
- Perform exploratory data analysis: summarize data, prepare and interpret visualizations.
- Formulate statistical questions.
- Define, perform and interpret appropriate statistical analyses.
- Write statistical reports.

Frequency of offering

The course will be designed and taught by Professor Rudy Guerra of the Statistics department. It will be offered every semester by Rice faculty members from the Statistics department. The core content of the course, and the lab plus lecture format, will remain consistent with the catalog description above. STAT 315 has been offered in Fall 2018 (150 students) and is being offered in Spring 2019 (90 students). Note that some students taking STAT 310, a substitute for DCSI 301, also have plans to complete the Data Science minor.

DSCI 302: Tools and models for data science

Catalog description

This course introduces key concepts in data management, preparation, and modeling and provides students with hands-on experience in performing these tasks using modern tools, including relational databases and distributed file systems. Models covered include linear and logistic regression and gradient descent. (3 credits) Prerequisites: DSCI 301, COMP 140/instructor permission. COMP140 highly recommended.

Work and Assessment

This second course in the data science quantitative core sequence focuses on systems to handle data and tools for building models. The first half of the course focuses on relational databases/SQL, and distributed file systems/Spark. The second half introduces Scipy and Numpy in Python for building models with labeled and unlabeled data. Concepts of experimental design, metrics for model evaluation, cross-validation, overfitting and regularization, data-preprocessing and feature engineering, outliers, and data quality will be covered. Theoretical concepts will be instantiated in the context of substantial implementation projects so students learn to do data science with the tools introduced in this class.

- 4 short exercises to support concepts introduced in class. Assessment: undergraduate grading under a typical homework rubric.

- 5 in-depth assignments where students implement models to develop key skills. Assessment: Students submit reports and code to be graded under a typical project rubric.
- Hands-on labs in which students install all necessary software and get systems running for use in projects.

Student learning outcomes

After taking this course, students will

- understand steps and tools that can be used to manage data
- prepare data for modeling using relational databases and Spark
- implement models, including linear and logistic regression and gradient descent
- evaluate model performance
- gain experience and confidence to implement more complex models on unique datasets

Frequency of offering

The course is already being offered in Spring 2019 (current enrollment 29) and is designed and taught by Risa Myers, Lecturer in Data Science in the Computer Science department. It will be taught every Spring term thereafter by a faculty member from the Computer Science department. The core content of the course will remain consistent with the catalog description above, independent of the instructor.

DSCI 303: Machine learning for data science

Catalog description

This course is an introduction to concepts, methods, best practices, and theoretical foundations of machine learning. Topics covered include regression, classification, kernels, dimensionality reduction, clustering, decision trees, ensemble learning, regularization, learning theory, and neural networks. 3 credits. prerequisites: DSC 301, DSCI 302. CAAM 334 is highly recommended.

Work and Assessment

- Homework every week or every other week to reinforce concepts introduced in class. Assessment: undergraduate grading under a typical homework rubric. Homework includes programming projects in which students build predictive models and perform exploratory analysis on datasets.
- Two exams: a midterm and a final which will be assessed on typical exam metrics.

Student learning outcomes

Students will learn how and when to apply machine learning algorithms, their comparative strengths and weaknesses, how to critically evaluate their performance, and their theoretical foundations. Students completing this course should be able to

- apply machine learning methods to build predictive models or perform exploratory analysis
- properly tune, select, and validate machine learning models
- interpret their results
- understand their limits

Frequency of offering

The course will be designed and taught by Reinhard Henkel, Assistant Professor in the ECE department. It will be taught every Fall term thereafter by a faculty member from the ECE department. The core content of the course will remain consistent with the catalog description above, independent of the instructor.

DSCI 304: Introduction to effective data visualization

This course will offer foundational instruction in data visualization for data science. The course will introduce students to visualization with an evidence-based foundation for best practices that is grounded in communication, statistics, psychology, and graphic design. The course will also provide instruction and practice in identifying meaningful patterns in data, critiquing visualization form and style, and presenting data to a variety of audiences. A 14-week content outline is in Appendix B.

Catalog description

This course teaches fundamental data visualization skills for data science. Students will learn how to create data visualizations in Python or R, how to design effective visualizations that account for visual perception, and how to explain and present data to technical and non-technical audiences. (3 credits) Lecture Lab course. Prerequisites: DSCI 301, DSCI 302.

Work and Assessment

Students will complete weekly data visualization assignments, a practical midterm, and a multi-week data visualization project and associated presentation. These tasks will be graded on a point system according to an appropriate rubric created by the instructor. The contribution of each component to the final grade is listed below:

Student learning outcomes

After completing the visualization lab, students will be able to:

- identify meaningful patterns and relationships in different types of data
- critique visualization form and style and articulate evidence-based revisions
- design and create compelling visualizations for written, oral, video, and online presentation
- explain data and analyses in a clear way to technical and non-technical audiences
- construct a data-driven narrative using real-world data

Frequency of offering

The initial offering of the course will be in Fall 2019 and will be designed and taught by Anneli Joplin, Instructor of Visual Communication and Design in the Program for Writing and Communication (PWC). It will be taught every Fall and Spring term thereafter by the NTT Instructor of Visual Communication and Design in the PWC. Based on demand for the course, additional sections will be offered in the Fall term with availability of instructional resources. The core content of the course will remain consistent with the catalog description above, independent of the instructor.

DSCI 305: Ethics

Catalog description

DSCI 305: Data, Ethics and Society. An examination of the ethical implications and societal impacts of choices made by data science professionals. The seminar course will provide practical guidance on evaluating ethical concerns, identifying the potential for harm, and applying best practices to protect privacy, design responsible algorithms, and increase the societal benefit of data science research. (3 credits)

Work and Assessment

- Active participation in class discussions and weekly reading responses (20%).
- Three short writing assignments designed to deepen students' engagement with ethical issues in data science (30%).
- A semester-long team project that: (1) critically evaluates the ethics and potential societal impacts of a real-world data science project; and (2) applies the techniques and best practices discussed in the course (and the team's cross-disciplinary expertise

and quantitative skills) to design a strategy for addressing the issues identified in part 1.

Student learning outcomes

- Understand the relationship between data, ethics, and society.
- Develop and apply ethical reasoning relevant to ethical, cultural, and social issues when gathering, processing, and analyzing data.
- Explore one's own social and ethical commitments as a future data scientist.

Frequency of offering

This course is being offered in Spring 2019 (current enrollment 9) and is designed and taught by Assistant Professor Elizabeth Roberto of the Sociology department. Subsequent offerings will occur every Spring term and will be taught by a Rice faculty member drawn from the School of Social Sciences and the School of Humanities. Although specific topics addressed in the course may vary according to the instructor, the core content will remain consistent with the stated catalog description. We expect the course content and delivery to evolve over time and to be led by the faculty member hired into the Data Ethics position allocated to the School of Humanities.

DSCI 400: Capstone Project

The capstone course is offered through the Rice Data-to-Knowledge (D2K) Lab (d2k.rice.edu) headed by Professor Genevera Allen of the Statistics Department.

Catalog description

DSCI 400: Data Science Projects. In this project-based course, student teams will complete semester-long data science research or analysis projects selected from a variety of disciplines and industries. Students will also learn best practices in data science. Prerequisites: DSCI 301, DSCI 302, DSCI 303, DSCI 304. (3 credits)

Work and Assessment

In this project-based course, student teams will complete a semester-long data science research or analysis project selected from a variety of disciplines and industries. Each student team will be mentored by an NTT co-instructor. There are three project types that will be offered.

- Option 1 – Client-sponsored projects at the D2K: This option will have students working on client-sponsored projects at the D2K. The D2K lab will select students

for these projects based on background and technical skills, as well as fit for the project. The selection process will be similar to that used by the OEDK to match students to client-sponsored projects.

- Option 2 – Curated dataset projects: : For students who cannot be matched to a D2K client-sponsored project, teams will be formed around carefully curated data sets known to be interesting and non-trivial to analyze. These data sets are derived from various open source data sets available (data.gov, Kinder Institute repositories, Kaggle competitions, etc.). These projects will be just as challenging as projects in Option 1, and perhaps more so, but they lack a specific paying client.
- Option 3 – Student-initiated, faculty-supported, disciplinary capstones: This path is for students with interest and initiative in specific data science problems that do not fit within Option 1 or Option 2. These students, either individually, or in a team, will need to identify a Rice faculty sponsor (or sponsors), and work with a D2K Lab instructor, and with the advice of these parties, identify a data set and a well formulated problem for analysis. This option offers students with opportunities for independence and flexibility. For example, for students who are already working on a disciplinary capstone, the data science capstone project can explore the uses of data science in analysis, communication, visualization and ethics. The project will be evaluated using the same rubrics as projects in the other options by a D2K instructor with input from the faculty sponsor.

The following apply to all students in all three options.

- Students will attend a weekly lecture during which they will learn best practices along each step of the data analysis pipeline: data wrangling, exploratory data analysis, modeling and validation.
- Student teams will be required to meet with their mentors once a week throughout the semester and set monthly project milestones.
- Students will document their milestones in terms of interim reports and short presentations.
- At the end of the semester, teams will produce a comprehensive project report and give an oral presentation of their work to a faculty assessment committee.
- Assessment: Teams will be assessed based on their effectiveness at meeting project objectives, clarity and meaningfulness of findings, creativity, adherence to computational and scientific reproducibility standards, and oral and written communication.
- Assessment: Individual students in a team will be evaluated based on lecture and meeting attendance, oral and written communication, and using research-based team grading rubrics (e.g., CATME `http://info.catme.org`).

Student learning outcomes

- Students will engage in experiential learning in which they learn best practices in data science by working on real-world data science projects.
- Students will learn how to wrangle, explore, and model data, as well as to how to validate their data analysis findings.
- Students will learn team-work, leadership, and communication skills.

Frequency of offering

The course will be coordinated by Professor Genevera Allen of the Statistics department. The first offering of the course is in Fall 2019. The core content of the course will remain consistent with the catalog description above, independent of the instructor.

9 Student Learning Outcomes

Upon completion of the data science minor, students will be able to

- formulate questions in a domain that can be answered with data;
- use tools and algorithms from statistics, applied mathematics, and computer science for analyses;
- visualize, interpret, and explain results cogently, accurately, and persuasively;
- understand the underlying social, political, and ethical contexts that are importantly and inevitably tied to data-driven decision-making.

A curriculum map and assessment plan are included in Appendix D.

To establish a baseline assessment, a subset of the final reports written by students who complete capstone projects will be examined by the Executive Director of the data science program. This output, will be compared in future years to the reports of students who declare the minor and complete the capstone course in earlier years. Each year, the Director and/or members of the advisory board of the program will examine work from the core courses and final reports from the capstones to see whether students are achieving the methodological and technical skills the program seeks to achieve. Using the Student Learning Outcomes Measure Rubric in Appendix D, the director and steering committee will keep numerical scores of skill level in the different areas, and compare them over time. The Director will also look at exit surveys from students graduating with the minor degree.

10 Appendix A: Faculty

The Minor in Data Science will be guided by the steering committee, which will convene at least once a year to assess the status of the minor, including the number of students enrolled, course listings, effectiveness of course offerings, and affiliated faculty. An Executive Director will administer the minor on a daily basis. The director will also update the steering committee, the relevant deans, and affiliated faculty of any programmatic or curricular developments and needs.

Program Administration

Chris Jermaine will head the administration of the Data Science Minor housed in the School of Engineering. Rudy Guerra of the Statistics department will serve as Acting Executive Director for the first year of the minor.

Faculty Steering Committee

The initial committee is composed of the Provost-appointed members of the curriculum committee for data science and the Dean of Engineering-appointed members for coordinating the implementation of the minor. Upon approval of the minor, a permanent committee will be established by the Dean of Engineering and the Provost.

- David Alexander, Physics and Astronomy
- Rudy Guerra, Statistics (DoE appointment)
- Matthias Heinkenschloss, Computational and Applied Mathematics
- Chris Jermaine, Computer Science
- Luay Nakhleh, Computer Science
- Barbara Ostdiek, Business
- Kirsten Ostherr, English
- Fred Oswald (co-chair), Psychology
- Renata Ramos, Bioengineering (DoE appointment)
- Devika Subramanian (co-chair), Computer Science
- Marina Vannucci, Statistics
- Ashok Veeraraghavan, Electrical and Computer Engineering
- Jennifer Wilson, Director of the Program for Writing and Communication

Faculty teaching core courses

As of January 2019, this list includes

- Rudy Guerra, Professor of Statistics (DSCI 301 instructor)
- Risa Myers, Lecturer in Computer Science (DSCI 302 instructor)
- Reinhard Henkel, Assistant Professor of Electrical and Computer Engineering (DSCI 303 instructor)
- Anneli Joplin, Center for Academic and Professional Communication, (DSCI 304 instructor)
- Elizabeth Roberto, Assistant Professor of Sociology (DSCI 305 instructor)
- Genevera Allen, Associate Professor of Statistics (DSCI 400 instructor)

11 Appendix B: Detailed 14-week Syllabi for Core Courses

DSCI 301/STAT 315: Statistics for data science

1. Course overview; introduction to statistical inference. R statistical software. Summary stats, visualization.
2. Axioms of Probability. Equally likely outcomes. Conditional probability. Bayes theorem; independence.
3. Random variables. Binomial. Poisson.
4. Continuous random variables and probability densities. Normal. Exponential.
5. Statistical expectation. Mean, variance, SD, higher-order moments. Moment generating functions (mgf)
6. Vectors, matrices and matrix algebra.
7. Covariance and correlation. Bivariate normal (covariance matrix).
8. Multivariate data.
9. Statistical inference. Heuristics of testing and estimation. Random samples. Statistics. Sampling distribution.
10. Bootstrap standard errors and bootstrap confidence intervals.
11. Large sample inference. Law of large numbers, Central limit theorem. Delta method.

12. Large sample testing and confidence intervals. Normal and binomial.
13. 1-and 2-sample problems: testing and estimation. Maximum likelihood estimation.
14. Simple linear regression. Linear systems and linear least squares.
15. Multivariable linear regression. Singular value decomposition.

DSCI 302: Tools and models for data science

Here is a week-by-week list of topics to be covered in this class.

1. Course overview; introduction to machine learning and data science. Relational databases: history, structure, motivation; the relational model; relational calculus
2. Relational algebra; Declarative SQL: queries and subqueries; aggregation
3. Database set operations, joins;
4. Database design and modeling, normalization;
5. Imperative SQL
6. Indexes; data definition language; MapReduce;
7. distributed file system (HDFS); Spark
8. More Spark; Introduction to modeling;
9. least squares, MLE, gradient descent; stochastic gradient descent;
10. Python: numpy, scipy, vectorization; common data formats, data types, encoding; strings: regular expressions, tokenization, stemming, stop words;
11. Supervised learning; unsupervised learning; experimental design and evaluation: confusion matrices; AUC, F-score, mean squared error;
12. training/validation/testing; cross-validation, overfitting; linear regression; logistic regression;
13. Data preprocessing and feature engineering: imputation; outliers; normalization; data quality; common feature transformations;
14. Regularization; bias/variance; challenges with big data, rare classes, etc.

DSCI 303: Machine learning for data science

Here is a week-by-week list of the topics to be covered in this class. Required linear algebra background will be introduced ahead of each core topic.

1. Overview, linear regression
2. Ridge regression and sparse regression (L2 and L1 regularization; elastic net)
3. Logistic regression and naive Bayes
4. Generalized linear models (unifies various statistical models including linear and logistic regression)
5. Nearest neighbor classifiers
6. Support vector machines and kernels
7. Model validation (cross validation, train/test splits)
8. Learning theory (bias/variance tradeoff; overfitting; generalization error)
9. Decision trees; strong and weak learning - boosting; random forests
10. Principal component analysis
11. Factor analysis (e.g., matrix factorization)
12. Clustering (k-means, spectral clustering)
13. Clustering, Mixture of Gaussians and EM Algorithm

Graphical models, active learning, ranking problems, deep learning and reinforcement learning are other topics which can be included to support the semester-long competition.

DSCI 304: Introduction to effective data visualization

The communication core course DSCI 304 teaches students how to effectively visualize, explain, and present data. This applied data visualization laboratory course is composed of seven modules, each of which will be accompanied by an assignment. Students will also draw on concepts from all seven modules to complete a final visualization project. The modules are

- Introduction to data visualization
    1. Introduction to exploratory data visualization
    2. Graphical perception theory and implications for visualization type

3. Evaluating patterns and trends in data
4. Recognizing how visualizations can distort data
5. Asking questions of data, and using data to answer questions
6. Introduction to modern visualization types

- Data visualization design

1. Accounting for context and audience
2. Balancing function and form
3. Using preattentive contrast to establish a visual hierarchy
4. Revising graphics to account for visual perception
5. Visual emphasis of important data and comparisons
6. Ethical considerations for visualization design
7. How to control the appearance of visualizations in Python / R

- Writing about data

1. Basic guidelines for technical writing (e.g. leading with claims, avoiding jargon)
2. Integrating text in data visualizations – data labels, titles, annotations, etc.
3. Writing interpretive visualization captions and / or titles
4. Establishing a visual narrative: symbols with meaning, visual legends, etc.
5. Formatting visualizations for print

- Storytelling through design

1. Design principles and applications to DS
2. Color theory and guidelines for use of color in DS
3. Cognitive processing of visual information and implications for design
4. Applying pattern recognition tendencies (Gestalt principles) to reinforce meaning
5. Formatting visualizations for posters (static storytelling)
6. Polishing visualizations using Adobe Illustrator

- Oral presentation of data

1. Basic guidelines for oral presentations (narrative structure, simple slides, etc.)
2. Strategies for discussing and explaining data (particular $\rightarrow$ general $\rightarrow$ particular)
3. Using animation to tell your data story

4. Ways to extract data from published figures (with citation of course)
5. Formatting visualizations for presentation

- Advanced visualization topics (1-2 topics per course selected based on student interests)

1. Visualizations of location data and maps (ArcGIS)
2. Data graphics and gifs for social media
3. Visualizing multidimensional data (e.g., 3D visualization in Blender)
4. Network and cluster visualizations (e.g., Gephi)
5. Information dashboard design (Tableau or PowerBI)
6. Creating an animated data story (Adobe After Effects)
7. Useful public data sets

- Interactive visualizations

1. Fundamentals of interaction design
2. Producing online, interactive visualizations using the D3.js library
3. Arranging visualizations to compose a cohesive story
4. Communicating data to a non-technical audience

Appendix C: Comparable Minor Curricula at Peer Institutions

MIT: Minor in Statistics and Data Science

MIT's Minor in Statistics and Data Science is available to MIT undergraduates from any major. It is housed in the MIT Statistics department and the MIT Institute for Data, Systems and Society. Through **six required** subjects, the Minor in Statistics and Data Science focuses on providing students with a working knowledge base in statistics, probability, and computation, along with an ability to perform data analysis. A minimum of four subjects taken for the statistics and data science minor cannot also count toward a major or another minor. The program has five components and students have to take one to two courses from each component.

- Foundation (select one)
  - 2.087 Engineering Mathematics: Linear Algebra and ODEs
  - 6.01 Introduction to EECS I

  - 6.0001 Introduction to Computer Science Programming in Python, and 6.0002 Introduction to Computational Thinking and Data Science
  - 18.03 Differential Equations
  - 18.06 Linear Algebra
- Statistics 1 (select one)
  - 1.010 Uncertainty in Engineering
  - 6.041A Introduction to Probability I and 6.041B Introduction to Probability II
  - 14.30 Introduction to Statistical Methods in Economics
  - 18.600 Probability and Random Variables
- Statistics 2 (select one)
  - 14.32 Econometrics
  - 15.075[J] Statistical Thinking and Data Analysis
  - 18.650 Statistics for Applications
- Computation and Data Analysis (select two)
  - 1.00 Engineering Computation and Data Science
  - 2.086 Numerical Computation for Mechanical Engineers
  - 6.008 Introduction to Inference
  - 6.036 Introduction to Machine Learning
  - 6.802[J] Foundations of Computational and Systems Biology
  - 6.819 Advances in Computer Vision
  - 9.07 Statistics for Brain and Cognitive Science
  - 14.31 Data Analysis for Social Scientists
  - 14.36 Advanced Econometrics
  - 15.053 Optimization Methods in Business Analytics
  - 15.0791 Introduction to Applied Probability
  - 16.09 Statistics and Probability
  - 18.065 Matrix Methods in Data Analysis, Signal Processing, and Machine Learning
  - 18.642 Topics in Mathematics with Applications in Finance
- Capstone (required)
  - IDS.012 Statistics, Computation and Applications

Georgia Tech: Minor in Computational Data Analysis

Housed in the College of Computing, Georgia Tech offers a minor in Computational Data Analysis for students in other disciplines who are looking to combine their area of study with the mathematical and statistical background to develop and apply data analysis techniques to real world datasets. The total credit hours to receive the minor is 15.

General requirements

- CS 1331 (Introduction to Object-Oriented Programming) must be completed with an A or B before applying for the Minor in Computational Data Analysis.
- In order to be accepted into the CS minor program, you must have a minimum of 48 credit hours remaining (not including CS 1331 and required minor coursework) in your major degree requirements, as seat availability in CS classes is limited.
- All courses must be completed with a letter-grade of C or better
- 9 hours must be 3000/4000 level.
- Course prerequisites are not a part of the minor; it is the student's responsibility to account for all required prerequisites understanding that they are also subject to change.
- If the student's major requires any of the minor courses listed below, the student should communicate with the minor advisor for course substitutions.
- This minor is not available to majors in Computer Science, Computational Media, or Industrial Engineering.
- CS 1331 and MATH through Multivariable Calculus (MATH 2401 or 2551), i.e. must be taken but not included in the 15 hours.

Required courses

- CX 4240 Introduction to Computing for Data Analysis, 3
- CX 4242 Data and Visual Analytics, 3
- Introduction to Probability & Statistics (pick one)
  - MATH 3215 Introduction to Probability & Statistics, 3
  - MATH 3225 Honors Probability and Statistics, 3
  - ECE 3077 Probability and Statistics for ECE, 3
  - ISYE 2027 Probability with Applications, 3

- Computational Methods (pick one)
  - CX 4010 Computational Problem Solving for Scientists and Engineers, 3
  - CS 4400 Introduction to Database Systems, 3
  - CS 4460 Introduction to Information Visualization, 3
- Computational Data Analysis Elective (pick one)
  - BIOL 4150 Genomics & Applied Bioinformatics, 3
  - CEE 3010 Geomatics, 3
  - CS 3630 Introduction to Perception and Robotics, 3
  - CS 4400 Introduction to Database Systems, 3
  - CS 4460 Introduction to Information Visualization, 3
  - CS 4476 Introduction to Computer Vision, 3
  - CX 4010 Computational Problem Solving for Scientists and Engineers, 3
  - EAS 4430 Remote Sensing and Data Analysis, 3
  - EAS 4480 Environmental Data Analysis, 3
  - ECE 4270 Fundamentals of Digital Signal Processing, 3
  - ECE 4560 Intro to Automation and Robotics, 3
  - ECE 4580 Computational Computer Vision, 3
  - ECE 4823 Game Theory and Multiagent Systems, 3
  - ISYE 4311 Capital Investment Analysis, 3
  - ISYE 3232 Stochastic Manufacturing and Service Systems, 3
  - MGT 4067 Financial Markets: Trading and Structure, 3
  - MGT 4068 Introduction to Fixed Income, 3
  - PYSC 4031 Applied Experimental Psychology, 3

Cornell: Concentration in Data Science

This minor/concentration is housed in the Department of Information Science. This concentration will equip undergraduate students in information science to learn about the world through data analytics. It is a four course concentration with one course chosen from four categories.

- Data Analysis (select one)
  - INFO 3300: Data-Driven Web Applications

  - CS 4780: Machine Learning for Intelligent Systems
  - CS 4786: Machine Learning for Data Science
  - ORIE 3120: Industrial Data and Systems Analysis
  - ORIE 4740: Statistical Data Mining I
  - STSCI 4740: Data Mining and Machine Learning
- Domain Expertise (select one)
  - INFO 2770: Excursions in Computational Sustainability
  - INFO 3350: Text Mining for History and Literature
  - INFO 4120: Ubiquitous Computing
  - INFO 4130: [Health and Computation]
  - INFO 4300: Language and Information
  - CS 4740: Natural Language Processing
- Big data, ethics, policy and society (select one)
  - INFO 3561: Computing Cultures
  - INFO 4200: [Information Policy: Research, Analysis, and Design]
  - INFO 4240: Designing Technology for Social Impact
  - INFO 4270: Ethics and Policy in Data Science
  - INFO 4561: Stars, Scores, and Rankings: Evaluation and Society
- Data Communication (select one)
  - INFO 4310: Interactive Information Visualization
  - COMM 3189: [Taking America's Pulse: Creating and Conducting a National Opinion Poll]
  - COMM 4200: [Public Opinion and Social Processes]
  - COMM 4860: [Risk Communication]

Berkeley: Data Science Minor (in progress)

The minor, housed in the Division of Data Science, would provide students with Data Science training to complement and integrate with coursework in their own major. Compared to the Data Science major, the minor would allow for a smaller number and broader range of lower division prerequisites. The minor is accessible to students across the university. This minor has not yet been formally approved.

Requirements: part 1

The current proposal includes 6 lower-division requirements

- one semester of Data 8 (COMPSCI/STAT/INFO C8)
- three semesters of lower division Math (Math 1A-1B or Math 10A-10B or Math 16A-16B for calculus, Math 54 or EE 16A-16B or possibly other alternatives to be determined for linear algebra)
- one semester of Computer Science (CS 88 or CS 61A or Engineering 7)
- and one semester of Probability (Math 10B or Stat 88 or CS 70 or an upper-division probability course)

Requirements: part 2

The current proposal includes 4-5 upper-division requirements

- one semester of Data 100 (COMPSCI/STAT C100)
- a selection of three or four additional upper-division courses. These might be specified to fall in particular areas, possibly including one course in your own major.

11.1 University of Pennsylvania: Data Science minor

Housed in the School of Engineering, the minor targets students with strong analytical abilities and some existing programming experience, and requires courses in statistics, data-centric programming, data management, and data analysis. It requires a total of six courses taken from the following list.

- CIS 120 Programming Languages and Techniques I
- CIS 419 Applied Machine Learning 1 or STAT 471 Modern Data Mining
- NETS 212 Scalable and Cloud Computing
- ENM 321 Engineering Statistics or ESE 302 Engineering Applications of Statistics or STAT 431 Statistical Inference
- Data Science Electives (select from two of the following required categories)
  - Data-Centric Programming
    - CIS 110 Introduction to Computer Programming
    - CIS 120 Programming Languages and Techniques I
    - OIDD 311 Business Computer Languages

    - ENGR 105 Introduction to Scientific Computing
    - STAT 405 Statistical Computing with R
    - STAT 470 Data Analytics and Statistical Computing
- Statistics
    - MATH 430 Introduction to Probability
    - ESE 301 Engineering Probability
    - ESE 302 Engineering Applications of Statistics
    - ENM 321 Engineering Statistics
    - STAT 430 Probability
    - STAT 431 Statistical Inference
    - STAT 471 Modern Data Mining
    - STAT 476 Applied Probability Models in Marketing
- Data Collection, Representation, Management and Retrieval
    - CIS 545 Big Data Analytics
    - CIS 450 Database and Information Systems or CIS 550 Database and Information Systems
    - NETS 212 Scalable and Cloud Computing
    - NETS 213 Crowdsourcing and Human Computation
    - OIDD 105 Developing Tools for Data Access and Analysis (VBA and SQL Programming)
    - STAT 434
    - STAT 475 Sample Survey Design
- Data Analysis
    - CIS 419 Applied Machine Learning or CIS 519 Introduction to Machine Learning or CIS 520 Machine Learning
    - CIS 421 Artificial Integlligence
    - MKTG 212 Data and Analysis for Marketing Decisions
    - MKTG 309 Special Topics: Experiments for Business Decision Making
    - OIDD 410 Decision Support Systems
    - STAT 422 Predictive Analytics for Business
    - STAT 435 Forecasting Methods for Management
    - STAT 471 Modern Data Mining
    - STAT 474 Modern Regression for the Social, Behavioral and Biological Sciences
    - STAT 520 Applied Econometrics I
- Modeling

    - NETS 312 Theory of Networks
    - MKTG 271 Models for Marketing Strategy
    - OIDD 325 Computer Simulation Models
    - OIDD 353 Mathematical Modeling and its Application in Finance
    - STAT 433 Stochastic Processes
    - STAT 436 Introduction to Large-Scale Data Science
  - Other electives
    - CIS 106 Visualizing the Past.
    - CIS 125 Technology and Policy

Appendix D: Curriculum Map and Assessment Plan

Measurement Rubric

Course Number and Title:
Instructor:
Term and Year:

Using a 4-point scale please rate how well students in this course demonstrated the following knowledge and skills. A zero indicates that students did not demonstrate this knowledge and/or skill, and a three indicates that students mastered this knowledge and/or skill. Through some form of oral or written communication, students who complete this minor will demonstrate that they can:

SLO1: formulate questions in a domain that can be answered with data	
Understanding the types of questions and information amenable to data science analysis	0 1 2 3
Methods for data description and curation	0 1 2 3
Methods for modeling domain-specific questions	0 1 2 3
Methods for collecting relevant, domain-specific data	0 1 2 3
Understand all components of the data analysis pipeline	0 1 2 3

SLO2: use tools and algorithms from statistics, applied mathematics and computer science for analyses	
Developing strong statistical and computational skills to *do* data science	0 1 2 3
Methods for handling a range of data volumes and velocity	0 1 2 3
Methods for integrating diverse data types (GIS, text, relational, image)	0 1 2 3
Methods for validation of models learned from unstructured data	0 1 2 3

SLO3: visualize, interpret, and explain results cogently, accurately, and persuasively	
Write a compelling report on the results from data analysis	0 1 2 3
Make oral presentations for specific audiences	0 1 2 3
Make poster presentations for specific audiences	0 1 2 3
Design online, interactive visualizations of data	0 1 2 3
Communicating data to a non-technical audience	0 1 2 3

SLO4: understand the underlying social, political, and ethical contexts of data and analysis	
Identify and characterize bias (social and political) in data collection, analysis and presentation	0 1 2 3
Understand privacy concerns in data collection, analysis and presentation	0 1 2 3
Understand importance of establishing data validity (falsification, unbalanced samples, incompleteness)	0 1 2 3
Identify and characterize ethical concerns (fairness, e.g.) in data collection, analysis and presentation	0 1 2 3
Ethical considerations in reuse and sharing of data	0 1 2 3
Reproducibility of results	0 1 2 3

Curriculum map

Course	**Faculty**	**SLO1: formulate questions in a domain that can be answered with data**	**SLO2: use tools and algorithms from statistics, applied mathematics and computer science for analyses**	**SLO3: visualize, interpret, and explain results cogently, accurately, and persuasively**	**SLO4: understand the underlying social, political, and ethical contexts of data and analysis**
DSCI 301	Guerra	Introduced	Introduced	Introduced	Not relevant
DSCI 302	Myers	Reinforced	Reinforced	Not relevant	Introduced
DSCI 303	Henckel	Not relevant	Reinforced	Reinforced	Not relevant
DSCI 304	Joplin	Not relevant	Not relevant	Reinforced	Not relevant
DSCI 305	Roberto	Not relevant	Not relevant	Not relevant	Reinforced
DSCI 400	Allen	Reinforced	Reinforced	Reinforced	Reinforced

Assessment Plan

Outcomes	**SLO1: formulate questions in a domain that can be answered with data**	**SLO2: use tools and algorithms from statistics, applied mathematics and computer science for analyses**	**SLO3: visualize, interpret, and explain results cogently, accurately, and persuasively**	**SLO4: understand the underlying social, political, and ethical contexts of data and analysis**
Embedded location	DSCI 400	DSCI 301, DSCI 302, DSCI 303	DSCI 304	DSCI 305
Measure	Project report in DSCI 400; exit interviews; rubric	DSCI 301, DSCI 302, DSCI 303 homework/exams/projects; exit interviews; rubric	DSCI 304 projects; exit interview; rubric	DSCI 305 final report exit interview; rubric
Standard	80% of students will meet or exceed expectations (2 or above in rubric)	80% of students will meet or exceed expectations (2 or above in rubric)	80% of students will meet or exceed expectations (2 or above in rubric)	80% of students will meet or exceed expectations (2 or above in rubric)
Responsibility	DSCI 400 instructor; director; faculty steering committee	DSCI 301, DSCI 302, DSCI 303 instructors; executive director; faculty steering committee	DSCI 304 instructor; executive director; faculty advisory boardsteering committee	DSCI 305 instructor; executive director; faculty steering committee

Appendix E: Letters of Support

Because the data science minor program will be housed in the School of Engineering, a letter of support from the Dean of Engineering, Reginald DesRoches, has also been included addressing the necessary resources to administer the program (program director, advising resources, resources for teaching quantitative core courses, and resources to support capstone requirement). Additionally, a letter of support from the Dean of Social Sciences, Antonio Merlo, as well as the Dean Canning of Humanities, has been included in reference to DSCI 305, which will be offered from the School of Social Sciences or the School of Humanities. Fred Higgs's letter commits needed resources for DSCI 304. Genevera Allen's

letter supports the key role the D2K lab will play in the delivery of DSCI 400, the capstone course.

References

[1] Data science community: A community for all things data science. http://datascience.community/colleges. accessed April 2017.

[2] K. D. Cooper, R. Baraniuk, K. Ensor, R. Kimbro, K. Ostherr, F. Oswald, P. Padley, and G. Phillips. Report of the Provost's advisory committee on data science. Technical report, Rice University, 2016.

[3] K. D. Cooper, M. L. Miranda, J. E. Odegard, and M. Y. Vardi. The Rice data science initiative: Vision paper. Technical report, Rice University, 2015.

[4] Engineering National Academies of Sciences and Medicine. Envisioning the data science discipline: The undergraduate perspective: Interim report. Technical report, The National Academies Press, 2018.

[5] Rebecca Nugent. The data science (r)evolution, April 2017. talk hosted by the Statistics Department at Rice University.

[6] D. Subramanian, F. Oswald, D. Alexander, M. Heinkenschloss, C. Jermaine, B. Ostdiek, K. Ostherr, and M. Vannucci. Proposal for the data science (X+DS) Minor at Rice University. Technical report, Rice University, May 2017.

Appendix A: Proposed General Announcement Text [for inclusion in 2019-2020 GA]

GA Text submitted is the Undergraduate Program Learning Outcomes and Requirements Tabs text.

Minor in Data Science

Outcomes | Requirements | Policies | Opportunities

Program Learning Outcomes for the Minor in Data Science

Upon completing the minor in Data Science, students will be able to:

1. Formulate questions in a domain that can be answered with data.
2. Use tools and algorithms from statistics, applied mathematics, and computer science for analyses.
3. Visualize, interpret, and explain results cogently, accurately, and persuasively.
4. Understand the underlying social, political, and ethical contexts that are importantly and inevitably tied to data-driven decision-making.

Minor in Data Science

Outcomes | **Requirements** | Policies | Opportunities

Requirements for the Minor in Data Science

Students pursuing the minor in Data Science must complete:

- A minimum of 9 courses (27-29 credit hours, depending on course selection) to satisfy minor requirements.
- A minimum of 6 courses (18 credit hours) taken at the 300-level or above.
- 3 courses (9-10 credit hours, depending on course selection) to satisfy prerequisite requirements.
- 5 courses (15-16 credit hours, depending on course selection) to satisfy core requirements.
- A capstone project (3 credit hours).

The courses listed below satisfy the requirements for this minor. In certain instances, courses not on this official list may be substituted upon approval of the minor's academic advisor, or where applicable, the Program Director. (Course substitutions must be formally applied and entered into Degree Works by the minor's Official Certifier). Students and their academic advisors should identify and clearly document the courses to be taken.

Summary

Total Credit Hours Required for the Minor in Data Science	27-29

Minor Requirements

Prerequisites		
Mathematics [1]		
MATH 101 or MATH 105 or MATH 211	SINGLE VARIABLE CALCULUS I AP/OTH CREDIT IN CALCULUS I ORDINARY DIFFERENTIAL EQUATIONS AND LINEAR ALGEBRA	3
MATH 102 or MATH 106 or MATH 212	SINGLE VARIABLE CALCULUS II AP/OTH CREDIT IN CALCULUS II MULTIVARIABLE CALCULUS	3
Programming [2]		
COMP 140 or CAAM 210	COMPUTATIONAL THINKING INTRODUCTION TO ENGINEERING COMPUTATION	3 or 4
Core Requirements		
Quantitative Core Requirements [3]		
DSCI 301 / STAT 315 or STAT 310	STATISTICS FOR DATA SCIENCE PROBABILITY AND STATISTICS	3 or 4
DSCI 302 or COMP 330	TOOLS AND MODELS FOR DATA SCIENCE TOOLS AND MODELS - DATA SCIENCE	3
DSCI 303 **(NEW)** or STAT 413	MACHINE LEARNING FOR DATA SCIENCE INTRO TO STAT MACHINE LEARNING	3
Additional Core Requirements		
DSCI 304	EFFECTIVE DATA VISUALIZATION	3
DSCI 305 **(NEW)**	DATA, ETHICS AND SOCIETY	3
Capstone Requirement		
DSCI 400 **(NEW)**	CAPSTONE PROJECT	3
Total Credit Hours		**27-29**

Footnotes and Additional Information

[1] Students may substitute a higher-level MATH course for the MATH 101/MATH 105 and MATH 102/MATH 106 prerequisites. See the DSCI Official Certifier for details.

[2] As a programming prerequisite, COMP 140 is highly recommended (due to its inclusion of Python programming in its course material). Students will only be able to register for DSCI 302 if they have COMP 140 in their academic history, or in exceptional circumstances by permission of the DSCI 302 instructor.

[3] In certain situations the DSCI Official Certifier may approve various and specific course substitutions; For example, COMP 540 may be substituted for DSCI 303, etc.

Marie Lynn Miranda, Ph.D.
Howard R. Hughes Provost
Professor of Statistics

2 November 2018

Dr. Jeffrey Fleisher
Chair, Committee for the Undergraduate Curriculum
Rice University

Dear Jeff:

I write in support of the minor in Data Science proposed and developed by the data science curriculum committee. The minor will be housed in the Brown School of Engineering, but has been designed to meet the needs of students from multiple schools and disciplines. The introduction of the minor will advance our offerings to many of our undergraduate students who will benefit from this new program.

I am grateful for the work that chairs Fred Oswald and Devika Subramanian undertook to ensure that faculty were actively engaged in the consideration of the minor and in all aspects of the design of the program. In addition to more technically oriented courses, the minor will also offer courses focused on the broader impact of the information age on our understanding of human activity, including discussion of privacy, ethics, decision-making, culture, and truth in the digital age.

The provost's office recognizes that the courses that count toward the data science minor, as with other majors and minors that Rice offers, have appropriate instructional resources. With regard to instructors, the most important faculty assets are those hired through the Data Science Initiative. These seven faculty, as well as their department chairs, are all aware that half of their teaching obligation is due to the data science curriculum. Chairs do have the latitude to substitute a different faculty member for the same course, but the teaching obligation is clear. We are also making provision to add an instructor line to the Program in Writing and Communications for the visual communication and design course.

With regard to teaching assistants (TAs), not all courses will require the same level of teaching assistant and undergraduate grader support. Working with the relevant deans and departments, we will expect departments to support TAs for existing courses being refocused for the minor. We are assessing the need for a central pool of funds, managed by the Dean of Graduate and Postdoctoral Studies' office, to assist with paying teaching assistants assigned to the Data Science minor courses. We will need to review the success of the funding model each year in light of actual enrollments and the impact of the data science minor on enrollments in other courses and the resulting impact on TA requirements.

I hope you and the CUC share my enthusiasm for this important addition to the Rice curriculum.

With my very best regards,

Marie Lynn Miranda, PhD
Howard R. Hughes Provost and Professor of Statistics

Email provost@rice.edu • Office 713-348-4026 • Rice University Office of the Provost–MS 2 • 430 Allen Center 6100 Main St. • Houston, TX 77005 • www.rice.edu/provost

Office of the Provost
Rice University

To: Committee on Undergraduate Curriculum and Members of the Faculty Senate

From: C. Fred Higgs III, VPAA

Date: November 1, 2018

Re: Support for new interdisciplinary Data Science minor

The office of the Vice Provost for Academic Affairs (VPAA), which is the supervising office of the Program in Writing and Communication (PWC), supports the formation of the interdisciplinary minor in data science. The VPAA and PWC recognize that this program aims to contextualize data science in society and decision-making in innovative ways that will benefit students across the Rice campus.

The PWC is prepared to support up to six sections of DSCI 304: *Introduction to Effective Data Visualization* per year, with a maximum of 30 students in each section. The PWC currently employs a full-time Instructor of Visual Communication and Design. This instructor wrote the curriculum for DSCI 304 and is prepared to teach one section per year, starting in Spring 2019. If the proposal for the new DSCI minor is accepted, the VPAA supports the PWC providing a full-time visual instructor who will teach the other five sections. Appropriate TA and undergraduate assistant support will also be provided.

As the Vice Provost for Academic Affairs, I endorse the creation of the minor in data science and am excited about this new program being initiated at Rice.

Regards,

C. Fred Higgs III, Ph.D.
Vice Provost for Academic Affairs
John & Ann Doerr Professor of Mechanical Engineering
Director, Particle Flow and Tribology Laboratory
RICE University (Houston, TX, USA)

Kathleen M. Canning
Dean, Andrew Mellon Professor of History

September 14, 2018

To: Committee on Undergraduate Curriculum and Members of the Faculty Senate

From: Kathleen Canning, Dean

The School of Humanities is pleased to support the proposed Data Science minor. As an interdisciplinary minor designed to be open to all students in any major on campus, this minor will provide valuable curricular offerings to Humanities students. The School of Humanities is particularly pleased that the minor includes a core, required course on the ethical and social dimensions of data science. We see inclusion of this course as essential to the minor and important in establishing the Rice program's distinctiveness from programs at other universities, few if any of which include required or elective courses in data ethics.

The School of Humanities is currently conducting a search to hire an open-rank professor in ethical and social dimensions of data science, to be appointed in a department in the School of Humanities. We expect that this faculty member will commit to teach the core required course in "Data, Ethics, and Society" (DSCI 305) once per year, while also participating in the DS minor in other ways (such as supervising student research projects). We understand that the course will be capped initially at 40 students, and we commit to providing the appropriate number of TAs from the school each time the course is offered. If the HUMA faculty member is unavailable to teach the course in a given year due to leaves, fellowships, etc., the School of Humanities will work to identify the resources needed to hire a temporary instructor to teach the course.

We also understand that the newly hired professor in the School of Social Sciences, (Elizabeth Roberto, Sociology) will teach DSCI in Spring 2019, and at least once per year thereafter. We expect that the new HUMA hire will collaborate with the SOCS faculty to ensure consistency in course content and pedagogy. We also recognize that anticipated demand for the DS minor is high. If and when demand exceeds capacity in the two sections of DSCI 305 to be offered per year (one in SOCS, one in HUMA, with capacity for a total of 80 students per year), we will work with the DSCI director to identify means to open another section or change the structure of the course to expand annual capacity to 120 or more. Our goal is to ensure that the course be taught in a format that allows for active student dialogue, multiple reflective and analytical writing assignments, and in-class presentations.

If we can provide you with additional information, do let us know.

Email kcanning@rice.edu • Office 713-348-4810 • Fax 713-348-5471 • Rice University School of Humanities–MS 33
P.O. Box 1892 • Houston, TX 77251-1892 • humanities.rice.edu

RICE

REGINALD "REGGIE" DESROCHES
WILLIAM AND STEPHANIE SICK DEAN OF ENGINEERING
PROFESSOR, CIVIL AND ENVIRONMENTAL ENGINEERING
PROFESSOR, MECHANICAL ENGINEERING

MEMO To: Faculty Senate and the University Curriculum Committee

From: Reginald DesRoches, Dean of Engineering

Re: Support for the Interdisciplinary Minor in Data Science

Date: September 12, 2018

I am writing to express my strong support for creating an interdisciplinary minor in data science (DSCI), to be housed in the School of Engineering. As stated in the proposal, this new minor will bring together a new set of six core courses that span three departments in the Schools of Engineering, the School of Social Sciences, the School of Humanities, and the Program in Writing and Communication (PWC). The proposal is well conceived, both in its innovativeness and rigor, and provides ample evidence that Rice has the faculty and administrative resources to support the new minor.

The DSCI minor will be administered by the School of Engineering under the leadership of a to-be-named Executive Director. Our initial commitment is to support the requirements of the minor assuming a demand of 100 minors from across the university. The Engineering school will be responsible for instructional resources to staff the quantitative core courses in the minor offered by departments in Engineering (i.e. DSCI 301, DSCI 302, DSCI 303, and DSCI 400).. If demand for the minor after AY19 exceeds the Engineering School's capacity to fund the program, the Dean's office will work with the Provost to seek additional lines (postdocs, NTT, and/or TT) from the central administration.

As Dean of Engineering, I support the new minor fully and urge the Faculty Senate to do the same.

E-MAIL RD27@RICE.EDU | OFFICE 713-348-4009 | FAX 713-348-5300 | GEORGE R. BROWN SCHOOL OF ENGINEERING
RICE UNIVERSITY-MS 364 | 6100 MAIN ST | HOUSTON, TX 77005 | WWW.ENGR.RICE.EDU

Antonio M. Merlo, Ph.D.
Dean of the School of Social Sciences and
George A. Peterkin Professor of Economics

October 2, 2018

MEMORANDUM

To: Faculty Senate and the Committee on Undergraduate Curriculum

From: Antonio Merlo, Dean of the School of Social Sciences

Re: Support for the Interdisciplinary Minor in Data Science

I enthusiastically endorse the proposal to create an interdisciplinary minor in data science (DSCI). This is an important initiative that exemplifies the wonderful environment of collaboration among the School of Engineering, the School of Social Sciences and the School of Humanities which transcends the somewhat artificial disciplinary boundaries between these schools in an innovative and creative way.

The School of Social Sciences will work closely with the School of Humanities to ensure that a tenured or tenure-track faculty member of either school will teach DSCI 305 -- Data Ethics, one of the six core courses for the minor. The School will also work collaboratively with the director of the data science minor to ensure smooth functioning of DSCI 305 within the core curriculum.

As Dean of the School of Social Sciences, I support the new minor wholeheartedly.

Email amerlo@rice.edu • Office 713-348-4824 • Rice University School of Social Sciences—MS 27
Baker Hall • Suite 180 • P. O. Box 1892 • Houston, TX 77251-1892 • socialsciences.rice.edu

Genevera I. Allen
Founder & Faculty Director,
Rice Data to Knowledge (D2K) Lab;
Associate Professor,
Departments of Statistics,
Electrical and Computer Engineering, &
Computer Science, Rice University.
Jan and Dan Duncan Neurological Research Institute,
Texas Children's Hospital & Baylor College of Medicine.
6100 Main St. MS-138, Houston, TX 77005.
(713)-348-6321; gallen@rice.edu.

October 26, 2018

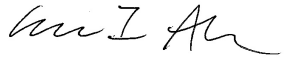
Dear Committee,

I am writing this letter of support for the proposal to establish a Data Science minor. The Rice Data to Knowledge (D2K) Lab is happy to teach the capstone requirement, DSCI 400, for the Data Science minor, provided adequate resources (e.g. teaching faculty) are provided to teach this course. The Rice D2K Lab (d2k.rice.edu) is a new center within the School of Engineering that provides students with experiential learning opportunities in data science while enhancing data-intensive research at Rice, and building partnerships with companies, institutions, and community organizations. I am the founder and faculty director for the D2K Lab.

The capstone for the Data Science minor will be a team-based data science project course that provides hands-on experience working on real data science challenges. Data Science minor students with a strong technical background will join D2K Learning Lab (d2k.rice.edu/about/d2k-learning-lab) teams who work on client-sponsored projects brought in by companies, government and community partners, and researchers from Rice or the Texas Medical Center. Data Science minor students with less technical background will be in a separate course where teams focus on working with cleaner, curated data sets. Both courses will teach best practices for complex, team-based data science and include modules on: designing a data science pipeline, communication (oral, written, and via code), technologies and practices for computational reproducibility and team-based software development, teamwork and project management, and scientific reproducibility and validation strategies for data science.

Please let me know if you have any further questions.

Sincerely,



Genevera I. Allen