DATA SCIENCE

Overview and Contact Information

The major in data science aims to guide students to be effective, ethical, and judicious users, interpreters, analyzers, and communicators of data and data-related concepts. The major offers students a foundational understanding of the data generating process, the appropriate and efficient translation of analytic strategies to specific data settings, the potential biases arising from missing data or data collection, the means for drawing accurate conclusions, and the techniques and principles of integrity in data visualization and communication. As part of their data science education, students will develop excellent communication skills and the ability to make clear and persuasive arguments framed by logic and supported by data. The data science curriculum reflects the increasingly collaborative and interdisciplinary academic landscape.

See Also

- Computer Science (http://catalog.mtholyoke.edu/areas-study/computer-science/)
- Statistics (http://catalog.mtholyoke.edu/areas-study/statistics/)
- Data Analytics and Society (http://catalog.mtholyoke.edu/areas-study/data-analytics-society/)

Contact Information

Martha Hoopes, Chair (Fall 2024)
Connell Heady, Academic Department Coordinator

415A Clapp Laboratory
413-538-2162
https://www.mtholyoke.edu/academics/find-your-program/data-science (https://www.mtholyoke.edu/academics/find-your-program/data-science/)

Learning Goals

- Apply core concepts of statistics, computing, and domain knowledge to extract insight from data sets.
- Understand the social and ethical issues surrounding data collection, analysis and use.
- Be able to communicate in multiple modalities the results of large scale data analysis.

Faculty

This area of study is administered by the Data Science Program Committee:

Maria Gomez, Elizabeth Page Greenawalt Professor of Chemistry

Martha Hoopes, Professor of Biological Sciences, Teaching Fall Only

Audrey Lee St. John, Professor of Computer Science

Heather Pon-Barry, Associate Professor of Computer Science

Benjamin Gebre-Medhin, Assistant Professor of Sociology

Laura Sizer, Senior Lecturer in Philosophy, Teaching Fall Only

Requirements for the Major

A minimum of 40 credits:

Code	Title	Credits
STAT-140	Introduction to the Ideas and Applications of Statistics	4
STAT-242	Intermediate Statistics	4
As a prerequisite for MATH-211:		
MATH-102	Calculus II (or above)	
MATH-211	Linear Algebra	4
COMSC-151	Introduction to Computational Problem Solving [1]	4
COMSC-205	Data Structures	4
12 credits at the 300 level from at least two departments or programs and chosen from the approved list of elective courses for Data Science. One course must be either: [2, 3, 4, 5]		12
COMSC-335	Machine Learning [6, 7]	
or STAT-340	Applied Regression Methods	
8 additional credits chosen from the approved list of elective courses for Data Science [2, 3, 4, 5]		8
Total Credits		**40**

[1] Any COMSC-151 offering, for example, COMSC-151CP, COMSC-151DS, or COMSC-151HC.

[2] Students who do not elect both COMSC-335 and STAT-340 will need to choose two other 300-level courses from this list, one of which is from a department other than their COMSC-335/STAT-340 choice.

[3] Many elective courses require prerequisites. Students are encouraged to plan their elective courses early in order to ensure that they meet the requirements to access chosen courses.

[4] Students are strongly encouraged to take an elective course in ethics.

[5] Other courses that focus on ethics, cover data analytic methods, or involve an independent project with data can be substituted with approval of the Data Science Program Committee.

[6] Students intending to attend graduate school in data science are advised to take both of these courses.

[7] COMSC-335 Machine Learning requires MATH-232 as a prerequisite.

Additional Specifications

- Students who declare a Data Science major automatically fulfill the College's "outside the major" requirement.

Course Advice

Students Considering a Major in Data Science

Data science is new and evolving; there are many important combinations of theoretical, applied, and field-specific knowledge that may provide a foundation for future work. If you are interested in a data science major, we recommend that you work with your advisor to choose a set of related courses that reflect your interests and priorities from the list of electives. Course combinations that focus on individual topics, disciplines, or domains are strongly recommended. We also strongly recommend substantial engagement with issues of ethics, which could be in one focused course or across multiple courses.

Students Considering Graduate School or a Career as a Data Scientist:

While there are many fields for which the combination of data analysis and computational tools may be valuable, we have particular recommendations for students seeking a future as a data scientist. We strongly recommend that you take both COMSC-335 Machine Learning and STAT-340 Applied Regression Methods. Ideally, at least one course

should involve an extended project requiring the analysis of data. We also recommend that you contextualize your data science preparation in the content of a domain or area of study that is theoretically and empirically cohesive.

Course Offerings

DATA-113 Introduction to Data Science
Fall and Spring. Credits: 4
Data scientists answer questions with scientific and social relevance using statistical theory and computation. We will discuss elementary topics in statistics and learn how to write code (in Python) to visualize data and perform simulations. We will use these tools to answer questions about real data sets. We will also explore ethical issues faced by data scientists today.
Applies to requirement(s): Math Sciences
K. Mulder, A. Shaus, L. Tupper

DATA-225 Topics in Data Science:

DATA-225AR Topics in Data Science 'Ethics and Artificial Intelligence'
Not Scheduled for This Year. Credits: 4
Artificially intelligent technologies are prominent features of modern life -- as are ethical concerns about their programming and use. In this class we will use the tools of philosophy to explore and critically evaluate ethical issues raised by current and future AI technologies. Topics may include issues of privacy and transparency in online data collection, concerns about social justice in the use of algorithms in areas like hiring and criminal justice, and the goals of developing general versus special purpose AI. We will also look at ethics for AI: the nature of AI 'minds', the possibility of creating more ethical AI systems, and when and if AIs themselves might deserve moral rights.
Crosslisted as: PHIL-260AR, EOS-299AR
Applies to requirement(s): Humanities
L. Sizer

DATA-295 Independent Study
Fall and Spring. Credits: 1 - 4
The department
Instructor permission required.

DATA-350 Advanced Topics in Data Science:

DATA-350TE Advanced Topics in Data Science: 'Technology, Ethics, and Public Policy'
Spring. Credits: 4
In this course, we study the most pressing ethical concerns relating to emerging technology and envision novel policy solutions to address them. Existing regulatory and policy instruments are often unable to provide sufficient oversight for emerging technology. Can legal anti-discrimination doctrine address biased algorithmic decision-making systems? How does generative artificial intelligence challenge traditional ways of thinking about intellectual property? Do we have rights over the personal data that private firms collect about us? We examine these gaps in the context of contemporary regulatory proposals on national, multinational, and international scales.
Crosslisted as: PHIL-350TE
Applies to requirement(s): Humanities
Other Attribute(s): Writing-Intensive
A. Ali
Prereq: 8 credits in Philosophy.

DATA-390 Data Science Capstone
Fall and Spring. Credits: 4
The Capstone is a research seminar that empowers students to design and execute a significant data science research project. Through group review of journal articles and targeted lectures, students will develop a thorough understanding of each of the components of a successful research project including defining their research question, conducting a literature review, identifying an appropriate data set, designing and implementing a defensible methodology, and presenting and interpreting their results in text, tables, and figures. There will be frequent opportunities for students to present their work, and their capstone will culminate in a written report. Concurrently, students will read and discuss several case studies that address issues of ethics involved with the collection, treatment, and analysis of data.
Applies to requirement(s): Math Sciences
K. Mulder, A. Shaus
Prereq: STAT-340 or COMSC-335. STAT-340 or COMSC-335 may be taken concurrently (contact instructor for permission).

DATA-395 Independent Study
Fall and Spring. Credits: 1 - 8
The department
Instructor permission required.

DATA-395P Independent Study w/Practicum
Fall and Spring. Credits: 1 - 8
The department
Instructor permission required.

Required Core Courses for the Data Science Major

Code	Title	Credits
Computer Science		
COMSC-151	Introduction to Computational Problem Solving	4
COMSC-205	Data Structures	4
COMSC-335	Machine Learning	4
Mathematics		
MATH-211	Linear Algebra	4
Statistics		
STAT-140	Introduction to the Ideas and Applications of Statistics	4
STAT-242	Intermediate Statistics	4
STAT-340	Applied Regression Methods	4

Note: Majors need to take either COMSC-335 or STAT-340.

Elective Courses for the Data Science Major

Code	Title	Credits
Biological Sciences		
BIOL-223	Ecology	4
BIOL-234	Biostatistics	4
BIOL-321GE	Conference Course: 'Genomics and Bioinformatics'	4
Chemistry		
CHEM-291	Scientific Illustration and Data Visualization	4
CHEM-328	From Lilliput to Brobdingnag: Bridging the Scales Between Science and Engineering	4
CHEM-348	Using Data Science to Find Hidden Chemical Rules	4

Computer Science

COMSC-133DV	Data Visualization: Design and Perception	4
COMSC-235	Applications of Machine Learning	4
COMSC-312	Algorithms	4
COMSC-334	Artificial Intelligence	4
COMSC-335	Machine Learning	4
COMSC-341CD	Topics: 'Causal Inference for Data Science'	4
COMSC-341NL	Topics: 'Natural Language Processing'	4
COMSC-341TE	Topics: 'Text Technologies for Data Science'	4

Data Science

DATA-113	Introduction to Data Science	4
DATA-225AR	Topics in Data Science 'Ethics and Artificial Intelligence'	4
DATA-350TE	Advanced Topics in Data Science: 'Technology, Ethics, and Public Policy'	4
DATA-390	Data Science Capstone	4

Economics

ECON-220	Introduction to Econometrics	4
ECON-320	Econometrics	4

Entrepreneurship, Orgs & Soc

EOS-299AR	Topic: 'Ethics and Artificial Intelligence'	4

Geography

GEOG-205	Mapping and Spatial Analysis	4
GEOG-210	GIS for the Social Sciences and Humanities	4

Mathematics

MATH-339PT	Topics in Applied Mathematics: 'Optimization'	4
MATH-339SP	Topics in Applied Mathematics: 'Stochastic Processes'	4
MATH-342	Probability	4

Philosophy

PHIL-260AR	Topics in Applied Philosophy: 'Ethics and Artificial Intelligence'	4

Psychology

PSYCH-326CP	Laboratory in Personality and Abnormal Psychology: 'Advanced Statistics in Clinical Psychology'	4

Sociology

SOCI-216TX	Special Topics in Sociology: 'Text as Data I: From Qualitative to Quantitative Text Analysis'	4
SOCI-316TX	Special Topics in Sociology: 'Text as Data II: Computational Text Analysis for the Social Sciences'	4

Statistics

STAT-244MP	Intermediate Topics in Statistics: 'Survey Sampling'	4
STAT-244NF	Intermediate Topics in Statistics: 'Infectious Disease Modeling'	4
STAT-244NP	Intermediate Topics in Statistics: 'Nonparametric Statistics'	4
STAT-244SC	Intermediate Topics in Statistics: 'Computational Statistics'	4
STAT-331	Design of Experiments	4
STAT-340	Applied Regression Methods	4
STAT-343	Mathematical Statistics	4
STAT-344ND	Seminar in Statistics and Scientific Research: 'Analysis of Neural Data'	4
STAT-344TM	Seminar in Statistics and Scientific Research: 'Time Series Analysis'	4
STAT-351	Bayesian Statistics	4