**DATA SCIENCE**

**Overview and Contact Information**

The major in data science aims to guide students to be effective, ethical, and judicious consumers, analysts 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 curriculum is flexible and innovative, broad enough to serve a student population that is diverse in their backgrounds and disciplinary interests, and deep enough to accommodate students who want ultimately to pursue advanced study in statistics and computer science. 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**

Dylan Shepardson, Chair
Connell Heady, Academic Department Coordinator

415A Clapp Laboratory
413-538-2162
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 ethical challenges and potential privacy issues involved in data analysis.
- 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, Teaching Fall Only

Dylan Shepardson, Robert L. Rooke Associate Professor of Mathematics
Benjamin Gebre-Medhin, Assistant Professor of Sociology

Marie Ozanne, Clare Boothe Luce Assistant Professorship in Statistics, On Leave 2022-2023

---

**Requirements for the Major**

A minimum of 40 credits:

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT-140</td>
<td>Introduction to the Ideas and Applications of Statistics</td>
<td>4</td>
</tr>
<tr>
<td>STAT-242</td>
<td>Intermediate Statistics</td>
<td>4</td>
</tr>
<tr>
<td>STAT-340</td>
<td>Applied Regression Methods</td>
<td>4</td>
</tr>
<tr>
<td>MATH-102</td>
<td>Calculus II (or above)</td>
<td></td>
</tr>
<tr>
<td>MATH-211</td>
<td>Linear Algebra</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-151</td>
<td>Introduction to Computational Problem Solving</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-205</td>
<td>Data Structures</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-335</td>
<td>Machine Learning</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>or 300-level alternative to COMSC-335</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Two courses at the 200 level or above within a single domain area</td>
<td>8</td>
</tr>
<tr>
<td>DATA-390</td>
<td>Data Science Capstone</td>
<td>4</td>
</tr>
</tbody>
</table>

Total Credits 40

---

1 The combinations of COMSC-150 and COMSC-121 or of COMSC-150 and COMSC-161 are equivalent to COMSC-151.
2 The combination of COMSC-205PY and COMSC-122 is equivalent to COMSC-205.
3 A domain area – chosen in consultation with the student’s Data Science advisor – is defined as any College-defined major excluding mathematics, statistics, and computer science. Course selection must be approved by the student’s Data Science advisor.
4 The study of ethics in relation to data science is integrated throughout the curriculum and emphasized in this integrative capstone course.

**Other Requirements**

- At the time of major declaration, a domain area will be selected by the student in consultation with an advisor from Data Science.
- Prior to the DATA-390 Capstone course, each Data Science major will submit to their advisor a brief document of reflection on the domain area, its connection to data science, and topics they might pursue for their capstone research. The Capstone will be offered in the spring term and be run as a research seminar.

**Additional Specifications**

- Course substitutions through the Five Colleges require pre-approval in writing by an advisor from Data Science.
- Independent studies cannot be used to satisfy any of the above requirements unless approved by the Data Science Program Committee (with the possible exception of the capstone).
- Students who declare a Data Science major automatically fulfill the College’s “outside the major” requirement.

**Sample Domain Pathways**

At the time of major declaration, the student selects a domain area in consultation with an advisor from Data Science. Some sample pathways are described below:
Chemistry
Analytical and physical chemists often generate and analyze significant amounts of data. Analysis methods learned in analytical or physical chemistry courses are regularly applied to organic, inorganic, or biochemical systems. Two course sequences highlighting both methods and systems could include (a) a course in analytical or physical chemistry and (b) a course with a focus on organic, inorganic, or biochemical materials. More data generation and analysis based two course sequences can be two courses in analytical and/or physical chemistry. All first courses in the above sub-areas of chemistry CHEM-150 General Chemistry: Foundations of Structure and Reactivity and some also require CHEM-202 Organic Chemistry I and/or MATH-203 Calculus III.

Economics
Data touches nearly all parts of economics by informing models and revealing patterns and causal relationships in data. Data science is becoming an essential part of every subfield in economics. For example students interested in: (1) finance might take ECON-270 Accounting and ECON-215 Economics of Corporate Finance; (2) development might take ECON-213 Economic Development; (3) theory might take ECON-201 Game Theory and ECON-212 Microeconomic Theory. Almost all all 200-level courses in economics require ECON-110 Introductory Economics as a prerequisite.

English
Digital Humanities and New Media Studies represent two humanities avenues for potential cross-pollination with data analysis. Topic modeling, text mining, and database construction for interactive editions of texts are examples of particular areas of digital humanities that lend themselves to asking interesting questions about large humanities corpora. Students interested in English and Data Science would like to take courses in literary analysis and at least one upper-level course in digital humanities in the Five Colleges. For example, students interested in: (1) text analysis of literature and the environment might take ENGL-240 American Literature I and ENGL-366 Love, Sex, and Death in the Anthropocene, or Living Through the Age of Climate Change and Other Disasters. Alternatively, (2) exploring large corpora might take survey courses offering breadth, e.g., ENGL-251 Contemporary African American Literature II or ENGL-241 American Literature II, ENGLISH-302 (UMass) Studies in Textuality and New Media or ENGL-390 (Amherst) Digital Humanities. Ideally, students would also take ENGL-199 Introduction to the Study of Literature.

Course Advice
The courses listed below form the core of the Data Science curriculum. In addition to core courses, students majoring in Data Science will take courses from their selected domain areas in consultation with their Data Science advisors.

Course Offerings

DATA-113 Introduction to Data Science
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

DATA-225 AR Topics in Data Science 'Ethics and Artificial Intelligence'
Spring. 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
Applies to requirement(s): Humanities
Other Attribute(s): Writing-Intensive
L. Sizer

DATA-239 Digital Image Processing
Spring. Credits: 4
An introduction to the mathematical aspects behind digital image processing. Students will learn how digital images can be represented using matrices and how matrix manipulations lead to image enhancements. Topics include converting images to black and white, adjusting image brightness via gamma correction, blurring images using smoothing masks, and sharpening images using derivative filters. Image restoration techniques such as the removal of noise or motion blur will be discussed. Students will apply theoretical knowledge to images using MATLAB. The methods in this course can be applied to several career fields including medical imaging, computer vision, forensics, security, and defense.
Crosslisted as: MATH-239
Applies to requirement(s): Math Sciences
S. Kirk
Prereq: MATH-102.

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

DATA-390 Data Science Capstone
Spring. Credits: 4
The Capstone is a research seminar that brings together the three pillars of the Data Science curriculum. The course will start with common readings about research projects across a range of disciplines, including readings that address issues of ethics involved with the collection, treatment, and analysis of data. Concurrently, each student will develop an individual research topic and identify relevant data resources. The remainder of the term will be dedicated to exploring these topics through extensive data analysis, visualization, and interpretation, leading to a final report with complete results and a presentation.
Applies to requirement(s): Math Sciences
K. Mulder
Prereq: COMSC-205 and STAT-340. STAT-340 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
Instructor permission required.

Courses in Other Departments Counting toward the Major in Data Science

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHEM-348</td>
<td>Using Data Science to Find Hidden Chemical Rules</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-151AA</td>
<td>Introduction to Computational Problem Solving: 'Algorithmic Arts'</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-151AR</td>
<td>Introduction to Computational Problem Solving: 'Artificial Intelligence'</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-151CP</td>
<td>Introduction to Computational Problem Solving: 'Computing Principles'</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-151DS</td>
<td>Introduction to Computational Problem Solving: 'Big Data'</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-151HC</td>
<td>Introduction to Computational Problem Solving: 'Humanities Computing'</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-151SG</td>
<td>Introduction to Computational Problem Solving: 'Computing for Social Good'</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-205</td>
<td>Data Structures</td>
<td>4</td>
</tr>
<tr>
<td>COMSC-335</td>
<td>Machine Learning</td>
<td>4</td>
</tr>
<tr>
<td>Data Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DATA-225AR</td>
<td>Topics in Data Science 'Ethics and Artificial Intelligence'</td>
<td>4</td>
</tr>
<tr>
<td>DATA-390</td>
<td>Data Science Capstone</td>
<td>4</td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MATH-211</td>
<td>Linear Algebra</td>
<td>4</td>
</tr>
<tr>
<td>Philosophy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHIL-260AR</td>
<td>Topics in Applied Philosophy: 'Ethics and Artificial Intelligence'</td>
<td>4</td>
</tr>
<tr>
<td>Statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAT-140</td>
<td>Introduction to the Ideas and Applications of Statistics</td>
<td>4</td>
</tr>
<tr>
<td>STAT-242</td>
<td>Intermediate Statistics</td>
<td>4</td>
</tr>
<tr>
<td>STAT-340</td>
<td>Applied Regression Methods</td>
<td>4</td>
</tr>
</tbody>
</table>