DATA SCIENCE (DATA)

undefined

DATA Courses

DATA 1000 Statistical and Data Literacy (3 units)
Term Typically Offered: F, SP
2026-28 or later catalog: GE Area 2
2020-26 catalogs: GE Area B4

Using data to answer questions, with emphasis on working with tabular data in spreadsheet software to provide insights via descriptions and visualizations. Data sources, sampling, collecting data. Multivariable thinking, analysis, and visualization. Decision-making when faced with uncertainty. Data ethics. 3 lectures. Fulfills GE Area 2 with a grade of C- or better (GE Area B4 for students on the 2020-26 catalogs). Crosslisted as DATA/STAT 1000. Formerly STAT 130.

DATA 1264 Calculus for Data Science I (4 units)
Term Typically Offered: F, SP
2026-28 or later catalog: GE Area 2
2020-26 catalogs: GE Area B4
Prerequisite: Appropriate Math Placement, MATH 1005 with a grade of C- or better, or MATH 1007 with a grade of C- or better.

Limits, continuity, and differentiation of rational, exponential, logarithmic, trigonometric, and inverse trigonometric functions. Optimization. Techniques of integration. Differential equations. Parametric equations and polar coordinates. Not open to students with credit in MATH 141 or MATH 1261. 4 lectures. Crosslisted as DATA/MATH 1264. Fulfills GE Area 2 with a grade of C- or better (GE Area B4 for students on the 2020-26 catalogs).

DATA 1265 Calculus for Data Science II (4 units)
Term Typically Offered: F, SP, SU
Prerequisite: MATH 121 or DATA/MATH 1264 with a grade of C- or better.

Vectors and vector functions. Partial derivatives. Multiple integrals. Lagrange Multipliers. Sequences and series. 4 lectures. Crosslisted as DATA/MATH 1265.

DATA 1810 Introduction to Statistical Computing with R (3 units)
Term Typically Offered: F, SP
Prerequisite: One of the following: DATA/STAT 1000, STAT 130, STAT 218, STAT 252, STAT 301, STAT 312, STAT 1110, STAT 1220, STAT 1510, or STAT 3210.

Development environment and core elements of statistical computing in R. Importing and managing tabular data. Objects. Data types. Data visualizations and numerical summaries. Logical operations, iteration, and function writing. Debugging. Reproducible documents. Simulation and random number generation. Statistical inference methods. Course may be offered in classroom-based, online, or hybrid format. 3 lectures. Crosslisted as DATA/STAT 1810. Replaced STAT 331.

DATA 2621 Introduction to Mathematical Optimization (3 units)
Term Typically Offered: F, SP
Prerequisite: One of the following: MATH 123, MATH 241, DATA/MATH 1265, or MATH 2263; and one of the following: MATH 206, MATH 244, MATH 1151, or MATH 2341.

Algorithms and mathematical analysis for solving optimization problems. One-dimensional search methods; Newton's method extended to multi-dimensional optimization problems and gradient methods. The basic mathematics of neural networks and examples of their application. Optimization problems with equality and inequality constraints. 3 lectures. Crosslisted as DATA/MATH 2621. Formerly MATH 253.

DATA 3301 Introduction to Data Science (4 units)
Term Typically Offered: F, SP
Prerequisite: MATH 206 or MATH 1151; CSC 101 or CSC 1001; and one of the following: STAT 217, STAT 218, STAT 1110, STAT 301, STAT 1510, STAT 312, or STAT 3210.

Introduction to the field of data science and associated tools and libraries. Algorithmic machine learning and modeling, types of data, data wrangling, data summary, validating and analyzing results, visualization, and applications of data science. 3 lectures, 1 laboratory. Formerly DATA 301.

DATA 3302 Data Visualization (4 units)
Term Typically Offered: F, SP
Prerequisite: CSC 101 or CSC 1001; and STAT 331 or STAT 1810.

Effectively communicate insights from data through visual displays. Basic and advanced chart types, data visualization design principles and critique visualizations in the media, multiple displays and interactive graphics. Geographical and time series, uncertain, and non-traditional data representation. 3 lectures, 1 laboratory.

DATA 3622 Mathematics of Data Science (3 units)
Term Typically Offered: F
Prerequisite: One of the following: MATH 206, MATH 244, MATH 1151, or MATH 2341; MATH 248 or MATH 2031; and CSC 101 or CSC 1001.

Introduction to mathematical foundations of data science including regression, data dimension reduction, clustering, community detection, and computational topology, with a focus on advanced linear algebra tools needed to establish the mathematical foundations of methods, to show convergence, and for error minimization. 3 lectures. Crosslisted as DATA/MATH 3622.

DATA 3800 Introduction to Statistical Computing with SAS and SQL (3 units)
Term Typically Offered: F, SP
Prerequisite: One of the following: STAT 252, STAT 1220, STAT 312, STAT 3210, STAT 302 or STAT 3520.

Using SAS to access and manage data, generate reports, and export results; graphical procedures, basic descriptive and inferential statistics. Introduction to SAS macros, and SQL for data management within the SAS environment. Course may be offered in classroom-based, online, or hybrid format. 3 lectures. Crosslisted as DATA/STAT 3800. Formerly STAT 330.

DATA 3820 Intermediate Statistical Computing with R (3 units)
Term Typically Offered: F, SP
Prerequisite: STAT 331 or STAT 1810; and one of the following: STAT 252, STAT 1220, STAT 312, STAT 3210, STAT 302 or STAT 3520.

Intermediate and advanced techniques for use of R Statistical Software to analyze data. Version control systems; reproducibility and documentation; data collection and wrangling; functional programming; randomization and bootstrapping; and dynamic data visualizations. Course may be offered in classroom-based, online, or hybrid format. 3 lectures. Crosslisted as DATA/STAT 3820. Formerly STAT 431.

DATA 4011 Optimization in Julia for Economics and Data Science (3 units)
Term Typically Offered: F
Prerequisite: One of the following: ECON 395, CSC 101, CSC 231, CSC 232, ECON 3015, CSC 1001, CSC 1031 or CSC 1032.

Optimization as a foundation for Economics and Data Science, using the Julia environment. Constrained and unconstrained least squares via QR factorization. Duality. Applications to game theory, resource allocation, econometrics, and machine learning. 3 lectures. Crosslisted as DATA/ECON 4011.

DATA 4011A Julia for Data Analysis Project (1 unit)
Term Typically Offered: TBD
Corequisite: ECON 4011.

Completion of coding exercises for data analysis in Julia under faculty supervision that complements the coursework on optimization in Julia for economics and data science. Course offered online only. 1 activity. Crosslisted as DATA/ECON 4011A.

DATA 4401 Data Science Process and Ethics (4 units)
Term Typically Offered: F
Prerequisite: DATA 4610.

Complete life cycle of a data science project. Requirements engineering and data acquisition. Management and integration of data of high volume, velocity, and variety. Deployment of data science products. Engagement with stakeholders. Ethical considerations, including privacy and fairness. 3 lectures, 1 laboratory. Formerly DATA 401.

DATA 4460 Senior Project - Data Science Capstone (2 units)
Term Typically Offered: SP
Prerequisite: DATA 4401.

Team-based design, implementation, deployment, evaluation, and delivery of a system or analytical methodology that involves working with and analyzing large quantities of data. Management of research and development teams. Documentation, visualization and presentation of results orally and in writing. 2 laboratories. Formerly DATA 451.

DATA 4461 Senior Project - Bioinformatics Capstone (2 units)
Term Typically Offered: SP
Prerequisite: One of the following: BIO 351, BIO 3351, CHEM 373, CHEM 3356, or MCRO 3351; one of the following: BIO 441, BIO 4451, CSC 448, or CSC 4448; and DATA 301 or DATA 3301.

Work with clients to design bioinformatics solutions to biological questions. Software requirements, elicitation techniques, data gathering, project planning, and project team organization. Ethics and professionalism. Course may be offered in classroom-based, online, or hybrid format. 2 laboratories. Formerly DATA 441.

DATA 4610 Fundamentals of Machine Learning (4 units)
Term Typically Offered: F, SP
Prerequisite: DATA 301 or DATA 3301; and MATH 253 or MATH 2621; or graduate standing in Statistics.

Theory and practice of Machine Learning, with focus on optimization-based methods and procedures. Likelihood and least-squares estimation. Linear, nonlinear, and penalized regression. Classification by linear separators. Gradient descent and boosting. Unsupervised and semi-supervised learning. Not open to students with credit in CSC 487 or CSC 4667. 3 lectures, 1 laboratory. Crosslisted as CSC/DATA 4610.

DATA 4620 Foundations and Applications of Deep Learning (4 units)
Term Typically Offered: F, SP
Prerequisite: DATA 4610.

Overview of modern machine learning techniques that relate to deep learning. Perceptrons, Feed-Forward Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks, Autoencoders and Decoders, Generative Adversarial Networks, and Transformers. Not open to students with credit in CSC 487 or CSC 4667. 3 lectures, 1 laboratory. Crosslisted as CSC/DATA 4620.

DATA 4632 Graph Mining (2 units)
Term Typically Offered: SP
Prerequisite: DATA 301 or DATA 3301; and CSC 349 or CSC 3449; or graduate standing in Statistics.

Graphs as data collections. Hubs and authorities. Node prestige in graphs. Markov chains. Perron-Frobenius theorem. PageRank algorithm and its extensions. Community discovery. Applications of graph mining. 2 lectures.

DATA 4720 Data Science Seminar (1 unit)
Term Typically Offered: TBD
CR/NC
Prerequisite: DATA 301 or DATA 3301.

Discussions of technical, societal and ethical aspects of modern data science theory and practice, concentrating on topics not covered in other courses. Total credit limited to 3 units. Credit/No Credit grading only. 1 seminar. Formerly DATA 472.

DATA 4810 SAS Certification Preparation: Base Programming (1 unit)
Term Typically Offered: TBD
Prerequisite: STAT 330 or STAT 3800; or graduate standing in Statistics.

Preparation for the Base Programming Specialist SAS certification exam offered by the SAS Institute. Includes using SAS to access and manage data, generate reports, and export results. 1 lecture. Crosslisted as DATA/STAT 4810. Formerly STAT 440.

DATA 4820 SAS Certification Preparation: Advanced Programming (2 units)
Term Typically Offered: TBD
Prerequisite: STAT 440 or STAT 4810.

Preparation for the Advanced Programming Professional SAS certification exam offered by the SAS Institute. Includes accessing and managing data using PROC SQL, macro processing, and using advanced SAS programming techniques such as arrays and hash objects. 2 lectures. Crosslisted as DATA/STAT 4820. Formerly STAT 441.

DATA 5550 Statistical Learning with R (3 units)
Term Typically Offered: SP
Prerequisite: Graduate standing; or STAT 331 or DATA/STAT 1810; one of the following: STAT 305, STAT 2610, STAT 350, or STAT 3310; and one of the following: STAT 334, STAT 513, STAT 3530, or STAT 5120.

Modern methods in predictive modeling. Supervised and unsupervised learning. Regression, classification, and clustering methods, including SVM, LASSO, splines, trees, and random forests. Model assessment and selection using cross validation, bootstrapping, and information criteria. Use of the R programming language. Course may be offered in classroom-based or hybrid format. 3 lectures. Crosslisted as DATA/STAT 5550. Formerly STAT 551.

DATA 5800 Introduction to SAS and SQL for Graduate Students (3 units)
Term Typically Offered: F, SP
Prerequisite: Graduate standing; and one of the following: STAT 252, STAT 1220, STAT 312, STAT 3210, STAT 302, STAT 3520, STAT 513, STAT 5120, STAT 542, or STAT 5210.

Using SAS to access and manage data, generate reports, and export results; graphical procedures, basic descriptive and inferential statistics. Introduction to SAS macros, and SQL for data management within the SAS environment. Not open to students with credit in STAT 330, STAT 3800 or Statistics majors. Course may be offered in classroom-based, online, or hybrid format. 3 lectures. Crosslisted as DATA/STAT 5800. Formerly STAT 530.

DATA 5820 Intermediate Statistical Computing with R for Graduate Students (3 units)
Term Typically Offered: F, SP
Prerequisite: STAT 331 or STAT 1810; one of the following: STAT 252, STAT 1220, STAT 312, STAT 3210, STAT 302, or STAT 3520; and graduate standing.

Intermediate and advanced techniques for use of R Statistical Software to analyze data. Version control systems; reproducibility and documentation; data collection and wrangling; functional programming; randomization and bootstrapping; and dynamic data visualizations. Not open to Statistics majors. Course may be offered in classroom-based, online, or hybrid format. 3 lectures. Crosslisted as DATA/STAT 5820. Formerly STAT 531.