Data Science (DATA)

Professors: P. L. Shick (Chair), B. Foreman, M. Kirschenbaum, D. W. Palmer, L. Seiter;
Assistant Professors: E. Manilich, W. Marget, R. Fang

Major Programs

The Department of Mathematics and Computer Science offers a Bachelor of Science in Data Science. The department also offers mathematics (MT), computer science (CS), and computer science with health care technology (CS HCT) programs that are described in separate sections.

Data science is an emerging academic discipline, a response to an increasing demand for people who are able to understand and analyze data. Data science provides powerful approaches for transforming large and complex data into information, knowledge, and intelligent decisions. John Carroll University's Data Science program utilizes an interdisciplinary approach with a focus on Statistics, Computer Science, and Mathematics applied to a specific discipline such as Communications, Digital Humanities, Entrepreneurship, Exercise Science, Biology, Physics, Political Science, Psychology, and Sociology.

The major in **Data Science** leading to the B.S. prepares students to explore the complex relationships between data, technology and society. Data is everywhere, so data science skills are increasingly critical in almost every discipline. Data science careers span the spectrum of possibilities from working for national companies with large data analytics departments to founding entrepreneurial start-up companies at the frontier of future technology, while filling critical roles in all fields including medicine, business, arts and entertainment, sports, government, law, manufacturing, and research.

****

Program Learning Goals in Data Science:

1. **Data Acquisition**: collect, store, preserve, manage and share data in a distributed environment through practical, hands-on experience with programming languages and big data tools;
2. **Problem Exploration**: develop problem solving skills through experiences that foster computational and data-analytic thinking;
3. **Analysis**: develop an in-depth understanding of the key technologies in data science: data mining, machine learning, visualization techniques, predictive modeling, and statistics;
4. **Domain knowledge**: experience discipline-specific data use cases in order to solve real-world problems of high complexity;
5. **Interpretation**: learn methods for effective data communication and visualization, and demonstrate their use in data representation;
6. **Social Value**: explore social and ethical implications of the use of data and technology.

****

Major and Minor Requirements

Major in Data Science: 46+ hours. CS 128, 128L; DATA 100, 122, 150, 200, 260, 470; EN300; MT288; 4 approved courses selected from one area of specialization; 2 electives selected from DATA 307, 322, 340, 350, 421, 422, 424;

Minor in Data Science: 22 hours. CS 128, 128L; DATA 100, 122, 150, 200. 2 electives selected from DATA 260, 307, 322, 340, 350, 421, 422, 424;

B.S. in Data Science - 4 courses selected from one of the following areas of specialization. Course list subject to change.		
Communications • CO201 Communications Research • CO225 Journalism • CO325 Investigative Reporting • CO346 Campaign Issues • CO315 Integrated Marketing Communications • CO360 IMC Research • CO455 Health & Environ. Writing	**Digital Humanities** • EN2xx Literature • EN311 Old English • EN312 Late Medieval Literature • EN488 History of the English Lang • EN498 Independent study	**Entrepreneurship** • ER201 Creativity, Innovation and Idea Development • ER301 Intro to Entrepreneurship • ER304 Social Entrepreneurship • ER305 Accounting & Finance for Entrepreneurs • ER306 Entrep. Marketing and Sales • ER480 Entrep. Field Experience
Exercise Science • EPA205/205L Human Anatomy • EPA206/206L Human Physiology • EPA230 Nutrition for Athletics and Physical Activity • EPA407 Exercise Physiology • EPA409 Kinesiology • EPA432 Motor Learning • EPA440 Independent Study	**Health Disparities&Social Justice** • BL155/157 Principles of Biology • BL156/158 Principles of Biology II • BL240 Epidemiology • BL260 Poverty & Disease • BL399 Special Problems Biology • PO160 Health Care and Social Justice in Latin America, and PO2xx Health Care Access in Latin America	**Physics** • PH135,135L Physics I • MT135 Calculus and Analytic Geometry I • PH136, 136L Physics II • MT136 Calculus and Analytic Geometry II
Political Science • PO200 Introduction to Methods • PO203 GIS I • PO300/L Research Methods & Lab • PO319 U.S. Elections • PO324 Crisis Mapping, New Media and Politics • PO337 Comparative Health Politics • PO399/498 Independent Study	**Psychology** • PS100 Introduction to Psychological Science • PS301/301L Experimental Design & Analysis • PS401/401L Research Methods • PS435 Tests & Measurements • PS499 Individual Research	**Sociology & Criminology** • SC101 Intro to Sociology • SC350 Sociological Res Methods I • SC351 Sociological Res Methods II • SC493 Independent Study • SC497 Undergraduate Research

100. INTRODUCTION TO DATA SCIENCE 3 cr. Data science capitalizes on big data and focuses on data analytics that turn information into actionable knowledge. This course will introduce students to the key ideas, practices, and challenges of modern data analysis. Students will get an overview of the data, questions, and tools that data scientists deal with in their practice. This course will introduce students to practical approaches to essential exploratory techniques, interactive data discovery, and predictive analytics including basic techniques for collecting, cleaning, and sharing data. Hands-on activities will enable students to learn the practical toolkit of a data scientist.

122. ELEMENTARY STATISTICS (MT 122) 3 cr. Describing data by graphs and measures, sampling distributions, confidence intervals and tests of hypotheses for one and two means and proportions, Chi-square tests, correlation and regression. Methods are illustrated in the context of quantitative research, with applications in disciplines such as sports, psychology, and social and natural sciences. Use of appropriate statistical software.

150. DATABASE SYSTEMS (CS 150) 3 cr. Introduction to relational database design and implementation. Topics include database systems concepts and architectures, structure query language (SQL), entity relationship (ER) modeling, relational database design, functional dependencies, and normalization.

200. INTERMEDIATE DATA SCIENCE 3 cr. Prerequisite: DATA 100, DATA/MT 122 or equivalent, CS 128. Pre or Co-requisite: DATA/CS 150. This course will provide a strong foundation in the field of data science and data analytics with a focus on computational approaches and experiential learning. Students will learn about processes and practice of data science that are developed to analyze diverse sources of data including data modeling, machine learning, and natural language processing. This course will present the fundamentals of inference in a practical approach. Students will build a portfolio of Big Data skills.

228. STATISTICS FOR BIOLOGICAL SCIENCES (MT 228) 3 cr. Exploratory data analysis, probability fundamentals, sampling distributions and the Central Limit Theorem, estimation and tests of hypotheses through one-factor analysis of variance, simple linear regression, and contingency tables using appropriate statistical software. Course content in biology context.

229. PROBABILITY AND STATISTICS (MT 229) 3 cr. Prerequisite: MT 135. Probability, discrete and continuous distributions, sampling distributions and the Central Limit Theorem, introduction to data analysis, estimation and hypothesis testing, simple linear regression and correlation; exact, normal-theory, and simulation-based inference; use of appropriate statistical software. Methods are illustrated in the context of quantitative research, with applications in disciplines including sports, psychology, and social and natural sciences.

260. INTERMEDIATE STATISTICS (MT 223) 3 cr. Prerequisite: DATA/MT 122 or MT228 or MT229 or EC208. Power analysis, factorial and repeated measures analysis of variance, nonparametric procedures, contingency tables, introduction to multiple regression. Use of appropriate statistical software.

307. BIOINFORMATICS (CS 307) 3 cr. Prerequisite: CS 228 or DATA 200. The application of computational methods and principles to solve data-intensive and pattern-discovery problems in biology, especially molecular and systems biology, without prior knowledge of college-level biology. Topics may include gene sequence assembly, sequence alignment, phylogenetic tree inference, gene expression, and protein interaction networks.

322. BIG DATA ANALYTICS (CS 322) 3 cr. Prerequisite: DATA/CS 150 and either CS 228 or DATA 200. What is "Big Data?" Data mining algorithms, machine learning algorithms. Emphasis on real analyses being performed every day by businesses, governments, and online social networks.

340. DATA VISUALIZATION (CS 340) 3cr. Prerequisite: CS 228 or DATA 200. Introduction to basic data visualization techniques. Discussion of different techniques to view data, and analysis of classic data representations. Students will use advanced tools for generating and exploring, static and dynamic visual representations of very large datasets.

350. ADVANCED DATABASE SYSTEMS (CS 350) 3 cr. Prerequisite: DATA/CS 150 and either CS 228 or DATA 200. Alternative data models and advanced database techniques, Big data support, Web-DBMS integration technology, data-warehousing and data-mining techniques, database security and optimization, and other advanced topics.

421. MATHEMATICAL STATISTICS (MT 421) 3 cr. Prerequisites: DATA/MT 229, MT 233. Moment generating functions, transformations, properties of estimators, foundations of hypothesis tests, one- and two-factor analysis of variance, and nonparametric analyses.

422. APPLIED STATISTICS (MT 422) 3 cr. Prerequisites: DATA 260/MT 223. Two factor analysis of variance; categorical data analysis, logistic regression, factor analysis, simulation, analysis of large datasets; use of appropriate statistical software.

424. APPLIED REGRESSION ANALYSIS (MT 424) 3 cr. Prerequisite: DATA 260/MT 223. Multiple linear regression, collinearity, model diagnostics, variable selection, nonlinear models; autocorrelation, time series, and forecasting; use of appropriate statistical software.

470. DATA SCIENCE PROJECT 3 cr. Prerequisite: DATA 200, DATA 260/MT 223, MT288, senior standing in the Data Science major and permission of instructor. Simulation of the environment of the professional data scientist working in a team on a large data project for a real client. Students will encounter a wide variety of issues that naturally occur in a project of scale, using their skills, ingenuity, and research abilities to address all issues and deliver a usable data product.