

# DATA SCIENCE (DATASCI)

---

## **DATASCI 2010 Data Science I 3 Credits**

This course will introduce students to the basics of programming using scripting languages (R and/or Python) in addition to developing competency with handling and analyzing datasets. The first third of the course focuses on learning the basics of coding through scripting languages as applied to exploratory data analysis (EDA). This will be accomplished through wrangling, analyzing, and summarizing/visualizing data. The second third of the course will focus on the continued development of programming and scripting skills through the creation of functions and algorithms in the form of simulation models. Stochastic and probabilistic methods will be incorporated to illustrate the use of uncertainty in decision making and data visualization. The third part of the course will focus on team project work and also examine different data types, data storage methods, and the use of documentation and organization of code, including version control, reproducibility, and peer review.

**Components:** Class, Laboratory

**Prereqs/Coreqs:** P. STAT 2030 with a grade of 'C-' or higher and (COMPUTER 1130 with a grade of 'C-' or higher or any COMPUTER class numbered higher than 1130 with a grade of 'C-' or higher) or consent of instructor

## **DATASCI 2510 Data Science II - Intermediate Data Science 3 Credits**

This course focuses on developing competency with advanced data analysis and visualization techniques. The first third of the course focuses on exploratory data analysis (EDA) applications using databases, advanced treatments of data cleaning, wrangling, and collection. The second third of the course focuses on statistical and inference skills. The third part of the course focuses on data visualization, including creation of dashboards and web-ready methods. While this course will use R, it will also focus on continuing to develop Python skills, including introducing the more important Python packages used in data science.

**Components:** Class, Laboratory

**Prereqs/Coreqs:** P. DATASCI 2010 with a 'C-' or higher

## **DATASCI 3010 Data Science Ethics 3 Credits**

This course investigates the ethics and justice of contemporary data science. Case studies are examined to determine whether data science models and their applications are ethical and just and it considers at least one new modeling strategy for ensuring that they are. This course will review several professional codes of ethics to see how ethical considerations are codified by professional organizations. It will then look at how these ethical considerations can be used to implement responsible data science throughout planning, data collection, analysis, and implementation of data science projects adding a level of "auditing" data science projects for ethical and socially just implementation. This course should be of interest to data scientists, statisticians, software engineers, managers, and all persons who are committed to responsible citizenship.

**Components:** Class

**Prereqs/Coreqs:** P. DATASCI 2010, STAT 3130 or MATH 4050 with a 'C-' or higher or consent of instructor

## **DATASCI 4900 Data Science Capstone 3 Credits**

Course under development at time of catalog publication.

**Components:** Class

**Prereqs/Coreqs:** P. Senior standing