Course Description:
This course is designed to provide students with an introduction to the key aspects of statistics and data science. It gives a general overview of the current status and future development of data science research and application and covers a broad spectrum of data analysis and collection methods, as well as the software implementation of those methods. The students would apply what they have learned through a capstone project designed by the instructor.

Enrollment: Undergraduate students majoring mathematics, statistics, or other similar fields.

Outlines:
1. Overview of Data Science
2. Fundamental Data Analysis Methods
3. Effective Data Collection Methods
4. Advanced Modeling Methods for Big Data
5. Capstone Project

Capstone Project
The instructor will design 3-5 project topics for the students, who can team up with each group of four members at most to work on one of the topics. The projects will be announced at the beginning of the course and the students can start work on it with the supervision of the instructor. At the end of the semester students will do oral presentations on the projects.
Modern Statistics in the Big Data Paradigm  
Instructor: Lulu Kang  
Department of Applied Mathematics, Illinois Institute of Technology

Course Description: This course is designed to provide students with an introduction to the key aspects of statistics and data science. It gives a general overview of the current status and future development of data science research and application and covers a broad spectrum of data analysis and collection methods, as well as the software implementation of those methods. The students would apply what they have learned through a capstone project designed by the instructor.

Enrollment: Undergraduate students majoring mathematics, statistics, or other similar fields.

References:

Software: R or MATLAB

Prerequisites: MATH 151, 152, 251, MATH 332, MATH 474, or MATH 475 and 476. Calculus I, II and III, Linear Algebra, Probability and Statistics.

Objective:
1. Students will learn modern statistical techniques for modeling and drawing inferences from large data sets,
2. Students will learn to use visual and numerical diagnostics to assess the soundness of their models,
3. Students will become familiar with the computational requirements and compromises to be made in analyzing large data sets, and
4. Students will gain experience in analyzing real data sets and communicating their results.
5. Students will learn how to implement and use these numerical methods in Matlab (or another similar software package) through the capstone project,
6. Students will improve their problem solving skills in computational mathematics,
7. Students will improve their presentation skills through the capstone project.

Lecture Schedule: TBA

Course Outlines:

1. Overview of Data Science (3 Hour)
   • Introduction of data science
   • Case studies of application of data science
   • Future development and application of data science

2. Fundamental Data Analysis Methods (10 Hour)
• Linear regression model
• Statistical inference of linear regression
• Model diagnostics
• Generalized linear model
• Classification methods

3. Effective Data Collection Methods (7 Hour)
• Design of experiments—Factorial Design
• Optimal experimental design
• Several subsampling methods
• Sequential modeling and design methods.

4. Advanced Modeling Methods for Big Data (10 Hour)
• Dimension reduction for supervised learning and non-supervised learning.
• Nonparametric regression models.
• Model selection methods for regression models.
• Deep Learning

5. Capstone Project

The instructor will design 3-5 project topics for the students, who can team up with each group of four members at most to work on one of the topics. The projects will be announced at the beginning of the course and the students can start work on it with the supervision of the instructor. In the end of the semester the students will present their work in the form of oral presentation.