

## MATH 571 DATA PREPARATION AND ANALYSIS

### Course Description

This course surveys industrial and scientific applications of data analytics, with case studies, including exploration of ethical issues. Students will learn how to prepare data for analysis, perform exploratory data analysis, and develop meaningful data visualizations. They will work with a variety of real world data sets and learn how to prepare data sets for analysis by cleaning and reformatting. Students will also learn to apply a variety of different data exploration techniques including summary statistics and visualization methods. 3-0-3

### Prerequisites:

CS 425 or equivalent, MATH 474 or equivalent

### Required Textbooks

*Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*, by Glenn J. Myatt

### Recommended Texts

- *The Visual Display of Quantitative Information*, by Edward R. Tufte
- *Visualizing Data: Exploring and Explaining Data with the Processing Environment*, by Ben Fry
- *Exploratory Data Mining and Data Cleaning*, by Tamraparni Dasu

### Detailed Course Topics

1. Defining data analysis problems
  - a. Knowing the client
  - b. Understanding the question(s)
2. Data gathering and preparation
  - a. Data formats, parsing, and transformation
  - b. Scalability and real-time issues
3. Data cleaning
  - a. Consistency checking
  - b. Heterogeneous and missing data
  - c. Data transformation and segmentation
4. Exploratory analysis
  - a. Descriptive and comparative statistics
  - b. Clustering and association
  - c. Hypothesis generation
5. Visualization
  - a. Designing visualizations
  - b. Time series
  - c. Geolocated data
  - d. Correlations and connections
  - e. Hierarchies and networks
  - f. Interactivity
6. Ethics in the profession
  - a. Cases in computing, statistics, and communication
  - b. Professional ethics codes: ACM, IEEE, Am Stat. Assoc.