DATA 1501 - Introduction to Data Science

Course Description
This course is intended to provide an introduction into the field of Data Science. Students will develop skills in appropriate technology and basic statistical methods by completing hands-on projects focused on real-world data and addresses the social consequences of data analysis and application.

Course Credit
3-0-3

Learning Outcomes

Required Outcomes for all Sections of the Course (should account for 70 – 80% of course content)

1. Explain the importance of and be able to formulate a data analysis problem statement that is clear, concise, and measurable.
2. Identify and appropriately acknowledge sources of data.
3. Be able to apply basic data cleaning techniques to prepare data for analysis.
4. Be able to identify the categorical and/or numerical data types in a given data set.
5. Apply appropriate descriptive and inferential methods to summarize data and identify associations and relationships.
6. Use appropriate tools and technology to collect, process, transform, summarize, and visualize data.
7. Be able to draw accurate and useful conclusions from a data analysis.
8. Effectively communicate methods and findings in a variety of modes.
9. Differentiate between ethical and unethical uses of data science.

Additional Optional Learning Outcomes (should account for 20 – 30% of course content)
1. Identify goals and methods of testing hypotheses.
2. Explain the bootstrap methods.
3. Identify legal issues surrounding the use of data.
4. Mine data to develop predictive models and evaluation.

Topics (70%-80% of course content):

What are data?
Sources of data, data collection and types of data
Sampling from a population
Data errors and appropriateness/Cleaning Data
The role of data in decision making at various levels of society

Methods of Data Analysis, including, but not limited to:
Distributions (including measures of central tendency and spread)
Expressions, names, and tables

Joins
Arrays
Functions
Modeling/mining the data

Using Computational Tools and Statistical Techniques for basic data manipulation

Interpreting results of the data analysis/Data Interpretation, possibly including, but not limited to the following:
Correlation
Chance
Decisions and error probabilities
Classification
Confidence intervals
Simulations
Empirical, Categorical, and Numerical Distributions
Assessing Models

Communicate data-driven insights in multiple media modes
Data visualization - (including graphs, charts, and histograms - univariate qualitative, univariate quantitative, bivariate)
Communication of the Data Science Findings and What It Means
Converting data into actionable information and the role of data in decision making at various levels of society

Ethical Aspects of Data Science
Accuracy
Misrepresentation
Privacy
Security

Additional topics (20%-30% of course content):
A/B Testing
Experiments
Hypothesis testing
Regression/Least squares
Prediction intervals
Inference for the true slope
Bootstrap
Bagging
Clustering
Frequent Patterns (Shopping Basket Analysis)
Information Retrieval
Anomaly Detection

Legal issues surrounding data
Causality and Experiments

Potential Textbooks

[The Foundations of Data Science](#) By Ani Adhikari and John DeNero, the OER that is currently used for the University of California - Berkeley Data 8 Course.

This is potentially one of many texts that would be appropriate for the course. Optimally, the text would be free or low-cost for students.