

SU DEPARTMENT OF COMPUTER SCIENCE SYLLABUS

COSC 311 – *Introduction to Data Visualization and Interpretation*

Description: Provides a foundation in the area of data science based on data curation and statistical analysis. Learn data analysis concepts and techniques that facilitate making decisions from a rich data set. Investigate data concepts, metadata creation and interpretation, machine learning, and basics of information visualization. Learn fundamentals about data standards and methods for organizing, curating, preserving data for reuse, drawing conclusions and making decisions from data. Understand how to use data analysis tools for data manipulation, analysis and visualization. Includes discussions on diverse issues around data, including technologies, behaviors, organizations, policies and society. Four hours per week.

Prerequisites: Grade of “C” or better in: COSC 120 and MATH 306

Required Text:

- ◁ Data Science from Scratch: First Principles with Python, By Joel Grus. ISBN: 978-1492041139.
- ◁ The Visual Display of Quantitative Information, 2nd edition, by Edward R. Tufte. ISBN: 978-0961392147.

References:

- ◁ Computational and Inferential Thinking: The Foundations of Data Science, online edition, by Ani Adhikari and John DeNero. <https://www.inferentialthinking.com>

	<i>Weeks</i>
<i>Python Programming Language Basics</i>	<i>2.0</i>
Jupyter, Variables, expressions, functions, classes, library usage, I/O	
<i>Data Processing and Organization</i>	<i>1.0</i>
Numerical tools, file processing, Pandas, Numpy	
<i>Tools for plotting and visualization</i>	<i>2.5</i>
Matplotlib, Scikit, categorical and numerical data, graph types and stylistic choices	
<i>Randomness and Sampling</i>	<i>1.5</i>
Simulating simple stochastic processes, synthetic data generation	
<i>Inferential Statistics</i>	<i>3.0</i>
Hypothesis testing, empirical distributions, estimators, non-parametric tools	
<i>Machine Learning Overview and Algorithms</i>	<i>3.0</i>
Linear methods, decision trees, clustering, sample complexity, supervised and unsupervised methods, Bayesian tools	
<i>Exams</i>	<i>1.0</i>

EVALUATION

Homework, labs, class participation: 20-30%

Projects and presentations: 40-60%

Exams and quizzes: 20