# EXPLORATORY DATA VISUALIZATION

# AND QUANTIFICATION

# STAT 4420/8426

**Course Description:**

Topics covered in this course include Exploratory Data Visualization for categorical/qualitative single/multivariate data, Grammar of Graphics, Organizing Data for Visualization, Methods of Displaying Data that include dynamic and interactive visualization, Visual Diagnostics of Statistical Models and Visual Statistical Inference. Students planning to enroll in this course should be comfortable with computer programming and have knowledge of data structures and preliminary statistical methods. **3 credits**

**Prerequisites:**

Undergraduate and Graduate: MATH 4750/8756 with a grade of C- or better or STAT 3800/8005 with a grade of C- or better or another introductory probability/statistics course with a grade of C- or better, and CSCI 1620 or equivalent with a grade of C- or better, or permission of instructor.

**Overview of content and purpose of the course:**

This course will help prepare students to be active citizens in the information technology age of the data-driven society. Students will develop critical thinking skills related to how information is visually presented, and they will learn how to accurately and attractively communicate quantitative information using statistical graphics.

This course explores exploratory data visualization using the grammar of graphics. It prepares students with a solid background in exploratory data analysis (EDA). Students will be able to work effectively in visualizing data while maintaining best ethical standards and avoiding misperception of data display.

**Anticipated audience/demand:**

Undergraduate or Graduate students in Statistics, Mathematics, Engineering, Computer Science or Business needing appropriate training on exploratory data analysis, dynamic and interactive data visualization and statistical modeling of data.

For the graduate students there will be additional requirements. They have to be engaged in a project with real data that produces substantial research interest or provides effective solution of a data problem.

**Major topics:**

1. Introduction
    - What is Exploratory Data Analysis (EDA)?
    - Visualizing Quantitative Information
    - History and Development of the terms
2. Grammar of Graphics
    - Building a statistical plot with a grammar
    - A layered grammar of graphics
        - Components
        - Hierarchy of plot defaults
3. Organizing Data for Visualization
    - Creating analysis data
    - Reshaping and recoding
    - Grammar of data manipulating (Merging, combining, sub-setting)
    - Handling missing data and dealing with outliers
4. Methods of displaying data
    - Data Display convention and evaluation
    - Tabular and textual display
    - Static display
        - Displaying categorical, quantitative or text data
        - Single, multivariate or repeated measure data
        - Geographic and temporal data
        - Big Data
    - Dynamic display
        - 2D or 3D Animation of visualization
        - Interactive visualization
    - Ethics in data display and visual (mis)perception
5. Visual Diagnosis of Statistical models
    - Linear models
        - Single of multiple covariates
        - Covariates with random coefficient
    - Time series and survival models.
    - Non-linear or generalized mixed models
    - Classification, trees, network graphs and other models
6. Visual Statistical Inference
    - Visual test and test statistic
    - Lineup
    - Visual p-value and decision rule
    -

**Methods:**

The class will be presented primarily in lecture form with student discussion encouraged. Questions are encouraged in class and out.

**Student role:**

Students must attend and participate in class in addition to completing course requirements. Students are expected to do reading and assignments as they are assigned.

**Textbook:**

Beautiful Visualization, edited by Julie Steele and Noah Iliinsky. O'Reilly Press: Sebastopol, CA, 2010.

January 13, 2016