

Introduction to Data Science Topic Outline

This outline describes only the scope of the course; the sequence is described in each unit.

I. Interpreting Data

- A. Types of data
- B. Numerical and graphical summaries
 - 1. Measures of center and spread, boxplots
 - 2. Bar plots
 - 3. Histograms
 - 4. Scatterplots
 - 5. Graphical summaries of multivariate data
- C. Simulation and visual inference
 - 1. Side-by-side bar plots and association
 - 2. Scatterplots
- D. Models
 - 1. Linear models
 - 2. k-means
 - 3. Smoothing
 - 4. Learning and tree-based models

II. Making Inferences and Justifying Conclusions

- A. Aggregating data
 - 1. Identification of sources
 - 2. Mechanics of Web 2.0
 - 3. Comparison of sources
- B. Data with special structures
 - 1. Random sampling
 - 2. Random assignment and A/B testing
 - 3. Simulation-based inference
- C. Participatory Sensing
 - 1. Designing a campaign
 - 2. Participation as a data collection strategy

III. Probability

- A. Computers and randomness
 - 1. Web services
 - 2. Pseudo-random numbers (optional)
- B. Frequency and probability
- C. Probability calculations

IV. Algebra in RStudio

- 1. Vectors
- 2. Algorithms
- 3. Functions
- 4. Evaluating and fitting models to data
- 5. Graphical representations of multivariate data
- 6. Numerical summaries of distributions and interpreting in context