Rationale

Many undergraduate data analytics courses focus on the mechanics of data analysis without first conveying fundamental data literacy. The result is rather like what would happen if we taught Hemingway, in English, to people who didn't speak English: their efforts to pick up the language along the way dramatically compound the challenges of absorbing the subject matter. This course conveys fundamental data literacy in the context of visualization, which provides obvious and engaging motivation. It then introduces students to the R statistical language, one of the most powerful tools in the world for statistical analysis. Finally, it shows them how to understand basic concepts in statistics and probability – uncertainty, generalizing from a sample to a population, etc. – using simulations and resampling. It is meant to provide both a foundation and a motivation for further quantitative training and research.

Learning objectives/outcomes

Students develop skills in finding and generating datasets, manipulating, summarizing, and visualizing data, critically evaluating arguments using data, and measuring the uncertainty of their conclusions.

The course will fulfill the General Education (GE) requirement in Data Analysis by helping students develop skills in drawing conclusions and critically evaluating arguments based on data. It will introduce students to basic concepts in statistics and probability, including sampling, data distributions, and the Central Limit Theorem, and it will teach students how to use iterated simulation and resampling (i.e., Monte Carlo simulation and bootstrapping) to obtain estimates of unknown probabilistic outcomes and to assign measures of accuracy to sample estimates. It will tie these elements together with the logic of research design in order to give students the ability to evaluate statistical arguments, and it will show them examples of how to do so using data on such topics as American partisanship and elections, drone strikes, and international conflict.