

Title: Hypothesize, Visualize, Analyze, Scrutinize, and Generalize: The Multi-Faceted Challenges of Examining Big Data

Presented by: Alexander Liss

Authors: Alexander Liss, Ruiruo Wu, and Elena Naumova, Department of Civil and Environmental Engineering, School of Engineering, Tufts University; Kenneth Chui, Department of Public Health and Community Medicine, School of Medicine, Tufts University

Abstract:

Big Data are emerging as an important source for public health studies. Its unprecedented richness in both quantitative and qualitative information allows for analytical breakthroughs in both breadth and depth. However, together with the immense benefits, Big Data present a significant set of challenges, ranging from increased demand to hardware and computational infrastructure to intensified validation and quality assessments. Using an analysis of the effects of extreme weather on hospitalizations of elderly in United States from 1991 through 2006 as an example, we highlight common challenges researchers need to be cognizant of when working with Big Data.

We propose three basic principles: i) maintain the theory-driven nature of the analysis in order not to be lost in the vastness of information, ii) actively integrate data visualization and analysis to guide the understanding of temporal, spatial and contextual relationships, and iii) draw mindful and critical conclusions that are applicable and relevant to the target audiences. We will list statistical methodology available for the researchers utilizing Big Data as well as free and user-friendly software for data analysis and visualization, provide tips for data validation and quality assurance, and offer suggestions on handling increasing technical requirements.