Supporting Reproducible Exploratory Data Analysis

David Koop

Computer and Information Science Department
University of Massachusetts Dartmouth

1:00 p.m., Tuesday, February 5, 2019

Abstract

As computational data analysis and visualization tools permeate more fields, the ability to generate results has often outpaced the ability to meaningfully record and reflect on them. Interactive visual applications on the web allow users to explore data and gain insight, but these findings are often ephemeral, lost when a new page is loaded. Computational notebooks combine cells of code with their outputs and explanatory text, and have become key tools in exploratory data analysis. However, the ability to modify and reorder cells often produces unknown or ambiguous dependencies, leading to problems with reproducing past results. In both settings, being able to rapidly explore data seems to be in tension with being able to reproduce and extend that work. My research aims to develop methods that improve the reproducibility of interactive web applications and computational notebooks without impeding the fast and flexible analyses these environments facilitate. Furthermore, making individual explorations more robust and reproducible can often enhance collaboration, allowing others to understand and extend that work.

Bio

David Koop is an Assistant Professor in the Computer and Information Science Department at UMass Dartmouth. His research interests include data visualization, computational provenance, and geographic data analysis. A focus of his research is on methods that support users in data exploration, analysis, and visualization tasks so they can focus on important ideas and decisions. During his work, he has collaborated with scientists in the fields of climate science, quantum physics, and invasive species modeling. He received his Ph.D. in Computing from the University of Utah in 2012.