Guest Editors' Introduction

# Visualization in Data Science

**Adam Perer**
Carnegie Mellon University

**Shixia Liu**
Tsinghua University

**Visualization is an** integral part of data science and essential to enable exploratory analysis of data. This Special Issue of *IEEE Computer Graphics and Applications* highlights six papers from the Symposium on Visualization in Data Science held in Berlin, Germany in October 2018. These six papers demonstrate how visualization can contribute to advances in data science.

Transformations in many fields are enabled by rapid advances in our ability to acquire and generate data. The bottleneck to discovery is now our ability to analyze and make sense of heterogeneous, noisy, streaming, and often massive datasets. Extracting knowledge or insights from this abundance of data lies at the heart of 21st century discovery, which can be used to inform decisions, coordinate activities, optimize processes, improve products and services, as well as enhance productivity and innovation across a wide range of business and scientific problems.

Data science is the practice of deriving insights from data, enabled by statistical modeling, computational methods, interactive visual analysis, and domain-driven problem solving. Data science draws from methodology developed in such fields as applied mathematics, statistics, machine learning, data management, visualization, and HCI. It drives discoveries in business, economy, biology, medicine, environmental science, the physical sciences, the humanities and social sciences, and beyond.

*Digital Object Identifier 10.1109/MCG.2019.2925493*
*Date of current version 21 August 2019.*

## IN THIS ISSUE

After three highly successful events, the fourth Symposium on Visualization in Data Science (VDS) was held at IEEE VIS 2018 in Berlin, Germany. VDS brought together hundreds of domain scientists and methods researchers (including visualization, usability and HCI, data management, statistics, machine learning, and software engineering) to discuss common interests, talk about practical issues, and identify open research problems in VDS. This Special Issue features extended papers from six submissions to VDS 2018, which we encouraged authors to cross disciplinary boundaries and show how VDS can be a catalyst for discovery.

In "BEAMES: Interactive Multimodel Steering, Selection, and Inspection for Regression Tasks," the authors present a technique that allows users to inspect and steer multiple machine learning models tailored to their domain and tasks. This technique demonstrates an advance over prior visualization systems that typically support steering a single model.

In "Data2Vis: Automatic Generation of Data Visualizations Using Sequence-to-Sequence Recurrent Neural Networks," the authors are the first to automatically generate visualizations by applying

deep neural translation to a set of visualization examples. The authors demonstrate that Data2Vis can generate visualizations comparable to user-generated ones in a fraction of the time.

In "DeepCompare: Visual and Interactive Comparison of Deep Learning Model Performance," the authors present a novel visual analytics approach to compare the results of deep learning models. The authors demonstrate how such a tool can help users evaluate model results, understand model behavior, and assess tradeoffs between competing models.

In "Tackling Similarity Search for Soccer Match Analysis: Multimodal Distance Measure and Interactive Query Definition," the authors present a visual search system to interactively identify similar contextual enhanced soccer moves in a dataset. The authors demonstrate how such a system allows users to influence the relevance of the results according to their needs.

In "Uncertainty-Aware Visualization for Analyzing Heterogeneous Wildfire Detections," the authors present a visualization approach for interpolating multiresolution, uncertain satellite detections of wildfires into intuitive visual representations. The authors demonstrate that such a tool provides a shared communication method being useful for both data scientists and emergency managers.

In "Visual Analytics of Volunteered Geographic Information: Detection and Investigation of Urban Heat Islands," the authors develop a visual analytics system for urban heat islands. The authors demonstrate how the developed tool helps contextualize the emergence of urban heat islands and facilitates the comprehension of the influencing causes and their effects.

## CONCLUSION

We hope this Special Issue has demonstrated how important visualization research is in the data science pipeline. However, there is still a large amount of work to do to reach the complex challenges of data scientists. We hope to witness some of these breakthroughs when the Symposium on VDS continues in its fifth iteration in October 2019, when it is held at IEEE VIS in Vancouver, Canada.

**Adam Perer** is an Assistant Research Professor with Carnegie Mellon University, Pittsburgh, PA, USA, where he is a member of the Human-Computer Interaction Institute. His research integrates data visualization and machine learning techniques to create visual interactive systems to help users make sense out of big data. Lately, his research focuses on human-centered data science and extracting insights from clinical data to support data-driven medicine. He received the Ph.D. degree in computer science from the University of Maryland, College Park, MD, USA. He has authored or coauthored more than 30 refereed papers in the leading journals and conferences such as IEEE TVCG, IEEE InfoVis, IEEE VAST, CHI, CSCW, and AMIA. He was the Paper Co-Chair in 2018 and Symposium Co-Chair in 2019 of IEEE Visualization in Data Science (VDS).

**Shixia Liu** is a tenured Associate Professor with the School of Software, Tsinghua University, Beijing, China. She has authored or coauthored more than 40 refereed papers in the leading journals and conferences such as IEEE TVCG, IEEE TKDE, IEEE InfoVis, IEEE VAST, KDD, and WWW. She is an Associate Editor-in-Chief of IEEE Transactions on Visualization and Computer Graphics and *Information Visualization* and an Associate Editor of IEEE Transactions on Big Data and *ACM Transactions on Interactive Intelligent Systems*. She was the Papers Co-Chair of IEEE VAST 2016 and 2017 and was the Program Co-Chair of PacifcVis 2014 and VINCI 2012. She was the Symposium Co-Chair of IEEE Visualization in Data Science in 2018 and 2019.