

# Foresight: Recommending Visual Insights

Çağatay Demiralp  
IBM Research  
cagatay.demiralp@us.ibm.com

Srinivasan Parthasarathy  
IBM Research  
spartha@us.ibm.com

Peter J. Haas  
IBM Research  
phaas@us.ibm.com

Tejaswini Pedapati  
IBM Research  
tejaswinip@us.ibm.com

## ABSTRACT

Current tools for exploratory data analysis (EDA) require users to manually select data attributes, statistical computations and visual encodings. This can be daunting for large-scale, complex data. We introduce Foresight, a system that helps the user rapidly discover visual insights from large high-dimensional datasets. Formally, an “insight” is a strong manifestation of a statistical property of the data, e.g., high correlation between two attributes, high skewness or concentration about the mean of a single attribute, a strong clustering of values, and so on. For each insight type, Foresight initially presents visualizations of the top  $k$  instances in the data, based on an appropriate ranking metric. The user can then look at “nearby” insights by issuing “insight queries” containing constraints on insight strengths and data attributes. Thus the user can directly explore the space of insights, rather than the space of data dimensions and visual encodings as in other visual recommender systems. Foresight also provides “global” views of insight space to help orient the user and ensure a thorough exploration process. Furthermore, Foresight facilitates interactive exploration of large datasets through fast, approximate sketching.

## KEYWORDS

Exploratory data analysis, guided data exploration, recommendation, statistical, insight, visualization, sketching, scalable analytics

**This paper was previously accepted for publication at the demo track of VLDB’17. The original draft can be accessed at: <https://arxiv.org/pdf/1707.03877.pdf>.**