# Feature-Based Visualization of Multivariate Data

***Daniel Engel***
*International Graduate School (IRTG) 1131*
*Department of Computer Science, University of Kaiserslautern, Germany*
*d_engel@cs.uni-kl.de*

**Abstract:**
Simulation and experimental data acquisition enable scientists and engineers to gather vast amounts of data. Thereby, more and more application domains are producing progressively larger and inherently more complex (multivariate) data sets. Visualization, the first and foremost step of exploratory data analysis, is presented acute challenges by multivariate data because its full depiction requires more degrees of freedom than what is feasible for physical display devices and human perception. Consequently, its full depiction is an idealistic goal. Due to this fact, the challenge for visualization thereby amounts to finding the optimal low-rank approximation of multivariate data that is depictable and comprehensible to the user, while conveying as much information within the data as possible.

Working closely with application scientists and engineers, we believe that it will be possible to identify core features within multivariate data sets that can be utilized to abstract, classify, and depict this data. These features are chosen such that they represent lower-dimensional, physically meaningful, and independent bases for the data. Our research aims to amalgamate domain expertise from key disciplines such as: statistics, machine learning, data mining, and visualization; to develop fundamental mathematical foundations that are required to support interactive visual data exploration.

This thesis addresses the development of novel techniques for feature-based visualization of multivariate data in interactive speed. The first step is to utilize multi-criteria optimization methods in order to extract independent data features that act as suitable lower-dimensional data bases. Subsequently, the data is classified using unsupervised machine learning methods and visualized by lower-dimensional abstractions. The above-mentioned techniques are to be integrated in an analytical framework for data exploration, such that it makes a significant contribution in the application area.

Several potential applications could benefit from these techniques, one such application is the investigation of air pollution. Atmospheric particles increase morbidity in our cities and cause serious implications for climate change. Innovative instruments are now capable of chemically analyzing these particles in real-time, providing the groundwork for air quality research. In order to gather reliable measurements, one such instrument typically analyzes hundreds of thousands of aerosols, each measured in over 250 attributes. As of now, no data analysis techniques are able to depict these results in a comprehensible manner.

*Advisor: Prof. Dr. Hans Hagen*