

# Tutorial in visualization

## Visualization & Data Mining for High-Dimensional Datasets

### Prof. Alfred Inselberg

---

**What?** A short course, open for everybody and for free. However, all participants (except those already registered for SIGRAD 2012) have to register at <http://cs.lnu.se/sigrad2012/index.php/tutorial>

**When?** Friday 30 November 2012 at 1.15-4.15 pm  
**Where?** Linné, building H, Linnaeus University, Växjö  
**Who?** Professor Alfred Inselberg, Computer Science and Applied Mathematics Departments, Tel Aviv University, Israel

**Abstract** A dataset with  $M$  items has  $2^M$  subsets any one of which may be the one fulfilling our objectives. With a good data display and interactivity our fantastic pattern-recognition can not only cut great swaths searching through this combinatorial explosion, but also extract insights from the visual patterns. These are the core reasons for data visualization. With parallel coordinates (abbr. ||-cs) the search for relations in multivariate datasets is transformed into a 2-D pattern recognition problem. The foundations are developed interlaced with applications. Guidelines and strategies for knowledge discovery are illustrated on several real datasets (financial, process control, credit-score, intrusion-detection etc) one with hundreds of variables. A geometric classification algorithm is presented and applied to complex datasets. It has low computational complexity providing the classification rule explicitly and visually. The minimal set of variables required to state the rule (features) is found and ordered by their predictive value. Multivariate relations can be modeled as hypersurfaces and used for decision support. A model of a (real) country's economy reveals sensitivities, impact of constraints, trade-offs and economic sectors unknowingly competing for the same resources. An overview of the methodology provides foundational understanding; learning the patterns corresponding to various multivariate relations. These patterns are robust in the presence of errors and that is good news for the applications. We stand at the threshold of breaching the gridlock of multidimensional visualization.

The parallel coordinates methodology has been applied to collision avoidance and conflict resolution algorithms for air traffic control (3 USA patents), computer vision (1 USA patent), data mining (1 USA patent), optimization, decision support and elsewhere.



 **Linnæus University**  
School of Computer Science, Physics and Mathematics

**Lnu.se**