

# Interactive Visual Data Exploration with Subjective Feedback — **SIDER**

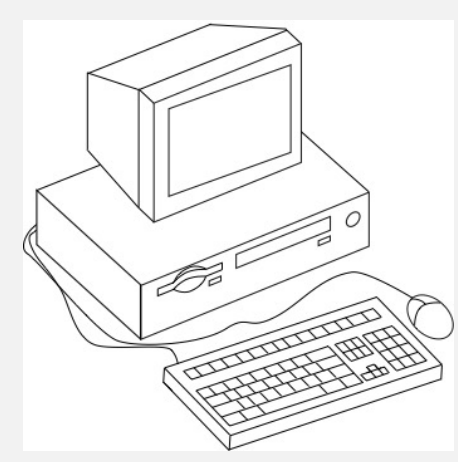
Kai Puolamäki<sup>ab</sup>, Emilia Oikarinen<sup>ab</sup>, Bo Kang<sup>c</sup>, Jeffrey Lijffijt<sup>c</sup>, Tijn De Bie<sup>c</sup>

<sup>a</sup> Department of Computer Science, Aalto University, Helsinki, Finland

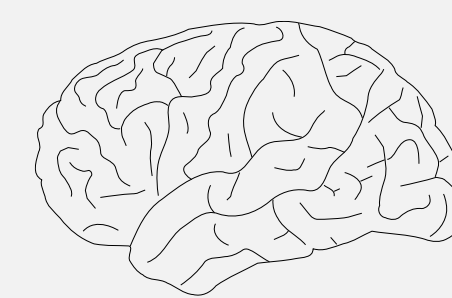
<sup>b</sup> Finnish Institute of Occupational Health, Helsinki, Finland

<sup>c</sup> Data Science Lab, Ghent University, Ghent, Belgium

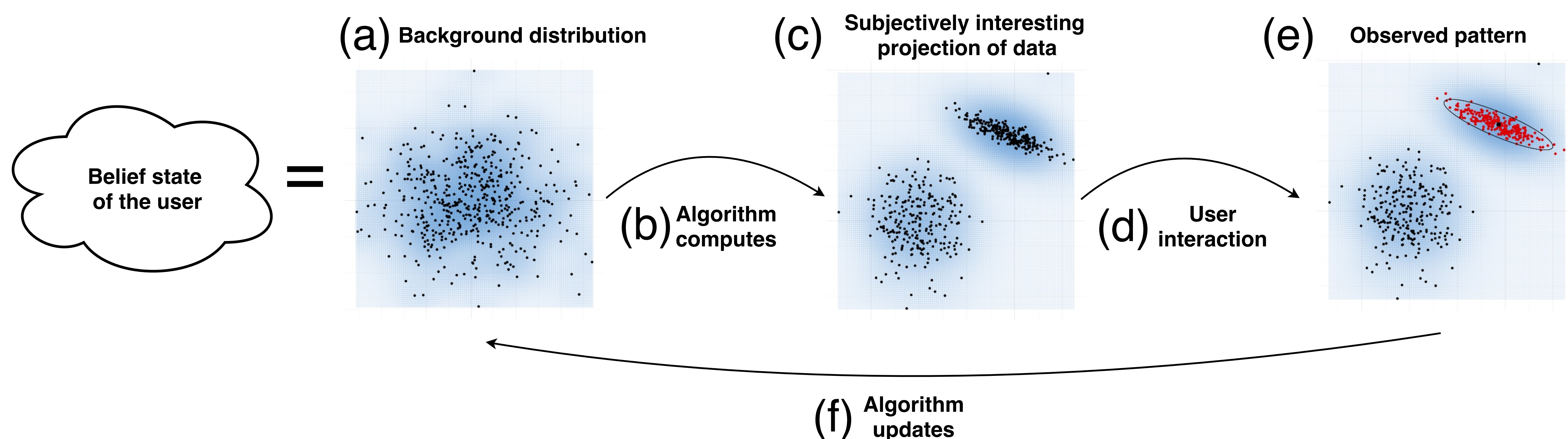
**Summary** We introduce a novel generic method for interactive visual exploration of high-dimensional data. We construct a theoretical model where identified patterns can be input as knowledge to the system. This background knowledge is used to find a **Maximum Entropy distribution of the data**, after which the system provides the user **data projections** in which the data and the Maximum Entropy distribution differ the most, hence showing the user aspects of the data **maximally informative** given the user's current knowledge.



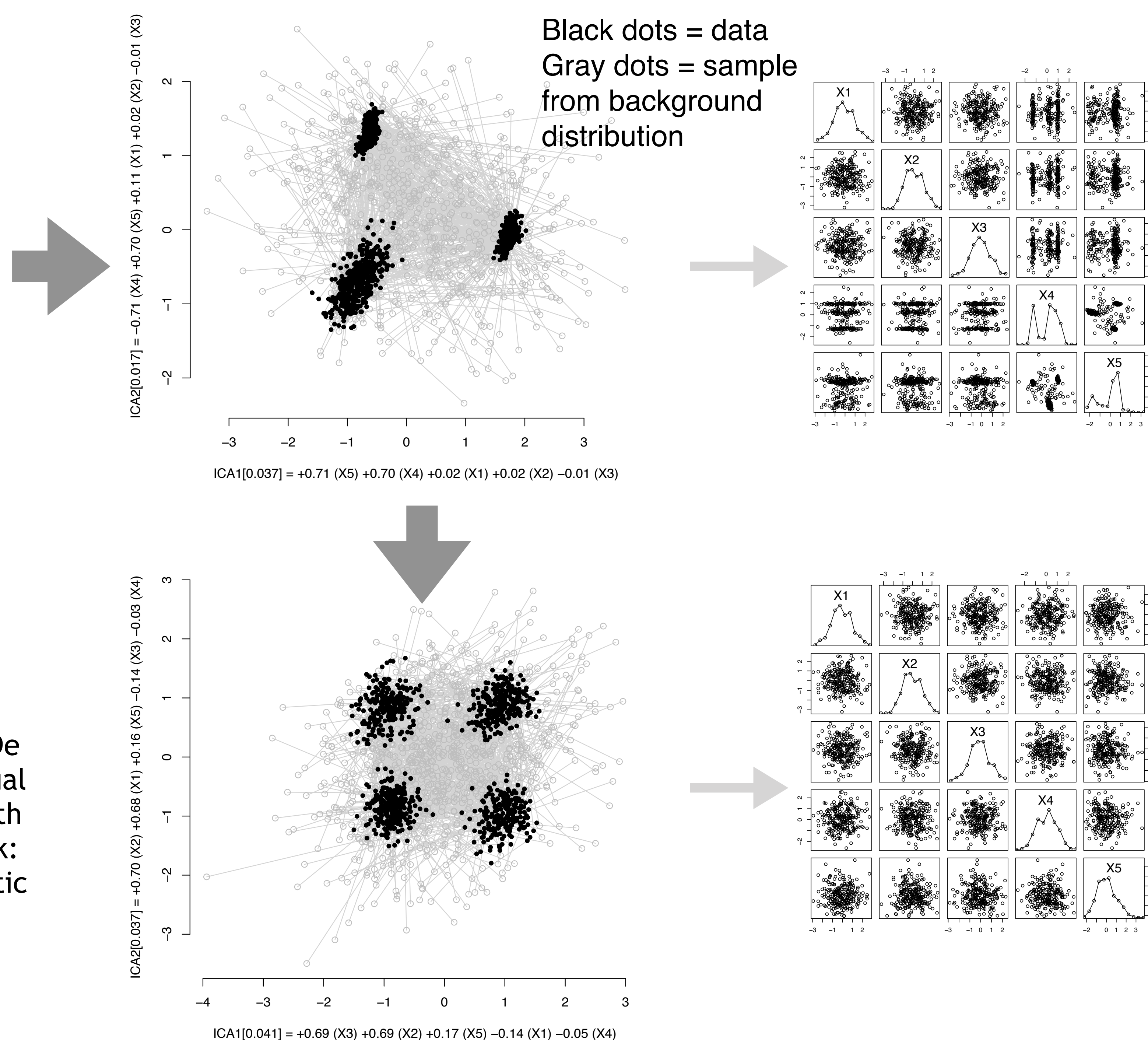
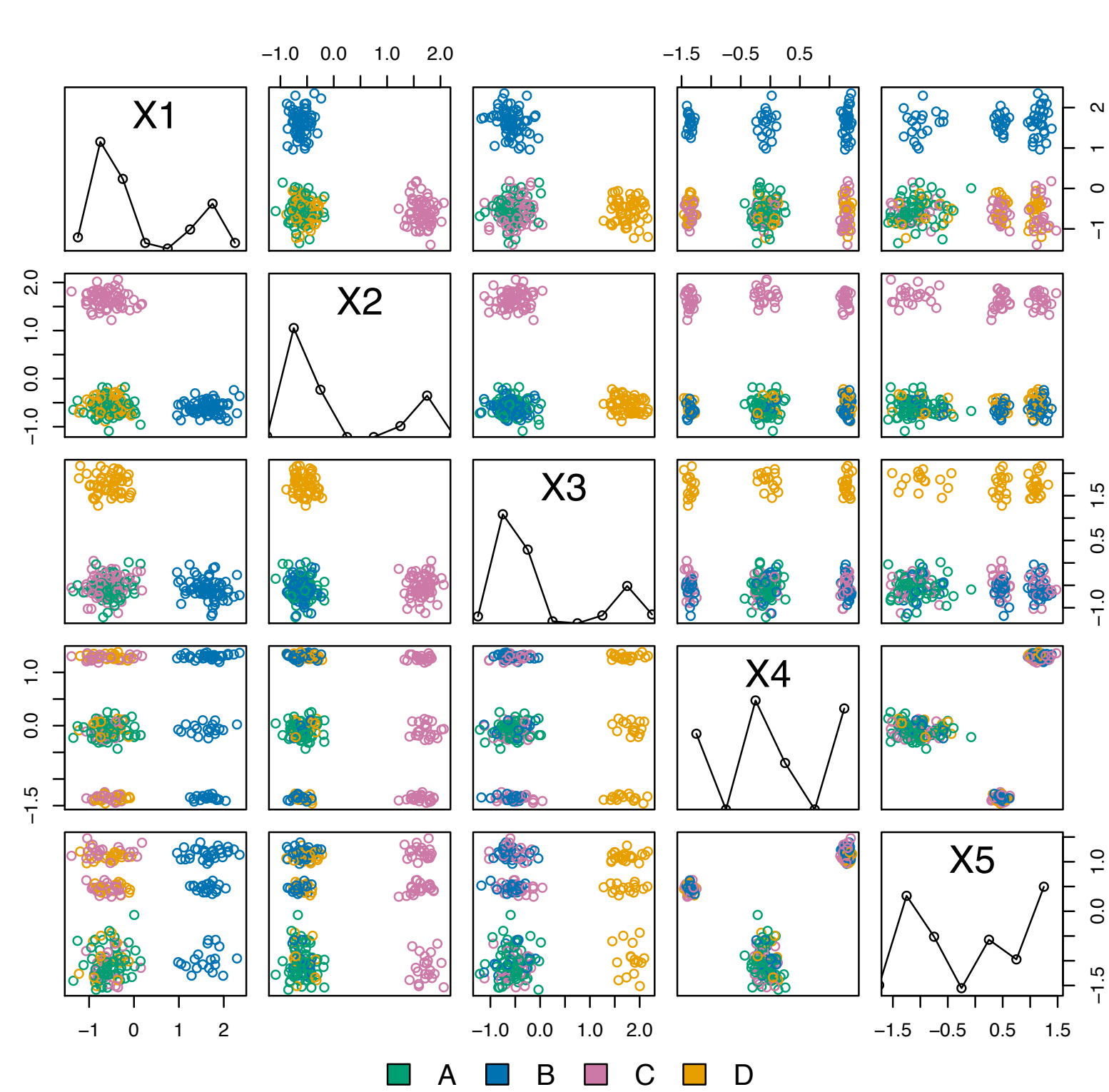
- + handling large data
- + handling high-dimensional data
- + making analytic comparisons
- identifying patterns truly relevant for the user
- black boxes, incomprehensible for the user



- + huge background knowledge
- + spotting patterns
- handling large, high-dimensional data
- making analytic comparisons



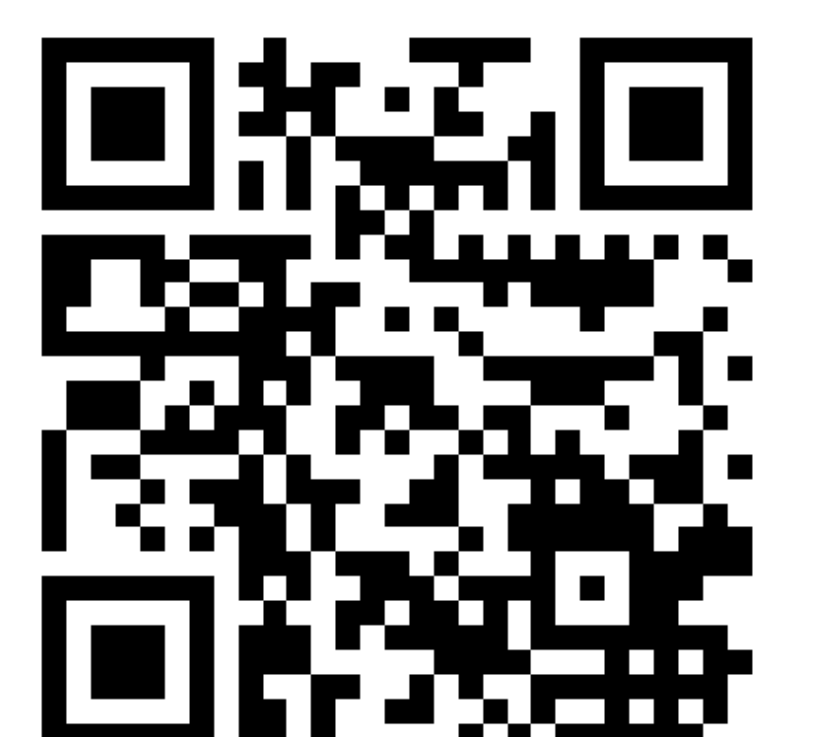
- Data = **real vectors**
- Background distribution (a) = **Maximum Entropy distribution satisfying constraints** (initially: no constraints, unit Gaussian spherical distribution with zero mean)
- Direction-preserving **whitening** transformation of the data results in a unit Gaussian spherical distribution, if the data follows the current background distribution
- **PCA/ICA** used to find non-Gaussian directions: **subjectively interesting projection of data (b,c)**
- User observes **patterns** and adds respective **constraints (d,e)**
- Background distribution is updated (f); the process is **iterative**
- Various constraints based on simple **linear and quadratic constraints**



Whitened data after adding cluster constraints for the visible clusters



For details, see: Kai Puolamäki, Emilia Oikarinen, Bo Kang, Jeffrey Lijffijt, Tijn De Bie. Interactive Visual Data Exploration with Subjective Feedback: Information-Theoretic Approach, arXiv: 1710.08167, 2017



Download for the **SIDER** tool