Information Visualization

High-Dimensional Data

Dr. David Koop

Schedule

- Today: High-Dimensional Data Lecture
- Tuesday, Oct. 26: No Class
- Thursday, Oct. 28: High-Dimensional Data Critique Due

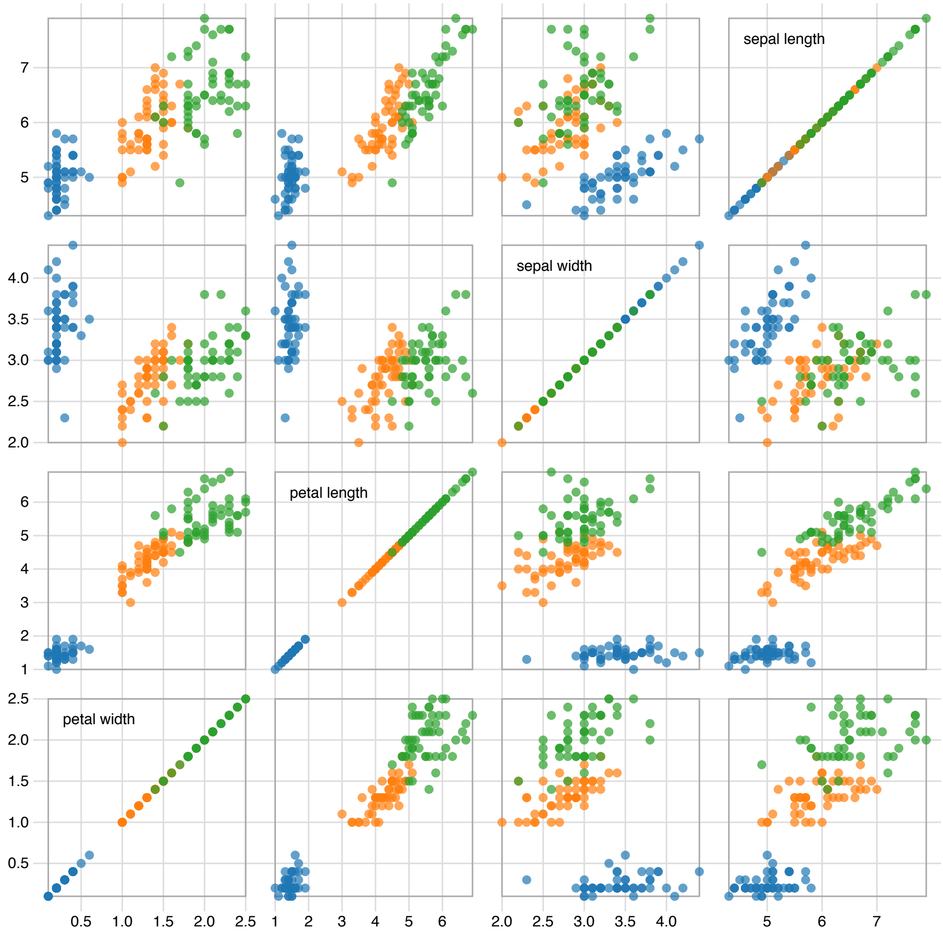
What techniques might you use for high-dimensional data visualization?

High-Dimensional Data Visualization Techniques

- Scatterplot Matrix (SPLOM)
- Parallel Coordinates Plot (PCP)
- Heatmap
- Interactive Elements:
  - Brushing (Linked Highlighting)
  - Tooltips

Scatterplot Matrix

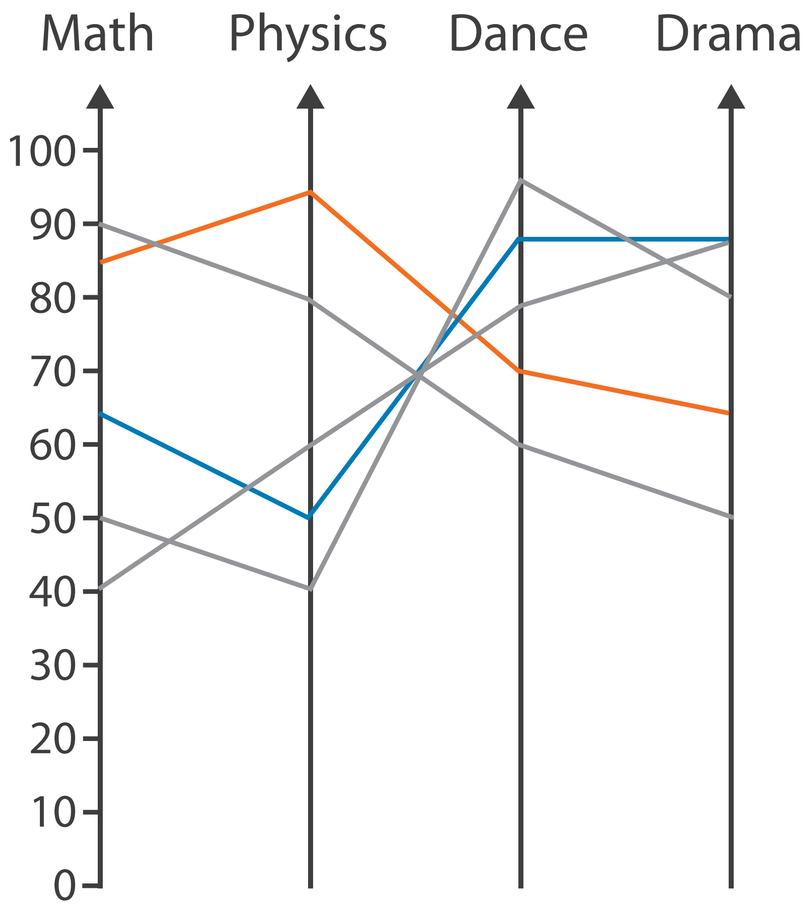
- Each pair of quantitative attributes has its own plot



[M. Bostock]

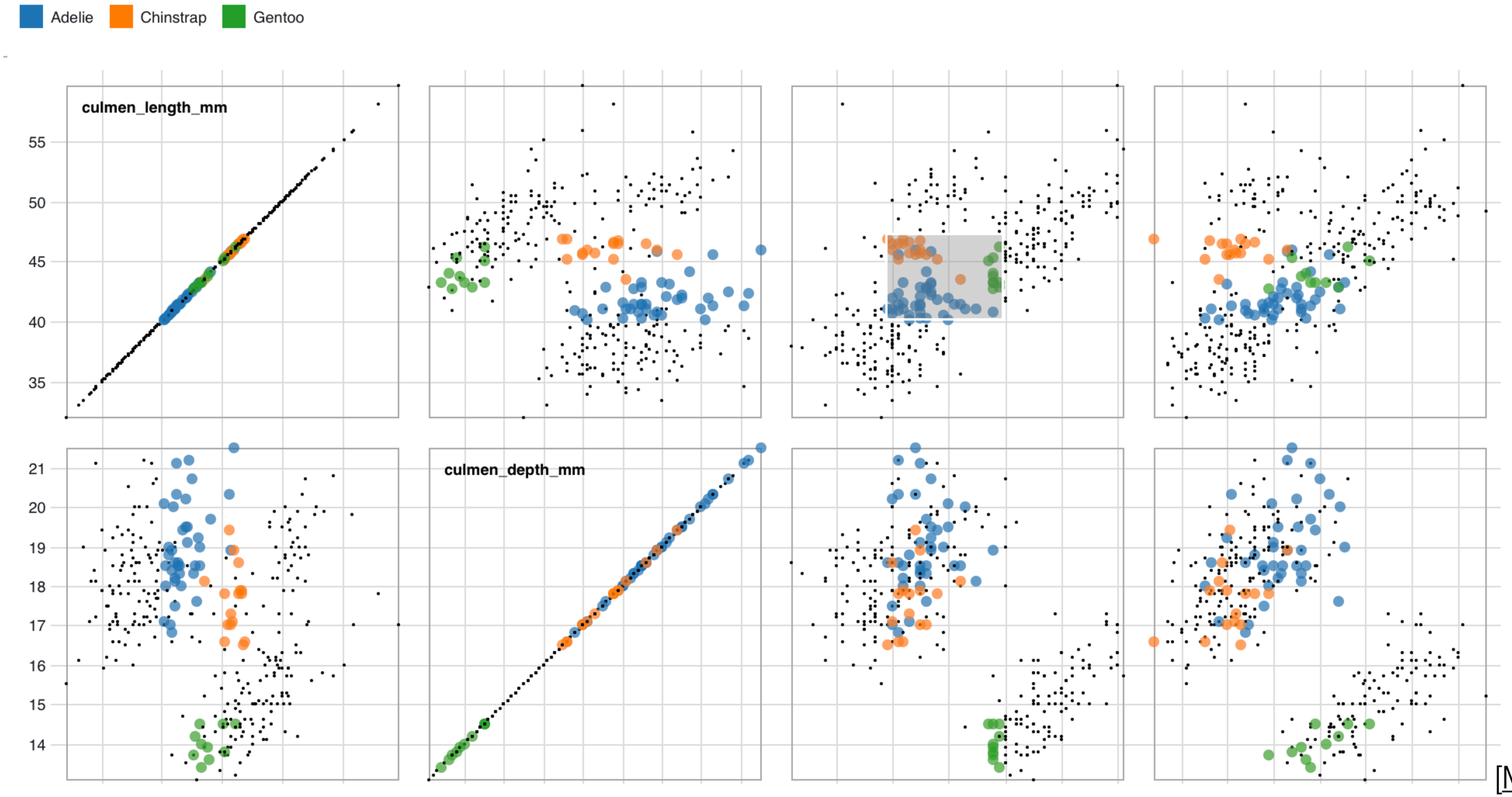
Parallel Coordinate Plots

- Use multiple parallel axes, one for each dimension
- Each data item is encoded as a line mark
- Positive and negative correlation can be seen in these plots…
- …but ordering becomes important



[Munzner (ill. Maguire), 2014]

Brushing



[M. Bostock]

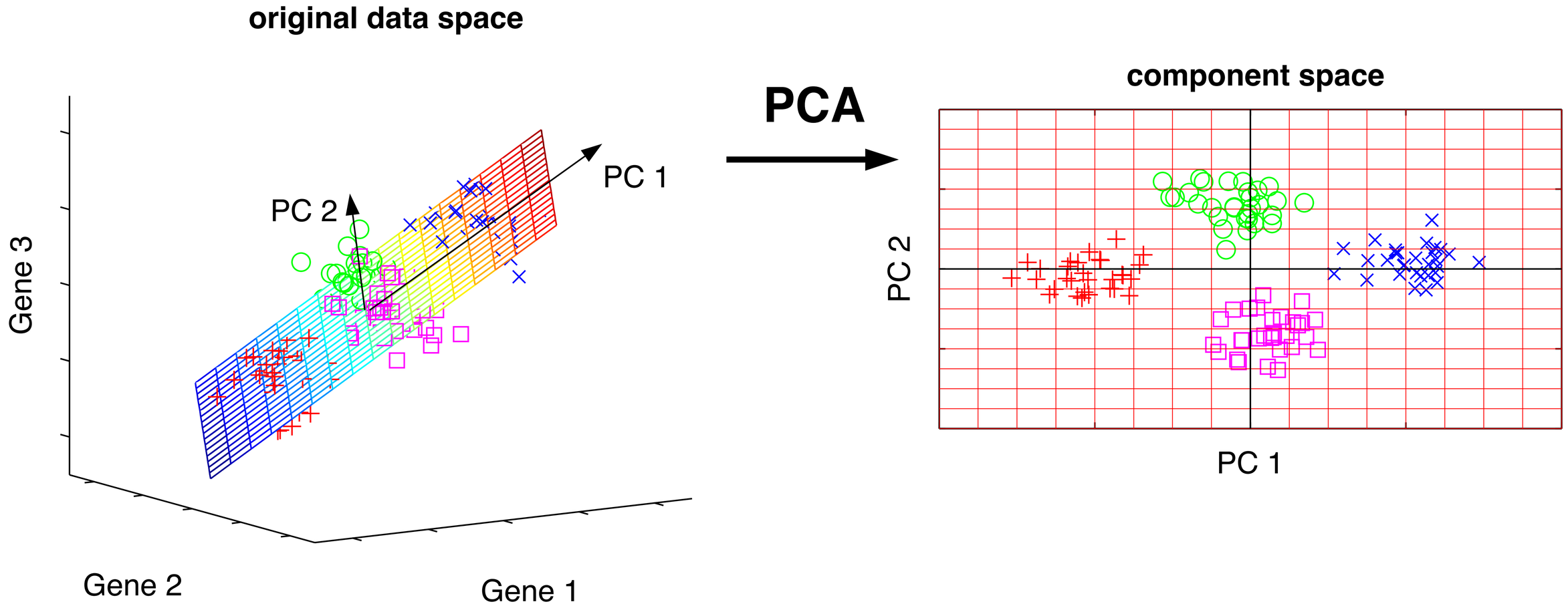
Our High-Dimensional Data Focus

- Projection Understanding
- Tours

Dimensionality Reduction

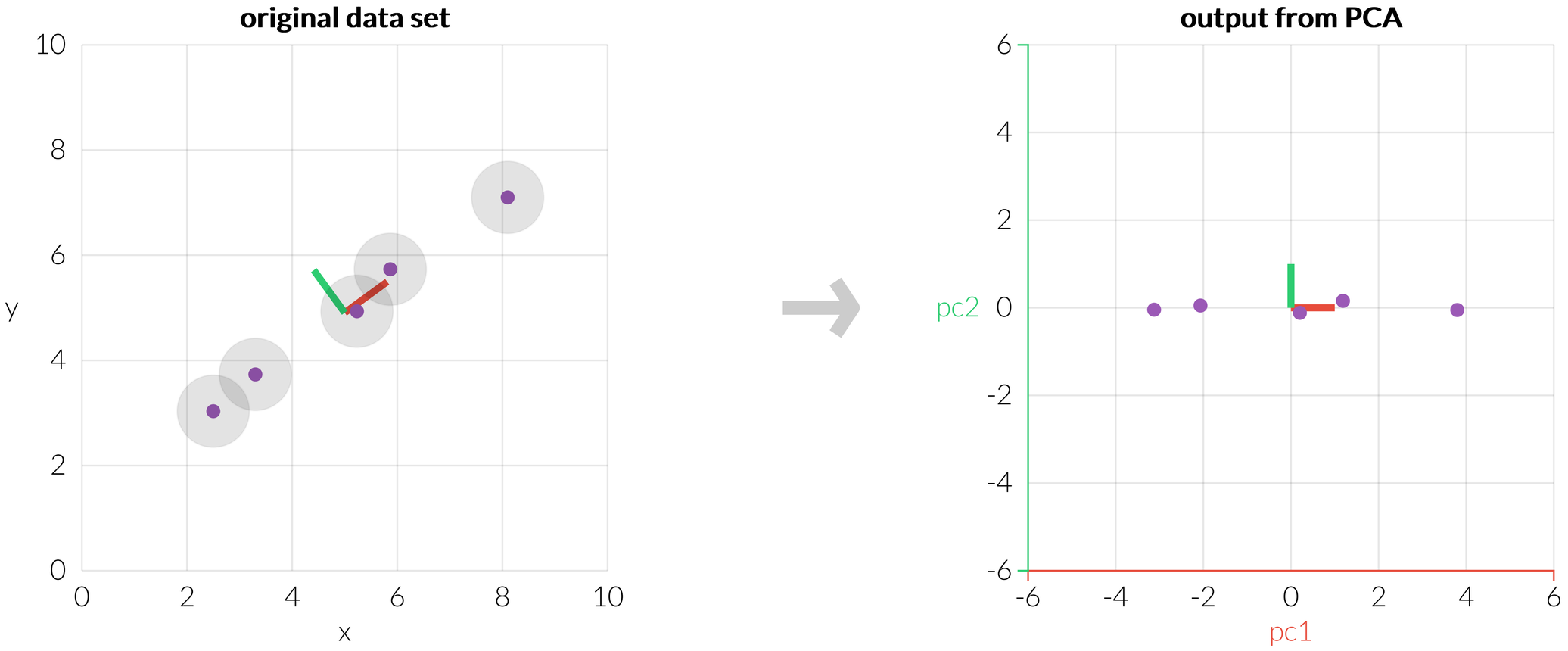
- Attribute Aggregation: Use fewer attributes (dimensions) to represent items
- Combine attributes in a way that is more instructive than examining each individual attribute
- Example: Understanding the language in a collection of books
  - Count the occurrence of each non-common word in each book
  - Huge set of features (attributes), want to represent each with an aggregate feature (e.g. high use of "cowboy", lower use of "city") that allows clustering (e.g. "western")
  - Don't want to have to manually determine such rules
- Techniques: Principle Component Analysis, Multidimensional Scaling family of techniques

Principle Component Analysis (PCA)



[M. Scholz, CC-BY-SA 2.0]

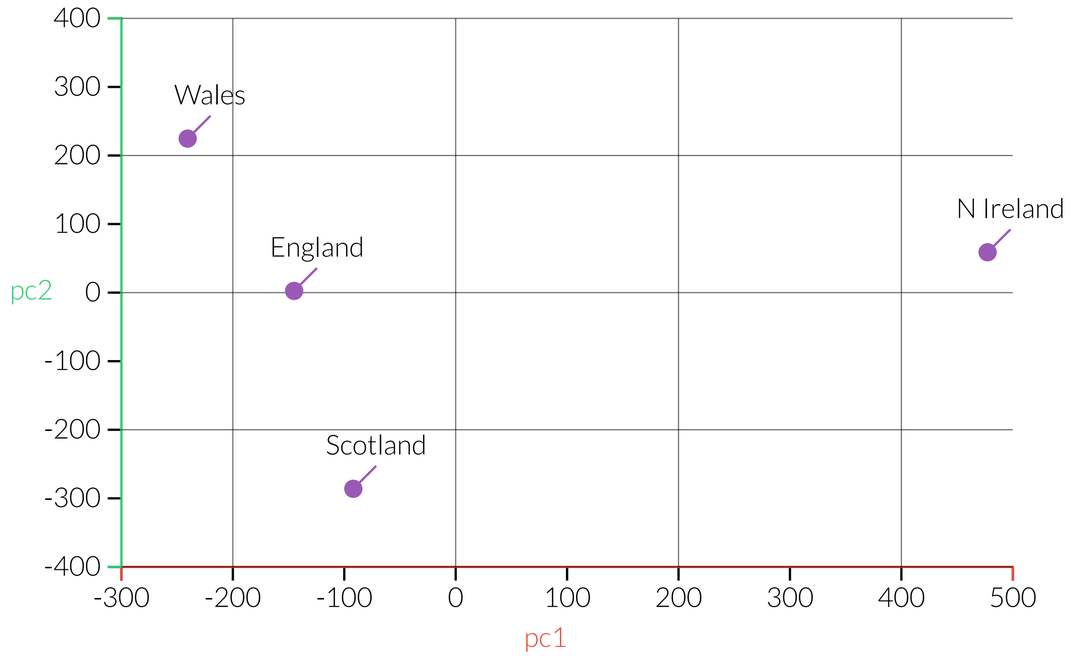
PCA



[Principle Component Analysis Explained, Explained Visually, V. Powell & L. Lehe, 2015]

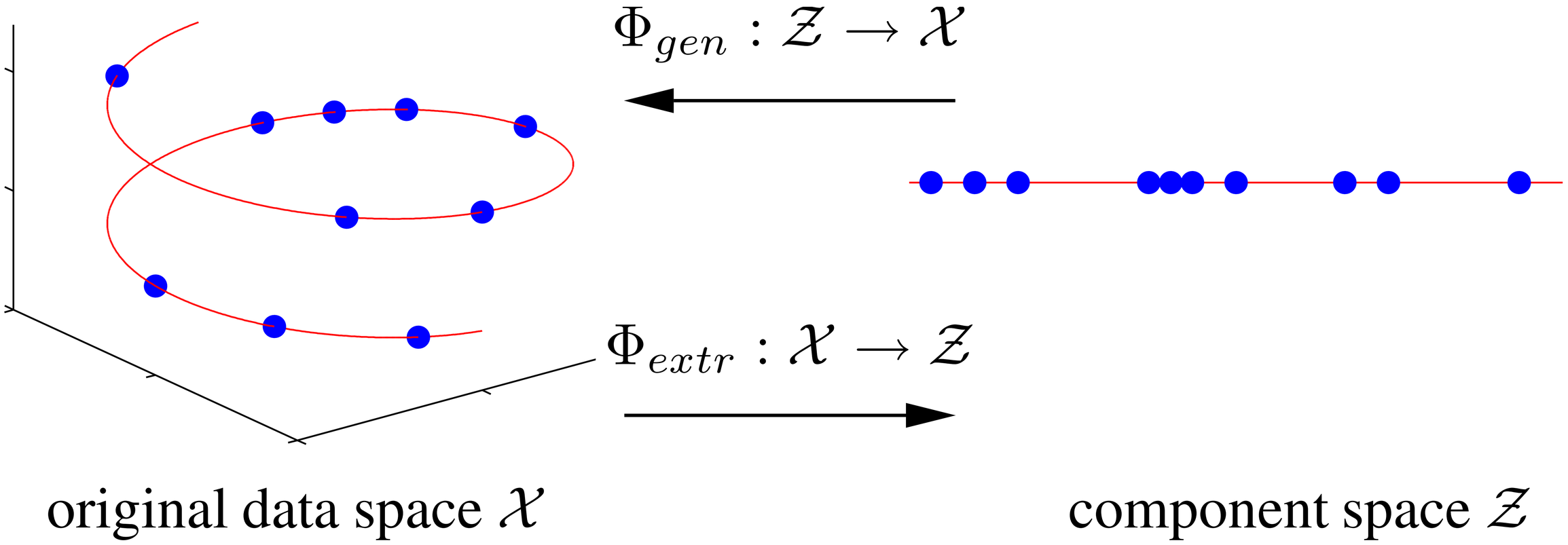
17 dimensions to 2

	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175



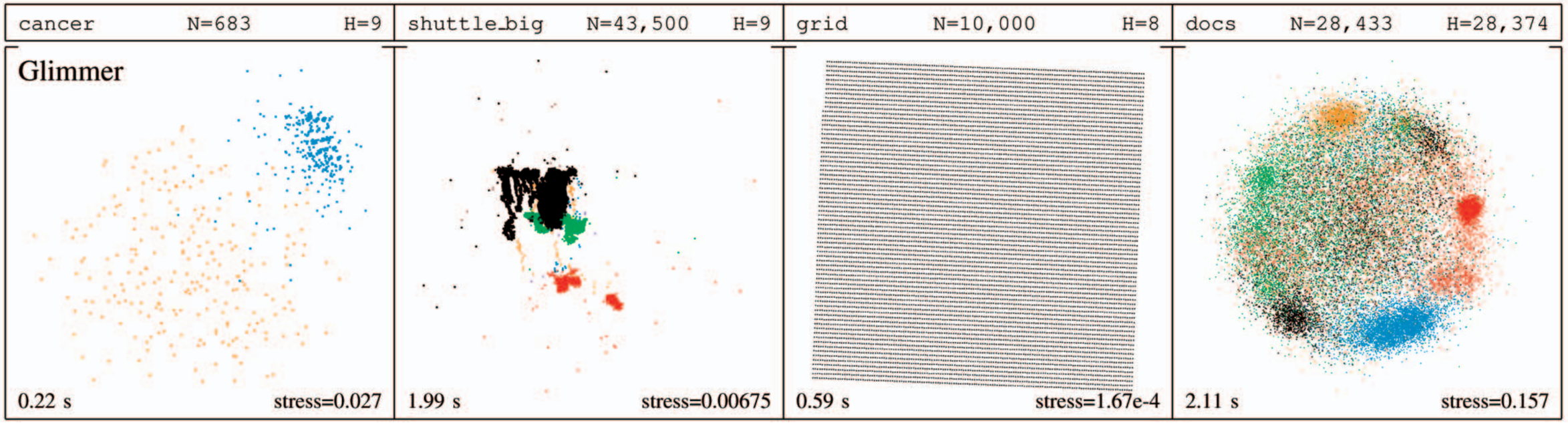
[Principle Component Analysis Explained, Explained Visually, V. Powell & L. Lehe, 2015]

Non-linear Dimensionality Reduction



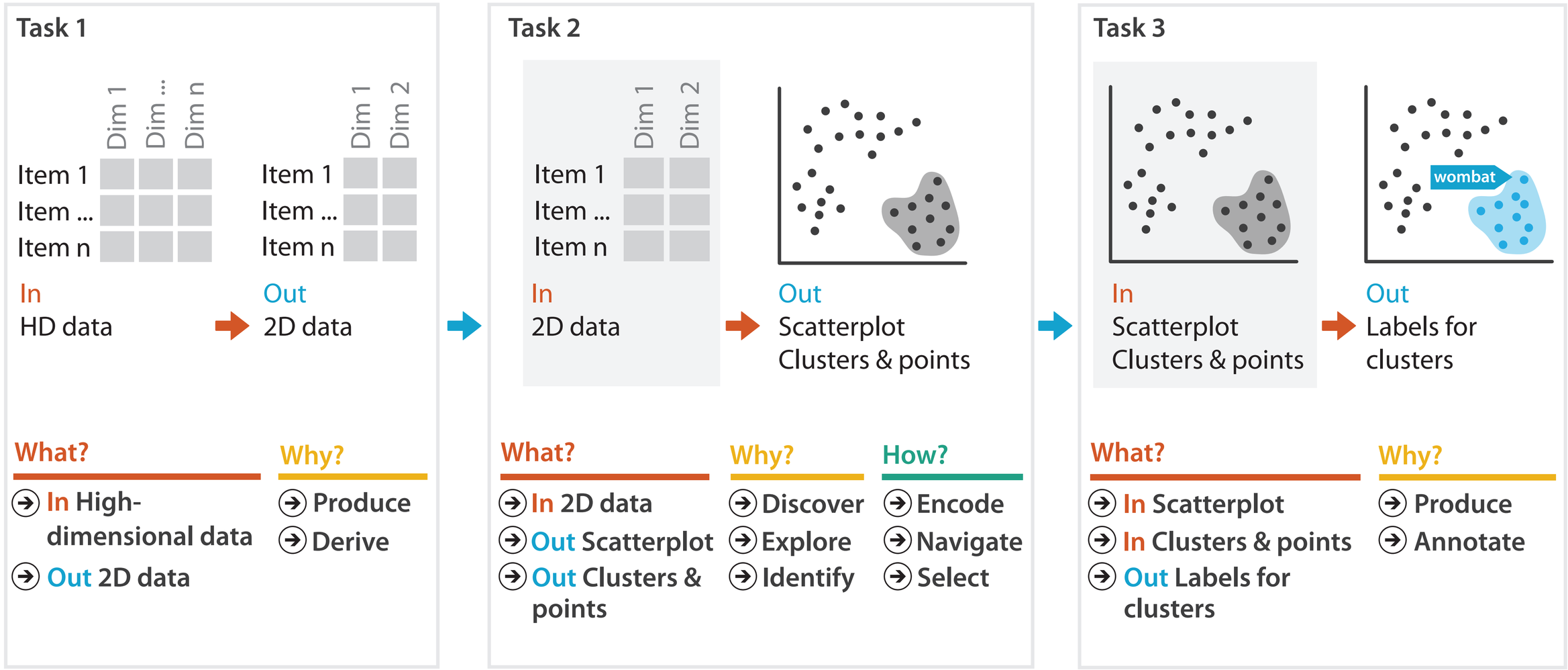
[M. Scholz, CC-BY-SA 2.0]

Dimensionality Reduction in Visualization



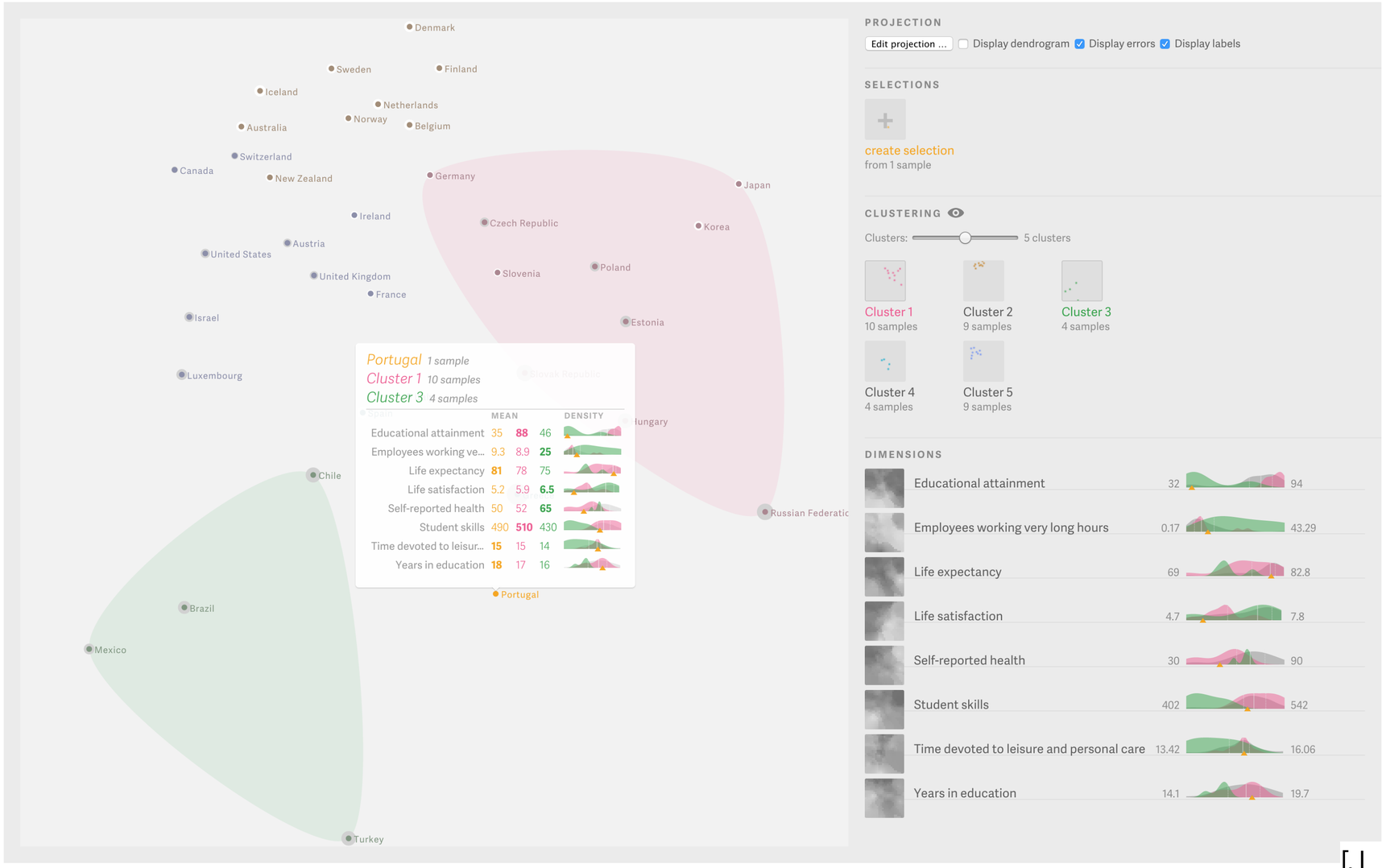
[Glimmer, Ingram et al., 2009]

Tasks in Understanding High-Dim. Data



[Munzner (ill. Maguire), 2014]

Probing Projections



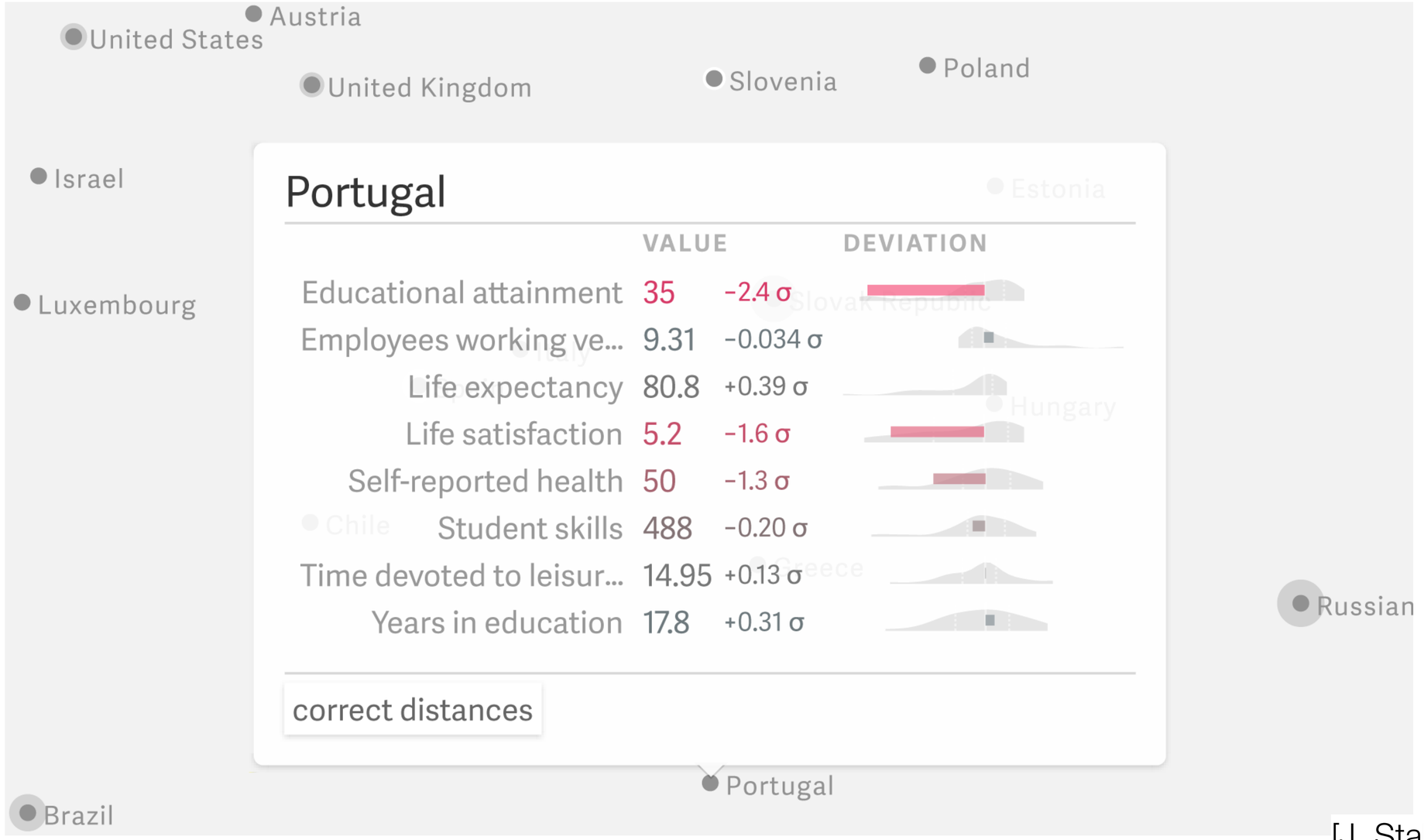
[J. Stahnke et al., 2015]

Probing Projection Goals

- Examining the Projection
- Exploring the Data
- Design Goals:
  - Show and correct approximation errors
  - Allow for multi-level comparisons
  - Spatial orientation
  - Consistent design
- Allow **grouping** of samples
  - Selections
  - Classes
  - Clusters

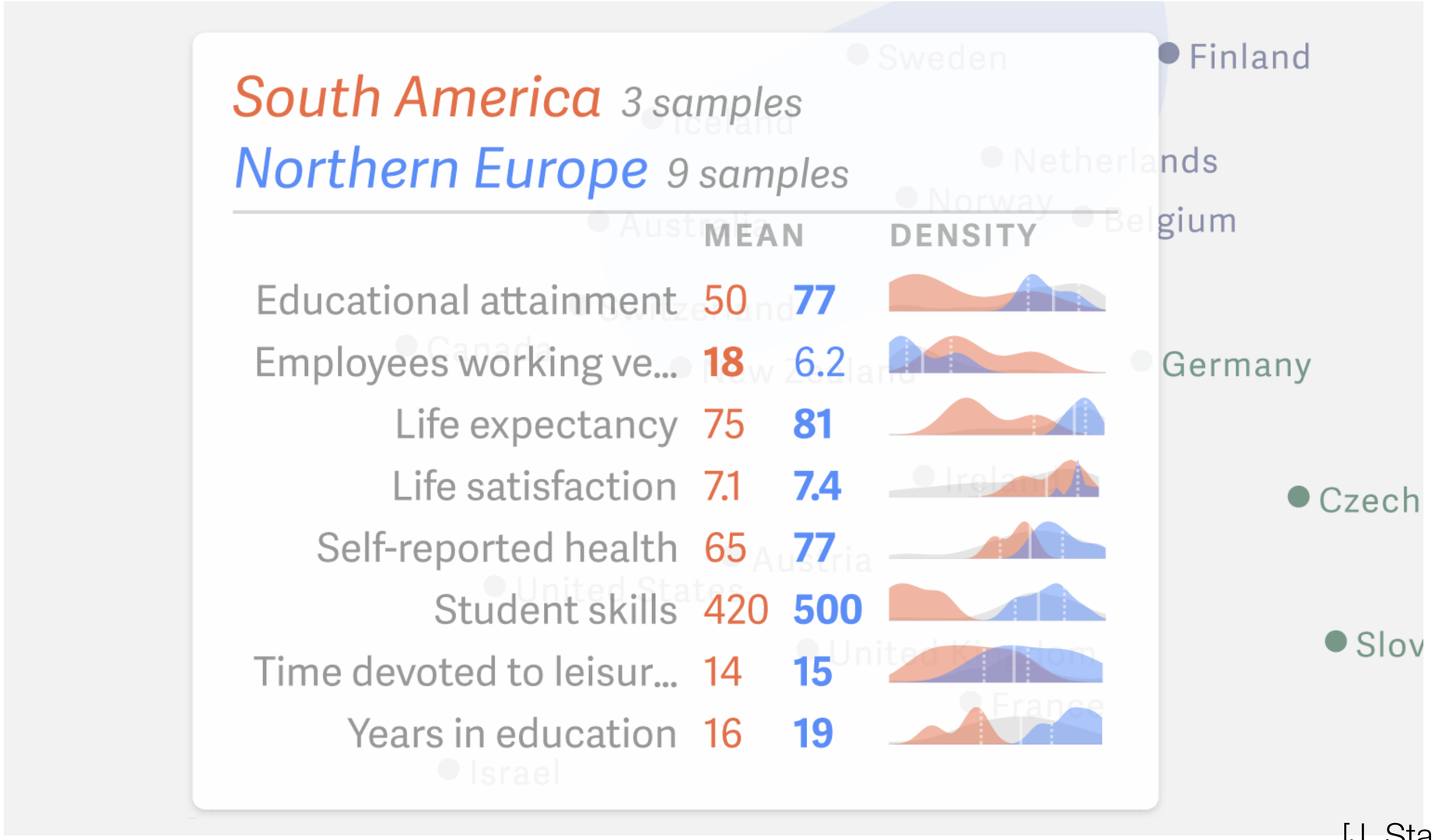
[J. Stahnke et al., 2015]

Tooltips with statistics



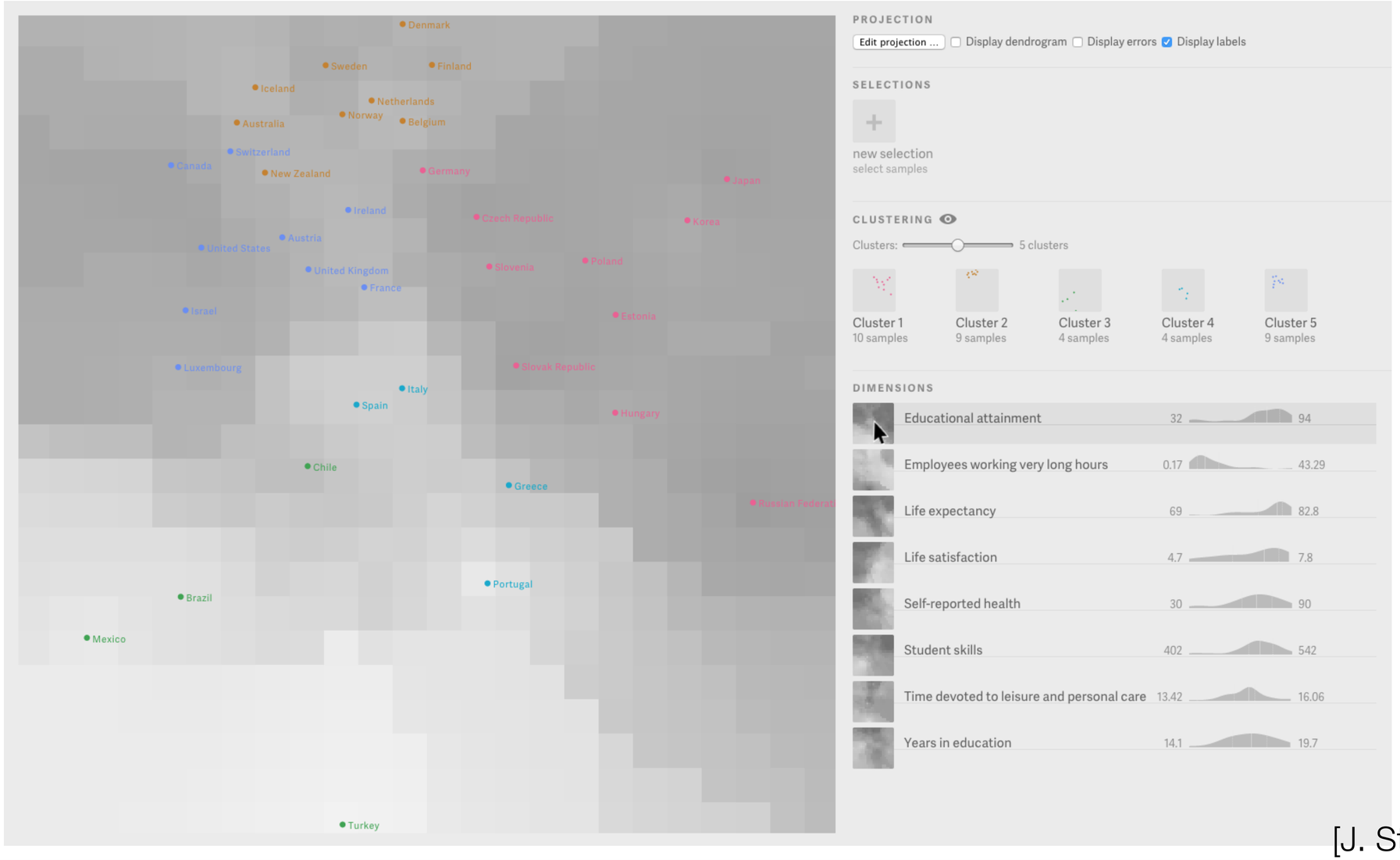
[J. Stahnke et al., 2015]

Comparing Two Groups



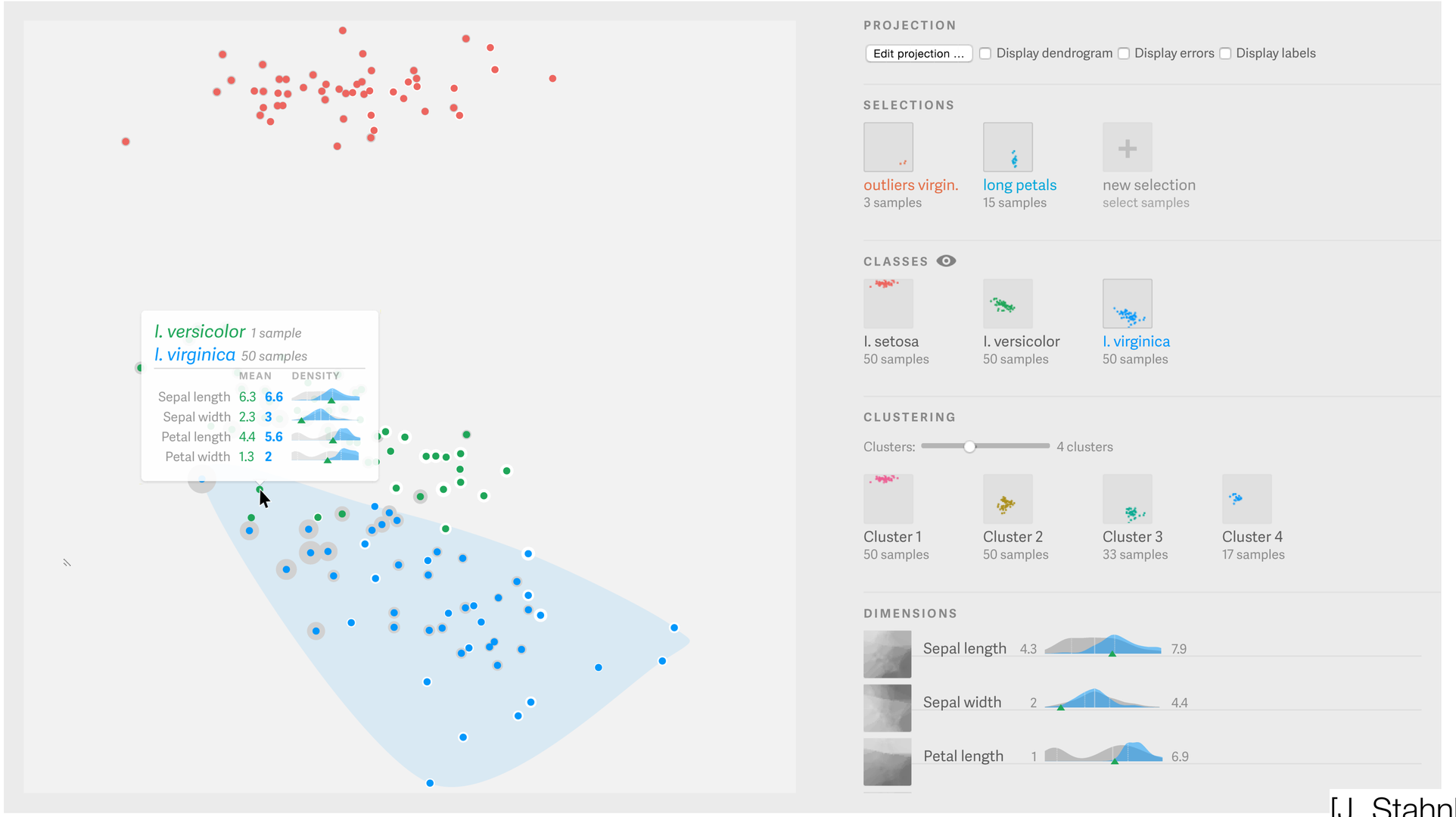
[J. Stahnke et al., 2015]

Heatmap from Dimension Hover



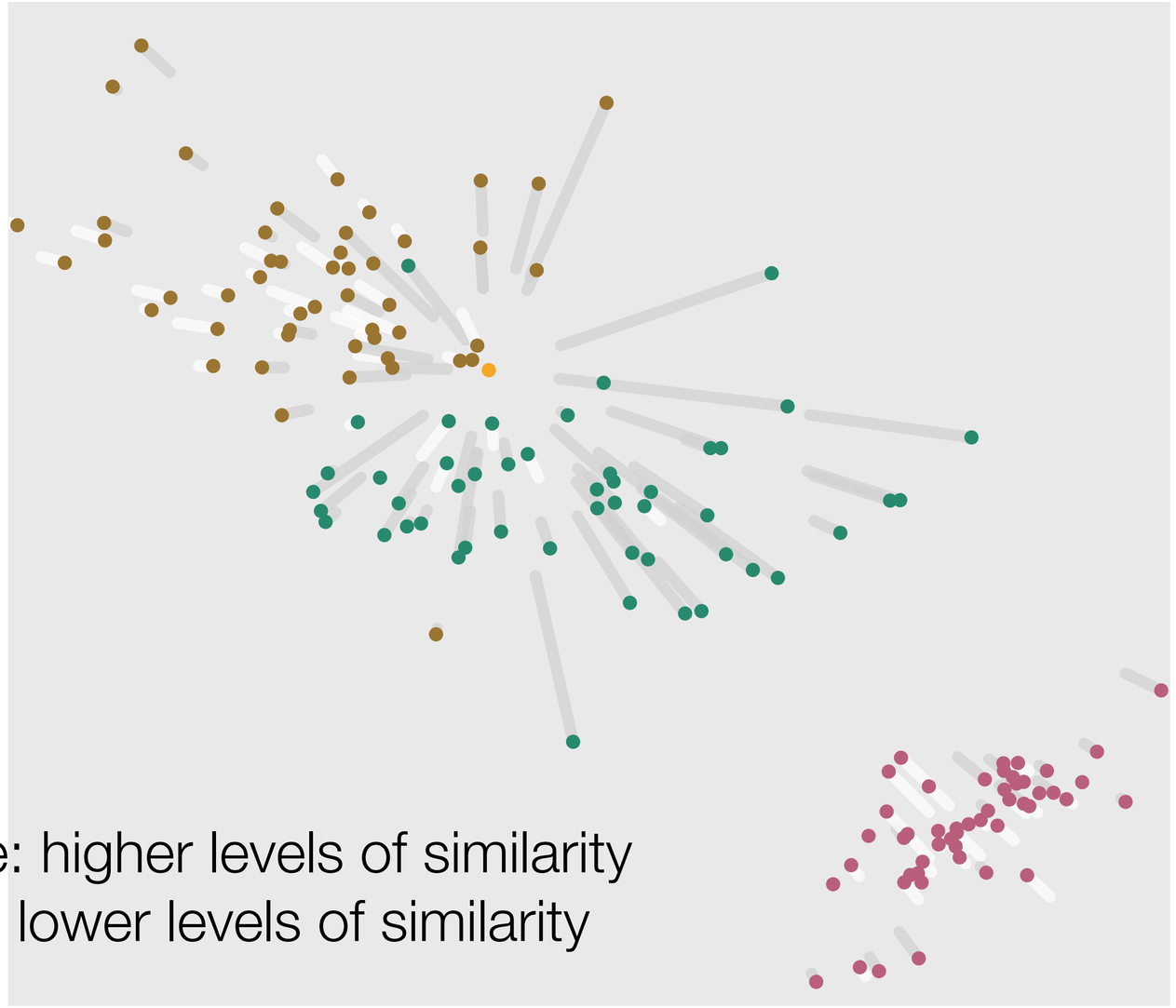
[J. Stahnke et al., 2015]

Showing Error via Sample-centric Halos



[J. Stahnke et al., 2015]

Showing Projection Errors



White: higher levels of similarity
Gray: lower levels of similarity

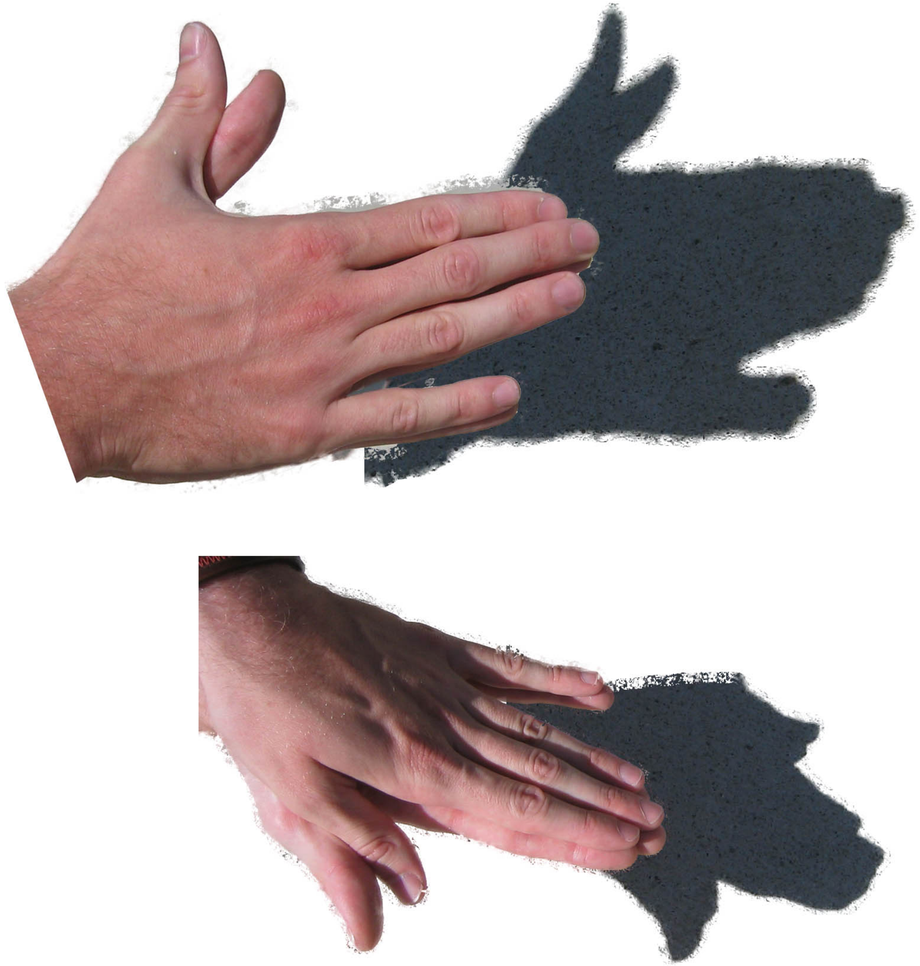
[J. Stahnke et al., 2015]

User Study & Results

- Types of Questions:
  - How would you try to characterize the type X?
  - In what way are X and Y different in their properties?
  - Are the projections of X and Y correct or do they deviate? How do you interpret this?
  - Can you discover which parts of the cluster combinations are A, B, and C?
- Discussion:
  - Learnability: need more effective mechanisms for grasping the concepts behind dimensionality reduction
  - Manipulation: What happens with results?
  - Large data: What about text corpora?

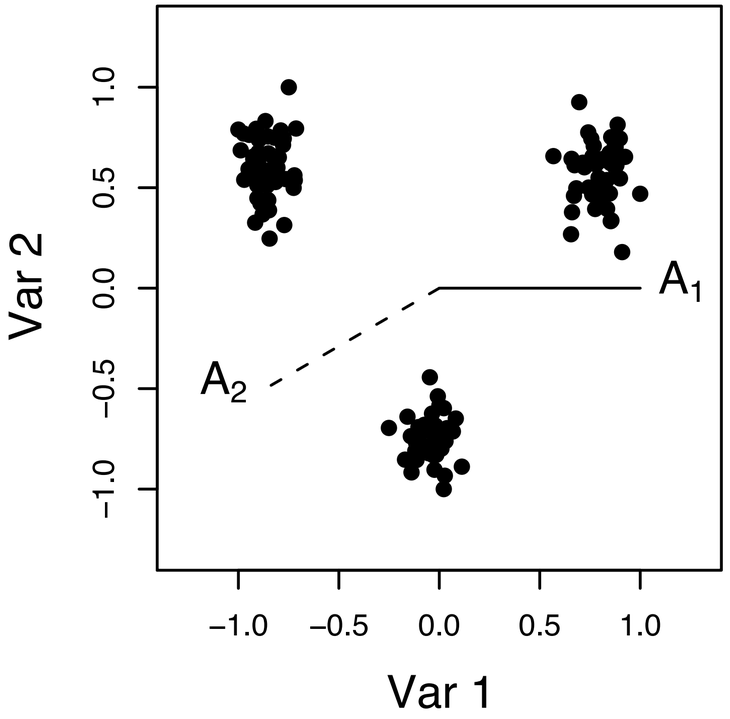
[J. Stahnke et al., 2015]

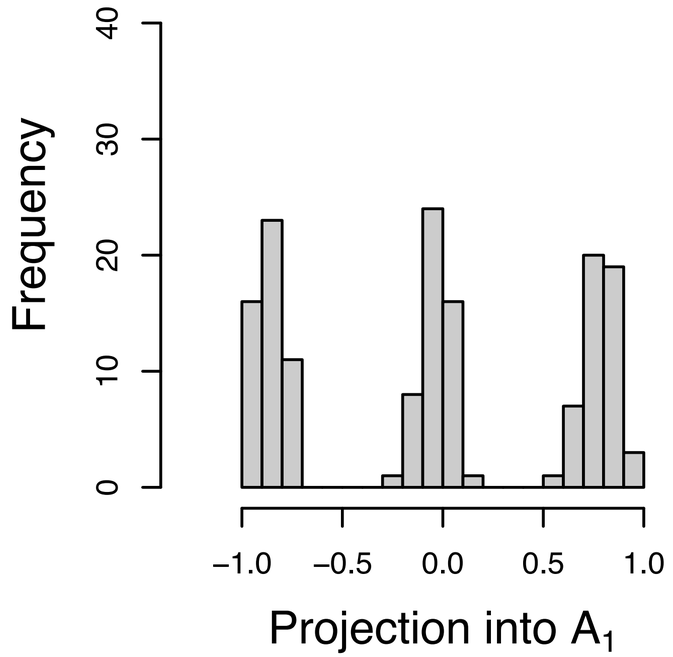
Different Projections

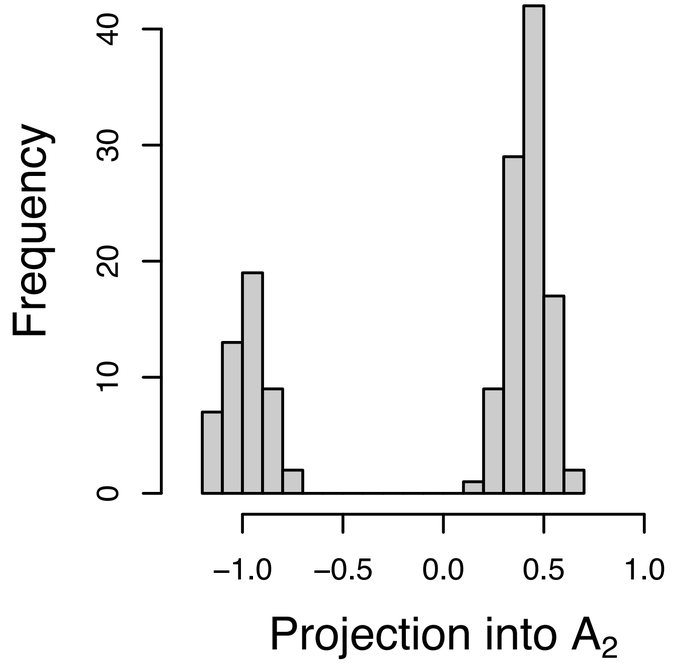


[D. Cook et al., 2008]

Leads to Different Conclusions

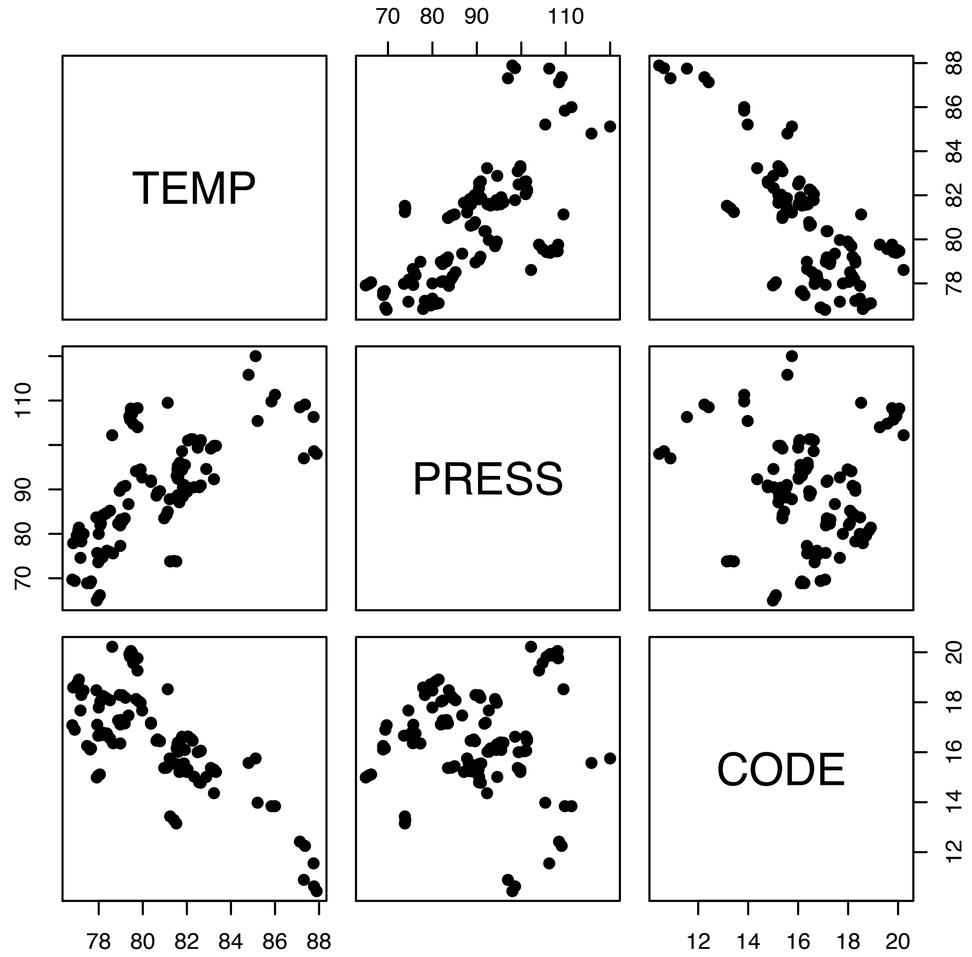


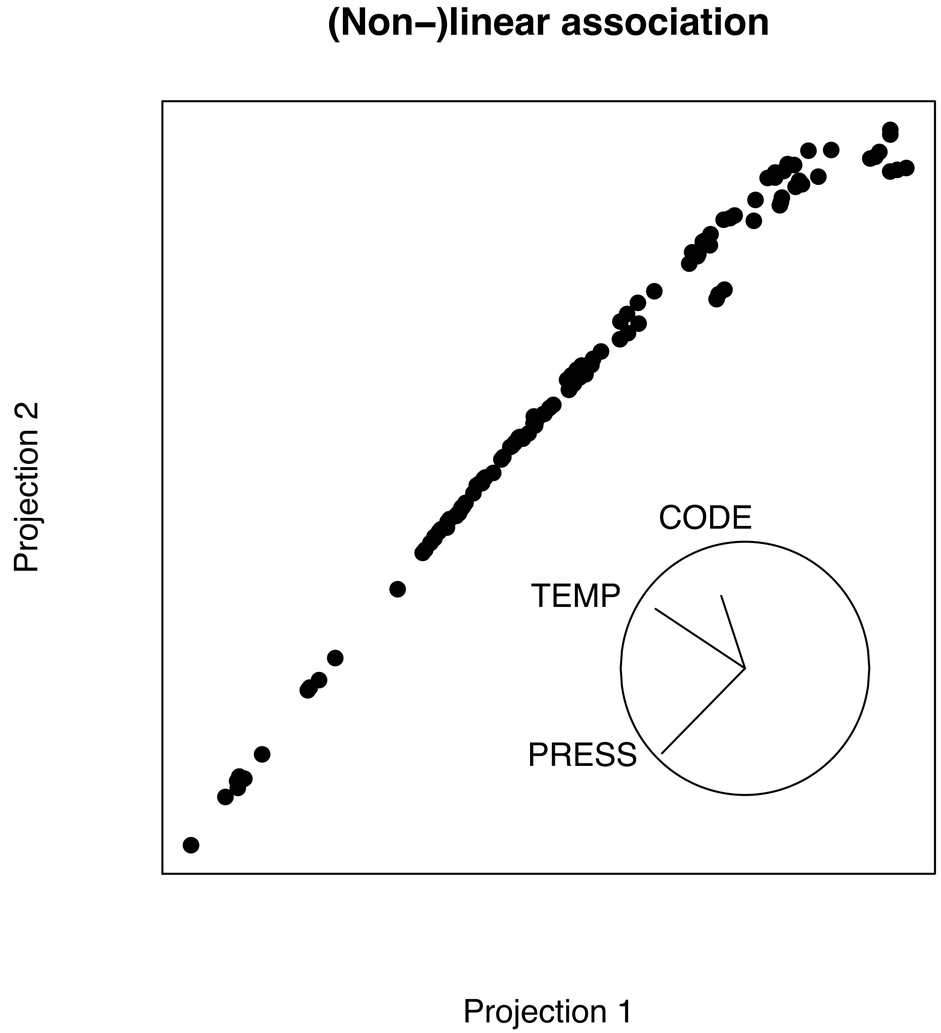




[D. Cook et al., 2008]

Tours help explore projections





[D. Cook et al., 2008]

Going beyond 2D and 3D to visualise higher dimensions, for ordination, clustering & other models

D. Cook

Toward Comparing DNNs with UMAP Tour

M. Li and C. Scheidegger