Information Visualization

High-Dimensional Data

Dr. David Koop





High-Dimensional Data Visualization Techniques

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

D. Koop, CSCI 628, Fall 2021





2

Principle Component Analysis (PCA)

original data space



D. Koop, CSCI 628, Fall 2021



PC 1















Probing Projections



D. Koop, CSCI 628, Fall 2021







4

Heatmap from Dimension Hover



	PROJECTION Edit projection Display dendrogram Display errors 🗸 Display labels					
	SELECTIONS					
	new selection select samples					
	CLUSTERING 💿					
	Clusters: 5 clusters					
			(w)		$\langle v_{i}\rangle$	74
ssian Federati	Cluster 1 10 sample	1 es	Cluster 2 9 samples	Cluster 3 4 samples	Cluster 4 4 samples	Cluster 5 9 samples
	DIMENSIONS					
		Educational attainment			32	94
	3	Employees working very long hours			0.17	43.29
		Life expe	ectancy		69	82.8
		Life satisfaction Self-reported health			4.7	7.8
					30	90
		Student skills			402	542
	-	Time devoted to leisure and personal care			13.42	16.06
	1	Years in education			14.1	19.7











Showing Projection Errors

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















Different Projections Lead to Different Conclusions



D. Koop, CSCI 628, Fall 2021

[D. Cook et al., 2008]





- A grand tour is...
 - "A space-filling curve in the manifold of lowdimensional projections of high-dimensional data spaces"
 - "A movie of low-dimensional projections" constructed in such a way that it comes arbitrarily close to showing all possible lowdimensional projections"
 - "A random walk through all possible planes"
 - Indexed by time











- A grand tour is...
 - "A space-filling curve in the manifold of lowdimensional projections of high-dimensional data spaces"
 - "A movie of low-dimensional projections" constructed in such a way that it comes arbitrarily close to showing all possible lowdimensional projections"
 - "A random walk through all possible planes"
 - Indexed by time











- Helps to see
 - Clusters
 - Outliers
 - Linear Dependence
 - Elliptical Clusters (different shapes)
 - Separated Elliptical Clusters











- Helps to see
 - Clusters
 - Outliers
 - Linear Dependence
 - Elliptical Clusters (different shapes)
 - Separated Elliptical Clusters











Different Views Help









Other Tours

- Guided: follows the optimization path for a projection pursuit index
- Local: rocks back and forth from a given projection, shows all possible projections within a radius
- Manual: start from best projection, control coefficient of one variable





Grand Tour of Image Data using UMAP



D. Koop, CSCI 628, Fall 2021



Northern Illinois University





Grand Tour of Image Data using UMAP



D. Koop, CSCI 628, Fall 2021



Northern Illinois University





Reading Critique

• Turn in by tonight





Project Updates

- Thursday
- If you want feedback, please bring demos, web pages





Reminder: IEEE VIS

- Lots of interesting work being presented
- Students have free admission
- Two more days: today and tomorrow
- <u>https://ieeevis.org</u>





Today's Paper

This paper is © IEEE and appears reformatted in IEEE Transactions on Visualization and Computer Graphics, 2017

The Subspace Voyager: Exploring High-Dimensional Data along a Continuum of Salient 3D Subspace

Bing Wang and Klaus Mueller, Senior Member, IEEE



Panel (a) and (c) are two projective views onto a 10-dimesional sales pipeline dataset with 900 points. The labels at the circle boundary indicate the data attributes and their axis directions in that view. The strength of the label fonts indicates how well the attributes are expressed in this view. Panel (b) shows a view generated by using our system's trackball interface to generate new projective views between view (a) and (c). The motion parallax clarified that there were not two but three clusters. Panel (d) shows the three clusters in different colors.



