SCIENTIFIC AND LARGE DATA VISUALIZATION

December 4, 2019

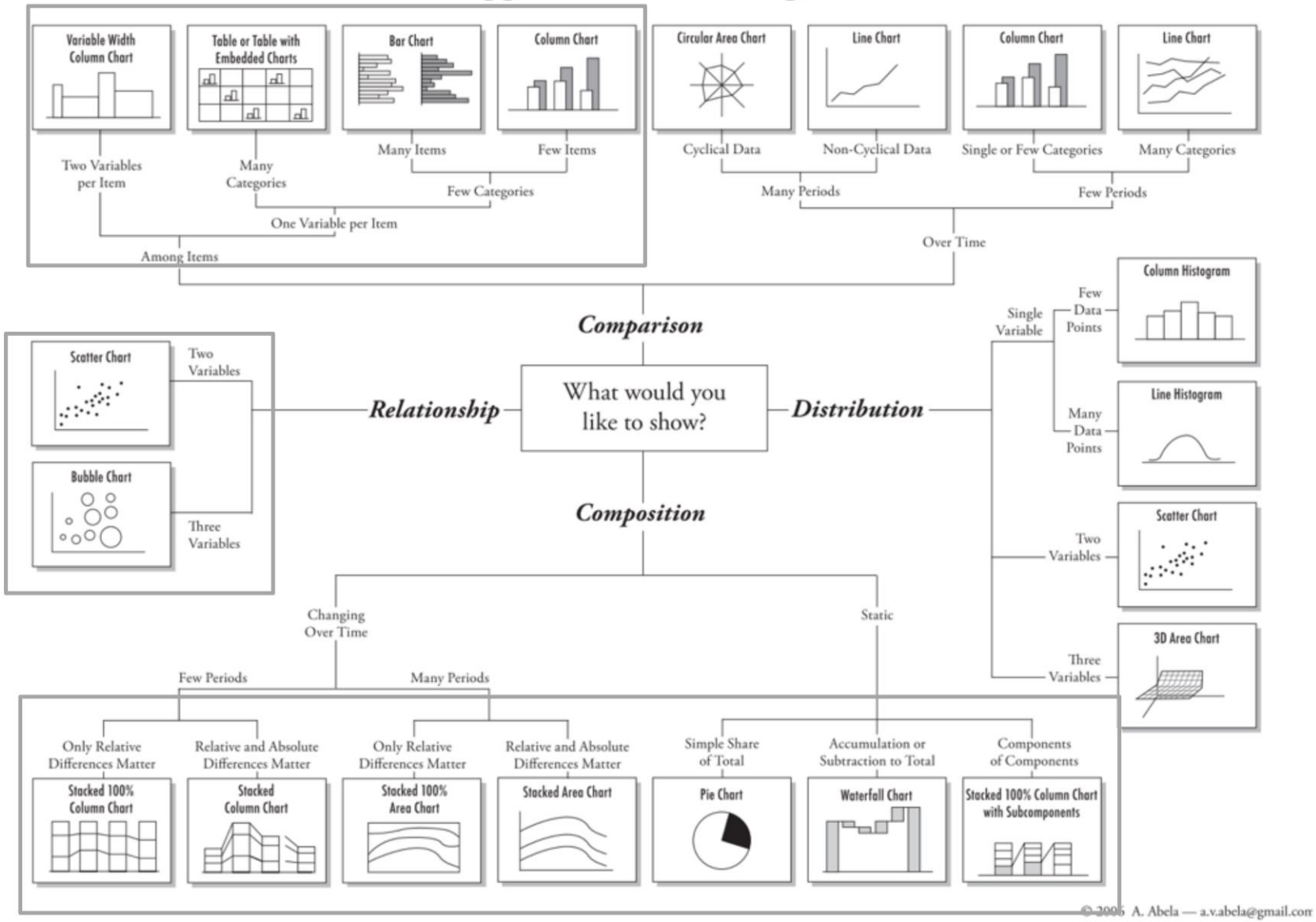
Multidimensional Data

Daniela Giorgi

Visual Computing Lab, CNR-ISTI

A brief recap

Chart Suggestions—A Thought-Starter

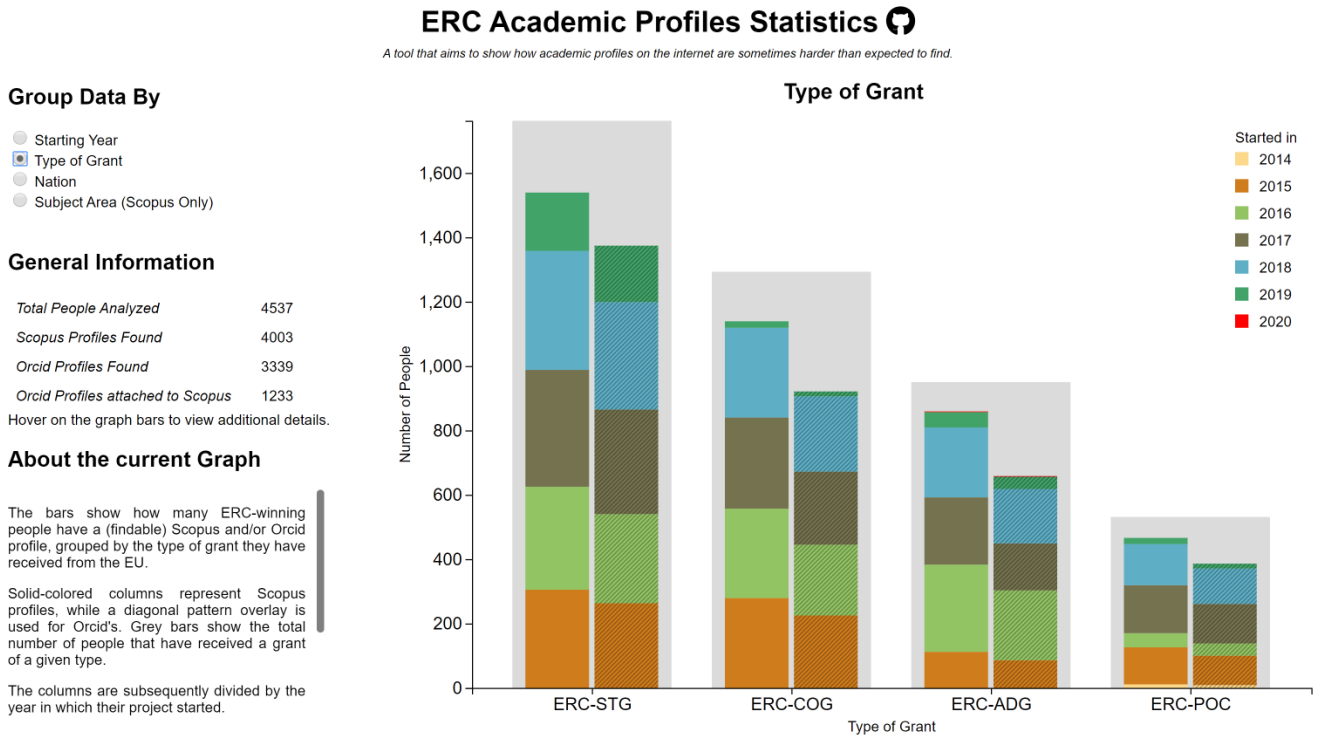


Dealing with multiple attributes

Stacking

Grouping

Small multiples



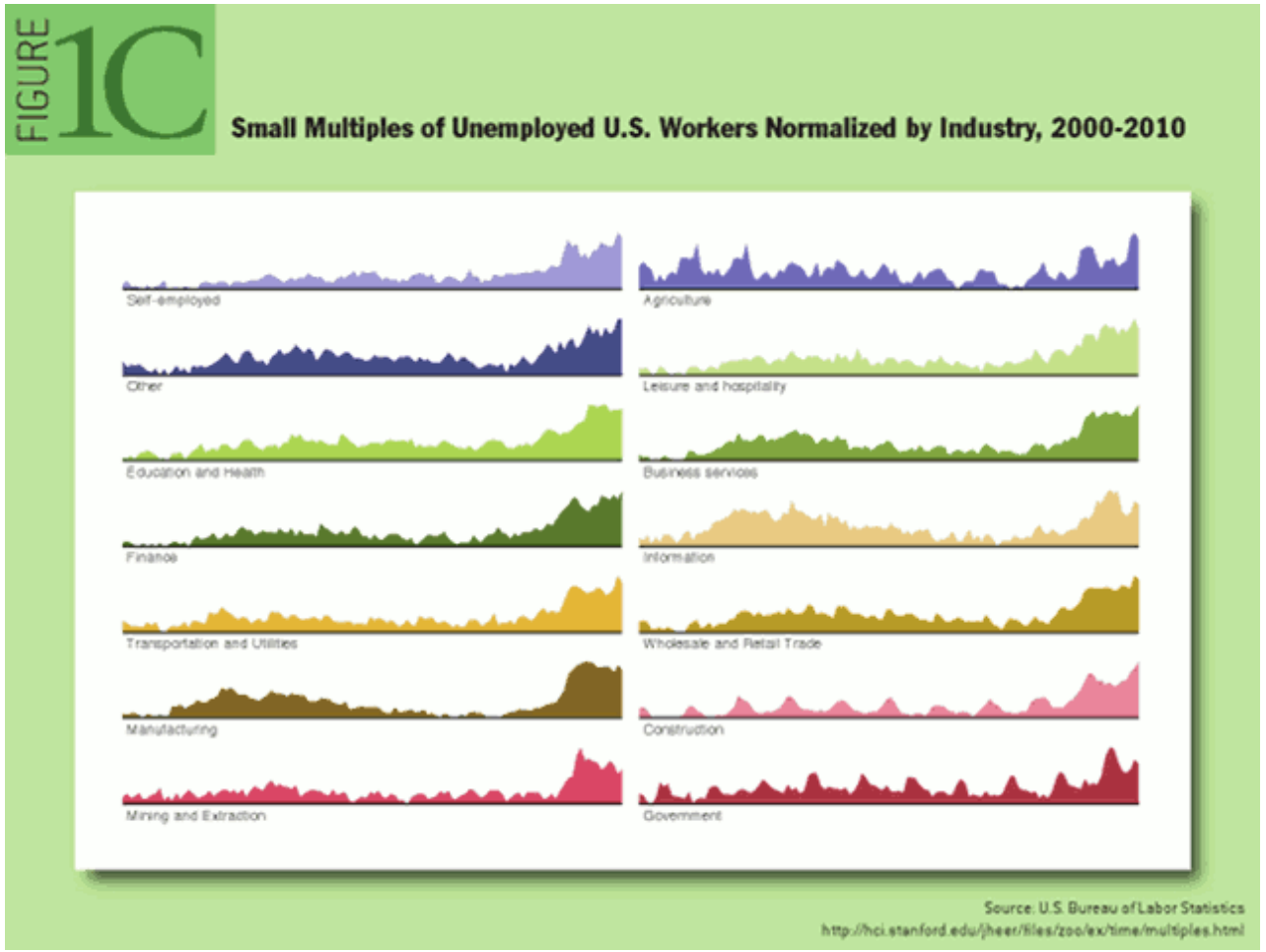
[*Davide Rucci, AA 2018-2019*, https://drdav.github.io/ERC-Academic-Profiles/]

Dealing with multiple attributes

Stacking

Grouping

Small multiples



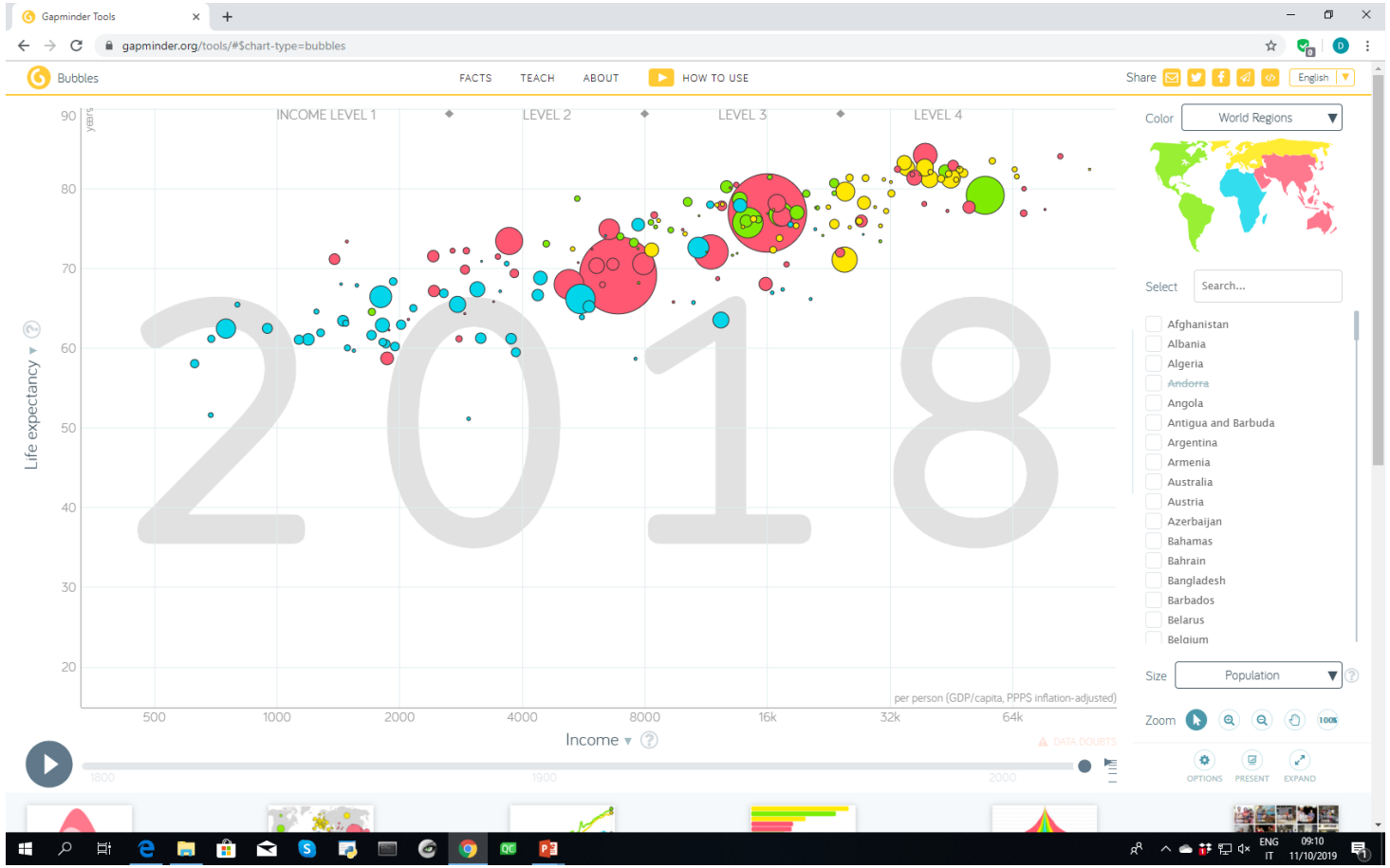
[*A tour through the visualization zoo*, https://queue.acm.org/detail.cfm?id=1805128]

Dealing with multiple attributes

Stacking

Grouping

Small multiples



[Gapminder https://gapminder.org/; Hans Rosling Ted Talk https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen?language=it#t-283414]

Iconographic displays

Chernoff faces

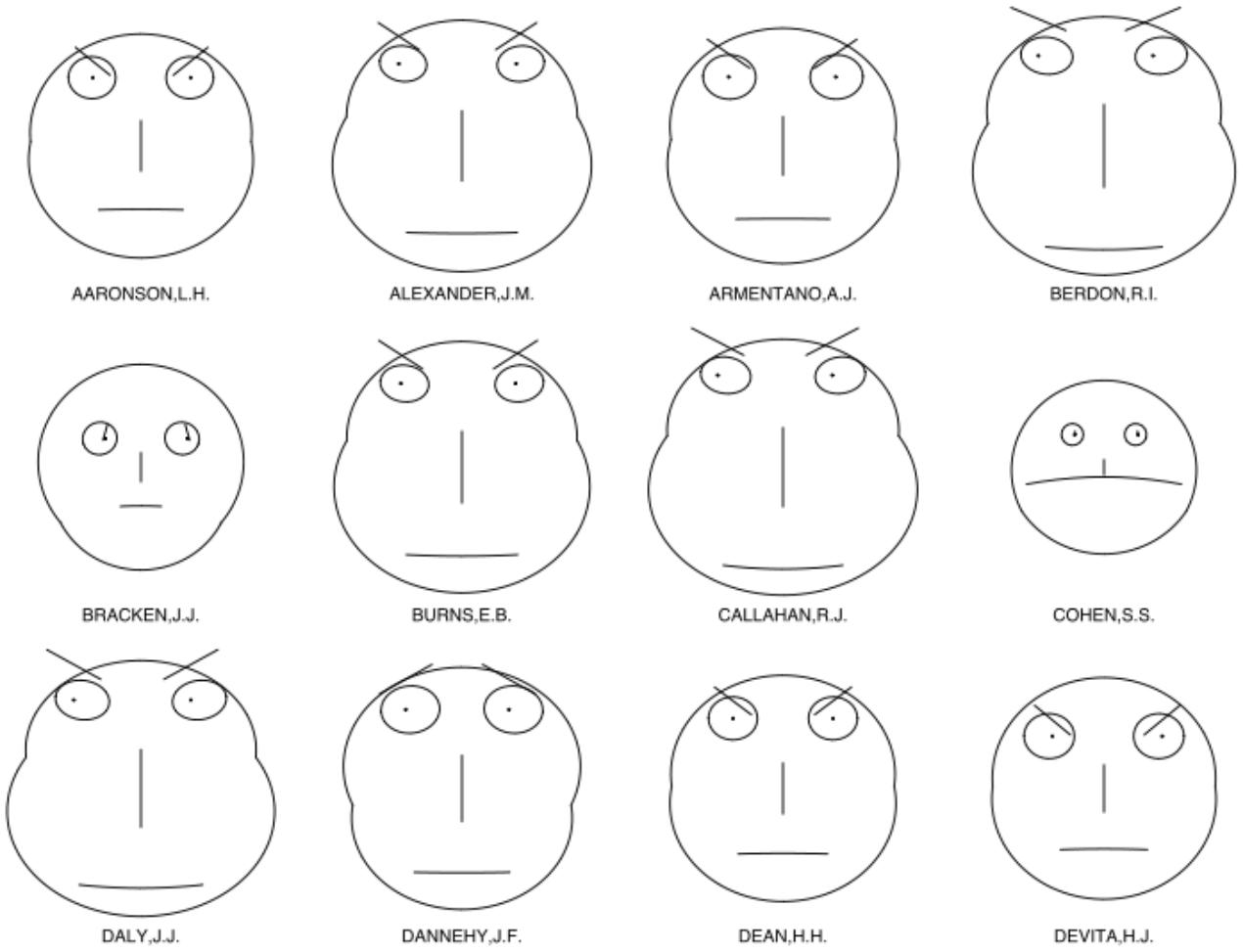
Rationale: we are very good at recognizing faces

Introduced by Herman Chernoff in 1973

Variables are mapped to facial features (width/curvature of mouth, vertical size of face, size/slant/separation of eyes, size of eyebrows, vertical position of eyebrows…)

Legend needed here…

…and not preattentive processes?



[lawyers' ratings of twelve judges]

Chernoff faces

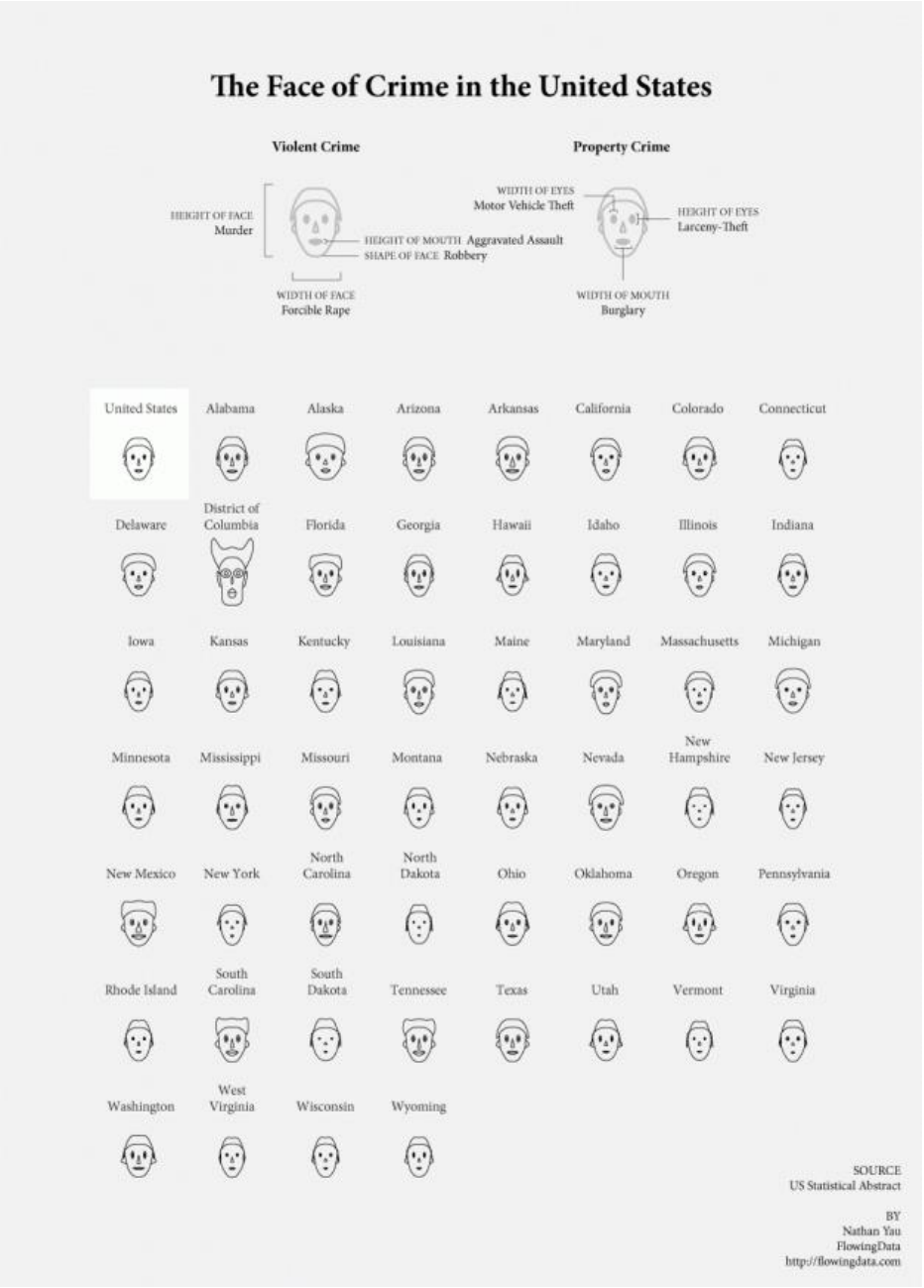
Rationale: we are very good at recognizing faces

Introduced by Herman Chernoff in 1973

Variables are mapped to facial features (width/curvature of mouth, vertical size of face, size/slant/separation of eyes, size of eyebrows, vertical position of eyebrows…)

Legend needed here…

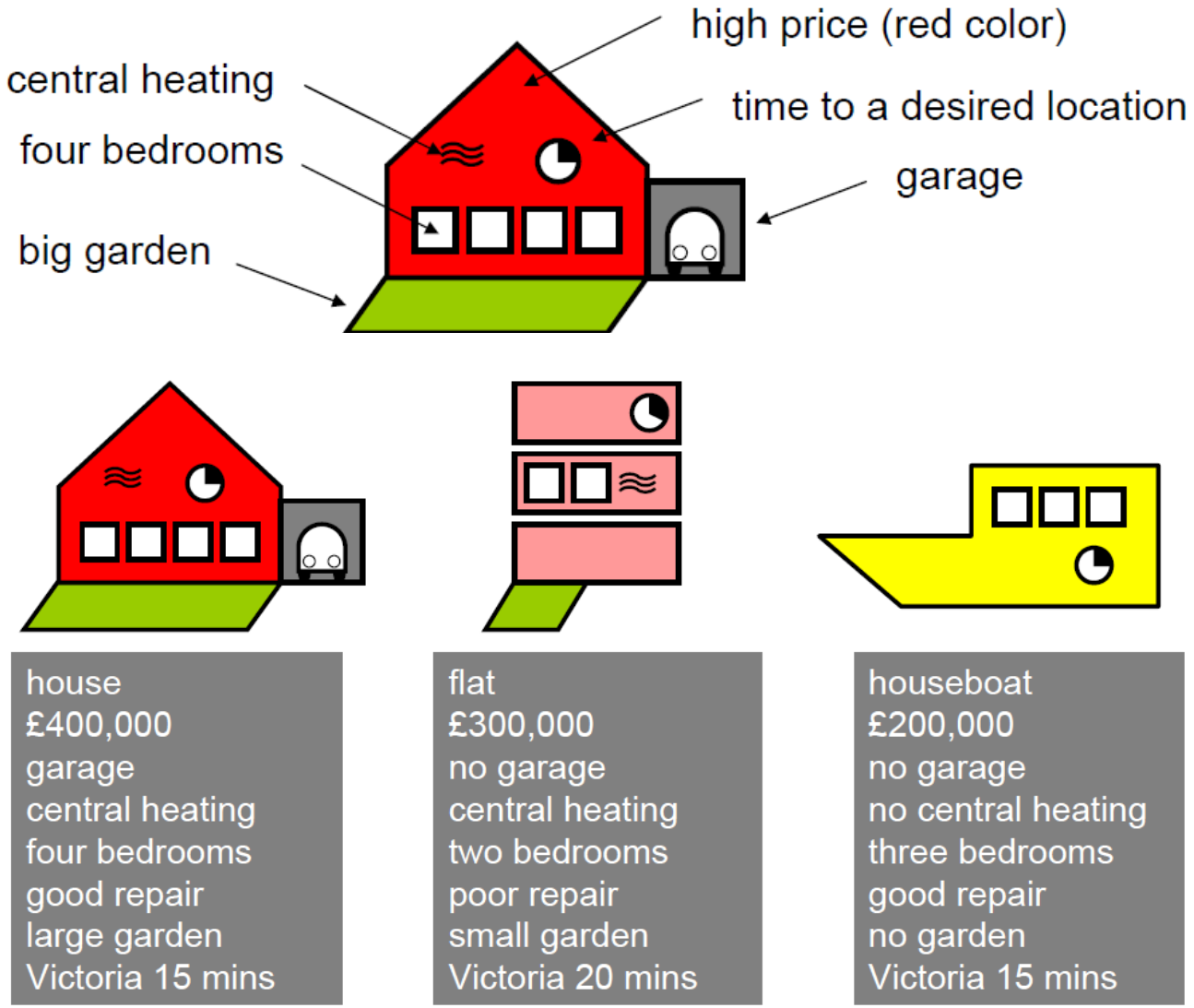
…and not preattentive processes?



Multidimensional icons

Spence and Parr (1991) proposed to encode properties of an object in a simple iconic representation

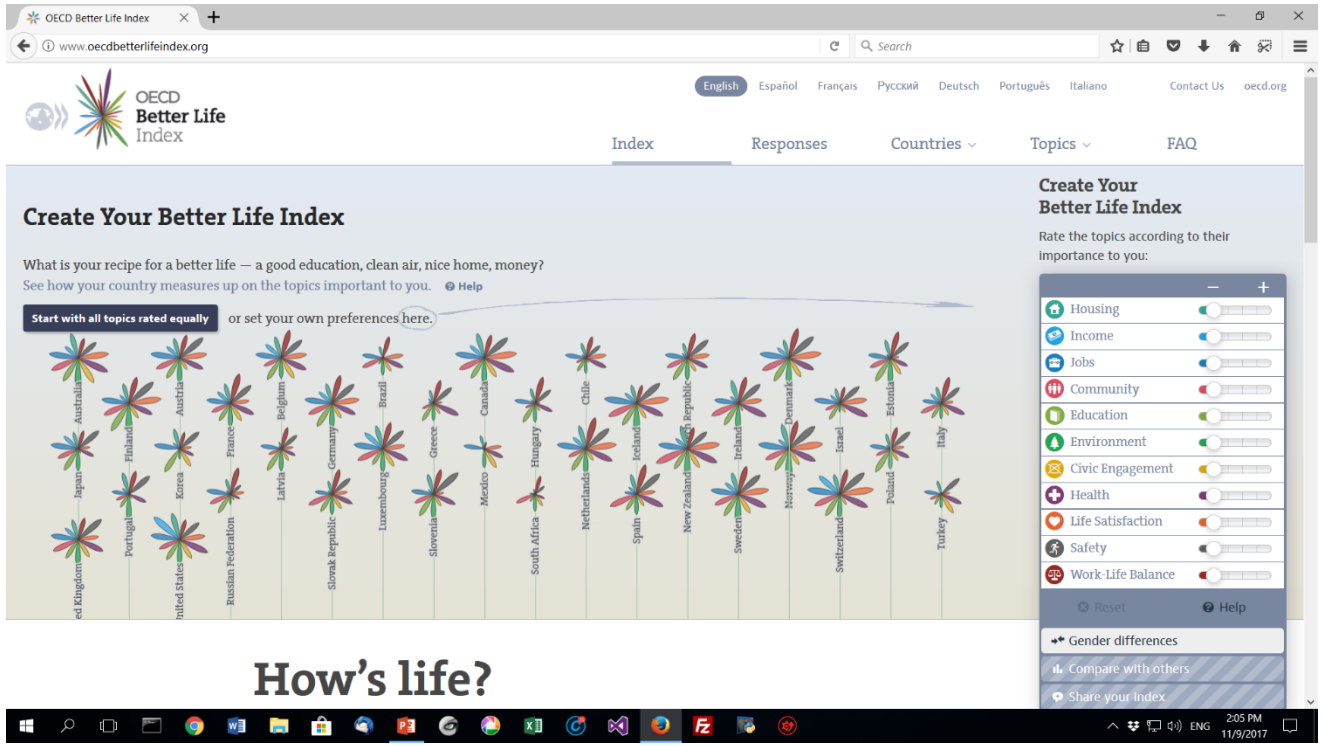
They applied this approach to check dwell offers



[Robert Spence and Maureen Parr, "Cognitive assessment of alternatives", Interacting with Computers 3, 1991]

Petals as a gliph

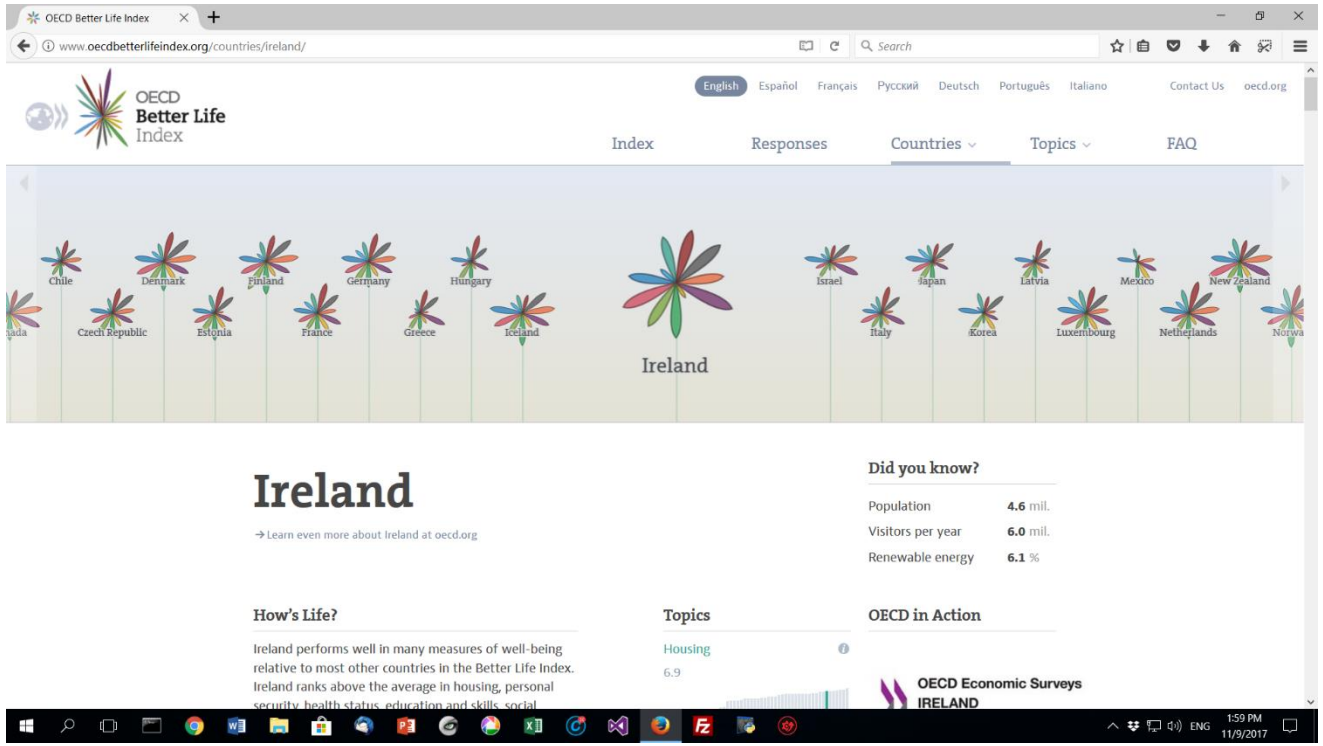
The idea of Moritz Stefaner to visualize a life index is to map several variables into petals of different size



[www.oecdbetterlifeindex.org]

Petals as a gliph

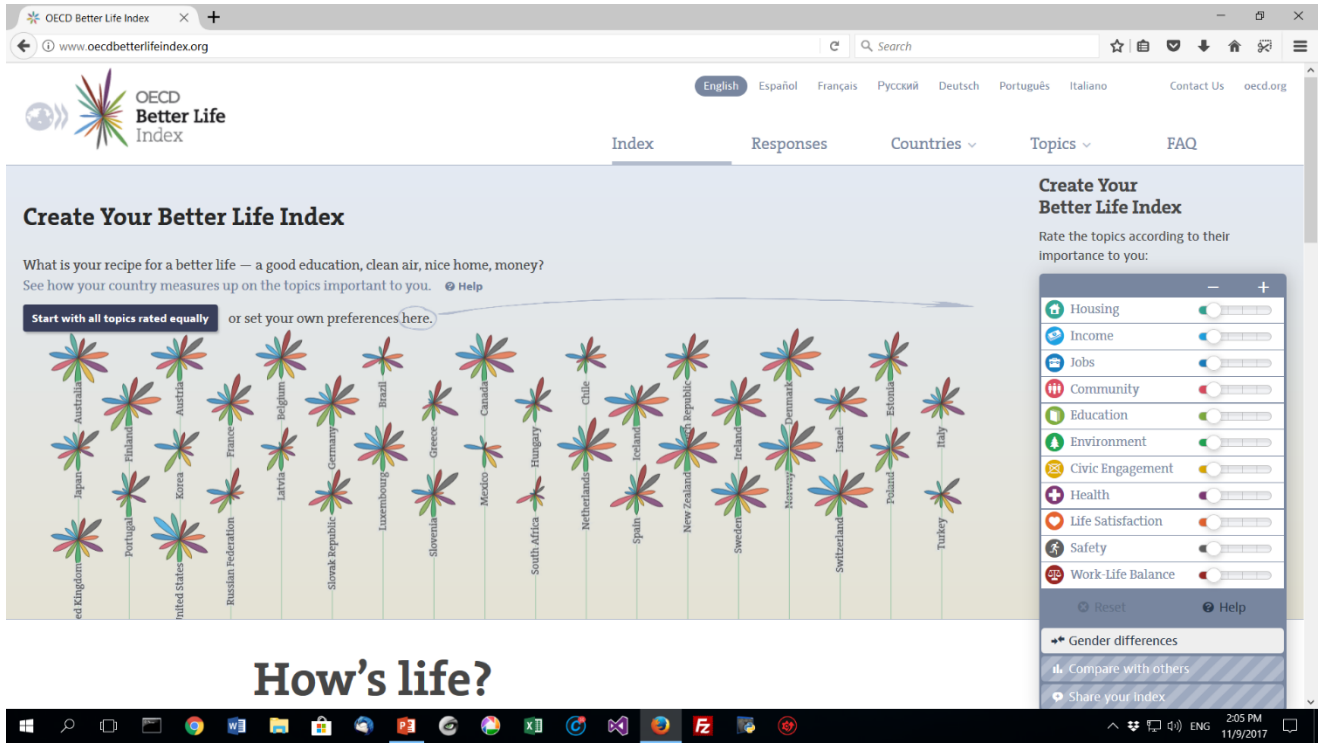
The idea of Moritz Stefaner to visualize a life index is to map several variables into petals of different size



[www.oecdbetterlifeindex.org]

Petals as a gliph

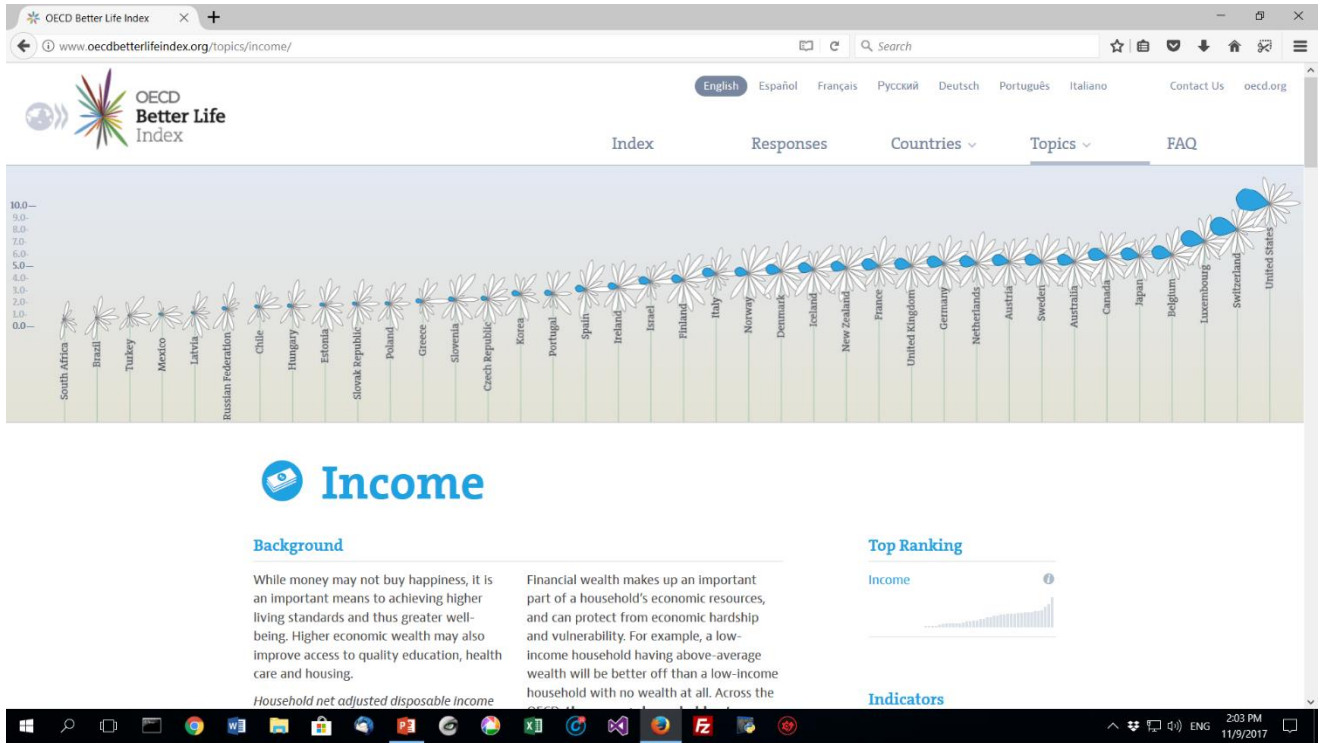
The idea of Moritz Stefaner to visualize a life index is to map several variables into petals of different size



[www.oecdbetterlifeindex.org]

Petals as a gliph

The idea of Moritz Stefaner to visualize a life index is to map several variables into petals of different size



[www.oecdbetterlifeindex.org]

Dimensionality reduction

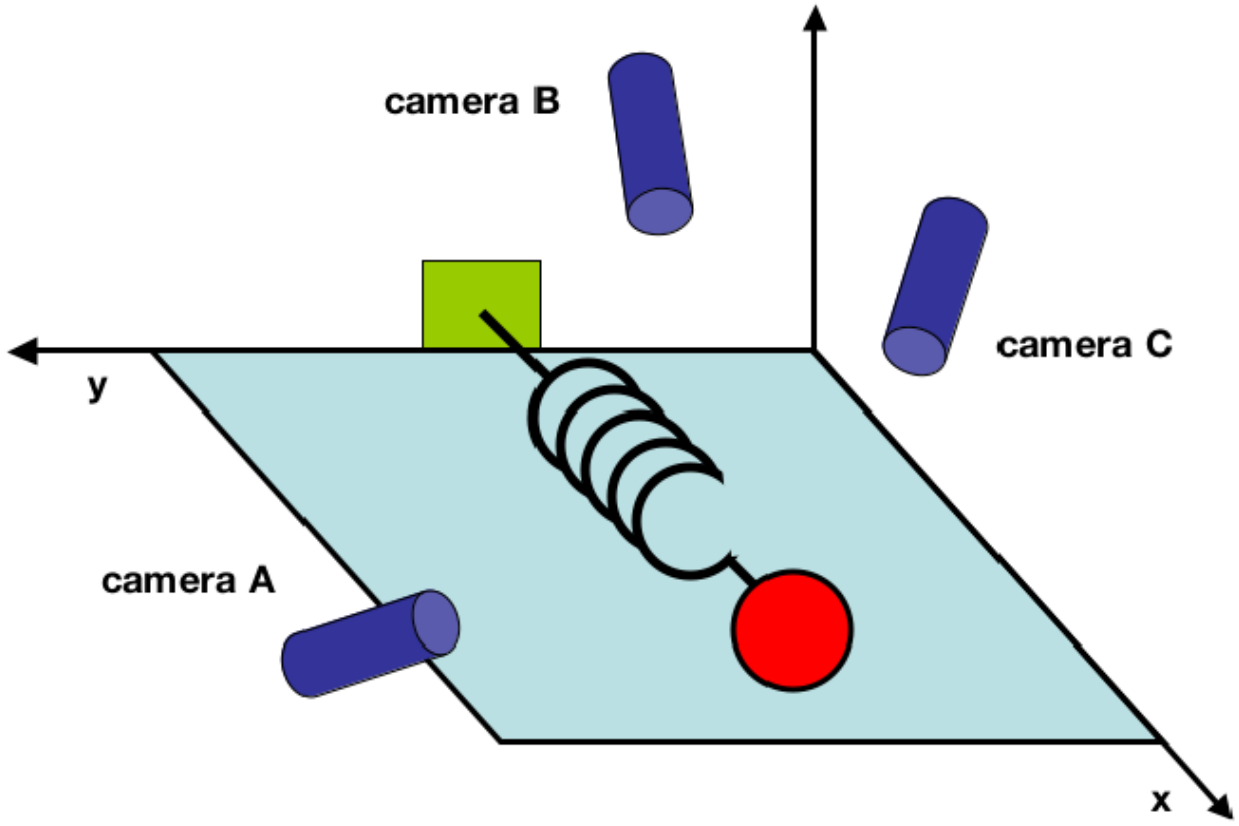
Dimensionality reduction

Finding meaningful low-dimensional structures hidden in high-dimensional observations

- Rationale: high-dimensional data are projected to a lower number of dimensions for better visualization and/or understanding
  - e.g., cell nuclei in breast cancer can be described by approximately 30 variables, as opposed to thousands of pixels in images [Street et al., 1993]
  - the human brain confronts the same problem in everyday perception, extracting from its high-dimensional sensory input a manageably small number of perceptually relevant features
- Many different mapping strategies, having different properties
  - *linear vs non-linear* underlying structure for the data

Principal Component Analysis (PCA)

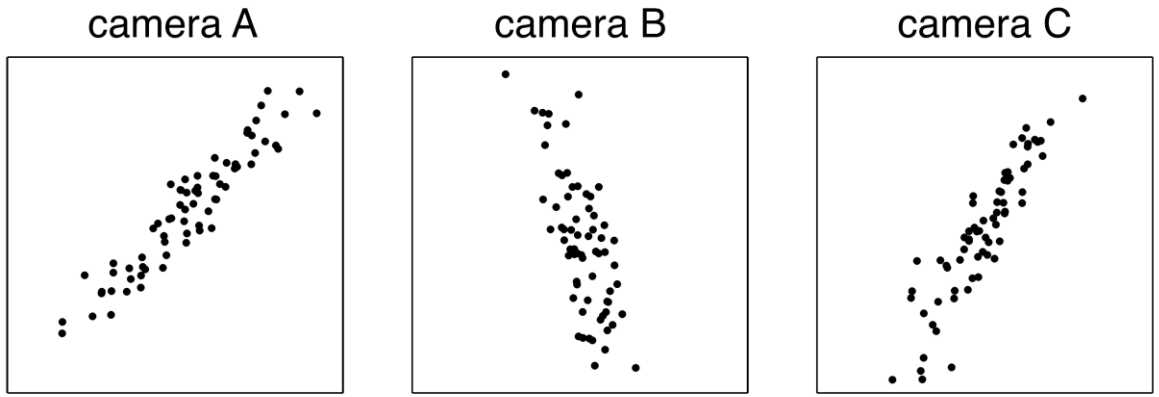
Motivation: we often end up measuring too many variables, also contaminated by noise, whereas the underlying relationships can be quite simple



[Toy example, from Jonathon Shlens, "A Tutorial on Principal Component Analysis", arXiv preprint arXiv:1404.1100, 2015]

Principal Component Analysis (PCA)

Motivation: we often end up measuring too many variables, also contaminated by noise, whereas the underlying relationships can be quite simple



$$m = \begin{bmatrix} x_A \\ y_A \\ x_B \\ y_B \\ x_C \\ y_C \end{bmatrix}$$

Each measure has 6 dimensions (and consider recording for several minutes…)

…but we know the ball moves along the *x*-axis only

Which measurements best reflect the dynamics of the system?

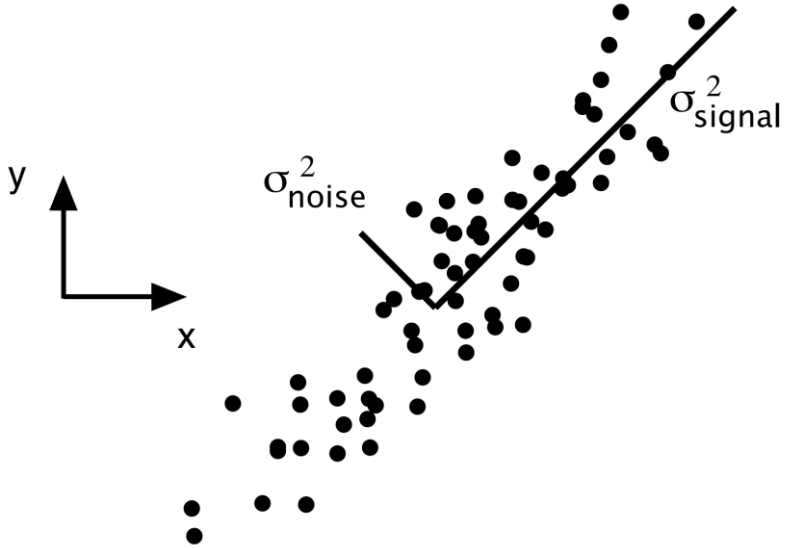
[Toy example, from Jonathon Shlens, "A Tutorial on Principal Component Analysis", arXiv preprint arXiv:1404.1100, 2015]

Principal Component Analysis (PCA)

Motivation: we often end up measuring too many variables, also contaminated by noise, whereas the underlying relationships can be quite simple

- Assumption 1: the dynamics of interest exist along directions with largest variance, and presumably with highest signal-to-noise ratio (SNR)

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}$$

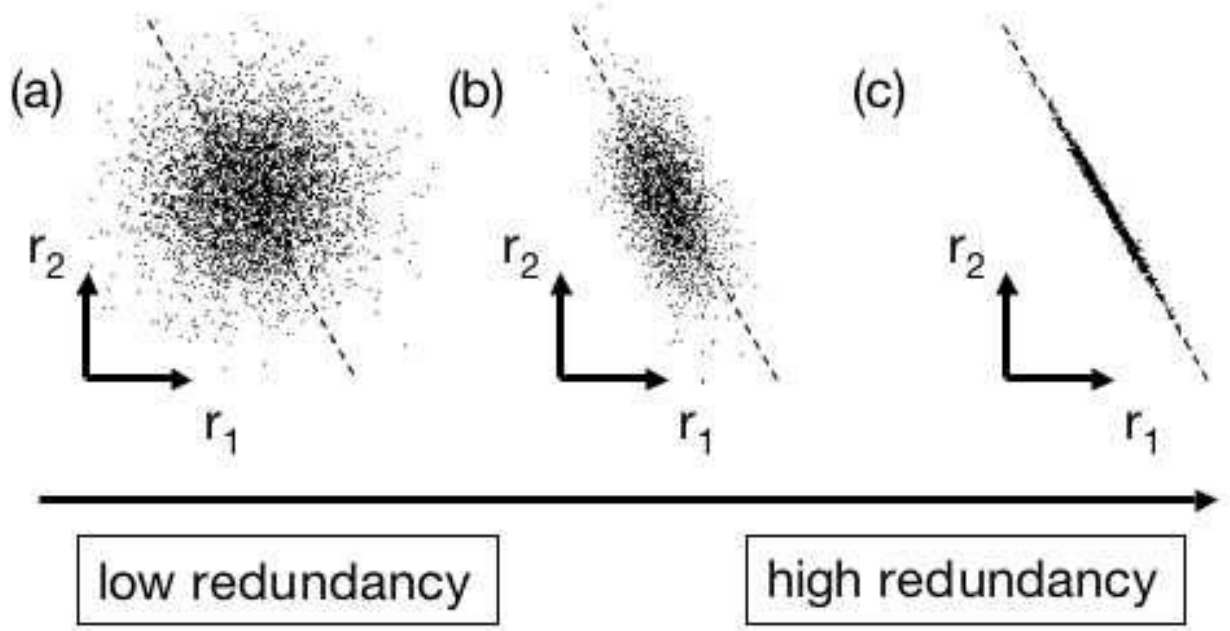


[Toy example, from Jonathon Shlens, "A Tutorial on Principal Component Analysis", arXiv preprint arXiv:1404.1100, 2015]

Principal Component Analysis (PCA)

Motivation: we often end up measuring too many variables, also contaminated by noise, whereas the underlying relationships can be quite simple

- Assumption 2: redundancy is a confounding factor
  - Redundant variables do not convey useful information



[Toy example, from Jonathon Shlens, "A Tutorial on Principal Component Analysis", arXiv preprint arXiv:1404.1100, 2015]

Principal Component Analysis (PCA)

Idea: look for the most meaningful basis to re-express a dataset

Linearity assumption: re-express the data as a linear combination of its basis vectors

- **X** is an (*m*x*n*) matrix representing the original dataset, where each row represents a type of measurement, and each column is a single sample (or moment in time)
- **Y** is an (*m*x*n*) matrix which stores the new representation of the dataset
- **P** is the matrix that transforms **X** into **Y**

$$\mathbf{PX} = \mathbf{Y}$$

Principal Component Analysis (PCA)

Idea: look for the most meaningful basis to re-express a dataset

Linearity assumption: re-express the data as a linear combination of its basis vectors

- **X** is an (*m*x*n*) matrix representing the original dataset, where each row represents a type of measurement, and each column is a single sample (or moment in time)
- **Y** is an (*m*x*n*) matrix which stores the new representation of the dataset
- **P** is the matrix that transforms **X** into **Y**
- Change of basis: The rows of **P** are a new set of basis vectors for representing the columns of **X**

$$\mathbf{PX} = \begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_m \end{bmatrix} [\mathbf{x}_1 \dots \mathbf{x}_n] \qquad \mathbf{y}_i = \begin{bmatrix} \mathbf{p}_1 \cdot \mathbf{x}_i \\ \vdots \\ \mathbf{p}_m \cdot \mathbf{x}_i \end{bmatrix}$$

Principal Component Analysis (PCA)

Idea: look for the most meaningful basis to re-express a dataset

Linearity assumption: re-express the data as a linear combination of its basis vectors

- How to define **P**?
  - maximize the variance
  - minimize redundancy
- Goal: find an orthonormal matrix **P** such that the covariance matrix of $\mathbf{Y}$, $\mathbf{C_Y} \equiv \frac{1}{n}\mathbf{YY}^T$, is a diagonal matrix
  - high values in the diagonal terms correspond to interesting structures
  - low values of the off-diagonal terms means that the redundancy between variables is minimized

Principal Component Analysis (PCA)

Idea: look for the most meaningful basis to re-express a dataset

Linearity assumption: re-express the data as a linear combination of its basis vectors

- How to define P?
  - maximize the variance
  - minimize redundancy
- Goal: find an orthonormal matrix **P** such that the covariance matrix of $\mathbf{Y}$, $\mathbf{C_Y} \equiv \frac{1}{n}\mathbf{YY}^T$, is a diagonal matrix
- We can select the matrix **P** to be a matrix where each row $\mathbf{p}_i$ is an eigenvector of

$$\mathbf{C_X} \equiv \frac{1}{n}\mathbf{XX}^T$$

Principal Component Analysis (PCA)

Idea: look for the most meaningful basis to re-express a dataset

Linearity assumption: re-express the data as a linear combination of its basis vectors

- How to define P?
  - Maximize the variance
  - Minimize redundancy
- Goal: find an orthonormal matrix **P** such that the covariance matrix of $\mathbf{Y}$, $\mathbf{C_Y} \equiv \frac{1}{n}\mathbf{YY}^T$, is a diagonal matrix
- We can select the matrix **P** to be a matrix where each row $\mathbf{p_i}$ is an eigenvector of

$$\mathbf{C_X} \equiv \frac{1}{n}\mathbf{XX}^T$$

- The principal components of $\mathbf{X}$ are the eigenvectors of $\mathbf{C_X} = \frac{1}{n}\mathbf{XX}^T$.
- The i^{th} diagonal value of $\mathbf{C_Y}$ is the variance of $\mathbf{X}$ along $\mathbf{p_i}$.

Principal Component Analysis (PCA)

Idea: look for the most meaningful basis to re-express a dataset

Linearity assumption: re-express the data as a linear combination of its basis vectors

- Organize the data as an *m x n* matrix.
- Subtract the corresponding mean to each row.
- Calculate the eigenvalues and eigenvectors of $\boldsymbol{XX^T}$
- Organize them to form the matrix $\boldsymbol{P}$

Principal Component Analysis (PCA)

Idea: look for the most meaningful basis to re-express a dataset

Linearity assumption: re-express the data as a linear combination of its basis vectors

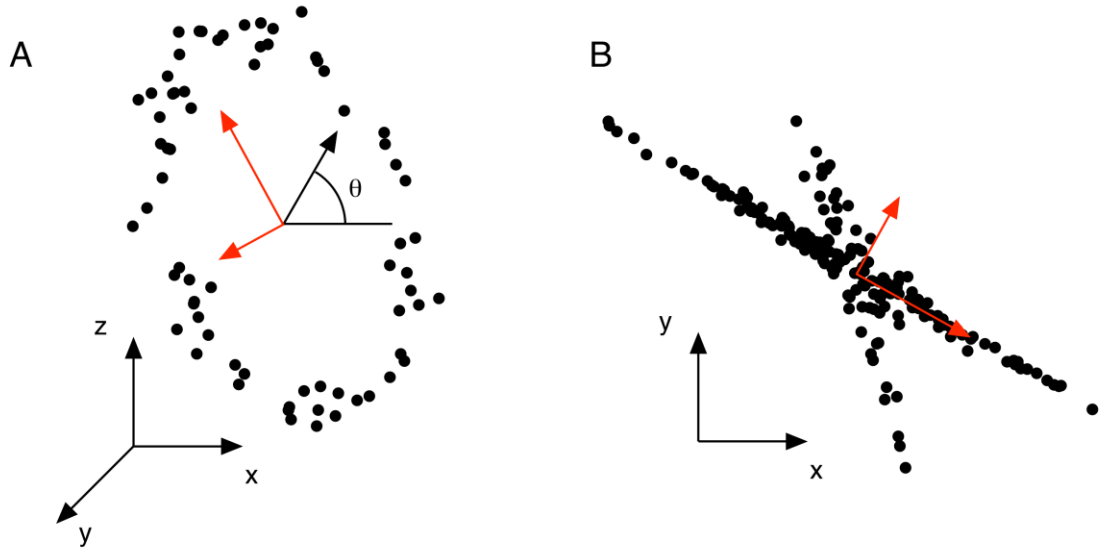
- The idea for dimensionality reduction is to find the k-th principal components ($k < m$), by sorting variances in decreasing order.
- Project the data on these directions and use such data instead of the original ones.
- This data are the best approximation w.r.t the sum of the squared differences.

Principal Component Analysis (PCA)

Idea: look for the most meaningful basis to re-express a dataset

Linearity assumption: re-express the data as a linear combination of its basis vectors

- PCA is non-parametric (a strength point but also a weak point)
- It fails for non-Gaussian distributed data
- It can be extended to account for non-linear transformation (kernel PCA)



(Classical) MultiDimensional Scaling (MDS)

Whereas PCA finds a low-dimensional embedding that best preserves variance, MDS finds an embedding that preserves the interpoint distances

- Find the minimizing linear mapping

$$\mathbf{y}_i = \mathbf{M}\mathbf{x}_i$$

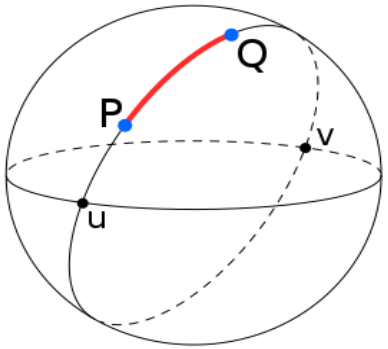
$$\phi(\mathbf{Y}) = \sum_{i,j} d_{ij}^2 - \underbrace{\|\mathbf{y}_i - \mathbf{y}_j\|^2}$$

Euclidean distance in low dimensional space

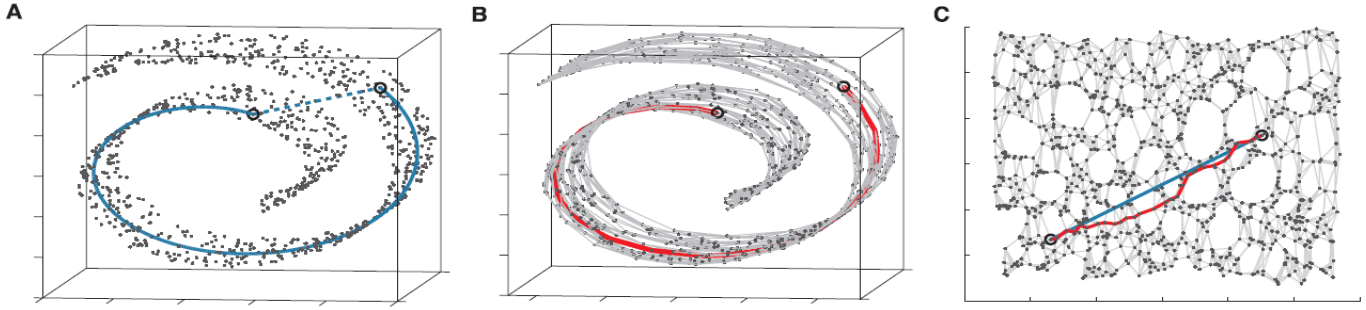
- The true dimensionality of data can be estimated by the decrease in error as the dimensionality of the embedding increases
- Both PCA and MDS better preserve large pairwise distances, but one may want to trust small distances more
- Non-linear methods needed

IsoMap

Core idea: preserving *geodesic* distances between data points, to recover the *global* geometry of a dataset



- Construct a weighted neighbourhood graph
  - connect each point with either all points within a ball of fized radius, or with K-NNs
- Evaluate distances between (faraway) points by computing shortest paths in the graph
- Compute an embedding in a *d*-dimensional Euclidean space by applying MDS to the matrix of graph distances



Euclidean distance *vs* geodesic distance

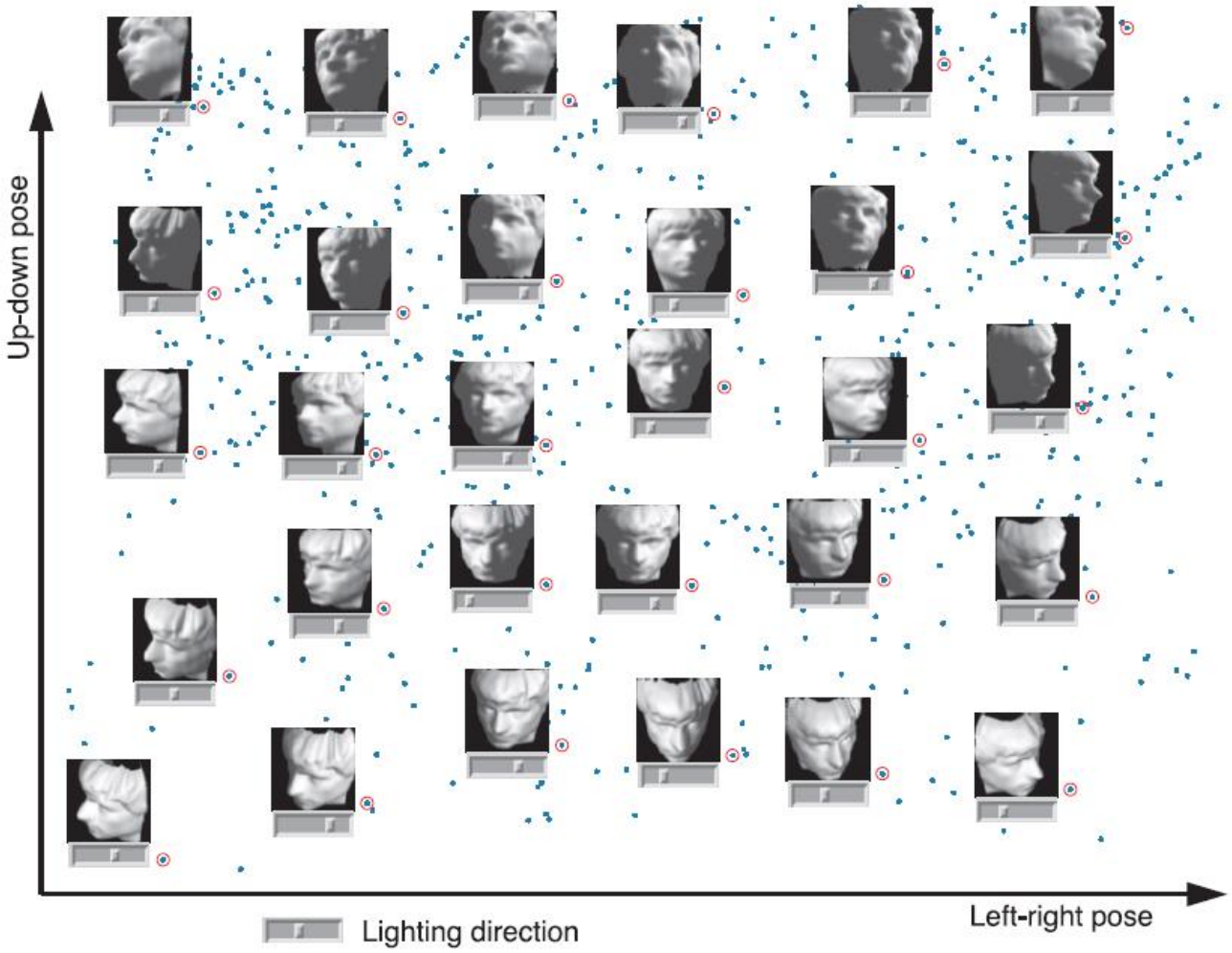
Neighbourhood graph and geodesic distance approximation

Geodesic distance vs approximated geodesic

[J. B. Tenenbaum et al.: A global geometric framework for non-linear dimensionality reduction, 2000]

IsoMap

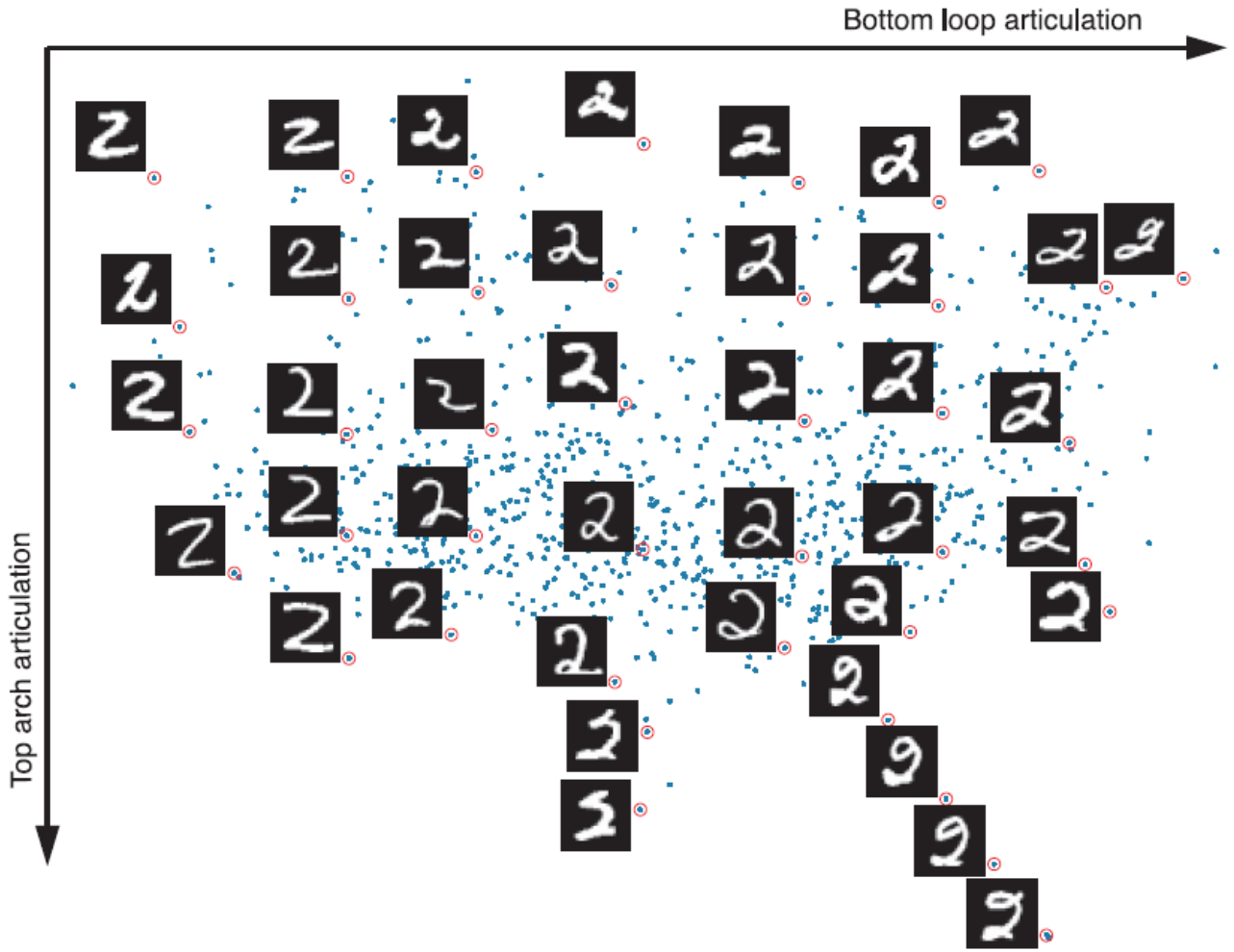
A canonical dimensionality reduction problem from visual perception: each coordinate axis of the embedding highly correlates with one degree of freedom underlying the original data



[J. B. Tenenbaum et al.: A global geometric framework for non-linear dimensionality reduction, 2000]

IsoMap

Application to 1K images in the MNIST dataset: the two most significant dimensions in the IsoMap embedding articulate the major features of the digit



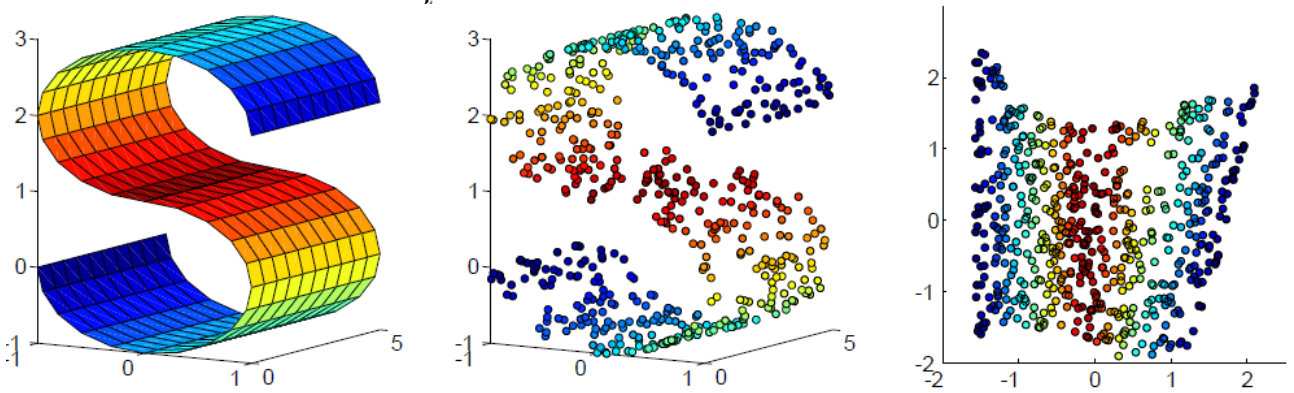
[J. B. Tenenbaum et al.: A global geometric framework for non-linear dimensionality reduction, 2000]

Locally Linear Embedding (LLE)

LLE attempts to discover *nonlinear* structures in high dimension by exploiting local linear approximations

- Intuition: assuming that there is sufficient data (well-sampled manifold) we expect each data point and its neighbors can be approximated by a local linear patch
- Patches are represented by a weighted sum of the K-NNs per data point, or points within a ball (Euclidean distances). The optimal weights are computed by minimizing the reconstruction error.
- The same weights should should also reconstruct its embedded manifold coordinates in a lower-dimensional space
- Construction of a neighbourhood-preserving mapping by minimization of an embedding cost function based on the above idea

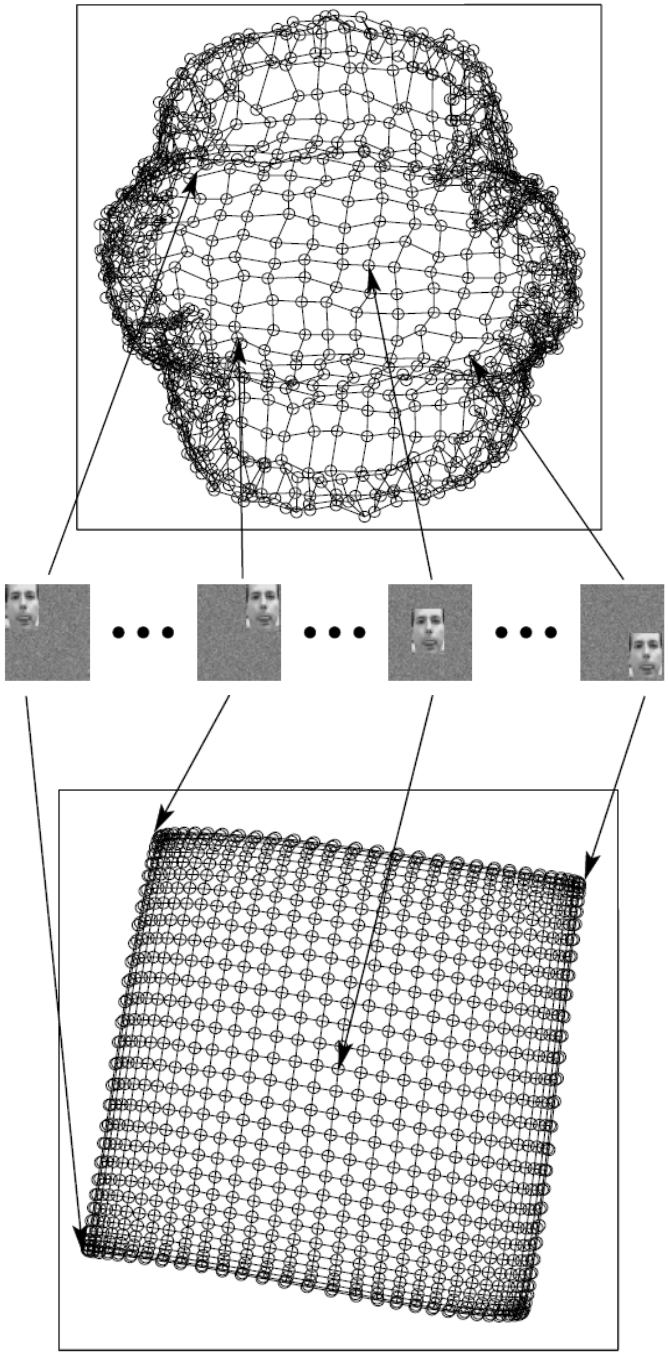
$$\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2$$



[L. K. Saul: An introduction to Locally Linear Embedding]

Locally Linear Embedding (LLE)

LLE attempts to discover *nonlinear* structures in high dimension by exploiting local linear approximations



[PCA (top) vs LLE (bottom), from L. K. Saul: An introduction to Locally Linear Embedding]

Stochastic Neighbour Embedding (SNE)

Technique for visualizing high-dimensional data by giving each data point a location on a 2- or 3-dimensional space

Again, focus on keeping the low-dim representation of similar datapoints close together

- Idea: convert the dataset into a matrix of pairwise similarities, modeled with conditional probabilities
- Conditional probability that the point x_i would peak x_j as its neighbor I neighbors were picked in proportion to their probability density under a Gaussian centered at x_i:

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)}$$

- For the low-dimensional points an analogous conditional probability is used:

$$q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|^2\right)}$$

Stochastic Neighbour Embedding (SNE)

Technique for visualizing high-dimensional data by giving each data point a location on a 2- or 3-dimensional space

Again, focus on keeping the low-dim representation of similar datapoints close together

- The goal is to minimizes the mismatch between $p_{j|i}$ and $q_{j|i}$.
- Using the Kullback-Leibler divergence this goal can be achieved by minimizing the function:

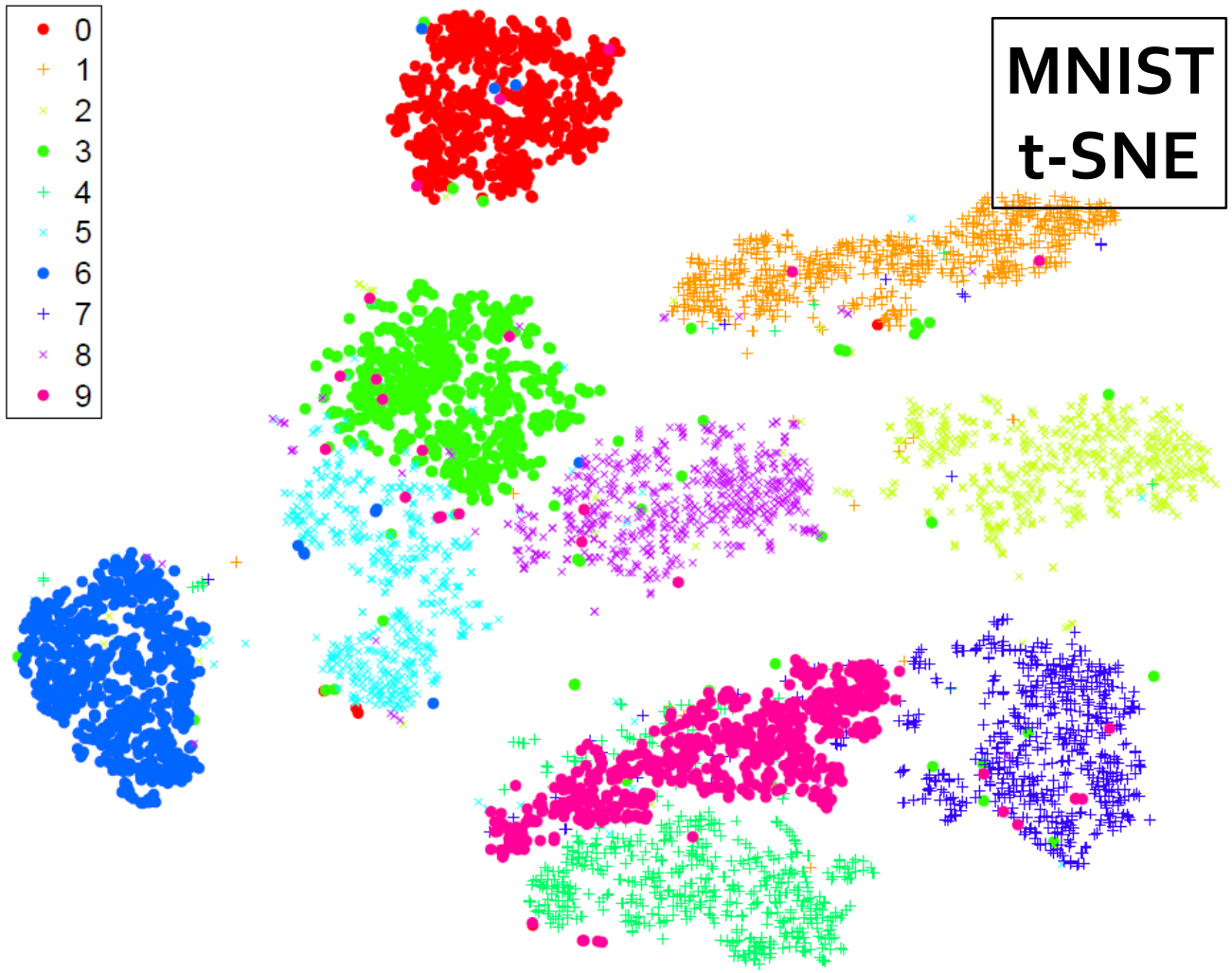
$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$$

- t-SNE variant: SNE made symmetric and different evaluation of distances in the embedding

Stochastic Neighbour Embedding (SNE)

Good for visualization puroposes

Unclear performance on general dimensionality reduction tasks

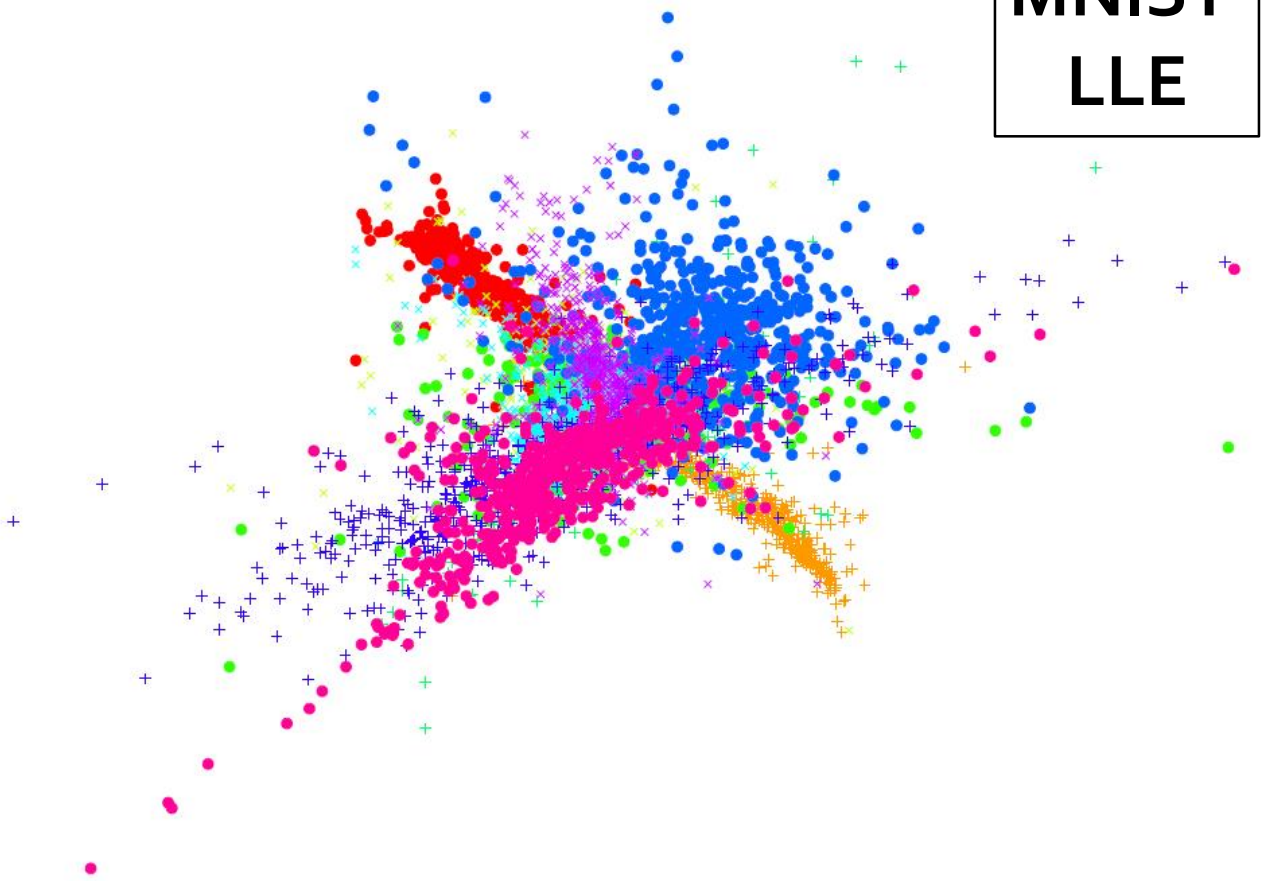


[L.J.P. van der Maaten and G.E. Hinton, "Visualizing High-Dimensional Data Using t-SNE", Journal of Machine Learning Research, Vol. 9, pp. 2579-2605, 2008]

Stochastic Neighbour Embedding (SNE)

Good for visualization puroposes

Unclear performance on general dimensionality reduction tasks

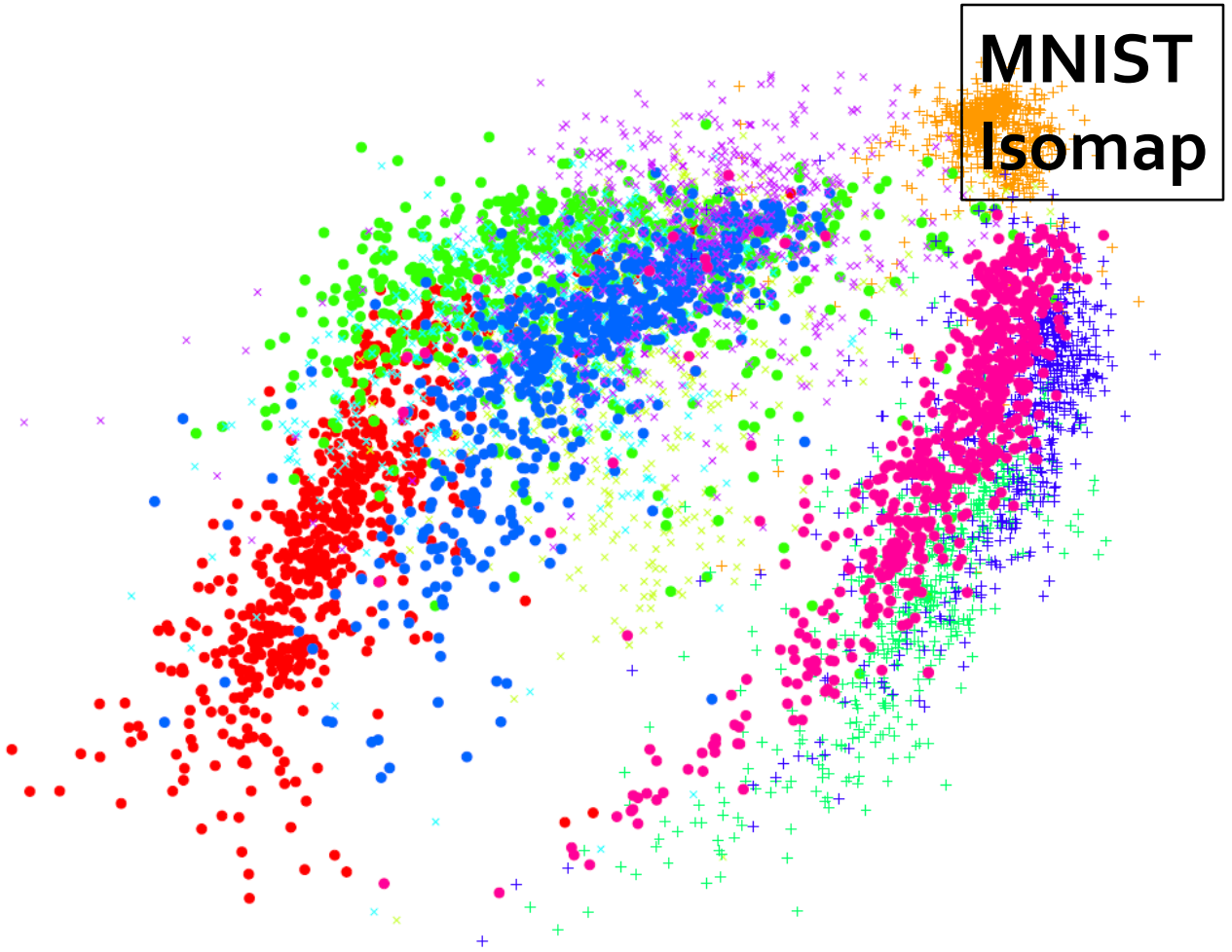


[L.J.P. van der Maaten and G.E. Hinton, "Visualizing High-Dimensional Data Using t-SNE", Journal of Machine Learning Research, Vol. 9, pp. 2579-2605, 2008]

Stochastic Neighbour Embedding (SNE)

Good for visualization puroposes

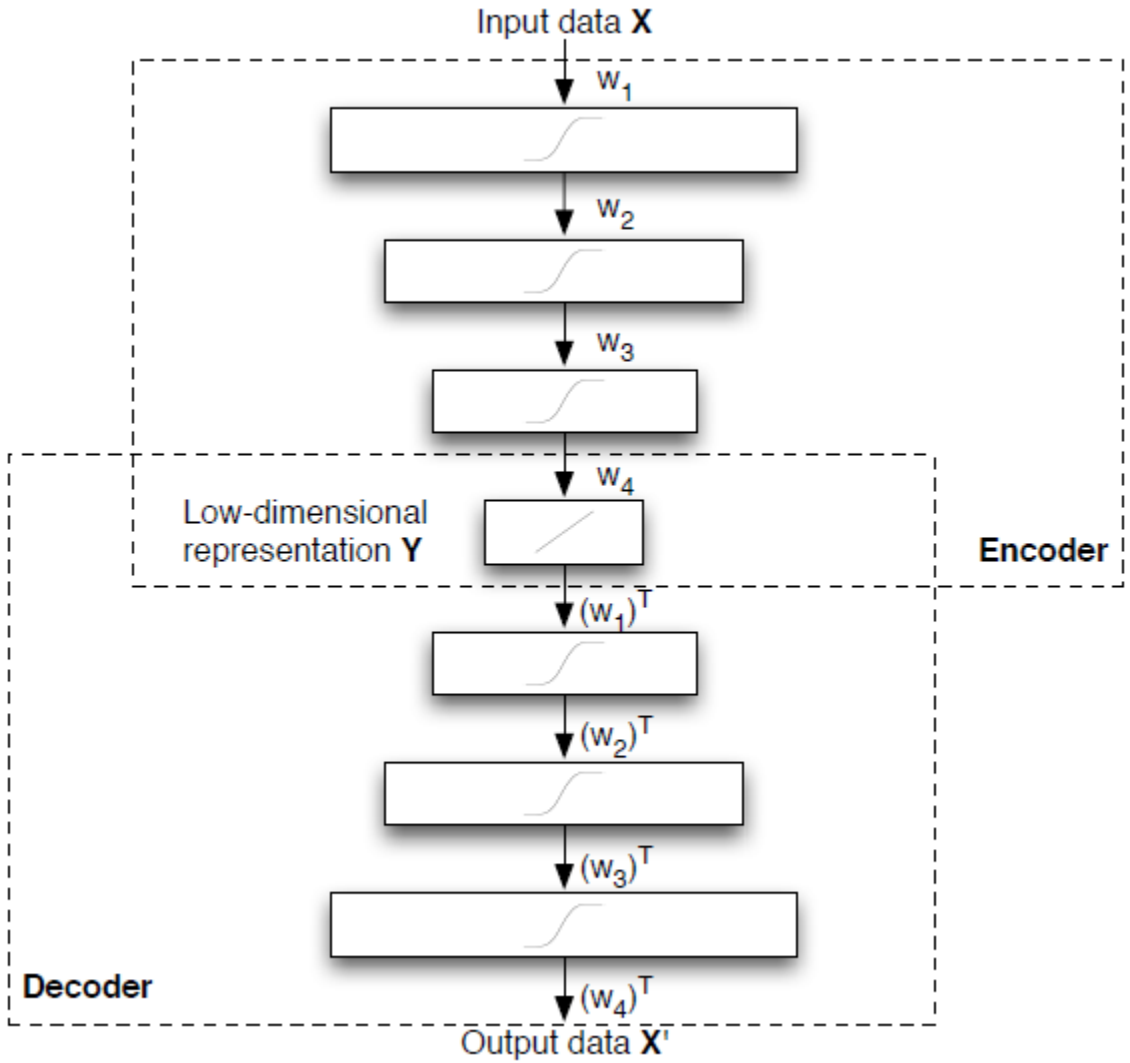
Unclear performance on general dimensionality reduction tasks



[L.J.P. van der Maaten and G.E. Hinton, "Visualizing High-Dimensional Data Using t-SNE", Journal of Machine Learning Research, Vol. 9, pp. 2579-2605, 2008]

Autoencoders

Autoencoders are types of neural networks that can perform dimensionality reduction

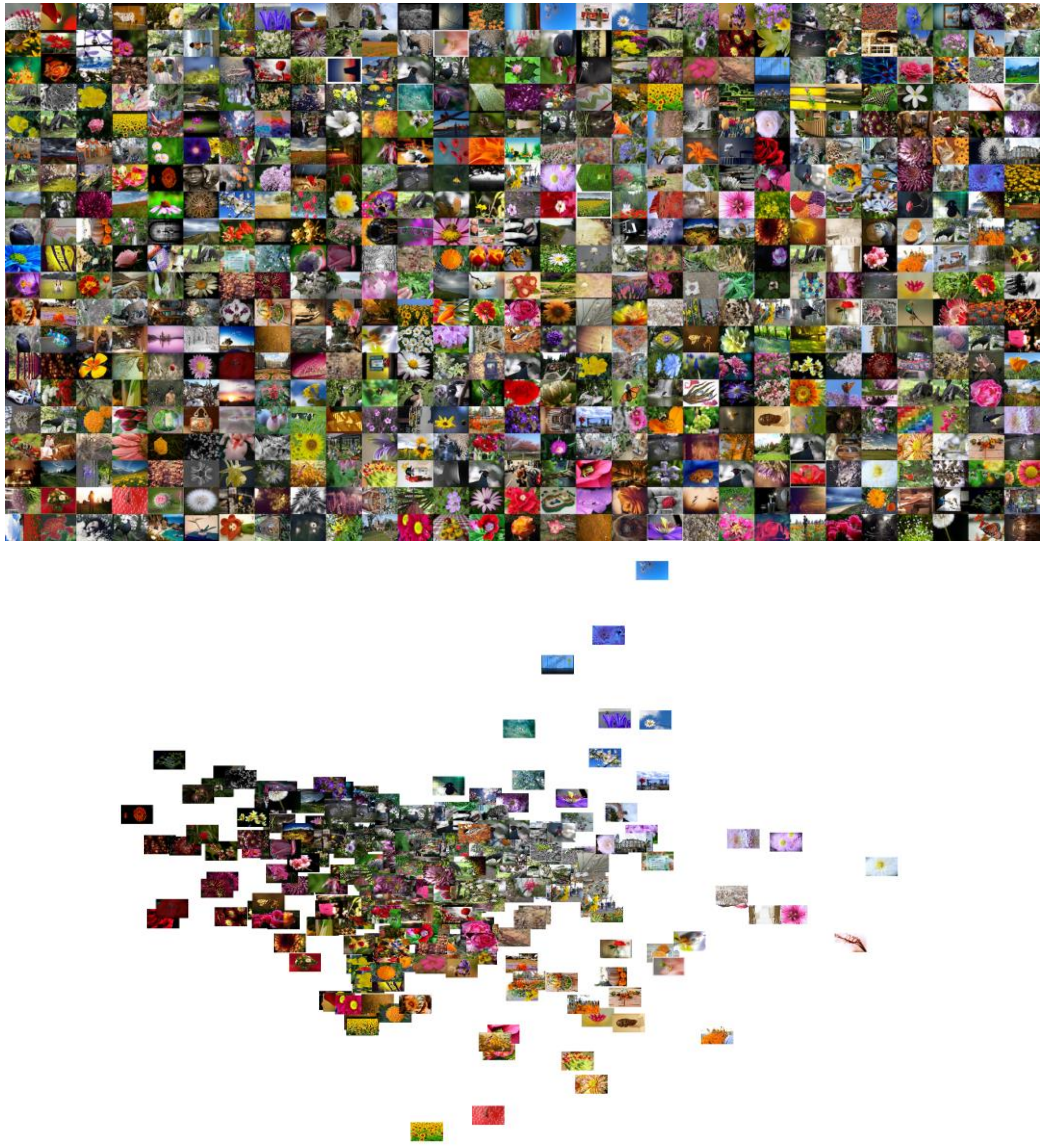


Object arrangement

Object arrangement

Motivation: dimensionality reduction can help visualizing objects in 2D or 3D while preserving pairwise distances, but the final placement is arbitrary

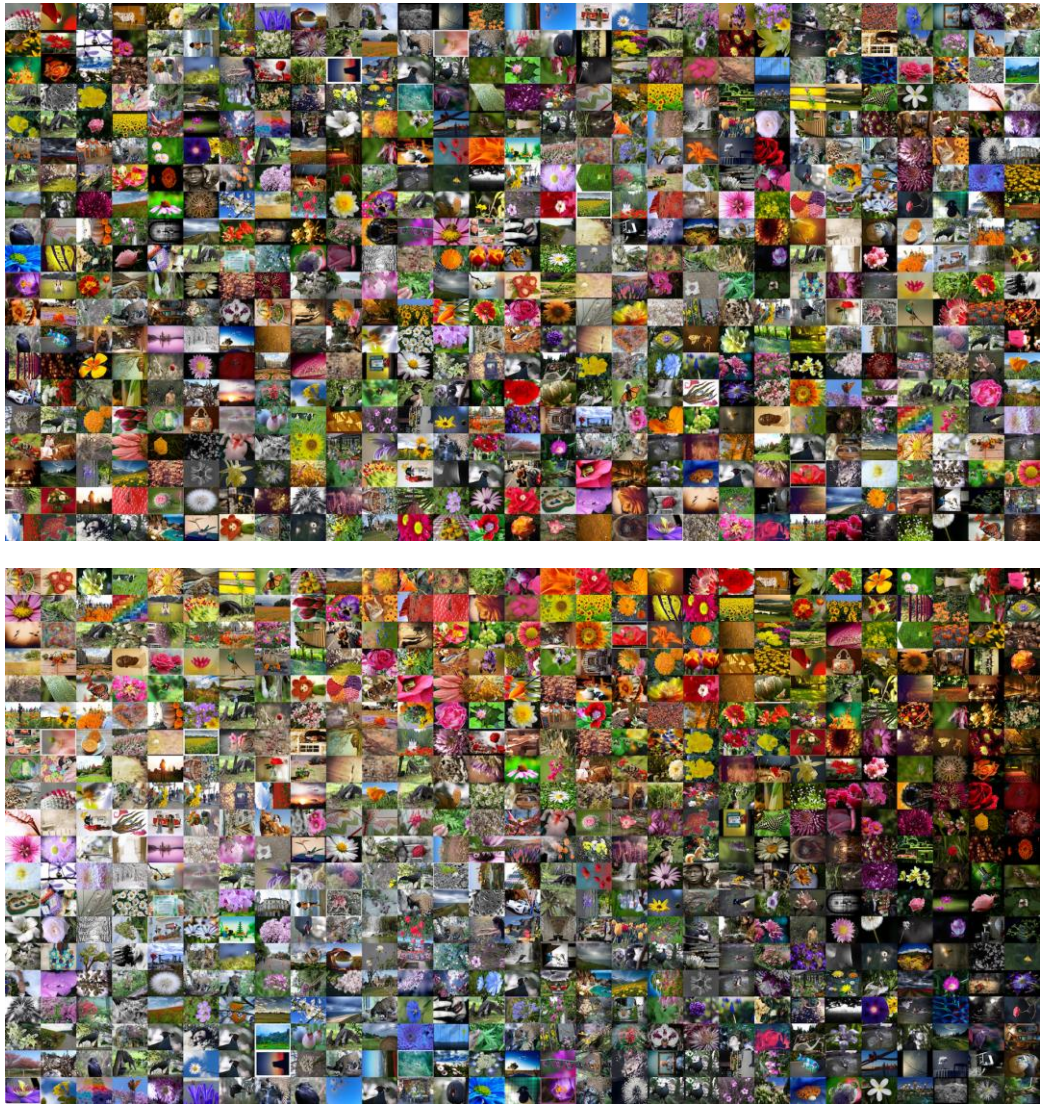
Many applications instead require to place the objects in a set of pre-defined, discrete, positions (e.g. on a grid)



Object arrangement

Motivation: dimensionality reduction can help visualizing objects in 2D or 3D while preserving pairwise distances, but the final placement is arbitrary

Many applications instead require to place the objects in a set of pre-defined, discrete, positions (e.g. on a grid)



IsoMatch

Takes as input a set of items together with a distance matrix, where each item is to be assigned to a set of spatial locations

- Problem statement: find the permutation (and scale) that minimize the energy

$$E_p(\pi) = \min \left(\sum_{i,j} cd(i,j) - d(\pi(i), \pi(j)) \right)^{\frac{1}{p}}$$

Permutation

Original pairwise distance

Euclidean distance in the grid

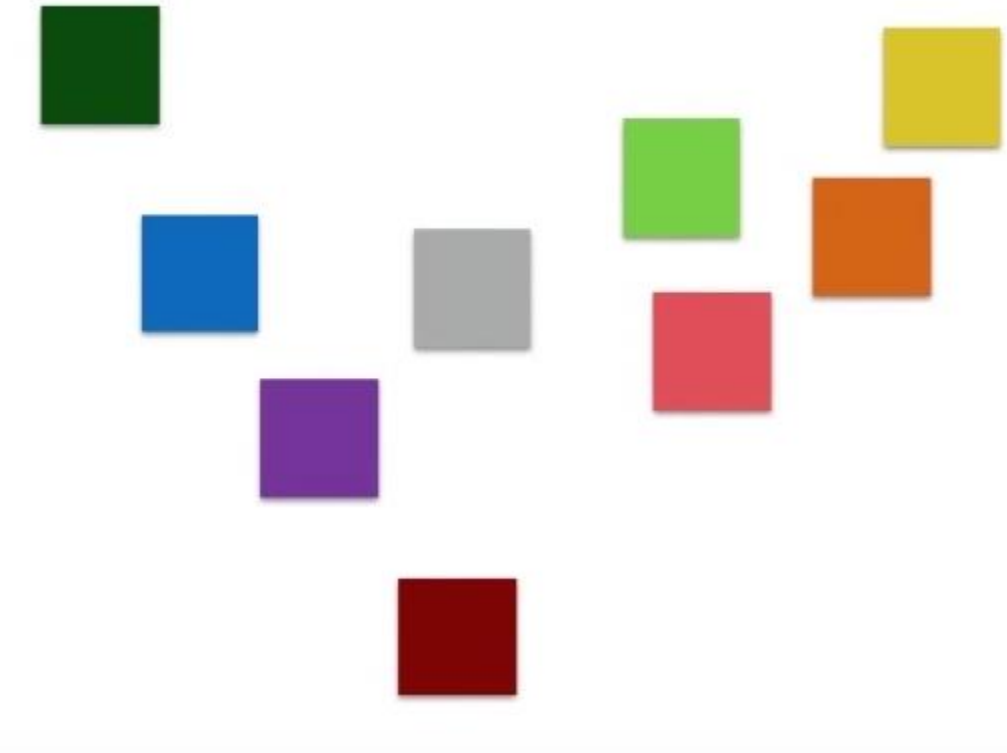
IsoMatch

Step I : Dimensionality Reduction (using Isomap)

Step II : Coarse Alignment (bounding box for the target arrangement)

Step III : Bipartite Matching

Step IV (optional) : Random Refinement (elements swap)



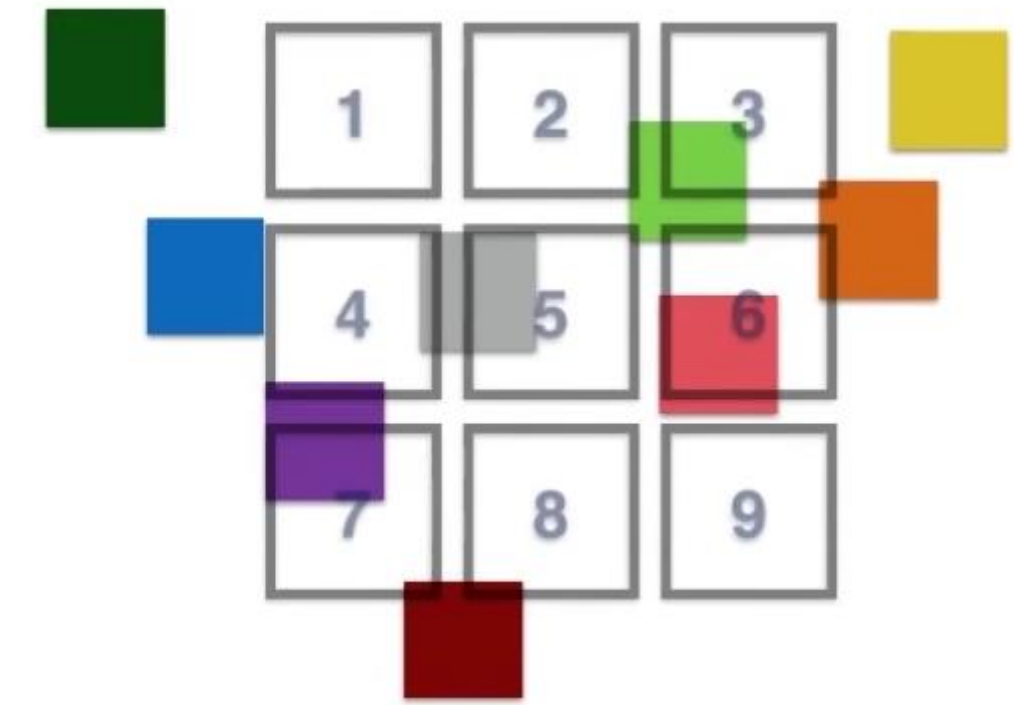
IsoMatch

Step I : Dimensionality Reduction (using Isomap)

Step II : Coarse Alignment (bounding box for the target arrangement)

Step III : Bipartite Matching

Step IV (optional) : Random Refinement (elements swap)



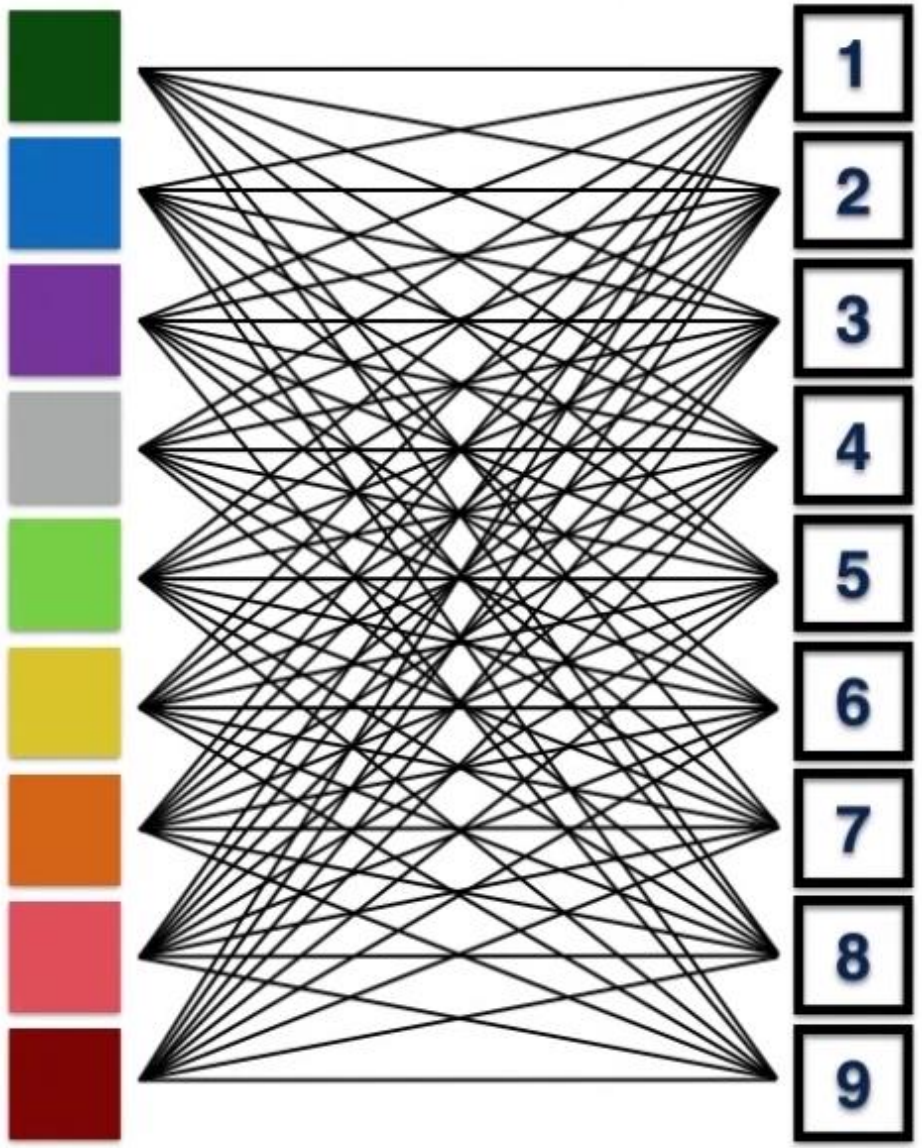
IsoMatch

Step I : Dimensionality Reduction (using Isomap)

Step II : Coarse Alignment (bounding box for the target arrangement)

Step III : Bipartite Matching

Step IV (optional) : Random Refinement (elements swap)



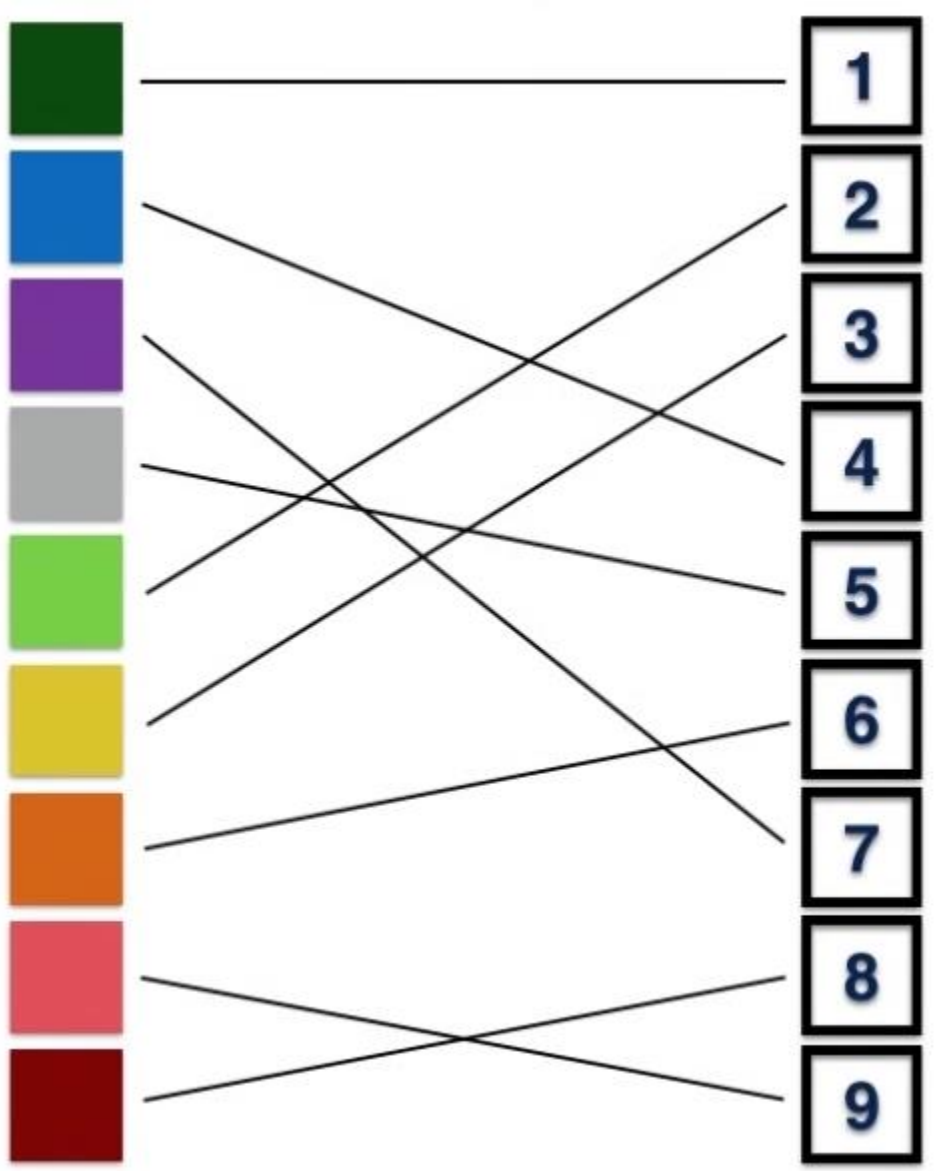
IsoMatch

Step I : Dimensionality Reduction (using Isomap)

Step II : Coarse Alignment (bounding box)

Step III : Bipartite Matching

Step IV (optional) : Random Refinement (elements swap)



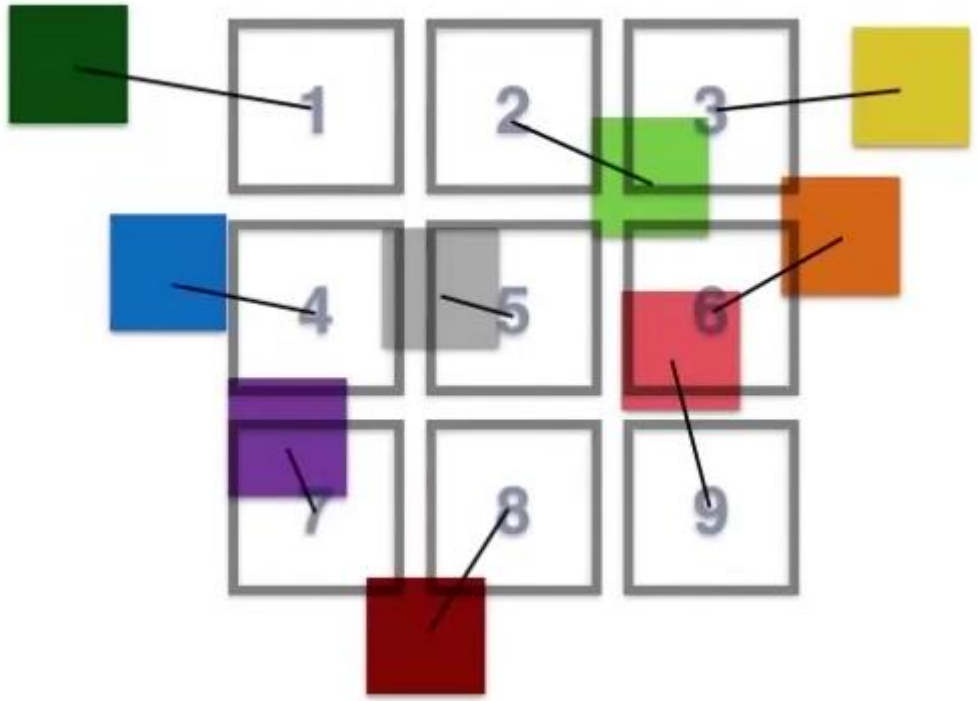
IsoMatch

Step I : Dimensionality Reduction (using Isomap)

Step II : Coarse Alignment (bounding box)

Step III : Bipartite Matching

Step IV (optional) : Random Refinement (elements swap)



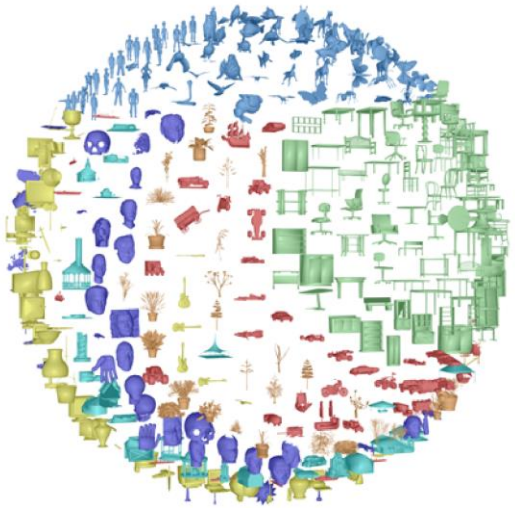
IsoMatch

[Average colors]

IsoMatch

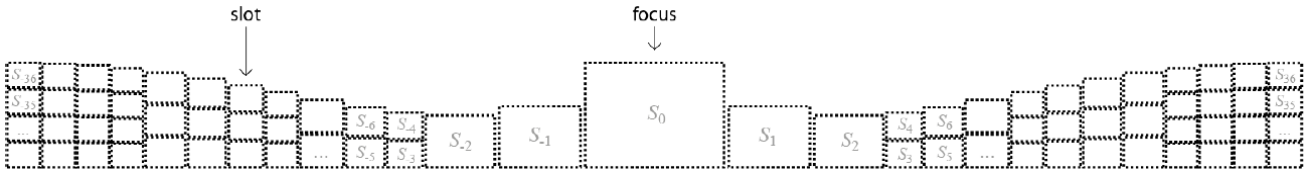
[Word similarity]

IsoMatch



IsoMatch

Pilebars



An alternative to regular grids

A new type of thumbnail bar, according to the *focus* + *context* paradigm

Thumbnails are dynamically rearranged, resized and reclustered adaptively during browsing, while ensuring smooth transitions

Any pairwise distance can be used

Example application: navigation of registered photographs

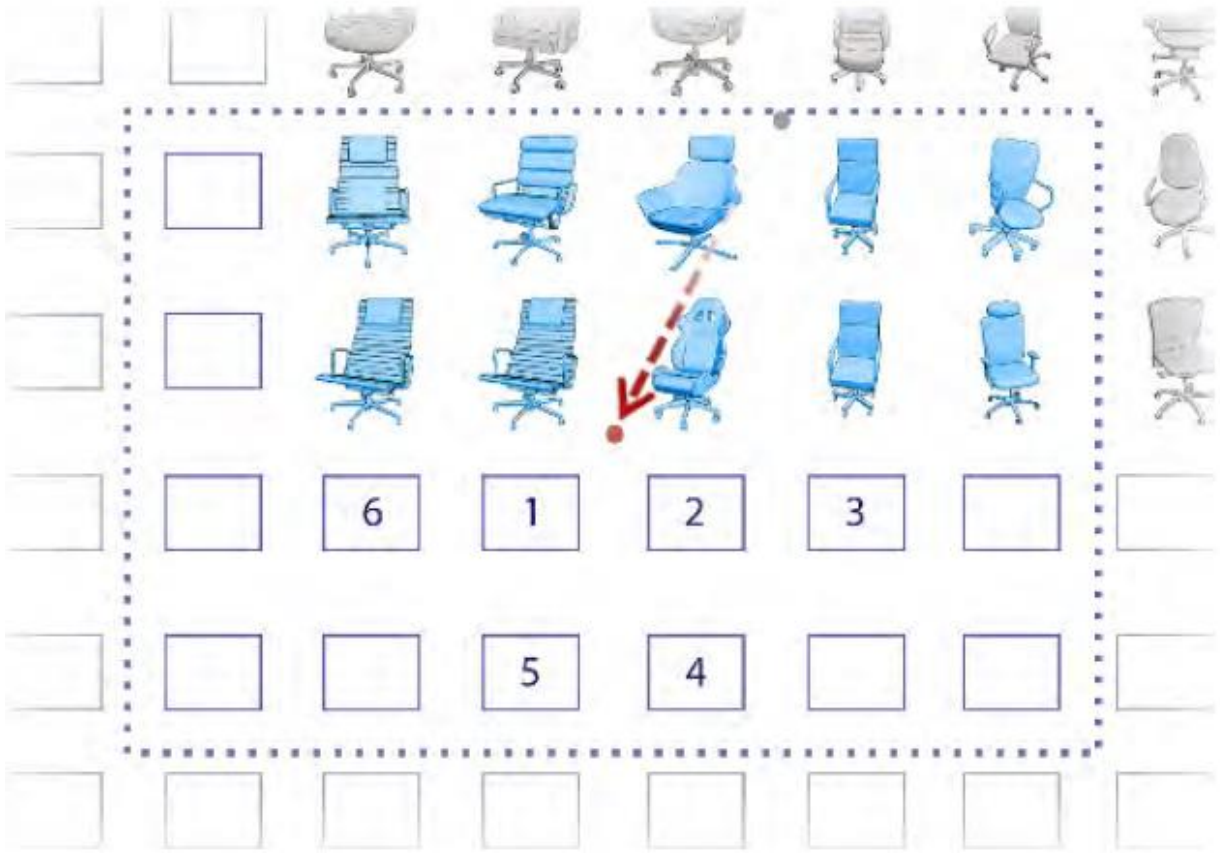
http://vcg.isti.cnr.it/photocloud

Browsing 3D collections

Motivation: Similarity between 3D models exists in a high dimensional space which cannot be accurately expressed in a two dimensional map.

Arrange the objects in a 3D dataset in a dynamic map that the user can spatially navigate

Similar shapes are placed next to each other. A local map with pan capabilities is provided, and a user interface that resembles an online experience of navigating through geographical maps. As the user navigates through the map, new shapes new shapes appear which correspond to the specific navigation tendencies and interests of the user



Kleiman et al.: Dynamic maps for exploring and browsing shapes. Computer Graphics Forum 32(5), 2013]

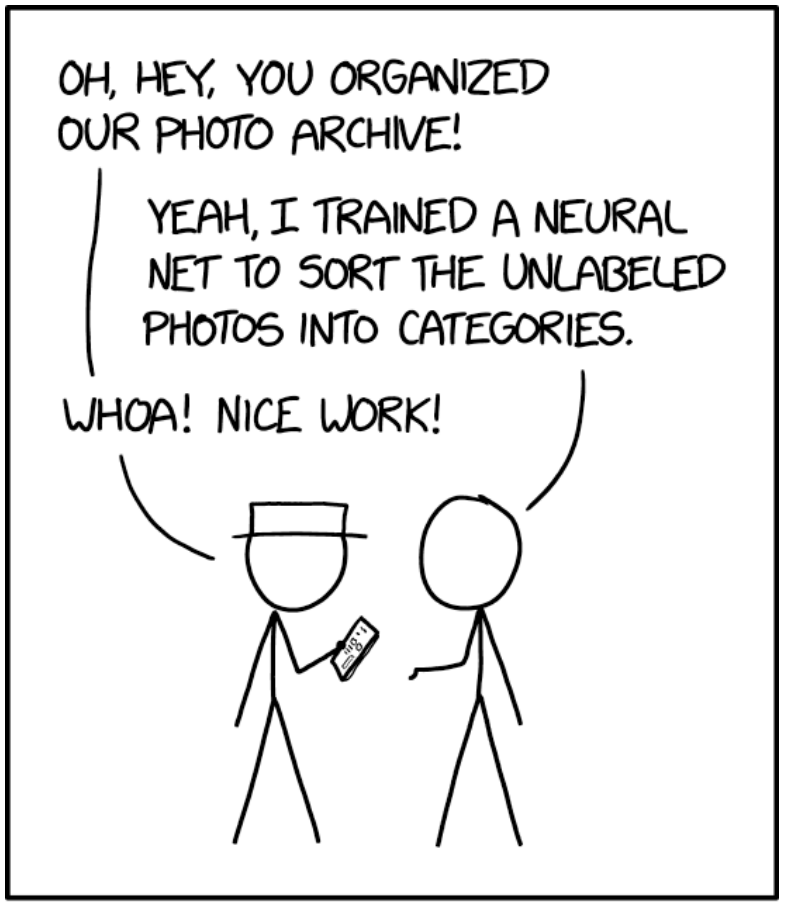
Browsing 3D collections

Once a query is selected, the rest of the shapes are automatically repositioned to form concentric circles around the selected one (shapes closer to the query are located at the inner circles)

It is a *Degree of Interest* (DOI) visualization technique, which magnify or highlight items of interest along with a subset of items which may provide explanatory context

[Huang et al.: Qualitative organization of collections of shapes via quartet analysis. ACM ToG 32,4 (2013)]

[*«I don't know what I'm looking for, but I'll know it when I find it »*]



ENGINEERING TIP:
WHEN YOU DO A TASK BY HAND, YOU CAN TECHNICALLY SAY YOU TRAINED A NEURAL NET TO DO IT.

[https://xkcd.com/2173/]

Questions?