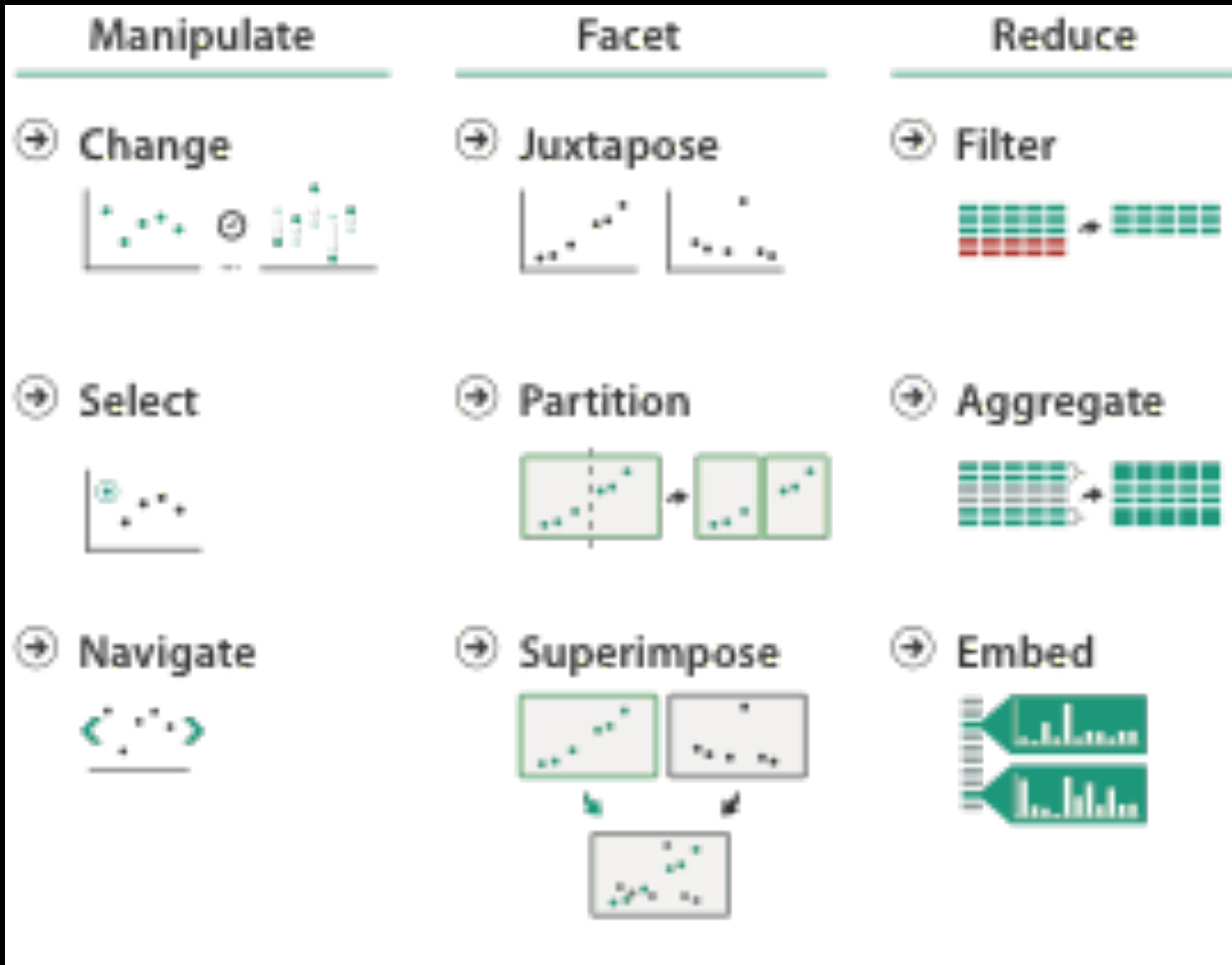
Pontificia Universidad Católica de Chile

Reducir

Visualización de Información
IIC2026

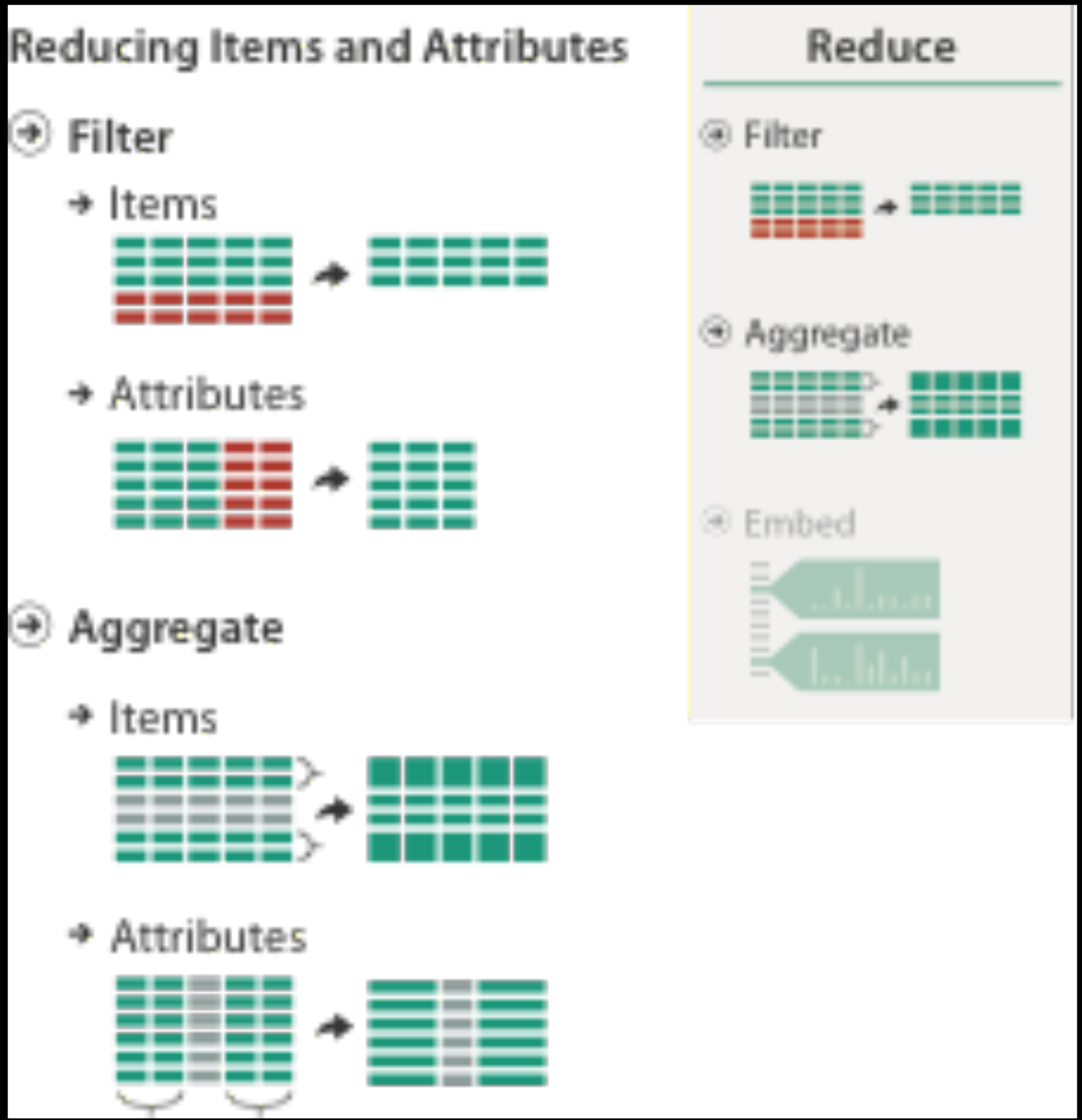
Profesor: Denis Parra

Elecciones de diseño de idioms: Parte 2



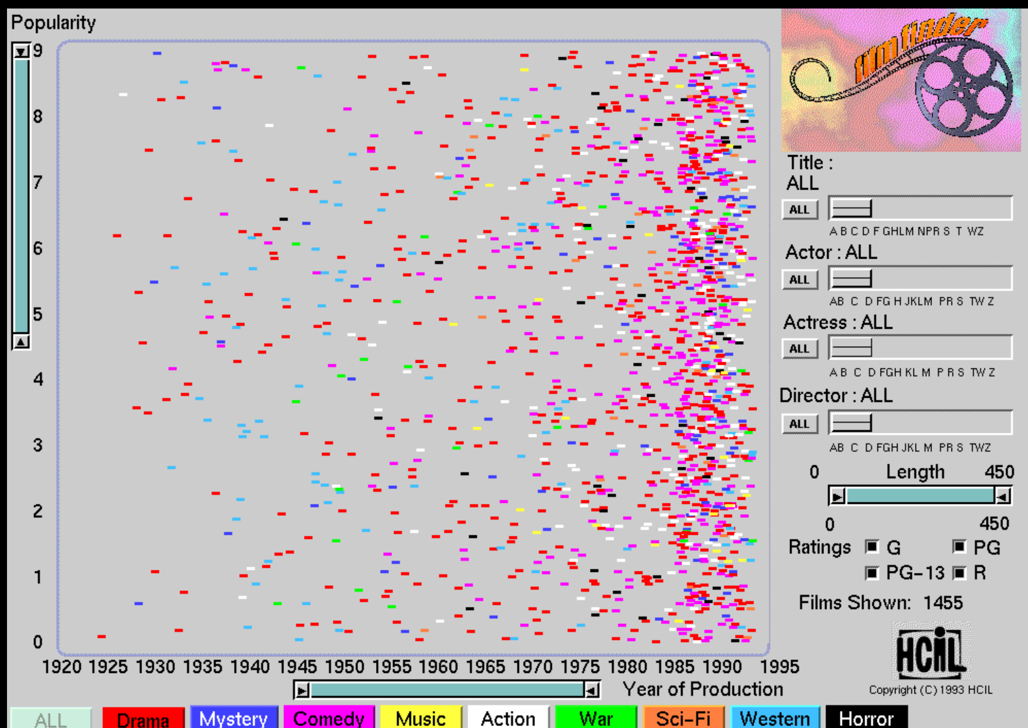
Reducir ítems y atributos

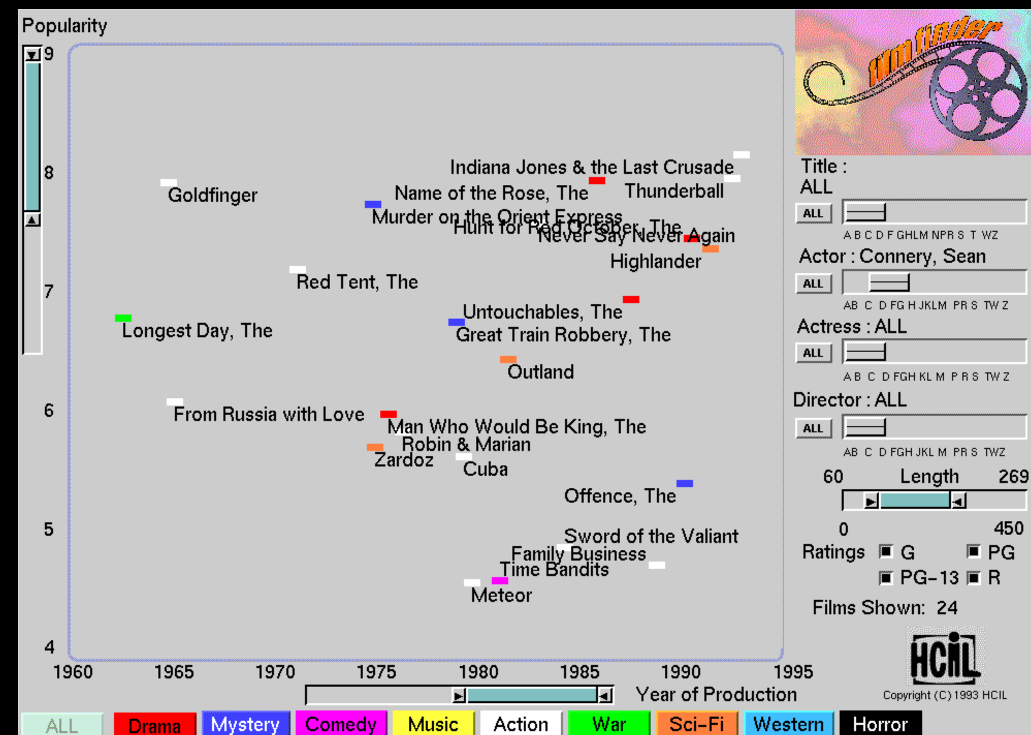
- Filtrar
  - pro: directo e intuitivo
    - Para comprender y calcular
  - con: "out of sight, out of mind"
- Agregación
  - pro: informa de todos los datos
  - con: difícil no perder información
- No mutuamente exclusivos
  - combinar filtrar, agregar
  - combinar reducir, cambiar, facet



Idiom: **filtrado dinámico** — Sistema: **FilmFinder**

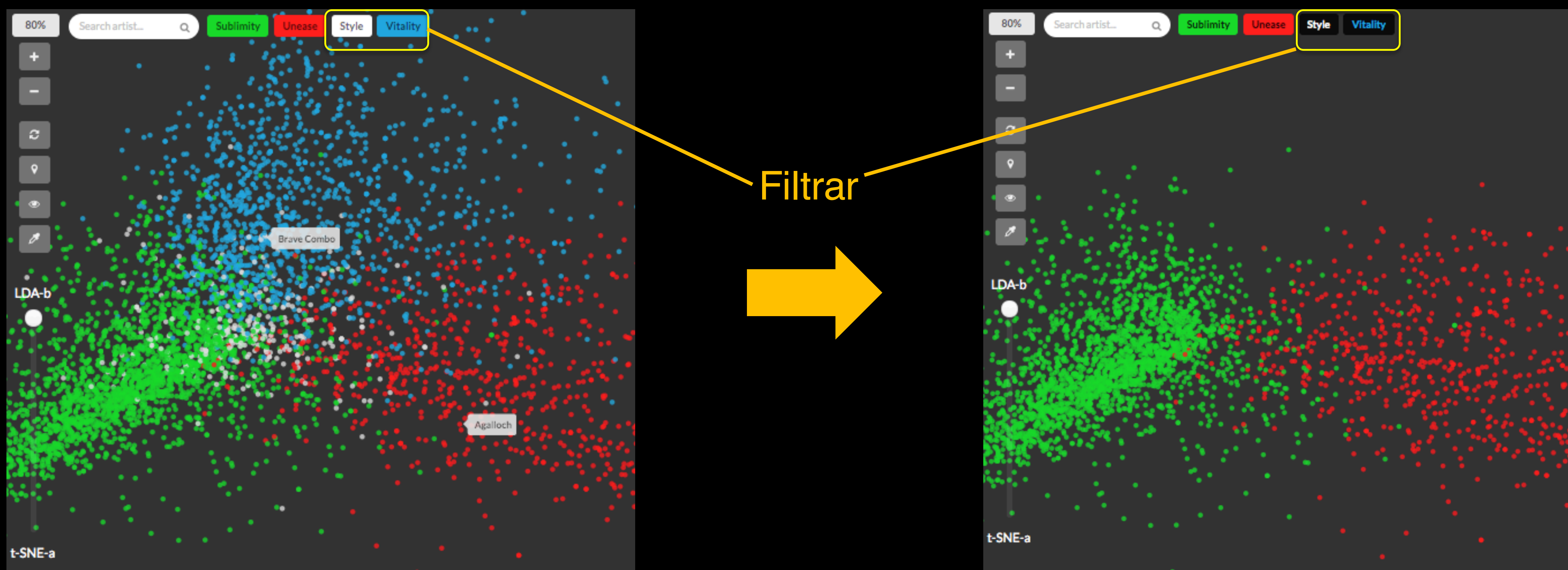
- Filtrado de ítems.
- "browse" cercanamente conectado con interacción.
  - Alternativa a "queries" que pueden retornar muchos o pocos





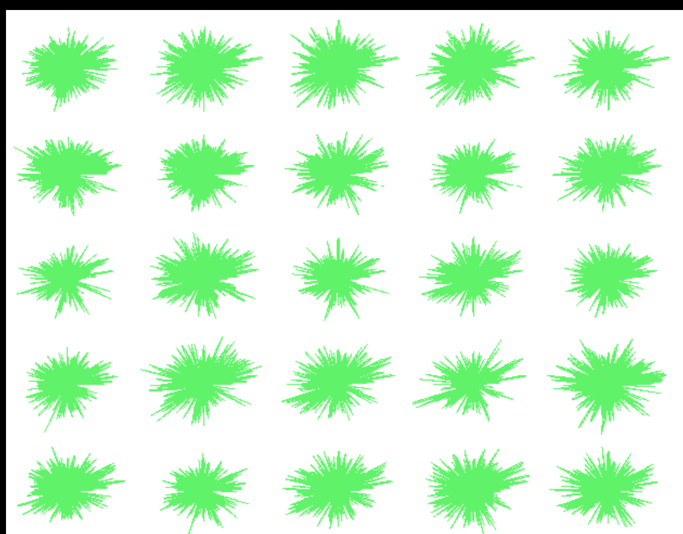
Idiom: **filtrado dinámico** Sistema: **Moodplay**

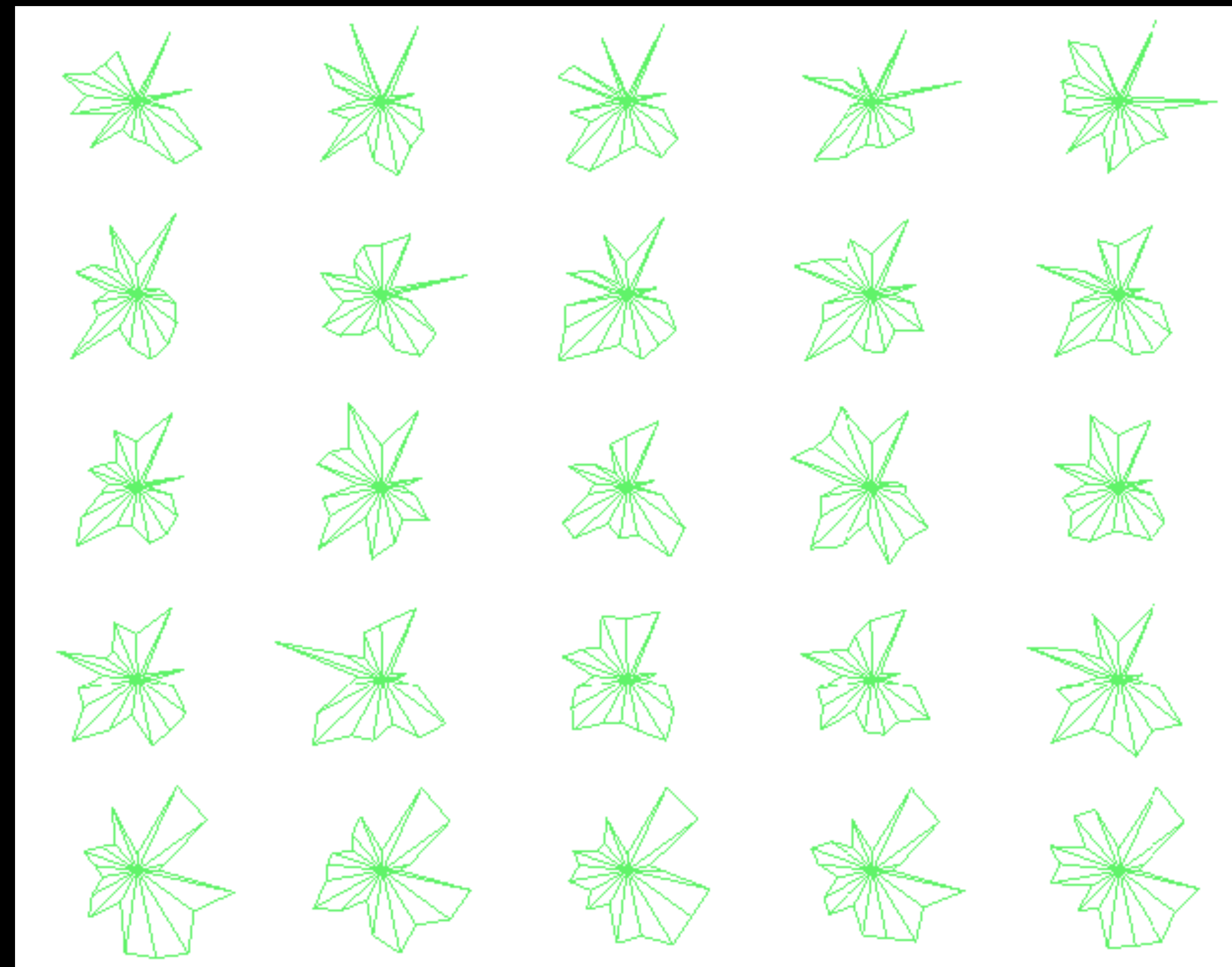
- Filtrado de artistas en base a "estado de ánimo" (mood)



Idiom: **DOSFA**

- Filtrado de atributos
- encoding: glifos de estrella

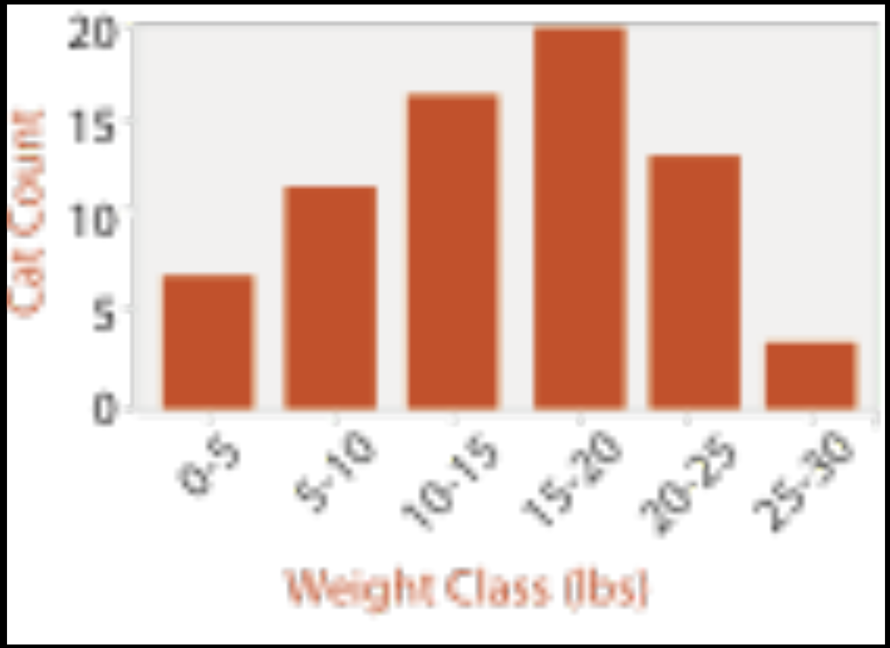




[Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration Of High Dimensional Datasets. Yang, Peng, Ward, and. Rundensteiner. Proc. IEEE Symp. Information Visualization (InfoVis), pp. 105–112, 2003.]

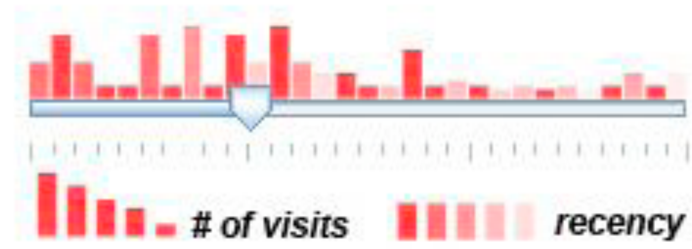
Idiom: **histograma**

- Agregación de ítems estática
- Tarea: encontrar distribución
- Datos: tabular
- Datos derivados
  - nueva tabla: llaves son bins, valores son cuentas
- Crucial: tamaño del bin
  - Patrón puede cambiar dramáticamente dependiendo del nivel de discretización
  - Oportunidad de interacción: controlar el tamaño del bin

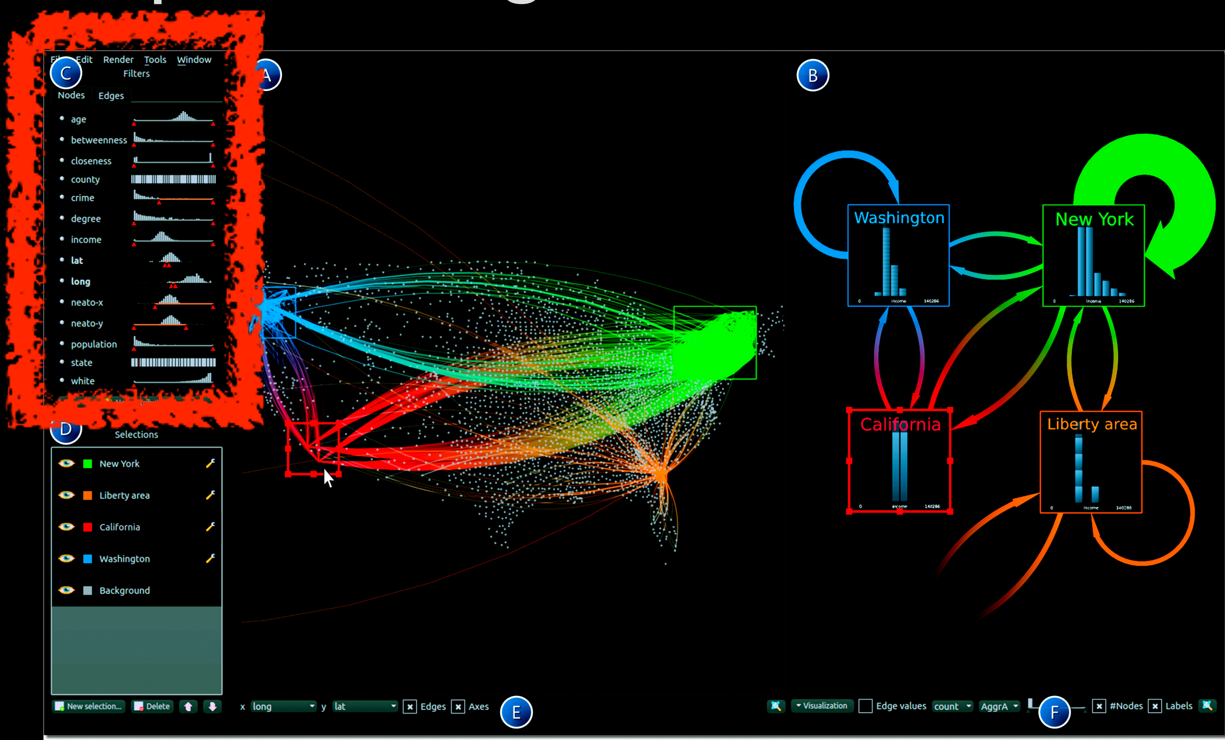


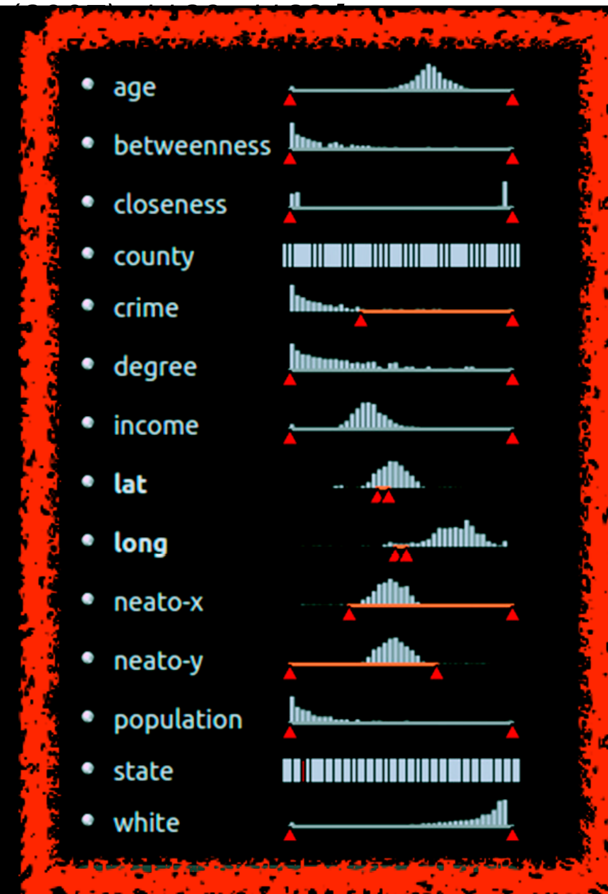
Idiom: **scented widgets**

- augmented widgets show information scent
  - cues to show whether value in drilling down further vs looking elsewhere
- concise use of space: histogram on slider



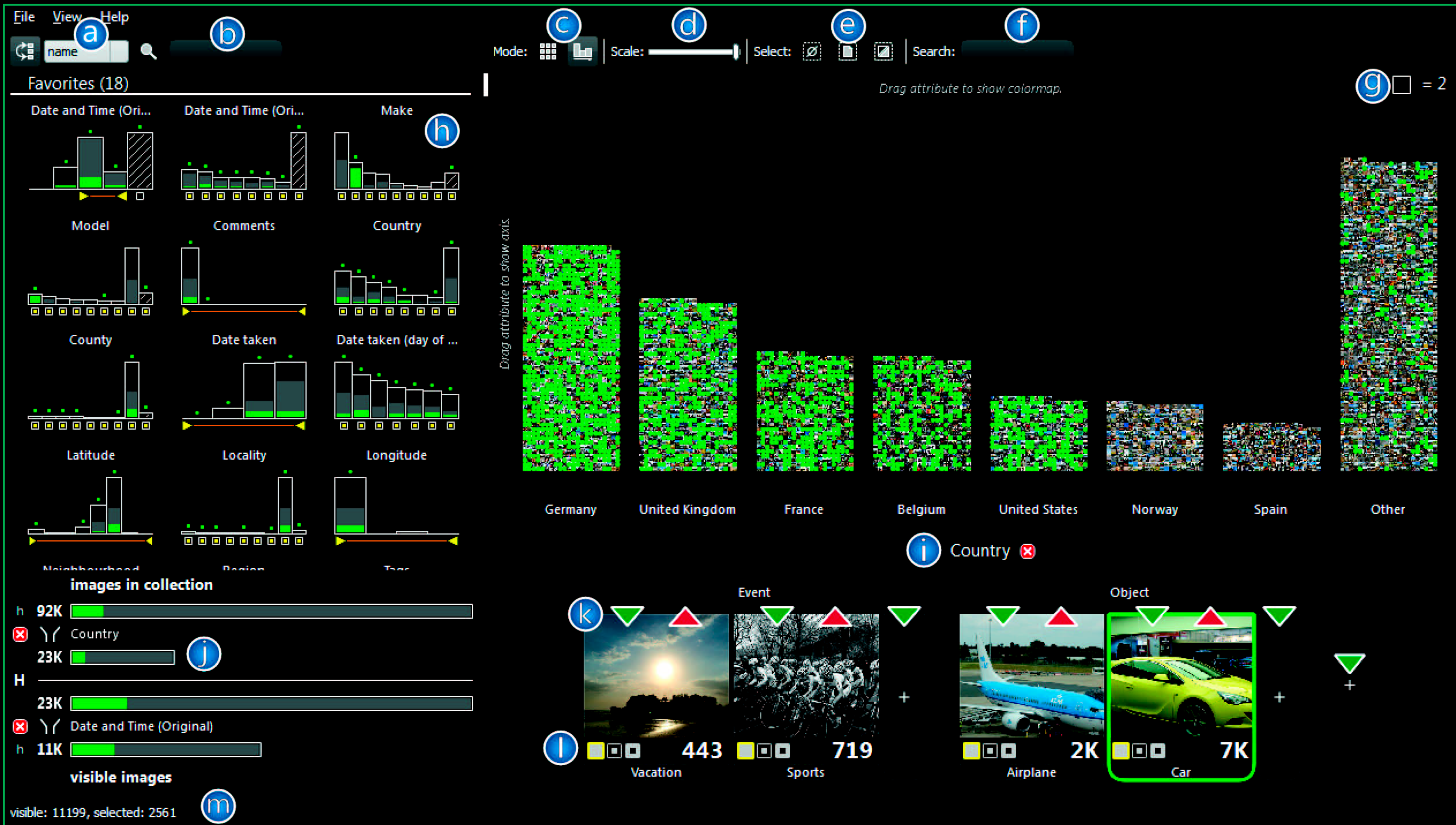
[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE TVCG (Proc. InfoVis 2007) 13:6





[Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. van den Elzen, van Wijk, IEEE TVCG 20(12): 2014 (Proc. InfoVis 2014).]

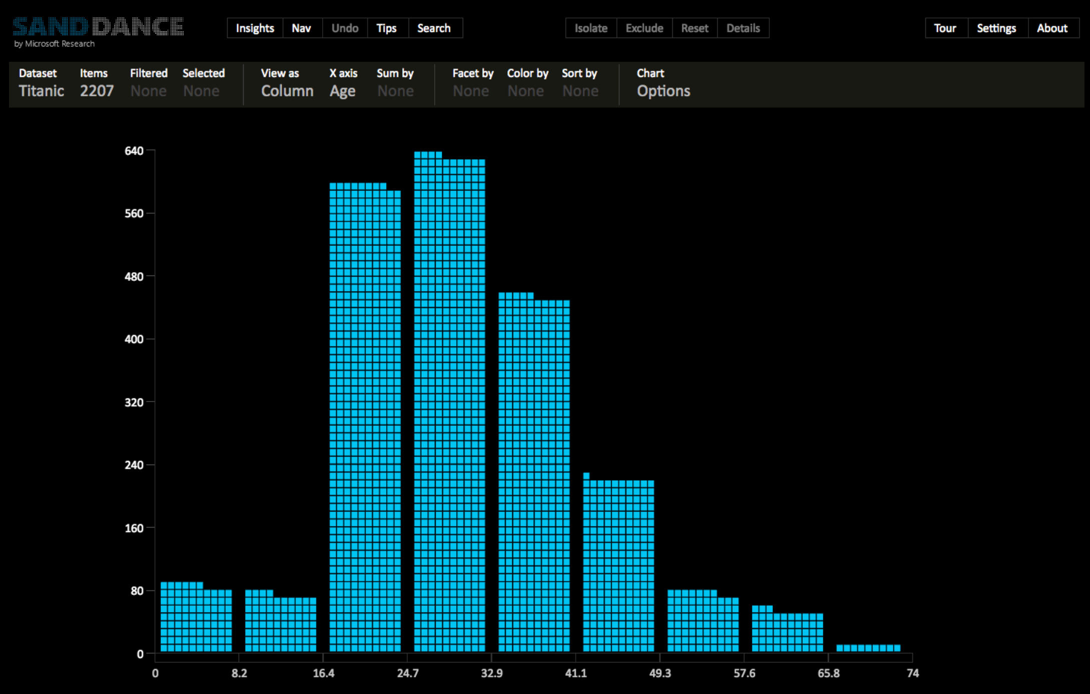
Scented histogram bisliders: detailed

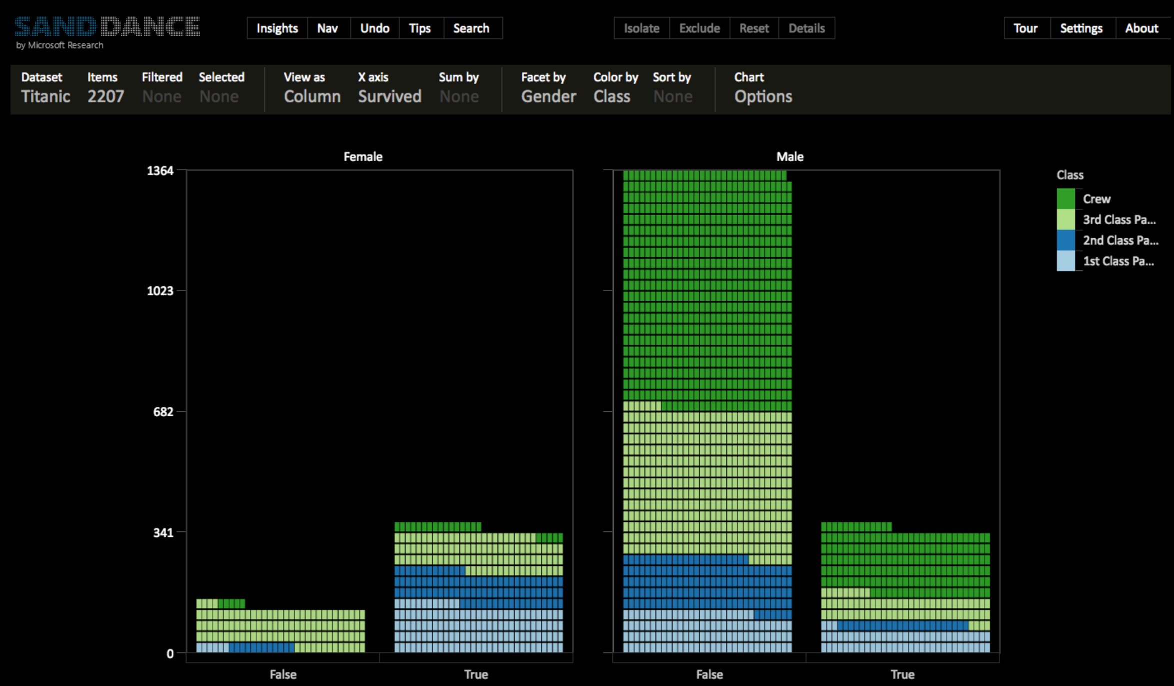


[ICLIC: Interactive categorization of large image collections. van der Corput and van Wijk. Proc. PacificVis 2016.]

Sistema: Microsoft SandDance

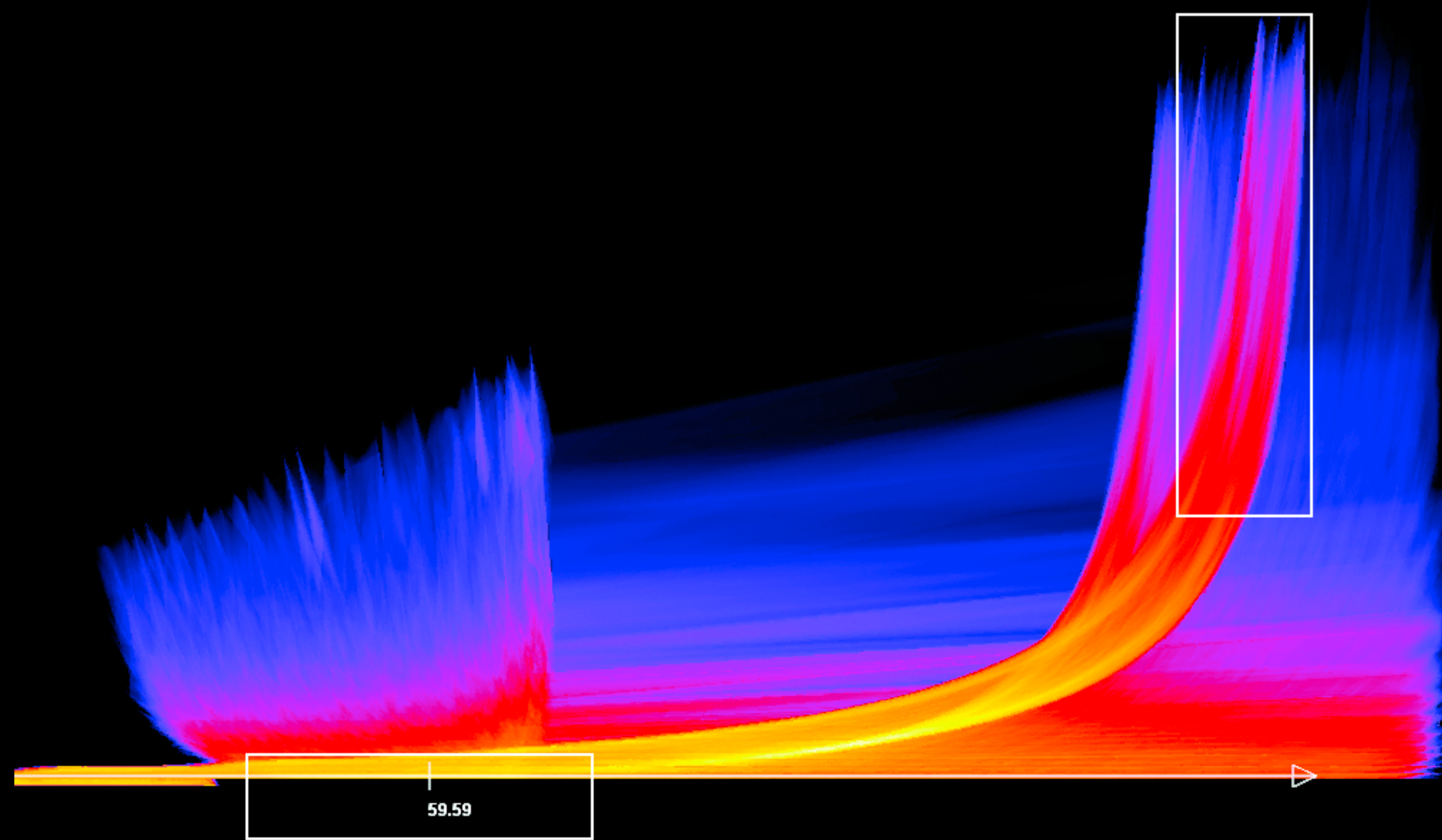
- https://sanddance.azurewebsites.net/BeachPartyApp/BeachPartyApp.html





Continuous scatterplot

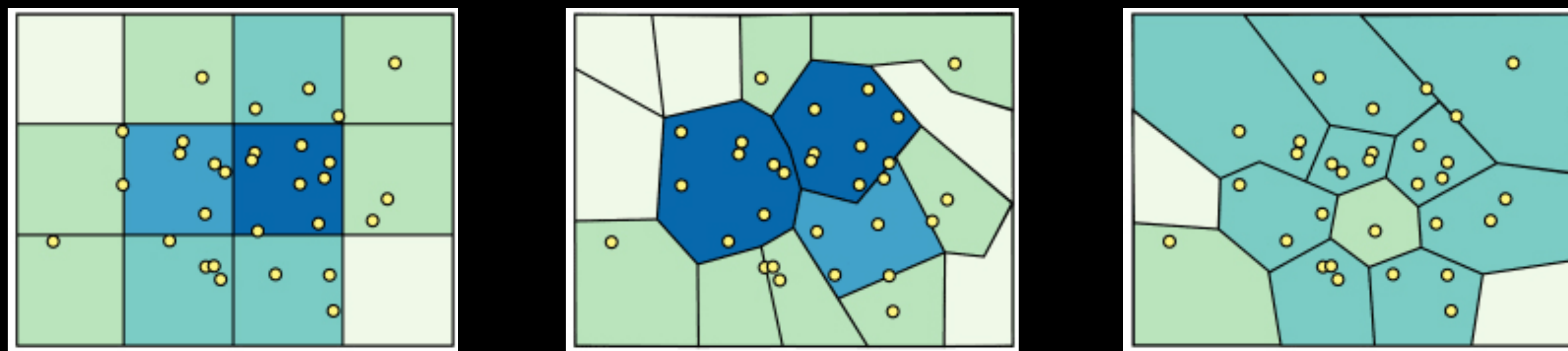
- static item aggregation
- data: table
- derived data: table
  - key attribs x,y for pixels
  - quant attrib: overplot density
- dense space-filling 2D matrix
- color: sequential categorical hue + ordered luminance colormap



[Continuous Scatterplots. Bachthaler and Weiskopf. IEEE TVCG (Proc. Vis 08) 14:6 (2008), 1428–1435. 2008.]

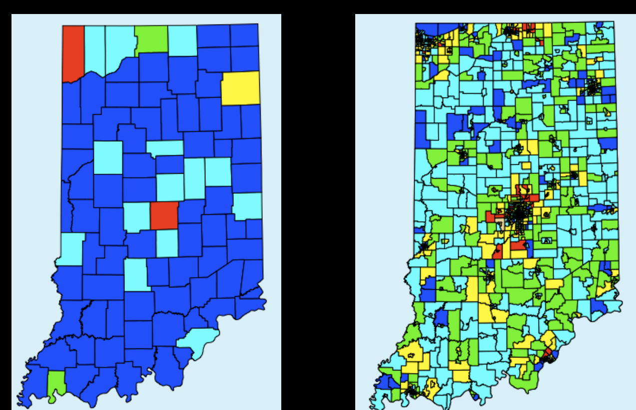
Spatial aggregation

- MAUP: Modifiable Areal Unit Problem
  - gerrymandering (manipulating voting district boundaries) is only one example!
  - zone effects



[http://www.e-education.psu/edu/geog486/l4_p7.html, Fig 4.cg.6]

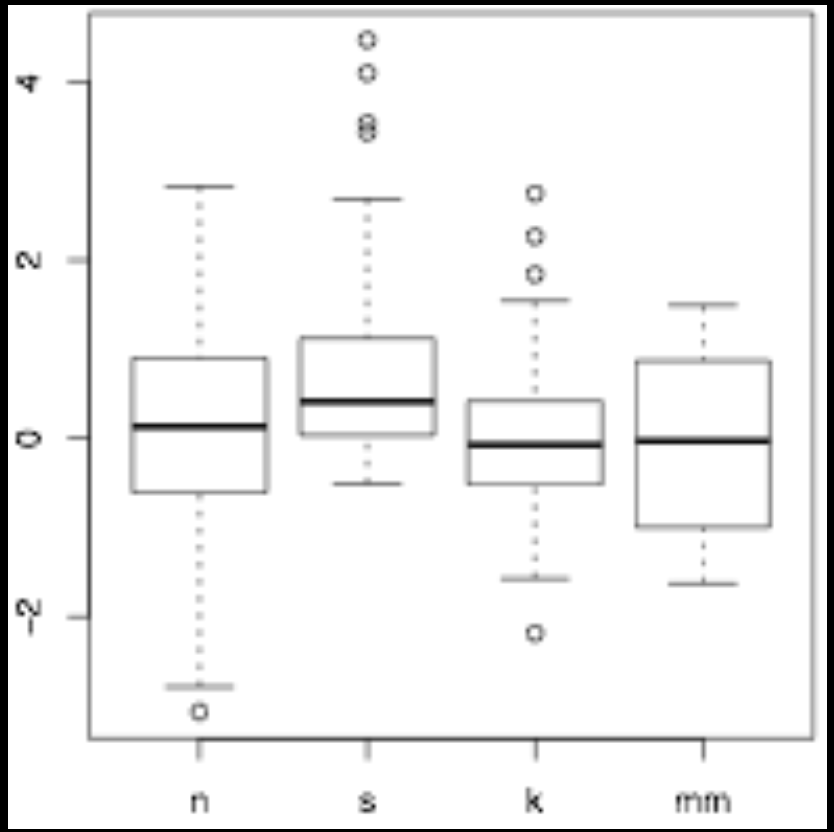
  - scale effects



https://blog.cartographica.com/blog/2011/5/19/the-modifiable-areal-unit-problem-in-gis.html

Idiom: **boxplot**

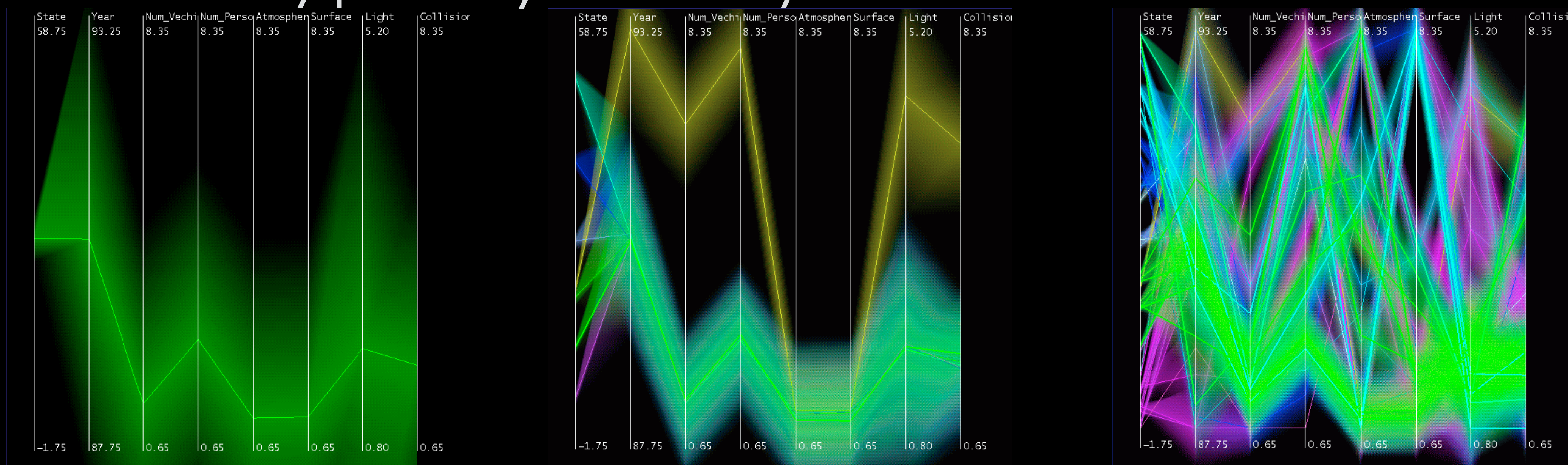
- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown



[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

Idiom: **Hierarchical parallel coordinates**

- dynamic item aggregation
- derived data: *hierarchical clustering*
- encoding:
  - cluster band with variable transparency, line at mean, width by min/max values
  - color by proximity in hierarchy



[Hierarchical Parallel Coordinates for Exploration of Large Datasets. Fua, Ward, and Rundensteiner. Proc. IEEE Visualization Conference (Vis '99), pp. 43– 50, 1999.]

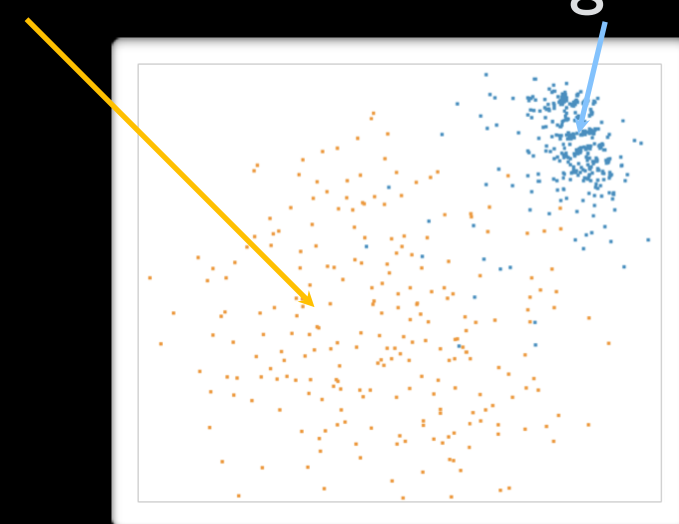
Dimensionality reduction

- attribute aggregation
  - derive low-dimensional target space from high-dimensional measured space
  - use when you can't directly measure what you care about
    - true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
    - latent factors, hidden variables

Tumor
Measurement Data

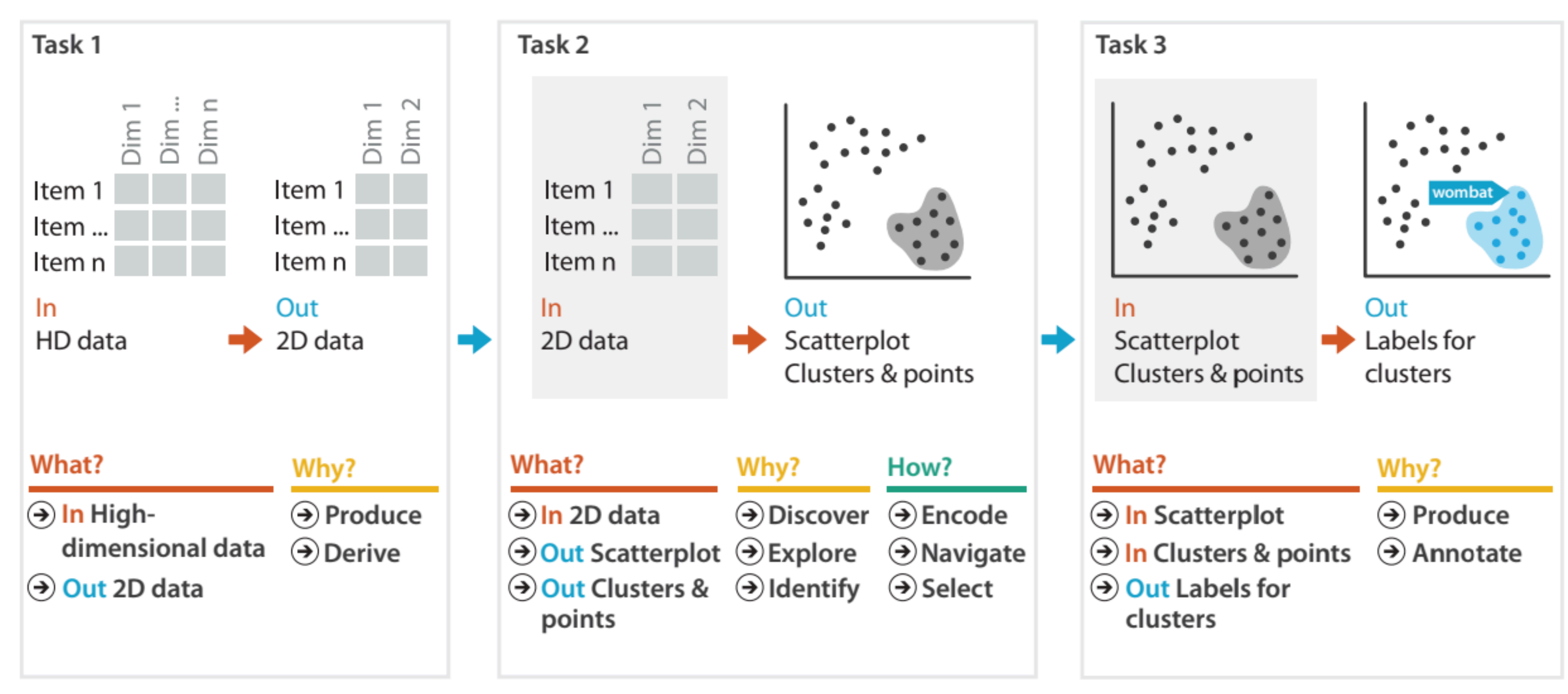
data: 9D measured space

DR



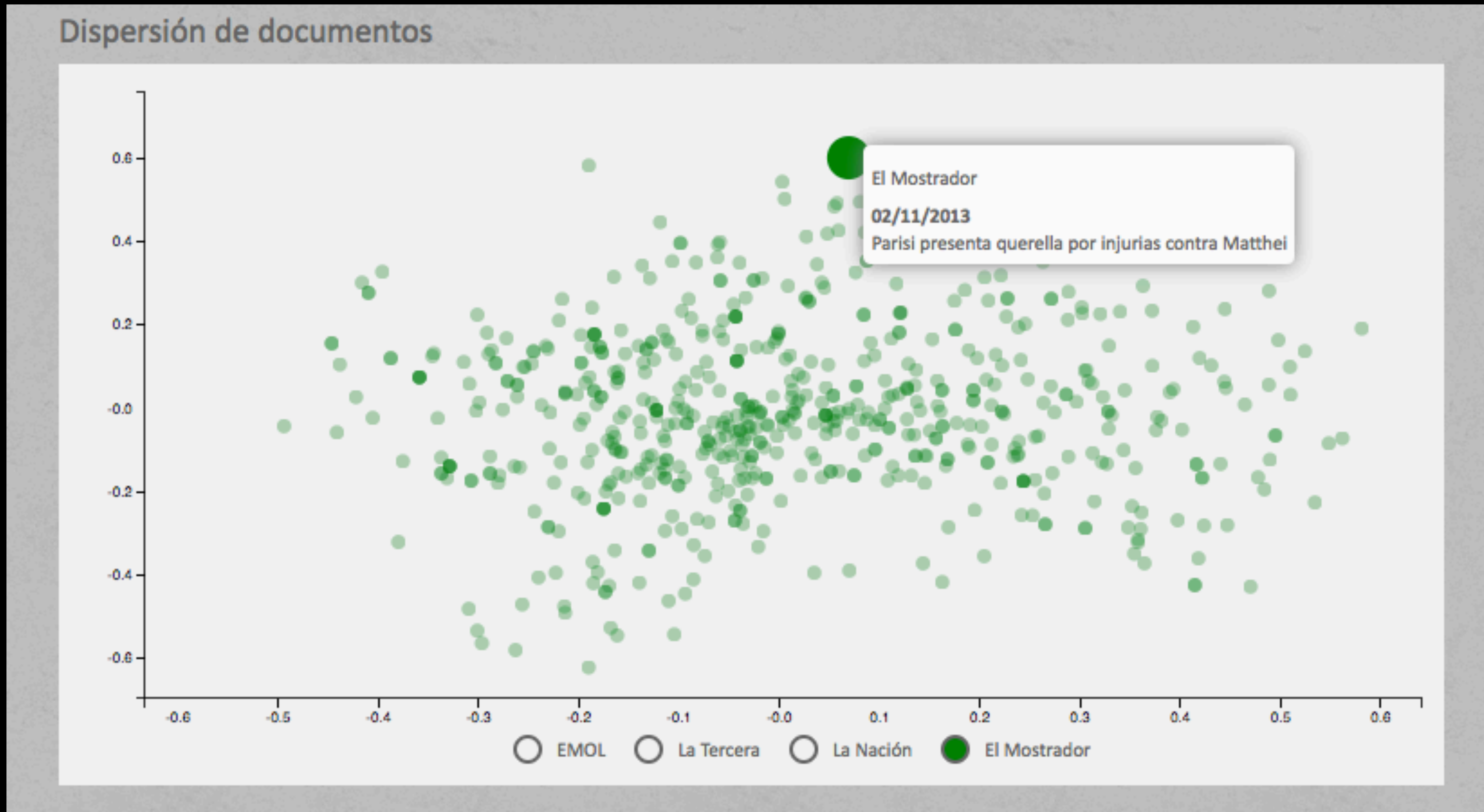
derived data: 2D target space

Dimensionality reduction for documents



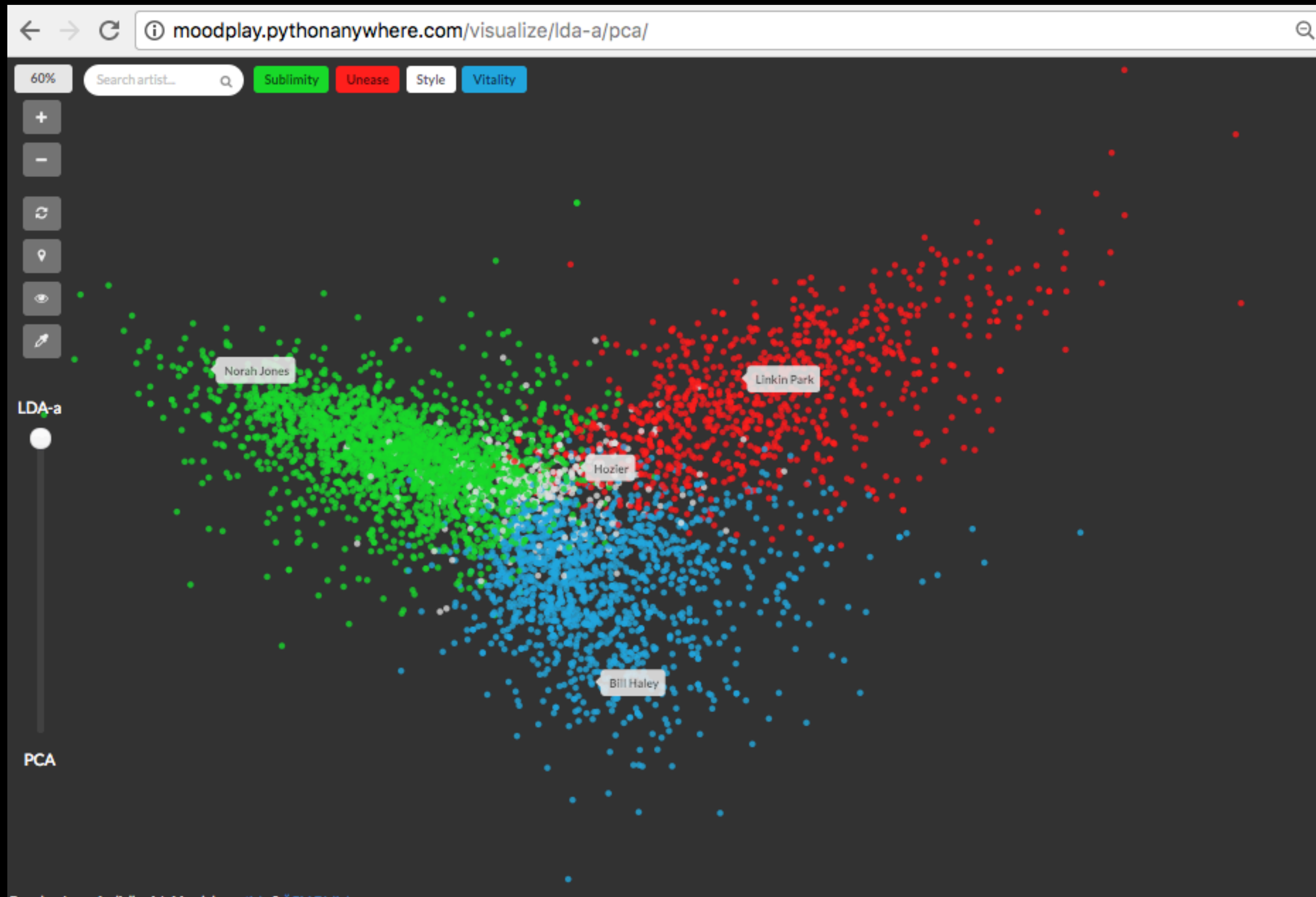
Ejemplo: Datos de noticias en última elección en Chile

http://dfao-uc.github.io/#articulo



Ejemplo 2: Moodplay (matrix artista x mood)

http://moodplay.pythonanywhere.com/visualize/lda-a/pca/



Dimensionality vs attribute reduction

- vocab use in field not consistent
  - dimension/attribute
- attribute reduction: reduce set with filtering
  - includes orthographic projection
- dimensionality reduction: create smaller set of new dims/attribs
  - typically implies dimensional aggregation, not just filtering
  - vocab: projection/mapping

Estimating true dimensionality

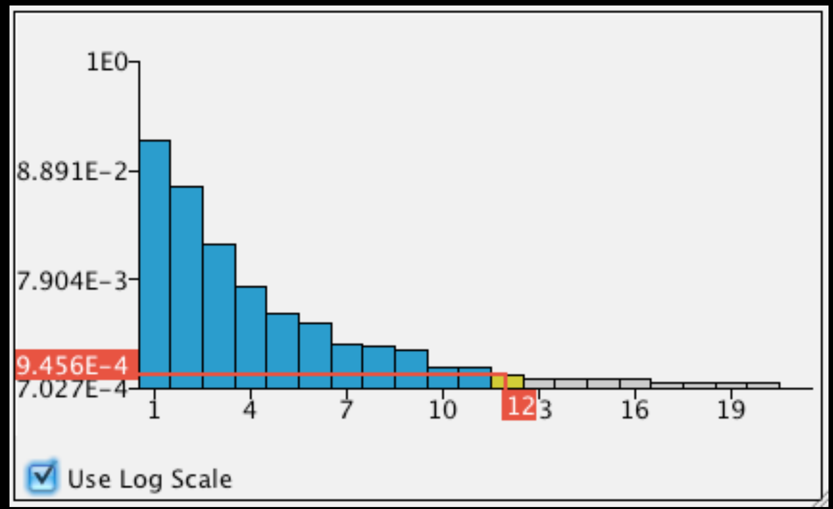
- how do you know when you would benefit from DR?
  - consider error for low-dim projection vs high-dim projection
- no single correct answer; many metrics proposed
  - cumulative variance that is not accounted for
  - strain: match variations in distance (vs actual distance values)
  - stress: difference between interpoint distances in high and low dims

$$stress(D, \Delta) = \sqrt{\frac{\sum_{ij}\left(d_{ij} - \delta_{ij}\right)^2}{\sum_{ij} \delta_{ij}^2}}$$

- D: matrix of lowD distances
- Δ: matrix of hiD distances δ_{ij}

Estimating true dimensionality

- scree plots as simple way: error against # attribs



  - original dataset: 294 dims
  - estimate: almost all variance preserved with < 20 dims

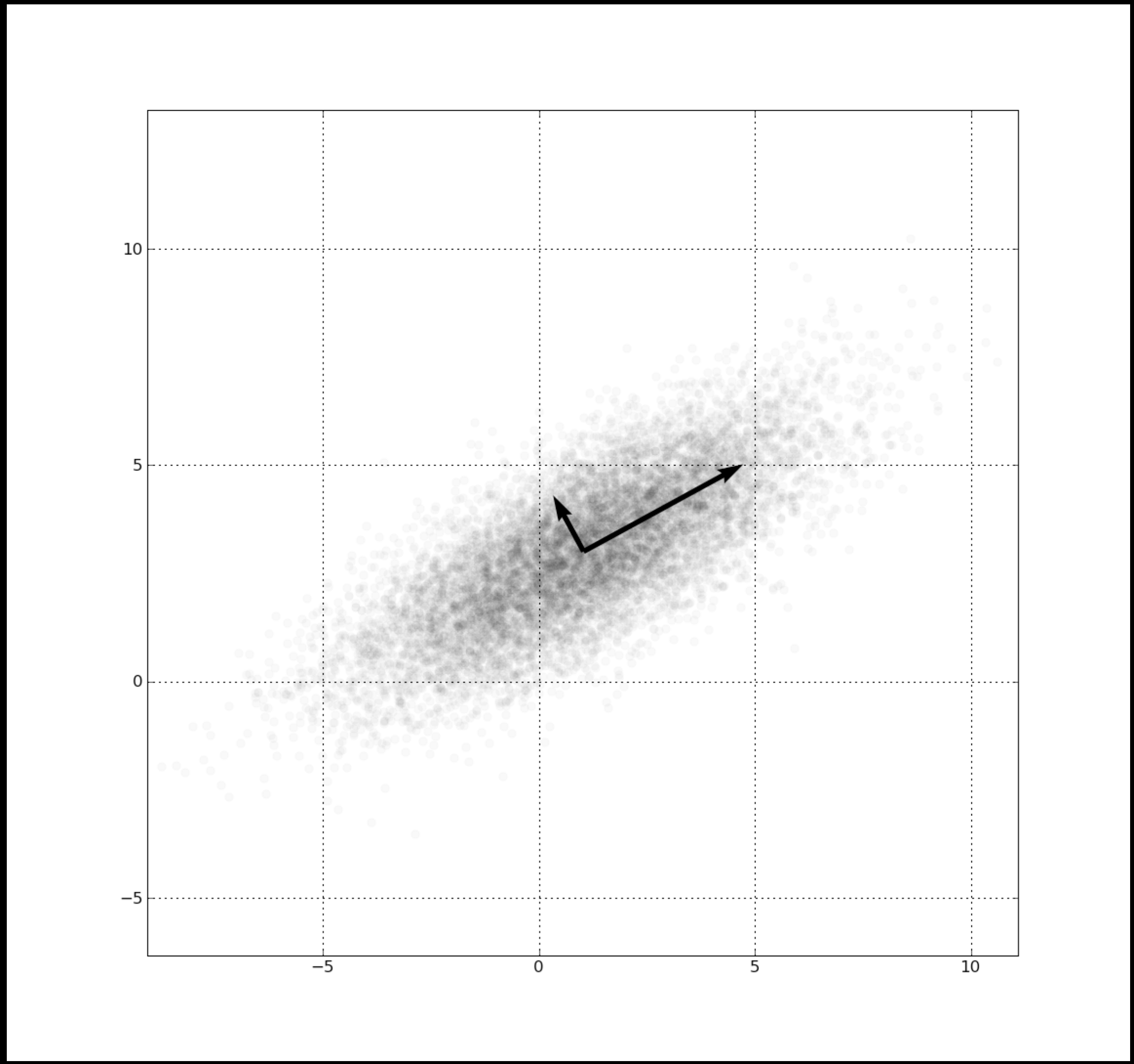
Dimensionality Reduction

- why do people do DR?
  - improve performance of downstream algorithm
    - avoid curse of dimensionality
  - data analysis
    - if look at the output: visual data analysis!
- DR tasks
  - dimension-oriented task sequences
    - name synthetic dimensions, map synthetic dims to original ones
  - cluster-oriented task sequences
    - verify clusters, name clusters, match clusters and classes

Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences

Reducción de dimensionalidad Lineal

- Principal components analysis (PCA)
  - Describir la ubicación de cada punto como una combinación lineal de los pesos de cada eje
  - Tarea: encontrar los ejes, desde el que captura más varianza en adelante



[http://en.wikipedia.org/wiki/File:GaussianScatterPCA.png]

PCA 1

Queremos lograr una transformación de nuestros datos originales a una base/proyección nueva, de forma que se maximice la varianza.

$$\mathbf{PX} = \mathbf{Y} \qquad (1)$$

Also let us define the following quantities.[2]

- $\mathbf{p_i}$ are the *rows* of $\mathbf{P}$
- $\mathbf{x_i}$ are the *columns* of $\mathbf{X}$ (or individual $\vec{X}$).
- $\mathbf{y_i}$ are the *columns* of $\mathbf{Y}$.

1. $\mathbf{P}$ is a matrix that transforms $\mathbf{X}$ into $\mathbf{Y}$.
2. Geometrically, $\mathbf{P}$ is a rotation and a stretch which again transforms $\mathbf{X}$ into $\mathbf{Y}$.
3. The rows of $\mathbf{P}$, $\{\mathbf{p_1}, \ldots, \mathbf{p_m}\}$, are a set of new basis vectors for expressing the *columns* of $\mathbf{X}$.

$$\mathbf{PX} = \begin{bmatrix} \mathbf{p_1} \\ \vdots \\ \mathbf{p_m} \end{bmatrix} \begin{bmatrix} \mathbf{x_1} & \cdots & \mathbf{x_n} \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{p_1} \cdot \mathbf{x_1} & \cdots & \mathbf{p_1} \cdot \mathbf{x_n} \\ \vdots & \ddots & \vdots \\ \mathbf{p_m} \cdot \mathbf{x_1} & \cdots & \mathbf{p_m} \cdot \mathbf{x_n} \end{bmatrix}$$

PCA 1

Queremos lograr una transformación de nuestros datos originales a una base/proyección nueva, de forma que se maximice la razón entre señal y ruido = razón entre varianzas.

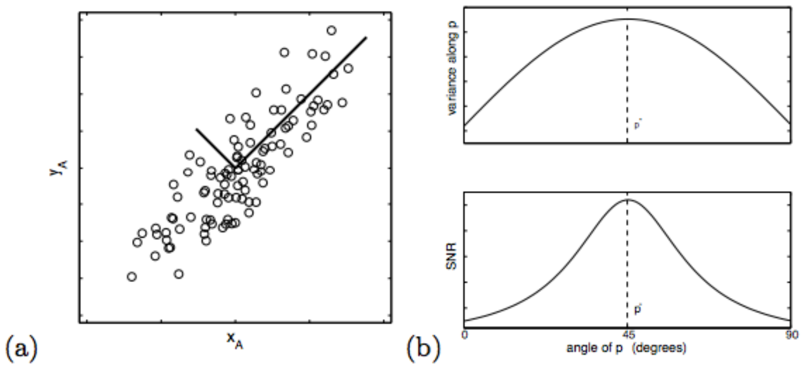


FIG. 2 (a) Simulated data of (x_A, y_A) for camera A. The signal and noise variances σ^2_{signal} and σ^2_{noise} are graphically represented by the two lines subtending the cloud of data. (b) Rotating these axes finds an optimal p^* where the variance and SNR are maximized. The SNR is defined as the ratio of the variance along p^* and the variance in the perpindicular direction.

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

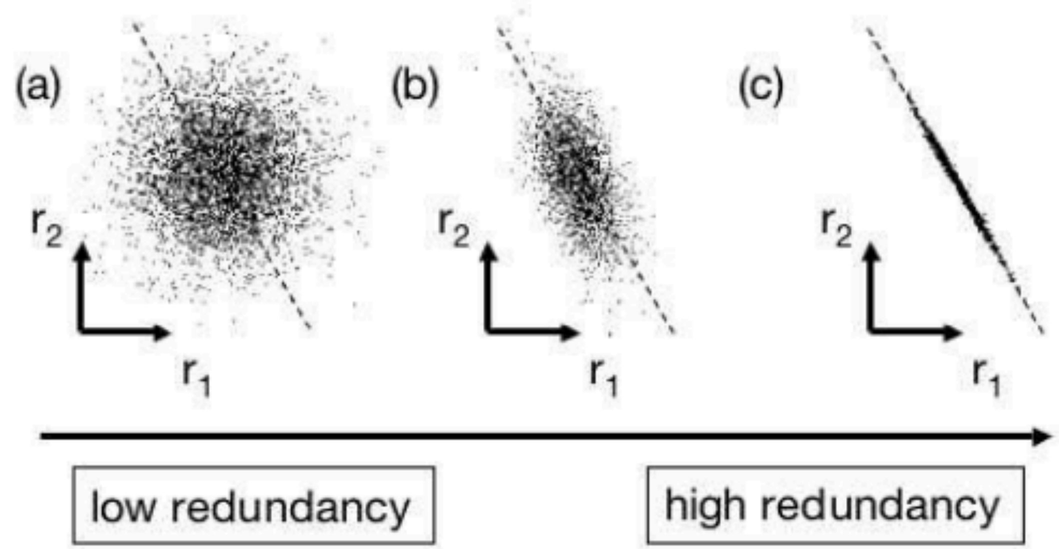


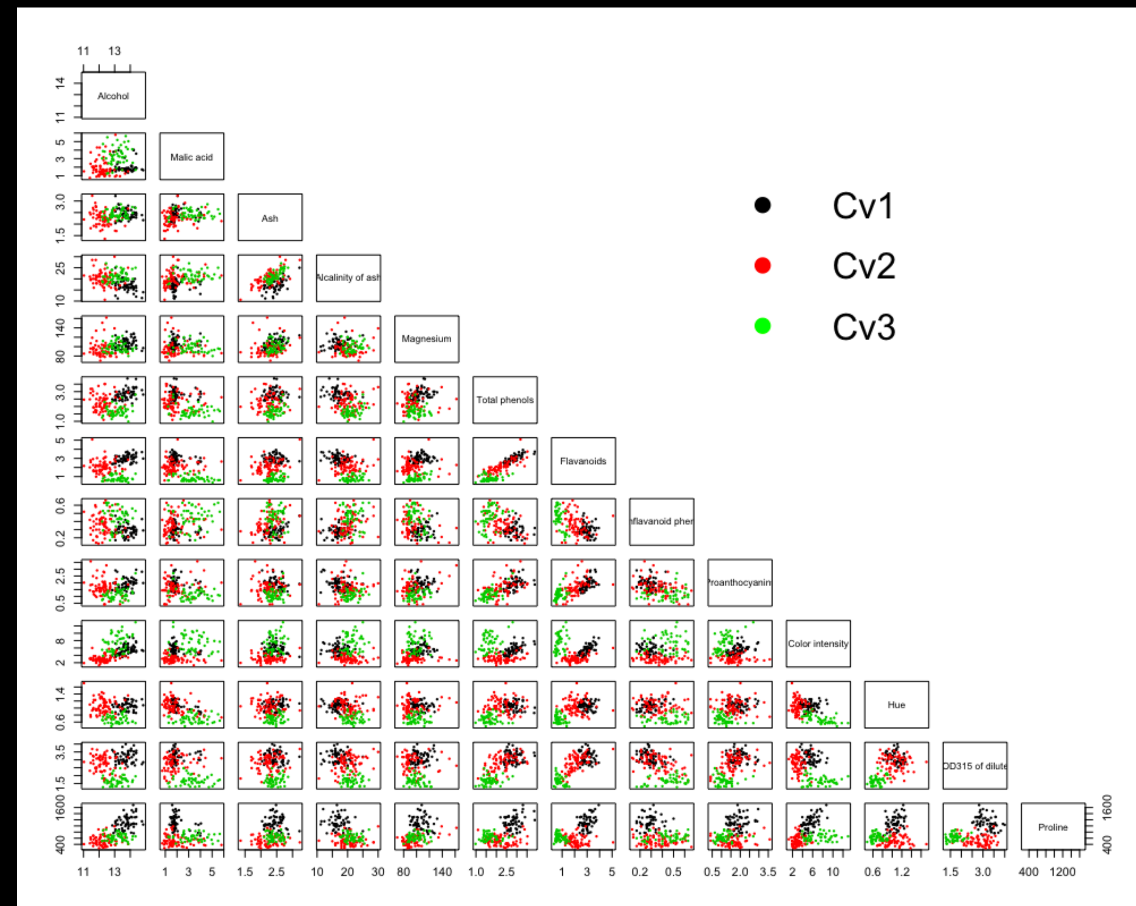
FIG. 3 A spectrum of possible redundancies in data from the two separate recordings r_1 and r_2 (e.g. x_A, y_B). The best-fit line $r_2 = kr_1$ is indicated by the dashed line.

PCA 1

Una forma de calcular la "redundancia" entre pares de atributos es a través de la covarianza. Todos los pares formarían una matriz de covarianza.

$$\mathbf{C} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \mu)(\mathbf{x}_n - \mu)^T$$

$$C = \begin{bmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{bmatrix}$$



PCA 1

Una formulación típica para calcular PCA es:

- Normalizar los datos (columnas)
- Calcular la matriz de covarianza
- Calcular valores y vectores propios de la matriz de covarianza
- Proyectar los datos originales en la nueva base (vectores propios)

PCA 2

Es posible demostrar que la formulación del problema en base a matriz de covarianza es equivalente a una descomposición matricial SVD:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

$$\begin{aligned}
\mathbf{X}^T\mathbf{X} &= (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\
&= \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\
&= \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T.
\end{aligned}$$

http://www.cs.cmu.edu/~tom/10601_fall2012/slides/pca.pdf

PCA 7

Ejemplo en R:

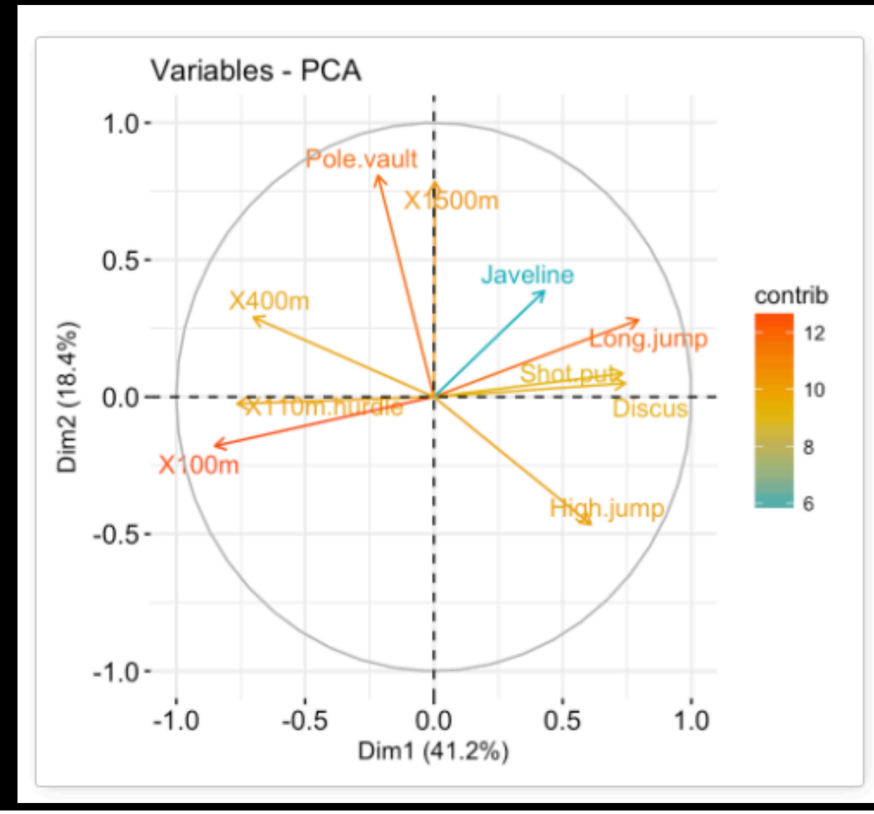
http://www.sthda.com/english/wiki/factoextra-r-package-easy-multivariate-data-analyses-and-elegant-visualization

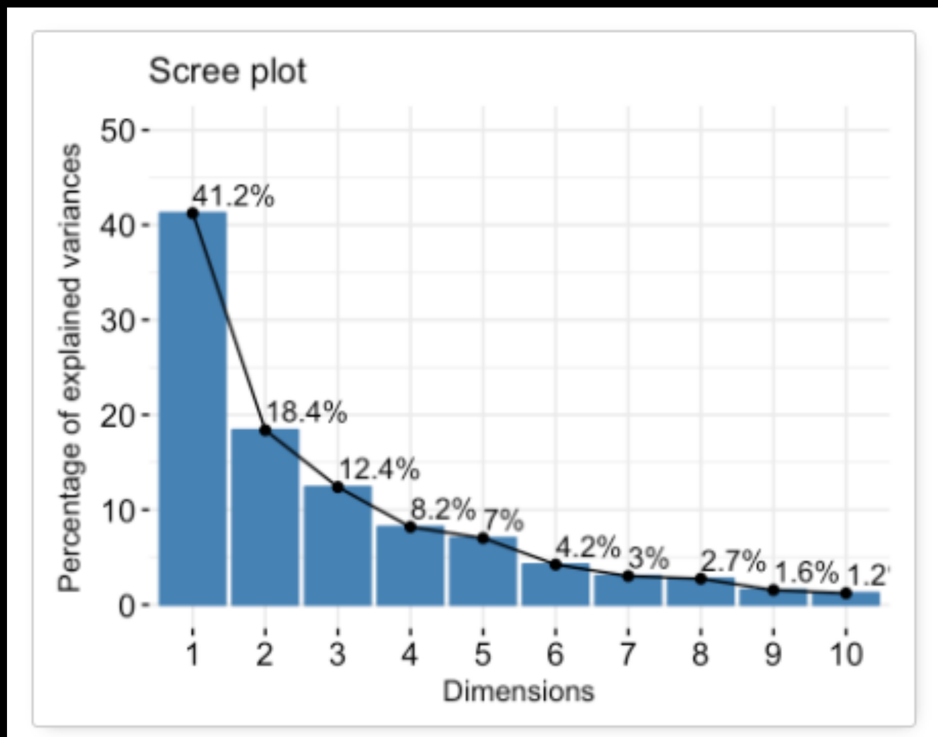
O

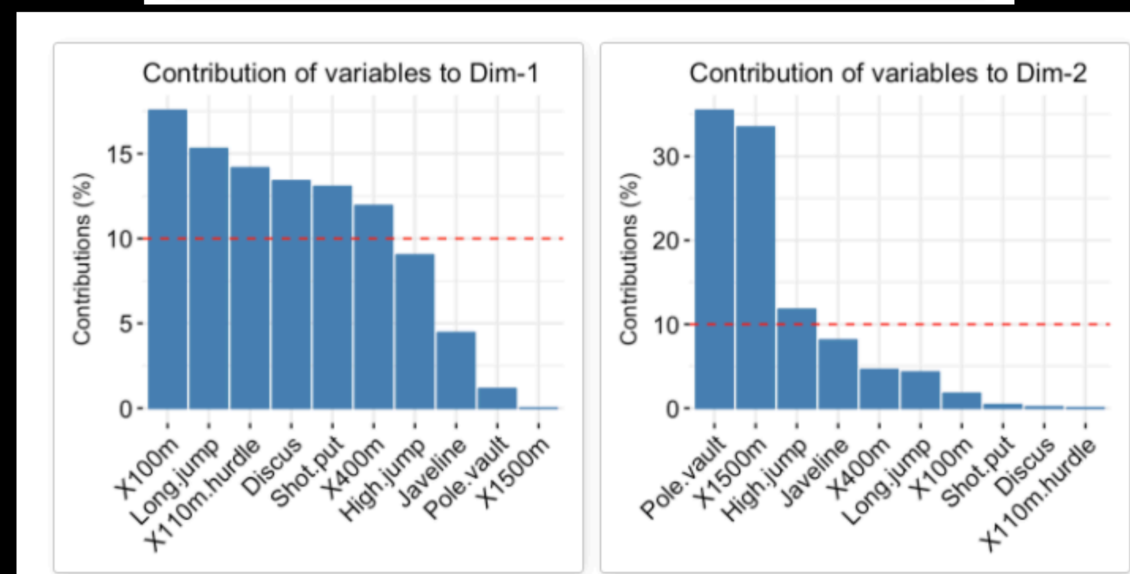
https://cran.r-project.org/web/packages/ggfortify/vignettes/plot_pca.html

PCA 7

Validación

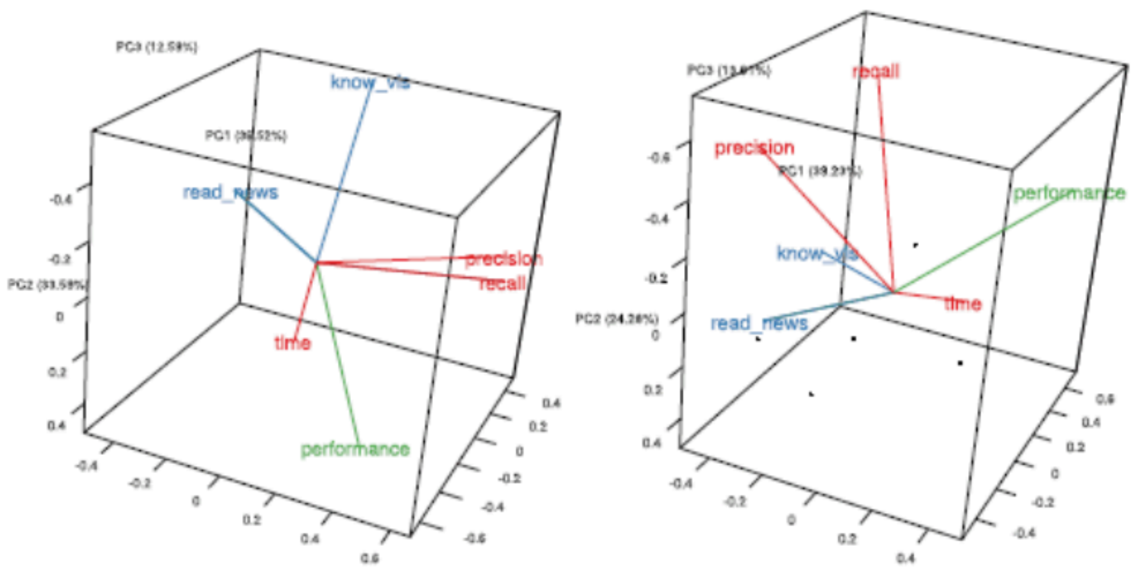






PCA 6

Ejemplo paper EPJ



(a) PCA Biplot for Galean

(b) PCA biplot for baseline interface.

Figure 9

3D biplots of the principal component analysis for objective and subjective metrics for both interfaces. Some metrics were removed for clarity. Subjects knowledge such as how familiar were participants with visualizations (know_vis) and how frequently they read news (read_news) are in blue. Objective metrics of time, precision and recall are in red. Subjective metrics of user's perceived performance are in green. While precision and recall are related to previous user knowledge on the baseline, loading on the same direction of PC1, in Galean precision and recall are more related to user's perceived performance and barely related to previous user knowledge.

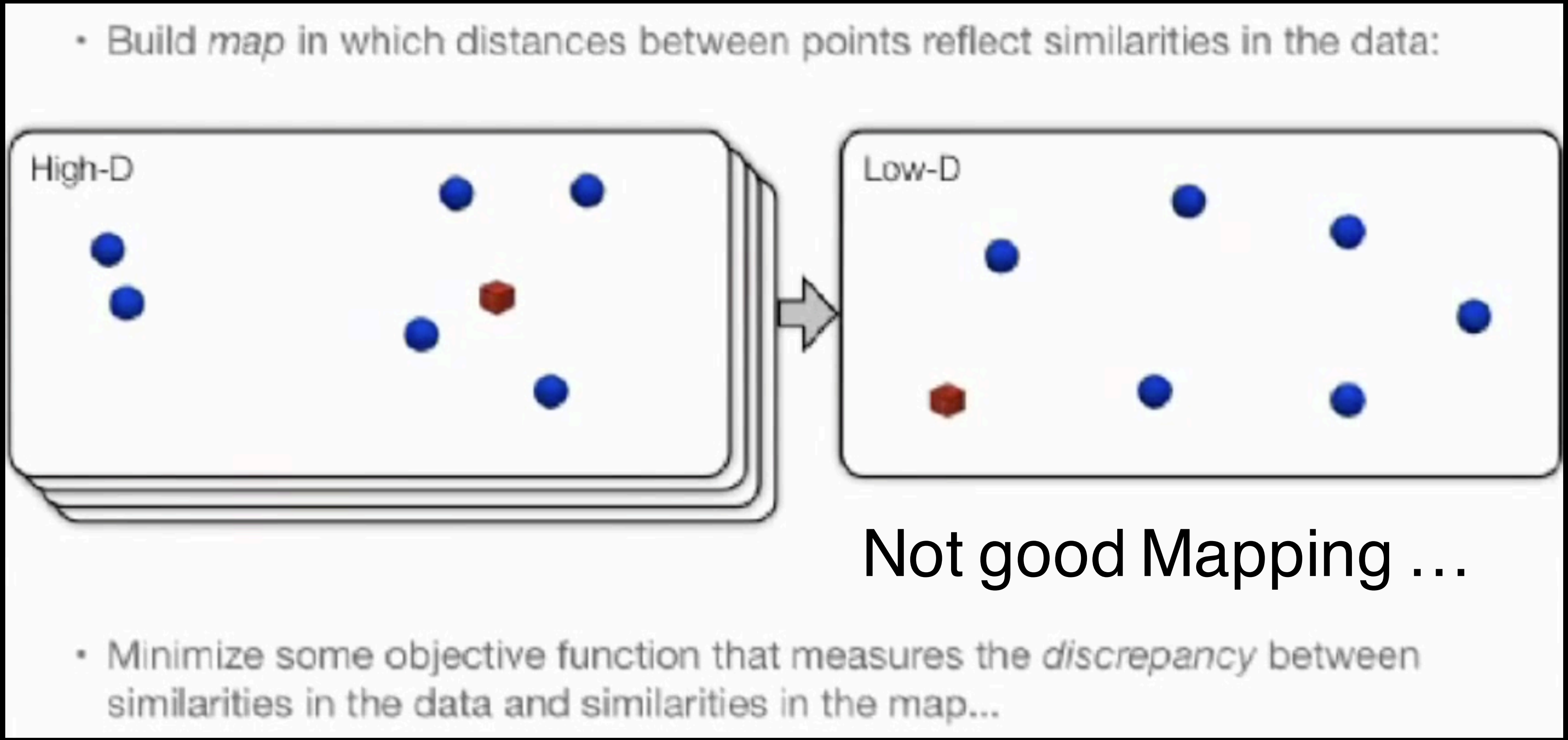
Gaining historical and international relations insights from social media: spatio-temporal real-world news analysis using Twitter, Peña-Araya et al. (2017)

Nonlinear dimensionality reduction

- many techniques proposed
  - MDS, charting, isomap, LLE, t-SNE
  - many literatures: visualization, machine learning, optimization, psychology, ...
- pro: can handle curved rather than linear structure
- cons: lose all ties to original dims/attribs
  - new dimensions cannot be easily related to originals

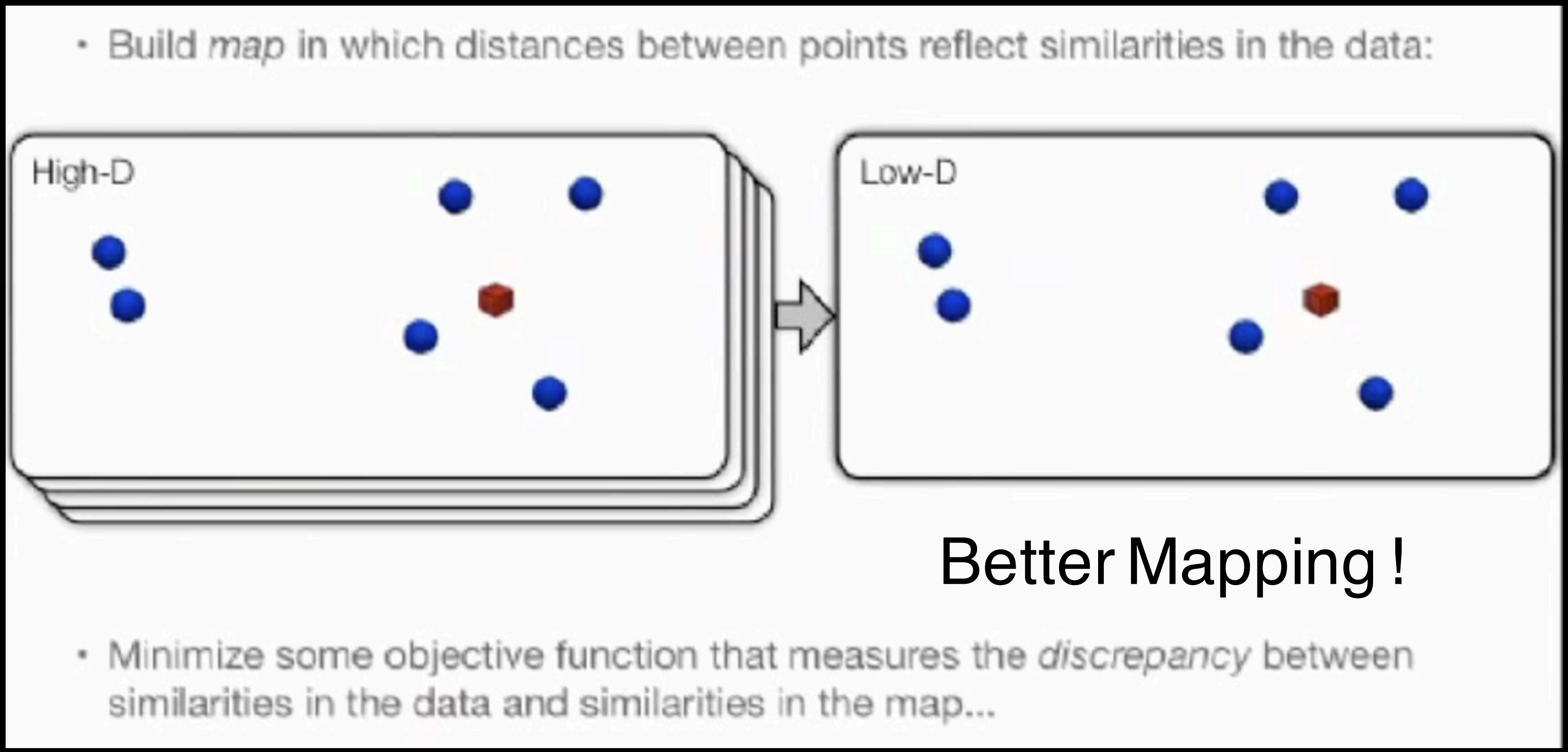
t-SNE

- T-distributed Stochastic Neighbor Embedding
- Used frequently to plot embeddings or Deep Neural Networks

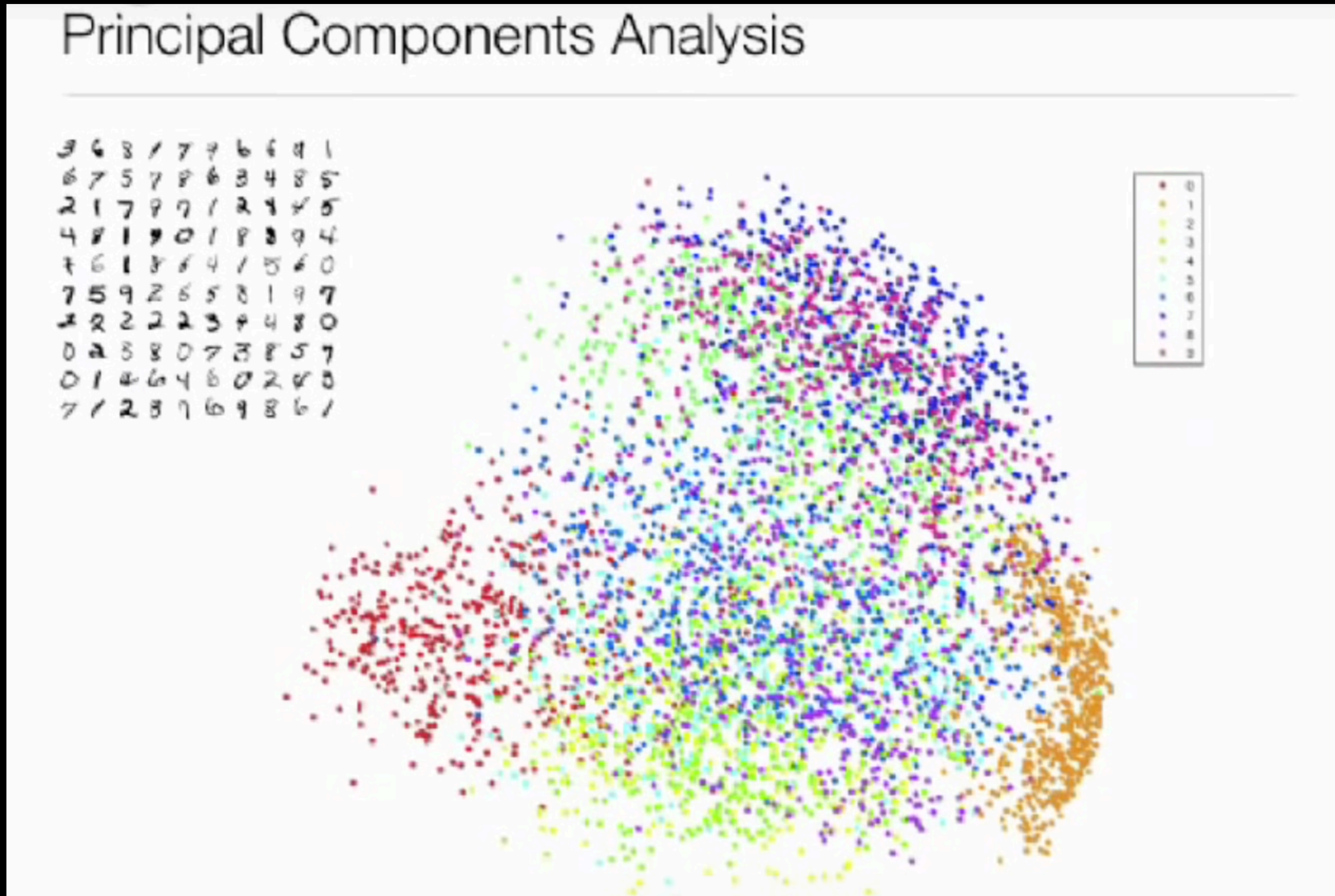


t-SNE

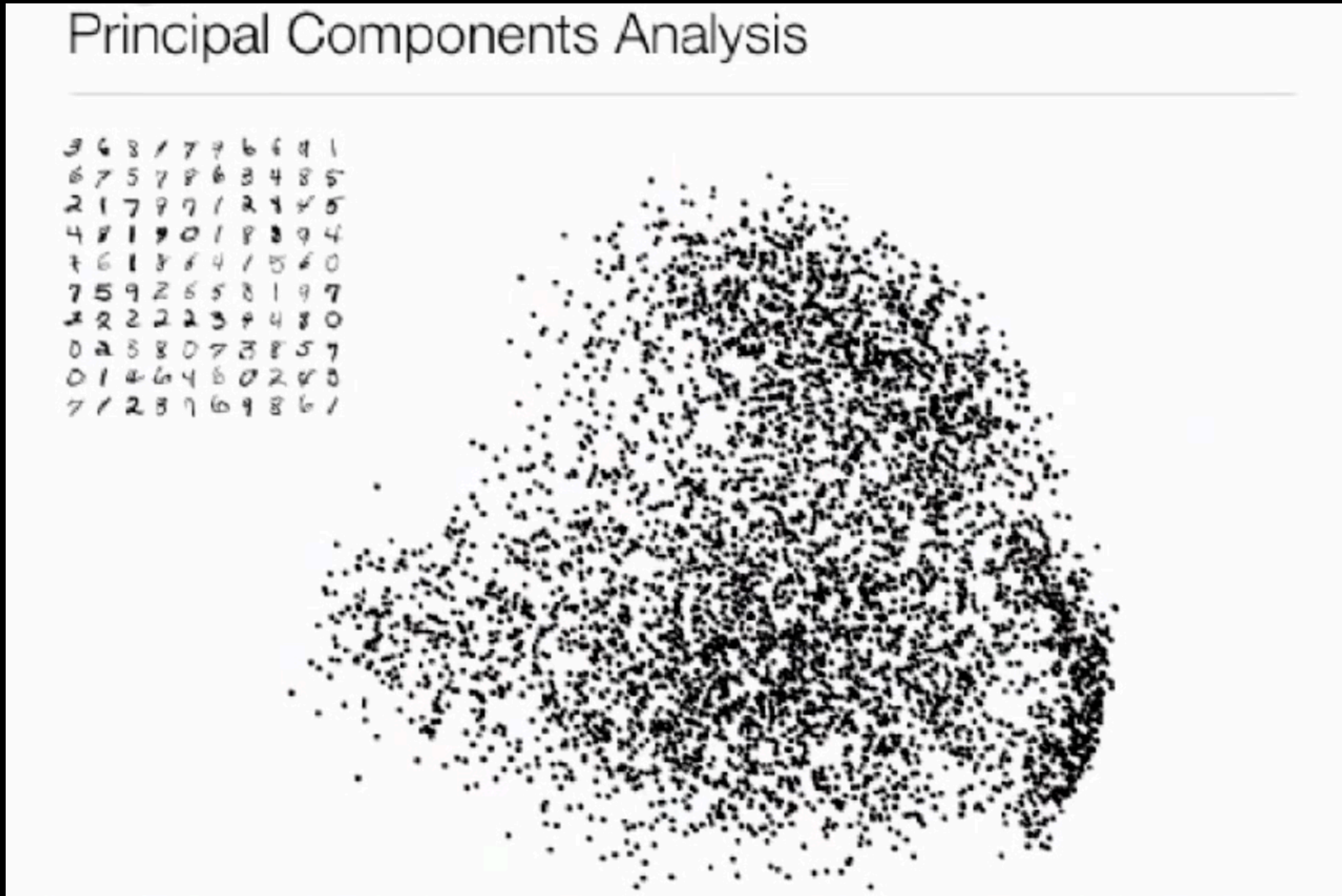
- T-distributed Stochastic Neighbor Embedding
- Used frequently to plot embeddings or Deep Neural Networks



Example 1: PCA

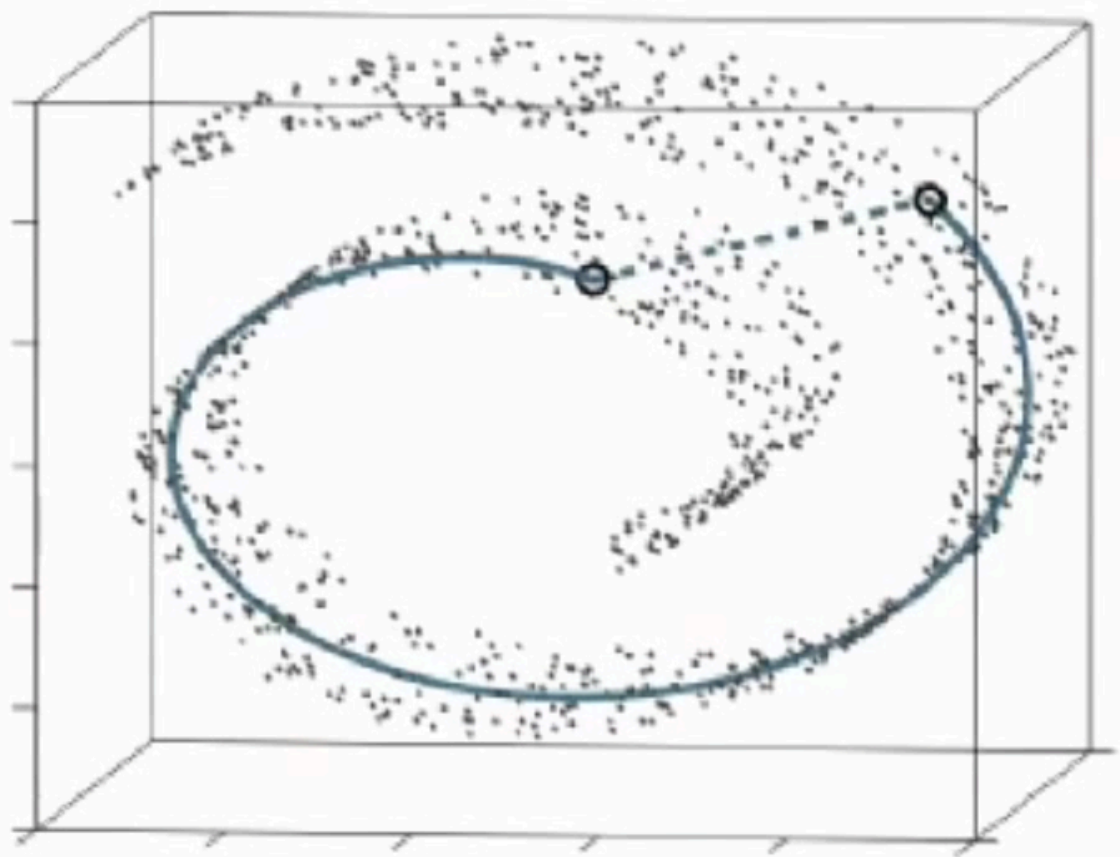


Example 1: Problem with PCA



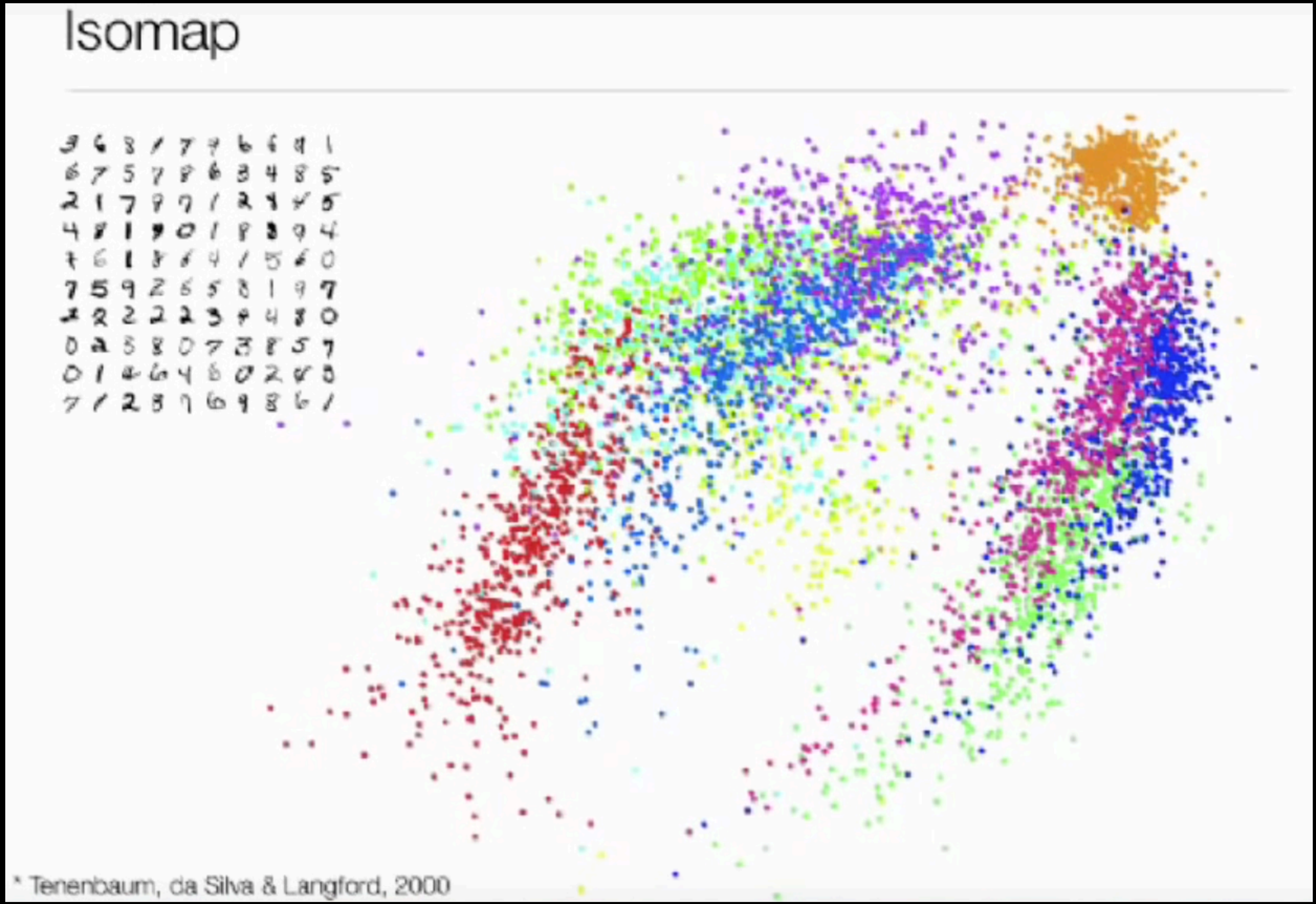
Example 1: Problem with PCA

- PCA is mainly concerned with preserving *large* pairwise distances in the map
- But are such distances very reliable?

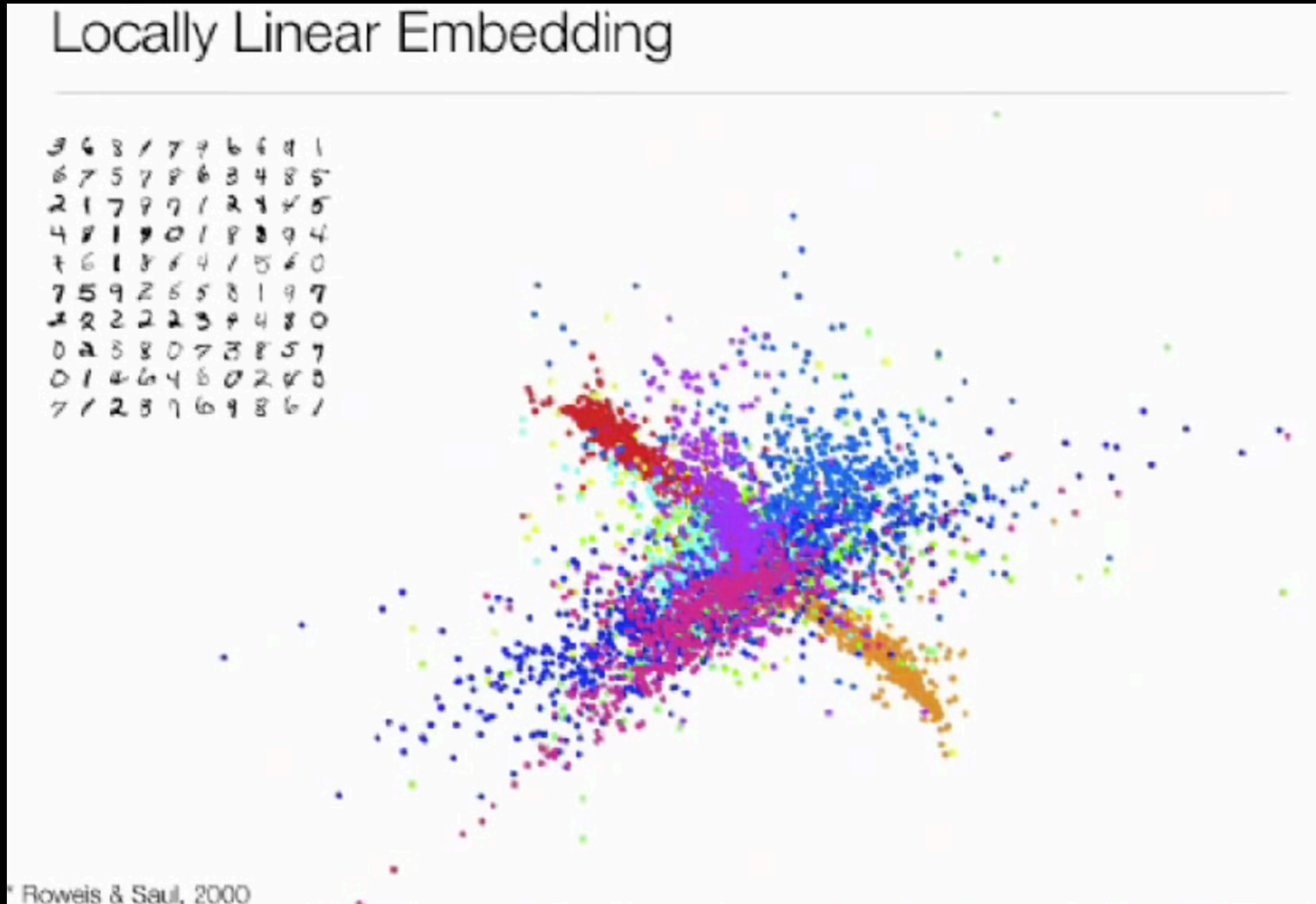


Example 1: Isomap

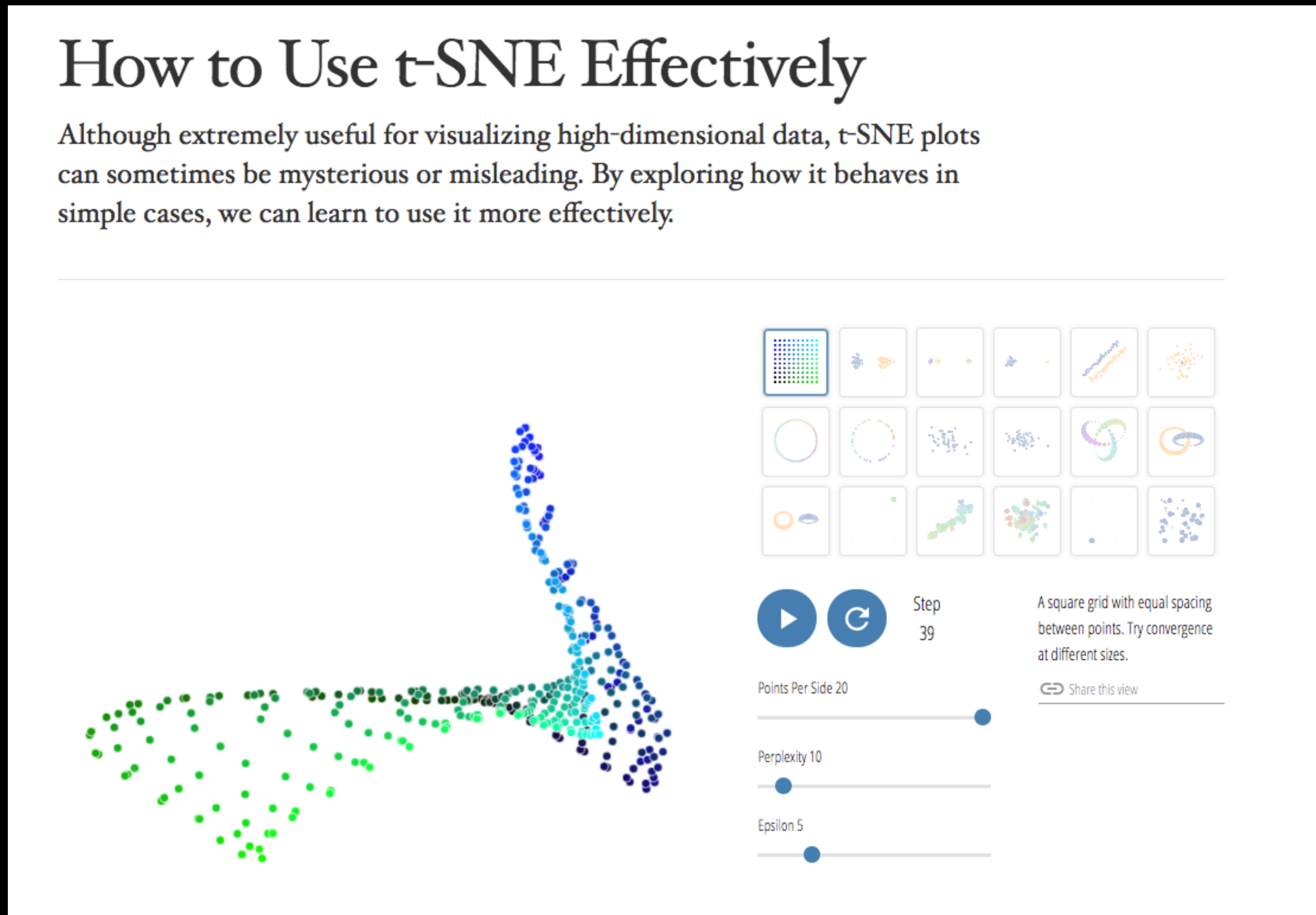
Usa estructura local



Example 1: Alternative 2: LLE



t-SNE: https://distill.pub/2016/misread-tsne/



t-SNE Details

- The goal is to take a set of points in a high-dimensional space and find a faithful representation of those points in a lower-dimensional space, typically the 2D plane.

- The algorithm is non-linear and adapts to the underlying data, performing different transformations on different regions.

- Those differences can be a major source of confusion.

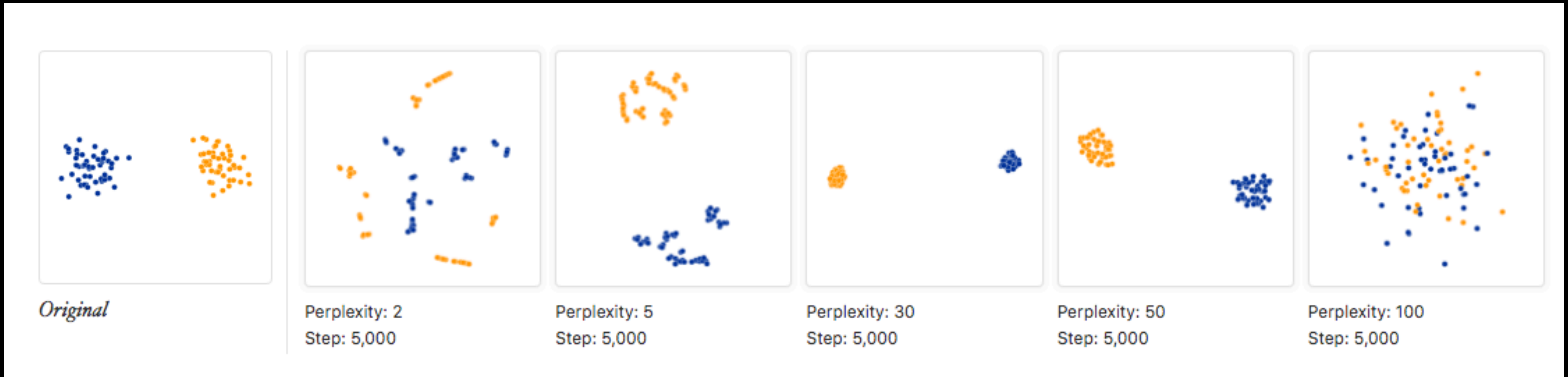
T-SNE details 2

- A second feature of t-SNE is a tuneable parameter, "perplexity," which says (loosely) how to balance attention between local and global aspects of your data.

- The parameter is, in a sense, a guess about the number of close neighbors each point has. The perplexity value has a complex effect on the resulting pictures.

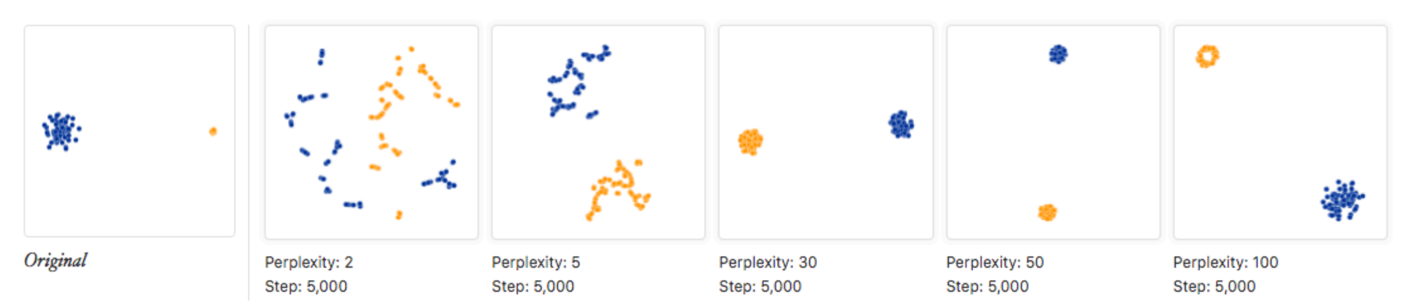
- The original paper says, *"The performance of SNE is fairly robust to changes in the perplexity, and typical values are between 5 and 50."* But the story is more nuanced than that.

Effect of perpexlity in t-SNE



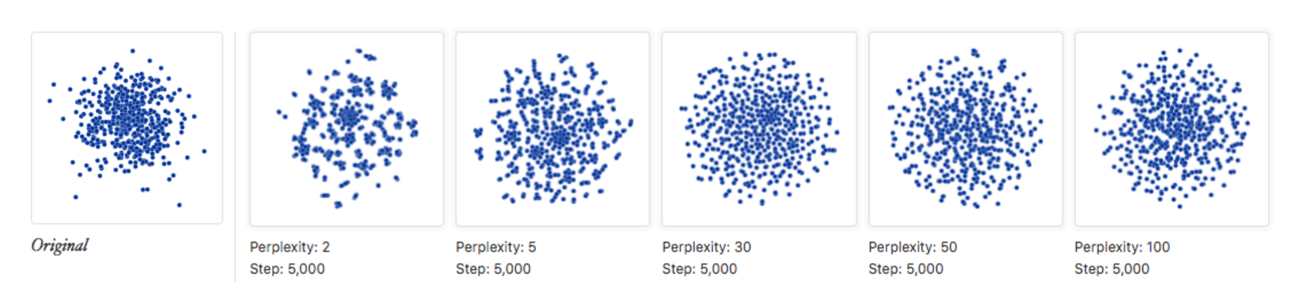
Other important aspects

2. Cluster sizes in a t-SNE plot mean nothing



3. Distances between clusters might not mean anything

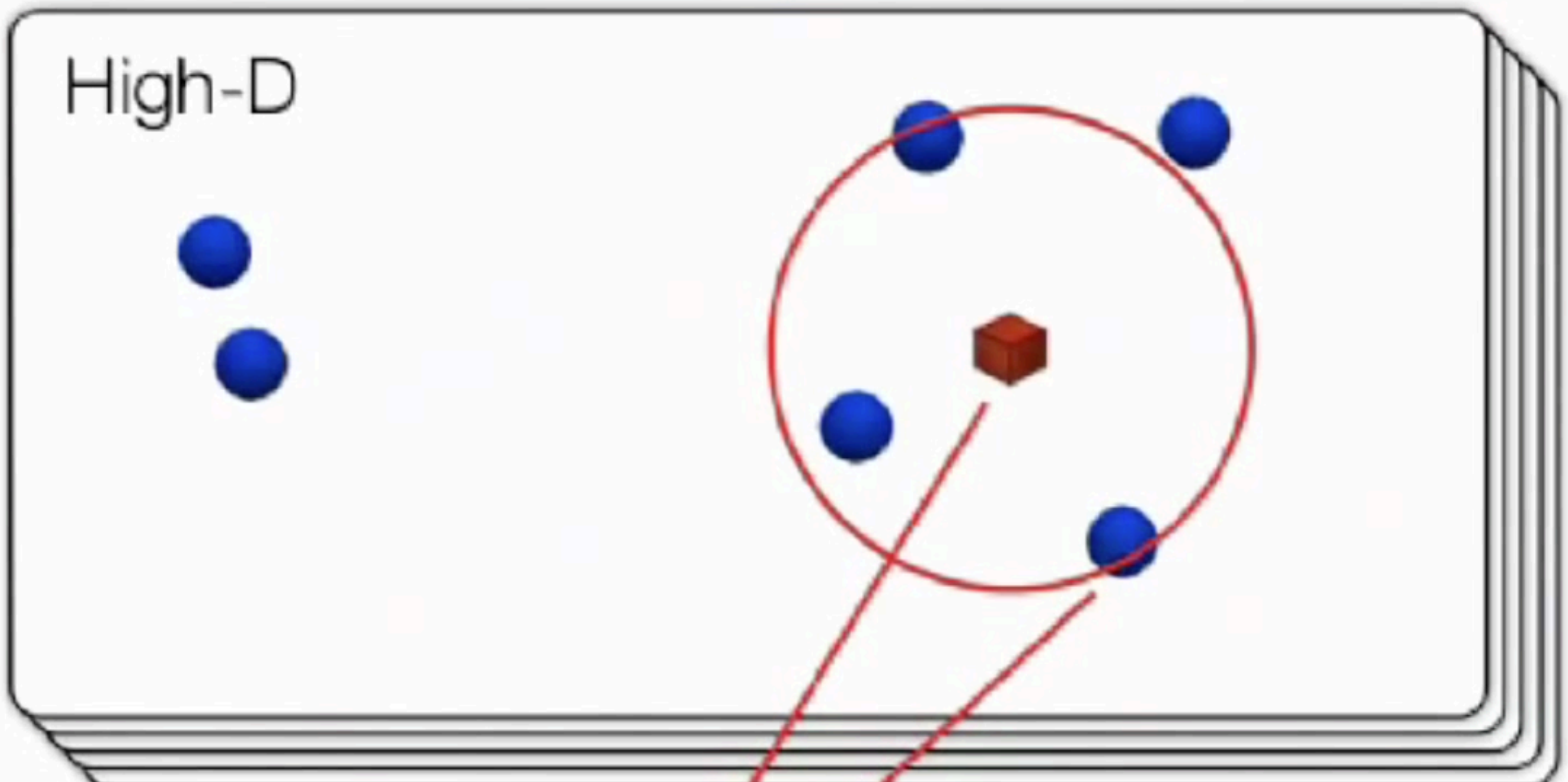
4. Random noise doesn't always look random.



Example 1: Alternative 3: t-SNE

t-Distributed Stochastic Neighbor Embedding

- Measure pairwise similarities between high-dimensional objects:



$$p_{ij} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)}{\sum_k \sum_{l \neq k} \exp(-\|\mathbf{x}_k - \mathbf{x}_l\|^2/2\sigma^2)}$$

Example 1: Alternative 3: t-SNE

t-Distributed Stochastic Neighbor Embedding

- In practice, we compute the input similarities slightly differently:

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma_i^2)}{\sum_{j' \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_{j'}\|^2/2\sigma_i^2)}$$

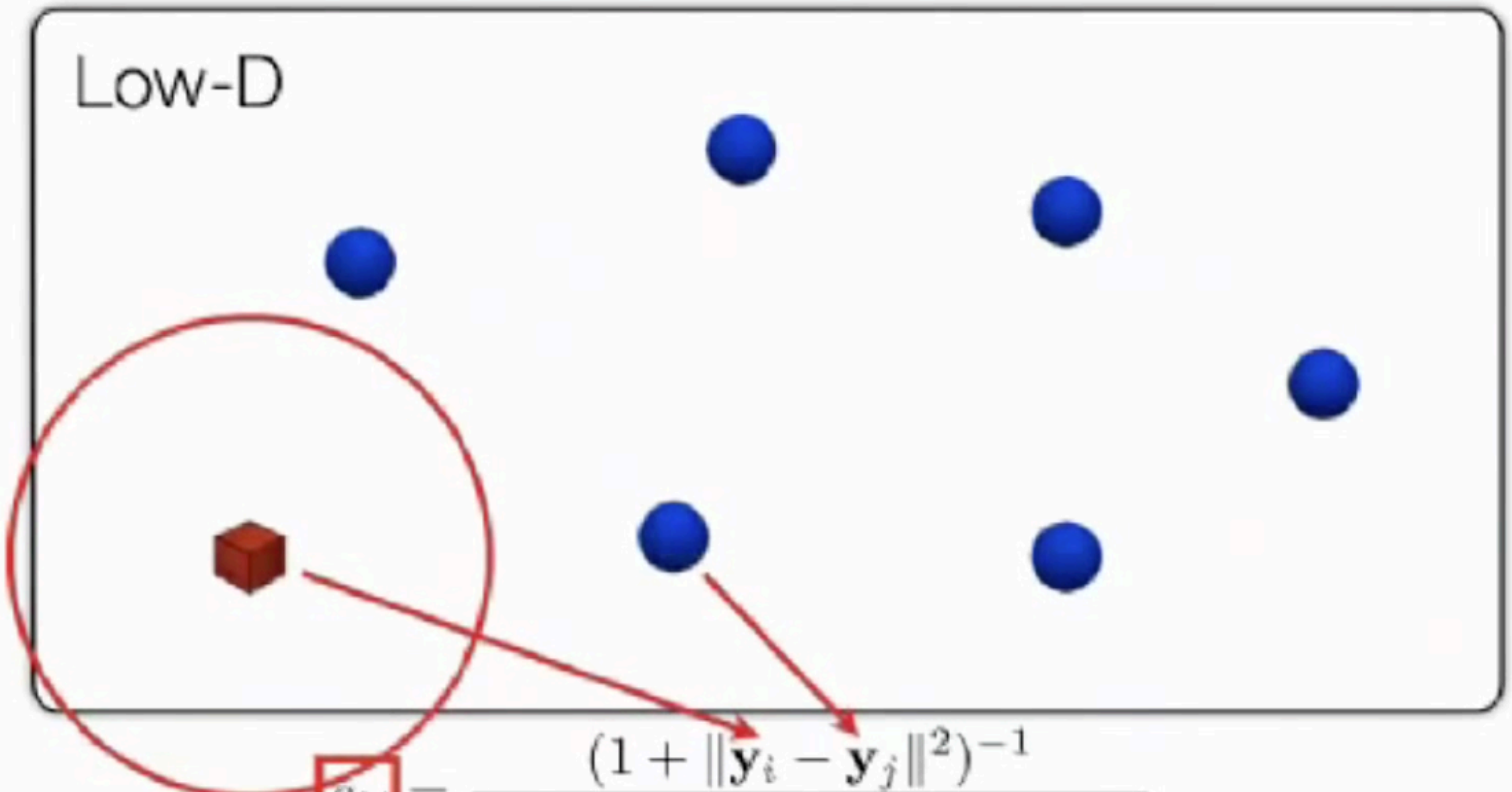
- We set the bandwidth σ_i such that the *conditional* has a *fixed perplexity*

- Finally, we *symmetrize* the conditionals: $p_{ij} = \dfrac{p_{j|i} + p_{i|j}}{2N}$

Example 1: Alternative 3: t-SNE

t-Distributed Stochastic Neighbor Embedding

- Measure pairwise similarities between low-dimensional map points:

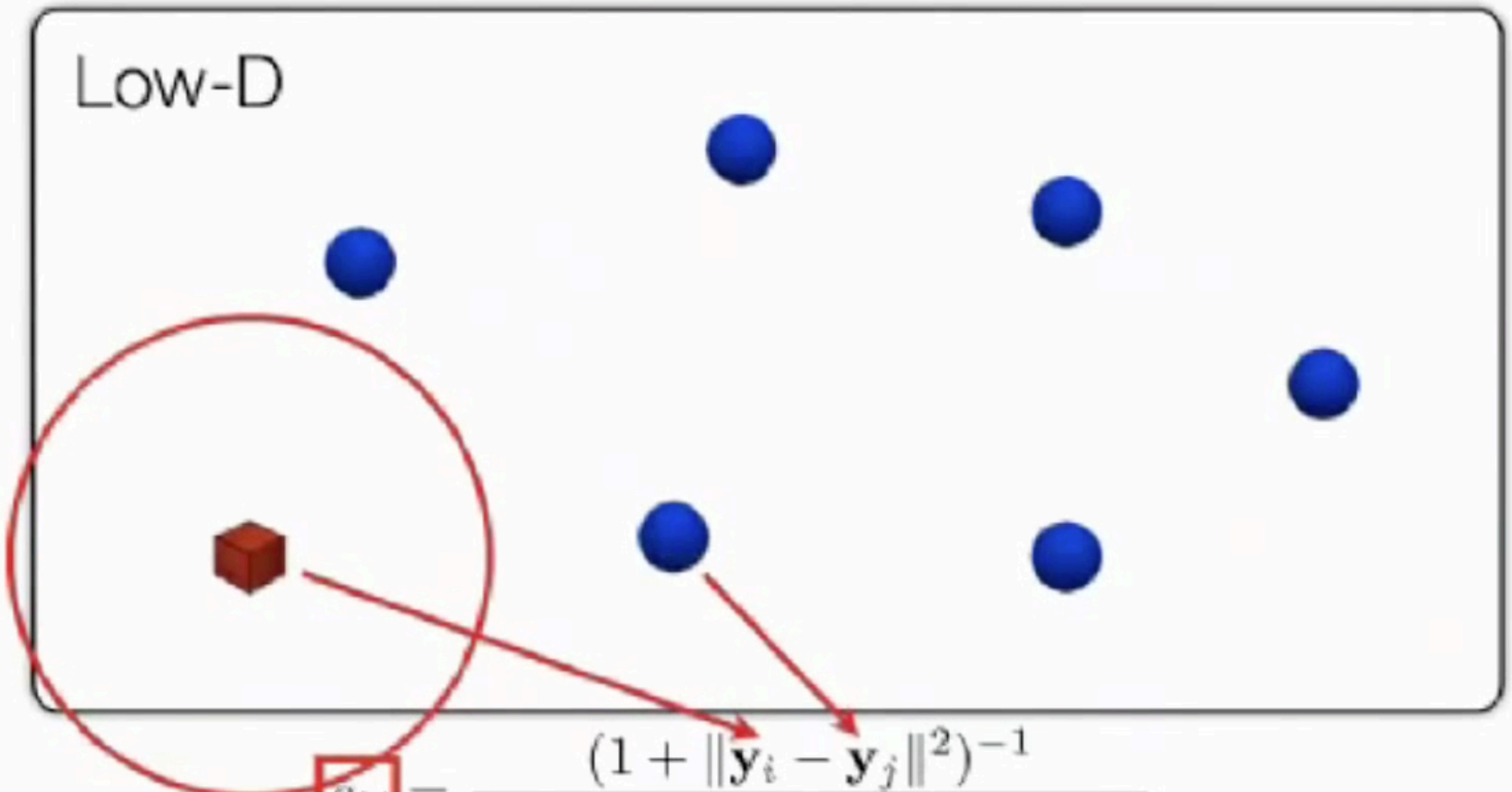


$$\boxed{q_{ij}} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_k \sum_{l \neq k} (1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}}$$

Example 1: Alternative 3: t-SNE

t-Distributed Stochastic Neighbor Embedding

- Measure pairwise similarities between low-dimensional map points:



$$\boxed{q_{ij}} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_k \sum_{l \neq k} (1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}}$$

Example 1: Alternative 3: t-SNE

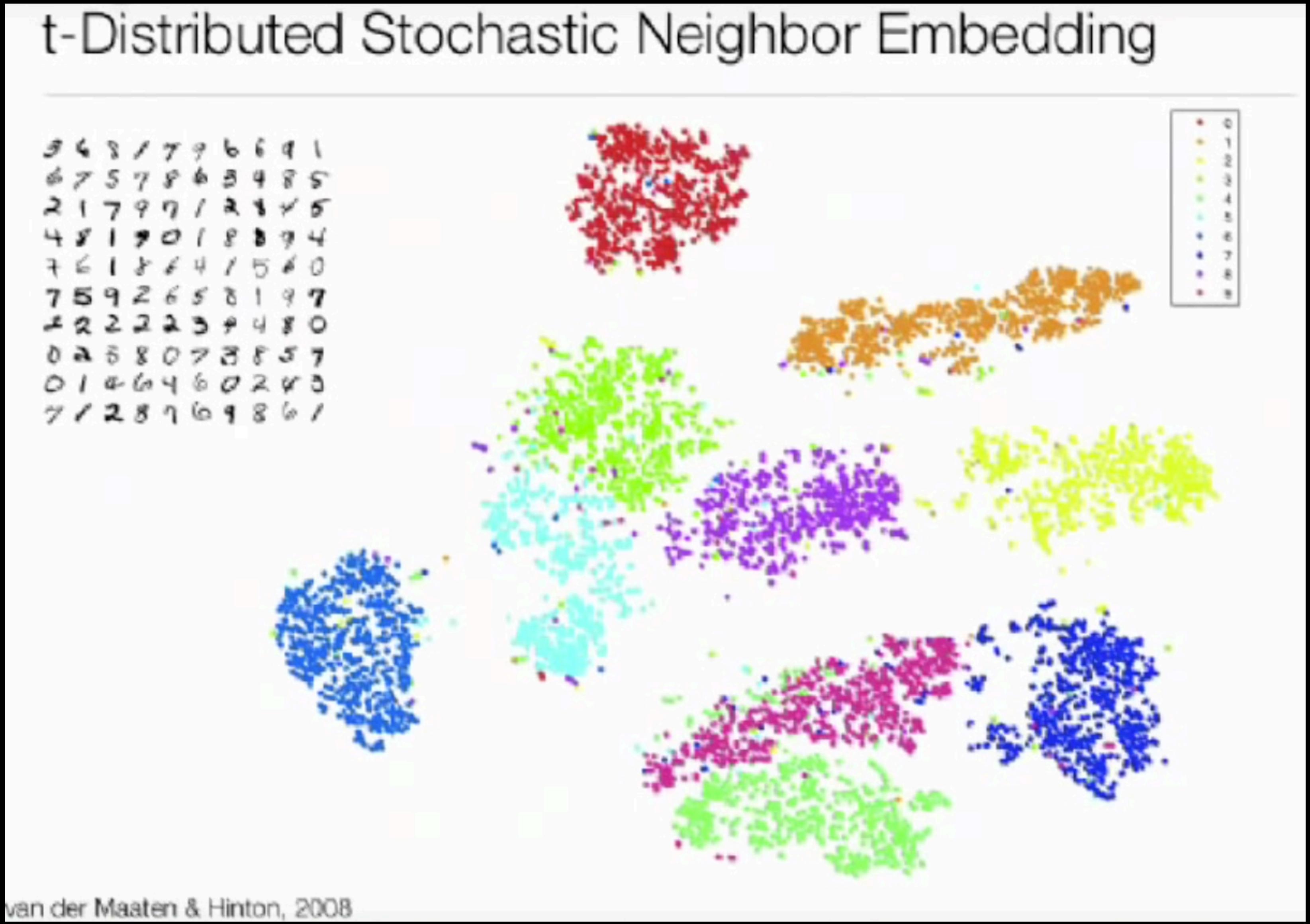
t-Distributed Stochastic Neighbor Embedding

- Kullback-Leibler divergence: $KL(P||Q) = \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}}$

  - Large p_{ij} modeled by small q_{ij}? Big penalty!

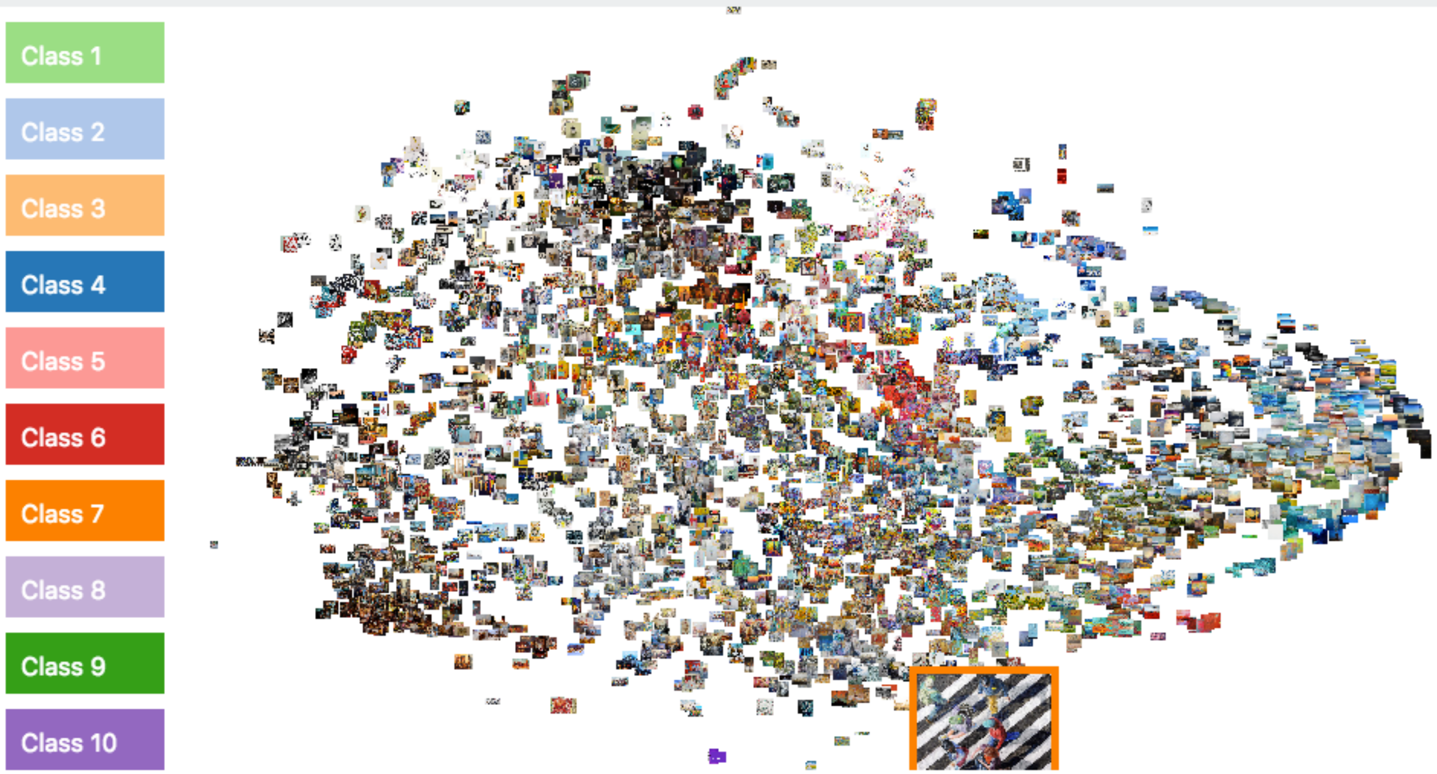
  - Small p_{ij} modeled by large q_{ij}? Small penalty!

t-SNE comparado con versiones anteriores



Ugallery I

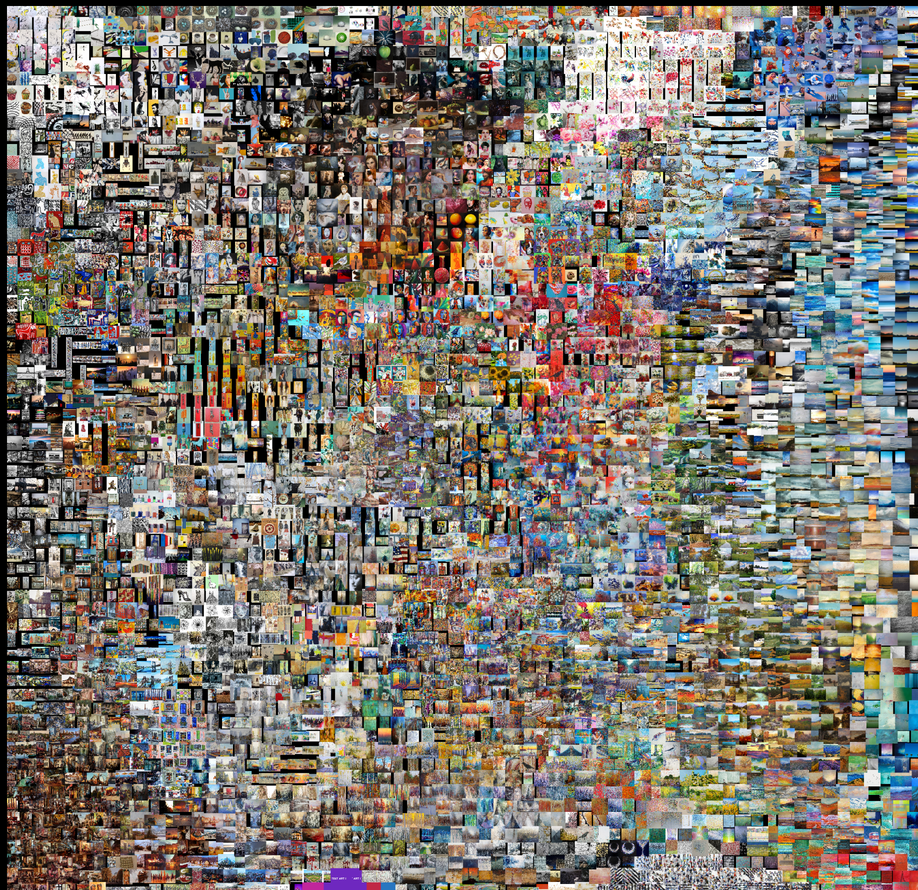
Image Embedding



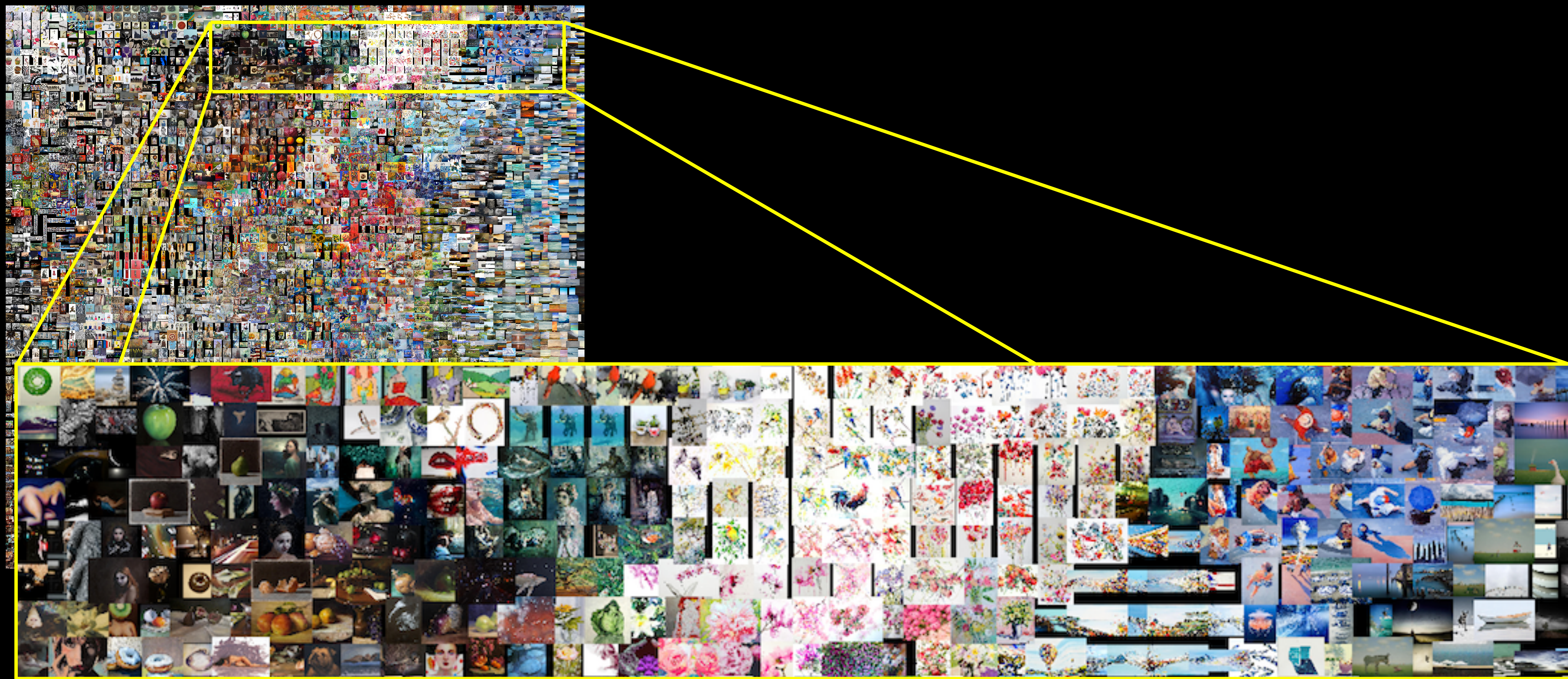
Detail

ID	File	Class	x	y
3008	img/37320.jpg	Clase 6	507	482

Ugallery 2: t-SNE + Jenker Volgenaut



Ugallery 2: t-SNE + Jonker Volgenaut



Vicente Dominguez, Pablo Messina, Denis Parra, Domingo Mery, Christoph Trattner, and Alvaro Soto. 2017.
Comparing Neural and Attractiveness-based Visual Features for Artwork Recommendation.
In *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems* (DLRS 2017). ACM,
https://doi.org/10.1145/3125486.3125495

t-SNE

- T-distributed Stochastic Neighbor Embedding
- Used frequently to plot embeddings or Deep Neural Networks

- Some resources:
- http://lvdmaaten.github.io/tsne/
- https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf

- Demos:
- https://cs.stanford.edu/people/karpathy/tsnejs/csvdemo.html
- http://scienceai.github.io/tsne-js/
- http://projector.tensorflow.org/