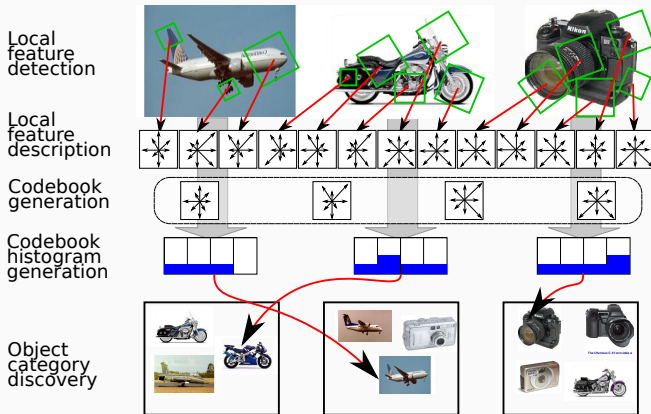


# Clustering approaches

---

Clustering and vector quantisation (coding)

# Quantisation example: visual “bag-of-words”



**Figure 4:** Generating visual words and the visual bag of words (BoW) approach to visual category detection.

## Other techniques

---

## Other techniques

---

### Self-Organising Map (SOM)

# The self-organising map (SOM) algorithm

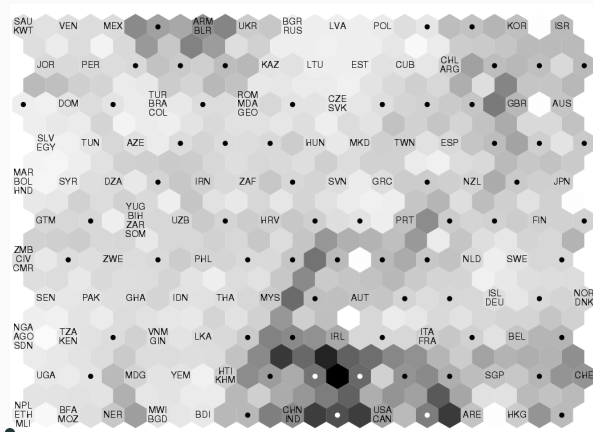
- Simple, robust and effective algorithm: organises N-dimensional data to fewer dimensions (typically 2-D: a map) retaining the topological structure: samples close each other in a SOM are close to each other in the original space (note: not vice versa)
- Developed by Finnish Academician Teuvo Kohonen in 80's
- Has competitors (MDS, LLE, ISOMAP, etc.) but remains as one of the best (and simplest)

## Example 1: Organizing countries (similar market areas?)

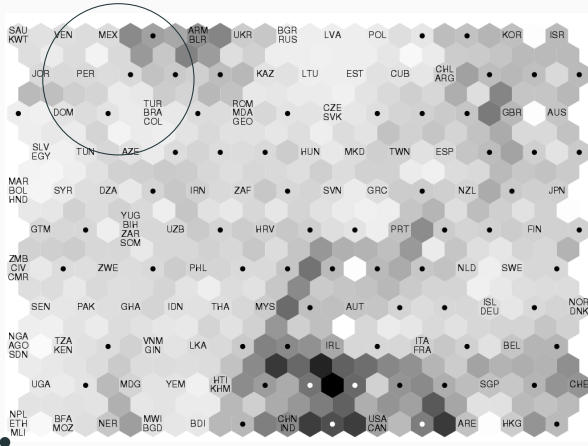
- Country specific information is recorded and we want to find similar countries
- U-matrix provides information on level of differences between the cells

	<i>GDP per capita</i>	<i>Population</i>	<i>Urbanisation percent</i>	<i>...</i>
Algeria	1779	25660000	50	...
Finland	24491	5029000	61	...
Japan	27436	124001000	77	...
...				...

## Example 1: Organizing countries (similar market areas?) (2)

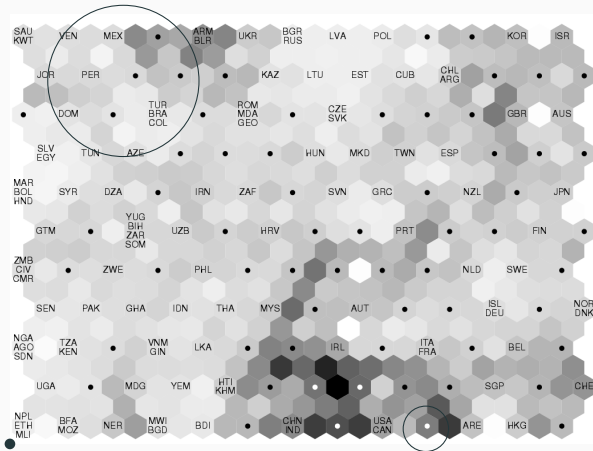


### Example 1: Organizing countries (similar market areas?) (2)

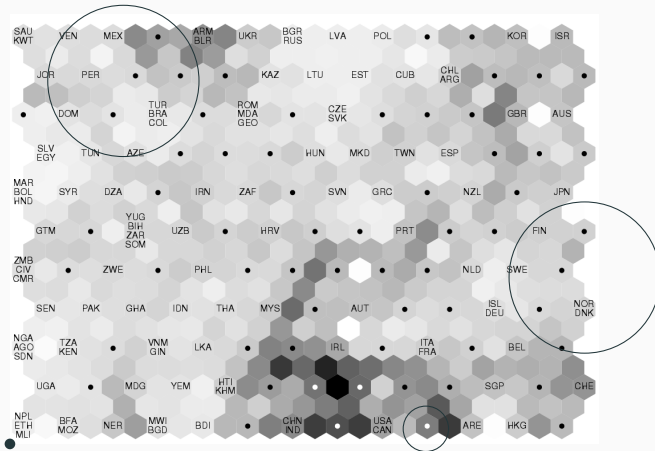




## Example 1: Organizing countries (similar market areas?) (2)



## Example 1: Organizing countries (similar market areas?) (2)

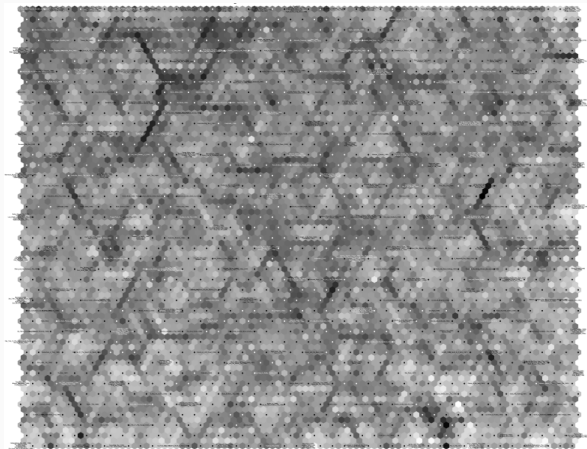


## Example 2: “What movies do you recommend?”

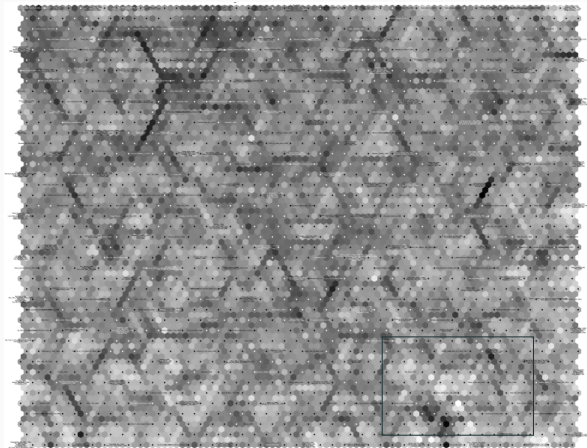
- Data are users' rates on films they have seen (available at the Internet movie database (<http://www.imdb.com>))

	<i>User 1</i>	<i>User 2</i>	<i>User 3</i>	<i>...</i>
Ace ventura	5	x	1	...
From dusk till dawn	4	2	x	...
Unforgiven	x	4	5	...
Toy Story	2	x	4	...
...				...

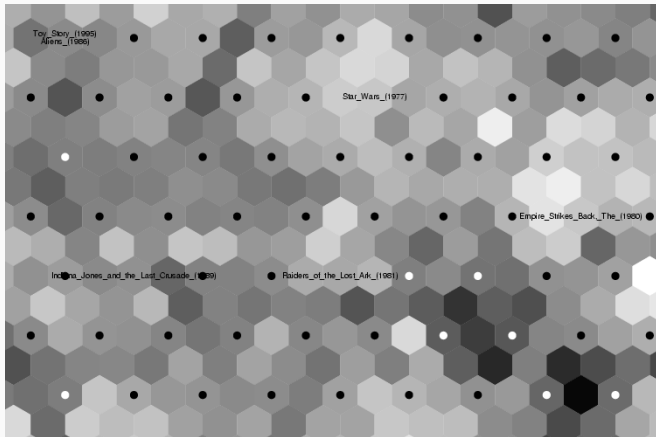
## Example 2: “What movies do you recommend?” (2)



## Example 2: “What movies do you recommend?” (2)



## Example 2: “What movies do you recommend?” (3)



## Example 3: “Who does look like me?” or “Do I look like a nerd?”

- Input is a large number of raw images of different persons
  - SOM structure 20x20 (*somstructure.png*)
  - U-matrix shows differences between “groups of similar looking people” (*somumatrix.png*)
  - We can draw faces stored in SOM (*somcodebook.png*)
  - We can query for any face image to which cell it is assigned (*somfaces01.png*)

## Example 4: Unsupervised Visual Classification

- SOM inputs are BoW codebook histograms for a large set of images
- Now we can inspect which images automatically locate near each other (*som\_80x50\_100best\_small.jpg*)



## Summary

---

- Clustering
  - Hierarchical
  - Mixture models
  - Sum-of-squares - K-means
  - Spectral
- Visualising high dimensional data: the self-organising map (SOM) algorithm