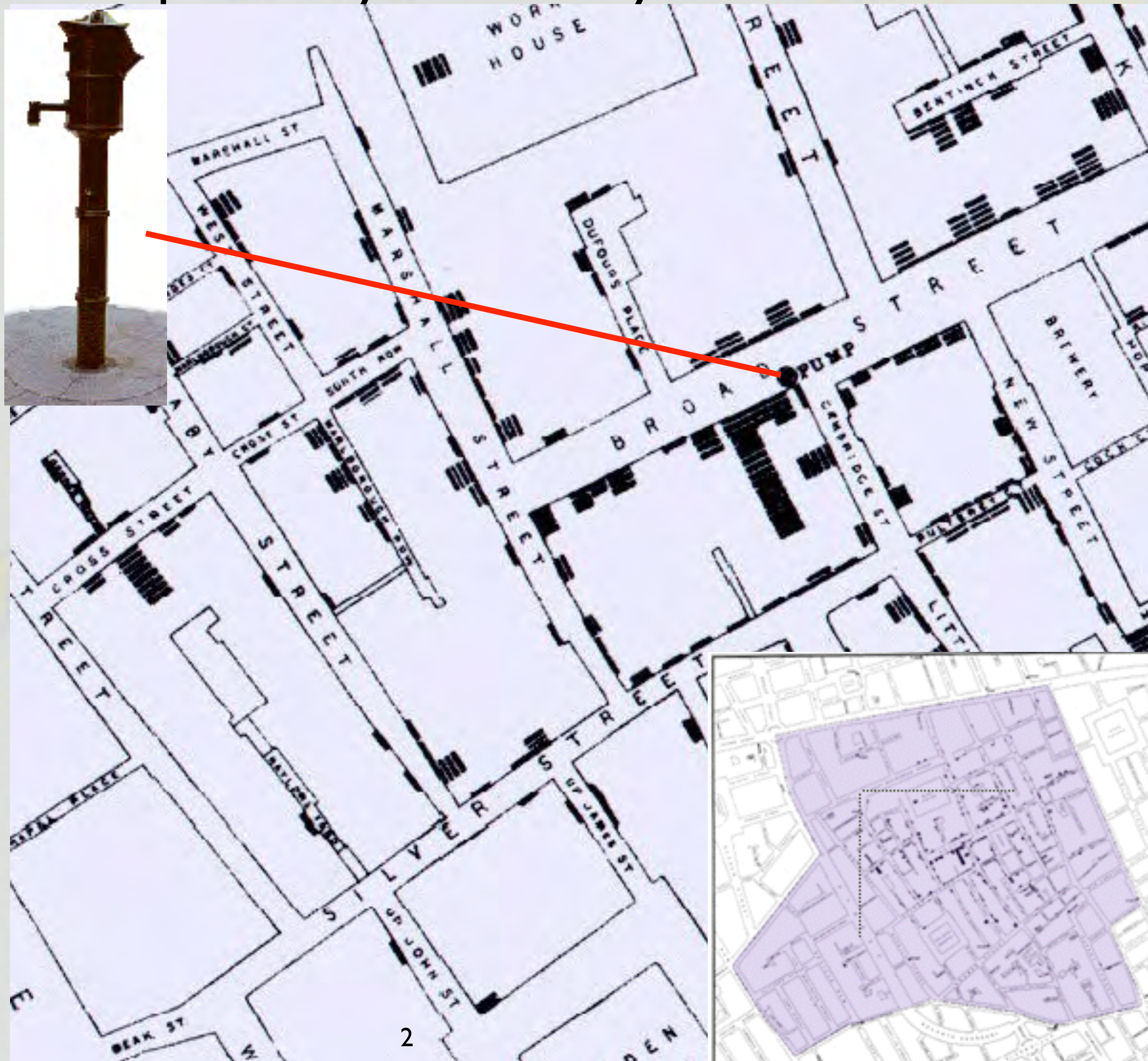
L5:Exploratory Data Analysis

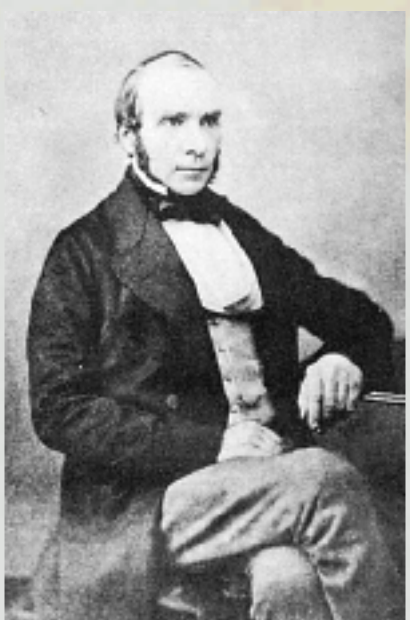
Demšar, Exploring geographical metadata by automatic and visual data mining, licenciate thesis, available for downloading at: http://www.infra.kth.se/~demsaru/publications/
- chapter 1: Data mining

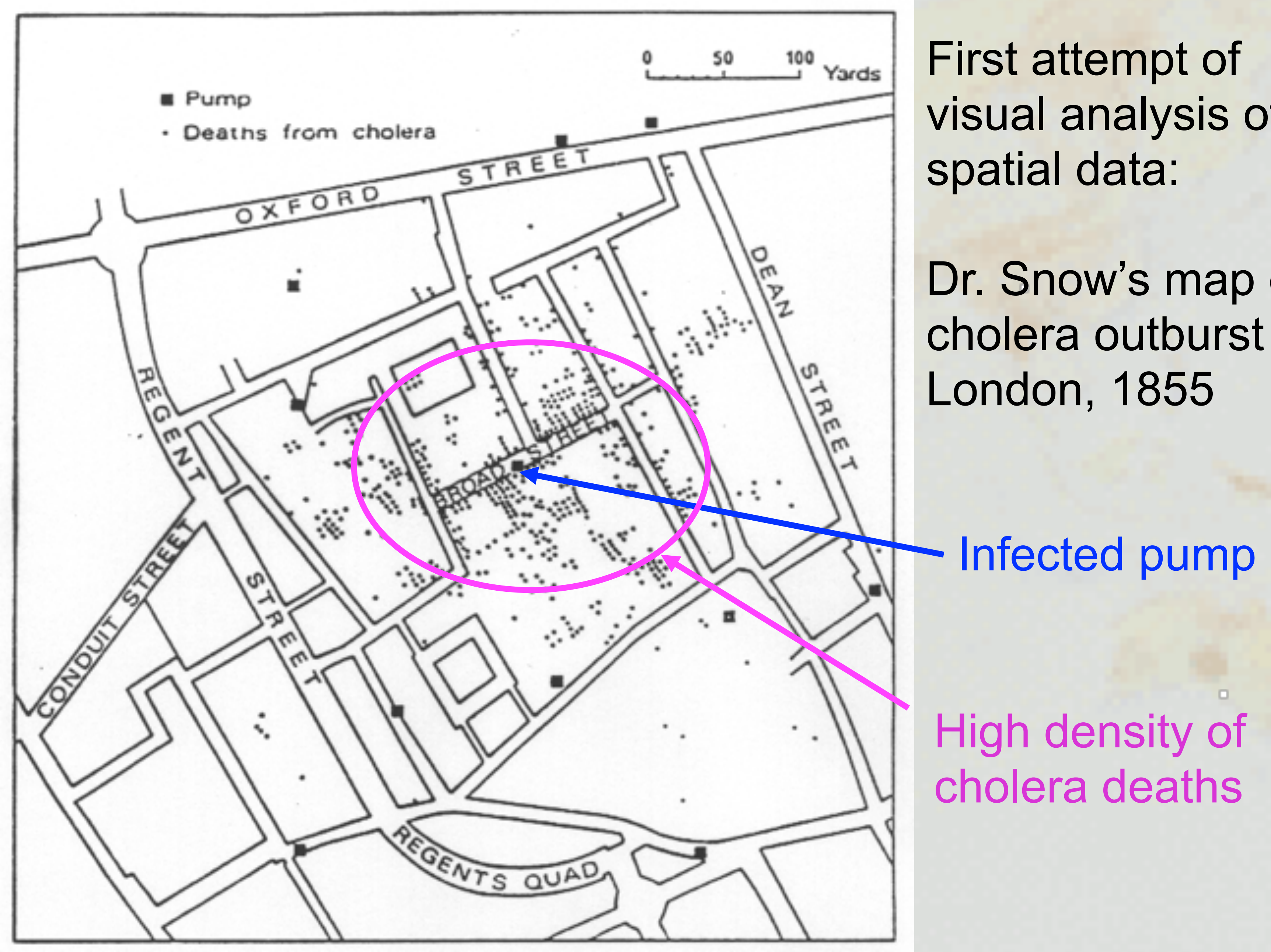
Kraak & Ormeling, Cartography – Visualization of Geospatial Data
- chapter 12: Geovisualization

Exploratory Data Analysis

First attempt of visual analysis of spatial data: **Dr. Snow's map of cholera outburst in London, 1855**







First attempt of visual analysis of spatial data:

Dr. Snow's map of cholera outburst in London, 1855

Exploratory Data Analysis

- Local (point)
- Focal (nearest neighbor)
- Regional (network)

- Profile data (2D, 2.5D, 3D, 3.5D,4D)

- Time series

Exploratory Data Analysis

- Questions (queries)
- Measurements
- Transformations
- Descriptive methods
- Optimisation
- Hypothesis testing

Exploratory data analysis then and now

Then (before GIS):
tools to study and explore geospatial data =
paper maps + statistics

Exploratory
geovisualisation

Now:

data mining

cartography

algorithms

database tools

Geospatial data → Knowledge

graphic tools

interaction

GIS

image analysis

exploratory data analysis

dynamics

information visualisation

Exploratory Data Analysis

Requirements for a geovisualisation system:

- basic display (map + pan, zoom, scale, transform, rotate)

- orientation & identificaiton (where the map is located, what the symbols mean)

- query data (perform logical queries in the database)

- multi-scale tools (combining data in different scales from different sources)

- re-expression (have possibilities to manipulate the data or choose between different mapping methods)

Exploratory Data Analysis

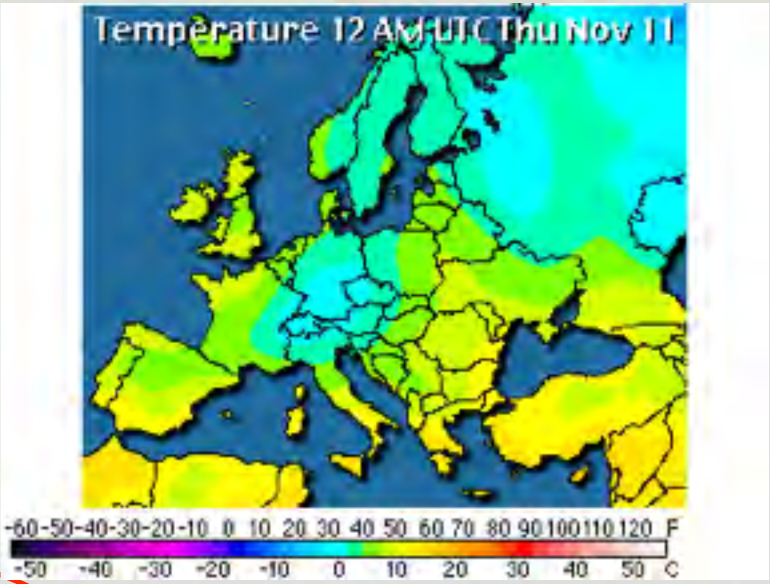
Database queries

- For spatial queries, we can also do other simple queries that used to require overlay analysis, which is even better in the object-oriented programs like ArcGIS
- Most of these searches must use some form of topology, logic, and advanced SQL (standard Query Language) to work
- Finally, have to know your dimensionality (0,1,2,3) for application, but most of these relate to the vector model
  - Equal – are the geometries the same?
  - Disjoint – do the geometries share a common point?
  - Intersects – do the geometries intersect?
  - Touches – do the geometries intersect at their boundaries?
  - Crosses – do the geometries overlap?
  - Within – is one geometry within another?
  - Contains – does one geometry completely contain another?
  - Overlaps – do the geometries overlap?
  - Relate – are there intersections between the interior, boundary, or exterior of the geometries?

Exploratory Data Analysis

- multiple dynamically linked views (brushing and linking)

- animation (temporal or non-temporal changes in data)



- exploratory tools (data mining, computational and visual)

Exploratory Data Analysis

Data mining

Data mining:
Identifying or discovering interesting and as yet undiscovered knowledge from the real-world data.

Part of **Knowledge discovery process:**

1. Data preparation and cleaning
2. Hypotheses generation
3. Interpretation and analysis of discovered knowledge

DM

Difference from statistical analysis
DM works on observational data as opposed to experimental data and has no role in the strategy of data collection.

Exploratory Data Analysis

Data mining:

Identifying or discovering interesting and as yet undiscovered knowledge from the real-world data.

Knowledge is in the form of (undiscovered) structure in the data.

Structure → pattern → description of a subset of data points

Structure → model → description of the whole dataset

Structure → relation → dependency between attributes over a subset of the data (deductive or inductive)

Exploratory Data Analysis

Data mining classification

Task

Predictive DM

Goal: build a model to predict a value of one attribute from other attributes

classification, regression

Exploratory DM

Goal: explore the data and identify the structure beneath

clustering, SOM

Reductive DM

Goal: data reduction – aggregate information in large datasets into smaller manageable subsets

principal component analysis (PCA)

Exploratory Data Analysis

Data mining classification

Data type and mining environment

Database data mining
Web data mining
Text data mining
Distributed data mining
Ubiquitous data mining
Hypertext and hypermedia data mining
Visual data mining
Multimedia data mining
Spatial and geographical data mining
Time series and sequence data mining

Exploratory Data Analysis

Forward and backward driven data mining

Forward (or data) driven:

Backward (or goal) driven:

Exploratory Data Analysis

Knowledge based or statistical data mining

Knowledge based methods:

- Expert systems (rule based)
- Decision trees
- Supervised classification

Statistical methods:

- Wavelet analysis (Fourier transformation)
- Principal Component Analysis
- Clustering
- Artificial Neural Networks
- Self Organising Maps (SOM)

Automatic (computational) data mining

Automatic algorithms: look for structural patterns in data

- rules
- decision trees
- clusters
- matching between a concept description and a data instance
- instance-based representations

Groups of automatic DM methodologies:

- Expert system
- Decision trees
- Association rules
- Classification models
- PCA, SOM, Wavelet
- Clustering
- Bayesian (apriori known)
- Artificial neural networks
- Instance-based learning
- Sequential pattern mining
- Time series mining

No data mining algorithm is universally best across all datasets!

↓

Choice of appropriate methodology is task-driven.

Common data mining algorithms

Classification rules - Expert systems

Divide the dataset into prespecified classes, defined by the values of the attribute that is predicted.

IF ….. **THEN** …..

Antecedent – a series of tests that compare an attribute value to a constant

Consequent – determines the class of the processed data instance by assigning a value to the predicted attribute.

Example (forward driven expert ruling):

IF outlook=“sunny” AND temp=“warm” AND wind=“no” **THEN** walk=“yes”.

IF outlook=“rainy” AND temp=“cold” AND wind=“yes” **THEN** walk=“no”.

Decision trees

A decision tree = a tree that classifies each data instance by applying to it a test at each node:
- enter the data instance in the tree at the root,
- let it “fall” down according to the tests,
- the leaf nodes give the classificiation.

Association rules

Similar to classification rules, but:
- can predict any attribute (not just one),
- can predict a combination of attributes.

IF X **THEN** Y (s, c%).

X and Y - sets of predicates
c – confidence of the rule – the number of instances that it predicts correctly, expressed as a proportion of all instances it applies to
s – support of the rule – the number of instances for which it predicts correctly

Example:
IF temperature = “cool” THEN humidity = “normal”.

c = the proportion of cool days that have normal humidity.
s = the number of days that are both cool and normally humid.

Numerical prediction: linear models, support vector machines, regression and model trees

For numeric attributes.
Methods:

- linear regression:
  predicted value = a linear combination of other attributes

- support vector machines:

Input data

f – non-linear function

Linear space

Lin.regression

Predicted value

f^{-1}

Original space: non-linear classification

- regression trees & model trees
  store the average value for each class or a linear regression model for each class in each leaf.

Exploratory Data Analysis

Instance-based learning

Searching the most similar already known data instances.

Methods:

- nearest neighbour:
  find the nearest training instance I_t to current data instance I and assign class of I_t to I.

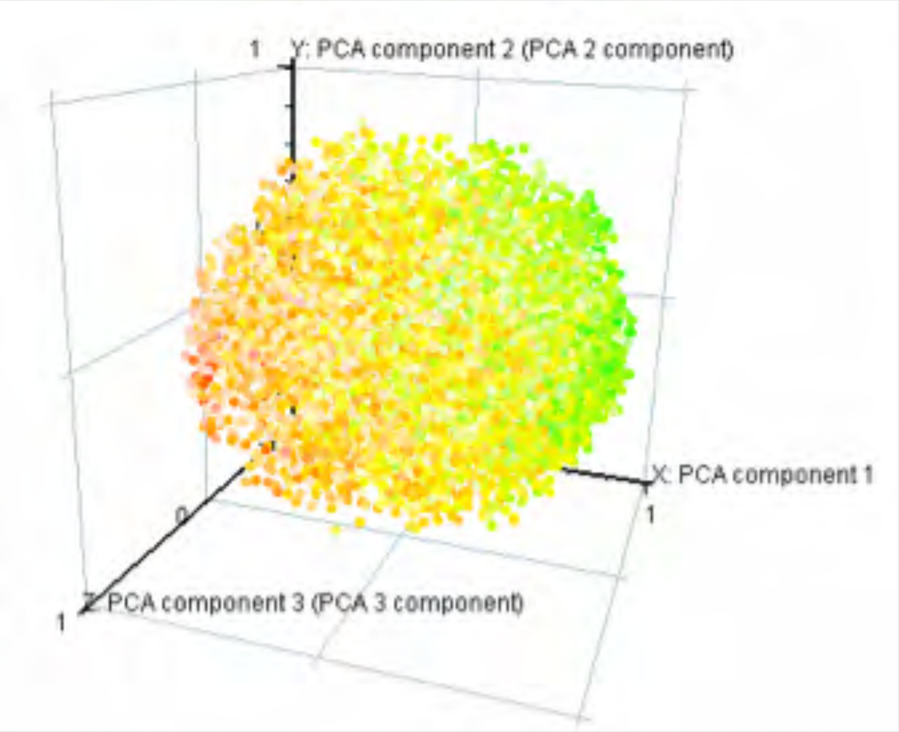
- k-nearest neighbours:
  find the k nearest training instances and assign the class according to their classes.

Exploratory Data Analysis

Principal Component Analysis

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

PCA can be used for dimensionality reduction in a data set by retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the "most important" aspects of the data. But this is not necessarily the case, depending on the application.

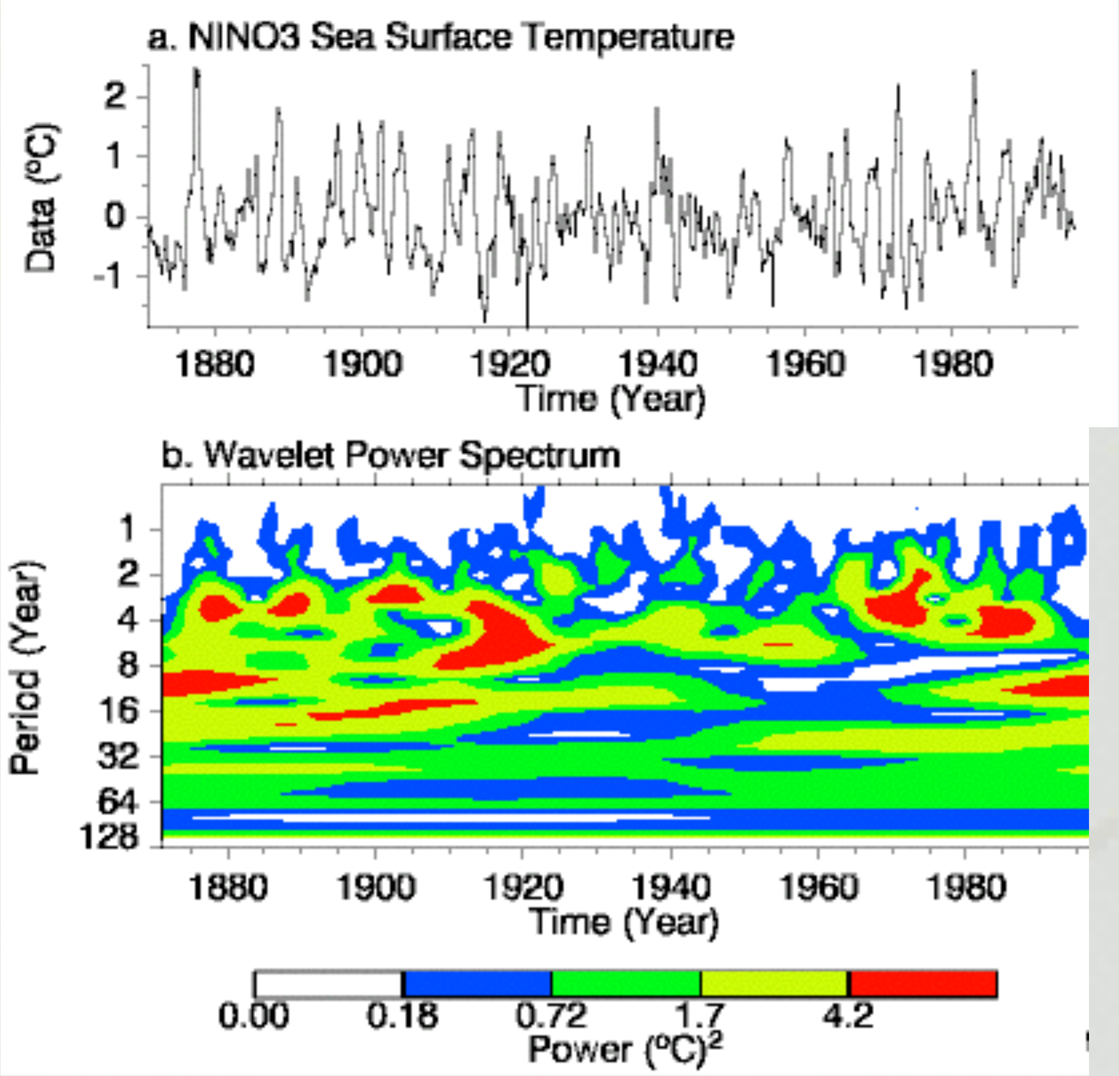


Exploratory Data Analysis

Wavelet analysis

The Niño3 SST index is defined as the seasonal SST averaged over the central Pacific (5°S–5°N, 90°– 150°W)

Interactive Wavelet analysis tool:
http://ion.researchsystems.com/cgi-bin/ion-p



Exploratory Data Analysis

Clustering

Unsupervised classification of data instances into groups/clusters according to similarity.

difference

Classification ⟷ Clustering

Classification → predefined groups → training data

Clustering → data-driven groups → obtained solely from data, without any training data

Clustering - example africa veg

Exploratory Data Analysis

Similarity measures: data type and mining task

- Euclidean distance
- Manhattan distance
- Mahalanobis distance
- count-based measures for nominal attributes
- syntactic measures for strings
- neighbourhood-measures

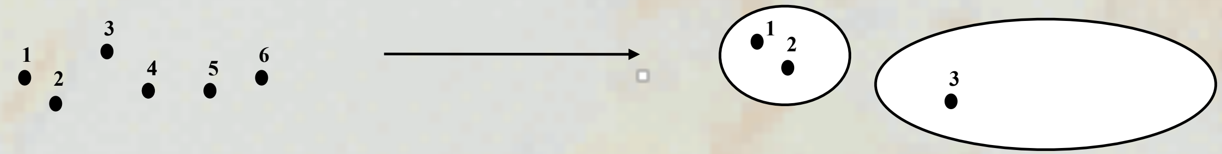
Types of clustering:

- Hierarchical vs. partitional clustering
- Agglomerative vs. divisive clustering
- Hard vs. fuzzy clustering
- Deterministic vs. stochastic clustering

Exploratory Data Analysis

Partitional clustering

produces one partition only.



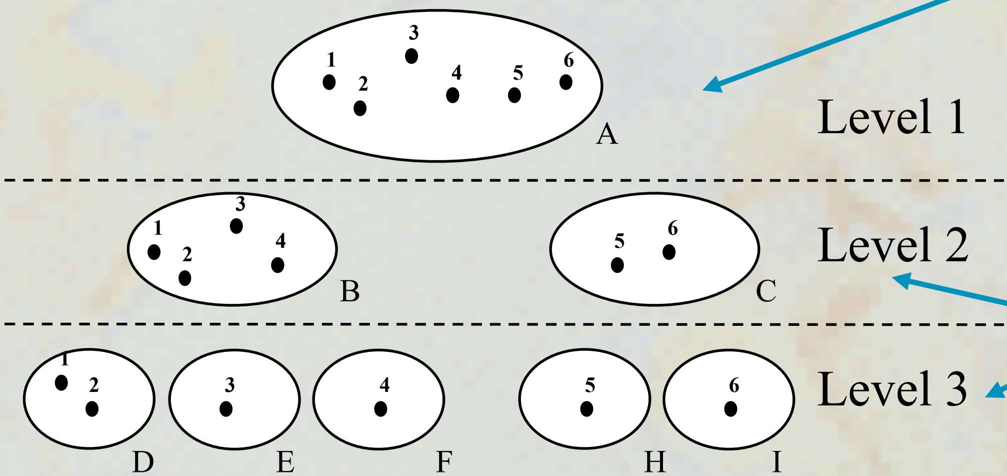
The set of elements is split into clusters only once.

Hierarchical (iterative) clustering

produces a nested structure of partitions – a hierarchy of clusters.

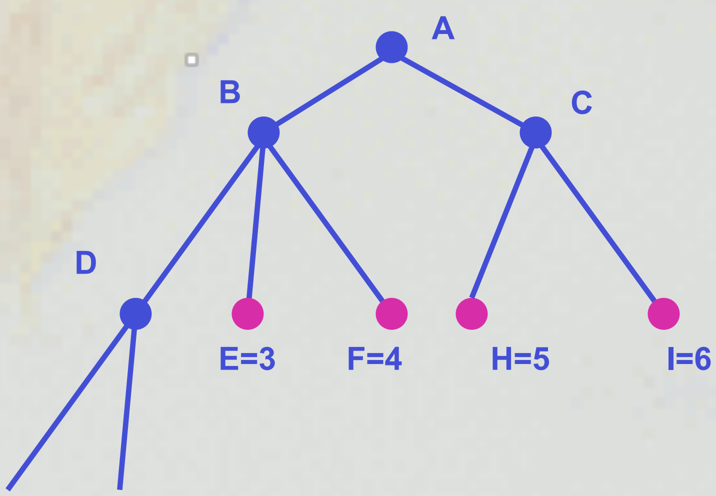
clusters divided in sub-clusters, similarly to a mathematical tree (a dendrogram)

Exploratory Data Analysis



higher level:
elements are less similar

lower levels:
elements are more similar

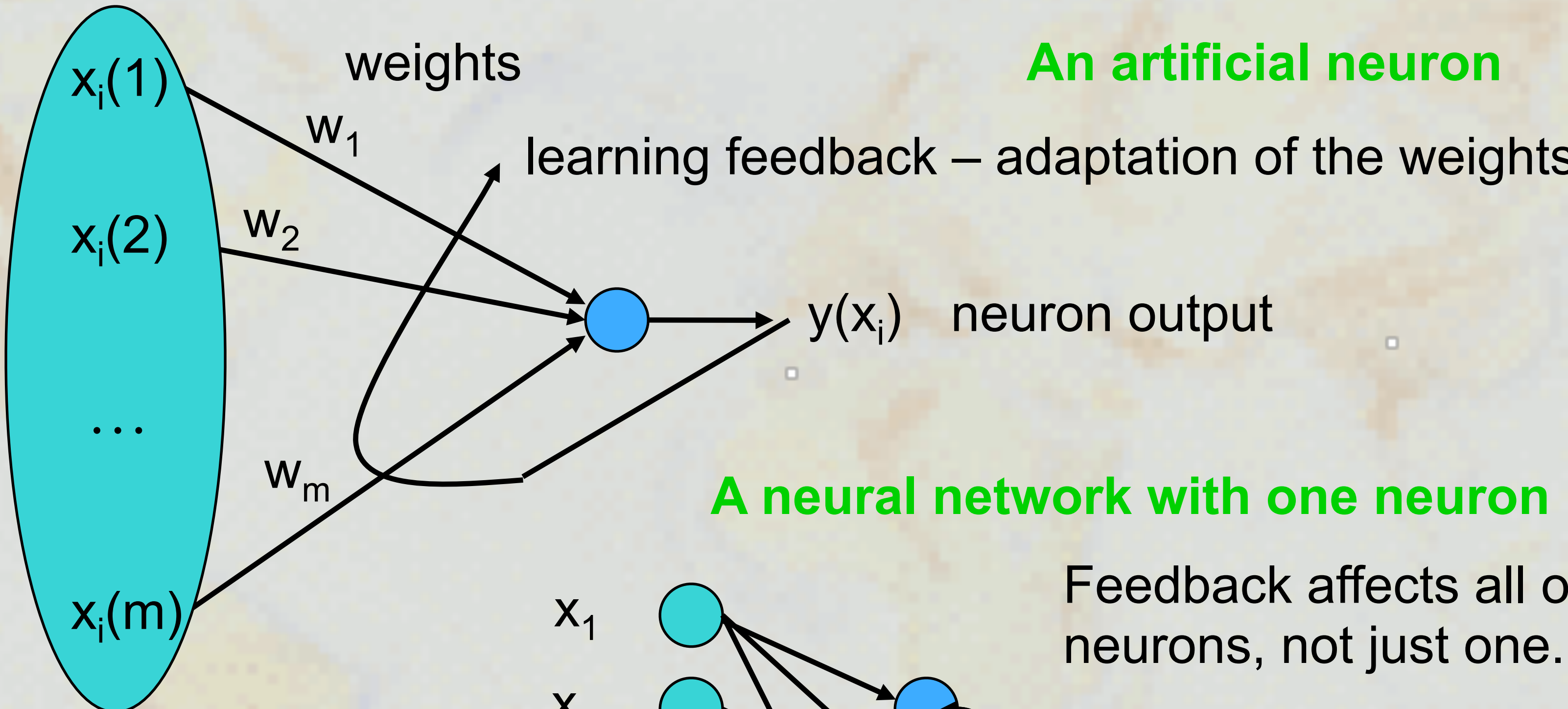


a dendrogram of this clustering

- data elements
- clusters with >1 element

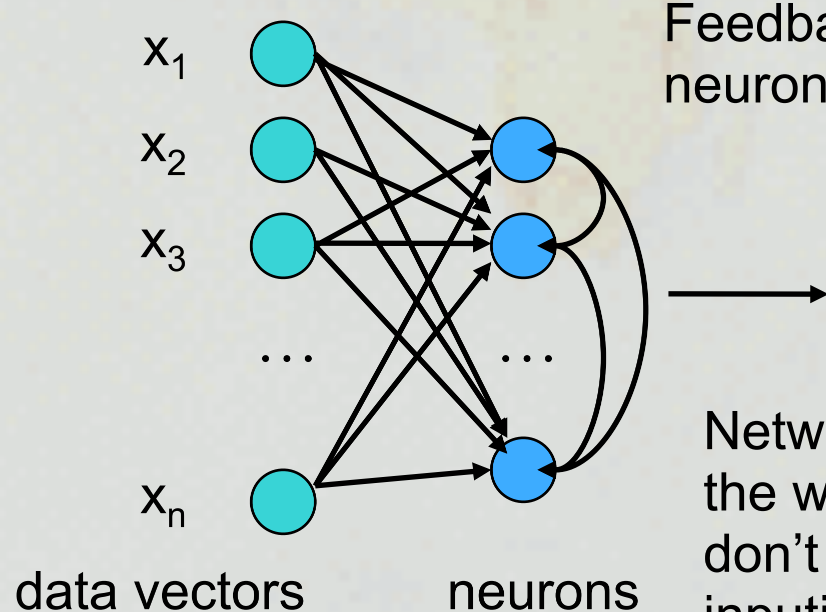
Artificial neural networks - ANN

An artificial neuron



input: data vector x_i, with m variables (dimensions)

A neural network with one neuron layer



Feedback affects all other neurons, not just one.

Network output = classification of the data input vector x_i.

Network is trained when the weights are stable: they don't change anymore when inputing a known data vector.

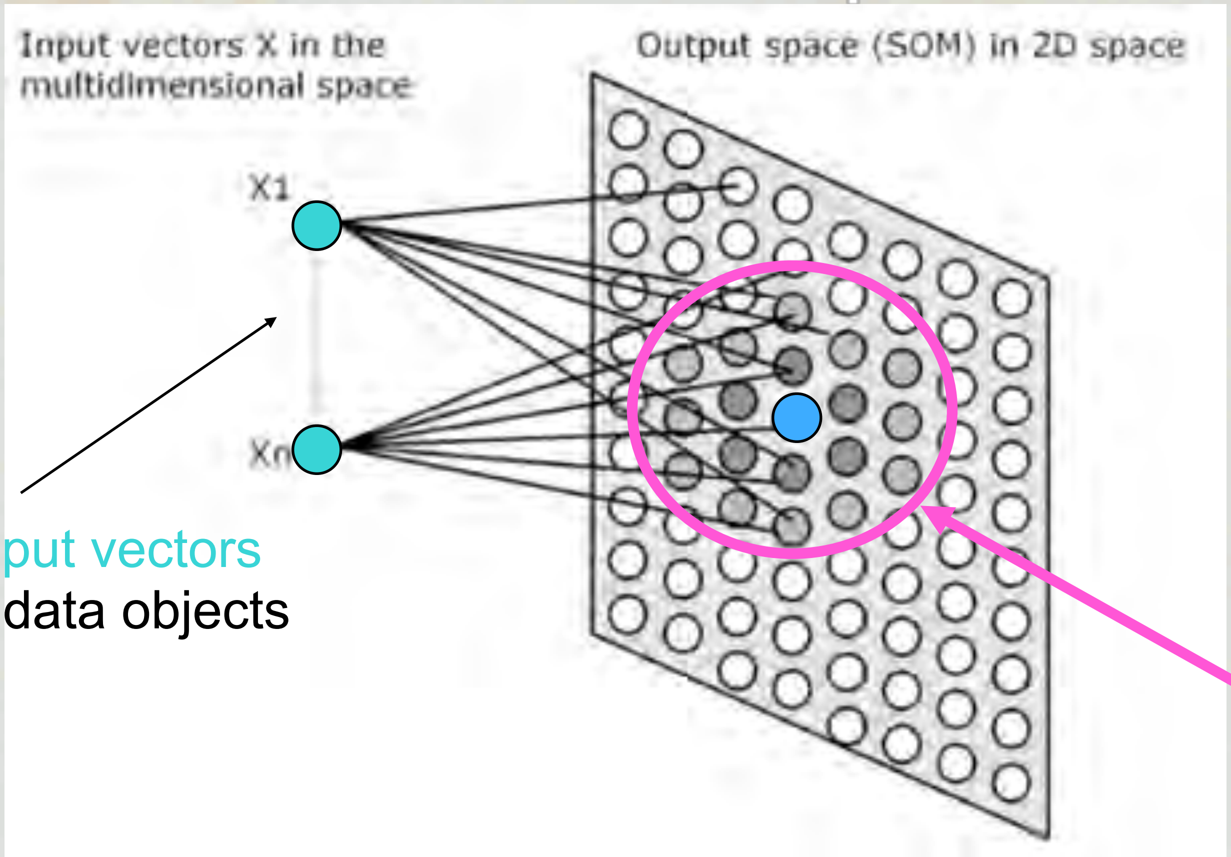
Self-organising map - SOM

Single-layer ANN – useful for spatial data

Non-linear projection of multidim. data on a 2-dim. lattice of neurons.

Neuron output = distance (similarity) from the neuron to the data input vector x_i.

Network output = location of the neuron that is most similar to the data input vector x_i.



Input vectors = data objects

Output cells = neurons in the SOM, distributed in the 2D space

Each neuron is connected to the neurons in its immediate neighbourhood – its output affects their weights.

Result: similar data vectors are mapped to similar locations.

Exploratory Data Analysis

Data mining of geospatial data

Problems and challenges

- Four dimensions of the information space provide the measurement framework for all other attributes.

- Spatial dependence - Tobler's 1st law of geography:
- Everything is related to everything else, but nearby things are more related than distant things.

- Large amount of data in geospatial databases:
- georeferenced RS imagery, socio-economic and statistical data, physical data, environmental data, etc..

- Heterogeneous data:
- semi-structured, unstructured, complex objects

Exploratory Data Analysis

Tobler's 1st law of geography & spatial dependence

Assumptions of independence and random distribution of variables in classical data mining algorithms **not valid!**

Data mining for geospatial data

Automatic data mining

Visual data mining

Combining visual and automatic mining

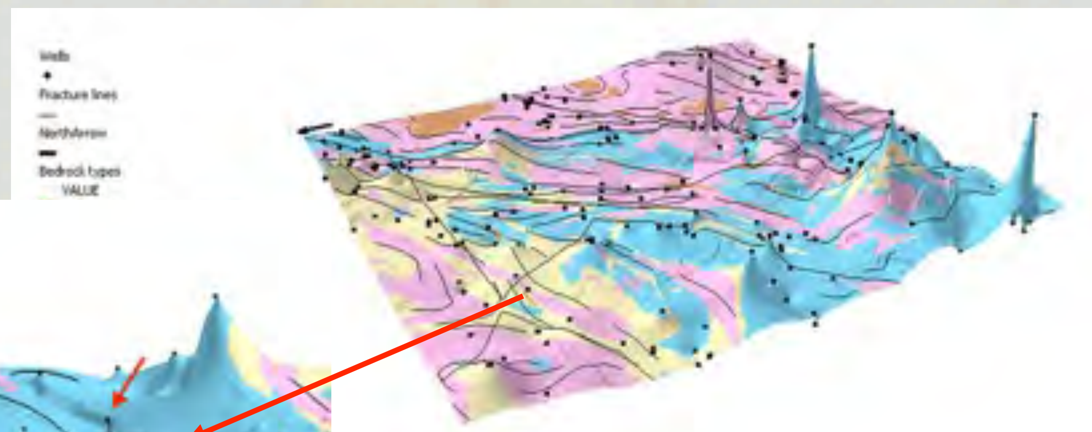
Current methods: spatial clustering, outlier analysis, spatial classification, spatial association rules.

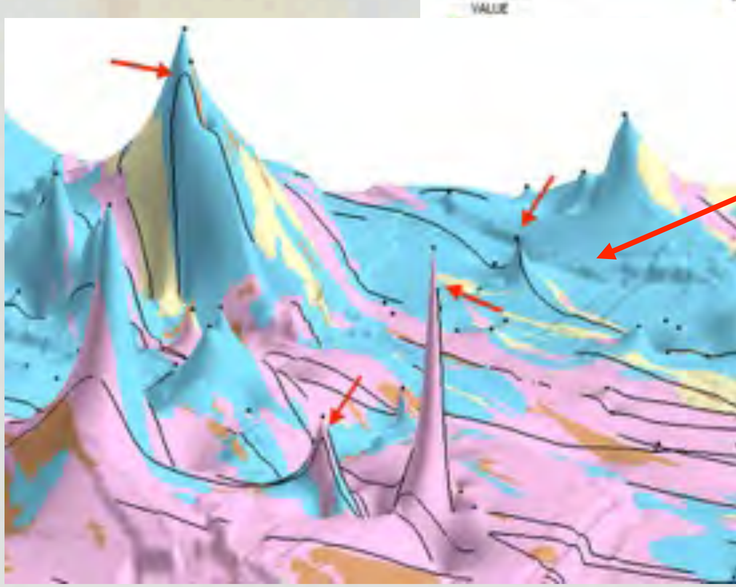
Visualisation and visual data mining

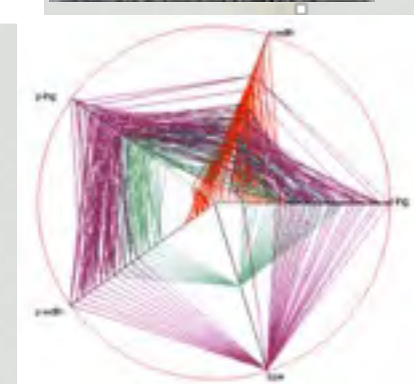
Visualisation

graphical communication of information

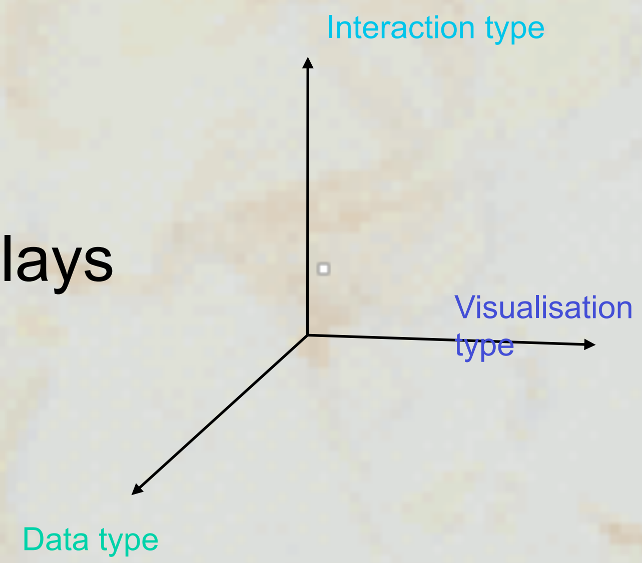
goal: present overview and summary of data, help to identify patterns and structures in the data







Classification of visualisations



Visualisation type:

- Standard 1D/2D/3D displays
- Geometrically transformed displays
- Iconic displays
- Dense pixel displays
- Hierarchical displays

Data type:

- 1-dim data
- 2-dim data
- multi-dim data
- text and hypertext
- hierarchies and graphs

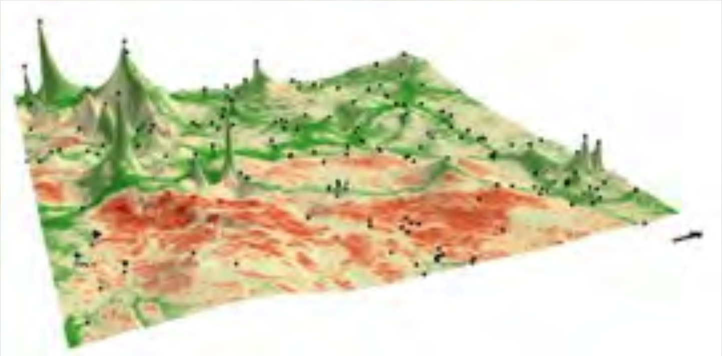
Interaction type:

- projection
- filtering
- zooming
- distortion
- brushing and linking

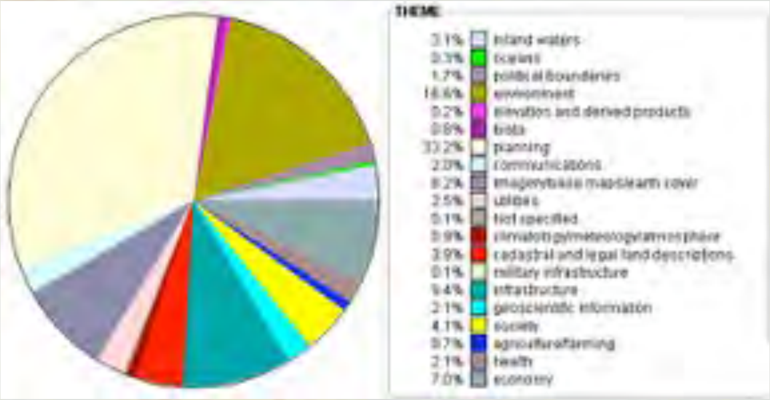
Examples of visualisations for Exploratory data analysis

Standard 1D,2D and 3D displays

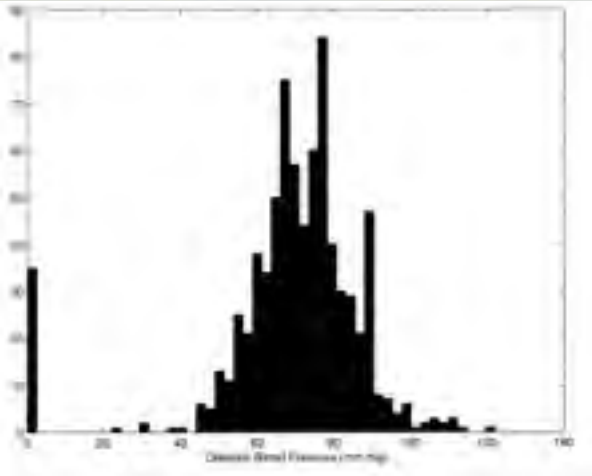
Line graphs, surfaces



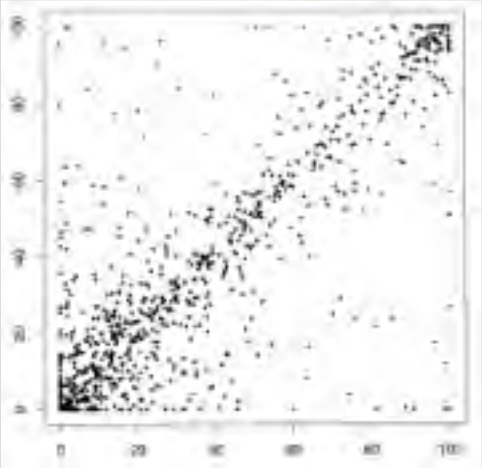
A pie chart



A histogram/bar chart



A scatterplot

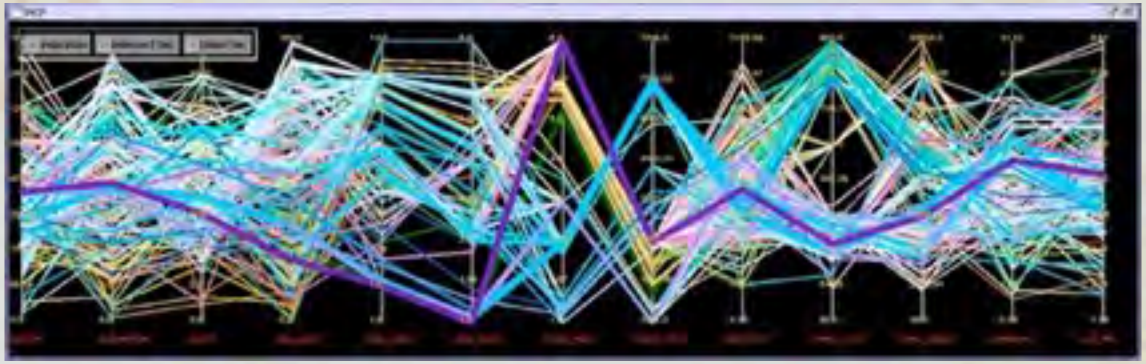


Examples of visualisations for Exploratory data analysis

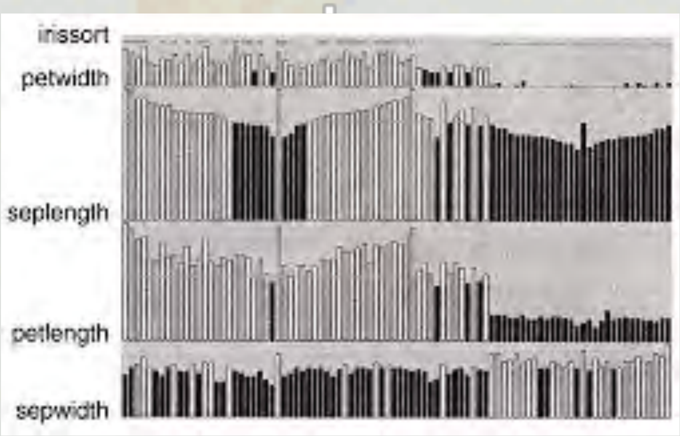
Geometrically transformed displays

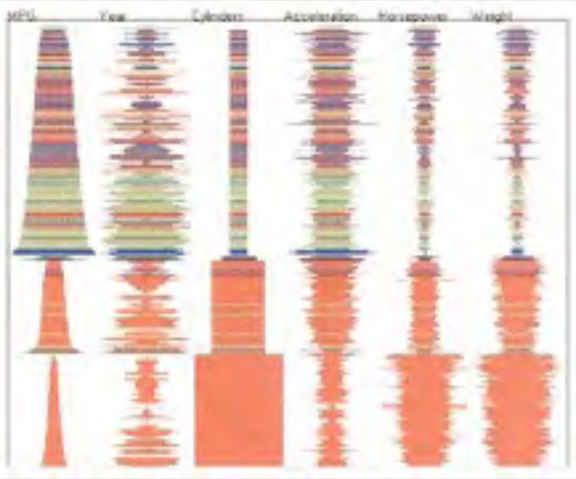
A scatterplot matrix

Parallel coordinates plot (PCP)



A permutation matrix and a survey plot



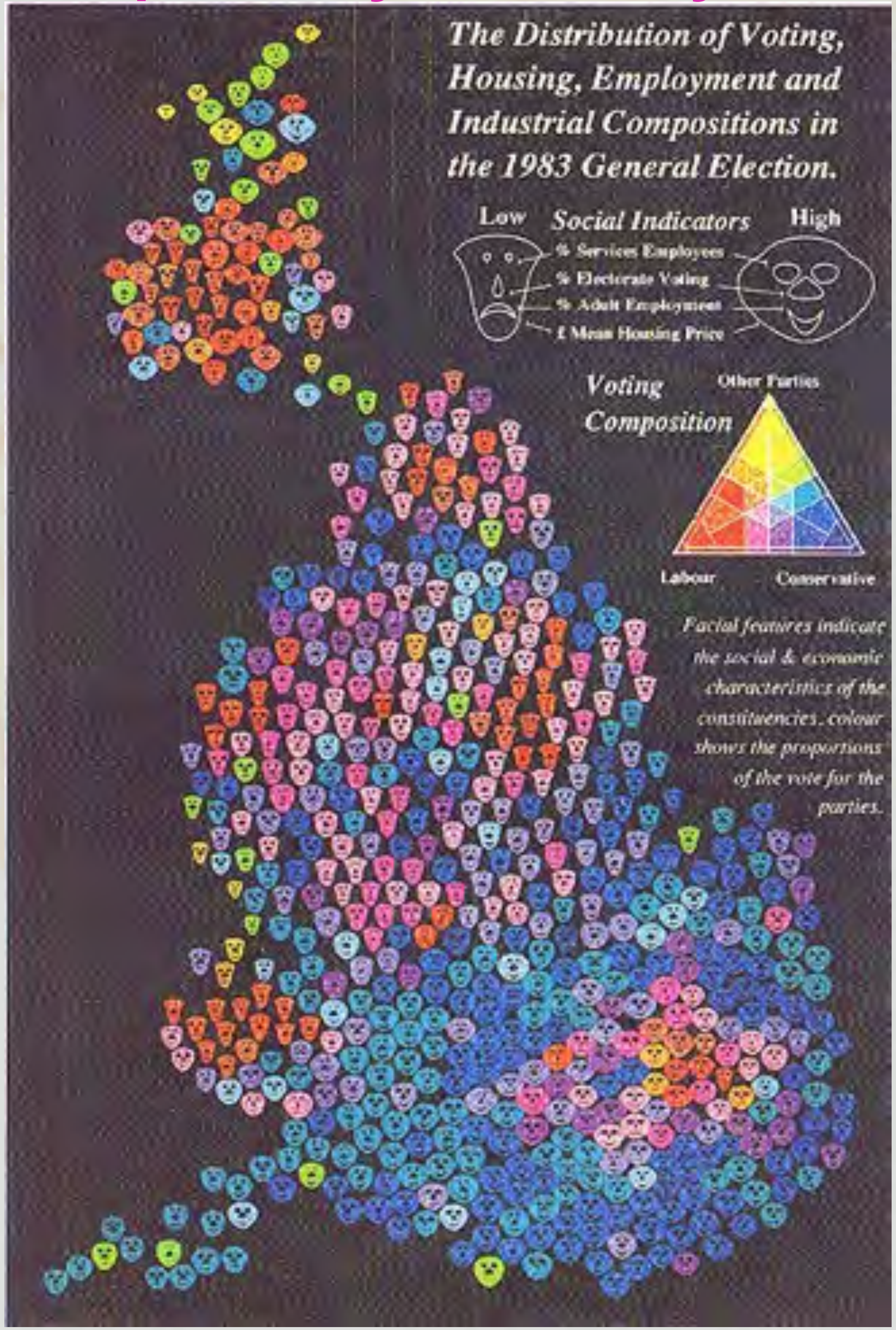


Examples of visualisations for Exploratory data analysis

Iconic displays

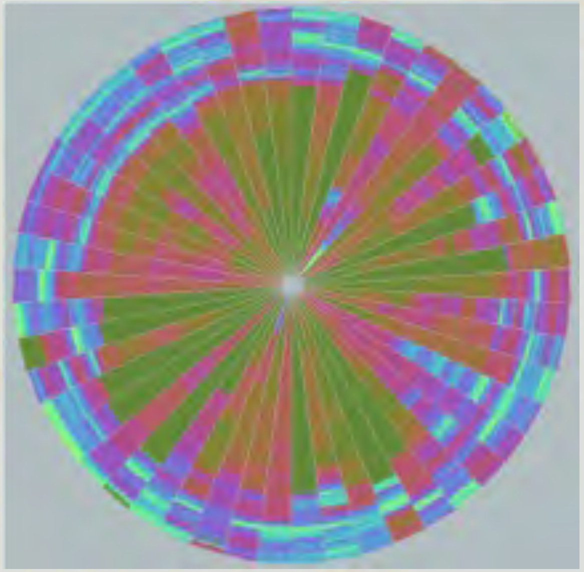
Chernoff faces

Star icons

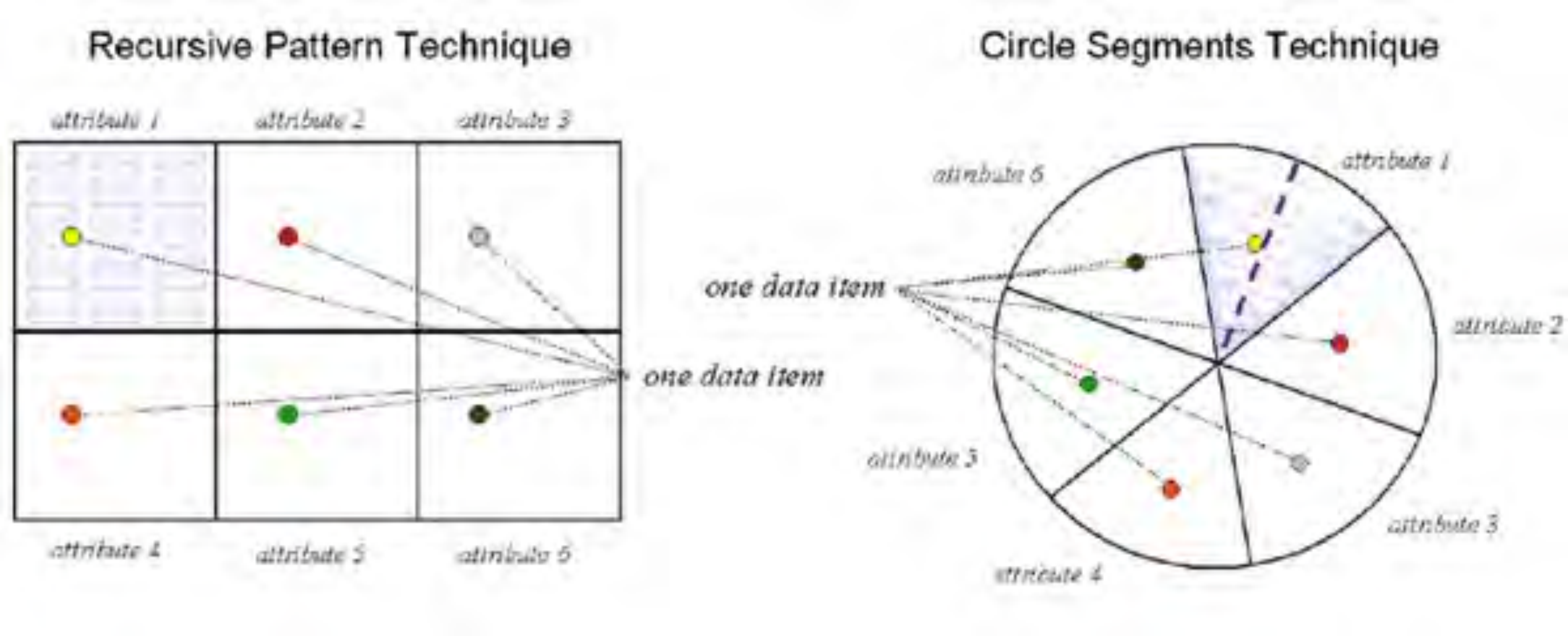


Examples of visualisations for Exploratory data analysis

Dense pixel displays



Recursive pattern Circle segments

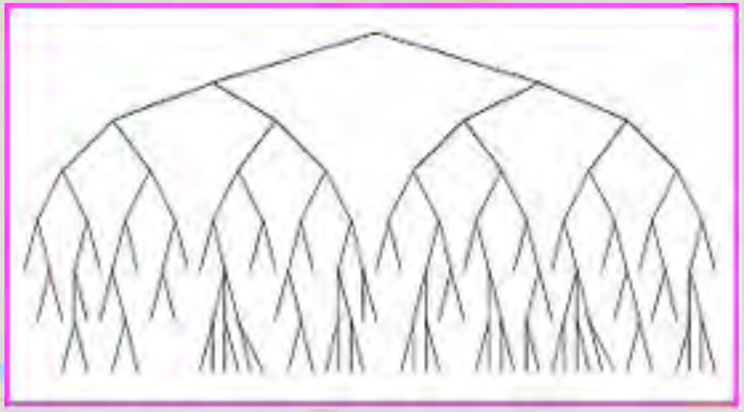


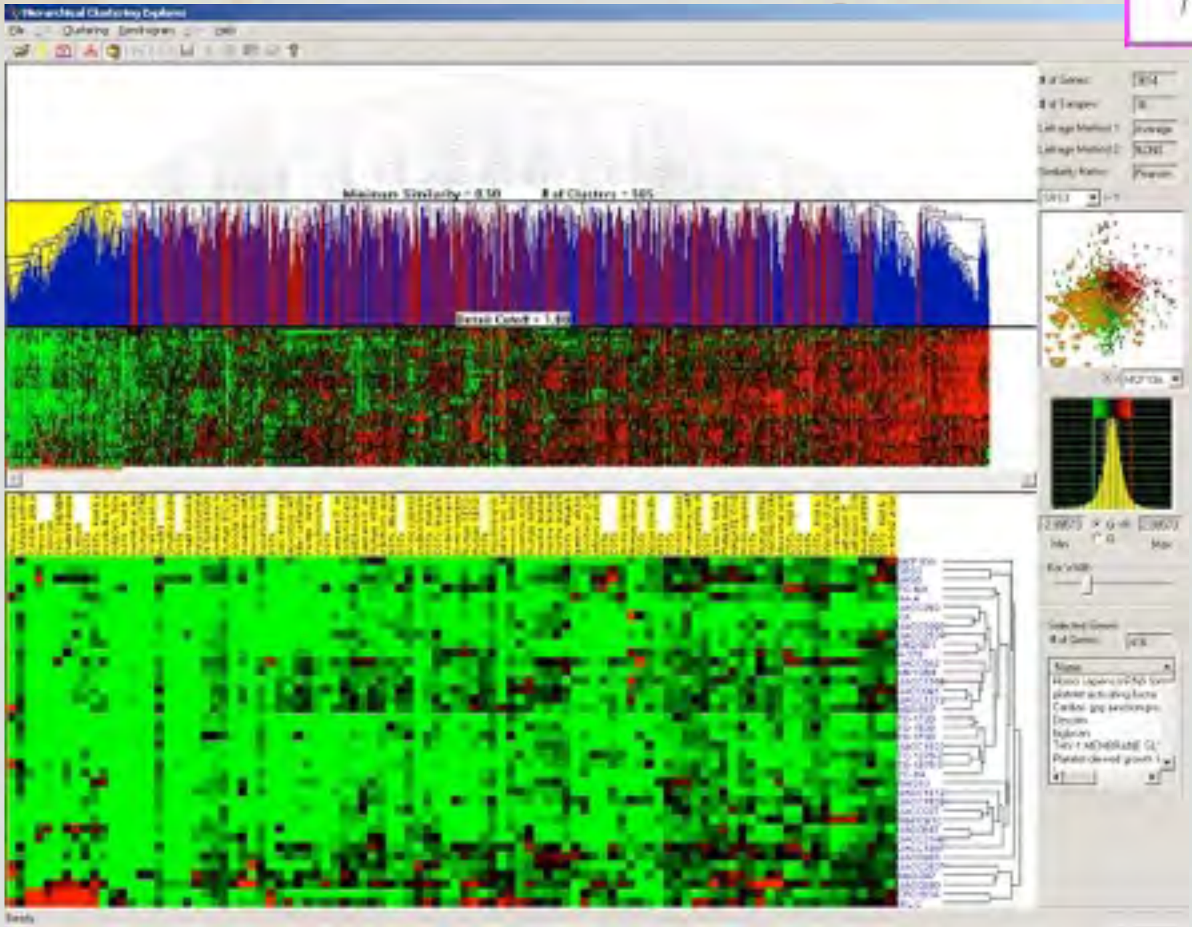
Examples of visualisations for Exploratory data analysis

Hierarchical visualisations

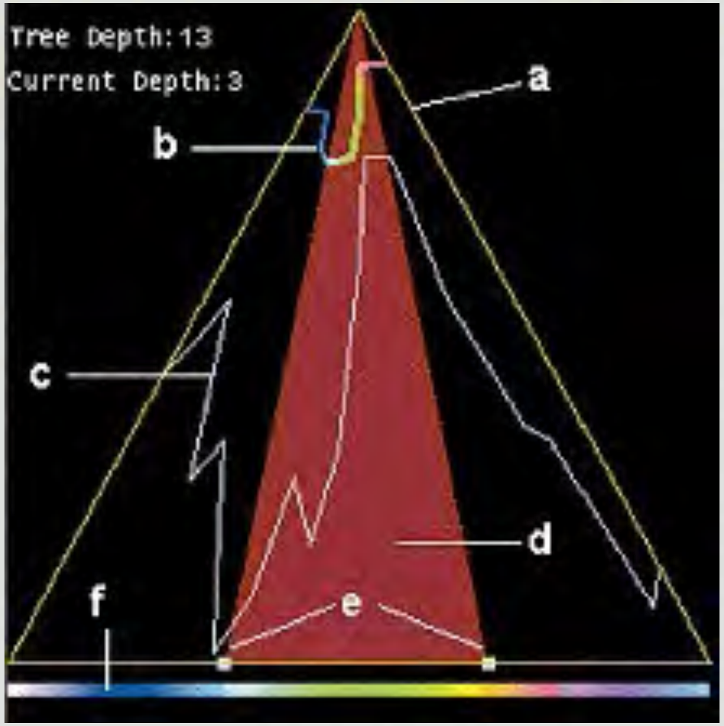
Visualising the result of hierarchical clustering

A dendrogram



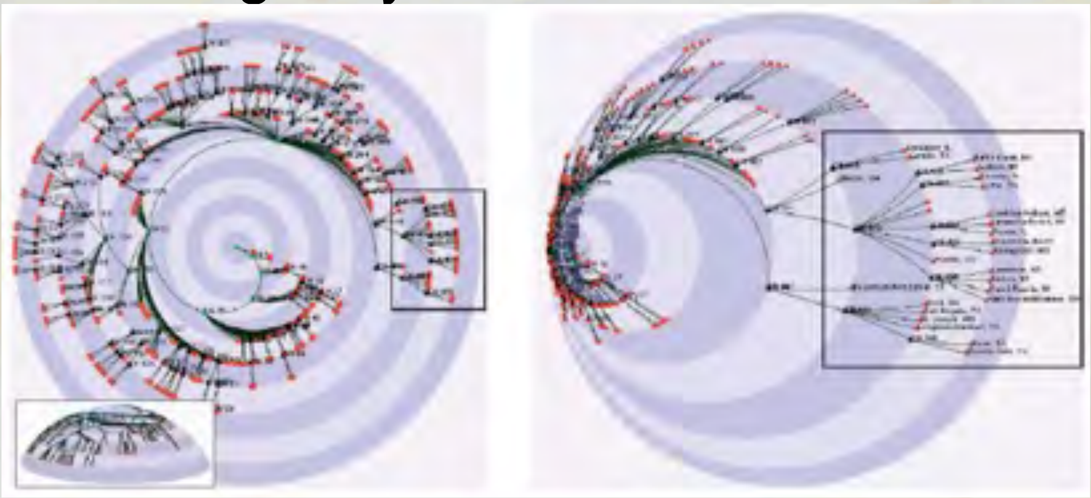


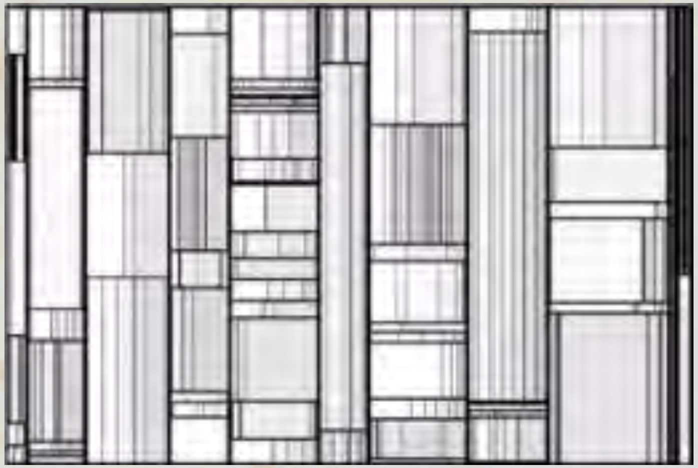
Structure-based brushes



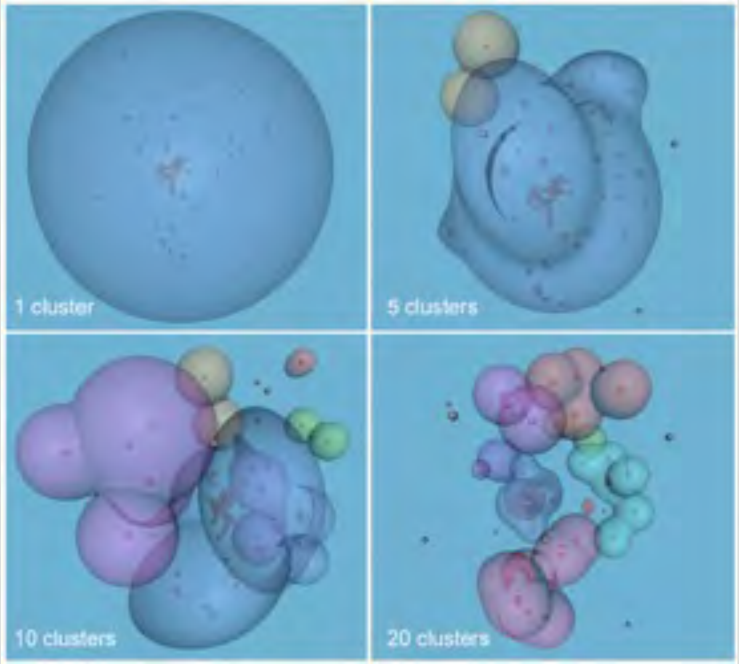
Examples of visualisations for Exploratory data analysis

The Magic Eye View



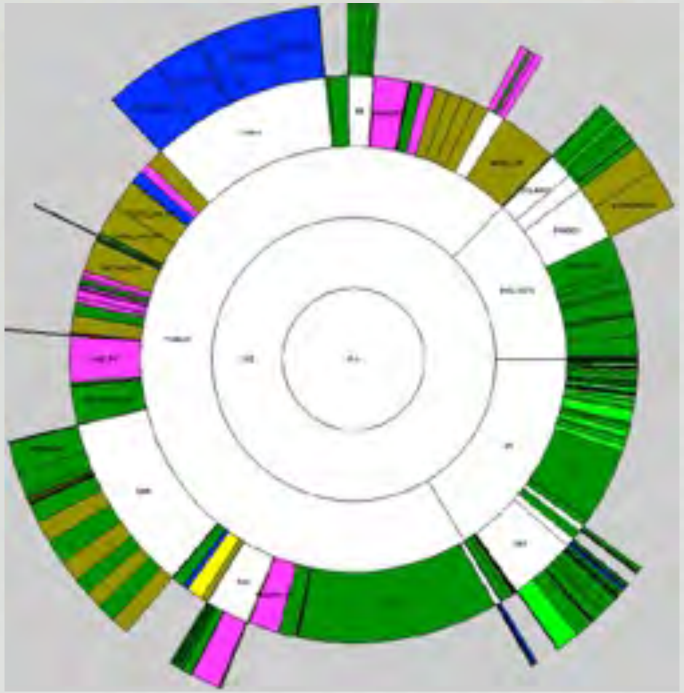


A treemap



H-BLOB

A sunburst



Examples of visualisations for Exploratory data analysis

SOM visualisations

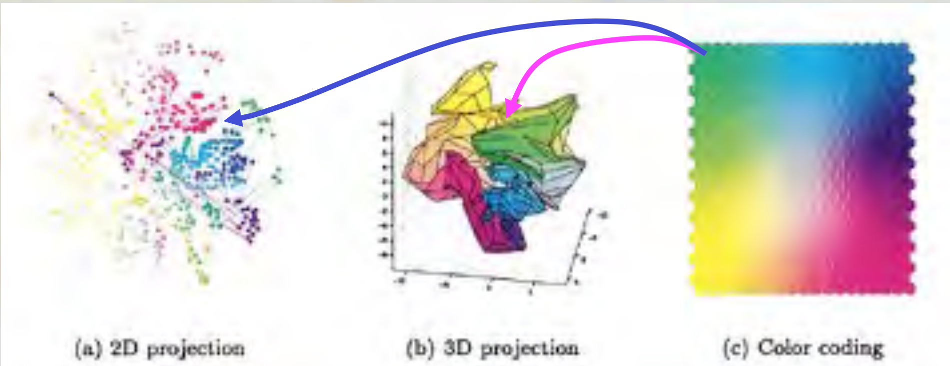
Visualisations based on **dist.matrices**

distance to neighbour units

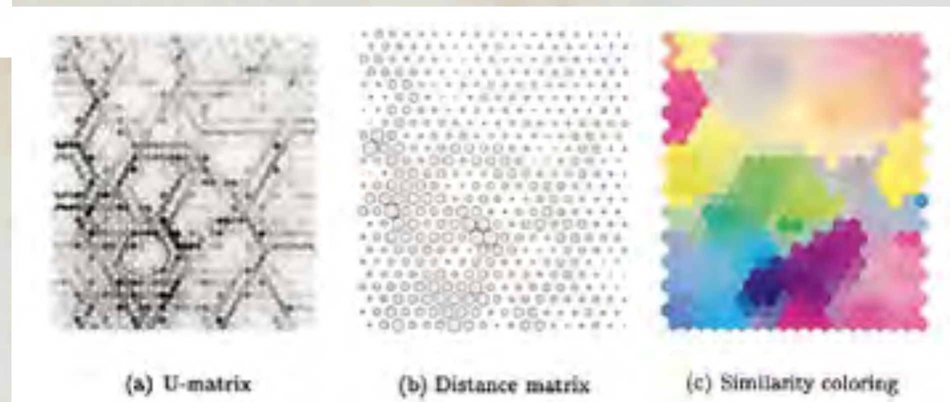
similarity in input space

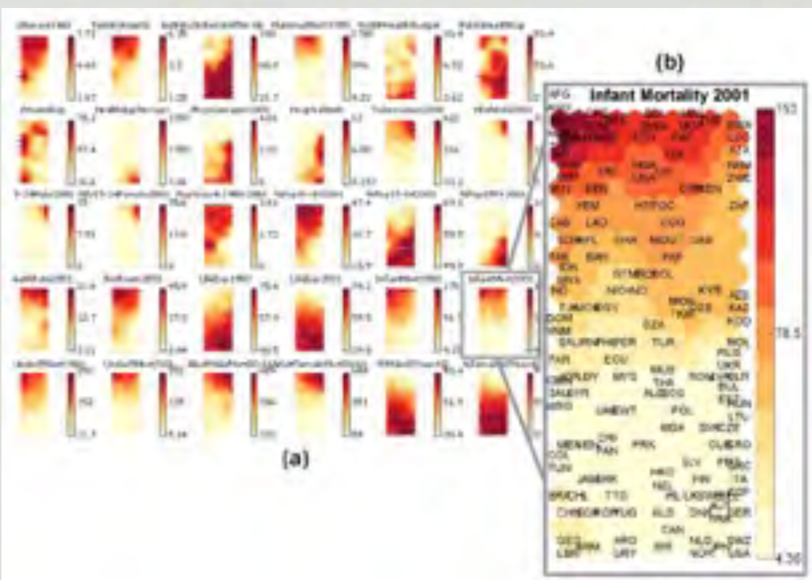
grey shade

neuron size



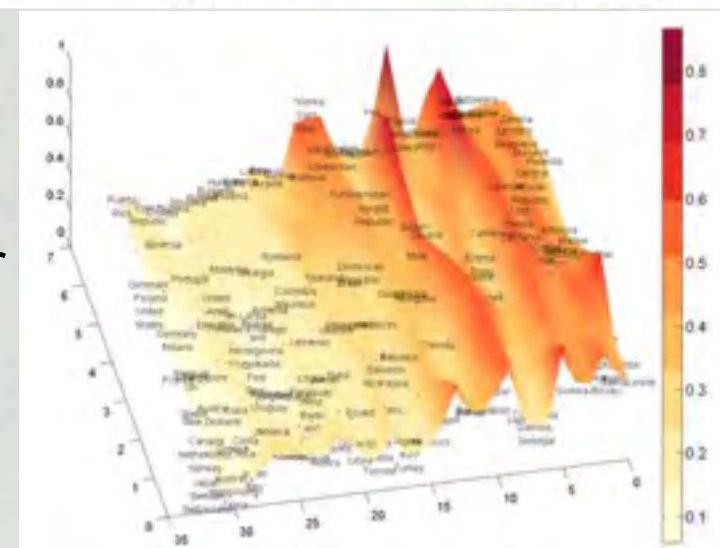
2D/3D projections of data elements with colour coding from SOM





Component planes – lattice from SOM and colours from different attributes

SOM lattice draped over **3D surface**



Exploratory Data Analysis - Interaction types

Interaction types

Projection:
- from multi-dim data to the 2D of the visualisation.

Filtering:
- select the data by using a filter or a query.

Zooming:
- get a closer/further away view of the data.

Distortion:
- transform the original data in order to display it in a better way.

Brushing and linking:
- interactive selection and linking.

Exploratory Data Analysis

Roles of visualisation in data mining

Visualising results of automatic data mining algorithms

Visual data mining

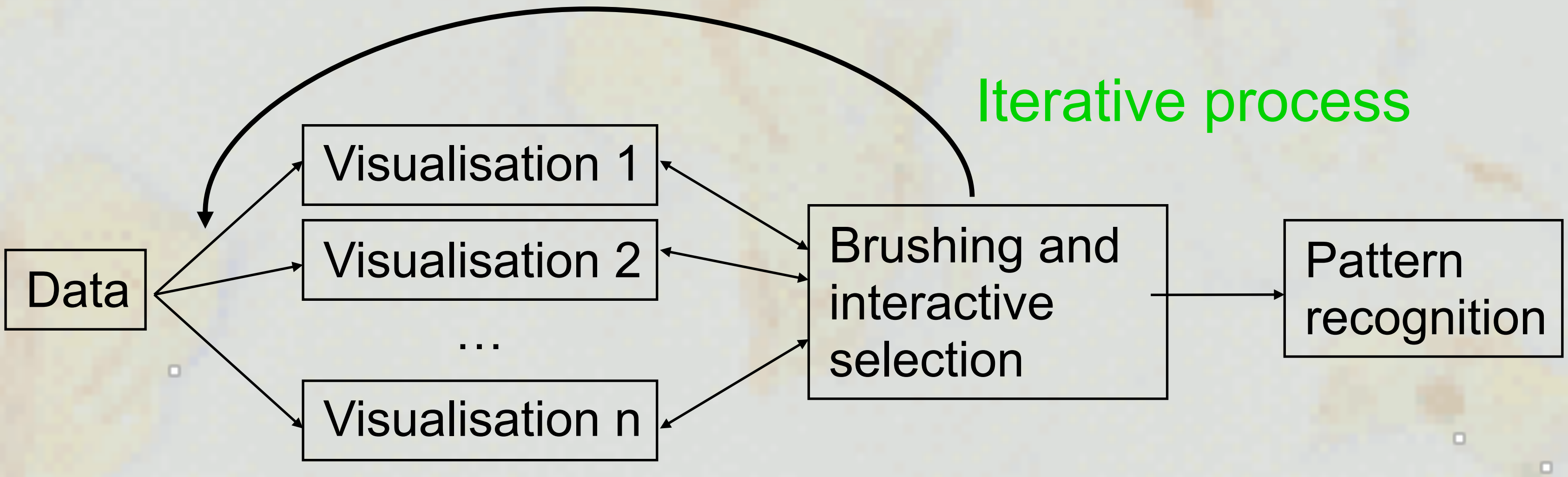
Combining automatic and visual data mining

Visual data mining:

a step in the knowledge discovery process that uses visualisation as a communication channel between the user and the computer in order to facilitate the knowledge discovery process.

Exploratory Data Analysis

Visual data mining process



Exploratory Data Analysis

Advantages over automatic mining

Interaction with the user, higher confidence

Faster and more effective exploration

Heterogeneous and noisy data

Effective when little is known about the data

Mainly used in **explorative analysis**:

Data → explorative analysis → Hypotheses about the data → confirmative analysis → Confirmed or rejected hypotheses

VDM

Exploratory Data Analysis

Combining Automatic and Visual data mining

Automatic data mining:

- can deal with large amounts of data
- user involvement minimal
- user needs to be familiar with complicated mathematics
- the domain knowledge of the user is not included in the exploration

Visual data mining:

- user interaction makes it easier to understand what is going on
- difficult to present large amounts of data
- difficult to include the multidimensionality of data
- human vision is too good a tool for pattern recognition: we may see patterns where there are none

Integration of visual and computational data mining methods

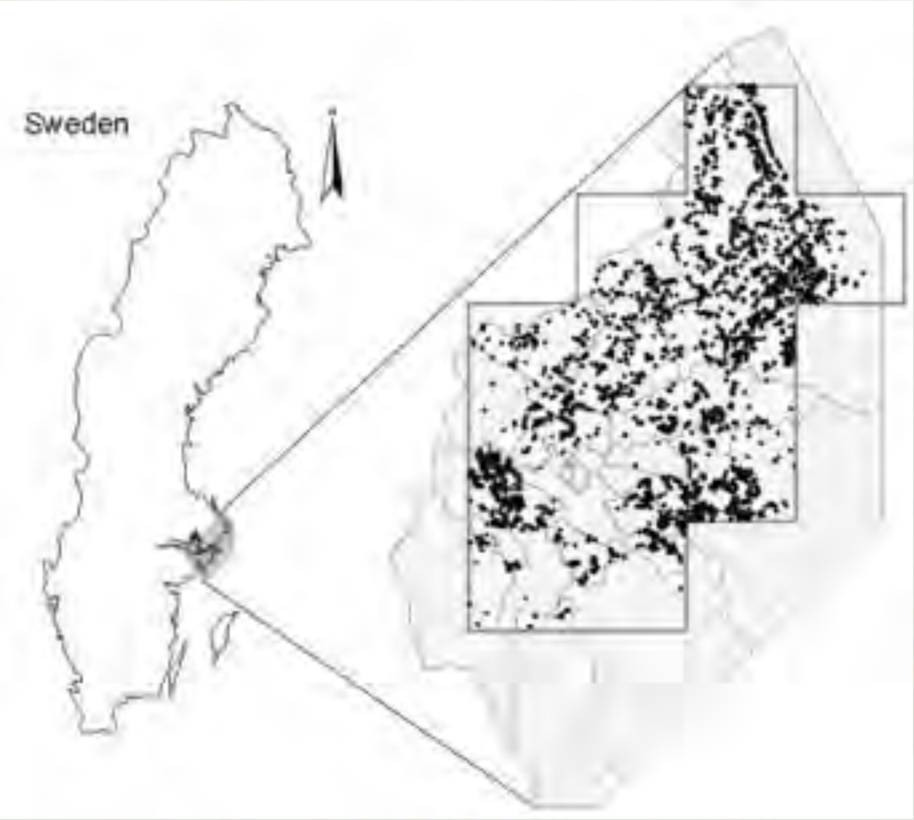
Ordinary visualisations	+	Visualisation(s) of the result of an automatic DM algorithm

Example - Drilled well water resources in Stockholm

Visual and automatic data mining for environmental data

Dataset with 4435 drilled wells from Stockholm county and 13 attributes:

- radon concentration in water
- elevation
- slope
- relative elevation
- soil (original)
- soil (reclassified)
- bedrock (original)
- bedrock (reclassified)
- distance to nearest fracture
- type of nearest fracture
- length of nearest fracture
- uranium concetration
- log Rn



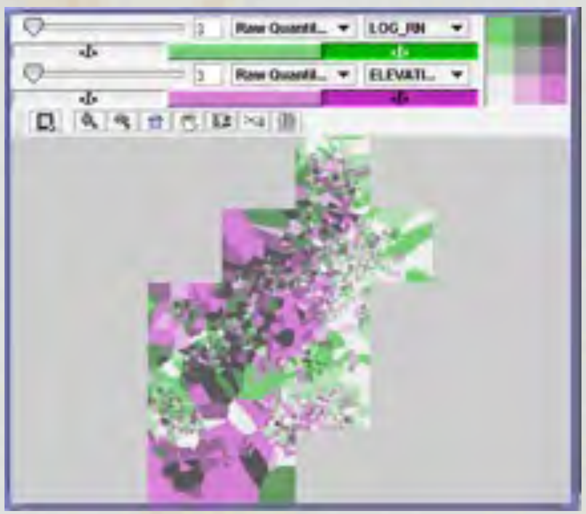
Exploration goal: find relationships between radon concentration and other attributes and look for global structure in the dataset.

Example - Drilled well water resources in Stockholm

Visualisations

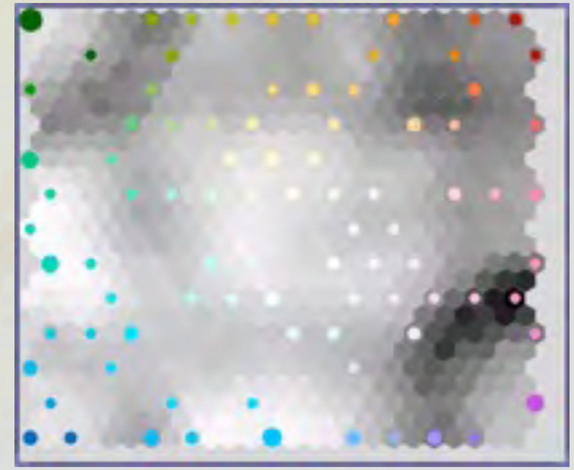
Table

geoMap

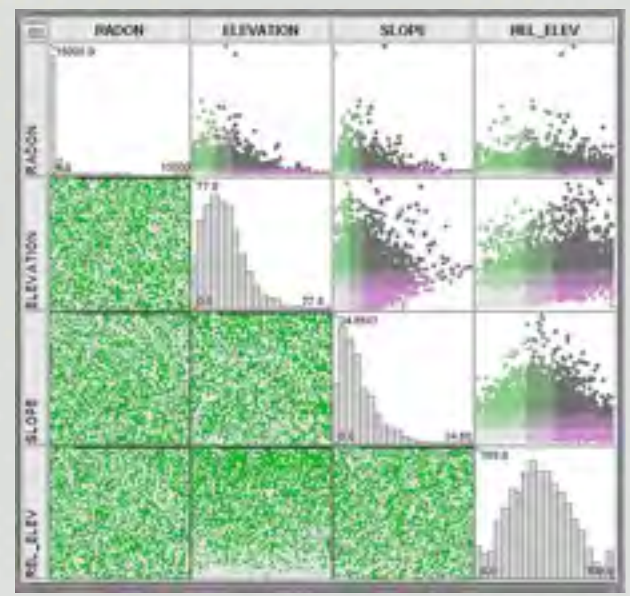


Brushing & linking + interactive selection

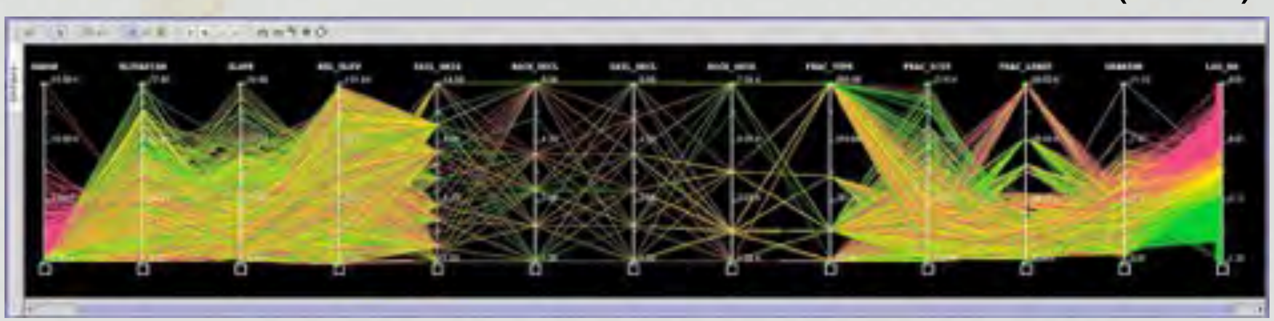
SOM visualisation



Bivariate multiplot matrix

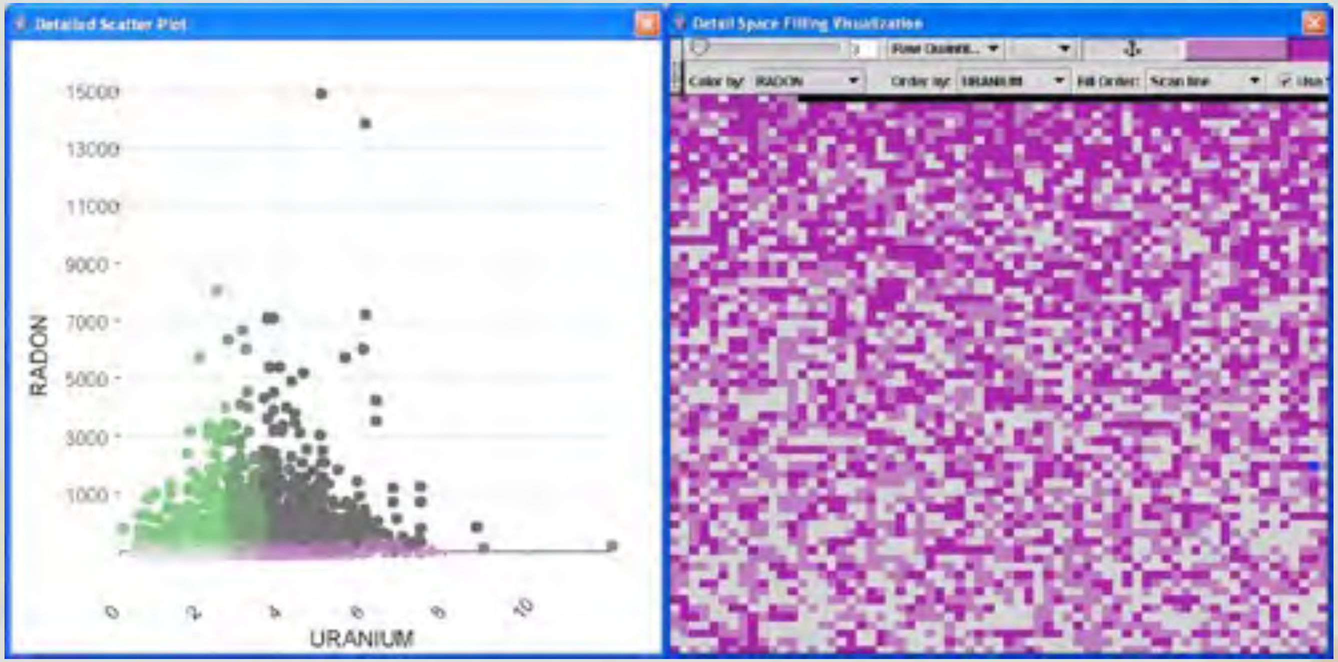


Parallel Coordinates Plot (PCP)



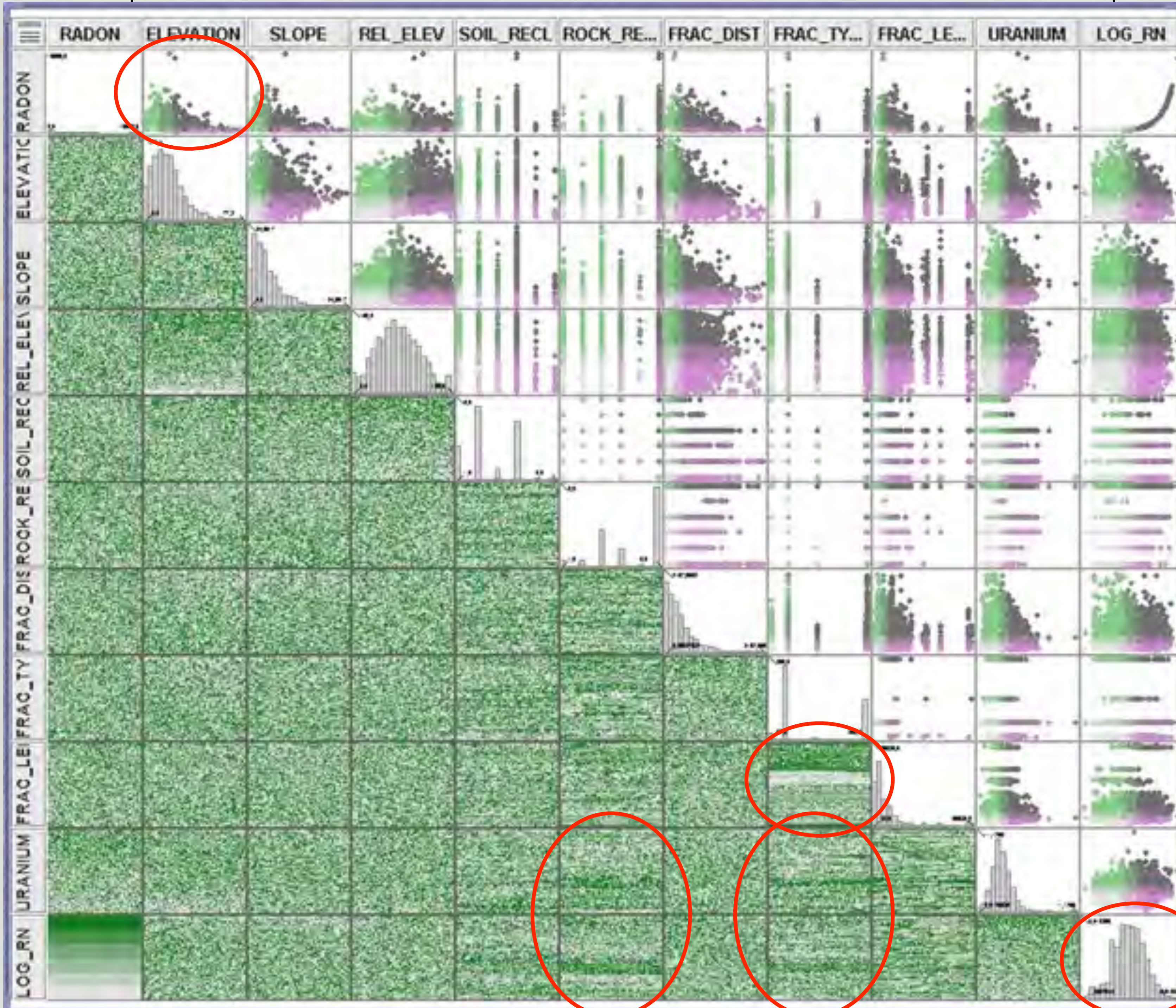
Example - Drilled well water resources in Stockholm

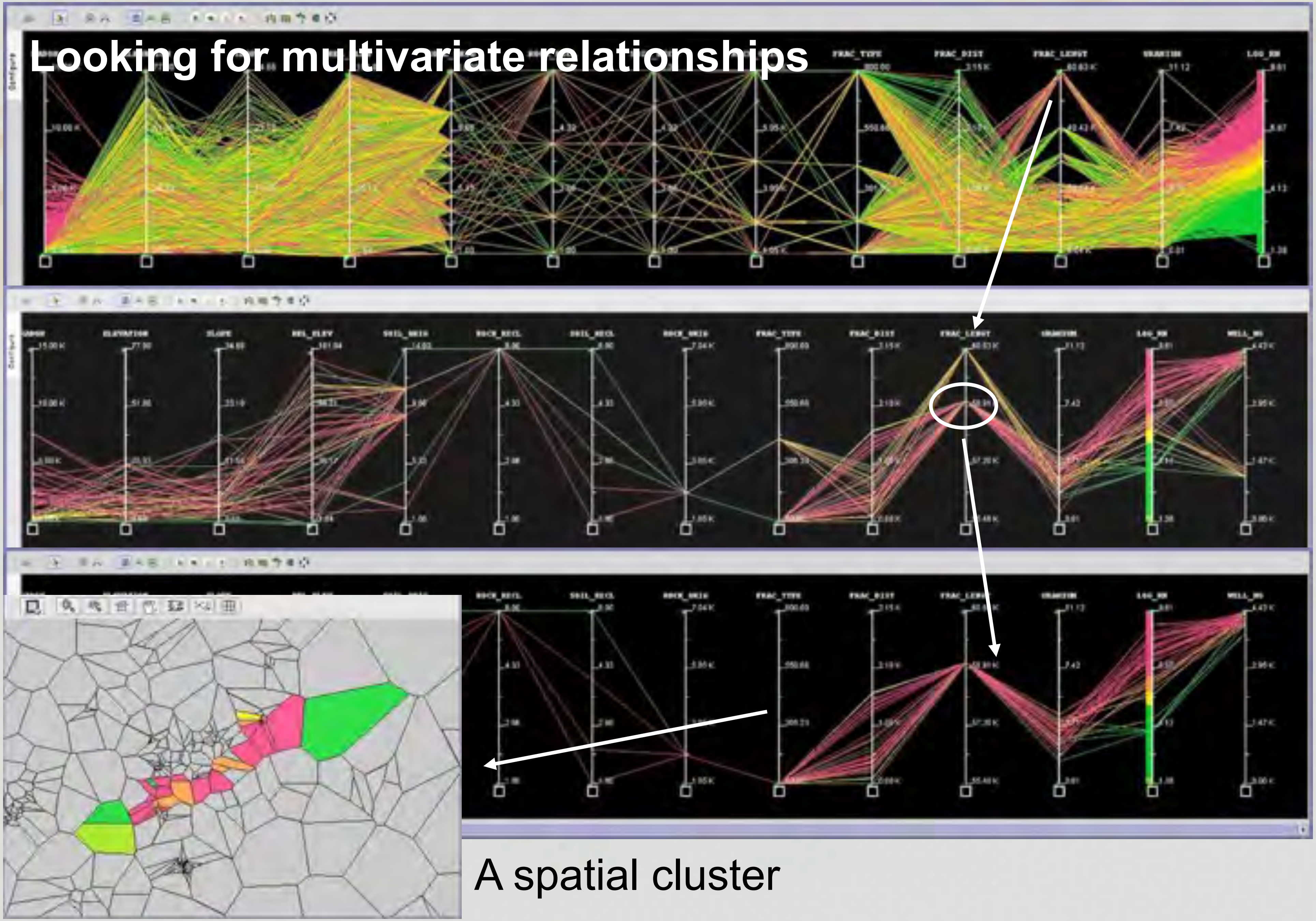
Looking for bivariate relationships



Scatterplot SpaceFill visualisation

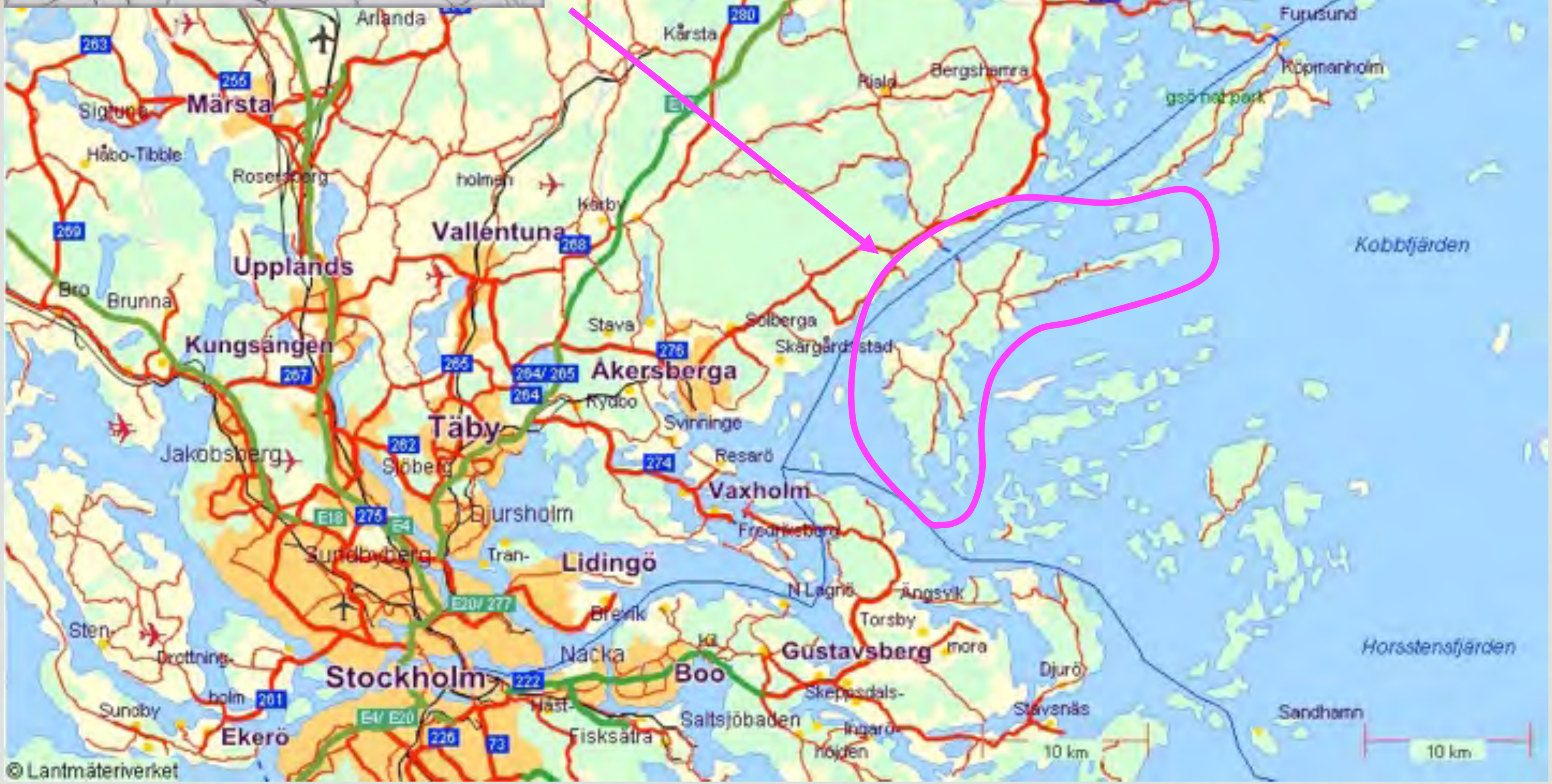
Used separately and in the multiform bivariate matrix


RADON
ELEVATION
SLOPE
REL_ELEV
SOIL_RECL
ROCK_RE...
FRAC_DIST
FRAC_TY...
FRAC_LE...
URANIUM
LOG_RN



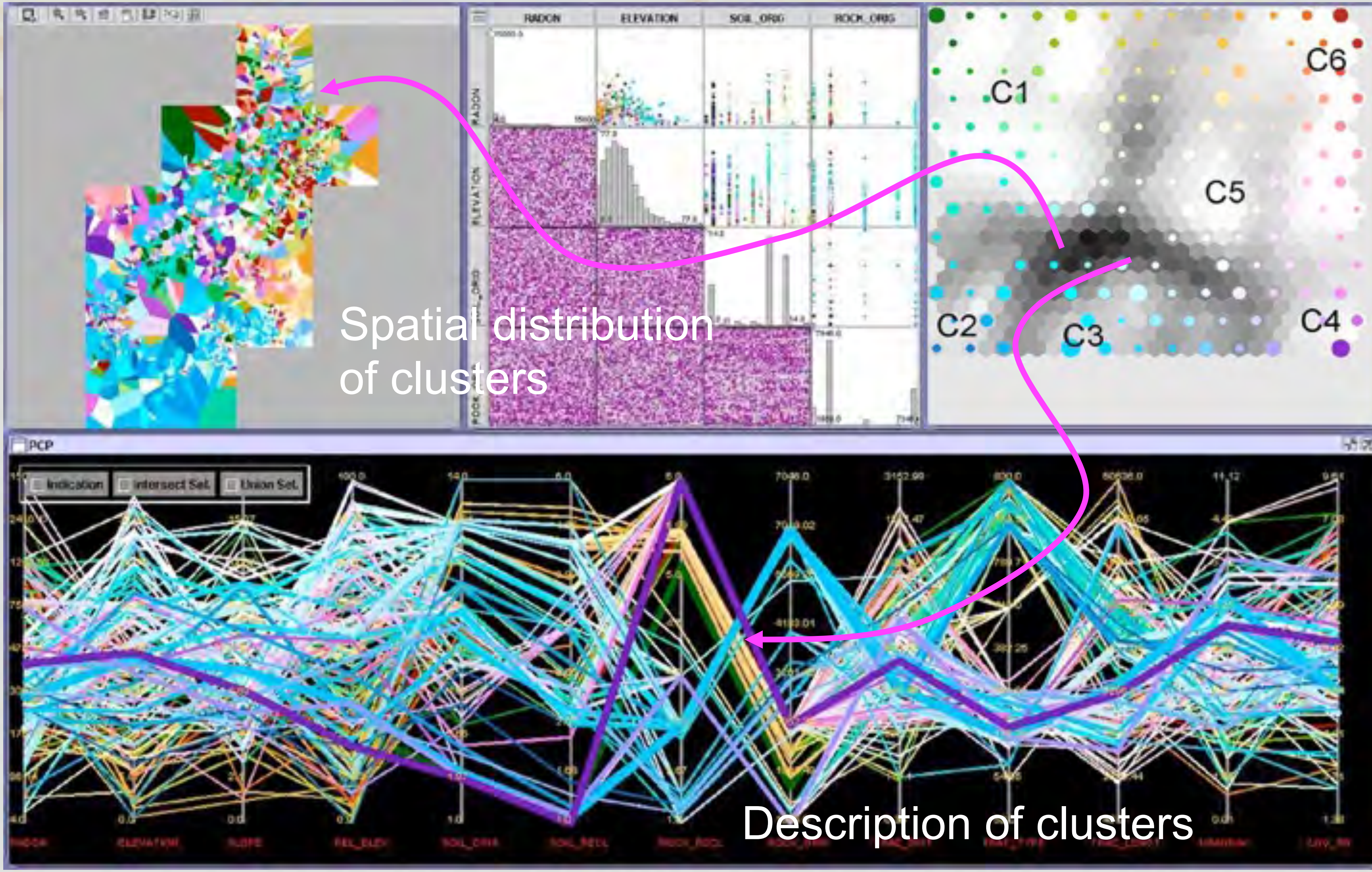
A spatial cluster

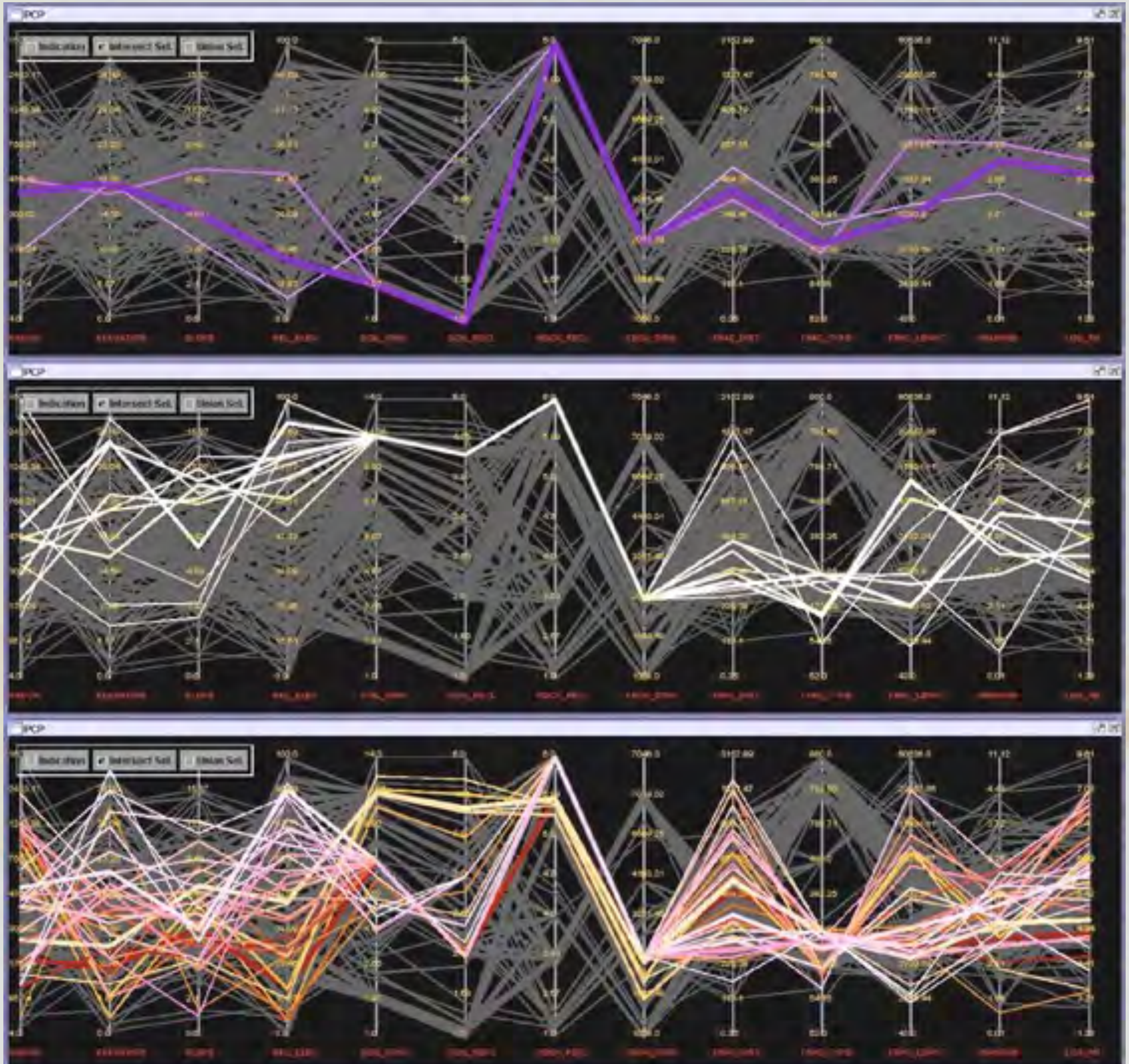
Comparison with topographical map



Global structure in the data

Clusters in the data





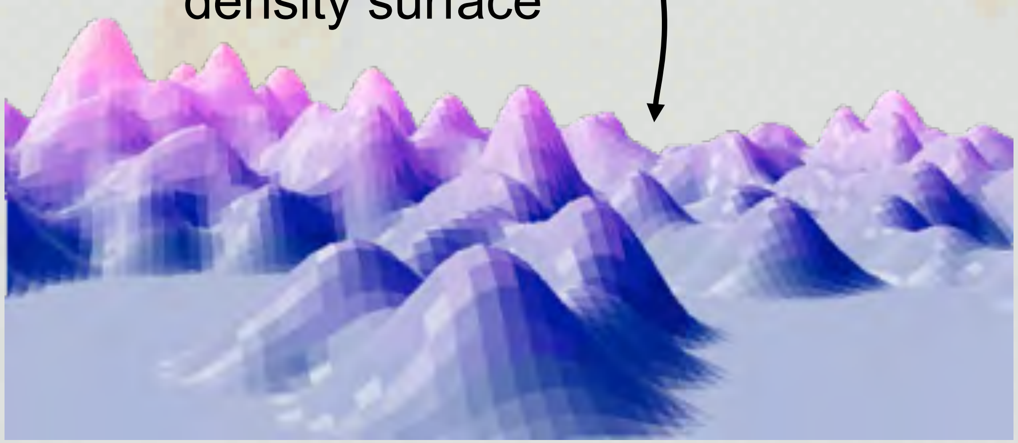
Spatio-visual data mining for fire&rescue incidents data

Spatio-visual DM = Spatial pre-processing + Visual data mining

Density analysis

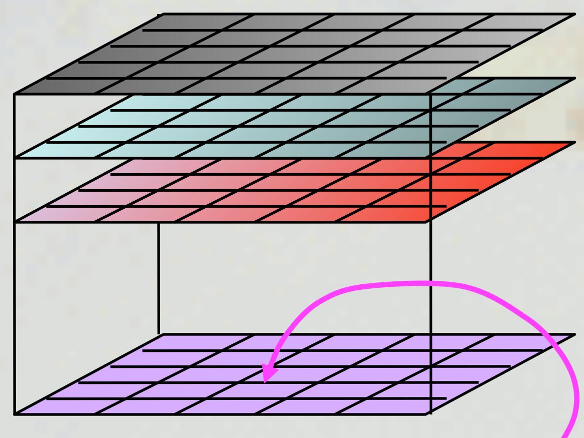
kernels

density surface



Vertical view approach

virtual overlay of densities and proximity surfaces



Multidimensional spatial database

Dataset with 3136 polygons (raster cells) with 51 attributes

36 temporal attributes

kernel densities of incidents according to

time

day

night

type

6 types

year

2001-2003

15 non-temporal attributes

3 aggregated incident densities

0/1 proximity surfaces

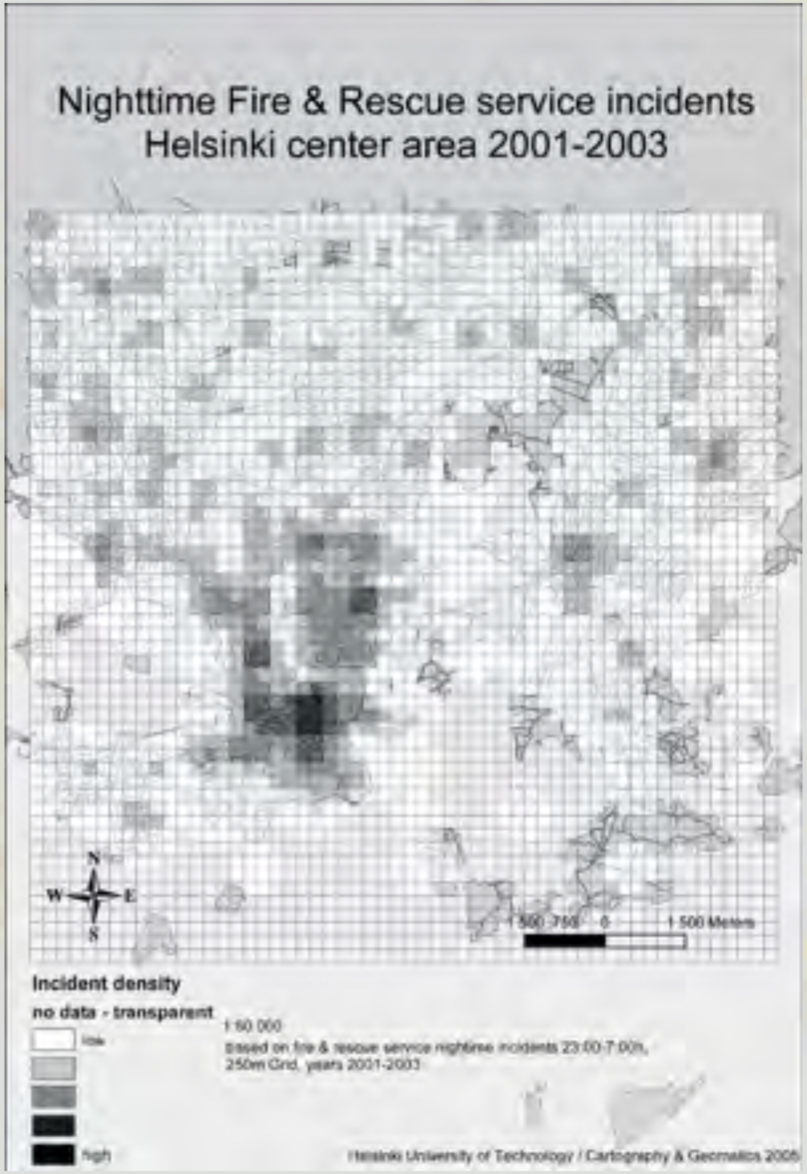
water areas

built-up areas

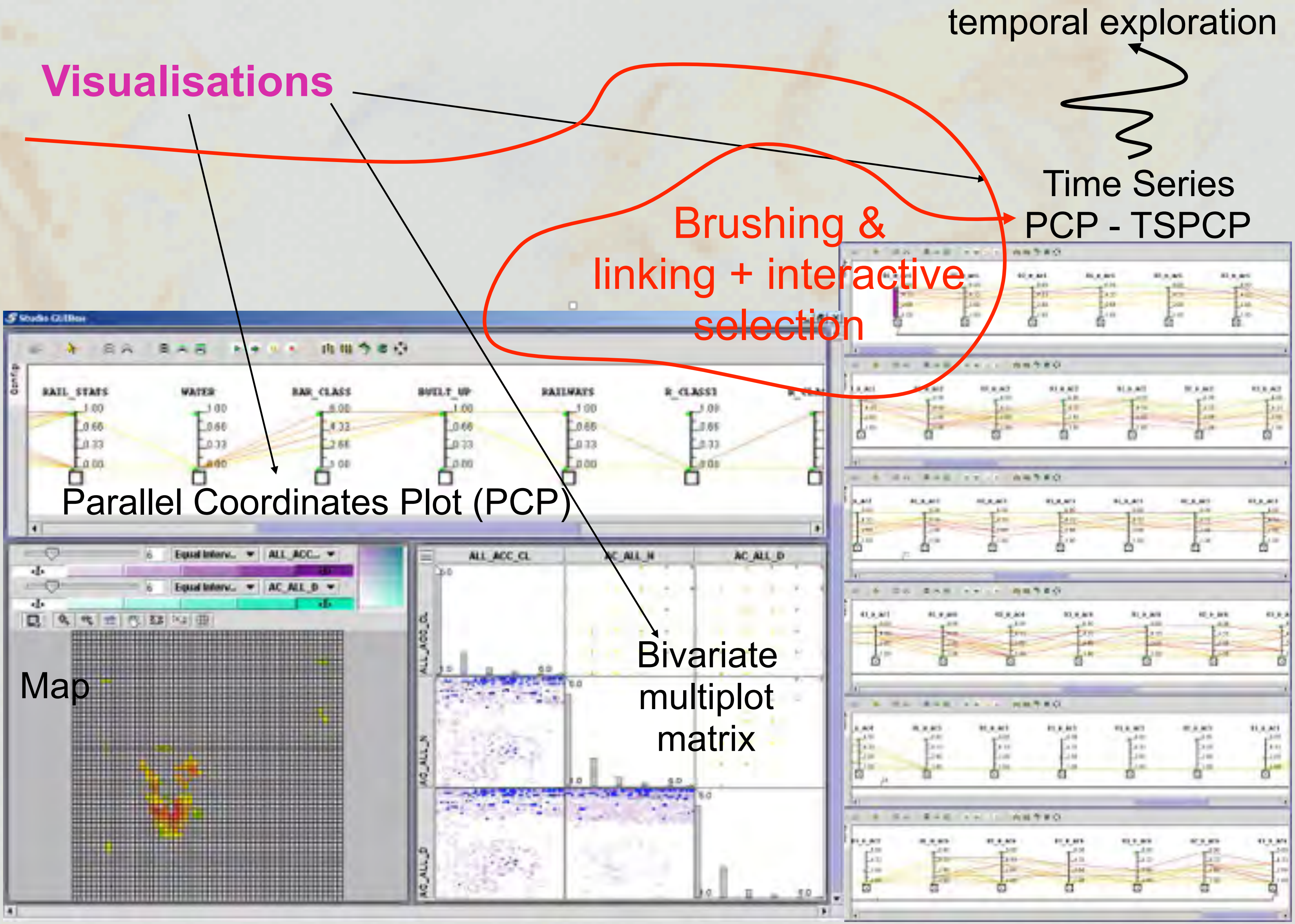
railway network

railway stations

6 road types



Exploration goal: discover connections between incidents' locations and other attributes


temporal exploration
Visualisations
Time Series
PCP - TSPCP
Brushing &
linking + interactive
selection
Parallel Coordinates Plot (PCP)
Map
Bivariate
multiplot
matrix

Visually estimating the strength of the bivariate relationships

Relatively smooth transition from white to black along the scanline

↓

stronger relationship

Not so smooth transition

↓

weaker relationship

Visual data mining – an example of application

Task:

find data instances that are

- in Italian,

- deal with site planning and

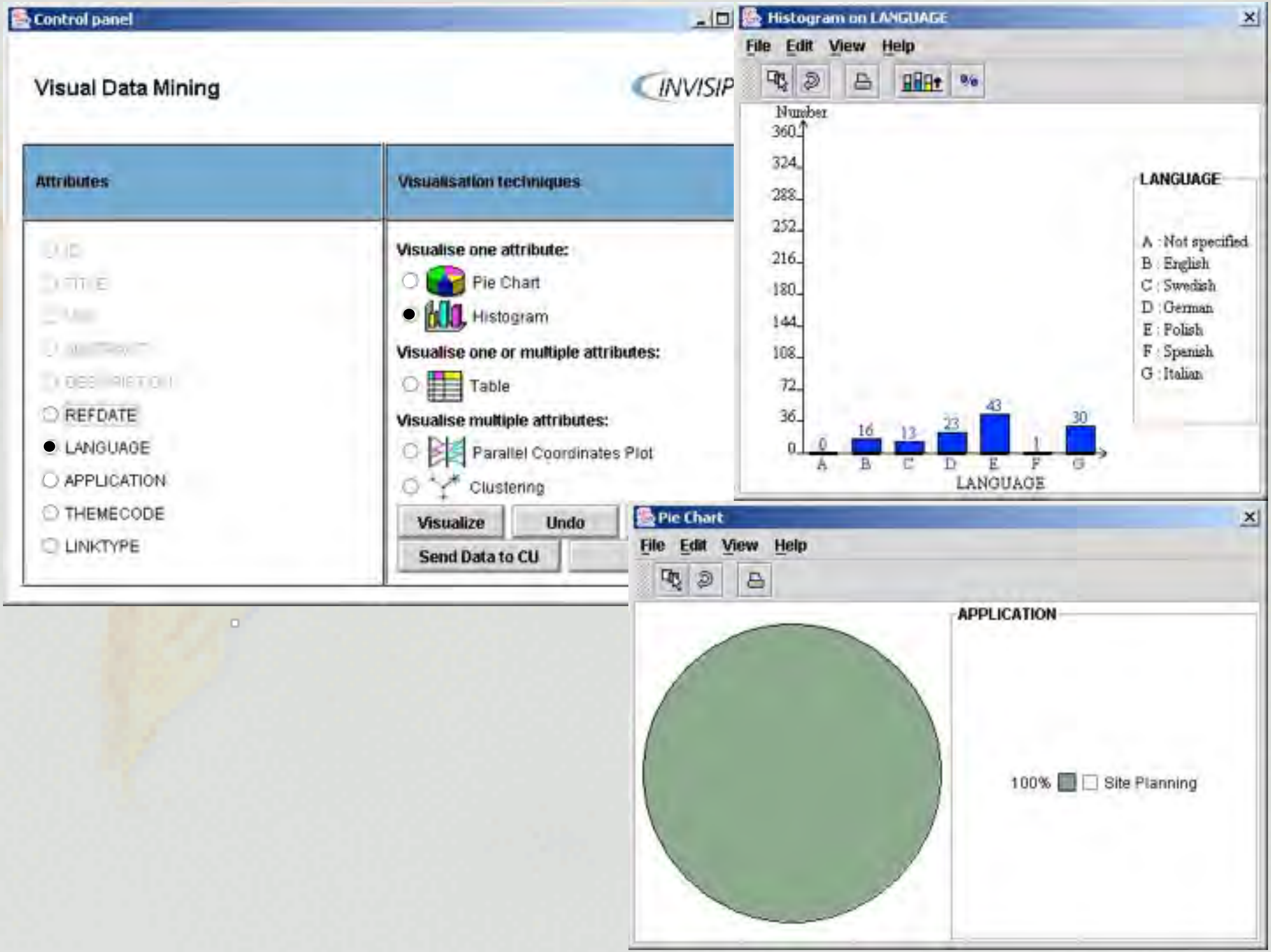
- are about society.

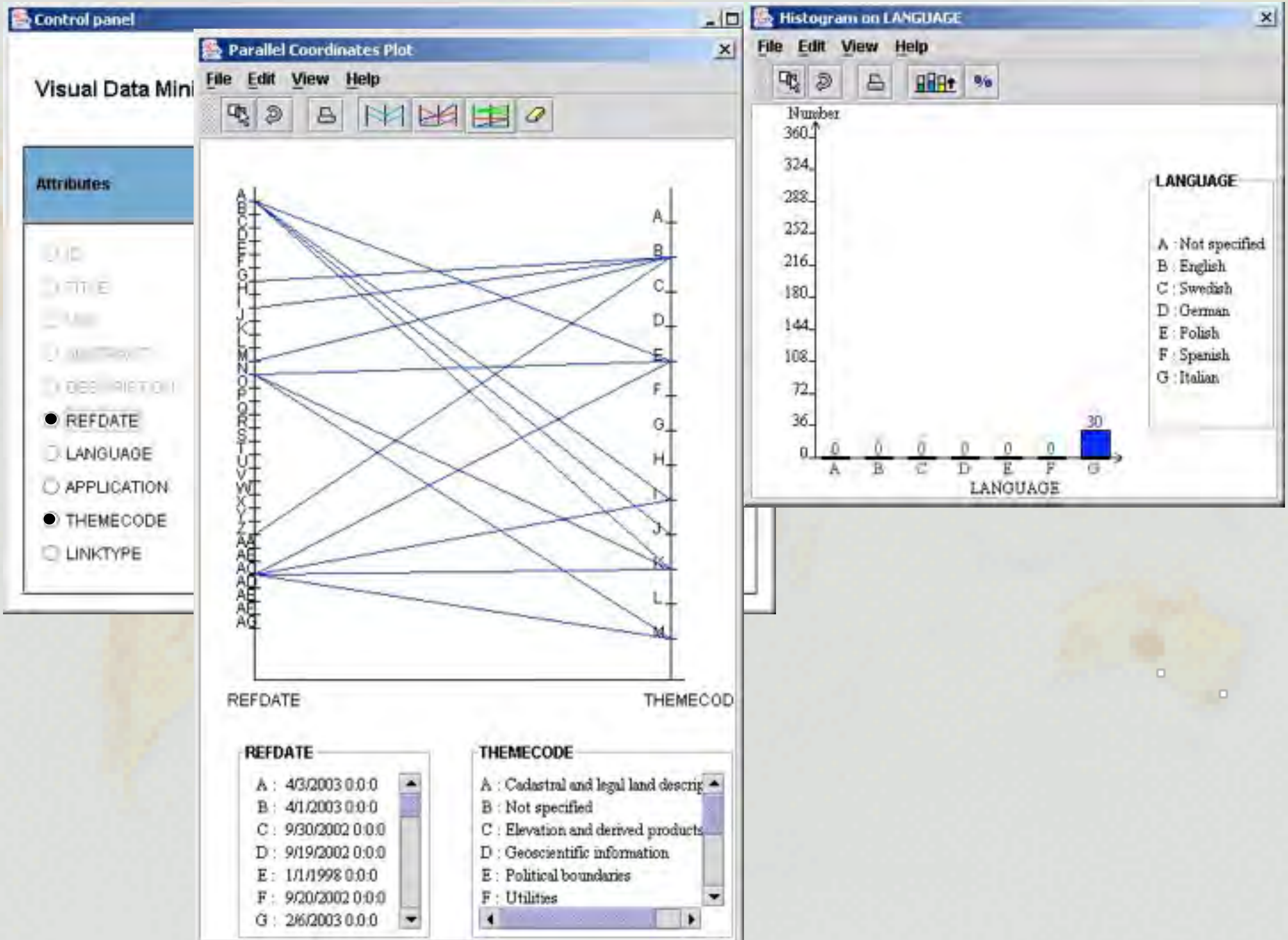
LANGUAGE

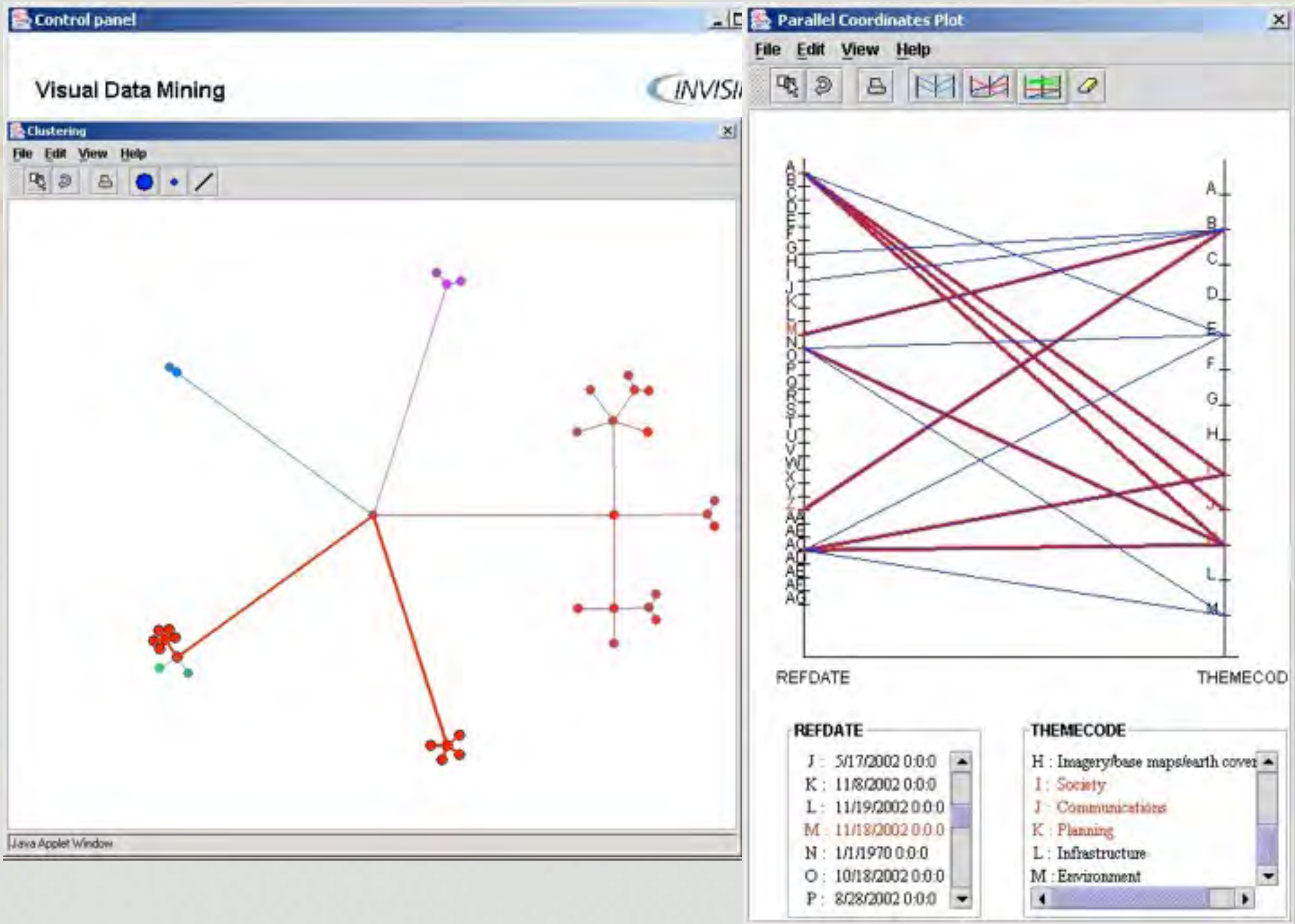
APPLICATION

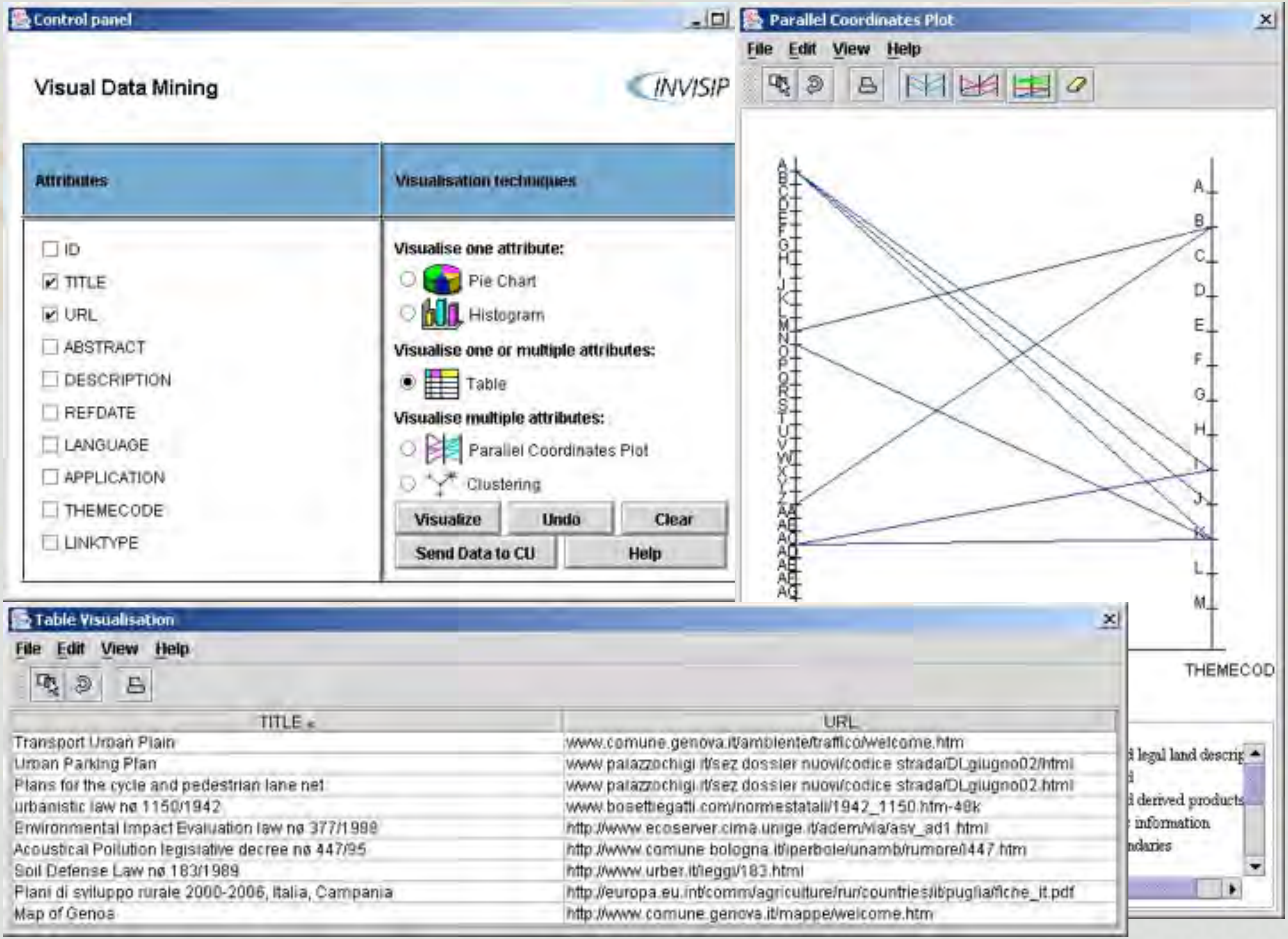
THEMECODE

Attributes


Control panel
Visual Data Mining
INVISIP
Attributes
REFDATE
LANGUAGE
APPLICATION
THEMECODE
LINKTYPE
Visualisation techniques
Visualise one attribute:
Pie Chart
Histogram
Visualise one or multiple attributes:
Table
Visualise multiple attributes:
Parallel Coordinates Plot
Clustering
Visualize
Undo
Send Data to CU
Histogram on LANGUAGE
File Edit View Help
Number
360
324
288
252
216
180
144
108
72
36
0
0 16 13 23 43 1 30
A B C D E F G
LANGUAGE
LANGUAGE
A : Not specified
B : English
C : Swedish
D : German
E : Polish
F : Spanish
G : Italian
Pie Chart
File Edit View Help
APPLICATION
100% Site Planning

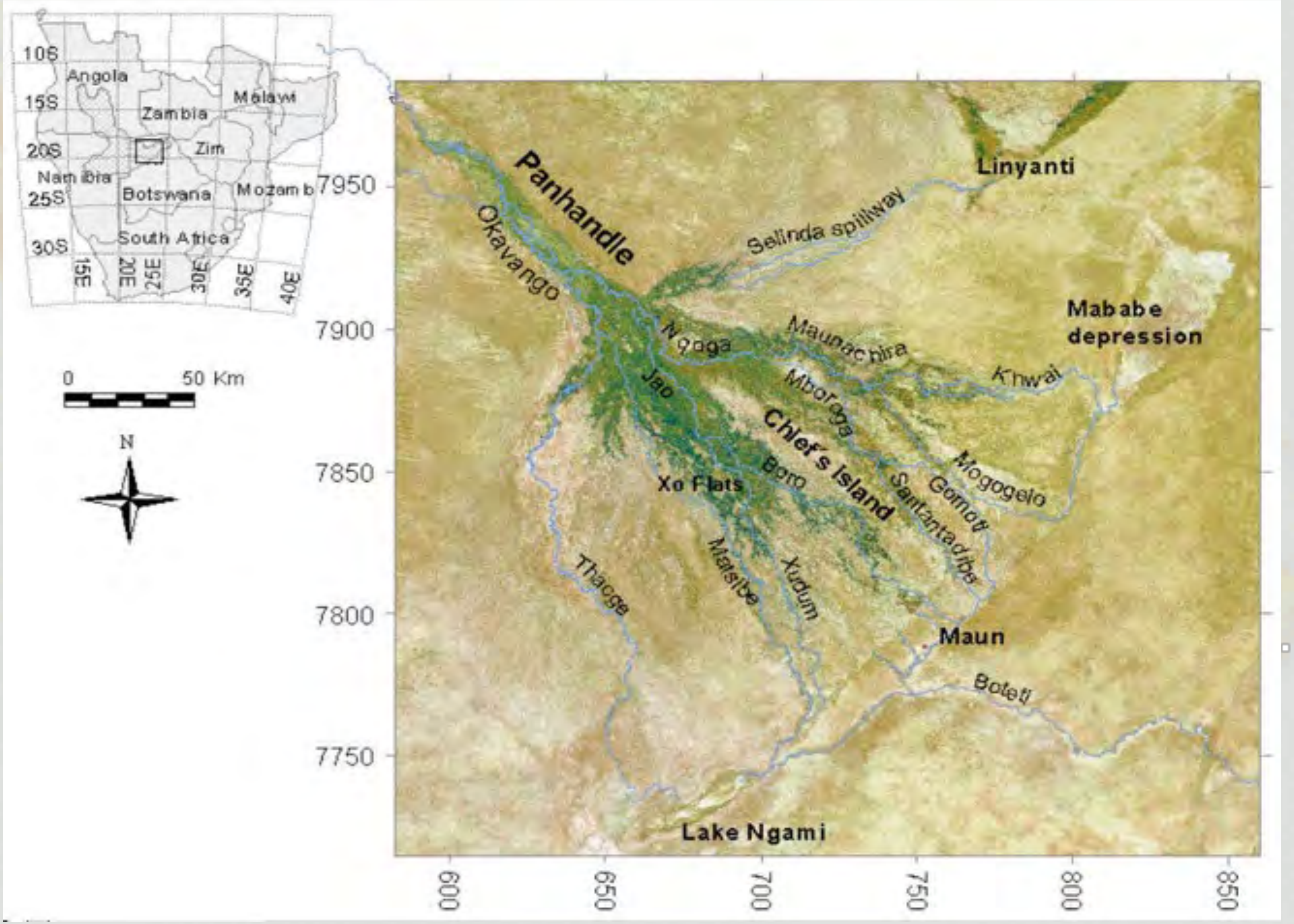

Control panel
Visual Data Mini
Attributes
REFDATE
LANGUAGE
APPLICATION
THEMECODE
LINKTYPE
Parallel Coordinates Plot
File Edit View Help
REFDATE
THEMECOD
REFDATE
A : 4/3/2003 0:0:0
B : 4/1/2003 0:0:0
C : 9/30/2002 0:0:0
D : 9/19/2002 0:0:0
E : 1/1/1998 0:0:0
F : 9/20/2002 0:0:0
G : 2/6/2003 0:0:0
THEMECODE
A : Cadastral and legal land descrip
B : Not specified
C : Elevation and derived products
D : Geoscientific information
E : Political boundaries
F : Utilities
Histogram on LANGUAGE
File Edit View Help
Number
LANGUAGE
A : Not specified
B : English
C : Swedish
D : German
E : Polish
F : Spanish
G : Italian
30
LANGUAGE


Control panel
Visual Data Mining
INVISI
Clustering
File Edit View Help
Java Applet Window
Parallel Coordinates Plot
File Edit View Help
REFDATE
THEMECOD
REFDATE
J : 5/17/2002 0:0:0
K : 11/8/2002 0:0:0
L : 11/19/2002 0:0:0
M : 11/18/2002 0:0:0
N : 1/1/1970 0:0:0
O : 10/18/2002 0:0:0
P : 8/28/2002 0:0:0
THEMECODE
H : Imagery/base maps/earth cover
I : Society
J : Communications
K : Planning
L : Infrastructure
M : Environment


Control panel
Visual Data Mining
INVISIP
Attributes
Visualisation techniques
ID
TITLE
URL
ABSTRACT
DESCRIPTION
REFDATE
LANGUAGE
APPLICATION
THEMECODE
LINKTYPE
Visualise one attribute:
Pie Chart
Histogram
Visualise one or multiple attributes:
Table
Visualise multiple attributes:
Parallel Coordinates Plot
Clustering
Visualize
Undo
Clear
Send Data to CU
Help
Parallel Coordinates Plot
File Edit View Help
THEMECOD
Table Visualisation
File Edit View Help
TITLE
URL
Transport Urban Plain
www.comune.genova.it/ambiente/traffico/welcome.htm
Urban Parking Plan
www.palazzochigi.it/sez dossier nuovi/codice strada/DLgiugno02/html
Plans for the cycle and pedestrian lane net
www.palazzochigi.it/sez dossier nuovi/codice strada/DLgiugno02.html
urbanistic law nø 1150/1942
www.bosettiegatti.com/normestatali/1942_1150.htm-48k
Environmental Impact Evaluation law nø 377/1988
http://www.ecoserver.cima.unige.it/adem/via/asv_ad1.html
Acoustical Pollution legislative decree nø 447/95
http://www.comune.bologna.it/iperbole/unamb/rumore/447.htm
Soil Defense Law nø 183/1989
http://www.urber.it/leggi/183.html
Piani di sviluppo rurale 2000-2006, Italia, Campania
http://europa.eu.int/comm/agriculture/rur/countries/it/puglia/fiche_it.pdf
Map of Genoa
http://www.comune.genova.it/mappe/welcome.htm

Example - orientation of islands in the Okavango delta, Botswana

Example - orientation of islands in the Okavango Delta, Botswana



Example - orientation of islands in the Okavango Delta, Botswana

Primary islands built from accumulation of clastic sediments

Island types

Inverted channel island

Example - orientation of islands in the Okavango Delta, Botswana

Primary islands built from accumulation of clastic sediments

Island types

Scroll bar island

Example - orientation of islands in the Okavango Delta, Botswana

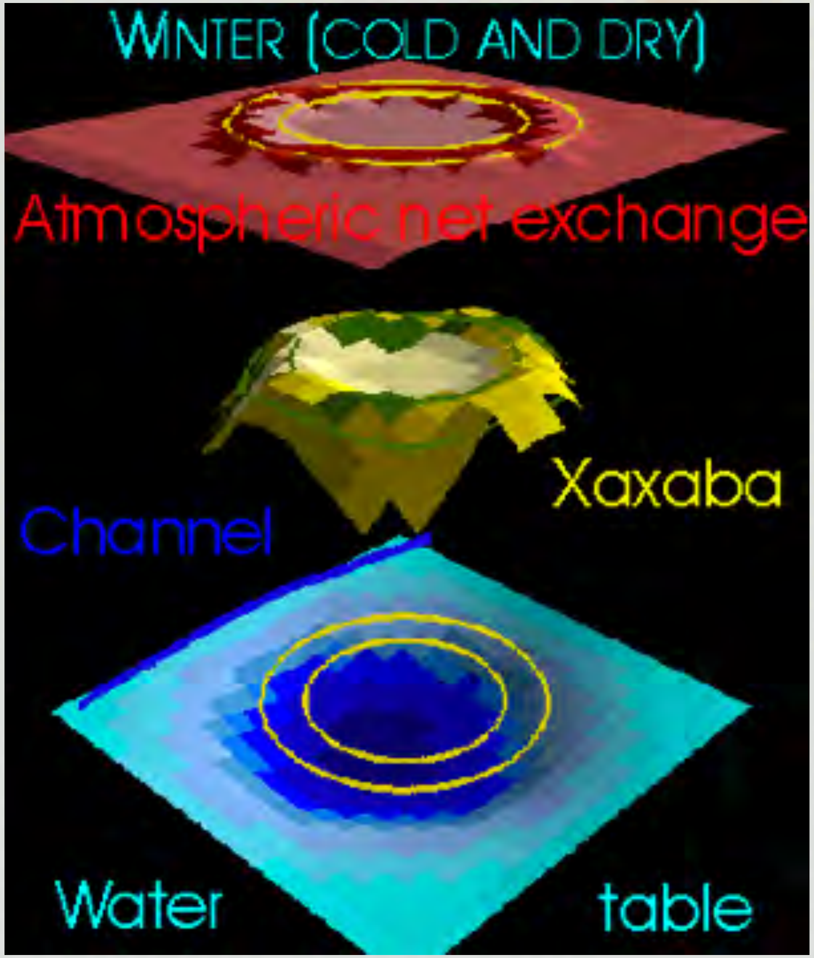
Primary islands built from accumulation of clastic sediments

Island types

Anthill island

Example - orientation of islands in the Okavango Delta, Botswana

Evapotranspiration, salinity balance and island secondary growth



Example - orientation of islands in the Okavango Delta, Botswana

Secondary islands grown from precipitation of chemical sediments

Island types

Riparian forest island

Example - orientation of islands in the Okavango Delta, Botswana

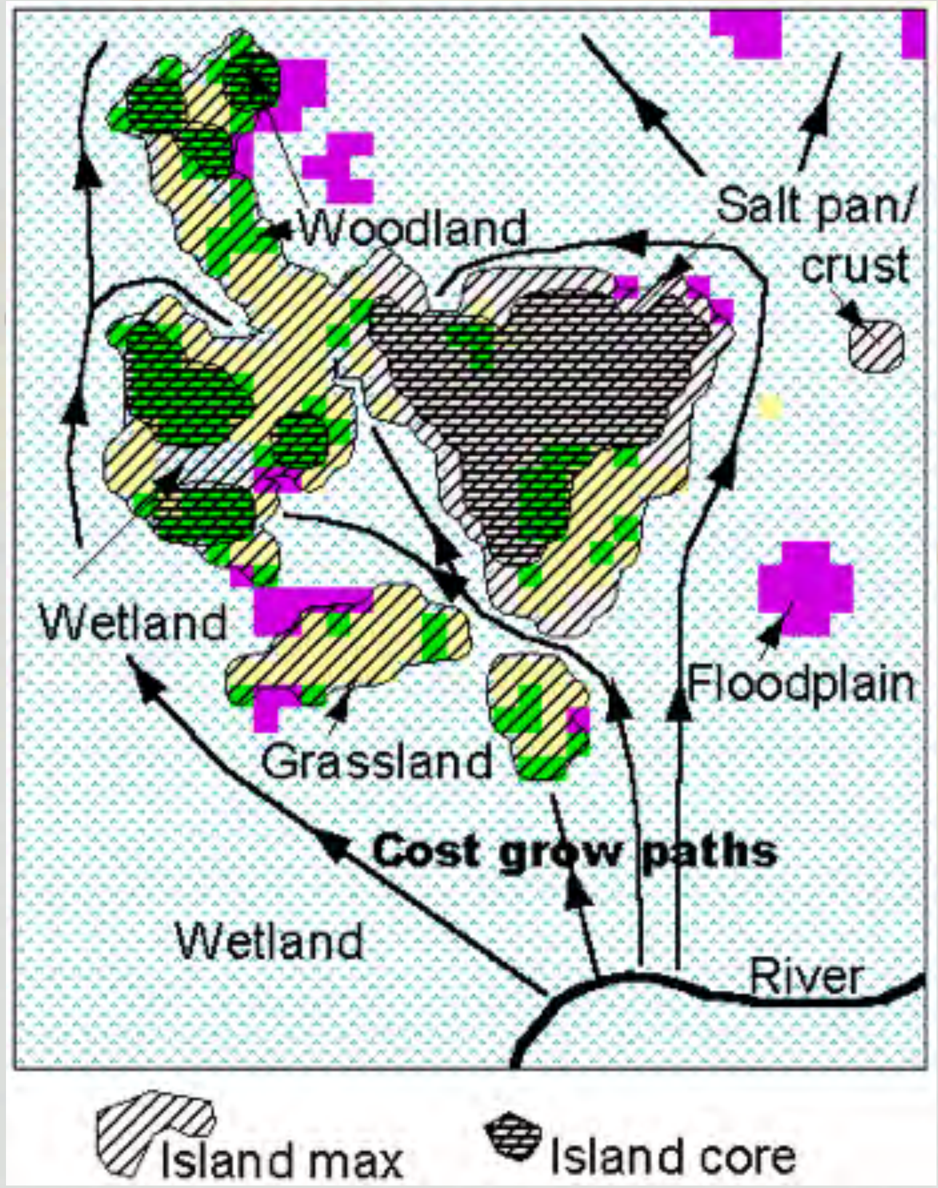
Secondary islands grown from precipitation of chemical sediments

Island types

Salt island

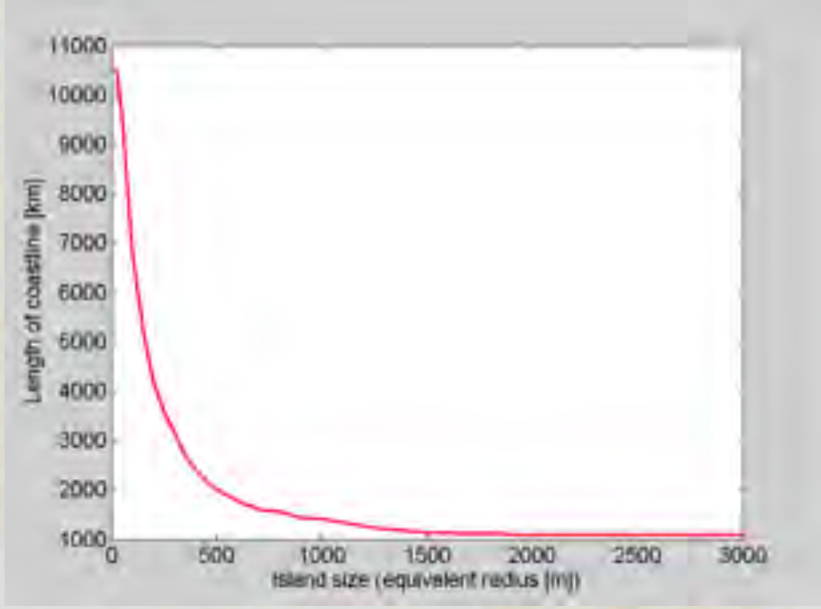
Example - orientation of islands in the Okavango Delta, Botswana

Exempel på Transformation raster till vektor



Example - orientation of islands in the Okavango Delta, Botswana

Salt Balance: Coastline from Remote Sensing

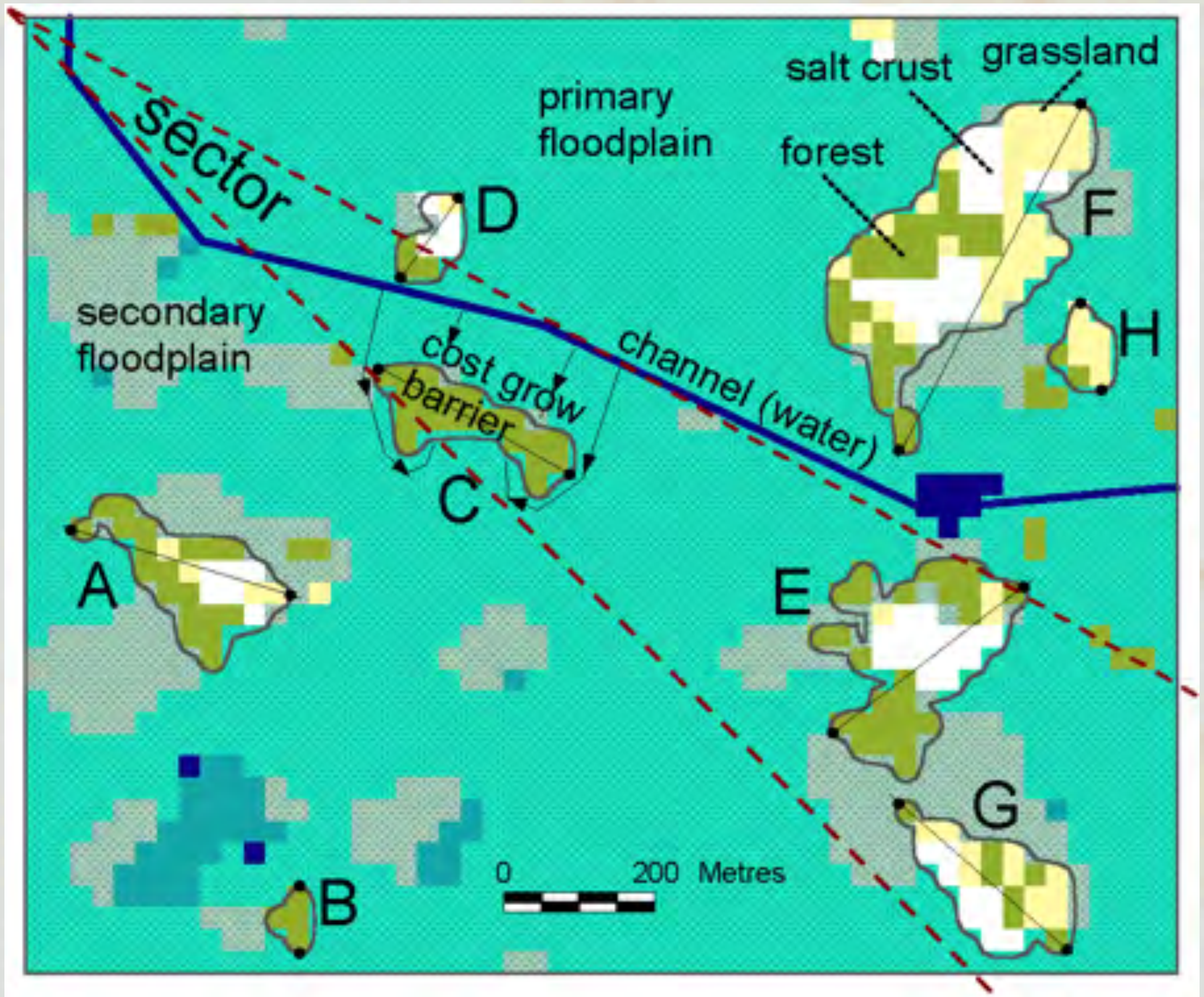


Exempel på
Hypotesprövning

Example - orientation of islands in the Okavango Delta, Botswana

Extraktion av längdaxel och beräkning av riktning

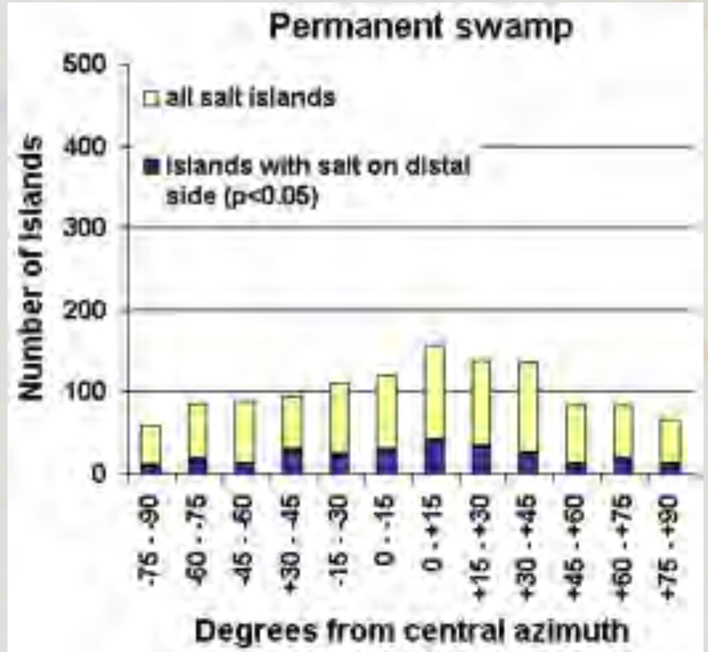
Exempel på mätning

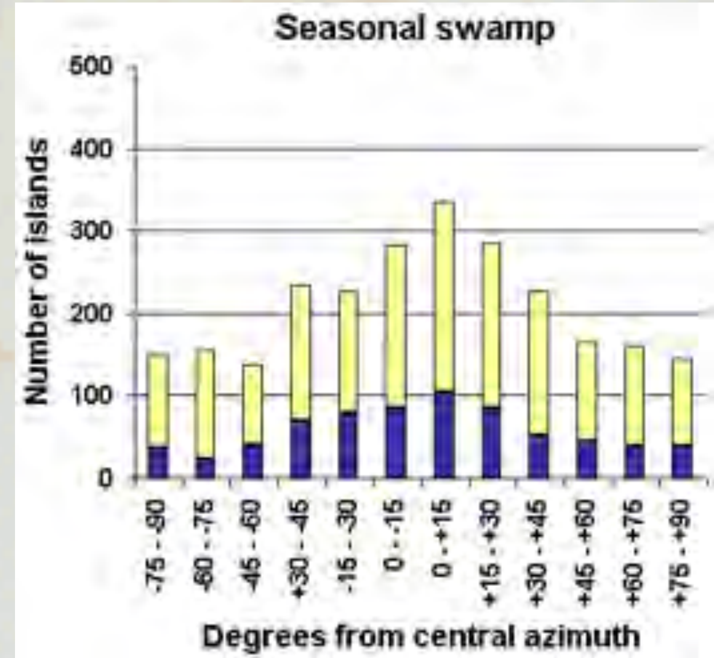


	A	B	C	D	E	F	G	H
Roundness	0.49	0.91	0.51	0.48	0.36	0.47	0.58	0.92
Regional salt position	distal*	na	na	proximal	distal	equal	proximal	na
Channel salt position	front	na	na	back	back	back	back*	na

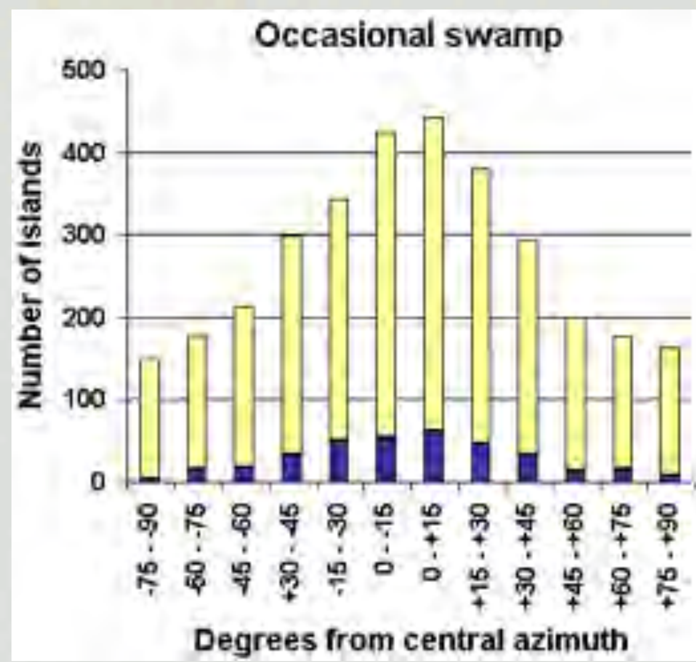
Example - orientation of islands in the Okavango Delta, Botswana

Öarnas längdriktning i relation till Deltats riktning



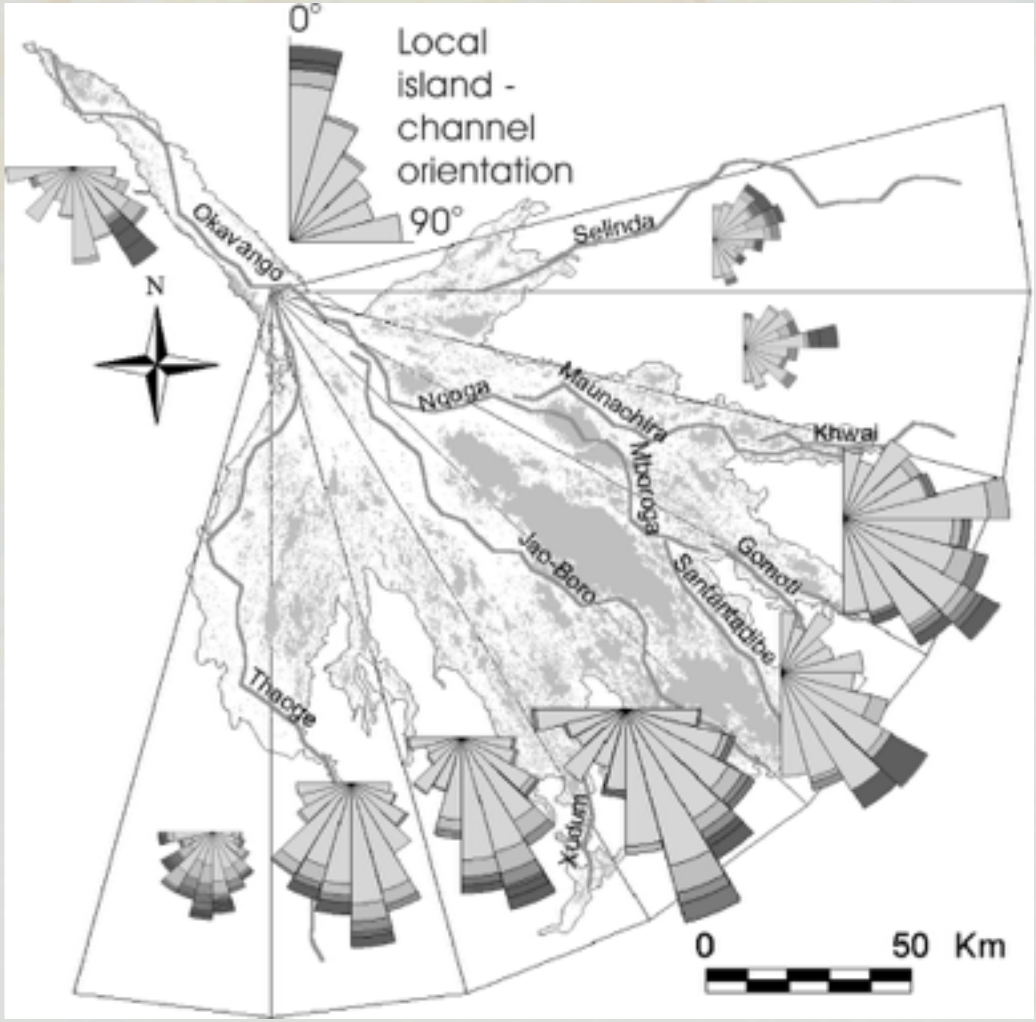


Exempel på deskriptiv metod



Example - orientation of islands in the Okavango Delta, Botswana

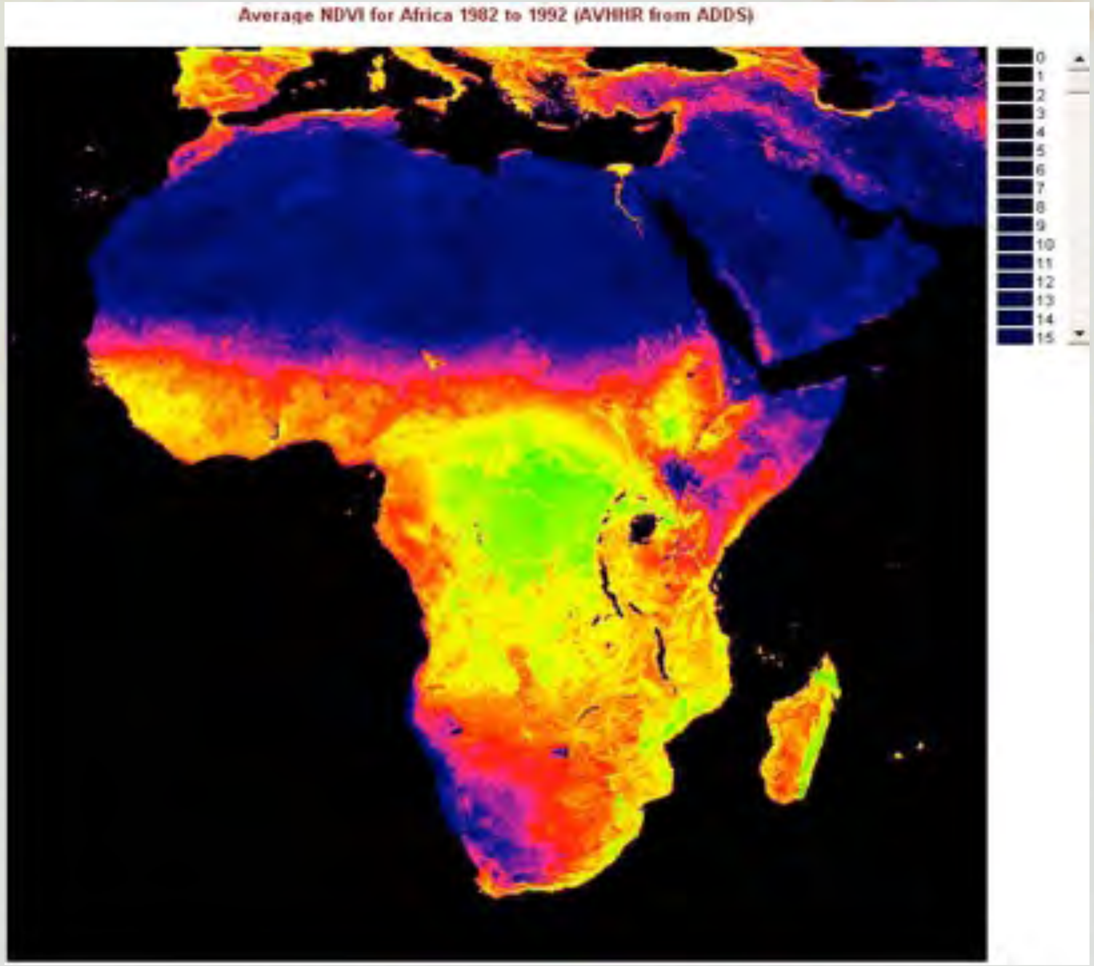
Öarnas betydelse för uppdelningen av vattenföring och indelning i bassänger



Exempel på deskriptiv metod

Example - mapping african vegetation using PCA

Data sources
- ➢NOAA AVHRR dekadal data (10day), 1982-2004



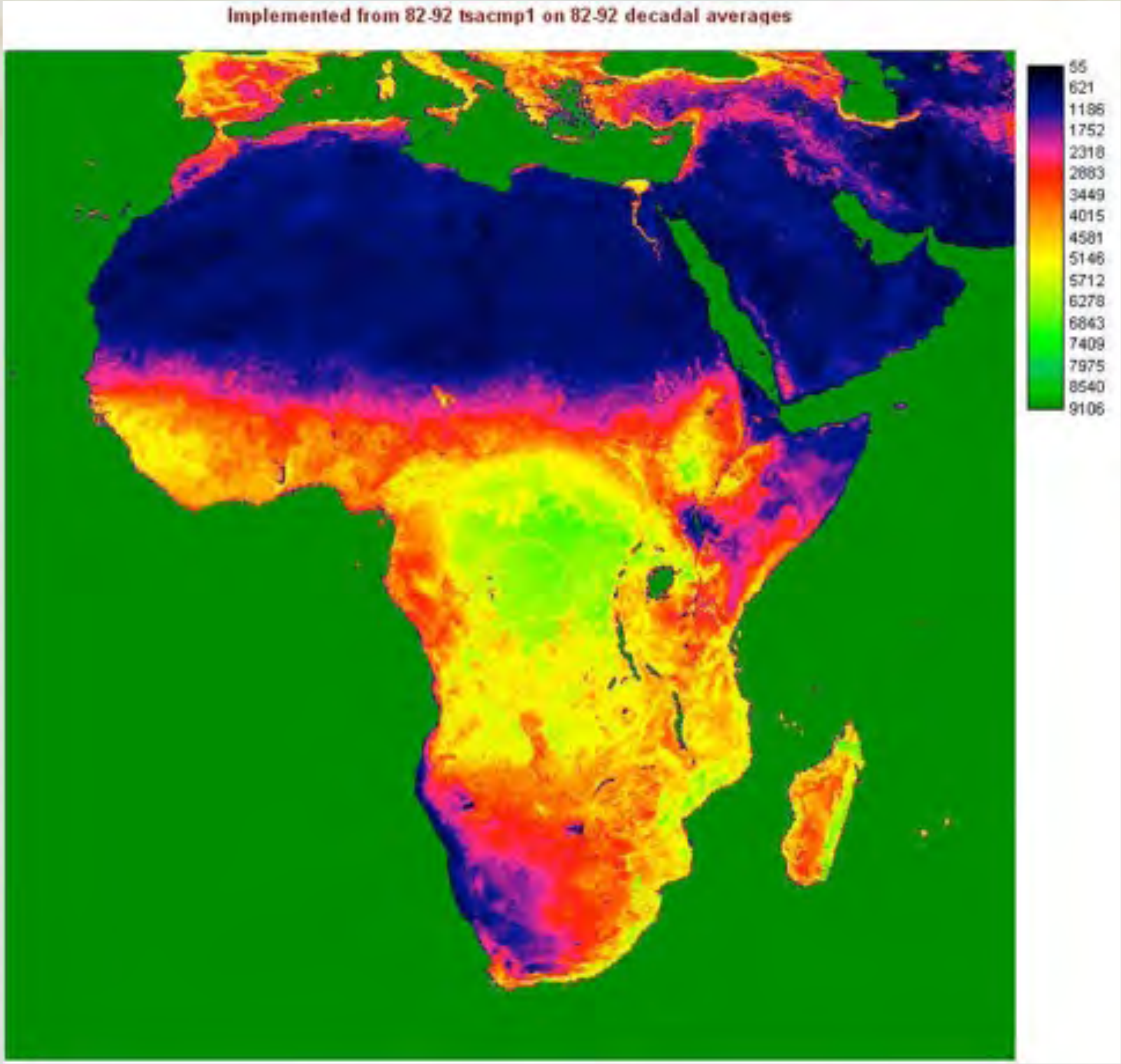
Average NDVI
1982-2004

Example - mapping african vegetation using PCA

	CMP 1	CMP 2	CMP 3	CMP 4	CMP 5	CMP 6
% Var	98.59048	1.035038	0.226096	0.085785	0.043496	0.006479
Loadings :						
	CMP 1	CMP 2	CMP 3	CMP 4	CMP 5	CMP 6
101	0.993337	-0.09901	0.048344	-0.0076	0.02786	0.007527
102	0.99266	-0.10829	0.044178	-0.01942	0.020256	0.000074
103	0.992226	-0.11372	0.039275	-0.02756	0.011667	-0.00679
201	0.992021	-0.11649	0.035049	-0.03118	0.002331	-0.01297
202	0.991694	-0.12038	0.027372	-0.03416	-0.00525	-0.01336
203	0.991518	-0.12283	0.018812	-0.03401	-0.01377	-0.01335
301	0.991634	-0.12221	0.008742	-0.02973	-0.0214	-0.00902
302	0.992108	-0.12055	-0.00472	-0.02132	-0.02721	-0.00099
303	0.992466	-0.11604	-0.01933	-0.01302	-0.03123	0.004552
401	0.993106	-0.10762	-0.03479	-0.00269	-0.03169	0.011244
402	0.993708	-0.09424	-0.04771	0.007093	-0.02751	0.014348
403	0.994404	-0.07714	-0.06425	0.019093	-0.02276	0.014018
501	0.995024	-0.0558	-0.07645	0.028797	-0.01345	0.007886
502	0.995332	-0.03208	-0.08377	0.033493	-0.00391	-0.00437
503	0.99563	-0.00827	-0.08646	0.031122	0.006785	-0.00982
601	0.995698	0.016546	-0.08375	0.021169	0.017201	-0.01315
602	0.995626	0.040324	-0.07756	0.005466	0.025648	-0.00583
603	0.995134	0.064109	-0.06652	-0.00879	0.030934	-0.00499
701	0.994307	0.08569	-0.05291	-0.02314	0.031373	-0.00013
702	0.992826	0.106176	-0.03746	-0.03154	0.02496	0.000984
703	0.991037	0.12434	-0.02106	-0.03915	0.016261	0.004884
801	0.989031	0.140057	-0.00475	-0.04163	0.004773	0.006784
802	0.987292	0.151797	0.009599	-0.03946	-0.00671	0.007163
803	0.986033	0.159555	0.020642	-0.03249	-0.01608	0.005776
901	0.986257	0.159637	0.028572	-0.02124	-0.0218	0.002952
902	0.987354	0.153173	0.033837	-0.0063	-0.02389	-0.00059
903	0.989098	0.14	0.037978	0.010968	-0.02323	-0.00404
1001	0.99124	0.11945	0.040607	0.028015	-0.02033	-0.00645
1002	0.993439	0.093659	0.042606	0.042171	-0.01558	-0.00733
1003	0.995419	0.064386	0.043987	0.051059	-0.00903	-0.00638
1101	0.996981	0.034115	0.044158	0.053318	-0.00076	-0.00367
1102	0.997644	0.0034	0.046174	0.049151	0.008654	0.000227
1103	0.997571	-0.02502	0.047341	0.039986	0.017996	0.004493
1201	0.996764	-0.0505	0.048067	0.027498	0.025797	0.00823
1202	0.995372	-0.07213	0.048576	0.013617	0.030637	0.010698
1203	0.994232	-0.08921	0.048688	0.000196	0.031375	0.011423

Loadings from a Principle Component Analysis representing an average annual vegetation cycle in Africa over 36 dekads.

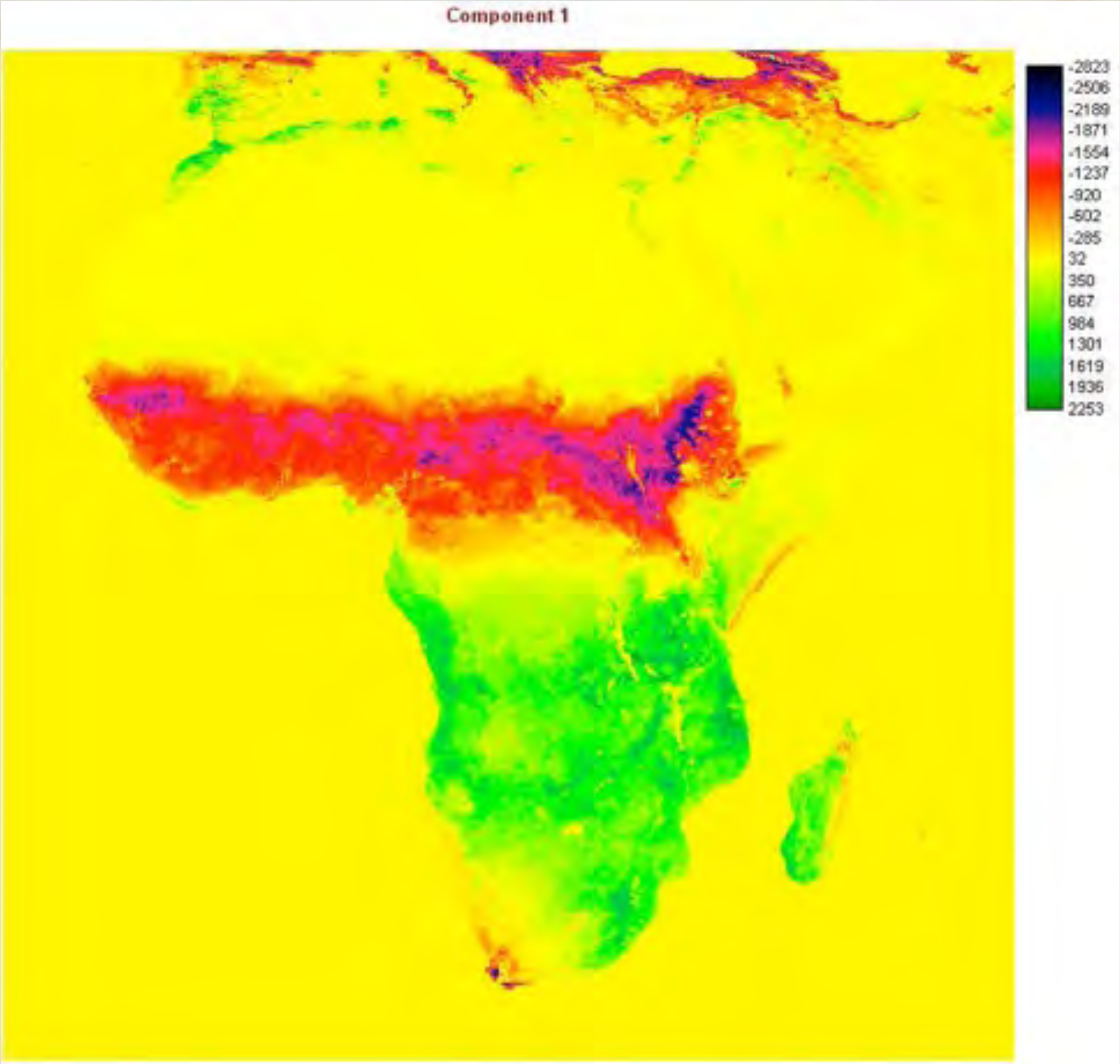
Example - mapping african vegetation using PCA



PCA component 1 representing an average annual vegetation cycle in Africa over 36 dekads. Component 1 show average vegetation.

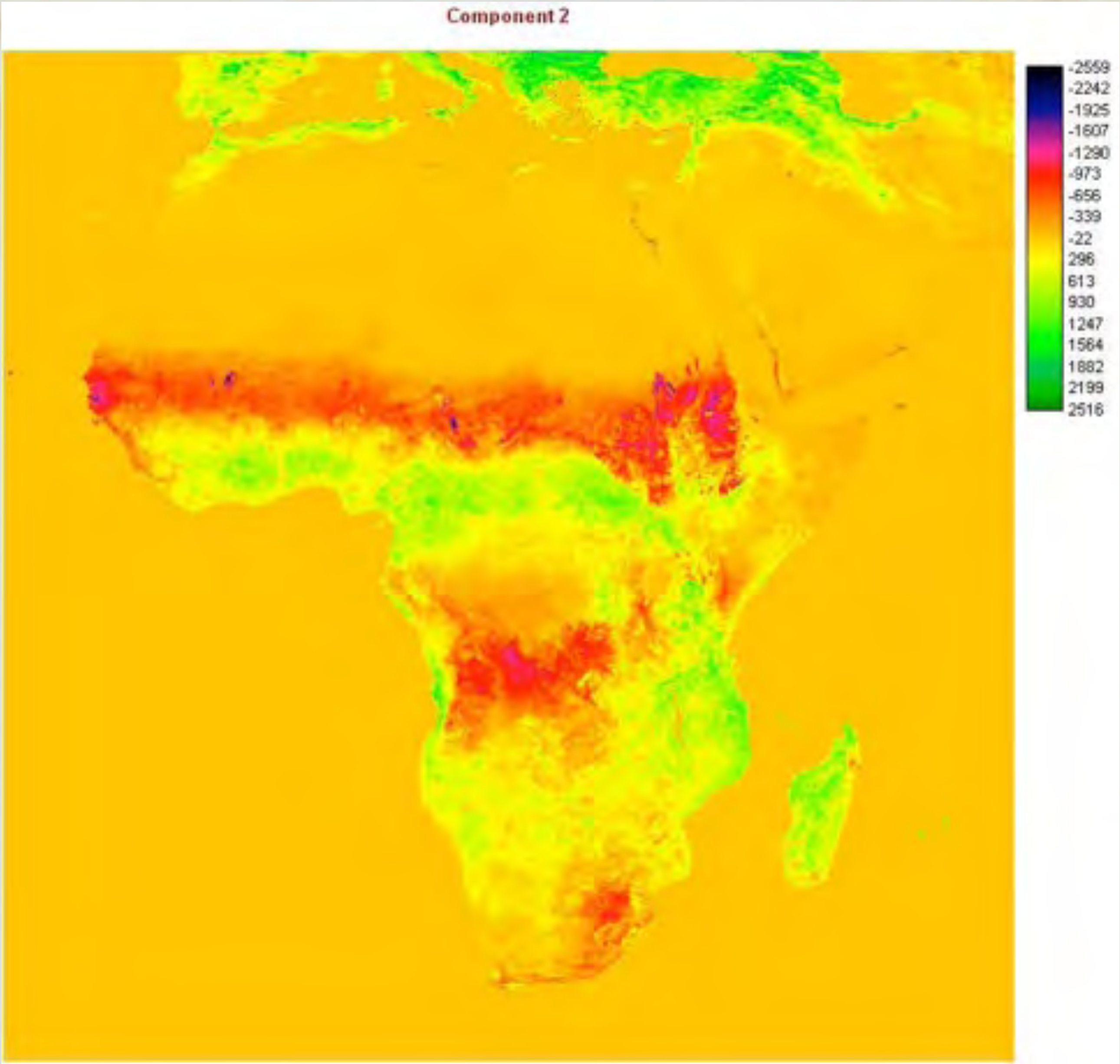
As PCA 1 is almost exclusively portraying the average, and carries > 99 % of the variation. I choose to use normalized data.

Example - mapping african vegetation using PCA



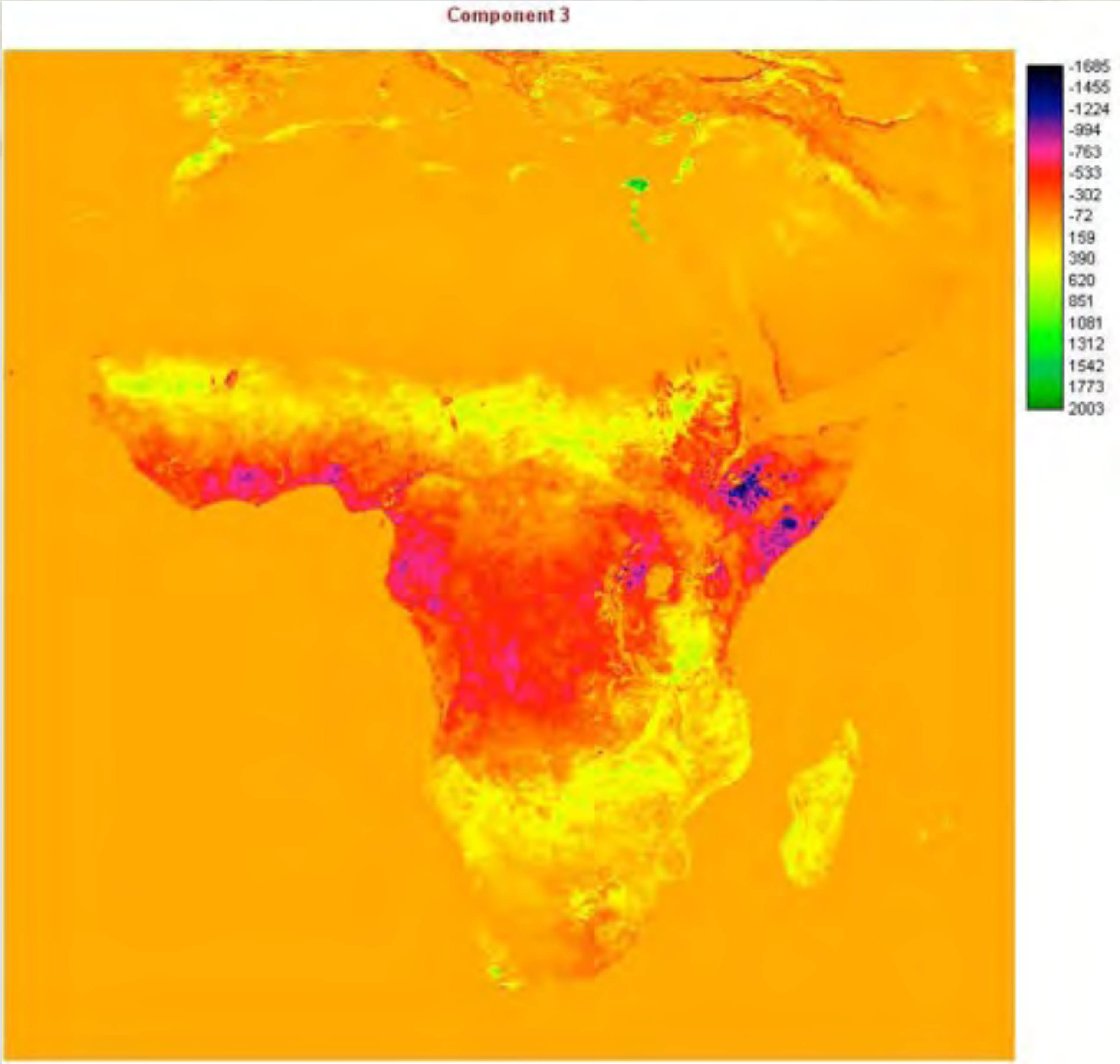
PCA component 1 representing an average annual vegetation cycle in Africa over 36 dekads.
Data normalised over total average per pixel.
Component 1 carries 63 % of the variation and show seasonal behavior.

Example - mapping african vegetation using PCA



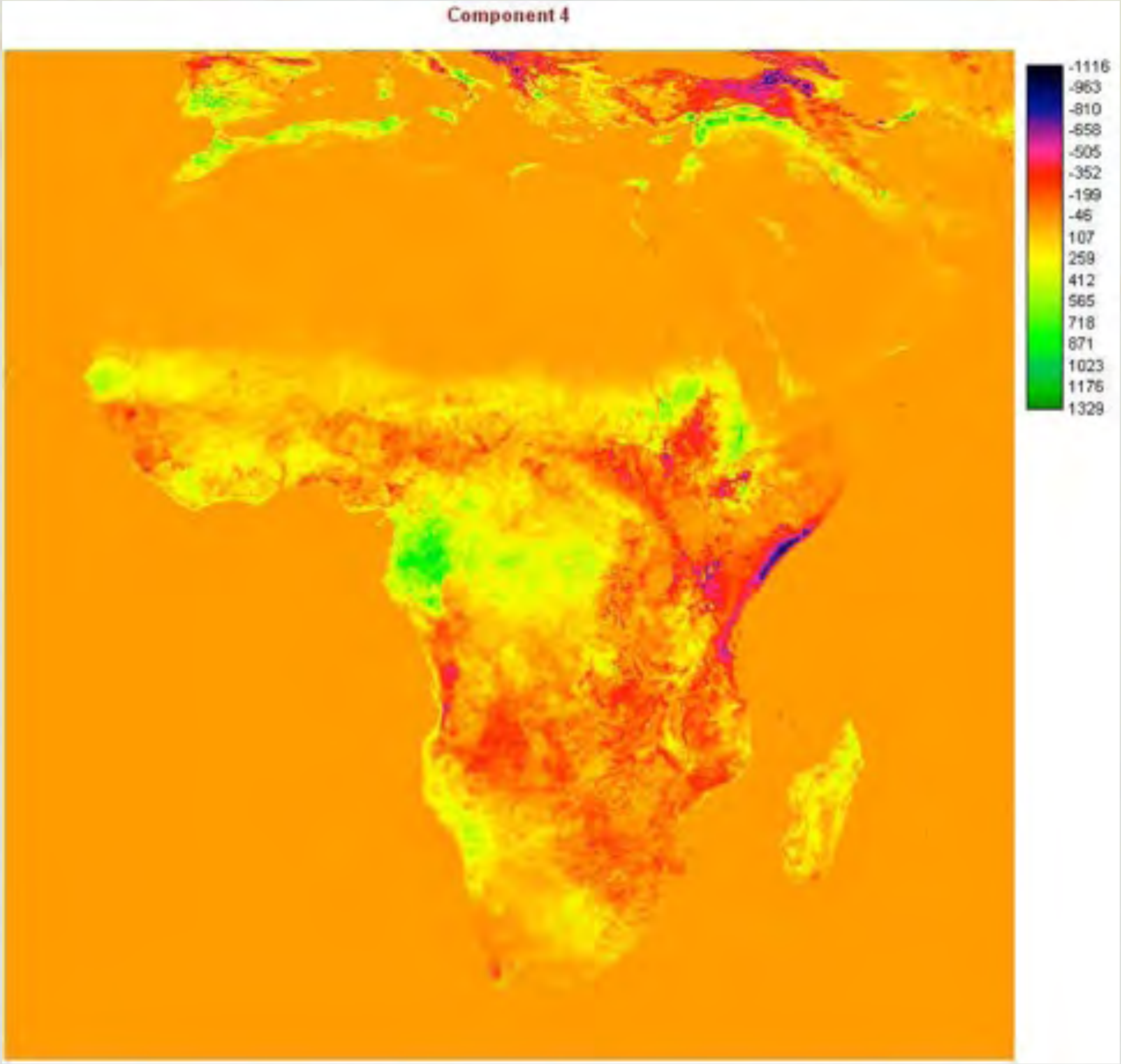
PCA component 2 representing an average annual vegetation cycle in Africa over 36 dekads.
Component 2 carries 24 % of the variation and show seasonal behavior.

Example - mapping african vegetation using PCA



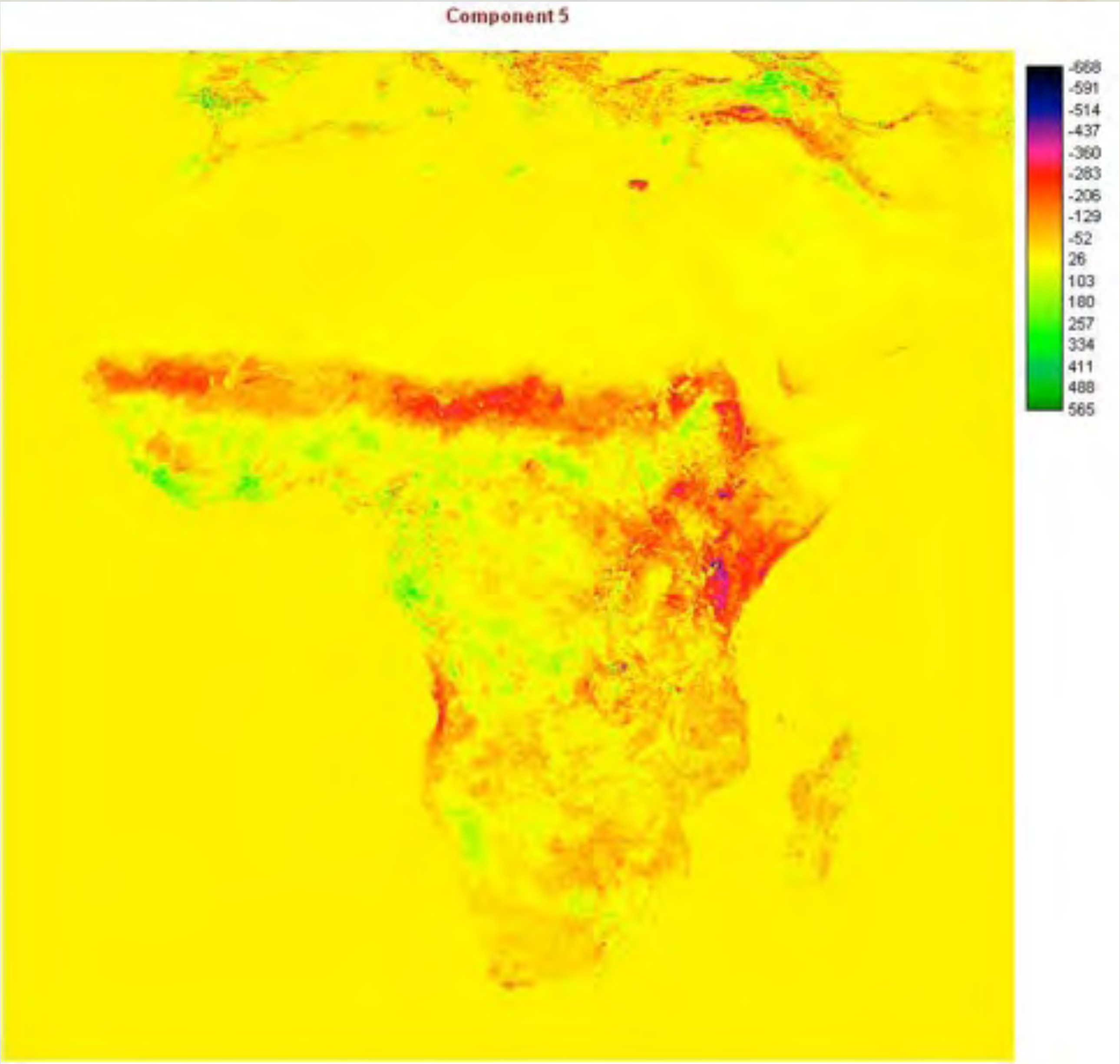
PCA component 3 representing an average annual vegetation cycle in Africa over 36 dekads. Component 3 carries 8.3 % of the variation and show clouds (low values) and vegetation zonation.

Example - mapping african vegetation using PCA



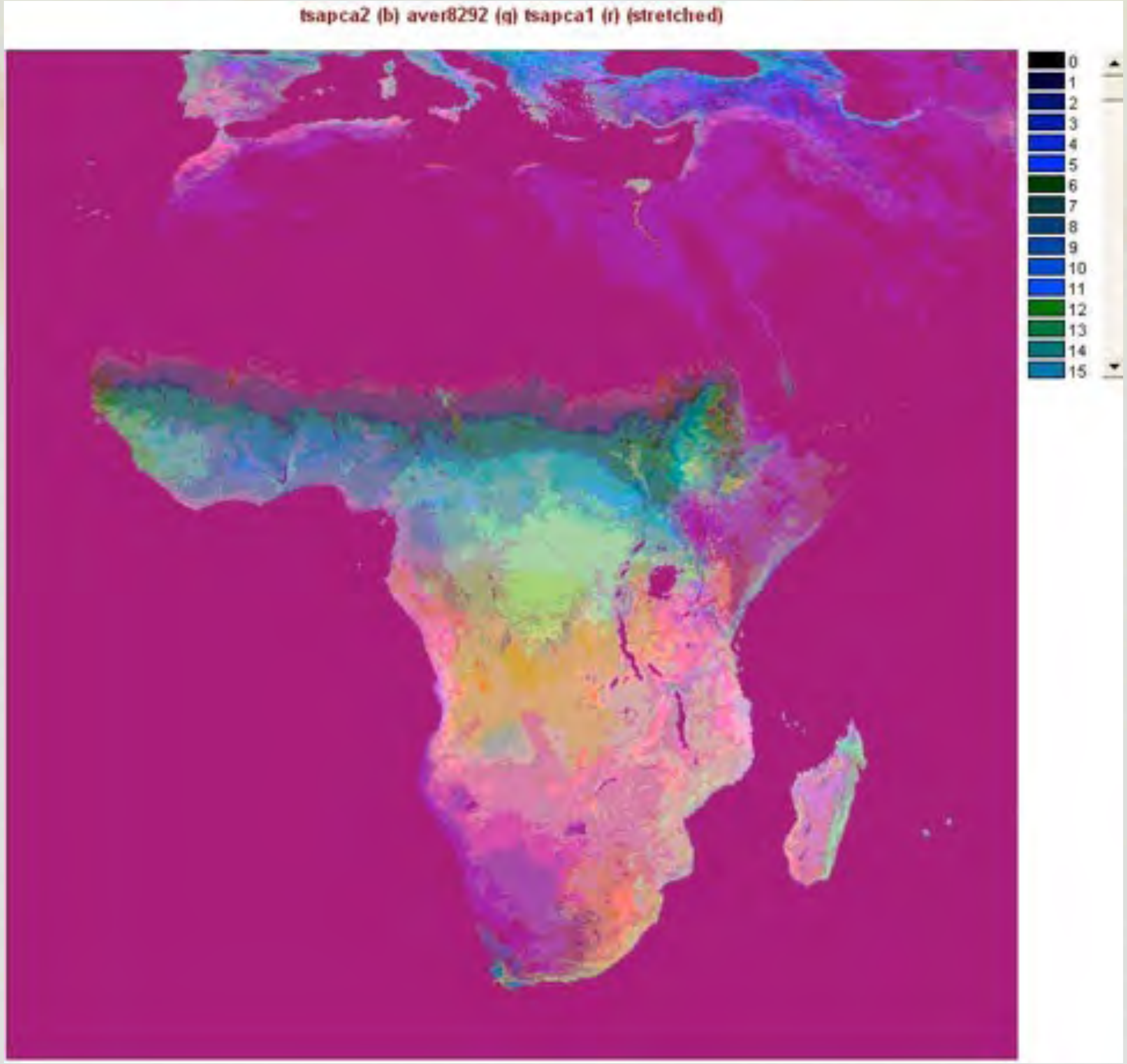
PCA component 4 representing an average annual vegetation cycle in Africa over 36 dekads. Component 4 carries 3.6 % of the variation.

Example - mapping african vegetation using PCA



PCA component 5 representing an average annual vegetation cycle in Africa over 36 dekads. Component 5 carries 0.7 % of the variation. I think it shows the drought prone areas (interpretation problem)

Example - mapping african vegetation using PCA



False color composite visualisation of the PCA timeseries data.

This is a color composite from the Normalised data
B = PCA 2
G = NDVI
R = PCA 1

Example - mapping african vegetation using PCA

Unsupervised ISOCLASS classification

Iterative Self-Organising Data Analysis Technique
• Steps in the algorithm
– Initial state selected; i.e. no. and center of clusters
– Each point in FS labelled to closest center (decision rule of closest distance to center)
– Mean calculated for cluster center
– Relabel points using new means
– Iterate until acceptable percentage of pixels don’t change between clusters

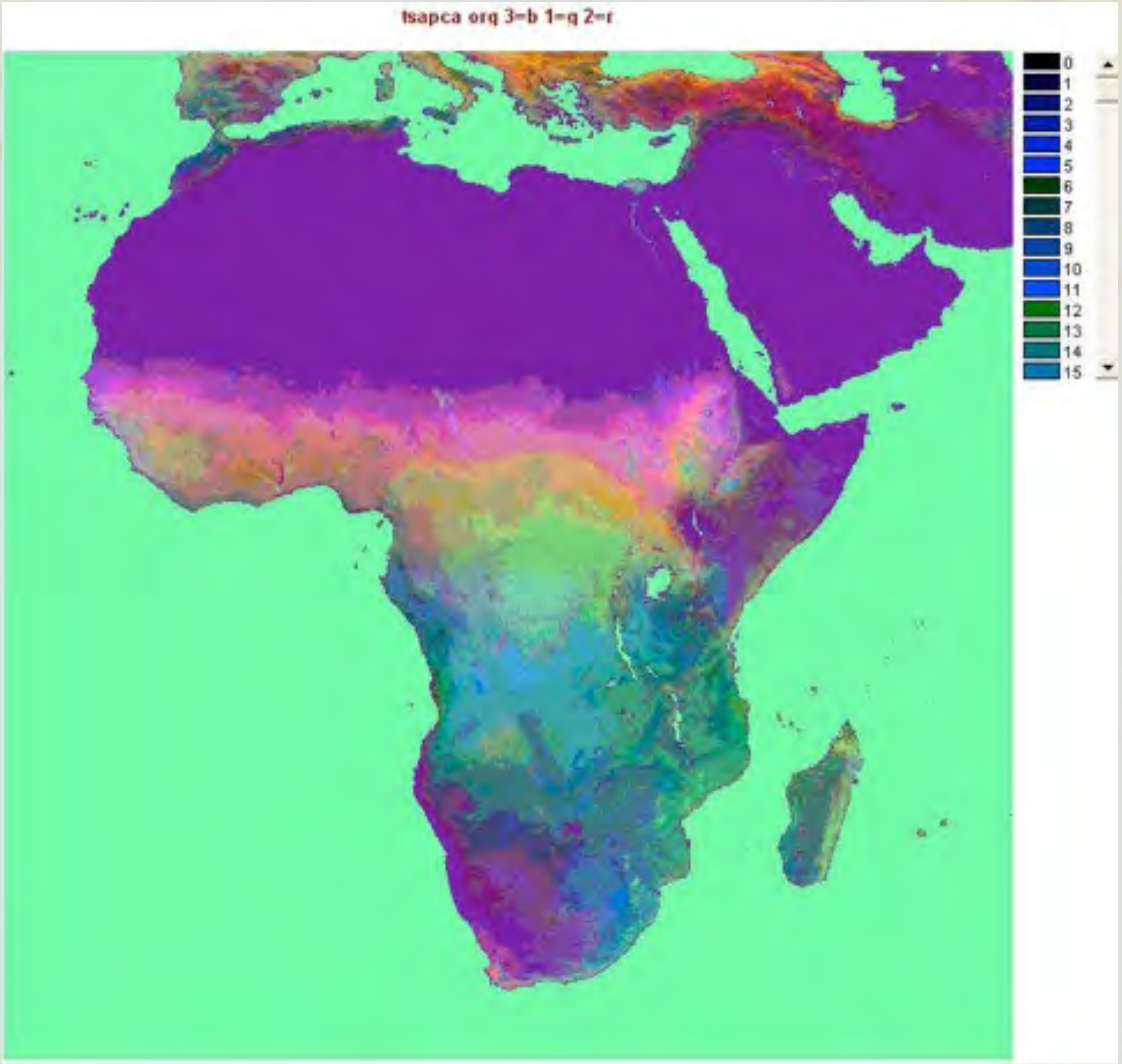
Example - mapping african vegetation using PCA

Unsupervised ISOCLASS classification

- Iterative process
- User defined variables

– Number of Classes this number specifies the exact number of thematic categories (classes) that will be produced
– Number Iterations this number will determine the maximum number of times the ISODATA process will be performed on a given data set.
– Convergence Threshold this setting will determine the percentage of pixels that must remain in a cluster from one iteration to the next in order to stop the ISODATA process.
– Classify Zeros this option specifies whether the classification will include pixels with a value of zero.
– Skip Factor this option will have the process skip the number of pixels for the 'X' and 'Y' set by the user. The higher the skip factors, the faster the process, but the lower the overall accuracy and the smaller the output thematic image.
– Initialize options; principal versus diagonal axis

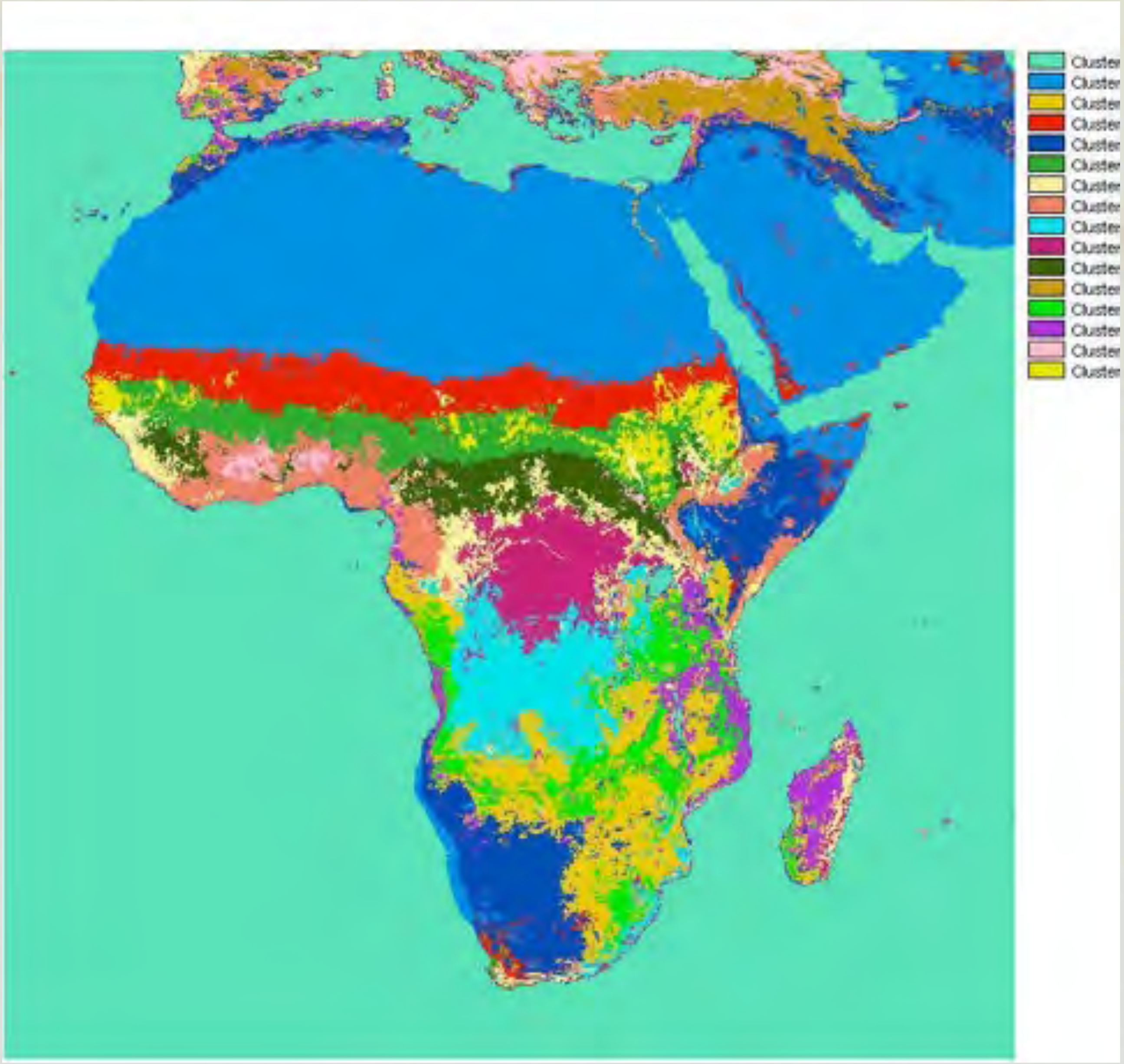
Example - mapping african vegetation using PCA



False color composite visualisation of the PCA timeseries data.

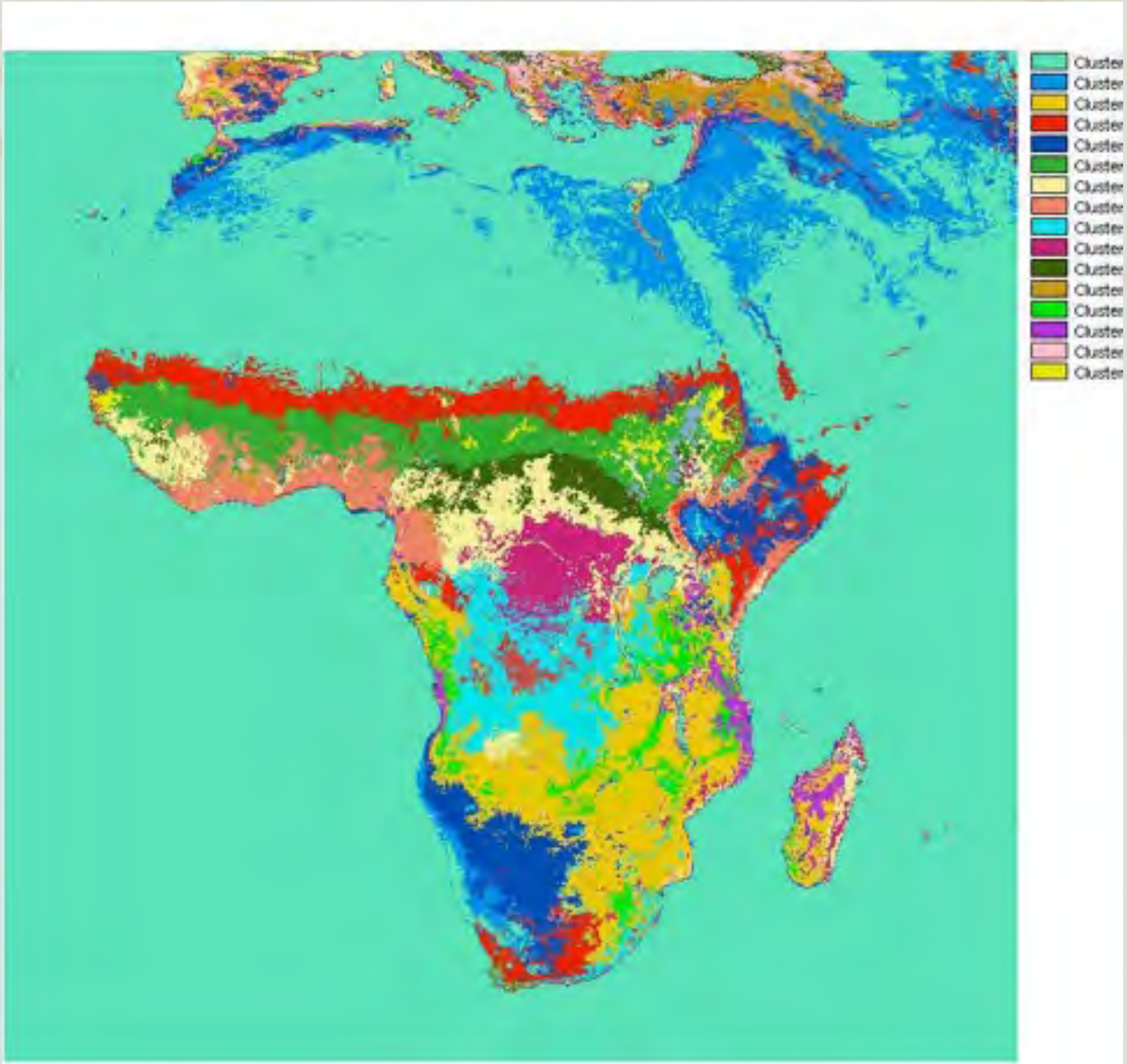
This is a color composite from the original data
B = PCA 3
G = PCA 1
R = PCA 2

Example - mapping african vegetation using PCA



32 classes from unsupervised clustering of the normalised data.

Example - mapping african vegetation using PCA



16 classes from unsupervised iso-clustering of
NDVI
PCA 1
PCA 2
from the normalised data.