What Is the Output of Visual Data Analysis?

Gennady & Natalia Andrienko
Fraunhofer Institute IAIS,
Sankt Augustin, Germany
http://www.ais.fraunhofer.de/and

Visual Analytics Panel @ InfoVis conference, 6/7/2006, London

---

Primary Task of Analysis

Data:

$f : \mathbf{R} \rightarrow \mathbf{C}$

e.g. time → temperature

➢ Explore data = characterise the behaviour of the data function (attribute, group of attributes) over the reference set **R**

= Represent the behaviour by an appropriate *pattern* (a.k.a. *model*)

increase   decrease

t_1

E.g. a verbal pattern: "increase from x_1 to x_2 over the period from t_0 to t_1, then decrease to x_3 over the period from t_1 to t_2". (a formula, a graphical pattern, …)

A *compound* pattern; consists of 2 subpatterns

Requirements to Visual Tools

- <u>Primary task</u>: characterise the behaviour of the data function over the reference set
    - ⇒ The analyst needs a tool allowing him to see simultaneously the entire reference set and all the corresponding characteristics
    - ⇒ The tool should represent the characteristics so that they perceptually merge into a single unit (to be seen as a behaviour rather than multiple separate characteristics)
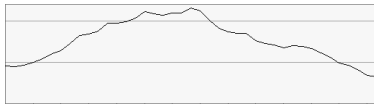
E.g. a good representation: all characteristics are represented by a single line, which is perceived as a unit

☞ But… such a representation is seldom achievable

Data Complexities

☹ Multi-dimensionality
- ➢ e.g. time $\times$ space $\rightarrow$ temperature

☹ Multiple attributes
- ➢ e.g. time $\times$ space $\rightarrow$ (temperature, wind, precipitation, …)

☹ Large data volume
- ➢ e.g. $(t_i, s_i, temp_i, wind_i, prec_i, …)$; i = 1, 2, …, 1 000 000, …

☹ Complex, heterogeneous nature of the reference set
- ➢ e.g. geographical space: land, water, mountains, cities, …

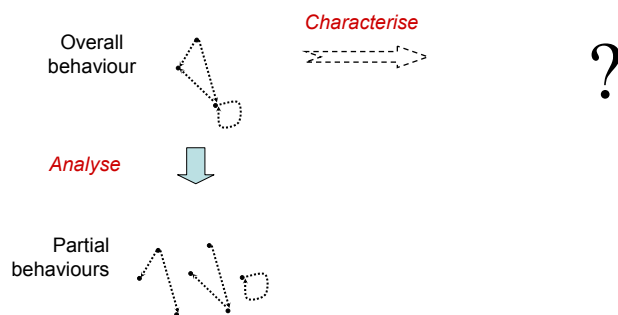☹ Outliers, discontinuities, …

Need for Analysis and Synthesis

The main task:

Overall behaviour

Characterise

?
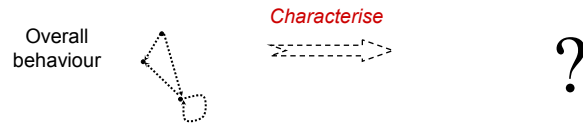
But… The data are too complex!

Need for Analysis and Synthesis

The main task:

Overall behaviour

Characterise

?

Analyse

Partial behaviours

Need for Analysis and Synthesis

The main task:

Overall
behaviour

Characterise

?

Analyse

Partial
behaviours

Characterise

Partial
patterns

Need for Analysis and Synthesis

Overall
behaviour

Overall
pattern

Analyse

Synthesise

Partial
behaviours
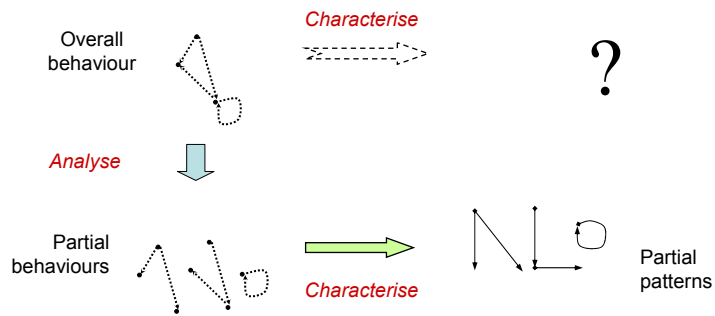
Characterise
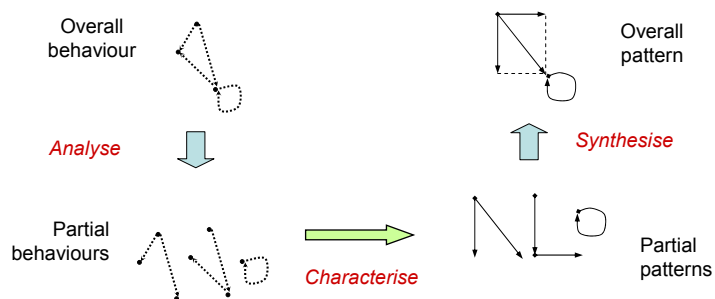
Partial
patterns

Tool requirements:

- Support analysis, i.e. division into partial behaviours
- Support the characterisation of the partial behaviours
- Support the synthesis of the overall pattern

4

Example: Behaviour over a Two-Dimensional Reference Set

Referrers

Space (set of the states of the USA)

Time (set of years from 1960 to 2000)

Attributes

• Property crime rate
• Violent crime rate
• …

This behaviour cannot be represented as a single unit
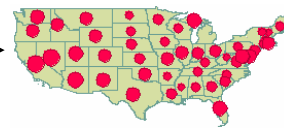

Slicing a Behaviour

Space as a whole

Specific time t

Spatial behaviour
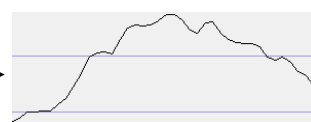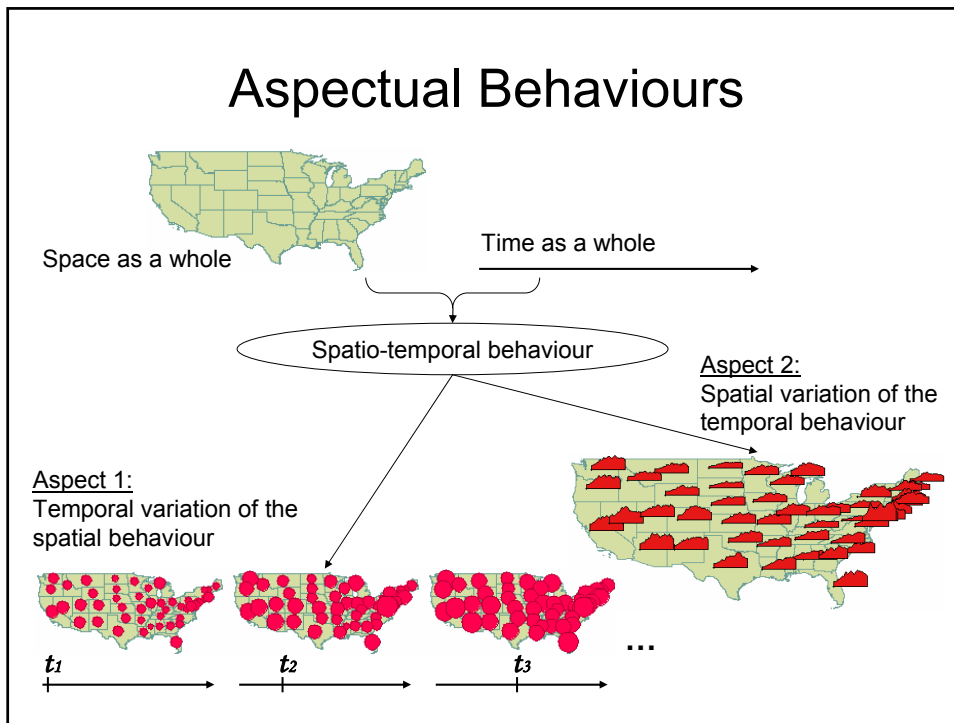(value distribution over the space)

Specific place

Time as a whole

Temporal behaviour
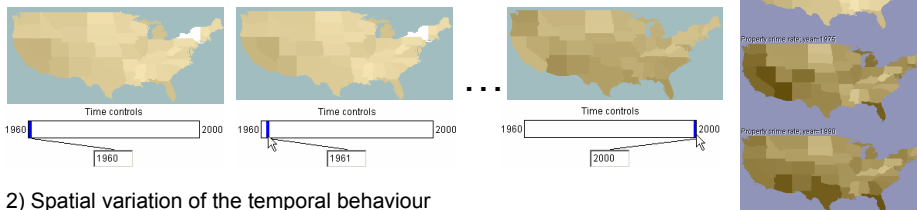(value variation over the time)

Aspectual Behaviours

Space as a whole

Time as a whole

Spatio-temporal behaviour

Aspect 1:
Temporal variation of the spatial behaviour

t_1 t_2 t_3 ...

Aspect 2:
Spatial variation of the temporal behaviour

Exploring Aspectual Behaviours

1)Temporal variation of the spatial behaviour

Time controls
1960   2000
1960

Time controls
1960   2000
1961

...

Time controls
1960   2000
2000

Property crime rate year=1960

Property crime rate year=1975

Property crime rate year=1990

2) Spatial variation of the temporal behaviour
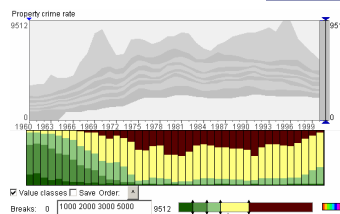
☞ The diagrams are perceived as separate entities → the map must be scanned and cannot be grasped as a single image

☞ Absence of ordering complicates seeking for specific behaviour patterns
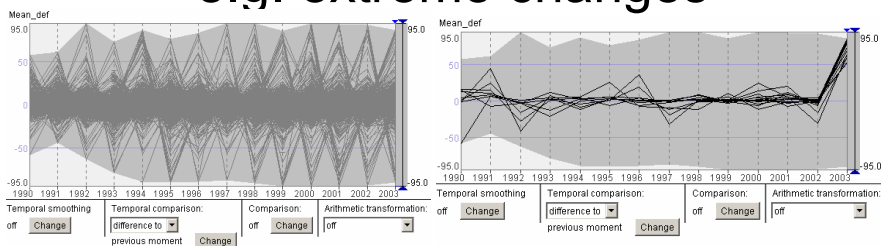
☞ Diagram overlapping is a serious problem

Property crime rate
9512   9512
0   0
1960 1963 1966 1969 1972 1975 1978 1981 1984 1987 1990 1993 1996 1999

☑ Value classes ☐ Save Order:
Breaks: 0   1000 2000 3000 5000   9512

☞ Highly synoptic views, but…

☞ no spatial context!
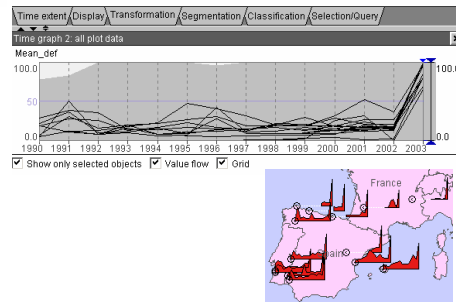
A Possible Approach: Group By Similarity

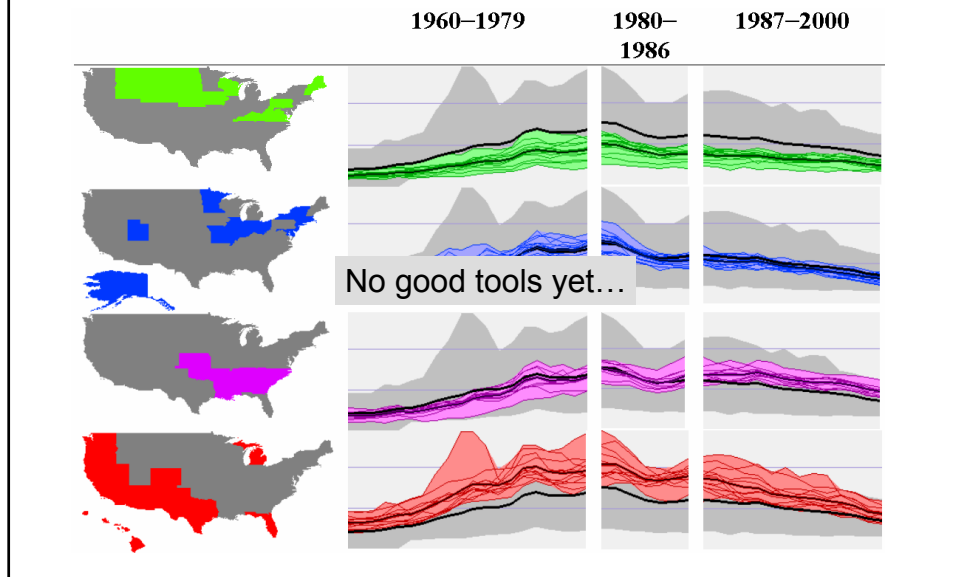E.g. by applying cluster analysis
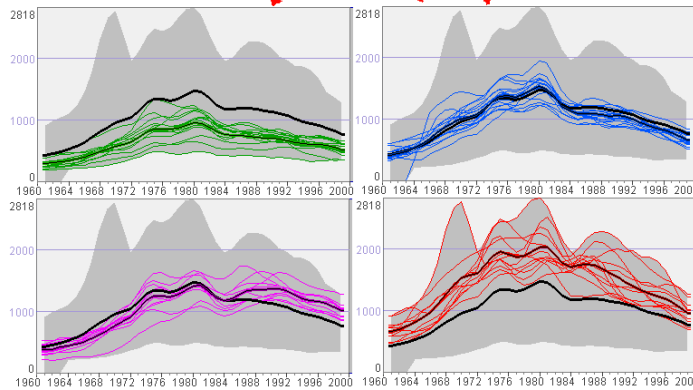
Attend to particulars
e.g. extreme changes

1. Transform the time graph to show changes
2. Select extreme changes in a specific year (here 2003)

Synthesising an Overall Pattern

No good tools yet…

Knowledge Synthesis Challenges

- Decomposition of a complex data analysis task results in multiple unconnected observations about various aspects of the overall behaviour and interesting particulars
- These are mostly visual impressions and tacit ideas
- How to capture these observations?
- How to put these together and to get something tangible?
- How could this "something" look like?

What Is the Output of Visual Data Analysis?

➢ This is an important scientific question

➢ This is not only a scientific question but a matter of survival of the InfoVis community:

- – People believe more in texts and numbers than in visualisations (R.Burkhard)
- – Possibly, because they do not see material results, e.g. models supporting their decisions
- – A consequence: lack of industrial and research funding (at least in Europe)