



**On exploration and mining of data
in education practice**

Dr. Erik-Jan van der Linden (MagnaView)
Ir. Thomas Lammers (MagnaView; TU/E)
Ir. Martijn Wijffelaars (MagnaView)



EDM conference
Eindhoven, July 6, 2011



visualize anything • visualize everything 2

The image is a slide with a dark blue background and a grid pattern. It features a central white text box with a dotted pattern. The text in the box is white and bold. To the right of the text box is a logo for "THE EUROPEAN ICT PRIZE NOMINEE" which includes a stylized blue and white figure and a yellow star. At the bottom left of the slide is the MagnaView logo, and at the bottom right is the tagline "visualize anything • visualize everything" followed by the number "2".

Speakers

Van der Linden, Lammers, Wijffelaars
MagnaView (company)

Our customers solve their business issues because we enable them to get the most out of their data, fast.



visualize anything • visualize everything 3

Speakers

Partly University Eindhoven spin-off

Lost for academia..?

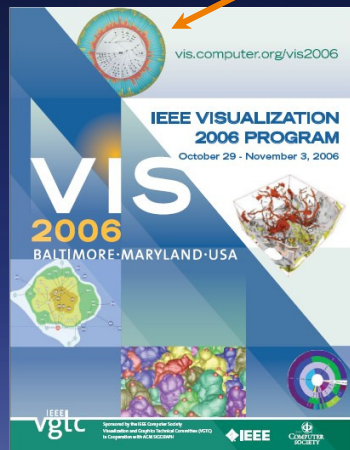
Yes... Please interpret company blabla as enthusiasm

No... R&D 'intensive': 10.000 hours per year.
Examples today.



visualize anything • visualize everything 4

Credentials: scientific publications Cooperation Eindhoven TU - MagnaView



IEEE Visualization, InfoVis 2006
Baltimore, Maryland

UDC Seminar 2007
The Hague, the Netherlands

EuroVis 2008
Eindhoven, the Netherlands

Overview

- A. Education practice: Context
- B. Brief introduction MagnaView
- C. Visual exploration in education
- D. Data mining meets visualization

A. Education practice: Context

A. Context

- Recognition of *knowledge* and *innovative power* as *the* drivers for economic development.
- Education is high on the political agenda.
- OECD Programme for International Student Assessment (PISA) considered important by politicians: *how does our national sector do?* → translate to policies which apply to institutions.

Context (ii)

- Insight of governments and politicians → Impossible to adapt system on basis of centralized regulations etc. → *schools run education*.
- Netherlands' conclusion: *lump sum* funding → right to spend
= **obligation to justify!**
Both spending and outcome.

Context (iii)

- N.B. International?
- *Germany*: visible shift. Ministeries and municipalities carry over budgets to schools
- *Belgium*: slowly shifting attention towards teachers
- *Finland*: best system in the world because of professional status of teachers; why change..?

Context (iv)

- Other development: data accumulates in school information systems.
- *If I were Coca Cola...*
- Data has become available → politicians: 'if they HAVE the data, we can ASK!'
- And Calvinist Dutch: '*we can publish...*'
 - 'Kwaliteitskaart'
 - 'Trouw'

Context (v)

- So, external pressure on schools to USE their data for *external* reporting...
- ... but as a consequence use for *internal* improvement.
- Impossible to change the school on basis of centralized regulations etc. → individual teachers teach/individual student learn
- Of course, the one-on-one communication between teacher and student is key.

This context....

... is our playground...

Companies only survive if they start from 'needs' of their customers...

... which is different from...



visualize anything • visualize everything 13

EDM (2011 preface)

Data-driven/curiosity-driven.

The increase of e-learning resources such as interactive learning environments, learning management systems, intelligent tutoring systems, and hypermedia systems, as well as the establishment of state databases of student test scores, has created large repositories of data that can be explored to understand how students learn.

MagnaView: issue-driven



visualize anything • visualize everything 14

B. Brief introduction to MagnaView



visualize anything • visualize everything 15

About MagnaView

- Started in 2004 in Eindhoven - The Netherlands
- **Our mission**
Our customers solve their business issues because we enable them to get the most out of their data, fast.
- **What we offer**
 - α General purpose Visual Analytics **Software**
 - β Turn key Analytics **Products** for specific sectors
 - γ Full **Service**



visualize anything • visualize everything 16

α General purpose visual analytics software



| | |
|--|---|
| Tell me <u>here</u>... | Web-based software for end user works on laptop, PC, Smartboard, Tablets, iPhone, iPad, Blackberry |
| .. and <u>now</u> (I really mean NOW) ... | Fastest tool in the marketplace |
| .. what <u>really</u> matters... | Visualize anything and easy to understand for everybody → create sense of urgency |
| ... allow me to gain <u>insight</u>.... | Easy mouse-click navigation to follow <i>train of thought</i> |
| ... and to decide where to <u>act</u>. | Any level of detail available: in-memory analysis of large datasets |



visualize anything • visualize everything 17

α General purpose visual analytics software



Desktop products



Enterprise products



OEM



Cloud / SaaS
Managed services



visualize anything • visualize everything 18

β Turnkey Products



| | |
|--|--|
| <i>We require information relevant for our sector...</i> | MagnaView offers Products for specific sectors |
| <i>... and we really mean <u>relevant</u>.</i> | Users are continuously involved in development of the MagnaView Products |
| <i>We don't want to bother about data</i> | MagnaView offers truly <u>turnkey</u> products which include queries, data handling, etc |
| <i>We want it integrated in our working environment</i> | MagnaView cooperates with your current software vendor |



β Turnkey Products

Health care: MagnaView MicroLab - Pathos



β Turnkey Products

Legal: MagnaView Patronus - Daan
 Education: MagnaView Cum Laude




γ Full service



| | |
|--|---|
| <i>We want <u>fast installation and integration with least effort</u></i> | MagnaView takes care of all aspects of installation and integration |
| <i>We require projects which lead to <u>direct results</u></i> | MagnaView manages full, <u>agile</u> implementation of projects |
| <i>We want to be able to do it ourselves for lower costs</i> | MagnaView effectively train your expert users, BI/MI staff |
| <i>We need support for creation of reports and analyses</i> | The MagnaView consultants create great reports in no time |
| <i>We want to maximize our ROI</i> | The MagnaView experienced consultants make your organization really profit from its data treasures |



Examples



visualize anything • visualize everything 23

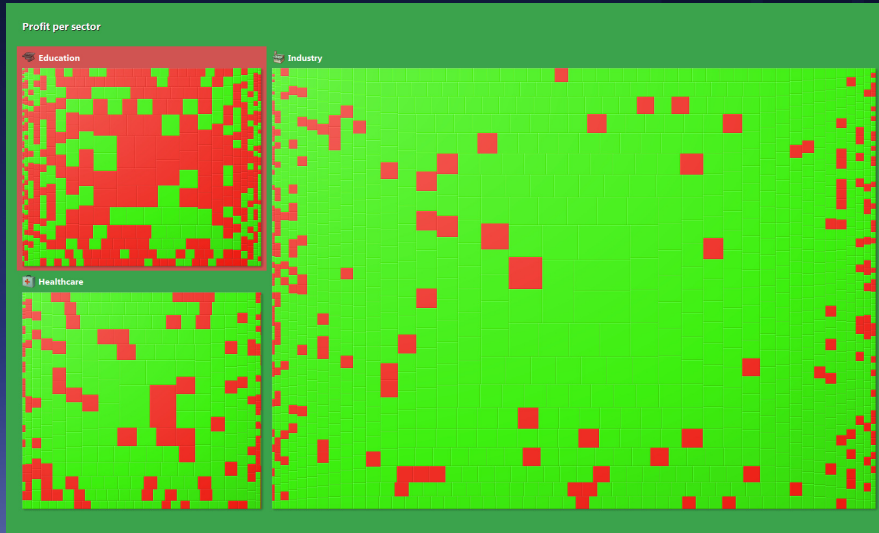
Profitability in legal firms - 1

| | Total Products | Total Map of Productgroups |
|--------------------|-------------------|-------------------------------|
| Company | Product 12 | € 1.074,- |
| | Product 8 | € 2.207 |
| | Product 6 | € 62.540 |
| | | € 63.073 |
| Personal | Product 5 | € 4.334 |
| | Product 2 | € 6.450,- |
| | Product 7 | € 9.910 |
| | Product 10 | € 3.585 |
| | Product 13 | € 17.513,- |
| | Product 11 | € 14.898,- |
| | | € 21.052,- |
| Real Estate | Product 9 | € 7.143 |
| | Product 4 | € 271.207 |
| | Product 3 | € 168.527 |
| | | € 446.877 |
| | € 488.898 | € 488.898 |

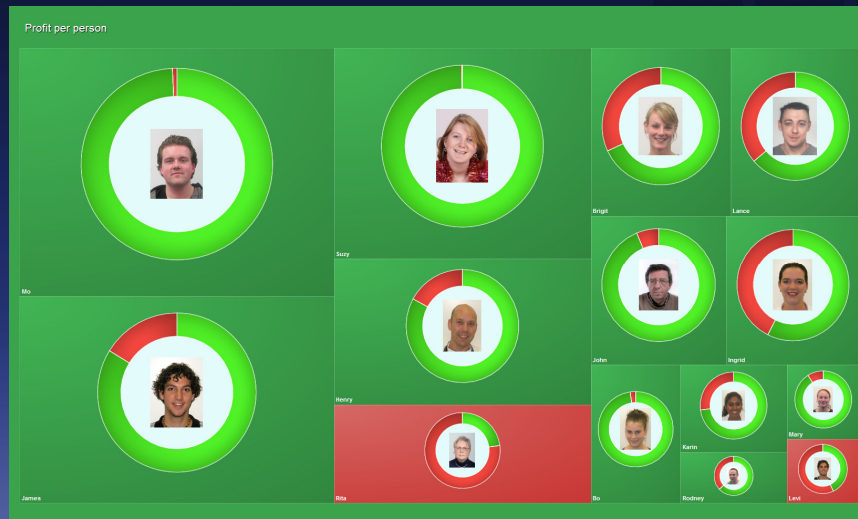


visualize anything • visualize everything 24

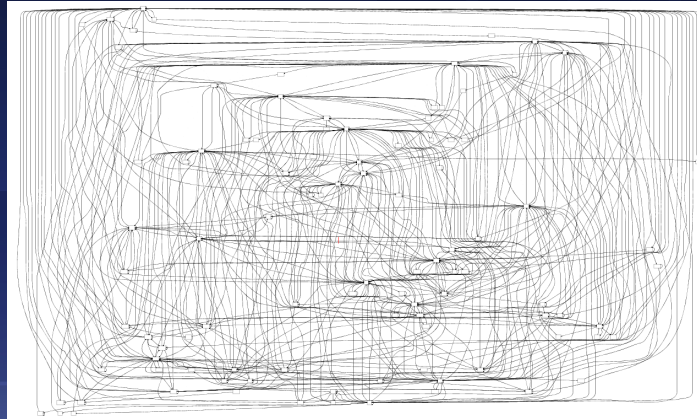
Profitability in legal firms - 2



Profitability in legal firms - 3



Example: hospital (i). Starts from an advanced technique: *Process mining*. Wil van der Aalst at TU/E

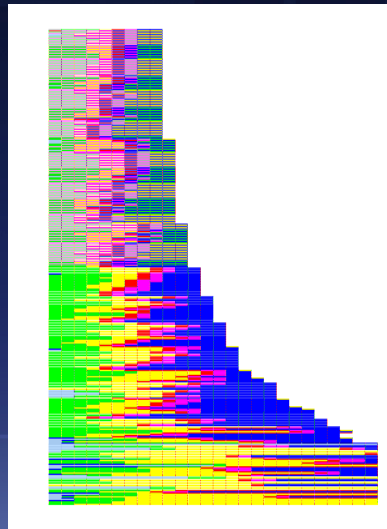
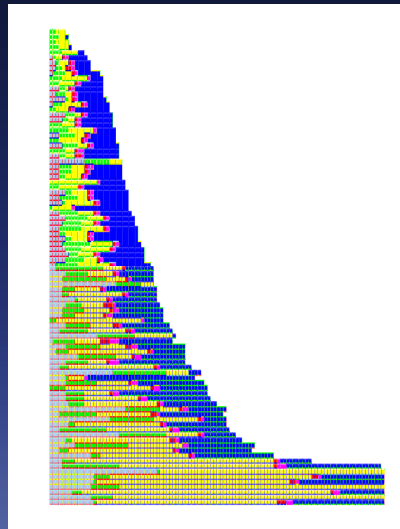


'Wiring scheme!'; 'Spaghetti!'



visualize anything • visualize everything 27

Example: hospital (ii)



visualize anything • visualize everything 28

Change...

➤ the way the **doctor** thinks and acts!!

→ the way the **teacher** thinks and acts...?

C. Visual exploration in education

Cum Laude

Tool for **management information**



What?

- In use by 25 schools
- 75.000 pupils; 100.000 by the end of 2011
- 6-10 years of data
- Secondary Education
- Netherlands (partner in Germany)
- Web-based SAAS-solution

What?

- Information about
 - Population. Pupils and teachers
 - Absence
 - 'Throughput'
 - Grades
 - 'Kwaliteitskaart'



visualize anything • visualize everything 33

Cum Laude

Demo



visualize anything • visualize everything 34

Actual use of Cum Laude

- **Longer term**
 - Evaluation of throughput
 - Regular evaluations of teaching staff
 - Communication with primary education
- **Short term**
 - Absence
 - Grades
 - Meetings of sections
 - Ad hoc information/analyses



visualize anything • visualize everything 35

Actual use of Cum Laude

P.S. MagnaView is interested in contacts with
Education Software Vendors for Cum Laude

martijn@magnaview.com

Tips are rewarded with free licenses ;-)



visualize anything • visualize everything 36

D. Data mining meets visualization

Follow the leader!



visualize anything • visualize everything 37

Data mining meets visualization

MSc Thesis Thomas Lammers. Now @MagnaView.

Not yet applied in schools, but show how general idea may apply to education data.

Reference: *Lammers, van der Linden, Vliegen and Wijffelaars, Interactive Grouping (in preparation).*

Original starting point...

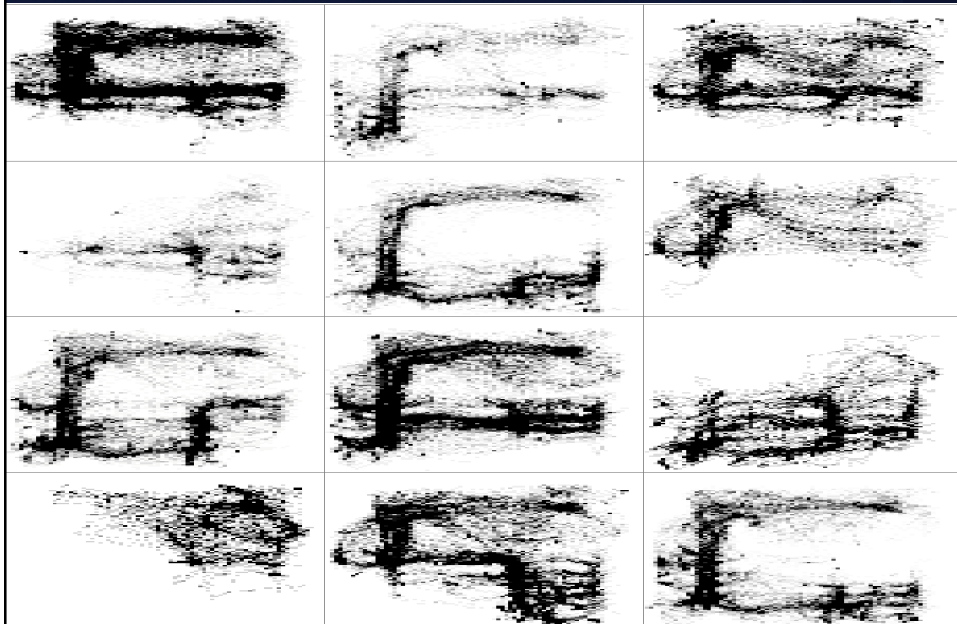


visualize anything • visualize everything 38

Which routes do my customers take?



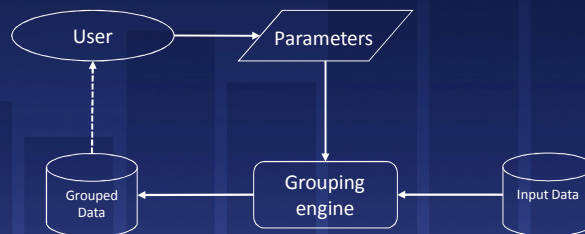
Which routes do my customers take?



Starting observations

- Strength of visual exploration:
 - The user can easily input domain knowledge by creating the best possible visual representation, choosing the attributes he/she is interested in.
- Weakness of purely visual exploration:
 - User has to do all pattern recognition and grouping himself.
- Strength of data mining:
 - Uncover these patterns, groups, etc.
- Weakness of data mining...

Typical approach



But...

- Users with domain knowledge typically do not have understanding of the way DM-techniques/algorithms work
- → DM is not going to be applied in practice (the MagnaView context)
- *Is there another way to combine DM and Viz?*



visualize anything • visualize everything 43

Previous work?

- There is little work on good combinations of DM and visualization.
- 'Visual Analytics' makes the claim, but the emphasis is on interaction and navigation. DM-techniques hardly found.



visualize anything • visualize everything 44

Quite some visual examples in EDM!

The collage displays several visualization types:

- Two heatmaps: one with 'Item' on the y-axis (0-40) and 'Examination' on the x-axis (0-100), and another with 'Item' on the x-axis (0-40) and 'Skill' on the y-axis (1-4).
- A large network graph with nodes and edges, overlaid with a grid.
- A detailed legend for 'Activity Data by Day' with categories like 'Skill', 'Examination', 'Study', 'Sleep', 'Work', 'Meal', 'Travel', 'Other'.
- A circular network graph labeled '(a) Term of'.
- A bar chart showing 'Activity by Day' with a legend for 'Activity'.
- A small line graph showing 'Activity by Day' with a legend for 'Activity'.

magnaview visualize anything • visualize everything 45

But...

- Not intended to serve as interface for user
- Exceptions: *Brick* and *EDM Vis Tool*...?

The left screenshot shows a legend for 'Activity Data by Day' with categories like 'Skill', 'Examination', 'Study', 'Sleep', 'Work', 'Meal', 'Travel', 'Other'. The right screenshot shows a network graph with nodes and edges, overlaid with a grid.

magnaview visualize anything • visualize everything 46

Observations (i)

- Grouping is an *incremental* activity. Users work in cycles
- Know it when they see it
- Browsing data can lead to discovery which groups are desired
- It is easier to critique than to construct.

Can we support that?

Observations (ii)

- *Any group can be a real group for a user.*
Women, Fire and Dangerous things (Lakoff).
How should we support that?
- When a user creates a visual representation, he/she in fact specifies what might be an interesting way to group data.
Could we use that?

Interactive grouping (i)

1. *Have the user interactively change the current grouping:*
 - Splitting all ungrouped entities into multiple groups using a clustering algorithm.
 - Splitting an existing group using a clustering algorithm.
 - Merging existing groups.
 - Creating a new group.
 - Moving an entity from one group to another.
 - Reducing the number of entities in a group.
 - Increasing the number of entities in a group.



visualize anything • visualize everything 49

Interactive grouping (ii)

2. *Automatically re-group all entities, using NN classification algorithm.*
3. *If the current grouping is not the intended one, go to the first step.*

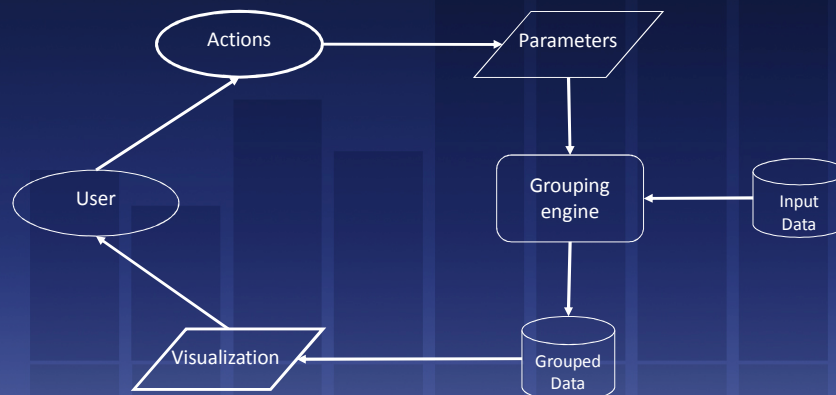


visualize anything • visualize everything 50

DM techniques used

- Off the shelf; no intention to invent new DM-techniques.
- For clustering: K-means.
- Re-group: Nearest Neighbor

Interactive grouping



Leaders and Followers

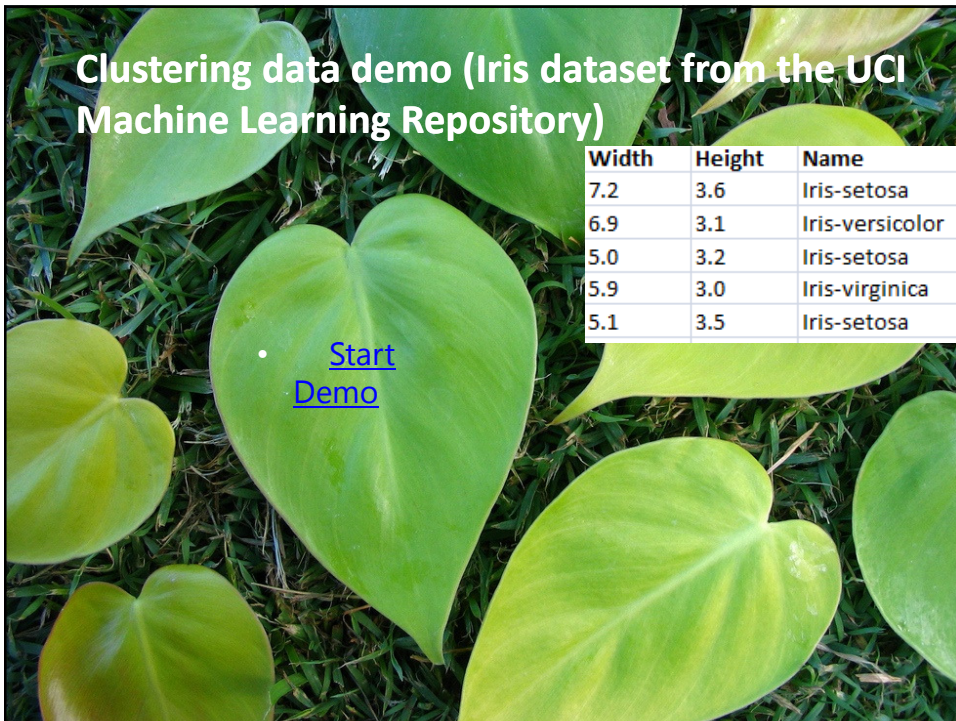
- Metaphor
- Leader is example, that is put manually in a group
- Followers follow leaders



Clustering data demo (Iris dataset from the UCI Machine Learning Repository)

| Width | Height | Name |
|-------|--------|-----------------|
| 7.2 | 3.6 | Iris-setosa |
| 6.9 | 3.1 | Iris-versicolor |
| 5.0 | 3.2 | Iris-setosa |
| 5.9 | 3.0 | Iris-virginica |
| 5.1 | 3.5 | Iris-setosa |

• [Start Demo](#)

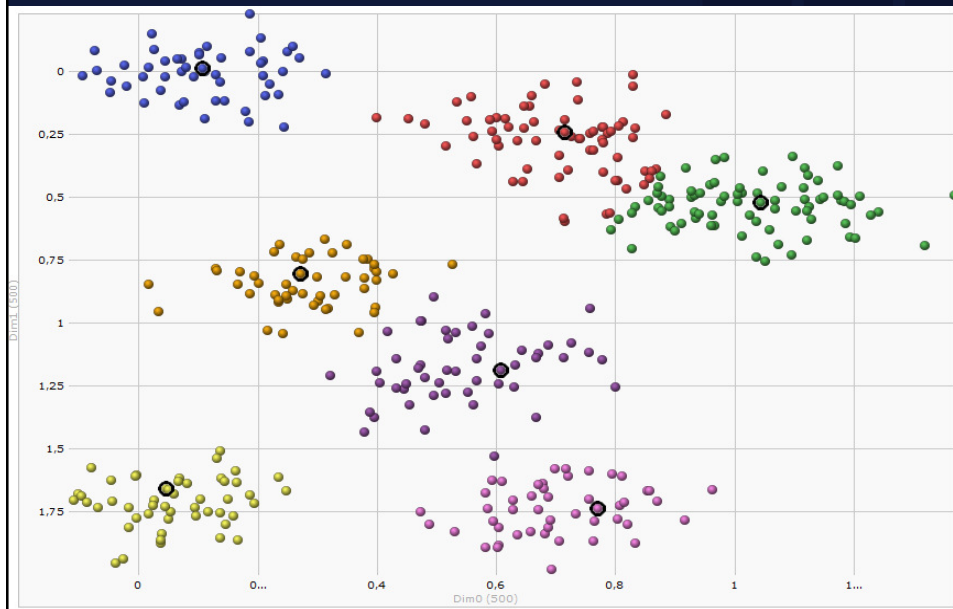


Grouping: summary

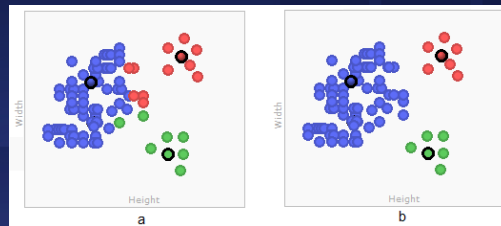
- Group manipulation
 - New group
 - Delete group
 - Merge group
 - Rename group
- Leader manipulation
 - Promote to leader
 - Move leader
 - Demote leader



Split group

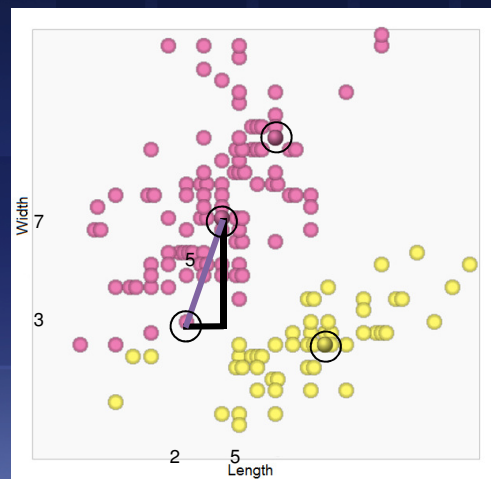


Change group Size



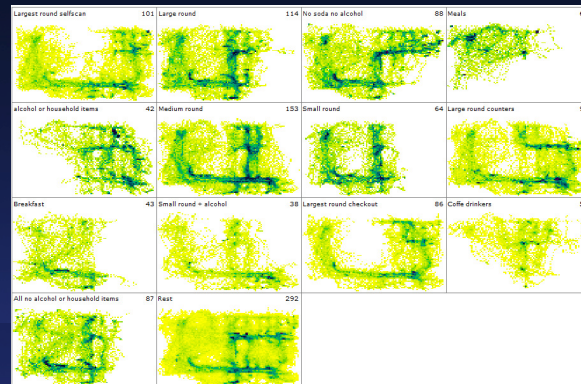
Similarity/Distance

- Data is labelled using Nearest neighbour; leaders are seed points.



Distance functions are *derived*

- Information in the view created by the user:
 - Attributes used to group data.
 - Attribute used for color.
 - Attribute use for size.

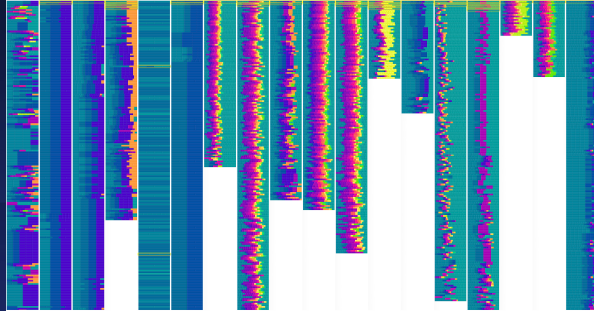


The attributes for X and Y-position were used, other attributes, such as **Speed**, The **shopping cart number** or **time of the day** were not visualized, and thus **not used** in our **derived** distance function.

User test

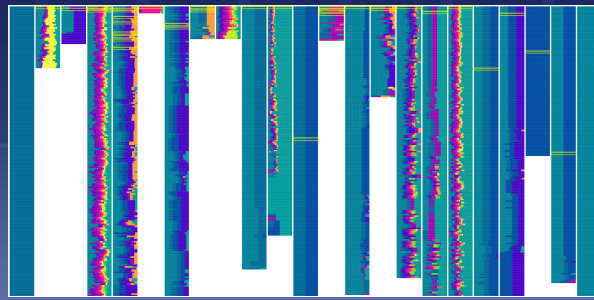
Dislikes

- Resize group
- Change order



Likes

- Dragging data
- Lead-Follow Metaphor



Application to education data

A most disappointing demo



Conclusions

- Interactive grouping is promising
- We are just starting with applications in education domain.
- Applications should help solve business issues
 - Prevent pupils from dropping out...?
 - Prevent pupils from doubling classes...?
 - From choosing the wrong program...?

Milou



Carlijn



Sasha Ilja

