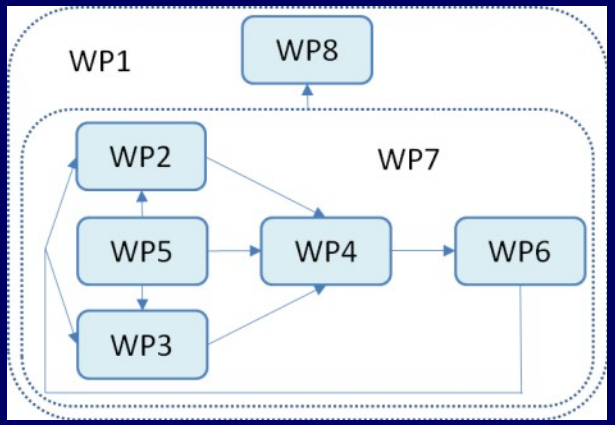
WP4 : Data visualization

Automatic Dashboard Generation

Praveen Soni[1], Cyril de Runz[1],
Fatma Bouali[1,2], Gilles Venturini[1],

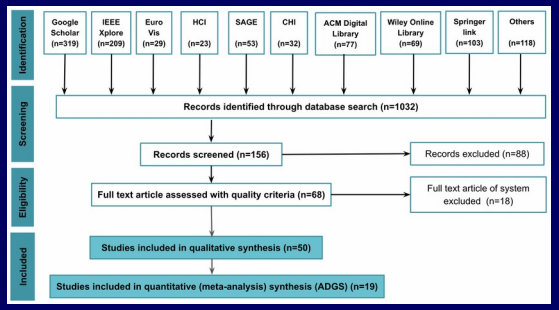
[1] LIFAT, University of Tours, Computer Science Lab., France
[2] University of Lille, France

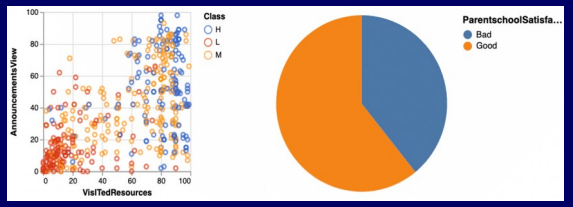
20/06/2023

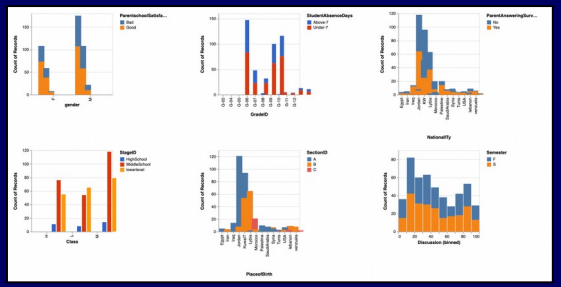


Talk overview

- Objectives
- State of the art
- Our proposal
  - Genetic representation of dashboards
  - Genetic operators
  - Fitness function
- First results
- Perspectives







Objectives

- Main goal: provide a tool to help novice users create dashboards

- Novice users requirements:
  - Explore/discover data (personnal, open, …)
  - Free tools, possibly on-line

  - Basic knowledge about BI domain (simple terminology, easy to learn UI)

  - Basic knowledge about visualizations
  - Difficulty for: data selection, vis. selection, data to vis. mapping
  - Assistance to create a visualization … dashboards (even more difficult)

- Integration in the BI4People platform

Grammel, L., Tory, M., & Storey, M. A. (2010). How information visualization novices construct visualizations. IEEE transactions on visualization and computer graphics, 16(6), 943-952.

State of the art: scope

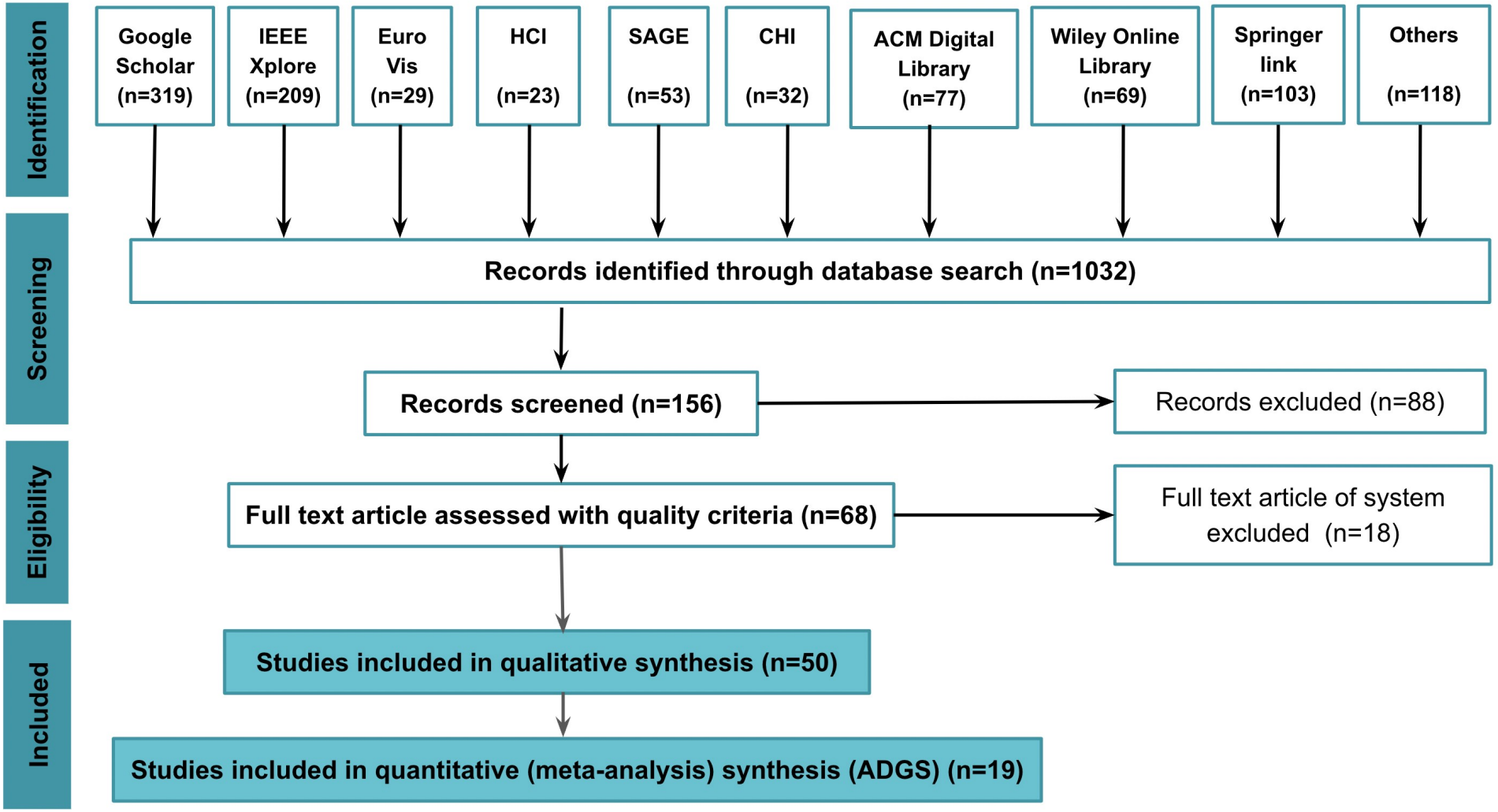
- User assistants for generating dashboards

- Definition of dasboard (consensus ?)
  - multiple visualizations interface
  - Possibly with linkage between visualizations
  - Intended to monitor/explore data, perform story telling?

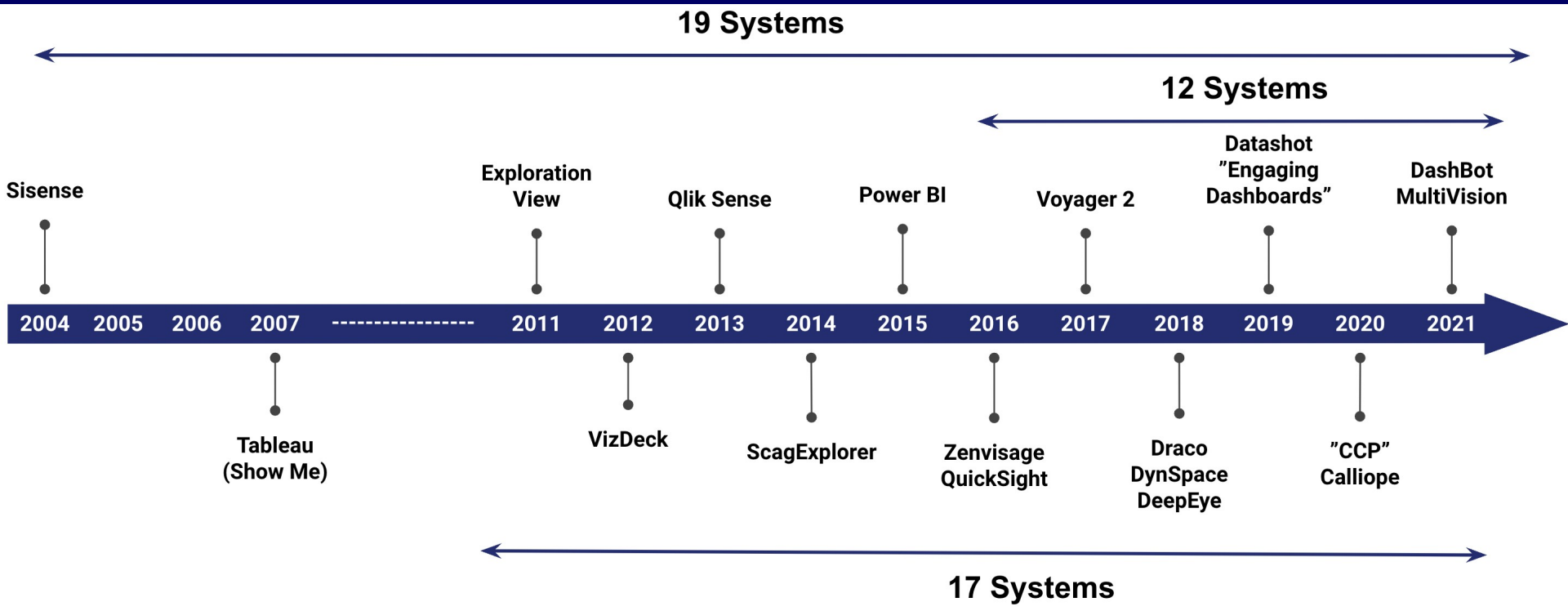
- Automatic dashboard generation
  - Must propose minimal conceptual design choices
  - Exclusion of systems that do not make suggestions (QualDash, LADV, …)
  - Exclusion of single visualization assistants

- Paper submited to *Visual Informatics* (2nd review): A survey on Automatic Dashboard Recommendation Systems, Praveen Soni, Cyril de Runz, Fatma Bouali, Gilles Venturini

State of the art: scope



State of the art: included systems



State of the art: scope

- Analysis along several main dimensions:

1 Work origin
  - System’s name, year and reference

2 Suggestion of visualizations
  - Considered visualizations
  - Suggestion method
  - Utility functions

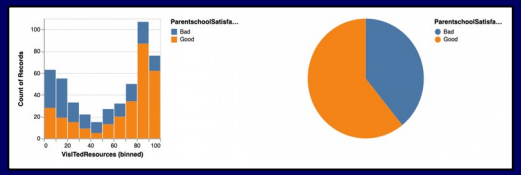
3 Layout

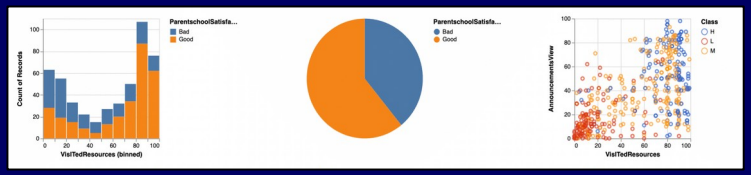
4 User’s viewpoint
  - User feedback
  - User interface and evaluation

5 Licence

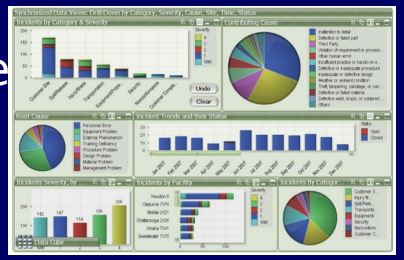
State of the art: suggestion methods

- One by one strategy (except for ScagExplorer)





- Simple suggestion methods:
  - Direct matching with predefined models (from expertise
  - A1 (num), A2 (num) -> Scatter plot(A1 -> X, A2 -> Y)



"Engaging Dashboards"

  - Predefined templates (from expertise)

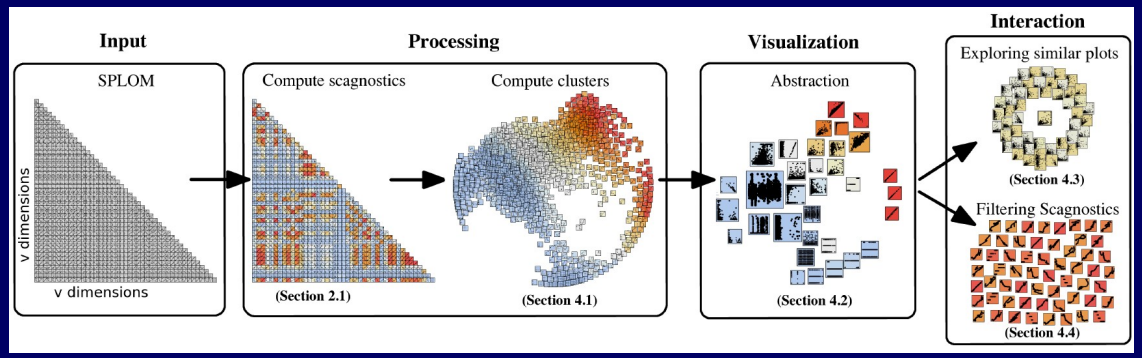
Exploration views"

State of the art: suggestion methods

- Advanced methods (statistics, knowledge-based):
  - Templates from large usage history:

CCP

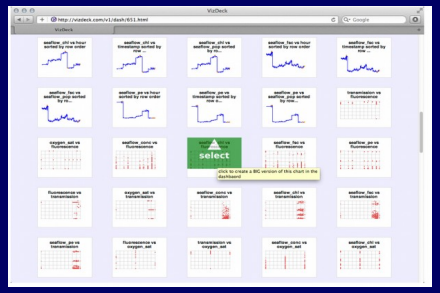
  - ScagExplorer, performs global optimization



  - Visual query languages (Zenvisage, Voyager 2) and constraint programming (Draco, …)

State of the art: suggestion methods

- Complex methods (machine learning):
  - Decision trees and behavioral feedback, or reinforcement learning

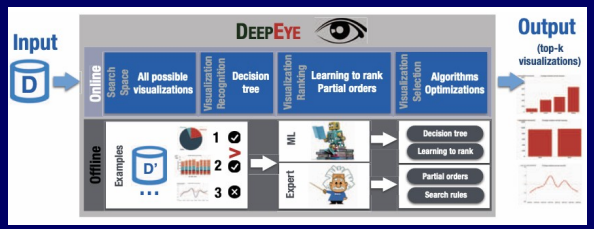


VizDeck

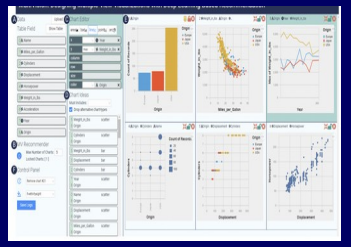
DashBot

  - Data intensive

DataShot, Calliope



DeepEye

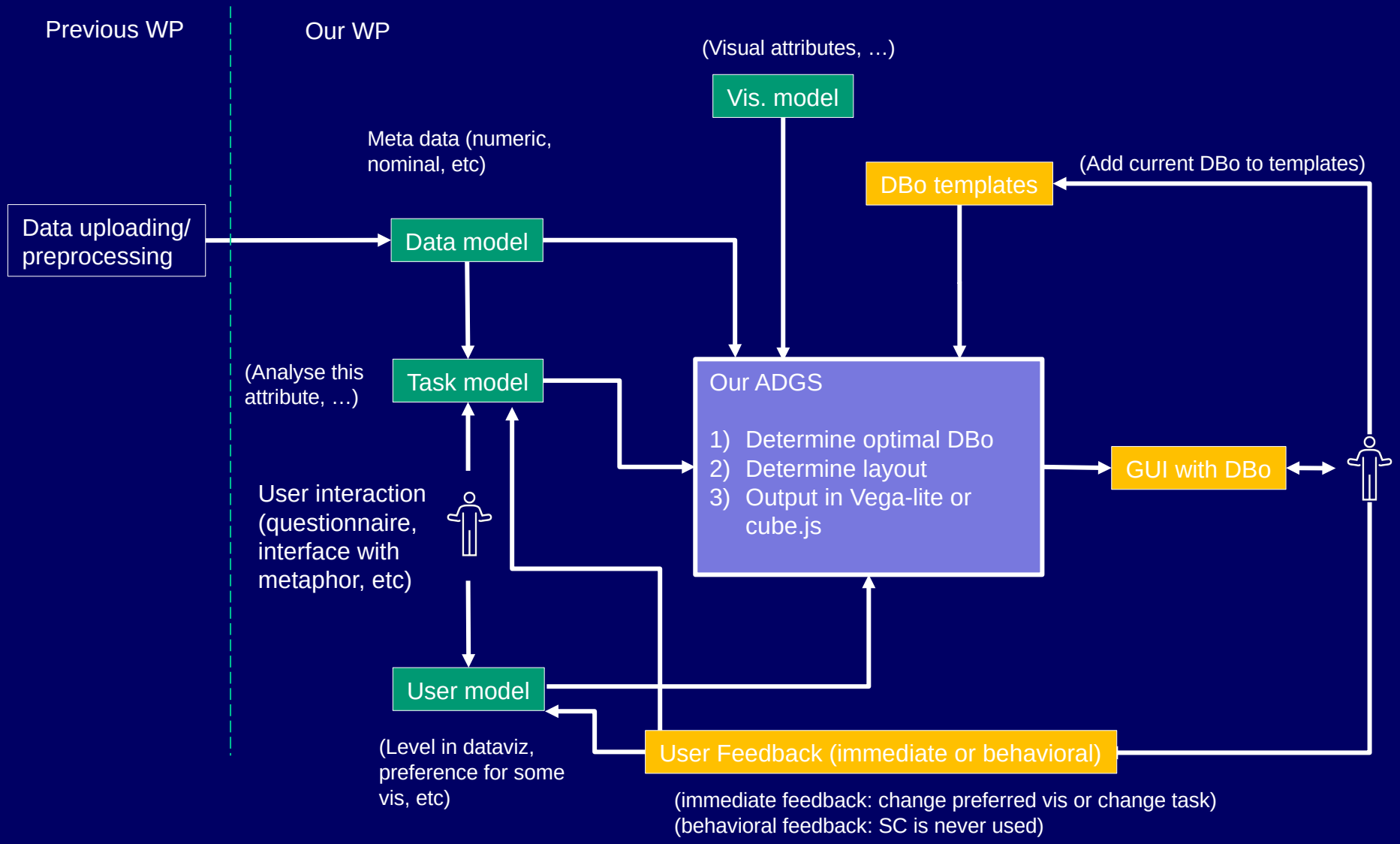


MultiVision
(deep learning)

State of the art: findings

- Weak points in existing systems:
  - No global optimization of dashboards
  - Limited (or absent) user feedback
  - Manual or predefined layout (grid, etc ...)
- Not so many systems for automatic dashboard generation
- Lack of comparison between systems

Our proposal: overview

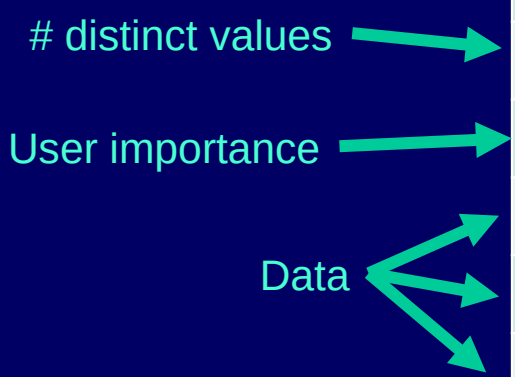


Data and task models

- Provided by previous WP
- Fixed (does not evolve)

- Representation:
  - List of attributes with type, number of values, etc

- Example :



A1	A2	A3	A4
Numeric	Numeric	Ordinal	Ordinal
230	210	6	10
50	80	100	20
10	5	A	C
2	6	B	D
…	…	…	…

Visualizations model

- List of visual attributes with their type, maximum values, …
- Complexity score
- Efficiency score (matching)

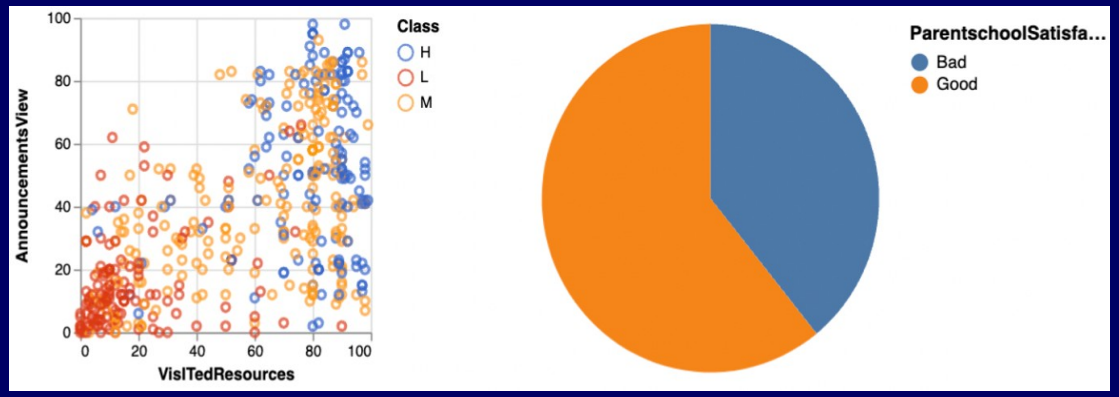
- Example: Scatter plot = (X,Y:position ; C:color hue)
- Complexity = 40 (over 100)
- Efficiency score =

imp_value	Numeric	Ordinal	Nominal
100	Position	Position	Position
95	Length/Height	Density	ColorHue
90	Angle	ColorSaturation	Texture
85	Slope	ColorHue	Connection
80	Area/Size	Texture	Containment
75	Volume	Connection	Density
70	Density	Containment	ColorSaturation
65	ColorSaturation	Length/Height	Shape
60	ColorHue	Angle	Length/Height
55	Texture	Slope	Angle
50	Connection	Area/Size	Slope
45	Containment	Volume	Area/Size
40	Shape	Shape	Volume

Mackinlay, J., 1986. Automating the design of graphical presentations of relational information. ACM Trans. Graph. 5, 110–141.

Dashboard genetic representation

- Use of a genetic algorithm: solution representation, genetic operators, evaluation



Scatter plot

Pie chart

Vis.	A1 (num)	A2 (num)	A3 (nomi)	A4 (nomi)
SP	X	Y		C
PC			X	

Visual attributes

Crosover operator

Vis.	A1 (num)	A2 (num)	A3 (nomi)	A4 (nomi)
SC	X	Y		C
PC				X

Vis.	A1 (num)	A2 (num)	A3 (nomi)	A4 (nomi)
AB	X	Y		C
CD				X
XY		A	B	
WZ	A	C		X

Cutting point

Vis.	A1 (num)	A2 (num)	A3 (nomi)	A4 (nomi)
SC	X	Y		C
CD				X
XY		A	B	
WZ	A	C		X

Vis.	A1 (num)	A2 (num)	A3 (nomi)	A4 (nomi)
AB	X	Y		C
PC				X

Mutation

- Random changes in a dashboard:
  - Add/remove one visualization

Vis.	A1 (num)	A2 (num)	A3 (nomi)	A4 (nomi)
SC	X	Y		C
PC				X
BC			X	

  - Change the mapping between visual attributes and data attributes

Vis.	A1 (num)	A2 (num)	A3 (nomi)	A4 (nomi)
SC	X	Y	C	
PC				X

Fitness function

- What is a good dasboard?

1. Low complexity: includes simple visualizations
2. High coverage: includes many important attributes
3. High matching quality: efficient visual rendering of data attributes

Complexity scores of included vis.

User importance scores of included attributes

Match scores from Mackinlay's Matrix

- Fitness = combination of three criteria

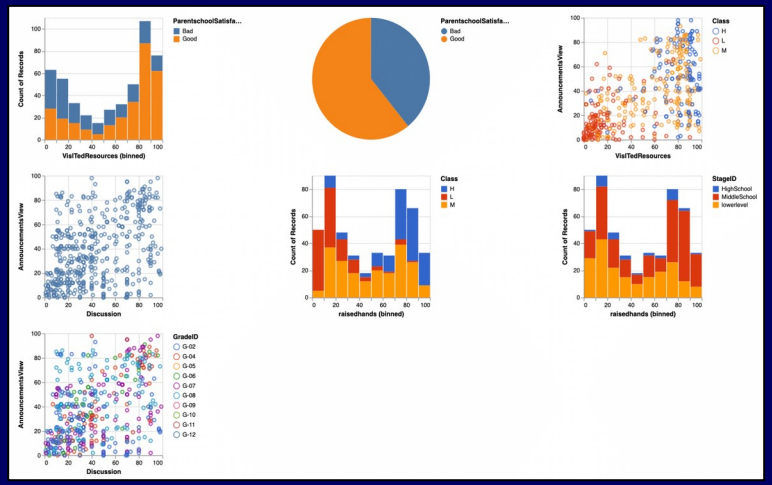
$$F(\text{Dashboard}) = -\text{Cplx} + \text{Covimp} \times \text{Match}$$

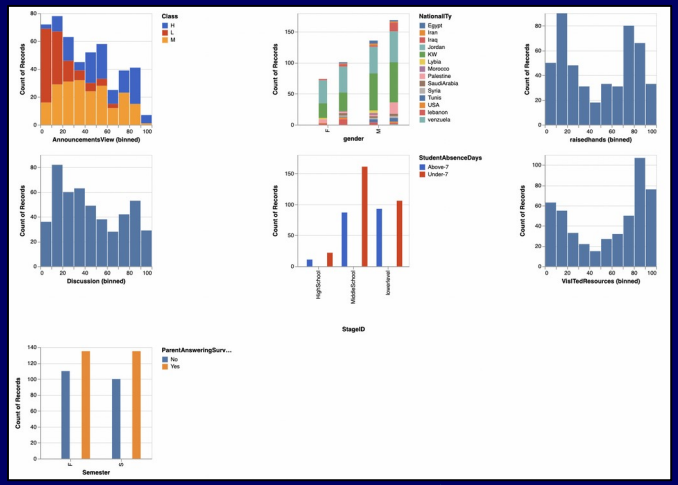
Population of dashboards

Dashboards	vis	sepal_length	sepal_width	petal_length	petal_width	species
DBO1	BCSO	X				C
DBO1	SP	X	L			
DBO2	BC		X			C
DBO2	BCCN	X	L			
DBO3	BC			X		C
DBO3	SP	X	L			
DBO4	DN				X	C
DBO4	SP	X	L			
DBO5	BCO	X				C
DBO5	HT		X	L		
DBO6	HT		L	X		
DBO7	BCSO	X		L		C
DBO8	BCCN					X

Results

- Grid search for best GA parameters
- GA is very fast: 20K generations with 20 dashboards in 1.4s
- Improves the quality of dashboards for several scenarios





Initial population

	Complexity	Coverage	Match	Fitness
DBo (Initial)	0.571	0.687	0.977	0.099
DBo (Final)	0.171	0.928	0.970	0.729

Final population

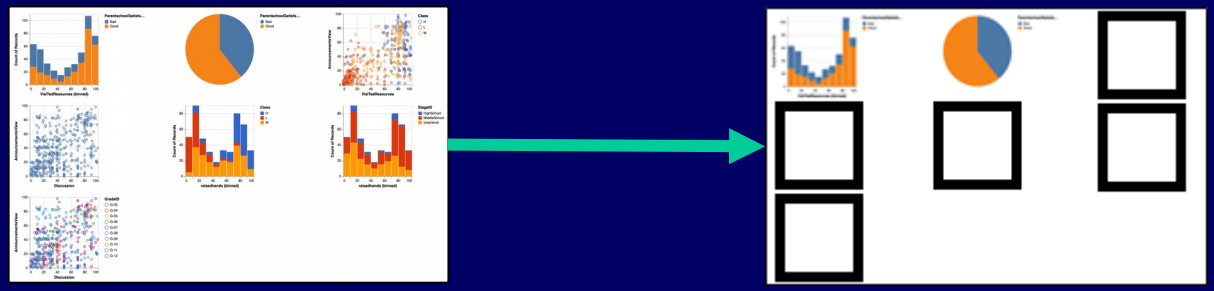
Results

- Additional work currently under study for this GA:
  - Improve the fitness function and operators
  - Add more visualizations
  - Adjust models and heuristics
  - Simple testing interface

- First conference paper to appear: Praveen Soni, Cyril de Runz, Fatma Bouali, Gilles Venturini, A genetic algorithm for automatic dashboard generation: first results, 27 International Conference Information Visualisation, 25-28 July 2023, Tampere University, Finland

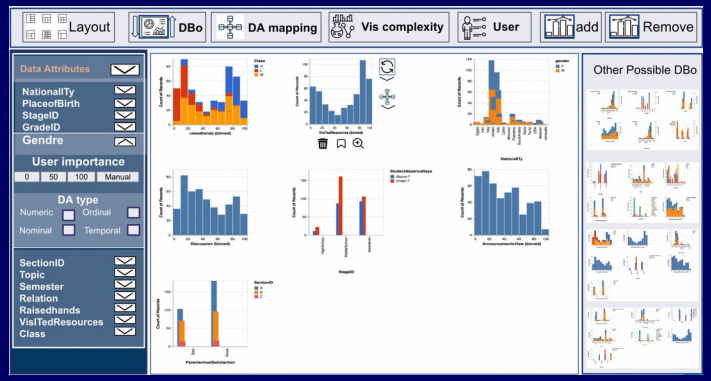
Perspectives: user feedback

- GA => great opportunities for user feedback
  - Fast + can cope with many different types of user feedback
  - Definition of « negative » use cases: *“Some vis are ok and I want to keep them, but I want to delete the others and replace them with other vis.“*



  - Definition of « positive » use cases: *“This dashboard is great.“* => Add dashboard to template

- Implementation and test of UI
- User evaluations



Perspectives: layout

What is a good layout?

- Low complexity: not too many visualizations
- Efficient use of computer screen
- Adapted to each visualization: minimum size, aspect ratio is important
- Place next to each other visualizations that share common dimensions
- C… Hconcat and Vconcat c…

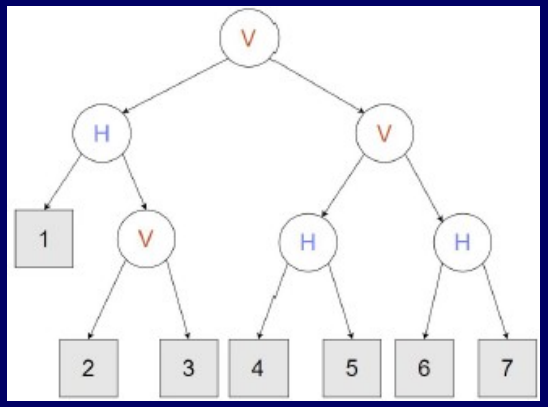
Table 2. Top 10 layouts: numbers (green bars) and percentages (blue bar) in VAST, InfoVis, SciVis, EuroVis, and PacificVis.

Layout	VAST	InfoVis	SciVis	EuroVis	PacificVis	Total
2A	17 8.6%	18 36.7%	6 22.2%	7 14.9%	2 5.0%	50 13.9%
3C	17 8.6%	8 16.3%	1 3.7%	8 17.0%	2 5.0%	36 10.0%
3A	16 8.1%	4 8.2%	0 0.0%	1 2.1%	2 5.0%	23 6.4%
3B	6 3.0%	5 10.2%	4 14.8%	2 4.3%	2 5.0%	19 5.3%
4E	10 5.1%	0 0.0%	1 3.7%	2 4.3%	4 10.0%	17 4.7%
3F	6 3.0%	1 2.0%	0 0.0%	4 8.5%	0 0.0%	11 3.1%
3E	6 3.0%	2 4.1%	0 0.0%	1 2.1%	1 2.5%	10 2.8%
4H	4 2.0%	1 2.0%	1 3.7%	2 4.3%	1 2.5%	9 2.5%
4C	4 2.0%	1 2.0%	1 3.7%	1 2.1%	1 2.5%	8 2.2%
2B	1 0.5%	2 4.1%	1 3.7%	2 4.3%	1 2.5%	7 1.9%
Total	87 44.2%	42 85.7%	15 55.6%	30 63.8%	16 40.0%	

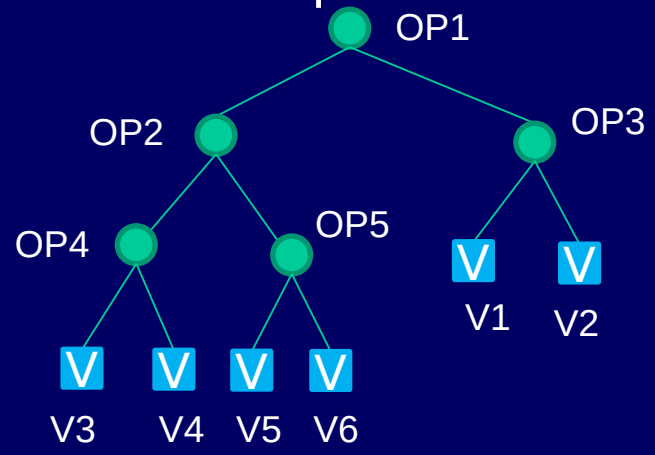
Chen, X., Zeng, W., Lin, Y., Ai-Maneea, H. M., Roberts, J., & Chang, R. (2020). Composition and configuration patterns in multiple-view visualizations. IEEE Transactions on Visualization and Computer Graphics, 27(2), 1514-1524.

Perspectives: layout

- Layout determined by slicing trees in Vega lite (inspired from Treemaps)



- Can be optimized with a GA



OP1	OP2	OP3	OP 4	OP 5	V1	V2	V3	V4	V5	V6

Many thanks, questions ?