Database / Data Mining Visualization

DataJewel: Tightly Integrating Visualization with Temporal Data Mining.

Mihael Ankerst, David H. Jones, Anne Kao, Changzhou Wang. ICDM Workshop on Visual Data Mining, Melbourne, FL, 2003

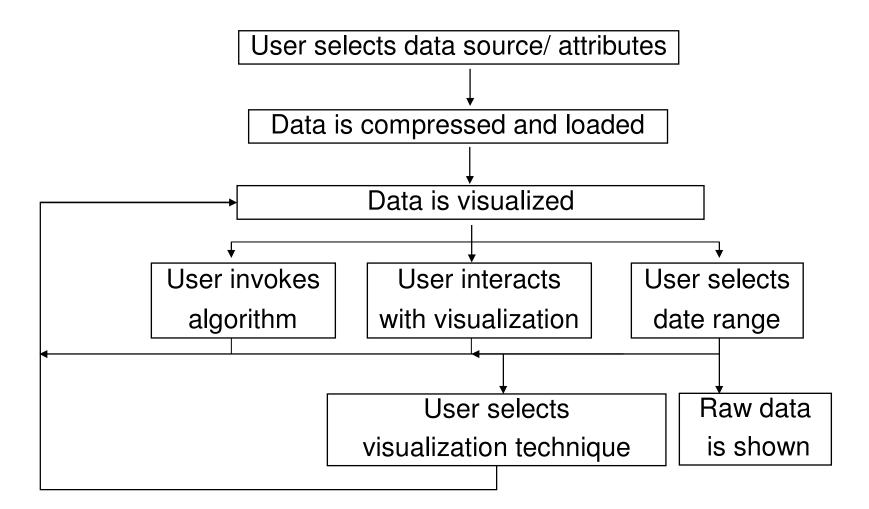
What is Data Mining ?

- Data mining, also known as knowledgediscovery in databases (KDD), is the practice of automatically searching large stores of data for patterns.
- data mining uses computational techniques from statistics and pattern recognition.

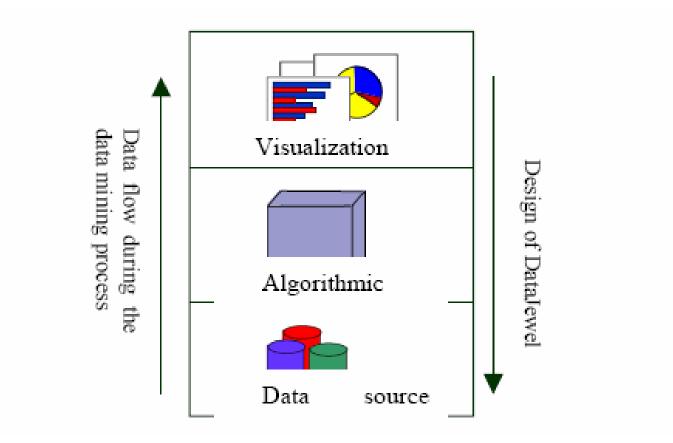
Temporal Data Mining

- Each record has a timestamp
- Databases evolve as a consequence of organizational need
- Inking together two databases with respect to time can give us a powerful tool to explore the union of attributes

User-centric data mining







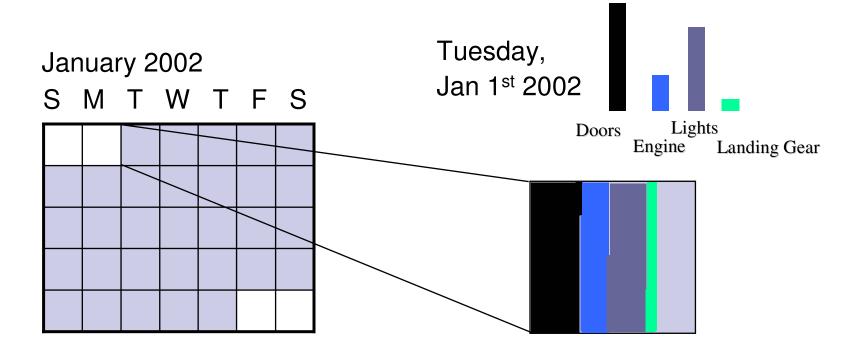
The Visualization Component

Calendar View

- □ Visual metaphor: Calendar.
- Structure of data is represented along the event dates is the frequency of events.
- Designed for domain experts intuitive and versatile design
- If there are few events the visualization is powerful since human's pre-attentive perception is very efficient in looking for variety of patterns

The Visualization Component

Time	Event type	Location	
09/11/2001	Door broken	Seattle	
09/12/2001			



The Visualization Component interaction

- Selection subset of dates
- Ascending/descending order frequency
- Interactive color assignment
- Zooming
- Detail on demand

The Temporal Mining Component

- Have algorithms that discover patterns
- Determine which events are involved in the patterns
- Automatically select colors based on the patterns
- Visualize not just data but also patterns
- Use of the same color assignment interface by user and algorithm.

The Temporal Mining Component

Discover one event of one event attribute
For example - highest variance, most interesting trend
give the event a unique color

Discover multiple events of one event attribute

 Set of events that together represent a pattern (for example - discovery of similar events) - each event that is part of the pattern receives a distinct color

Discover one event for each event attribute

Look for patterns relating event attributes to each other instead of analyzing them separately. (for example – finding similar events across different event attributes) – update the color assignments of each event attribute accordingly.

The Database component

- Each event is stored in one record
- Data resides in tables in one or more relational databases
- Aggregate database events according to event date (using select count(*) ... group by ...)
- Access the raw data of all attributes

	🙀 Categories for Airline1_767.ATA		×
-	Category name	Frequency Category co	blor
	05 - MAINTENANCE CHECKS	374	
	21 - AIR CONDITIONING/PRESSURIZATION	10381	
	22 - AUTO FLIGHT	9161	
	23 - COMMUNICATIONS	50226	
DataJev	23 - ELECTRICAL POWER	8160	jorithm
	25 - EQUIPMENT/FURNISHINGS	82968	_ / · · · · · · · · · · · · · · · · · ·
	26 - FIRE PROTECTION	4685	
ataJewel	27 - FLIGHT CONTROLS	7725	
	28 - FUEL	9382	
	29 - HYDRALIC POWER	3959	
	30 - ICE AND RAIN PROTECTION	1845	1998
May June	31 - INDICATING/RECORDING SYSTEM	6037	December January
May June	32 - LANDING GEAR	30678	Airline1_767.A
	33 - LIGHTS	29777	
	34 - NAVIGATION	21238	
	35 - OXYGEN	8087	
	36 - PNEUMATICS	5327	
	38 - WATER/WASTE	10649	
	49 - AIRBORNE AUXILIARY POWER	10737	
	51 - STRUCTURES	800	
Attribute :	52 - DOORS	6189	
Noluo :	53 - FUSELAGE	1484	
	54 - NACELLES/PYLONS	1016	
Date .	55 - STABILIZERS	171	
	56 - WINDOWS	2496	
Lettinodoc batton for octobilon. Fight p	57 - WINGS	1153	
	71 - POWER PLANT	3398	
	72 - ENGINE GENERAL	12929	
	73 - ENGINE FUEL/CONTROL	4568	
	74 - ENGINE IGNITION	890	
	75 - ENGINE AIR	916	
	76 - ENGINE CONTROL	925	
	77 - ENGINE INDICATING	4347	
	78 - ENGINE EXHAUST	4979	
	79 - ENGINE OIL	5633	
	80 - ENGINE STARTING	1731	
	OK Reset	Cancel	

DataJewel – Scenario: Mining Algorithm

Me Da	ataJewel										
<u>F</u> ile	<u>O</u> ptions	Operatio	ns <u>T</u> ools	<u>V</u> iew	<u>H</u> elp						
2	May		June		July	August	September	October	November	December	1998 January Airline1_767.ATA
	51 SU 30 50 SA 27 50 SI 50 70 SI 50	14 70 91 97 7 96 91 91 91 92 91 91 92 91 91 92 91 91 92 91	90 N 20 7 71 77 81 20 9 75 77 91 70 9 71 92 10 1 7 7	1 91 91 1 91 91 7 93 91 7 9 7			NA GU 21/ NO 21 AU 10 31 54 50 30 30 51 AL 30 50 50 50 51 41 51 51 51 50 50 51 51 50 50 51 54 50 51				
	•										
	Attribute : Value : Date :										
	Left mou:	se button f	or selectior	i. Right	mouse button for a	hange of color ma	ipping.				
	Press here for running mining algorithm										

DataJewel – Scenario: Mining Algorithm

🕌 Da	taJewel										_ _ _ ×
<u>F</u> ile	<u>O</u> ptions	Operatio	ns <u>T</u> ool	ls <u>V</u> iew	<u>H</u> elp						
20	May		June		July	August	September	October	November	December	1998 January Airline1_767.ATA
	•							3333333333			
	Attribute : Value : Date :				(V - Longe	st Deviation] Attribu	ite: Airline1_767.A1	"A Category: 49 - A	IRBORNE AUXILIAI	RY POWER	
	Left mous	se button f	or selecti	ion. Right	mouse button for	change of color ma	ipping.				

	🚰 Categories for Airline1_767.ATA		×	1	
-	Category name	Frequency	Category color	I	
	05 - MAINTENANCE CHECKS	374		1	
	21 - AIR CONDITIONING/PRESSURIZATION	10381		1	
	22 - AUTO FLIGHT	9161		1	
Data la	23 - COMMUNICATIONS	50226		┝╴╸╶╬┹┠╸╴	~~
DataJe	23 - COMMUNICATIONS 24 - ELECTRICAL POWER	8160		jorithi	n
	25 - EQUIPMENT/FURNISHINGS	82968		,	
	26 - FIRE PROTECTION	4685			
ataJewel	27 - FLIGHT CONTROLS	7725			_
Options Operations Tools View	28 - FUEL	9382			
	29 - HYDRALIC POWER	3959			
	30 - ICE AND RAIN PROTECTION	1845		1	1998
May June	31 - INDICATING/RECORDING SYSTEM	6037		December	January
indy cane	32 - LANDING GEAR	30678			Airline1_767.AT
	33 - LIGHTS	29777			
	34 - NAVIGATION	21238			
	35 - OXYGEN	8087			
	36 - PNEUMATICS	5327			
	38 - WATER/WASTE	10649			
•	49 - AIRBORNE AUXILIARY POWER	10737			
	51 - STRUCTURES	800		LIARY POWER	
Attribute :	52 - DOORS	6189			
Value :	53 - FUSELAGE	1484			
Date :	54 - NACELLES/PYLONS	1016			
	55 - STABILIZERS	171			
Left mouse button for selection. Right	56 - WINDOWS	2496			
,	57 - WINGS	1153			
	71 - POWER PLANT	3398			
	72 - ENGINE GENERAL	12929			
	73 - ENGINE FUEL/CONTROL	4568			
	74 - ENGINE IGNITION	890			
	75 - ENGINE AIR	916			
	76 - ENGINE CONTROL	925			
	77 - ENGINE INDICATING	4347			
	78 - ENGINE EXHAUST	4979			
	79 - ENGINE OIL	5633			
	80 - ENGINE STARTING	1731			

_	Categories for Airline1_767.ATA		×		
- í	Category name	Frequency	Category color		
	25 - EQUIPMENT/FURNISHINGS	82968		Ì	
	23 - COMMUNICATIONS	50226		1	
	32 - LANDING GEAR	30678		1	
	33 - LIGHTS	29777			
DataJe	34 - NAVIGATION	21238		jorithi	m
	72 - ENGINE GENERAL	12929		2	
	49 - AIRBORNE AUXILIARY POWER	10737			
ataJewel	38 - WATER/WASTE	10649			_
Options Operations Tools View	21 - AIR CONDITIONING/PRESSURIZATION	10381			
	28 - FUEL	9382			
	22 - AUTO FLIGHT	9161			1998
May June	24 - ELECTRICAL POWER	8160		December	January
may ounc	35 - OXYGEN	8087		December	Airline1_767.A
	27 - FLIGHT CONTROLS	7725			
	52 - DOORS	6189			
	31 - INDICATING/RECORDING SYSTEM	6037			
	79 - ENGINE OIL	5633			
	36 - PNEUMATICS	5327			
•	78 - ENGINE EXHAUST	4979			
	26 - FIRE PROTECTION	4685		LIARY POWER	
Attribute :	73 - ENGINE FUEL/CONTROL	4568		ENTER OTTER	
Value :	77 - ENGINE INDICATING	4347			
Date :	29 - HYDRALIC POWER	3959			
	71 - POWER PLANT	3398			
Left mouse button for selection. Right	56 - WINDOWS	2496			
,	30 - ICE AND RAIN PROTECTION	1845			
	80 - ENGINE STARTING	1731			
	53 - FUSELAGE	1484			
	57 - WINGS	1153		1	
	54 - NACELLES/PYLONS	1016			
	76 - ENGINE CONTROL	925			
E CONTRACTOR OF CONTRACTOR	75 - ENGINE AIR	916			
	74 - ENGINE IGNITION	890			
E Contraction of the second	51 - STRUCTURES	800			
	05 - MAINTENANCE CHECKS	374			
	55 - STABILIZERS	171			

۰.	10							
	-			Categories for Airline1_767.ATA	-	×	1	
				Category name	Frequency	Category color		
				25 - EQUIPMENT/FURNISHINGS	82968			
				23 - COMMUNICATIONS	50226			
				32 - LANDING GEAR	30678			
			+-7-	33 - LIGHTS	29777		ho atio	
		νd	laje	33 - LIGHTS 34 - NAVIGATION	21238		ractio	
		_		72 - ENGINE GENERAL	12929			
				49 - AIRBORNE AUXILIARY POWER	10737			
Da	taJewel			38 - WATER/WASTE	10649			
e	Options	Operations	Tools View	21 - AIR CONDITIONING/PRESSURIZATION	10381			
	201010	- En anono	1000 100	= 28 - FUEL	9382			
:				22 - AUTO FLIGHT	9161			1998
_	May	.10	ine	24 - ELECTRICAL POWER	8160		December	January
Мау				35 - OXYGEN	8087			Airline1_767.ATA
				27 - FLIGHT CONTROLS	7725			
				52 - DOORS	6189			
				31 - INDICATING/RECORDING SYSTEM	6037			
				79 - ENGINE OIL	5633			
				36 - PNEUMATICS	5327			
	•			78 - ENGINE EXHAUST	4979			
	<u> </u>	•		26 - FIRE PROTECTION	4685		LIARY POWER	
	Attribute :			73 - ENGINE FUEL/CONTROL	4568		LIARTFOWER	
	Value :			77 - ENGINE INDICATING	4347			
	Date :			29 - HYDRALIC POWER	3959			
	Date .			71 - POWER PLANT	3398			
	Loftmour	eo hutton for a	selection. Righ		2496			
			selection. Righ	30 - ICE AND RAIN PROTECTION	1845			
				80 - ENGINE STARTING	1731		1	
				53 - FUSELAGE	1484		1	
				57 - WINGS	1153			
				54 - NACELLES/PYLONS	1016			
				76 - ENGINE CONTROL	925			
				75 - ENGINE AIR	916			
				74 - ENGINE IGNITION	890			
				51 - STRUCTURES	800			
				05 - MAINTENANCE CHECKS	374			
				55 - STABILIZERS	171			
					171			
				OK Reset	Cancel			

DataJewel – Scenario: User Interaction

崔 Da	taJewel								_ _ ×
	Options Operatio	ns <u>T</u> ools <u>V</u> iew	<u>H</u> elp						
2	Мау	June	July	August	September	October	November	December	1998 January Airline1_767.ATA
	•					333333333			
	Attribute : Value : Date :								
	Left mouse button f	for selection. Right	mouse button for	change of color ma	ipping.				

DataJewel – Scenario: User Interaction

M D	ataJewel										_D×
<u>F</u> ile	<u>O</u> ptions	Operatio	ns <u>T</u> ool	s <u>V</u> iew	<u>H</u> elp						
2	Мау		June		July	August	September	October	November	December	2001 January Airline1_767.ATA
	•									300000000	
	Attribute : Value :										
	Date :					6		<u></u>			
	J Left mous	se button f	or selecti	on. Right	mouse button	for change of color ma	apping.				

DataJewel – Scenario: User Interaction

		崔 Selected Ra	w Data:					×
<mark>e</mark> Da	taJewel	Arms	s_Demo.Airline	e1_767				
File	Options O				AP_MFG	CMPLNT_		MAINT_RESOLN_TX
P 3		Jul 30, 2000	78 - ENGINE E			DEF: FIT NO.1 ENG REVERSER LOCK		
1.7		Jul 30, 2000		UEL/CONTR	25442	DEF: CARRY OUT SECOND INSPECTION	ON OF NO.2 ENGINE CORE W.	DEF: CARRY OUT 🚃
Q	May	Jul 30, 2000		UEL/CONTR	25832	DEF: ACTION EOD 767 732573X PART	A NO 1 ENG FUEL MANIFLOW	DEF: ACTION EOD
4		Jul 30, 2000	73 - ENGINE F	UEL/CONTR	24341	DEF:CARRY OUT REQMTS OF EOD 76	7-732573X NO 1 ENG FUEL M.	DEF:CARRY OUT R
		Jul 30, 2000	73 - ENGINE P	UEL/CONTR	24341	DEF:CARRY OUT SECOND INSPECTIC	IN OF FITMENT AND SECURI	. DEF:CARRY OUT S
		Jul 30, 2000	73 - ENGINE P	UEL/CONTR	29231	DEF:ACTION EOD-767-73273X PRT A N	NR 1 ENGINE FUEL MANIFOL	. DEF:ACTION EOD 👻
				11.				
			1.1					_
ļ	•						200200000	
	Attribute :						- 1 -	
	Value :							
	Date :							
	Left mouse I	button for selectio	n. Right mouse	button for chan	ge of color i	mapping.		

Critique (+)

- Combine data mining algorithms with visualization
- Can work with several databases
- Scalable handles large databases
- Intuitive and easy to use don't need a data mining expert

Critique (-)

- Hard to see patterns over weeks or months or within a single day
- Only one event attribute for each calendar presentation
- Not easily transferable to other domains like author claims.
- Only for categorical attributes
- Does not handle other types of databases other than relational
- No user studies

DEVise: Integrated Querying and Visual Exploration of Large Datasets

Miron Livny, Raghu Ramakrishnan, Kevin Beyer, Guangshun Chen, Donko Donjerkovic, Shilpa Lawande, Jussi Myllymaki, and Kent Wenger. Proc. .SIGMOD 1997

What is DEVise?

- A data exploration system that allows users to develop, browse, and share visual representations of datasets from several sources.
- A framework which describes a set of querying and visualization primitives that is combined to develop a visual presentation.

Basic concepts

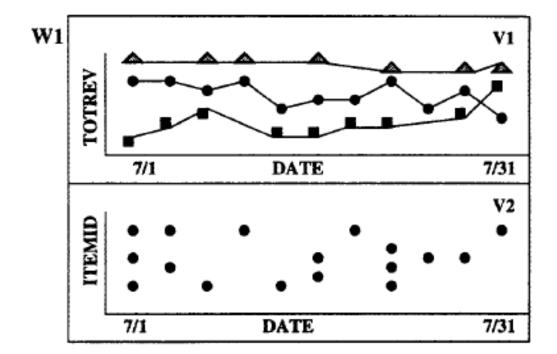
- Mapping each source data record to a visual symbol on screen
- **TData** (Textual Data) a collection of records with one or more attributes (along with a schema).
- <u>GData</u> (Graphical Data) high level representation of the screen (x, y, size, color, pattern, orientation, shape
- <u>Mapping</u> a function that is applied to the TData record to produce a GData record.

Basic concepts - presentation

View – basic display unit

- TData
- mapping
- □ Background (title, axes)
- data display
- cursor display additional data independent information
- visual filter set of selection (a query) on the GData of a view
- Window collection of views
- Visual presentation collection of windows

Visualization model

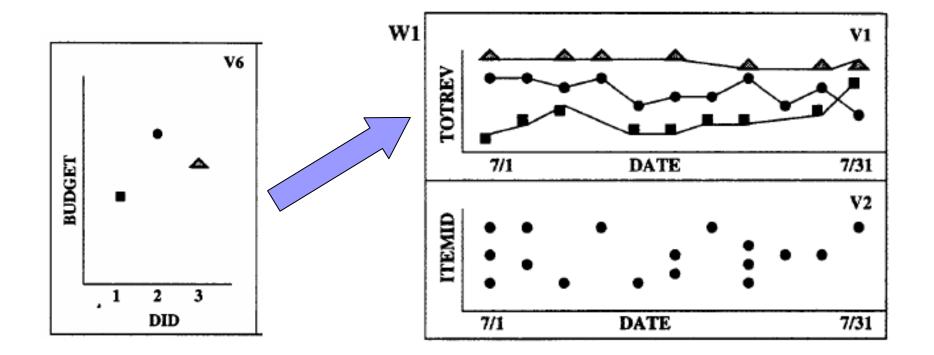


Overall_sales (date, Did, totRev) Sales (date, itemid, custid, number)

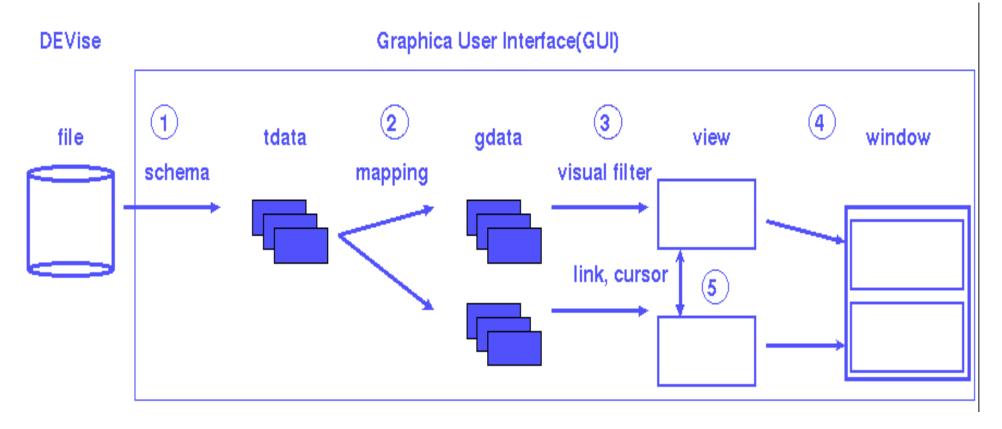
Some more concepts...

- Cursors allows the visual filter of one view to be seen as a highlight in another view
- Links constraints that allows the contents of two views to be coordinated.
 - Visual associate visual filters of two views
 - Record the projection of the data in one view (on the linked attributes) will act as a filter on the TData of the other view
 - Operator
 - aggregate

Record link example



DEVise Model



Semantics of a visual display

A mapping function is applied from the TData record to produce a Gdata record:

A view can then be represented as: (B, σ^G, μ, T, C)

B – Background

Sigma – visual filter

Mu – mapping

T – TData

C - cursor layer

Visual Queries and SQL

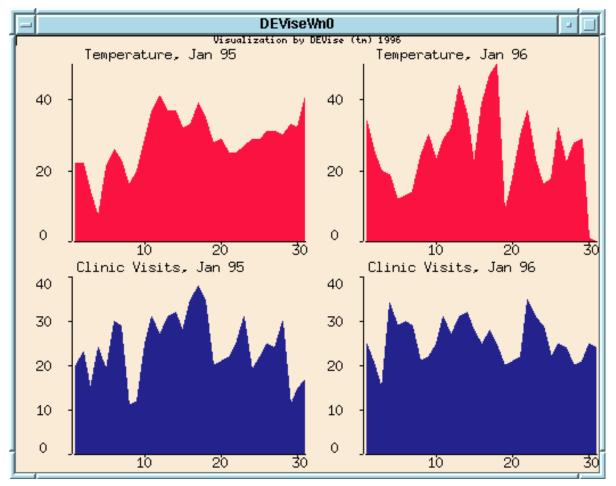
- Visual queries user selection on visual attributes of a view. (zoom in/out, scroll, point selection)
- Can save and transfer a visual query
- Enables users to generate sophisticated SQL queries through intuitive graphical operations
- Can be used as an SQL front-end (but not only!)

Achievements

- Visual presentation capabilities users can render their data. Simple mapping between data and presentation
- Ability to handle large distributed databases (not limited to available memory)
- Collaborative data analysis
- Support for interactively exploring the data visually at any level of detail

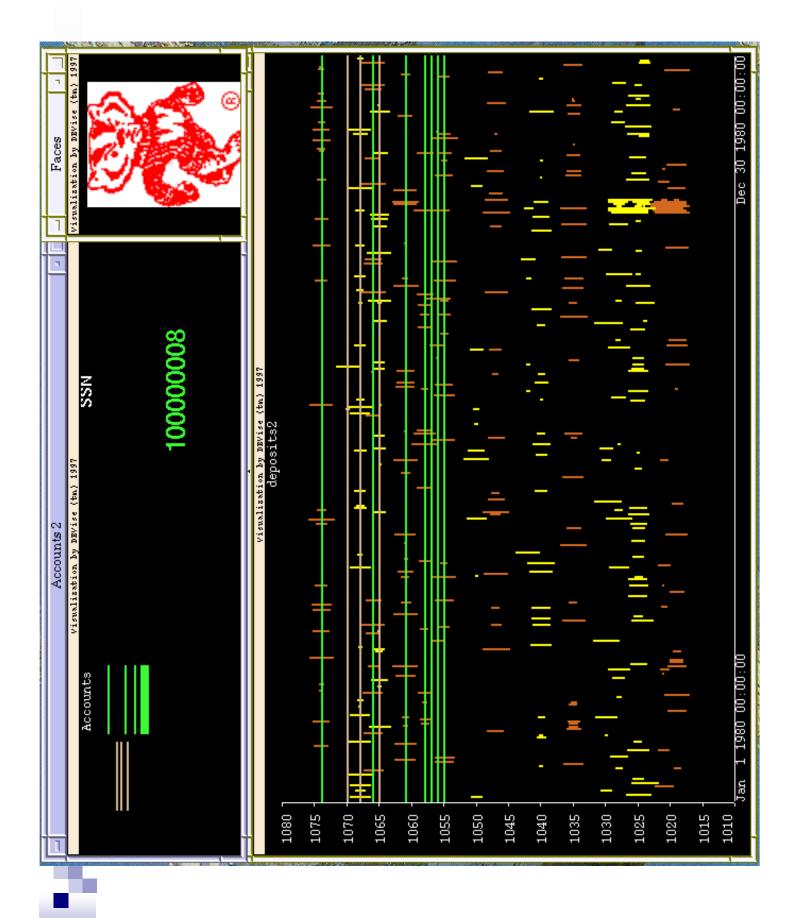
Example

Input two data sources: clinic information about number of visits, and information about temperature



Another Example:

- Input data: has information about deposits into various accounts at 2 different banks:
 Account (bankNum, SSN, accNum, pic, ...)
 Deposit (accNum, date, amount)
- problem: We want to analyze the transactions to find out who has a suspiciously large number of transactions within a short period of time.



critique

+

Very thorough well-defined framework
Many examples of implementations in real application

- Leaves the visualization decisions to the user (but that's the idea...)
- Some visualizations are very hard or impossible to do

