CSE 564: Visualization

Visual Analytics

Klaus Mueller

Computer Science Department

Stony Brook University

Explosion of Data

82,000 fingerprints are matched every day against INS database with 40 million records

100 million VISA credit card transactions per day

300 million phone long distance calls on ATT's network per day

7 million IP packets per second on DE-CIX backbone

→ there is *NO* chance to visualize all these data

Problems With Scalability

Must be scalable to

- number of data points
- number of dimensions
- data sources
- diversity of data sources
- number of users
- diversity of users and tasks
- quality of the data

Visual Analytics comes to the rescue…

The Goal of Visualization

Ease understanding of the data by providing an effective visual representation

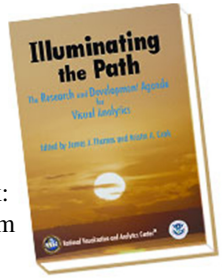
Amplify Perception

Detect the Expected, Discover the Unexpected™

What is Visual Analytics

Visualization plus…

- interaction (HCI)
- data processing (analytics)
- story telling
- scientific approach



Agenda setting book:
http://nvac.pnl.gov/agenda.stm

but also…

- intelligent computing (AI, machine learning)
- behavioral psychology (cognitive science, human factors)

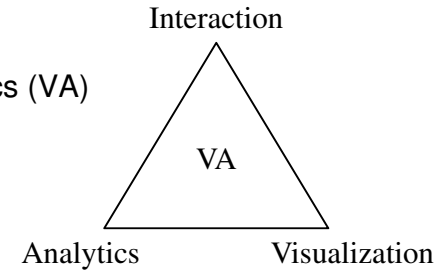
Visual Analytics is the science of analytical reasoning supported by a highly interactive visual interface

Visual Analytics Method

The Daniel Keim Mantra of Visual Analytics

"Analyze First - Show the Important – Zoom, Filter and Analyze Further - Details on Demand"

The triangle of Visual Analytics (VA)



Intelligence Analysis

Intelligence analysis is challenging

Huge amounts of data

Low signal vs. noise (SNR)

Many data types

- text, images, video, sensor data, etc.

Uncertainty

Contradictions

Omissions

Use of Visualization

Visual perception

- high bandwidth
- fast screening of a lot of data
- pattern recognition
- higher-level cognition

Interaction

- direct manipulation
- two-way communication

Recall intro lecture on the human visual system…

Use of Visualization

Visual perception

- high bandwidth
- fast screening of a lot of data
- pattern recognition
- higher-level cognition

Interaction

- direct manipulation
- two-way communication

Recall intro lecture on the human visual system…

But… humans are imperfect

Focus vs. Periphery

Humans tend to overlook/ignore non-focus (and unexpected) objects even when very close and obvious

- note the Visual Analytics slogan: *Detect the Unexpected*

Humans also have limited working memory

- fine details are quickly forgotten when focus changes
- big effect in animated or interactive visualizations
- need to preserve temporal context

Change Blindness

Thoroughly studied by Dan Simons (U Illinois)

- see http://www.dansimons.com/index.html

Videos

Change Blindness

Thoroughly studied by Dan Simons (U Illinois)

- see http://www.dansimons.com/index.html

Videos

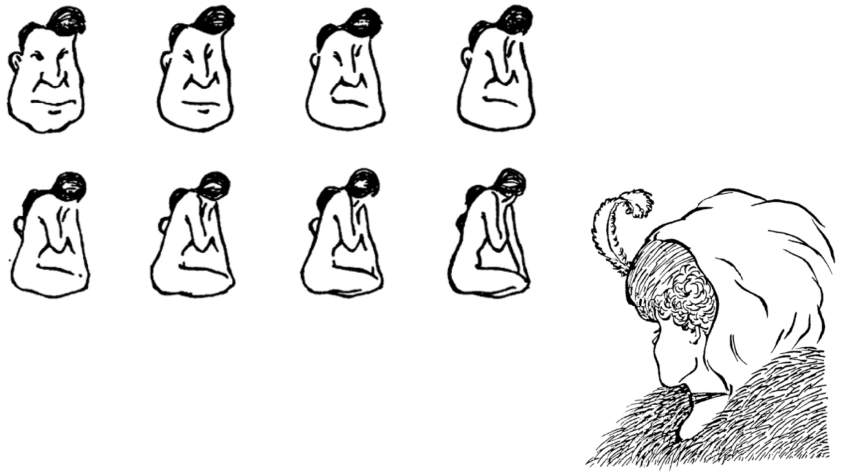
Visual Analytics tools

- help human analysts cope with insufficient memory
  → visualizations externalize memory
  → allow humans to perform *visual queries* (see C. Ware book)
- help human analysts deal with change blindness
  → analytics can detect changes
  → visualization can highlight/emphasize these changes

Persistence of Mindset (Priming)

Another deficiency of humans

- humans tend to stick with an “opinion” for a long time



Persistence of Mindset (Priming)

Another deficiency of humans

- humans tend to stick with an “opinion” for a long time

man/woman

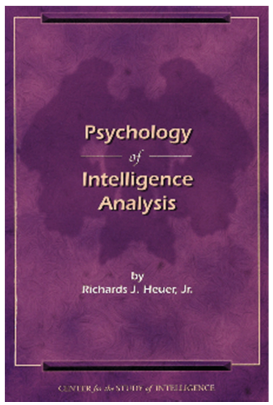
Young/old woman

Human Limitations

The Magic Number Seven

- ± 2 : the number of things most people can keep in working memory at one time
- causes problems for complicated analysis

Excellent book



the next slides follow it

Strategies for Dealing with Complexity

Decomposition

- decompose a complex problem into simpler problems
- get one’s thinking straight in these simpler problems

Externalization

- get the decomposed problem out of one’s head and down on paper or on a computer screen in some simplified form
- shows the main variables, parameters, or elements of the problem and how they relate to each other

Strategies for Dealing with Complexity

Decomposition

- decompose a complex problem into simpler problems
- get one's thinking straight in these simpler problems

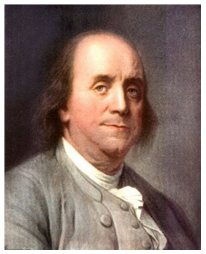
Externalization

- get the decomposed problem out of one's head and down on paper or on a computer screen in some simplified form
- shows the main variables, parameters, or elements of the problem and how they relate to each other

Recall principles of information visualization and visual analytics

- overview and detail
- focus and context
- analyze, filter, zoom,…

200 Years Ago… Benjamin Franklin's Letter.



Mentioned his method of solving decision problems

Why is the decision problem so difficult?

- folks cannot keep all pros and cons in mind at the same time

Solution?

- write down all the pros and cons onto paper in some visible, shorthand form
- allows you make a global judgment effectively

Visual analytics = many ☺

Multiattribute Utility Analysis

You want to choose the best car among various cars.

What is the best car?

- Lowest maintenance cost? Highest resale value? Slickest styling? Best gas mileage? Largest trunk space?

How to make a decision?

- Car purchase matrix

1. List the important attributes you want to maximize

Price

Maintenance Cost

Styling

Gas Mileage

Comfort

Handling

2. Quantify the relative importance of each attributes

Price	30%
Operating Cost	10%
Styling	20%
Comfort	20%
Handling	15%
Safety	5%
Total	100%

3. Identify the cars you are considering and judge each one ranks on each of attributes

	% Value	Car 1	Car 2	Car 3
Price	30%	3.5%	3.0%	3.5%
Operating Cost	10%	3.5%	2.0%	4.5%
Styling	20%	2.5%	4.5%	3.0%
Comfort	20%	4.0%	2.5%	3.5%
Handling	15%	3.0%	4.0%	3.0%
Safety	5%	3.5%	2.5%	4%

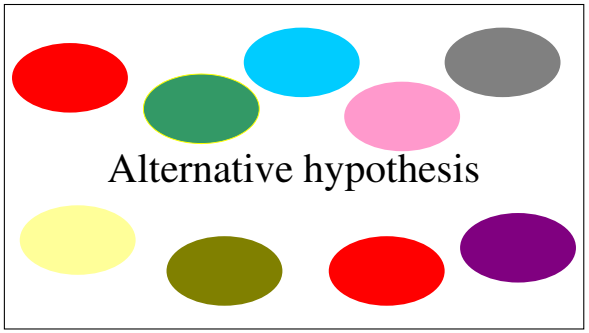
4. Multiply the percentage value by the value of each cars

	% Value	Car 1	Car 2	Car 3
Price	30%	105	90	105
Operating Cost	10%	35	20	45
Styling	20%	50	90	60
Comfort	20%	80	50	70
Handling	15%	45	60	45
Safety	5%	17.5	12.5	20
Totals		332.5	322.5	345

More Formally: Problem Description

When working on difficult intelligence issues

- which is the correct explanation?
- which is the most likely outcome?



Analysis of Competing Hypotheses (ACH)

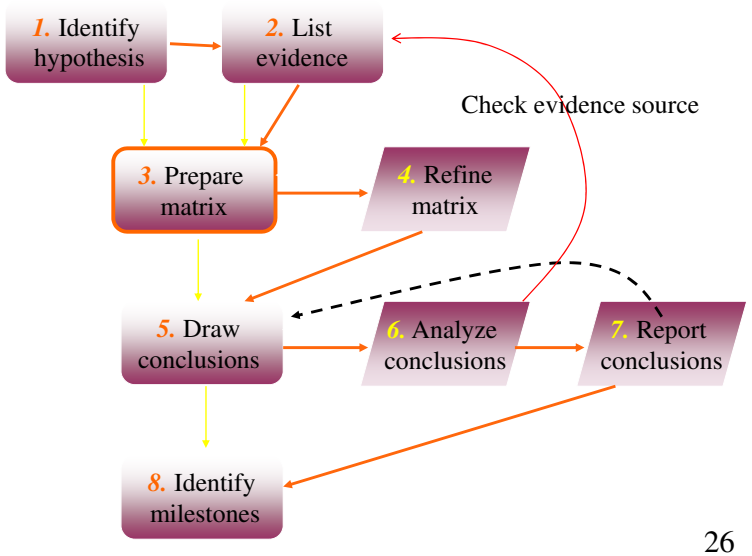
Used to

- aid judgment on important issues
- minimize cognitive limitations

Basic insights from

- cognitive psychology
- decision analysis
- scientific method

Eight-Step of ACH



Step 1: Identify Hypothesis

Hypothesis generation *vs.* hypothesis evaluation

- generation: bring together all possibilities
- evaluation: focus on

Disproved *vs.* unproven

- for a disproved hypothesis there is positive evidence that it is wrong
- for an unproven hypothesis, there is no evidence that it is correct

Step 2: List Evidence

Don't limit to the evidences current available

For each hypothesis, list supporting and contradicting factors

Absence and presence of evidence

- for example: If the dog barked in the night?
  no. nobody heard it barked (absence)

Step 3: Prepare Matrix

	H1	H2	H3
E1. Saddam public statement of intent not to retaliate.	+	+	+
E2. Absence of terrorist offensive during the 1991 Gulf War.	+	+	–
E3. Assumption that Iraq would not want to provoke another US attack.	+	+	–
E4. Increase in frequency/length of monitored Iraqi agent radio broadcasts.	–	+	+
E5. Iraqi embassies instructed to take increased security precautions.	–	+	+
E6. Assumption that failure to retaliate would be unacceptable loss of face for Saddam.	– –	+	+

Question: will Iraq Retaliate for US Bombing?

- H1: Iraq will not retaliate
- H2: It will sponsor some minor terrorist actions.
- H3: Iraq is planning a major terrorist attack, perhaps against one or more CIA installations

	H1	H2	H3
E1. Saddam public statement of intent not to retaliate.	+	+	+
E2. Absence of terrorist offensive during the 1991 Gulf War.	+	+	–
E3. Assumption that Iraq would not want to provoke another US attack.	+	+	–
E4. Increase in frequency/length of monitored Iraqi agent radio broadcasts.	–	+	+
E5. Iraqi embassies instructed to take increased security precautions.	–	+	+
E6. Assumption that failure to retaliate would be unacceptable loss of face for Saddam.	– –	+	+

Step 4: Refine Matrix

Diagnostic value – likeliness of hypothesis

- high-temp indicates sickness, but can't determine which illness

Reconsider the hypotheses

- add, or need finer distinction
- combine

Reconsider the evidences

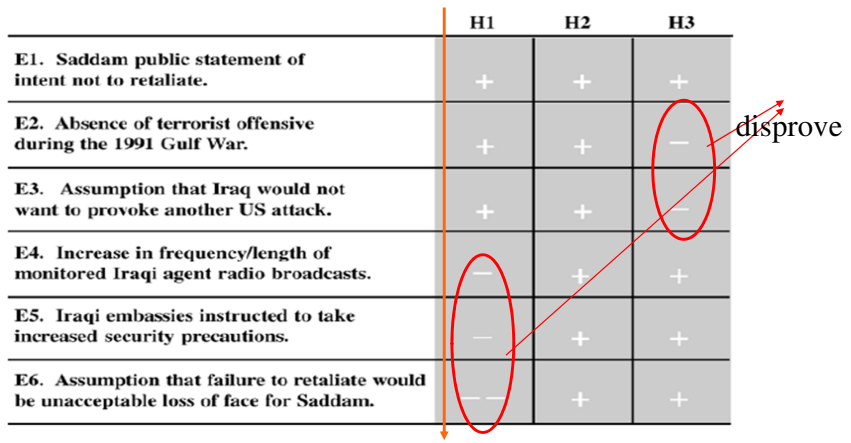
- put in missing factors
- delete evidence that have no diagnostic value

Step 5: Draw Conclusions

Work down the matrix, looking at each hypothesis

Proceed by trying to disprove the hypotheses rather than prove them

	H1	H2	H3
E1. Saddam public statement of intent not to retaliate.	+	+	+
E2. Absence of terrorist offensive during the 1991 Gulf War.	+	+	–
E3. Assumption that Iraq would not want to provoke another US attack.	+	+	–
E4. Increase in frequency/length of monitored Iraqi agent radio broadcasts.	–	+	+
E5. Iraqi embassies instructed to take increased security precautions.	–	+	+
E6. Assumption that failure to retaliate would be unacceptable loss of face for Saddam.	– –	+	+



Step 6: Analyze Conclusions

Analyze how sensitive your conclusion is to a few critical items of evidence

- the consequences if the evidence were wrong
- check the original source

Step 7: Report Conclusions

Decision-maker needs to make decisions on the basis of a full set of alternative possibilities

The importance is eliminating not confirming

Discuss the relative likelihood of all the hypotheses

Step 8: Identify Milestones

Analytical conclusion should always be regarded as tentative

Specify in advance things will change possibly

Summary and Conclusion

Key differences b/t competing hypotheses from conventional intuitive analysis

	Competing hypotheses	conventional
Num of possibilities	Full set	Most likely one
Diagnostic value	Greatest	Maybe no
Use of evidence	Refute	Confirm

Things to Remember

ACH's Way to analyze

- not by satisfying strategy
- but by simultaneous evaluation

Note the important difference b/t

- disproof and no proof

The Sense-Making Loop

Support visualization with computations for data processing

Form a loop: visualize - refine

Gather (forage) information

Re-Represent

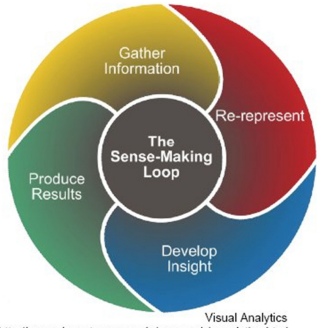
- choose form that aids analysis

Develop insight

- through manipulation of representations

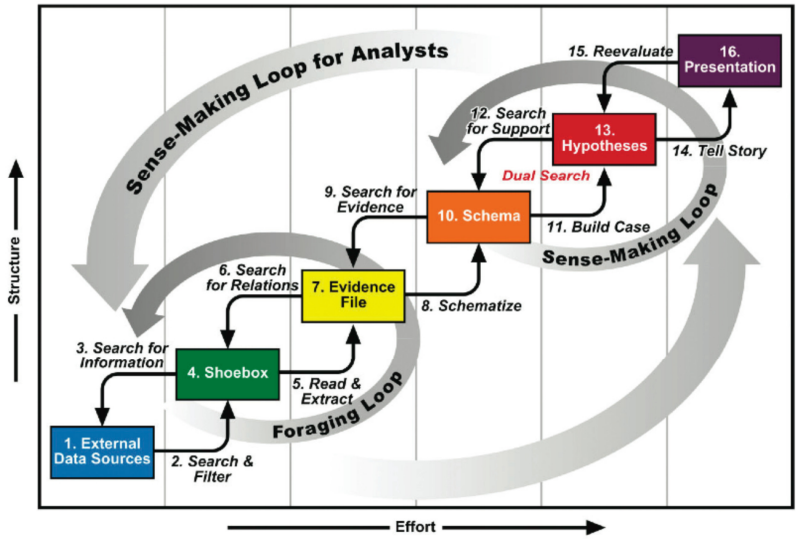
Produce results

- "product"



Visual Analytics
http://www.viscenter.uncc.edu/courses/visanalytics.html

Nominal Sense-Making Process



Reasoning Artifacts

Elemental artifacts

- source intelligence, evidence, assumptions

• Pattern artifacts

- relationships, temporal and spatial structure

• Higher-order knowledge constructs

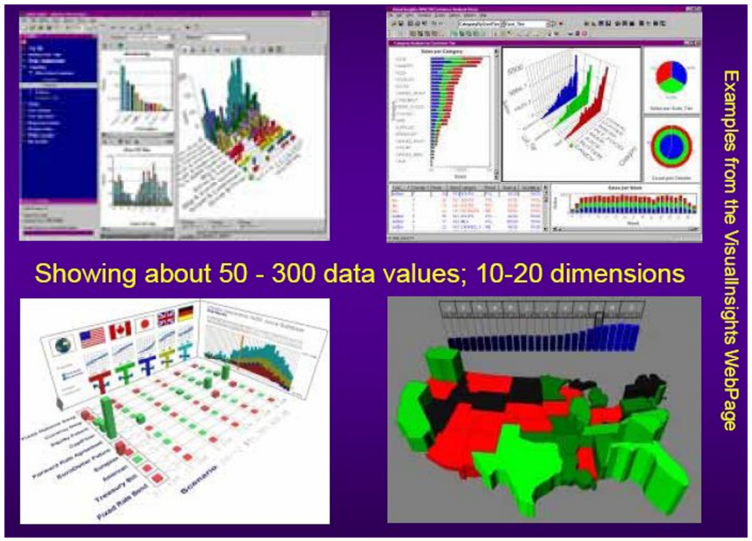
- arguments, causality, models

• Complex reasoning constructs

- hypotheses, scenarios

All these become part of the Visual Analytics sense-making (reasoning) process

Standard Information Displays



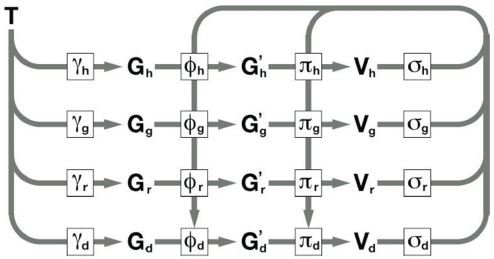
Multiview Visualization

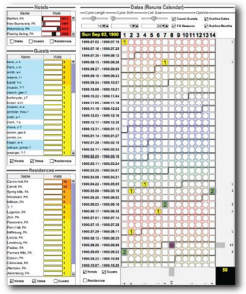
Cross-filtering

- a method for interactively expressing sequences of multidimensional set queries by selecting and filtering unique data values across pairs of views
- the next 2 slides are due to Chris Weaver (U Oklahoma) – check out his *IEEE Trans Visualization and Computer Graphics* (16(2): 192-204) paper

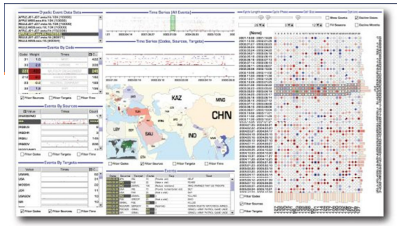
Cross-Filtering

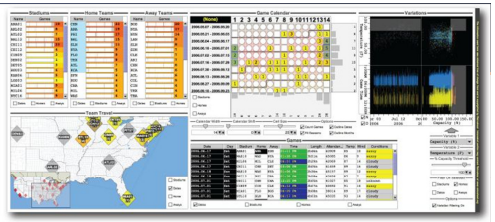
a general pattern for constructing an interdependent set of data transformation operations that support the method





- Group (γ) data records into sets for each unique attribute value.
- Filter (ϕ) each set, keeping records whose attribute values match those selected in other views.
- Project/visually encode (π) each value and its filtered set.
- Select (σ) values/sets corresponding to brushed glyphs in the view.





Design Variations

		KEDS	Hotels	Retrosheet	Cinegraph
Attributes	Nominal	code (event)	name (guest)	name (home team, away team)	name (movie, genre, oscar, person, role)
	Temporal	date (event)	date (visit)	date & time (game)	date (release)
	Spatial	region (countries)	location (hotel, residence)	location (stadium)	-
	Numerical	cooperative/conflictual weight	-	capacity, attendance, temperature, wind speed	box office, rating average, rating count
Auxiliary Views	Pre-filter	list (data sources)	-	-	sliders (ratings & roles thresholds)
	Post-filter	map (world)	map (Pennsylvania)	map (North America), rich drill-down table	attribute relationship graph
	Detail	drill-down table, split time series	drill-down table	-	movie viewer
	Nested	scatter plot (date vs. weight)	1-D heatmap (visit count by date)	1-D heatmap (game count by date)	histogram (rating distribution)