Topic Notes

Cognitive Issues & User Tasks

CS 7450 - Information Visualization
January 25, 2011
John Stasko

Outline

- Overview
- 1. Role – How visualizations aid cognition?
- 2. Tasks – What does the visualization assist?

Basic Premise

- Understanding (the cognitive aspects) is the crucial part of InfoVis
- Visualization is simply a tool useful for aiding analysis, exploration, comprehension and understanding

- Discussed the role of external cognition aids briefly earlier in intro, more now

How Are Graphics Used?

- What does a visualization or graphic image provide for us?

How Are Graphics Used?

- Larkin & Simon '87 investigated usefulness of graphical displays
- Graphical visualization could support more efficient task performance by:
  - Allowing substitution of rapid perceptual influences for difficult logical inferences
  - Reducing search for information required for task completion
- (Sometimes text is better, however)

Norman's Thoughts

- Cognitive Artifacts
  - Wonderful discussion on p. 49
- Matching Representation to Task
  - Tic-tac-toe, flight schedules
- Representations Aid Info Access and Computation
  - Medical prescriptions, Roman numerals, maps & legends
- Naturalness and Experiential Cognition

Revisit

Visualization

- Often thought of as process of making a graphic or an image
- Really is a cognitive process
  - Form a mental image of something
  - Internalize an understanding
- "The purpose of visualization is insight, not pictures"
  - Insight: discovery, decision making, explanation

Revisit

Main Idea

- Visuals help us think
  - Provide a frame of reference, a temporary storage area
- Cognition → Perception
- Pattern matching
- External cognition aid
  - Role of external world in thinking and reason

Larkin & Simon '87
Card, Mackinlay, Shneiderman '98

Revisit

Visualization

- Definition
  - "The use of computer-supported, interactive visual representations of data to **amplify cognition**."

    From [Card, Mackinlay Shneiderman '98]

Examine More Closely

- What does "amplify cognition" mean?
- Discuss

Another View

- Leverage Hutchins' theory of distributed cognition (DCog) to explain the value and utility of infovis
- Use DCog as a supporting theoretical framework for infovis

Liu, Nersessian, Stasko
TVCG '08

Amplifying Cognition

- Hutchins argues that tools don't amplify or scaffold cognition (a more traditional cognitive science view)
  - Eg, Our memory isn't amplified
- Instead, tools help transform the analytic process into another more doable one

Distributed Cognition

- Cognitive system is composed of people and the artifacts they use
  - Cognition isn't only internal
- Changes in external representation spur changes in internal representation and understanding
- It is **interaction** with the external representations that drives this process

More Details

- OK, so now let's talk about the analytic process in more detail, and specifically, how visualization can play a role

Understanding

- People utilize an mental/internal model that is generated based on what is observed
- B. Tversky calls the internal model a *cognitive map*
  - Think about that term

Example

- You're taking the MARTA train to get to Georgia State University
  - You have some existing internal model of the system, stops, how to get there
  - On train, you glance at MARTA map for help
  - Refines your internal model, clarifying items and extending it
  - Note that it's still not perfect, no internal model ever is

Cognitive Map

- Just don't have one big one
- Have large number of these for all different kinds of things
- Collection of cognitive maps --> *Cognitive collage*

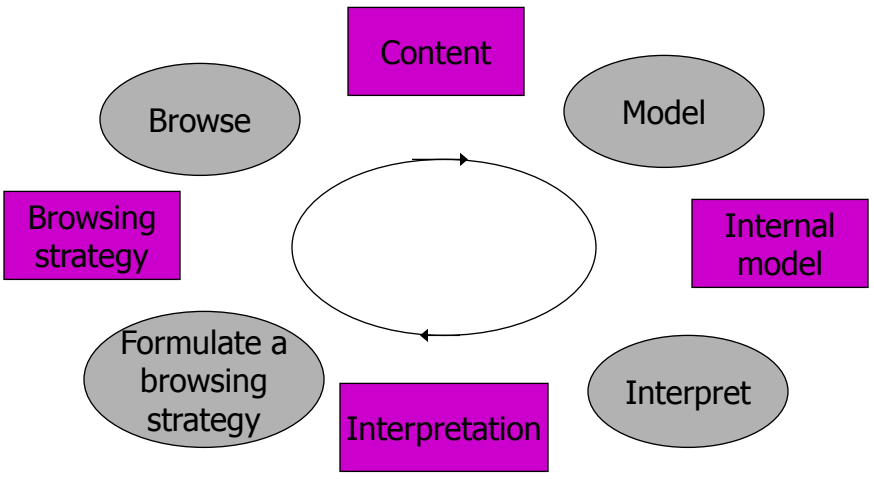
1. Process Models

- (Recall the user and cognitive models from HCI?)
- Process by which a person looks at a graphic and makes some use of it
  - A number of substeps probably exist
- Can you describe process?

Process Model 1

- Robert Spence
- *Navigation* - Creation and interpretation of an internal mental model

Navigation



Interpretation

- Can someone explain that?

Interpretation

- - Content is the display on screen
  - Modeling of that pattern results in cognitive map
  - Interpretation (ah, variables x and y are related) leads to new view, that generates an idea for a new browsing strategy
  - Look at the display again with that

Process Model 2

- Card, Mackinlay, Shneiderman book
- Knowledge crystallization task
  - Gather info for some purpose, make sense of it by constructing a representational framework, and package it into a form for communication or action

Knowledge Crystallization

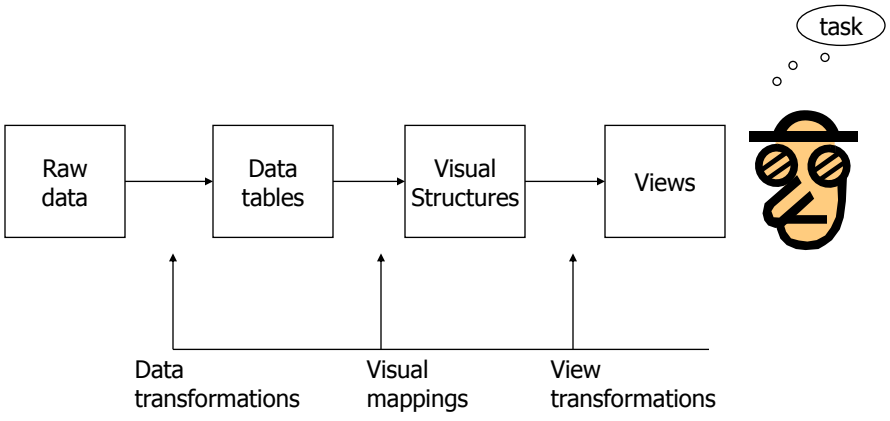
- Information foraging
- Search for schema (representation)
- Instantiate schema
- Problem solve to trade off features
- Search for a new schema that reduces problem to a simple trade-off
- Package the patterns found in some output product

From CMS '98

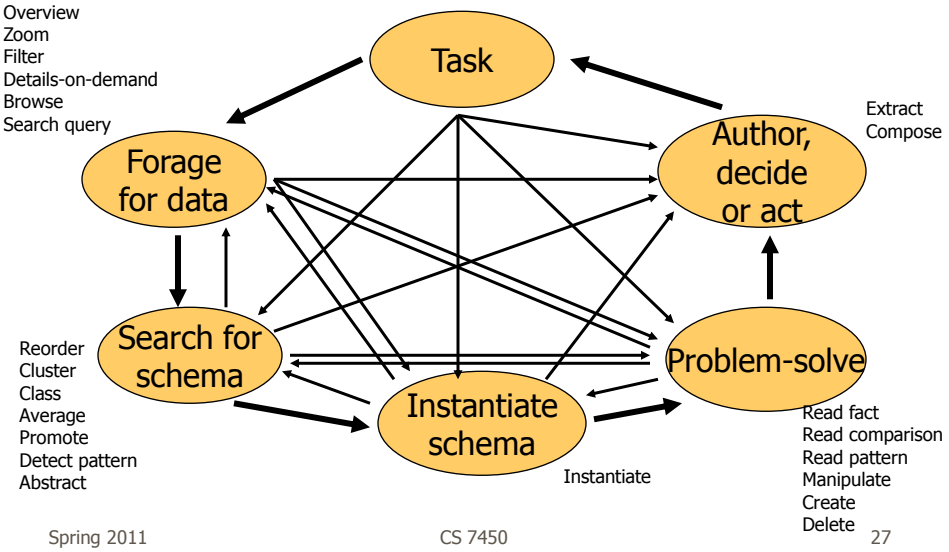
How Vis Amplifies Cognition

- Increasing memory and processing resources available
- Reducing search for information
- Enhancing the recognition of patterns
- Enabling perceptual inference operations
- Using perceptual attention mechanisms for monitoring
- Encoding info in a manipulable medium

Process



Knowledge Crystallization



2. User Tasks

- What things will people want to accomplish using information visualizations?
- Earlier, we briefly discussed
  - search vs. browsing

Browsing vs. Search

- Important difference in activities
- Appears that information visualization may have more to offer to browsing

- But…browsing is a softer, fuzzier activity
- So, how do we articulate utility?
  - Maybe describe when it's useful
  - When is browsing useful?

Browsing

- Useful when
  - Good underlying structure so that items close to one another can be inferred to be similar
  - Users are unfamiliar with collection contents
  - Users have limited understanding of how system is organized and prefer less cognitively loaded method of exploration
  - Users have difficulty verbalizing underlying information need
  - Information is easier to recognize than describe

Lin '97

Thought

- Maybe infovis isn't about answering questions or solving problems… hmmm
- Maybe it's about asking better questions

Tasks

- OK, but browsing and search are very high level
- Let's be more specific…

Example from Earlier

Questions:
- Which cereal has the most/least potassium?
- Is there a relationship between potassium and fiber?
  - If so, are there any outliers?
- Which manufacturer makes the healthiest cereals?

	A	B	C	D
1	Cereal	Manufacturer	Fiber	Potassium
2	100% Bran	N	10	280
3	100% Natural Bran	Q	2	135
4	All-Bran	K	9	320
5	All-Bran with Extra Fiber	K	14	330
6	Almond Delight	R	1	0
7	Apple Cinnamon Cheeric	G	1.5	70
8	Bran Chex	R	4	125
9	Bran Flakes	P	5	190
10	Cap'n'Crunch	Q	0	35
11	Cheerios	G	2	105
12	Cocoa Puffs	G	0	55
13	Corn Chex	R	0	25
14	Corn Flakes	K	1	35
15	Count Chocula	G	0	65
16	Cracklin' Oat Bran	K	4	160
17	Cream of Wheat (Quick)	N	1	0
18	Crispy Wheat & Raisins	G	2	120
19	Double Chex	R	1	80
20	Froot Loops	K	1	30
21	Frosted Flakes	K	1	25
22	Fruit & Fibre Dates, Wal	P	5	200
23	Fruitful Bran	K	5	190
24	Fruity Pebbles	P	0	25
25	Golden Grahams	G	0	45
26	Grape Nuts Flakes	P	3	85
27	Honey Nut Cheerios	G	1.5	90
28	Honey-comb	P	0	35
29	Just Right Fruit & Nut	K	2	95
30	Life	Q	2	95
31	Lucky Charms	G	0	55
32	Maypo	A	0	95
33	Muesli Raisins, Dates, &	R	3	170
34	Multi-Grain Cheerios	G	2	90
35	Nutri-Grain Almond-Rais	K	3	130
36	Nutri-grain Wheat	K	3	90
37	Oatmeal Raisin Crisp	G	1.5	120
38	Post Nat. Raisin Bran	P	6	260
39	Product 19	K	1	45
40	Quaker Oatmeal	Q	2.7	110
41	Raisin Bran	K	5	240
42	Raisin Nut Bran	G	2.5	140
43	Rice Krispies	K	0	35
44	Shredded Wheat	N	3	95
45	Shredded Wheat 'n'Bran	N	4	140
46	Shredded Wheat spoon	N	3	120
47	Smacks	K	1	40
48	Special K	K	1	55
49	Strawberry Fruit Wheats	N	3	90
50	Total Corn Flakes	G	0	35
51	Total Raisin Bran	G	4	230
52	Total Whole Grain	G	3	110
53	Trix	G	0	25
54	Wheaties	G	3	110
55	Wheaties Honey Gold	G	1	60

Exercise

- What are the (types of) tasks being done here?
- Can you think of others?
  - Let's develop a list

Task Taxonomies

- Number of different ones exist, important to understand what process they focus on
  - Creating an artifact
  - Human tasks
  - Tasks using visualization system
  - …

User Tasks

- Wehrend & Lewis created a low-level, domain independent taxonomy of user tasks in visualization environments
- Eleven basic actions
  - identify, locate, distinguish, categorize, cluster, distribution, rank, compare within relations, compare between relations, associate, correlate

Wehrend & Lewis
Vis '90

Another Perspective

- Shneiderman proposed task × data type taxonomy to understand what people do with visualization
- Mantra: "Overview first, zoom and filter, then details on demand"
  - Design paradigm for infovis systems

Shneiderman
VL '96

Taxonomy

- Data Types
  1. 1D
  2. 2D
  3. 3D
  4. Temporal
  5. ND
  6. Tree
  7. Network
- Tasks
  1. Overview
  2. Zoom
  3. Filter
  4. Details-on-demand
  5. Relate
  6. History
  7. Extract

Another Task Taxonomy

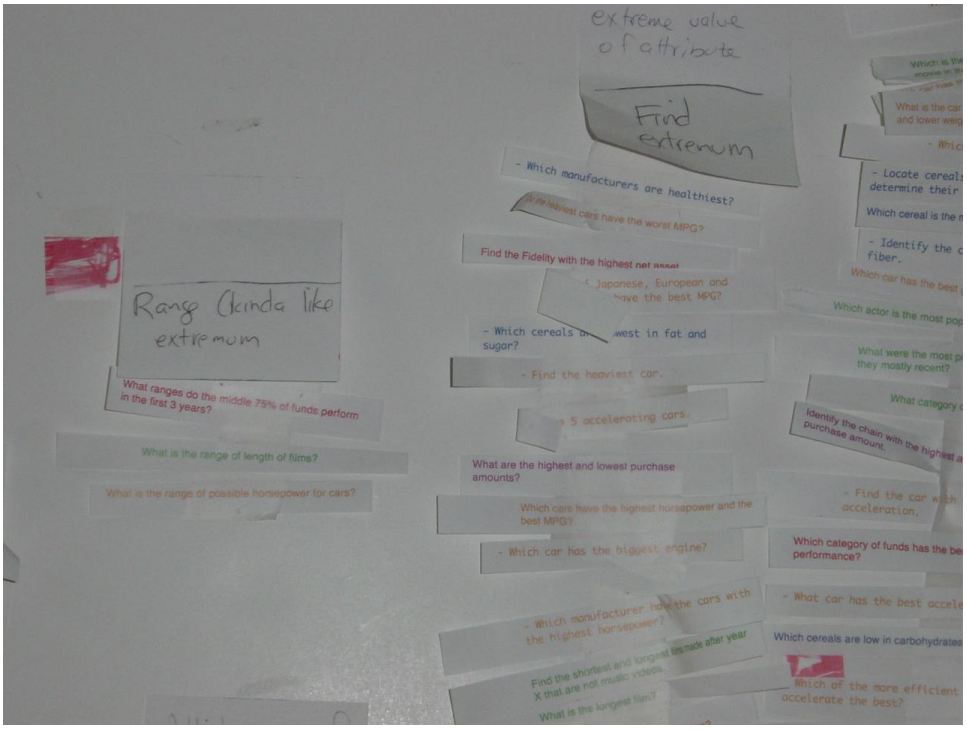
- Amar, Eagan, & Stasko – InfoVis '05

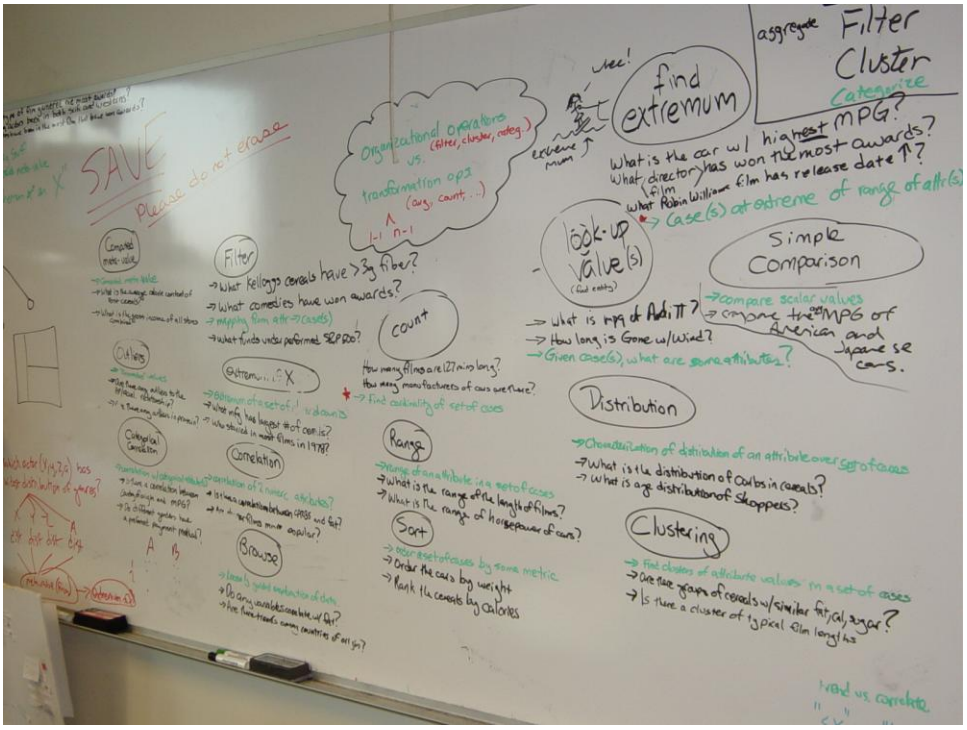
Background

- Use "commercial tools" class assignment from this class
- Students generate questions to be answered using commercial infovis systems
- Data sets:

Domain	Data cases	Attributes	Questions Generated
Cereals	78	15	107
Mutual funds	987	14	41
Cars	407	10	153
Films	1742	10	169
Grocery surveys	5164	8	126

- Generated 596 total analysis tasks


extreme value of attribute
Find extremum
Range (kinda like extremum
- Which manufacturers are healthiest?
Find the Fidelity with the highest net asset
- Find the heaviest car.
What ranges do the middle 75% of funds perform in the first 3 years?
What is the range of length of films?
What is the range of possible horsepower for cars?
What are the highest and lowest purchase amounts?
Which cars have the highest horsepower and the best MPG?
- Which car has the biggest engine?
Find the shortest and longest films made after year X that are not music videos.
What is the longest film?
Identify the chain with the highest purchase amount.
- Find the car with acceleration.
Which category of funds has the be performance?
- What car has the best accele
Which cereals are low in carbohydrates
Which of the more efficient accelerate the best?



Terminology

- *Data case* – An entity in the data set
- *Attribute* – A value measured for all data cases
- *Aggregation function* – A function that creates a numeric representation for a set of data cases (eg, average, count, sum)

1. Retrieve Value

General Description:
Given a set of specific cases, find attributes of those cases.

Examples:
- What is the mileage per gallon of the Audi TT?
- How long is the movie Gone with the Wind?

2. Filter

General Description:
Given some concrete conditions on attribute values, find data cases satisfying those conditions.

Examples:
- What Kellogg's cereals have high fiber?
- What comedies have won awards?
- Which funds underperformed the SP-500?

3. Compute Derived Value

General Description:
Given a set of data cases, compute an aggregate numeric representation of those data cases.

Examples:
- What is the gross income of all stores combined?
- How many manufacturers of cars are there?
- What is the average calorie content of Post cereals?

4. Find Extremum

General Description:
Find data cases possessing an extreme value of an attribute over its range within the data set.

Examples:
- What is the car with the highest MPG?
- What director/film has won the most awards?
- What Robin Williams film has the most recent release date?

5. Sort

General Description:
Given a set of data cases, rank them according to some ordinal metric.

Examples:
- Order the cars by weight.
- Rank the cereals by calories.

6. Determine Range

General Description:
Given a set of data cases and an attribute of interest, find the span of values within the set.

Examples:
- What is the range of film lengths?
- What is the range of car horsepowers?
- What actresses are in the data set?

7. Characterize Distribution

General Description:
Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute's values over the set.

Examples:
- What is the distribution of carbohydrates in cereals?
- What is the age distribution of shoppers?

8. Find Anomalies

General Description:
Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers.

Examples:
- Are there any outliers in protein?
- Are there exceptions to the relationship between horsepower and acceleration?

9. Cluster

General Description:
Given a set of data cases, find clusters of similar attribute values.

Examples:
- Are there groups of cereals w/ similar fat/calories/sugar?
- Is there a cluster of typical film lengths?

10. Correlate

General Description:
Given a set of data cases and two attributes, determine useful relationships between the values of those attributes.

Examples:
- Is there a correlation between carbohydrates and fat?
- Is there a correlation between country of origin and MPG?
- Do different genders have a preferred payment method?
- Is there a trend of increasing film length over the years?

Discussion/Reflection

- Compound tasks
  - "Sort the cereal manufacturers by average fat content"
    Compute derived value; Sort
  - "Which actors have co-starred with Julia Roberts?"
    Filter; Retrieve value

Discussion/Reflection

- What questions were left out?
  - Basic math
    "Which cereal has more sugar, Cheerios or Special K?"
    "Compare the average MPG of American and Japanese cars."
  - Uncertain criteria
    "Does cereal (X, Y, Z…) sound tasty?"
    "What are the characteristics of the most valued customers?"
  - Higher-level tasks
    "How do mutual funds get rated?"
    "Are there car aspects that Toyota has concentrated on?"
  - More qualitative comparison
    "How does the Toyota RAV4 compare to the Honda CRV?"
    "What other cereals are most similar to Trix?"

Concerns

- InfoVis tools may have influenced students' questions
- Graduate students as group being studied
  - How about professional analysts?
- Subjective – Not an exact science
- Data was really quantitative so may get a different set of tasks for relational/graph data
  - See Lee et al, BELIV '06

Contributions

- Set of grounded low-level analysis tasks
- Potential use of tasks as a language/vocabulary for comparing and evaluating infovis systems

Can We Be More?

- Is InfoVis helping people enough?
- What do we need to do to provide even more value?

Providing Better Analysis

- Combine computational analysis approaches such as data mining with infovis
  - Too often viewed as competitors in past
- Each has something to contribute

Shneiderman
Information Visualization '02

Issues

- Issues influencing the design of discovery tools:
  - Statistical Algorithms vs. Visual data presentation
  - Hypothesis testing vs. exploratory data analysis
- Pro's and Con's?

Differing Views

- Hypothesis testing
  - Advocates:
    By stating hypotheses up front, limit variables and sharpens thinking, more precise measurement
  - Critics:
    Too far from reality, initial hypotheses bias toward finding evidence to support it
- Exploratory Data Analysis
  - Advocates:
    Find the interesting things this way, we now have computational capabilities to do them
  - Skeptics:
    Not generalizable, everything is a special case, detecting statistical replationships does not infer cause and effect

Recommendations

- Integrate data mining and information visualization
- Allow users to specify what they are seeking
- Recognize that users are situated in a social context
- Respect human responsibility

Another Question?

- Are the visualizations helping with exploratory analysis enough?
- Are they attempting to accomplish the right goals?

Status Quo Limitations

- Current Information Visualization systems inadequately support decision making:
  - Limited Affordances
  - Predetermined Representations
  - Decline of Determinism in Decision-Making
- "Representational primacy" versus "Analytic primacy"

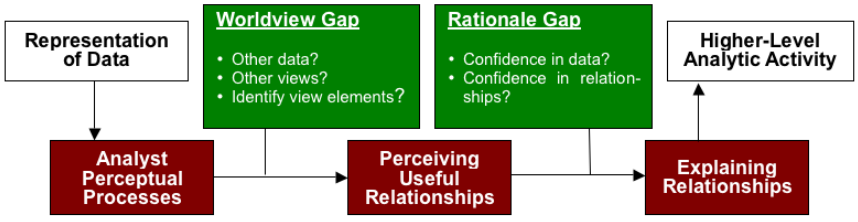
Amar & Stasko
TVCG '05

High-Level Tasks

- Complex decision-making, especially under uncertainty
- Learning a domain
- Identifying the nature of trends
- Predicting the future
- …

Analytic Gaps

- Analytic gaps – "obstacles faced by visualizations in facilitating higher-level analytic tasks, such as decision making and learning."
  - Worldview Gap
  - Rationale Gap



Knowledge Precepts

- For narrowing these gaps
  - Worldview-Based Precepts ("Did we show the right thing to the user?")
    - Determine Domain Parameters
    - Expose Multivariate Explanation
    - Facilitate Hypothesis Testing
  - Rationale-Based Precepts ("Will the user believe what they see?")
    - Expose Uncertainty
    - Concretize Relationships
    - Expose Cause and Effect

Application of Precepts

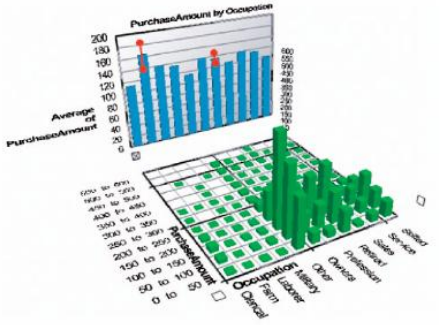


Fig. 2. Error bars (which we have added in red) would be a simple way to increase confidence in the degree of difference between two aggregations. (Picture taken from the SeeIt system by Visible Decisions, Inc.)

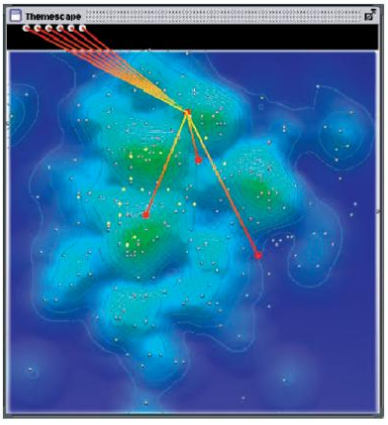


Fig. 3. This themescape variation allows documents with missing metadata, shown as dots in the upper black region, to participate in analysis, such as the reference relationship shown. (Picture courtesy of Nicholas Diakopoulos.)

Application of Precepts

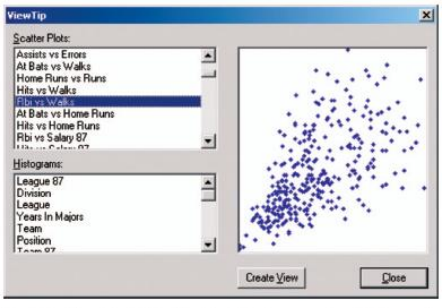


Fig. 4. The View Tips in SpotFire Pro 4.0 allow users to quickly examine possible sources of correlation for further examination.

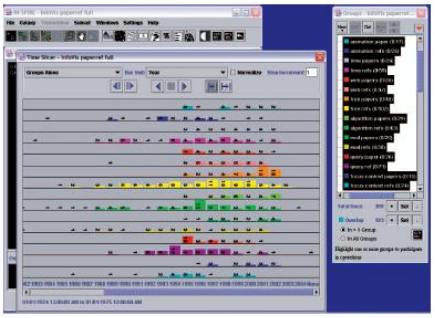


Fig. 5. IN-SPIRE uses horizontal scrolling to navigate time slices of user-defined content groups. (Picture produced at and provided with permission of Pacific Northwest National Laboratory, which is managed and operated by the Battelle Memorial Institute on behalf of the US Department of Energy.)

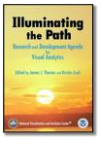
Put Them Together

- Combine the ideas:
  - Use computational, statistical analysis more
  - Cater to the user's analytic reasoning needs
- And put together with infovis

- Leads to…

Visual Analytics

- "The science of analytical reasoning facilitated by interactive visual interfaces"
- Combines
  - Data analysis
  - Infovis
  - Analytical reasoning
- Grew from view that infovis was neglecting these other aspects
  - True?



Thomas & Cook
Illuminating the Path

Visual Analytics

- Grew from stimulus in the homeland security area
  - Need for better data analysis methods
  - Really big data
- Topic for entire day later in term…

HW 2

- Reminder: Due on Thursday

Design Project

More details on website

- Group of 2-4 students
- Understand problem, design, build
- You pick the topic/domain/data
  - **Absolutely crucial!!!**
  - NY Times vizs are good examples
- First milestone: Teams and topics in 2 weeks (Feb 8th)

Upcoming

- Multivariate visual representations
  - Reading:
    Inselberg '97
- Overview and Detail (Focus + Context)
  - Reading:
    Bederson '04

References

- Spence & CMS texts
- All referred to papers