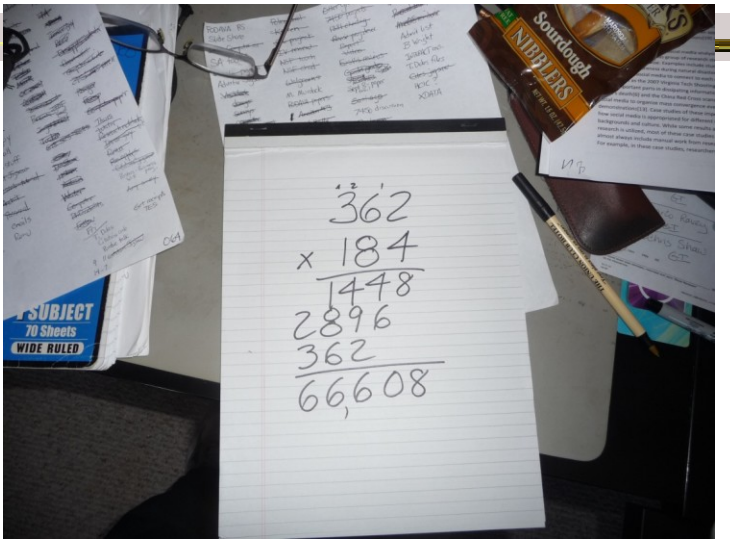
Value/Benefits of Visualization

CS 7450 - Information Visualization
August 28, 2013
John Stasko

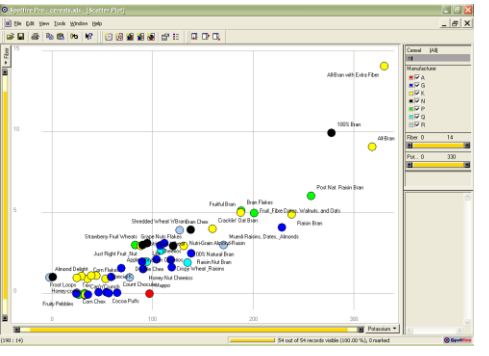
Premise

- Visualization is still not as well accepted in various communities as we might hope
  - Why?
- Goals
  - 1. Understand how vis helps
  - 2. Why is showing value important and difficult?
  - 3. Identify what vis does best
  - 4. Characterize (quantify?) the value of vis



Example from Intro Day

	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



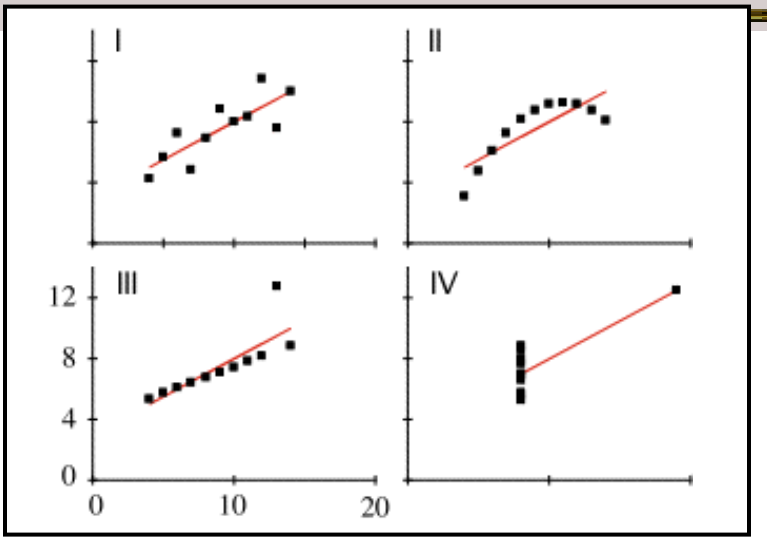
Another Illustrative Example

Four Data Sets

- Mean of the x values = 9.0
- Mean of the y values = 7.5
- Equation of the least-squared regression line is: y = 3 + 0.5x
- Sums of squared errors (about the mean) = 110.0
- Regression sums of squared errors (variance accounted for by x) = 27.5
- Residual sums of squared errors (about the regression line) = 13.75
- Correlation coefficient = 0.82
- Coefficient of determination = 0.67

http://astro.swarthmore.edu/astro121/anscombe.html

The Data Sets



The Values

1	2	3	4
10.0, 8.04	10.0,9.14	10.0, 7.46	8.0, 6.58
8.0, 6.95	8.0,8.14	8.0, 6.77	8.0, 5.76
13.0, 7.58	13.0,8.74	13.0,12.74	8.0, 7.71
9.0, 8.81	9.0,8.77	9.0, 7.11	8.0, 8.84
11.0, 8.33	11.0,9.26	11.0, 7.81	8.0, 8.47
14.0, 9.96	14.0,8.10	14.0, 8.84	8.0, 7.04
6.0, 7.24	6.0,6.13	6.0, 6.08	8.0, 5.25
4.0, 4.26	4.0,3.10	4.0, 5.39	19.0,12.50
12.0,10.84	12.0,9.13	12.0, 8.15	8.0, 5.56
7.0, 4.82	7.0,7.26	7.0, 6.42	8.0, 7.91
5.0, 5.68	5.0,4.74	5.0, 5.73	8.0, 6.89

1

How Does Vis Help?

- What does a visualization or graphic image provide for us in terms of cognitive benefits?

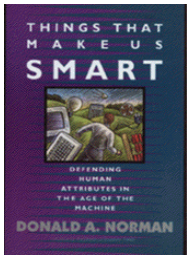
(Discuss)

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

- Book chapter that argues how good representations of data help us to understand it better



"The Power of Representation"
Chapter 3 from
Things That Make Us Smart

Overview

- Opening paragraph
  - "The power of the unaided mind is highly overrated. Without external aids, memory, thought, and reasoning are all constrained. But human intelligence is highly flexible and adaptive, superb at inventing procedures and objects that overcome its limits. The real powers come from devising external aids that enhance cognitive abilities. How have we increased memory, thoughts, and reasoning? By the invention of external aids: It is things that make us smart. Some assistance comes through cooperative social behavior; some arises through exploitation of the information present in the environment; and some comes through the development of tools of thought – cognitive artifacts – that complement abilities and strengthen mental powers."

p. 43

Further

- "A good representation captures the essential elements of the event, deliberately leaving out the rest. … The critical trick is to get the abstractions right, to represent the important aspects and not the unimportant."

  p. 49

Examples

- Matching Representation to Task
  - Tic-tac-toe, flight schedules
- Representations Aid Info Access and Computation
  - Medical prescriptions, Roman numerals, maps & legends

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?

Does your brain get bigger?
Do you get smarter?

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

Can anyone explain DCog?

Liu, Nersessian, Stasko
TVCG (InfoVis) '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, (what?)
  - 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

Intermission

- Surveys

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*

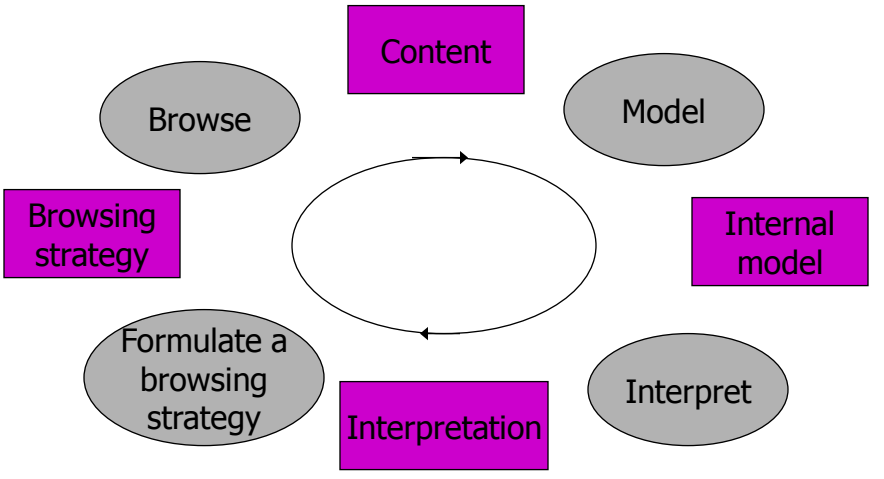
1. Process Models

- (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

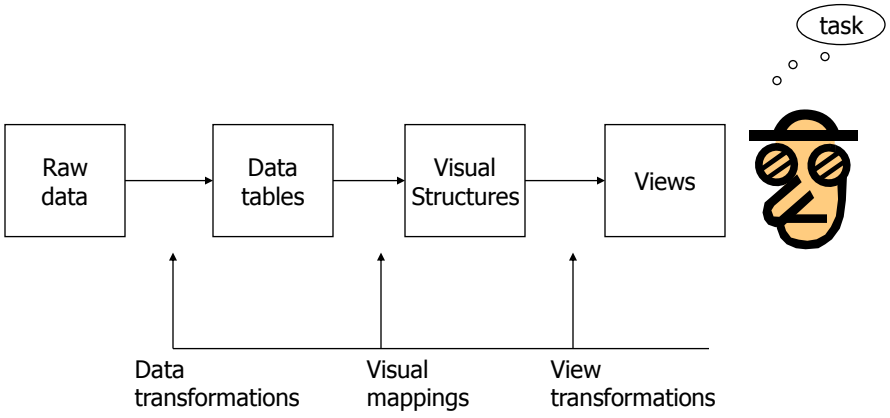
- Information foraging
- Search for schema (representation)
- Instantiate schema
- Problem solve to trade off features
- Search for a new schema that reduces problem to a simple trade-off
- Package the patterns found in some output product

From CMS '98

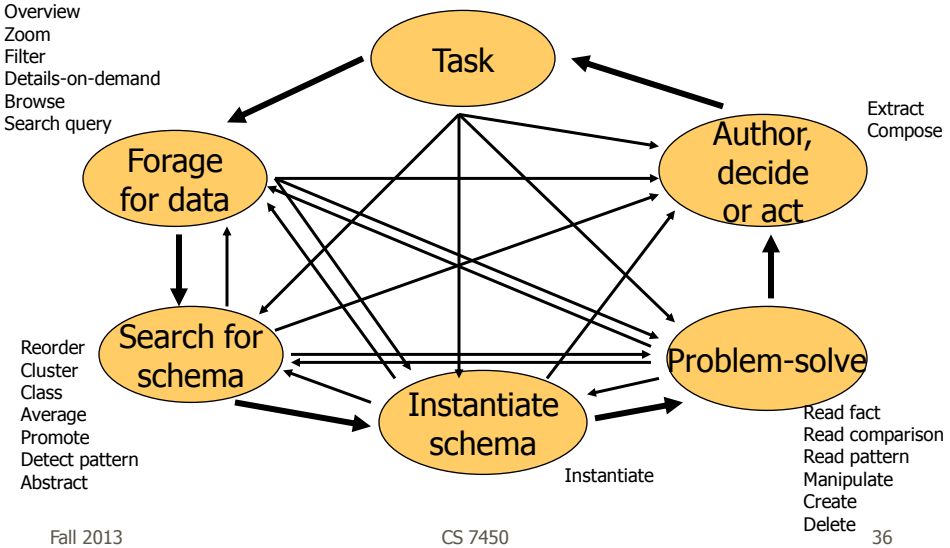
How Vis Amplifies Cognition

- 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



Alternate Model

Acquire ⟶ Parse ⟶ Filter ⟶ Mine ⟶ Represent ⟶ Refine ⟶ Interact

From: Fry '02

2

Showing Value

- Why is it important?
  - Organizations want measurable outcomes and benefits
  - Easier in other fields

Fekete, van Wijk, Stasko, North
Springer '08

Showing Value

- Why is it difficult?
  - Benefits of vis not easily quantifiable
    - Ask better questions
    - Facilitate exploration
    - Generate "insights"

Epistemological Issues

- Science theories can't be proved true, only falsified
  - We select best theory at the time

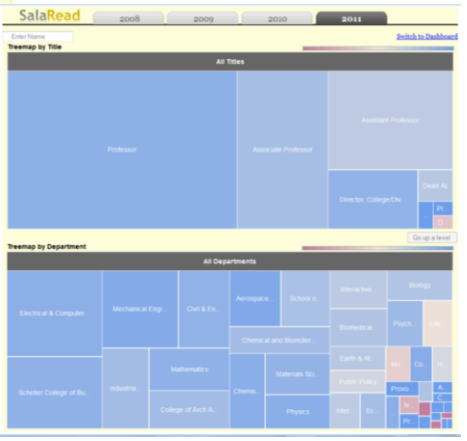
What Vis Does Best

- What does visualization uniquely contribute (compared to other approaches)?
  - Presentation
  - Analysis

Presentation

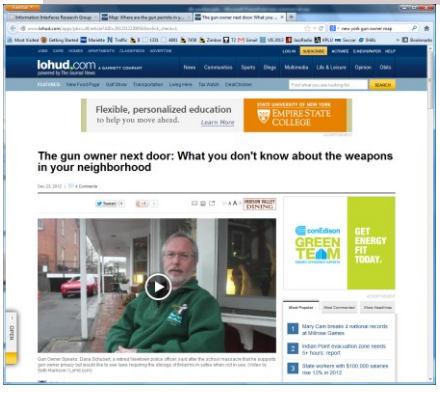
- Simply communicating data visually can have a profound effect
  - Examples

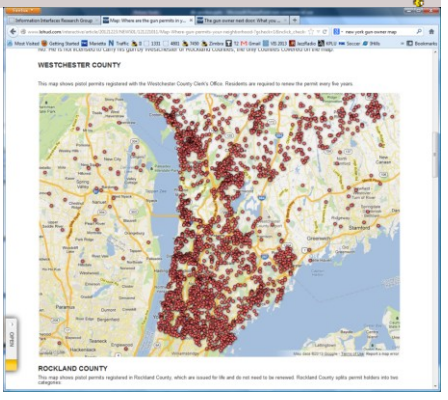
CS 7450
Fall '12
Design project



Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri

Gun Ownership in NY Counties





http://www.lohud.com/apps/pbcs.dll/article?AID=20123122300560&nclick_check=1

http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/
Map-Where-gun-permits-your-neighborhood-?gcheck=1&nclick_check=1

Analysis

understand, compare, decide, judge, evaluate, assess, determine, …

Many Data Analysis Approaches

Statistics
Database & information retrieval
Data mining
Machine learning

Think Carefully

- Preconceptions about visualization's utility
  - Answering specific questions and accomplishing specific analytic tasks
  - Generating unexpected, serendipitous discoveries and insights

    "Finding a needle in a haystack"

Yes, but not what it's best for

1. Visualization is more than just answering a specific question (as is often the case for automated analysis methods)
It also is about the investigative analysis process, which helps us to learn about, develop awareness of, and generate trust in the data, its domain, and its context.

Learning, awareness, trust, context

2. Visualization, primarily through its interactive capabilities, promotes a dialog between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction

3. Visualization rapidly and efficiently facilitates flexible exploration to foster both broad and deep understanding of the information in a data collection.

Broad and deep understanding quickly

Visualization most useful in **exploratory data analysis**

Don't know what you're looking for
Don't have a priori questions
Want to know what questions to ask

Insights

spontaneous aha! moments
vs.
knowledge-building & model confirmation

Chang et al
IEEE CG&A '09

Measuring Value

Can we model & quantify the value of vis?

$$I(t) = V(D,S,t)$$

D – data
S – specification (hardware, vis technique, …)
t – time

I(t) – time varying image

van Wijk
Vis '05

More

$$\frac{dK}{dt} = P(I,K)$$

K – knowledge
P – perception and cognition of user

$$K(t) = K_0 + \int_0^t P(I,K,t)\, dt$$

Economic Model

We simplify this by assuming that there is a homogeneous user community, consisting of n users which use a certain visualization V to visualize a data set m times each, where each session takes k explorative steps and time T. This is a crude simplification of course. In the real world, the user community will often be highly varied, with different K_0's and also with different aims. The costs associated with using V come at four different levels:

- $C_i(S_0)$: *Initial development costs*. The visualization method has to be developed and implemented, possibly new hardware has to be acquired.
- $C_u(S_0)$: *Initial costs per user*. The user has to spend time on selection and acquisition of V, understanding how to use it, and tailoring it to his particular needs.
- $C_s(S_0)$: *Initial costs per session*. Data have to be converted, and an initial specification of the visualization has to be made.
- C_e: *Perception and exploration costs*. The user has to spend time to watch the visualization and understand it, as well as in modification and tuning of the specification, thereby exploring the data set.

The total costs are now given by

$$C = C_i + nC_u + nmC_s + nmkC_e.$$

The return on these investments consists of the value $W(\Delta K)$ of the acquired knowledge $\Delta K = K(T) - K(0)$ per session, multiplied by the total number of sessions:

$$G = nmW(\Delta K)$$

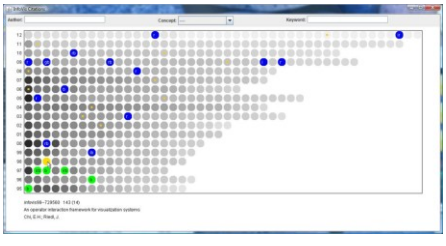
and hence for the total profit $F = G - C$ we find

$$F = nm(W(\Delta K) - C_s - kC_e) - C_i - nC_u.$$

This gives us a recipe to decide on the value of a visualization method. Positive are high values for n, m, $W(\Delta K)$, and low values for C_s, C_e, C_i, C_u, and k. Or, in other words, a great visualization method is used by many people, who use it routinely to obtain highly valuable knowledge, without having to spend time and money on hardware, software, and effort. Indeed, quite obvious.

Design Project

- Team advertising
  - Wiki pages in t-square
- Example projects
  - Saw one earlier
  - CiteVis, by my lab



Upcoming

- Labor Day Holiday
- Data & table/graph design
  - Paper:
    Few '06

References

- Spence & CMS texts
- All referred to papers