Elective in Software and Services
(Complementi di software e servizi per la società dell'informazione)

Section **Information Visualization**

Numbers of credit : 3

Giuseppe Santucci

Overture : a bad story
4 - Main issues : what, why, how

Thanks to John Stasko, Robert Spence, Ross Ihaka, Marti Hearst, Kent Wittemburg

Outline

- A bad story
- Issues to take into account designing an Infovis application
  - Task, subtasks, & problem (why)
  - Data (how)
  - Representation & Presentation (how)
  - Overview & interactive object/attribute selection (how)
  - Filtering (how)
  - Significant/reference data (how)
  - Navigational guidance (how)
  - Movement in the information space (how)

The Challenger disaster

- The Space Shuttle Challenger exploded shortly after take-off in January 1986. Subsequent investigation determined that the cause was failure of the O-ring seals used to isolate the fuel supply from burning gases. The figure shows O-ring failures at different launch temperatures collected by NASA before 1986.

History of O-Ring Damage in Field Joints (Cont)

MORTON THIOKOL, INC.
Wasatch Operations

* No Erosion
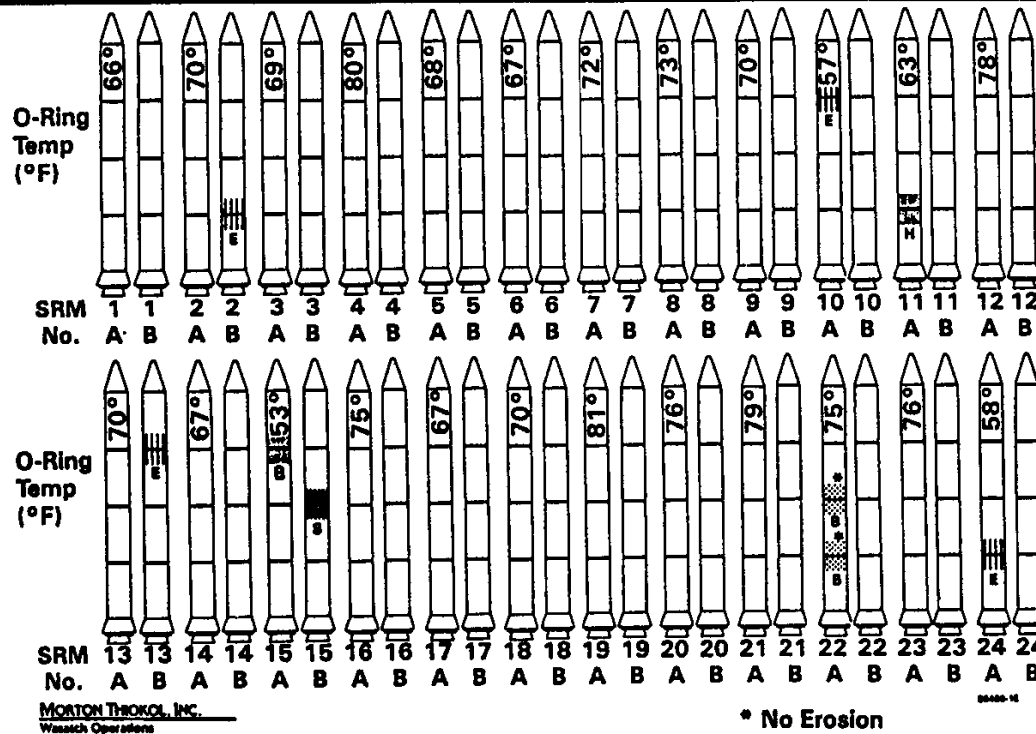
INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

http://www.math.yorku.ca/SCS/Gallery/

3

The Challenger disaster

- NASA staff had analysed the data plotting ambient temperature and number of O-ring failures but they had <u>excluded observations where no O-rings failed</u>, believing that they were uninformative. Unfortunately, those observations had occurred when the launch temperature was relatively warm (65-80 deg F).
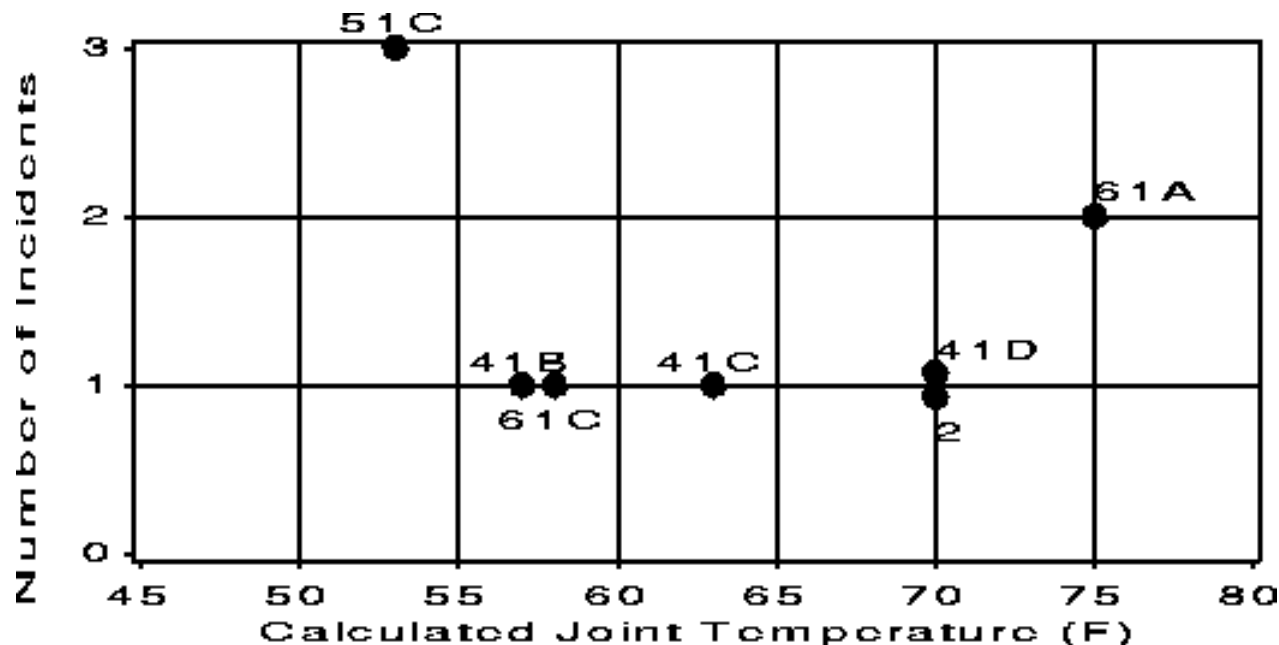
The Challenger disaster

- Reanalysis of the O-ring data involved fitting a logistic regression model. This provides a predicted extrapolation (black curve) of the probability of failure to the low (31 degF) temperature at the time of the launch and confidence bands on that extrapolation (red curves). There's not much data at low temperatures (the confidence band is quite wide), but the predicted probability of failure is uncomfortably high. Would you take a ride on Challenger when the weather is cold?

NASA Space Shuttle O-Ring Failures
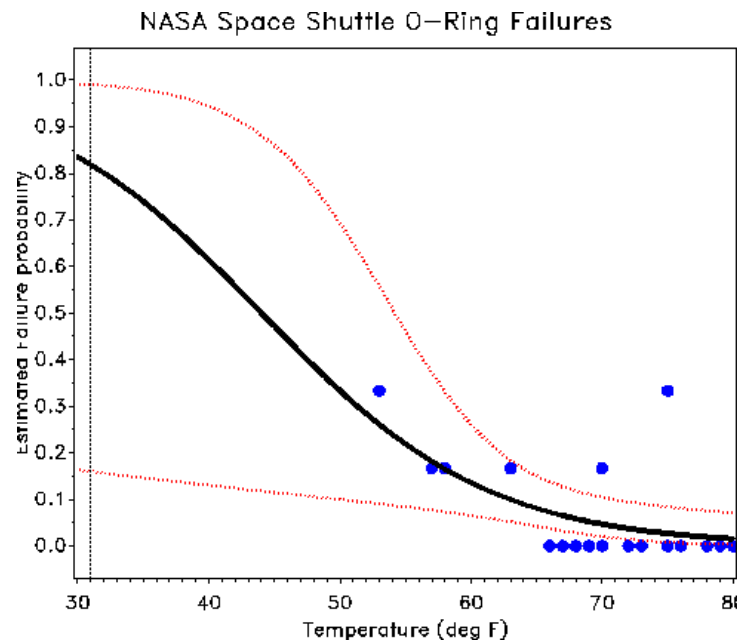
The Challenger disaster

- It is perhaps unreasonable to expect that a sophisticated statistical analysis of the data should have been carried out, given the time pressure for a launch / no-launch decision.

- Nevertheless, it is of interest to ask whether a re-design of the original graph might have signaled that something was amiss. Apart from the disastrous blunder of omitting the observations with 0 failures, two steps:
  1. drawing a smoothed curve to fit the points (to show the trend)
  2. removing the background grid lines (which obscure the data)
  gives a graph which should have caused any engineer to conclude that either (a) the data were wrong, or (b) there were excessive risks associated with both high and low temperatures.
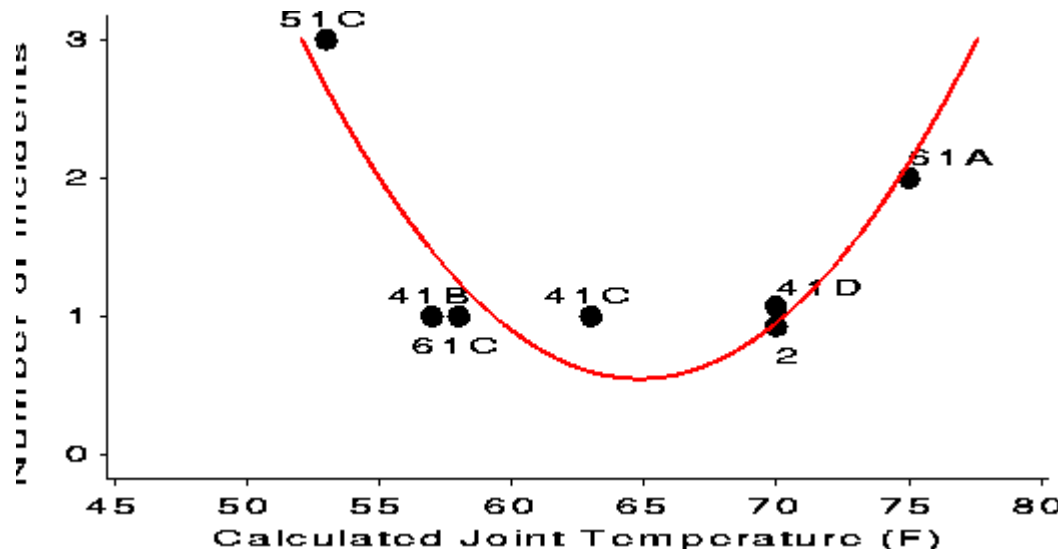
Outline

- Issues to take into account designing an Infovis application
  - Task, subtasks, & problem (why)
  - Data (how)
  - Representation & Presentation (how)
  - Overview & interactive object/attribute selection (how)
  - Filtering (how)
  - Significant/reference data (how)
  - Navigational guidance (how)
  - Movement in the information space (how)
- A case study (Explorative: We do not know the answer)

# Issues to take into account designing an Infovis application

- Different Infovis applications share common concepts, like user task, data, etc.

- This lecture will introduce them using a running example

- A very common Infovis application is a system that allows the user to select one item to buy (☺); we use, as a running example, a system that allows to buy a used car
  - Note that, in this case, the data cardinality is not very high, it is the combination of the options (model, price, age, fuel consumption, etc.) that makes it a complex activity and comparing, e.g., even 50 cars is not so easy

Outline

- Issues to take into account designing an Infovis application
  - Task, subtasks, & problem (why)
  - Data (how)
  - Representation & Presentation (how)
  - Overview & interactive object/attribute selection (how)
  - Filtering (how)
  - Significant/reference data (how)
  - Navigational guidance (how)
  - Movement in the information space (how)
- A case study (Explorative: We do not know the answer)

Task, subtask, & problem

- The running example task is to search for a used car within a predefined (not necessarily very large) set

- This task can be abstracted as : select a subset of "interesting" objects within a (even large) collection
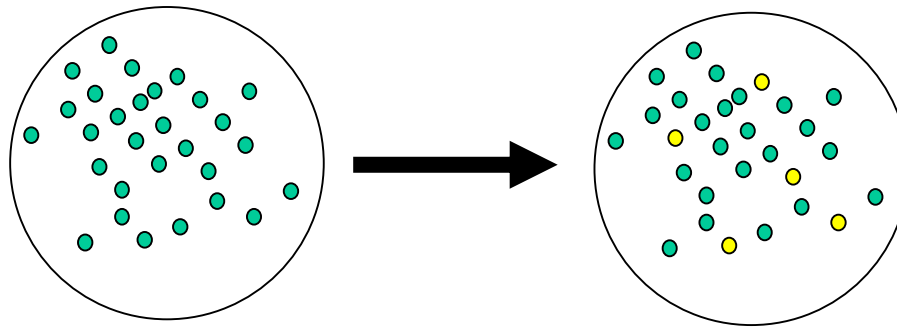
- N.B.: There exist completely different tasks, like:
  - There is a correlation between x and y?
  - There are "strange" data items (a Rolls Royce for just 100 Euros ?)
  - …

Task, subtask, & problem

- Obviously, to perform the main task, some, additional, subtasks are needed

- The most important one is to gain <u>insights</u> into a collection:
  - Ah ah, this seller has only very old cars;
  - Ah ah, there is not the information about cars' HP;
  - Ah ah, the price range is [10k .. 33k];
  - …

- The idea is: in order to make a decision on a set, I have to understand some set "characteristics"
  - At a global level (overview)
  - At the object level (details)

Task, subtask, & problem

- … select a subset of "interesting" objects within a (even large) collection
- What does "interesting" mean?
- We witness, very often, a lack of precision
  - Nice looking car ?
  - Inexpensive ?
  - …
- Or for other tasks
  - I'm looking for suspicious people passing the USA border…
  - Strange DNA patterns…
- Very often "a problem is formulated as it is being solved" (Schon, 1983)"
- In order to cope with vague problems we have to provide for a <u>high degree of interactivity</u>

Outline

- Issues to take into account designing an Infovis application
  - Task, subtasks, & problem (why)
  - Data (how)
  - Representation & Presentation (how)
  - Overview & interactive object/attribute selection (how)
  - Filtering (how)
  - Significant/reference data (how)
  - Navigational guidance (how)
  - Movement in the information space (how)
- A case study (Explorative: We do not know the answer)

Data

- Roughly speaking we can assume that our data is a table of "abstract data"
- Complexity rises from the number of attributes and the number of rows
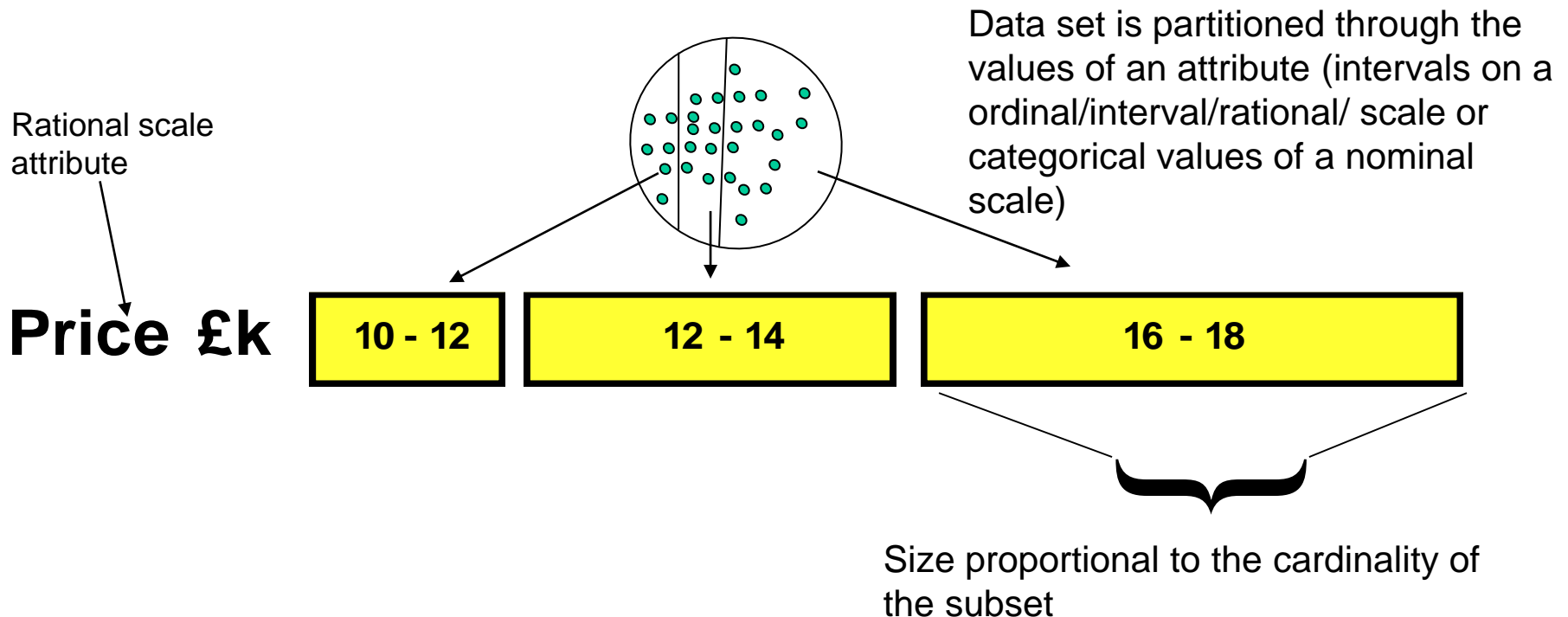- The used car dataset is as follows

Make	Price (£)	Miles per gallon MPG	Rating	Age (yrs)
Ford	15,450	31	*****	3
Chevy	12,450	27	***	4

Outline

- Issues to take into account designing an Infovis application
  - Task, subtasks, & problem (why)
  - Data (how)
  - Representation & Presentation (how)
  - Overview & interactive object/attribute selection (how)
  - Filtering (how)
  - Significant/reference data (how)
  - Navigational guidance (how)
  - Movement in the information space (how)

Representation & Presentation

- The first step is to map <u>data values</u> to <u>visual attributes</u>
- We will detail this activity in the next lectures
- Now we assume to use a common representation method, the <u>bargrams</u> (a partition of the data set)

Data set is partitioned through the values of an attribute (intervals on a ordinal/interval/rational/ scale or categorical values of a nominal scale)

Rational scale attribute

Price £k

10 - 12	12 - 14	16 - 18

Size proportional to the cardinality of the subset

Representation for vague goals

- Nice looking car???
- Bargrams + pictures

Price £k

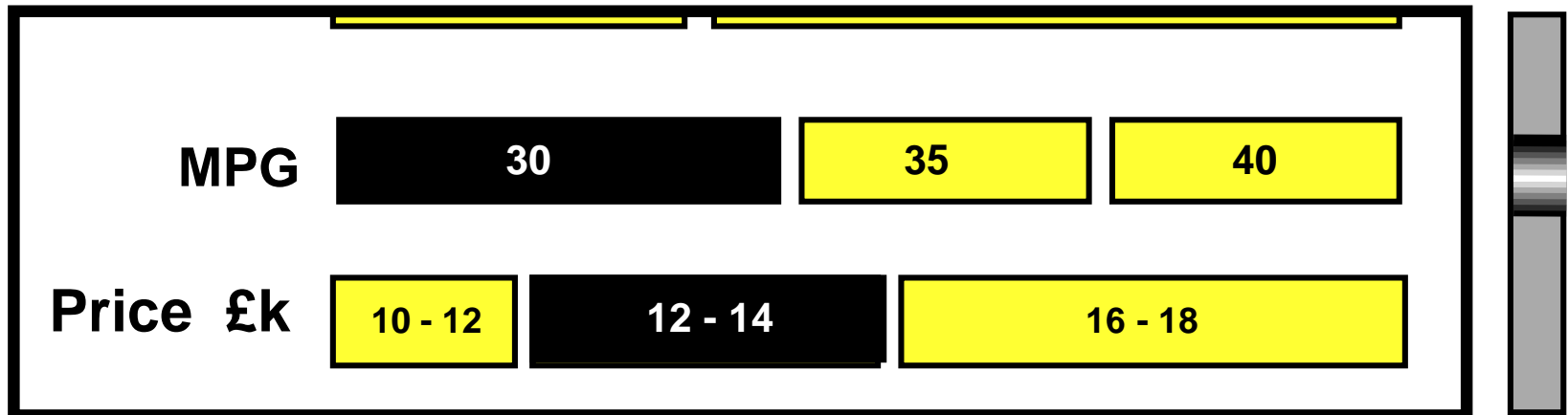| 10 - 12	12 - 14	16 - 18

Representation & Presentation

- The chosen representation is presented on the screen
- Limitations in <u>time</u> and <u>space</u> may require to adapt the representation
- Usual techniques (e.g., scrollbars) may partially solve the problem

MPG	30	35	40
Price £k	10 - 12	12 - 14	16 - 18

Presentation

- Several presentation techniques use zoom/distortion to better solve space limitations
- Here we use a techniques called "semantic zoom": the enlarged zone presents details that are not available at all elsewhere
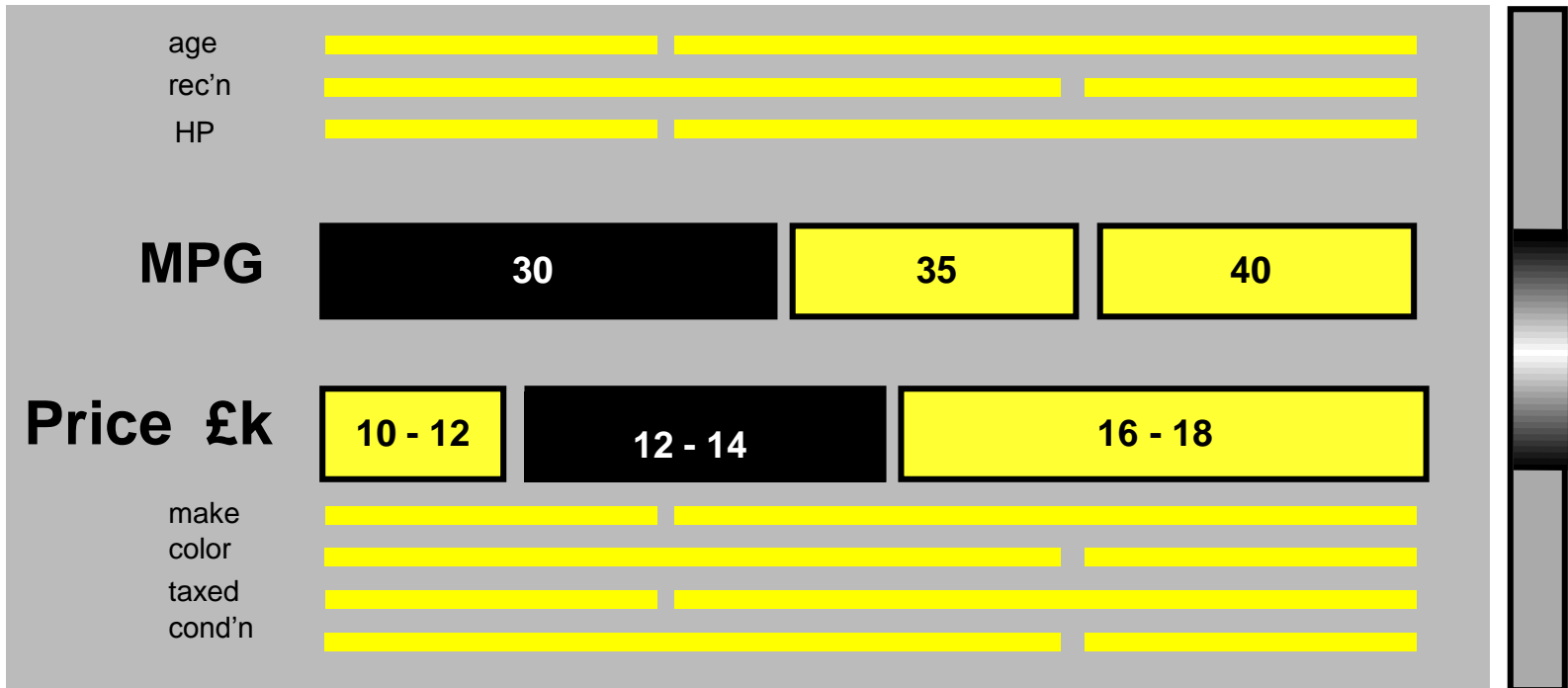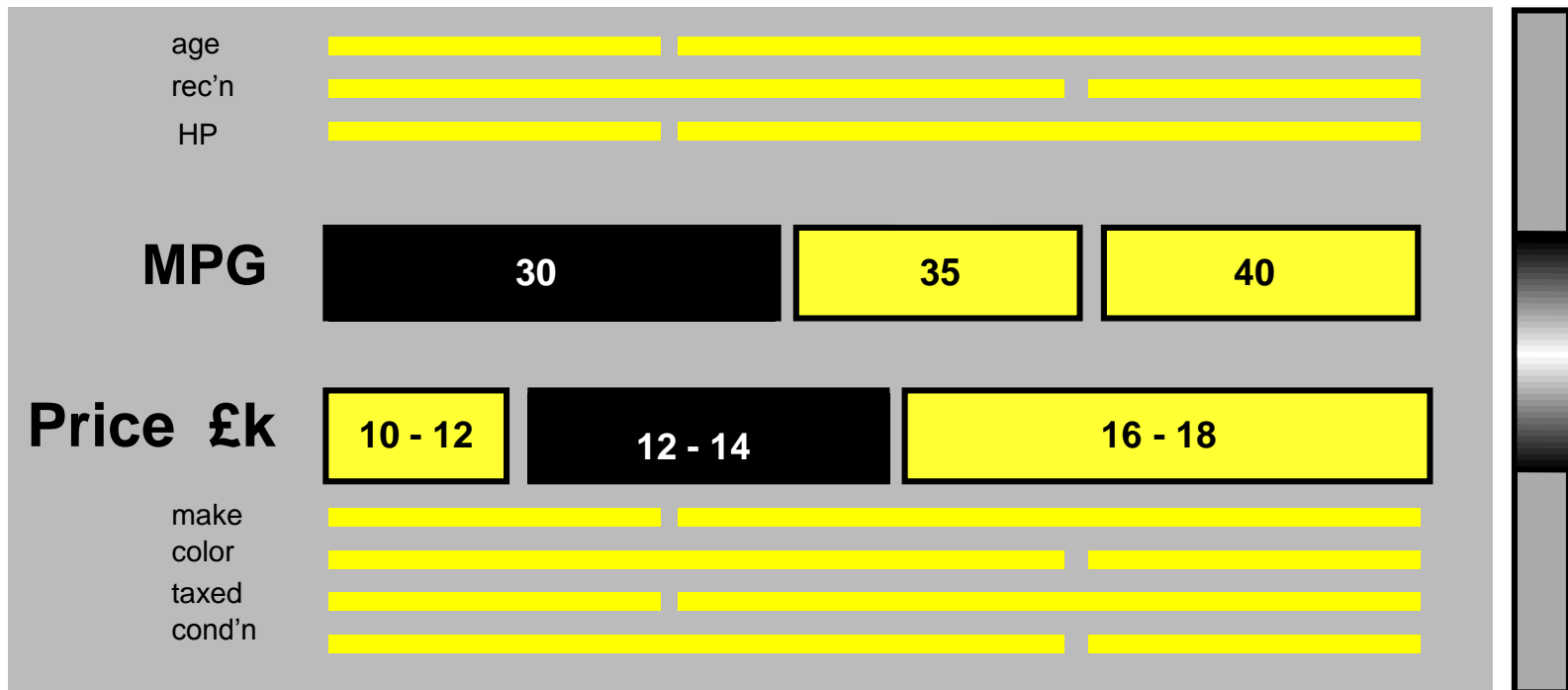
Outline

- Issues to take into account designing an Infovis application
    - Task, subtasks, & problem (why)
    - Data (how)
    - Representation & Presentation (how)
    - Overview & interactive object/attribute selection (how)
    - Filtering (how)
    - Significant/reference data (how)
    - Navigational guidance (how)
   - Movement in the information space (how)

Overview

- Bargrams allow for presenting a data overview
  - Qualitative awareness of one aspect
  - Quickly (even better, pre-attentively)
- Half of the cars consume 30 MPG !
- Price ranges in 10-18 and half of the cars cost 16-18

Interactive object/attribute selection

- Beside overview a means is needed for selecting **single** items

Price £k

| 10 - 12 | 12 - 14 | 16 - 18 |
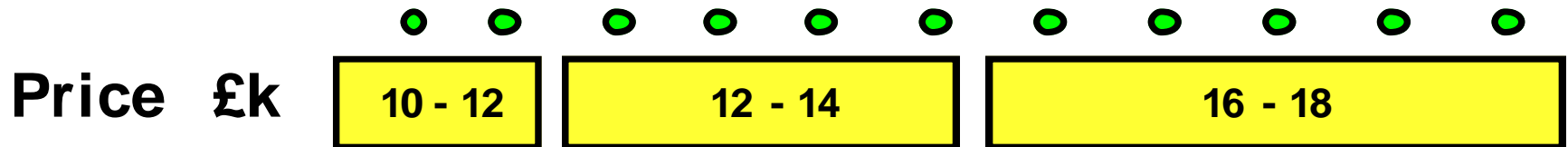
- Icons positioned above a bargram represent individual cars
- Selecting a single item allows for accessing details and managing single objects

Interactive object/attribute selection

Price £k

| 10 - 12	12 - 14	16 - 18
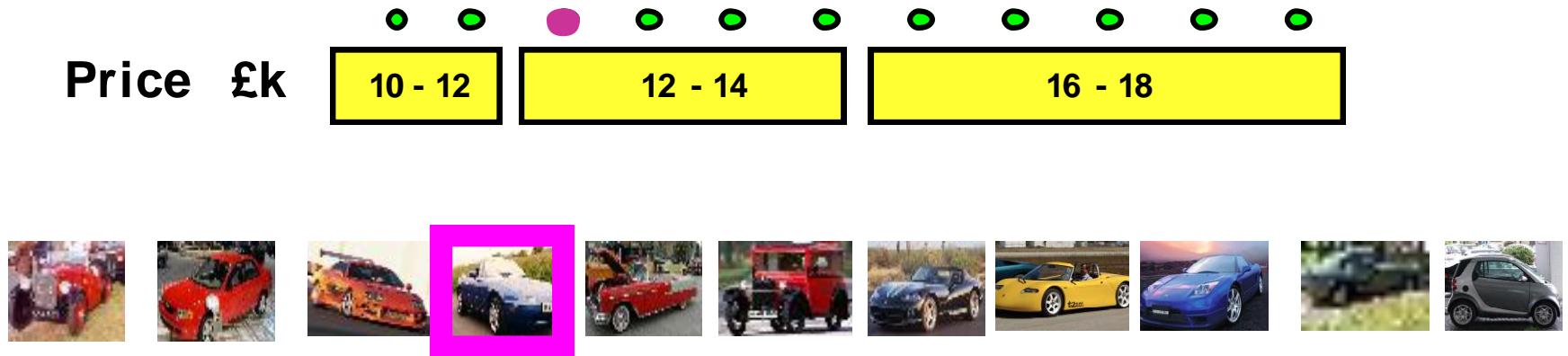
- Interaction and pre-attentive coding allow for exploring single items in efficient way
- Showing at the same time multiple attributes

Outline

- Types of Symbolic Displays (what)
- Issues to take into account designing an Infovis application
  - Task, subtasks, & problem (why)
  - Data (how)
  - Representation & Presentation (how)
  - Overview & interactive object/attribute selection (how)
  - Filtering (how)
  - Significant/reference data (how)
  - Navigational guidance (how)
  - Movement in the information space (how)

Filtering

- A system for **suppressing** not relevant data is required
- For instance it is possible to focus only on a specific price range
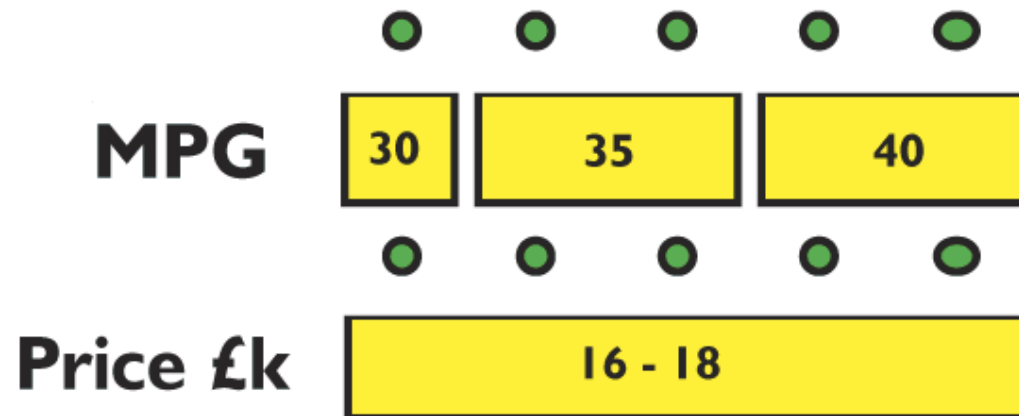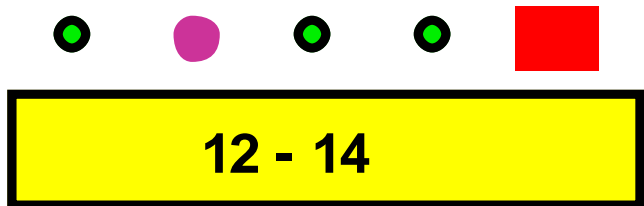

Ambiguity !

Outline

- Issues to take into account designing an Infovis application
    - Task, subtasks, & problem (why)
    - Data (how)
    - Representation & Presentation (how)
    - Overview & interactive object/attribute selection (how)
    - Filtering (how)
    - Significant/reference data (how)
    - Navigational guidance (how)
    - Movement in the information space (how)

Significant objects

- To simplify the exploration of the data set it is useful to mark/define some **significant** objects

- Significant means
  - Mmm, interesting, I'll look at it again…
  - This is my ideal car!
  - …

12 - 14

A car that is potentially of interest and worth remembering can be 'tagged' for later re-examination

12 - 14

An icon above a bargram can represent an 'ideal' (and possibly nonexistent) car to act as a point of reference
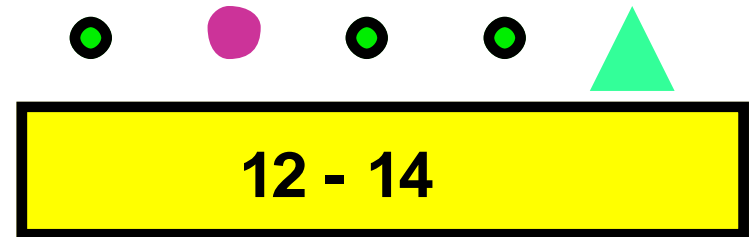The system can compute how existing cars are **far** form the ideal one

Outline

- Issues to take into account designing an Infovis application
  - Task, subtasks, & problem (why)
  - Data (how)
  - Representation & Presentation (how)
  - Overview & interactive object/attribute selection (how)
  - Filtering (how)
  - Significant/reference data (how)
  - Navigational guidance (how)
  - Movement in the information space (how)

Navigational guidance

Select *
From cars
Where price<16 AND MPG > 35 AND Color="Red"

….. NO cars !

So what?

Navigational guidance

What if question:
- what if I had 1k more to spend?

Obviously such questions can be answered changing the filtering option on price, including more expensive cars

But that must be done again for each new what if question:
- what if I decide to consider a lower MPG ?

A very helpful solution is to make the user aware that some **potential** items are around, e.g., showing items that does not satisfy **only one** requirement
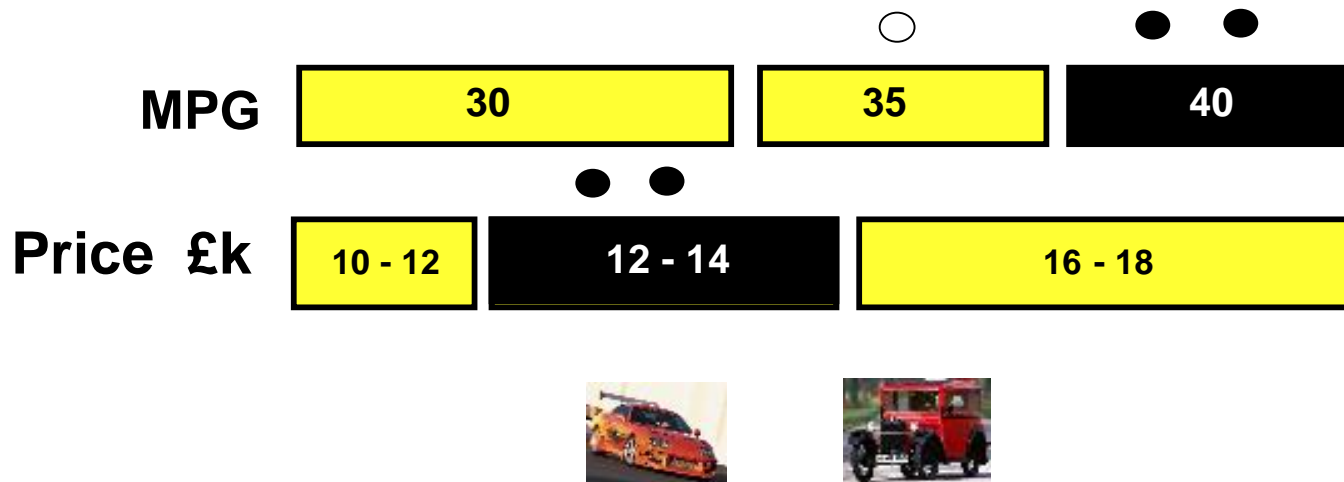
MPG

| 30 | 35 | 40 |

Price £k

| 10 - 12 | 12 - 14 | 16 - 18 |

Outline

- Issues to take into account designing an Infovis application
    - Task, subtasks, & problem (why)
    - Data (how)
    - Representation & Presentation (how)
    - Overview & interactive object/attribute selection (how)
    - Filtering (how)
    - Significant/reference data (how)
    - Navigational guidance (how)
    - Movement in the information space (how)

Movement in the information space

What is the main difference between Minard's visualization and this system?

?
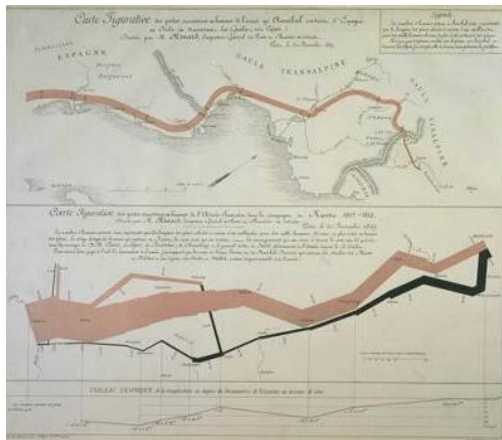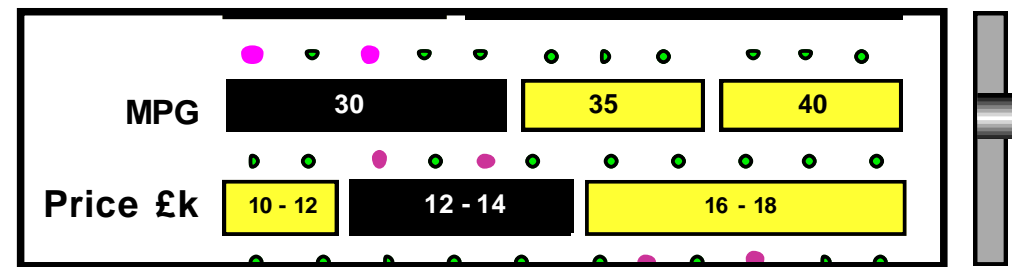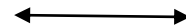
Interaction!
Computers allows for changing, switching, filtering, arranging, tagging, etc.
This is a key issue!

EzChooser
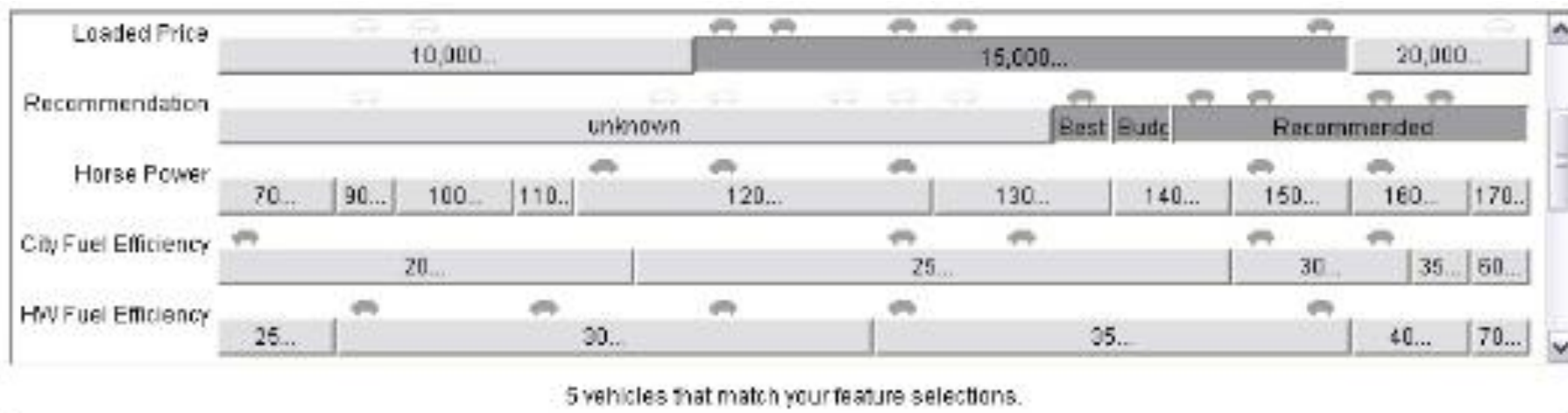
Summarizing

DATA

REPRESENTATION of data

PRESENTATION of the
represented data

INTERACTION to select
the required view of data

PERCEPTION

INTERPRETATION

**HIGHER-ORDER
COGNITIVE
PROCESSES**

Problem (re)formulation
Evaluation of options
Strategy formulation
Internal modelling
etc.

Decision making

Spence's book

Ware's book
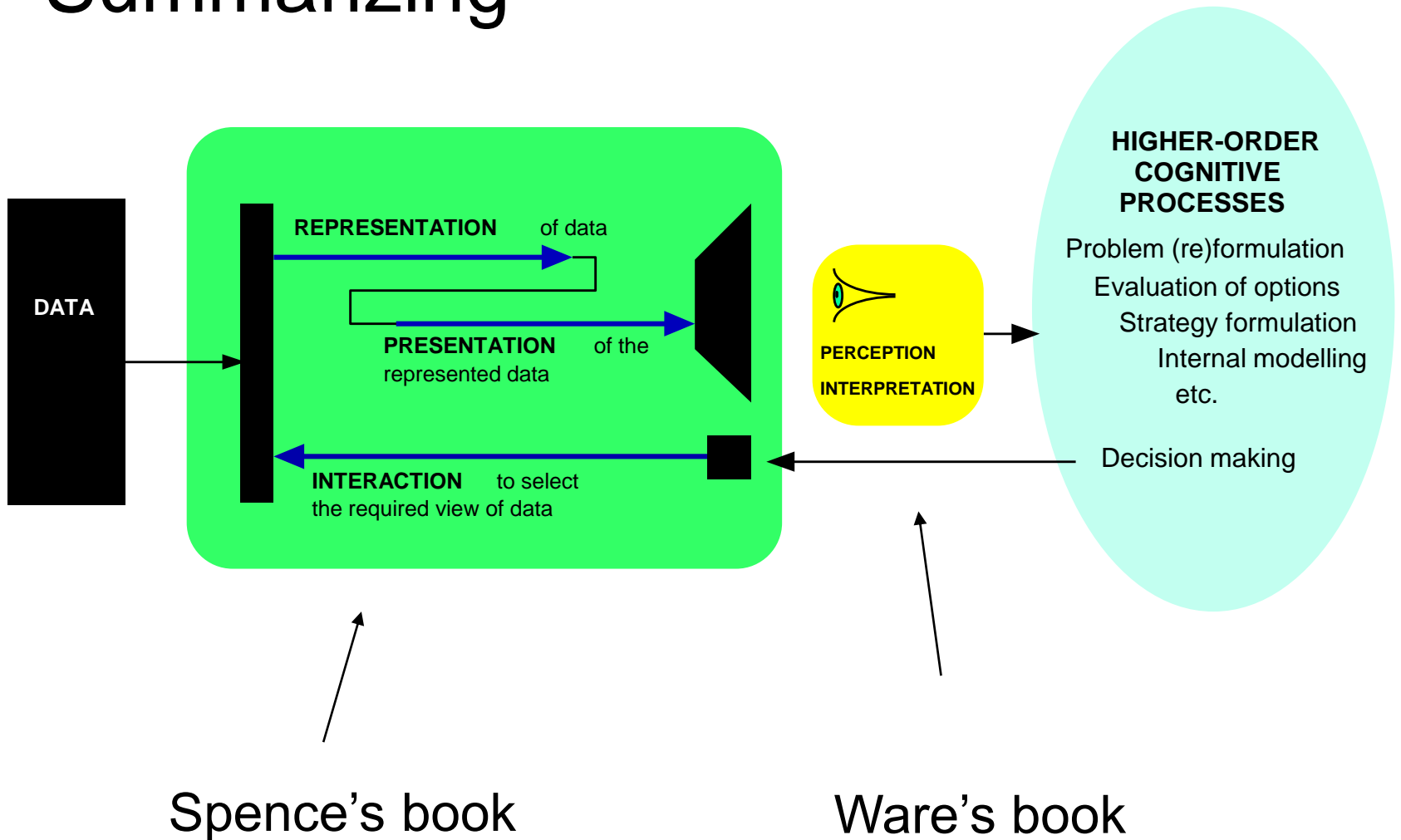
Case Study:
The Journey of the TreeMap

Trees waste horizontal space !
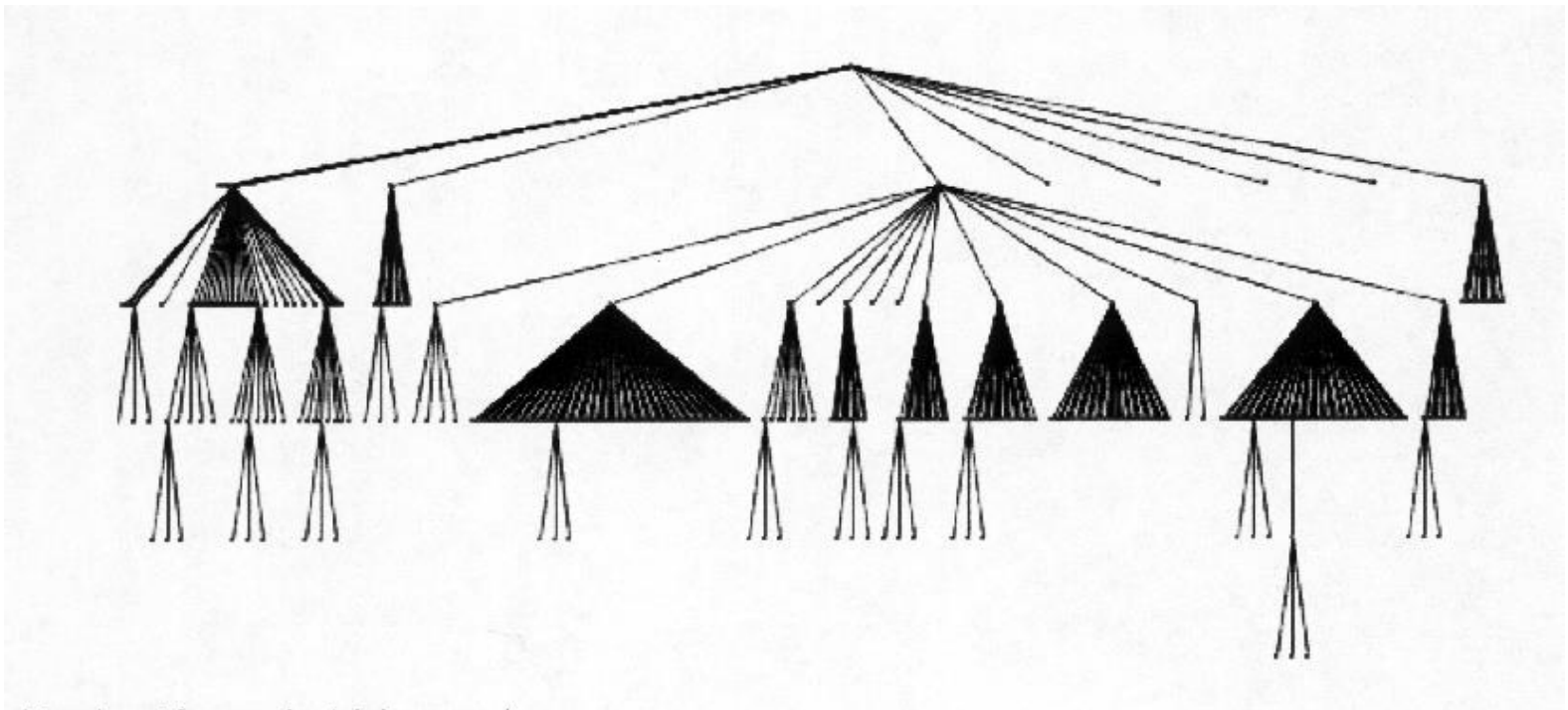
Case Study:
The Journey of the TreeMap

(Johnson & Shneiderman '91)
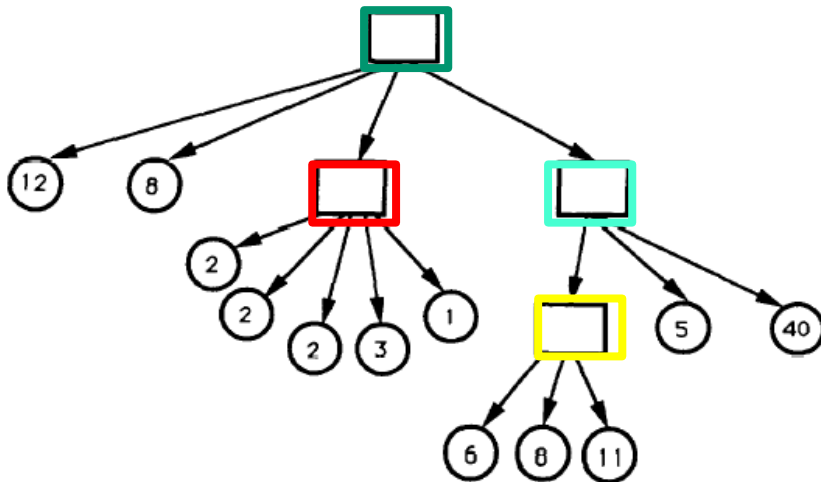
Fig. 1. Typical 3-level tree structure with numbers indicating size of each leaf node.
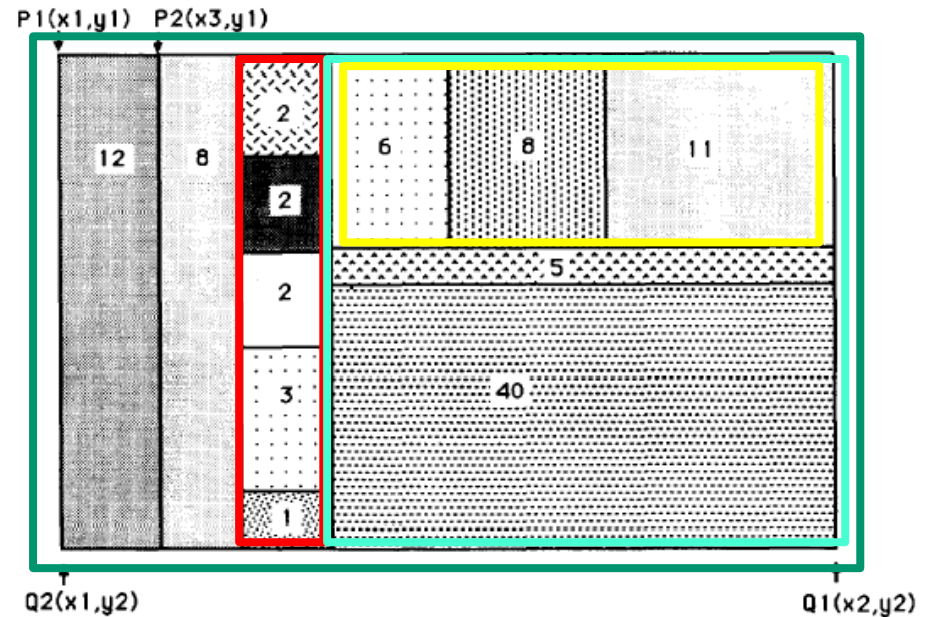
Fig. 2. Tree-map of Figure 1.

Take your time while describing a visualization !

Case Study:
The Journey of the TreeMap

- The TreeMap **(Johnson & Shneiderman '91)**
- Idea:
  - Show a hierarchy as a 2D layout
  - Fill up the space with rectangles representing objects
  - Nested rectangles indicated levels of hierarchy
  - Size on screen indicates relative size of underlying objects.

Visualizing the hierarchical content of an Hard disk

Colors code the types of files
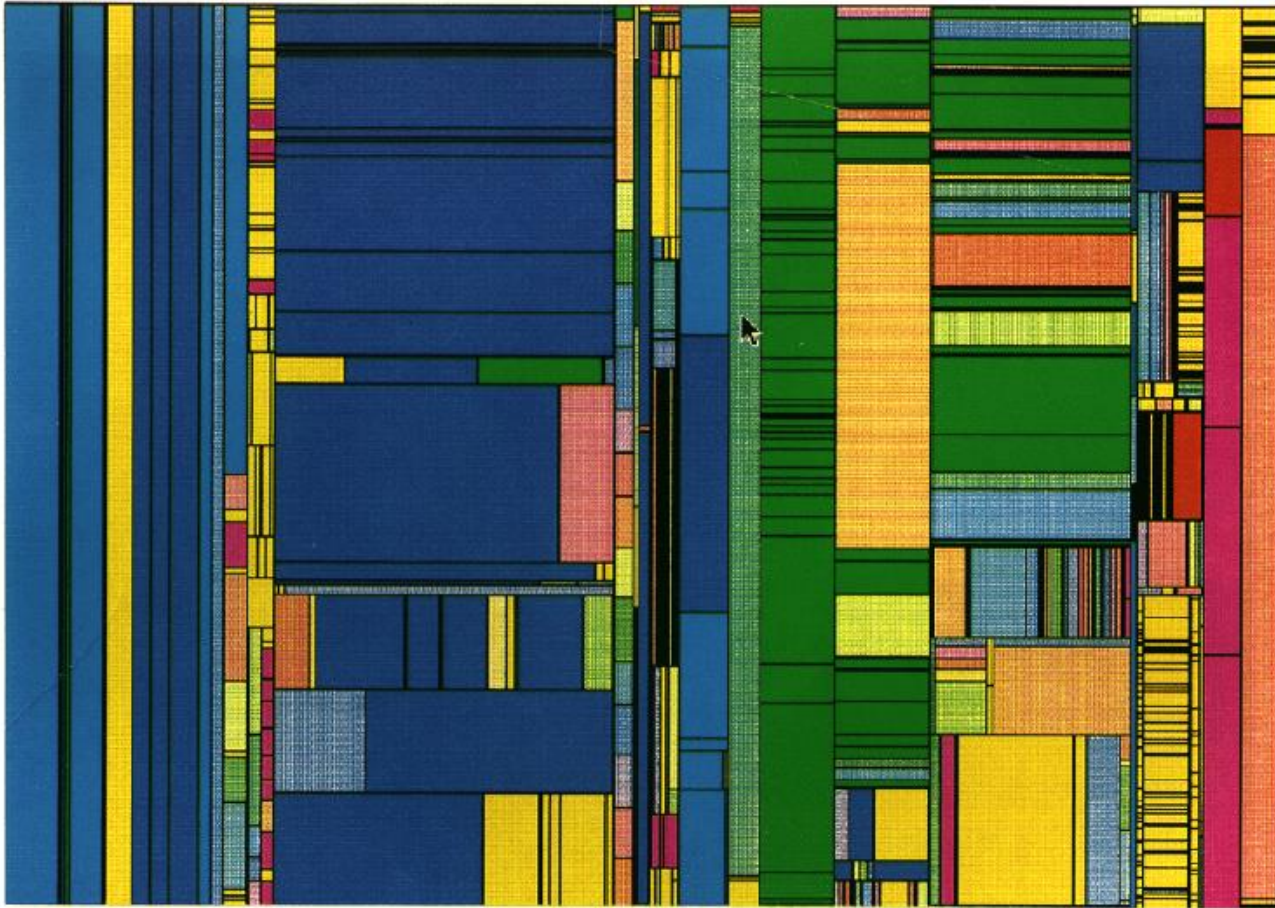
Fig. 4.   850 files at four levels with color coding by tile type. File name pops up when cursor rests on a file.

VM Storage

eWorld · eWorld Color Art · Com · Audio Help · HyperCard Player · Macinta · Speech · AppleS · TTS Victor · TTS Bruce · QuicKeys™ · Plain · Recognit · Speech · TTS Agnes · Introdu · Test Dr · Install · PowerTalk · Inst Using · M PowerTalk Sh · Insta · Quick · User G · Ma · Mail*Li · Fax · Insta · Quick · User Guide · Mail*L · SuToPi · Power Fax™ · Installer · Power Fax™ · STF Applic · Installer · QuickDraw™ · Hoefler · ATM™ GX · QuickDraw™ · Skia

bankscapemd · bankscapoe4 · goodtempom · MacroMind P · WBdemo-for · minidemoformakingprints · slides/pictures

Director 4.0 · Director 4.0 Help · Bella Rosa · Liq · Rob · The Ha · QuickTim · Backgrou · R · Sony VTR · Behind the Scenes · Navigator · BEGIN.DIR · FURNITUR.DIR · HELP.DIR · LEARNING.DIR · LUCK.DIR · PHILANTH.DIR · MECH · Noh_Tale · NTBranch · Shared.DIR · Chair · Cur · Rollover · Use · INT · wallcov · Hypertext · Ink_FX · shared.dir · Weeping

Adobe Photoshop® 2.01 · GIFCony · PageMaker 4.2 · Came · Simp

Clipboard · Finder · System · PM42R · App · Fin · Graphing Calc · Jig · Aldus I · Diction · Diction · Desk · Maci · MacTC · PCE · Find · Macinto Mou Opening Animatio Shared · Macintosh Guide · QuickTime™ · Palati · Times

Print Util · printworl · Gloss · Norton Util Nort SP · Maci · ETC · ETC · MathT · Tattle · Apple DocViewer · Cou · Hel Pal · Tim · Norton · Nort SP

Equa · Exce · Gram Mac · Spel · Tool · U.S. U.S. Englis U.S. · Wind · Word · Word 5 Wo Wo Wo Wor · Tattle · Microsoft

Name
Path Information

Unknown | Text | Graphics | Archives/Stacks | Programming | Applications | System

Size · % Total · Type · Creator · Creation Date · Modification Date

Treemap Problems

- ## Too disorderly
  - What does adjacency mean?
  - Aspect ratios uncontrolled leads to lots of skinny boxes that <u>clutter</u>

- ## Hard to understand
  - Must mentally convert nesting to hierarchy descent

Successful Application of Treemaps

- Think more about the use
  - Break into meaningful groups
- Make appearance more usable
  - Fix these into a useful aspect ratio
- Use visual properties properly
  - Use color to distinguish meaningfully
    - Use only two colors:
      - Can then distinguish one thing from another
    - When exact numbers aren't very important
- Provide excellent interactivity
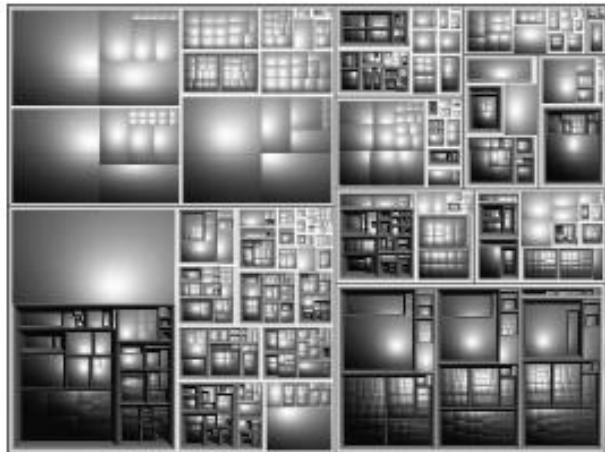 - Access to the real data

Squarified Treemaps
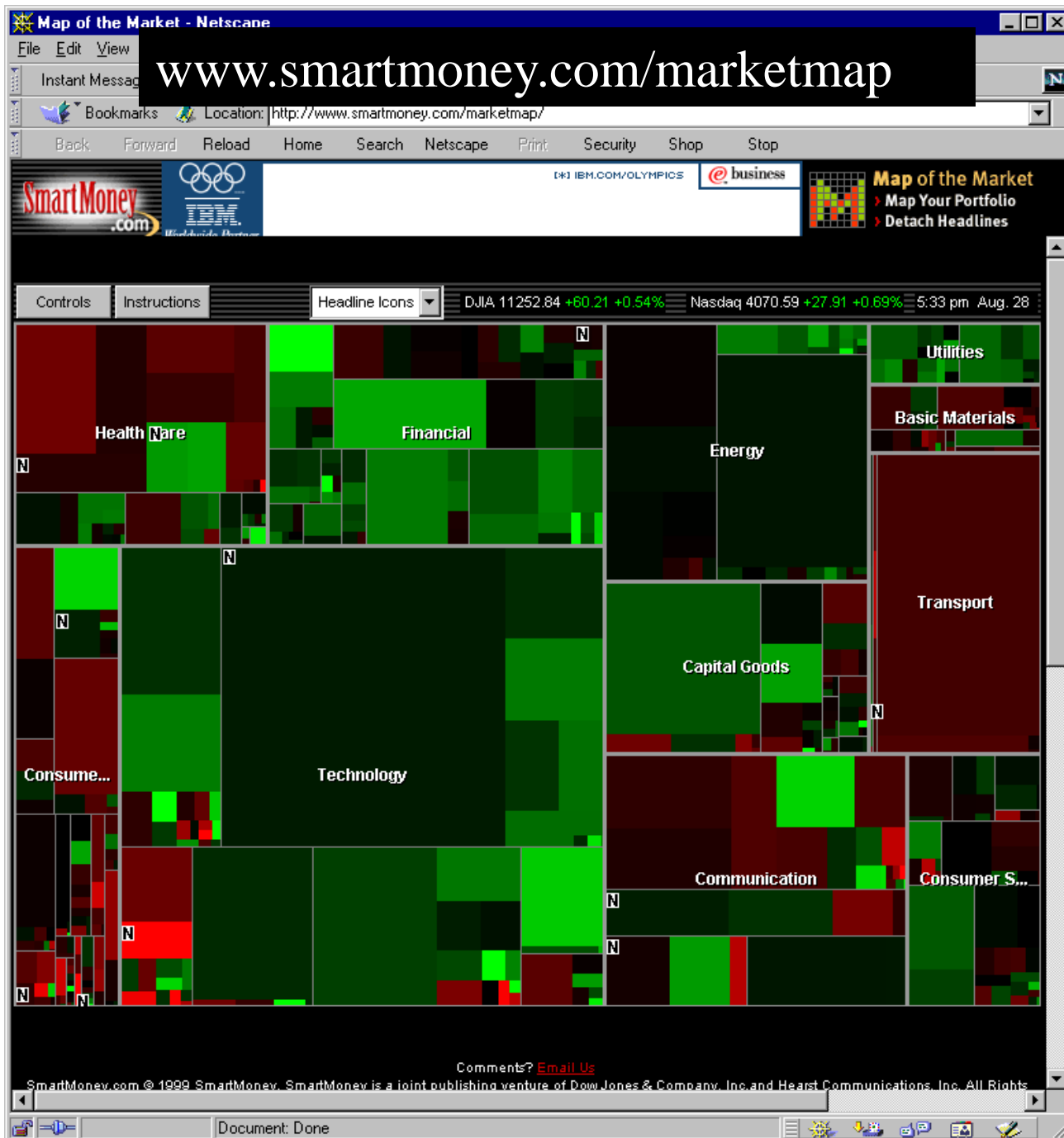
Bruls, Huizing, van Wijk, 1999

(a) File system
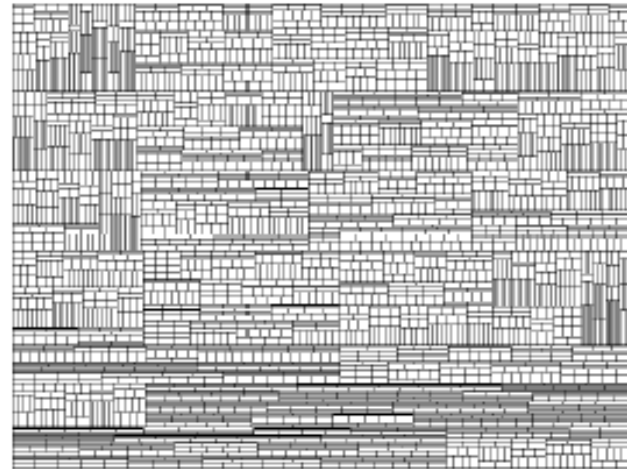
(b) Organization

(a) File system

(b) Organization

www.smartmoney.com/marketmap

Commercial Applet ☹

Treemaps in Peets site

Analysis vs. Communication

- MarketMap's use of TreeMaps allows for sophisticated analysis
- Peets' use of TreeMaps is more for presentation and communication
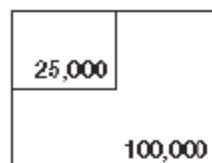- This is a key contrast

NY Times uses treemaps!

Truck Sales Slip, Tripping Up Chrysler

Amanda Cox, Heather Fairfield, NY Times

lead of Toyota
ut as American
continued to
the country.

READING THE CHART

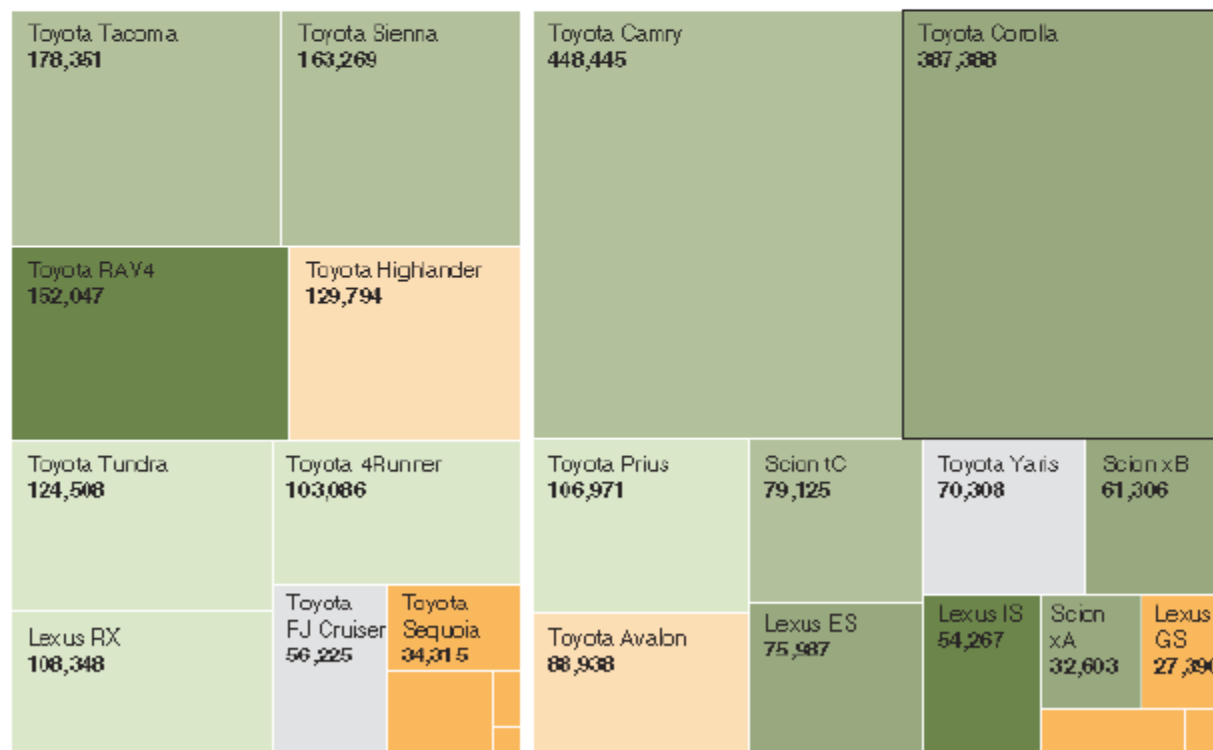Boxes are scaled proportionally according to number of cars sold in 2006

25,000

100,000

Change in sales from 2005 to 2006

Below | Above industry average

−10%   −2.6   0   +10   +100

No 2005 sales

Many of these vehicles were introduced in 2005.

◄ **TRUCKS, VANS, S.U.V.'S | CARS** ►

Toyota Tacoma
178,351

Toyota Sienna
163,269

Toyota Camry
448,445

Toyota Corolla
387,388

Toyota RAV4
152,047

Toyota Highlander
129,794

Toyota Tundra
124,508

Toyota 4Runner
103,086

Toyota Prius
106,971

Scion tC
79,125

Toyota Yaris
70,308

Scion xB
61,306

Lexus RX
108,348

Toyota FJ Cruiser
56,225

Toyota Sequoia
34,315

Toyota Avalon
88,938

Lexus ES
75,987

Lexus IS
54,267

Scion xA
32,603

Lexus GS
27,390

Toyota **+12.5%**

Trucks/vans/S.U.V.'s 1.1 million

Cars 1.5 million

Toyota rolled out a new version of the Camry, and once again it was the country's best-selling car.

Toyota Corolla

Corolla sales also jumped, along with gas prices. Toyota could not escape the decline in sales of supersized S.U.V.'s like its Sequoia.

Amanda Cox, Heather Fairfield, NY Times