Visual Analytics

Giuseppe Santucci

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

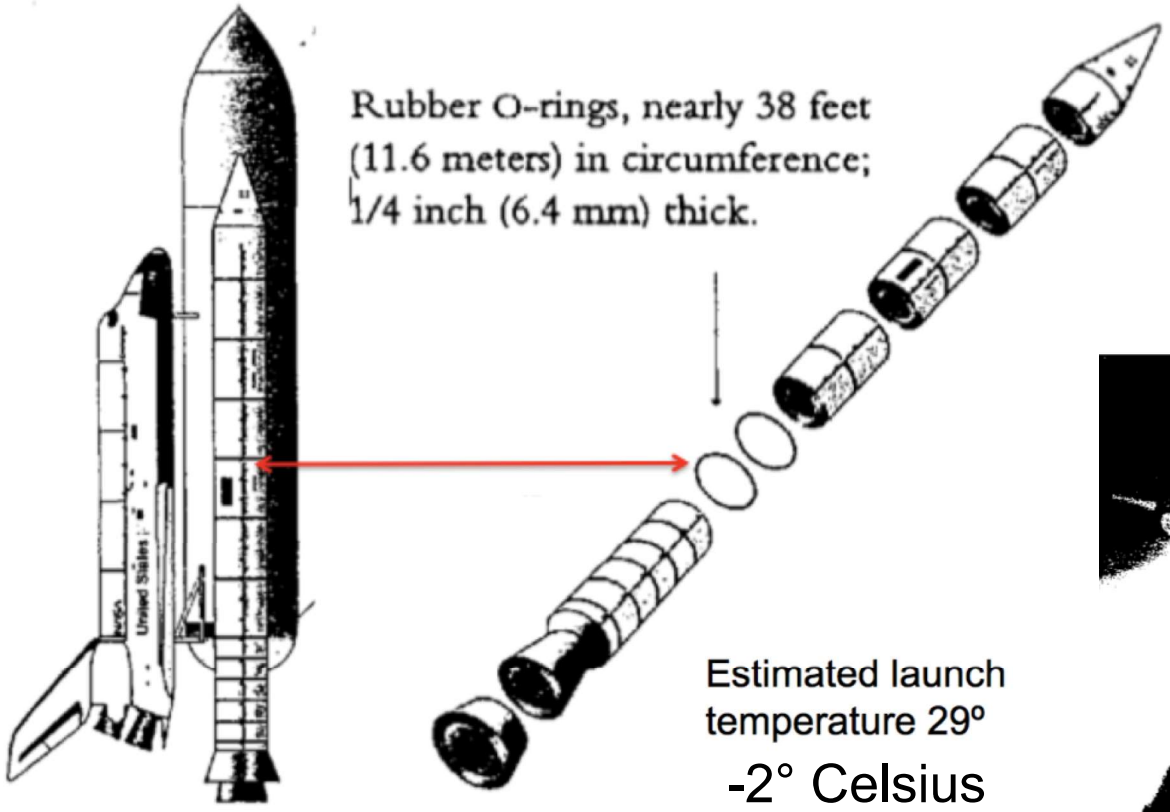
Thanks to John Stasko, Robert Spence, Ross Ihaka, Marti Hearst, Kent Wittemburg

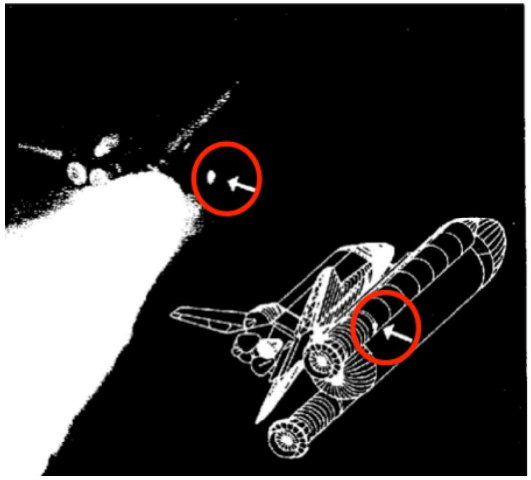
Outline

- 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 1986 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.

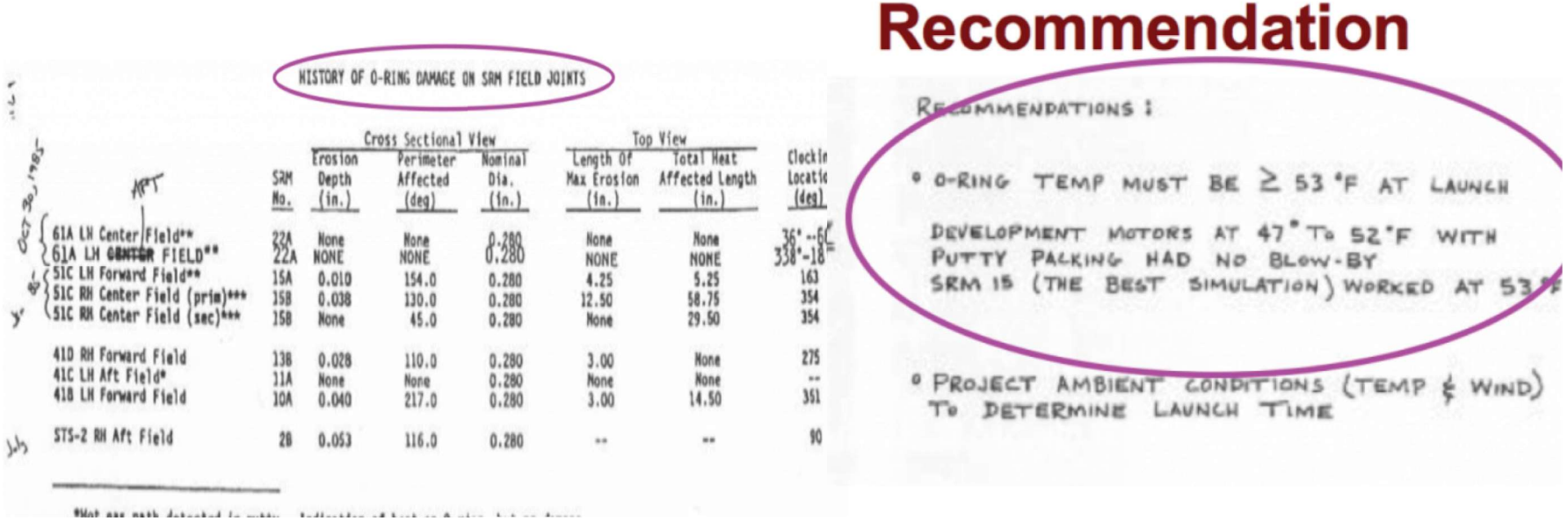




http://www.math.yorku.ca/SCS/Gallery/

The Challenger disaster

- NASA staff had analysed the data, plotting ambient temperature and number of O-ring failures on test rockets
- A lot of faxes were around the day before the lunch, looking for correlation between temperature and o-ring failures.
- But they had excluded observations where no O-rings failed, believing that they were uninformative.
- Even if the conclusion was clear



HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clockin Locatio (deg)
61A LH Center Field**	22A	None	None	0.280	None	None	36°-6
61A LH CENTER FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338°-18
51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	5.25	163
51C RH Center Field (prim)***	15B	0.038	130.0	0.280	12.50	58.75	354
51C RH Center Field (sec)***	15B	None	45.0	0.280	None	29.50	354
41D RH Forward Field	13B	0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	11A	None	None	0.280	None	None	--
41B LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	2B	0.053	116.0	0.280	--	--	90

Recommendation

RECOMMENDATIONS :

- O-RING TEMP MUST BE ≥ 53 °F AT LAUNCH

  DEVELOPMENT MOTORS AT 47° TO 52°F WITH PUTTY PACKING HAD NO BLOW-BY
  SRM 15 (THE BEST SIMULATION) WORKED AT 53°F

- PROJECT AMBIENT CONDITIONS (TEMP & WIND) TO DETERMINE LAUNCH TIME

Disaster was no a surprise!

that cool temperatures were not linked to O-ring problems.

Thus the *exact cause* of the accident was intensely debated during the evening before the launch. That is, for hours, the rocket engineers and managers considered the question: *Will the rubber O-rings fail catastrophically tomorrow because of the cold weather?* These discussions concluded at midnight with the decision to go ahead. That morning, the Challenger blew up 73 seconds after its rockets were ignited.

REGARDLESS of the indirect cultural causes of the accident, there was a clear proximate cause: an inability to assess the link between cool temperature and O-ring damage on earlier flights. Such a pre-launch analysis would have revealed that this flight was at considerable risk.[27]

Weak data presentation and visualization (?)

On the day before the launch of Challenger, the rocket engineers and managers needed a quick, smart *analysis* of evidence about the threat of cold to the O-rings, as well as an effective *presentation* of evidence in order to convince NASA officials not to launch. Engineers at Thiokol prepared 13 charts to make the case that the Challenger should *not* be launched the next day, given the forecast of very chilly weather.[28] Drawn up in a few hours, the charts were faxed to NASA and discussed in two long telephone conferences between Thiokol and NASA on the night before the launch. The charts were unconvincing; the arguments against the launch failed; the Challenger blew up.

?

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clocking Location (deg)
61A LH Center Field**	22A	None	None	0.280	None	None	36° --66°
61A LH CENTER FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338°-18°
51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	5.25	163
51C RH Center Field (prim)***	15B	0.038	130.0	0.280	12.50	58.75	354
51C RH Center Field (sec)***	15B	None	45.0	0.280	None	29.50	354
41D RH Forward Field	13B	0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	11A	None	None	0.280	None	None	--
41B LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	2B	0.053	116.0	0.280	--	--	90

*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.

??

Blow By History

SRM-15 worst Blow-By

- 2 case joints (80°), (110°) Arc
- Much worse visually than SRM-22

SRM 22 Blow-By

- 2 case joints (30-40°)

SRM-13A, 15, 16A, 18, 23A 24A

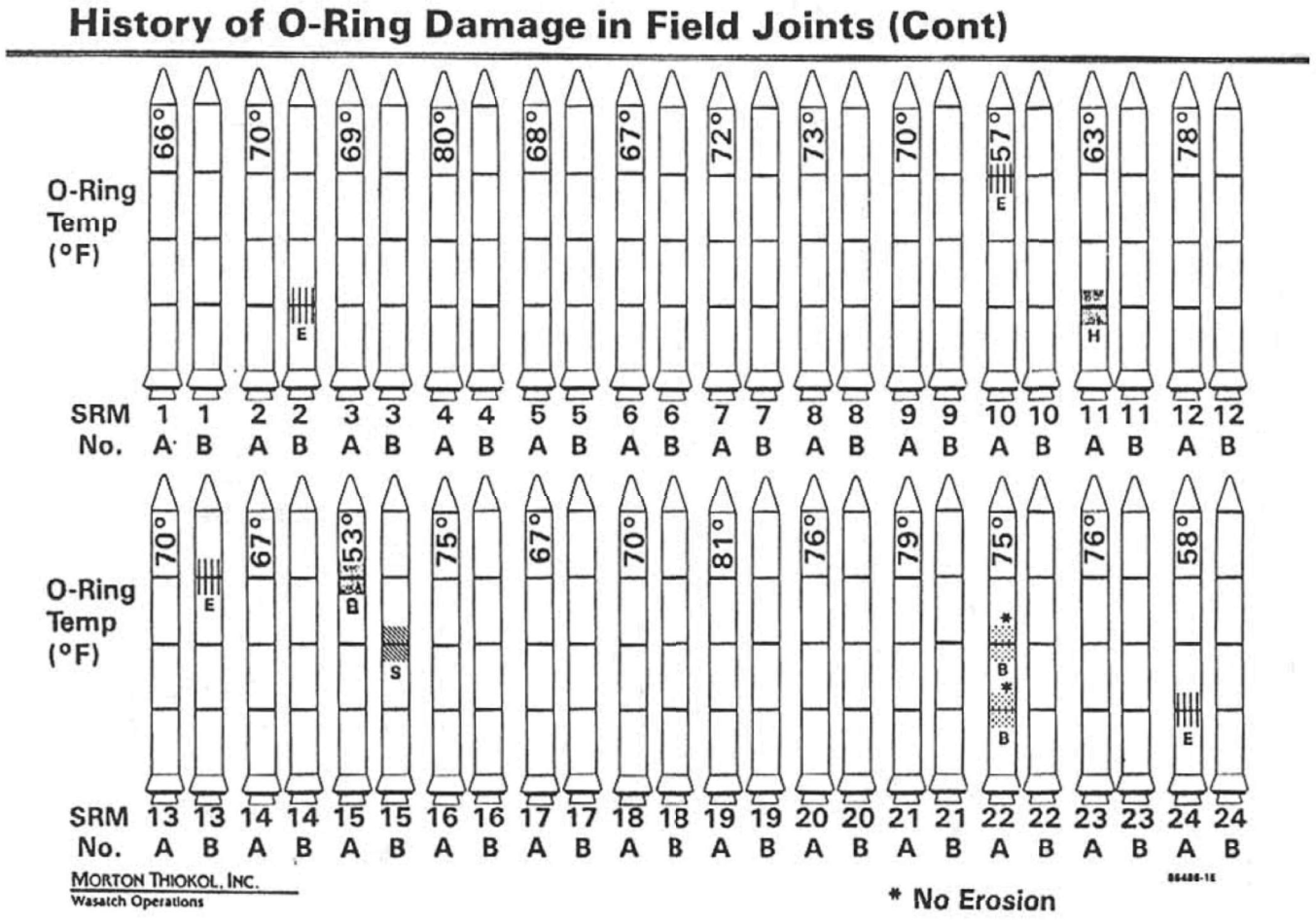
- Nozzle Blow-By

History of O-ring Temperatures (Degrees-F)

Motor	MBT	Amb	O-ring	Wind
DM-1	68	36	47	10 mph
DM-2	76	45	52	10 mph
QM-3	72.5	40	48	10 mph
QM-4	76	48	51	10 mph
SRM-15	52	64	53	10 mph
SRM-22	77	78	75	10 mph
SRM-25	55	26	29 27	10 mph 25 mph

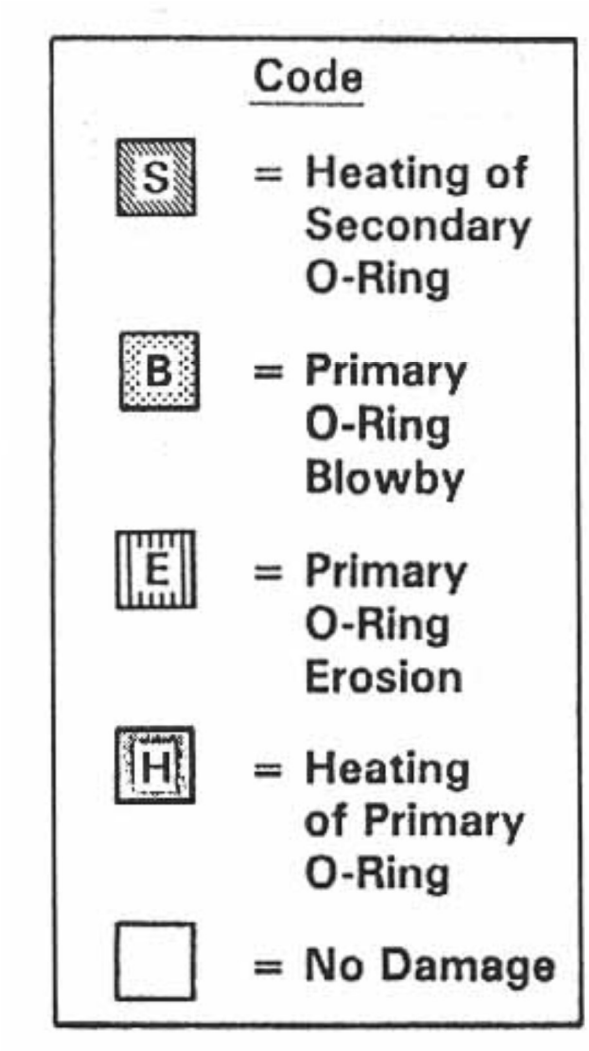
Bad infovis/number viz continued during the investigation

- The figure shows O-ring failures at different launch temperatures collected by NASA before 1986 (24 flights)



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

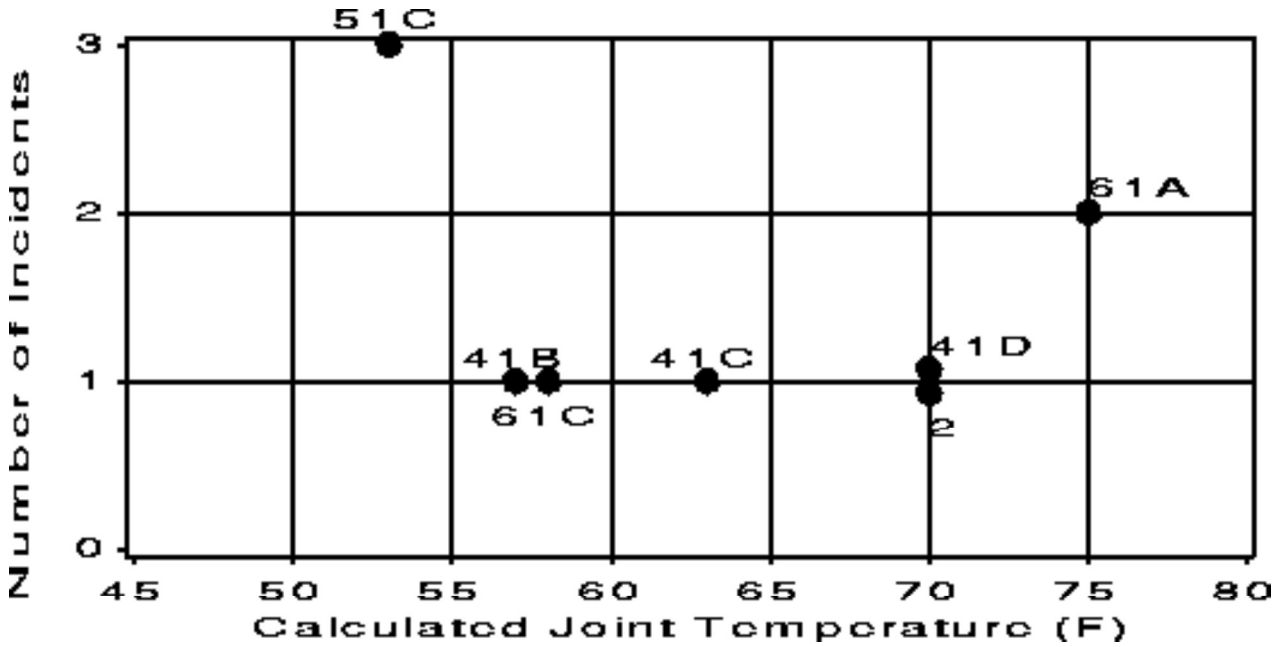
Legend on the next page...



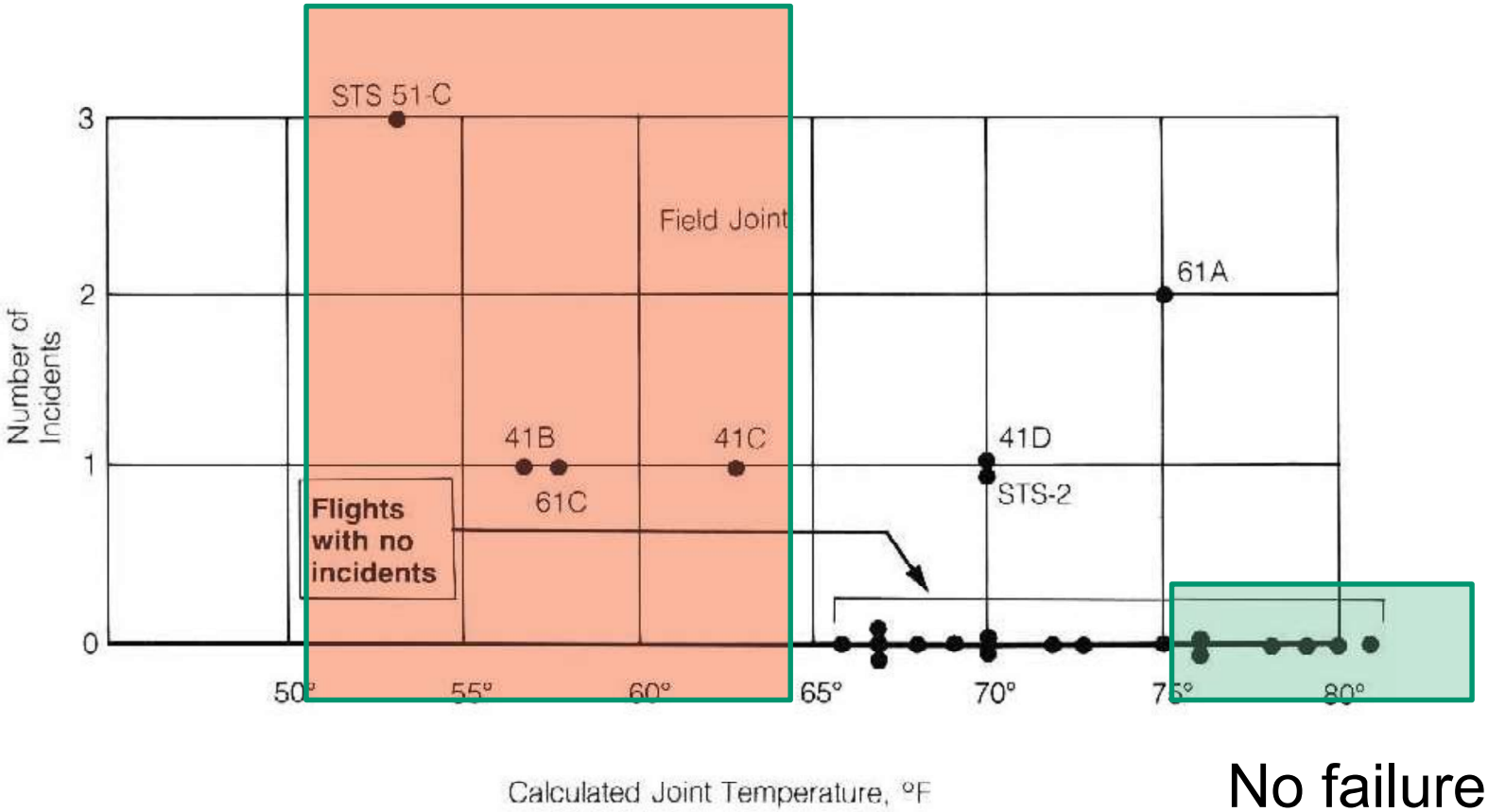
http://www.math.yorku.ca/SCS/Gallery/

The visualization of the fax data **during the investigation**

- Making decisions also on fax data was not easy (and engineers were under pressure...). They perceived a risk for low temperature but they were no able to quantify it nor to convince NASA managers!
- So, the political decision was: GO!
- And viz during investigation were bad as well...



The whole data (scatterplot correlation)

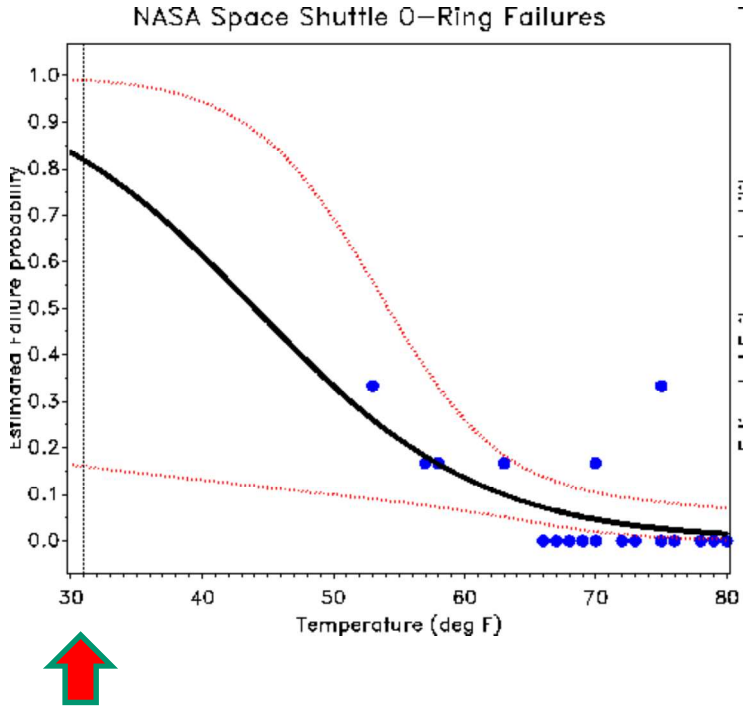


Failure on any flight below 65° F

No failure above 75° F

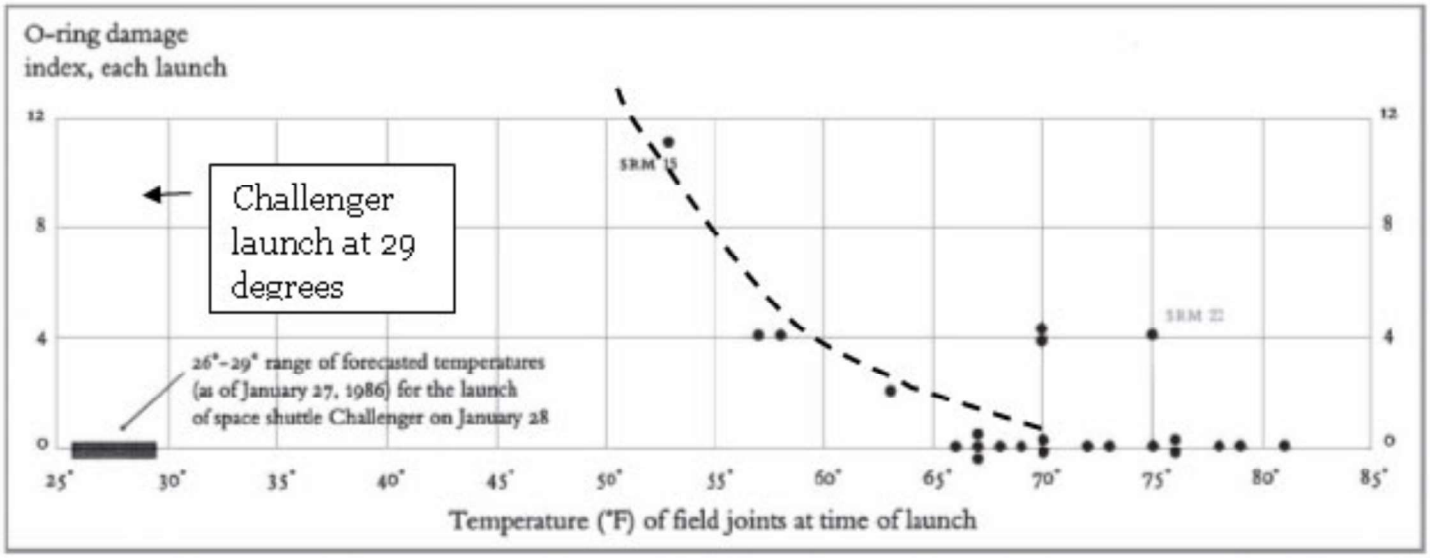
The Challenger disaster

- Reanalysis of the O-ring data using a regression model. This provides a predicted extrapolation (black curve) of the probability of failure to the low (31° F) 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?**



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 better graph might have signaled that something was wrong
- The following graph should have caused any engineer to conclude that there were excessive risks associated with 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

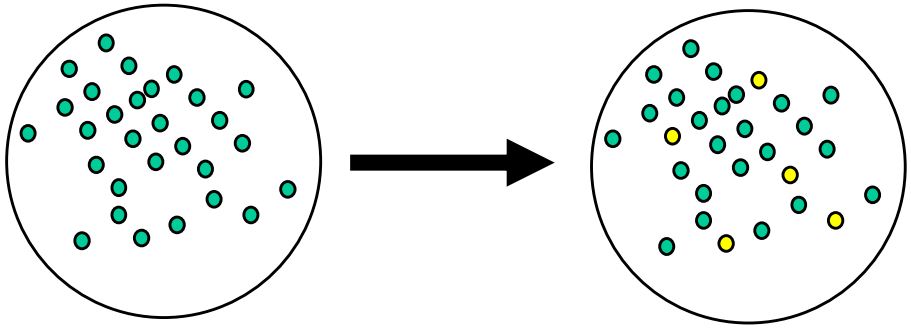
- Different Infovis applications share common concepts, like user task, data, etc.
- This lecture will introduce them using a running example
- A very common Infovis application is a system that allows the user to select one item to buy (☺); we use, as a running example, a system that allows to buy a 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 insights 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, we 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 high degree of interactivity

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 car dataset is as follows

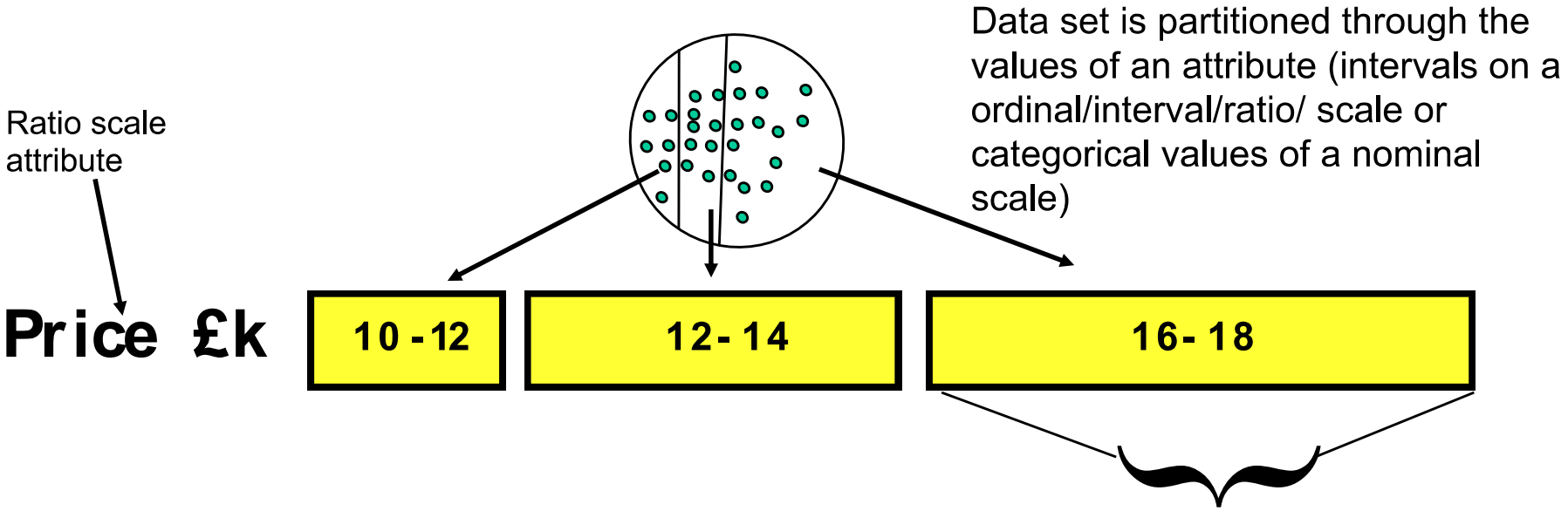
Make	Price (£)	Miles per gallon MPG	Rating	Age (yrs)
Ford	15,450	31	*****	3
Chevy	12,450	27	***	4

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 data values to visual attributes
- We will detail this activity in the next lectures
- Now we assume to use a common representation method, the bargrams (a partition of the data set)



Representation for vague goals

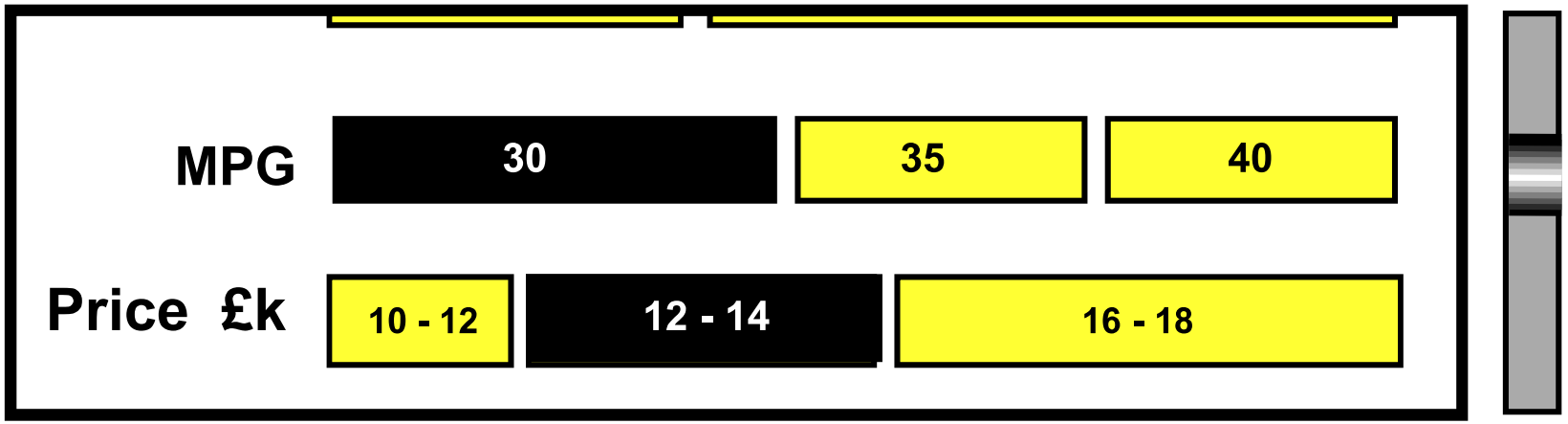
- Nice looking car???
- Bargrams + pictures

Price £k

10 - 12	12 - 14	16 - 18

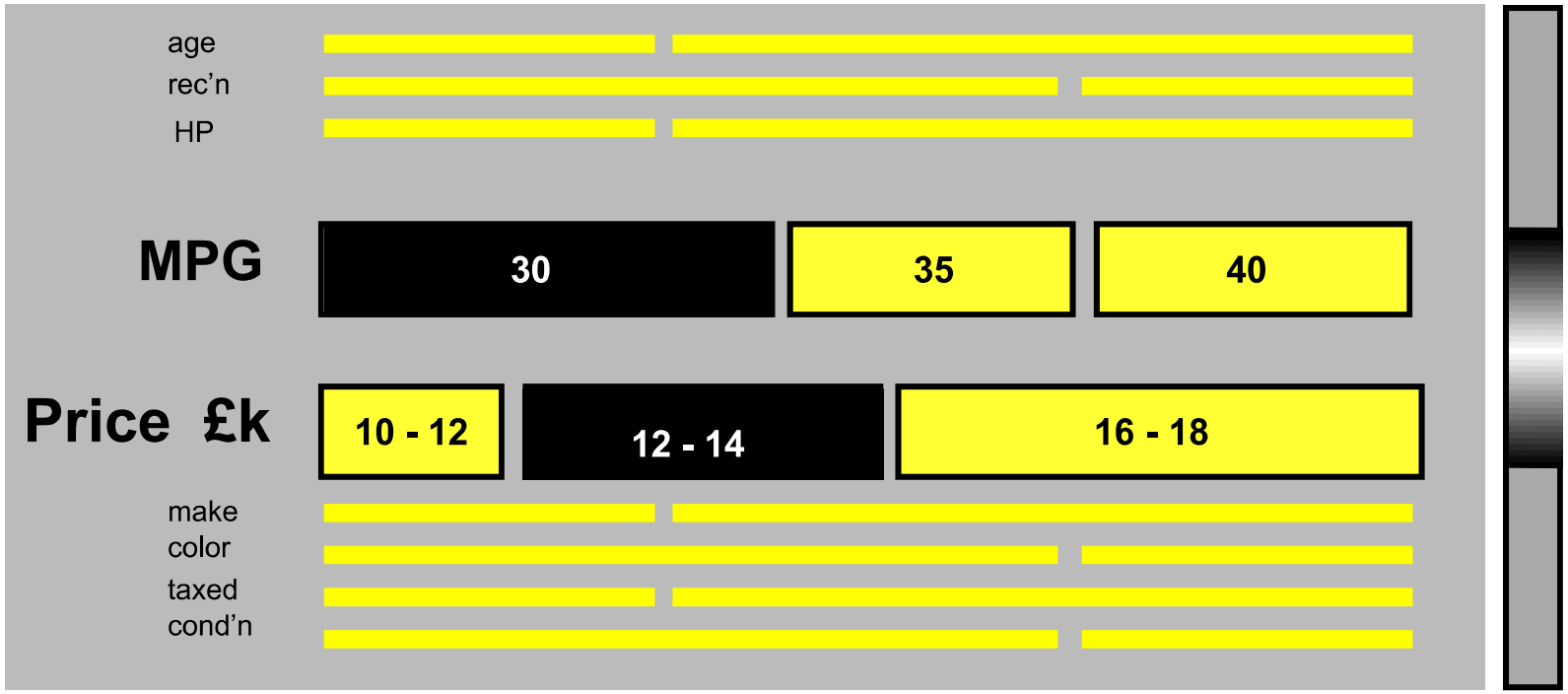
Representation & Presentation

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



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

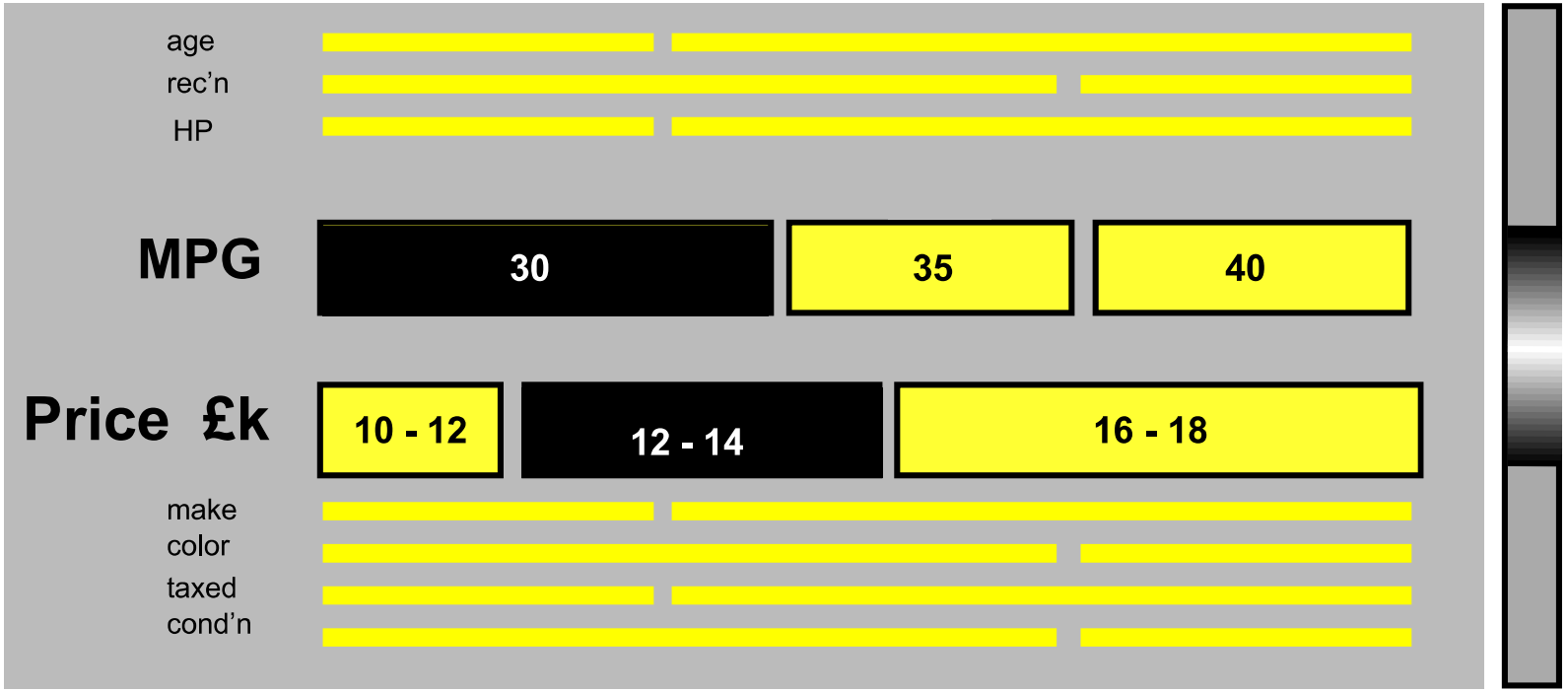


Outline

- 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)

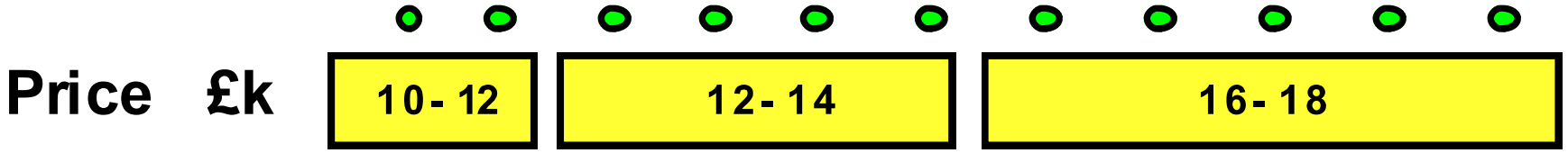
Overview

- 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 10k-18k and half of the cars cost 16k-18k



Interactive object/attribute selection

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



- Icons positioned above a bargram represent individual cars
- Selecting a single item allows for accessing details and managing single objects

Interactive object/attribute selection

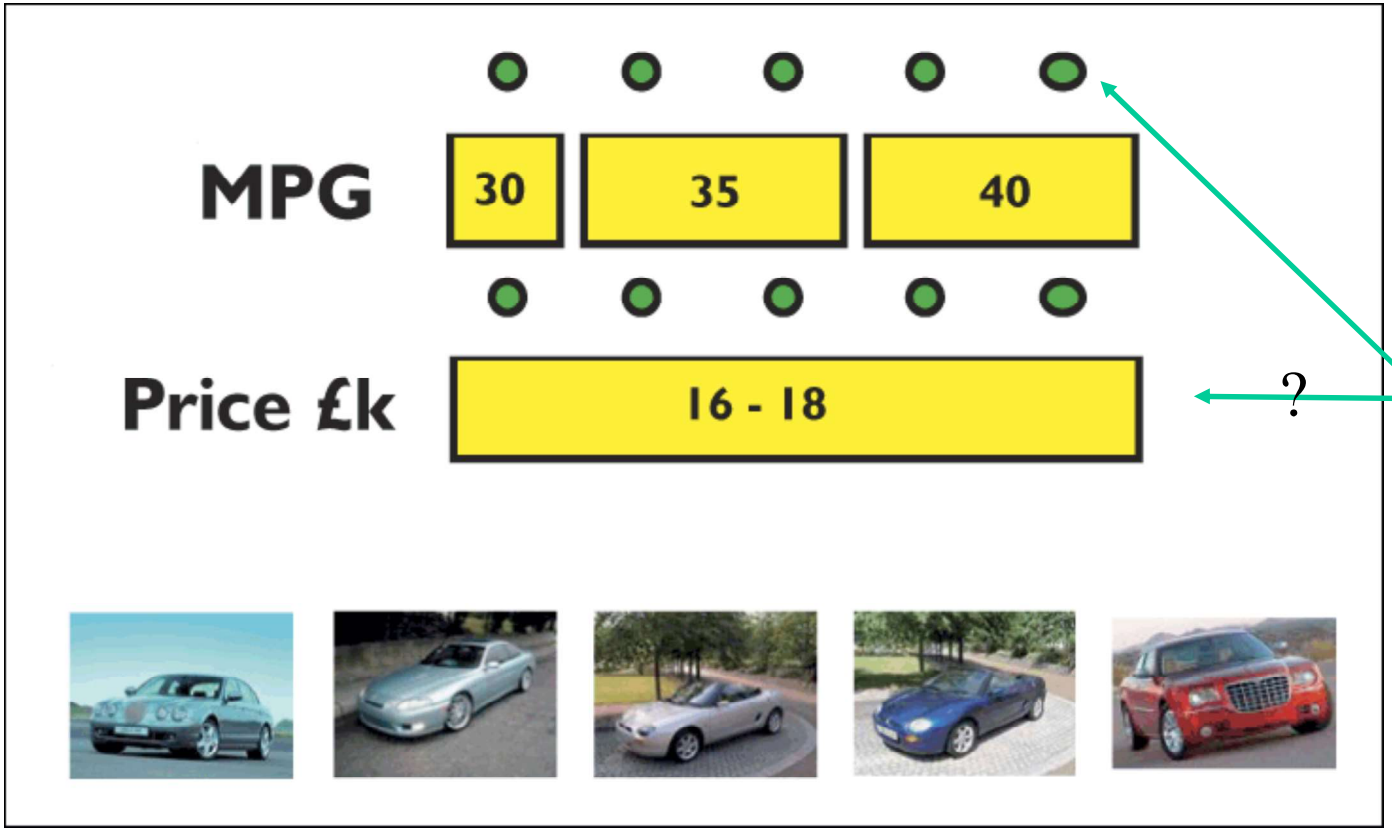
- Interaction and pre-attentive coding allow for exploring single items in efficient way
- Showing at the same time multiple attributes

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

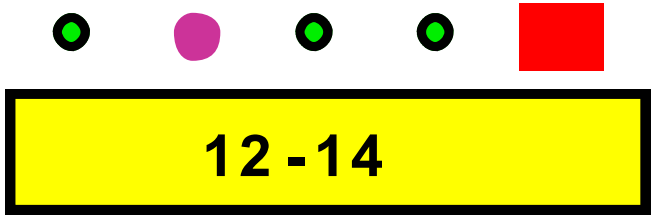


Outline

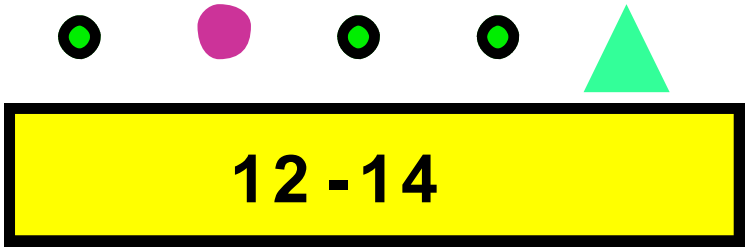
- Issues to take into account designing an Infovis application
  - Task, subtasks, & problem (why)
  - Data (how)
  - Representation & Presentation (how)
  - Overview & interactive object/attribute selection (how)
  - Filtering (how)
  - Significant/reference data (how)
  - Navigational guidance (how)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



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



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:

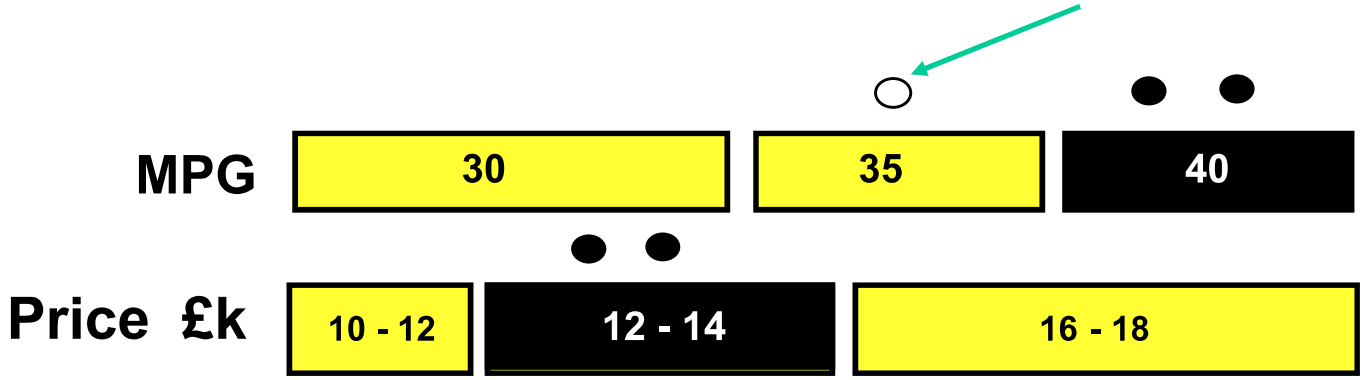
- what if I have 1k more to spend?

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

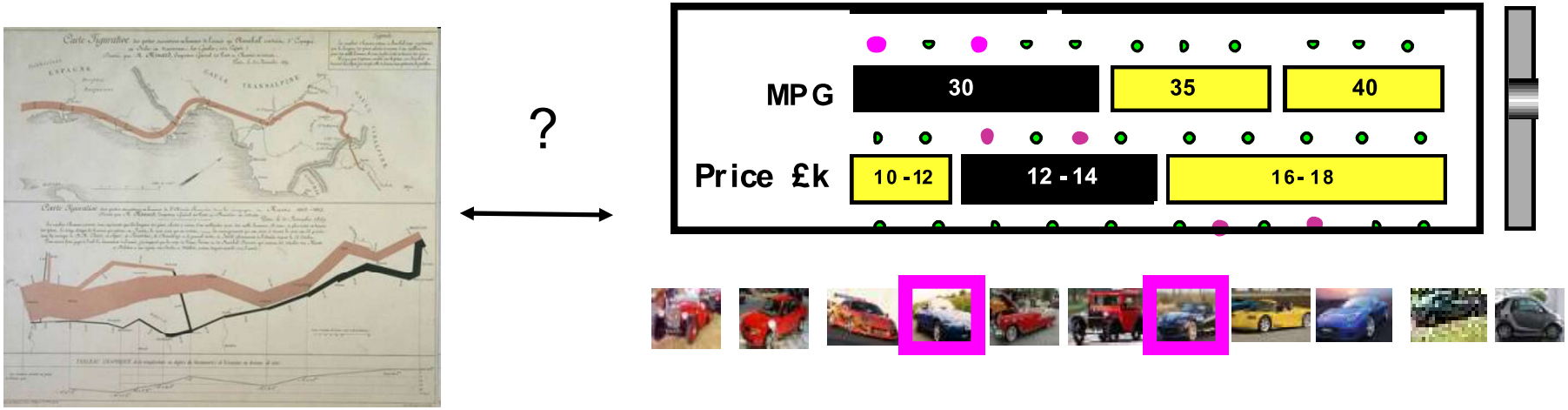


Outline

- 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?

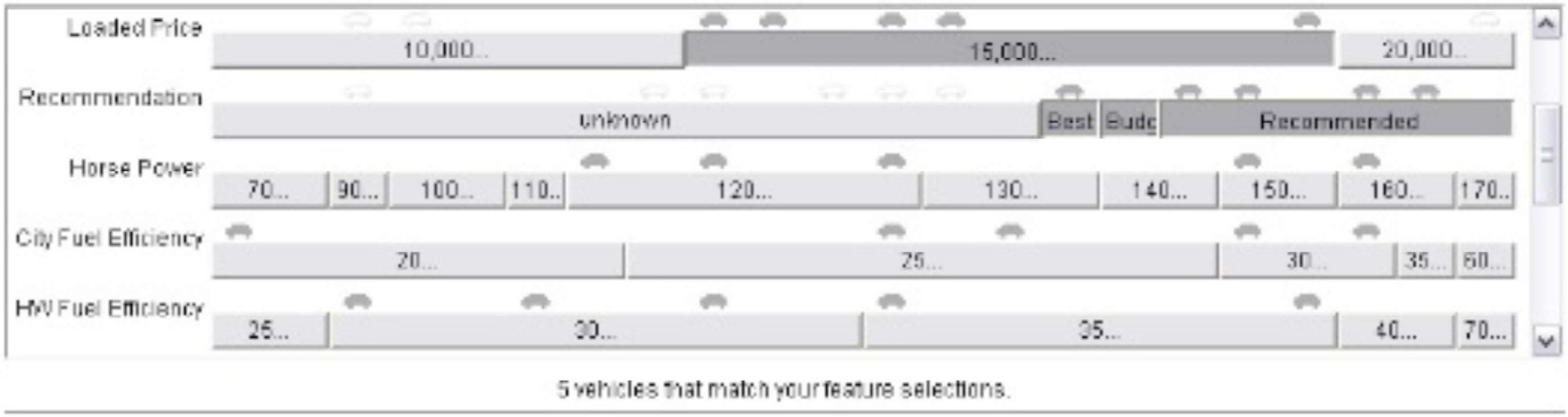


Interaction!

Computers allows for changing, switching, filtering, arranging, tagging, etc.

This is a key issue!

EzChooser




Chevrolet Cavalier

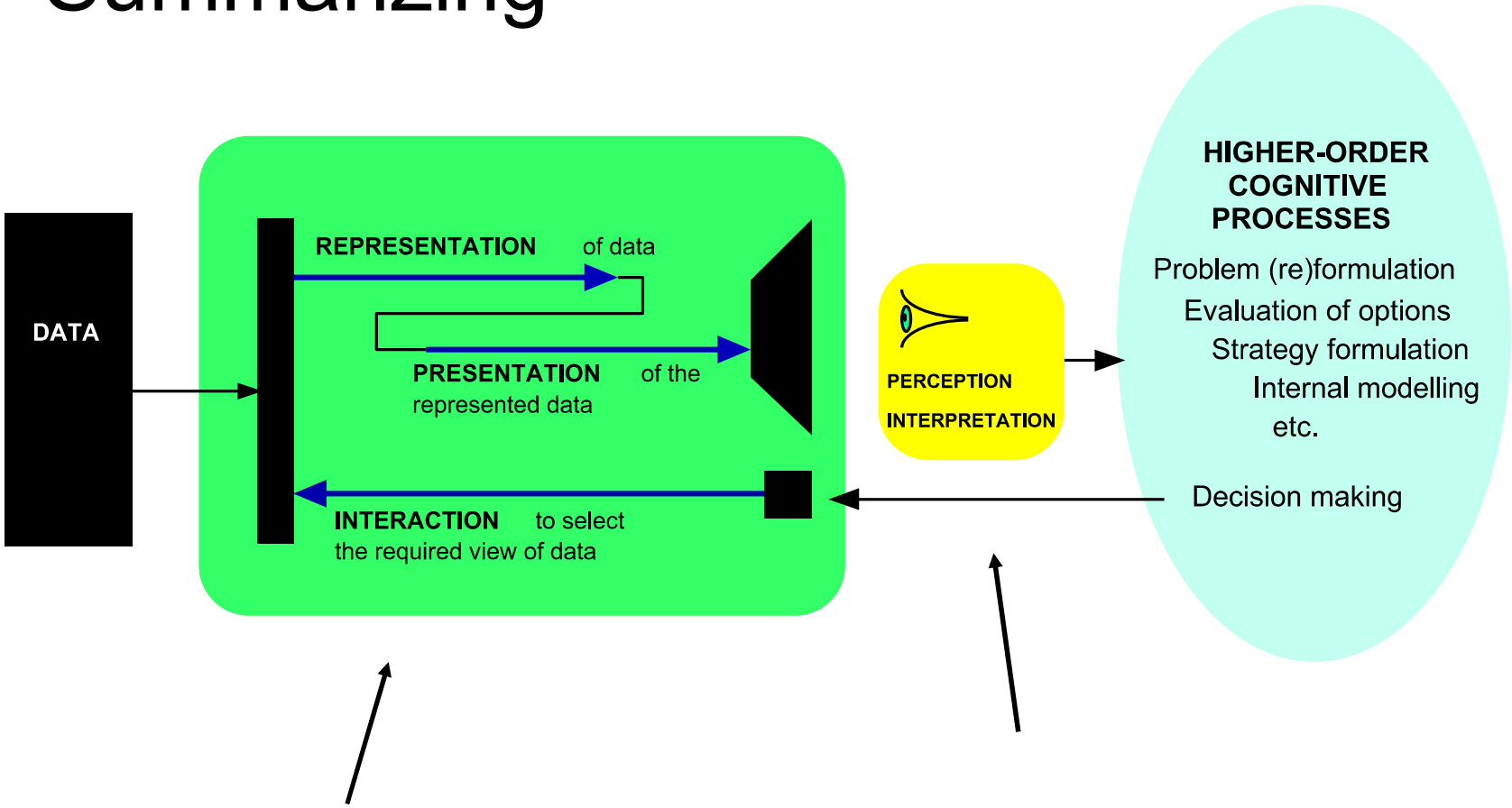

Chevrolet Prizm


Honda Civic


Mazda Protege


Toyota Corolla

Summarizing



Spence's book

Ware's book