information visualisation

Alan Dix

talis

~~Birmingham~~ Lancaster University and Talis

www.hcibook.com/alan/teaching/Promise2012/

PROMISE Winter School 2012
Zinal, Switzerland

example
Map your moves

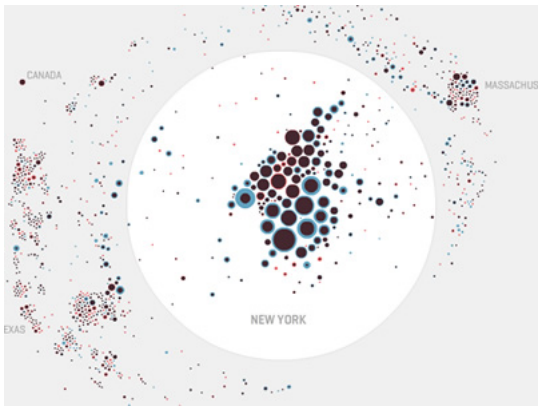
where New Yorkers move (10 years data)

distorted map

circle = moves for one zip code

red – out
blue – in

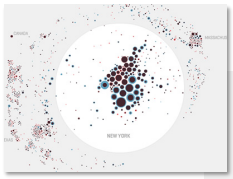
overlaid

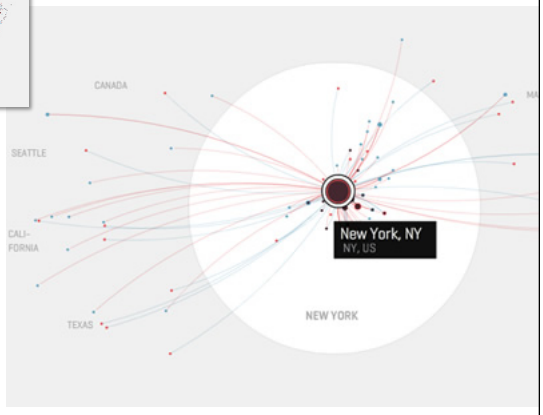


http://moritz.stefaner.eu/projects/map%20your%20moves/

PROMISE Winter School 2012
Zinal, Switzerland

example
Map your moves





interactive:
- selecting a zip code shows where movements to/from

also hiding:
- what you don't show also important

http://moritz.stefaner.eu/projects/map%20your%20moves/

what is visualistion?

making data easier to understand
using **direct** sensory experience

especially visual!

but can have aural, tactile 'visualisation'

PROMISE Winter School 2012
Zinal, Switzerland

direct sensory experience

N.B. sensory rather than linguisitic

sort of right/left brain stuff!

but ... may include text, numbers, etc.

visualising in text alignment - numbers

think purpose!

which is biggest?

```
532.56
179.3
256.317
15
73.948
1035
3.142
497.6256
```

PROMISE Winter School 2012
Zinal, Switzerland

visualising in text
alignment - numbers

visually:

long number = big number

align decimal points

or right align integers

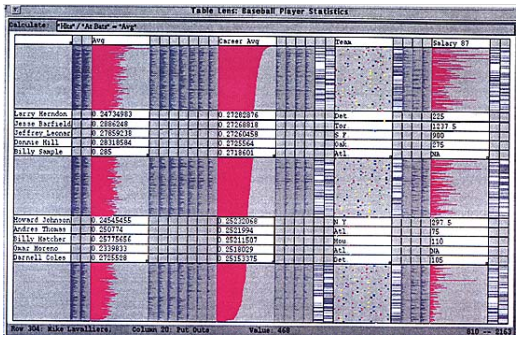
```
627.865
  1.005763
382.583
2502.56
432.935
  2.0175
652.87
 56.34
```


visualising in text
TableLens

like a‘spreadsheet’ ...

... but some rows squashed to one pixel high

numbers become small histogram bars



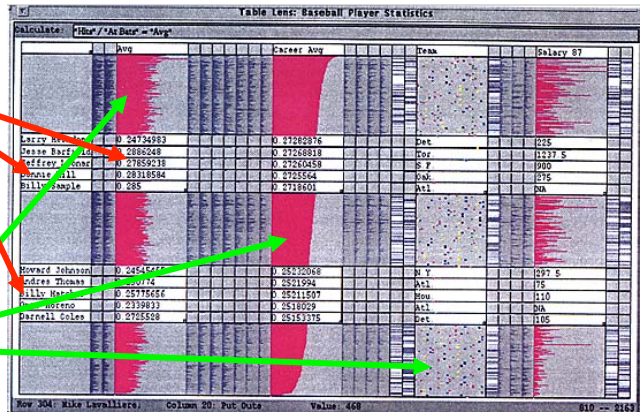
PROMISE Winter School 2012
Zinal, Switzerland

visualising in text
TableLens

N.B. also an example of **focus+context**

focus
some rows
in full detail

context
whole dataset
can also be seen
in overview



especially visual

visual cortex is 50% of the brain!

... but disability, context, etc.,
may mean non-visual forms needed

why visualisation?

for the data analyst

scientist, statistician, probably you!

for the data consumer

audience, client, reader, end-user

why visualisation?

understanding

consumer

rhetoric

focus on well understood, simple representations

why visualisation?

understanding

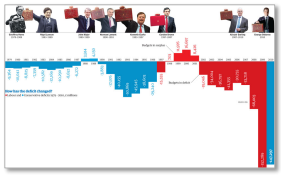
consumer

rhetoric

to help others see what the analyst has already seen

infographics

data journalism



http://www.guardian.co.uk/news/datablog/2010/oct/18/deficit-debt-government-borrowing-data

PROMISE Winter School 2012
Zinal, Switzerland

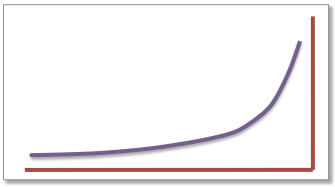
why visualisation?

understanding

consumer

rhetoric

the business plan hockey stick!



to persuade readers of particular point (and not others!)

lies, damn lies, and graphs

PROMISE Winter School 2012
Zinal, Switzerland

why visualisation?

understanding

analyst

exploration

powerful, often novel visualisations, training possible

why visualisation?

understanding

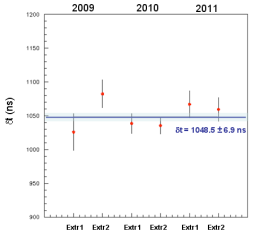
consumer

exploration

to make more clear particular aspects of data

confirming hypotheses

noticing exceptions



e.g. box plots in stats

graph from: Measurement of the neutrino velocity with the OPERA detector in the CNGS beam

why visualisation?

seeking the unknown

avoiding the obvious

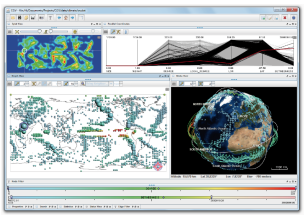
wary of happenstance

understanding

consumer

exploration

to find new things
that have not been
previously considered



a brief history of visualisation

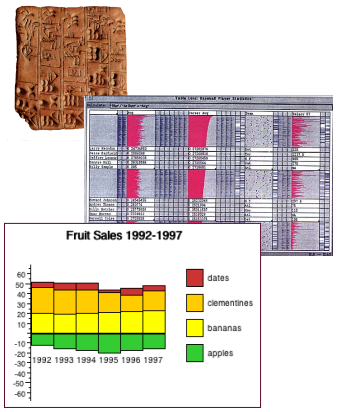
from 2500 BC to 2012

a brief history ...

static visualisation
- the first 2500 years

interactive visualisation
- the glorious '90s

and now?
- web and mass data
- visual analytics



PROMISE Winter School 2012
Zinal, Switzerland

static visualisation

from clay tablets to Tufte

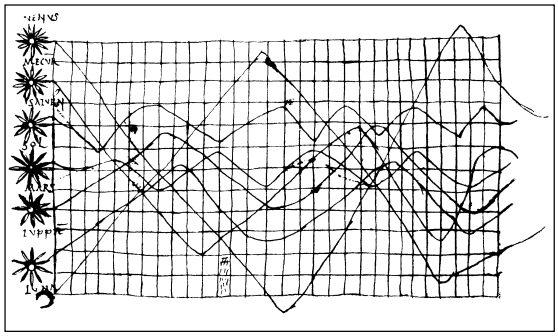
Mesopotamian tablets

static visualisation

from clay tablets to Tufte

Mesopotamian tablets

10th Century time line



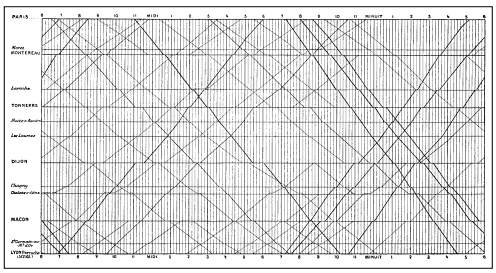
static visualisation

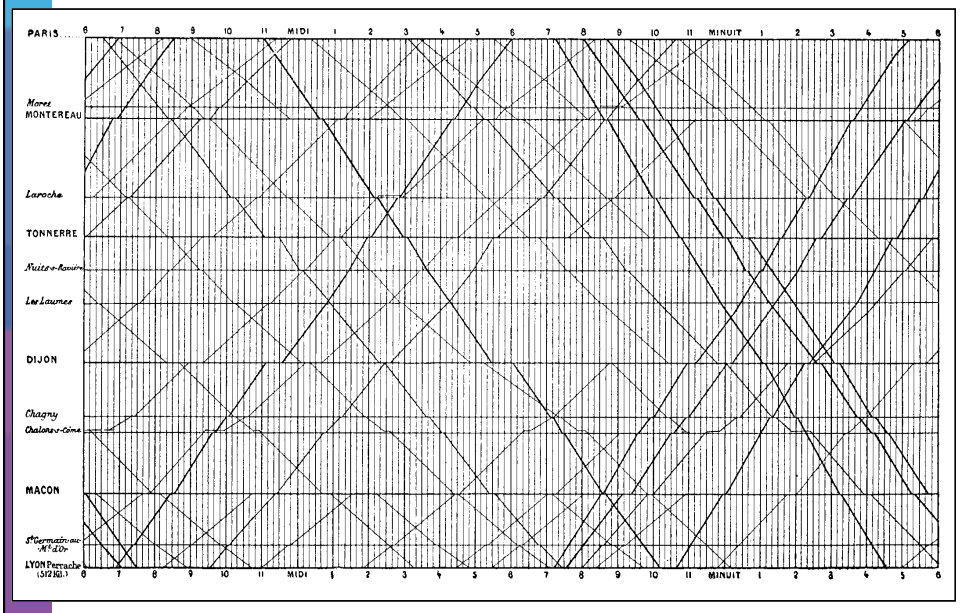
from clay tablets to Tufte

Mesopotamian tablets

10th Century time line

1855 Paris-Lyon train timetable





PROMISE Winter School 2012
Zinal, Switzerland

static visualisation

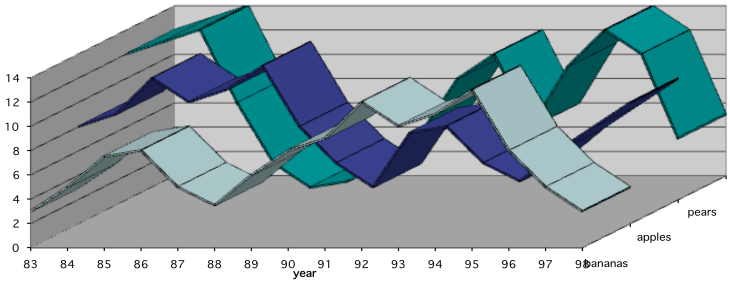
from clay tablets to Tufte

Mesopotamian tablets

10th Century time line

1855 Paris-Lyon train timetable

Excel etc.



Trends in fruit sales

static visualisation

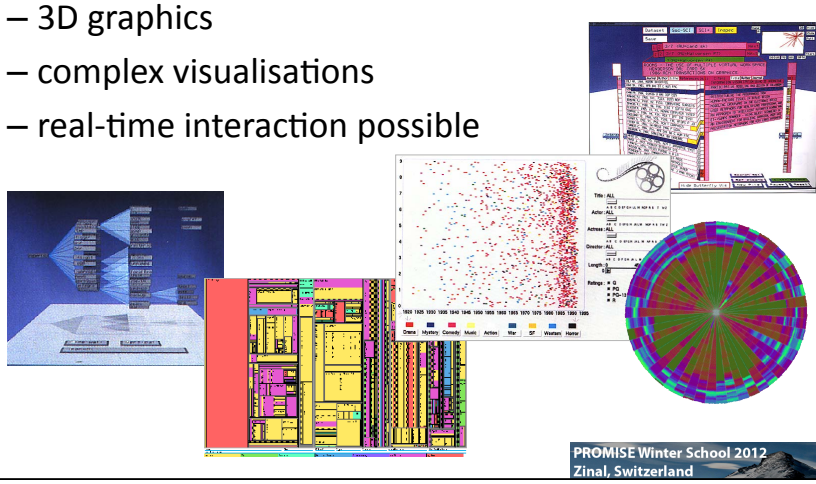
read Tufte's books ...

- *The Visual Display of Quantitative Information*
- *Envisioning Information*
- *Visual Explanations*

interactive visualisation

early 1990s growing graphics power

- 3D graphics
- complex visualisations
- real-time interaction possible

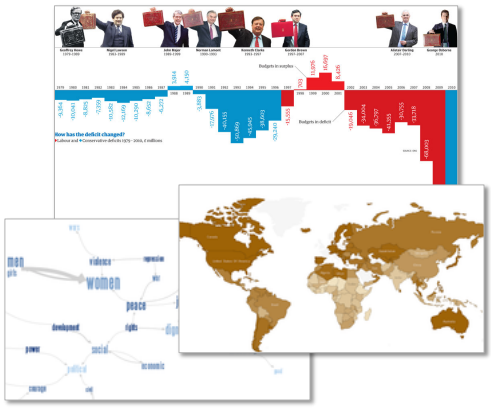


... and now

loads of data

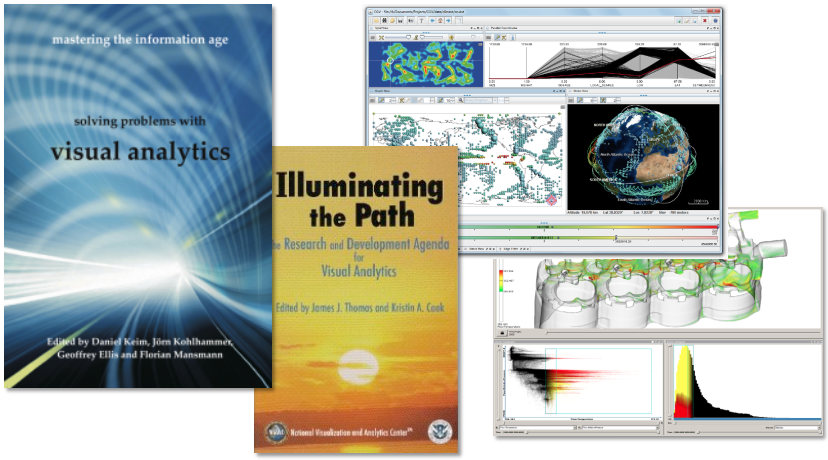
web visualisation

data journalism



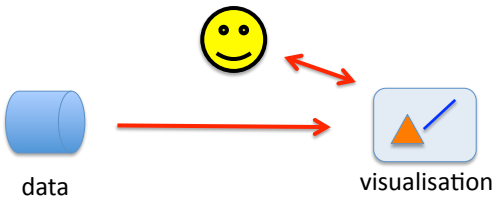
http://www.guardian.co.uk/news/datablog/2010/oct/18/deficit-debt-government-borrowing-data
http://www-958.ibm.com/software/data/cognos/manyeyes/

and visual analytics!

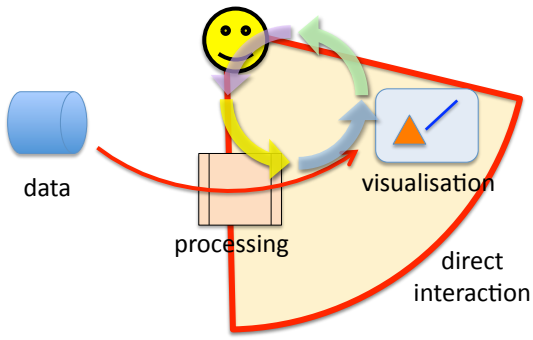


visualisation in context

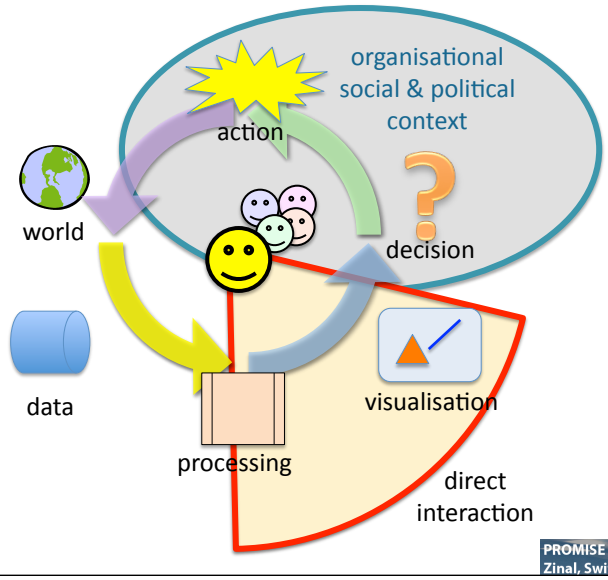
plain visualisation

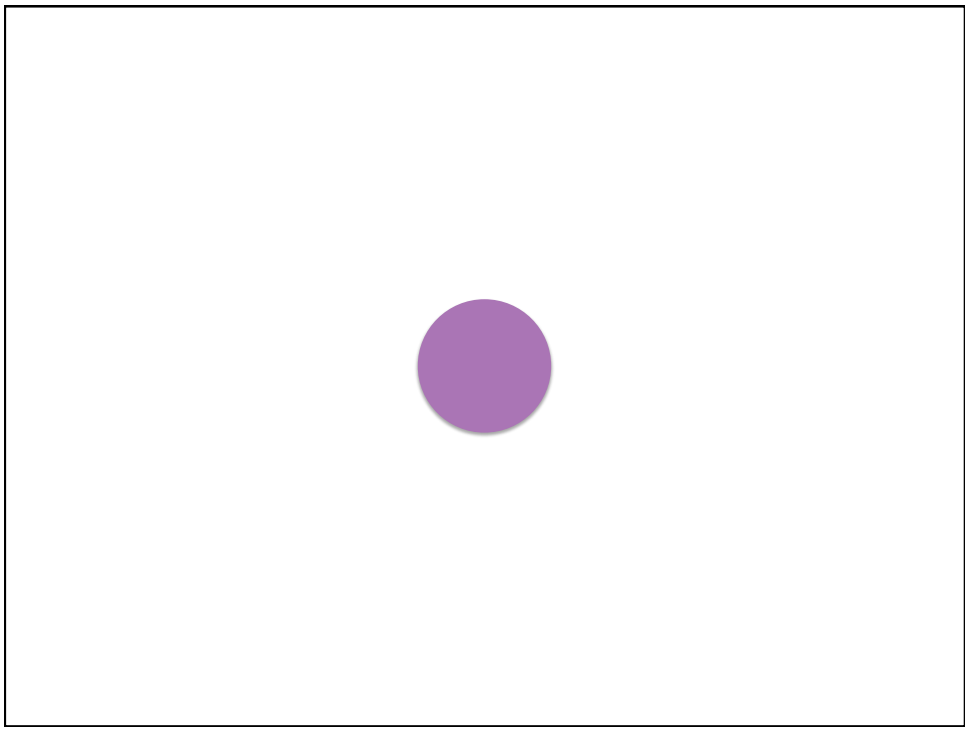


visual analytics



the big picture





designing visualisation

choosing representations

visualisation factors

- visual 'affordances'
  what we can see
- objectives, goals and tasks
  what we need to see
- aesthetics
  what we like to see

trade-off

visualisation factors

- visual affordances
- objectives, goals and tasks
- aesthetics

static representation ⇒ trade-off

interaction reduces trade-off

- stacking histogram, overview vs. detail, etc. etc.

relaxing constraints

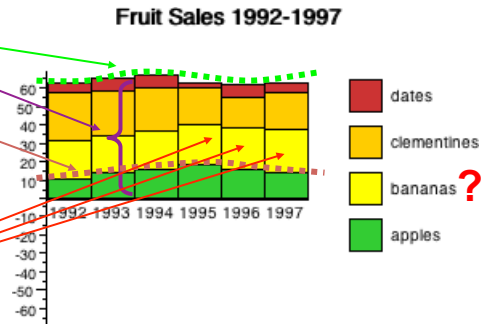
normal stacked histogram

good for:

- –overall trend
- –relative proportions
- –trend in bottom category

bad for others

- –what is happening to bananas?

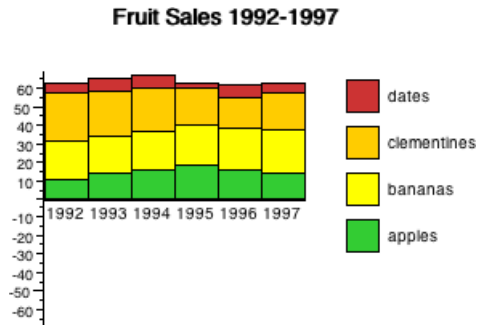


make your own (iii)

relaxing constraints

interactive stacking histograms ...
or ... **dancing histograms**

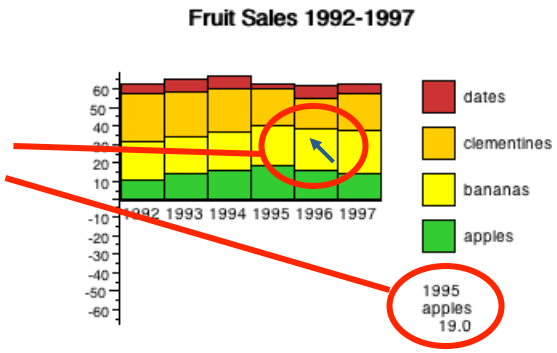
normal histogram
except ...



make your own (iii)
relaxing constraints

interactive stacking histograms ...
or ... **dancing histograms**

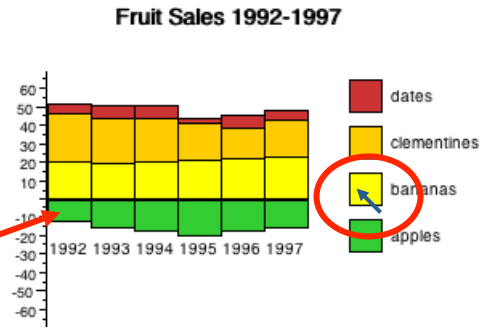
normal histogram
except ...
hover over cell
to show detail



make your own (iii)
relaxing constraints

interactive stacking histograms ...
or ... **dancing histograms**

normal histogram
except ...
hover over cell
to reveal detail
click on legend
to change
baseline



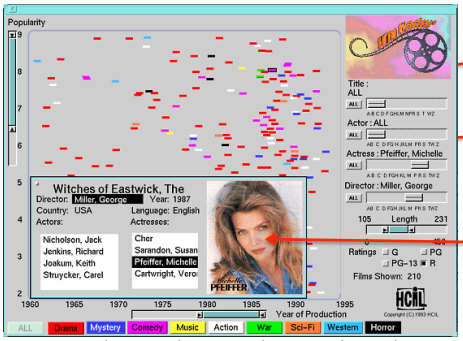
demonstration

kinds of interaction

highlighting and focus
drill down and hyperlinks
overview and context
changing parameters
changing representations
temporal fusion

Shneiderman's visualisation mantra

overview first,
zoom and filter,
then details on demand

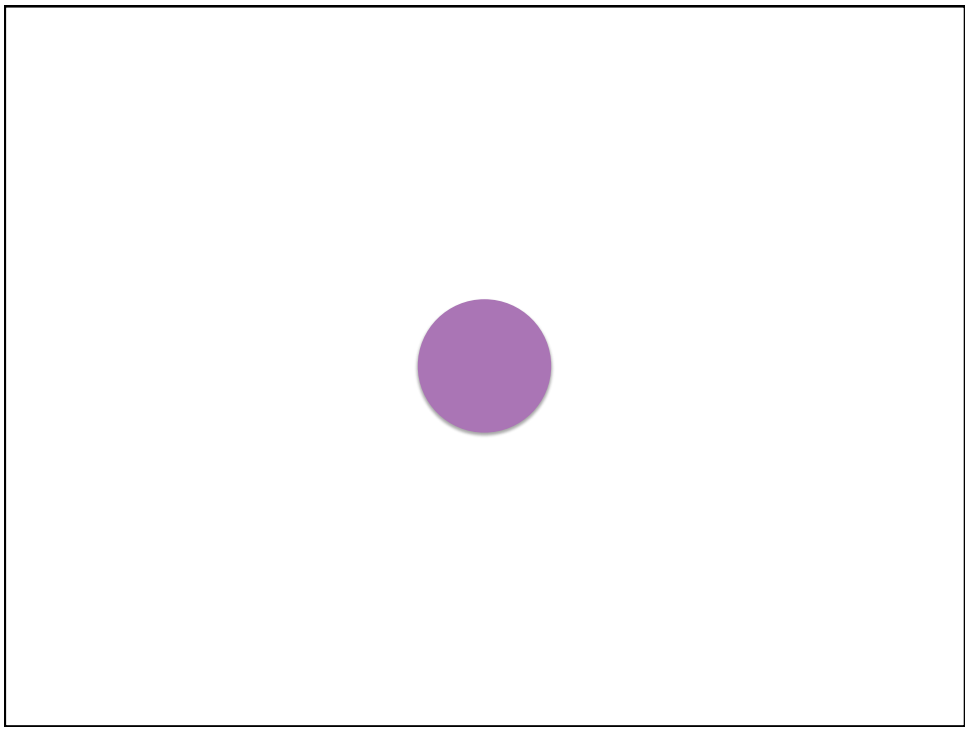


overview

zoom and filter using sliders

details on demand

http://www.sapdesignguild.org/community/book_people/visualization/controls/FilmFinder.htm



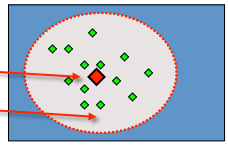
classic visualisations

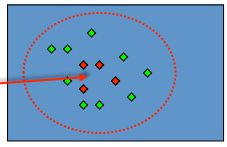
displaying groups/clusters

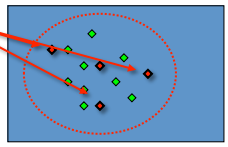
numeric attributes
- use average or region

categorical attributes
- show values of attributes common to cluster

text, images, sound
- no sensible 'average' to display
- use typical documents/images
- central to cluster ... or spread within cluster







using clusters

the scatter/gather browser

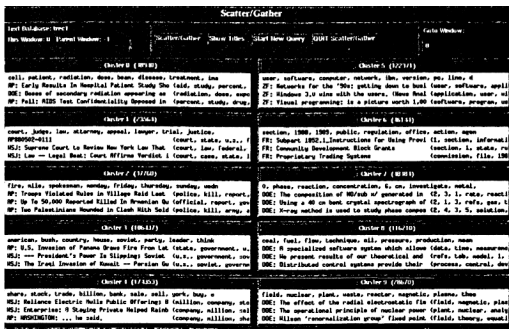
take a collection of documents

scatter:
- group into fixed number of clusters
- displays clusters to user

gather:
- user selects one or more clusters
- system collects these together

scatter:
- system clusters this new collection

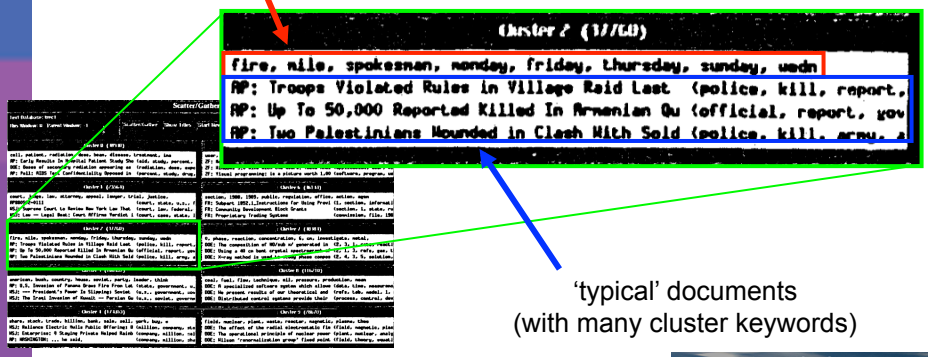
...



displaying clusters

scatter-gather browser

keywords (created by clustering algorithm)

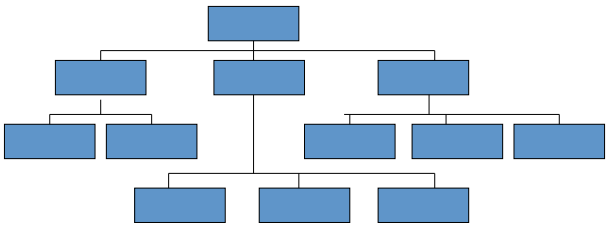


'typical' documents
(with many cluster keywords)

hierarchical data

hierarchies are everywhere!

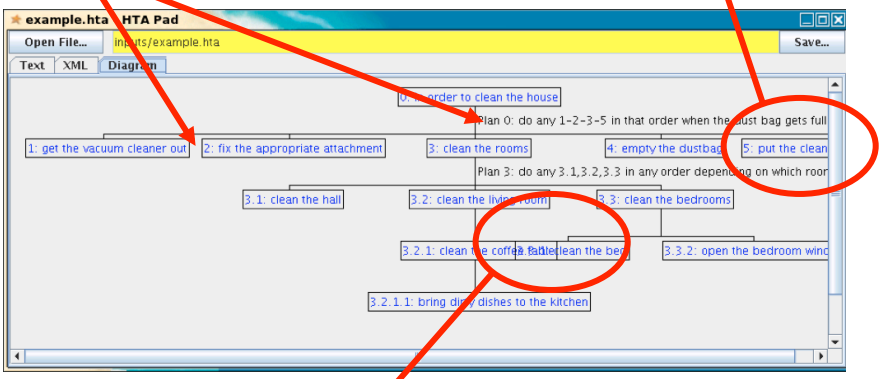
- file systems
- organisation charts
- taxonomies
- classification trees
- ontologies
- xml



problems with trees ...

hard to fit text labels

width grows rapidly

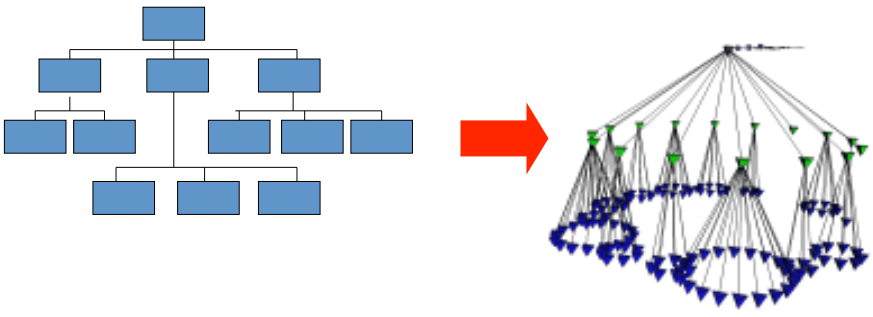


overlapping low level nodes

use 3D?

cone tree

– use stacked circles of subtrees

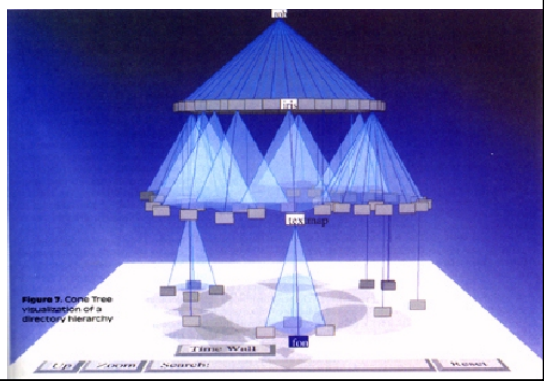


good use of 3D

still have occlusion ...
but 'normal' in 3D

shadows help to disambiguate

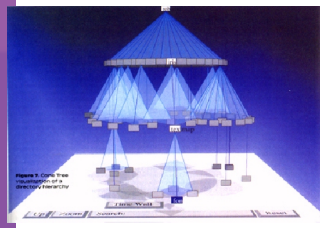
but text labels difficult

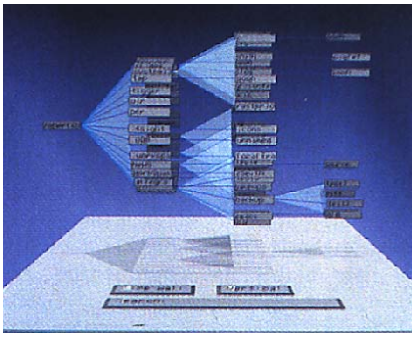


cone trees → cam trees

horizontal layout makes labels readable

small things matter!





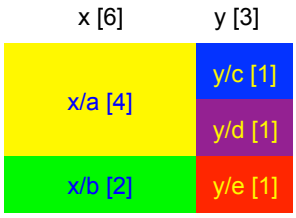
disect 2D space - treemaps

takes tree of items with some 'size'

- e.g. file hierarchy, financial accounts

alternatively divides space horizontally/vertically for each level, proportionate to total size

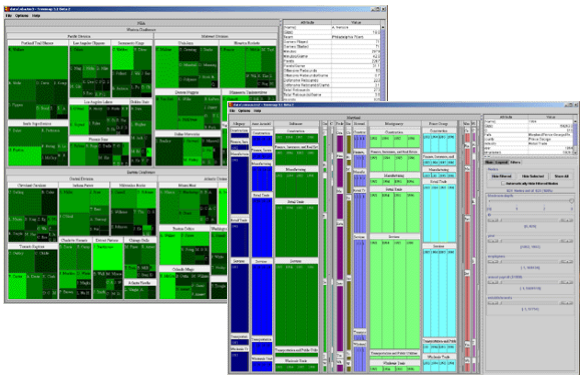
x
  x/a – 4
  x/b – 2
y
  y/c – 1
  y/d – 1
  y/e – 1



http://www.cs.umd.edu/hcil/treemap-history/

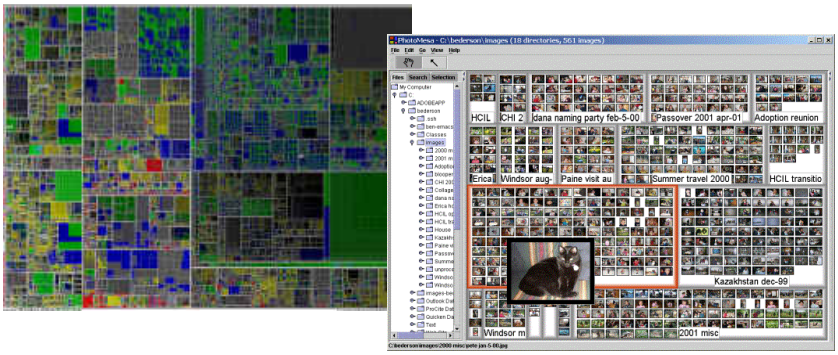
treemaps (2)

later variants improved the shape and appearance of maps



treemaps (3)

plus algorithms for vast data sets, for thumbnail images, etc. etc.



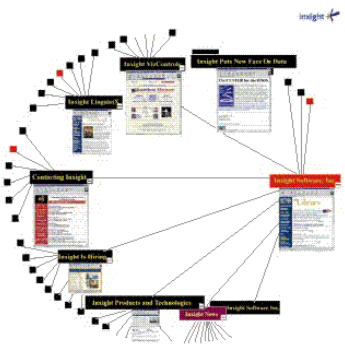
distort space ...

tree branching factor b:
- number of nodes at depth d = b^d

Euclidean 2D space:
- amount of space at radius r = $2\pi r$
- not enough space!

non-Euclidean hyperbolic space:
- exponential space at radius r

hyperbolic browser
- lays out tree in hyperbolic space
- then uses 2D representation of hyperbolic space



multiple attributes

often data items have several attributes

e.g. document:

- type (journal, conference, book)
- date of publication
- author(s)
- multiple keywords (perhaps in taxonomy)
- citation count
- popularity

traditional approach ... boolean queries

>new query

?type='journal' and keyword='visualisation'

=query processing complete - 2175 results
list all (Y/N)

>N

>refine query

refine: type='journal' and keyword='visualisation'

+author='smith'

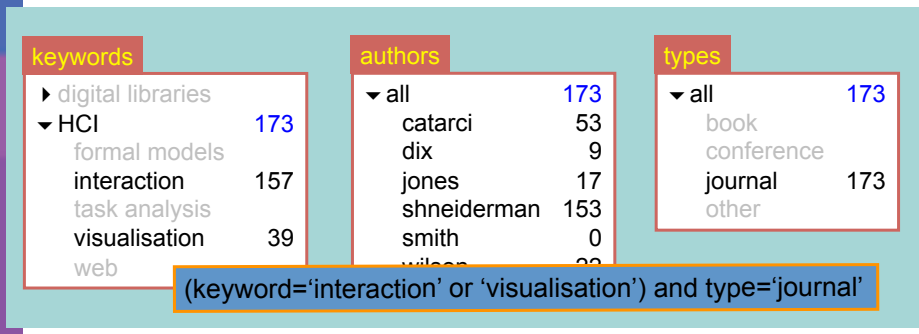
=query processing complete - 0 results

faceted browsing

e.g. HiBrowse (one of the earliest)

multiple selection boxes

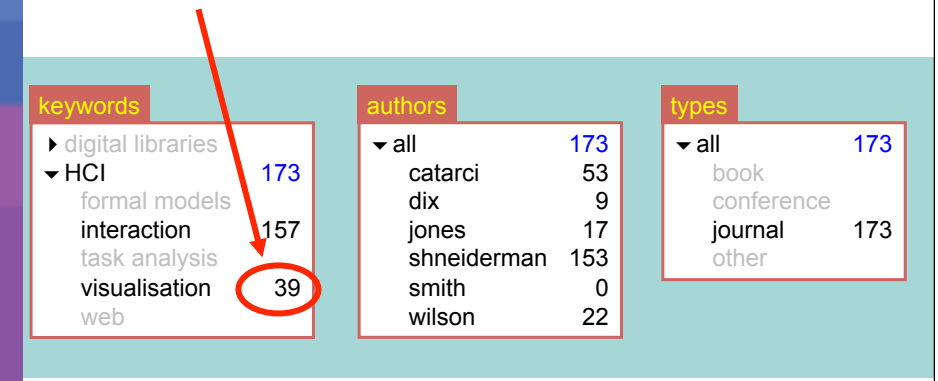
– 'or' within box - 'and' between boxes



HiBrowse (ii)

shows how many items with particular value

e.g. 39 documents with keyword='visualisation' and type='journal'



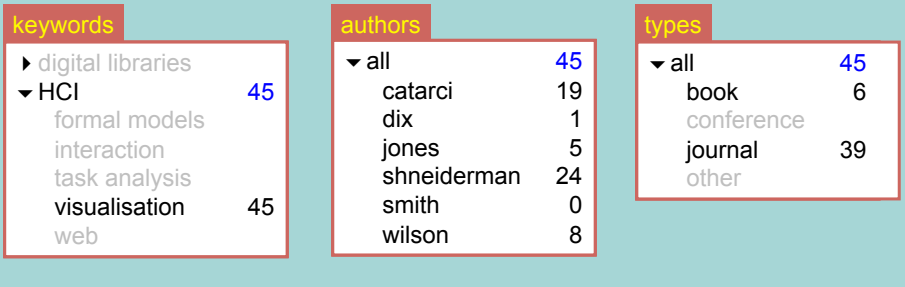
HiBrowse (iii)

can predict the effect of refining selection

e.g. selecting 'smith' would give empty result

HiBrowse (iv)

refining selection updates counts in real time

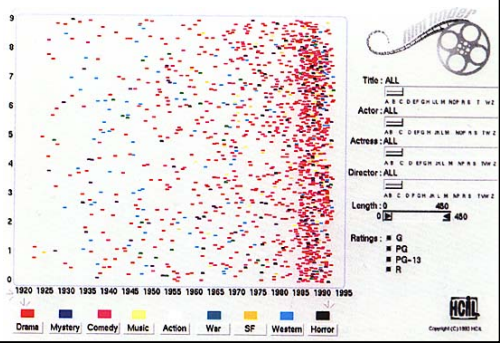


starfield (i)

scatter plot for two attributes
colour/shape codes for more

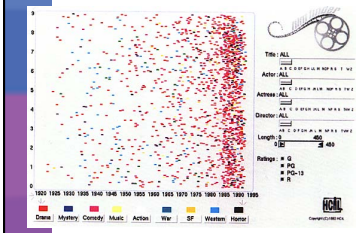
adjust rest with sliders
dots appear/disappear as slider values change

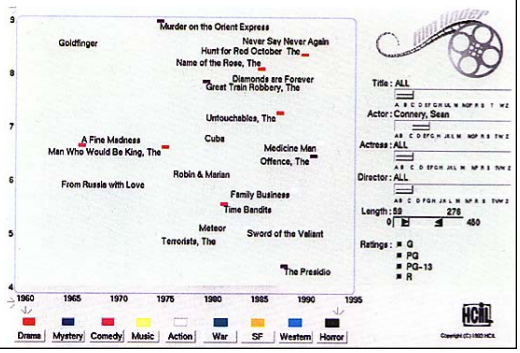
dynamic filtering



starfield (ii)

when few enough points more details appear





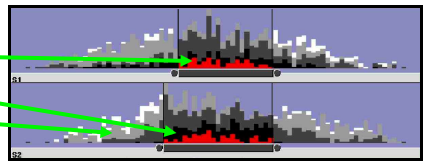
Influence Explorer (i)

developed for engineering models

like Starfield ...
but sliders show histogram

how many in category (like HiBrowse)
... and how many 'just miss'

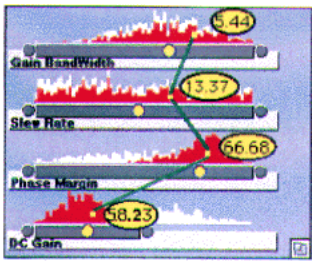
red = full match
black = all but one attribute
greys = fewer matching attr's



Influence Explorer (ii)

some versions highlight individual items
in each histogram

similar technique has
been used to match
multiple taxonomic
classifications



Information Scent

Starfield
shows what is *currently* selected
- explore using trial and error

HiBrowse and Influence Explorer
show what *would* happen

Pirolli et al. call this Information Scent
– things in the interface that help you know what actions to take to find the information you want

very large datasets

too many points/lines to see

solutions ...

space-filling single-pixel per item
Keim's VisD

random selection
(see Geoff Ellis' thesis)

clustering
visualise groups not individuals