Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

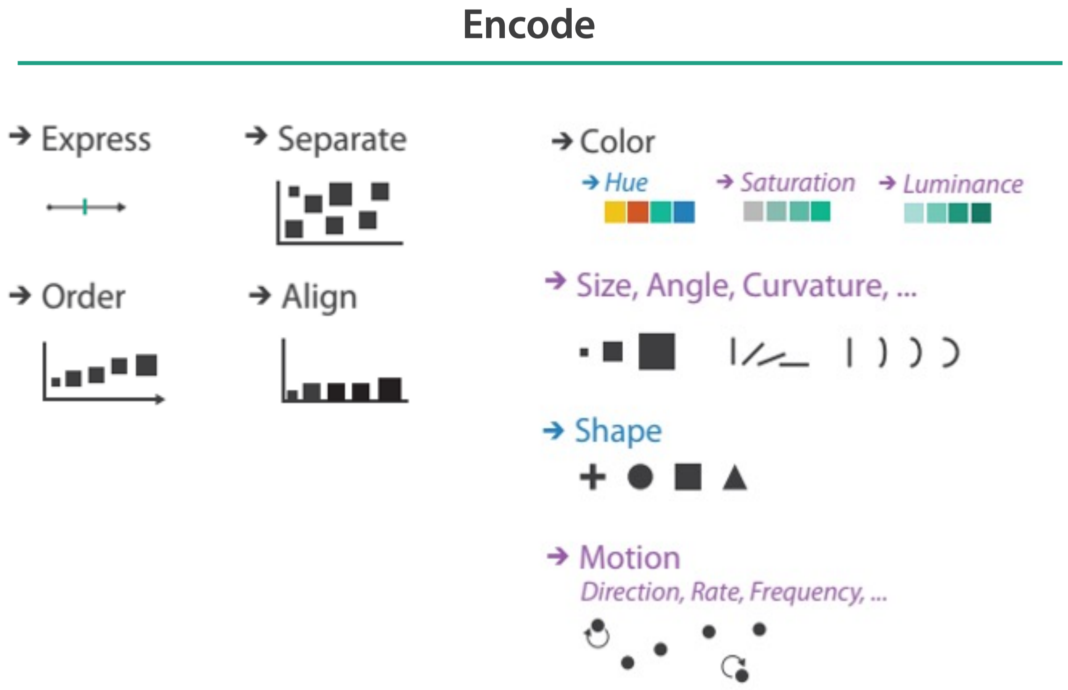
Transforming Visualizations

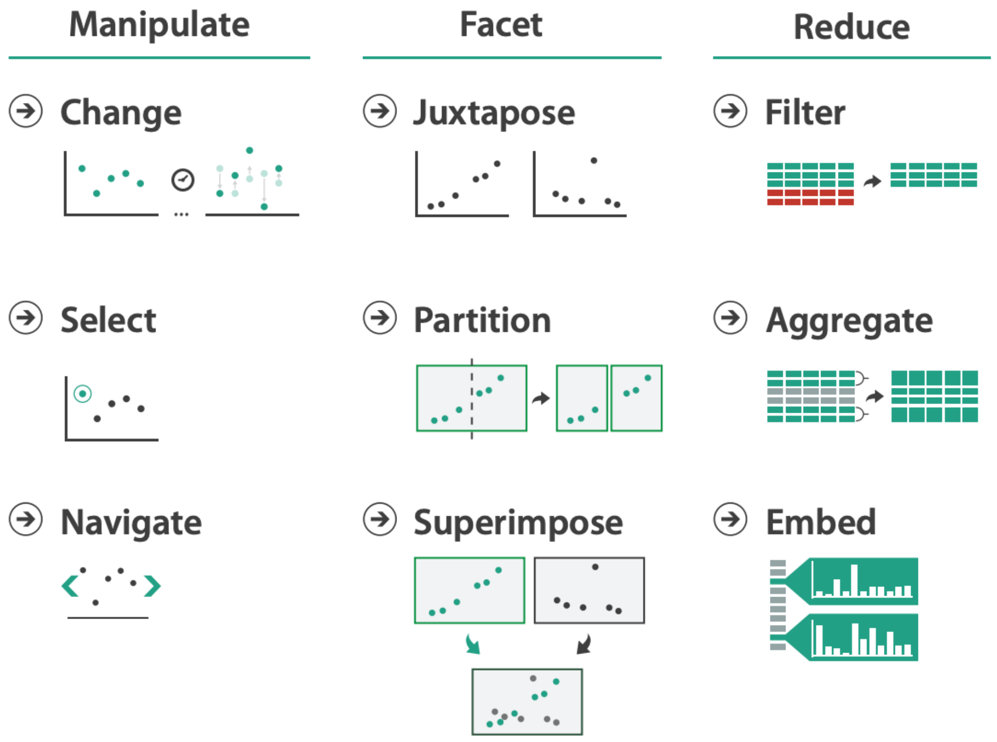
Dealing with complexity

13/10/2023

Visual Encoding Idioms

Images © Munzner 2014





Visual encodings

Encode: use visual channels to directly create representations of data from specific "views". **Sometimes this is not sufficient when data/task is complex.**

Transformations/interactions with visual encodings

Manipulate: navigate or change view(s) of data

Facet: different aspects of data presented through multiple co-ordinated views

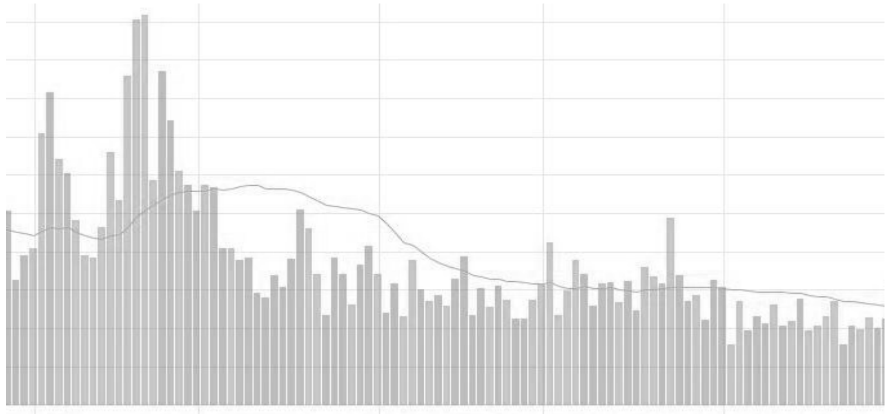
Reduce: decrease (or increase) what is shown within a view

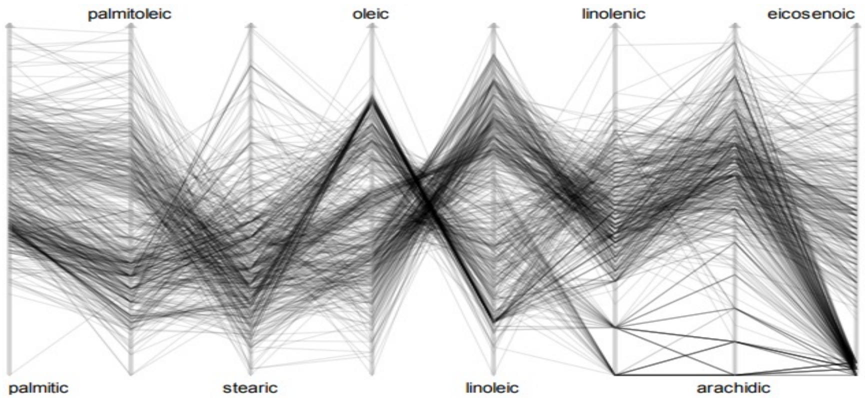
Motivations

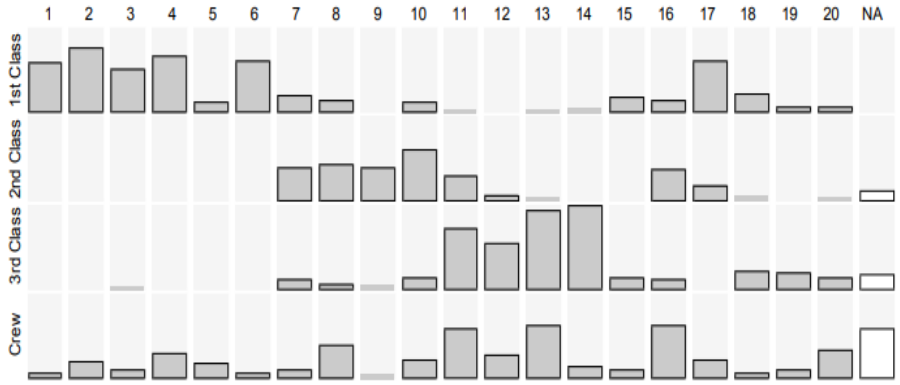
Some tasks cannot be addressed using standard static charts

Need to address data complexity

- Large data (many items)
- High-dimensional data or many attributes
- Complex interactions between elements







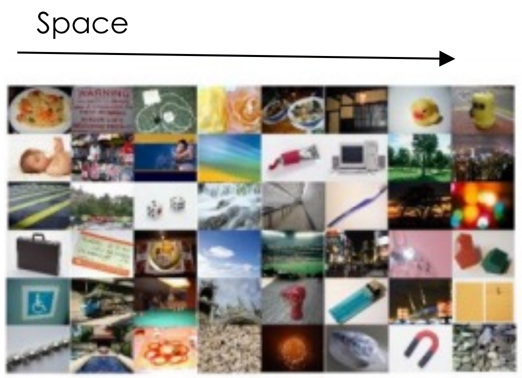
Motivations

Some tasks cannot be addressed using standard static charts

Need to manage visual complexity

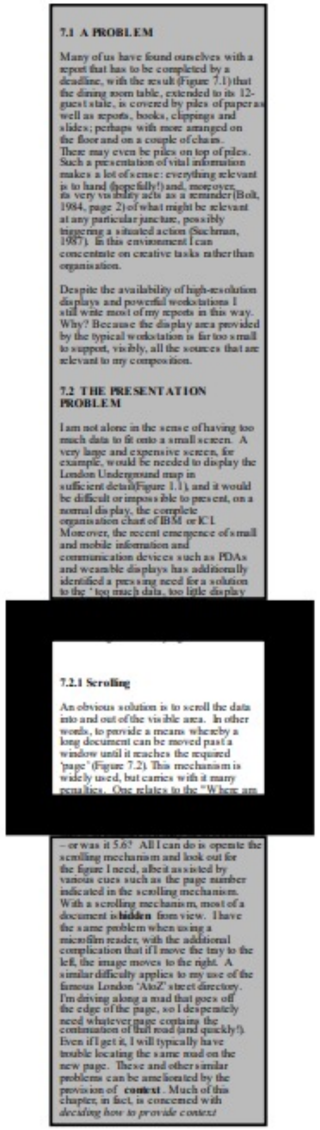
- Spatial limitations (can't fit on screen)
- Time limitations (no time to read)

- Perceptual/cognitive limitations (too confusing)
- Platform limitations (can't fit in Primary Memory or too slow to process)



Concurrent Presentation

Serial Presentation



7.1 A PROBLEM

Many of us have found ourselves with a report that has to be completed by a deadline, with the result (Figure 7.1) that the dining room table, extended to its 12-guest state, is covered by piles of paper as well as reports, books, clippings and slides; perhaps with more arranged on the floor and on a couple of chairs. There may even be piles on top of piles. Such a presentation of vital information makes a lot of sense: everything relevant is to hand (hopefully!) and, moreover, its very visibility acts as a reminder (Bolt, 1984, page 2) of what might be relevant at any particular juncture, possibly triggering a situated action (Suchman, 1987). In this environment I can concentrate on creative tasks rather than organisation.

Despite the availability of high-resolution displays and powerful workstations I still write most of my reports in this way. Why? Because the display area provided by the typical workstation is far too small to support, visibly, all the sources that are relevant to my composition.

7.2 THE PRESENTATION PROBLEM

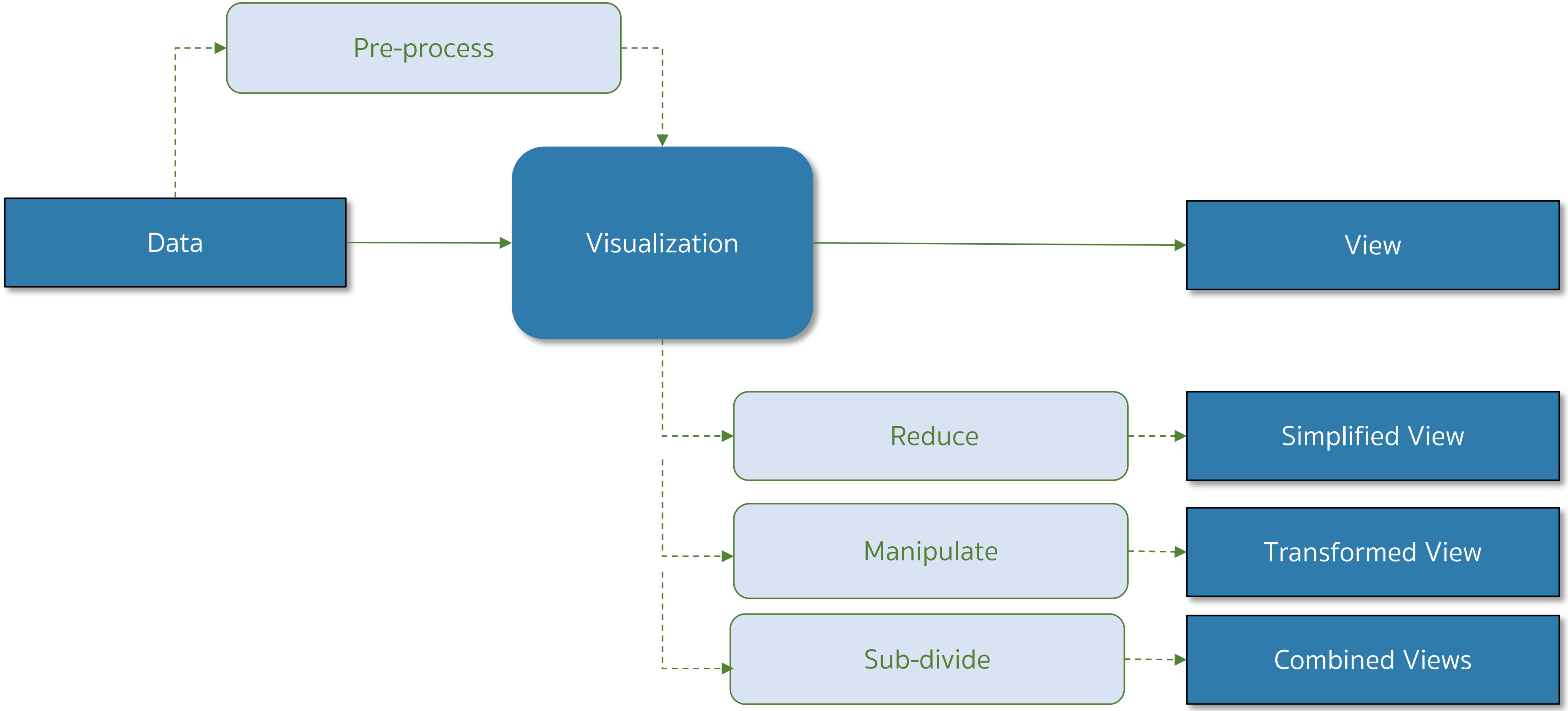
I am not alone in the sense of having too much data to fit onto a small screen. A very large and expensive screen, for example, would be needed to display the London Underground map in sufficient detail (Figure 1.1), and it would be difficult or impossible to present, on a normal display, the complete organisation chart of IBM or ICI. Moreover, the recent emergence of small and mobile information and communication devices such as PDAs and wearable displays has additionally identified a pressing need for a solution to the 'too much data, too little display

7.2.1 Scrolling

An obvious solution is to scroll the data into and out of the visible area. In other words, to provide a means whereby a long document can be moved past a window until it reaches the required 'page' (Figure 7.2). This mechanism is widely used, but carries with it many penalties. One relates to the "Where am

– or was it 5.6? All I can do is operate the scrolling mechanism and look out for the figure I need, albeit assisted by various cues such as the page number indicated in the scrolling mechanism. With a scrolling mechanism, most of a document is **hidden** from view. I have the same problem when using a microfilm reader, with the additional complication that if I move the tray to the left, the image moves to the right. A similar difficulty applies to my use of the famous London 'AtoZ' street directory. I'm driving along a road that goes off the edge of the page, so I desperately need whatever page contains the continuation of that road (and quickly!). Even if I get it, I will typically have trouble locating the same road on the new page. These and other similar problems can be ameliorated by the provision of **context**. Much of this chapter, in fact, is concerned with *deciding how to provide context*

Operations on Data Visualizations



1. Data Pre-Processing

Aims: Change data or derive additional attributes that will shape the visualization.

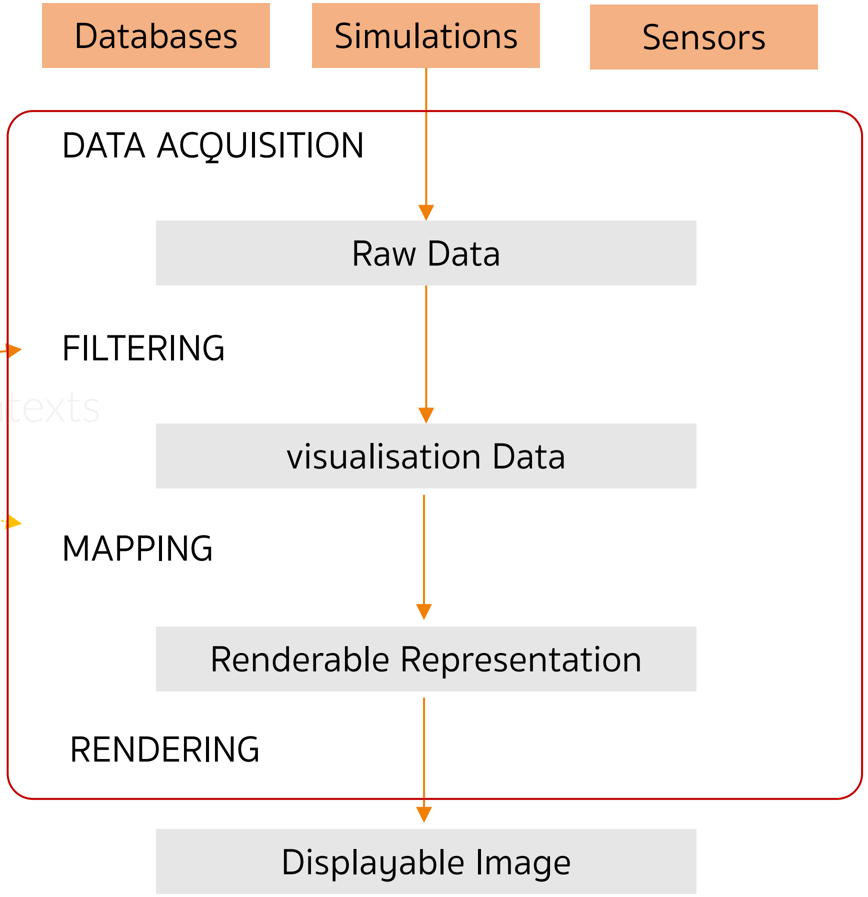
Main techniques (Ward 2012):

- Metadata and statistical analysis e.g. mean, variance
- Data-cleansing and completion e.g. filter, smoothing
- Normalization : scale data for comparing across different instances / contexts
- Mapping nominal dimensions to numbers : induce ordering
- Segmentation : split into subsets that will be handled differently
- Data Reduction *e.g. summarize, resample*

When and when not to modify data? Not applicable for instance in many applications in medical domain.

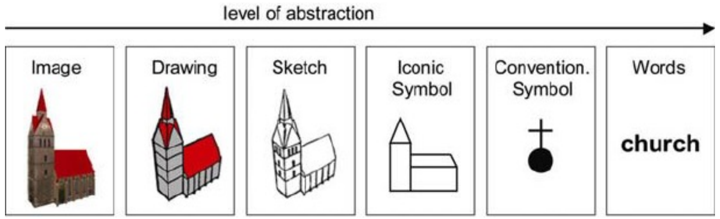
1. Data Pre-Processing

- Metadata and statistical analysis e.g. mean, variance
- Data-cleansing and completion e.g. filter, smoothing
- Normalization : scale data for comparing across different instances / contexts
- Mapping nominal dimensions to numbers : induce ordering
- Segmentation : split into subsets that will be handled differently
- Data Reduction e.g. summarize, resample



visualisation Pipeline by Weiskopf (2006)

2. **Reduce** (Simplify views)



Aims: Reduce visual complexity; optimise for human capability

Main Approaches:

- **Filter**: Eliminate some elements from visualization
- **Aggregate**: Groups of elements represented by a <u>derived</u> element; summarizing

Forms of reduction

- Reducing attributes: assumes task doesn't require all aspects of items
- Reducing items: assumes global/abstract task, usually not applicable for search/query of specific items
- Persistent vs Transient reduction (temporarily eliminate some elements from **current** view)

Attribute

Item

Cell

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

Filtering Items

Aim: Reduce items, retain attributes

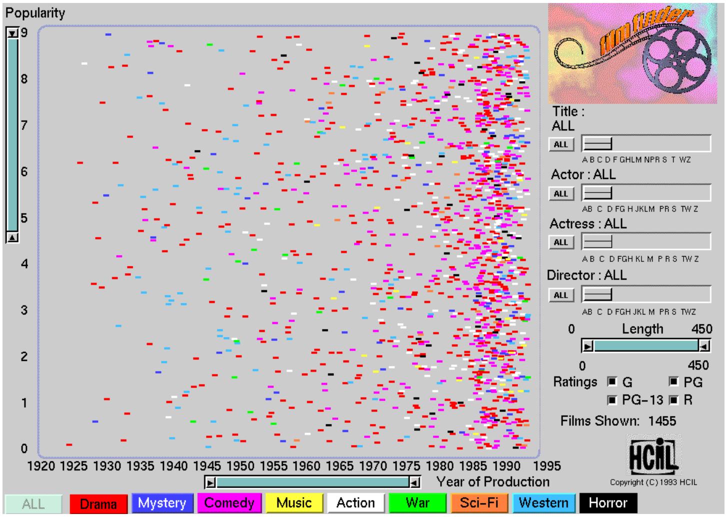
#	dim1	dim2	dim3	dim4	...	dimM
1						
2						
3						
4						
5						
...						
N						

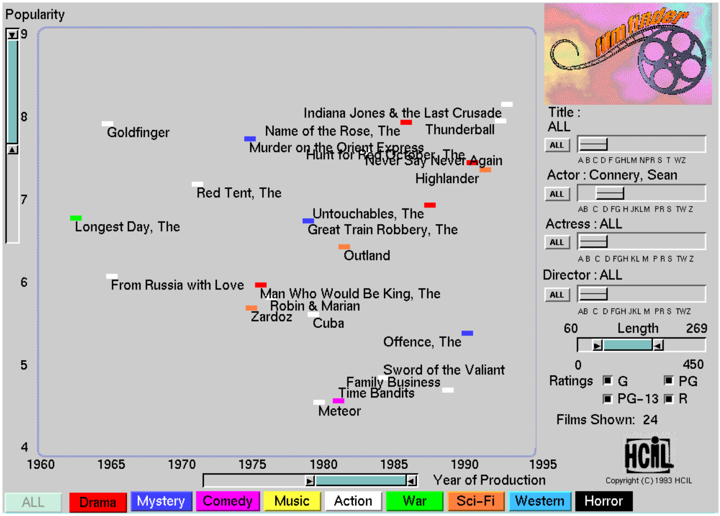
How: Eliminate some items based on values of a specific attribute

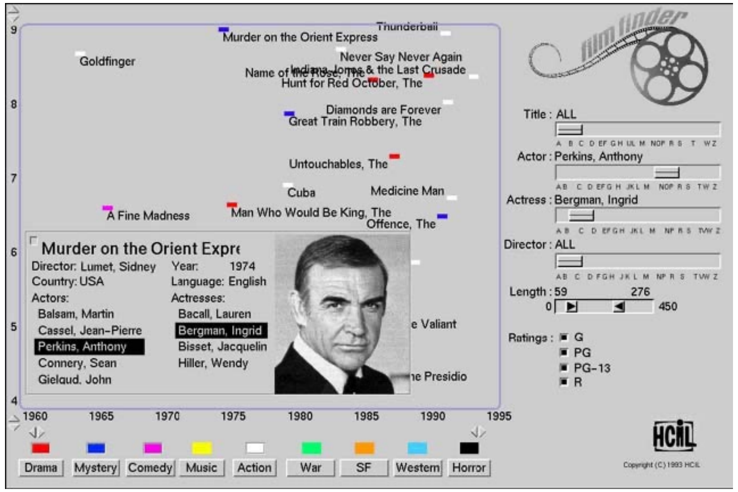
- e.g. show only selected ranges of certain attributes

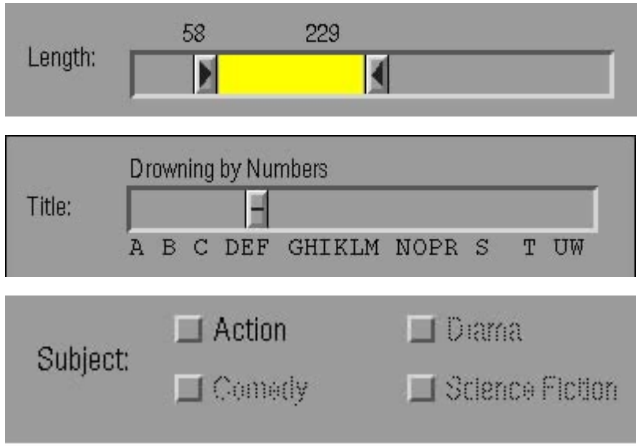
Why: aid visual search by reducing clutter/complexity

Assumes global/abstract task; usually not applicable for search/query of specific items [unless filtering is interactive]









Filmfinder, Ahlberg & Shneiderman (1994)
[https://infovis-wiki.net/wiki/Film_Finder]

Filtering Attributes

Aim: Retain items, reduce attributes

#	dim1	dim2	dim3	dim4	...	dimM
1						
2						
3						
4						
5						
...						
N						

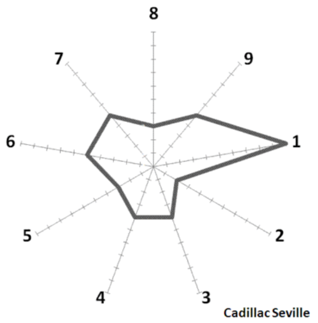
How: Simplify dataset by omitting some attributes

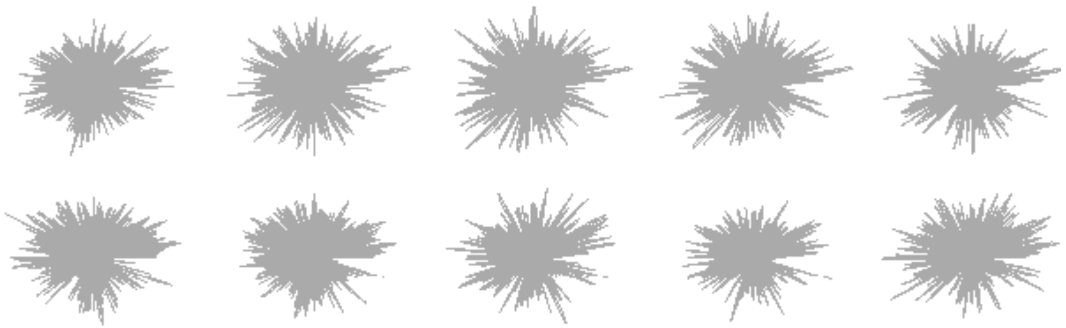
Combine with attribute ordering

- Based on similarity
- Based on variance

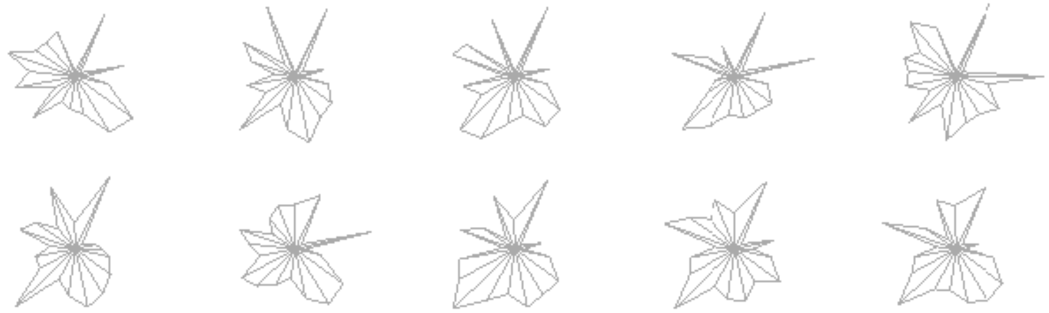
Starplots (similar to radar chart) Yang et al (2003)

1. Price
2. Mileage (MPG)
3. 1978 Repair Record (1 = Worst, 5 = Best)
4. 1977 Repair Record (1 = Worst, 5 = Best)
5. Headroom
6. Rear Seat Room
7. Trunk Space
8. Weight
9. Length



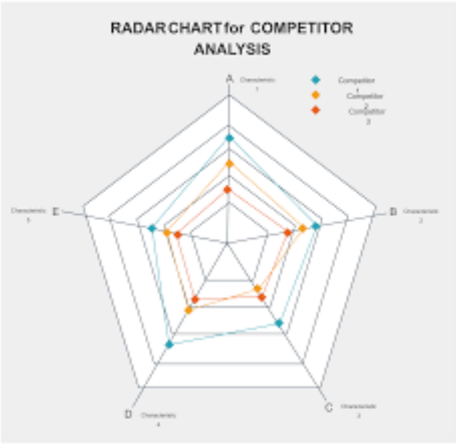


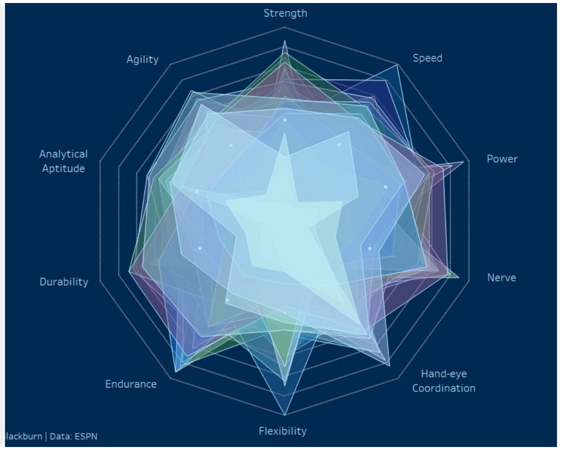
Full dataset too dense to perceive meaningful patterns

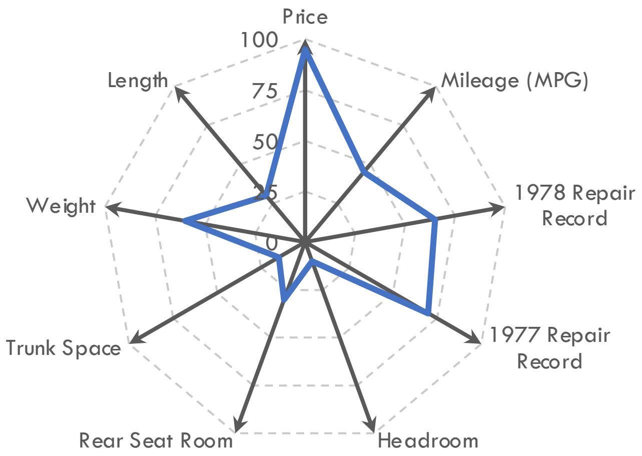


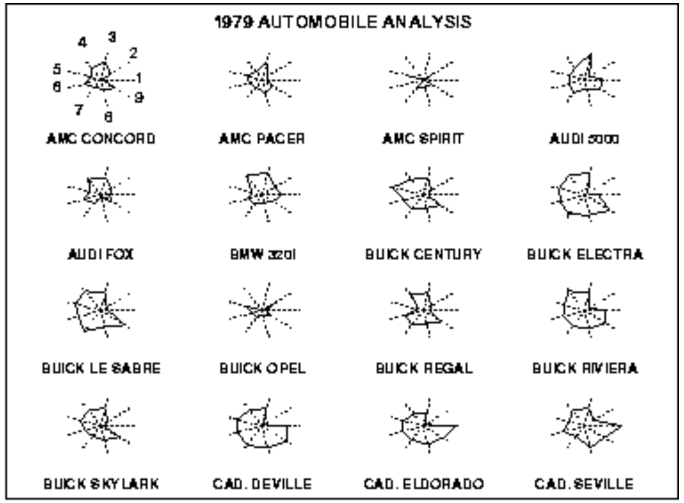
After filtering on similarity and importance

Aside: **Radar Chart / Star Plot**









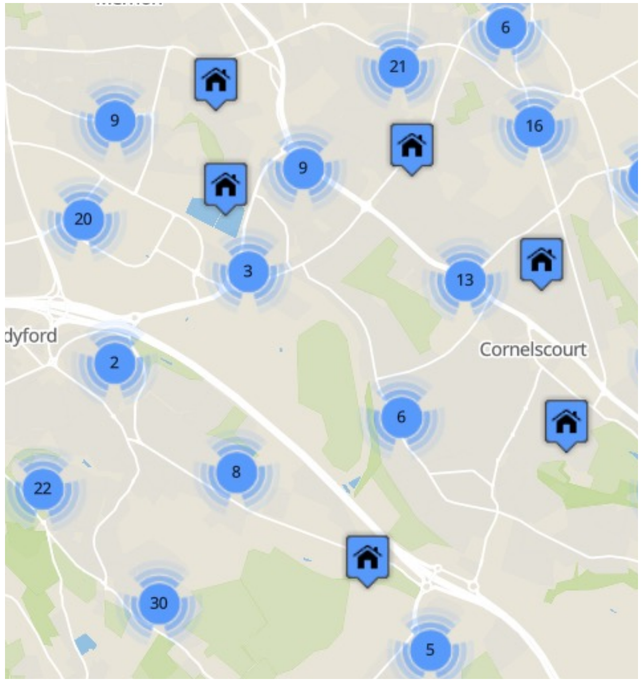
Idiom	Radar Chart / Spiderweb Chart
Data	Table with one categorical key attribute, and many quantitative/ordered attributes
Encode	Angular arrangement of axes, each of which encodes a specific attribute by position. Line marks connect the same item across different axes. Colour of line encodes category.
Task	Compare multi—featured similarities, lookup values

Idiom	Star Plot
Data	Table with one categorical key attribute, and many quantitative/ordered attributes
Encode	Positional arrangement of a sub-chart for each category. Angular arrangement of axes, each of which encodes a specific attribute by position. Line marks connect the same item across different axes.
Task	Summarize and compare multi—featured similarities.

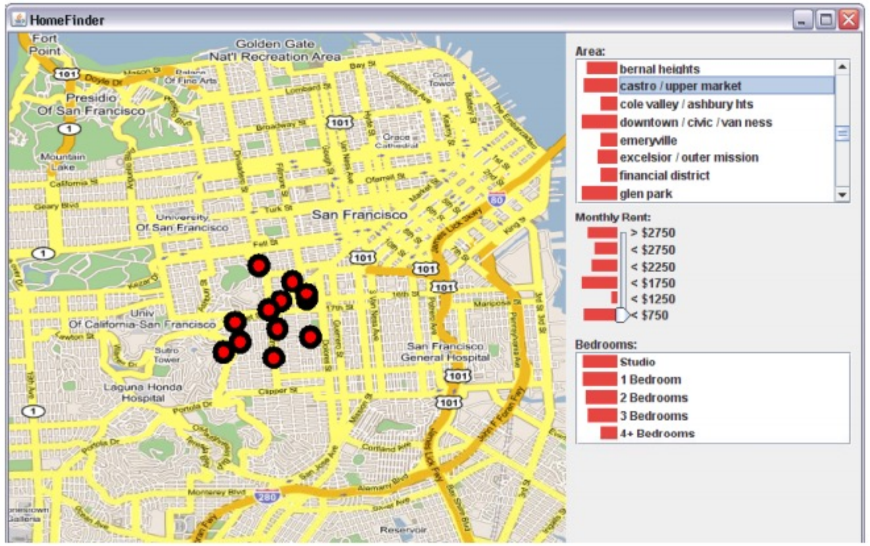
Information Scent

How not to "lose" relevant data amidst reduction:

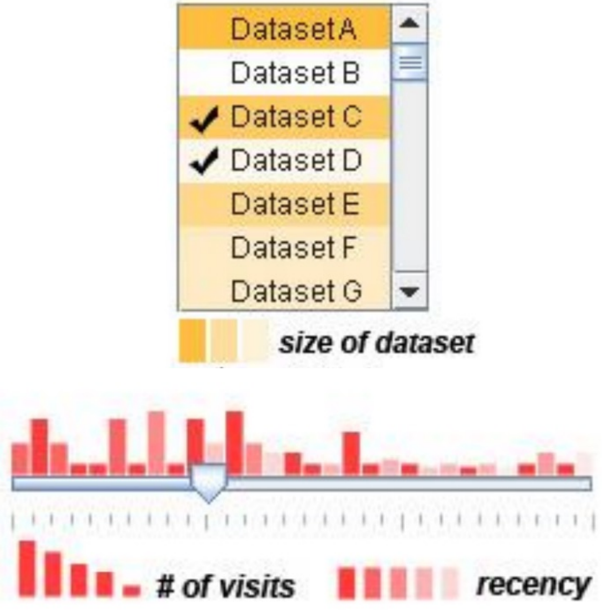
- Augmented encoding cues that guide user to drill down in areas that may be of importance
- Implemented as concise graphic, or secondary visual channels e.g. hue



Visualization on DAFT.ie

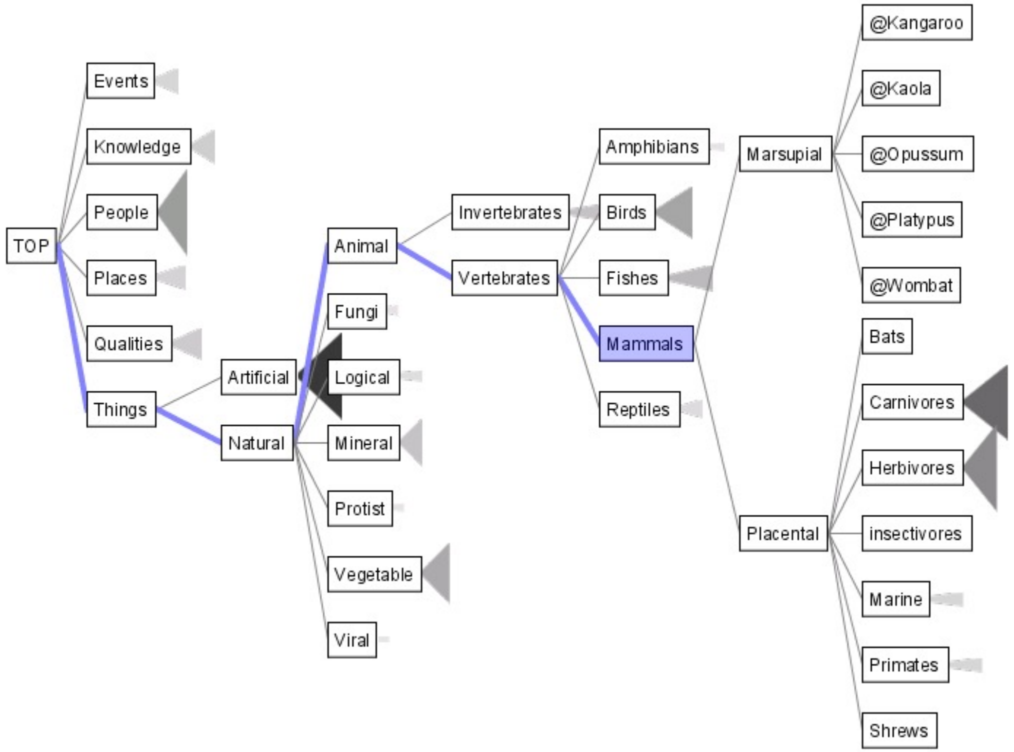


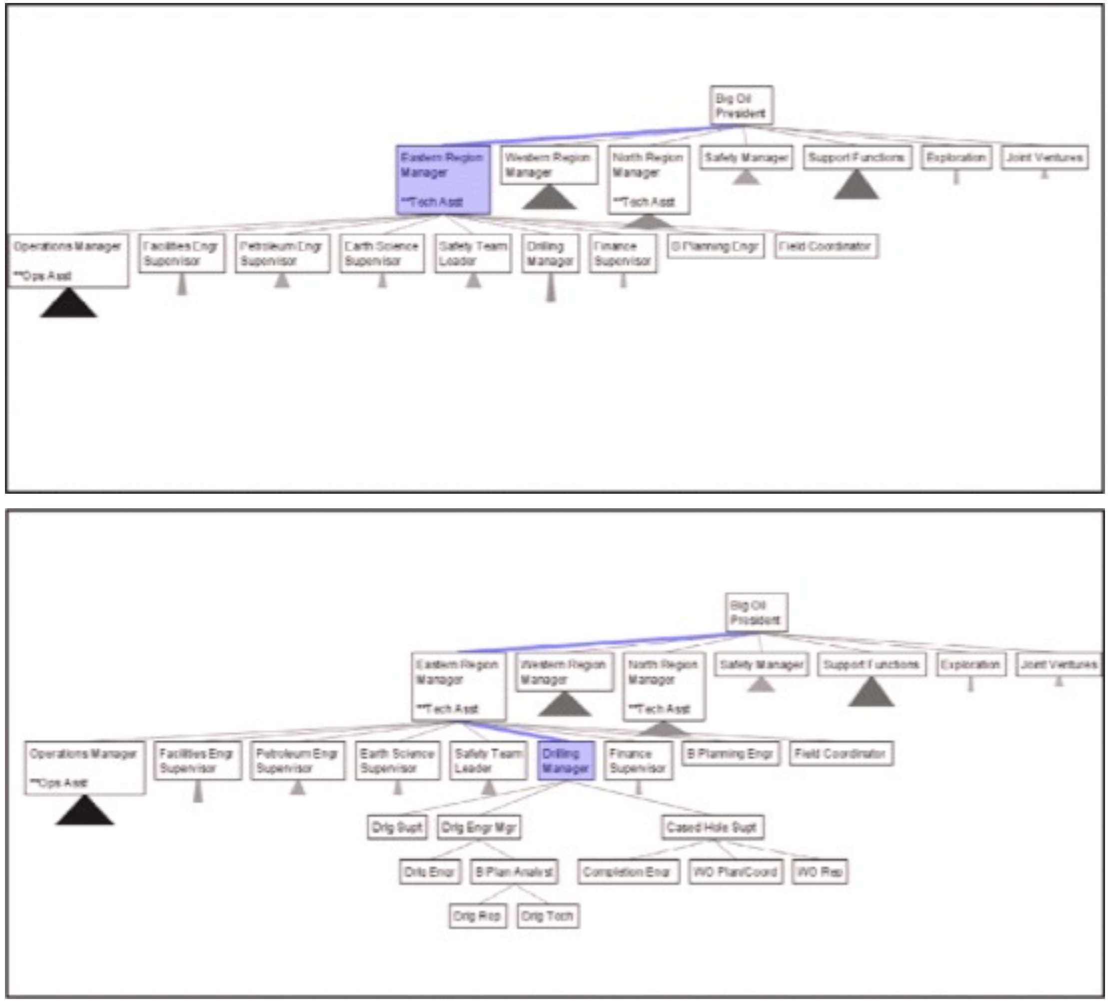
Homefinder [Williamson & Shneiderman 92]



Scented Widgets [Willet et al 2007]

Information Scent





SpaceTree (Grosjean et al 2002) includes integrated search and filter functions and automatic scaling to optimize use of available screen space. Preview icons summarize the topology of the branches that are not expanded.

Interactive Filtering and Highlighting

Selection: define a subset of the data

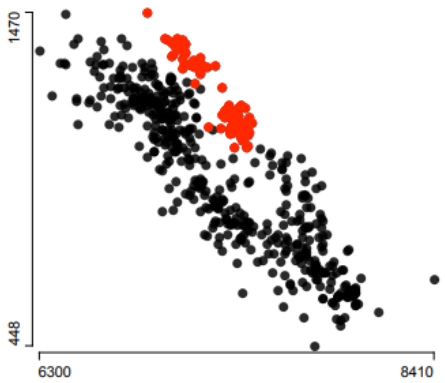
- **Brushing :** dynamic and transient
- **Painting :** persistent

Hot-selection: select subset and maximize view to visualize this

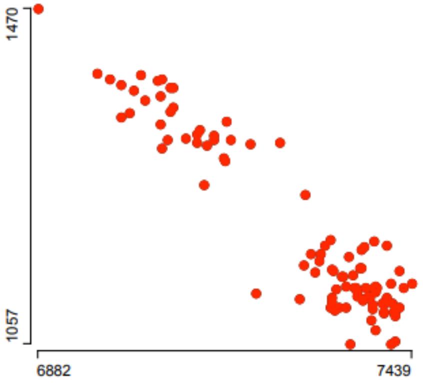
Shadowing: select a subset that should be hidden/de-emphasized

Highlighting: enhance a specific subset by a colour

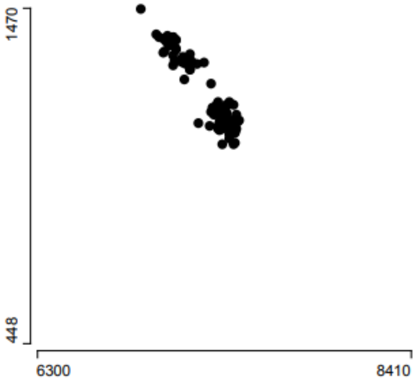
Colour Brushing: assign a colour map for several different subsets



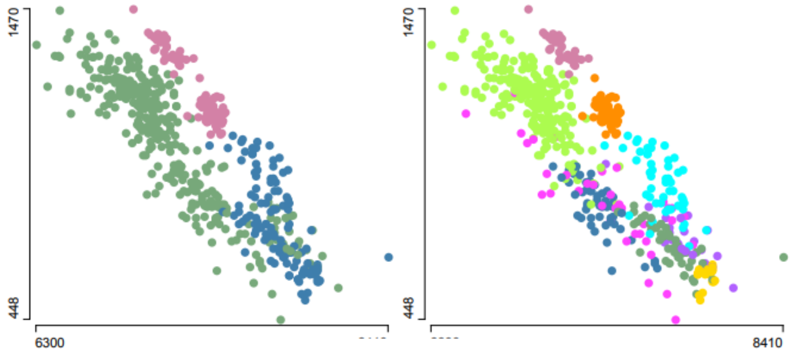
Selection + highlighting



Hot Selection



Shadowing

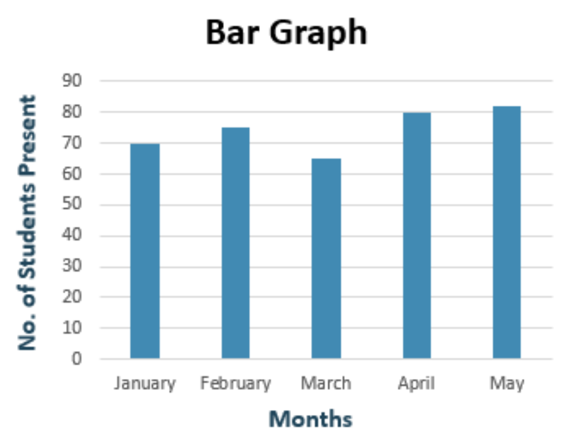


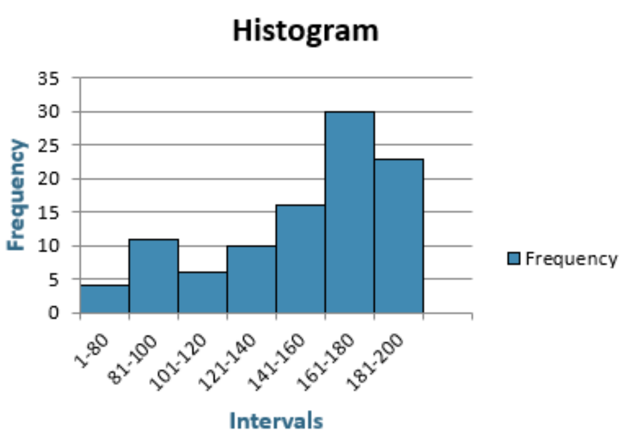
Colour brushing

Aggregation

Groups of elements represented by a derived element; summarizing

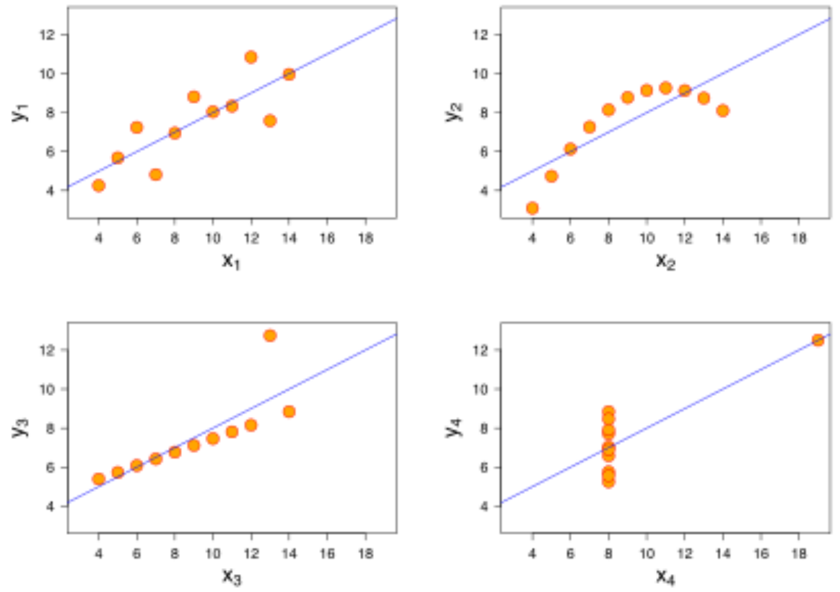
- Elements are merged together or abstracted rather than eliminated
- Can be used for both items and attributes
- e.g. averaging, max, min, count, sum





Challenges: avoid eliminating information of interest

e.g. Anscombe's Quartet (1973)



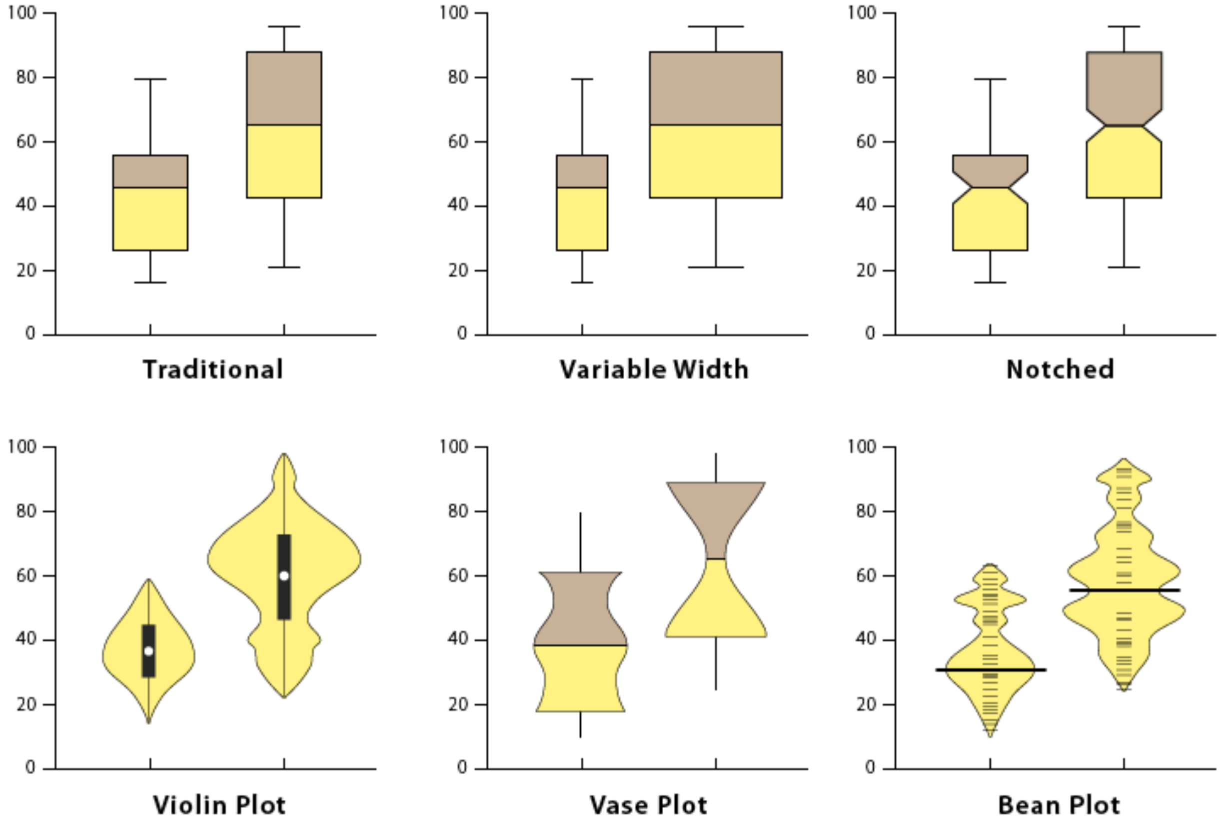
Aggregating Items: Box plot

a.k.a. box-and-whisker chart

Reduce visual complexity by item aggregation

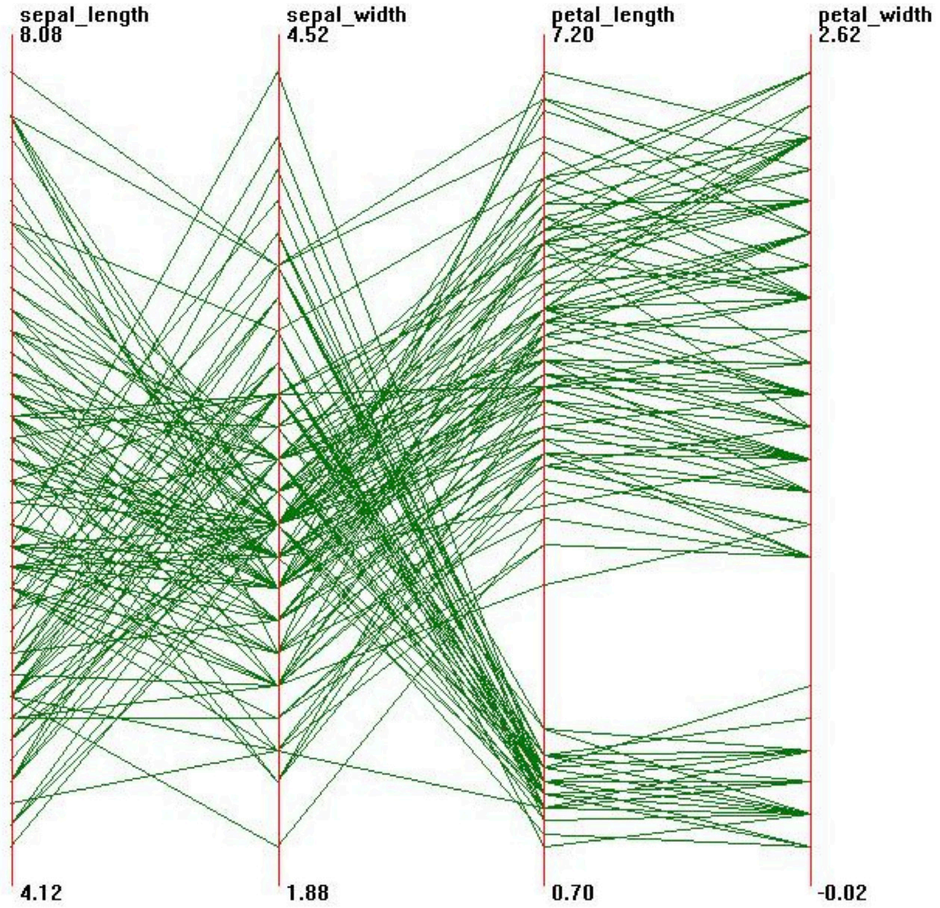
- Median encoded by central line
- Lower and upper quartile: boxes
- Lower upper fences: whiskers

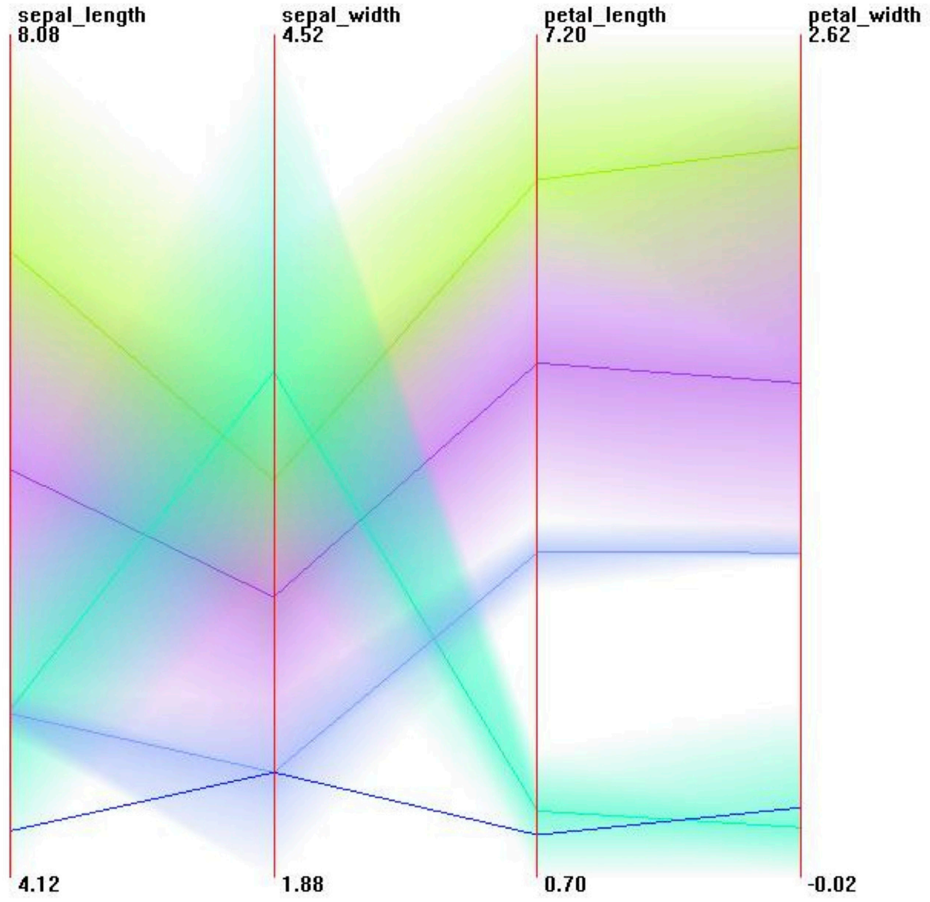
	Box Plot + variants
Data	Table: at least 1 categorical key attribute and 1 quantitative value attribute. Typically large no. of items.
Derived Data	5 quantitative attributes: Median, quartiles, min, max , etc.
Encoding	Position + area encoding of derived quant. data. (Size and shape for further details e.g. density)
Tasks	Summarize, derive, explore distribution



Box plot & variants: variable width (size of group), notched plot (confidence), violin plot (probability density), vase plot (density), bean plot (density over bands)

Aggregating Items: Parallel Coordinates

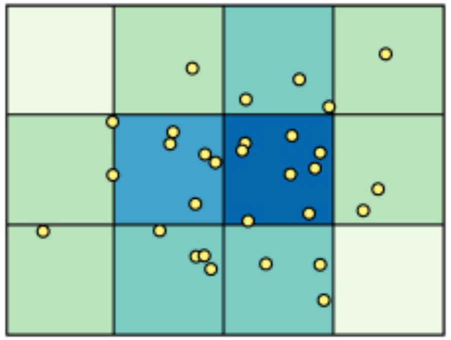




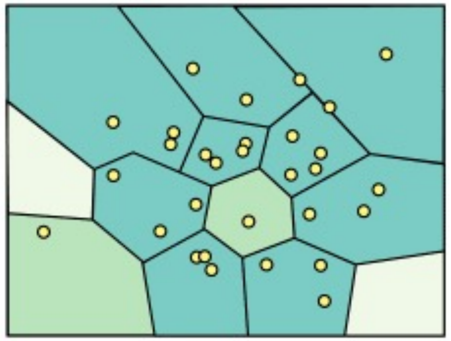
Ideally suited for tasks involving analysis of trends, outliers, extremes, correlation. Thus some visual abstraction techniques can address overplotting problems and enhance visualization for such tasks

Spatial Aggregation

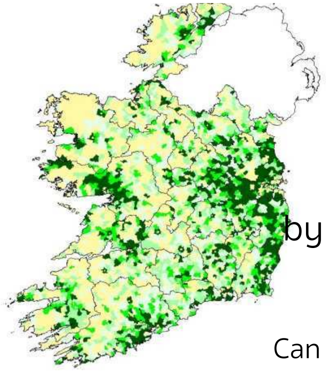
Take spatial information into account for aggregation e.g. proximity, locality

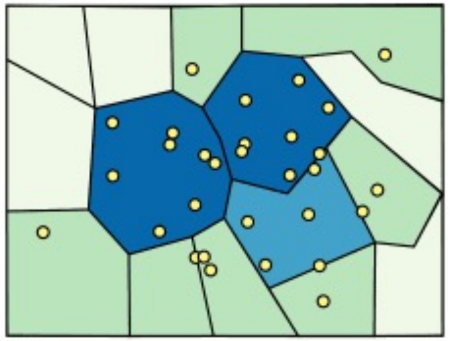


by space



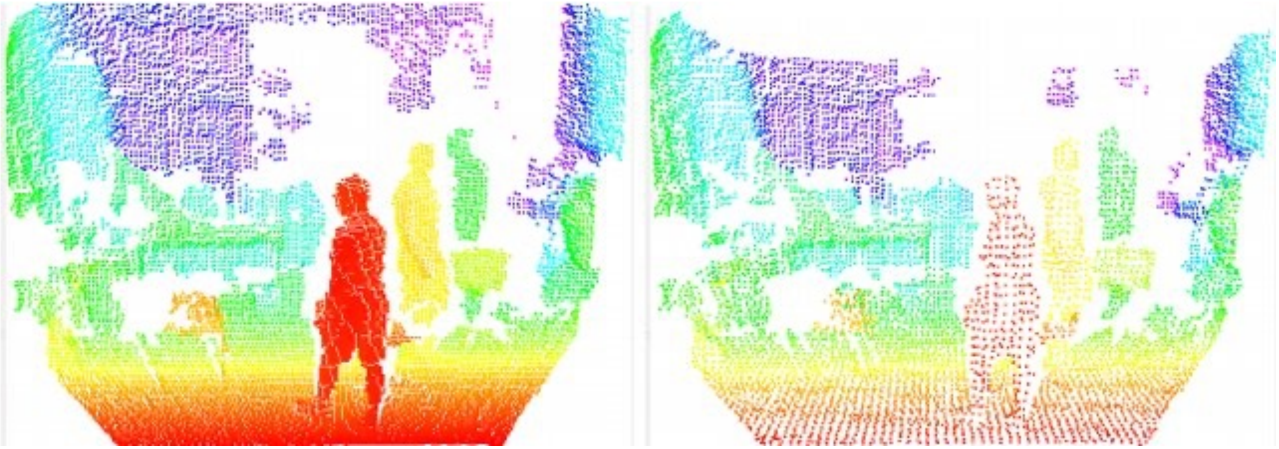
by density



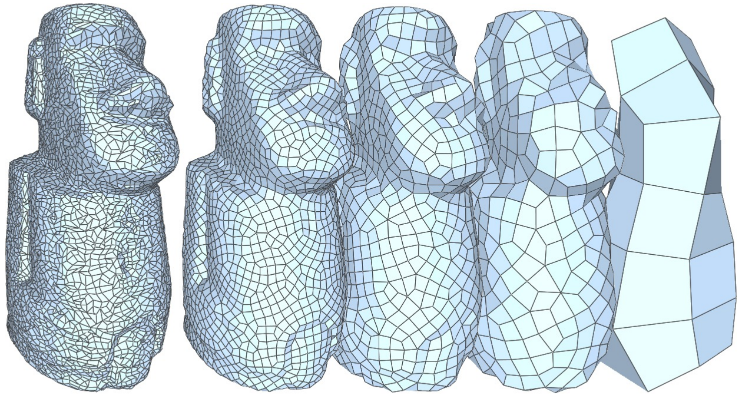


by conventional regions (e.g. district)

Can lead to bias e.g. Modifiable aerial unit problem (MAUP)



by value e.g. voxel filtering [Munaro et al 2013]



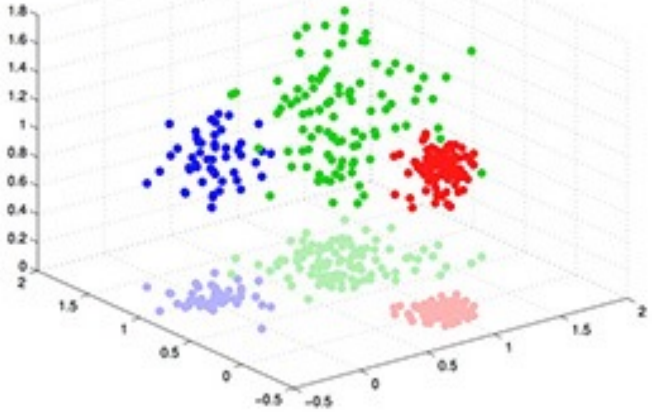
by feature
e.g. Geometrical simplification [Tarini et al 2010]

Aggregating Attributes

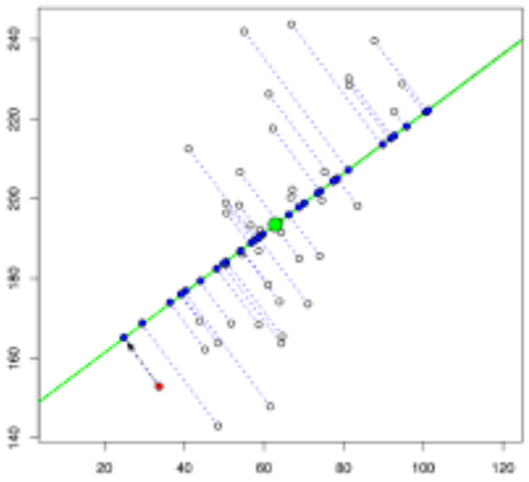
a.k.a. Dimension Reduction: derive low-dimensional target space from high-dimensional measured space

- Aim: Reduce elements (attributes) yet preserve meaning.
- Assumptions:
  - There exists some latent information (hidden structure): can't directly measure what is relevant to task
  - Redundancy in original data set: true dimensionality conjectured to be smaller than that of measurements

Straightforward if some homogeneity between attributes e.g. projection

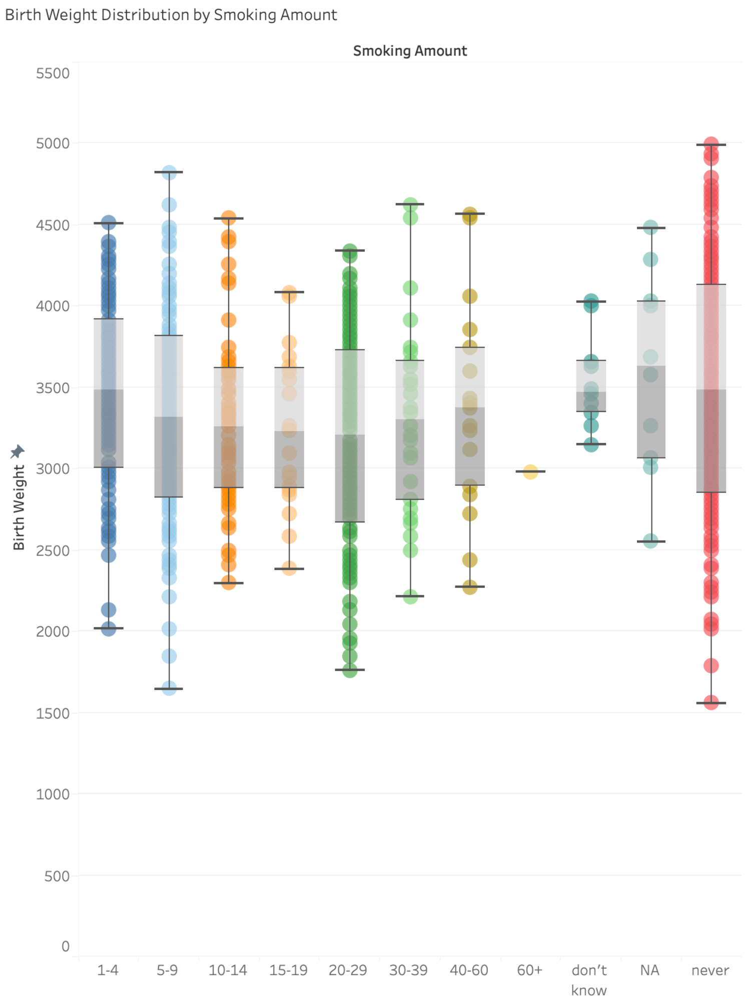


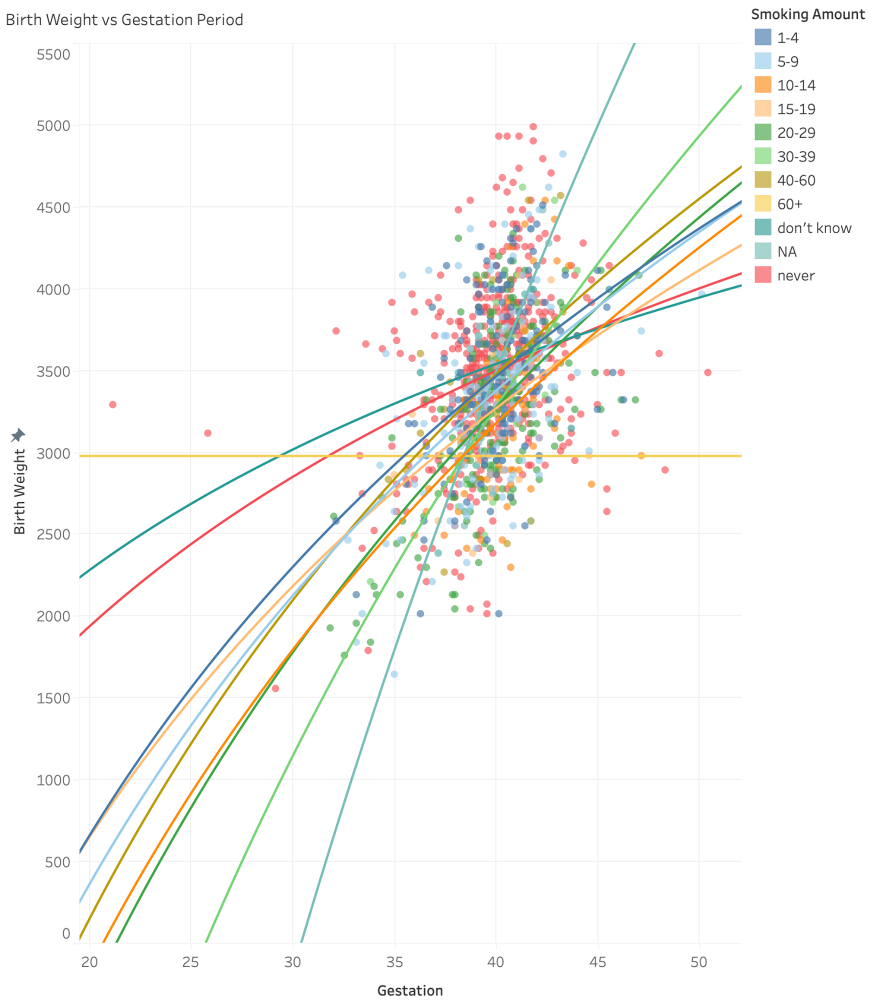
Projection



PCA

Example: **Aggregation in Tableau**





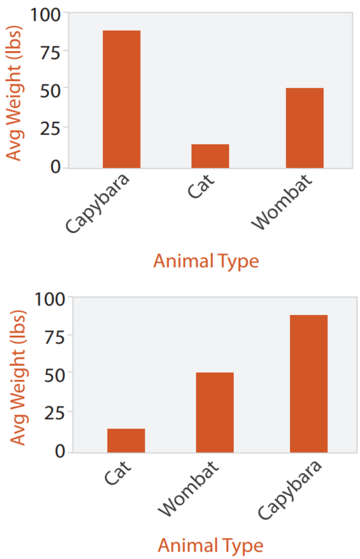
3. **Manipulate** (Modify Views)

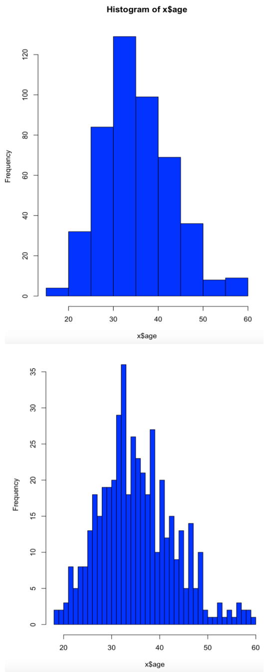
Aims: modify visualization idiom/parameters to improve expressiveness of data relevant to task

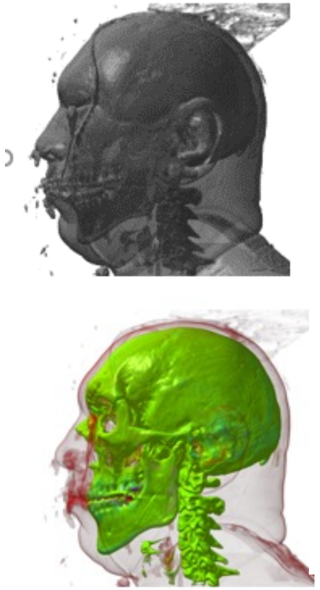
Changes can be driven by visualization designer OR user i.e. interactive navigation (discussed later)

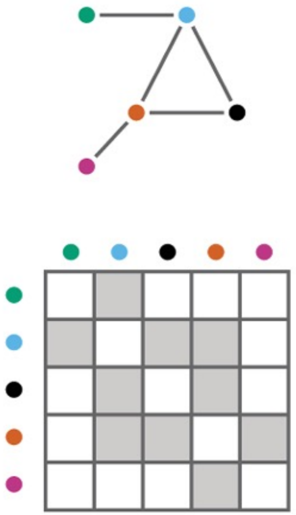
What can be changed:

- Rearrange / sort
- Pan and Zoom
- Level of detail
- Encoding parameters
- Idiom used

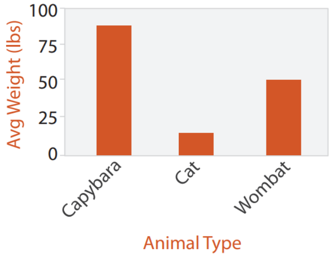


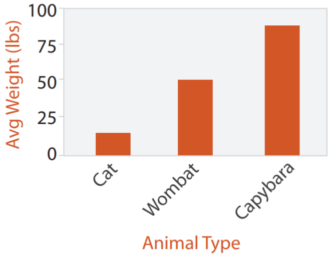


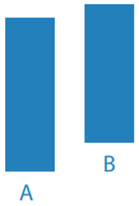


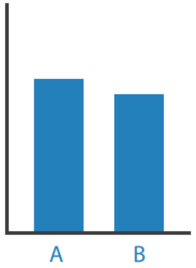


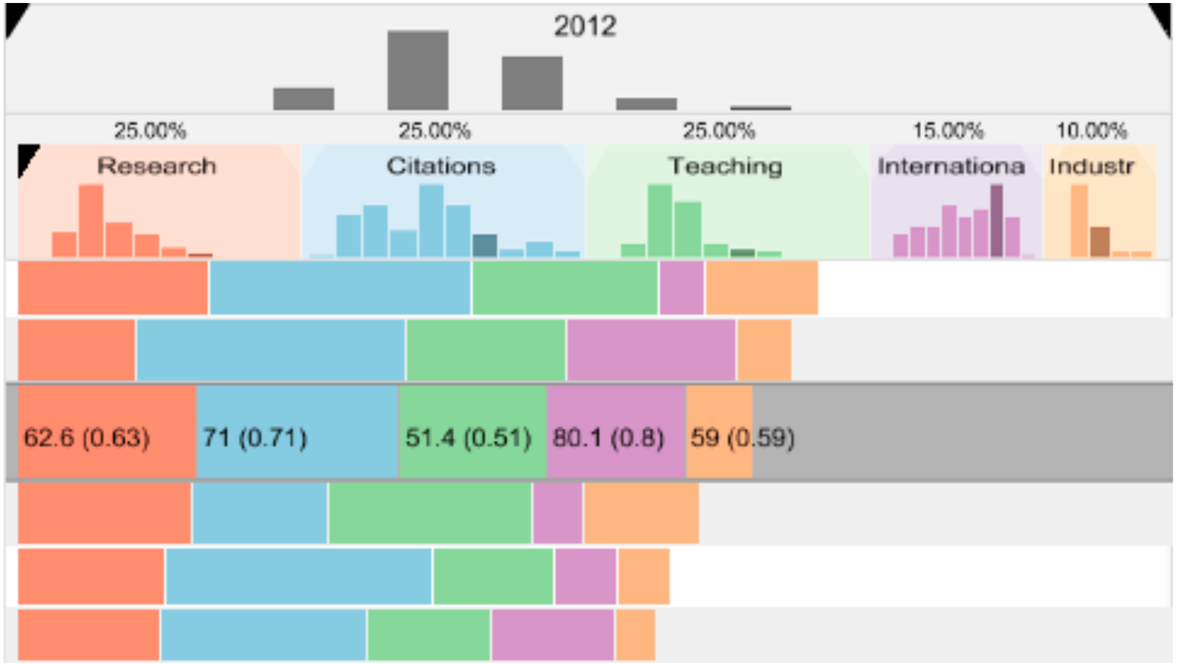
Rearrange and Sort

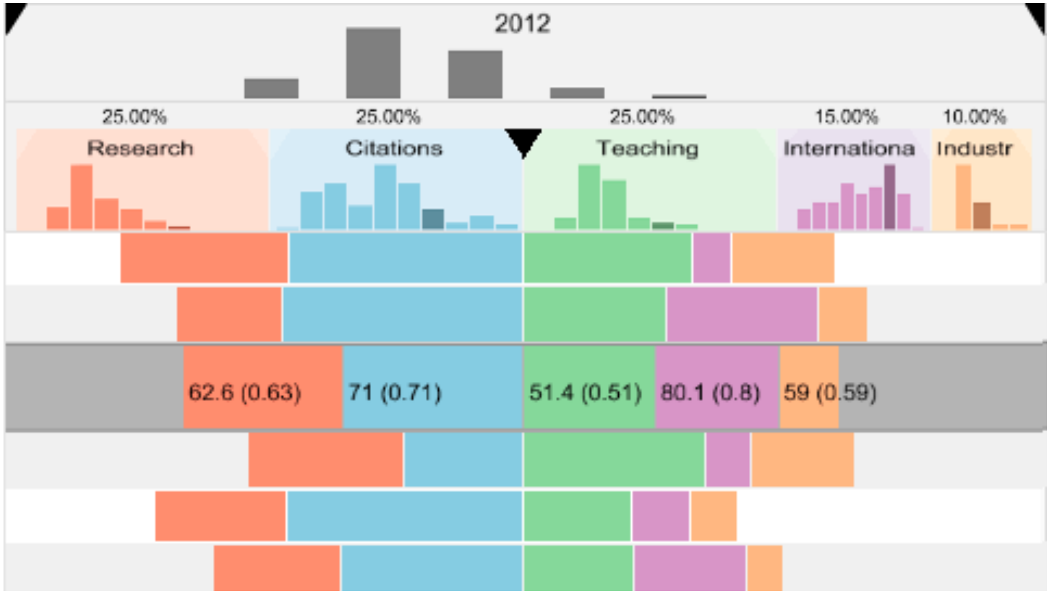


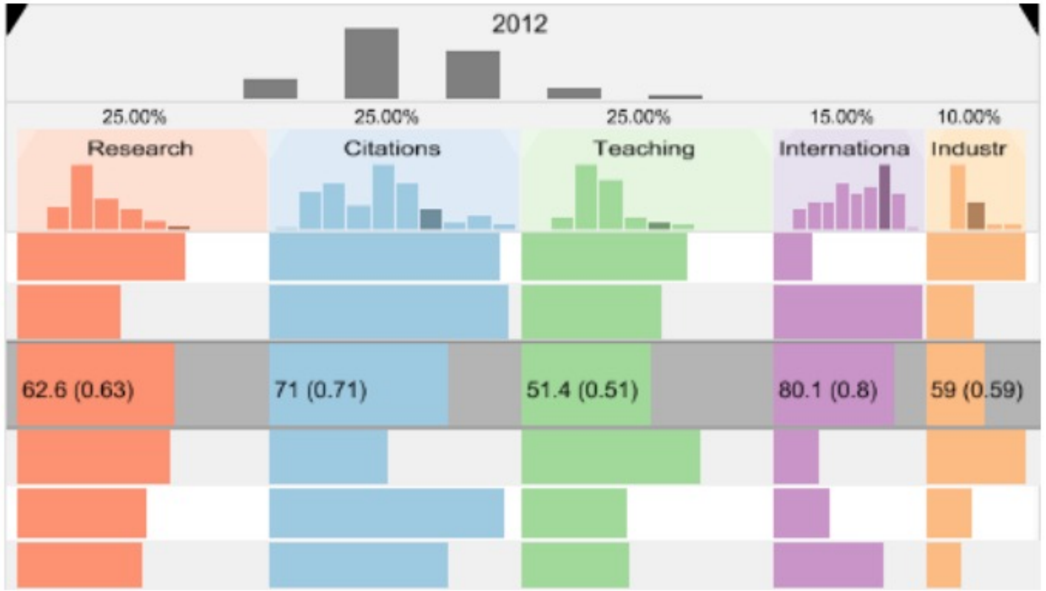






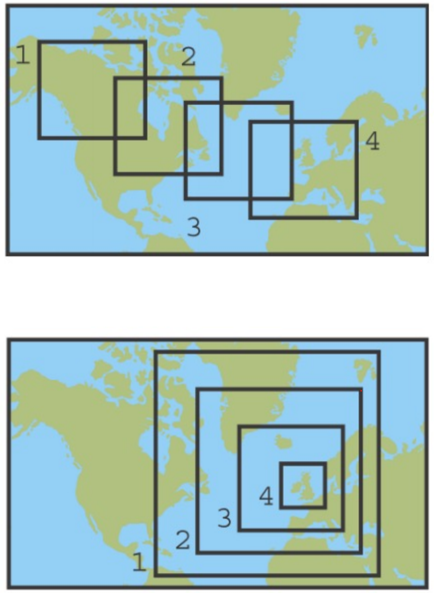






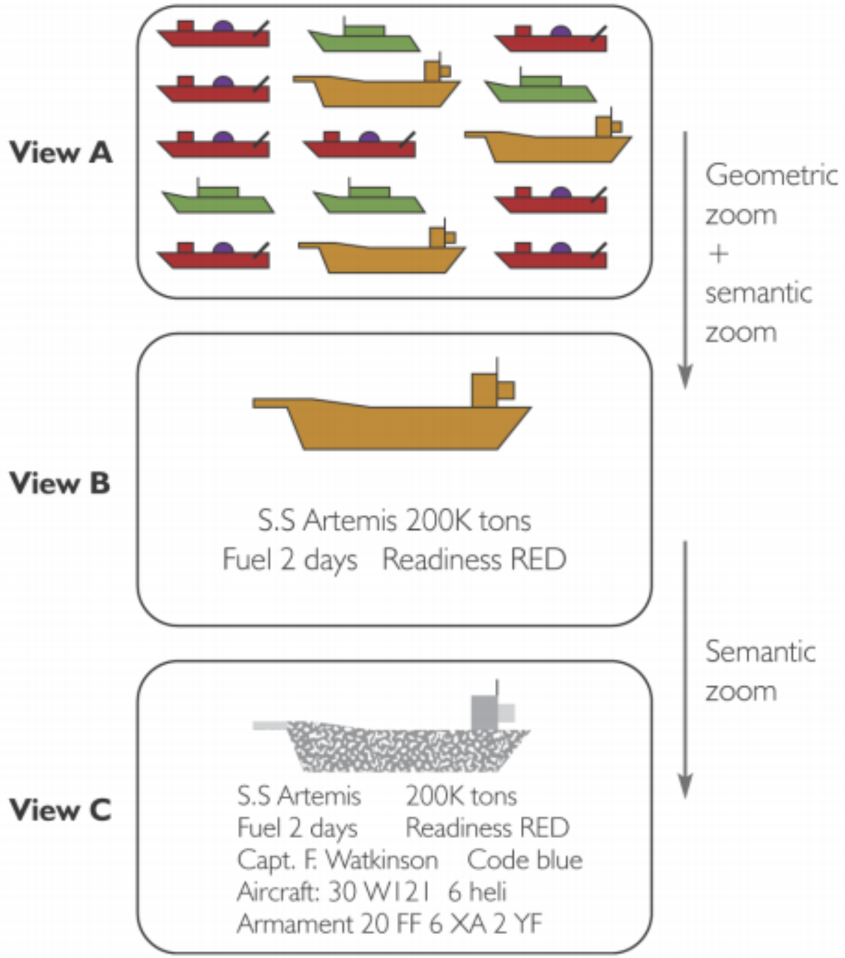
From: **LineUp,** a tool for Visual Analysis of Multi-Attribute Rankings (Gratzl 2003)

Pan and Zoom



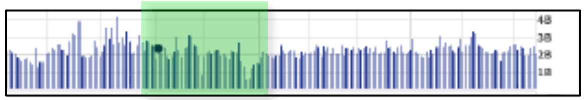
1
2
3
4

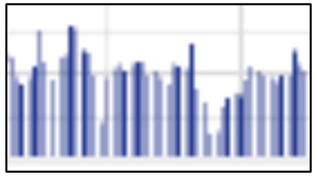
Hudson Bay
Canada
MANITOBA
SASKATCHEWAN
ONTARIO
NORTH DAKOTA
MINNESOTA
Canada
MANITOBA

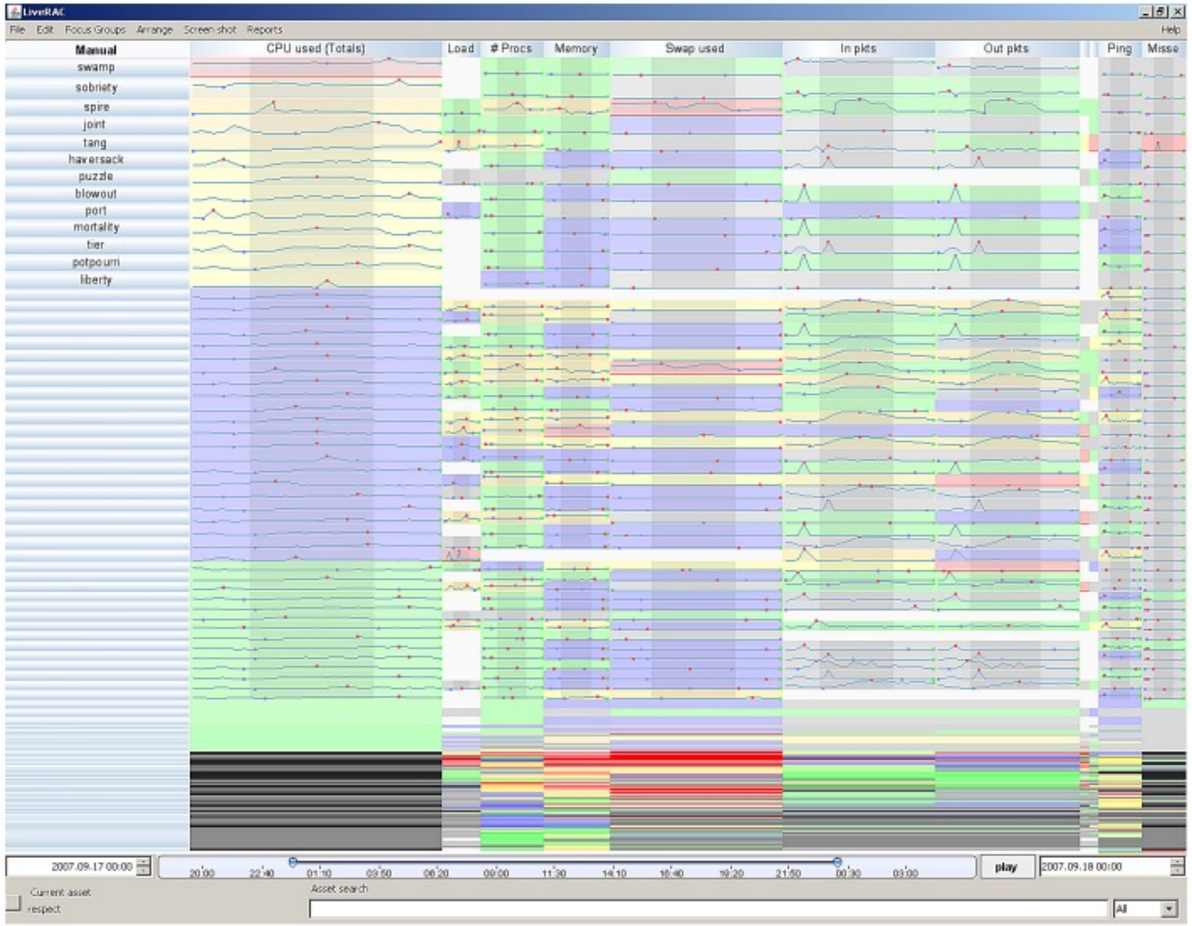


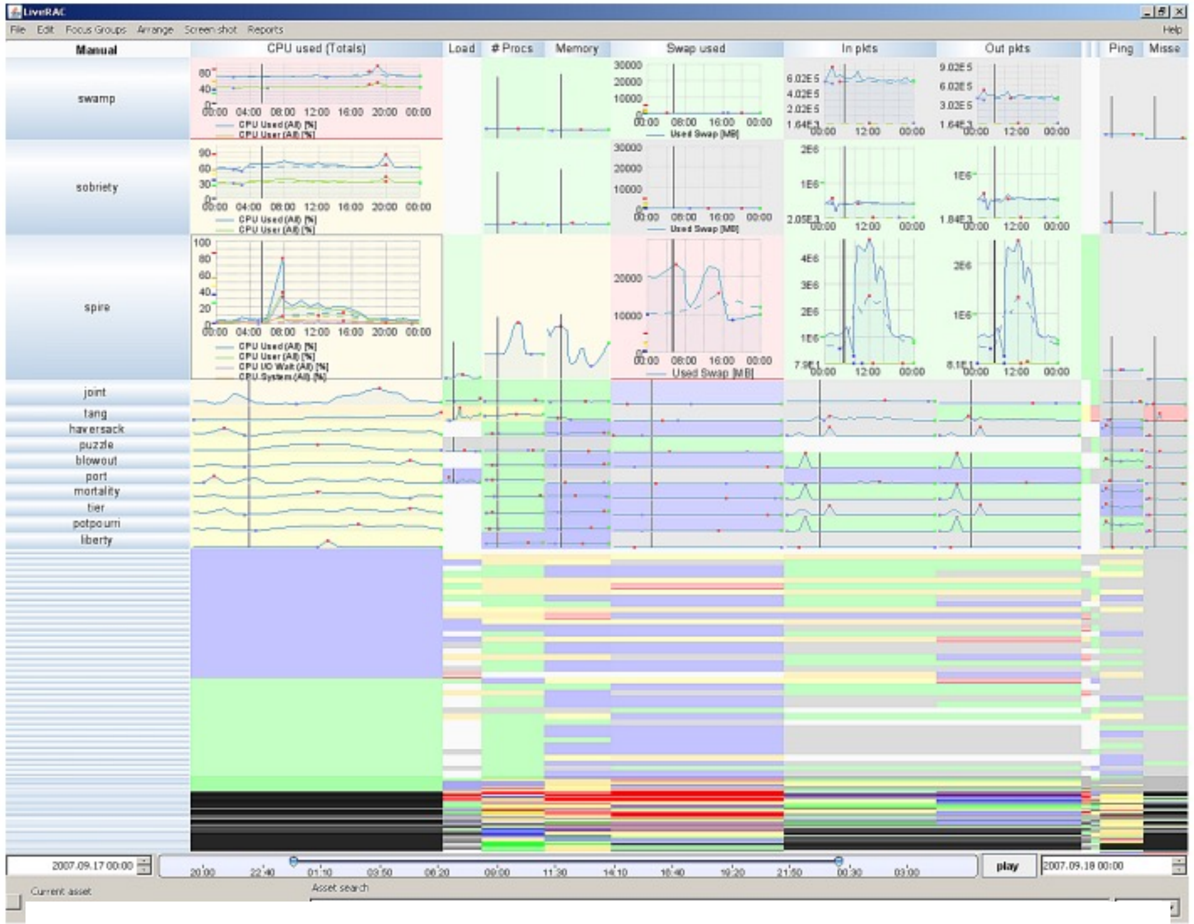
View A
View B
View C
Geometric zoom + semantic zoom
Semantic zoom
S.S Artemis 200K tons
Fuel 2 days Readiness RED
S.S Artemis 200K tons
Fuel 2 days Readiness RED
Capt. F. Watkinson Code blue
Aircraft: 30 W121 6 heli
Armament 20 FF 6 XA 2 YF

Semantic Pan and Zoom



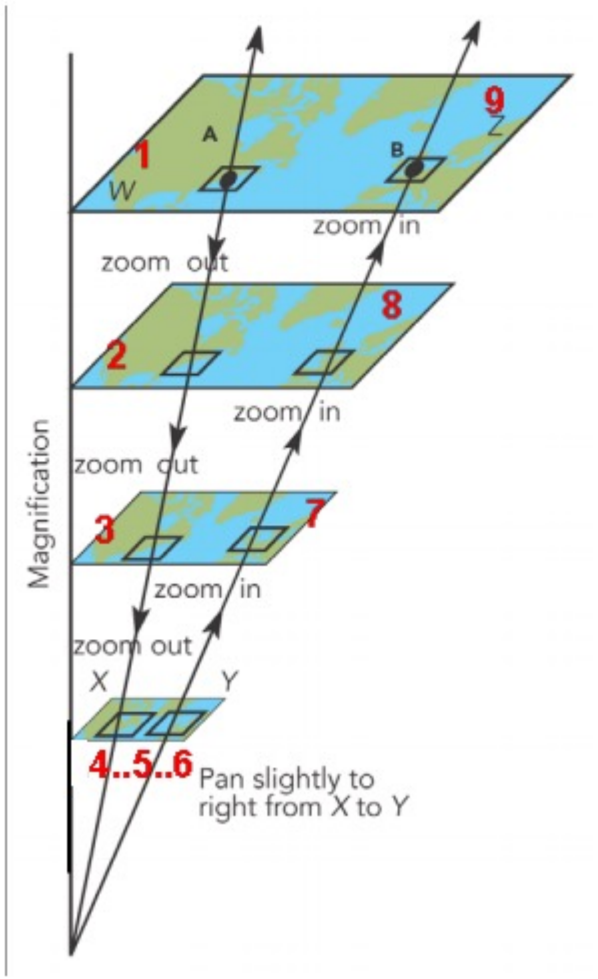




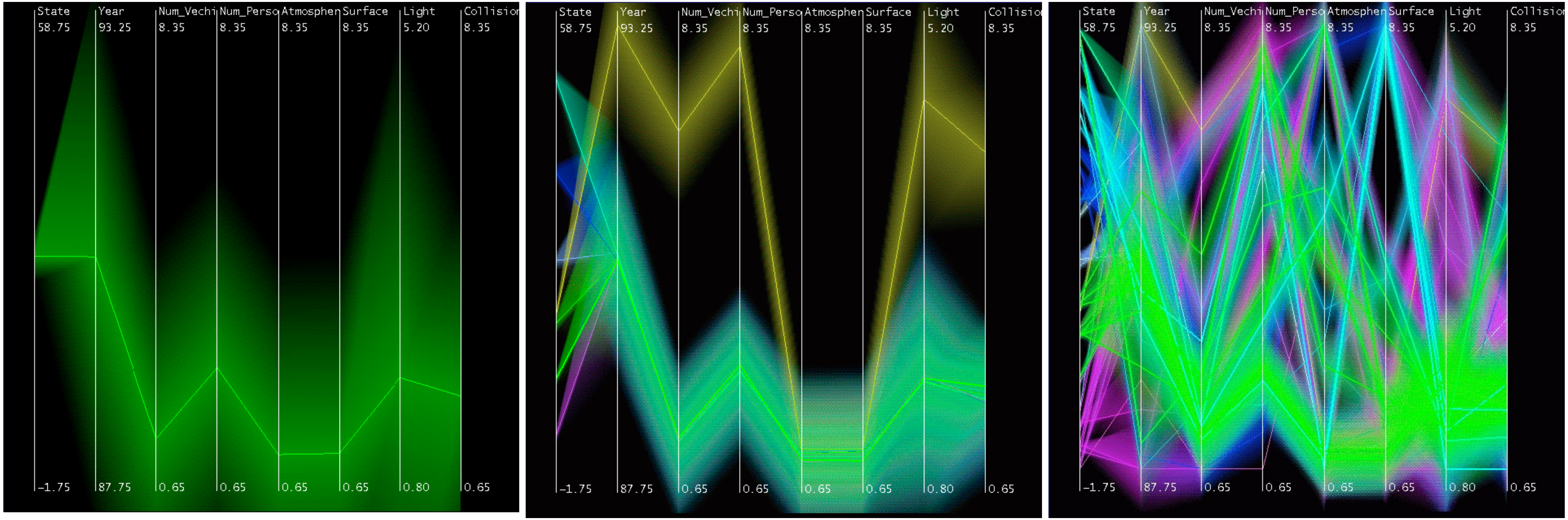


SEMANTIC PAN AND ZOOM : LiveRAC (Maclachlan et al 2008)

Change Level of Detail

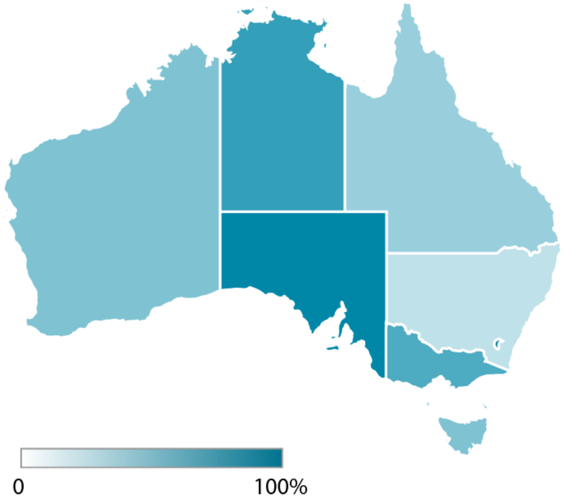


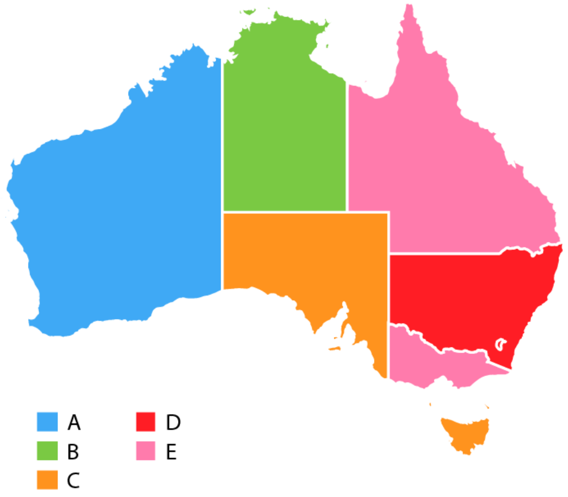
Change Level of Detail

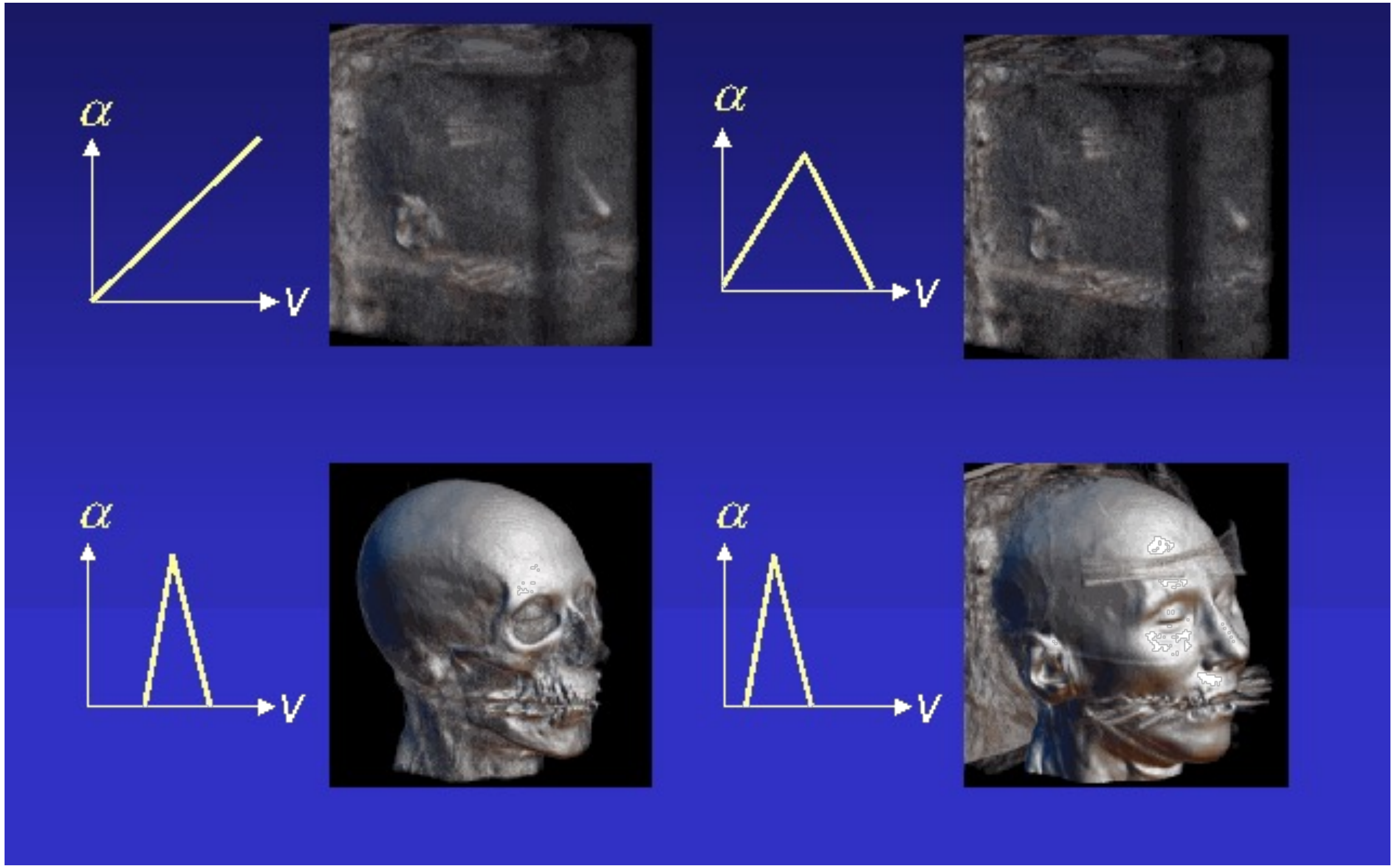


Multiple levels of detail in a hierarchical parallel coordinate visualization (Fua et al 1999)

Change Encoding Parameters



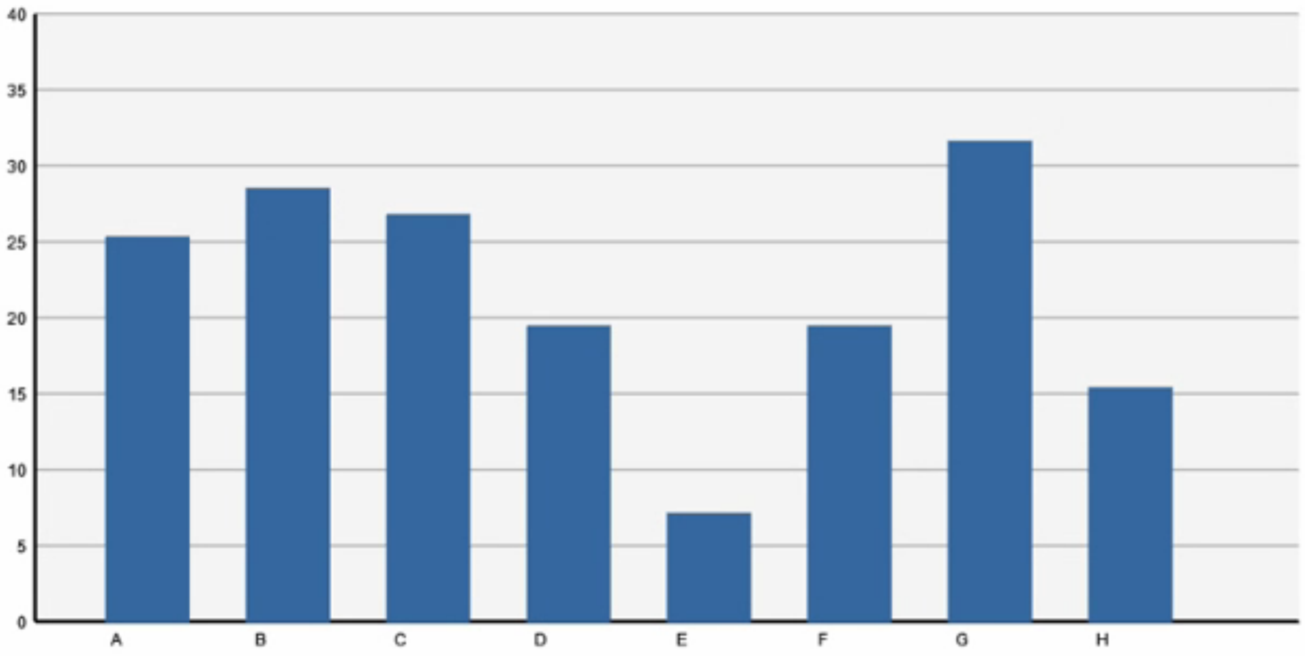




Change Idiom

Use different choices of encodings for the same data

Important to maintain a sense of context between idioms otherwise the benefits of using multiple encodings is lost



Animated transitions to preserve context across idioms

Video © Stanford Visualization Group

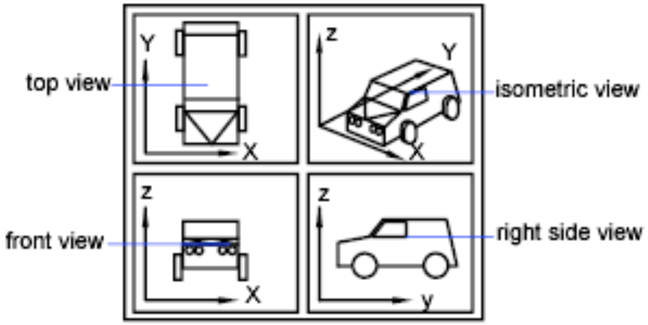
4. **Facet** (Combining Views)

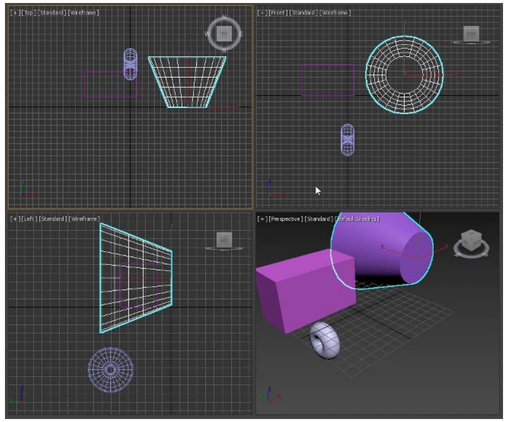
Aims: Combine views to improve expressiveness of visualization

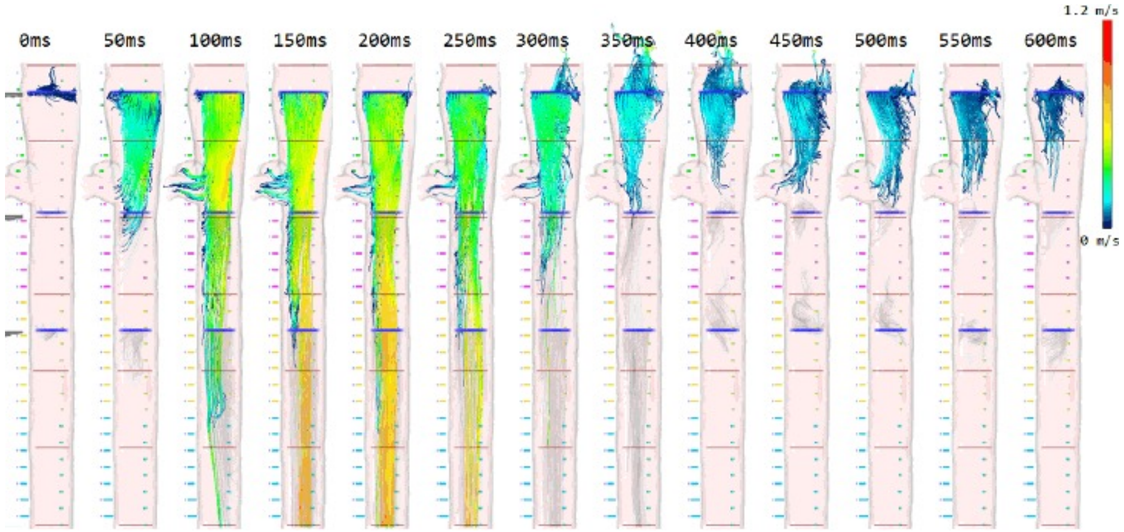
When a single view is not effective, split visualization into multiple simultaneous views (or "*facets*")

- Advantages over animated/changed views (working memory)
- Need to balance against perceptual clutter; expressiveness
- Need to decide how to split elements usefully
- Need to coordinate views: e.g. common elements identifiably linked across views

Main strategies: Juxtapose; Superimpose; Embed







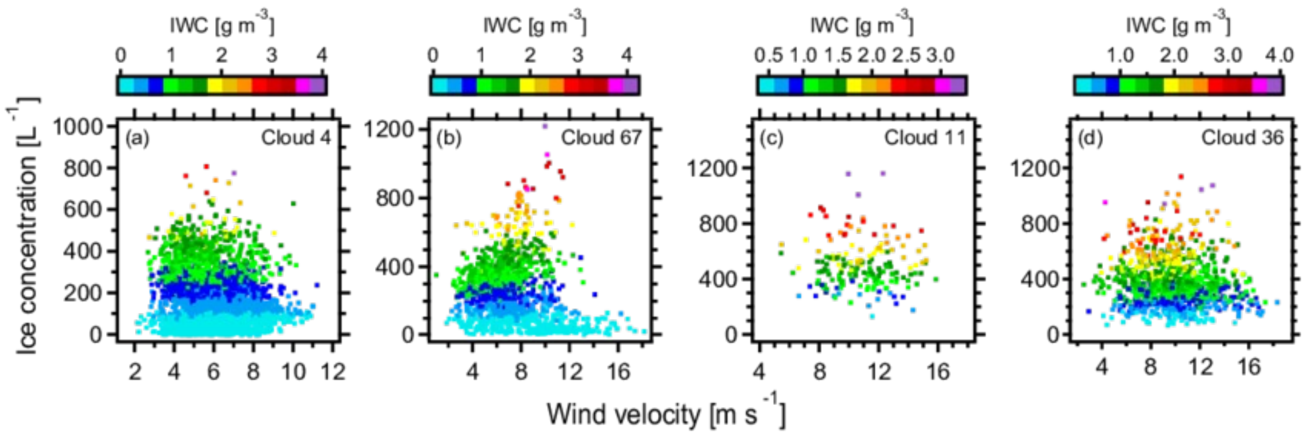
Juxtaposed Views

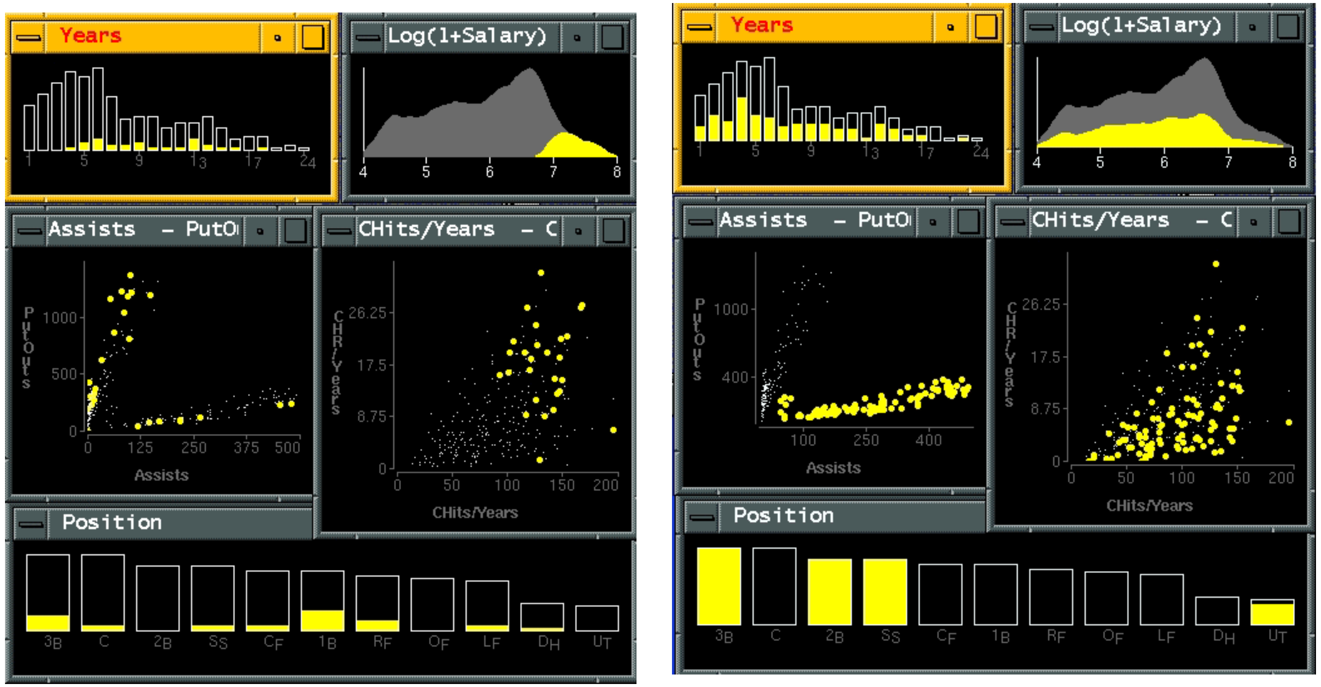
Coordinated views shown spread out in space side by side

- Unform : identical encoding shared across facets – different elements only a.k.a. *small multiples*
- Multiform : different encoding idioms across facets

Challenges:

- effective use of screen real-estate
- linking meaning between views

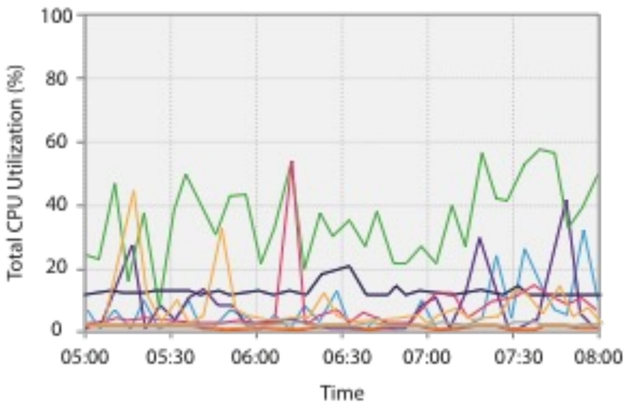


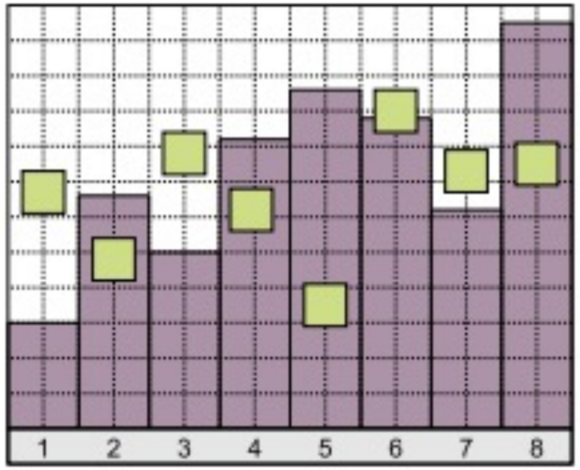


(Wills 95)

Superimposed Views

Combine multiples layers by stacking them over each other in a composite view



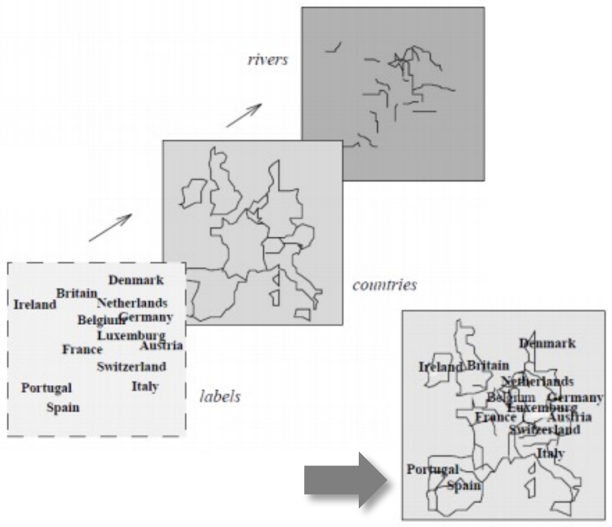


All elements have same horizontal & vertical extents

Quite common in 2D spatial Maps

Challenges:

- perceptual clutter
- non-uniform density across layers



Embedded Views

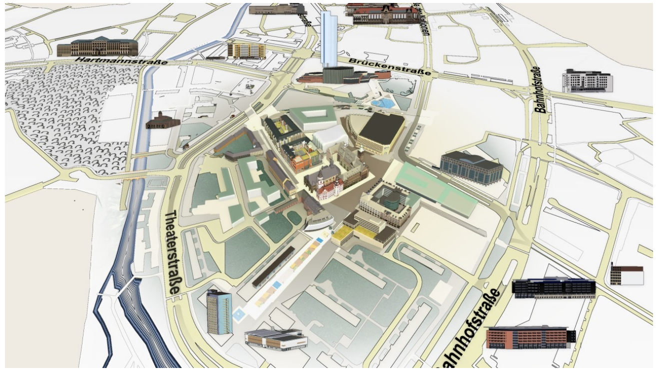
Detailed information of a selected subset of data visualized in a subset of the display space

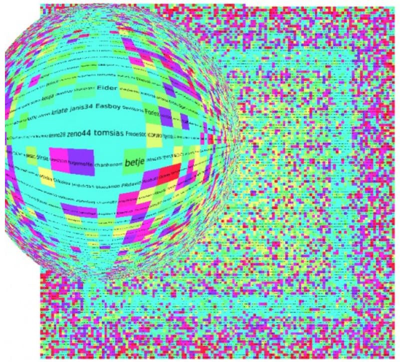
Most common example: focus & context

- Embed detailed information about a selected subset of the data (*the focus*) within a single view that also contains overview information (*the context*)
- Combines Reduction and Multiple views

Main strategies for composition

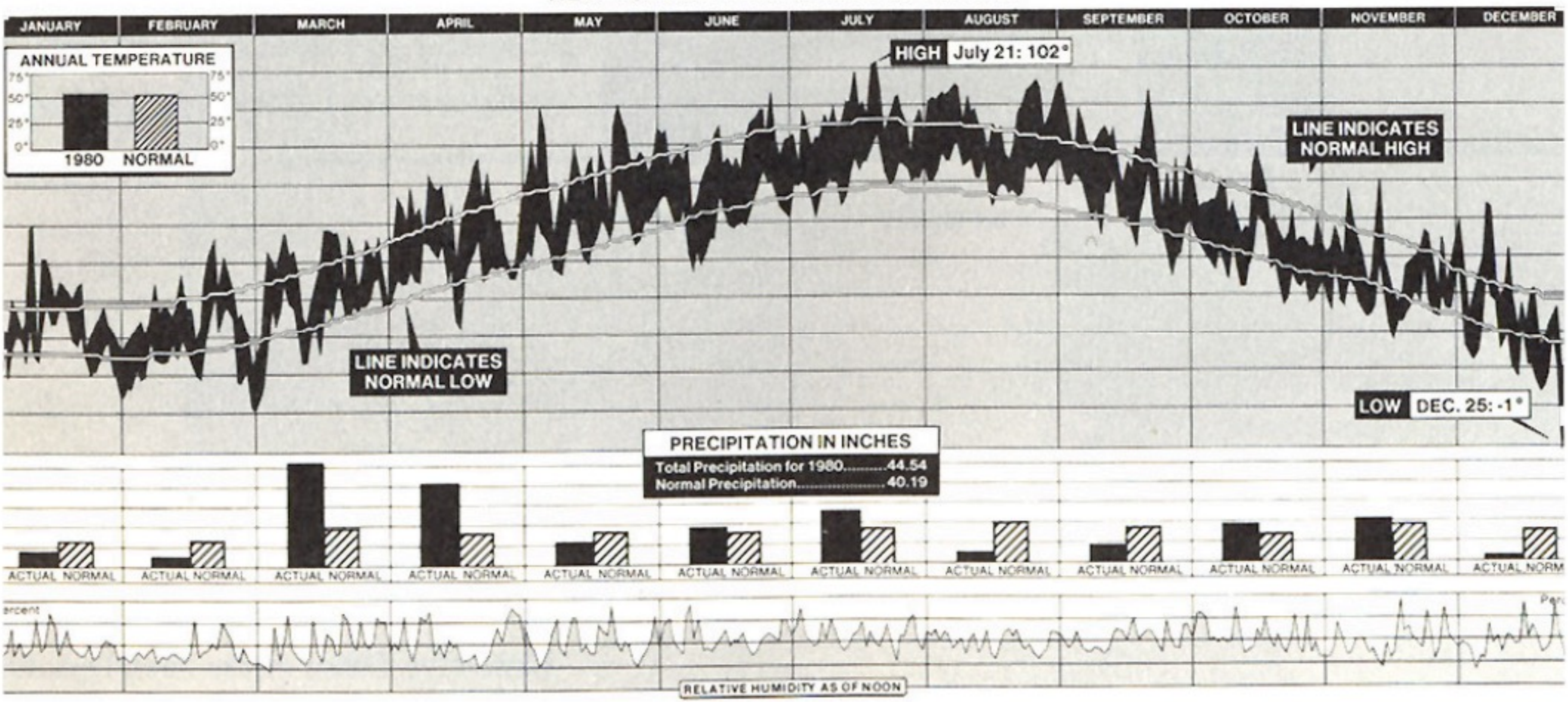
- Nesting
- Distortion





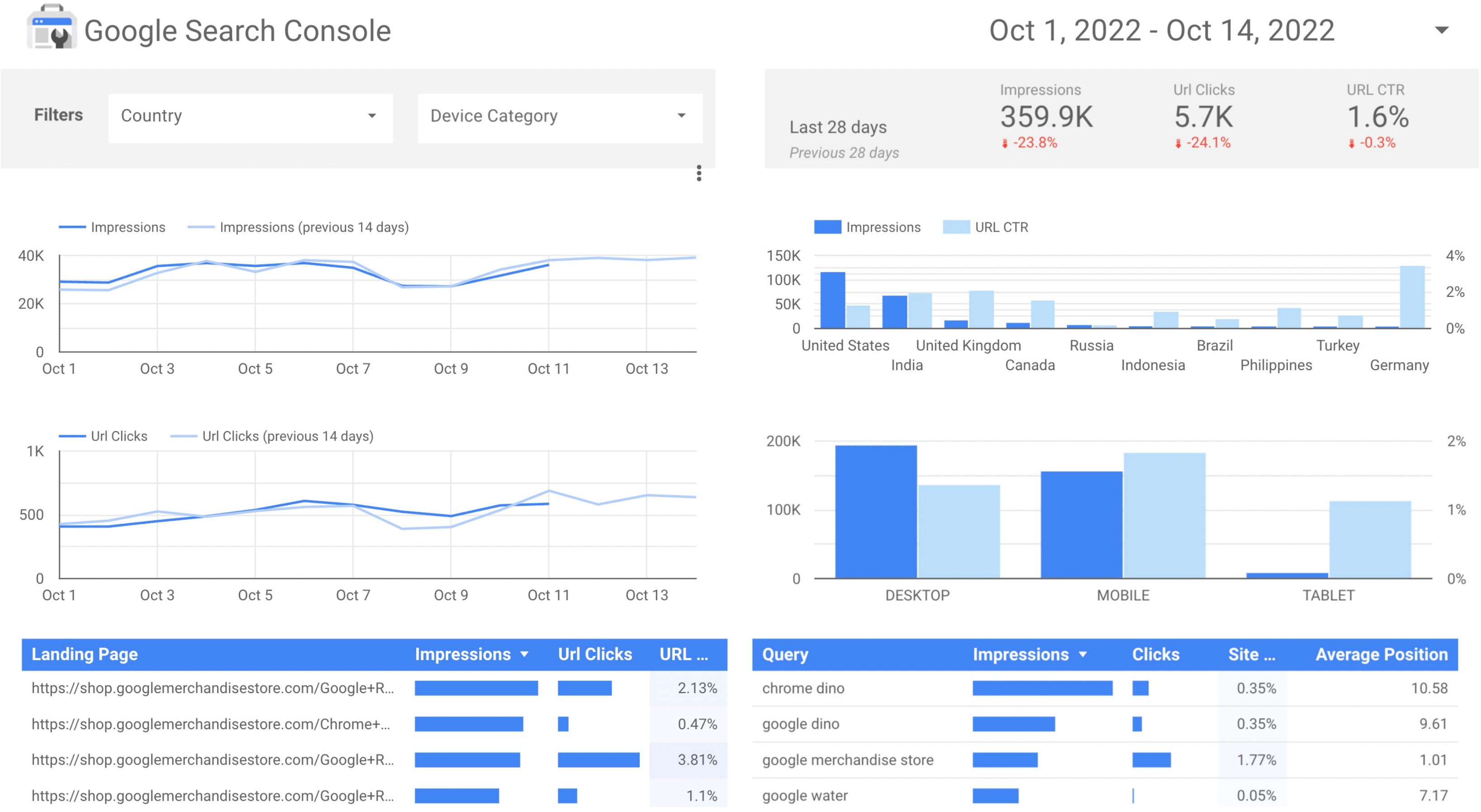
Hybrid Example

Isomorphism and visual juxtapositions should be used (if possible) to reveal connections



New York Weather History 1980 (from Tufte 1983)

Hybrid Example



Simplify Views: item filtering using drop down menu by ordinal attribute, date; categorical attributes, country and category. Also the scrollable-bar chart at bottom is reduction (based on which ever attribute is used to sort the bars)

Combine views: juxtaposition of several chart types, and some elision (in stacked line chart, bar charts with combined y-axis on the right panel)

Change views: sorting/alignment of bars in bottom panels, automatically changing encoding parameters in the adaptive axis ranges for bar/line charts

Google Search Console Statistics - Google Data Studio Example (2022) https://datastudio.google.com/u/0/reporting/0B_U5RNpwhcE6QXg4SXFBVGUwMjg/preview/

Related Readings

On Navigating data, Juxtaposition, Partitioning, Superimposition, Focus and Context

- Munzner (2014) Visualization Analysis and Design. Chapters 11, 12, 13*, 14. [Available in Library Reading Rooms]
- Also in Munzner (2012) Visualization Design and Analysis: Abstractions, Principles, and Methods Draft. Sec 8.4-8.6 and Ch 9.

Available Freely at [https://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/ClassLectures/PDF-old/book.120803.pdf#page=187]