Visualization

Interaction and Information Navigation

Klaus Mueller

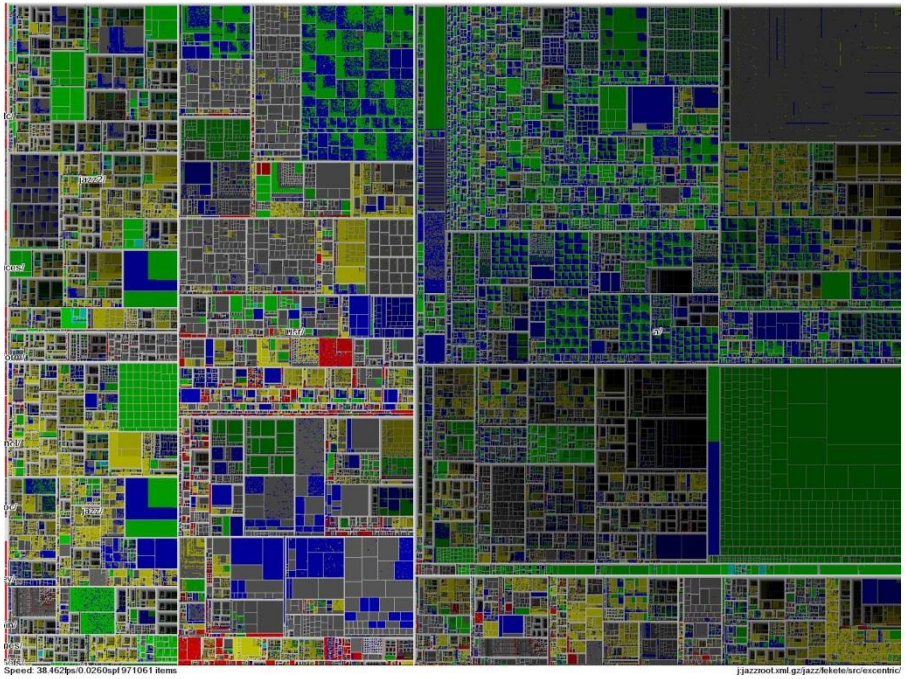
Computer Science Department

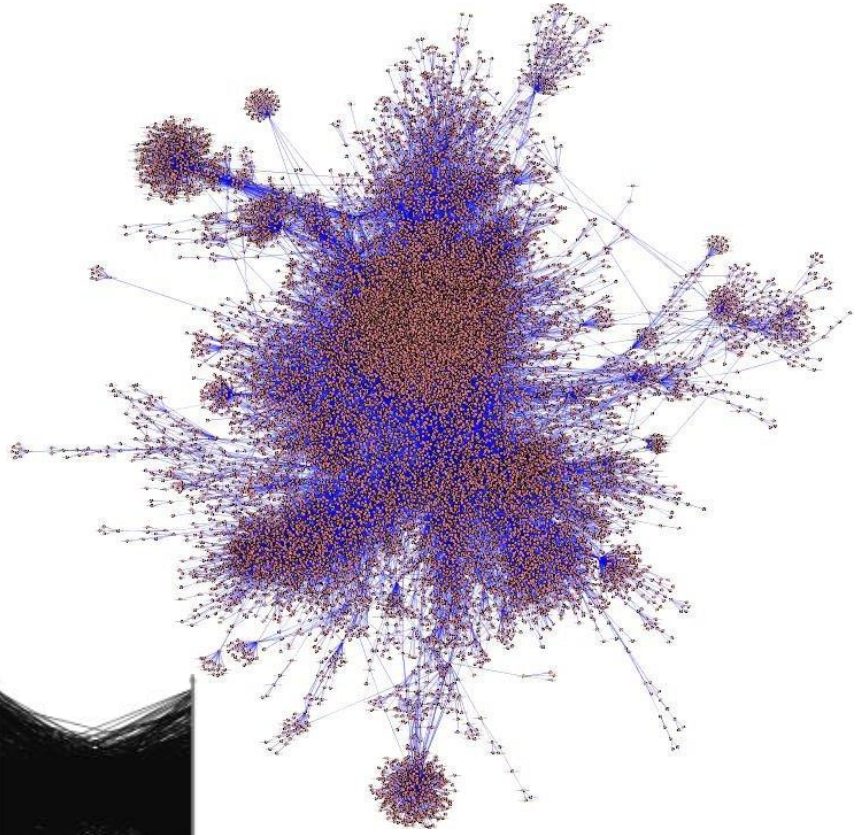
Stony Brook University

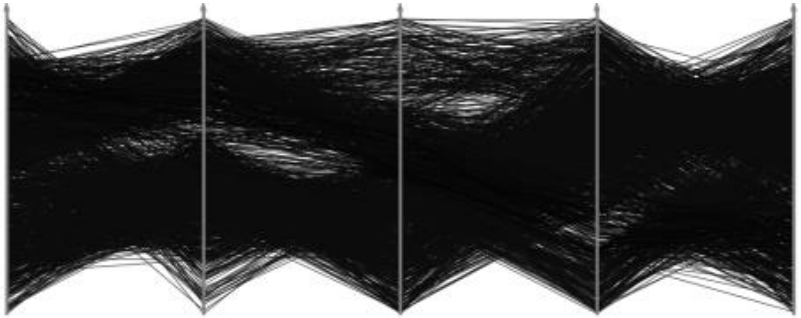
Too Much Data?

How can we deal with data overload

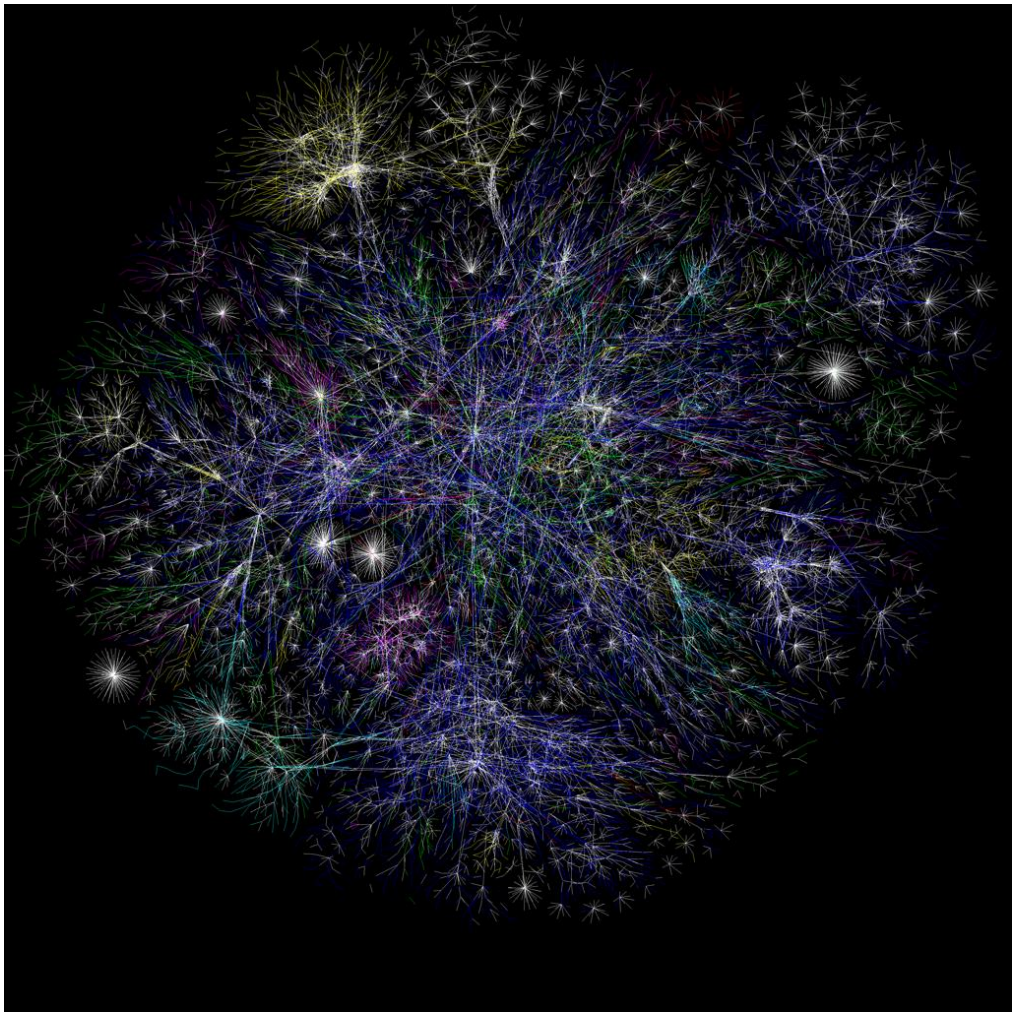
- see the forest for the trees (or the other way around)







Too Much Data?



Internet routes (1/15/2005)

(NY Museum of Modern Art)

The Key to Overcome the Data Deluge: Interaction

Allow users to control what is currently shown:

- level of detail
- extent of the data (spatial, values)
- aspects of the data (attributes)

But do not leave the user lost in the forest

- provide navigation hints

Two powerful paradigms:

- overview, and detail on demand (forest and trees)
- focus and context (trees and forest)

Interaction needs to be interactive (as in responsive)

- user needs get quick visual feedback on actions

Interaction: Key to Visual Analytics

Puts the human in the loop

- appeals to human's expertise and intuition

Requires a suitable human-computer interface

- recall the lectures on color and perception

Interaction can help with:

- making sense of it all
- putting things in proper context
- data overload (scalability)
- telling stories with data (explain findings to others)

Evaluate effectiveness

- do human users actually benefit?
- user studies!

A Taxonomy of Fundamental Interaction Types

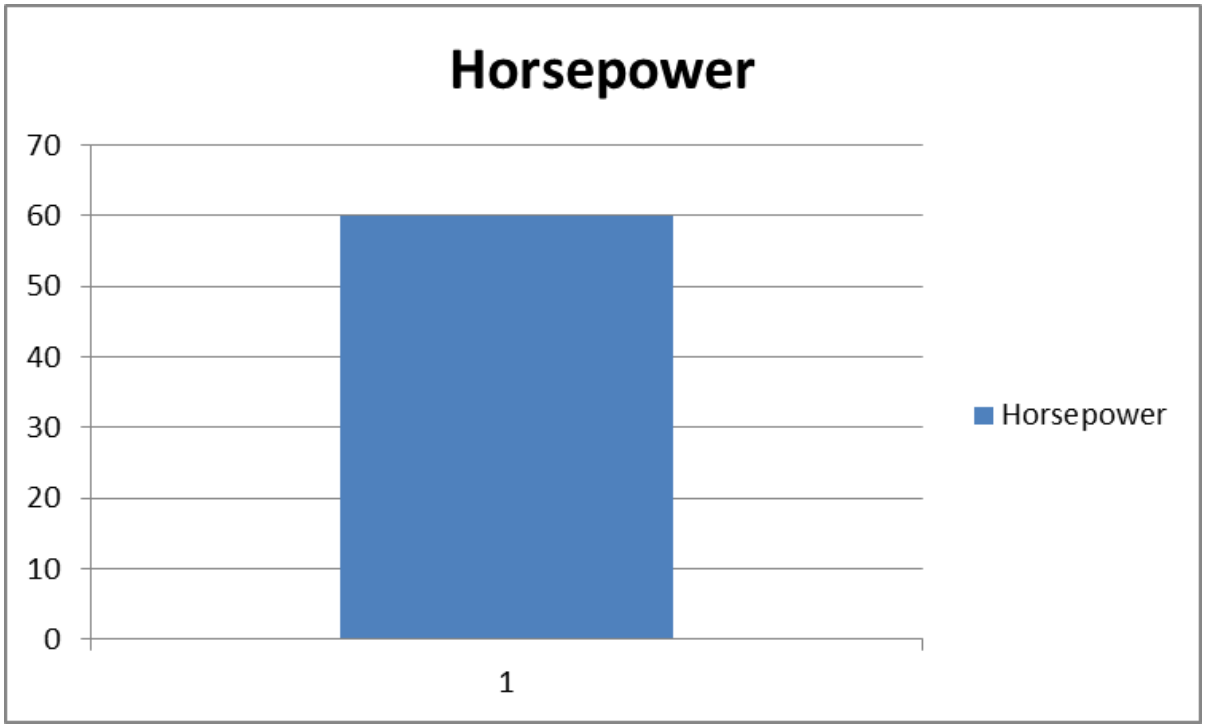
Stephen Few (chapter 4):

- compare
- sort
- add variables
- re-scale
- re-express
- filter
- highlight
- annotate
- bookmark
- aggregate
- re-visualize
- zoom and pan
- details on demand

Example

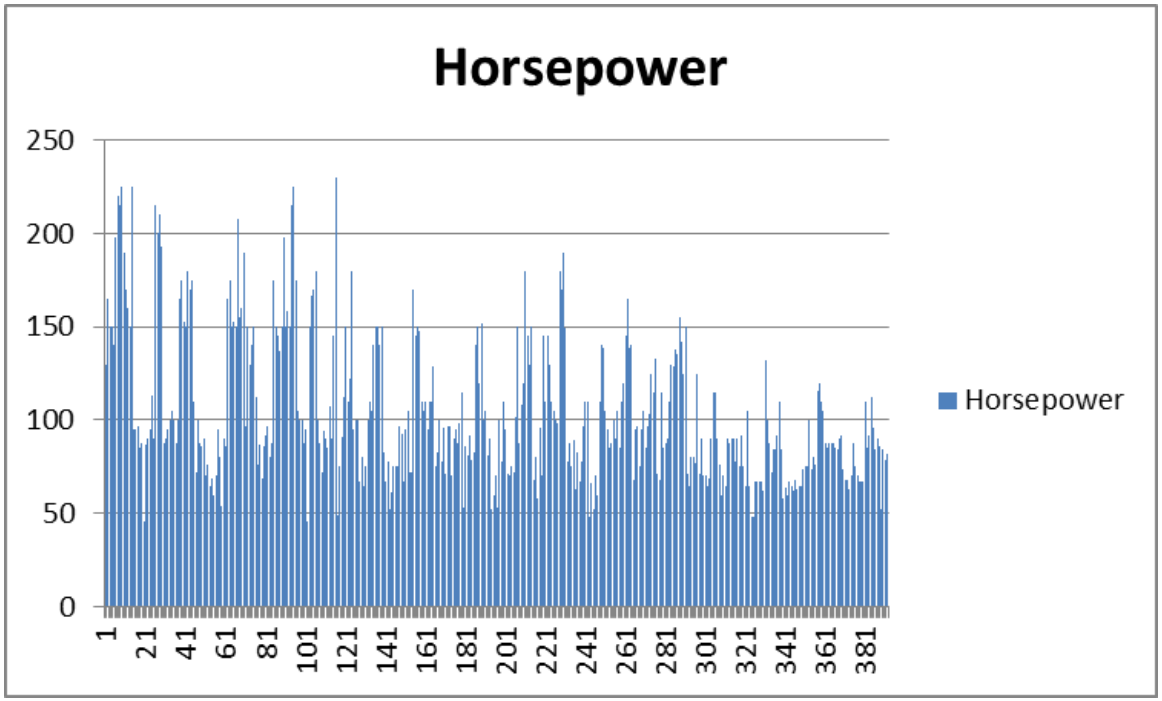
Assume you have been offered a car to buy

- assume you are mostly interested in horsepower, weight, acceleration
- the car you have been offered has 60 hp, 1834 kg, 8 s



Compare

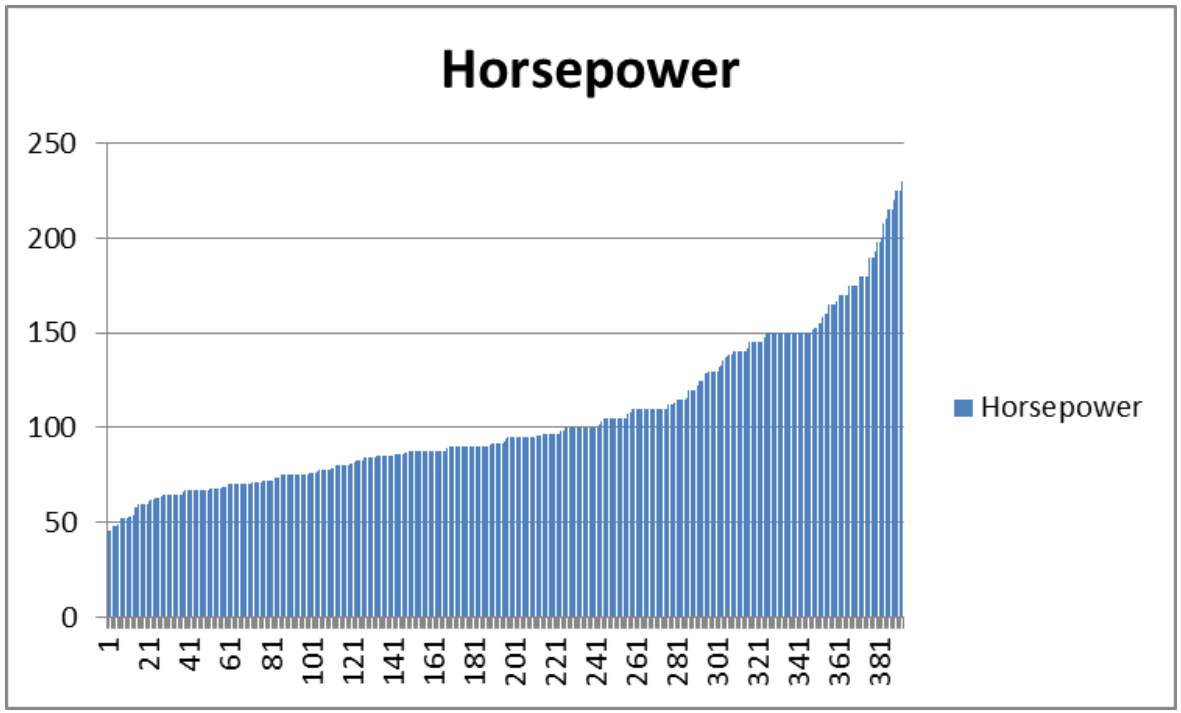
See the car with other available cars



hard to see how it really ranks

Sort

See the car in the context of other available cars

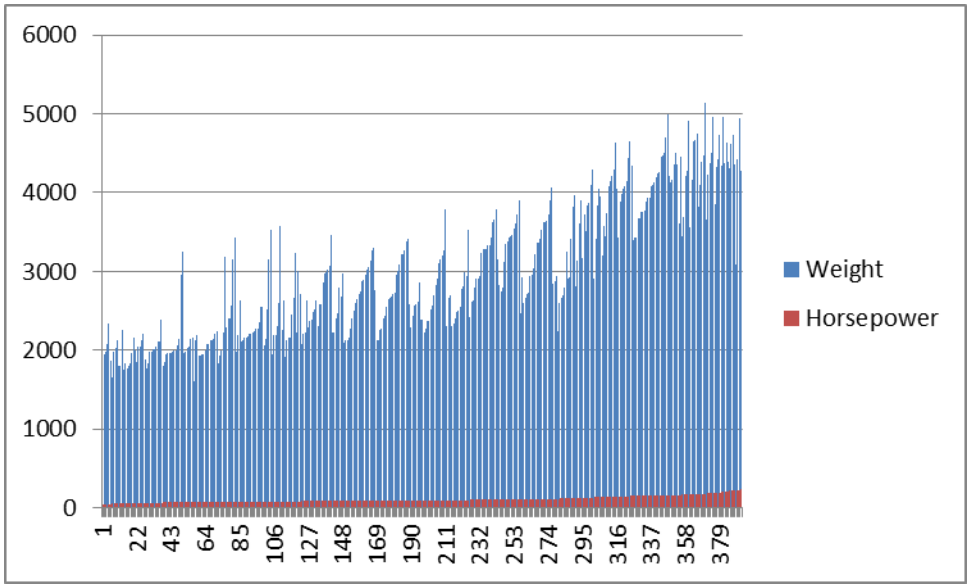


it is a low-horsepower car

Additional Variables

Is horsepower correlated to weight?

- are there trade-offs?

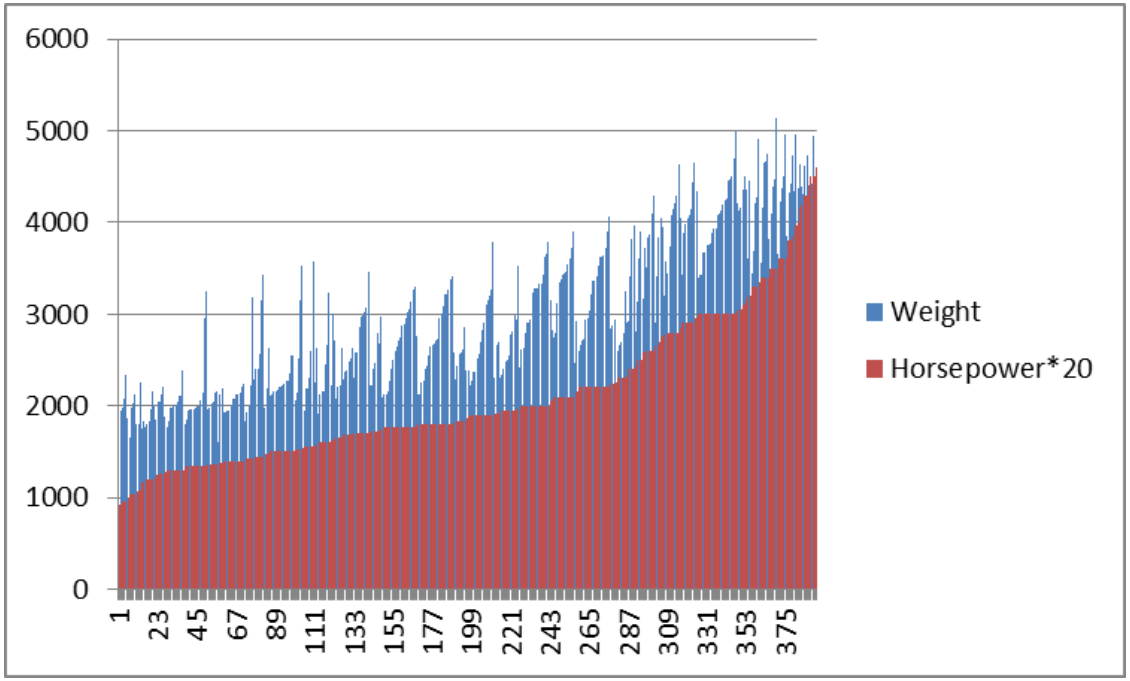


hard to see what is going on

Re-Scale

Scale horsepower into the same range than weight

- could also normalize each to (0.0, 1.0)

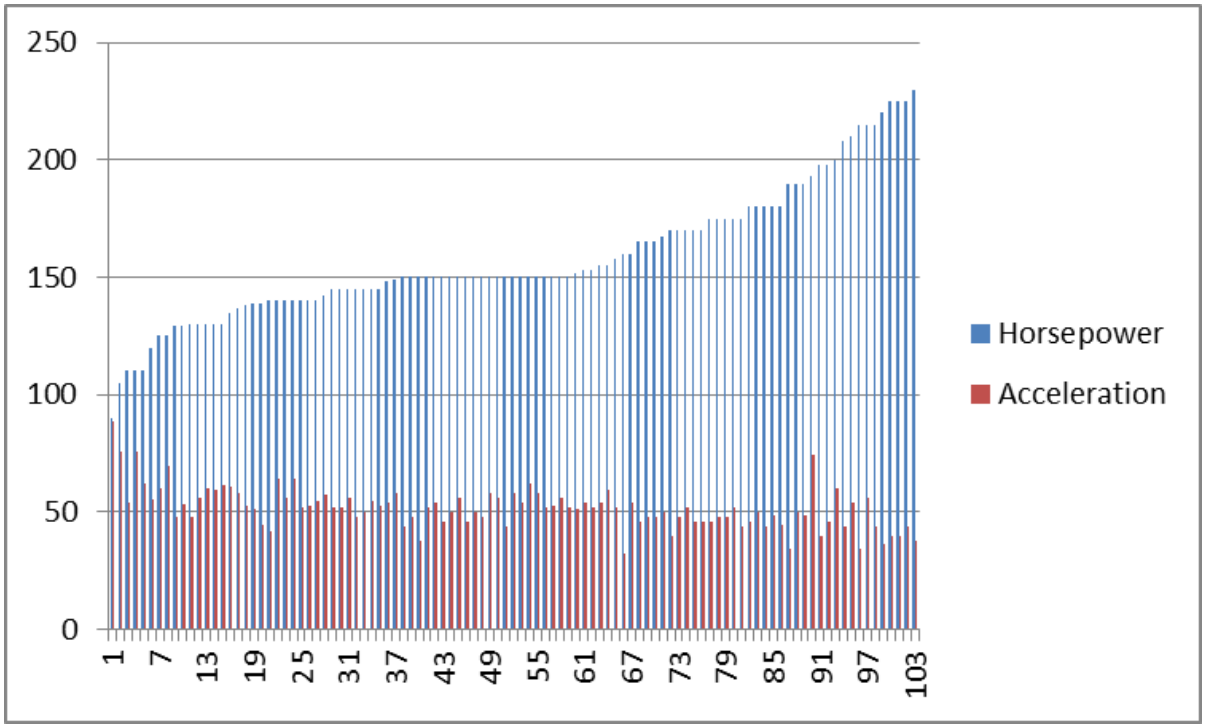


There seems to be a positive correlation

- cars with higher horsepower are also heavier

Another Variable

How does it relate to acceleration?

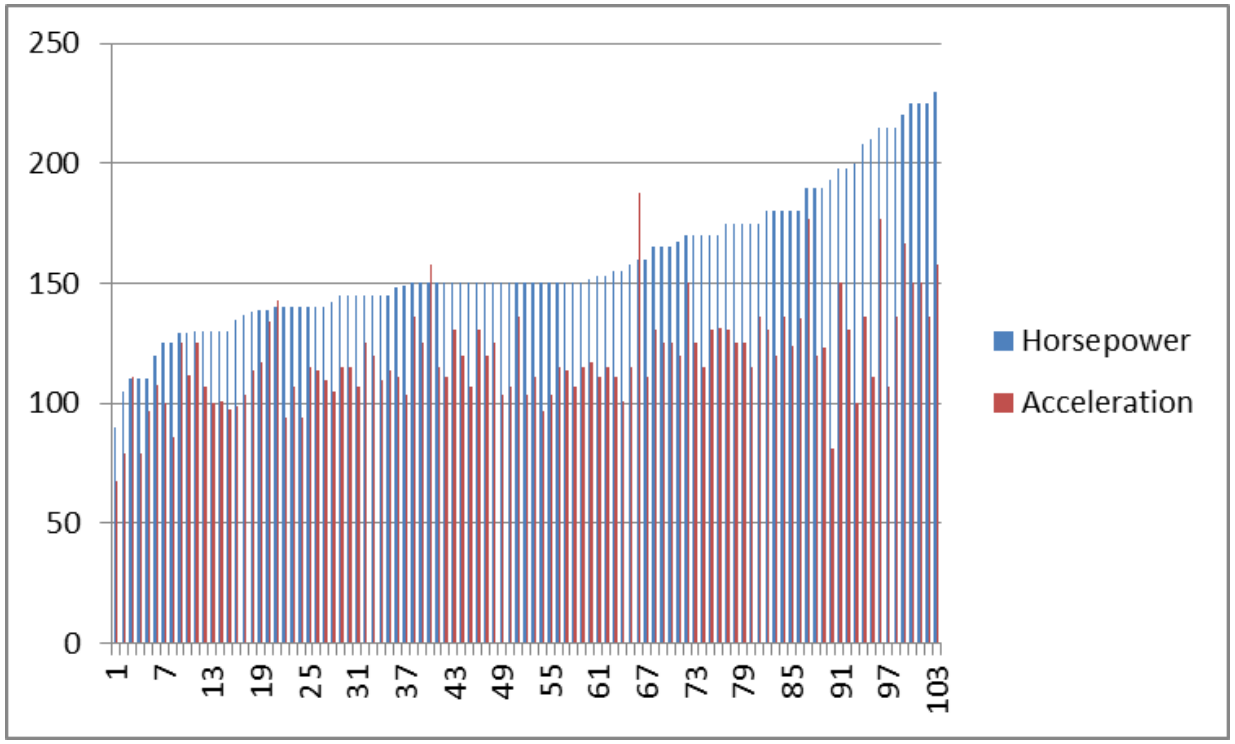


non-intuitive that acceleration is less for high horsepower cars

Re-Express

Acceleration should really be 1/acceleration

- should be measured in 1/sec (and not sec)



- now higher horsepower cars also seem to have higher acceleration (but the influence is quite minor) → is there a higher-D relationship?

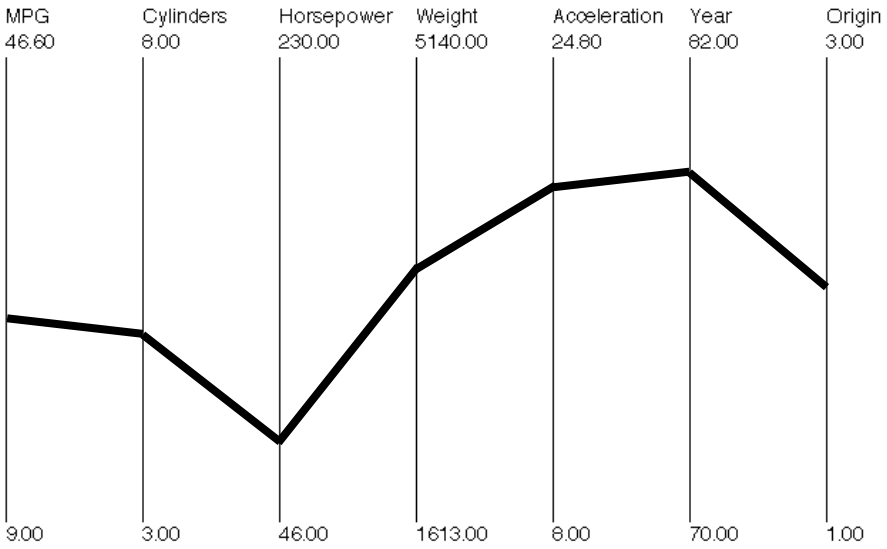
Example Dataset

Let’s look at the entire dataset

392 cars with 7 attributes:

- MPG
- Cylinders
- Horsepower
- Weight
- Acceleration
- Year
- Origin

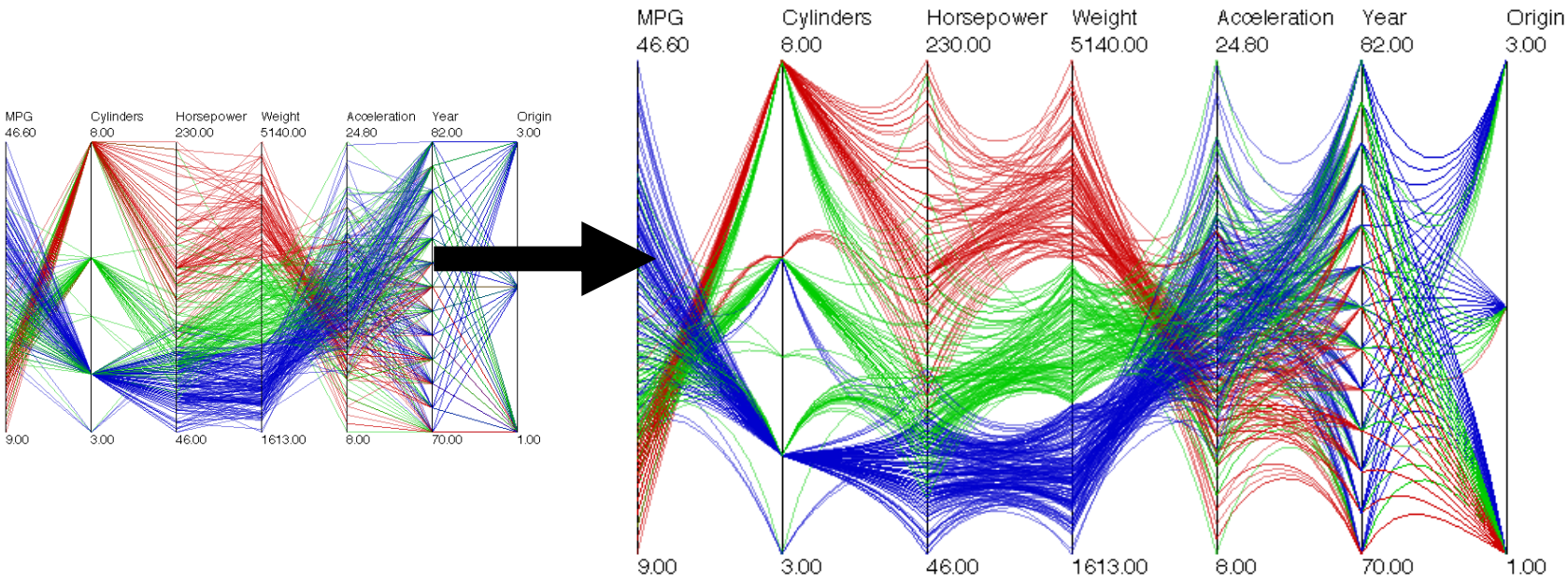
Show All Attributes: Parallel Coordinates



a car as a 7-dimensional data point

Clustering and Highlighting

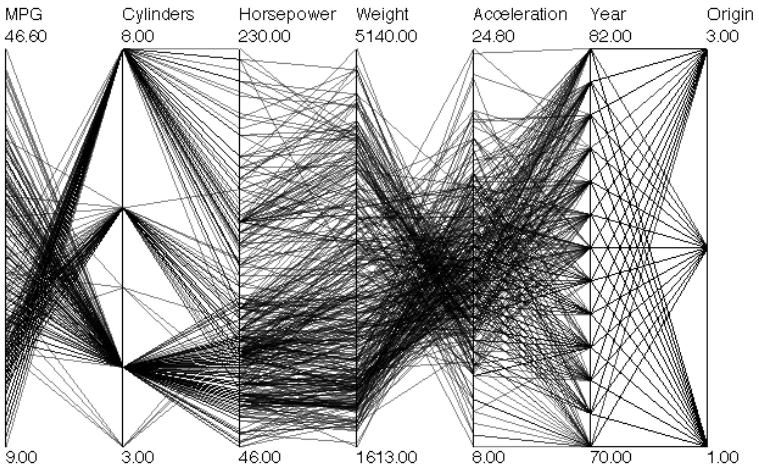
Cluster the data and highlight clusters with different colors



there are actually 3 distinct classes of cars, each with different (weight, horsepower, acceleration, …) patterns (signatures)

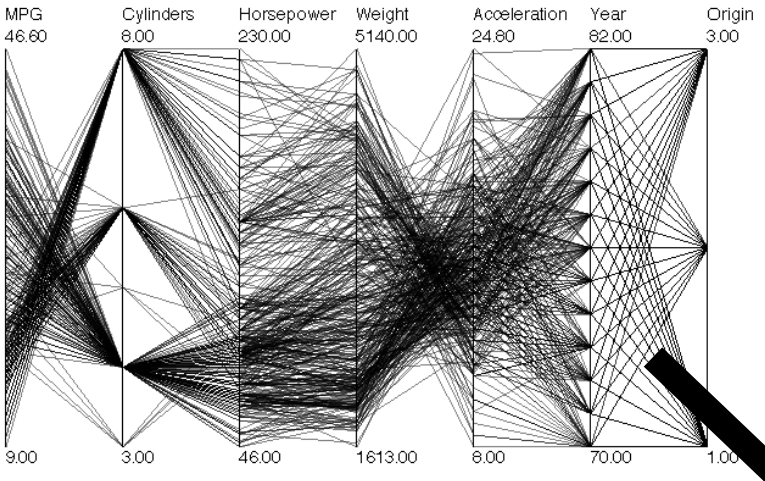
- only the high-D clusters distinguish these 3 classes
- the projection into 2D hides these relationships

Scalability Problems: Clutter



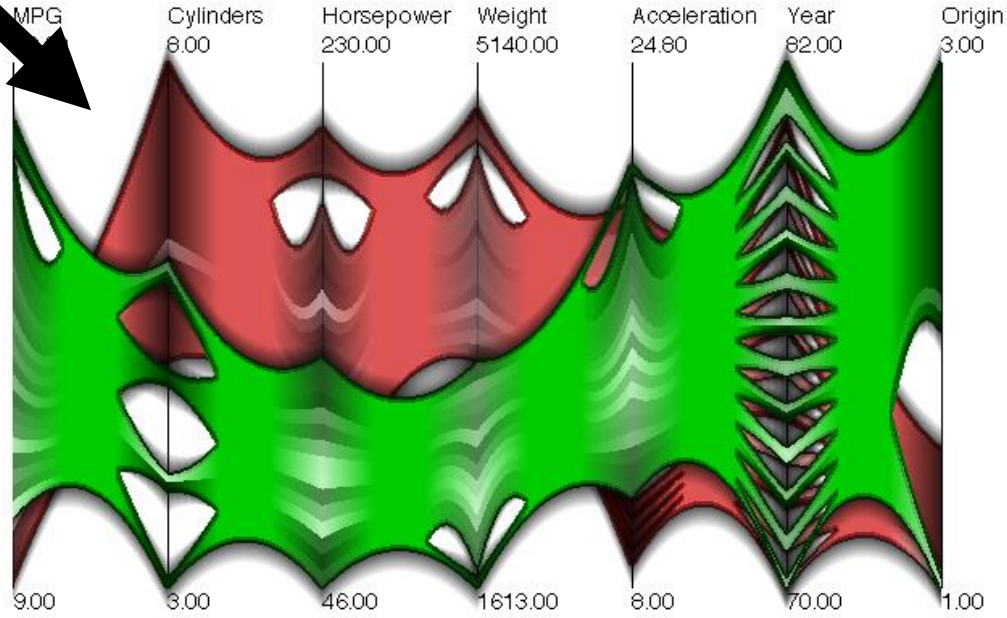
Traditional parallel coordinates plot

Clutter Reduction by Analysis-Guided Abstraction



Traditional parallel coordinates plot

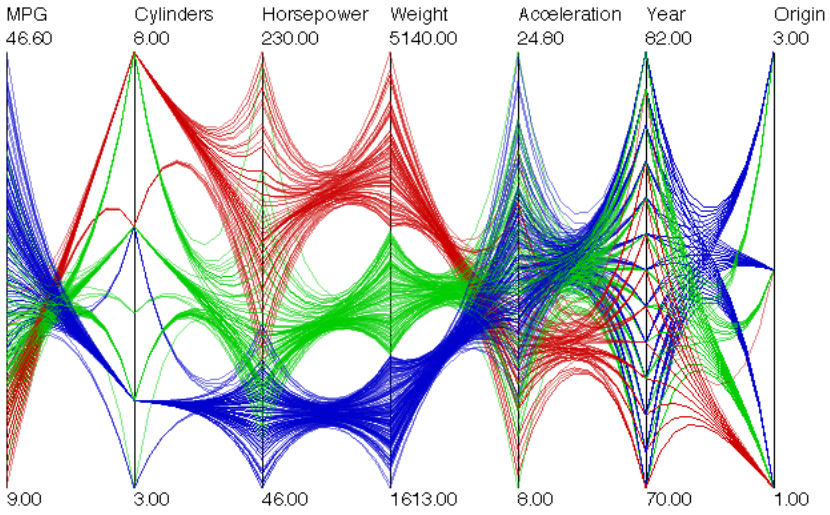
Illustrative parallel coordinates plot

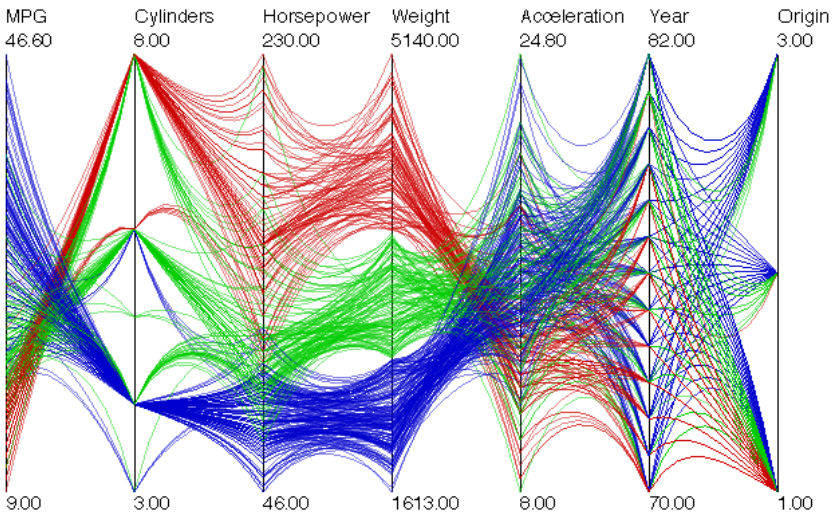


Clutter Reduction by Edge Bundling

The user can change the tension to control the amount of clutter reduction

Examples of low and medium tension, respectively:

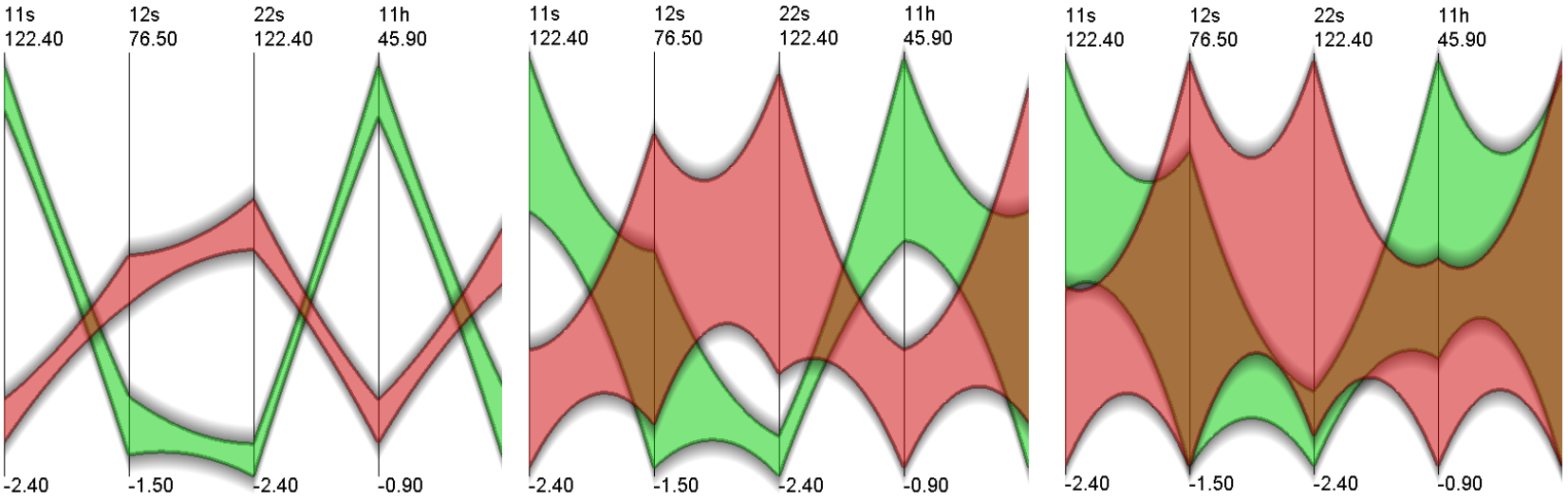




Level of Detail: Cluster Rendering

In traditional PC, clusters are often rendered as heavy line segments on top of the dataset

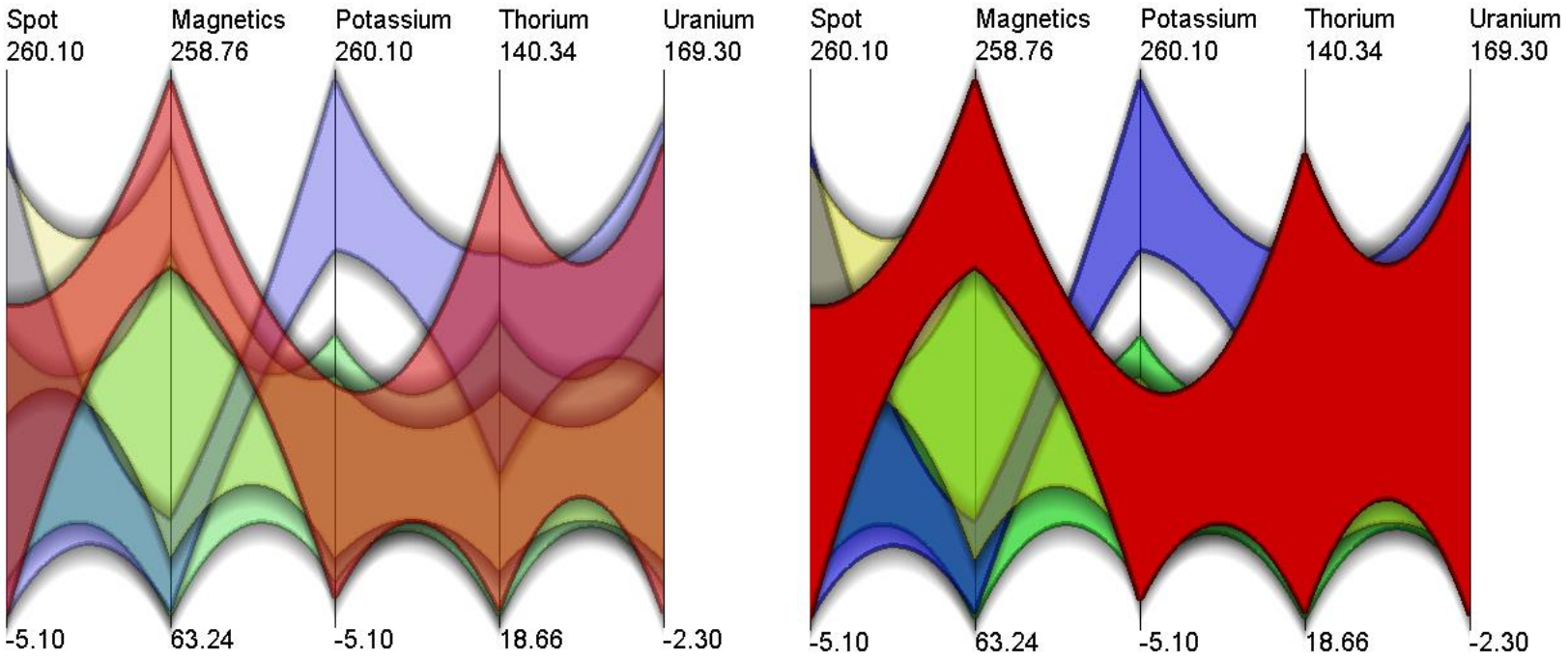
- in IPC we render the clusters as polygonal meshes
- helps to show the ranges of each cluster along axes



Highlighting Importance by Opacity Hints

Allows context to be preserved

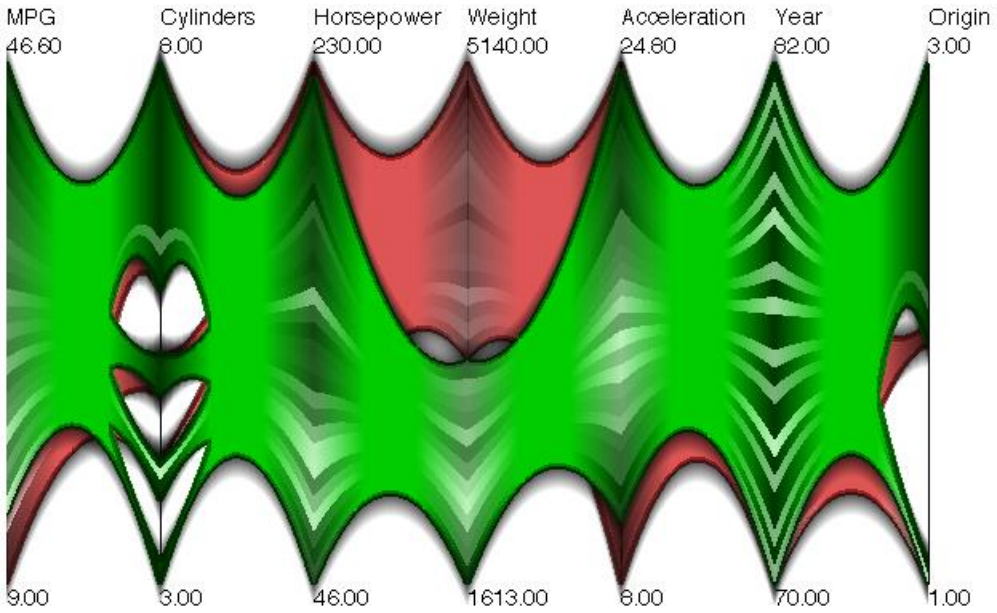
Important clusters can be made more opaque



Incorporate More Data Statistics by Per-Axis Histograms

Histograms are typically used in parallel coordinate plots to show distributions along individual axes

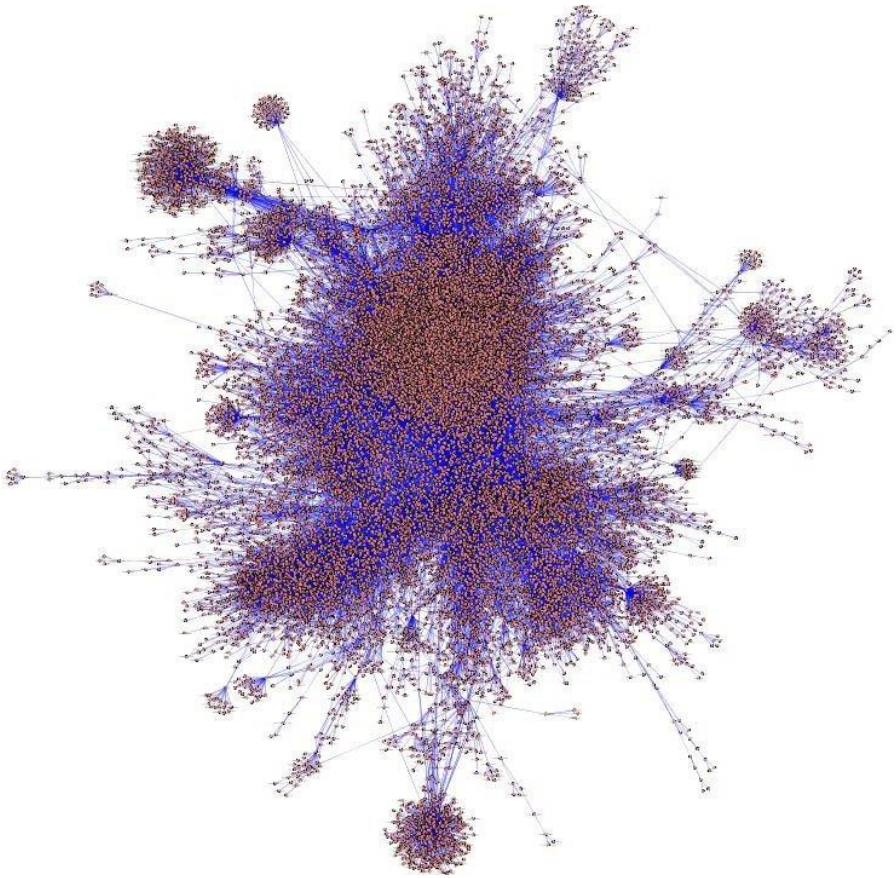
We introduce the idea of using histograms on a per-cluster basis to reveal distribution



Filtering

Example: Graph of concepts

- related concepts group closer

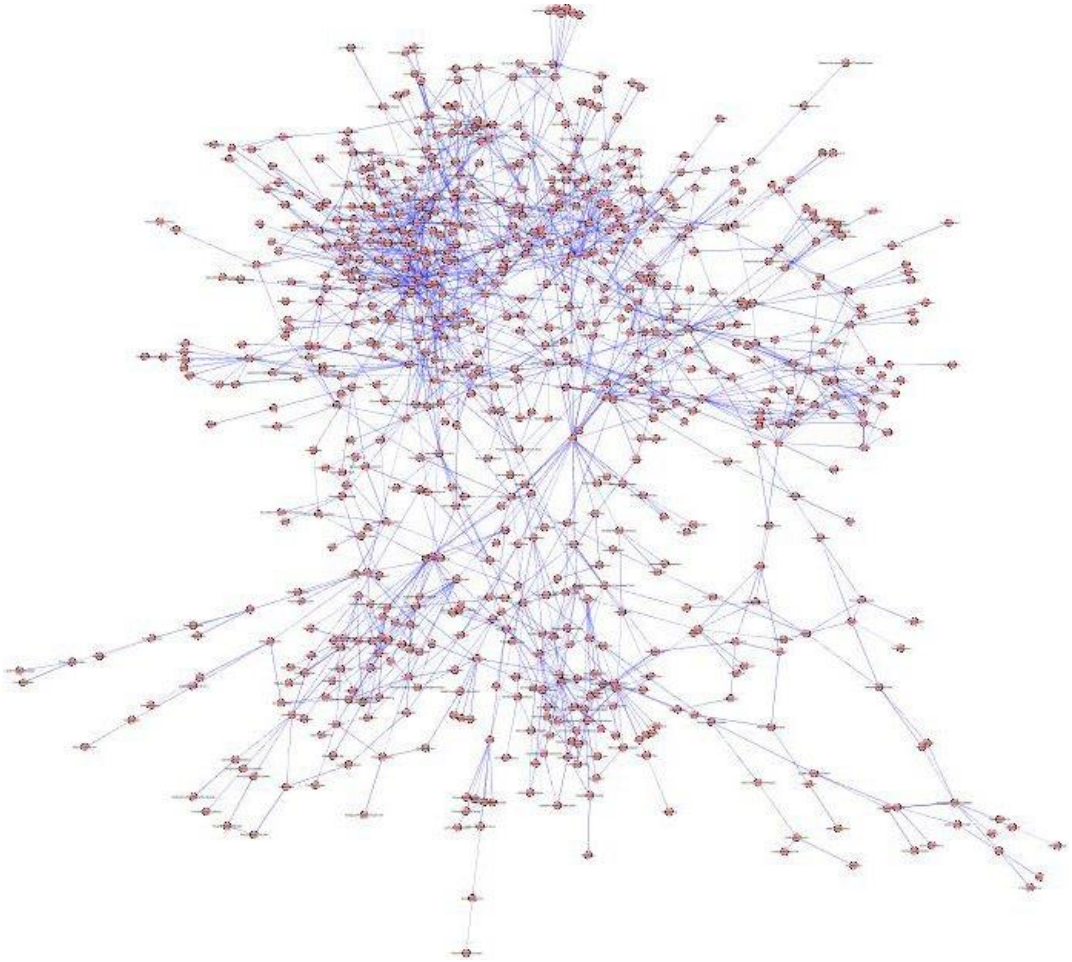


from: http://www.mkbergman.com/date/2008/02/

Filtering

Example: Graph of concepts

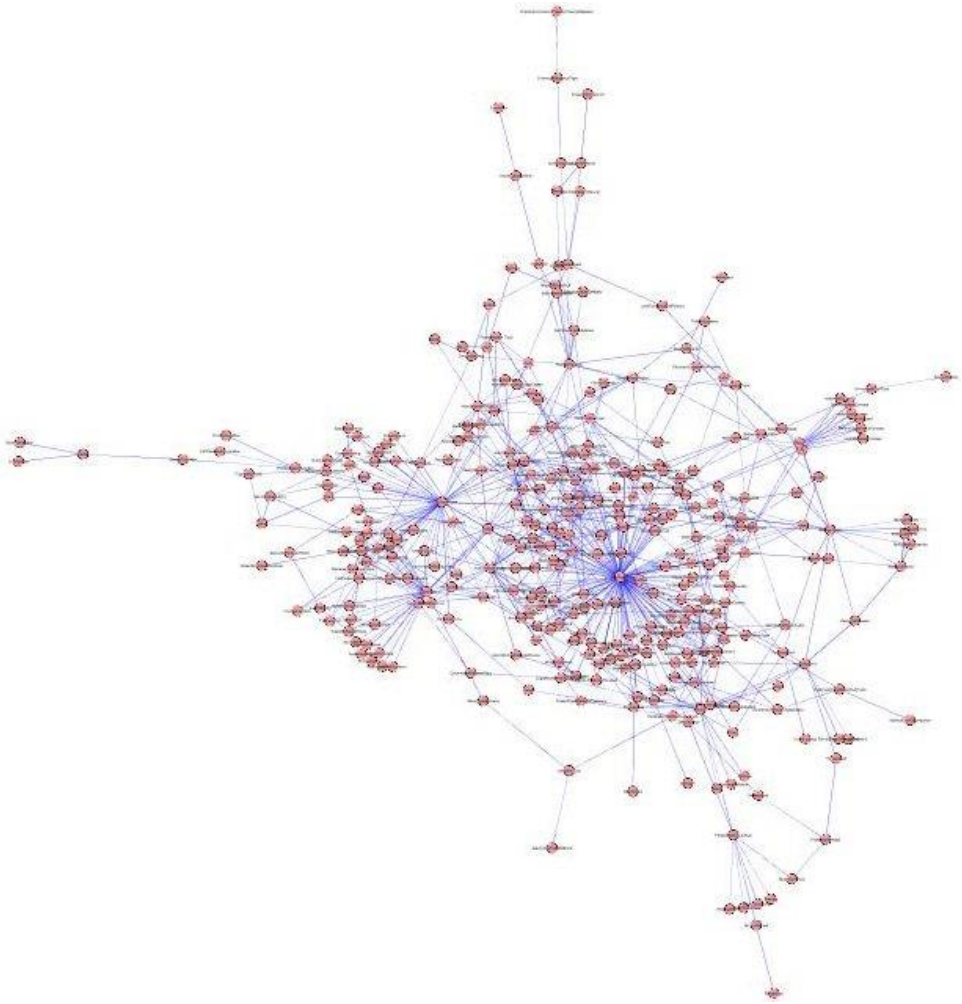
- only keep top 750 connected nodes



Filtering

Example: Graph of concepts

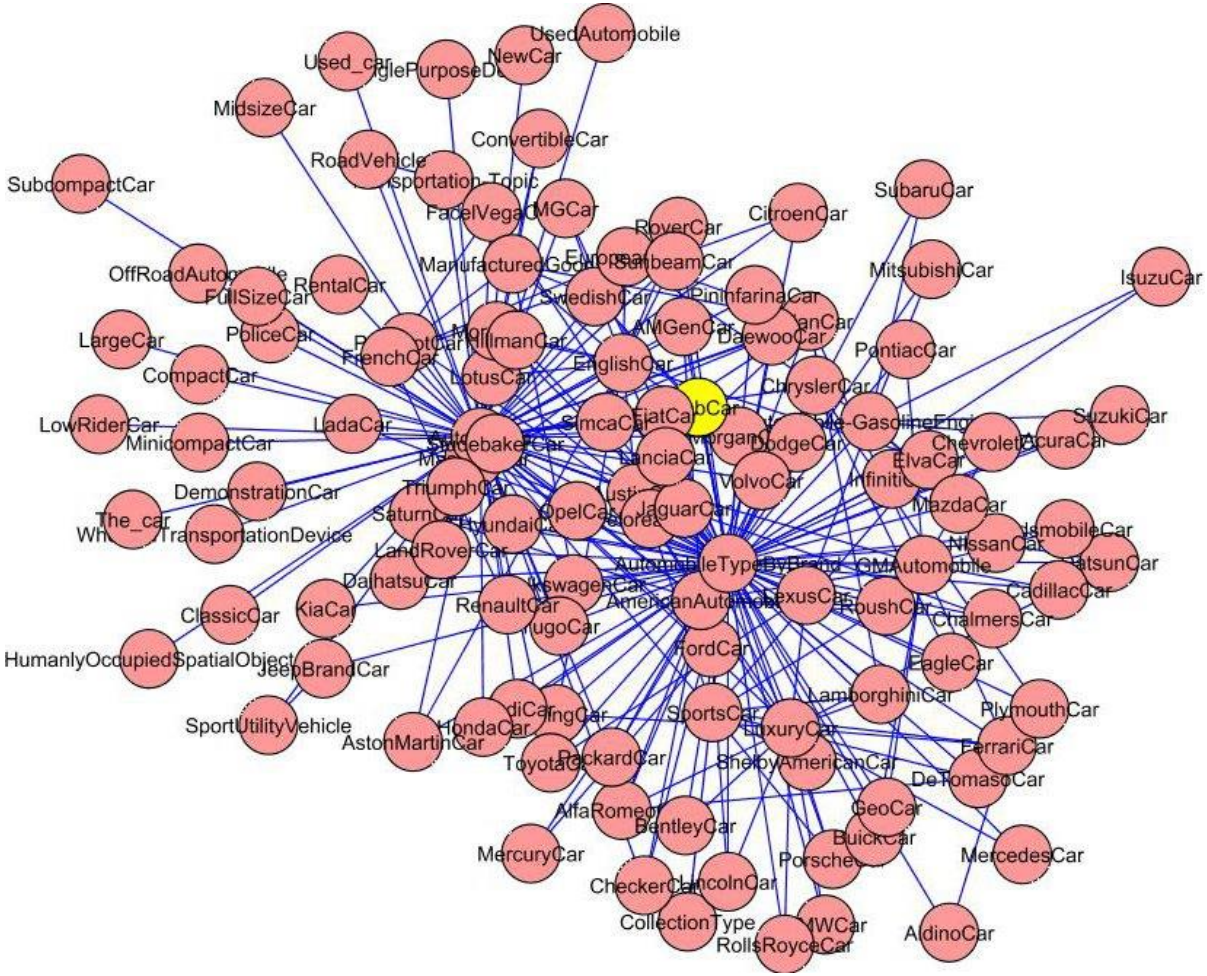
- only keep top 350 connected nodes



Zooming

Example: Graph of concepts

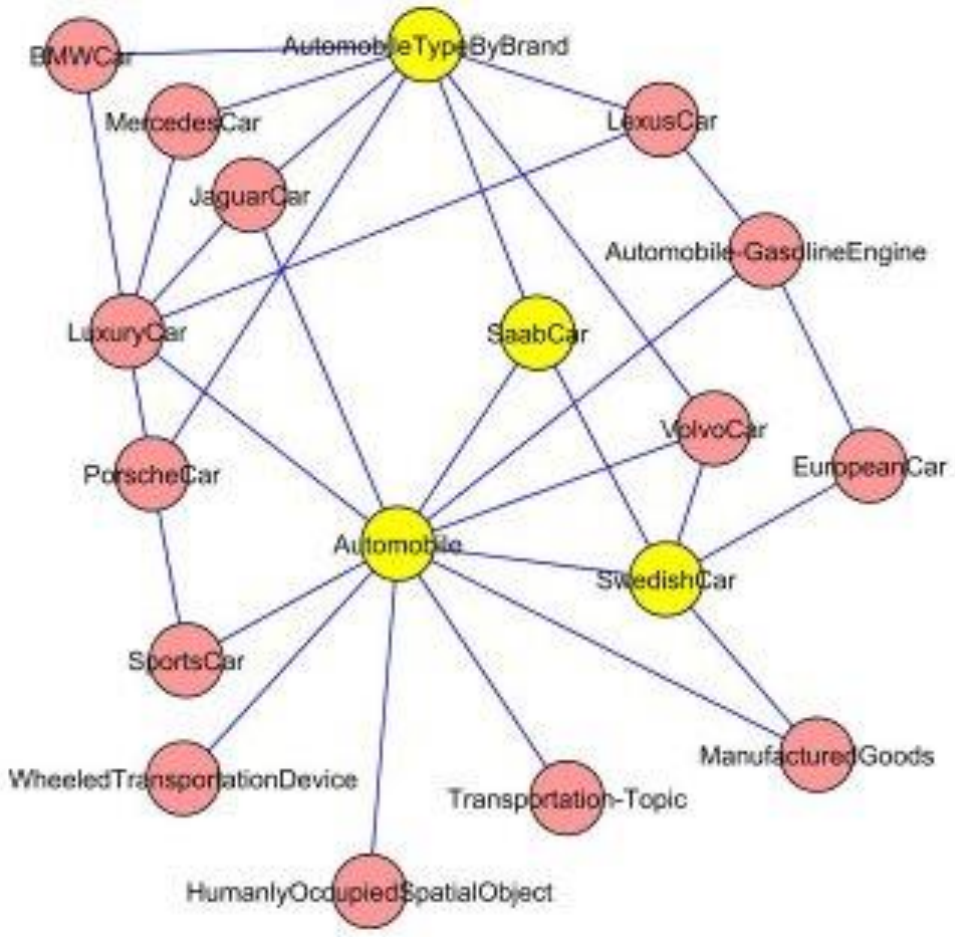
- only keep Saab neighborhood



Zooming

Example: Graph of concepts

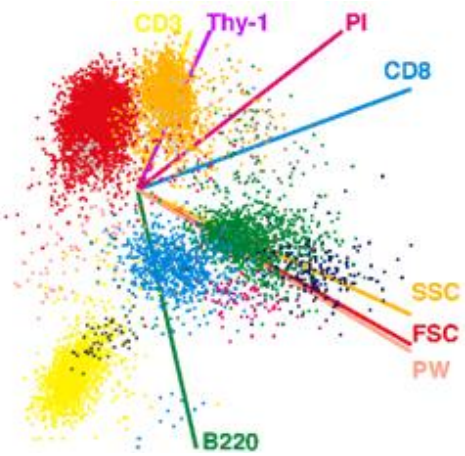
- only keep Saab neighborhood, zoom in more



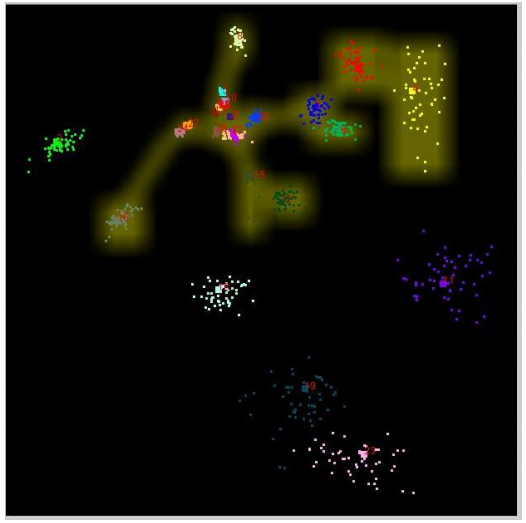
Aggregate

As discussed, good ways to aggregate all data into a single display are:

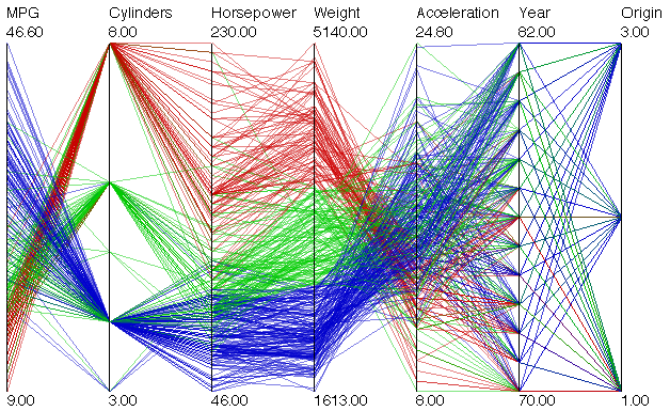
- biplots (project all data into a PCA vector basis)
- multidimensional Scaling (MDS)
- parallel coordinates



biplot



MDS



parallel coordinates

Overview and Detail

The Visual Information-Seeking Mantra

- devised 1996 by Ben Shneiderman (U Maryland, College Park)
- summarizes many visual design guidelines
- in some ways inspired by human vision/behavior
- provides an excellent framework for designing Information visualization applications.

Overview, zoom and filter, then details-on-demand

….

Overview, zoom and filter, then details-on-demand

Overview, zoom and filter, then details-on-demand

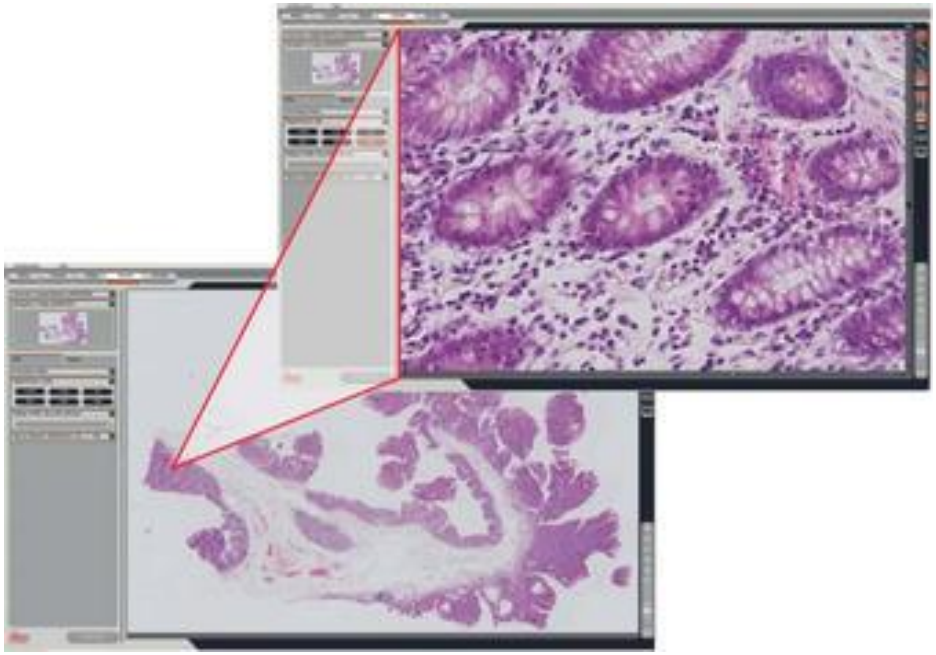
Overview, zoom and filter, then details-on-demand

Overview, zoom and filter, then details-on-demand

Overview and Detail

Information space overview plus some detail

- maintains (some) context with the detail currently focused on



Leica Microsystems

WikiViz

Focus + Context

Overview and detail

- disjoint views, maybe connected by a fan
- but: they simultaneously shows both overview and details
- require the viewer to consciously shift his/her focus of attention

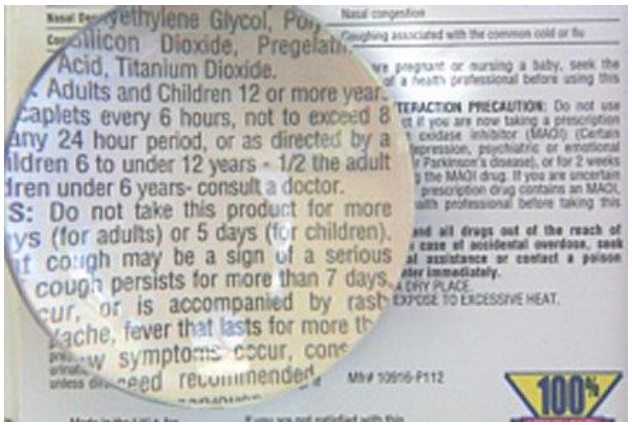
Focus + Context

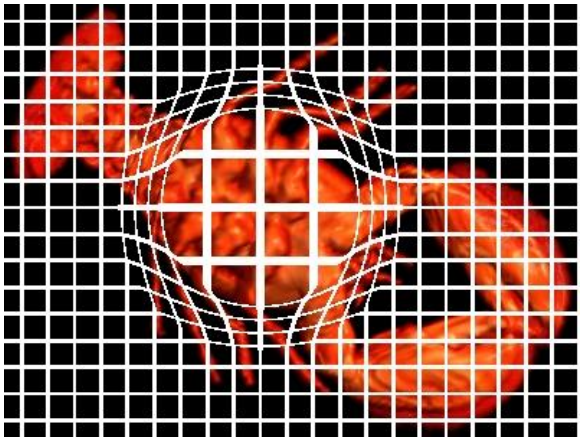
Overview and detail

- disjoint views, maybe connected by a fan
- but: they simultaneously shows both overview and details
- require the viewer to consciously shift his/her focus of attention

Focus + context

- one single view which shows information in direct context
- maintains continuity across the display
- do not require viewer to shift back and forth
- a good F+C paradigm is the *lens*
- but: there will be distortion

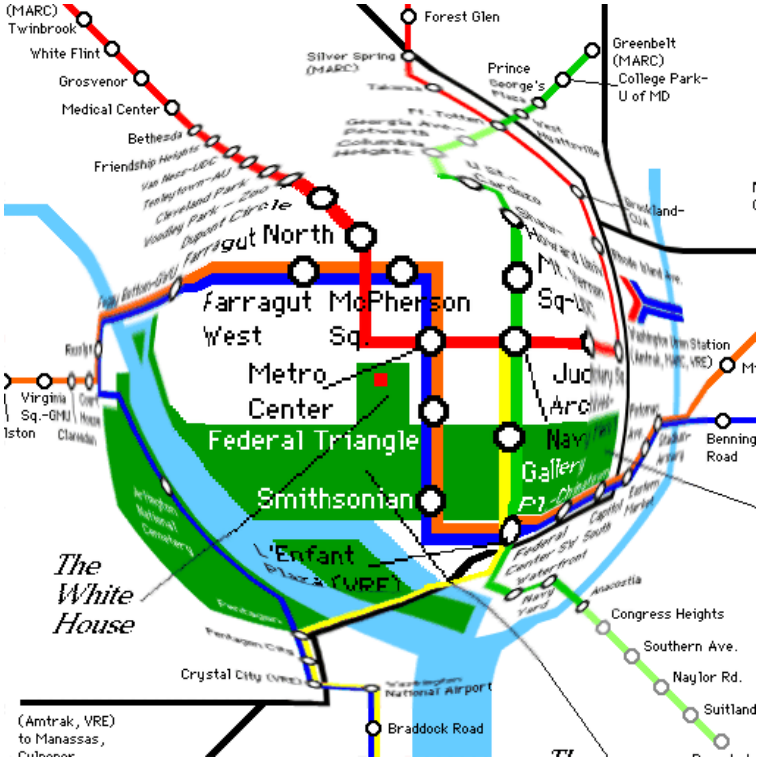




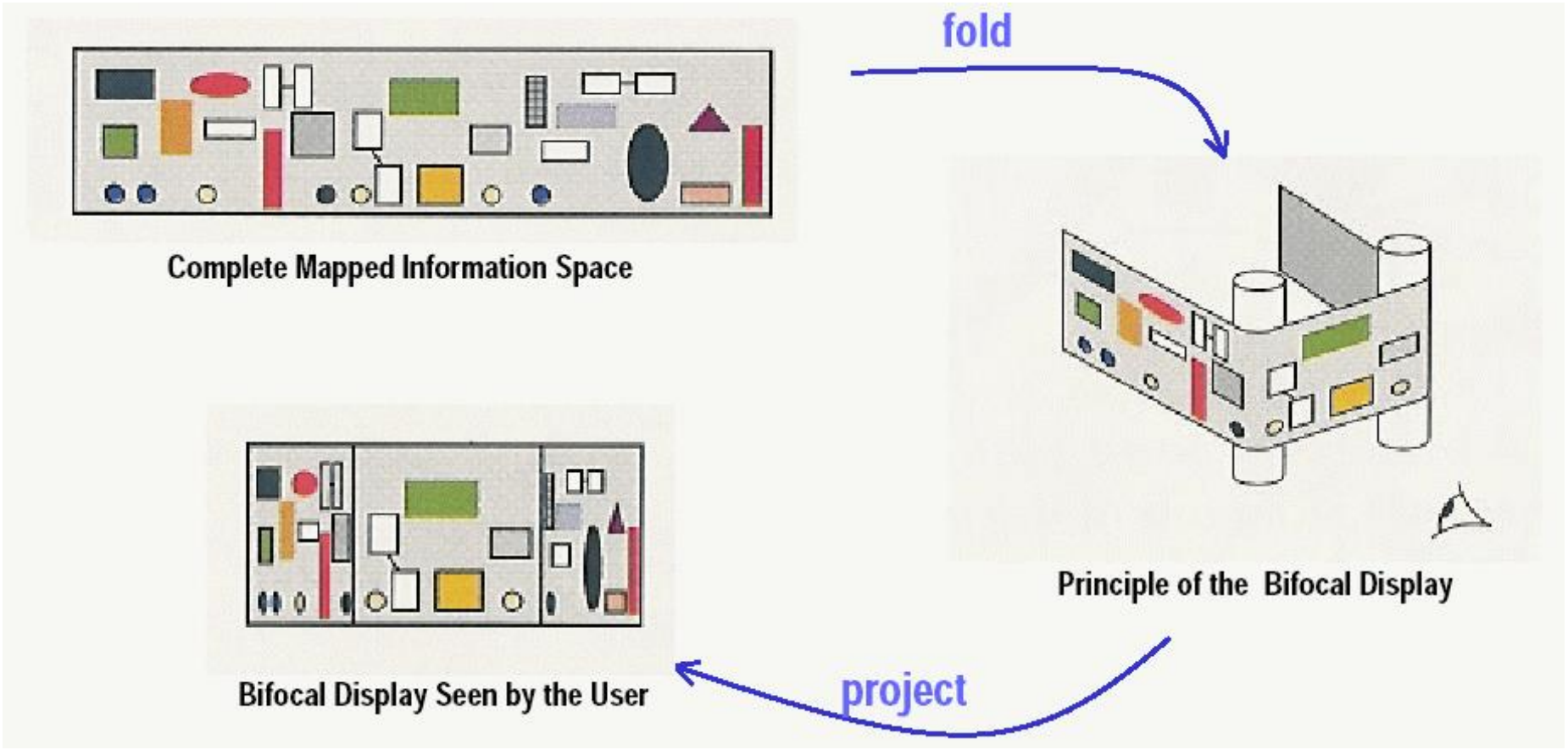
Fisheye Lenses

Fisheye lenses

- physically correct and therefore familiar
- keep target item in focus
- less relevant peripheral items are dropped or reduces in size
- distortion

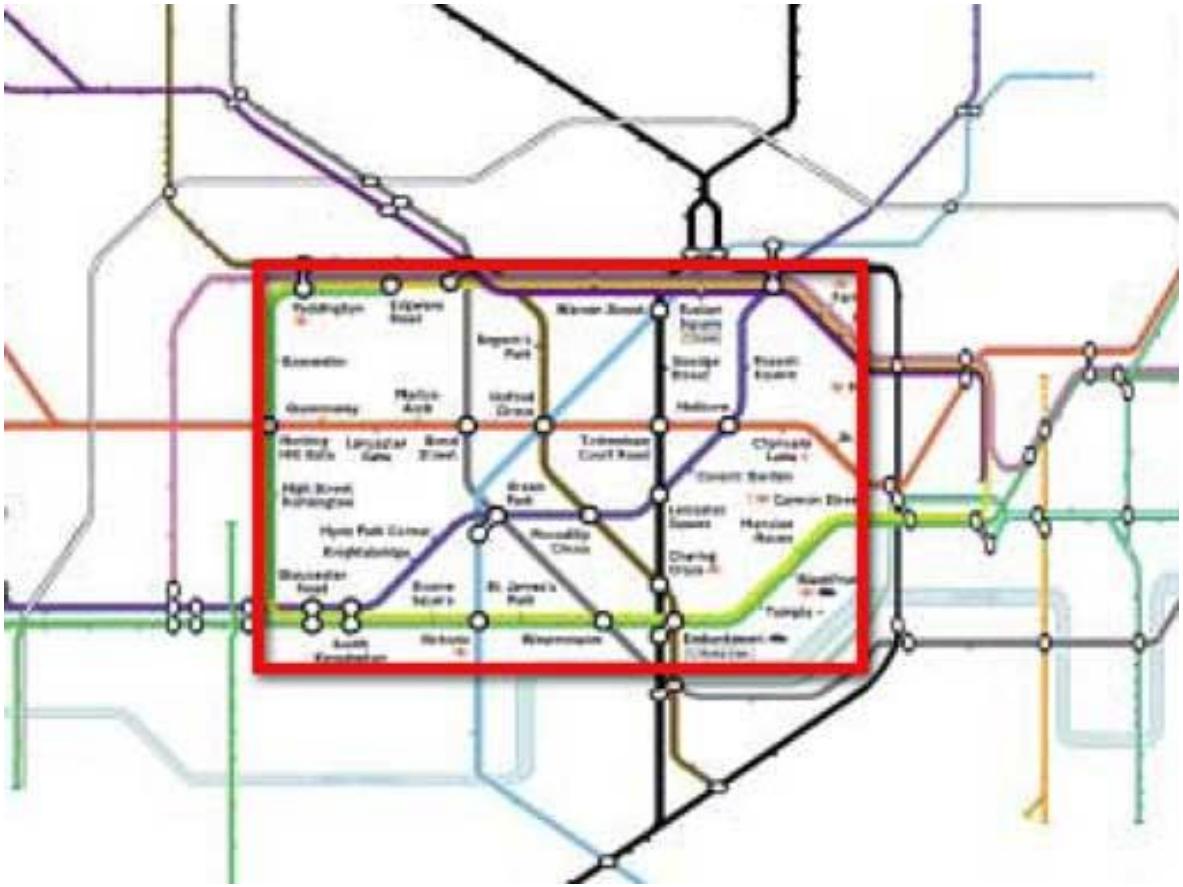


Bifocal Lens

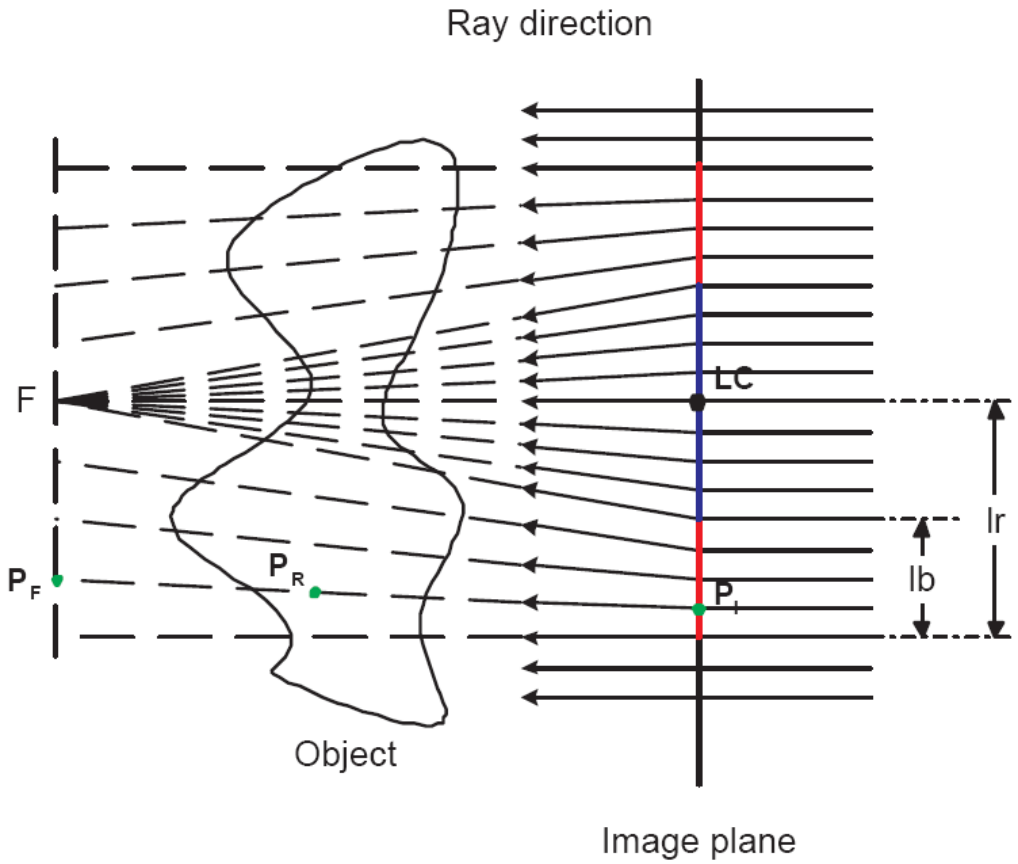


Bifocal Lens

London subway map



(Volumetric) Magnification Lenses



normal region

transition region

magnification region

transition region

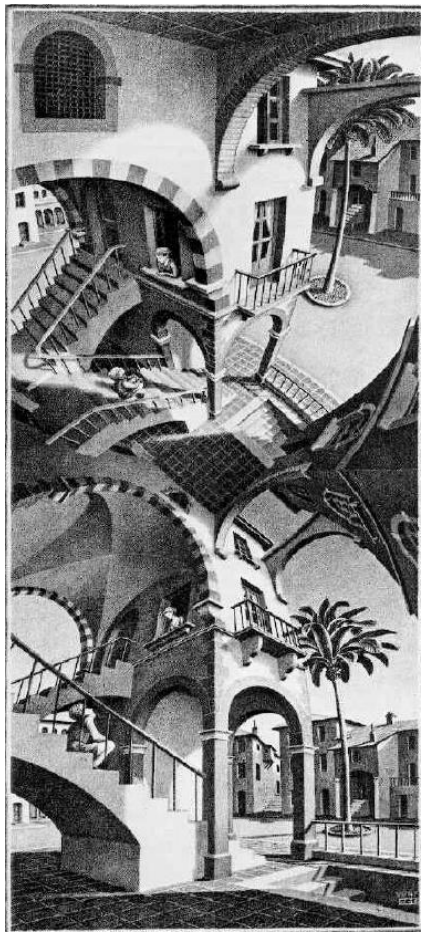
normal region

Avoid aliasing in transition regions by low-pass filtering

Generalized Lenses

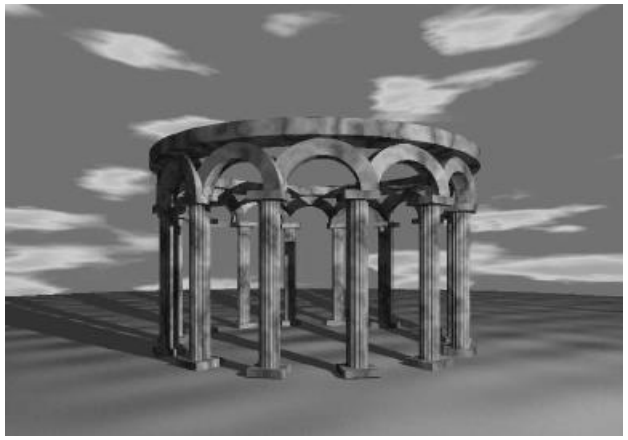
Computers can go beyond (stretch) the laws of physics

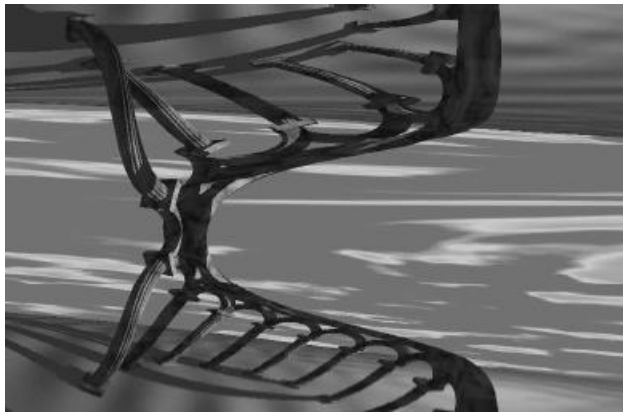
- example: multi-perspective lens rendering, gaze-directed, …



MC Escher

Rademacher/Bishop

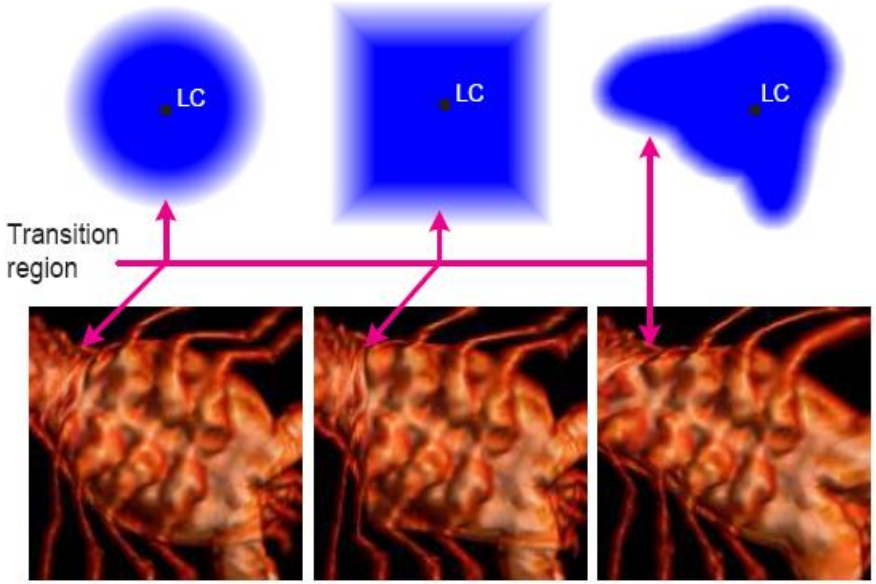




Loeffelmann/Groeller

Generalized Lenses

no lens

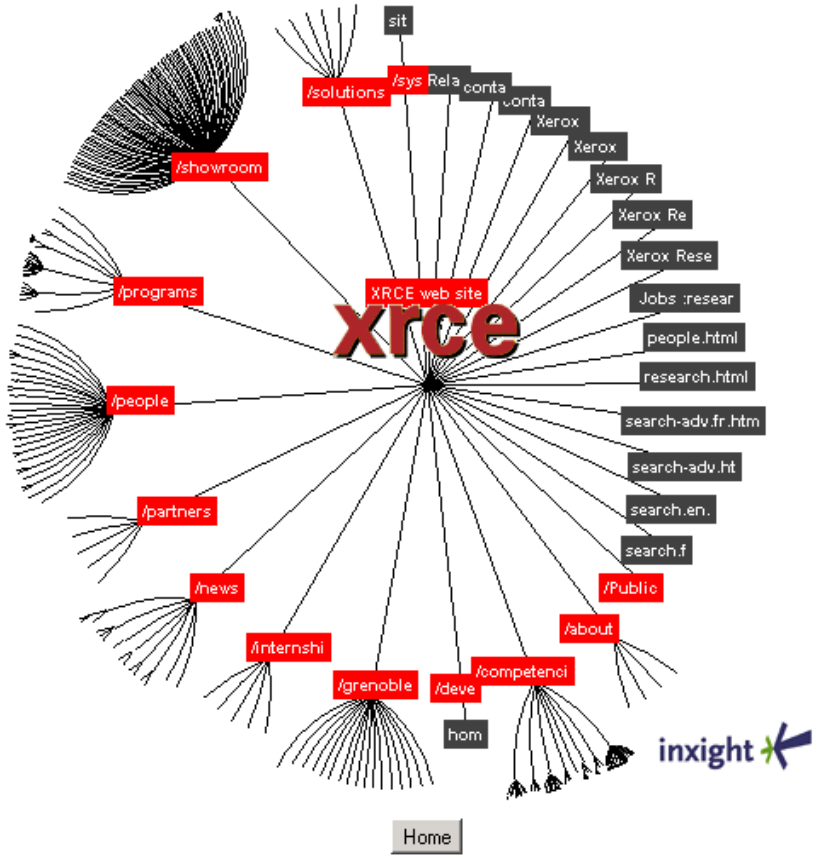


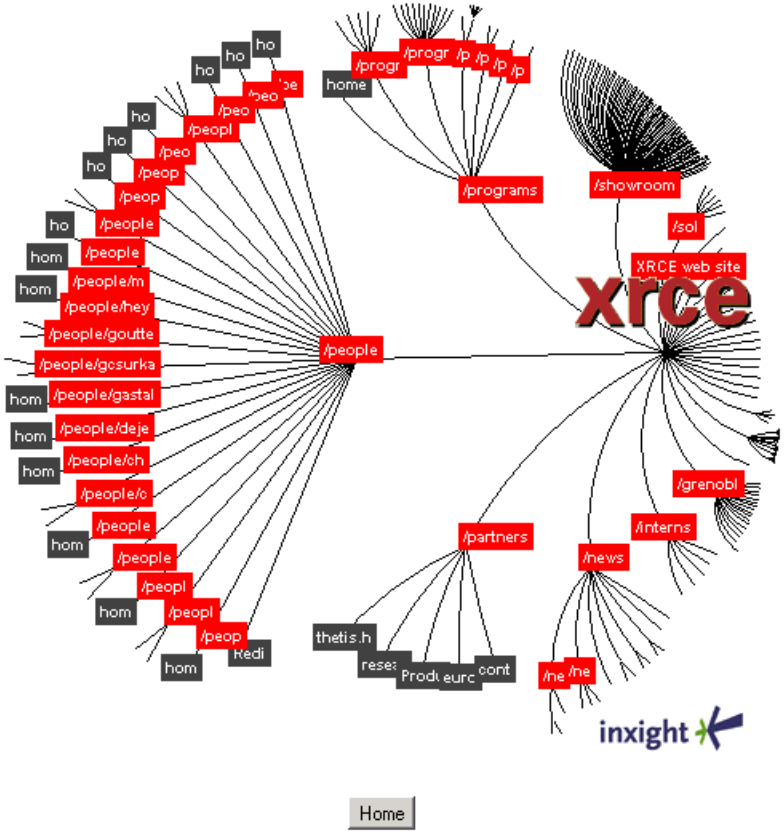
Wang et al., 2005

Lenses in Information Visualization

Hyperbolic Tree fisheye lens

- Xerox PARC/Inxight
- selectively and smoothly reduce complexity as user focus changes

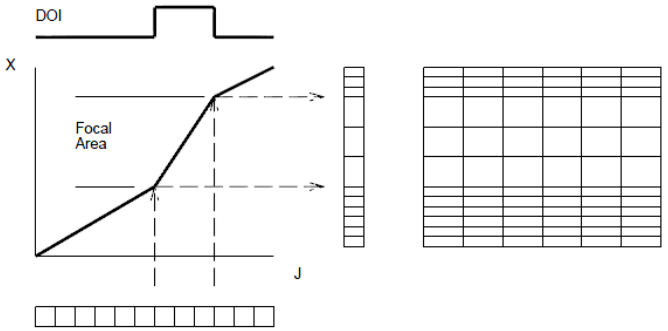


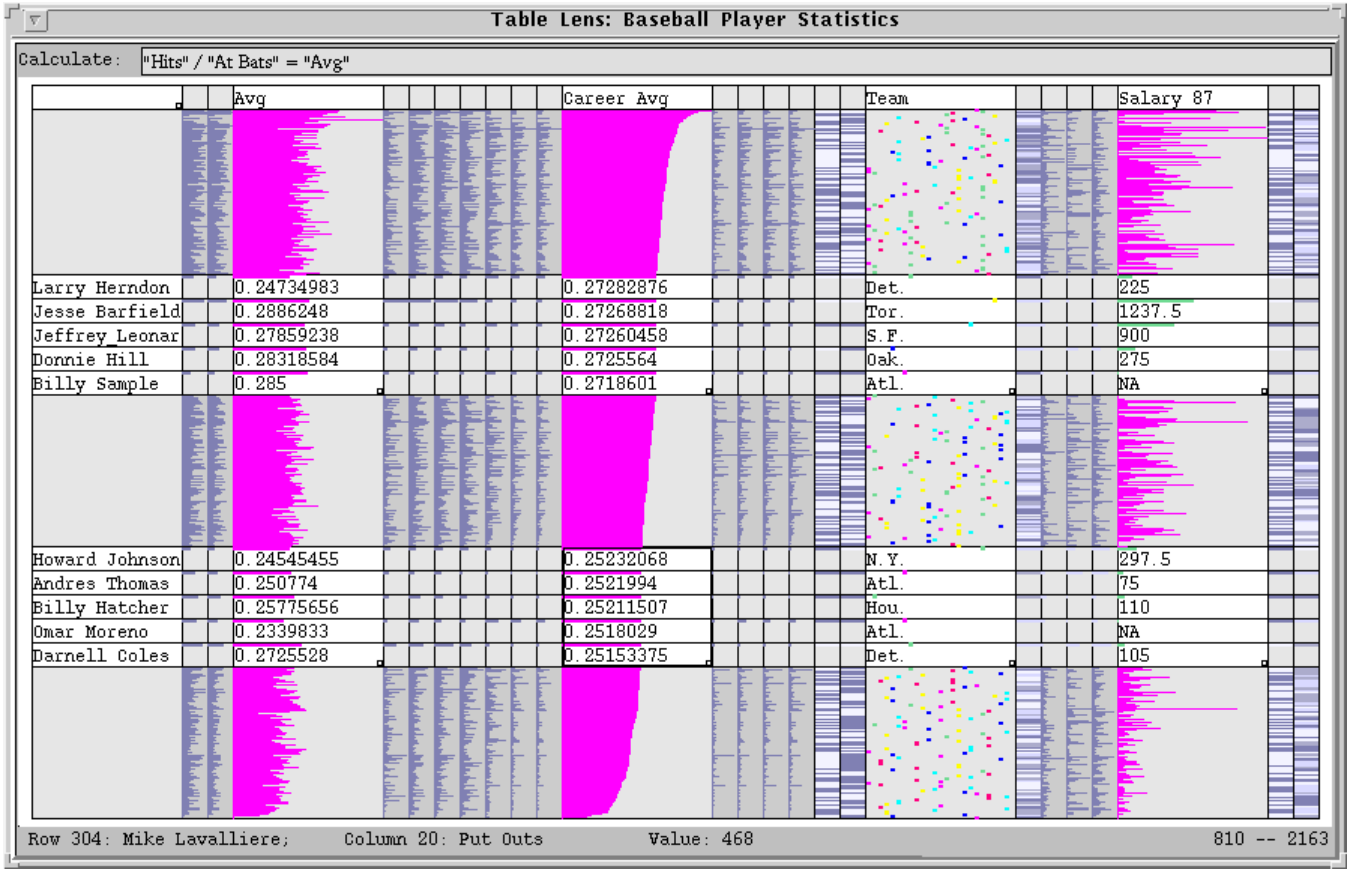


Lenses in Information Visualization

Table Lens (Rao and Card, 1994)

- uses a DOI (degree of interest) lens
- fuses symbolic and graphical detail driven by the DOI lens

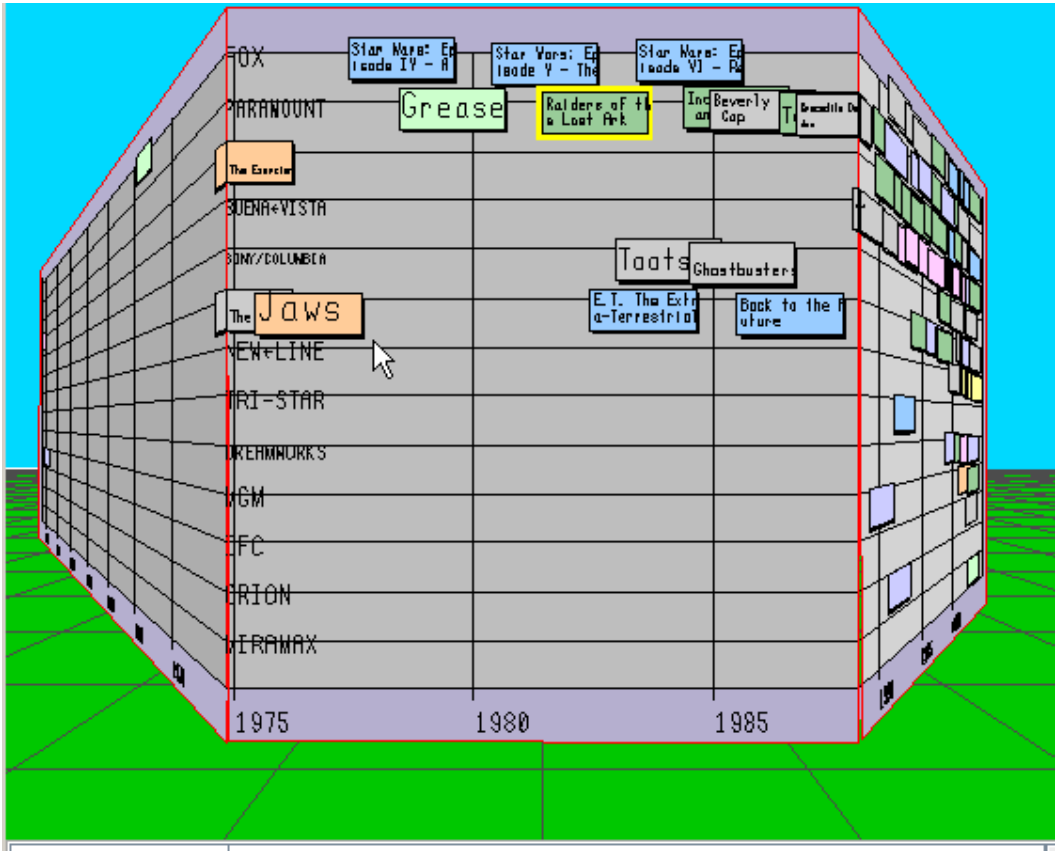




Lenses in Information Visualization

Perspective Wall

- Xerox PARC/Inxight
- details on the center panel are at least three times larger than the details on a flat wall that fits the field of view



Zoom and Pan

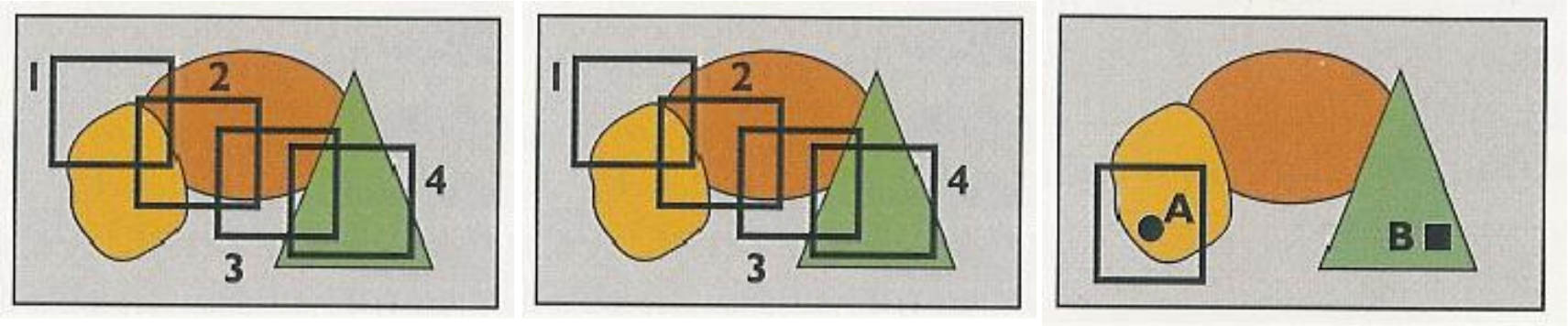
Panning

- smooth movement of a viewing frame over a 2D image of greater size

- Zooming
  - increasing magnification of a decreasing fraction (or vice-versa) of a 2D image under the constraint of a viewing frame of constant size

Transfer of the focus of attention:

- zoom out → pan → zoom in

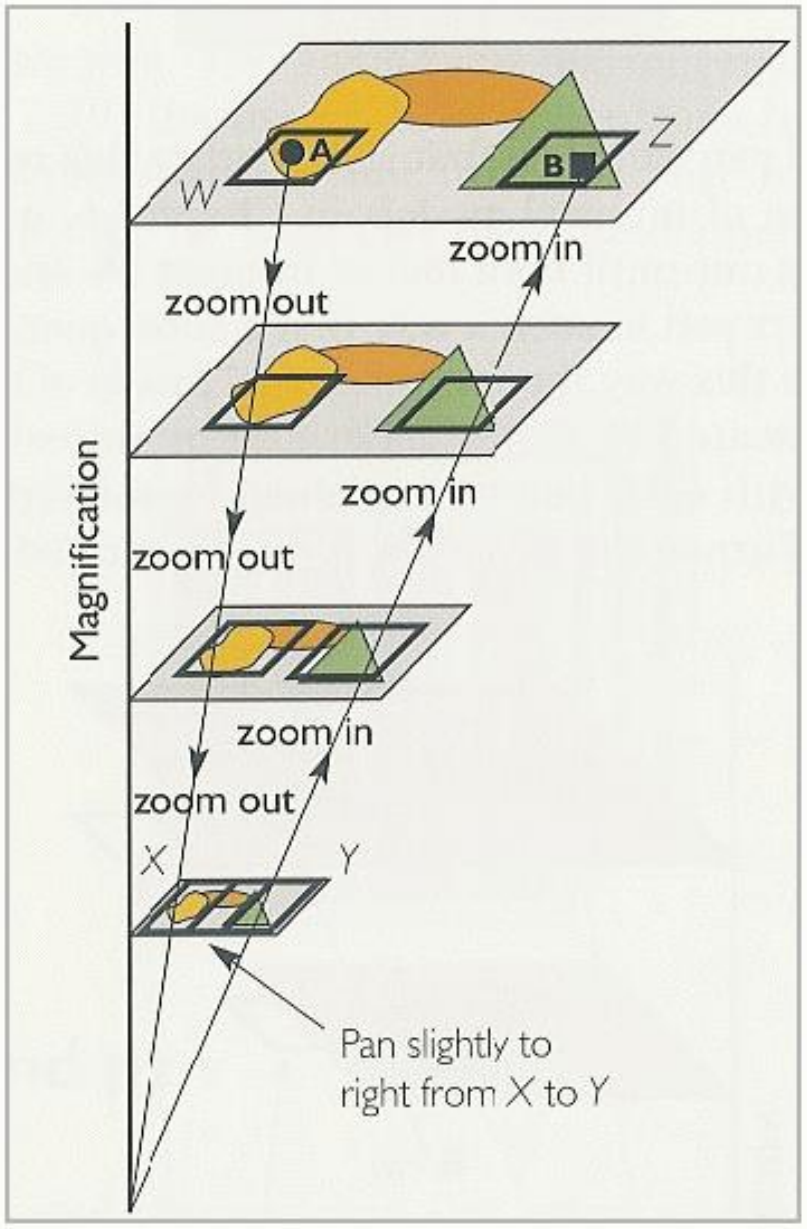


Scale-Space Diagrams

Efficient transfer of the focus of attention:

- zoom out → pan → zoom in

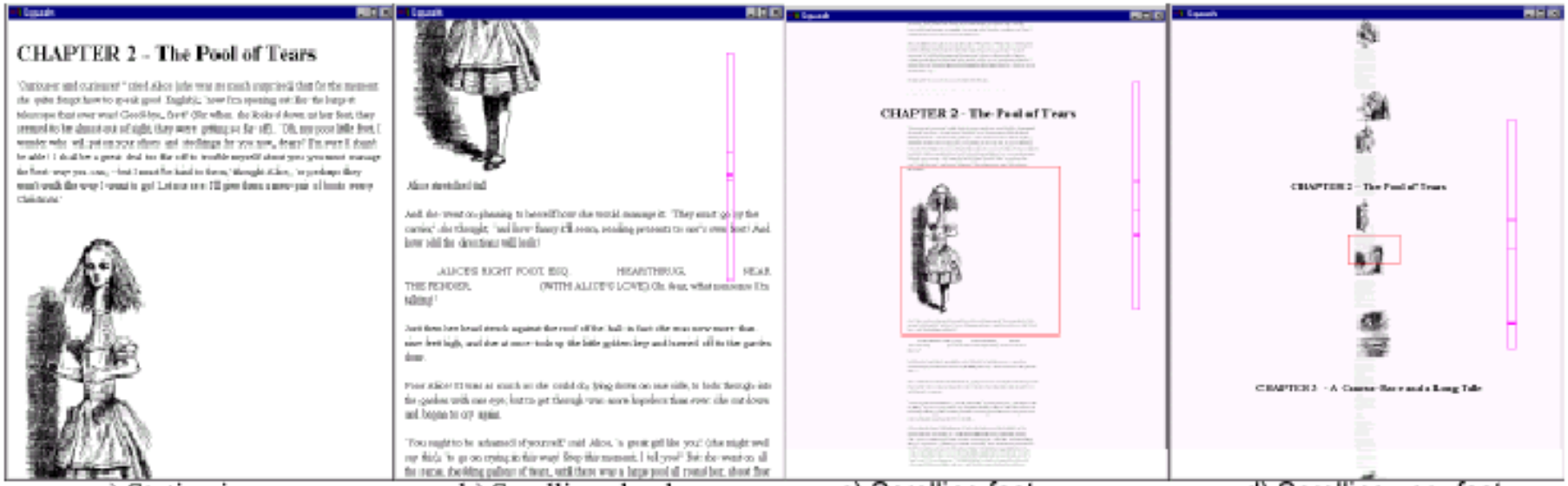
Furnas, Bederson, 1995



Intelligent Zooming

Depending on scrolling speed, zoom more or less

- allows efficient navigation of large documents
- employs semantic zooming



a) Static view b) Scrolling slowly c) Scrolling fast d) Scrolling very fast

Igarashi, Hinckley, 2000

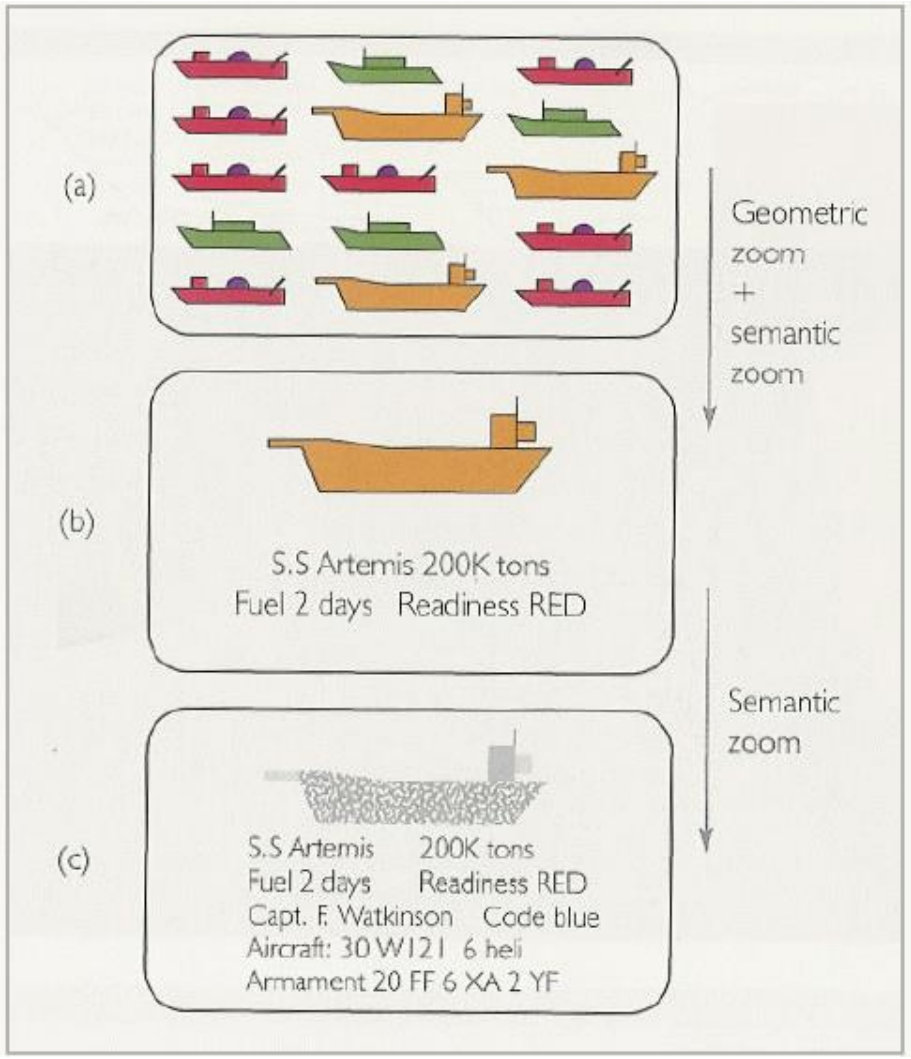
Semantic Zoom

Standard zoom:

shows a down/up scaled version of the object/image

Semantic zoom:

- shows a different representation determined by the space available



Semantic Zooms: Maps

Semantic Zooms: Information Visualization

Could show different levels/aspects of information

- on a map, show either parking lots, bars, or restaurants
- zoom in by price range (cheap first, then more expensive…)
- zoom in by preference (favorite food first, then less favorite…)
- may combine these criteria into a preference function

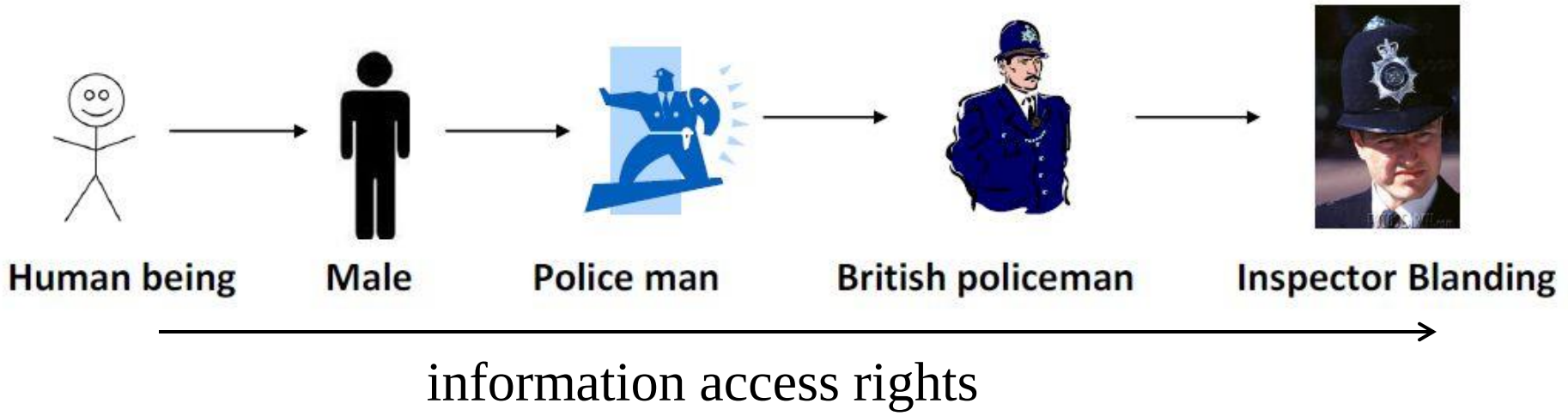
Semantic Zooms: Information Visualization

Could show different levels/aspects of information

- on a map, show either parking lots, bars, or restaurants
- zoom in by price range (cheap first, then more expensive…)
- zoom in by preference (favorite food first, then less favorite…)
- may combine these criteria into a preference function

Zoom levels may require access rights

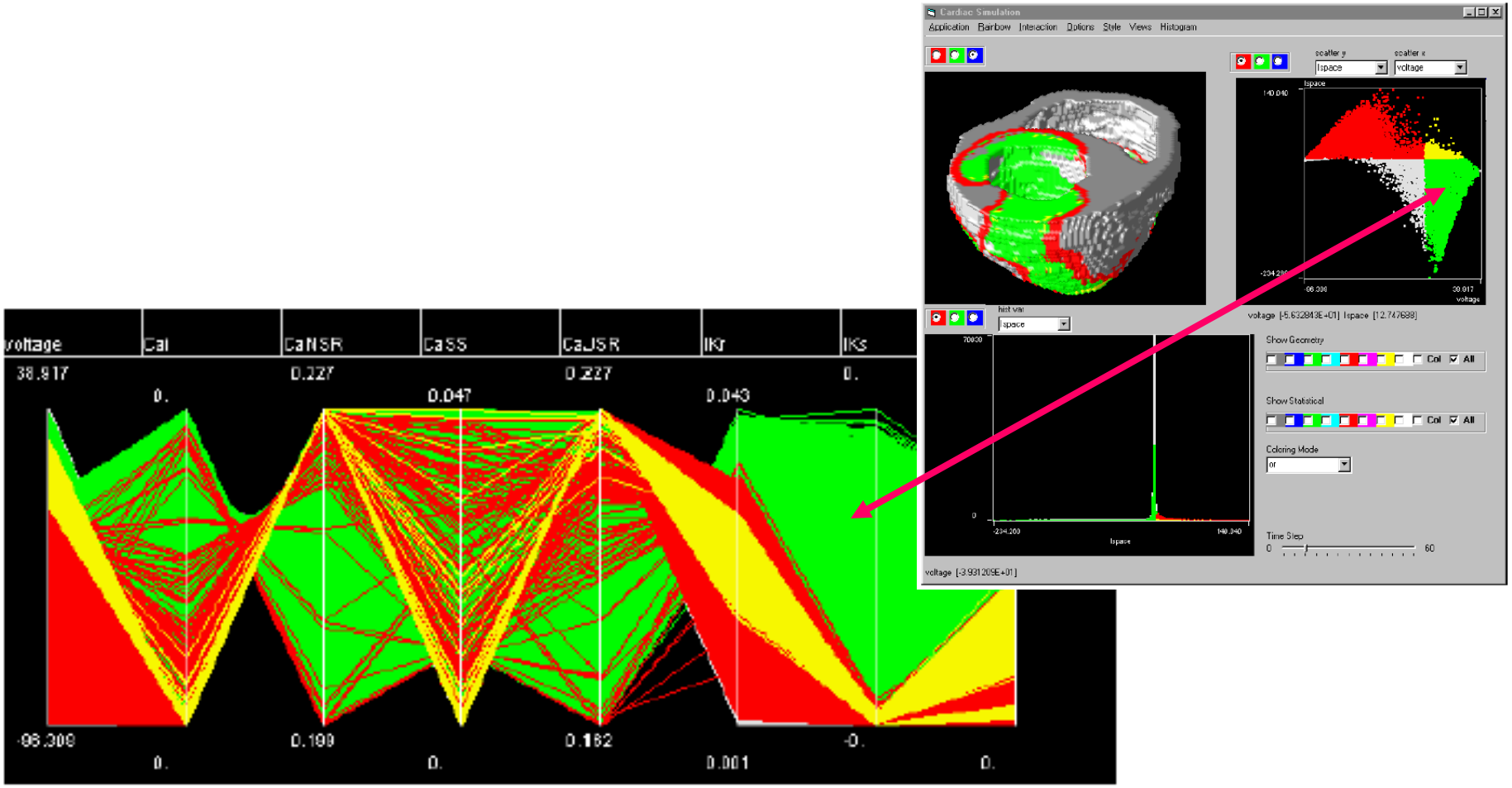
- members only
- big wallets only
- classified information



Selection by Brushing

Brushing

- users interactively select a subset of points and see how these changes are updated in other linked (related) views

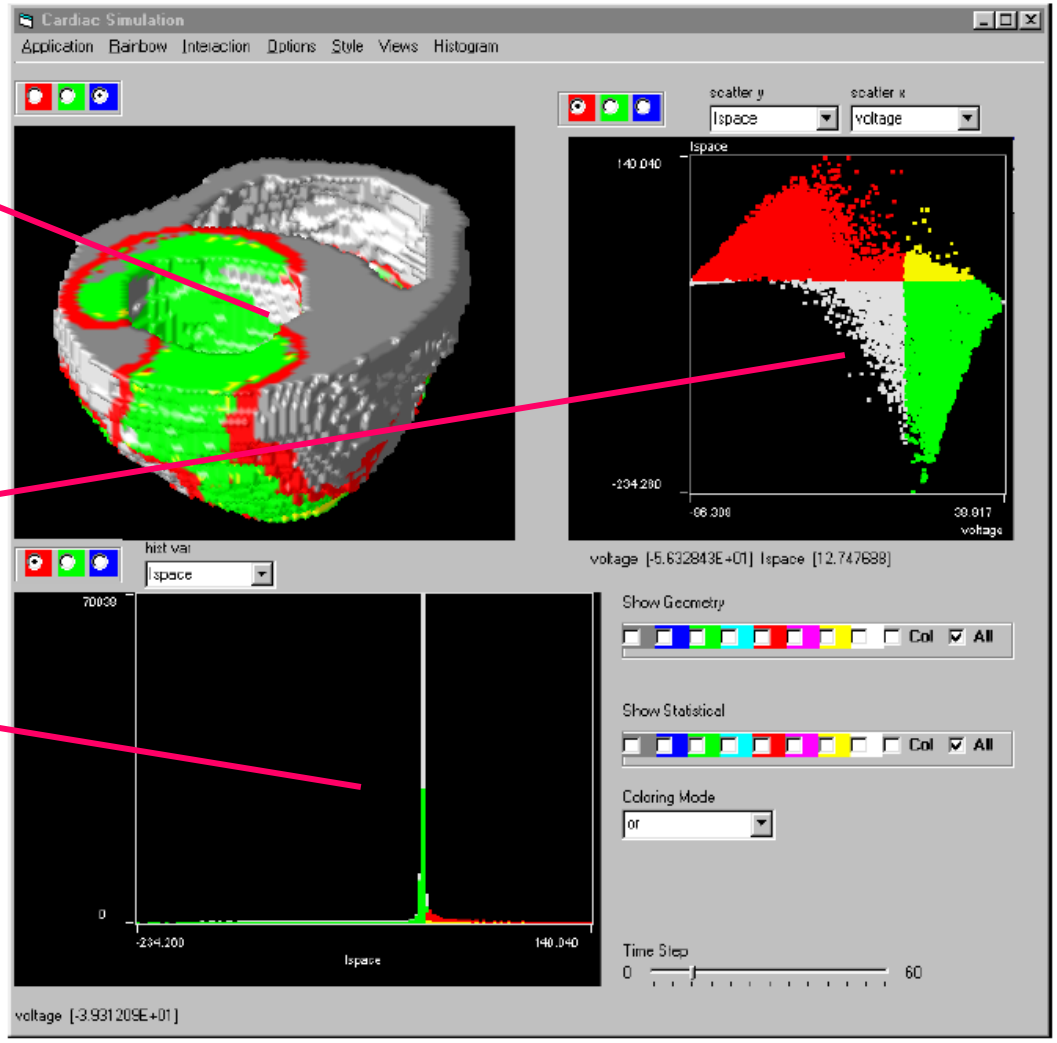


Selection by Brushing

Structure of the heart using 3-D anatomical data

Scatterplot, using measurement data.

Histogram of *Is, using measurement data*

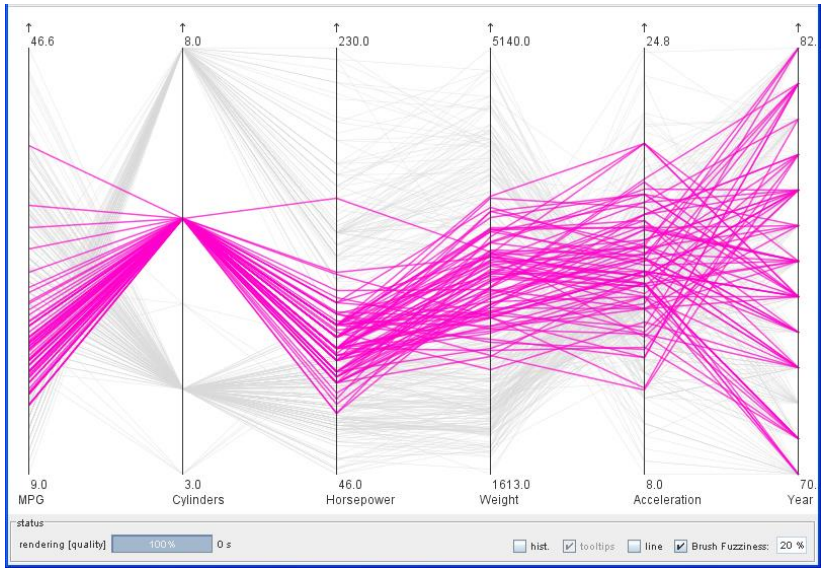


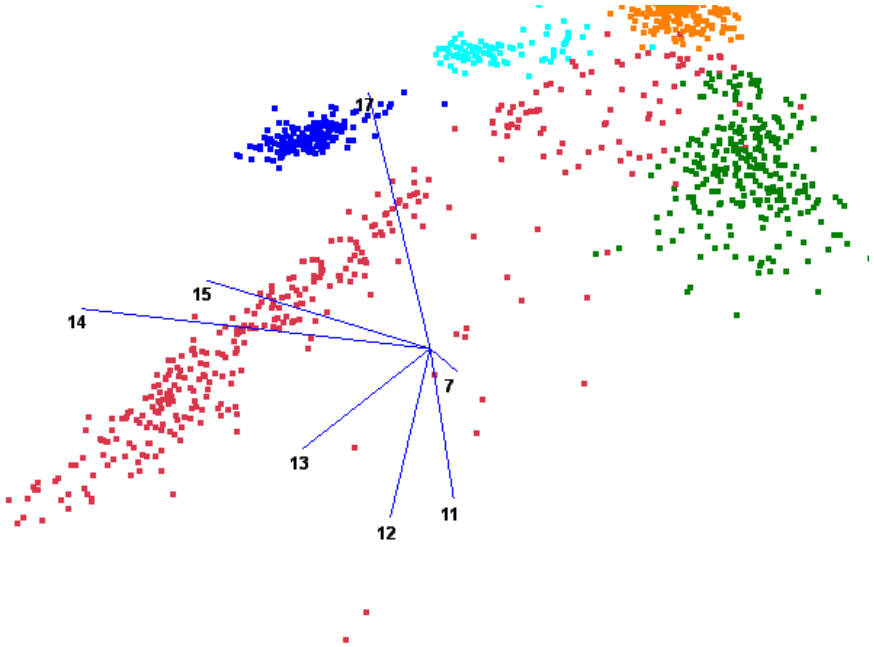
WEAVE, Gresh et al. 2000

Brushing: Highlighting

Use mouse interaction to highlight points and lines in

- parallel coordinates
- scatterplots

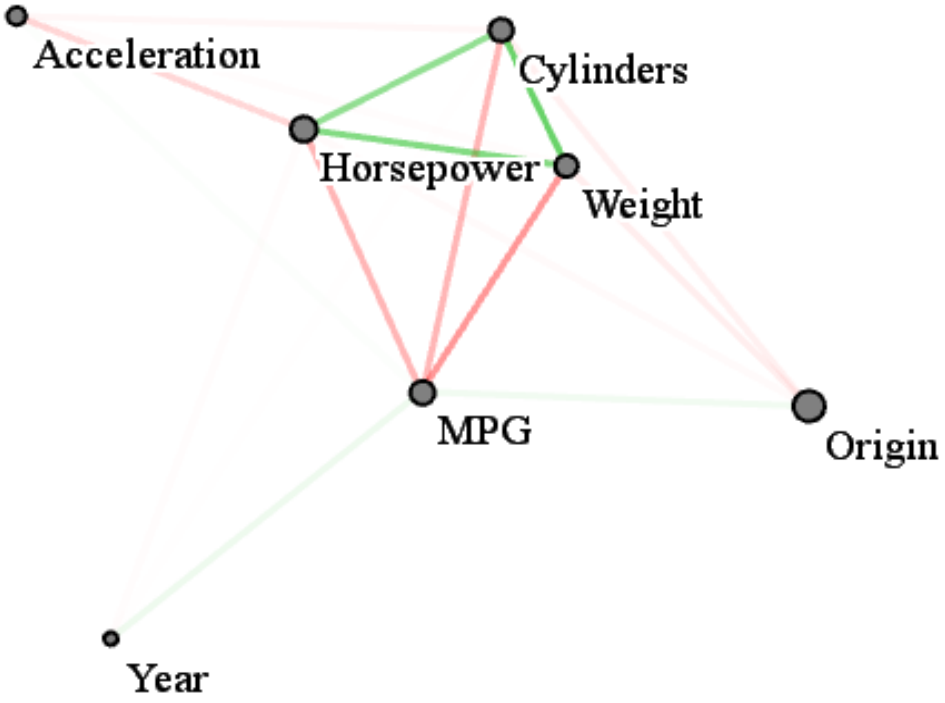




Show Meta Information: Correlation of Attributes

Based on Pearson's correlation for all attribute pairs

- vertices: attributes
- edges: correlation between attributes



Correlation Graph

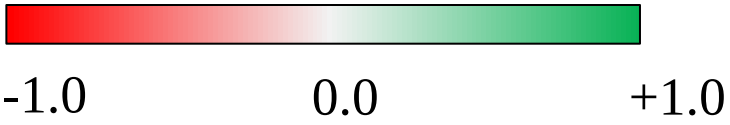
Use Mass-Spring Model for layout

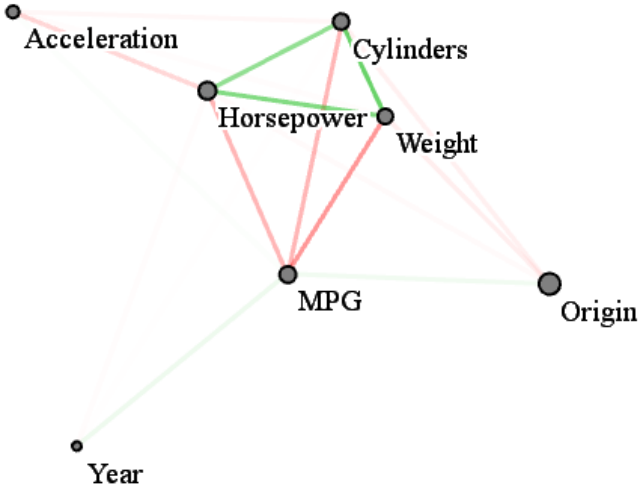
Edge is weighted by the correlation

- negative, positive, or absolute

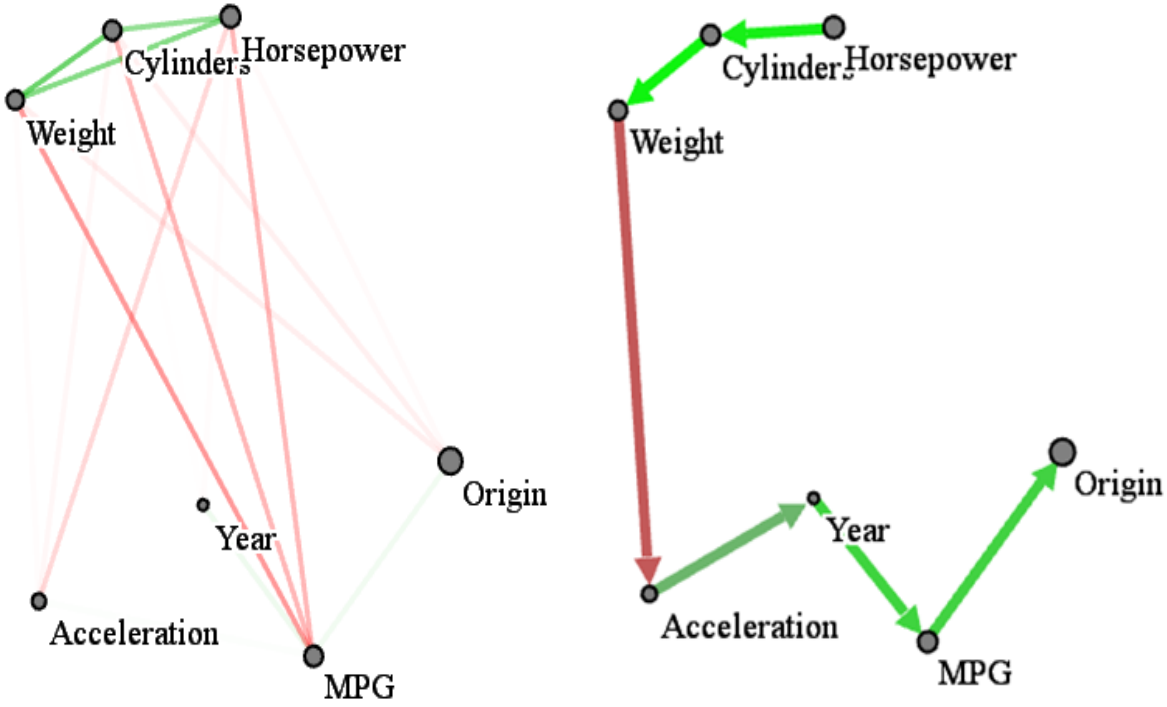
The correlation itself is encoded by color

- green : correlation = 1.0
- red : correlation = -1.0
- linear interpolation computes the colors in between.

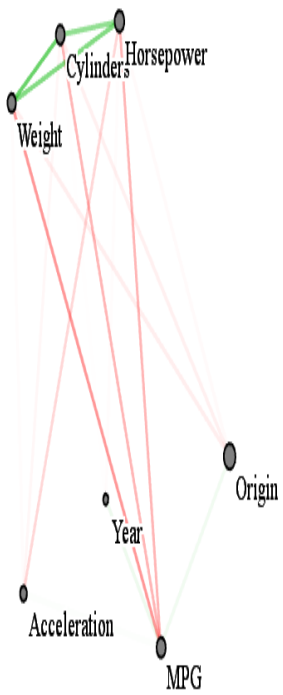


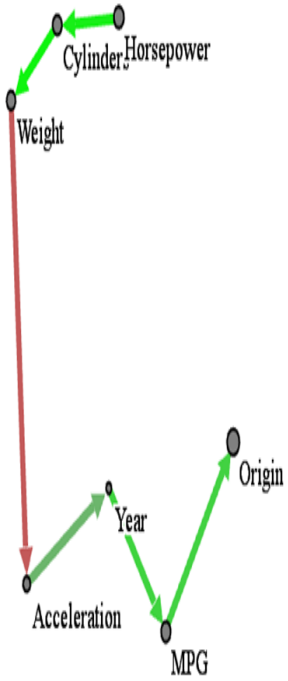


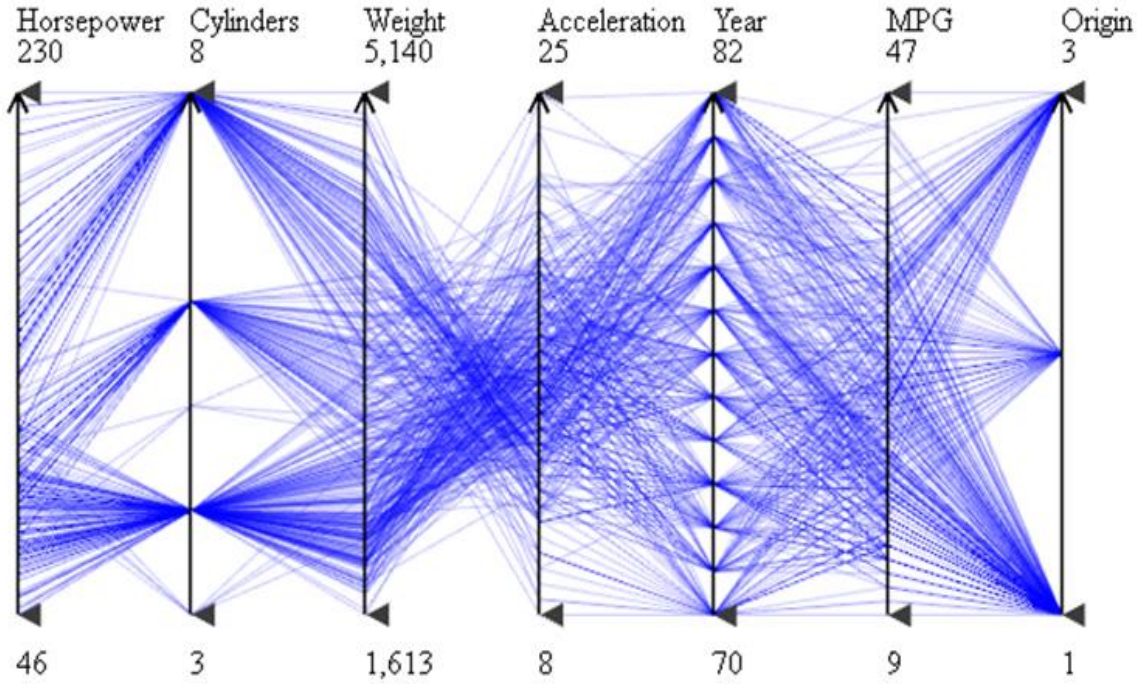
Compute Optimal Path



Re-Order Parallel Coordinate Display



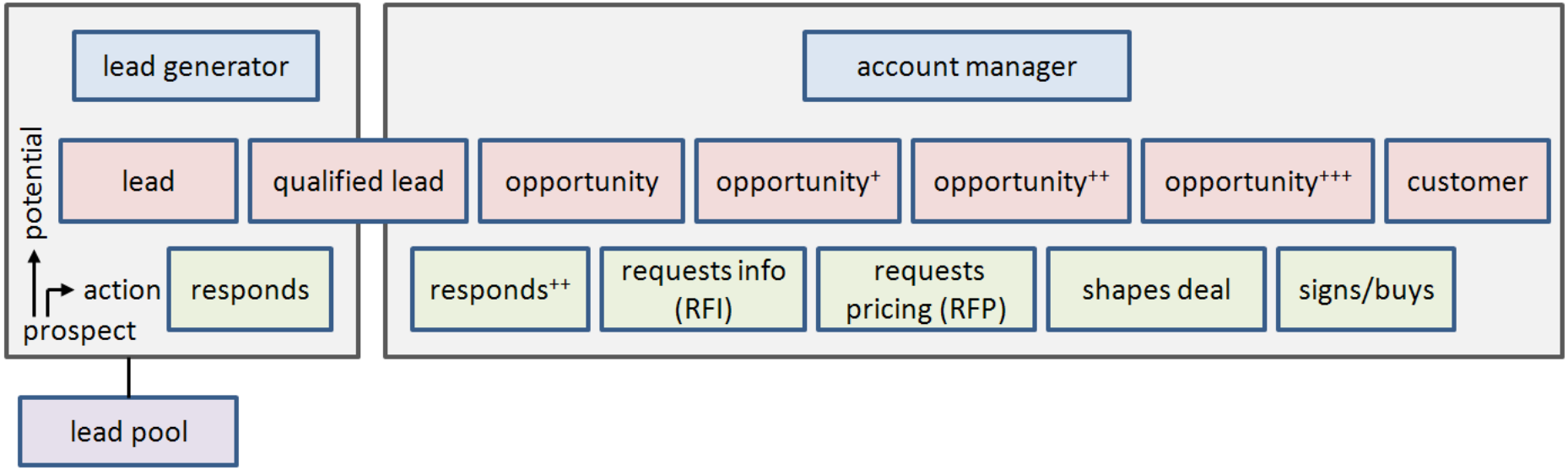




Practical Case Studies

Sales Campaign Analysis

Anatomy of a Sales Pipeline



The Setup

Scene:

- a meeting of sales executives of a large corporation, Vandelay Industries

Mission:

- review the strategies of their various sales teams

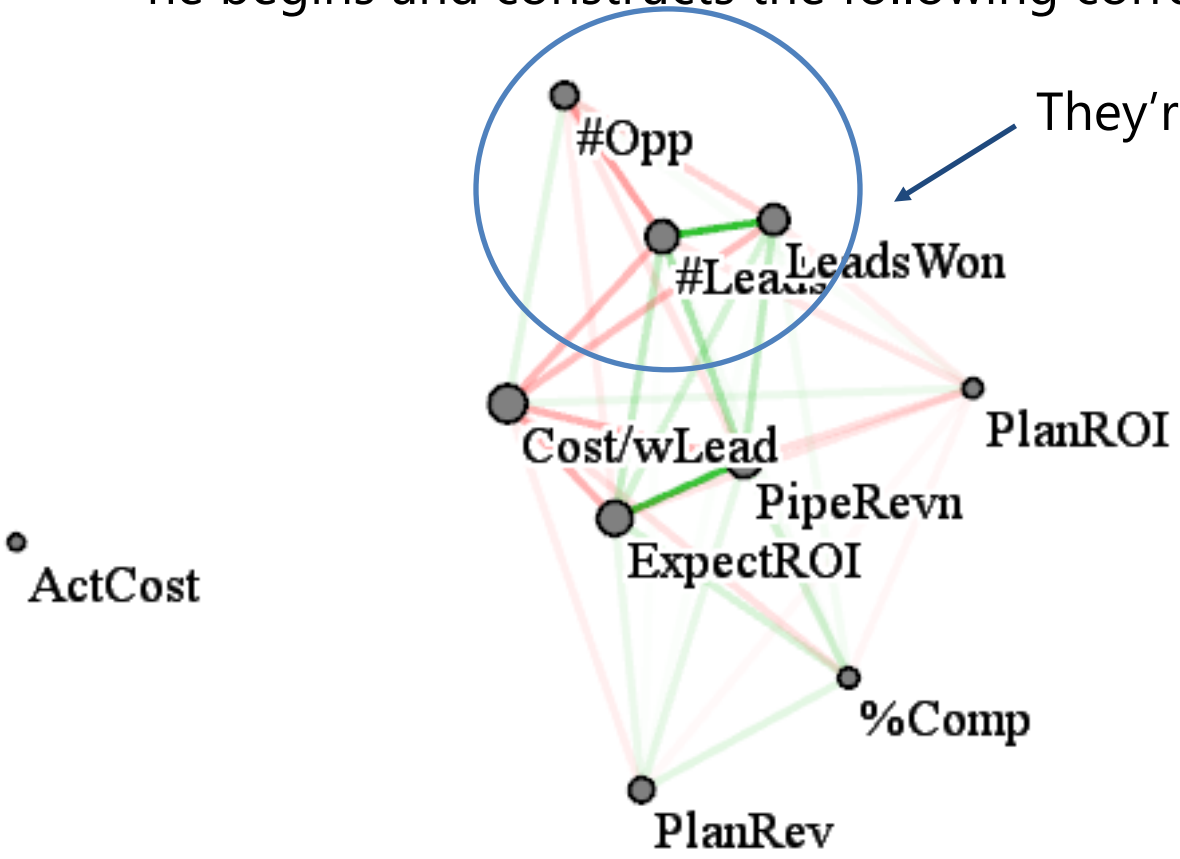
Evidence:

- data of three sales teams with a couple of hundred sales people in each team

Jim Begins

Meet Jim, one of the sales strategy analysts

- he begins and constructs the following correlation graph



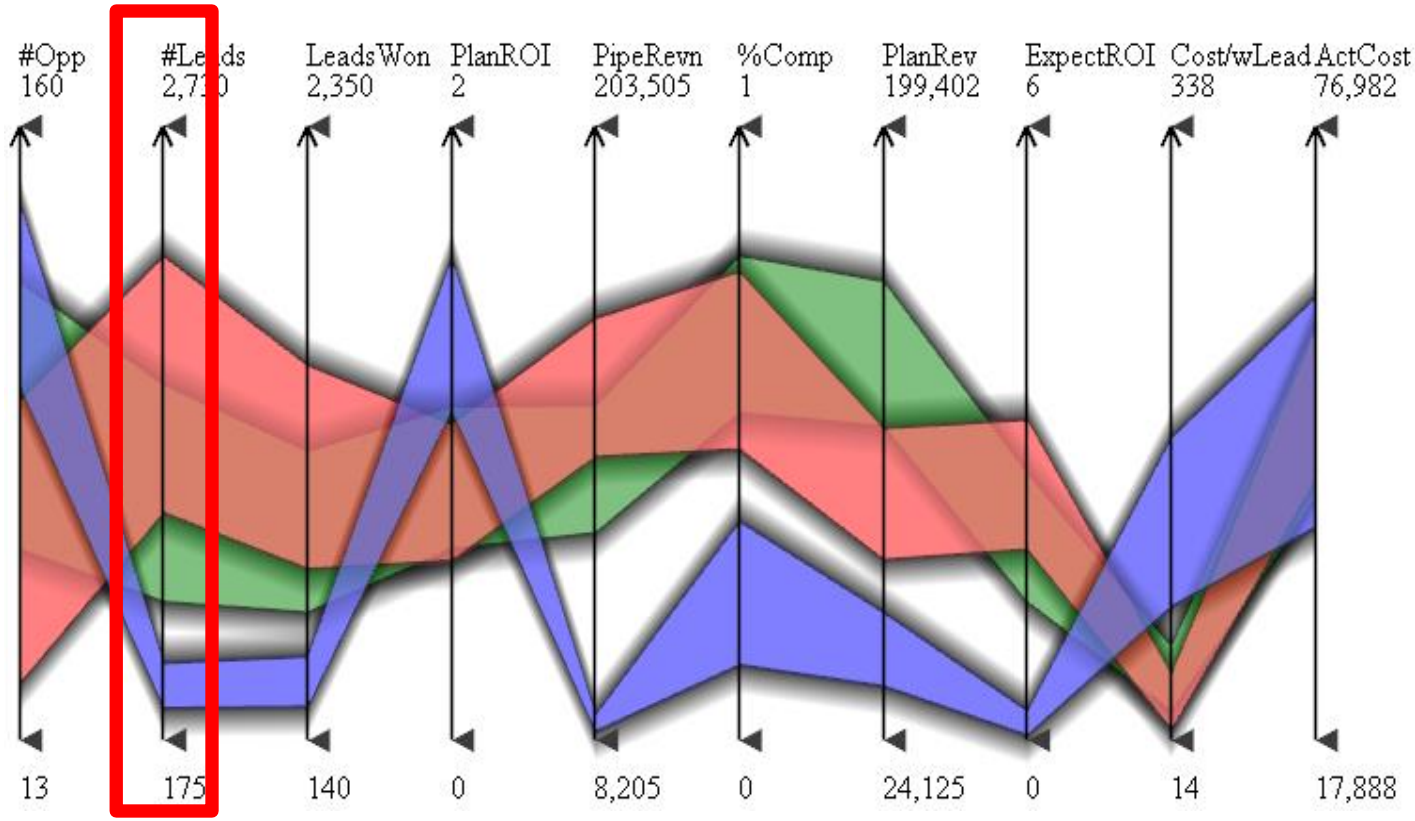
They're related!

Jim

Jim's Story

He asks the TSP to compute an initial route

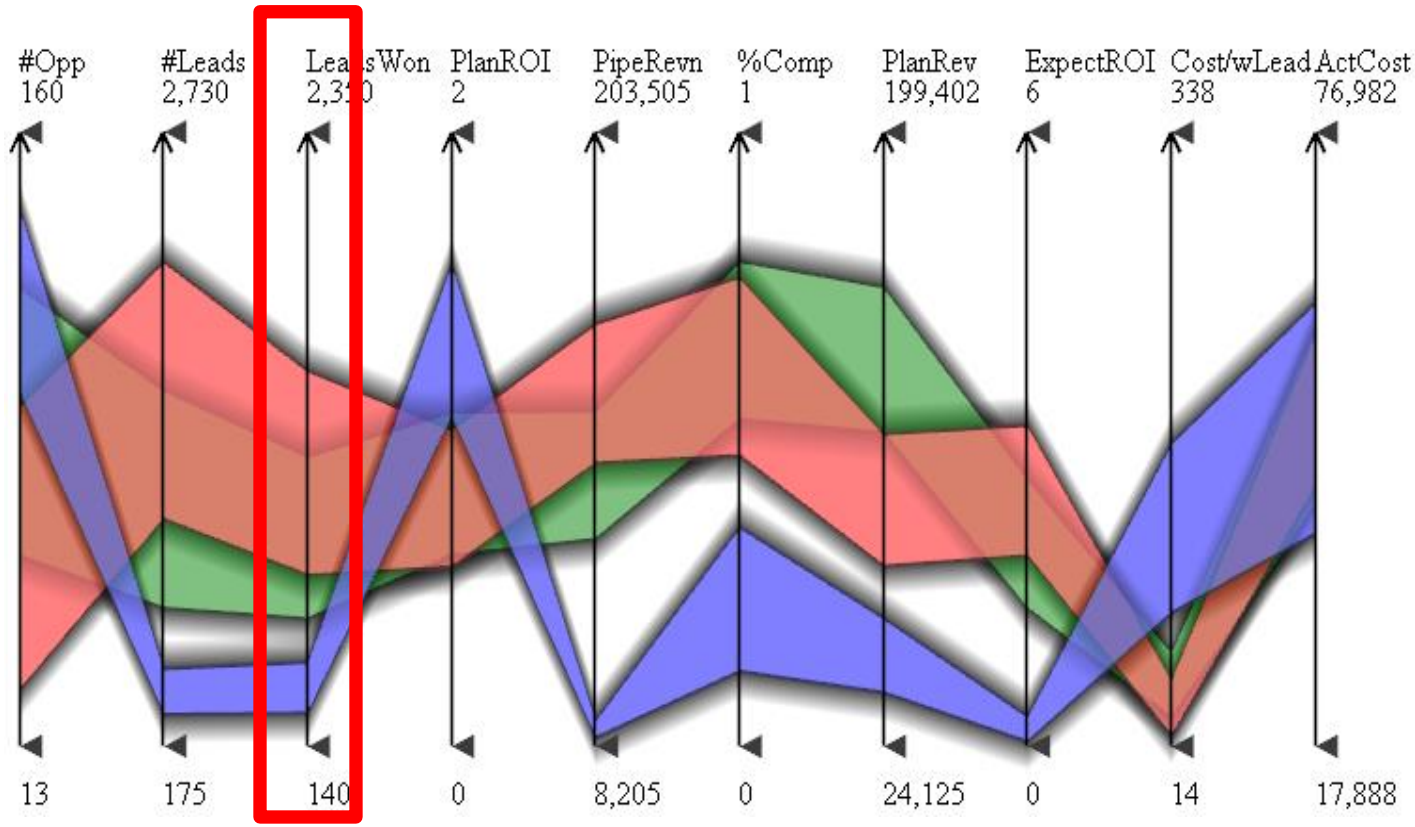
It gives rise to this parallel coordinate display



Jim's Story

He asks the TSP to compute an initial route

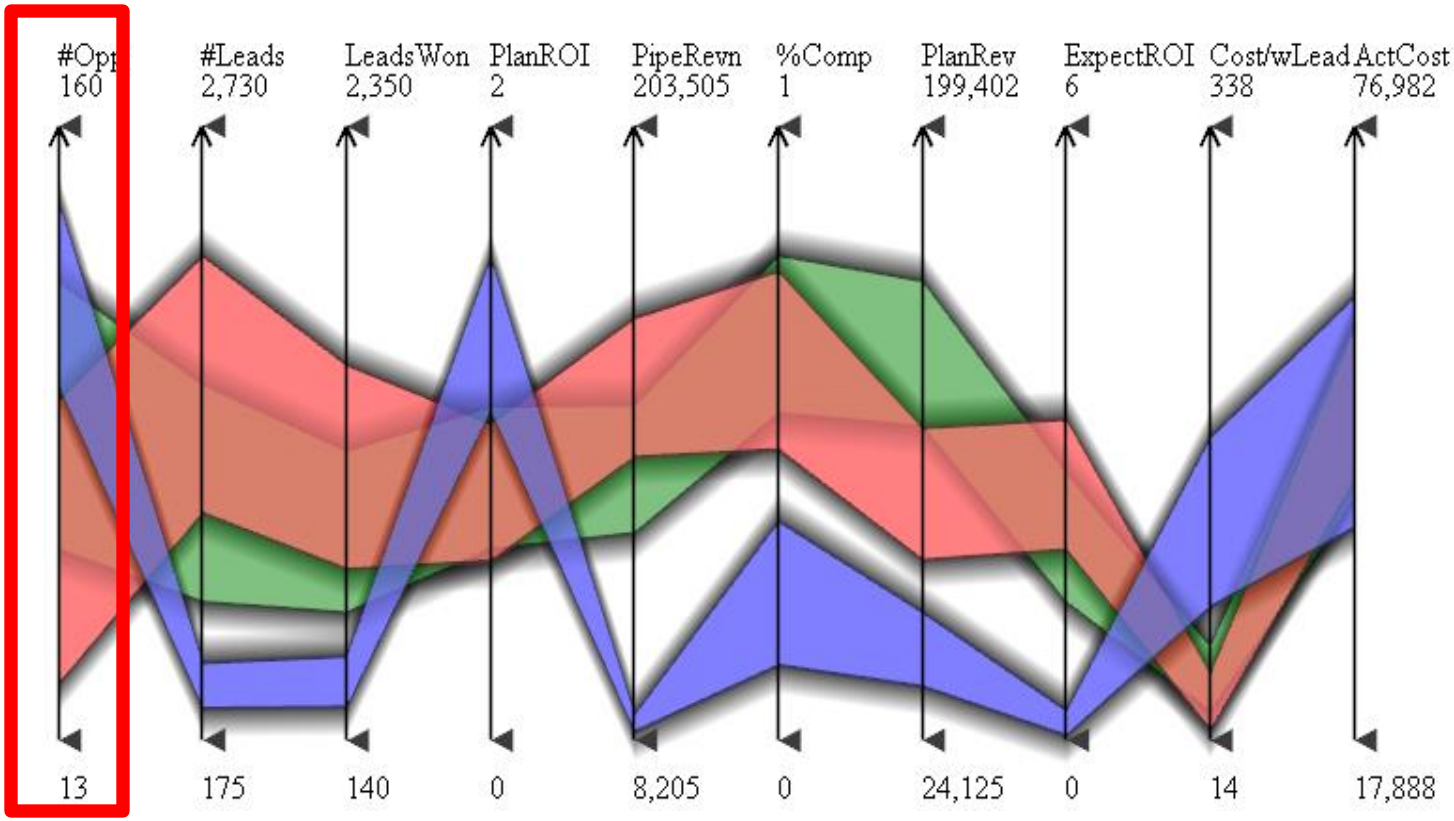
It gives rise to this parallel coordinate display



Jim's Story

He asks the TSP to compute an initial route

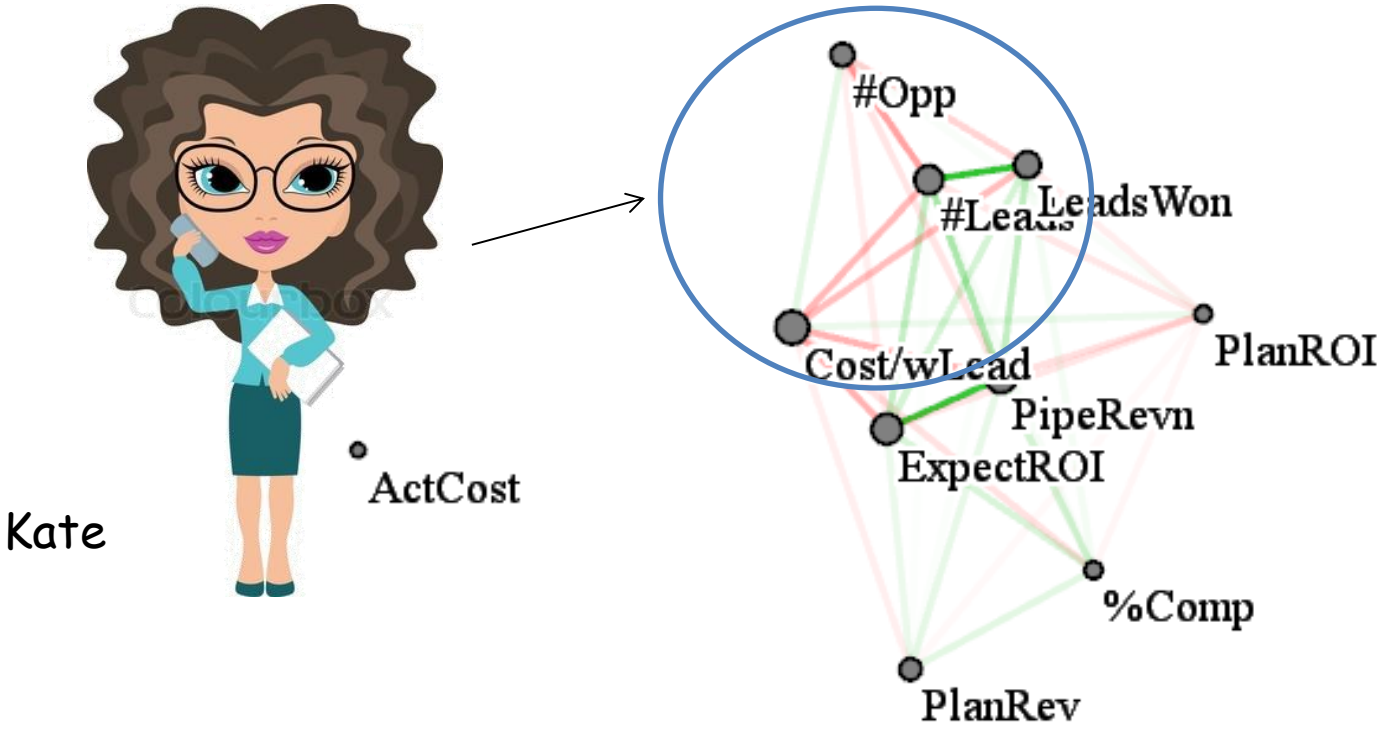
It gives rise to this parallel coordinate display



Kate Steps In

Now meet Kate, another sales analyst in the meeting room:

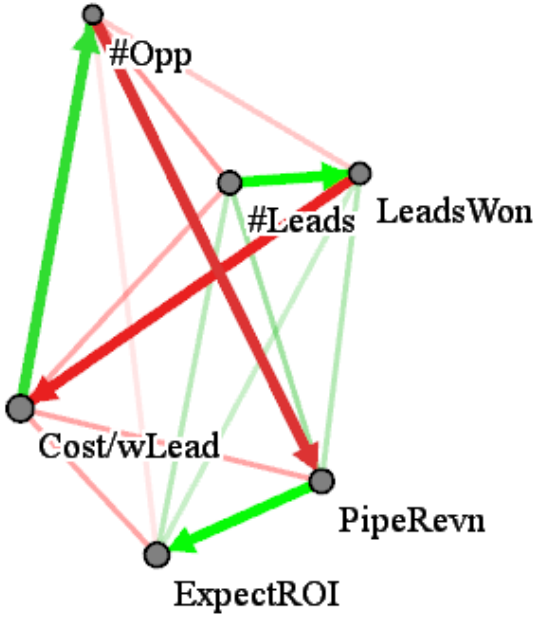
"Hey, cost/won lead is nearby and it has a positive correlation with #opportunities but also a negative correlation with #won leads"



Kate's Story

"Let's go and make a more revealing route!"

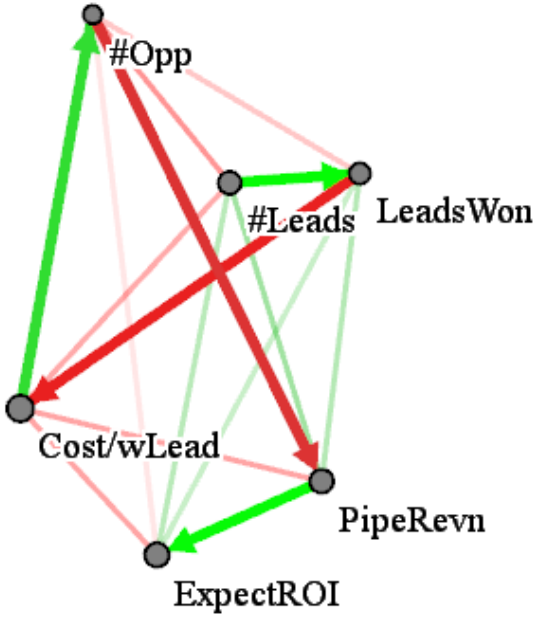
- so she uses the mouse and designs the route shown

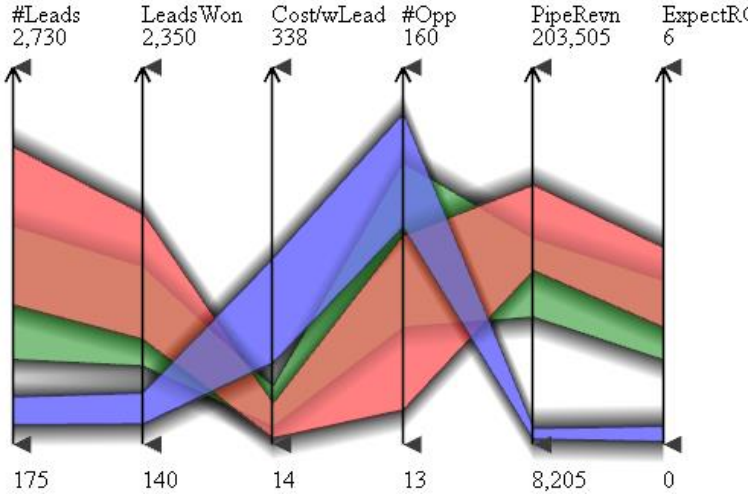


Kate's Story

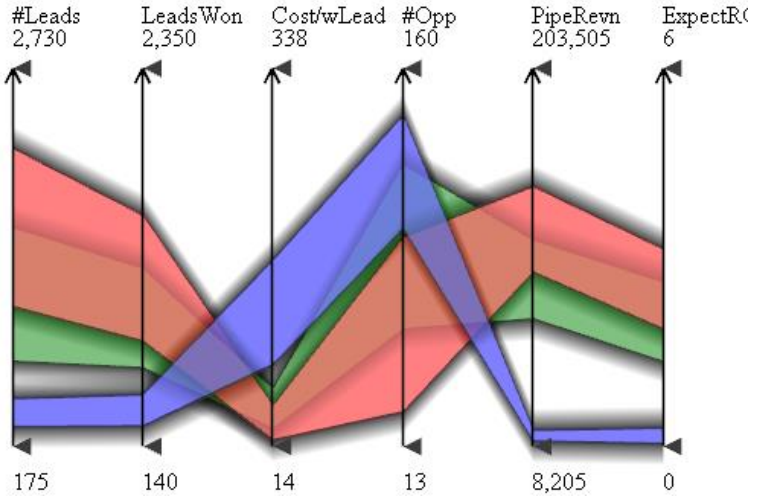
"Let's go and make a more revealing route!"

- so she uses the mouse and designs the route shown





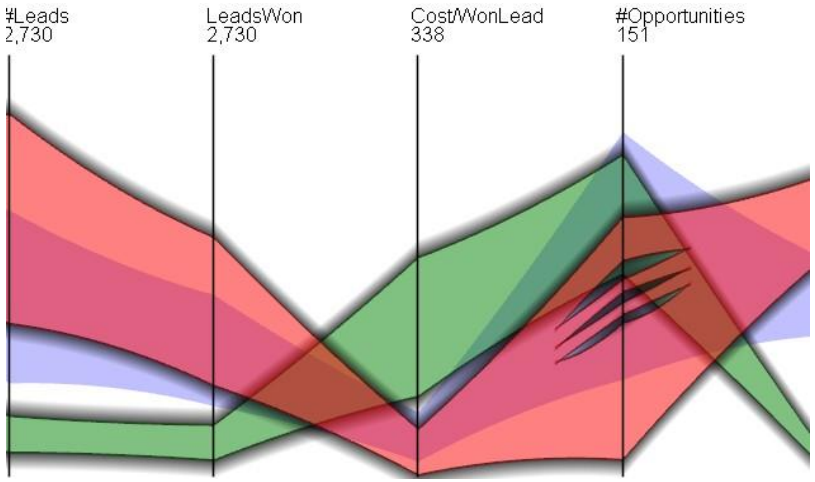
The Big Insight



It is now immediately obvious:

- the blue team employs a very different strategy than the green and the red teams.
- it generates far fewer leads but spends much more resources on each → this gives it an advantage in the final outcome.
- the blue team is also much more consistent than the other teams, as indicated by the much narrower band

Further Insight



Kate notices something else:

- now looking at the red team
- there seems to be a spread in effectiveness among the team
- the team splits into three distinct groups

She recommends: “Maybe fire the least effective group or at least retrain them”