Information Visualization and Knowledge Management

Dr. Margaret-Anne Storey
Dept of Computer Science

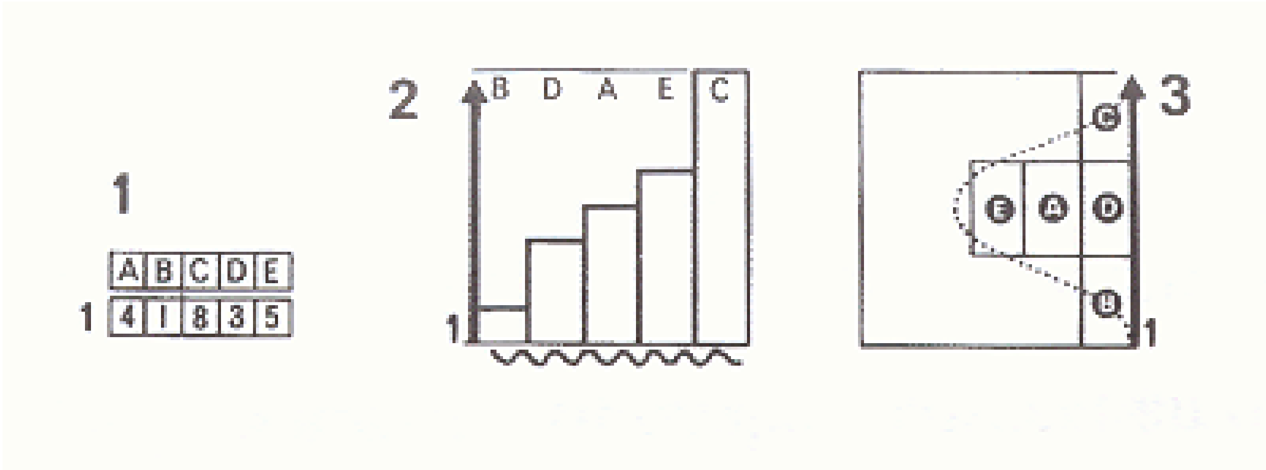
1D, 2D, 3D and > 3D

How much can we see?

- We can perceive only in 3D (4D if we include time)
- But we can see more than 3 or 4 dimensions if we make use of Bertin's retinal variables
- But how many can we effectively use?

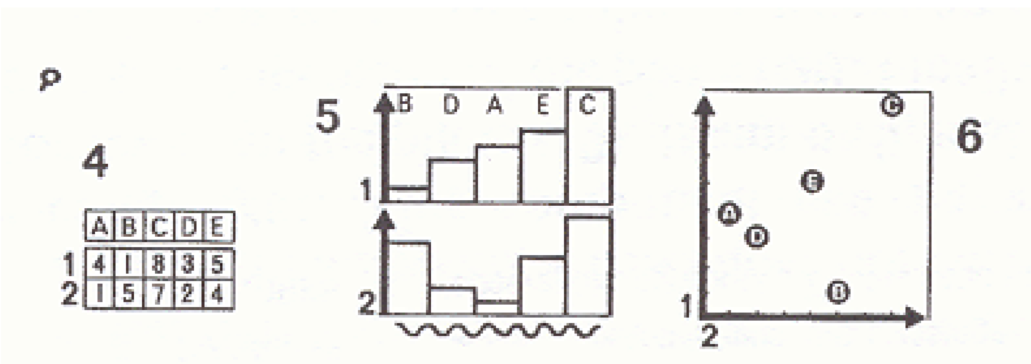
Univariate data, 1D

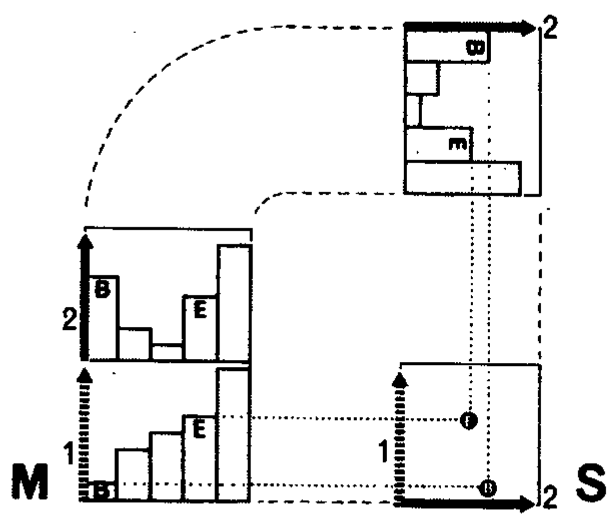
- We can use bar charts, point plots….



Bivariate data, 2D

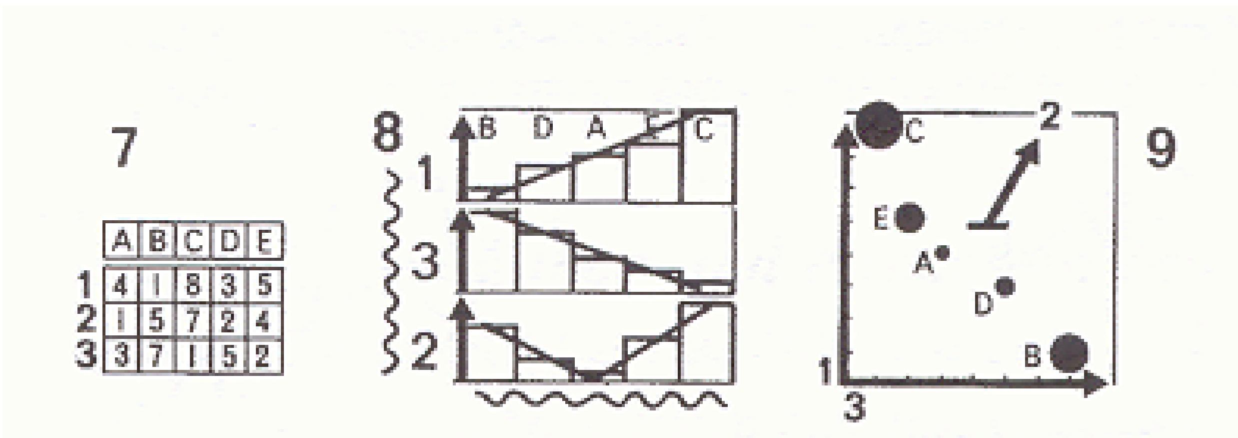
- We can use scatterplot (really two barcharts projected onto a single image)
- Two barcharts side by side
- Or perhaps a barchart and one retinal variable



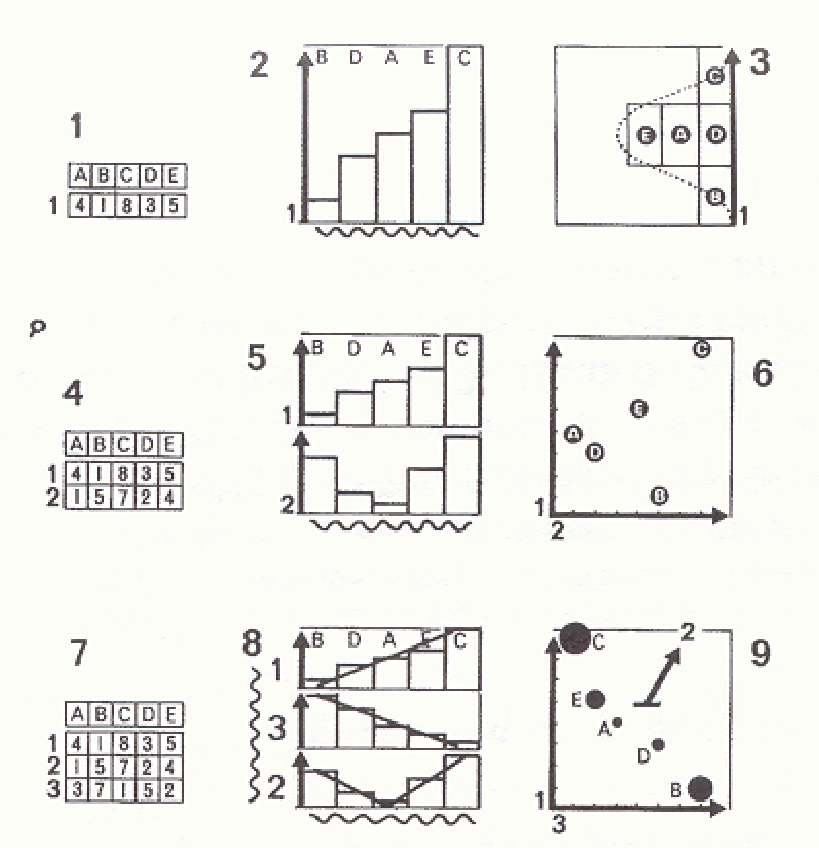


Trivariate data

- Scatter plot + retinal variable
- Scatter plot in 3D (using perspective)
- Or show multiples of 1D drawings

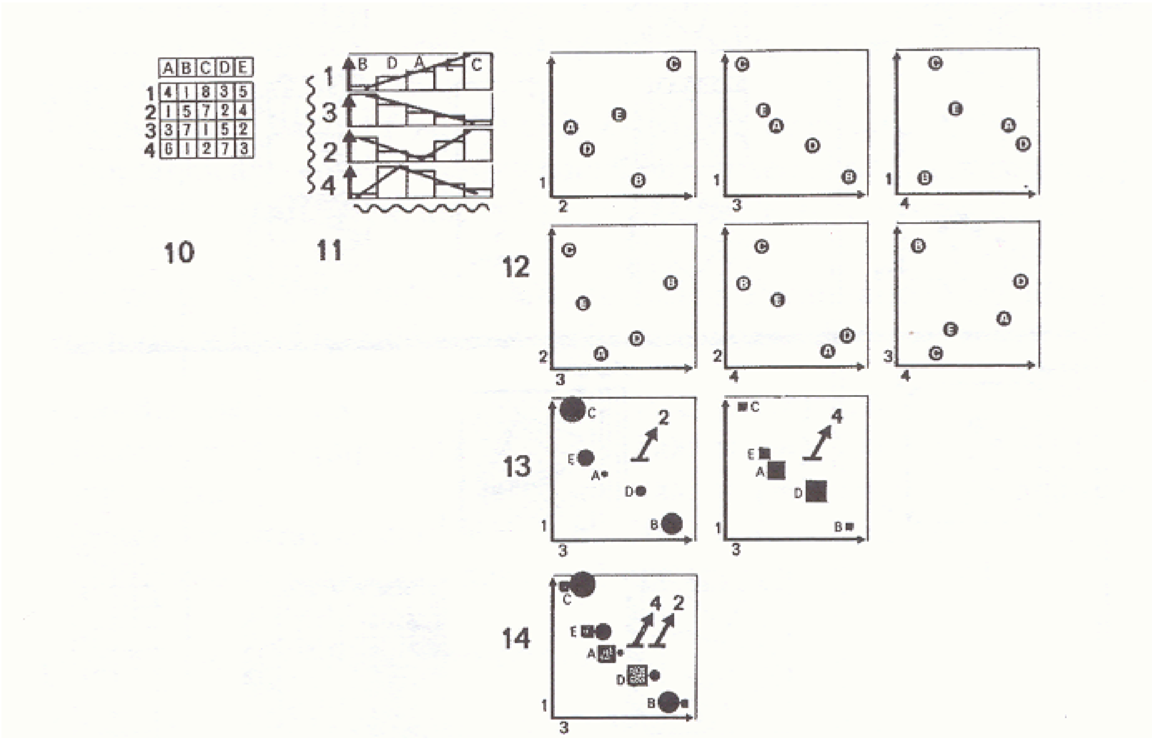


1, 2 and 3 rows



- Axes are the most effective dimension for encoding data in space
- 1, 2, 3 dimensions of data are also special cases of n dimensions

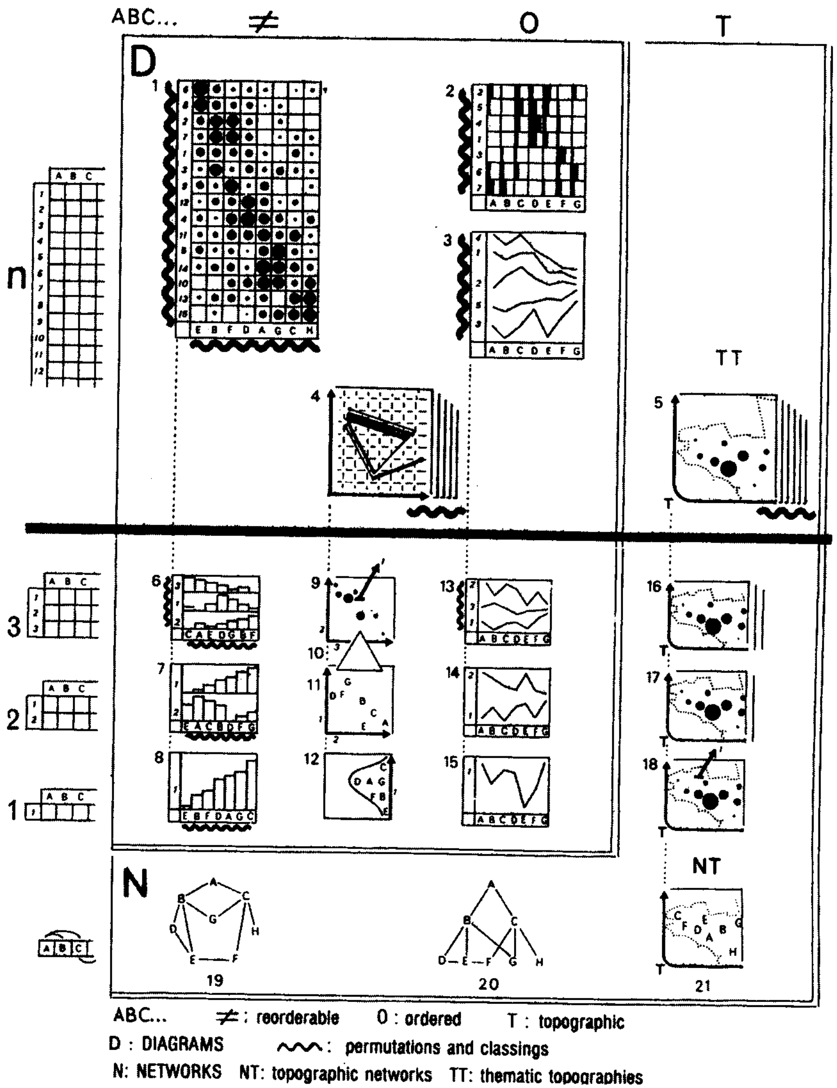
Multivariate data, >3D



"Impassable" barrier of 3 (or 4) dimensions…??

A table with more than 3 rows cannot be constructed directly as a single signifying image – we need a series of images

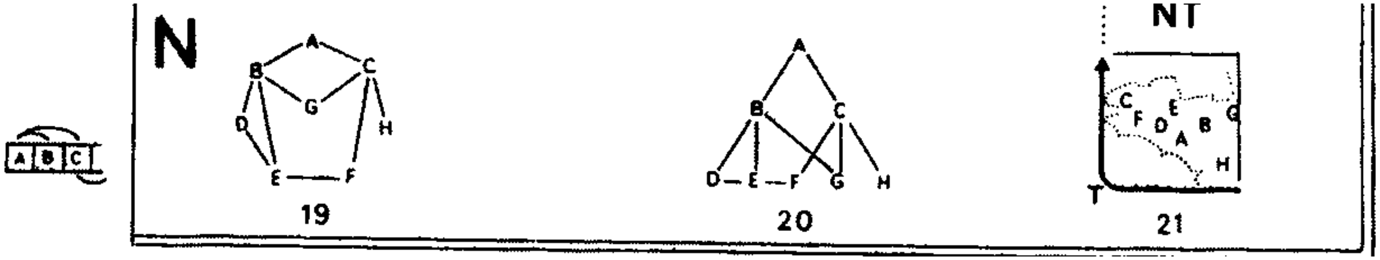
Multivariate data



- We can use a matrix permutation and discover all relationships through permutations (dynamic graphics)
- But we can permute rows only if the data is reorderable
- If the data is ordered, we can create an image file or array of curves (but only if the slopes are meaningful)
- Pick the series of images and select an important relationship to reveal
- We can also use interaction and dynamic manipulation to reveal relationships

Networks

- A Network (N) portrays the relationships which exist among the elements of a single component
- Can be unordered, so the image can be transformed any way in the plane
- If it is ordered, it can't be transformed (a tree is ordered in 1D, a map is ordered in 2D)
- Networks can be represented in matrix form (which can be permuted if it is reorderable)

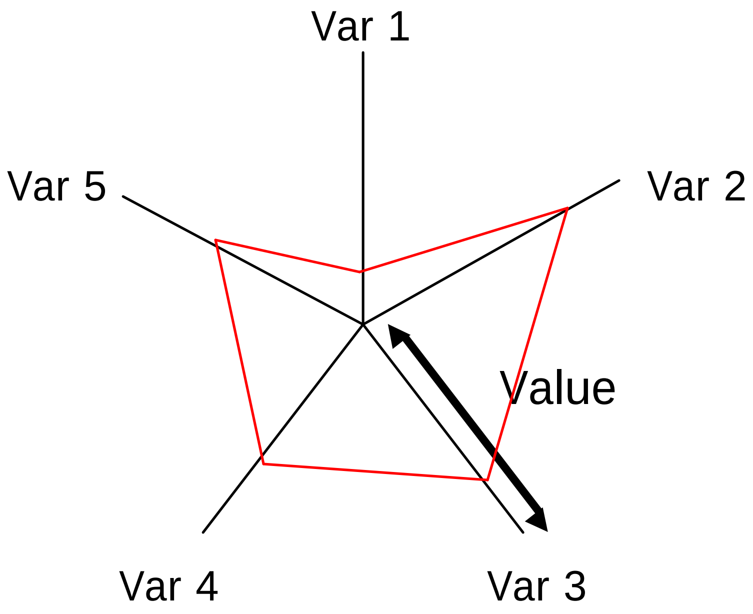


Bertin’s synoptic

Synoptic -- *affording a general view of the whole*

- Start with the # of rows and nature of the components
- Select the image construction corresponding to the structure of the data (ordered, unordered, ordinal, nominal etc)
- Decide which is the important relationship to show first
- Use permutations to disclose other relationships

Star Plots

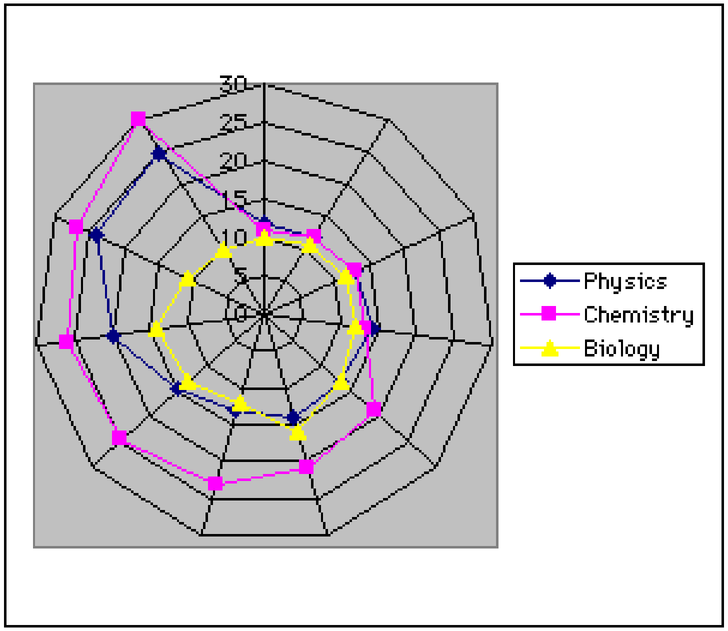


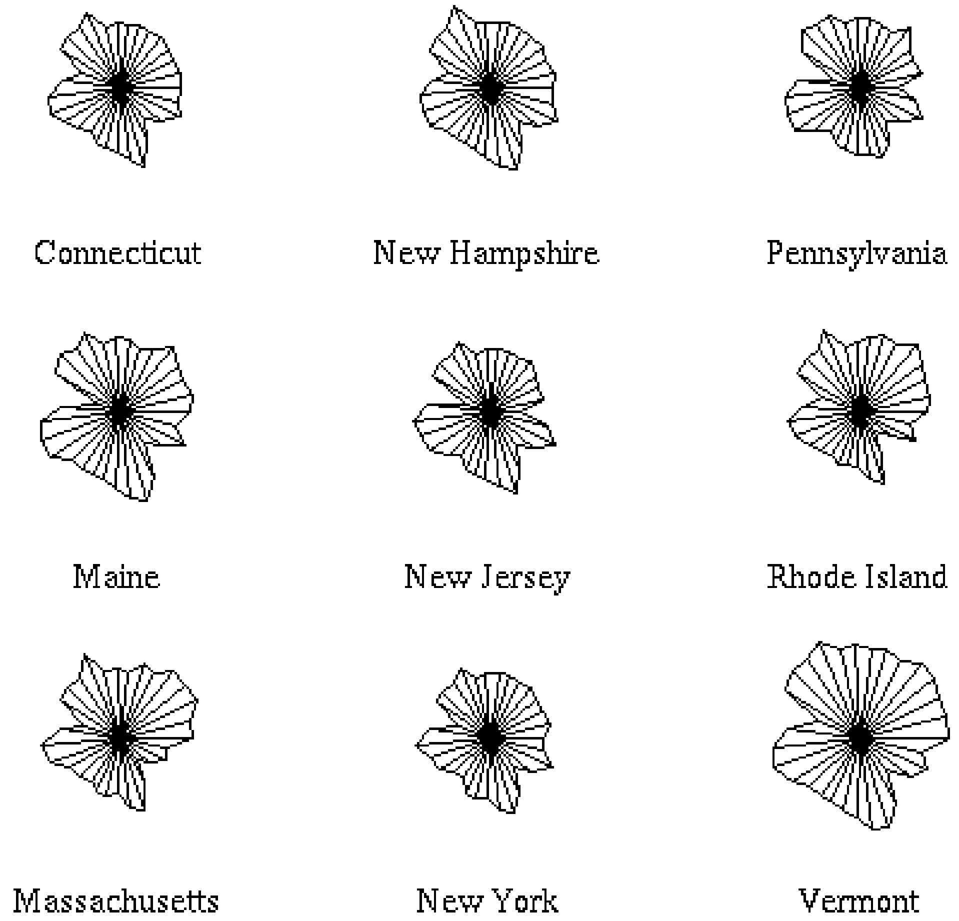
Space out the n variables at equal angles around a circle

Each “spoke” encodes a variable’s value

Based on John Stasko slide

Starplots





- http://seamonkey.ed.asu.edu/~behrens/asu/reports/compre/radar_plot.GIF

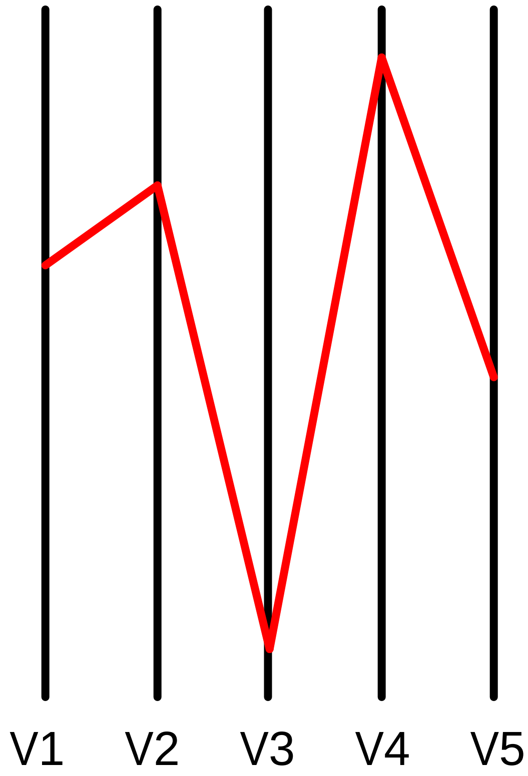
Multidimensional Detective

- System that uses parallel coordinates for information analysis and discovery
- Interactive tool
  - Can focus on certain data items
  - Color

Taken from:
A. Inselberg, “Multidimensional Detective”
InfoVis ‘97, 1997.

Based on John Stasko slide

Explicit Variables: Parallel Coordinates (Inselberg)



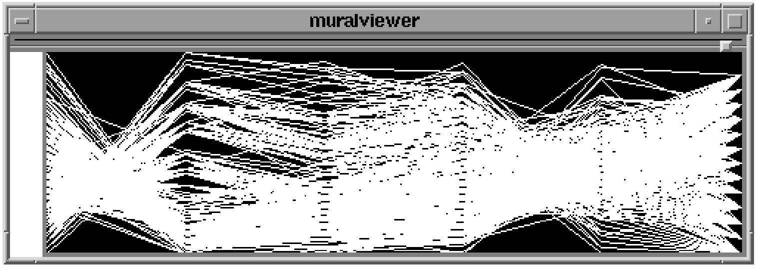
Encode variables along a horizontal row

Vertical line specifies single variable

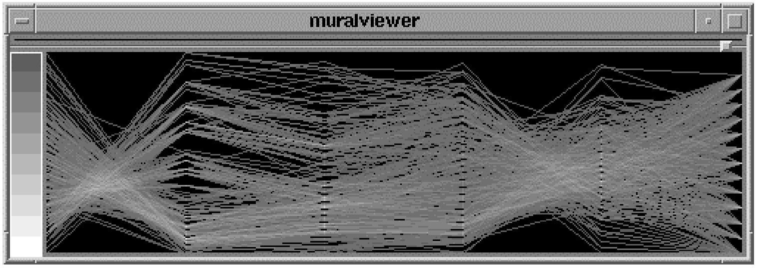
Red line specifies a case

Based on John Stasko slide

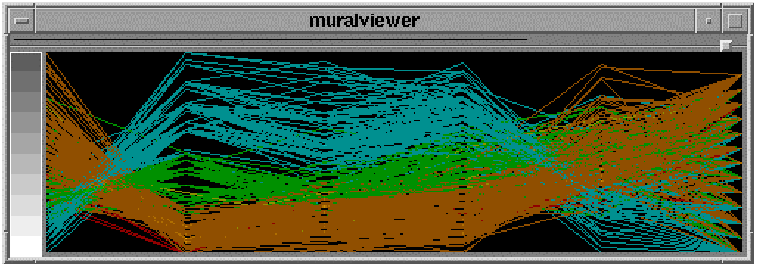
Parallel Coordinates Example



Basic



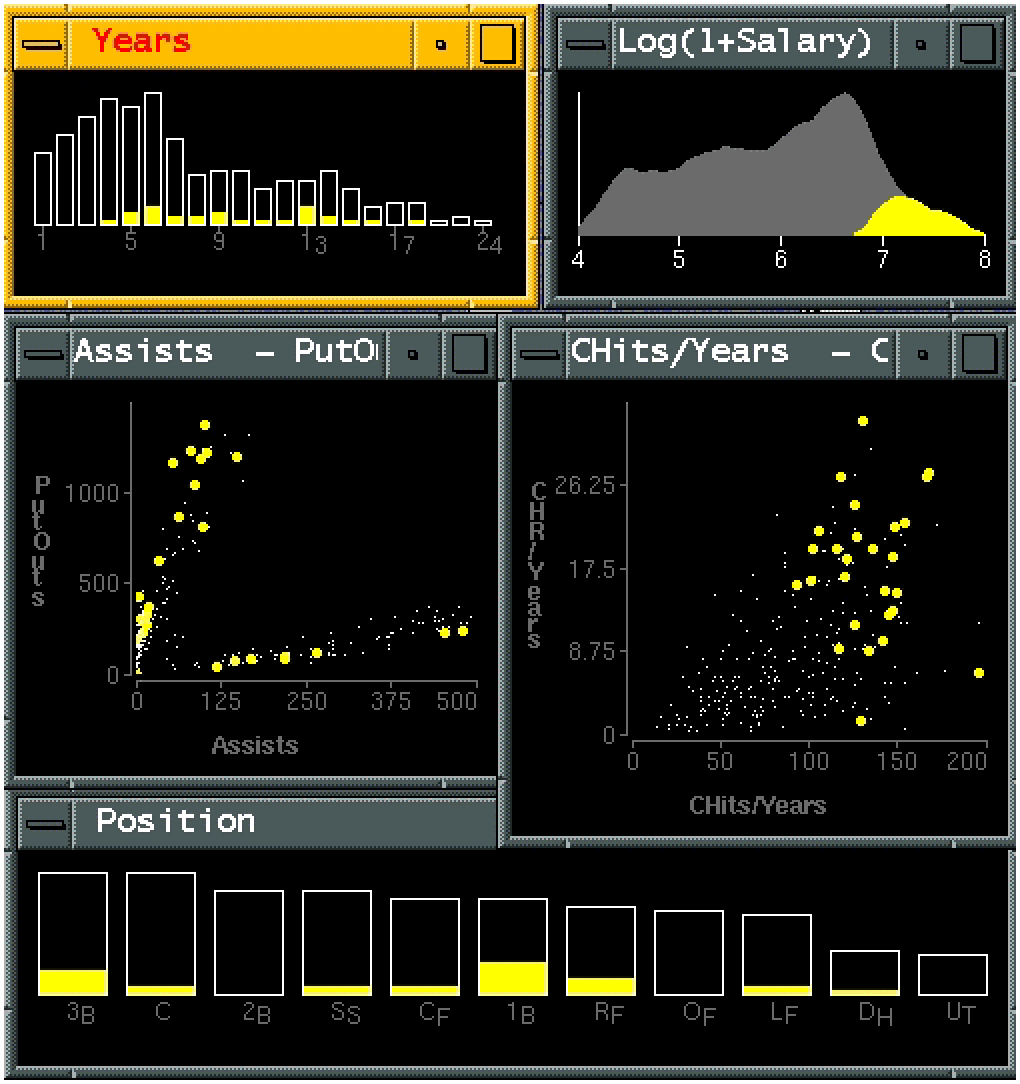
Grayscale



Color

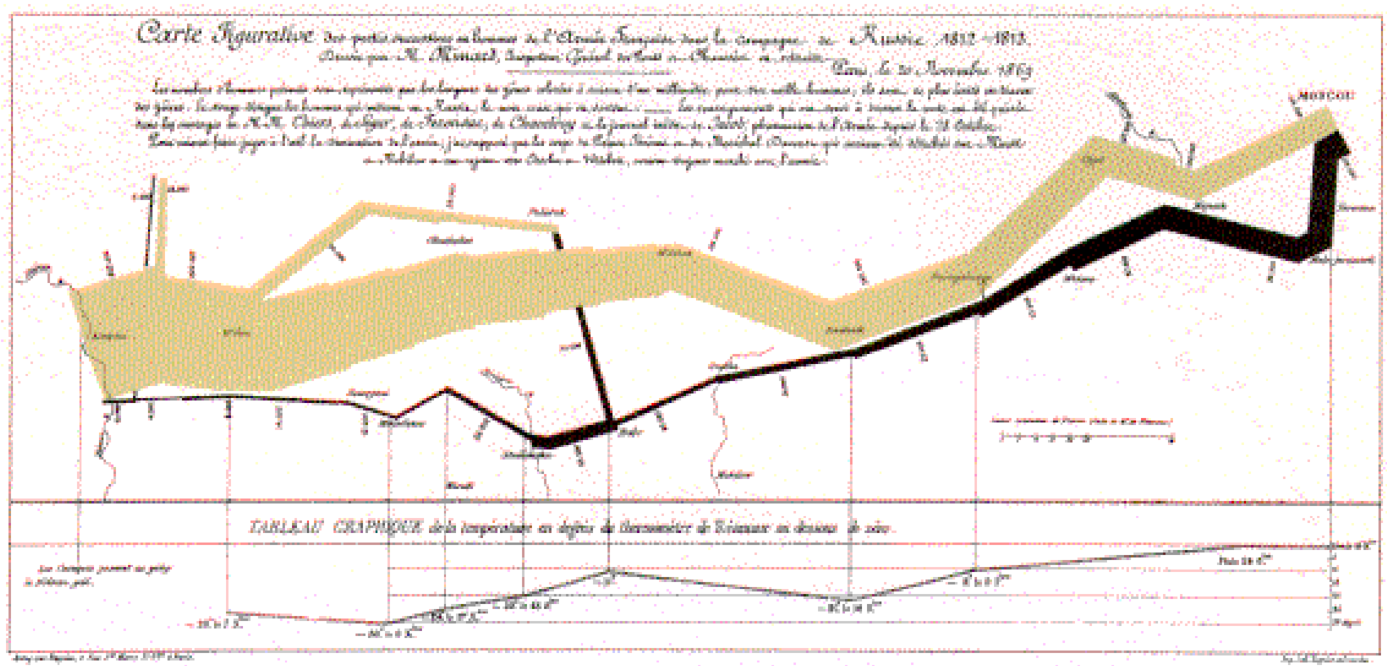
Based on John Stasko slide

Link different types of graphs: Scatterplots and histograms and bars (from Wills 95)

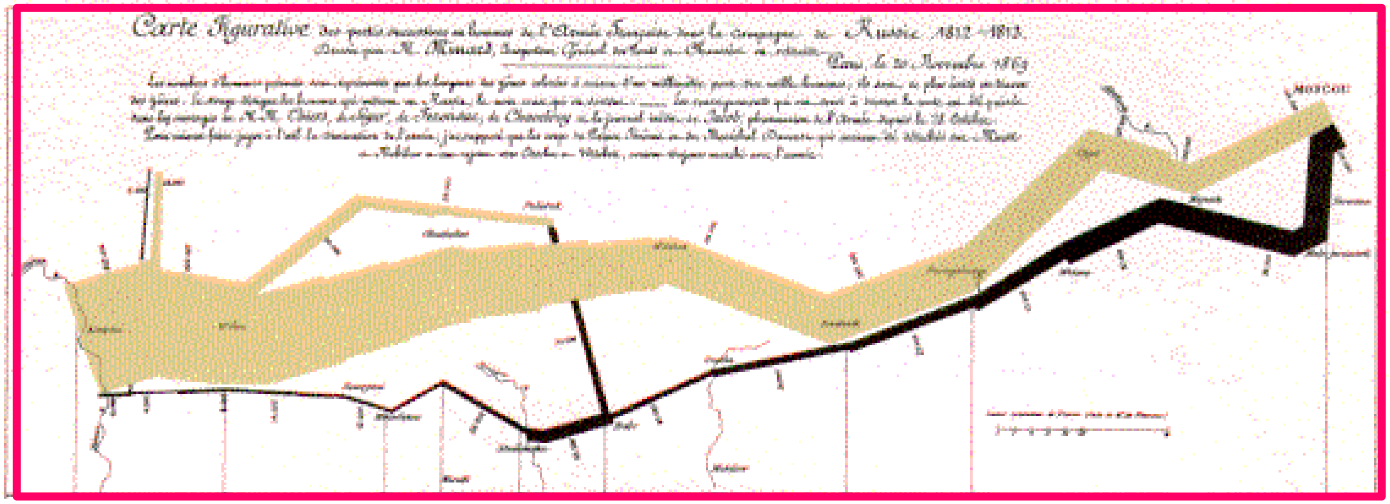


Composition / decomposition

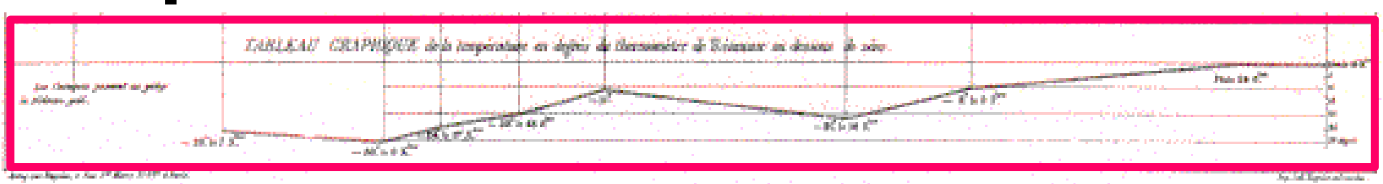
- Minard’s 1869 Napoleon’s march



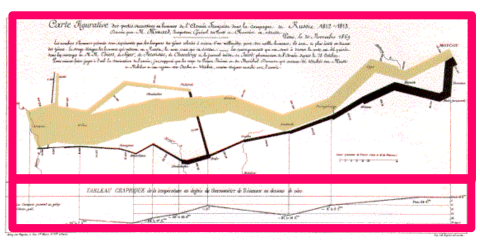
Single axis composition



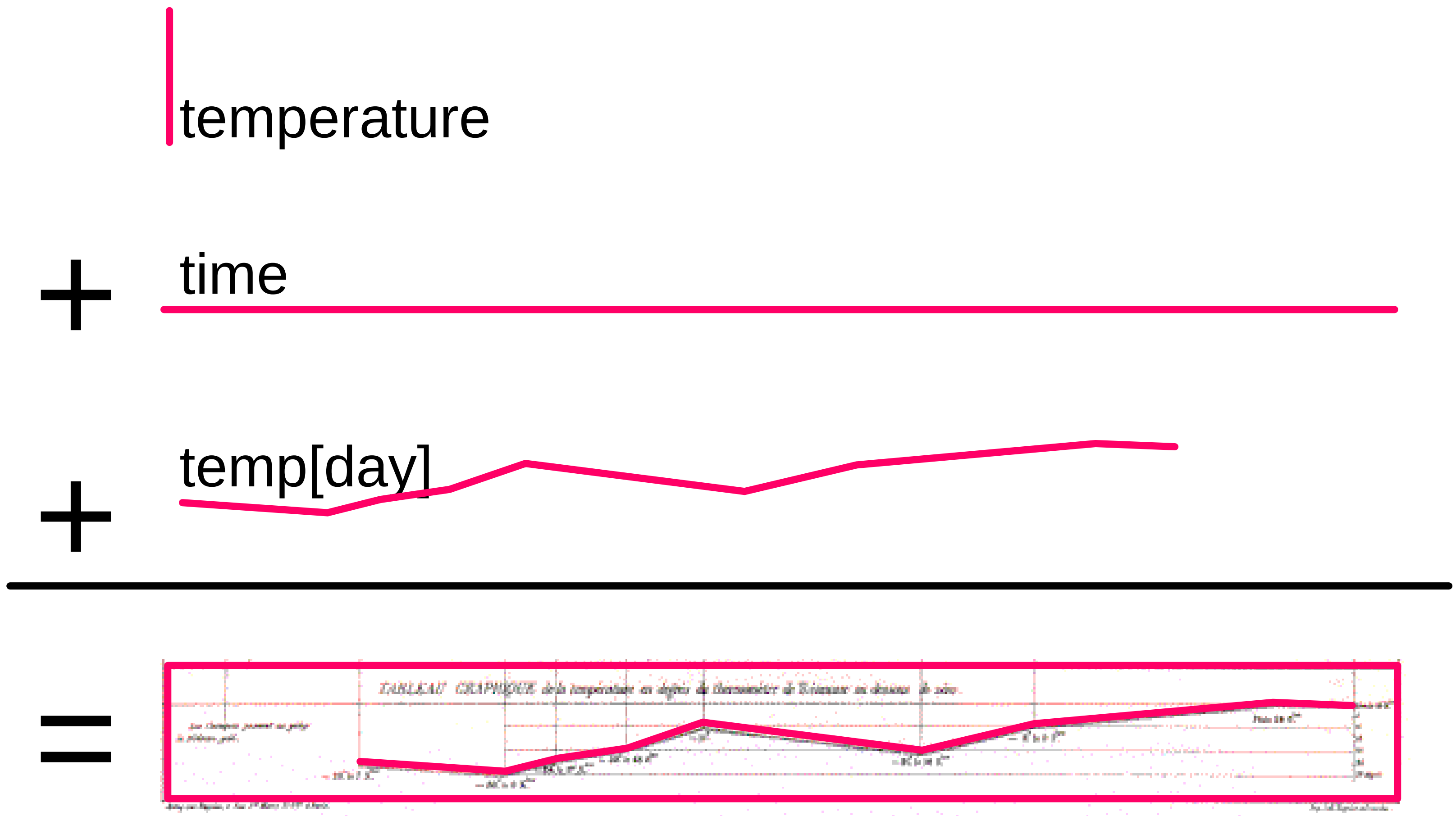
+



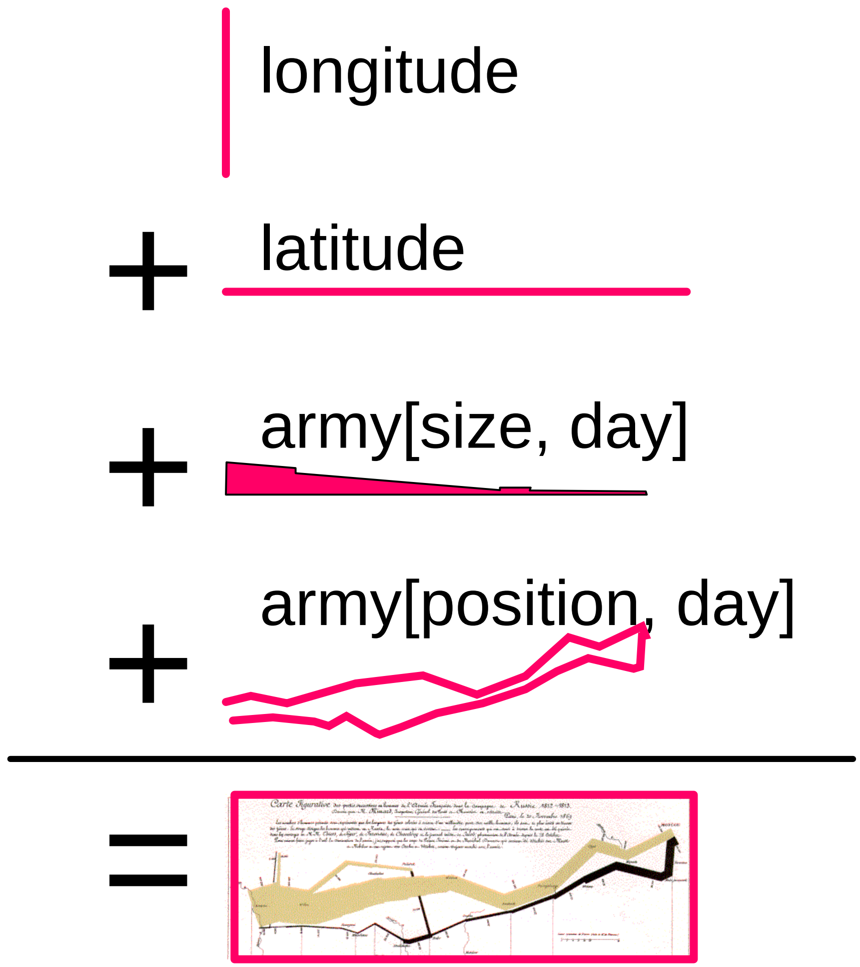
=



Mark Composition



Mark composition

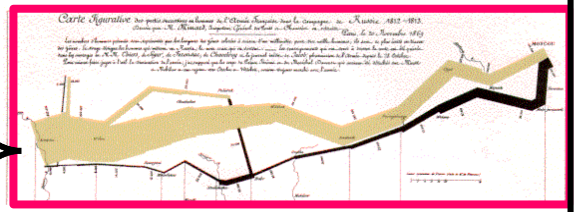


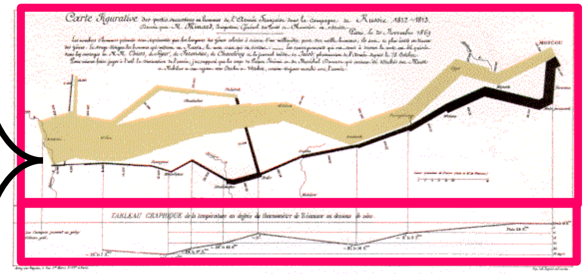
longitude

latitude

army[size, day]

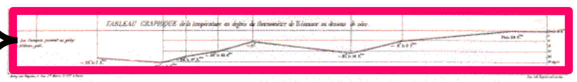
army[position, day]





temperature

time



temp[day]

- Chernoff faces: http://people.cs.uchicago.edu/~wiseman/chernoff/
- Preattentive Processing: http://www.csc.ncsu.edu/faculty/healey/PP/PP.html
- Table Lens: http://www.tablelens.com/

- Discussion point:
  - How many retinal variables can we use in a scatterplot to encode information?

Tool support for automatically designing effective graphical presentations

- APT tool
- Polaris

Mackinlay's APT Environment

- Goal: to develop an application independent presentation tool that automatically designs effective graphical presentations (such as bar charts, graphs, scatter plots) of relational information
- Approach: Graphical presentations are "sentences of graphical languages"
- Mackinlay's work was the first attempt on automatic design

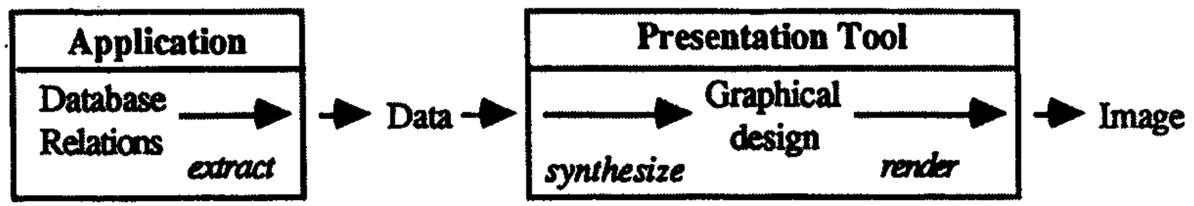


Fig. 1. A linear model for generating presentations. This simplified model, which does not include feedback loops that are required for difficult design problems, describes the fundamental process of generating a graphical presentation. A graphical design synthesized by a presentation tool describes the basic structure and meaning of a graphical presentation. The rendering process fills in the details that are required to form the final image.

Mackinlay's APT Environment (2)

- The tool extracts information from the database, analyzes it and suggests a graphic design
- 'A graphic design' is an abstract description of an image that indicates the graphical techniques (e.g. color, position) that are used to encode information
- But graphic design criteria must be codified before the presentation tool can synthesize effective designs
- The graphical presentation problem is to synthesize a graphical design that expresses a set of relations and their structural properties effectively

Fig. 8. Bertin's graphical objects and graphical relationships.

Marks:	Points, lines, and areas
Positional:	1-D, 2-D, and 3-D
Temporal:	Animation
Retinal:	Color, shape, size, saturation, texture, and orientation

- A problem: Present the price and mileage relations, omit the car details

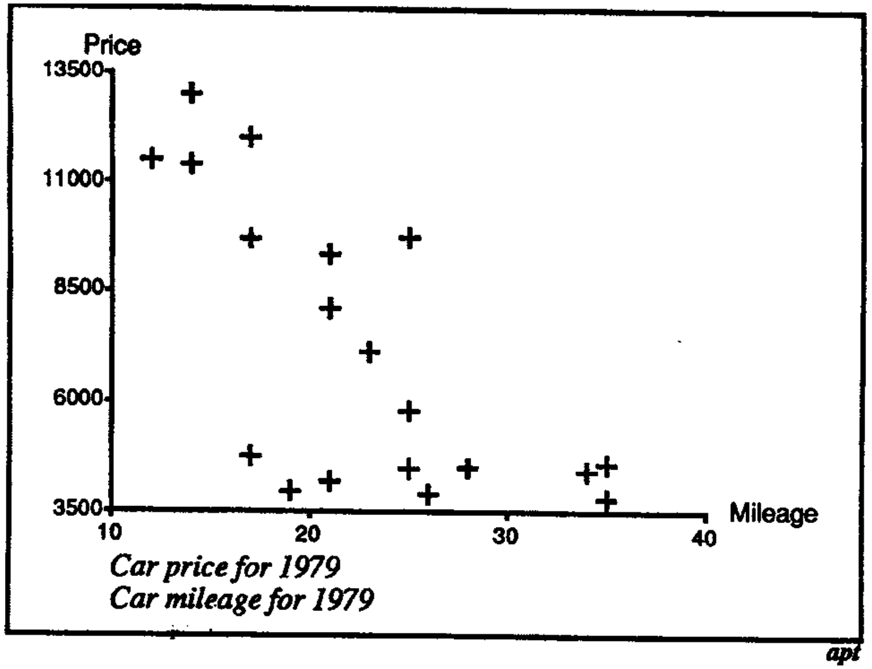


Fig. 5. Scatter plot of the price/mileage input. The graphical design for this image is in Figure 4. The design expresses the relations only if the application permits the details about the cars to be omitted. The *apt* in the lower right corner indicates that APT designed and rendered this diagram.

- But if the car details are needed what are the options?

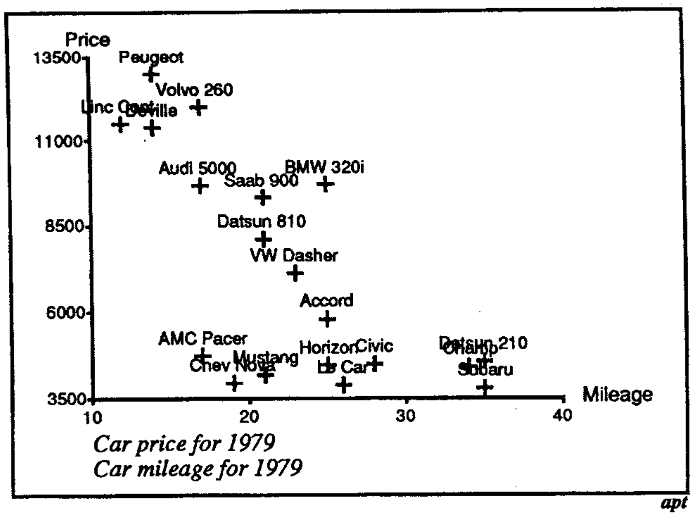


Fig. 6. Labeled scatter plot for the price/mileage input. Although a more sophisticated rendering could avoid the overlapping of the labels, two basic problems of a labeled scatter plot design reduce its effectiveness. First, labels make it difficult to perceive the positions of the points. Second, a given label is difficult to find.

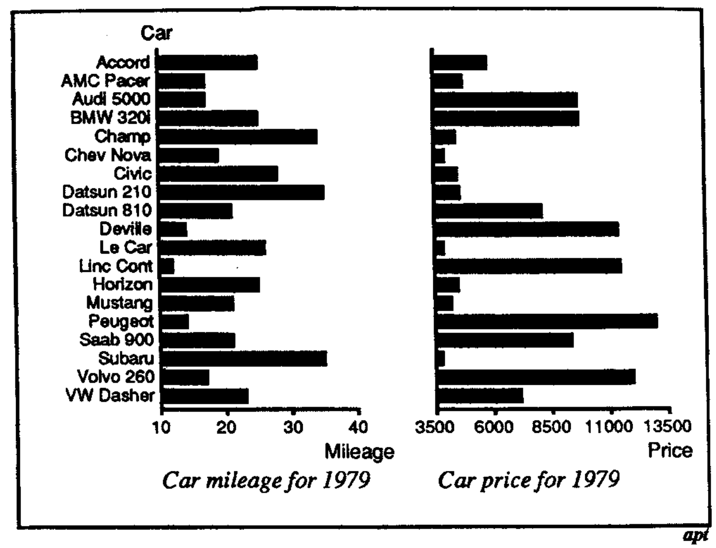


Fig. 7. Aligned bar chart for the price/mileage input. This diagram shows the detailed properties of the cars better than a scatter plot. However, the general relationships are not so easy to see.

Criteria in MacKinlay's framework

- Expressiveness criteria:
  - Determines whether a graphical language can express the desired information and *only* the desired information
- Effectiveness criteria:
  - Determines whether the desired information exploits the capabilities of the output medium and the human visual system

Incorrect use of a bar chart – Violates expressiveness criteria

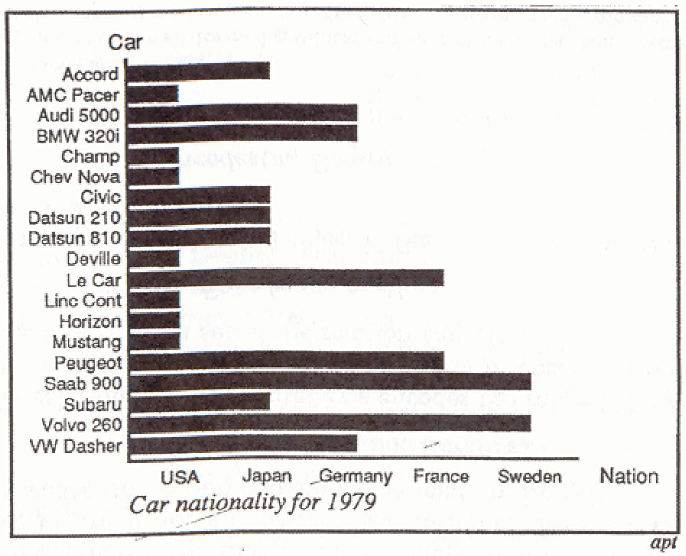


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

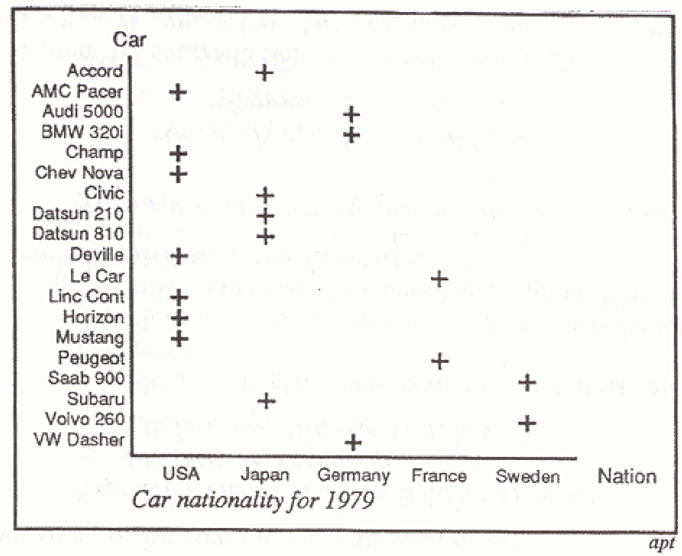


Fig. 12. Correct use of a plot chart for the *Nation* relation. Since bar charts encode ordered domain sets, plot charts are conventionally used to encode nominal domain sets. The ordering of the labels on the axes is ignored.

Is this ok?

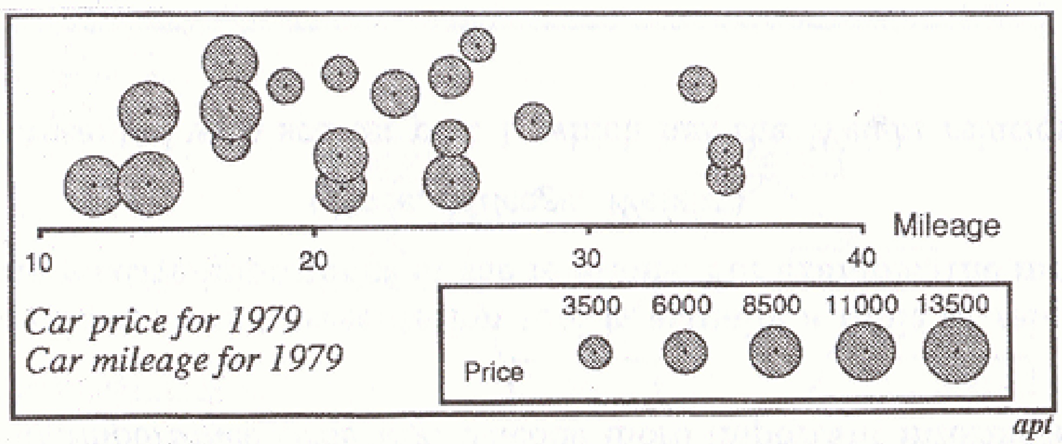


Fig. 13. Area/position presentation of the *Price* and *Mileage* relations. The vertical positioning of the marks reduces the chance that a mark is covered. This technique is called jittering; the vertical positioning does not encode any information.

Effectiveness criteria

- Effectiveness criteria:
  - Determines whether the desired information exploits the capabilities of the output medium and the human visual system
  - But there is no empirically verified theory of human perceptual capabilities… so we build theories based on what is known…

Ranking of perceptual tasks

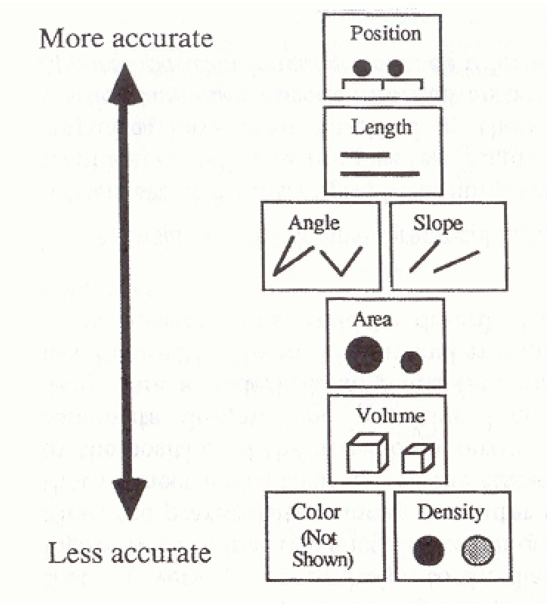


Fig. 14. Accuracy ranking of quantitative perceptual tasks. Higher tasks are accomplished more accurately than lower tasks. Cleveland and McGill empirically verified the basic properties of this ranking.

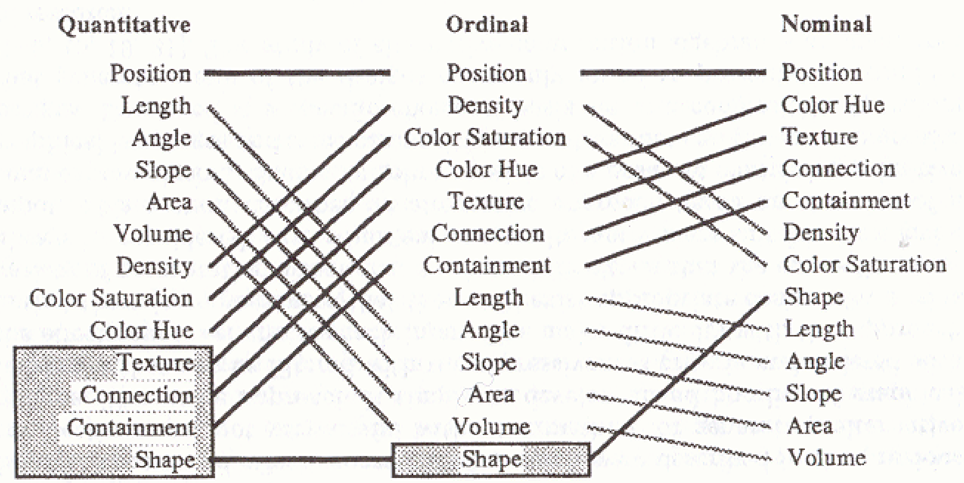


Fig. 15. Ranking of perceptual tasks. The tasks shown in the gray boxes are not relevant to these types of data.

Encoding Technique	Primitive Graphical Language
Single-position	Horizontal axis, vertical axis
Apposed-position	Line chart, bar chart, plot chart
Retinal-list	Color, shape, size, saturation, texture, orientation
Map	Road map, topographic map
Connection	Tree, acyclic graph, network
Misc. (angle, contain, . . .)	Pie chart, Venn diagram, . . .

Fig. 22. A basis set of primitive graphical languages.

Length and appropriateness of the retinal variables…

- It is often ok to confuse 2 quantitative values close in value, but usually a bigger mistake to misinterpret ordinal or nominal values
- Is area a suitable variable for nominal data?

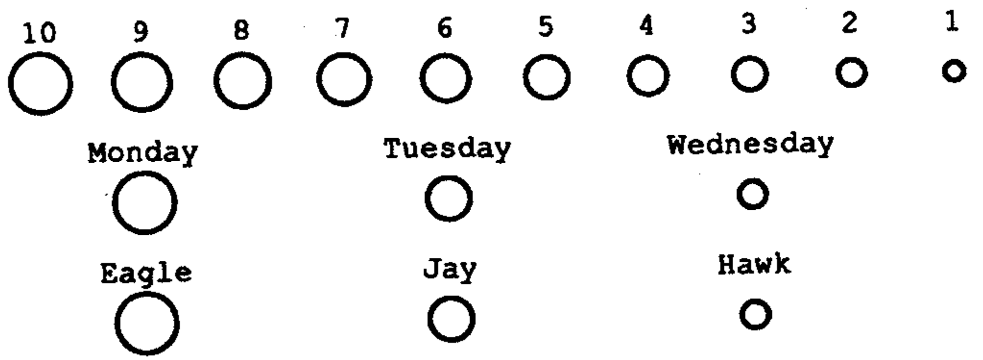


Fig. 16. Analysis of the area task. The top case shows that area is moderately effective for encoding quantitative information. The middle case shows that it is possible to encode ordinal information as long as the step size between areas is large enough so that the values are not confused. The bottom case shows that it is possible to encode nominal information, but people may perceive an ordinal encoding.

Which representations are better?

- Which scatterplot is preferred?

	Price	Mileage
Scatter plot	position	position
Area/Position	area	position

Fig. 17. Comparison of perceptual tasks for the price/mileage designs.

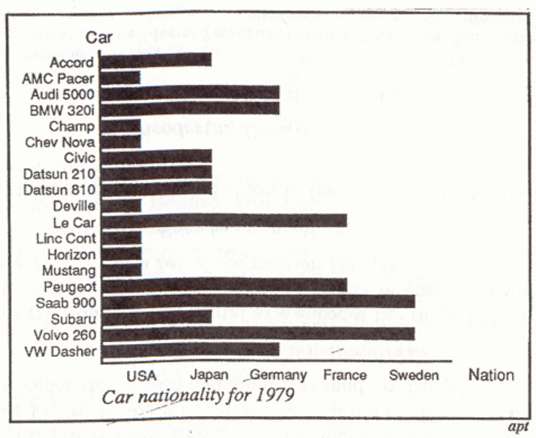


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

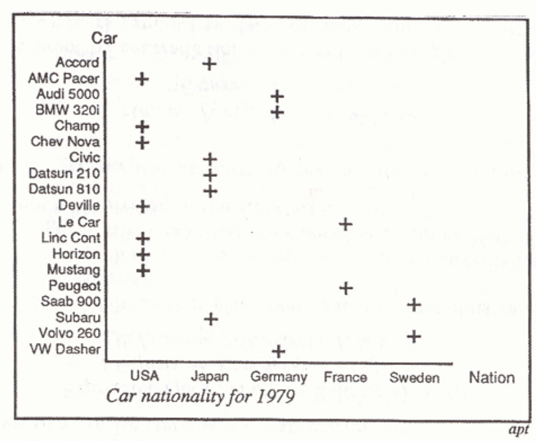


Fig. 12. Correct use of a plot chart for the *Nation* relation. Since bar charts encode ordered domain sets, plot charts are conventionally used to encode nominal domain sets. The ordering of the labels on the axes is ignored.

Which representations is better? (2)

- Which scatterplot is preferred?

	Price	*Mileage*	*Weight*
Scatter plot 1	position	position	area
Scatter plot 2	area	position	position

Fig. 18. Example of designs not ordered by the effectiveness ranking.

- General principle: encode more important information more effectively (so if price was more important use position rather than area for price)

Combining position and network representations

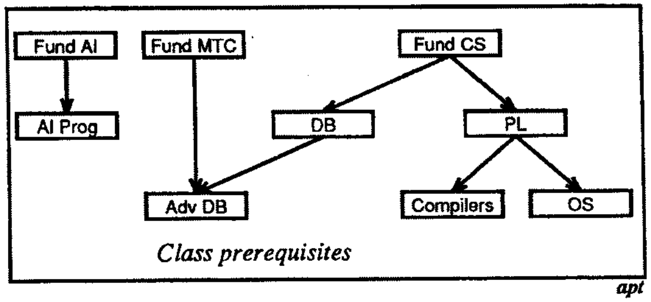


Fig. 20. Network presentation for the prerequisite relation.

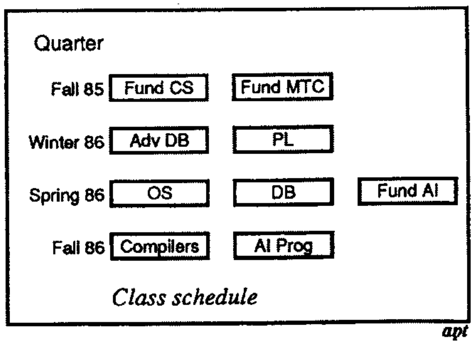


Fig. 21. Vertical-axis presentation for the schedule relation.

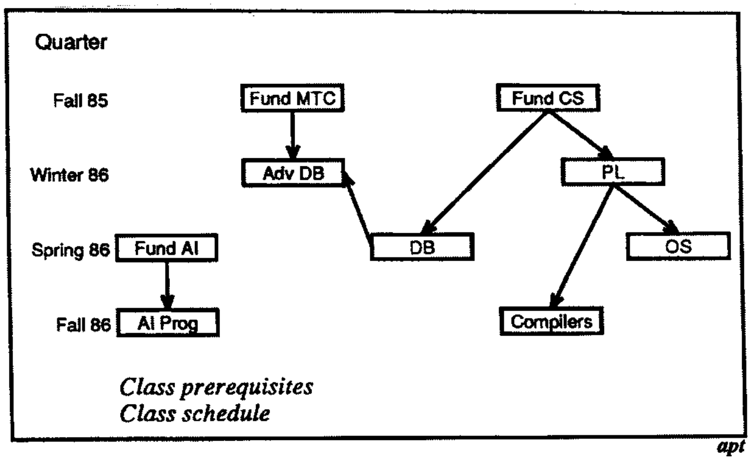


Fig. 19. Composite presentation for the prerequisite and schedule relations. The links encode the prerequisite relationships between computer science classes. The position on the vertical axis encodes the scheduling of the classes. Note that the advanced database class is scheduled before its prerequisite.

Expressiveness of retinal techniques

	Nominal	Ordinal	Quantitative
Size	–	•	•
Saturation	–	•	•
Texture	•	•	
Color	•	*	
Orientation	•		
Shape	•		

Fig. 25. Expressiveness of retinal techniques. The – indicates that size and saturation should not be used for nominal measurements because they will probably be perceived to be ordered. The * indicates that the full color spectrum is not ordered. However, parts of the color spectrum are ordinally perceived [23].

Encoding more than 3 variables….

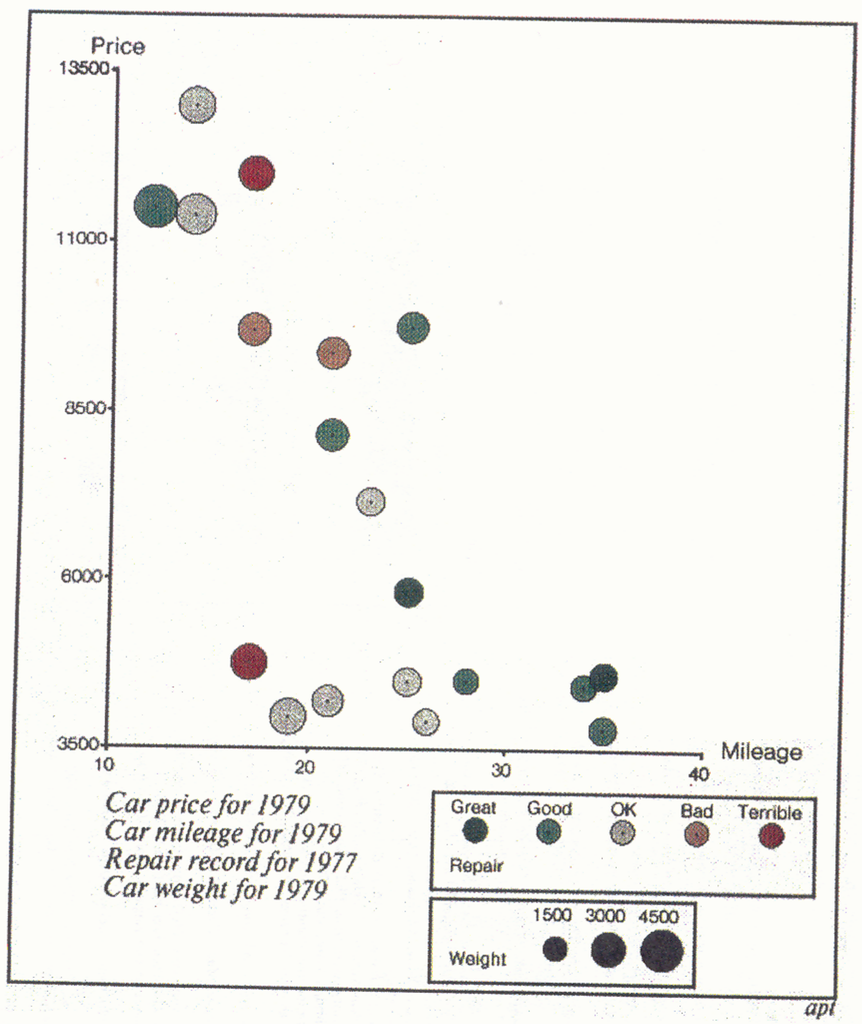


FIGURE 30

Color scatter plot for four automobile relations. The design expresses the relations only if the application permits the details about the cars to be omitted.

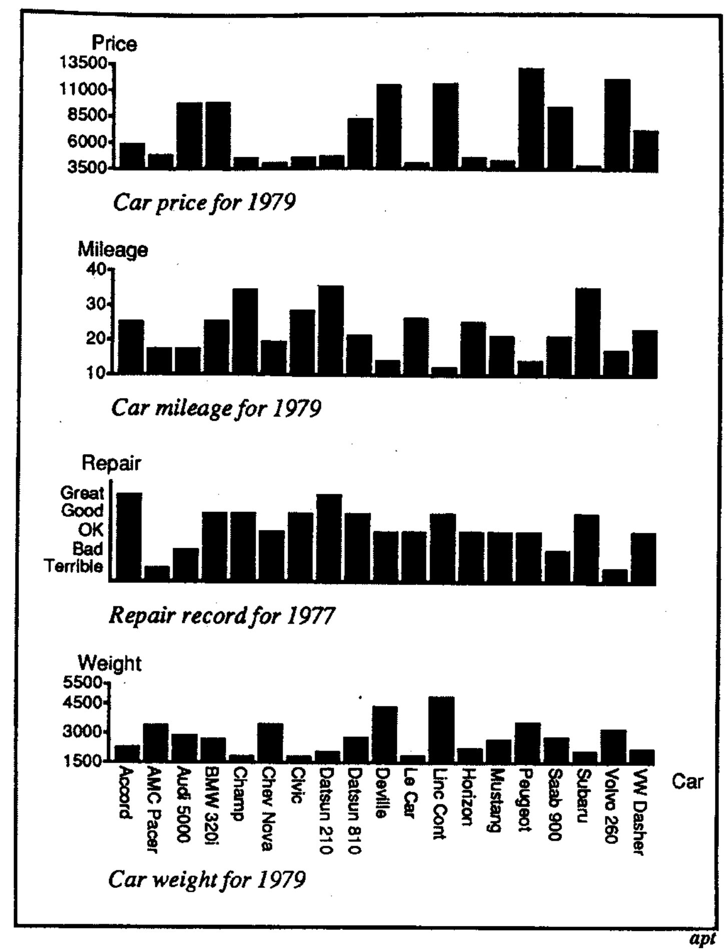


Fig. 31. Aligned bar chart of automobile data. This diagram shows the details about the car domain set. However, the general relationships are not so easy to see as in the scatter plot design.

Polaris system

Visualization of Multi-dimensional
Relational Databases

Motivation

- Over the past few years, multi-dimensional large databases have become very common in a variety of domains/applications
- A major challenge is to extract meaning, discover structure, find patterns and derive causal relationships
- Exploratory analysis is one of:
  - Hypothesis
  - Experiment
  - Discovery
  - Iterate – analysts need to rapidly change both *what* data they are viewing and *how* they are viewing that data

Polaris

- Polaris creates visualizations using a set of tables based on fields of the database
- The use of tables to organize multiple graphs is a well known technique by statisticians in their analysis of data
- Each table has layers and panes (each pane may have a different graphic)

Relational databases

- In a relational database: Each row in a table corresponds to a basic entity or fact and each column represents a property of that entity
- A row is referred to as a tuple, and a column is a field
- A single relational database will contain many heterogeneous but interrelated tables
- Fields in a database may be nominal, ordinal or quantitative
- Fields in a database can be partitioned in 2 types:
  - Dimensions (e.g. name or type of a product, independent variables)
  - Measures (e.g. price or size, dependent variables)
- Polaris assumed that all nominal fields are dimensions and all quantitative fields are measures

Requirements for visualizing multidimensional databases

1. Data dense displays
2. Multiple display types
3. Exploratory interface

- Polaris addresses these requirements by providing an interface for rapidly and incrementally generating table-based displays
  - Each table axis may have multiple nested dimensions
  - Each table entry (or pane) contains a set of records that are visually encoded as a set of marks to create a graphic

Why tables?

- Multivariate – multiple dimensions of the data can be explicitly encoded in the structure of the table
- Comparative – tables generate “small multiple displays of information” [Tufte] which can be compared, exposing patterns and trends across dimensions of the data
- Tables are familiar…

Interface

- Interface approach is to drag and drop fields from the database schema onto shelves throughout the display
- Interface supports “brushing”
- Some screenshots…

Polaris User Interface

Axis Shelves:
The fields placed here determine the structure of the table and the types of graphs in each table pane.

Database Schema:
The user drags fields from the database schema to shelves to define the visual specification.

Layer Shelf:
The fields placed here determine how records are partitioned into layers.

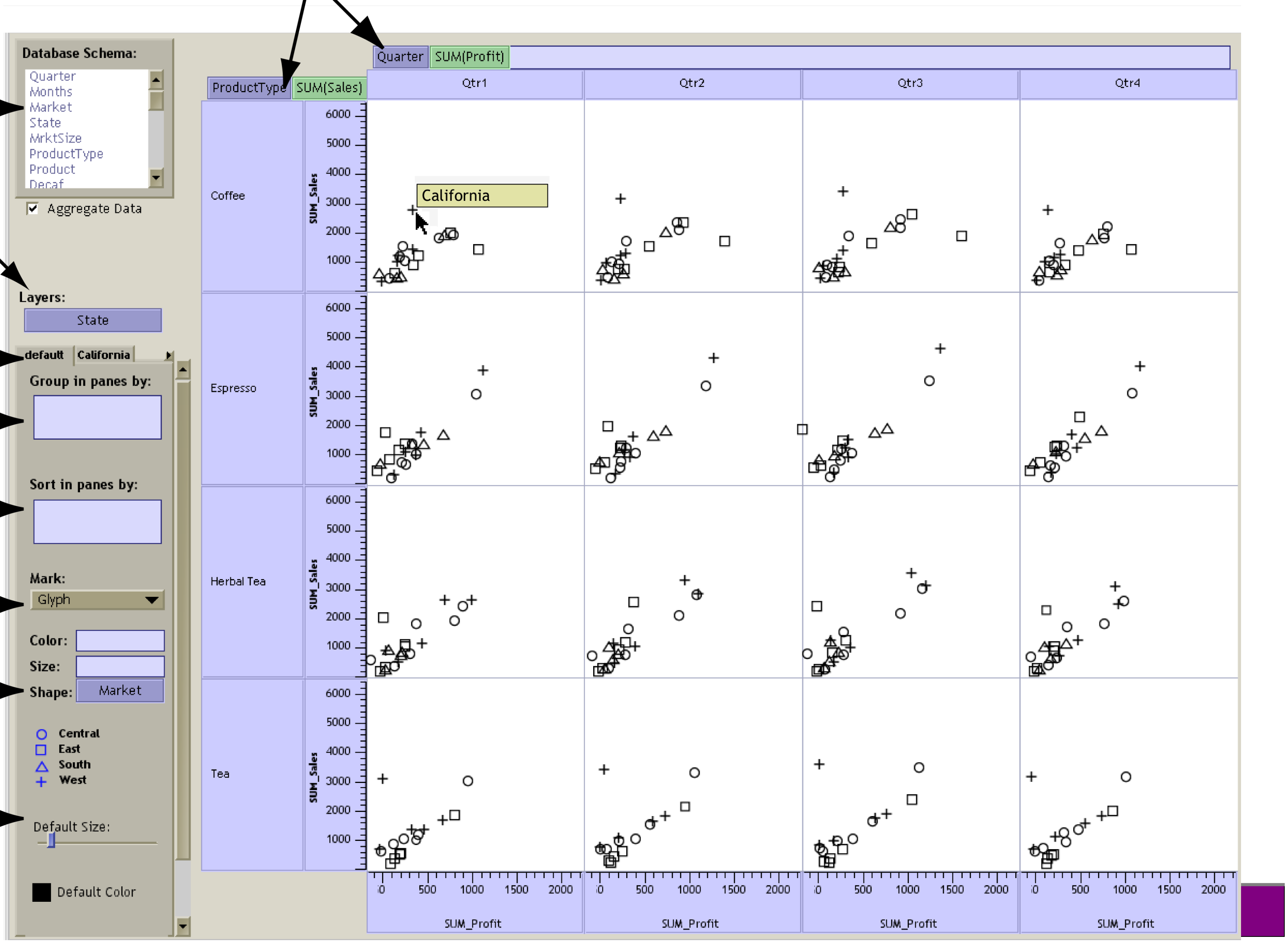
Layer Tabs:
Each layer has its own tab; different transformations and mappings can be specified for each layer.

Grouping and Sorting Shelves:
The fields placed here determine how records are grouped and sorted within the table panes.

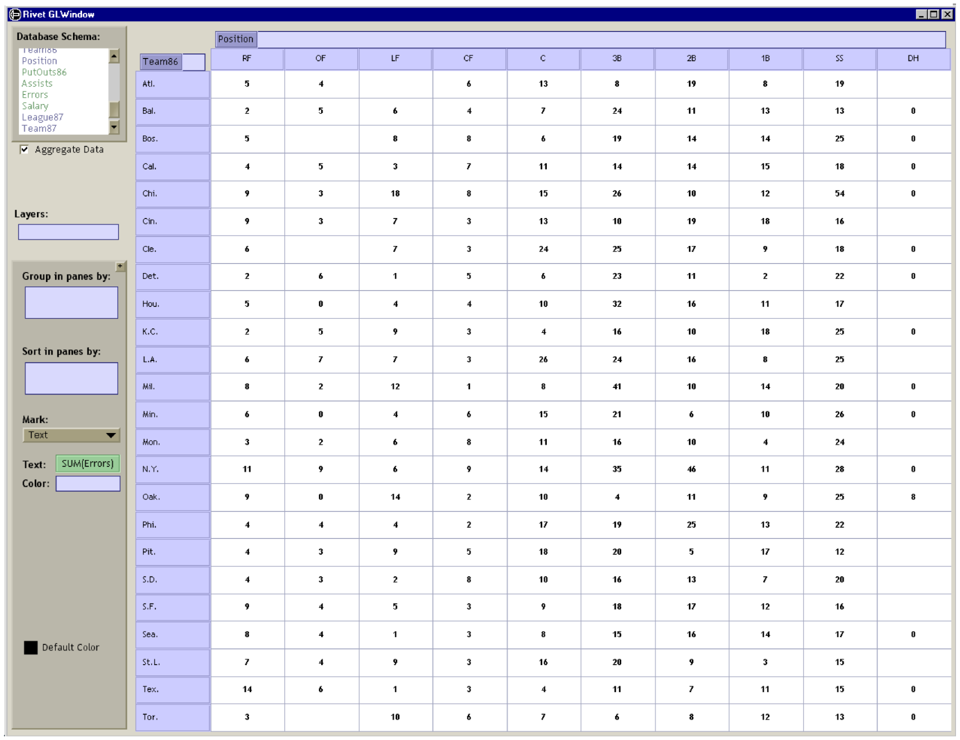
Mark Pulldown:
Relations in each pane are mapped to marks of the selected type.

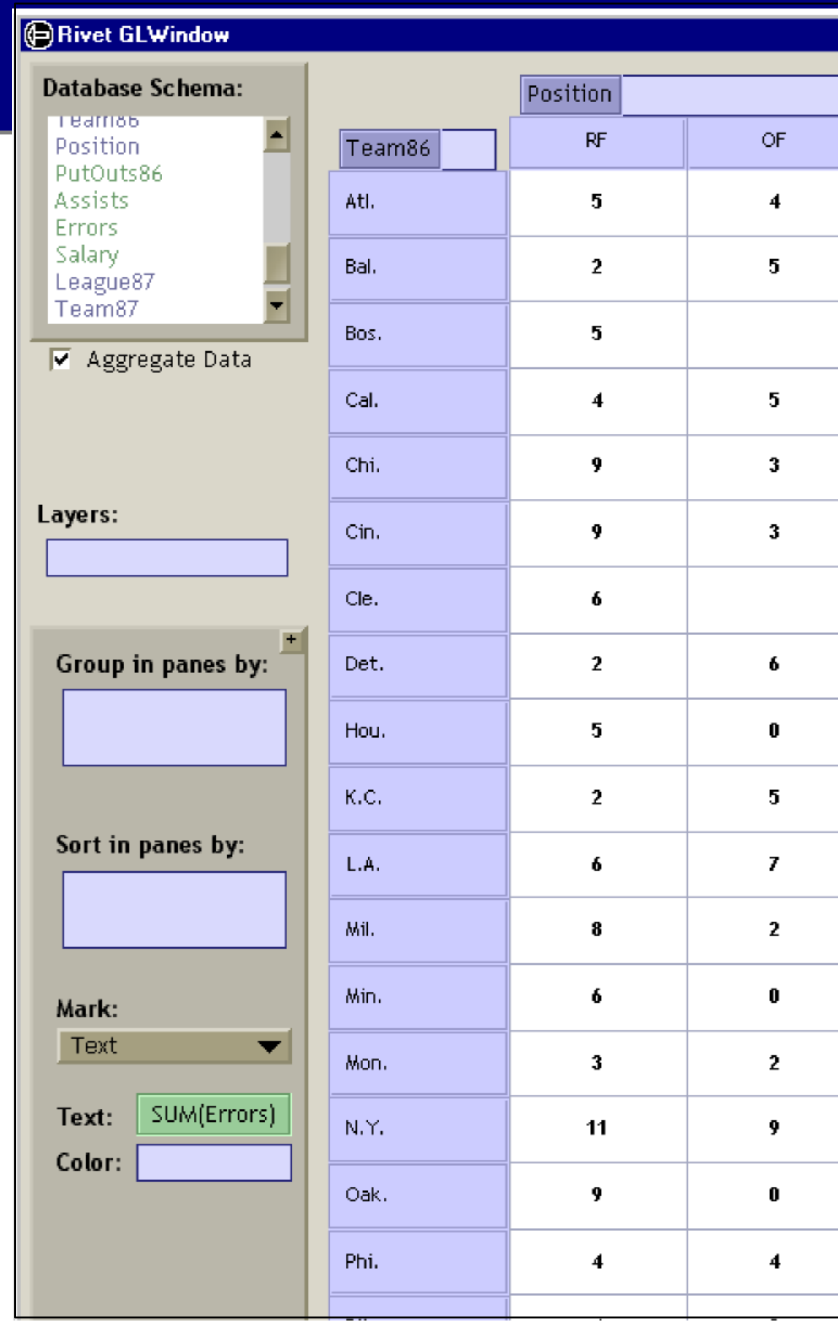
Retinal Property Shelves:
The fields placed here determine how data is encoded in the retinal properties of the marks.

Legends:
Legends enable the user to see and modify the mappings from data to retinal properties.

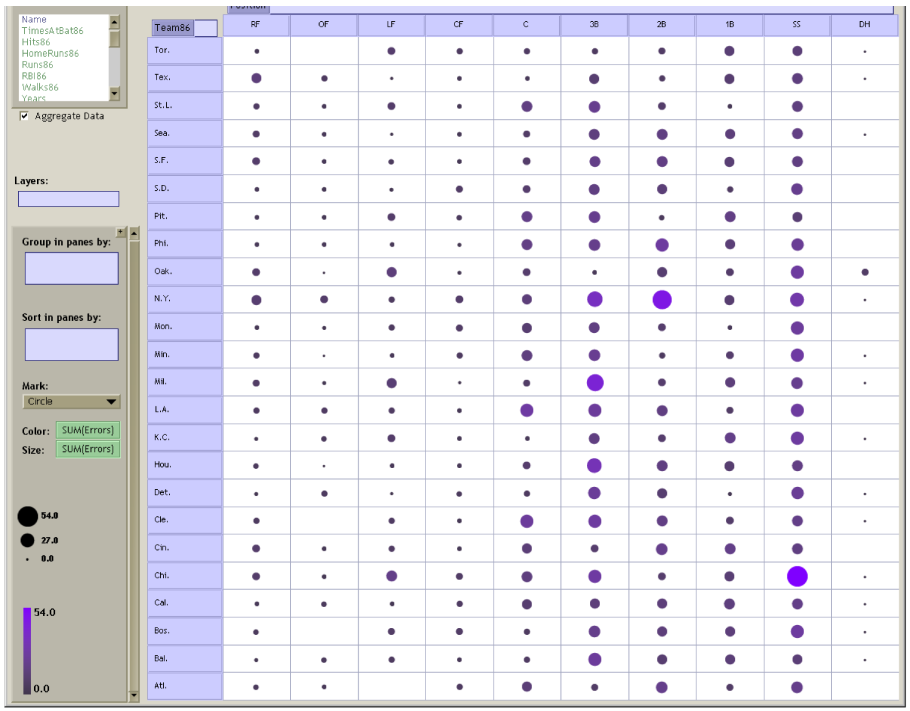


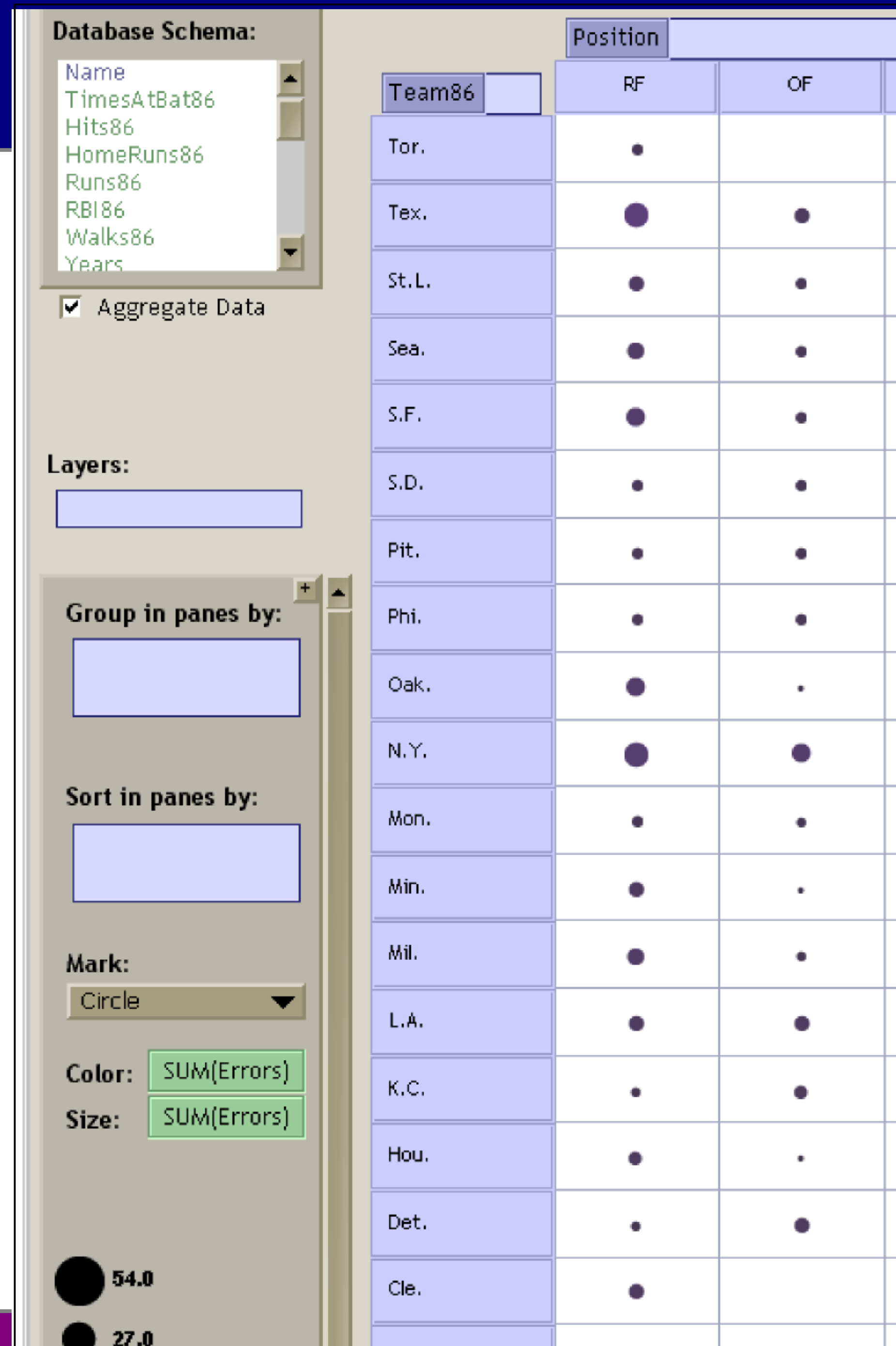
Ordinal-Ordinal



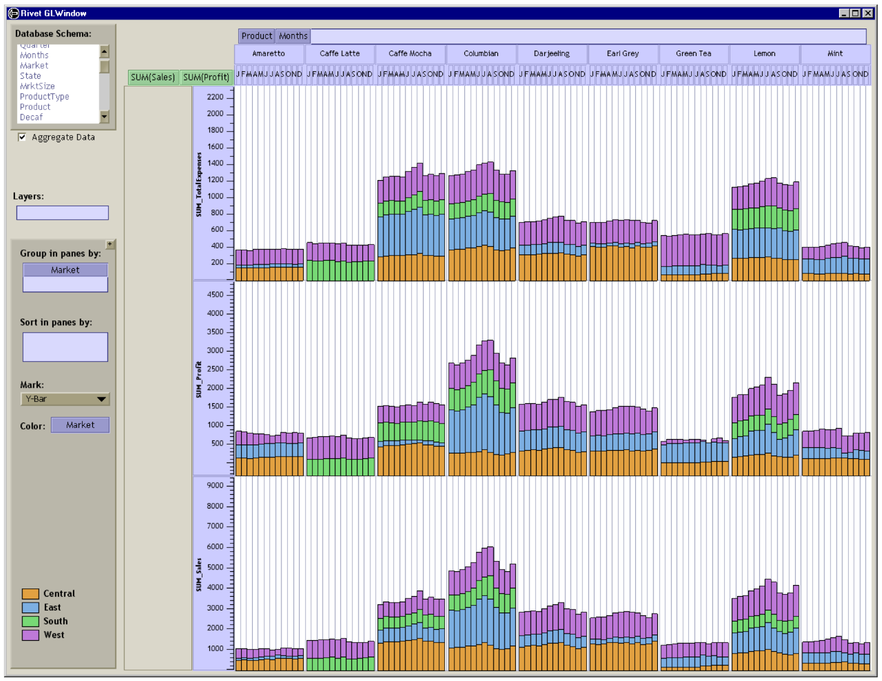


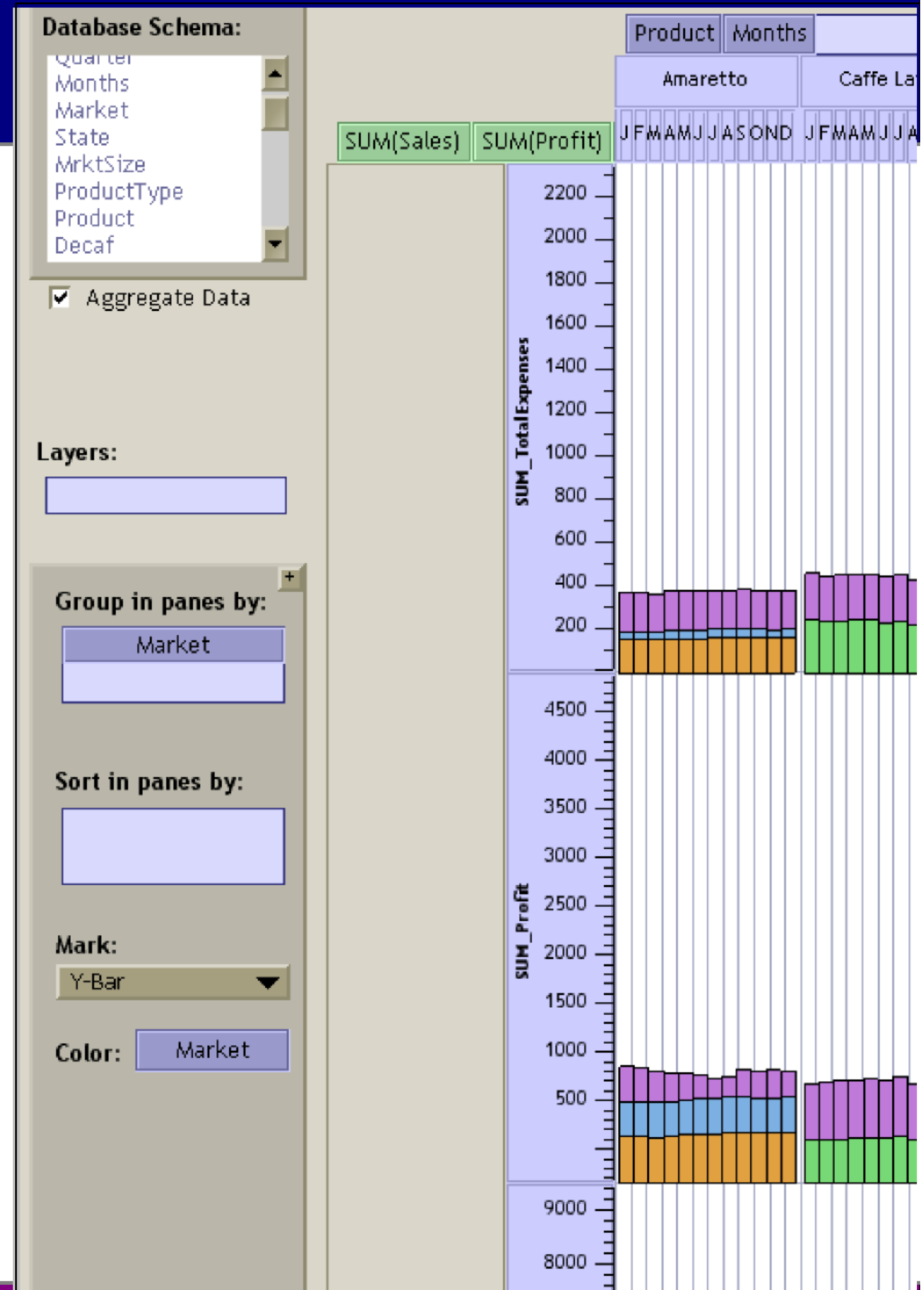
Ordinal-Ordinal



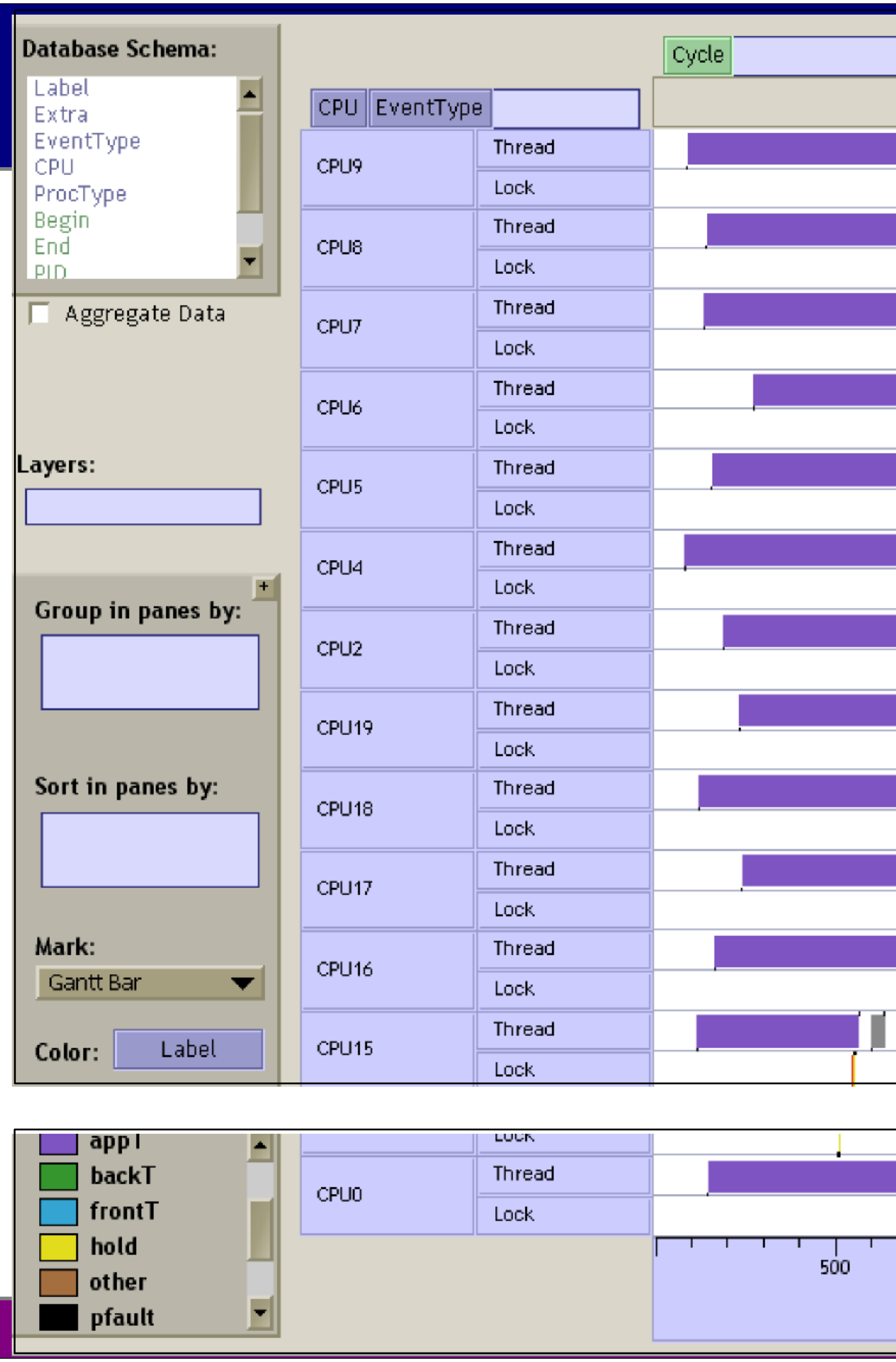


Ordinal-Quantitative

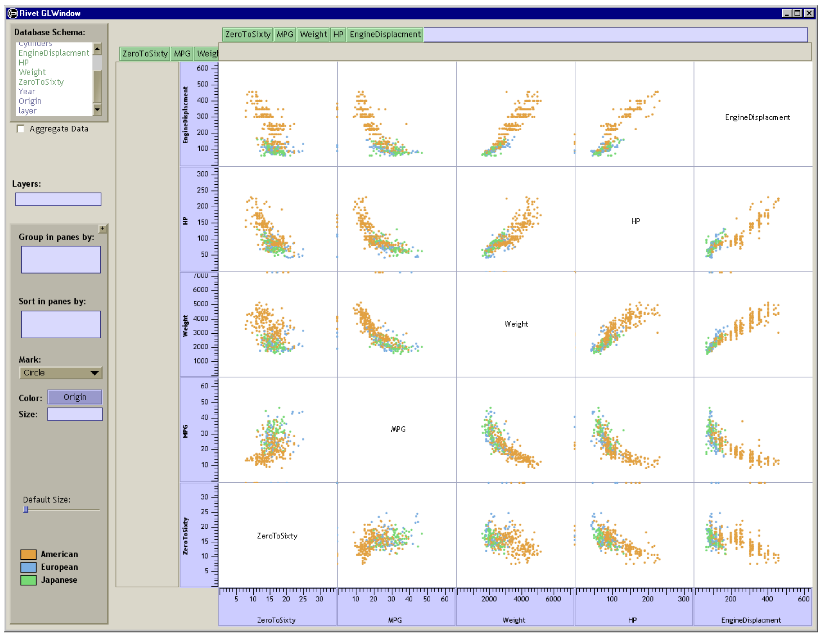


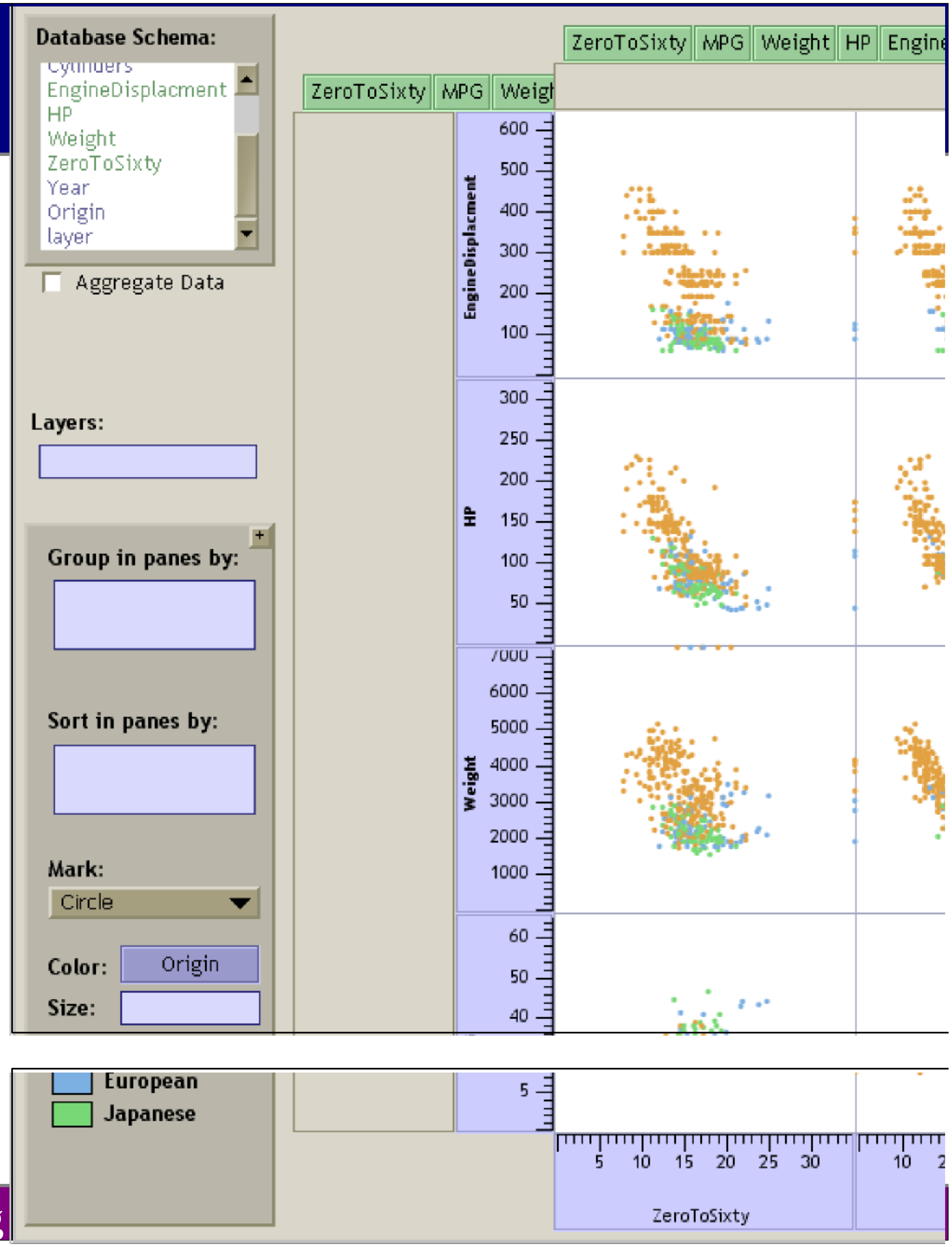


Ordinal-Quantitative

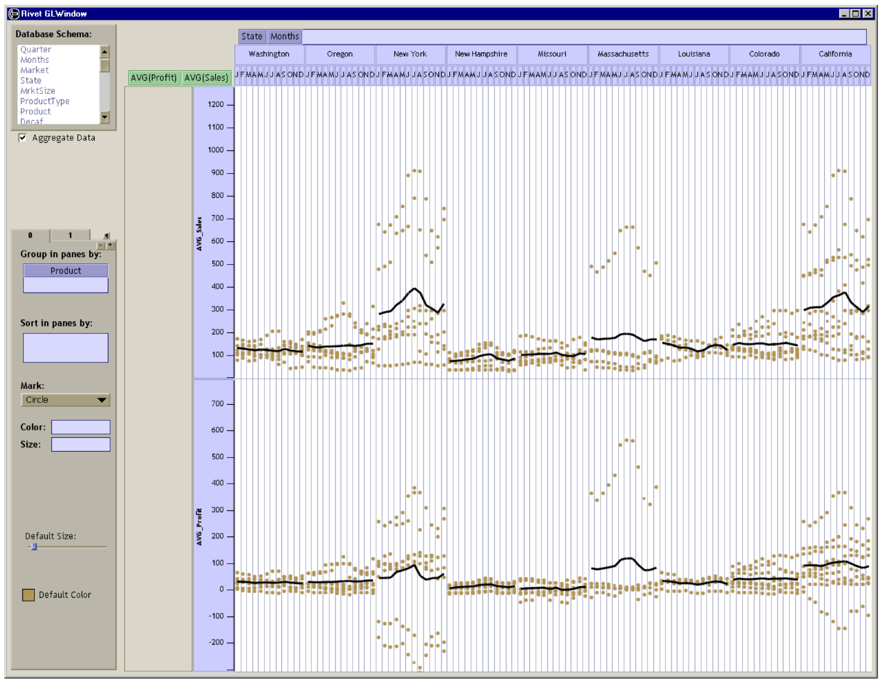


Quantitative-Quantitative





Quantitative-Quantitative



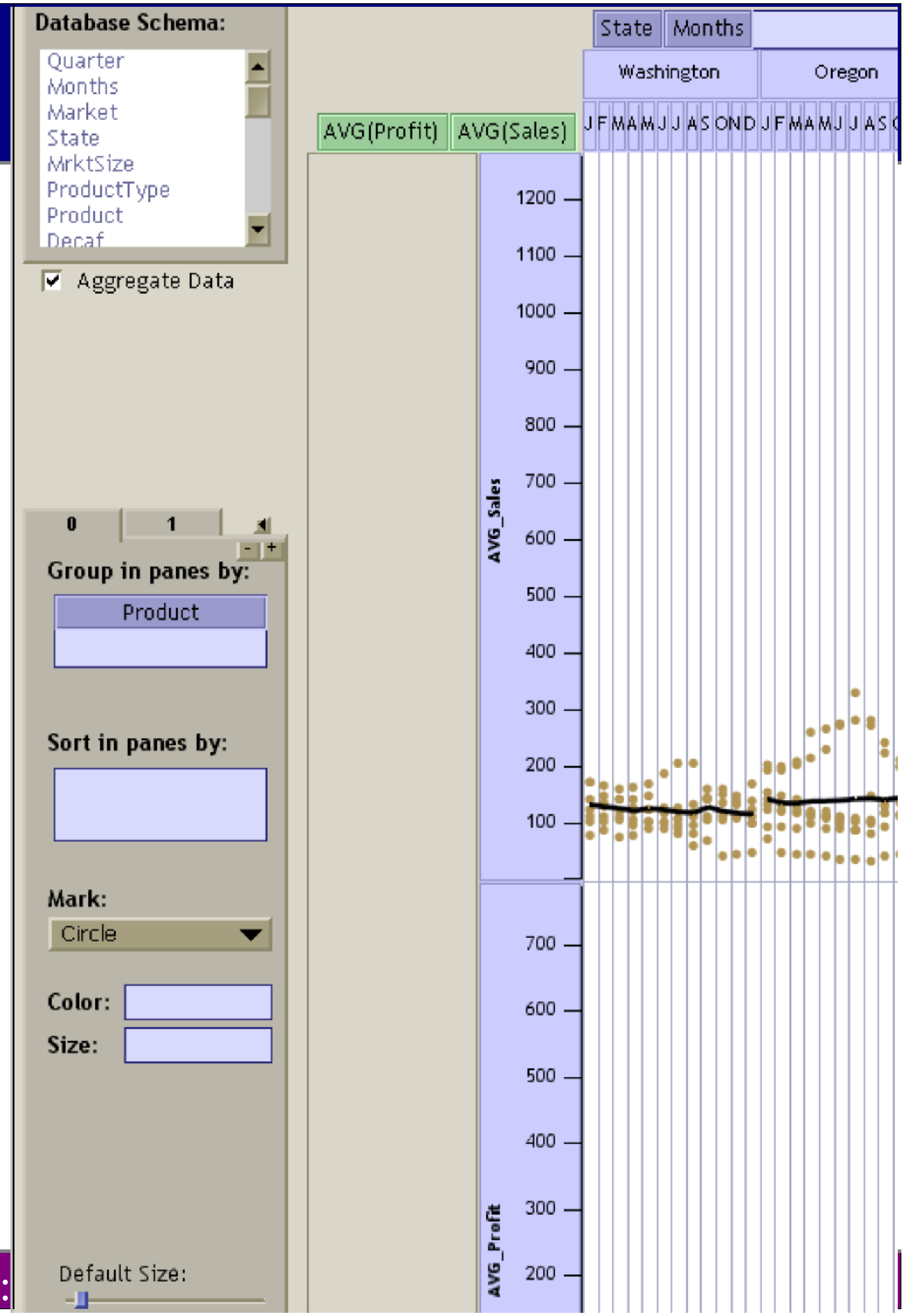


Table Algebra: Ordinal

Ordinal fields: Quarter, Months, Product   Quantitative fields: Profit, Sales

O = Quarter = {Qtr1, Qtr2, Qtr3, Qtr4} = Qtr1 + Qtr2 + Qtr3 + Qtr4:

Qtr1	Qtr2	Qtr3	Qtr4

O + O = Quarter + Product = {Qtr1, Qtr2, Qtr3, Qtr4, Coffee, Espresso, Herbal Tea, Tea}:

Qtr1	Qtr2	Qtr3	Qtr4	Coffee	Espresso	Herbal Tea	Tea

O x O = Quarter x Product = {(Qtr1,Coffee), (Qtr1,Espresso), (Qtr1,Hebral Tea), (Qtr1, Tea), (Qtr2, Coffee) ... (Qtr4, Tea)}:

Qtr1				Qtr2				Qtr3				Qtr4			
Coffee	Espresso	Herbal Tea	Tea	Coffee	Espresso	Herbal Tea	Tea	Coffee	Espresso	Herbal Tea	Tea	Coffee	Espresso	Herbal Tea	Tea

O/O = Quarter / Month = {(Qtr1,Jan), (Qtr1,Feb), (Qtr1,Mar), (Qtr2, Apr), (Qtr2, May) ... (Qtr4, Dec)}:

Qtr1			Qtr2			Qtr3			Qtr4		
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec

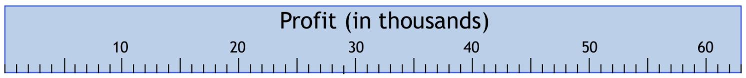
\+ Concatenate

x Cross

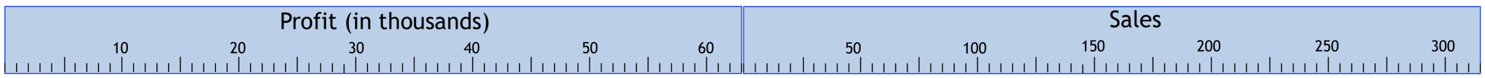
/ Nest

Table Algebra: Quantitative

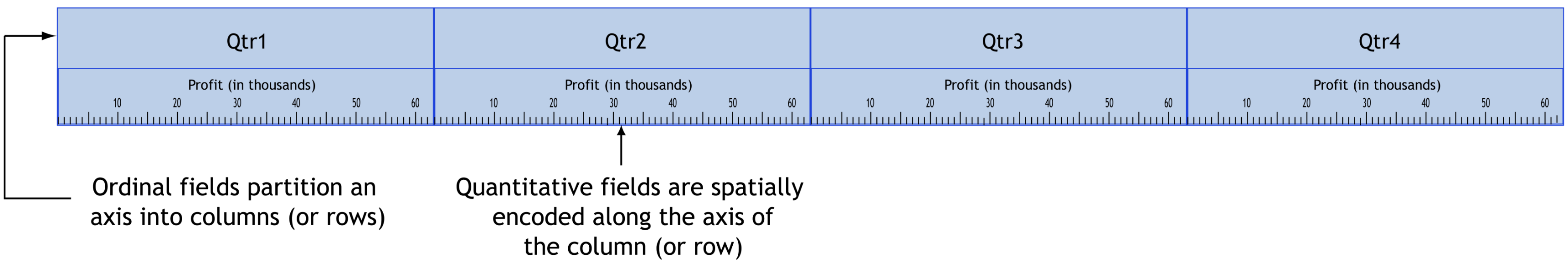
Q = Profit = {Profit}:



Q + Q = Profit + Sales = {Profit, Sales}:



O×Q = Quarter × Profit = {(Qtr1,Profit), (Qtr2, Profit), (Qtr3, Profit), (Qtr4, Profit)}:



+ Concatenate

x Cross

/ Nest

Summary:

- Visualizaion barrier on 3 or 4 dimensions of data…
- But interaction and permutations can be used to increase the number of relationships we can see
- Small multiple displays of information that are linked can reveal important relationships
- Appropriateness of the retinal variables to nominal, ordinal and quantitative data
- Some images can be automatically generated using tools such as Polaris and other tools from Visual Inxight
- Example: http://www.inxight.com/products/st_viewer/