Information Visualization

Text: Information visualization, Robert Spence, Addison-Wesley, 2001

What Visualization?

- Process of making a computer image or graph for giving an insight on data/information
  - Transforming abstract, physical data/information to a form that can be seen
  - Interpreting in visual terms or putting into visual forms (i.e., into pictures)
- Cognitive process
  - Form a mental image of something -- an internal image
  - Internalize an understanding
- What is information?
  - Items, entities, things which do not have a direct physical relevance, e.g, stock trends, baseball statistics, car attributes, train routes, text

Topics

- Internal models
  - Visualization goes on in mind and results in something called a mental model or internal model
- Data representation
  - Visualization represents abstract things (data/information) in someway graphically
- Interaction and exploration
  - Visualization allows one to extract useful information by interacting with and exploring data/information graphically
- Presentation
  - Visualization deals with problem of displaying too much data onto a small screen
- Connectivity
  - Visualization deals with cases of connectivity (networks, trees)

Internal Models

Internal Model - Definition

- We use an internal model that is generated based on what is observed

- The internal model is called a *cognitive map*
  - You just don't have one big one
  - You have a large number of these for all different kinds of things

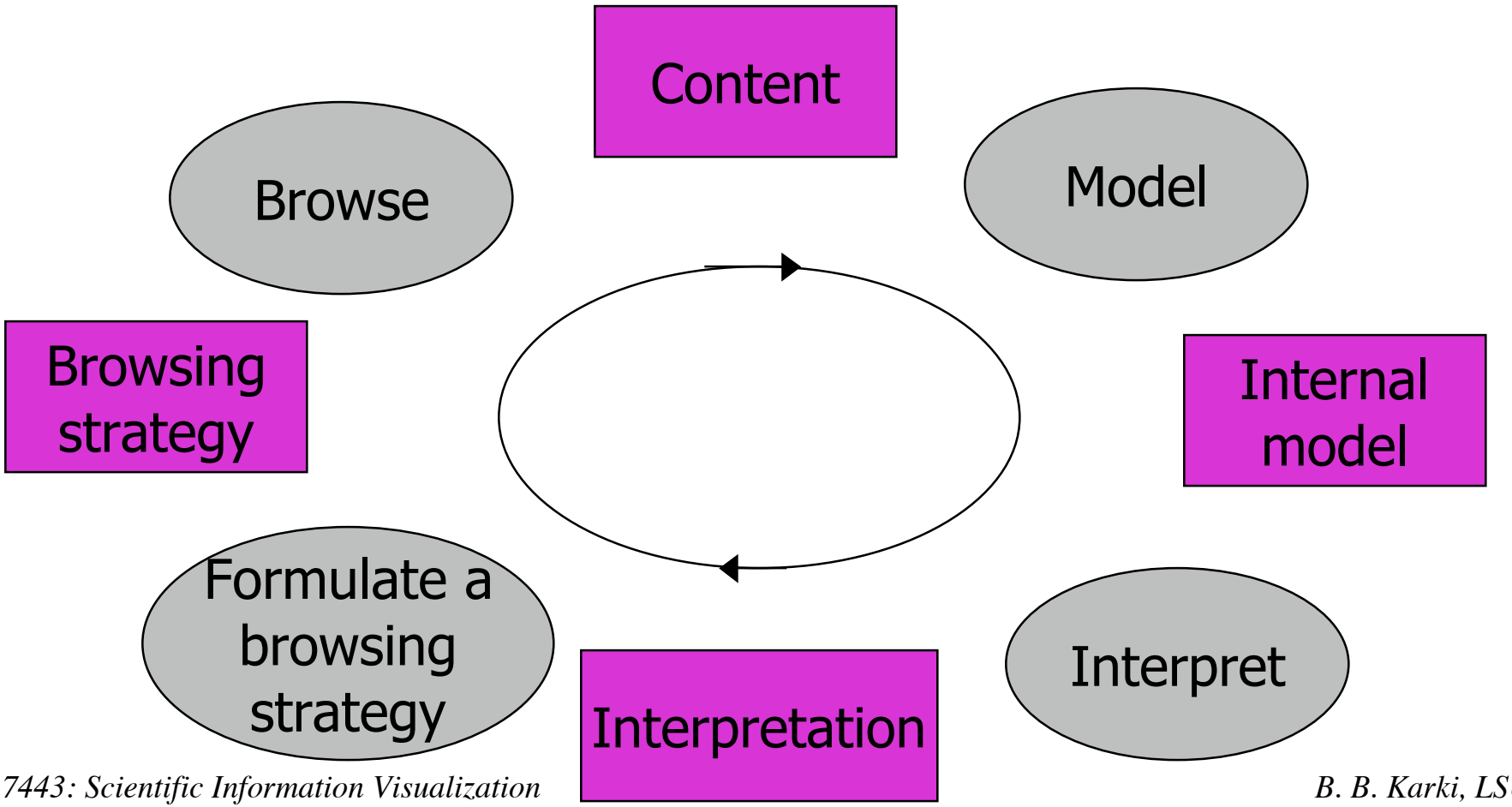
  Collection of cognitive maps --> *Cognitive college*

- London underground railway system:
  - If you are in Imperial College for sometime, you will have some existing internal model of the system
  - To make short journeys from the College, you need not to look at map
  - But less familiar journeys, you may glance at map to be sure

    Refines your internal model, clarifying items and extending it
  - Note that it's still not perfect, no internal model ever is

Navigation: Framework

- Navigation of information space -- a framework for the human activity -- creation and interpretation of an internal model



Navigation: Explanation

- **Browsing:** An user scans a display to 'see what's there'. It causes registration of content
  - Look at the content on the display
- **Modeling:** The content acquired by browsing is soon integrated to begin forming an internal model
  - Modeling of that pattern seen on the display results in cognitive map
- **Interpretation:** One then interprets the internal model to decide as to how and whether further browsing should proceed
  - Leads to new view that generates an idea for a new browsing strategy
- **Formulation of browsing strategies:** The process can be cognitive (driven by interpretation or a new idea) or perceptual (influenced by what is displayed) determinant
  - Look at the display again

Data Representation

A Data Example

- Students in class

Cases →

Variables ↓	Mary	John	Sally	Peter	….
SSN	138	179	286	843	
Age	20	17	23	19	
GPA	3.5	3.1	2.9	2.5	
Hair	black	red	brown	blonde	
….					

- Individual items are called cases
- Cases have variables (attributes)

Dimensionality

- Dimensions: Number of variables or attributes

- Univariate data - 1 variable

  Car: cost

- Bivariate data - 2 variables

  Car: cost, model

- Trivariate data - 3 variables

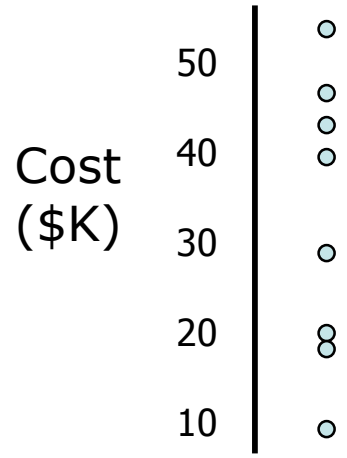
  Car: cost, model, year

- Hypervariate or multivariate data - more than 3 variables

  Car: cost, model, year, make, miles for gallon, no. of cylinders, weight, ….

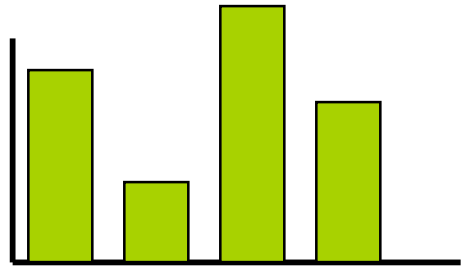
Univariate Data

- Different representations



- In form of points against some scale

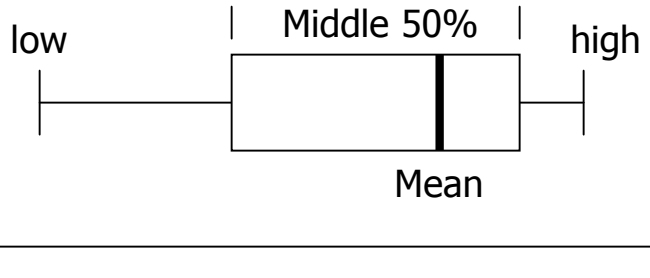
  (points can be labeled)



- In forms of aggregation:

  Histogram

  Tukey box plot

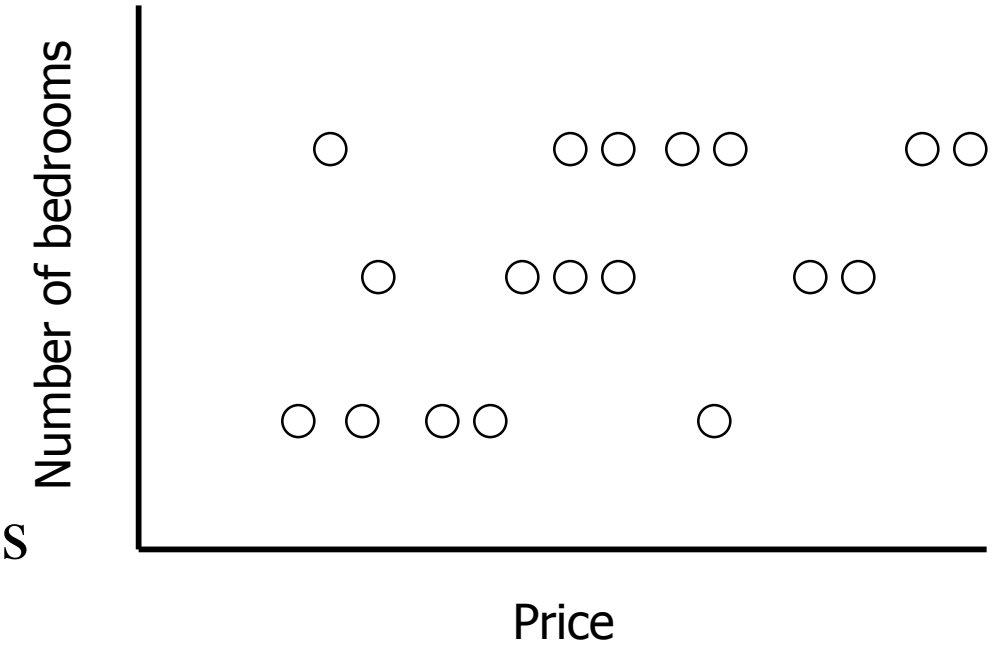


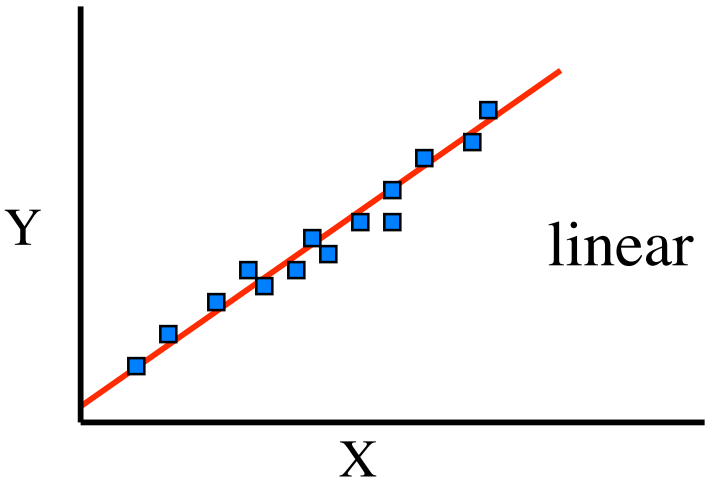
Bivariate Data

- Scatter plot of one variable against other
- In forms of aggregations or groups

  Two histograms

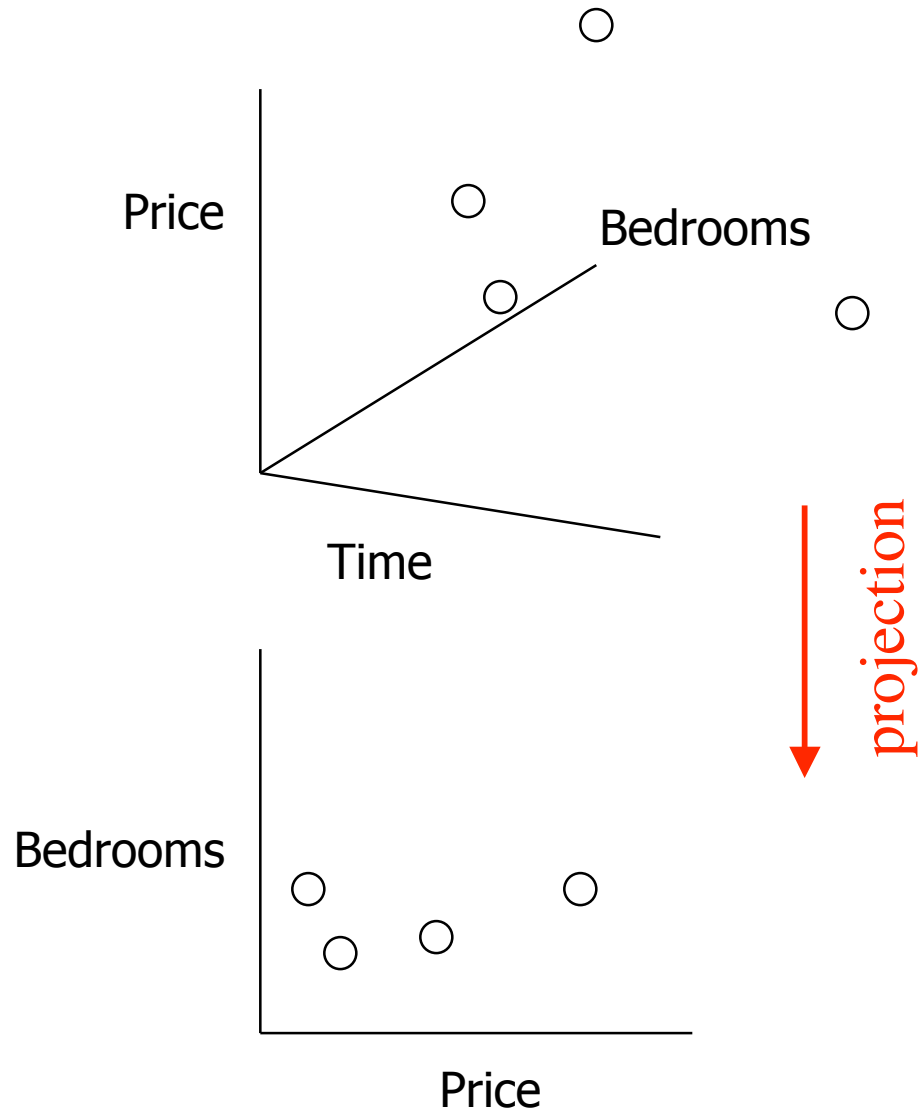
  Two box plots





Trivariate Data

- 3D world in 2D graphic representation
- Scatter plot showing three axes
- Projection onto all pair of axes
  - 3 projections
- Spinplot [Fisherkeller et al. 1974]
  - To allow viewing in any direction



Hypervariate Data

- Hypervariate or multivariate data
- Multiple views
  - Give each variable its own display
  - Use techniques for datasets of 1 - 3 dimensions

    histograms, scatter plots, line graphs
- Interrelationships between many variables shown simultaneously

  Starplot

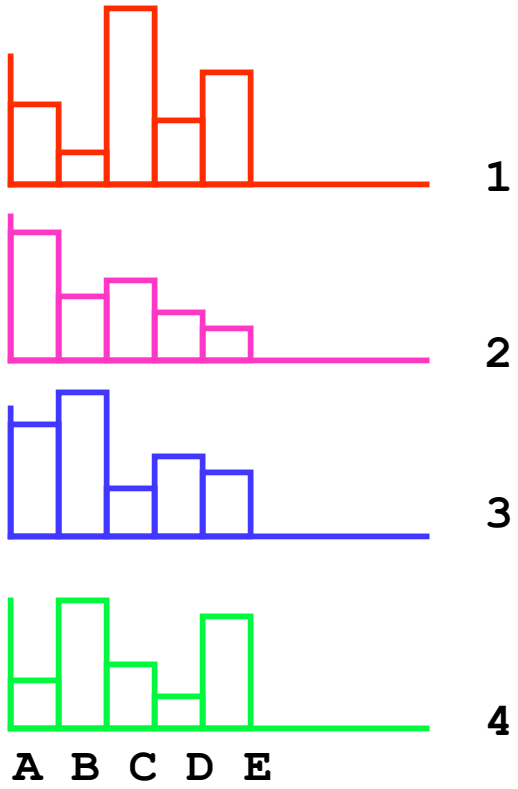
  Star coordinates

  Parallel coordinates

  Hyperbox

Multiple Views

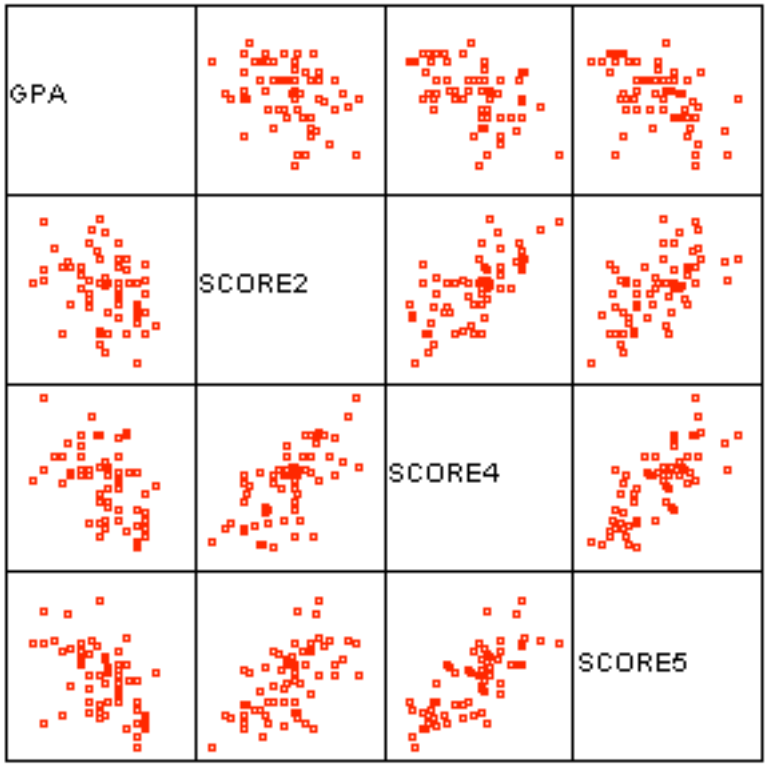
	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5

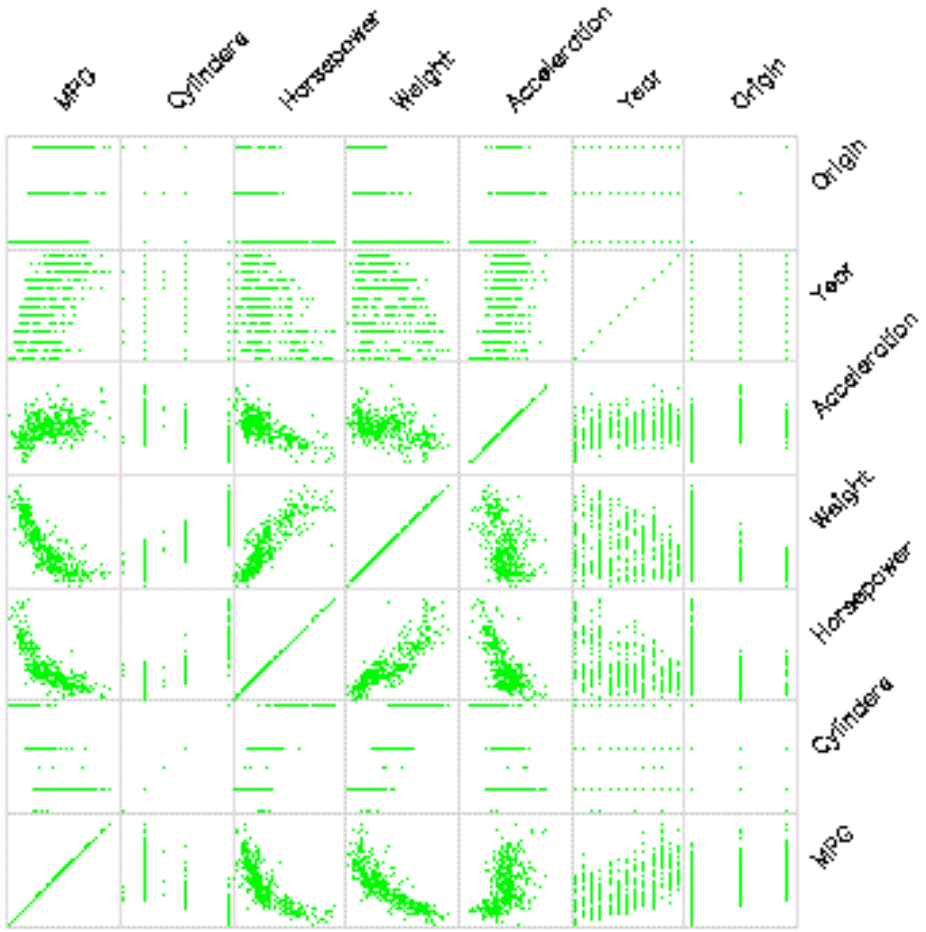


Each variable is shown separately

Scatterplot Matrix

Represent each possible pair of variables in their own 2D scatter plot



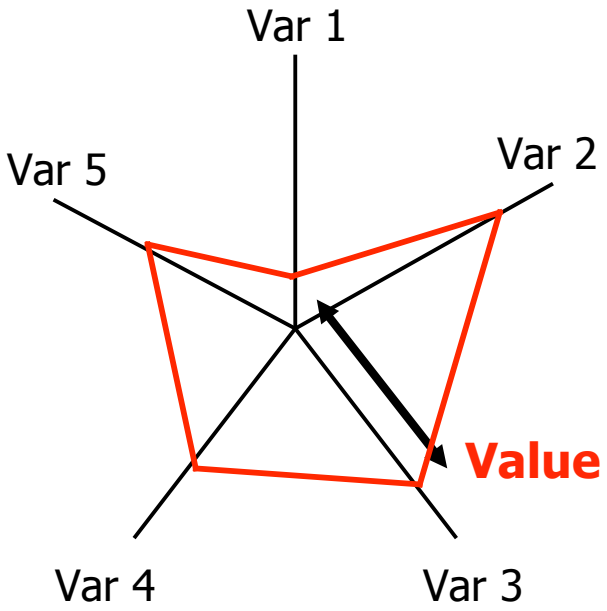


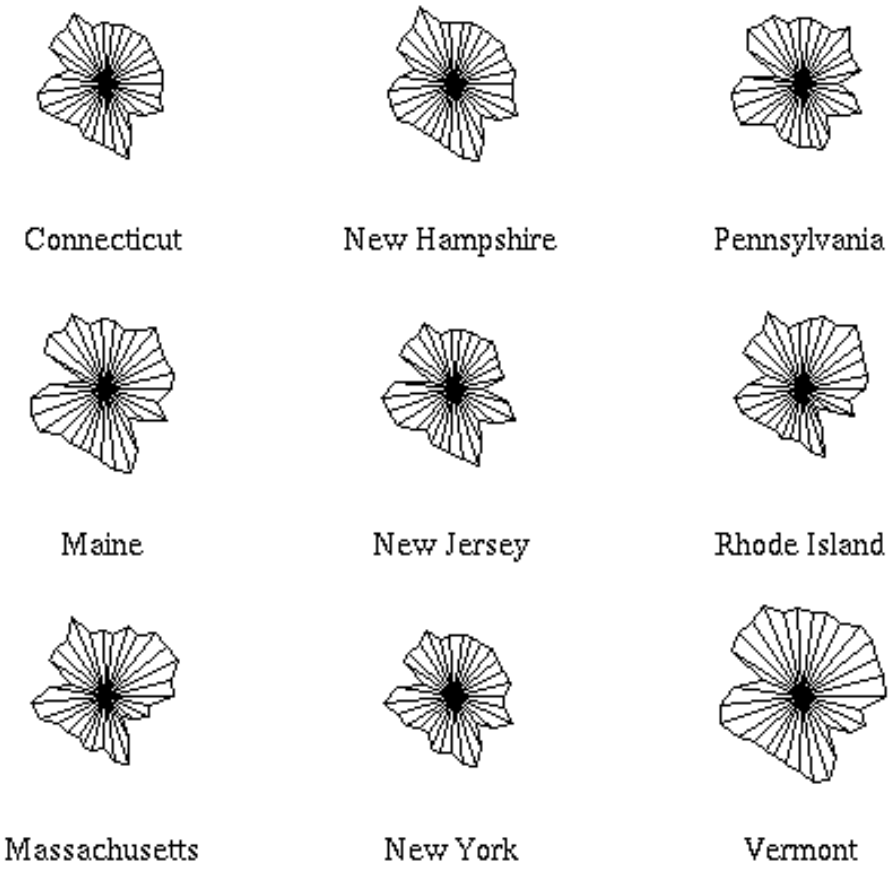
Brushing can aid interpretation:

Identify a group of points in one of the plots whereupon those objects are highlighted in all other plots

Star Plots

- Space out the *n* variables at equal angles around a circle
- Each spoke encodes a variable's value

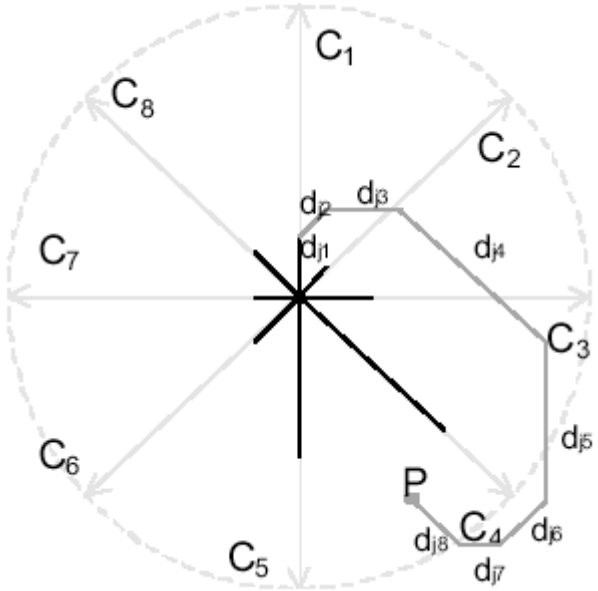


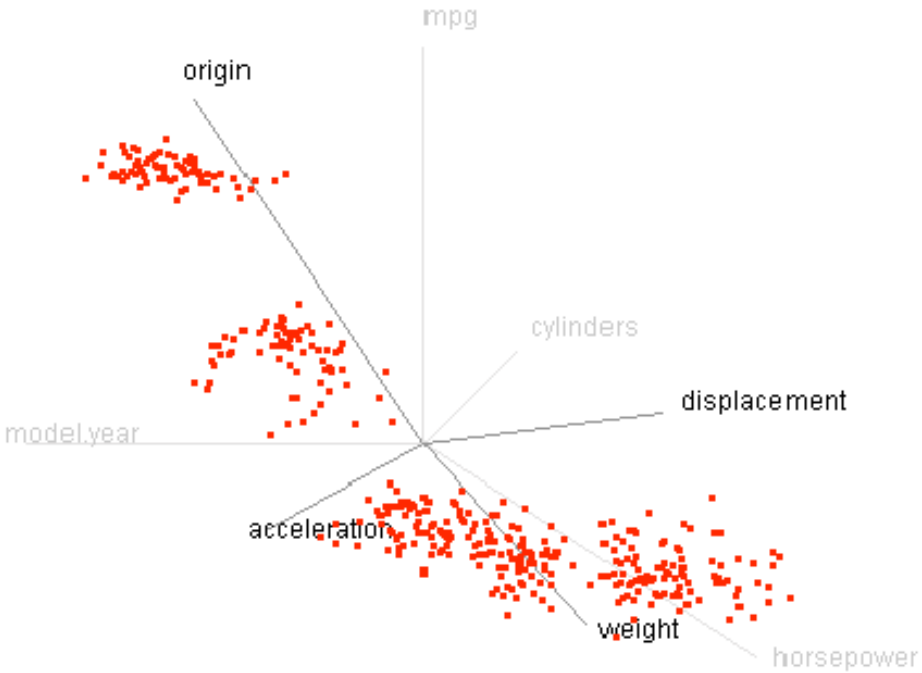


31 variables measured in nine states

Star Coordinates [Kandogan, 2000]

The basic idea of Star Coordinates is to arrange the coordinate axes on a circle on a two-dimensional plane with equal (initially) angles between the axes with an origin at the center of the circle (Figure 1). Initially, all axes have the same length. Data points are scaled to the length of the axis, with the minimum mapping to the origin and the maximum to the other end of the axis. Unit vectors are calculated accordingly.



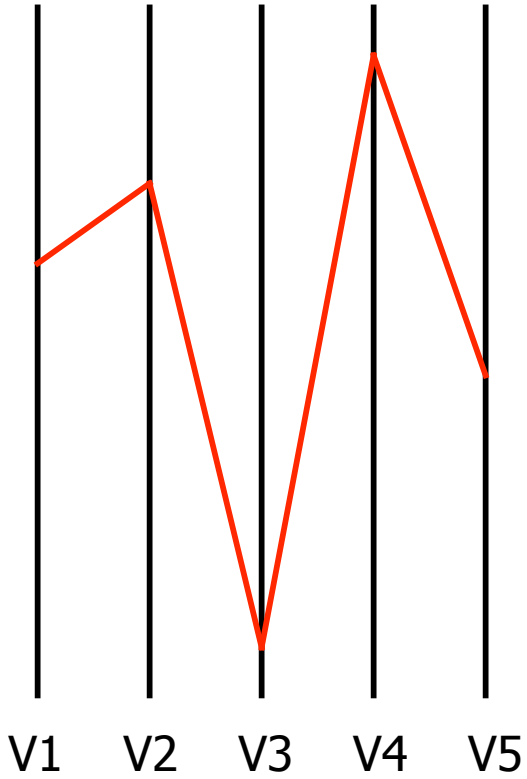


Cluster analysis in Cars data:
Four major clusters are discovered after playing with the data (by scaling, rotating, turning off some coordinates)

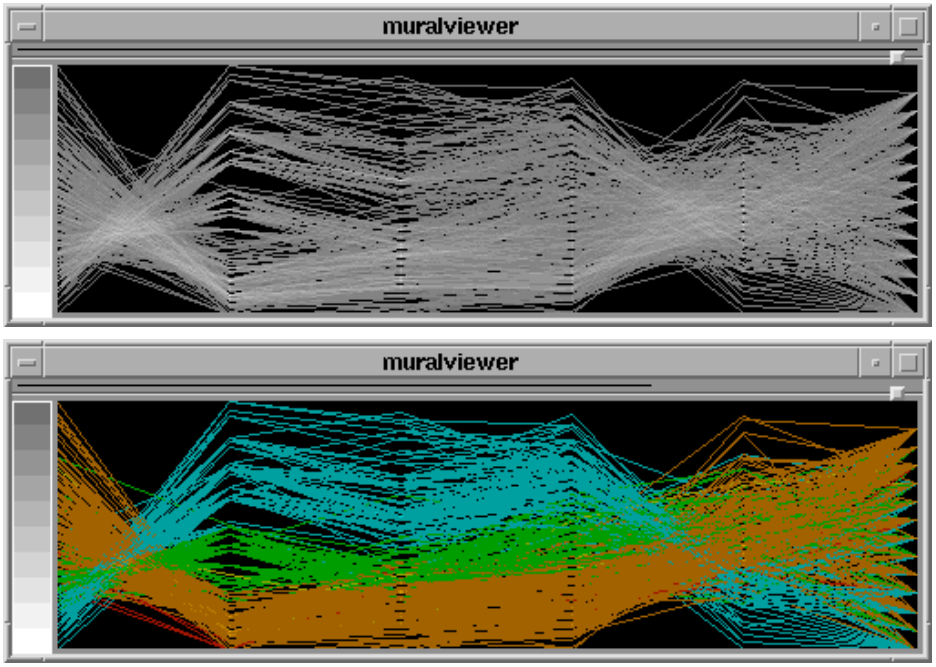
Scaling the 'origin' coordinate moves the only top two clusters

Parallel Coordinates [Inselberg, 1985, 1998]

- Parallel axes (vertical lines) specify dimensionality (*n* variables)
- A point in *n*-dimensional space is represented as a polyline with vertices (coordinates) on the parallel axes.



Five variables

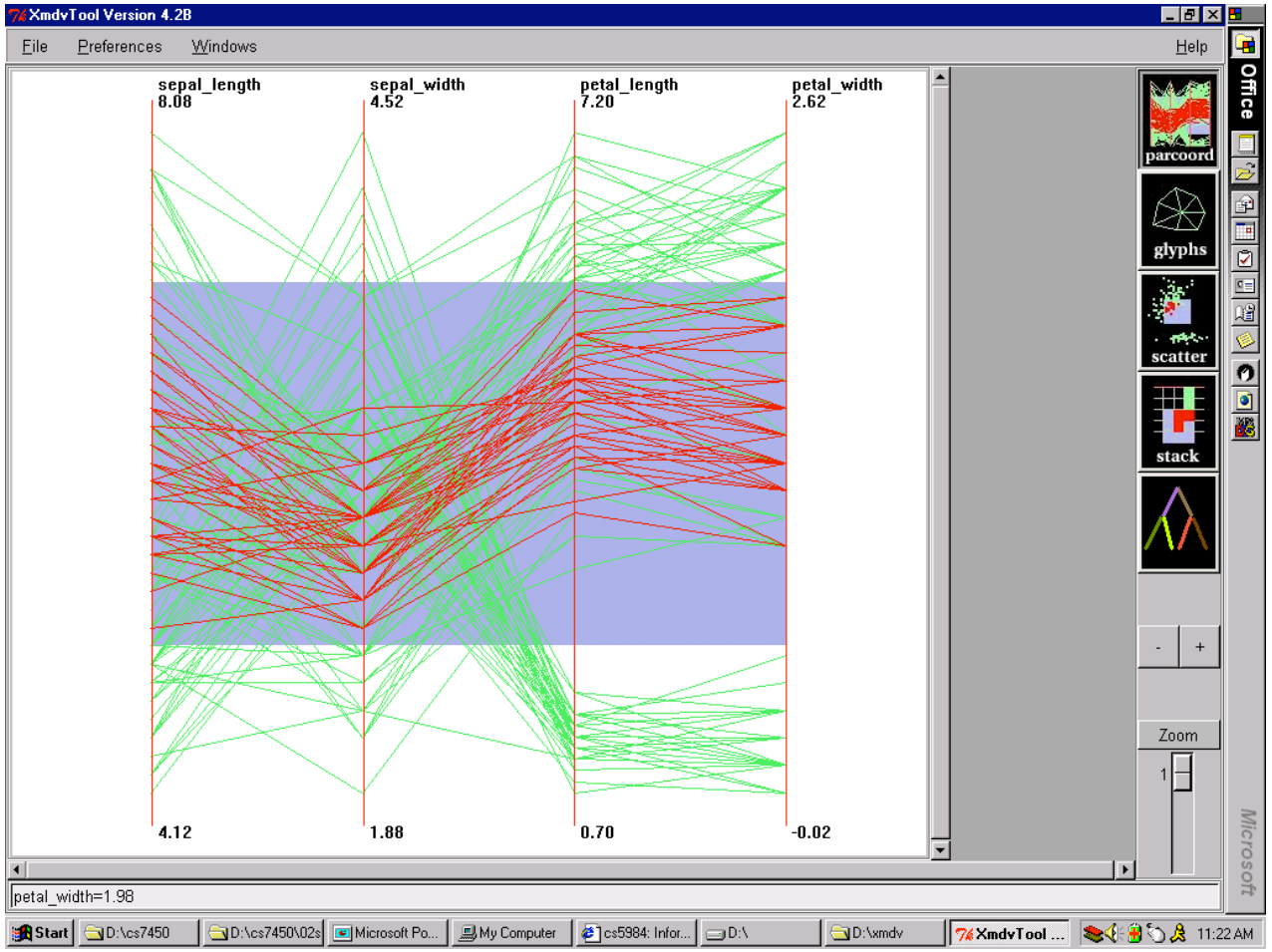


Mural of a parallel coordinate view of automobile data showing MPG, engine displacement, horsepower, weight, acceleration, and model year (1970-1982)

Xmdv

XmdvTool is a public domain software for interactive visual exploration of multivariate datasets

Includes parallel coordinates

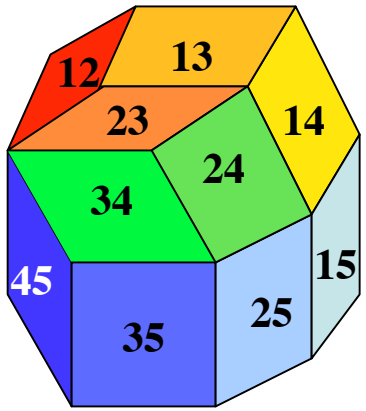


http://davis.wpi.edu/~xmdv

Hyperbox

- Hyperbox -- all possible pairs of variables are plotted against each other [Alpern and Carter, 1991]

- Any pair can be brought to front with Cartesian axes, with all others still visible



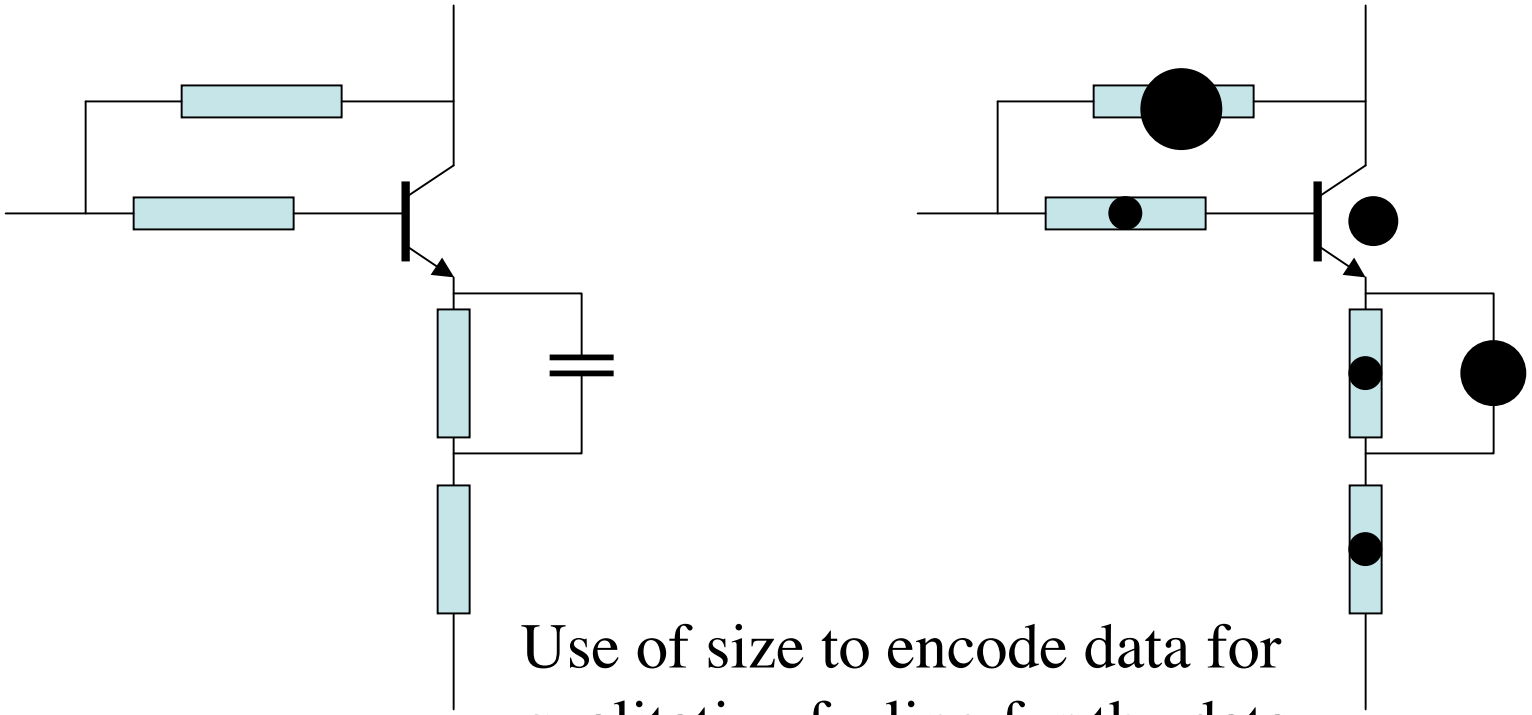
A 5-dimensional hyperbox

Other Representations

- Size
- Length and Height
- Color
- Face
- Multidimensional icons
- Pattern
- Virtual worlds

Size

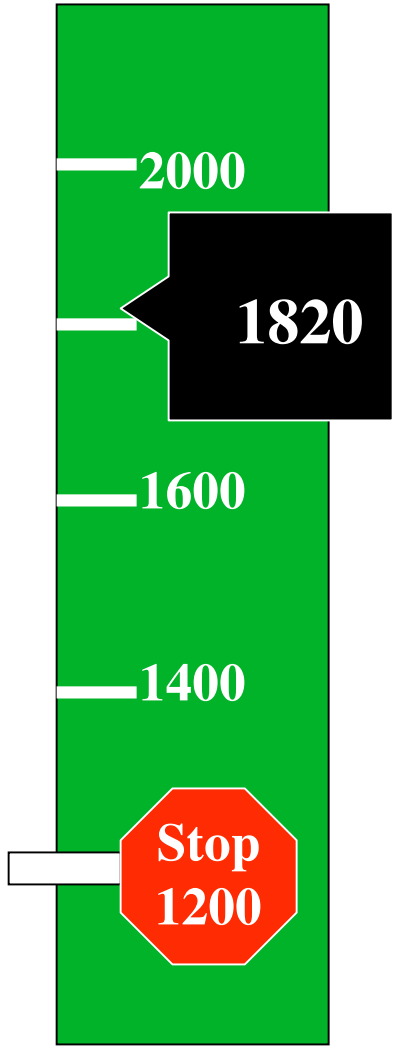
- Circles provide a qualitative indication of the sensitivity of the circuit's performance to a change in each component [Spence and Apperley, 1977]



Use of size to encode data for qualitative feeling for the data

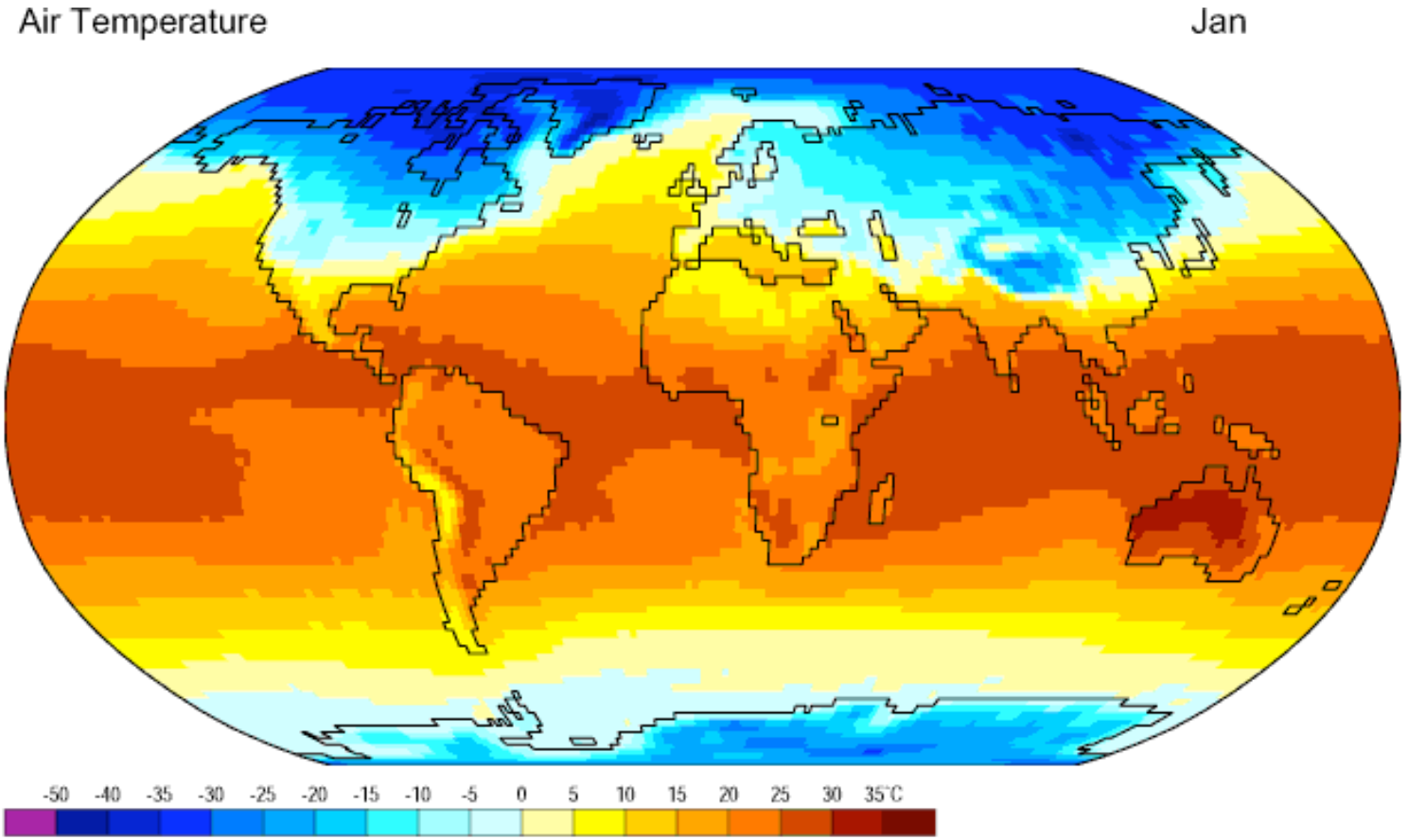
Length and Height

- Design of an altimeter (for the cockpit of a light aircraft) which provides both qualitative and quantitative indications of altitude [Matthew, 1999]



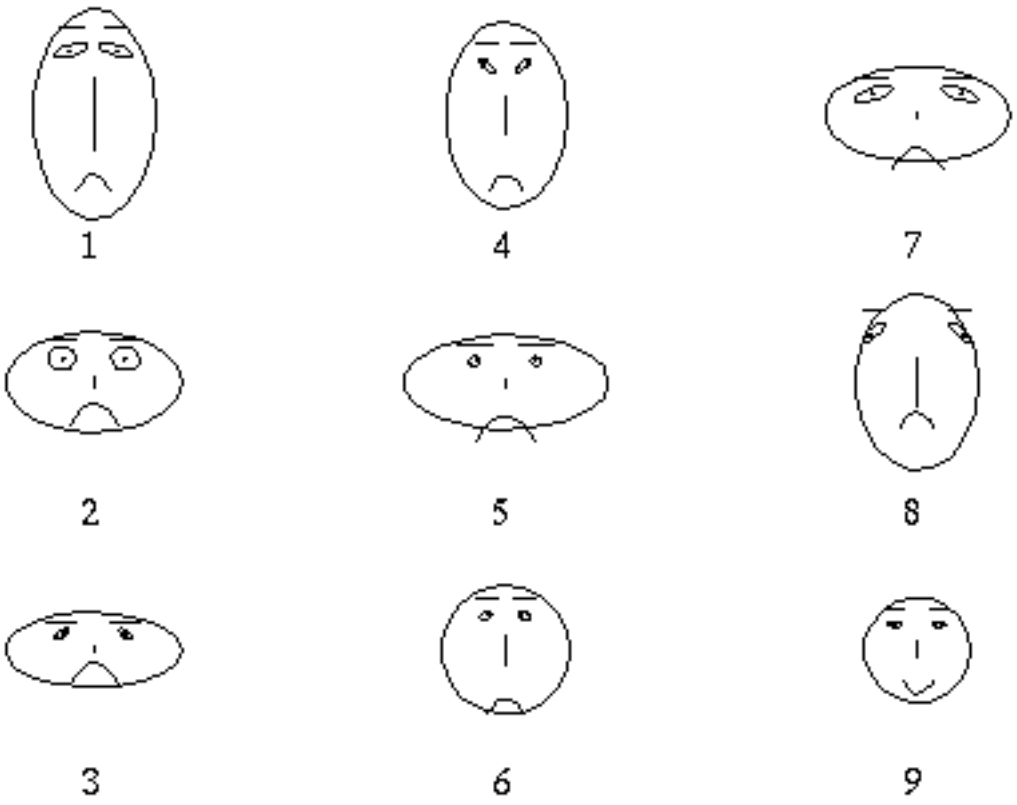
Color

- Mean January air temperature for the Earth's surface



Chernoff Faces

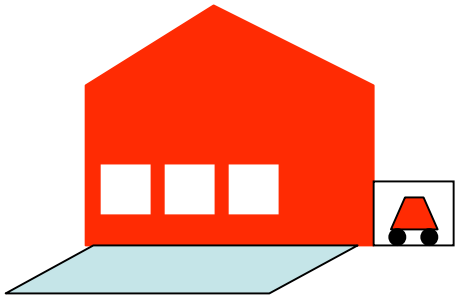
- Visualizing multivariate data developed by statistician H. Chernoff [1973]
- Chernoff faces map data to facial characteristics
- Applied to the study of geological samples (characterized by 18 attributes, e.g., salt content, water content)
- Identification of interesting groups of samples
- Use of Asymmetrical faces

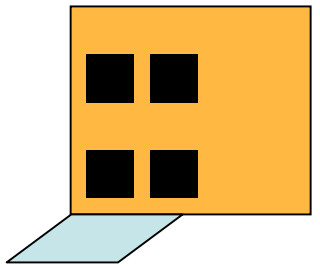


Applet in java: http://people.cs.uchicago.edu/~wiseman/chernoff/

Multidimensional Icons

- Multidimensional icons for different tasks and domains
- Selecting a house satisfying certain requirements [Spence and Parr, 1991]
  - Color encodes price band (red is over $400,000, orange between $300,000 and $400,000), yellow between $200,000 and $300,000 and white between $100,000 and 200,000)
  - Number of bedrooms indicated by windows
  - Black or white windows means bad or good state of repair
  - Shape encodes a categorical variable (house, apartment, and cottage)
  - Garden size is indicated by size
  - Garage is represented by a symbol

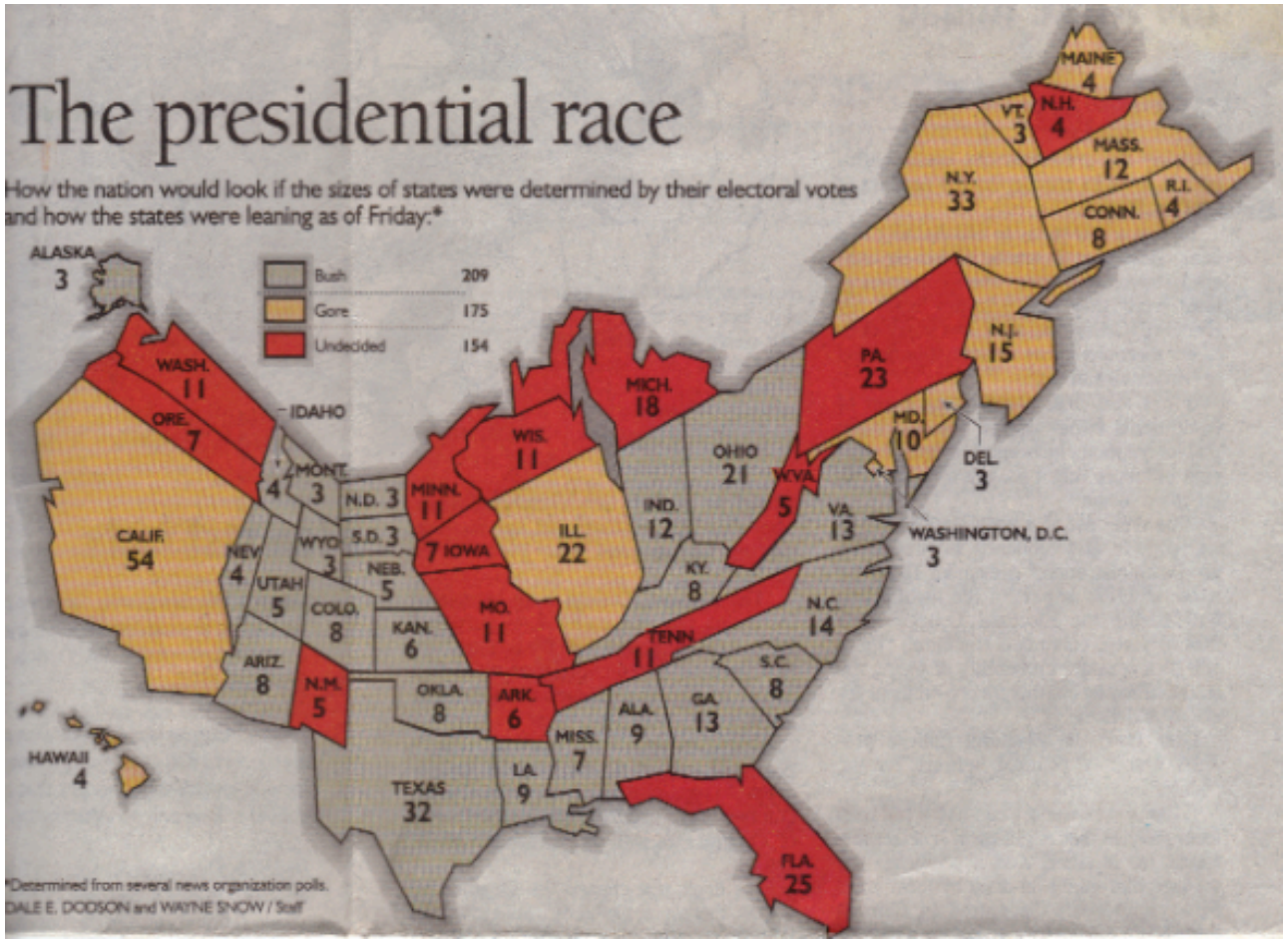




Six dimensions are represented

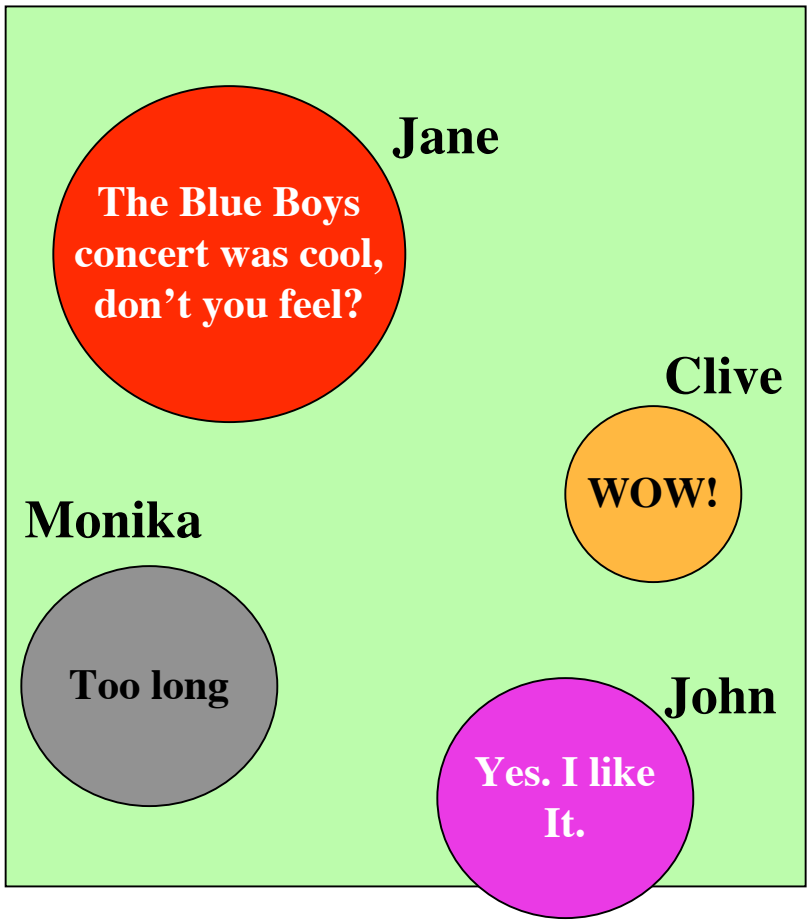
Magnification

- Magnification as an encoding scheme for geographic data
- Electoral College



Patterns

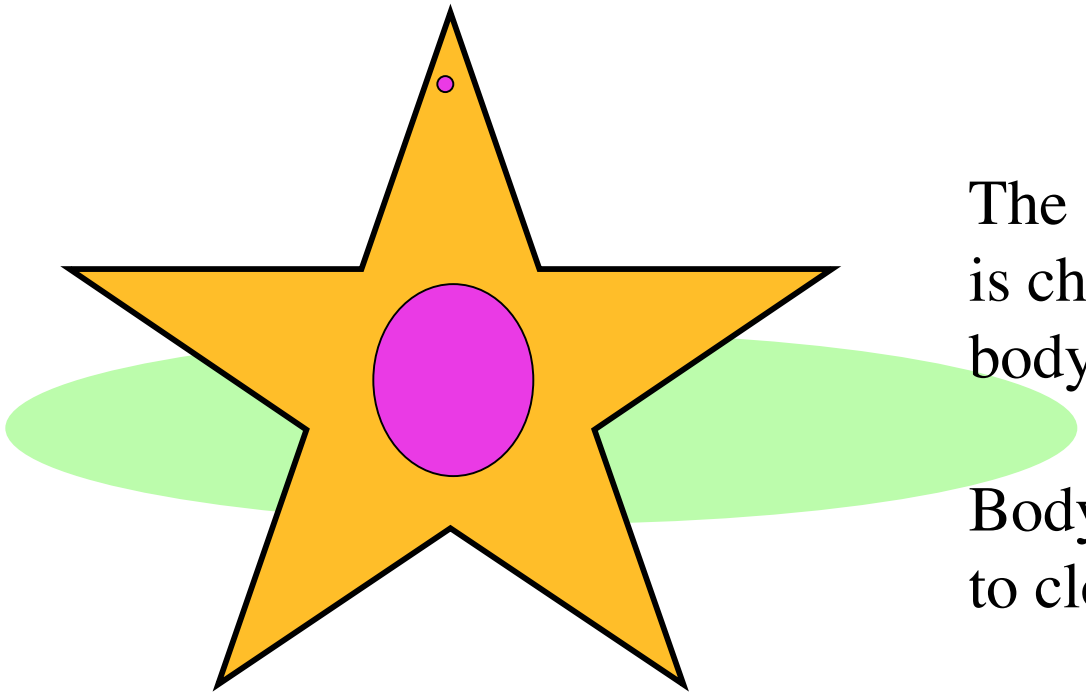
- Chart circles allow the visualization of an internet discussion [Viegas and Donath, 1999]



Human pattern recognition

Virtual Worlds

- Electronic imaginary worlds -- Virtual worlds
- A StarCursor representing a human being in a virtual world [Rankin et al., 1998]



The anthropomorphic StarCursor is characterized by eye, heart, body, limbs, aura.

Body can be colored according to clothing