visualizing multidimensional data

John Brosz and Uta Hinrichs, CPSC 583 – Nov 3, 2011
terminology

• multidimensional data
  – data that has many independent /variables attributes
  – houses: garden size, # of bedrooms

• multivariate/hypervariate data
  – data that has many dependent attributes/variables
  – philosophers’ occupation & interest
linked histograms
linked histograms

→ Attribute Explorer
linked histograms

- price
- bedrooms
- garden size
linked histograms

- selection of a subset of houses in the desired price range
- black = houses outside of the price range
linked histograms

• adjusting price range and # of bedrooms
• black = house fails one limit
• dark grey = houses fail 2 limits
linked histograms

- knowing how to readjust the attributes if no desired solution is available
parallel coordinates
parallel coordinates
advantages

• complexity correlates with # of variables
• can be applied to any # of dimensions
• all variables receive uniform treatment
limitations

• not immediately visible:
  – trade-off between B and E
  – strong correlation between C and G

• ordering of attributes affects how easy relationships can be identified
interaction techniques

- selection and highlighting
type of insight to be gained

- attribute visibility
- relations between attributes
parallel coordinates in data mining

A. Inselberg. Multidimensional Detective.
parallel coordinates – discovery process

• don’t panic
• understand the objectives and use them to obtain visual cues
parallel coordinates – discovery process

- carefully scrutinize the picture

A. Inselberg. Multidimensional Detective.
parallel coordinates – discovery process

- test assumptions

A. Inselberg. Multidimensional Detective.
star plots
similarities to parallel coordinates

- attribute value = point on coordinate axis
- attribute points of one object joined by lines
star plots

- attribute axes radiate from a common origin
comparing different data

• high object visibility
3D parallel coordinates