Data Mining & Data Warehouse

Associate Professor Dr. Raed Ibraheem Hamed

University of Human Development,
College of Science and Technology
Department of Information Technology

2016 – 2017

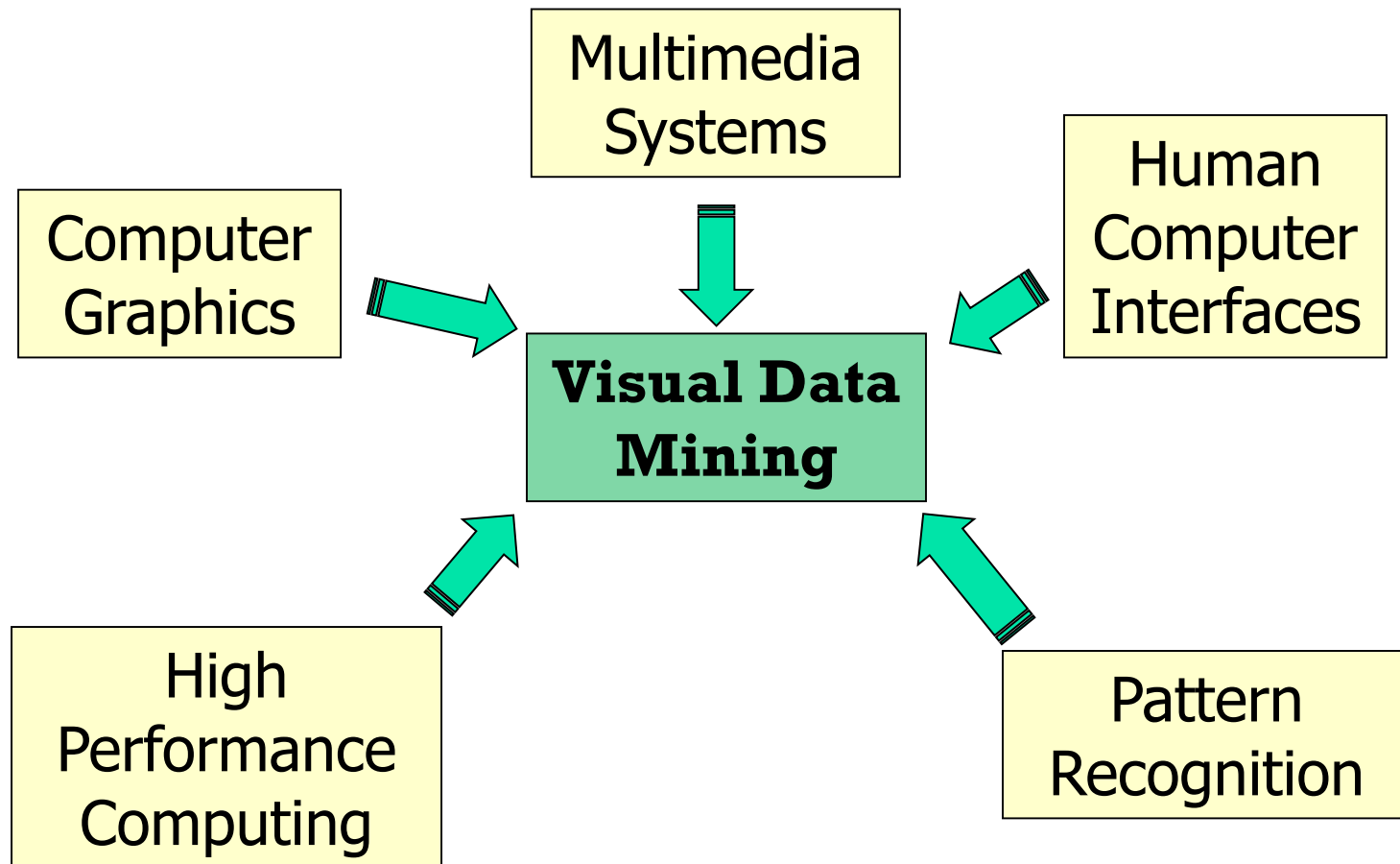Department of IT- DMDW - UHD

Introduction

- Visual Data Mining?

- Purpose of Visualization for Data Mining

- Visual Data Mining & Data Visualization

- Data Mining Result Visualization

- Data Mining Process Visualization

- Visual Classification

- Visualization of Data Mining Processes by Clementine

- Methods of Data Visualization

What is Visual Data Mining

- **Visualization**: Use of computer graphics to create visual images which aid in the understanding of complex, often massive representations of data.

- **Visual Data Mining**: Visual Data Mining presents the data in some visual form, allowing users to mine and gain insight into the data, draw conclusions and directly interact with the data.

Visual Data Mining is Closely Related to :

Multimedia Systems

Computer Graphics

Human Computer Interfaces

Visual Data Mining

High Performance Computing

Pattern Recognition

Purpose of Visualization for Data Mining

1. Gain insight into an information space by mapping data onto graphical primitives
2. Provide qualitative overview of large data sets
3. Search for patterns, trends, structure, irregularities, relationships among data.
4. Help find interesting regions and suitable parameters for further quantitative analysis.
5. Provide a visual proof of computer representations derived

Computer Representations Of Reality
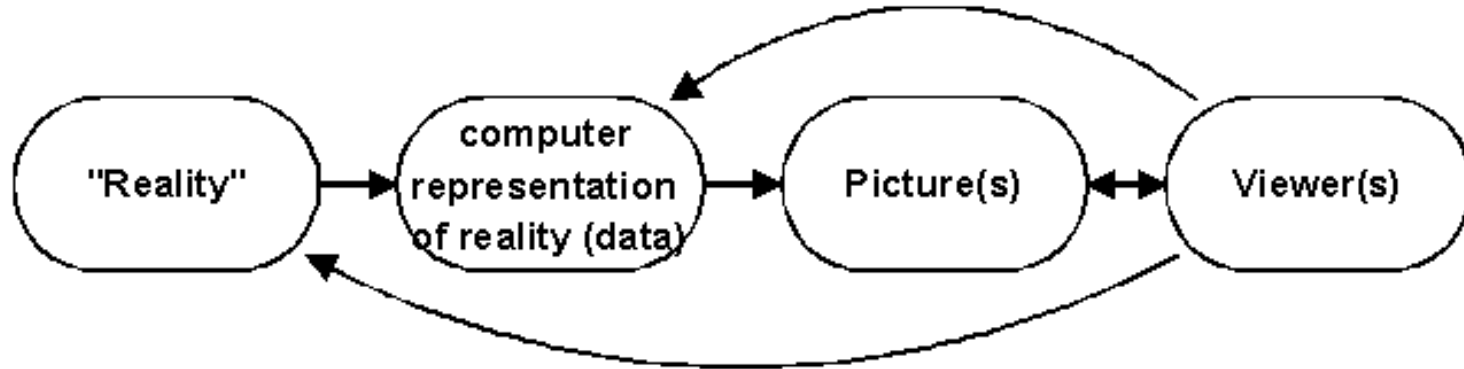
As depicted by the above figure, visualization is essentially a mapping process from computer representations to perceptual representations, choosing techniques to maximize human understanding. The goal of a viewer might be a deeper understanding of physical phenomena, but it also might be a visual proof of computer representations derived from such an initial stage.

Data Mining Algorithms Verses Visualization

	Data Mining Algorithms	Visualization
Actionable	+	−
Evaluation	+	−
Flexibility	−	+
User Interaction	−	+

Visual Data Mining & Data Visualization

- Integration of visualization and data mining
  - data visualization
  - data mining result visualization
  - data mining process visualization
- Data visualization
  - Data in a database or data warehouse can be viewed
    - at different levels of abstraction
    - as different combinations of attributes or dimensions
  - Data can be presented in various visual forms

Data Mining Result Visualization

- Presentation of the results or knowledge obtained from data mining in visual forms

- Examples
  - Decision trees
  - Association rules
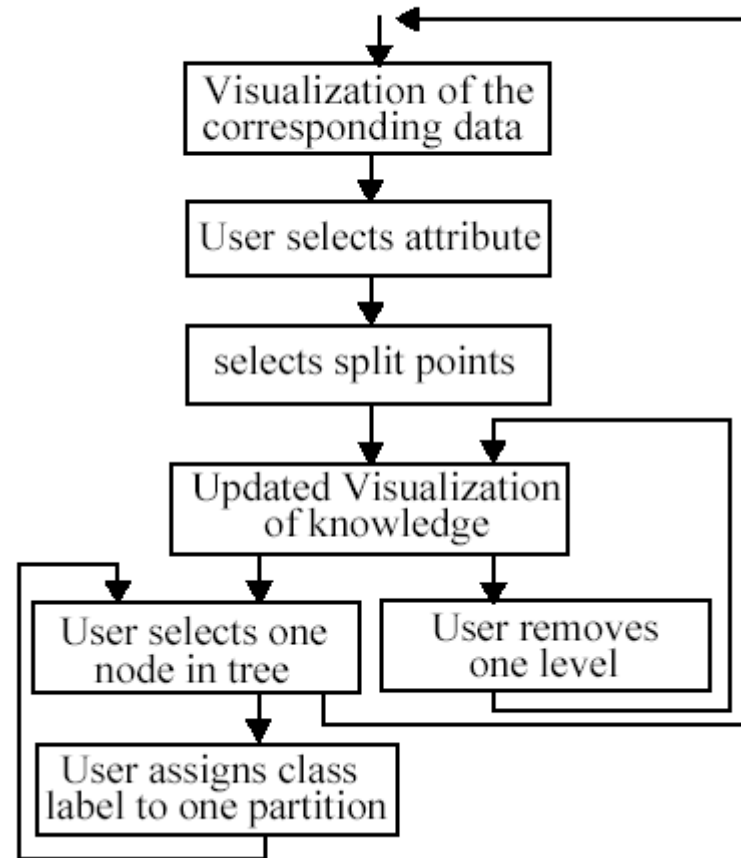 - Clusters
  - Outliers
  - Generalized rules
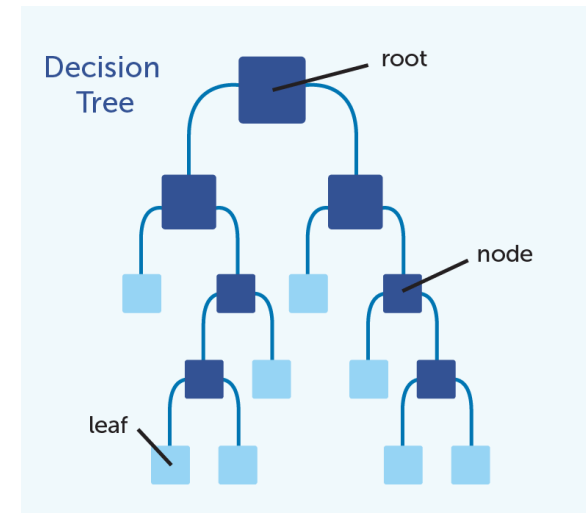
Data Mining Process Visualization

- Presentation of the various processes of data mining in visual forms so that users can see

  - Data extraction process

  - Where the data is extracted

  - How the data is cleaned, integrated, preprocessed, and mined

  - Method selected for data mining

  - Where the results are stored

  - How they may be viewed

Visual Classification

"Visual Classification: An Interactive Approach to Decision Tree Construction"

INTERACTIVE CLASSIFICATION

```
        ┌──────────────────────────┐
        │ Visualization of the     │◄──────────┐
        │ corresponding data       │           │
        └────────────┬─────────────┘           │
                     │                         │
        ┌────────────▼─────────────┐           │
        │ User selects attribute   │           │
        └────────────┬─────────────┘           │
                     │                         │
        ┌────────────▼─────────────┐           │
        │ selects split points     │           │
        └────────────┬─────────────┘           │
                     │          ┌──────────────┤
        ┌────────────▼──────────▼──┐           │
        │ Updated Visualization    │           │
        │ of knowledge             │           │
        └──┬──────────┬──────────┬─┘           │
           │          │          │             │
   ┌───────▼──────┐   │   ┌──────▼────────┐    │
   │ User selects │   │   │ User removes  │────┤
   │ one node in  │   │   │ one level     │    │
   │ tree         │   │   └───────────────┘    │
   └───────┬──────┘   │                        │
           │          │                        │
   ┌───────▼──────────▼──┐                     │
   │ User assigns class   │─────────────────────┘
   │ label to one partition│
   └──────────────────────┘
```

Audio Data Mining

- Uses audio signals to indicate the patterns of data or the features of data mining results

- An interesting alternative to visual mining

- An inverse task of mining audio (such as music) databases which is to find patterns from audio data

- Visual data mining may disclose interesting patterns using graphical displays, but requires users to concentrate on watching patterns

VDM Approach

VDM takes advantage of both,

- The power of automatic calculations, and
- The capabilities of human processing.
  - Human perception offers phenomenal abilities to extract structures from pictures.

Levels of VDM

- No or very limited integration
  - Corresponds to the application of either traditional information visualization or automated data mining methods.

- Loose integration
  - Visualization and automated mining methods are applied sequentially.

  - The result of one step can be used as input for another step.

- Full integration
  - Automated mining and visualization methods applied in parallel.

  - Combination of the results.

Methods of Data Visualization

Different methods are available for visualization of data based on type of data

Data can be

- Univariate

- Bivariate

- Multivariate

Univariate data

- Measurement of single quantitative variable

- Characterize distribution

- Represented using following methods

    - Histogram

    - Pie Chart

Histogram

Histogram — Frequency vs. Queue Time (minutes)
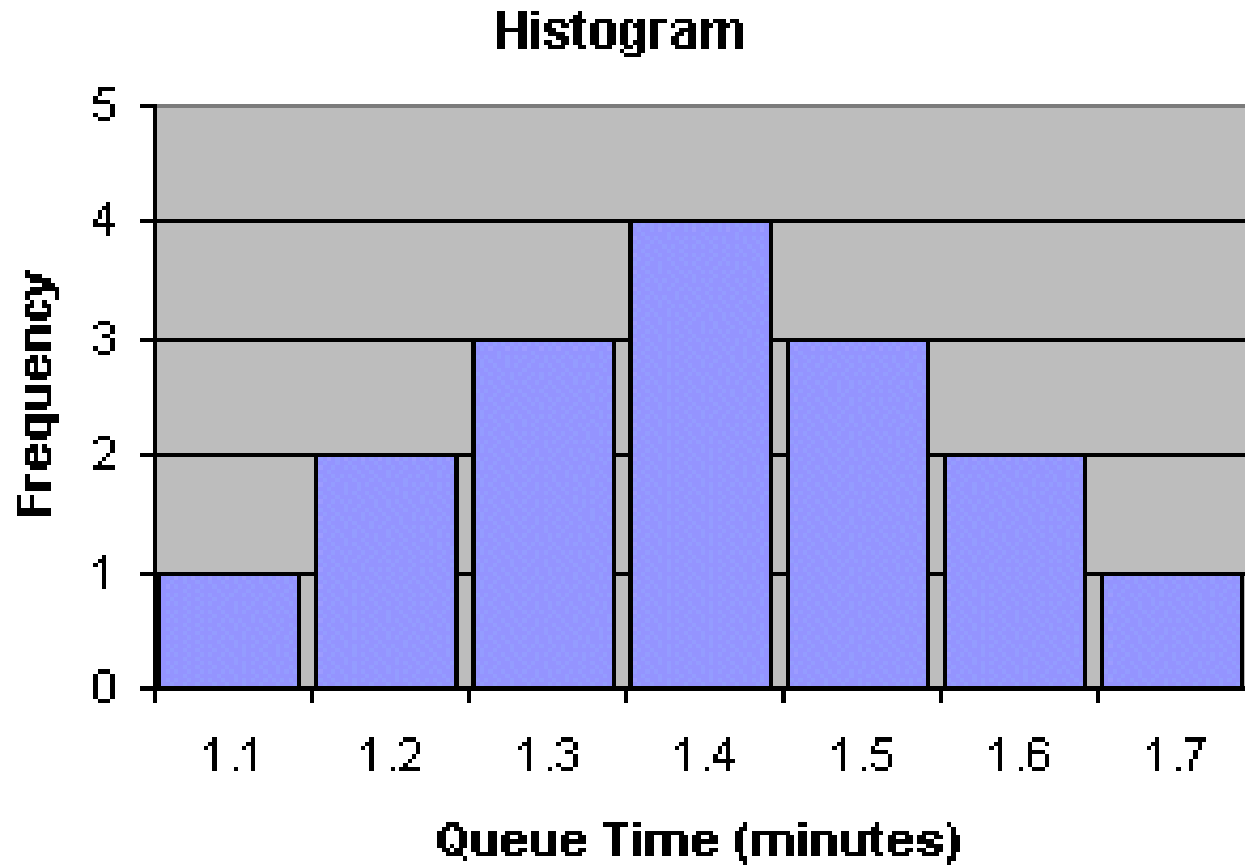
Pie Chart
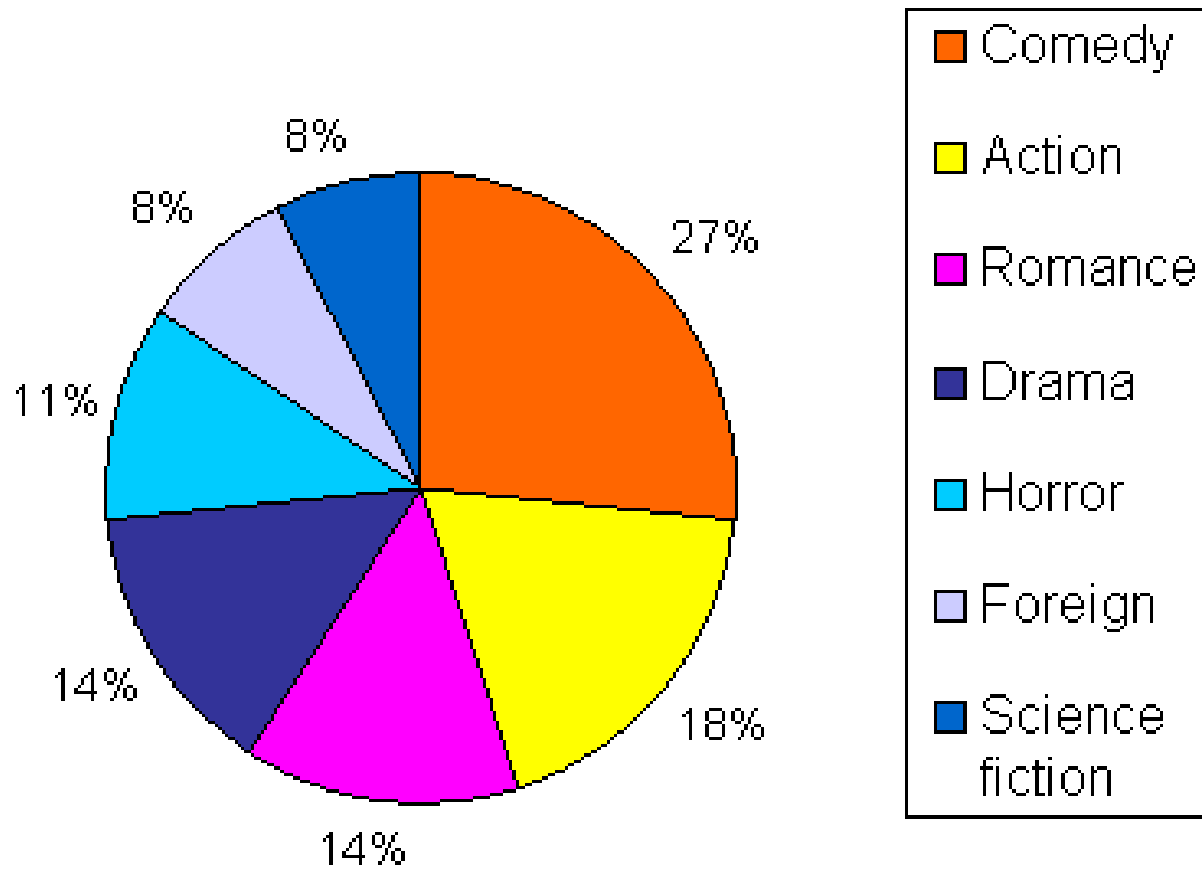
Bivariate Data

- Constitutes of paired samples of two quantitative variables

- Variables are related

- Represented using following methods

  - Scatter plots

  - Line graphs

Scatter plots

Scatter Plot, SAT vs. Beta Test
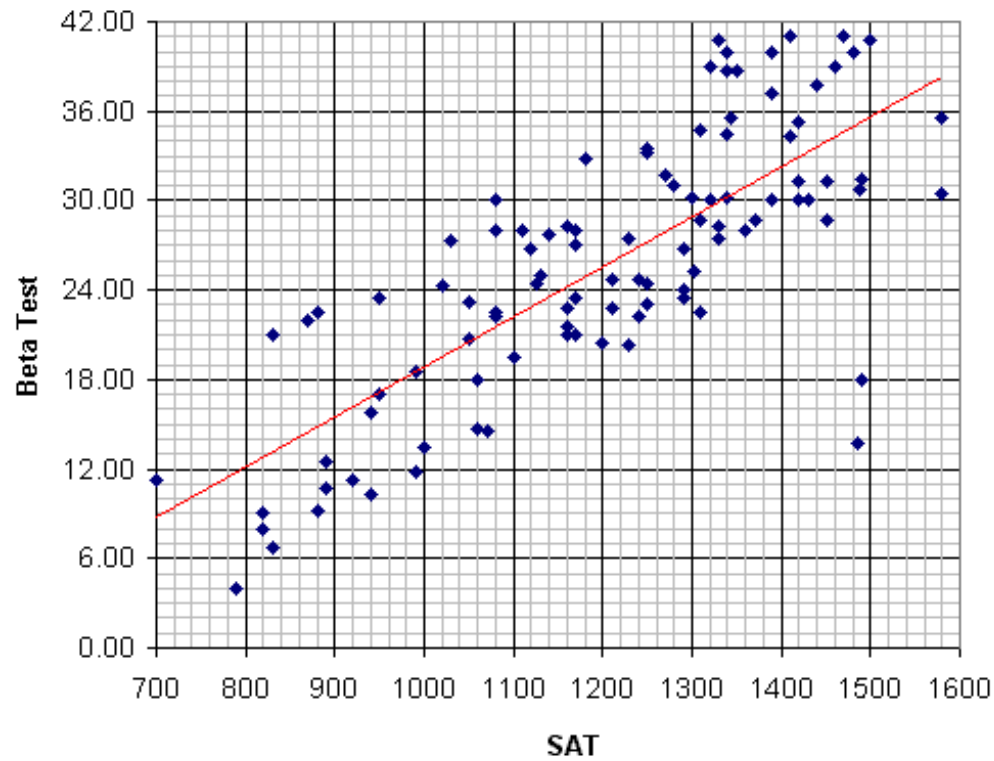
Line graphs
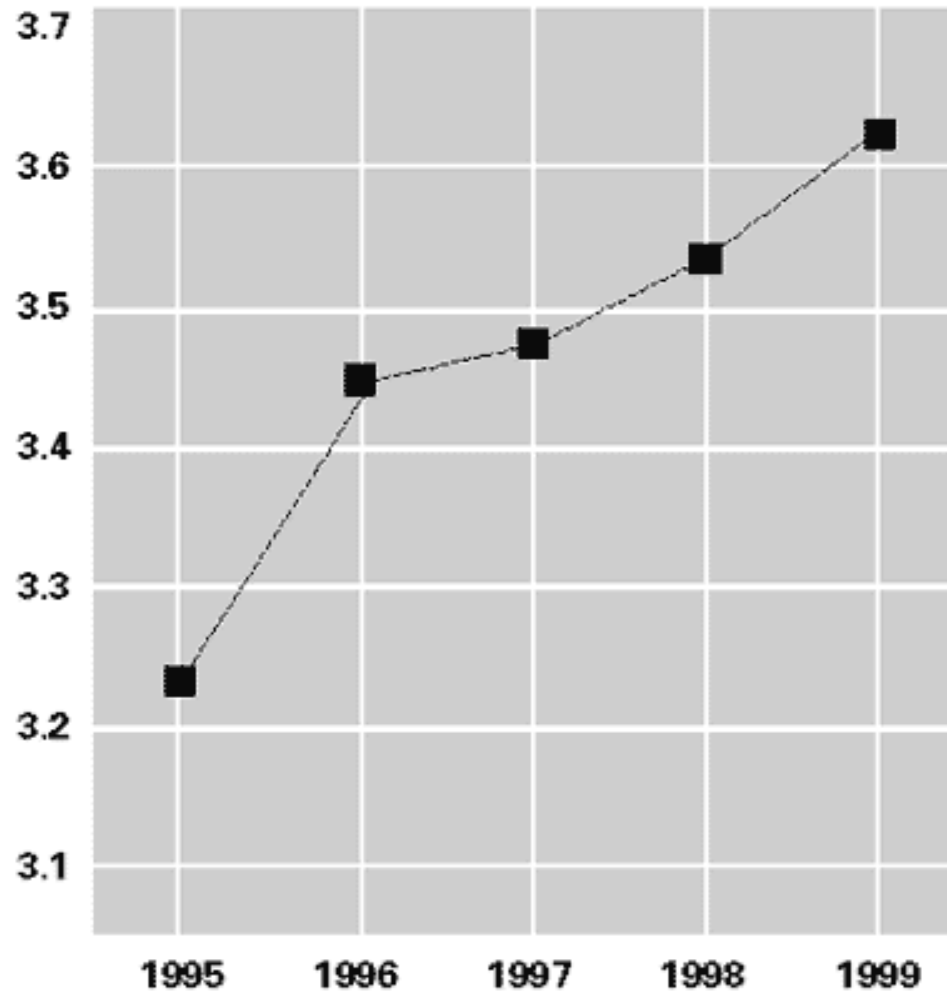
Multivariate Data

- Multi dimensional representation of multivariate data

- Represented using following methods

  - Icon based methods

  - Pixel based methods

  - Dynamic parallel coordinate system

Icon based Methods
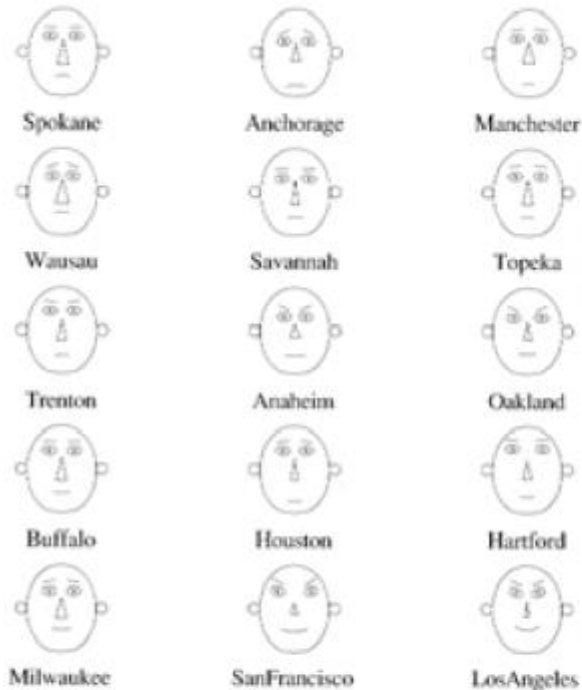
Visualizing town data with Chernoff-faces, (Spence, 2001)
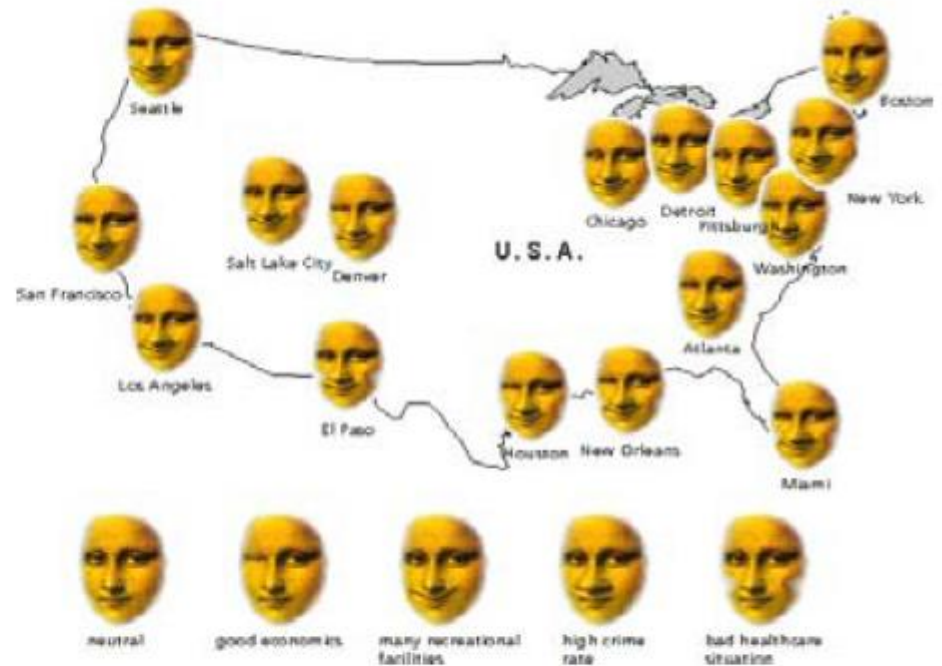
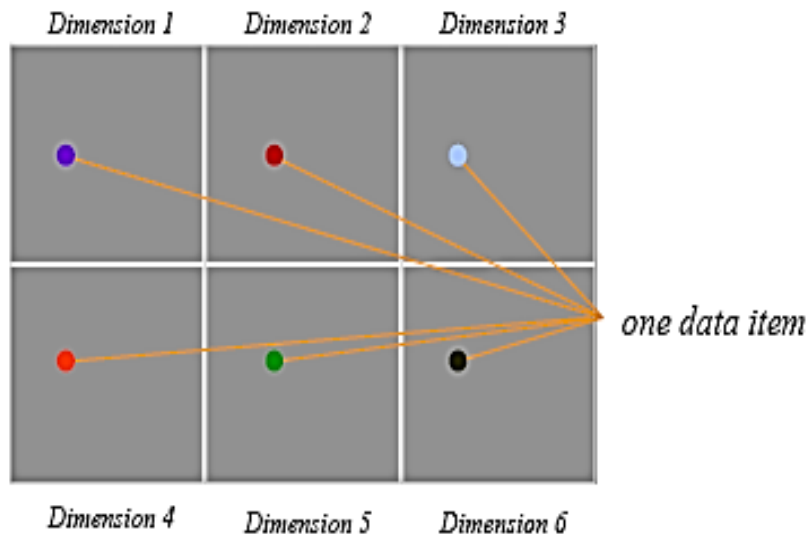Visualization of town data using morphed faces (Alexa, Müller, 1998)

Pixel Based Methods

- Approach:

    - Each attribute value is represented by one colored pixel (the value ranges of the attributes are mapped to a fixed color map).
    - The values of each attribute are presented in separate sub windows.

- Examples:
    - Dense Pixel Displays

Dense Pixel Display

Approach:

- Each attribute value is represented by one colored pixel.

Thank You