Data & Tasks

Thursday 4 Oct 2007
Polle Zellweger

Overview

Information Visualization model & process

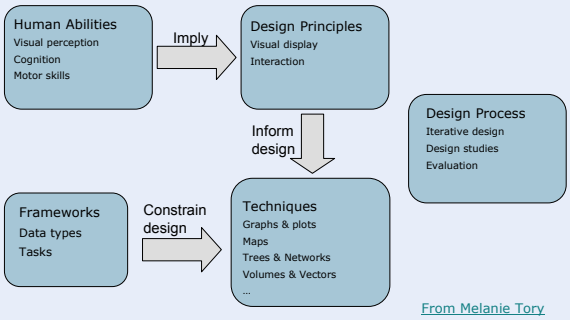
Data

Influence on representation (encodings)

Tasks & data structure/complexity

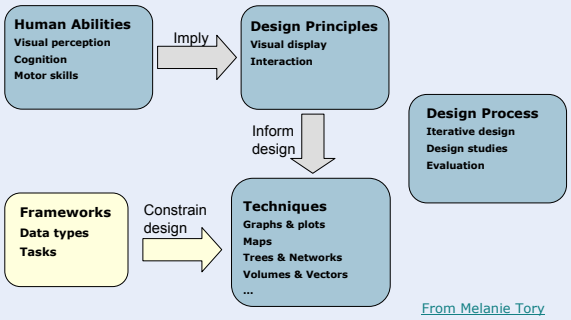
Influence on representation

Visualization Components



From Melanie Tory

Visualization Components



From Melanie Tory

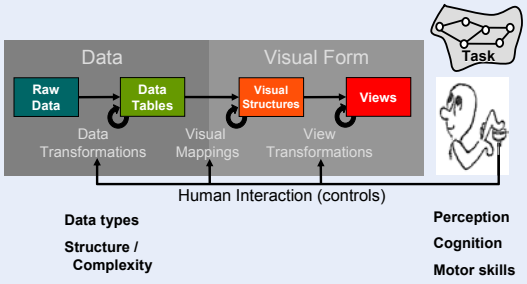
The message is central

"The impact and appeal of information, quantitative or not, flows naturally from the significance and relevance of the message it contains."

"When you design the display of quantitative information, whether you use a table or graph, the specific type of table or graph you use depends primarily on your message."

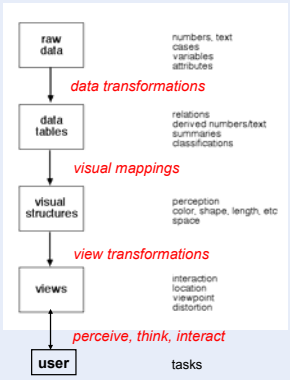
-Few, *Show Me the Numbers*, page 15

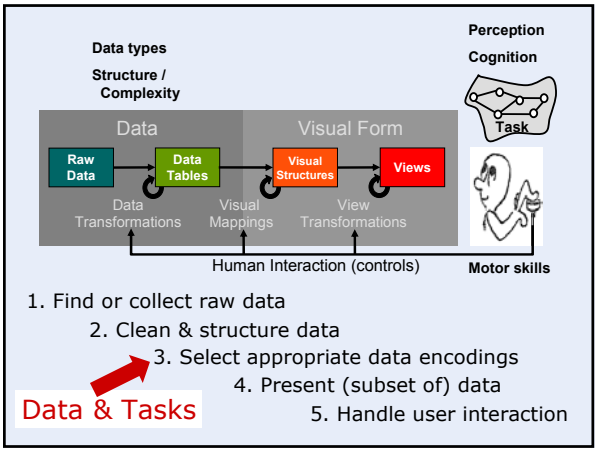
Information Visualization Model



Also describes the process for developing a visualization

Information Visualization Model (Revealed)





1. Find or collect raw data
2. Clean & structure data
3. Select appropriate data encodings
4. Present (subset of) data
5. Handle user interaction

Data & Tasks

Data

Few
Chapter 2: Numbers Worth Knowing

Basic Elements of a Data Model

A *data model* represents some aspect of the world

Data models consist of these basic elements:

- objects
- values (also called *attributes*)
- relations

Basic Elements: Objects

Objects are items of interest

- people, plants, cars, films, etc…

Objects allow you to define and reason about a domain

- ecosystem: ponds, streams, woodlands, mountains, plants, animals, etc.

Basic Elements: Values

*Values (*or *attributes)* are properties of objects

Two major types

- quantitative
- categorical

Appropriate visualizations often depend upon the type of the data values

Basic Elements: Relations

Relations relate two or more objects
- *leaves* are part of a *plant*
- a *department* consists of *employees*

Ecosystem
- connections between streams and lakes
- predator/prey network of what eats what
- …

Types of quantitative information

Quantitative (allows arithmetic operations)

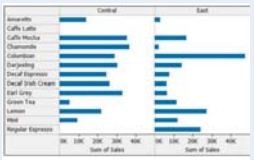
Categorical (group, identify & organize; no arithmetic)
- Nominal
- Ordinal
- Interval
- Hierarchical

Types of quantitative information

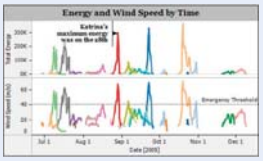
Quantitative (allows arithmetic operations)
- *123, 29.56, …*

Categorical (group, identify & organize; no arithmetic)
- Nominal (name only, no ordering)
  - *Region: North, East, South, West*
- Ordinal (ordered, < operator)
  - *Time: Jan, Feb, Mar, …*
- Interval (subdivide into ordered ranges)
  - *0-999, 1000-4999, 5000-9999, 10000-19999, …*
- Hierarchical (successive inclusion)
  - *Region: North=(AK, WA, OR, ID, …)*
  - *West=(CA, NV, AZ, UT, …)*

Nominal



Ordinal (time series)



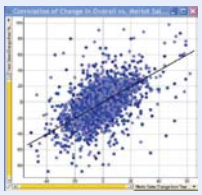
Hierarchical

Derived Values

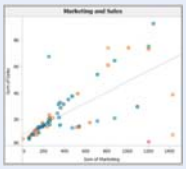
Relationships between quantitative values
- Ranking
- Ratio
- Correlation

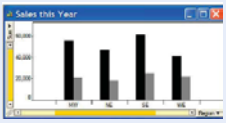
Simple statistics
- Average: mean, median, mode, midrange
- Distribution: range, standard deviation
- Correlation
  - linear correlation coefficient=
  - -1 negative correlation; 0 no correlation; +1 positive

Ranking



Correlation (trend line)





Total sales vs merlot sales (grey)

Merlot sales as % of total sales

Ratios

Types of Data Values

Other infovis authors use a simplified set of basic types…

Nominal (categorical)
- no ordering (= or ≠ only)
- *film type: Action, Comedy, Romance, …*

Ordinal
- can be ordered in sequence (< operator)
- *film rating: G, PG, PG-13, R*

Quantitative
- allow arithmetic operations
- *film length: 120 minutes*

Data Tables

VARIABLES

CASES

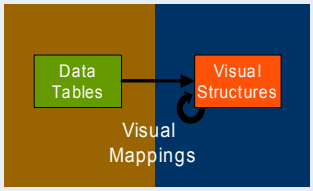
FilmID	Title	Director	Actor	Actress	Year	Length	Popularity	Rating	Film Type
230	Goldfinger	Hamilton	Connery	Blackman	1964	112	7.7	PG	Action
105	Ben Hur	Wyler	Heston	Harareet	1959	212	8.2	G	Action
540	Ben Hur	Niblo	Novarro	McAvoy	1926	133	7.4	G	Drama
…	…	…	…	…	…	…	…	…	…

Examples of Data Types

Nominal Ordinal Quantitative

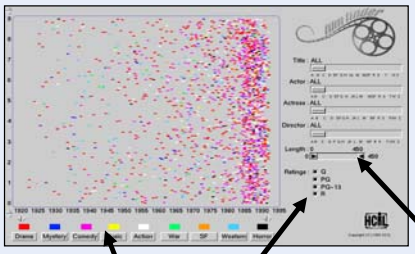
FilmID	Title	Director	Actor	Actress	Year	Length	Popularity	Rating	Film Type
230	Goldfinger	Hamilton	Connery	Blackman	1964	112	7.7	PG	Action
105	Ben Hur	Wyler	Heston	Harareet	1959	212	8.2	G	Action
540	Ben Hur	Niblo	Novarro	McAvoy	1926	133	7.4	G	Drama
…	…	…	…	…	…	…	…	…	…

Visual Mappings



Visual Structure

Film Finder, U of Maryland Video



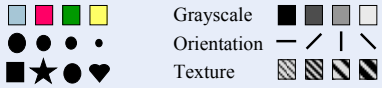
Nominal: Color Ordinal: List Quantitative: Axis

Bertin's Graphical Vocabulary

Position

Marks
- Points
- Lines
- Areas

Retinal variables
- Color
- Size
- Shape
- Grayscale
- Orientation
- Texture



Tasks & Data Structure/complexity

Shneiderman

The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization

Visual Information Seeking Mantra

Overview, zoom & filter, details-on-demand
Overview, zoom & filter, details-on-demand
Overview, zoom & filter, details-on-demand
Overview, zoom & filter, details-on-demand
Overview, zoom & filter, details-on-demand
Overview, zoom & filter, details-on-demand
Overview, zoom & filter, details-on-demand
Overview, zoom & filter, details-on-demand
Overview, zoom & filter, details-on-demand

Tasks (abstract)

- Overview
- Zoom
- Filter
- Details on demand
- Relate
- History
- Extract

Tasks

Overview: see overall patterns, trends
Zoom: see a smaller subset of the data
Filter: see a subset based on values, …
Details on demand: see values of objects
Relate: see relationships, compare values
History: keep track of actions & insights
Extract: mark & capture data

Sample user tasks (concrete)

Find value of specific object
Find object with given value
See patterns, trends, outliers
Compare attributes of two objects

Data structure / complexity

1D
2D
3D
Temporal

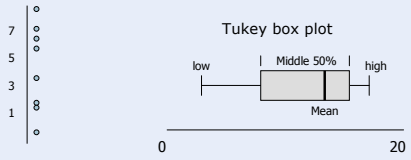
Multi-D class Oct 18
Trees class Nov 1
Networks class Nov 1

Data complexity + sample tasks

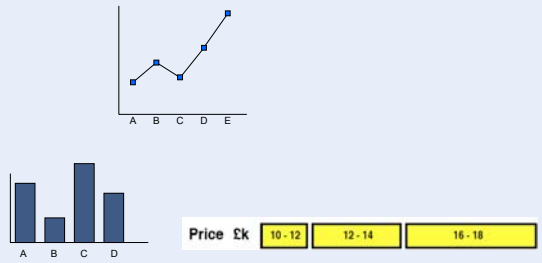
1D *find # of items, see items with certain values*
2D *find adjacent items, counting, filtering*
3D *adjacency + above/below, inside/outside*
Temporal *find events before, after, during*

Multi-D *find patterns, clusters, outliers*
Trees *how many levels, children*
Networks *shortest path between points*

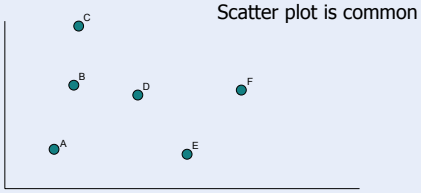
Univariate Data



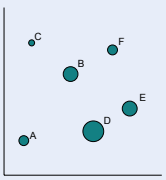
Univariate Data



Bivariate Data



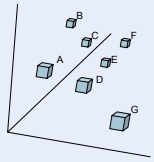
Trivariate Data



Scatter plot with retinal variable is common

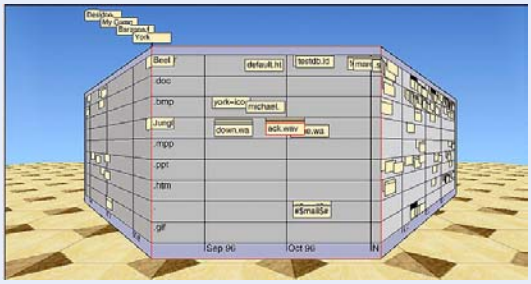
Trivariate Data

3D scatter plot is possible



Exercise: Under what conditions?

Temporal Data



Perspective Wall [Robertson+ 1993]

Summary

Information Visualization model & process

Data

Influence on representation (encodings)

Tasks & data structure/complexity

Influence on representation

Questions?