

Representation of Data

INFO 424
Lecture 5
October 10, 2006
Polle Zellweger

Overview

- ◆ Data representation depends upon
 - ❖ type of data
 - ❖ its complexity
 - ❖ the way in which the user interprets the encoded data
- ◆ Visual mappings
 - ❖ Choice of encoding
- ◆ Data complexity
 - ❖ Univariate data
 - ❖ Bivariate data
 - ❖ Trivariate data
 - ❖ Multivariate data
- ◆ Relationships will be discussed Oct 31
 - ❖ graphs, trees, networks

Fall 2006 #4

INFO 424

2

Data Tables

Variables

Cases

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

Fall 2006 #4

INFO 424

3

Types of Data Values

- ◆ *Nominal (categorical)*
 - ❖ no ordering (= or ≠ only)
 - ❖ Ex: film type: Action, Comedy, Romance, ...
- ◆ *Ordinal*
 - ❖ can be ordered in sequence (< operator)
 - ❖ Ex: film rating: G, PG, PG-13, R
- ◆ *Quantitative*
 - ❖ allow arithmetic operations
 - ❖ Ex: film length: 120 minutes

Fall 2006 #4

INFO 424

4

Data Types

Nominal

Ordinal

Quantitative

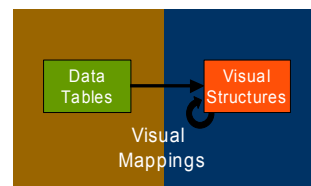
FilmID	230	105	540	...
Title	Goldfinger	Ben Hur	Ben Hur	...
Director	Hamilton	Wyler	Niblo	...
Actor	Connery	Heston	Novarro	...
Actress	Blackman	Harareet	McAvoy	...
Year	1964	1959	1926	...
Length	112	212	133	...
Popularity	7.7	8.2	7.4	...
Rating	PG	G	G	...
Film Type	Action	Action	Drama	...

Fall 2006 #4

INFO 424

5

Visual Mappings



Fall 2006 #4

INFO 424

6

Visual Structure

◆ Film Finder, University of Maryland

Nominal: Color Ordinal: List Quantitative: Axis

Fall 2006 #4 INFO 424 7

Where do you start designing?

Fall 2006 #4 INFO 424 8

Bertin's Graphical Vocabulary

Position

Marks

Points

Lines

Areas

Retinal variables

Color

Size

Shape

Grayscale

Orientation

Texture

Fall 2006 #4 INFO 424 9

Effectiveness rankings for encoding quantitative data

[Cleveland & McGill 1984, adapted from Spence 2006]

Fall 2006 #4 INFO 424 10

Space and Position

Nominal Presence
N

Nominal Grid
N

Ordinal Grid
O (A<B)

Quantitative Scale
Q (B/A=2.2)

Spatial Coordinates
Qs (A=2.5 cm)

Geographical Scale
Qg (A=24° 5' W)

Fall 2006 #4 INFO 424 11

Composition of space

- ◆ 1D axis
- ◆ 2D axis
- ◆ 3D axis
- ◆ Single axis
- ◆ Double axis
- ◆ Triple axis
- ◆ Mark composition

Fall 2006 #4 INFO 424 12

Design criteria: Expressiveness

- Show all the data
- Show only the data

Drama
Mystery
Comedy
Music
Action
War
SF
Western
Horror

Ordering?

Fall 2006 #4 INFO 424 13

Expressiveness rankings for Info Vis tasks

	Association <small>The marks can be grouped as follows</small>	Selection <small>The marks are highlighted, leaving others</small>	Order <small>The marks are ordered as follows</small>	Quantity <small>The marks are arranged in a specific order</small>
Size				
Value				
Texture				
Colour				
Orientation				
Shape				

[Bertin, adapted from Spence 2006] 14

Design criteria: Effectiveness

- faster to interpret
- more distinctions
- fewer errors

Fall 2006 #4 INFO 424 15

Effectiveness rankings

	Spatial	Q	O	N	Object	Q	O	N
Extent	Position	●	●	●	Gray Scale	◐	◑	○
	Size	●	●	●				
Differential	Orientation	◐	◑	●	Color	◐	◑	●
					Texture	◐	◑	●
					Shape	○	○	●

● Good ◐ Fair ○ Poor

Fall 2006 #4 INFO 424 16

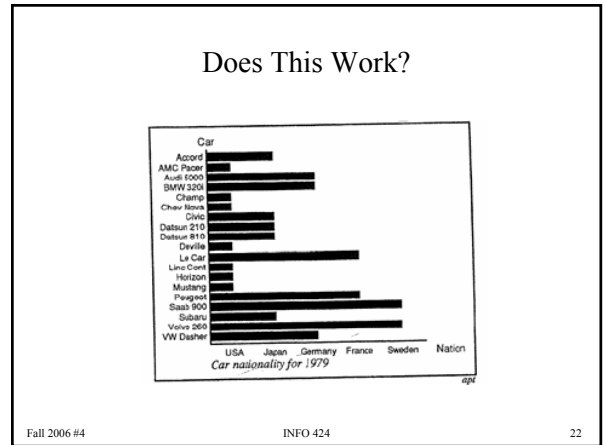
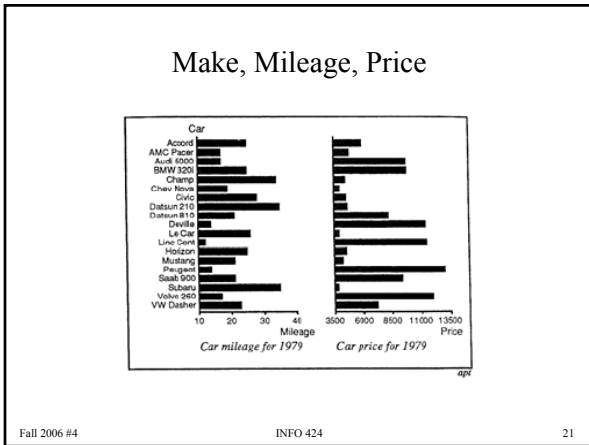
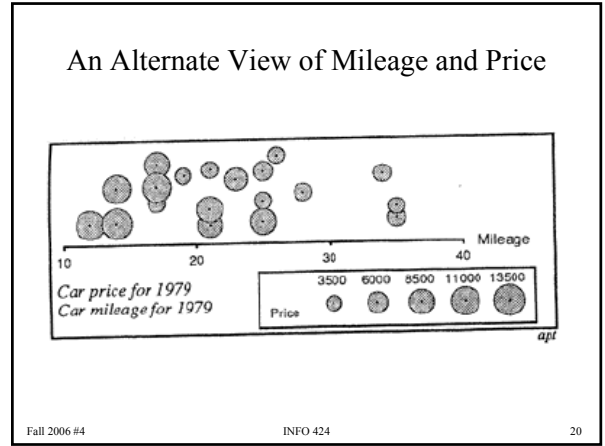
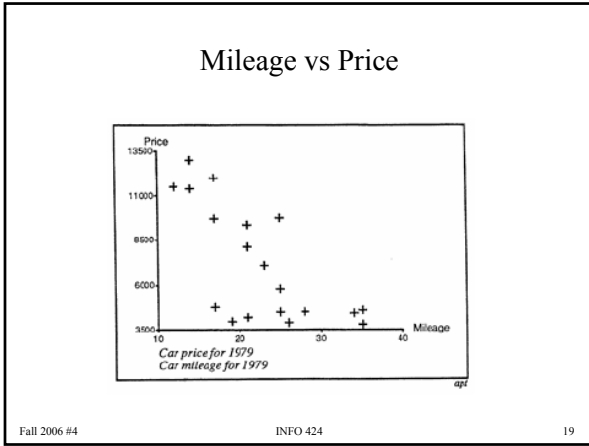
Examples

Fall 2006 #4 INFO 424 17

Automobile Relations

- Price: cars → [12000, 60000]
- Mileage: cars → [10, 40]
- Weight: cars → [1500, 5000]
- Repair: cars → [Great, Good, OK, Poor, Bad]
- Nation: cars → [USA, Germany, France, ...]
- Cars = {Accord, Audi, BMW, ...}

Fall 2006 #4 INFO 424 18



Key Idea

- ◆ How should data of various types be encoded into visual features?
- ◆ Mapping quantities into shapes does not work!
 - ❖ 10 ● 100 ■
- ◆ But using extent works well

Fall 2006 #4 INFO 424 23

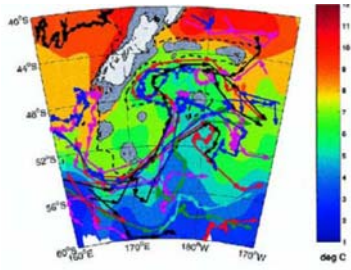
Attribute Types

Latitude	Yi	Yj	Yk
Longitude	Xi	Xj	Xk
Values	Qi	Qj	Qk

- ◆ What is the difference between...
 - ❖ Numbers that represent temperatures
 - ❖ Numbers that represent car accidents
- ◆ How could this issue be addressed?

Fall 2006 #4 INFO 424 24

Temperature: A Continuous Variable

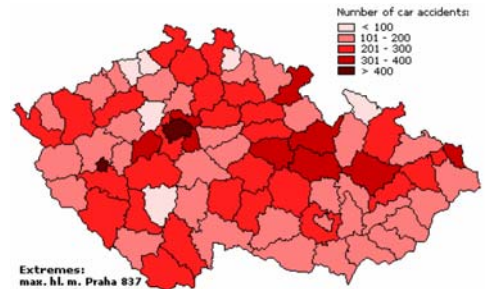


Fall 2006 #4

INFO 424

25

Car Accidents: Discrete Values



Extremes:
max. hl. m. Praha 837
min. Kladno 74

Fall 2006 #4

INFO 424

26

Data Complexity

- ◆ Univariate data
- ◆ Bivariate data
- ◆ Trivariate data
- ◆ Multivariate data

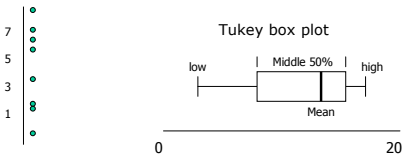
Fall 2006 #4

INFO 424

27

Univariate Data

	cases			
	A	B	C	
1				variable



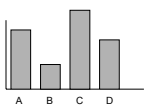
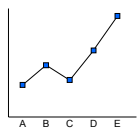
Fall 2006 #4

INFO 424

28

Univariate Data

	cases			
	A	B	C	
1				variable



Price €k 10 - 12 12 - 14 16 - 18

Fall 2006 #4

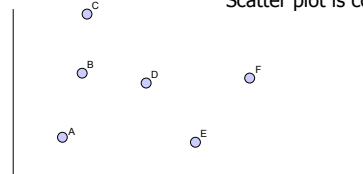
INFO 424

29

Bivariate Data

	cases			
	A	B	C	
1				
2				

Scatter plot is common



Fall 2006 #4

INFO 424

30

Trivariate Data

	A	B	C
1			
2			
3			

Scatter plot with retinal variable is common

Fall 2006 #4 INFO 424 31

Trivariate Data

	A	B	C
1			
2			
3			

3D scatter plot is possible

Exercise: Under what conditions?

Fall 2006 #4 INFO 424 32

Multivariate or Hypervariate Data

	A	B	C
1			
2			
3			
4			
5			
6			
7			
8			

Fall 2006 #4 INFO 424 33

Multivariate Data

- ◆ ≤ 3 variables can be mapped directly to space
 - ❖ line graphs, bar graphs, scatter plots
 - ❖ We see a 3-D world (4-D with time)
- ◆ With more than 3 variables we add:
 - ❖ Time
 - ❖ Retinal Variables (color, size, shape, ...)
- ◆ What about data sets with many variables?
 - ❖ Often the interesting ones
 - ❖ Are there limits to how many variables you make visual?

Fall 2006 #4 INFO 424 34

Bertin's Synoptic

Syn•op•tic "affording a general view of a whole"
Merriam Webster Dictionary

"With up to three rows, a data table can be constructed directly as a single image ... However, an image has only three dimensions. And this barrier is impassible."

Bertin's two choices for multiple variables:

- ❖ Construct several images and sacrifice the overall relationship
- ❖ Construct a matrix and discover relationships through permutation

Fall 2006 #4 INFO 424 35

Approach 1: Multiple Views

Give each variable its own display

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

Fall 2006 #4 INFO 424 36

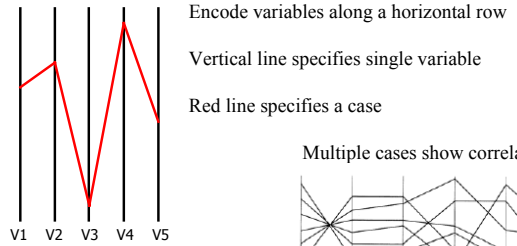
Approach 2: Matrix TableLens demo

- ◆ Explicit variables and cases
 - ❖ To see (up to 1000 rows x 50 columns)
 - ❖ To manipulate (sort, rearrange, focus, filter)
- ◆ Focus+context
- ◆ Spotlights
- ◆ Demo

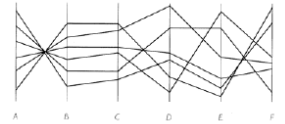
Fall 2006 #4

37

Explicit Variables: Parallel Coordinates (Inselberg)



Multiple cases show correlations



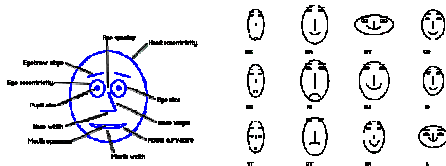
Fall 2006 #4

INFO 424

38

Chernoff Faces

Encode different variables' values in characteristics of human face



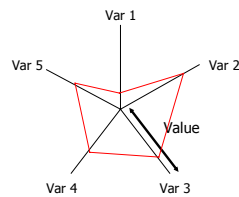
Cute applet: <http://people.cs.uchicago.edu/~wiseman/chernoff/>

Fall 2006 #4

INFO 424

39

Star Plots



Space out the n variables at equal angles around a circle

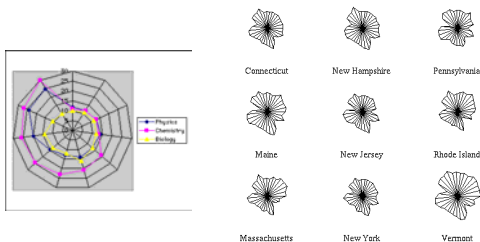
Each "spoke" encodes a variable's value

Fall 2006 #4

INFO 424

40

Star Plot examples



<http://seamonkey.ed.asu.edu/~behrens/asu/reports/compre/compl.html>

Fall 2006 #4

INFO 424

41

Summary: Data Representation

- ◆ Visual mappings
- ◆ Data complexity
 - ❖ univariate -> multivariate data

Syn•op•tic "affording a general view of a whole"

Merriam Webster Dictionary

- ◆ Questions
 - ❖ How many variables can we visualize?
 - ❖ How many cases can we visualize?
 - ❖ How often do we need to see the whole?
 - ❖ When do we need to see the whole?

Fall 2006 #4

INFO 424

42

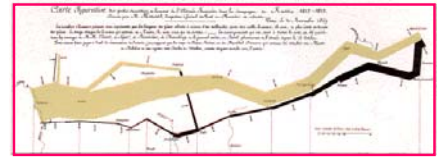
Analyzing Minard's Napoleonic Map

Fall 2006 #4

INFO 424

43

Single axis composition



+



=



Fall 2006 #4

INFO 424

44

Mark Composition

temperature

+ time

+ temp[day]

=



Fall 2006 #4

INFO 424

45

Mark composition

longitude

+ latitude

+ army[size, day]

+ army[position, day]

=



Fall 2006 #4

INFO 424

46

longitude

latitude

army[size, day]

army[position, day]

temperature

time

temp[day]



INFO 424

47

Fall 2006 #4

Minard's data table and synoptic

	Day1	Day2
Army size		
Army longitude		
Army latitude		
Temperature		

Issue: seeing the relationships



Fall 2006 #4

INFO 424

48

Sources Used

- ◆ CMS Chapter 1
- ◆ Spence textbook 2nd edition 2006
- ◆ Jock Mackinlay
 - ❖ U of Aarhus InfoVis course Fall 2000
- ◆ Dave Hendry
 - ❖ UW Info 424 Autumn 2005
- ◆ John Stasko
 - ❖ Georgia Tech InfoVis course Spring 2006