Data and Image Models

Maneesh Agrawala

CS 558: Visualization
Winter 2005

Lecture adapted from Hanrahan 2004

The big picture

task

data
- physical type
  - int, float, etc.
- abstract type
  - nominal, ordinal, etc.

domain
- metadata
- semantics
- conceptual model

processing algorithms

→

mapping
- visual encoding
- visual metaphor

image
- visual channel
- retinal variables

[based on slide from Munzner]

Topics

- Properties of data or information
- Properties of the image
- Mapping data to images


Jacques Bertin

Data

Taxonomy by data type

- 1D (sets and sequences)
- Temporal
- 2D (maps)
- 3D (shapes)
- nD (relational)
- Trees (hierarchies)
- Networks (graphs)

Are there others?

The eyes have it: A task by data type taxonomy for information visualization [Schneiderman 96]

Data models vs. Conceptual models

Data models are mathematical abstractions

- Sets with operations on them
- Example: integers with + and × operators

Conceptual models are mental constructions

- Include semantics and support reasoning
- Example: navigating through a city using landmarks

Examples (data vs. conceptual)

- 1D vs. Time
- nD vs. Space

Relational data model

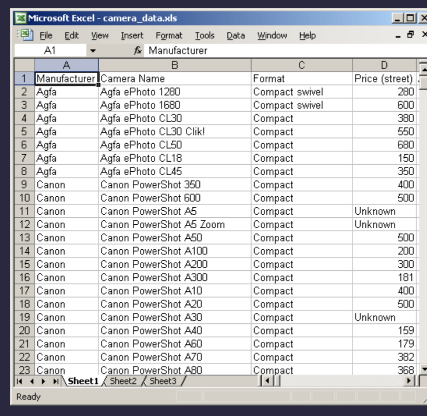
Records are fixed-length tuples

Each column (attribute) of tuple has a domain (type)

Relation is schema and a table of tuples

Database is a collection of relations

Example: Digital Cameras



	A	B	C	D
1	Manufacturer	Camera Name	Format	Price (street)
2	Agfa	Agfa ePhoto 1280	Compact swivel	280
3	Agfa	Agfa ePhoto 1680	Compact swivel	600
4	Agfa	Agfa ePhoto CL30	Compact	380
5	Agfa	Agfa ePhoto CL30 Clik!	Compact	550
6	Agfa	Agfa ePhoto CL50	Compact	680
7	Agfa	Agfa ePhoto CL18	Compact	150
8	Agfa	Agfa ePhoto CL45	Compact	350
9	Canon	Canon PowerShot 350	Compact	400
10	Canon	Canon PowerShot 600	Compact	500
11	Canon	Canon PowerShot A5	Compact	Unknown
12	Canon	Canon PowerShot A5 Zoom	Compact	Unknown
13	Canon	Canon PowerShot A50	Compact	500
14	Canon	Canon PowerShot A100	Compact	200
15	Canon	Canon PowerShot A200	Compact	300
16	Canon	Canon PowerShot A300	Compact	181
17	Canon	Canon PowerShot A10	Compact	400
18	Canon	Canon PowerShot A20	Compact	500
19	Canon	Canon PowerShot A30	Compact	Unknown
20	Canon	Canon PowerShot A40	Compact	159
21	Canon	Canon PowerShot A60	Compact	179
22	Canon	Canon PowerShot A70	Compact	382
23	Canon	Canon PowerShot A80	Compact	368

Relational algebra [Codd]

Data transformations (SQL)

- Selection (SELECT)
- Projection (WHERE)
- Sorting (ORDER BY)
- Aggregation (GROUP BY, SUM, MIN, …)
- Set operations (UNION, …)
- Join (INNNER JOIN)

Statistical data model

Variables or measurements (~ float attribute)
Categories or factors (~ int attribute)
Observations or cases (~ record)

Microsoft Excel - fischer.iris.1.xls

ID	Case	Species_No	Species	Petal_width	Petal_length	Sepal_width	Sepal_length
1	1	1	I. Setosa	2	14	33	50
2	1	3	I. Verginica	24	56	31	67
3	1	2	I. Versicolor	13	45	28	57
4	2	1	I. Setosa	2	10	36	46
5	2	3	I. Verginica	23	51	31	69
6	2	2	I. Versicolor	16	47	33	63
7	3	1	I. Setosa	2	16	31	48
8	3	3	I. Verginica	20	52	30	65
9	3	2	I. Versicolor	14	47	32	70
10	4	1	I. Setosa	1	14	36	49
11	4	3	I. Verginica	19	51	27	58
12	4	2	I. Versicolor	12	40	26	58
13	5	1	I. Setosa	2	13	32	44
14	5	3	I. Verginica	17	45	25	49
15	5	2	I. Versicolor	10	33	23	50
16	6	1	I. Setosa	2	16	38	51
17	6	3	I. Verginica	19	50	25	63
18	6	2	I. Versicolor	10	41	27	58
19	7	1	I. Setosa	2	16	30	50
20	7	3	I. Verginica	18	49	27	63
21	7	2	I. Versicolor	15	45	29	60
22	8	1	I. Setosa	4	19	38	51
23	8	3	I. Verginica	21	56	28	64
24	8	2	I. Versicolor	10	33	24	49
25	9	1	I. Setosa	2	14	30	49
26	9	3	I. Verginica	19	51	27	58
27	9	2	I. Versicolor	14	39	27	52
28	10	1	I. Setosa	2	14	36	50
29	10	3	I. Verginica	18	55	31	64
30	10	2	I. Versicolor	12	39	27	58
31	11	1	I. Setosa	4	15	34	54

Sepal and petal lengths and widths for three species of iris [Fisher].

Microsoft Excel - fischer.iris.2.xls

ID	Case	Species_No	Species	Organ	Width	Length
1	1	1	I. Setosa	Petal	2	14
2	1	3	I. Verginica	Petal	24	56
3	1	2	I. Versicolor	Petal	13	45
4	1	1	I. Setosa	Sepal	33	50
5	1	3	I. Verginica	Sepal	31	67
6	1	2	I. Versicolor	Sepal	28	57
7	2	1	I. Setosa	Petal	2	10
8	2	3	I. Verginica	Petal	23	51
9	2	2	I. Versicolor	Petal	16	47
10	2	1	I. Setosa	Sepal	36	46
11	2	3	I. Verginica	Sepal	31	69
12	2	2	I. Versicolor	Sepal	33	63
13	3	1	I. Setosa	Petal	2	16
14	3	3	I. Verginica	Petal	20	52
15	3	2	I. Versicolor	Petal	14	47
16	3	1	I. Setosa	Sepal	31	48
17	3	3	I. Verginica	Sepal	30	65
18	3	2	I. Versicolor	Sepal	32	70
19	4	1	I. Setosa	Petal	1	14
20	4	3	I. Verginica	Petal	19	51
21	4	2	I. Versicolor	Petal	12	40
22	4	1	I. Setosa	Sepal	36	49
23	4	3	I. Verginica	Sepal	27	58
24	4	2	I. Versicolor	Sepal	26	58
25	5	1	I. Setosa	Petal	2	13
26	5	3	I. Verginica	Petal	17	45
27	5	2	I. Versicolor	Petal	10	33
28	5	1	I. Setosa	Sepal	32	44
29	5	3	I. Verginica	Sepal	25	49
30	5	2	I. Versicolor	Sepal	23	50
31	6	1	I. Setosa	Petal	2	16

fischer.iris

Sepal and petal lengths and widths for three species of iris [Fisher].

Microsoft Excel - fischer.iris.3.xls

ID	Case	Species	Organ	Measure	Value
1	1	I. Setosa	petal	width	2
2	1	I. Verginica	petal	width	24
3	1	I. Versicolor	petal	width	13
4	1	I. Setosa	petal	length	14
5	1	I. Verginica	petal	length	56
6	1	I. Versicolor	petal	length	45
7	1	I. Setosa	sepal	width	33
8	1	I. Verginica	sepal	width	31
9	1	I. Versicolor	sepal	width	28
10	1	I. Setosa	sepal	length	50
11	1	I. Verginica	sepal	length	67
12	1	I. Versicolor	sepal	length	57
13	2	I. Setosa	petal	width	2
14	2	I. Verginica	petal	width	23
15	2	I. Versicolor	petal	width	16
16	2	I. Setosa	petal	length	10
17	2	I. Verginica	petal	length	51
18	2	I. Versicolor	petal	length	47
19	2	I. Setosa	sepal	width	36
20	2	I. Verginica	sepal	width	31
21	2	I. Versicolor	sepal	width	33
22	2	I. Setosa	sepal	length	46
23	2	I. Verginica	sepal	length	69
24	2	I. Versicolor	sepal	length	63
25	3	I. Setosa	petal	width	2
26	3	I. Verginica	petal	width	20
27	3	I. Versicolor	petal	width	14
28	3	I. Setosa	petal	length	16
29	3	I. Verginica	petal	length	52
30	3	I. Versicolor	petal	length	47
31	3	I. Setosa	sepal	width	31

Sheet1

Sepal and petal lengths and widths for three species of iris [Fisher].

	I. Setosa				I. Verginica				I. Versicolor			
	petal		sepal		petal		sepal		petal		sepal	
	length	width	length	width	length	width	length	width	length	width	length	width
1	14	2	50	33	56	24	67	31	45	13	57	28
2	10	2	46	36	51	23	69	31	47	16	63	33
3	16	2	48	31	52	20	65	30	47	14	70	32
4	14	1	49	36	51	19	58	27	40	12	58	26
5	13	2	44	32	45	17	49	25	33	10	50	23
6	16	2	51	38	50	19	63	25	41	10	58	27
7	16	2	50	30	49	18	63	27	45	15	60	29
8	19	4	51	38	56	21	64	28	33	10	49	24
9	14	2	49	30	51	19	58	27	39	14	52	27
10	14	2	50	36	55	18	64	31	39	12	58	27
11	15	4	54	34	50	15	60	22	42	15	59	30
12	14	2	55	42	57	23	69	32	44	13	63	23
13	14	2	44	29	49	20	56	28	49	15	63	25
14	14	1	48	30	58	18	67	25	30	11	51	25
15	17	3	57	38	54	21	69	31	36	13	56	29
16	15	4	51	37	61	25	72	36	44	14	66	30
17	13	2	55	35	55	21	68	30	50	17	67	30
18	13	2	44	30	56	22	64	28	45	15	62	22
19	16	2	47	32	51	15	63	28	46	14	61	30
20	12	2	50	32	59	23	68	32	39	11	56	25

Format of the data in Appendix 14, pp. 365-366

Chambers, Cleveland, Kleiner, Tukey, *Graphical Methods for Data Analysis*

Types

Physical types

- **Characterized by storage**
- **Characterized by machine operations**

Example:

bool, short, int32, float, double, string, …

Abstract types

- **Characterized by methods/attributes**
- **Organized into a class hierarchy**

Example:

nominal, ordinal, cardinal, …,
plants, animals, metazoans, …

Measurements

N - Nominal (labels)

- Fruits: Apples, oranges, …

O - Ordered

- Days: Mon, Tue, Wed, Thu, Fri, Sat, Sun
- Quality of meat: Grade A, AA, AAA

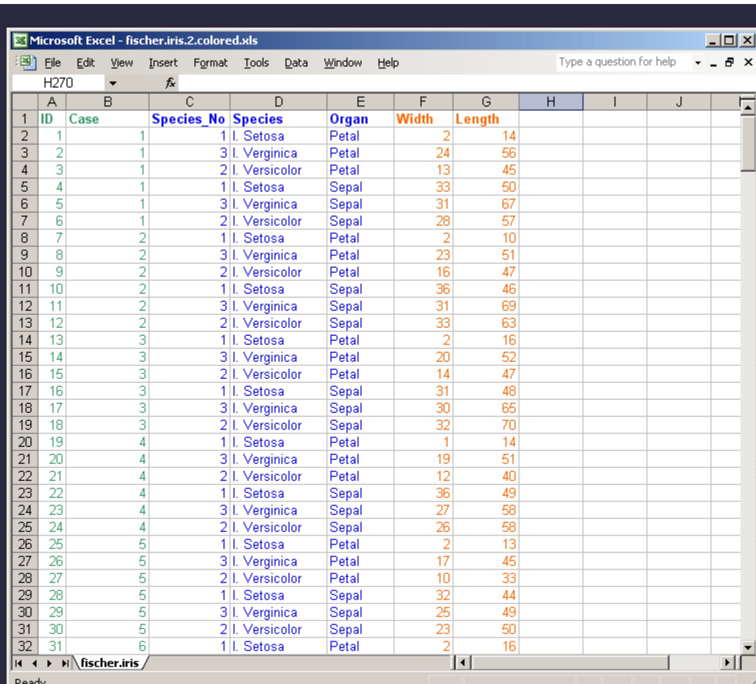
Q - Interval (Location of zero arbitrary)

- Periods of time: second, minute, …

Q - Ratio (zero fixed)

- Counts
- Physical measurement: Kelvin, …

S. S. Stevens, On the theory of scales of measurements, 1946



ID	Case	Species_No	Species	Organ	Width	Length
1	1	1	I. Setosa	Petal	2	14
2	1	3	I. Verginica	Petal	24	56
3	1	2	I. Versicolor	Petal	13	45
4	1	1	I. Setosa	Sepal	33	50
5	1	3	I. Verginica	Sepal	31	67
6	1	2	I. Versicolor	Sepal	28	57
7	2	1	I. Setosa	Petal	2	10
8	2	3	I. Verginica	Petal	23	51
9	2	2	I. Versicolor	Petal	16	47
10	2	1	I. Setosa	Sepal	36	46
11	2	3	I. Verginica	Sepal	31	69
12	2	2	I. Versicolor	Sepal	33	63
13	3	1	I. Setosa	Petal	2	16
14	3	3	I. Verginica	Petal	20	52
15	3	2	I. Versicolor	Petal	14	47
16	3	1	I. Setosa	Sepal	31	48
17	3	3	I. Verginica	Sepal	30	65
18	3	2	I. Versicolor	Sepal	32	70
19	4	1	I. Setosa	Petal	1	14
20	4	3	I. Verginica	Petal	19	51
21	4	2	I. Versicolor	Petal	12	40
22	4	1	I. Setosa	Sepal	36	49
23	4	3	I. Verginica	Sepal	27	58
24	4	2	I. Versicolor	Sepal	26	58
25	5	1	I. Setosa	Petal	2	13
26	5	3	I. Verginica	Petal	17	45
27	5	2	I. Versicolor	Petal	10	33
28	5	1	I. Setosa	Sepal	32	44
29	5	3	I. Verginica	Sepal	25	49
30	5	2	I. Versicolor	Sepal	23	50
31	6	1	I. Setosa	Petal	2	16

Q
O
N

Dimensions and measures

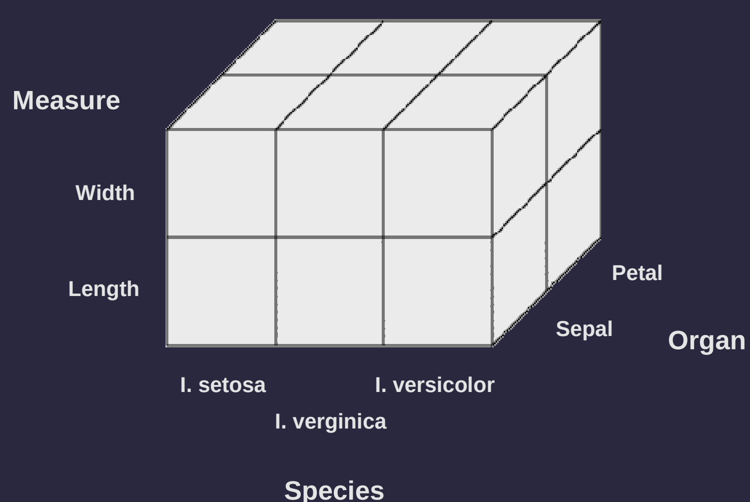
Independent vs. dependent variables

- Example: y = f(x,a)
- Dimensions: Domain(x) × Domain(a)
- Measures: Range(y)

Common techniques for summarizing data

- GroupBy dimensions
- Aggregate measures

Data cube



Summary of basic properties

- Multidimensional
  - Number of columns (dimensions)
- Type
  - Nominal, ordinal, quantitative
- Cardinality (levels)
  - Number of values possible within a column

Image

Visual language is a sign system

Image is perceived as a set of signs

Sender encodes information in these signs

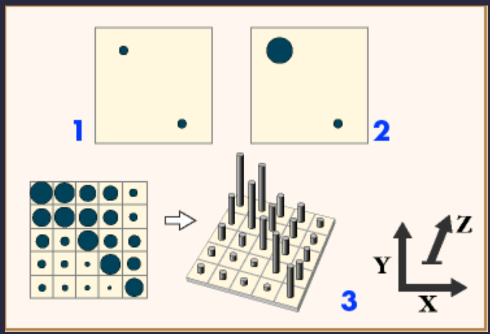
Receiver decodes information from these signs

Visual variables

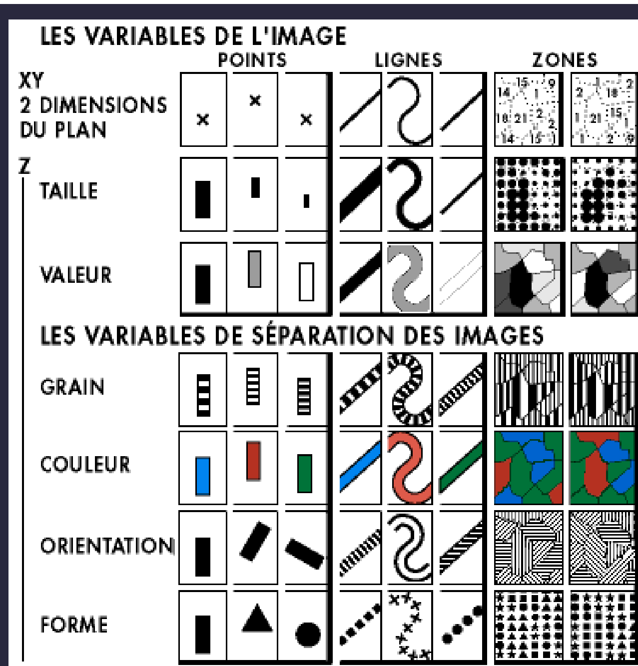
J. Bertin, Semiology of Graphics, 1967

[x,y]
- Position

[z]
- Size
- Value
- Color
- Texture
- Orientation
- Shape

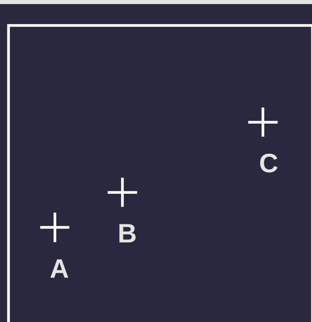


Note: Bertin does not consider 3D or time
Note: Card and Mackinlay extend the number of vars.



[Bertin, Semiology of Graphics, 1983]

Information in position



1. A, B, C are distinguishable
2. B is between A and C.
3. BC is twice as long as AB.

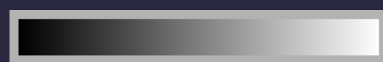
"Resemblance, order and proportional are the three signfields in graphics." - Bertin

Information in color and value

Value is perceived as ordered

∴ Encode ordinal variables (O)

∴ Encode continuous variables (Q) [not as well]



Hue is normally perceived as unordered

∴ Encode nominal variables (N) using color

Bertins' "Levels of Organization"

Position	N	O	Q
Size	N	O	Q
Value	N	O	Q
Texture	N	O	
Color	N		
Orientation	N		
Shape	N		

N Nominal
O Ordered
Q Quantitative

Note: Q < O < N

Note: Bertin actually breaks visual variables down into differentiating (≠) and associating (≡)

Encoding rules

Depicting information

2 variables can be mapped directly to space

- Graphs, maps, …

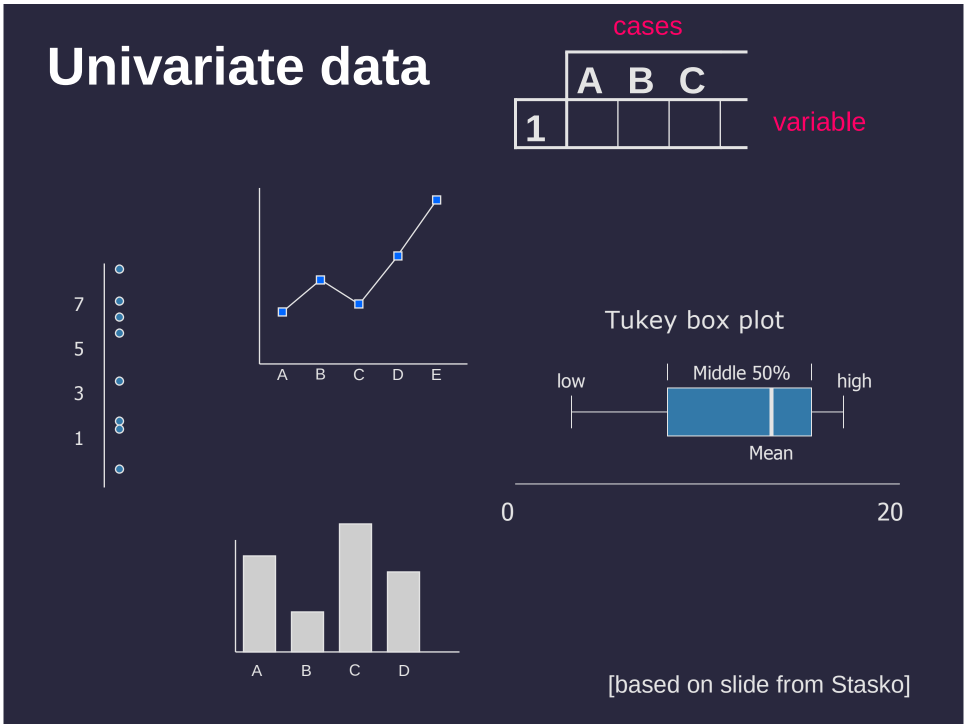
Other variables must be a visual variables

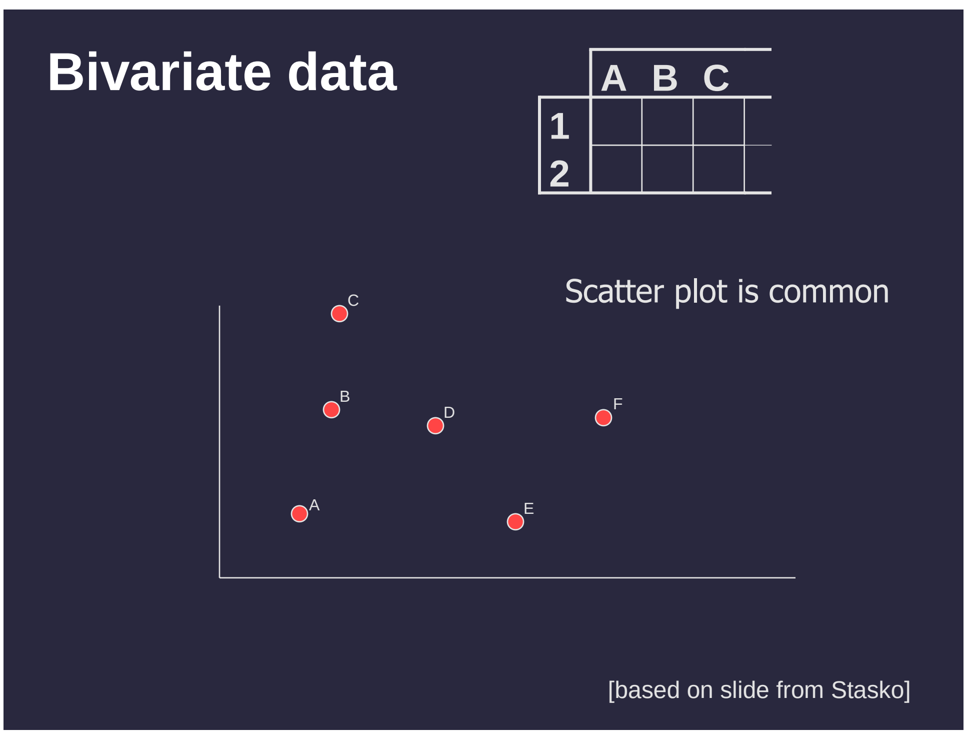
- Color, size, shape, …

What about data sets with many variables?

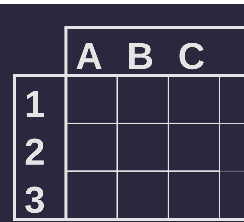
- How many variables may be shown?
- What are the best ways to show those variables?

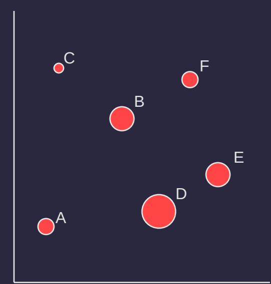
[Exclude 3D and T for now …]


Univariate data
cases
A B C
1
variable
7
5
3
1
A B C D E
Tukey box plot
low
Middle 50%
high
Mean
0
20
A B C D
[based on slide from Stasko]

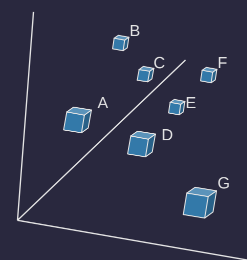

Bivariate data
A B C
1
2
Scatter plot is common
C
B
D
F
A
E
[based on slide from Stasko]

Trivariate data



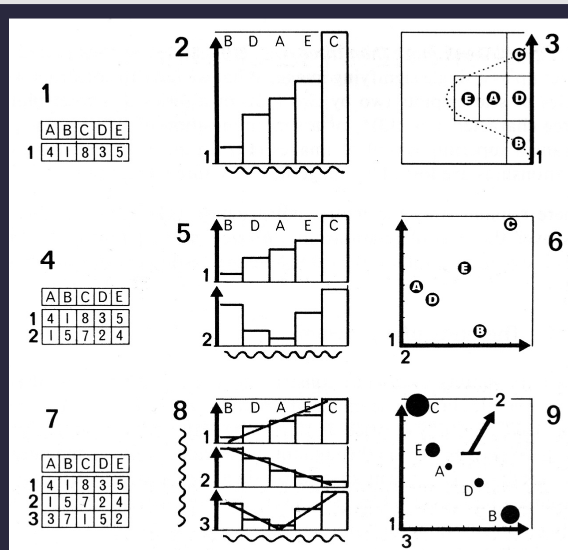


3D scatter plot is possible



[based on slide from Stasko]

Large design space (visual metaphors)



[Bertin, Graphics and Graphic Info. Processing, 1981]

Multidimensional data

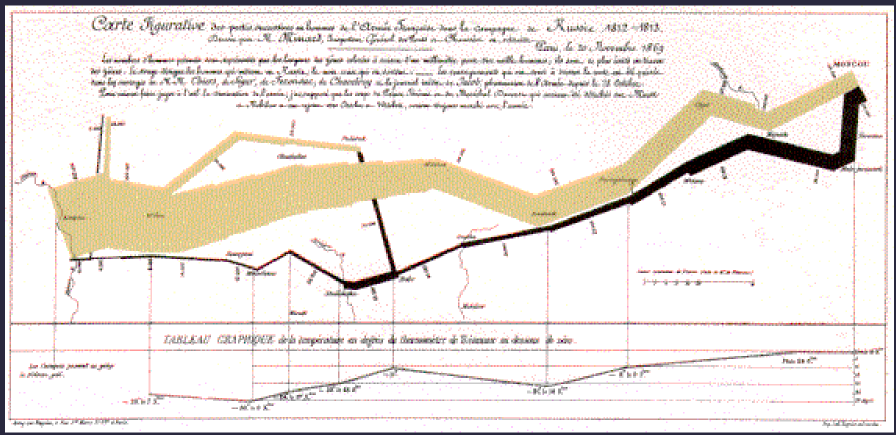
"With up to three rows, a data table can be constructed directly as a single image … However, an image has only three dimensions. And this barrier is impassible."

Bertin

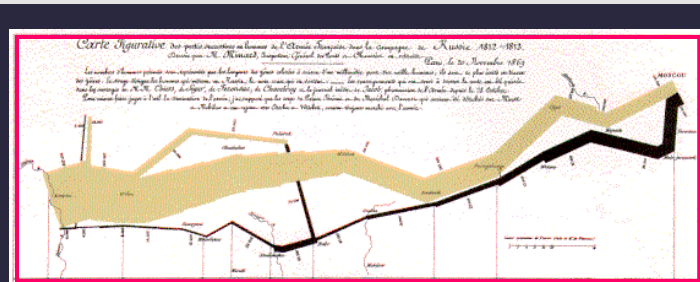
	A	B	C	
1				
2				
3				
4				
5				
6				
7				
8				

Composition/Decomposition

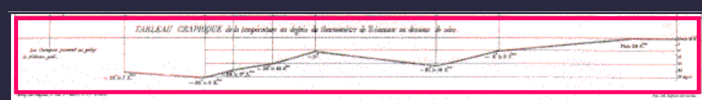
Minard's 1869 Napoleon's march



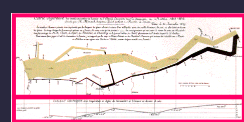
Single axis composition



+

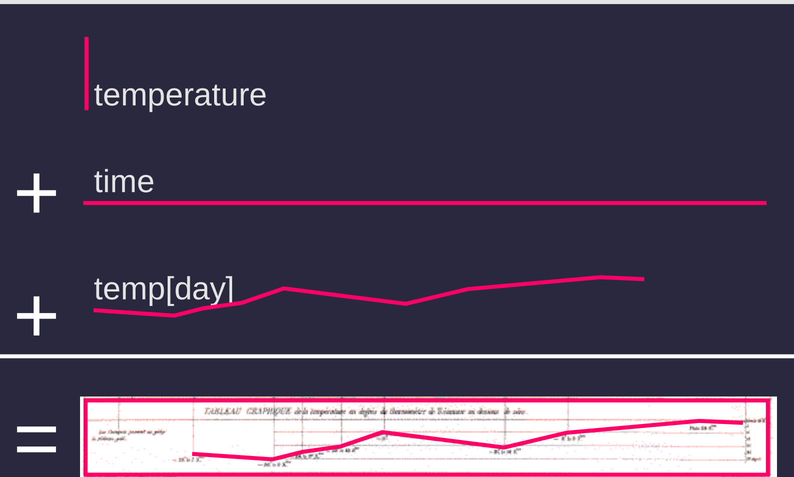


=

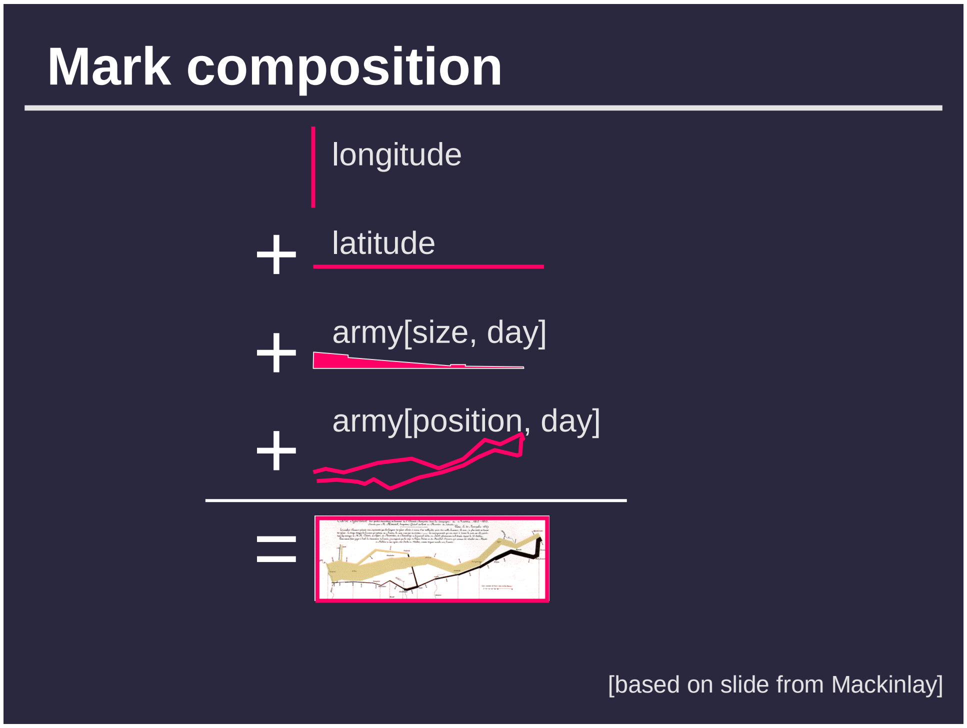


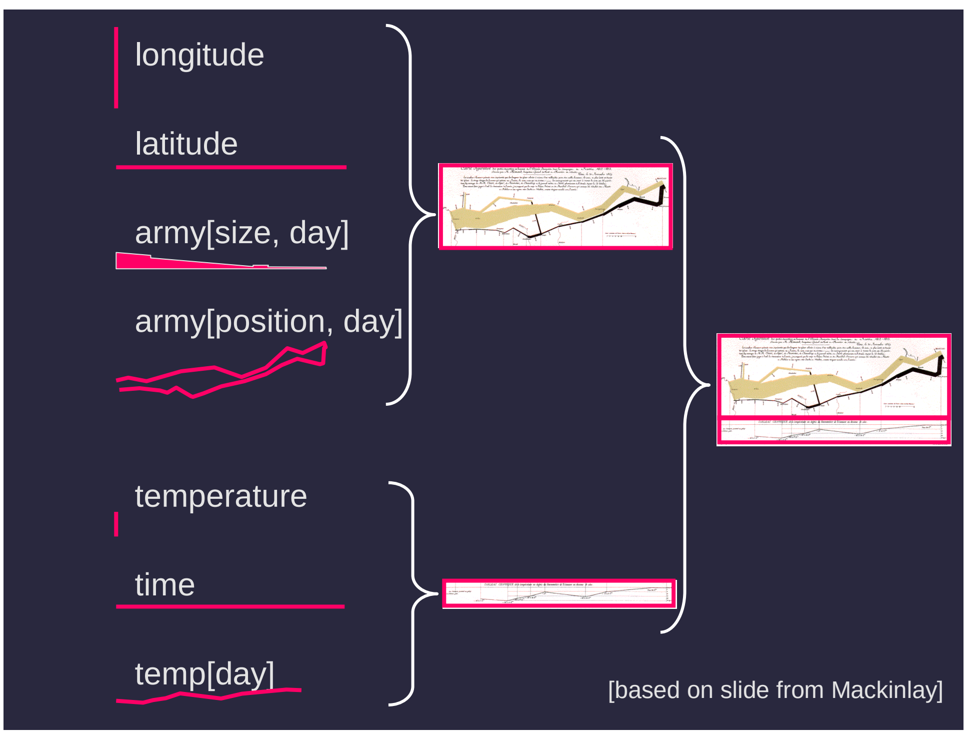
[based on slide from Mackinlay]

Mark composition



[based on slide from Mackinlay]


Mark composition
longitude
+ latitude
+ army[size, day]
+ army[position, day]
=
[based on slide from Mackinlay]


longitude
latitude
army[size, day]
army[position, day]
temperature
time
temp[day]
[based on slide from Mackinlay]

Automated design

J. Mackinlay's APT 86

Combinatorics of encodings

Challenge:

Pick the best encoding from the exponential number of possibilities $(n+1)^8$

Principle of Consistency:

The properties of the image should match the properties of the data.

Principle of Importance Ordering:

Encode the most important information in the most effective way.

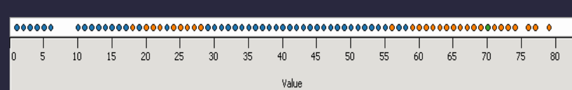
Mackinlay’s expressiveness criteria

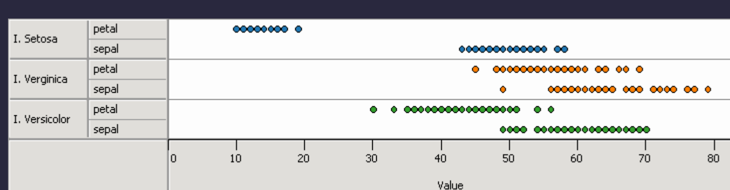
Expressiveness

A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Cannot express the facts

A one-to-many (1 → N) relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position





Expresses facts not in the data

A length is interpreted as a quantitative value;

∴ Length of bar says something untrue about N data

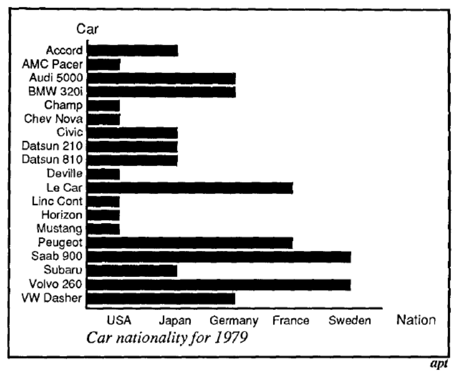


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[Mackinlay, APT, 1986]

Mackinlay’s effectiveness criteria

Effectiveness

> **A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.**

Subject of next lecture

Mackinlay ranking

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Val)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Val)
Density (Val)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Conjectured *effectiveness* of the encoding

Mackinlay's design algorithm

- User formally specifies data model
- APT searches over design space
  - Tests expressiveness of each visual encoding
  - Generates image for encodings that pass test
  - Tests perceptual effectiveness of resulting image
- Outputs most effective visualization

Summary

- Formal specification
  - Data model
  - Image model
  - Encodings mapping data to image
- Choose expressive and effective encodings
  - Formal test of expressiveness
  - Experimental tests of perceptual effectiveness