From Data To Image

Pat Hanrahan

Topics

The properties of the data or information
The properties of the image
The rules mapping data to images

Bertin 101

The Data

Taxonomy by Data Type

- 1D (sets and sequences)
- Temporal
- 2D (maps)
- 3D (shapes)
- nD (relational)
- Trees (hierarchies)
- Networks (graphs)
- Text and documents [mine]

B. Schneiderman, The eyes have it: A task by data type taxonomy for information visualization, 1996

Data Models vs. Conceptual Models

Data models are mathematical abstractions

- Sets with operations on them

  For example, integers with + and × operators

Conceptual models are mental constructions

- Include semantics and support reasoning

  For example, navigating through a city using landmarks

Examples (data vs. conceptual):

1D vs. Time

nD vs. Space

Types of Data Models

Discrete

- Relations
- Topology

Continuous

- Fields*
- Manifolds

* Treinish, A function-based datamodel for visualization

Relational Data Model

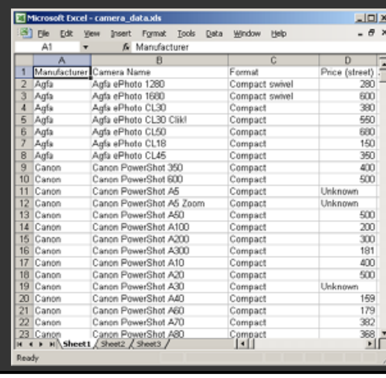
Records are fixed-length tuples

Each column of a tuple has a domain (type)

Relation is a schema plus 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 (INNER JOIN)

Statistical Data Model

Variables or Measurements
Categories or Factors
Observations or Cases

Month	Control	Placebo	300 mg	450 mg
March	165	163	166	168
April	162	159	161	163
May	164	158	161	153
June	162	161	158	160
July	166	158	160	148
August	163	158	157	150

Blood Pressure Study (4 treatments, 6 months)

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

fischer.iris

Ready

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

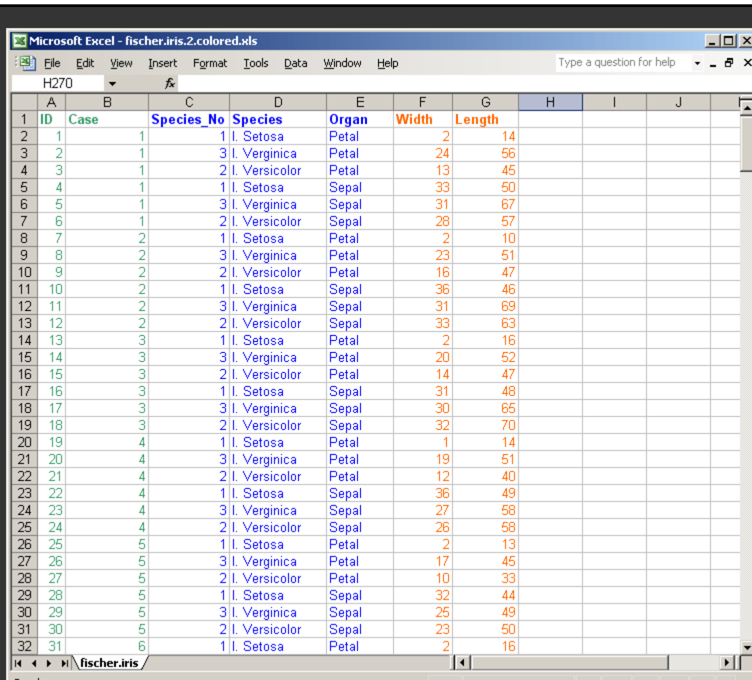
N - Nominal (labels or types)
- Fruits: Apples, oranges, ...

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

Q - Interval (Location of 0 arbitrary)
- Periods of time: second, minute, ...

Q - Ratio (0 fixed)
- Counts
- Physical measurement: Kelvin, L, M, R, ...

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



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

Q
O
N

Dimensions and Measures

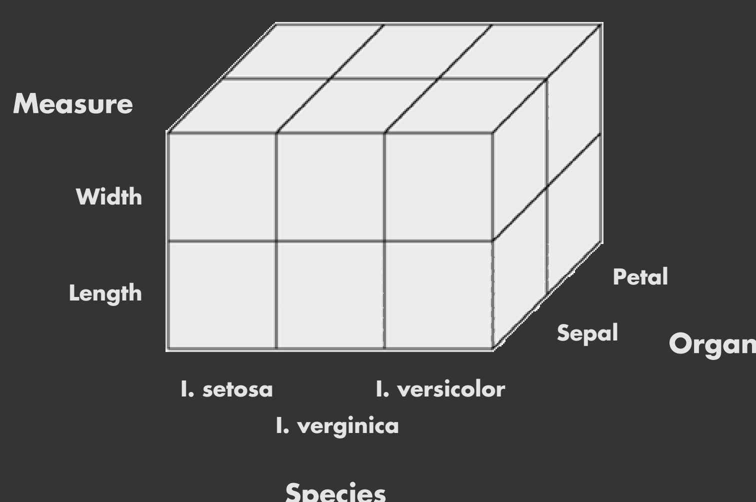
Independent vs. dependent variables

- Example: y=f(x,a)
- Infer causality
  - Response ~ factors
  - Functional dependency in databases [Ullman]

Extrinsic vs. intrinsic variables

- Example: mass vs. density (mass/vol)
- Summarize
  - Groupby dimensions and aggregate measures

Data Cube



Summary of Basic Properties

- Multidimensional
  - Number of columns
- Type
  - Type of column (N, O, Q)
- Cardinality (levels)
  - Number of different column values

The Image

Image Information

Graphical primitives and attributes (Marks)

- Attributes are parameters that control the appearance of geometric primitives

Visual channels

- Separable channels of information flowing from the retina to the brain

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

8 Visual Variables

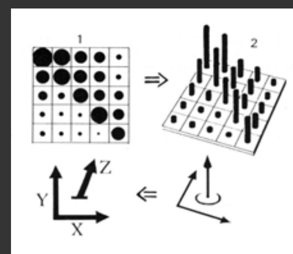
J. Bertin, Semiology of Graphics, 1967

[x,y]

- Position

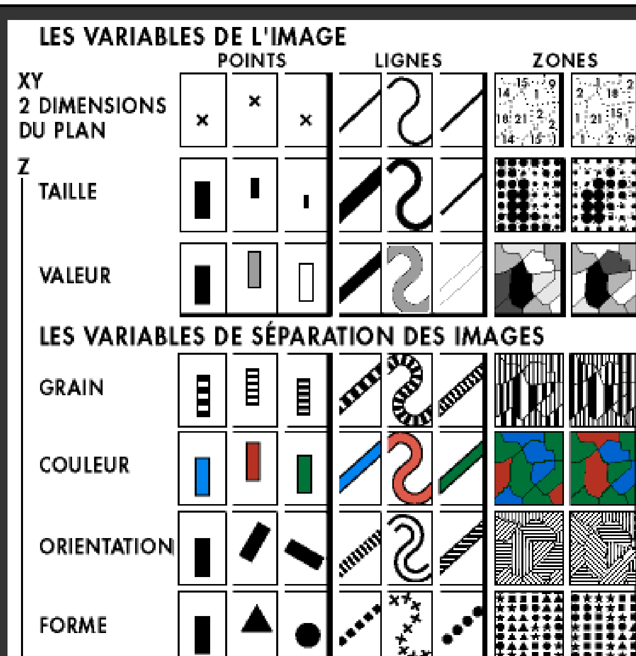
[z]

- Size
- Value
- Color
- Texture
- Orientation
- Shape



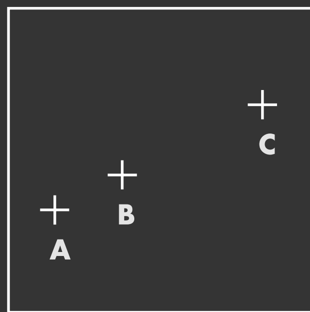
[Bertin, Graphics, 1983]

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



[Bertin, 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		

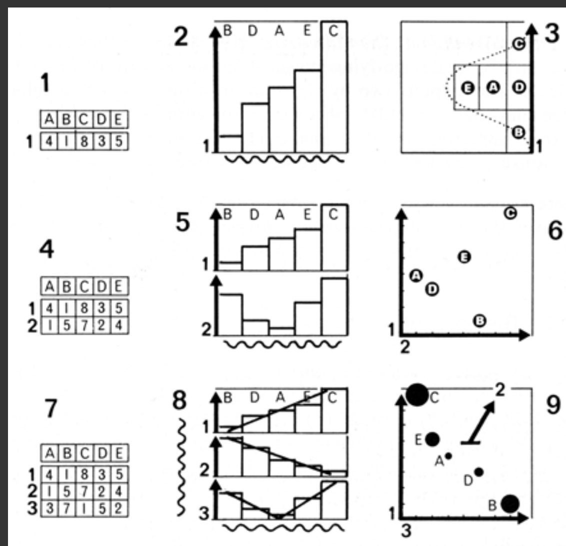
N Nominal
O Ordered
Q Quantitative

Note: Q<O<N

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

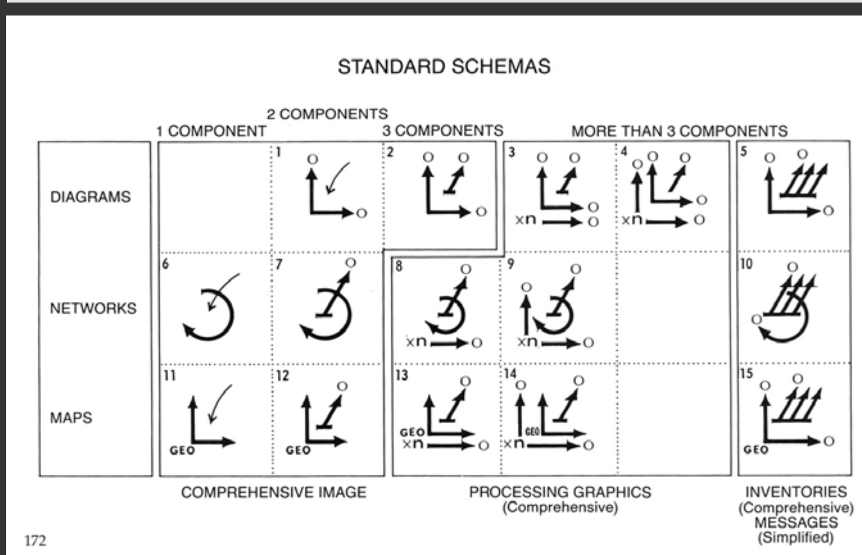
The Rules

Design Space = Visual Metaphors



[Bertin, Semiology, 1967]

Bertin's Specification



[Bertin, Semiology, 1967]

Polaris

C. Stolte

Fields Create Tables and Graphs

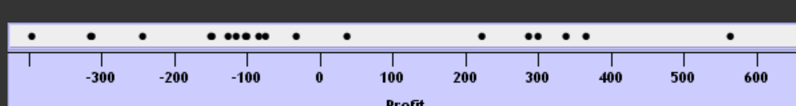
Ordinal fields: interpret field as a sequence that partitions table into rows and columns:

Quarter = {(Qtr1),(Qtr2),(Qtr3),(Qtr4)} →

Qtr1	Qtr2	Qtr3	Qtr4
95892	101760	105282	98225

Quantitative fields: treat field as single element sequence and encode as an axis:

Profit = {(Profit)} →



Union (+) Operator

Quarter + ProductType

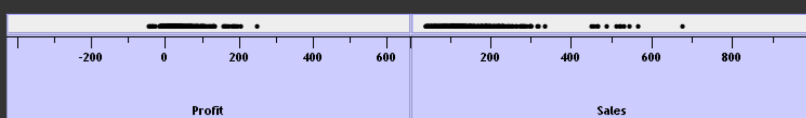
= {(Qtr1),(Qtr2),(Qtr3),(Qtr4)}+{(Coffee),(Espresso)}

= {(Qtr1),(Qtr2),(Qtr3),(Qtr4),(Coffee),(Espresso)}

Qtr1	Qtr2	Qtr3	Qtr4	Coffee	Espresso
48	59	57	53	151	21

Profit + Sales

= {(Profit),(Sales)}



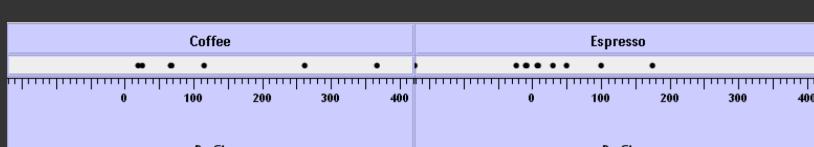
Cross (×) Operator

Quarter × ProductType

= {(Qtr1, Coffee), (Qtr1, Tea), (Qtr2, Coffee), (Qtr2, Tea), (Qtr3, Coffee), (Qtr3, Tea), (Qtr4, Coffee), (Qtr4, Tea)}

Qtr1		Qtr2		Qtr3		Qtr4	
Coffee	Espresso	Coffee	Espresso	Coffee	Espresso	Coffee	Espresso
131	19	160	20	178	12	134	33

ProductType × Profit = {(Coffee, Profit), (Tea, Profit)}



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.

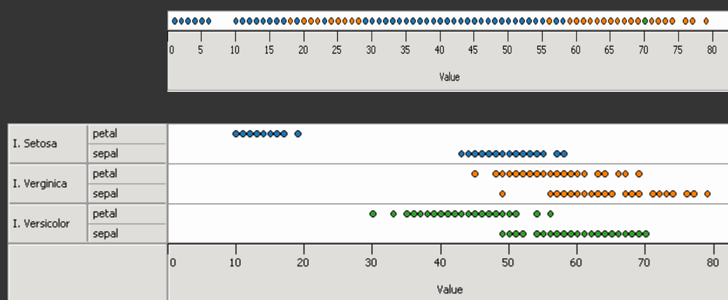
Mackinlay's Expressiveness Criteria

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 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 says something untrue about N data

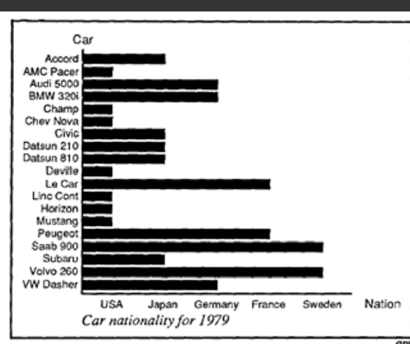


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[Mackinlay, APT, 1986]

Mackinlay's Criteria 2

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.

The subject of the next lecture.

Summary

Formal approach to picture specification

- Declare the picture you want to see
- Compile query, analysis, and rendering commands needed to make the picture
- Automatically generate presentations by searching over the space of designs

Bertin's vision still not complete

- Formalize data model
- Formalize the specifications
- Experimentally test perceptual assumptions

Much more research to be done in this area …