Data and Image Models

Maneesh Agrawala

CS 294-10: Visualization
Spring 2010

Last Time: The Purpose of Visualization

Three functions of visualizations

Record information
- Photographs, blueprints, …

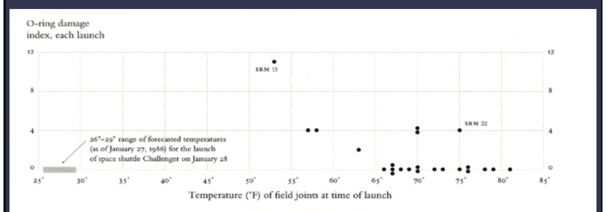
Support reasoning about information (analyze)
- Process and calculate
- Reason about data
- Feedback and interaction

Convey information to others (present)
- Share and persuade
- Collaborate and revise
- Emphasize important aspects of data

Record information

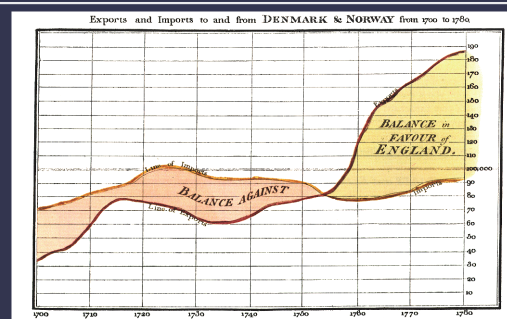
Gallop, Bay Horse "Daisy" [Muybridge 1884-86]

Analysis: Challenger



Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]

Communicate: Exports and Imports



[Playfair 1786]

Announcements

Auditors, *please* enroll in the class (1 unit, P/NP)

- Requirements: Come to class and participate (online as well)
- Requirements: Assignment 1 and 2

Class participation requirements

- Complete readings before class
- In-class discussion
- Post at least 1 discussion substantive comment/question on wiki within a week of each lecture

All, add yourself to participants page on the wiki

Class wiki
http://vis.berkeley.edu/courses/cs294-10-sp10/wiki/

Assignment 1: Good and Bad Vis.

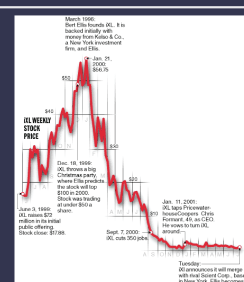
Find two visualizations one *good* and one *bad*

Use original sources

- Journals
- Science magazines
- Newspapers
- Textbooks

Make wiki page

- Clearly mark as good or bad
- Provide short explanation
- Be prepared to succinctly describe in class on Mon Sep. 8



Due before class today

Assignment 2: Visualization Design


News › Datablog

DATA BLOG

Facts are sacred

Previous | Blog home

Weekend update

Haiti earthquake aid pledged by country

Haiti's quake has apparently galvanised the world. Find out how much different countries and organisations have pledged to the aid effort - and how much has actually been handed over

• Get the data

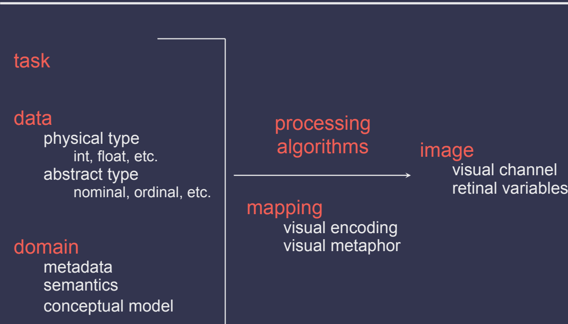

Haiti pledges of aid by country and organisation
Click headings to sort

Country/organisation	Funding, committed and uncommitted, $	$ per person	% of total
Others	636958619		36.32
Private (individuals & organisations)	363486895		20.73
United States	167769881	**0.53**	9.57
Canada	130733775	**3.89**	7.45
World Bank (emergency grant)	100000000		5.7
Spain	45880231	**1.02**	2.62
European Commission	44877340		2.56
United Kingdom	32590138	**0.53**	1.86
France	31313132	**0.50**	1.79

Due before class on Feb 8, 2010

Data and Image Models

The big picture



[based on slide from Munzner]

Topics

Properties of data or information

Properties of the image

Mapping data to images

Data

Data models vs. Conceptual models

Data models are low level descriptions of the data

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

Conceptual models are mental constructions

- Include semantics and support reasoning

Examples (data vs. conceptual)

- (1D floats) vs. Temperature
- (3D vector of floats) vs. Space

Taxonomy

- **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]

Types of variables

Physical types

- Characterized by storage format
- Characterized by machine operations

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

Abstract types

- Provide descriptions of the data
- May be characterized by methods/attributes
- May be organized into a hierarchy

Example:
plants, animals, metazoans, …

Nominal, ordinal and quantitative

N - Nominal (labels)
- Fruits: Apples, oranges, …

O - Ordered
- Quality of meat: Grade A, AA, AAA

Q - Interval (Location of zero arbitrary)
- Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)
- Like a geometric point. Cannot compare directly
- Only differences (i.e. intervals) may be compared

Q - Ratio (zero fixed)
- Physical measurement: Length, Mass, Temp, …
- Counts and amounts
- Like a geometric vector, origin is meaningful

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

Nominal, ordinal and quantitative

N - Nominal (labels)
- Operations: =, ≠

O - Ordered
- Operations: =, ≠, <, >, ≤, ≥

Q - Interval (Location of zero arbitrary)
- Operations: =, ≠, <, >, ≤, ≥, -
- Can measure distances or spans

Q - Ratio (zero fixed)
- Operations: =, ≠, <, >, ≤, ≥, -, ÷
- Can measure ratios or proportions

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

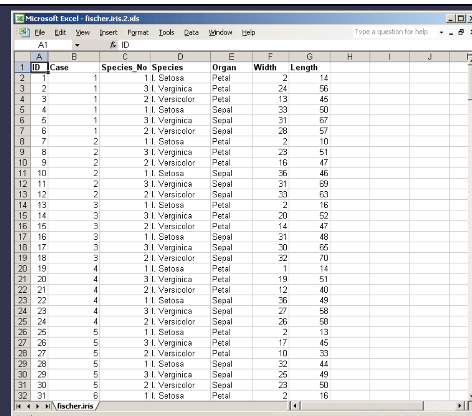
From data model to N,O,Q data type

Data model
- 32.5, 54.0, -17.3, …
- floats

Conceptual model
- Temperature

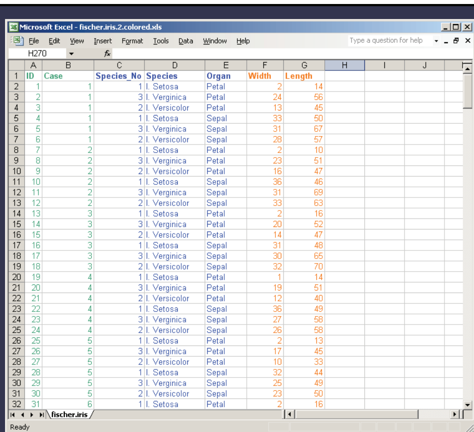
Data type
- Burned vs. Not burned (N)
- Hot, warm, cold (O)
- Continuous range of values (Q)

[based on slide from Munzner]



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

Sepal and petal lengths and widths for three species of iris [Fisher 1936].



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

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

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
Categories or factors or dimensions
Observations or cases

Statistical data model

Variables or measurements
Categories or factors or dimensions
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)

Dimensions and measures

Independent vs. dependent variables

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

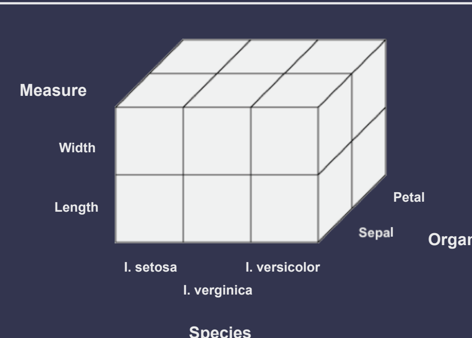
Dimensions and measures

Dimensions: Discrete variables describing data
dates, categories of values (independent vars.)

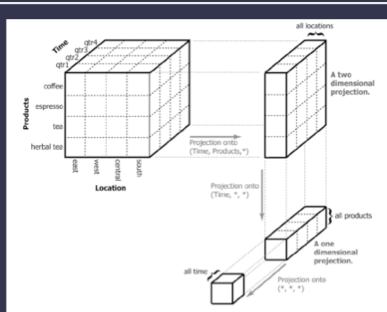
Measures: Data values that can be aggregated
numbers to be analyzed (dependent vars

Aggregations: sum, count, average, std. dev.

Data cube



Projections summarize data



Multiscale visualization using data cubes [Stolte et al. 02]

Image

Visual language is a sign system

Jacques Bertin

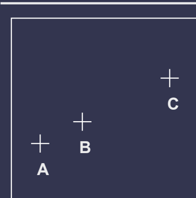
Images perceived as a set of signs

Sender encodes information in signs

Receiver decodes information from signs

Semiology of Graphics, 1983

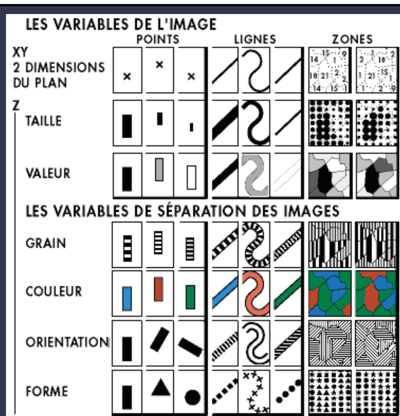
Information in position



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

∴ Encode quantitative variables

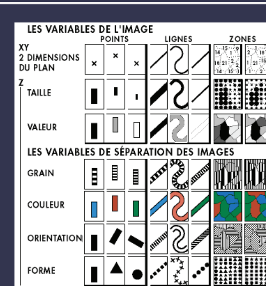
"Resemblance, order and proportional are the three signfields in graphics." - Bertin



[Bertin, Semiology of Graphics, 1983]

Visual variables

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



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

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

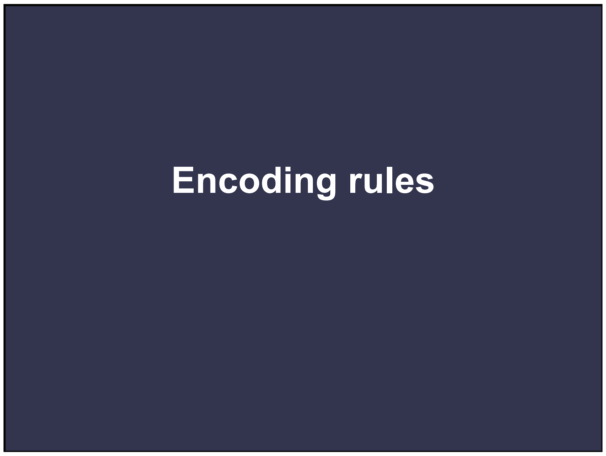
Bertins' "Levels of Organization"

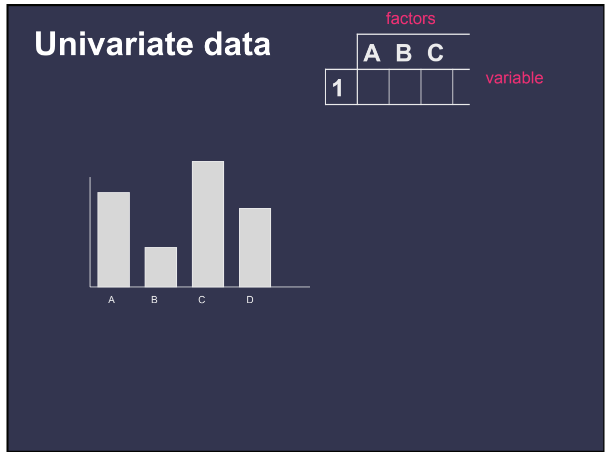
Position	N	O	Q
Size	N	O	Q
Value	N	O	Q
Texture	N	O	
Color	N		
Orientation	N		
Shape	N		

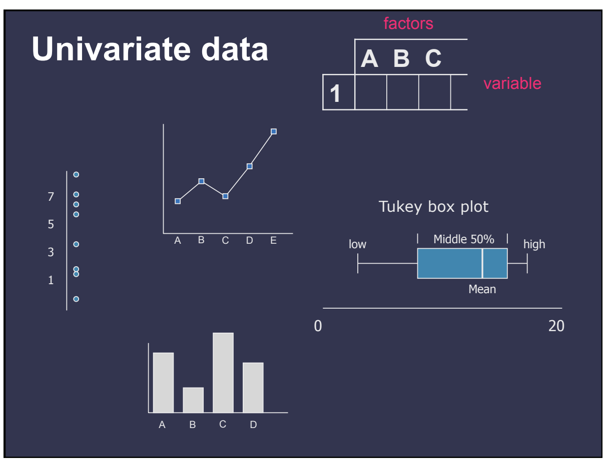
N Nominal
O Ordered
Q Quantitative

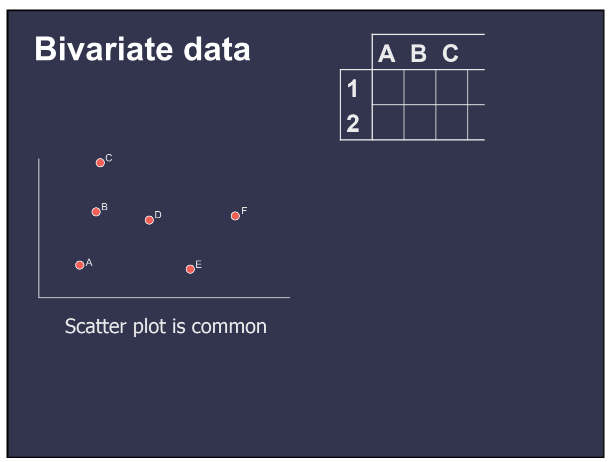
Note: Q < O < N

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


Encoding rules

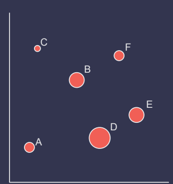

Univariate data
factors
A B C
1
variable
A
B
C
D


Univariate data
factors
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

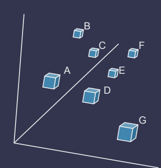

Bivariate data
A B C
1
2
C
B
D
F
A
E
Scatter plot is common

Trivariate data

	A	B	C
1			
2			
3			



3D scatter plot is possible



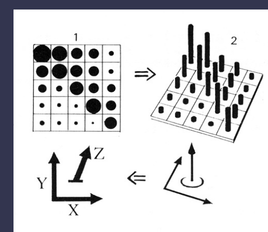
Three variables

Two variables [x,y] can map to points

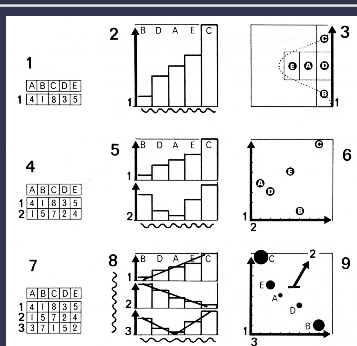
- Scatterplots, maps, …

Third variable [z] must use …

- Color, size, shape, …



Large design space (visual metaphors)



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

Multidimensional data

How many variables can be depicted in an image?

	A	B	C
1			
2			
3			
4			
5			
6			
7			
8			

Multidimensional data

How many variables can be depicted in an image?

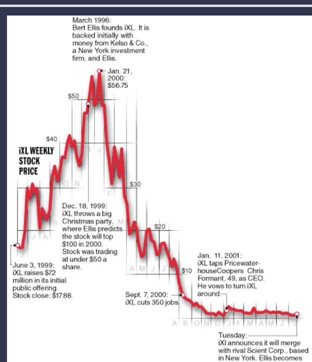
	A	B	C
1			
2			
3			
4			
5			
6			
7			
8			

"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."

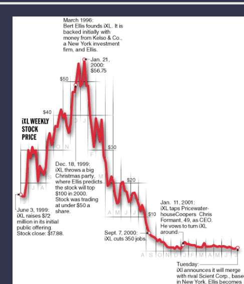
Bertin

Deconstructions

Stock chart from the late 90s

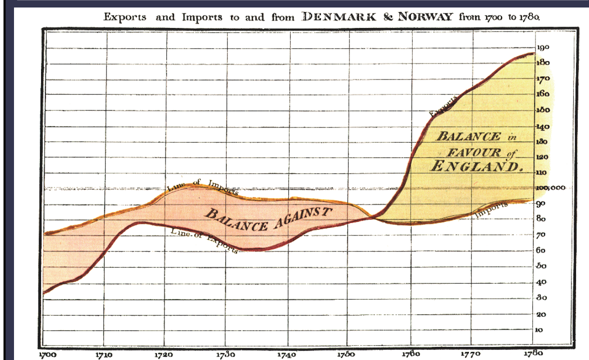


Stock chart from the late 90s

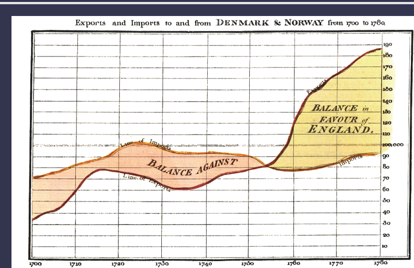


- x-axis: time (Q)
- y-axis: price (Q)

Playfair 1786

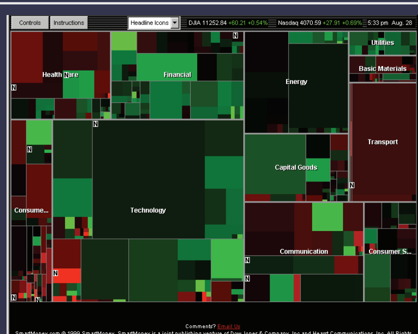


Playfair 1786



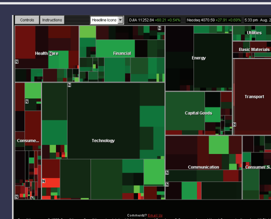
- x-axis: year (Q)
- y-axis: currency (Q)
- color: imports/exports (N, O)

Wattenberg 1998



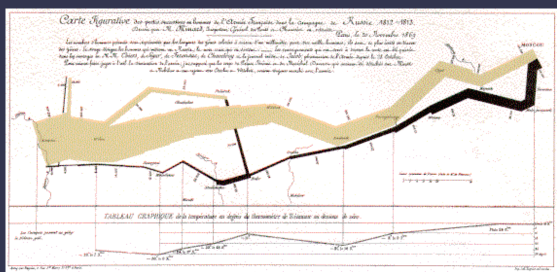
http://www.smartmoney.com/marketmap/

Wattenberg 1998

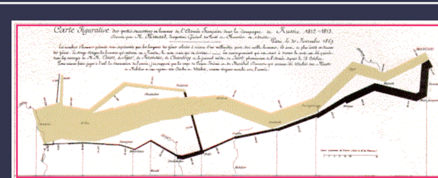


- rectangle size: market cap (Q)
- rectangle position: market sector (N), market cap (Q)
- color hue: loss vs. gain (N, O)
- color value: magnitude of loss or gain (Q)

Minard 1869: Napoleon's march

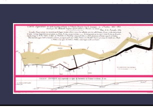


Single axis composition



+

=



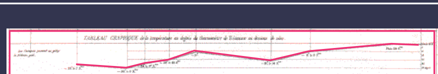
[based on slide from Mackinlay]

Mark composition

y-axis: temperature (Q)

+ x-axis: time (Q)

=



temp over time (Q x Q)

[based on slide from Mackinlay]

Mark composition

y-axis: longitude (Q)

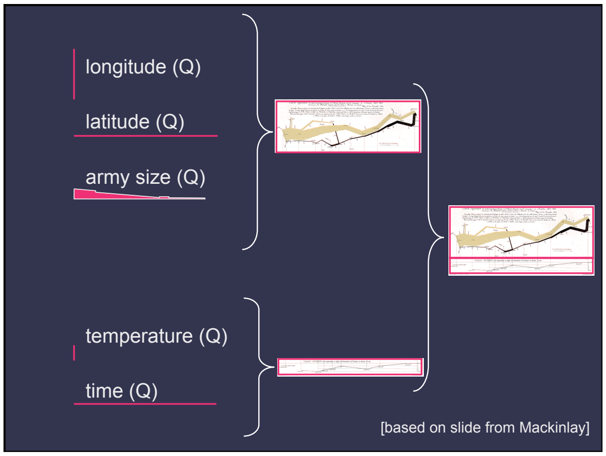
+ x-axis: latitude (Q)

+ width: army size (Q)

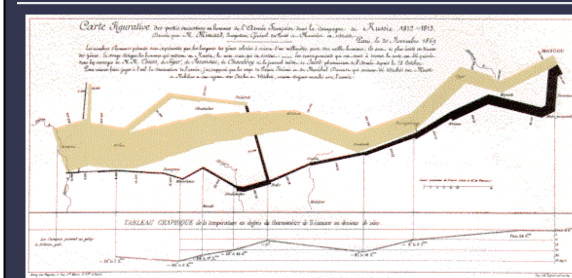
=

army position (Q x Q) and army size (Q)

[based on slide from Mackinlay]



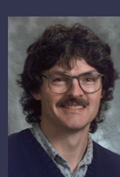
Minard 1869: Napoleon's march



Depicts at least 5 quantitative variables
Any others?

Automated design

Jock 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 (visual variables) should match the properties of the data.

Principle of Importance Ordering:

Encode the most important information in the most effective way.

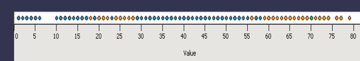
Mackinlay's expressiveness criteria

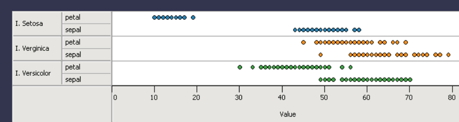
Expressiveness

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

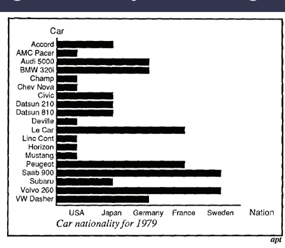


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[Mackinlay, APT, 1986]

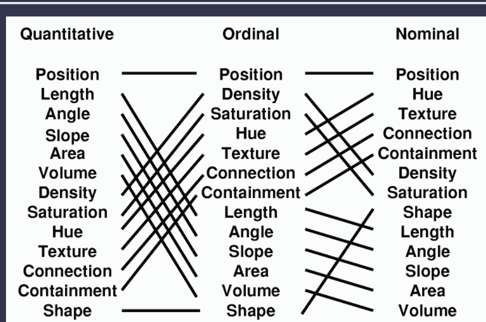
Mackinlay's effectiveness criteria

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 perception lecture

Mackinlay's ranking



Conjectured *effectiveness* of the encoding

Mackinlay's design algorithm

- User formally specifies data model and type
- 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

Limitations

Does not cover many visualization techniques

- Bertin and others discuss networks, maps, diagrams
- They do not consider 3D, animation, illustration, photography, …

Does not model interaction

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