Data and Image Models

Maneesh Agrawala

CS 294-10: Visualization
Fall 2007

Announcements

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

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

Enrollees, 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-fa07/wiki/

Assignment 1a: Good and Bad Vis.

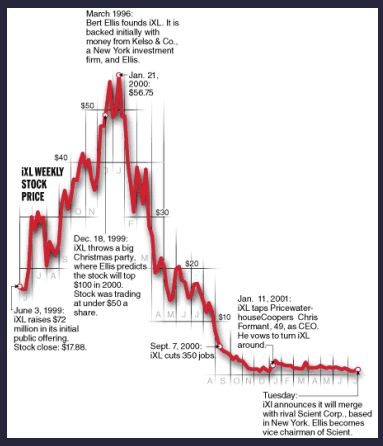
Find two visualizations one *good* and one *bad*

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 Wed Sep. 5



Due before class on Sep 5, 2007

Data and Image Models

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

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

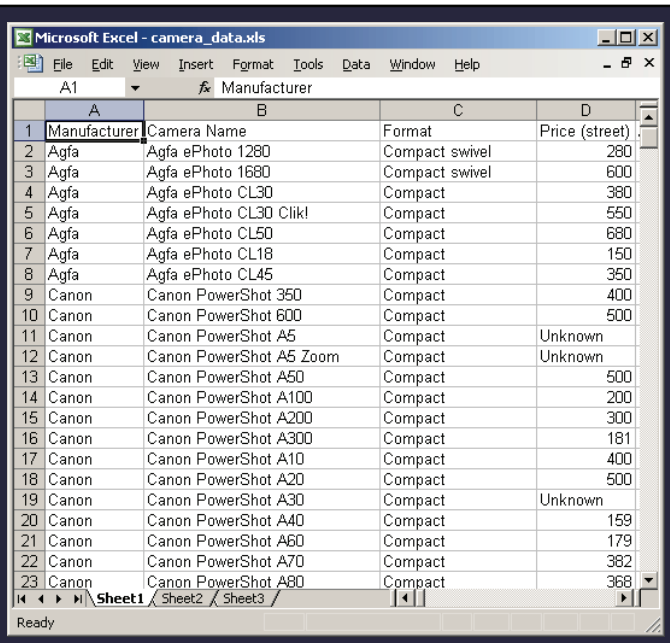
- Include semantics and support reasoning

Examples (data vs. conceptual)

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

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



	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

Example: Digital cameras

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) $\times$ Domain(a)
- Measures: Range(y)

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

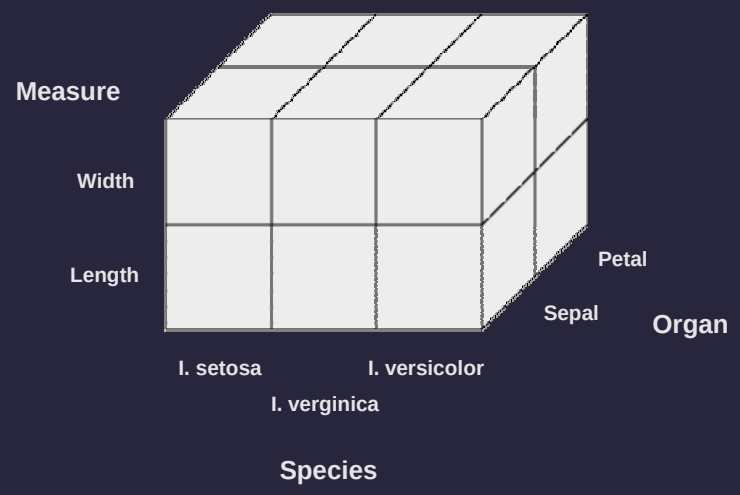
Sepal and petal lengths and widths for three species of iris [Fisher 1936].

	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
21	11	1	43	30	54	23	62	34	45	15	64	32

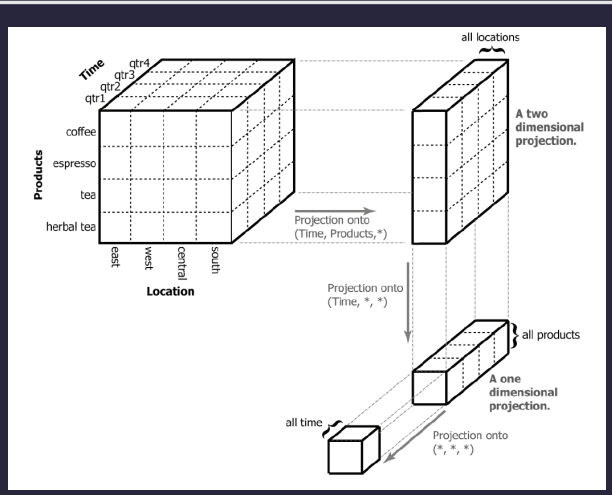
Format of the data in Appendix 14, pp. 365-366

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

Data cube



Projections summarize data



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

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

From data model to 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]

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

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

Microsoft Excel - fischer.iris.2.colored.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

Q

O

N

Image

Visual language is a sign system

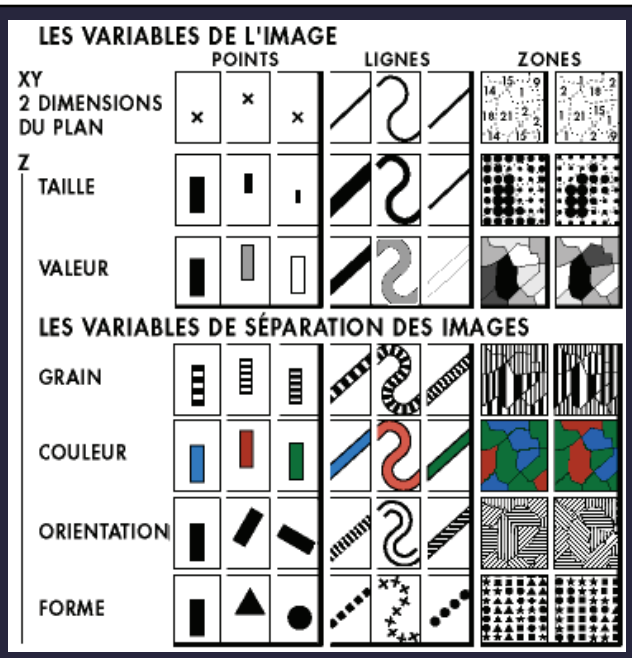
Jacques Bertin

Images perceived as a set of signs

Sender encodes information in signs

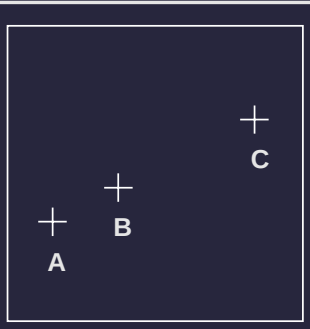
Receiver decodes information from signs

Semiology of Graphics, 1983



[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.

∴ Encode quantitative variables (Q)

"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)

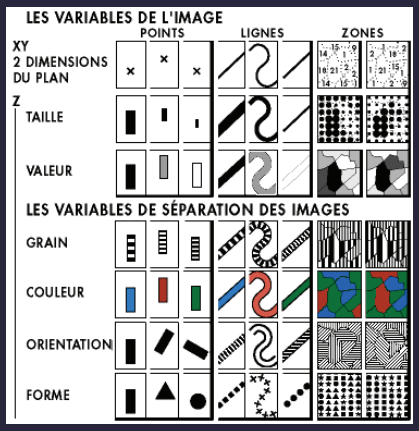
∴ Encode continuous variables (Q) [not as well]

Hue is normally perceived as unordered

∴ Encode nominal variables (N) using color

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.

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 (≡)

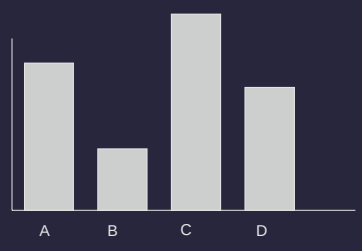
Encoding rules

Univariate data

factors

	A	B	C	
1				

variable

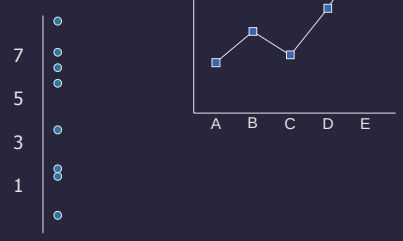


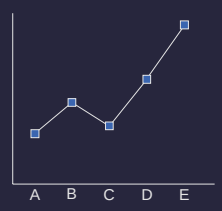
Univariate data

factors

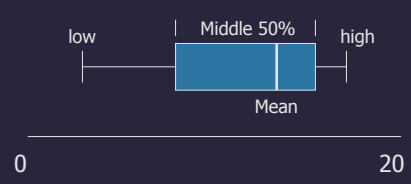
	A	B	C	
1				

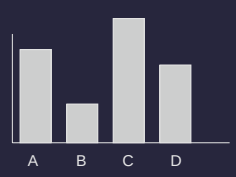
variable





Tukey box plot

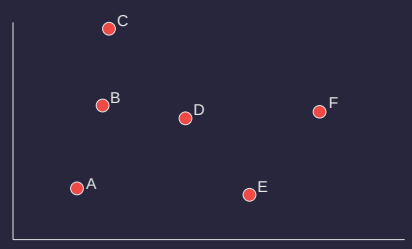




[based on slide from Stasko]

Bivariate data

	A	B	C
1			
2			

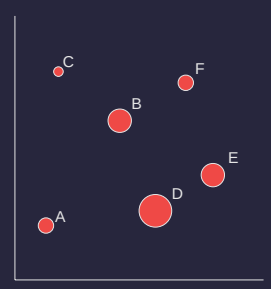


Scatter plot is common

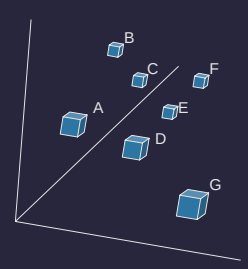
[based on slide from Stasko]

Trivariate data

	A	B	C
1			
2			
3			



3D scatter plot is possible



[based on slide from Stasko]

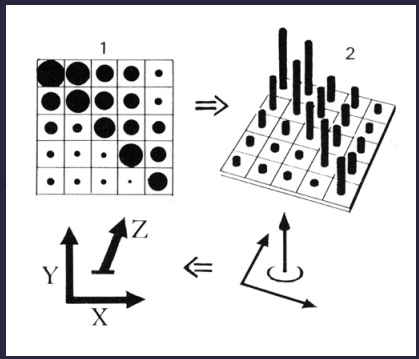
Three variables

Two variables [x,y] can map to points

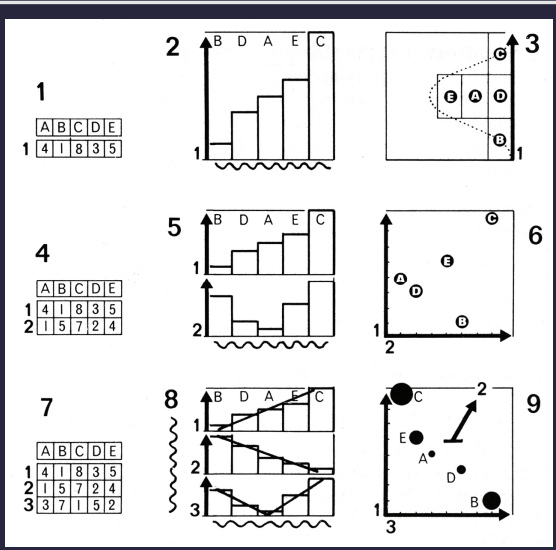
- Scatterplots, maps, …

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?

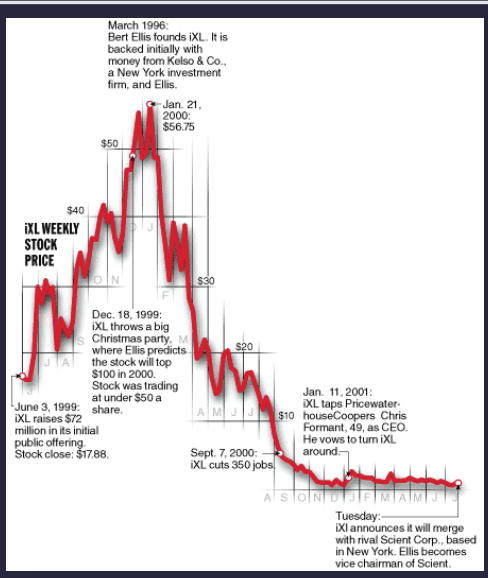
	A	B	C	
1				
2				
3				
4				
5				
6				
7				
8				

"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."

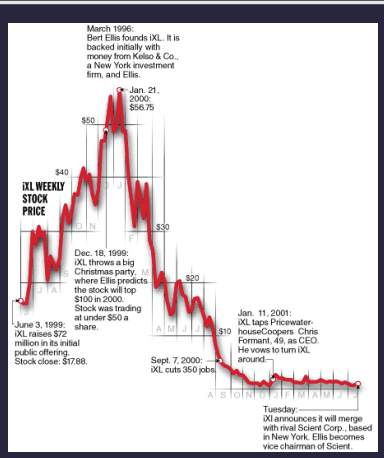
Bertin

Deconstructions

Stock chart from the late 90s

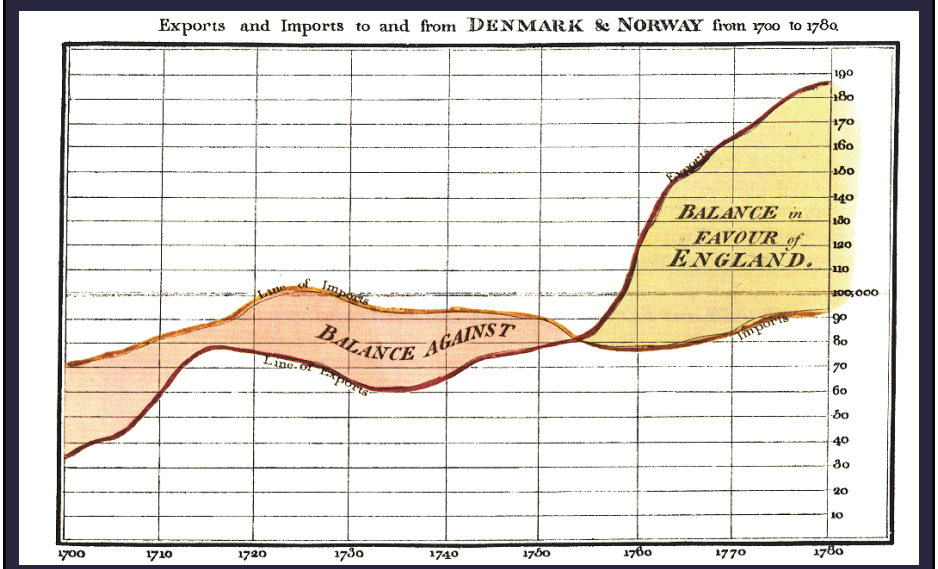


Stock chart from the late 90s

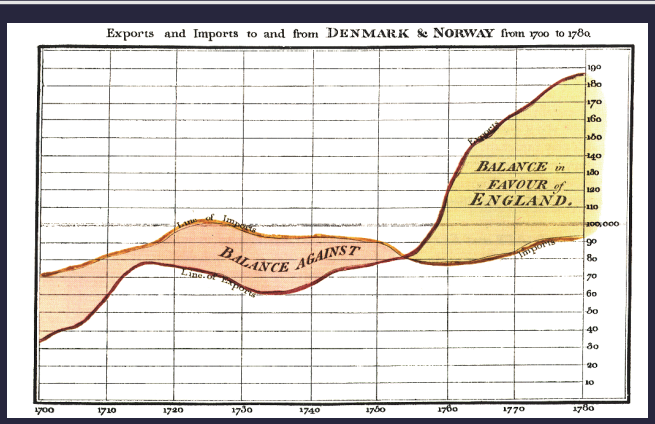


- x-axis: time (Q)
- y-axis: price (Q)

Playfair 1786

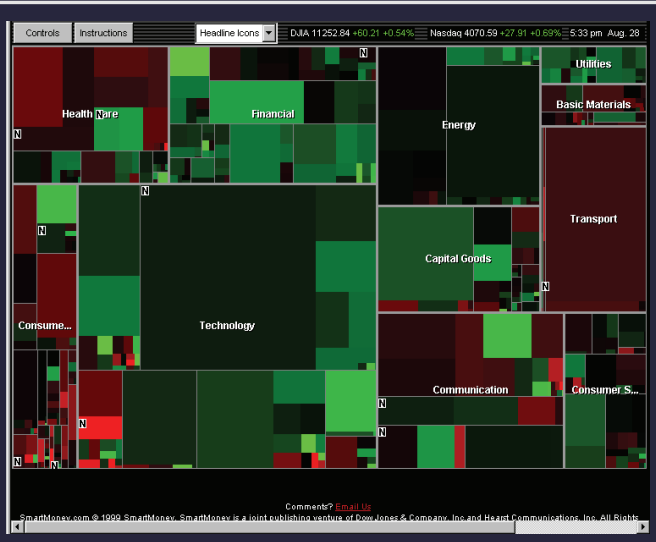


Playfair 1786



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

Wattenberg 1998



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