Information Visualization and Visual Analytics (M1522.000500)

Introduction to Information Visualization

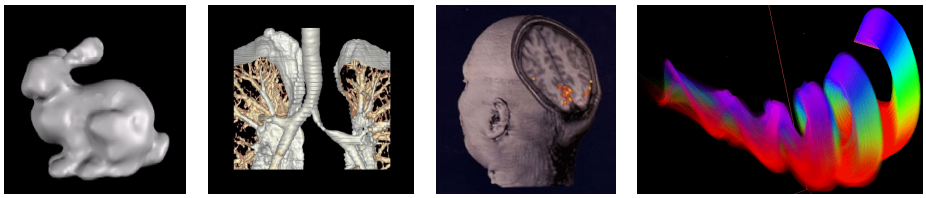
Jinwook Seo, Ph. D.
Professor, Dept. of Computer Science and Engineering
Director, Human-Computer Interaction Lab
Seoul National University

What is it?

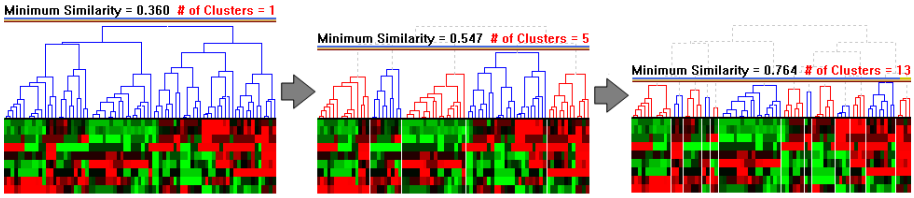
- Information Visualization vs. Scientific Visualization
- Data Visualizations – its power
- Data Visualization – definition

Visualization

- Scientific Visualization



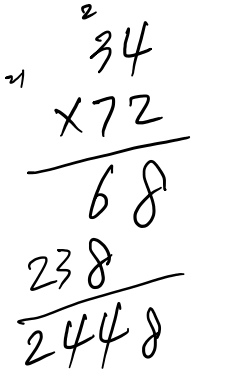
- Information Visualization - **Abstract data**

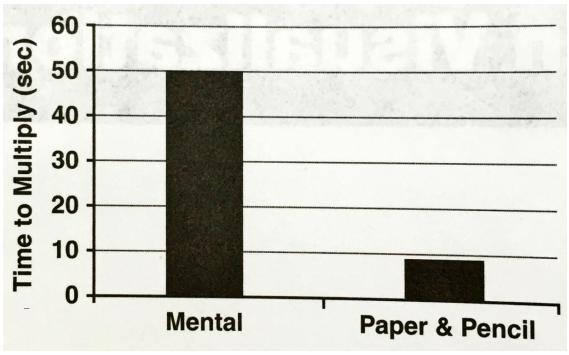


Scientific Visualization: the use of interactive visual representations of scientific data, typically physically based, to amplify cognition. Information Visualization: the use of interactive visual representa- tions of abstract, non-physically based data to amplify cognition.

Why Use an External Representation?

- Finding the *artificial memory* that best supports our natural means of perception
  - Bertin, 1983





Visualizations Reveal Structures

- Statistical characterization of datasets is a very powerful approach
  - losing information through summarization → hide the true structure of the dataset

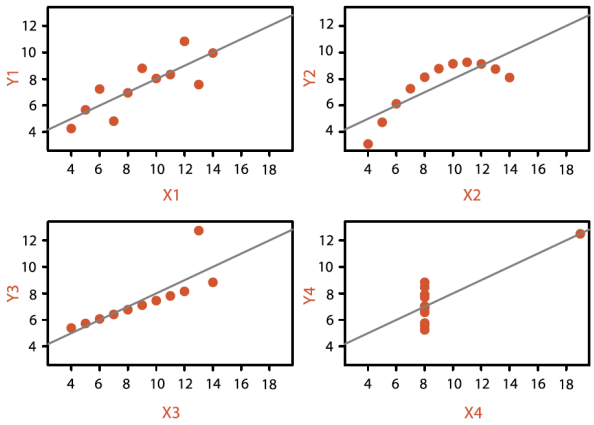
Anscombe's Quartet: Raw Data

	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

Descriptive statistics => Over-simplification

Why Show the Data in Details?

- Identical descriptive statistics → very different structures
- what about features hidden in larger and/or more complex datasets?

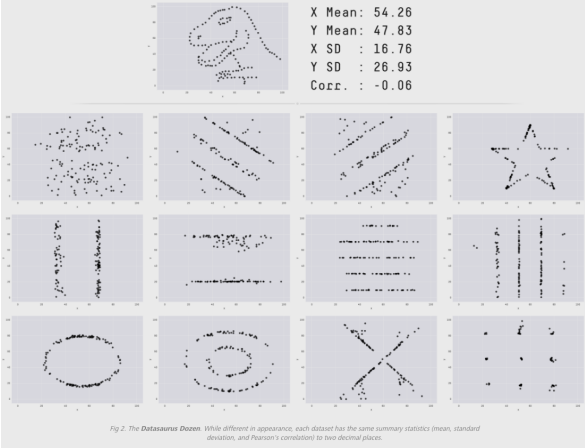


*...make **both** calculations **and** graphs. Both sorts of output should be studied; each will contribute to understanding.*

F. J. Anscombe, 1973

Same Stats, Different Graphs

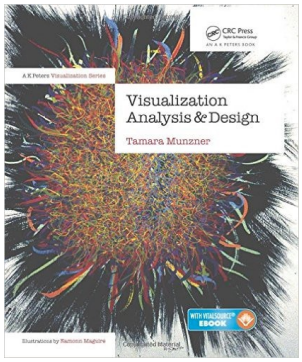
- Generating Datasets with Varied Appearance and Identical Statistics

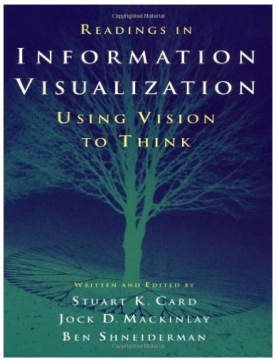


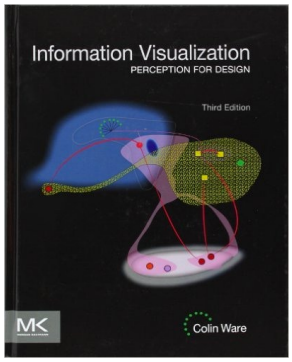
https://www.autodeskresearch.com/publications/samestats

Definition of Visualization

- The use of computer-supported, interactive, visual representations of **abstract** data to amplify **cognition**
  - Stuart Card, Jock Mackinlay, Ben Shneiderman, 1999







Definitions

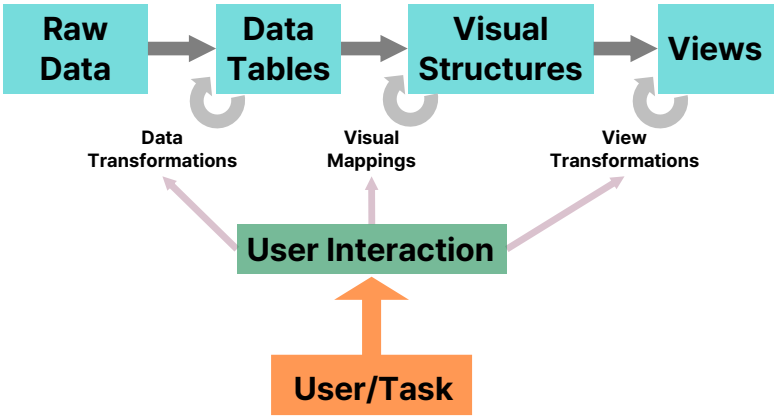
Definition of Visualization

- The use of computer-supported, interactive, visual representations of abstract data to amplify cognition
  - Stuart Card, Jock Mackinlay, Ben Shneiderman, 1999

Definition of external representation

- Finding the *artificial memory* that best supports our natural means of perception
  - Bertin, 1983
- Provide tools that present data in a way to help people understand and gain insight from it

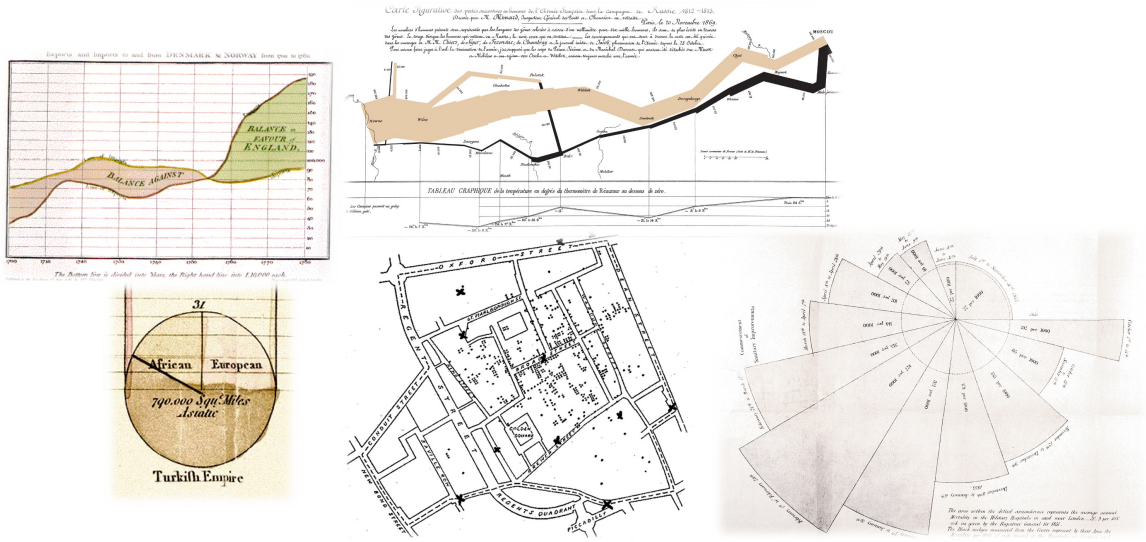
InfoVis Reference Model



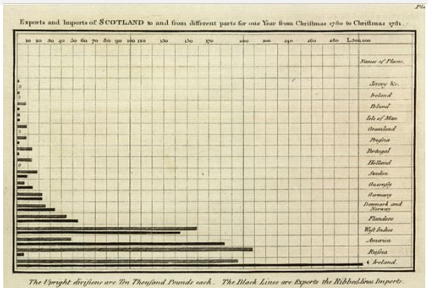
InfoVis is Interdisciplinary

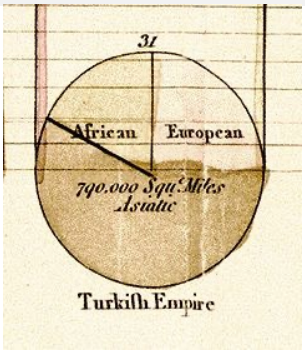
- Graphics: drawing in real time (<100 ms)
- Cognitive psychology: appropriate representation
- HCI: using users and tasks to guide design and evaluation

Data Visualizations – Historical Examples



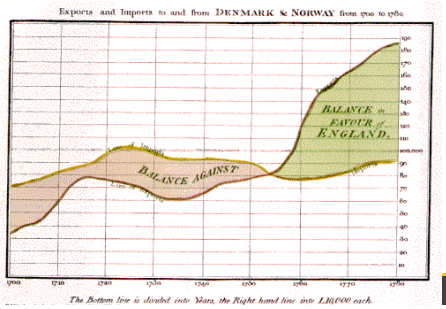
Data Visualizations – Historical Examples





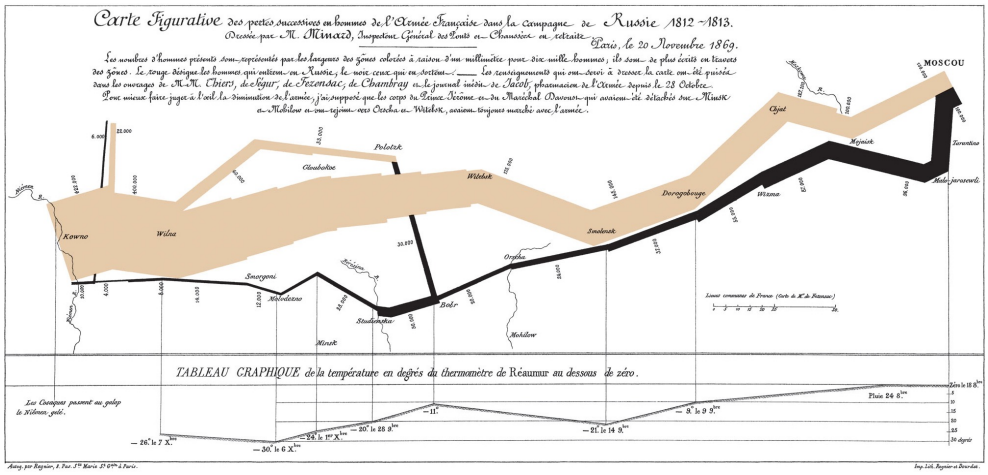
"charts communicated better than tables of data"

William Playfair (1759~1823)
Scottish civil engineer
Playfair has been credited with inventing the **line**, **bar**, **area**, and **pie** charts.



source: https://en.wikipedia.org/wiki/William_Playfair

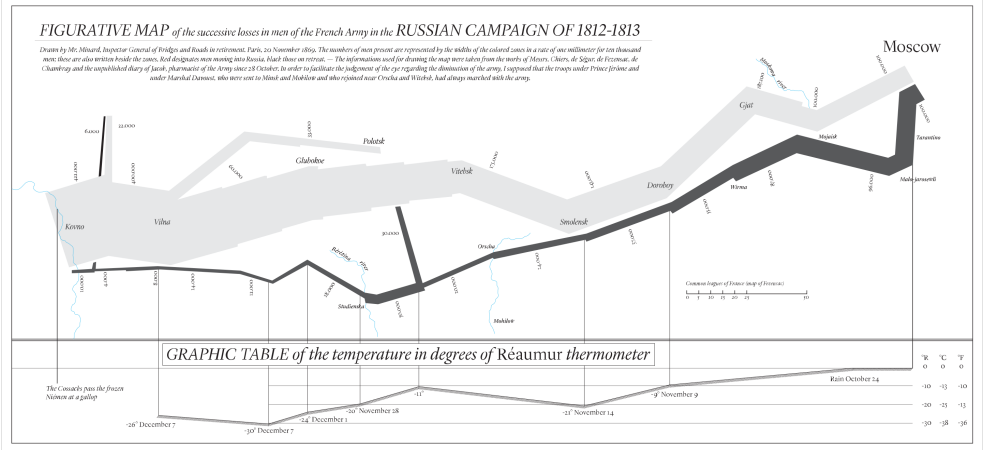
Data Visualizations – Historical Examples



Size of army
Position
Direction of movement
Temperature
Time

Advance of Napoleon's Grande Armée into Russia in 1812
Charles Joseph Minard, 1861

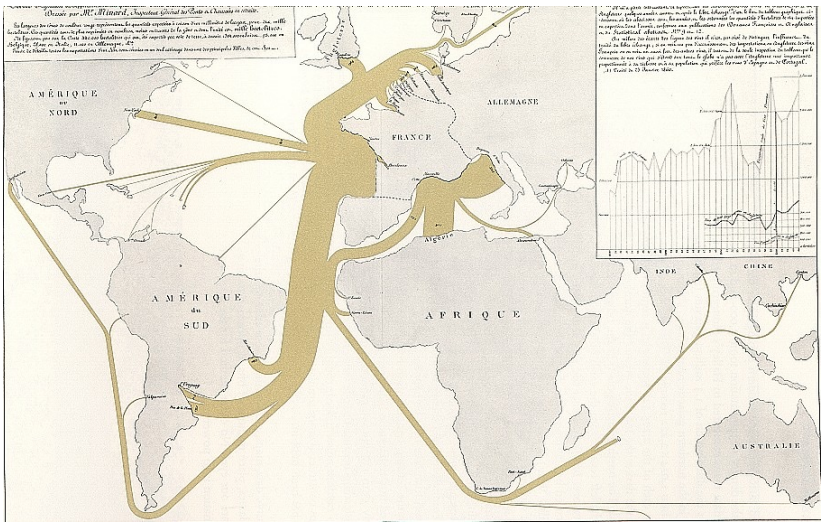
Data Visualizations – Historical Examples



Modern redrawing of Charles Joseph Minard's figurative map of the 1812 French invasion of Russia, including a table of temperatures converting degrees Réaumur to degrees Fahrenheit and Celsius

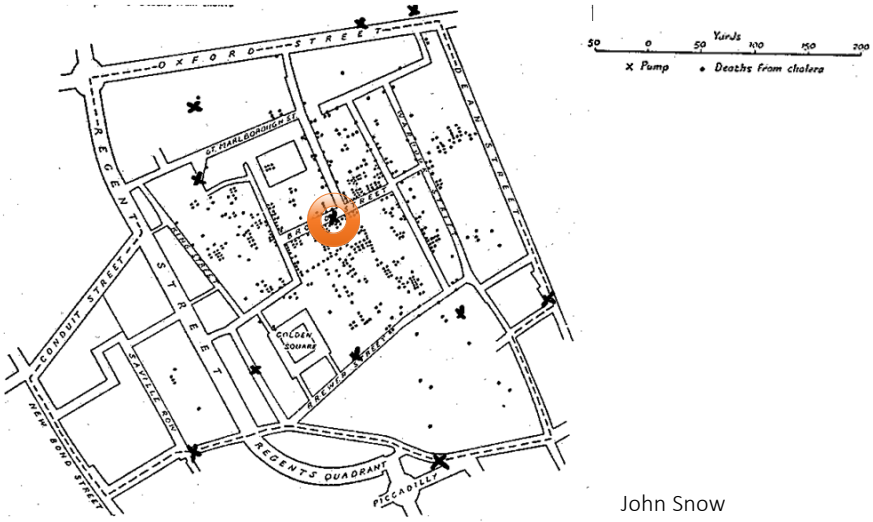
https://en.wikipedia.org/wiki/Charles_Joseph_Minard

1864 Exports of French Wine



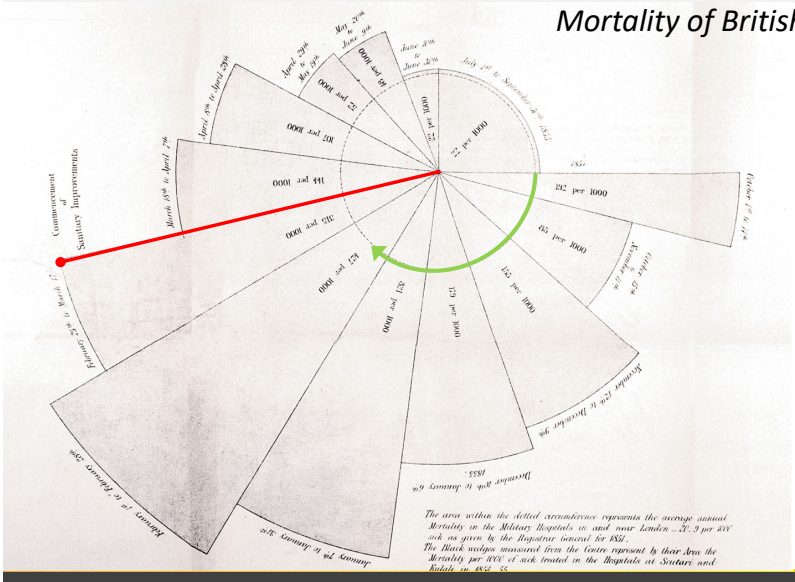
Charles Minard
(economic cartographer)

1854 London Cholera Epidemic



John Snow

Rose-petal diagram



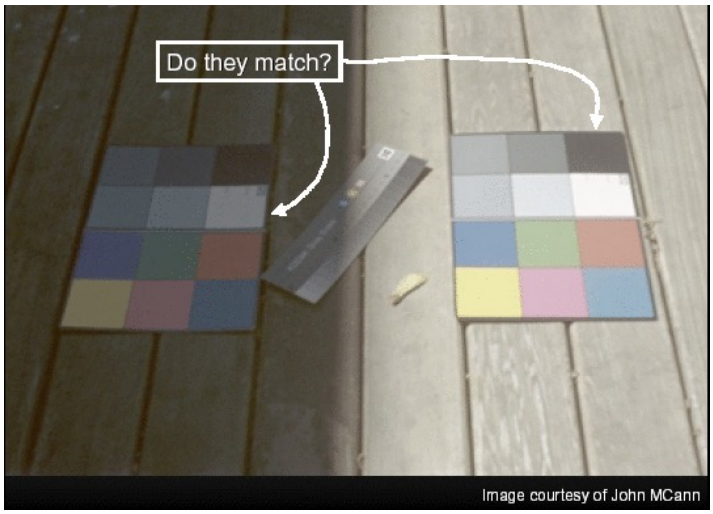
Florence Nightingale's diagram showing the dramatic reduction in death rates in the hospitals of Scutari following the changes she introduced

Source: Nightingale (1858)

Perception for InfoVis

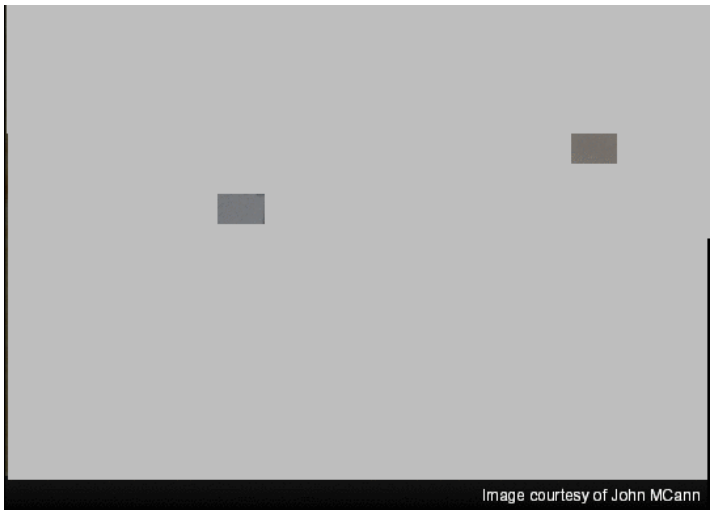
- Visual Perception
- Visual Encodings of Quantitative Data
- Data Types and Tasks for InfoVis

Relative Perception



courtesy of John McCann, from Stone 2001 SIGGRAPH course graphics.stanford.edu/courses/cs448b-02-spring/04cdrom.pdf

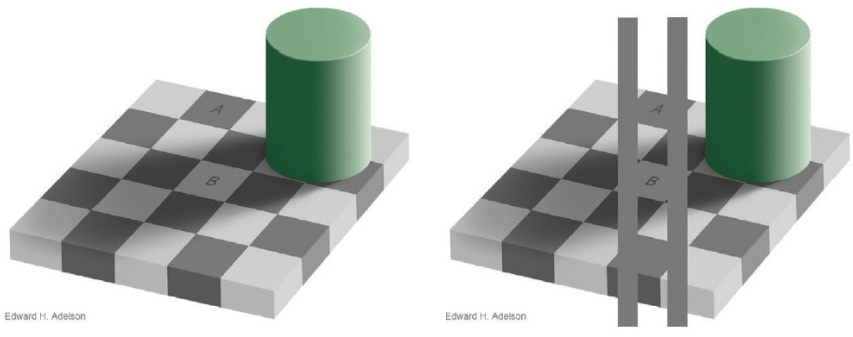
Relative Perception



courtesy of John McCann, from Stone 2001 SIGGRAPH course graphics.stanford.edu/courses/cs448b-02-spring/04cdrom.pdf

Relative versus Absolute Judgements

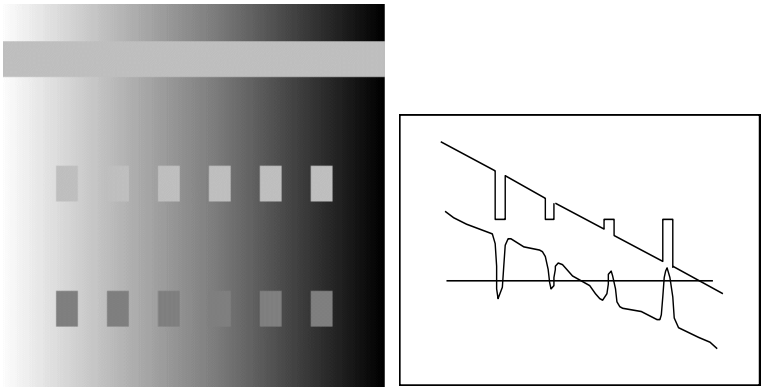
- Luminance contrast – Simultaneous Brightness Contrast
- Luminance perception is based on relative judgements



(a) (b)

Relative versus Absolute Judgements

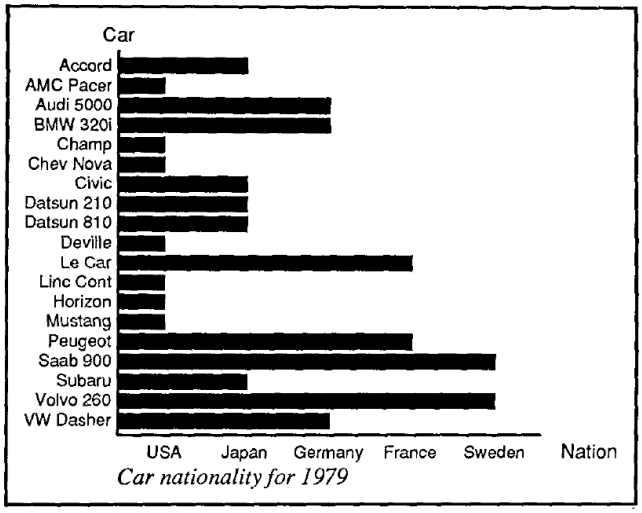
- Luminance contrast – Simultaneous Brightness Contrast
- Luminance perception is based on relative judgements



Two criteria for evaluating graphical designs

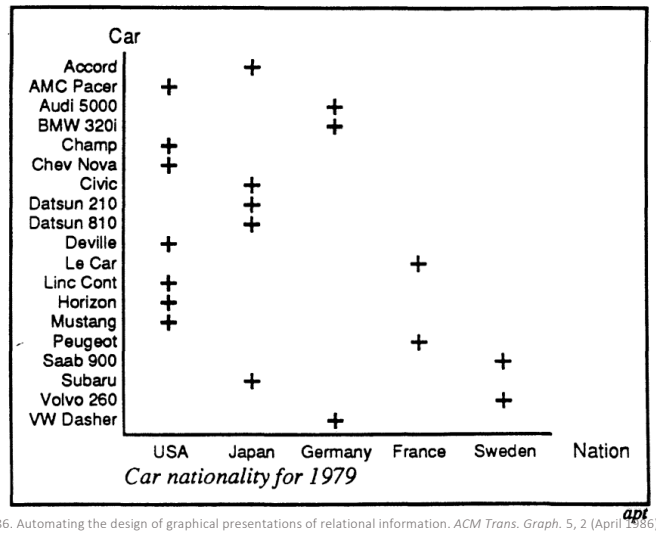
- **Expressiveness**
  - Vis idiom should express **all of**, and **only**, the information in the dataset attributes
- **Effectiveness**
  - Most important attributes should be encoded with the most effective channels
    → ranking of channels

Any better encodings?



Jock Mackinlay. 1986. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.* 5, 2 (April 1986), 110-141DOI=http://dx.doi.org/10.1145/22949.22950

Better Expressiveness! More Effective?



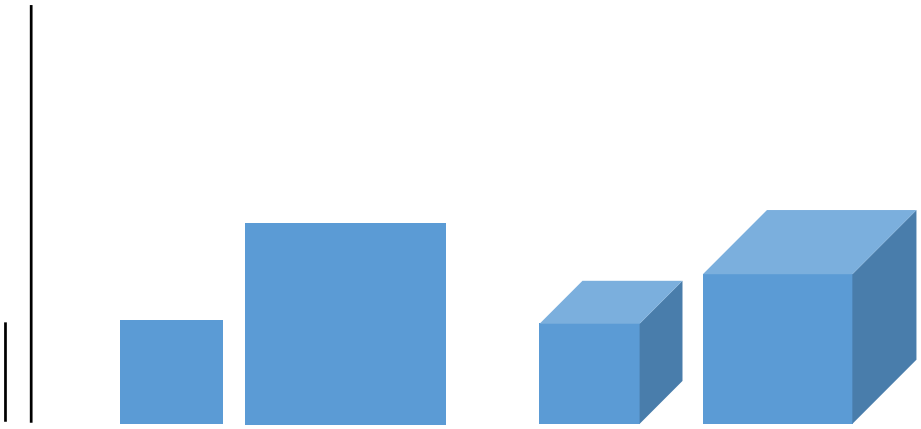
Jock Mackinlay. 1986. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.* 5, 2 (April 1986), 110-141DOI=http://dx.doi.org/10.1145/22949.22950

Two criteria for evaluating graphical designs

- Expressiveness
  - Vis idiom should express **all of**, and **only**, the information in the dataset attributes
- Effectiveness
  - Most important attributes should be encoded with the most effective channels
    → ranking of channels

1D, 2D, 3D

- size ratio for each pair ➔ 1:4



INTERACTIVE DATA VISUALIZATION: FOUNDATIONS, TECHNIQUES, AND APPLICATIONS, Matthew O. Ward; Georges Grinstein; Daniel Keim, A K Peters Ltd (July 1, 2010)

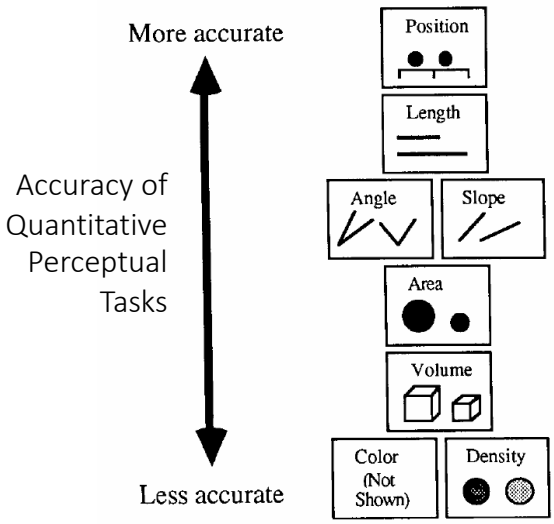
Steven's Power Law

- p: perceived magnitude
- a: actual magnitude
- $p = ka^{\alpha}$
- $p_1/p_2 = (a_1/a_2)^{\alpha}$
- length judgment: $\alpha \approx 1$
- area judgment: $\alpha < 1$
- volume judgment: $\alpha \ll 1$

Continuum	Exponent (a)	Stimulus condition
Loudness	0.67	Sound pressure of 3000 Hz tone
Vibration	0.95	Amplitude of 60 Hz on finger
Vibration	0.6	Amplitude of 250 Hz on finger
Brightness	0.33	5° target in dark
Brightness	0.5	Point source
Brightness	0.5	Brief flash
Brightness	1	Point source briefly flashed
Lightness	1.2	Reflectance of gray papers
Visual length	1	Projected line
Visual area	0.7	Projected square
Redness (saturation)	1.7	Red-gray mixture
Taste	1.3	Sucrose
Taste	1.4	Salt
Taste	0.8	Saccharin
Smell	0.6	Heptane
Cold	1	Metal contact on arm
Warmth	1.6	Metal contact on arm
Warmth	1.3	Irradiation of skin, small area
Warmth	0.7	Irradiation of skin, large area
Discomfort, cold	1.7	Whole body irradiation
Discomfort, warm	0.7	Whole body irradiation
Thermal pain	1	Radiant heat on skin
Tactual roughness	1.5	Rubbing emery cloths
Tactual hardness	0.8	Squeezing rubber
Finger span	1.3	Thickness of blocks
Pressure on palm	1.1	Static force on skin
Muscle force	1.7	Static contractions
Heaviness	1.45	Lifted weights
Viscosity	0.42	Stirring silicone fluids
Electric shock	3.5	Current through fingers
Vocal effort	1.1	Vocal sound pressure
Angular acceleration	1.4	5 s rotation
Duration	1.1	White noise stimuli

http://en.wikipedia.org/wiki/Stevens'_power_law

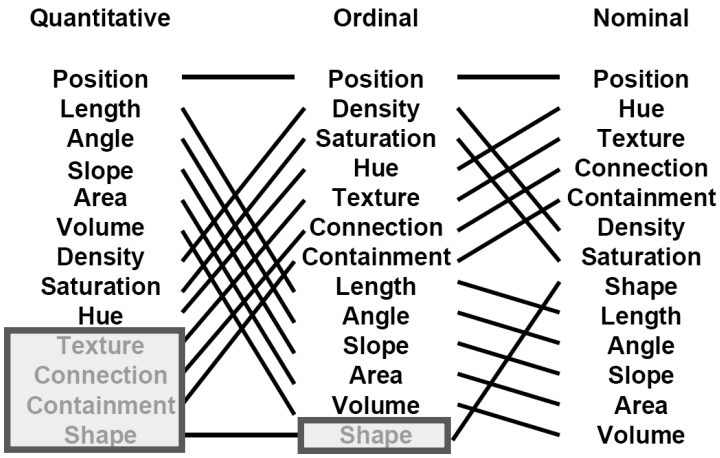
Effectiveness of Visual Encoding



Cleveland & McGill 1984

Visual Encoding (Effectiveness) Principles

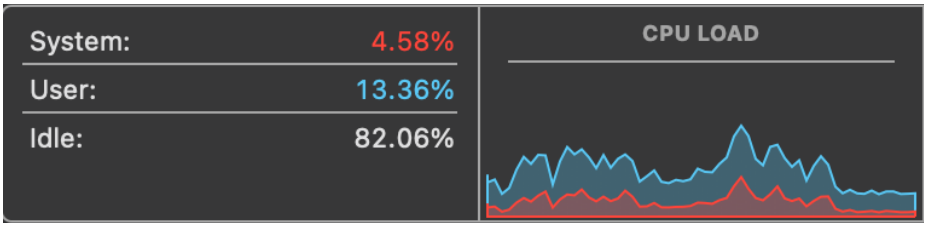
- Channel Ranking Varies by Data Type



Automating the Design of Graphical Presentations of Relational Information ,Jock Mackinlay, ACM Transaction on Graphics, 1986

Which representation is best?

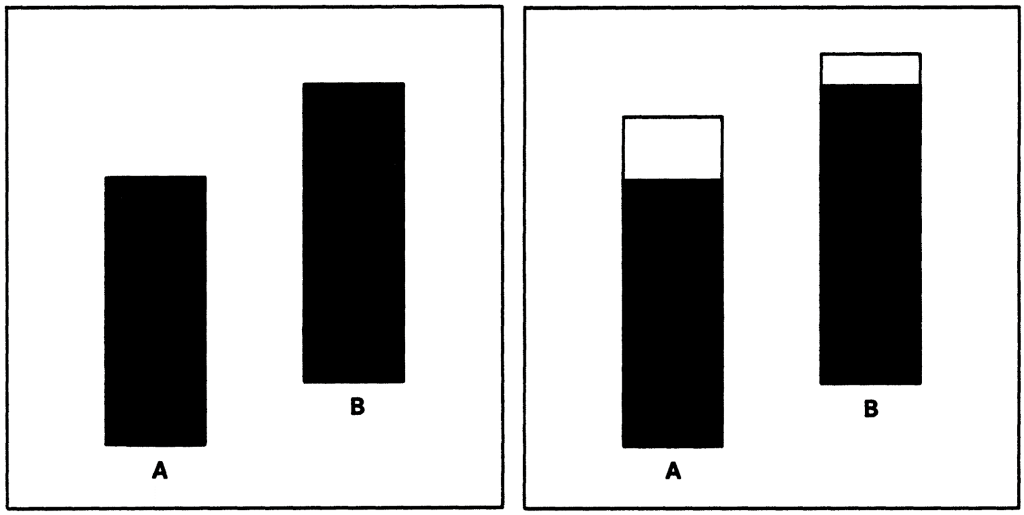
- Depends heavily on TASK



Mac Activity Monitor

Weber's law is the principle that states that the just-noticeable difference between two stimuli is a function of the magnitude of the original stimulus

Weber's Law



Graphical Perception: Theory, Experimentation and the Application to the Development of Graphical Models. William S. Cleveland, Robert McGill, J. Am. Stat. Assoc. 79:387, pp. 531-554, 1984.

Preattentive Processing

- Cognitive operations done preattentively, without the need for focused attention
  - less than 200-250 ms
  - eye movements take 200 ms
    - minimum time to initiate eye movement
    - involves only information available in a single glance
- Popout effects
- Segmentation effects

Preattentive tasks

- visual features that are detected very rapidly by low-level, fast-acting visual processes
- seems to precede focused attention
  - occurring within a single fixation
  - attention plays a critical role in what we see in this early stage
- "pop out" of a display
  - easily detected *regardless of the number of distractors*
  - vs. time-consuming visual search

Christopher G. Healey, James T. Enns: Attention and Visual Memory in Visualization and Computer Graphics. IEEE Trans. Vis. Comput. Graph. 18(7): 1170-1188 (2012)

How many sevens?

239841930921398587450620934895203480950239842102938419092387401293846103298492384265293845013945594858601239480234958728459639405864059823948580239489502934865845610245962348516045698283094586730495614591645908613049529864665895640519680958663045986830495618356018304596803459076283486510465183560241620945613045618304596823045963045986039562034956820438536259682304596304598603956203495682043853625968230459630459860395620349568204385362

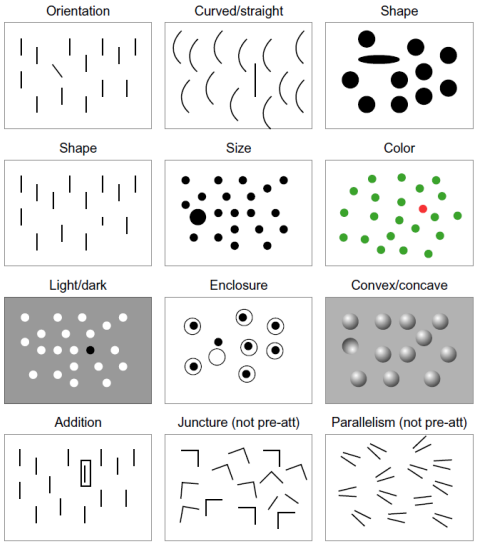
Slide Idea from Colin Ware

Color Makes Them Pop Out

239841930921398587450620934895203480950239842102938419092387401293846103298492384265293845013945594858601239480234958728459639405864059823948580239489502934865845610245962348516045698283094586730495614593045916459086130495298646658956405196809586630459868304956183560183045968034590762834865104651835602416209456130456183045968230459630459860395620349568204385362

Slide Idea from Colin Ware

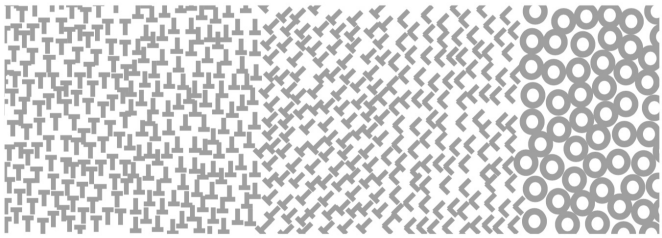
Example



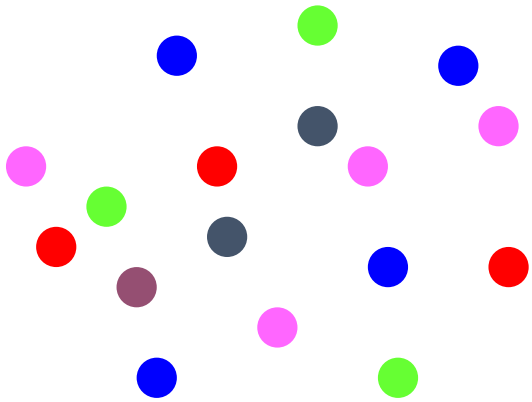
https://www.csc2.ncsu.edu/faculty/healey/PP/

What Kinds of Tasks?

- Target detection
  - Is something there?
- Segmentation (Boundary detection)
  - Can the elements be grouped?
- Region tracking
  - Can a distinctive moving group be traced?
- Counting
  - How many elements of a certain type are present?



Surrounded colors do not pop out



Laws of Preattentive display

- Must stand out on some simple dimension
  - color,
  - simple shape = orientation, size
  - motion,
  - depth
- Lessons for highlighting – one of each

Colin Ware

- Shneiderman's Guidelines
- Tufte's Design Principles
- The Feynman-Tufte Principle

Design Guidelines / Principles

- Visual presentation of query components
- Visual presentation of results
- Rapid, incremental and reversible actions
- Immediate and continuous feedback
- Selection by pointing (not typing)
- Reduces errors
- Encourages exploration
- Visual Information Seeking Mantra
  - Overview first, zoom and filter, details on demand

Ben Shineiderman

Tufte's Design Principles

- Tell the truth
  - Graphical integrity
- Do it effectively with clarity, precision…
  - Design principles/aesthetics
- "simple design, intense content"
  - The Feynman-Tufte Principle, April 2005 *Scientific American*

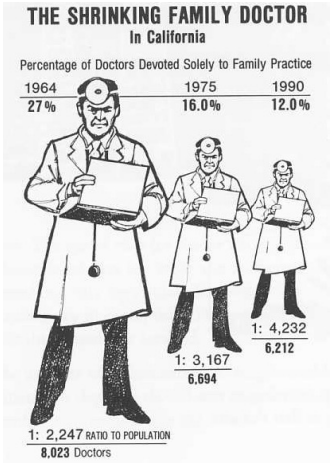
E. Tufte, *The Visual Display of Quantitative Information* (1983)
E. Tufte, *Envisioning Information* (1990)
E. Tufte, *Visual Explanations* (1997)
E. Tufte, *Beautiful Evidence*(2006)

Jinwook Seo / John Stasko

Measuring Misrepresentation

- Visual attribute value should be directly proportional to data attribute value
- Height/width vs. area vs. volume

$$\text{Lie factor} = \frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}$$



"Lie factor" = 2.8

Design Principles

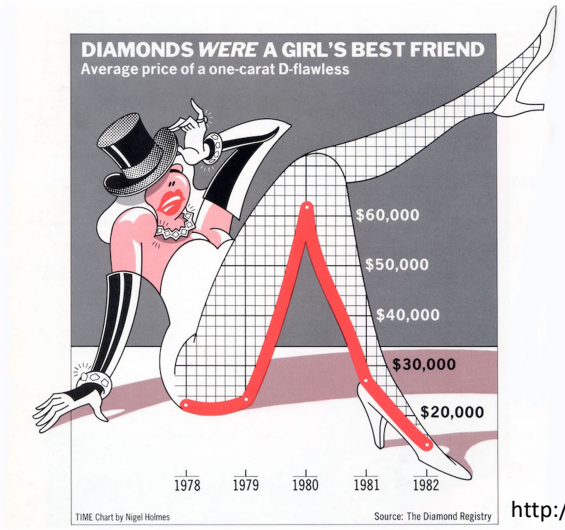
- Maximize data-ink ratio

$$\text{Data-ink ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$

= proportion of graphic's ink devoted to the non-redundant display of data-information

Design Principles

- Avoid chartjunk
  - Extraneous visual elements that detract from information

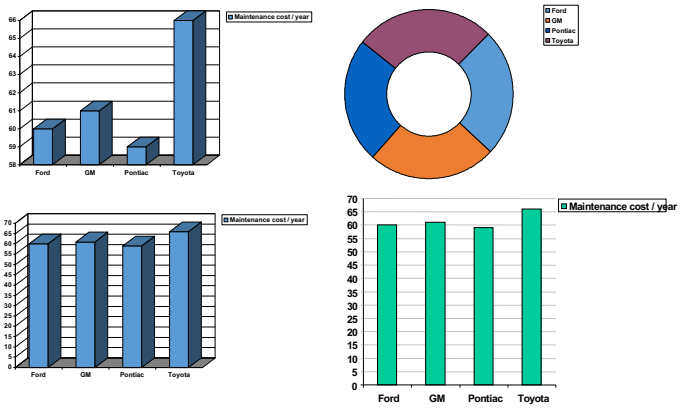


http://nigelholmes.com/

John Stasko

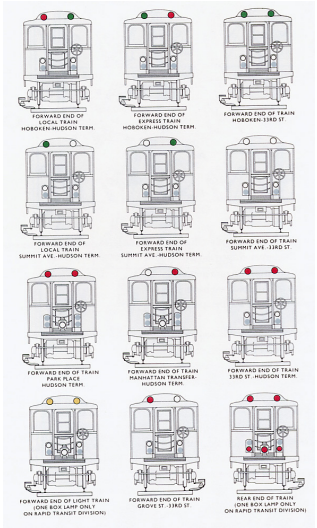
Chartjunk

- All visual elements in charts and graphs that are not necessary to comprehend the information represented on the graph, or that distract the viewer from this information



Design Principles

- Use Small multiples
  - Repeat visually similar graphical elements nearby rather than spreading far apart
  - The same graphical design structure is repeated
  - Learn once and compare
    → invite comparisons
  - Reveal, all at once, a scope of alternatives, a range of options
    → overview



Rules and Regulations for the Government of Employees of the Operating Department of the Hudson & Manhattan Railroad Company, 1923

Design Principles

- Utilize narratives of space and time
  - Tell a story of position and chronology through visual elements