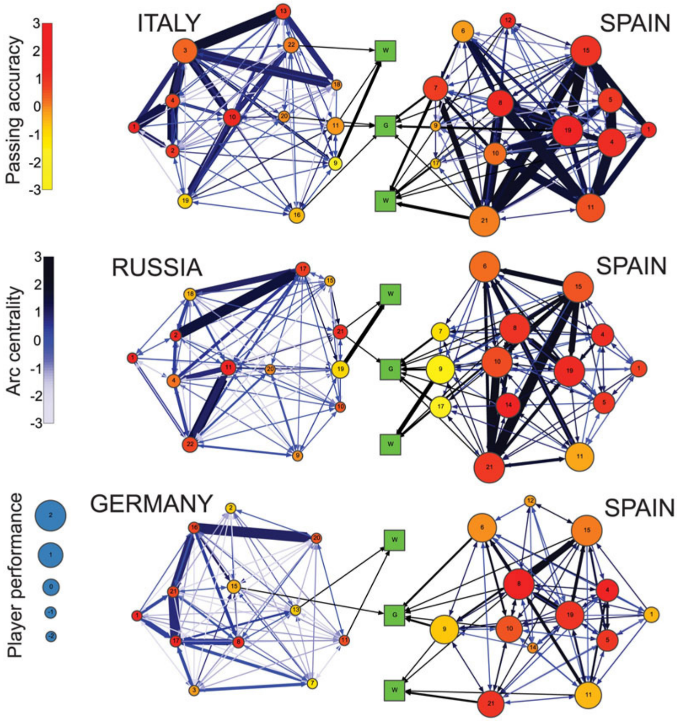


IAT 814

Visual Encoding and Image Models

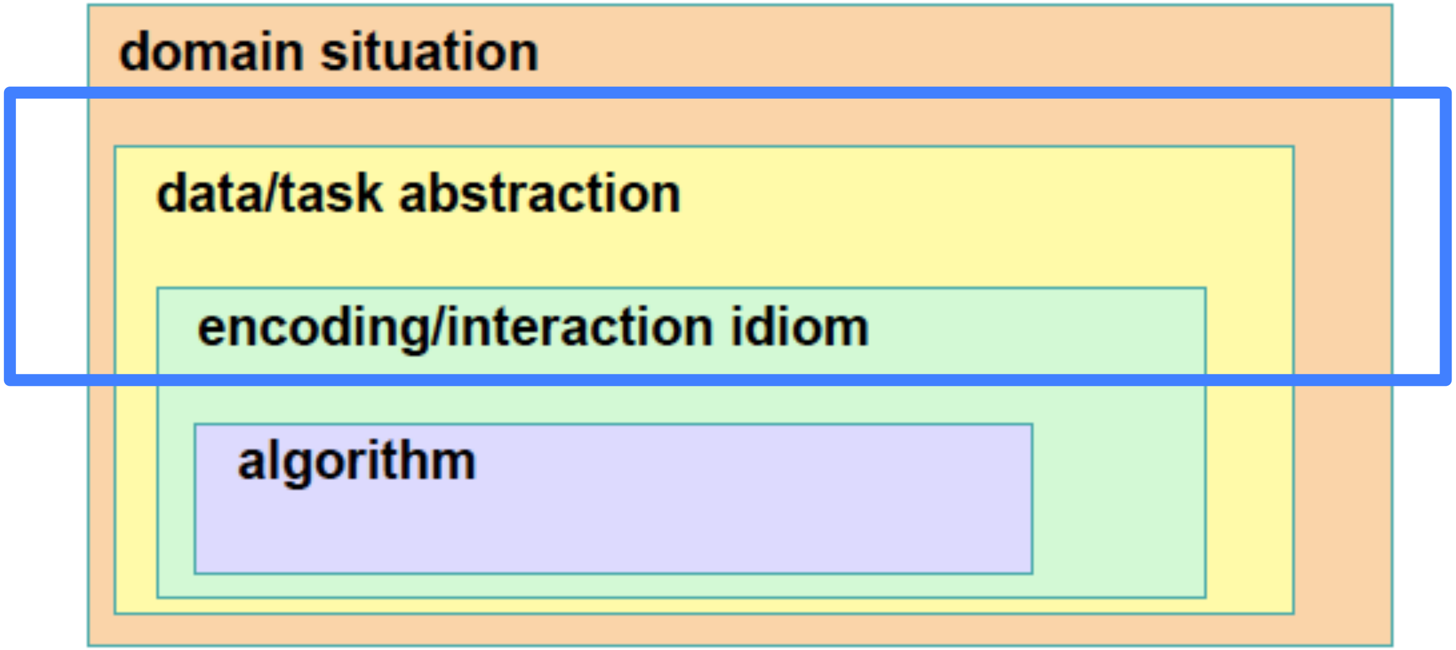
Lyn Bartram

Note: Many of these slides have been borrowed and adapted from T. Munzner and J.Heer

SFU

SCHOOL OF INTERACTIVE ARTS + TECHNOLOGY

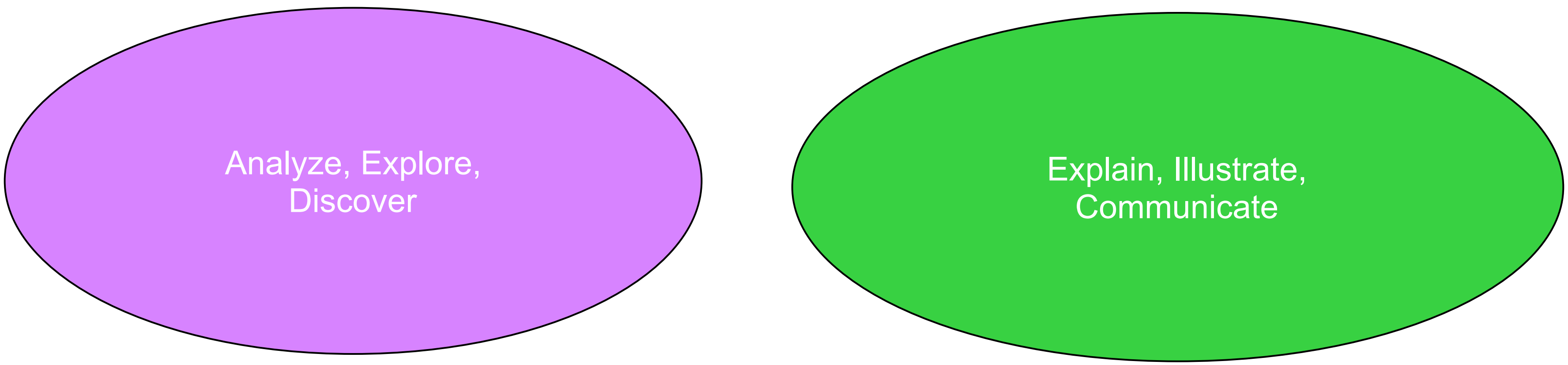
4 stages of visualization design



Data Abstractions

- Tables
  - Data item (row) with attributes (columns) : row=key, cells = values
- Networks
  - Item (node) with attributes (features) and relations (links)
  - Trees (hierarchy)
  - Node = key, node-node, link = key, cell = value
- Text/Logs
  - Grammar
  - Bag of words
  - Derived values
- Image
  - 2d location = key, pixel value expresses single attribute or combo of attributes according to coding (RGB)

Visualization: Why?



Task Abstractions: [Munzner]

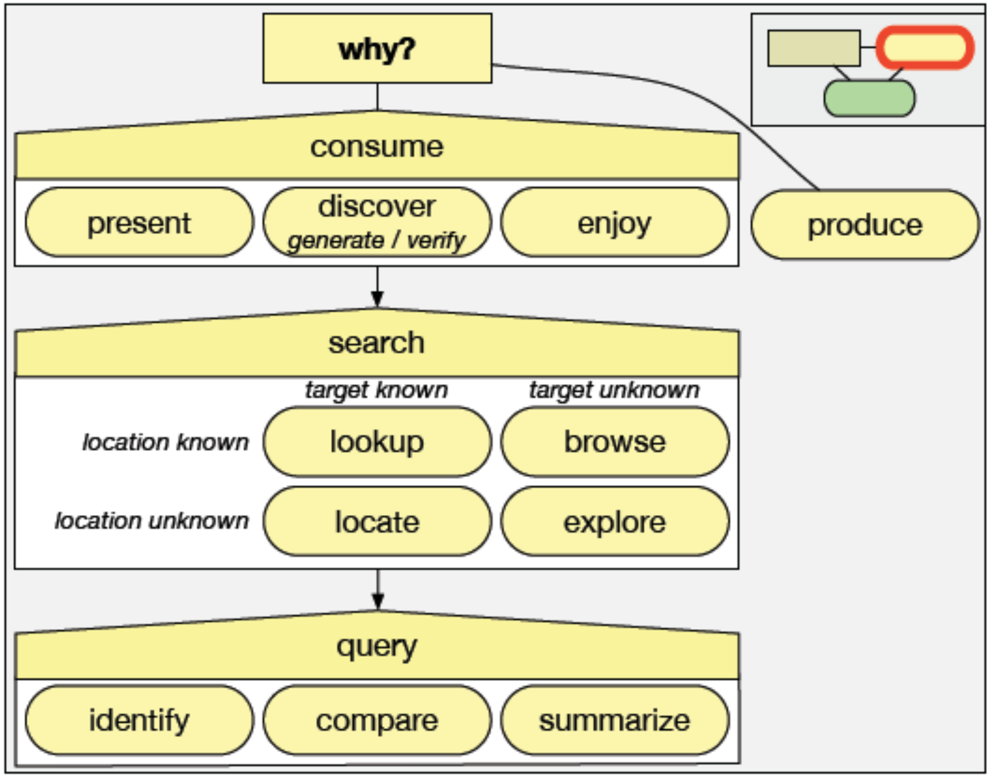


Image models

- Representation
  - The visual, aural or haptic (ie sensory) encoding of the data
  - This is often termed *mapping*
- Presentation
  - Selection, layout and organisation of encoded data
  - May involve *multiple representations*
- Interaction
  - Manipulation to acquire different views of the data

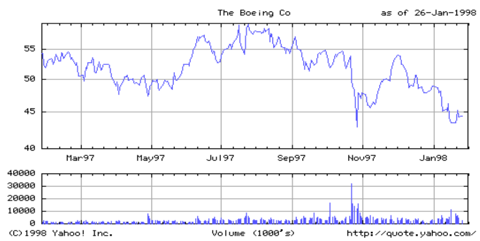
Representation

- What's a common way of visually representing multivariate data sets?
- Graphs! (not the vertex-edge ones)

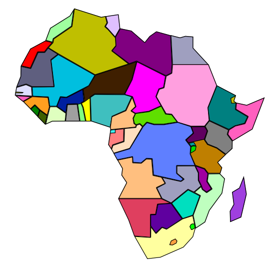
- More accurately, symbolic display

Types of Symbolic Displays (Kosslyn 89)

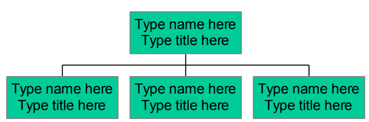
- Graphs



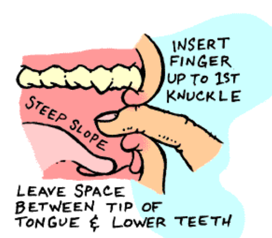
- Maps



- Charts



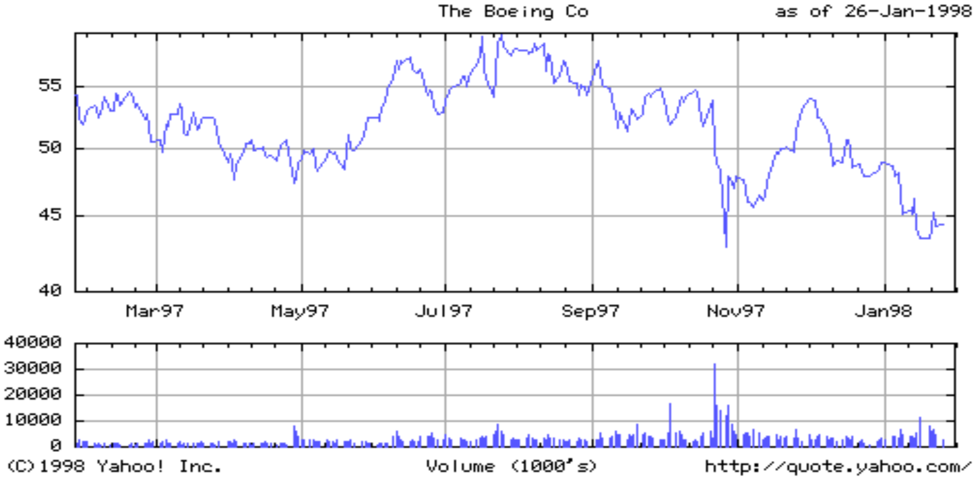
- Diagrams



Slide aapted from Marti Hearst

Types of Symbolic Displays

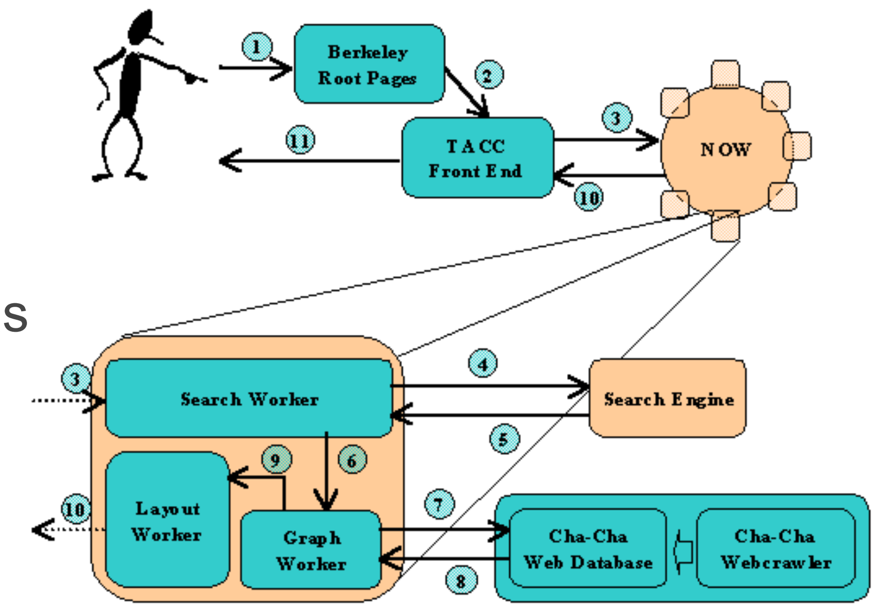
- Graphs
  - at least two scales required
  - values associated by a symmetric “paired with” relation
    - Examples: scatter-plot, bar-chart, layer-graph



Slide adapted from Marti Hearst

Types of Symbolic Displays

- Charts
  - discrete relations among discrete entities
  - structure relates entities to one another
  - lines and relative position serve as links
- Examples:
  - Family tree
  - Flow chart
  - Network diagram

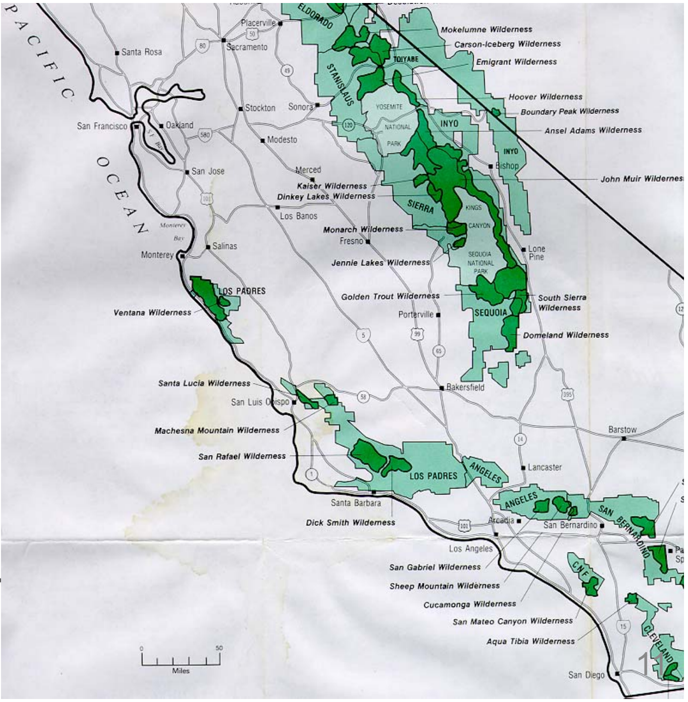


Slide aapted from Marti Hearst

Types of Symbolic Displays

- Maps
  - Internal relations determined (in part) by the spatial relations of what is pictured
  - Labels paired with locations

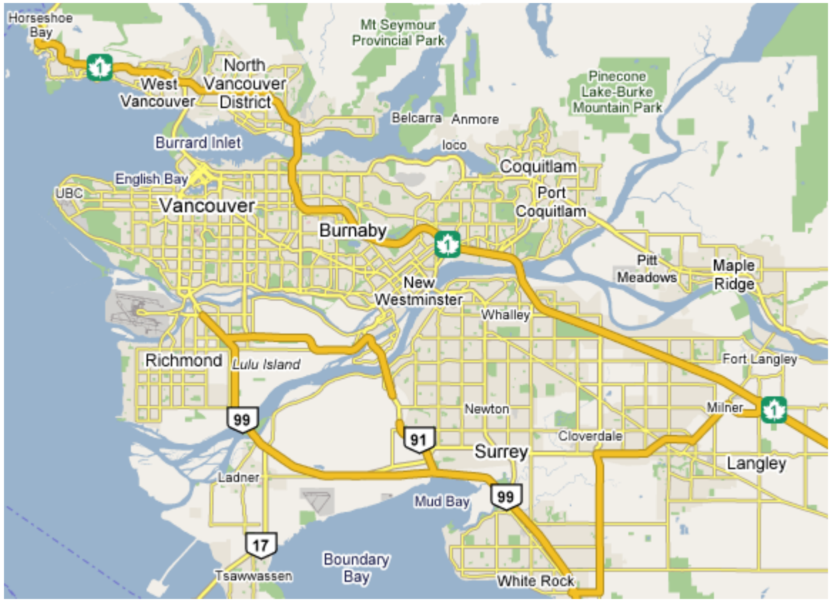
- Examples:
  - Map of census data
  - Topographic maps



Map

- Internal relations determined (in part) by the spatial relations of what is pictured
  - Grid: *geometric* metadata
- Locations identified by labels
- *Nominal* metadata

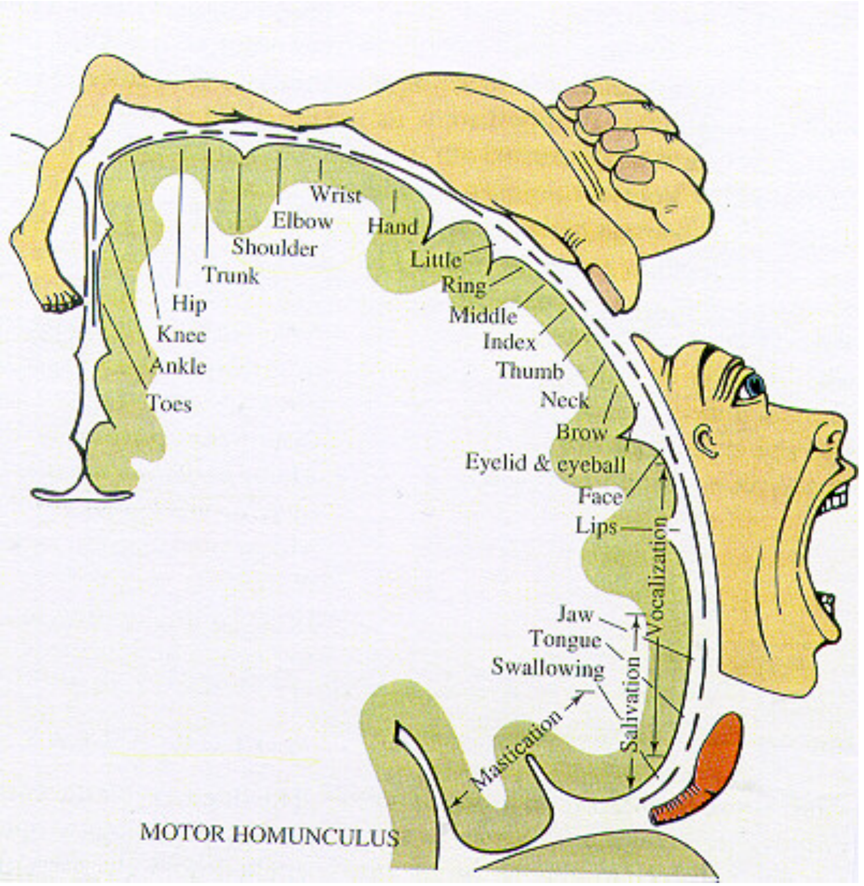
- Examples:
- Map of census data
- Topographic maps



SFU

Types of Symbolic Displays

- Diagrams
  - Schematic pictures of objects or entities
  - Parts are symbolic (unlike photographs)
    - how-to illustrations
    - figures in a manual



From Glietman, Henry. Psychology. W.W. Norton and Company, Inc. New York, 1995

What is the “real” taxonomy for visual representations? [Lohse et al.’ 94]

- empirical investigation (Only used static, 2D graphics)
  - 16 participants
    - Half had a graphic design background
  - First, looked at 60 images and scored them along 10 scales.
    - These were used to compute statistical similarity
  - organized the 60 images into categories according to similarity.
    - Were asked to name the groups
    - Then they grouped these into higher-level groups, repeatedly, until they were in one large group.

Lohse, G L; Biolsi, K; Walker, N and H H Rueter, A Classification of Visual Representations, CACM, Vol. 37, No. 12, pp 36-49, 1994

Scales that Participants Used
(and percentage of variance explained)

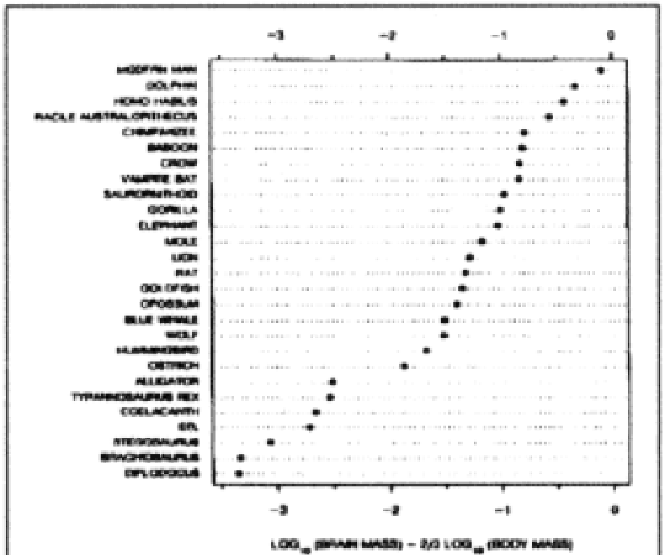
16.0	emphasizes whole – parts
11.3	spatial – nonspatial
10.6	static structure – dynamic structure
10.5	continuous – discrete
10.3	attractive – unattractive
10.1	nontemporal – temporal
9.9	concrete – abstract
9.6	hard to understand – easy
9.5	nonnumeric – numeric
2.2	conveys a lot of info – conveys little

Resulting Categories (Lohse et al. 94)

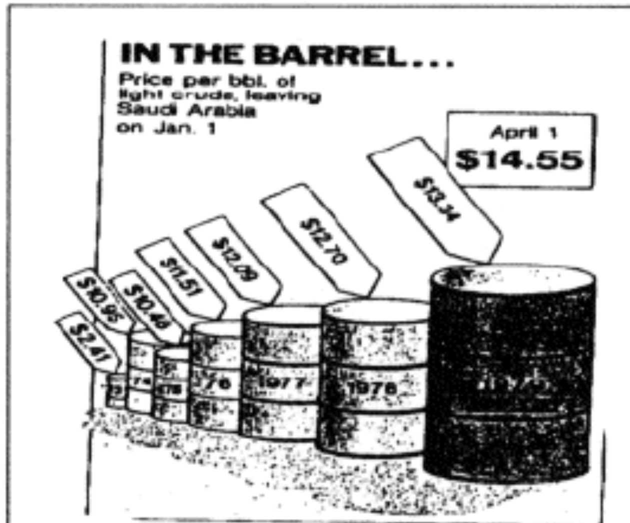
- Graphs
- Tables (numerical)
- Tables (graphical)
- Charts (time)
- Charts (network)
- Diagrams (structure)
- Diagrams (network)
- Maps
- Cartograms
- Icons
- Photo-realistic images

Graphs

- Encode quantitative information using position and magnitude of geometric objects.
- Examples: scatter plots, bar charts.



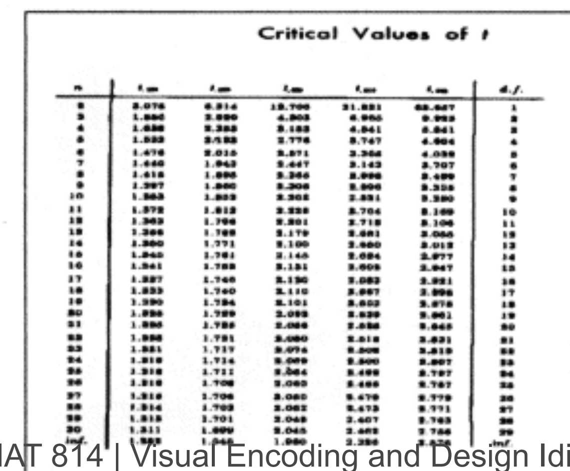
38. dot chart



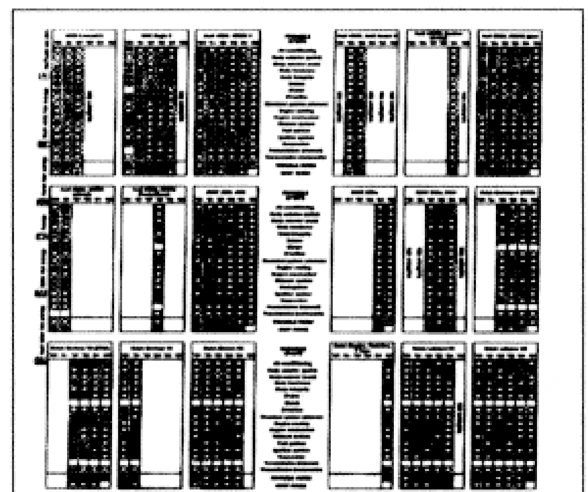
34. oil barrels

Tables

- An arrangement of words, numbers, signs, or combinations of them to exhibit a set of facts or relationships in a compact fashion.
- Less abstract symbolic notation than graphs.
  - Graphical tables and numerical tables



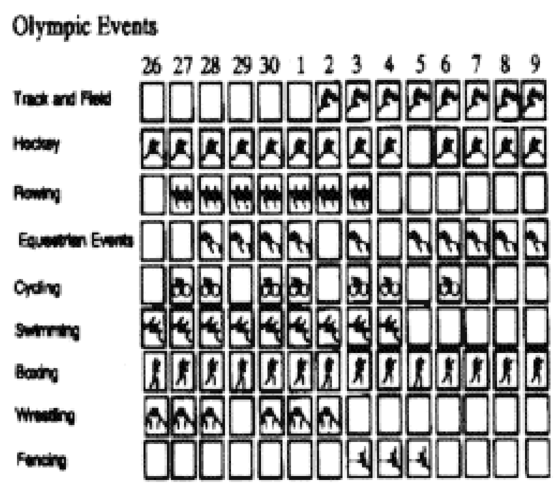
21. t-table



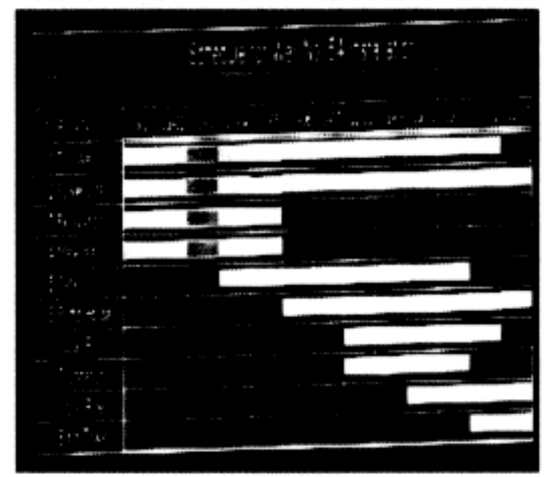
7. auto repair records

Time Charts

- Display temporal data.
  - Gantt chart, time schedule.



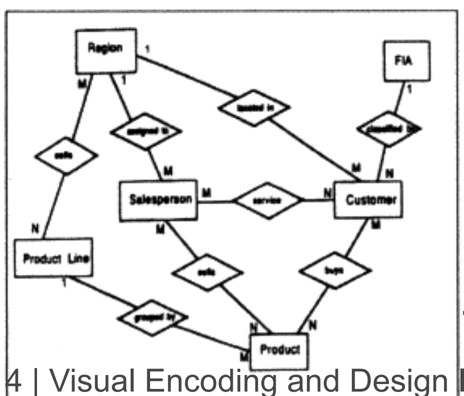
55. Olympics schedule



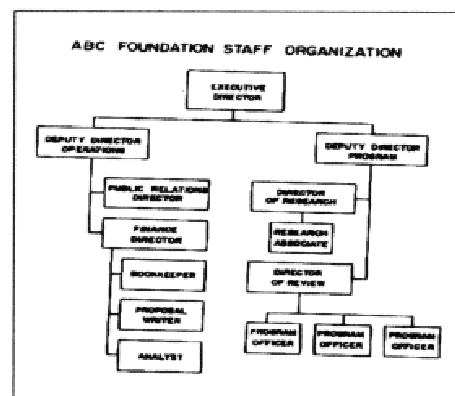
14. Gantt chart

Network Charts

- Show the relationships among components
- Symbols indicate the presence or absence of components.
- Correspondences are shown by lines, arrows, proximity, similarity, or containment.
  - Flow charts, org charts, pert charts, decision trees.



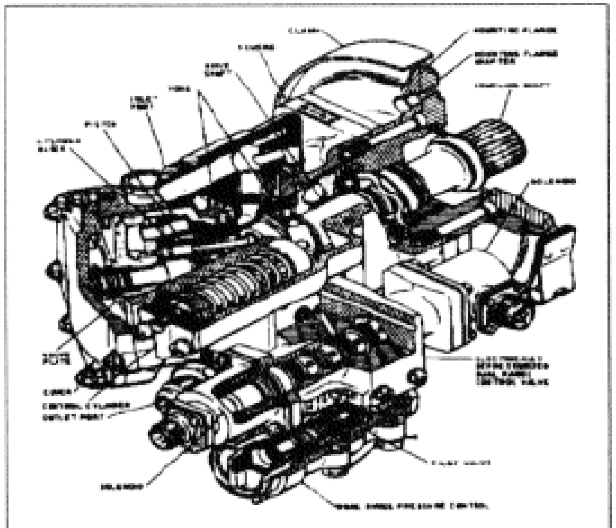
18. data model



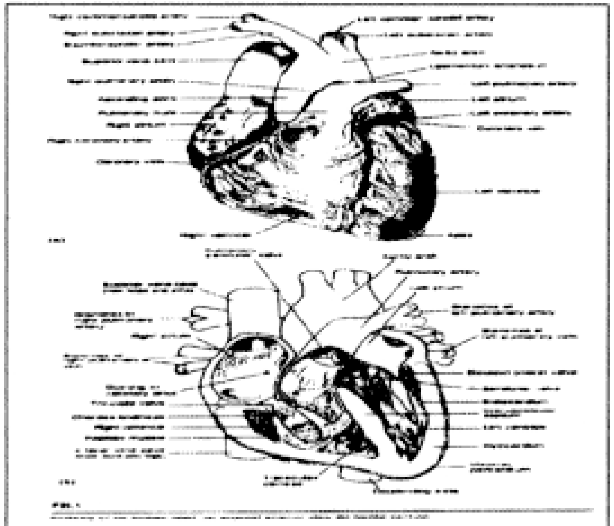
36. organizational chart

Structure Diagrams

- A static description of a physical object.
- Spatial layout expresses true coordinate dimensions of the object.
  - Cross-sections



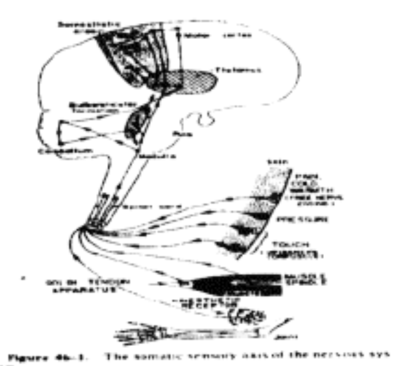
59. engine block



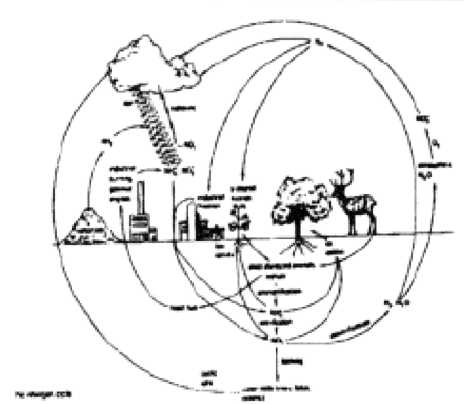
11. heart

Process Diagrams

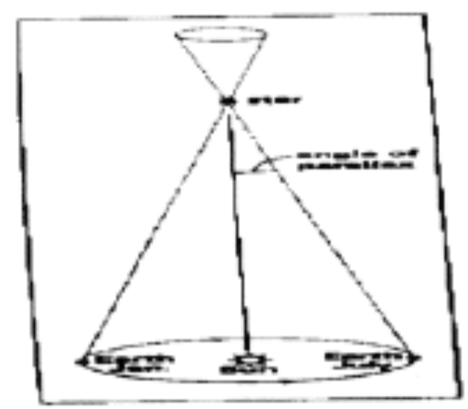
- Describe interrelationships and processes associated with physical objects.
- Spatial layout expresses dynamic, continuous, or temporal relationships among the objects.



41. nervous system



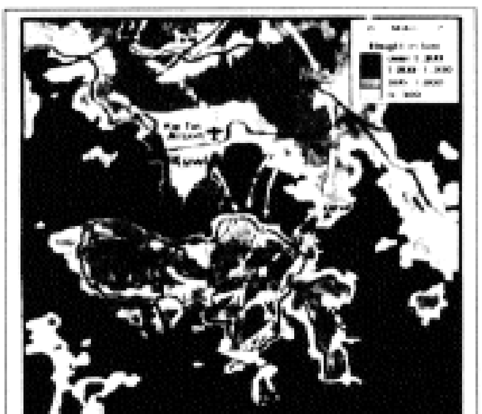
27. nitrogen cycle



54. equinox cycle

Maps

- Symbolic representations of physical geography.
  - Marine charts, topo maps, projections of world maps.
- Differ from cartograms in that cartograms super-impose quantitative data over a base map.



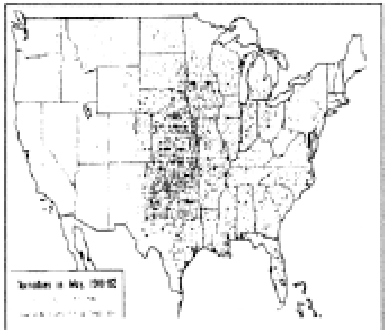
50. map of Hong Kong

26. UM campus

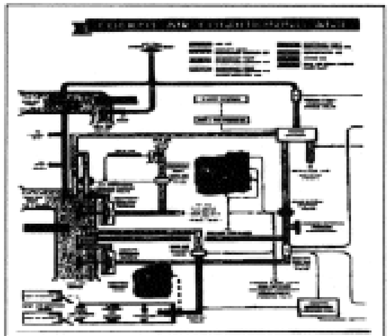
Cartograms

- Spatial maps that show quantitative data.
- Show more quantitative information than structure diagrams.
  - Chloropleths, dot maps, flow maps.

22. USA murder rates



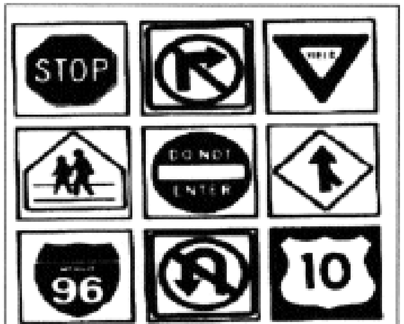
3. USA tornadoes



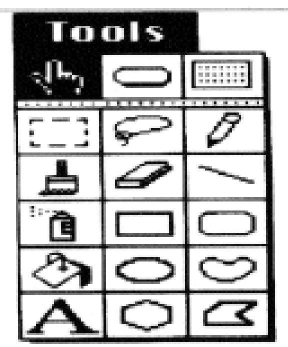
32. cockpit air conditioner

Icons

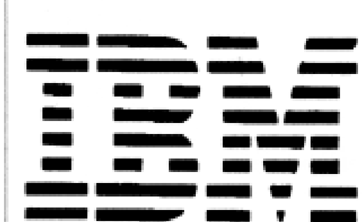
- Impart a single interpretation or meaning for a picture; a unique label for a visual representation.



30. highway signs



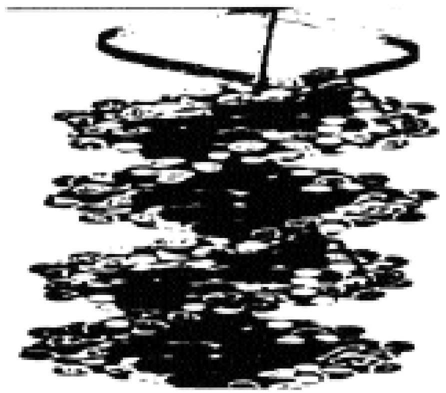
58. iconic menu



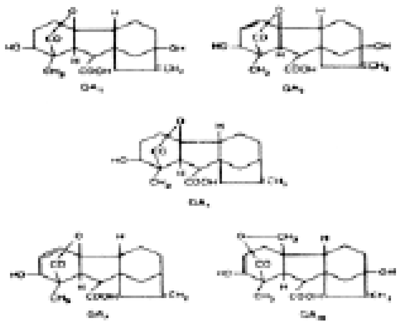
4.IBM

Put into Multiple Categories

- No real agreement on these.

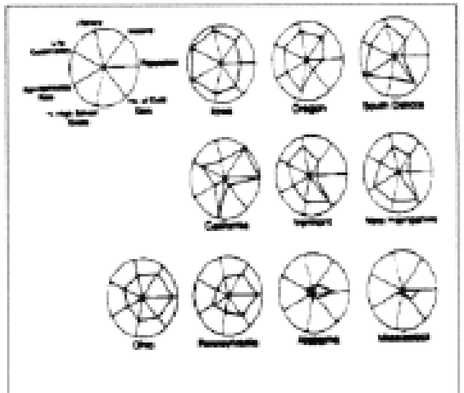


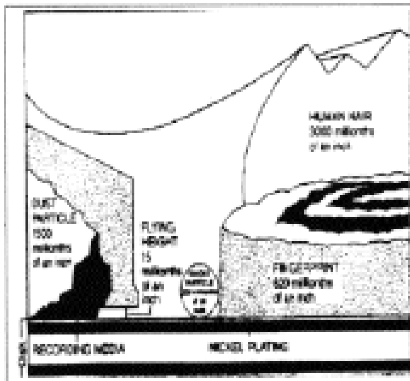
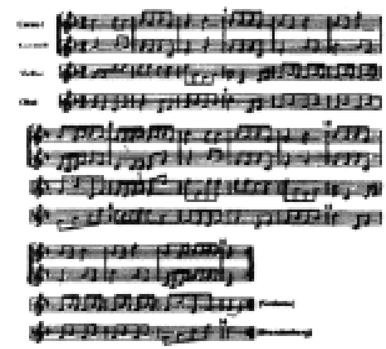
57. DNA double helix



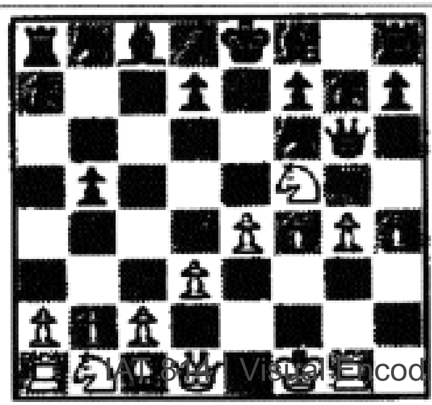
10. gibberellins

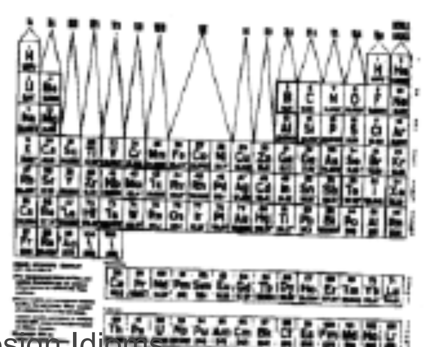
Where should these go?

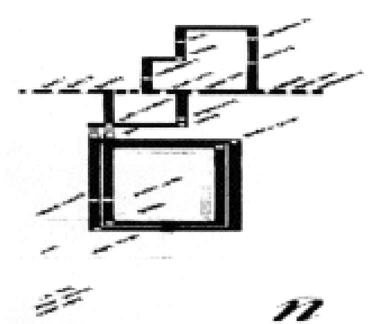

56. star diagrams


60. diskette cross section


39. music


51. chess board


35. periodic table


29. bus routes

THE BIG QUESTION

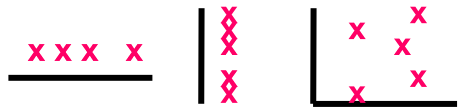
- How do you decide what kind of chart is best for what kind of data?

- Image models and visual language
  - Data → Visual feature

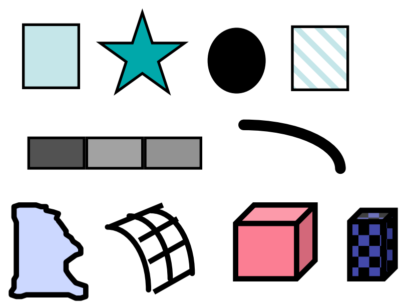
- Semiology [Bertin] : an image is perceived as a set of signs and “retinal variables”

Bertin’s Graphical Vocabulary

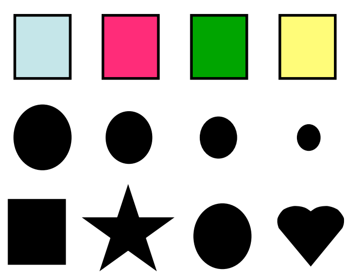
- Position



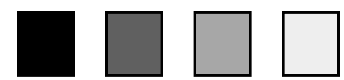
- Marks
  - Points
  - Lines
  - Areas



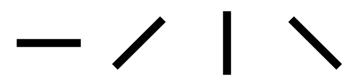
- Retinal variables
  - Color
  - Size
  - Shape



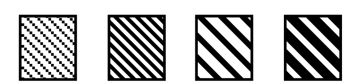
- Grayscale



- Orientation

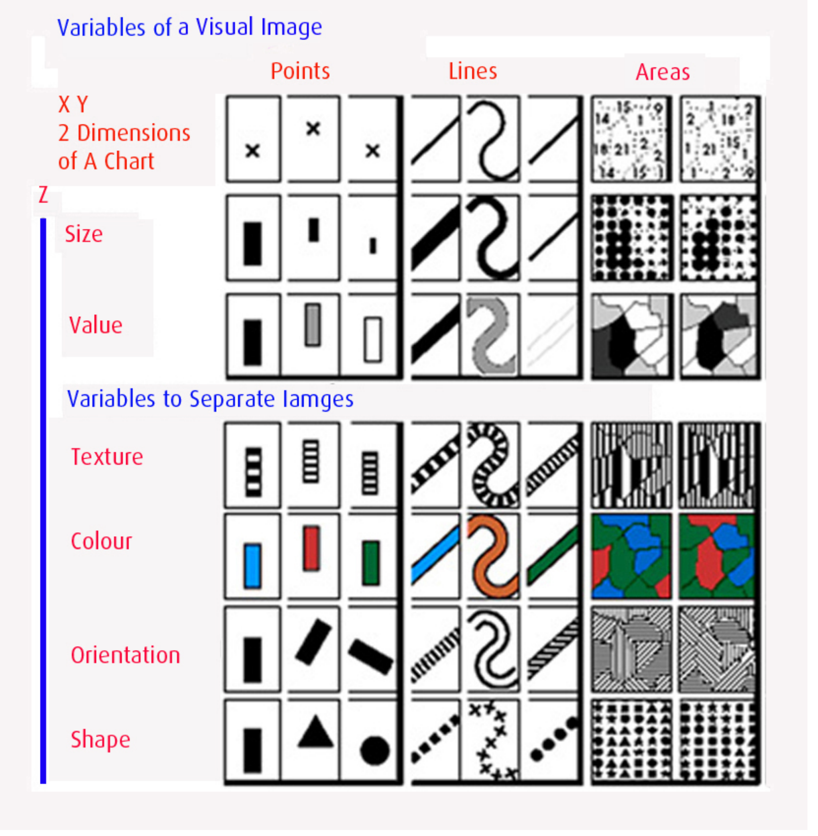


- Texture



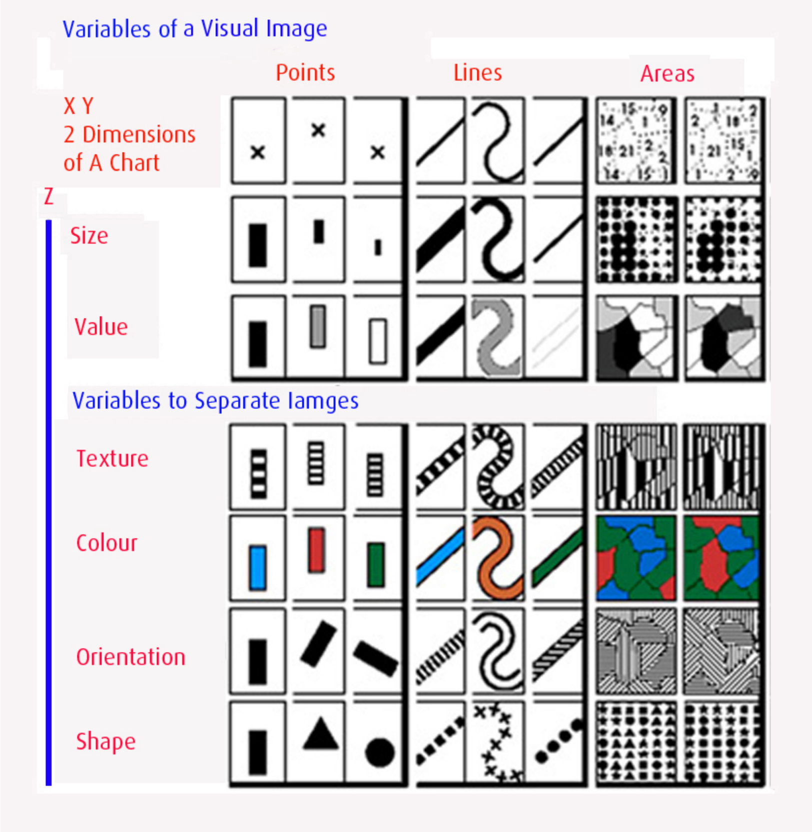
Visual encoding variables

- Position (x 2)
- Size
- Value
- Texture
- Color
- Orientation
- Shape



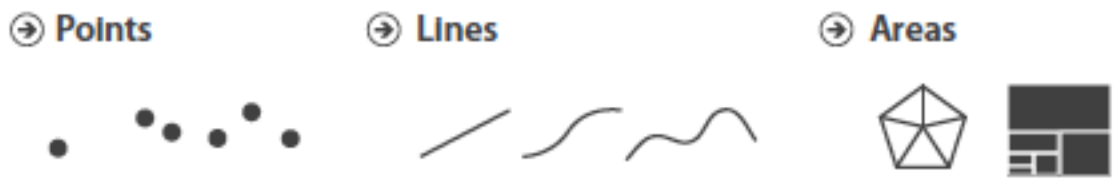
Visual encoding variables

- Position
- Length
- Area
- Volume
- Value
- Texture
- Color
- Orientation
- Shape
- Transparency
- Blur / Focus …….

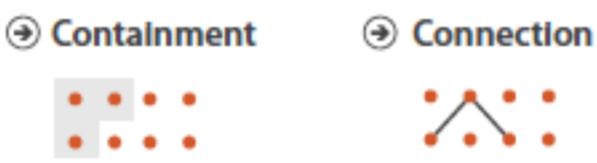


Marks and channels [Munzner]

- Marks are geometric primitives (items , tabular data)

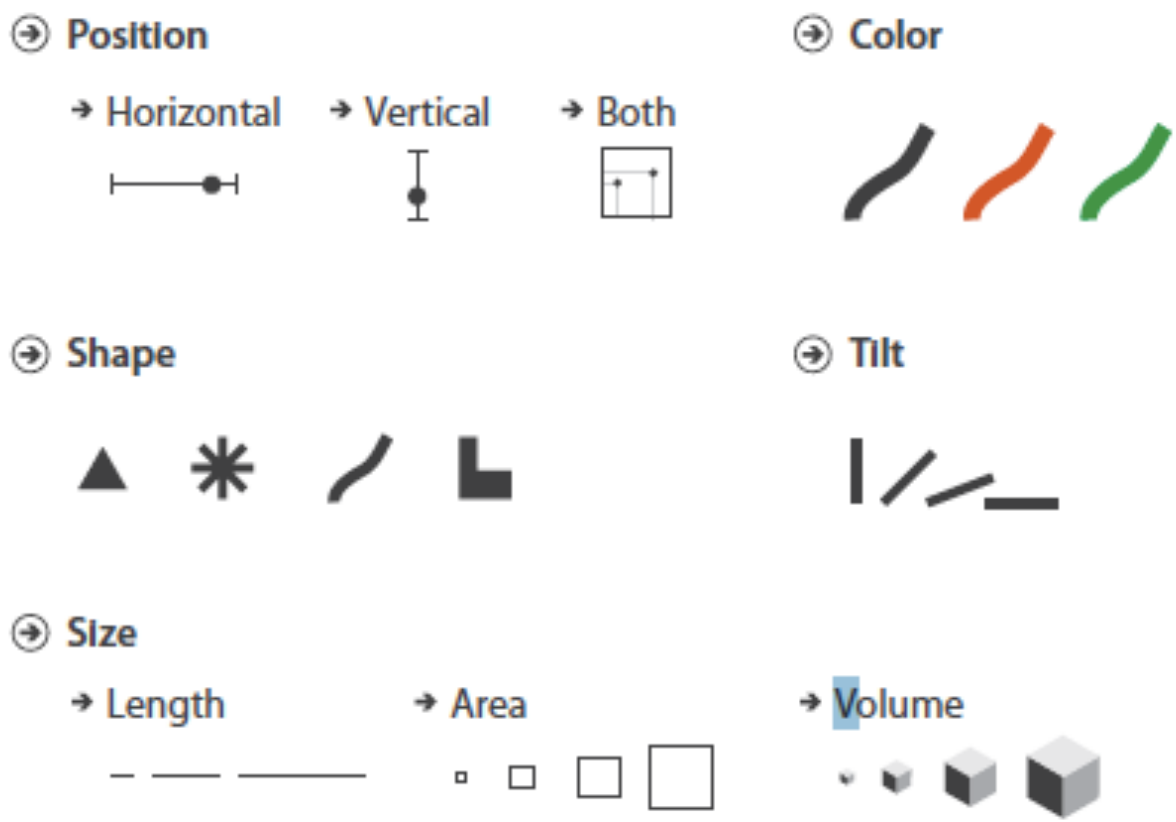


- Or links



Marks and channels

- Channels control the appearance of Marks



Channel types [Munzner]

Identity

- What and where (single)
- Shape, colour, stipple/ texture, motion

Magnitude

- How much (more)?
- Size (length, area, height); luminance and saturation; tilt; speed; position (relative)

Key visual encoding tasks

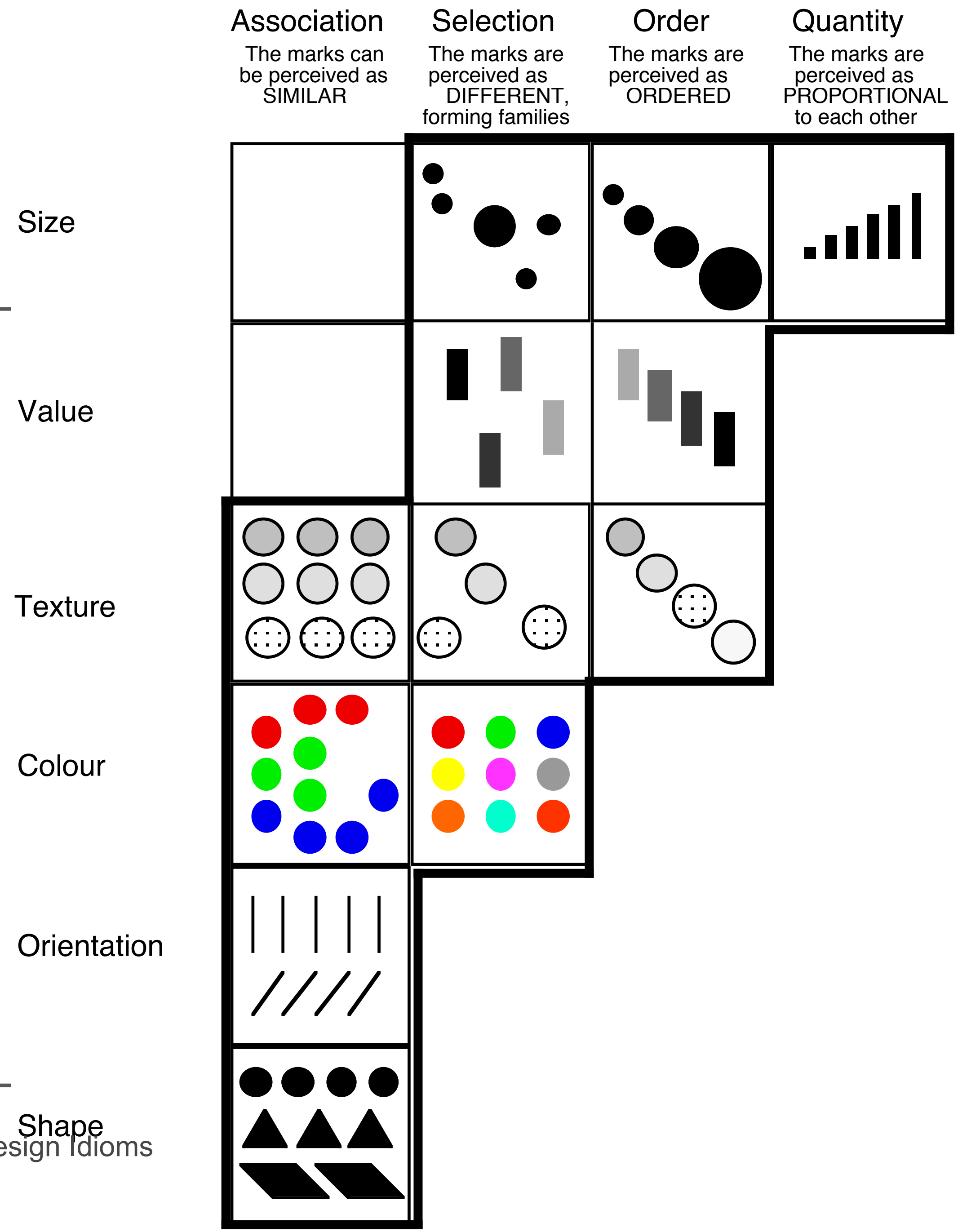
1. Selection/Discrimination:
   - Is A different from B?
2. Association:
   - Are A and B similar (related in some way)?
3. Order
   - Is A > B?
4. Quantification: value
   - How much is A?

Key visual encoding tasks

5. Quantification: a number can be deduced from differences
   - How much bigger is A than B?

6. Capacity (length) [Carpendale]
   - The number of distinctions possible using the variable
   - How many different things can we represent with this variable?

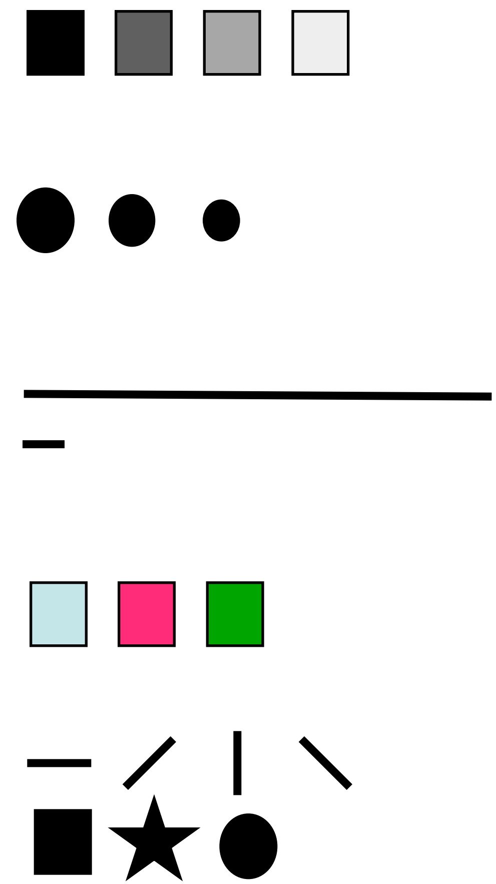
Interpretation of Bertin's guidance regarding the suitability of various encoding methods to support common tasks



Interpretations of Graphical Vocabulary

Discrimination vs ordering semantics (Senay & Ingatious 97, Kosslyn, others)

- Density (Greyscale)
  - Darker -> More
- Size / Length / Area
  - Larger -> More
- Position
  - Leftmost -> first,
  - Topmost -> first
- Hue
  - no intrinsic meaning;
  - good for highlighting
- Slope / Shape
  - no intrinsic meaning;
  - good for contrast



Visual variables: selectivity

Selectivity: Different values are easily seen as different

“Is A different from B?”

Worst case: visual properties of all objects need to be looked at one by one

Visual variables: Associativity

- **Associativity:** Similar values can easily be grouped together

  "Is A similar to B?"

Full selectivity / associativity ⟷ No selectivity / associativity

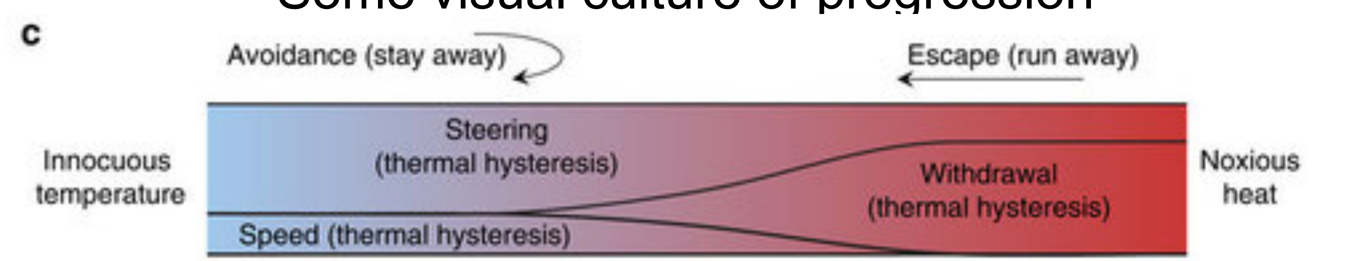
Positioning > {size, brightness} > {color, orientation (for points)} > texture > shape

Visual variables: Order

Order: Different values are perceived as ordered

"Is A more/greater/bigger than B?"

- Size and brightness are ordered
- Orientation, shape, texture are not ordered
- Hue is "not really" ordered
  - Some visual culture of progression



Visual variables: quantity

Quantity: A number can be deduced from differences

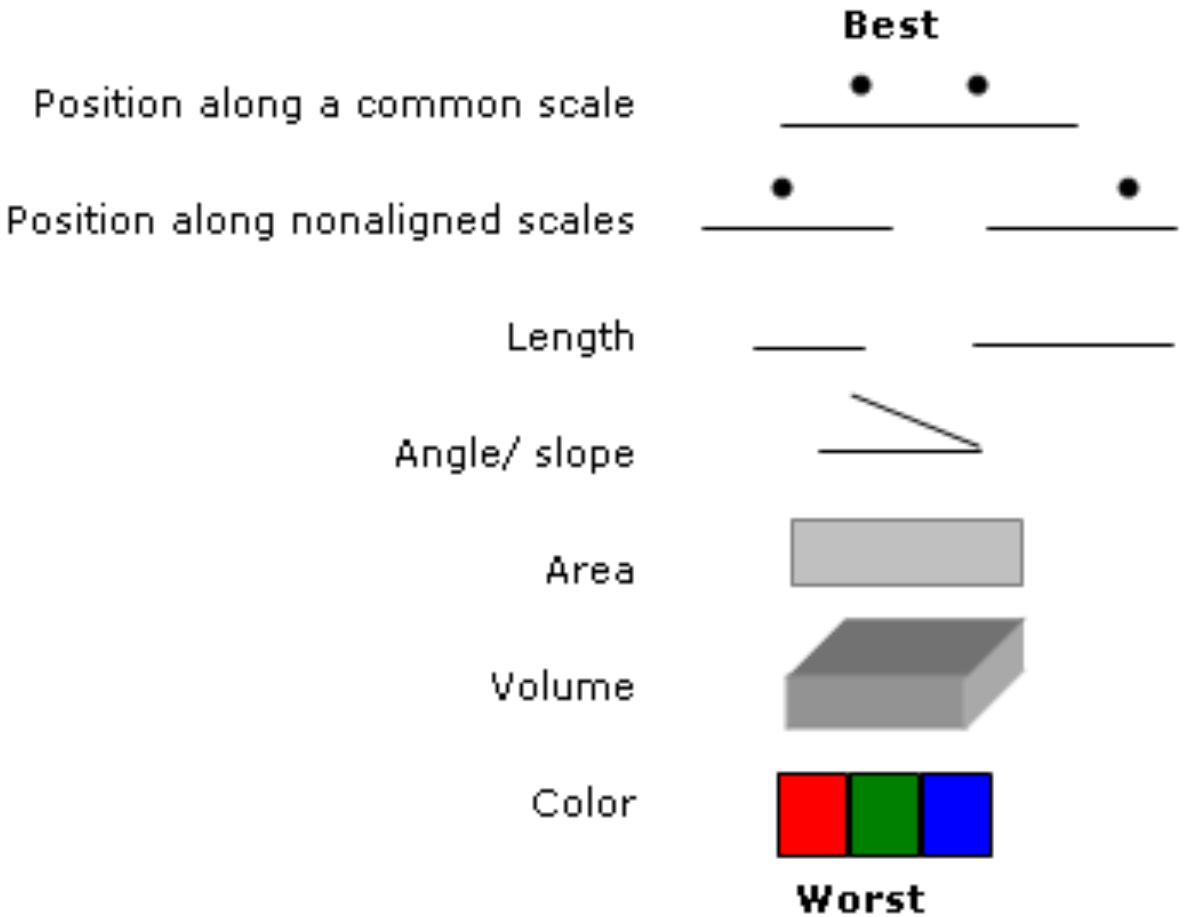
- “How much is the difference between A and B?”
    - Position is quantitative, size is somewhat quantitative
    - The other variables are not quantitative

Visual variables: capacity

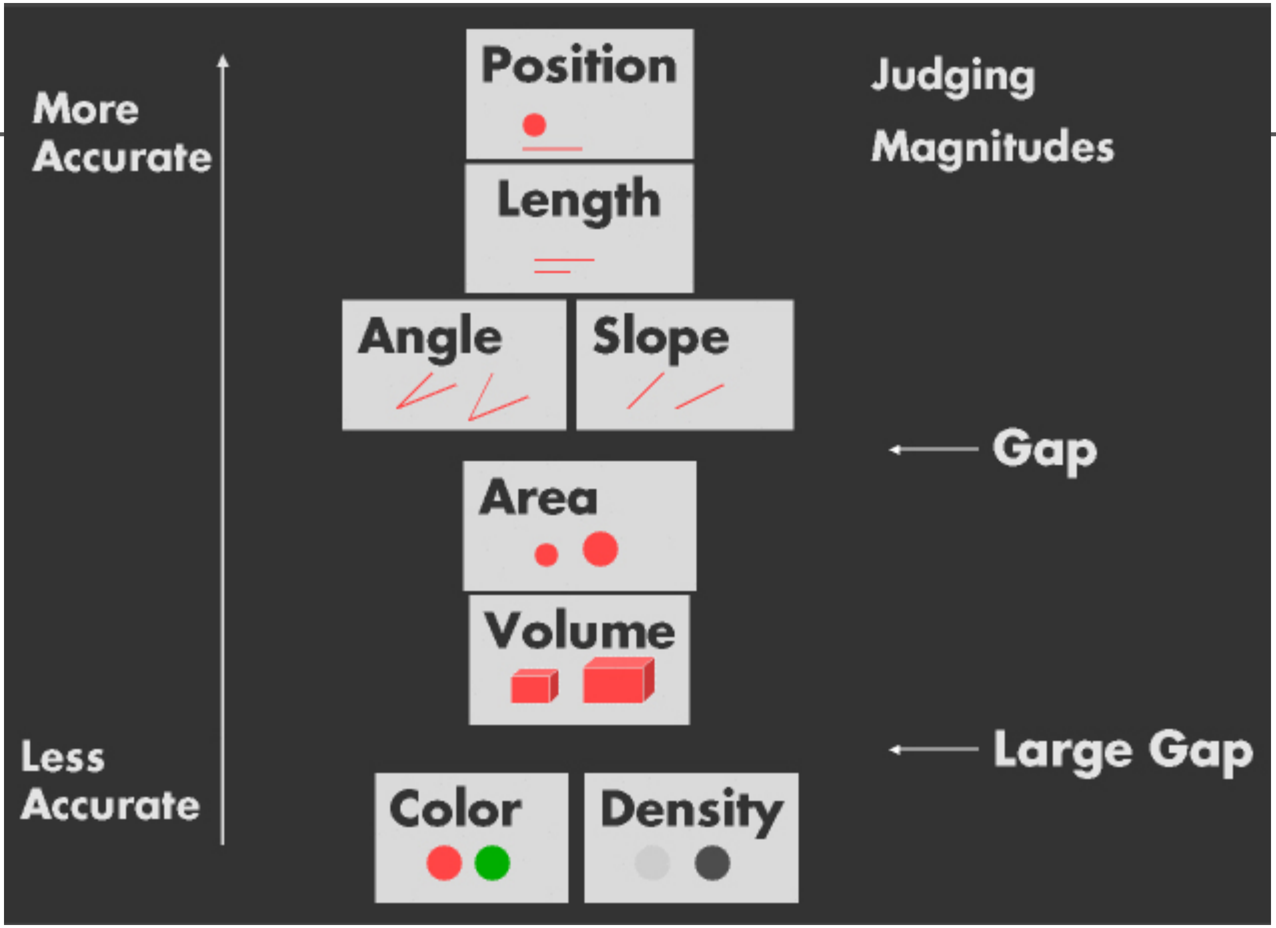
Length: The number of distinctions possible using the variable

- "How many different things can we represent with this variable?"
  - Shape, Texture: infinite, but …
  - Brightness, hue: 7 (Association) – 10 (Distinction)
  - Size: 5 (Association) -20 (Distinction)
  - Orientation: 4

Cleveland's (1984) Graphical Feature Interpretation Hierarchy



Based on graphic (Figure 2) in Presentation Graphics (white paper) by Leland Wilkinson, SPSS, Inc and Northwestern Uiv.


More Accurate
Less Accurate
Position
Length
Angle
Slope
Area
Volume
Color
Density
Judging Magnitudes
Gap
Large Gap

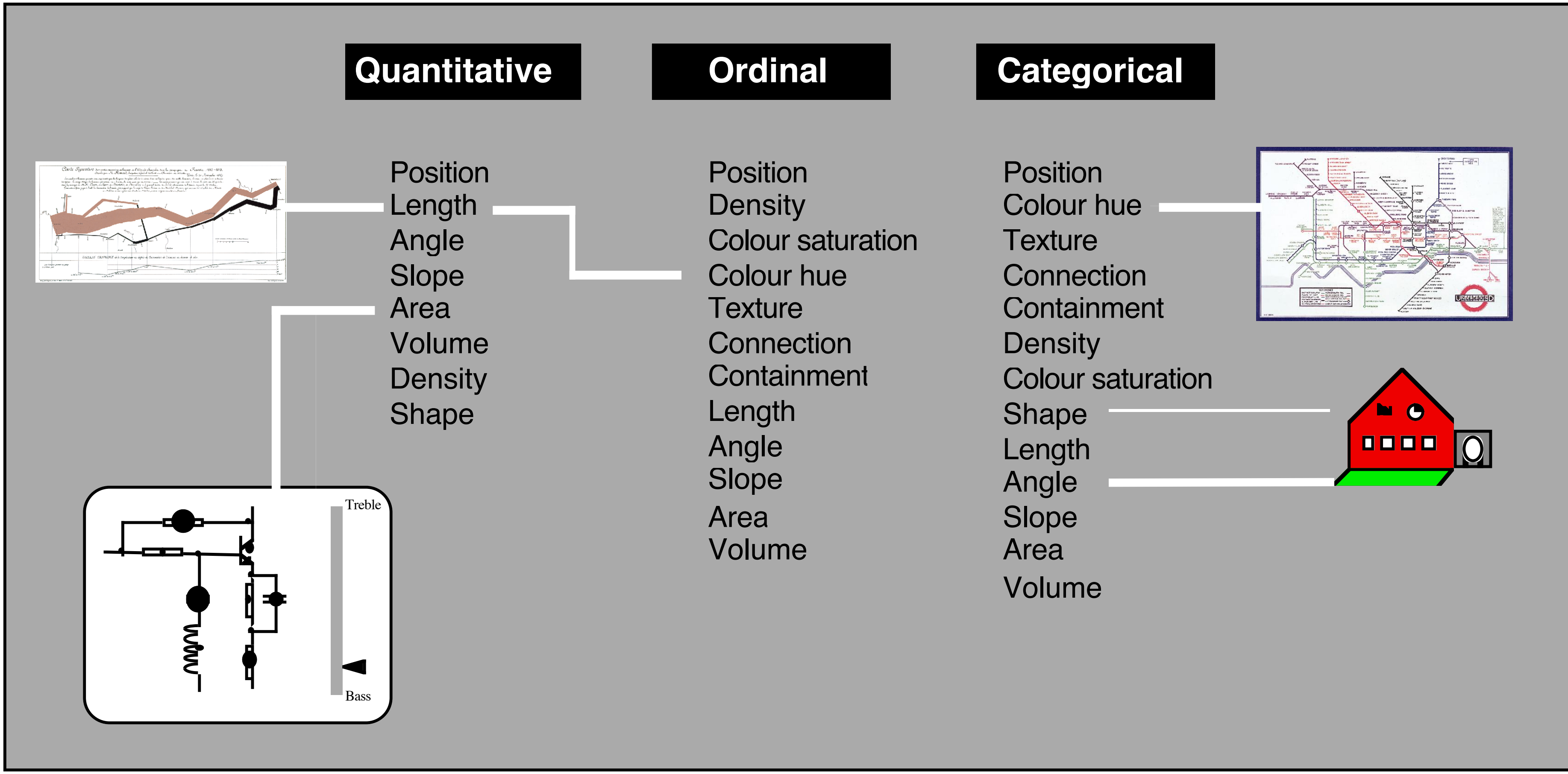


Figure 3.45 Mackinlay's guidance for the encoding of quantitative, ordinal and categorical data

Few's Table:

Attribute	*Quantitative*	*Qualitative*
Line length	●	
2-D position	●	
Orientation		●
Line width		●
Size		●
Shape		●
Curvature		●
Added marks		●
Enclosure		●
Hue		●
Intensity		●

Choosing Visual Encodings

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 important way

Using visual variables: Tufte 1

- "Sameness of a visual element implies sameness of what the visual element represents" (Tufte, 2006)

1. Characteristics of visual variables determine their use
   - e.g. Ordered values have to be represented by ordered visual variables
2. Be consistent concerning relations of similarity, proportion and configuration
3. Adhere to conventional uses of visual variables

   e.g. in cartography use blue color for water
4. Scales should be made up of visually equidistant values of a variable

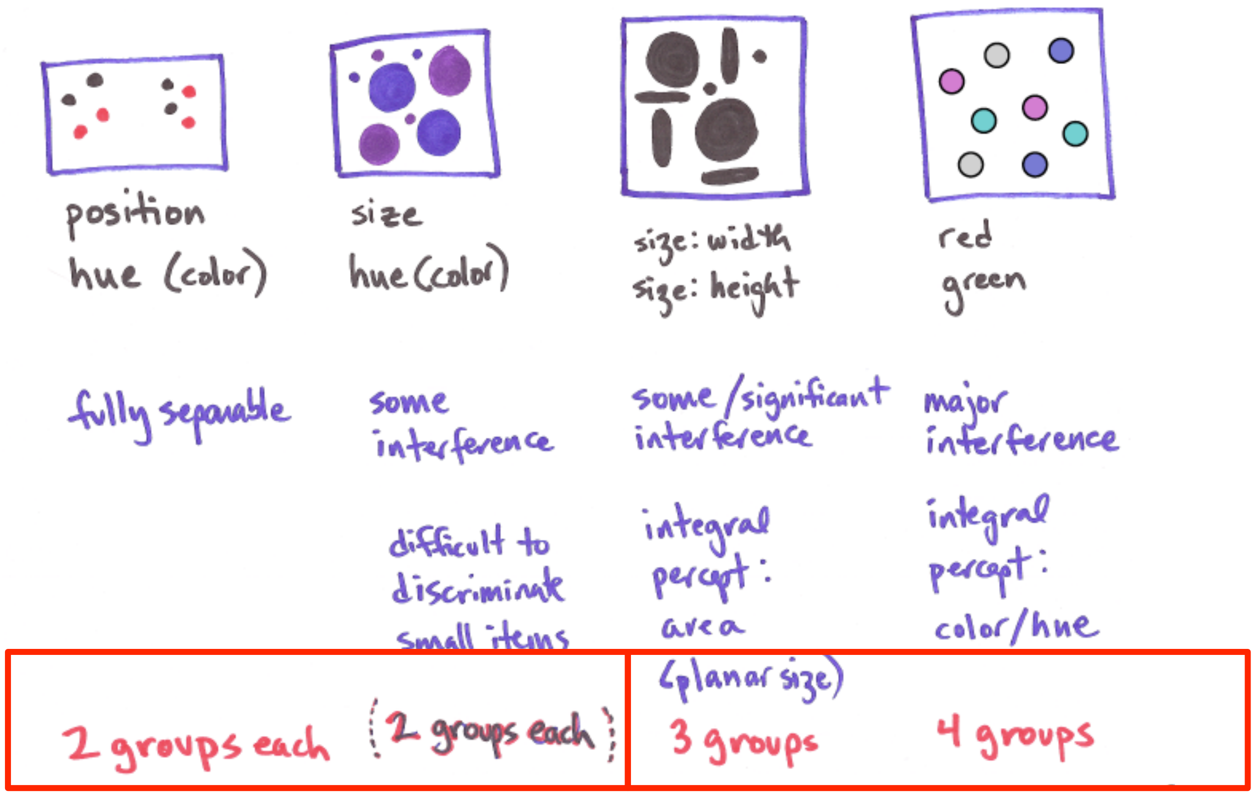
Using visual variables : Tufte 2

5. The full range of a visual variable should be used
   - e.g. when using shades of gray, use from white to black

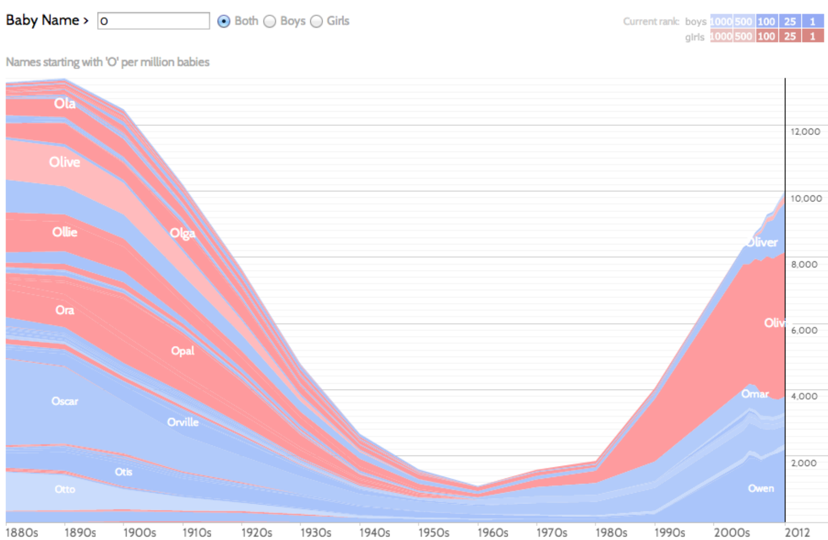
6. The number of visual variables of a visualization should correspond to the dimensionality of the represented information

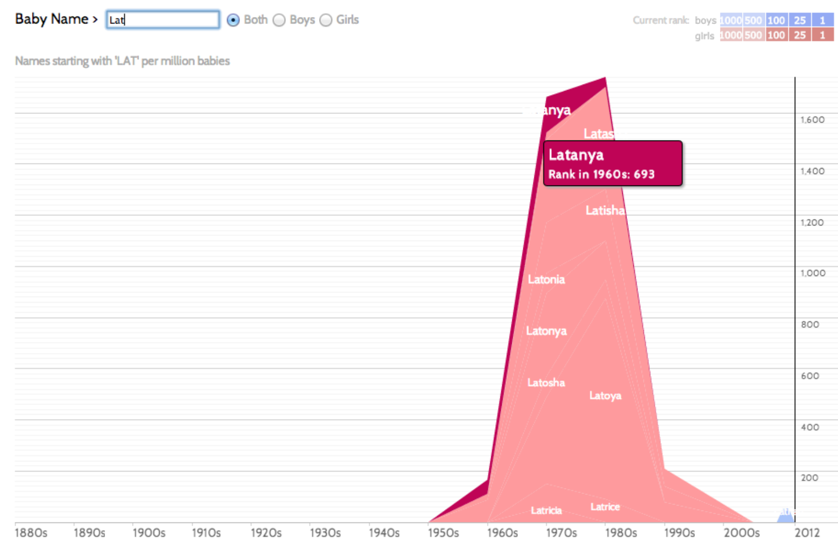
7. When combining two visual variables, if people should be able to analyze the two attributes independently, then **separable** variables should be used

Integral vs separable: recap



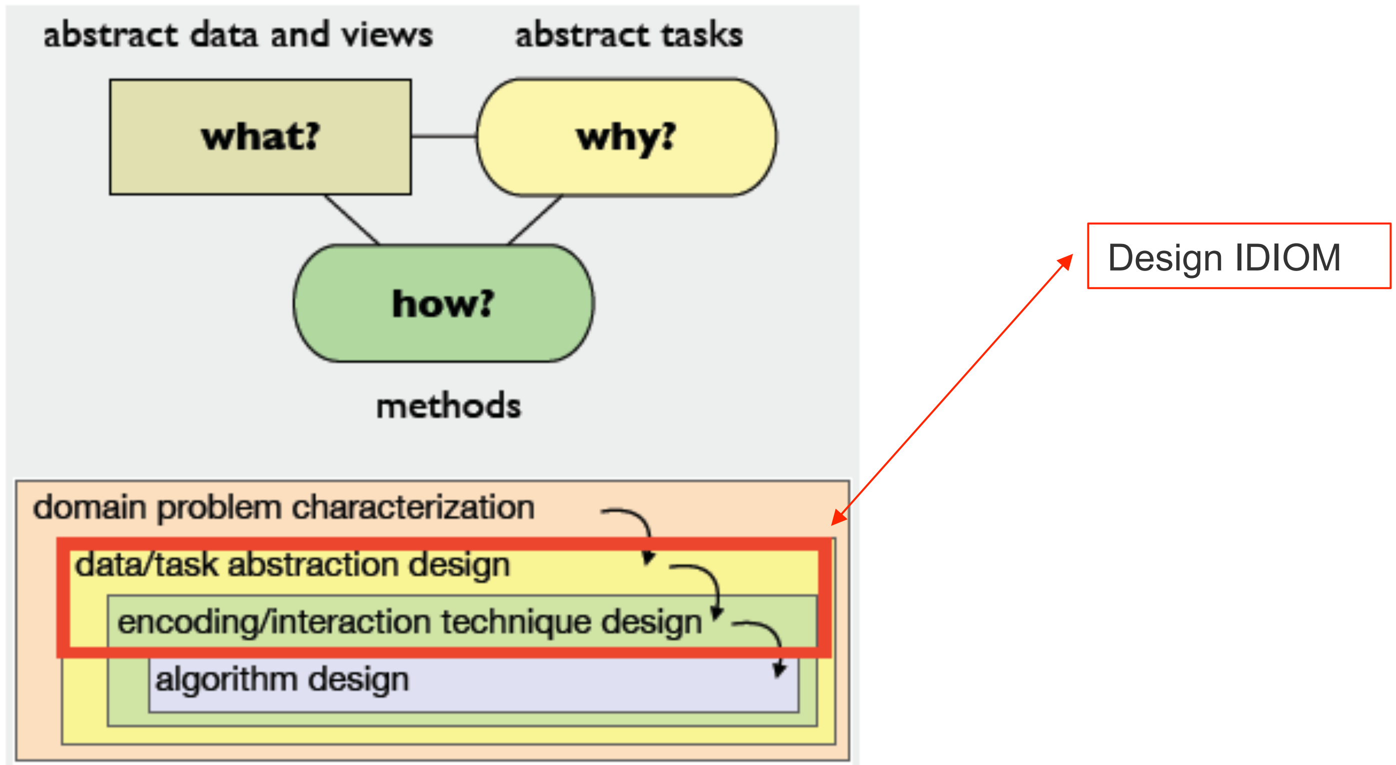
Why does not define how completely!



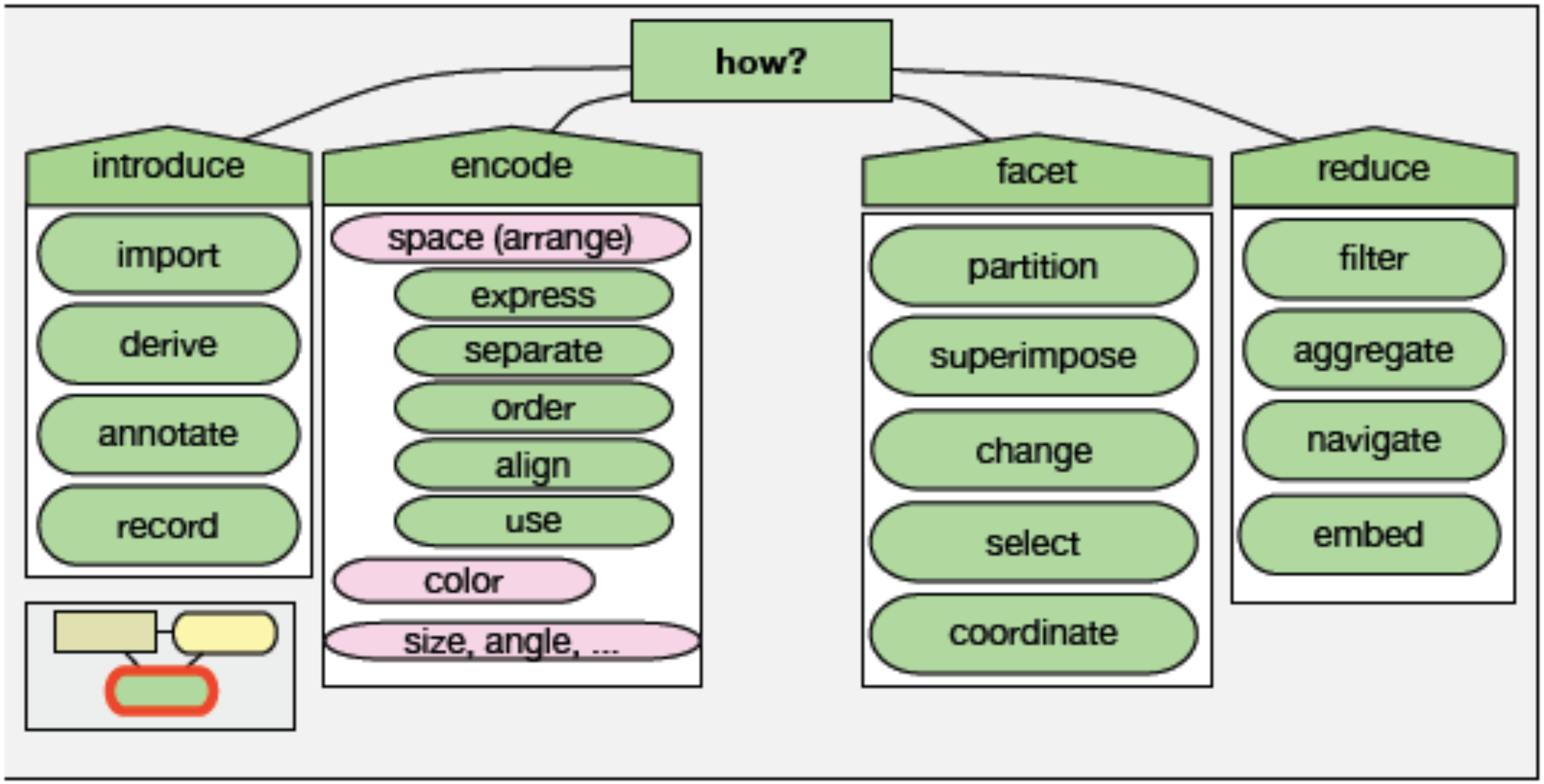


the decision about **why** is separable from **how** the idiom is designed: discovery can be supported through a wide variety of idiom design choices.

A Framework for Analysis (Munzner)

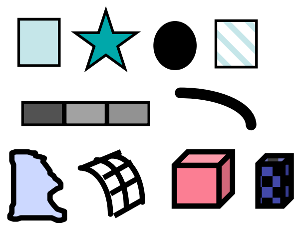


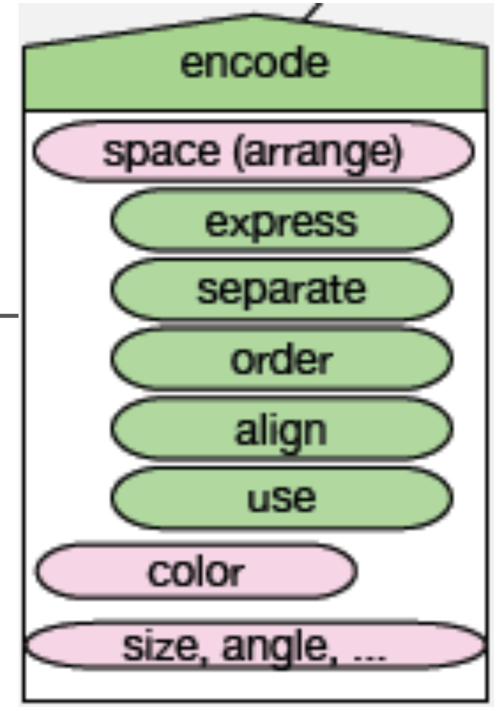
Designing Vis Idioms (Munzner)



Review: (Munzner)

- Marks
  - Points
  - Lines
  - Areas





- Channels

Position	hue
Size	saturation
Shape	lightness
orientation	texture …

What is this?

How Much/many of something is there?

Credit: T. Munzner, 2014

Fundamental principles

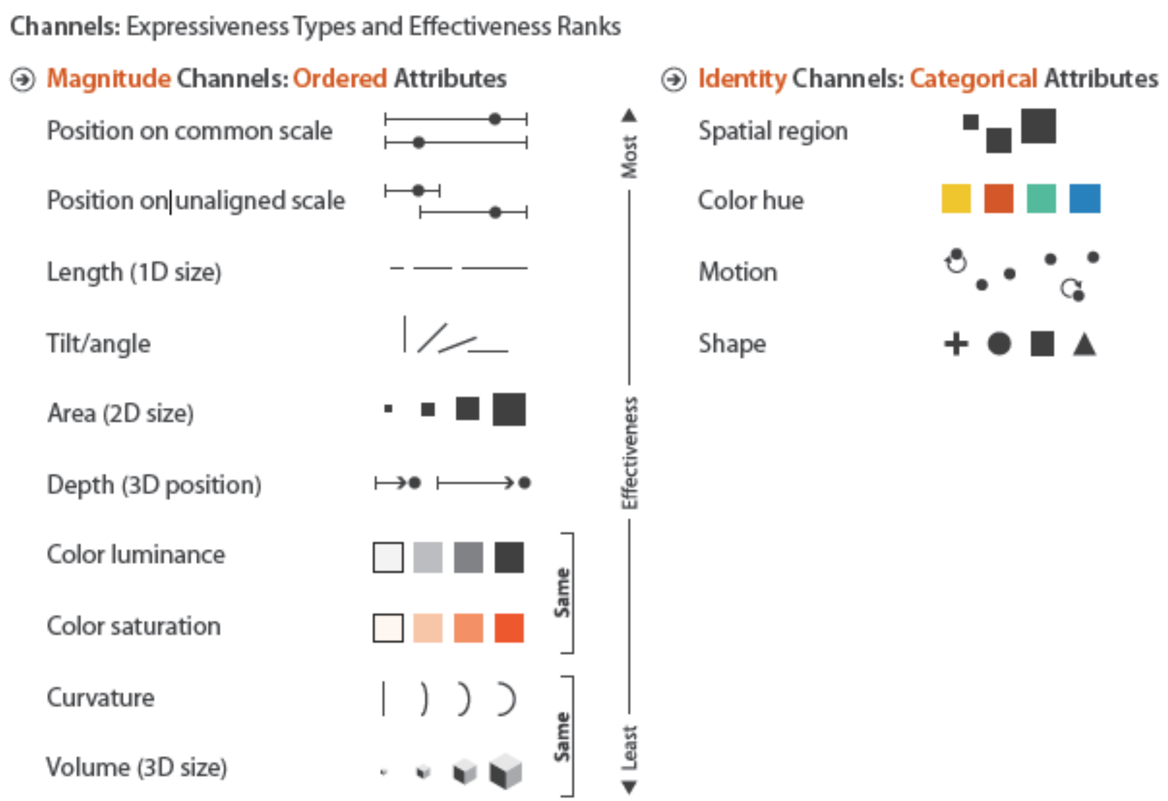
Expressiveness:

- the visual encoding should express all of, and **only**, the information in the dataset attributes

Effectiveness:

- the **importance** of the attribute should match the **salience** of the channel :Use the strongest accurate channels for the most important interpretation tasks accuracy,discriminability, separability, popout
- Channel visual precedence

Channels: Expressiveness and Effectiveness



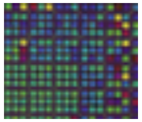
Credit: T. Munzner, 2014

Space

Spatial Channels

- Values
  - Express
- Regions
  - Separate
  - order
  - Align
    - 1D
    - 2D
- Given Use
  - Geographic
  - Fields
    - scalar

Spatial Layouts

- Parallel
- Rectilinear
- Radial
- Spacefilling 
- Dense 

Categorical

What/where

Planar position

Hue

Shape

Stipple/texture

Relational/Same category

Grouping

Containment (2D)

Connection

Similarity (other channels)

Proximity (position)

Ordered/Quantitiative

How Much

Position common scale

Position unaligned scale

Length

Tilt/angle

Area

Curvature

Lightness

Saturation

Texture

density

Categorical

What/where

Planar position

Hue

Shape

Stipple/texture

Relational/Same category

Grouping

Containment (2D)

Connection

Similarity (other channels)

Proximity (position)

Ordered/Quantitiative

How Much

Position common scale

Position unaligned scale

Length

Tilt/angle

Area

Curvature

Lightness

Saturation

Texture

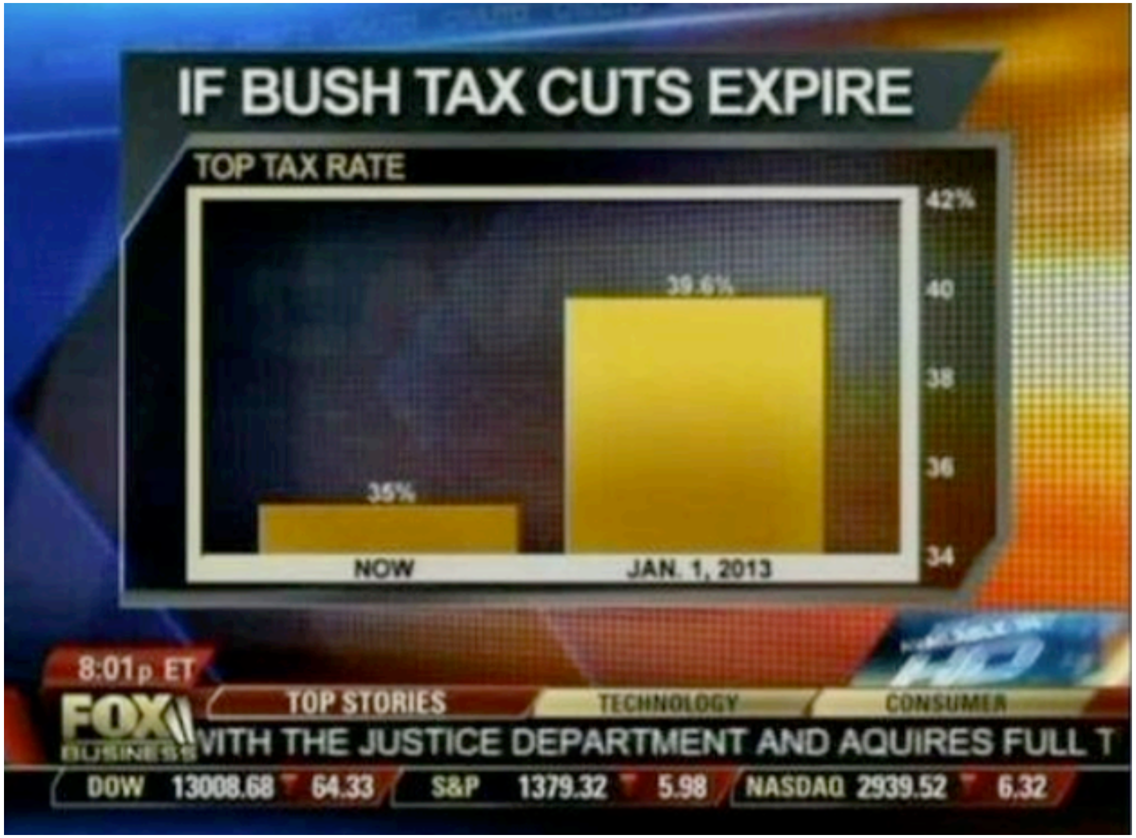
density

Revisiting some examples

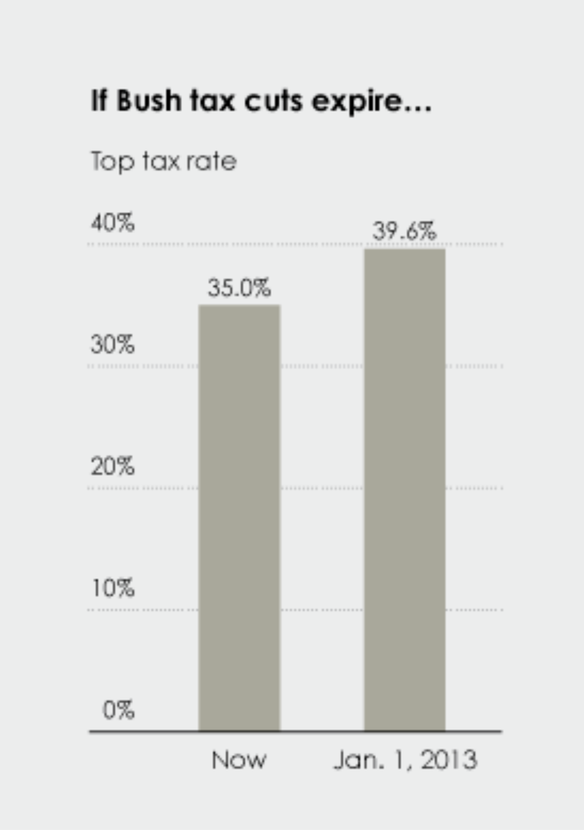
What are these showing with respect to data TYPE ?

What channels are they using and are they effective ?

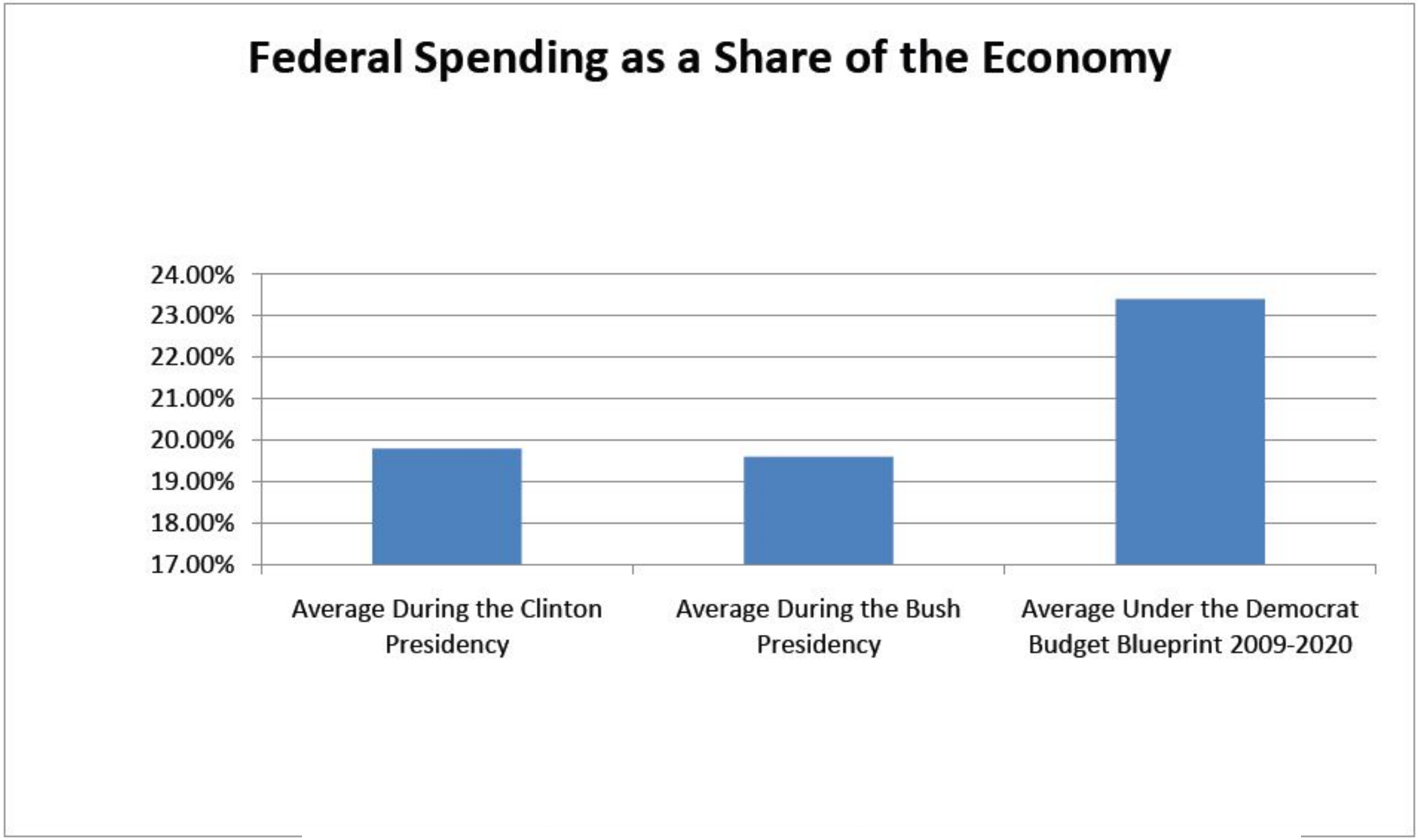
Fox News to the (lie) – what is this??



The real story



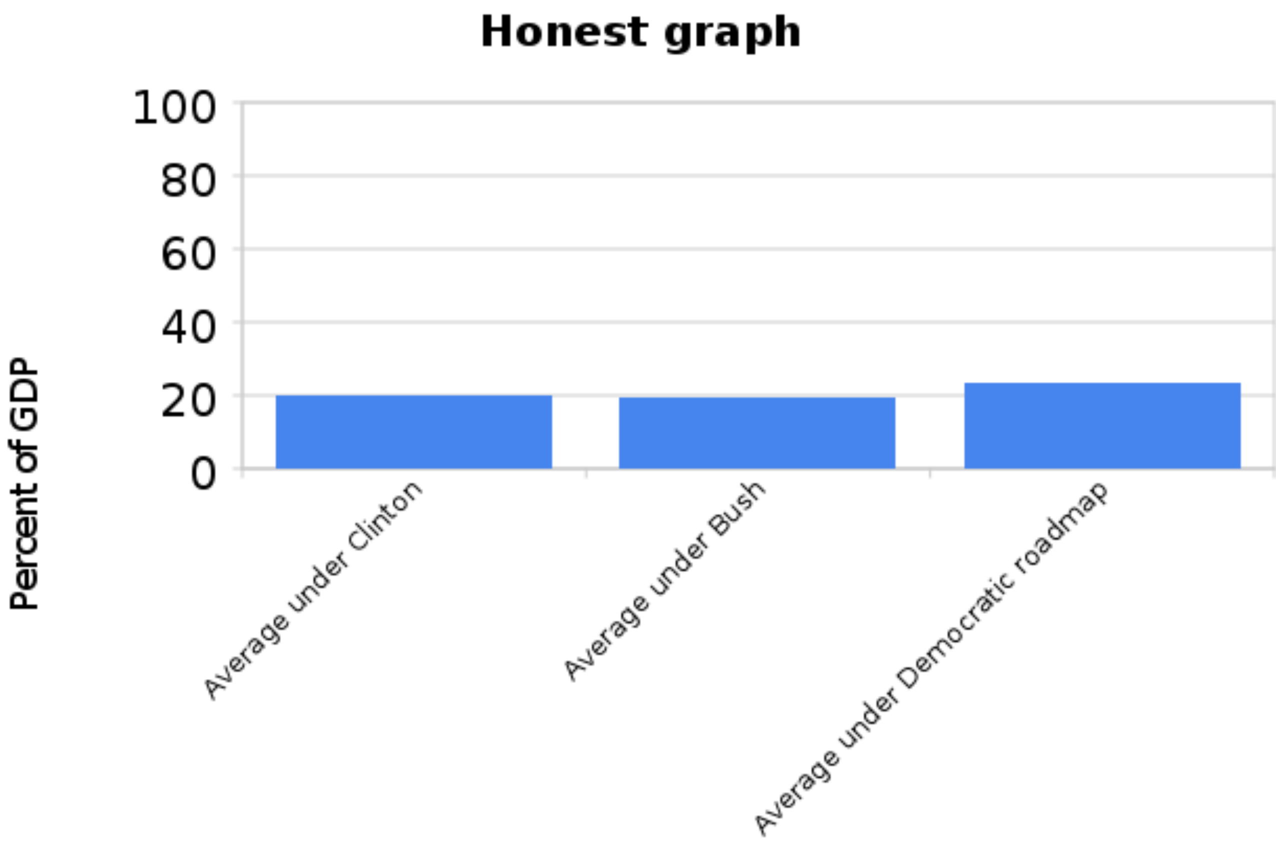
Lies, Damn Lies, and Bad Graphs



http://www.tnr.com/blog/jonathan-cohn/77893/lying-graphs-republican-style

SFU

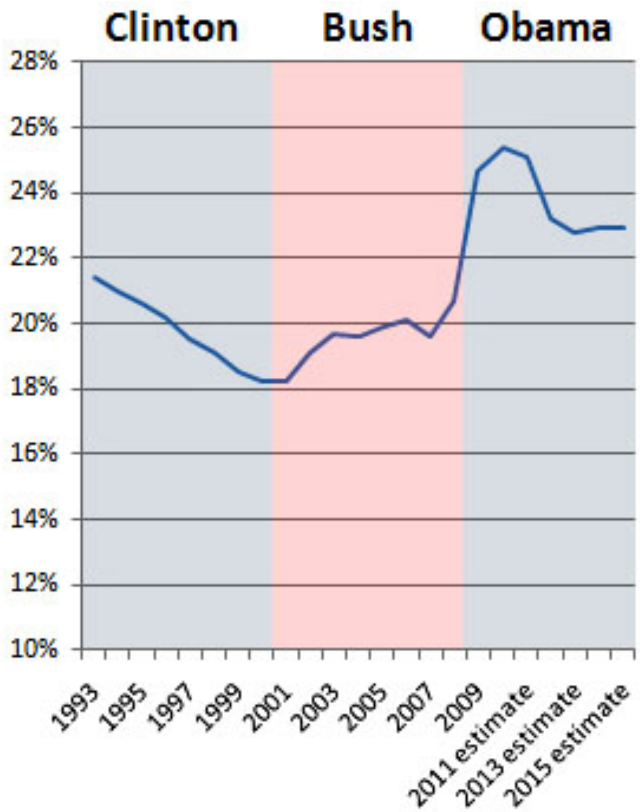
More honest



http://www.tnr.com/blog/jonathan-cohn/77893/lying-graphs-republican-style

SFU

Better



http://www.tnr.com/blog/jonathan-cohn/77893/lying-graphs-republican-style

SFU

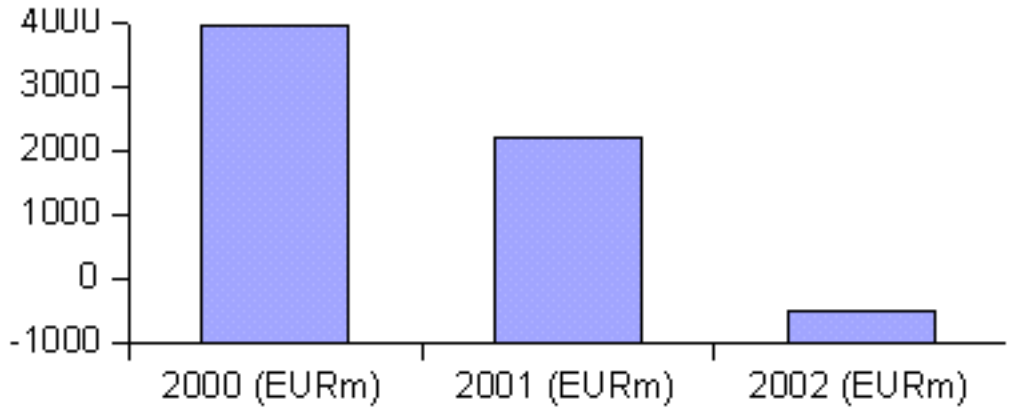


Chart 8: Imaginative net profits for Nokia 2000-2002, with x-axis at -1000

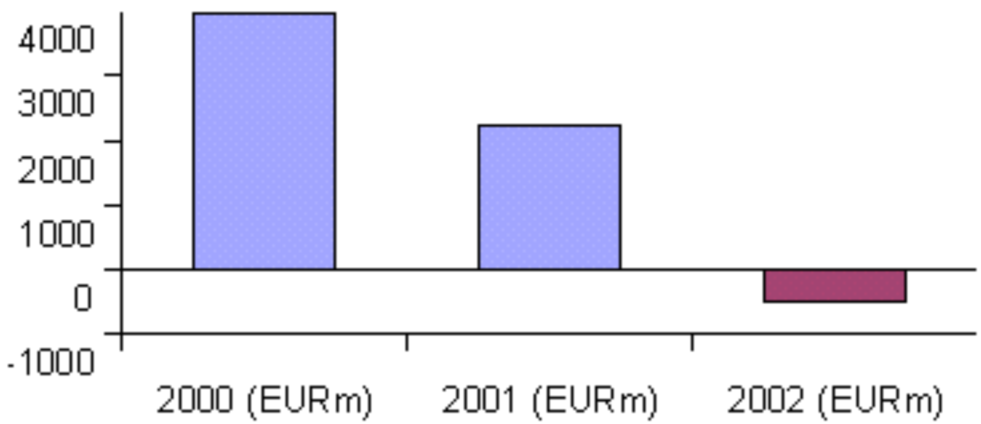


Chart 7: Imaginative net profits for Nokia 2000-2002

Perspective as well as value scales

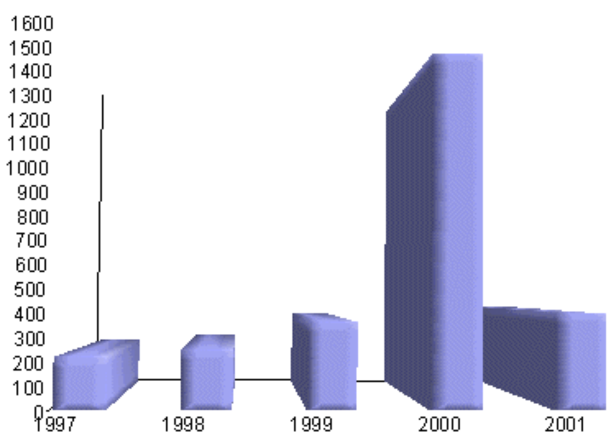


Chart 19: Net profits for Sonera 1997 - 2001 (EURm), nice 3d effect

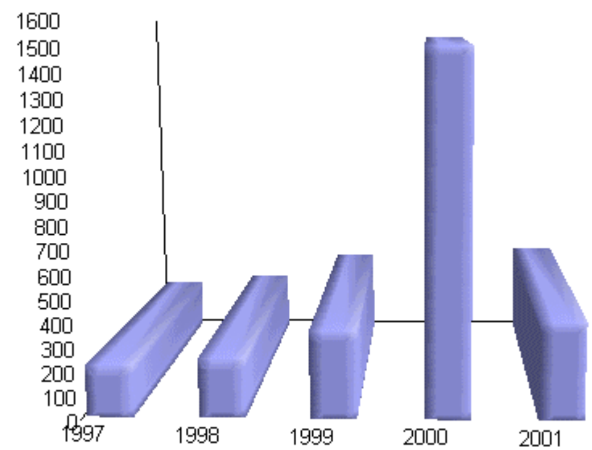
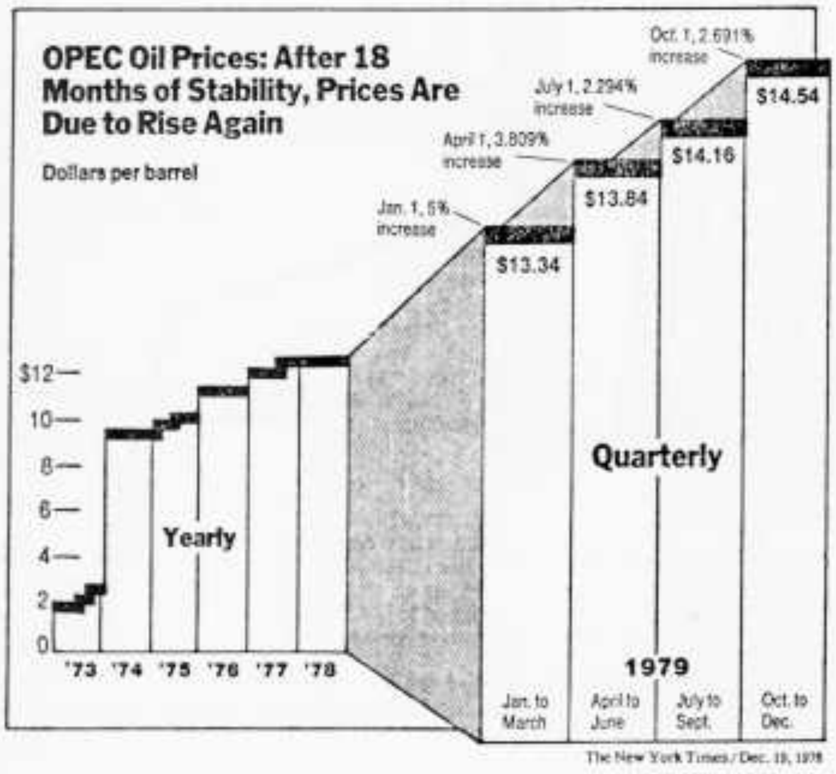


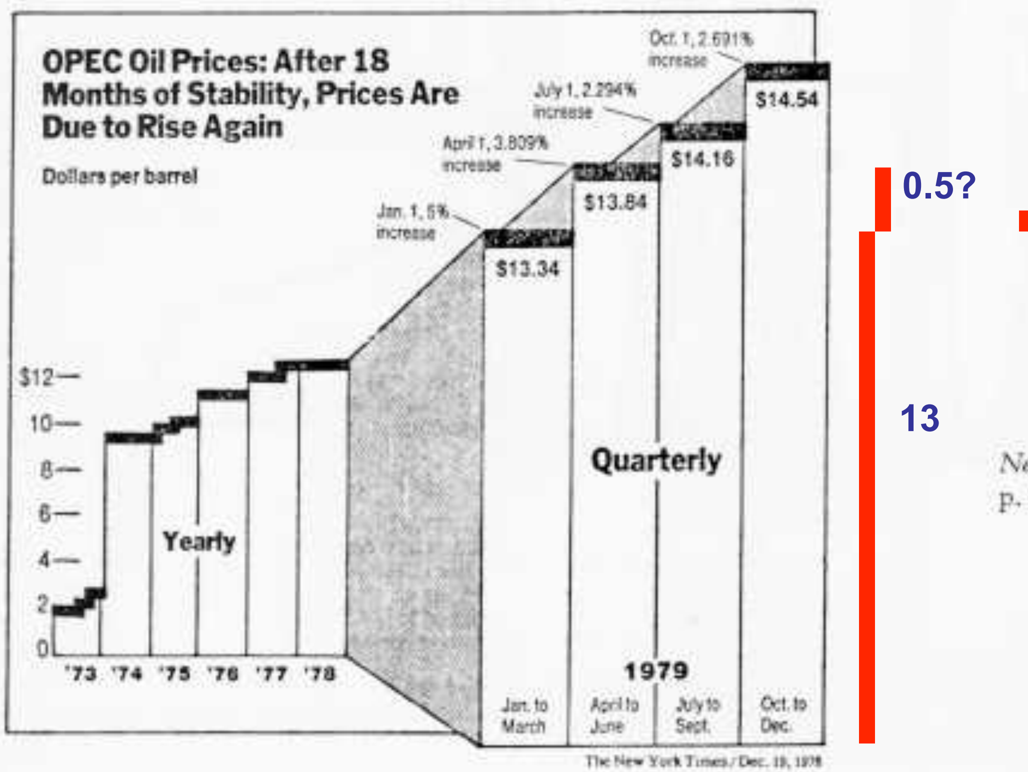
Chart 20: Net profits for Sonera 1997 - 2001 (EURm), another perspective into same 3d effect



New York Times, December 19, 1978, p. D-7.

Changing Scale

Design variation corrupts this display:



New York Times, December 19, 1978, p. D-7.

Changing Scale

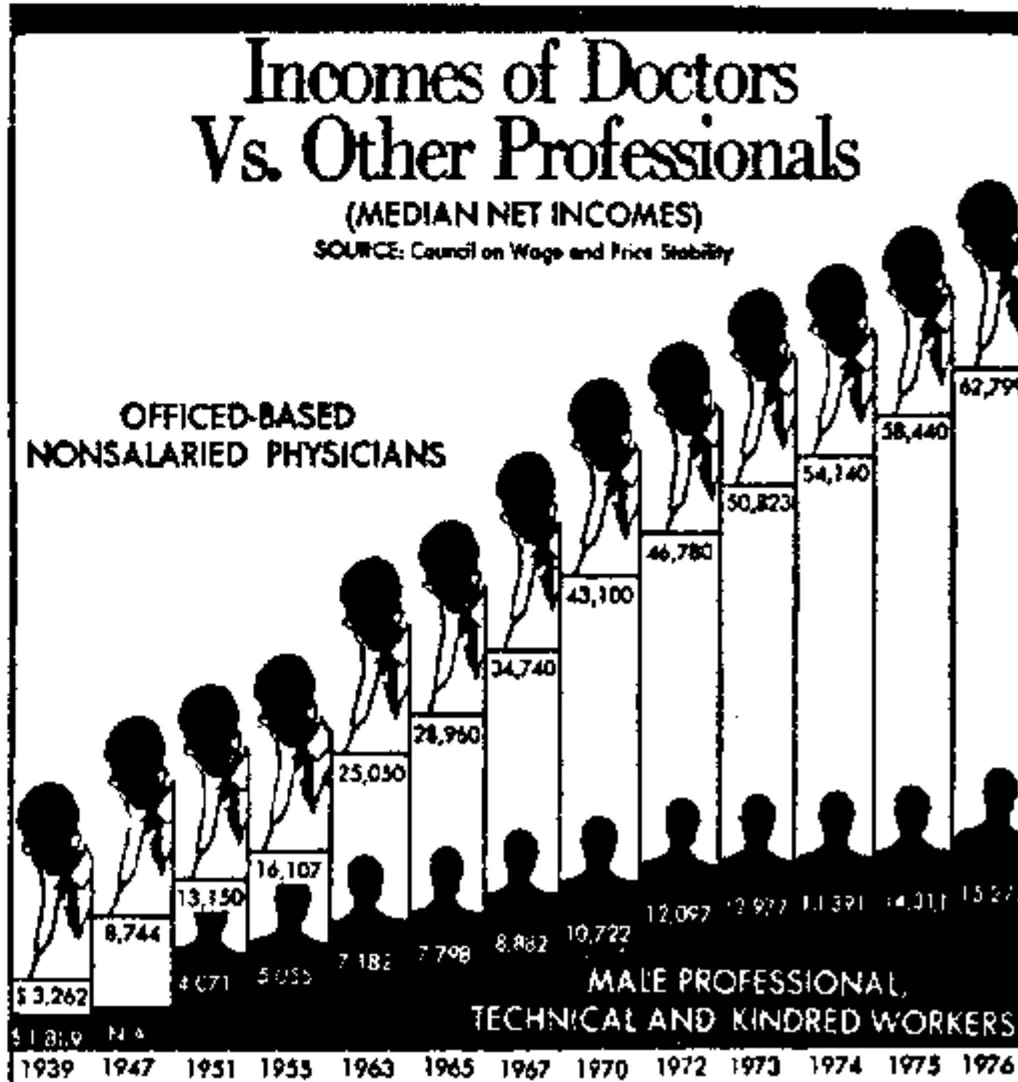
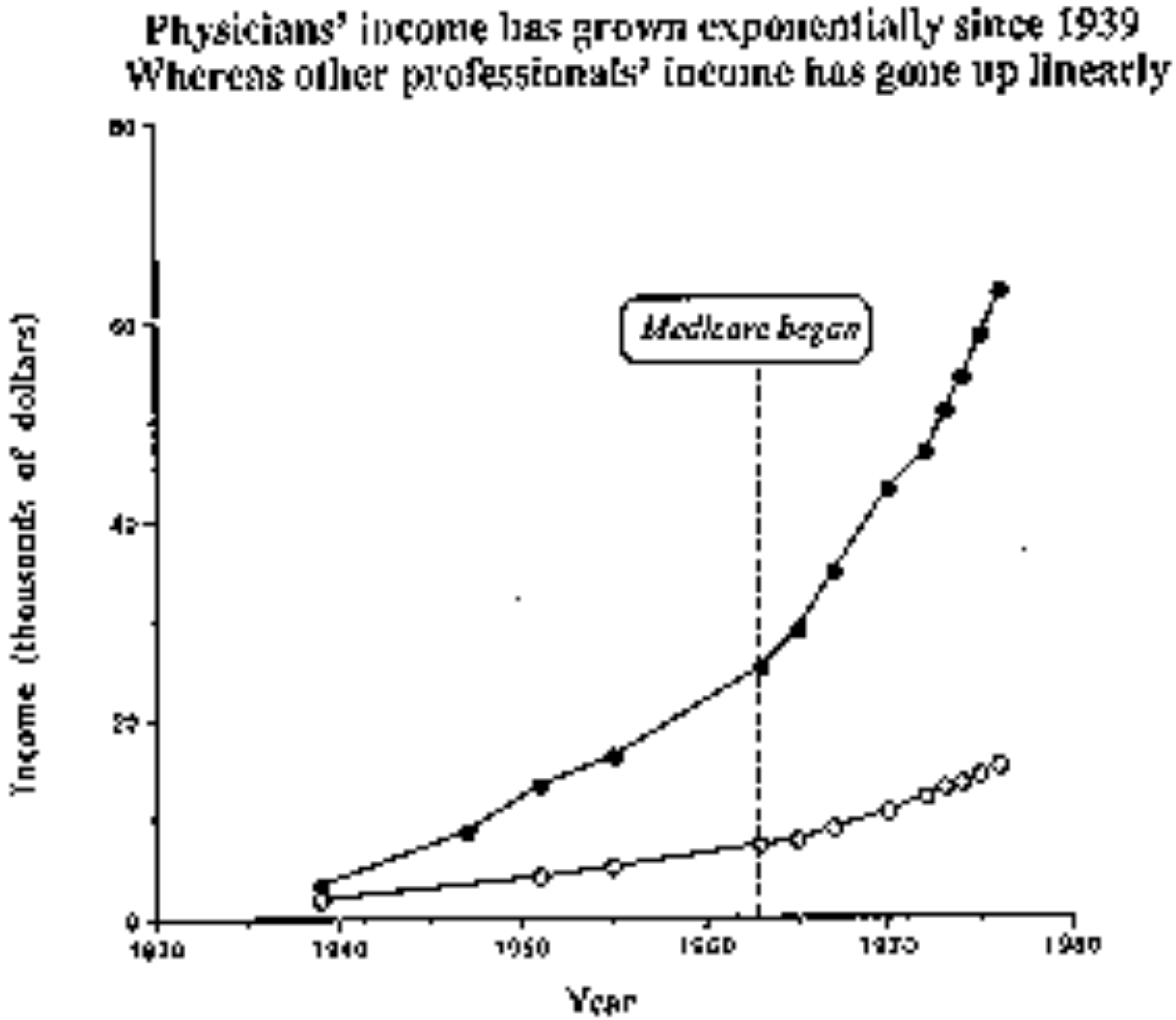


Figure 13. Changing scale in mid-axis to make exponential growth linear (© The Washington Post).

SFU

…with linear time scale

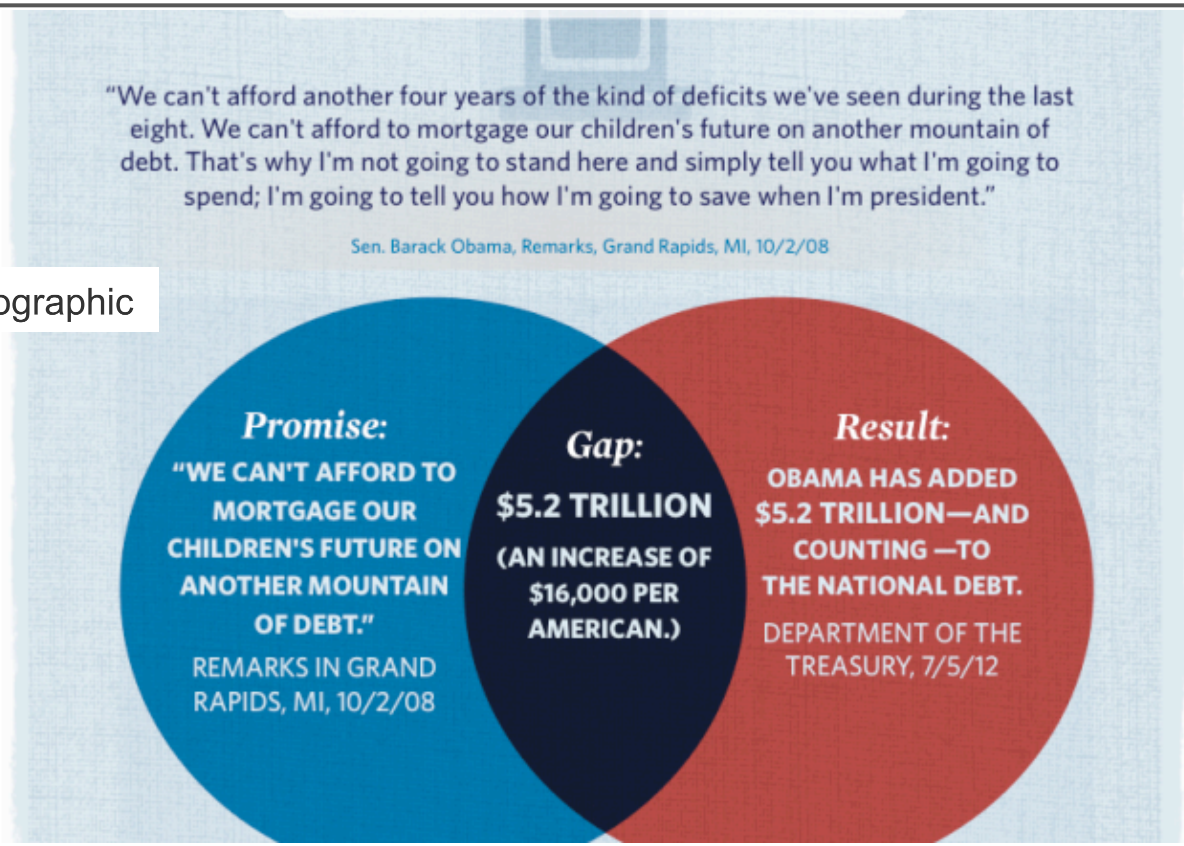


SFU

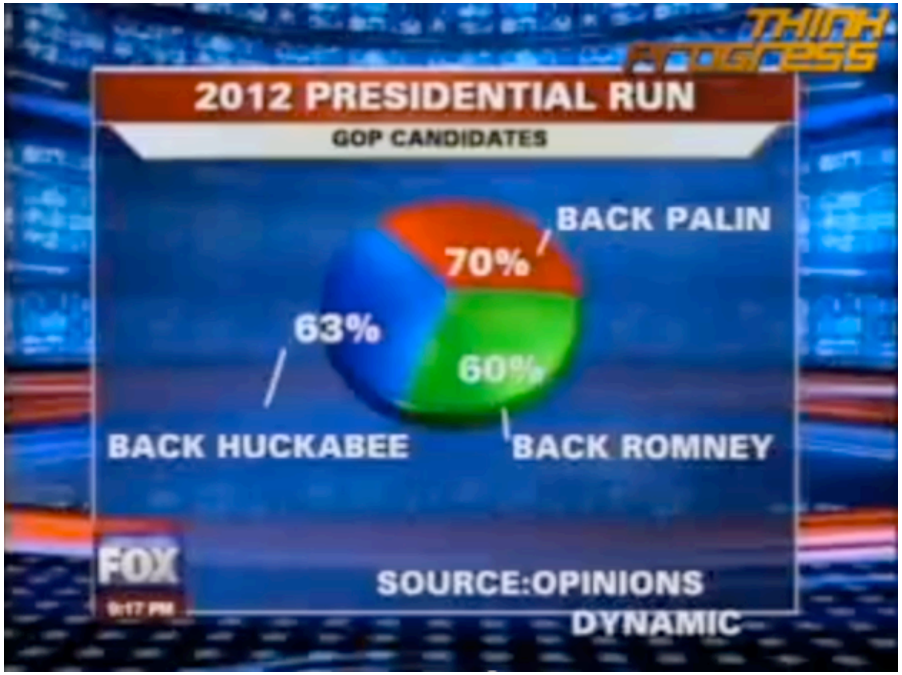
And more

Umm, what's wrong here?

Mitt Romney campaign infographic

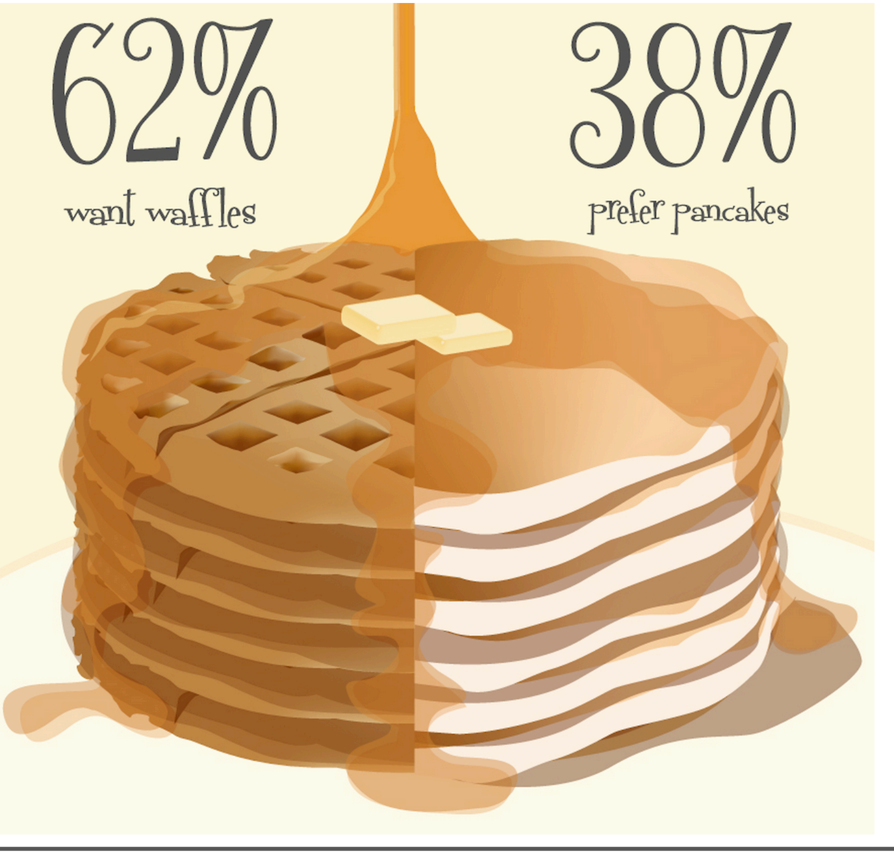


Fox news again: best pie chart ever

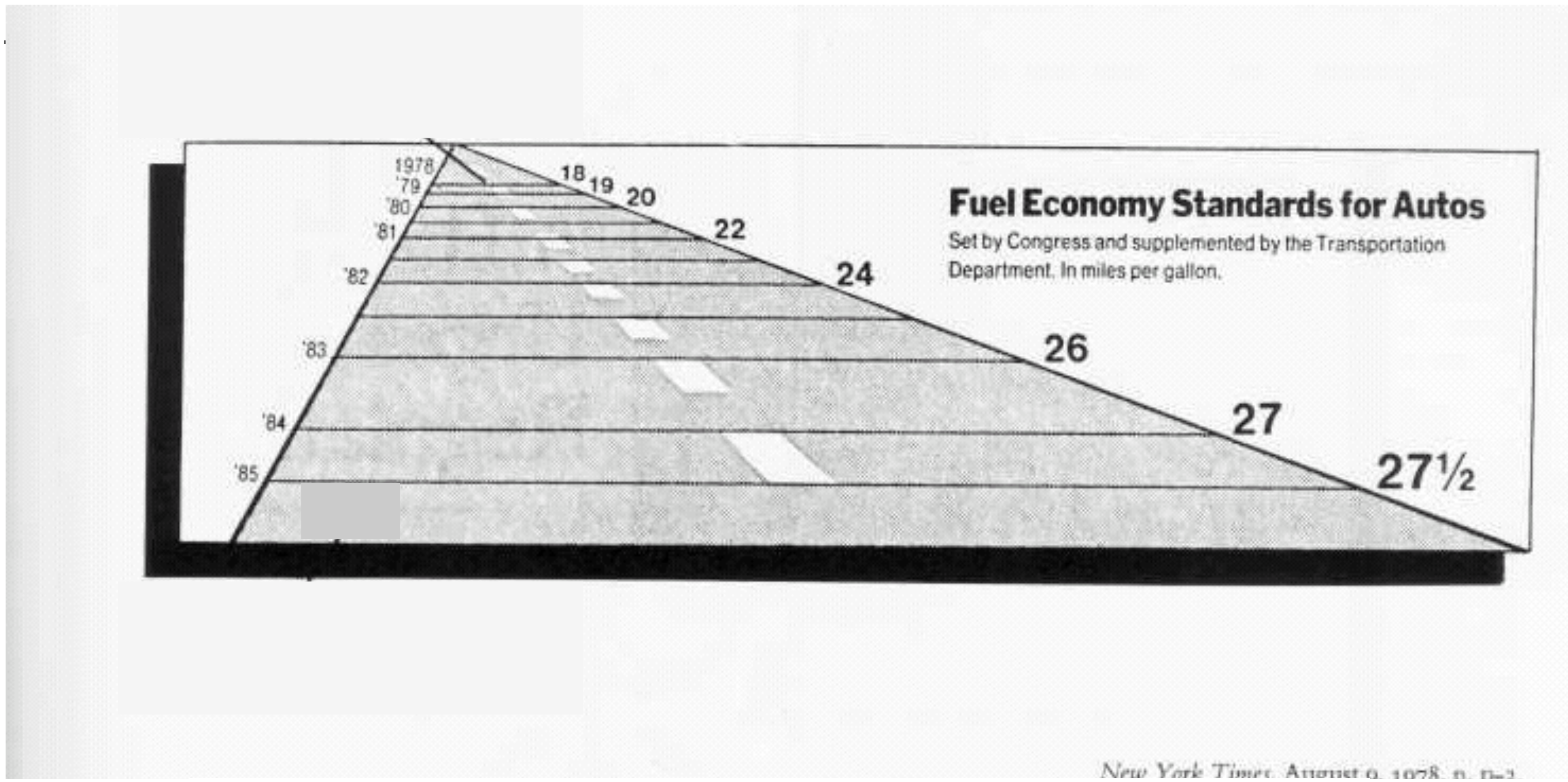


http://flowingdata.com/2009/11/26/fox-news-makes-the-best-pie-chart-ever/

??

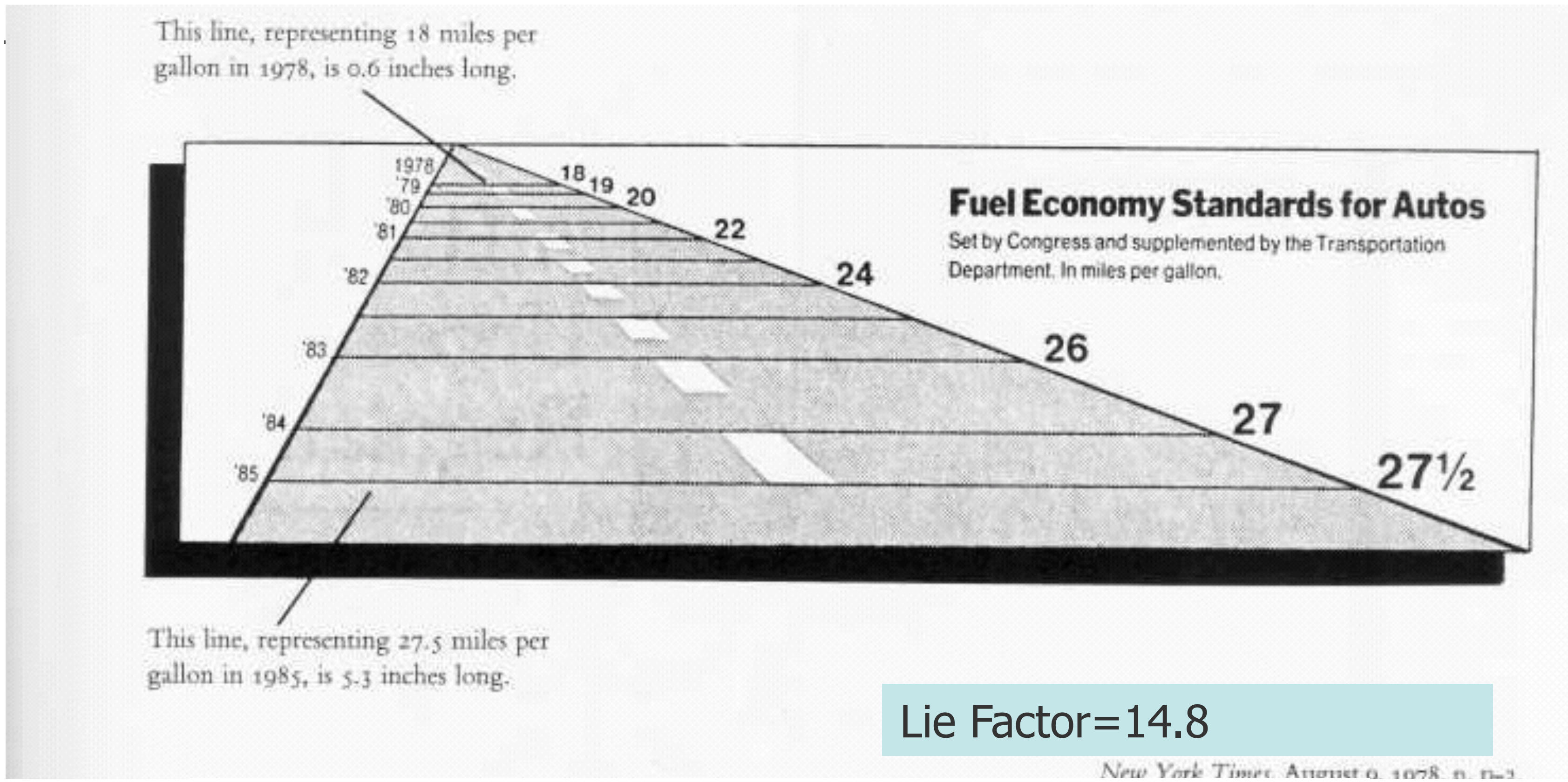

62%
want waffles
38%
prefer pancakes

What about this?



E.R. Tufte, "The Visual Display of Quantitative Information", 2nd edition

Lie factor



Lie Factor=14.8

E.R. Tufte, "The Visual Display of Quantitative Information", 2nd edition

Lie Factor

$$Lie\ Factor = \frac{size\ of\ effect\ shown\ in\ graphic}{size\ of\ effect\ in\ data} =$$

$$= \frac{\frac{(5.3 - 0.6)}{0.6}}{\frac{(27.5 - 18.0)}{18}} = \frac{7.833}{0.528} = 14.8$$

Tufte requirement: 0.95<Lie Factor<1.05

(E.R. Tufte, "The Visual Display of Quantitative Information", 2nd edition)

Tufte's Principles of Graphical Excellence

- Give the viewer
  - the greatest number of ideas
  - in the shortest time
  - with the least ink in the smallest space.
- Tell the truth about the data!

E.R. Tufte, "The Visual Display of Quantitative Information", 2nd edition

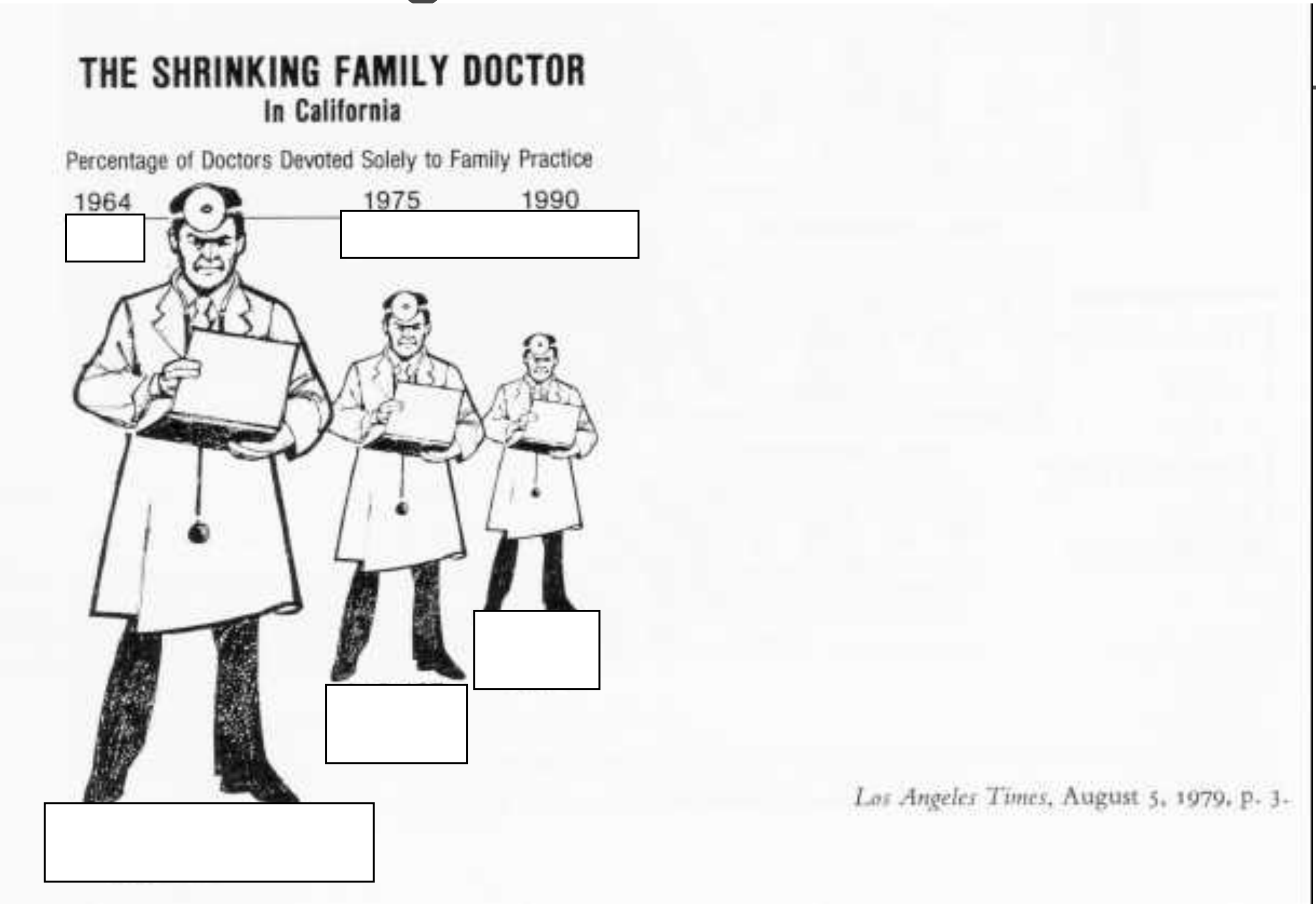
Two Principles

- The representation of numbers, as physically measured on the surface of the graphics, should be directly proportional to the numerical quantities represented

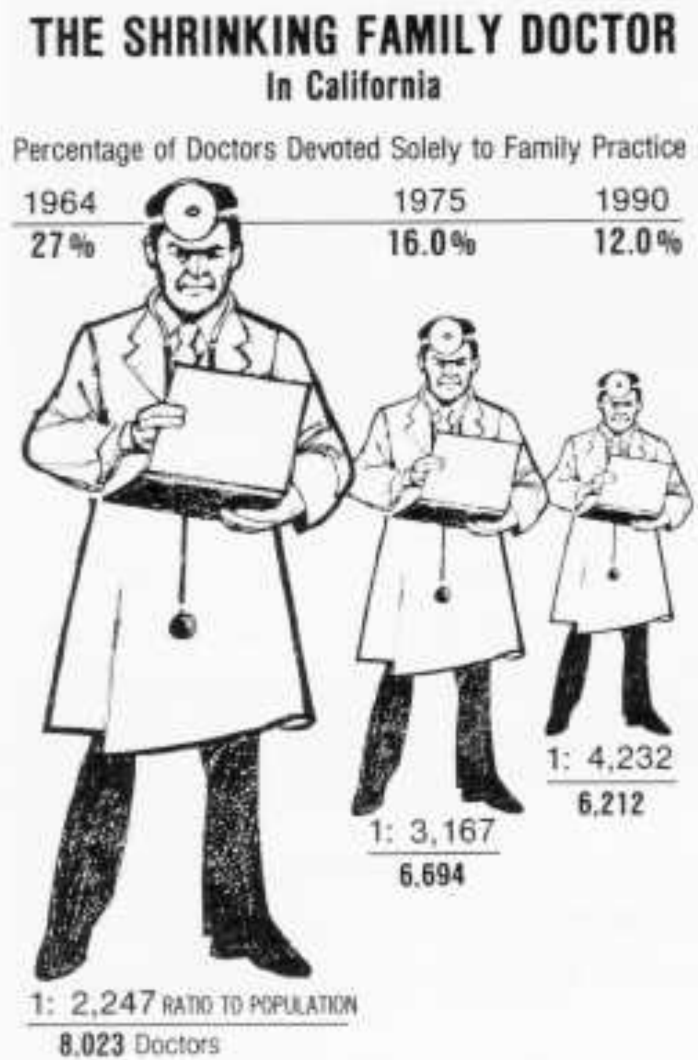
- Clear, detailed and thorough **labeling** should be used to defeat **distortion**

SFU

Size Encoding

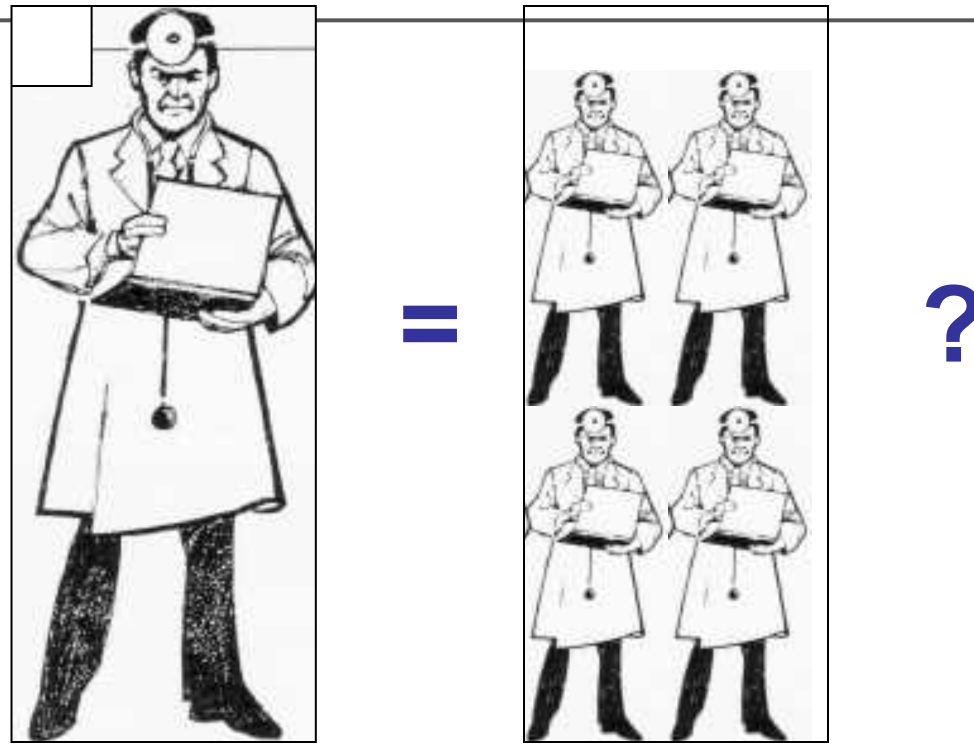


Size Encoding



Los Angeles Times, August 5, 1979, p. 3.

Size Encoding: height or area?



Height or Area

- Height = value
  Width = value
  Area = $value^2$

  or

- Area = value
  height*width = value
  height = width = $value^{0.5}$

Problem:
Using 2 dimensions to represent 1 dimension.

SFU

Height = value

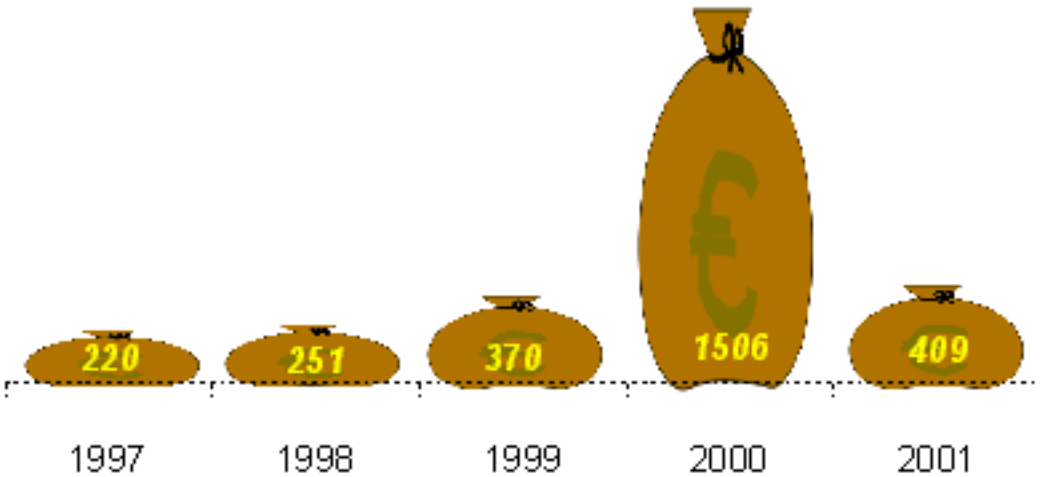


Chart 15: Soneras result drawn with moneybags as bars

With radius=value

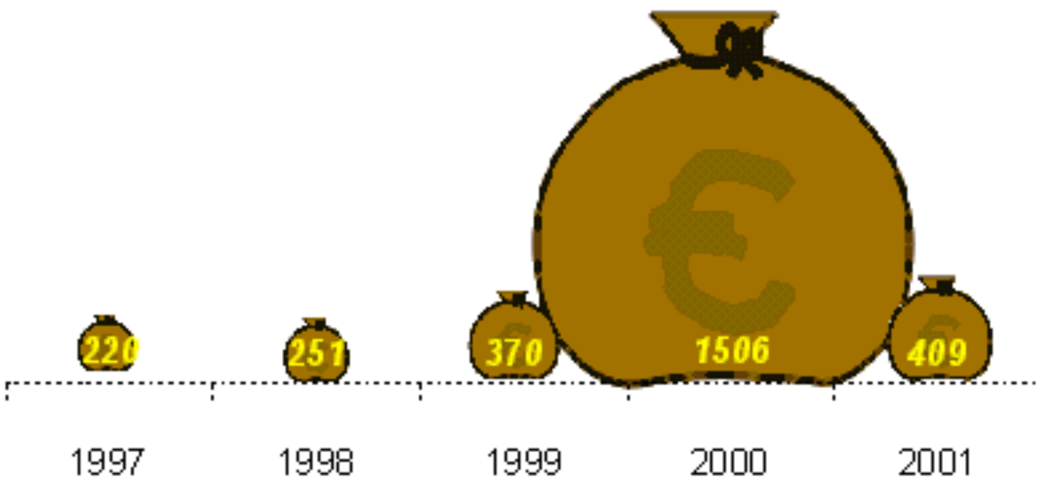


Chart 16: Moneybags that grow in both directions

Area = value

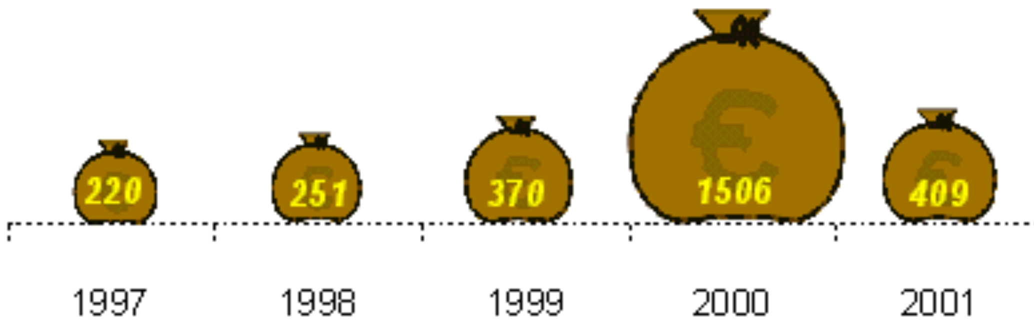


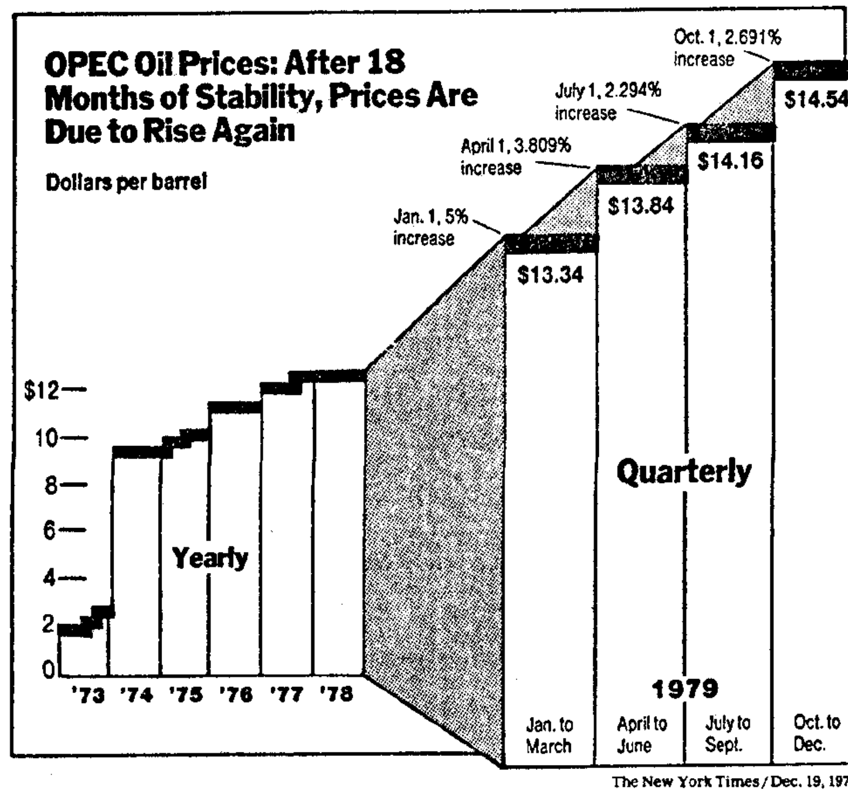
Chart 17: True 2-dimensional moneybags

Design and Data Variation

- Show data variation, not design variation

• 1973-1978: one vertical inch equals to $8.00. In 1979, One vertical inch equals $3-4

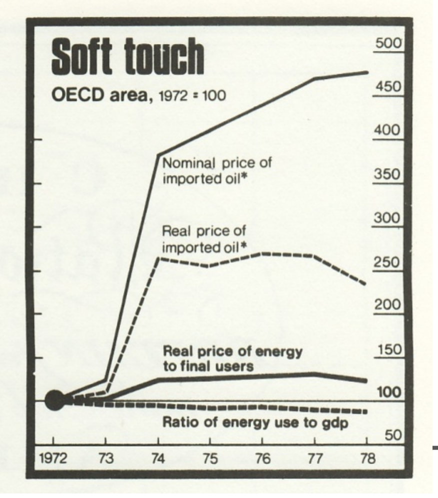
• 1973-1978: one horzontal inch equals 3.7 years, while 1979 equals 0.57 year

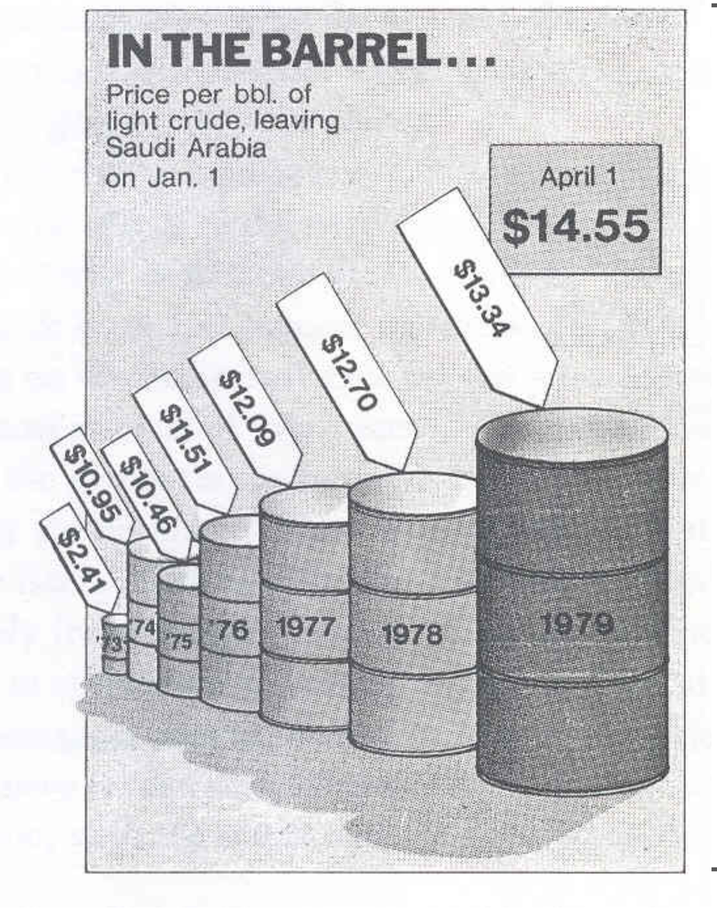


SFU

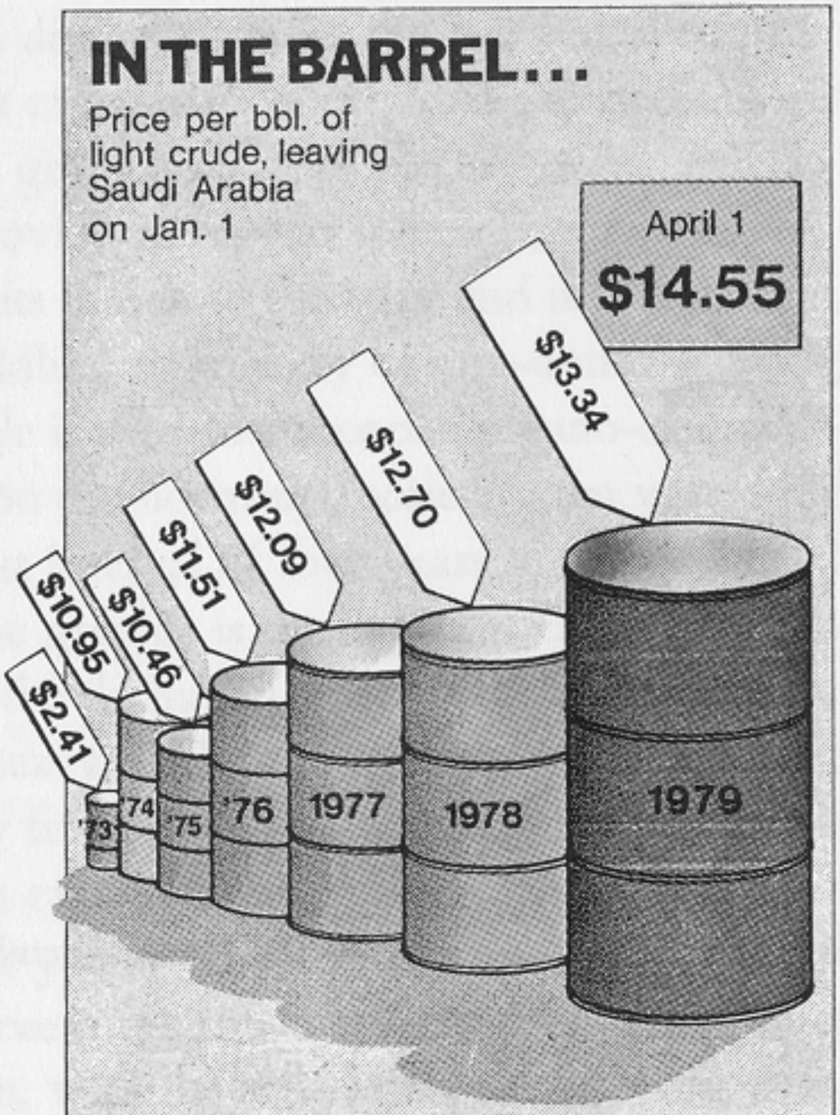
Example

- Lie factor: 9.5
- The price of oil is inflated so needs to be repaired.





SFU

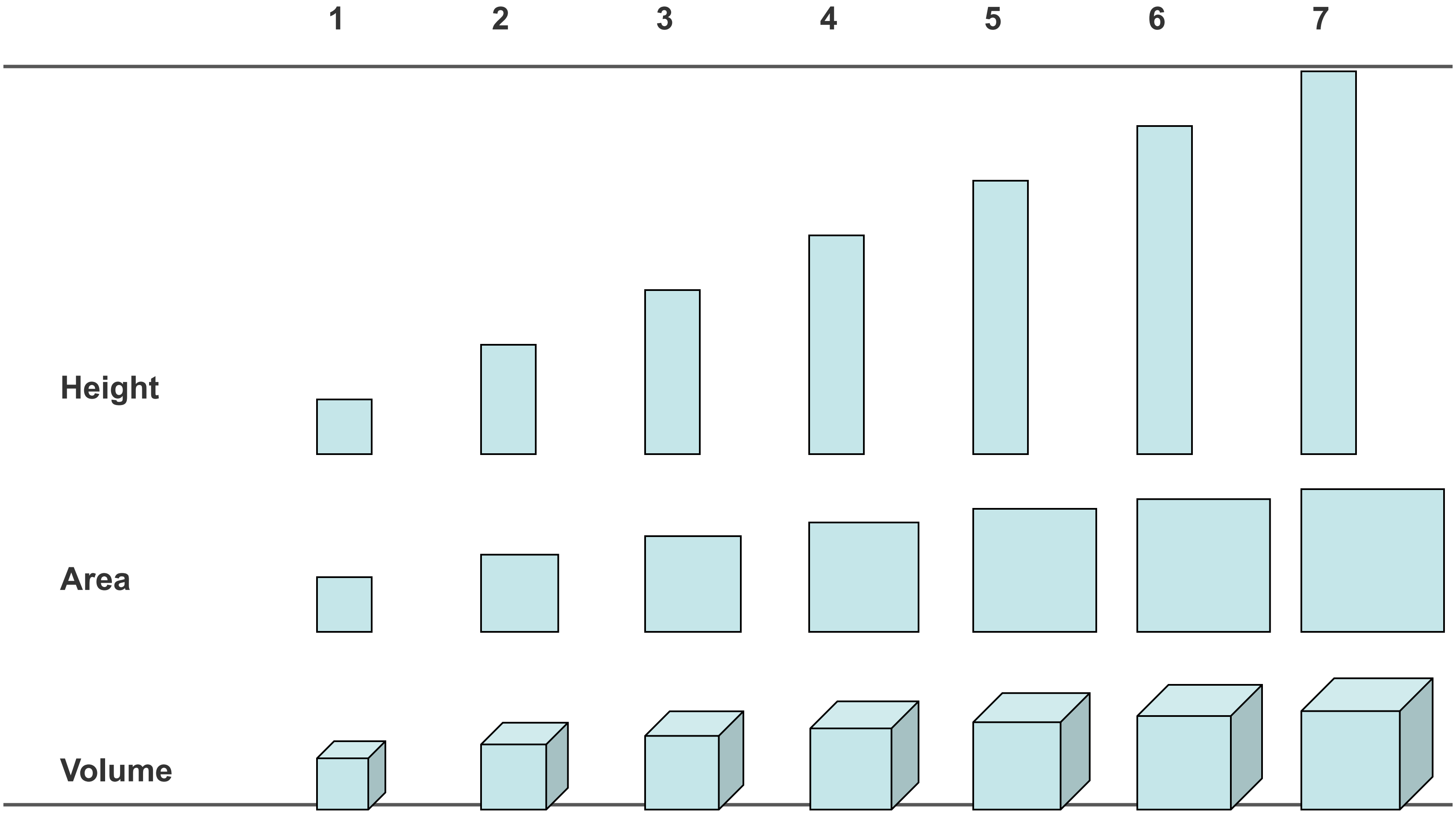


Time, April 9, 1979, p. 57.

Height?
Diameter?
Surface area?
Volume?

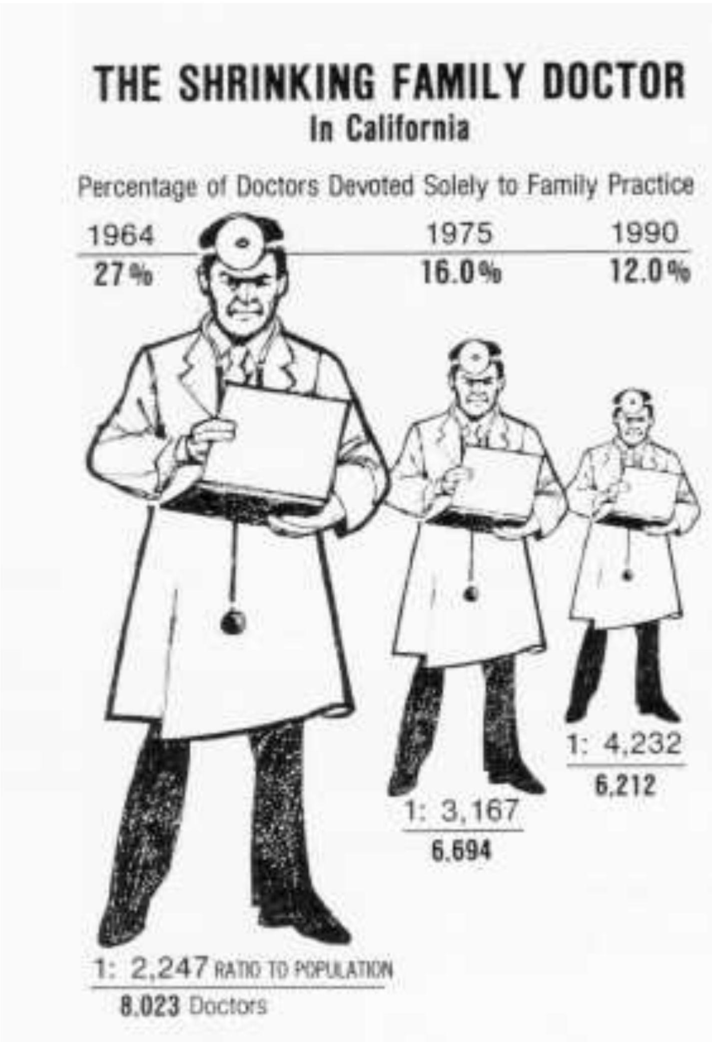
73 – 79 data difference = 5.5x
73 – 79 volume difference = 270x

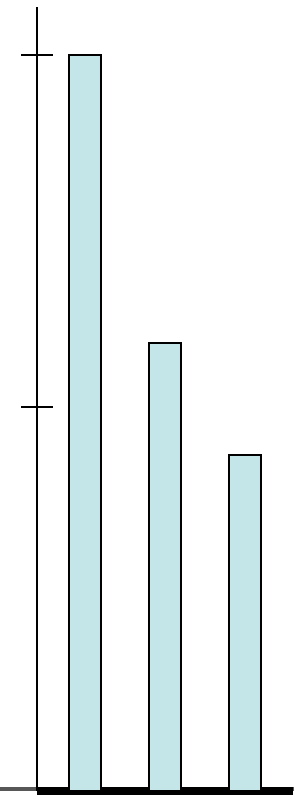
Problem with size encoding



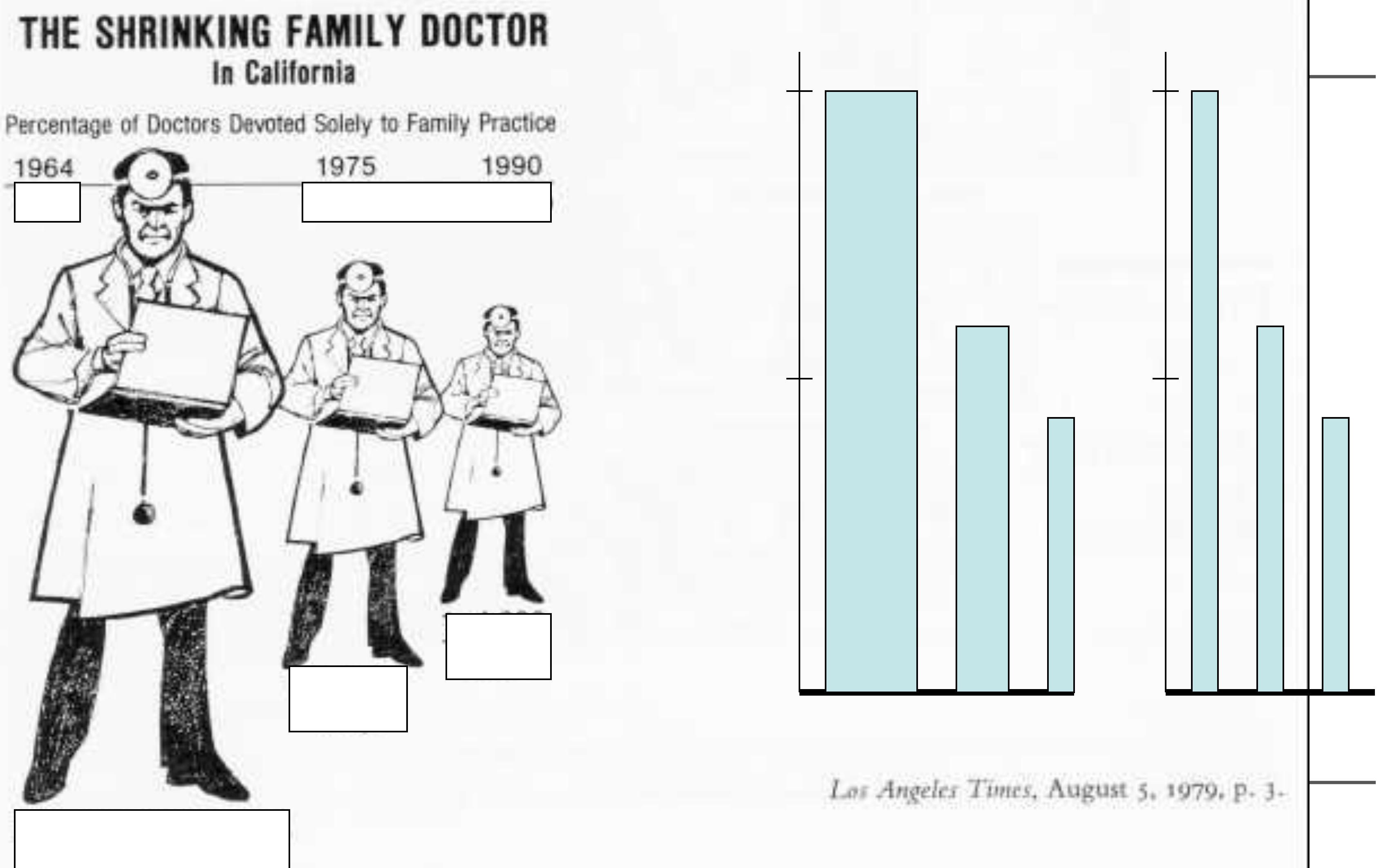
Height & width encoding

Solution: just use height





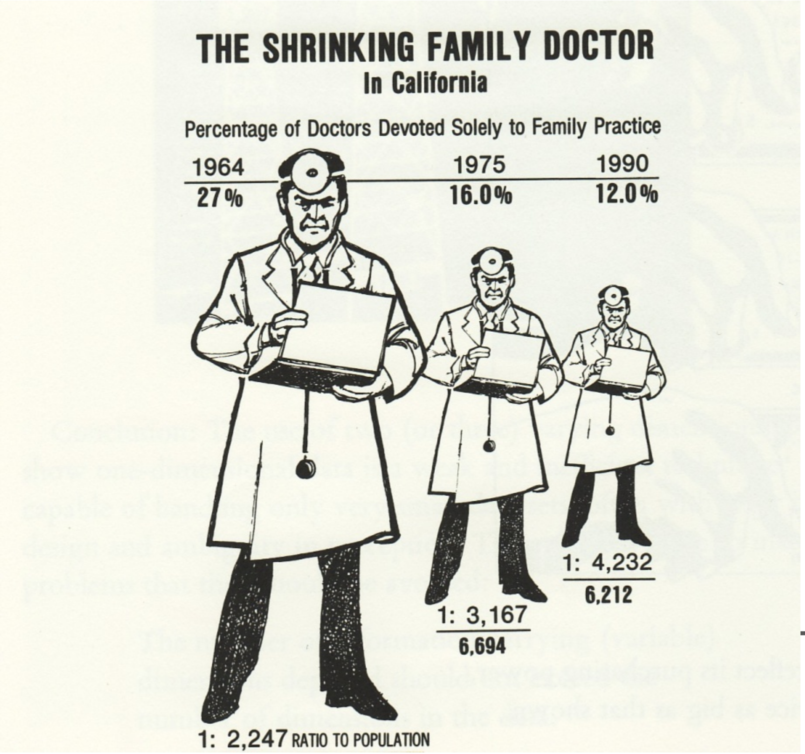
Size Encoding

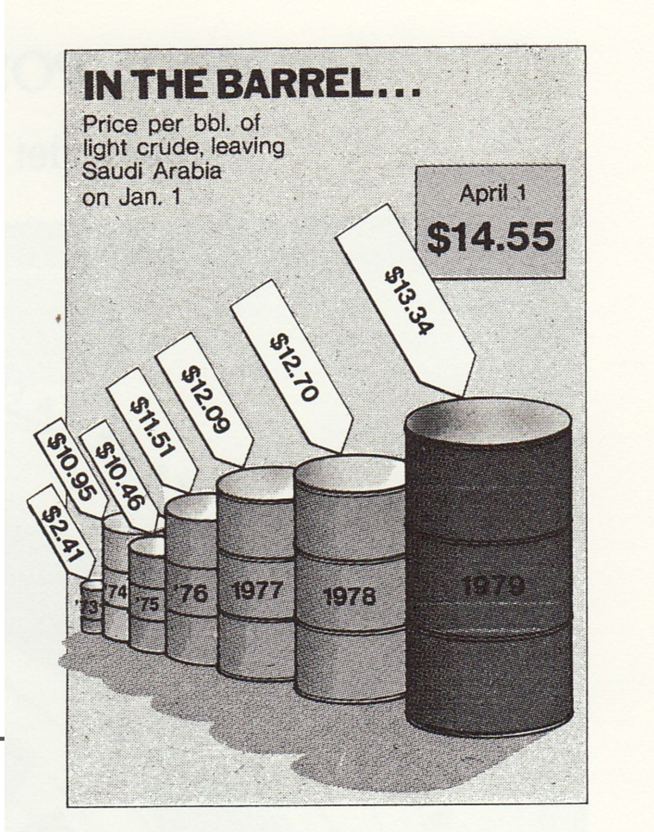


Los Angeles Times, August 5, 1979, p. 3.

Visual Area and Numerical Measure

- Tricking the reviewer with design variation is to use areas to show 1D data

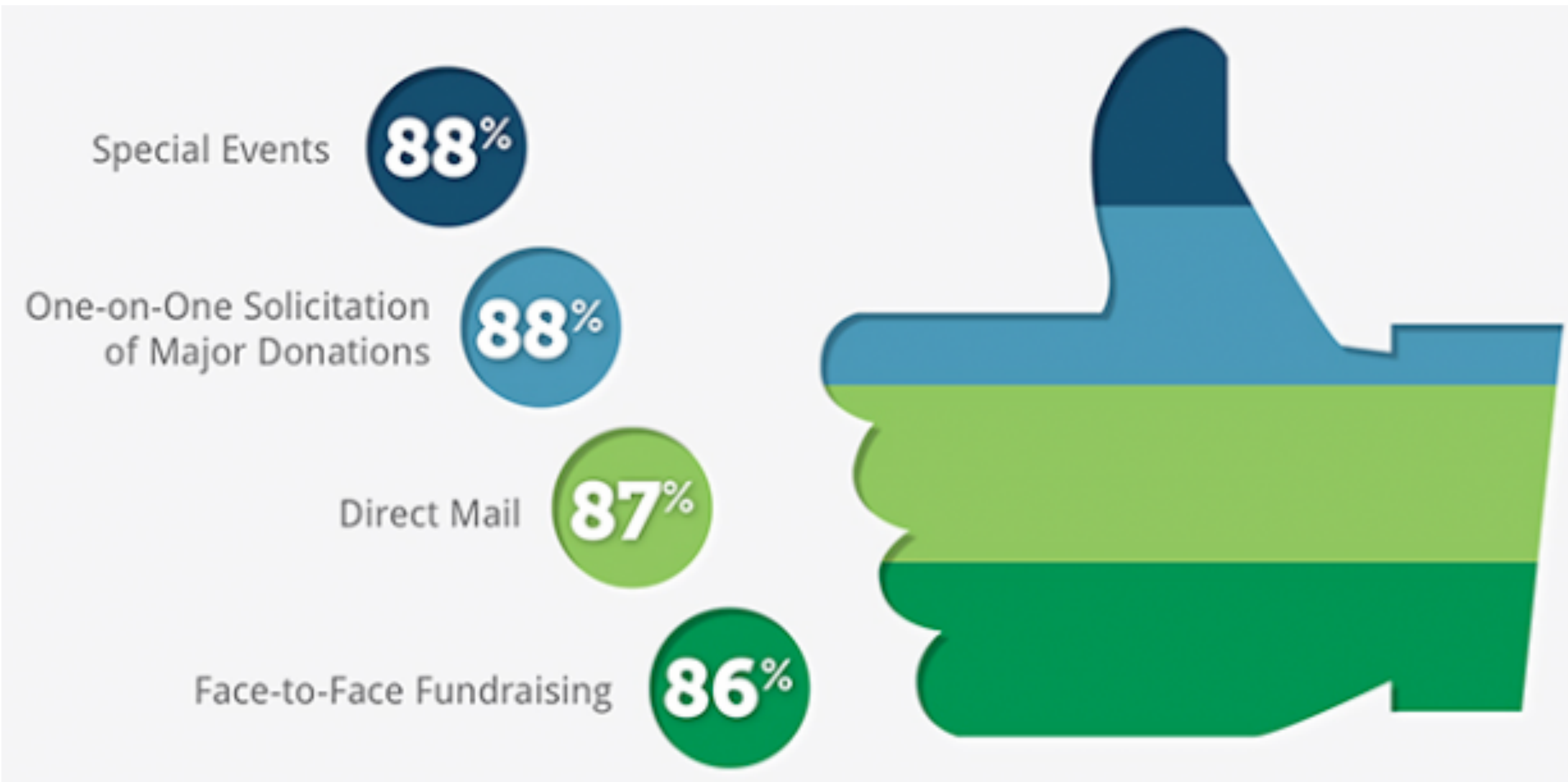




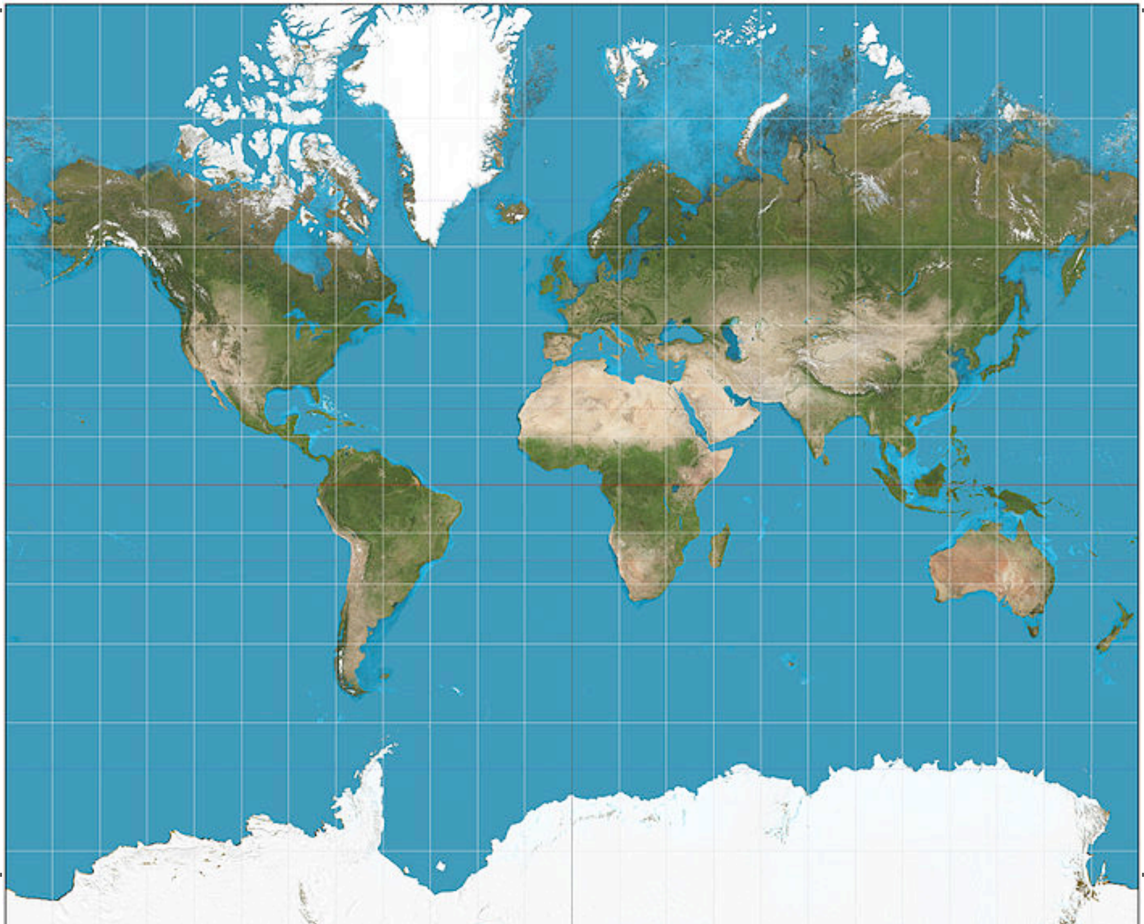
Lie factor: 2.8

SFU

What's being used here?



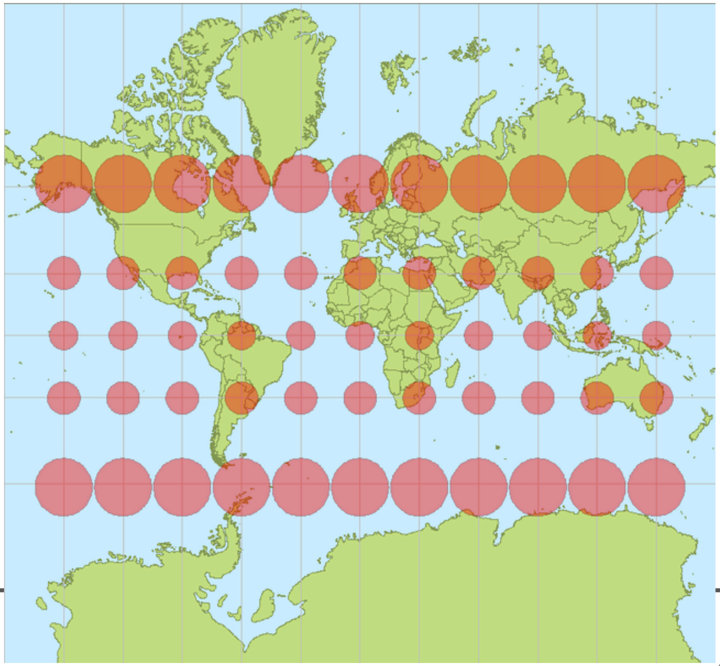
Something more familiar



Mercator projection

- Most maps use a version of the Mercator projection
- Designed for navigators to represent lines of constant course as straight segments

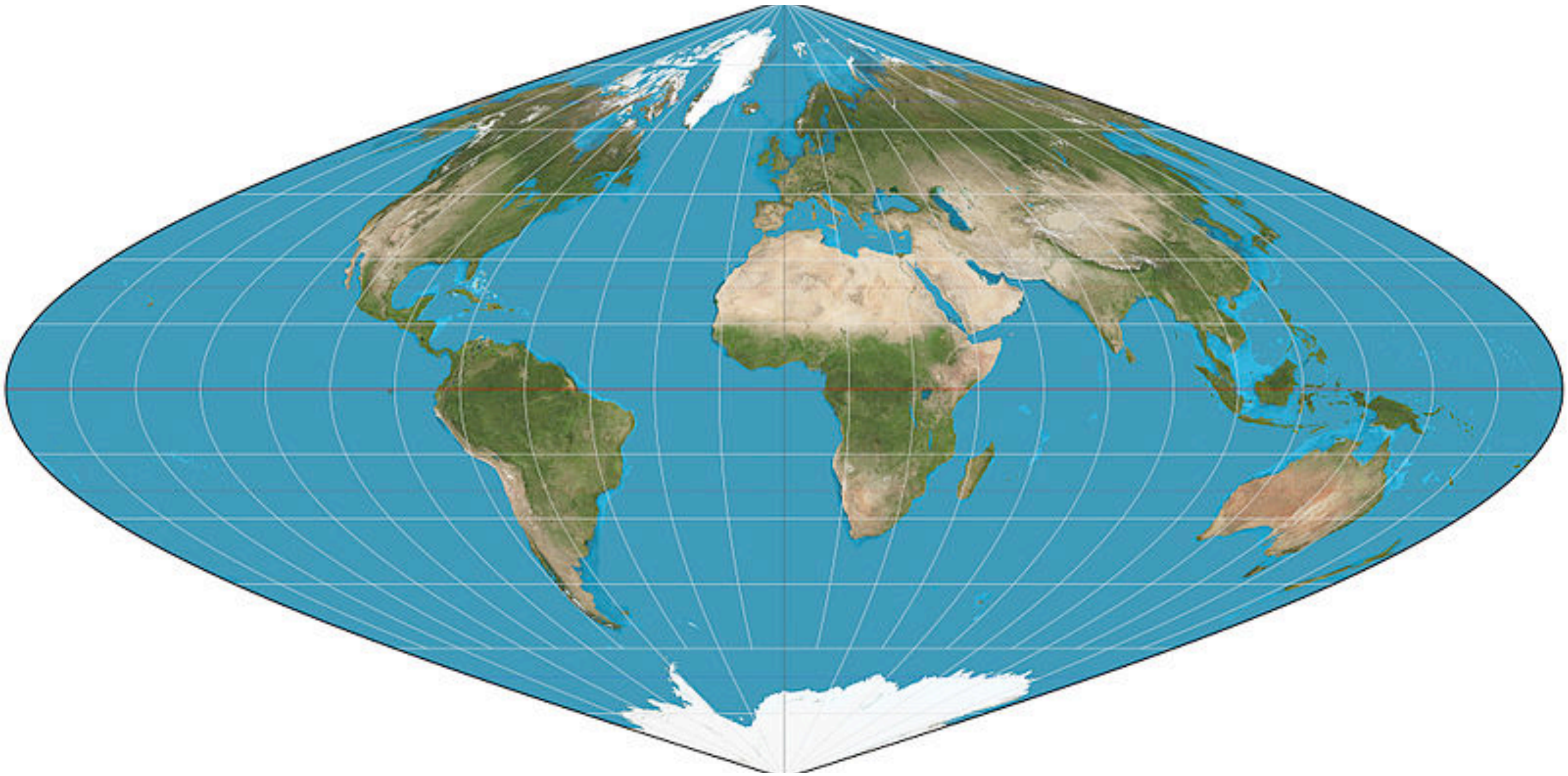
- So what's wrong?



Projections distort

- Greenland takes as much space on the map as Africa, when in reality Africa's area is 14 times greater and Greenland's is comparable to Algeria's alone.
- Alaska takes as much area on the map as Brazil, when Brazil's area is nearly five times that of Alaska.
- Finland appears with a greater north-south extent than India, although India's is greater.
- Antarctica appears as the biggest continent, although it is actually the fifth in terms of area.

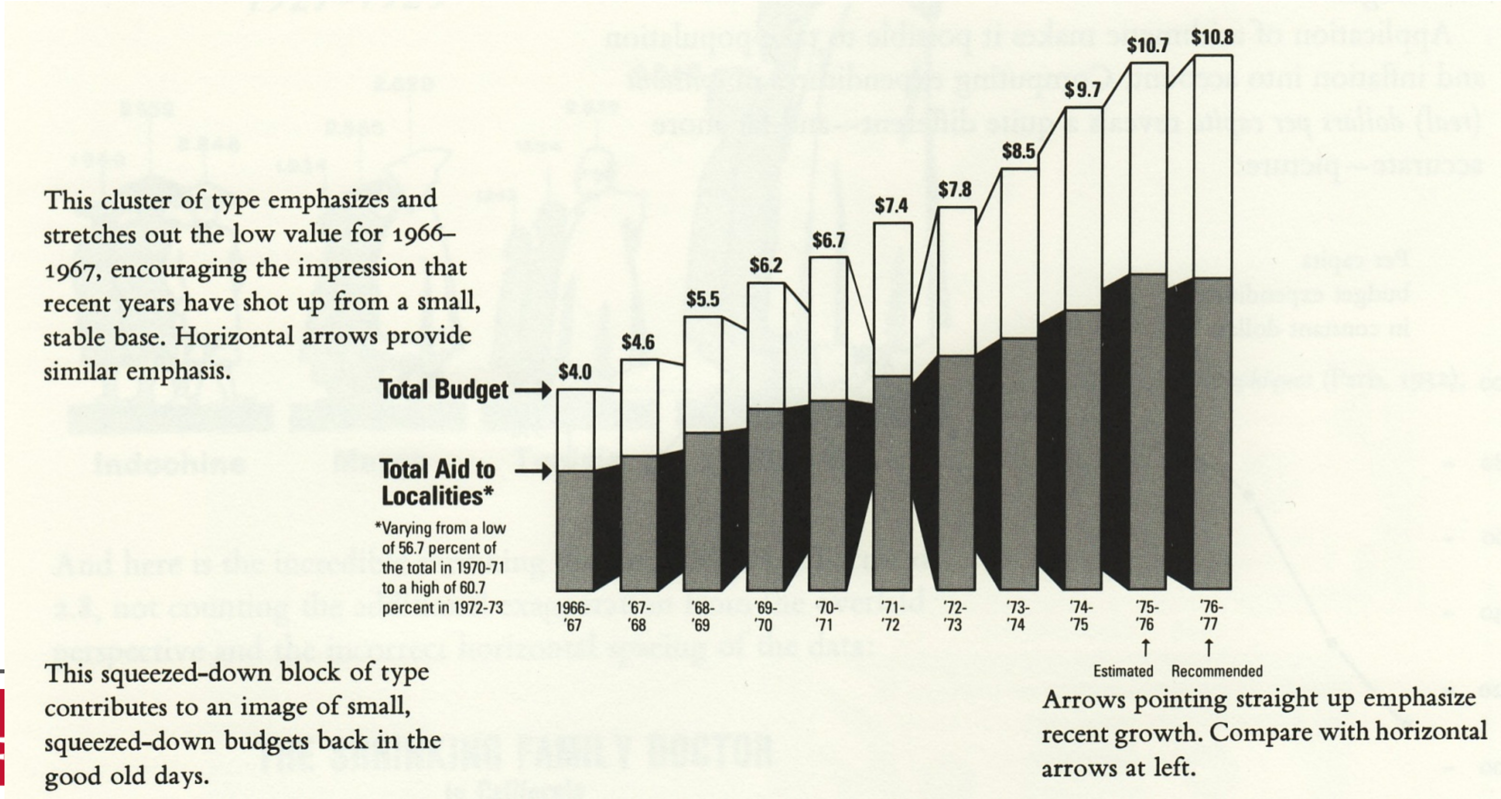
More accurate … but unfamiliar!



And more

Government Spending

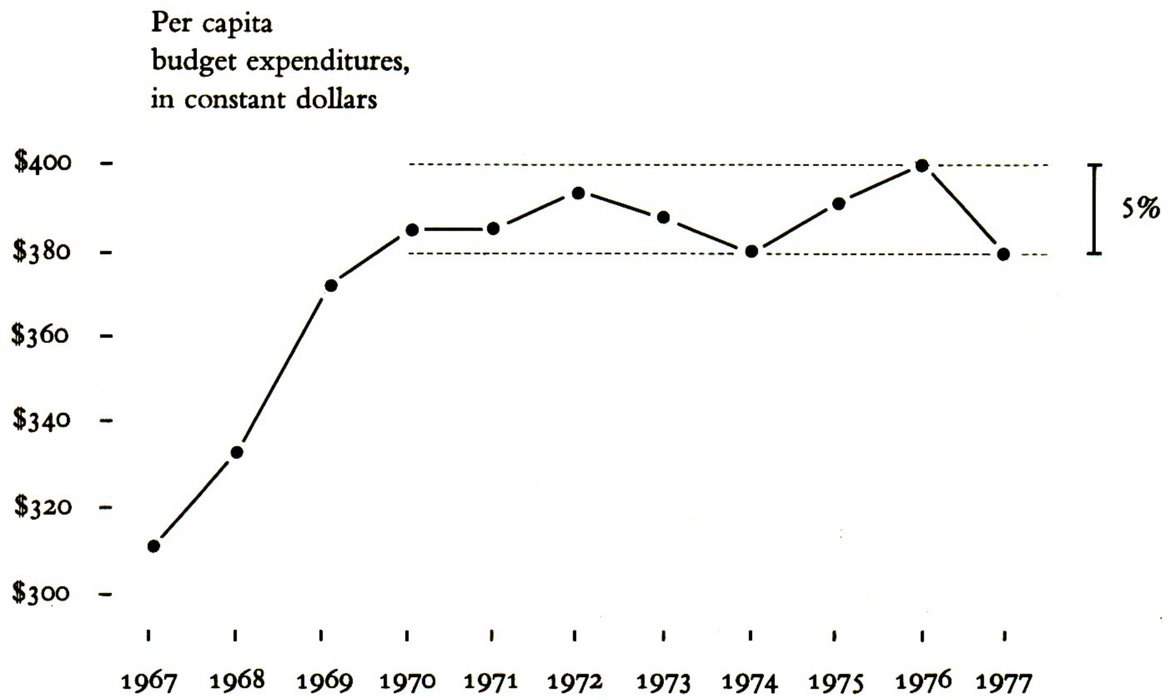
- Tricks to exaggerate the growth of spending



SF

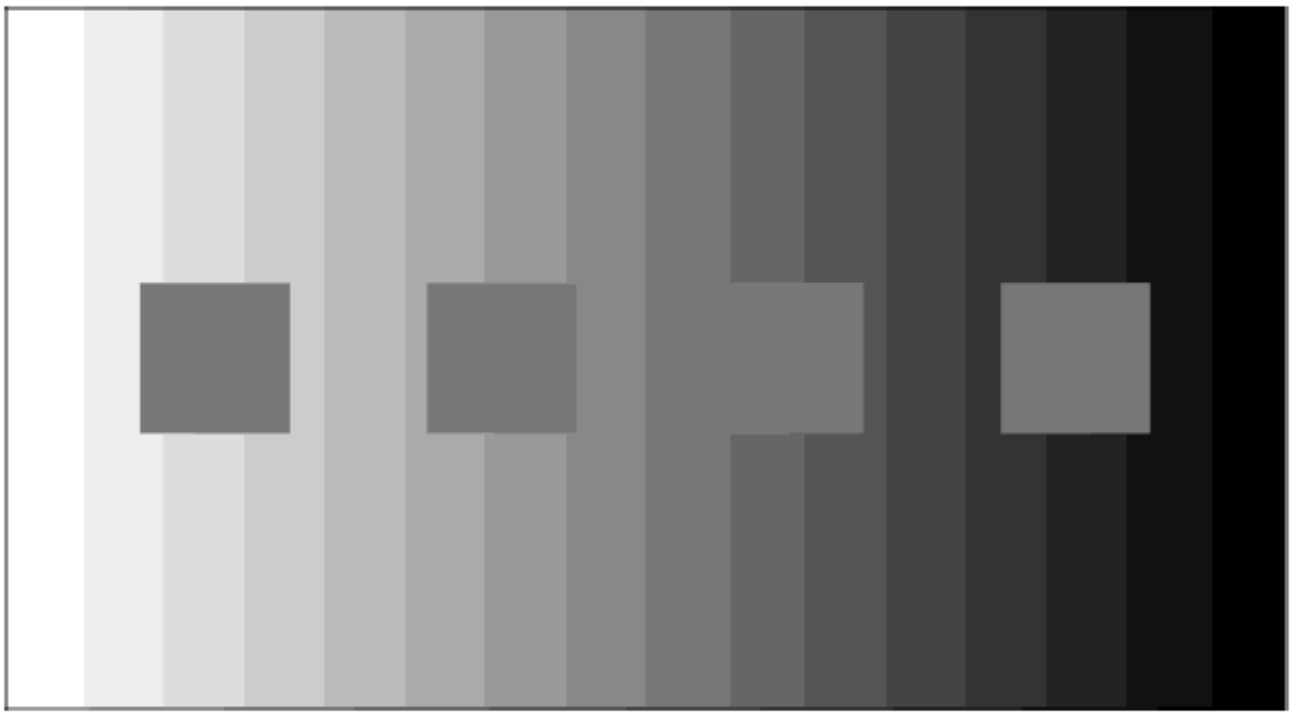
Real Government Spending

- Tricks to exaggerate the growth of spending



SFU

Remember the perceptual issues of colour



Contrast rules

- Rule 1: if you want objects of the same colour to look the same, make sure the (surrounding) background is consistent.
- Rule 2: if you want them to be easily see, use a background colour that contrasts sufficiently.

SFU

Be careful about implied meaning

- Certain representations impute meanings

- Cultural
- perceptual

Verbal-visual conflict: the Stroop effect