Visualization

COS 323

Slides based on CHI 2003 tutorial by Marti Hearst

What is Information Visualization?

"Transformation of the symbolic into the geometric"

(McCormick et al., 1987)

"... finding the artificial memory that best supports our natural means of perception."

(Bertin, 1983)

The depiction of information using spatial or graphical representations, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system.

Information Visualization

- Problem
  - Big datasets: How to understand them?
- Solution
  - Take better advantage of human perceptual system
  - Convert information into a graphical representation.
- Issues
  - How to convert abstract information into graphical form?
  - Do visualizations do a better job than other methods?

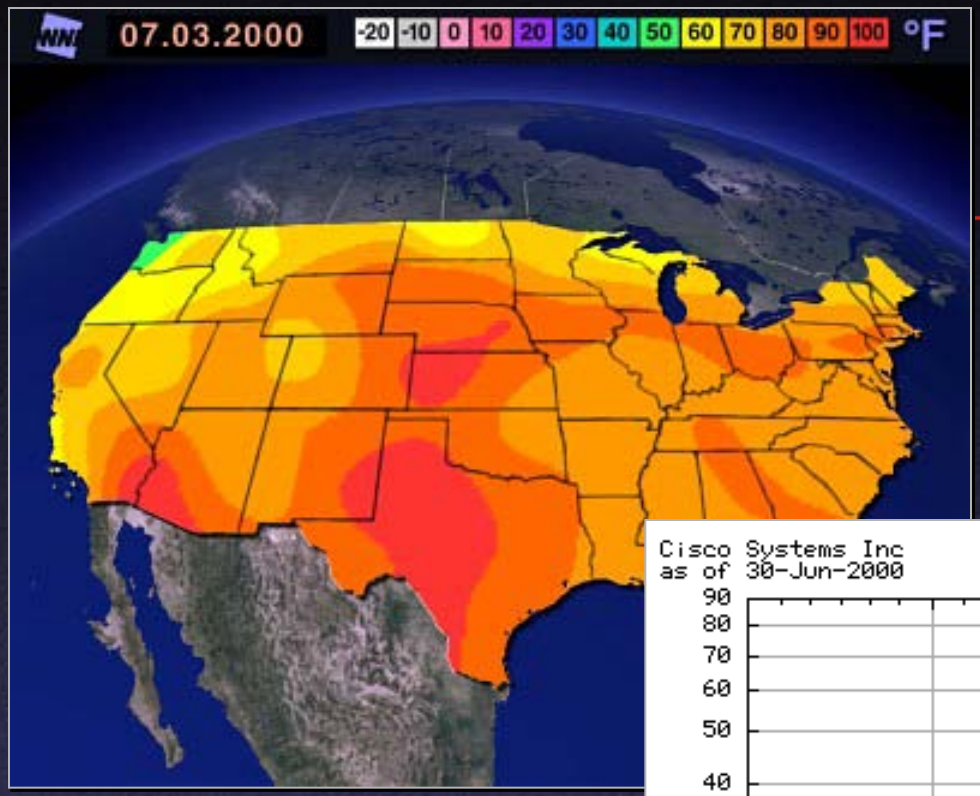
Goals of Information Visualization

- More specifically, visualization should:
  - Make large datasets coherent

    (Present huge amounts of information compactly)
  - Present information from various viewpoints
  - Present information at several levels of detail

    (from overviews to fine structure)
  - Support visual comparisons
  - Tell stories about the data

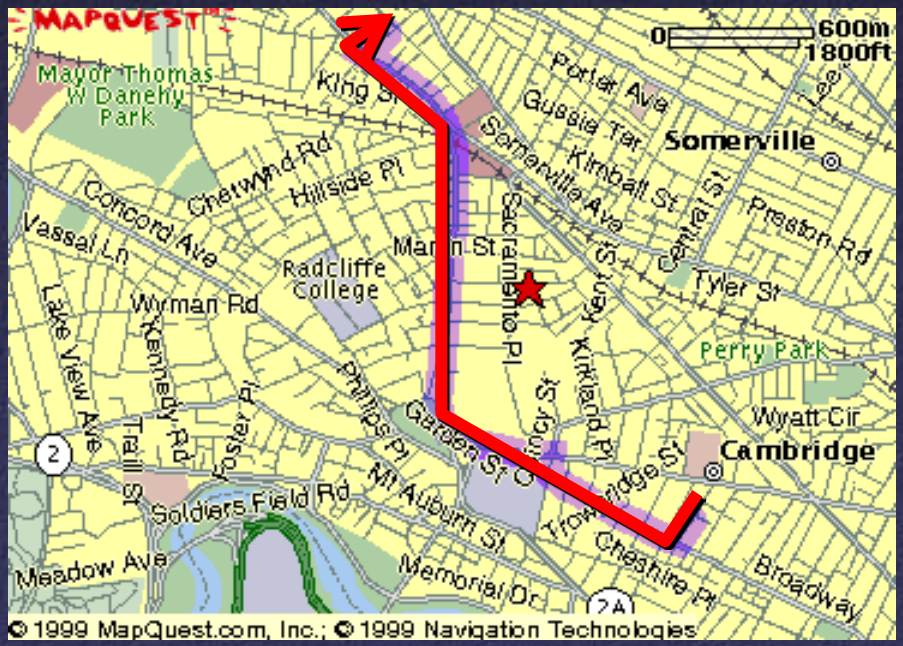
Visualization Success Stories



yahoo.com

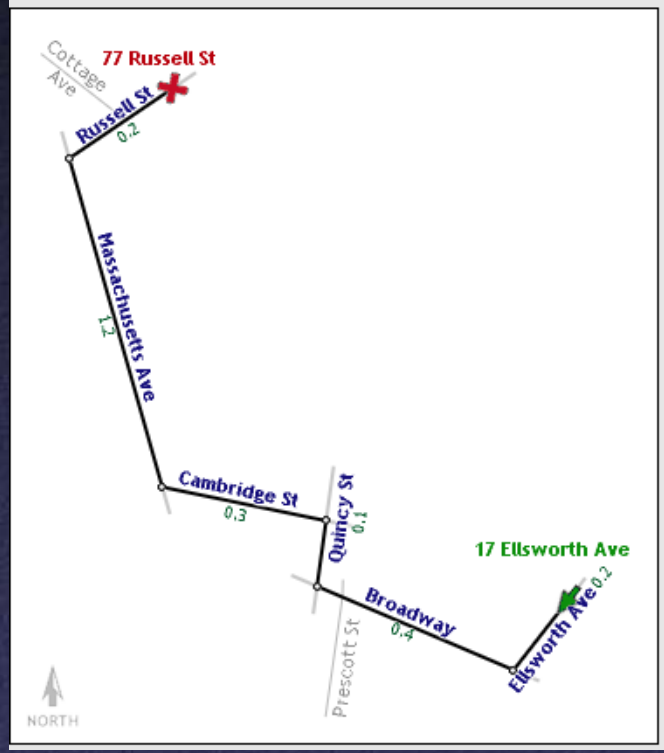
The Power of Visualization

1. Start out going Southwest on ELLSWORTH AVE
   Towards BROADWAY by turning right.
2: Turn RIGHT onto BROADWAY.
3. Turn RIGHT onto QUINCY ST.
4. Turn LEFT onto CAMBRIDGE ST.
5. Turn SLIGHT RIGHT onto MASSACHUSETTS AVE.
6. Turn RIGHT onto RUSSELL ST.



The Power of Visualization

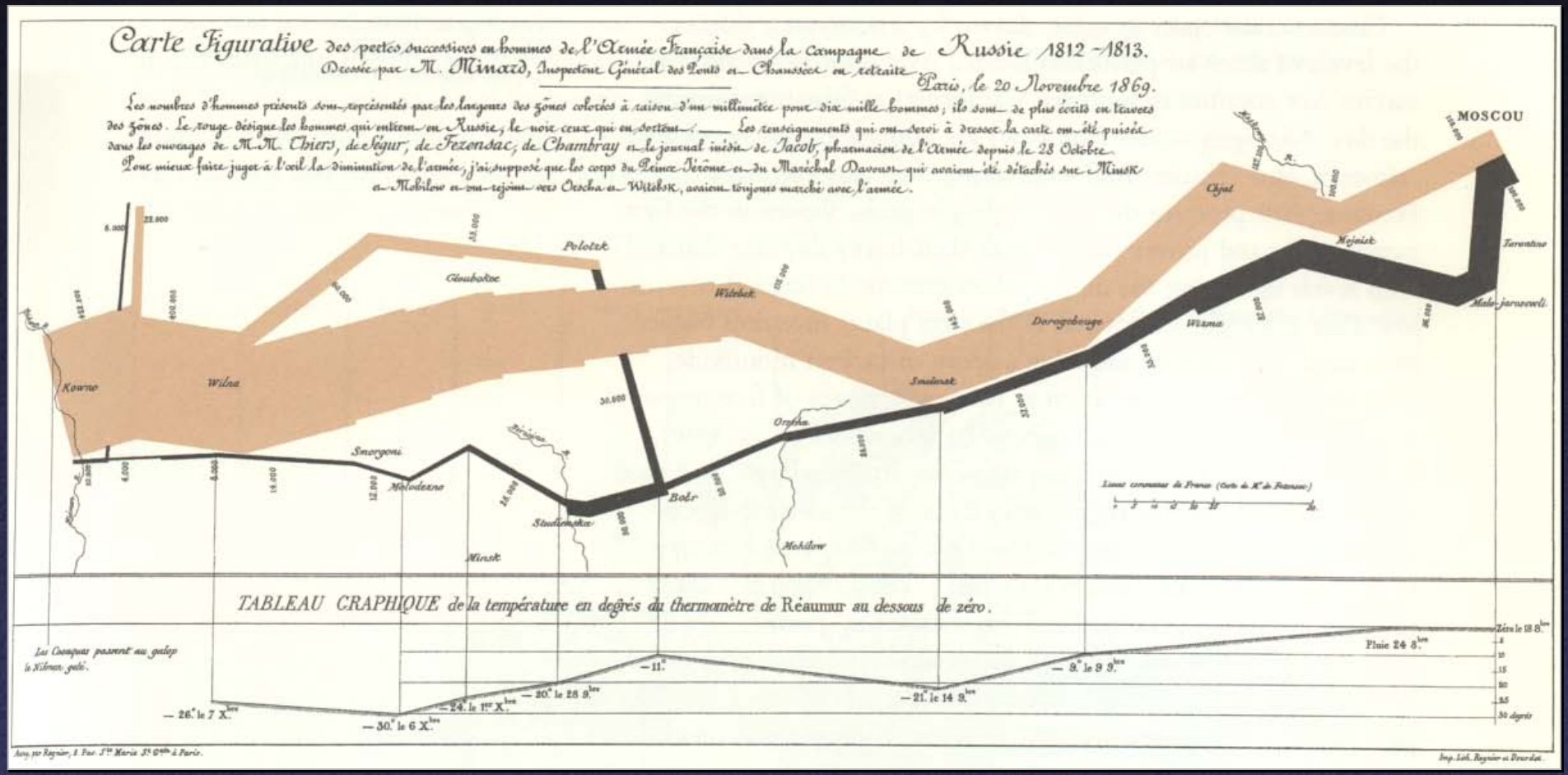
The estimated travel time is 5 minutes for 2.16 miles of travel, total of 6 steps.



	Directions	Elapsed Distance
1	Begin at **17 Ellsworth Ave** on **Ellsworth Ave** and go Southwest for 500 feet	0.1
2	Turn right on **Broadway** and go Northwest for 0.4 miles	0.5
3	Turn right on **Quincy St** and go North for 200 feet	0.5
4	Turn left on **Cambridge St** and go West for 0.3 miles	0.8
5	Bear right on **Massachusetts Ave,Mass Ave,RT-2A** and go North for 1.2 miles	2.0
6	Turn right on **Russell St** and go Northeast for 1000 feet to **77 Russell St**	2.2

Maneesh Agrawala – http://graphics.stanford.edu/~maneesh/

Napoleon's 1812 March by Charles Joseph Minard

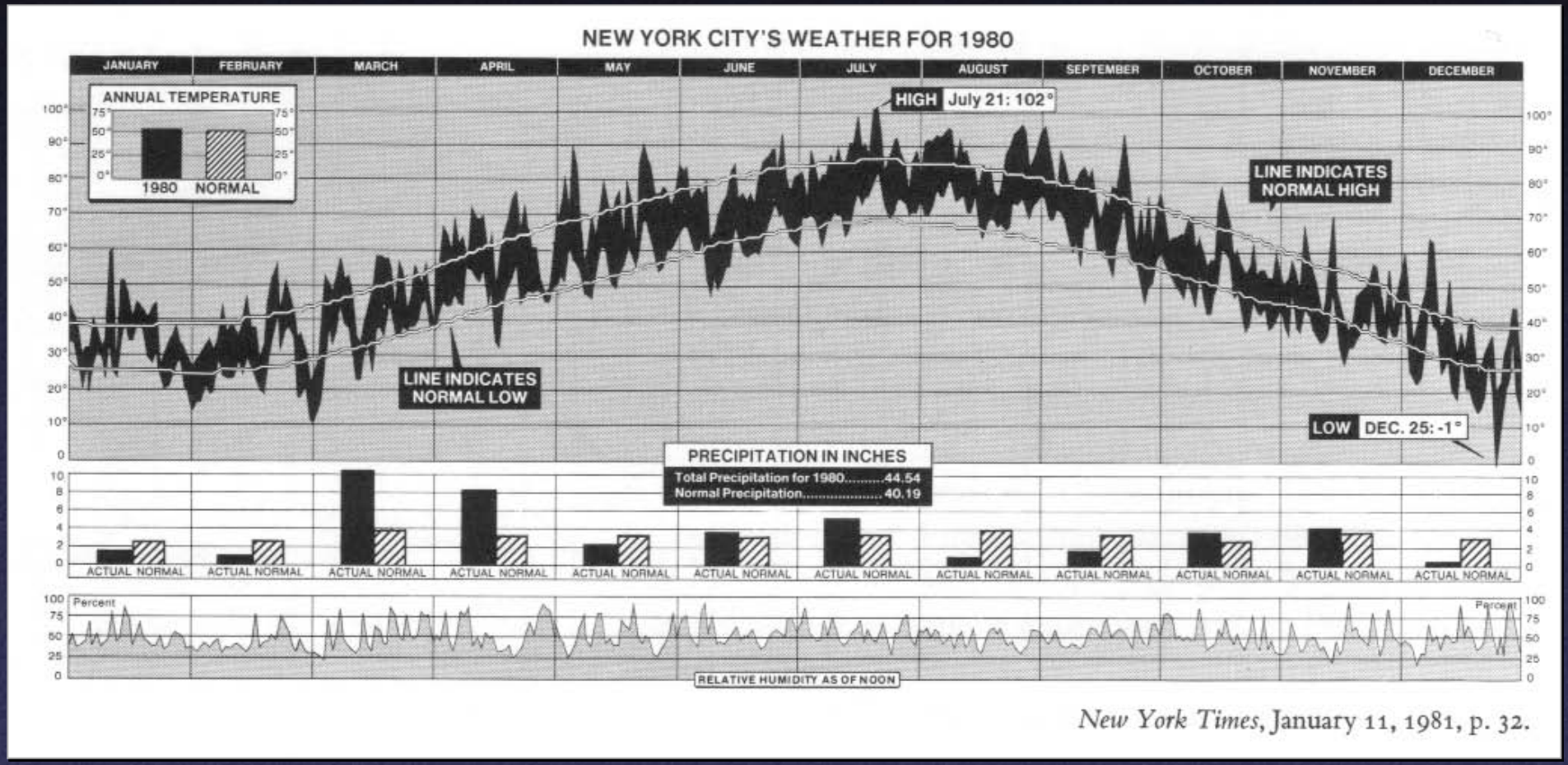


Variables shown:

- size of army
- direction
- latitude
- longitude
- temperature
- date

[Tufte]

NYC Weather



New York Times, January 11, 1981, p. 32.

2220 numbers

[Tufte]

Visualization Success Story

Mystery: what is causing a cholera epidemic in London in 1854?

Visualization Success Story

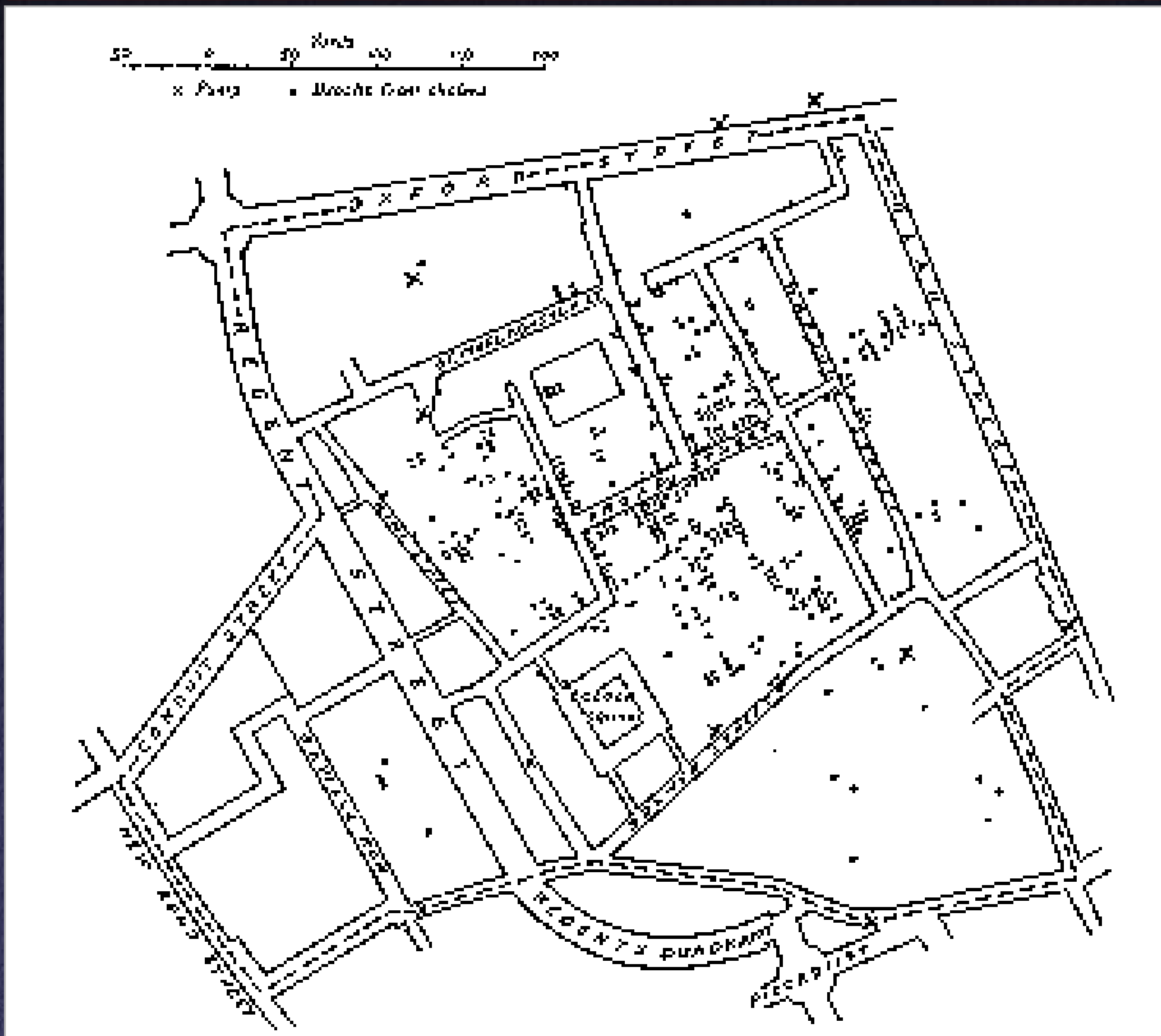
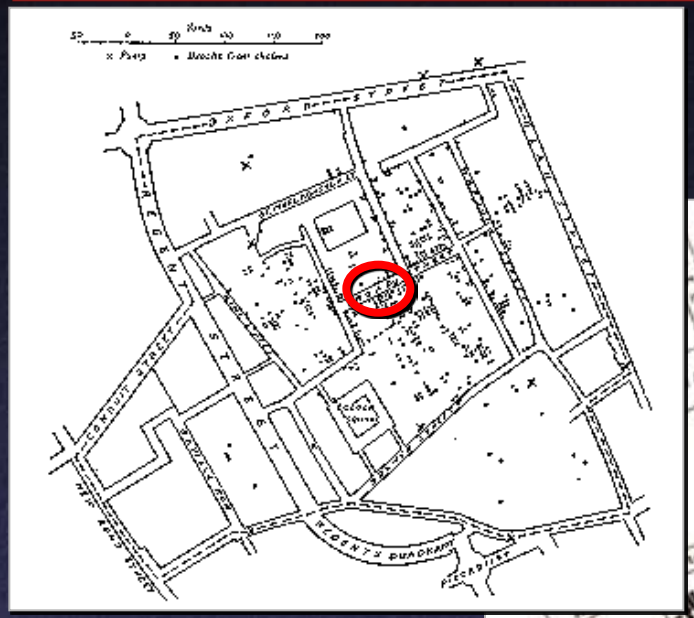


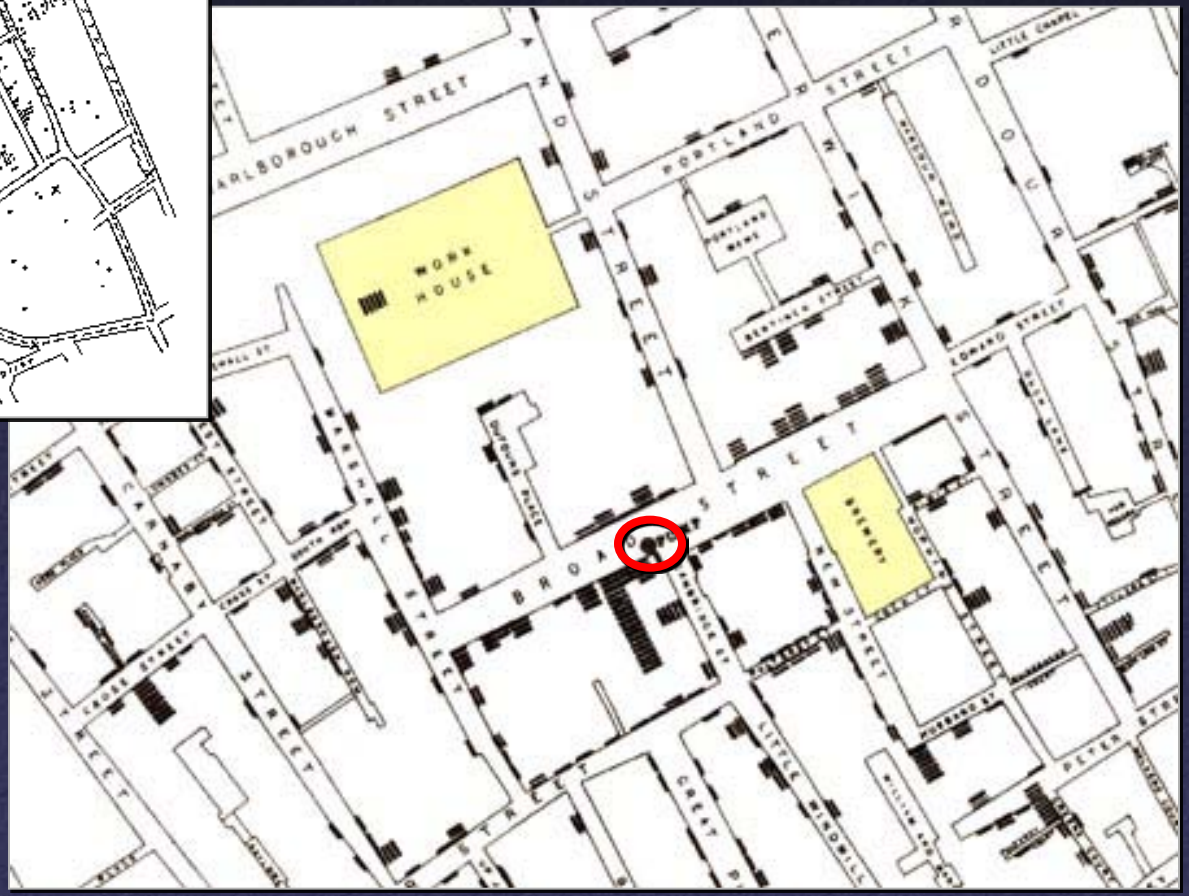
Illustration of
John Snow's
deduction that a
cholera epidemic
was caused by
a bad water pump,
circa 1854.

Horizontal lines
indicate
locations of deaths.

[Tufte]

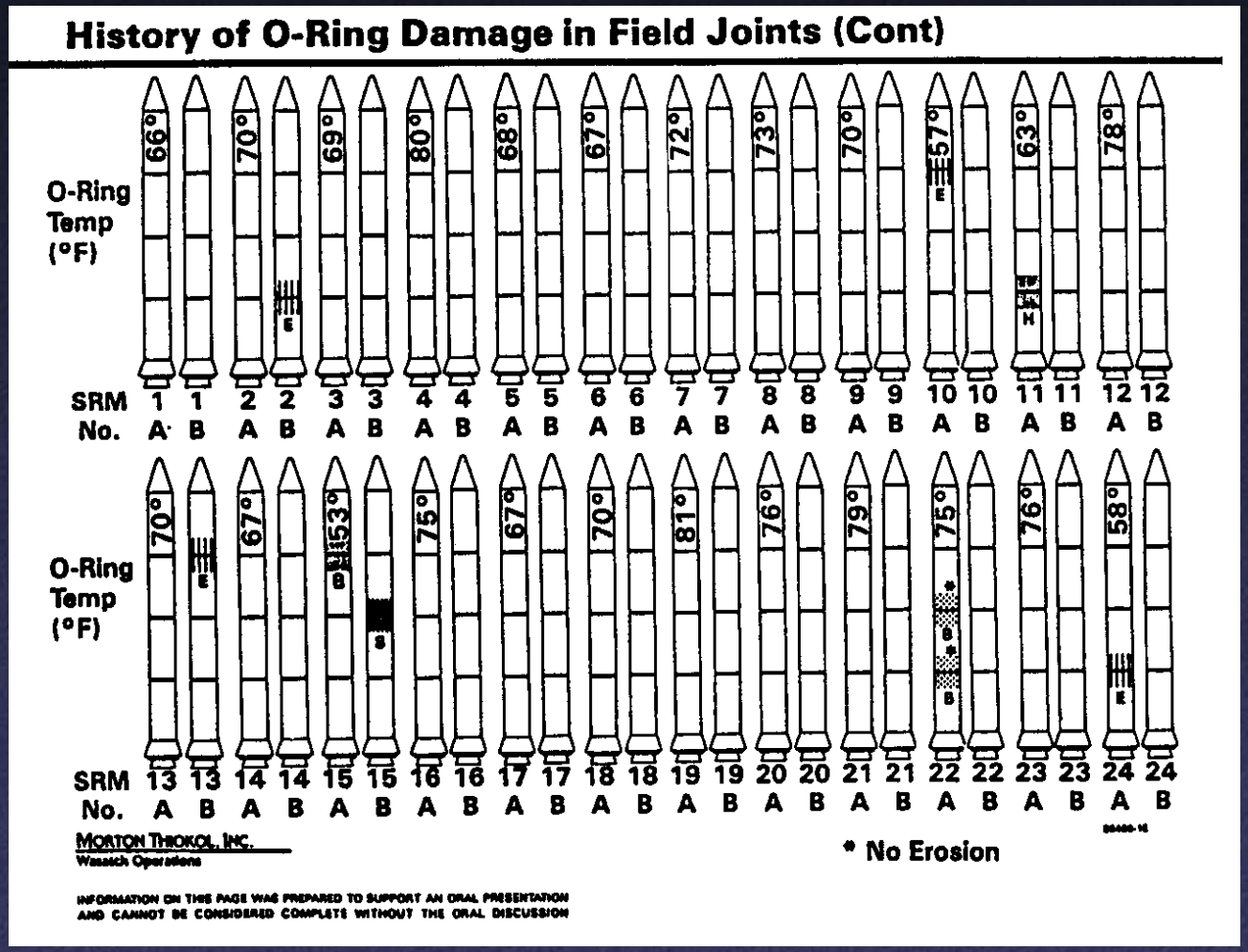
Visualization Success Story





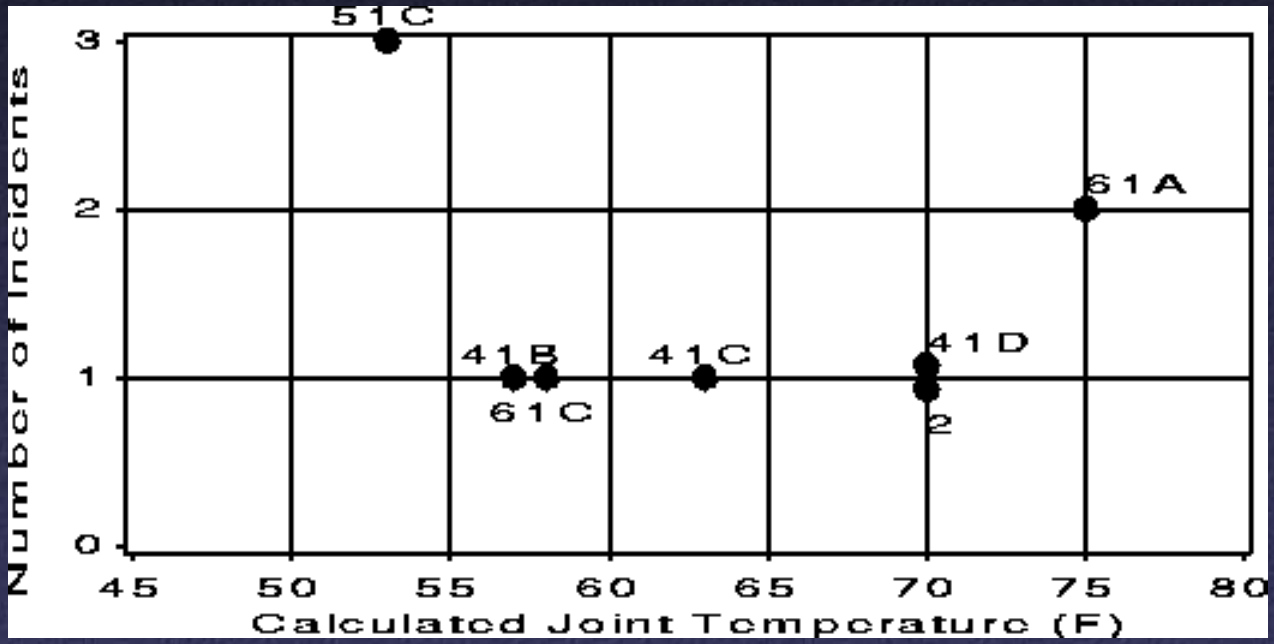
[Tufte]

Visualization Failure



Visualization Failure

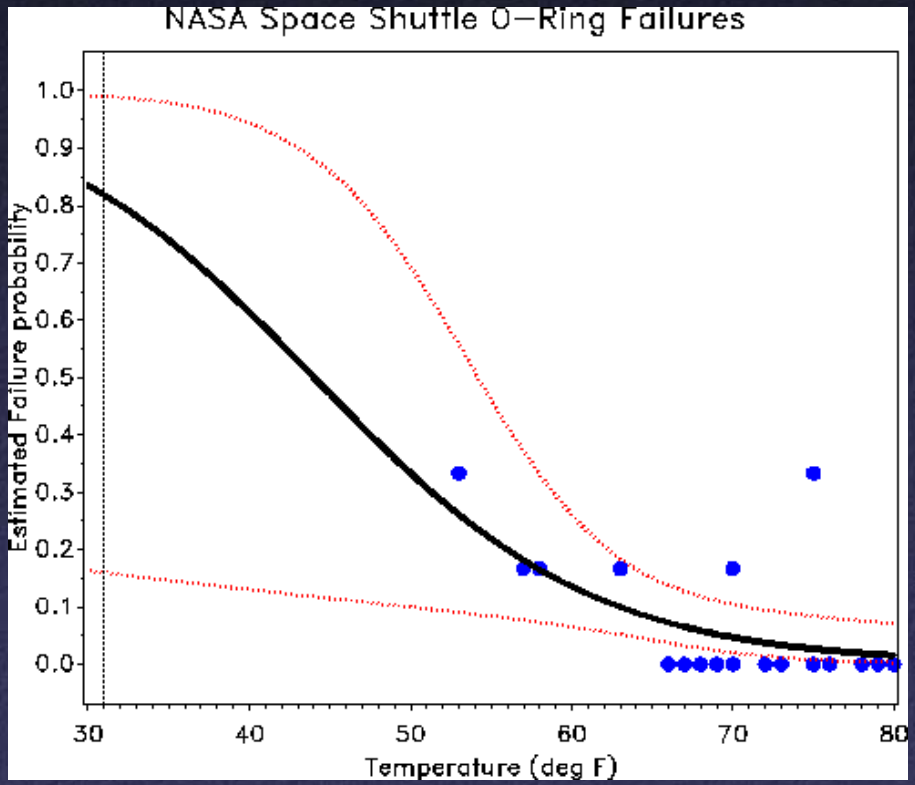
The visualization they made…



http://www.math.yorku.ca/SCS/Gallery/

Visualization Failure

The one they should have made…



http://www.math.yorku.ca/SCS/Gallery/

Why Visualization?

- Use the eye for pattern recognition; people are good at
  - scanning
  - recognizing
  - remembering images
- Graphical elements facilitate comparisons via
  - length
  - shape
  - orientation
  - texture
- Animation shows changes across time
- Color helps make distinctions
- Aesthetics help maintain interest

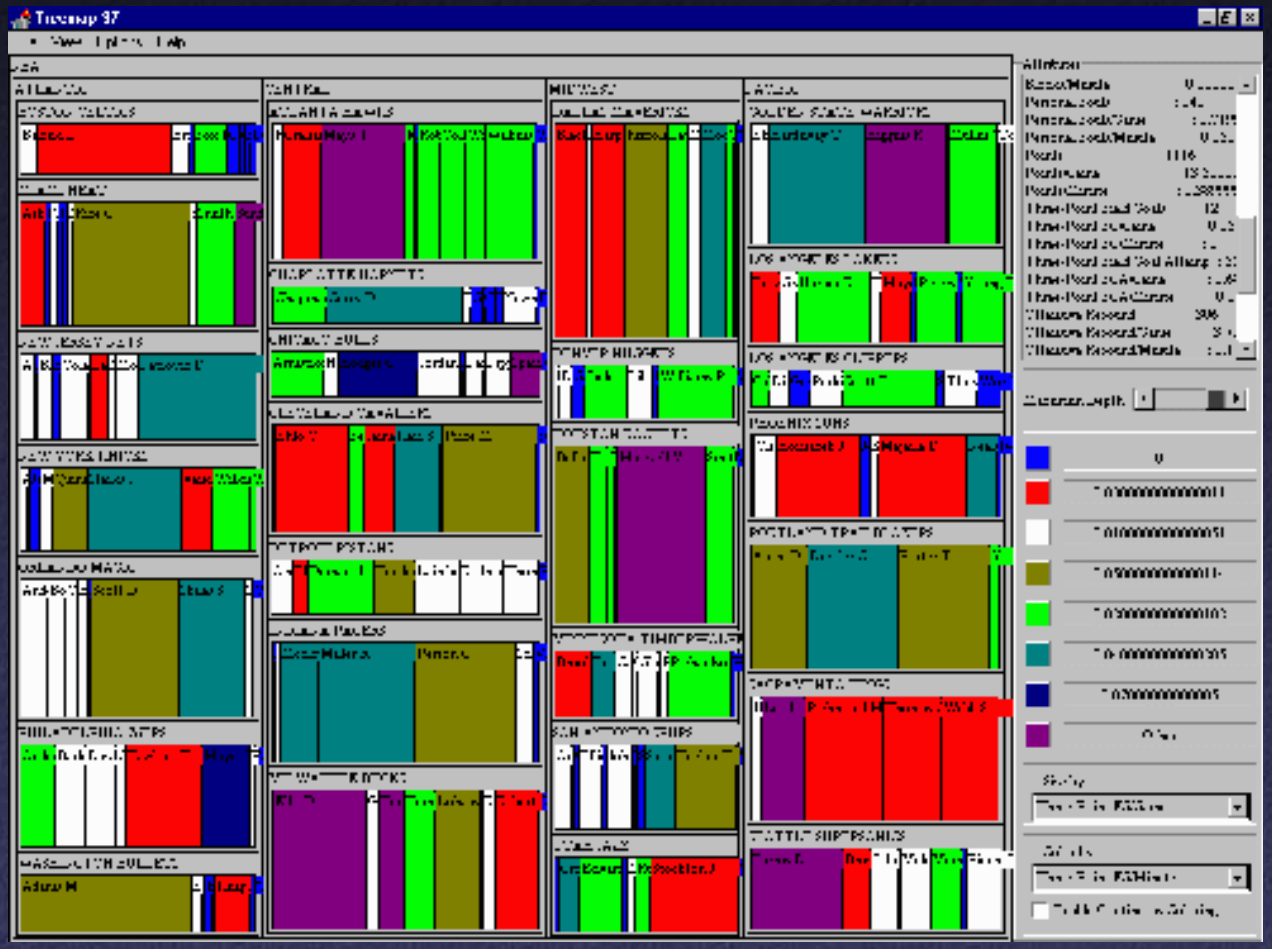
Two Different Primary Goals: Two Different Types of Viz

- Explore/Calculate
  - Analyze
  - Reason about Information
- Communicate
  - Explain
  - Make Decisions
  - Reason about Information

Case Study: The Journey of the TreeMap

- The TreeMap [Johnson & Shneiderman '91]
- Idea:
  - Show a hierarchy as a 2D layout
  - Fill up the space with rectangles representing objects
  - Size on screen indicates relative size of underlying objects

Early Treemap Applied to File System



Treemap Problems

- Too disorderly
  - What does adjacency mean?
  - Aspect ratios uncontrolled leads to lots of skinny boxes that clutter
- Color not used appropriately
  - In fact, is meaningless here
- Wrong application
  - Don't need all this to just see the largest files

Successful Application of Treemaps

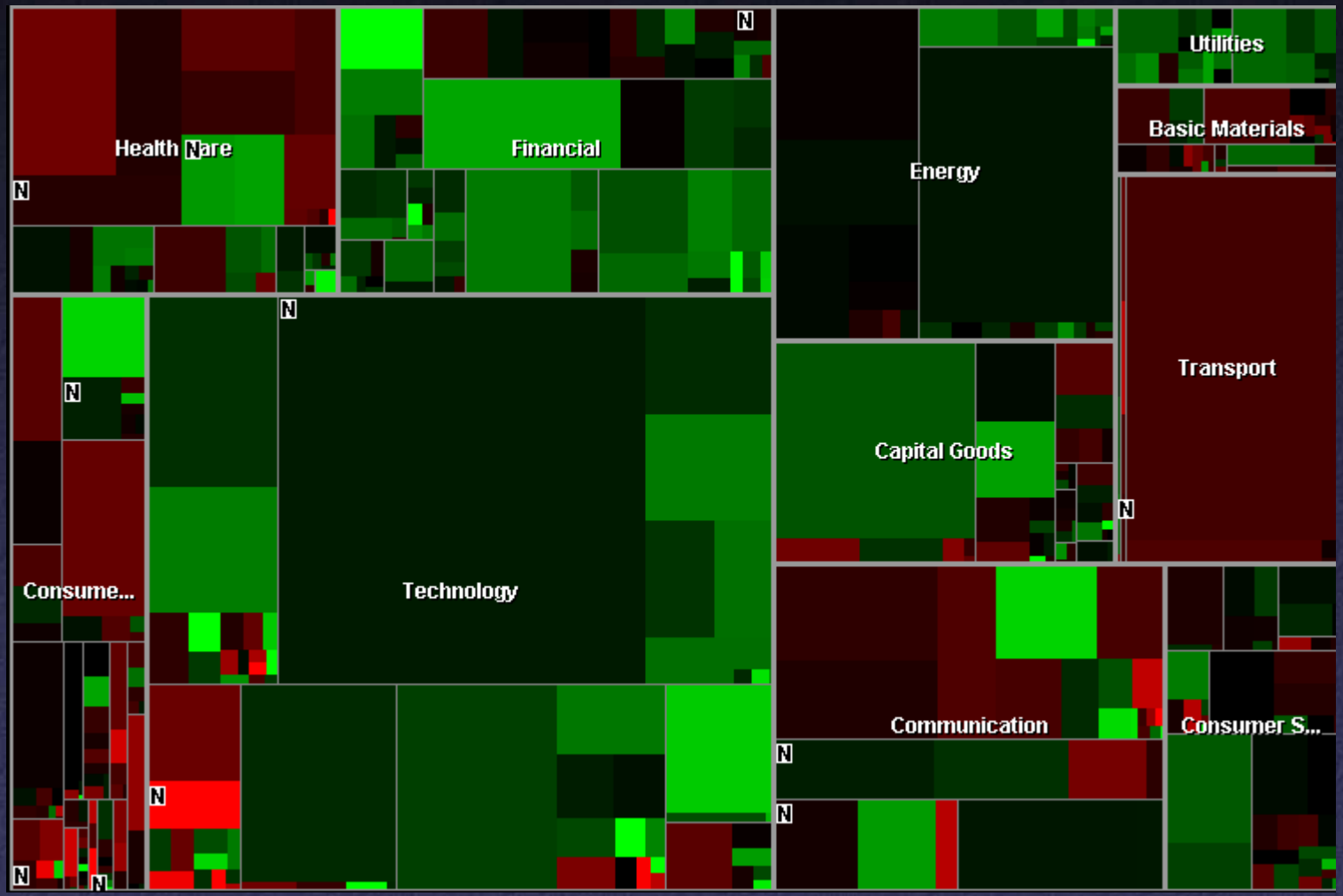
- Think more about the use
  - Break into meaningful groups
  - Fix these into a useful aspect ratio
- Use visual properties (e.g. color) properly
  - Use only two colors: easily visible tagging of qualitative properties
- Provide interactivity
  - Access to the real data
  - Makes it into a useful tool

TreeMaps in Action

http://www.smartmoney.com/maps

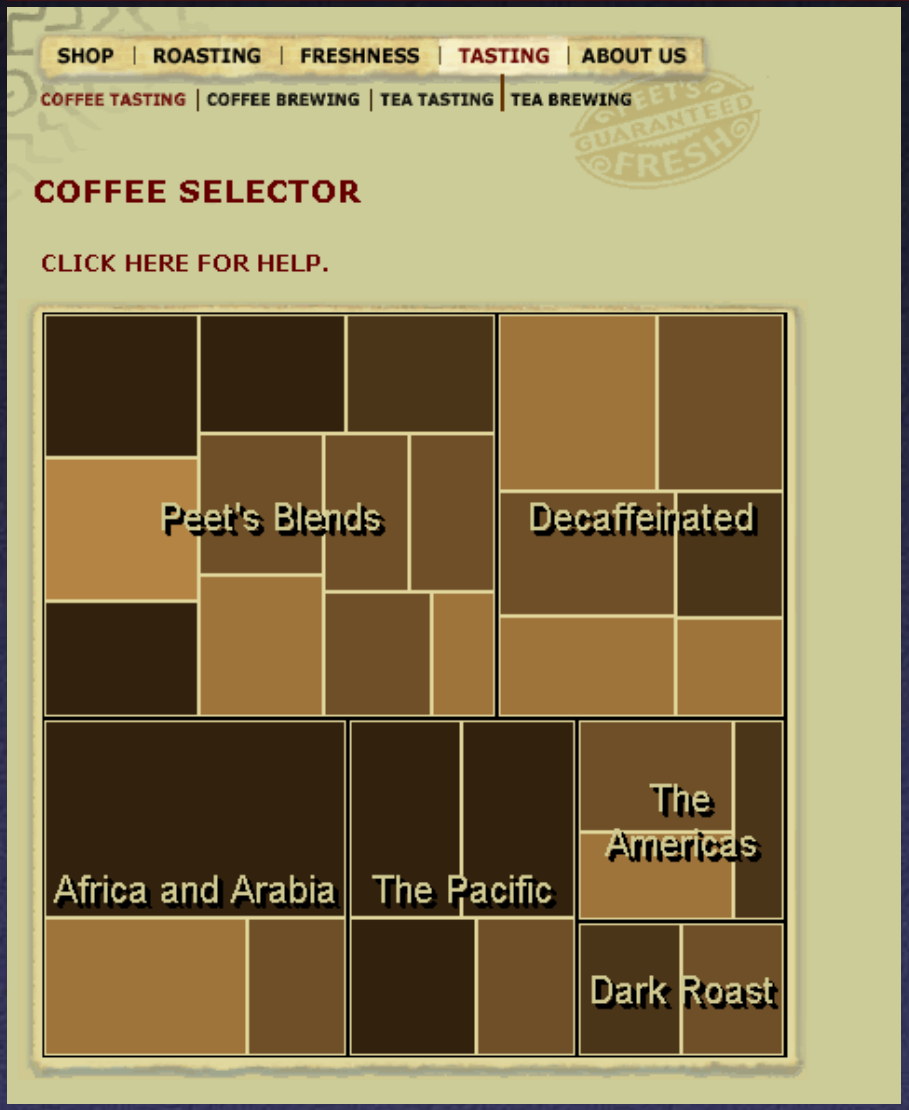
http://www.peets.com/tast/11/coffee_selector.asp

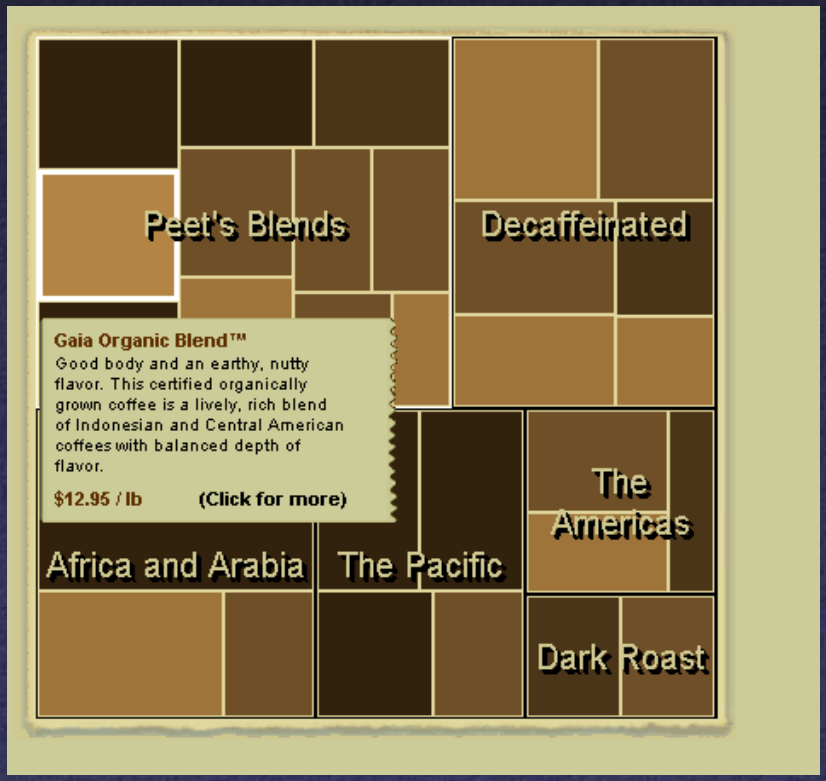
A Good Use of TreeMaps and Interactivity



http://www.smartmoney.com/marketmap

Treemaps in Peet's site





Analysis vs. Communication

- MarketMap's use of TreeMaps allows for sophisticated analysis
- Peet's use of TreeMaps is more for presentation and communication

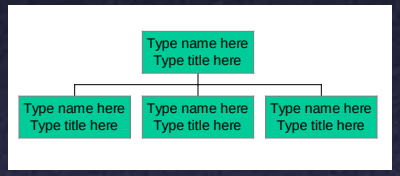
Visual Principles

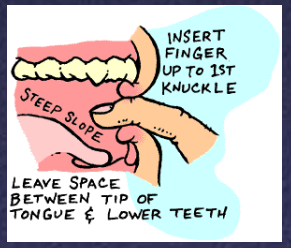
Visual Principles

- Types of Graphs
- Pre-attentive Properties
- Relative Expressiveness of Visual Cues
- Visual Illusions
- Tufte's notions
  - Graphical Excellence
  - How to Lie with Visualization
  - Data-Ink Ratio Maximization

Types of Symbolic Displays

- Graphs
- Charts
- Maps
- Diagrams



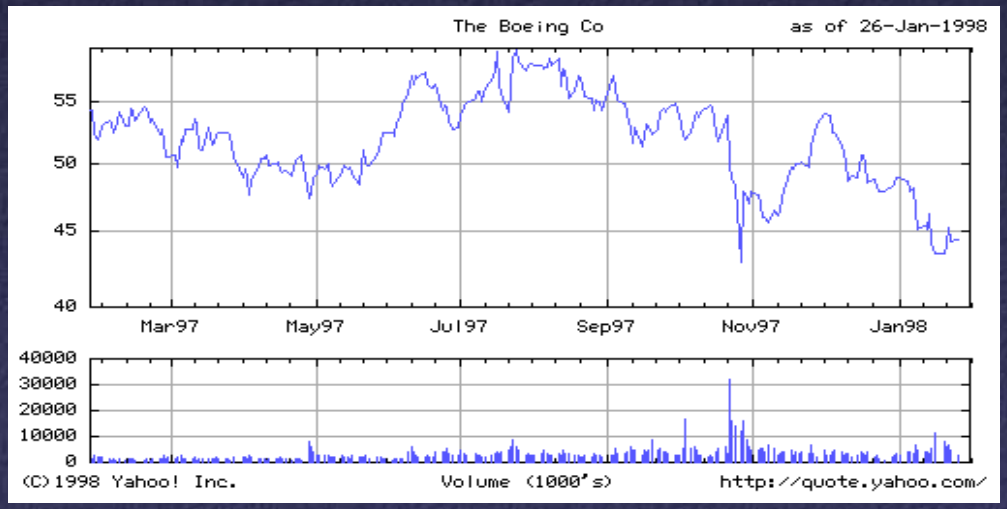


[Kosslyn]

Types of Symbolic Displays

Graphs

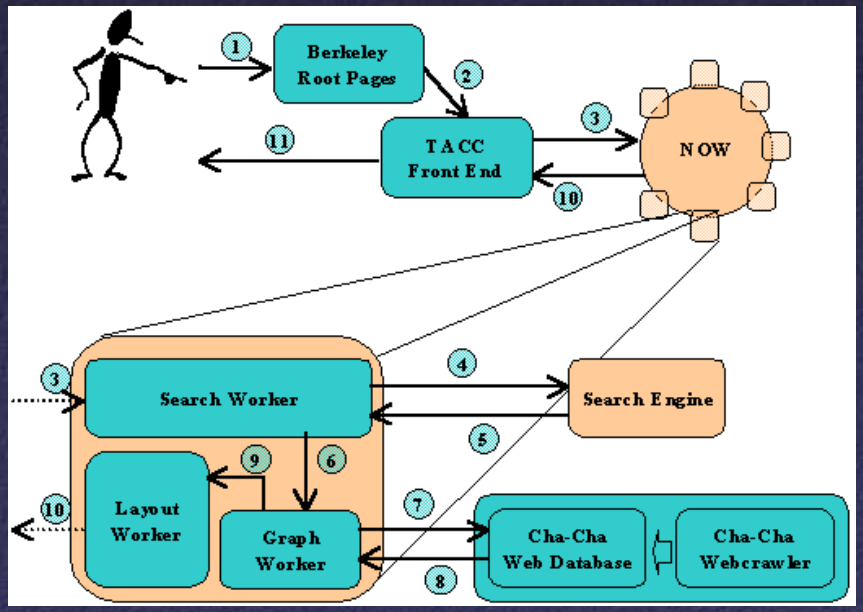
- at least two scales required
- values associated by symmetric “paired with” relation
- Examples: scatter-plot, bar-chart, layer-graph



Types of Symbolic Displays

Charts

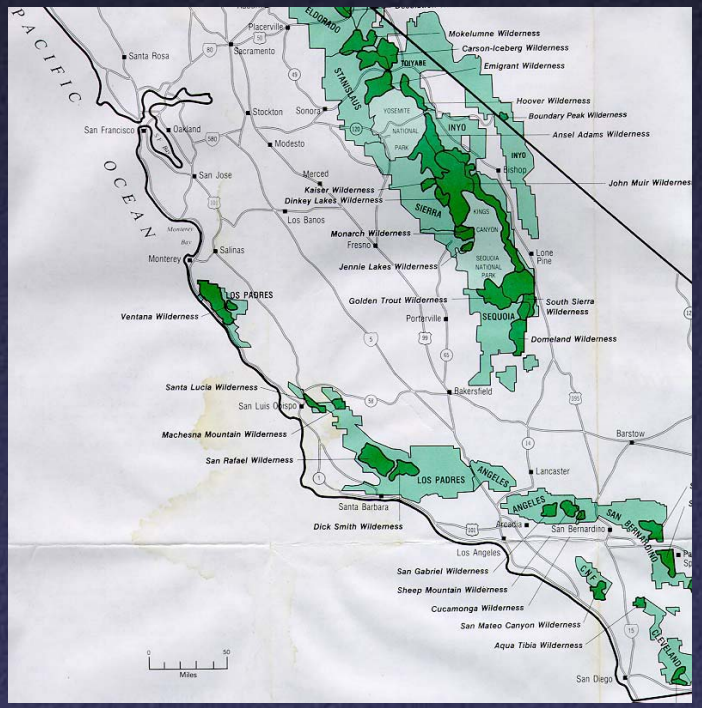
- discrete relations among discrete entities
- structure relates entities to one another
- lines and relative position serve as links
- Examples:
  family tree,
  flow chart,
  network diagram



Types of Symbolic Displays

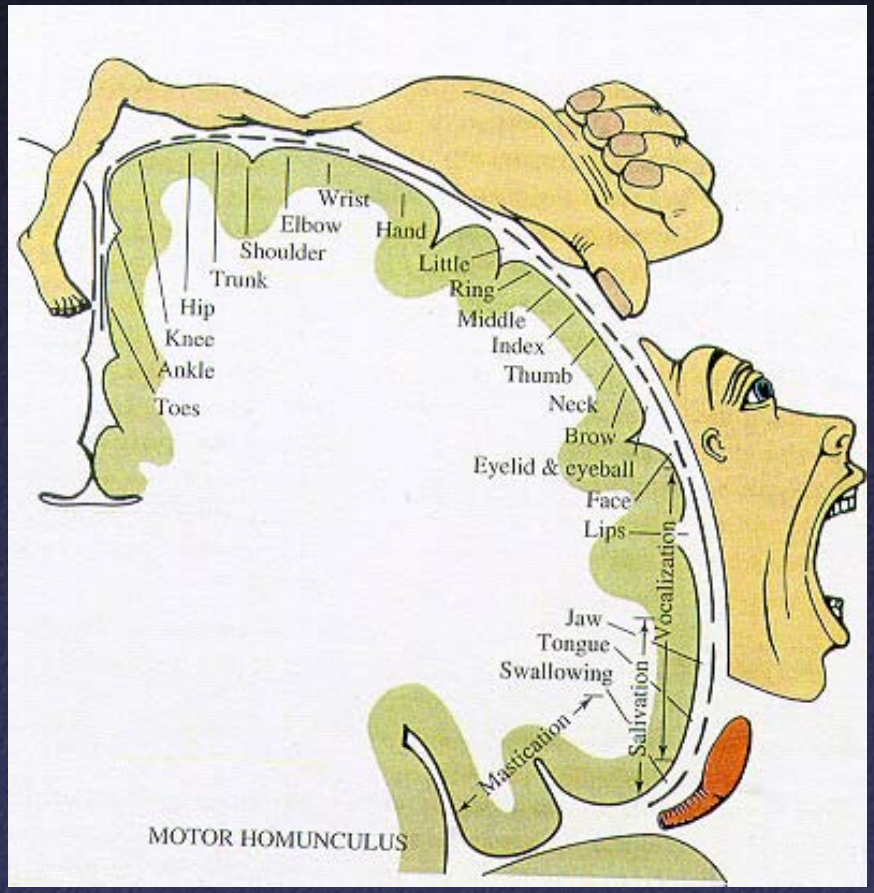
Maps

- internal relations determined (in part) by the spatial relations of what is pictured
- labels paired with locations
- Examples:
  physical maps,
  topographic maps,
  political maps,
  maps of census data



Types of Symbolic Displays

- Diagrams
  - schematic pictures of objects or entities
  - parts are symbolic (unlike photographs)
  - Examples:
    how-to illustrations, figures in a manual



[Glietman]

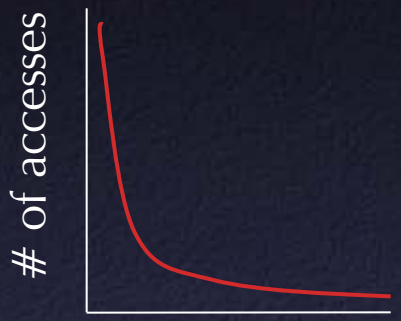
Anatomy of a Graph [Kosslyn 89]

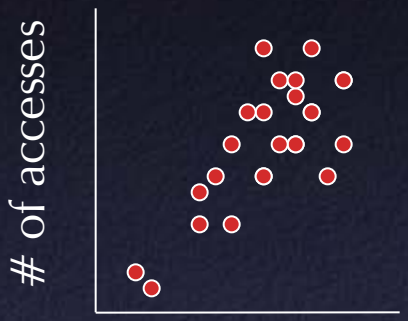
- Framework
  - sets the stage
  - kinds of measurements, scale, ...
- Content
  - marks
  - point symbols, lines, areas, bars, …
- Labels
  - title, axes, tic marks, ...

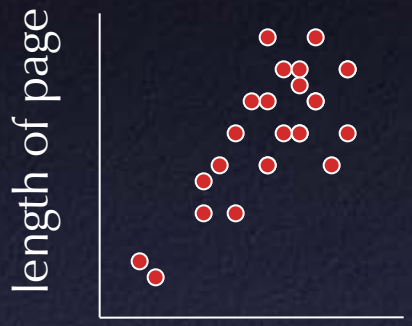
Basic Types of Data

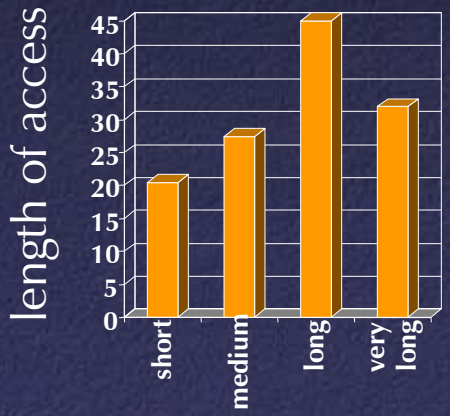
- Nominal (qualitative)
  - (no inherent order)
  - city names, types of diseases, ...
- Ordinal (qualitative)
  - (ordered, but not at measurable intervals)
  - first, second, third, …
  - cold, warm, hot
- Interval (quantitative)
  - list of integers or reals

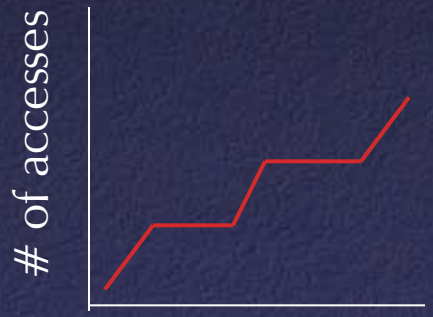
Common Graph Types











When to use which type?

- Line graph
  - x-axis requires quantitative variable
  - Variables have contiguous values
  - familiar/conventional ordering among ordinals
- Bar graph
  - comparison of relative point values
- Scatter plot
  - convey overall impression of relationship between two variables
- Pie Chart?
  - Emphasizing differences in proportion among a few numbers

Classifying Visual Representations

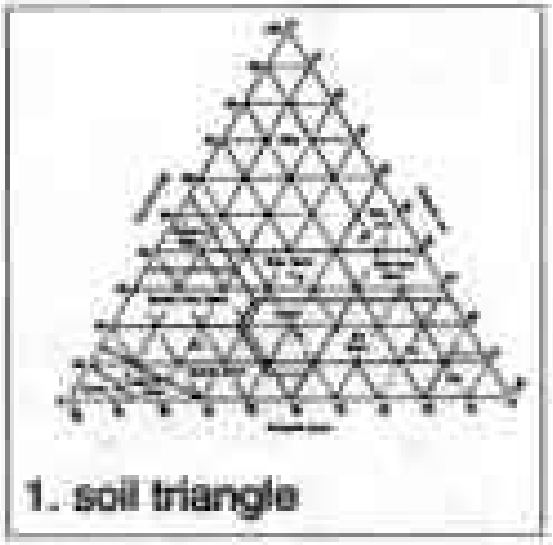
Lohse, G L; Biolsi, K; Walker, N and H H Rueter,

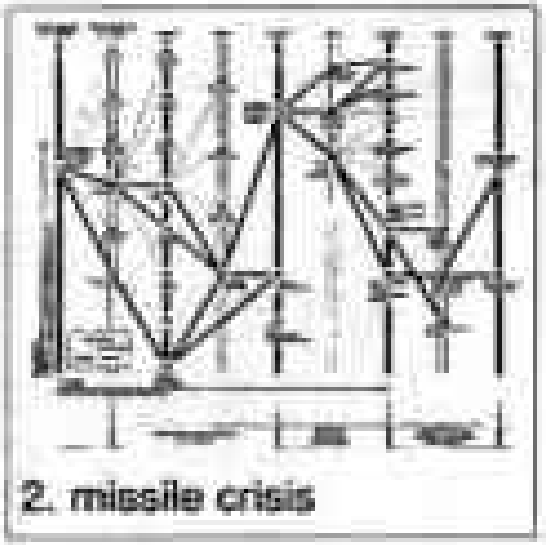
A Classification of Visual Representations

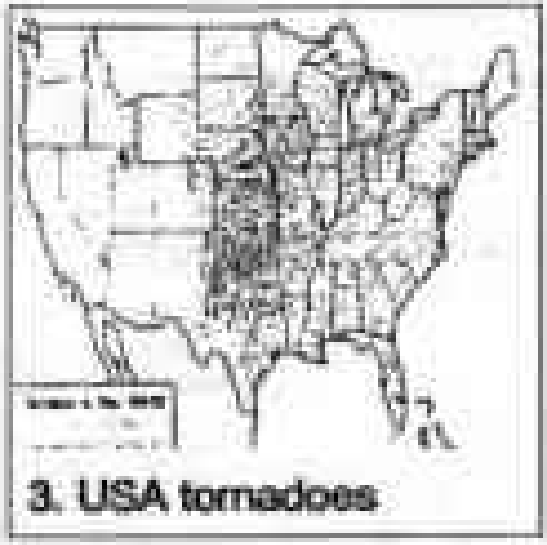
CACM, Vol. 37, No. 12, pp 36-49, 1994

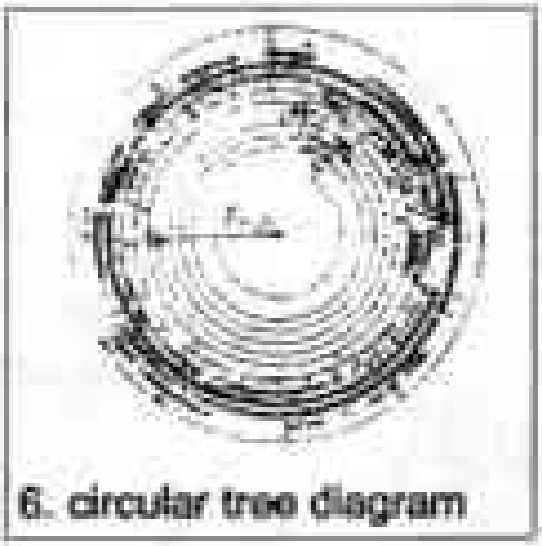
- Participants sorted 60 items into categories
- Others assigned labels from Likert scales
- Experimenters clustered the results various ways.

Subset of Example Visual Representations From Lohse et al. 94





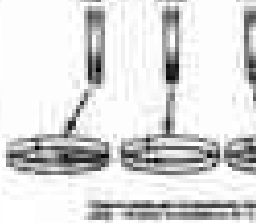




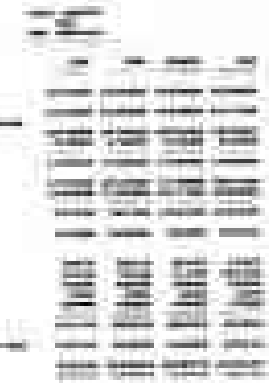
Lake

10. gibberellins

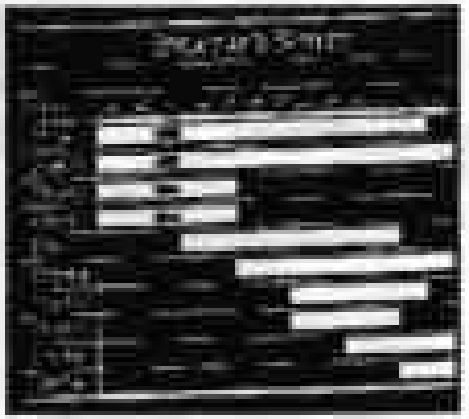
11. heart



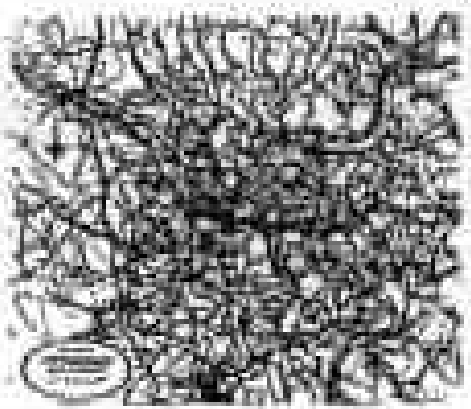
12. microbiolo



dsheet budget

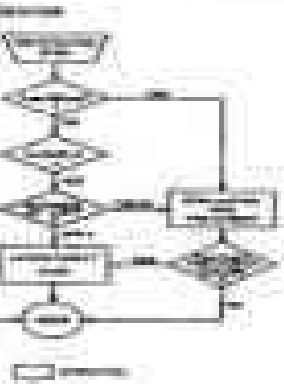


14. Gantt chart

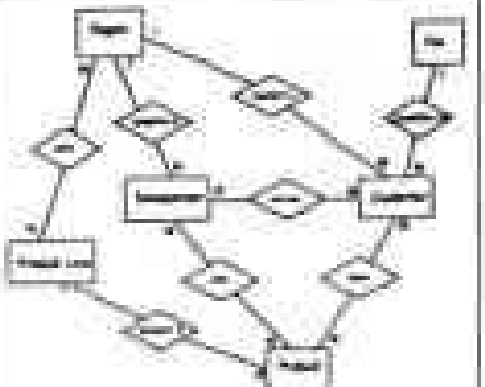


15. Pittsburgh map

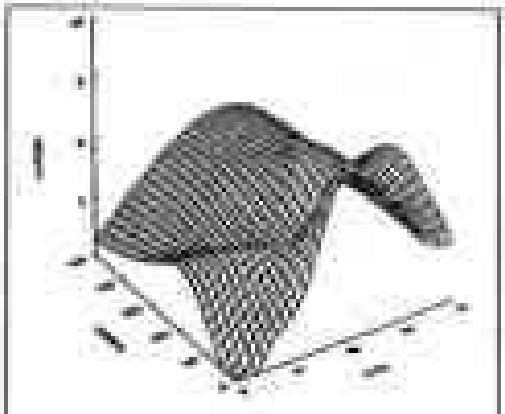
16. list of integ



hart



18. data model



19. response surface

20. wheelbarr

Interesting Findings

Lohse et al. 94

- Photorealistic images were least informative
  - Echos results in icon studies – better to use less complex, more schematic images
- Graphs and tables are the most self-similar categories
  - Results in the literature comparing these are inconclusive
- Temporal data more difficult to show than cyclic data
  - Recommend using animation for temporal data

Visual Properties

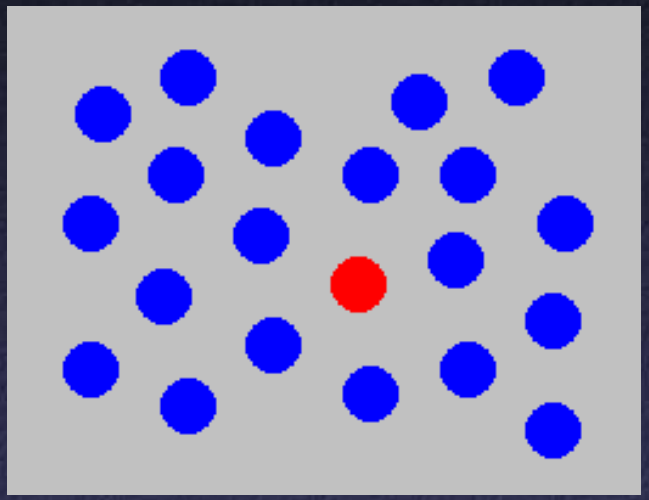
- Preattentive Processing
- Accuracy of Interpretation of Visual Properties
- Illusions and the Relation to Graphical Integrity

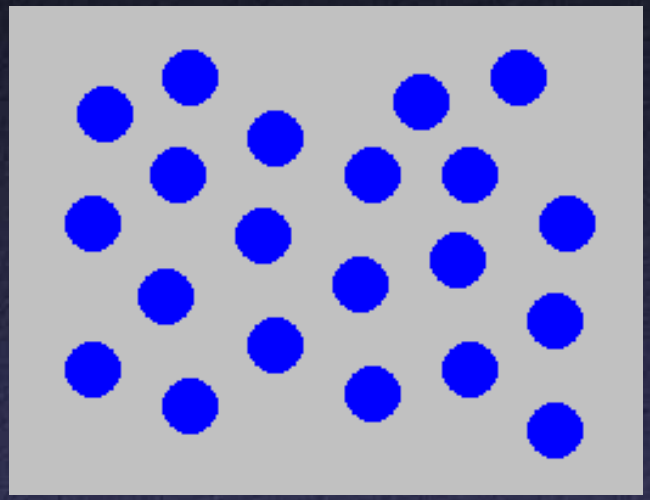
Preattentive processing sildes from Healey
http://www.csc.ncsu.edu/faculty/healey/PP/PP.html

Preattentive Processing

- Some properties are processed preattentively
  (without need for focusing attention).
- Important for design of visualizations
  - what can be perceived immediately
  - what properties are good discriminators
  - what can mislead viewers

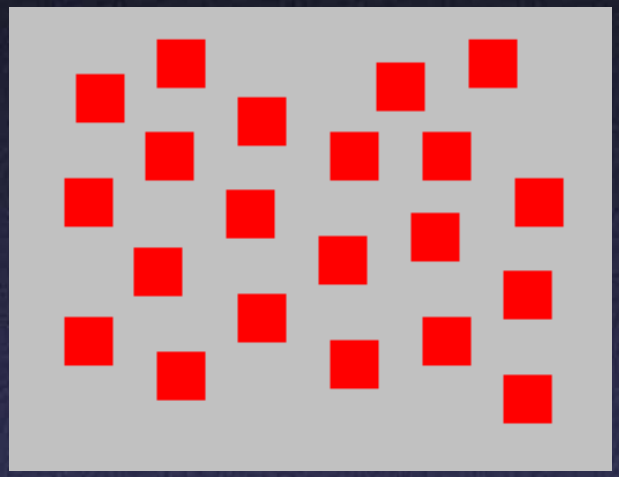
Example: Color Selection

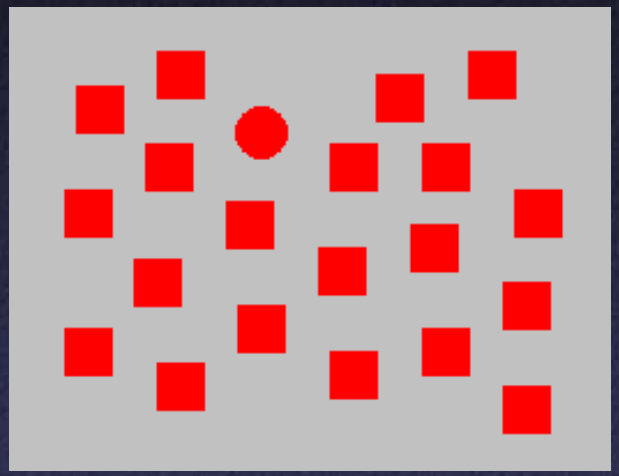




Viewer can rapidly and accurately determine whether the target (red circle) is present or absent. Difference detected in color.

Example: Shape Selection



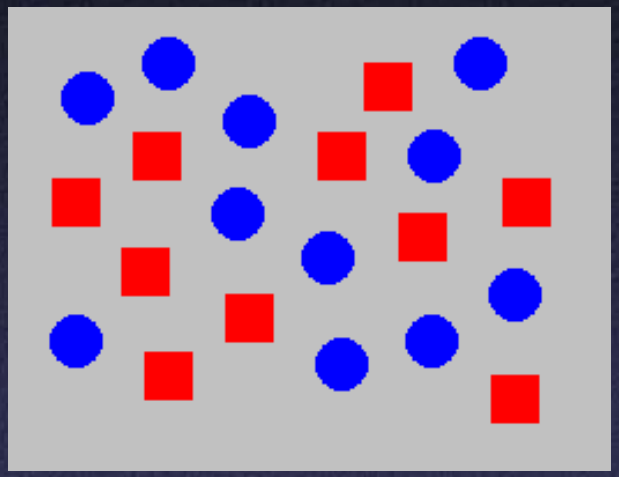


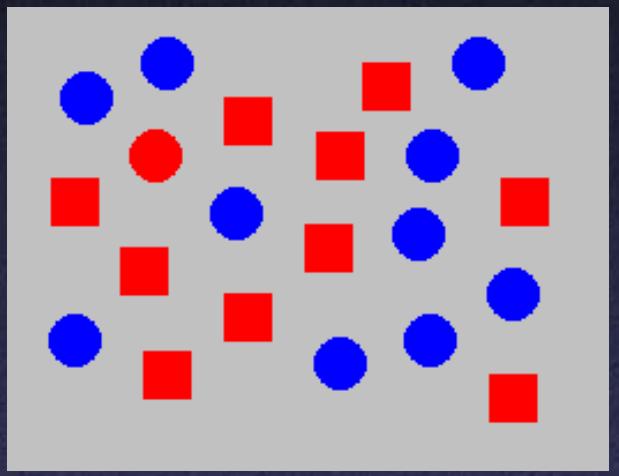
Viewer can rapidly and accurately determine whether the target (red circle) is present or absent. Difference detected in form (curvature)

Pre-attentive Processing

- < 200–250 ms qualifies as pre-attentive
  - eye movements take at least 200ms
  - yet certain processing can be done very quickly, implying low-level processing in parallel
- If a decision takes a fixed amount of time regardless of the number of distractors, it is considered to be preattentive

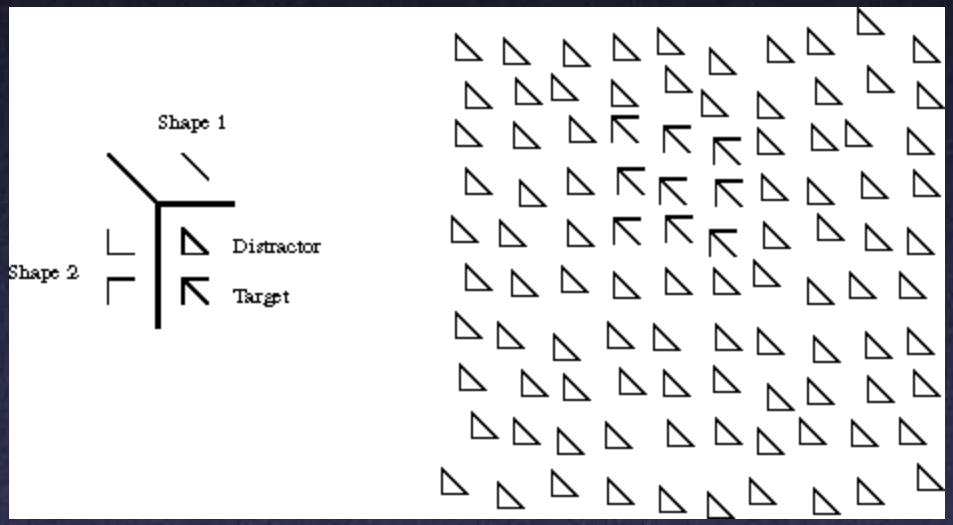
Example: Conjunction of Features





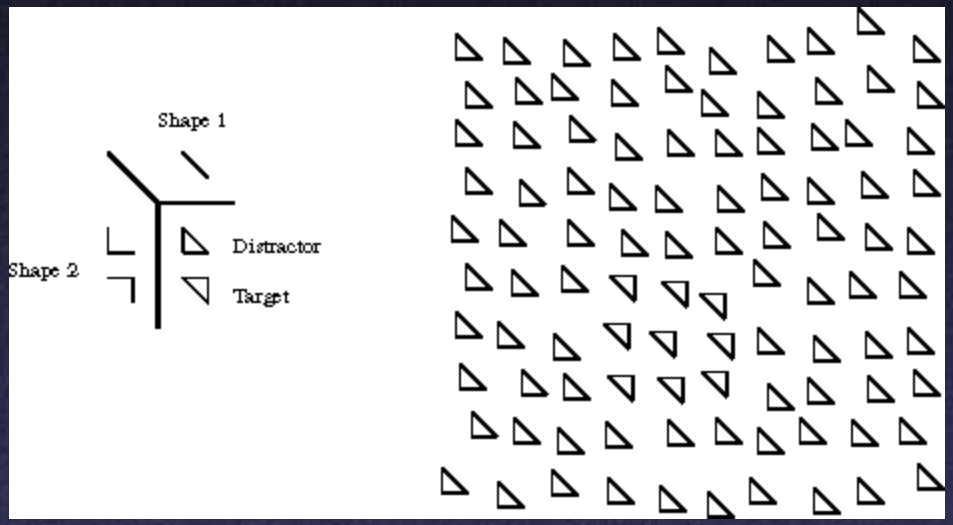
Viewer *cannot* rapidly and accurately determine whether the target (red circle) is present or absent when target has two or more features, each of which are present in the distractors. Viewer must search sequentially.

Example: Emergent Features



Target has a unique feature with respect to distractors (open sides) and so the group can be detected preattentively.

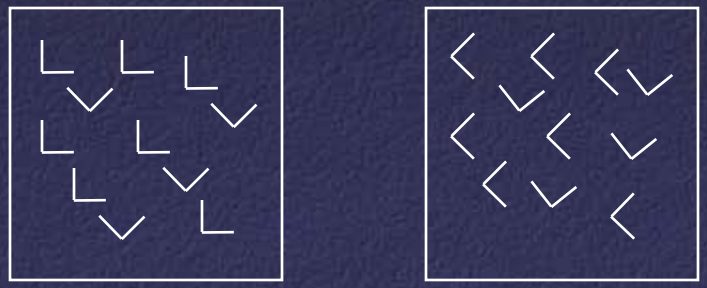
Example: Emergent Features



Target does not have a unique feature with respect to distractors and so the group cannot be detected preattentively.

Asymmetric and Graded Preattentive Properties

- Some properties are asymmetric
  - a sloped line among vertical lines is preattentive
  - a vertical line among sloped ones is not
- Some properties have a gradation
  - some more easily discriminated among than others



SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC

Text NOT Preattentive

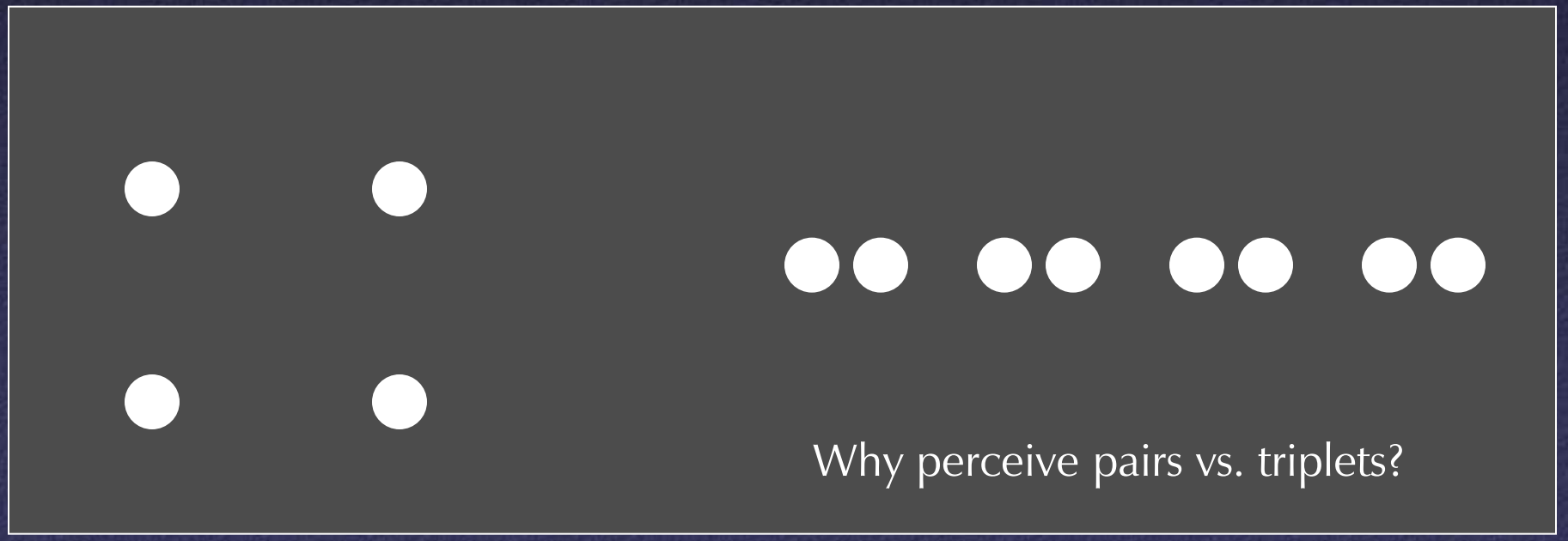
SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC

Preattentive Visual Properties [Healey 97]

length	Triesman & Gormican [1988]
width	Julesz [1985]
size	Triesman & Gelade [1980]
curvature	Triesman & Gormican [1988]
number	Julesz [1985]; Trick & Pylyshyn [1994]
terminators	Julesz & Bergen [1983]
intersection	Julesz & Bergen [1983]
closure	Enns [1986]; Triesman & Souther [1985]
colour (hue)	Nagy & Sanchez [1990, 1992]; D'Zmura [1991] Kawai et al. [1995]; Bauer et al. [1996]
intensity	Beck et al. [1983]; Triesman & Gormican [1988]
flicker	Julesz [1971]
direction of motion	Nakayama & Silverman [1986]; Driver & McLeod [1992]
binocular lustre	Wolfe & Franzel [1988]
stereoscopic depth	Nakayama & Silverman [1986]
3-D depth cues	Enns [1990]
lighting direction	Enns [1990]

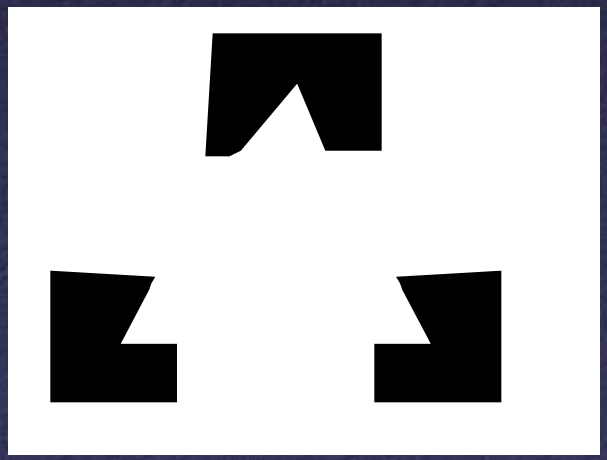
Gestalt Properties

- *Gestalt:* form or configuration
- Idea: forms or patterns transcend the stimuli used to create them
  - Why do patterns emerge? Under what circumstances?



Gestalt Laws of Perceptual Organization [Kaufman 74]

- Figure and Ground
  - Escher illustrations are good examples
  - Vase/Face contrast
- Subjective Contour



More Gestalt Laws

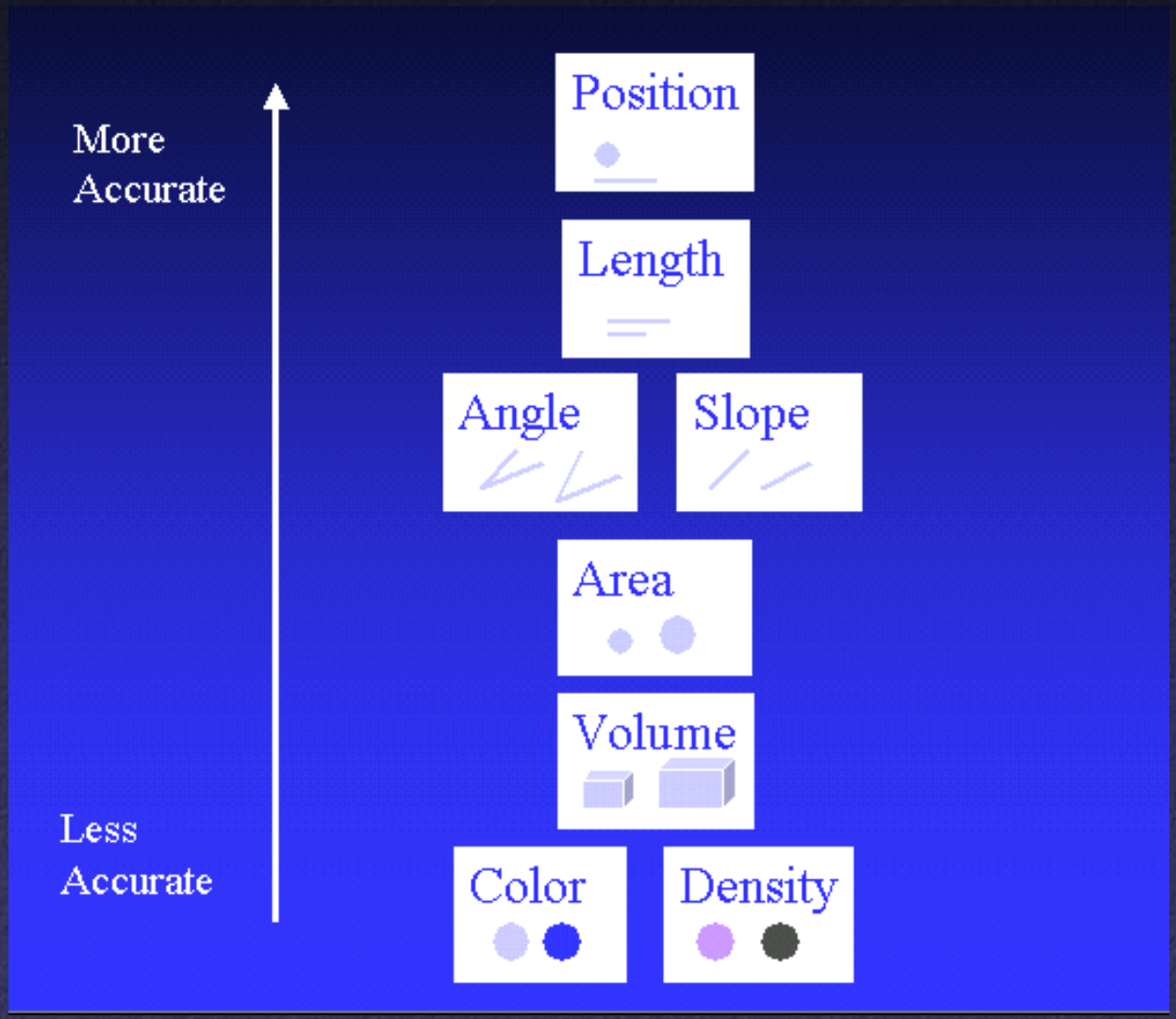
- Law of Proximity
  - Stimulus elements that are close together will be perceived as a group
- Law of Similarity
  - like the preattentive processing examples
- Law of Common Fate
  - like preattentive motion property
    - move a subset of objects among similar ones and they will be perceived as a group

Which Properties are Appropriate for Which Information Types?

Accuracy Ranking of Quantitative Perceptual Tasks

Estimated; only pairwise comparisons have been validated

[Mackinlay 88 from Cleveland & McGill]



Interpretations of Visual Properties

Some properties discriminated more accurately but have no intrinsic meaning [Senay & Ingatious 97, Kosslyn, others]

- Density (Greyscale)
  - Darker → More
- Size / Length / Area
  - Larger → More
- Position
  - Leftmost → first, Topmost → first
- Hue
  - ??? no intrinsic meaning
- Slope
  - ??? no intrinsic meaning

Ranking of Applicability of Properties for Different Data Types

[Mackinlay 88, Not Empirically Verified]

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Color Hue
Angle	Color Saturation	Texture
Slope	Color Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Color Saturation
Color Saturation	Length	Shape
Color Hue	Angle	Length

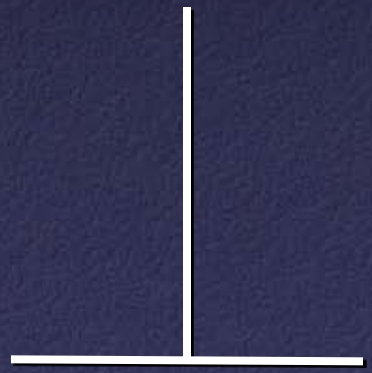
Visual Illusions

- People don't perceive length, area, angle, brightness they way they "should"
- Some illusions have been reclassified as systematic perceptual errors
  - e.g., brightness contrasts (grey square on white background vs. on black background)
  - partly due to increase in our understanding of the relevant parts of the visual system
- Nevertheless, the visual system does some really unexpected things

Illusions of Linear Extent

- Mueller-Lyon (off by 25-30%)

- Horizontal-Vertical



Illusions of Area

- Delboeuf Illusion

- Height of 4-story building overestimated by approximately 25%

Tufte's Principles of Graphical Excellence

Graphical excellence

- is the well-designed presentation of interesting data – a matter of substance, of statistics, and of design
- consists of complex ideas communicated with clarity, precision and efficiency
- is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
- requires telling the truth about the data

Tufte Principles

- Use multifunctioning graphical elements
- Use small multiples
- Show mechanism, process, dynamics, and causality
- High data density
  - Number of items/area of graphic
  - This is controversial
    - White space thought to contribute to good visual design
    - Tufte's book itself has lots of white space

Tufte's Graphical Integrity

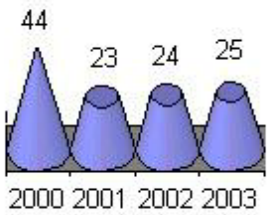
- Some lapses intentional, some not
- Lie Factor = $\frac{\text{size of effect in graph}}{\text{size of effect in data}}$
- Misleading uses of area
- Misleading uses of perspective
- Leaving out important context
- Lack of taste and aesthetics

How to Lie With Visualizations

A common example of a high lie factor occurs when both dimensions of a two-dimensional figure are made proportional to the same data, so that the size of the figure is proportional to the square of the data; for instance,

Year	Books circulated
2001	100 
2002	141 
2003	200 

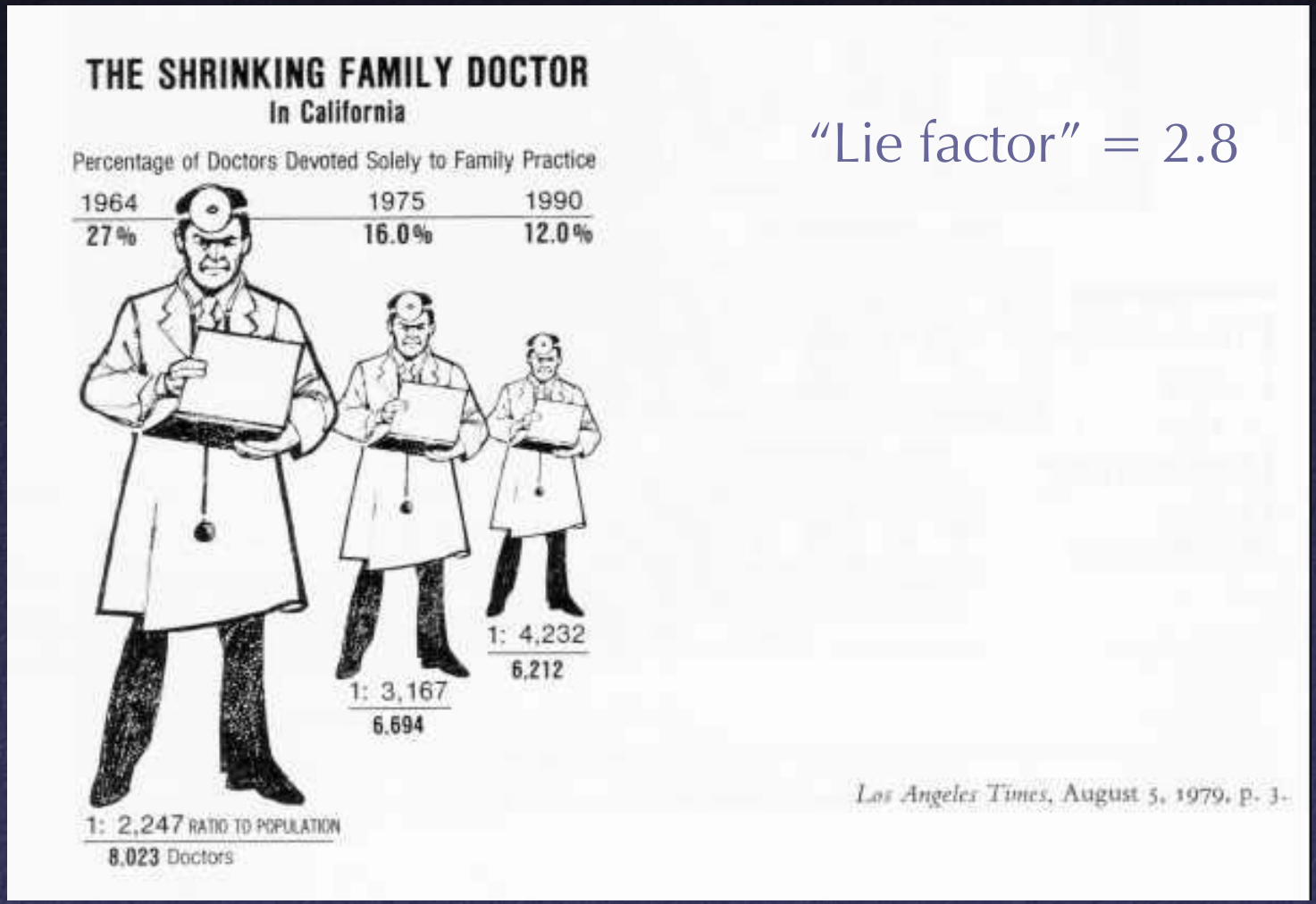
An example of a **low** lie factor can be seen in the "Cones" custom chart format in Microsoft Excel.



The heights of the (truncated) cones are proportional to the data, but their areas on the screen and their apparent volumes make the larger data values seem relatively small.

Tim Craven – http://instruct.uwo.ca/fim-lis/504/504gra.htm#data-ink_ratio

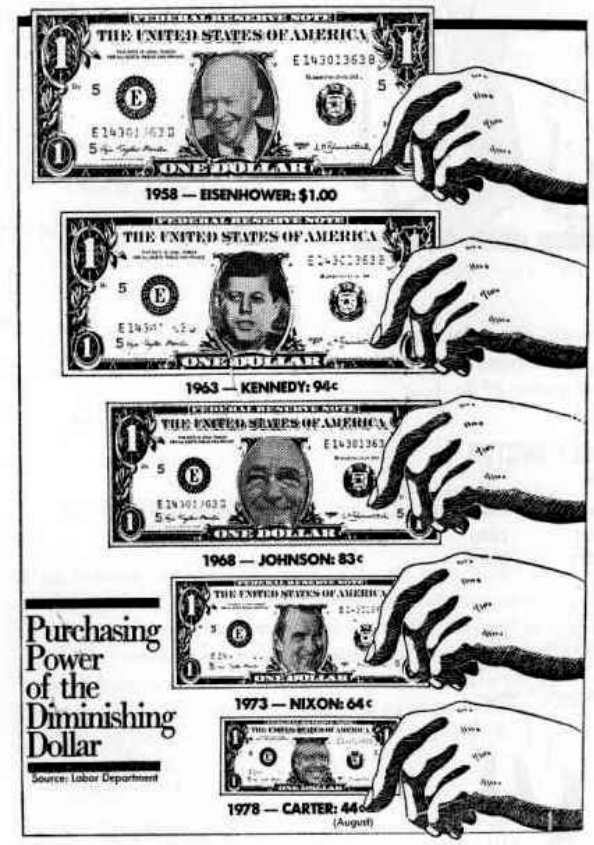
How to Lie With Visualizations



"Lie factor" = 2.8

[Tufte]

How to Lie With Visualizations



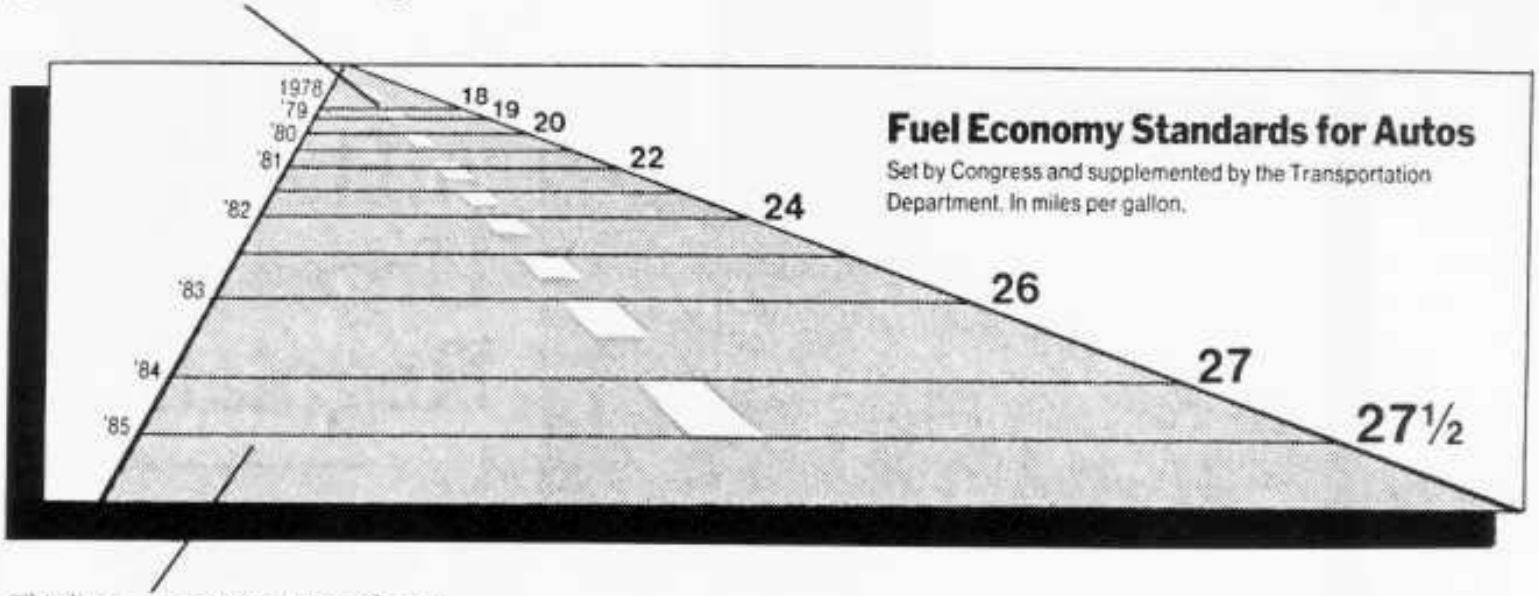
Washington Post, October 25, 1978, p. 1.

Error: Shrinking along both dimensions

[Tufte]

How to Lie With Visualizations

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

New York Times, August 9, 1978, p. D-2.

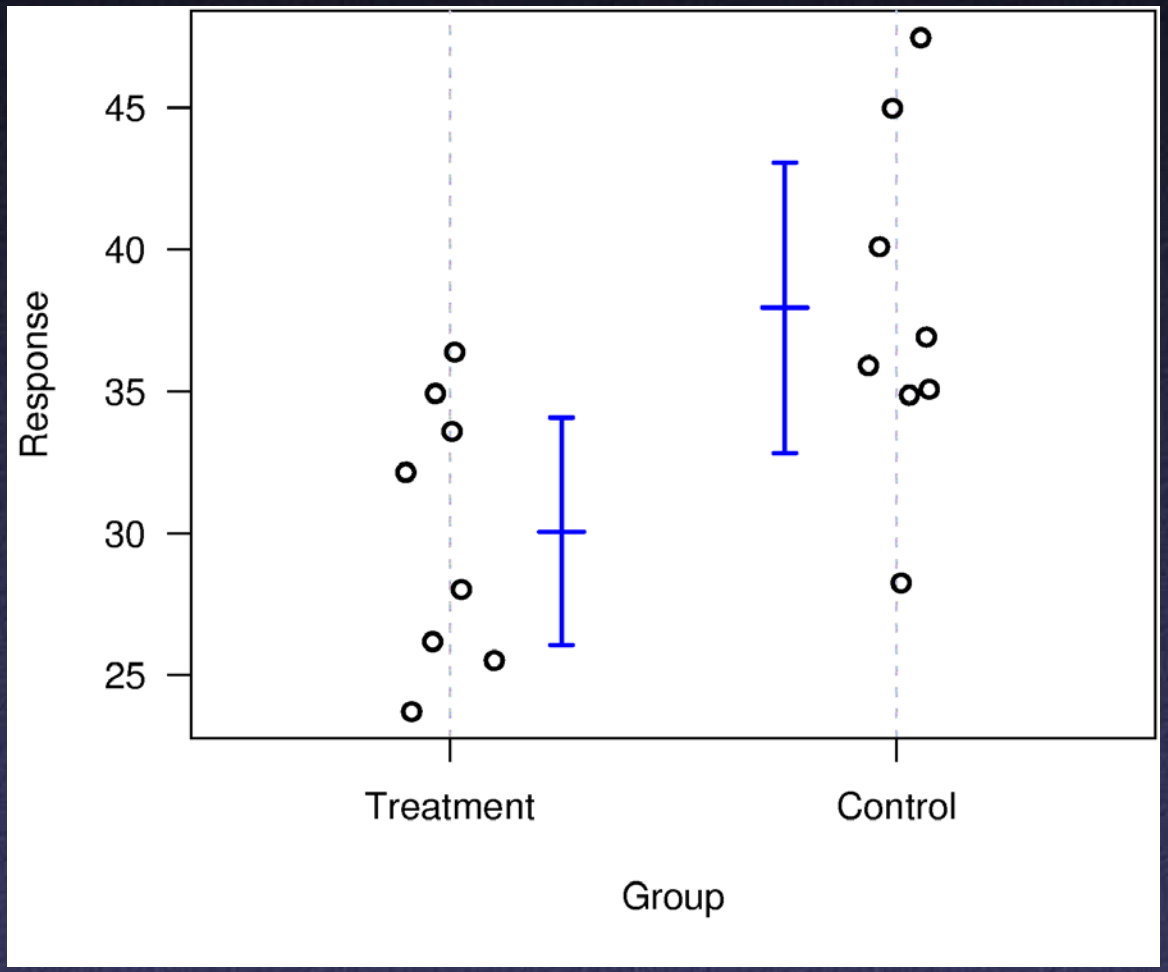
[Tufte]

Tufte's Principle of Data Ink Maximization

- Goal: maximize ratio of "data ink" to total ink
  - draw viewers' attention to the substance of the graphic
  - the role of redundancy
  - principles of editing and redesign
- What's wrong with this?
  What is he really getting at?

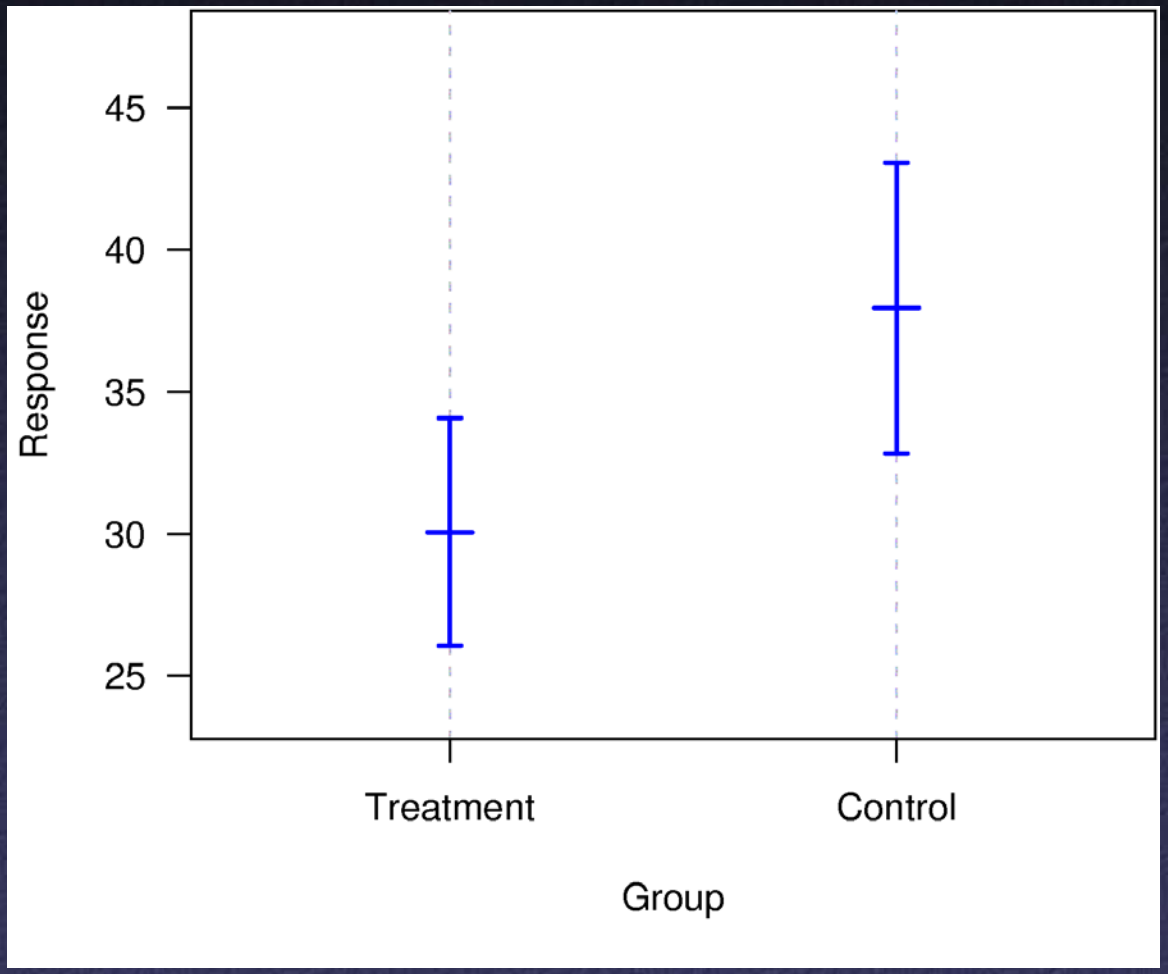
Avoid "chart junk"

Example 1



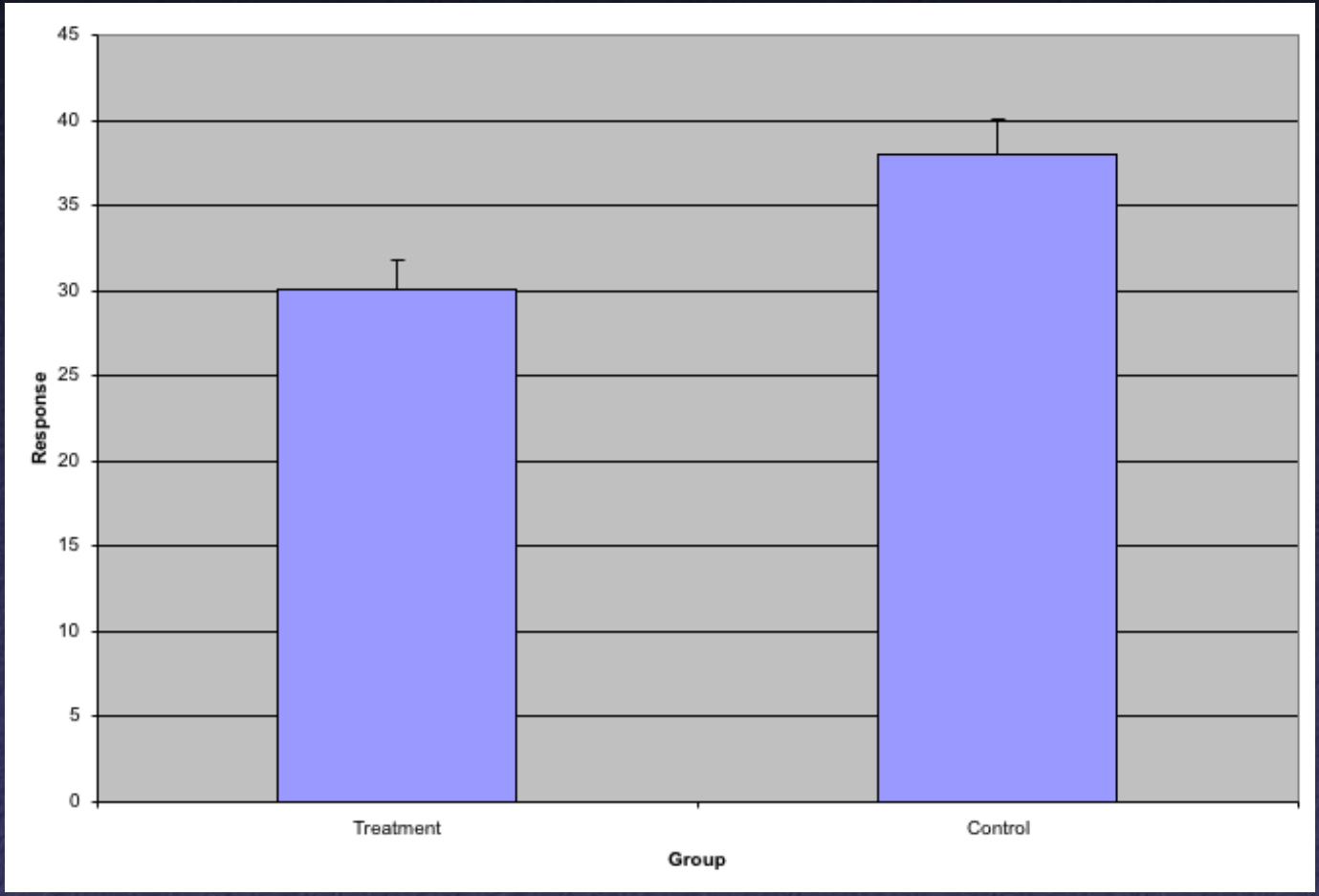
[Karl Broman]

Example 1



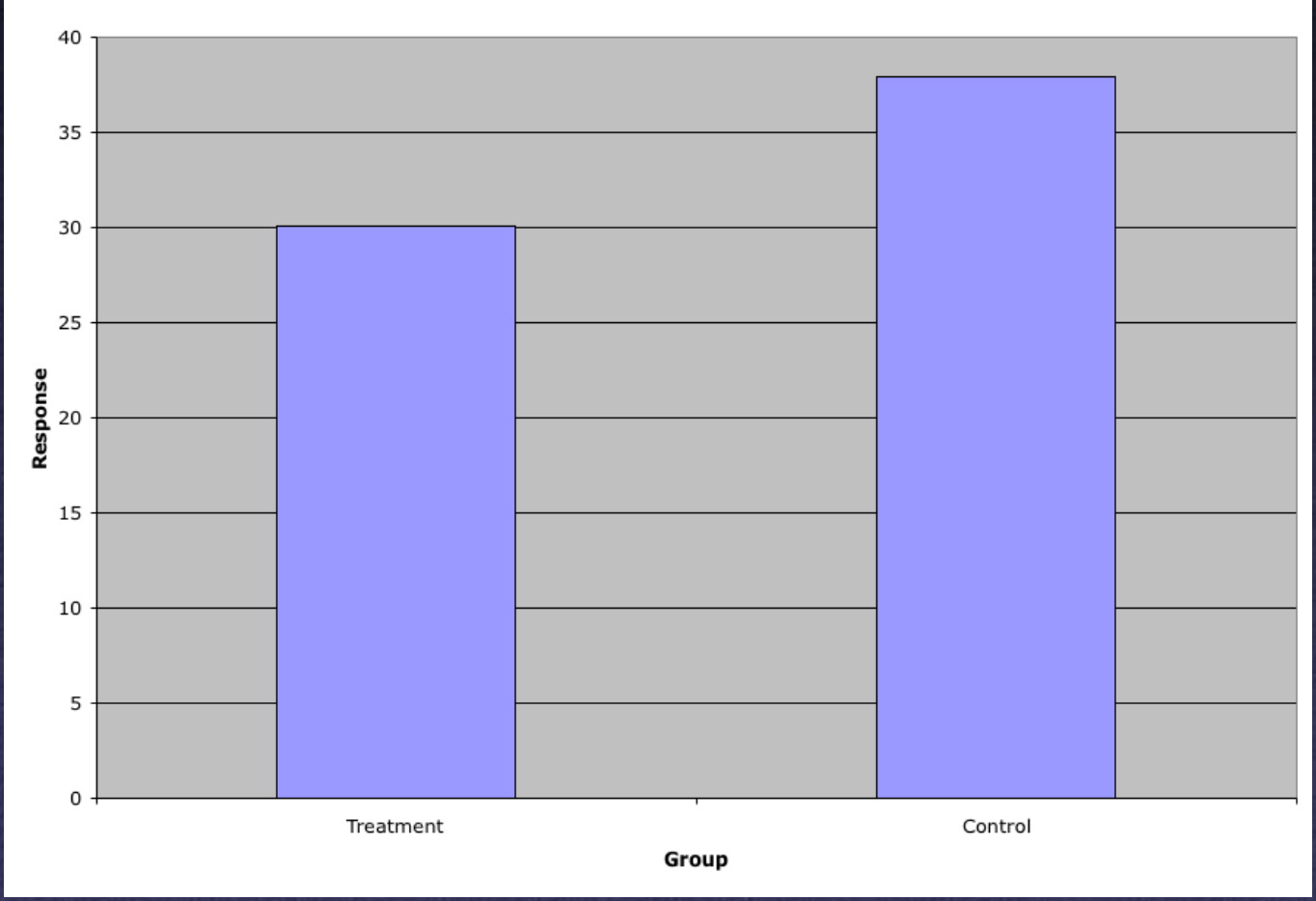
[Karl Broman]

Example 1



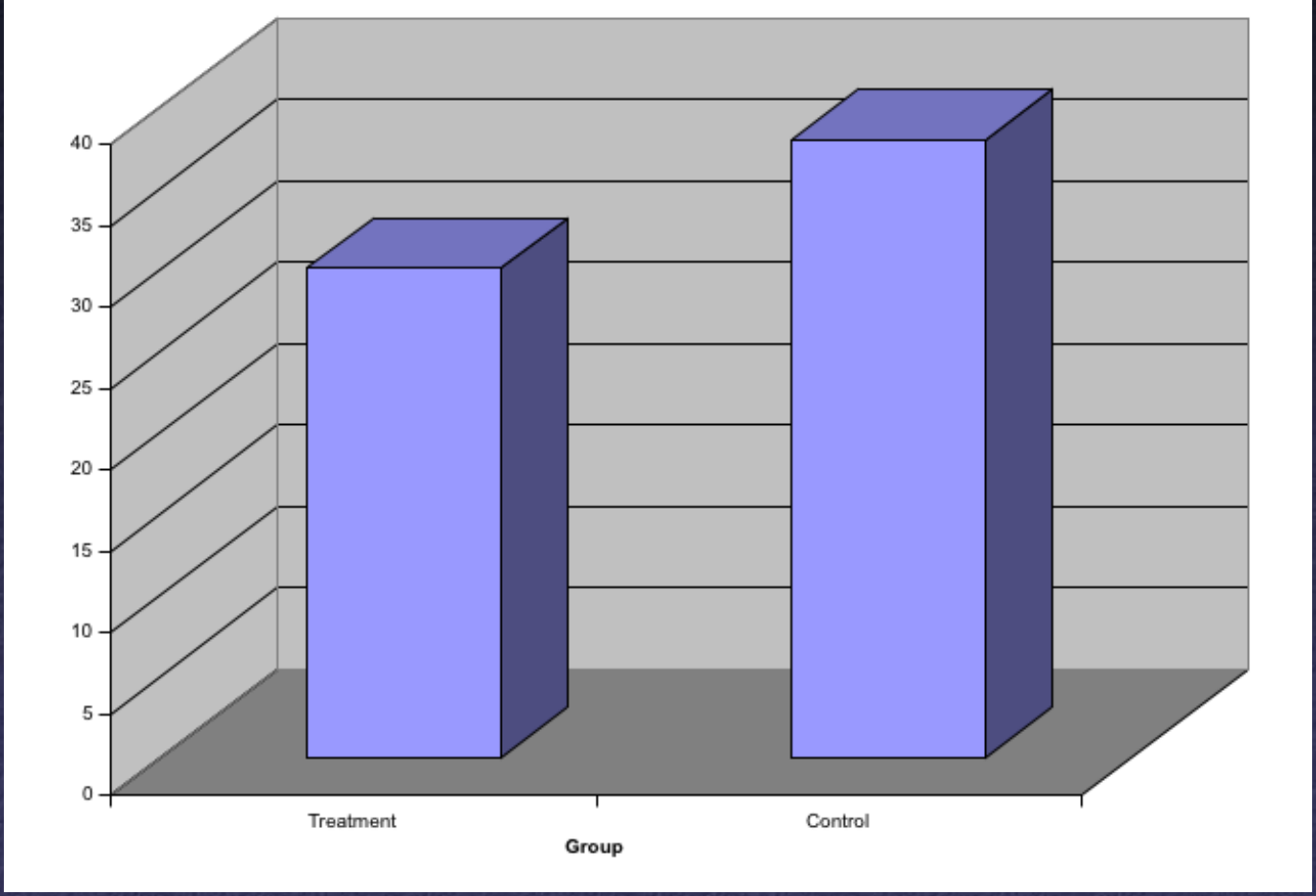
[Karl Broman]

Example 1



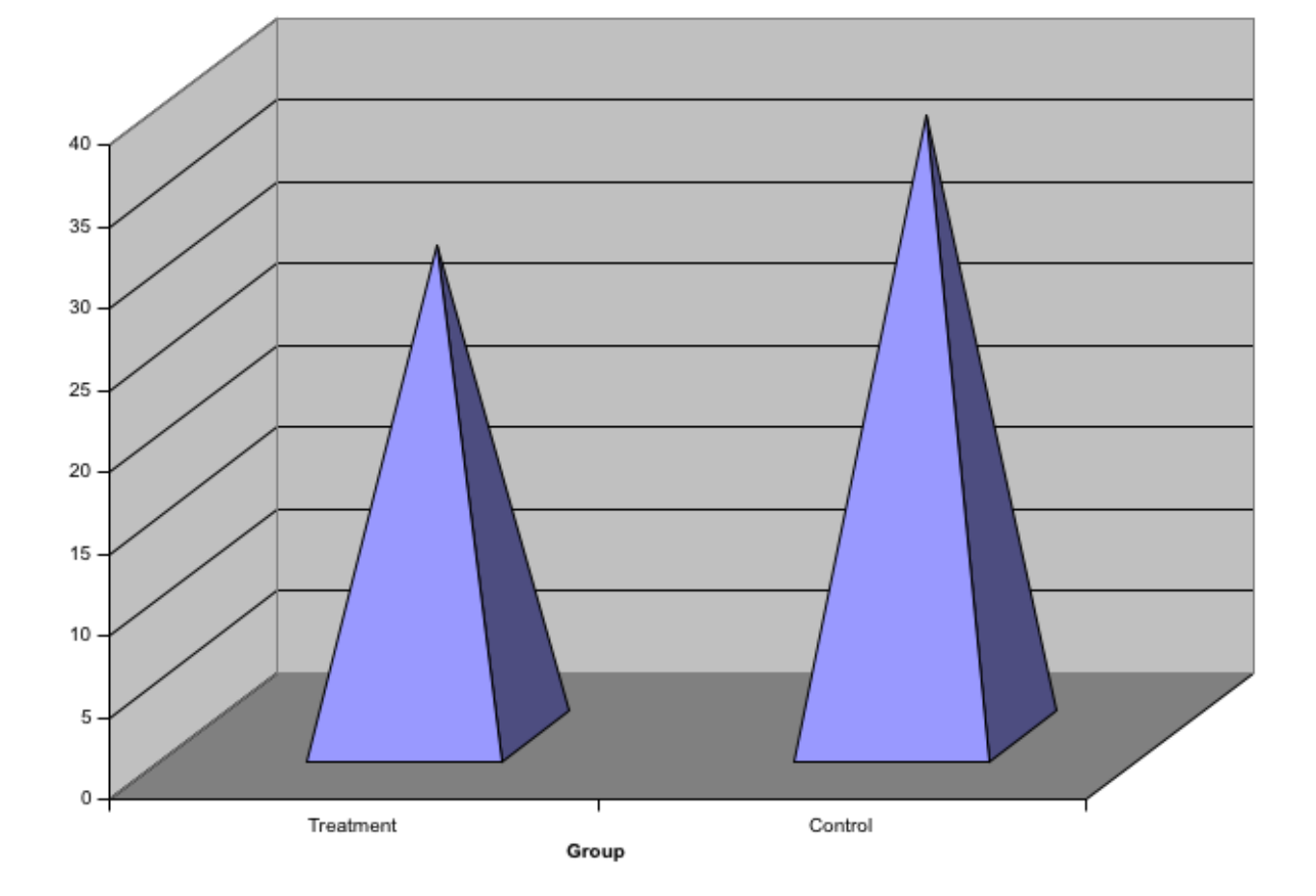
[Karl Broman]

Example 1



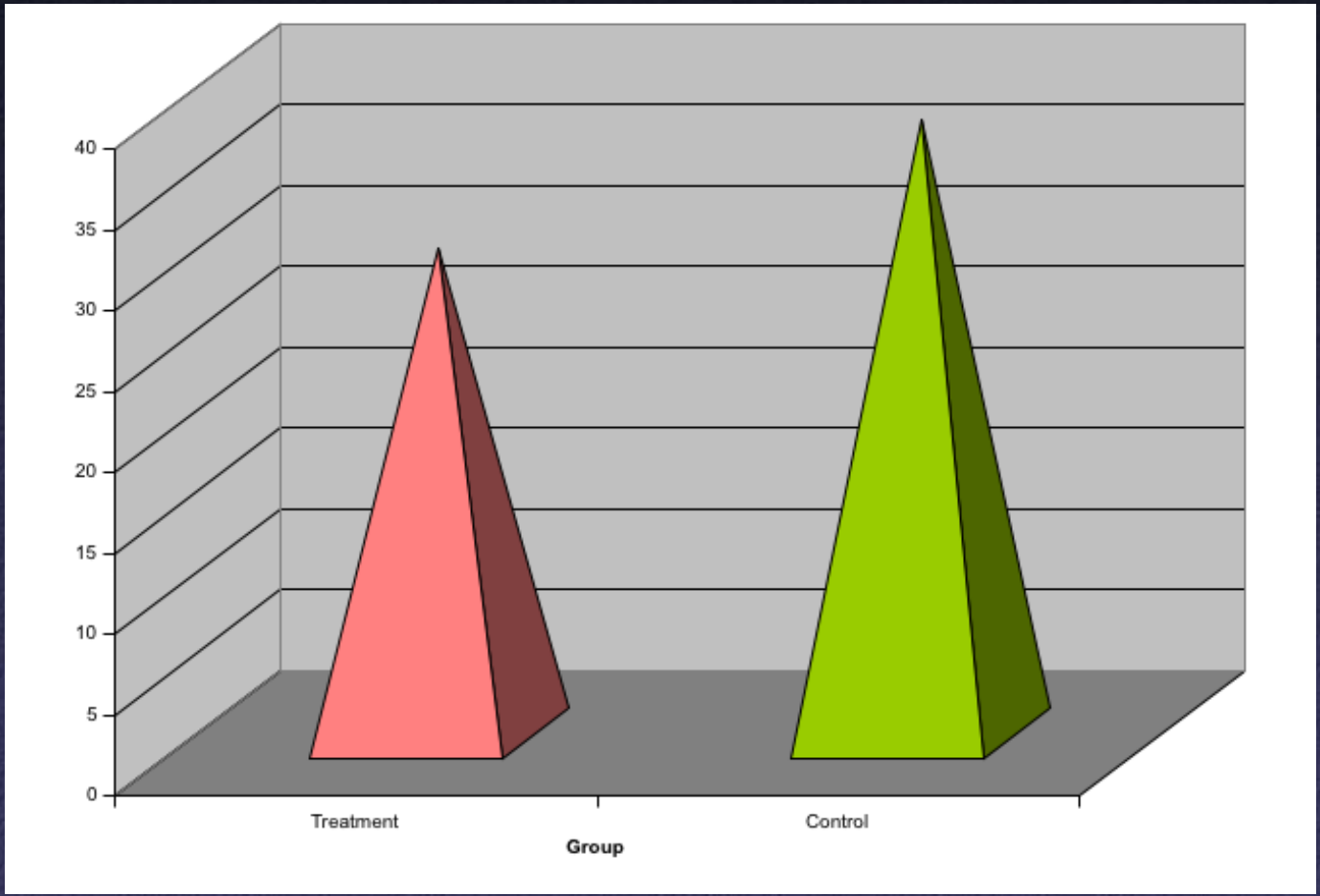
[Karl Broman]

Example 1



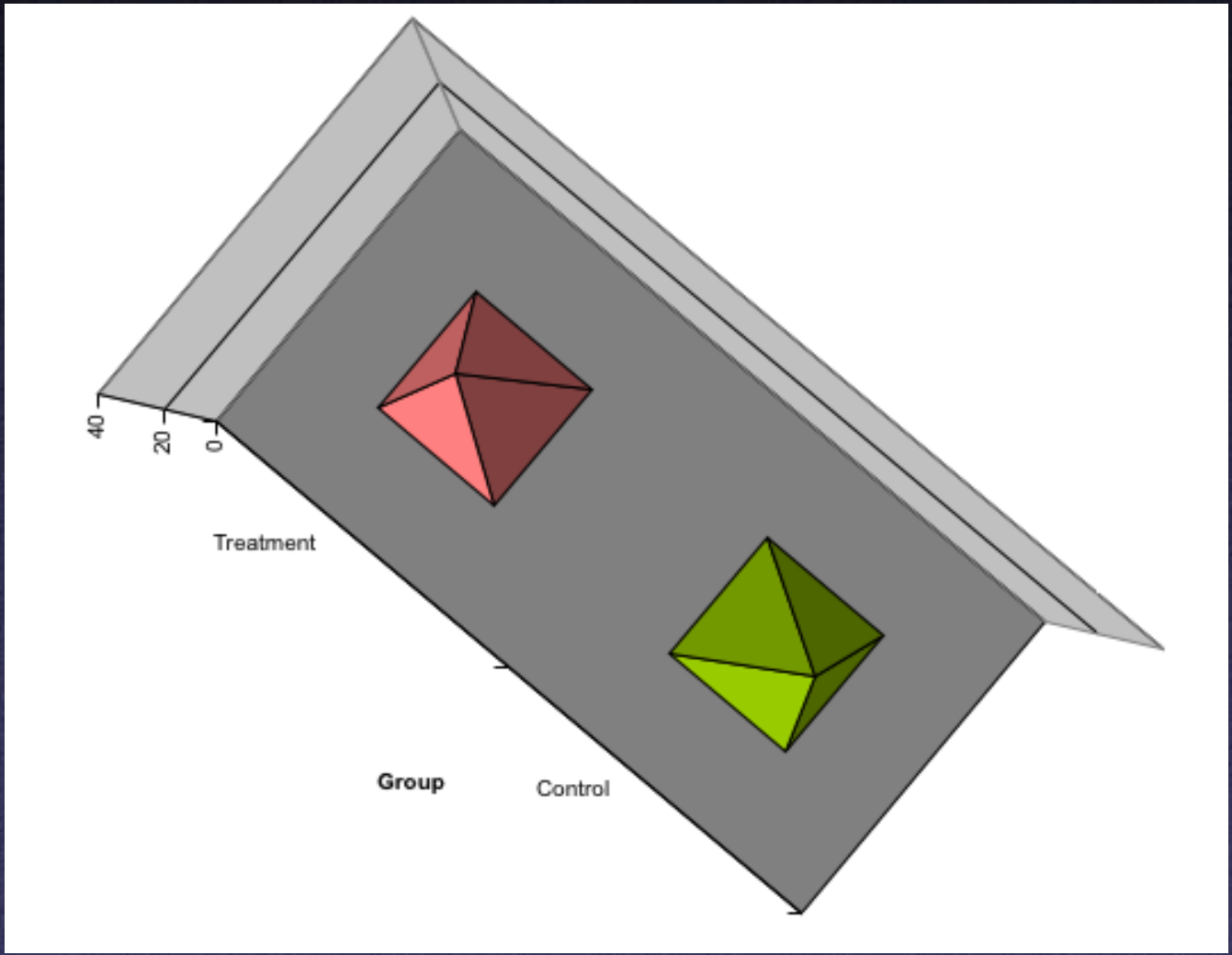
[Karl Broman]

Example 1



[Karl Broman]

Example 1



[Karl Broman]

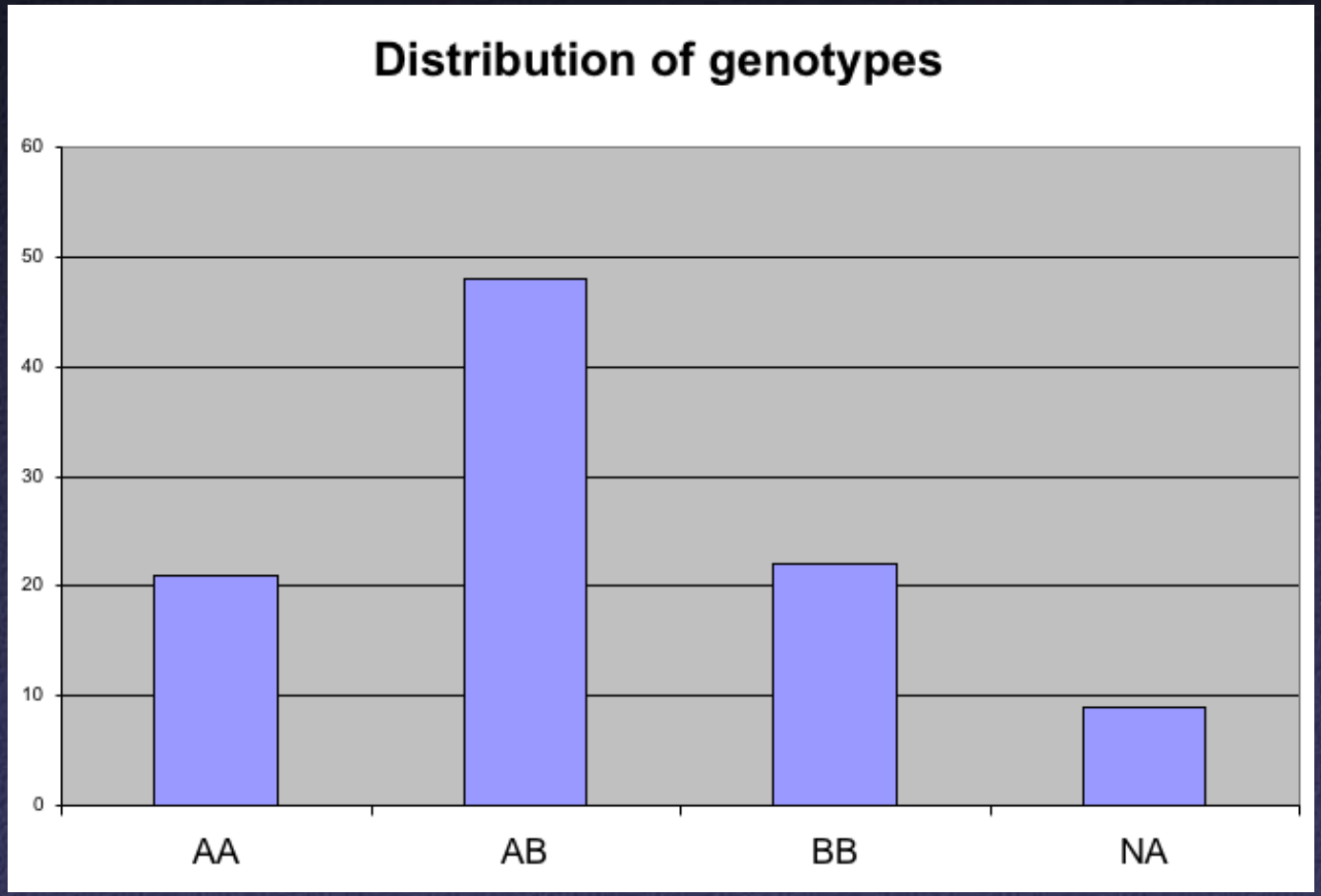
Example 2

Distribution of genotypes

AA	21%
AB	48%
BB	22%
missing	9%

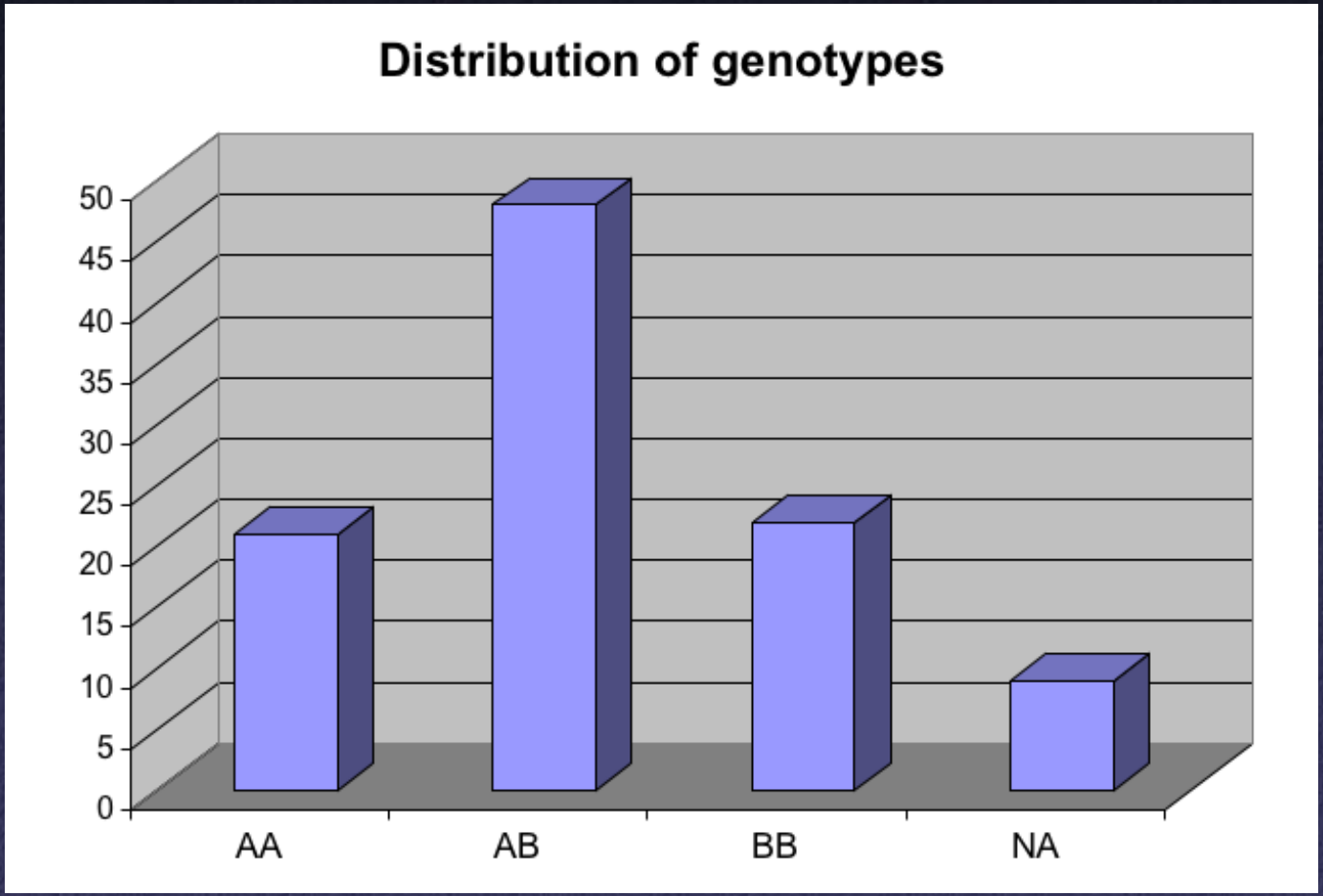
[Karl Broman]

Example 2



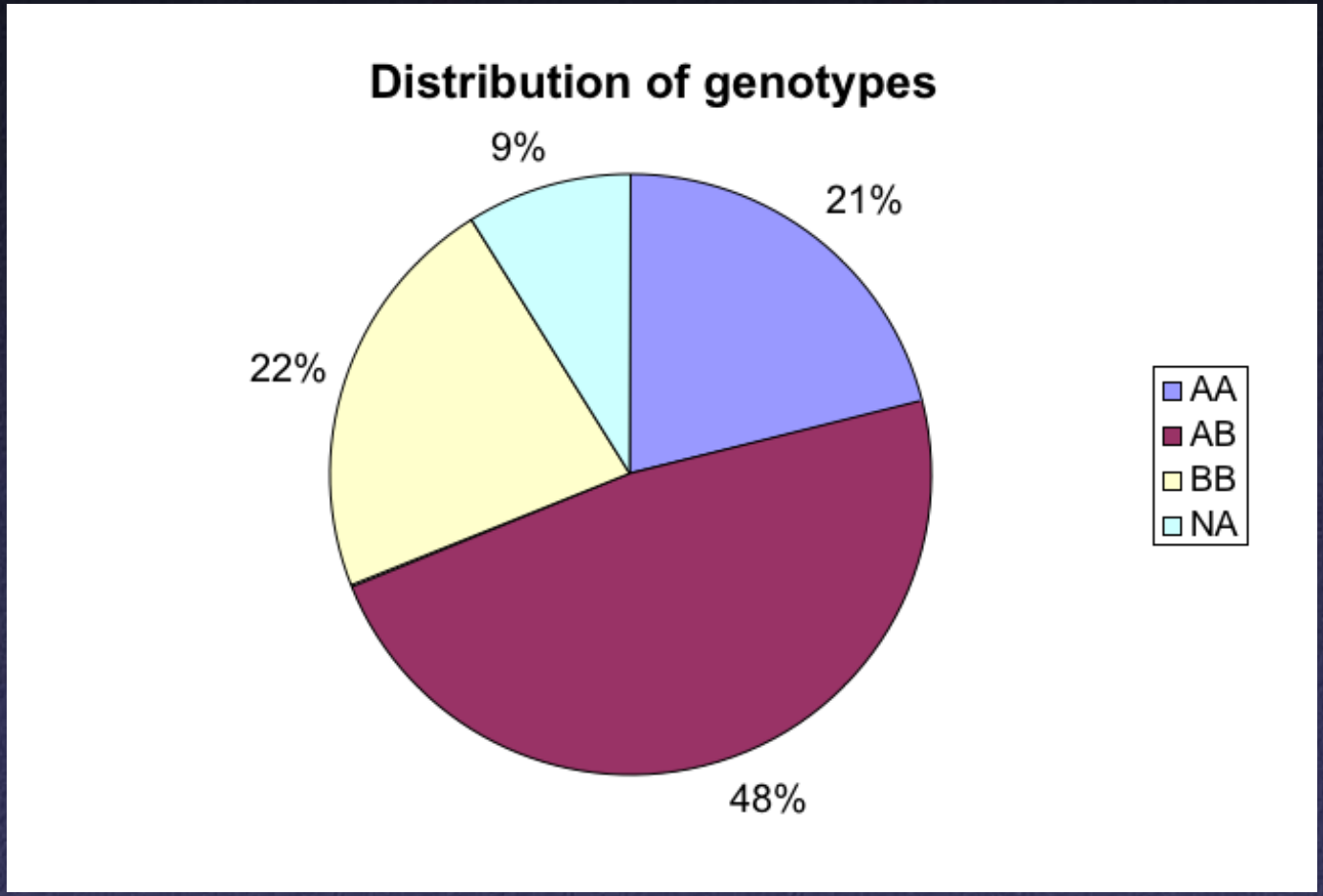
[Karl Broman]

Example 2



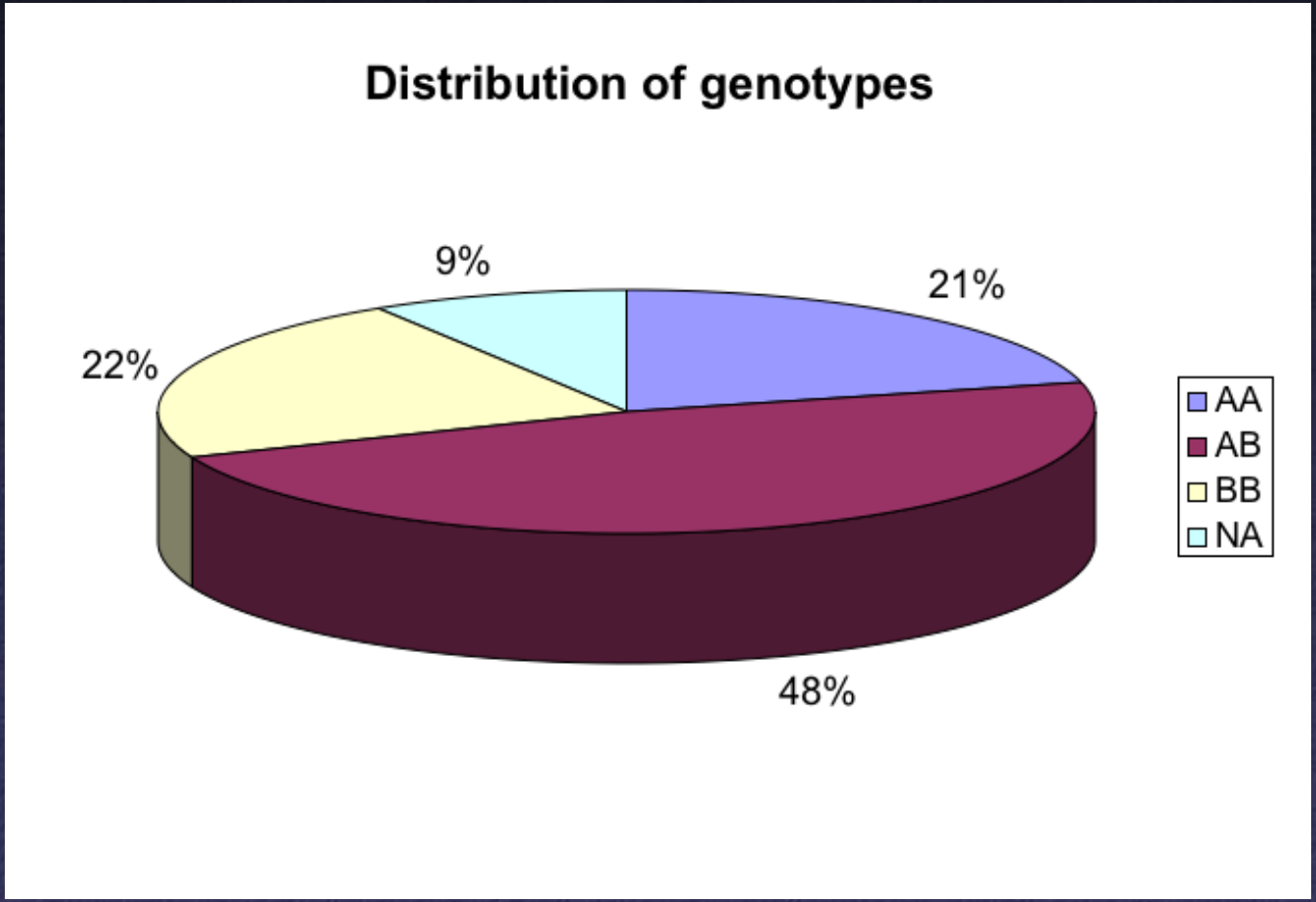
[Karl Broman]

Example 2



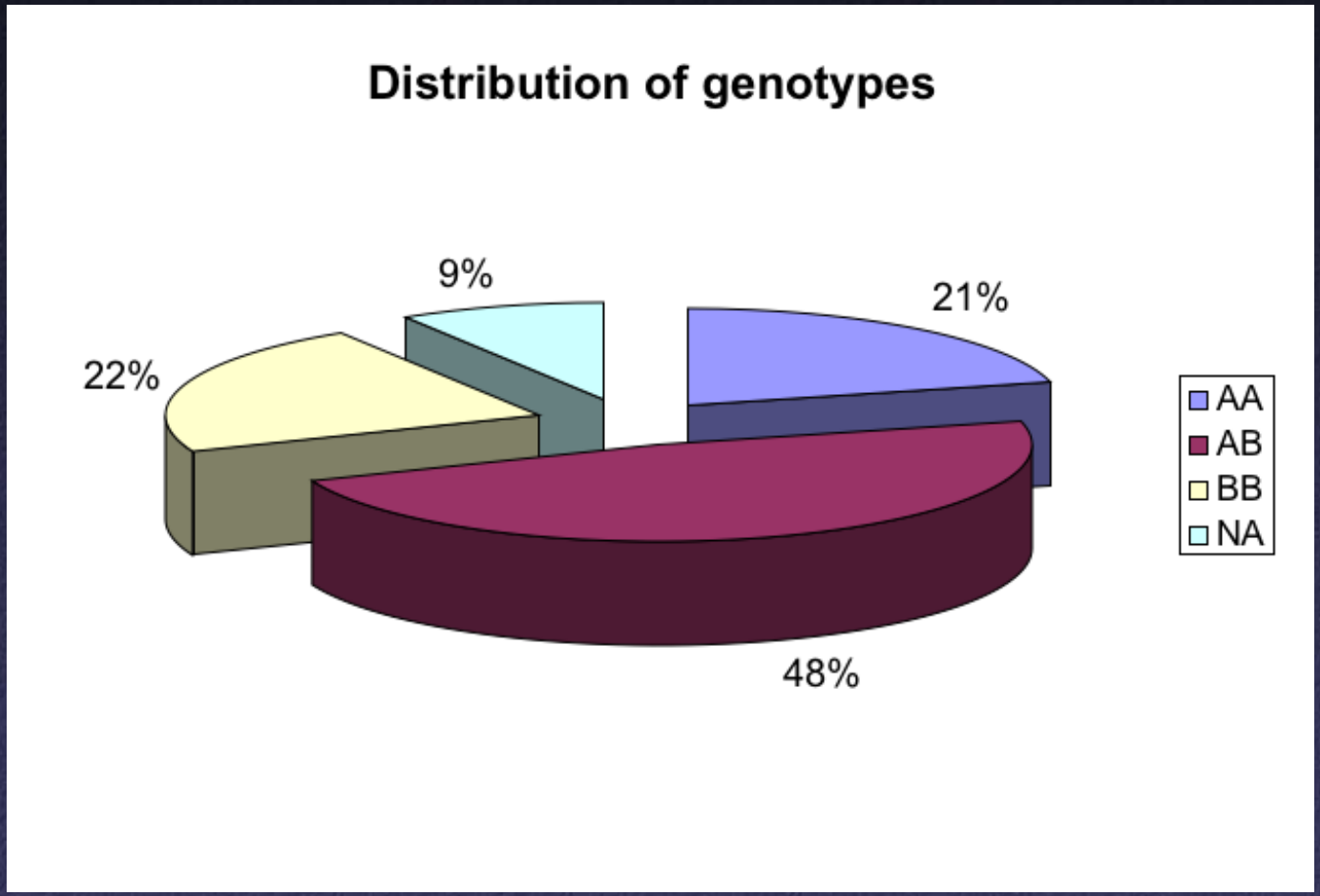
[Karl Broman]

Example 2



[Karl Broman]

Example 2



[Karl Broman]

Recent Example

AOL News, March 15, 2006