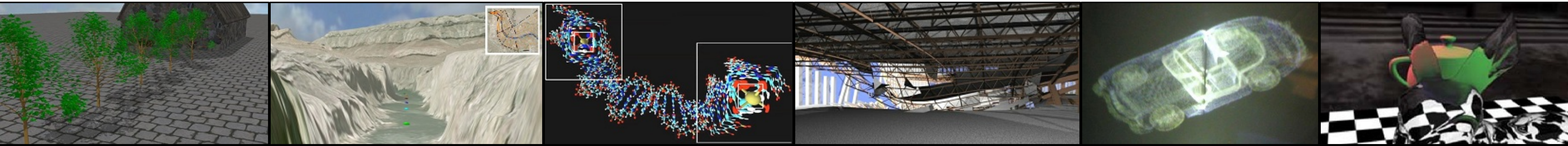
CIS 4930/6930-002
DATA VISUALIZATION

Visual Design

Paul Rosen
Assistant Professor
University of South Florida

slides credits Miriah Meyer (U of Utah), Hanspeter Pfister (Harvard), John Stasko (Georgia Tech), & Josh Levine (Clemson)



Reminders…

1/24/2017 – Project 1 Peer and Self Critiques due

1/29/2017 - Project #2 due

TODAY . . .

Four Levels of Visualization Design

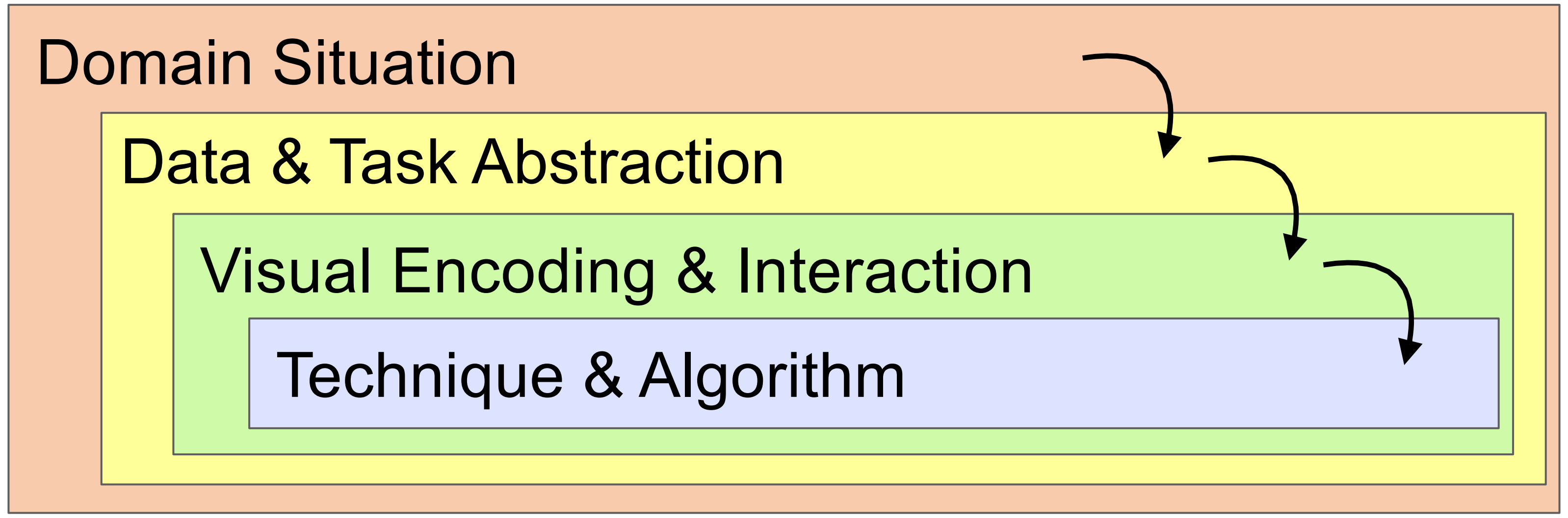
Tufte's Principles (Integrity & Design)

Critiques

THE FOUR LEVELS OF VISUALIZATION DESIGN

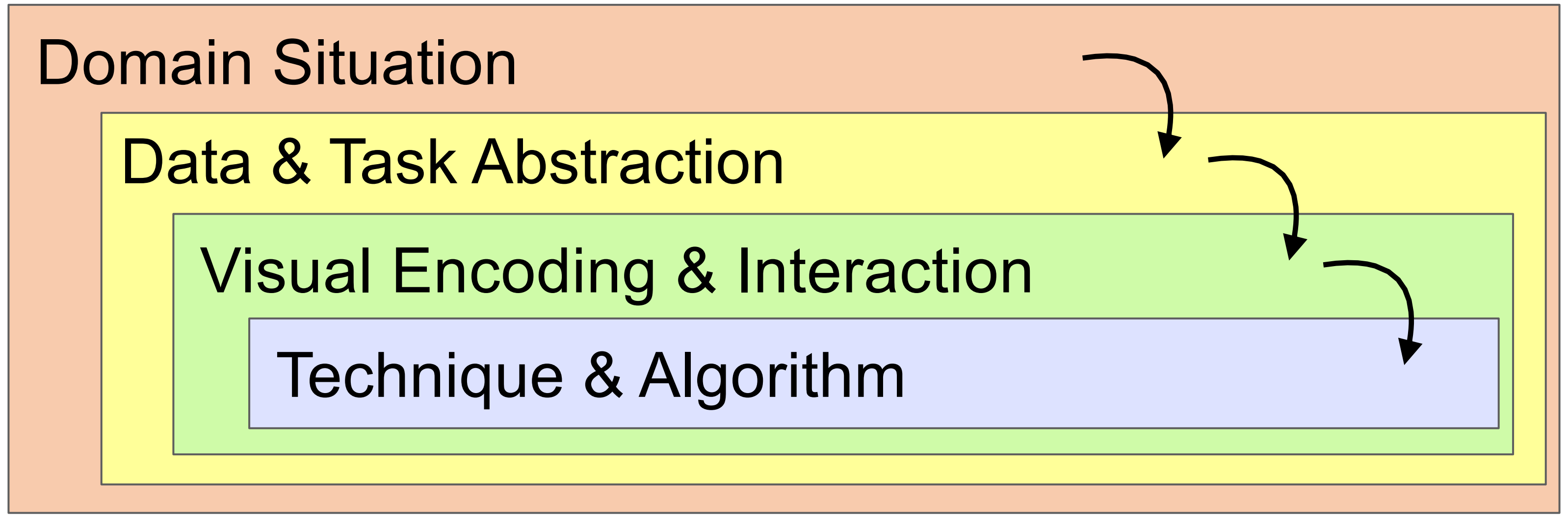
MUNZNER'S NESTED MODEL

design model—describes levels of design inherent to, and that should be considered in, the creation of a visualization



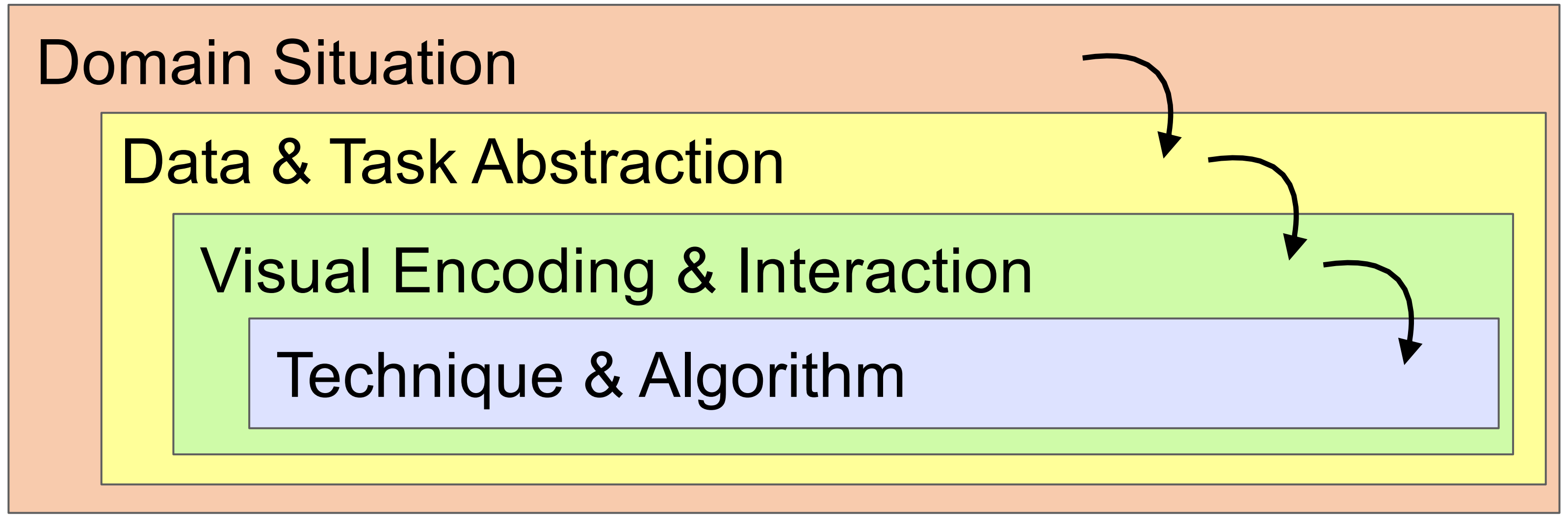
MUNZNER'S NESTED MODEL

domain situation—describing a group of target users, their domain of interest, their questions, and their data



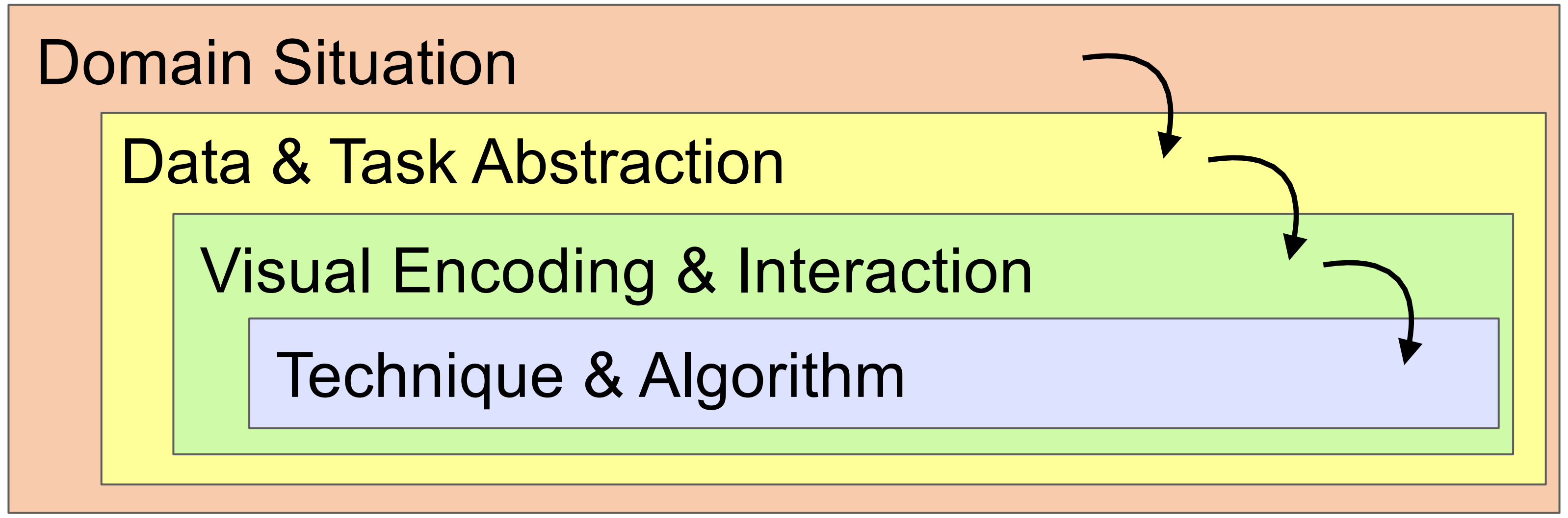
MUNZNER'S NESTED MODEL

data/task abstraction—abstracting the specific domain questions and data from the domain-specific form into a generic, computational form



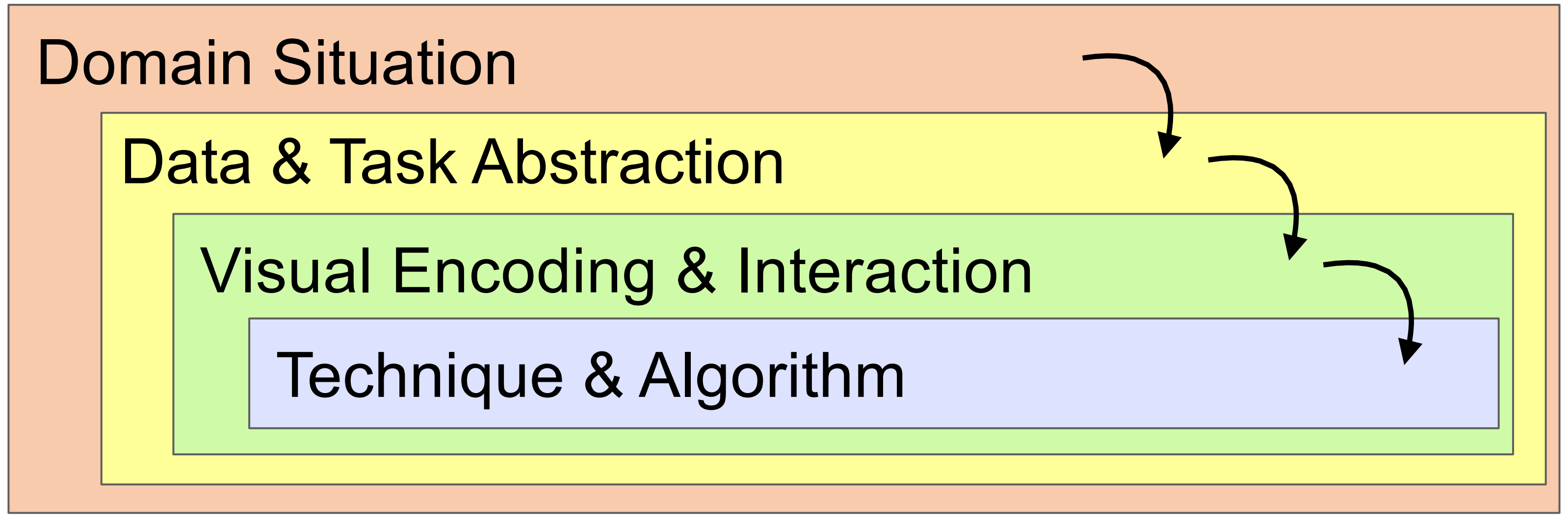
MUNZNER'S NESTED MODEL

visual encoding & interaction—decide on the specific way to create and manipulate the visual representation of the abstraction

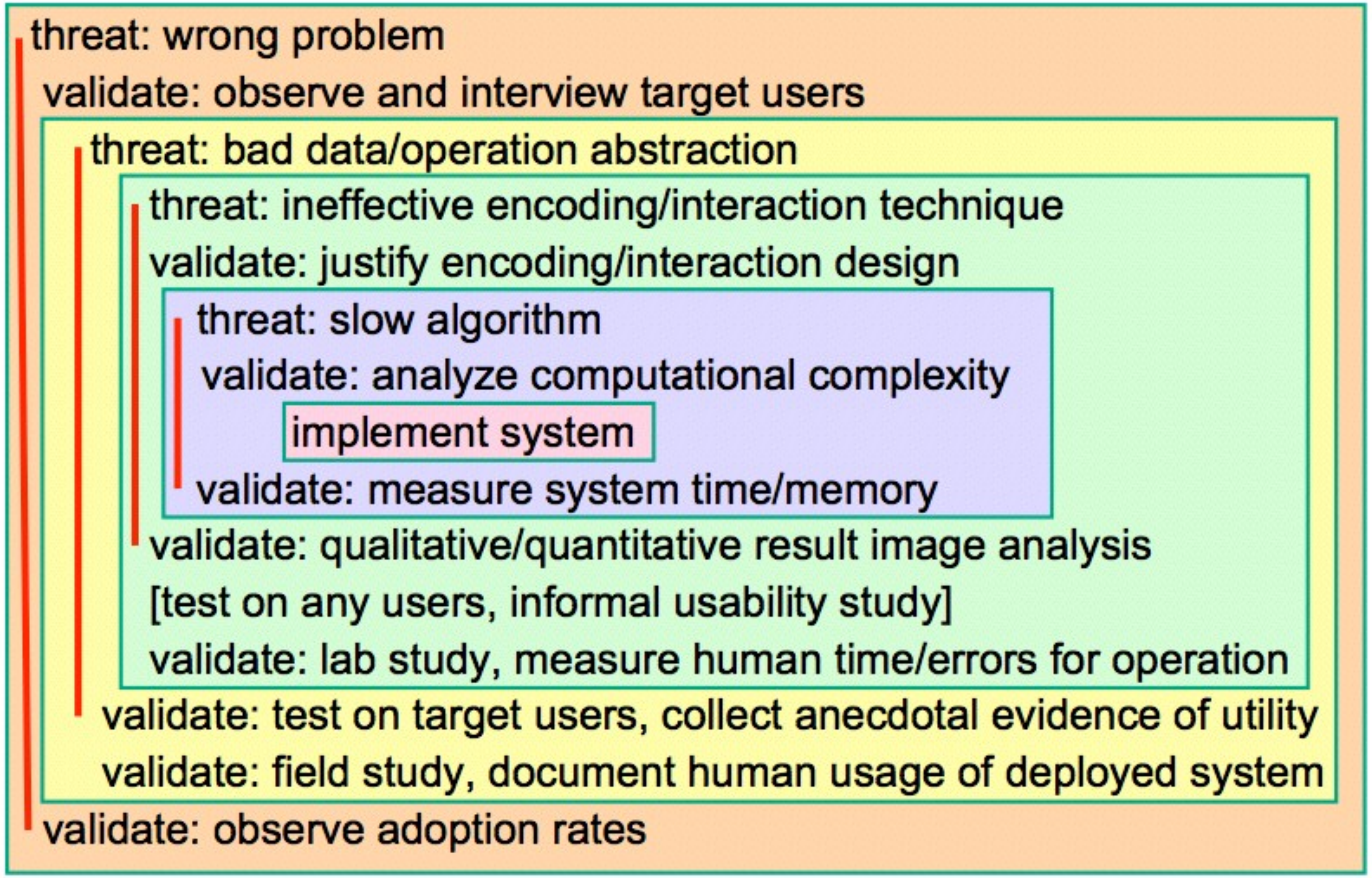


MUNZNER'S NESTED MODEL

algorithm—crafting a detailed procedure that allows a computer to automatically and efficiently carry out the desired visualization goal

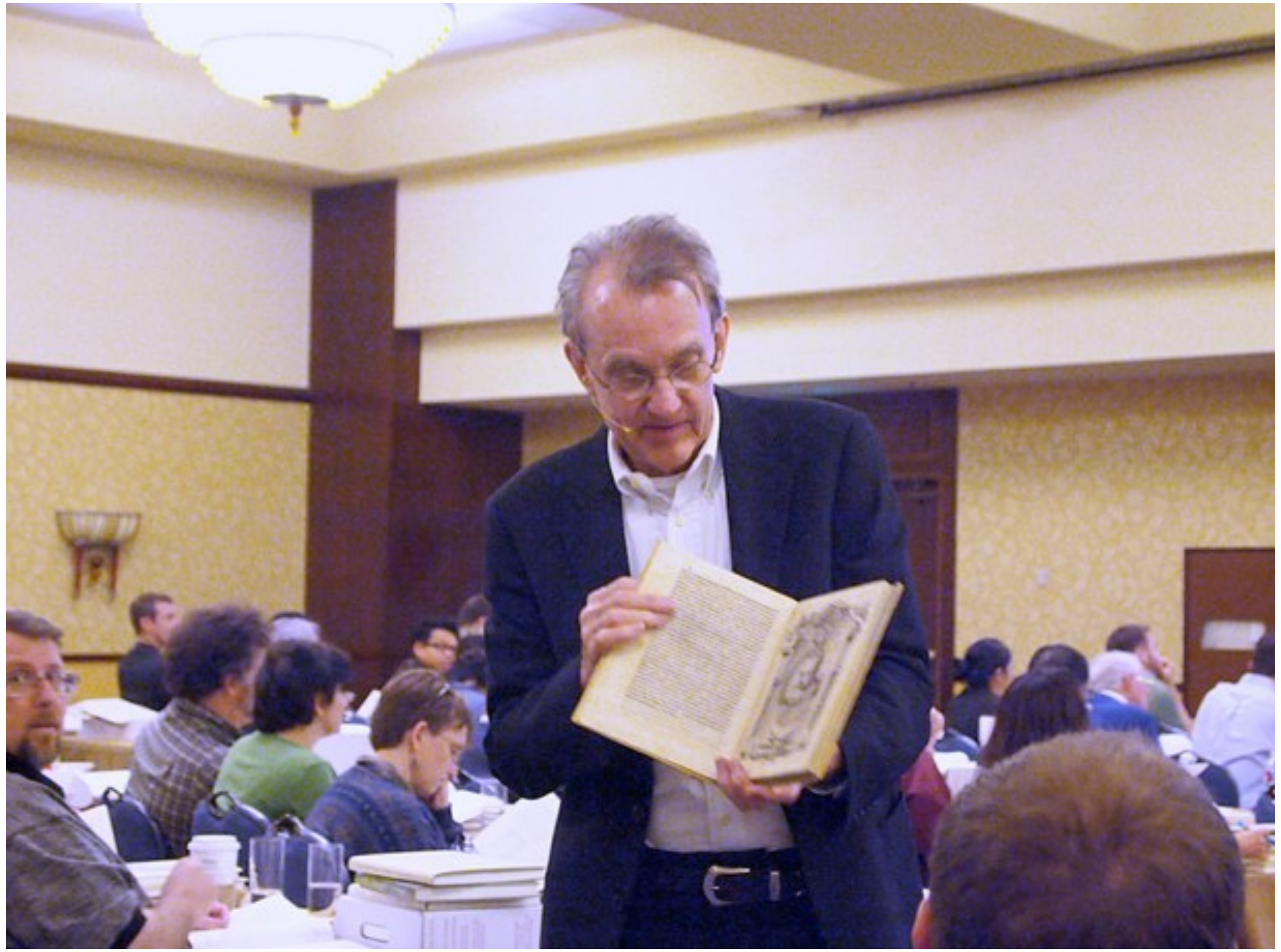


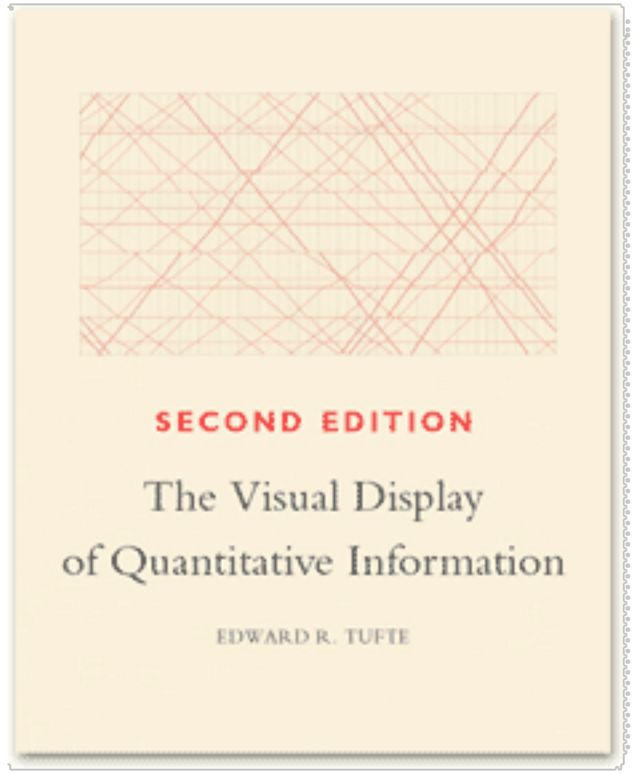
MUNZNER'S NESTED MODEL

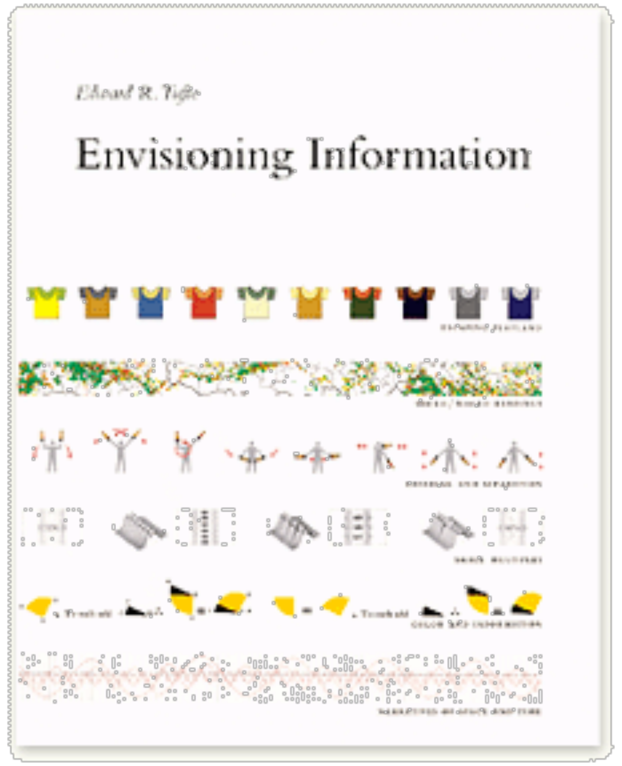


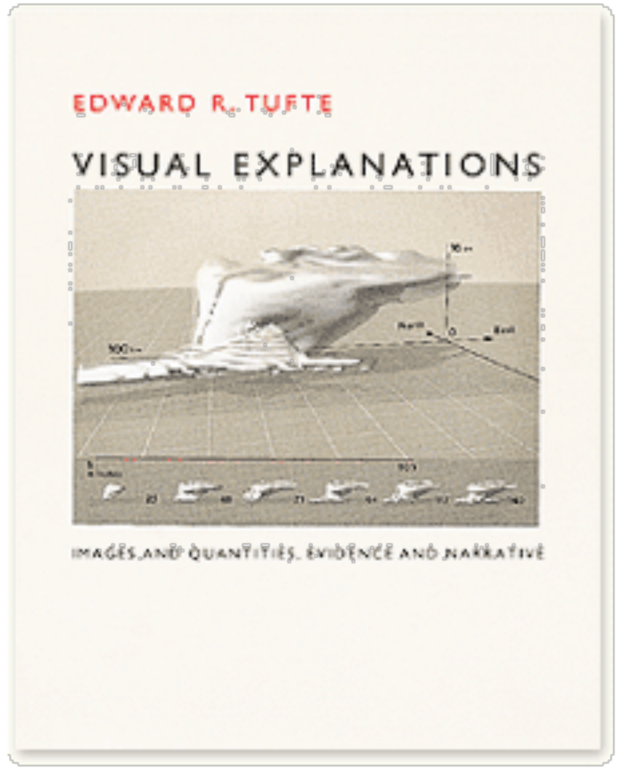
TUFTE

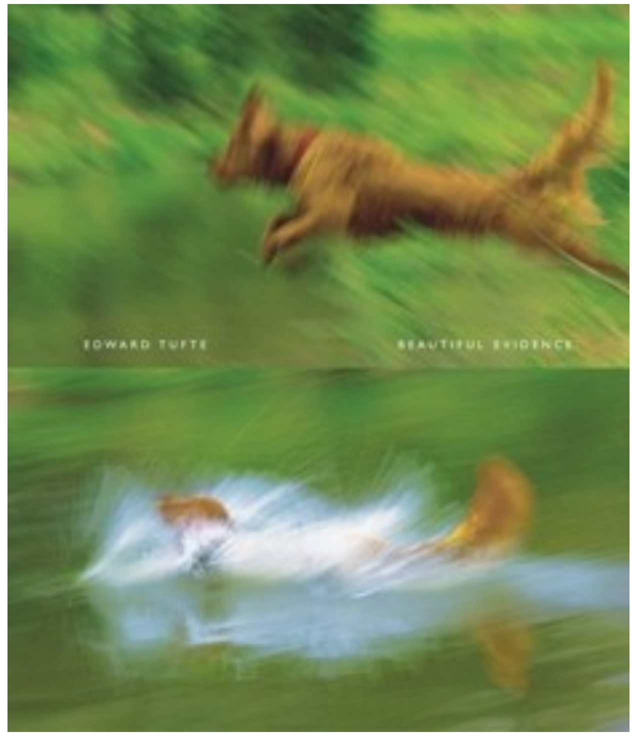
design excellence




SECOND EDITION
The Visual Display
of Quantitative Information
EDWARD R. TUFTE


Edward R. Tufte
Envisioning Information


EDWARD R. TUFTE
VISUAL EXPLANATIONS
IMAGES AND QUANTITIES, EVIDENCE AND NARRATIVE


EDWARD TUFTE
BEAUTIFUL EVIDENCE

TUFTE'S LESSONS

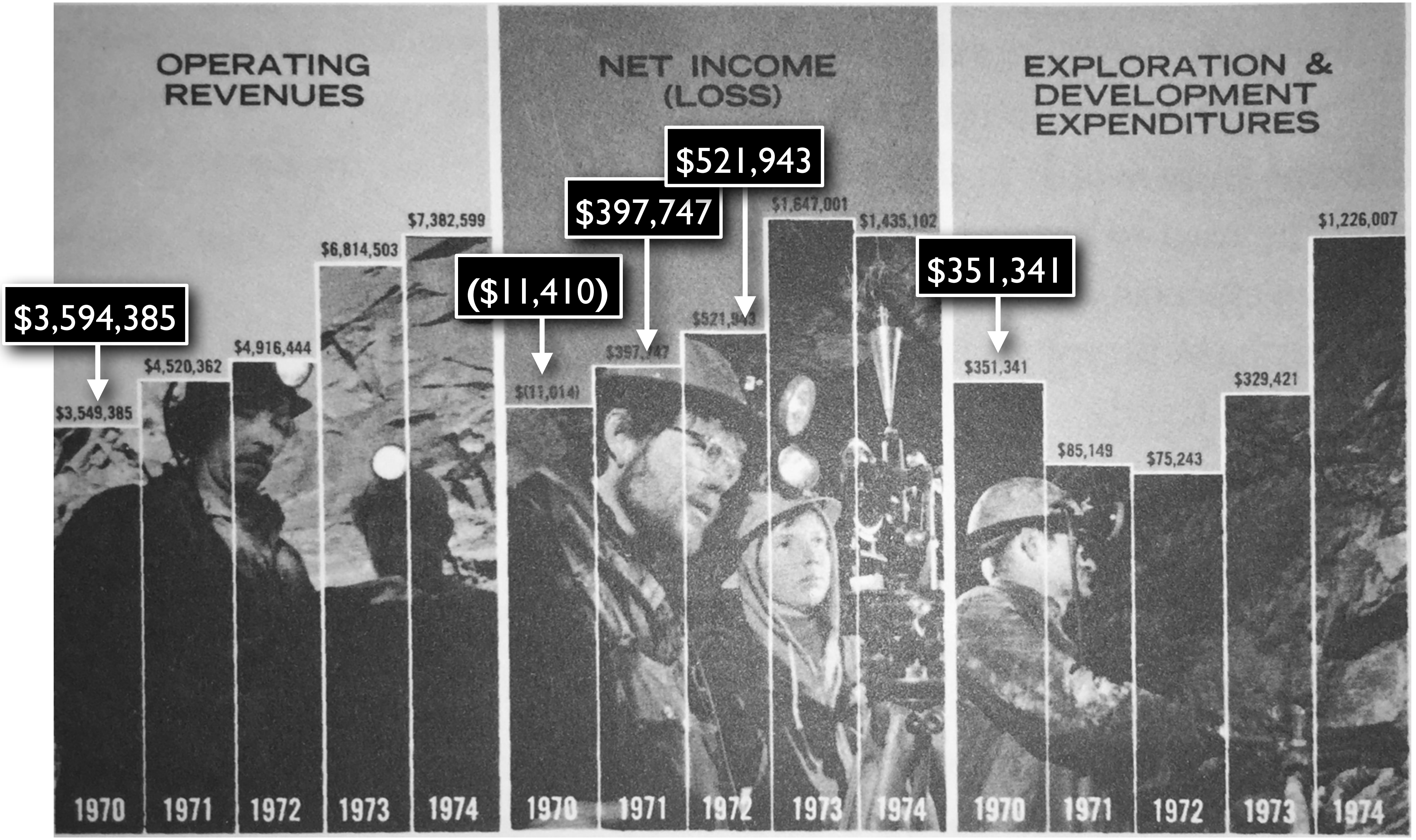
practice—graphical integrity and excellence

theory—design principles for data graphics

GRAPHICAL INTEGRITY

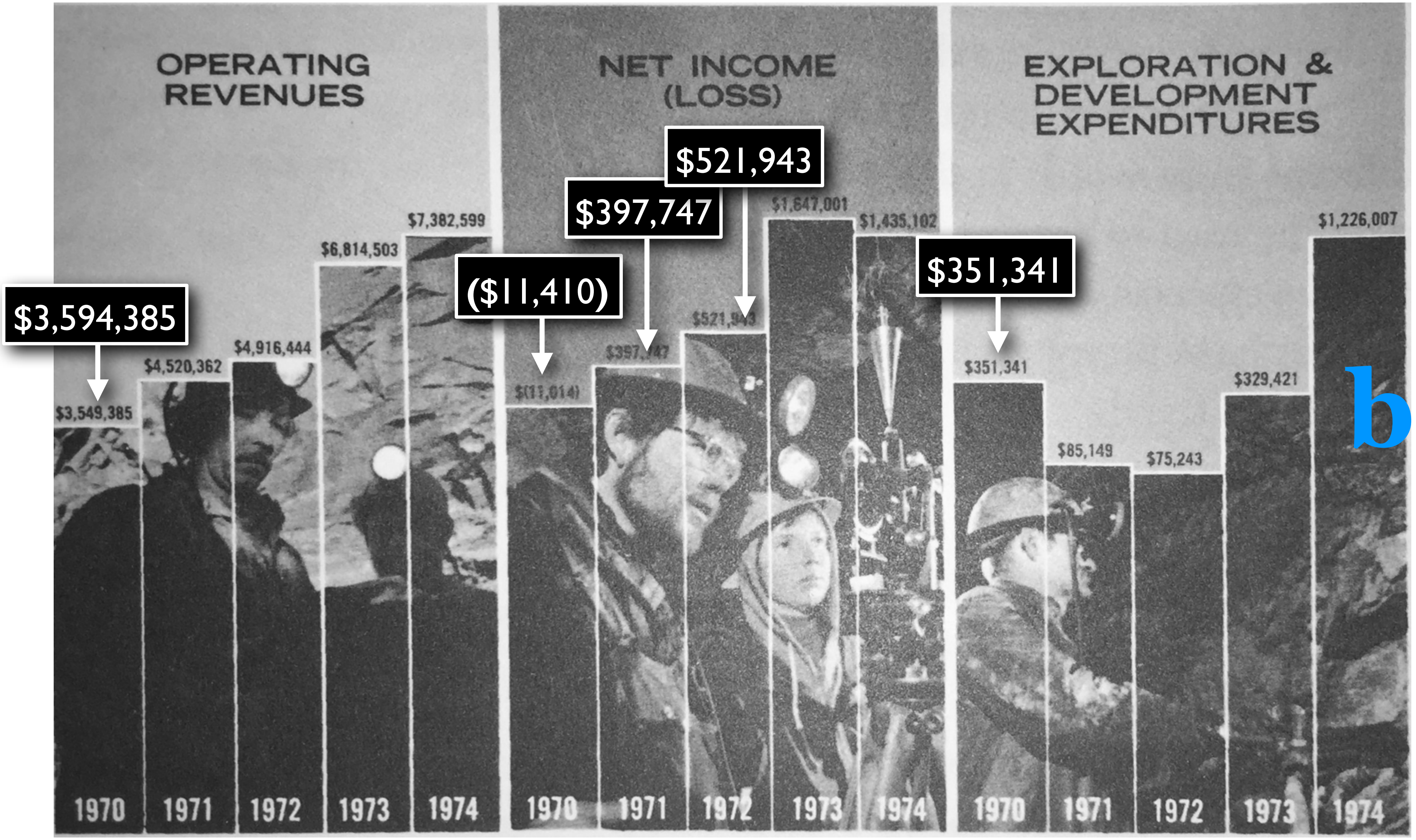
clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity

MISSING SCALES



TUFTE 2001

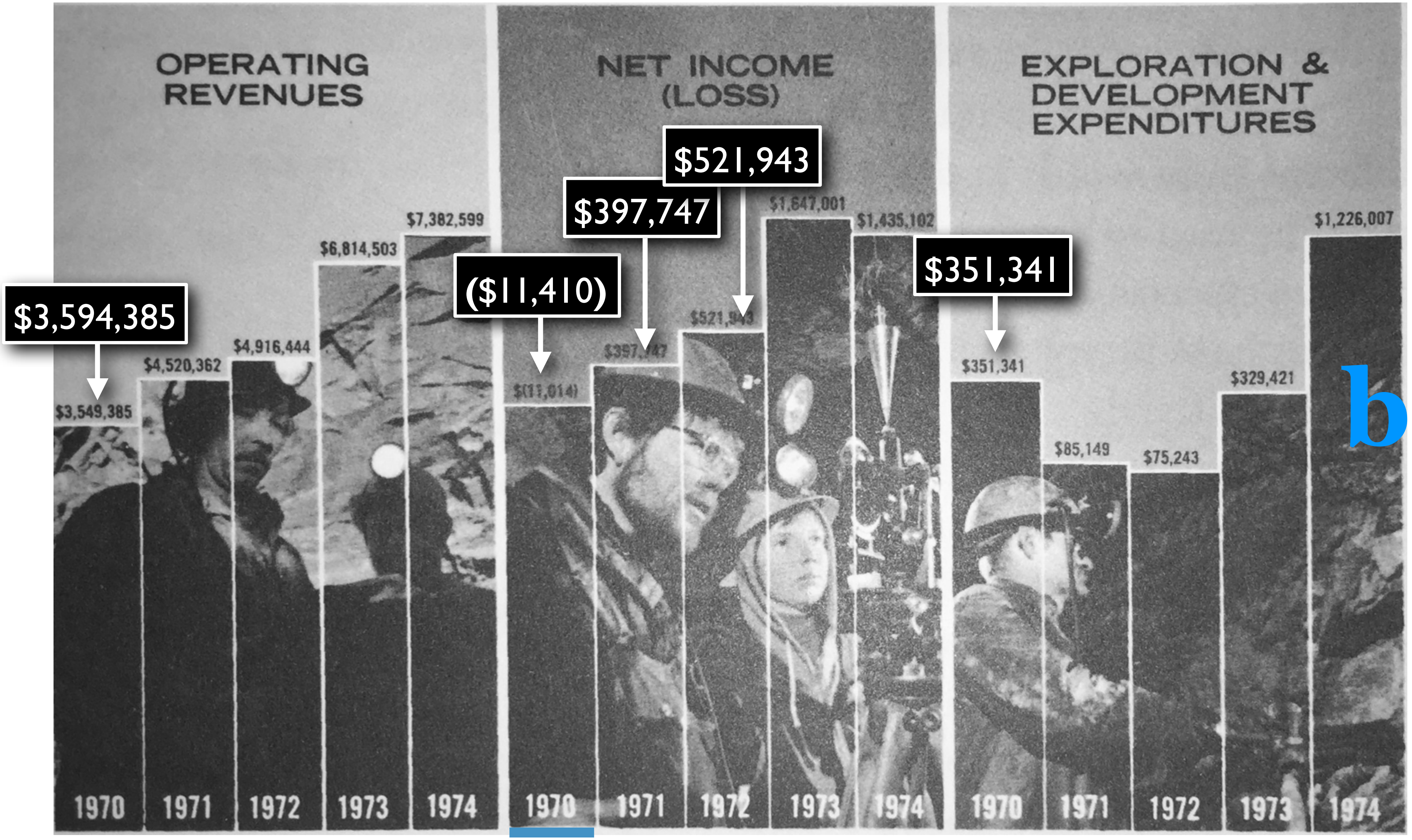
MISSING SCALES



baseline?

TUFTE 2001

MISSING SCALES

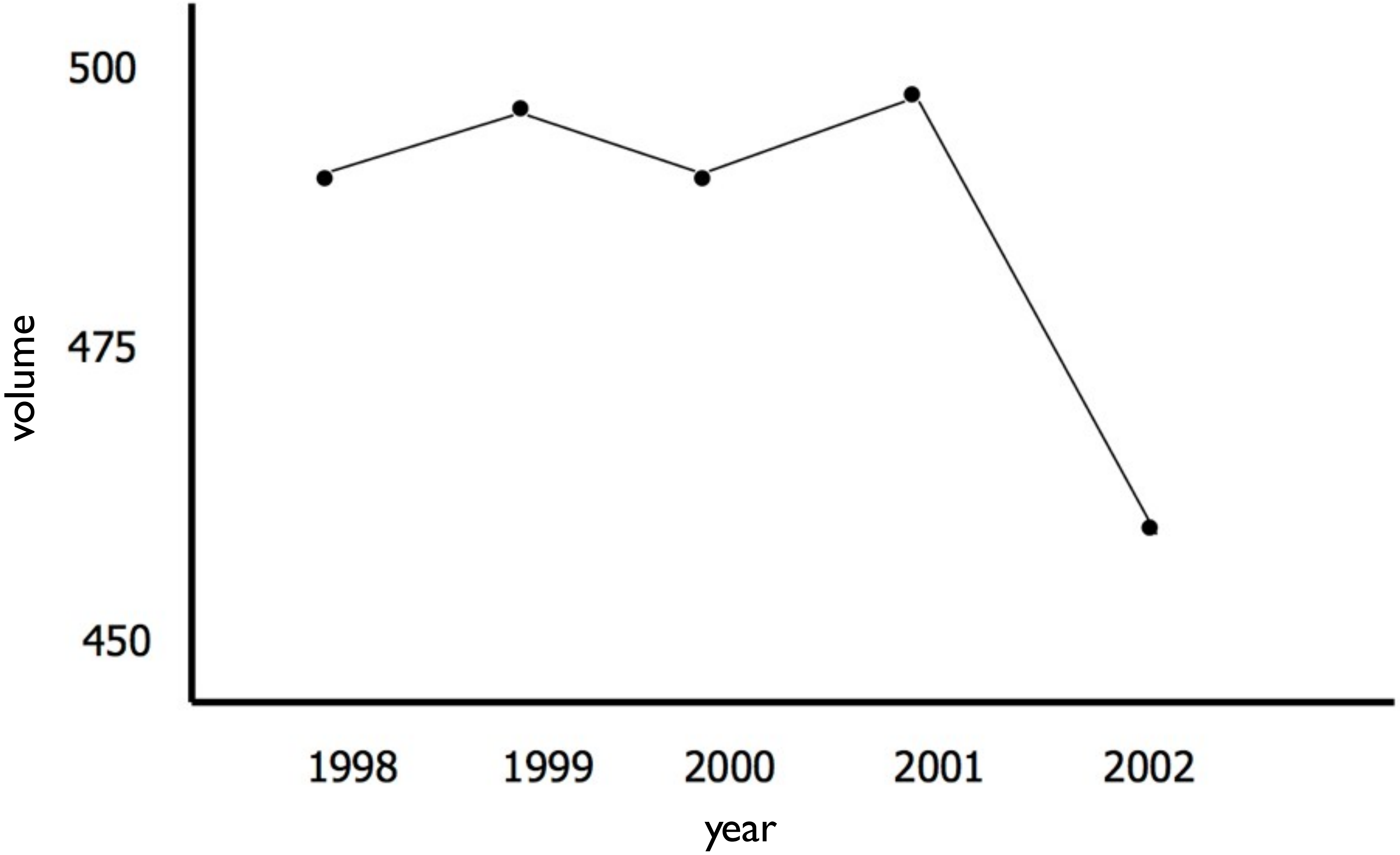


baseline?

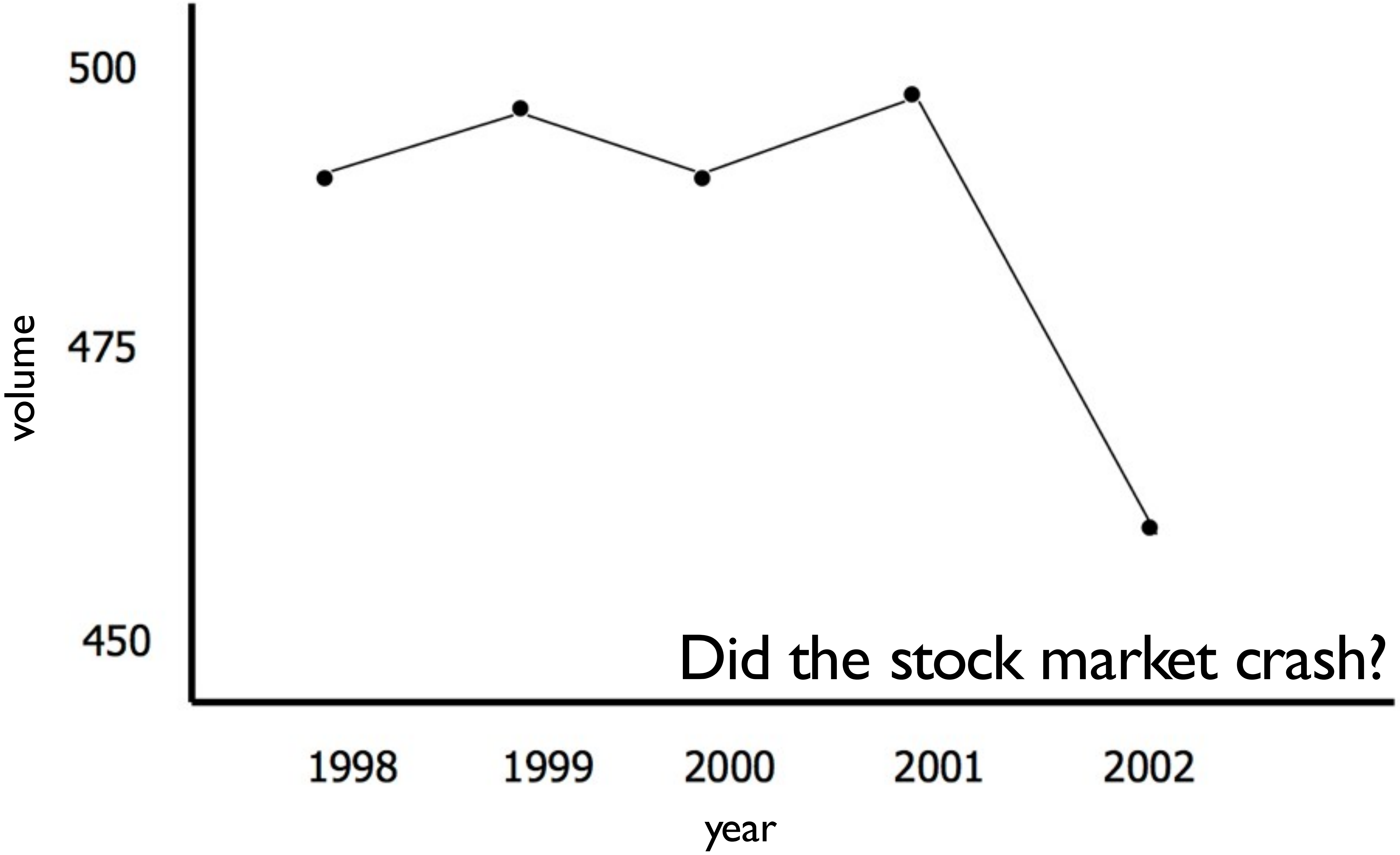
-$4,200,000

TUFTE 2001

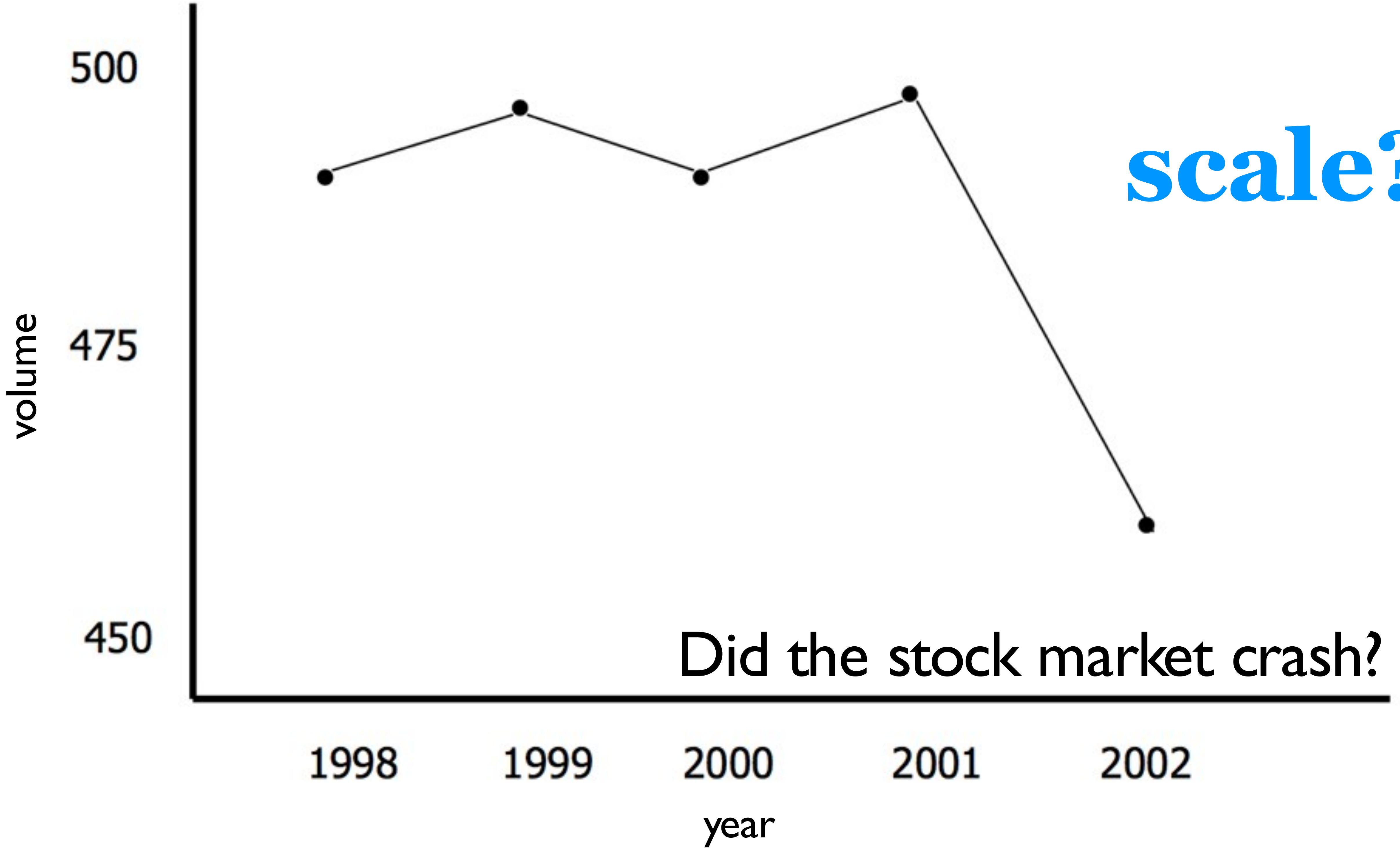
SCALE DISTORTION



SCALE DISTORTION



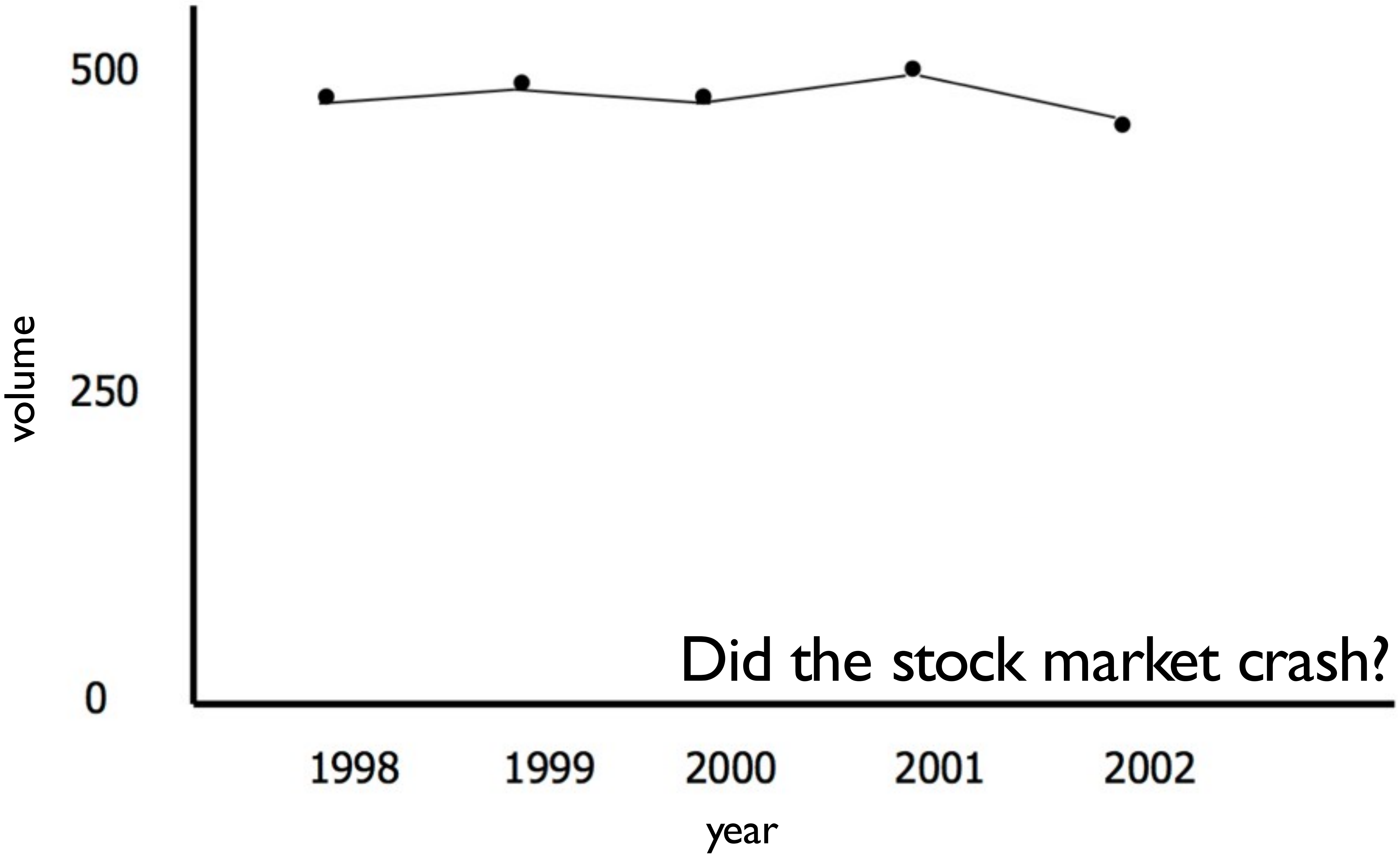
SCALE DISTORTION



scale?

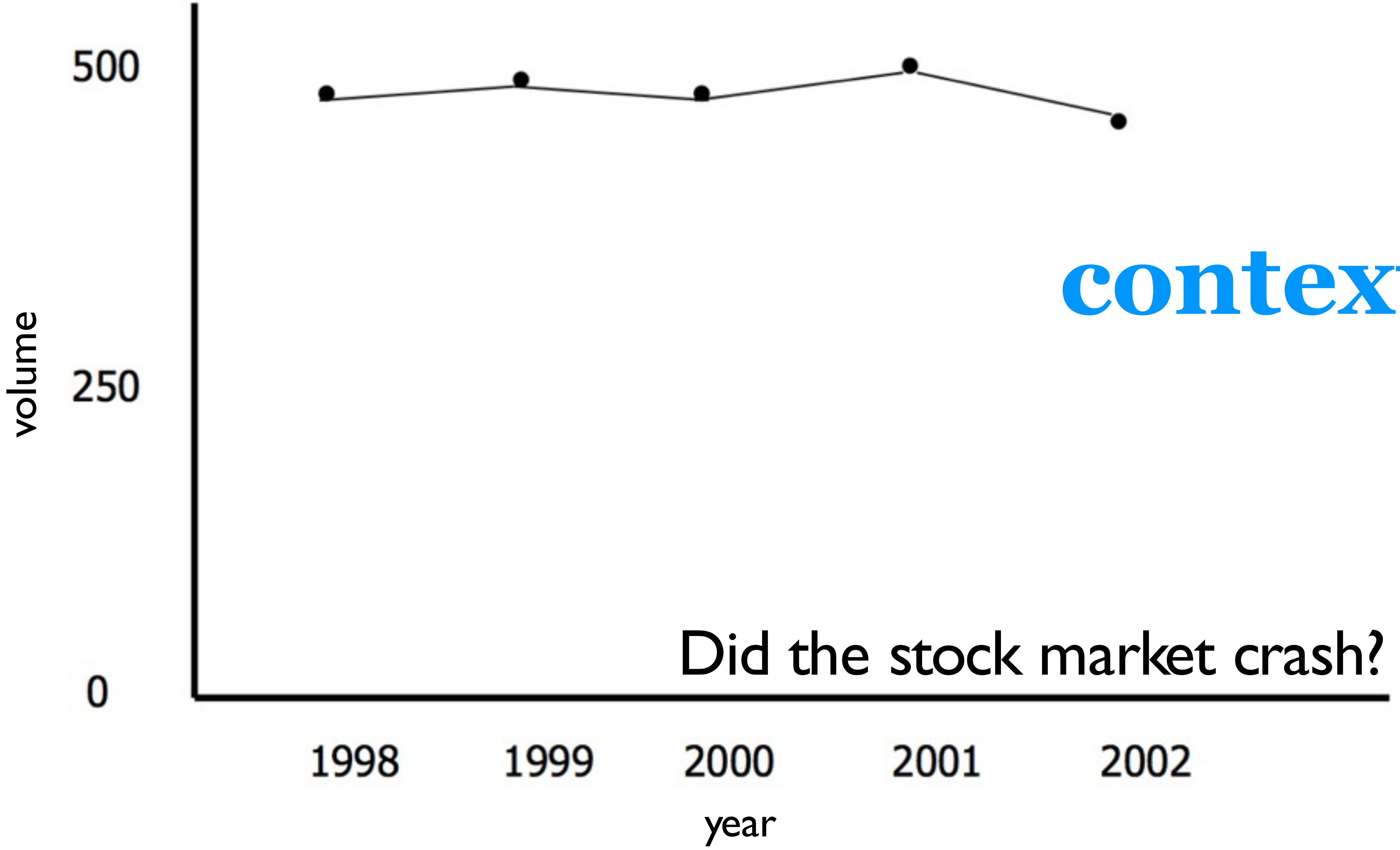
Did the stock market crash?

SCALE DISTORTION



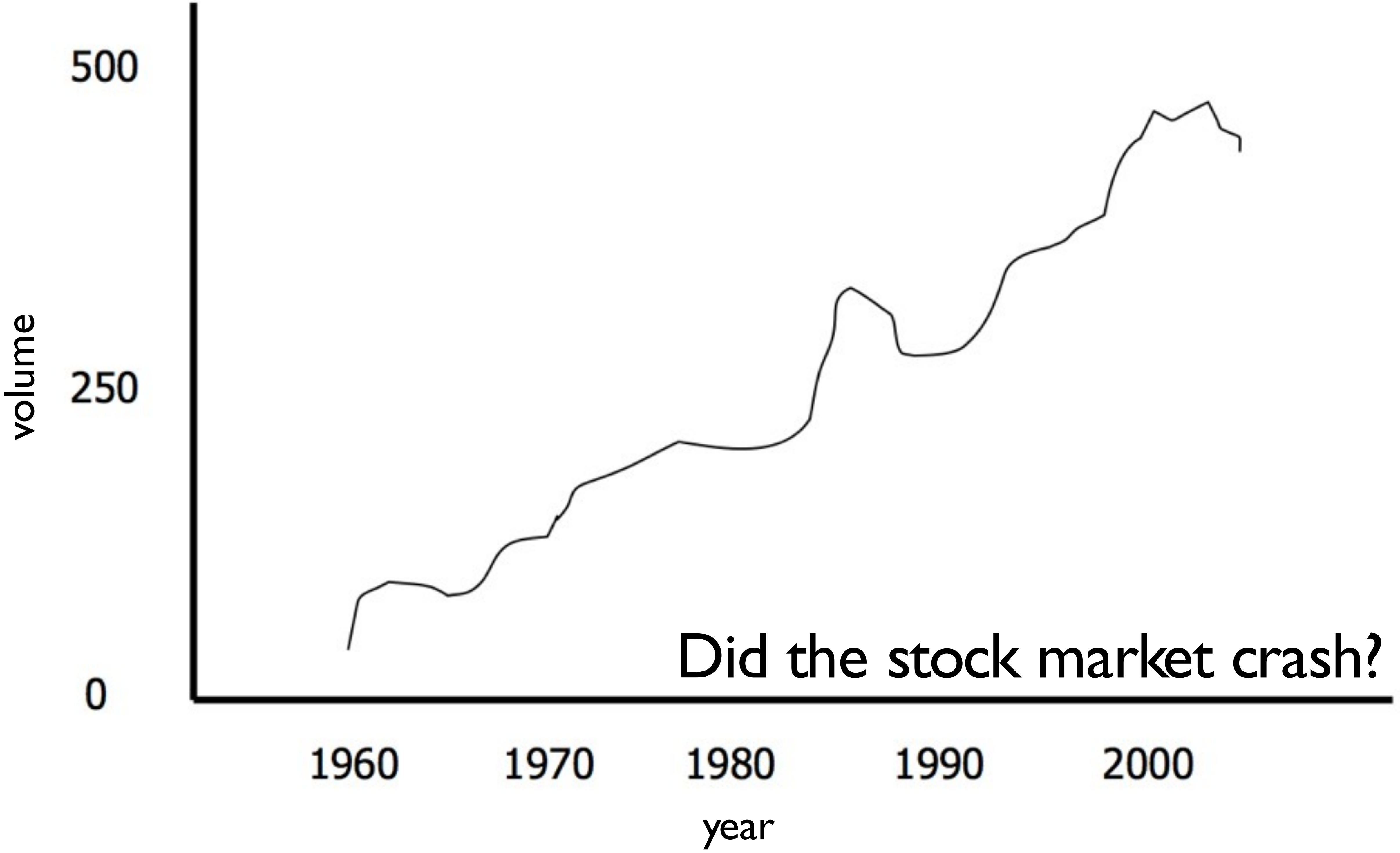
Did the stock market crash?

SCALE DISTORTION

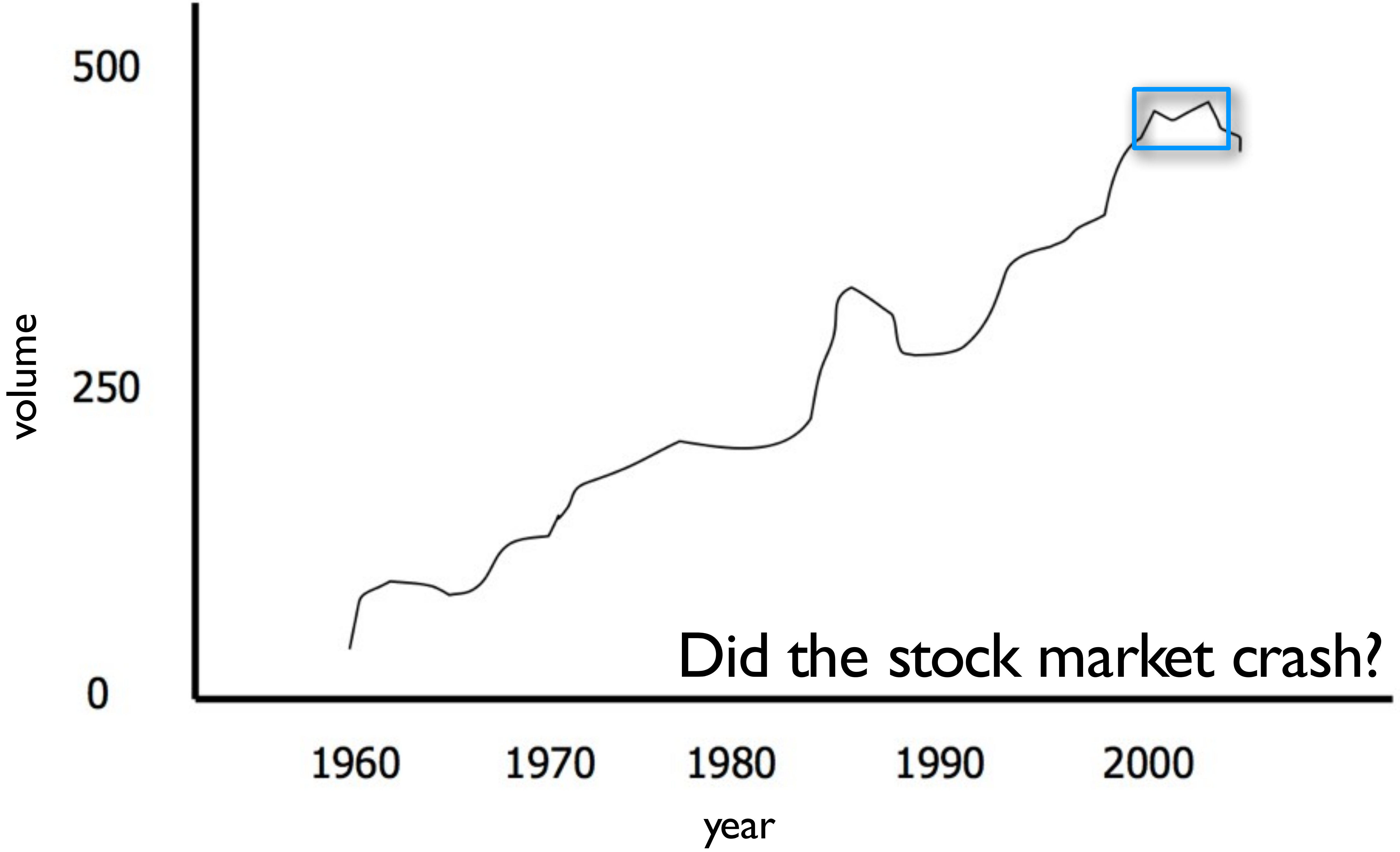


context?

SCALE DISTORTION



SCALE DISTORTION

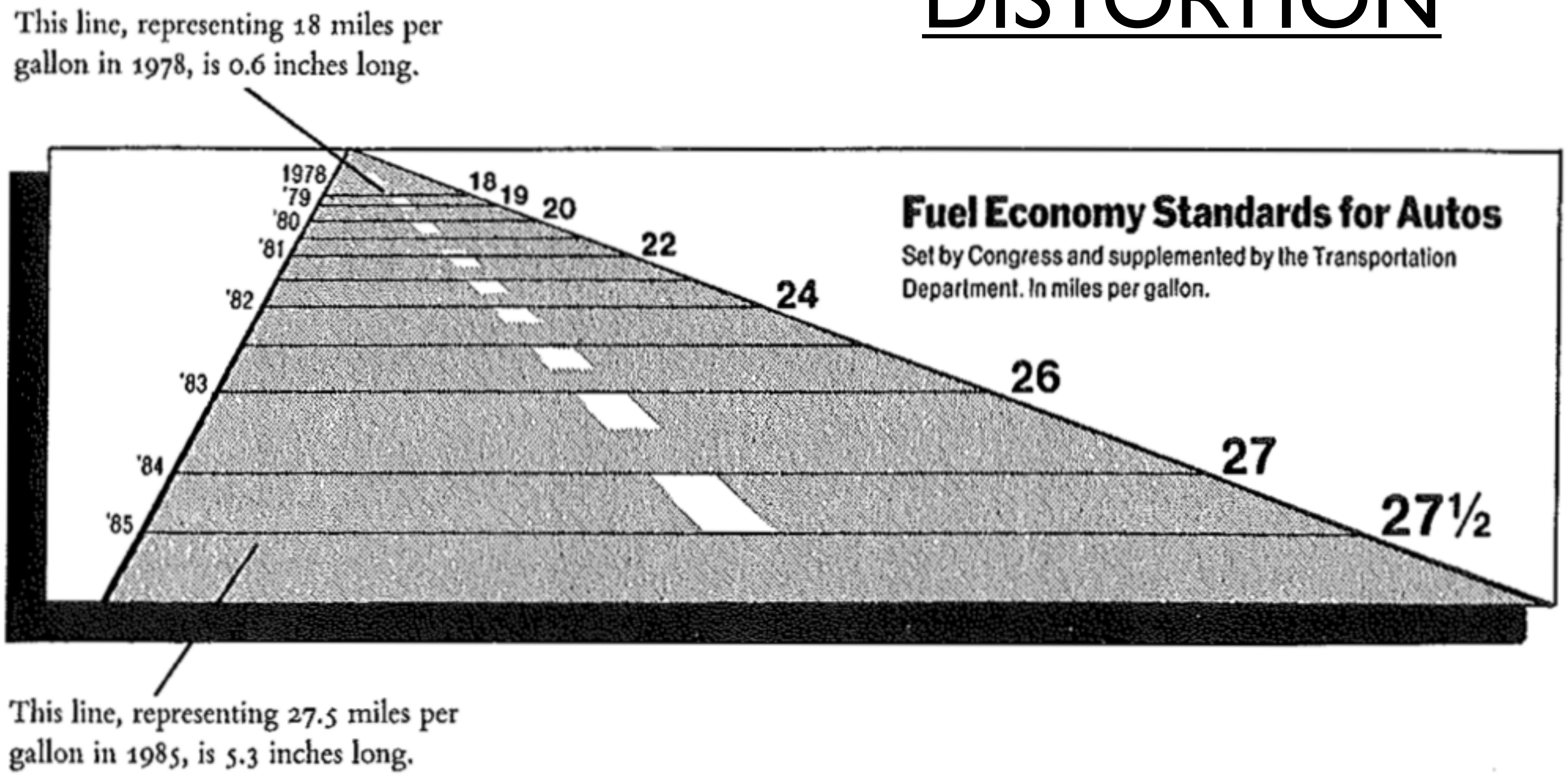


TUFTE'S INTEGRITY PRINCIPLES

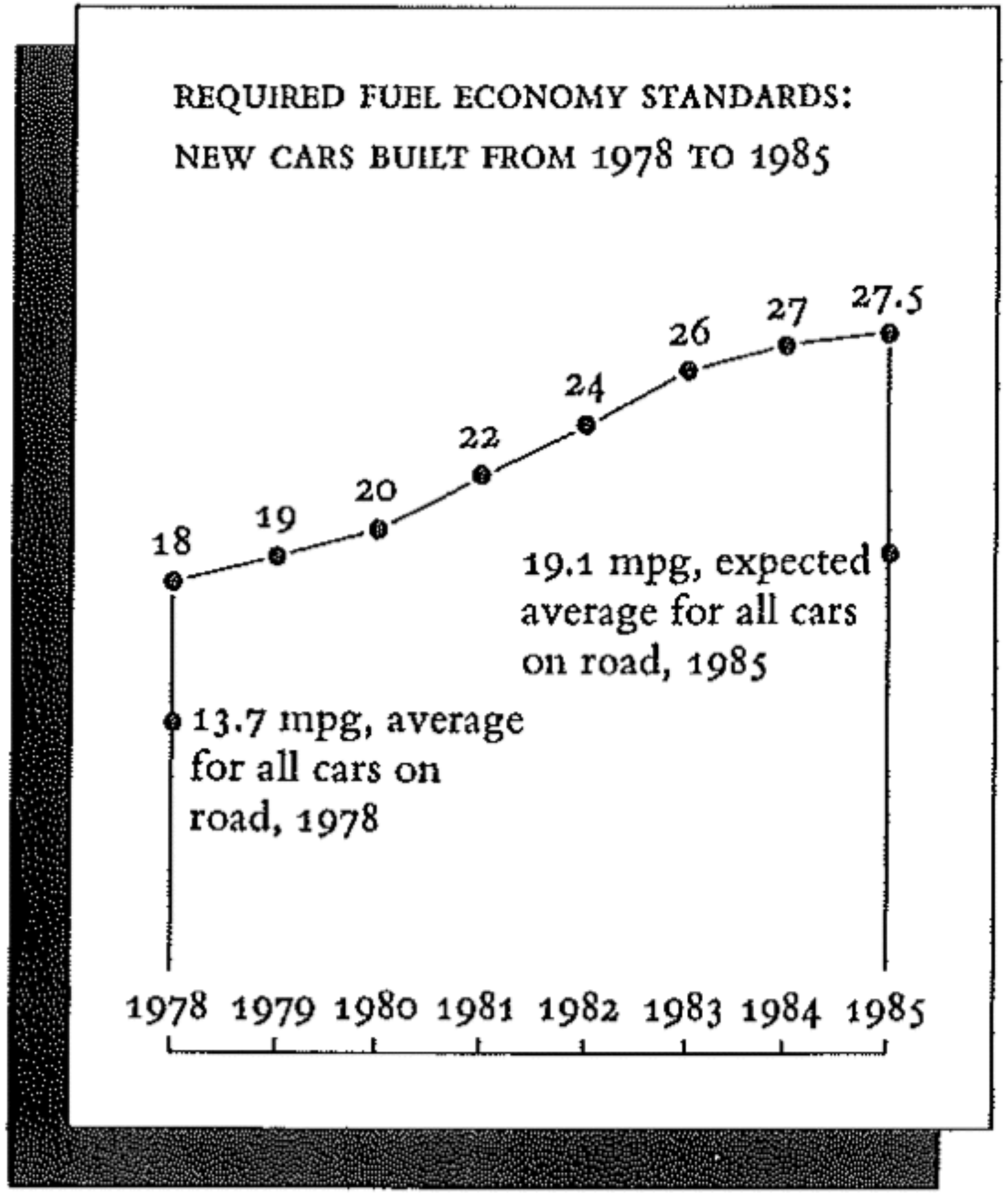
the representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.

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

DISTORTION



TUFTE 2001

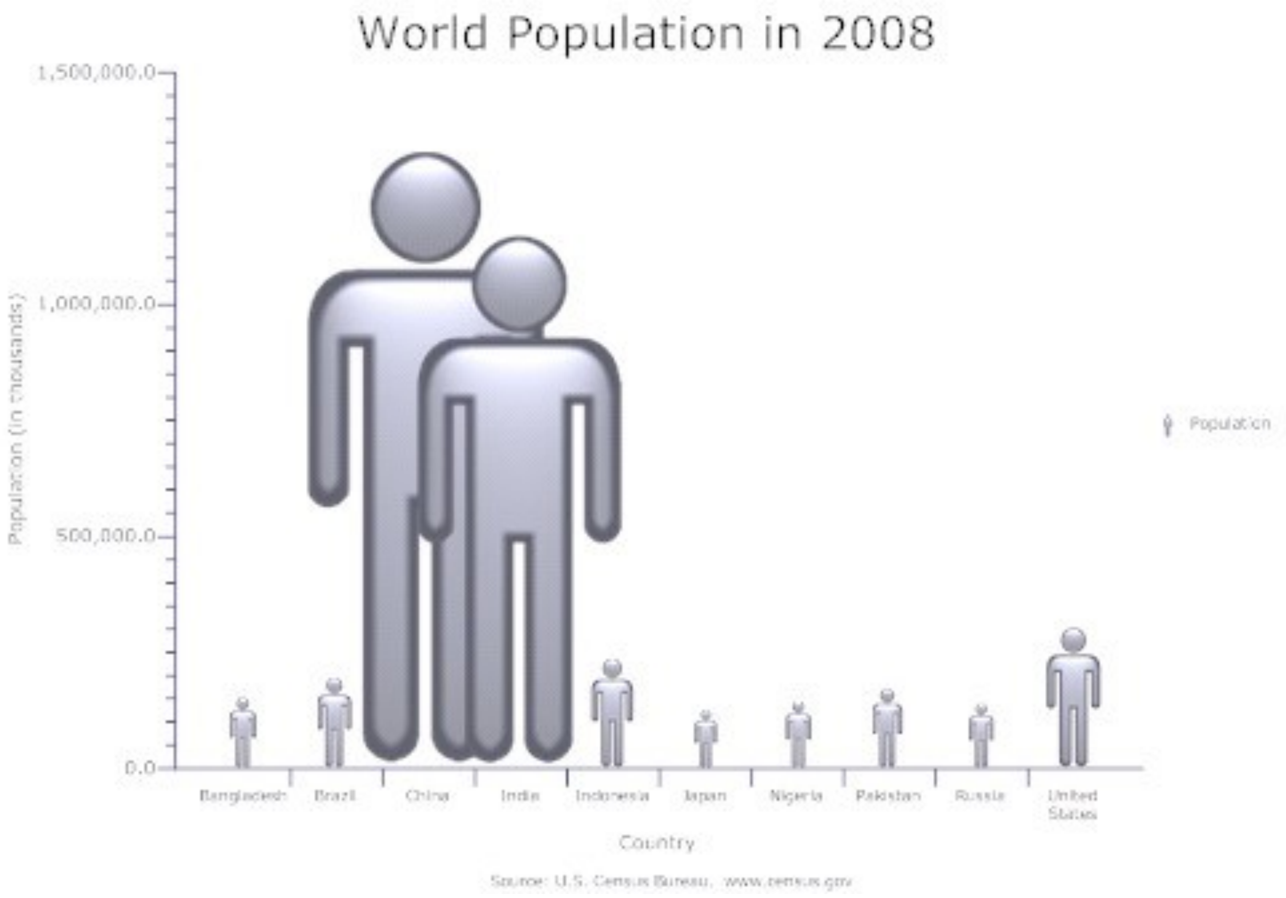

REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985
18
19
20
22
24
26
27
27.5
19.1 mpg, expected average for all cars on road, 1985
13.7 mpg, average for all cars on road, 1978
1978
1979
1980
1981
1982
1983
1984
1985

TUFTE 2001

TUFTE'S INTEGRITY PRINCIPLES

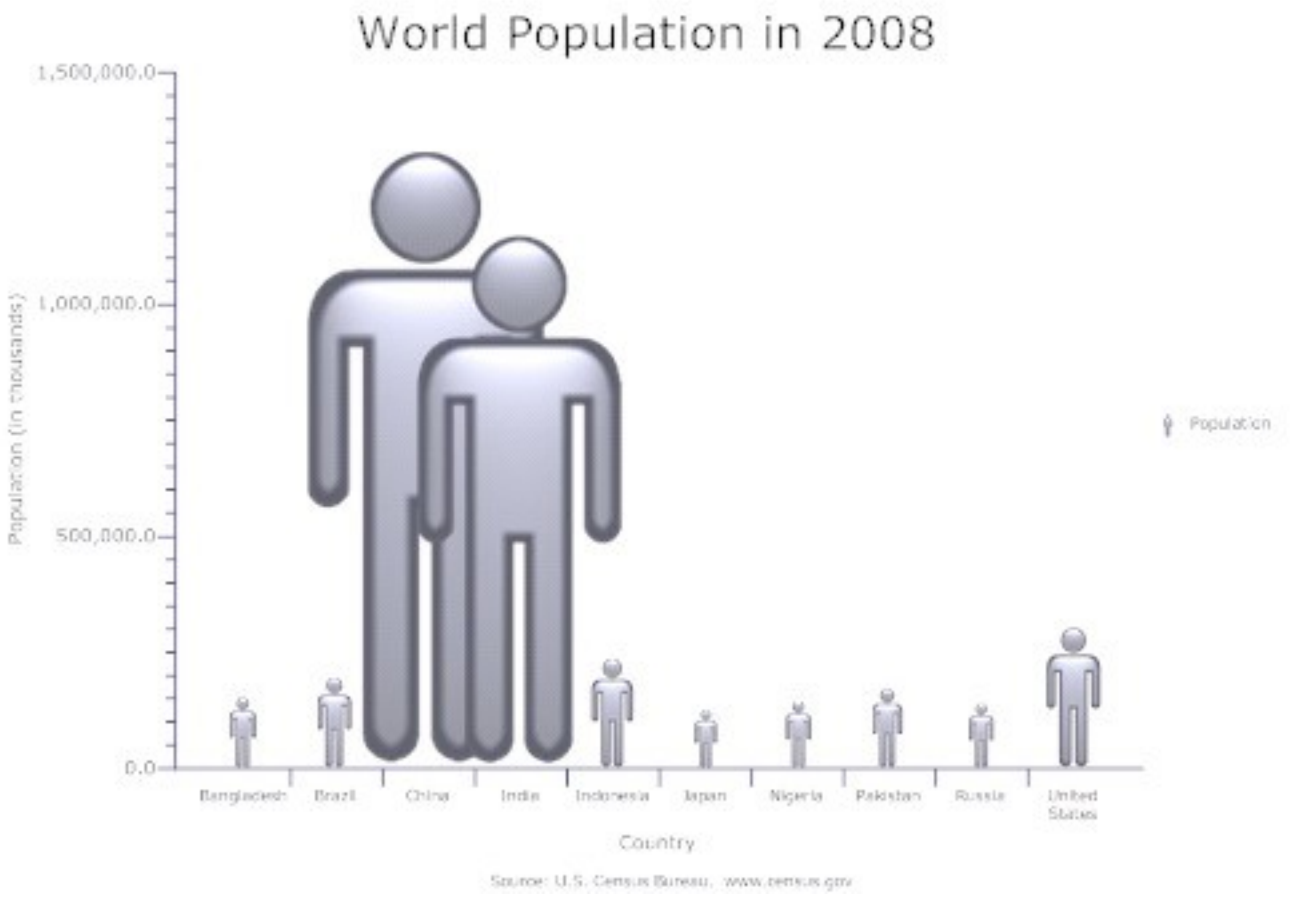
show *data* variation, not *design* variation

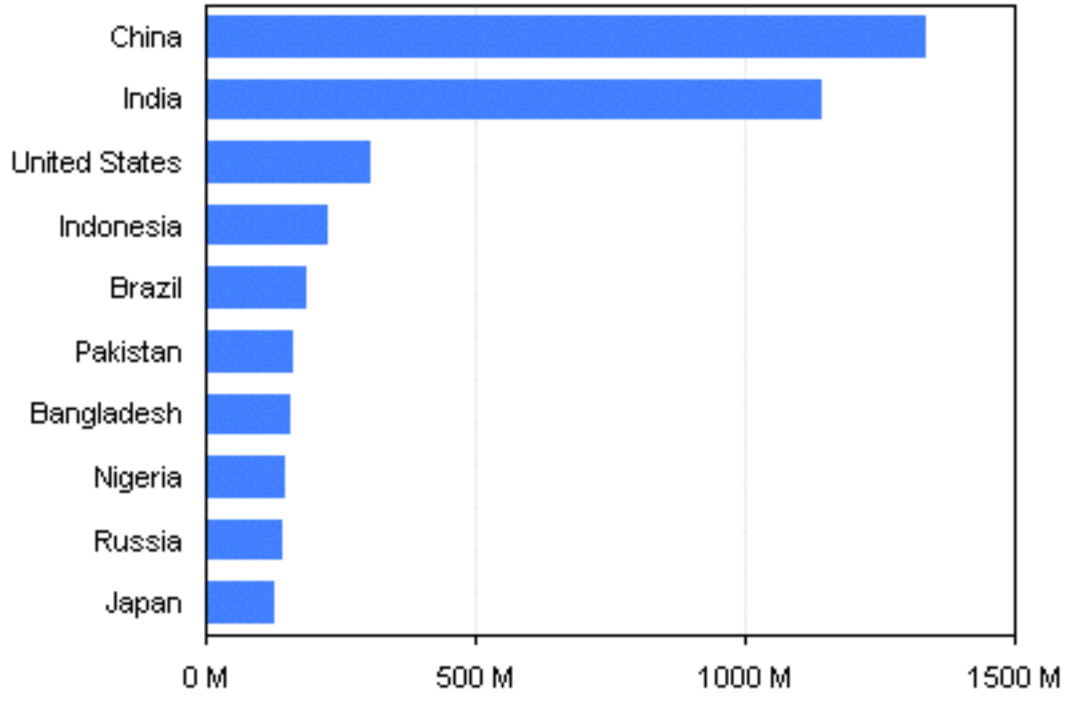
UNINTENDED SIZE CODING



[HTTP://PELTIERTECH.COM/WORDPRESS/BAD-BAR-CHART-PRACTICES-OR-SEND-IN-THE-CLOWNS/](http://peltiertech.com/WordPress/bad-bar-chart-practices-or-send-in-the-clowns/)

UNINTENDED SIZE CODING





HTTP://PELTIERTECH.COM/WORDPRESS/BAD-BAR-CHART-PRACTICES-OR-SEND-IN-THE-CLOWNS/

Tufte's integrity principles
Graphical excellence is that which

gives the viewer the greatest number of ideas

in the shortest time

with the least ink

in the smallest space

A. Einstein, "An explanation should be as simple as possible, but no simpler."

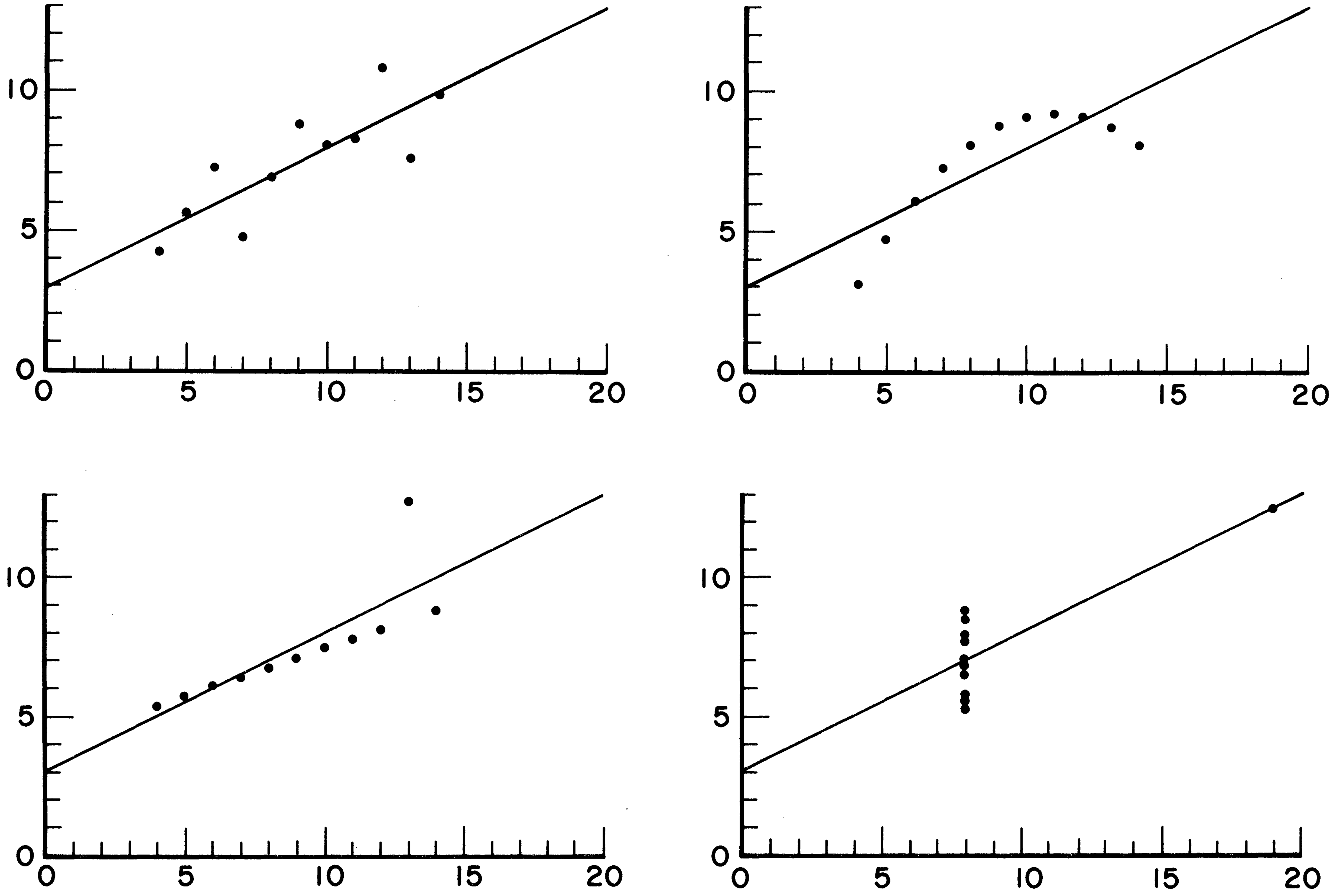
ANSCOMBE'S QUARTET

Data set		1-3	1	2	3		4	4
Variable		x	y	y	y		x	y
Obs. no.	1 :	10.0	8.04	9.14	7.46	:	8.0	6.58
	2 :	8.0	6.95	8.14	6.77	:	8.0	5.76
	3 :	13.0	7.58	8.74	12.74	:	8.0	7.71
	4 :	9.0	8.81	8.77	7.11	:	8.0	8.84
	5 :	11.0	8.33	9.26	7.81	:	8.0	8.47
	6 :	14.0	9.96	8.10	8.84	:	8.0	7.04
	7 :	6.0	7.24	6.13	6.08	:	8.0	5.25
	8 :	4.0	4.26	3.10	5.39	:	19.0	12.50
	9 :	12.0	10.84	9.13	8.15	:	8.0	5.56
	10 :	7.0	4.82	7.26	6.42	:	8.0	7.91
	11 :	5.0	5.68	4.74	5.73	:	8.0	6.89

TABLE. Four data sets, each comprising 11 (x, y) pairs.

Number of observations $(n) = 11$
Mean of the x's $(\bar{x}) = 9.0$
Mean of the y's $(\bar{y}) = 7.5$
Regression coefficient (b_1) of y on $x = 0.5$
Equation of regression line: $y = 3 + 0.5\,x$
Sum of squares of $x - \bar{x} = 110.0$
Regression sum of squares $= 27.50$ (1 d.f.)
Residual sum of squares of $y = 13.75$ (9 d.f.)
Estimated standard error of $b_1 = 0.118$
Multiple $R^2 = 0.667$

GRAPHS IN STATISTICAL ANALYSIS. F. J. ANSCOMBE. THE AMERICAN STATISTICIAN, VOL. 27, NO. 1. (FEB., 1973), PP. 17-21.
HTTP://WWW.SJSU.EDU/FACULTY/GERSTMAN/STATPRIMER/ANSCOMBE1973.PDF



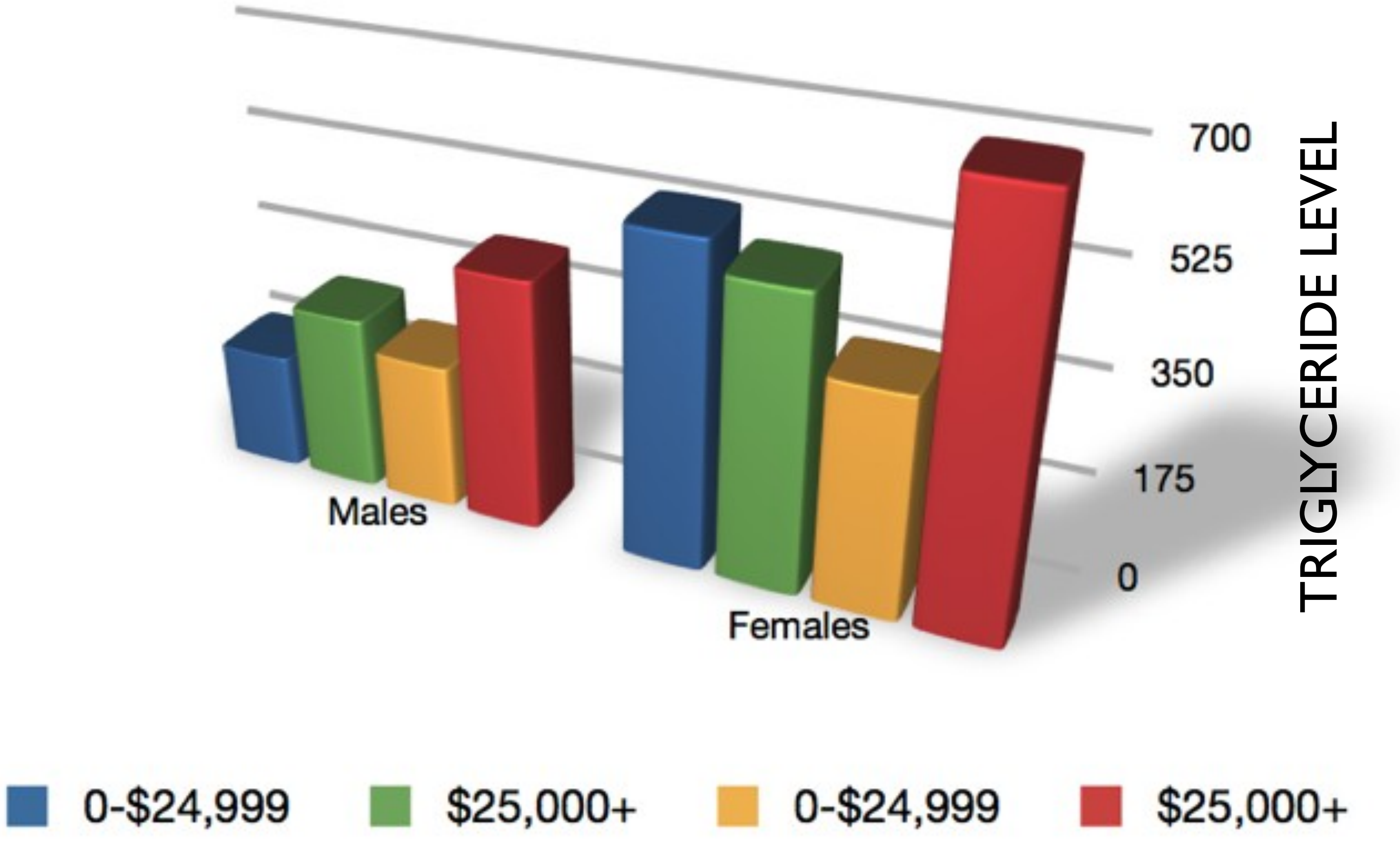
10
5
0
0
5
10
15
20
10
5
0
0
5
10
15
20
10
5
0
0
5
10
15
20
10
5
0
0
5
10
15
20

DESIGN PRINCIPLES

(or how to achieve integrity and excellence)

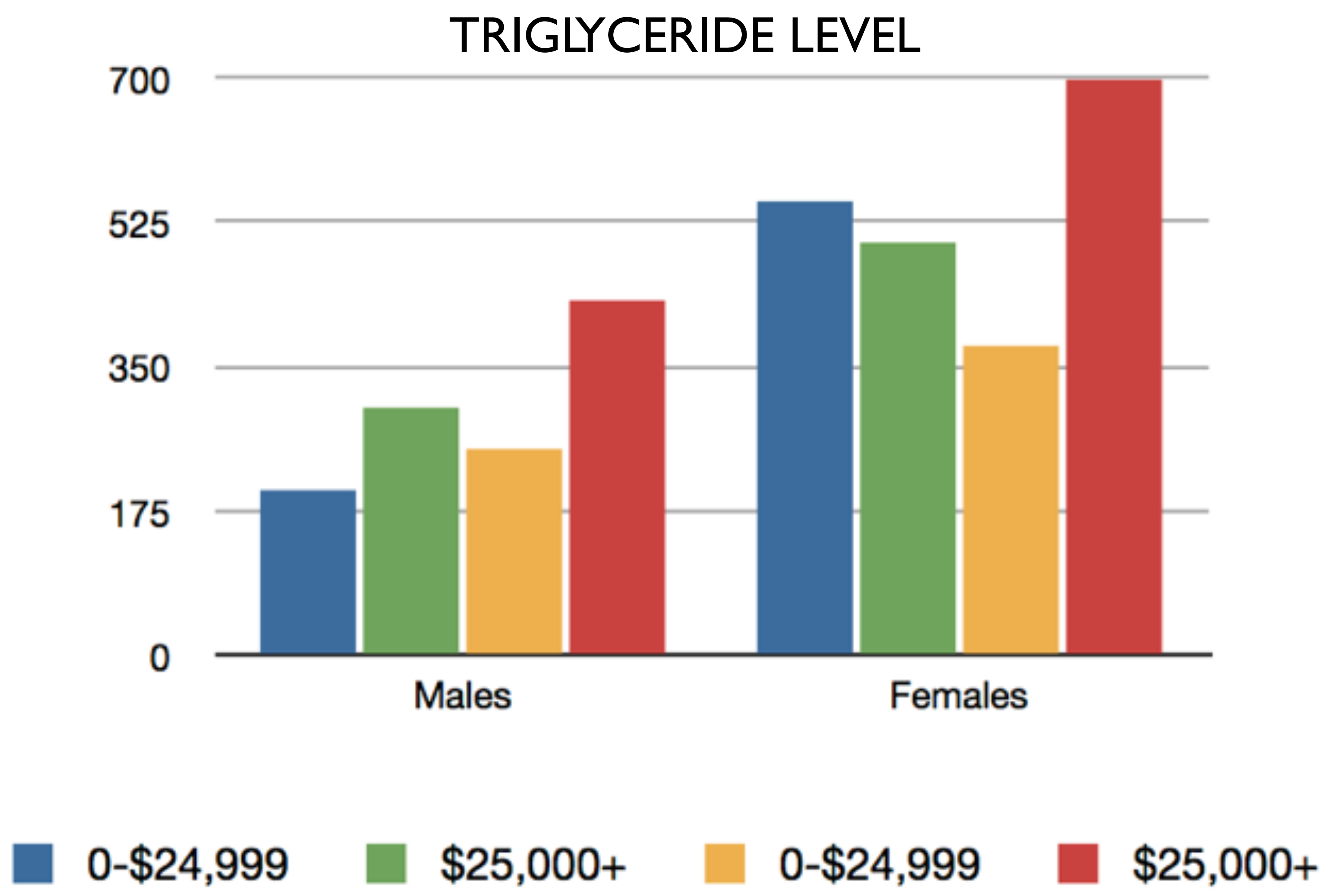
maximize the

$$\textbf{Data-ink Ratio} = \frac{\text{data-ink}}{\text{total ink used in graphic}}$$



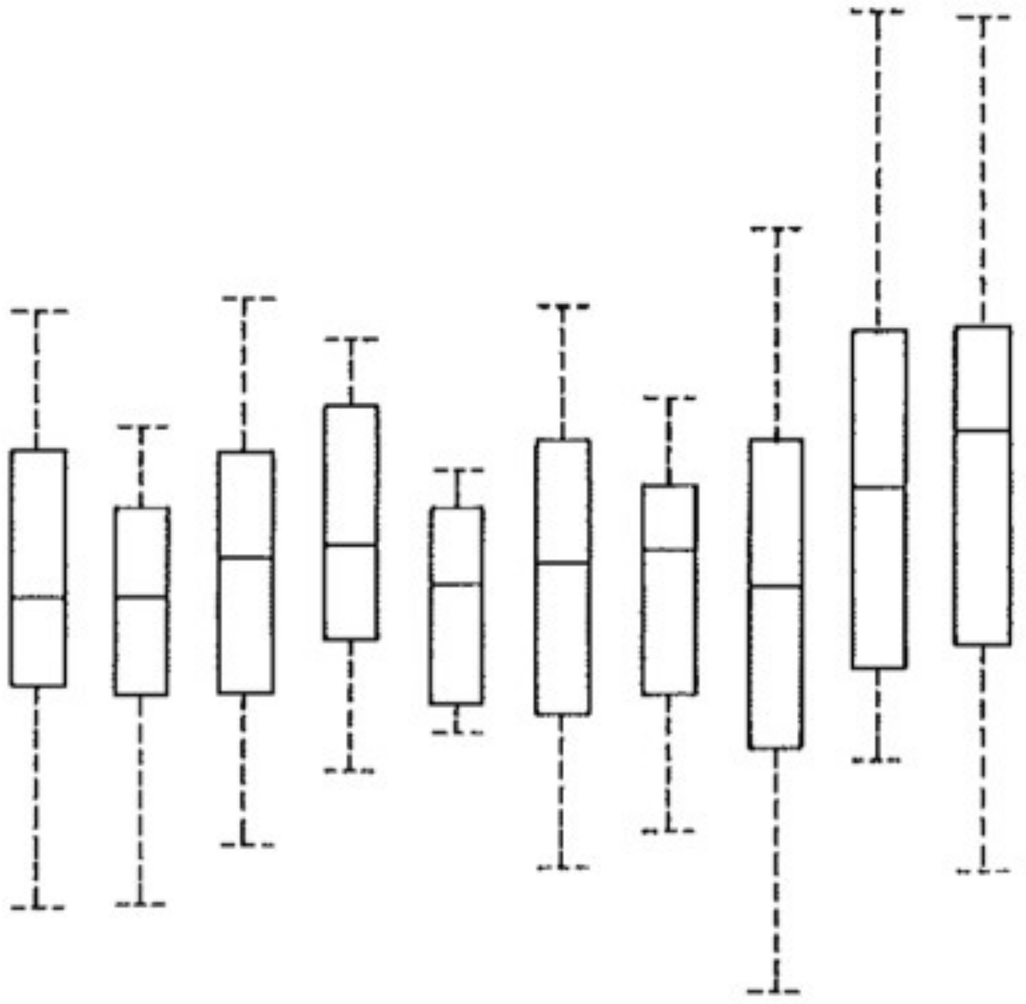
maximize the

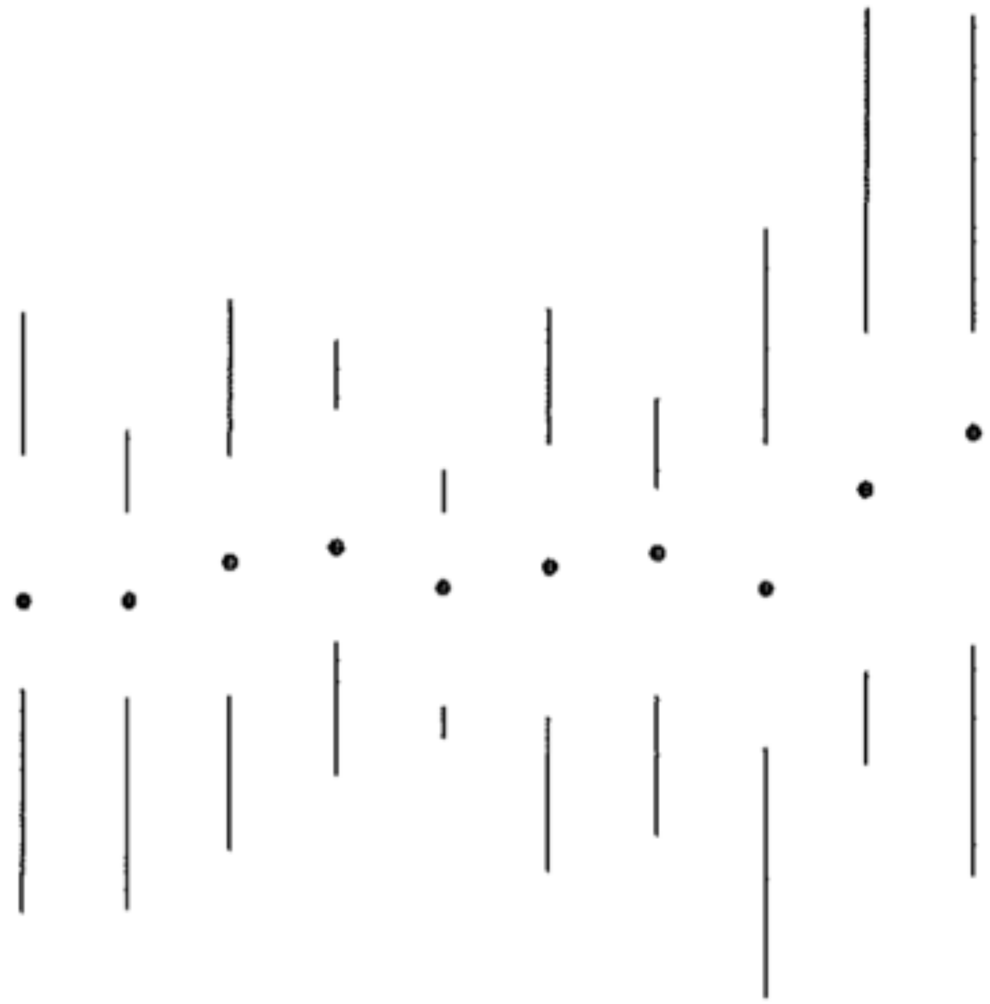
Data-ink Ratio = $\dfrac{\text{data-ink}}{\text{total ink used in graphic}}$



maximize the

Data-ink Ratio = $\dfrac{\text{data-ink}}{\text{total ink used in graphic}}$





TUFTE 2001

Eurographics / IEEE Symposium on Visualization 2011 (EuroVis 2011)
H. Hauser, H. Pfister, and J. J. van Wijk
(Guest Editors)

Volume 30 (2011), Number 3

A User Study of Visualization Effectiveness Using EEG and Cognitive Load

E. W. Anderson[1], K. C. Potter[1], L. E. Matzen[2], J. F. Shepherd[2], G. A. Preston[3], and C. T. Silva[1]

[1]SCI Institute, University of Utah, USA
[2]Sandia National Laboratories, USA
[3]Utah State Hospital, USA

Abstract

Effectively evaluating visualization techniques is a difficult task often assessed through feedback from user studies and expert evaluations. This work presents an alternative approach to visualization evaluation in which brain

COUNTER-POINT

This information is processed to provide insight into the cognitive load imposed on the viewer. This paper describes the design of the user study performed, the extraction of cognitive load measures from EEG data, and how those measures are used to quantitatively evaluate the effectiveness of visualizations.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: General—Human Factors, Evaluation, Electroencephalography

1. Introduction

Efficient visualizations facilitate the understanding of data sets through an appropriate choice of visual metaphor

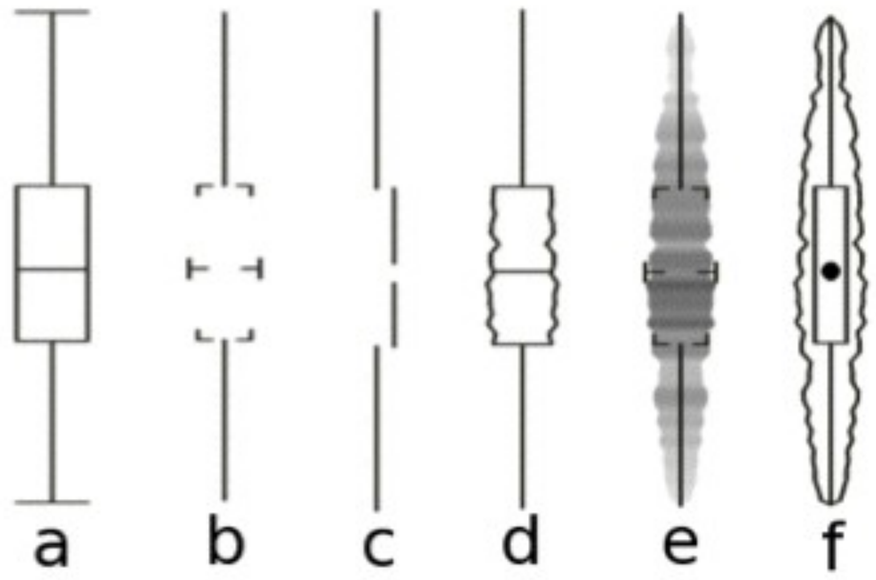
this paper strives to evaluate visualization techniques objectively by using passive, non-invasive monitoring devices to measure the burden placed on a user's cognitive resources.

EXPERIMENT

asked participants to choose box plot with largest range from a set

varied representation

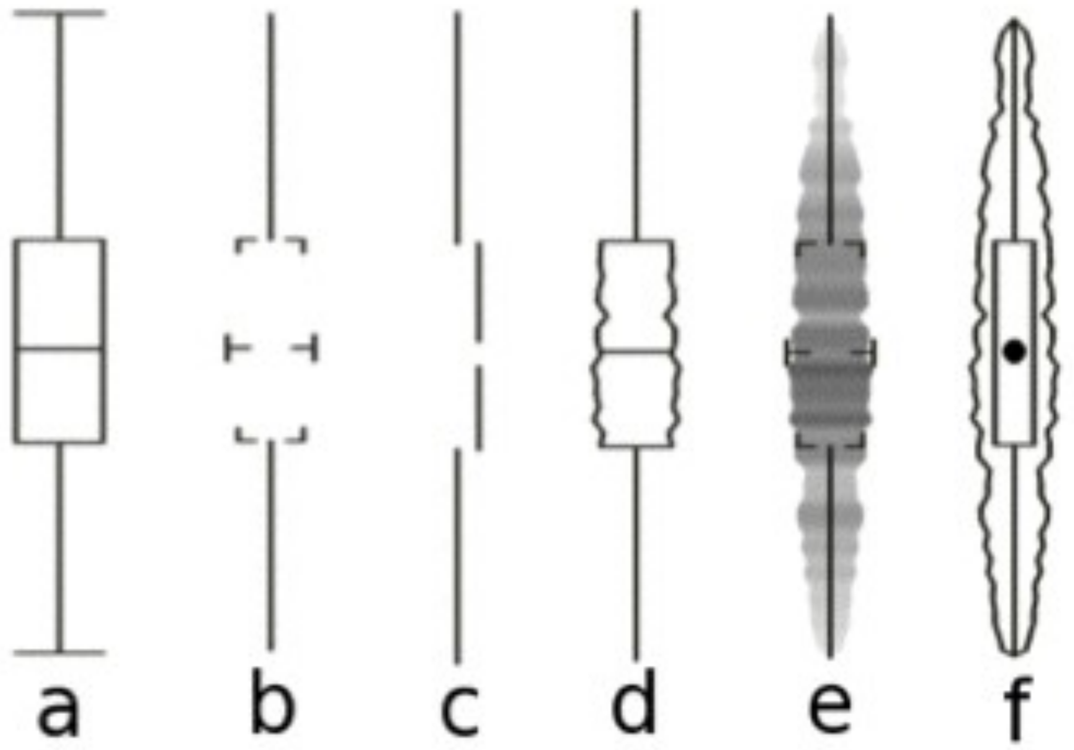
measured cognitive load from EEG brain waves



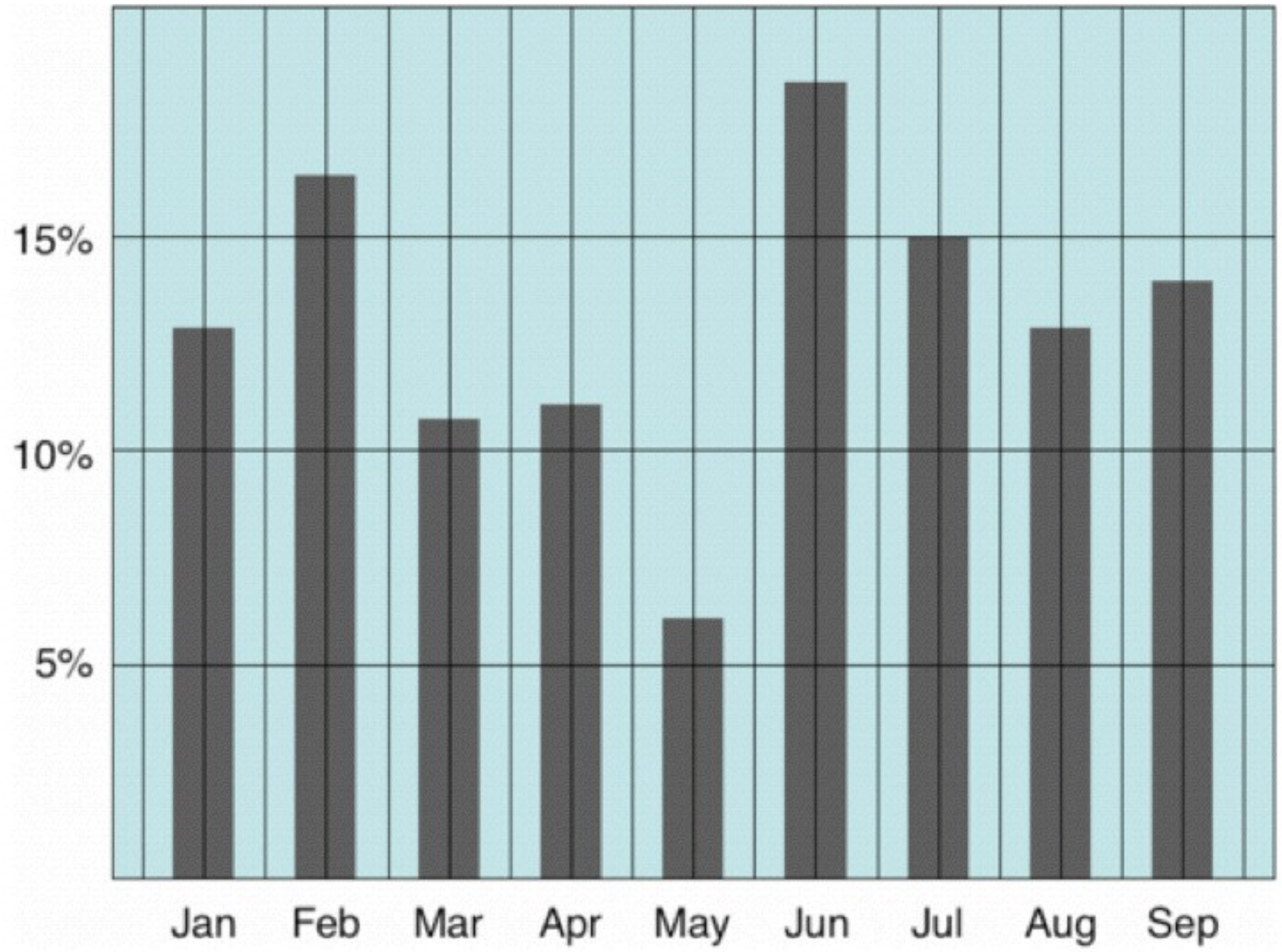
EXPERIMENTAL RESULTS

studies showed that the simplest box plot is hardest to interpret

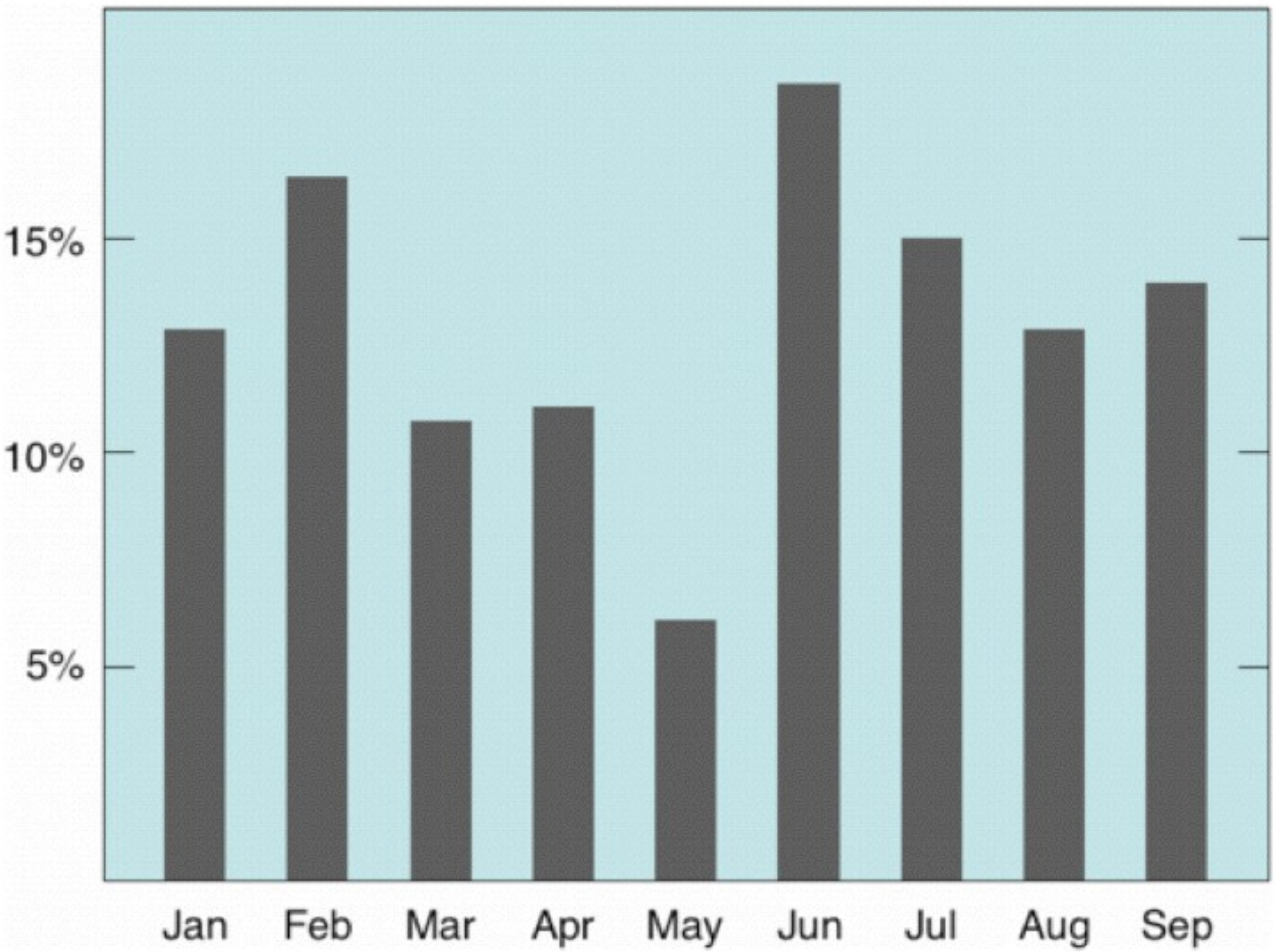
paper focused on cognitive load as an evaluation method



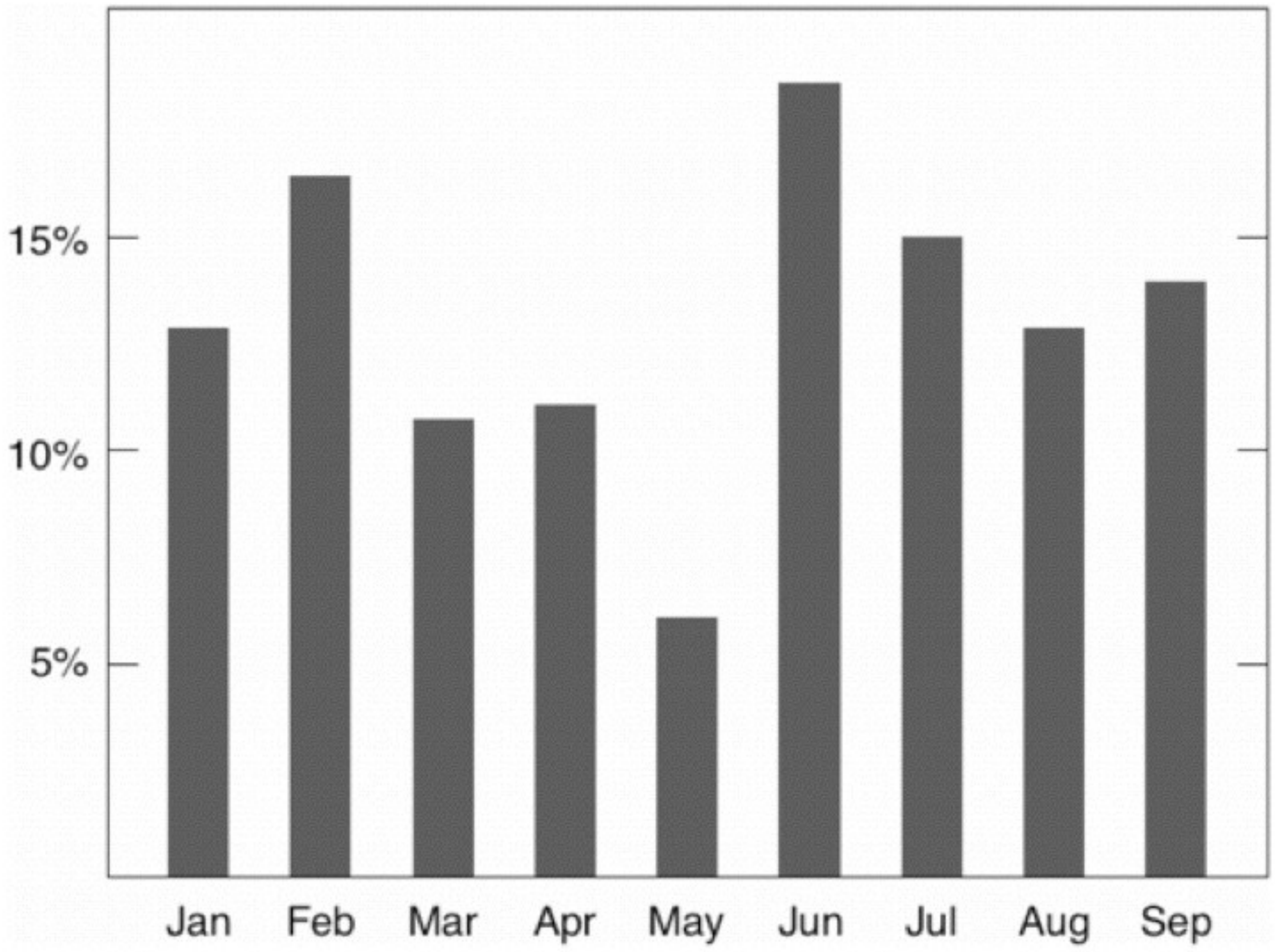
AVOID CHART JUNK



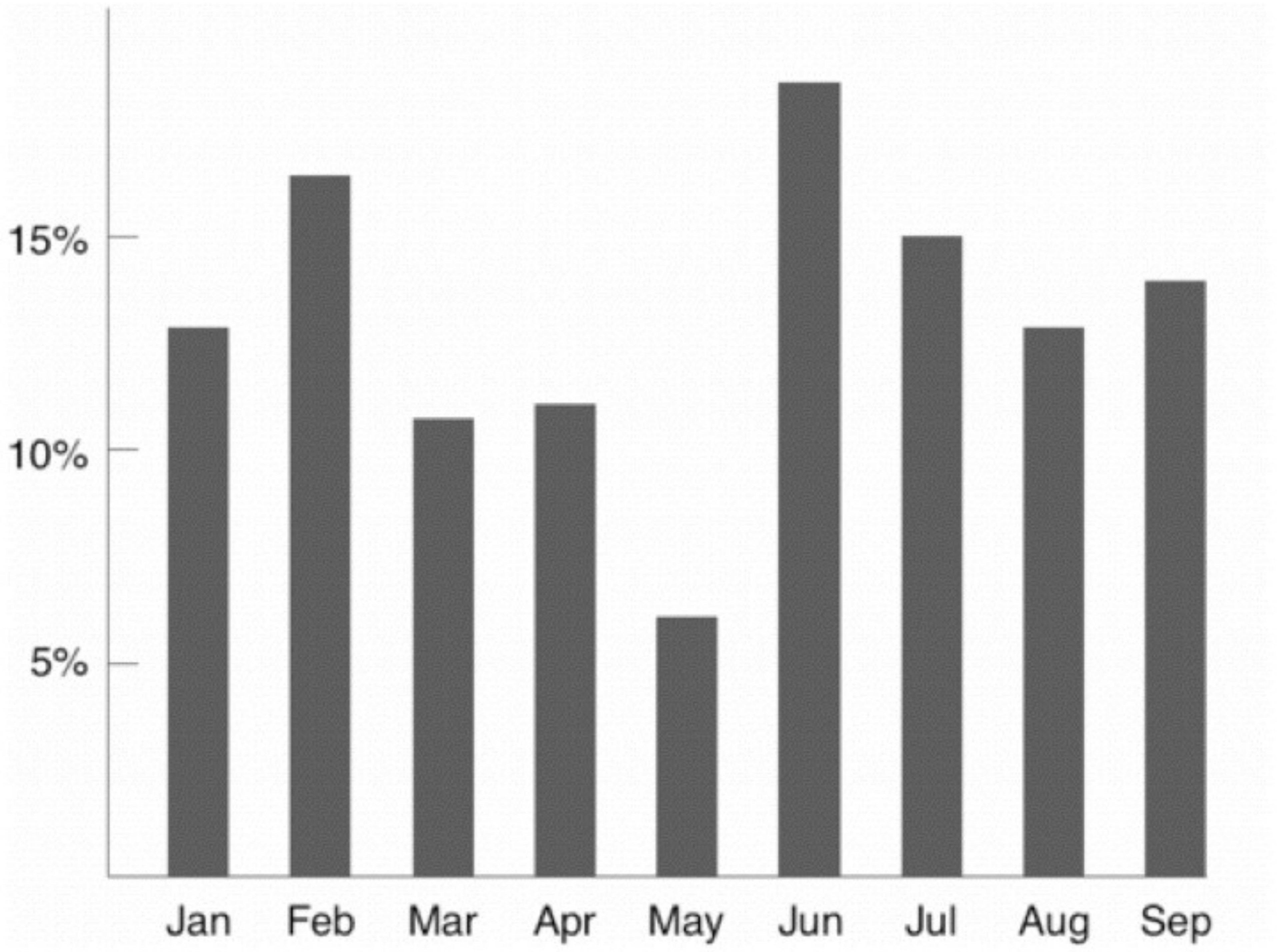
AVOID CHART JUNK



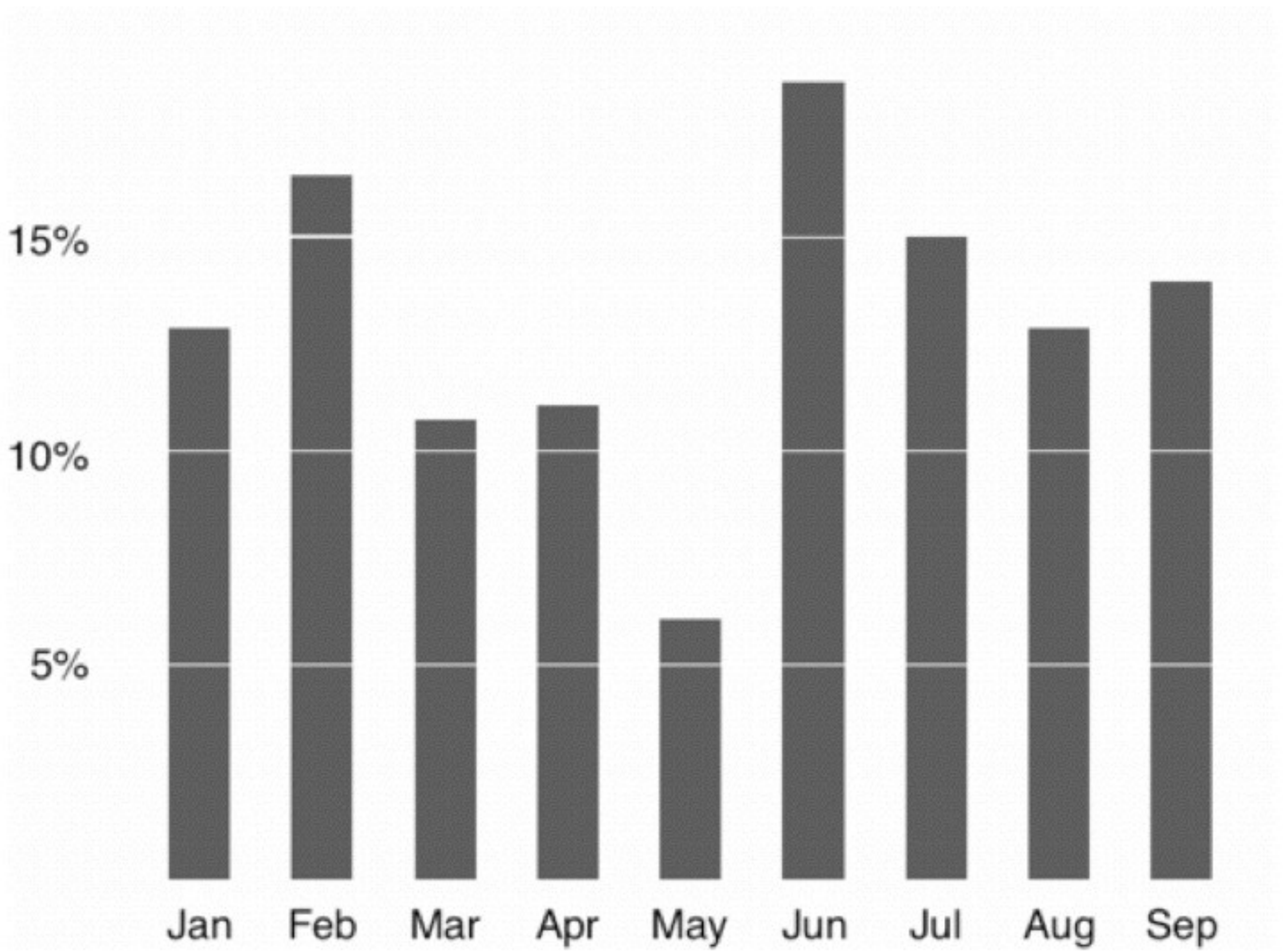
AVOID CHART JUNK



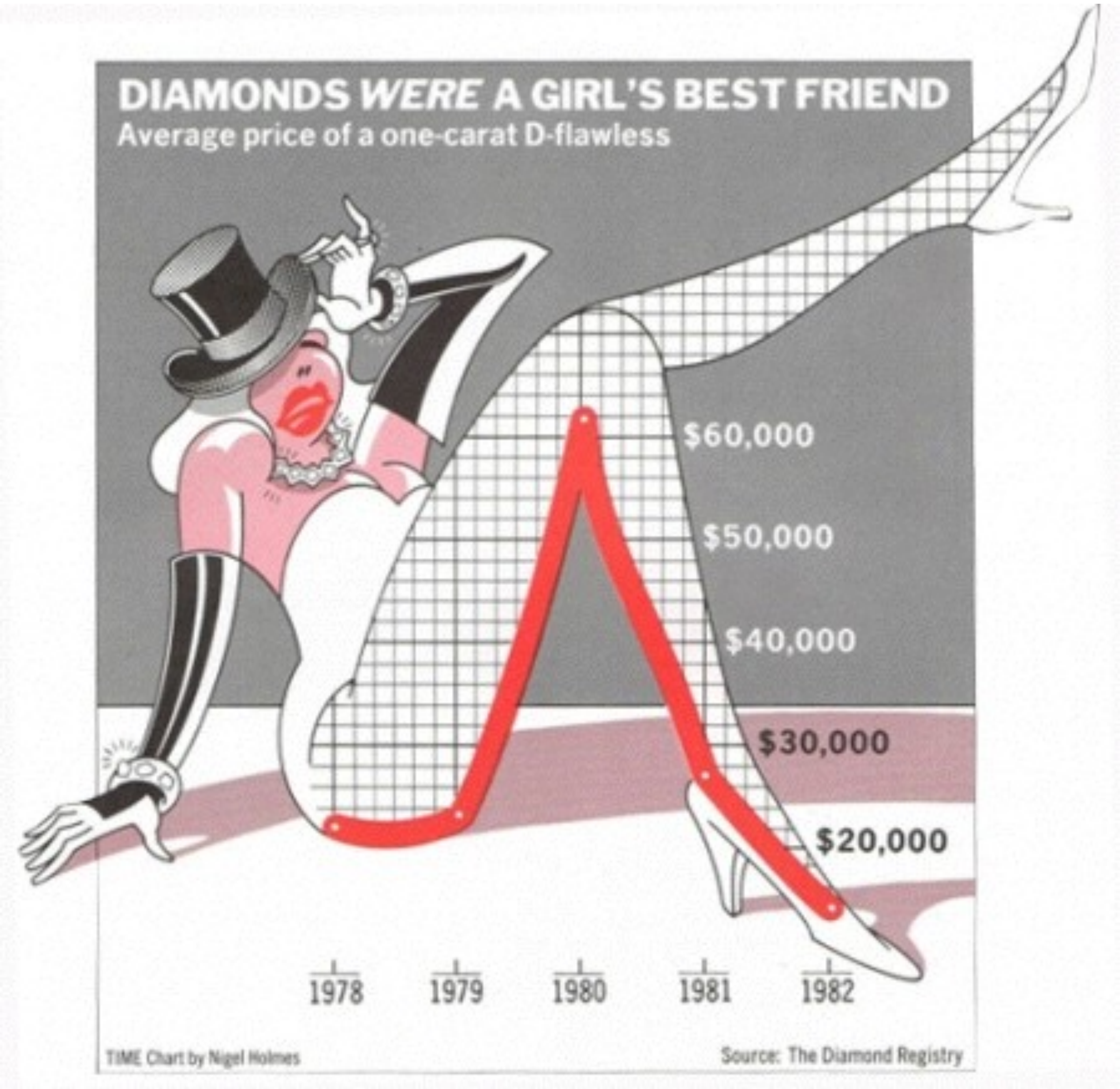
AVOID CHART JUNK

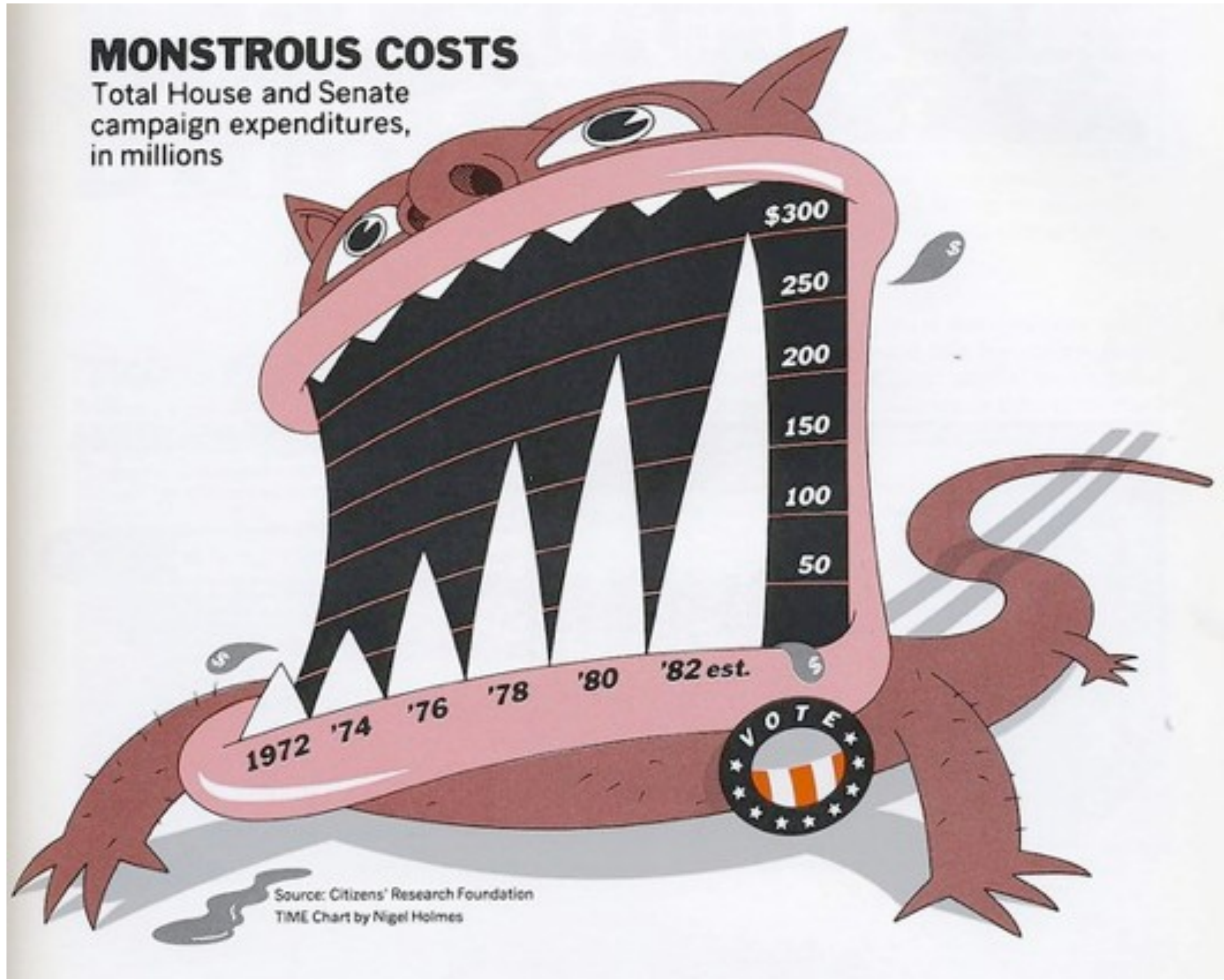


AVOID CHART JUNK



ATTRACTION OR DISTRACTION?





NIGEL HOLMES, TIME MAGAZINE

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

COUNTER-POINT

ABSTRACT
Guidelines for designing information charts often state that

presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords
Charts, information visualization, imagery, memorability.

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their

whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).

EXPERIMENTAL QUESTIONS

do visual embellishments cause comprehension problems?

do embellishments provide additional information that is valuable for the reader?

EXPERIMENTAL RESULTS

No significant difference between plain and embellished charts for interactive interpretation accuracy

No significant difference in recall accuracy after a five-minute gap

EXPERIMENTAL RESULTS

Significantly better recall for embellished charts of both the chart topic and the details (categories and trend) after long-term gap (2-3 weeks)

Participants saw value messages in the embellished charts significantly more often than in the plain charts

Participants found the embellished charts more attractive, most enjoyed them, and found that they were easiest and fastest to remember

What Makes a Visualization Memorable?

Michelle A. Borkin, *Student Member, IEEE*, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, *Student Member, IEEE*,
Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister, *Senior Member, IEEE*

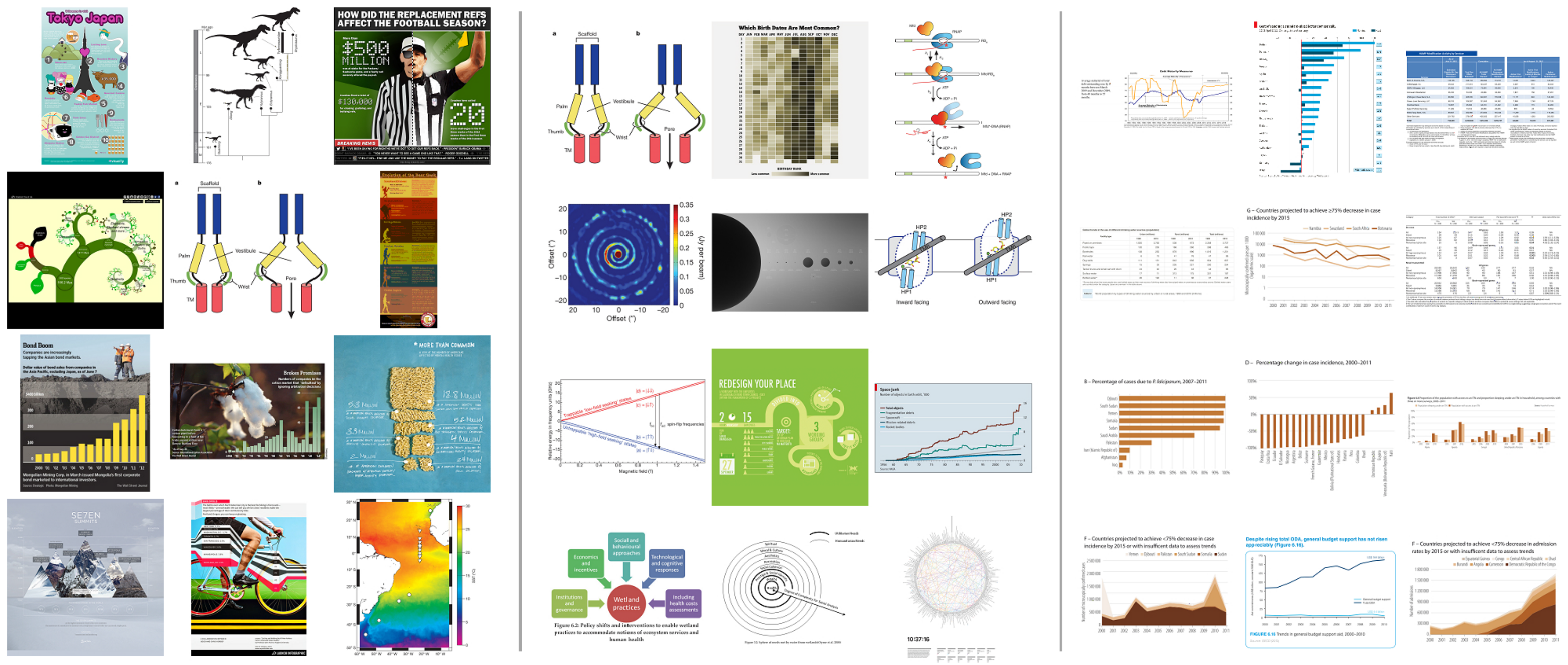


Fig. 1. **Left:** The top twelve overall most memorable visualizations from our experiment (most to least memorable from top left to bottom right). **Middle:** The top twelve most memorable visualizations from our experiment when visualizations containing human recognizable cartoons or images are removed (most to least memorable from top left to bottom right). **Right:** The twelve least memorable visualizations from our experiment (most to least memorable from top left to bottom right).

Abstract—An ongoing debate in the Visualization community concerns the role that visualization types play in data understanding. In human cognition, understanding and memorability are intertwined. As a first step towards being able to ask questions about impact and effectiveness, here we ask: “What makes a visualization memorable?” We ran the largest scale visualization study to date using 2,070 single-panel visualizations, categorized with visualization type (e.g., bar chart, line graph, etc.), collected from news media sites, government reports, scientific journals, and infographic sources. Each visualization was annotated with additional attributes, including ratings for data-ink ratios and visual densities. Using Amazon’s Mechanical Turk, we collected memorability scores for hundreds of these visualizations, and discovered that observers are consistent in which visualizations they find memorable and forgettable. We find intuitive results (e.g., attributes like color and the inclusion of a human recognizable object enhance memorability) and less intuitive results (e.g., common graphs are less memorable than unique visualization types). Altogether our findings suggest that quantifying memorability is a general metric of the utility of information, an essential step towards determining how to design effective visualizations.

Index Terms—Visualization taxonomy, information visualization, memorability

RESULTS

color and human recognizable objects enhance memorability

common graphs are less memorable than unique visualization types

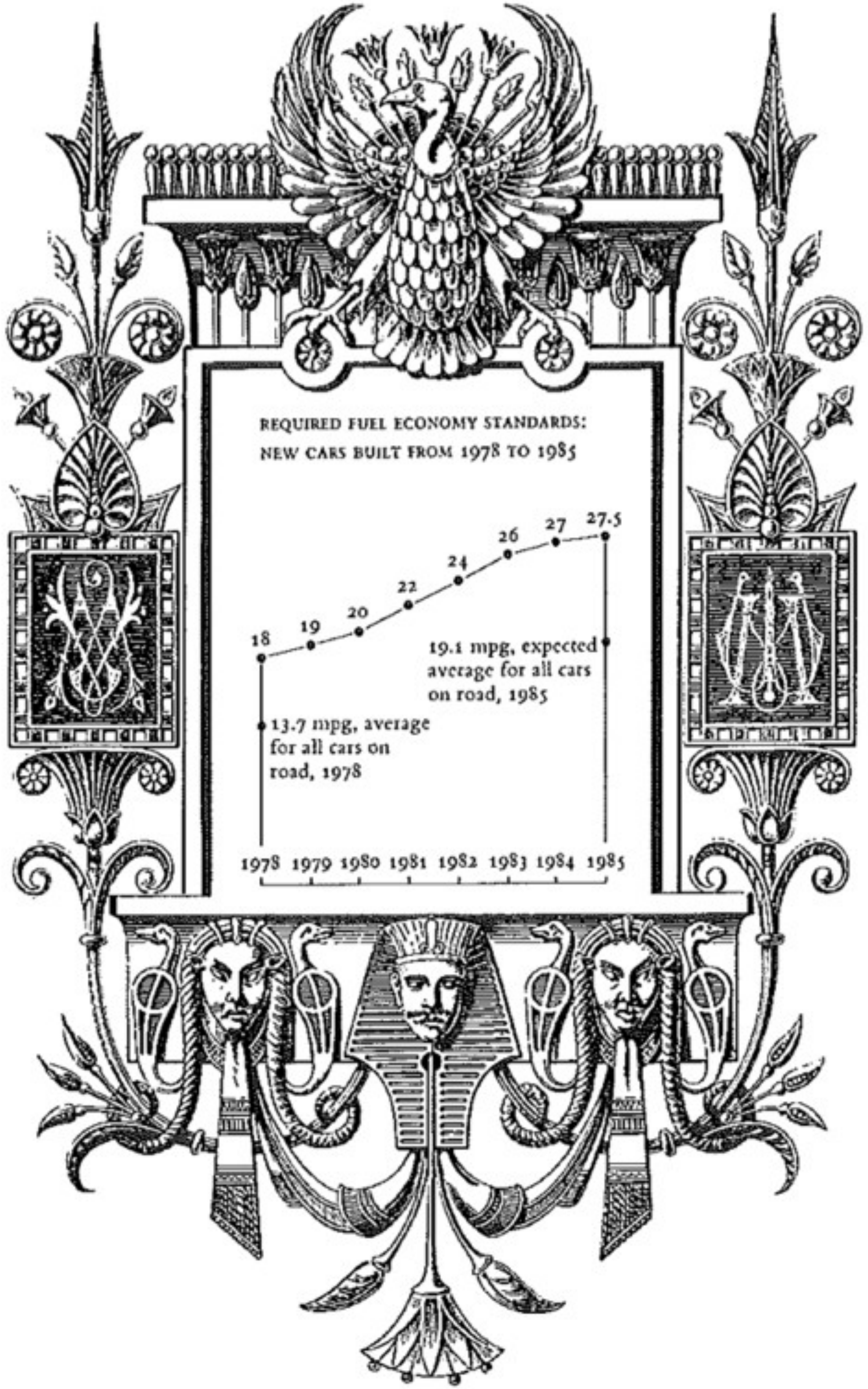
CHART JUNK? IT DEPENDS

persuasion
memorability
engagement

PROS

unbiased analysis
trustworthiness
interpretability
space efficiency

CONS


REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985
18
19
20
22
24
26
27
27.5
19.1 mpg, expected average for all cars on road, 1985
13.7 mpg, average for all cars on road, 1978
1978 1979 1980 1981 1982 1983 1984 1985

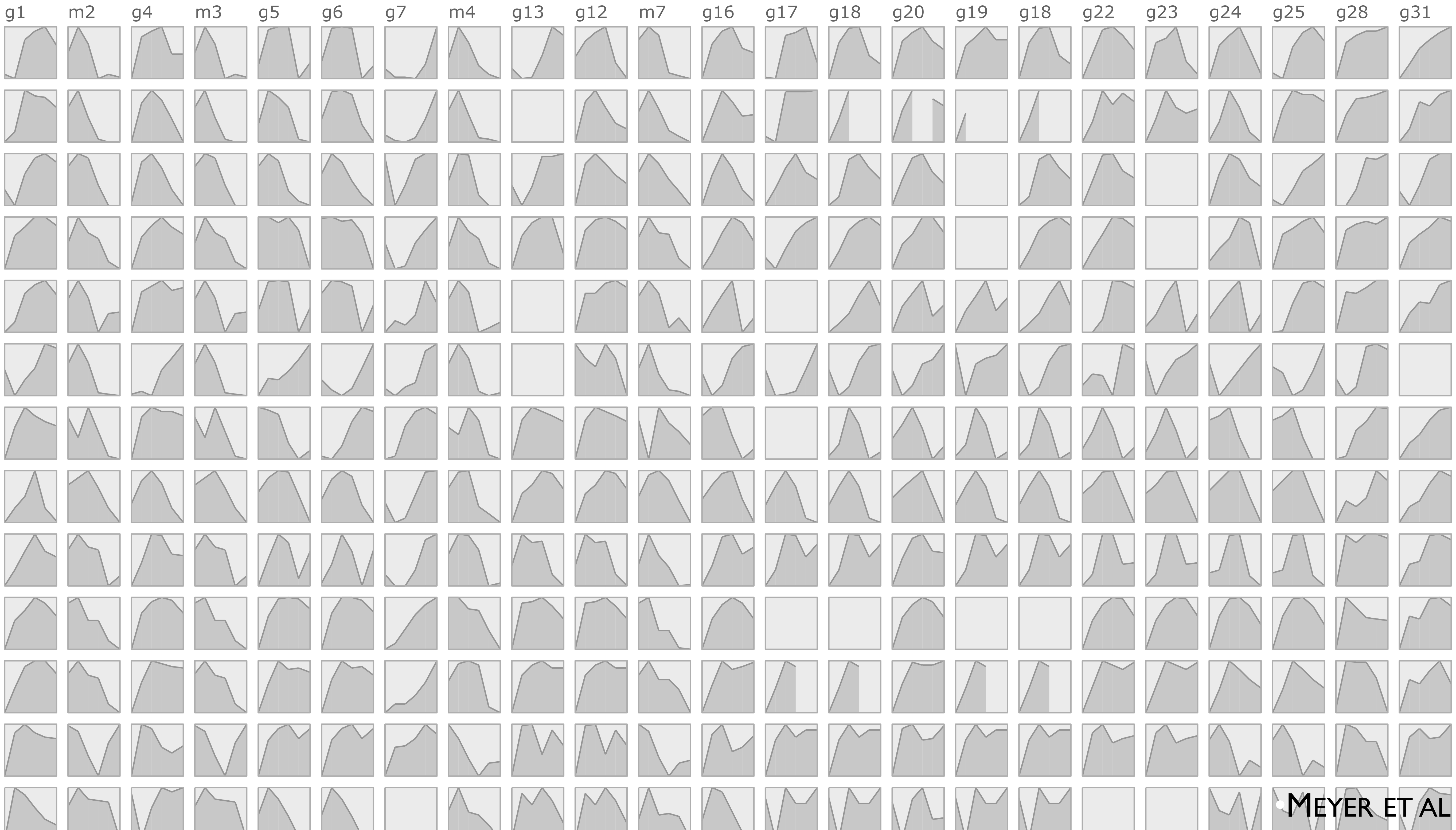
TUFTE 2001

maximize the

$$\textbf{Data Density} = \frac{\text{number of entries in data array}}{\text{area of data graphic}}$$

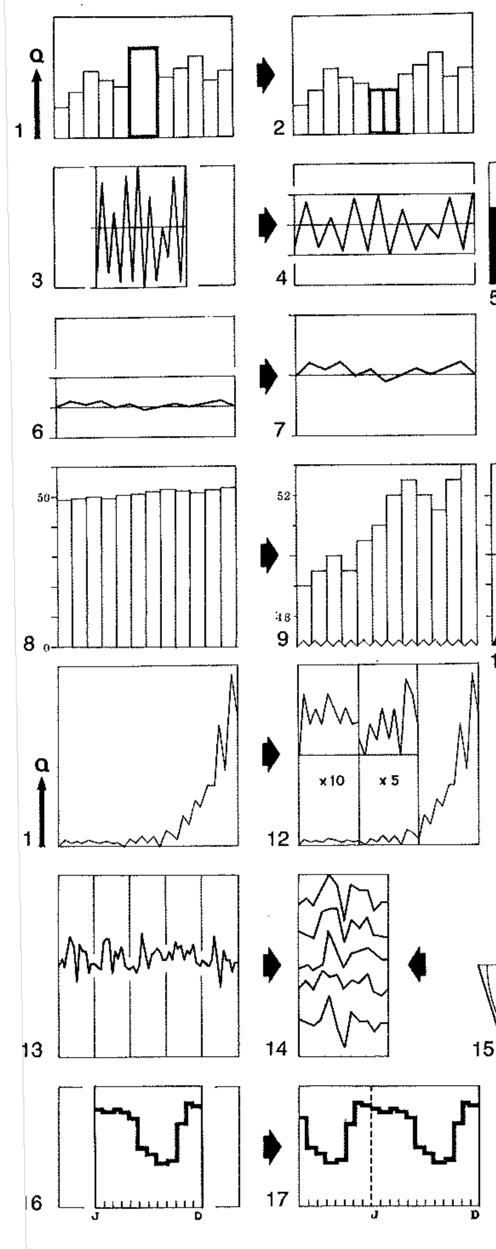
SHRINK THE GRAPHICS

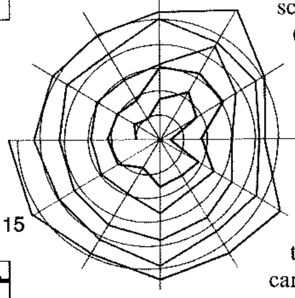
with small multiples



MEYER ET AL 2010

SHRINK THE GRAPHICS





GRAPHIC PROBLEMS POSED BY TIME SERIES

Scale in years
With a scale in years, a two-year total (figure 1) should be divided by 2 (figure 2). A total for six months should be multiplied by 2.

Pointed curves
For overly pointed curves (figure 3), the scale of the Q should be reduced; optimum angular perceptibility occurs at around 70 degrees (figure 4).
If the curve is not reducible (large and small variations), filled columns can be used (figure 5).

Flat curves
For overly flat curves (figure 6), the scale of the Q should be increased (figure 7).

Small variations
For small variations in relation to the total (figure 8), the total loses its importance, and the zero point can be eliminated, provided the reader is made aware of this elimination (figure 9). The graphic can be interpreted as an acceleration if a precise study of the variations is necessary; here, we use a logarithmic scale (figure 10). (See also page 240.)

Large range
For a very large range between the extreme numbers (figure 11), we must either:
(1) leave out the smallest variations;
(2) be concerned only with relative differences (logarithmic scale), without knowing the absolute quantities;
(3) select different parts (periods) within the ordered component and treat them on different scales above the common scale (figure 12).

Obvious periodicity
If there is obvious periodicity (figure 13), and the study involves a comparison of the phases of each cycle, it is preferable to break up the cycles in order to superimpose them (figure 14). A polar construction can be used, preferably in a spiral shape (figure 15), but we should not begin with too small a circle. As striking as it seems, it is less efficient than an orthogonal construction.

Annual curves
For annual curves of rainfall or temperature, if a cycle has two phases (figure 17), why depict only one (figure 16)?

A contrast
Unlike what we see in figure 18, the pertinent or "new" information must be separated from the background or "reference" information. The background involves: (a) the invariant, highlighted by a heading (Port St. Michel); (b) the highly visible identification of each component (tonnage and dates). The new information (the curve) must stand out from the background (figure 19).

Reference points
It is impossible to utilize a graphic such as figure 20, except in a general manner. There is confusion concerning the position of the points, and no potential comparison is possible, as it is in figure 21.

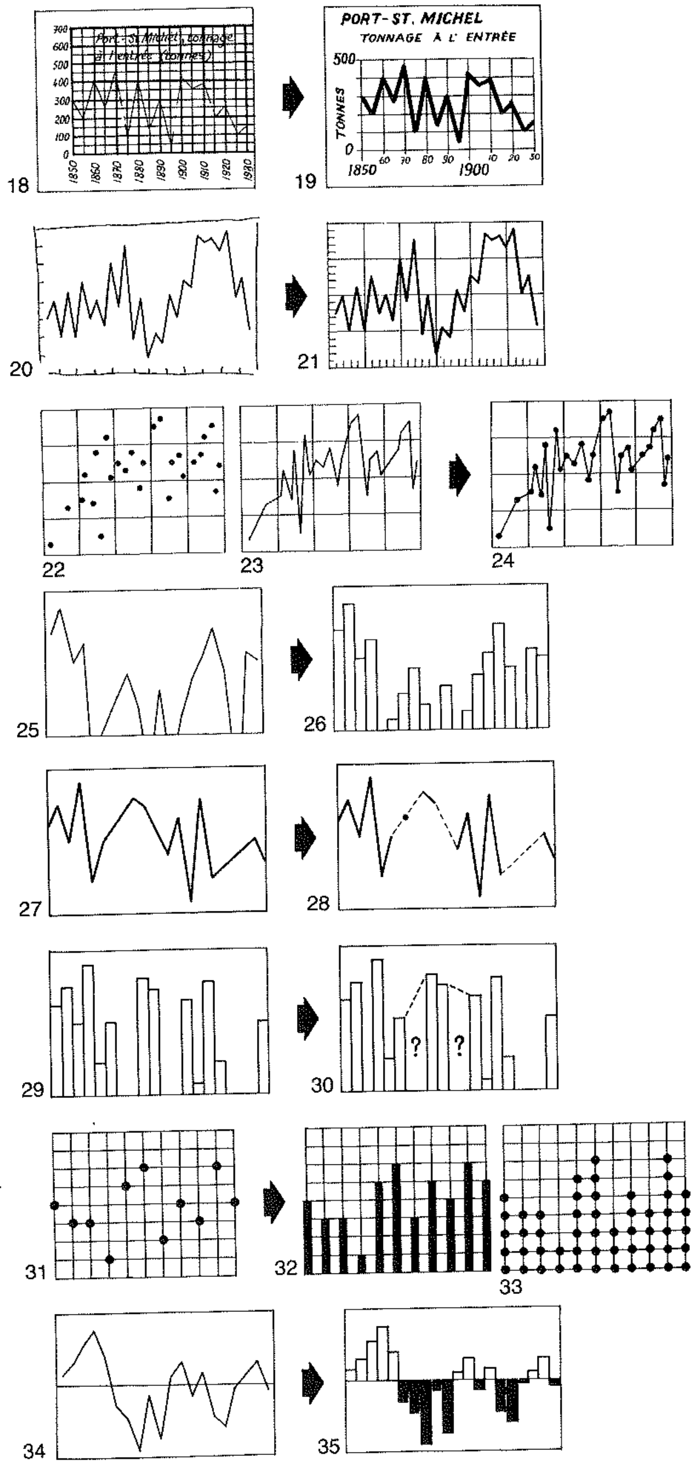
Precision reading
A precision reading (utilization on the elementary level, as in figure 24) is difficult in figure 22, which results in a poor reading of the order of the points, and in figure 23, where there is ambiguity concerning the position of the points. On the other hand, figure 22 does favor overall vision (correlation).

Null boxes
Curves accommodate null boxes poorly (figure 25). Columns (figure 26) are preferable.

Unknown boxes
The drawing must indicate the unknowns of the information in an unambiguous way (figures 28 and 30).The reader might interpret figure 27 as a change in the structure of the curve and figure 29 as involving null values.

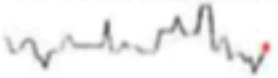
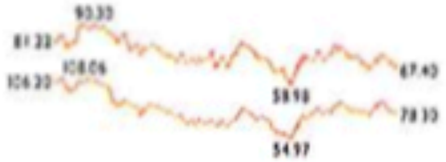
Very small quantities
Except in seeking a correlation (quite improbable here) the number of ships entering into a port is represented better by figure 33 than by figures 31 or 32. The reader can perceive the numerical values at first glance.

Positive-negative variation
This is in fact a problem involving three components O, Q, ≠ (+ –), and it must be visually treated as such. Figure 34 can be improved by utilizing a retinal variable (in figure 35 a value difference: black–white) to differentiate the ≠ component and thus highlight positive-negative variation.



BERTIN 1967

SHRINK THE GRAPHICS
with sparklines

Dequantification In exchange for an enormous increase in graphical resolving power, the wordlike size of sparklines precludes the overt labels and scaling of conventional statistical displays. Most of our examples have, however, depicted *contextual methods* for quantifying sparklines: the gray bar for normal limits and the red encoding to link data points in sparklines to exact numbers  glucose 6.6 ; global scale bars and labels for sparkline clusters; and, probably best of all, surrounding a sparkline with an implicit data-scaling box formed by nearby numbers that label key data points (such as beginning/end, high/low) 1.1025  1.1907 1.0783 1.2858. And now and then sparklines might be scaled by very small type: 

Production methods Data lines produced by conventional statistical graphics programs must be gathered together, rescaled, and resized into sparklines. Sometimes this can be quickly done by cutting and pasting data lines, then resizing the printed output to sparkline resolutions. To produce and display really elegant sparklines, however, currently requires elaborate software: (1) a *page layout* program, (2) a *graphic design* program that gives complete control over type, tables, linework, and (3) a *statistical analysis* program to generate hundreds of chartjunk-free sparklines for export into design and layout operations. Once the basic templates for sparklines are worked out, then ongoing production and

TUFTE 2006

Unseen and Unaware: Implications of Recent Research on Failures of Visual Awareness for Human–Computer Interface Design

D. Alexander Varakin and **Daniel T. Levin**
Vanderbilt University

Roger Fidler
Kent State University

COUNTER-POINT

ABSTRACT

Because computers often rely on visual displays as a way to convey information to a user, recent research suggesting that people have detailed awareness of only a small subset of the visual environment has important implications for human–computer interface design. Equally important to basic limits of awareness is the fact that people often over-predict what they will see and become aware of. Together, basic failures of awareness and people's failure to intuitively understand

ILLUSIONS OF VISUAL BANDWIDTH

people over-predict what they will see and become aware of

OVERESTIMATE OF BREADTH

belief that viewers can take in all (or most) of the details of a scene at once

adding extra visual features makes it harder to find specifics bits of information

OVERESTIMATE OF COUNTENANCE

belief that user will attend to a higher proportion of the display than they do

users typically have expectations about where in a display to look

OVERESTIMATE OF DEPTH

belief that attending to an object leads to more complete and deep understanding than is the case

TUFTE'S DESIGN PRINCIPLES

maximize the data-ink ratio

avoid chart junk (sometimes)

use multifunctioning elements

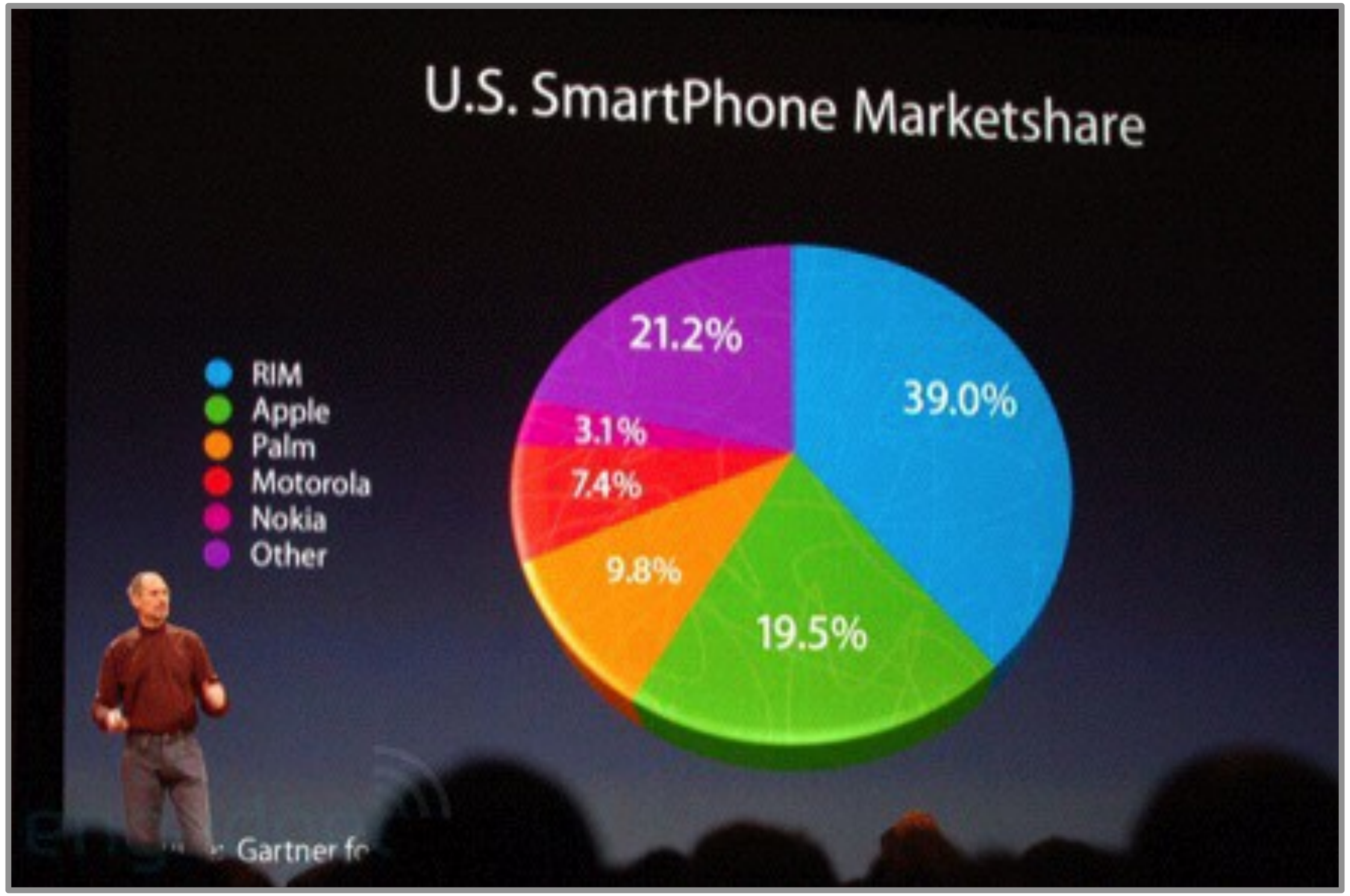
layer information

maximize the data density

shrink the graphics

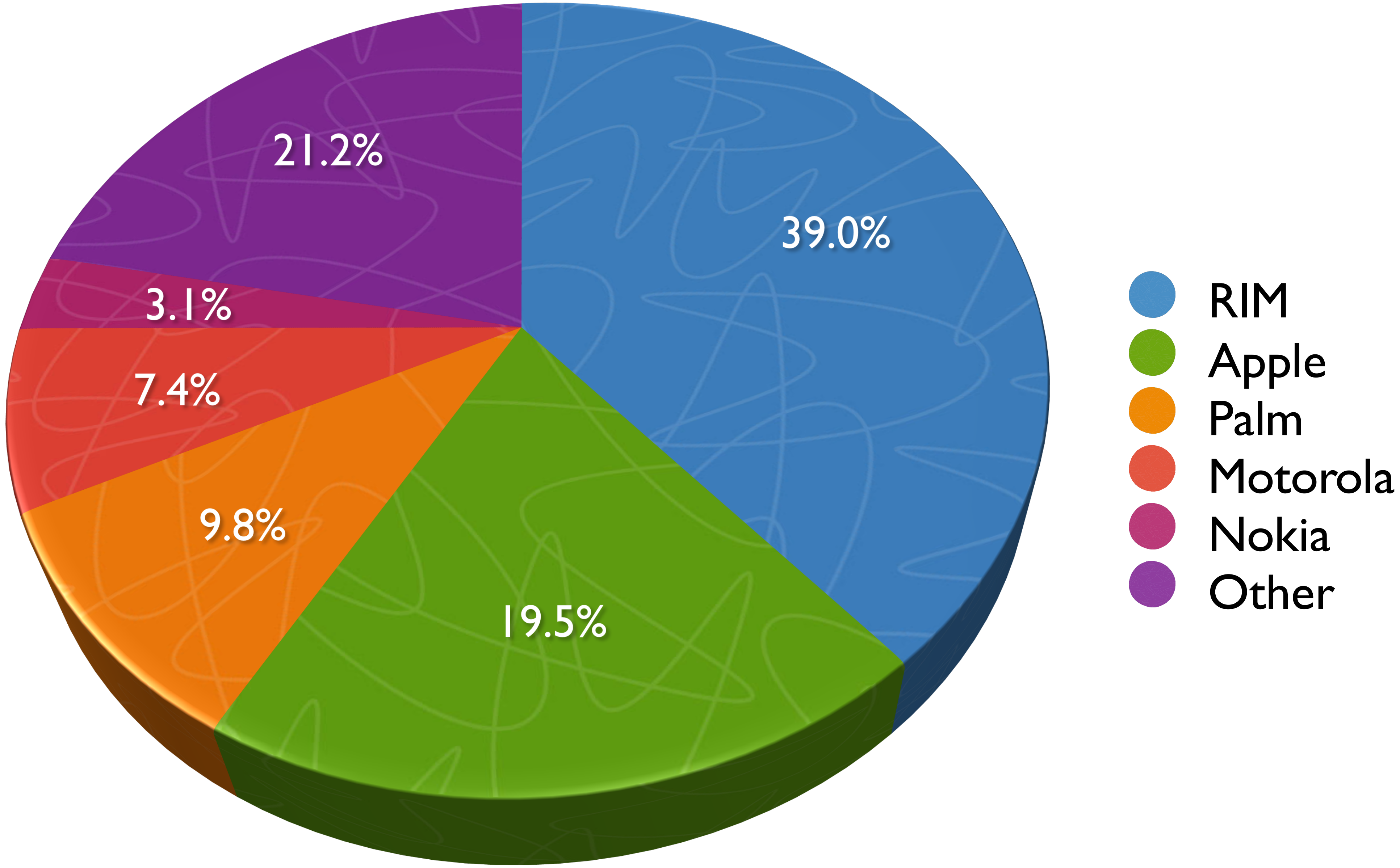
maximize the amount of data shown (sometimes)

DESIGN CRITIQUES

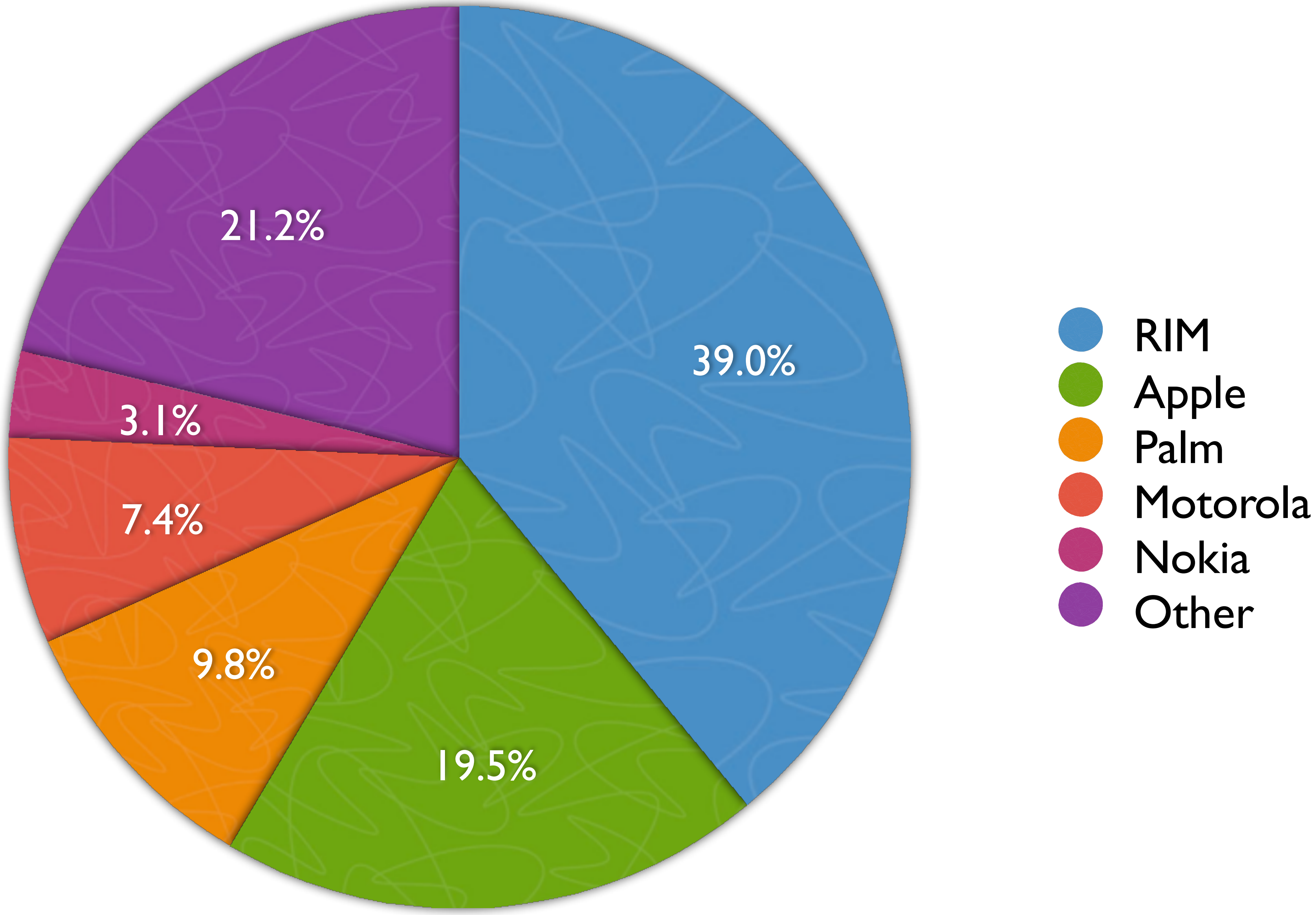


HTTP://WWW.GUARDIAN.CO.UK/TECHHNOLOGY/BLOG/2008/JAN/21/LIESDAMNLIESANDSTEVEJOBS

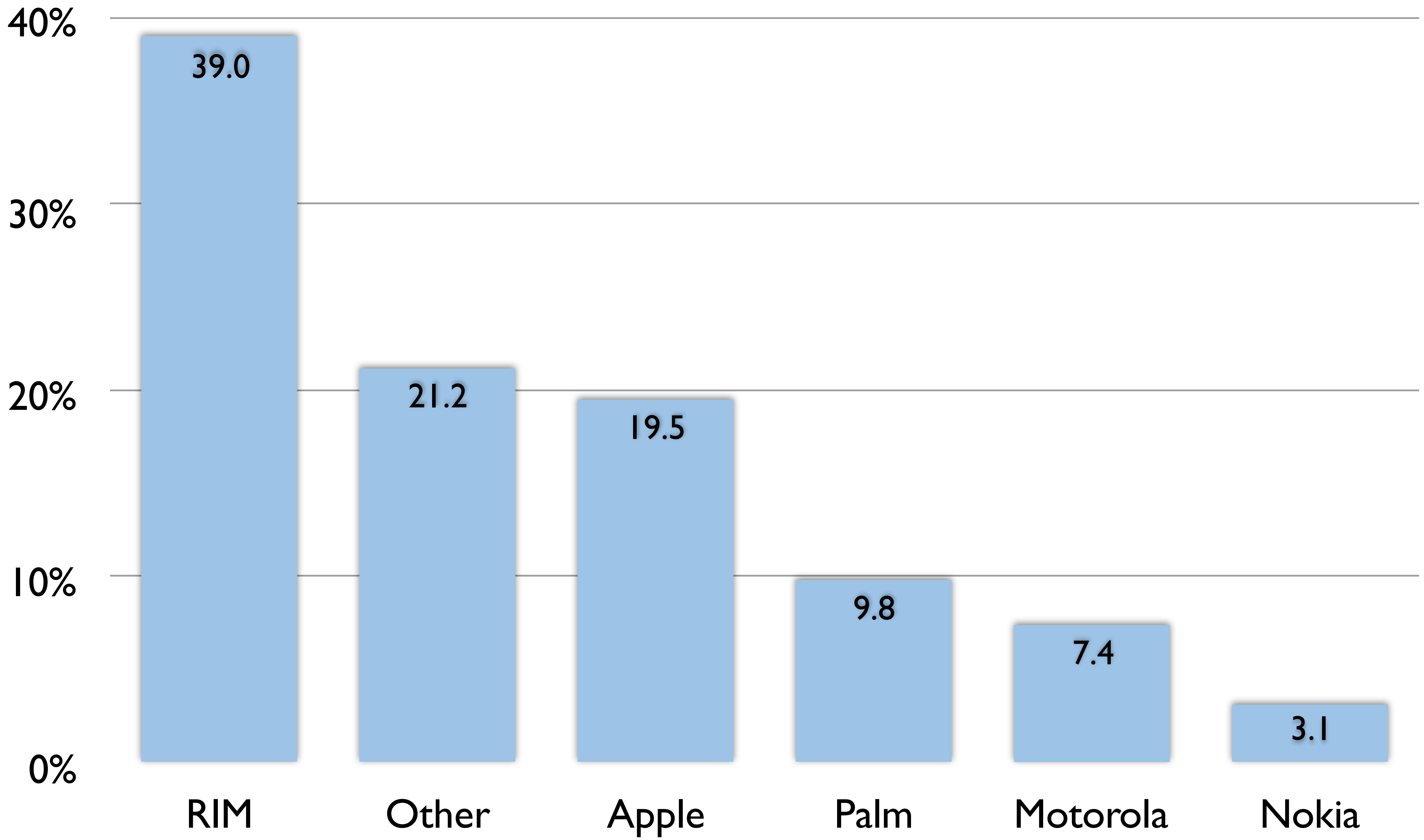
U.S. SmartPhone Marketshare



U.S. SmartPhone Marketshare



U.S. SmartPhone Marketshare

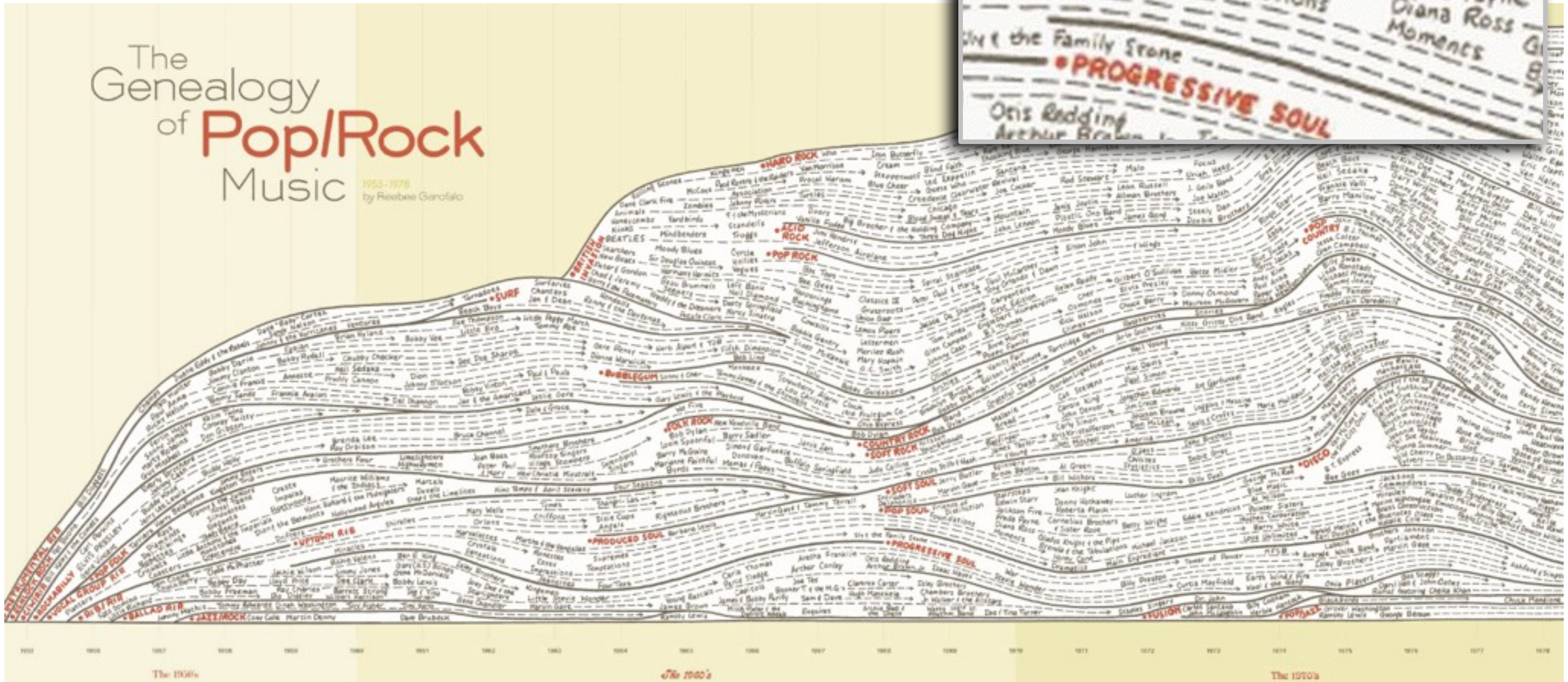


Recommended Reading

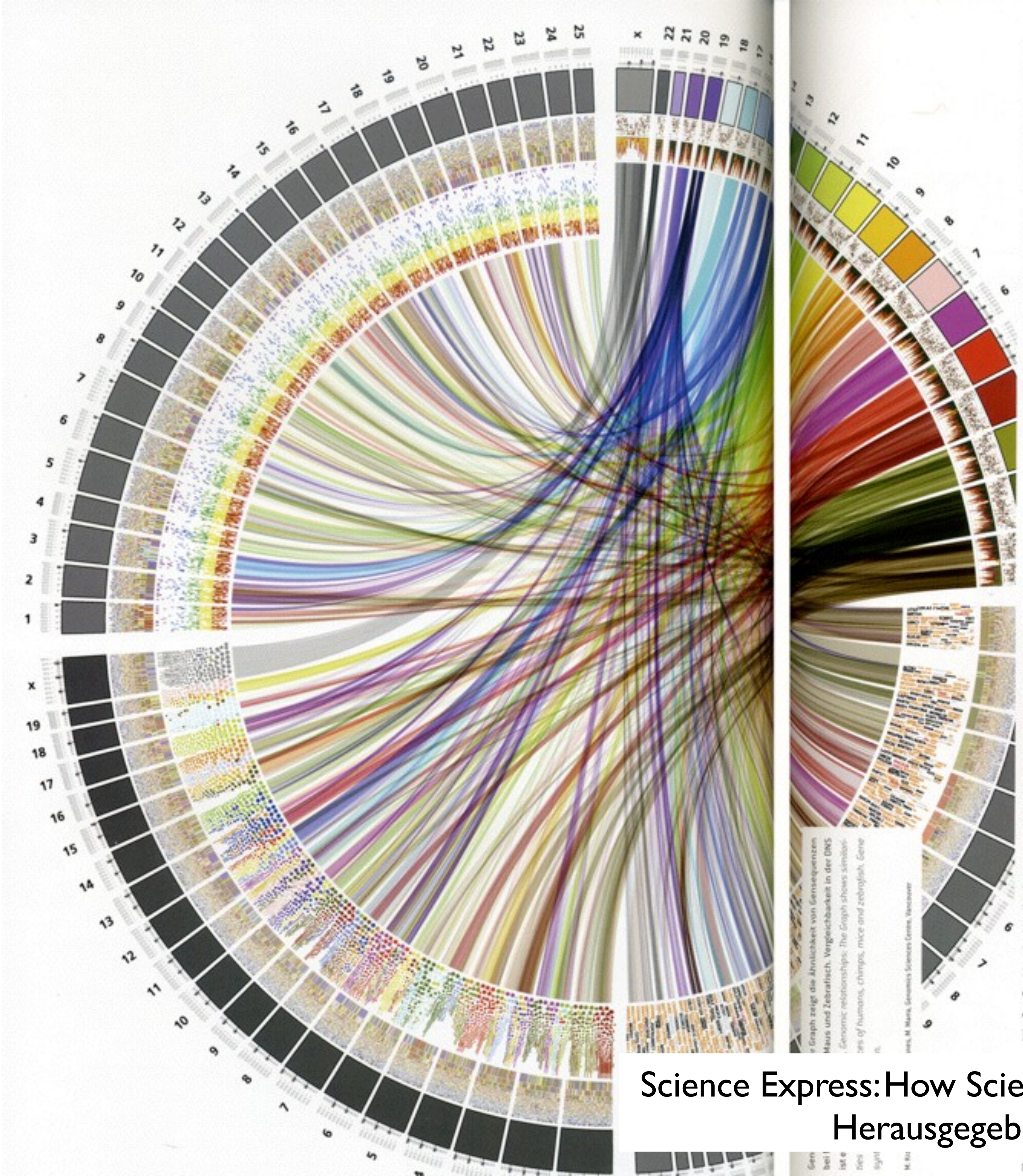
Visualization Analysis & Design: Chapter 4 (pp. 66-93)

The Visual Display of Quantitative Information: all

MAXIMIZE AMOUNT OF DATA SHOWN



STEVE CHAPPEL AND REEBE GAROFALO
IN ROCK 'N' ROLL IS HERE TO PAY: THE HISTORY AND POLITICS OF THE MUSIC INDUSTRY, 1977



On the road to a digital society ——— Computer technology is an ubiquitous element of our world, and fast networks are spanning the globe. This is changing the way we live and work and communicate. A new digital world is emerging, an environment in which creativity and innovation can flourish in many new ways. As a result, science and research have a greater influence on our life in the 21st century than ever before. This is attributable to massive investments in research and development, but also to intensive cooperation and tough competition. The convergence of nano-, bio-, information- and neurotechnologies facilitates completely new applications. Taking its place beside the more traditional factors of land, capital and employment, knowledge is fast becoming the decisive factor for prosperity – and also for the resolution of global problems. In this, the appropriate balance between digital freedom and digital security must be maintained. ——— **Science 2020: Systematically surveying the world** *Millions of scientists are getting to the bottom of the secrets of our world, across the whole spectrum of space, time, energy and complexity. Fundamentally new knowledge is emerging from research into inter-disciplinary topics or extreme states of matter. Science long ago escaped the constraints of working only in the realm of our natural living conditions and our perceptions. Considerable investment is flowing into efforts to decode the smallest building blocks of our world and to understand how their interplay produces brand new qualities. The drivers of innovation in research today are data capture via digital sensors; storage, analysis and visualisation via computer and software; and the global exchange of information and knowledge.* ——— **The cost of new knowledge is rising** There is now no part of our life that is not the subject of research. At the same time, it is becoming ever more difficult to generate new knowledge. These days, new research methods and technologies enable us to study even the ›farthest frontiers‹ of the world: extremely fast or slow processes, the tiniest building blocks or the largest structures, extreme cold or extreme heat. ——— **Networked knowledge takes on global challenges** *Thanks to worldwide information and communication networks, the challenges our civilisation faces in the long term are known to us sooner and more clearly than ever before. We can start developing solutions together at an earlier stage. Research on many topics is global – taking place in close cooperation or in international competition for the fastest and best solutions. National boundaries are becoming irrelevant. Millions of scientists work across countries, continents and time zones in thousands of labs. Their global networking enhances the diversity and efficiency of science and technology. And this, in turn, reinforces globalisation and networking. In a world changing at such a pace, each country must redefine its place.* ——— **The end of distance** Mankind faces enormous challenges both locally and globally – the challenge of using resources sustainably and of organising a global economy. Across the globe, complex processes are being recorded in detail, collated in databases and analysed in computer networks. New visualisation techniques make it possible to analyse larger and larger data records and to draw conclusions from the results. ——— **Global networking as the driving force of science** In the early days, the Internet linked up scientists, large-scale equipment and information; now it networks computational power and enormous amounts of data through grid and cloud computing. A global Semantic Web is emerging, bringing together data, expertise and knowledge that had previously been distributed among virtual libraries and observatories. The information is being intelligently developed,

Science Express: How Science and Technology change our life.
Herausgegeben von der Max-Planck-Gesellschaft

must have access to the Internet. This is where new knowledge is freely available.