

ARC²S Group

Applied Research on Computational Complex Systems

Perception, Design and Evaluation

Prof. Giancarlo **Ruffo**

“Analisi e Visualizzazione di Reti Complesse” (9 credits)

Laurea Magistrale in **Informatica**

Università degli Studi di Torino

A.A. 2018/19

@giaruffo



Perception

How is our perception distorted?

Recap (from lecture 1)

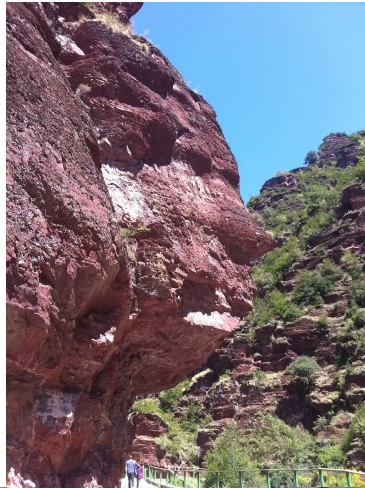
- Pre-attentive attributes
- Inattentional Blindness
- Gestalt Principles

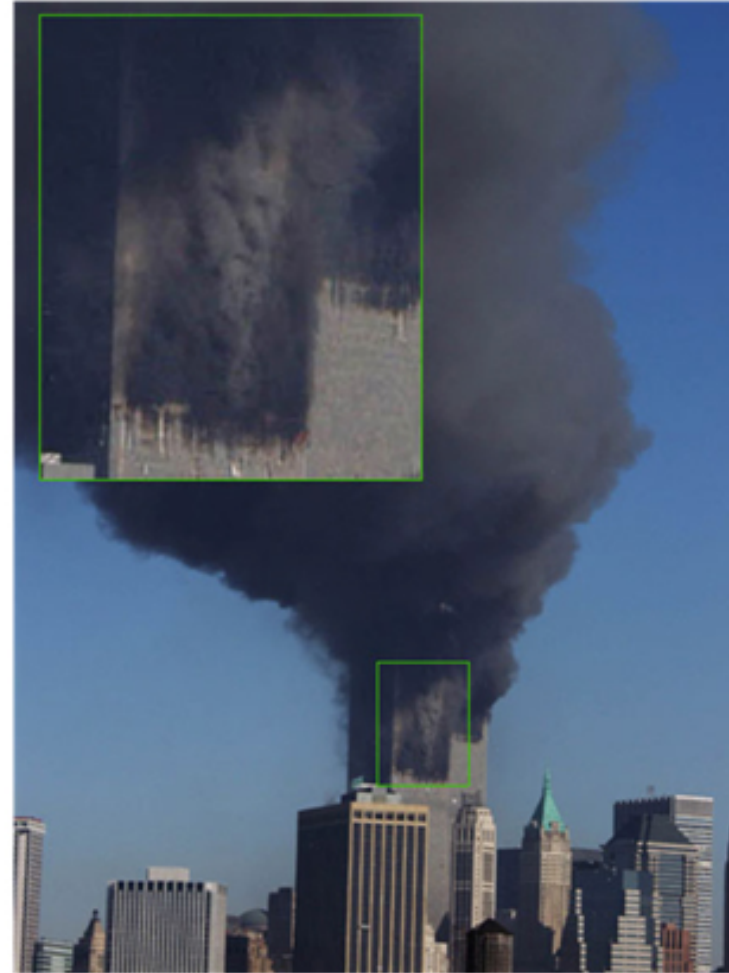
Pareidolia

A psychological phenomenon involving a stimulus (an image or a sound) wherein the mind perceives a familiar pattern of something where none actually exists.

Common examples are perceived images of animals, faces, or objects in cloud formations, the "*man in the moon*", the "*moon rabbit*", and hidden messages within recorded music played in reverse or at higher- or lower-than-normal speeds.

Examples





Often used to support religious beliefs and/or conspiracy theories

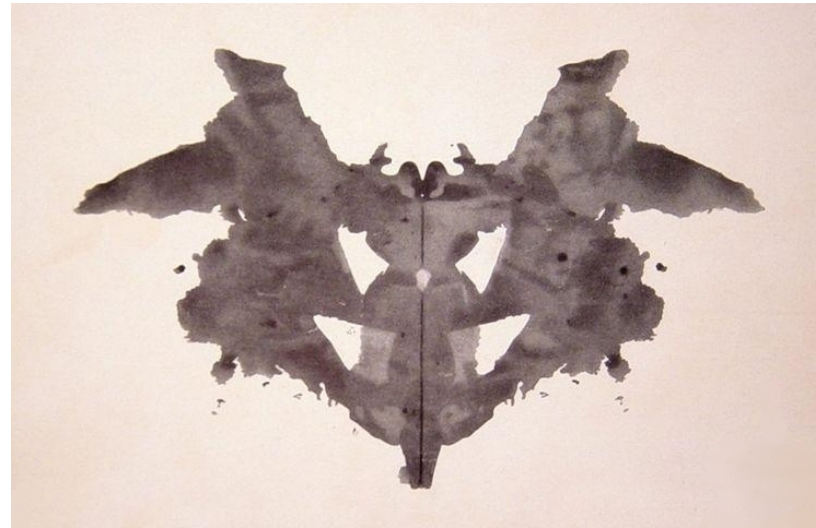
Pareidolia as a device for an artist

“If you look at any walls spotted with various stains or with a mixture of different kinds of stones, if you are about to invent some scene you will be able to see in it a resemblance to various different landscapes adorned with mountains, rivers, rocks, trees, plains, wide valleys, and various groups of hills. You will also be able to see divers combats and figures in quick movement, and strange expressions of faces, and outlandish costumes, and an infinite number of things which you can then reduce into separate and well conceived forms.”

Leonardo da Vinci



The Jurist by Giuseppe Arcimboldo, 1566



The Rorschach inkblot test uses pareidolia in an attempt to gain insight into a person's mental state. The Rorschach is a projective test, as it intentionally elicits the thoughts or feelings of respondents which are "projected" onto the ambiguous inkblot images. Projection in this instance is a form of "*directed pareidolia*"

Pareidolia and computer vision



When passed through the **DeepDream** program, an image of pies on a market stall shows eyes and the faces of dogs

References

<https://en.wikipedia.org/wiki/Pareidolia>

<http://gestaltrevision.be/en/master-index/66-what-we-do/overview/research-areas/mid-level/multistability/71-pareidolia>

SHORT-TERM MEMORY

Short-Term Memory

- Attention/focus transfers information from sensory memory to short-term memory
- Lasts from a few seconds to a minute
- Limited storage capacity
 - Minimum: 5 elements
 - Average: 7 elements
 - Maximum: 9 elements

Practical Indications

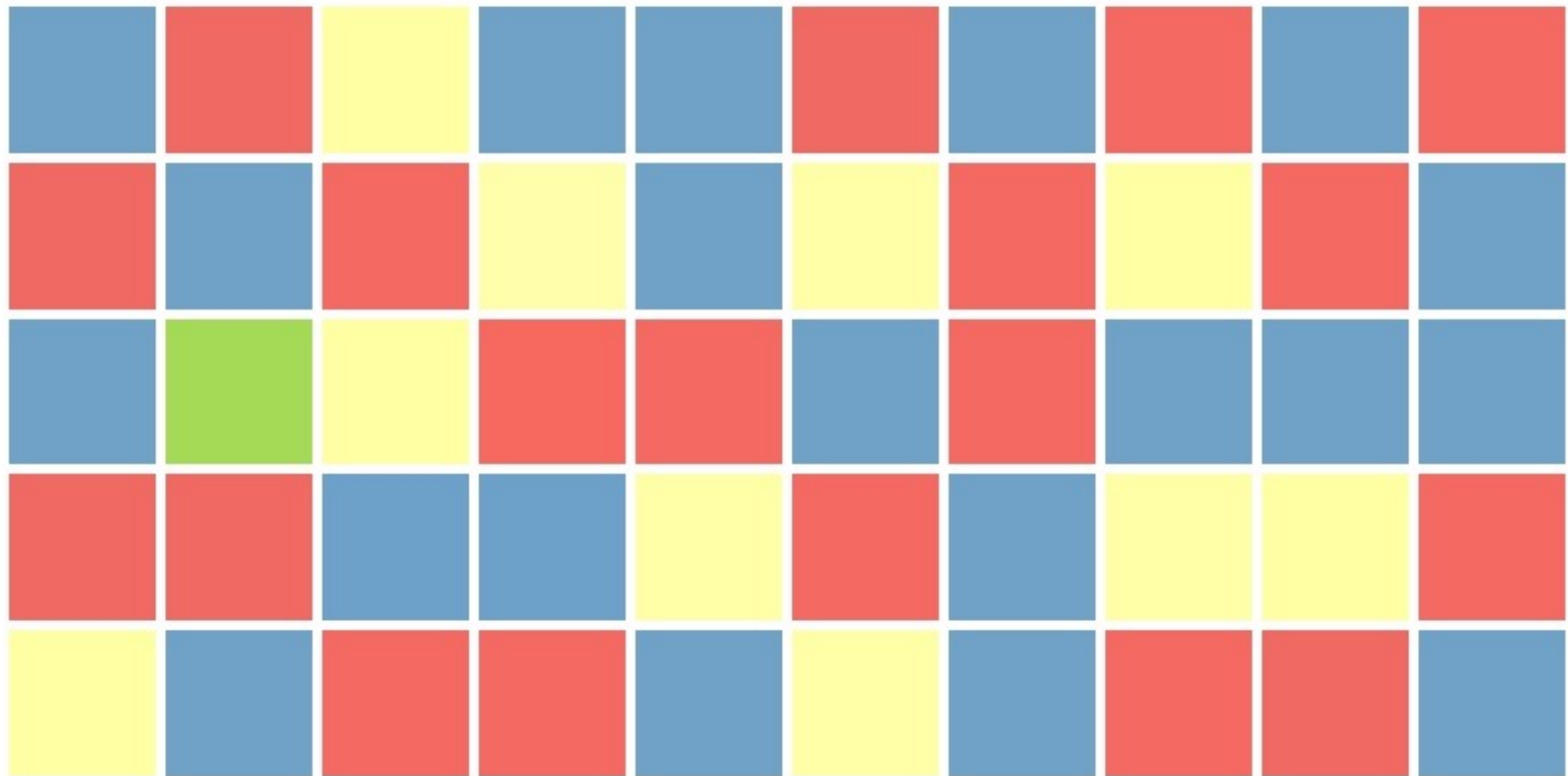
- Can reliably use 5 distinct attributes
- Should use no more than 7 to be accessible
 - No more than 7 distinct colors or shapes
- Attributes are cumulative
 - 3 shapes, 4 colors = 7 attributes
- Once lose focus, forget information
 - Distraction is costly

Grouping

- Grouping/chunking can increase capacity
 - 4154224174 versus (415) 422-4174
- Group sizes must be kept small
- Grouping can also improve speed of processing

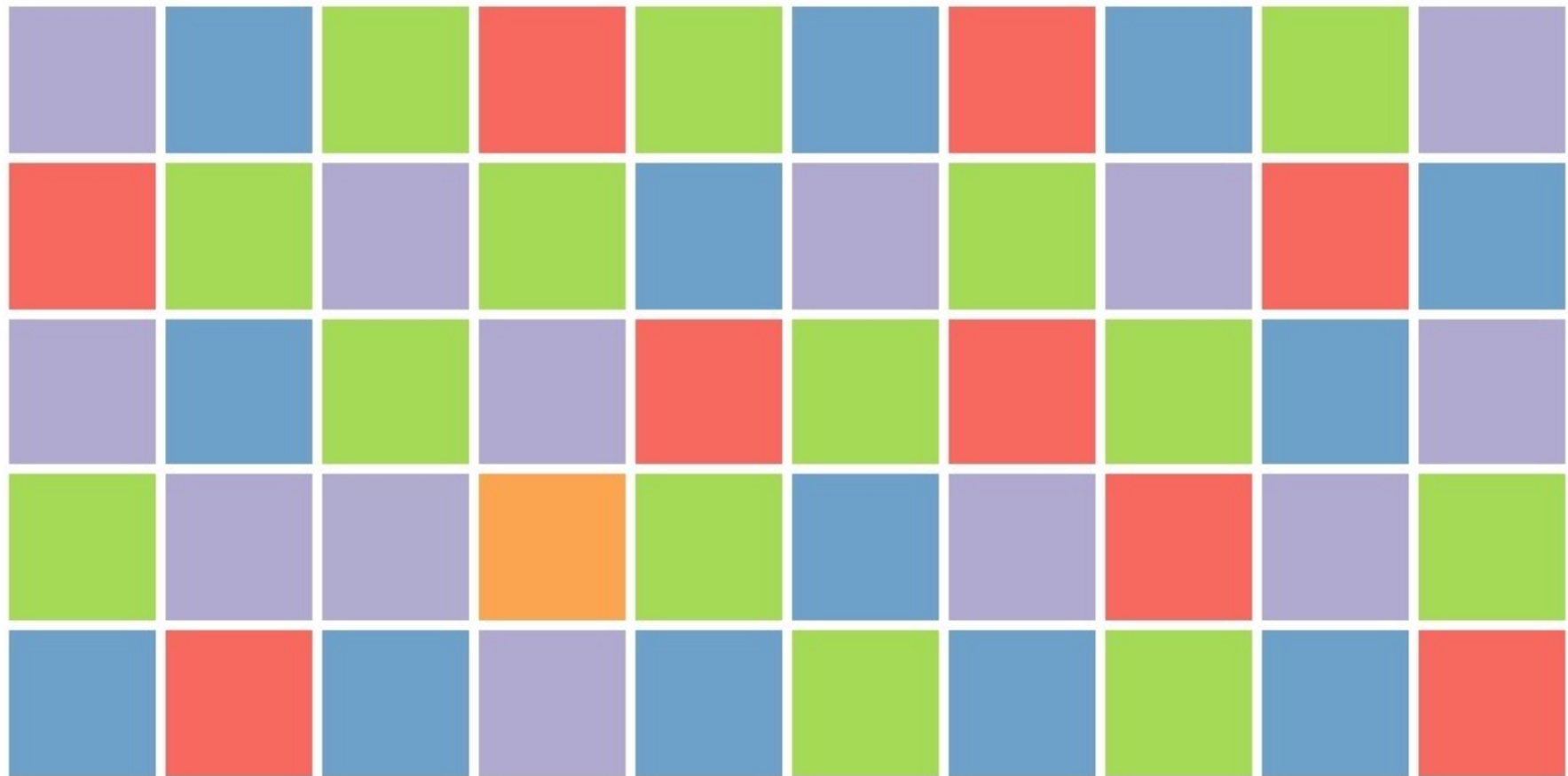
Grouping

- Find the unique color



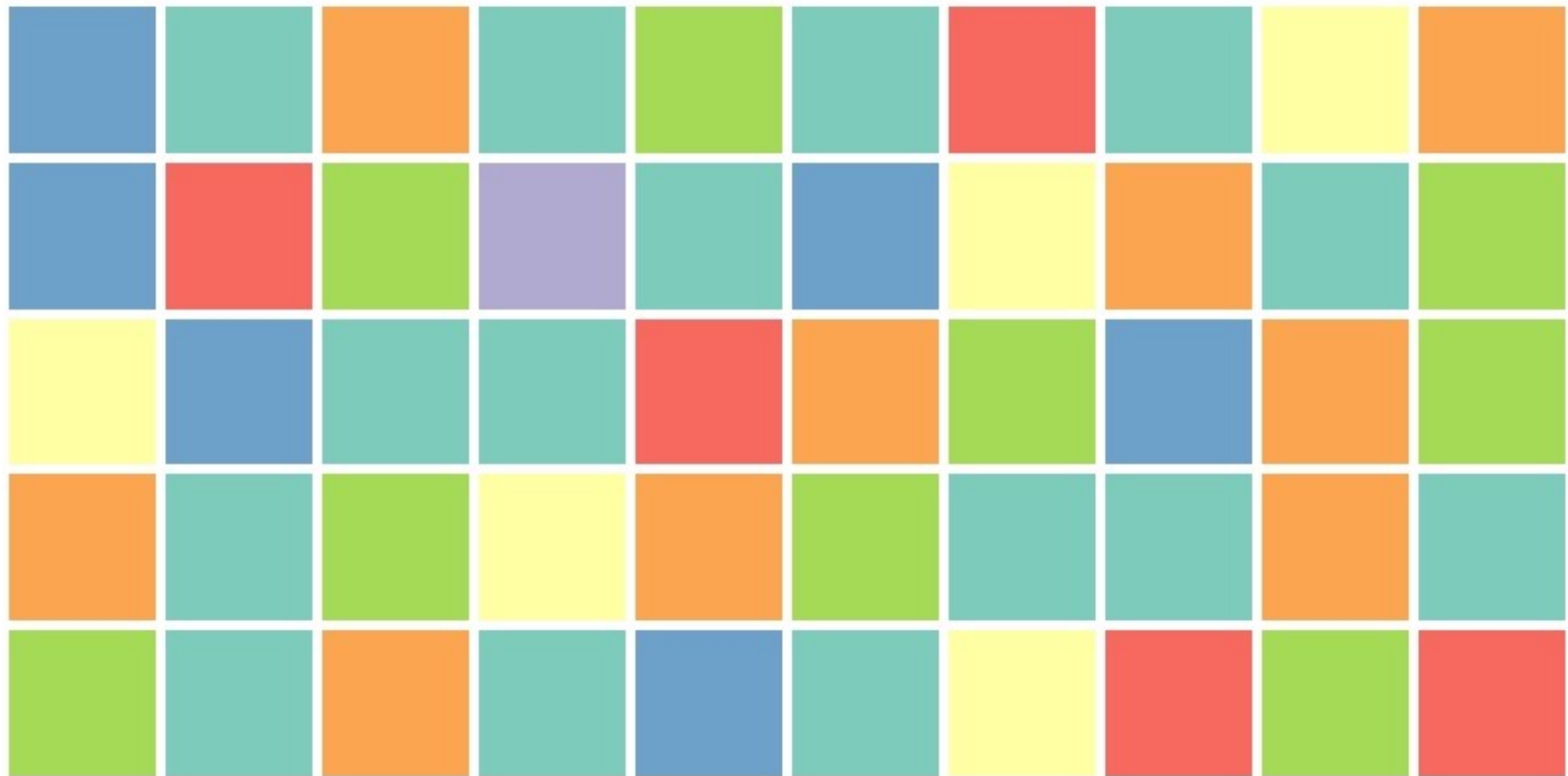
Grouping

- Find the unique color



Grouping

- Find the unique color



Grouping

- Find the unique color

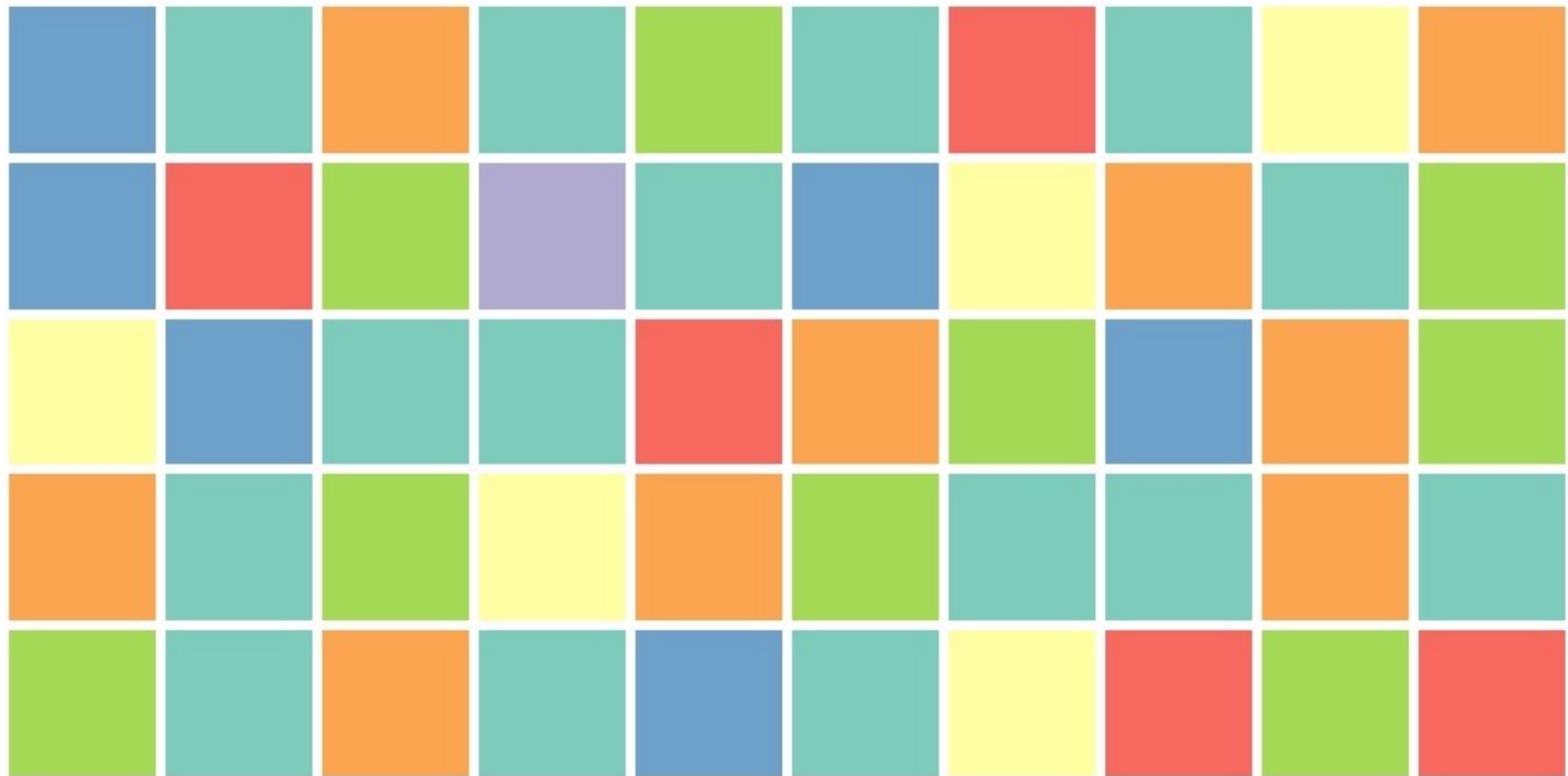


Grouping

- Improves ability to detect outliers
- Especially important as short-term capacity is strained (approaching 7 colors)
- Works for other pre-attentive attributes (e.g. motion video)
- Does not seem to help with search tasks

Grouping

- Find all of the red squares



Grouping

- Find all of the red squares



Change Blindness

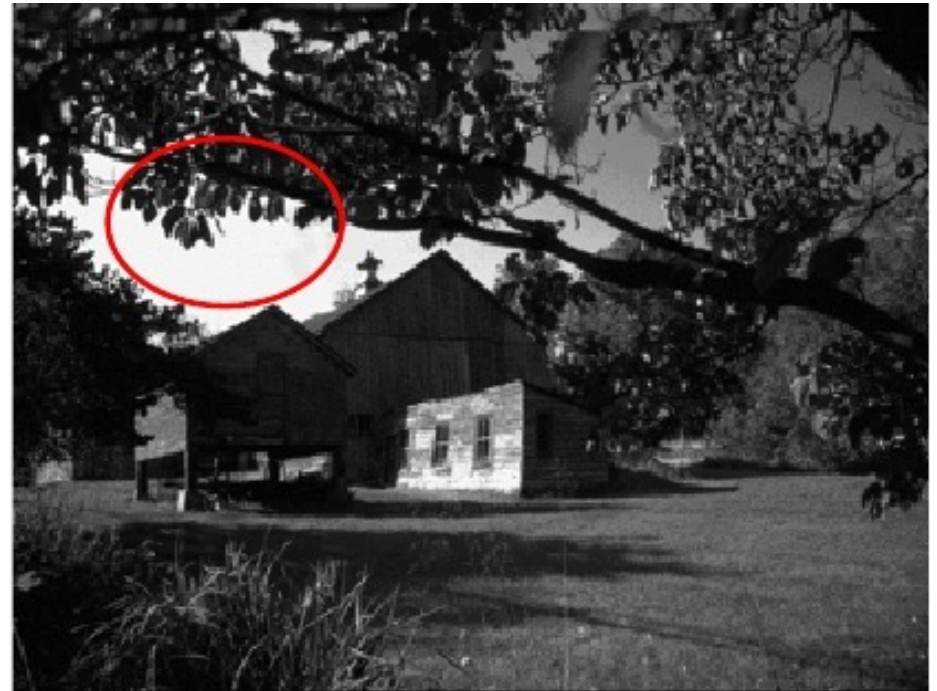
- To notice change, must pay attention to or focus on area of change
- Can break focus with flicker, making it difficult to detect change
- For visualization, must be careful to direct the eye where it is important

Change Blindness



<http://www.csc.ncsu.edu/faculty/healey/PP/>

Change Blindness



REFERENCES

Stephen Few, *Now You See It*, Analytics Press
Chapter 3: Thinking with our eyes

References

Attention and Visual Memory in Visualization and Computer Graphics

Christopher Healey and James. T. Enns, in IEEE Transactions on Visualization and Computer Graphics (IEEE TVCG), Volume 18, Issue 7, Pages 1170–1188, July 2012.

DOI: [10.1109/TVCG.2011.127](https://doi.org/10.1109/TVCG.2011.127) URL: <http://steveharoz.com/research/attention/>

How Capacity Limits of Attention Influence Information Visualization Effectiveness

Steve Haroz and David Whitney, in IEEE Transactions on Visualization and Computer Graphics (IEEE TVCG), Volume 18, Issue 12, Pages 2402–2410, December 2012.

DOI: [10.1109/TVCG.2012.233](https://doi.org/10.1109/TVCG.2012.233) URL: <http://www.csc.ncsu.edu/faculty/healey/PP/>

Questions?

Design Principles

How to design visualizations

Designing Visualizations

- What
 - What data are you visualizing?
- Why
 - Why are you visualizing that data?
 - Why would others use your visualization?
- How
 - How will you encode the data?
 - How will you implement the visualization?

Type of Data

- Multivariate Data
- Text Data
- Temporal Data
- Geospatial Data
- Hierarchical Data
- Network Data
- Numerical Data
- Categorical Data

Visualization Purpose

- Convey complex information
- Capture attention and raise awareness
- Create something aesthetically pleasing
- Encourage exploration

Visualization Task

- Quickly identify outliers
- Quickly identify groups/classes
- Quickly identify problems
- Explore data to gain insight
- Identify complex patterns

Encoding Data

- Map data to pre-attentive attributes
- Keep in mind perception
 - Which attributes are stronger?
 - How many distinct attributes can you use?
- Revisit if how encoding is perceived matches underlying data

Implementation

- Static Visualizations
 - Python with matplotlib
 - R with ggplot2
- Interactive Visualizations
 - D3.js
 - Python with plot.ly
 - R with Shiny
 - Processing.js
- Animated Visualizations and new metaphors
 - Processing

Evaluation

- Does the visualization achieve your purpose?
- Can the users achieve their visualization task?
 - Quickly?
 - Accurately?
- Evaluate and iterate

GUIDELINES

Information Visualization by Colin Ware

Random Selection of Tips

- [G1.2] Important data should be represented by graphical elements that are more visually distinct than those representing less important information.
- [G1.6] Consider adopting novel design solutions only when the estimated payoff is substantially greater than the cost of learning to use them.
- [G3.1] Avoid using grayscale as a method for representing more than two to four values.

Random Selection of Tips

- [G4.1] Use more saturated colors when color coding small symbols, thin lines, or other small areas. Use less saturated colors for coding large areas.
- [G5.6] Use strong pre-attentive cues before weak ones where ease of search is critical.
- [G10.6] Consider providing a small overview map to support navigation through a large data space.


General Guidelines

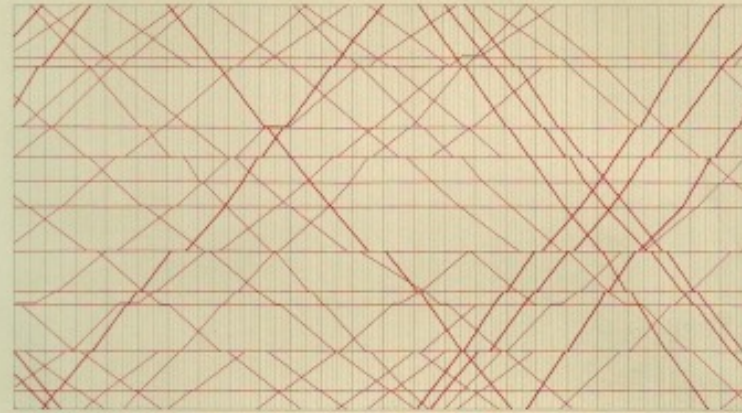
- More found in book by Colin Ware
 - Information Visualization: Perception for Design, 3rd Edition, 2013

Graphical Excellence

Tufte on Design and Evaluation

Resources

- Envisioning Information
by Edward R. Tufte, Graphics Press, 1990
- Visual Explanations
by Edward R. Tufte, Graphics Press, 1997 
- The Visual Display of Quantitative Information
by Edward R. Tufte, Graphics Press, 2001



SECOND EDITION

The Visual Display
of Quantitative Information

EDWARD R. TUFTE

GRAPHICAL EXCELLENCE

VDQI Chapter 1

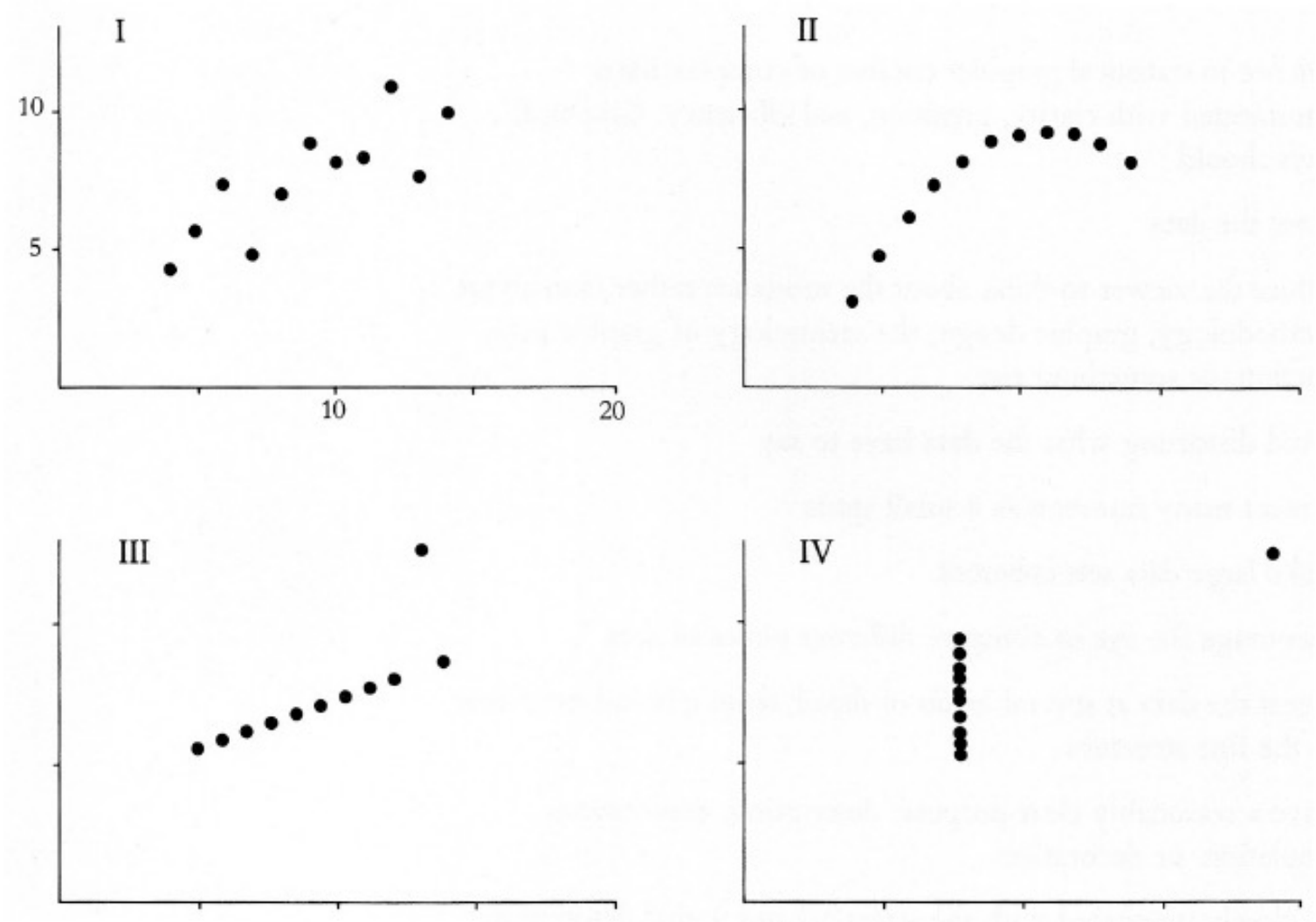
Graphics reveal data.

VDQI Example (p13)

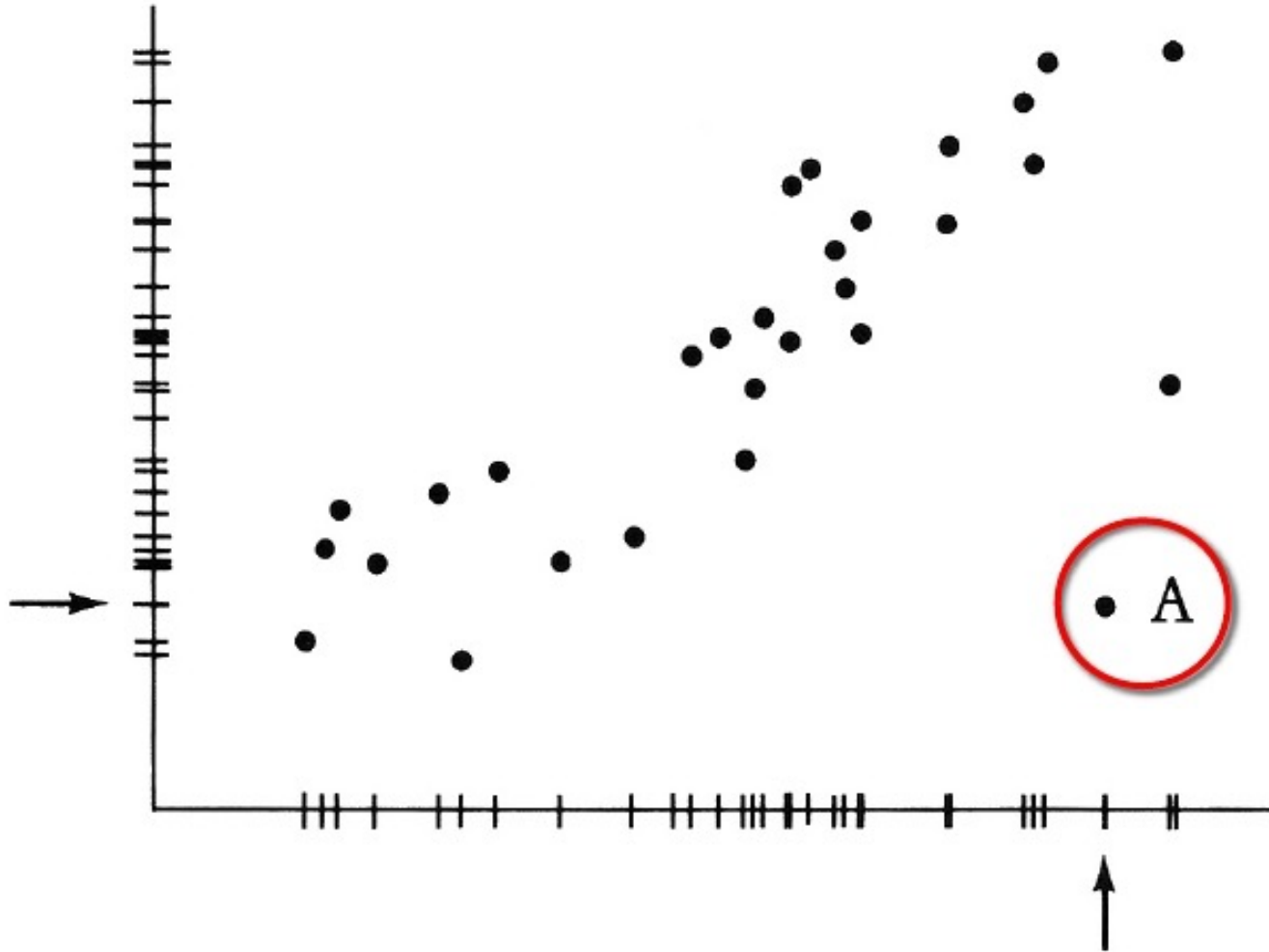
I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0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

$N = 11$
 mean of X's = 9.0
 mean of Y's = 7.5
 equation of regression line: $Y = 3 + 0.5X$
 standard error of estimate of slope = 0.118
 $t = 4.24$
 sum of squares $X - \bar{X} = 110.0$
 regression sum of squares = 27.50
 residual sum of squares of Y = 13.75
 correlation coefficient = .82
 $r^2 = .67$


VDQI Example (p13)



VDQI Example (p14)

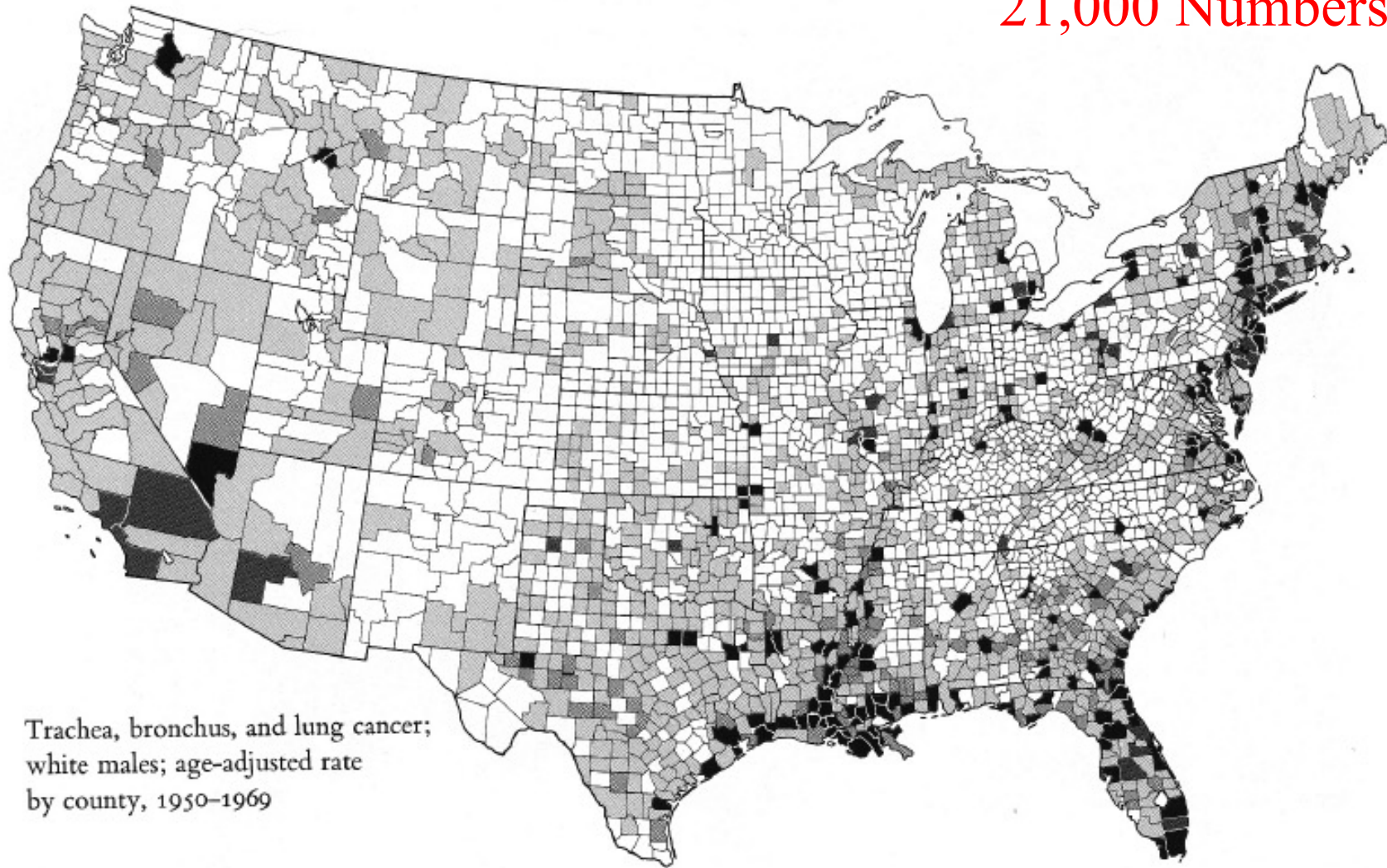


Graphical Excellence

- Avoid distorting the data
- Reveal data at different levels of detail
- Present many numbers in a small space 
- Encourage comparison
- Encourage viewers to think about the data, not graphic design, etc.

VDQI Example (p18)

21,000 Numbers



Trachea, bronchus, and lung cancer;
white males; age-adjusted rate
by county, 1950-1969

VDQI Example (p37)

700 Words

FRANKED MAIL TIE TO VOTING SHOWN

Testimony Finds the Volume Rises Before Elections

WASHINGTON, June 1 (AP) —New court testimony and documents show that much of the mail Congress sends at taxpayer expense is tied directly to the re-election campaigns of Senate and House members. According to material filed in a lawsuit in Federal Court:

¶Senate Republicans put two direct-mail experts on the public payroll to advise them on how to use their free mailing privileges to get votes.

¶An election manual prepared for Senate Democrats refers to newsletters as a "free forum," and sets up a timetable

for sending them as an integral part of a model re-election campaign.

¶Senator John G. Tower, Republican of Texas, mailed more than 800,000 special-interest letters at taxpayer expense as part of his 1972 re-election effort and received campaign volunteer offers and donations in response.

¶Senator Jacob K. Javits, Republican of New York, gave written approval in 1973 for a tax-paid mail program intended to better his image and pay off at the polls. He focused his mail on areas where he needed votes.

¶The volume of "official" Congressional mail rises in election years and peaks just before the general election.

None of this activity necessarily violates any law or regulation, since Congress has wide discretion in the use of tax-paid mail. Congress gave itself the right to send official mail at Government expense at the

founding of the republic, and only Congress polices against abuses of the free mailings.

Complaints of political use of the free-mailing privilege, called the franking privilege, are heard every election year. Recently, however, the volume and cost of franked mail has multiplied. A new Federal law will limit what out-of-office challengers can spend to unseat incumbents.

In 1972, Congress passed a law prohibiting mass franked mailings within 28 days before an election. The sponsor of that legislation, Representative Morris K. Udall, Democrat of Arizona, said in an interview that further changes were needed to curtail political abuse of the frank.

Mr. Udall urged a 60-day pre-election cutoff for mass mailings and said he favored closing a loophole that recently allowed defeated Representative Frank M. Clark, Democrat of Pennsylvania, to send a

franked newsletter to his old constituents after he had left office. Mr. Clark is seeking to regain his old post.

Practice Documented

Seldom has the political use of franked mail been so well documented as in recent testimony and documents filed in a Federal Court by Common Cause, the lobby group, which is suing for an end to tax-financed mass mailings by Congress.

For example, Joyce P. Baker, a political mail specialist, said in a 1973 job proposal that she wanted to set up direct-mail programs for Republican Senators using franked mail.

"The purpose of such a program is to help an incumbent Senator get re-elected," she said.

She was put on the Senate payroll at \$18,810 a year in 1973 and 1974 and testified that during that time she aided Republican Senators Robert J.

Dole of Kansas, Peter H. Dominick of Colorado, Charles McMathias Jr. of Maryland

Another political mail specialist, Lee W. MacGregor, wrote a proposal for the use of franked mail by his chief, Senator Javits, in 1973.

"The over-all objective of the franked mail program can be to get the recipient of the mail to identify positively with a particular stand you have taken or a bill you have introduced; the kind of identification that can be translated into a vote at the polls on election day" Mr. MacGregor said.

Mr. Javits was out of the country and could not be reached. His administrative assistant, Donald Kellerman, defended the use of franked mail.

"It is a standard device to let voters, not voters but citizens, know what the Senator is doing here in Washington," he said.

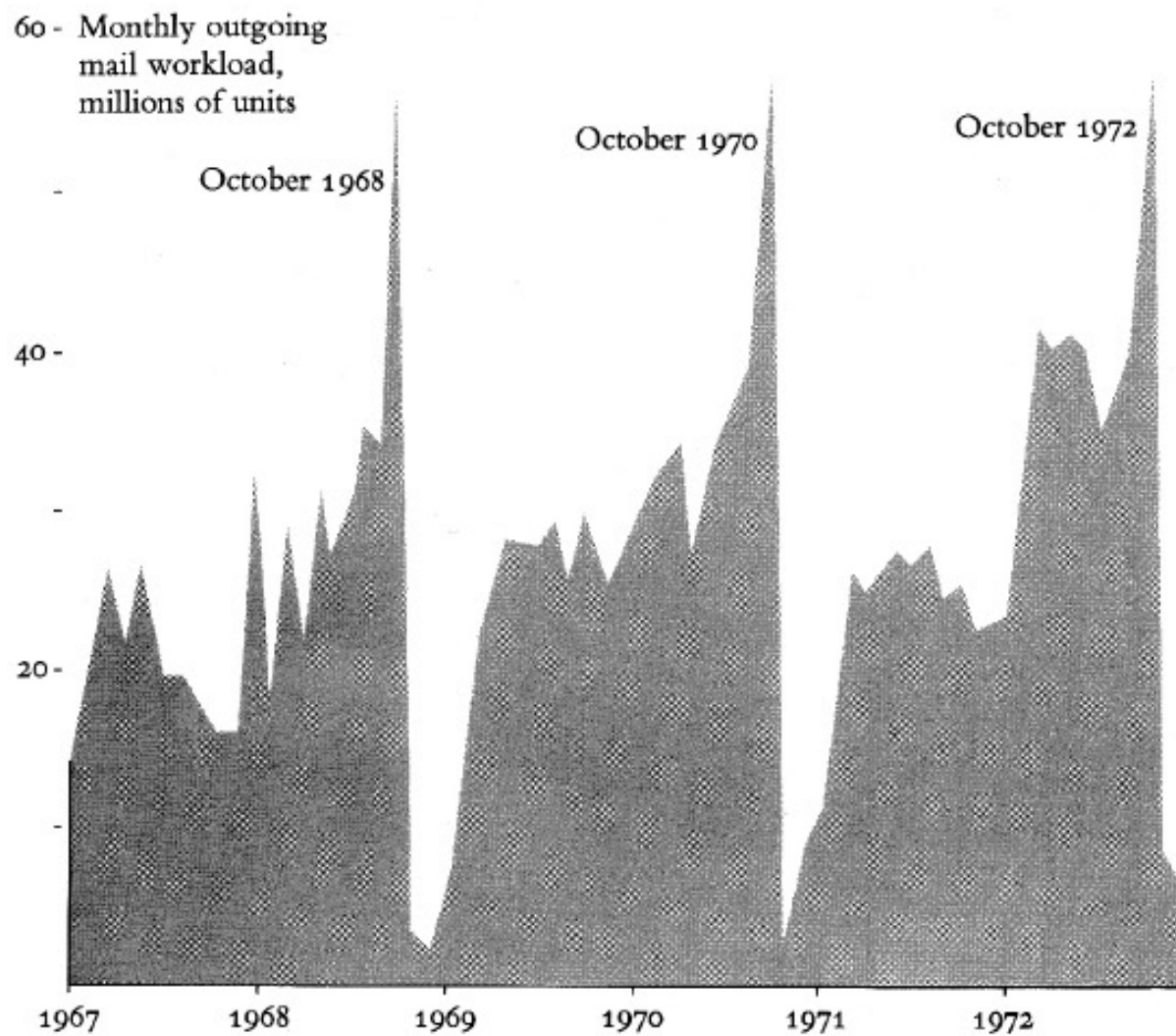
Senator Tower's use of franked mail in his 1972 campaign was documented by memorandums.

Tom Loeffler, a high-ranking campaign aide, wrote in a memorandum dated Oct. 27, 1972, that during the campaign Senator Tower had sent "31 special interest letters totaling approximately 803,333 franked mailings."

Mr. Tower was not available for comment. His administrative assistant, Elwin Skiles, said the Senator's use of franked mail in 1972 was within the law, and he defended the free-mailing privileges.

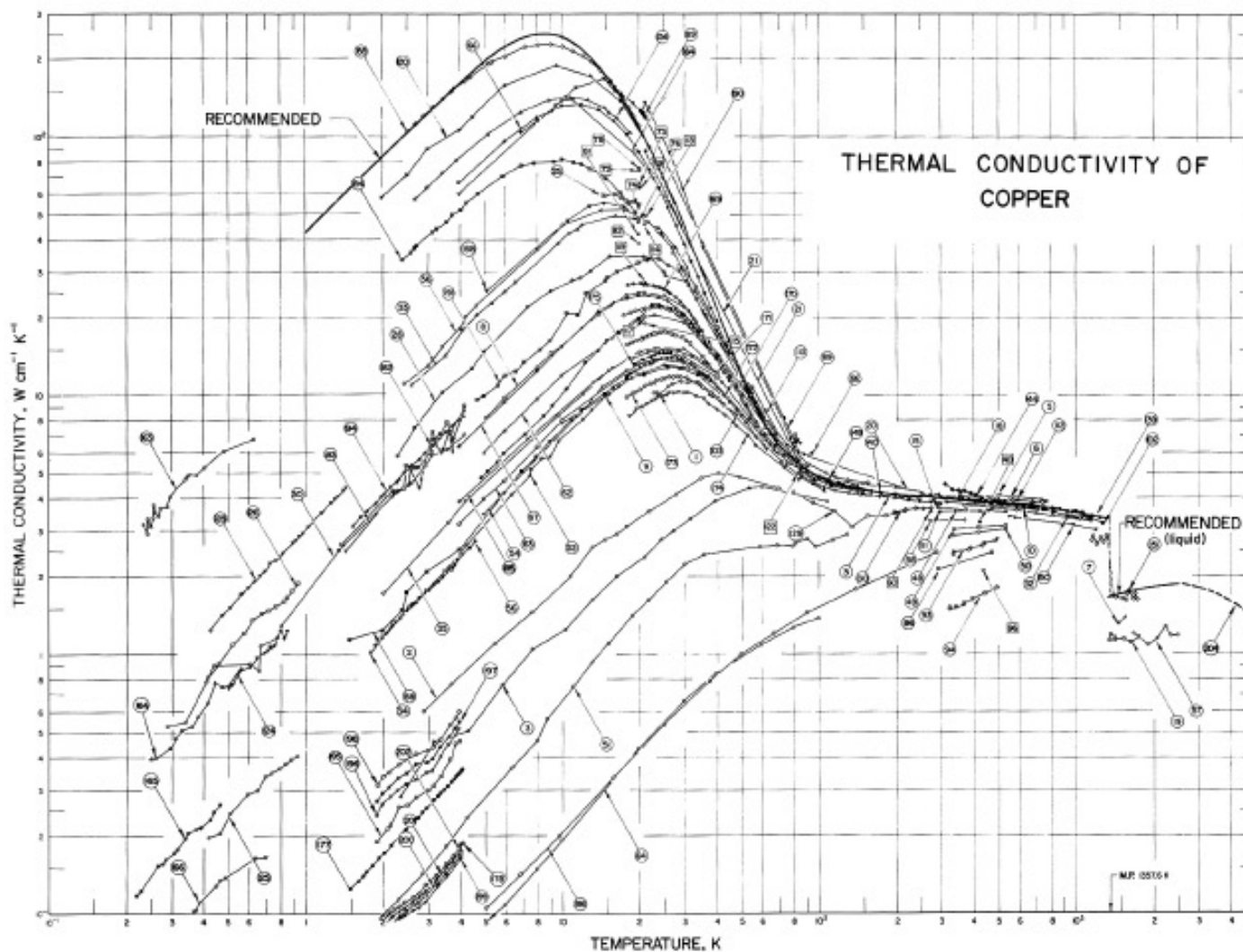
Postal Service figures show that in the 12 months before November, 1973, Congress sent 222.9 million franked pieces of mail. But in the next 12 months, covering the election season of 1974, Congress sent 350.6 million, a jump of 57 per cent about what's happening," Mr. Skiles said.

VDQI Example (p37)




VDQI Example (p49)

Hundreds of Publications



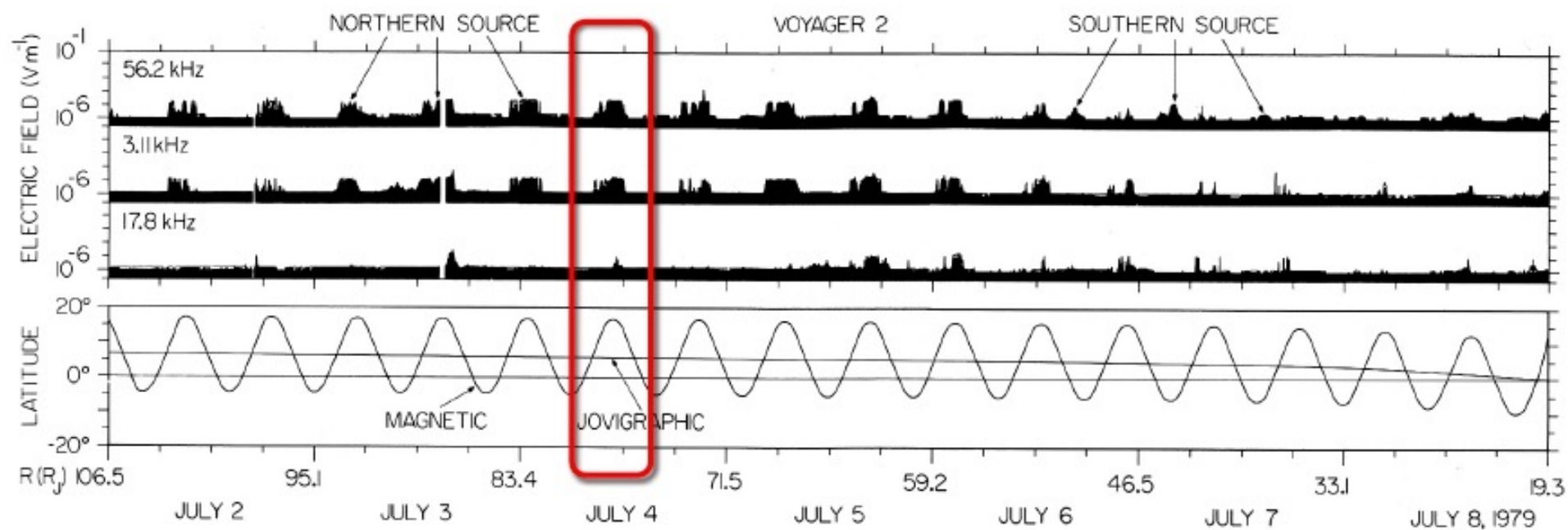
Graphical Excellence

- Avoid distorting the data
- Reveal data at different levels of detail
- Present many numbers in a small space
- Encourage comparison 
- Encourage viewers to think about the data, not graphic design, etc.

VDQI Example (p24)



VDQI Example (p29)



Graphical Excellence

- A matter of *substance, statistics, and design*
- Consists of ideas communicated with *clarity, precision, and efficiency*
- *Greatest* number of ideas in *shortest* time and *smallest* space
- Nearly always *multivariate*


Graphical Excellence

“...requires telling the truth about the data.”

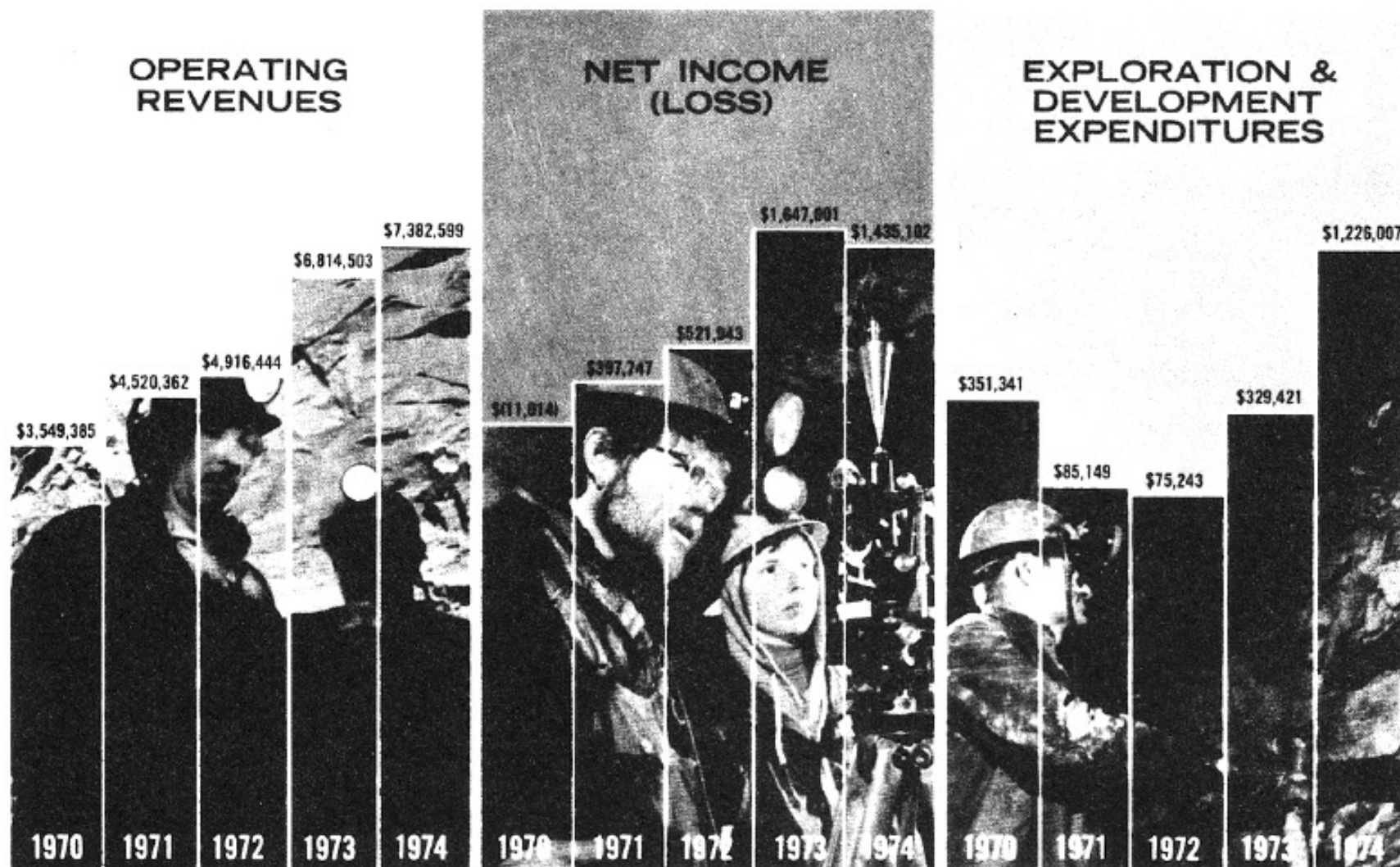
GRAPHICAL INTEGRITY

VDQI Chapter 2

Graphical Integrity

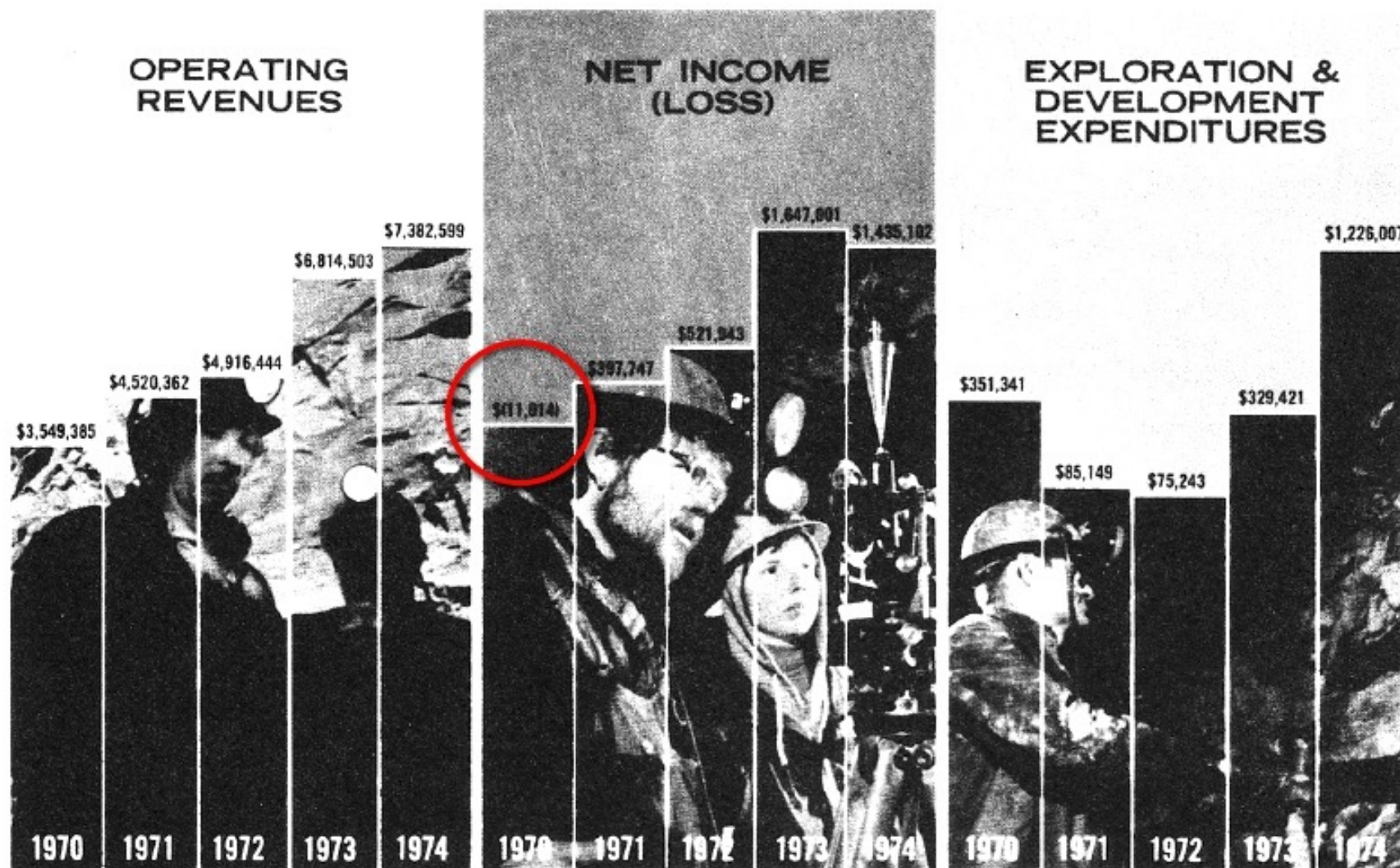
- Missing scales and labels 
- Missing context
- Distorting scales
- Distorting design

VDQI Example (p54)



VDQI Example (p54)

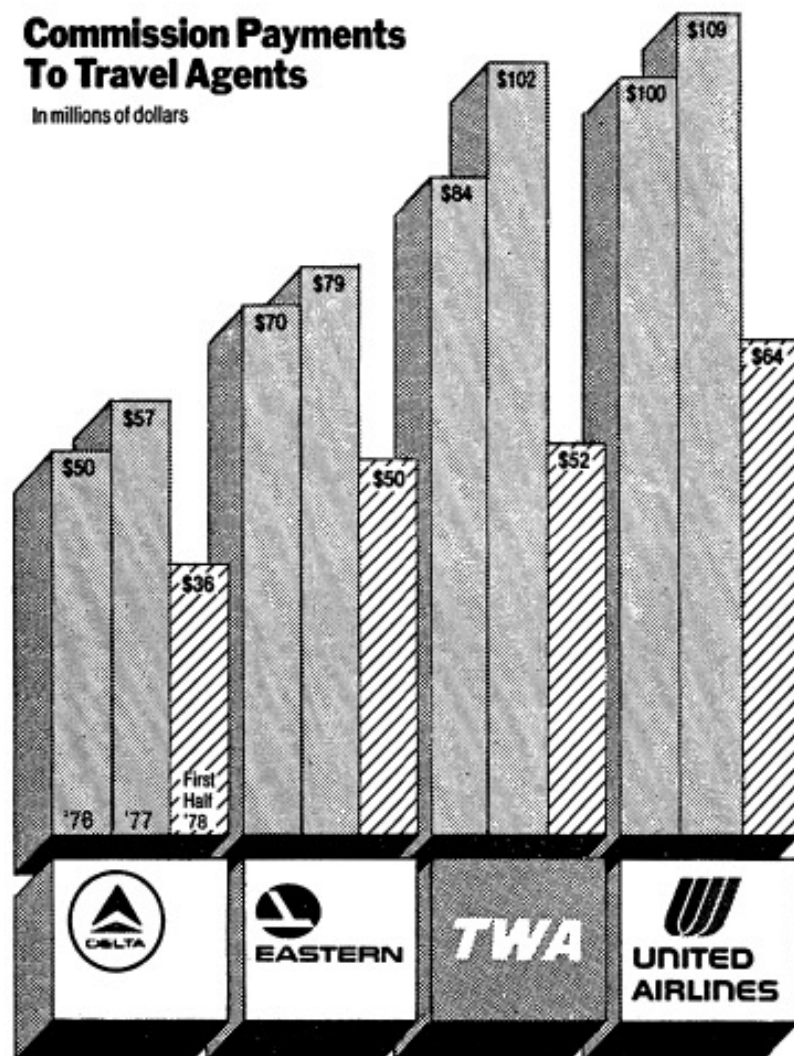
bottom of scale \approx $-\$4,200,000$



VDQI Example (p54)

Commission Payments To Travel Agents

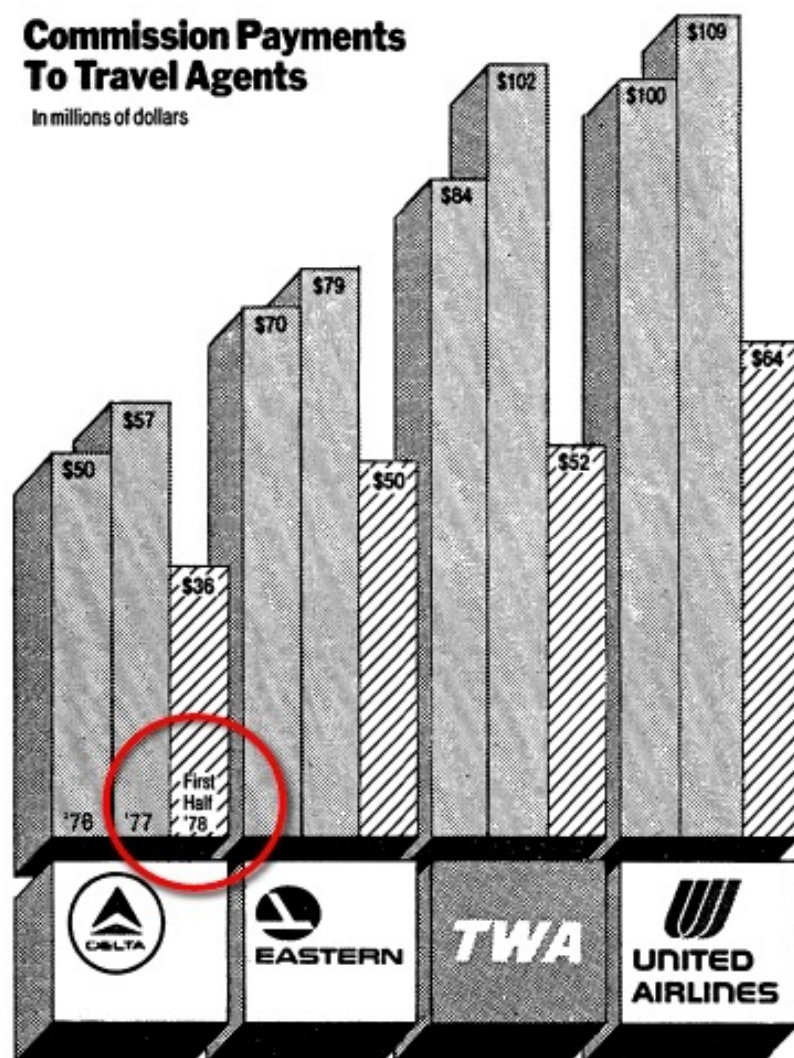
In millions of dollars



VDQI Example (p54)


Commission Payments To Travel Agents

In millions of dollars

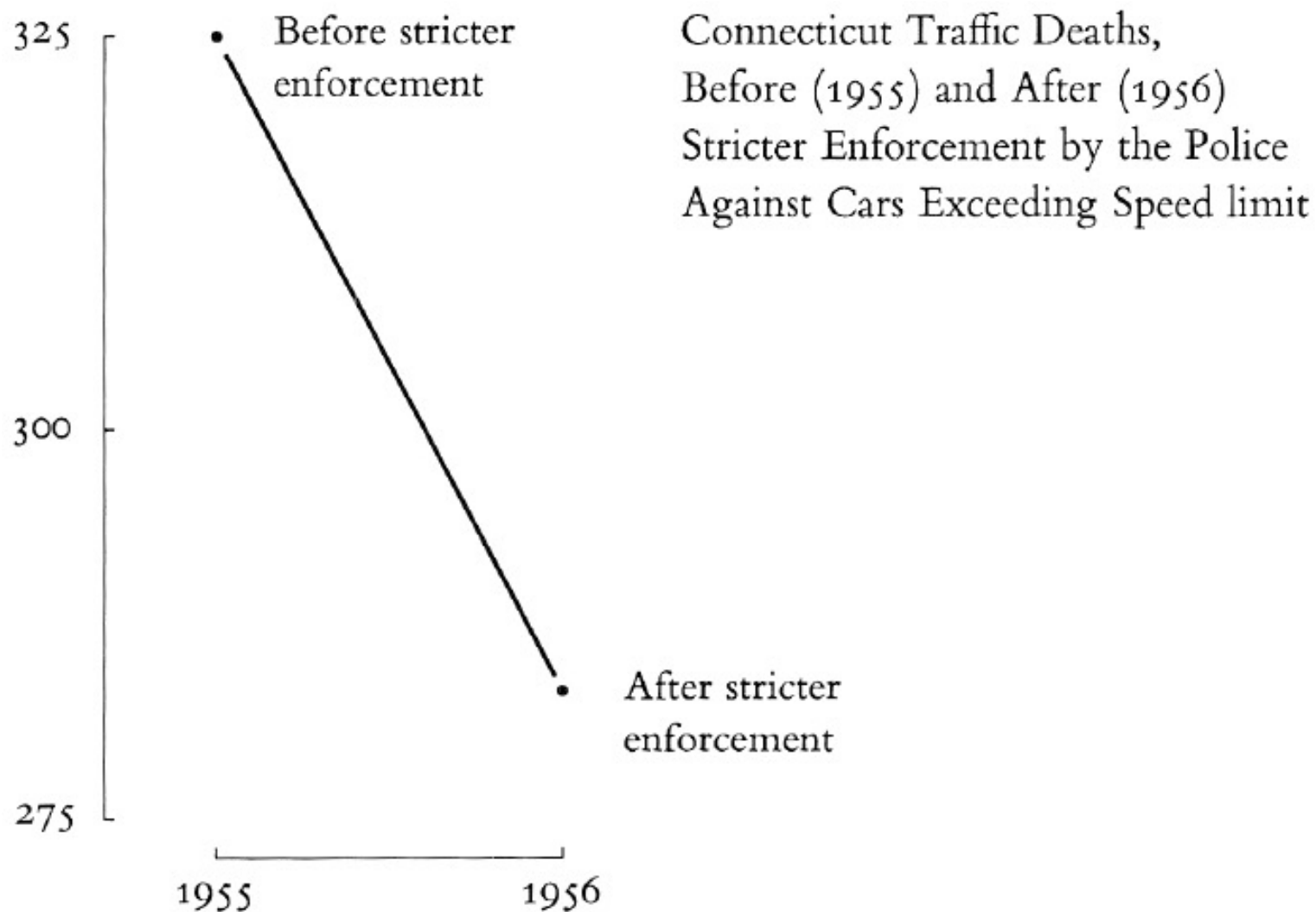


- “Pseudo-decline”
- Comparing full years (1976 and 1977) to half year (1978)

Graphical Integrity

- Missing scales and labels
- Missing context 
- Distorting scales
- Distorting design

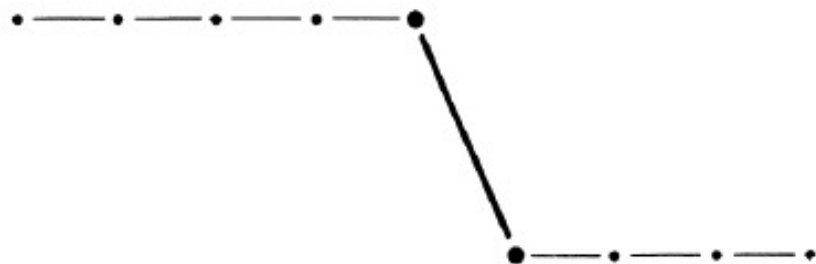
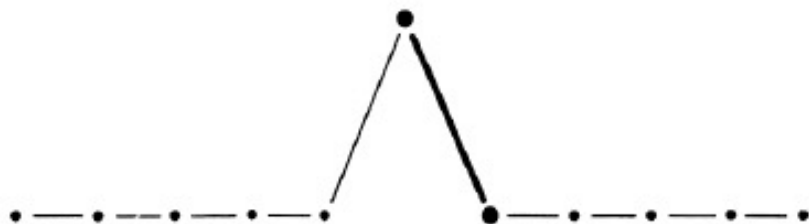
VDQI Example (p74)



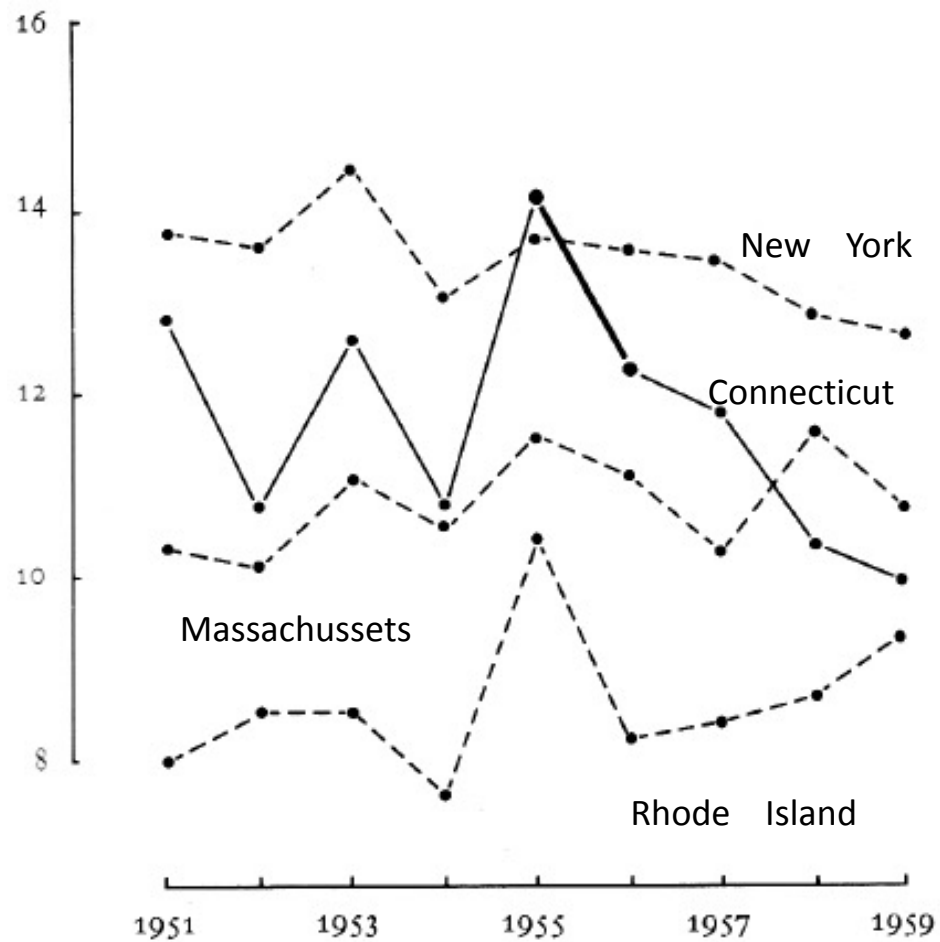
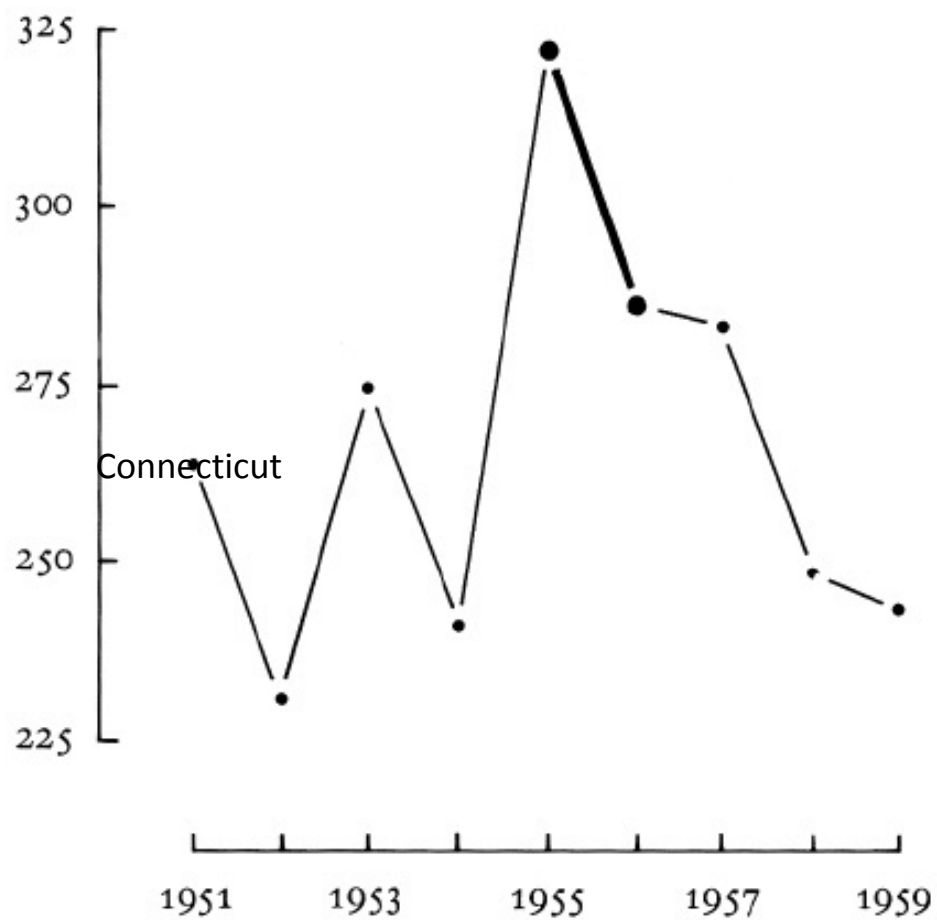
VDQI Example (p74)




- Missing context
- Which pattern did this segment come from?



VDQI Example (p74)

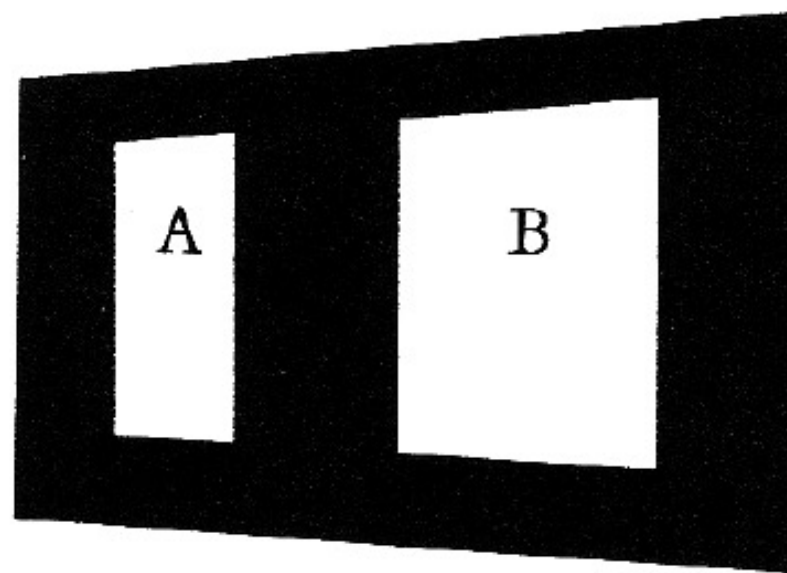
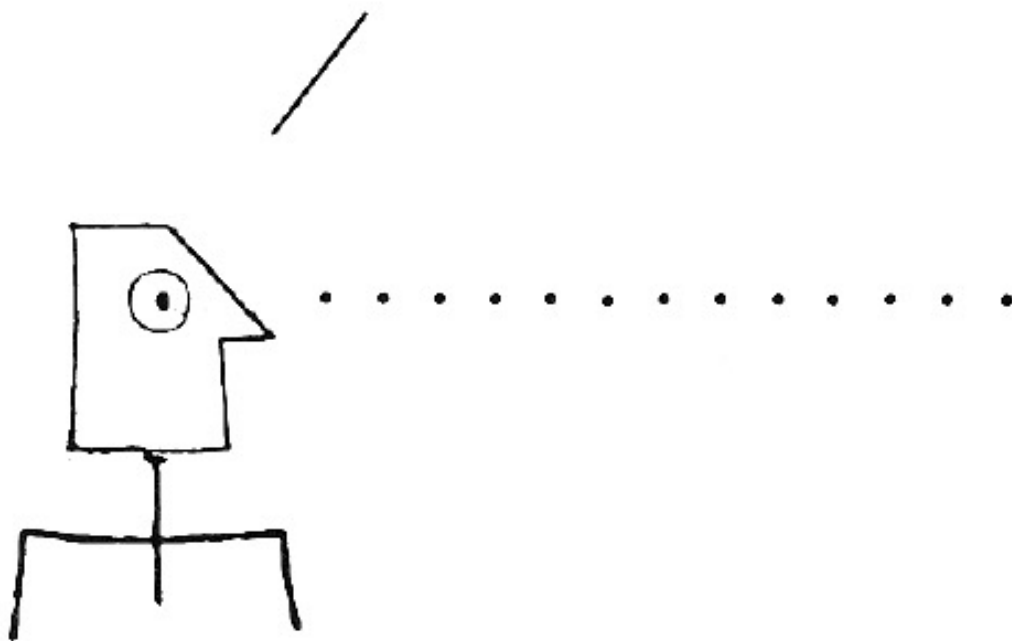


Graphical Integrity

- Missing scales and labels
 - Missing context
 - Distorting scales
 - Distorting design
- 

Distortion

I think I see that area B is 3.14 times bigger than area A. Is that correct?



Distortion: Problem

- Perceived area grows more slowly than actual area
- Perception changes per user
- Perception changes with experience
- Perception is context-dependent

Distortion: Solution*

- Physical representation should be directly proportional to numerical quantities
- Use clear and detailed labeling
- Label important events in data



Lie Factor

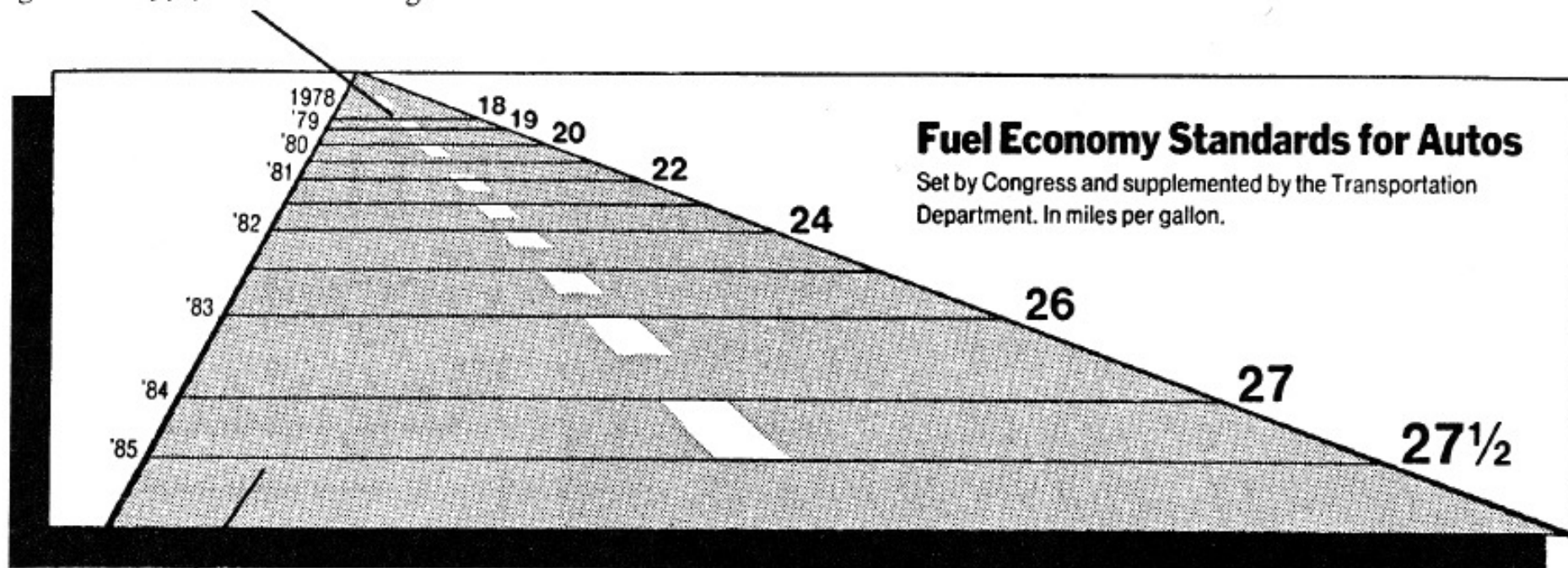
$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

Lie factor should be close to one
i.e. $1.05 > \text{Lie Factor} > 0.95$

VDQI Example (p57)

$$\frac{27.5-18.0}{18.0} = 0.53 \quad \frac{5.3-0.6}{0.6} = 7.83 \quad \text{Lie Factor} = \frac{7.83}{0.53} = 14.8$$

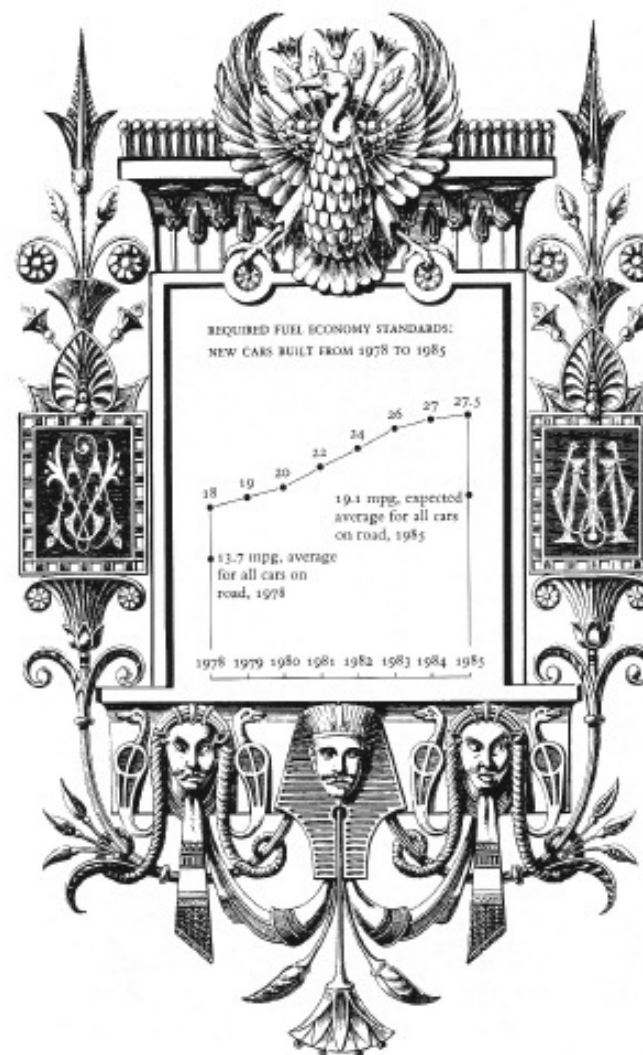
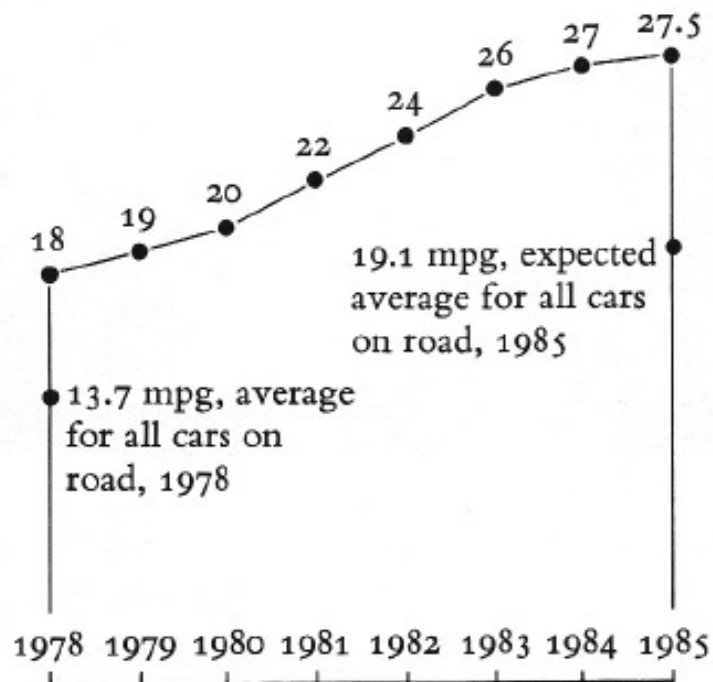
This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



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

VDQI Example (p58) – No Lies!

REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985

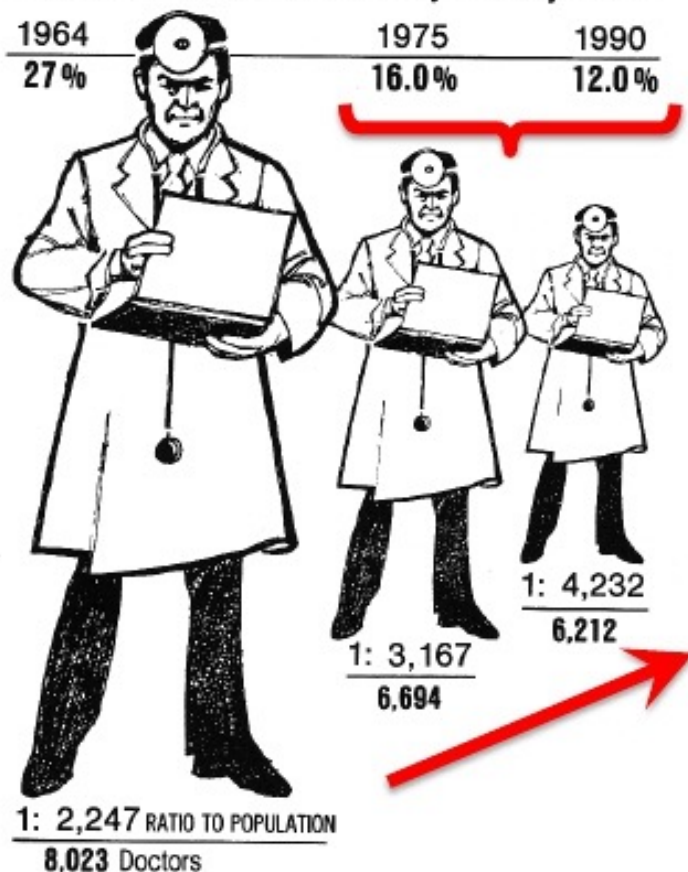


VDQI Example (p69)

THE SHRINKING FAMILY DOCTOR In California

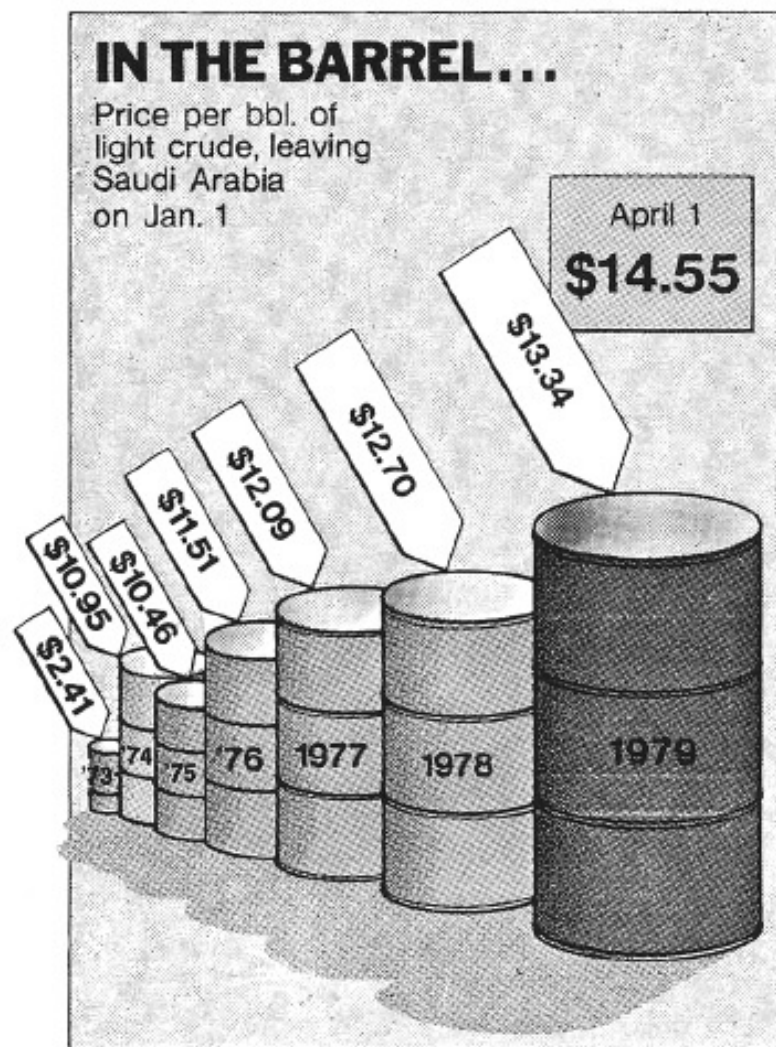
Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%




- Lie factor of 2.8
- Additionally:
 - Exaggeration from perspective
 - Incorrect horizontal spacing

VDQI Example (p71)



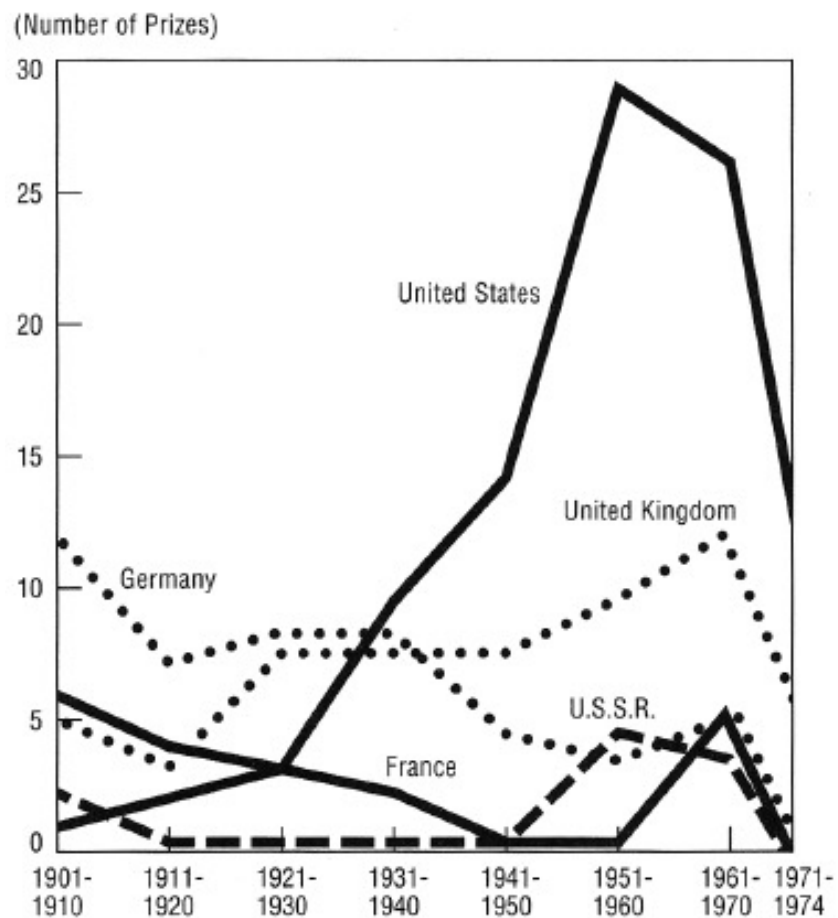
- Surface area lie factor:
9.4
- Volume lie factor:
59.4
- Don't use 2D or 3D to show 1D data!

Graphical Integrity

- Missing scales and labels
 - Missing context
 - Distorting scales
 - Distorting design
- 

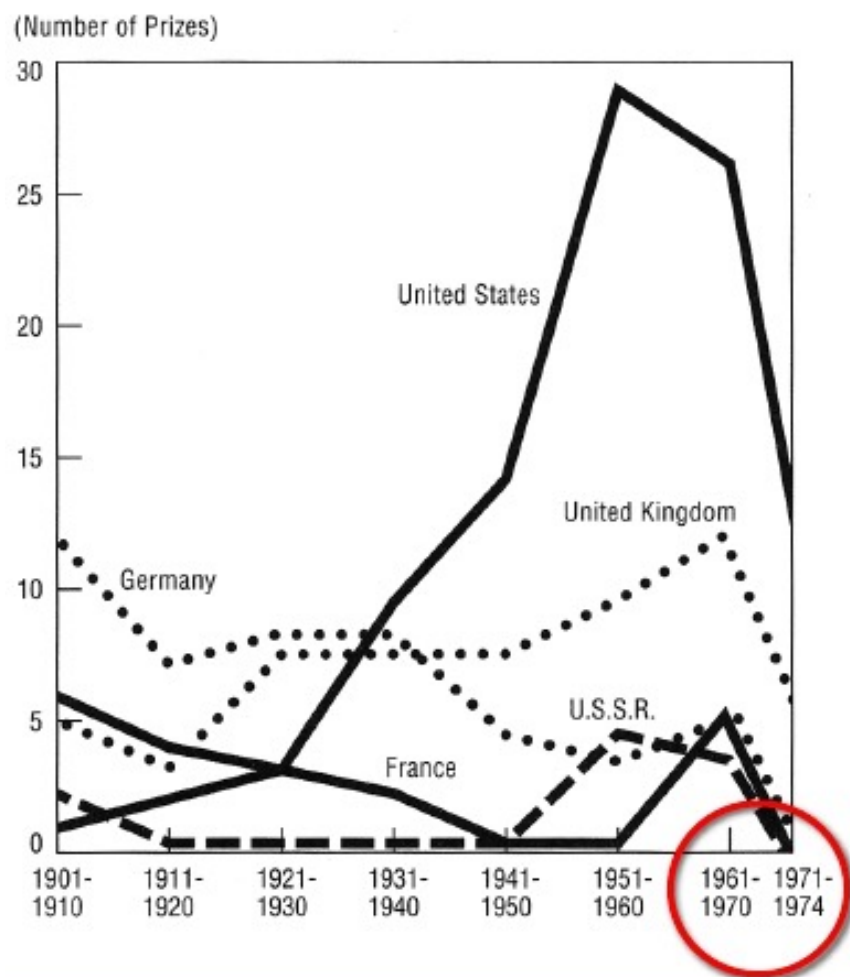
VDQI Example (p60)

**Nobel Prizes Awarded in Science,
for Selected Countries, 1901-1974**

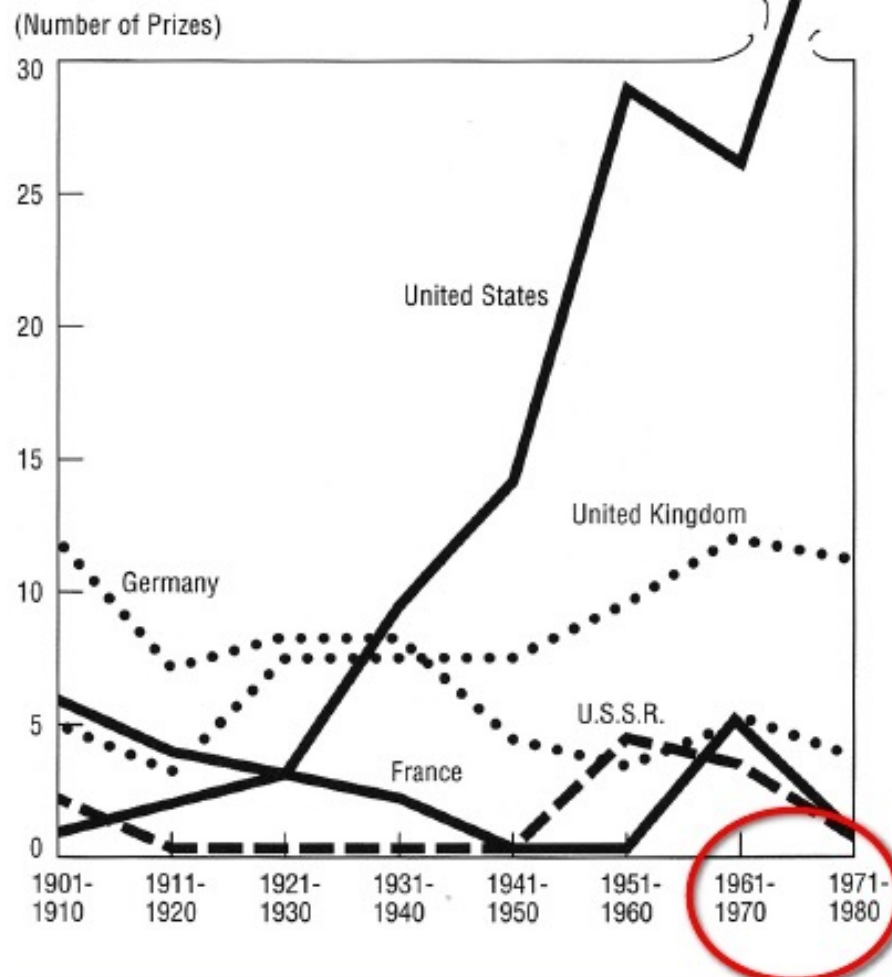


VDQI Example (p60)

**Nobel Prizes Awarded in Science,
for Selected Countries, 1901-1974**



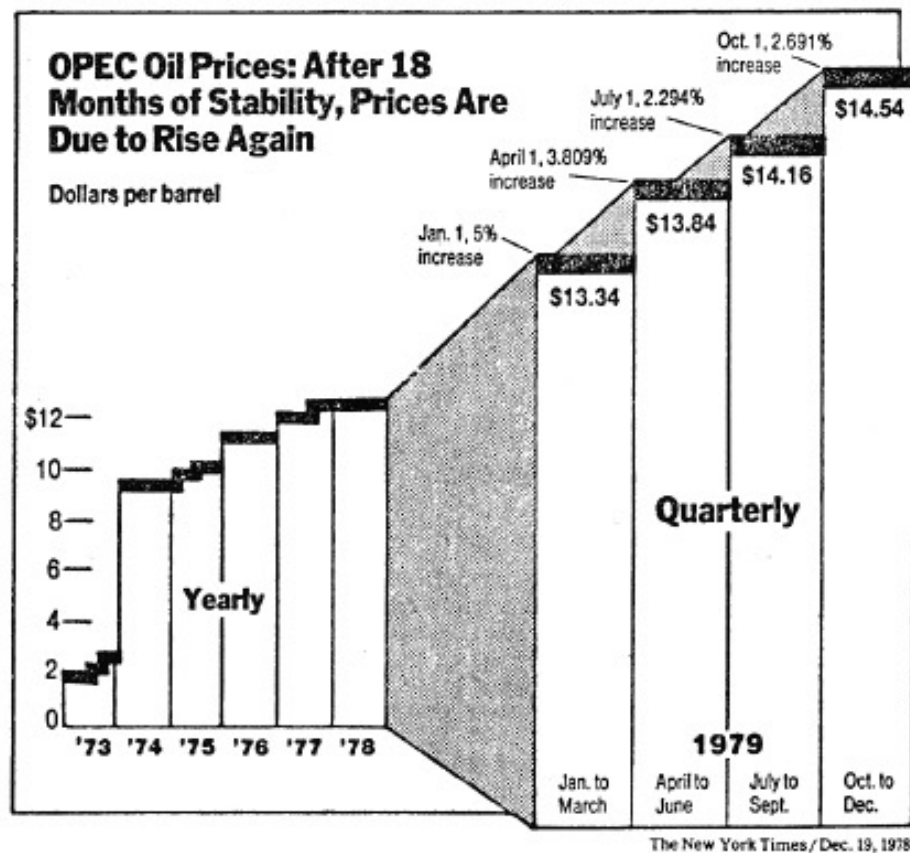
**Nobel Prizes Awarded in Science,
for Selected Countries, 1901-1980**



Graphical Integrity

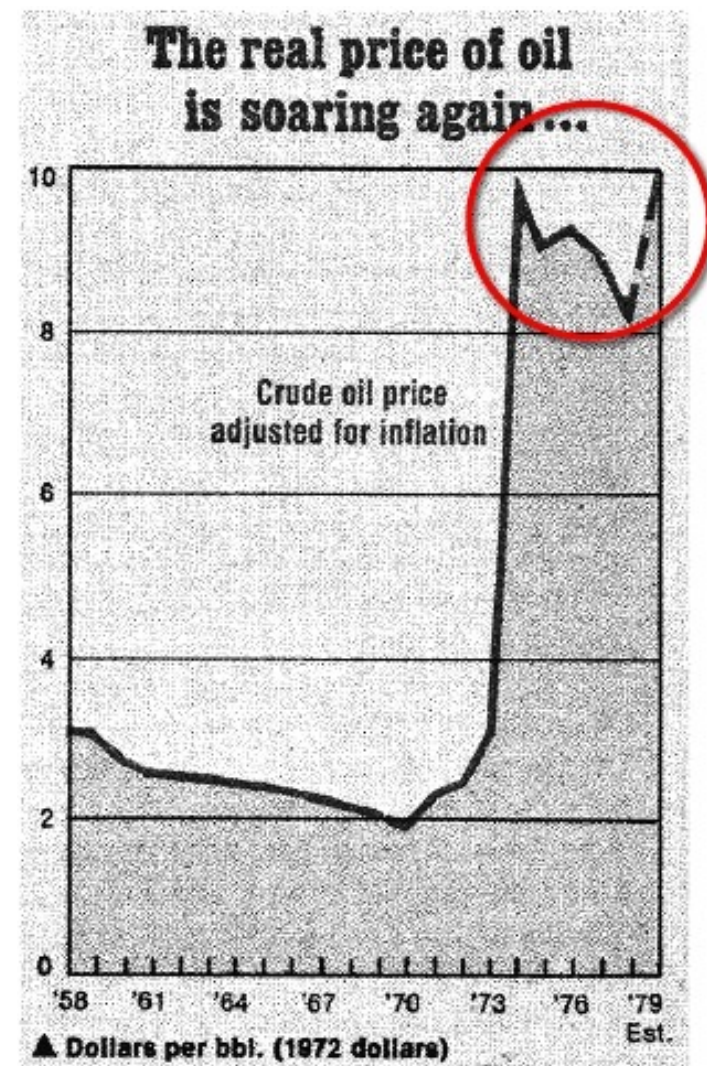
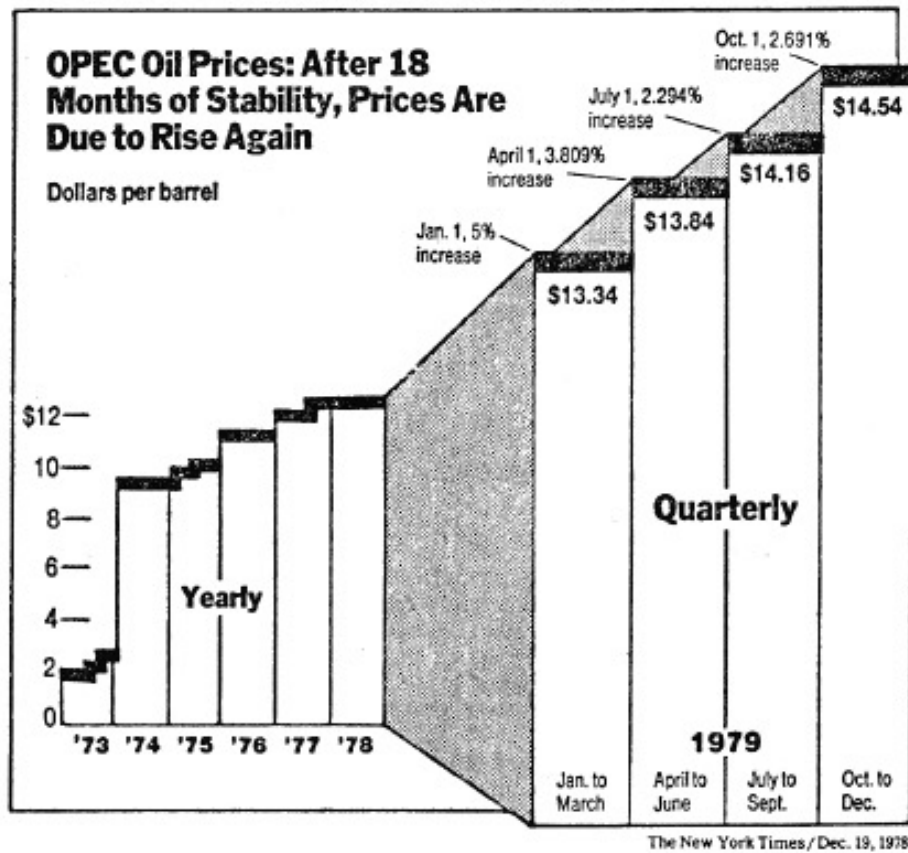
- Missing scales and labels
- Missing context
- Distorting scales
- Distorting design
 - Show *data* variation, not *design* variation

VDQI Example (p61)



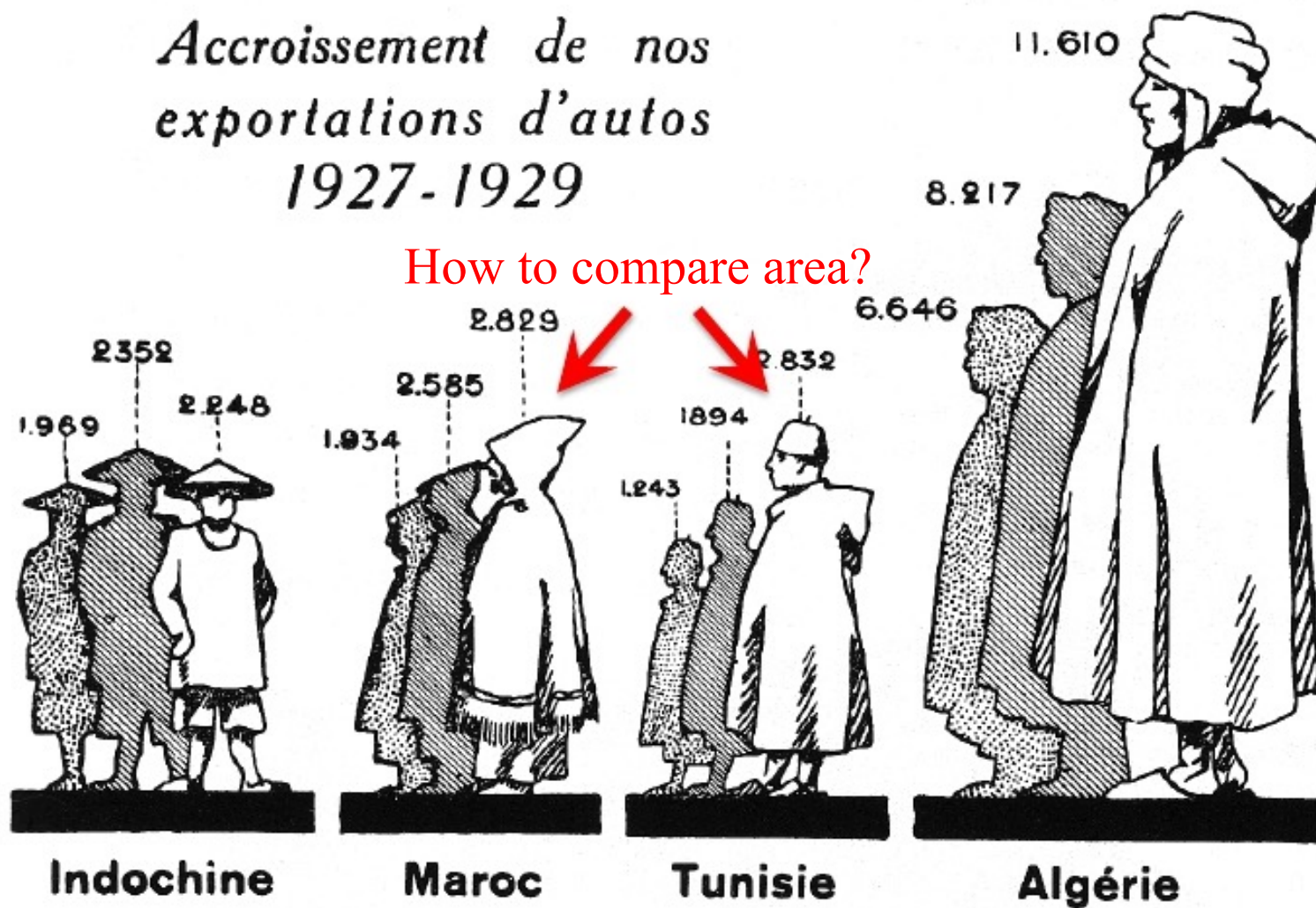
- Vertical Scales
 - 1973 to 1978: \$8.00 per inch
 - 1979: \$4.00 per inch
- Horizontal Scales
 - 1973 to 1978: 3.8 years per inch
 - 1979: 0.57 years per in

VDQI Example (p63)



VDQI Example (p69)

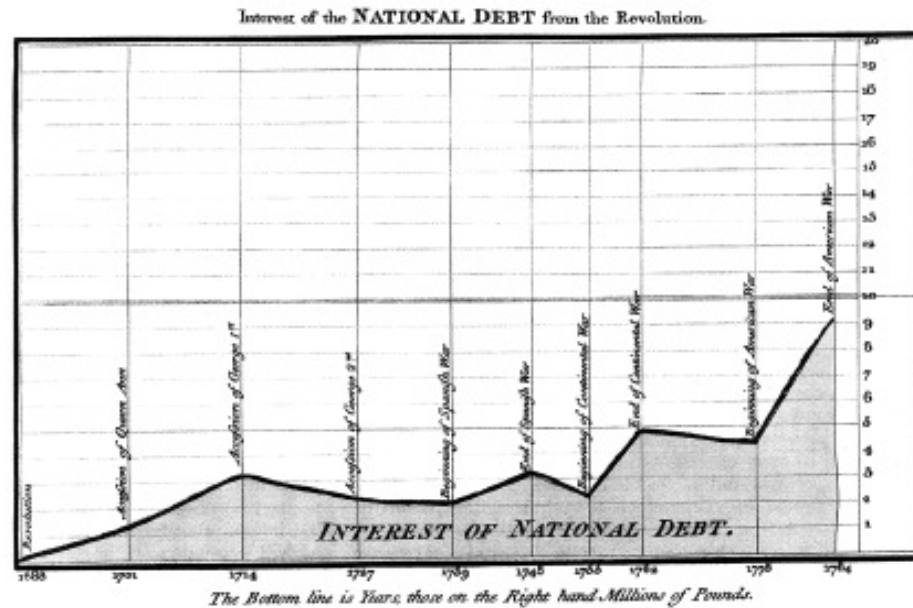
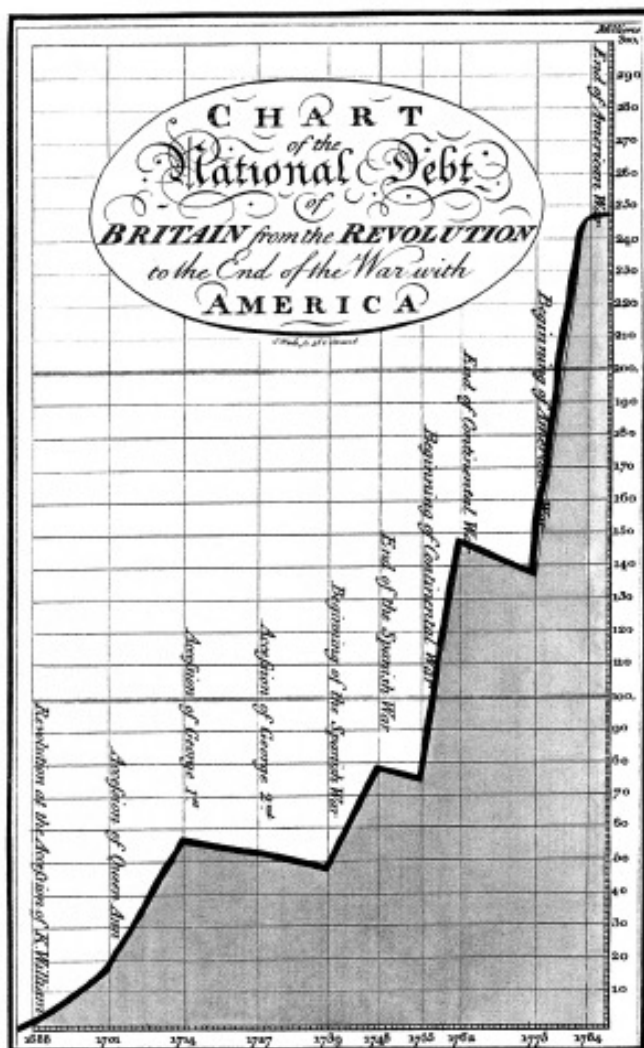
*Accroissement de nos
exportations d'autos
1927-1929*



GRAPHICAL INTEGRITY

Government Spending Example

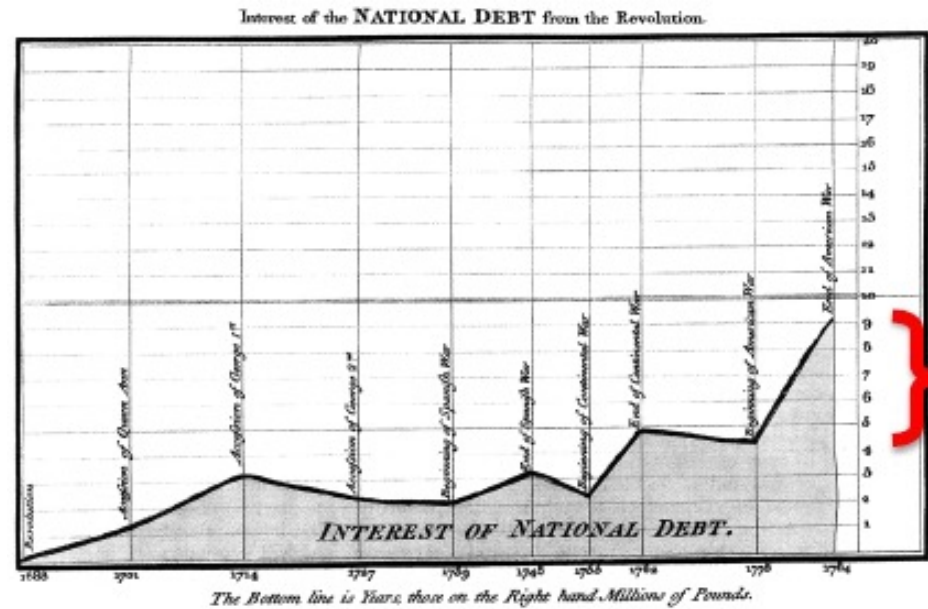
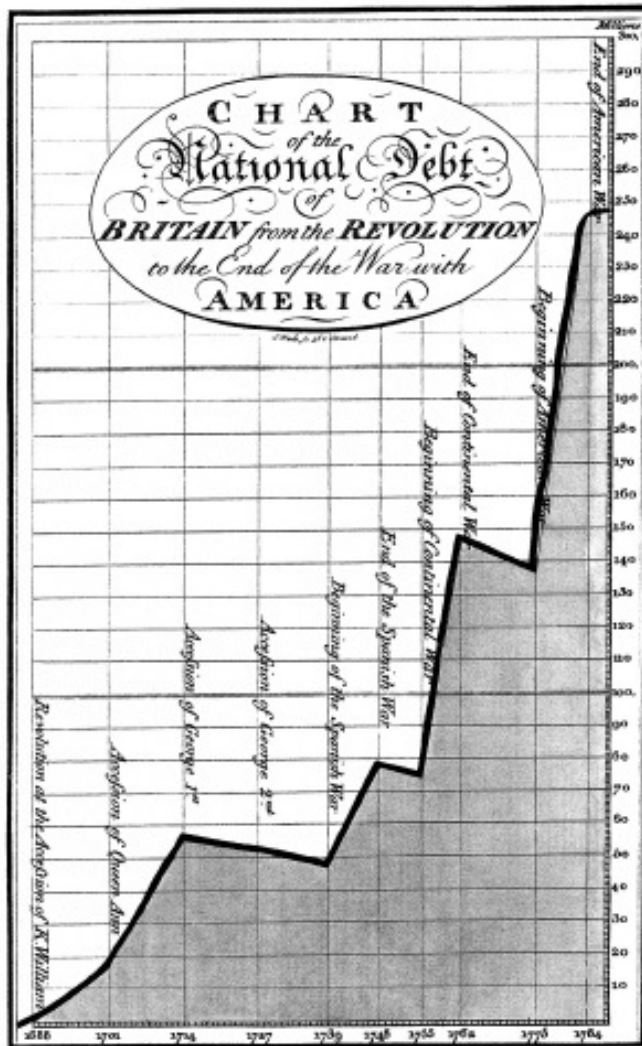
VDQI Example (p64)



Graphical Integrity

- Tall versus wide graphic emphasizes growth

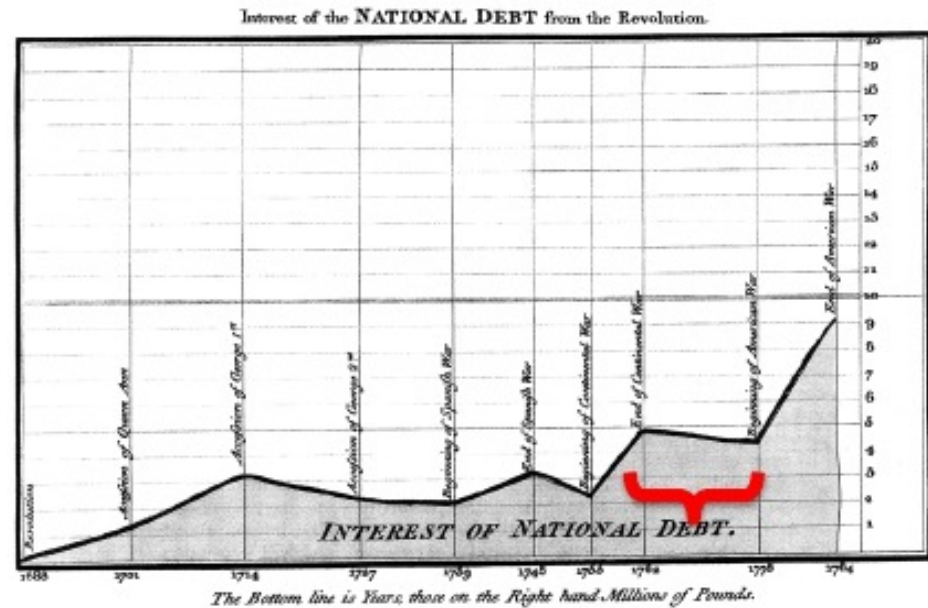
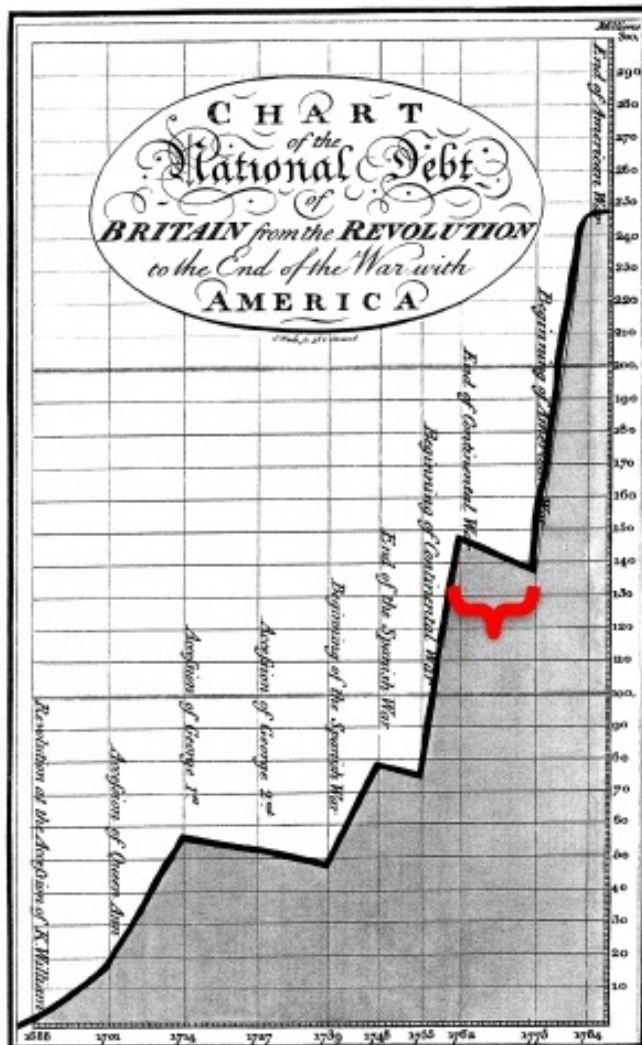
VDQI Example (p64)



Graphical Integrity

- Tall versus wide graphic emphasizes growth
- Wide versus tall graphic emphasizes time

VDQI Example (p64)



Graphical Integrity

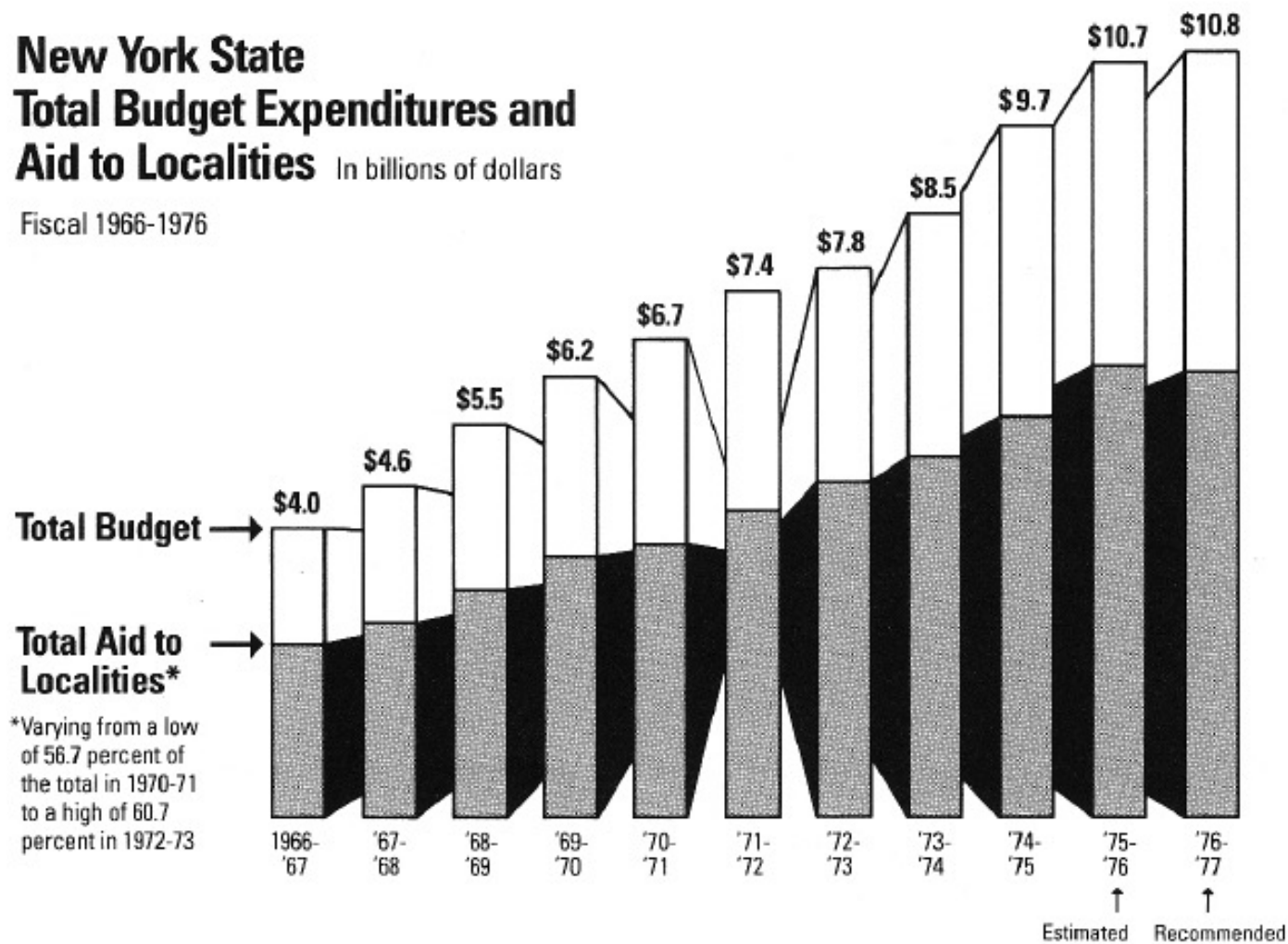
- Tall versus wide graphic emphasizes growth
- Wide versus tall graphic emphasizes time
- Gimmicks emphasize different aspects

VDQI Example (p66)

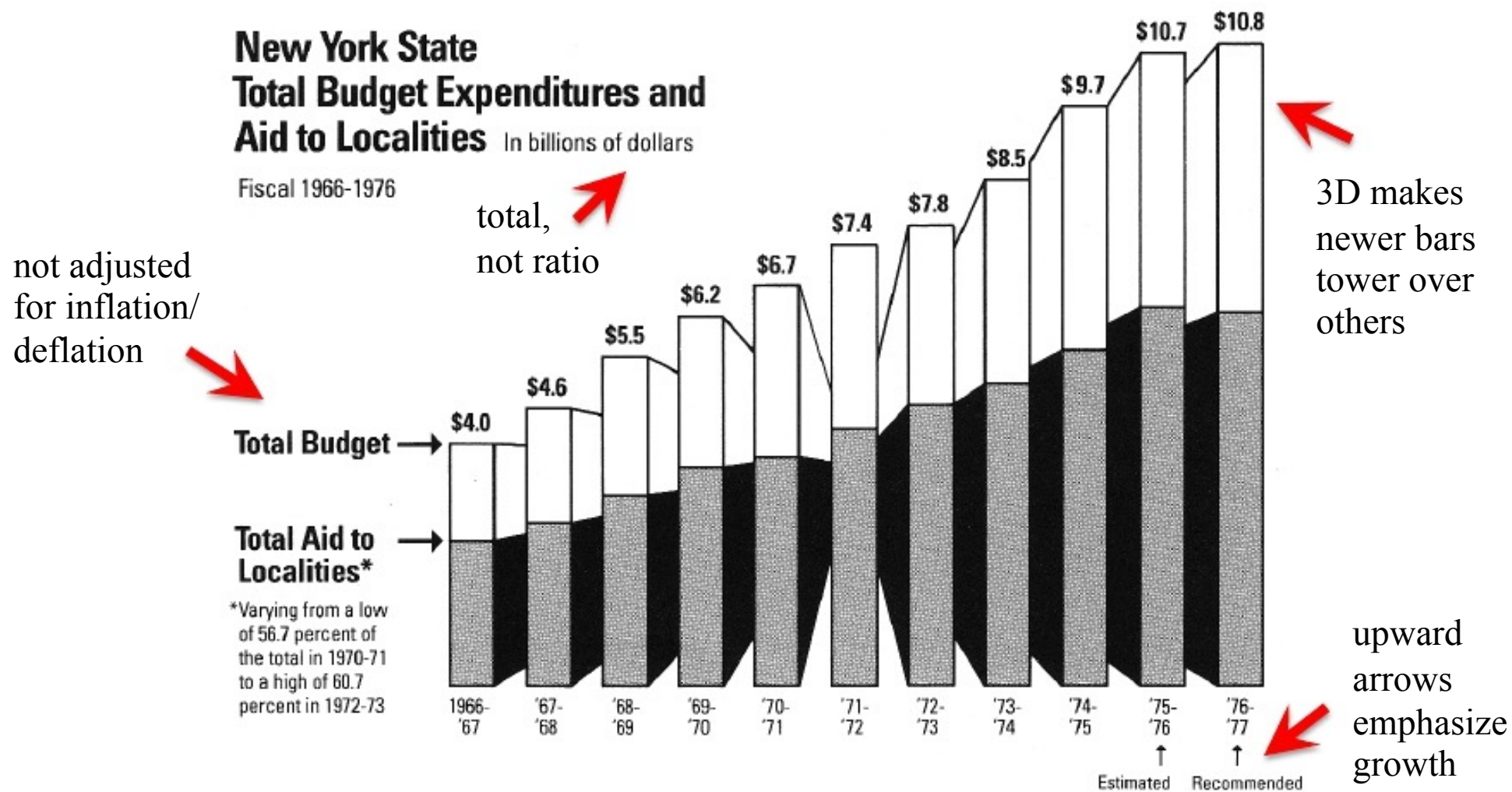
New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

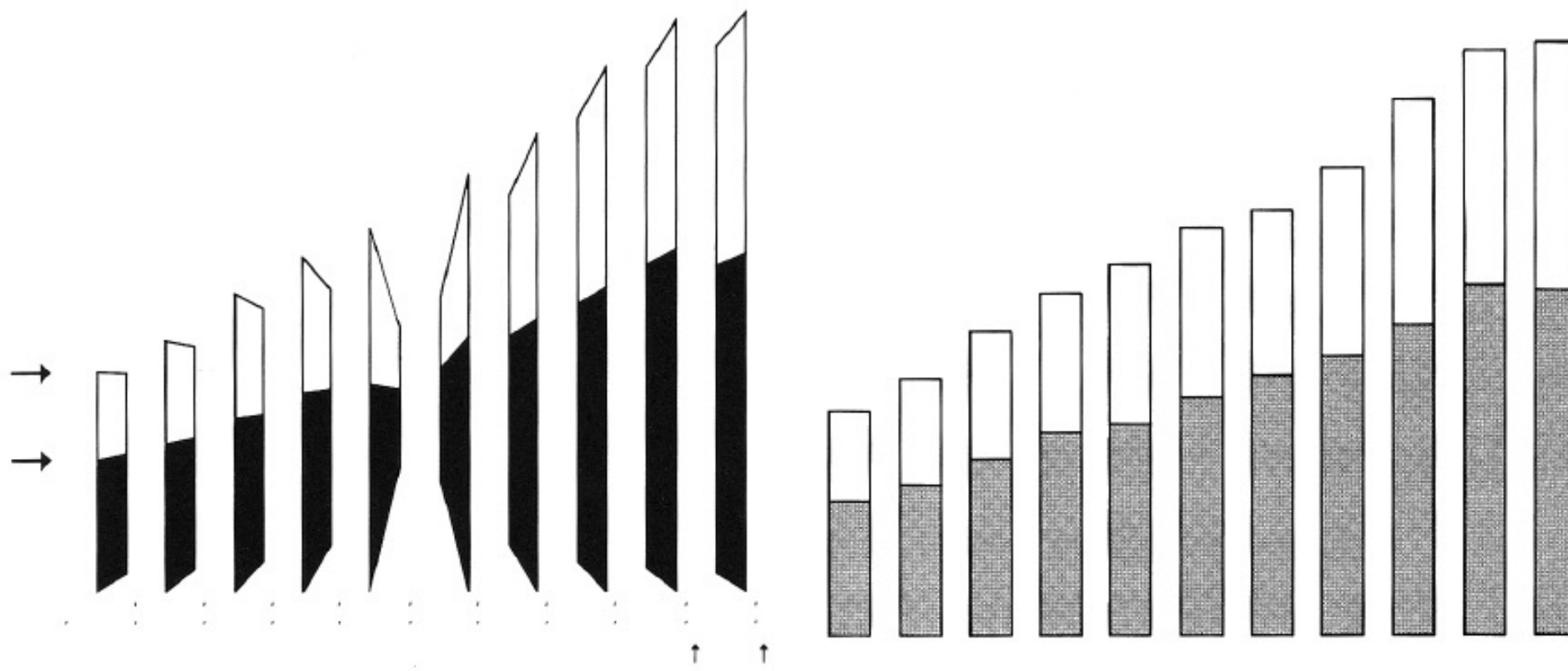
Fiscal 1966-1976



VDQI Example (p66)



VDQI Example (p66)

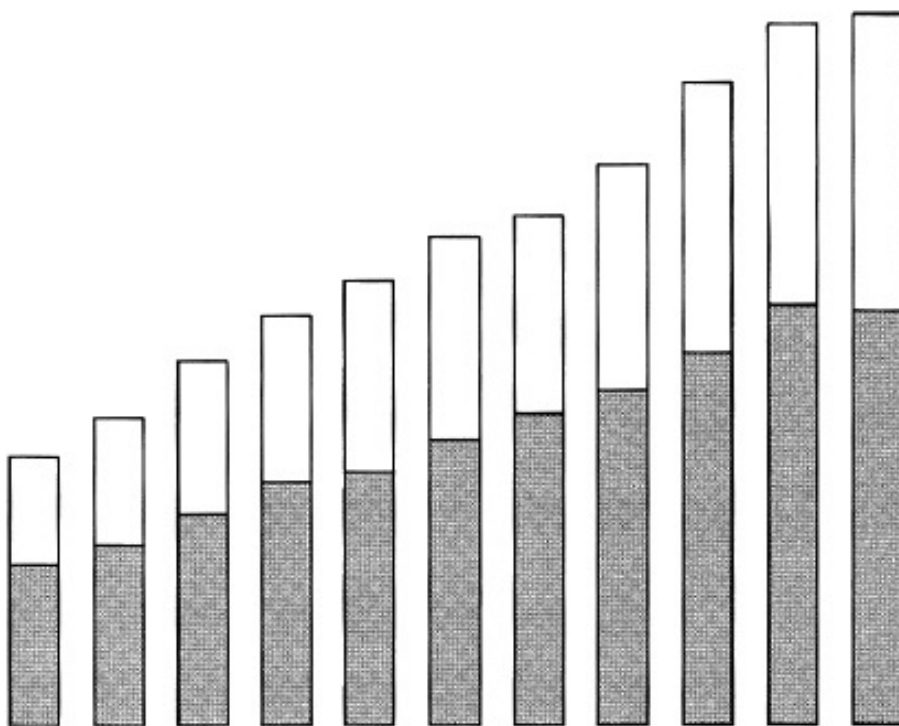


remove design distortion

Graphical Integrity

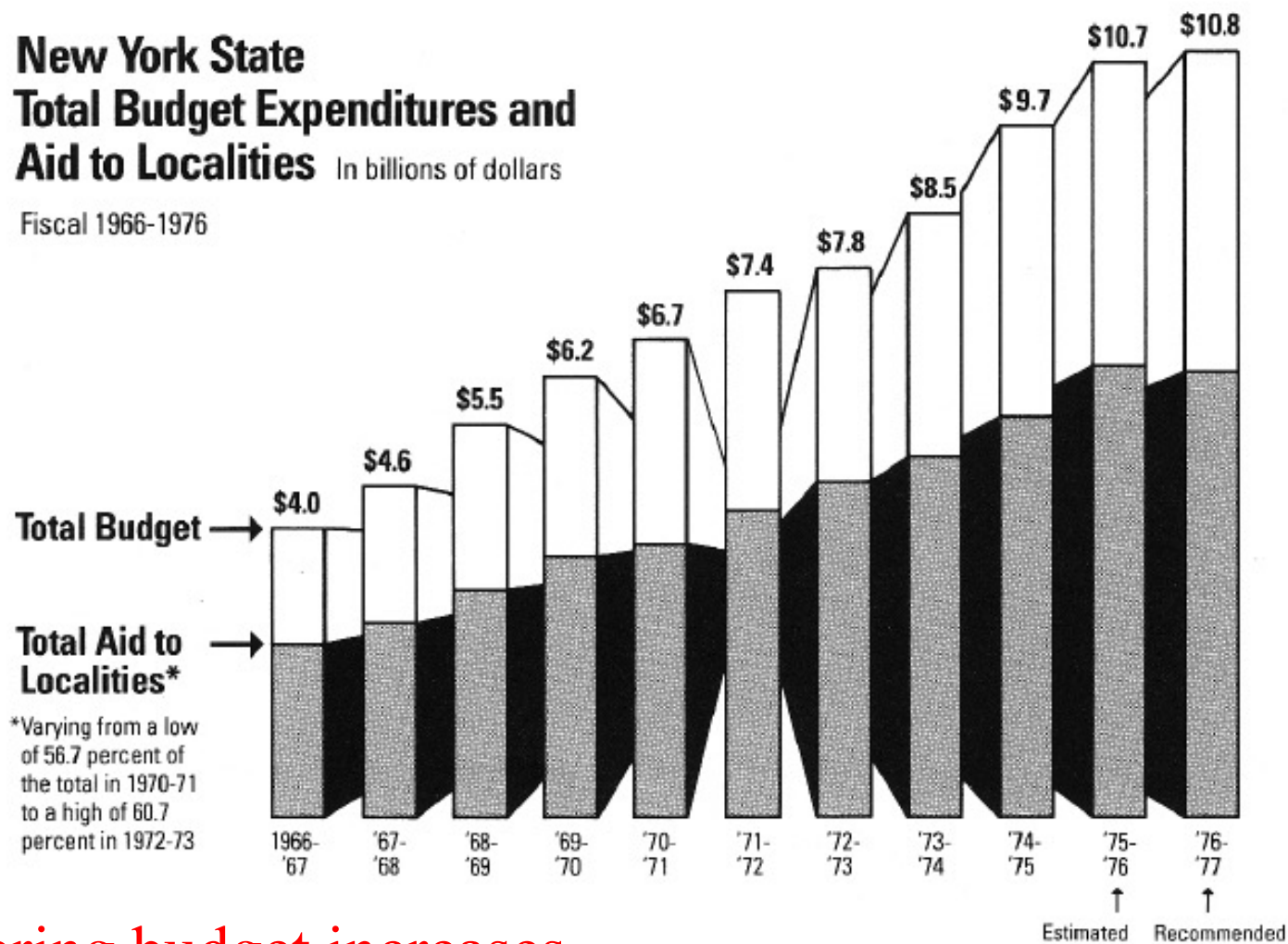
- Tall versus wide graphic emphasizes growth
- Wide versus tall graphic emphasizes time
- Gimmicks emphasizes different aspects
- Counts versus ratios don't provide context

VDQI Example (p66)



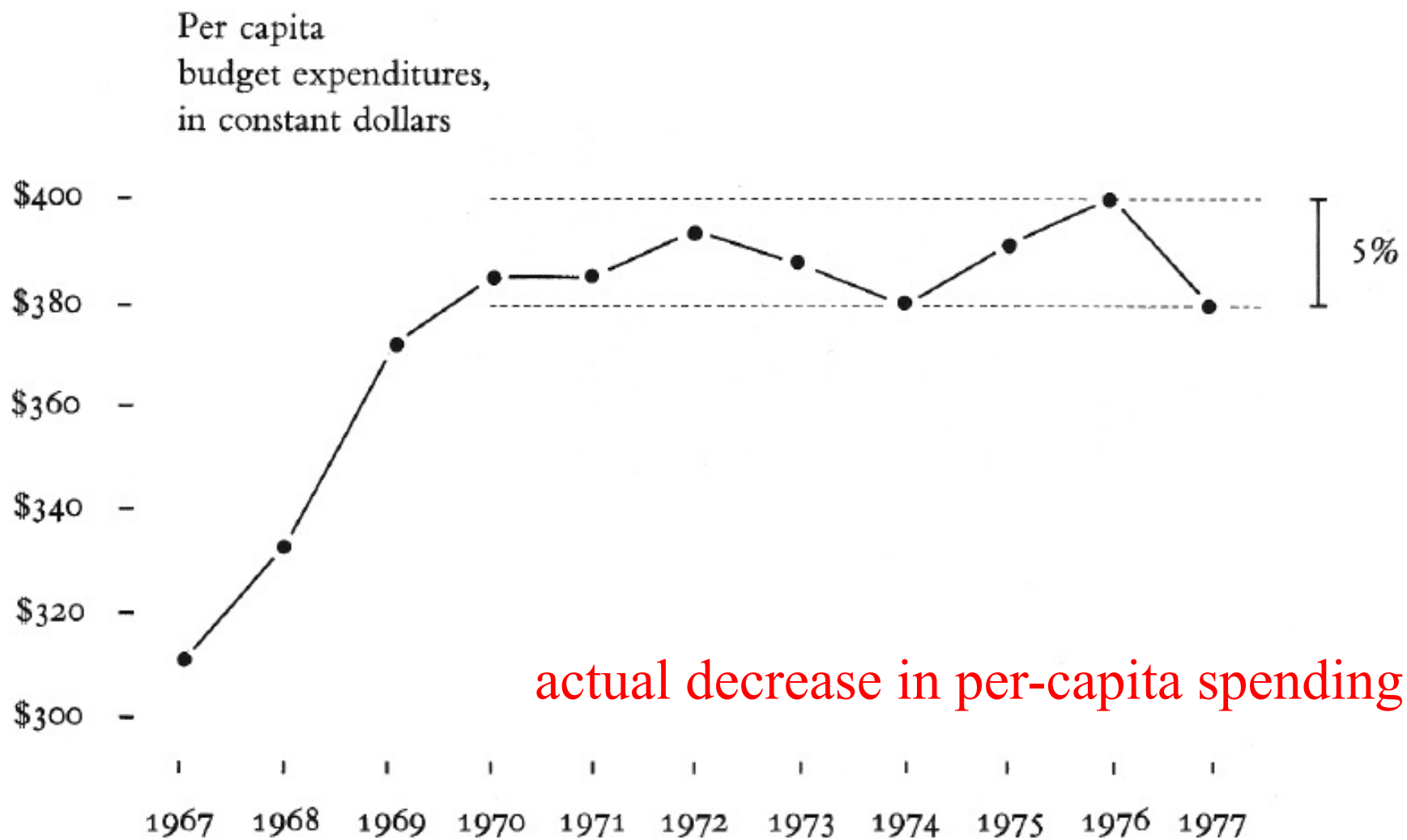
adjust for population growth and inflation

VDQI Example (p66)




towering budget increases

VDQI Example (p66)



Lessons Learned

- Beware emphasis caused by tall versus wide graphs
- Beware gimmicks over emphasizing data
- Beware counts versus ratios
- Beware un-adjusted numbers (e.g. dollars versus dollars adjusted for inflation)



*Thanks to
Sophie J. Engle
San Francisco University*

for ideas, suggestions, slides, links, and much other stuff