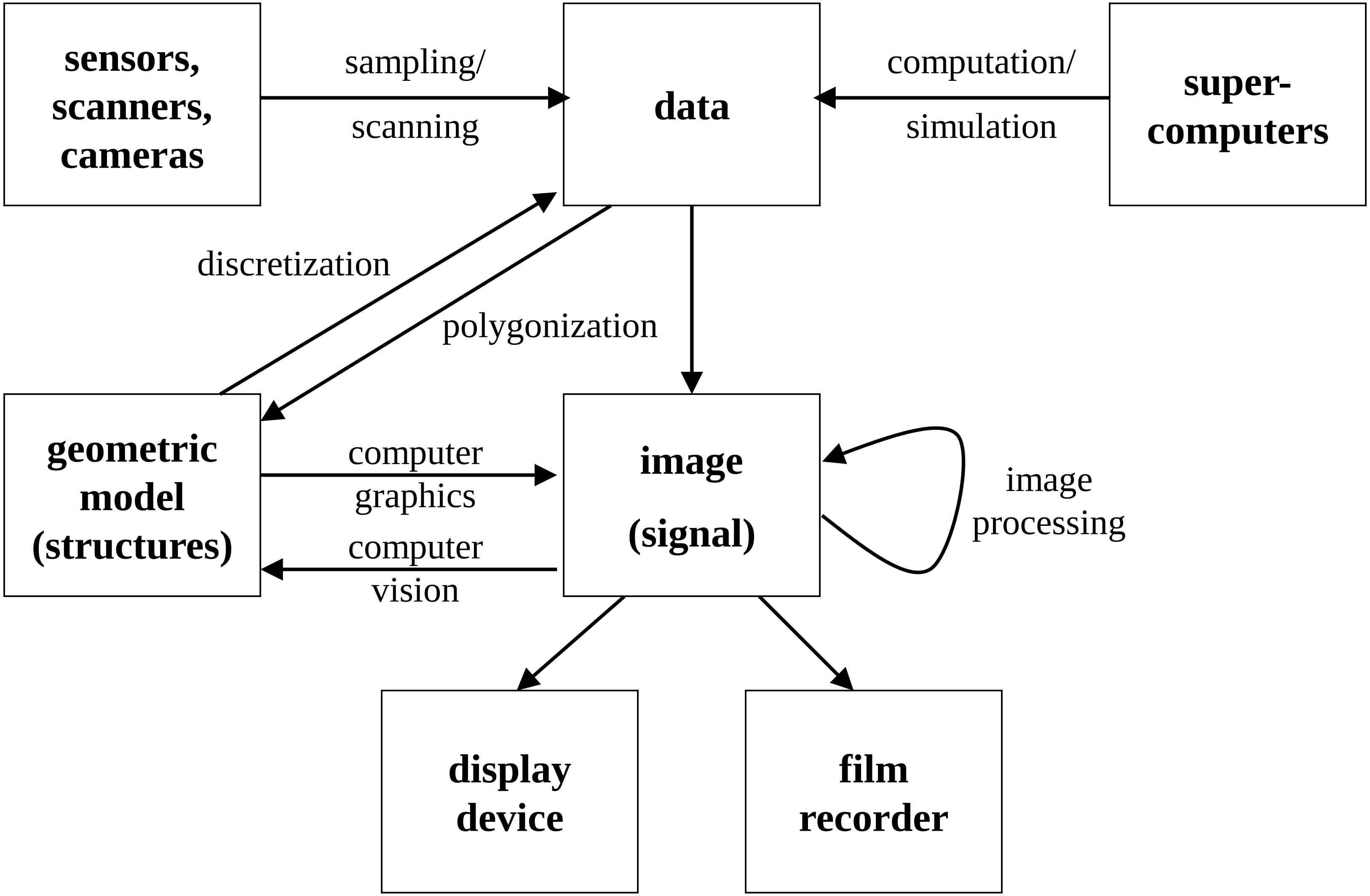
Visualization Taxonomy



Review: Visualization Highlights

- Visualization consists of methods for transforming the symbolic or numeric into the visual
- Many applications spanning science, engineering, medicine, analysis, simulation, etc.
- Draws on many fields, such as computer graphics, imaging, signal processing, computer vision, etc.
- Potential danger: deliberate misuse of visualization to deceive

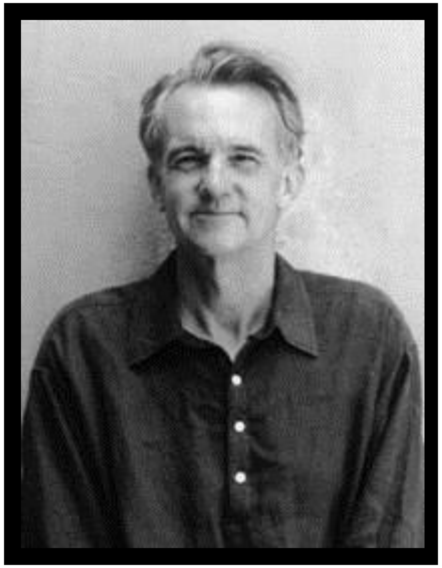
Traditional Visualization: Historical Perspectives

Traditional Visualization

- Where does Visualization come from?
- What are the origins of visualization?
- What are some of the troubles inherent in trying to visualize data?
- What makes a visual representation of some data faithful, helpful, accurate, etc.?
- "Those ignorant of history are doomed to repeat it."
- Well, maybe not doomed, but it certainly helps understand where we're coming from in this field
- Surprisingly, we can learn a lot about 3D computer-driven visualization by looking at early attempts at effective 2D hand-drawn visualization

Graphical Display (i.e., 2D)

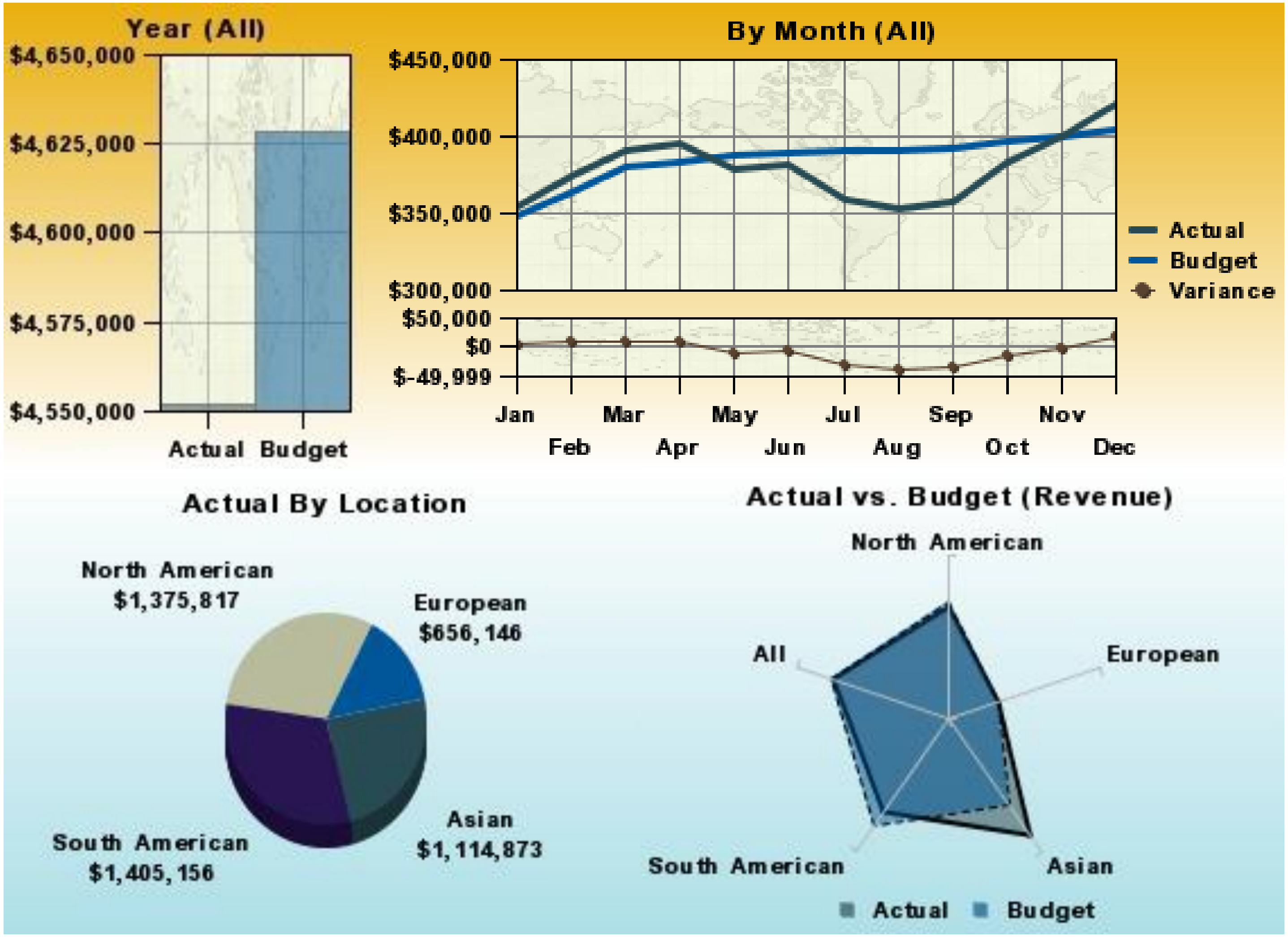
- Some material from Edward R. Tufte's "The Visual Display of Quantitative Information"
- Who? Tufte? Author of some of the best-known texts on the origins, application, and mis-application of visualization
- Professor at Yale (retired)
- Fall 2003 *Wired* article written by him called "Powerpoint Is Evil" (uh oh)
- Maintains that the Challenger shuttle disaster would not have occurred if data had been presented in tabular format correctly
- In summary, he's a big-shot in the visualization field and has some interesting things to say on the subject ("information graphics")



STONY BROOK
STATE UNIVERSITY OF NEW YORK

Earlier Days of Computer Graphics

- Visual display of data (graphs and charts)



Graphical Display

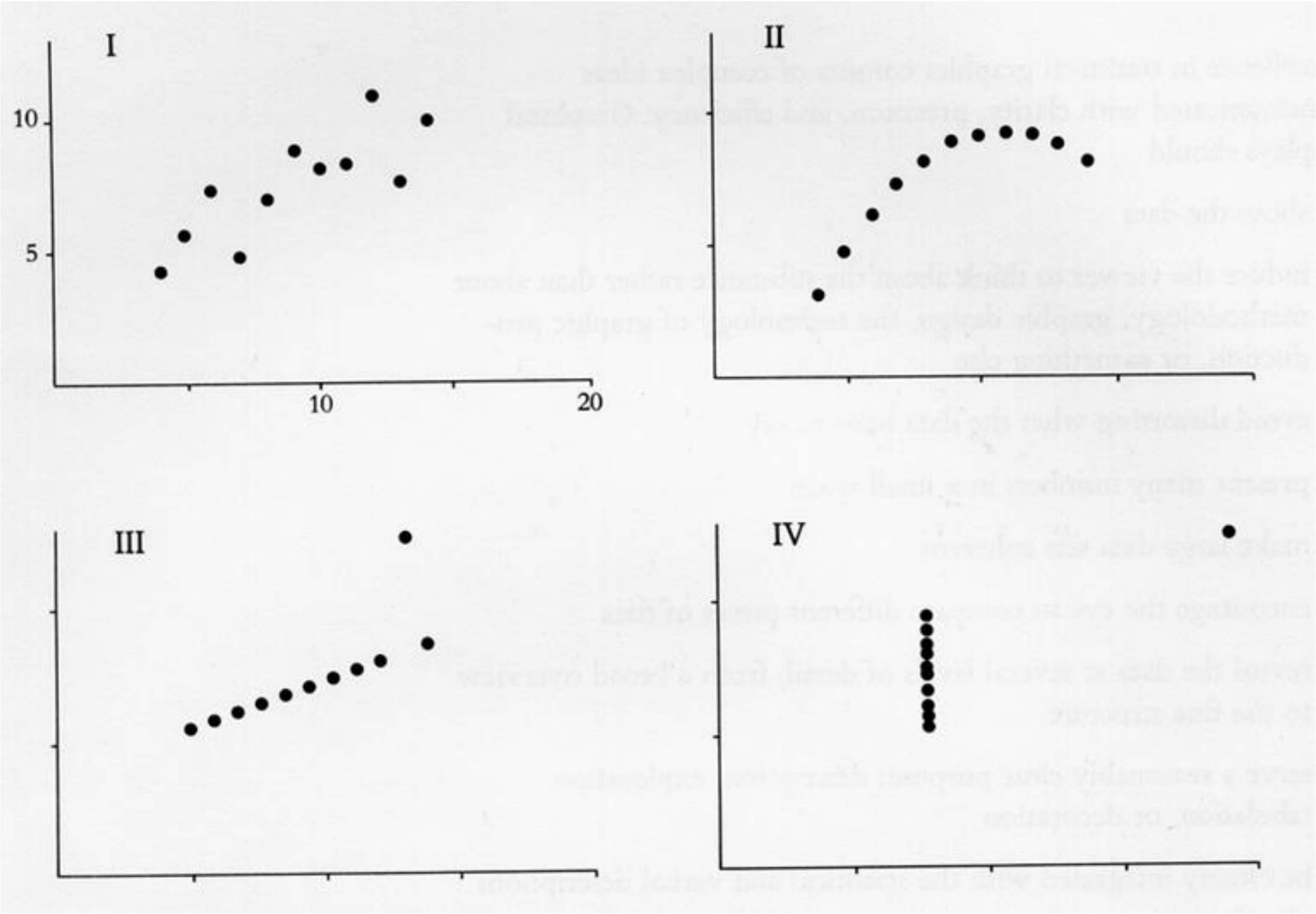
- Fundamental question: why bother with visualization? What do we gain? Why aren't words and numbers enough?
- Graphics (i.e., pictures) can be more precise and revealing than numerical display

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 $X - \overline{X} = 110.0$
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
$r^2 = .67$

STONY BROOK
STATE UNIVERSITY OF NEW YORK

Surprise in the Data!



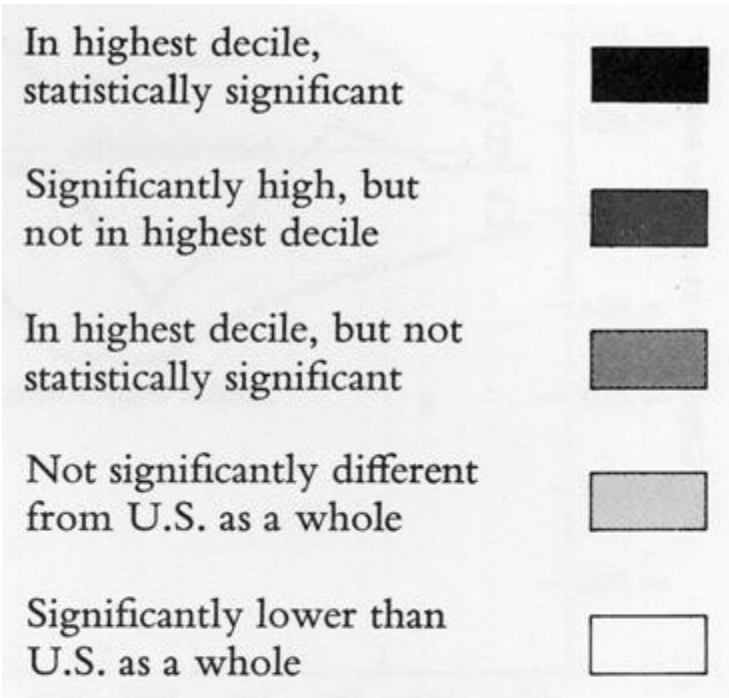
Dr. John Snow's Cholera Map of London (1854)

- **Dot indicates cholera death**
- **X indicates water pump (circled)**
- **What does this visualization tell us?**

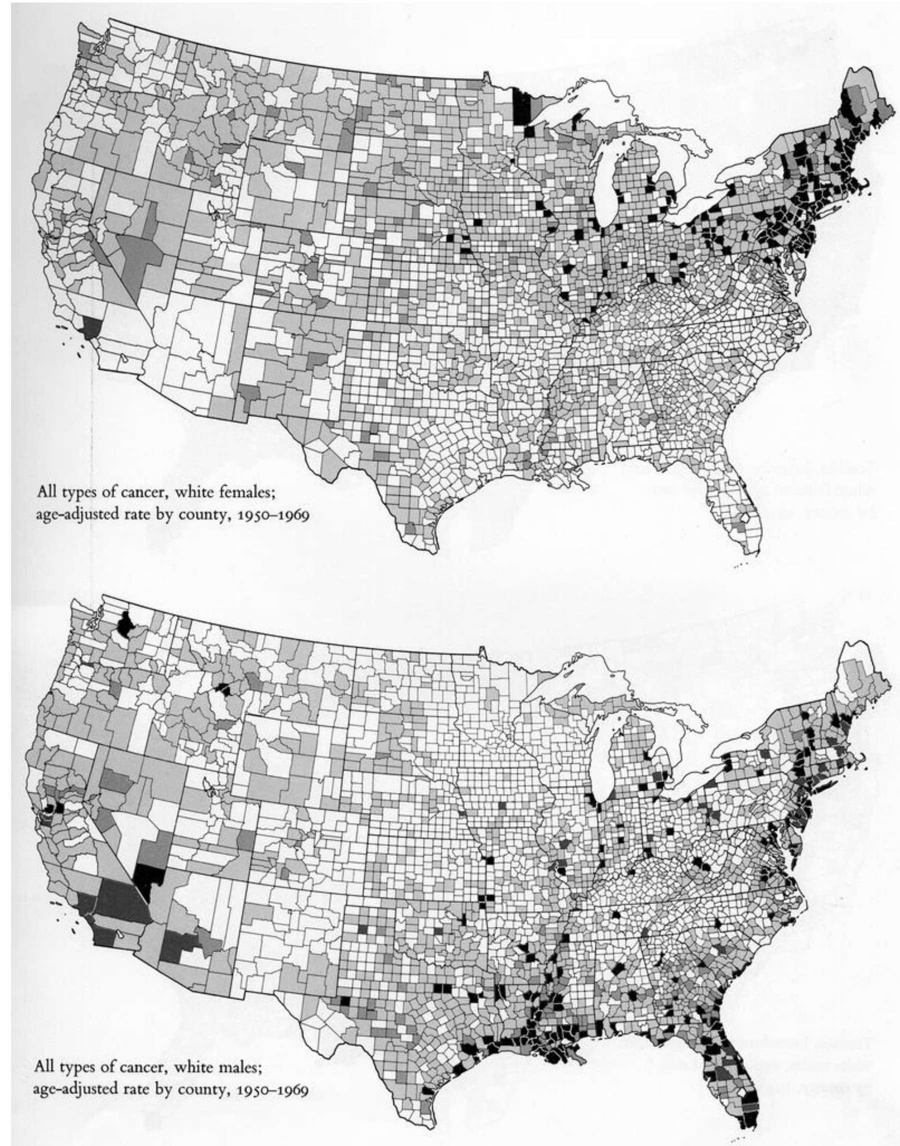
STONY BROOK
STATE UNIVERSITY OF NEW YORK

Graphical Display – Large Datasets

- Total cancer deaths, 1950-1969 (top: white women; bottom: white men)



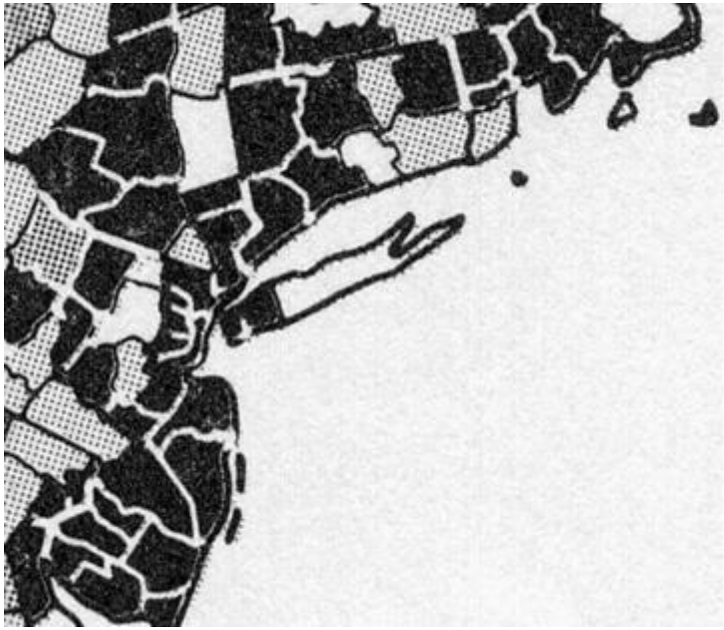
- Can capture a large amount of information in a very small space (billions of bits on one page)



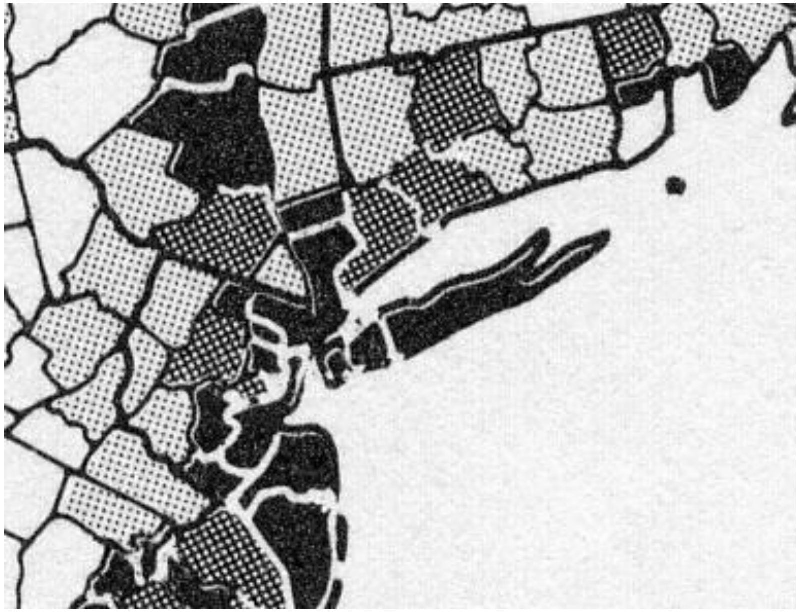
STONY BROOK
STATE UNIVERSITY OF NEW YORK

Graphical Display – Large Datasets

- The local situation



women

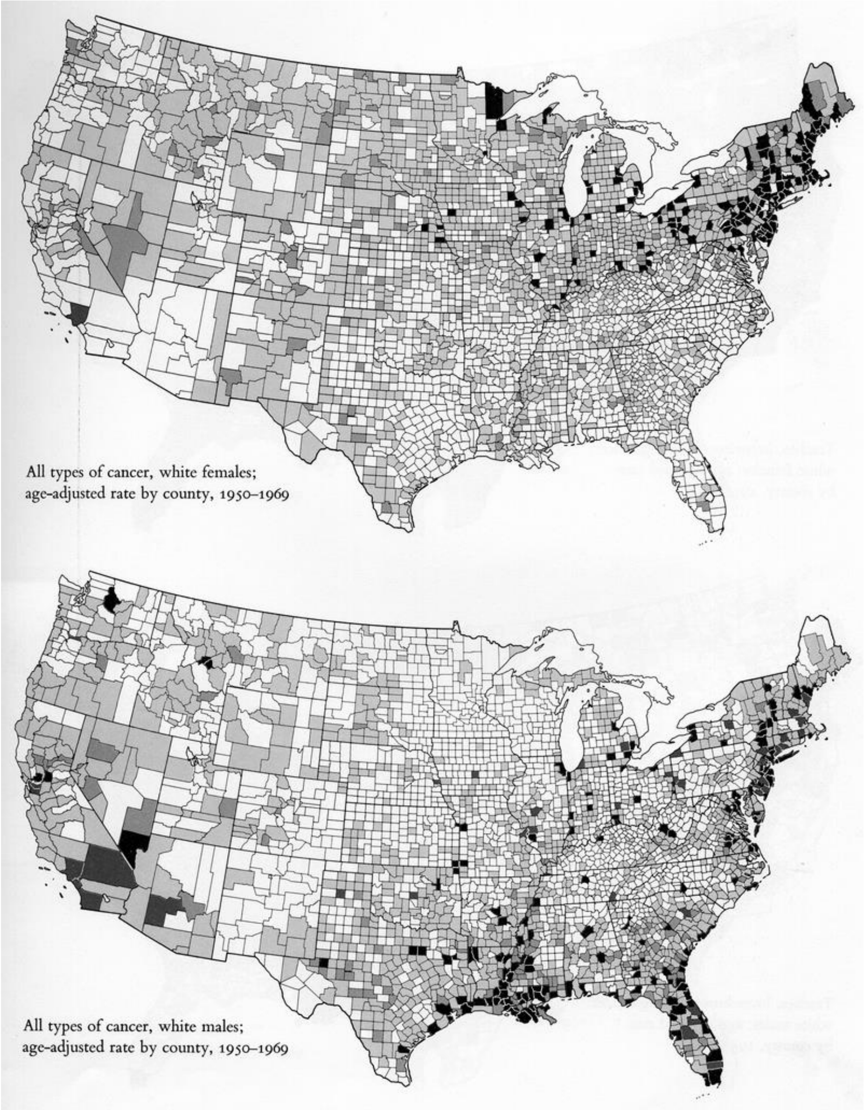


men

Let's hope things have improved since then!

Graphical Display – Large Datasets

- Important questions:
  - Where are the highest death rates?
  - Lowest rates?
  - Rate for men vs. women
  - Any anomalies?
  - What to do with the knowledge?
- Do you see any possible problems with visualizing the data in this manner? (hint: consider land area)
- Focus is incorrectly drawn to land area rather than number of people actually living in county
- A large county may have only a few people living in it
- Are there any other considerations besides the *display* of the data?

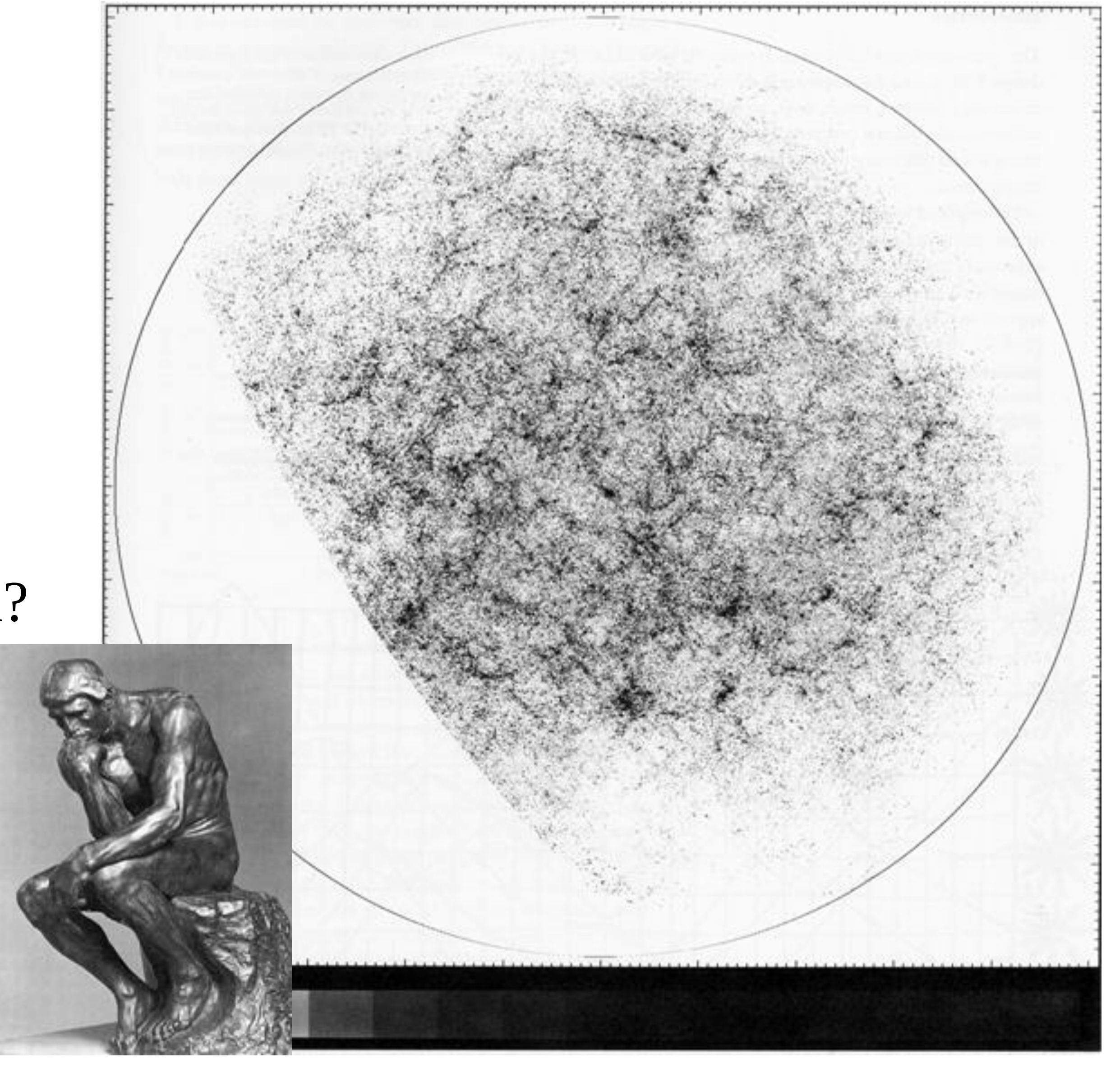


Traditional Visualization ("Information Graphics")

- The fundamental goal of visualization is to reveal the substance of the data – i.e., what we can learn from the raw data and what we should do with our knowledge
- Our concern is with the data and not so much the techniques, algorithms or methodologies used to draw the image
- We also have to make reasonable assumptions that the data itself is not corrupt or skewed in some manner
- Traditionally, image generation (IG) has been the domain of artists who are skilled at drawing 2D images
- Today we might say IG is the domain of (3D) computer graphics experts and those involved in computer-driven IG
- Visualization has moved from using hand-drawn illustration to Computer-Generated Imagery (CGI)

Graphical Display – Large Datasets

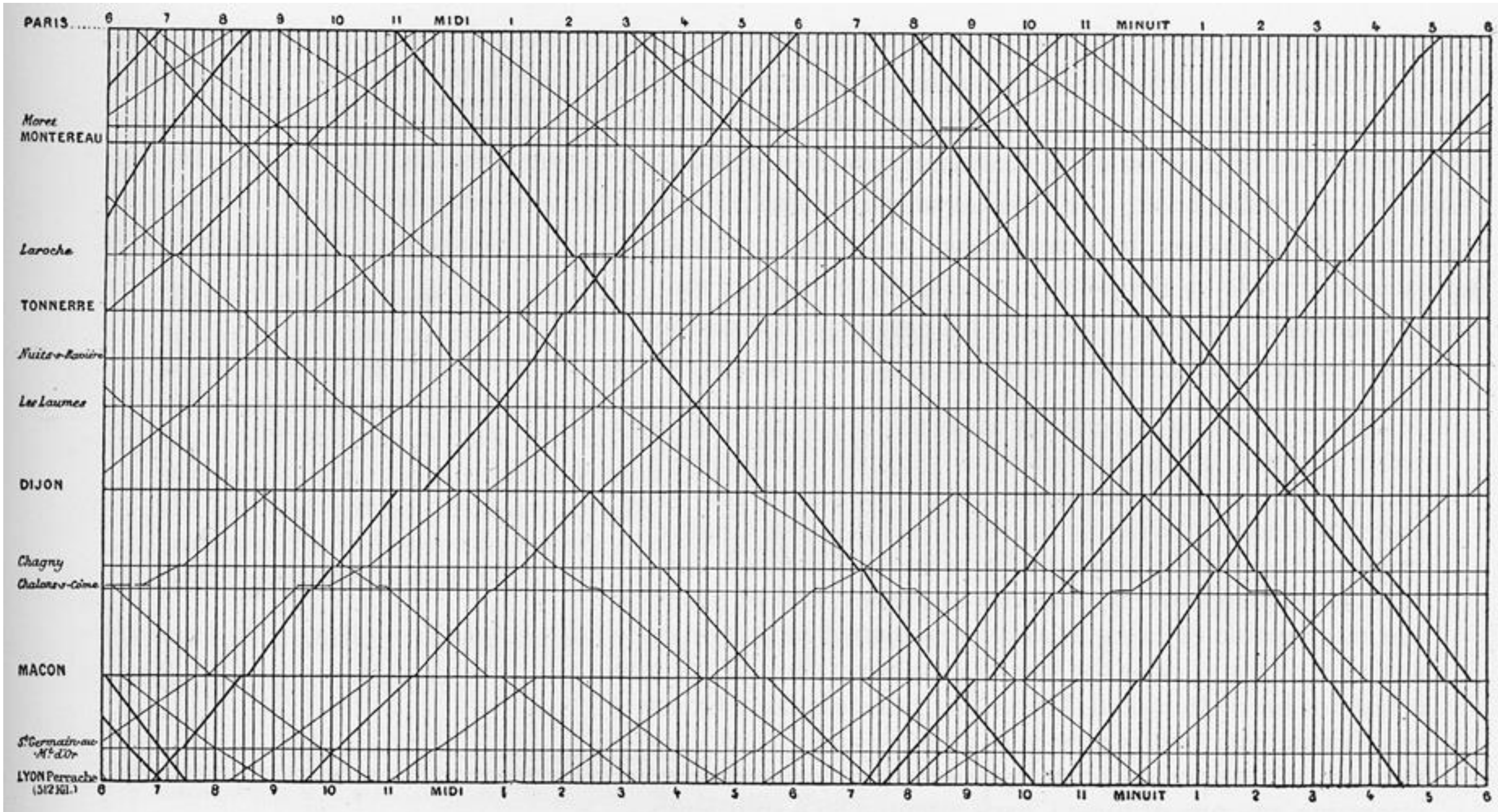
- Galaxy map
- Each dot represents a collection of galaxies
- 1.3 million total
- 1024x2022 grid
- Can you see any structure in the data?
- Clusters, filaments
- What might these observations tell us?



STONY BROOK
STATE UNIVERSITY OF NEW YORK

Time-Series Display

- Paris-Lyon train schedule from 1880s



STONY BROOK
STATE UNIVERSITY OF NEW YORK

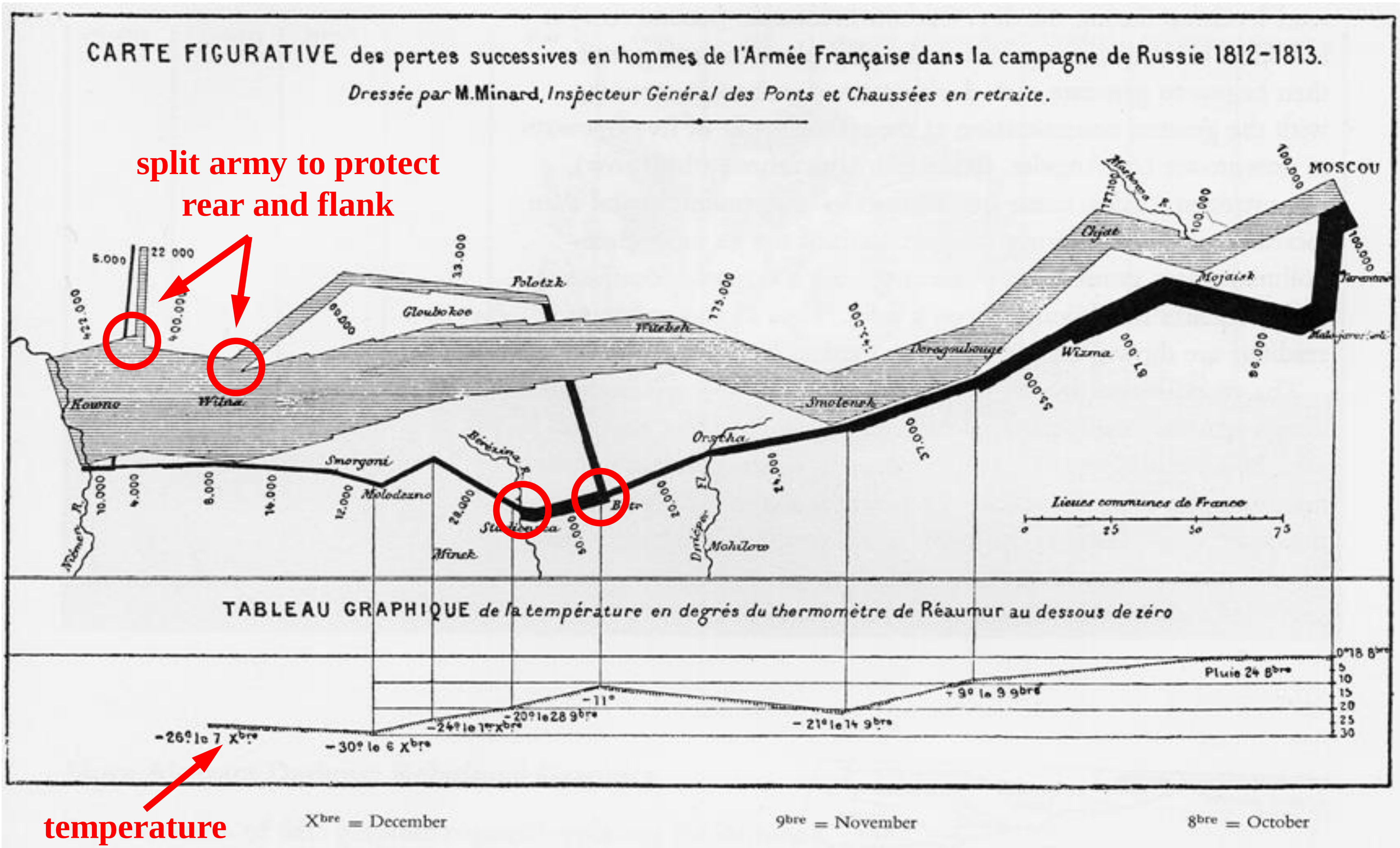
Compare with this way…

- The modern method is more precise, but is it better?
- What are the advantages and disadvantages of each method?
- What does each visualization tell us that the other doesn't or can't?
- French map: can see how trains overlap – perhaps easier to make connections?
- LI table: more precise, contains annotations
- Consider issues like these in developing your own visualization algorithms and systems

To Long Island Monday to Friday Except Holidays				
	Leave			Arrive
Notes	Penn Station	Flatbush Avenue	Jamaica	Stony Brook
	J1:44 am	J1:44 am	2:11 am	3:29 am
	5:47 am	J5:51 am	6:12 am	AT7:46 am
	J7:49 am	J7:50 am	8:15 am	A9:28 am
	9:15 am	J9:09 am	9:37 am	E11:00 am
	10:41 am	J10:37 am	11:04 am	T12:30 pm
	12:15 pm	J12:10 pm	12:37 pm	E2:00 pm
	1:42 pm	J1:37 pm	2:04 pm	T3:30 pm
	2:52 pm	J2:33 pm	3:13 pm	T4:42 pm
	J3:34 pm	J3:32 pm	4:00 pm	5:15 pm
Peak	4:19 pm	J4:20 pm	4:40 pm	5:54 pm
Peak	4:49 pm	J4:47 pm	5:12 pm	6:24 pm
Peak	J5:04 pm	J5:06 pm	5:30 pm	6:49 pm
Peak	J5:38 pm	J5:45 pm	6:07 pm	7:19 pm
Peak	6:01 pm	......	......	E7:49 pm
Peak	......	J6:08 pm	6:28 pm	7:49 pm
Peak	J6:27 pm	J6:27 pm	6:49 pm	8:04 pm
Peak	7:22 pm	J7:23 pm	7:44 pm	T9:07 pm
	8:42 pm	J8:42 pm	9:03 pm	T10:30 pm
	9:42 pm	J9:38 pm	10:03 pm	T11:30 pm
Notes	Penn Station	Flatbush Avenue	Jamaica	Stony Brook
	10:42 pm	J10:30 pm	11:03 pm	T12:22 am
	11:42 pm	J11:30 pm	12:03 am	T1:29 am

Visualization in Narrative Form

- Napolean’s march to Moscow – War of 1812 (drawn in 1861)



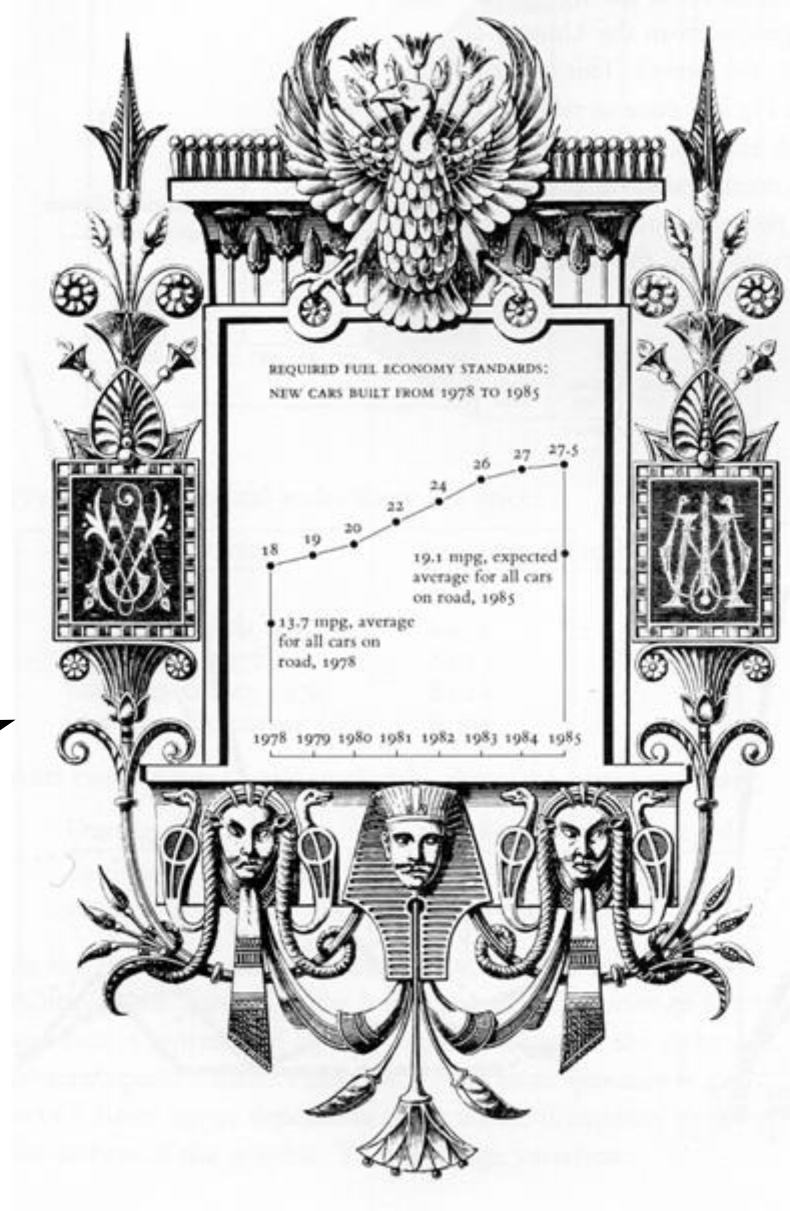
STONY BROOK
STATE UNIVERSITY OF NEW YORK

Tufte's Principles of Graphical Excellence

- Graphical excellence is the well-designed presentation of interesting data. It is a matter of:
  - substance
  - statistics
  - design
- Substance – convey the data and nothing more
- Statistics – secondary about the information itself that aids in understanding the data
- Design – present the data in a way that makes it as easy as possible to understand and extract salient aspects

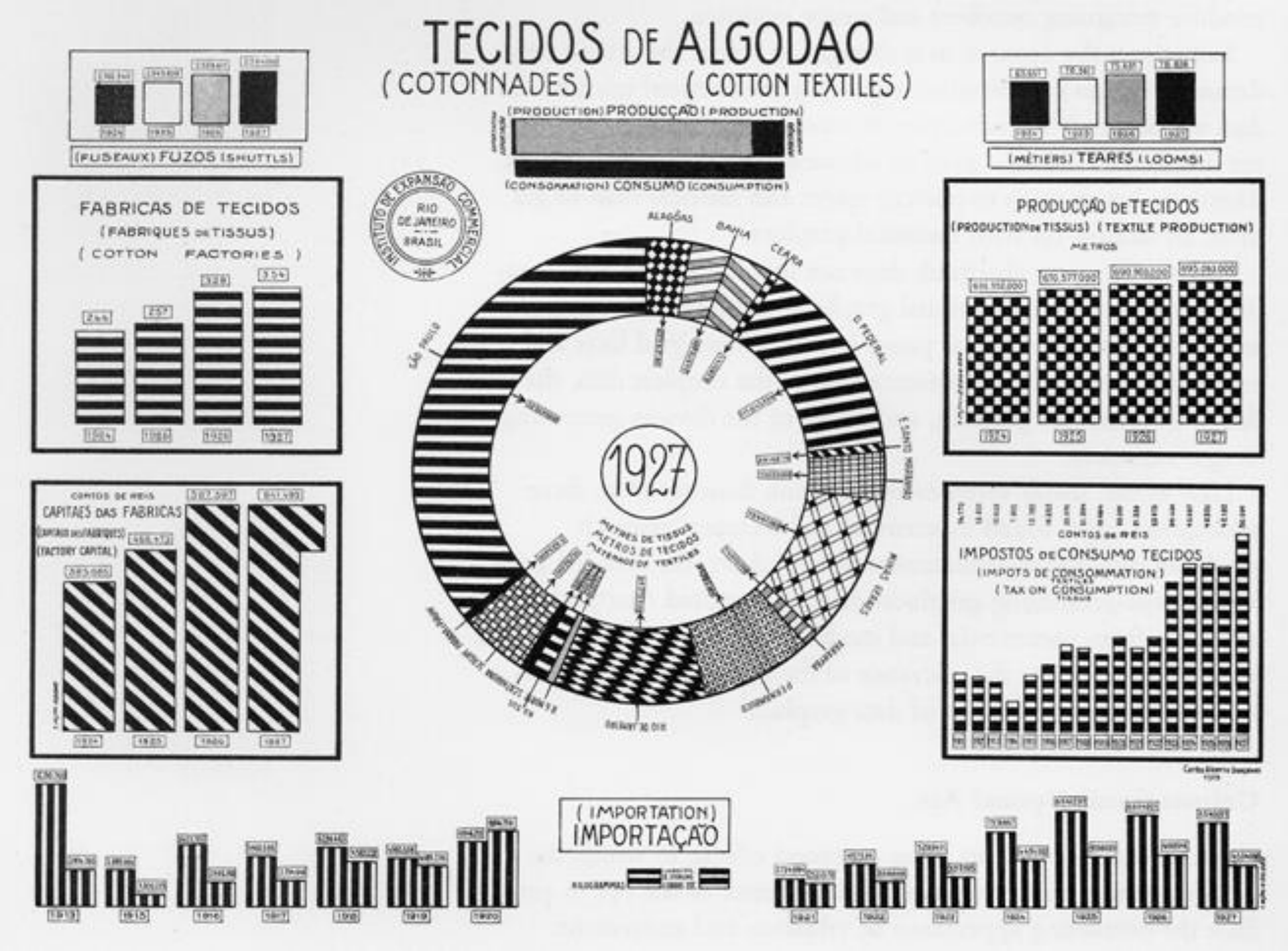
Tufte's Principles of Graphical Excellence

- Graphical excellence is what gives the user
  – the **greatest number of ideas**
  – in the **shortest time**
  – with the **least ink**
  – in the **smallest space**
- In other words, don't burden the viewer with extraneous material or "chartjunk," as Tufte calls it
- Tufte's idea of a joke



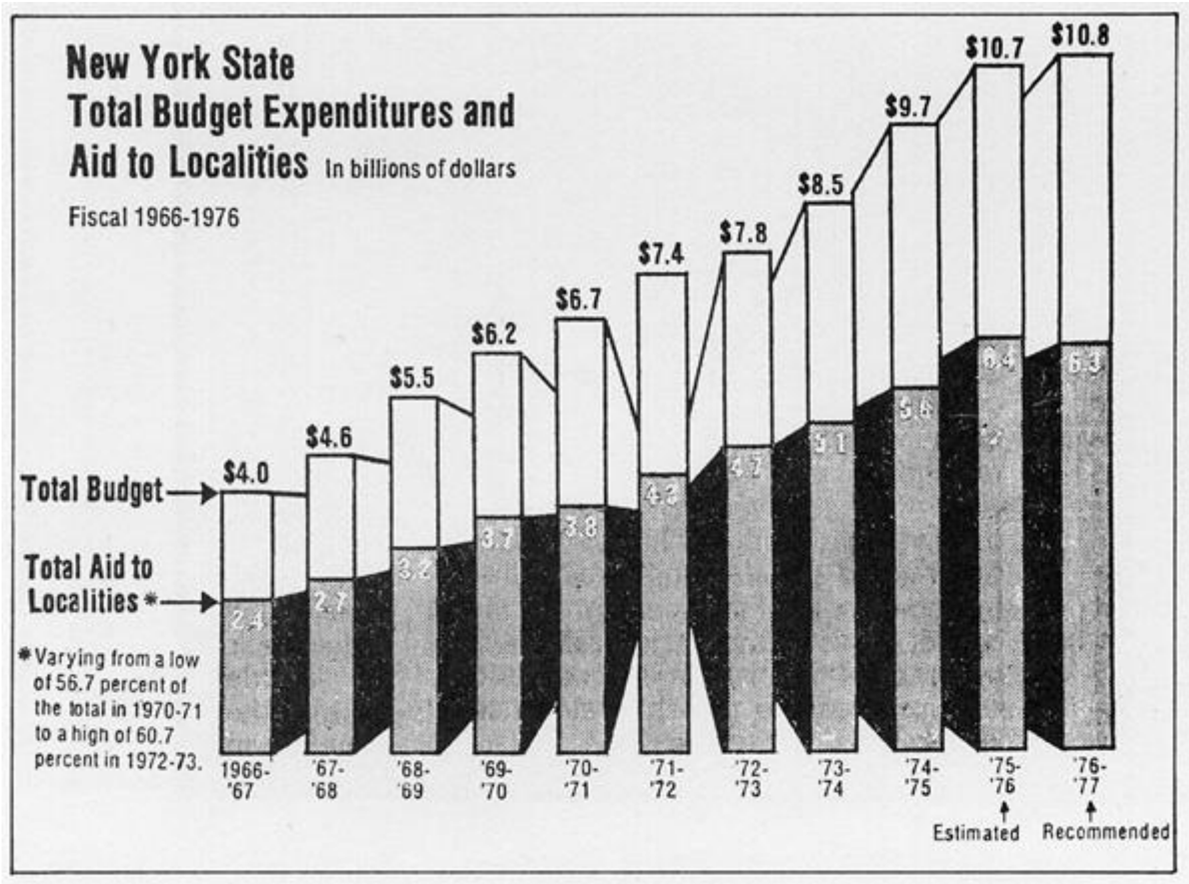
Chartjunk

- What is wrong with this graphic?
- How could it be improved?
- What's the lesson to be learned here?
- Don't use fancy designs for the sake of being fancy
- Doing so distracts you from the task of visualization!



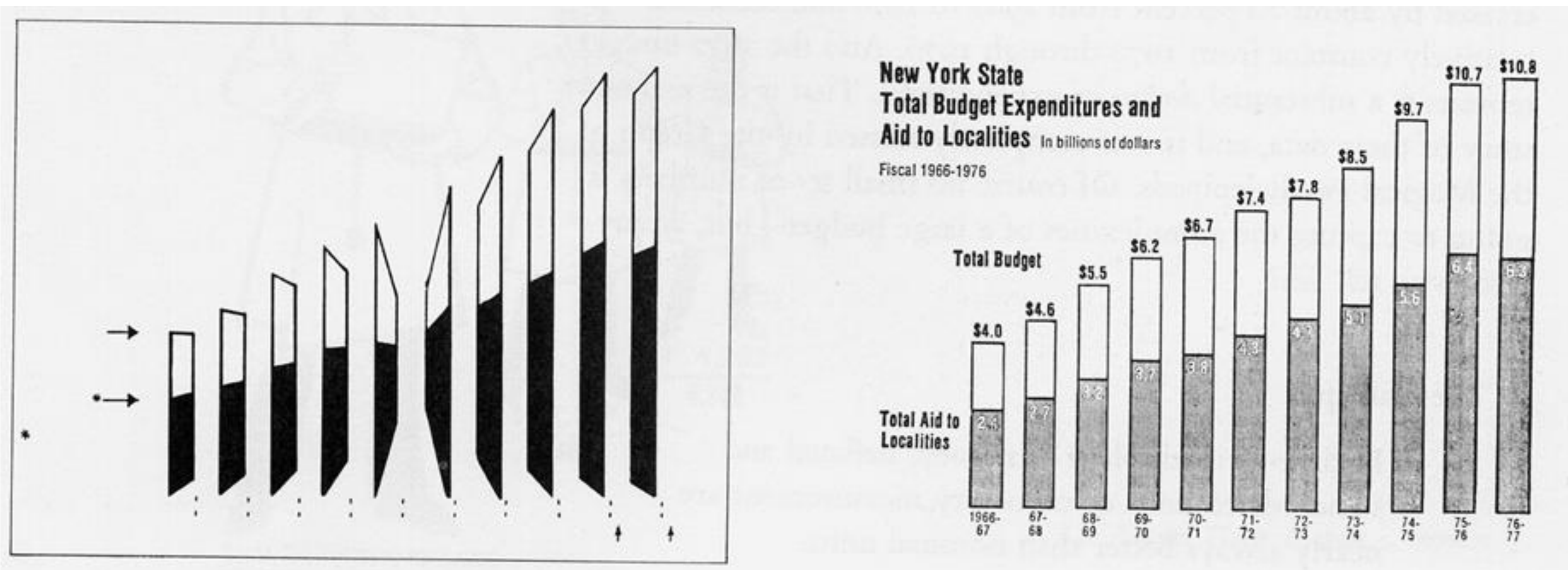
Graphical Integrity

- Oftentimes, the addition of chartjunk deceives the user (sometimes intentionally!)
- This chart is intended to convey the idea that the state budget has increased dramatically during the last nine years, when it actually hasn't grown much
- Let's remove the junk and have a fresh look



Graphical Integrity

- Now that we've removed the chartjunk, the budget increases are revealed not to be so great after all
- 3D tricks were used to make it seem like budgets of later years literally "towered over" those of earlier years



STONY BROOK
STATE UNIVERSITY OF NEW YORK

How to Measure Graphical Integrity

- It would help to have a systematic way of measuring the integrity of a graphic
- Tufte's solution: the **Lie Factor:**

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

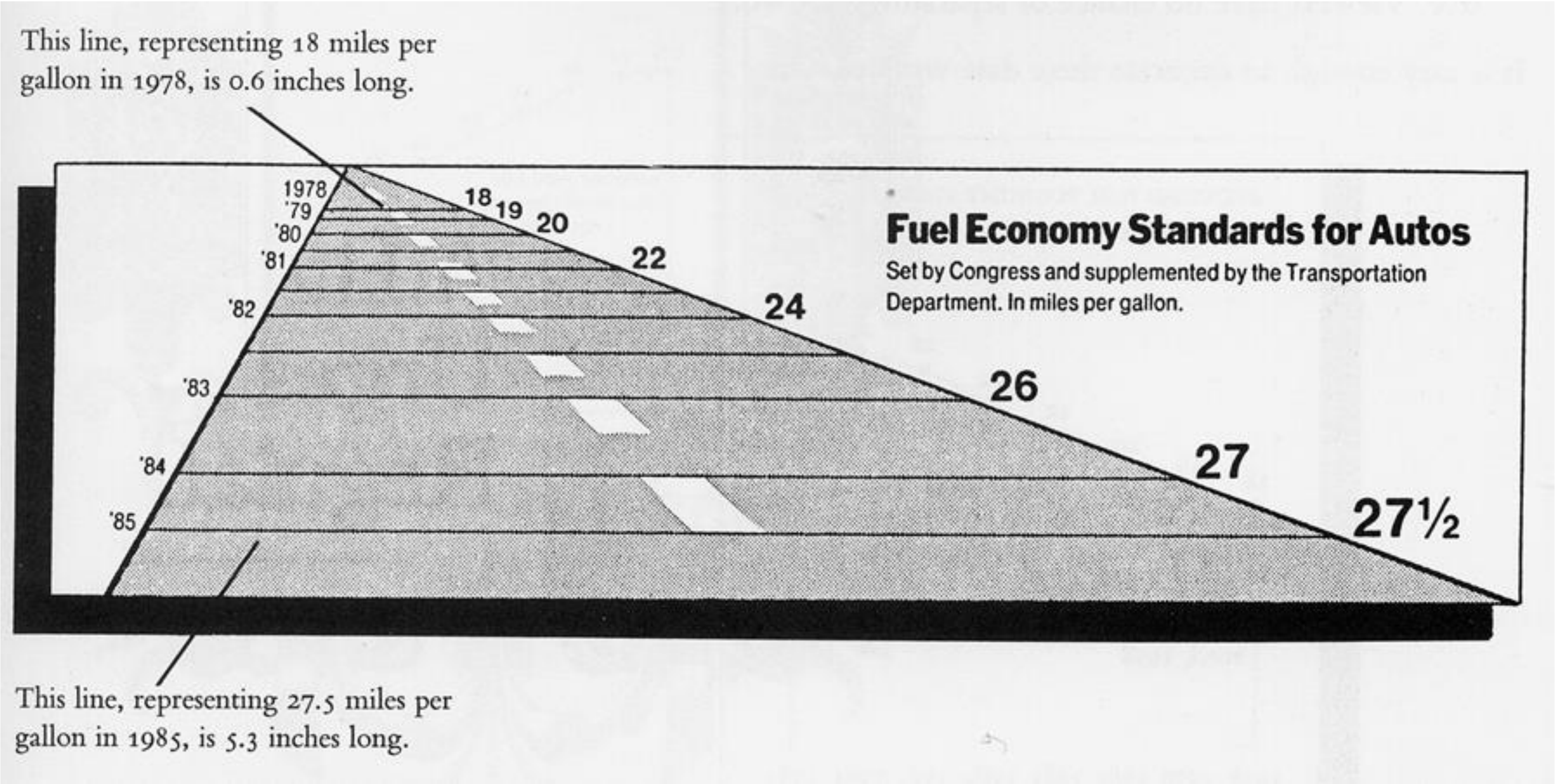
- In other words, any change suggested by the graphic should be reflected by the data
- We seek a lie factor in the range [0.95, 1.05]

- After looking at the following examples, I guarantee you will never again look at a chart in a newspaper/article/website the same way

STONY BROOK STATE UNIVERSITY OF NEW YORK

Lie Factor

- Consider the following graphic, which shows how minimum fuel standards (miles per gallon) have changed over the years:



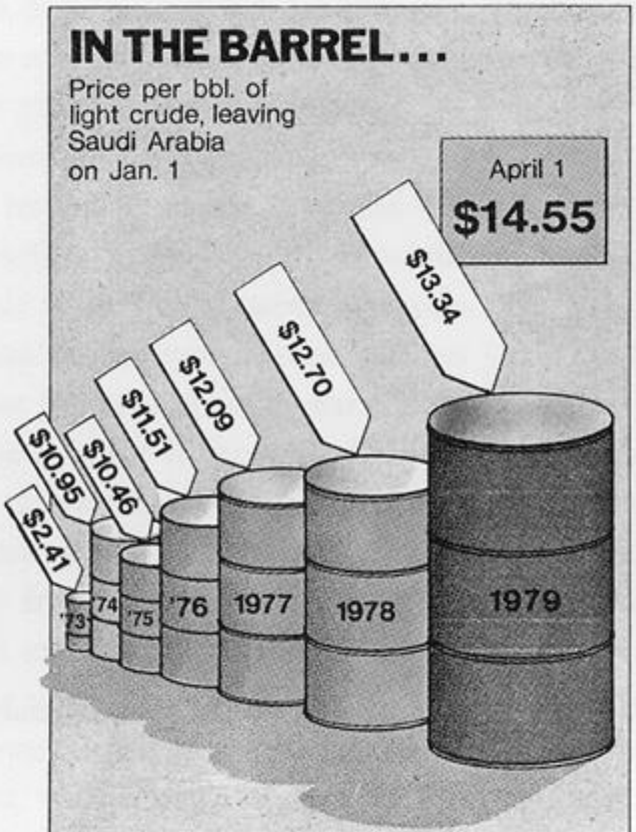
Lie Factor

- Without looking at the numbers, one would think that fuel standards have increased dramatically between 1978 and 1985
- Let's look more closely at the data
- Data: (27.5 – 18.0) / 18.0 X 100 = 53%
- Graphic: (5.3 – 0.6) / 0.6 X 100 = 783%
- This yields a lie factor of (783 / 53) or 14.8!
- The graphic depicts a change that is 14.8 times greater than what occurred in reality
- Other subtle tricks include the use of perspective, some numbers are rendered in varying type size
- Also note that the left and right diagonal lines are drawn at different angles to further mislead the viewer

Lie Factor Example

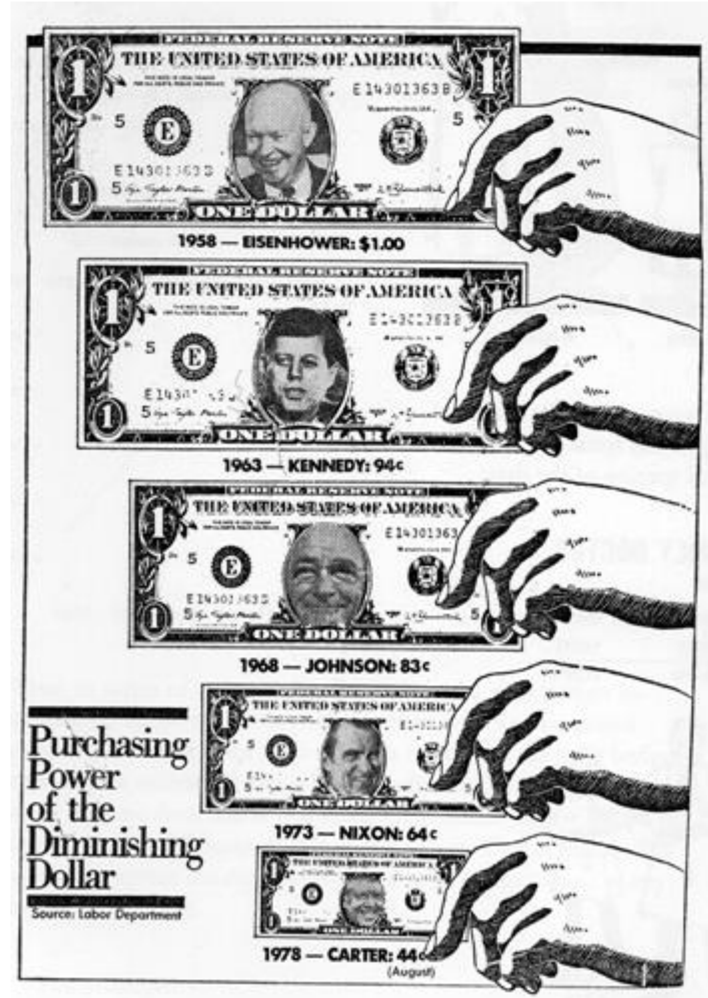
- The growing barrel
- An increase of 454% is depicted as an increase of 4280%, yielding a lie factor of 9.4
- Note also use of progressively larger labels that also seek to trick the eye and the mind

Design variation infected similar graphics in other publications. Here an increase of 454 percent is depicted as an increase of 4,280 percent, for a Lie Factor of 9.4:



Lie Factor Example

- The incredible shrinking dollar
- This figure is intended to show the diminishing purchasing power of the U.S. dollar from 1958 – 1978
- Yet, in truth, the 1978 dollar should be twice its size
- This is a very poor use of graphics – the illustrator used a 2D device (dollar bills) to show a change in a 1D quantity (money)
- Lesson: use no more dimensions than exist in the data



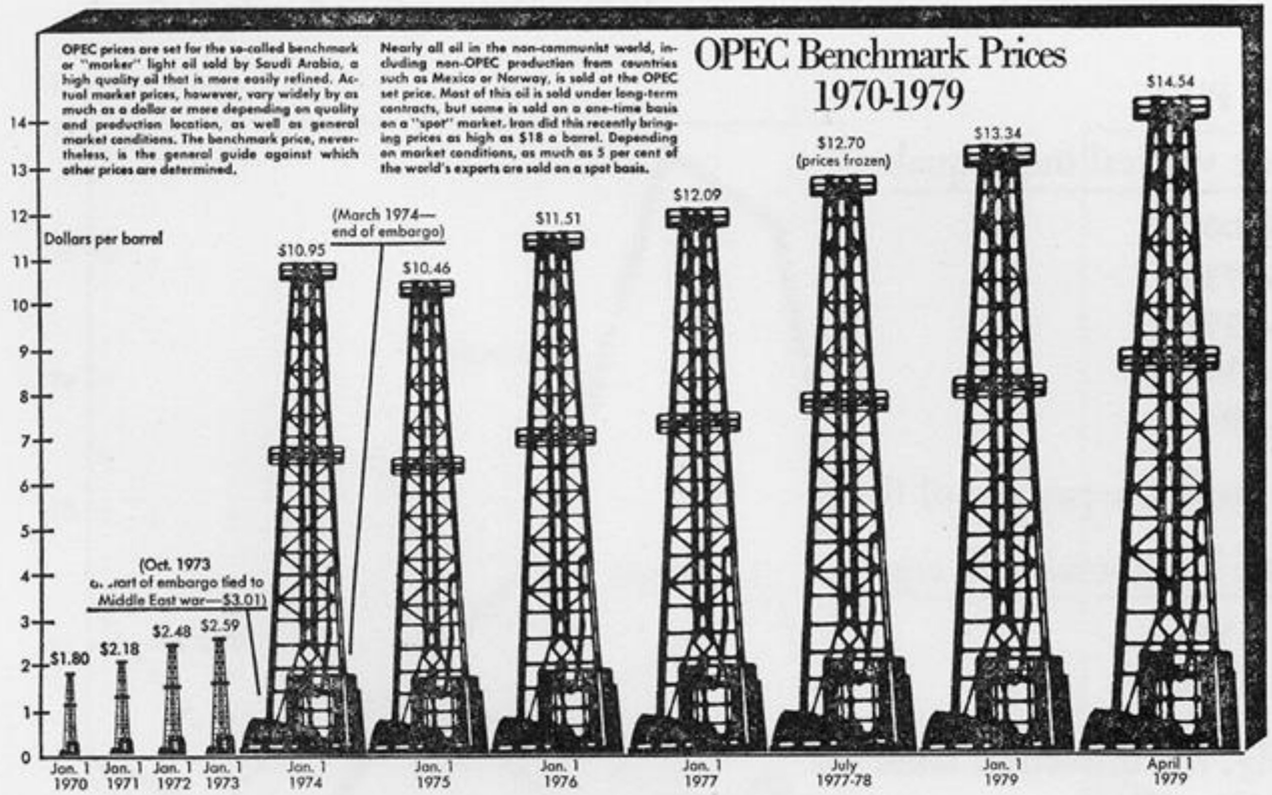
If the area of the dollar is accurately to reflect its purchasing power, then the 1978 dollar should be about twice as big as that shown.

STONY BROOK
STATE UNIVERSITY OF NEW YORK

Lie Factor Example

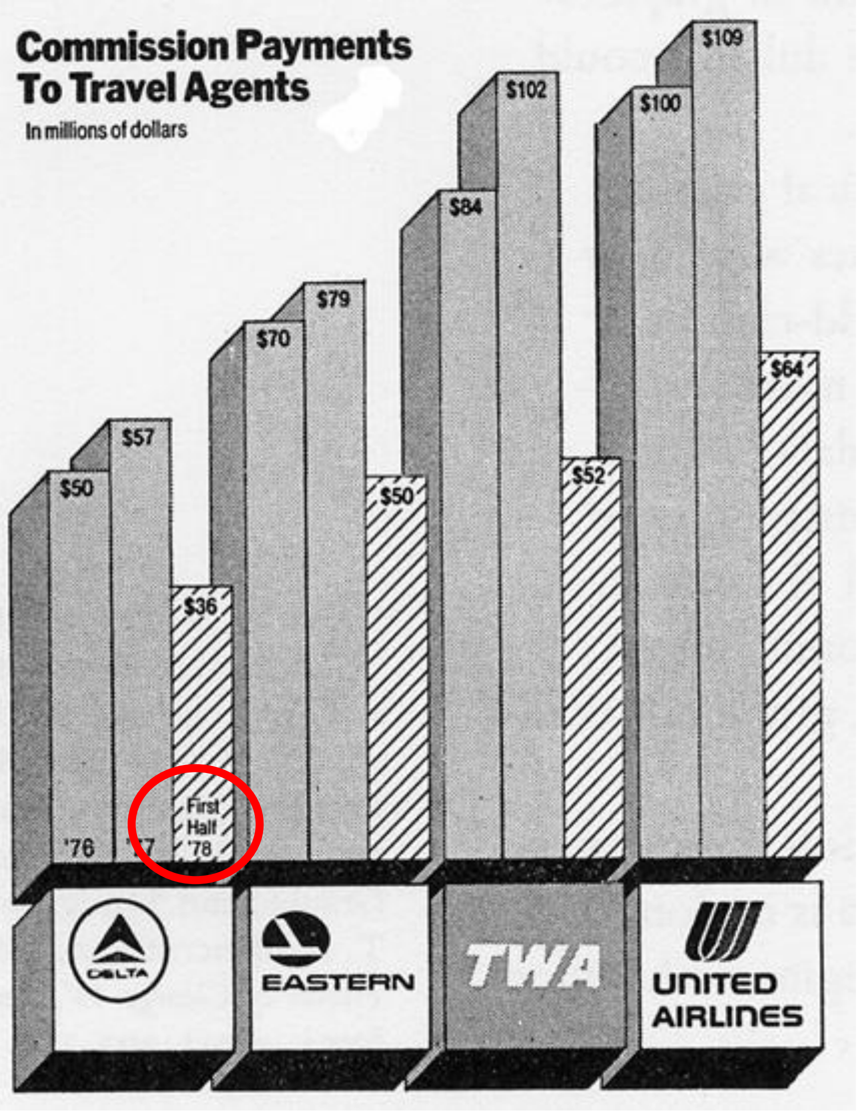
- This graph depicting the rise price of oil shows a lie factor of 9.5
- An increase of 708% is shown as a whopping increase of 6700%!

And an increase of 708 percent is shown as 6,700 percent, for a Lie Factor of 9.5:



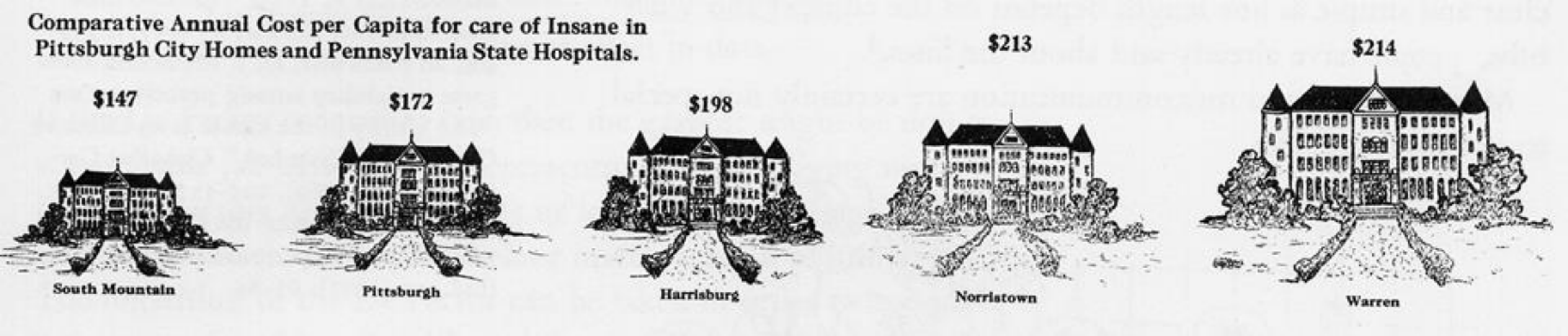
Lie Factor Example

- This graph shows how commission payments to travel agents have changed over time
- Looks like their commissions have suddenly dropped
- Oops, wait a second!
- No need to worry about travel agents after all – they are doing quite fine



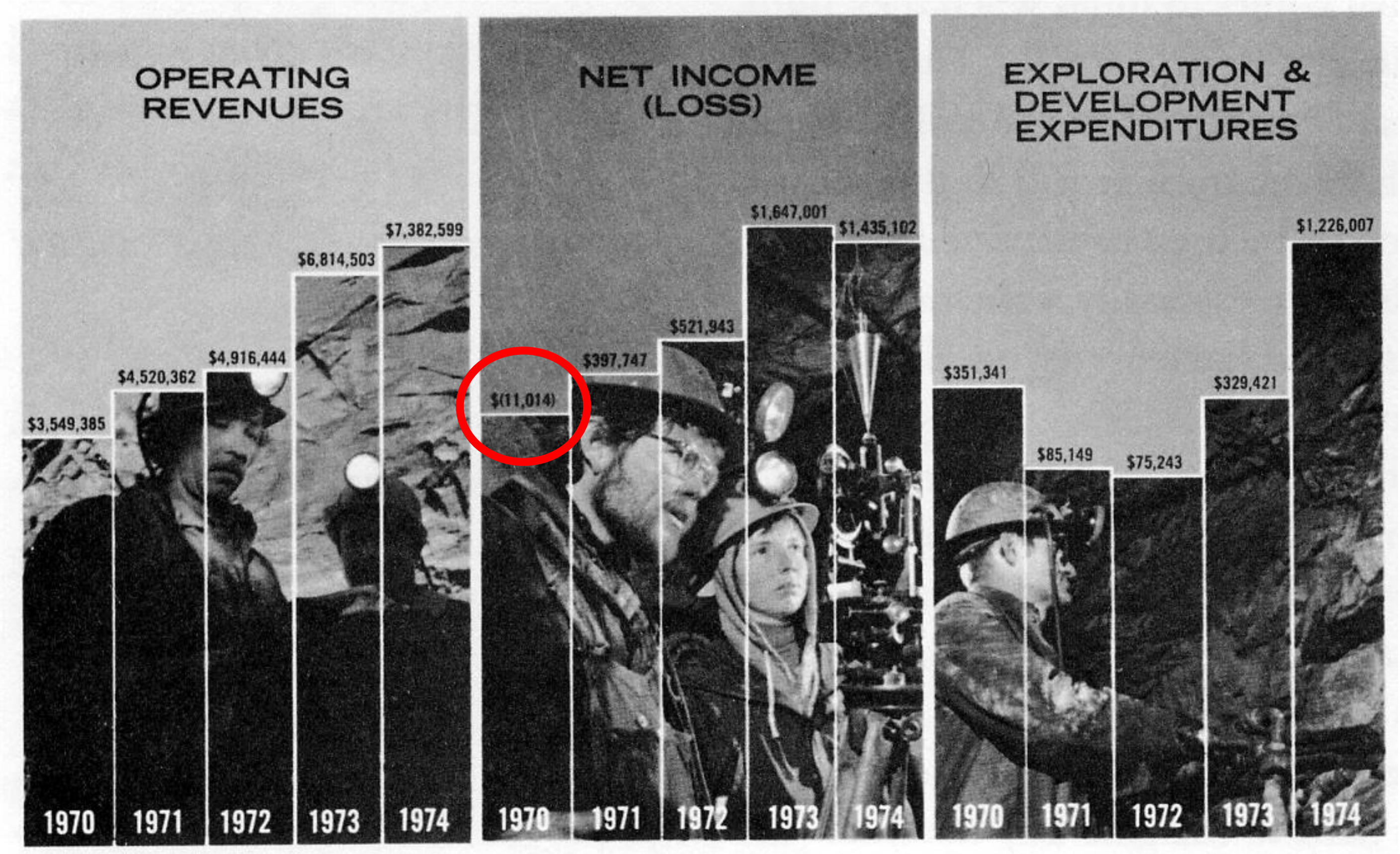
Other Misuses of Graphics

- This graphic is supposed to show the annual cost to the government to care for a mental patient in Pittsburgh City Homes and Pennsylvania State Hospitals
- The illustrator has seem to forgotten that numbers actually have magnitudes
- e.g., compare the two rightmost figures



Other Misuses of Graphics

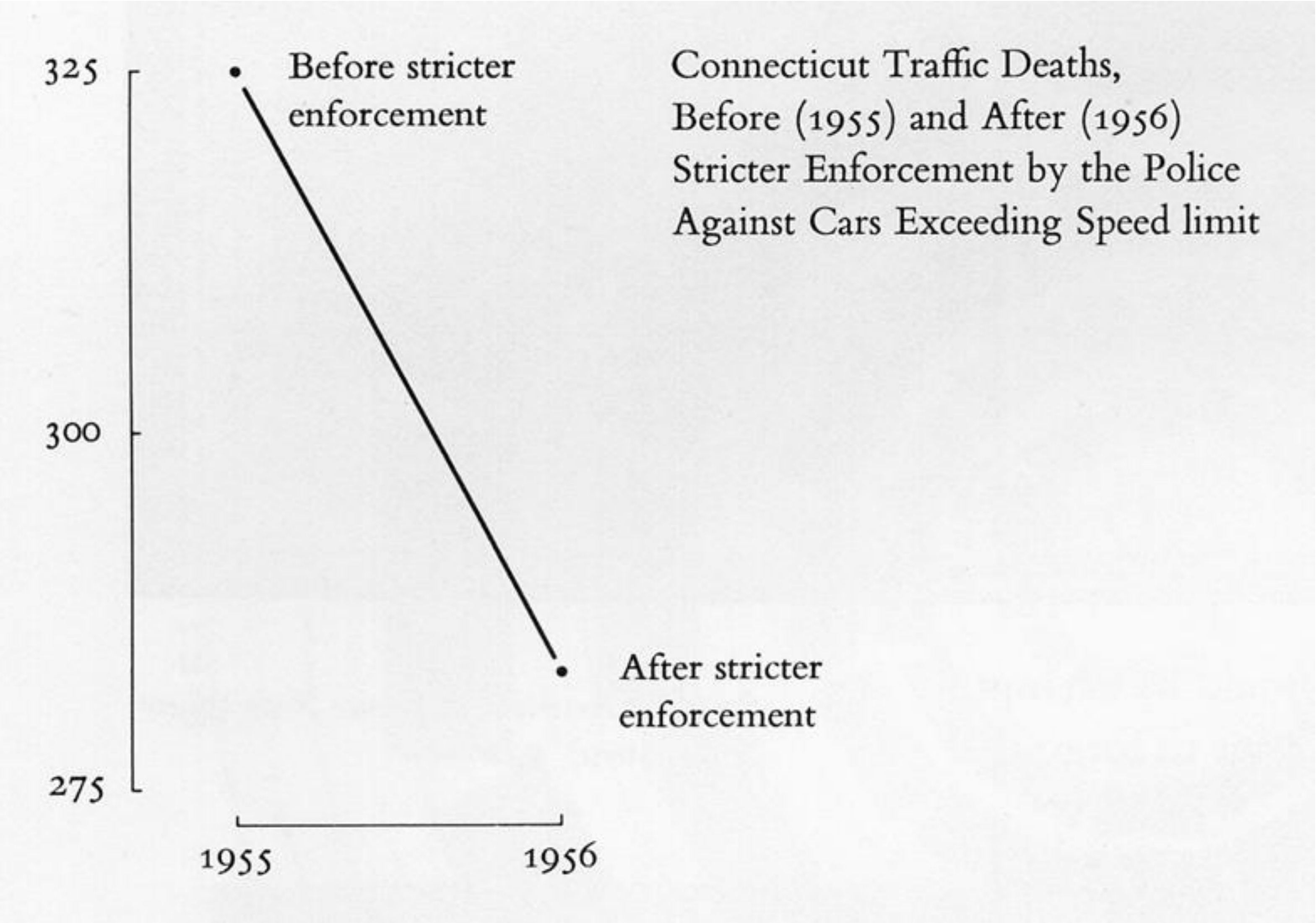
- Based on these charts, this company seems to be doing quite well
- But wait, what's this?
- In 1970 this company actually suffered a loss!



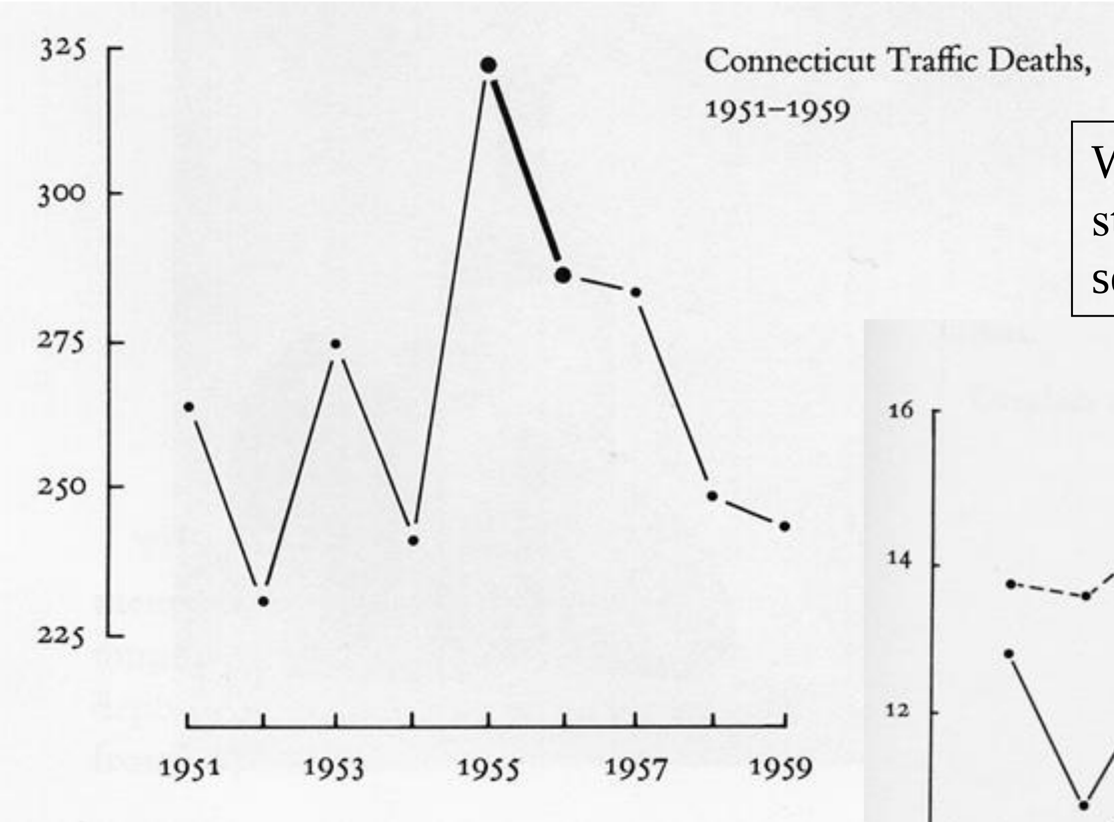
Other Misuses of Graphics

- Can you figure out what the trick is?
- The baseline at the bottom is not $0, as one would expect, but actually -$4,200,000!

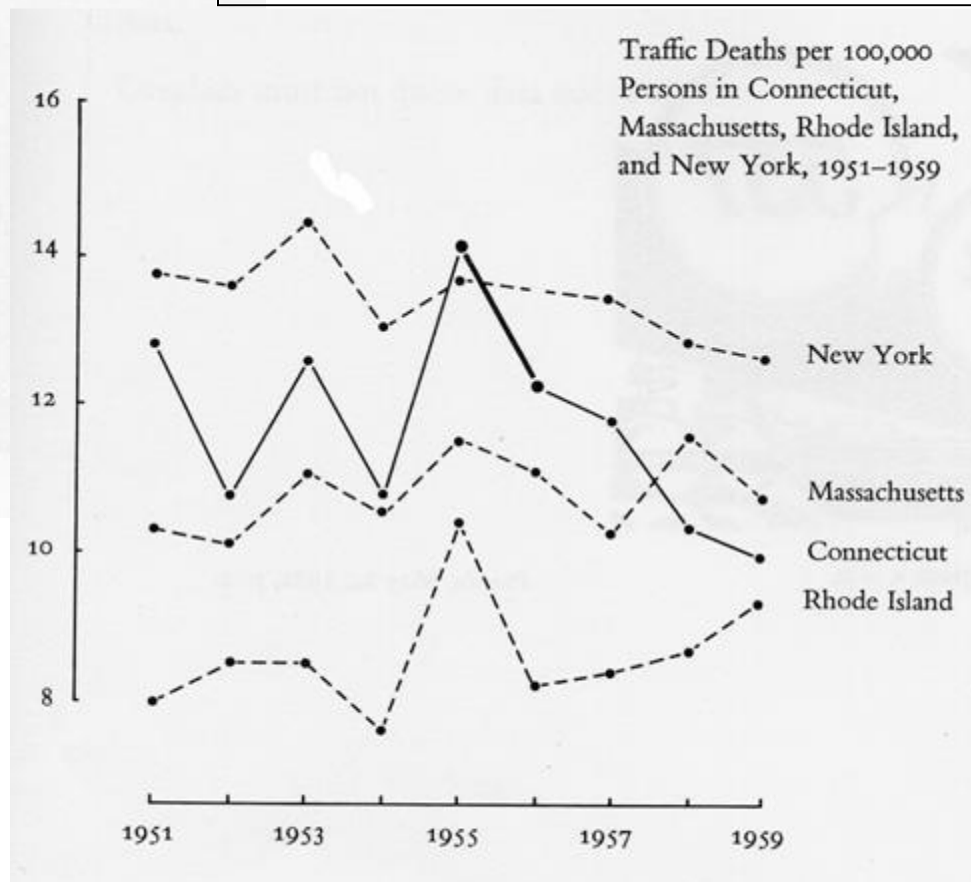
Beware of Data Taken Out of Context



Beware of Data Taken Out of Context



When compared to neighboring states, the Connecticut drop doesn't seem that special



The spike and corresponding drop might just have been part of a continuing trend and have had nothing to do with the new stricter enforcement rules

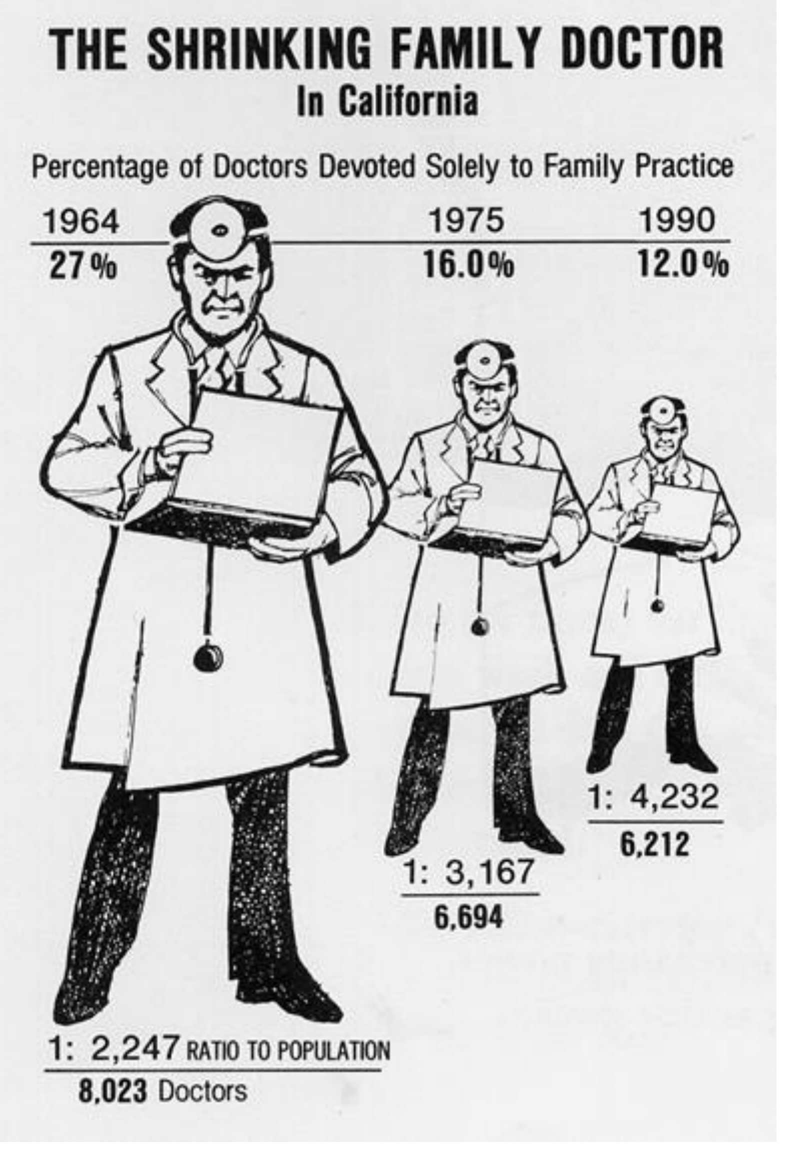
STONY BROOK
STATE UNIVERSITY OF NEW YORK

Lessons Learned

- The representation of numbers should be directly proportional to the numerical quantities represented (see the growing barrels)
- Clear and detailed labeling should be used to defeat graphical distortion and ambiguity
- Show data variations and not design variations (see the fuel economy graph)
- In time-series displays of money, show deflated and standardized units
- The number of information carrying dimensions should not exceed the data dimensions (see the growing barrels)
- Graphics must not quote data out of context (see Connecticut traffic deaths)
- Convincing graphics must demonstrate cause and effect (Challenger disaster)

Review: Traditional Visualization

- Visualization of data should convey to the user/viewer the greatest amount of information in the shortest time and in the smallest place
- Have a discerning eye – be wary of fancy graphics and figures; there may be a hidden agenda or just the exercise of poor judgment
- Lie Factor concept – change suggested by graphic should represent the change reflected by the data
- Lie factor of 2.4 in this example:



Visualization Enters the Digital Age

- Since the advent of powerful computers, visualization has really just entered its golden age
- Medical imaging, for instance, is a field growing approximately 40% a year (increased applications, growing/aging populations)
- Now we'll look at the fundamentals of image generation and display
  - How the eye perceives light and color
  - How images are constructed
  - How color is manifested on-screen and in computer code
- Next we'll apply this knowledge in a critical aspect of visualization: the use of image processing techniques to extract important information from digital images