CSE 564: Visualization

The Views of Edward Tufte (and Some Others)

Klaus Mueller

Computer Science Department

Stony Brook University

Seminal Books by Edward Tufte

Standard literature for every visualization enthusiast

- written 1983, 1990, 1997, 2006

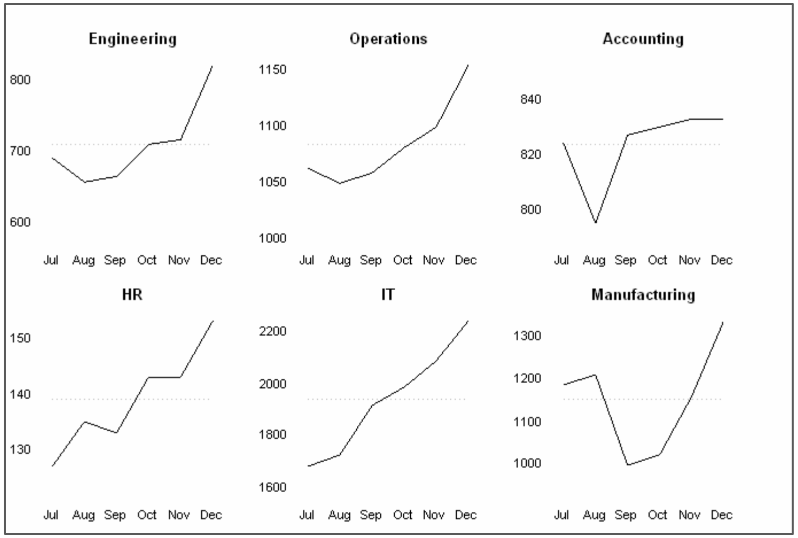
Edward Tufte

Well recognized for his writings on information design

- a pioneer in the field of data visualization
- taught information design at Princeton University
- now a professor at Yale University

Popularized concept of "small multiples"

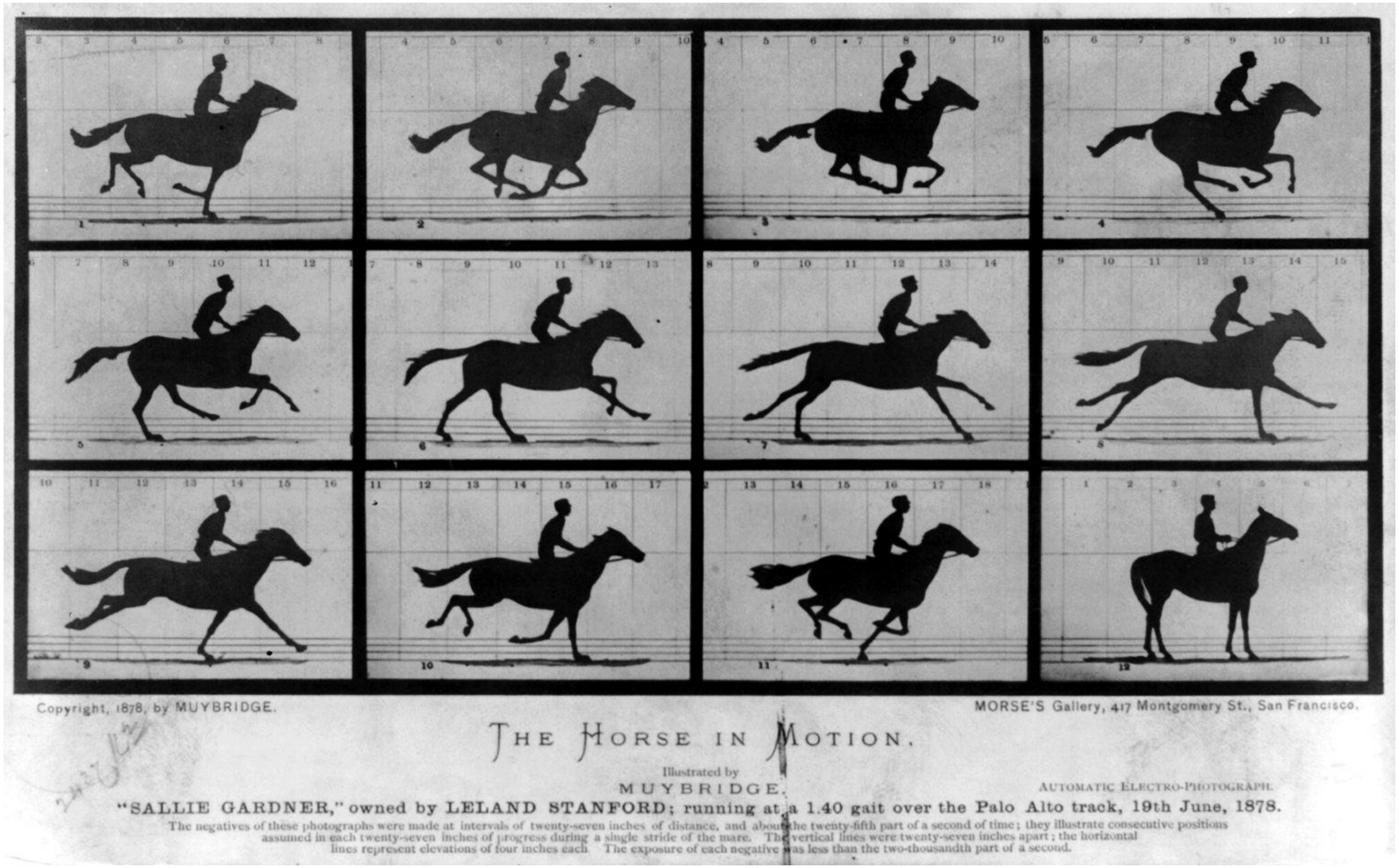
- aka trellis chart or panel chart
- similar charts of same scale + axes
- allows them to be easily compared
- use multiple views to show different partitions of a dataset



Small Multiples – Historical Reference

E. Muybridge's Horses in Motion (1886)

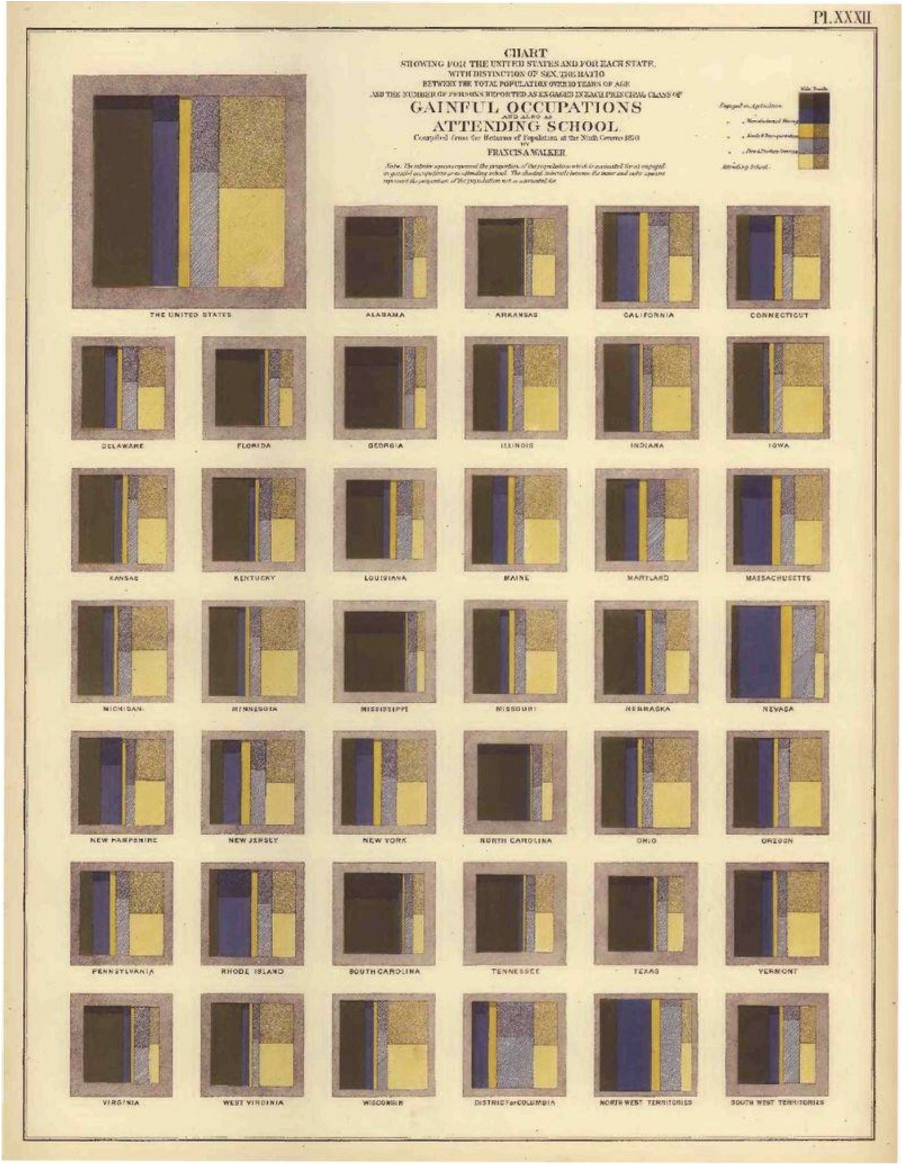
- proofed for the first time that horses CAN have all 4 legs in the air
- work was also foundational to the development of the motion picture



Small Multiples – Historical Reference

FA Walker's census charts (1870)

- population is broken down by state and then occupation, including a count of those attending school
- also has tree maps!



Edward Tufte

Also popularized “sparklines”

- small integrative visualizations

Symbol	Bid	Ask	Last	Change	T	Chart	Volume	High	Low	Value Change		Value	Gain	
DELL	89 3/4	89 13/1	89 3/4	+ 1 1/4			10,310,100	90 1/8	88 1/2	+1.41%	250	17,950	+273.72	13,147
CPQ	48 7/16	48 9/16	48 7/16	- 13/16			25,628,700	51 1/4	1/4	-1.65%	-81	4,844	+60.79%	1,831
SDTI	26 1/4	26 3/8	26 3/8	+ 1/2			504,600	27 3/8	25 5/8	+1.93%	250	13,188	+133.15	7,531
COMS	46 1/2	46 9/16	46 9/16	- 25/32			3,191,100	47 15/1	45 3/4	-1.65%	-102	6,053	+29.79%	1,389
LU	111 5/8	111 11/1	111 9/16	+ 1 9/16			5,104,600	112 5/8	110	+1.42%	78	5,578	+22.76%	1,034
YHOO	368 1/16	368 1/2	368 1/2	+ 17 1/4			3,787,800	381 3/16	280	+4.91%	431	9,213	-0.41%	-38
AOL	162 13/16	163	163	+ 8			10,008,500	164	158 1/2	+5.16%	280	5,705	+73.06%	2,408
CMGI	97 3/8	97 1/2	97 1/2	+ 5 7/8			1,323,800	98 1/2	93	+6.41%	705	11,700	+186.76	7,620
SPLN	33 13/16	33 15/1	33 13/16	+ 7/16			300,200	34 3/4	33 5/8	+1.31%	88	6,763	+94.60%	3,288
BEAS	13 1/2	13 5/8	13 5/8	- 7/16			389,200	14 1/4	13 1/8	-3.11%	-44	1,363	-9.17%	-138
GNET	102	103 3/16	101 5/16	+ 6 1/8			307,600	108	97	+6.43%	613	10,131	+130.26	5,731
RNWK	67	67 1/4	67	+ 2 3/4			1,233,900	69	64 15/1	+4.28%	275	6,700	+79.87%	2,975
MSFT	173 1/8	173 1/4	173 5/16	+ 1 3/4			13,284,500	174 7/16	170	+1.02%	175	17,331	+54.74%	6,131
INTC	133 3/4	133 13/1	133 13/16	- 3 1/8			8,094,300	137 1/2	133 3/8	-2.28%	-625	26,763	+65.20%	10,563
TOTAL								205,302	80,993	+1.63%	2,293	143,280	+79.41%	63,377

Sparklines inspired “word size visualizations”

- charts or graphs tightly integrated into text or even computer code

Although Tufte is said to have invented sparklines, in actuality he invented only the name and popularized it as technique.[15] Sparklines are a condensed way to present trends and variation, associated with a measurement such as average temperature or stock market activity, often embedded directly in the text; for example: The Dow Jones index for February 7, 2006 [16][17] These are often used as elements of a small multiple with several lines used together. Tufte explains the sparkline as a kind of "word" that conveys rich information without breaking the flow of a sentence or paragraph made of other "words" both visual and conventional. To date, the earliest known implementation of sparklines was done by interaction designer Peter Zelchenko and programmer Mike Medved in early 1998.[18]

Tufte on Graphical Excellence

According to Tufte (pg. 51):

- Graphical excellence is the well-designed presentation of interesting data

    - a matter of s**ubstance, statistics**, and **design**

- Graphical excellence consists of complex ideas communicated with:

    - **clarity**, **precision**, and **efficiency**

- Graphical excellence is that what gives the viewer:

    - the **greatest number of ideas**
    - in the **shortest time**
    - with the **least ink**
    - in the **smallest space**

- Graphical excellence is nearly always multivariate
- Graphical excellence requires telling the truth about the data

(Nevertheless, visualizations should be visually pleasing and may very well have an artistic touch)

The Need for Visualization: Anscombe Quartet

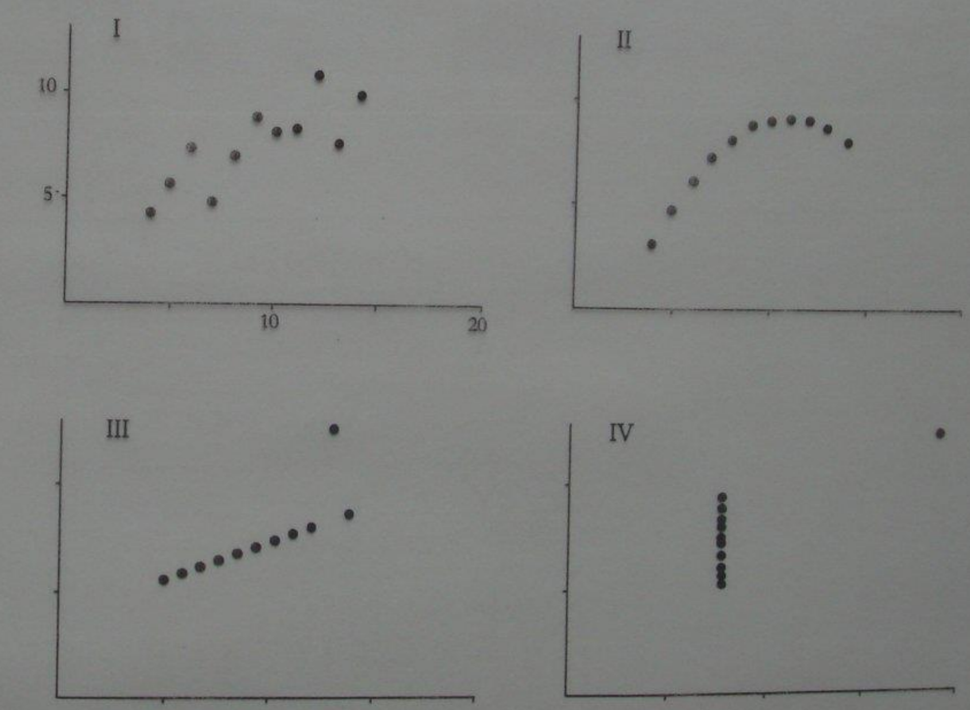
I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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
$r^2 = .67$

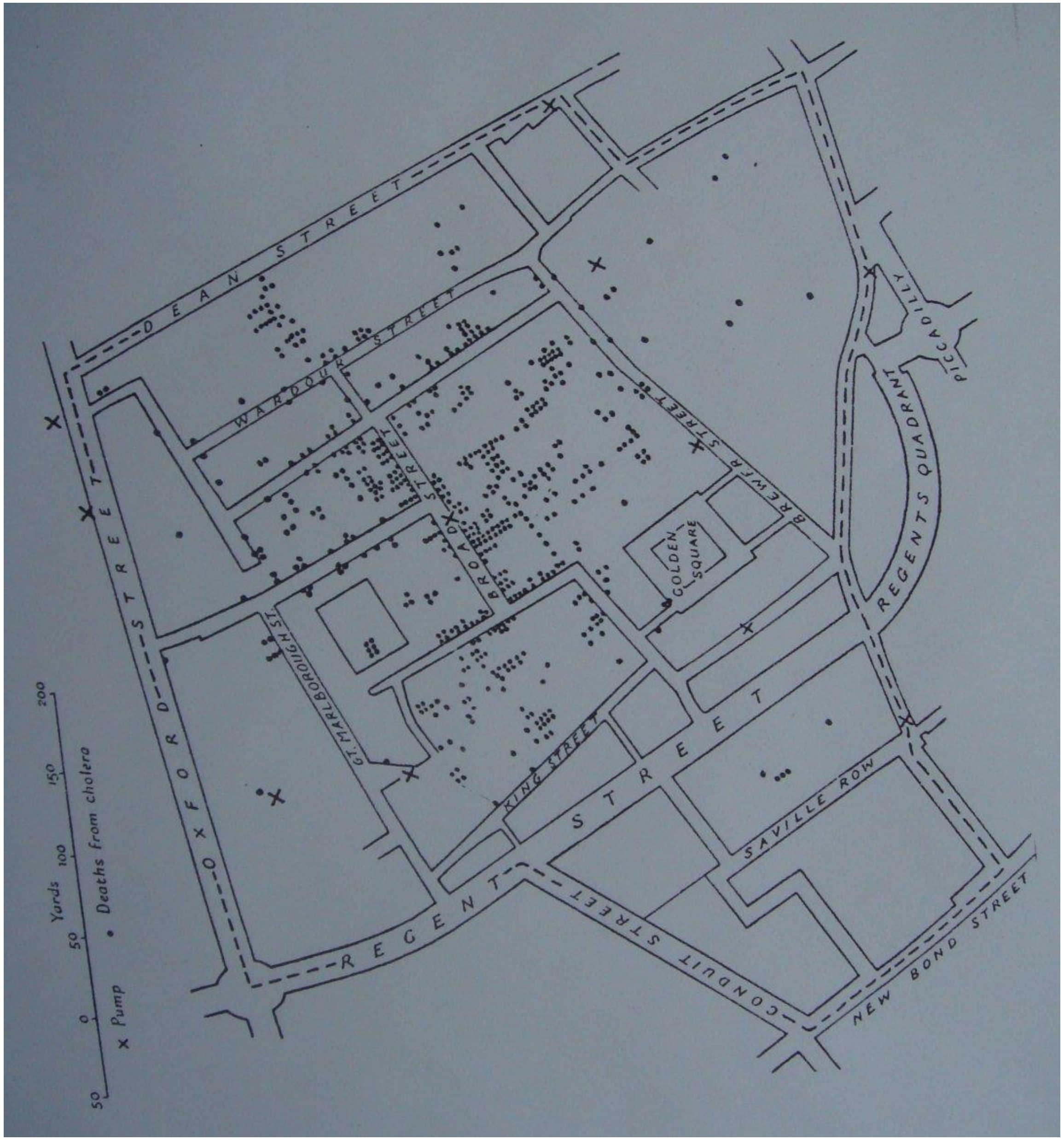
The Need for Visualization: Anscombe Quartet

I X	I Y	II X	II Y	III X	III Y	IV X	IV 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

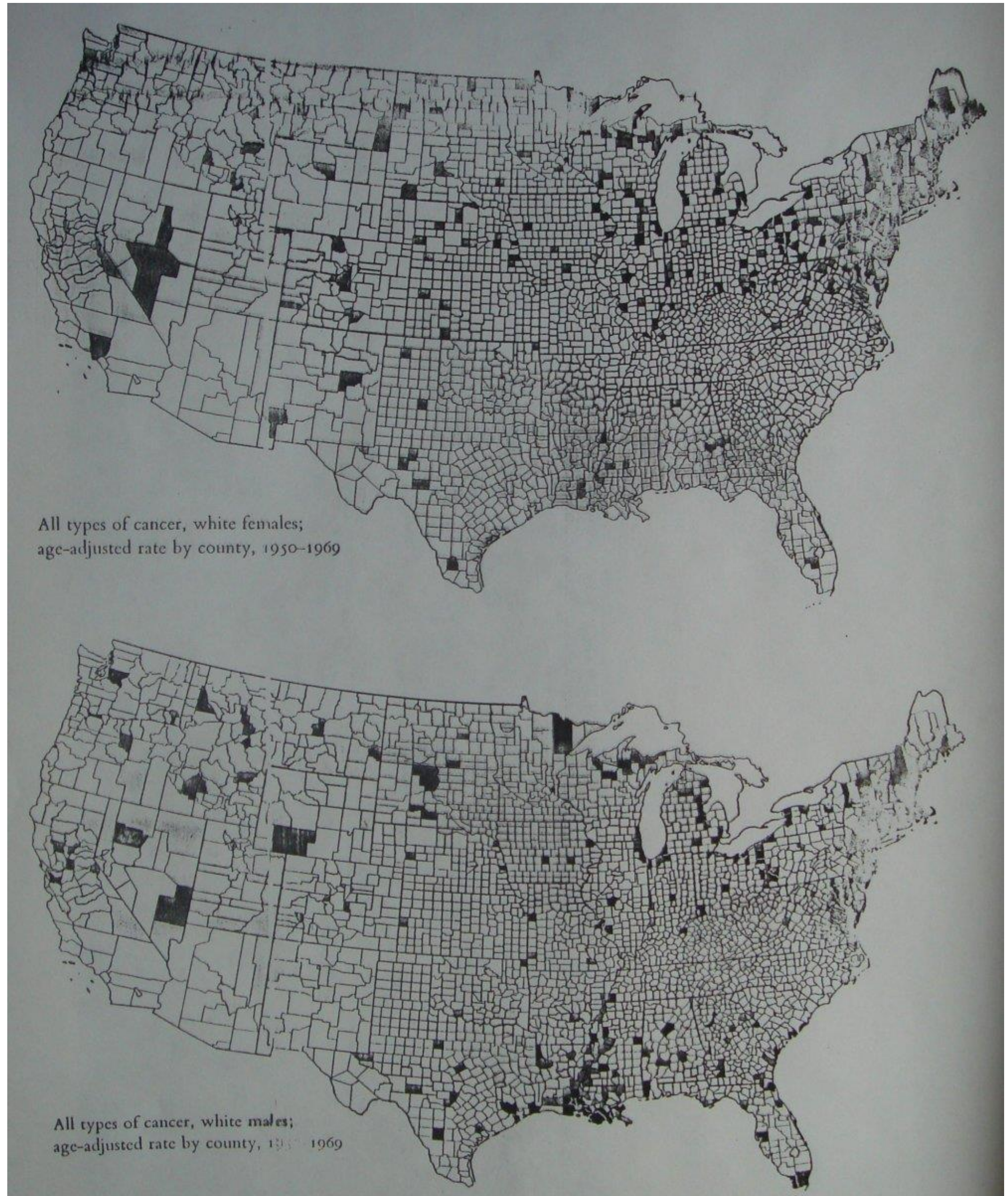
N = 11
mean of X's = 9.0
mean of Y's = 7.5
equation of regression line: Y = 3 + 0.5X
standard error of estimate of slope = 0.118
t = 4.24
sum of squares $X - \overline{X}$ = 110.0
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
r^2 = .67



John Snow: London Cholera Map (1854)



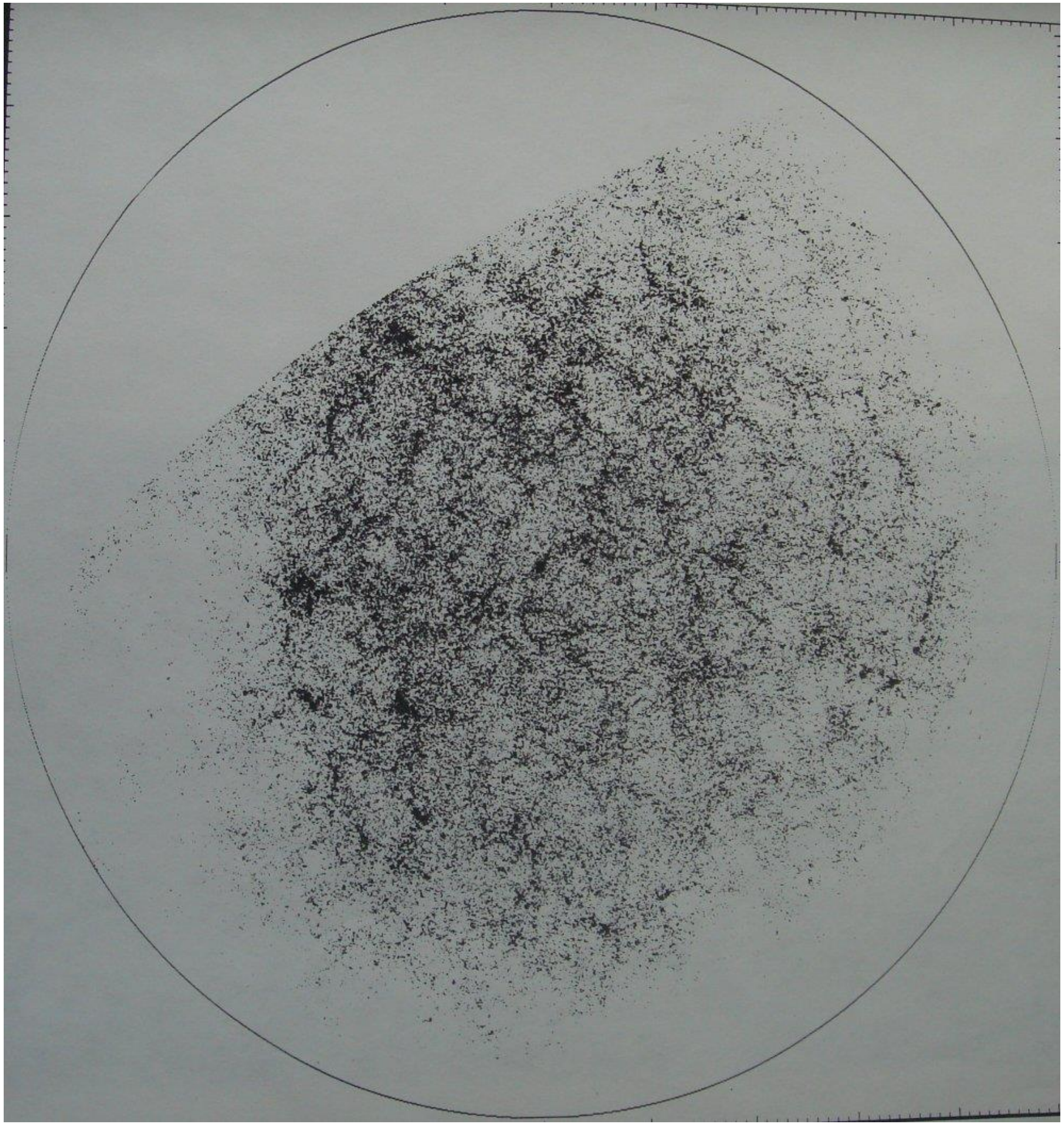
Age-Adjusted Cancer Rates (by County)



21,000 numbers
3056 counties
7 numbers per county:
- size (4)
- location (2)
- cancer rate (1)

1950-1969

Galaxy Maps

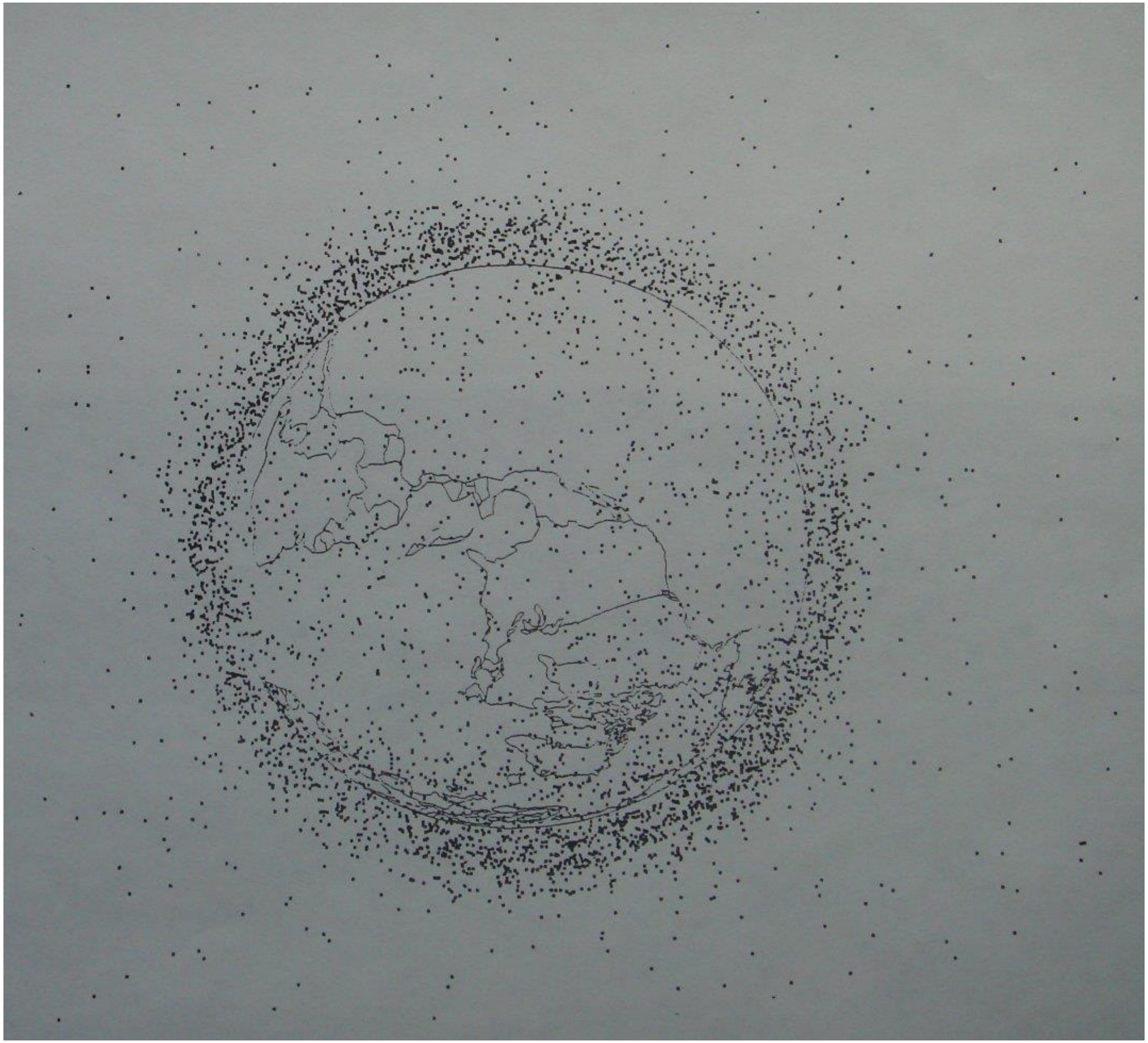


divide sky into
1,024 x 2,222 rectangles

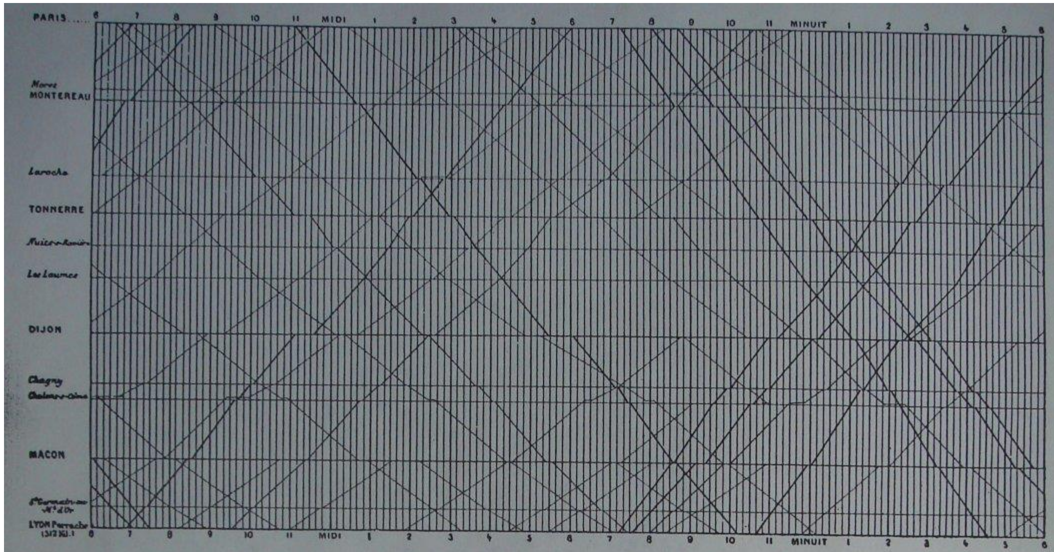
tone = number of galaxies
per rectangle

Space Debris Map (1990)

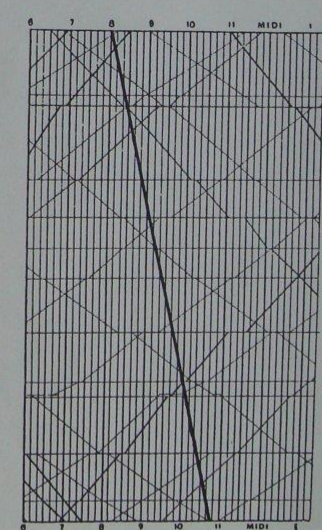
7,000 objects > 10 cm
doubles every 5 years



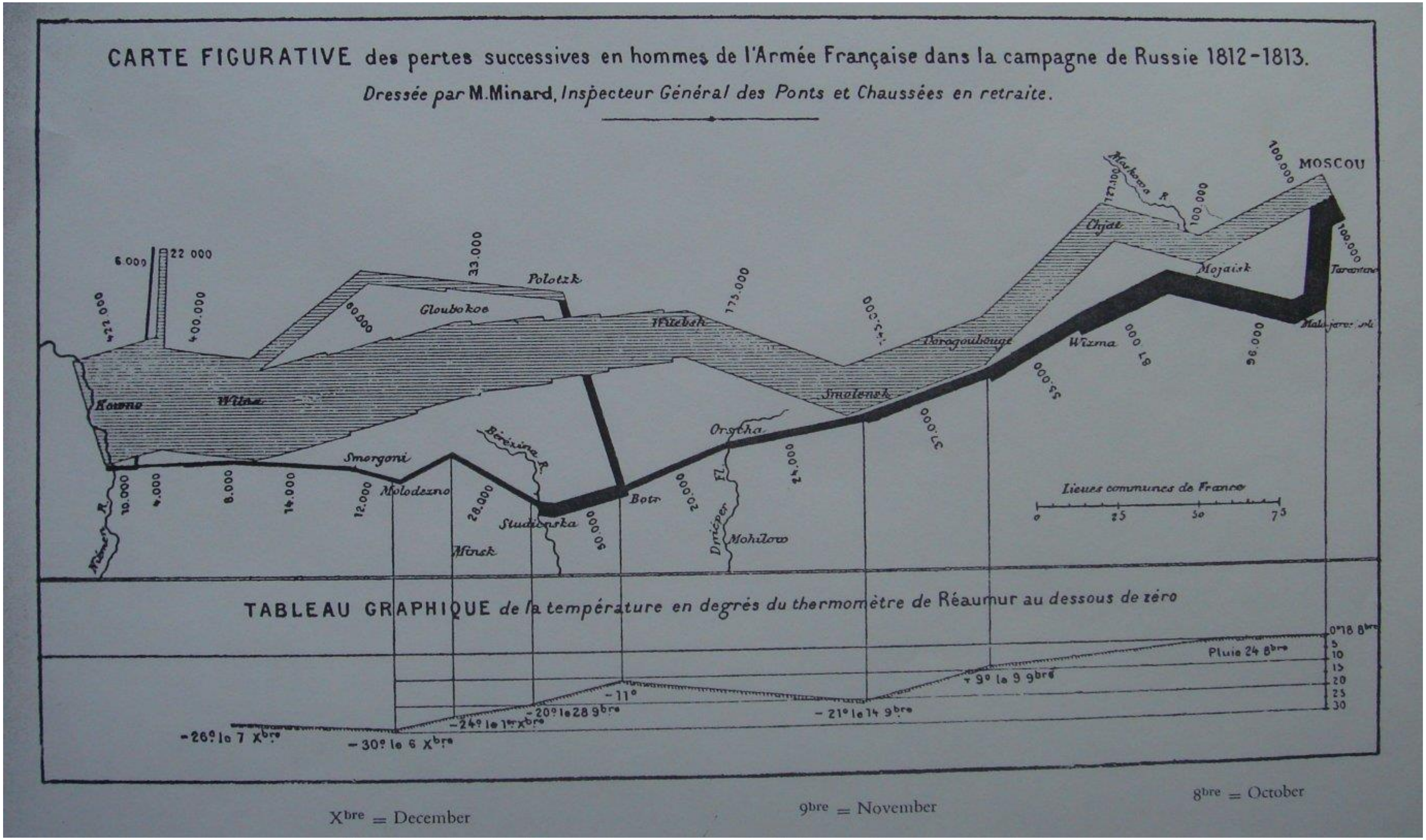
Train Schedule: Paris – Lyon, 1880s



E. J. Marey, *La Méthode Graphique* (Paris, 1885), p. 20. The method is attributed to the French engineer, Ibry.

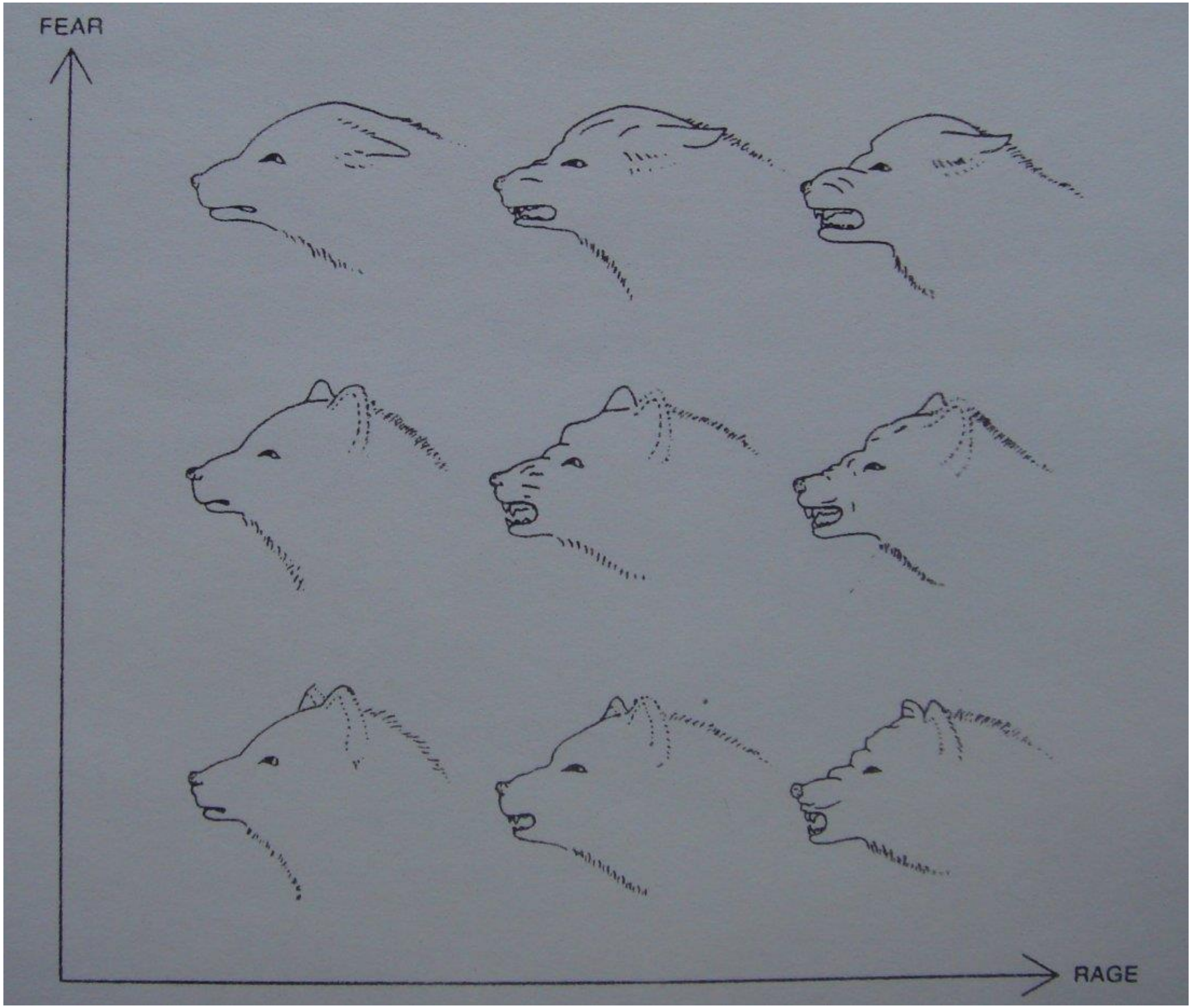


Minard: Visualization of Napoleon's Russia Campaign (1812)

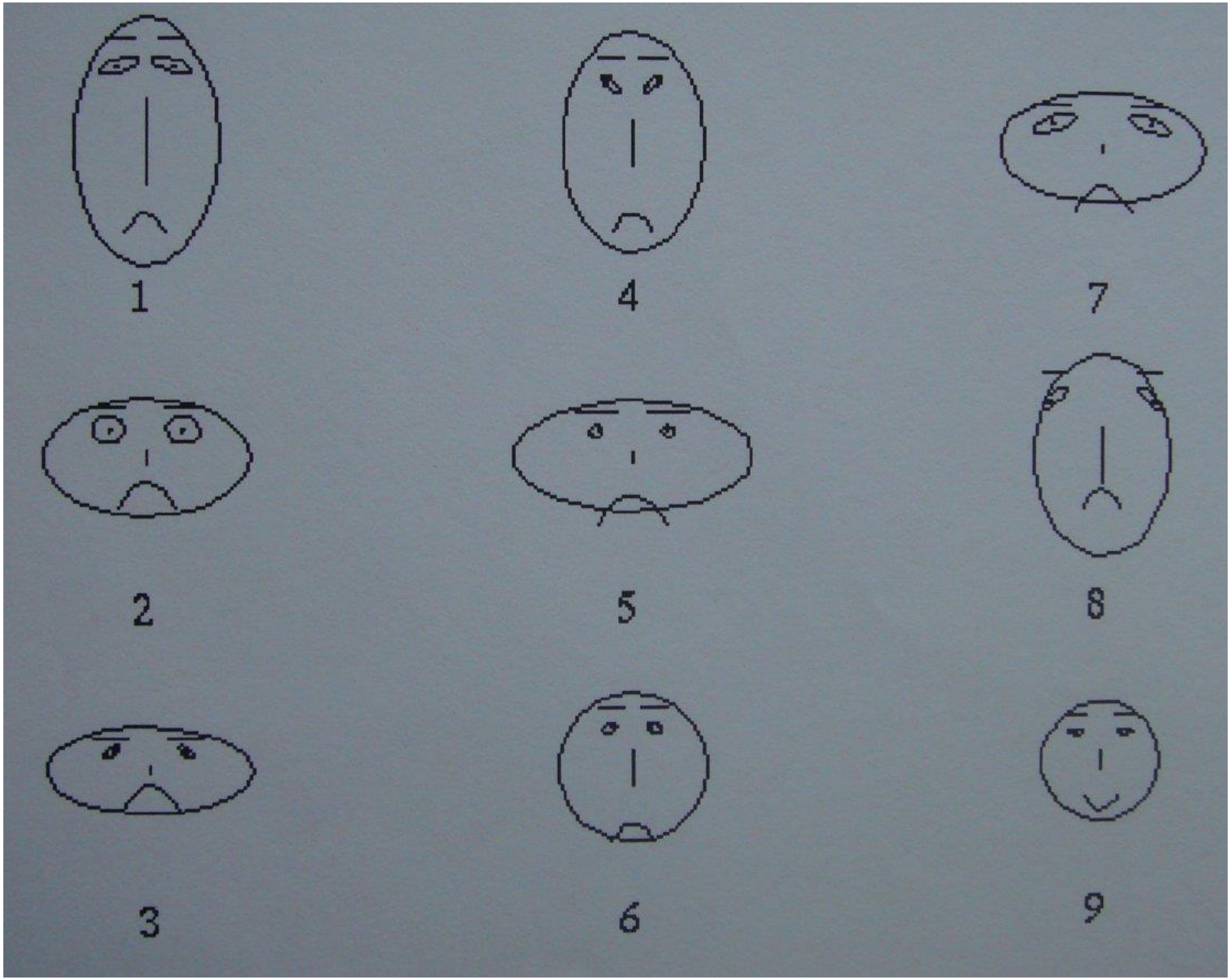


plots 6 variables: army size, 2D location, direction vector, temperature, time

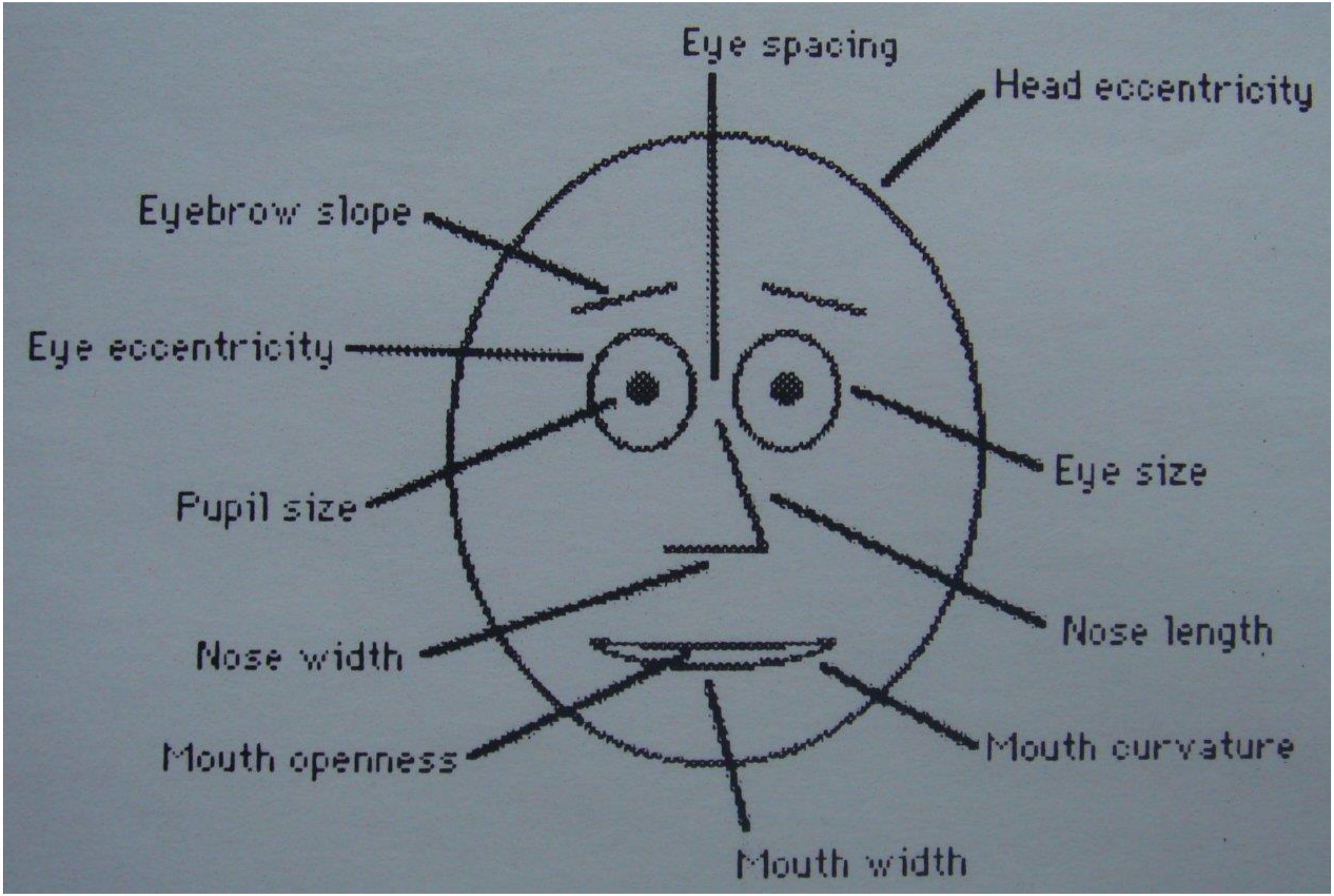
Rage Fear Graph: Expressive Glyphs



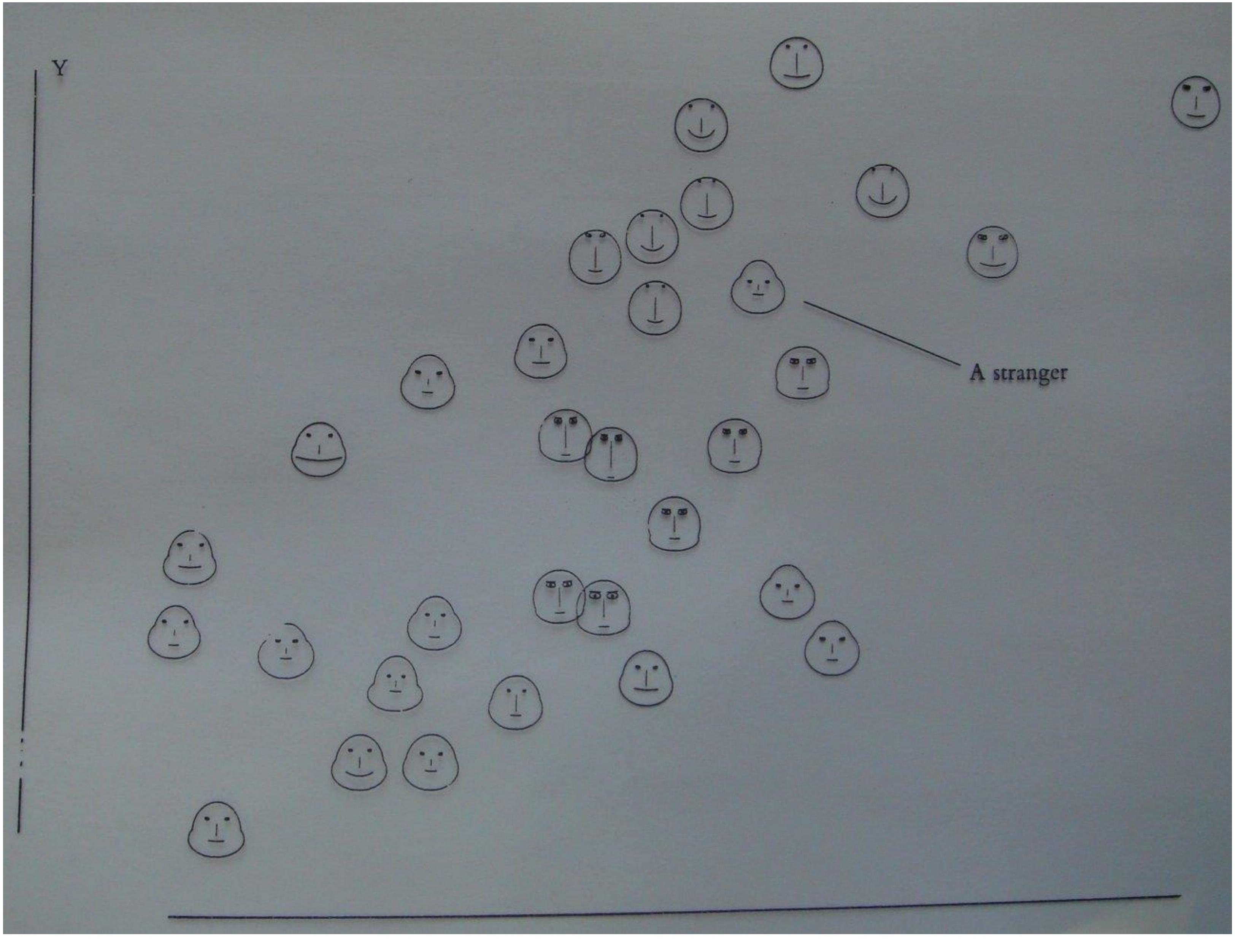
Chernoff Faces: Multi-Variable Display



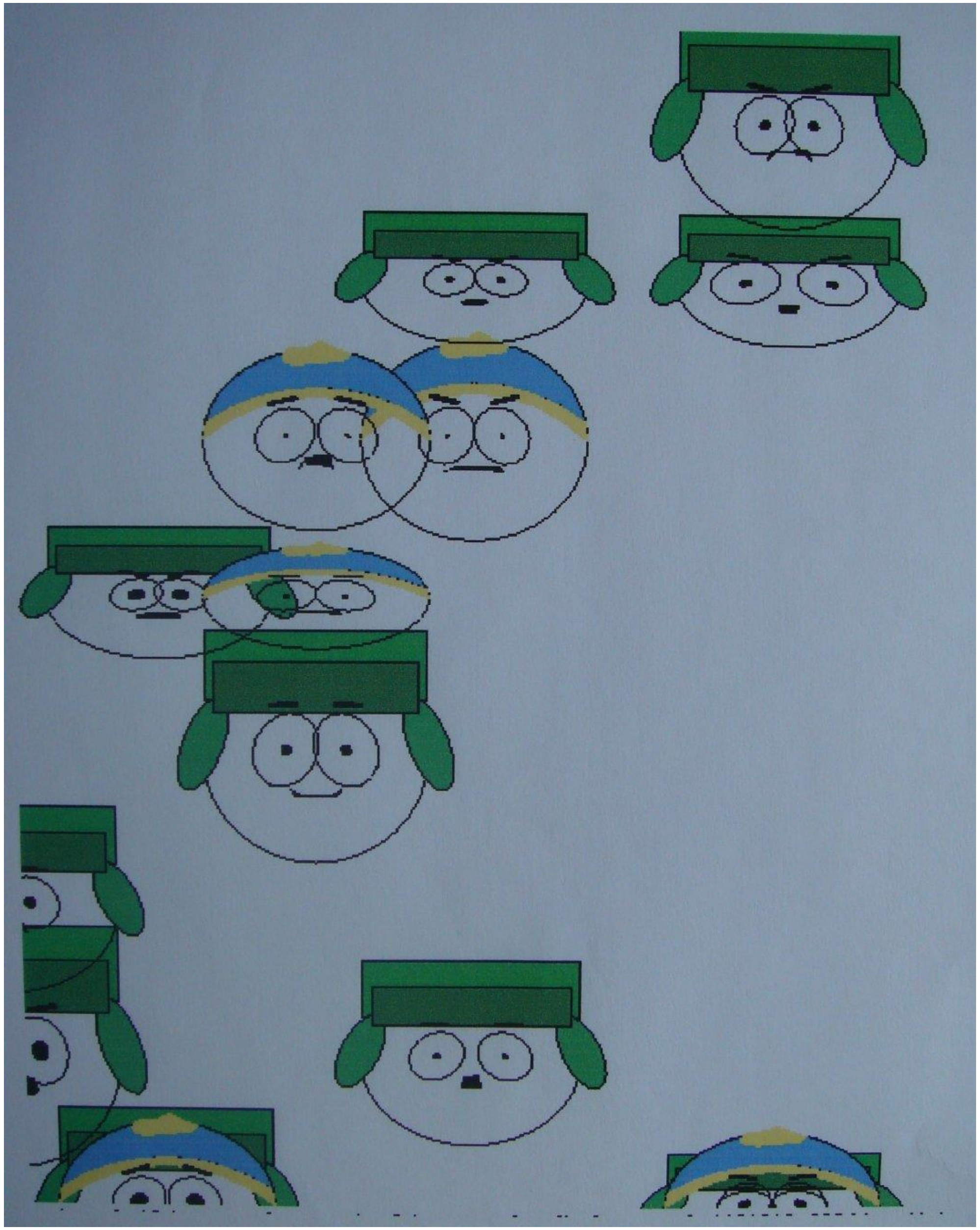
Chernoff Faces



Chernoff Faces



Chernoff Faces



Graphical Display: History

- Can be more precise and revealing than numerical display

  - example: Anscombe's quartet (pg. 13/14)
  - example: cholera map of central London, 1854, by Dr. John Snow (pg. 24)

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

  - example: data maps for cancer incidence (pg. 17)
  - example: galaxy maps (pg. 27)
  - example: space debris (pg. 48, Tufte "Envisioning Information")

- Can extend to time-series display

  - example: train schedule Paris-Lyon, 1880s (pg. 31)

- Can be narrative

  - example: Napoleon's Russia campaign, 1812, plots 6 variables on a 2D graph (pg. 41)

- Can represent each datapoint by visual information (graphic, icon, image, color, pattern)

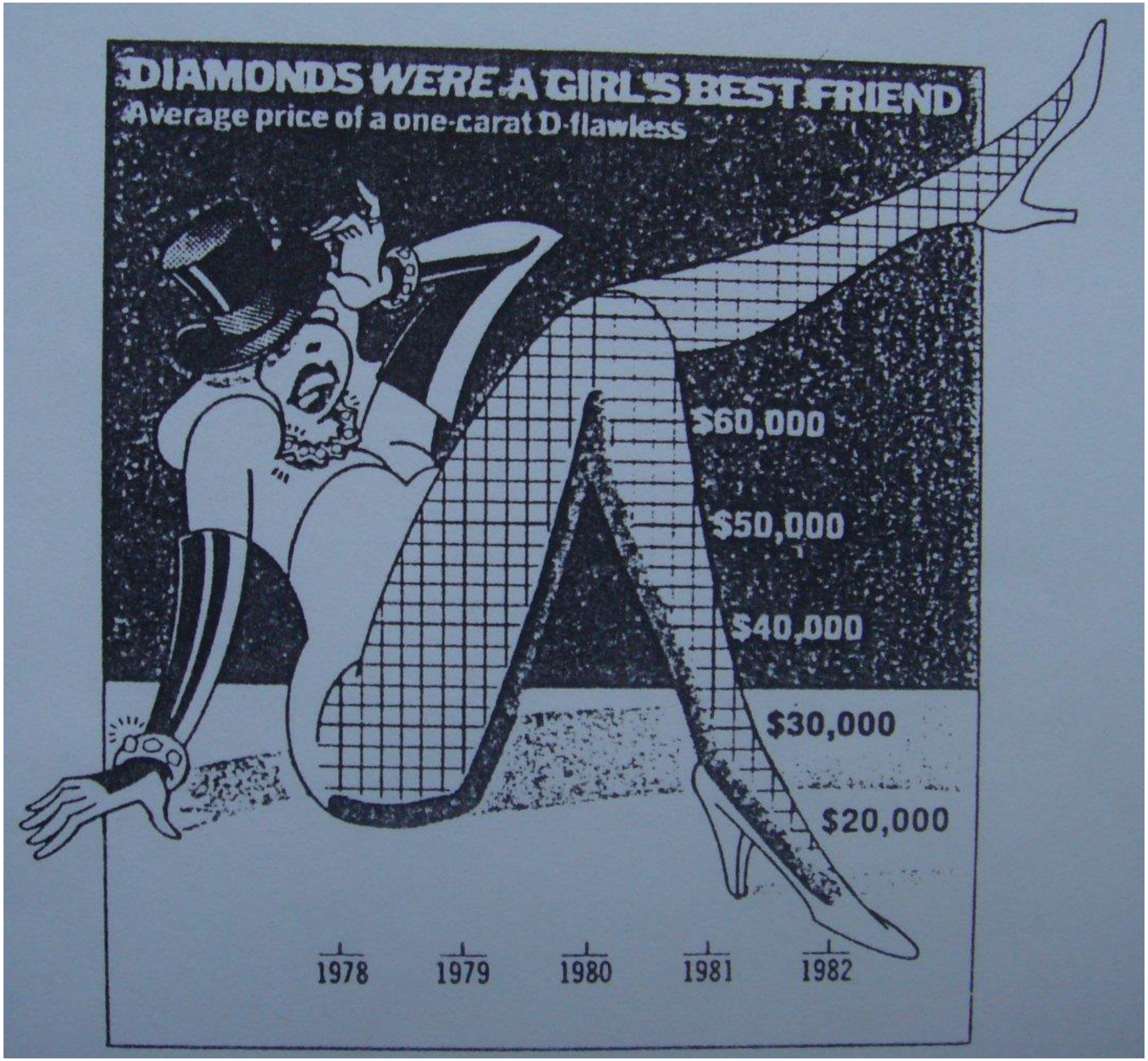
  - examples: fear-rage graph (pg. 50), Chernoff faces (pg. 97, 142)

Next Slides…

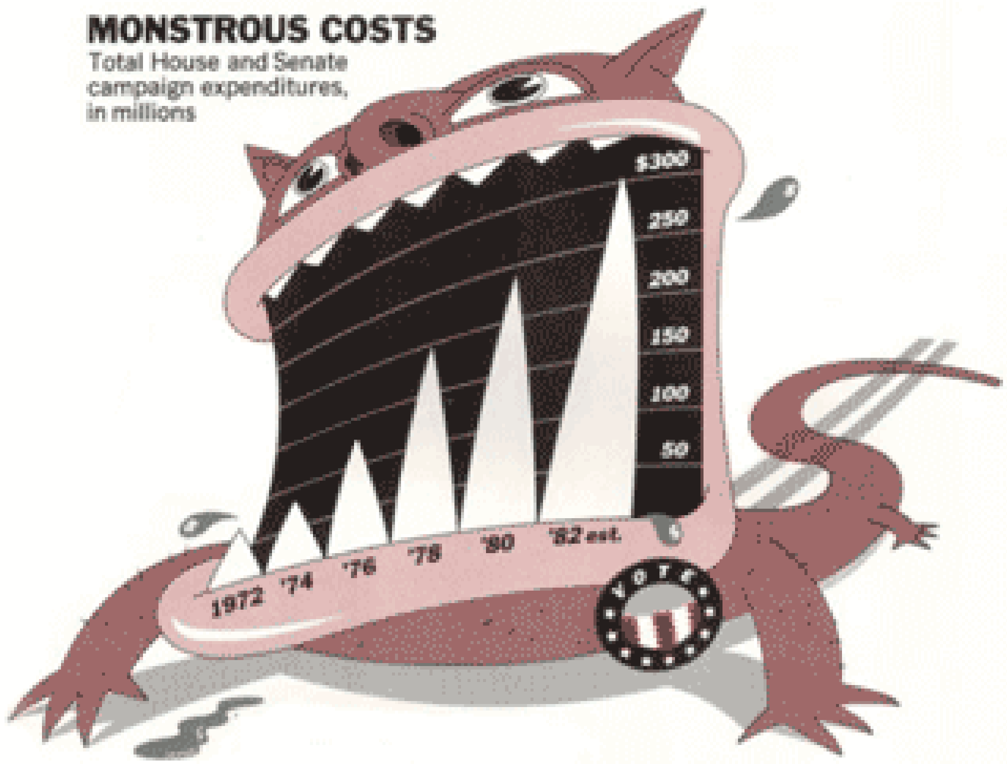
Tufte's views on

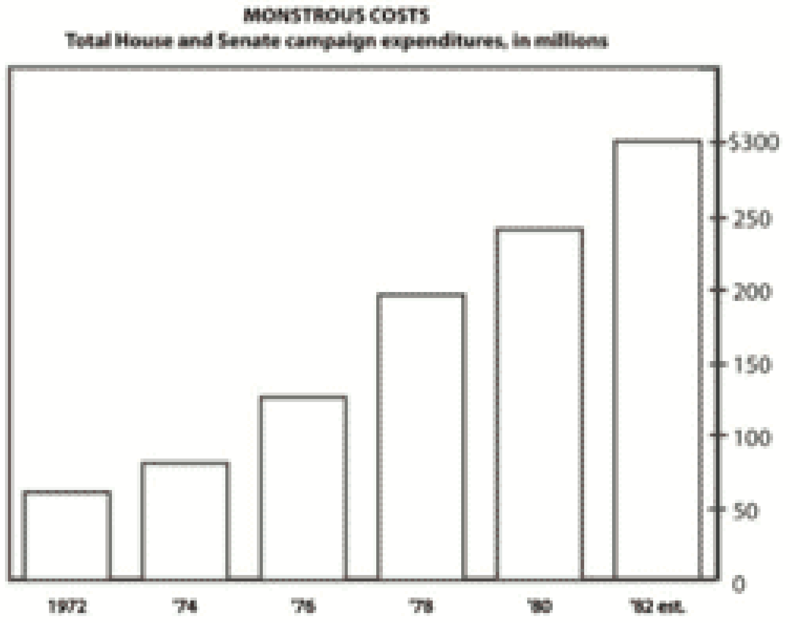
- visual embellishments → "chart junk"
- abuse of physically-motivated distortions → "lie factor"

Avoid Misleading Embellishments = Chart Junk

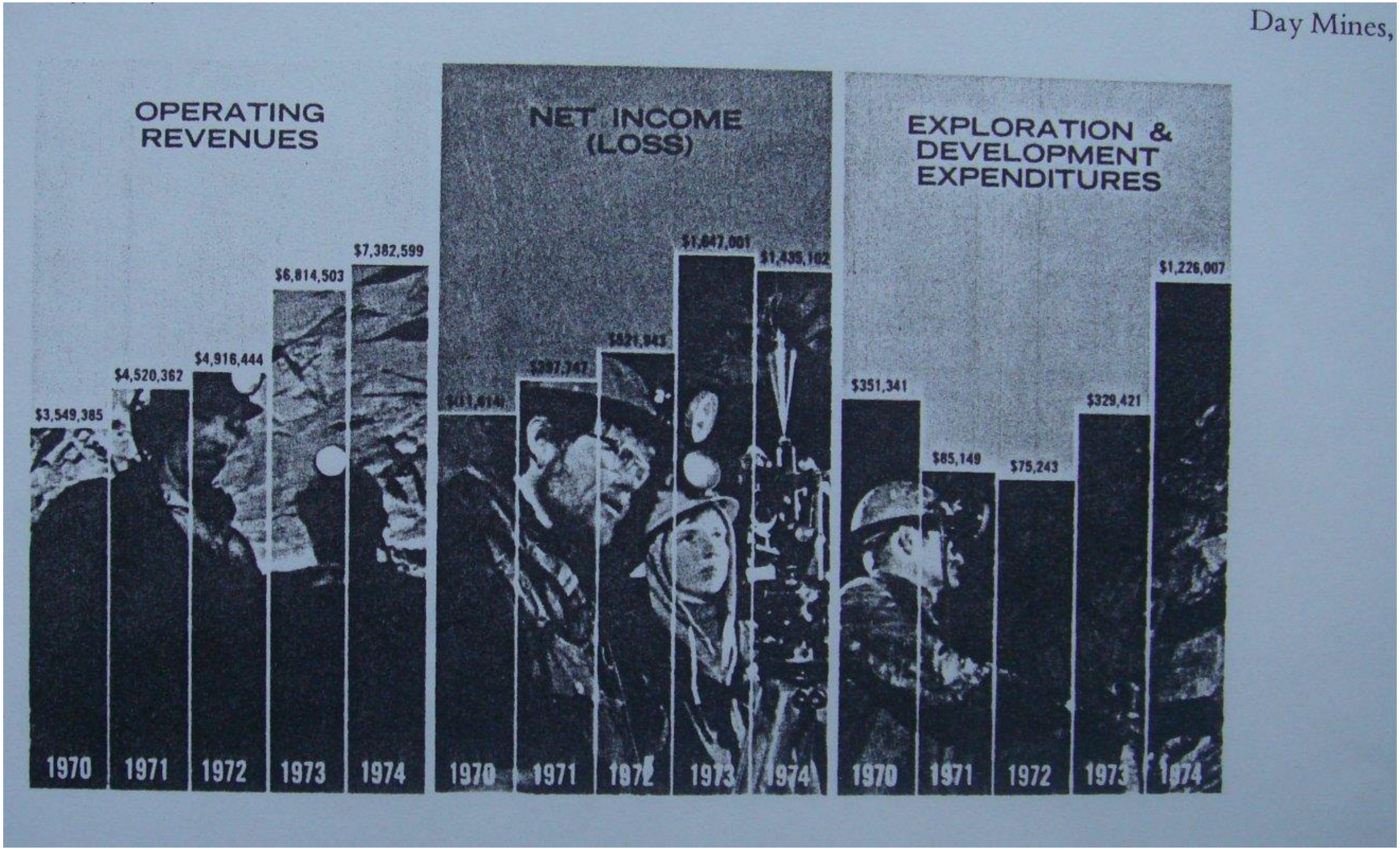


Nigel Holmes' Famous Chart

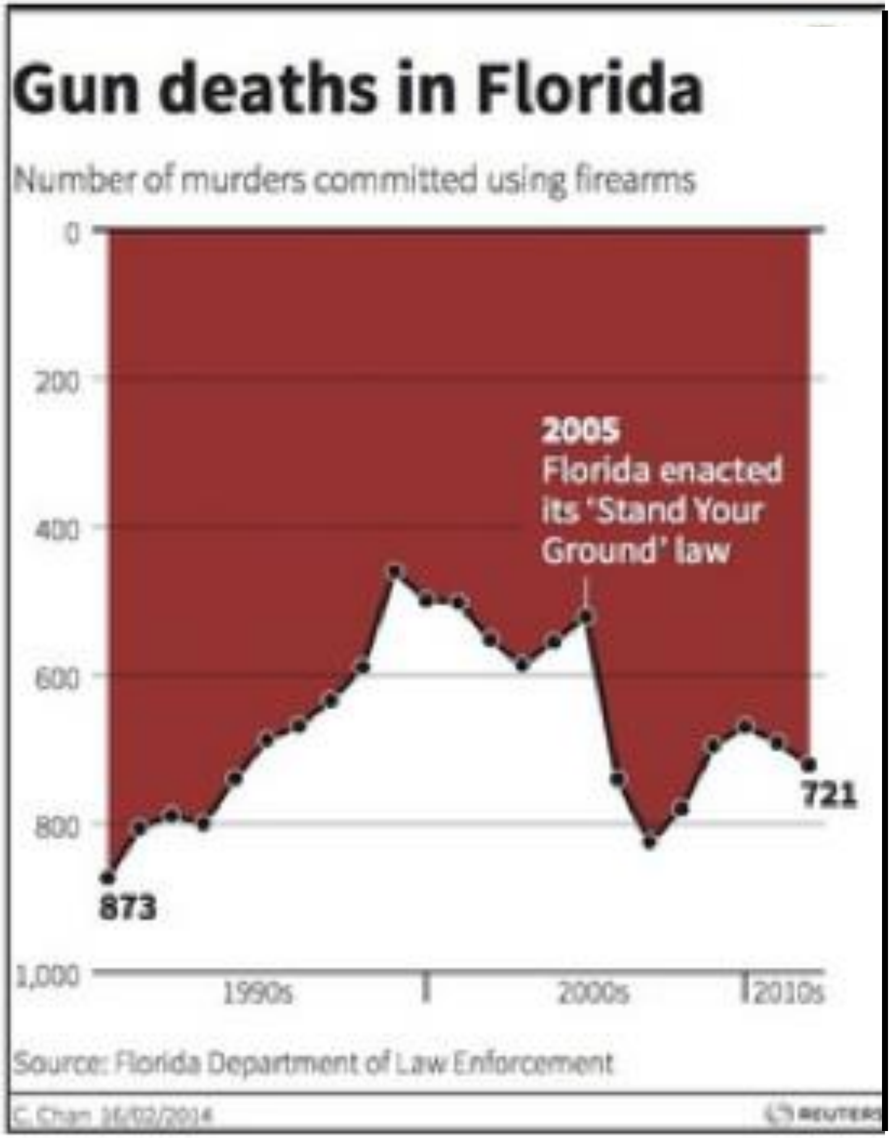




Avoid Misleading Scaling

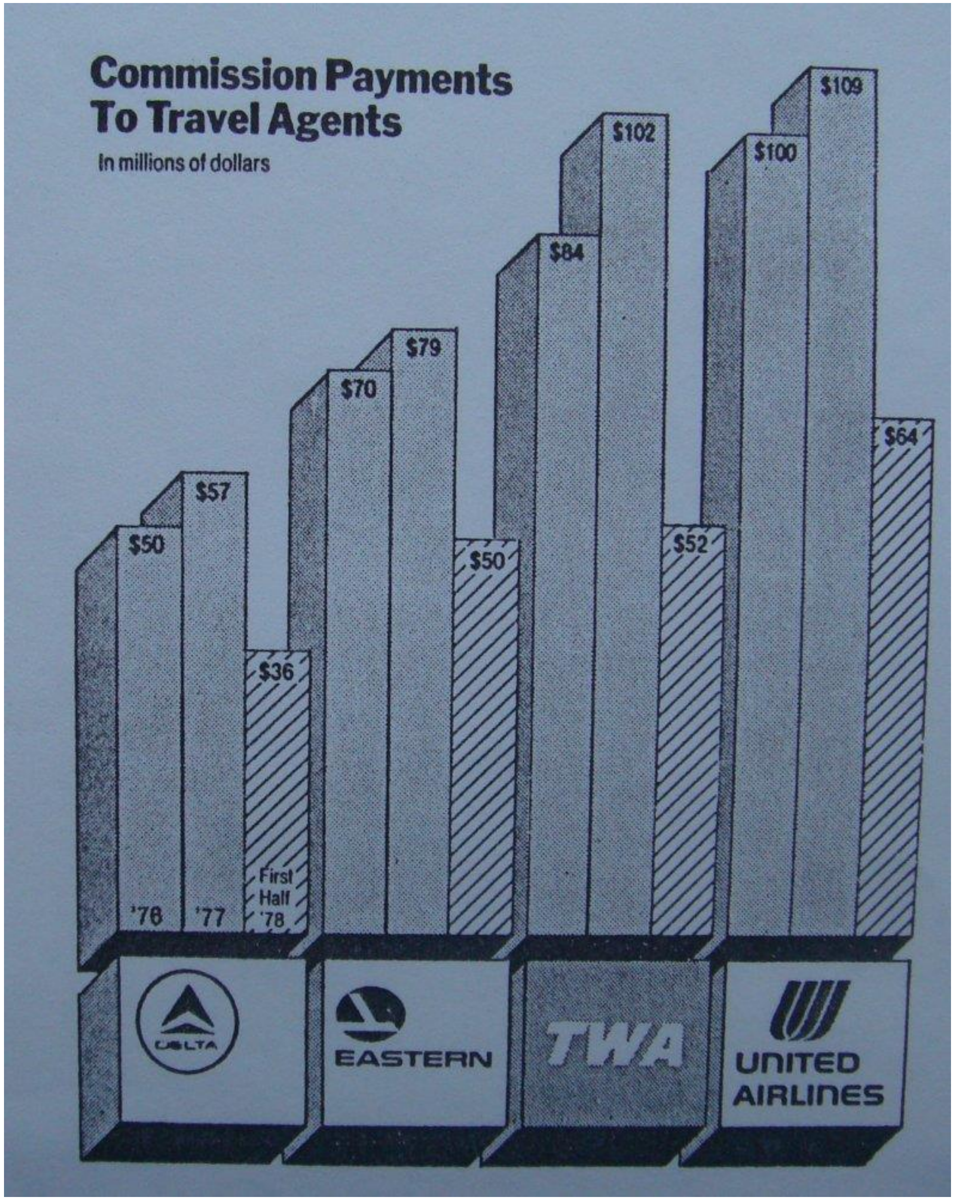


Manipulation of Axis Orientation

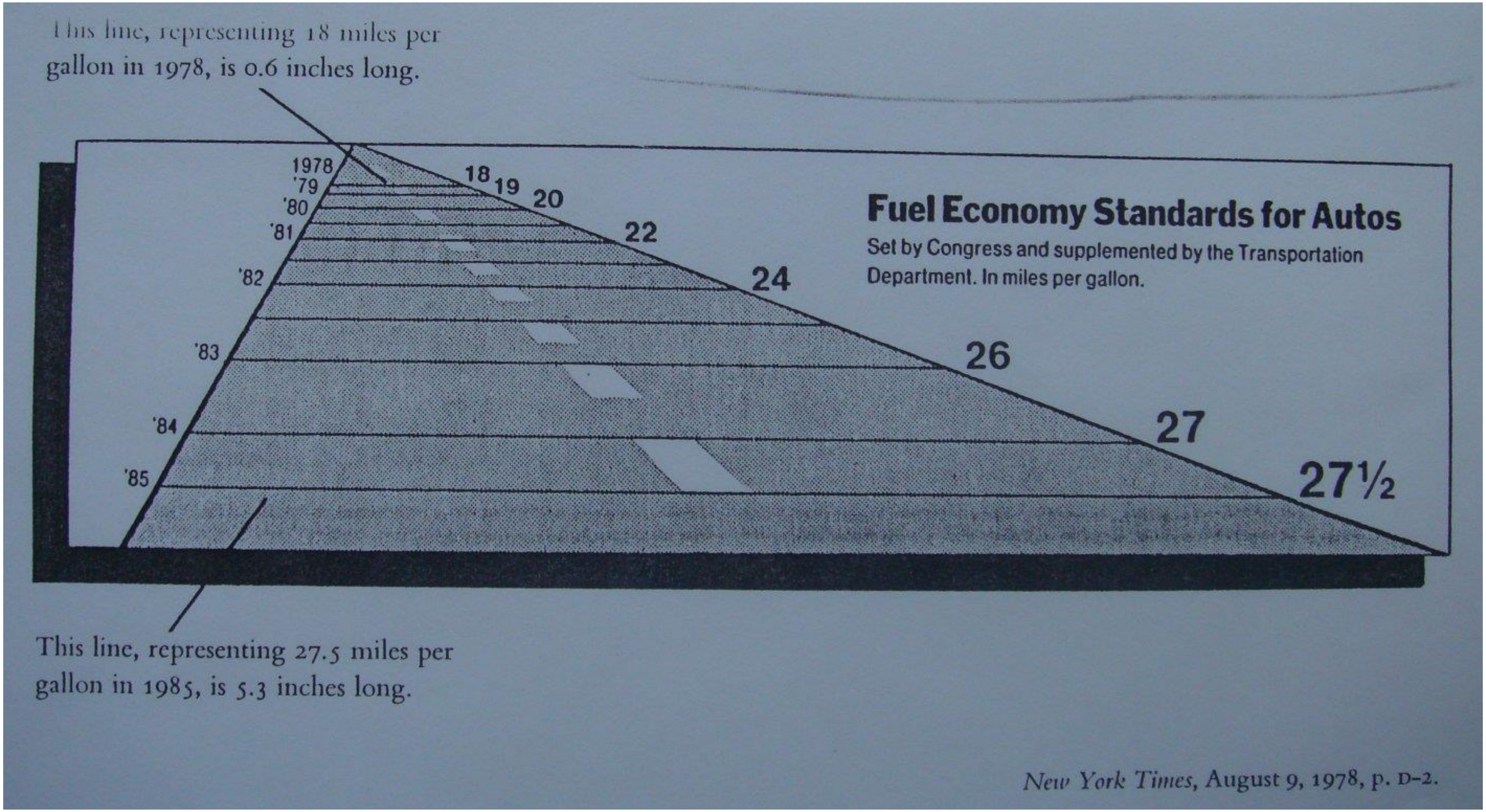


from Panday at al. (CHI 2015)

Avoid Misleading Scaling



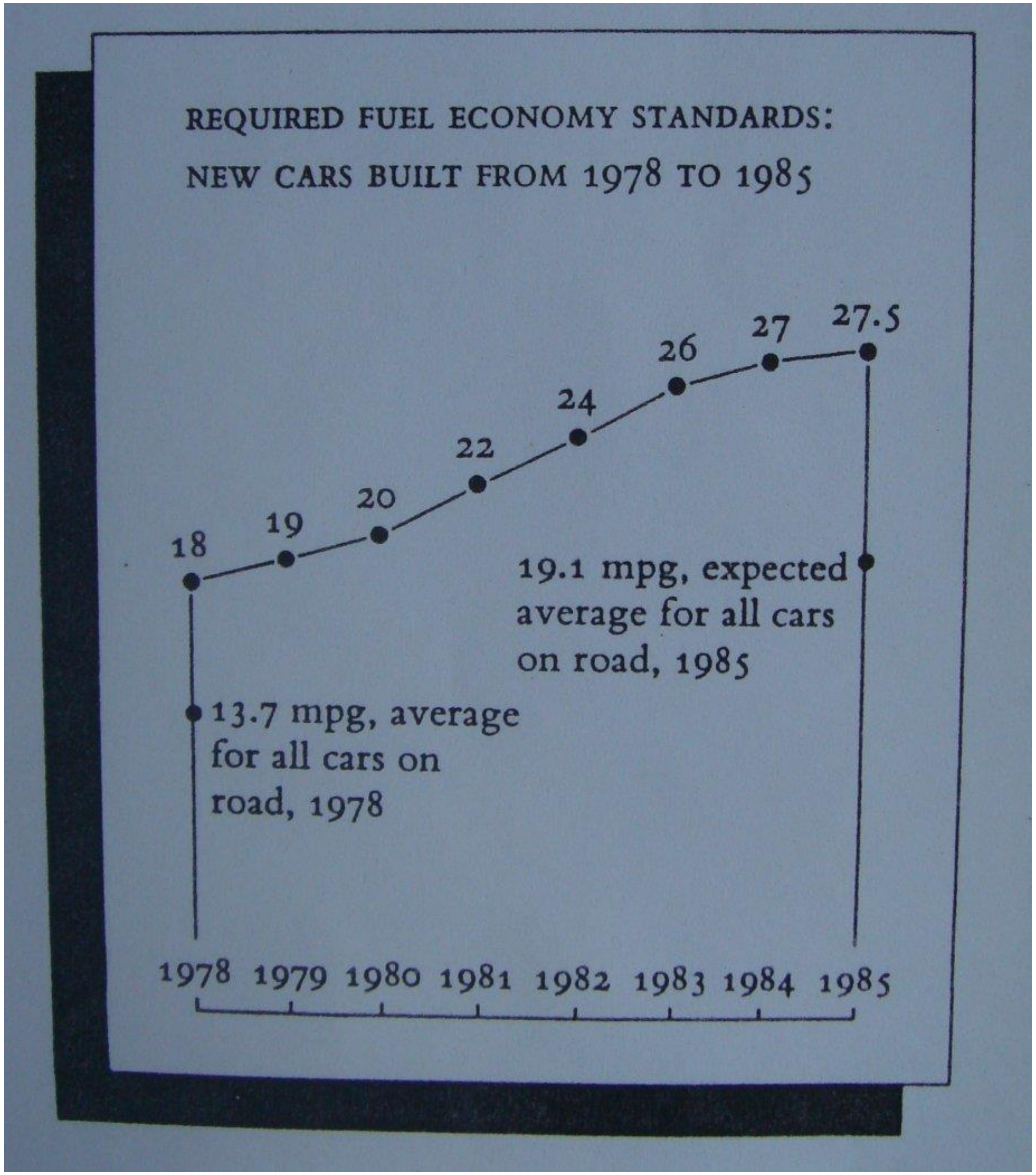
Avoid Misleading Use of Graphics Effects



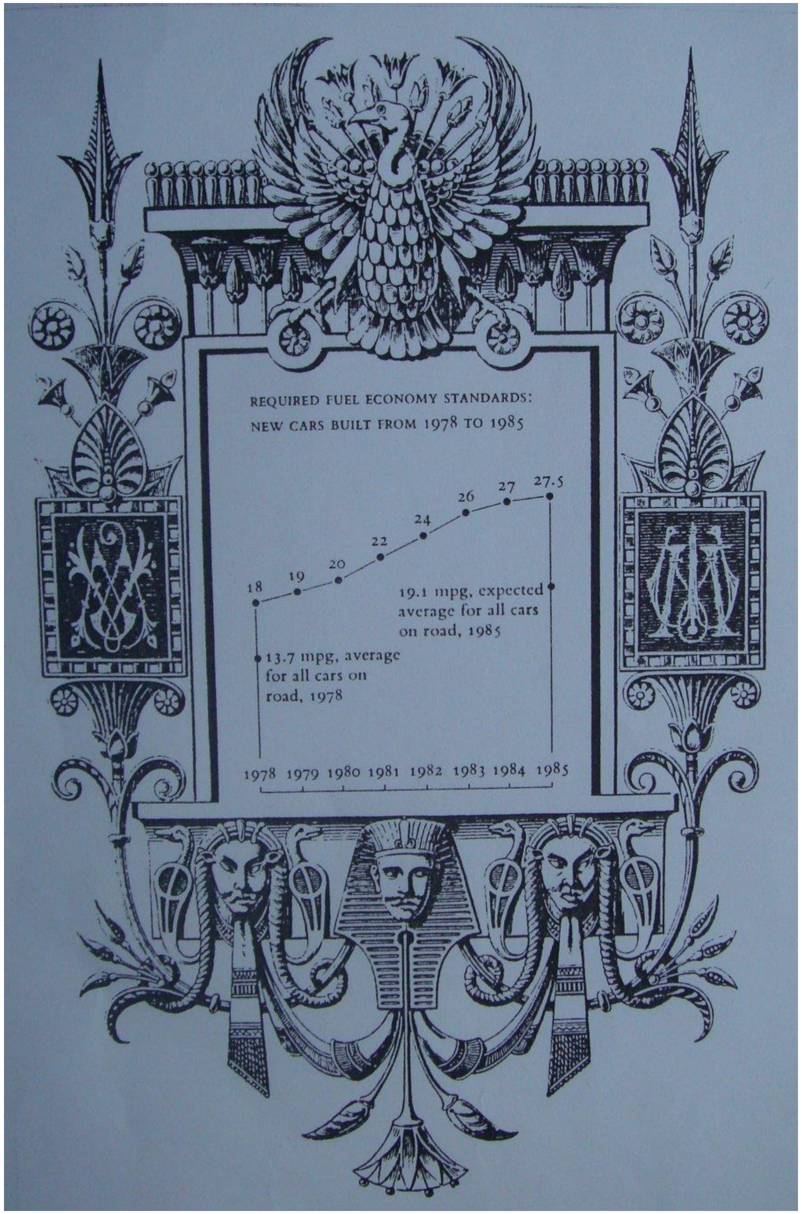
real effect: (27.5-18) / 18 = 53 %

graphical effect: (5.3”-0.6”) / 18 = 783 % → lie factor: 783/53 = 14.8

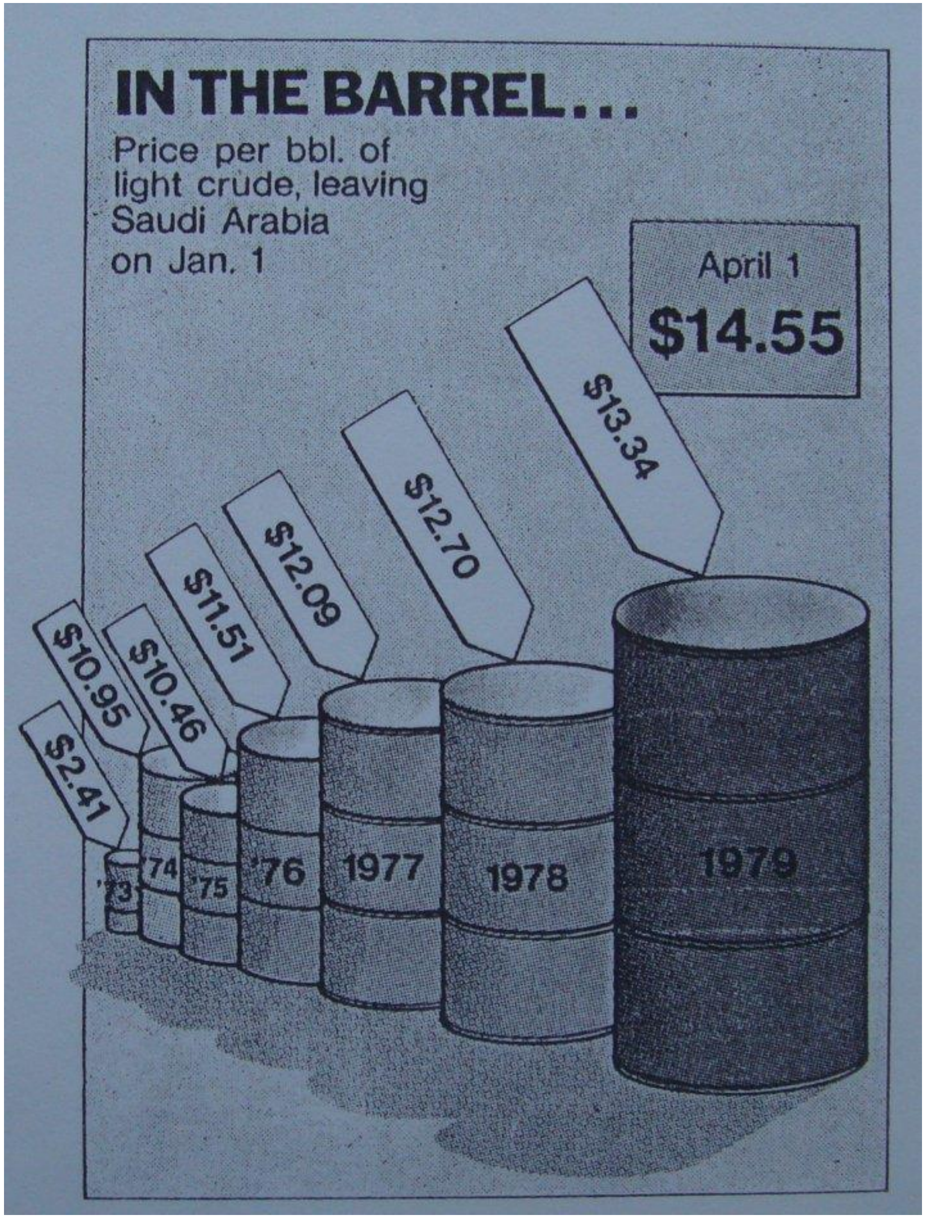
Tell the Truth About the Data



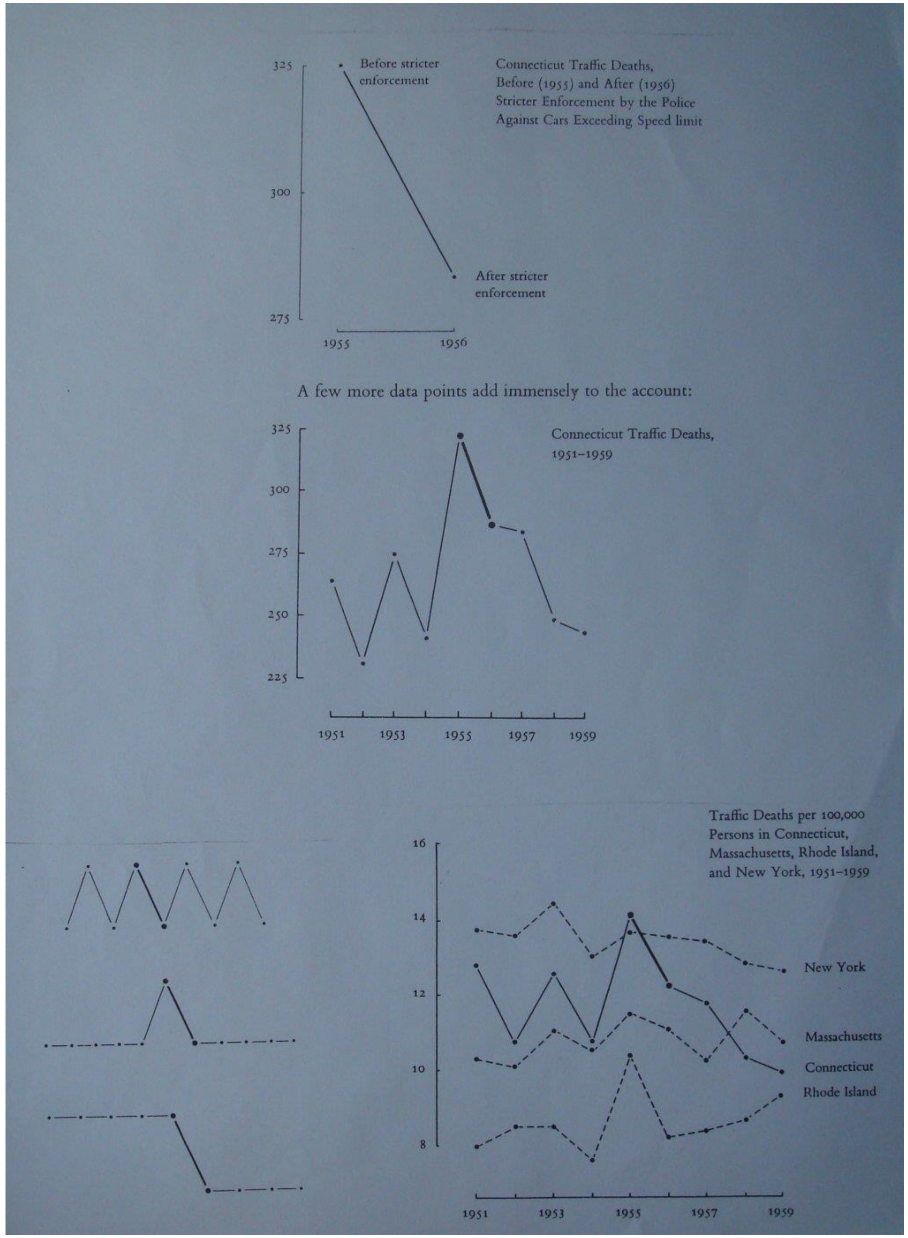
If You Must Embellish…



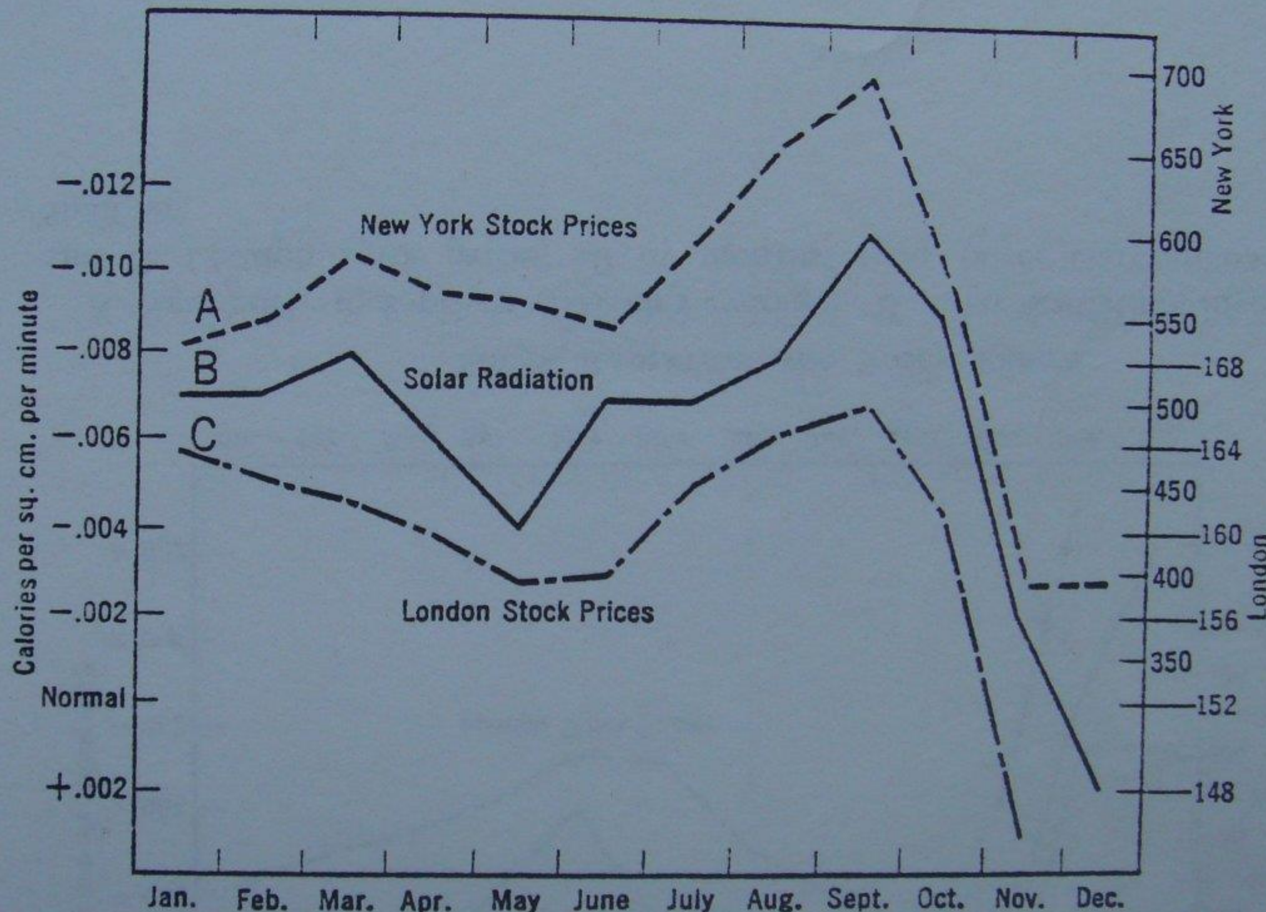
Avoid Suggestive Distortions



Show the Data in Their Proper Context



Avoid Display of Out-of-Context Data



SOLAR RADIATION AND STOCK PRICES

A. New York stock prices (Barron's average). B. Solar Radiation, inverted, and C. London stock prices, all by months, 1929 (after Garcia-Mata and Shaffner).

Graphical Excellence

- Is cosmetic decoration really needed to make data more interesting (may only distract):

  - example: diamond graph (adds a useless 3rd dimension)

- Misleading graphical representation

  - example: missing baseline in Day Mines, Inc. annual report (pg. 54)
  - example: non-uniform data spans in Commision Payments graph (pg. 54)
  - example: non-uniform scaling of icons in Pittsburgh Civic Commission report (pg. 55)

- The Lie Factor = $\frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$ (should be within [0.95, 1.05])

  - example: graph on fuel economy standards for autos (lie factor = 14.8) (pg. 57)

- Visualizing data bearing some dimension by means of objects of higher dimensions:

  - example: the *growing barrel* (lie factor: 9.4 (2D), 59.4 (3D)) (pg. 62)
  - example: the *growing oil pump* (lie factor: 9.5) (pg. 62)
  - example: the *shrinking dollar bill* (lie factor: ~6) (pg. 70)
  - example: the *incredibly shrinking family doctor* (pg. 69)

→ the number of information carrying dimensions should not exceed the data dimensions

Graphical Integrity

- Quoting data out of context and/or too sparse (recall: graphics allows high data density)

  - example: Connecticut traffic deaths (pg. 74/75)

Principles that ensure graphical integrity:

- 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, the shrinking doctor)
- Graphics must not quote data out of context (see the Connecticut traffic deaths)
- Convincing graphics must demonstrate cause and effect (see Challenger disaster)

But Wait… There is More

Do these bare graphs engage a human audience?

- are they memorable?

A recent (research) trend

- will embellishment help memorability, engagement?
- do we need what Tufte char junk

Video

MemViz: A Tool for Creating Memorable Visualizations

Darius Coelho, Sungsoo Ha, Shenghui Cheng, Salman Mahmood, Jisung Kim, and Klaus Mueller

Visual Analytics and Imaging Lab, Computer Science Department, Stony Brook University and SUNY Korea

Stony Brook University

SUNY Korea
Stony Brook University
Graduate Degree Programs
The State University of New York

Memorability Experiment by Borkin et al.

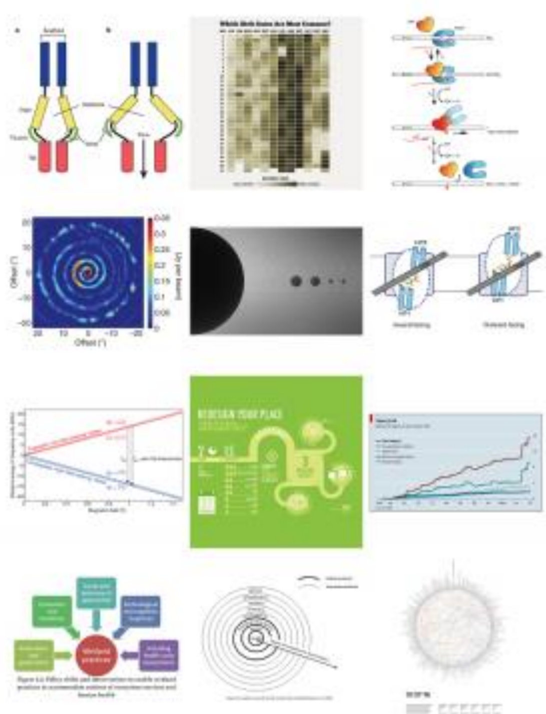
Experiment set up as a game on Amazon Mechanical Turk

- workers were presented with a sequence of images (about 120)
- presented for 1 second, with a 1.4 second gap between consecutive images
- workers had to press a key if they saw an image for the second time in the sequence (spacing 1-7 images with “filler” images in between)

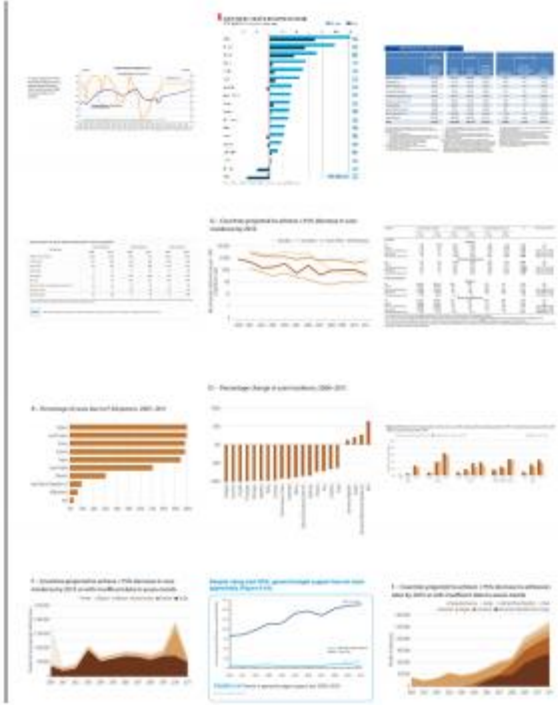
Borkin et al. IEEE TVCG 2014

Memorability Experiment by Borkin et al.

most memorable



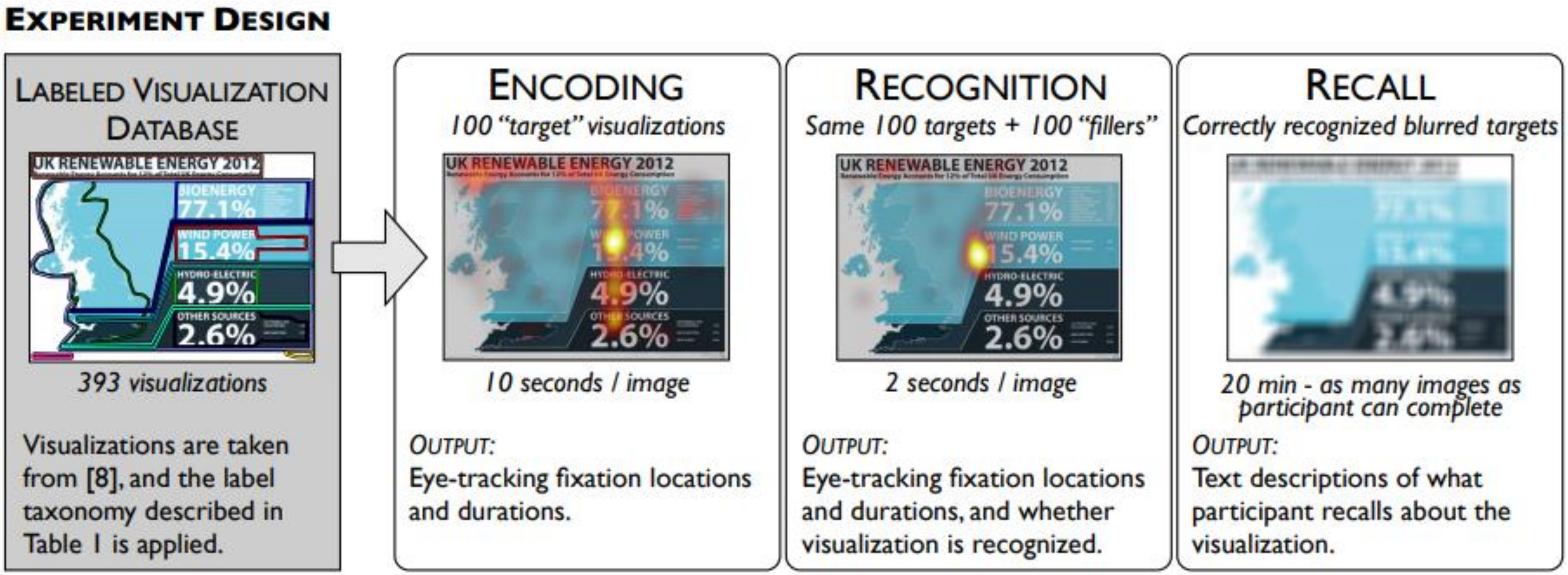
most memorable after removing human recognizable cartoons



least memorable

Borkin et al. IEEE TVCG 2014

What Do People Remember?



Borkin et al. IEEE TVCG 2016

Eye Tracking Experiments

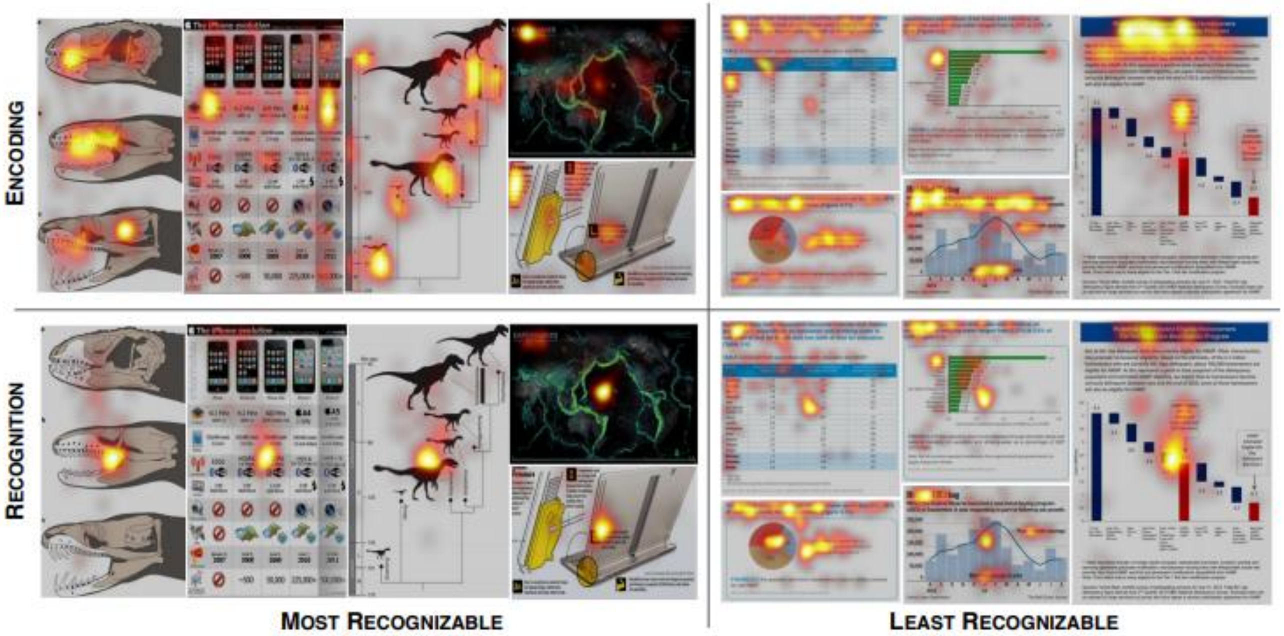


Fig. 7. Examples of the most and least recognizable visualizations from [8]. **TOP**: Eye-tracking fixation heat maps (i.e., average of all participants' fixation locations) from the *encoding* phase of the experiment in which each visualization was presented for 10 seconds. The fixation patterns demonstrate visual exploration of the visualization. **BOTTOM:** Eye-tracking fixation heat maps from the *recognition* phase of the experiment in which each visualization was presented for 2 seconds or until response. The most recognizable visualizations all have a single focus in the center indicating quick recognition of the visualization, whereas the least recognizable visualizations have fixation patterns similar to the encoding fixations indicative of visual exploration (e.g., title, text, etc.) for recognition.

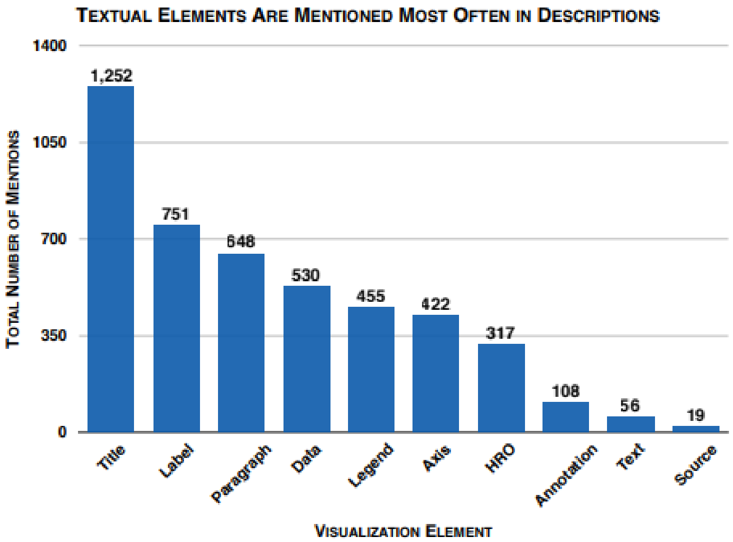
Borkin et al. IEEE TVCG 2016

Practical Rules for Visualization Design

Takeaways:

- 393 visualizations and eye movements of 33 participants and 1,000s of participant-generated text descriptions of the visualizations
- titles and supporting text should convey the message of a visualization
- if used appropriately, pictograms do not interfere with understanding and can improve recognition
- redundancy helps effectively communicate the message
- visualizations that are memorable “at-a-glance” are also capable of effectively conveying the message of the visualization

→ thus, a memorable visualization is often also an effective one



Important for Memorability

Important are:

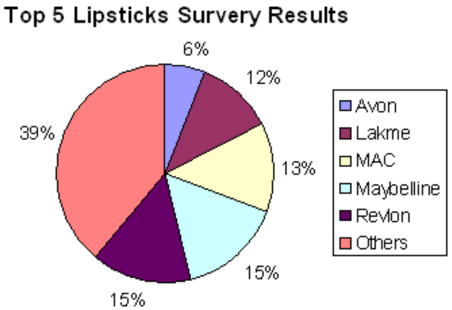
- attributes like color
- inclusion of a human recognizable object

However, link to human engagement not explicitly established

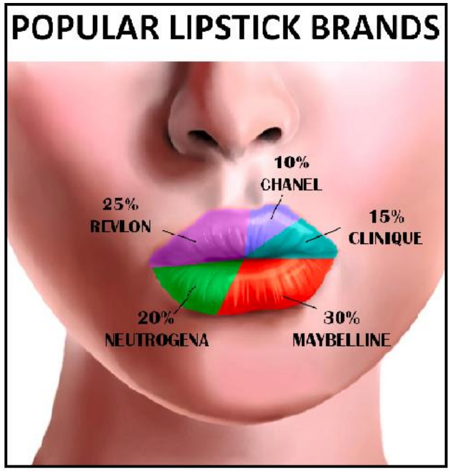
- “just” memorability

Our own studies show that embellishments can get humans interested in studying an image

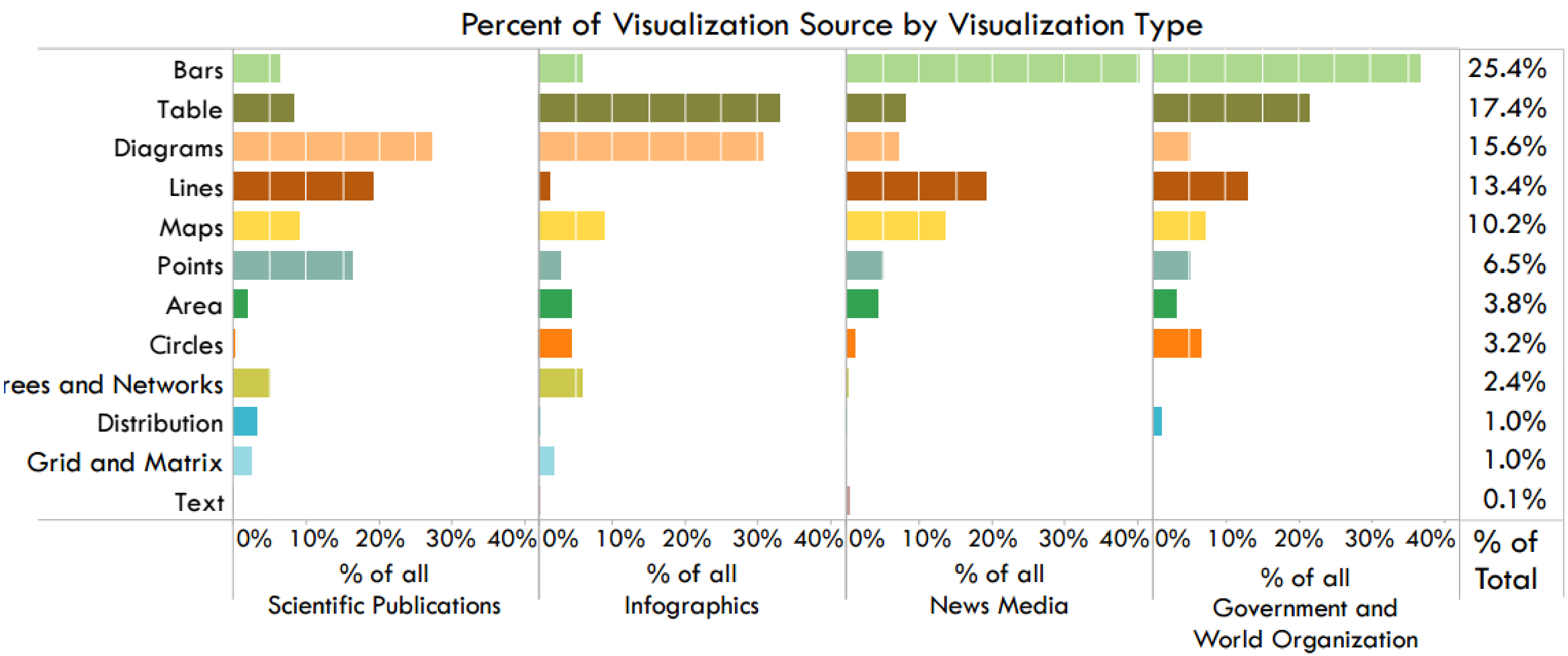
- but prefer conventional charts for problem solving



vs.



Visualizations Sources and Origins



Borkin et al. IEEE TVCG 2014

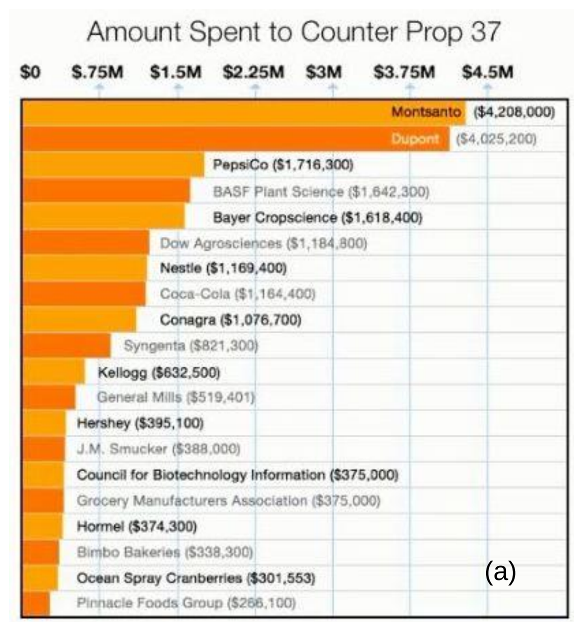
Infographic

Graphic visual representations of information, data or knowledge intended to present information quickly and clearly

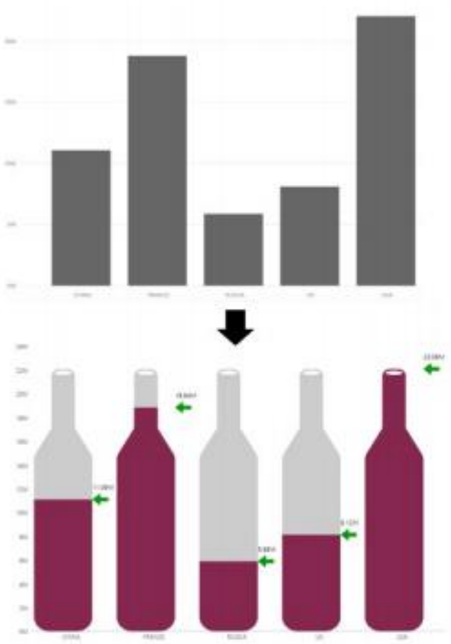
Evolved in recent years to be for mass communication

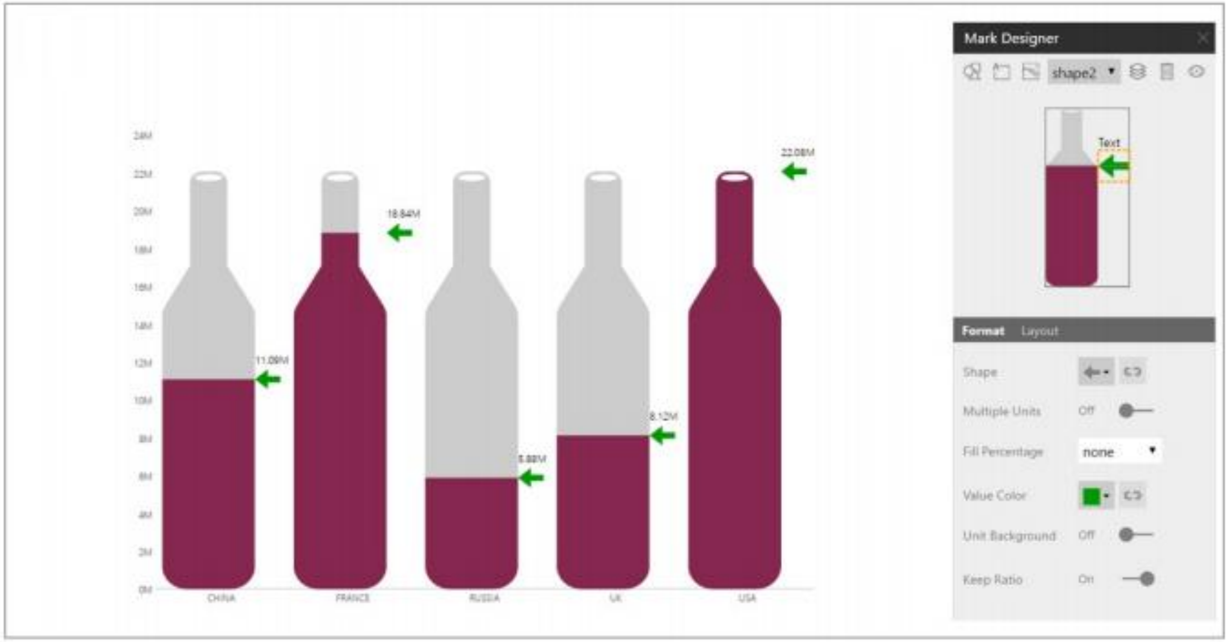
- designed with fewer assumptions about the readers knowledge base than other types of visualizations
- but can be misleading and express the opinion of the author

vs.



Using Icons as Bar Graphs

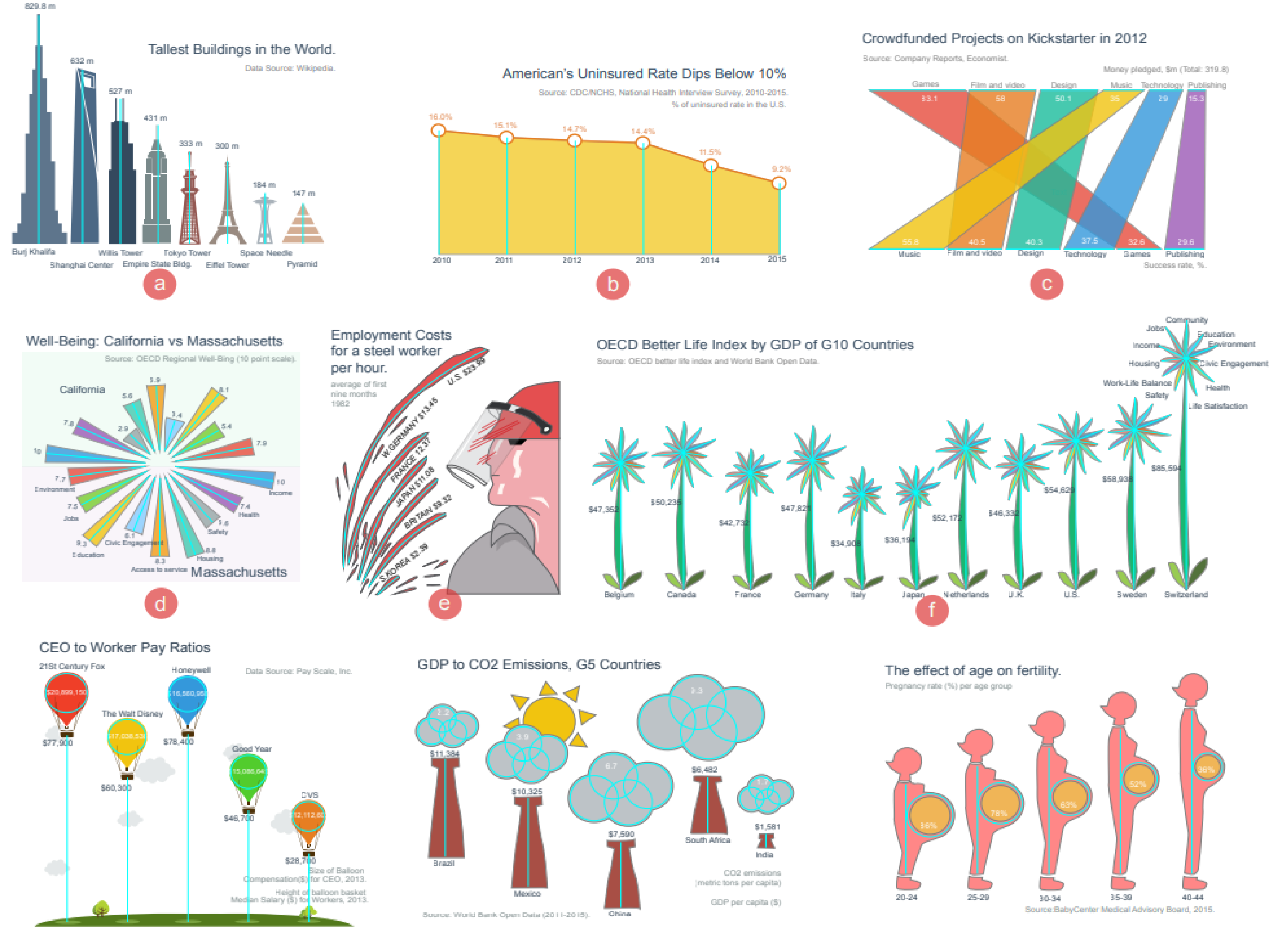




Wang et al. CHI 18

Data-Driven Design Guides

Kim et al. TVCG 17



Kim et al. TVCG 17