

CSE 664  
VISUALIZATION & VISUAL ANALYTICS

THE VIEWS OF EDWARD TUFTE  
(AND SOME OTHERS)

**KLAUS MUELLER**

COMPUTER SCIENCE DEPARTMENT  
STONY BROOK UNIVERSITY

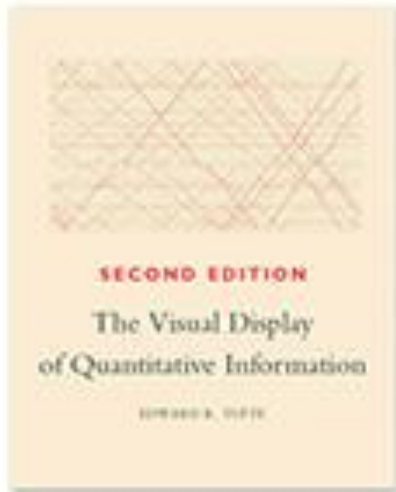
Lecture	Topic	Projects
1	Intro and logistics	
2	Basic visualizations and tasks, data types, examples, ethical considerations	
3	Data preparation (cleaning, imputation, data set integration)	
4	AI-assisted coding for VIS applications (design, debugging, refactoring)	Project #1 out
5	Big data and data reduction (distance/sim metrics, intro to clustering)	
6	High-D data: concept, subspaces, dimension reduction, PCA	
7	Cluster analysis: hierarchical, density, model, embedding, temporal	
8	Perception and cognition (human visual system, color, contrast)	Project #2(a) out
9	Visual design and aesthetics	
10	Visualization of multivariate and high-D data: linear methods, projections	
11	Vis. of multivariate and high-D data: non-linear methods, embeddings	
12	Visualization and AI: mutual support and capabilities (VIS4AI, AI4VIS)	Project #2(b) out
13	Principles of interaction: drive what is visualized, analyzed & how (HCI4VIS)	
14	Visual analytics, human-centered AI, mixed-initiative & collaborative VA	
15	Midterm #1 (tentative date)	
16	VA system design and evaluation, the nested model	
17	Midterm #1 discussion (tentative date)	Final proj. proposal call out
18	Visualization of hierarchical data	
19	Visualization of maps and data with geo-reference	
20	Vis. of time-varying, time-series, streaming data, progressive visualization	Final project proposal due
21	Applications in Visualization Research	
22	Applications in Visualization Research	
23	Ed Tufte's principles and critiques, responsible visualization, uncertainty	
24	Design of effective infographics	Final proj. prelim report due
25	Foundations scientific and medical visualization	
26	Intro to volume rendering	Bonus project out (Vol Ren)
27	Scientific visualization	
28	Midterm #2 (tentative date)	
Final	Final project demo on zoom (public)	All final proj. materials due

# Seminal Books by Edward Tufte

Standard literature for every visualization enthusiast

- written 1983, 1990, 1997, 2006

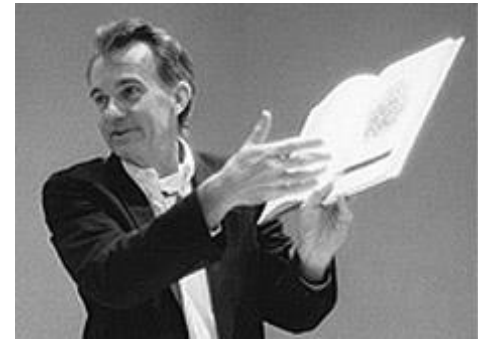
## EDWARD TUFTE TAKES HIS COURSE ON THE ROAD



# Edward Tufte

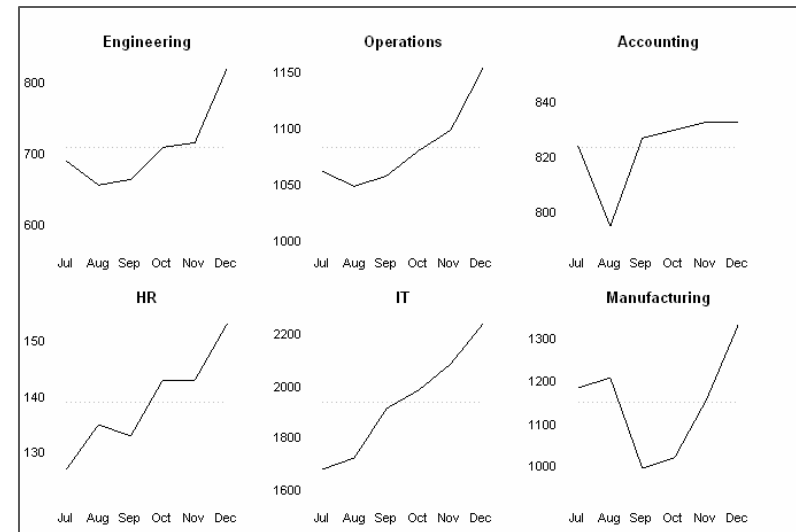
Well recognized for his writings on information design

- 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”

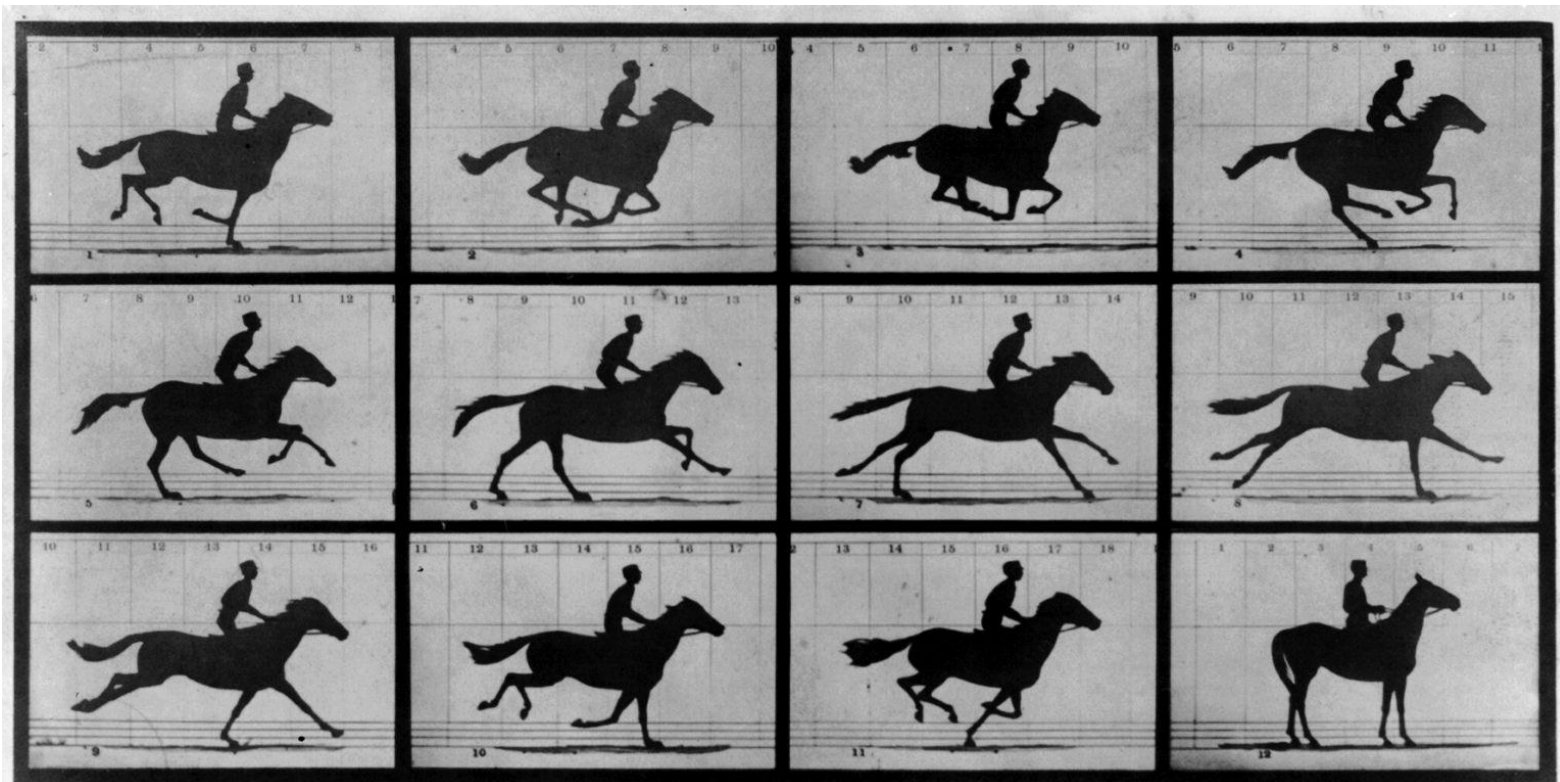
- aka trellis chart or panel chart
- similar charts of same scale + axes
- allows them to be easily compared
- use multiple views to show different partitions of a dataset



# 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



Copyright, 1878, by MUYBRIDGE.

MORSE'S Gallery, 417 Montgomery St., San Francisco.

### THE HORSE IN MOTION.

Illustrated by  
MUYBRIDGE.

AUTOMATIC ELECTRO-PHOTOGRAPH.

"SALLIE GARDNER," owned by LELAND STANFORD; running at a 1.40 gait over the Palo Alto track, 19th June, 1878.

The negatives of these photographs were made at intervals of twenty-seven inches of distance, and about the twenty-fifth part of a second of time; they illustrate consecutive positions assumed in each twenty-seven inches of progress during a single stride of the mare. The vertical lines were twenty-seven inches apart; the horizontal lines represent elevations of four inches each. The exposure of each negative was less than the two-thousandth part of a second.

# 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!

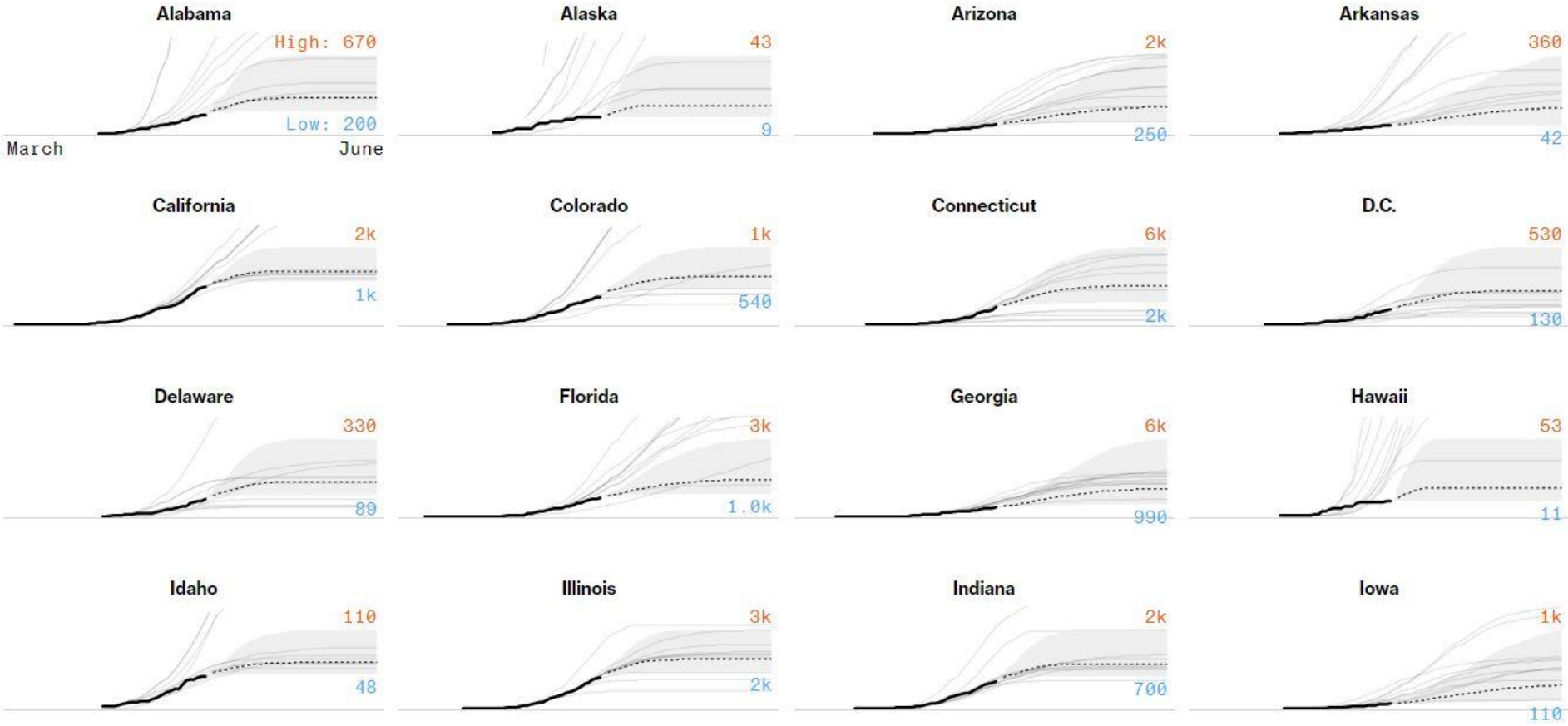


# Small Multiples

Tracking Covid-19 [Global Cases ↗](#) [Cases in the U.S. ↗](#)

## Cumulative projected deaths in each state

▬ Deaths    ⋯ Projected Deaths    ▭ Previous Estimates (Mean Projection)



[Bloomberg page](#)

# Small Multiples

Tracking Covid-19 [Global Cases ↗](#) [Cases in the U.S. ↗](#)

Maryland

2k

740

Massachusetts

9k

2k

Michigan

5k

3k

Minnesota

900

180

Mississippi

1k

210

Missouri

920

260

Montana

41

13

Nebraska

1k

63

Nevada

470

180

New Hampshire

160

42

New Jersey

12k

5k

New Mexico

300

58

New York

32k

20k

North Carolina

610

250

North Dakota

1k

31

Ohio

1k

590

Oklahoma

770

Oregon

240

Pennsylvania

5k













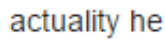
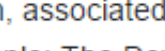
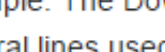
Puerto Rico

110




## 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 13/16	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 15/16	45 3/4	-1.65%	-102	6,053 +29.79% 1,389
LU	111 5/8	111 11/16	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 15/16	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
RNVK	67	67 1/4	67	+ 2 3/4	↓		1,233,900	69	64 15/16	+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 13/16	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.<sup>[15]</sup> 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 .<sup>[16][17]</sup> 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.<sup>[18]</sup>

# Tufte on Graphical Excellence

According to Tufte (pg. 51):

- Graphical excellence is the well-designed presentation of interesting data
  - a matter of **substance, 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

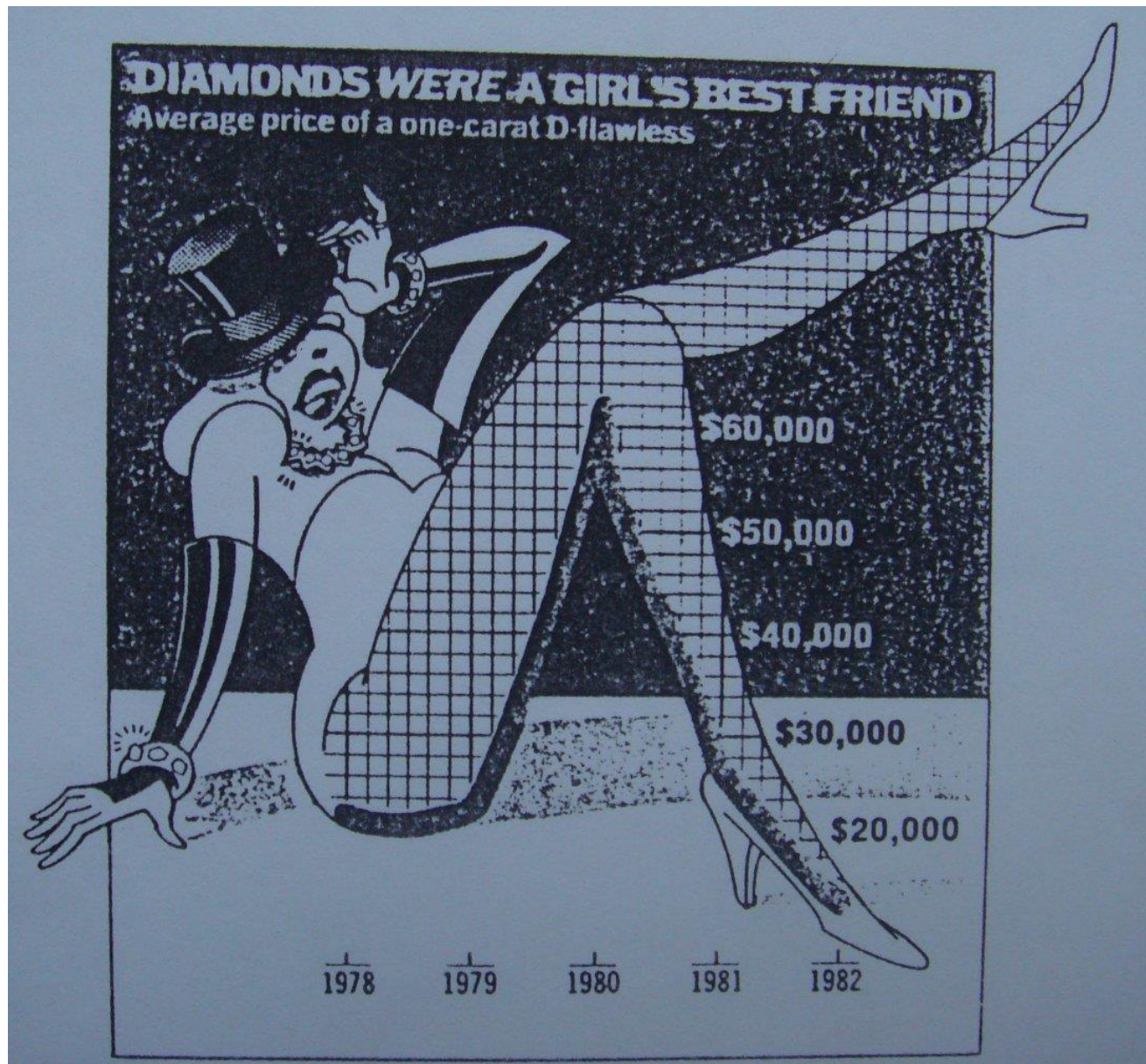
(Nevertheless, visualizations should be visually pleasing and may very well have an artistic touch)

## Next Slides...

### Tufte's views on

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

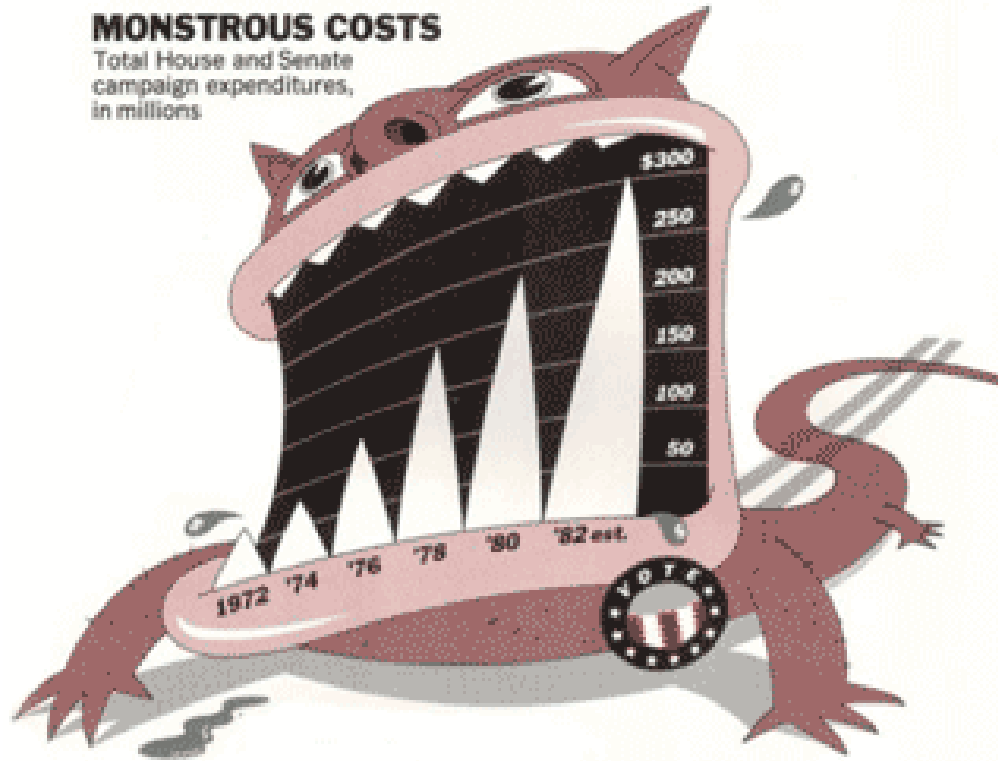
# Avoid Misleading Embellishments = Chart Junk



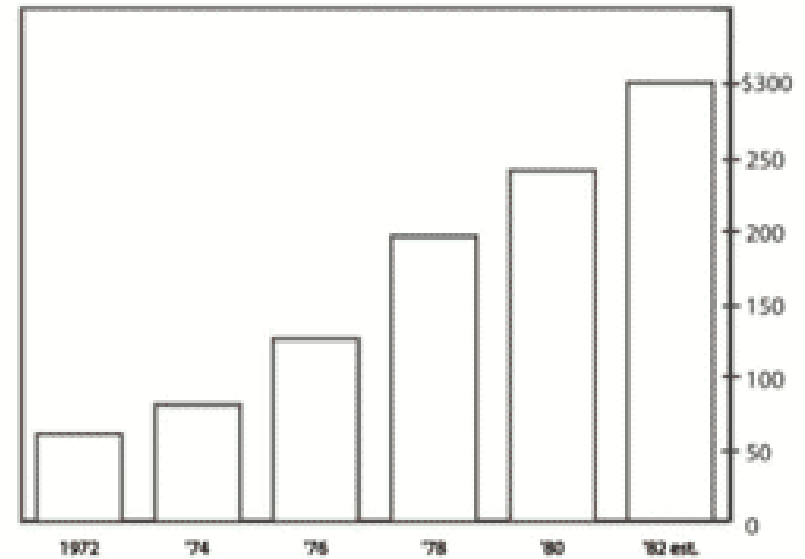
# Nigel Holmes' Famous Chart

## MONSTROUS COSTS

Total House and Senate campaign expenditures, in millions

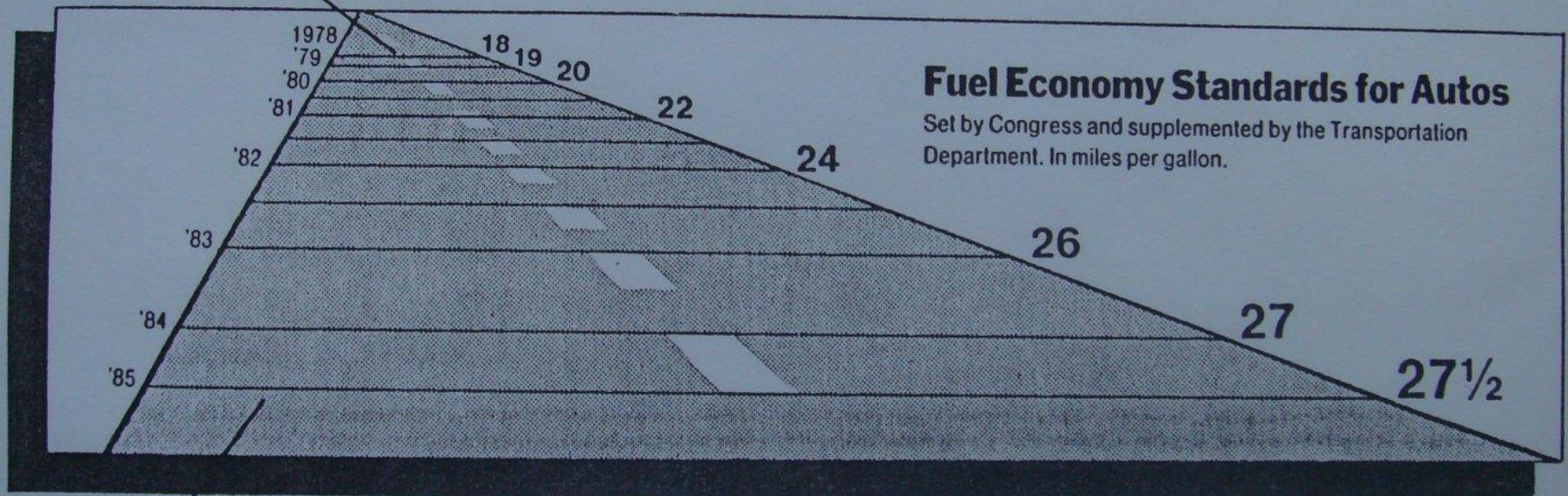


MONSTROUS COSTS  
Total House and Senate campaign expenditures, in millions



# Avoid Misleading Use of Graphics Effects

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



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

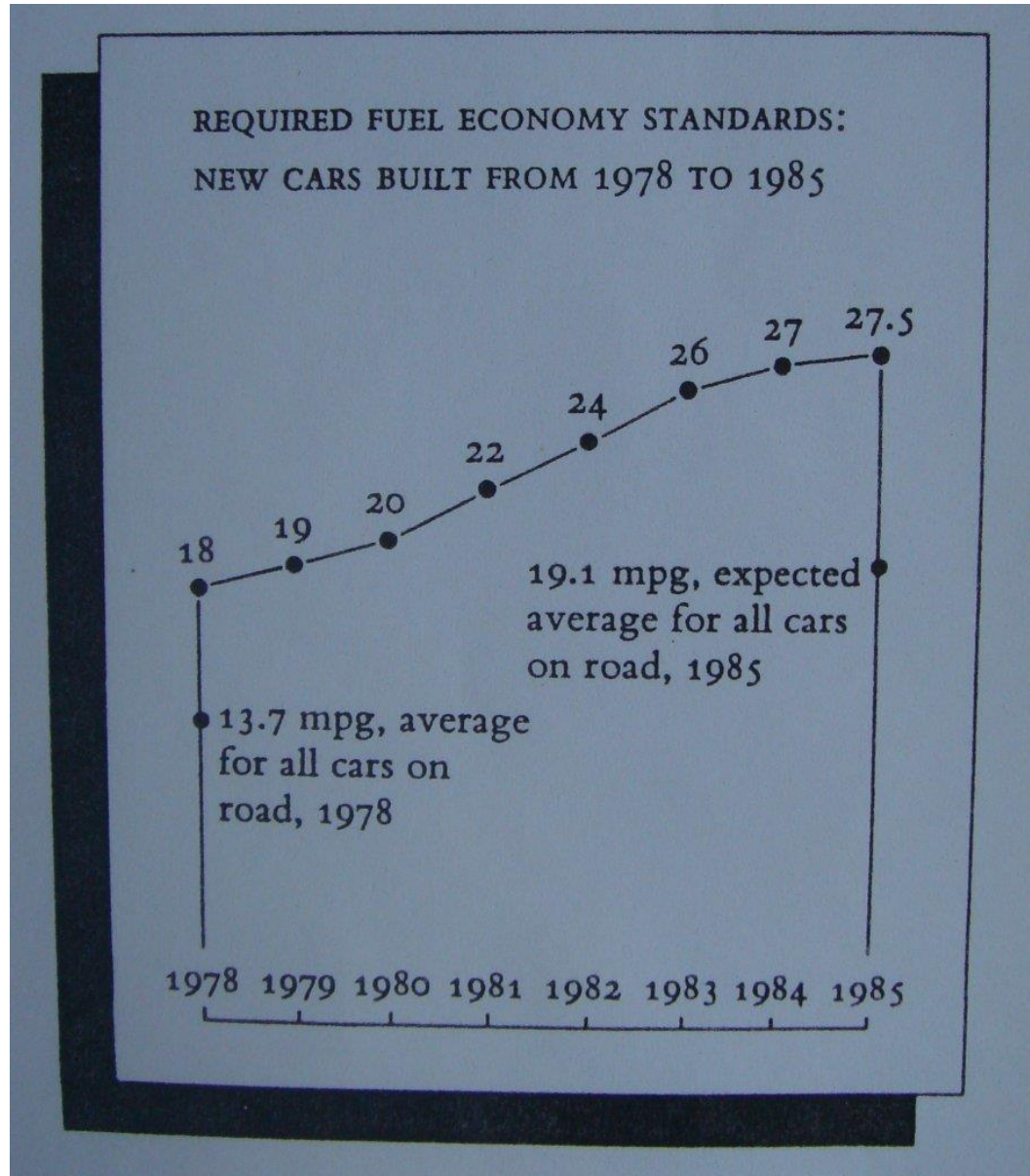
*New York Times*, August 9, 1978, p. D-2.

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

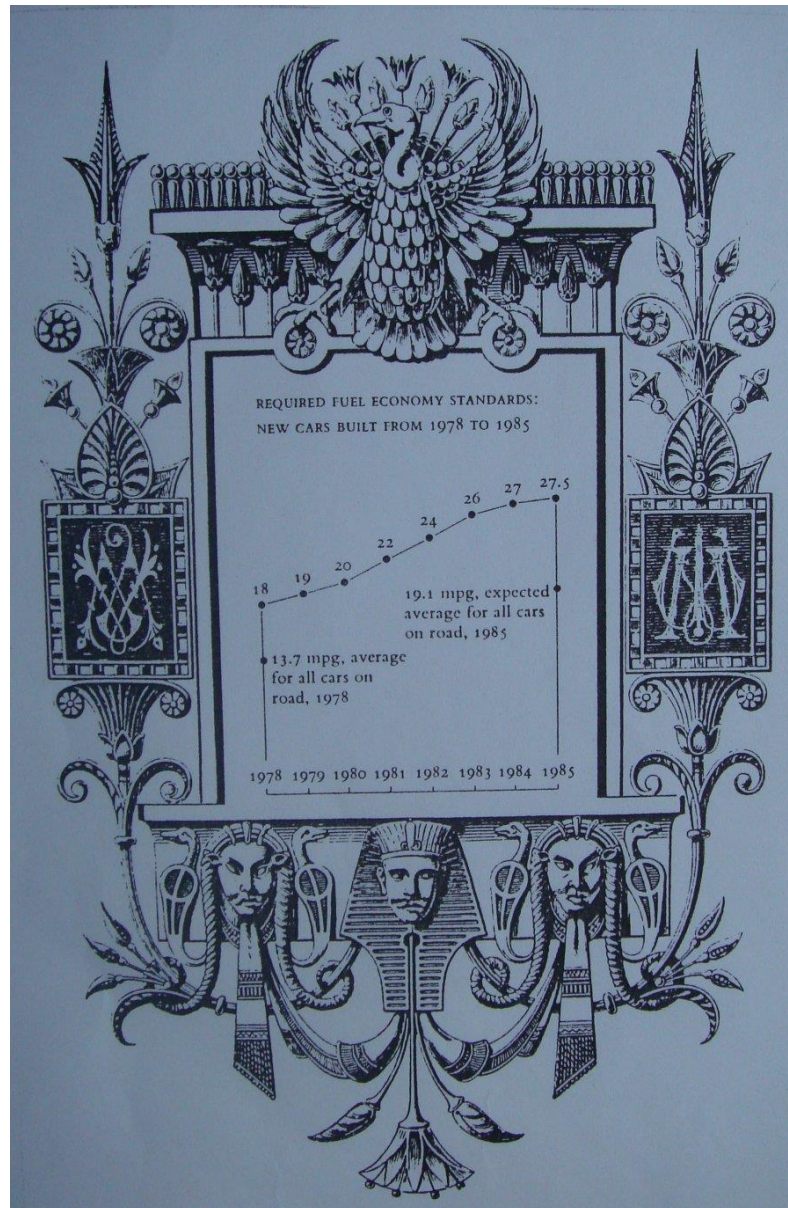
graphical effect:  $(5.3'' - 0.6'') / 0.6'' = 783\%$

→ lie factor:  $783 / 53 = 14.8$

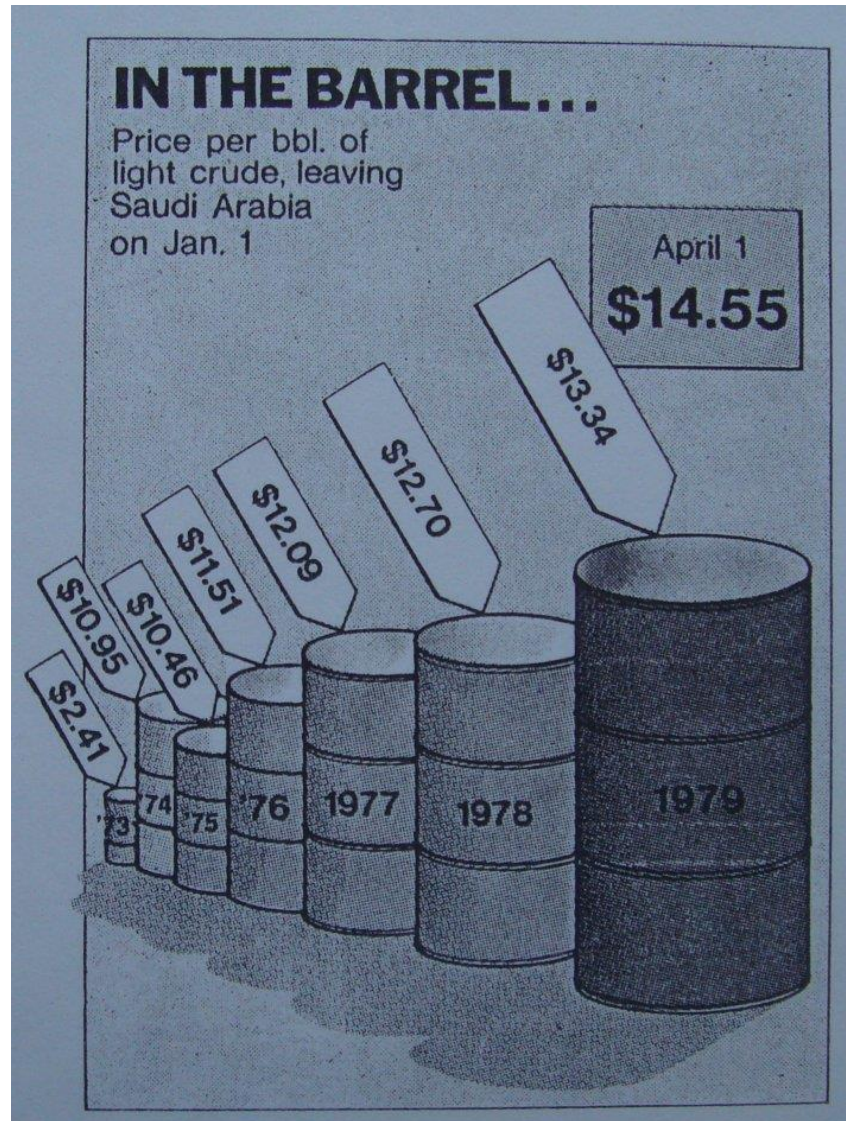
# Tell the Truth About the Data



# If You Must Embellish...

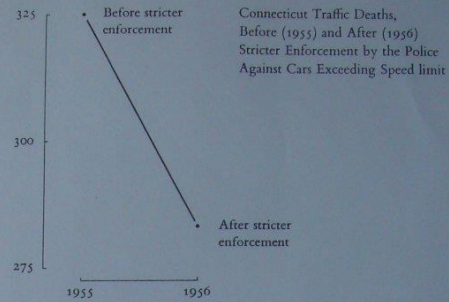


# Avoid Suggestive Distortions

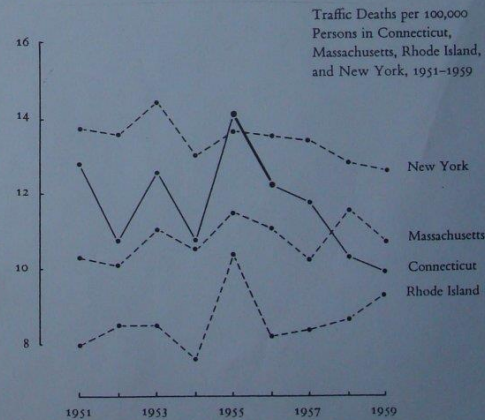
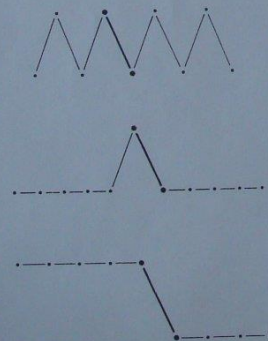
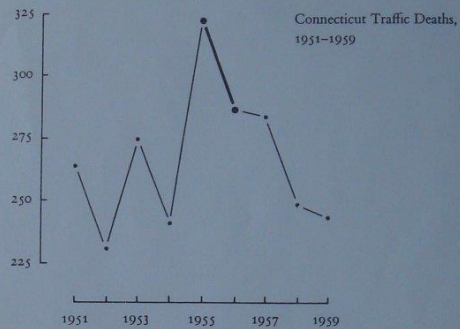


Lie factor: 9.4 (2D), 59.4 (3D)

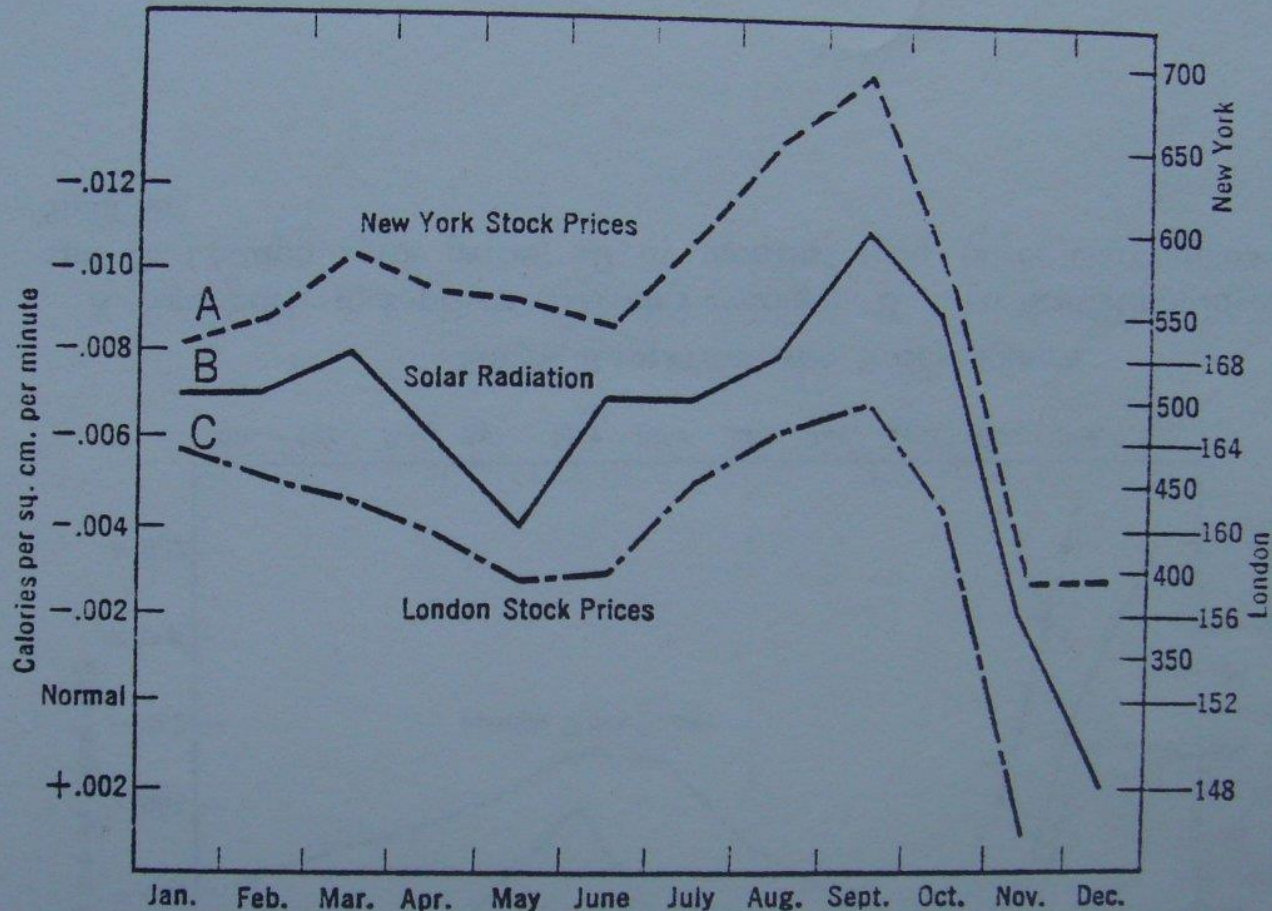
# Show the Data in Their Proper Context



A few more data points add immensely to the account:



# 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 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

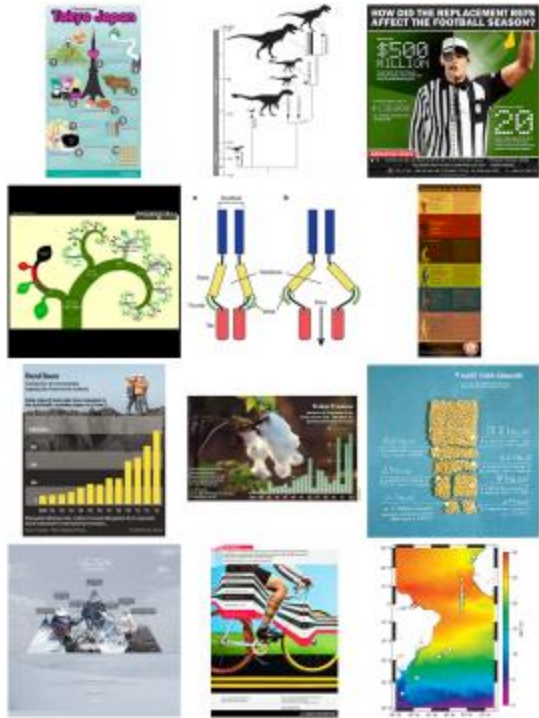
- will embellishment help memorability, engagement?
- do we need what Tufte calls “chart junk”

# Memorability Experiment by Borkin et al.

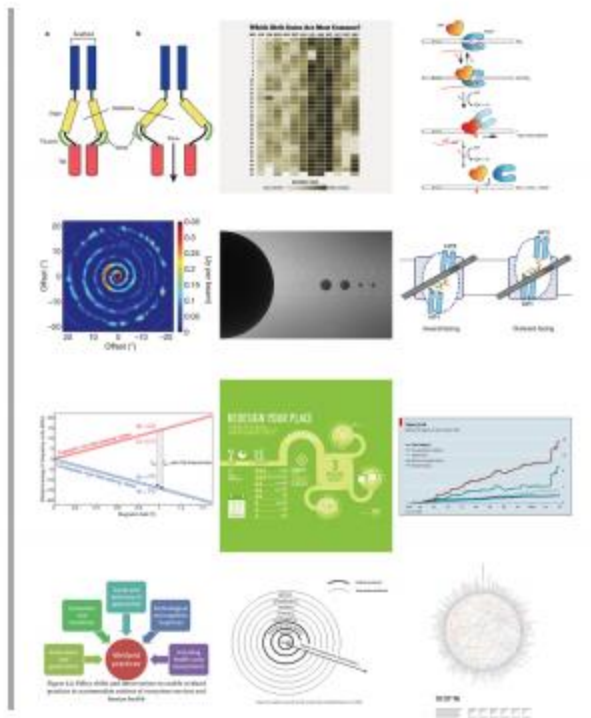
Experiment set up as a game on Amazon Mechanical Turk

- 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)

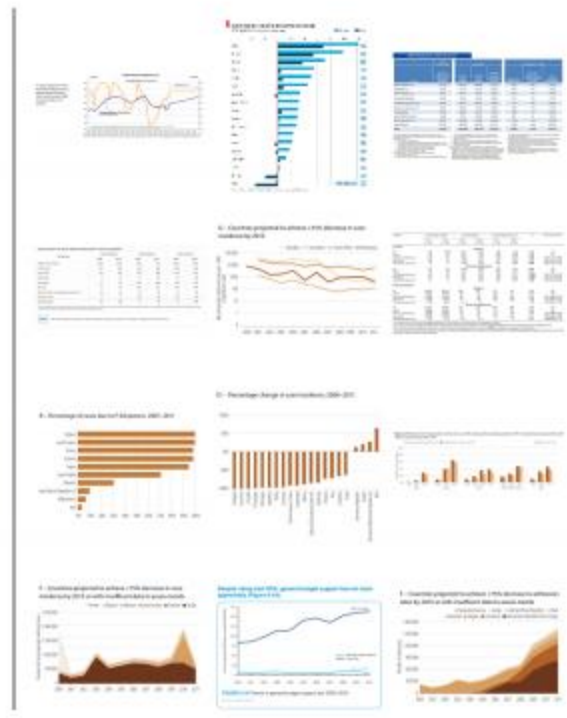
# Memorability Experiment by Borkin et al.



most memorable



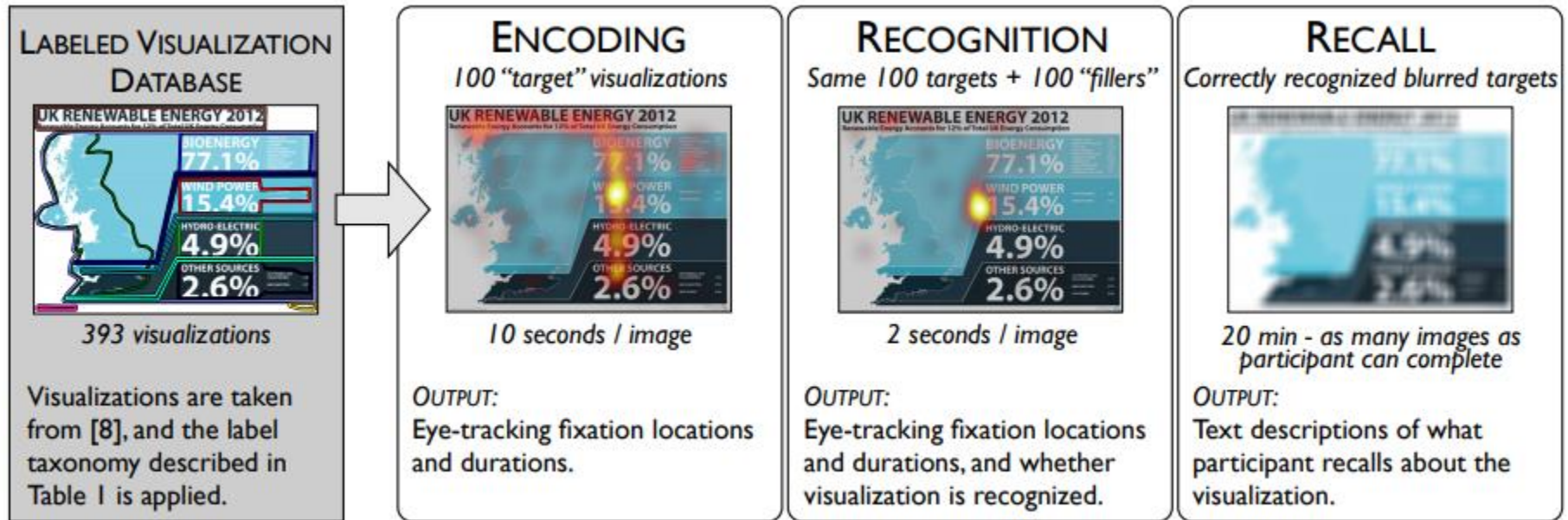
most memorable  
after removing  
human recognizable  
cartoons



least memorable

# What Do People Remember?

## EXPERIMENT DESIGN



# Eye Tracking Experiments

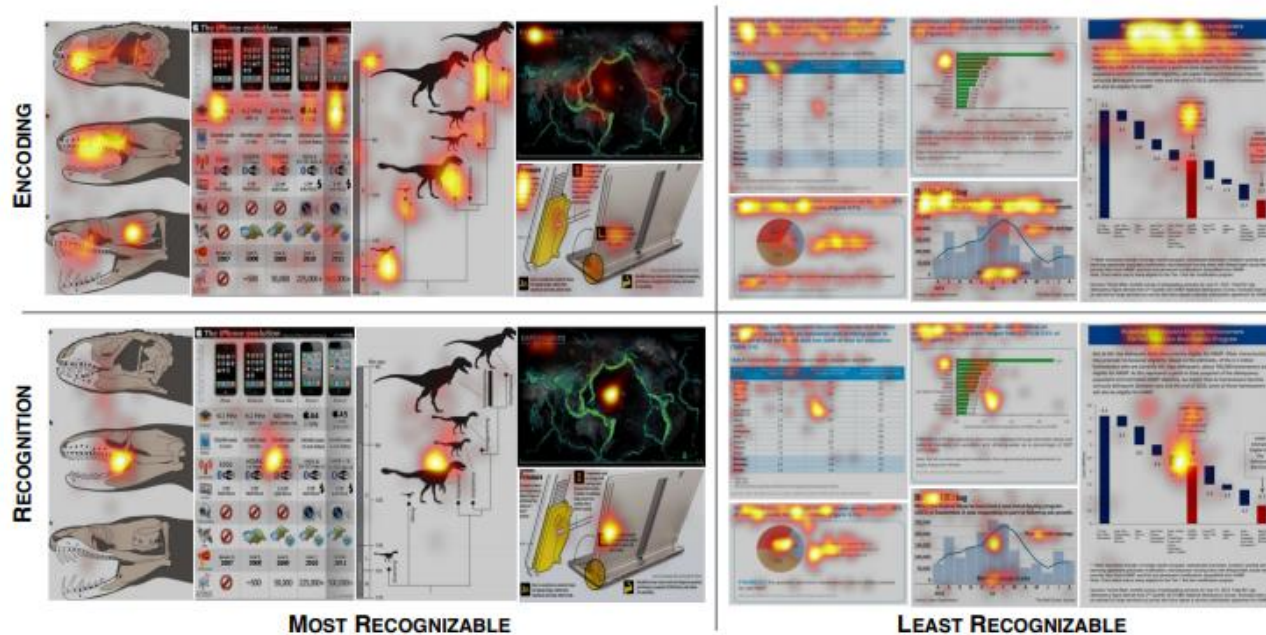
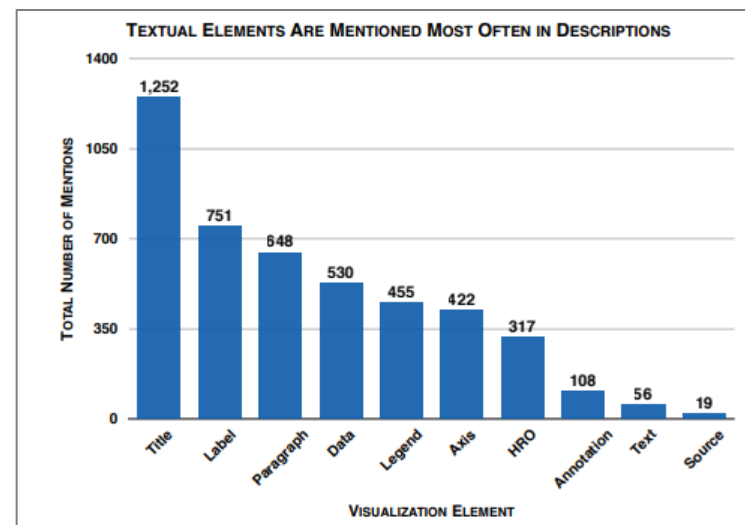


Fig. 7. Examples of the most and least recognizable visualizations from [8]. **TOP:** Eye-tracking fixation heat maps (i.e., average of all participants' fixation locations) from the *encoding* phase of the experiment in which each visualization was presented for 10 seconds. The fixation patterns demonstrate 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.

# 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

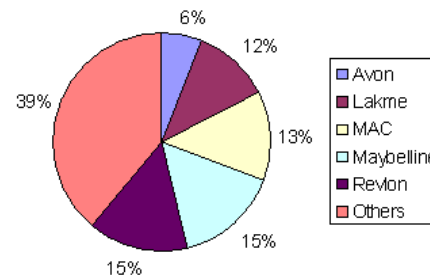
However, link to human engagement not explicitly established

- “just” memorability

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

- but prefer conventional charts for problem solving

Top 5 Lipsticks Survey Results

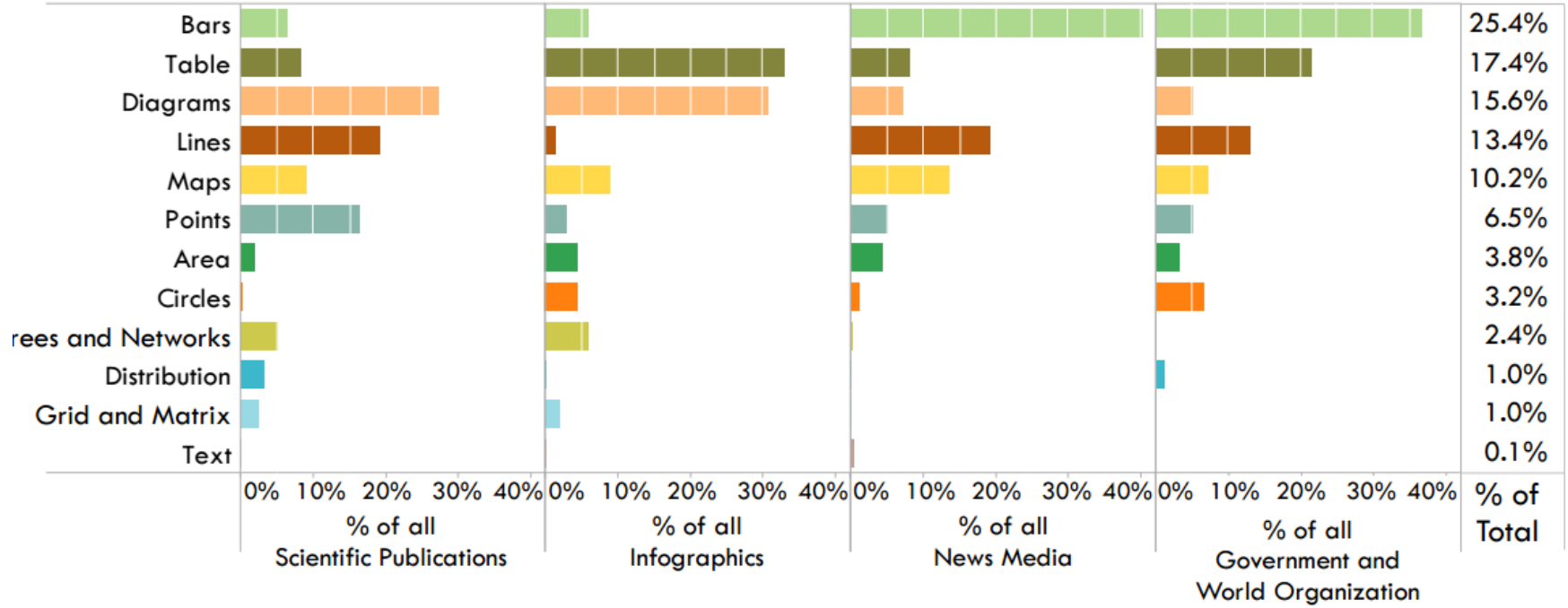


VS.



# Visualizations Sources and Origins

Percent of Visualization Source by Visualization Type



# 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

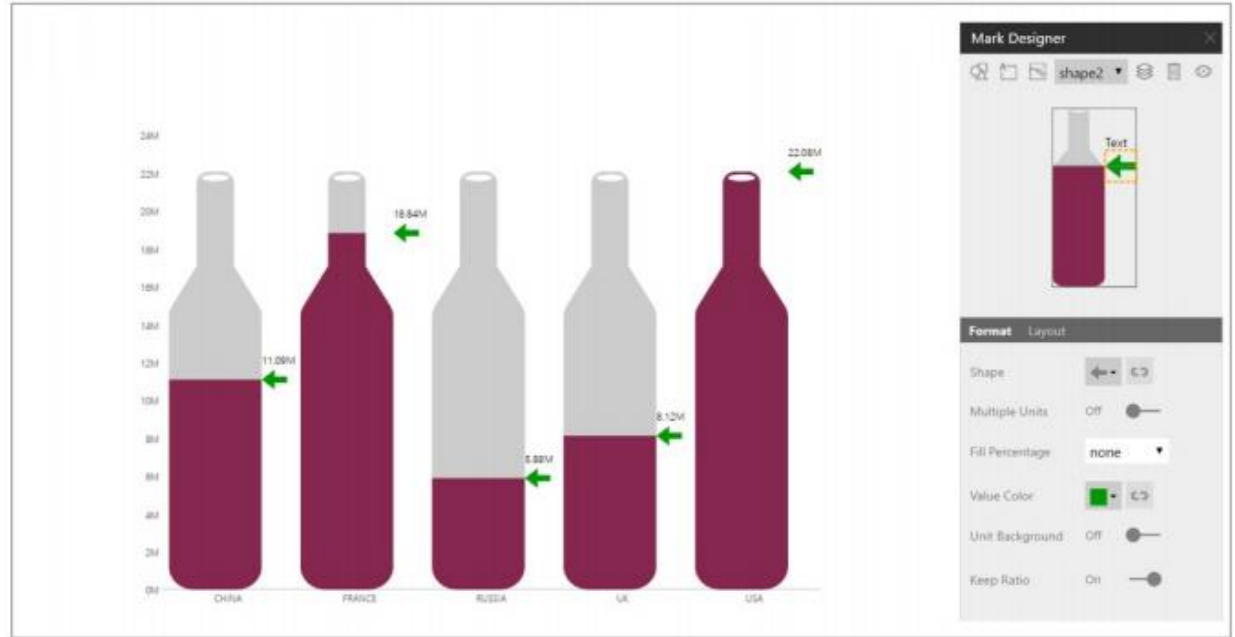
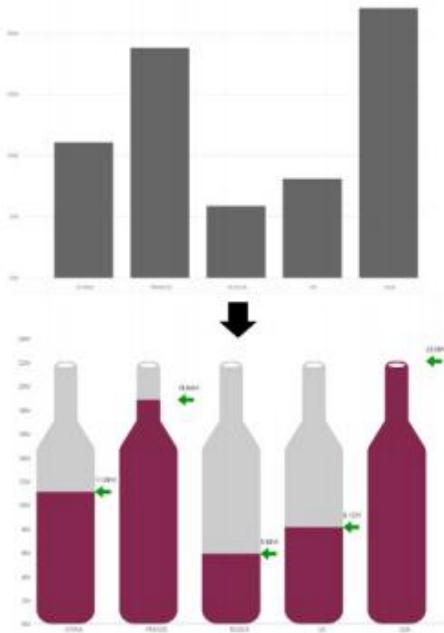
- designed with fewer assumptions about the readers knowledge base than other types of visualizations
- but can be misleading and express the opinion of the author



vs.

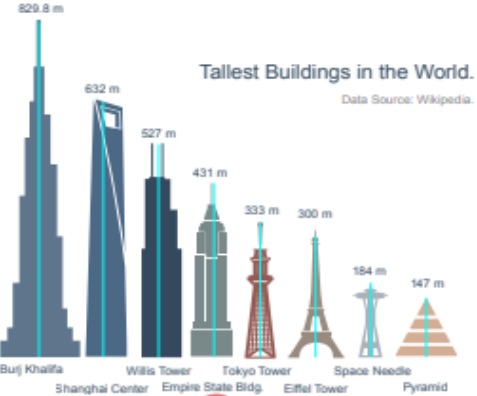


# Using Icons as Bar Graphs



# Data-Driven Design Guides

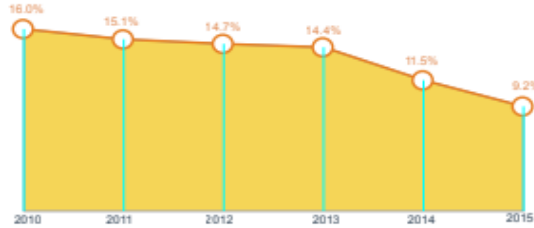




a

### American's Uninsured Rate Dips Below 10%

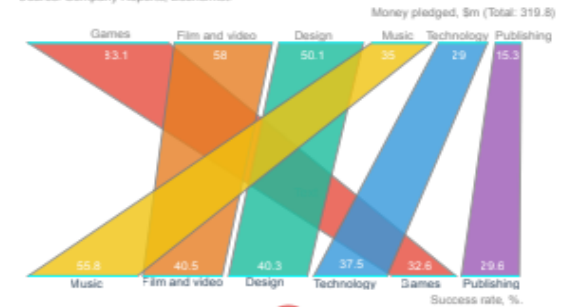
Source: CDC/NCHS, National Health Interview Survey, 2010-2015.  
% of uninsured rate in the U.S.



b

### Crowdfunded Projects on Kickstarter in 2012

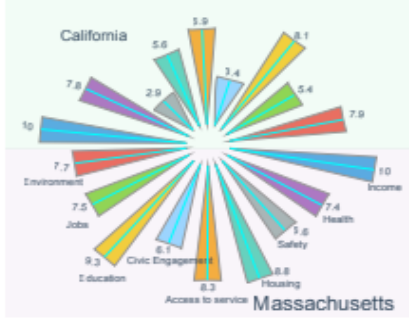
Source: Company Reports, Economist.



c

### Well-Being: California vs Massachusetts

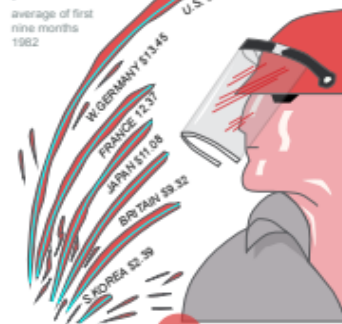
Source: OECD Regional Well-Being (10 point scale).



d

### Employment Costs for a steel worker per hour.

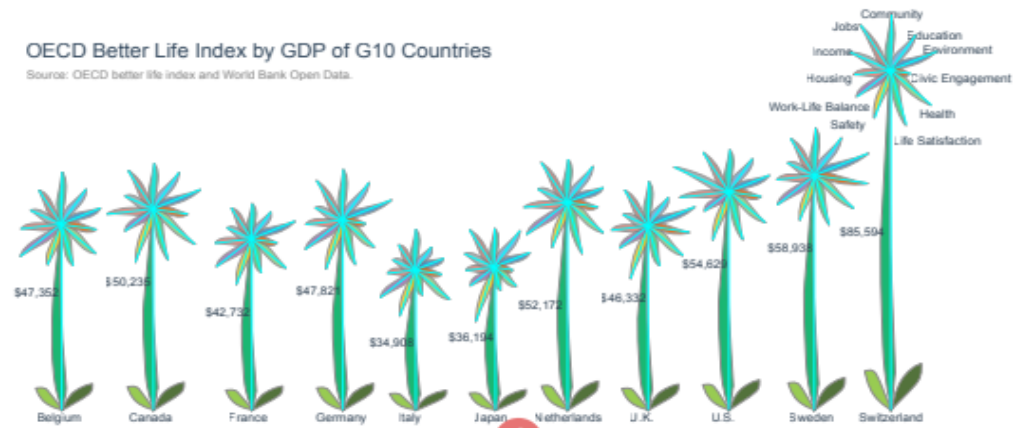
average of first nine months 1992



e

### OECD Better Life Index by GDP of G10 Countries

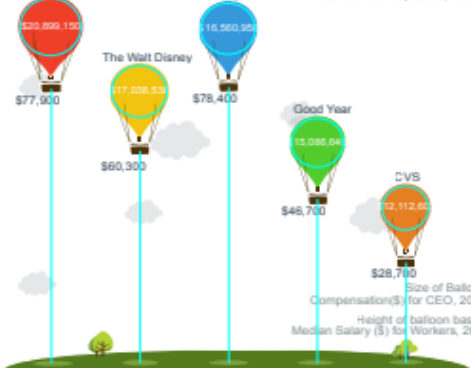
Source: OECD better life index and World Bank Open Data.



f

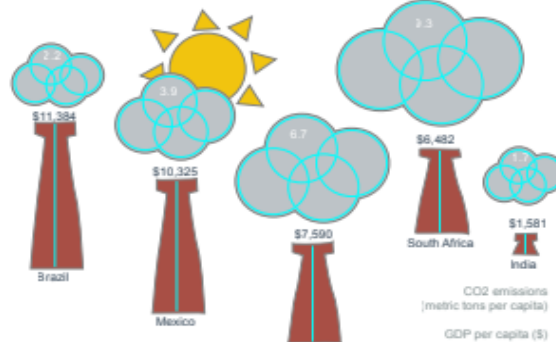
### CEO to Worker Pay Ratios

21st Century Fox, Honeywell, The Walt Disney, Good Year, DVS. Data Source: Pay Scale, Inc.



g

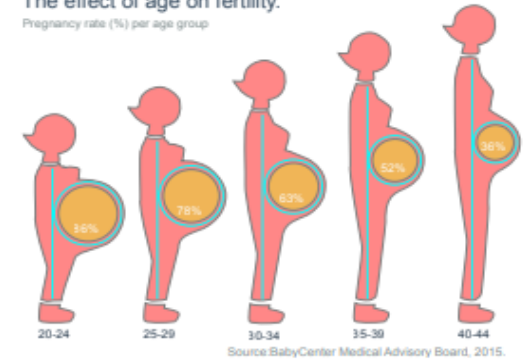
### GDP to CO2 Emissions, G5 Countries



h

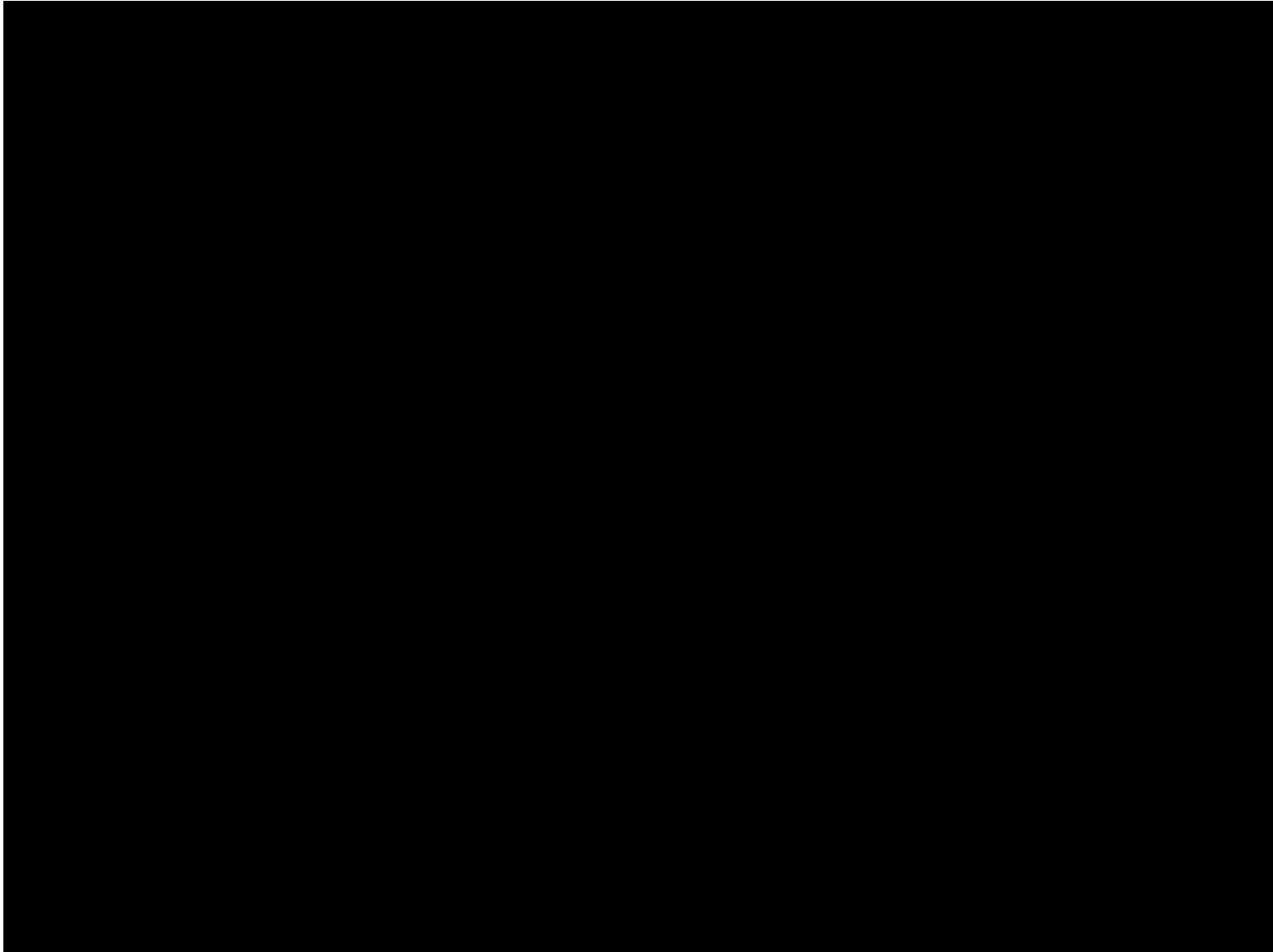
### The effect of age on fertility.

Pregnancy rate (%) per age group



i

# Video



video and more is [here](#)