ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Data Visualization
The Good, the Bad, the Weird

Credit: prof. Nam Wook Kim

Goal

- Rules of thumbs to critique visualization design

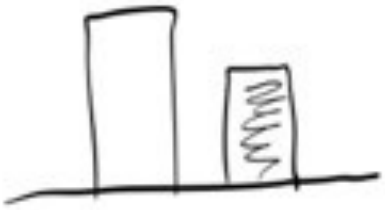
Activity

- Create at least three sketches to visualize these 2 quantities

42 23

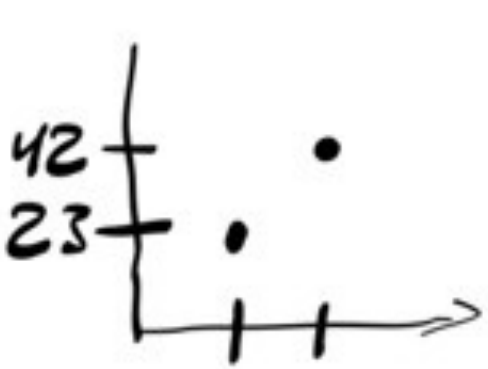
Most likely results

Pie Chart



Bar Chart

Scatterplot



23

42

Arabic Numbers

There are numerous ways to draw even just
two numbers…

Which one is the best and why?

Is there an ideal way to visualize a data set?

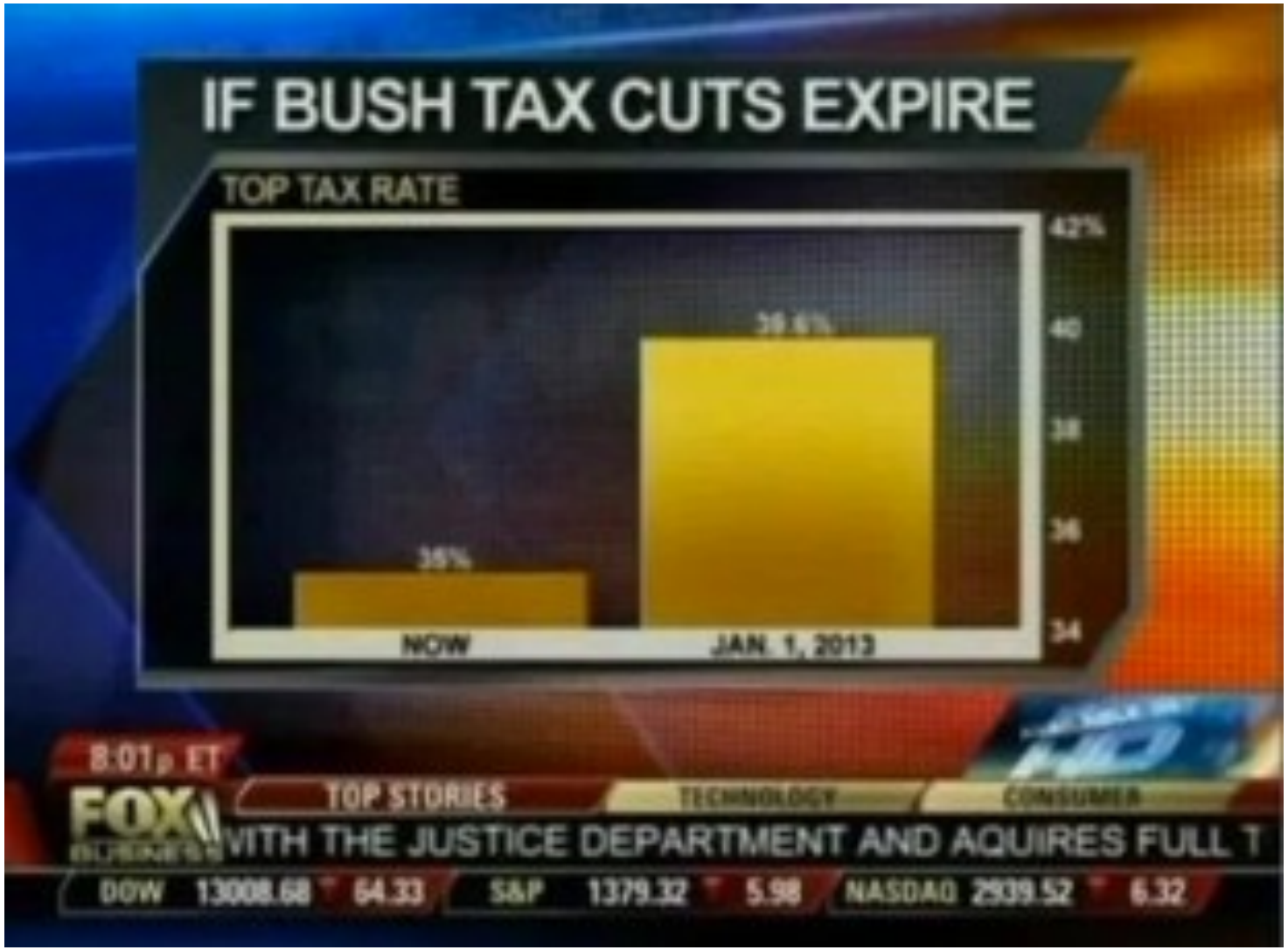
It depends on

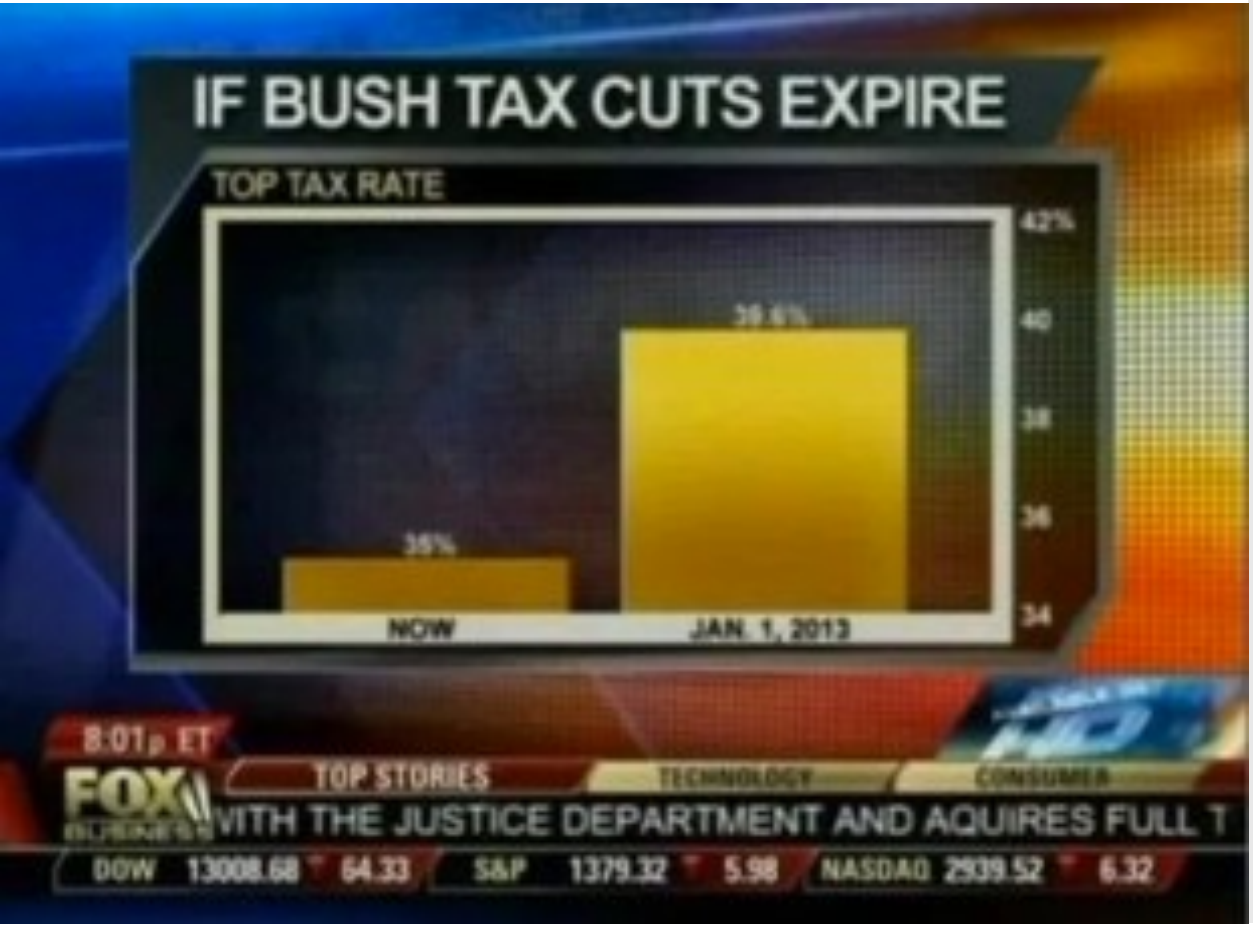
- Data types
  - e.g., table, network, spatial, temporal
- Context of the data
- Tasks to perform
  - e.g., identify trends, compare values
- Questions to answer
- Messages to deliver

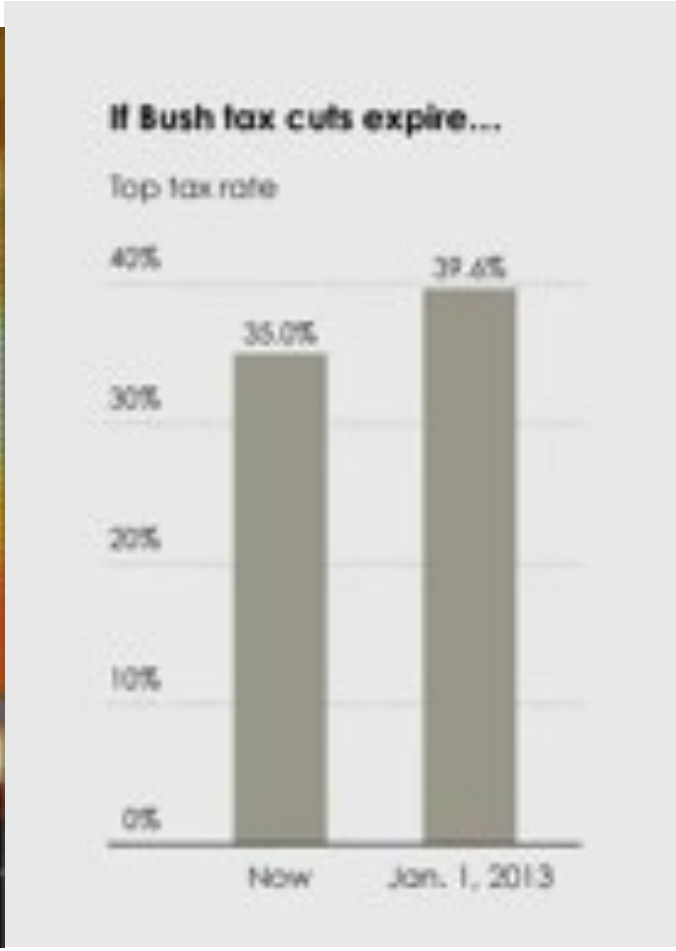
But, is there at least a guide for visualization design?

Edward Tufte's design principles

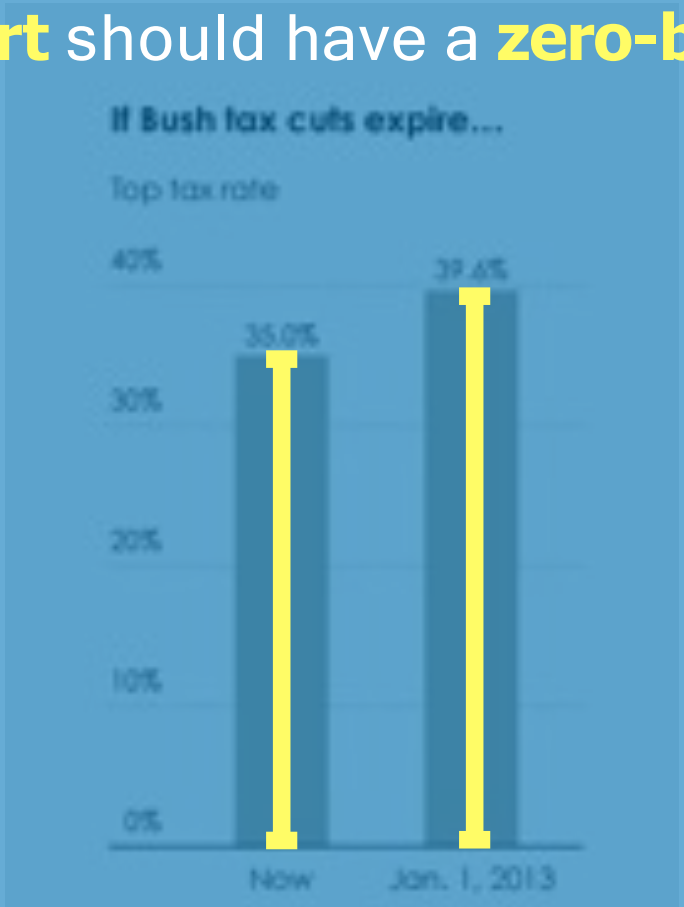
- Above all else show the data
- Maximize data-ink ratio
- Eliminate non-data ink
- Eliminate redundant data ink
- Revise and Edit


IF BUSH TAX CUTS EXPIRE
TOP TAX RATE
42%
40
38
36
34
39.6%
35%
NOW
JAN. 1, 2013
8:01 p. ET
FOX BUSINESS
TOP STORIES
TECHNOLOGY
CONSUMER
VITH THE JUSTICE DEPARTMENT AND AQUIRES FULL T
DOW 13008.68 64.33
S&P 1379.32 5.98
NASDAQ 2939.52 6.32


IF BUSH TAX CUTS EXPIRE
TOP TAX RATE
35%
39.6%
NOW
JAN. 1, 2013
42%
40
38
36
34
8:01p ET
TOP STORIES
TECHNOLOGY
CONSUMER
WITH THE JUSTICE DEPARTMENT AND AQUIRES FULL T
DOW 13008.68 64.33
S&P 1379.32 5.98
NASDAQ 2939.52 6.32

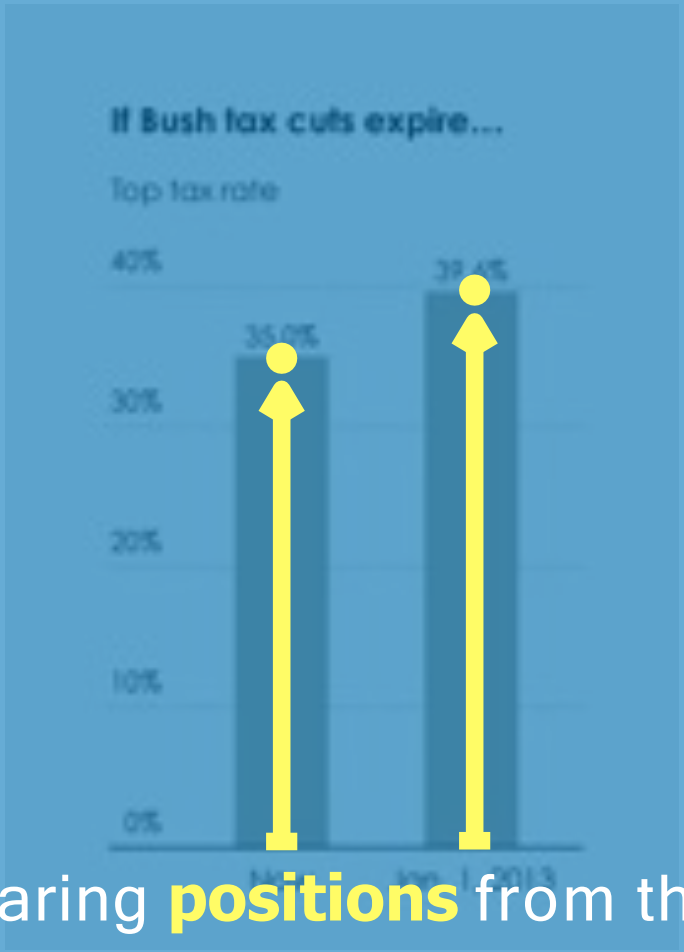

If Bush tax cuts expire...
Top tax rate
40%
30%
20%
10%
0%
35.0%
39.6%
Now
Jan. 1, 2013

Bar Chart should have a **zero-baseline**.



Because you are comparing the **lengths**

Flowing data



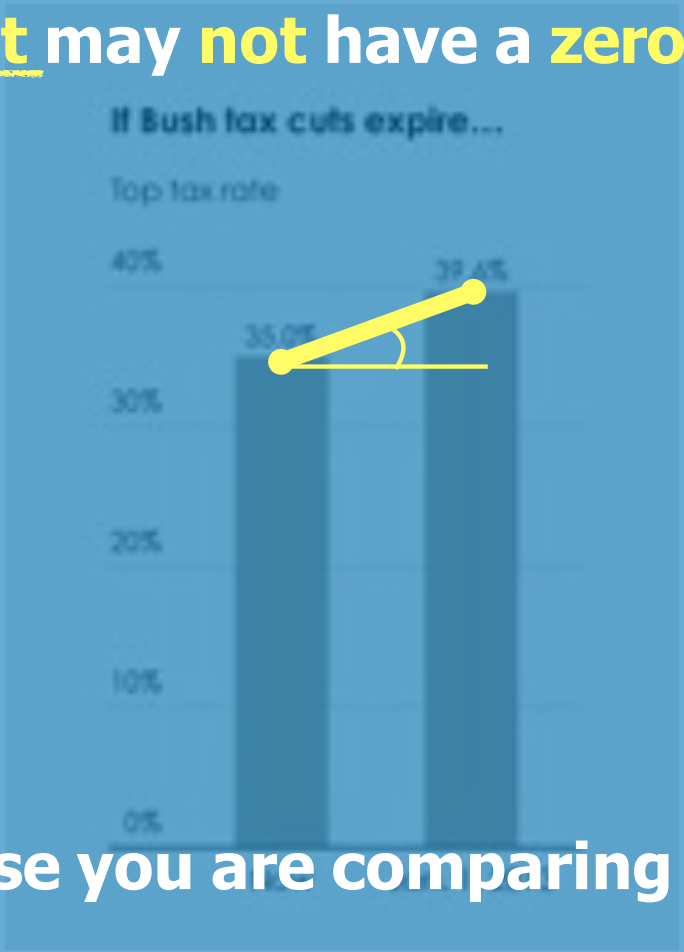
Or comparing **positions** from the **baseline**.

Do we always need a zero-baseline?

"USE A BASELINE THAT SHOWS THE DATA, NOT THE ZERO POINT."

Edward Tufte

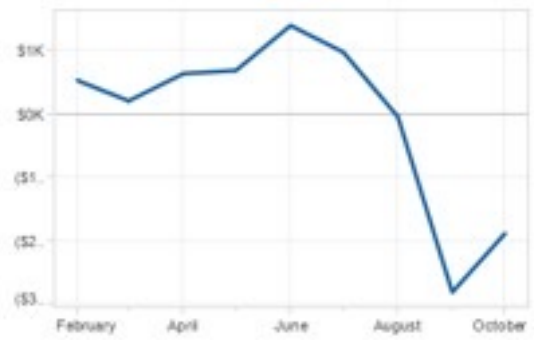
Line Chart may not have a zero-baseline.

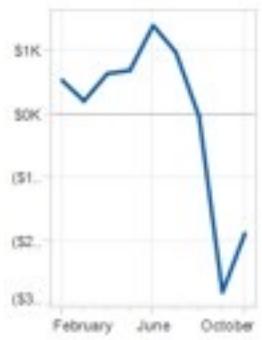


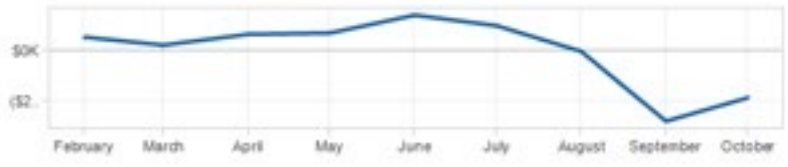
Because you are comparing the angle

Flowing data

Line chart's aspect ratios can matter too

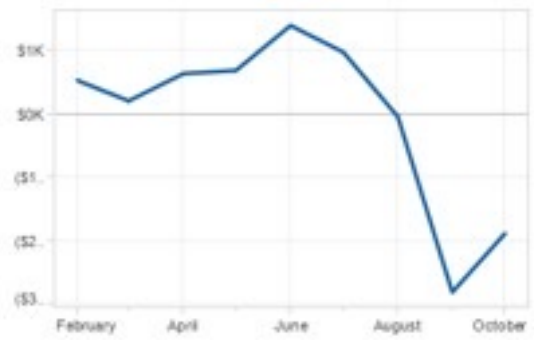


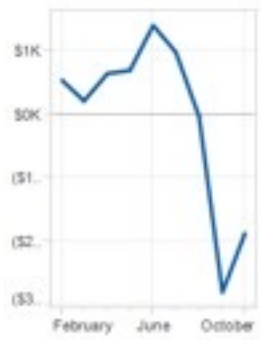


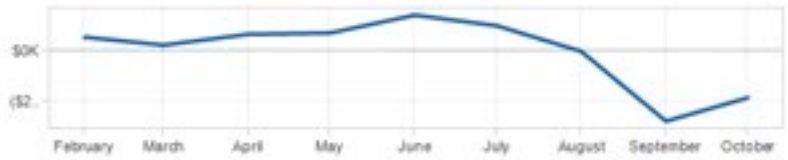


https://eagereyes.org/basics/banking-45-degrees

Line chart's aspect ratios can matter too





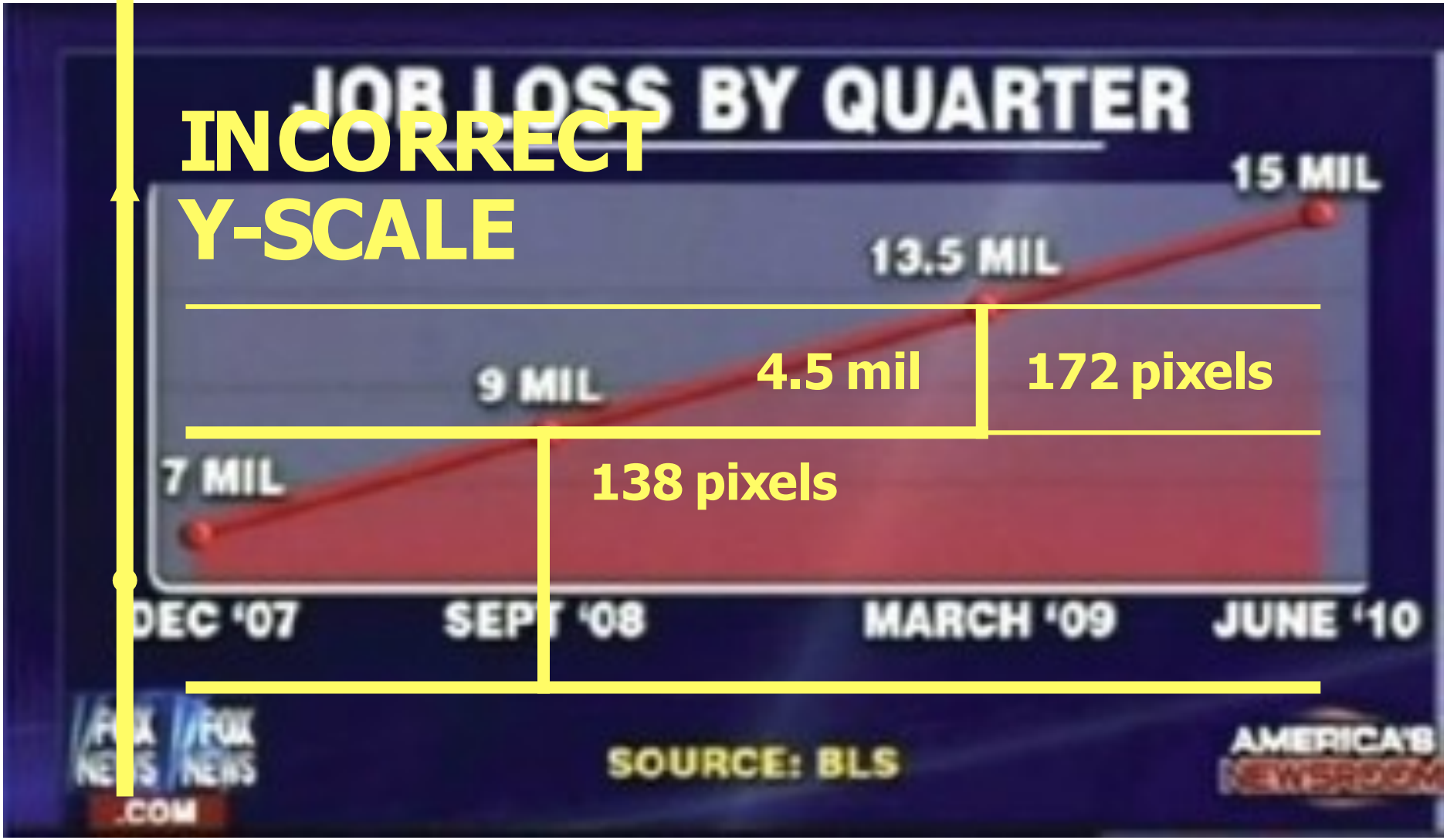


A rule of thumb is banking to 45 degrees (the average line slope in a line chart should be 45)

to minimize errors in visual judgments of slope ratios.


JOB LOSS BY QUARTER
15 MIL
13.5 MIL
9 MIL
7 MIL
DEC '07
SEPT '08
MARCH '09
JUNE '10
FOX NEWS .COM
SOURCE: BLS
AMERICA'S NEWSROOM


JOB LOSS BY QUARTER
15 MIL
13.5 MIL
9 MIL
7 MIL
RANDOM QUARTER
DEC '07
SEPT '08
MARCH '09
JUNE '10
FOX NEWS .COM
SOURCE: BLS
AMERICA'S NEWSROOM


JOB LOSS BY QUARTER
INCORRECT
Y-SCALE
15 MIL
13.5 MIL
9 MIL
7 MIL
4.5 mil
172 pixels
138 pixels
DEC '07
SEPT '08
MARCH '09
JUNE '10
SOURCE: BLS
AMERICA'S

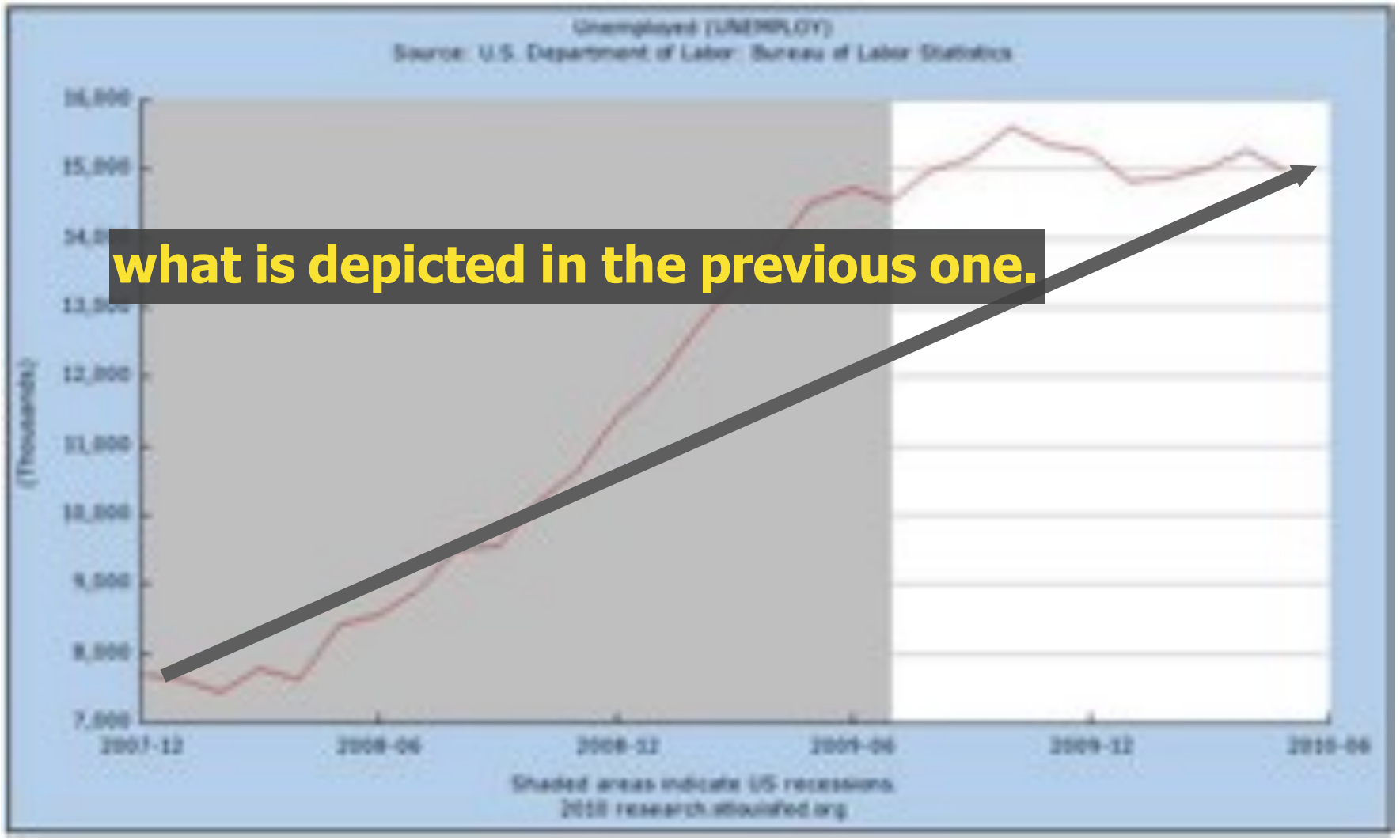

JOB LOSS BY QUARTER
MISLEADING TITLE
15 MIL
13.5 MIL
9 MIL
7 MIL
15 MIL JOB LOSS?
DEC '07
SEPT '08
MARCH '09
JUNE '10
FOX NEWS .COM
SOURCE: BLS
AMERICA'S NEWSROOM

The truth is…

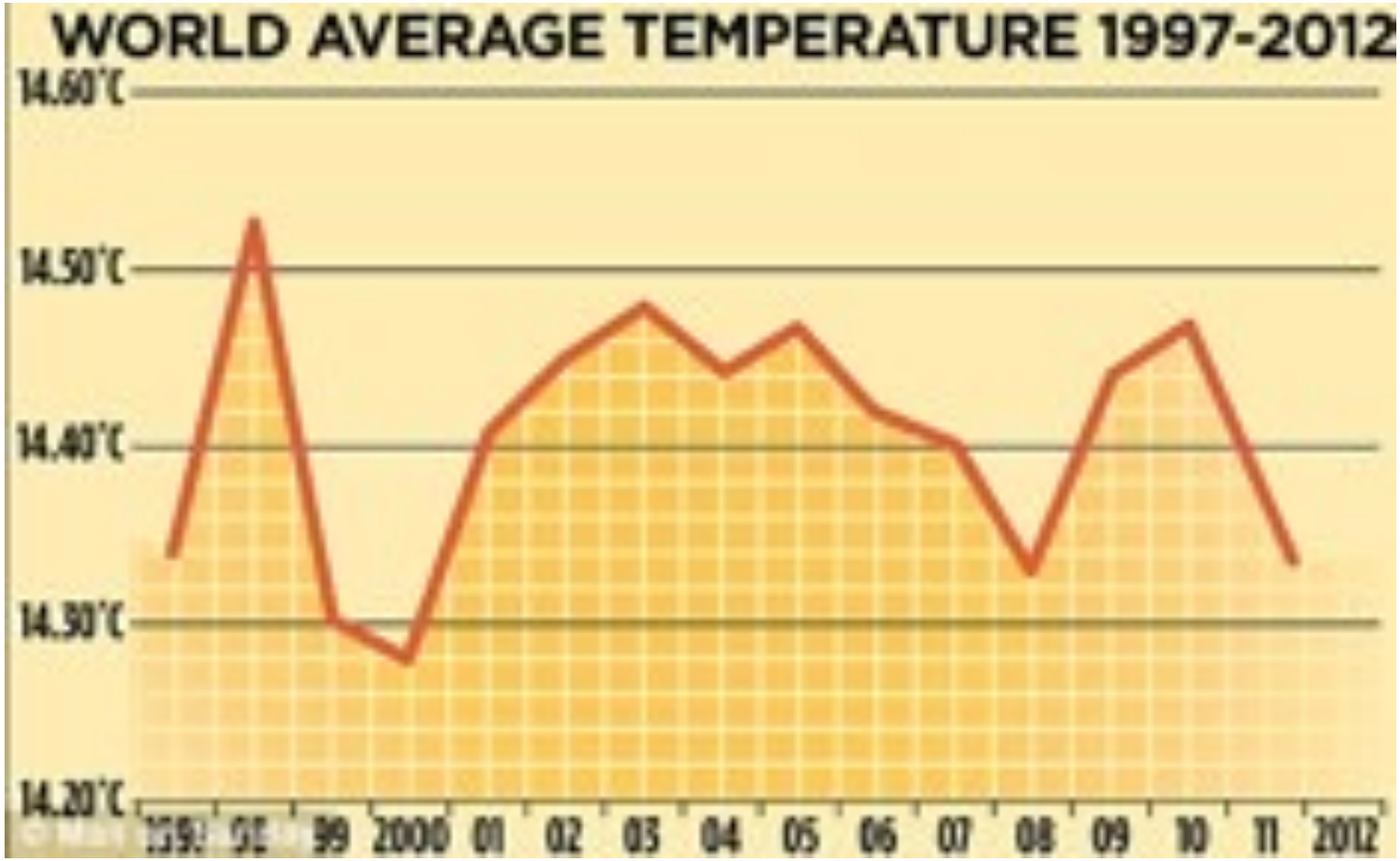
unemployed not job loss!

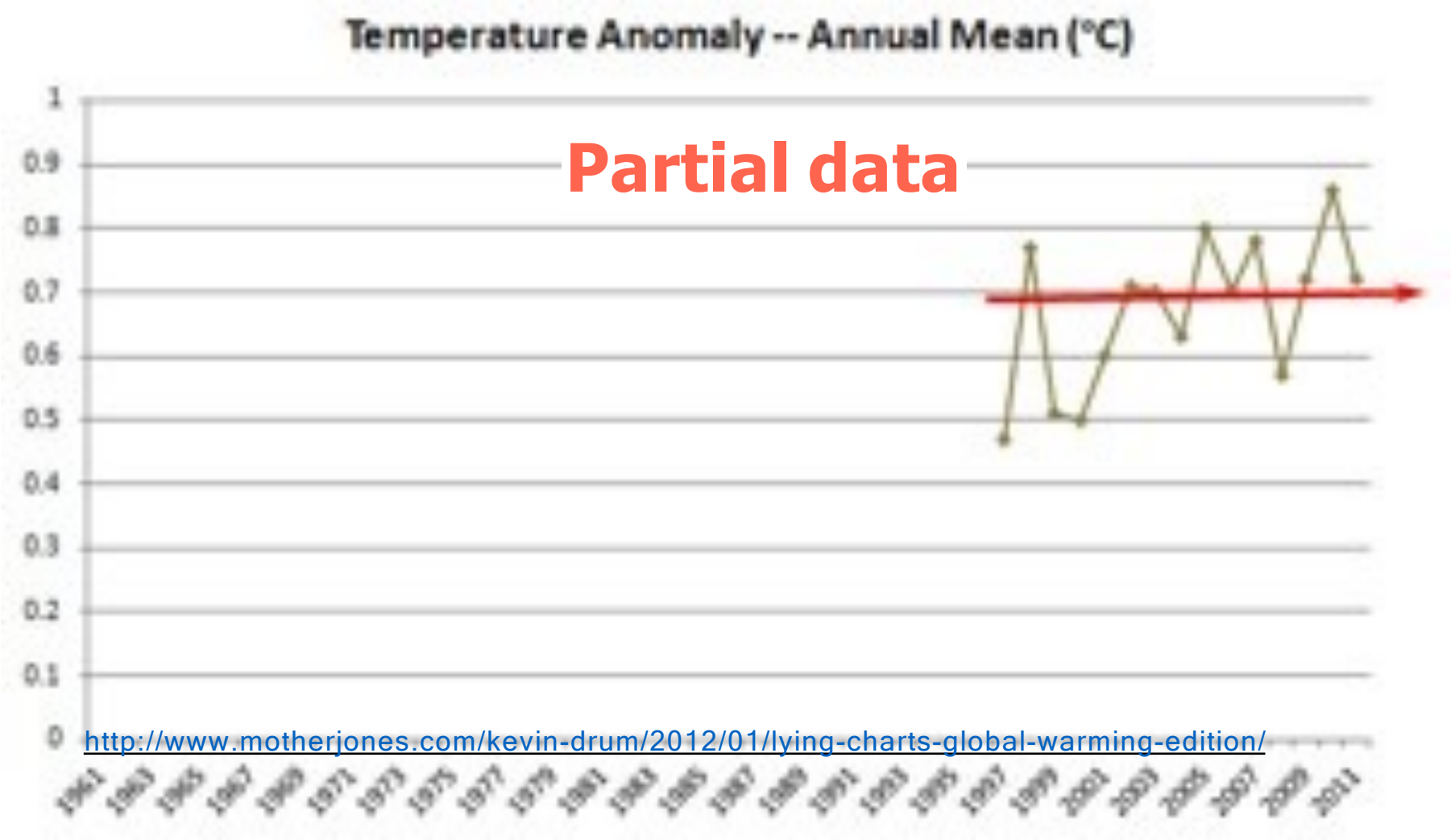
stabilized after 2009

SOURCE: Worst chart I've seen all day, Jocelyn Fong (2010)

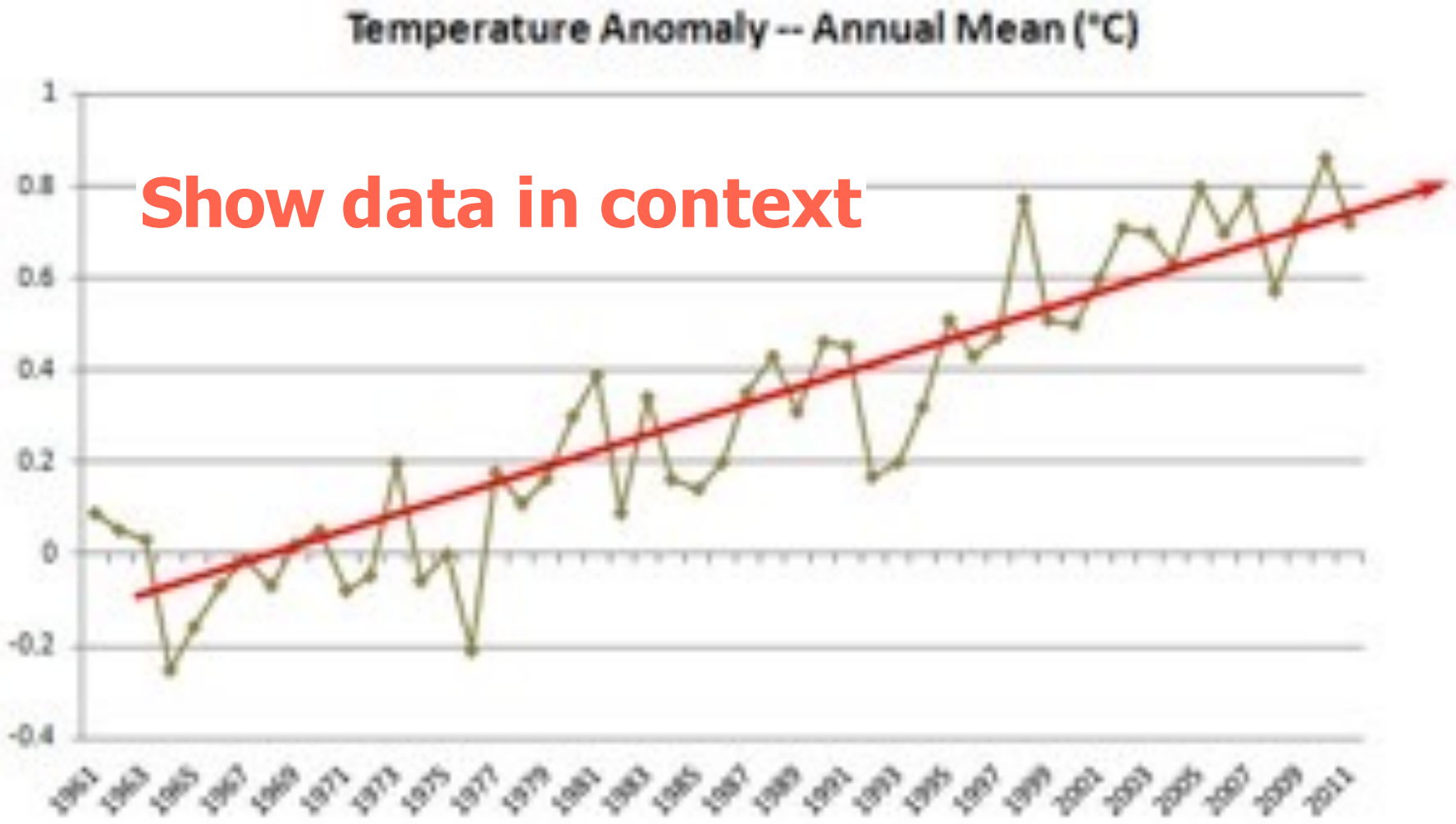


what is depicted in the previous one.

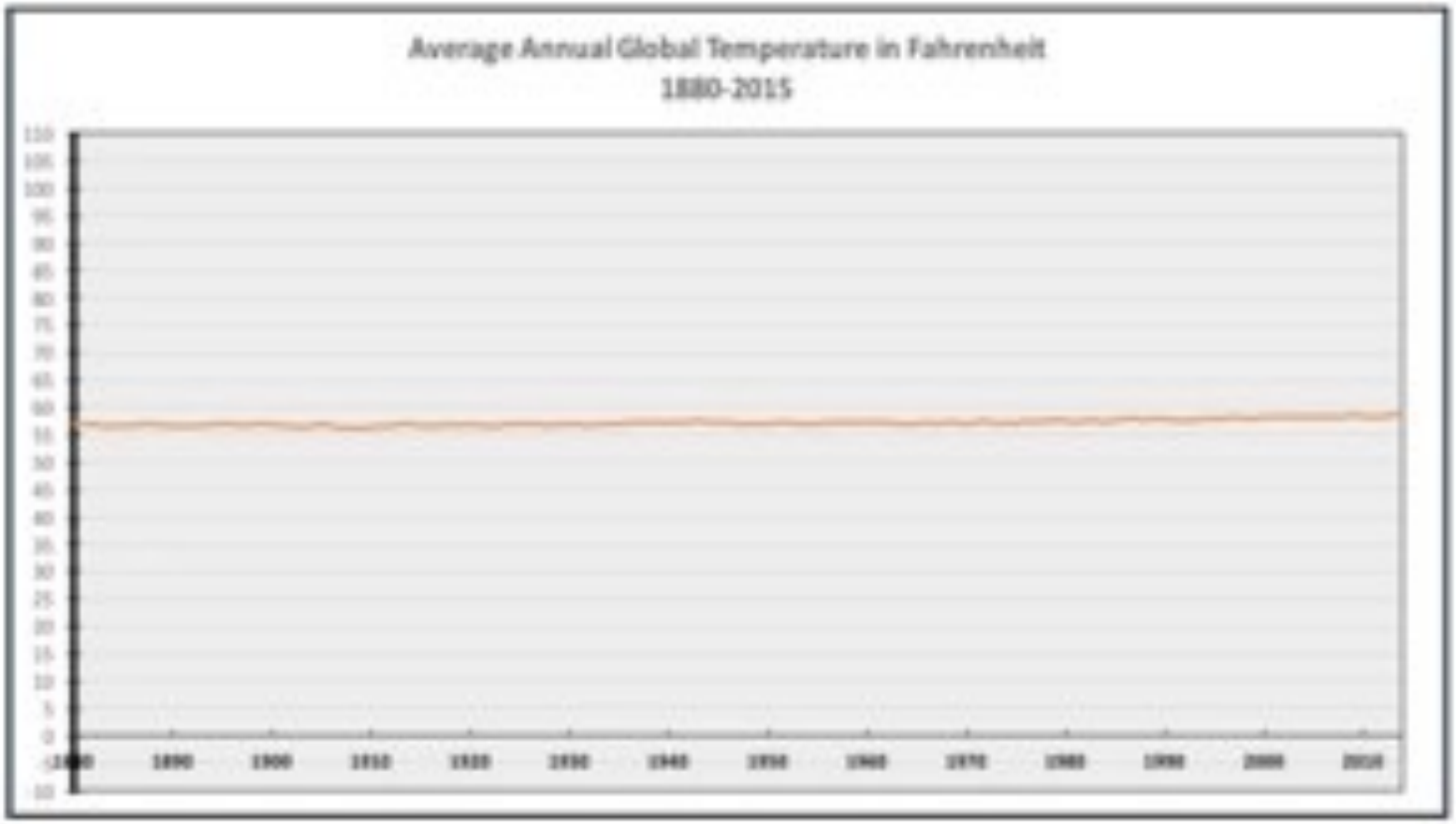

WORLD AVERAGE TEMPERATURE 1997-2012
14.60°C
14.50°C
14.40°C
14.30°C
14.20°C
99
2000
01
02
03
04
05
06
07
08
09
10
11
2012



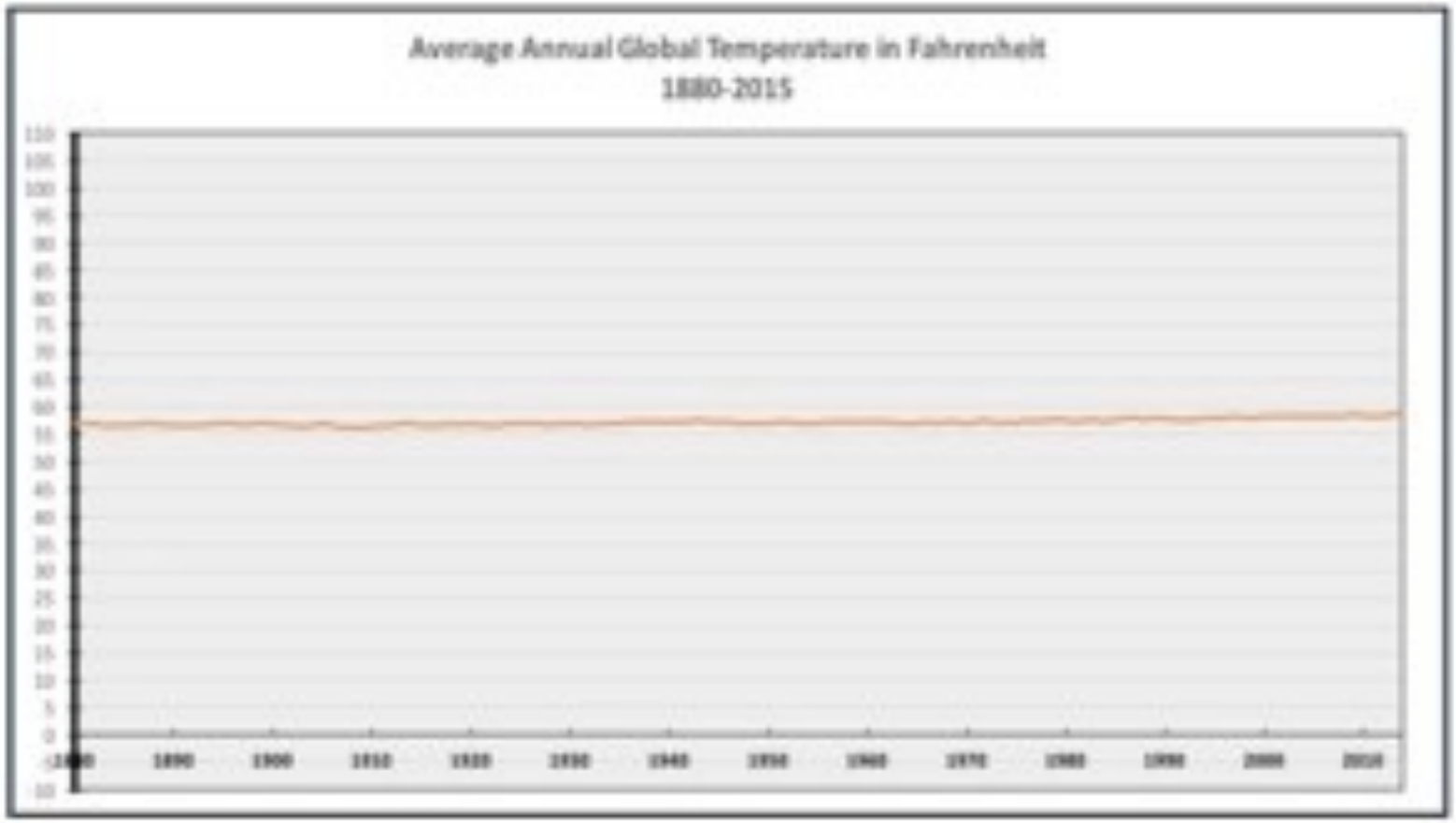
http://www.motherjones.com/kevin-drum/2012/01/lying-charts-global-warming-edition/


Temperature Anomaly -- Annual Mean (°C)
Show data in context
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4

Average Annual Global Temperature (°F) 1880-2015



Choose axis scales wisely



Principles: Tell the truth

- The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities measured.
  - A classic case of misrepresentation of data is when a chart is created with a scale that is completely off.

IFTV

STATS VIA YAHOO AND FOX SPORTS

MARKETING MACHINE

TOTAL JUVE JERSEY'S SOLD IN 2016

850K

RONALDO JERSEYS SOLD IN THE FIRST DAY

520K

NEYMAR JERSEYS SOLD FROM BARCA TO PSG

10K

ĐẠI HỌC BÁCH KHOA 25 YEARS ANNIVERSARY SOICT VIỆN

Lie factor

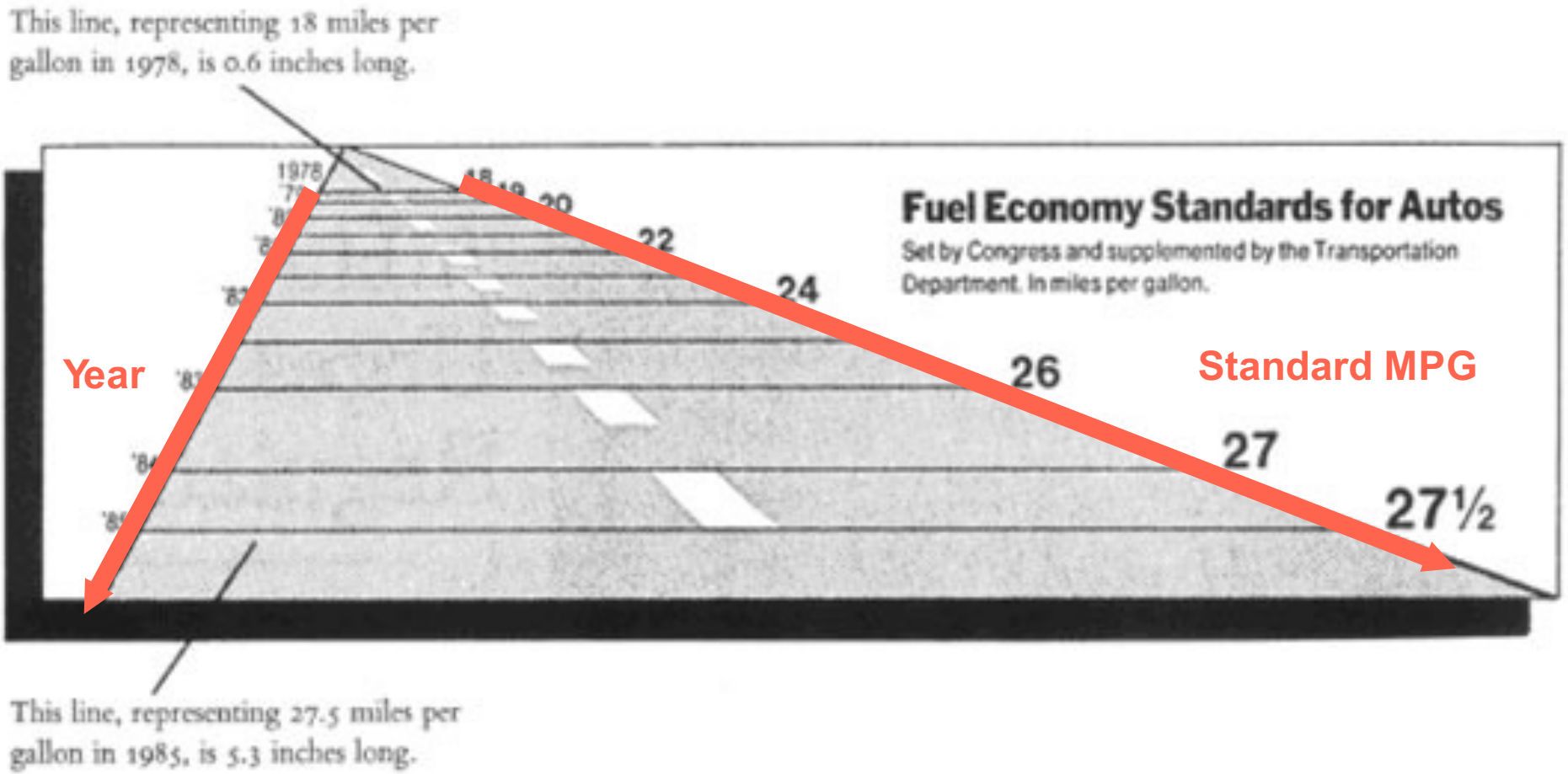
- The “Lie Factor” is a value to describe the relation between the size of effect shown in a graphic and the size of effect shown in the data.

This results in the following formula:

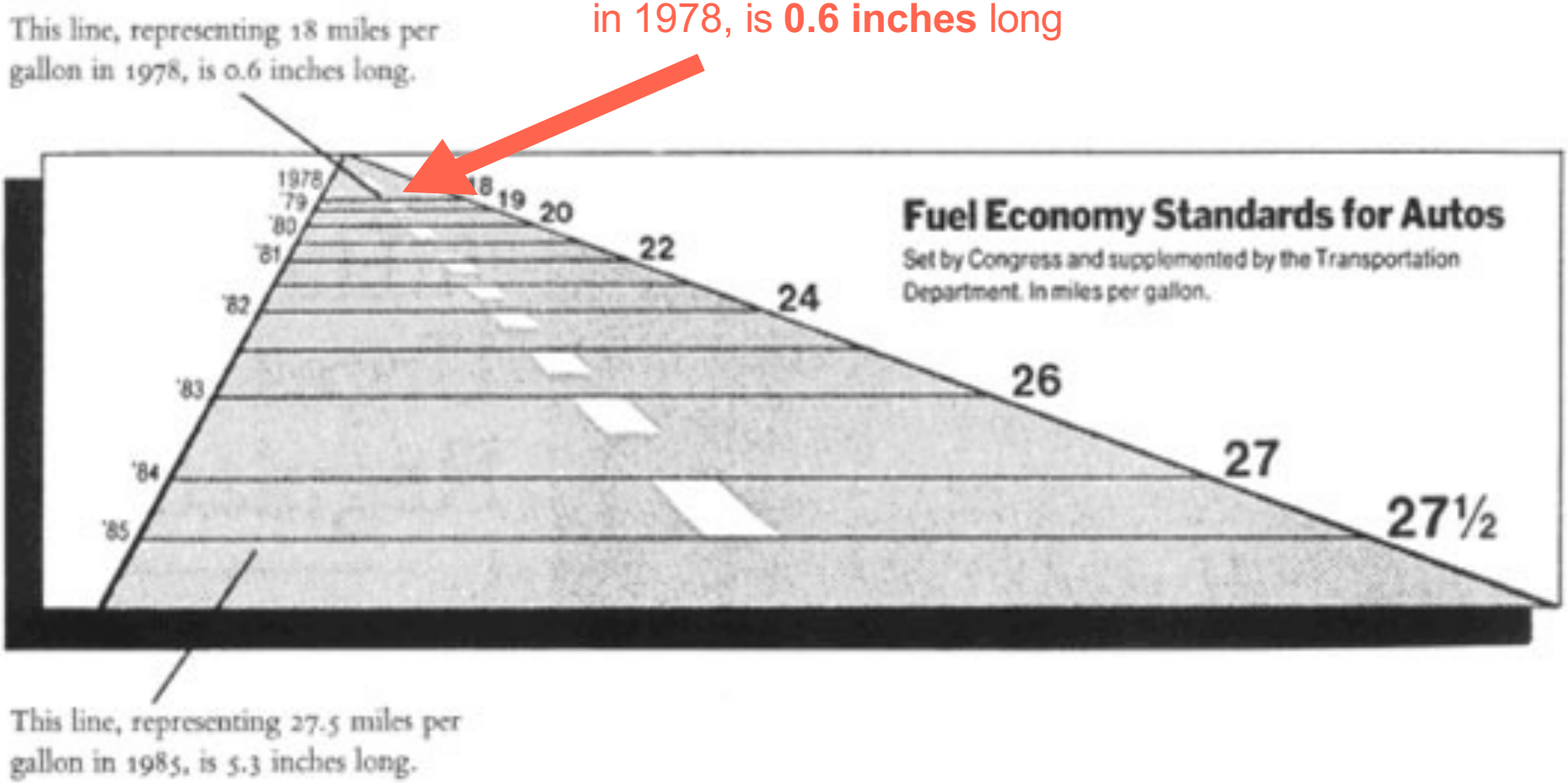
$$Lie\ Factor = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

where

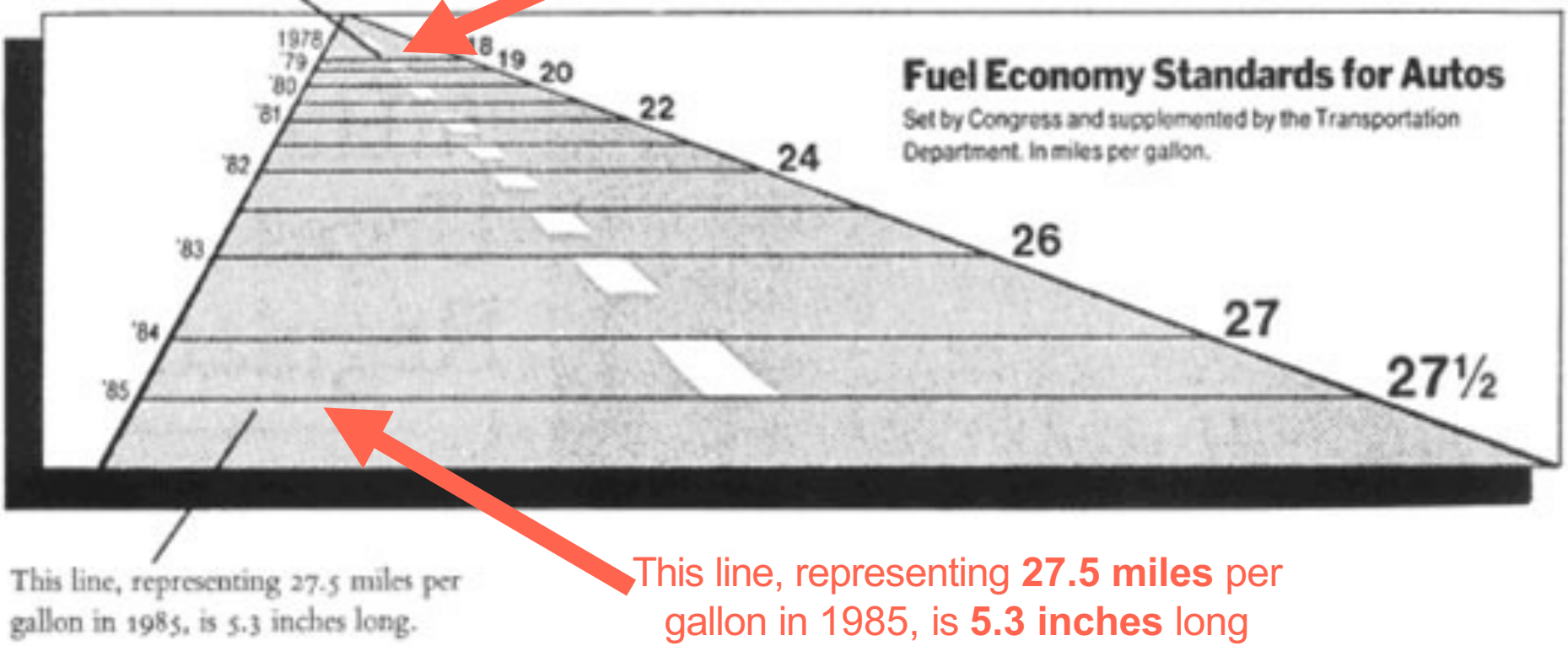
$$size\ of\ effect = \frac{|\text{second value - first value}|}{\text{first value}}$$


This line, representing 18 miles per gallon in 1978, is 0.6 inches long.
Fuel Economy Standards for Autos
Set by Congress and supplemented by the Transportation Department. In miles per gallon.
Year
Standard MPG
1978
18
20
22
24
26
27
27½
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

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

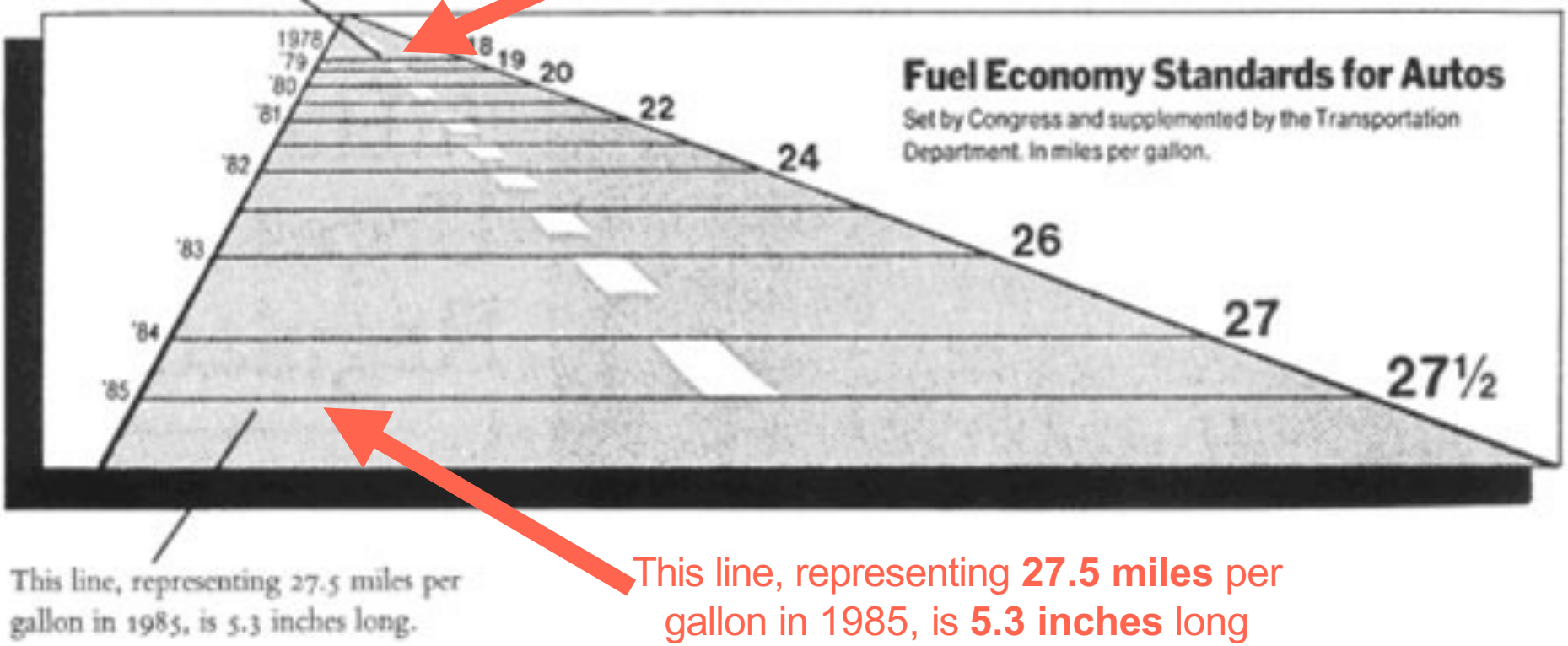


This line, representing **18 miles** per gallon in 1978, is **0.6 inches** long



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

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



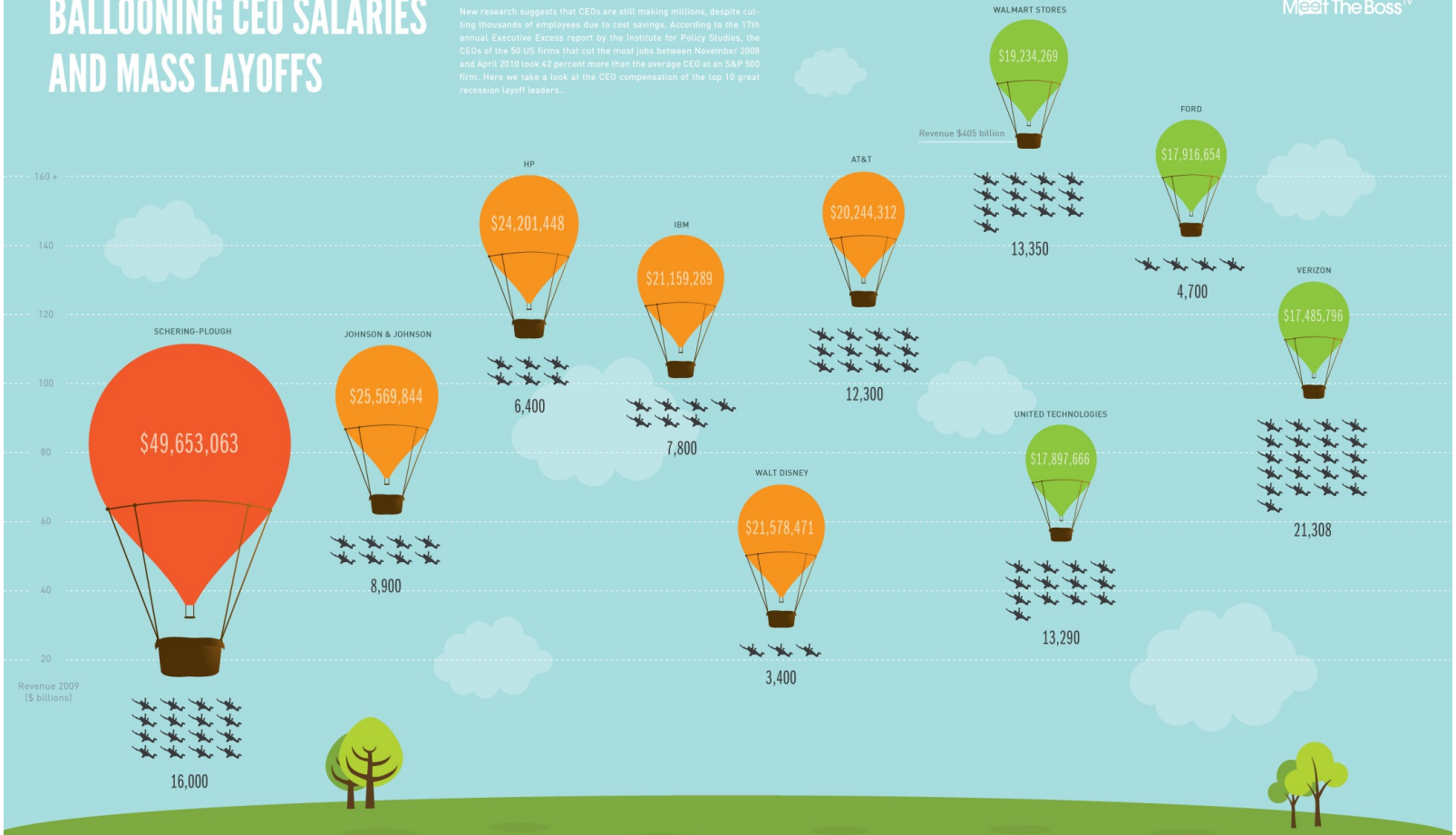
This line, representing **27.5 miles** per gallon in 1985, is **5.3 inches** long

$$\text{Lie Factor} = \frac{(5.3-0.6) / 0.6}{(27.5-18) / 18} = 14.8$$

BALLOONING CEO SALARIES AND MASS LAYOFFS

New research suggests that CEOs are still making millions, despite cutting thousands of employees due to cost savings. According to the 17th annual Executive Excess report by the Institute for Policy Studies, the CEOs of the 50 US firms that cut the most jobs between November 2008 and April 2010 took 42 percent more than the average CEO at an S&P 500 firm. Here we take a look at the CEO compensation of the top 10 great recession layoff leaders...

Meet The Boss TV



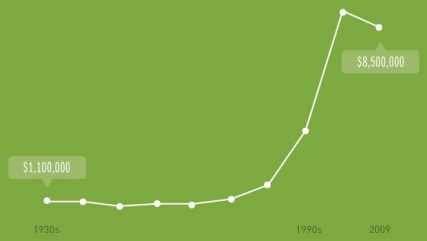
KEY

SLASHING JOBS PAYS

CEOs of the top 50 firms that have laid off the most workers took home on average 42% more than the CEO pay average at S&P 500 firms.

CEO SALARIES OVER TIME

Average annual CEO salary (adjusted to 2000 dollars)



CEO PAY AND UNEMPLOYMENT INSURANCE

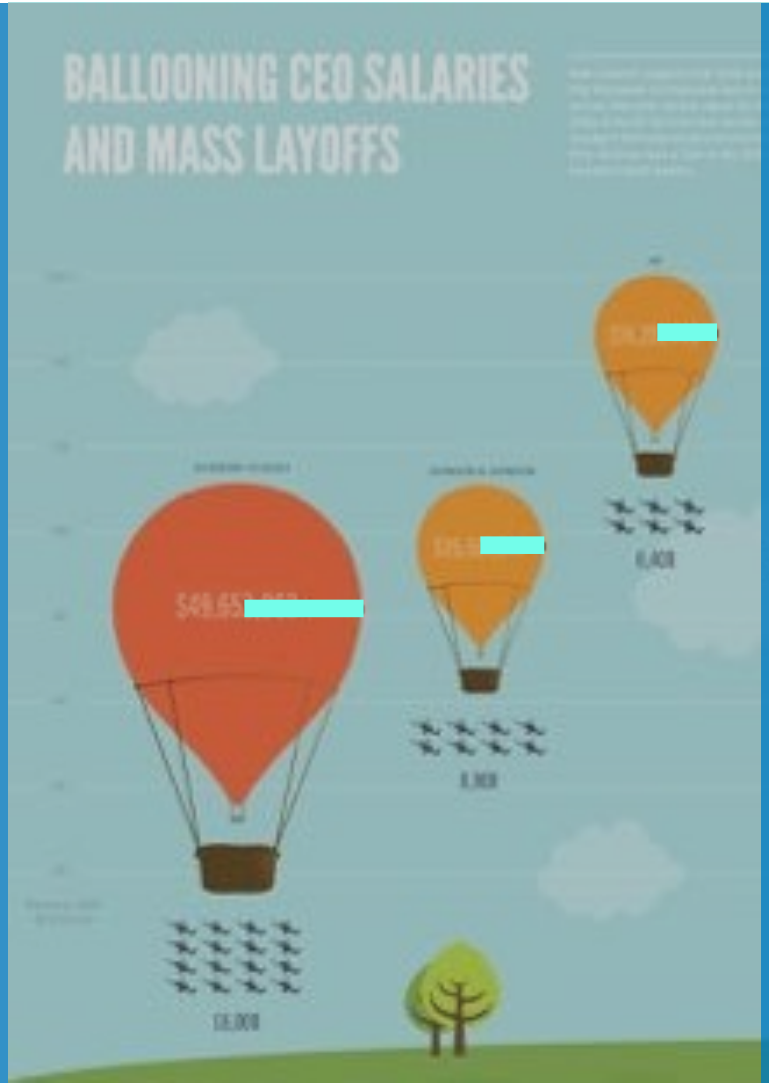
Combined compensation of top 50 CEOs in the layoff leader survey

could pay for a month of benefits to the

531,363

workers their companies laid off

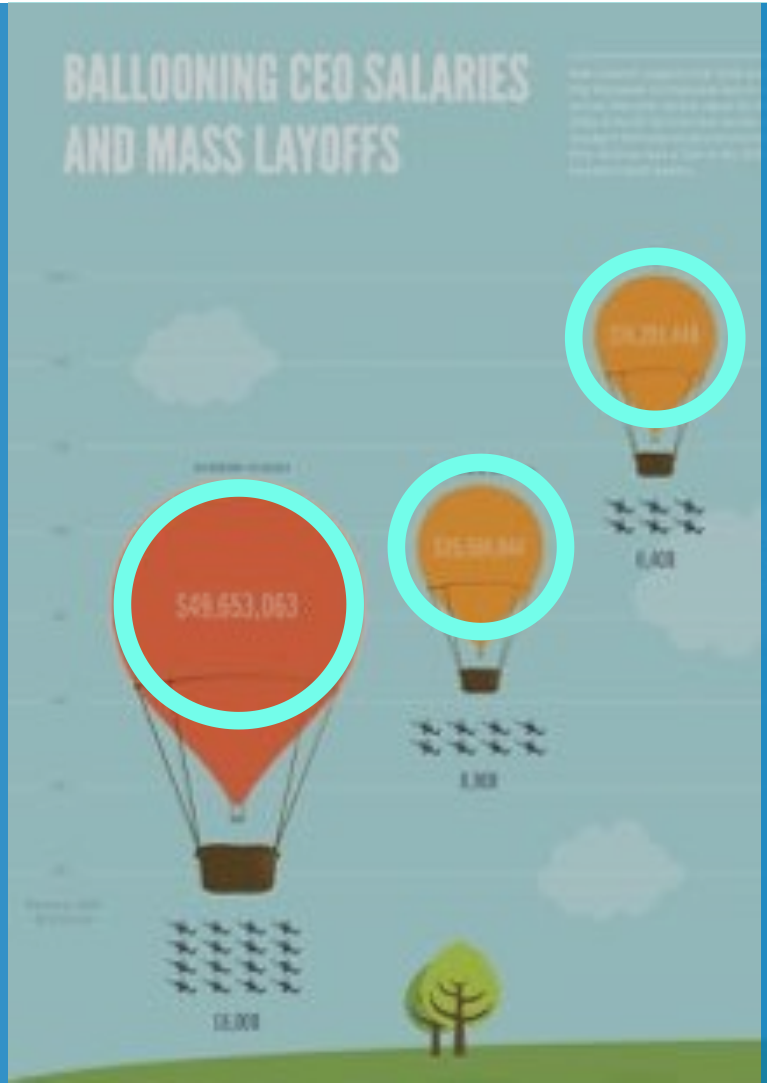
Source: CEO Pay and the Great Recession, 17th Annual Executive Compensation Survey by the Institute of Policy Studies (IPS)
Graphic for Meet the Boss TV | www.meettheboss.tv
Graphic by Tiffany Farrant | Twitter @fallenblossom



This chart uses radius of the balloon to encode the data

Doubling the radius (or data) increases the perceived area by four.

Lie factor = (4-1)/(2-1) = 3.



The size of the balloon should be something like this.

Principles: Maximize data-ink ratio

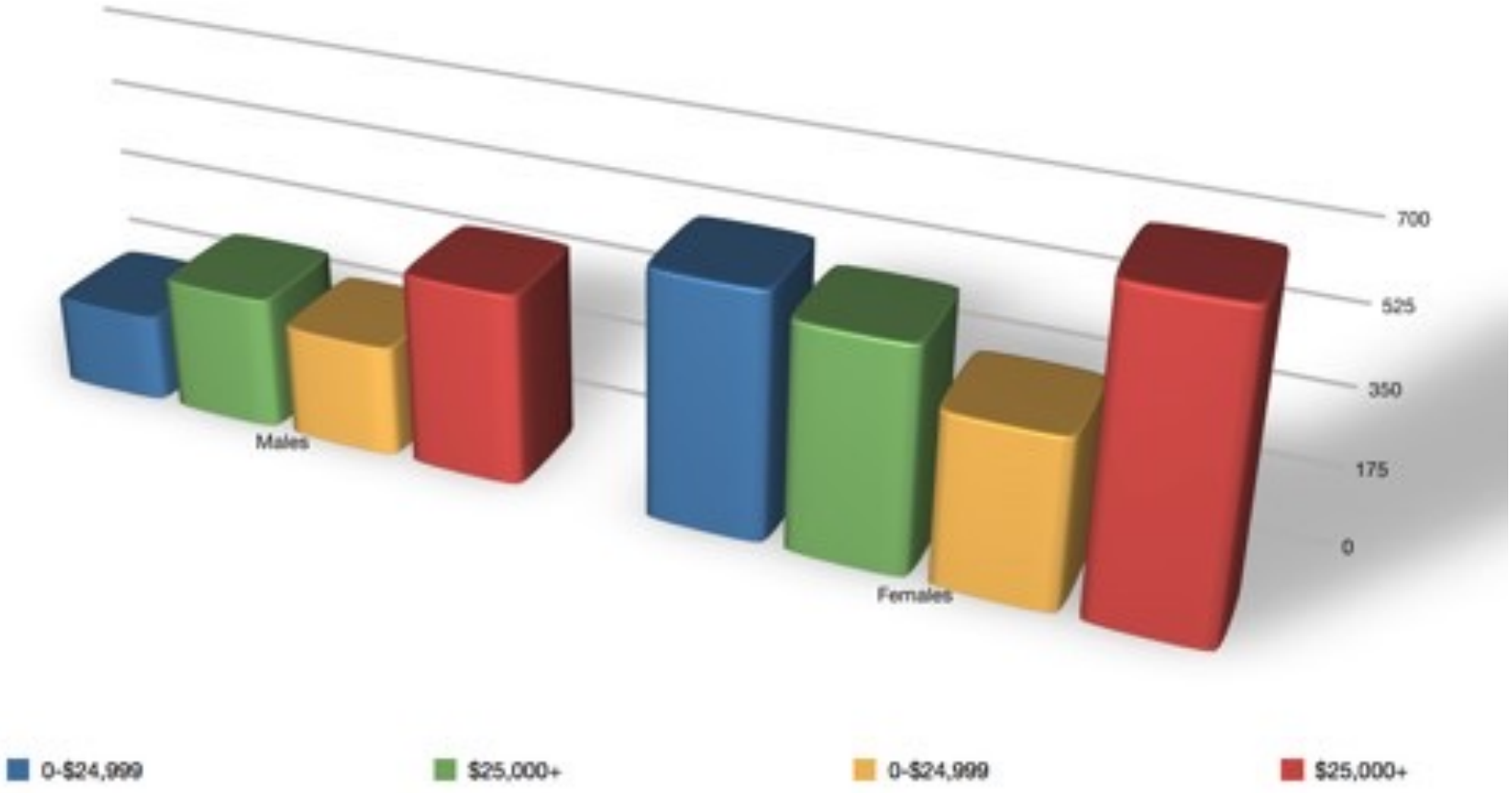
- Use the least ink to represent the greatest amount of information in the smallest space

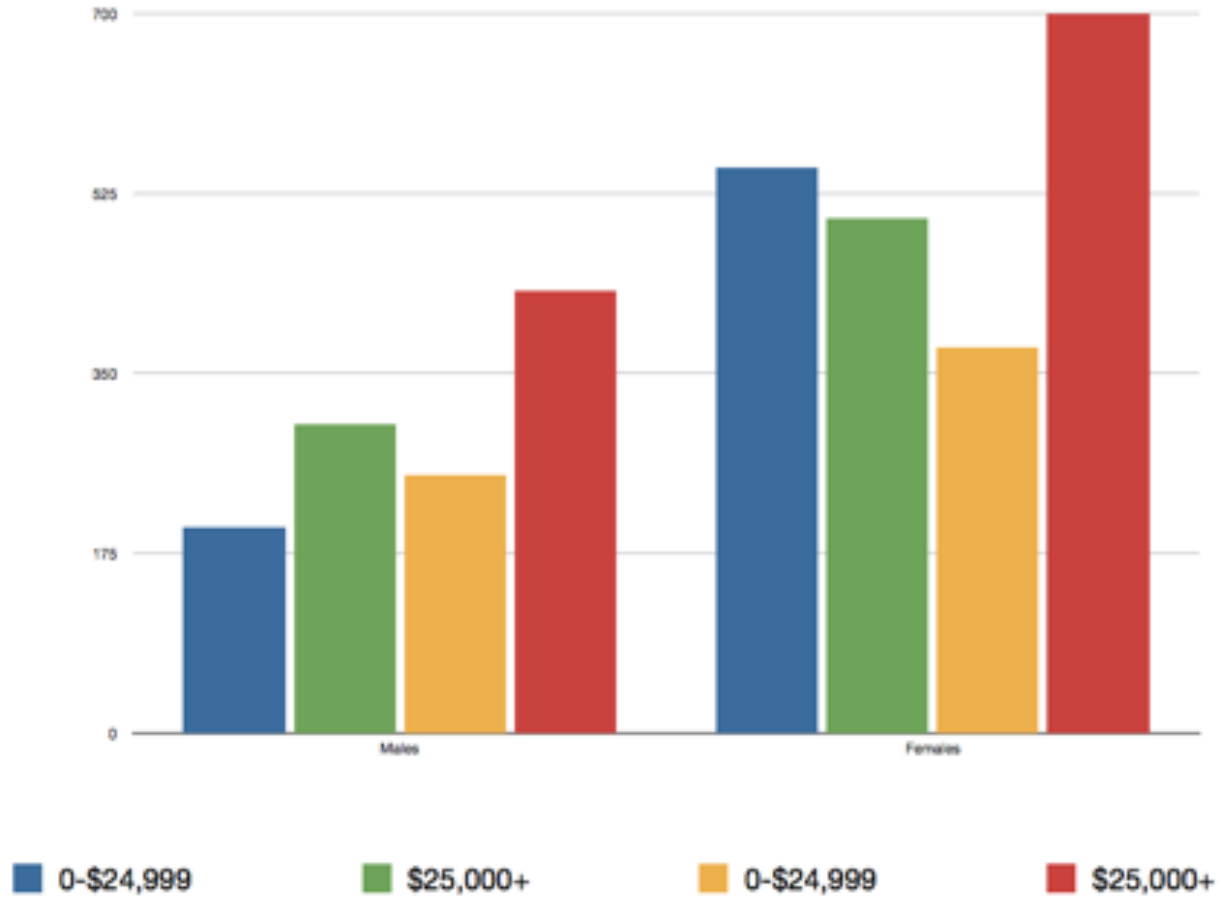
Maximize Data-Ink Ratio

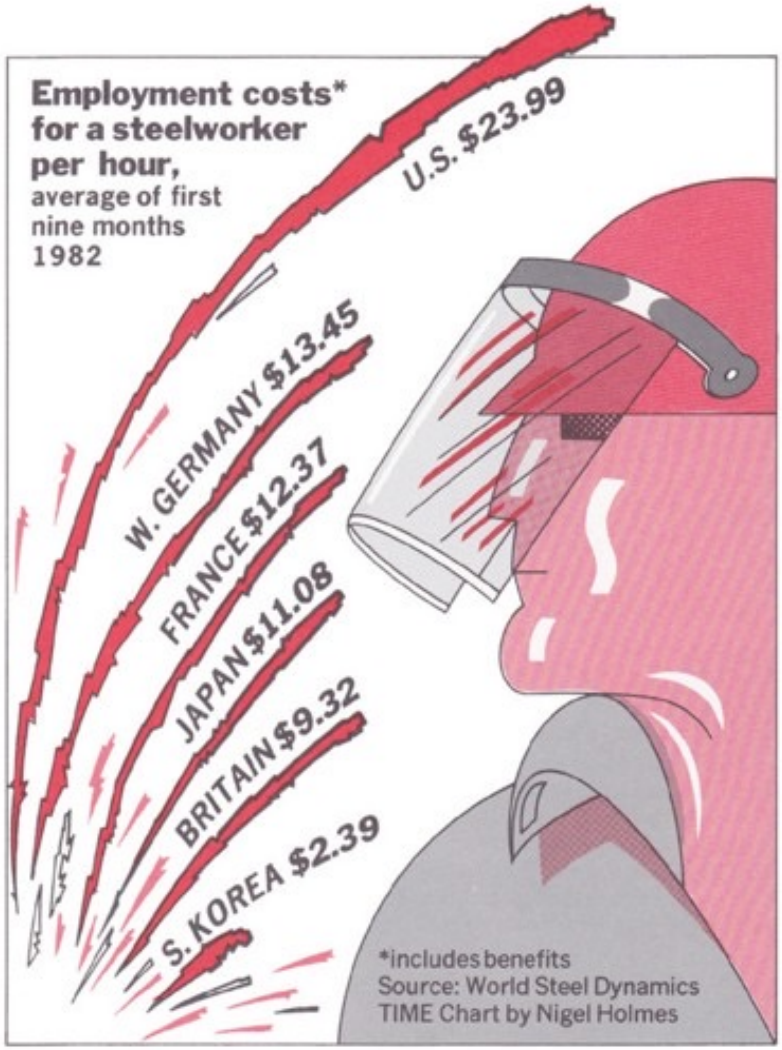
$$\text{Data-Ink Ratio} = \frac{\text{Data-Ink}}{\text{Total "ink" used to print the graphic}}$$

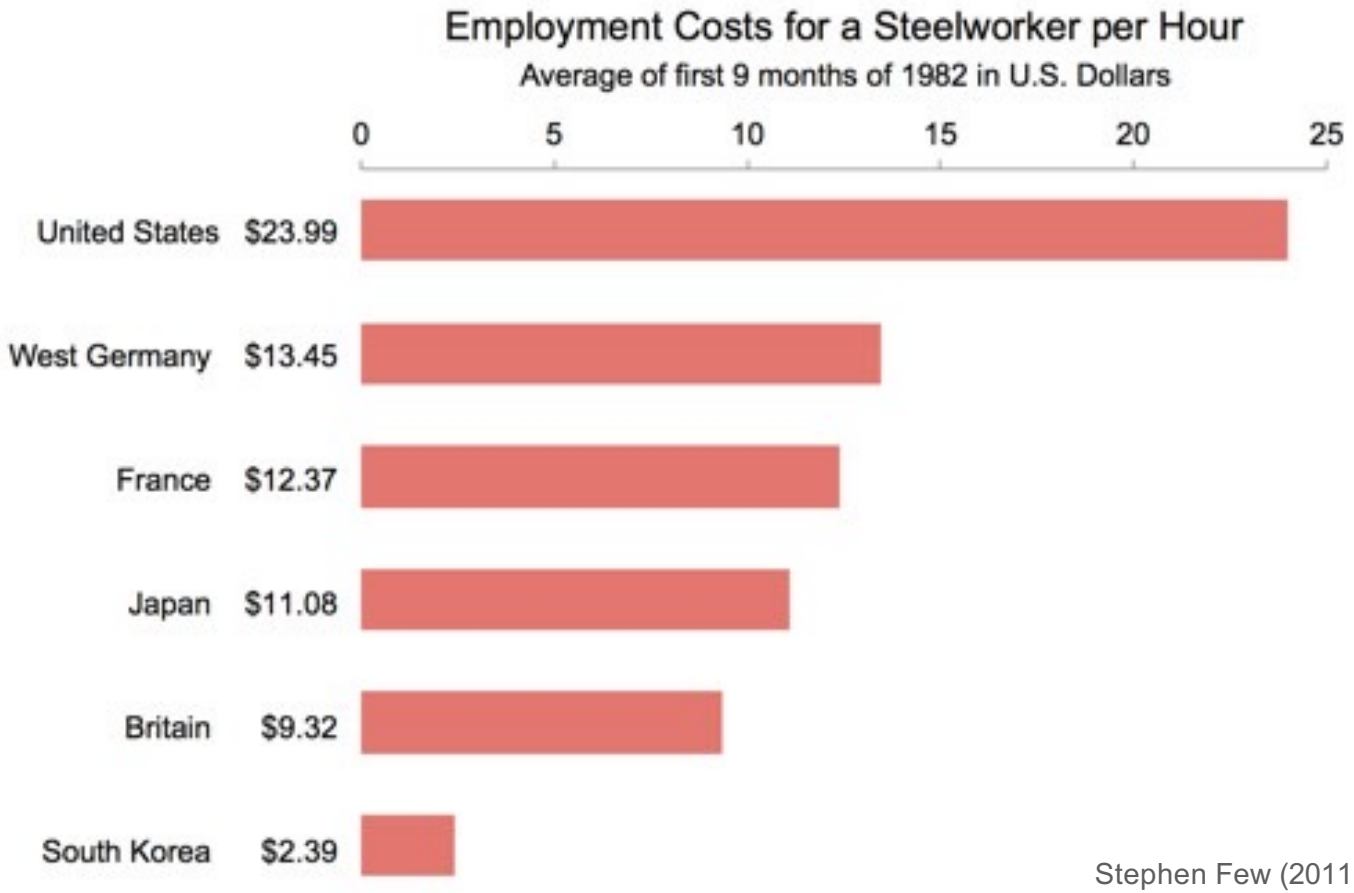
$$= \text{Proportion of a graphic's "ink" devoted to the non-redundant display of data-information}$$

$$= \text{1.0 – proportion of a graphic that can be erased without loss of data-information}$$


700
525
350
175
0
Males
Females
0-$24,999
$25,000+
0-$24,999
$25,000+


700
525
350
175
0
Males
Females
0-$24,999
$25,000+
0-$24,999
$25,000+

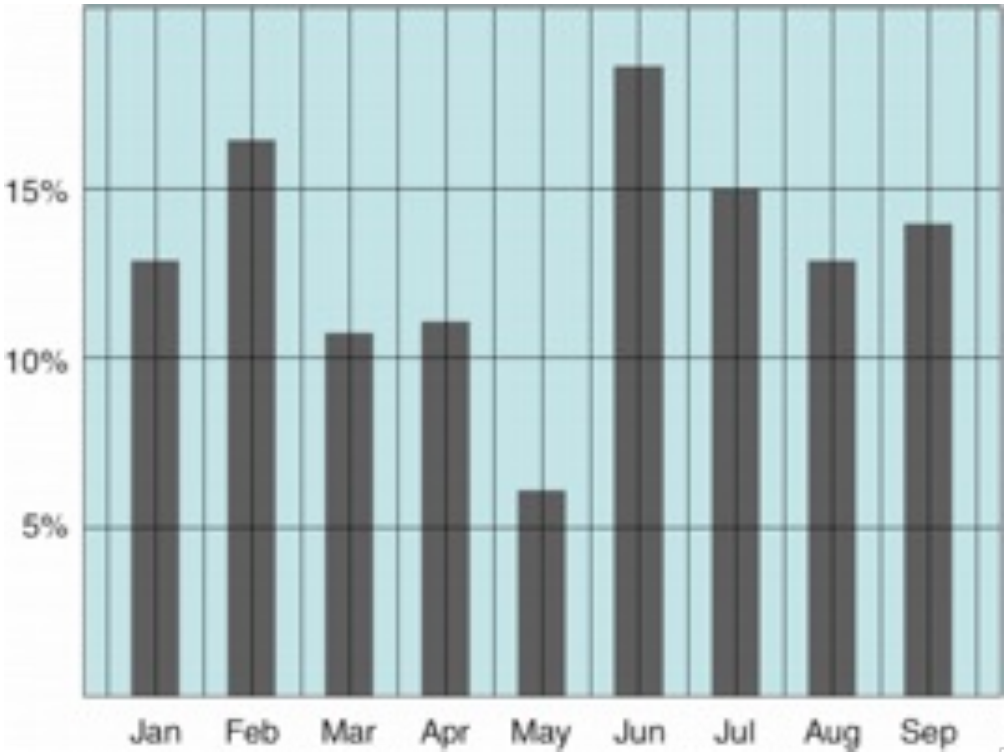

Employment costs* for a steelworker per hour, average of first nine months 1982
U.S. $23.99
W. GERMANY $13.45
FRANCE $12.37
JAPAN $11.08
BRITAIN $9.32
S. KOREA $2.39
*includes benefits
Source: World Steel Dynamics
TIME Chart by Nigel Holmes

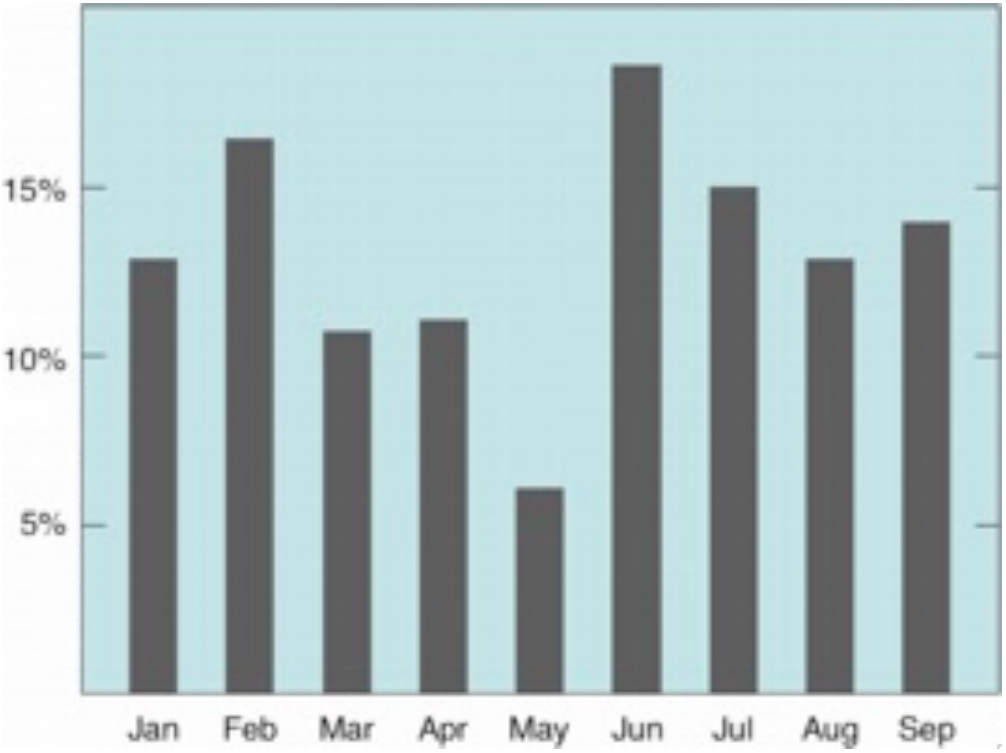


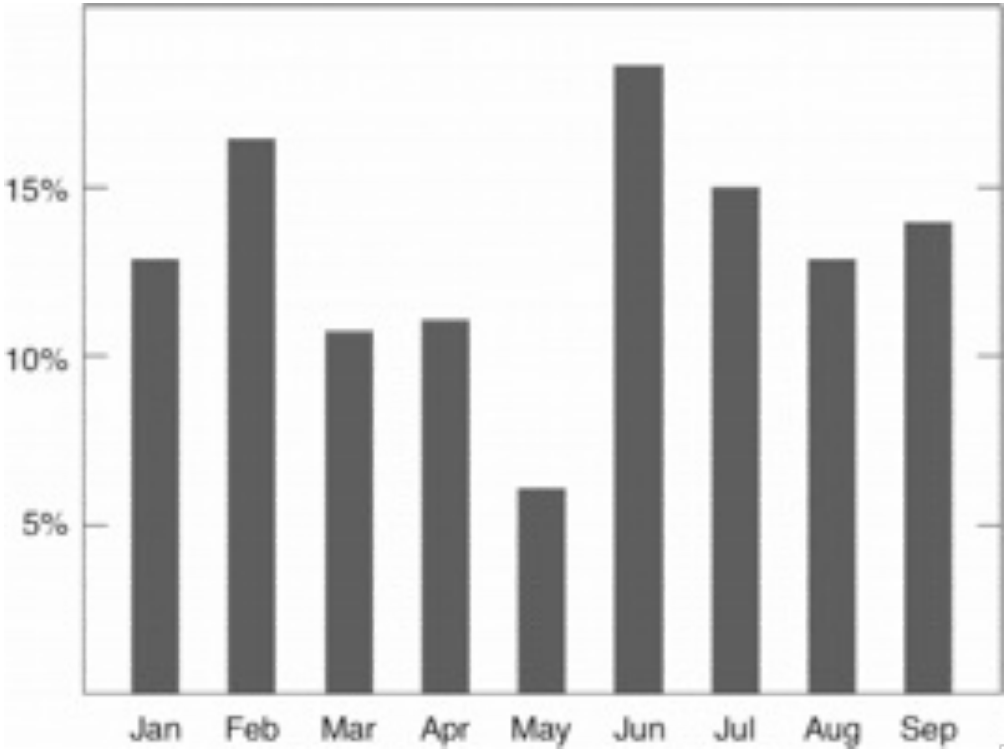
Stephen Few (2011)

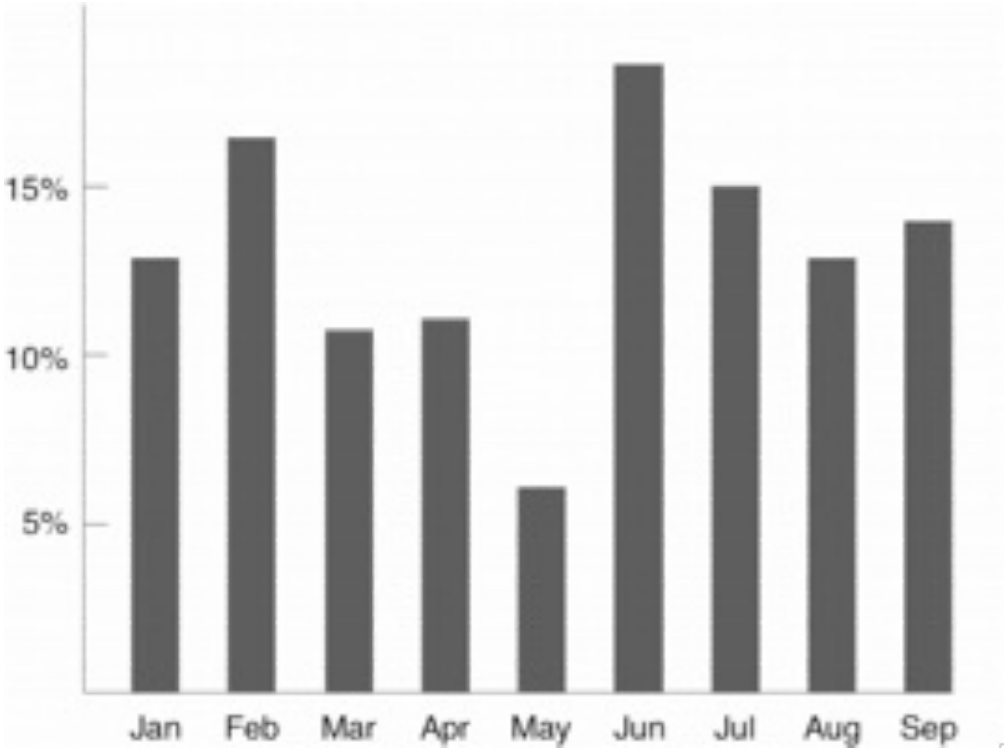
Principles: Avoid chart junks

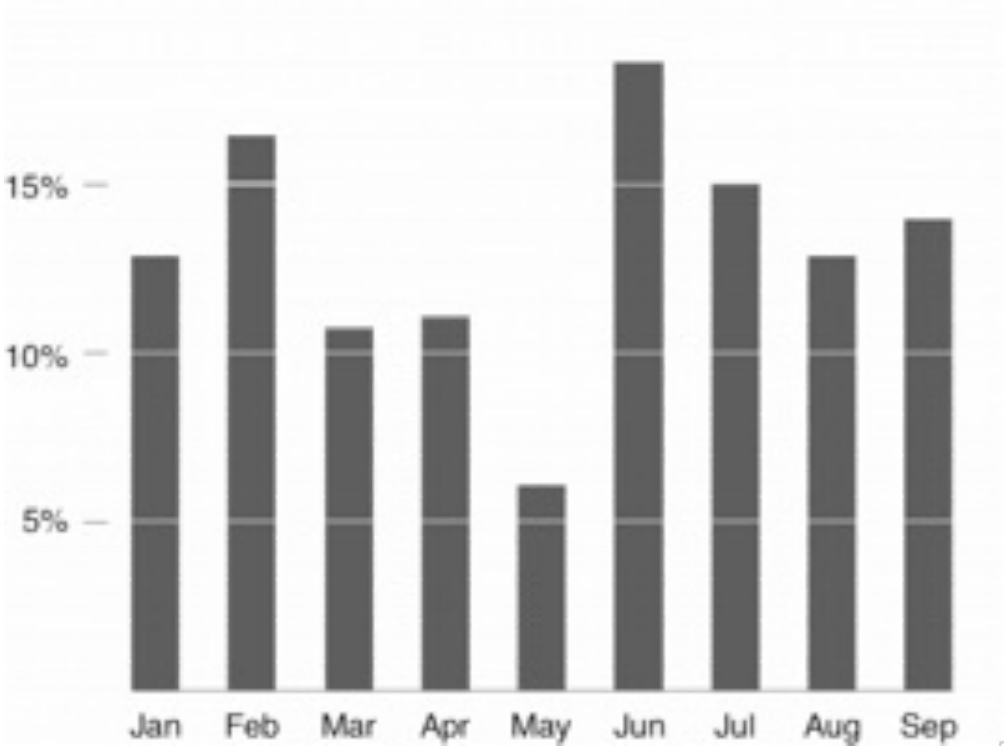
- Unnecessary visual elements in charts that distracts the viewer from the information.

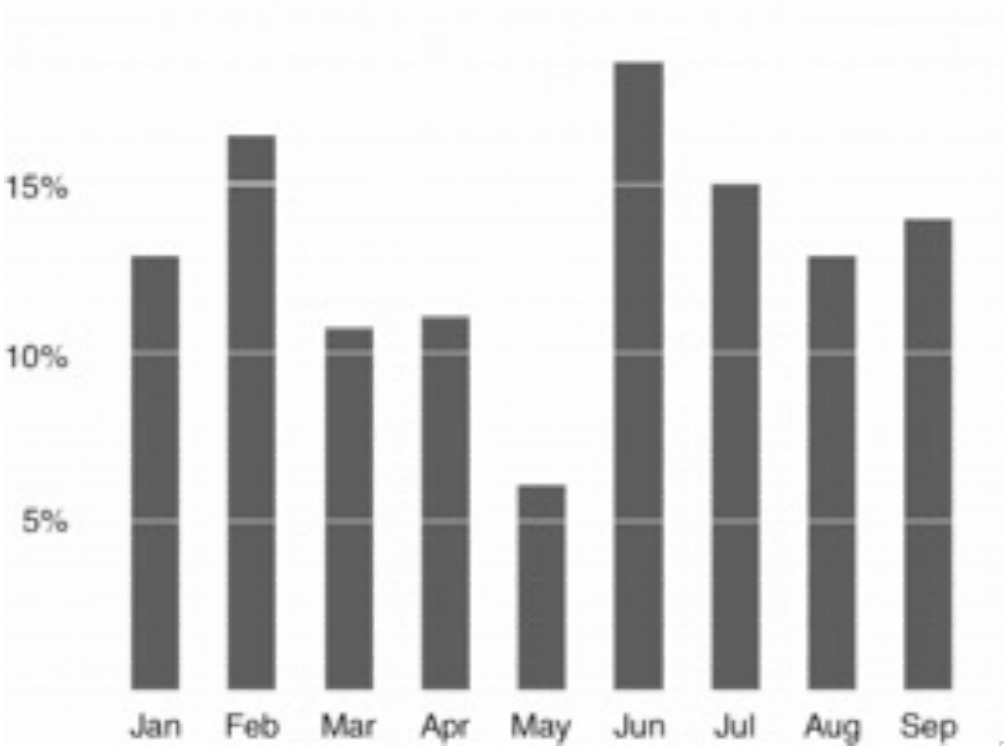

15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep


15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep

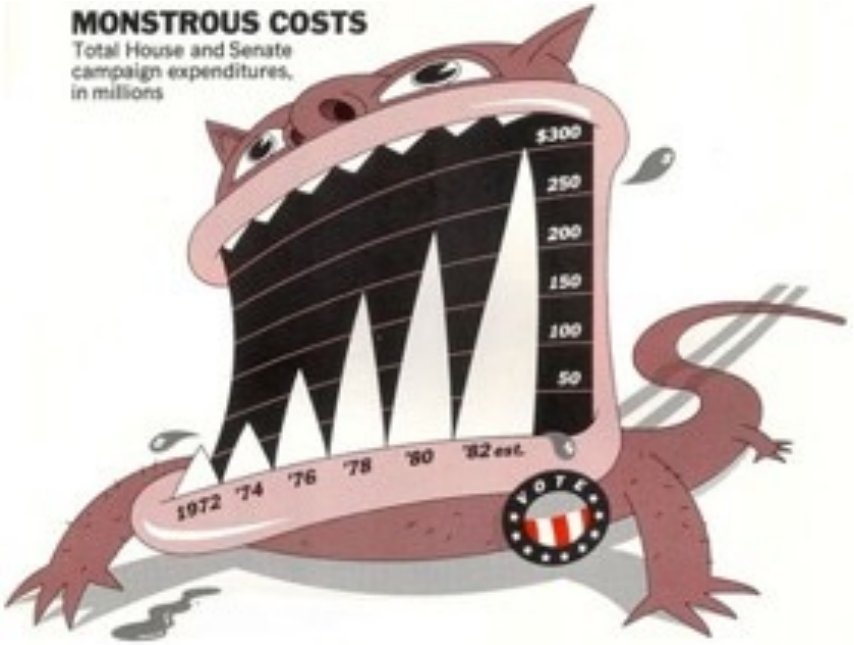

15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep

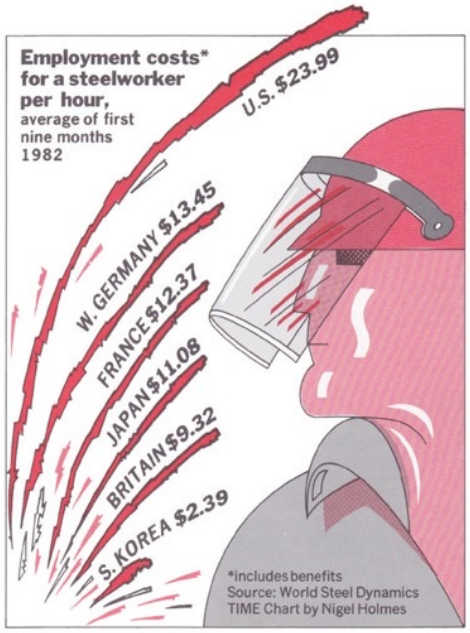

15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep


15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep


15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep

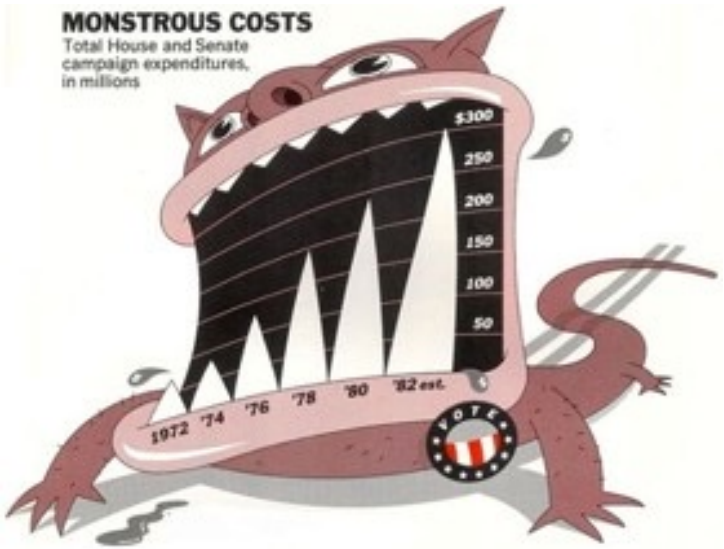
Are these chart junks?

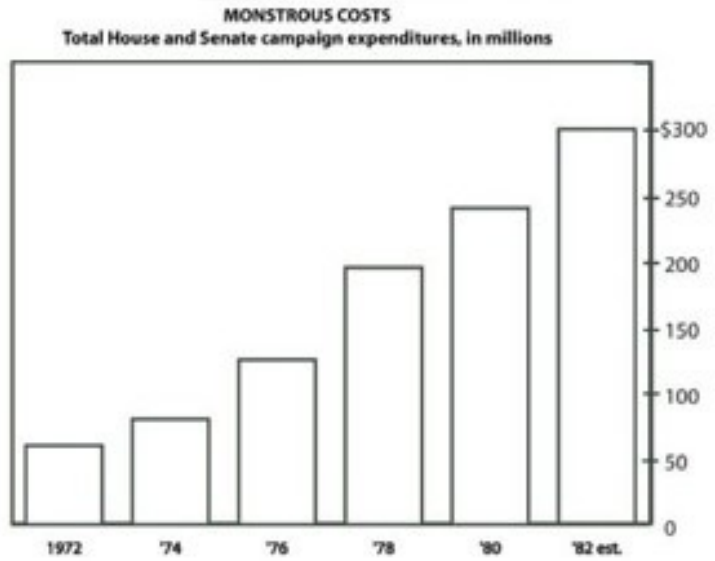




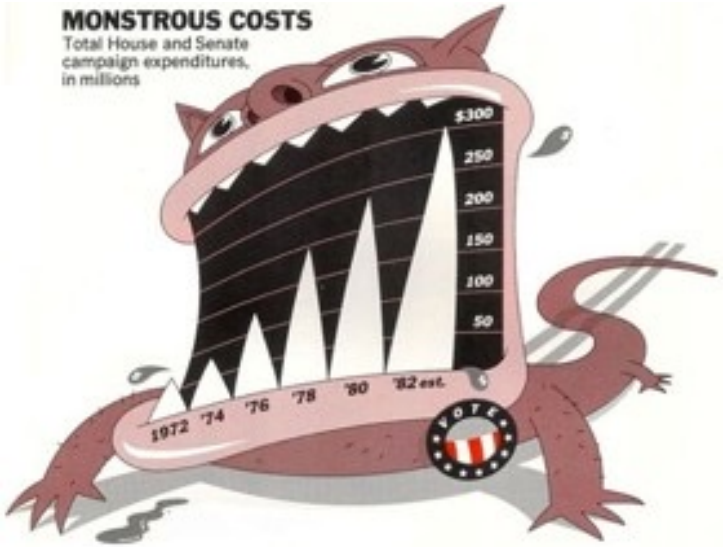
Not all chart junks are the same.

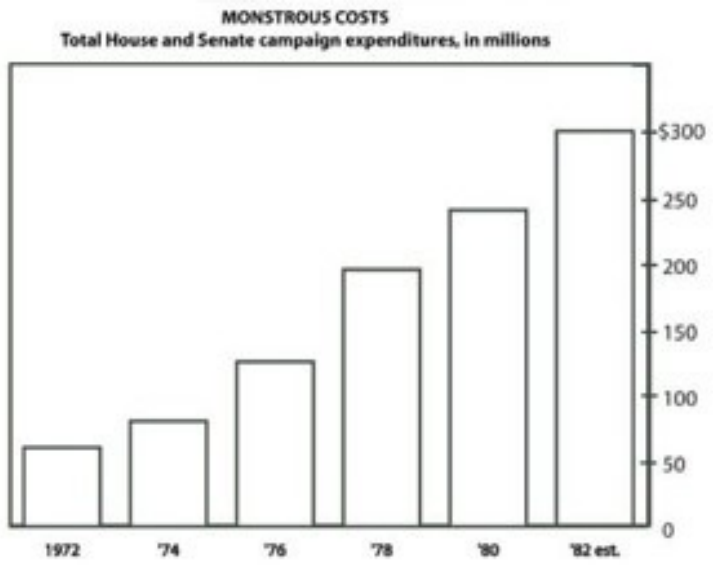
Useful chart junks?





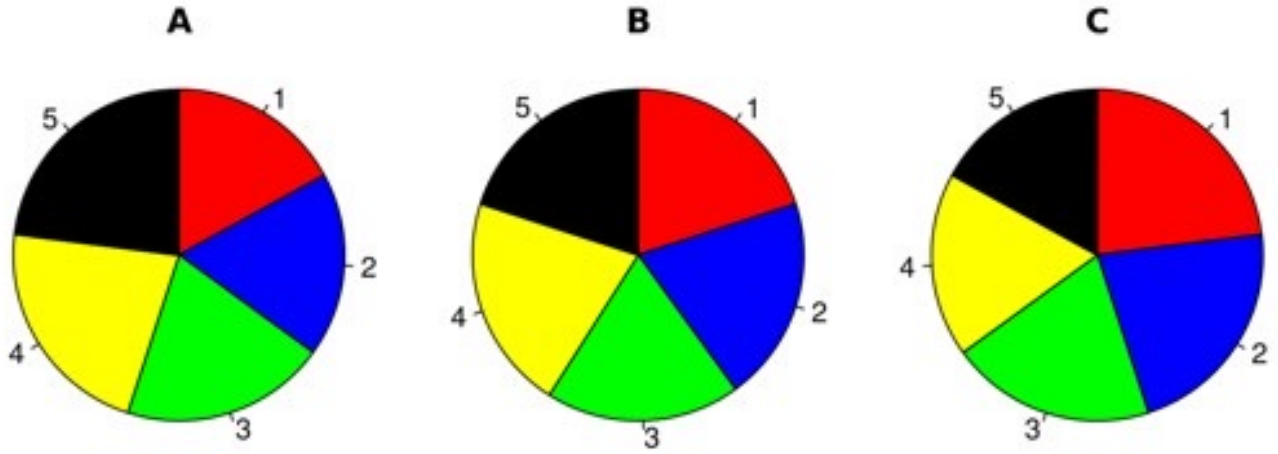
Not harmful in comprehension but more engaging & memorable



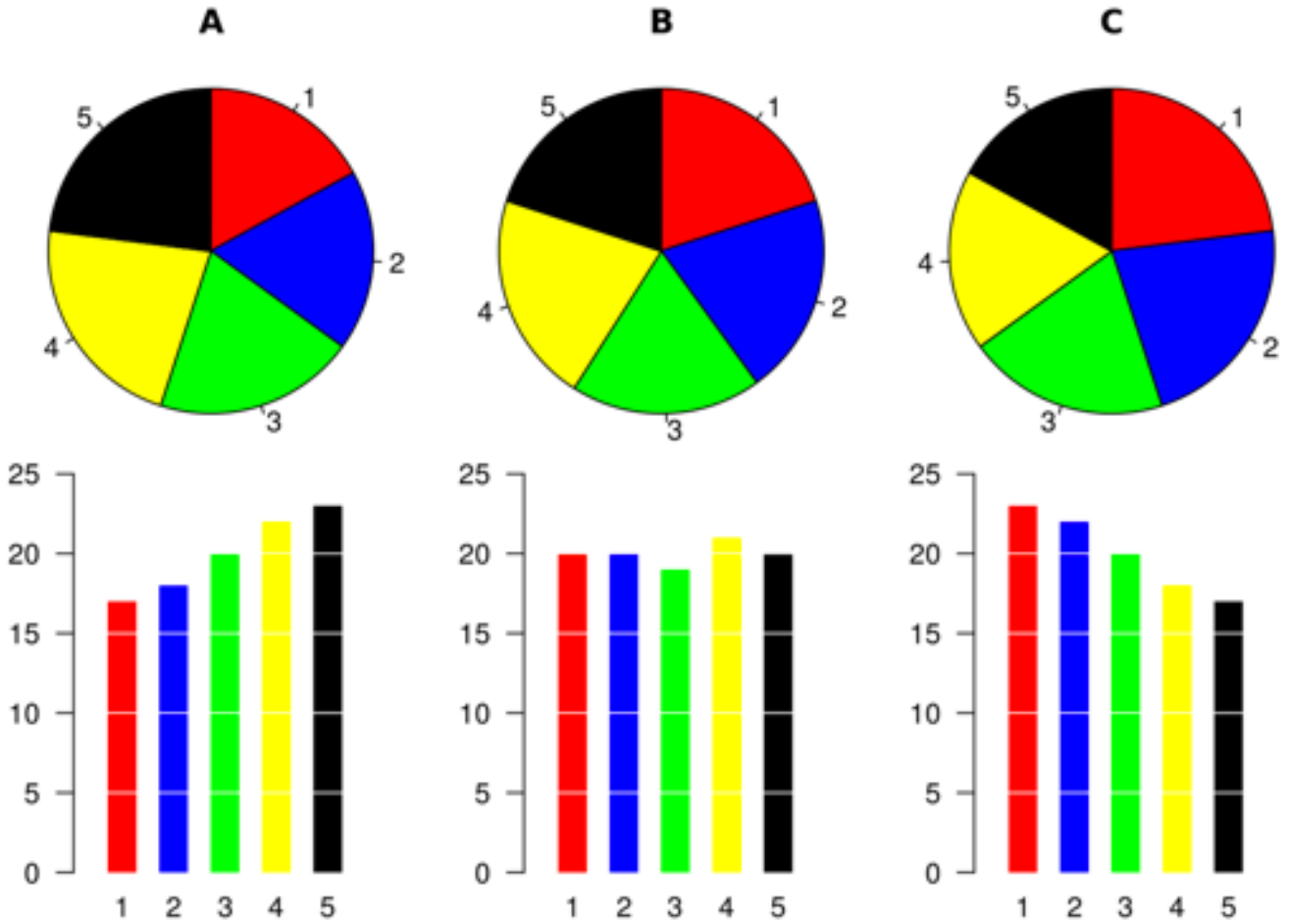


Pie Charts

Challenge: Find the biggest pie slice in each pie chart!

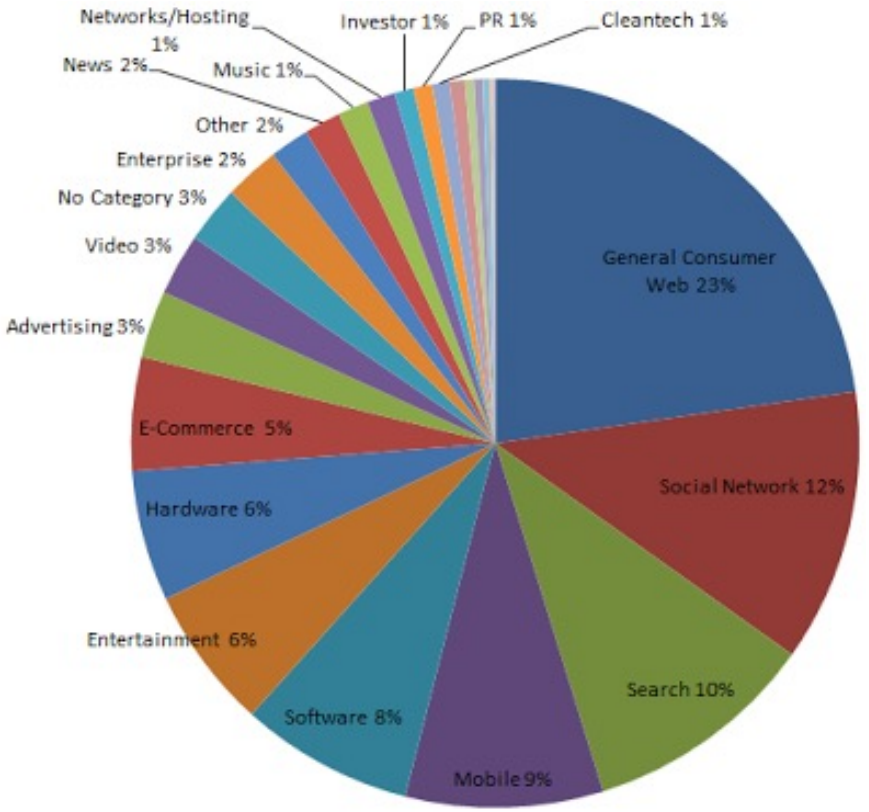


[Schutz 2007]



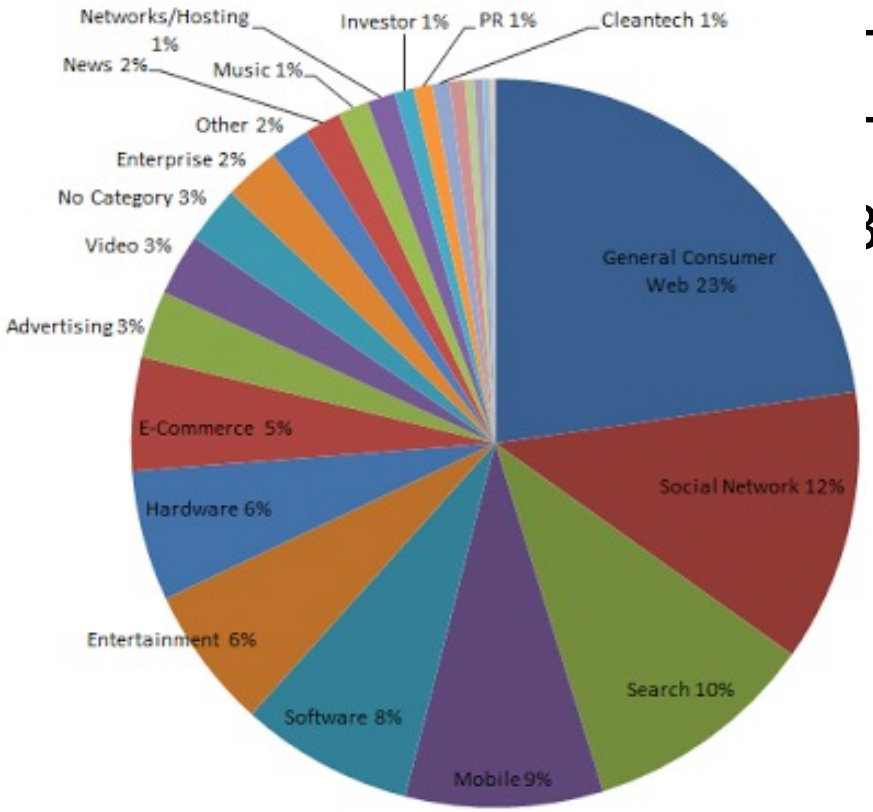
[Schutz 2007]

Share of coverage by topic on TechCrunch



[http://www.storytellingwithdata.com/blog/2011/07/death-to-pie-charts]

Share of coverage by topic on TechCrunch



Too many slices!

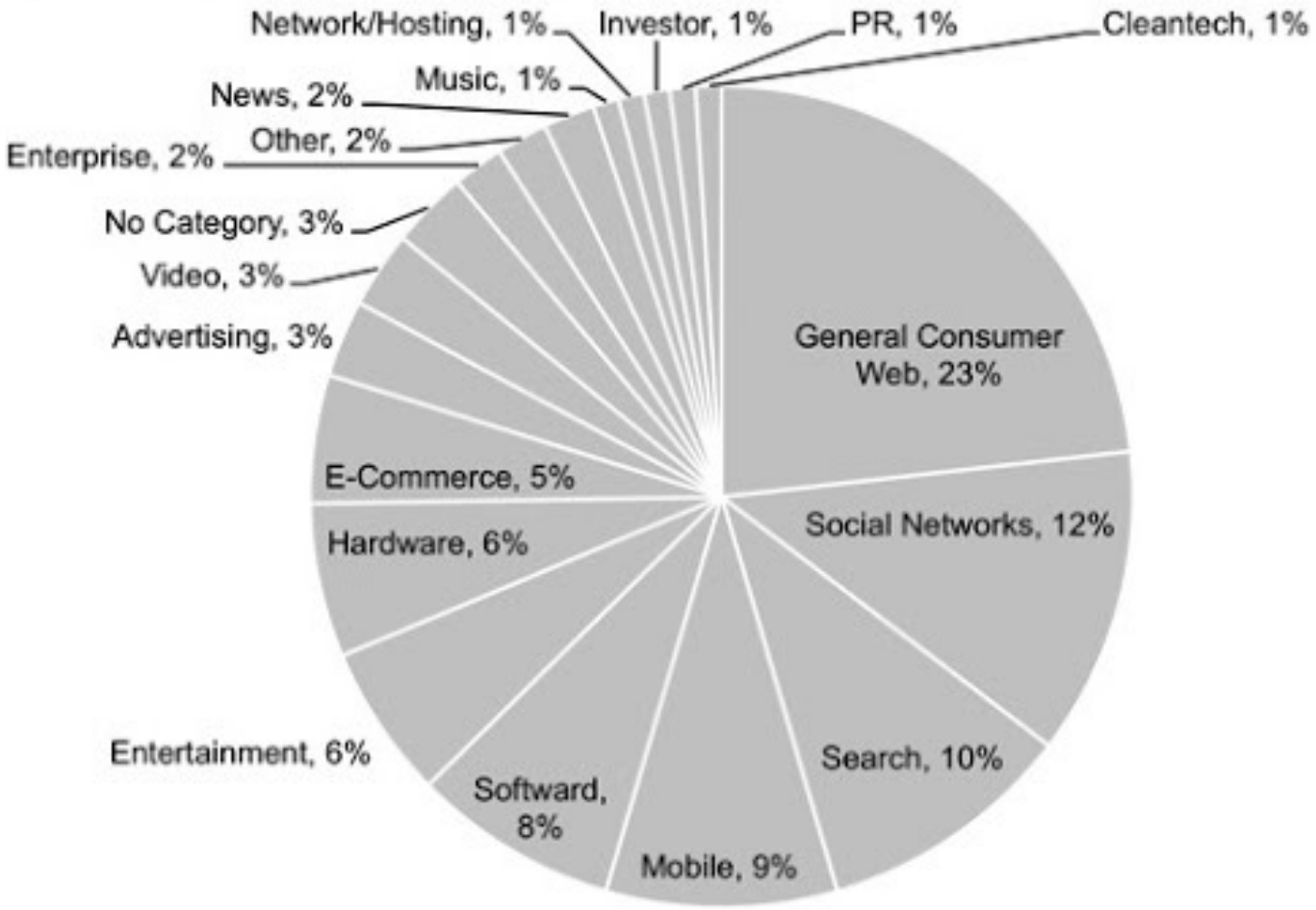
Too many colors!

Bad color contract!

[http://www.storytellingwithdata.com/blog/2011/07/death-to-pie-charts]

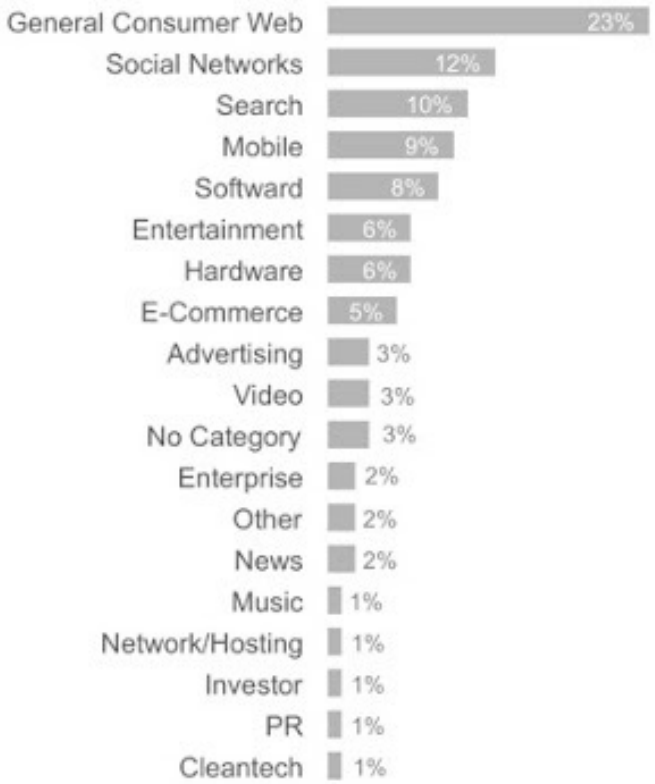
TechCrunch Coverage: 2005 - 2011

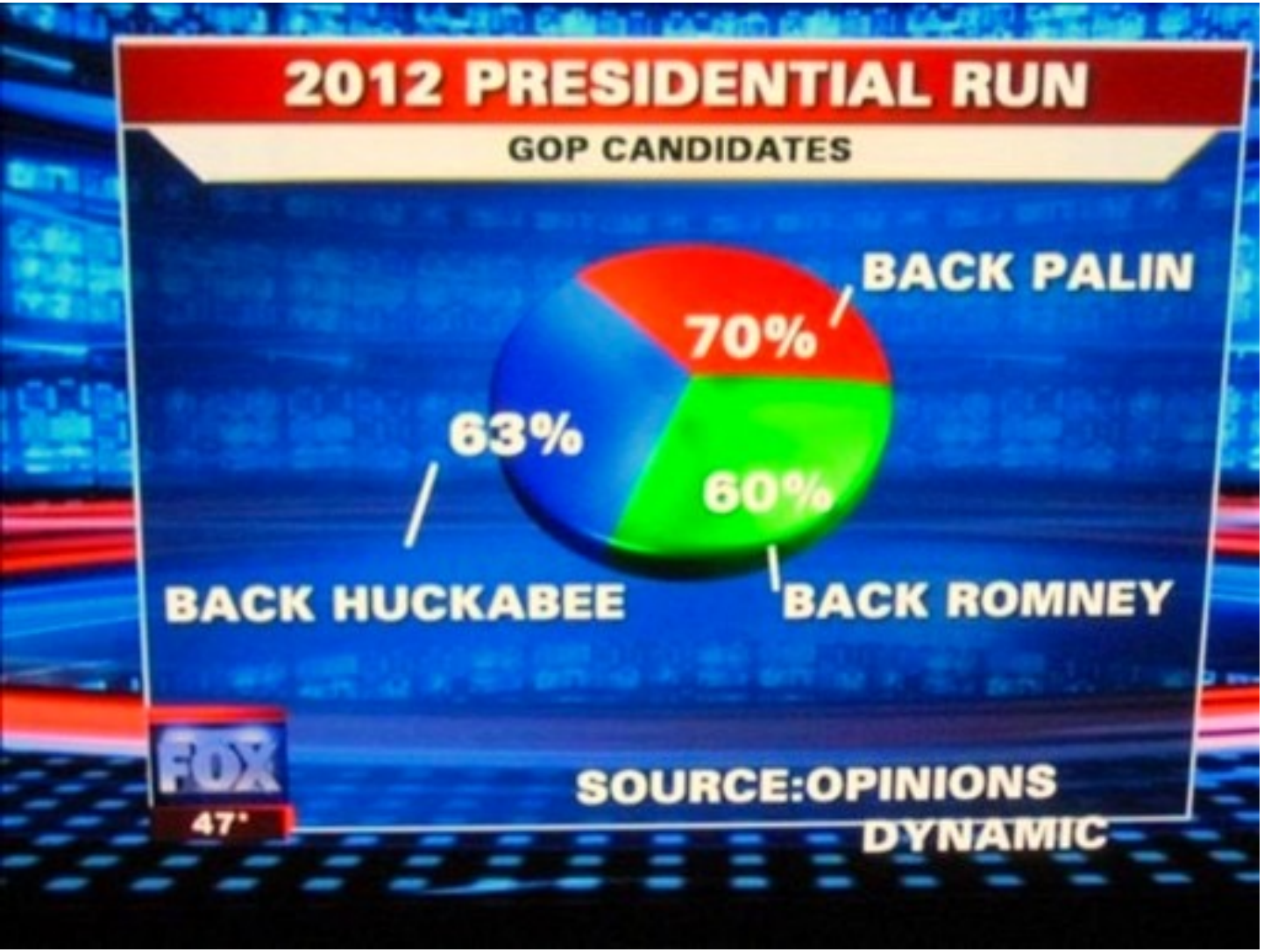
A slightly better pie?



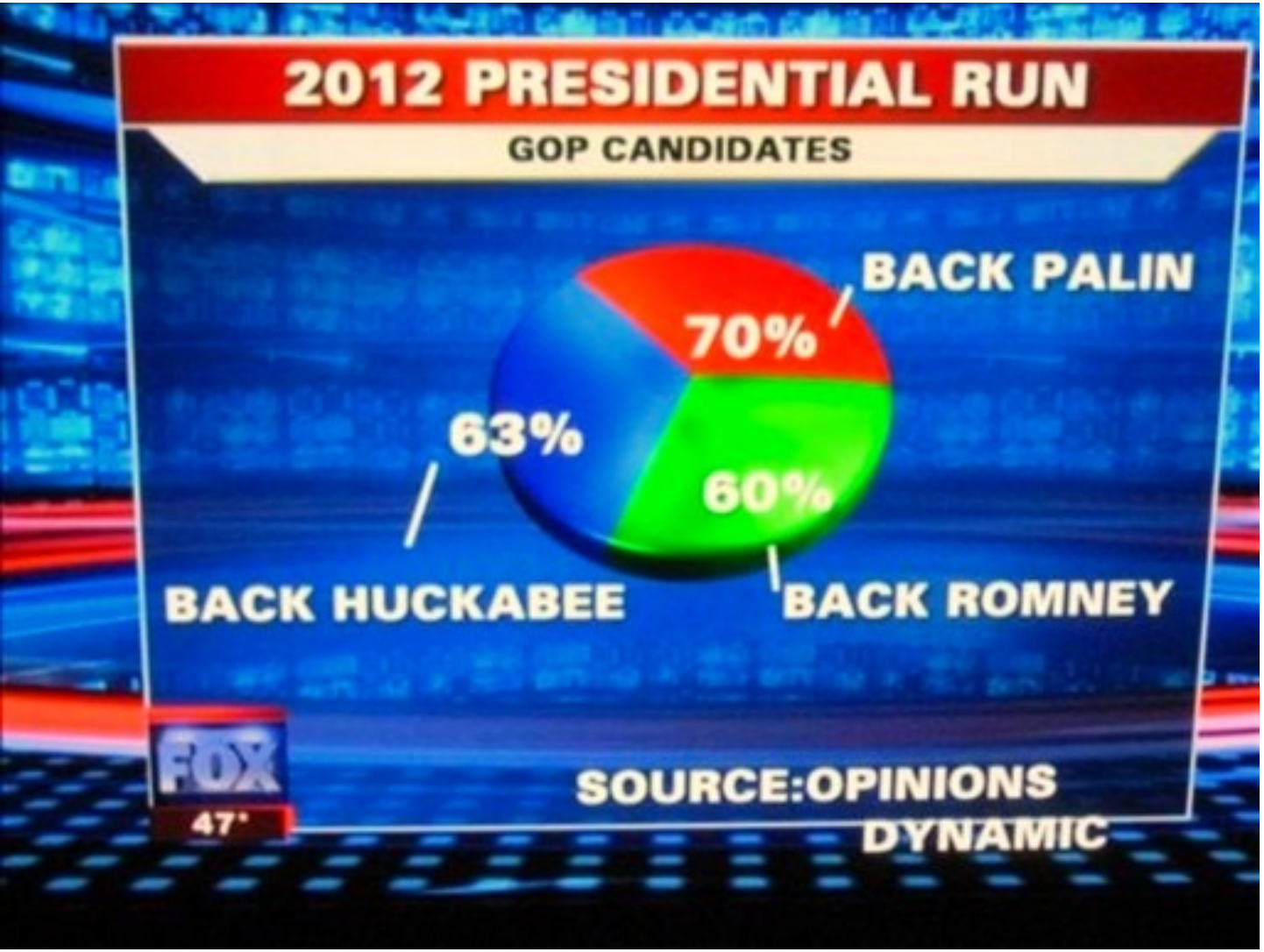
TechCrunch Coverage: 2005 - 2011

Bars are best!




2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
47'
SOURCE:OPINIONS
DYNAMIC

?!


2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
47'
SOURCE:OPINIONS
DYNAMIC

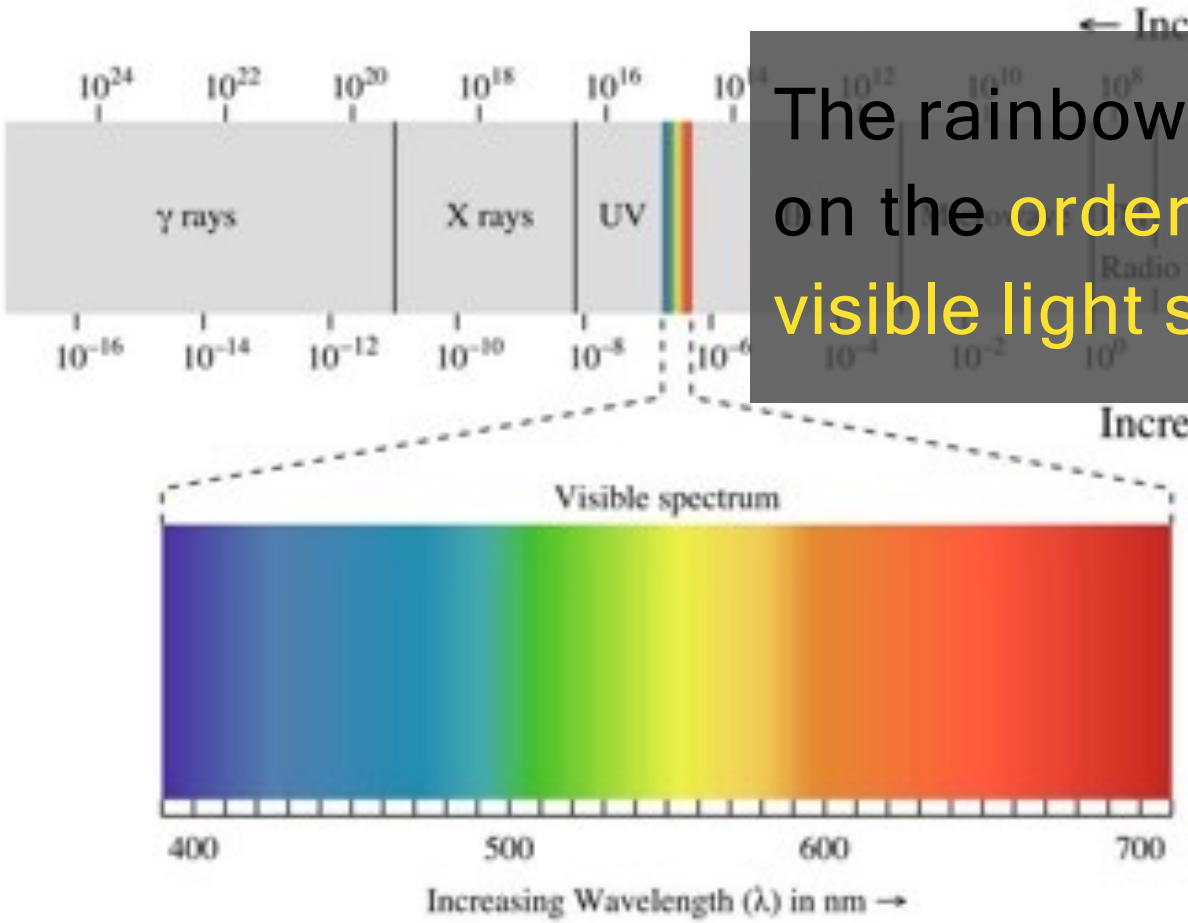
Math fail?!

World's Most Accurate Pie Chart

Some rules of thumb

- You're comparing the parts to the whole.
- There's a small number of slices.
- Sort the values.
- Start at 12 o’clock.
- Use alternatives (e.g., bar chart)
- Pie charts are space efficient though

Rainbow Colormap



The rainbow color map is based on the order of colors in the the visible light spectrum.

[http://factmyth.com/factoids/visible-light-is-electromagnetic-radiation/]

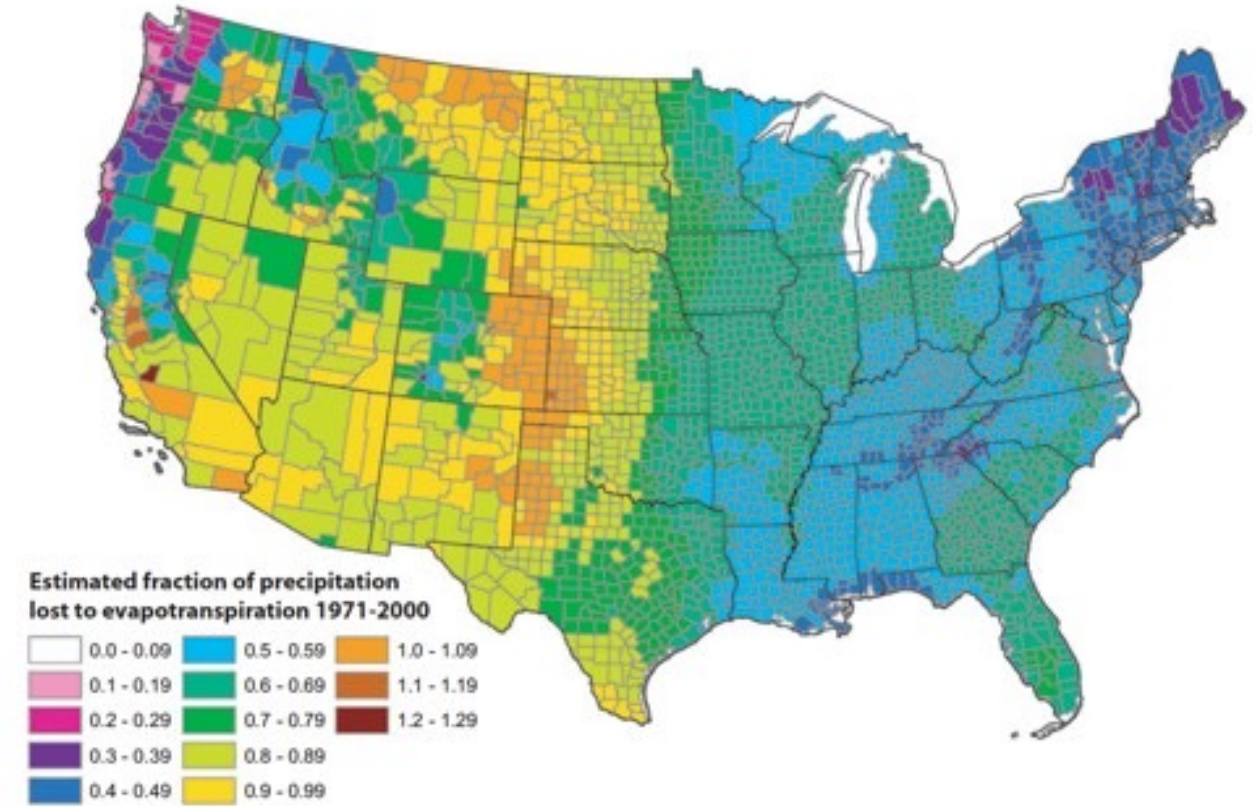
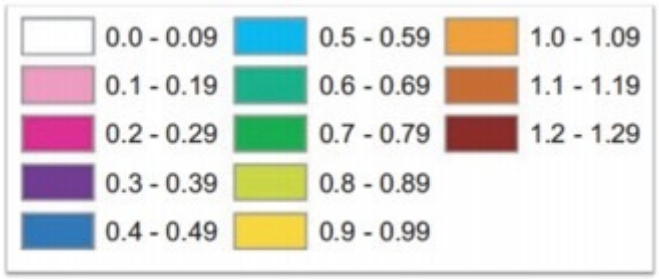
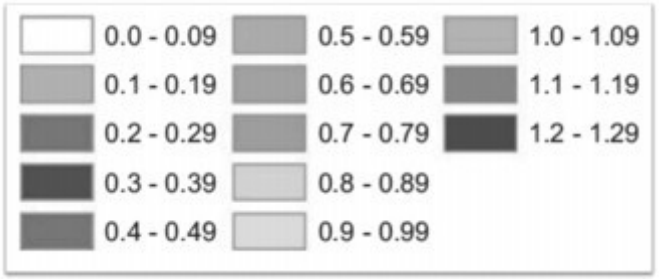


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (*P*) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/*P* were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

Can you say which color represents a higher or lower value group?

Lack of perceptual ordering





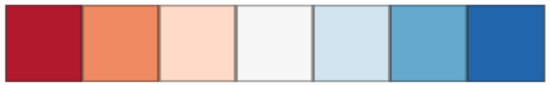
How The Rainbow Color Map Misleads (Kosara 2013)

Some rules of thumb

Qualitative (rainbow) scheme – categorical data.

Sequential scheme – ordered categories or numerical data

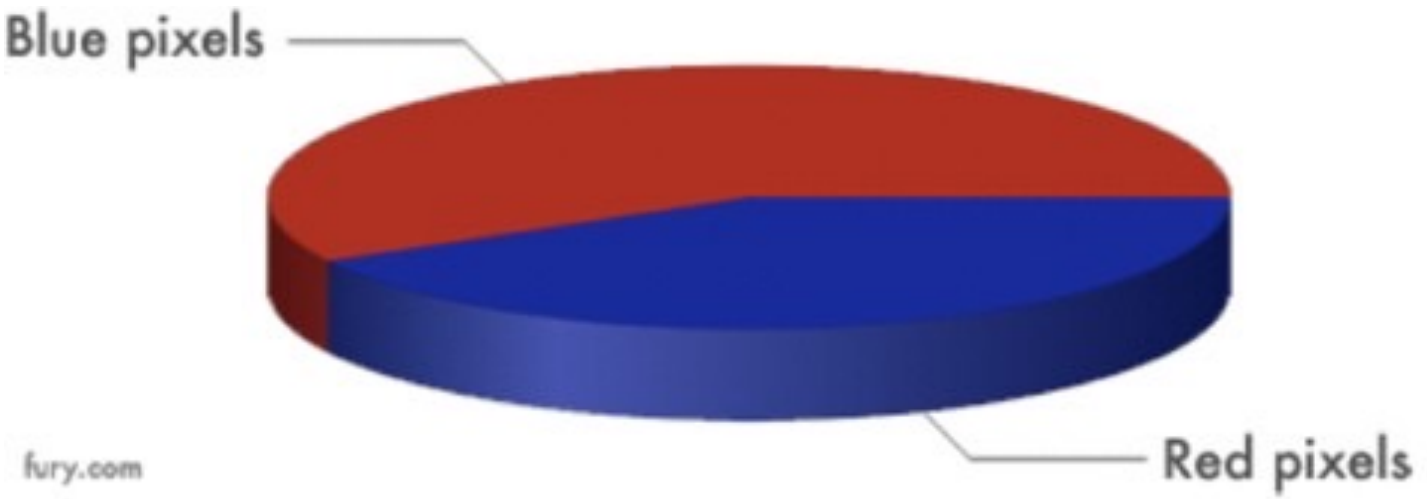
Diverging scheme – numerical data with a meaningful mid-point.



[Color Brewer]

3D charts

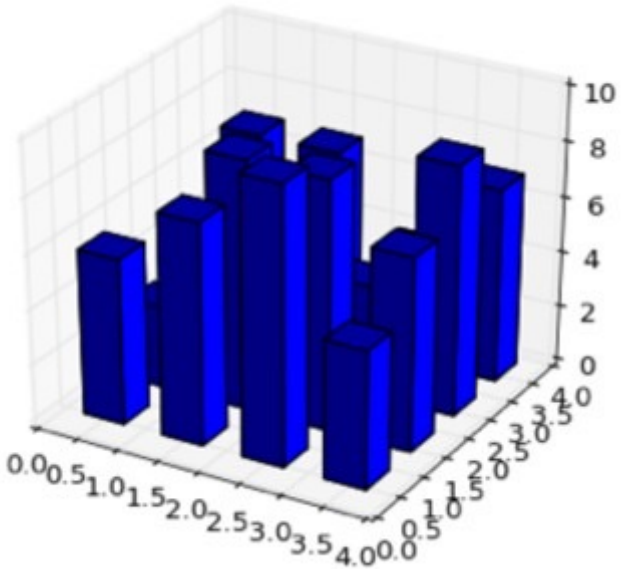
Perspective distorts information

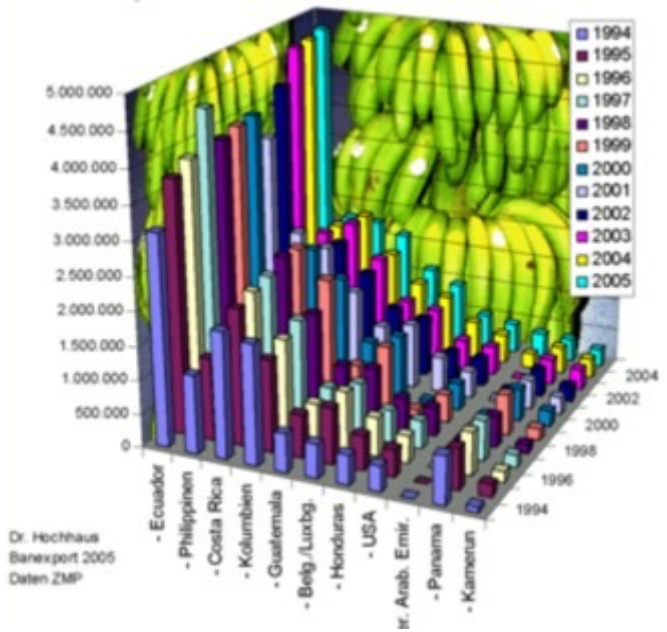


Kevin Fox

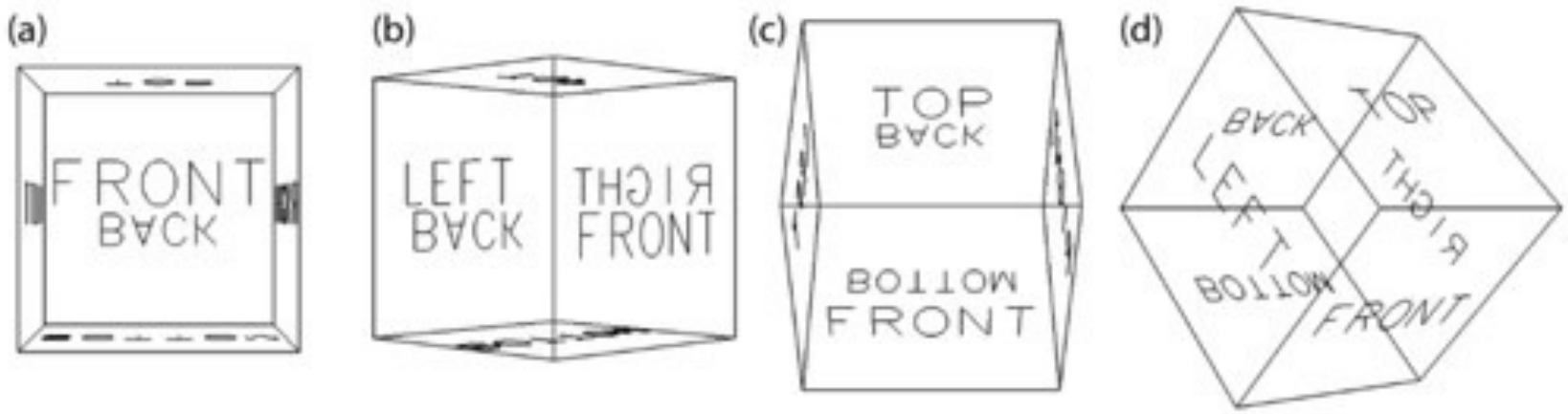
Occlusion hides information

Can rotate, but still no picture at once.

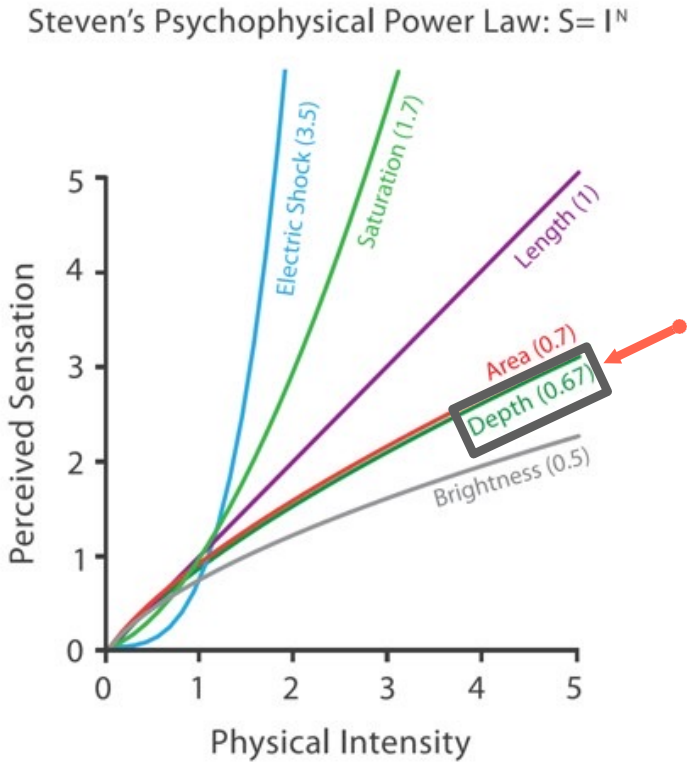




Tilted text isn't legible



Depth judgment is bad



Human underestimate depth!

Actual intensity change

vs

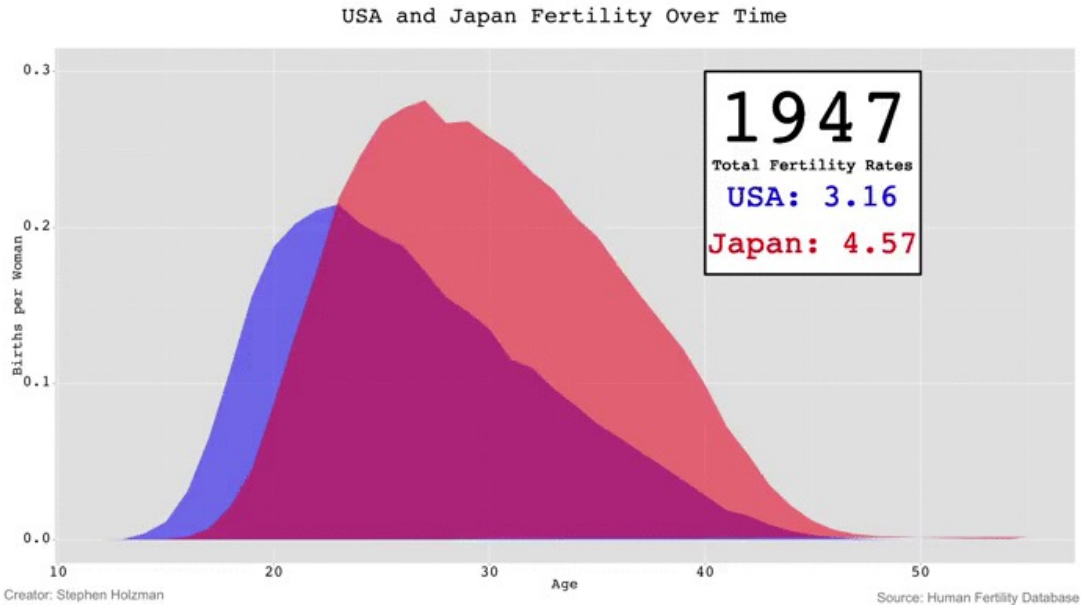
Perceived Sensation

External cognition vs Internal cognition

Eyes beats memory

Animation

Harder to compare visible item to memory of what you saw

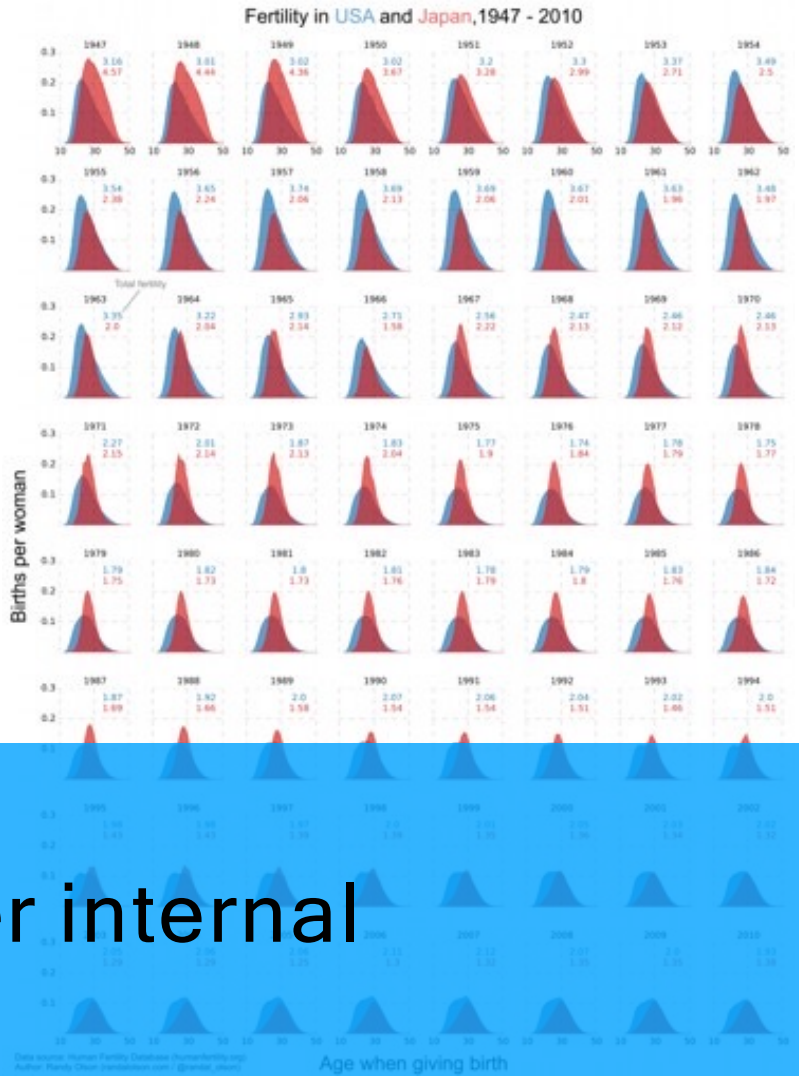


http://www.randalolson.com/2015/08/23/small-multiples-vs-animated-gifs-for-showing-changes-in-fertility-rates-over-time/

Eyes beats memory

Small multiples

Easy to compare by moving eyes between side-by-side views



Use external cognition over internal cognition

Subjective Dimensions

- Aesthetics: Attractive things are perceived as more useful.
- Style: Communicates brand, process, who the designer is.
- Playfulness: Encourages experimentation and exploration.
- Vividness: Can make a visualization more memorable.