Data Visualization
The Good, the Bad, the Weird

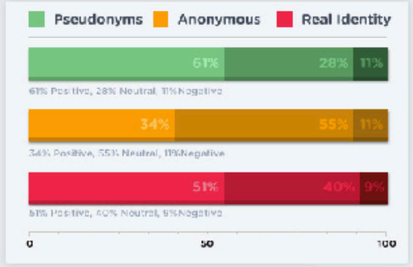
Nam Wook Kim

Mini-Courses — January @ GSAS
2018

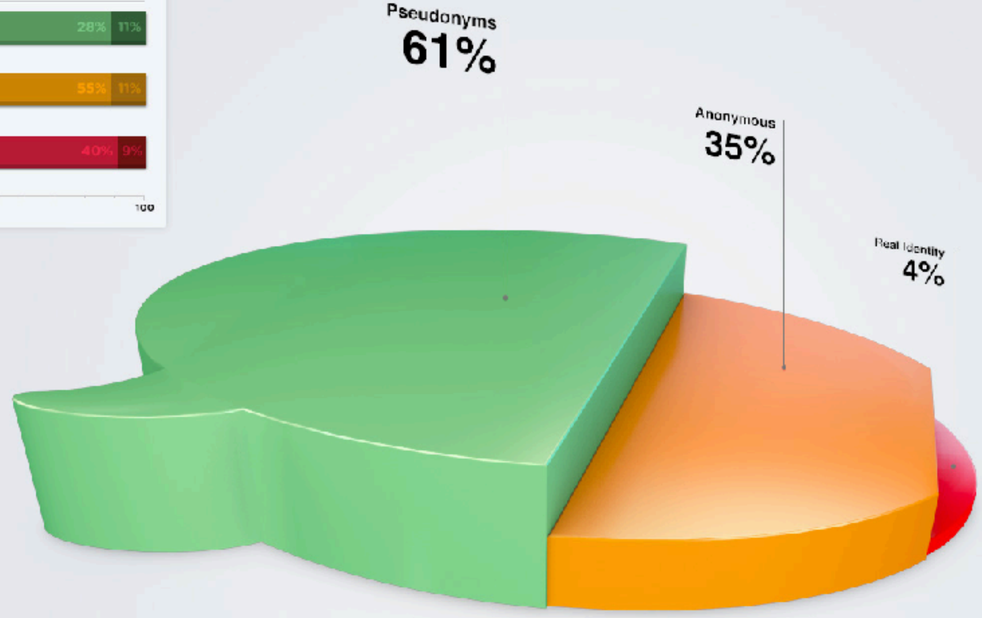
Goal

Rules of thumbs
to critique visualization design

Quality Signals by Identity



Percentage of Comments by Identity



Average Comments Per User By Identity

The average commenter using a pseudonym contributed **6.5** times more than anonymous commenters and **4.7** times more than commenters identifying with **Facebook**.

May 3, 2008

SIGN IN TO E MAIL OR SAVE THIS | FEEDBACK

All of Inflation's Little Parts

Each month, the Bureau of Labor Statistics gathers 84,000 prices in about 200 categories — like gasoline, bananas, dresses and garbage collection — to form the Consumer Price Index, one measure of inflation.

It's among the statistics that the Federal Reserve considered when it cut interest rates on Wednesday. The categories are weighted according to an estimate of what the average American spends, as shown below.

An Average Consumer's Spending

Each shape below represents how much the average American spends in different categories. Larger shapes make up a larger part of spending.

Color shows change in prices from March 2007 to March 2008

-10% -2 0 +2 +4 +8 +10 +20 +40%

ZOOM IN ZOOM OUT

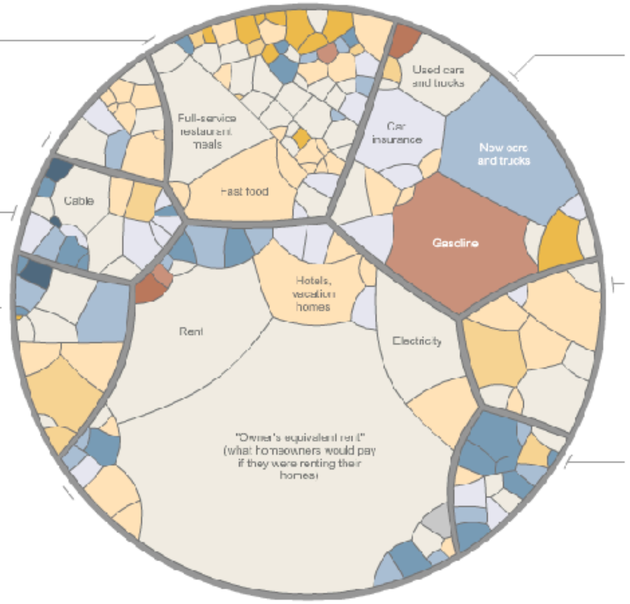
Food and beverages 15%
The high price of oil is a factor that has made food prices rise quickly.

Miscellaneous 3%

Recreation 6%

Education/Communication 6%
Cellphones were added to the index in 1997. Because the Consumer Price Index can be slow to add new goods, which are often cheaper, it may overstate parts of inflation.

Housing 42%
In the C.P.I., home ownership costs track rent prices more closely than housing prices. This means inflation may have been understated when home prices were rising faster than rents.



Transportation 18%
Gas is 5.2 percent of spending nationwide, but only 3.8 percent in the New York area.

Health care 6%
As a group, the elderly spend about twice as much of their budget on medical care.

Apparel 4%
The ratio of spending on women's clothes to that on men's clothes is about 2 to1

Sources: Bureau of Labor Statistics; Michael Balzer, University of Konstanz (Germany)

Matthew Bloch, Shan Carter and Amanda Cox/The New York Times

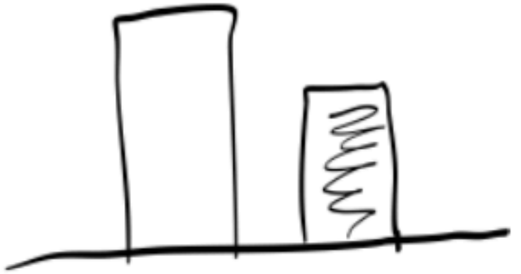
Activity

CS 171

Create at least three sketches to visualize these two quantities. (1 min)

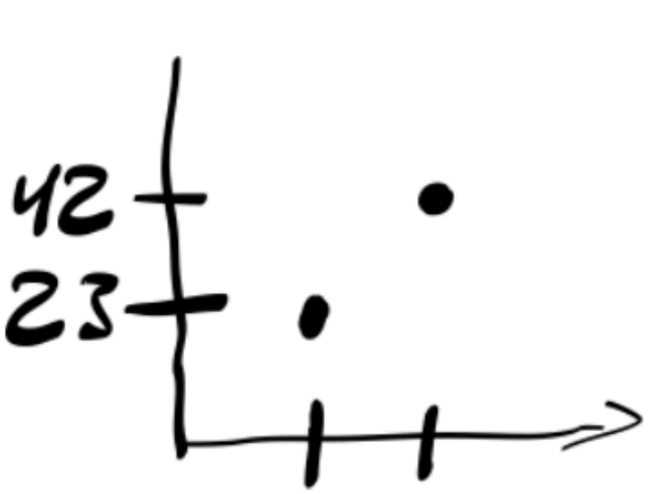
Most likely results

Pie Chart



Bar Chart

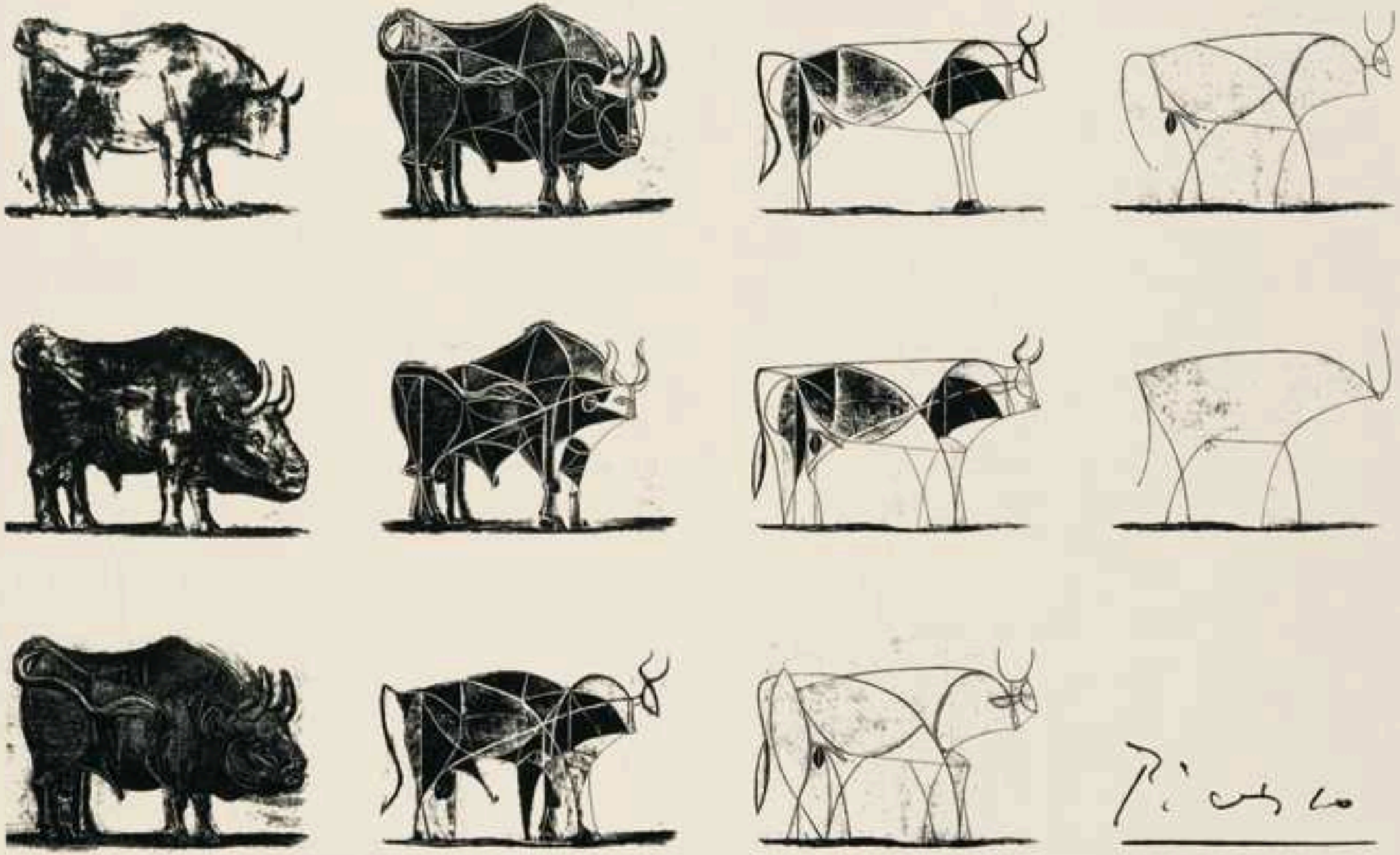
Scatterplot

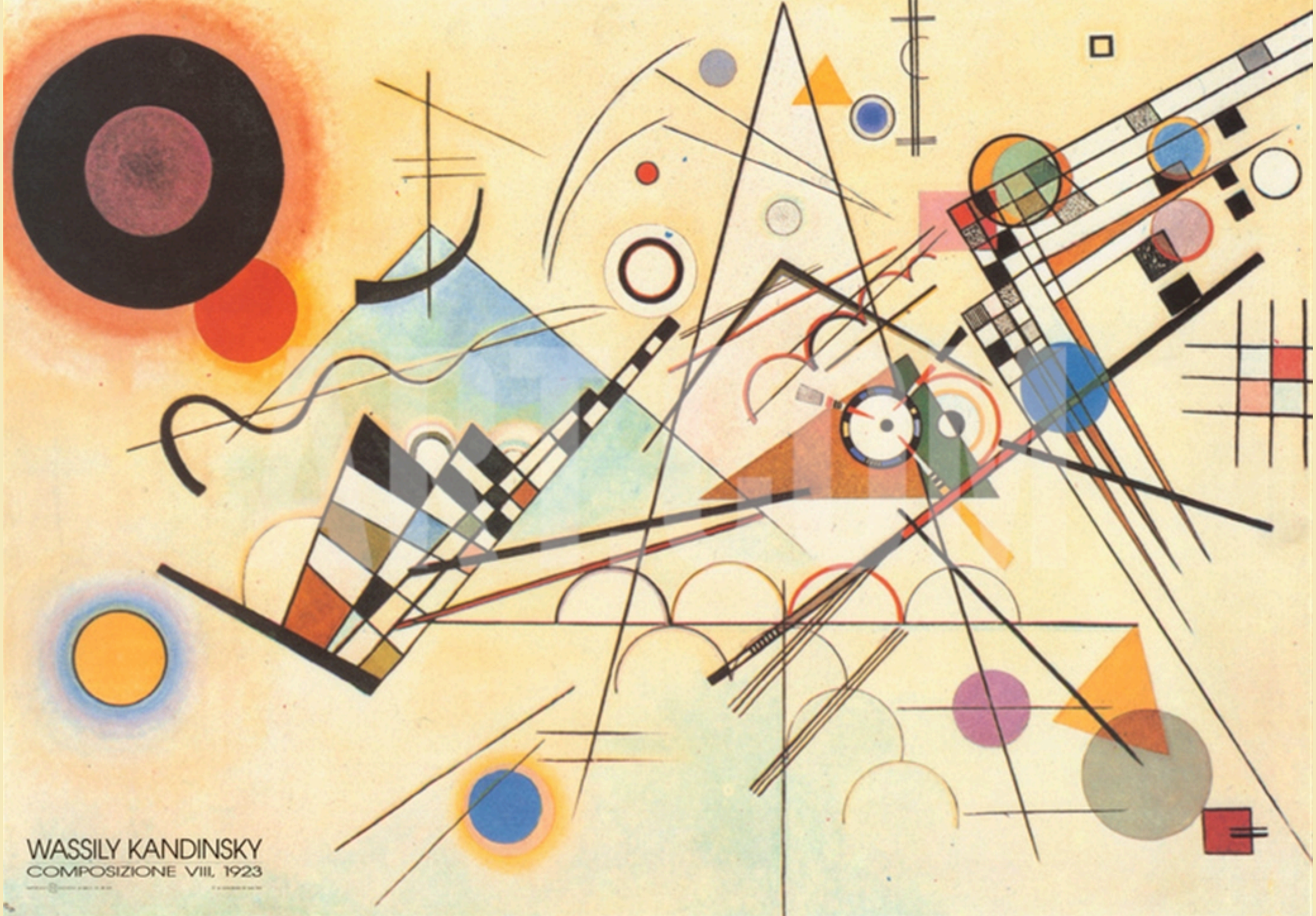


23

42

Arabic Numbers



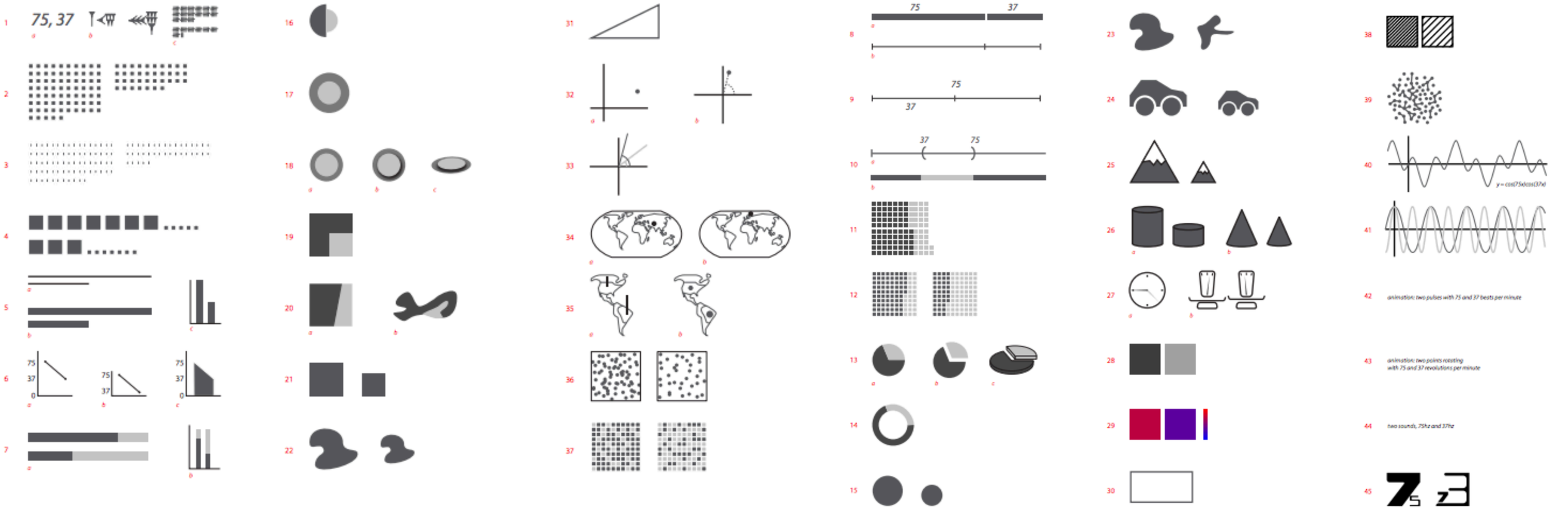

WASSILY KANDINSKY
COMPOSIZIONE VIII, 1923

Activity

CS 171

Create at least three new visualizations that are different from your previous ideas using a constraint - e.g., one line, only black/white, only round objects,...
(2 mins)

45 Ways to Communicate Two Quantities



[Santiago Ortiz 2012]

Which one is the best and why?

There are so many ways to draw just two numbers…

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

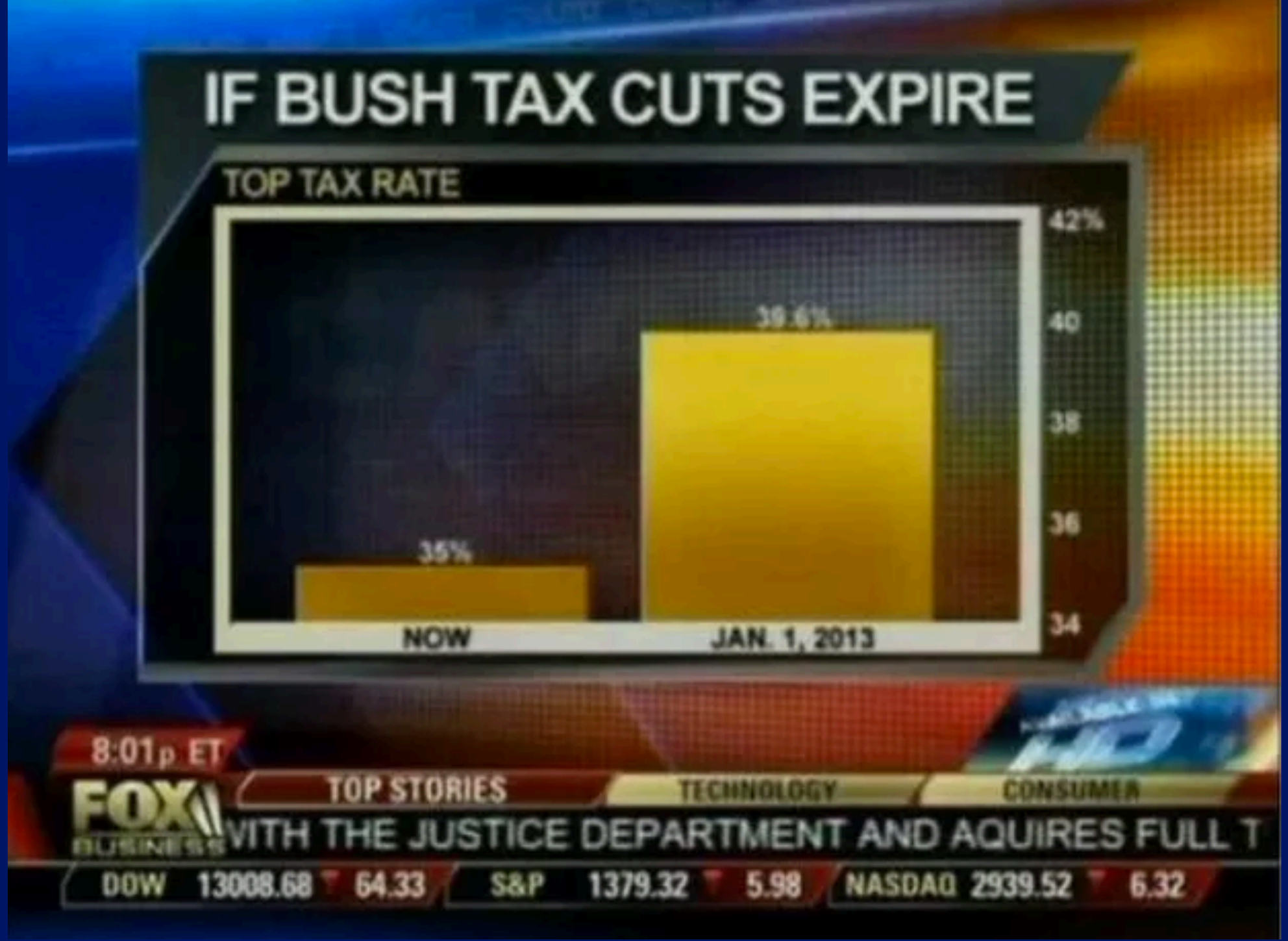
Questions to answer

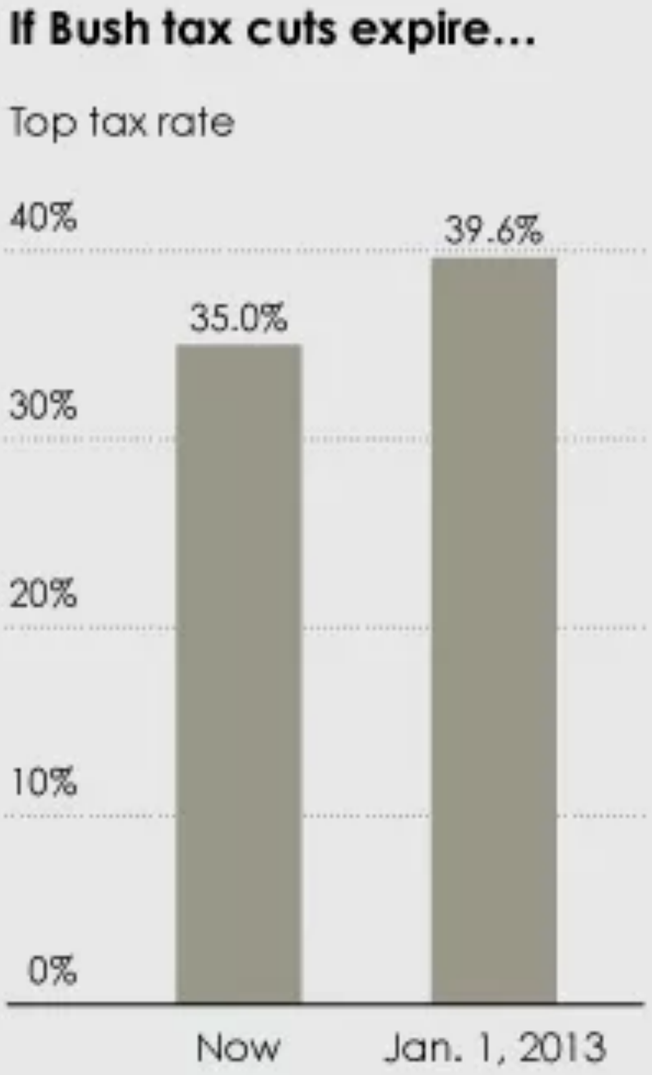
Messages to deliver

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

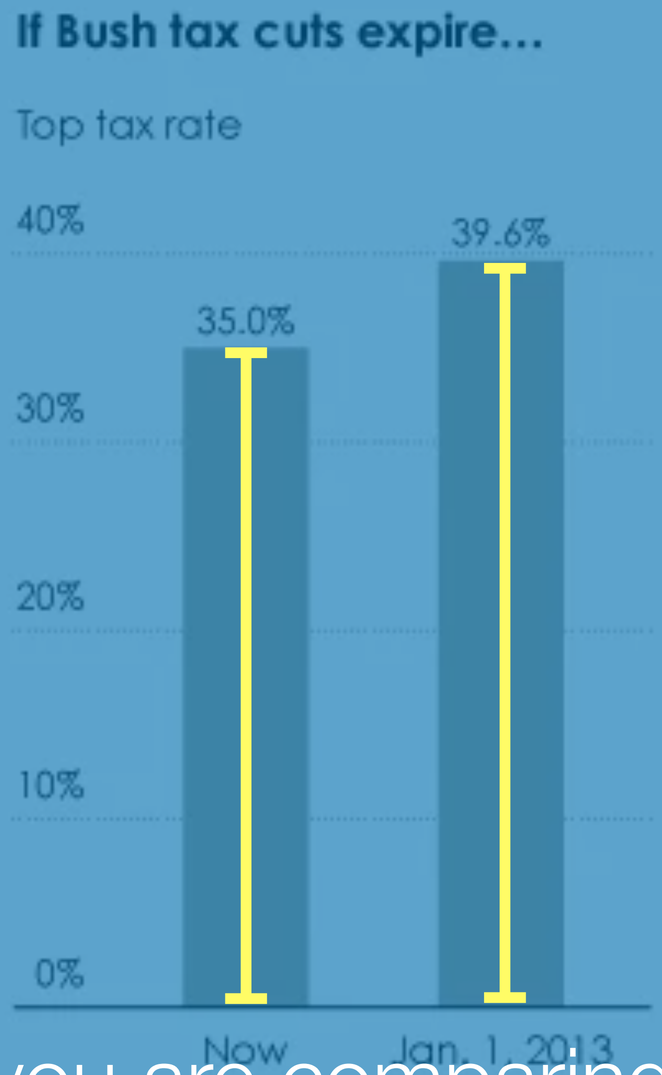
Edward Tufte's Design Principles

Graphical Integrity

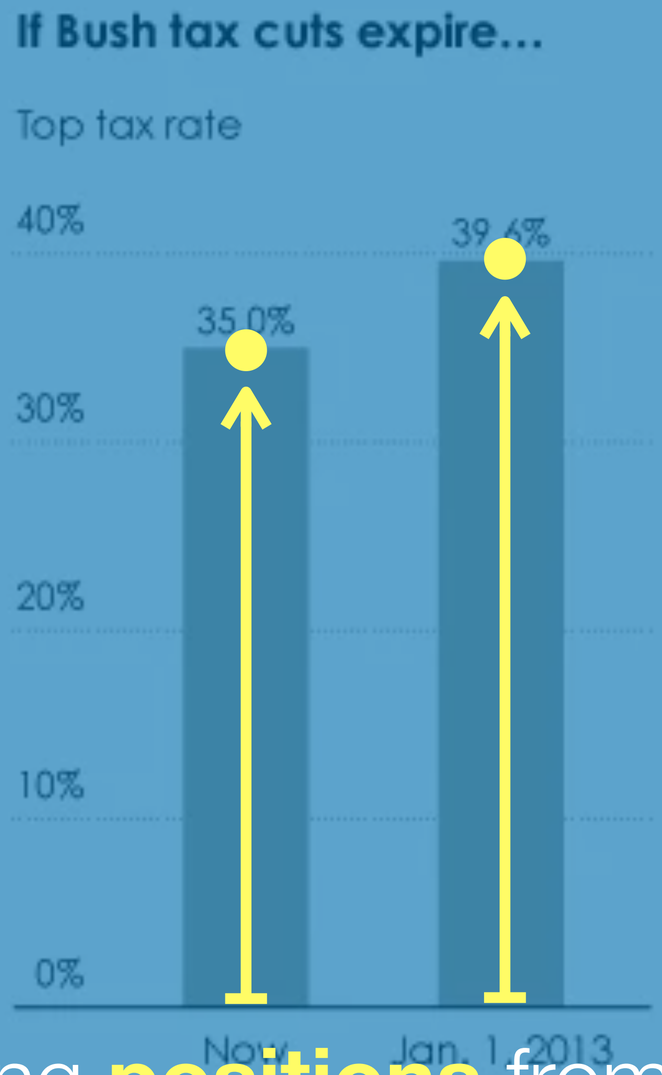

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE
42%
40
38
36
34
39.6%
35%
NOW
JAN. 1, 2013
8:01p ET
FOX BUSINESS
TOP STORIES
TECHNOLOGY
CONSUMER
VITH THE JUSTICE DEPARTMENT AND AQUIRES FULL T
DOW 13008.68 64.33
S&P 1379.32 5.98
NASDAQ 2939.52 6.32


If Bush tax cuts expire...
Top tax rate
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**



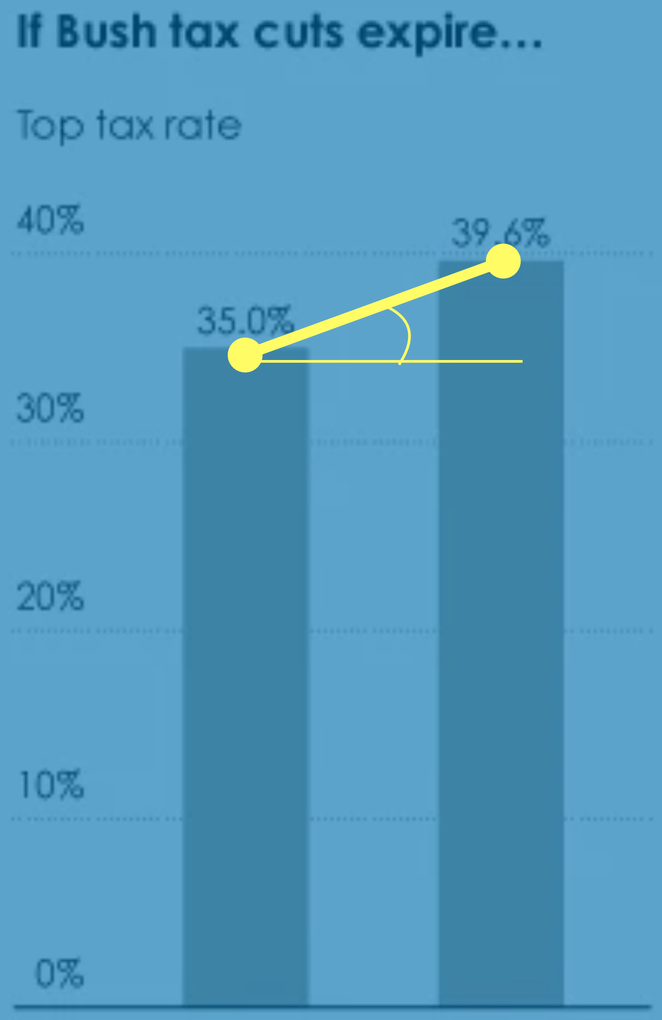
Or comparing **positions** from the **baseline**.

Do we always need a zero-baseline?

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

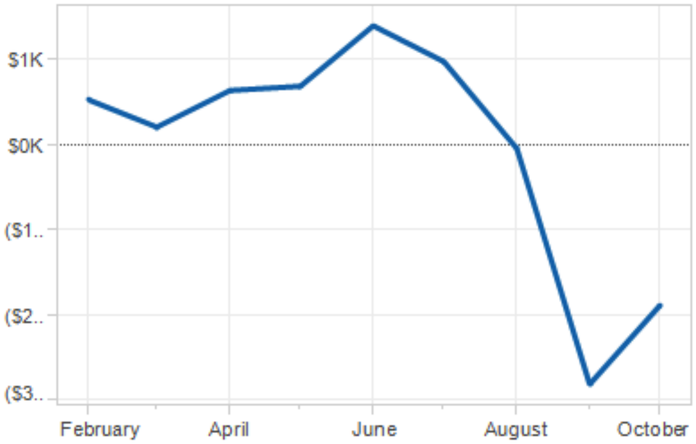
Edward Tufte

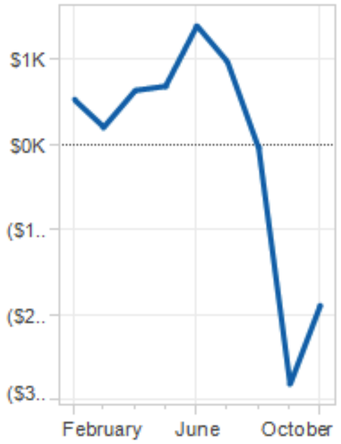
Line Chart may not have a zero-baseline.

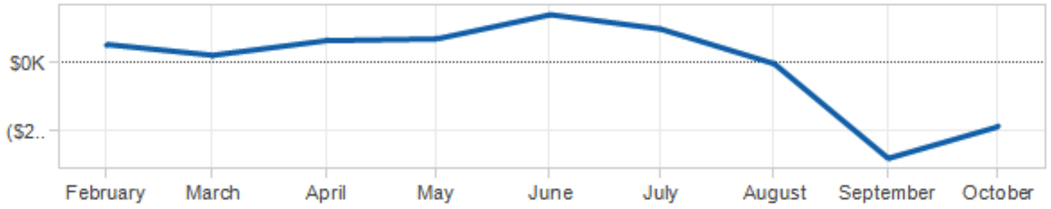


Because you are comparing the angle

Line chart’s aspect ratios can matter too.

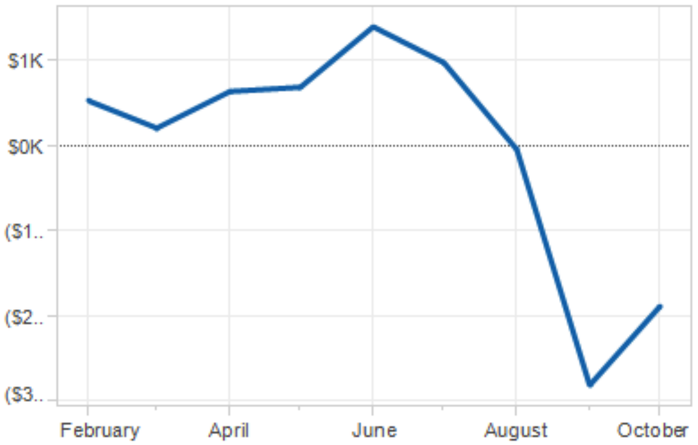


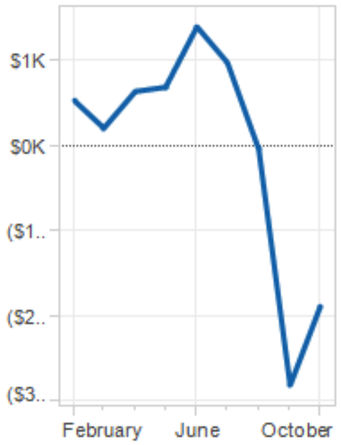


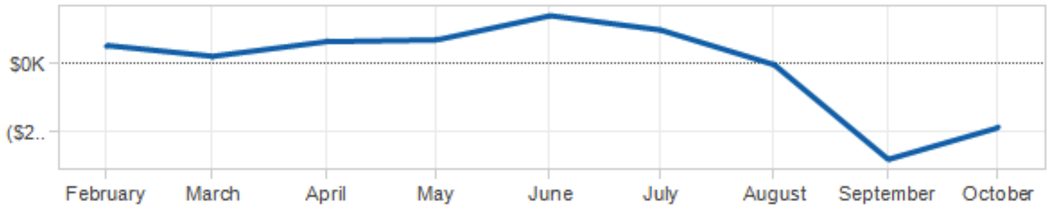


https://eagereyes.org/basics/banking-45-degrees

Line chart's aspect ratios can matter too.







A rule of thumb is banking to 45 degrees
to minimize errors in visual judgments of slope ratios.

Comes down to the
message you want to deliver

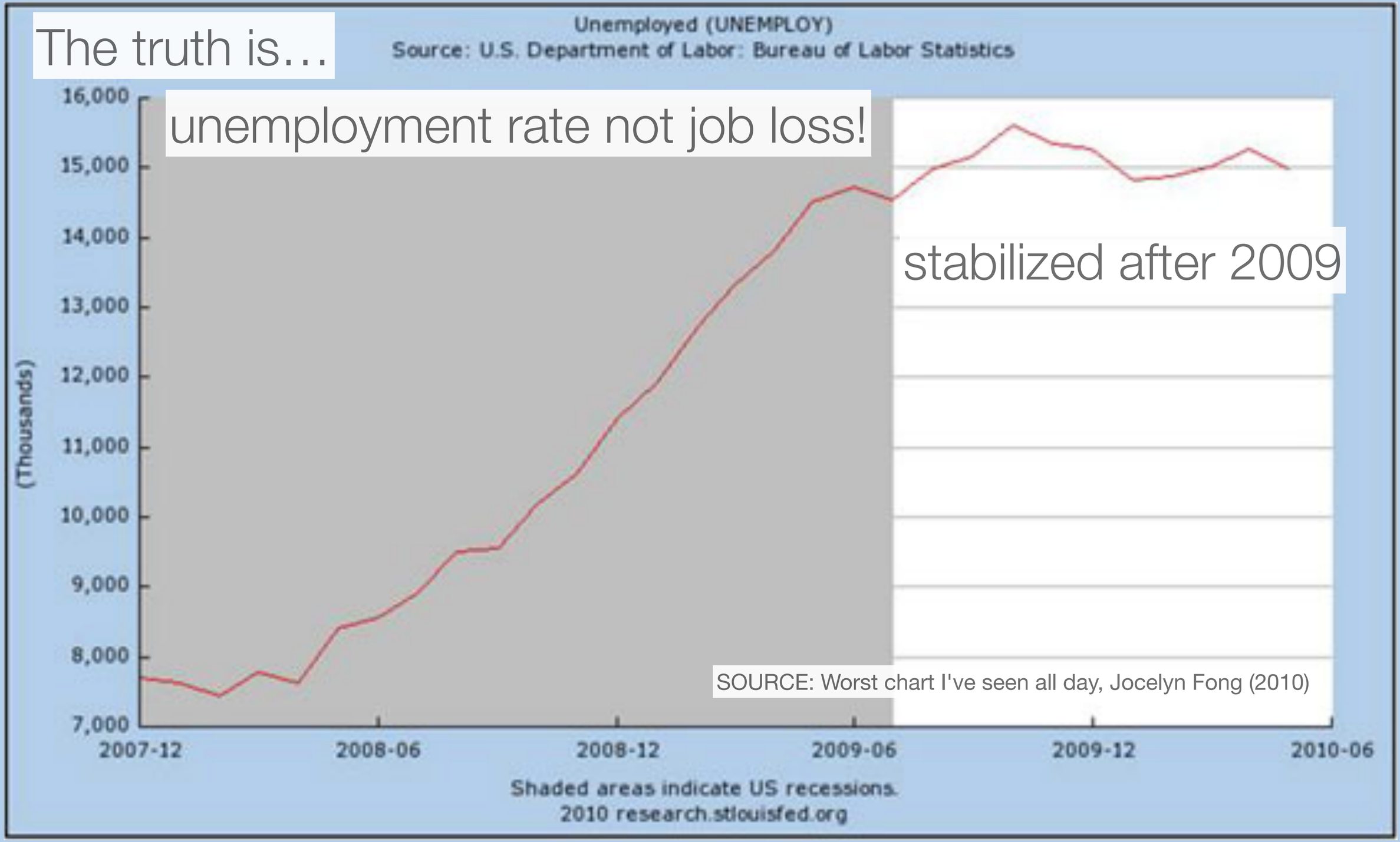
JOB LOSS BY QUARTER
15 MIL
13.5 MIL
9 MIL
7 MIL
DEC '07
SEPT '08
MARCH '09
JUNE '10
FOX NEWS .COM
SOURCE: BLS
AMERICA'S NEWSROOM

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
FOX NEWS
.COM
SOURCE: BLS
AMERICA'S NEWSROOM

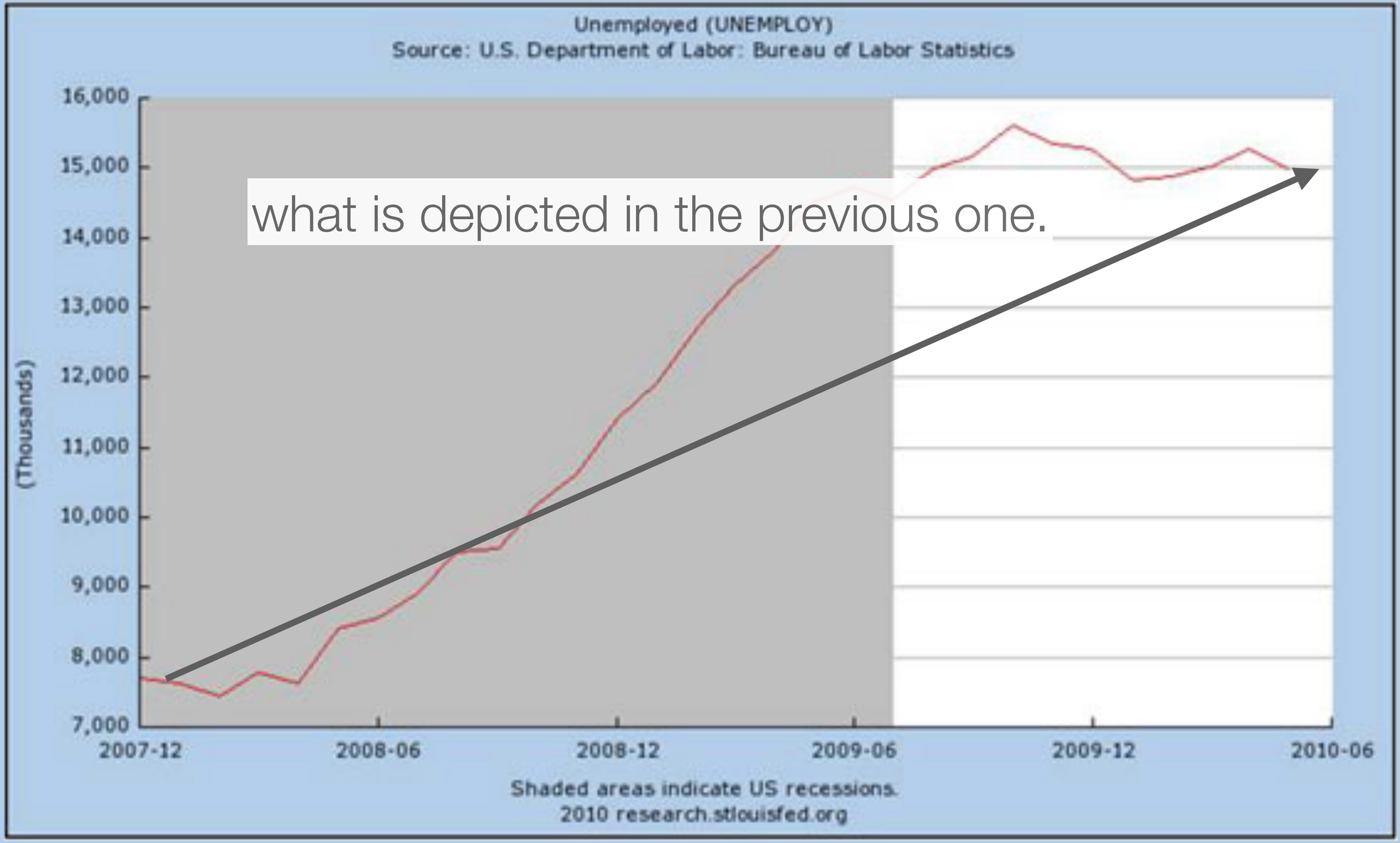
JOB LOSS BY QUARTER
INCORRECT Y-SCALE
15 MIL
13.5 MIL
9 MIL
7 MIL
4.5 mil
172 pixels
2 mil
138 pixels
DEC '07
SEPT '08
MARCH '09
JUNE '10
FOX NEWS
FOX NEWS
.COM
SOURCE: BLS
AMERICA'S NEWSROOM

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

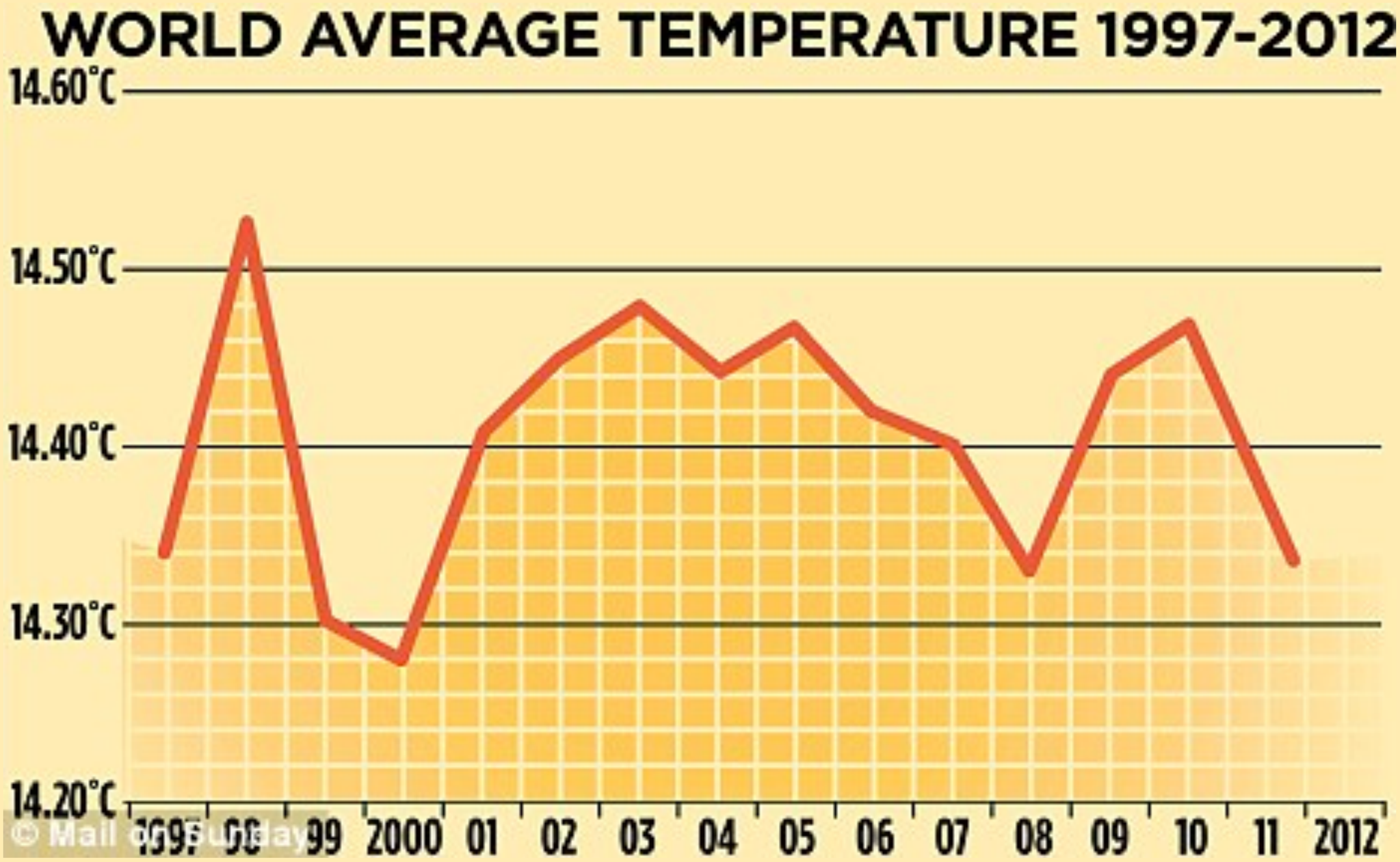
The truth is…

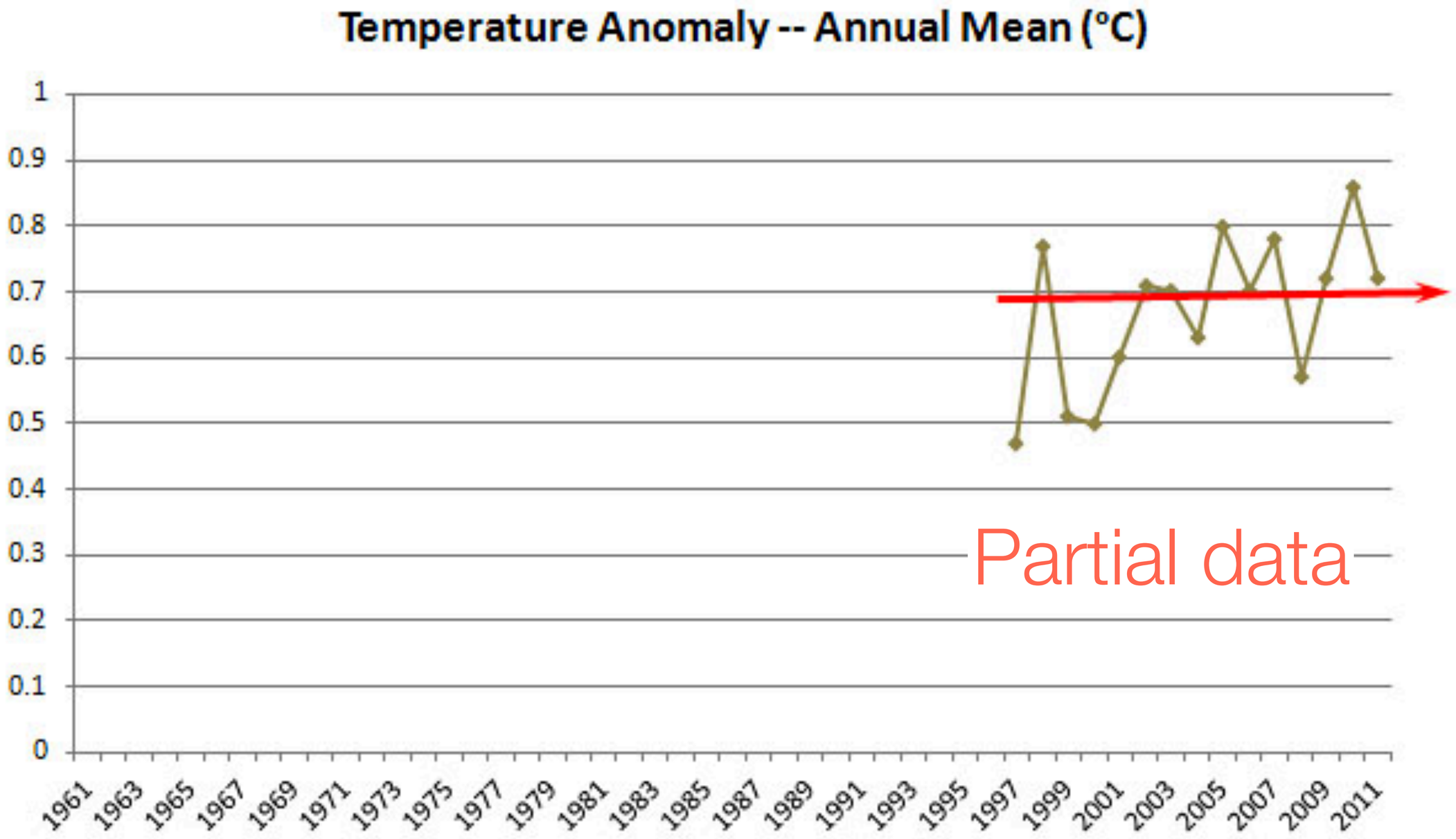


SOURCE: Worst chart I've seen all day, Jocelyn Fong (2010)

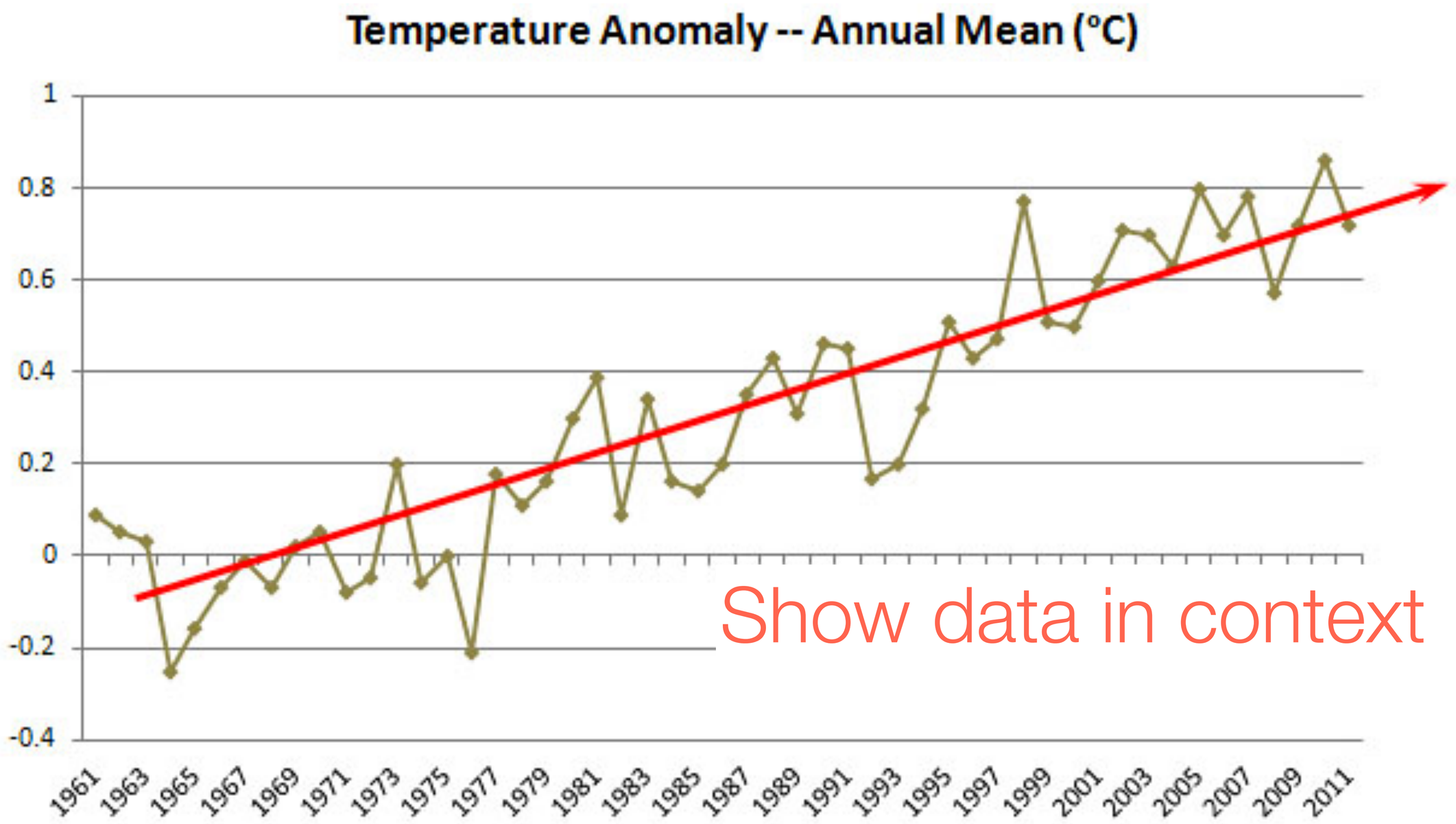


what is depicted in the previous one.

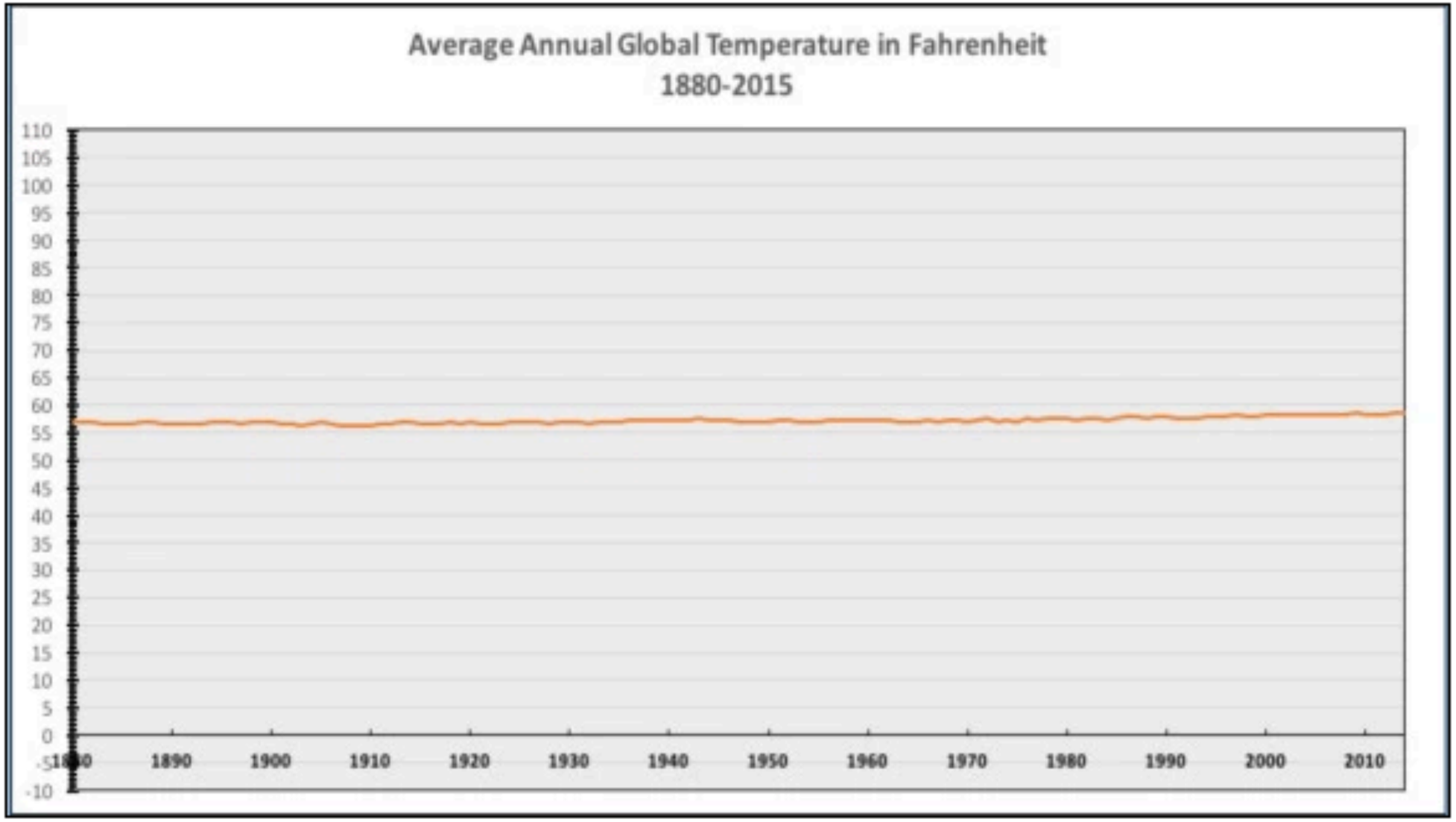

WORLD AVERAGE TEMPERATURE 1997-2012
14.60°C
14.50°C
14.40°C
14.30°C
14.20°C
97
98
99
2000
01
02
03
04
05
06
07
08
09
10
11
2012
© Mail on Sunday

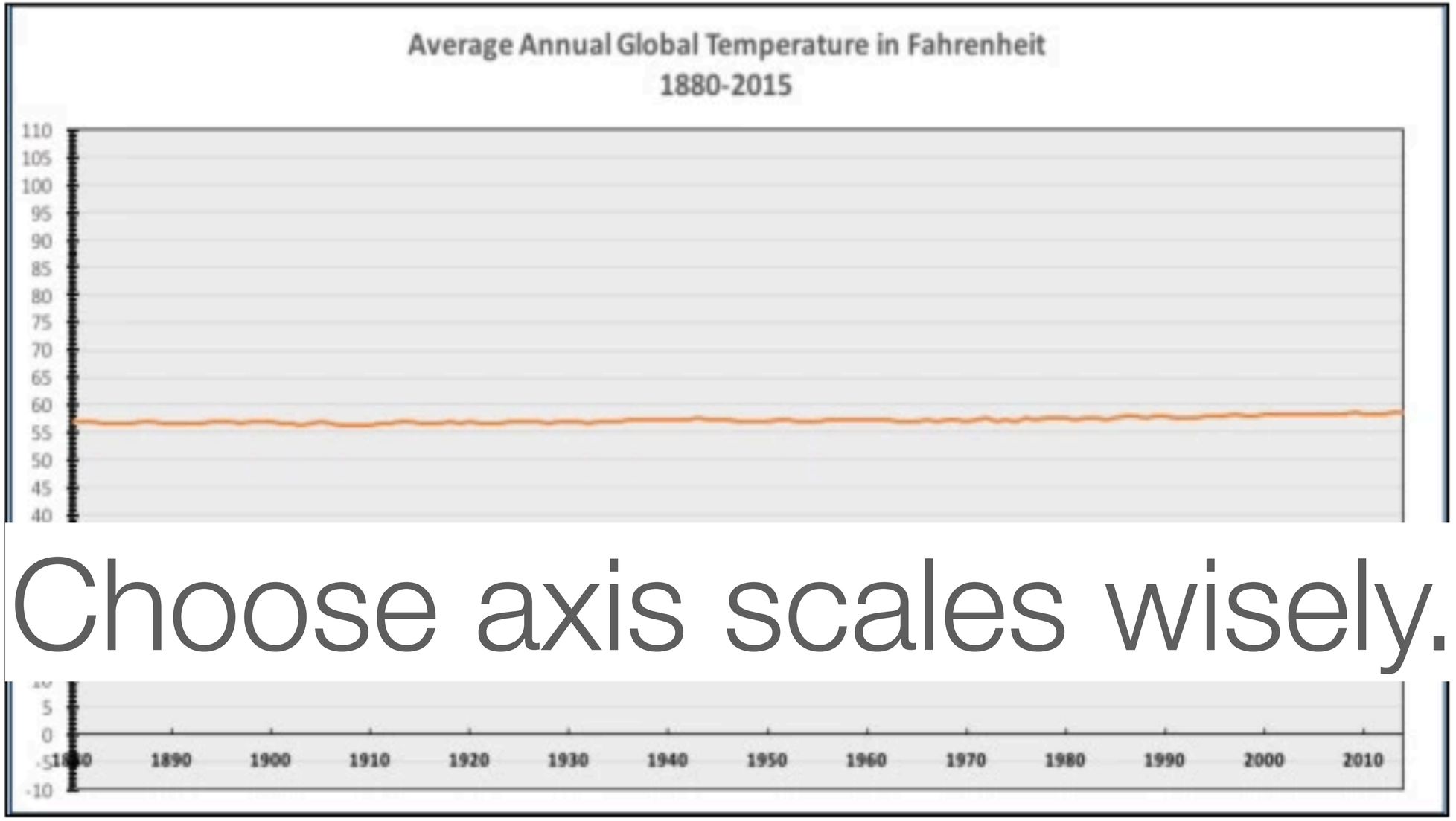


http://www.motherjones.com/kevin-drum/2012/01/lying-charts-global-warming-edition/


Temperature Anomaly -- Annual Mean (°C)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
Show data in context

Average Annual Global Temperature (°F) 1880-2015





Choose axis scales wisely.

Same here…

World population from 10,000 BC to present
5,000,000,000,000
0
Population

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.

— [Edward Tufte 83]

Lie Factor

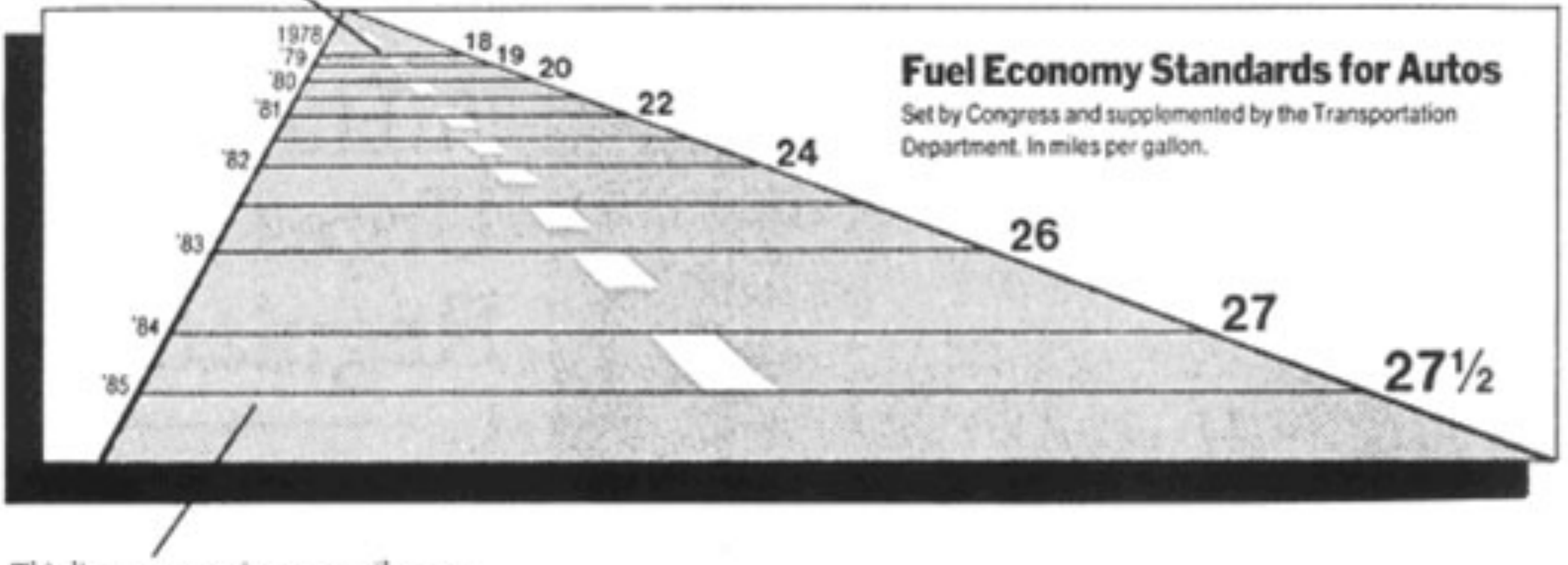
$$\text{Lie Factor} = \frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}$$

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

where size of effect = percentage change

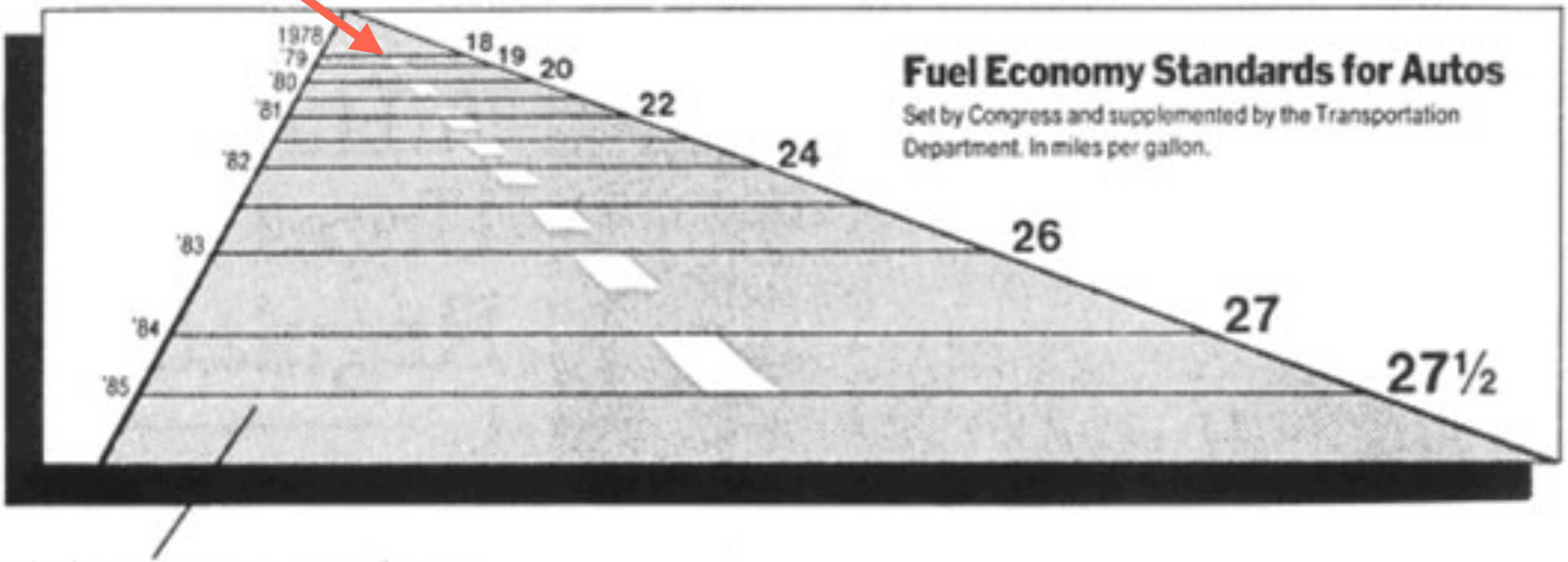
$$= \frac{|\text{first value} - \text{second value}|}{|\text{first value}|}$$

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



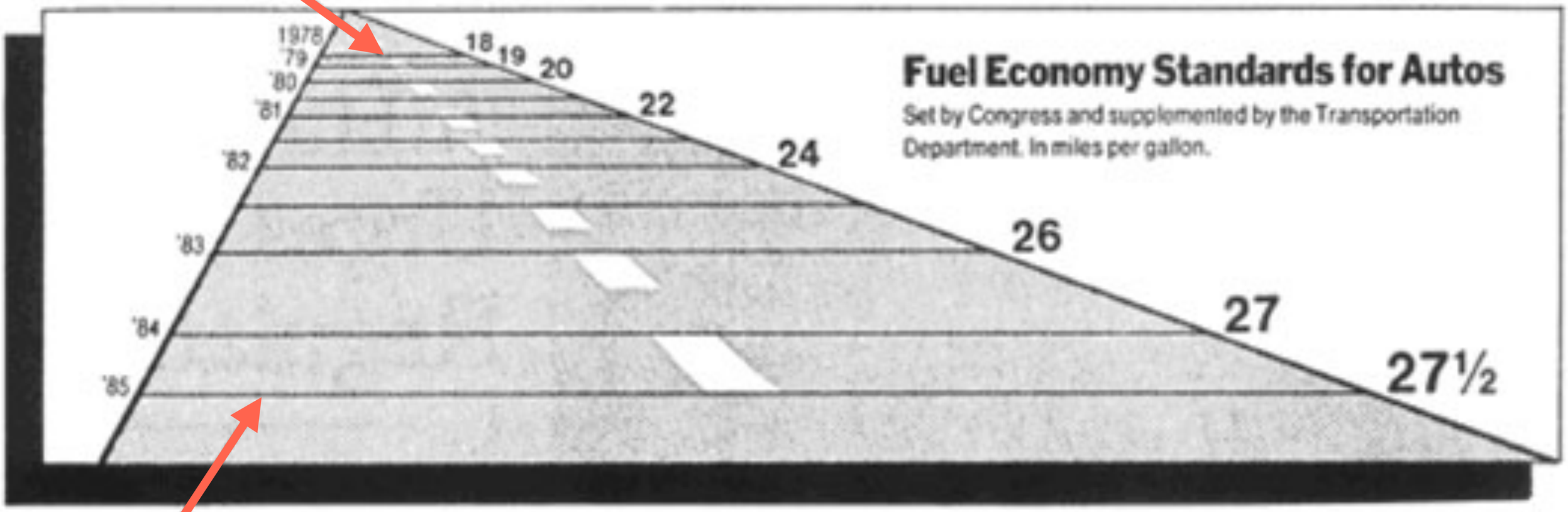
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

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



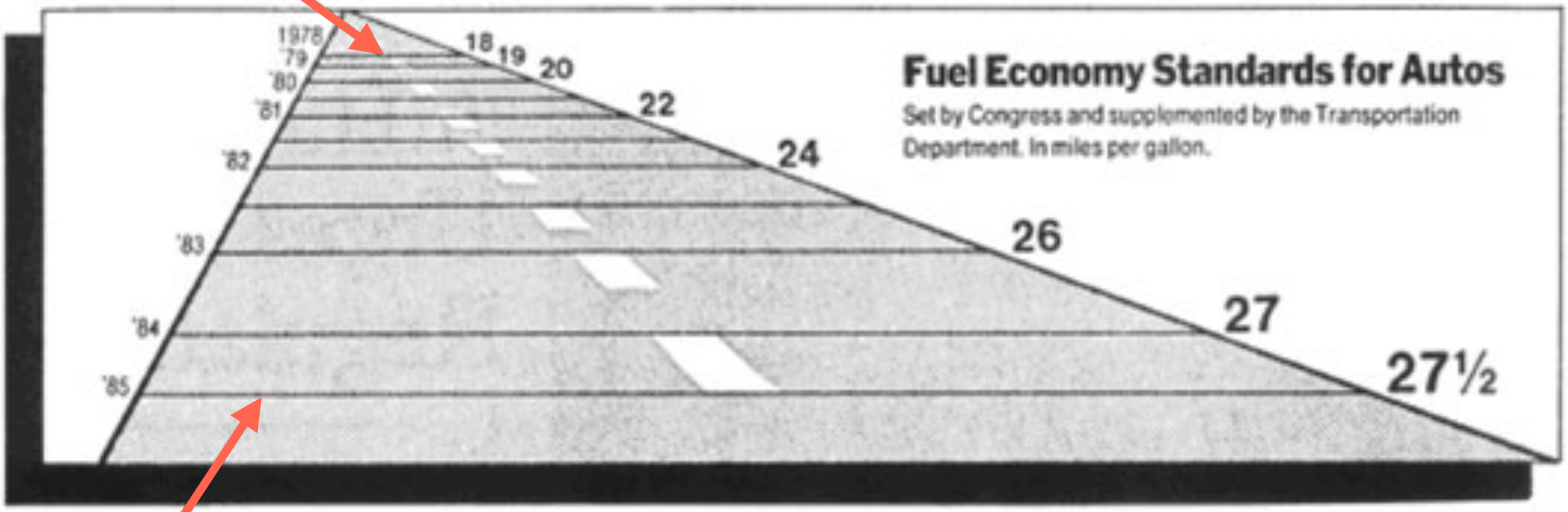
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

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



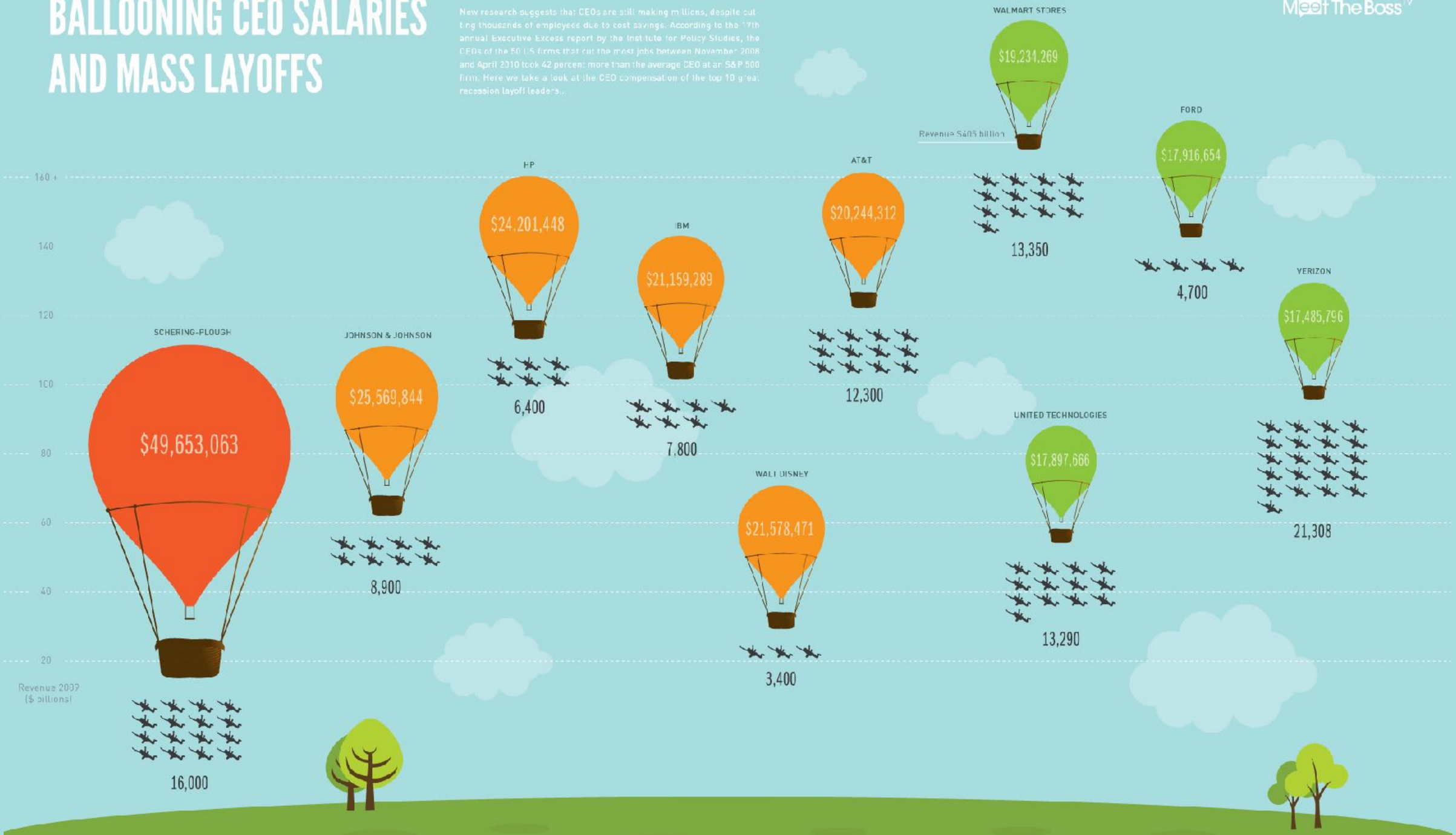
This line, representing **27.5 miles** per gallon in 1985, is **5.3 inches** long

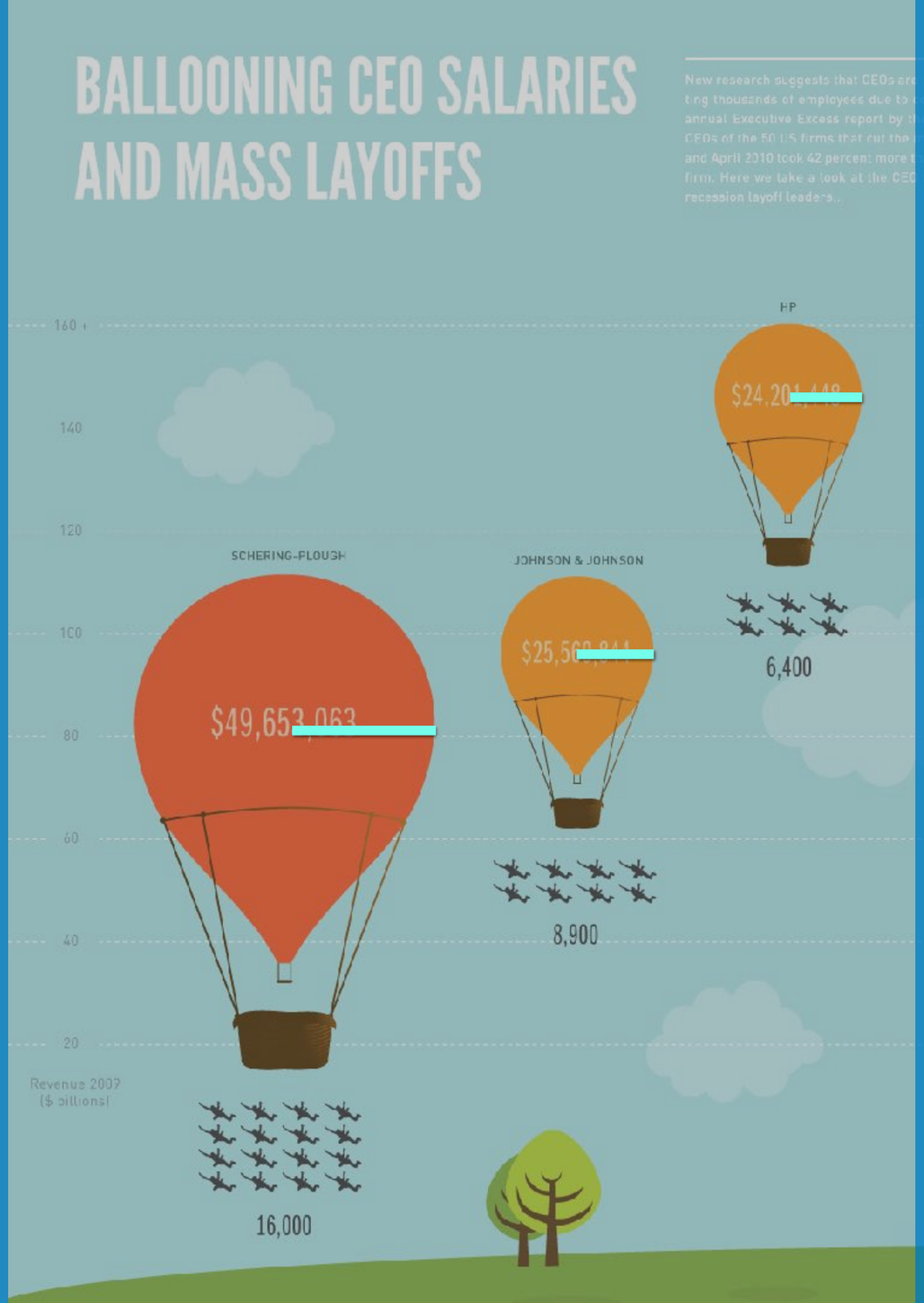
$$\textbf{Lie Factor} = \frac{(5.3-0.6) / 0.6}{(27.5-18) / 18} = \textbf{14.8}$$

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 U.S 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

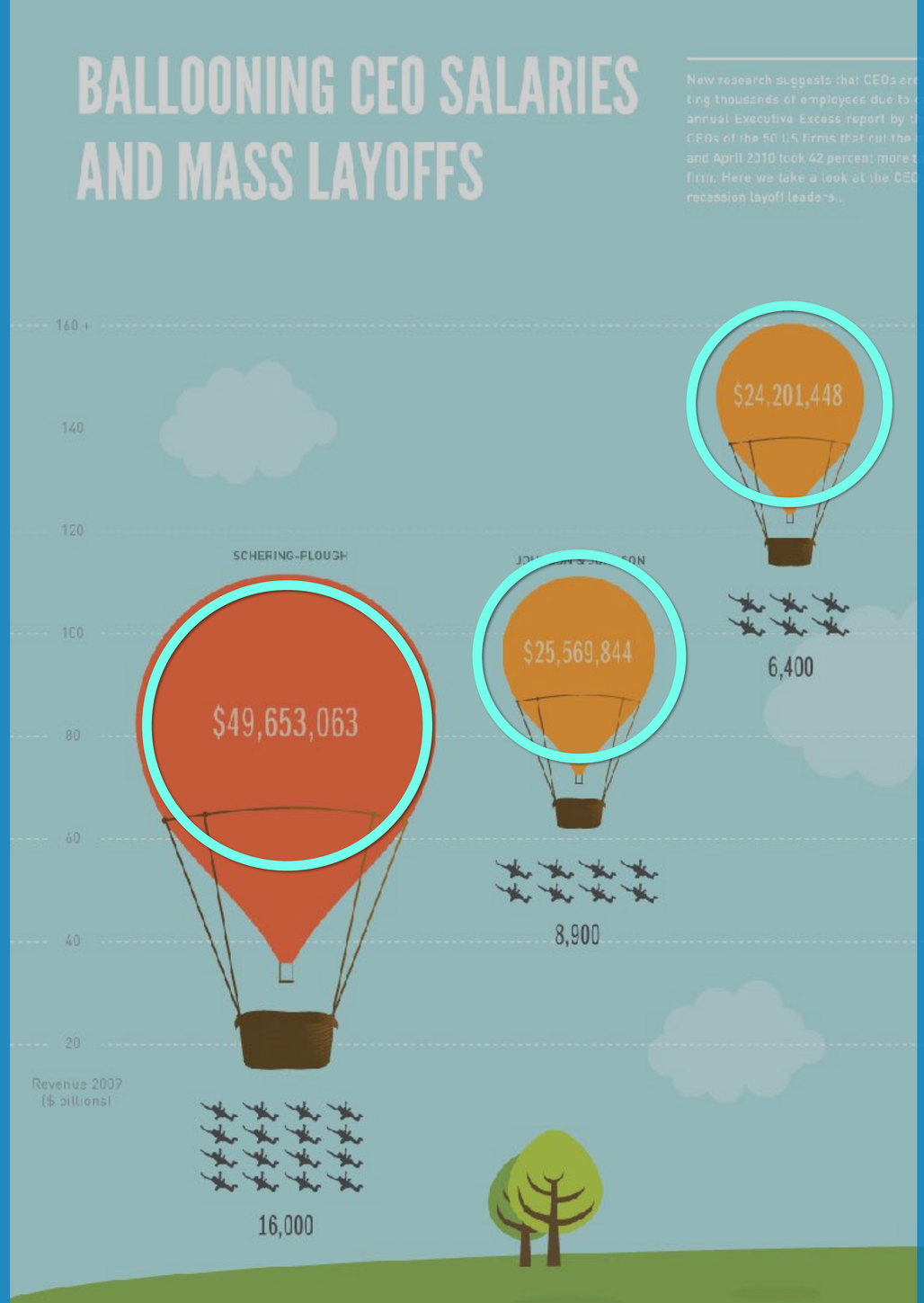




This chart uses radius of the balloon to encode the data

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

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

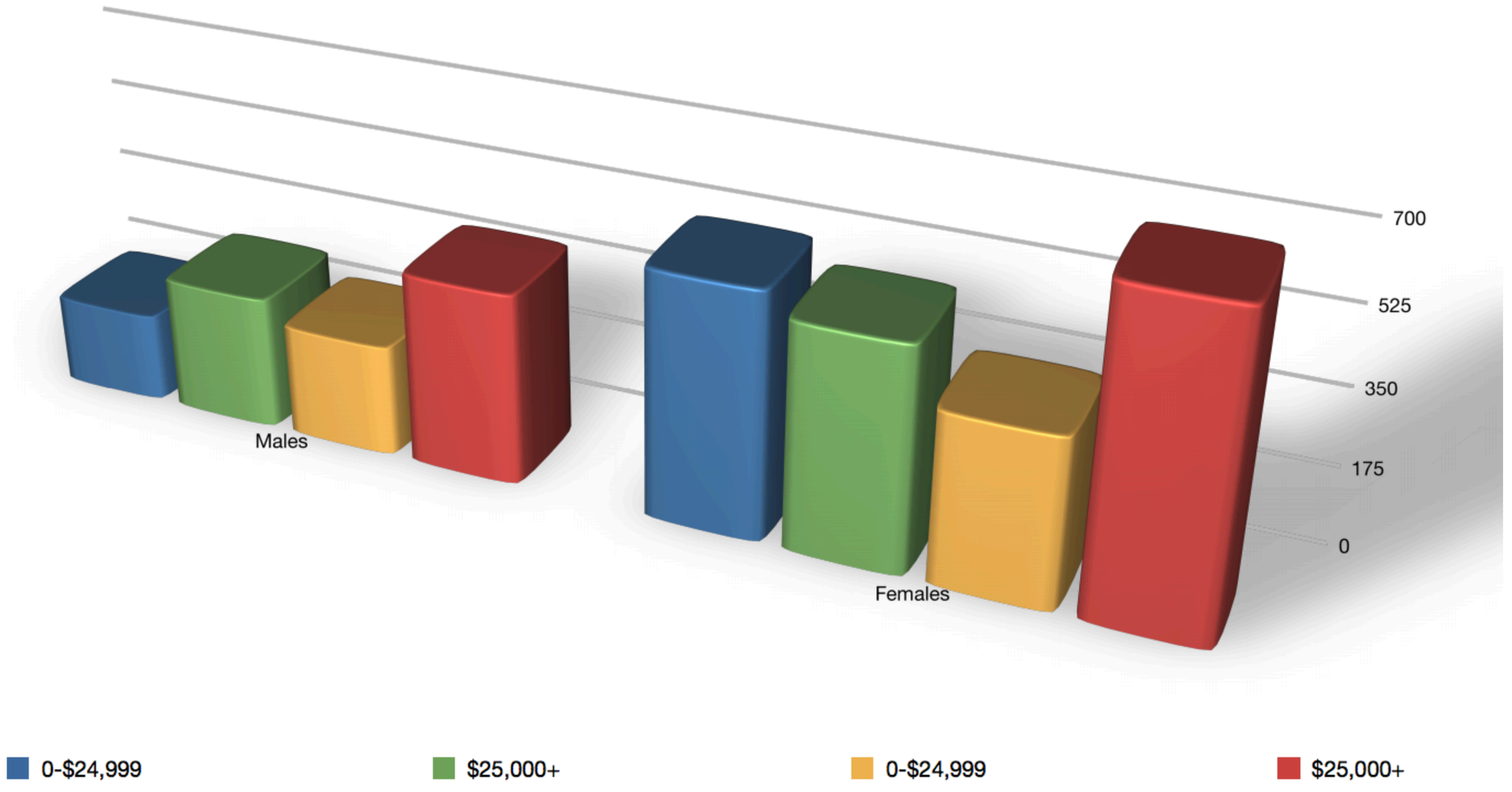


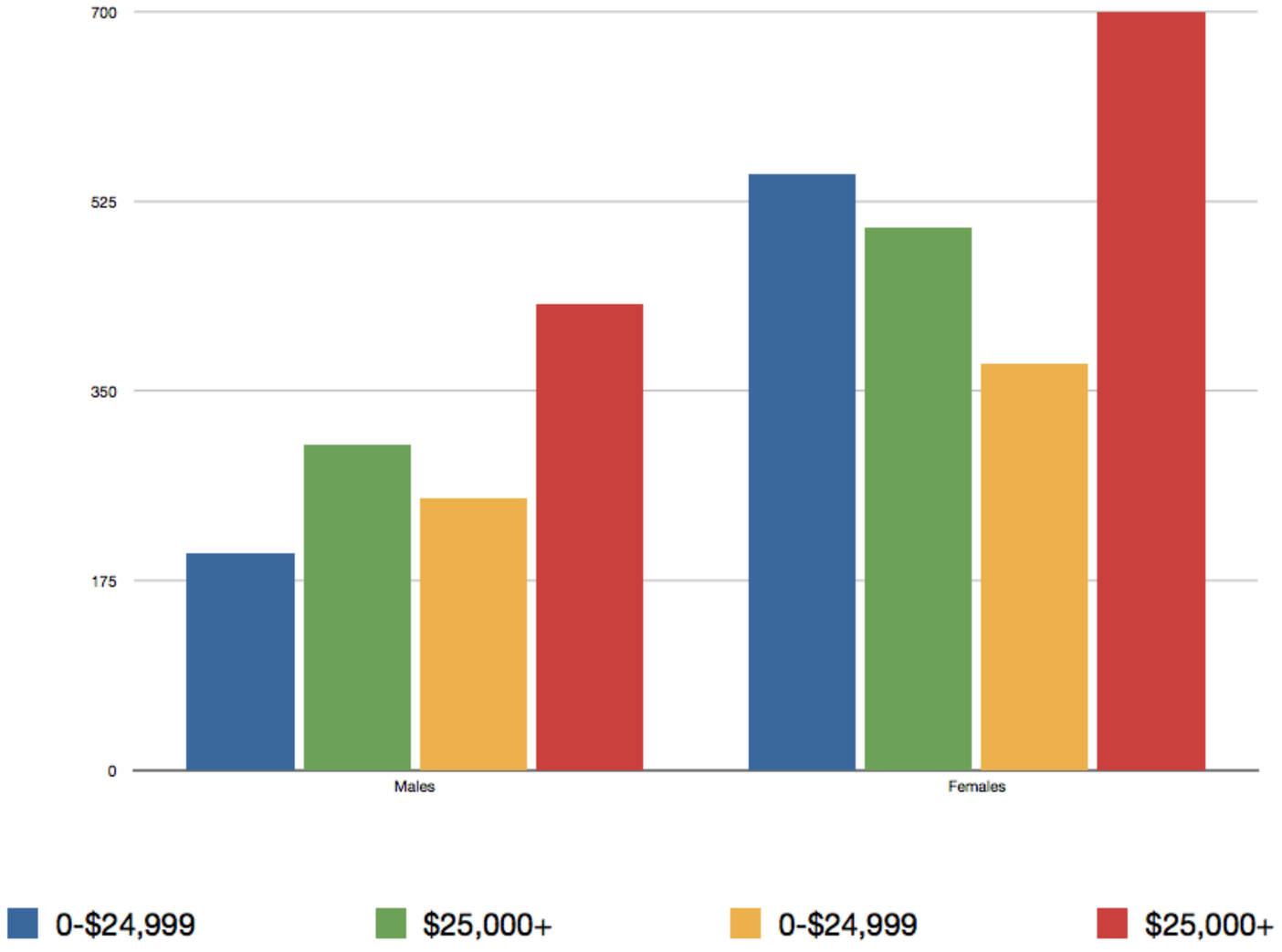
The size of the balloon should be something like this.

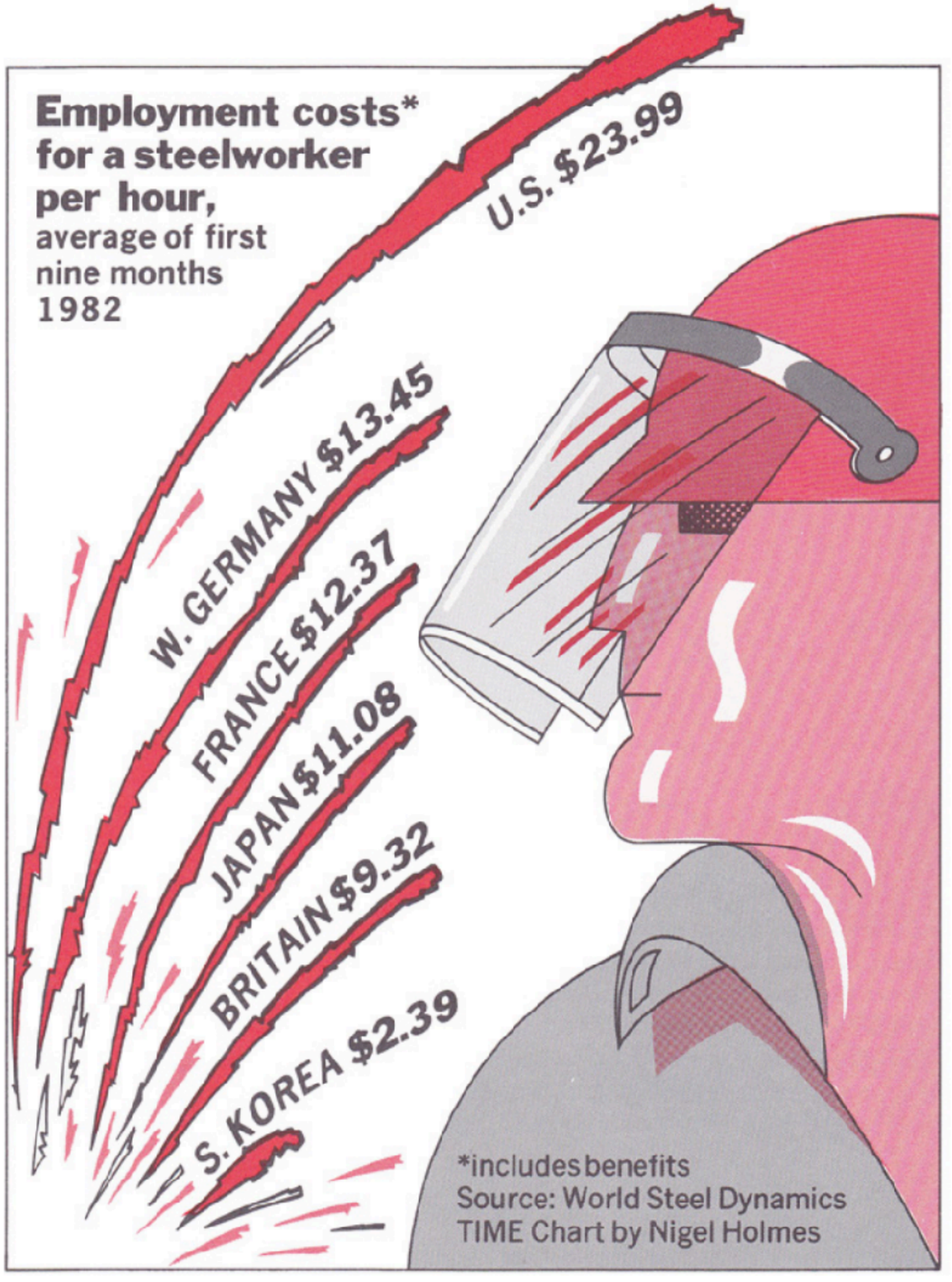
Avoid Distortion!

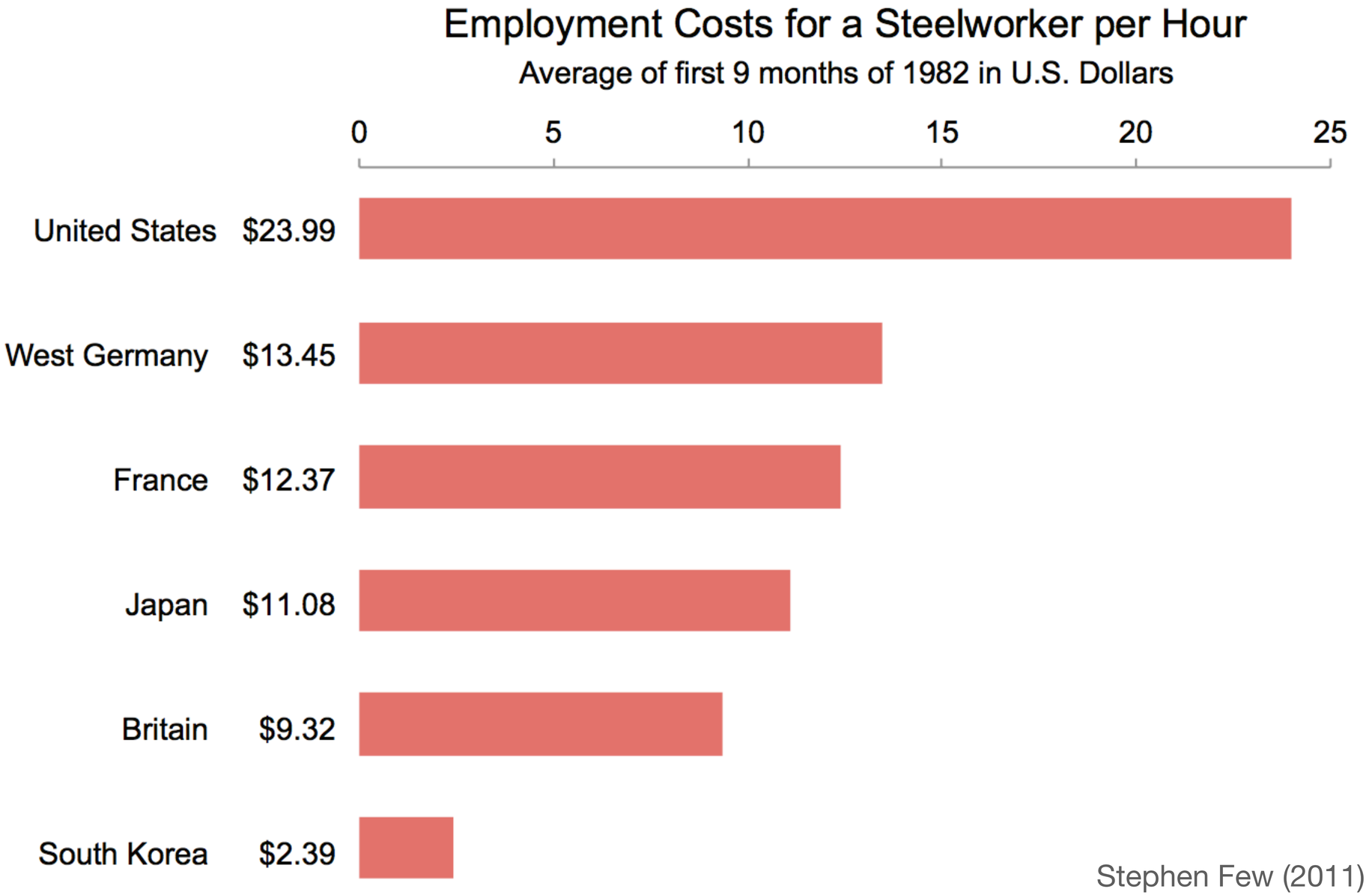
Maximize Data-Ink Ratio

$$\text{Data-Ink Ratio} = \frac{\text{Data-Ink}}{\text{Total ink used in graphic}}$$


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+

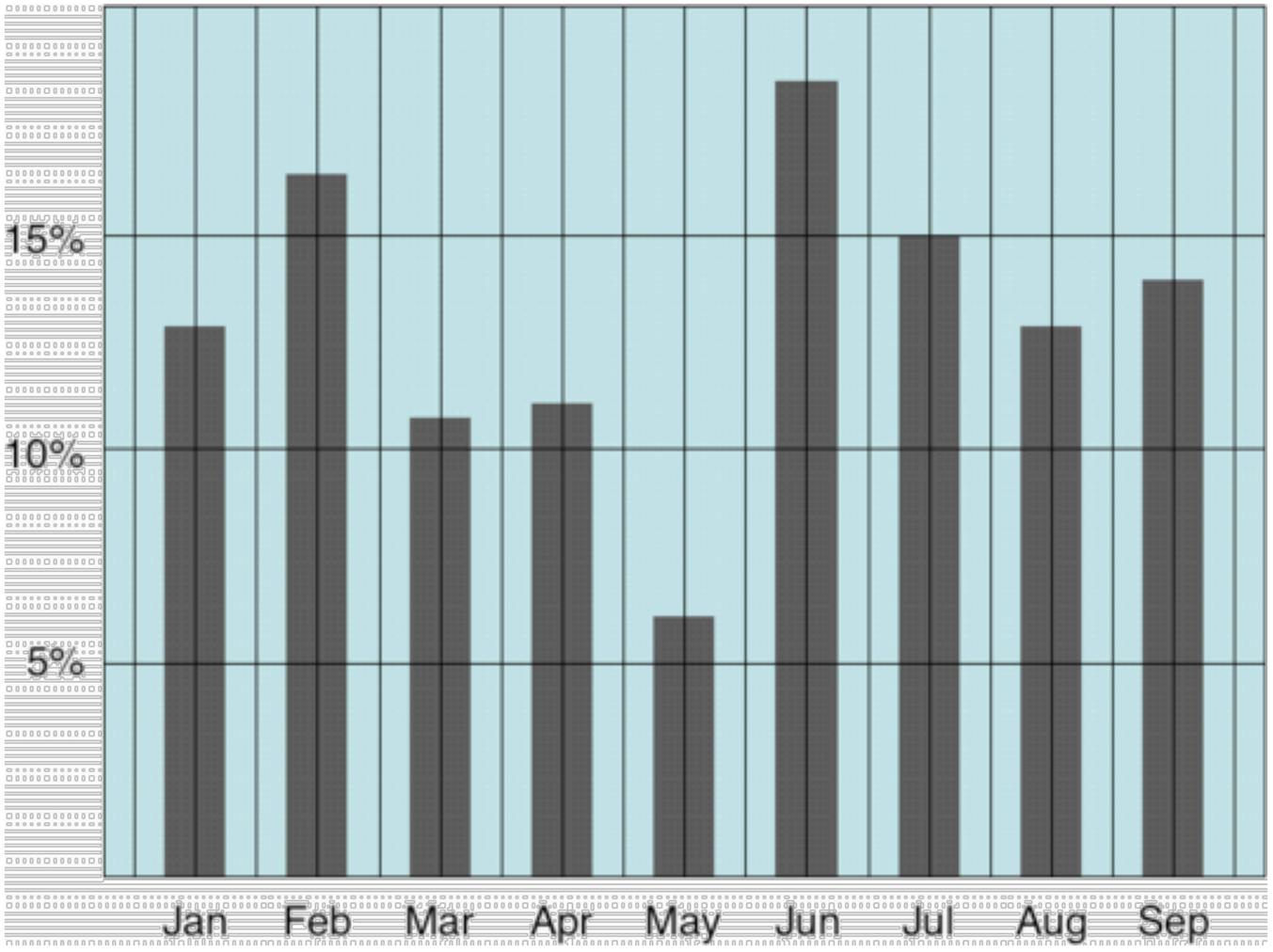

Employment costs* for a steelworker per hour,
average of first nine months 1982
U.S. $23.99
W. GERMANY $13.45
FRANCE $12.37
JAPAN $11.08
BRITAIN $9.32
S. KOREA $2.39
*includes benefits
Source: World Steel Dynamics
TIME Chart by Nigel Holmes

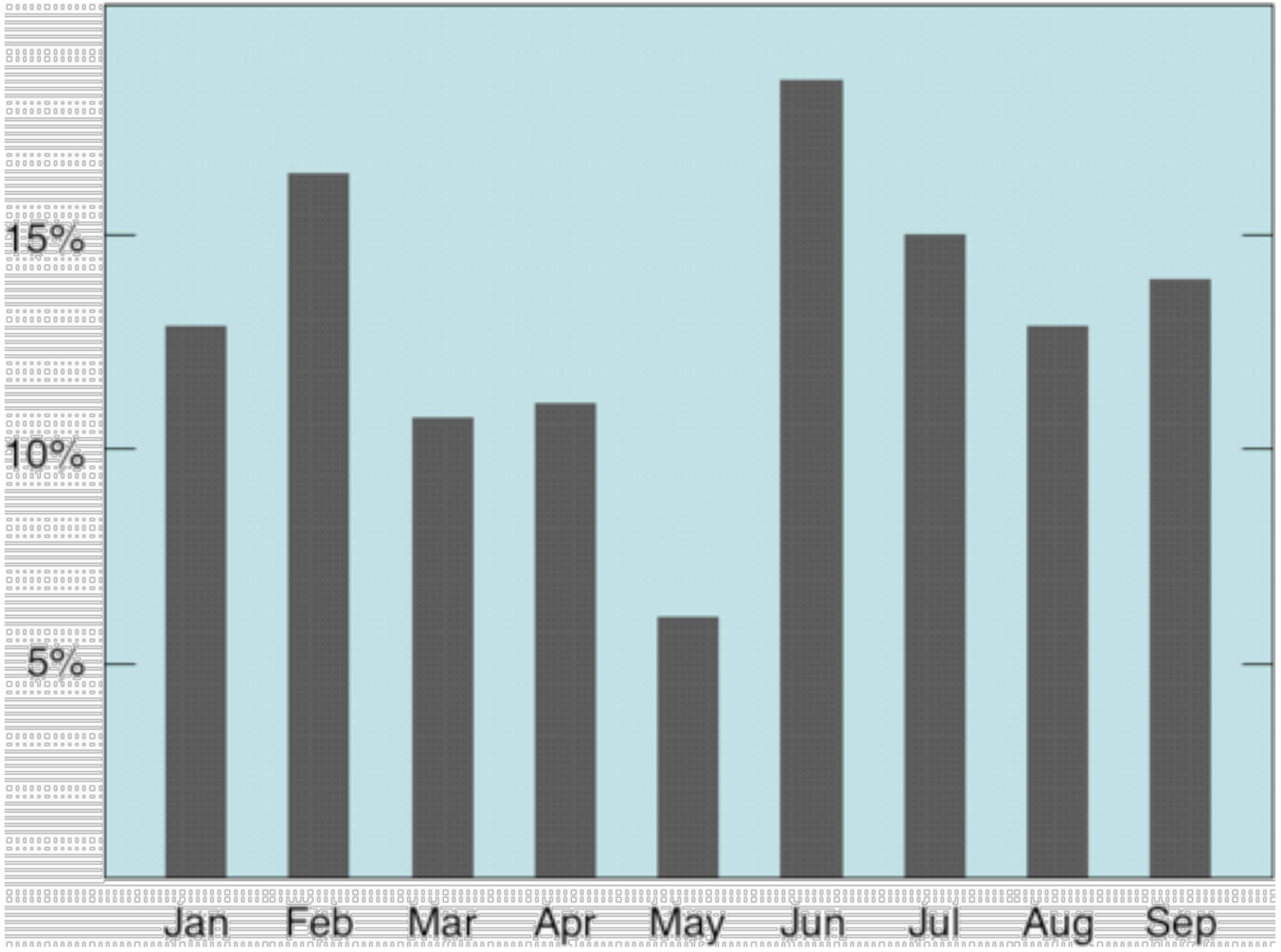


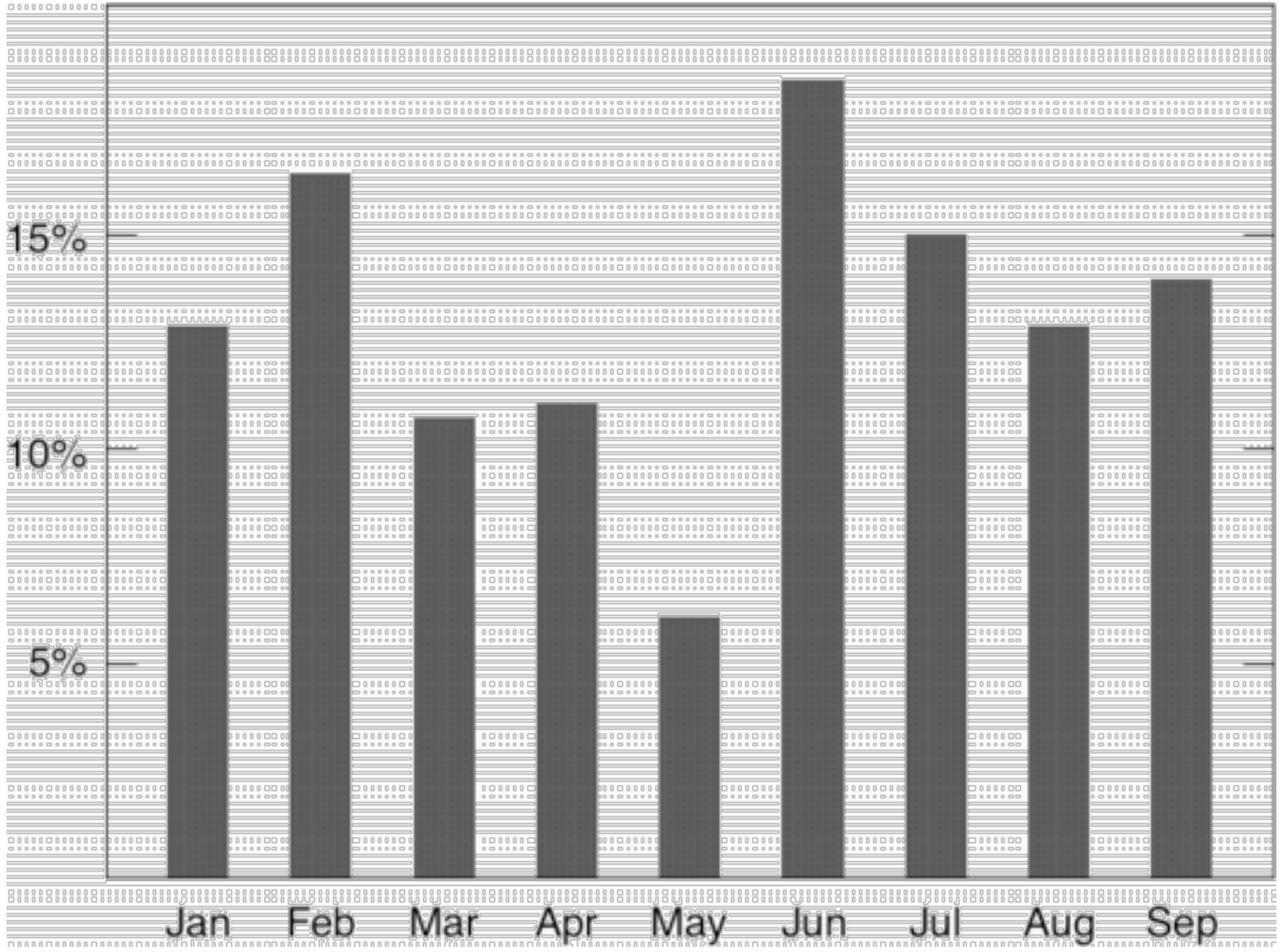
Stephen Few (2011)

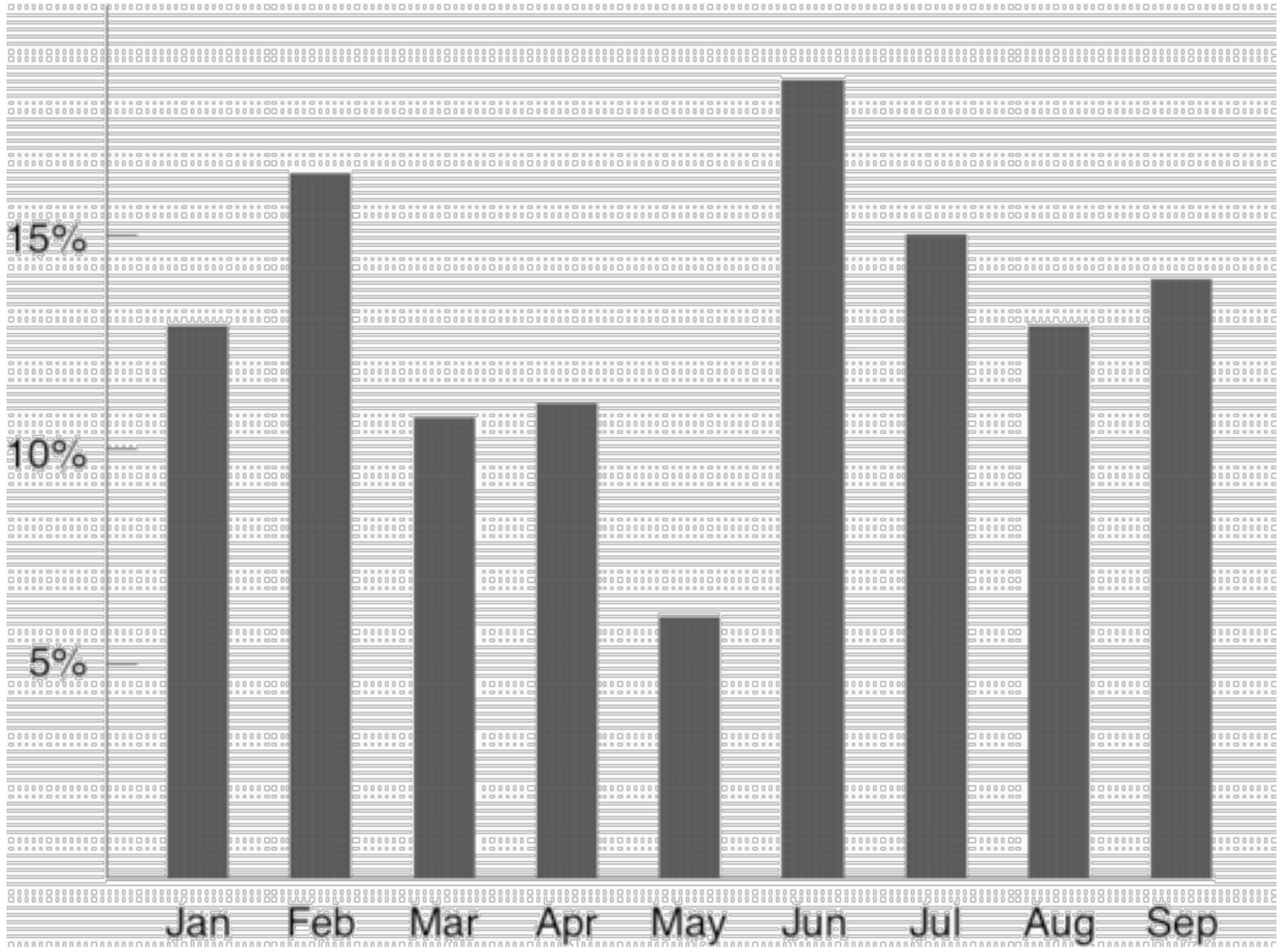
Avoid Chart Junks

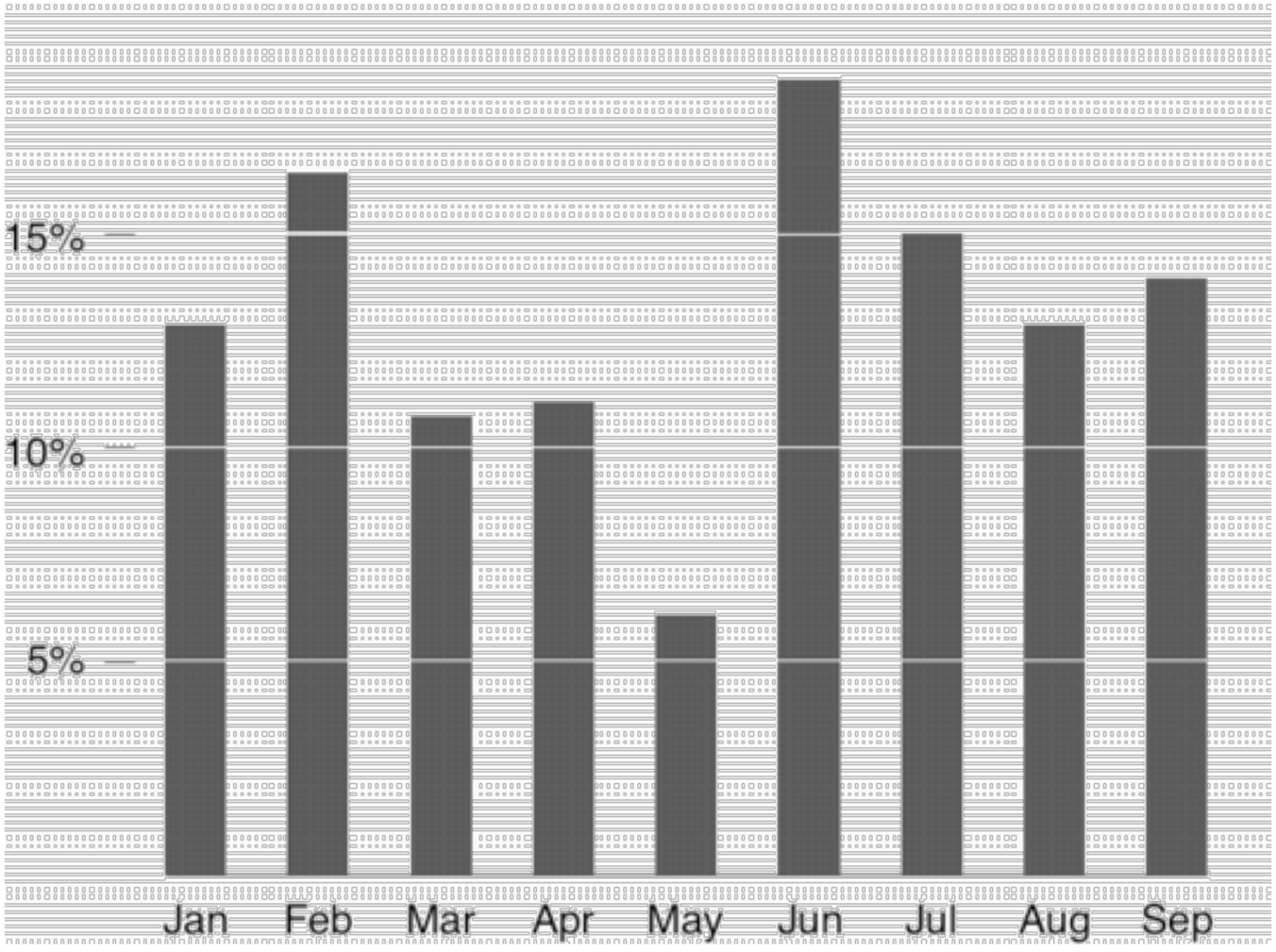
Chart Junks = Unnecessary visual elements in charts that distracts the viewer from the information

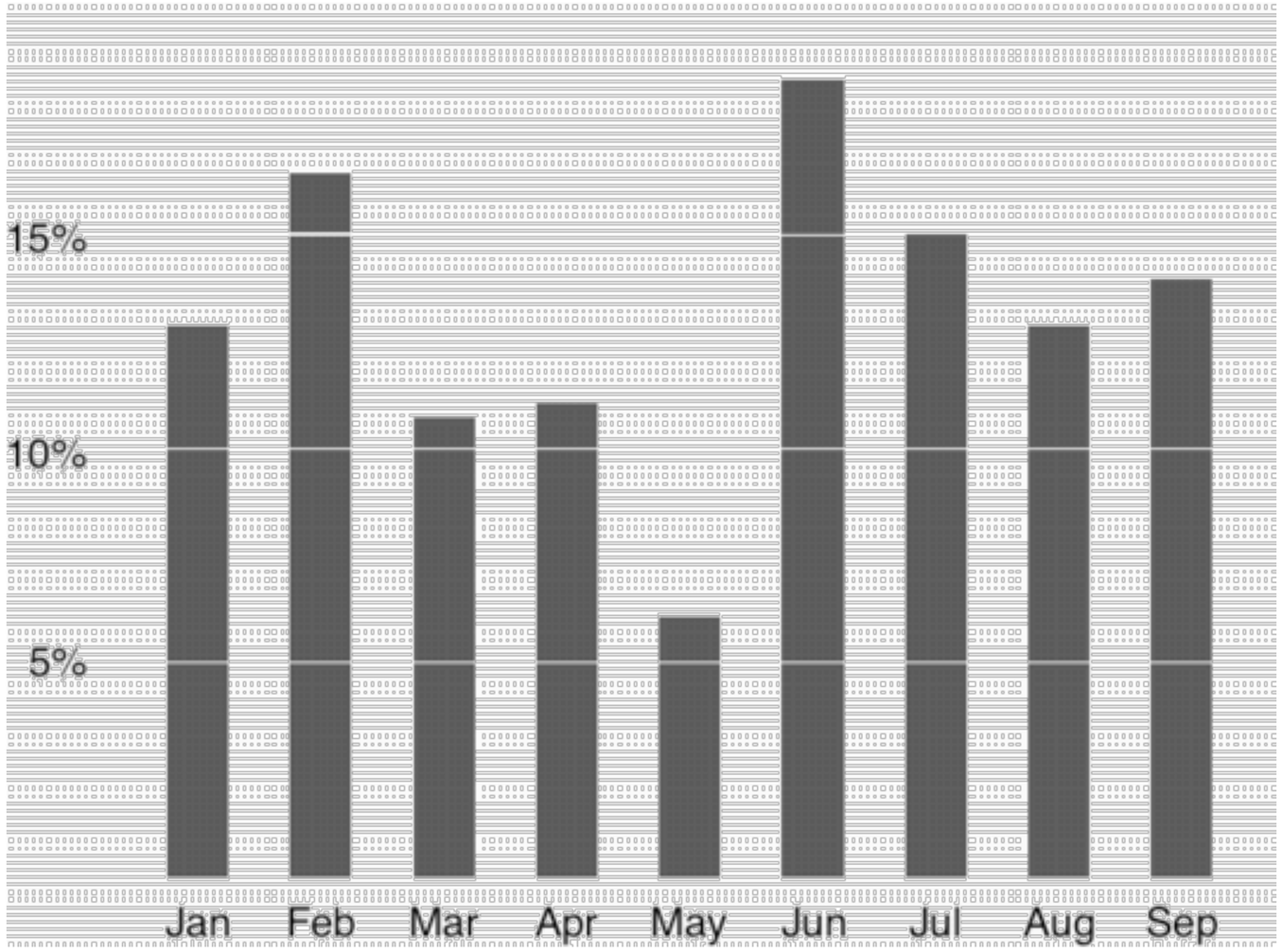

15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep


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

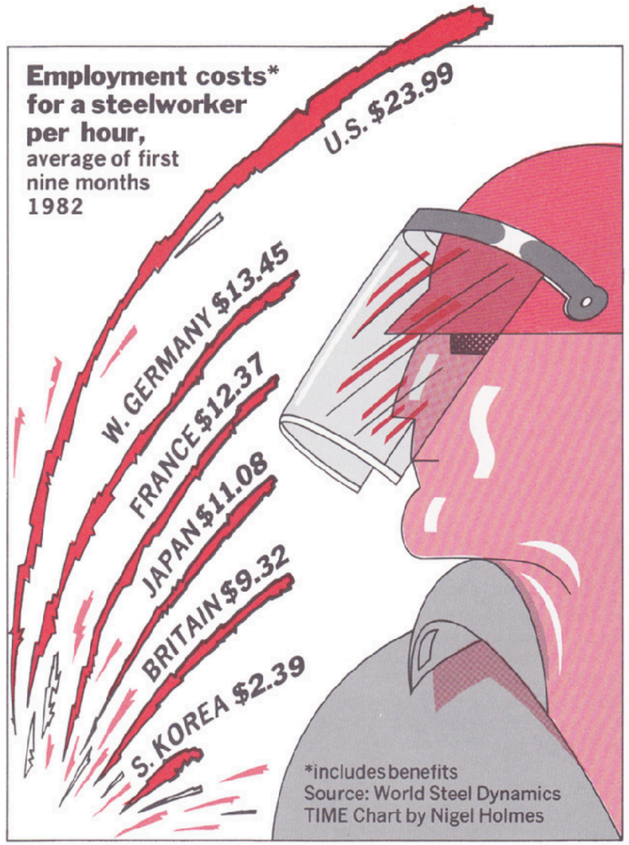

15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep


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


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

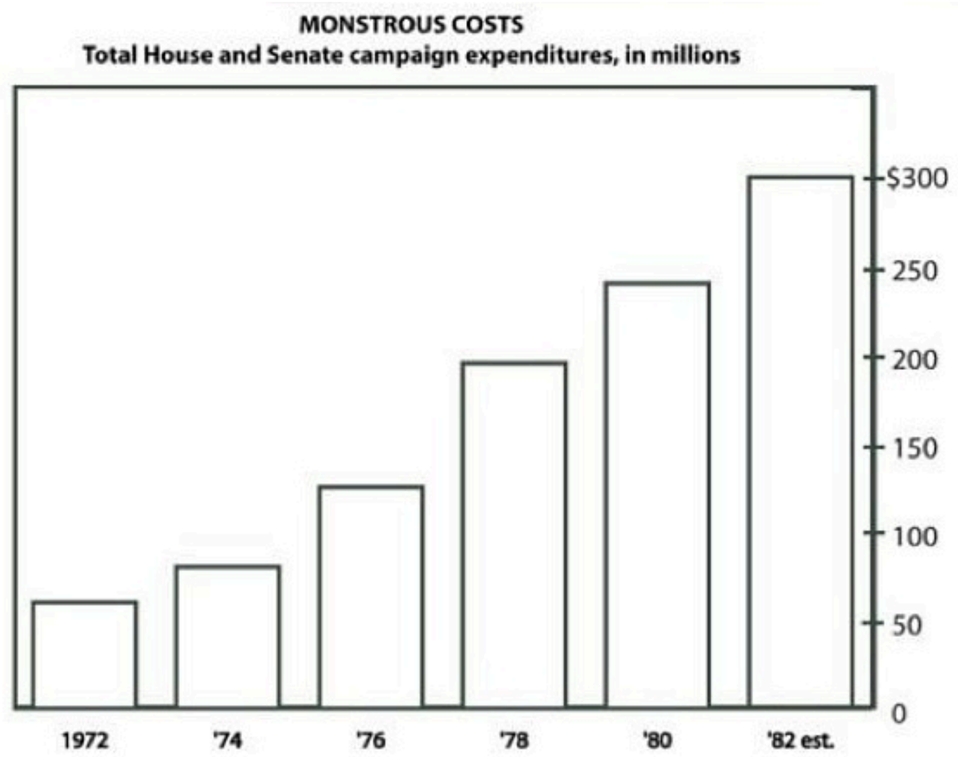

15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep

Are these chart junks?



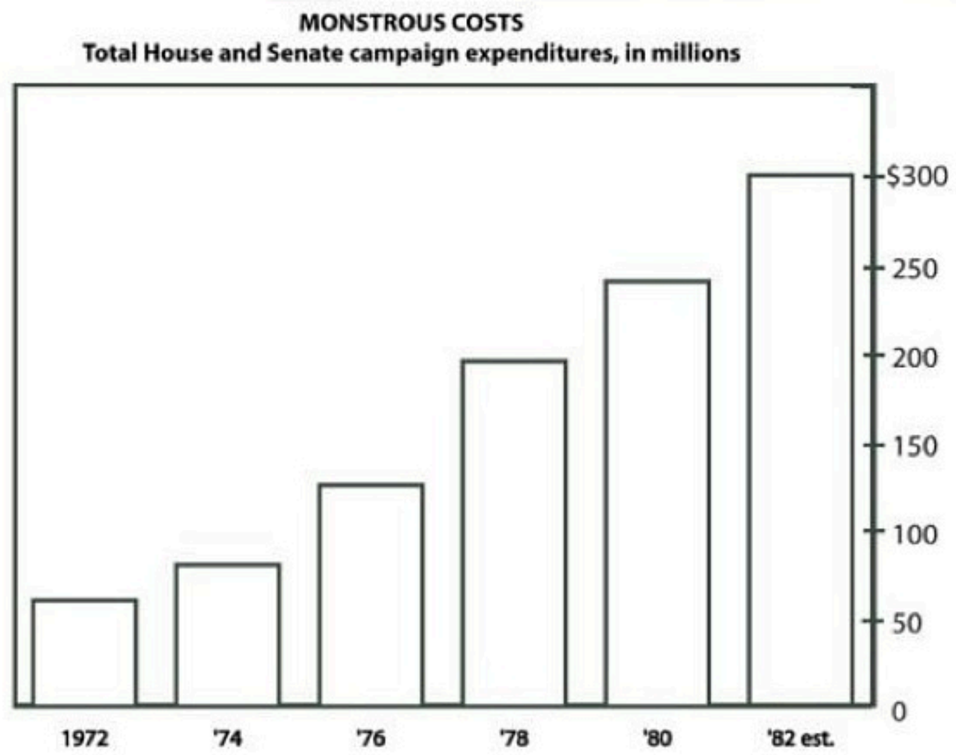
Not all chart junks are the same.

Useful chart junks?



Source: Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts (CHI 2010)

Not harmful in comprehension but more **engaging** & **memorable**



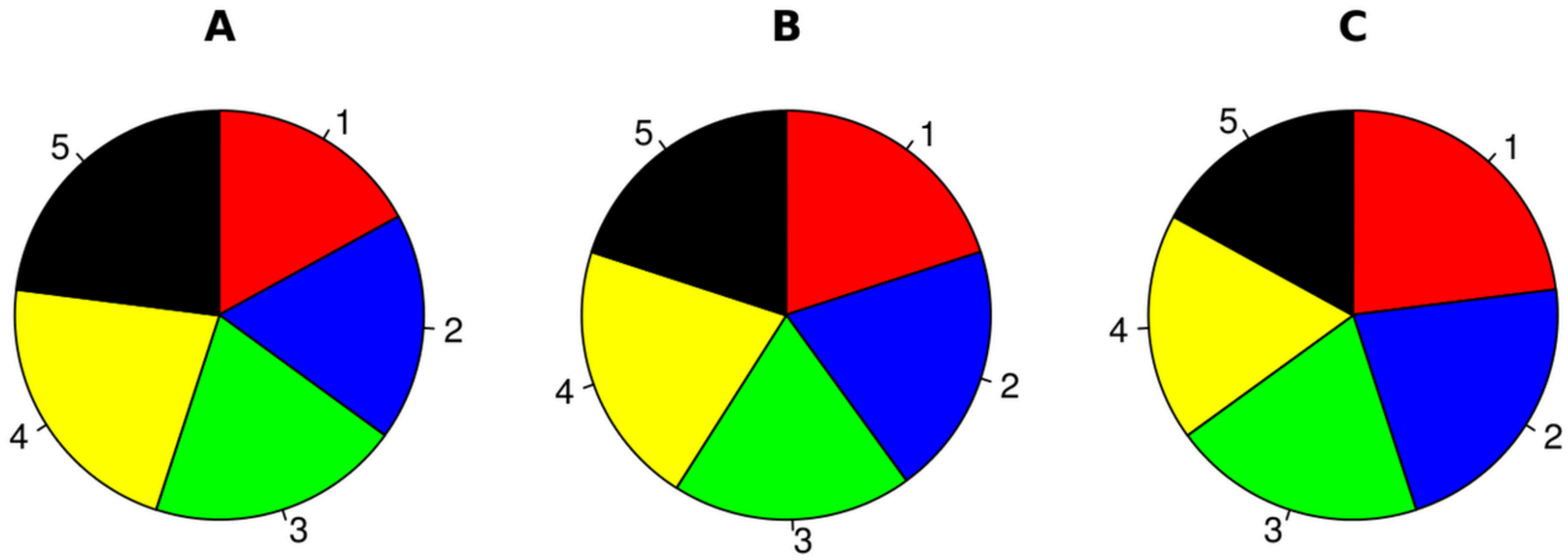
“The same ink should often serve more than one graphical purpose. A graphical element may carry data information and also perform a design function usually left to non-data-ink.” — [Edward Tufte 83]

Contextual representation can be helpful

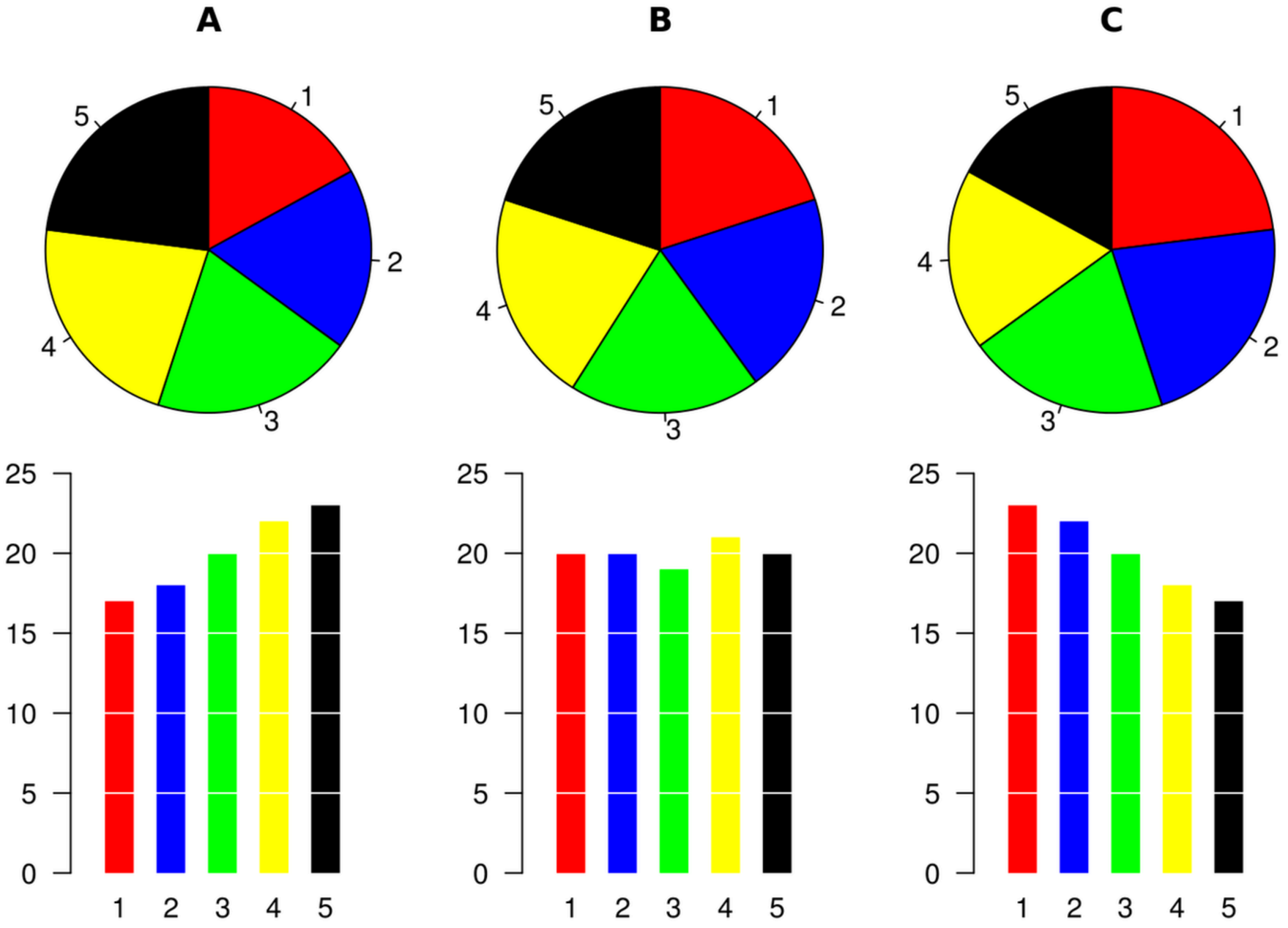
Pie Charts

Challenge:

Find the biggest pie slice in each pie chart!

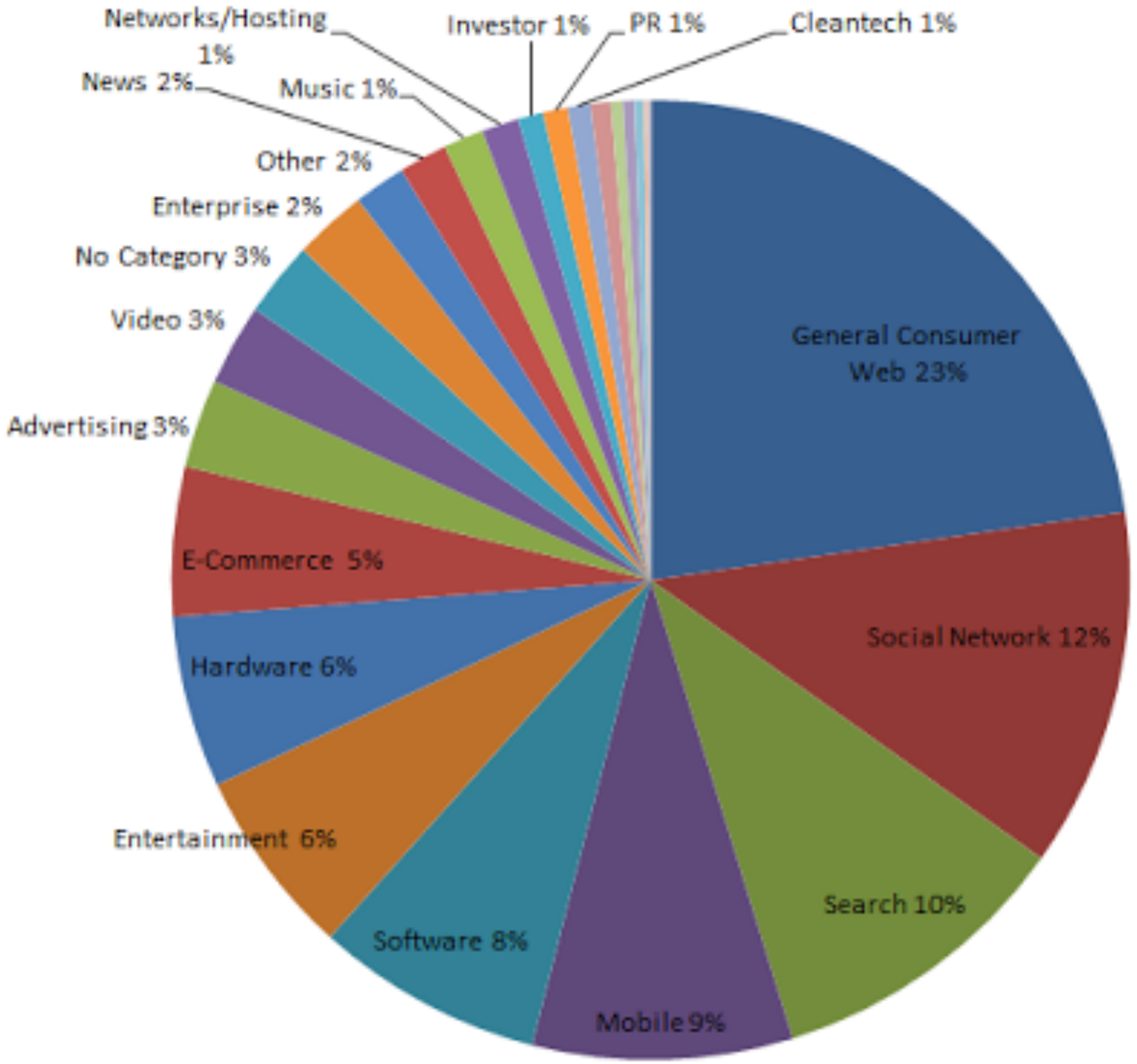


[Schutz 2007]



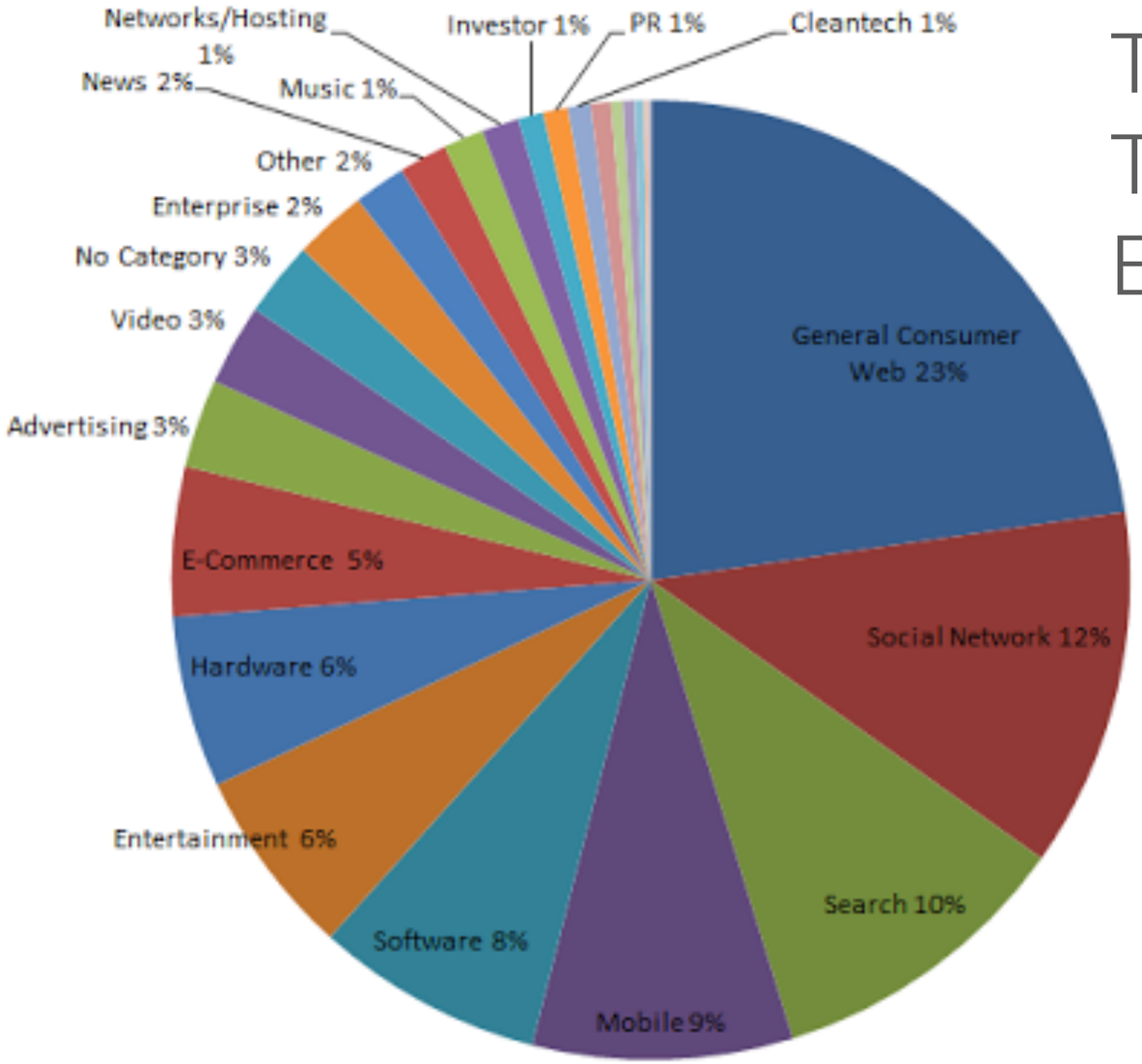
[Schutz 2007]

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

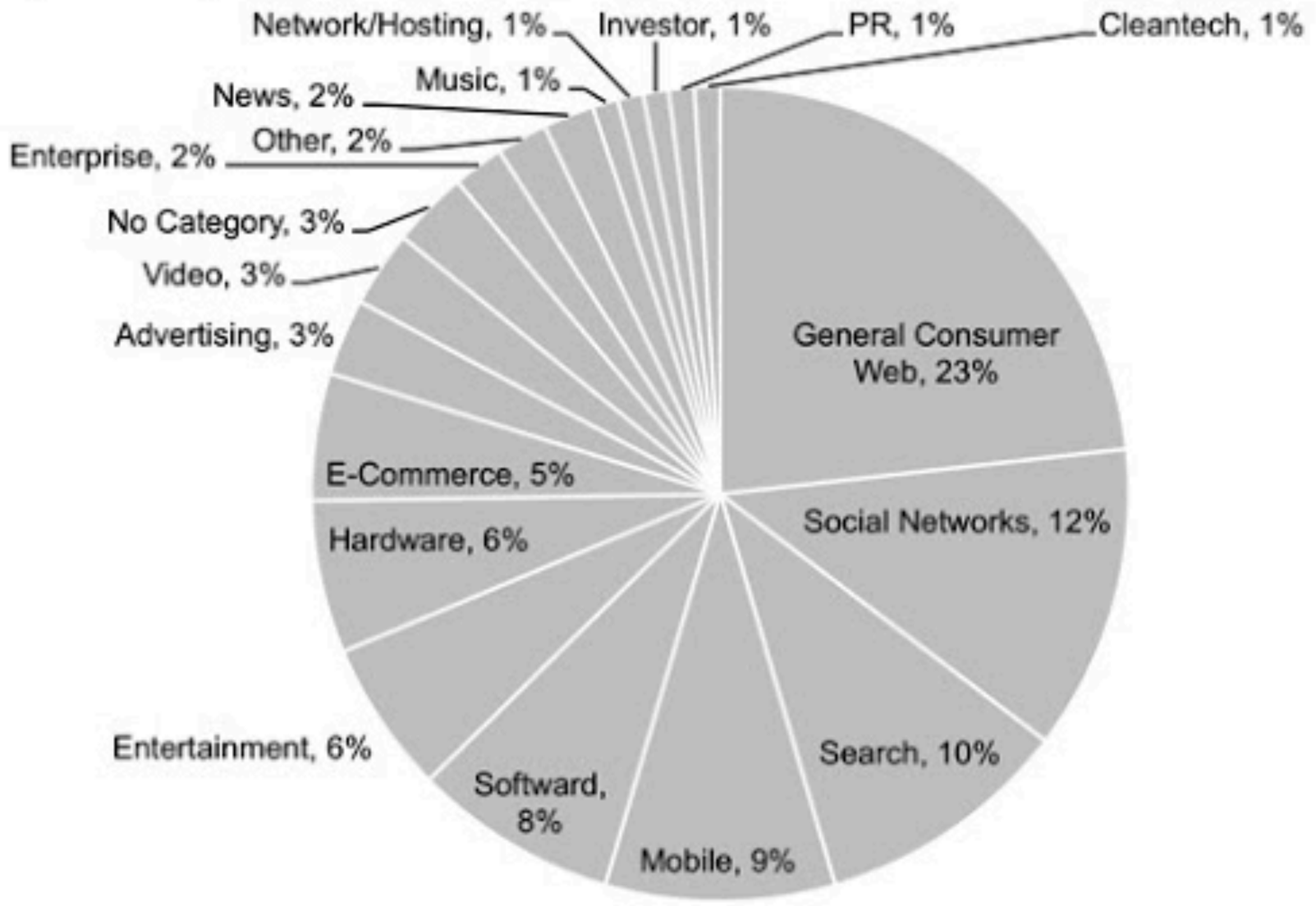


Too many slices!
Too many colors!
Bad color contract!

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

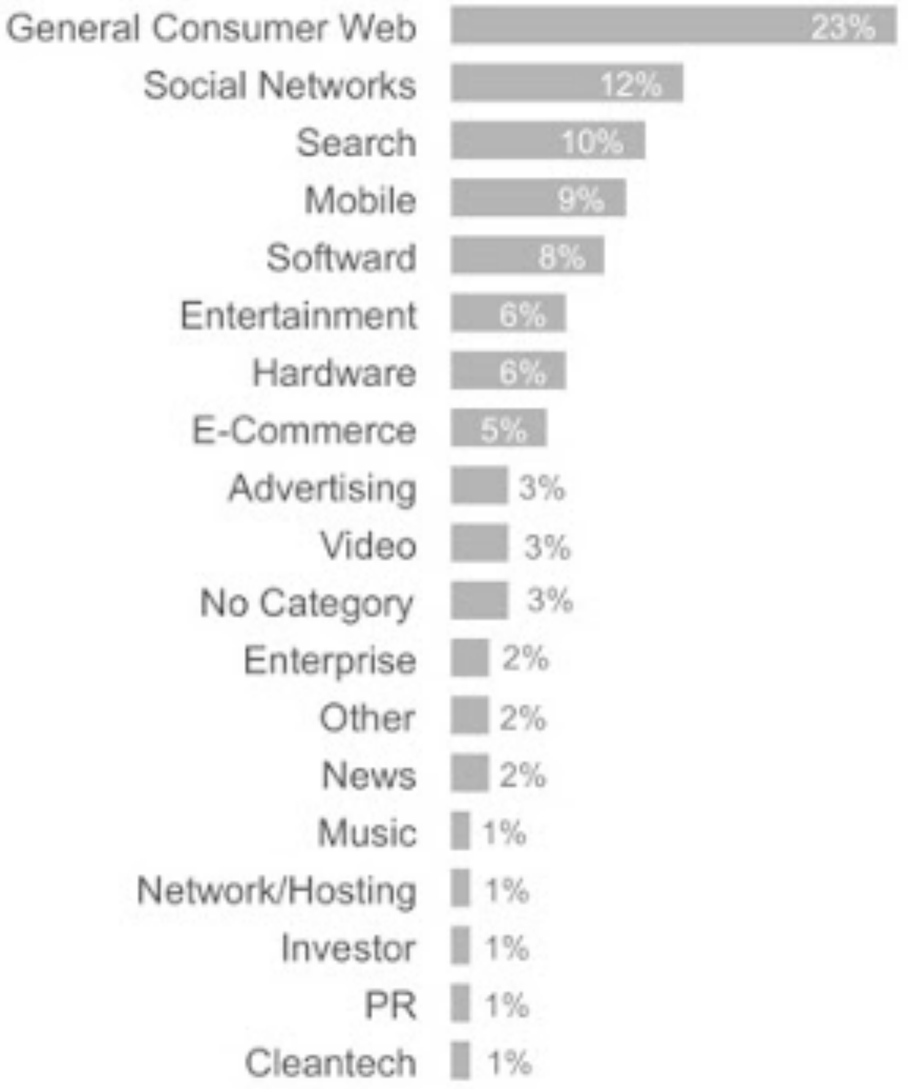
TechCrunch Coverage: 2005 - 2011

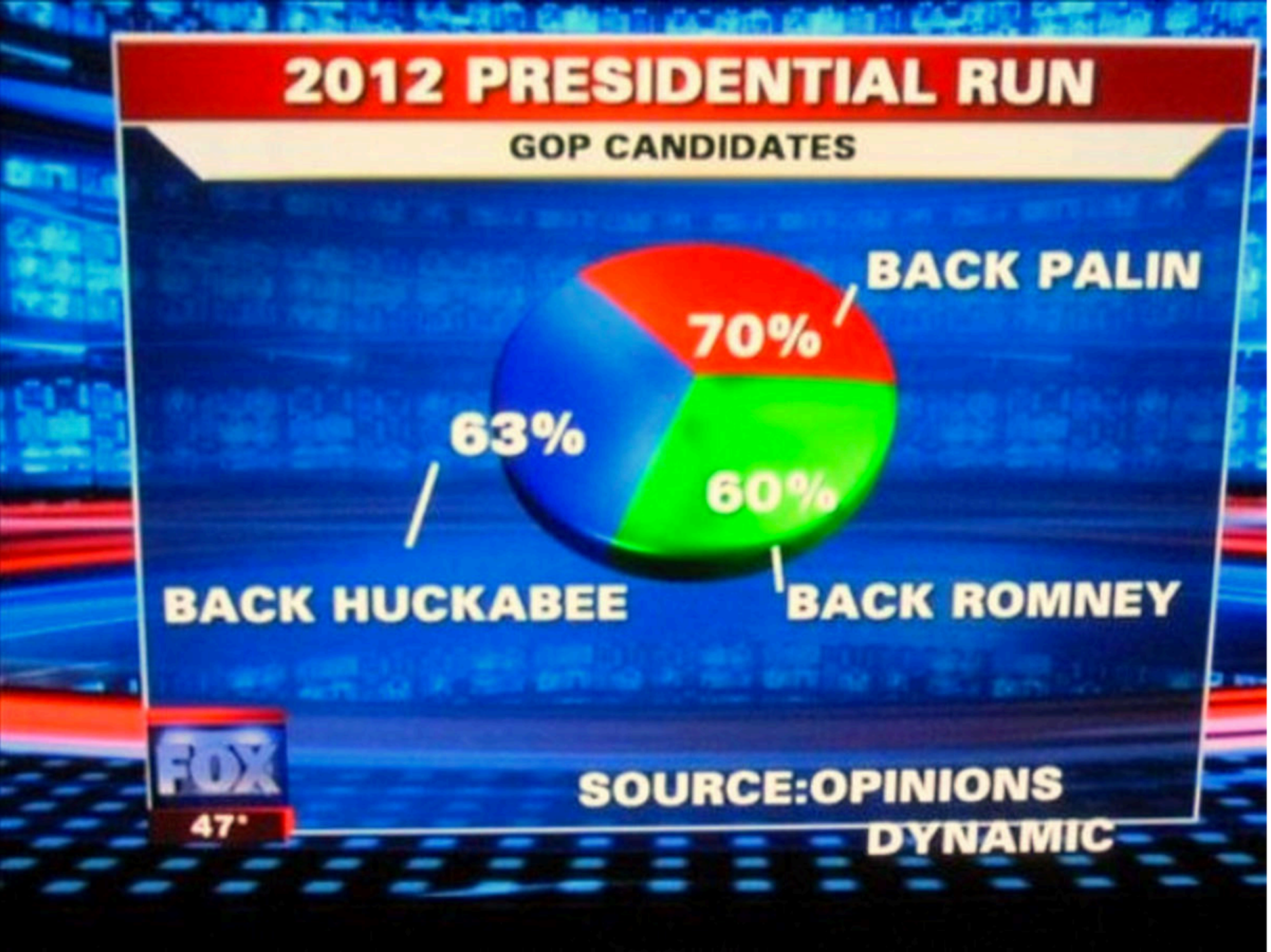
A slightly better pie?



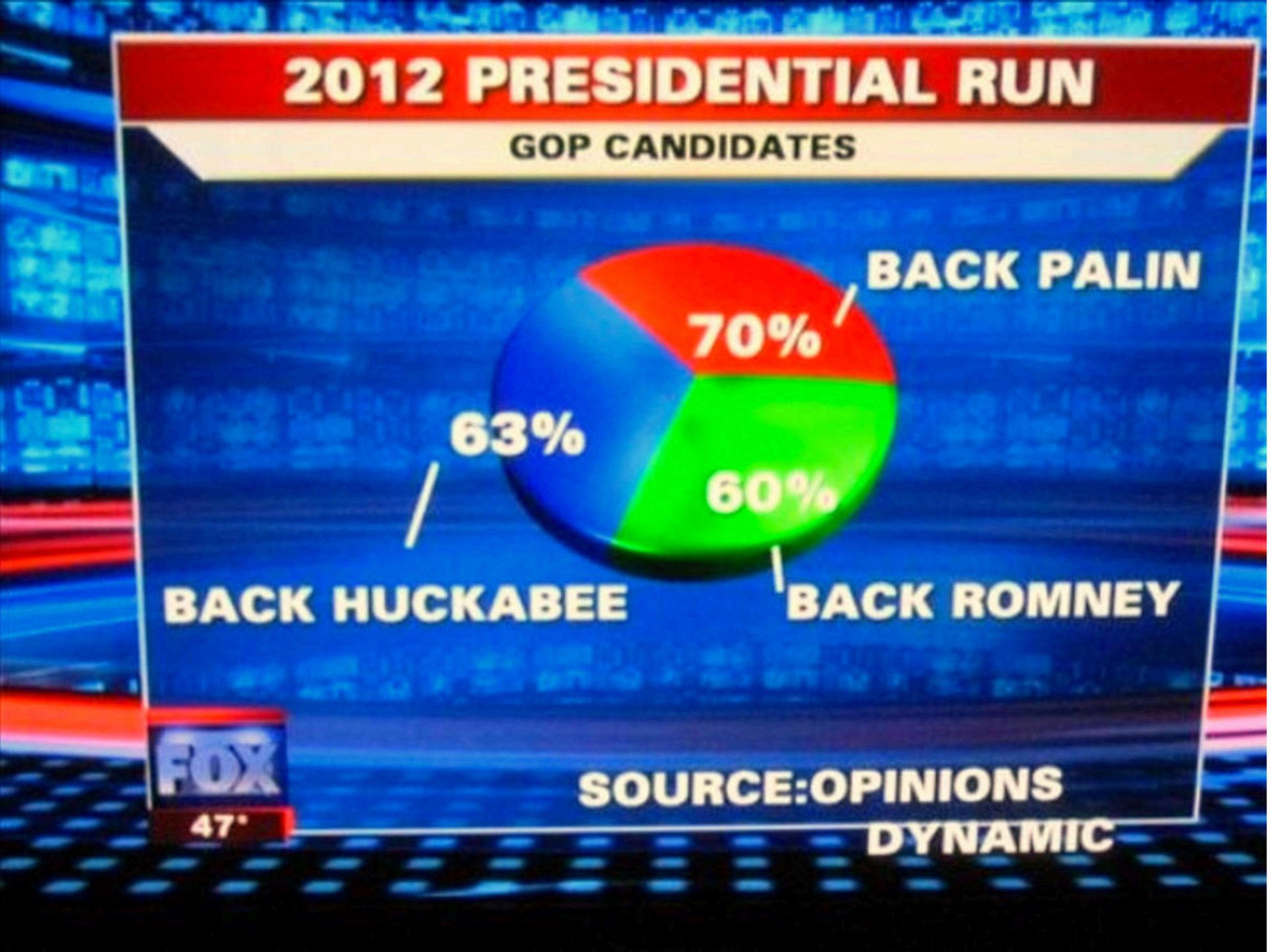
TechCrunch Coverage: 2005 - 2011

Bars are best!



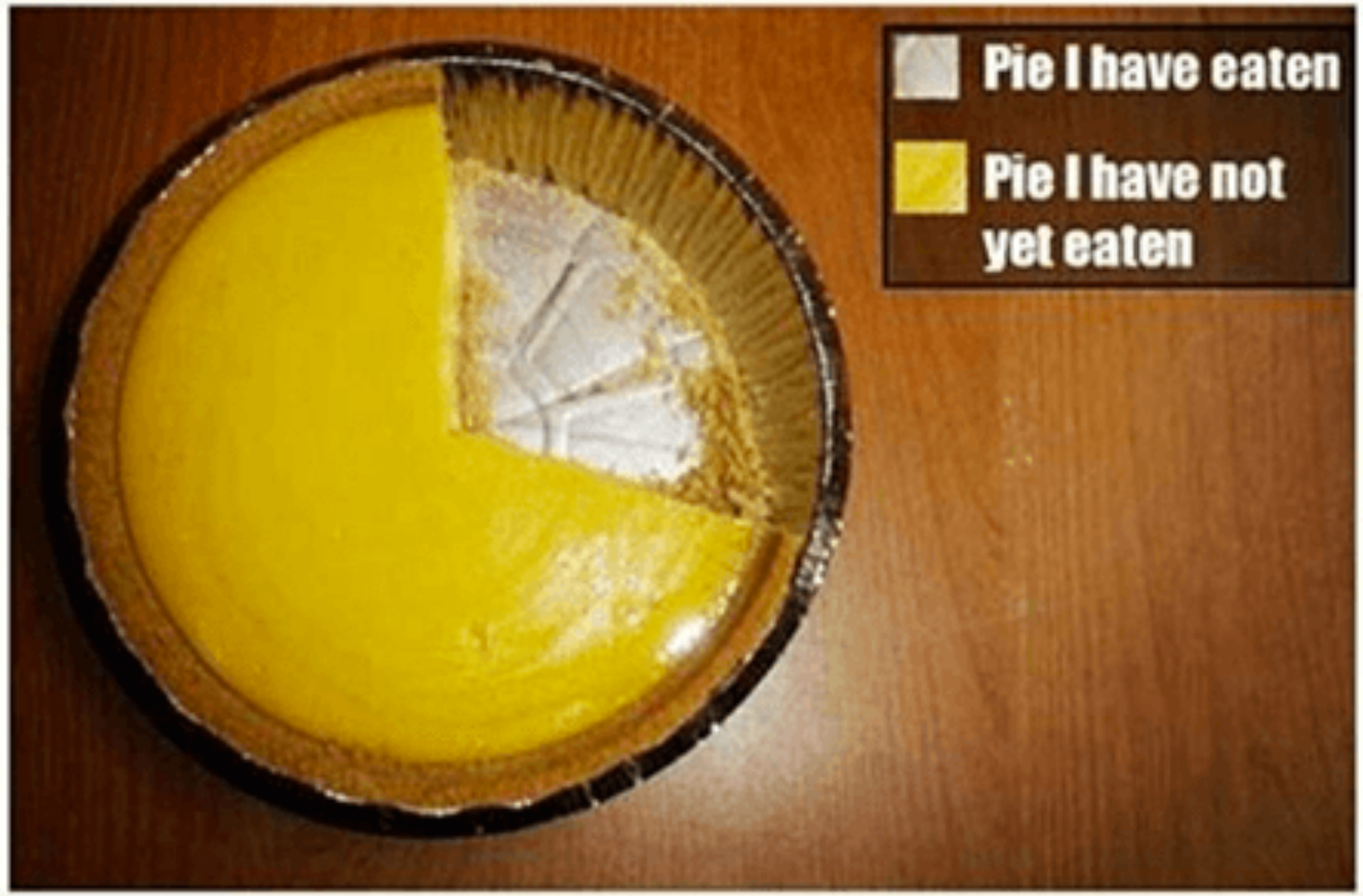

2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
47
SOURCE:OPINIONS
DYNAMIC

?!


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

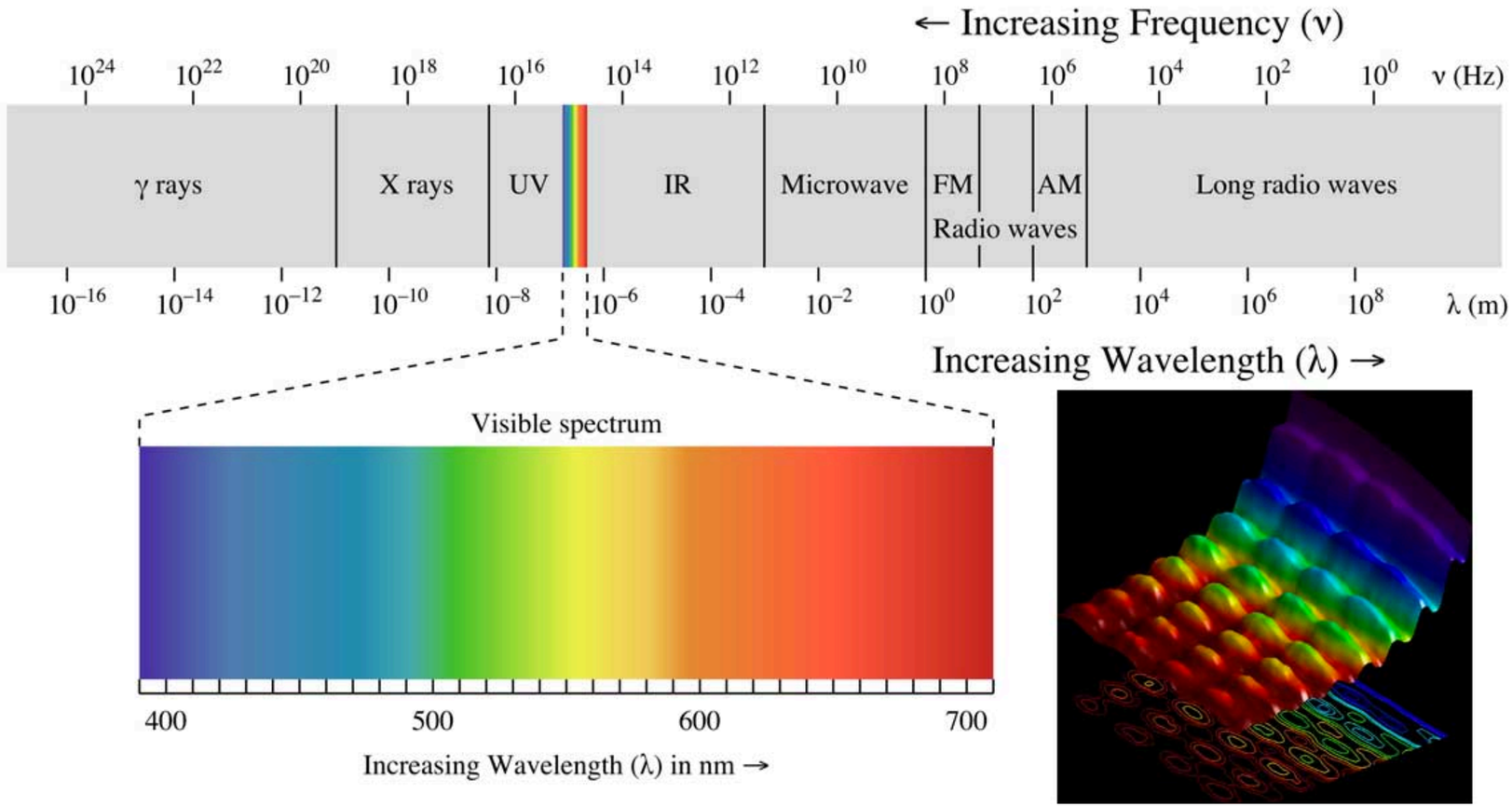


Some rules of thumb

- 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)

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/]

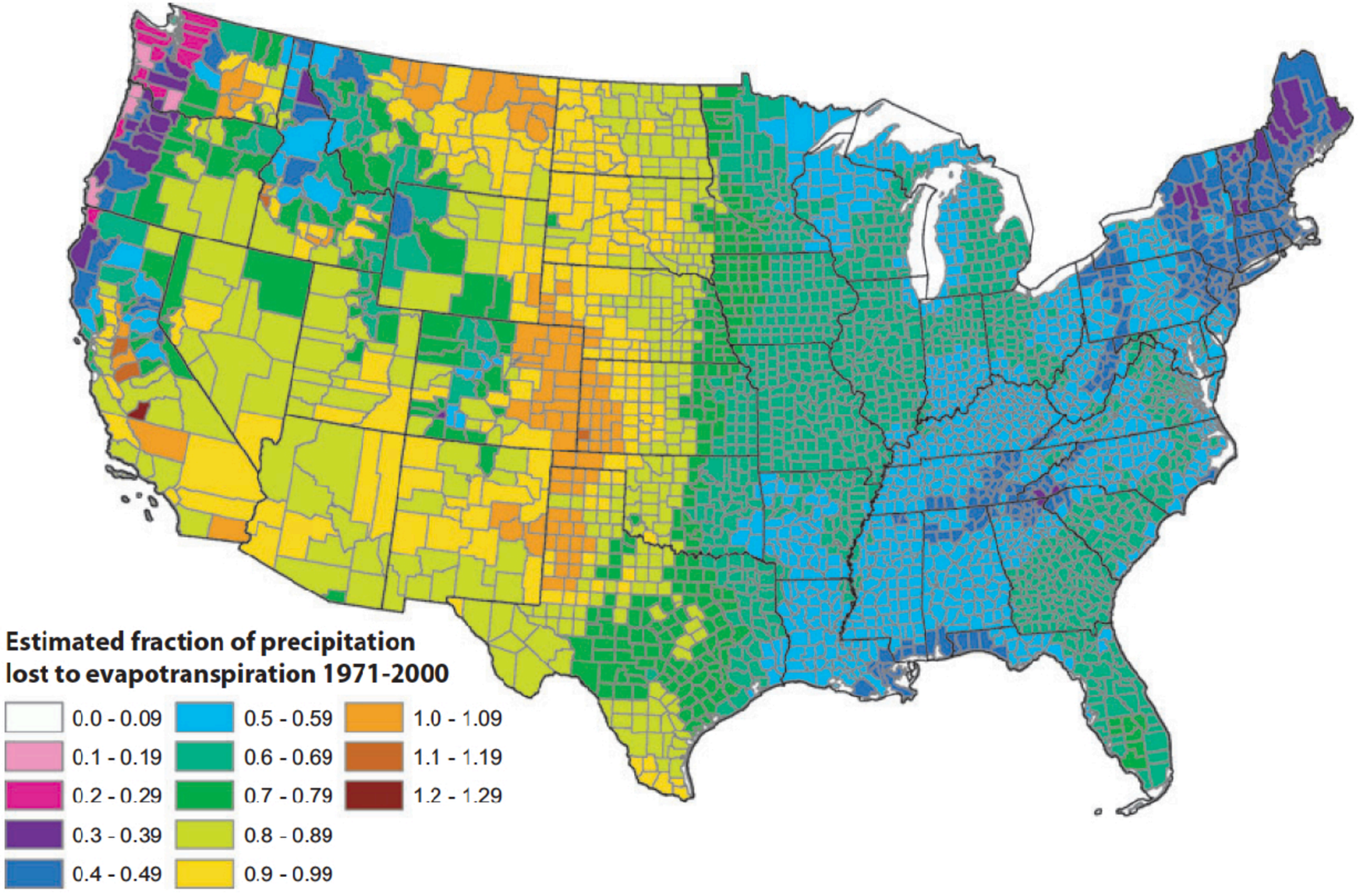


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM 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?

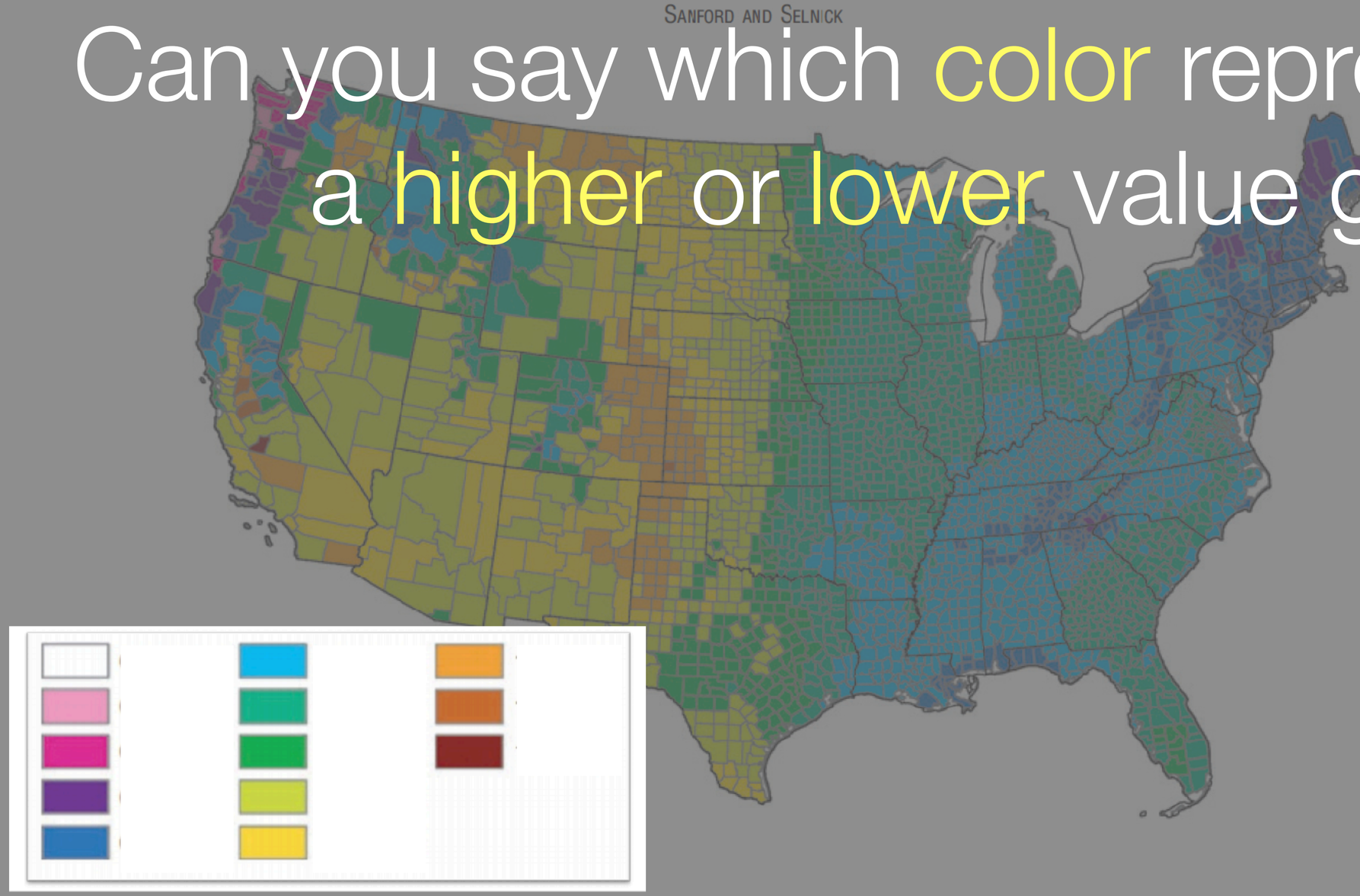
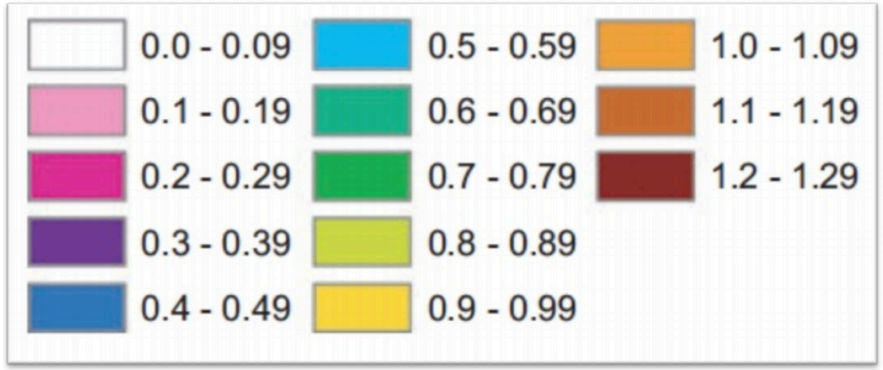


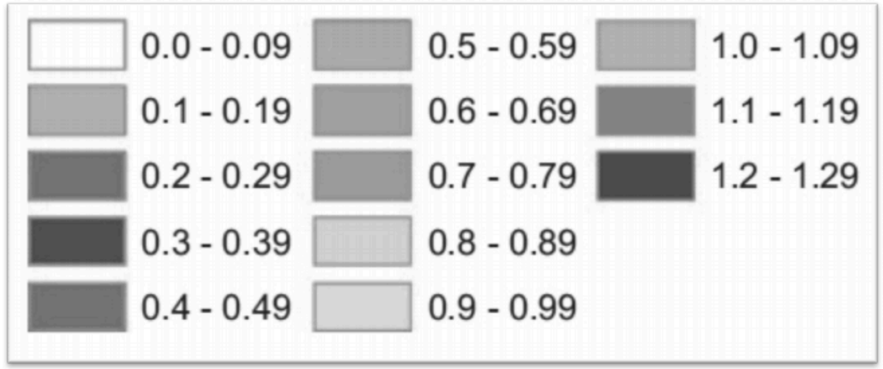
FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM 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.

Lack of perceptual ordering

Hue



Luminance



How The Rainbow Color Map Misleads (Kosara 2013)

Example: How many low ESS regions found?

ESS: Endothelial shear stress

World's Most Accurate Pie Chart

39% → 71%

10.2 sec/region → 5.6 sec/region

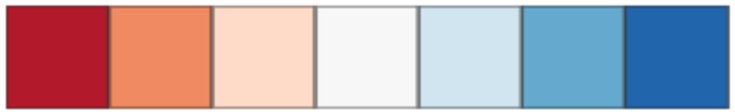
[M. Borkin et al 2011]

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**.

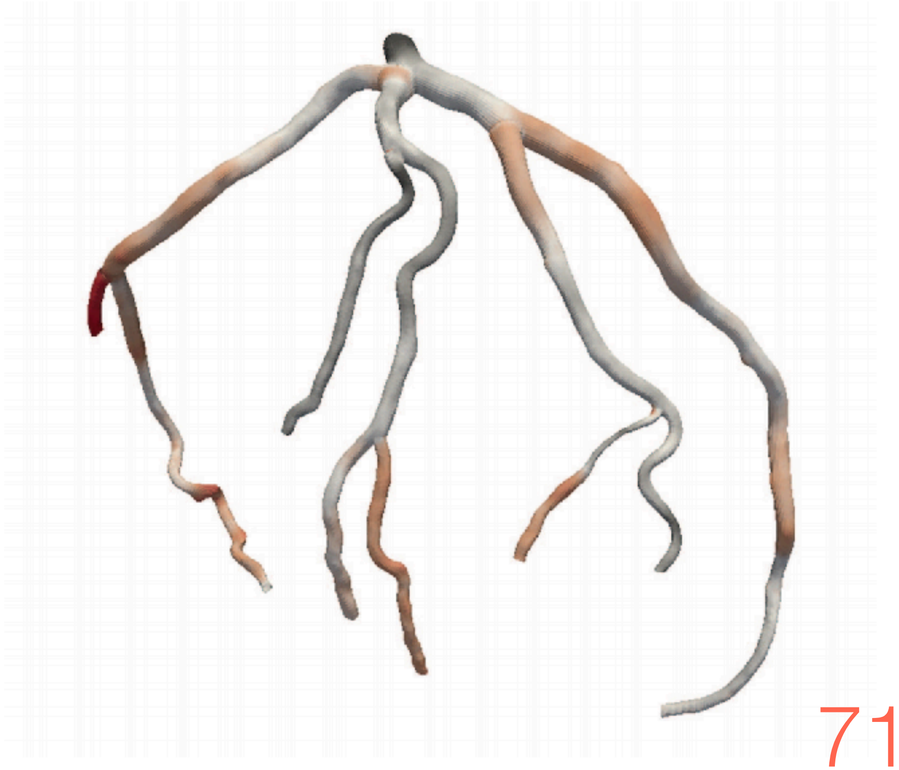


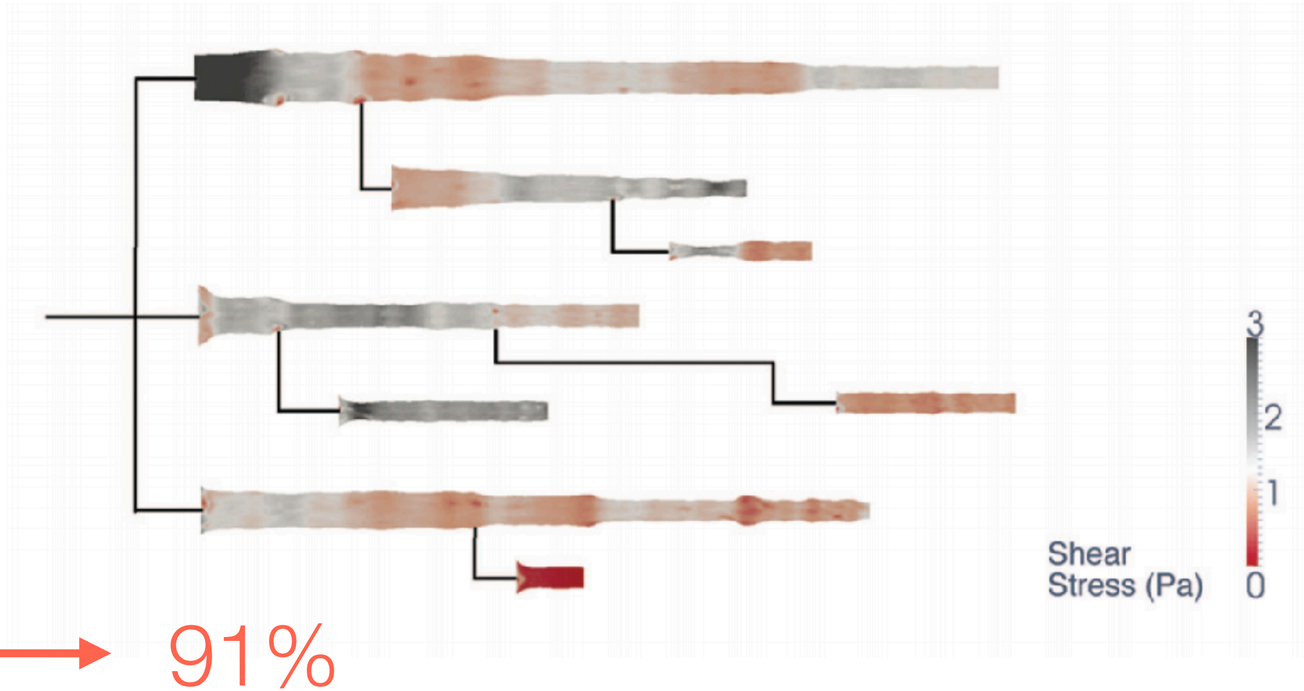
[Color Brewer]

3D charts

Perspective distorts information

2D is more accurate and efficient in identifying problematic regions.



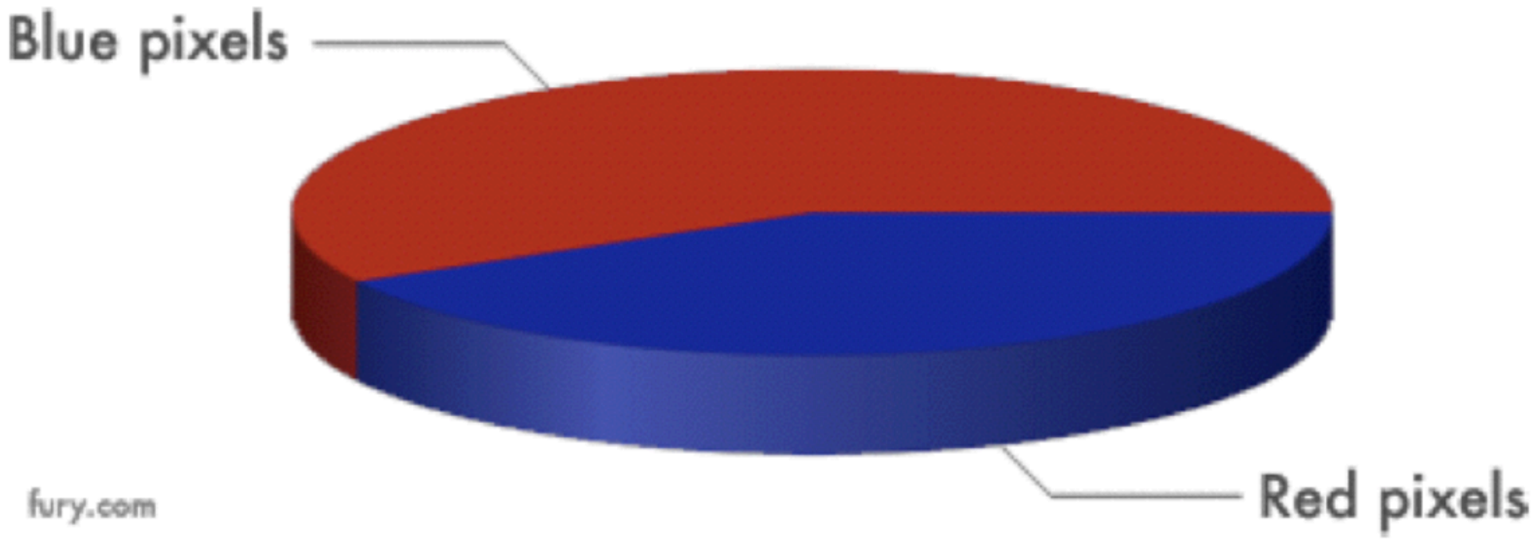


71% → 91%

5.6 sec/region → 2.4 sec/region

[M. Borkin et al 2011]

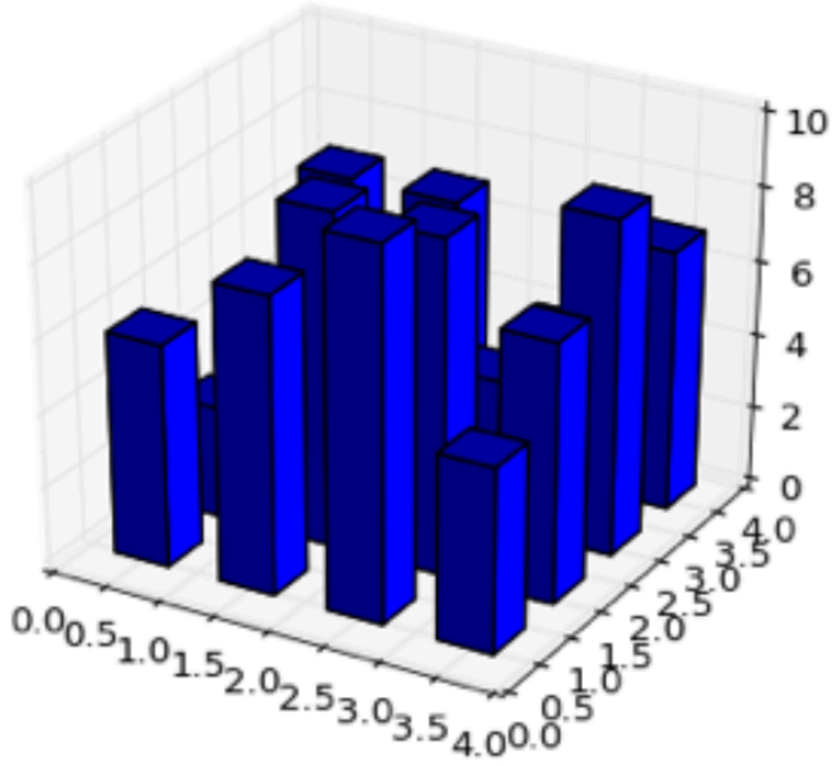
Perspective distorts information

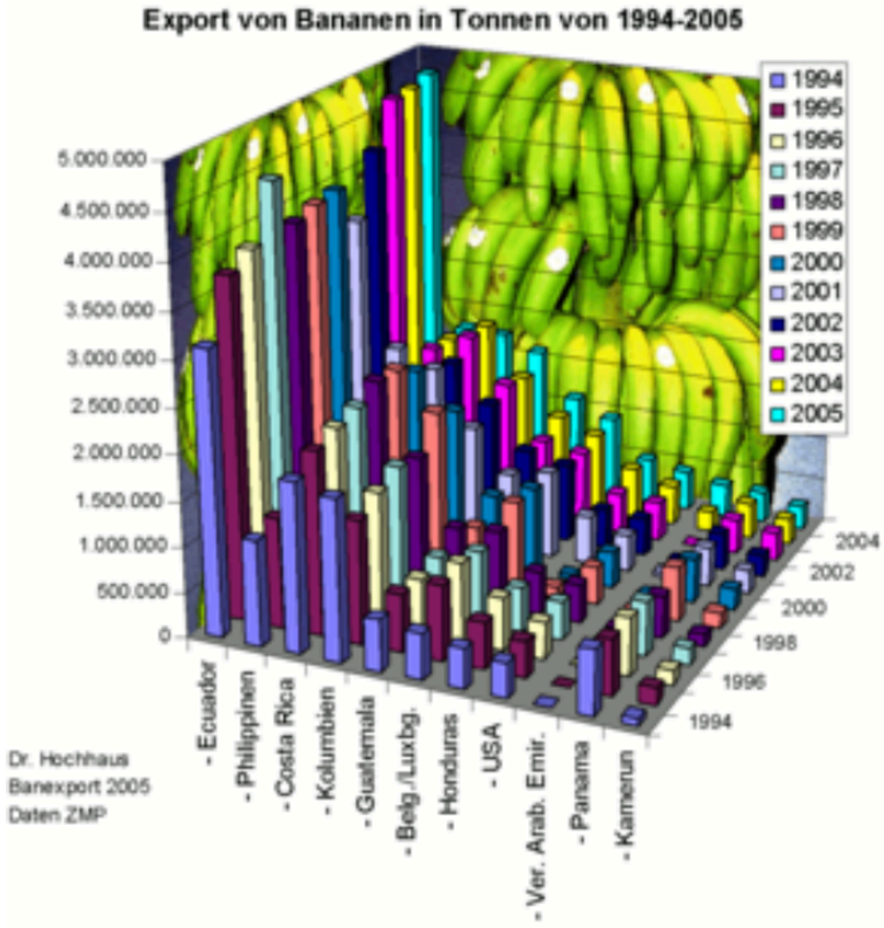


Kevin Fox

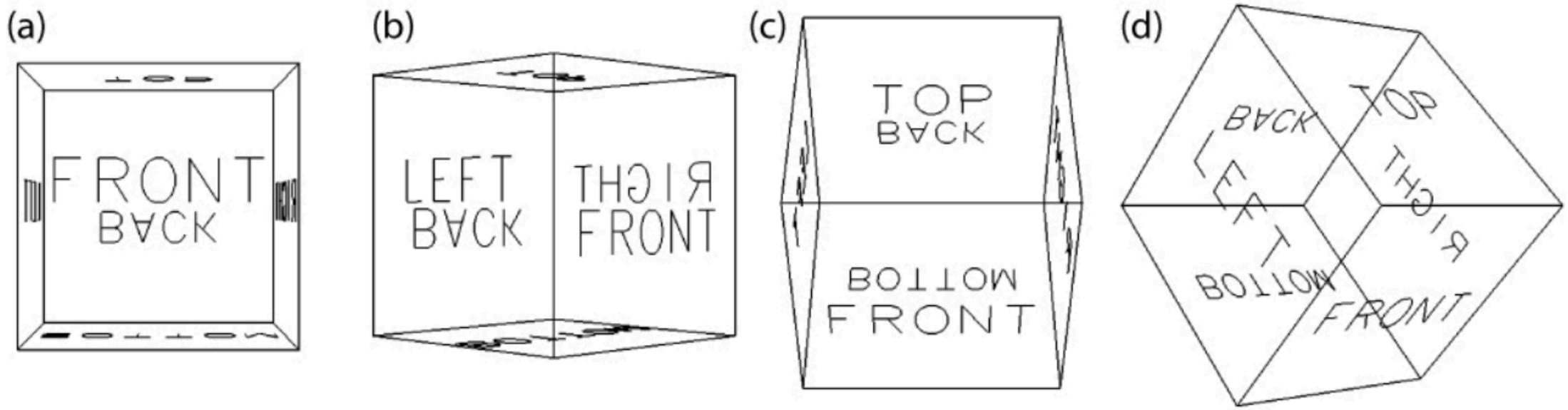
Occlusion hides information

Can rotate, but still no picture at once.



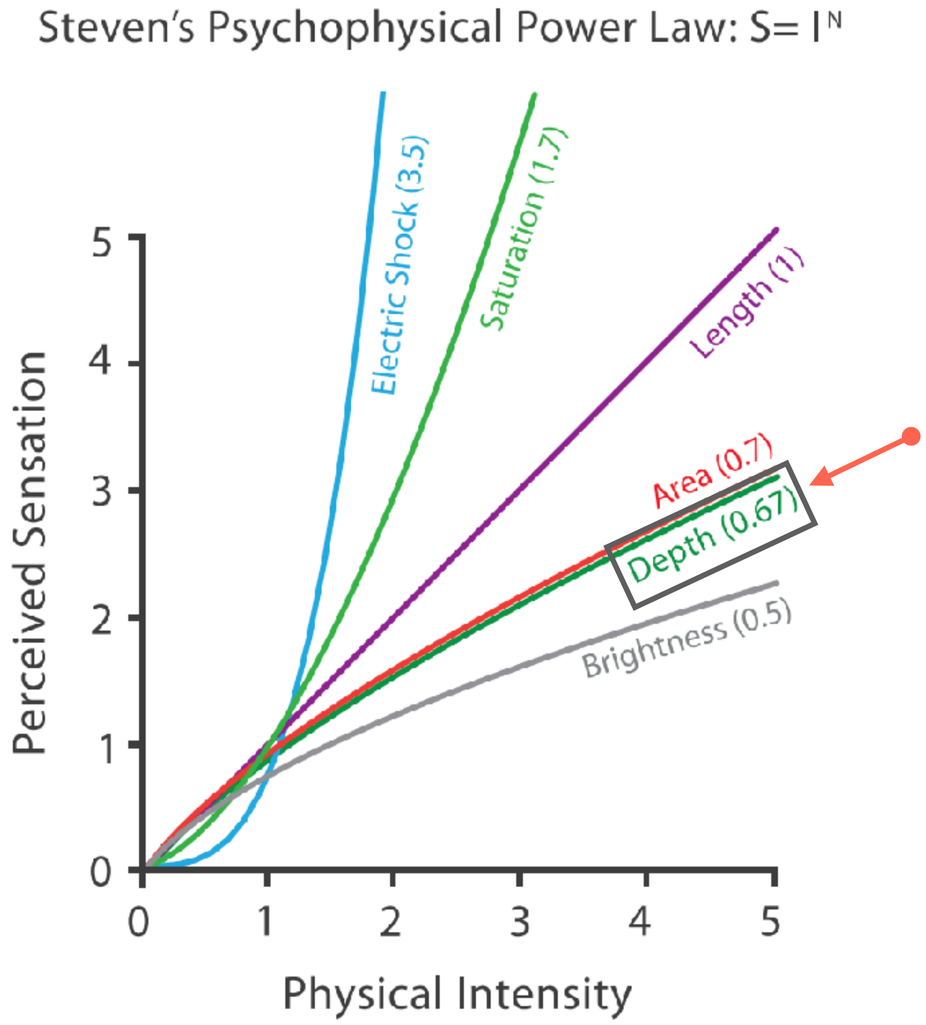


Tilted text isn’t legible



Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays (CHI 2007)

Depth judgment is bad



Actual intensity change

vs

Perceived Sensation

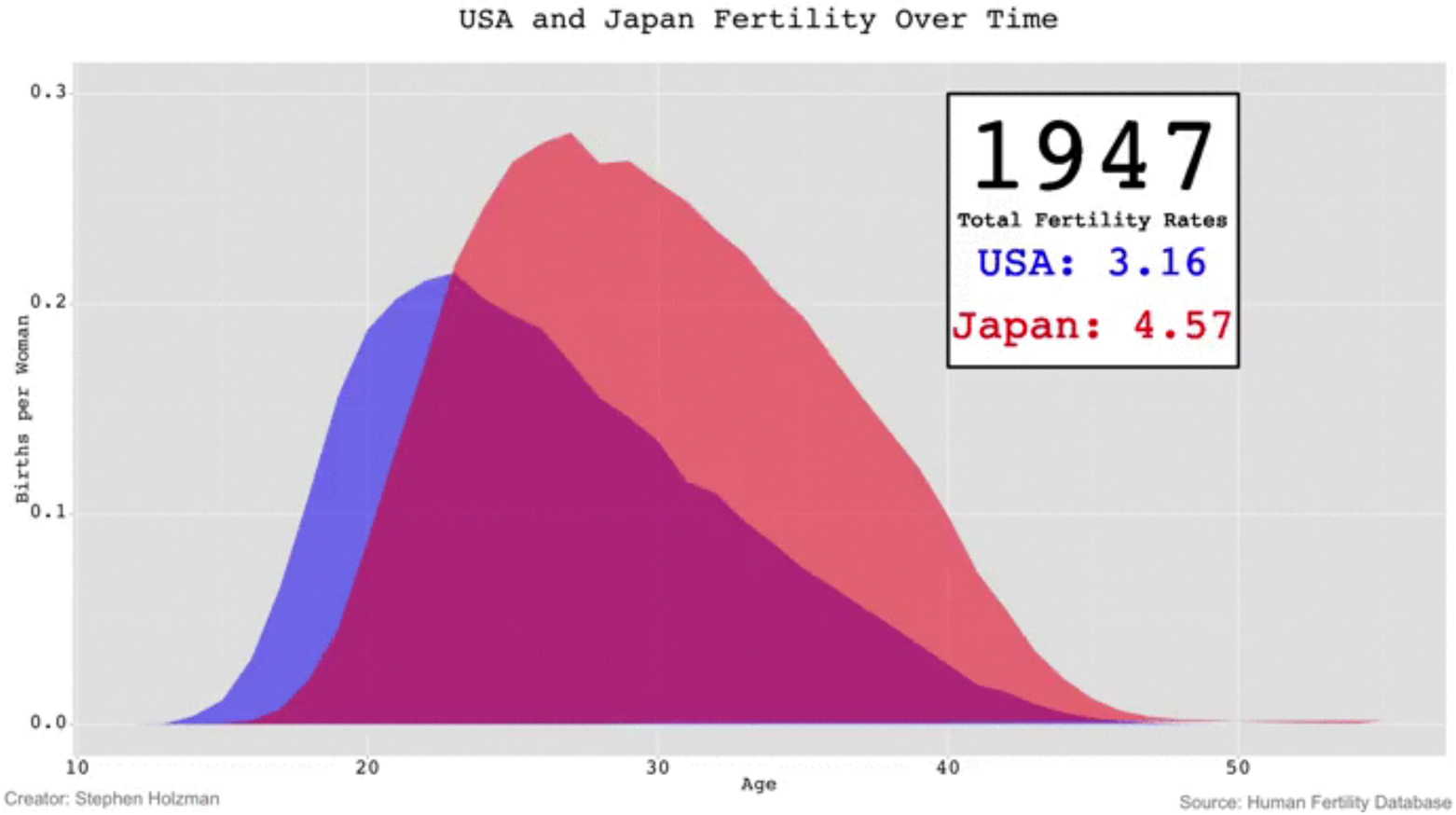
Human underestimate depth!

External cognition vs Internal cognition

Eyes beats memory

Harder to compare visible item to memory of what you saw

Animation—
narration can help

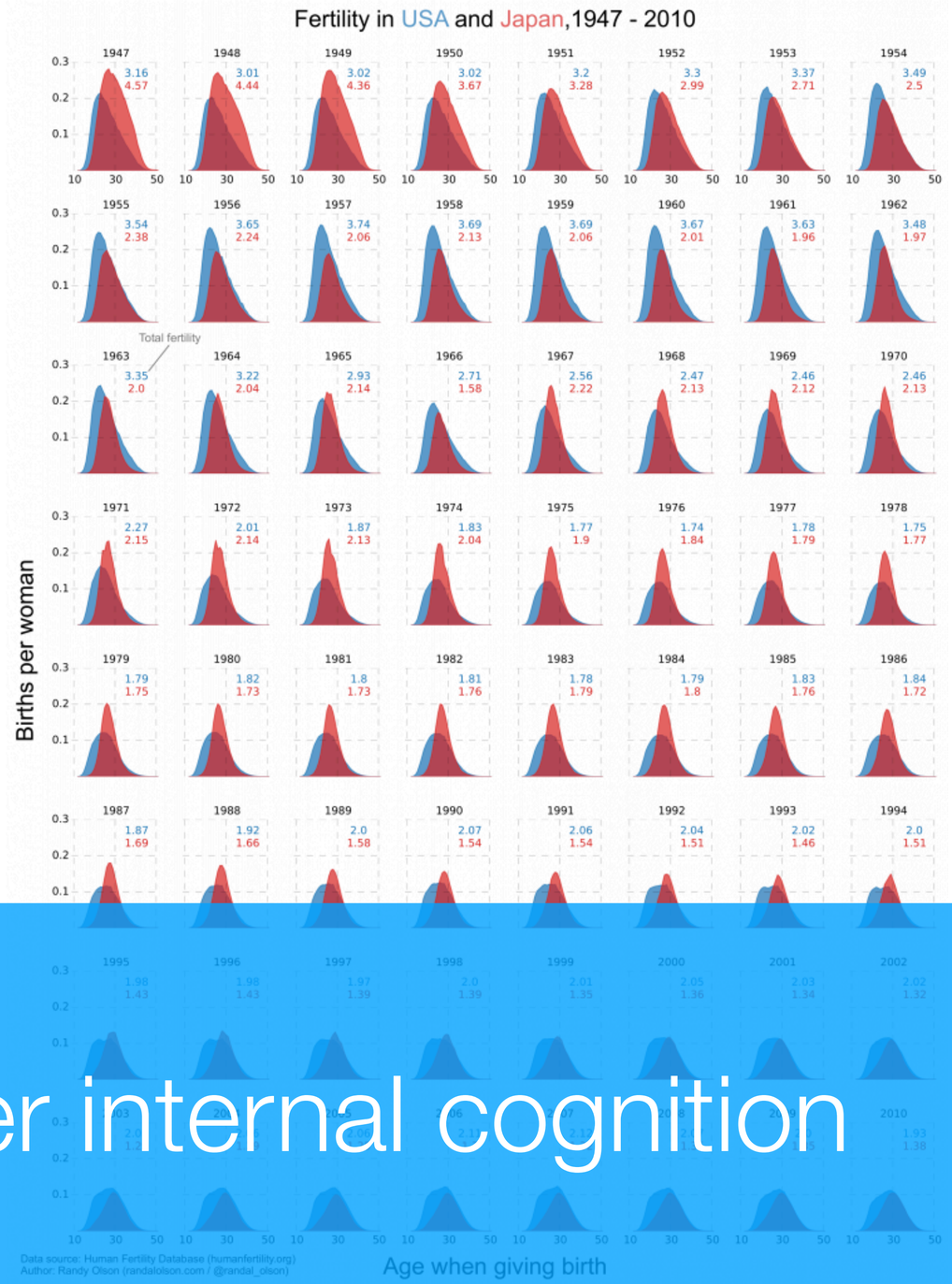


http://www.randalolson.com/2015/08/23/small-multiples-vs-animated-gifs-for-showing-changes-in-fertility-rates-over-time/

Eyes beats memory

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

Small multiples— high data density



Use external cognition over internal cognition

Tufte Principles

- Graphical integrity
- The lie factor
- Maximize data-ink ratio
- Avoid *harmful* chart junk

Other considerations

- Pie charts
- Rainbow colormap
- 3D charts
- Eyes beats memory

Subjective Dimensions

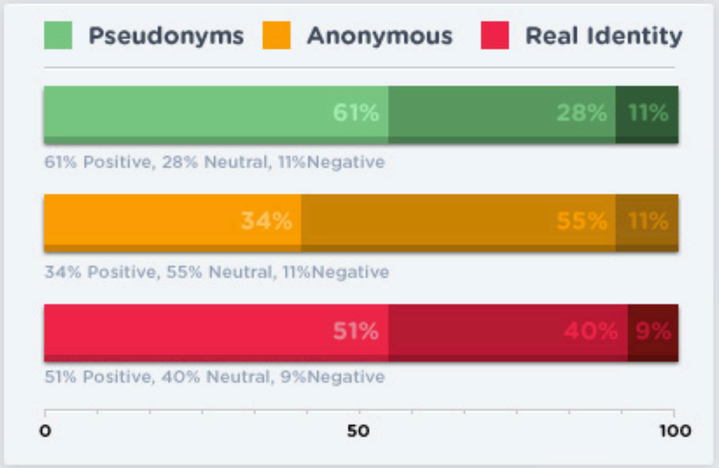
- Aesthetics: Attractive things are perceived as more useful.
- Style: Communicates brand, process, who the designer is.
- Playfulness: Encourages experimentation and exploration.
- Vividness: Can make a visualization more memorable.

Subjective Dimensions

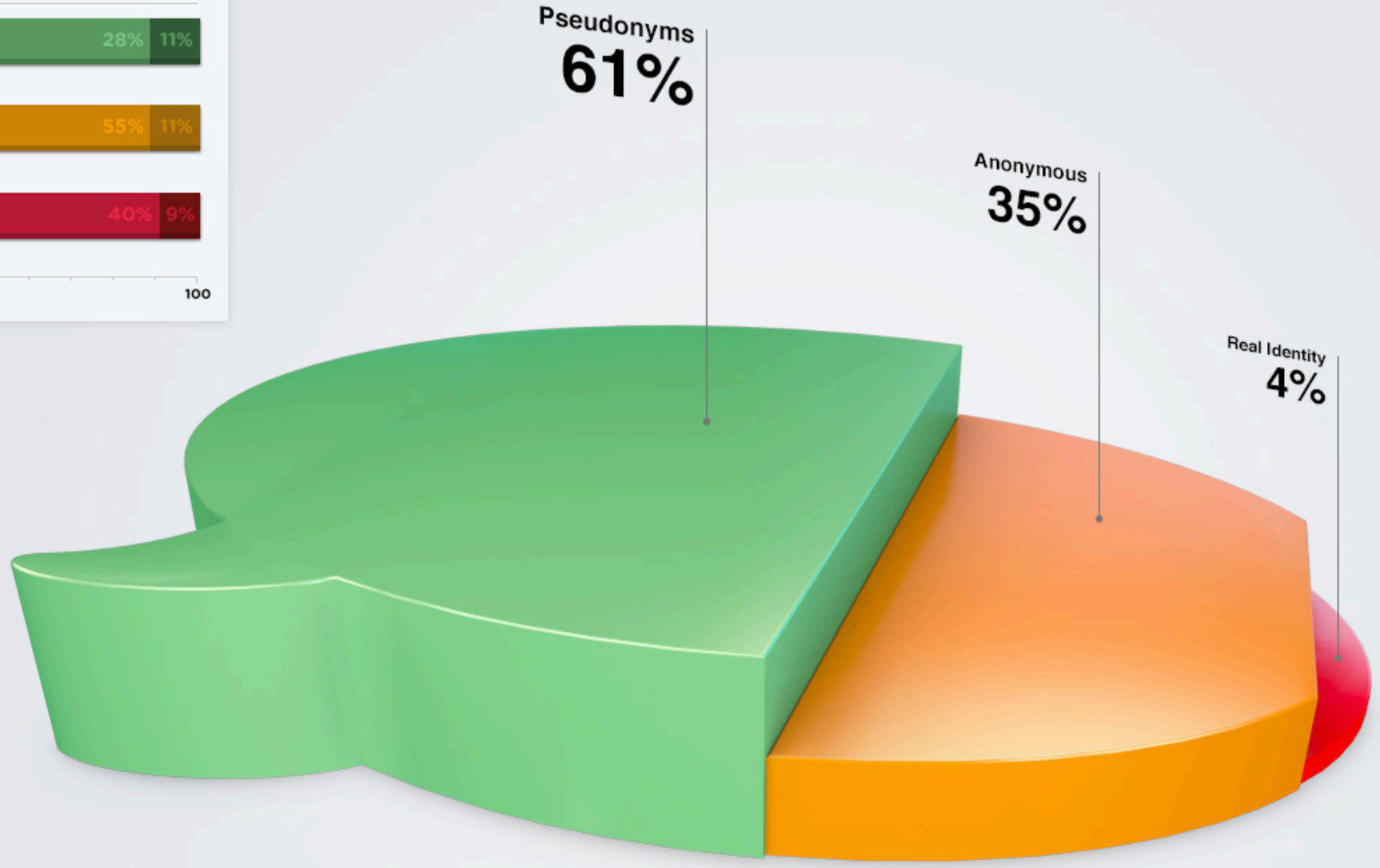
- Aesthetics: Attractive things are perceived as more useful.
- Style: Communicates brand, process, who the designer is.
- Playfulness: Encourages experimentation and exploration.
- Vividness: Can make a visualization more memorable.

Important if you want others to look at your chart!

Quality Signals by Identity



Percentage of Comments by Identity



May 3, 2008

SIGN IN TO E-MAIL OR SAVE THIS | FEEDBACK

All of Inflation's Little Parts

Each month, the Bureau of Labor Statistics gathers 84,000 prices in about 200 categories — like gasoline, bananas, dresses and garbage collection — to form the Consumer Price Index, one measure of inflation.

It's among the statistics that the Federal Reserve considered when it cut interest rates on Wednesday. The categories are weighted according to an estimate of what the average American spends, as shown below.

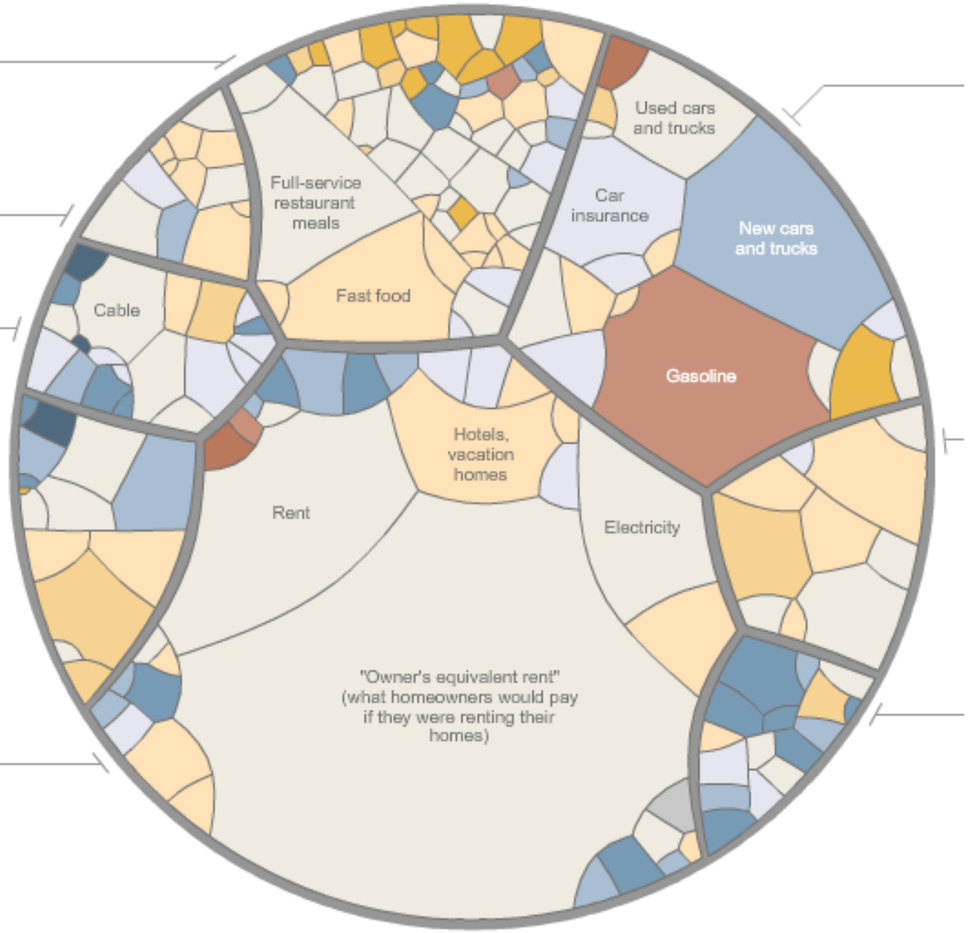
An Average Consumer's Spending

Each shape below represents how much the average American spends in different categories. Larger shapes make up a larger part of spending.

Color shows change in prices from March 2007 to March 2008

-10% -2 0 +2 +4 +6 +10 +20 +40%

ZOOM IN ZOOM OUT



Food and beverages 15%
The high price of oil is a factor that has made food prices rise quickly.

Miscellaneous 3%

Recreation 6%

Education/Communication 6%
Cellphones were added to the index in 1997. Because the Consumer Price Index can be slow to add new goods, which are often cheaper, it may overstate parts of inflation.

Housing 42%
In the C.P.I., home ownership costs track rent prices more closely than housing prices. This means inflation may have been understated when home prices were rising faster than rents.

Transportation 18%
Gas is 5.2 percent of spending nationwide, but only 3.8 percent in the New York area.

Health care 6%
As a group, the elderly spend about twice as much of their budget on medical care.

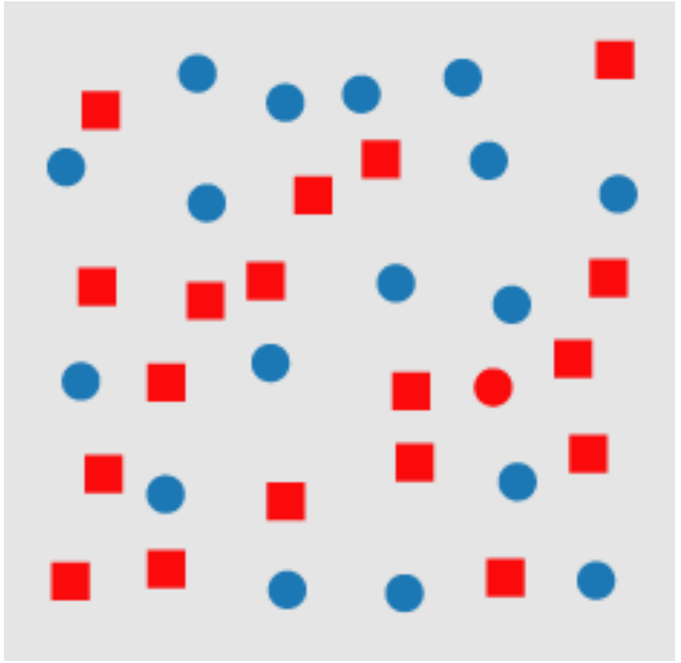
Apparel 4%
The ratio of spending on women's clothes to that on men's clothes is about 2 to1.

Sources: Bureau of Labor Statistics; Michael Balzer, University of Konstanz (Germany)

Matthew Bloch, Shan Carter and Amanda Cox/The New York Times

Next

Graphical Perception



Where is a red circle?

10 min break