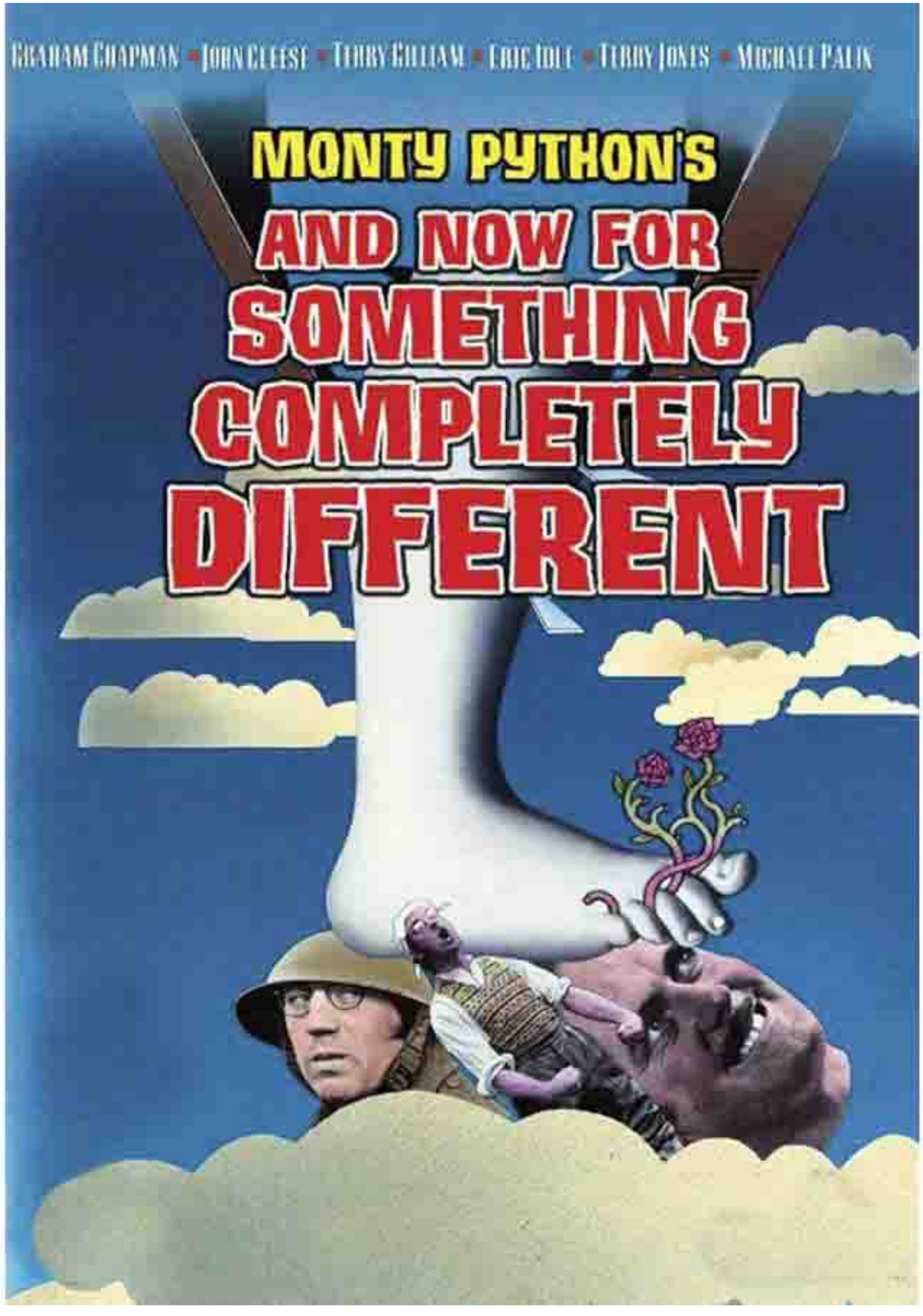
# Effective Visualizations

Hanspeter Pfister
pfister@seas.harvard.edu


GRAHAM CHAPMAN • JOHN CLEESE • TERRY GILLIAM • ERIC IDLE • TERRY JONES • MICHAEL PALIN
MONTY PYTHON'S
AND NOW FOR
SOMETHING
COMPLETELY
DIFFERENT

1971

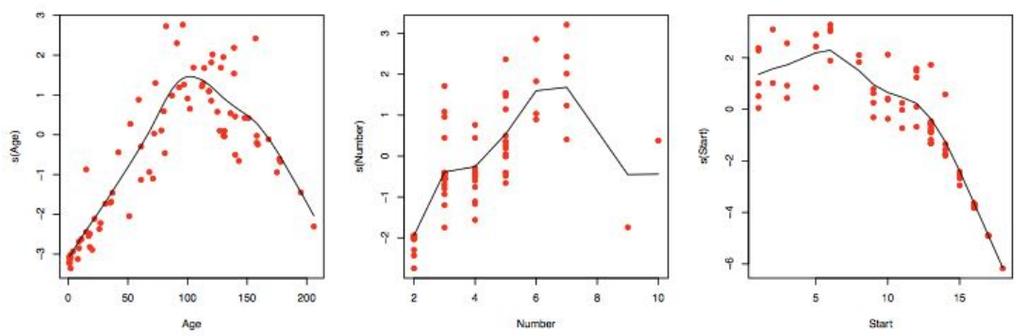
# Visualization

To convey insights through graphical representations of data

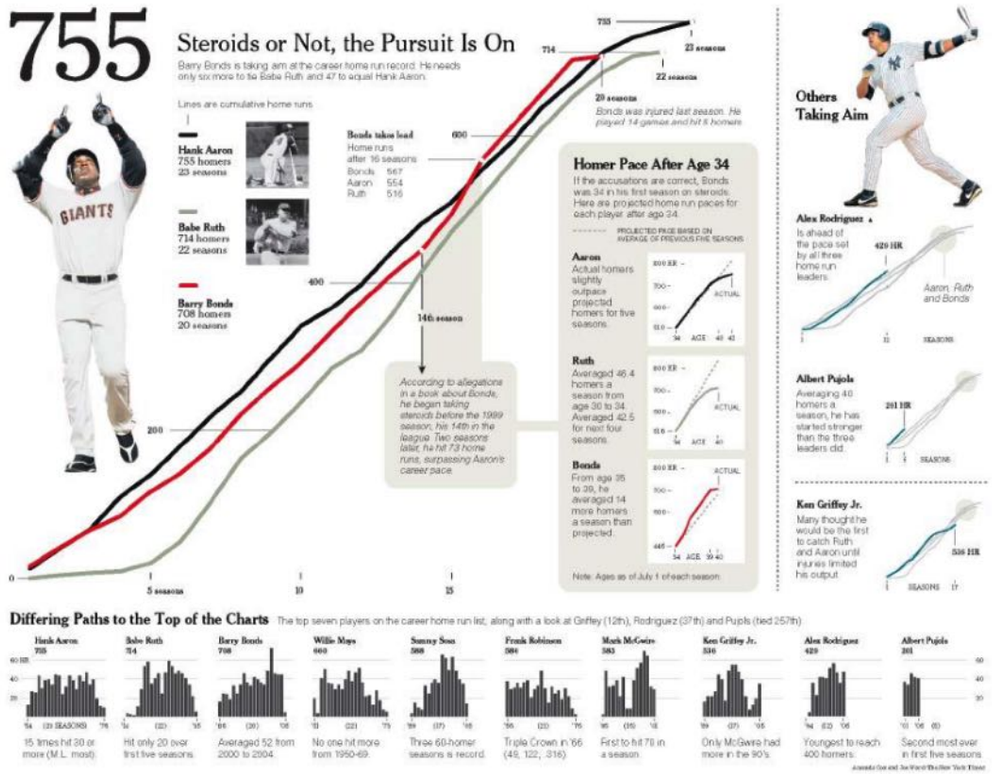
# Visualization Goals

**Data Exploration** - find the unknown

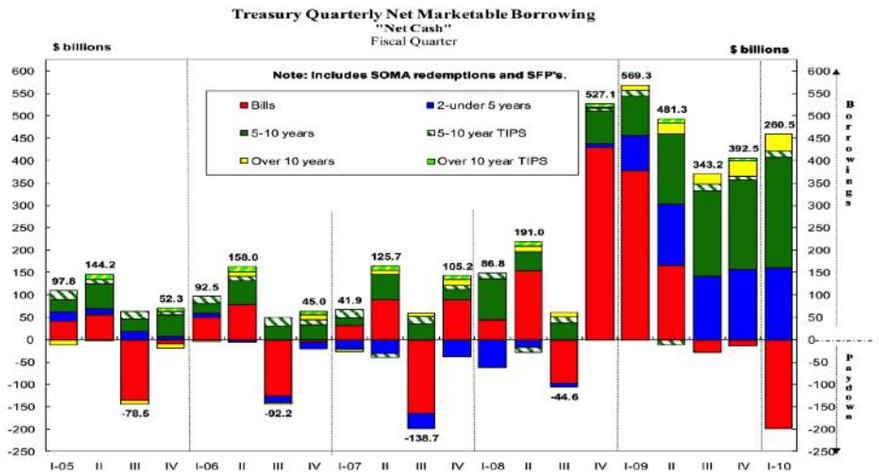
**Data Presentation** - communicate and disseminate




```
> summary(kyph1.gam)

(Dispersion Parameter for binomial family taken to be 1)

    Null Deviance: 83.2345 on 80 degrees of freedom
Residual Deviance: 40.526 on 67.9997 degrees of freedom
AIC: 66.5266

Number of Local Scoring Iterations: 11

DF for Terms and Chi-squares for Nonparametric Effects

            Df Npar Df Npar Chisq P(Chi)
(Intercept)  1
s(Age)       1       3     5.7937 0.1221
```



# Unfortunately…



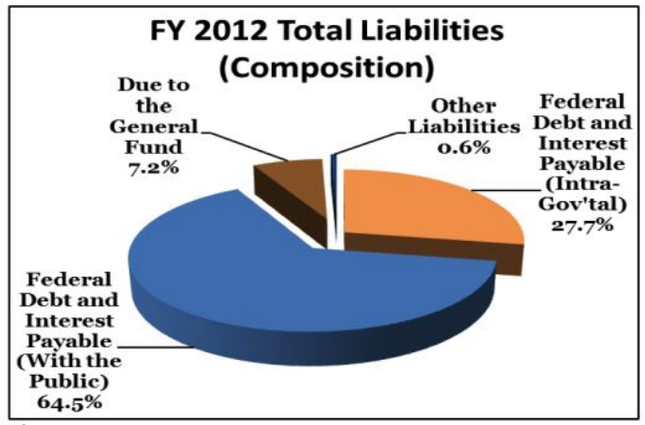
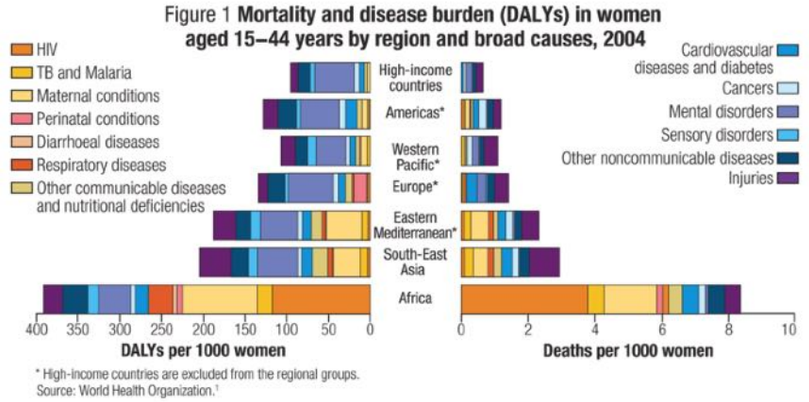
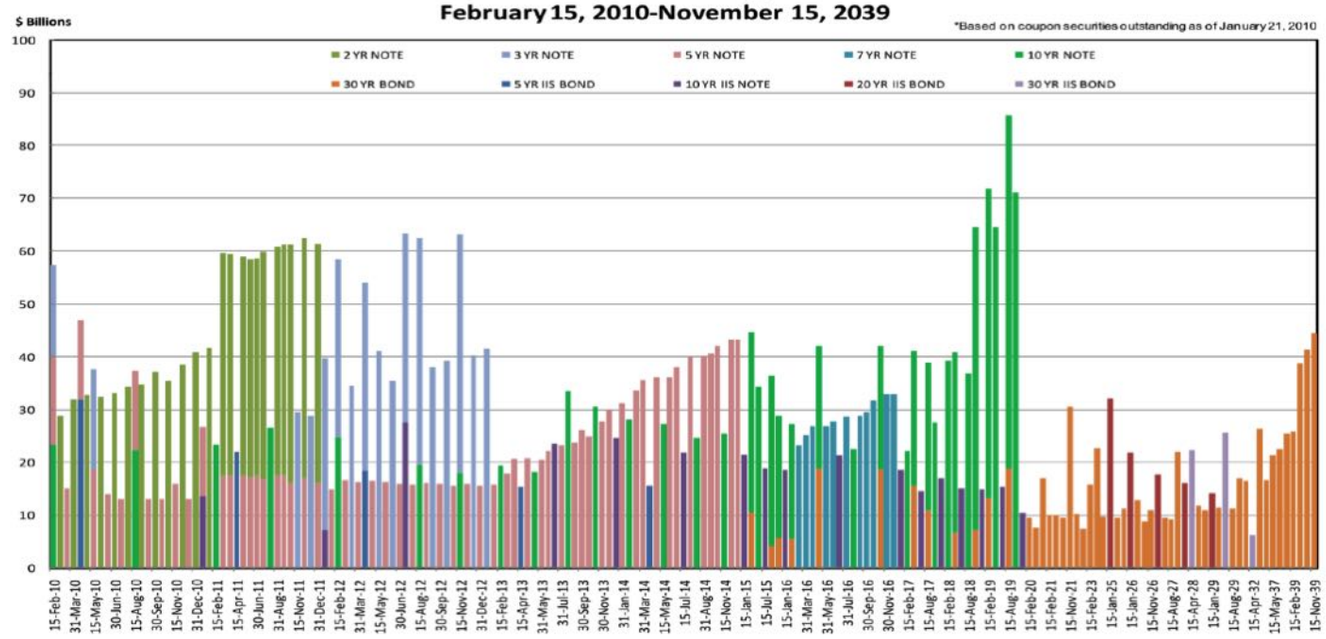




Figure 10



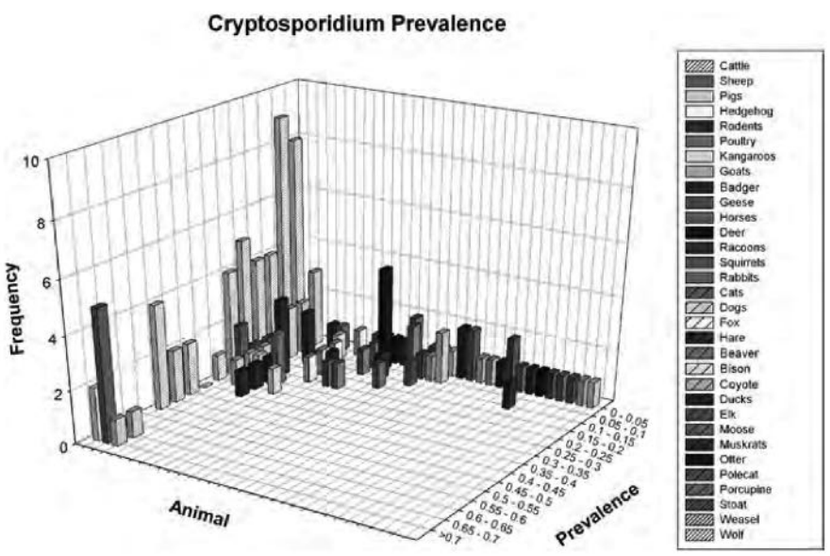



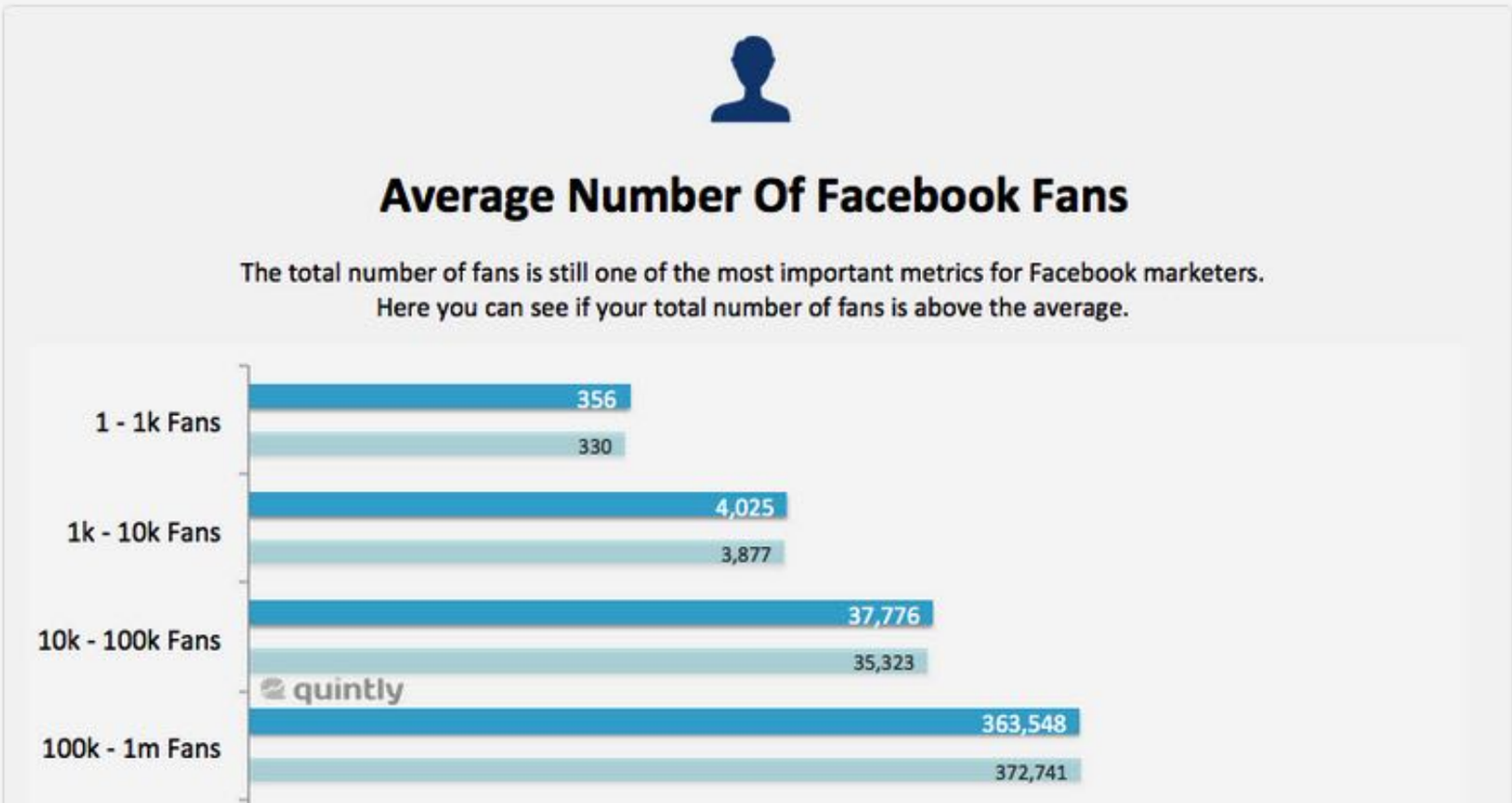

**Figure 5.2** Mean prevalence rates of *Cryptosporidium* oocysts by animal species.

Sources: US Treasury and WHO reports

http://viz.wtf

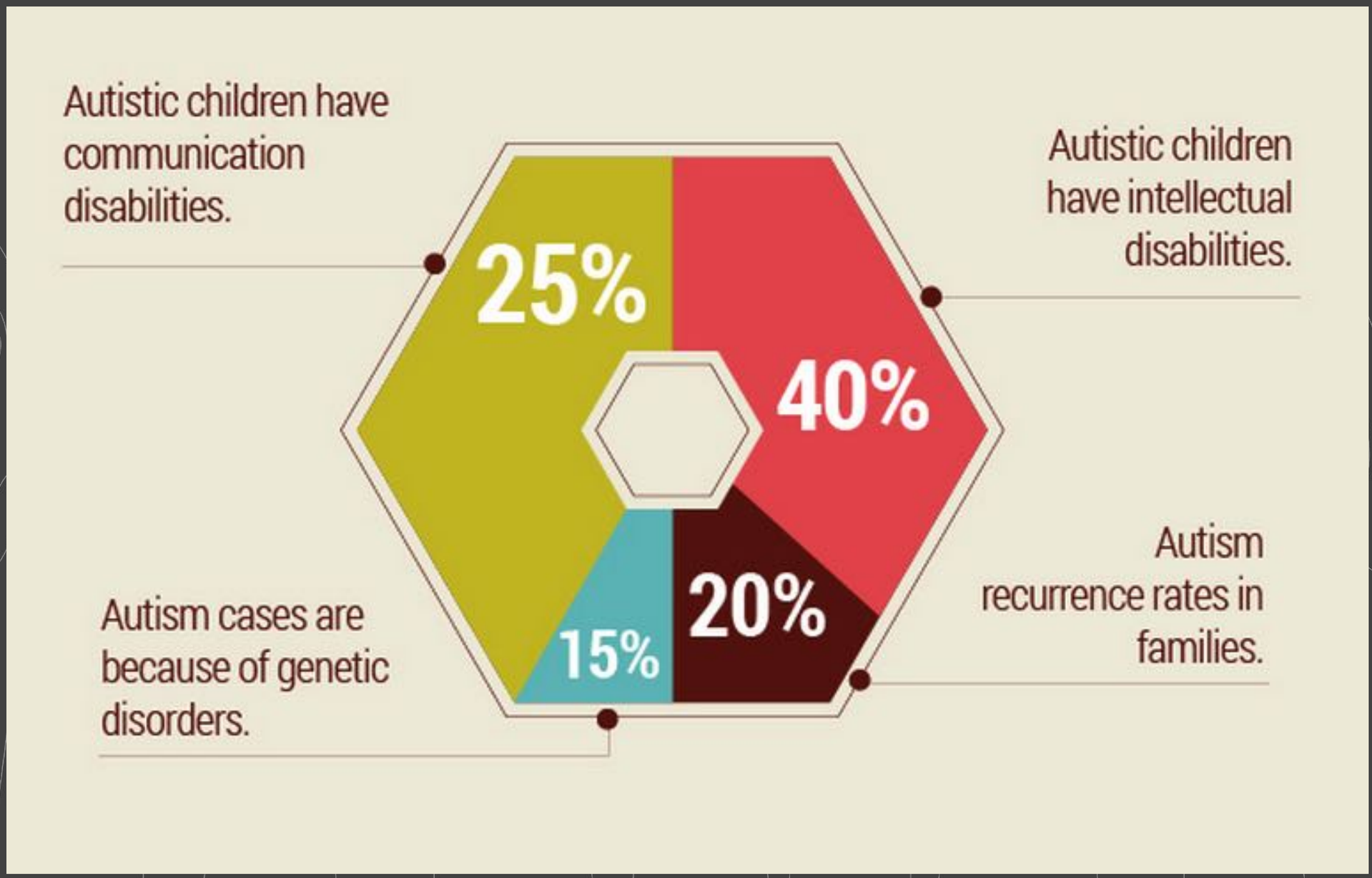
# WTF Visualizations

Visualizations that make no sense.
For a discussion of what is wrong with a particular visualization, tweet at us @WTFViz.
Check out our friends Thumbs Up Viz and accidental aRt, or submit.



# Activity

What do you like / dislike about this visualization?



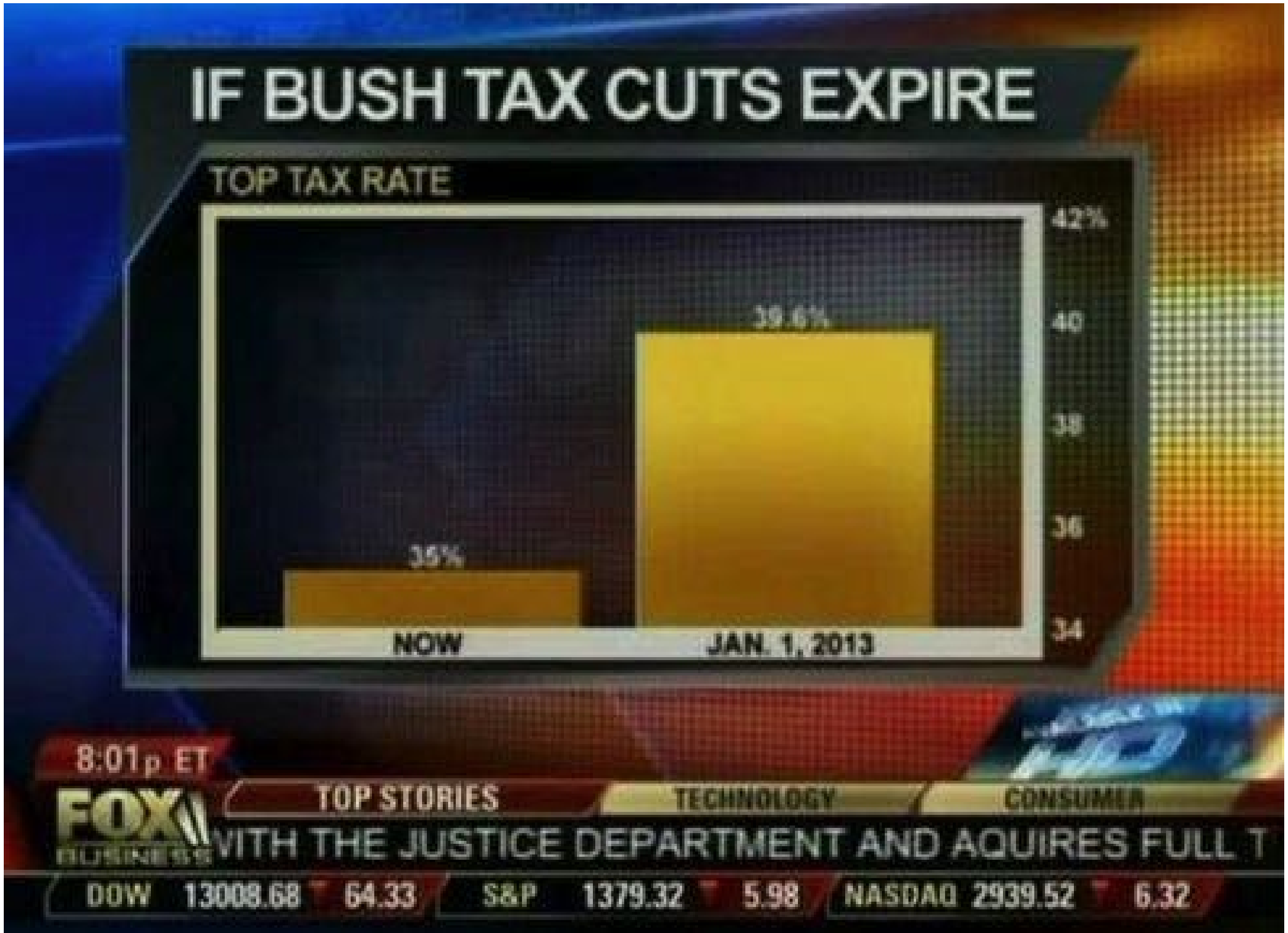

http://wtfviz.net/

# Effective Visualizations

1. Have graphical integrity
2. Keep it simple
3. Use the right display
4. Use color strategically
5. Tell a story with data

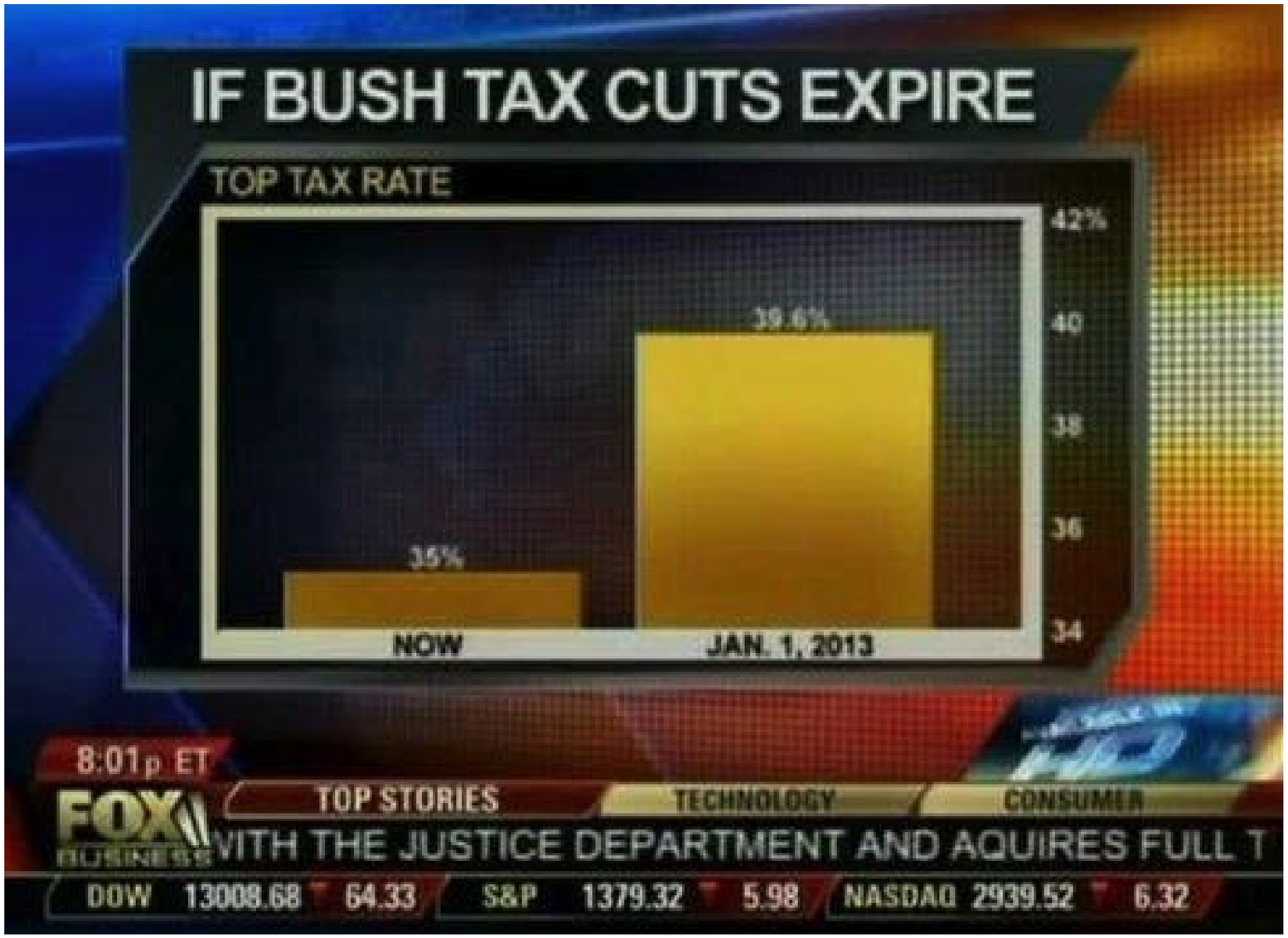
# Graphical Integrity

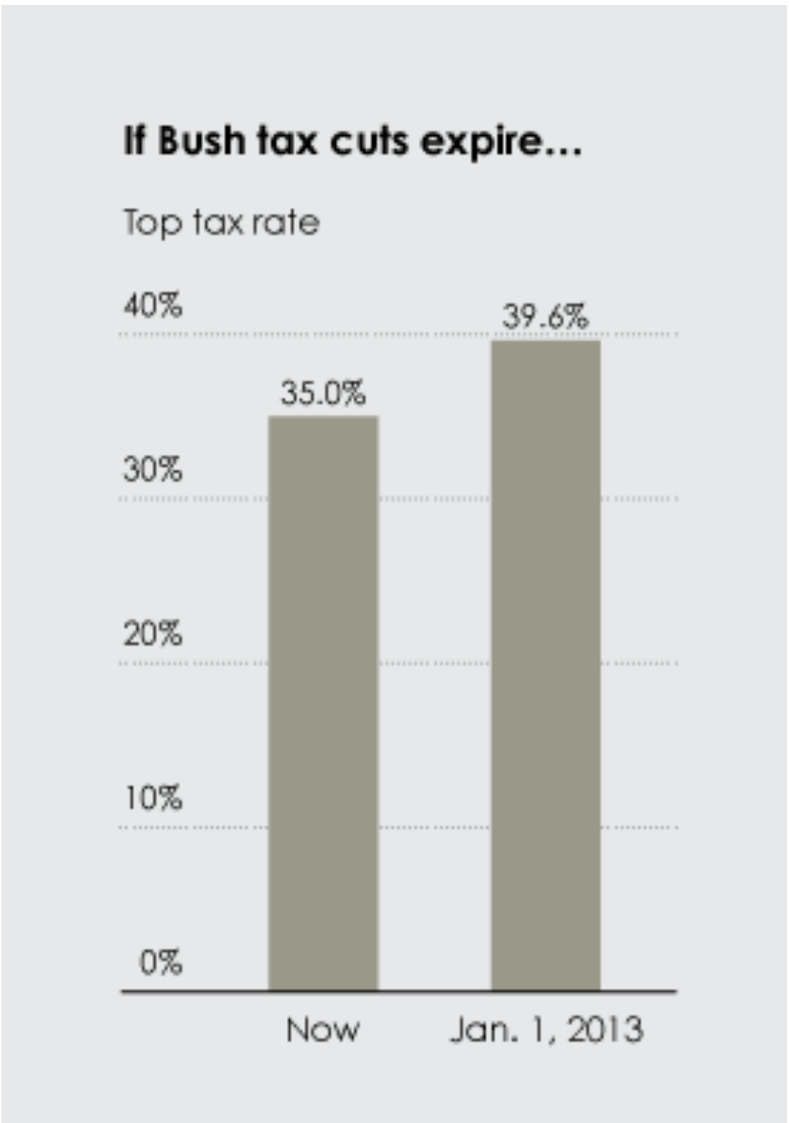
# Graphical Integrity




Flowing Data

# Scale Distortions

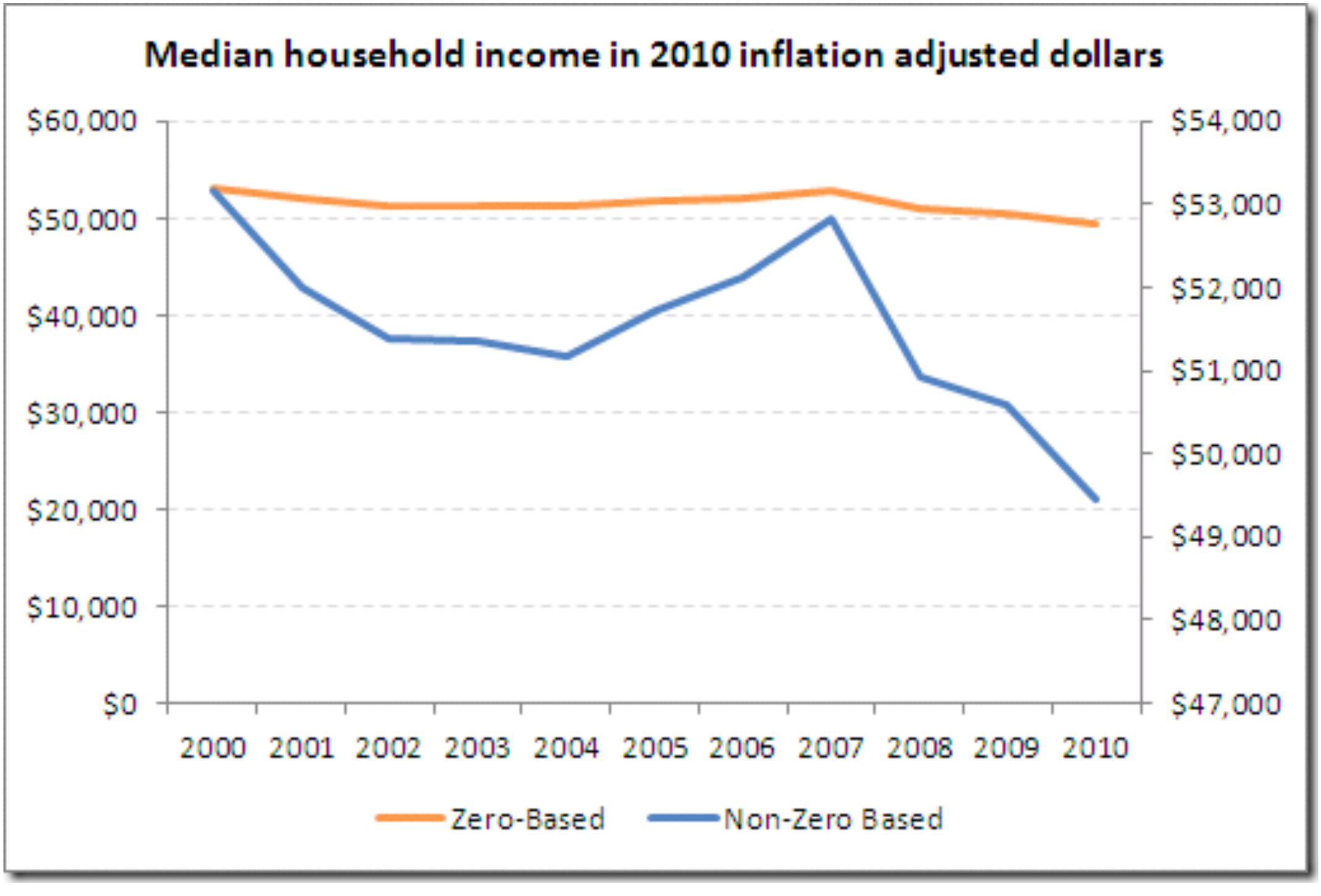


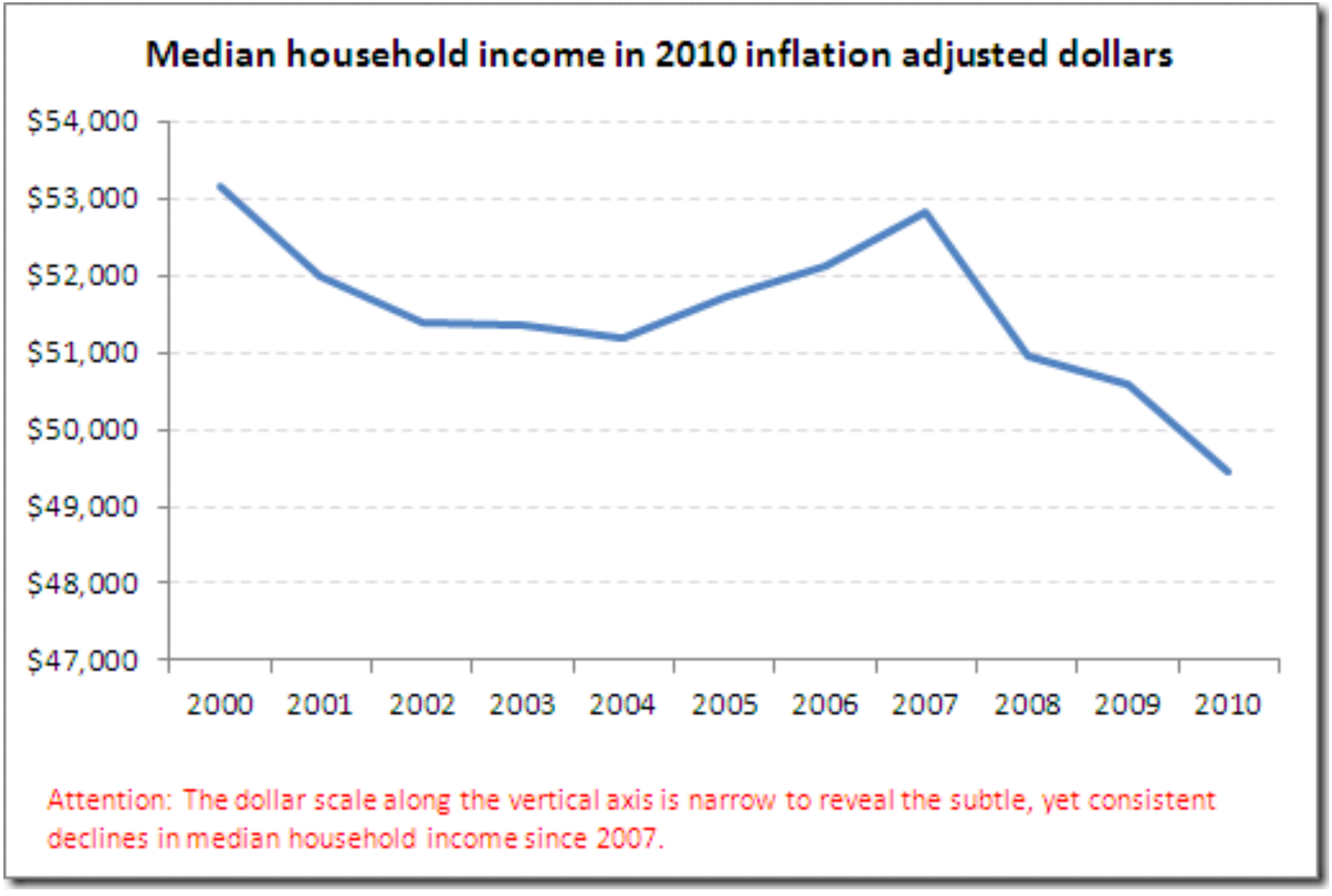
# Scale Distortions

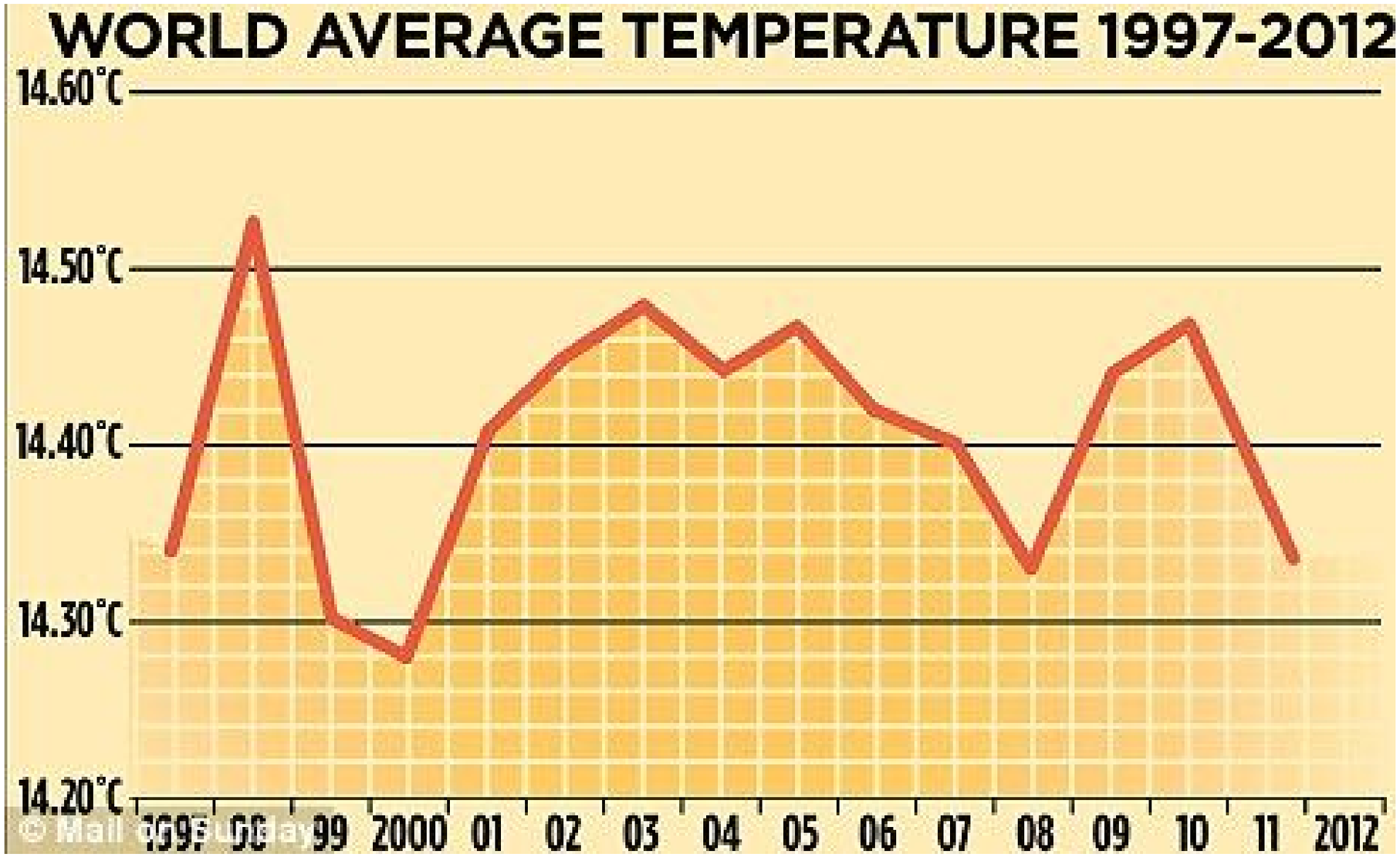
# Scale Distortions

# Scale Distortions



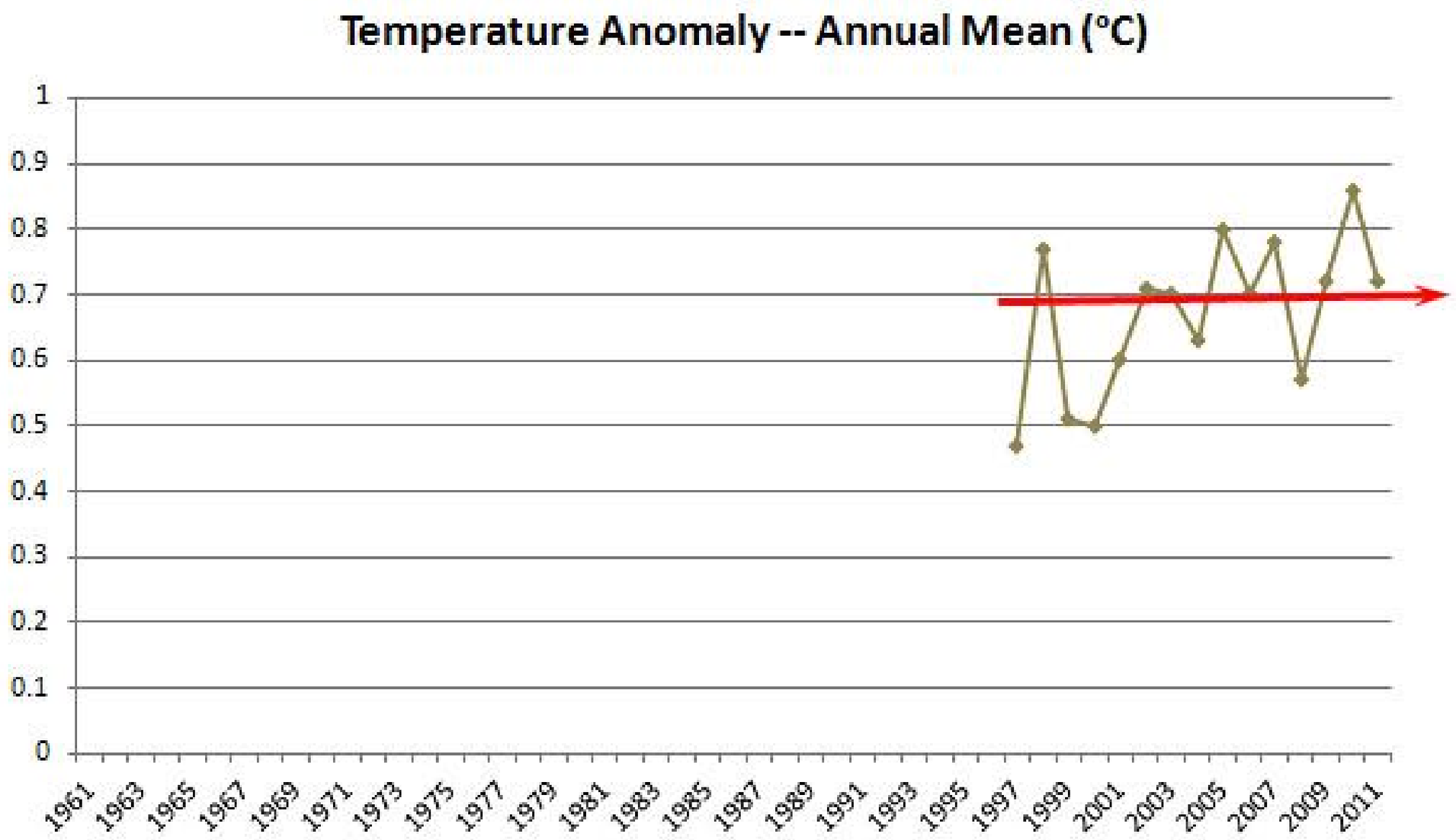

A. Kriebel, VizWiz


Median household income in 2010 inflation adjusted dollars
$54,000
$53,000
$52,000
$51,000
$50,000
$49,000
$48,000
$47,000
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Attention: The dollar scale along the vertical axis is narrow to reveal the subtle, yet consistent declines in median household income since 2007.



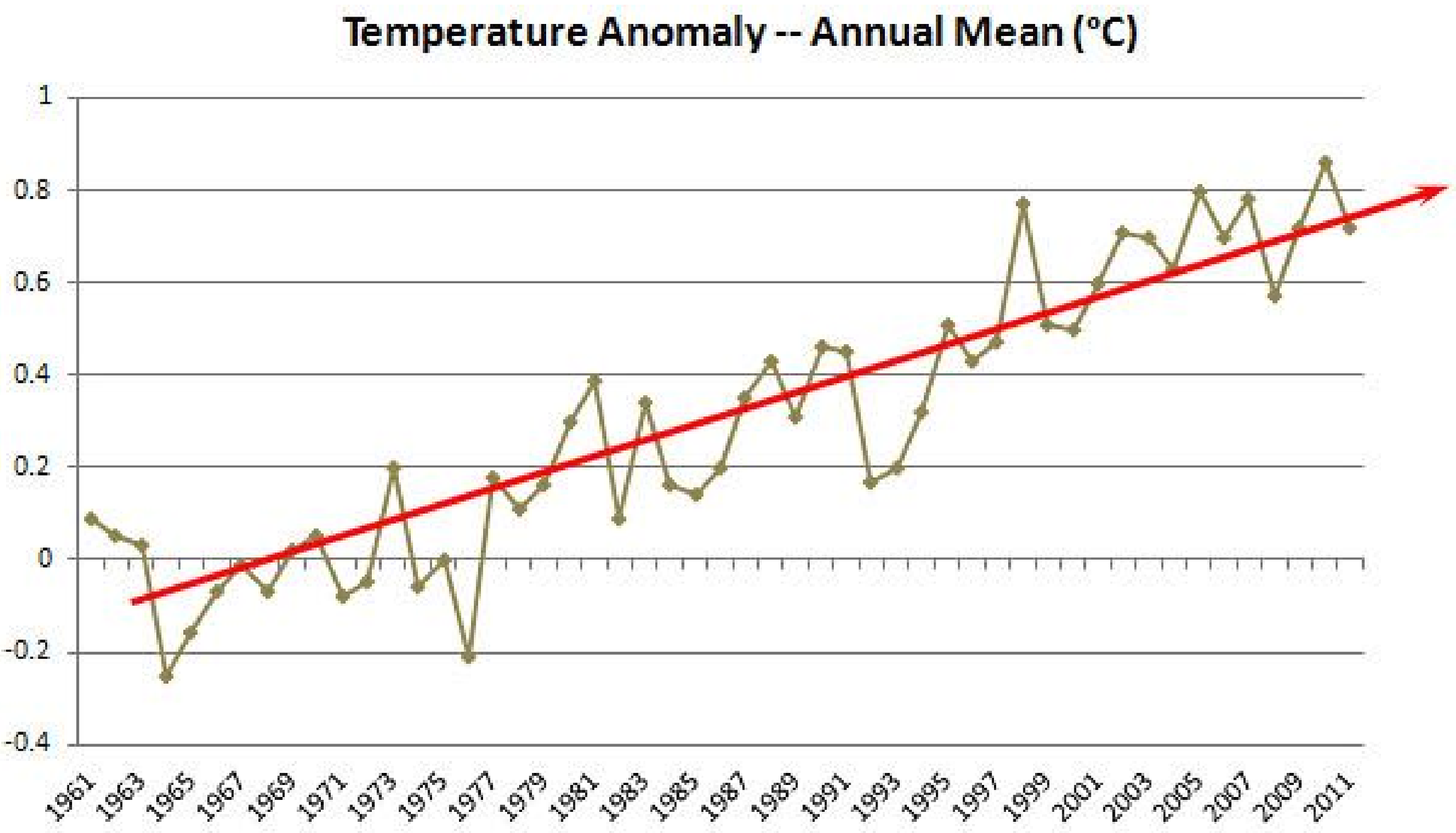

The Daily Mail, UK, Jan 2012

# Global Warming?

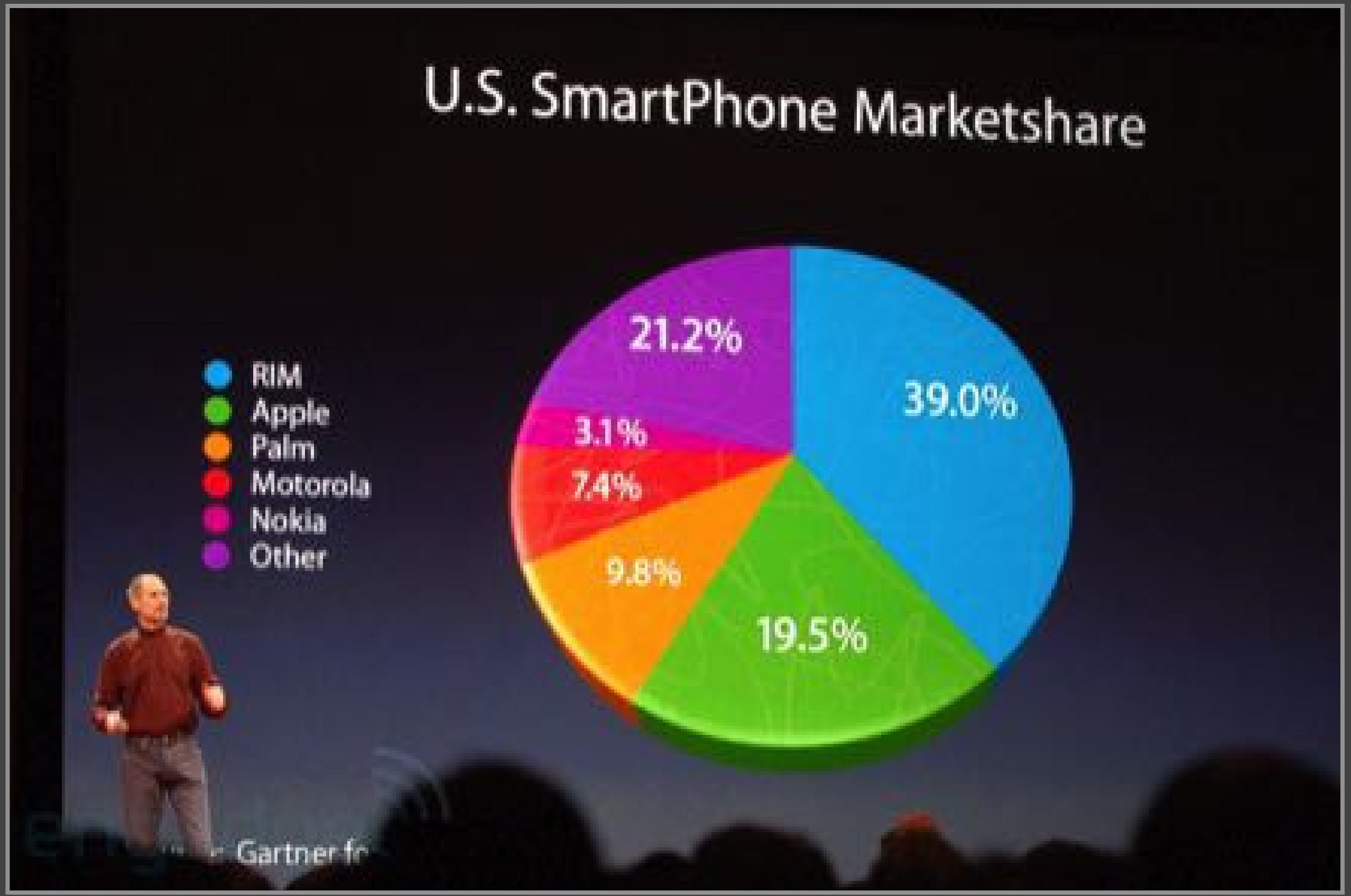



Mother Jones

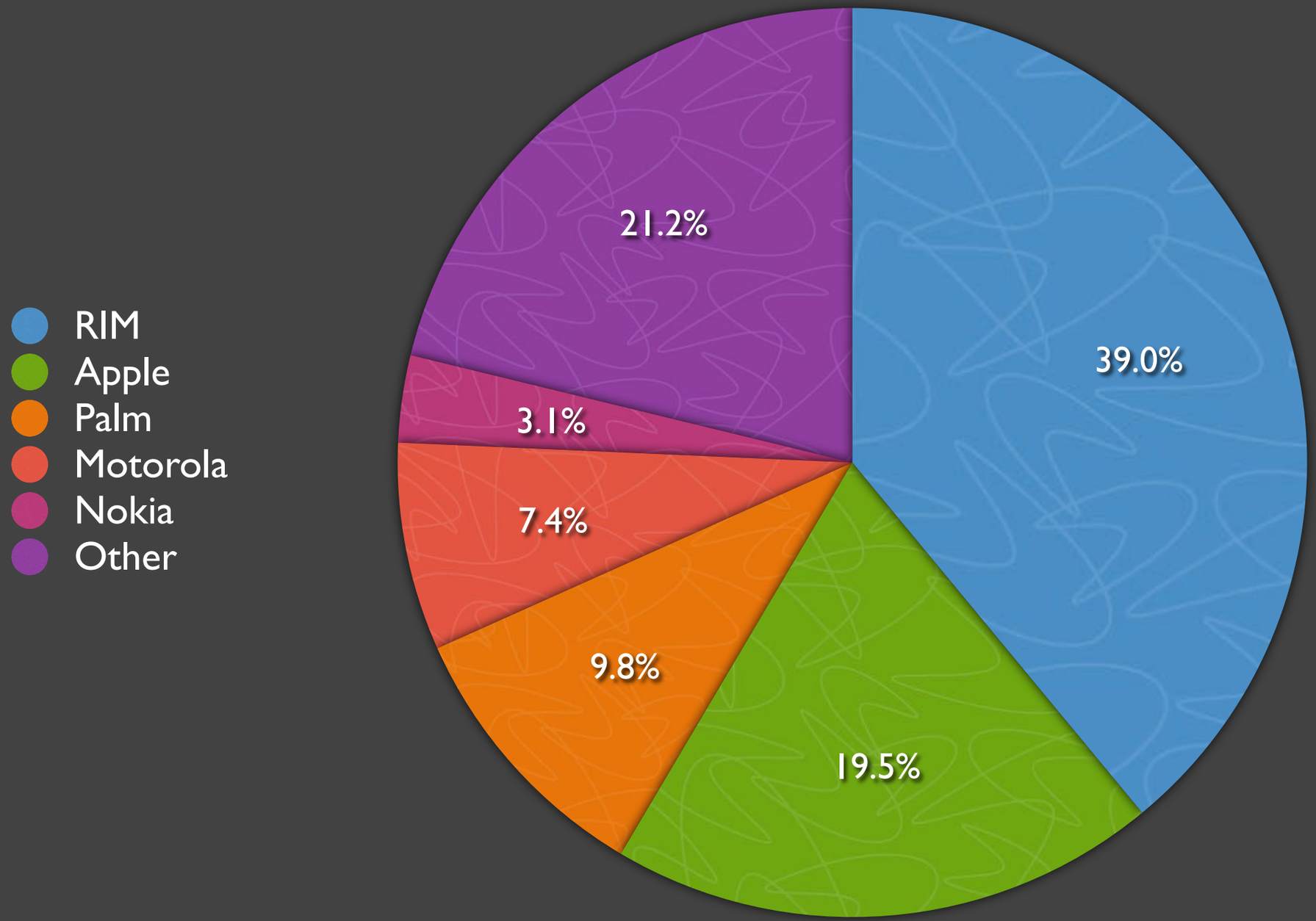
# Global Warming!



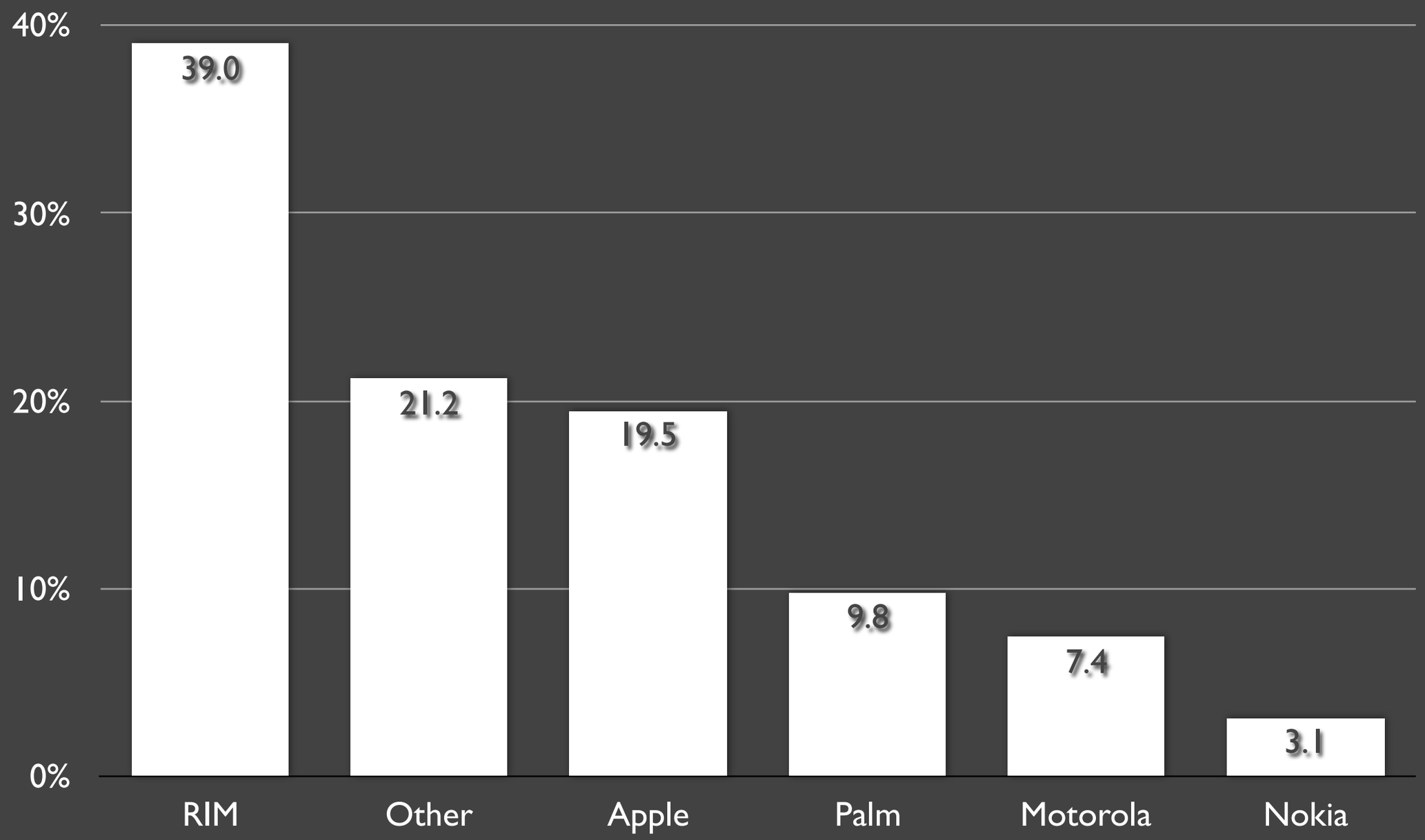

Mother Jones


U.S. SmartPhone Marketshare
RIM
Apple
Palm
Motorola
Nokia
Other
21.2%
39.0%
3.1%
7.4%
9.8%
19.5%
Gartner fr

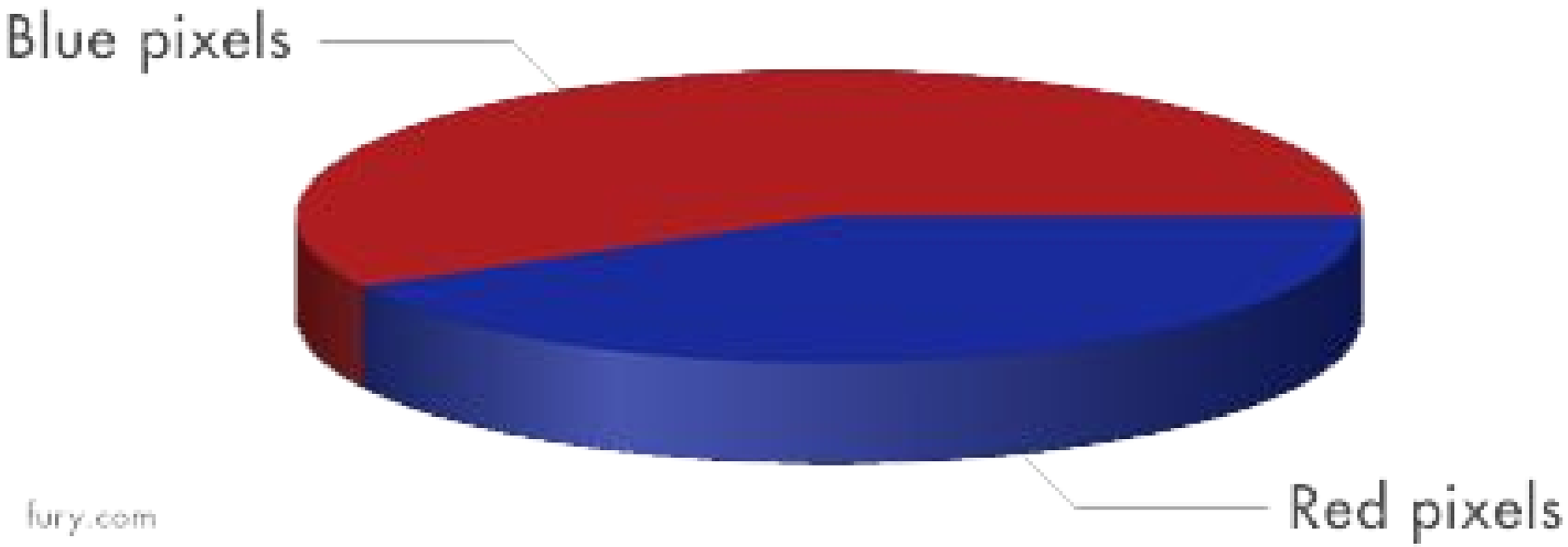
# U.S. SmartPhone Marketshare



# U.S. SmartPhone Marketshare



# Why 3D pie charts are bad



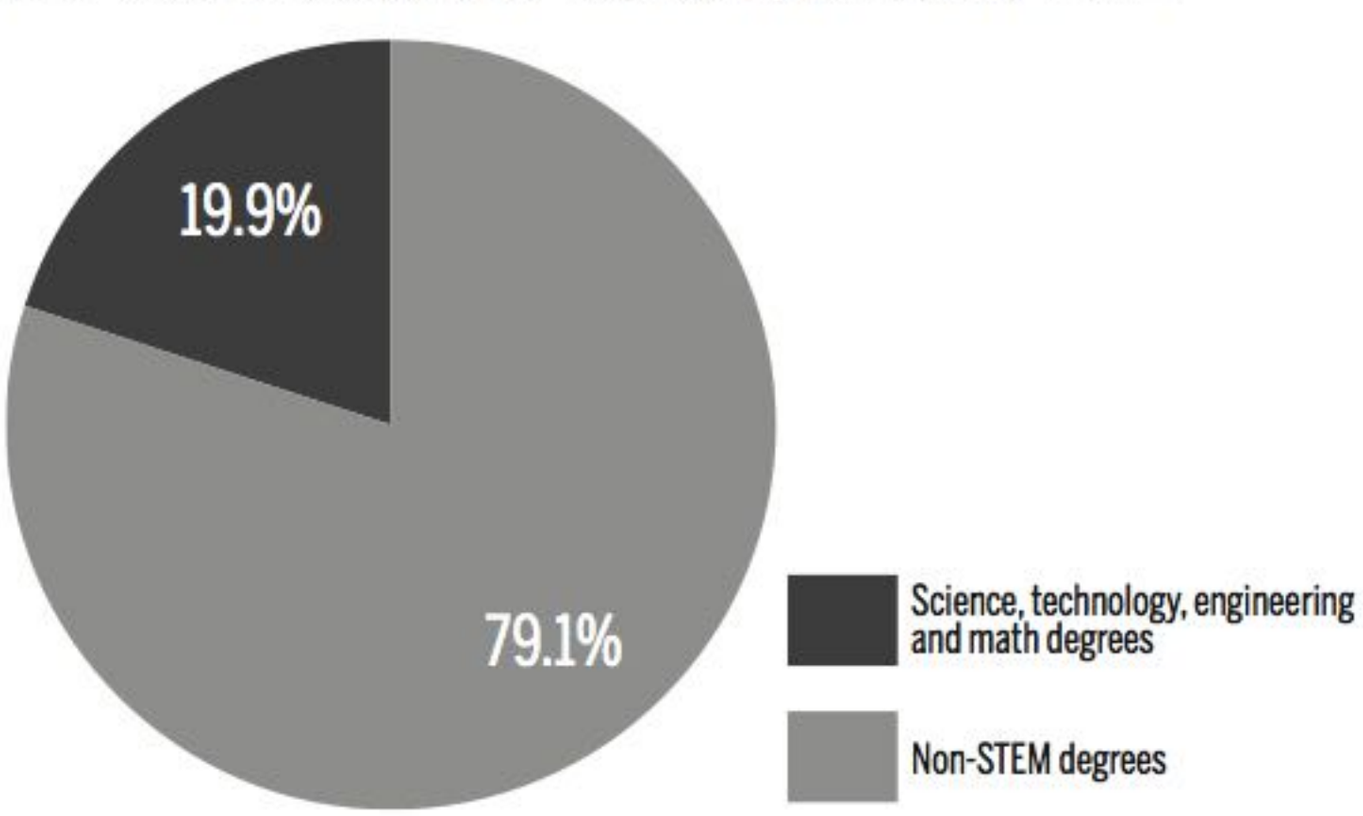

Kevin Fox

Yale Daily News | THE OLDEST COLLEGE DAILY ESTABLISHED 1878



Yale Summer Session
Over 200 full-credit courses.
Same Veritas. More Lux.
June 4 – July 6 , July 9 – Aug 10
2012 experience Yale

# CHART YALE GRADUATES' MAJORS, CLASS OF 2011




September 13, 2011

Facebook Recommendations

Shake Shack to open in New Haven
277 people recommend this.

Popular anti-religion creates false dichotomy
15 people recommend this.

Friends remember Foucher LAW '14
10 people recommend this.

AIDS activist speaks about documentary film
8 people recommend this.

Panel outlines changes in hip-hop
30 people recommend this.

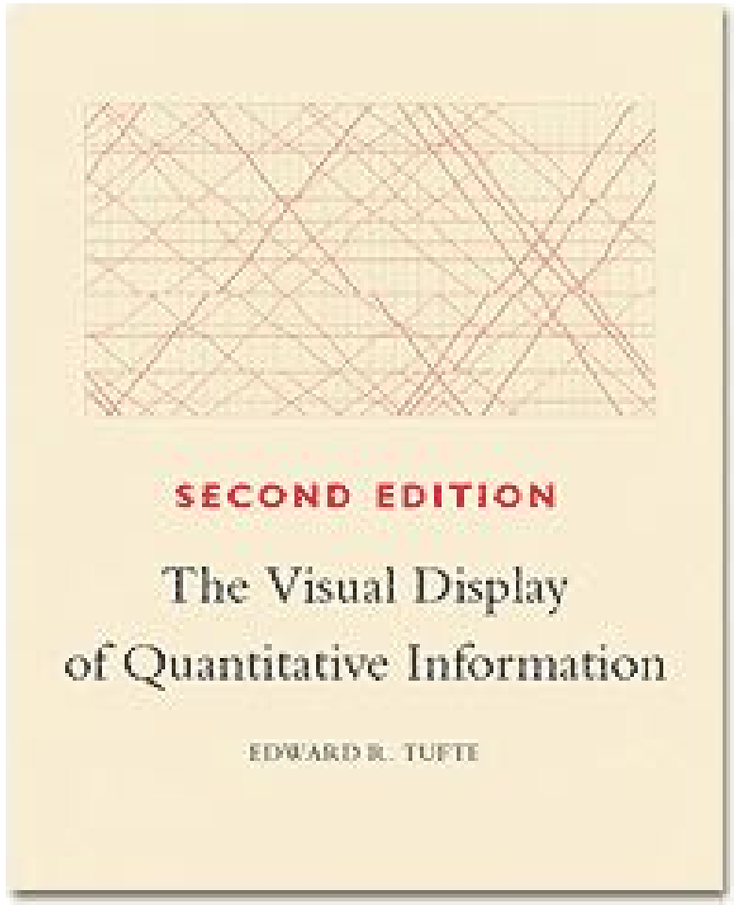
Facebook social plugin

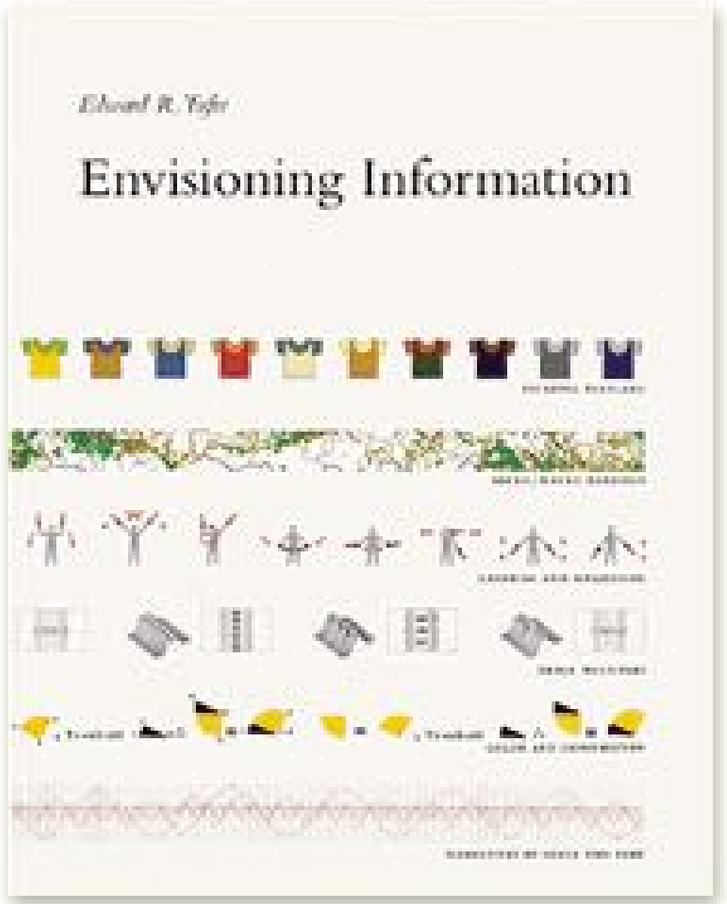
Advertisement

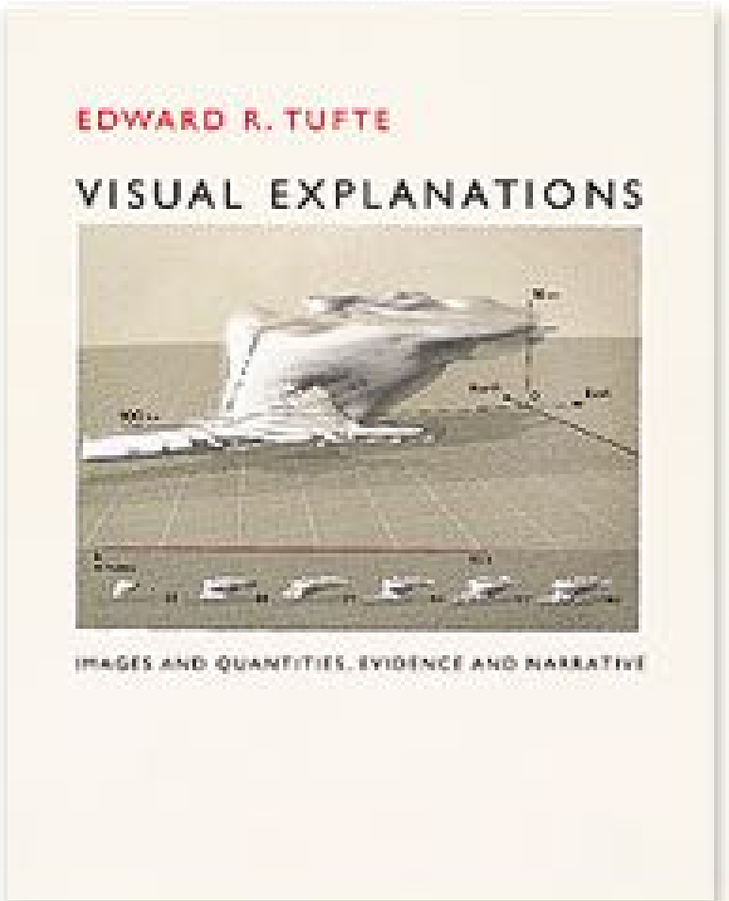
Featured Jobs

# Keep It Simple

# Edward Tufte

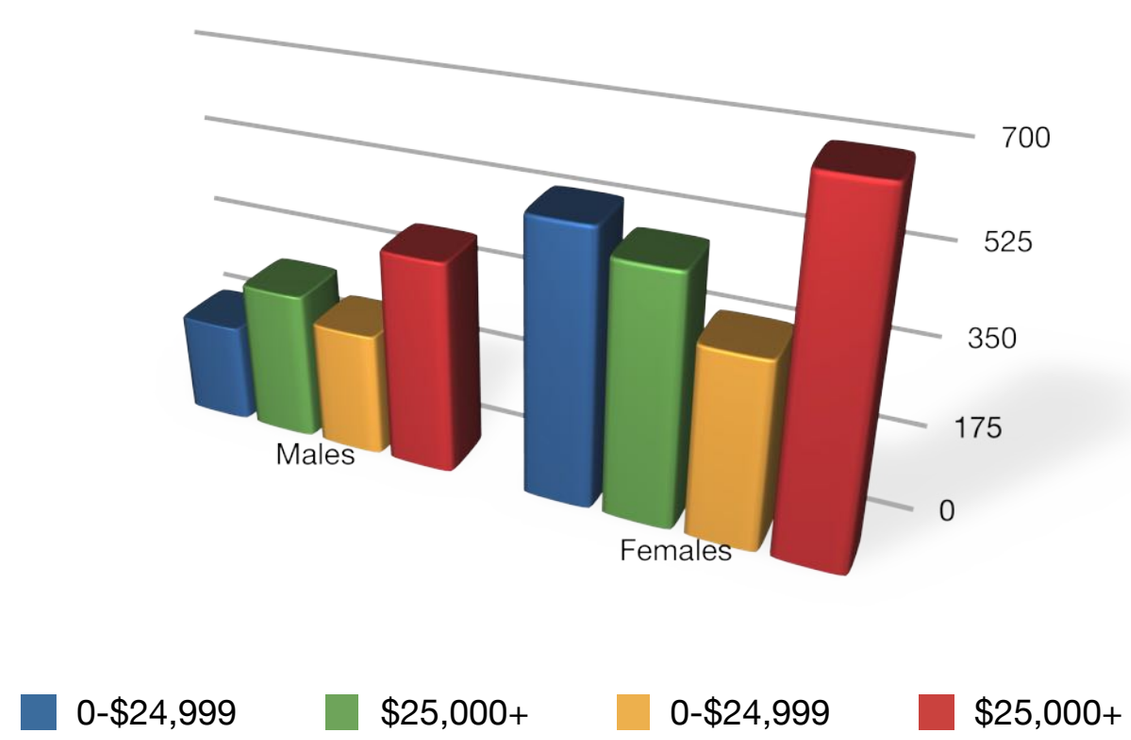


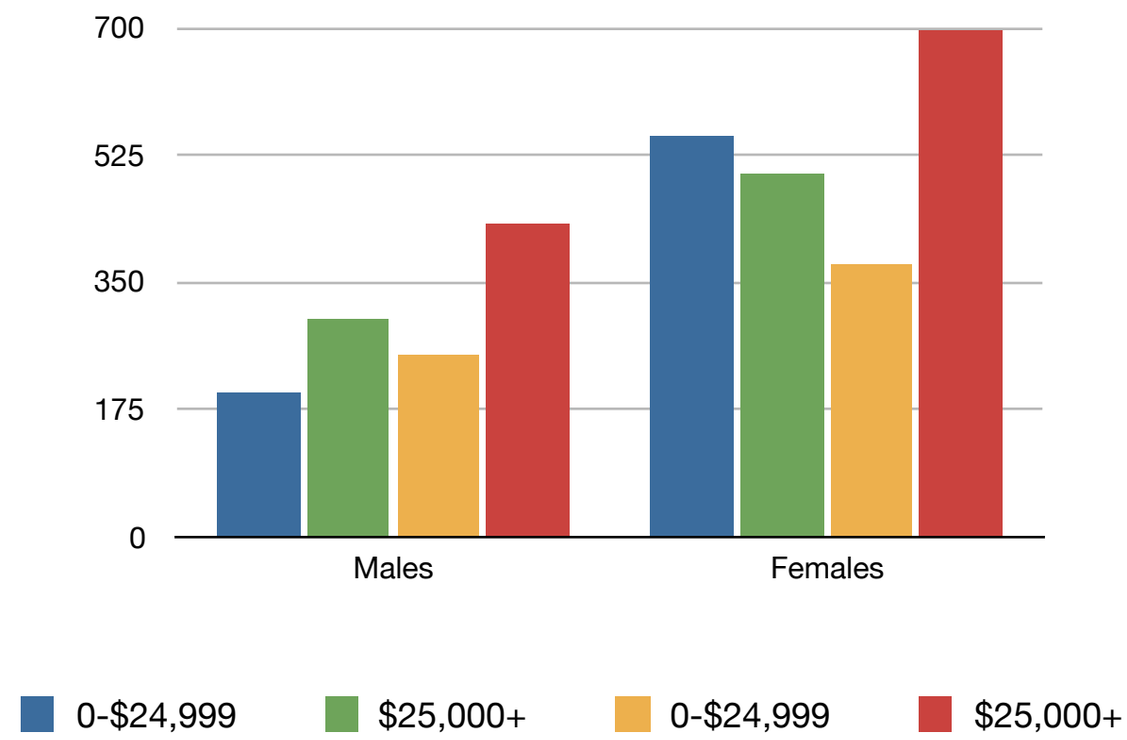
# Maximize Data-Ink Ratio

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



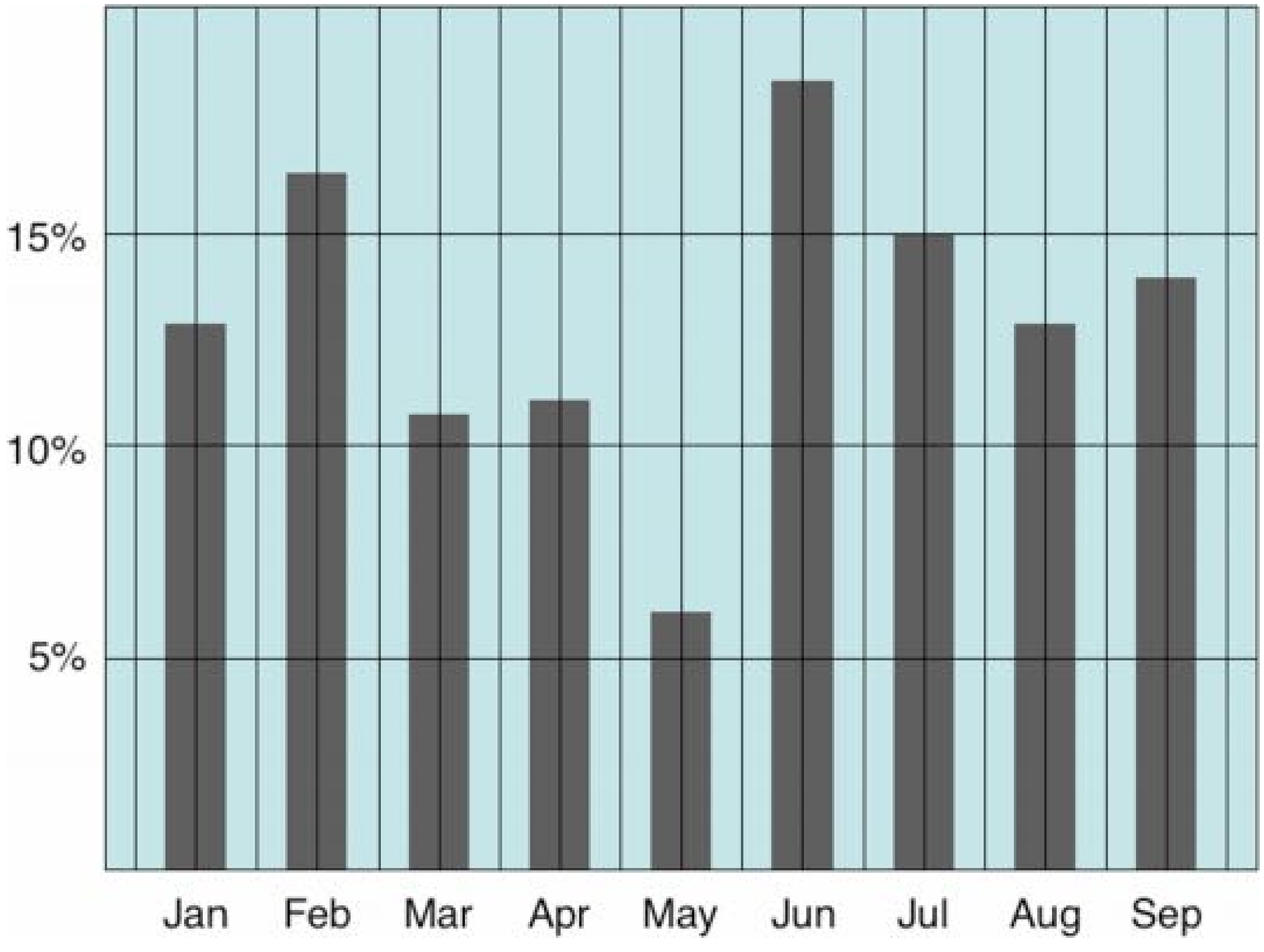
# Maximize Data-Ink Ratio

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



# Avoid Chartjunk

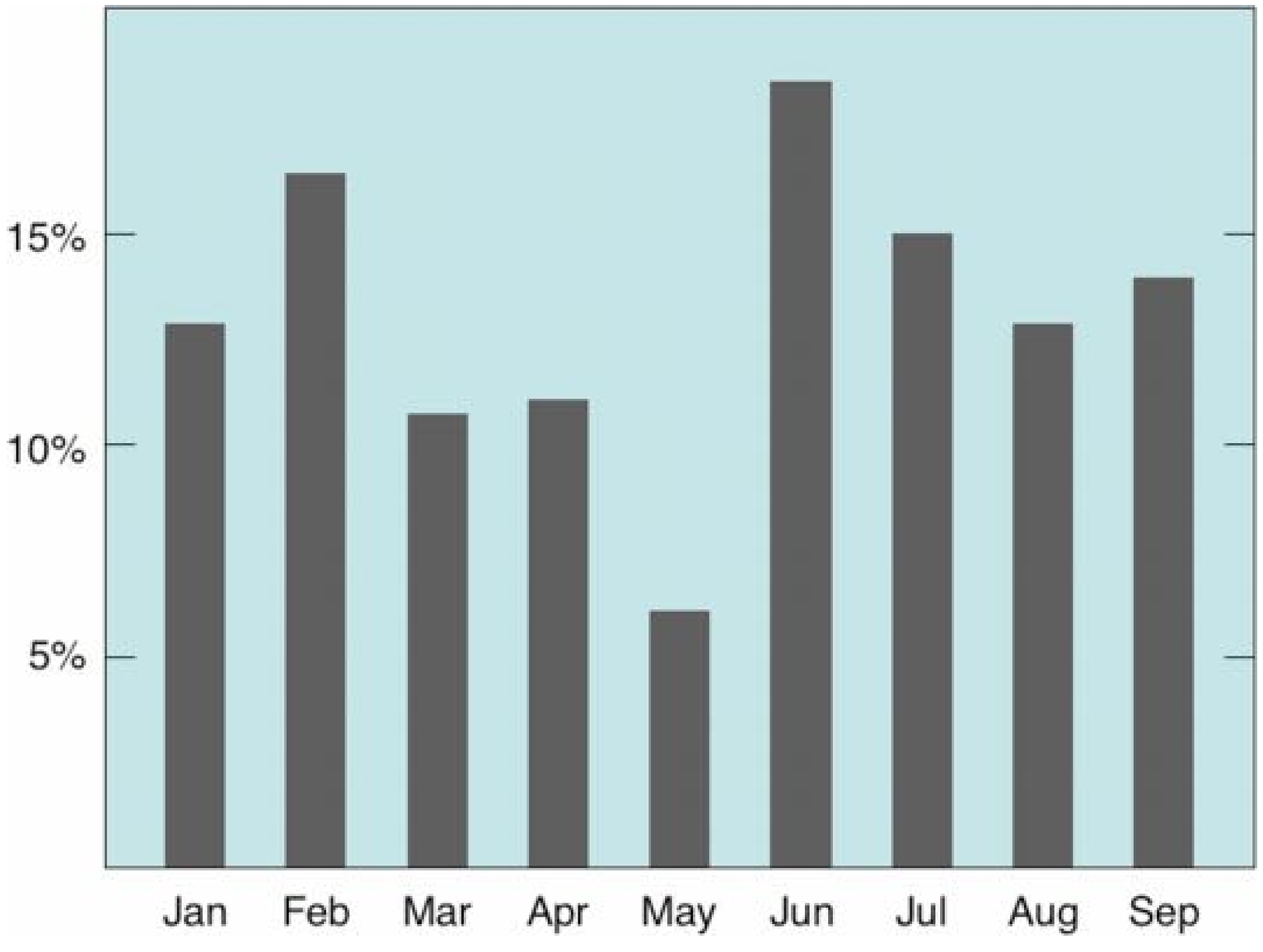
Extraneous visual elements that distract from the message




ongoing, Tim Brey

# Avoid Chartjunk

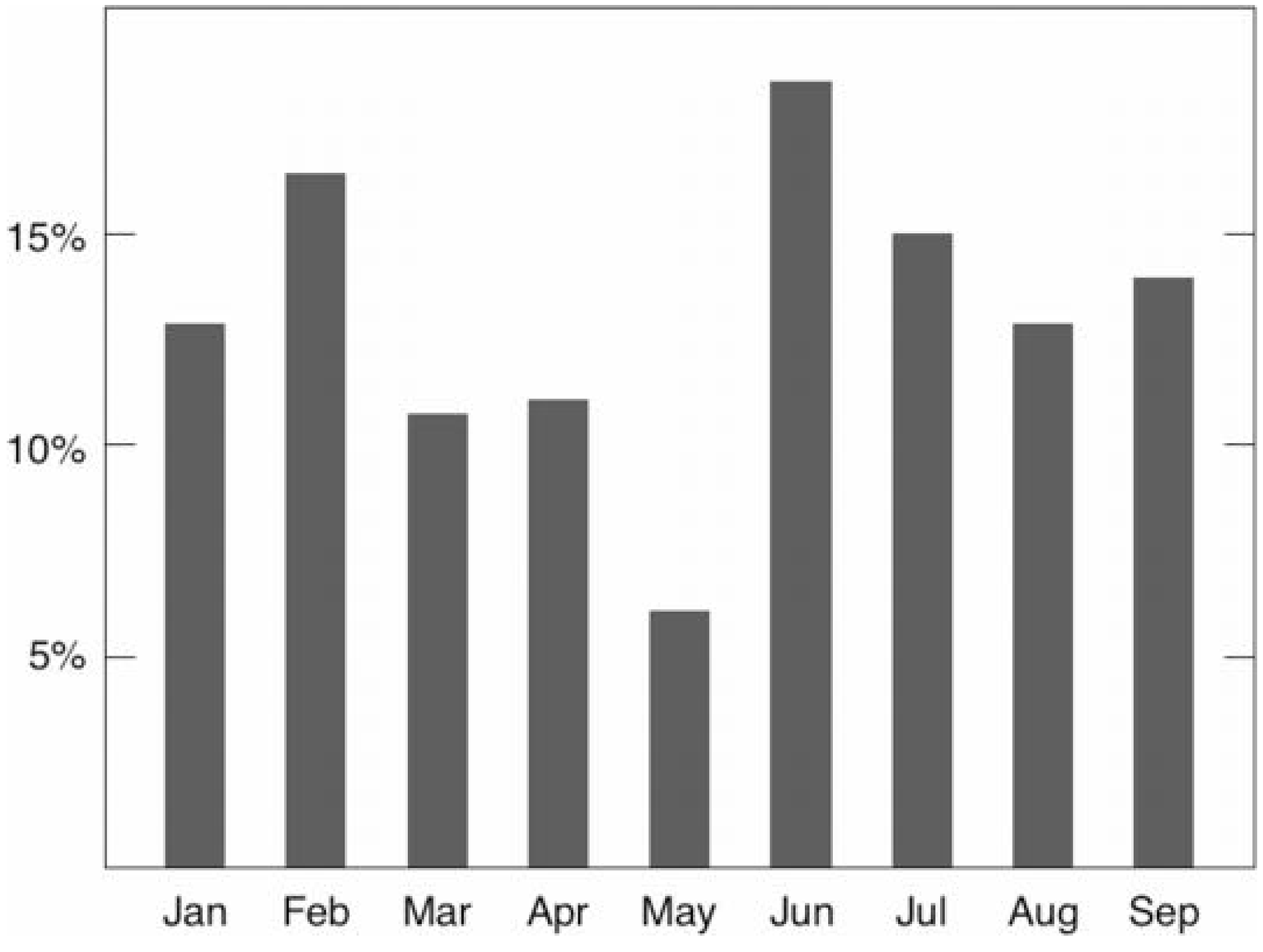
Extraneous visual elements that distract from the message




ongoing, Tim Brey

# Avoid Chartjunk

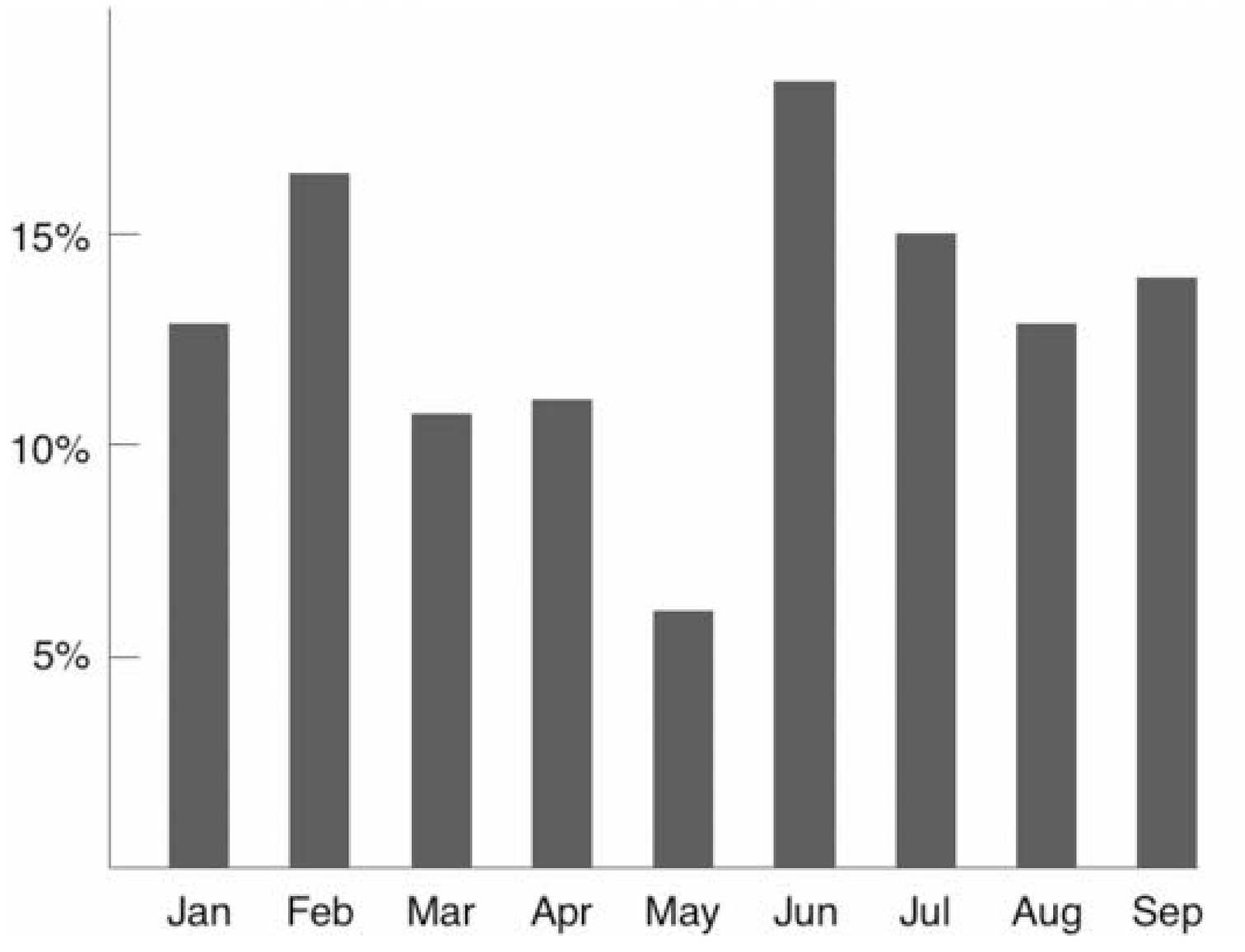
Extraneous visual elements that distract from the message




ongoing, Tim Brey

# Avoid Chartjunk

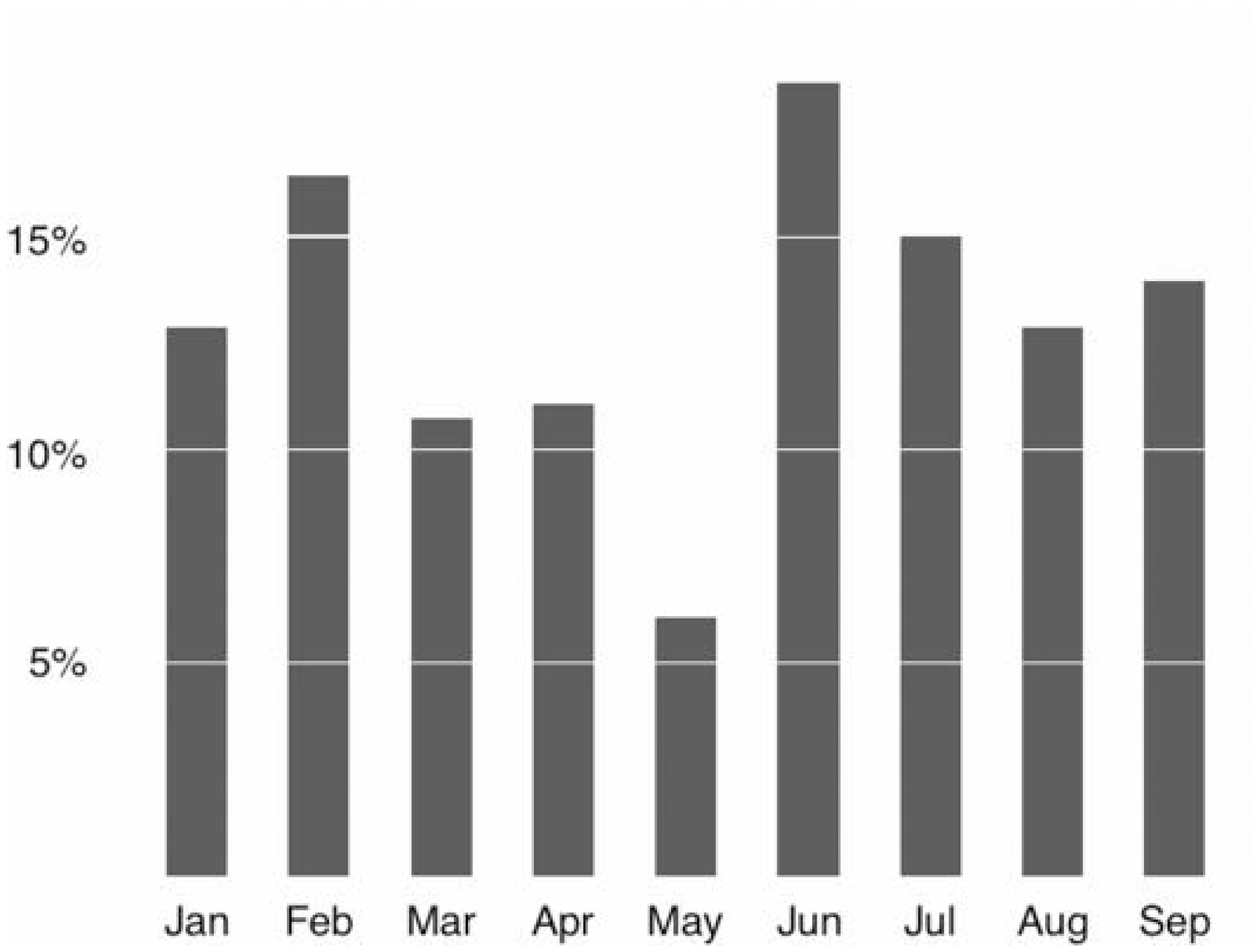
Extraneous visual elements that distract from the message




ongoing, Tim Brey

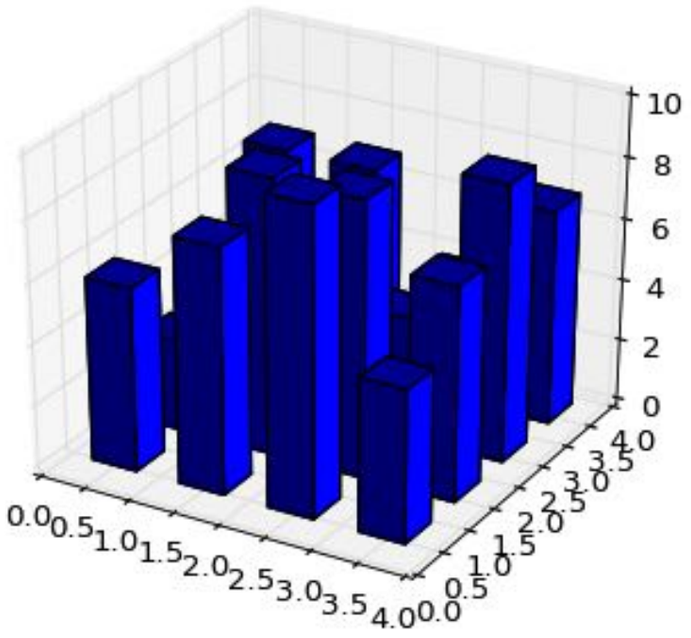
# Avoid Chartjunk

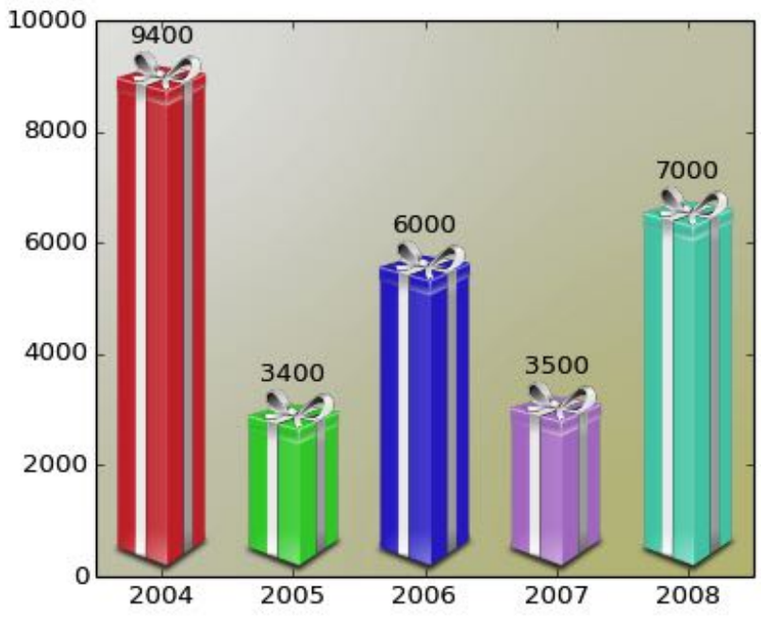
Extraneous visual elements that distract from the message



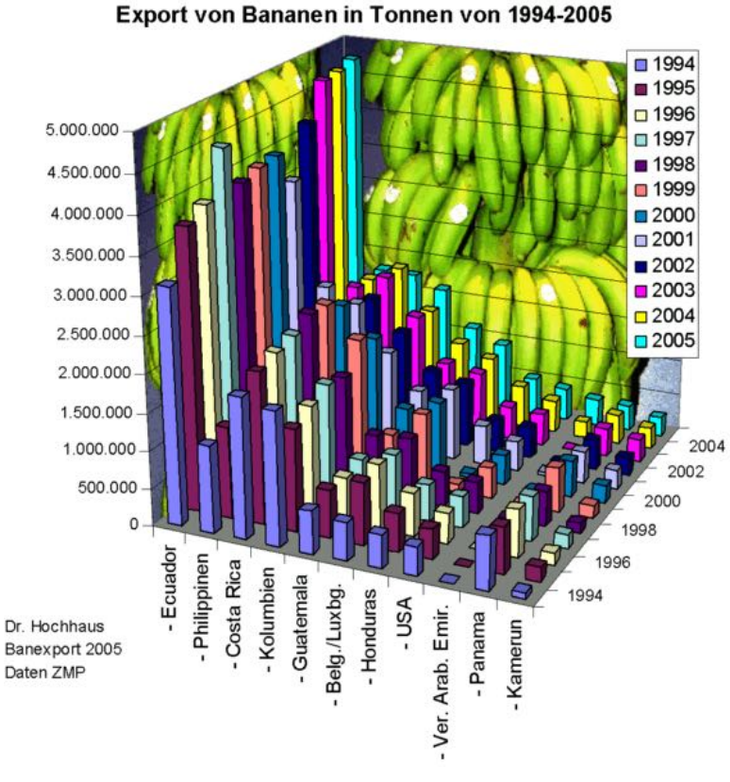

ongoing, Tim Brey

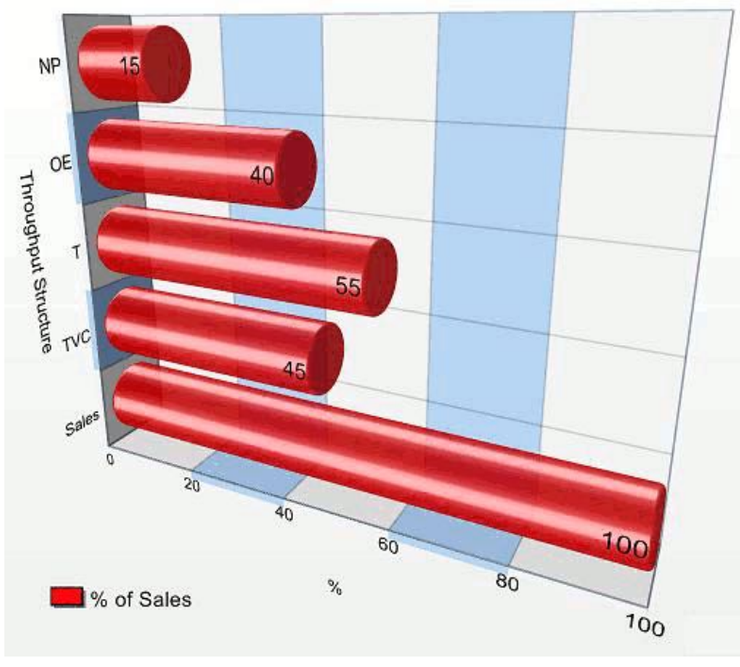
# Don’t!



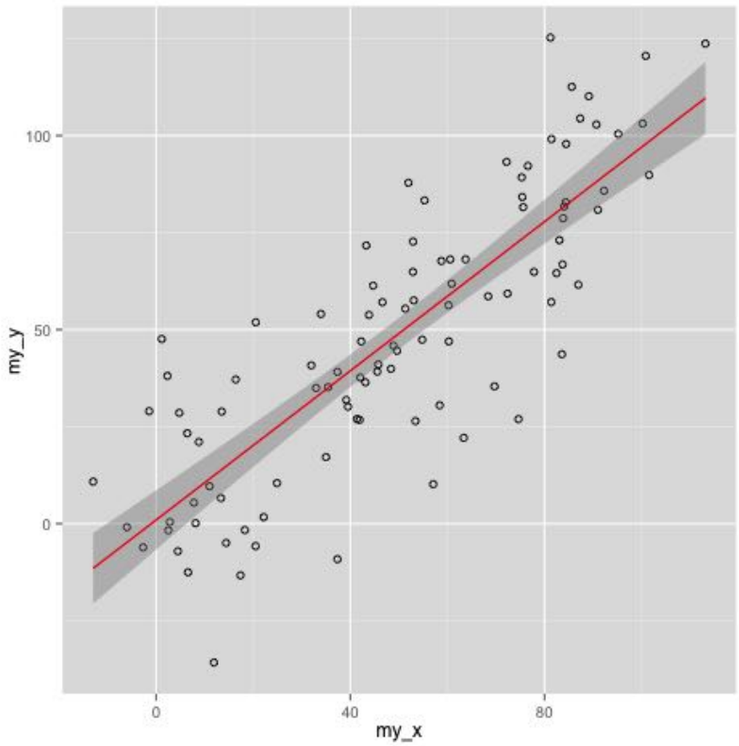

matplotlib gallery

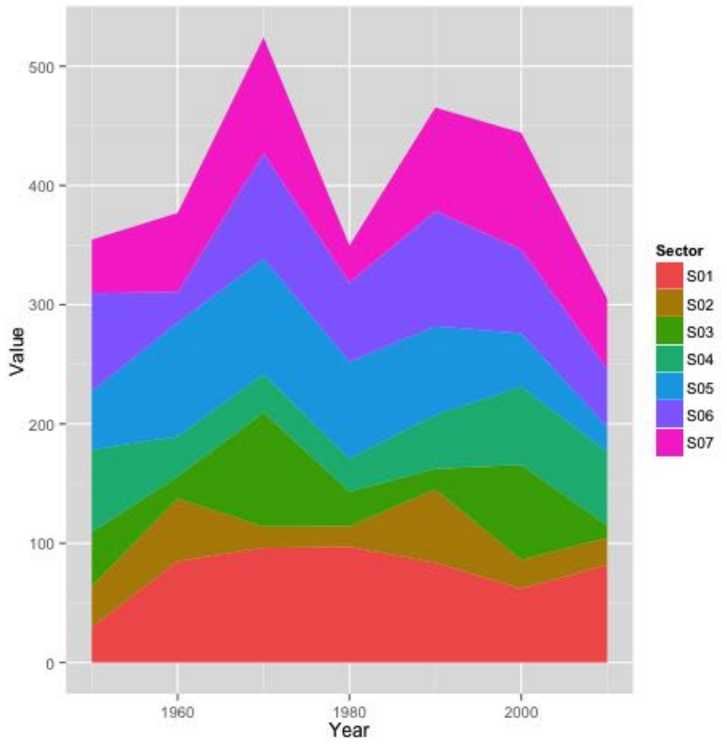


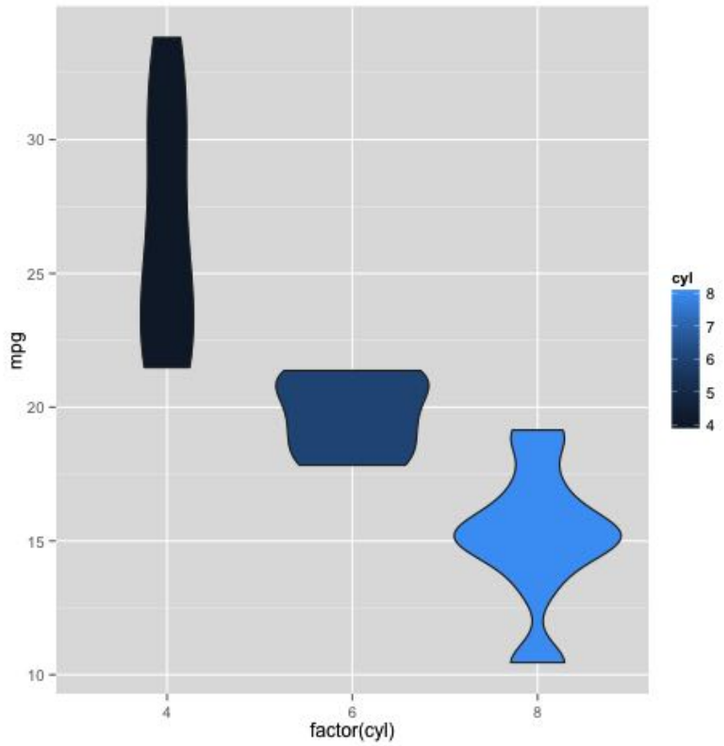



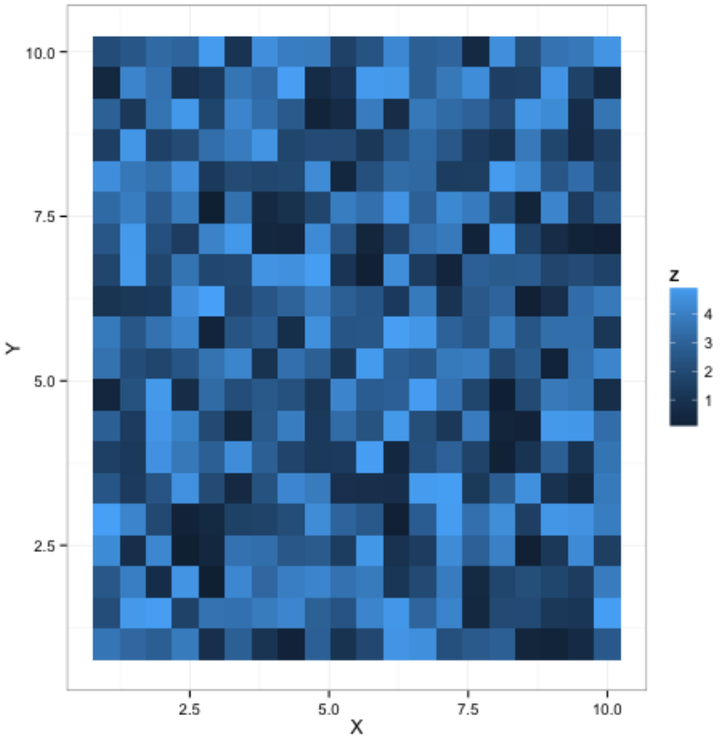

Excel Charts Blog

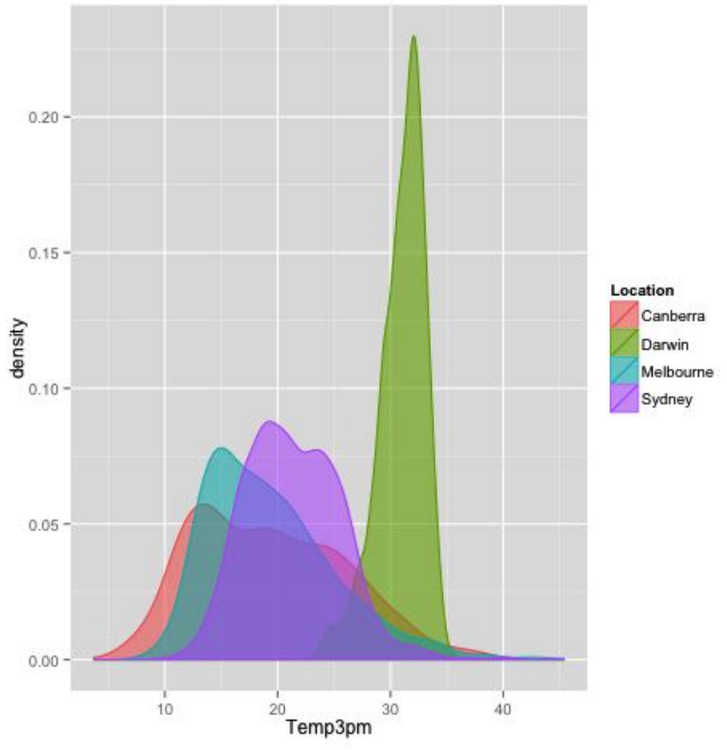
# ggplot2

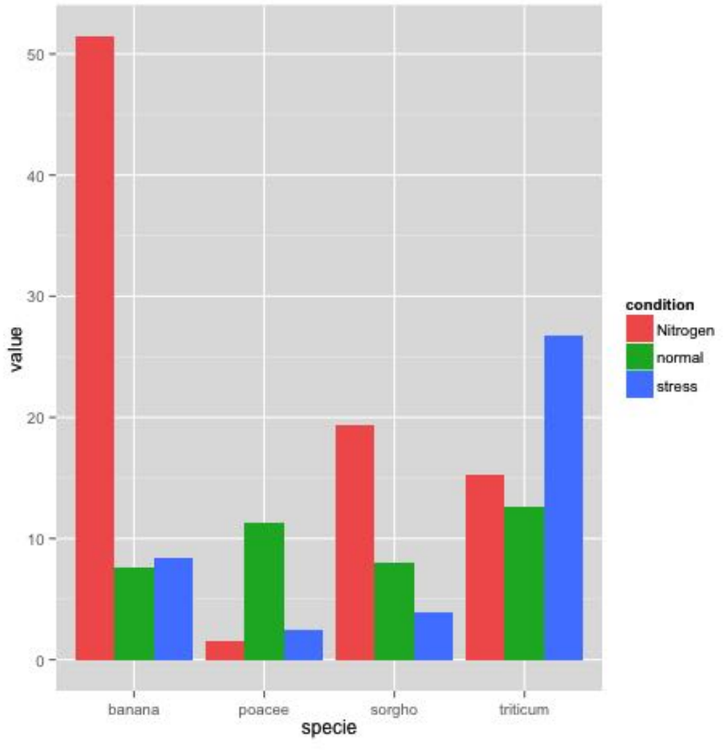

my_y
100
50
0
0
40
80
my_x


Value
500
400
300
200
100
0
1960
1980
2000
Year
Sector
S01
S02
S03
S04
S05
S06
S07


mpg
30
25
20
15
10
4
6
8
factor(cyl)
cyl
8
7
6
5
4

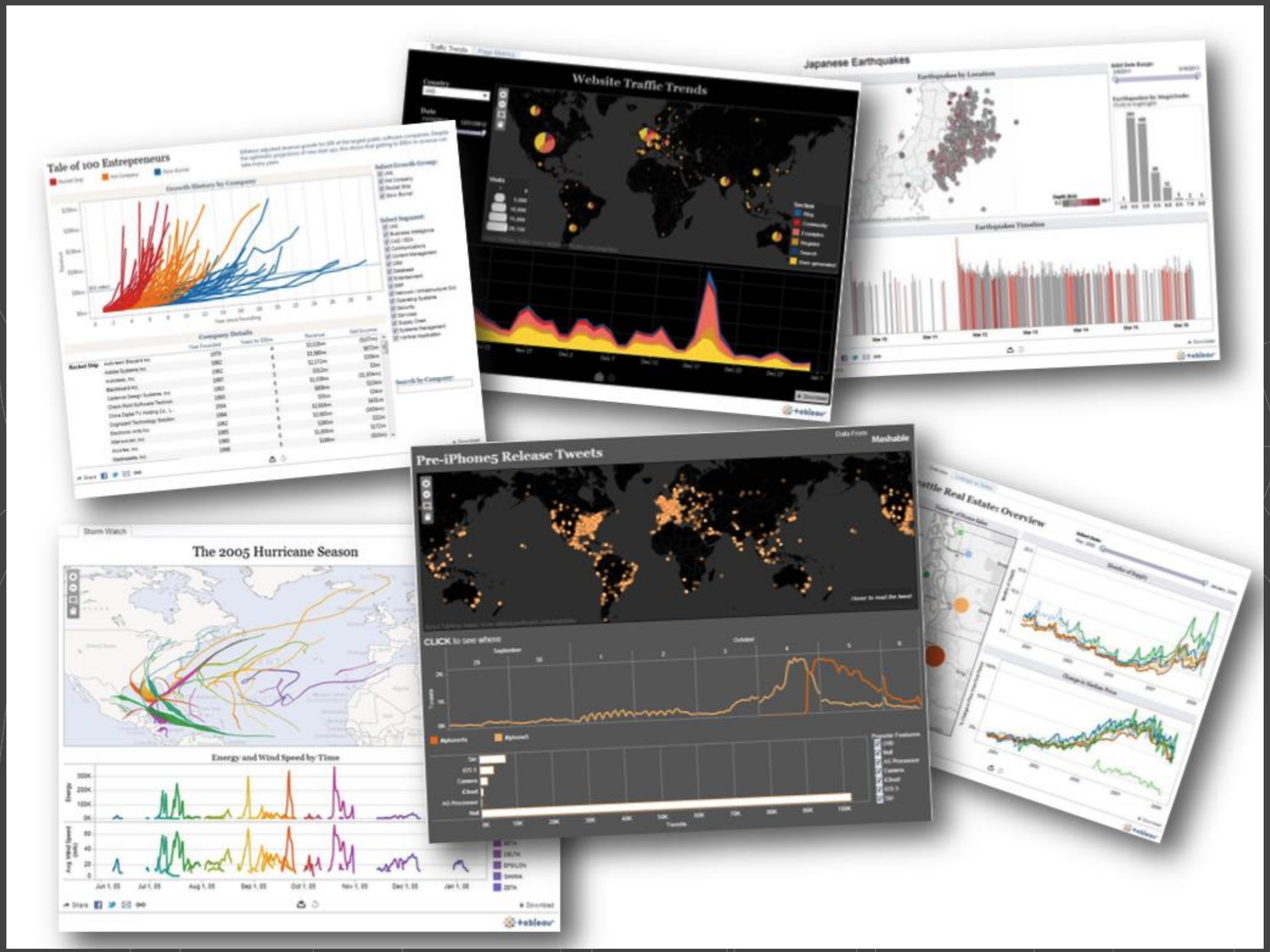

Y
10.0
7.5
5.0
2.5
2.5
5.0
7.5
10.0
X
z
4
3
2
1


density
0.20
0.15
0.10
0.05
0.00
10
20
30
40
Temp3pm
Location
Canberra
Darwin
Melbourne
Sydney


value
50
40
30
20
10
0
banana
poacee
sorgho
triticum
specie
condition
Nitrogen
normal
stress

# Activity

## Introduction to Tableau



# Activity

Look at the dataset called MusicStore.csv in Tableau.

Which genre sells the most in Europe and Asia?

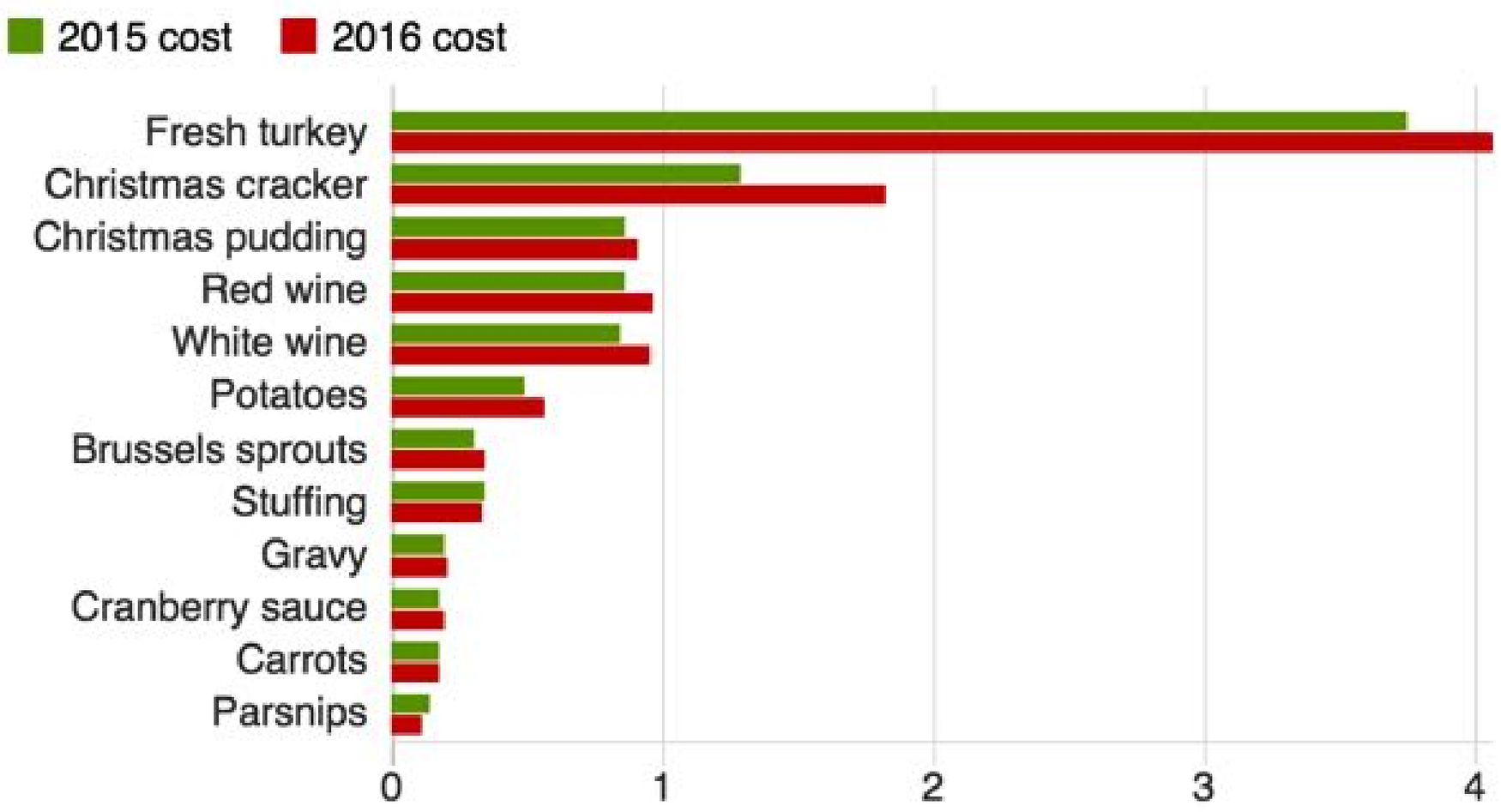
How does the sales trend of Rock compare with Heavy Metal?

What does the geographic sales distribution look like in Europe and Asia?

# Use The Right Display

## How the cost of your Christmas dinner has changed

Price (£) per item per person




Source: mySupermarket (prices adjusted for inflation)

BBC

# Is the Cost of Christmas Dinner Rising?

55% of products have seen a cost increase since 2006

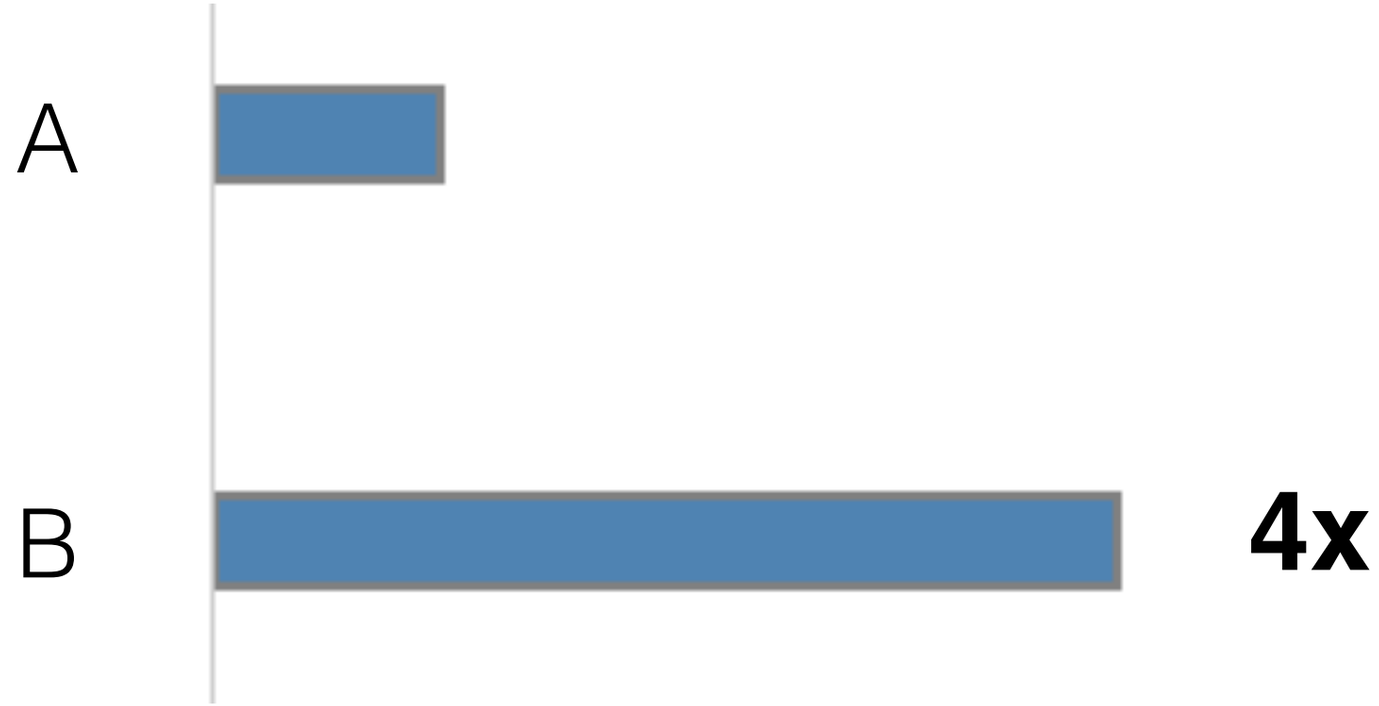
• Less than 2006 • More than 2006 • Peak Cost

| Product Family | 2016 | % Change | |
|---|---|---|---|
| Total Cost | £119.94 | +3.3% | |
| Gammons & Hams | £26.46 | -15.2% | |
| Blended Whisky | £24.98 | +36.8% | |
| Christmas Crackers | £10.90 | -5.5% | |
| Fresh Turkey | £8.34 | +8.0% | |
| Stuffing | £5.98 | +19.6% | |
| Red Wine | £5.75 | +14.8% | |
| White Wine | £5.69 | +14.0% | |
| Sponge Pudding | £5.44 | -10.7% | |
| Frozen Turkey | £5.37 | +57.9% | |
| Beers & Lagers | £5.26 | -21.8% | |
| Brussels Sprouts | £4.07 | +11.2% | |

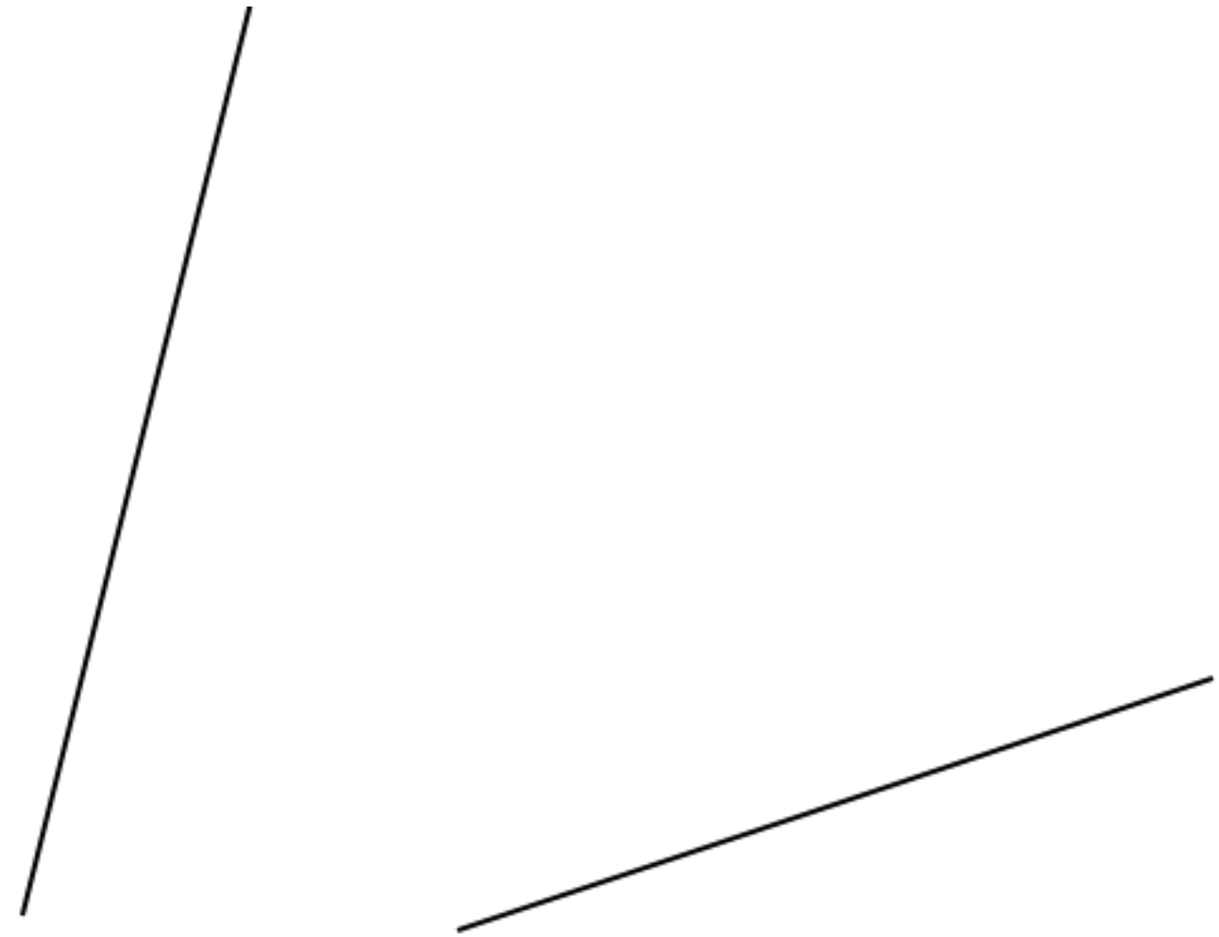
# Activity

Please answer the following quiz questions by holding up fingers on your hand.

# How much longer?



# How much steeper?

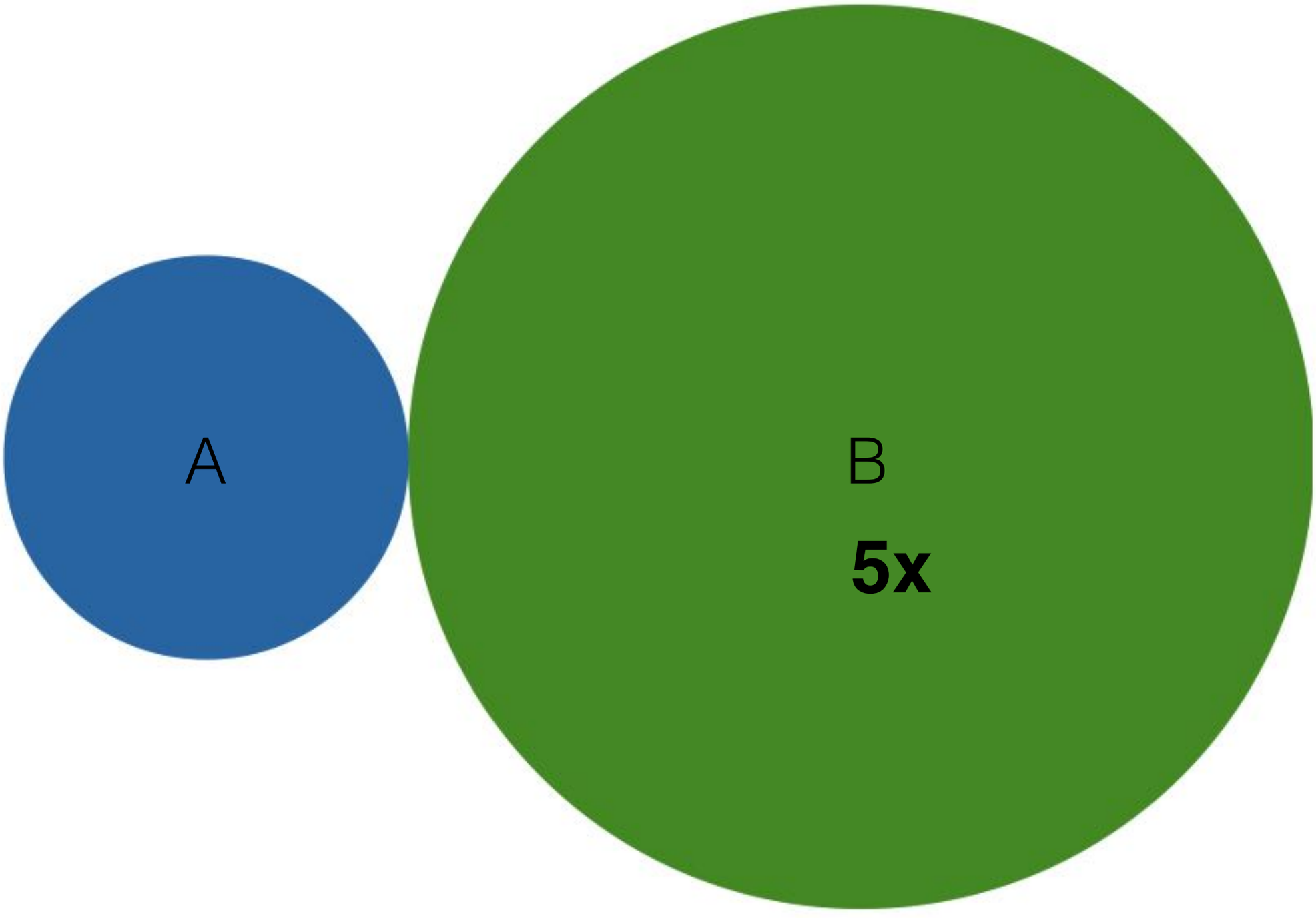


A

B

**4x**

# How much larger area?



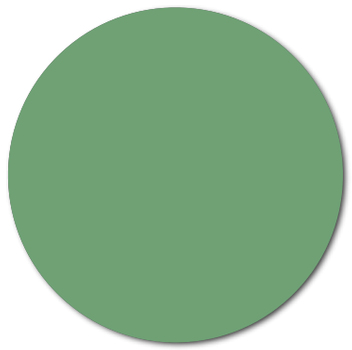
# How much darker?

A

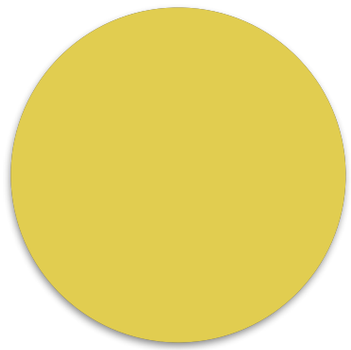
B

**2x**

# How much bigger value?



A

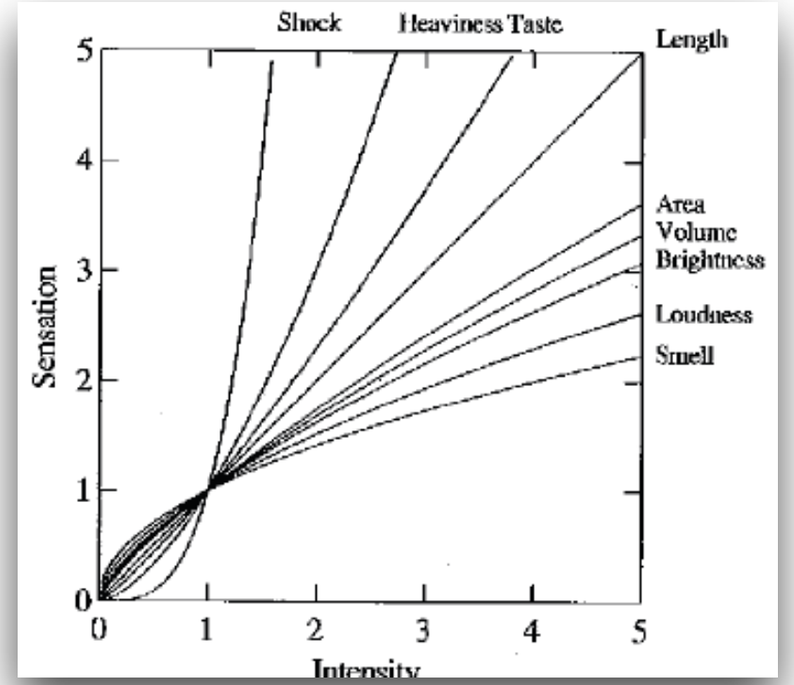


B

**4x**

2

16




Stephen's Power Law, 1961

| | Nominal | Ordinal | Quantitative |
|---|---|---|---|
| Position | ✔ | ✔ | ✔ |
| Size | ✔ | ✔ | ~ |
| (Grey)Value | ✔ | ✔ | ~ |
| Texture | ✔ | ~ | ✖ |
| Color | ✔ | ✖ | ✖ |
| Orientation | ✔ | ✖ | ✖ |
| Shape | ✔ | ✖ | ✖ |

✔ = Good
~ = OK
✖ = Bad

J. Bertin, 1967

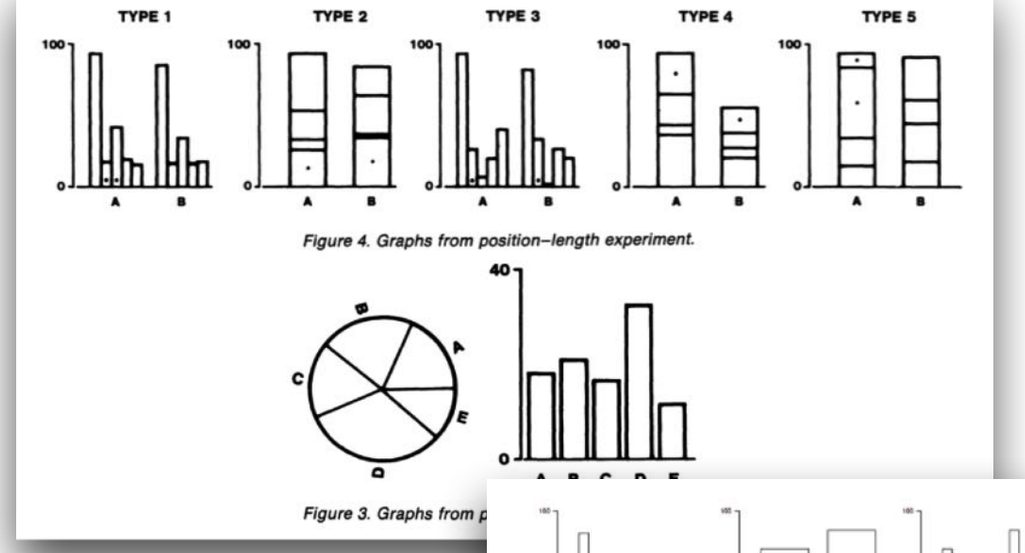
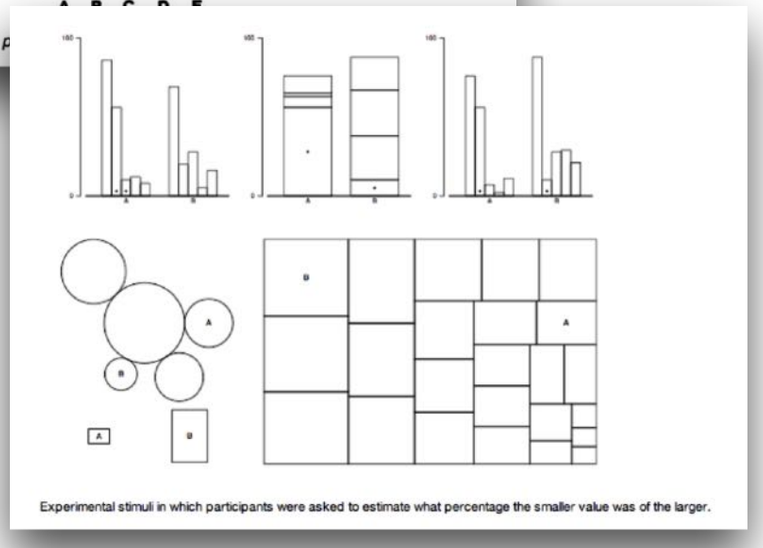



Figure 4. Graphs from position–length experiment.



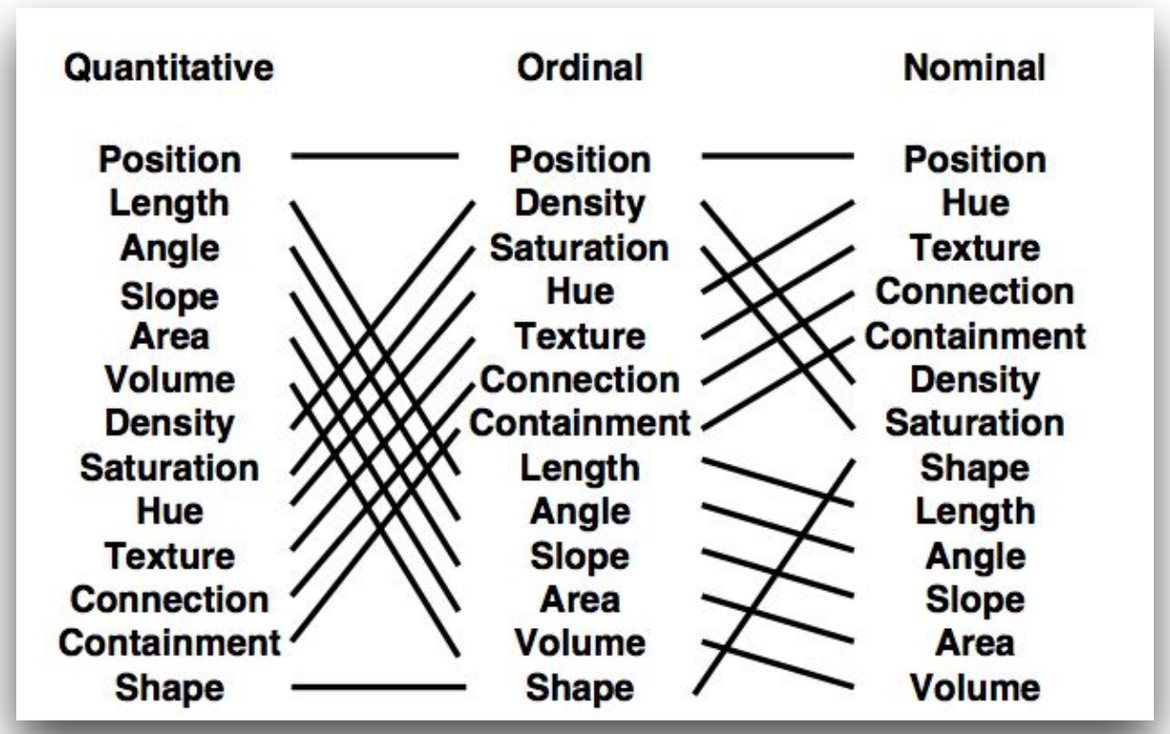

Figure 3. Graphs from p

Cleveland / McGill, 1984



Experimental stimuli in which participants were asked to estimate what percentage the smaller value was of the larger.

Heer / Bostock, 2010




J. Mackinlay, 1986

Most Efficient

Position

Length

Slope

Angle

Quantitative

Area

Intensity

Ordered

Least Efficient

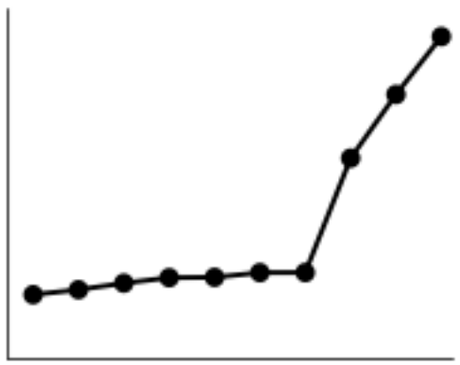
Color

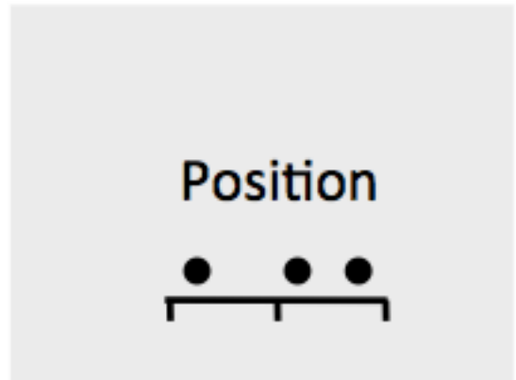
Shape

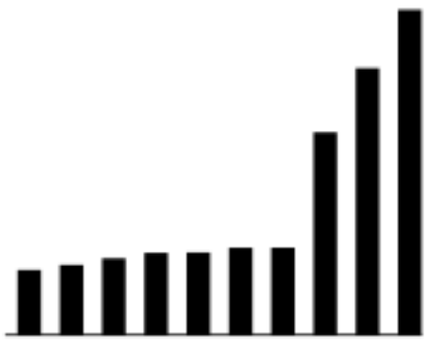
Categorie

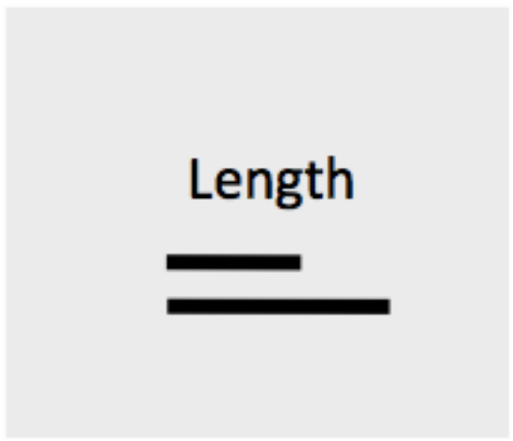
C. Mulbrandon
VisualizingEconomics.com

# Most Effective



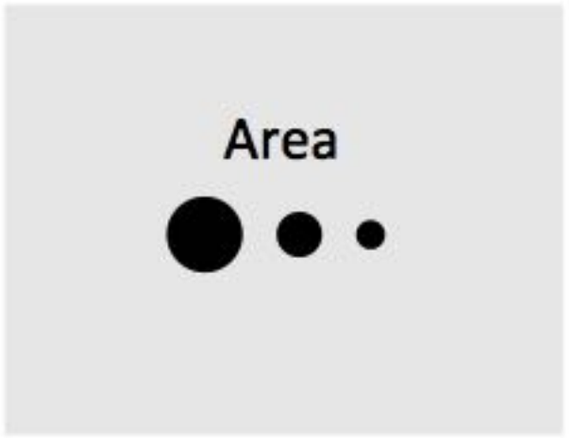







VisualizingEconomics.com

# Less Effective



# Least Effective

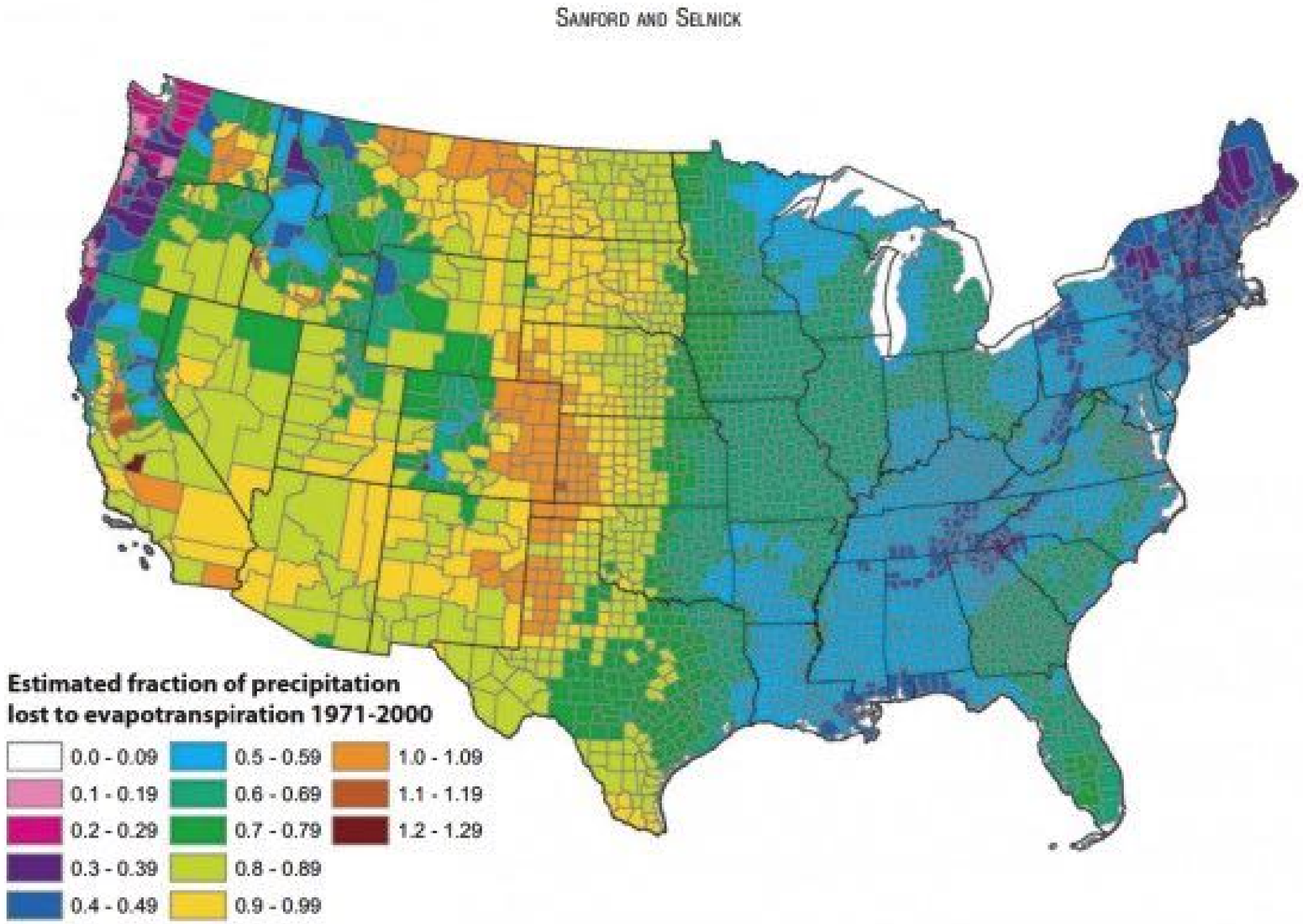
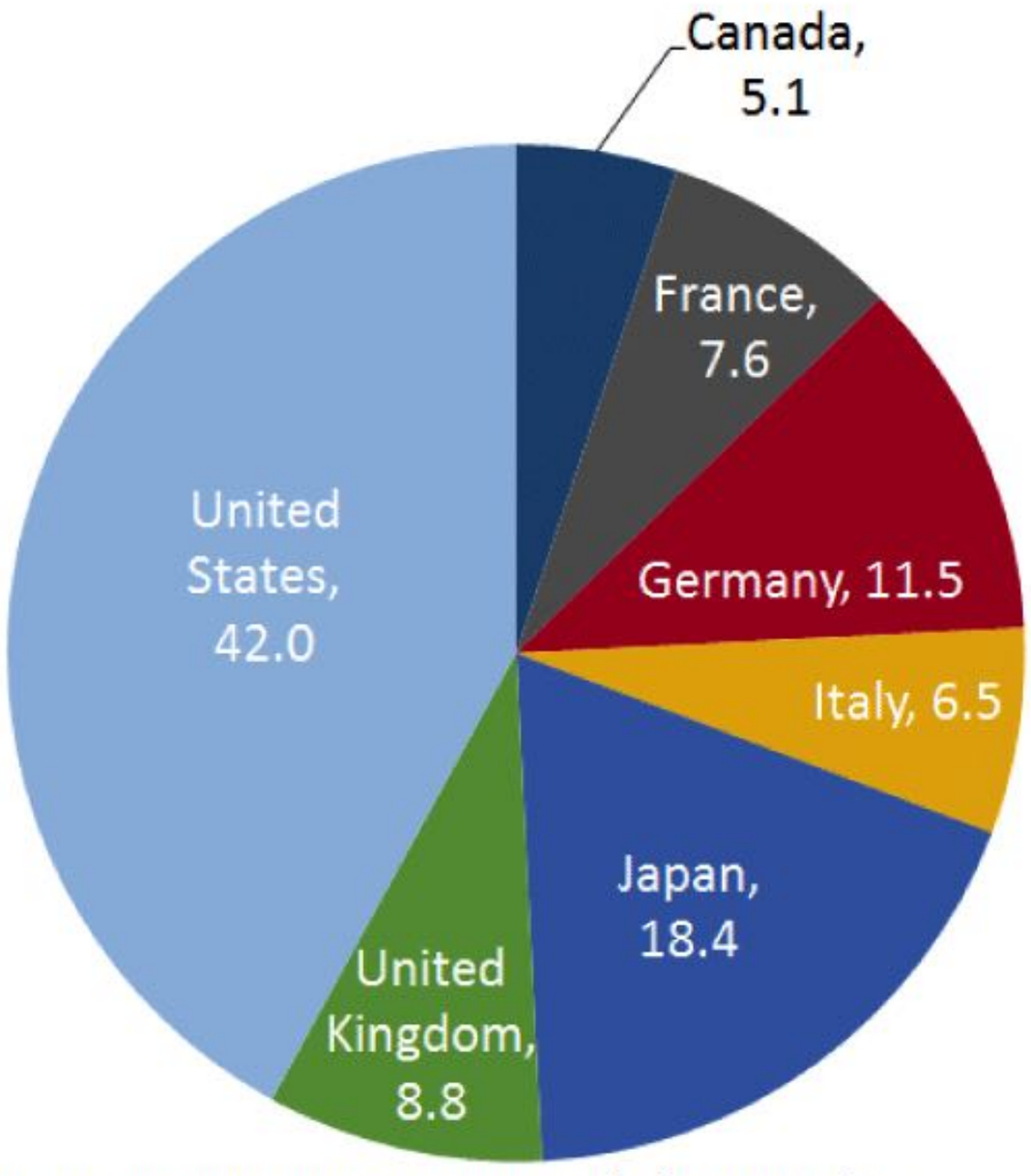


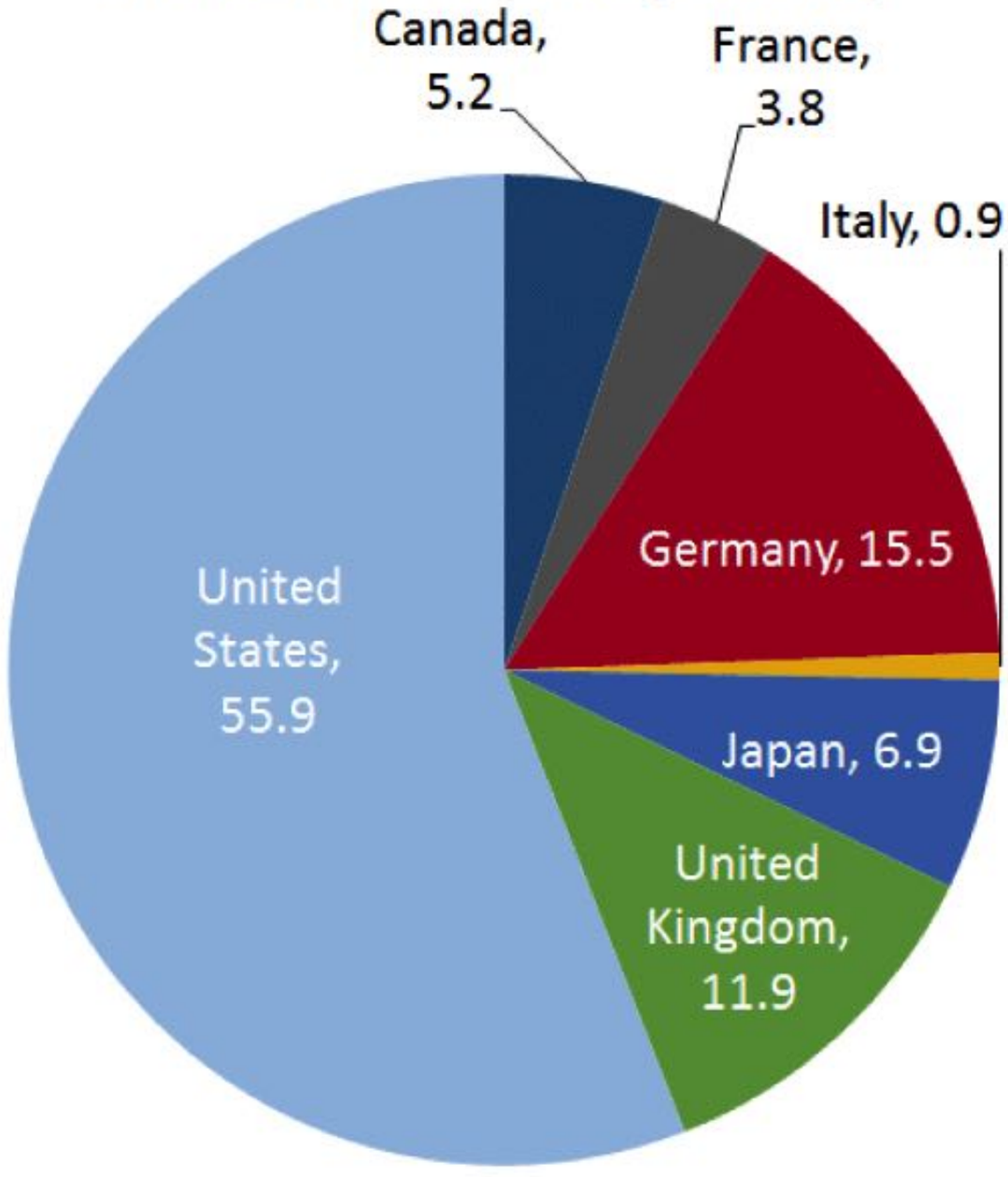

FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ($P$) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of $ET/P$ were made first at the 800-m resolution of the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

# Employment Growth in G-7 Countries, 2010:Q1-2016:Q3

**Share of Total Employment, Average 2010:Q1-2016:Q3 (Percent)**




**Share of Net Growth in Employment, 2010:Q1 to 2016:Q3 (Percent)**




Note: Components may not sum to totals due to rounding.
Source: Organisation for Economic Co-operation and Development; CEA calculations.

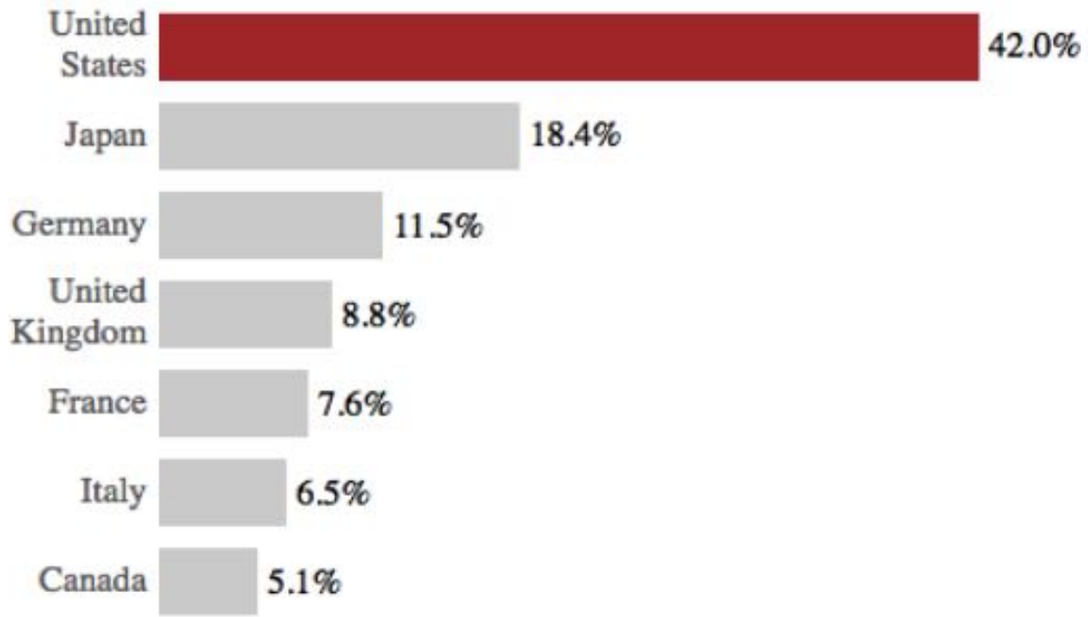
BUSINESS INSIDER

http://uk.businessinsider.com/us-economy-added-jobs-faster-than-all-g7-economies-combined-2017-1?r=US&IR=T

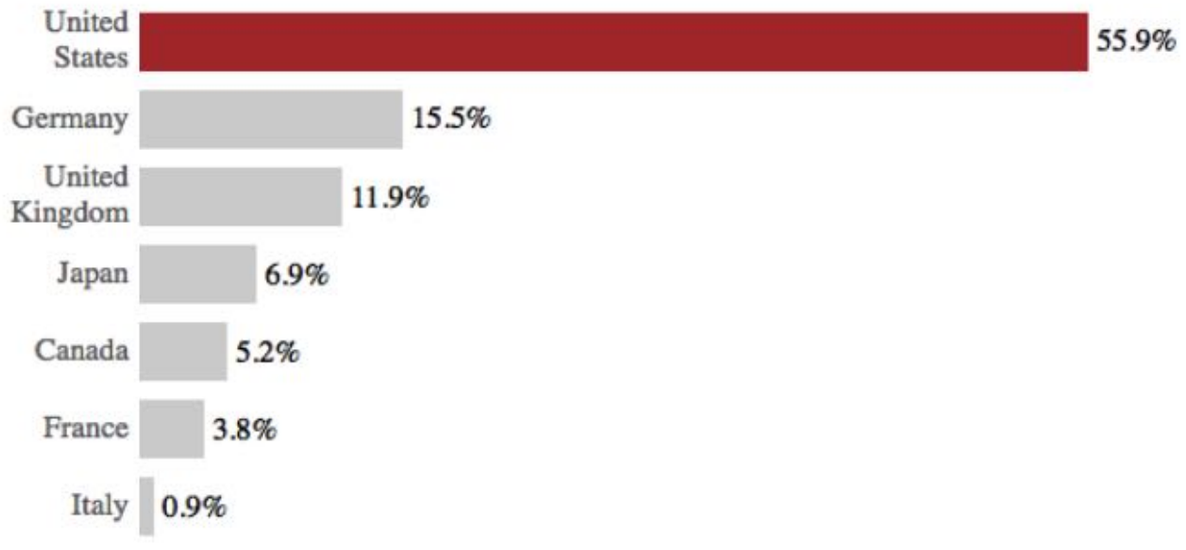
Since 2010, the United States has put more people back to work than all the other G-7 economies combined.

---

Although the United States comprises about two-fifths of total employment in the G-7...




...it has been responsible for more than 55% of the net employment growth since 2010.




Data source: Business Insider

Created by Andy Kriebel @VizWizBI

Undo Redo Reset

+ableau

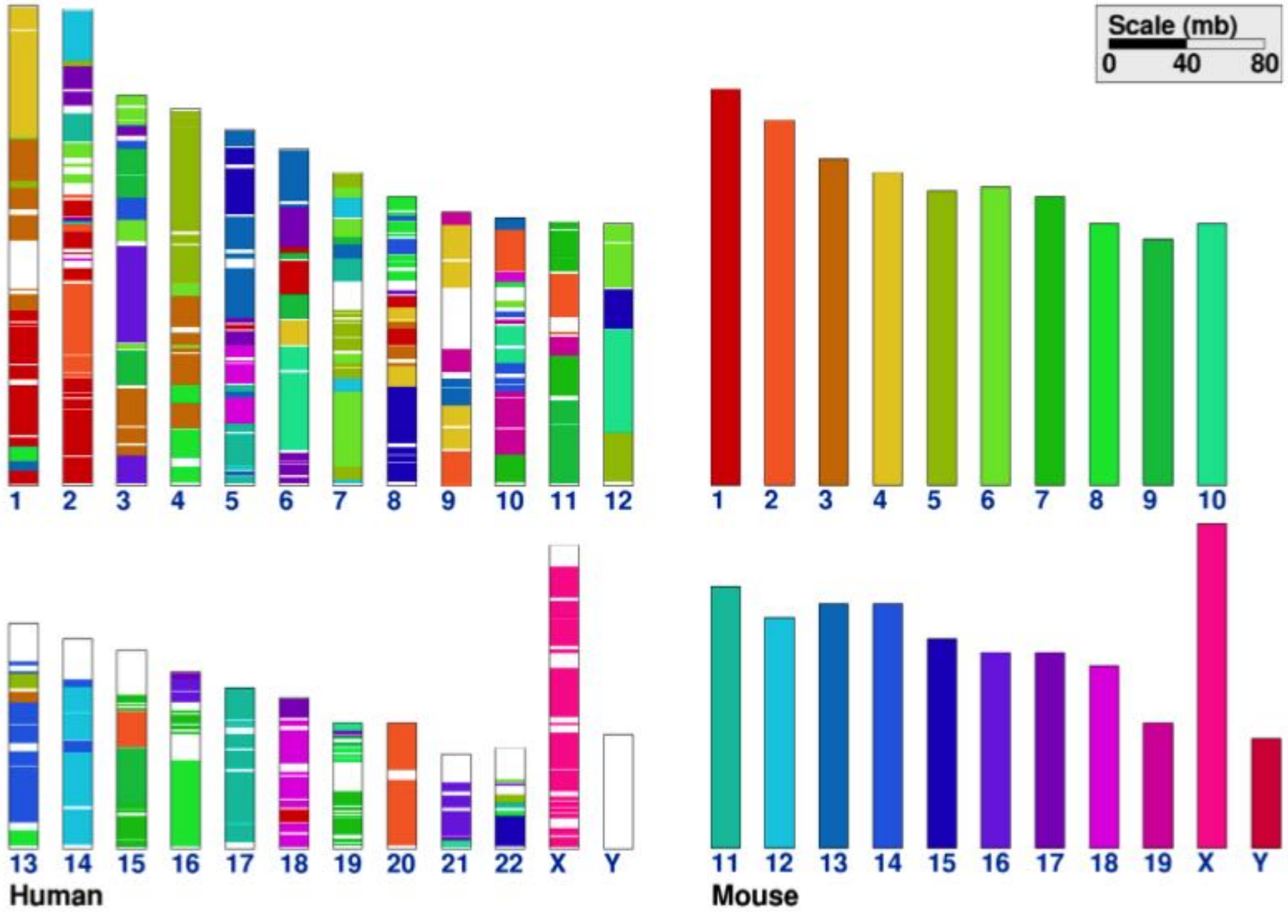
452 views | more by this author

Share Download Full Screen

A. Kriebel

# Use Color Strategically

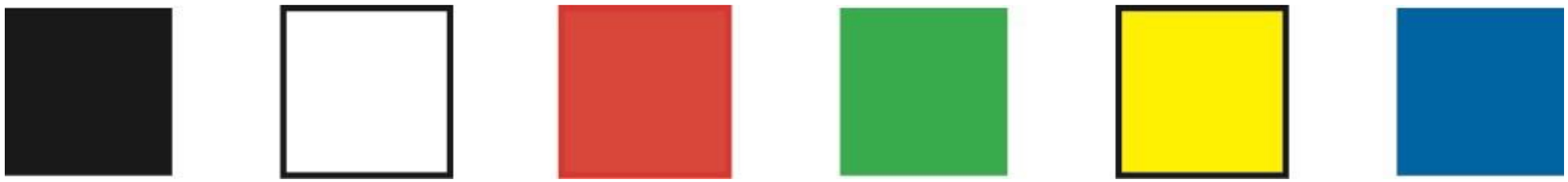
# Color Discriminability



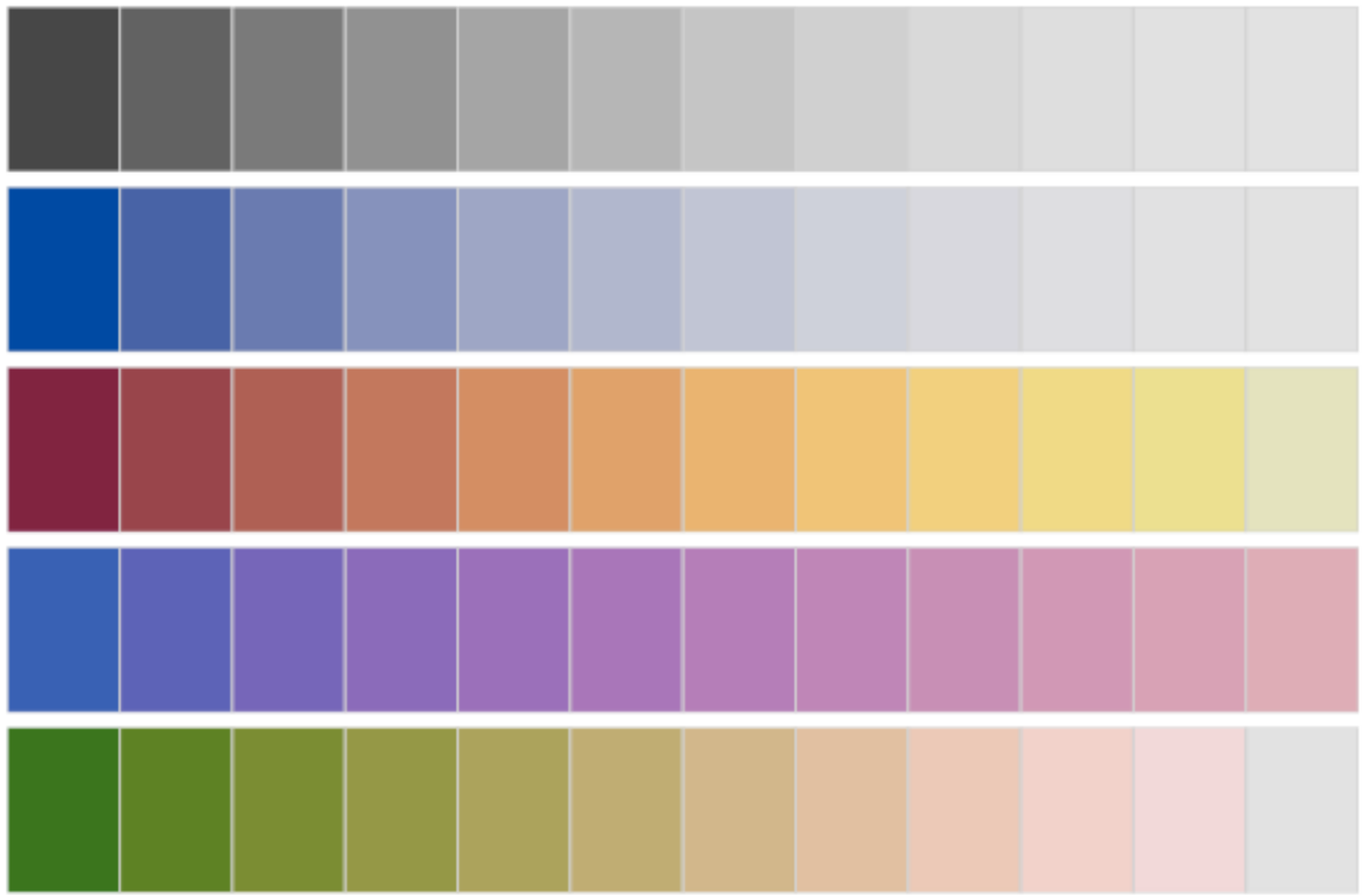

Sinha 2007

# Colors for Categories

Use no more than 5-8 colors at once

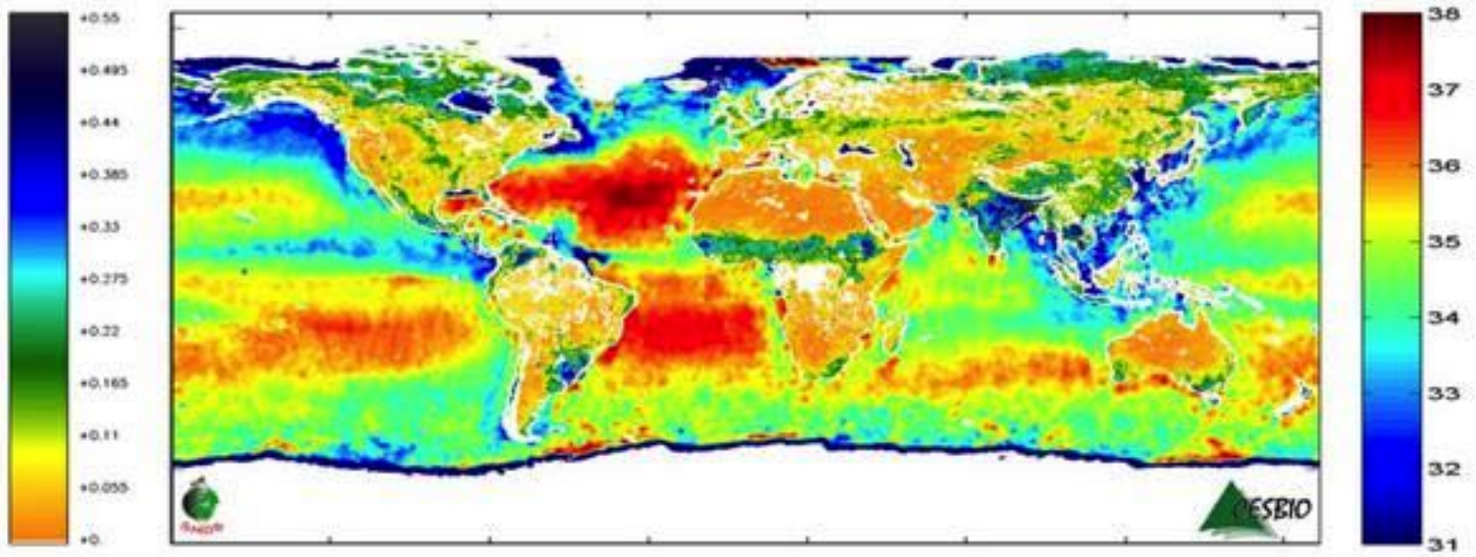


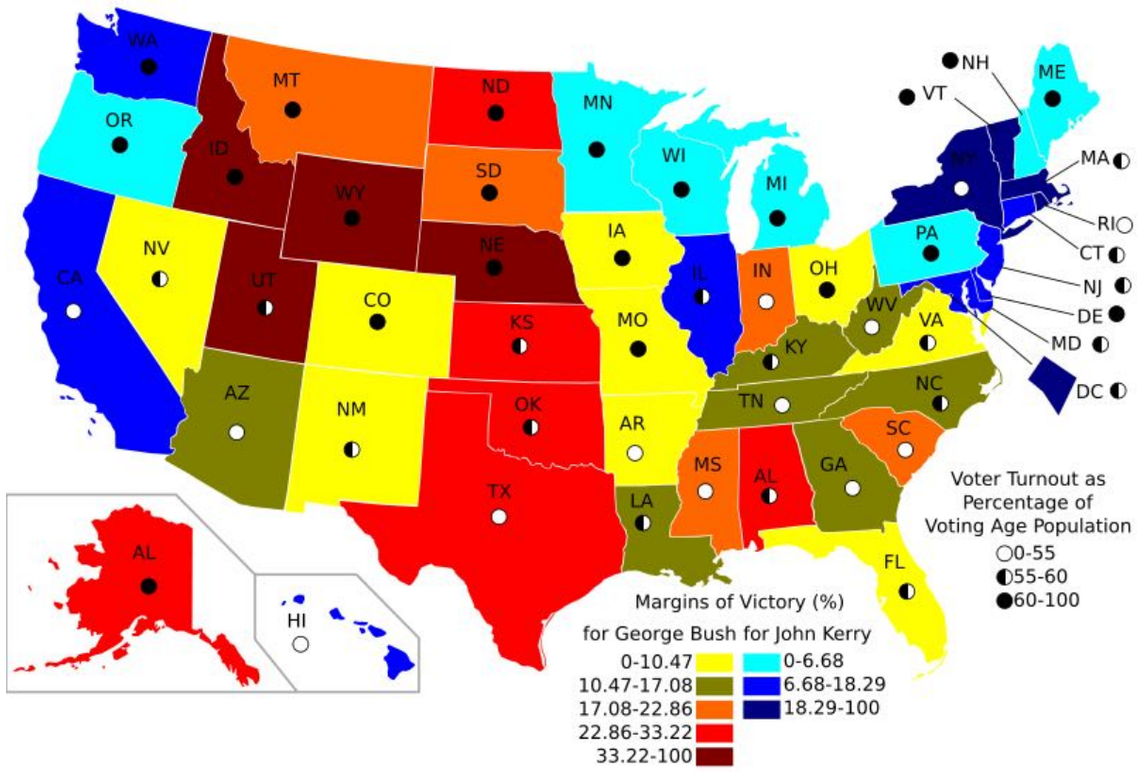
# Colors for Ordered Data



Zeilis et al, 2009, “Escaping RGBland: Selecting Colors for Statistical Graphics”

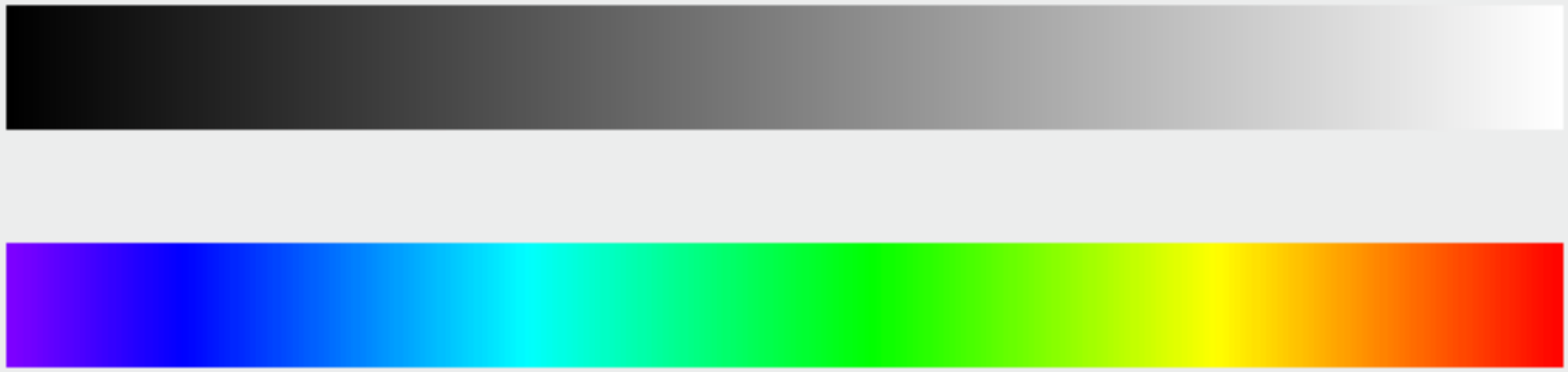
# Colors for Maps



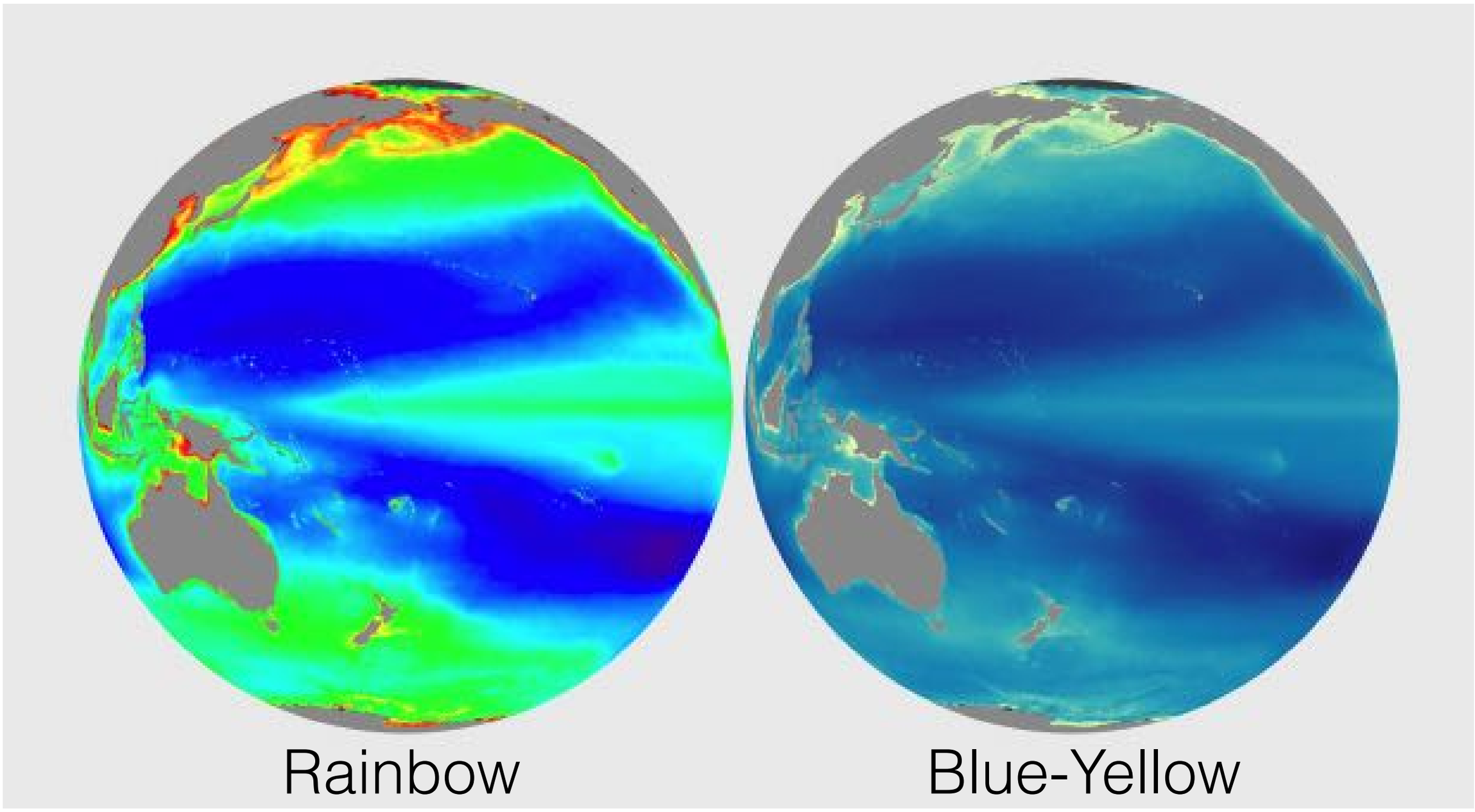



# Rainbow Colormap

Perceptually nonlinear



R. Simmon

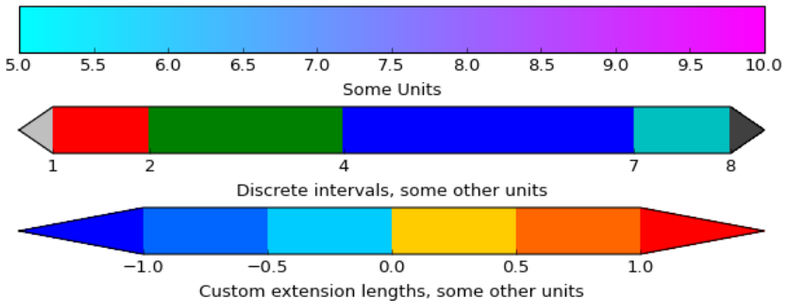


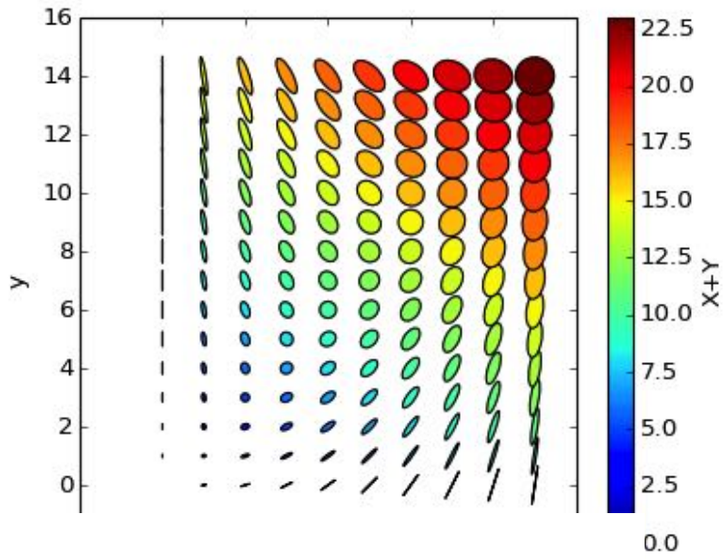
Rainbow

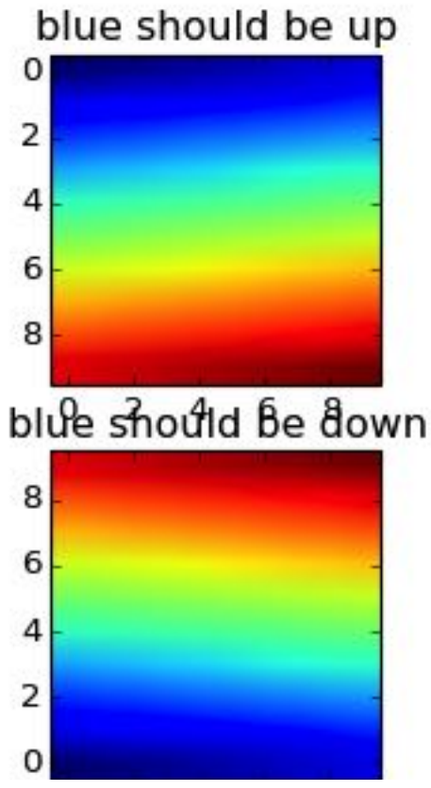
Blue-Yellow

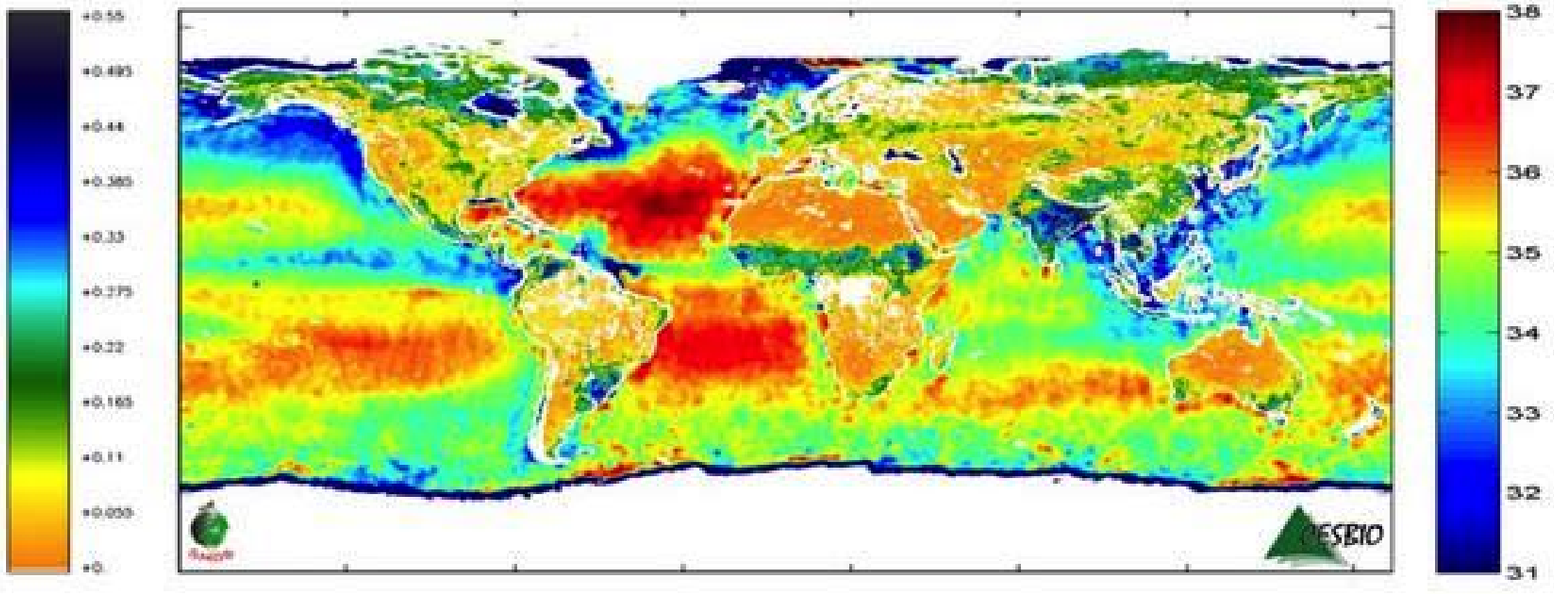
R. Simmon

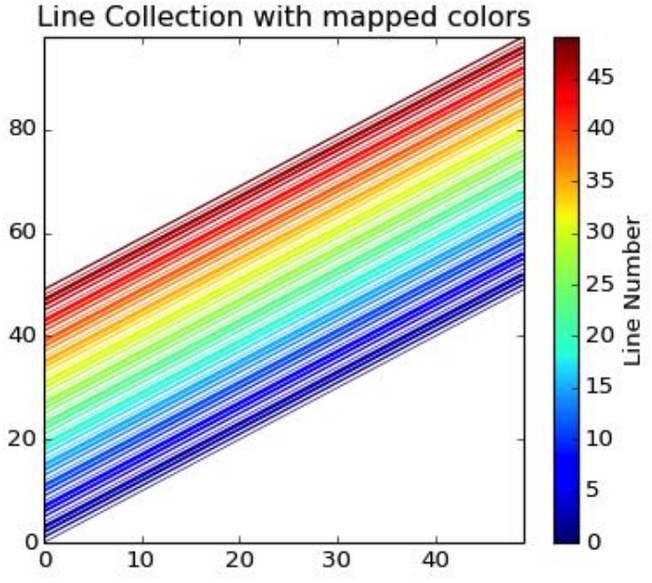
# Avoid Rainbow Colors!



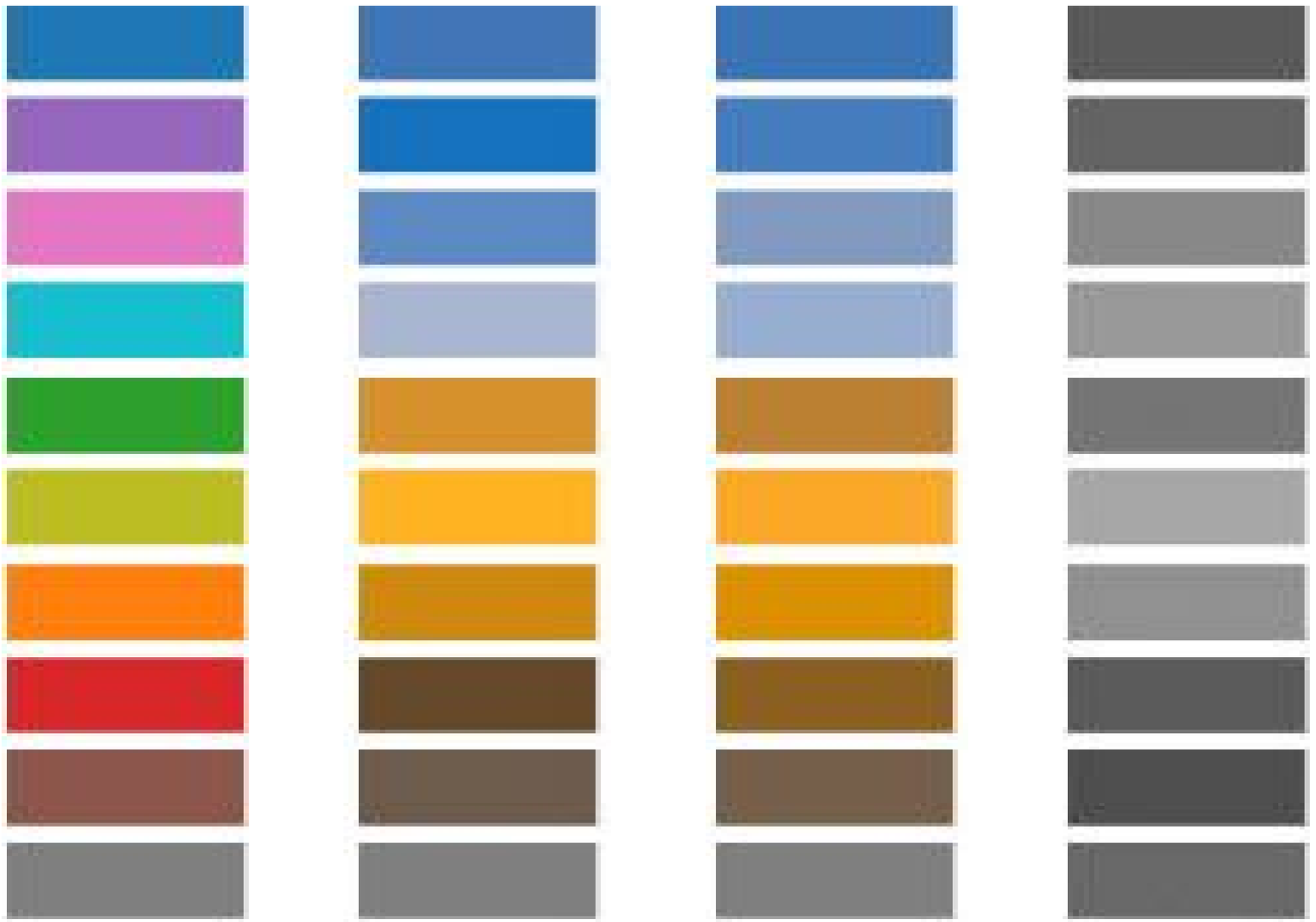




matplotlib

# Color Blindness



Normal | Protanope | Deuteranope | Lightness

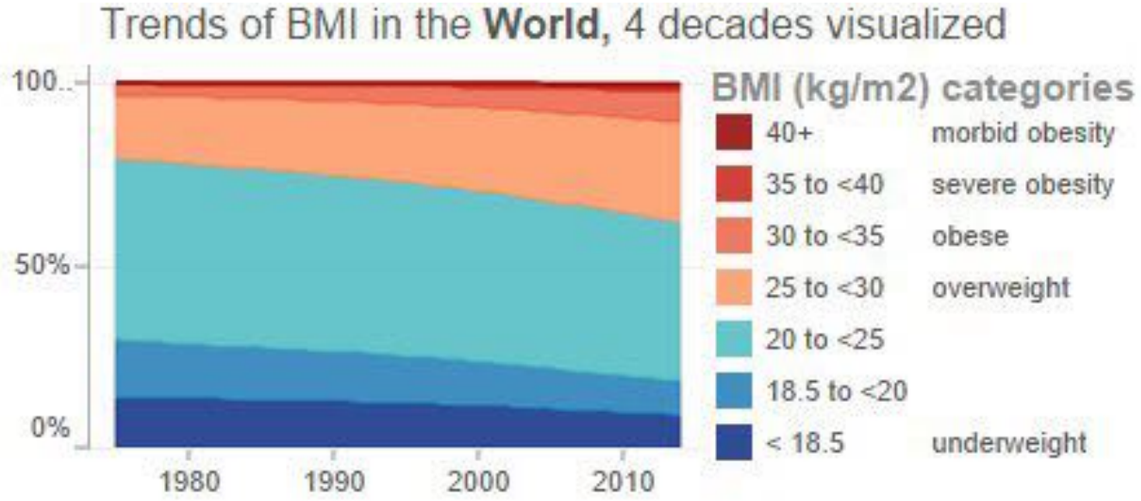
Based on slide from Stone

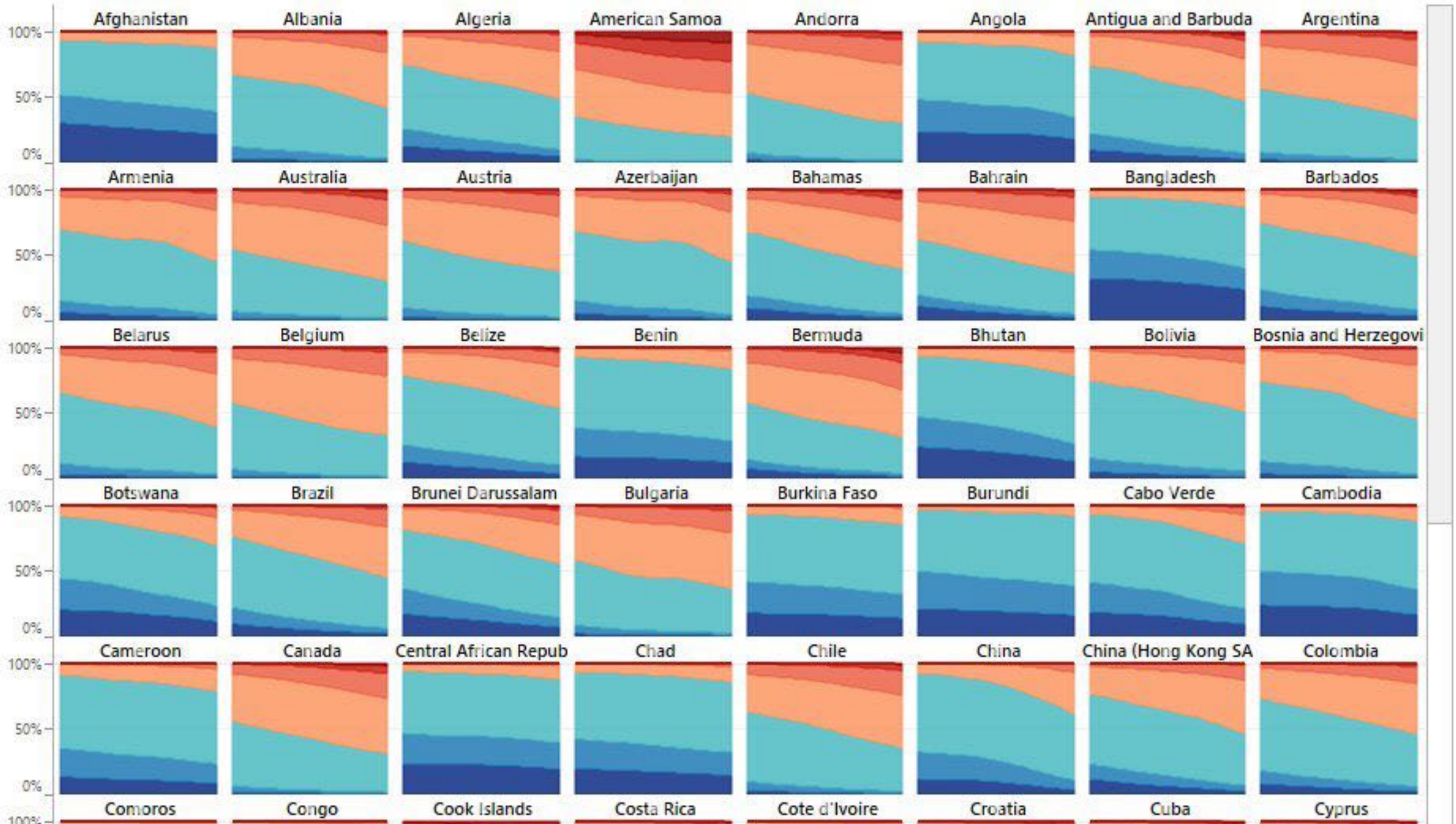
# Trends in Adult Body-Mass Index in 200 Countries from 1975 to 2014

Over the past 40 years, we have changed from a world in which **underweight** prevalence was more than double that of obesity, to one in which **more people are obese than underweight**

**Select sex to show**
Men

**Sort countries by**
Alphabetic



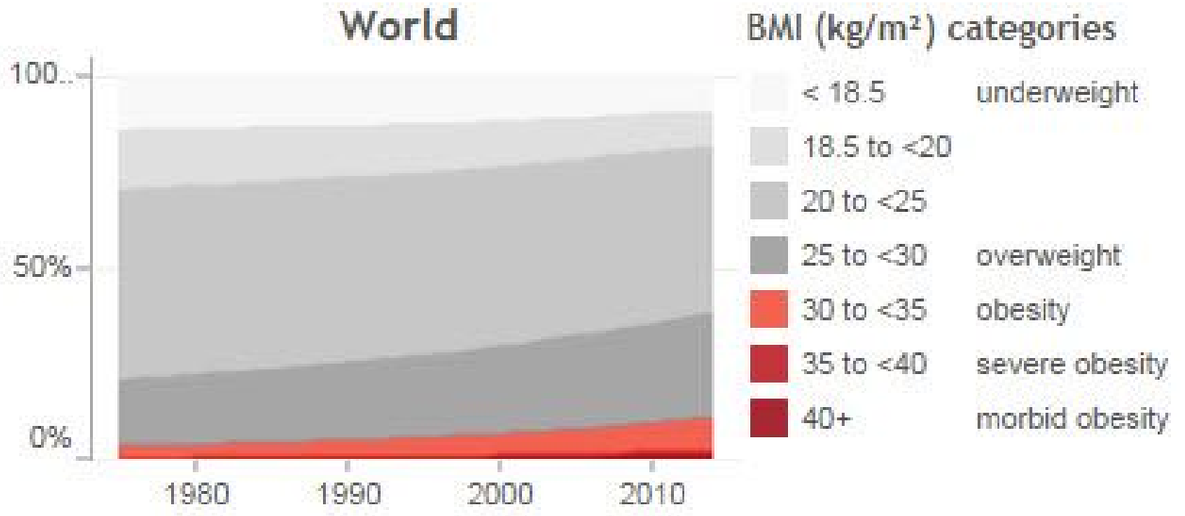
# Trends in Adult Body-Mass Index in 200 Countries from 1975 to 2014

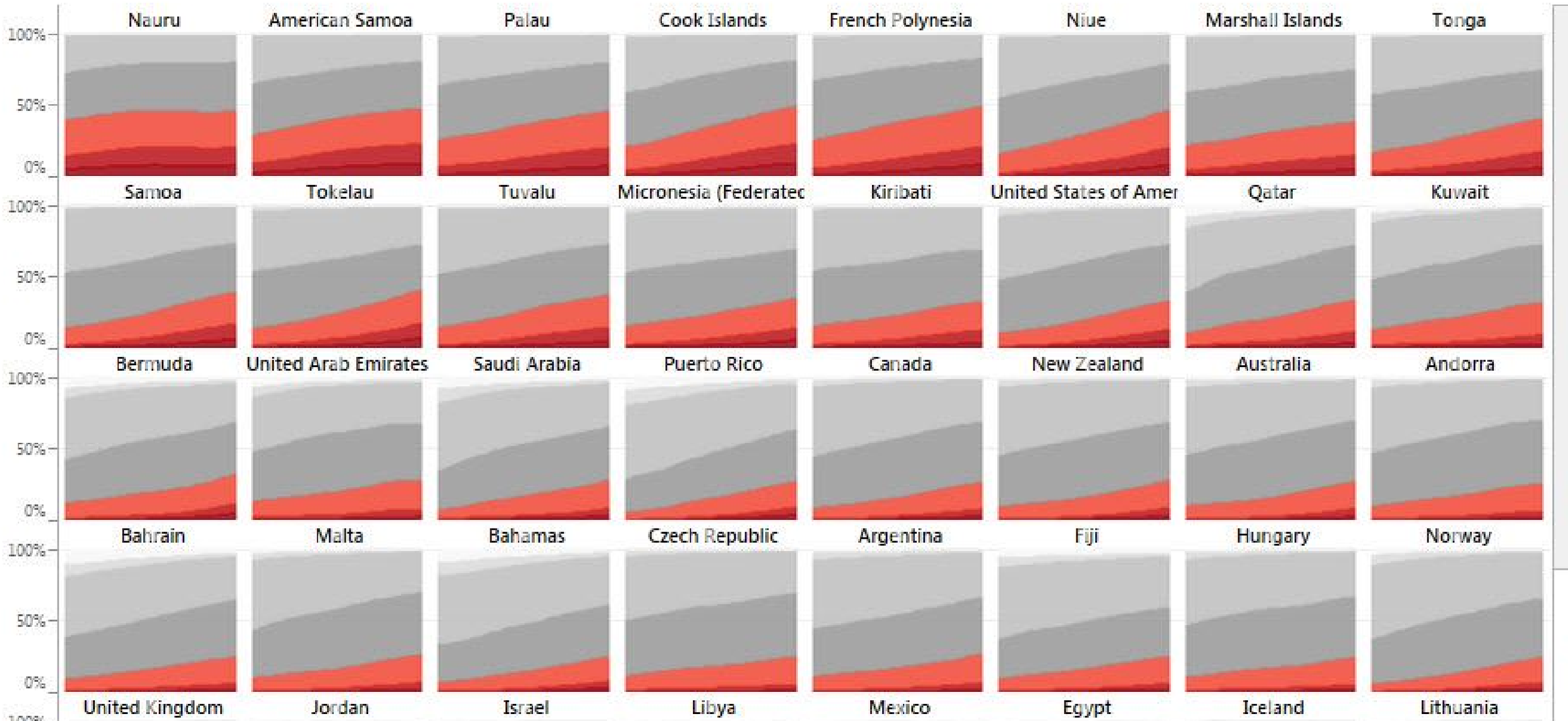
## Age-standardized prevalence (%) by BMI categories

Over the past 40 years, we have changed from a world in which **underweight** prevalence was more than double that of **obesity**, to one in which **more people are obese than underweight**

**Select sex to show**
Men

**Sort countries by**
Obesity, Descending



# Tell a Story with Data

New York Times

# Effective Visualizations

1. Have graphical integrity
2. Keep it simple
3. Use the right display
4. Use color strategically
5. Tell a story with data