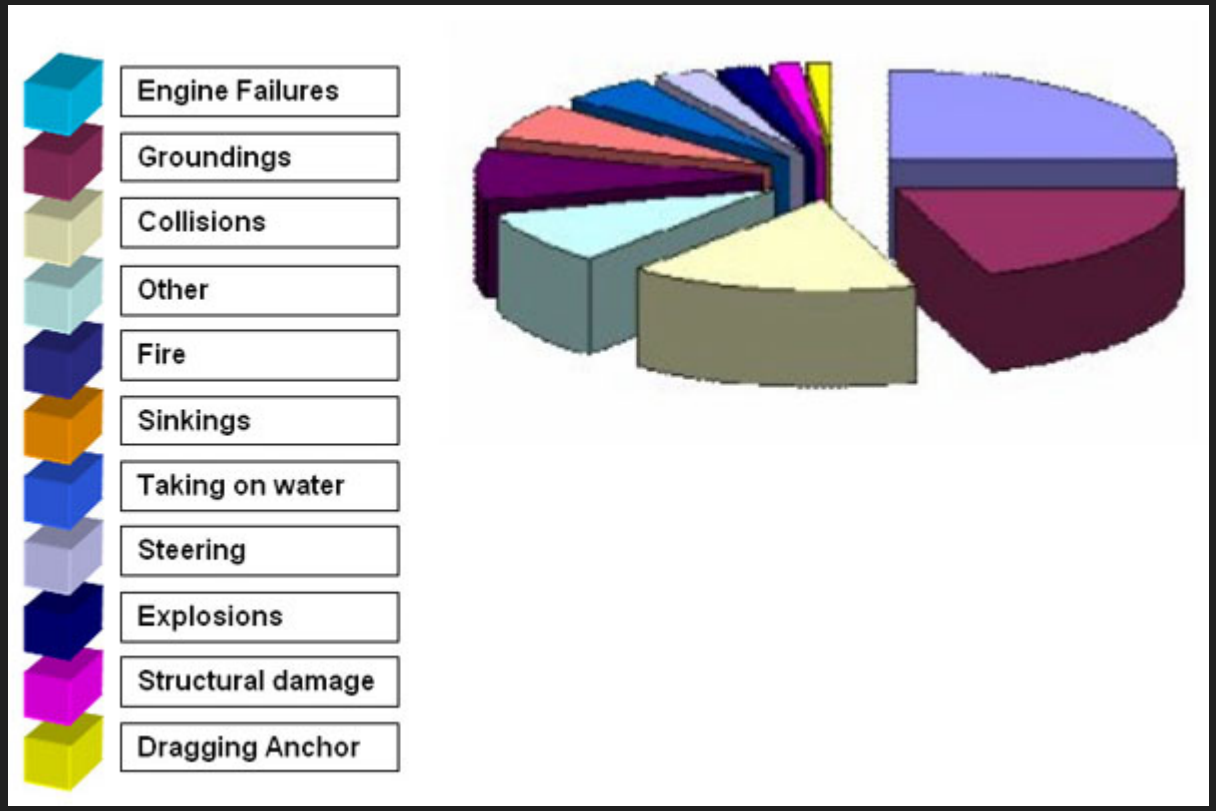
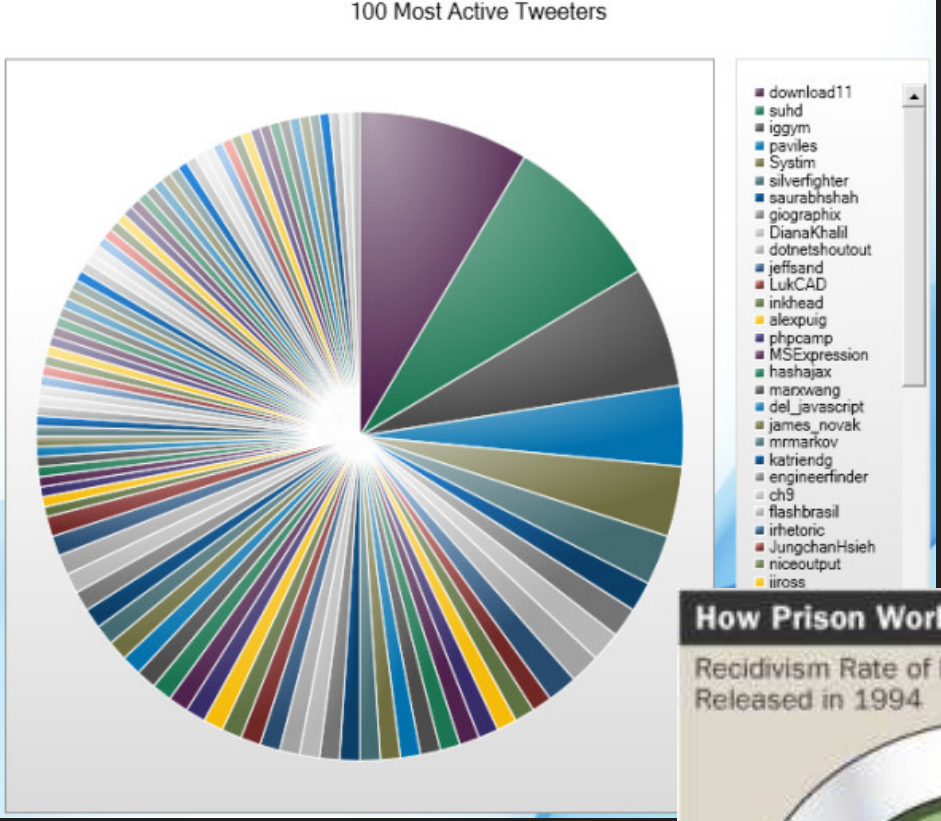
Graphical Integrity

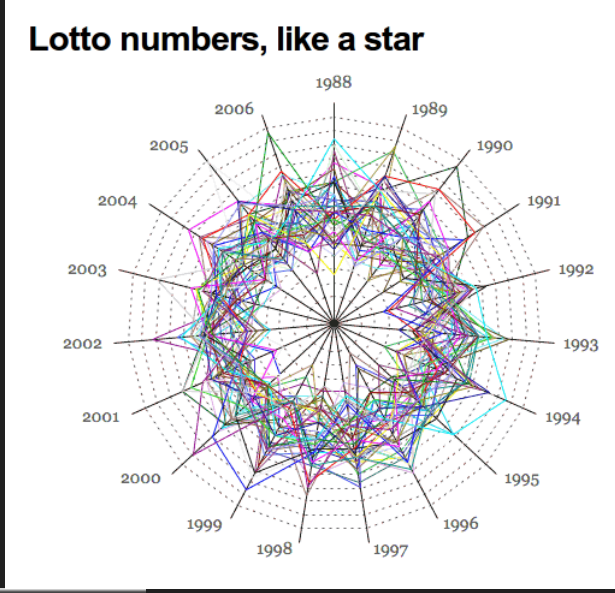
Alark Joshi

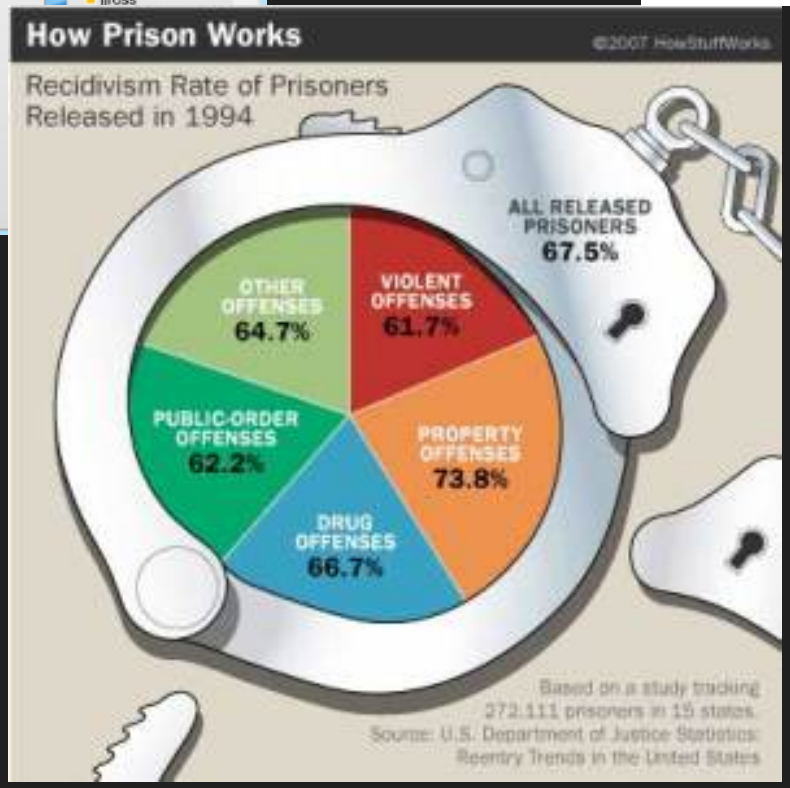
Maritime and Coastguard Agency



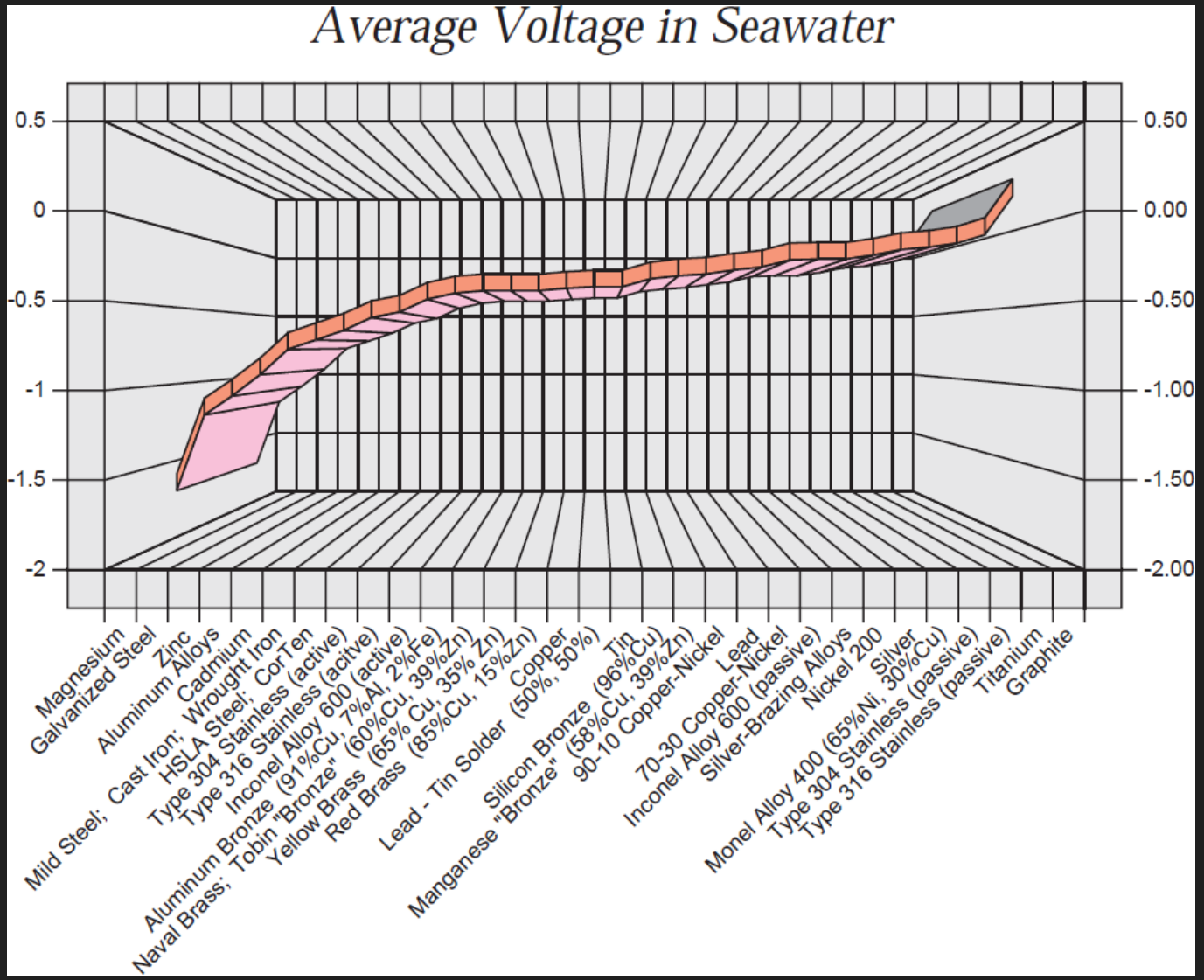
- Some of the colors in the legend don't match the slices.


100 Most Active Tweeters
download11
suhd
iggym
paviles
Systim
silverfighter
saurabhshah
giographix
DianaKhalil
dotnetshoutout
jeffsand
LukCAD
inkhead
alexpuig
phpcamp
MSExpression
hashajax
marxwang
del_javascript
james_novak
mrmarkov
katriendg
engineerfinder
ch9
flashbrasil
irhetoric
JungchanHsieh
niceoutput
iiross

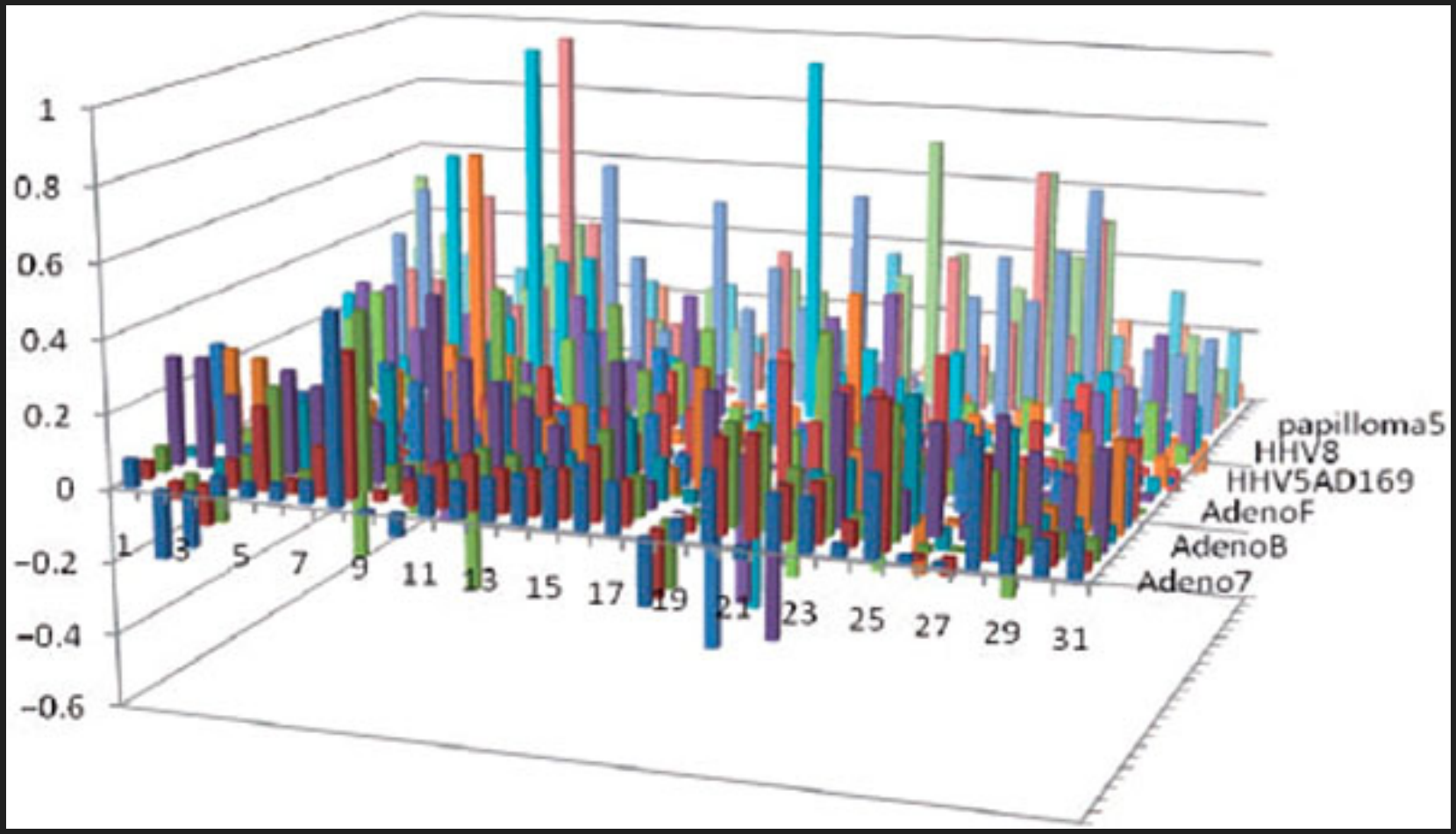

Lotto numbers, like a star
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006

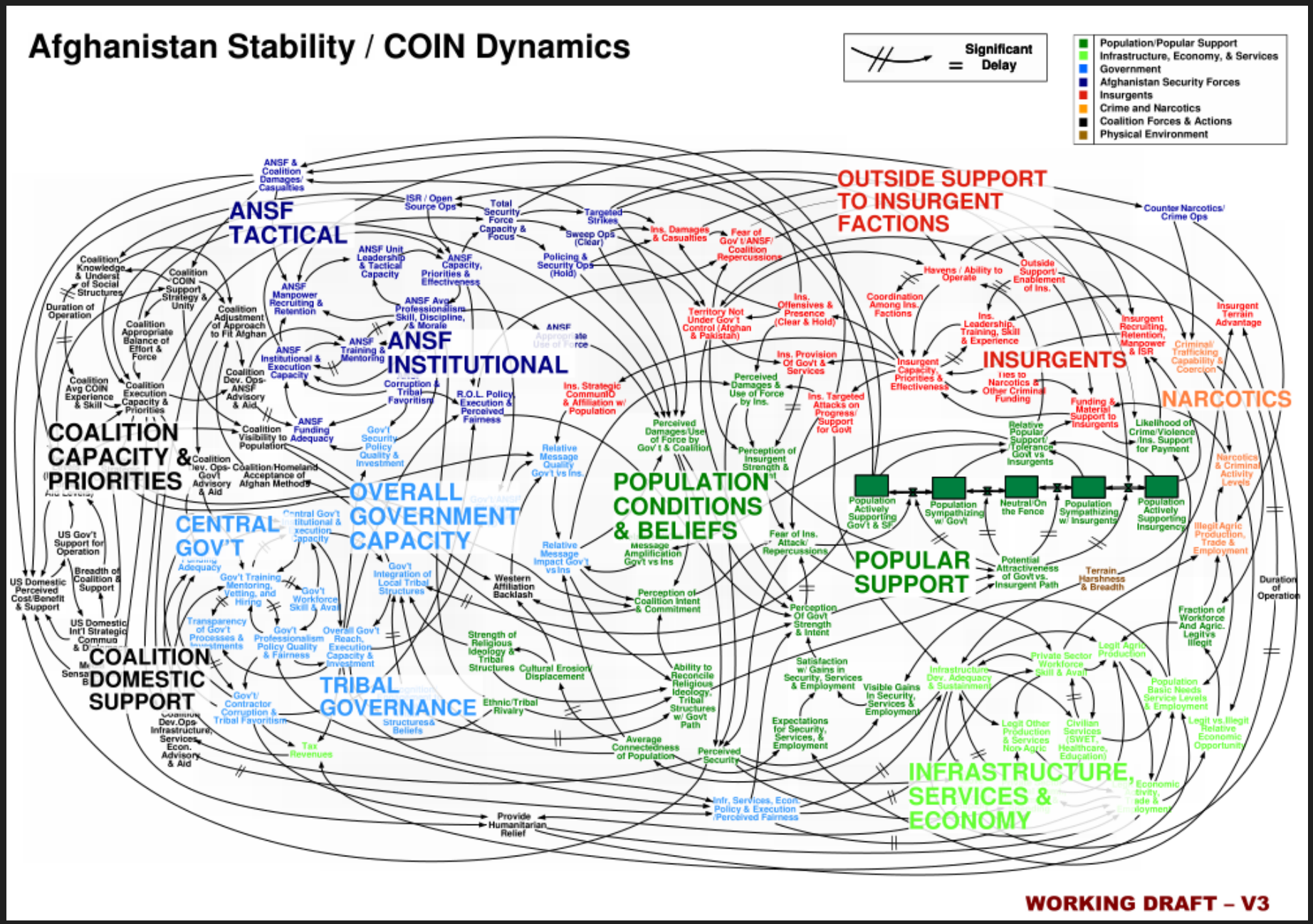

How Prison Works
©2007 HowStuffWorks
Recidivism Rate of Prisoners Released in 1994
ALL RELEASED PRISONERS 67.5%
OTHER OFFENSES 64.7%
VIOLENT OFFENSES 61.7%
PUBLIC-ORDER OFFENSES 62.2%
PROPERTY OFFENSES 73.8%
DRUG OFFENSES 66.7%
Based on a study tracking 272,111 prisoners in 15 states.
Source: U.S. Department of Justice Statistics: Reentry Trends in the United States

Overuse of 3D in a visualization

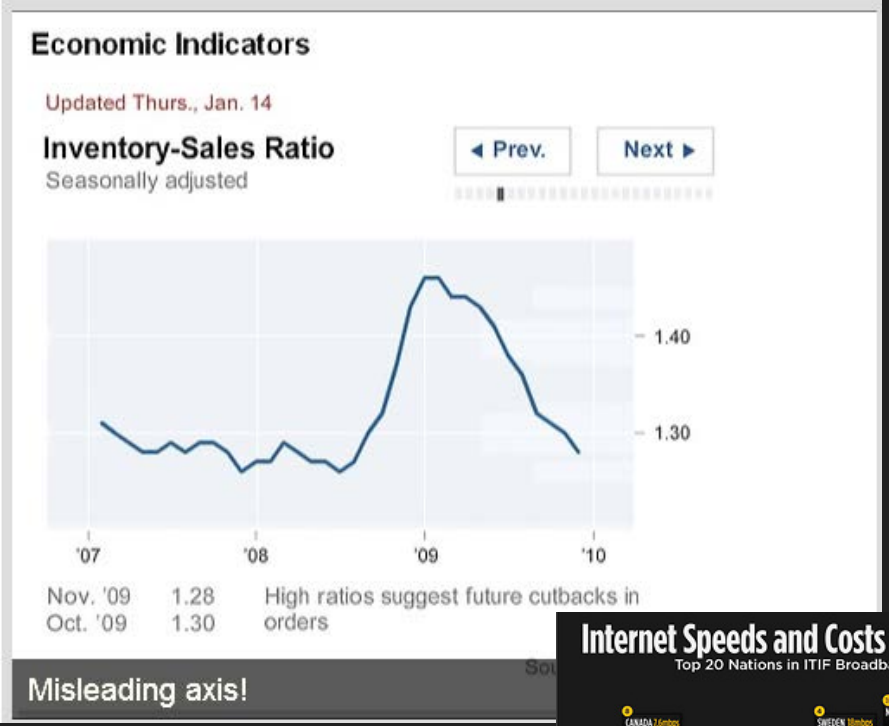


SIR Score – Oxford Journals

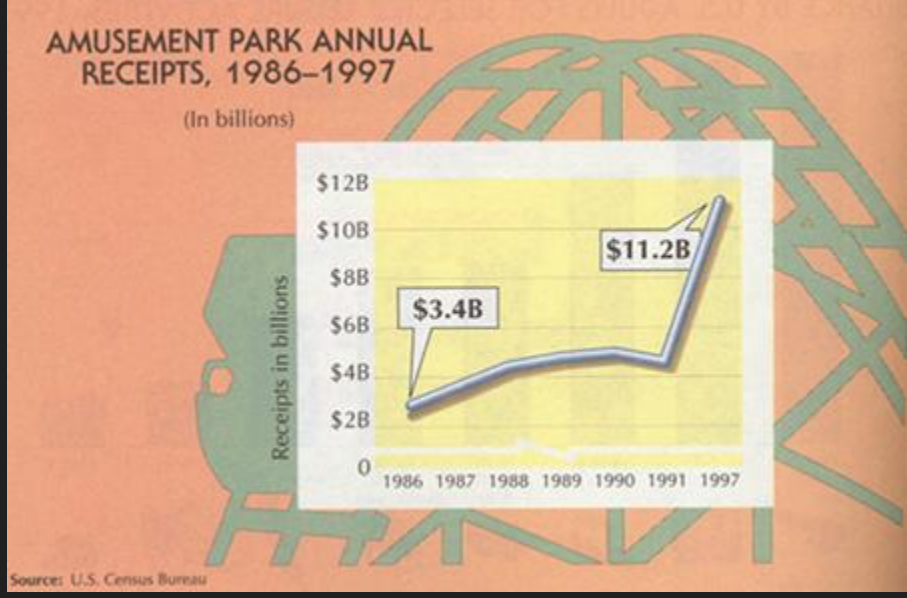




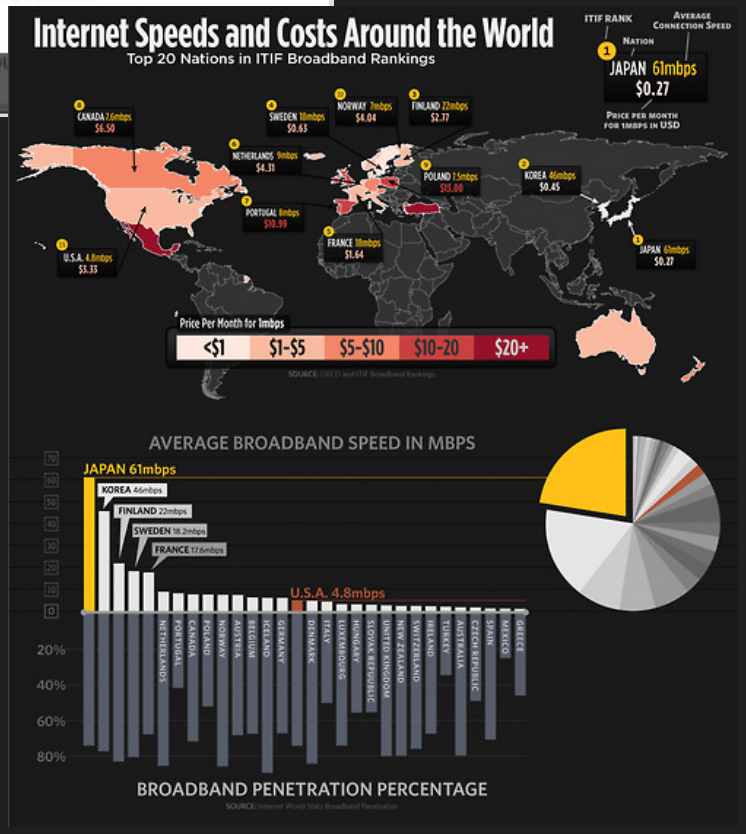
Credits: *yooniehoh.tumblr.com*



Credits: *eyegames.tumblr.com*



Credits: *chartbum.tumblr.com*



Credits: *kylemiller.tumblr.com*

Graphical Integrity

- Statistical charts = Lies?
- John Tukey, one of the first statisticians to make graphics respectable
- Unfortunately, with the widespread use of technology, making graphical representations is far too easy
- Deception is easy and convenient
  - Must be identified

Graphical Integrity

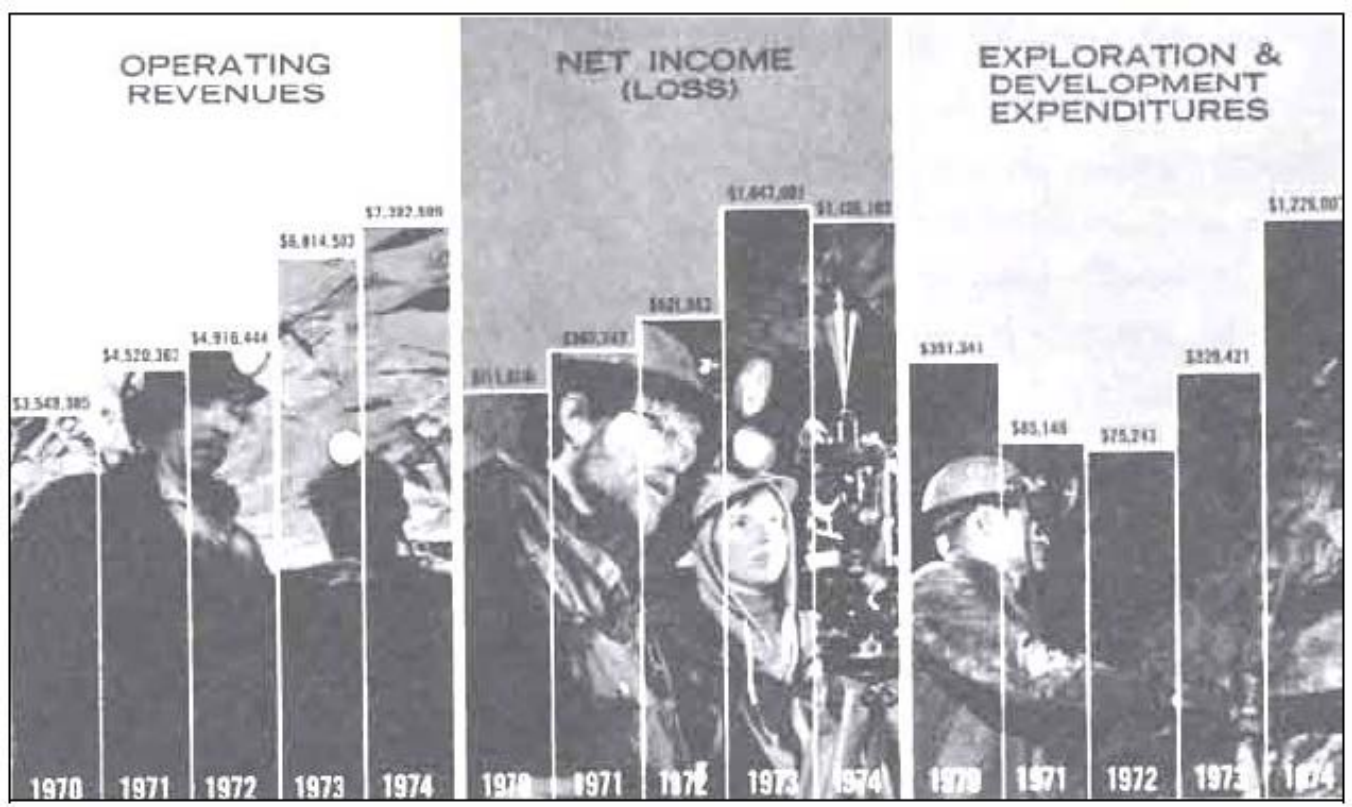
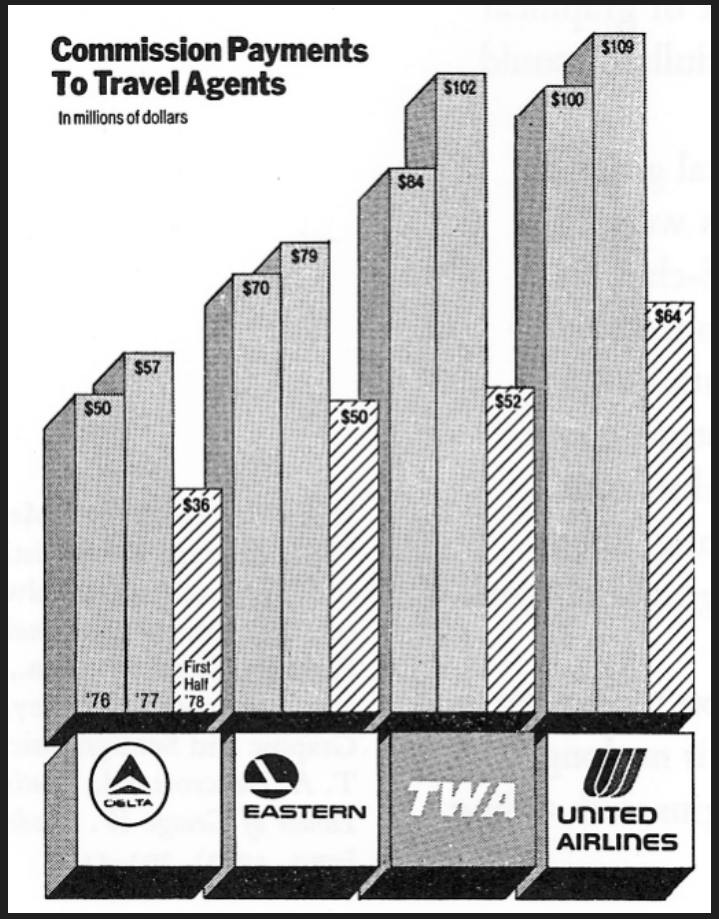


Figure 1: Day Mines, Inc., 1974 Annual Report, pg. 1
Source: Tufte, Edward R., *The Visual Display of Quantitative Information*, 2001

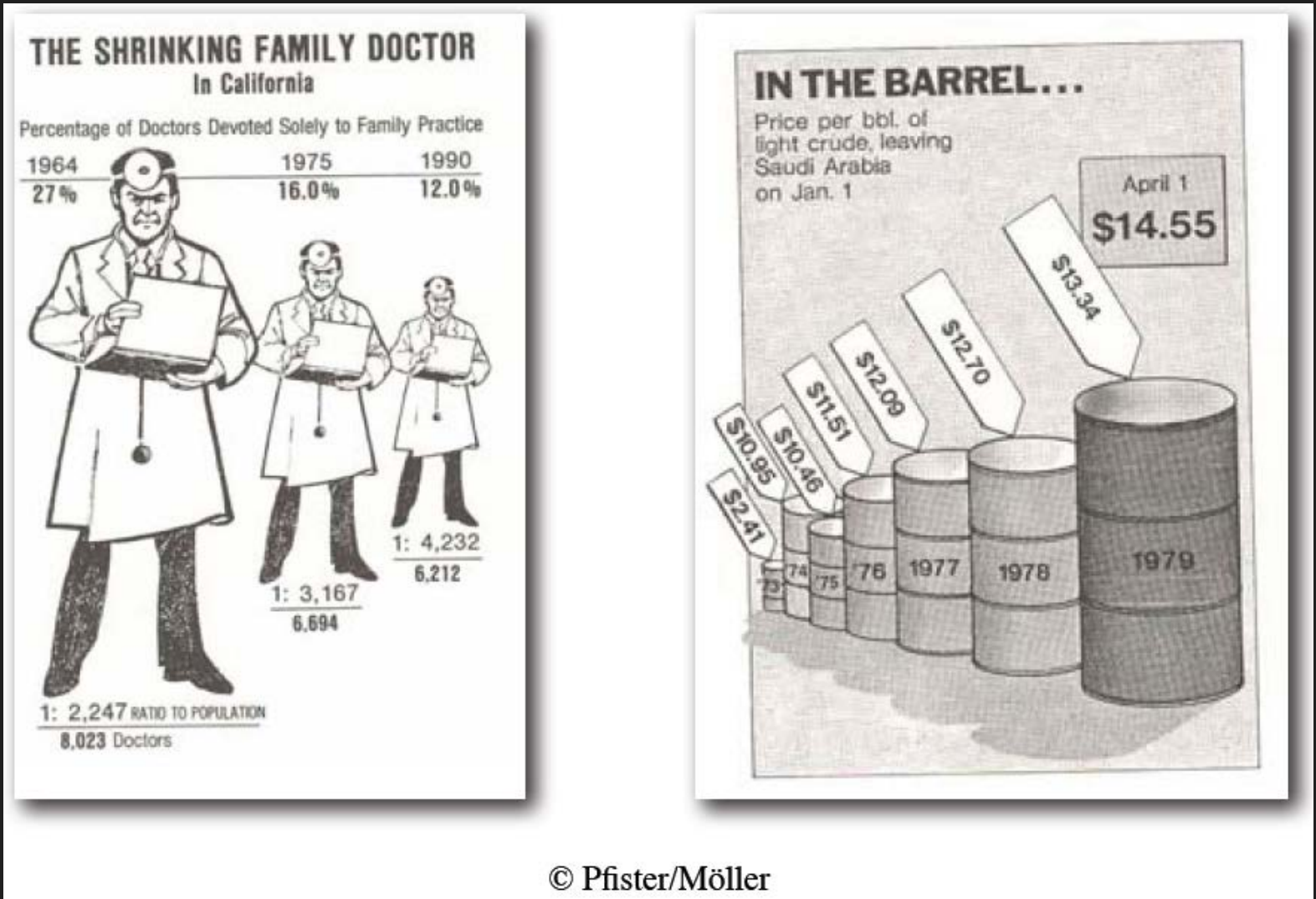
- Negative income in 1970 in the middle panel, masked by having the bars begin at approximately $4,200,000

Graphical Integrity



- Comparing six months of 1978 with 1976, 1977.

Graphical Integrity

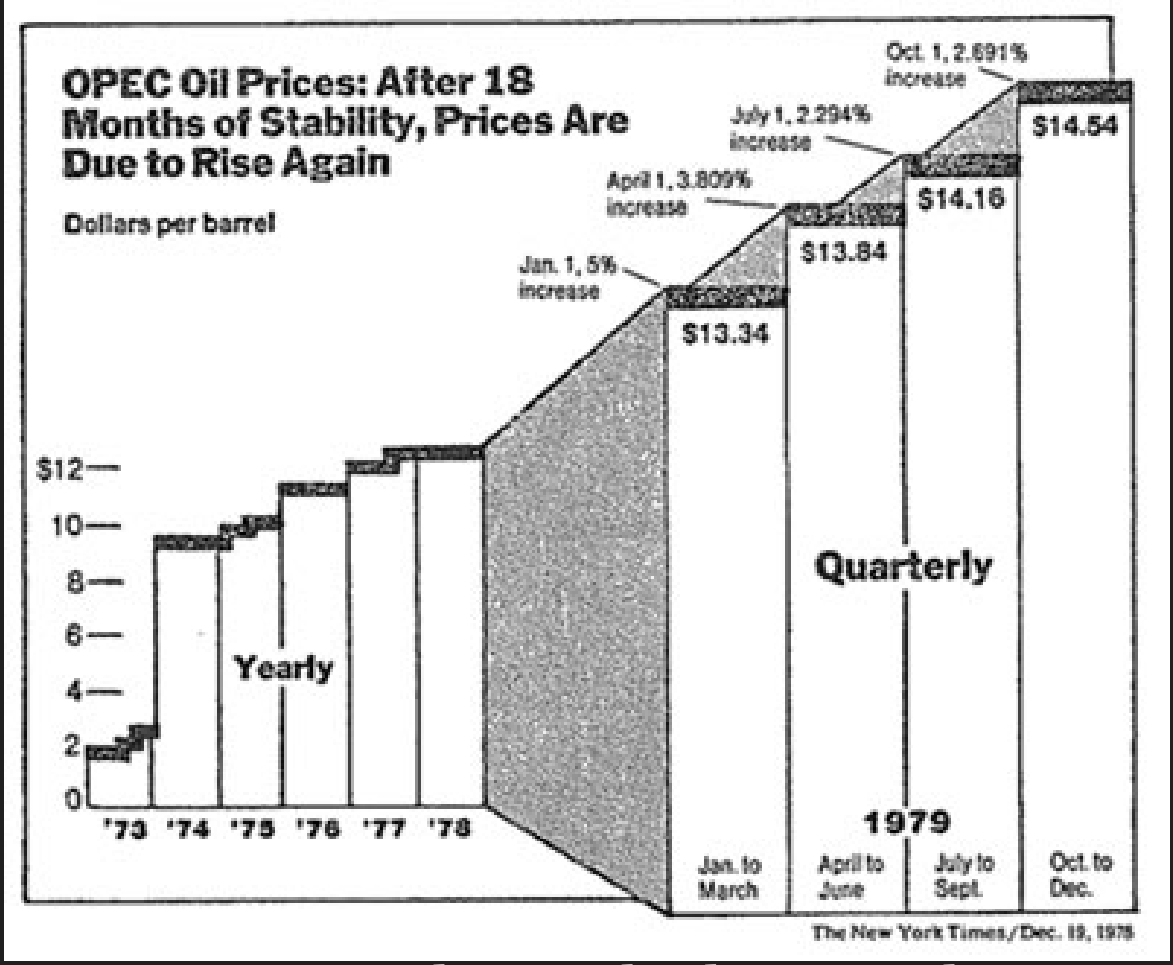


Inconsistent Visual area and numeric measure

Graphical Integrity

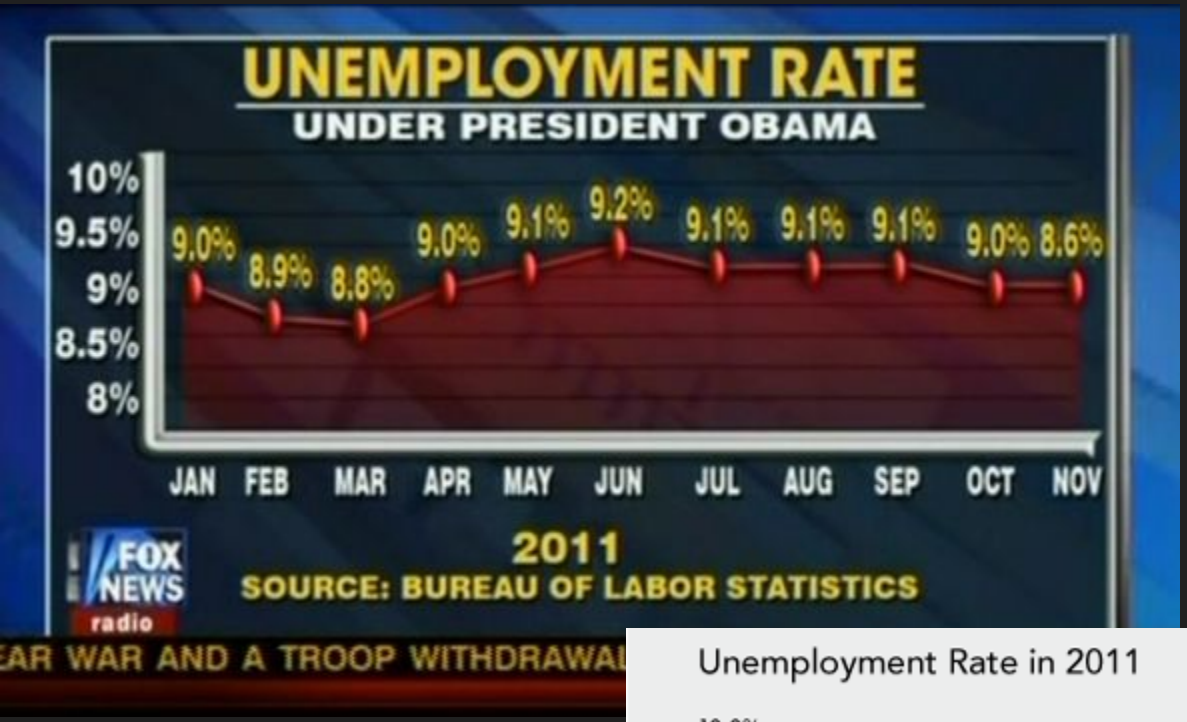
- Close valued numbers should be appropriately represented

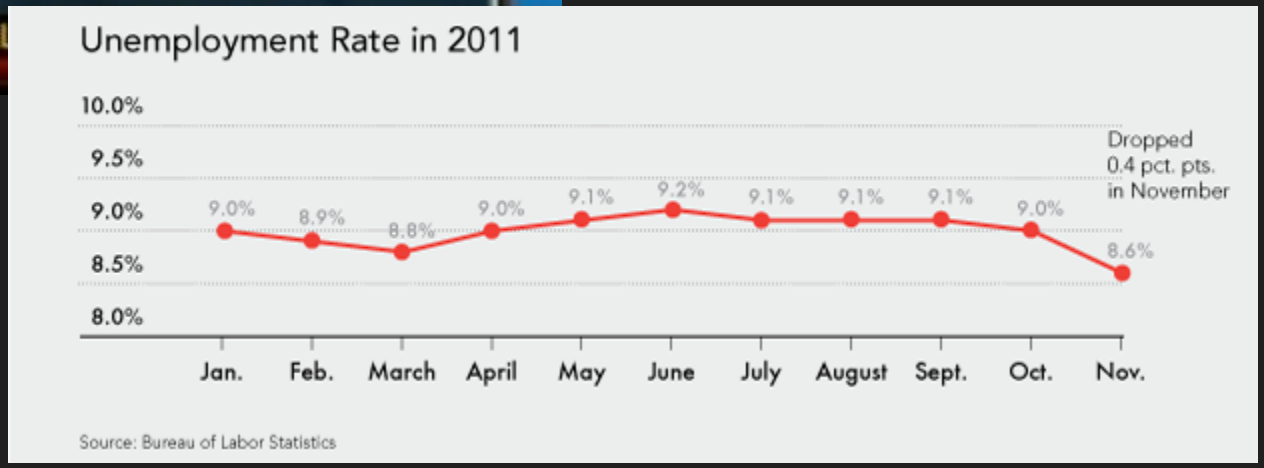
Graphical Integrity



- Inconsistent axis through the graph

Graphical Integrity

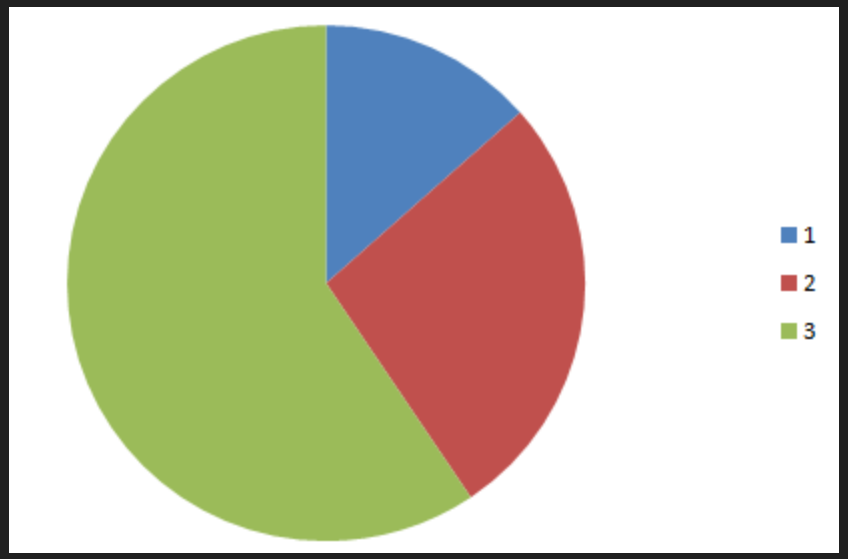




Redesigned by Nathan Yau at FlowingData.com

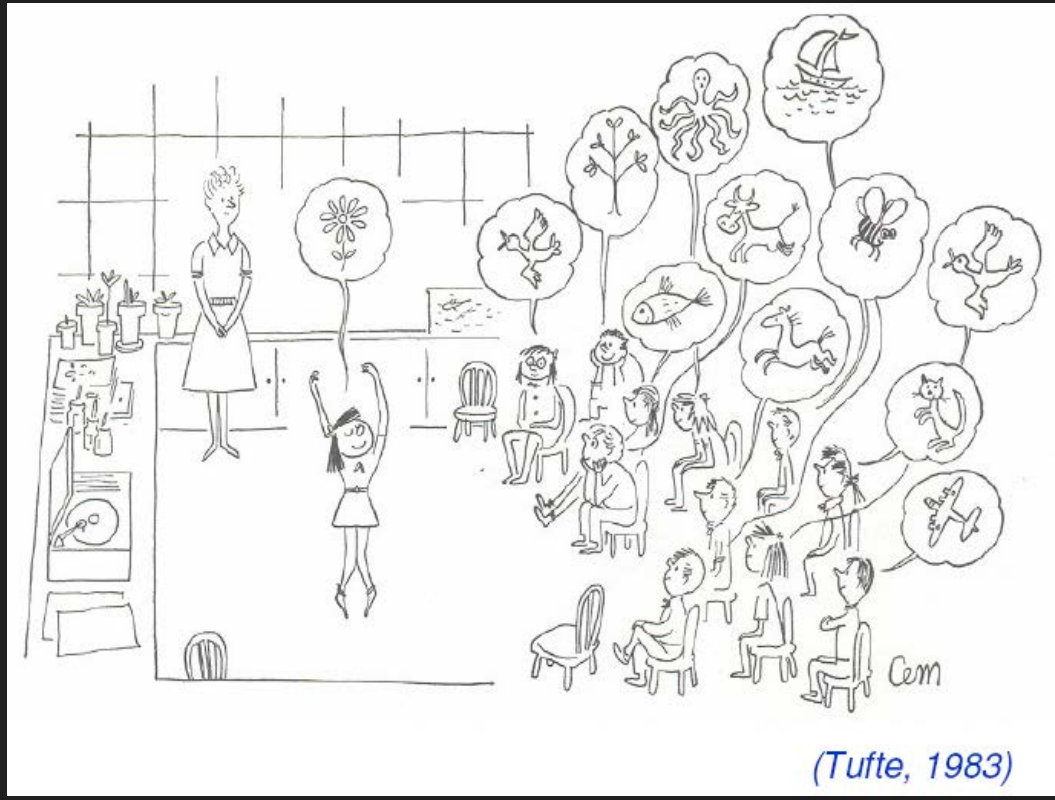
Avoiding Distortion in a Data Graphic

- Visual Representation of data should be consistent with the numerical representation
- Visual differences/perceived differences should match data differences



Misperception

- Misperception and miscommunication are not limited to statistical graphics
- Conduct user evaluation to evaluate efficacy of visual representations



(Tufte, 1983)

Graphical Integrity

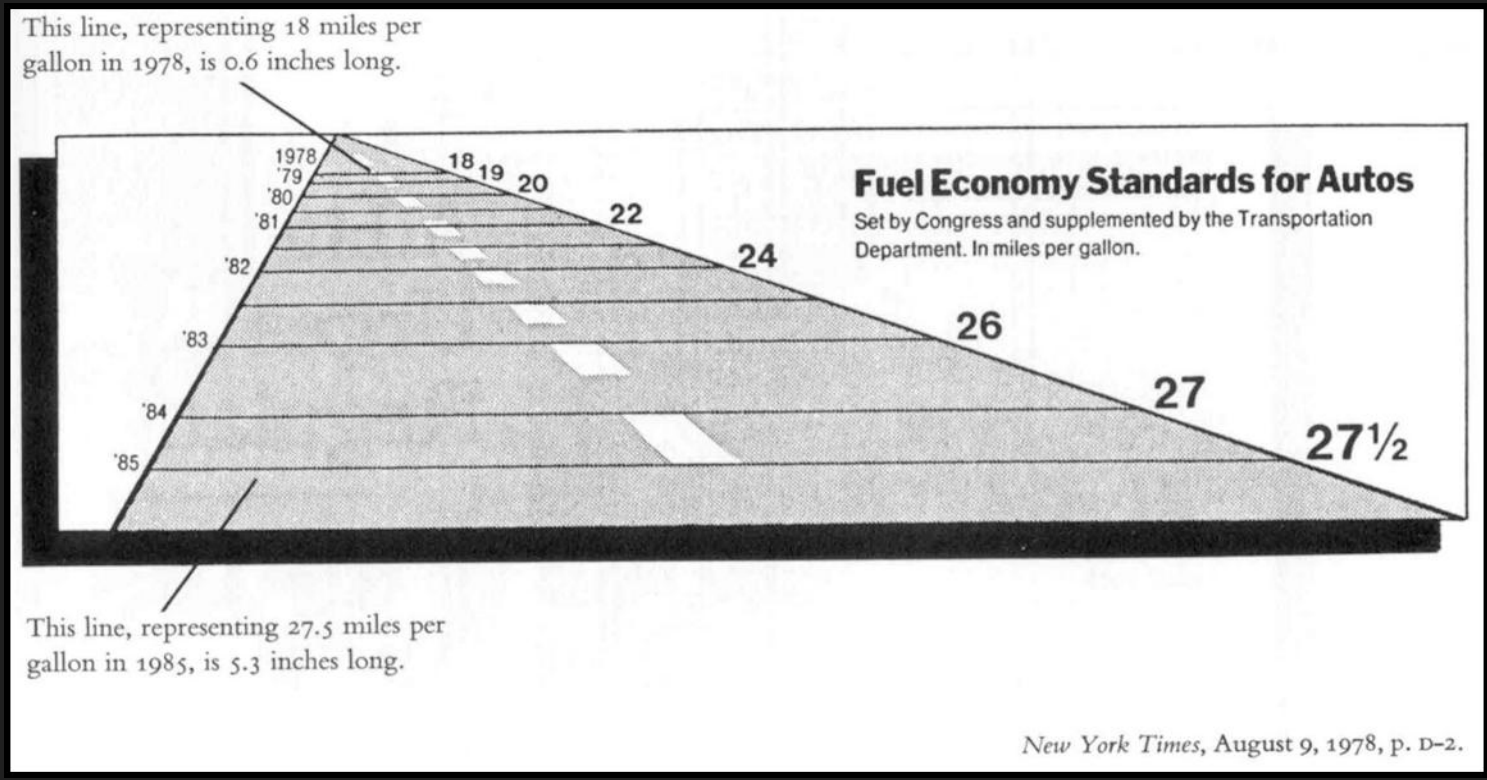
- How should you represent data to take into account personal differences?
- Use a table of numbers to show the data!!
  - Applicable for small datasets of 20 numbers of less
- Edward Tufte's first principle
  - Representation of numbers, as physically measured on the surface of the graphic itself, must be directly proportional to the numerical quantities represented

Lie Factor

- Violation of the principle causes misrepresentation (willful/unintentional)
- Lie Factor = $\frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$

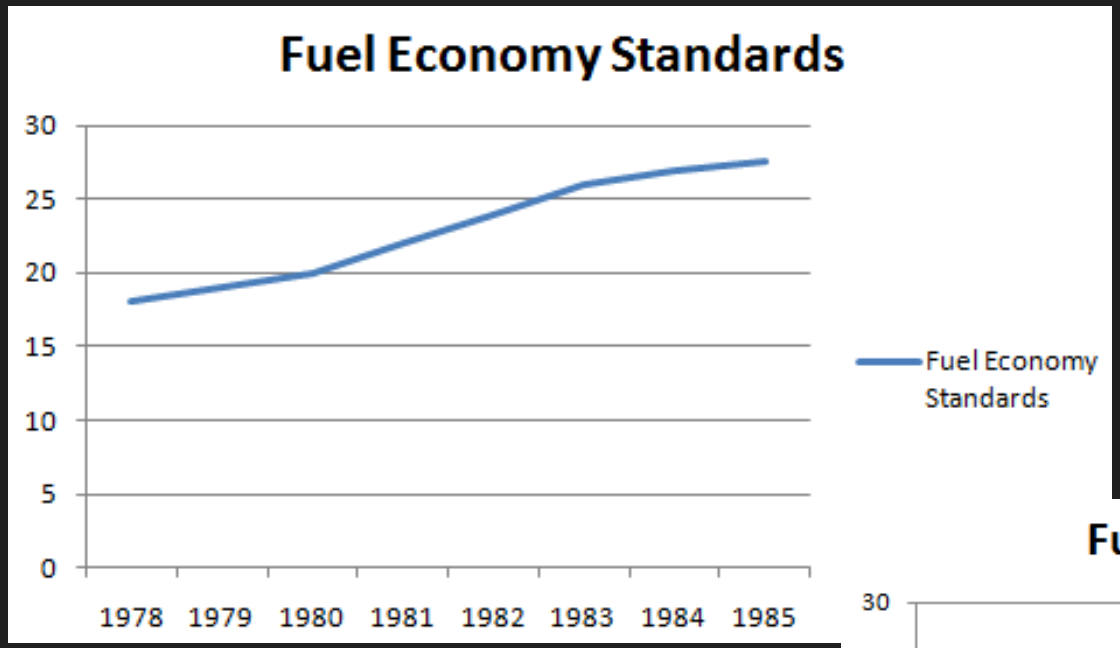
- Lie Factor = 1 implies that the graphic represents the data reasonably accurately
- Lie Factors > 1.05 or <0.95 indicates substantial inaccuracies

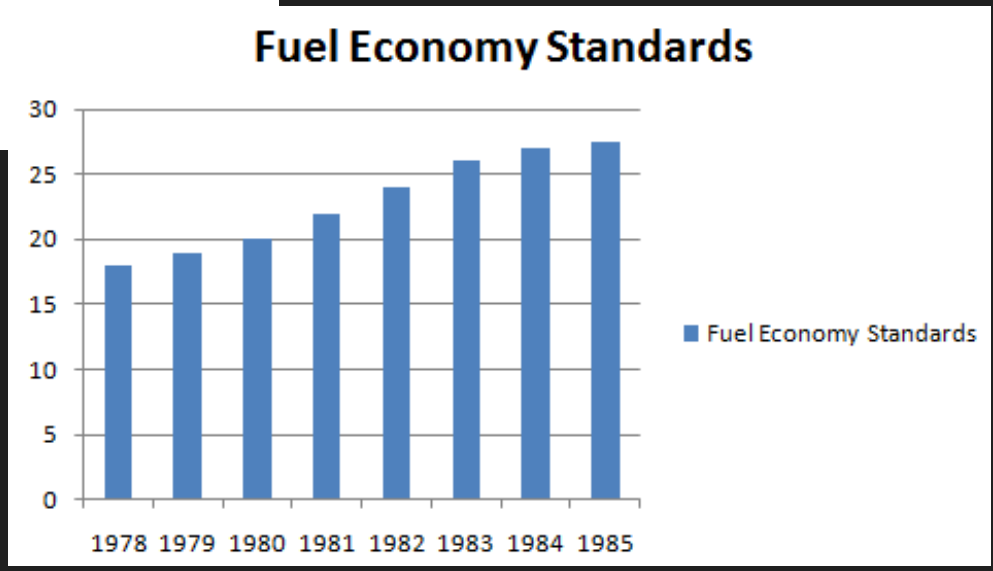
Lie Factor = 14.8%



- Change of 53% is shown by using a perspective view
- Easily represented in the form of a bar chart/line chart

Fuel Economy Standards





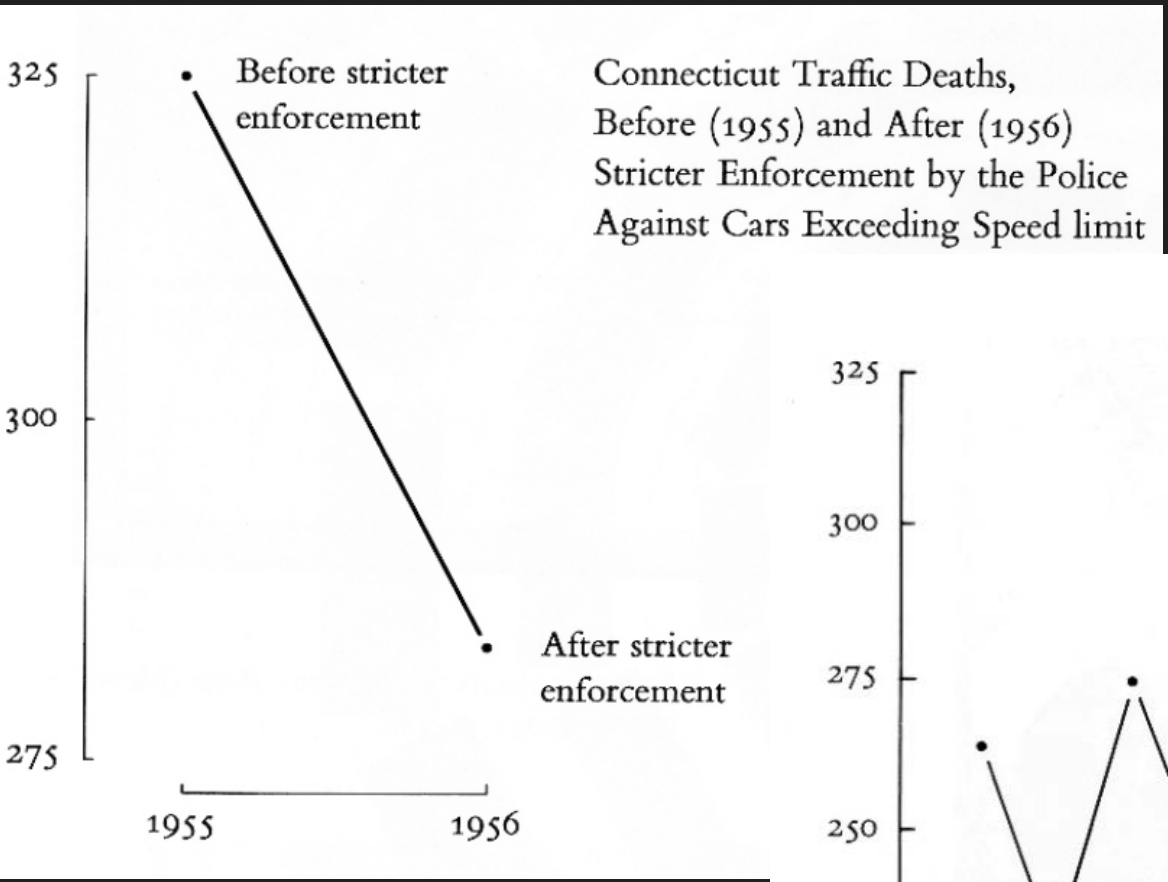
Data vs. Design Variation

- Show Data Variation, not Design Variation

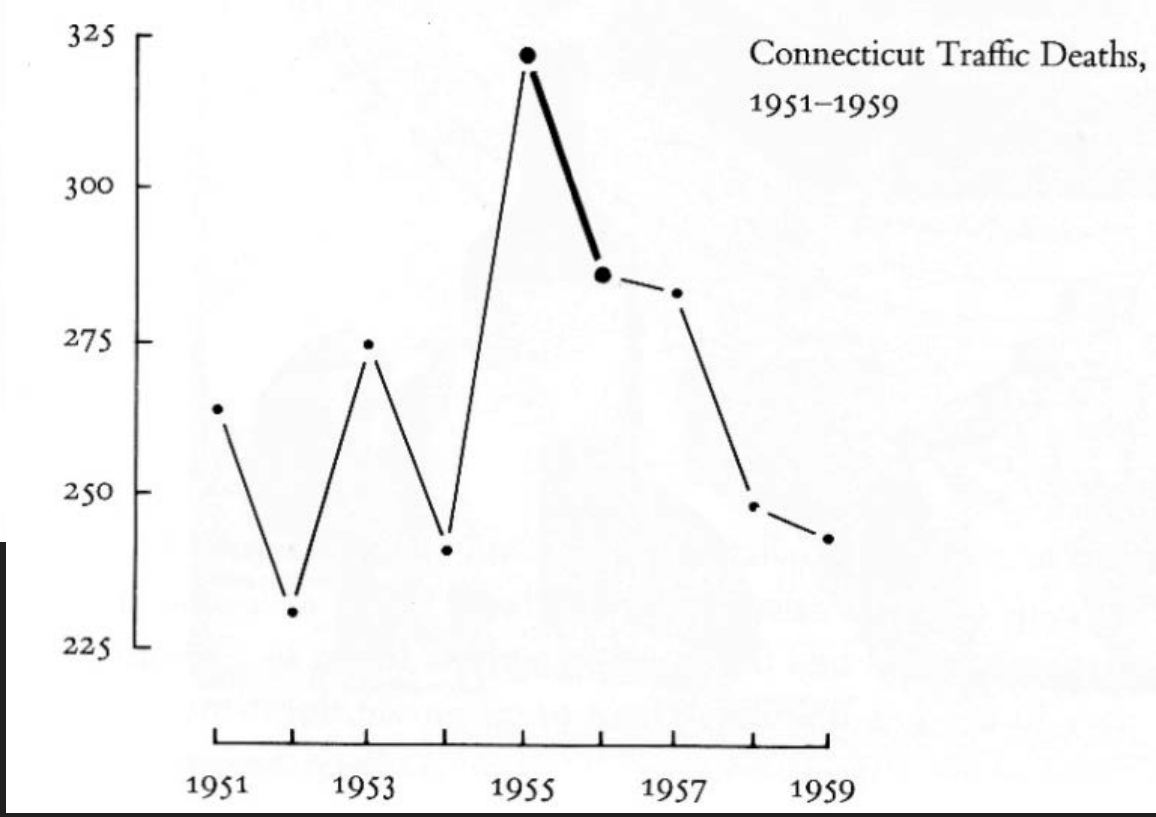
- Lie Factor = 15.1 for the *OPEC Oil Prices*
- Lie Factor = 9.4 for *In the Barrel*
- Lie Factor = 9.5 for the *OPEC Benchmark Prices*

- Stay consistent with your design through the data representation.
- Let the data variation present itself

Provide Context



Without Context

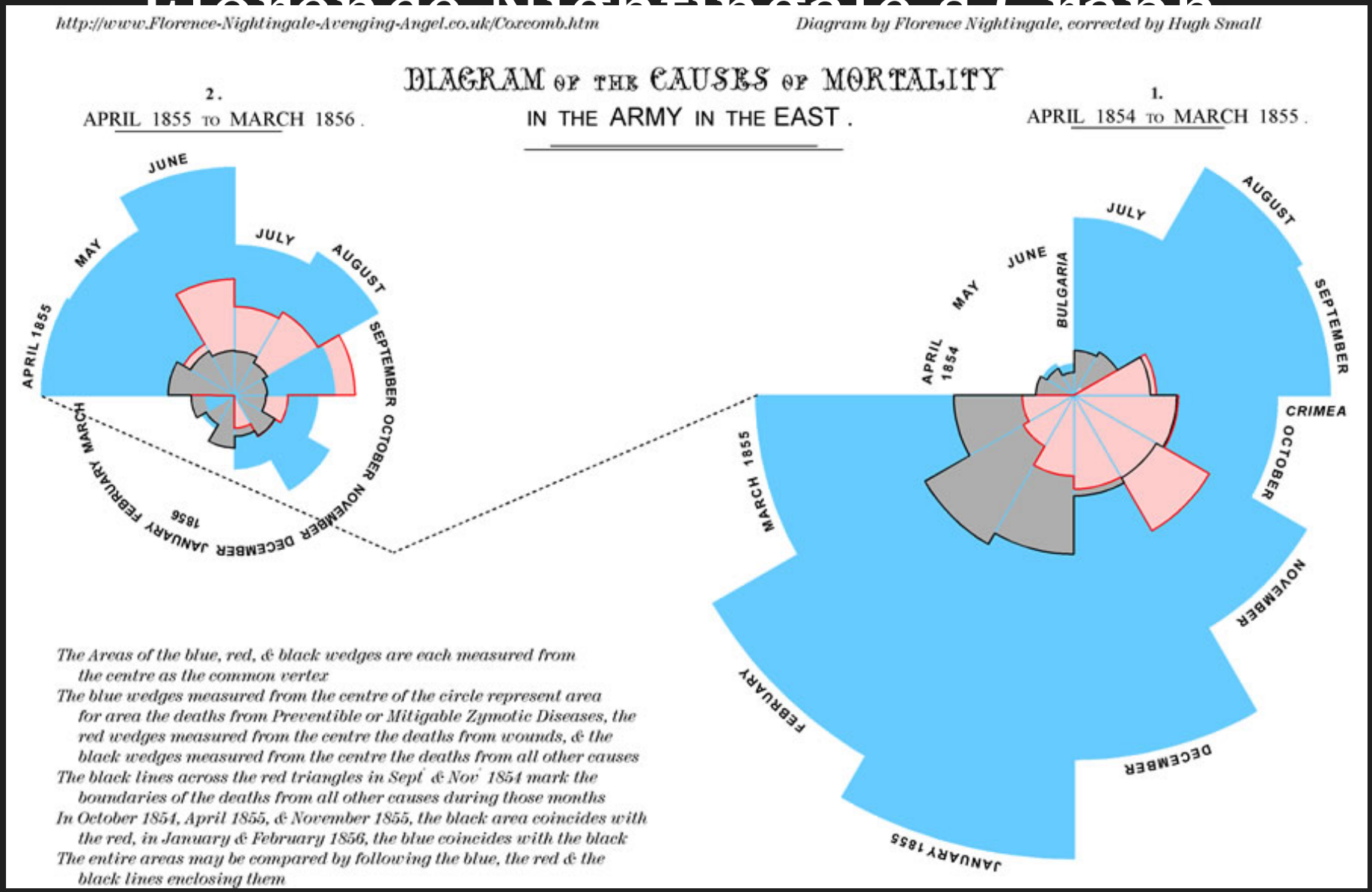


With Context

Context is Essential

- Context plays a crucial part in maintaining graphical integrity
- Provide sufficient context to the viewer
- Present data and make observations in relation to the context

Florence Nightingale's Graph



http://www.sciencenews.org/view/generic/id/38937/title/Math_Trek__Florence_Nightingale_The_passionate_statistician

Tufte's Principles for Graphical Integrity

1. The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
2. Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity.
3. Write out explanations of the data on the graphic itself. Label important events in the data.

Tufte's Principles for Graphical Integrity

4. In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
5. Show data variation not design variation
6. The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data

Design Principles

- Identify 10 design principles for creating effective data visualizations

For next class

- Look through some of the popular media (websites/newspapers etc.) and post an image of a **good** and a **bad** visualization on your blog before next class
  - State why the visualizations are good/bad
- Watch Hans Rosling's TED talk
- Read assigned papers and post your reaction on your blog before next class