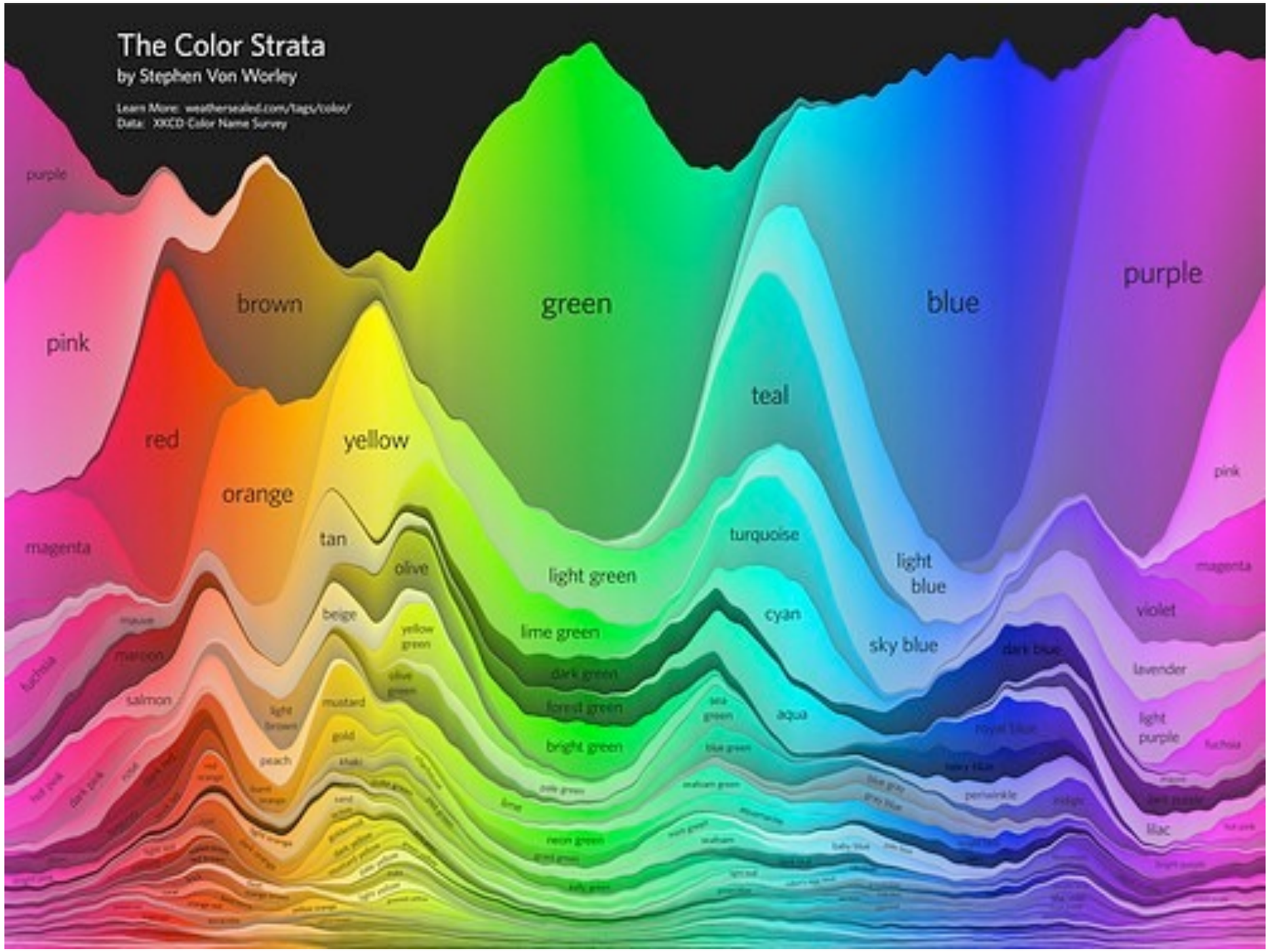
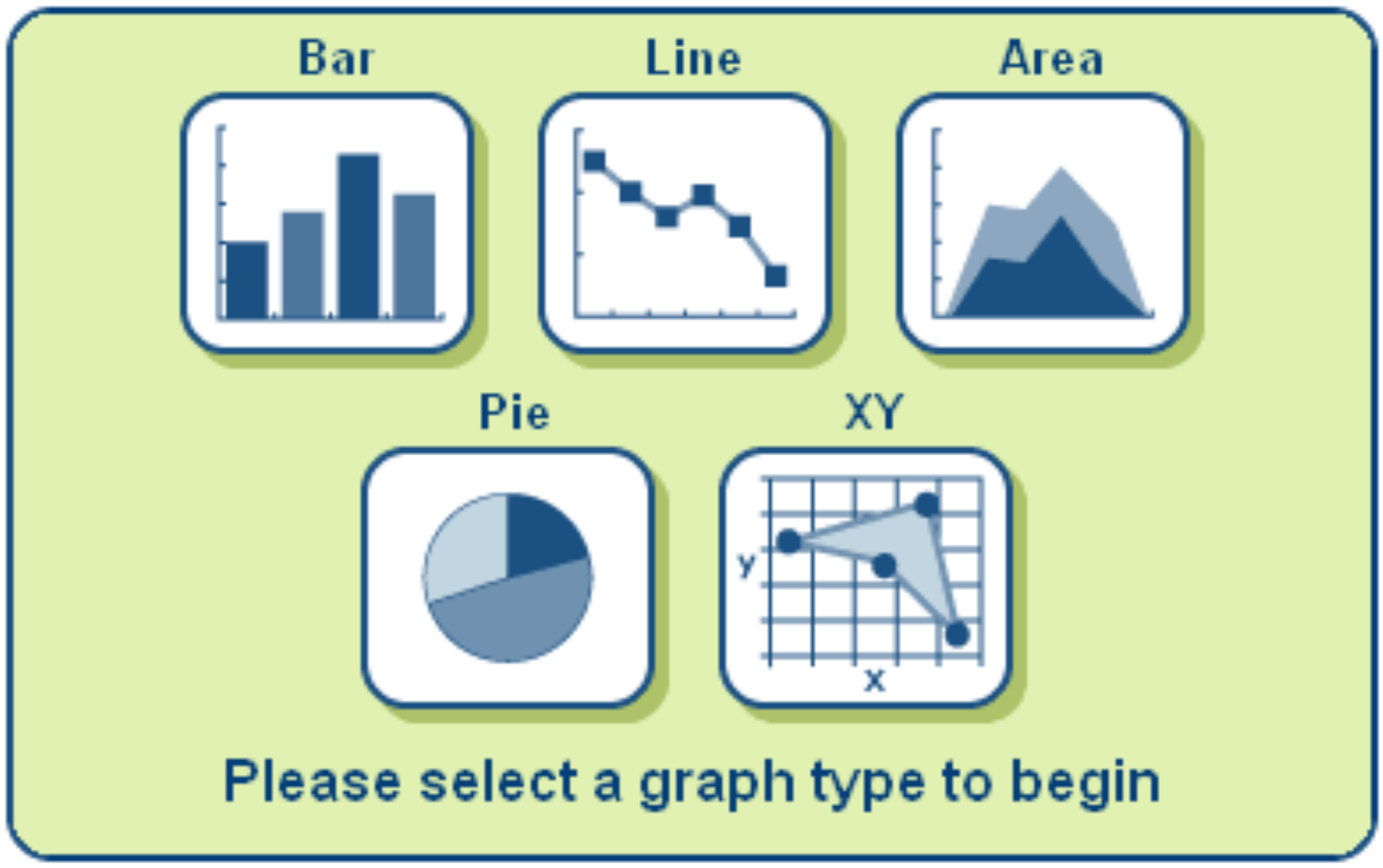
Visualising Data



Microsoft Office



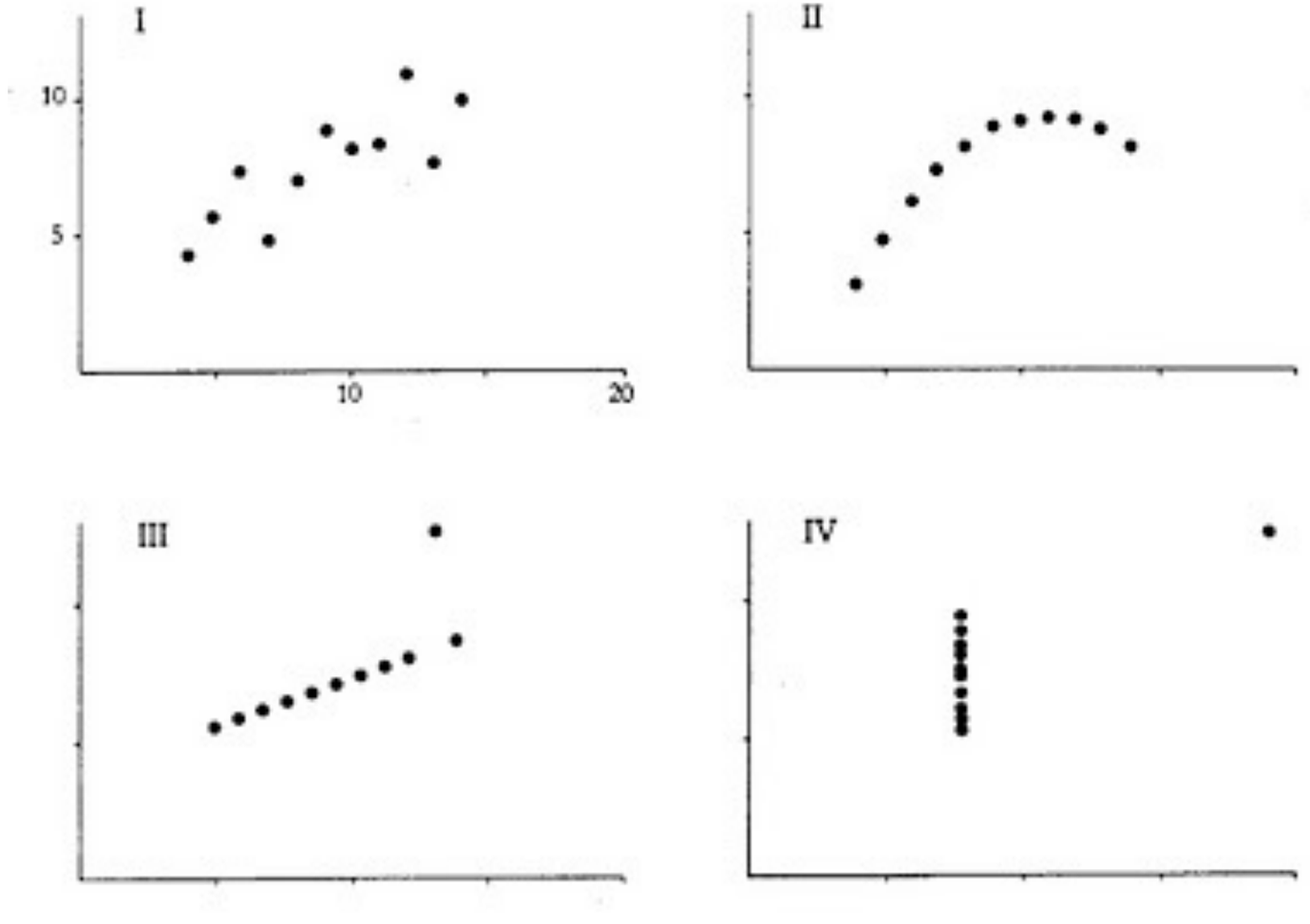
Raw data

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

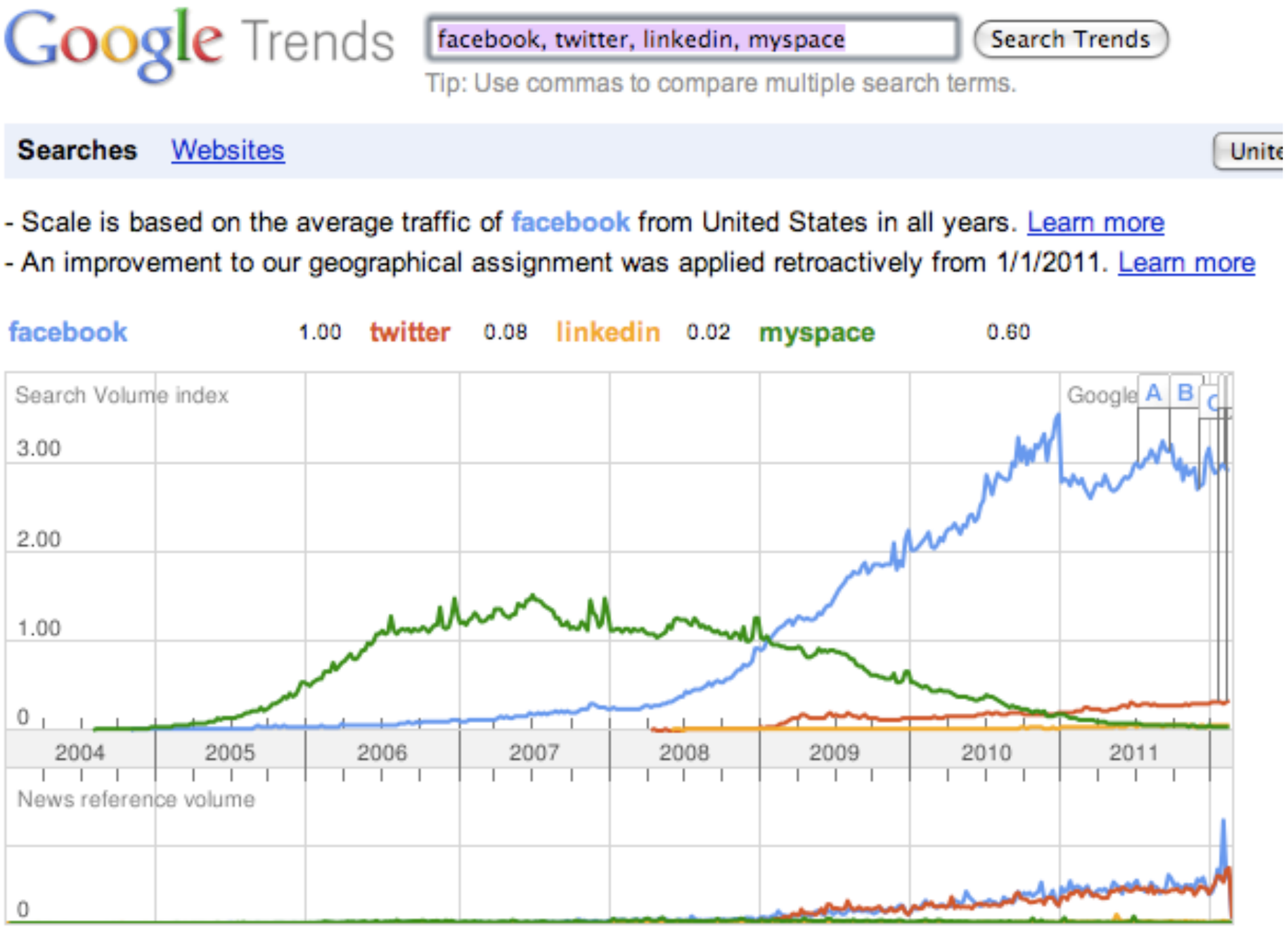
→

N = 11
mean of X's = 9.0
mean of Y's = 7.5
equation of regression line: Y = 3 + 0.5X
standard error of estimate of slope = 0.118
t = 4.24
sum of squares $X - \overline{X} = 110.0$
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
$r^2 = .67$

Visualisation

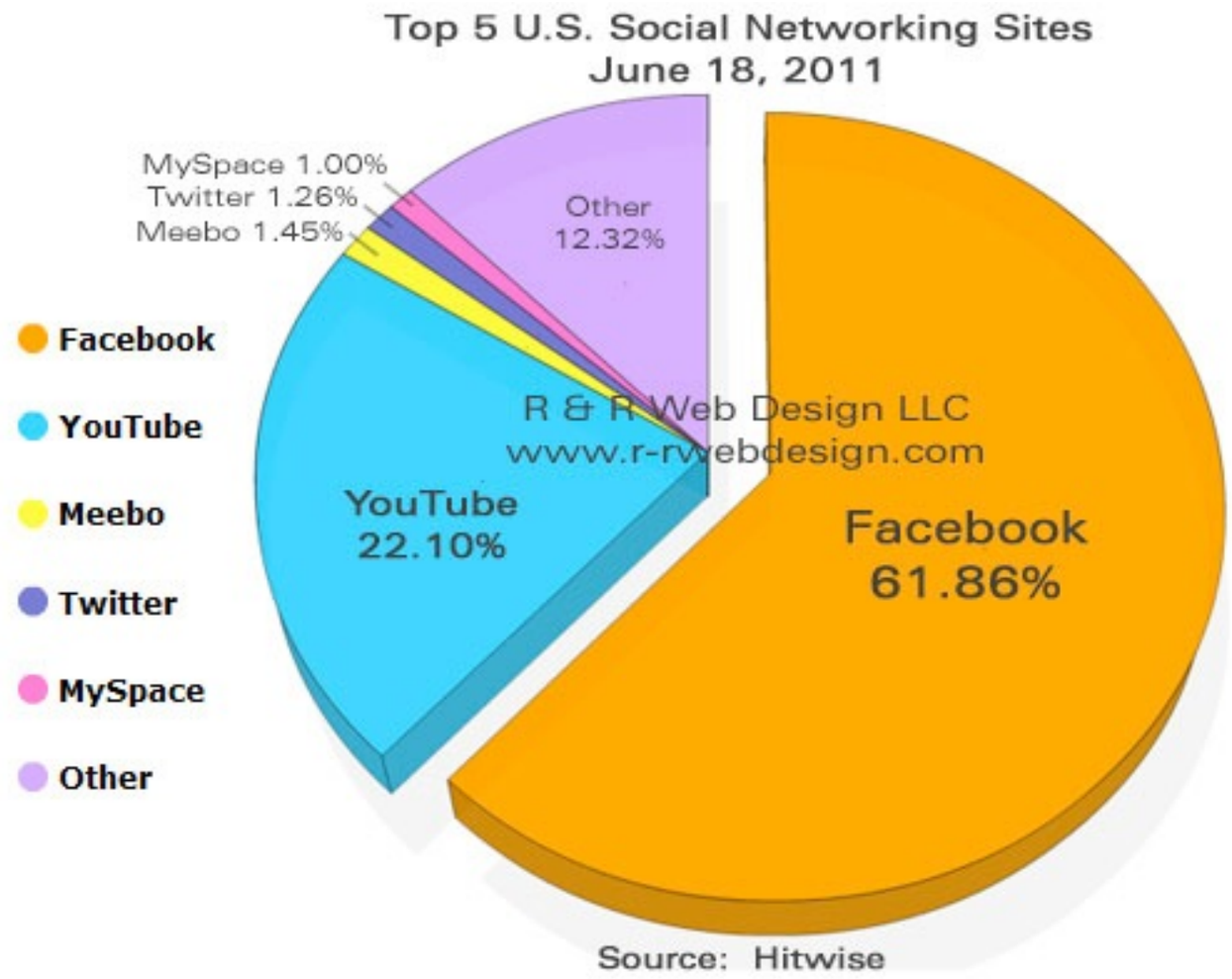


Graph of continuous numerical data (x-y)

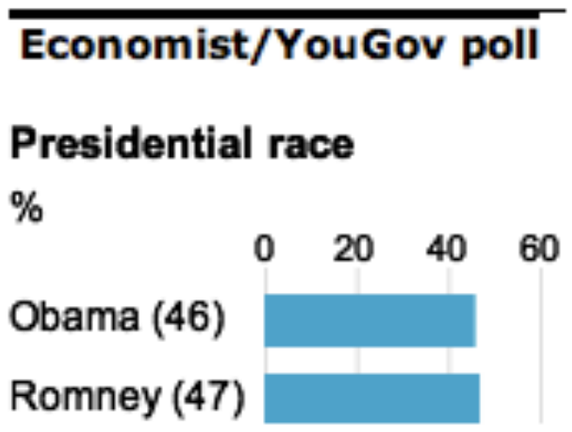


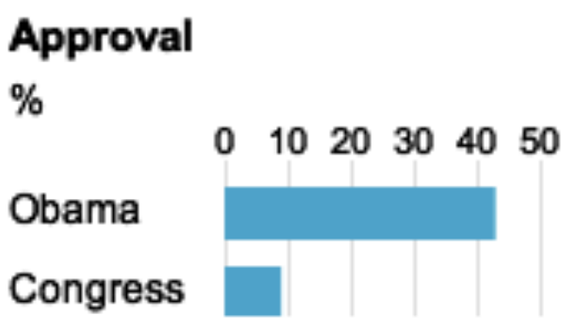
https://www.google.com.au/trends/

Pie Charts

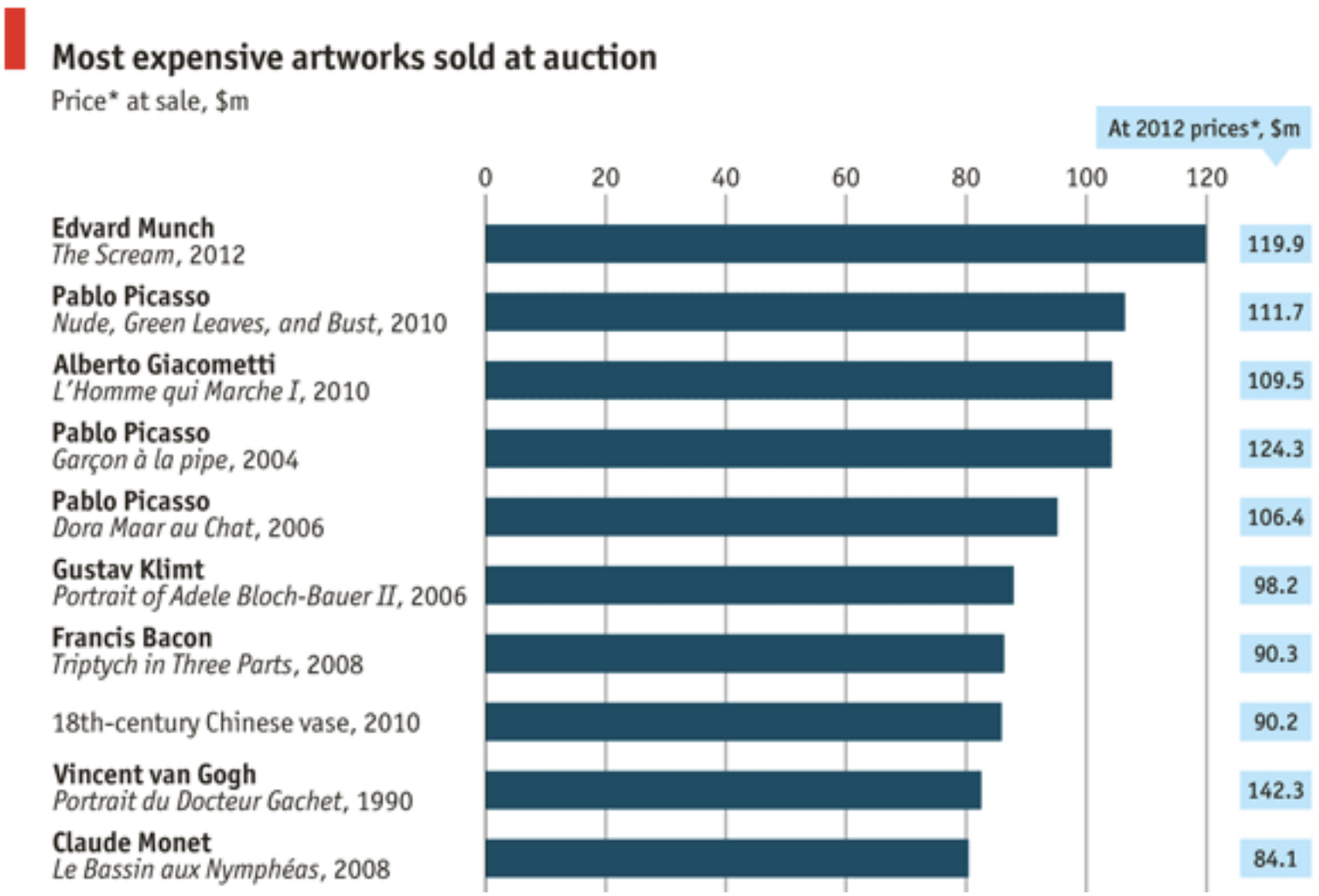

Top 5 U.S. Social Networking Sites
June 18, 2011
MySpace 1.00%
Twitter 1.26%
Meebo 1.45%
Other
12.32%
Facebook
YouTube
Meebo
Twitter
MySpace
Other
R & R Web Design LLC
www.r-rwebdesign.com
YouTube
22.10%
Facebook
61.86%
Source: Hitwise

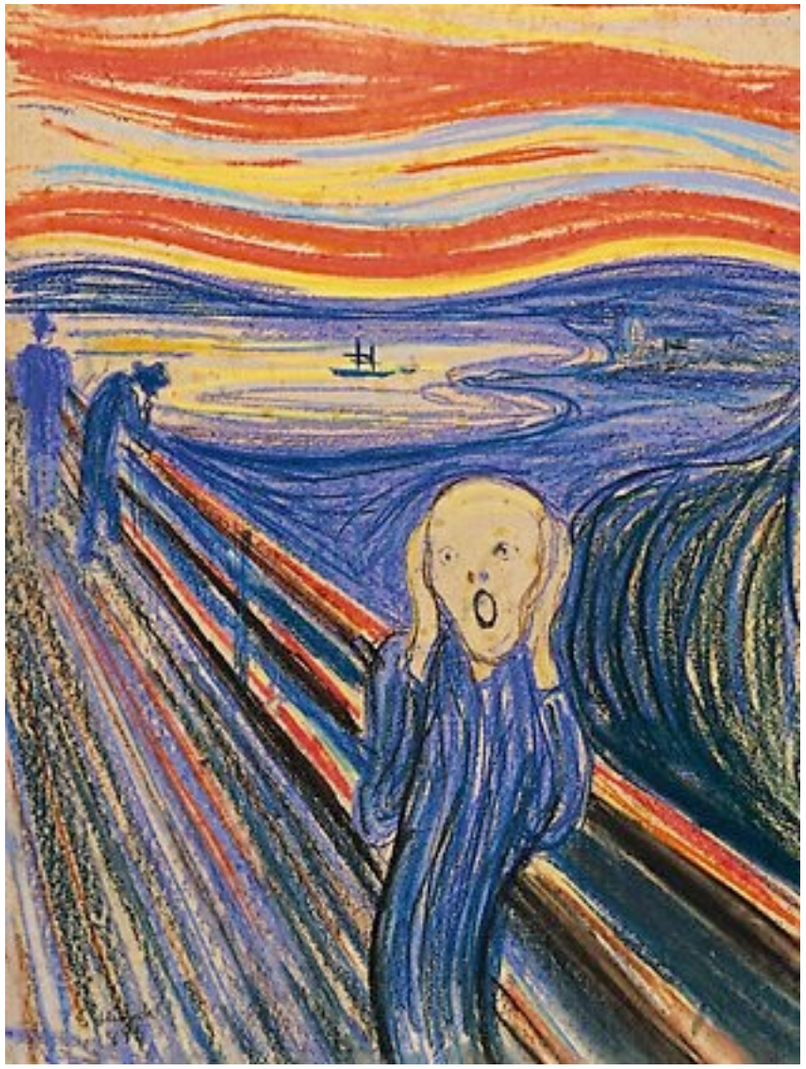
Simple bar graphs





Bar Graphs are useful for seeing comparisons

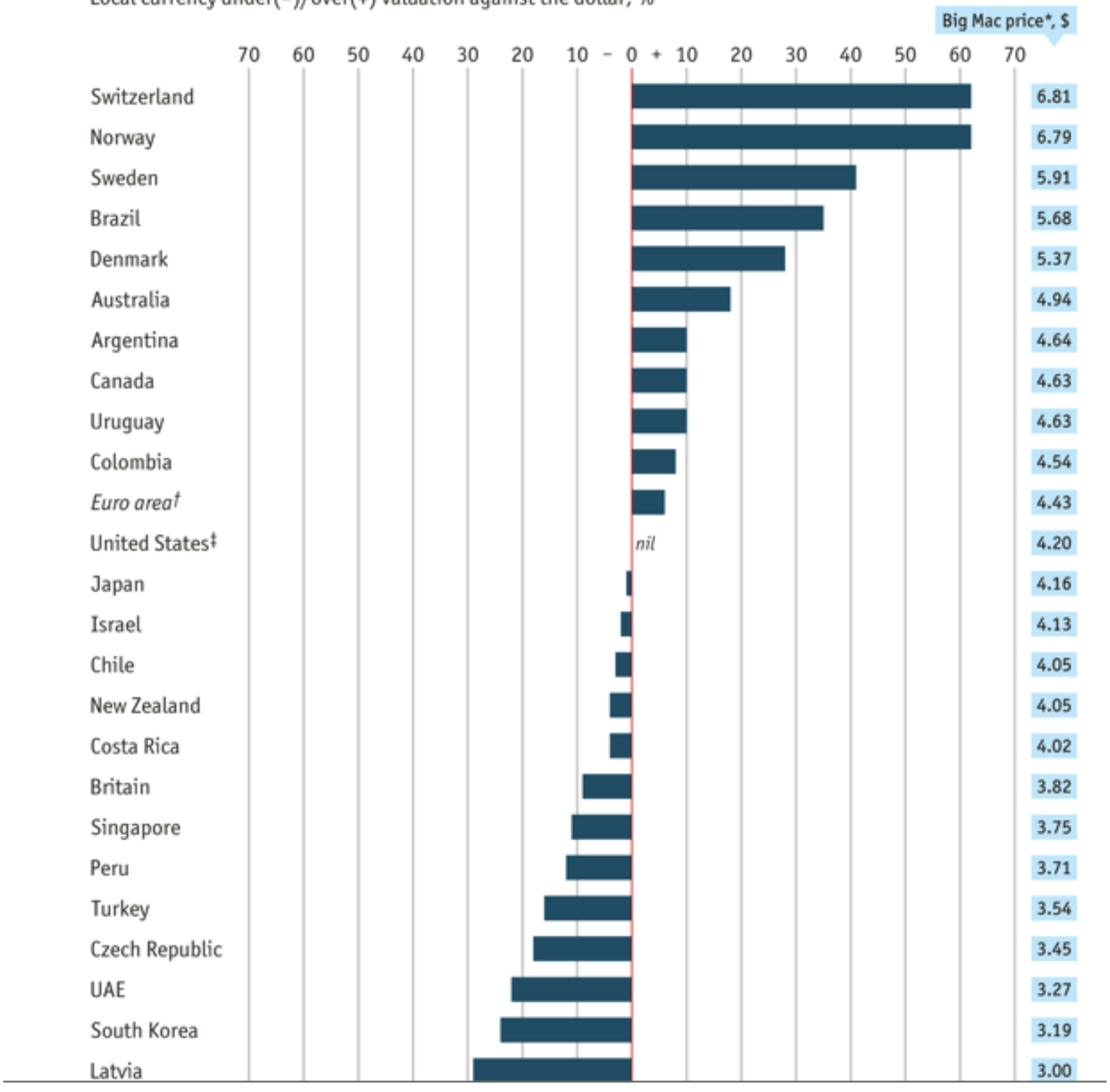




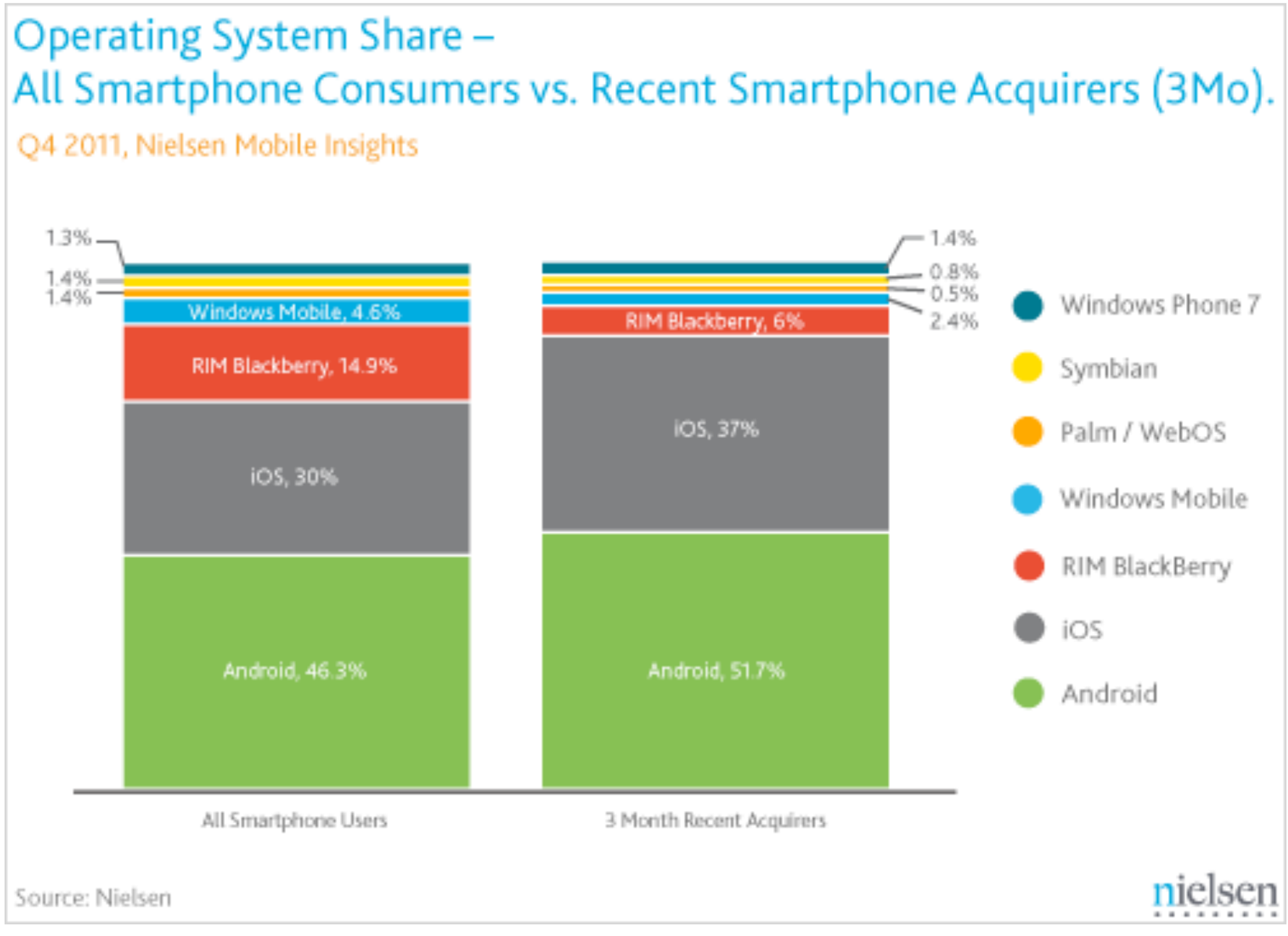
If data is presented vertically, sometimes called a "column graph"

The Big Mac index

Local currency under(−)/over(+) valuation against the dollar, %

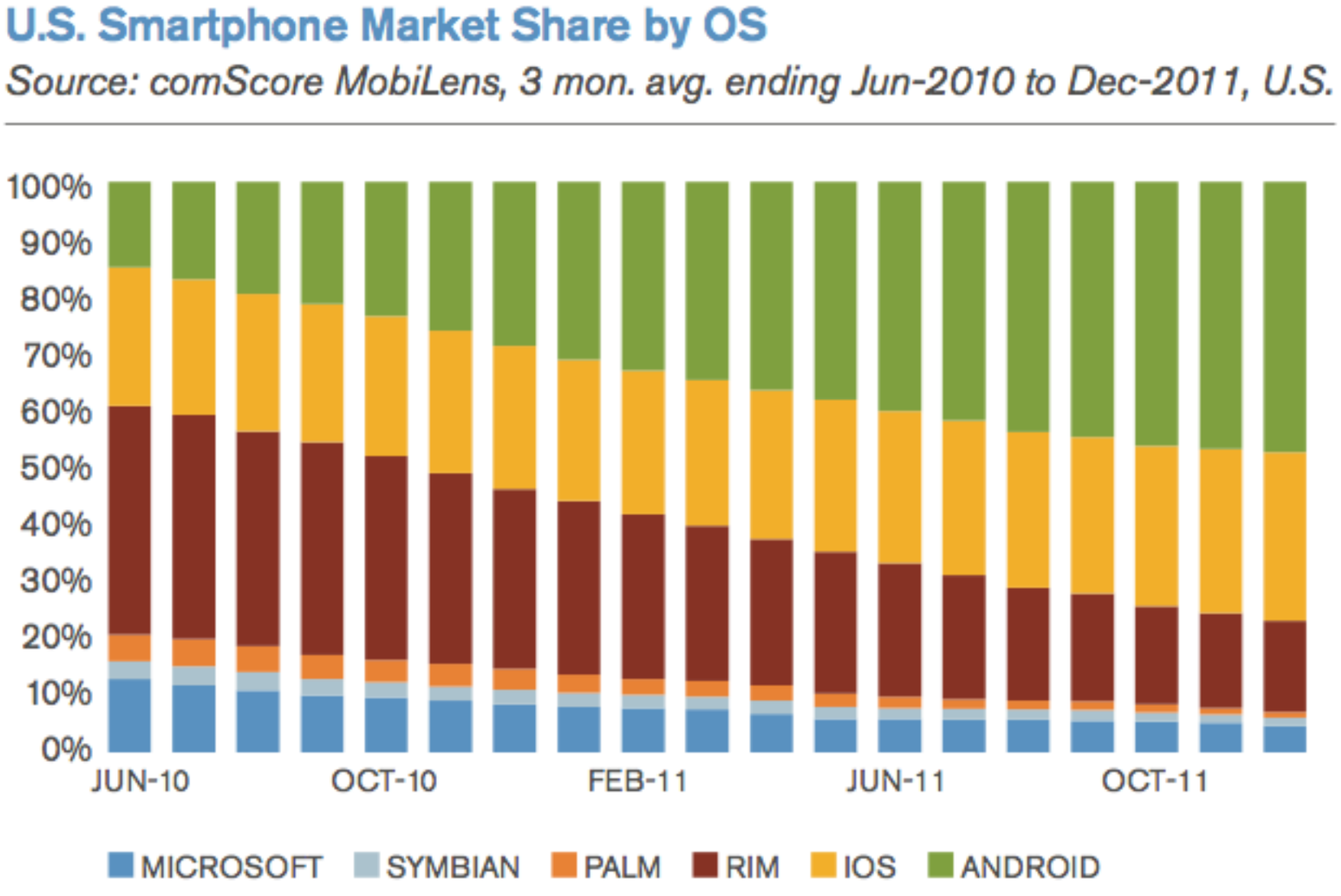


Possible to "stack" data

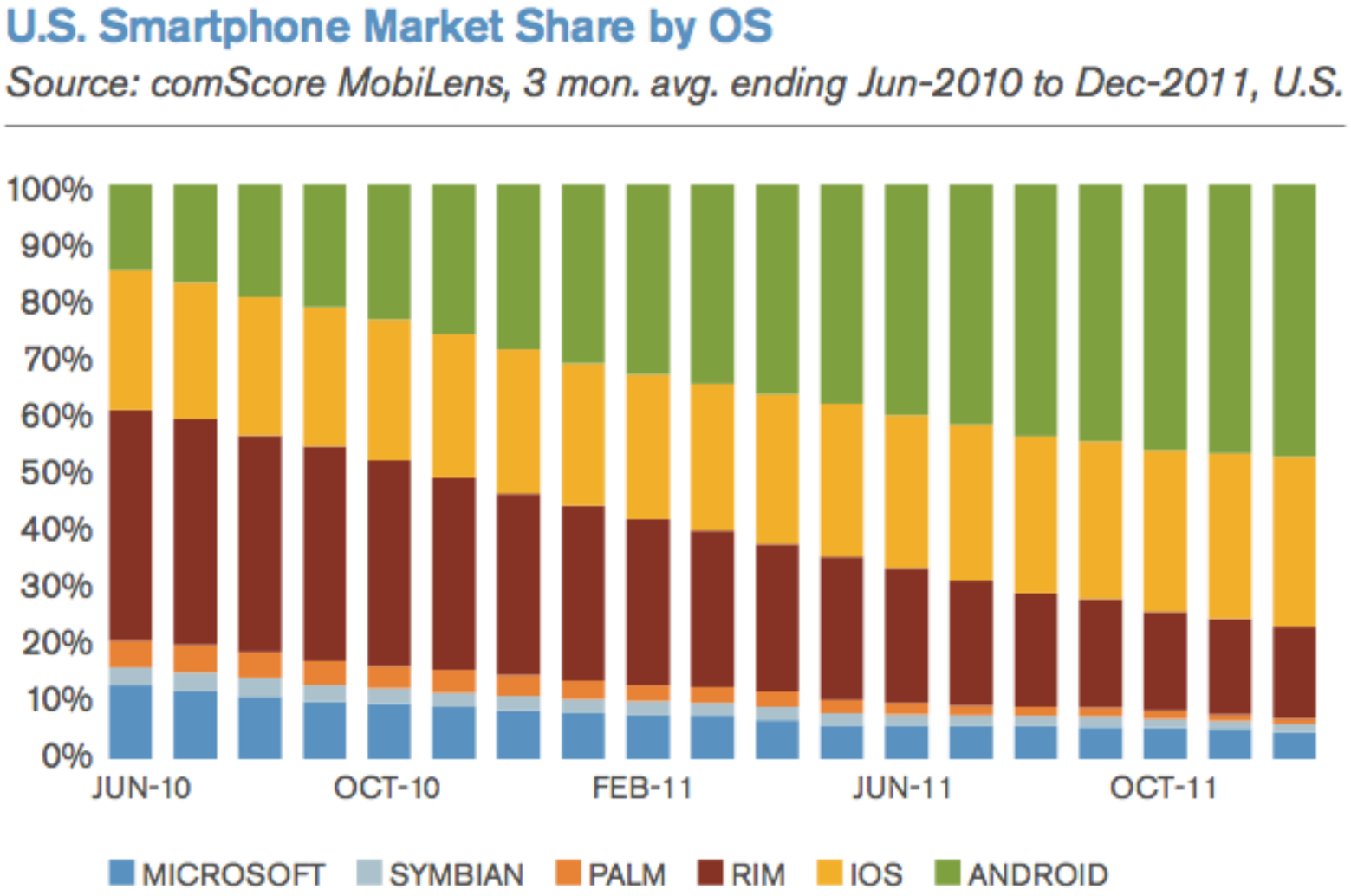


What kind of graph is this?

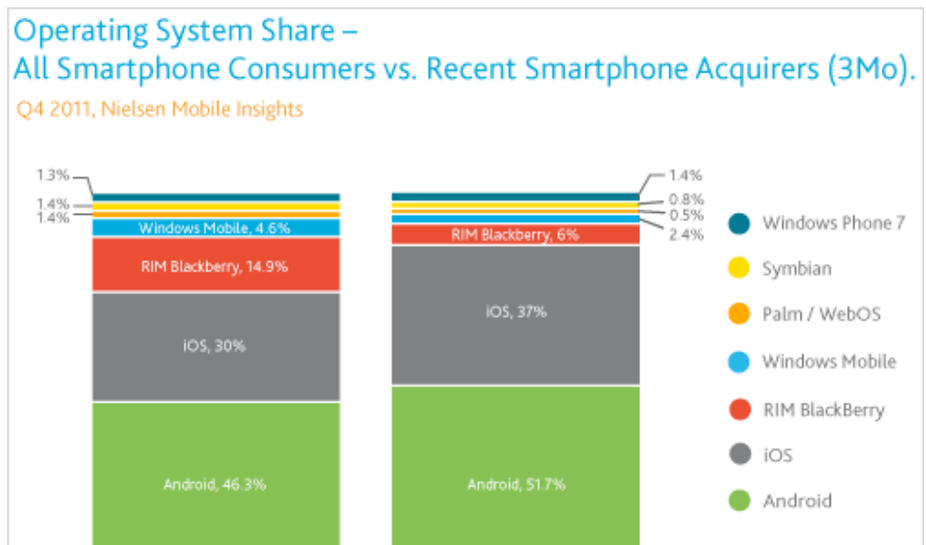
Another data representation of the previous theme.



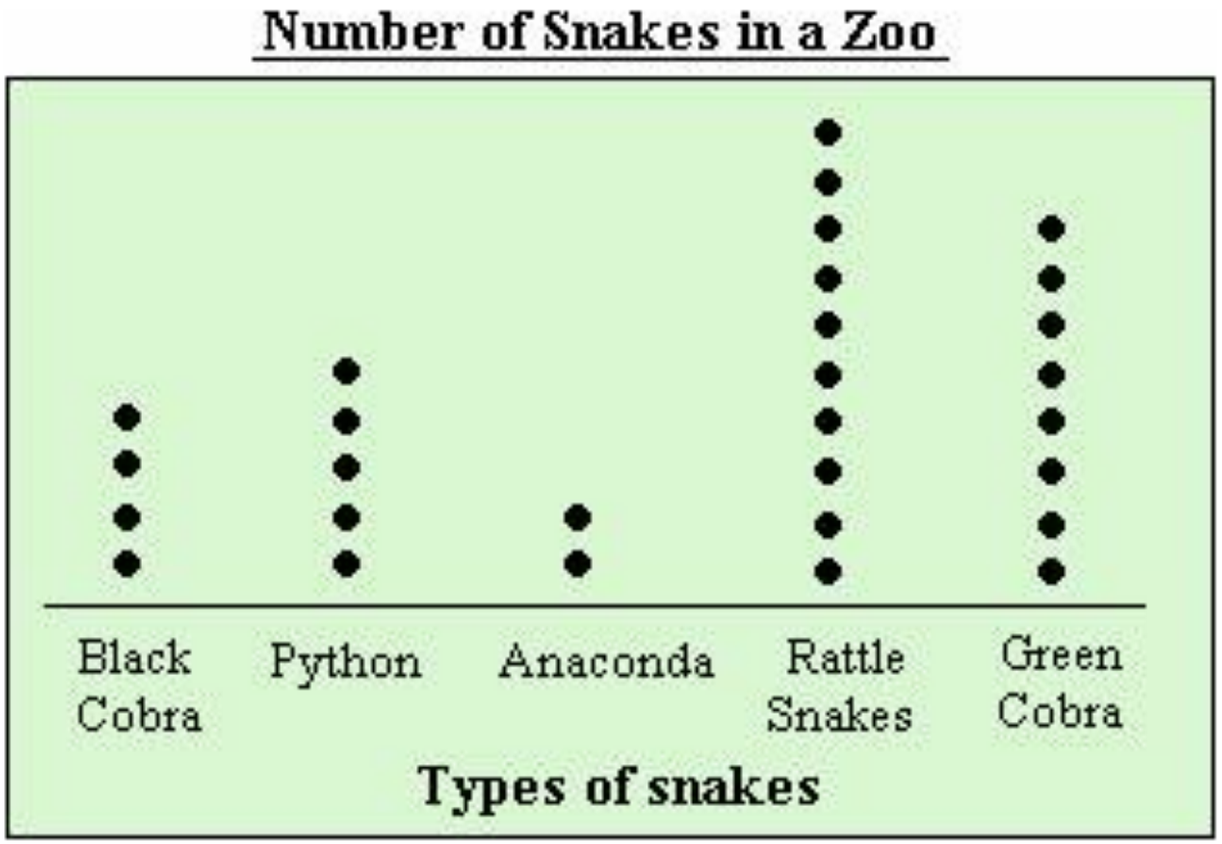
Another data representation of the previous theme.



How does this differ from the previous slide?

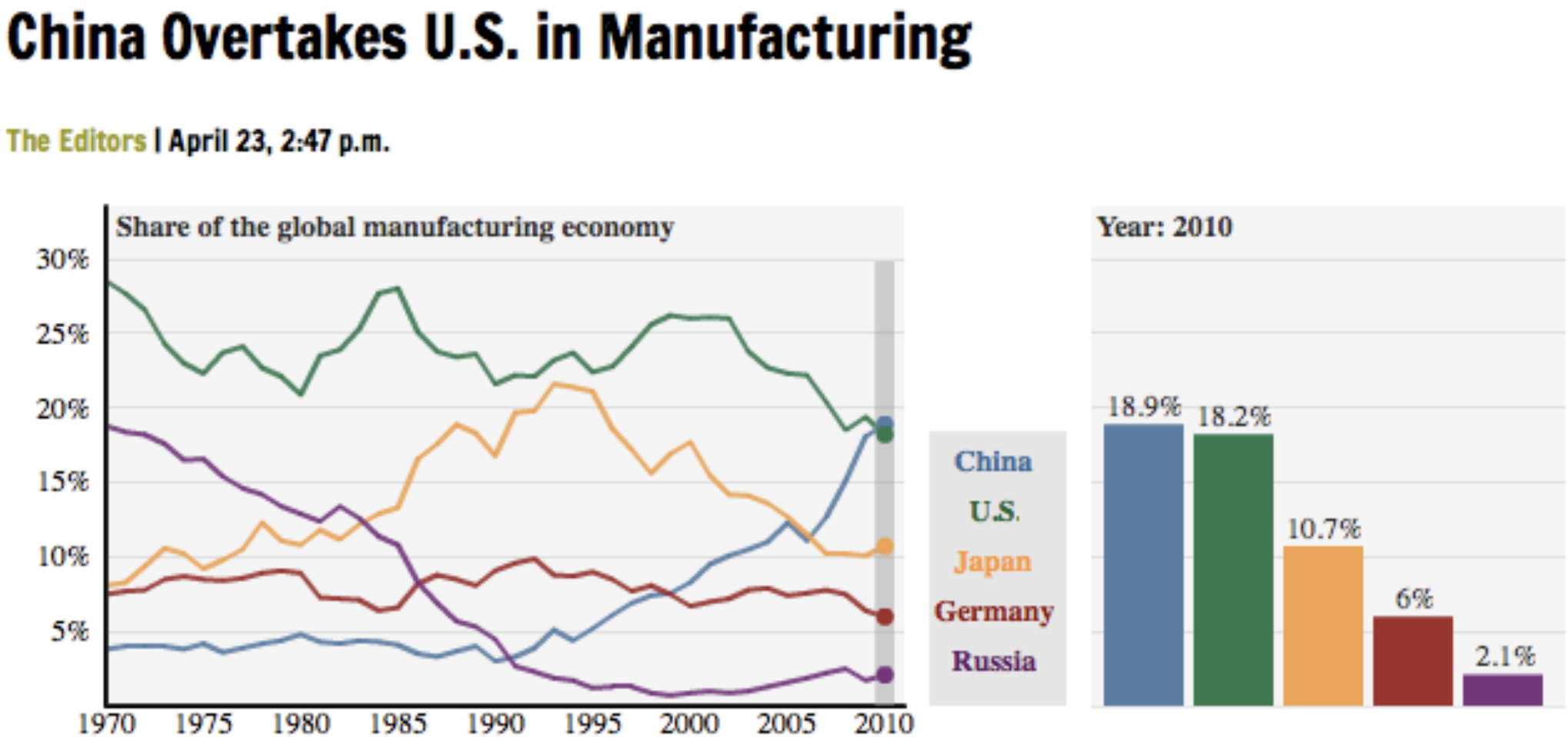


Dot Plots --similar to column/bar graphs and can be presented vertically or horizontally



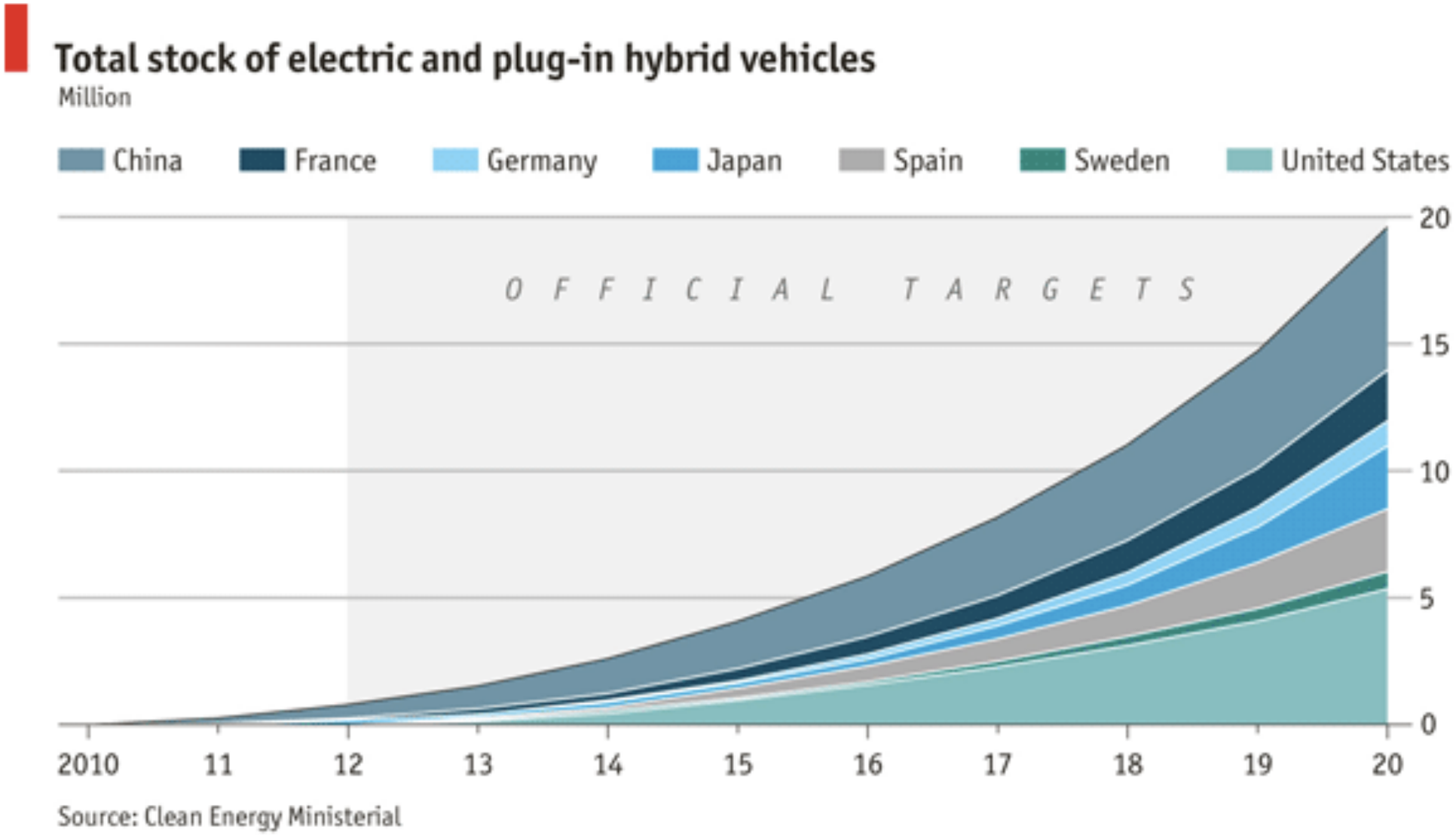
Die Roll	Frequency (50 total)
1	••••••••
2	••••••
3	••••••
4	•••••••••
5	•••••••
6	••••••••••••••

Line Graph (x-y) and column graph



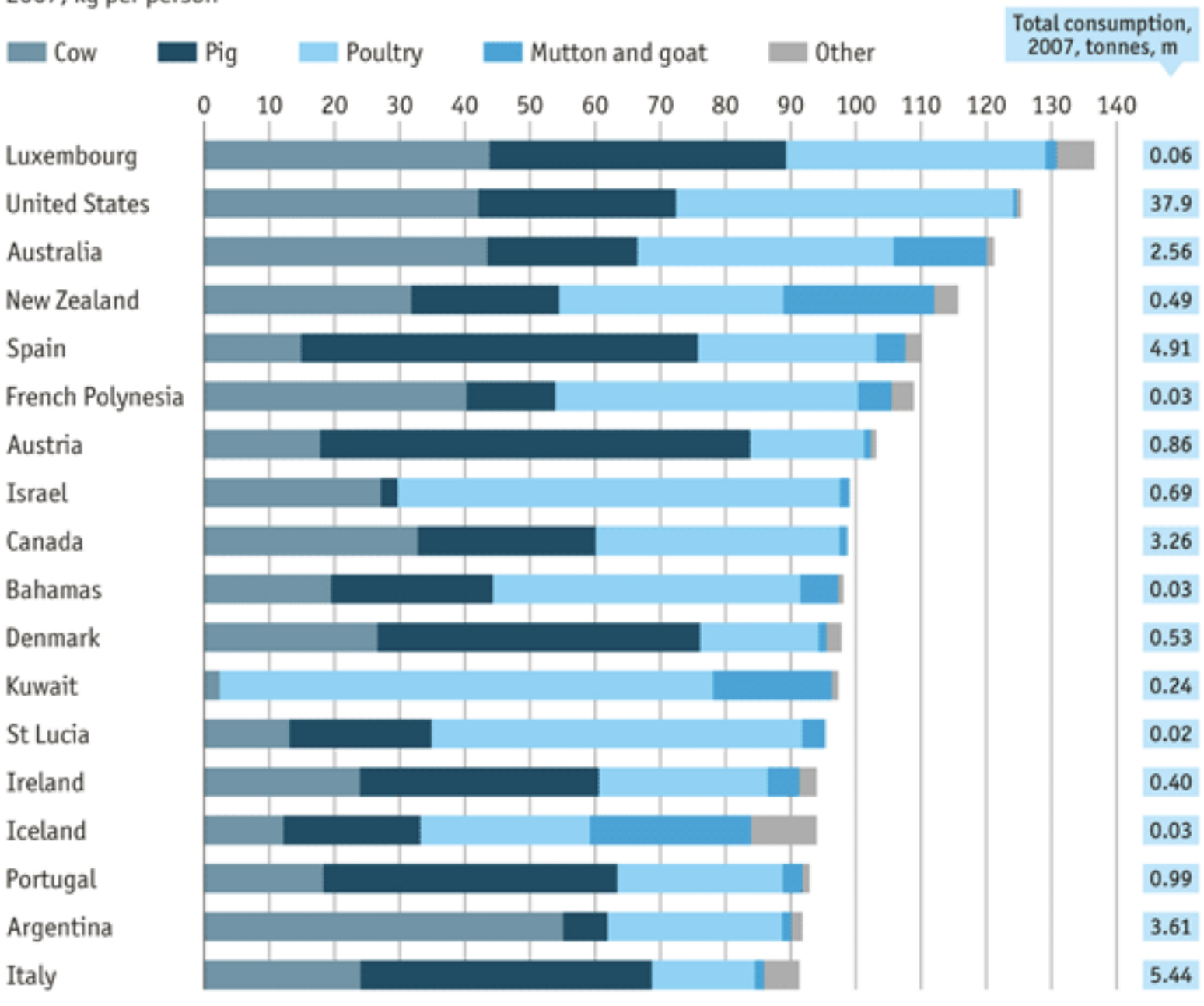
How do these relate?

More complex graphs showing multiple data



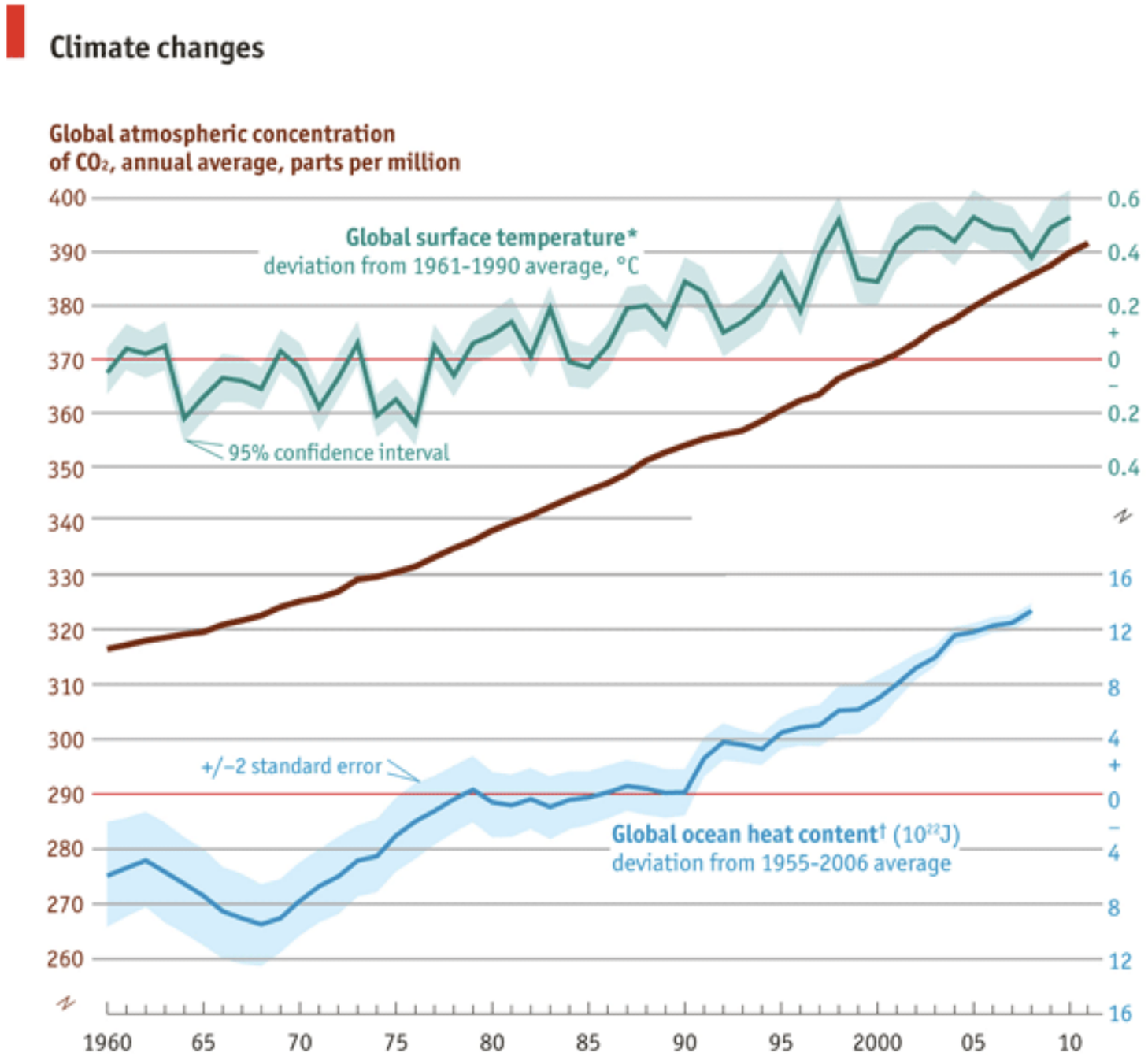
World's biggest meat-eaters

2007, kg per person



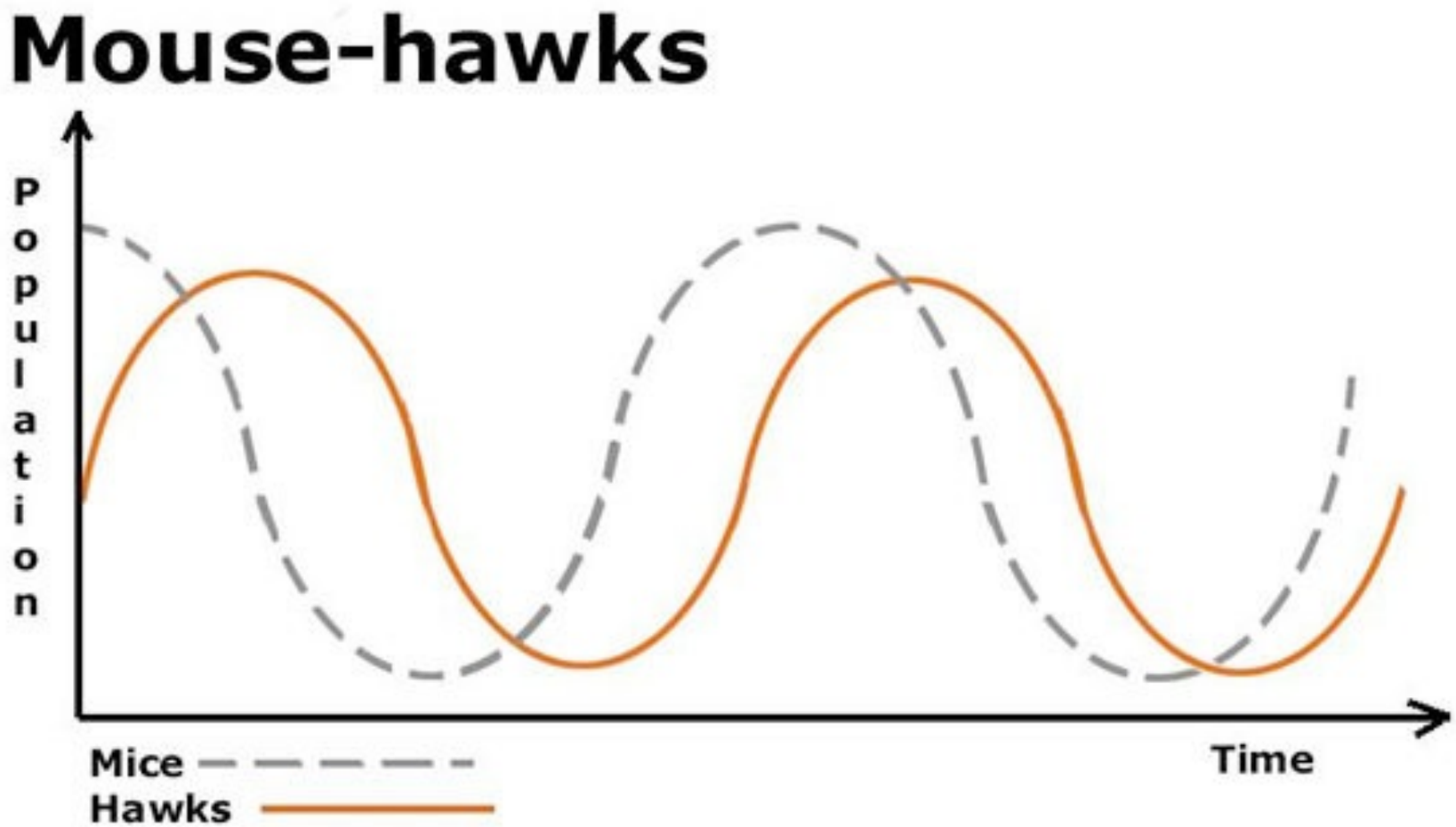
Sources: UN Food and Agriculture Organisation; *The Economist*

Graphs with multiple y axes (common x axis)



Sources: Met Office Hadley Centre; NOAA; Scripps Institute of Oceanography; Sydney Levitus *et al*, GRL

Population graphing--predator-prey relationship



As the mice population increases, there are more easily-caught mice as prey for the predatory hawks. The hawk population will increase too, but slightly later than the increase in mice population. (Note: one hawk eats many mice, so the total numbers of hawks is always lower.)

More hawks eat more mice, so mice numbers drop. Mice become harder to catch, so hawk numbers drop too. With fewer predators, the mice can breed up again. The number of prey goes up and down, mirrored by the later rise and fall of predators. The time delay is related to the life cycle of the organisms.

■ STATE *of* THE NATION *the* UNION ■ IN NUMBERS

How the Recession Changed Us

WHAT A DIFFERENCE TWO YEARS MAKES.

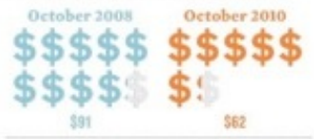
By Timothy Lavin
Graphics by Amanda Buck

OFFICIALLY, THE GREAT Recession lasted from December 2007 to June 2009. A mere 18 months—about average, as recessions go. Yet if the trauma this time feels deep and lasting, that may be because, as the figures on these pages show, so many disruptions have upended national life at once.

Millions of Americans have lost their jobs, nearly every state faces a budget shortfall, and hundreds of banks have shut their doors. The young are unemployed, living at home, and playing video games. The ranks of third-party candidates have swollen, militias have proliferated, and national leaders of both parties have seen their support decline. Of course, times of flux are often times of anxiety and unrest. But as the economy begins its slow and stuttering recovery, the vast changes wrought by this recession will continue to reverberate for many years—in ways predictable and otherwise. ◢

Timothy Lavin is an Atlantic senior editor.

DAILY CONSUMER SPENDING



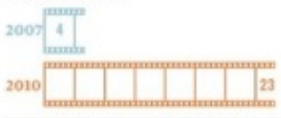
U.S. BOOK SALES

PERCENTAGE INCREASE *in NYT* ARTICLES *mentioning the* WORD *"uncertainty," 2007 vs. 2010**

TAX COLLECTED *by the* TREASURY *on sales of* FIREARMS *and* AMMUNITION

PERCENTAGE *who believe it's more important to preserve* GUN OWNERS' RIGHTS *than to control* GUNS

FEATURE-LENGTH MOVIES *released in* 3-D



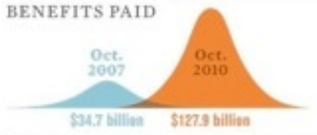
AVERAGE MINUTES SPENT READING *per* WEEKEND DAY, *15-to-19-year-olds*

AVERAGE MINUTES SPENT PLAYING A GAME *or* USING A COMPUTER *for leisure per* WEEKEND DAY, *15-to-19-year-olds*

UNEMPLOYMENT RATE

PERCENTAGE *of* UNEMPLOYED WHO HAVE BEEN UNEMPLOYED *for* 27 WEEKS *or more*

GOVERNMENT UNEMPLOYMENT BENEFITS PAID



PERCENTAGE *of 18-to-29-year-olds* WHO MOVED BACK IN WITH THEIR PARENTS *during recession*

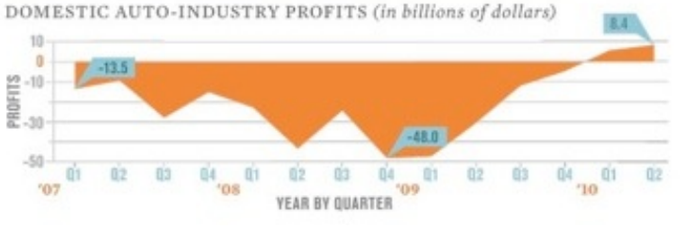
PERCENTAGE *of 18-to-29-year-olds* WHO THINK AMERICA *is still a* LAND *of* PROSPERITY

NUMBER *of* PEOPLE *in* POVERTY

FOOD-STAMP RECIPIENTS *by state*

	2007	2009	% increase
FLORIDA	1,232,803	1,952,362	58%
NEVADA	122,224	200,056	63%
NEW YORK	1,801,984	2,322,742	29%

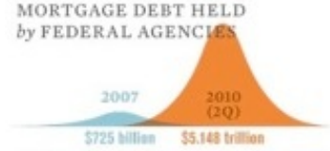
DOMESTIC AUTO-INDUSTRY PROFITS *(in billions of dollars)*



ALL STATISTICS ARE THE MOST RECENT AVAILABLE AT PRESS TIME, SEASONALLY ADJUSTED WHERE POSSIBLE.

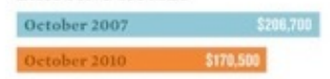
RESIDENTIAL PROPERTIES *in* FORECLOSURE

MORTGAGE DEBT HELD *by* FEDERAL AGENCIES



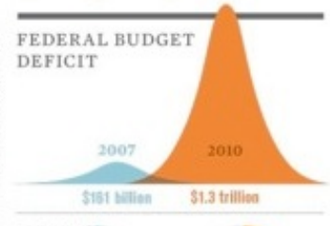
AVERAGE RATE *on* 30-YEAR FIXED MORTGAGE

MEDIAN SALE PRICE *of* EXISTING HOMES



STATES *facing* BUDGET SHORTFALLS

FEDERAL BUDGET DEFICIT



DEFICIT *as* PERCENTAGE *of* GDP

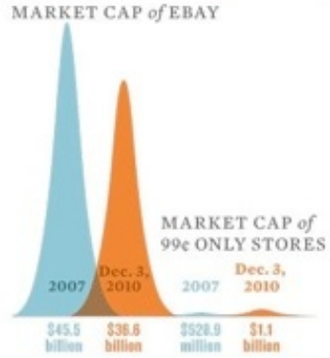
NUMBER *of* FEDERAL EMPLOYEES*

*Excludes certain agencies and offices not provided by the U.S. Office of Personnel Management

FOR A COMPLETE LIST OF SOURCES, VISIT WWW.THEATLANTIC.COM/RECESSIONMAP.

BANK FAILURES

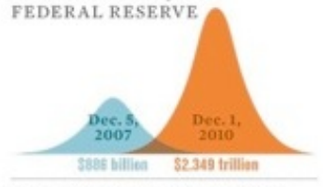
MARKET CAP *of* EBAY



MARKET CAP *of* 99¢ ONLY STORES

INTEREST RATE *on a 2-year* TREASURY NOTE

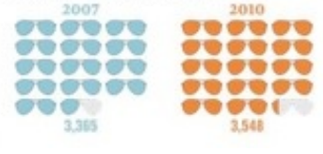
TOTAL ASSETS *of the* FEDERAL RESERVE



CURRENCY *in* CIRCULATION

ACTIVE MILITIAS

SECRET-SERVICE AGENTS

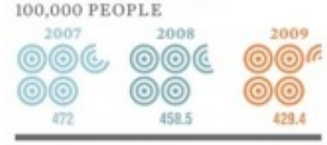


SAVINGS RATE, *as percentage of disposable income*

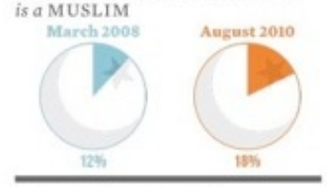
AVERAGE INTEREST PAID *on* DOMESTIC DEPOSITS

PERSONAL BANKRUPTCIES

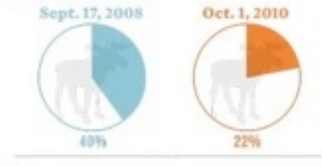
VIOLENT CRIMES *per* 100,000 PEOPLE



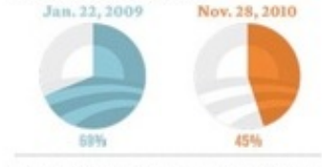
PERCENTAGE *of* AMERICANS *who think* PRESIDENT OBAMA *is a* MUSLIM



FAVORABLE VIEW *of* SARAH PALIN

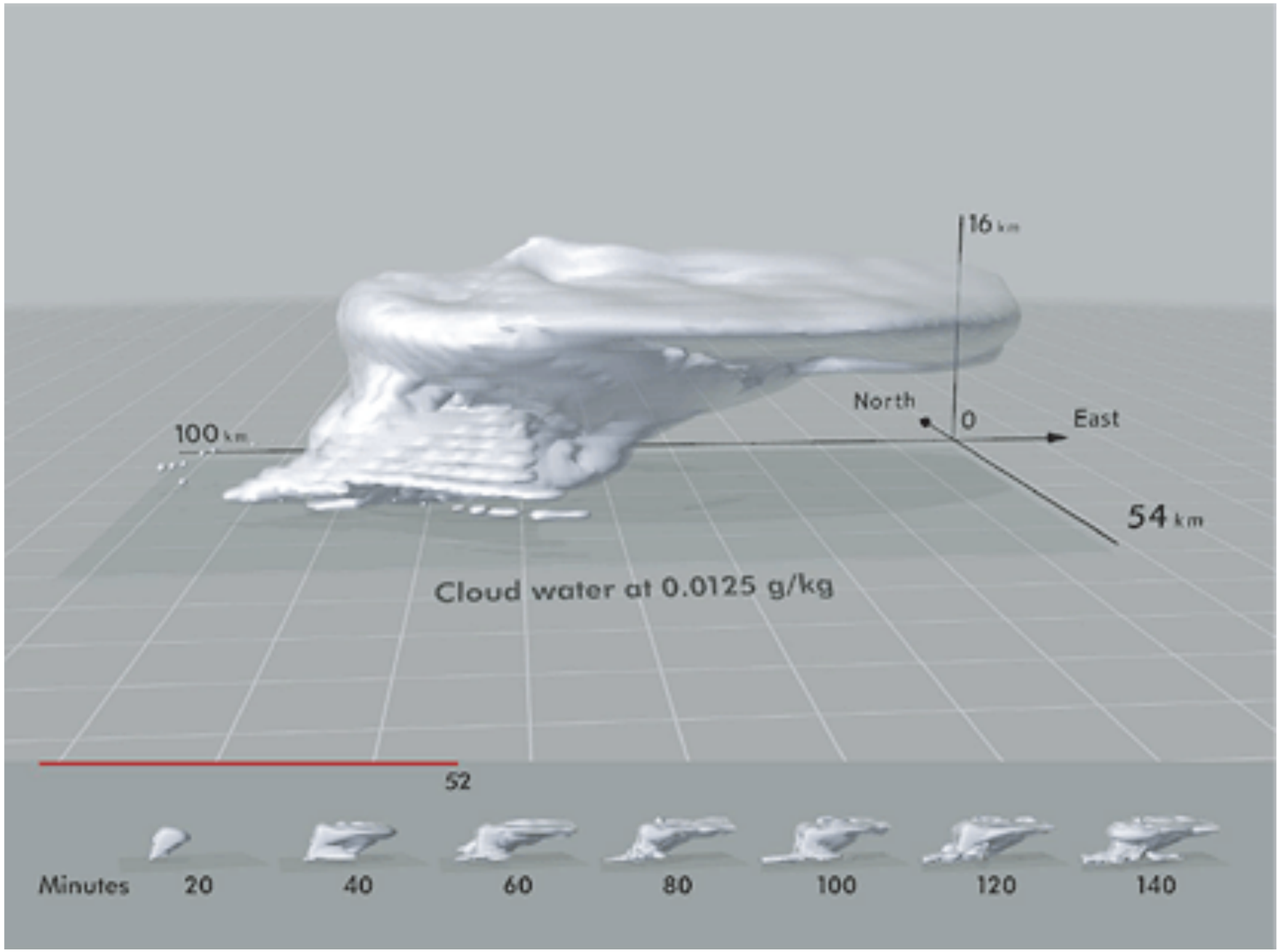


APPROVAL RATING *for* PRESIDENT OBAMA



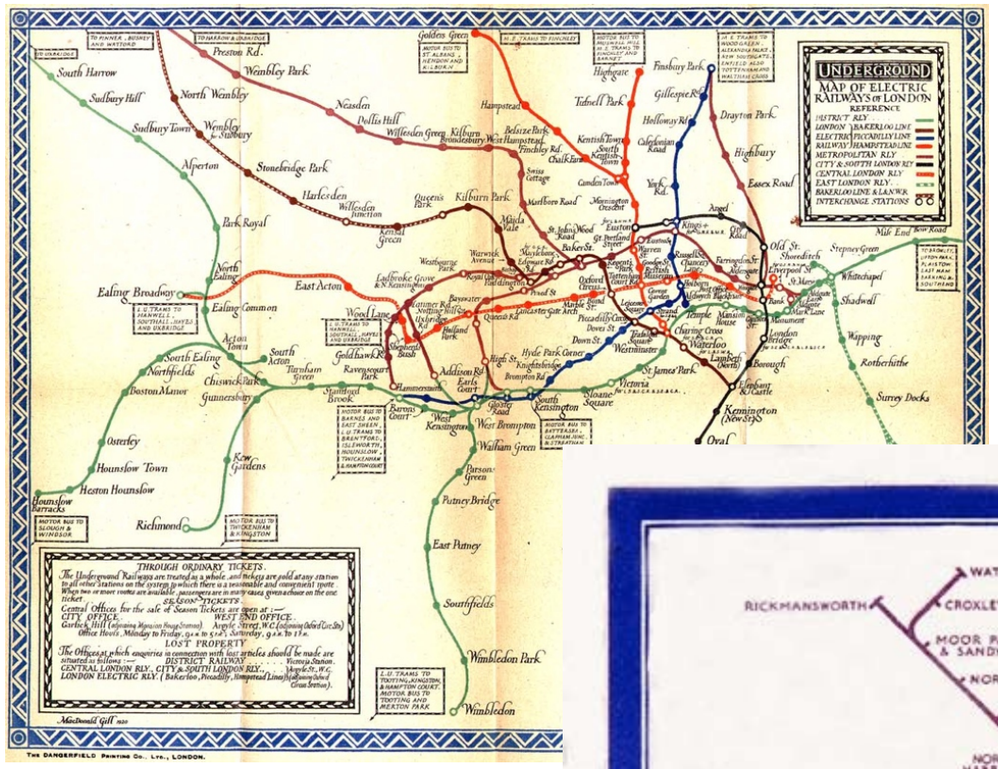
THIRD-PARTY HOUSE *candidates*

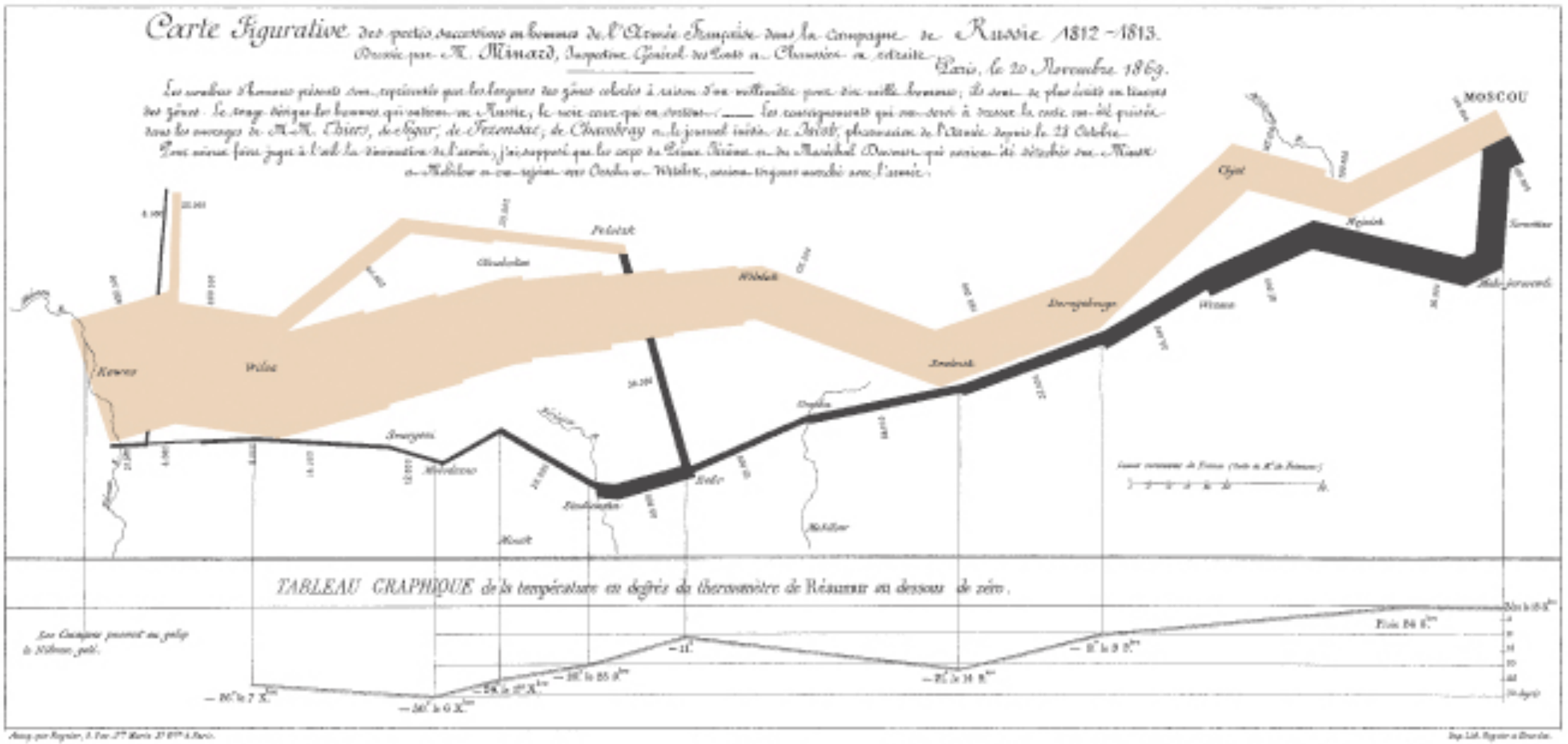
Data Visualisation--Edward Tufte



Each multiple maintains a consistent frame of reference—size, color, fonts—with changing data—cloud shape, number of minutes. The resulting information provides a complete narrative of how the storm changed over time.

Harry Beck 1933



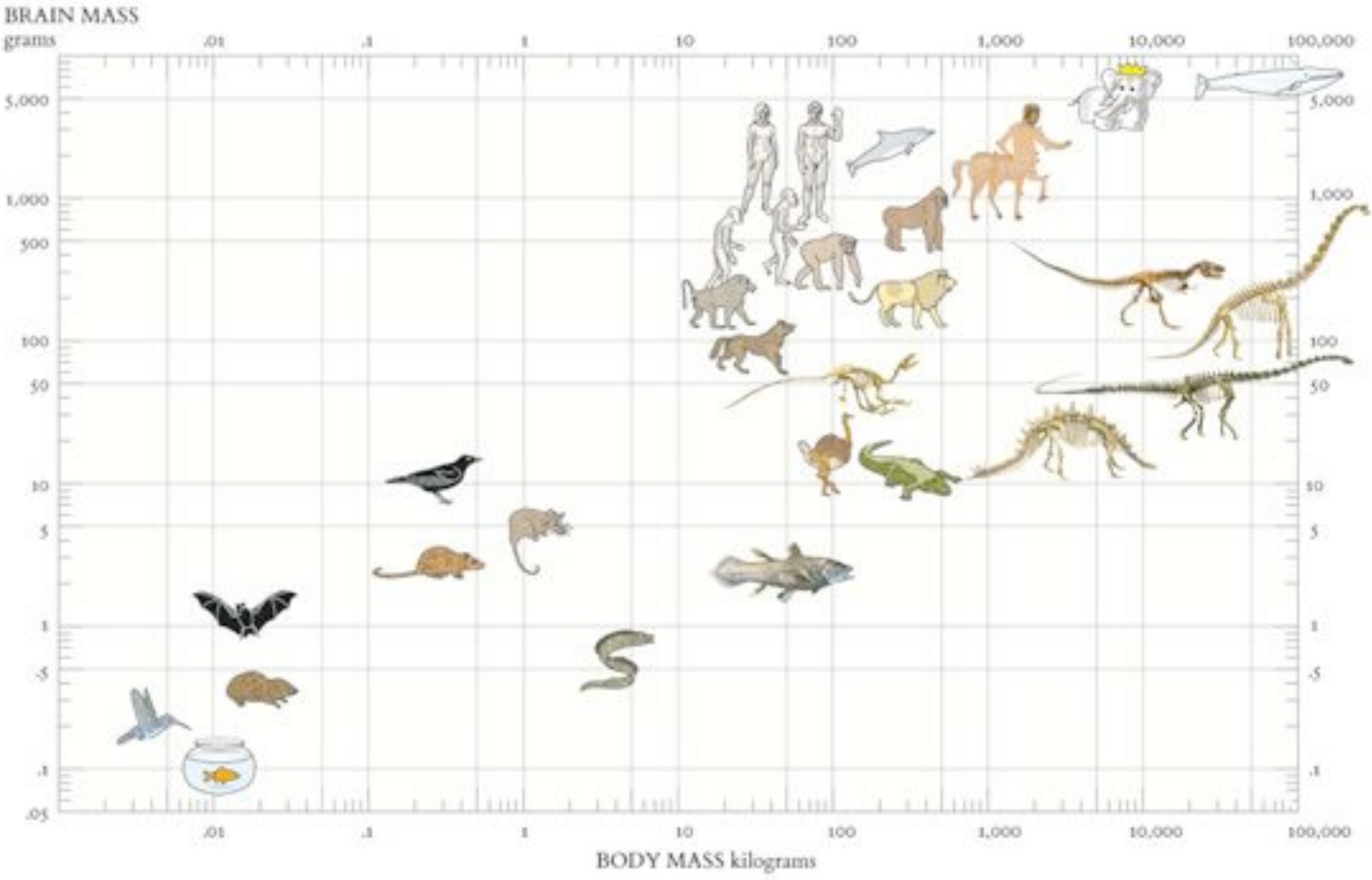


Napoleon's March to Moscow The War of 1812

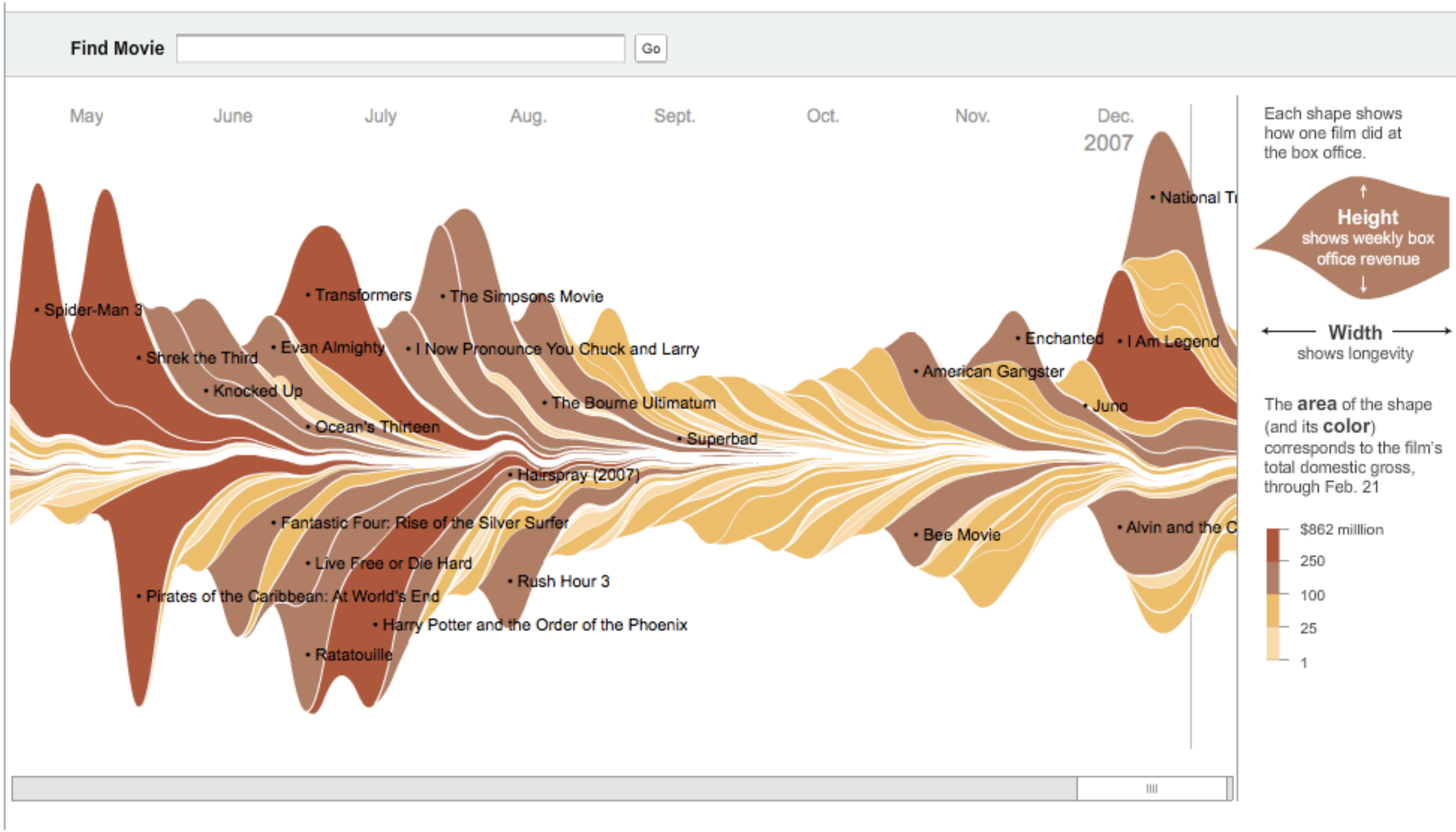
Charles Joseph Minard

This classic of Charles Joseph Minard (1781–1870), the French engineer, shows the terrible fate of Napoleon's army in Russia. Described by E. J. Marey as seeming to defy the pen of the historian by its brutal eloquence, this combination of data map and time-series, drawn in 1869, portrays the devastating losses suffered in Napoleon's Russian campaign of 1812. Beginning at the left on the Polish-Russian border near the Niemen River, the thick band shows the size of the army (422,000 men) as it invaded Russia in June 1812. The width of the band indicates the size of the army at each place on the map. In September, the army reached Moscow, which was by then sacked and deserted, with 100,000 men. The path of Napoleon's retreat from Moscow is depicted by the darker, lower band, which is linked to a temperature scale and dates at the bottom of the chart. It was a bitterly cold winter, and many froze on the march out of Russia. As the graphic shows, the crossing of the Berezina River was a disaster, and the army finally straggled back into Poland with only 10,000 men remaining. Also shown are the movements of auxiliary troops, as they sought to protect the rear and the flank of the advancing army. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. Six variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. It may well be the best statistical graphic ever drawn.

Brain Mass vs. Body Mass



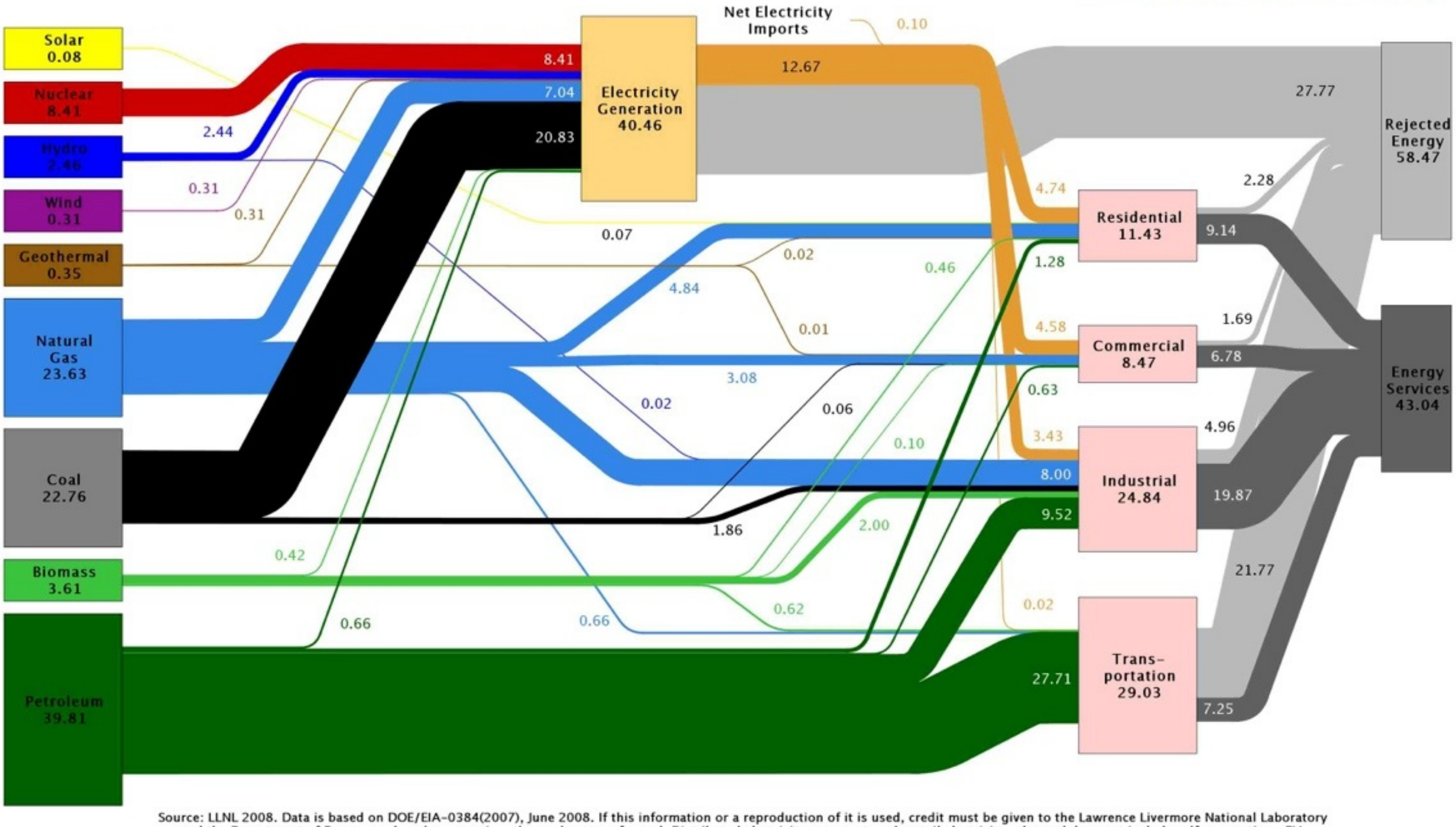
The Ebb and Flow of Movies: Box Office Receipts



http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html

Estimated U.S. Energy Use in 2007: ~101.5 Quads

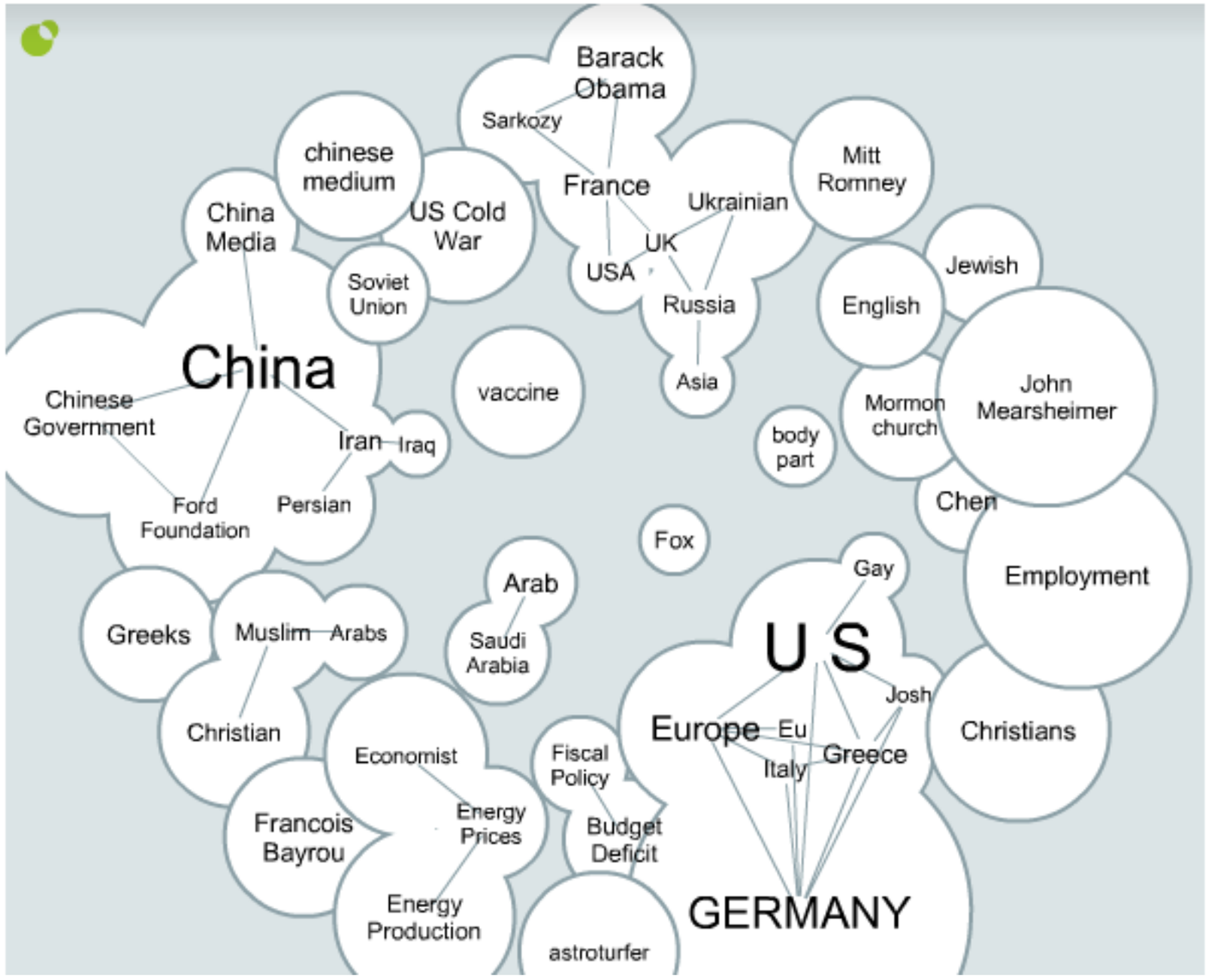
Lawrence Livermore National Laboratory

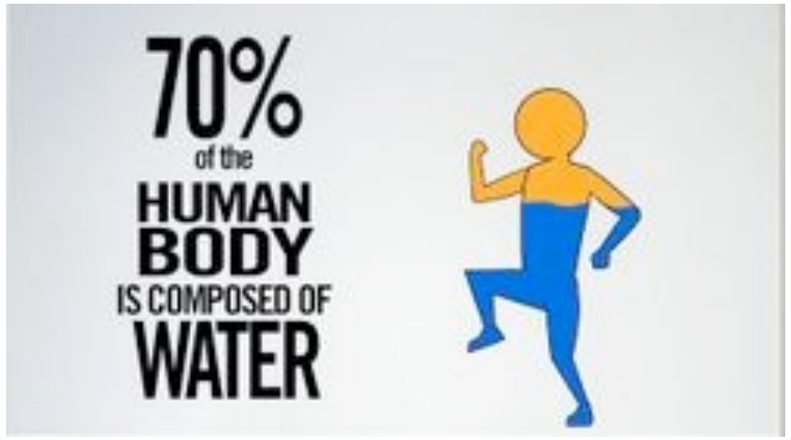


Source: LLNL 2008. Data is based on DOE/EIA-0384(2007), June 2008. If this information or a reproduction of it is used, credit must be given to the Lawrence Livermore National Laboratory and the Department of Energy, under whose auspices the work was performed. Distributed electricity represents only retail electricity sales and does not include self-generation. EIA reports flows for non-thermal resources (i.e., hydro, wind and solar) in BTU-equivalent values by assuming a typical fossil fuel plant "heat rate." The efficiency of electricity production is calculated as the total retail electricity delivered divided by the primary energy input into electricity generation. End use efficiency is estimated as 80% for the residential, commercial and industrial sectors, and as 25% for the transportation sector. Totals may not equal sum of components due to independent rounding. LLNL-MI-410527

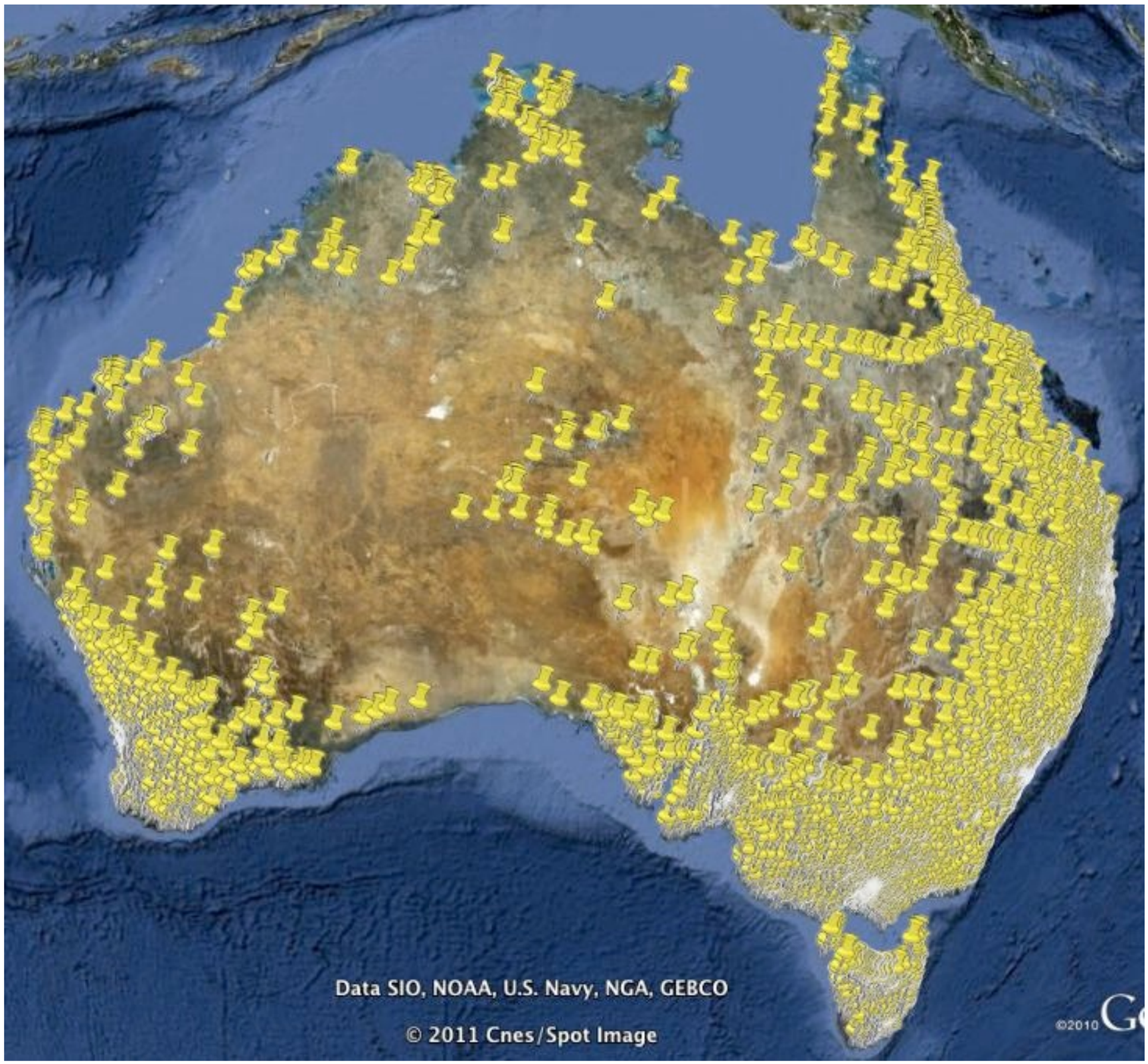
Topics most commented on

Read comments on the site's most popular topics

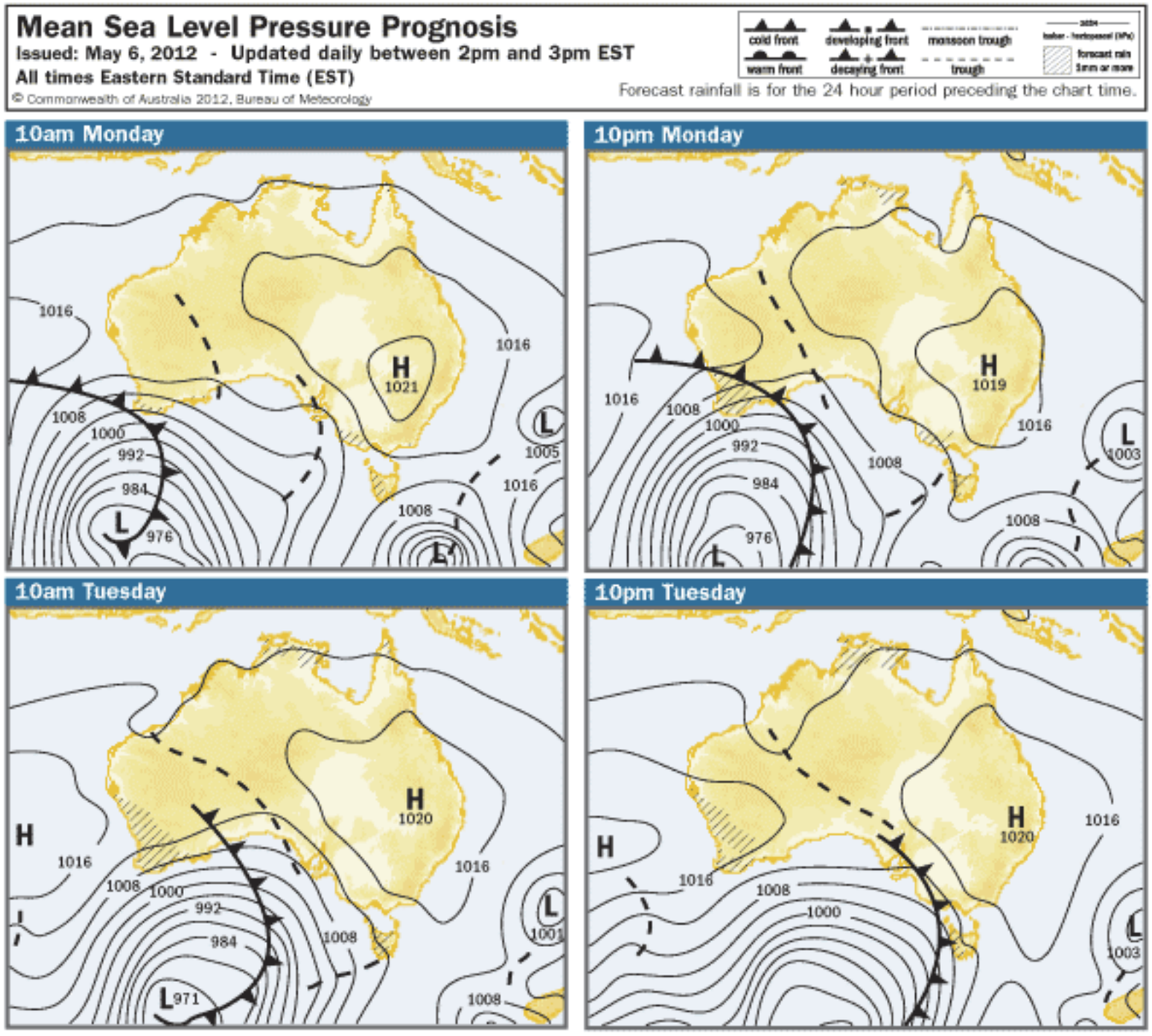
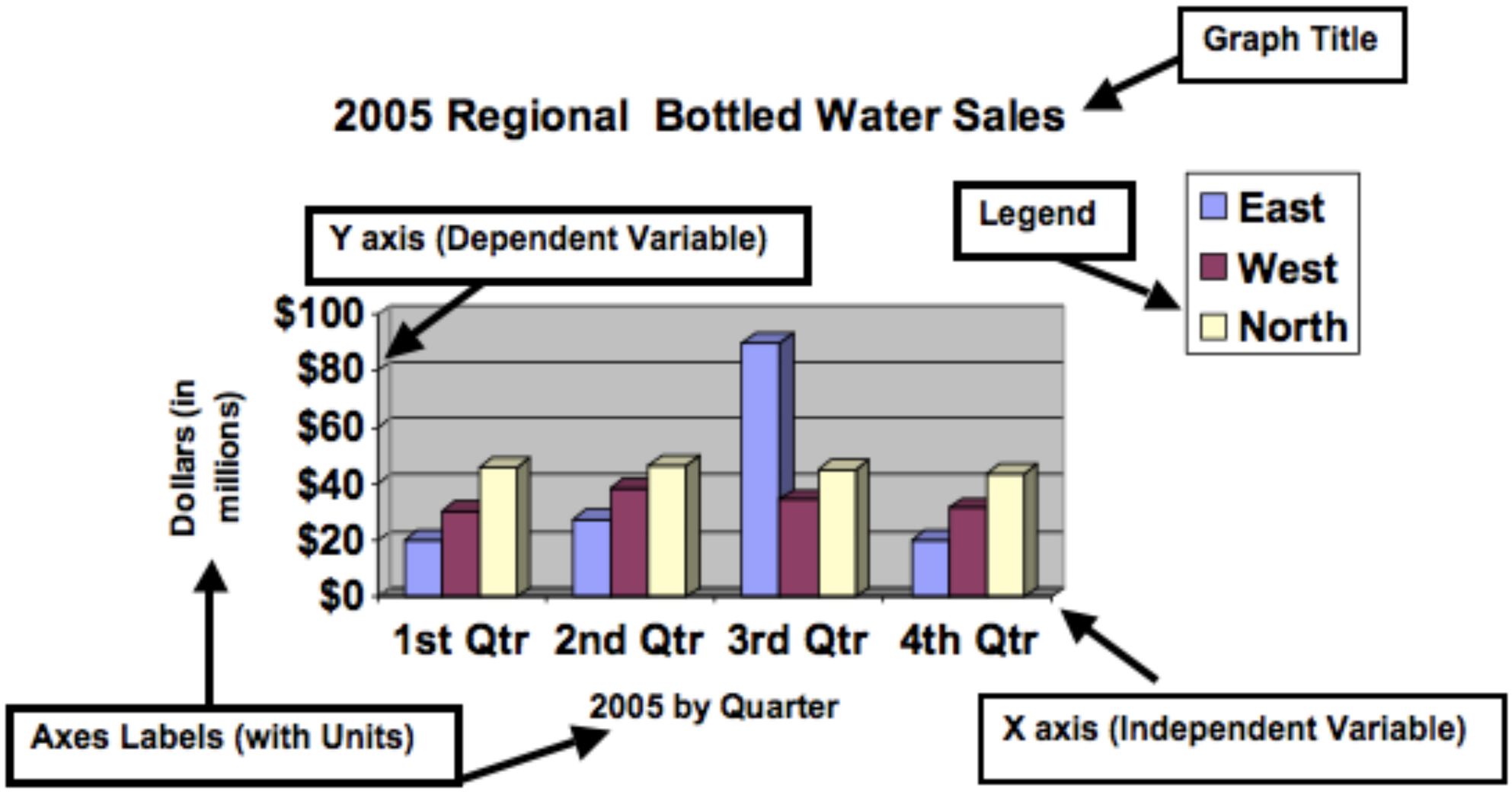




Public Loos in Australia



Weather


Mean Sea Level Pressure Prognosis
Issued: May 6, 2012 - Updated daily between 2pm and 3pm EST
All times Eastern Standard Time (EST)
© Commonwealth of Australia 2012, Bureau of Meteorology
cold front
developing front
monsoon trough
warm front
decaying front
trough
isobar - horizontal (hPa)
forecast rain 5mm or more
Forecast rainfall is for the 24 hour period preceding the chart time.
10am Monday
10pm Monday
10am Tuesday
10pm Tuesday

Required elements of a graph