GRAPHING DATA
FOR DECISION-MAKING

Tibor Tóth, Ph.D.
Center for Applied Demography and
Survey Research (CADSR)
University of Delaware

Fall, 2006

TABLE OF CONTENTS

INTRODUCTION

Chart presentations and narrated text are the two of the most common forms of presenting data to inform and aid decision-making. Which specific type of chart or display you use, however, can impact the interpretation of data itself. Ideally, the selected form should aid in achieving these goals:
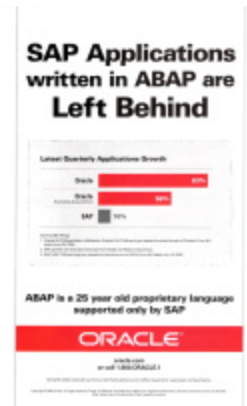
- **enhance** the clarity of the data
- increase the reader's ability to **perceive ideas** encoded in the data and point out relationships that otherwise are not self evident

Common problems in chart design that prevent the above goals from being achieved are:

- elements that **distract** from the evidence presented in the chart
- **poor choices that communicate** distorted numbers or relationships

The Economist, August 26th-September1st, 2006, p.84

This document is a synopsis of best practices in chart design for decision making, drawn from the work of Edward Tufte, Howard Wainer and John Tukey. They are not intended to guide the creation of marketing materials that sell (see examples at right). Most charts presented on the following pages are borrowed from reports created by the US Census Bureau, the Government Accountability Office and the United Nations Development Programme. When applicable, the chart is re-designed and the explanation of improvements is presented.

Business Week, September 4, 2006, p. 10

These guidelines are not strict rules. The main goals in chart design should always be to make the data clearer and help the reader to better perceive the embedded relationships
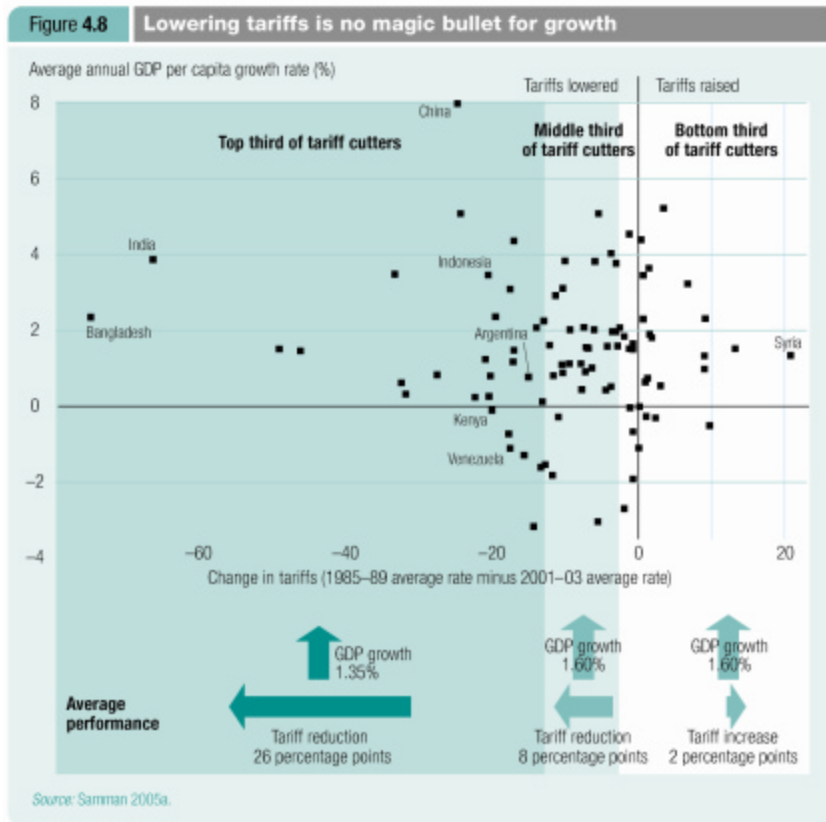
USE HIGH INFORAMTION GRAPHICS

Suitably designed information graphics with high information density are preferred to low information density graphs. A good example of a high information den sity chart appears below.


UNDP, *Human Development Report 2005,* 2005, p. 119.

The chart summarizes the relationship between the percentage change in tariffs and the growth of GDP per capita for 92 countries over the period 1985-89 to 2001-2003. Country to country comparison reveals that a diversity of outcomes is associated with decreased tariffs (both increases an d decreases of GDP are shown) - a conclusion that is expected. However, in addition to simply tabulating 2 variables against each other, the chart clusters the countries into three groups (top, middle and bottom) on the basis of the depth of their tariff cuts. The GDP change across groups is positive and comparable (even though the average percentage point change in tariffs varies greatly), thus the chart helps to support the argument that lowering tariffs is not a magic bullet for GDP growth and that links to growth are more complex than often argued. These insights would be hard to argue without a well designed high information graphic.

MAXIMIZE DATA INK RATIO

Tufte suggests: "A large share of ink on a graphic should present data-information, the ink changing as the data change".   One should not waste ink on printing embellishments that do not carry any data-information. The original figure on the left has a distracting background image and horizontal gridlines, both of which are unnecessary.
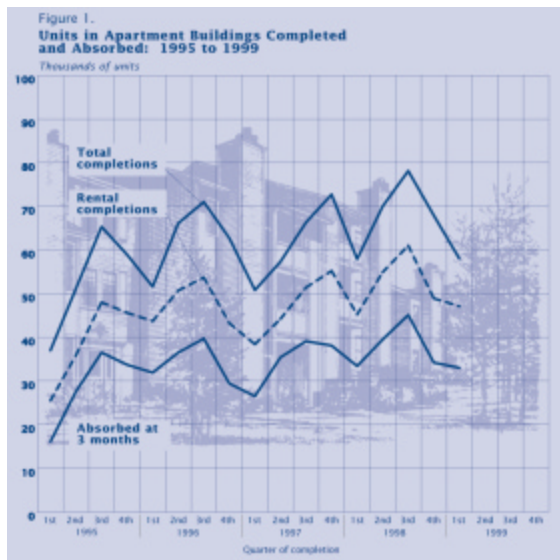

Census Bureau, *Market Absorption of Apartments Second Quarter 1999-Absorptions*, U.S. Government Printing Office, September 1999, p.1.
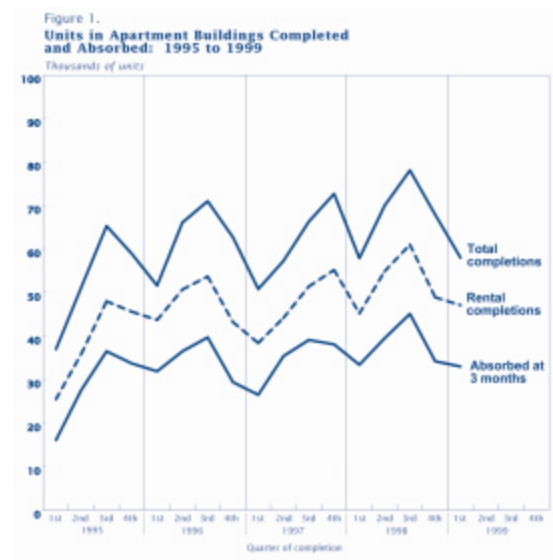

The same chart redesigned: both background image and horizontal grid lines removed

Stay away from chart designs where elements have been introduced that carry no information. See examples below: drawing of cigarettes on the left and the depiction of the house and furnace on the right. Although these graphics share the theme of the information displayed, they are distracting and do not add value. Such embellishments often miss present data, and both Tufte and Wainer argue against their use. The data by itself should be interesting and conclusive enough without the need to over-produce it.
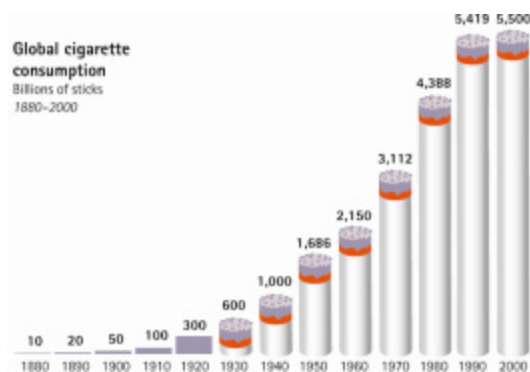

WHO, *The Tobacco Atlas*, 2002, p.31
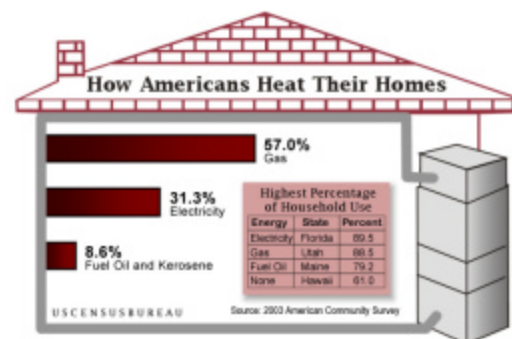
Census Bureau, *Gas Most Popular Home Heating Fuel*, Press Release CB04-CN.08, November 16, 2004

SUPPRESS THE GRID

Grid lines were originally used to aid the charting of data. Tufte suggests that they carry no information, they clutter up the graphic, and they even foil the observer's ability to recognize patterns in the data that would otherwise be apparent. Thus, in most cases, gridlines should be muted or suppressed.
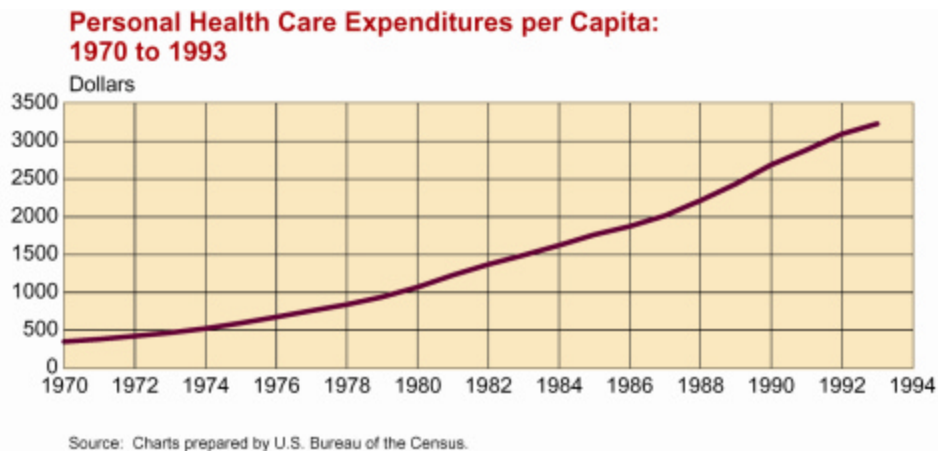
Census Bureau, *Statistical Abstract of the United States,* 1995, p. 106

The above chart is either incomplete or cluttered. If it includes the grid to allow a lookup of numerical values, then it is missing the precision to do that effectively. If it attempts to simply show a trend from 1970 to 1994, then the vertical gridlines and the background color are distracting elements that carry no valuable information. They are not increasing the clarity of the data nor do they enhance the ability of the reader to perceive the embedded relationships. Below is the same data, presented without the vertical gridlines and without the background color. Some of the horizontal gridlines were retained to allow the reader to quickly perceive the trend over time (below $1,000 in 1970, above $3,000 in 1993).
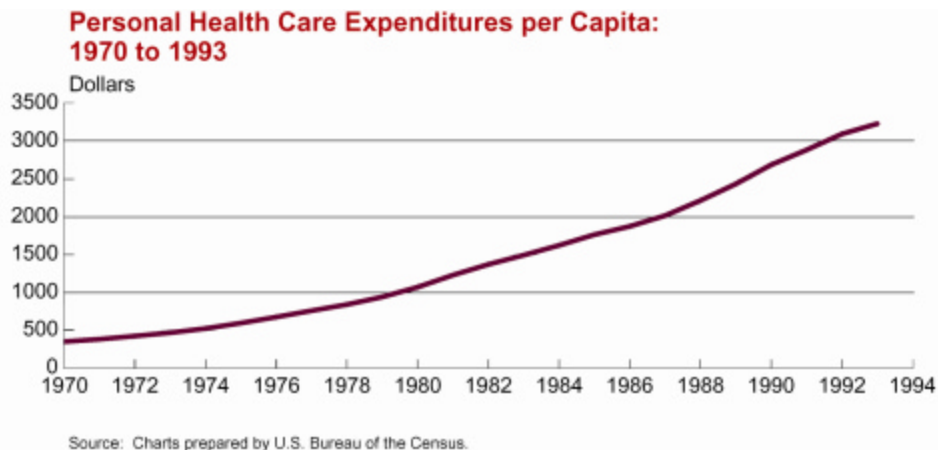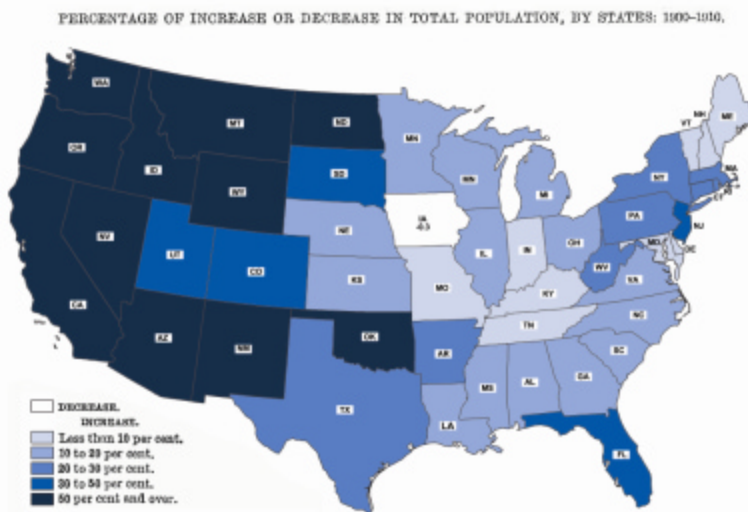

Same chart redesigned but without the grid line.

MINIMIZE THE MOIRÉ EFFECT

Usually a side effect of overzealous graphic designers that intend to embellish charts and graphs with different fill effects, the moiré effect is an optical design that interacts with the physiological tremor of the eye. It prevents the viewer from focusing on the chart and thus decreases their ability to perceive the presented data. Tufte suggests replacing cross hatching with different density and shades of gray, using different shades of the same color, or by labeling the areas with words.
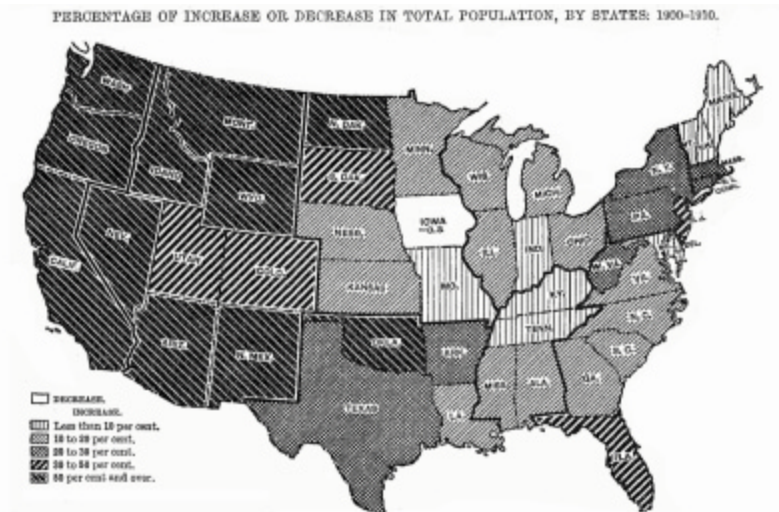
Department of Commerce, Bureau of the Census, *Population 1920,* U.S.Government Printing Office, 1921, p.19.

The same chart redesigned without the Moiré effect using color gradation

The various stripes in the first map cause a moiré effect (focus on UT and NV). The same data is remapped in the second map using different shades of blue. The human brain can readily perceive and effectively interpret the meaning of lighter vs. darker and the categories associated with this variation. An added benefit of using different shades of the same color is that the graph retains its information value whether viewed in color or in grayscale. Please note: Using more than one color usually makes the interpretation more difficult.

TELL THE STORY WITH A TABLE

Tables are often used to present data for lookup purposes. Even though tables usually do not accommodate visual perception of data they can be designed to do so. Tables that present data visually are often a good choice in situations where little data and lots of words are to be presented. The example below is an effective combination of a data table and a chart.

Figure 5.1 Top-grossing films of all time at the international (non-US) box office were US films, April 2004

Rank	US Films	Year	Country of origin	Total gross revenue (millions of US$)
1	Titanic	1997	US	1,235
2	Lord of the Rings: The Return of the King	2003	US	696
3	Harry Potter and the Sorcerer's Stone	2001	US	651
4	Harry Potter and the Chamber of Secrets	2002	US	604
5	Lord of the Rings: The Two Towers	2002	US	581
6	Jurassic Park	1993	US	563
7	Lord of the Rings: The Fellowship of the Ring	2001	US	547
8	Finding Nemo	2003	US	513
9	Independence Day	1996	US	505
10	Star Wars: Episode I: The Phantom Menace	1999	US	491
	Non-US Films			
44	Sen to Chihiro no kamikakushi	2001	Japan	254
69	The Full Monty	1997	UK	211
86	Four Weddings and a Funeral	1994	UK	191
96	Bridget Jones' Diary	2001	UK	183

Source: The Internet Movie Database 2004.

UNDP, *Human Development Report 2004,* 2004, p. 119.

Table that effectively aid the visual perception of data do the following:

- only include elements that support the claim (otherwise you are only creating a lookup table)
- order the table by an aspect of data
- decrease the number of decimal places
- use clear labels

ORDER BY AN ASPECT OF DATA

The analytical value of a properly designed chart is not measured in its ability to present categories in alphabetical order but to present insights into the changes to the variable across categories.   Wainer suggests creating charts that have been ordered by an aspect of the data.

Figure 2-7.
Aging Index: 2000 and 2030
(People aged 65 and over per 100 people aged 0-14)

DEVELOPED COUNTRIES
Western Europe
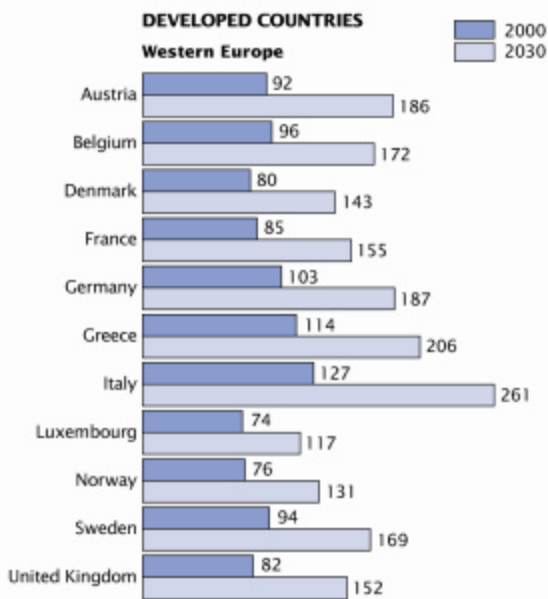
Census Bureau, *An Aging World: 2001,* November 2001, p. 14

Figure 2-7.
Aging Index: 2000 and 2030
(People aged 65 and over per 100 people aged 0-14)

DEVELOPED COUNTRIES
Western Europe
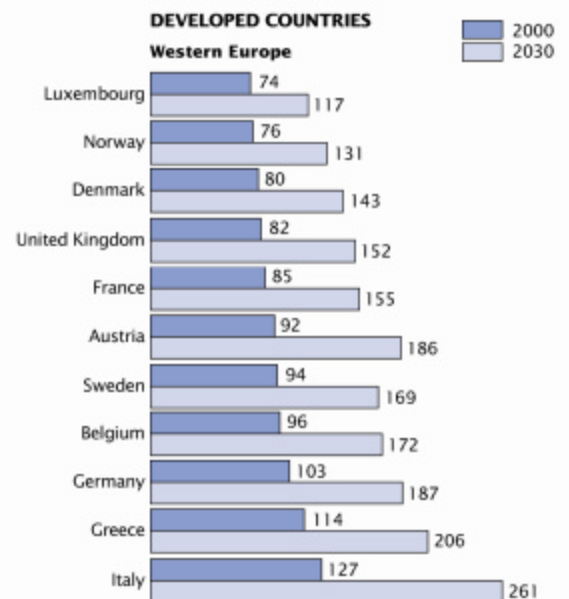
The same figure with a different order

The original chart on the left presents the aging index for selected developed countries, shown in alphabetical order.   This is usually done to allow for an easy lookup of specific values for specific countries. Tables are more suitable for that task. What tables do not allow us to do is to visually perceive the data and recognize patterns. Once the bar chart is ordered by an aspect of the data, the outliers jump out. Both Austria and Italy are outliers, their populations will age faster than the populations of the other countries presented in the chart.

STAY AWAY FROM PIE CHARTS

In general, pie charts are used because of their visual appeal and their ability to clearly indicate that the numbers presented ad up to 100%. However, pie charts perform very poorly in presenting the data and empowering the viewer to make comparisons. Different angles of the slices along with different colors make it very difficult to analytically comprehend the presented information. It is often the case that perception and comparison of categories presented in a pie chart is not based on the graphic itself, but by comparing the values provided in the legend. Both Wainer and Tufte suggest that data from pie charts are more effectively presented using bar charts or point charts, both better suited for cross comparisons.
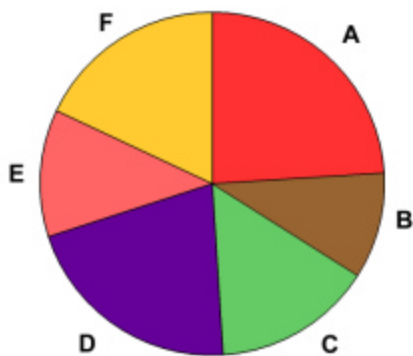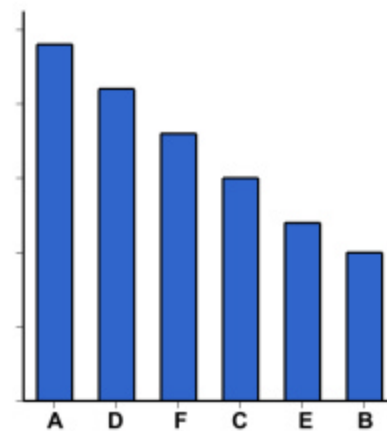
Simulated data presented in a pie chart   Simulated data presented in a bar graph

Above is an example of simulated data presented in a pie chart, and the same data presented using a bar graph. Looking at the pie chart first, it is rather difficult to make conclusive comparisons among categories. It is obvious that A is larger than B and A is larger than E. It is also clear that D is larger than B and D is larger than E. However, conclusions about relative sizes of slices D and A, B and E and C and F are hard to make. The changing colors, slopes of the lines dividing the slices and the relative positioning of the slices all make it difficult to clearly perceive the data. On the right, the same data is presented in a bar graph, with comparisons and relative sizes easy to distinguish.

Please note: Data labels were purposefully omitted from these charts to indicate the point about the inefficacy of the pie chart to communicate data visually. Annotating the pie chart with data values does not solve the inconclusiveness of the design – it would just suggest that the data should be presented in a table and not a pie chart.
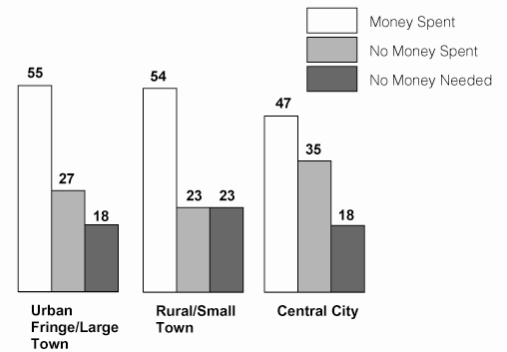
NO DROP SHADOWS AND NO 3D CHARTS

No need to get overly fancy! Graphs often include elements that don't help to convey the message. Creating charts with drop shadows and adding a 3rd dimension is a bad idea. Both of these elements decrease the clarity of the chart and the reader's ability to comprehend the information.
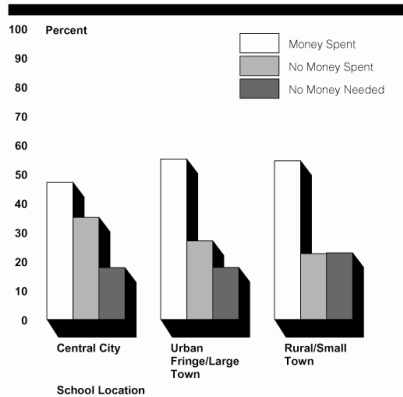
GAO, *School Facilities,* December 1995, p. 6. (adapted)

The same figure without the drop shadow effect

The bar chart above on the left has been embellished by adding a drop shadow. This feature, however, carries no additional information and does not make the chart more informative. The same data is presented above on the right without the drop shadow effect, bars are re-ordered and data labels added.
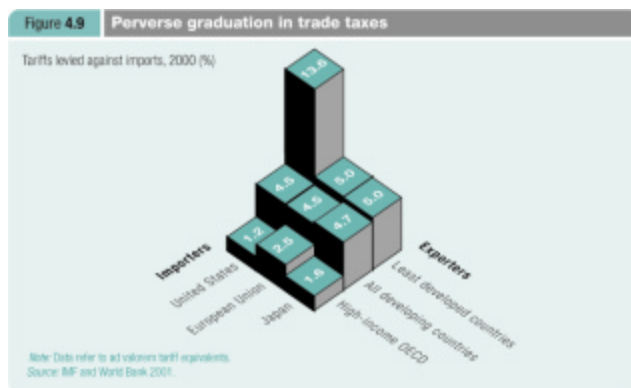
UNDP, *Human Development Report 2005,* 2005, p. 127.

Table **4.9** Perverse graduation in trade taxes			
Tariffs levied against imports, 2000 (%)		IMPORTERS	
EXPORTERS	United States	European Union	Japan
Least developed countries	13.6	5.0	5.0
All developing countries	4.5	4.5	4.7
High income OECD	1.2	2.5	1.6

Note: Data refer to ad valorem equivalents.
Source: Imf and World Bank 2001.

The same data presented in a table without the 3rd dimension

The 3D bar chart on the left exaggerates the differences in tariffs (see the bar representing the value 13.6%) because of the shading and the perspective. Comparisons can be more effectively and accurately done within a small look-up table as presented on the right.

BEWARE OF STACKED BAR CHARTS

The stacked bar chart makes it very difficult to estimate the values presented and to make conclusive comparisons.

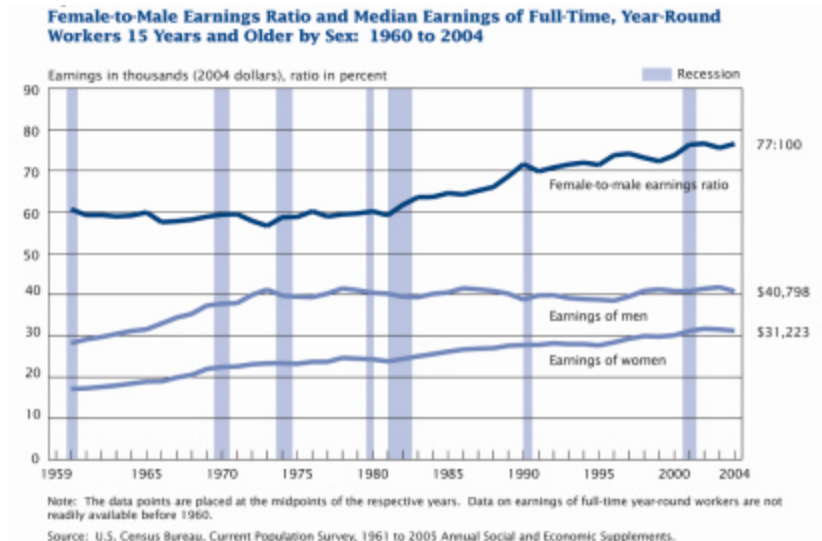Figure 6: Fiscal Year 2006 Authorizations for Acquisition Personnel by Rank at SMC and for the Air Force Overall
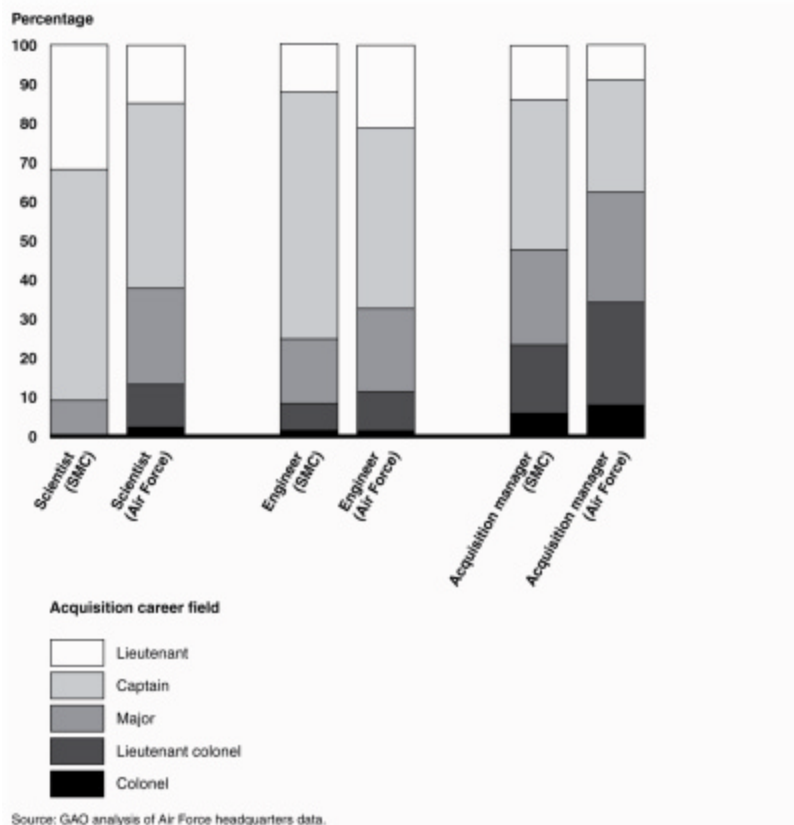
GAO, *Defense space activities,* September 2006, p. 24

The above stacked bar chart presents the authorizations for acquisition of personnel by rank at the Space and Missile Systems Center and the Air Force. The data is presented in percentages. Unfortunately, the use of bar charts prevents the reader from making meaningful comparisons. Attempt to answer the following questions:

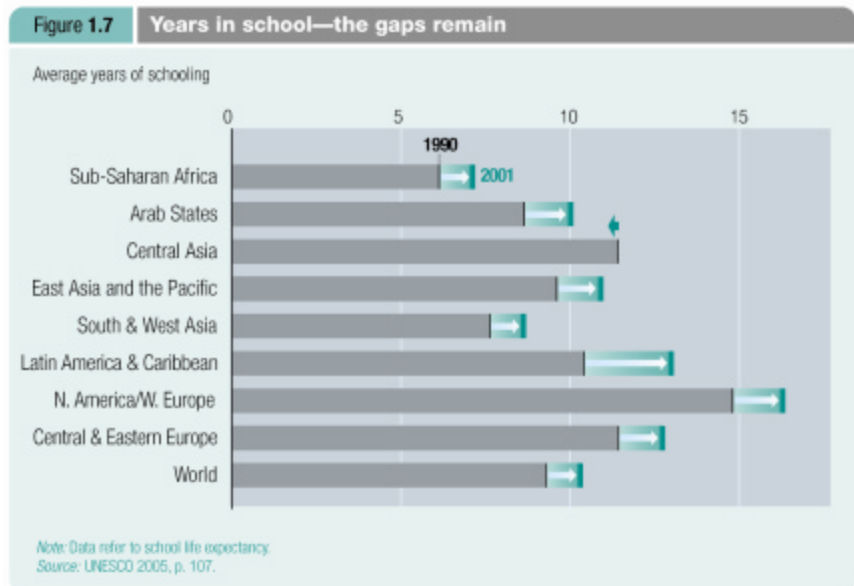? Percentage wise are there more captains among scientists in the Air force or SMC?

? Is the percentage of Majors among Acquisition managers higher in the Air Force or in SMC?

? Estimate the percentage of Lieutenants among Engineers at SMC and Acquisition Managers at the Air Force.

The ambiguity of the answers comes not from the unavailability of the data but is directly related to the choice of the graph.
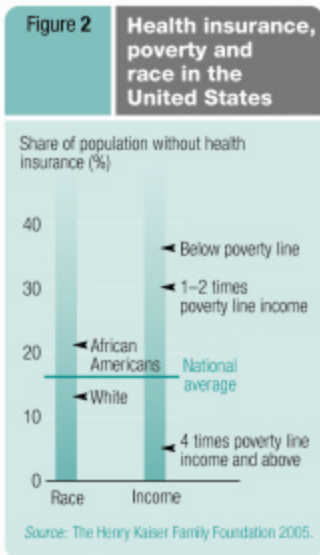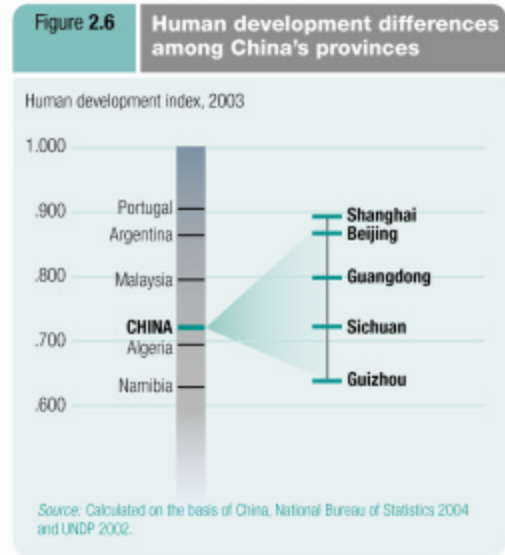
EXAMPLES OF WELL DESIGNED GRAPHS

Female-to-Male Earnings Ratio and Median Earnings of Full-Time, Year-Round Workers 15 Years and Older by Sex: 1960 to 2004

Earnings in thousands (2004 dollars), ratio in percent

Note: The data points are placed at the midpoints of the respective years. Data on earnings of full-time year-round workers are not readily available before 1960.

Source: U.S. Census Bureau, Current Population Survey, 1961 to 2005 Annual Social and Economic Supplements.

Census Bureau, *Income, Poverty, and Health Insurance Coverage in the United States: 2004*, U.S. Government Printing Office, 2005, p.7.

Figure 1.7 Years in school—the gaps remain

Average years of schooling

Note: Data refer to school life expectancy.
Source: UNESCO 2005, p. 107.

UNDP, *Human Development Report 2005,* 2005, p. 25.

Figure 2 — Health insurance, poverty and race in the United States

Share of population without health insurance (%)

Source: The Henry Kaiser Family Foundation 2005.

UNDP, *Human Development Report 2005,* 2005, p. 58.

Figure 2.6 — Human development differences among China's provinces

Human development index, 2003

Source: Calculated on the basis of China, National Bureau of Statistics 2004 and UNDP 2002.

UNDP, *Human Development Report 2005,* 2005, p. 59.

Figure 1.11 — Child mortality—gaps between rich and poor are widening

Under-five mortality rate
Index (high-income OECD average = 1)
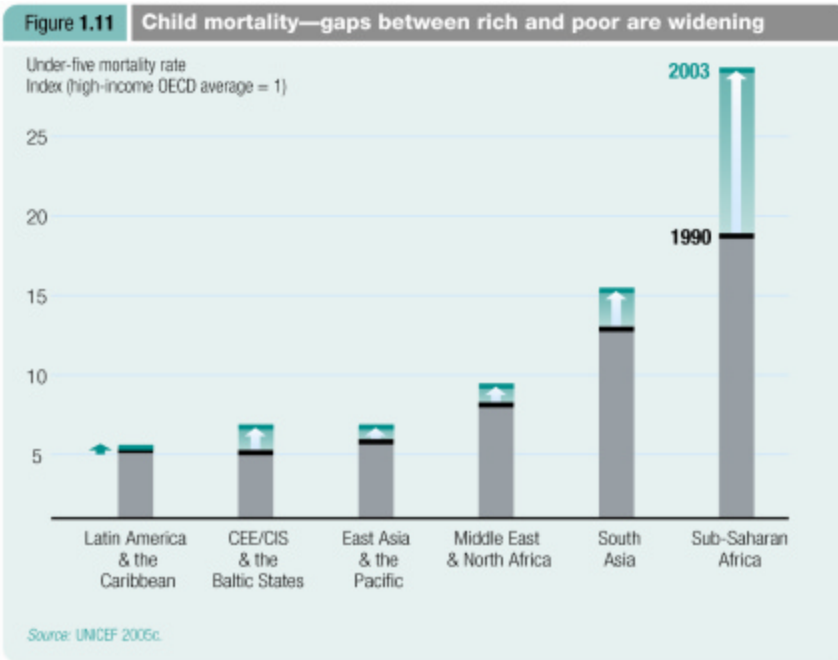
Source: UNICEF 2005c.

UNDP, *Human Development Report 2005,* 2005, p. 28.

UNDP, *Human Development Report 2005,* 2005, p. 62.

UNDP, *Human Development Report 2005,* 2005, p. 89.

UNDP, *Human Development Report 2004,* 2004, p. 67.

Figure
2.4 **Many lack access to primary education in their mother tongue**

| Region or group	Number of spoken languages	Population with access to education in mother tongue in 2000	Total population (millions)
Sub-Saharan Africa	2,632	13%	641
East Asia and the Pacific	2,815	62%	1,918
South Asia	811	66%	1,480
Central and Eastern Europe and the CIS	625	74%	409
High-income OECD	1,299	87%	912
Latin America and the Caribbean	1,086	91%	530

Source: SIL International 2004b.

UNDP, *Human Development Report 2004,* 2004, p. 34.

Figure
2.2 **Indigenous people can expect a shorter life**

Gap in life expectancy in selected developing countries, 1997–2000

- Indigenous population
- Non-indigenous population
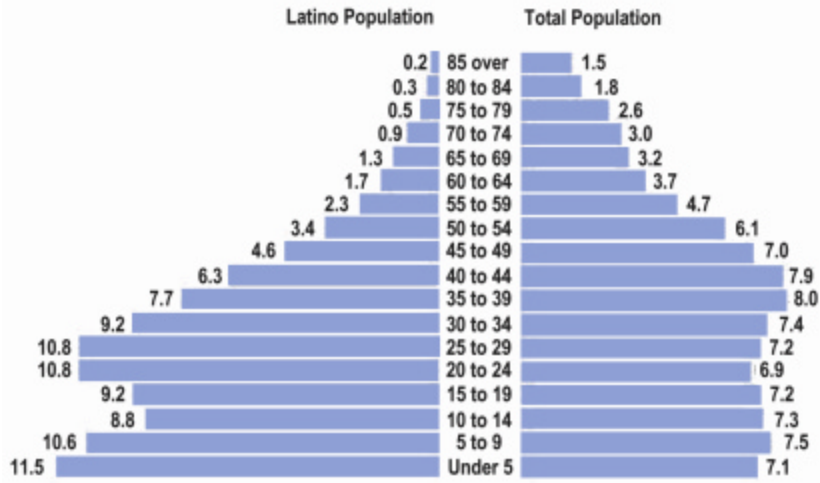- Gap in life expectancy
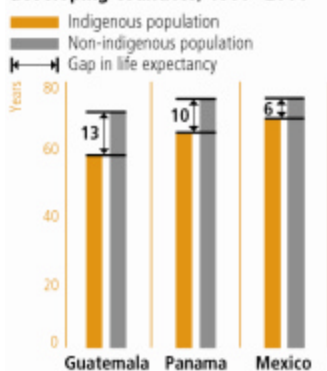
Source: During 1993 (share of indigenous peoples);
WHO 2001 (Guatemala); UNDP 2002b
(Panama); Mexico, Ministry of Health 2004
(Mexico).

UNDP, *Human Development Report 2004,* 2004, p. 29.

Figure 2.
Age of Latino and Total Population in Illinois: 2000

	Latino Population	Age	Total Population	
	0.2	85 over	1.5	
	0.3	80 to 84	1.8	
	0.5	75 to 79	2.6	
	0.9	70 to 74	3.0	
	1.3	65 to 69	3.2	
	1.7	60 to 64	3.7	
	2.3	55 to 59	4.7	
	3.4	50 to 54	6.1	
	4.6	45 to 49	7.0	
	6.3	40 to 44	7.9	
	7.7	35 to 39	8.0	
	9.2	30 to 34	7.4	
10.8		25 to 29	7.2	
10.8		20 to 24	6.9	
	9.2	15 to 19	7.2	
	8.8	10 to 14	7.3	
10.6		5 to 9	7.5	
11.5		Under 5	7.1	

Source: U.S. Census Bureau, Census 2000.

Census Bureau, *Using Census Data to Help Local Communities: Census Information Centers at Work,* October 2003, p. 17

SOURCES

Wainer, H. 1997. *Visual revelations - Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot.* New York: Copernicus Books.

Wainer, H. 2005. *Graphic discovery.* New Jersey: Princeton University Press.

Tufte, E.R. 1983. *The Visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tufte, E.R. 1996. *Visual Explanations*. Cheshire, CT: Graphics Press.

Tukey, J. W. 1977. *Exploratory Data Analysis.* Reading, MA: Addison-Wesley