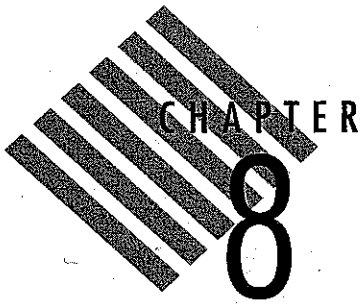
ANALYZING PUBLIC POLICY

Concepts, Tools, and Techniques

Dipak K. Gupta
San Diego State University

CQ PRESS
A Division of Congressional Quarterly Inc.
Washington, D.C.



MAKING SENSE OF NUMBERS

Much of a busy executive's workday involves making decisions. These days executives find themselves increasingly surrounded by information, often expressed in numbers. The spectacular advances in technology have made collecting and storing information inexpensive and data retrieval and the display of their analyses quite simple. A large series of numbers, however, tends to numb our senses and push us beyond our cognitive capabilities. Therefore, decision makers prefer seeing large sets of numbers in an understandable form.

The purpose of this chapter is to provide students and practicing analysts with a guide for rendering social, political, or economic phenomena in graphs or tables. It is important to be creative without being misleading or deceptive. I will take you through some familiar terrain and expose you to the advantages and pitfalls of the most commonly used and abused methods of presenting numerical arguments to decision makers.

Descriptive statistics and graphical techniques are useful in assessing social conditions, such as per capita income, rate of population growth, and crime rates. These techniques are often neglected in more sophisticated statistics and operations research textbooks because they seem too simplistic. However, an overwhelming

number of decisions in both public and private organizations are made on the basis of simple decision rules: a brief assessment of relative desirability based on quick impressions rather than thorough research.

A PICTURE'S WORTH: THE GRAPHICAL METHODS OF ANALYSIS

Let us consider a hypothetical situation. Suppose the city manager of a medium-size city, Masters, Pennsylvania, would like to know how much the city depends on state and federal grants. The financial management division for the city gives the city manager breakdowns of the city's state and federal grants for 1990 through 1999. (see Table 8.1).

Current vs. Constant Dollars

The numbers in Table 8.1 indicate that state and federal funding for the city is growing every year. A better appreciation of the historical trend can be obtained by plotting the data. However, as you will soon discover, there is more than one way of drawing a picture.

Of course, the plotted information shown in Figure 8.1 is a great improvement over the table. The diagram clearly shows the generosity of the state and federal governments to the city. The dollar amount of the grants has increased steadily over the years, with the greatest increase coming during the early 1990s. However, the city manager is skeptical; this diagram does not take into account the rate of inflation for the period. Therefore, the price deflator (the consumer price index for the United States) was obtained for the period, and the yearly numbers were converted to constant dollars (see Table 8.2).

Time series data requires the conversion of current dollars into constant dollars because the value of money does not remain the same over time. The books that you purchased last year probably cost more today. Such increases make the value

Table 8.1 State and Federal Grants to Masters, Pennsylvania, in Current Dollars, 1990–1999

Year	*State and federal grants in current $*
1990	71,000
1991	75,000
1992	85,000
1993	89,000
1994	91,000
1995	91,500
1996	93,000
1997	95,000
1998	98,000
1999	102,000

Figure 8.1 State and Federal Grants to Masters, Pennsylvania, in Current Dollars

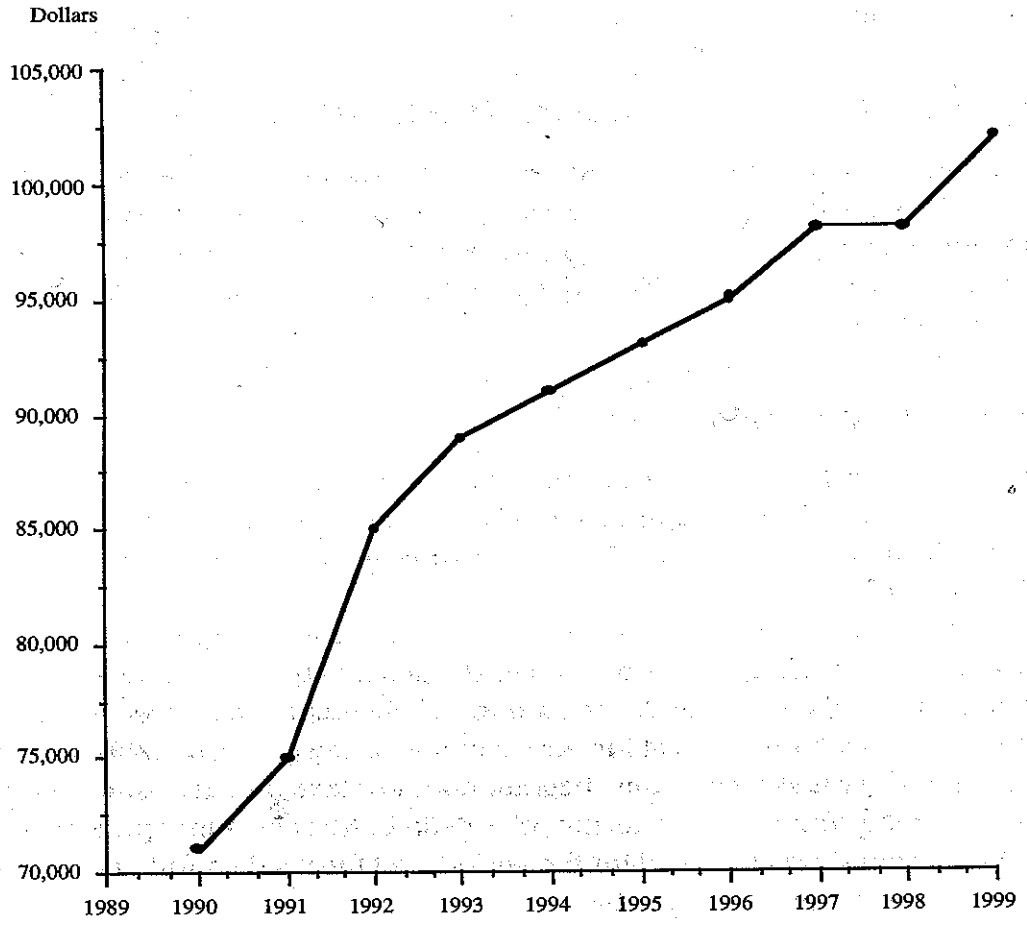


Table 8.2 State and Federal Grants to Masters, Pennsylvania, in Current and Constant Dollars, 1990–1999

Year	*State and federal grants in current $*	*Consumer price index number*[a]	*State and federal grants in constant $*
1990	71,000	100	71,000
1991	75,000	110	68,182
1992	85,000	117	72,650
1993	89,000	122	72,951
1994	91,000	127	70,930
1995	91,500	129	67,883
1996	93,000	137	68,345
1997	95,000	139	68,534
1998	98,000	143	68,531
1999	102,000	147	68,388

[a] The consumer price index data are hypothetical.

of money decrease with the passage of time. Therefore, by using the actual dollars received by Masters from the state and the federal government, we paint a deceptive picture. The rate at which the dollar loses its value is measured by a price index number. In the United States the Bureau of Labor Statistics (BLS) collects data on price changes and other employment-related information.[1] Price changes are measured in terms of a base year, which is expressed as 100. For example, most of the current series in price indexes measured by the BLS hold the period 1982–1984 as 100.

A price index is calculated on the basis of what is known as **variable weight,** or **fixed weight.** The variable-weight price index states price change as a ratio of a set of goods and services in the current period and their costs in the base year. More precisely, it is written as:

$$\text{Variable-weight price index} = \frac{\textit{value of goods at current-year prices}}{\textit{value of those goods at base-year prices}}$$

The effects of price changes are, of course, not universal. If you do not like turnips, you are not affected if their prices go through the roof. However, if you practically live on turnips, your well-being will be affected. If you are making the same amount of money as last year, you are poorer this year than you were before. Your income in current dollars is called the **nominal income,** and the true value of your reduced income is called the **real income** (meaning, adjusted for price change). You may note that the variable-weight price index takes an overall picture and does not consider whether the price changes are going to affect any single segment of the population. As a result, the most commonly used variable-weight price index is called the **GDP deflator,** which is used to compare the GDP of the past year with that of the current year.

In contrast, the fixed-weight price index considers a typical basket of goods that a consumer would consume and tracks its prices over the years. The fixed-weight price index is used to measure the **consumer price index,** or the changes in prices that will affect a typical consumer. The BLS produces data series on the consumer price index for various regional centers (urban, rural, a specific city, and so forth) as well as GDP deflators. For example, suppose you have a job offer from two different cities. By comparing the price levels of the two, you can decide which one is offering you a higher salary in "real" dollars. In Table 8.3 I show the change of prices facing all urban consumers.

From this table you can see that between 1982–1984 and 1999, a dollar lost 66.6 percent of its value. That is, a dollar in 1999 was worth about 34 cents in 1982–1984.

You can also calculate the amount of inflation by changing the base. For example, if you want to use 1990 as the base year, you can recalculate the series by using the following formula:

$$\pi_{t+1} = \frac{P_{t+1} - P_t}{P_t} \times 100 \qquad (8.1)$$

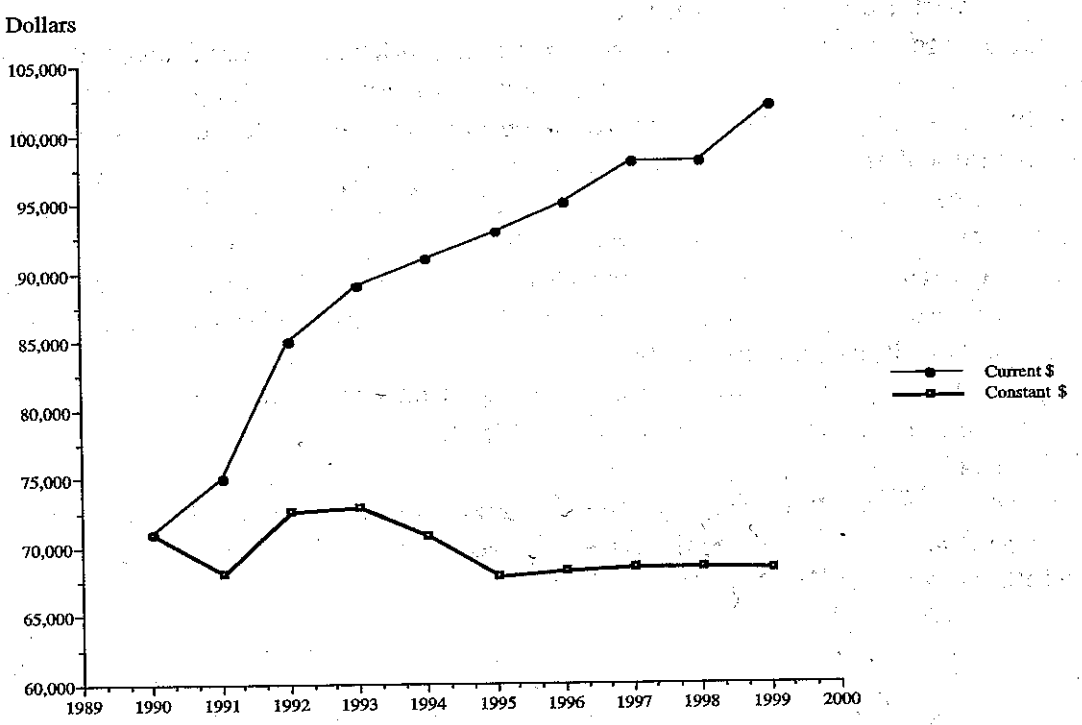
Table 8.3 Consumer Price Index—All Urban Consumers, 1990–1999

Year	*CPI (1982–1984 = 100)*	*CPI (1990 as the base year)*	*Rate of inflation*
1990	130.7	100.00	—
1991	136.2	104.21	4.21
1992	140.3	107.34	7.34
1993	144.5	110.56	10.56
1994	148.2	113.39	13.39
1995	152.4	116.60	16.60
1996	156.9	120.05	20.05
1997	160.5	122.80	22.80
1998	163.0	124.71	24.71
1999	166.6	127.47	27.47

where π_{t+1} is the index of price change for the year $t+1$, P_t is the consumer price index for the year t, and P_{t+1} is the consumer price index for the year $t+1$. The results are shown in the table under the heading "Rate of inflation."

Looking again at the case of Masters, Pennsylvania, it is obvious from the calculation of constant dollar figures that the city is not doing as well with state and federal grants as was assumed. In fact, the inflationary forces in the early 1990s have eroded so much of the purchasing power that they caused an actual decline in the grant money in **real terms** (see Figure 8.2).

Figure 8.2 Comparison of State and Federal Grants, in Current and Constant Dollars



As you can clearly see, we can get radically different conclusions each time we transform the data. You can further transform the data by comparing it with other variables, such as population or the size of the city budget; the data may also be compared with a national or regional average, or the series can be looked at by its rate of increase. In each case, the data will tell us a different story. In a sense, the data are like a kaleidoscope, in which you can see a completely different picture by slightly changing the angle of the device. Let us consider some other ways of looking at the same information.

We can expand the data presented in Table 8.2 to include information on total city revenue during the period of study. If we express the dollar amount of grants received by the city as a percentage of its total revenue, we will be comforted by the fact that we have done well over the years (see Table 8.4). We can be further comforted if we look at the national trend of the ratio of state and federal government assistance to cities as a percentage of their total revenue. While external assistance to cities across the nation was going down in the 1990s, our city held its ground and was doing better than the national average during the latter part of the decade (see Figure 8.3).

Each of these transformations tells a slightly different story. Through them we get glimpses of different facets of the situation. Therefore, the question is not which one is telling the "true" story, but which one contains the most important message from the perspective of the inquirer.

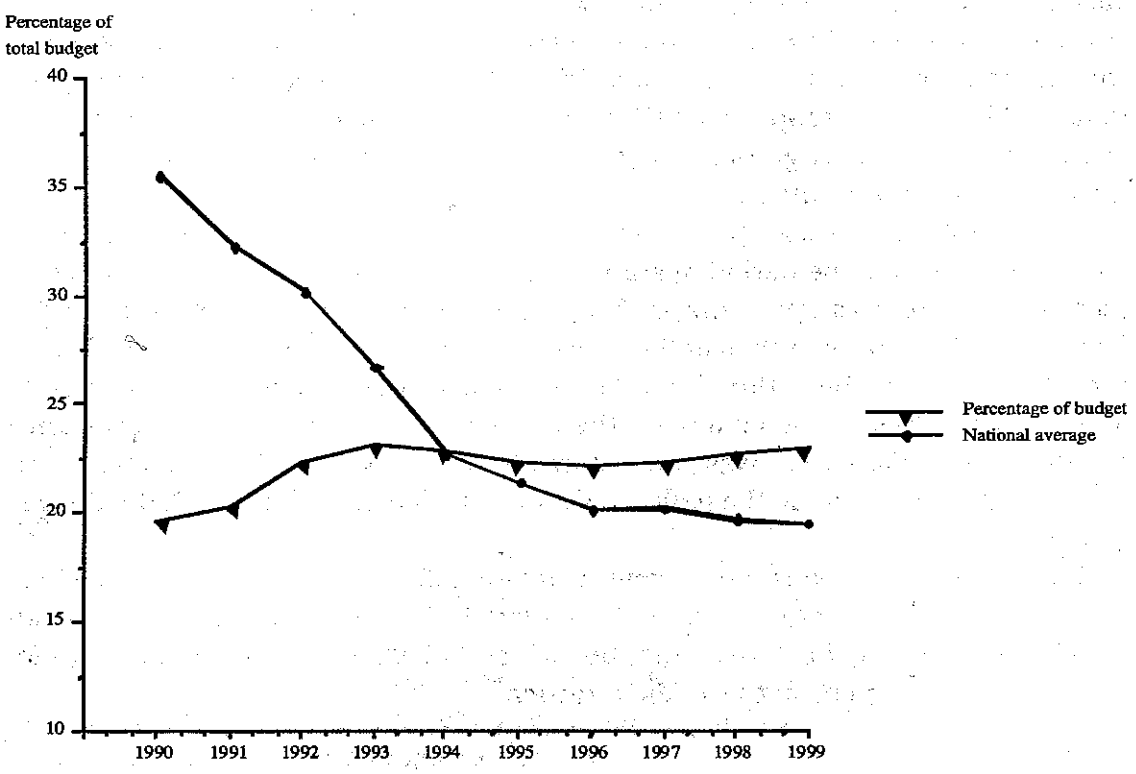
Percentage Change

You may want to look at the information in yet another way. You may calculate the yearly percentage change in constant dollar grants to Masters. This information, presented in Table 8.5 and plotted in Figure 8.4, can be quite useful in discerning year-to-year changes in state and federal assistance to the city.

Table 8.4 State and Federal Grants to Masters, Pennsylvania, as Percentage of Revenue, 1990–1999

Year	*State and federal grants in current $*	*Total government revenue in current $*	*State and federal grants as ratio of total government revenue*	*National average of grants as a percentage of local government revenue*
1990	71,000	360,000	19.7	35.5
1991	75,000	369,000	20.3	32.3
1992	85,000	382,000	22.3	30.2
1993	89,000	385,000	23.1	26.8
1994	91,000	398,000	22.9	22.8
1995	91,500	410,000	22.3	21.5
1996	93,000	419,000	22.2	20.2
1997	95,000	425,000	22.4	20.3
1998	98,000	432,000	22.7	19.8
1999	102,000	444,000	23.0	19.5

Figure 8.3 Comporison of State and Federal Grants as a Percentage of Local Government Tax Revenue



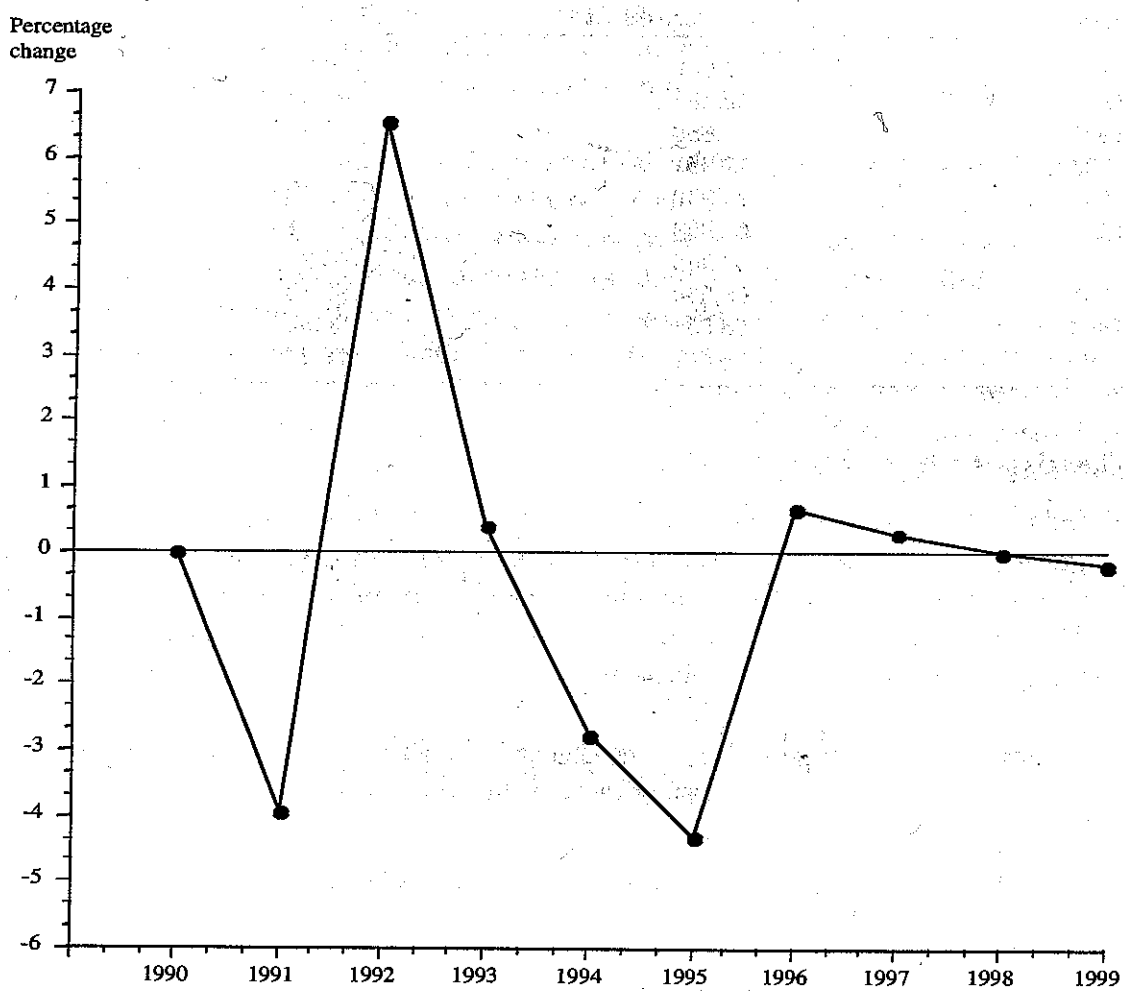
Creating an Index

Finally, it may be useful to look at the data with the help of an index. As noted earlier, an index is created when we take a particular figure as the **base** and then express the series in relation to this particular number. For example, if we take the

Table 8.5 Yearly Percentage Change in State and Federal Grants to Masters, Pennsylvania, 1990–1999

Year	*State and federal grants in constant $*	*Yearly percentage change (1990 = 100)*
1990	71,000	—
1991	68,182	-3.97
1992	72,650	6.55
1993	72,951	0.41
1994	70,930	-2.77
1995	67,883	-4.30
1996	68,345	0.68
1997	68,534	0.28
1998	68,531	0.00
1999	68,388	-0.21

Figure 8.4 Plot of Yearly Percentage Change in State and Federal Grants to Masters, Pennsylvania, in Constant Dollars



grants figure for 1990 as the base (expressed as 100), we can calculate the index by dividing each year's data by this number and then multiplying it by 100. That is,

$$\text{Index for 1991} = \frac{1991}{1990} \times 100$$

or

$$\frac{68{,}182}{71{,}000} \times 100 = 96.03$$

The data presented in Table 8.6 are plotted in Figure 8.5. As you can see, each presentation of the same information tells a slightly different story. Therefore, how you present your case will depend on your need.

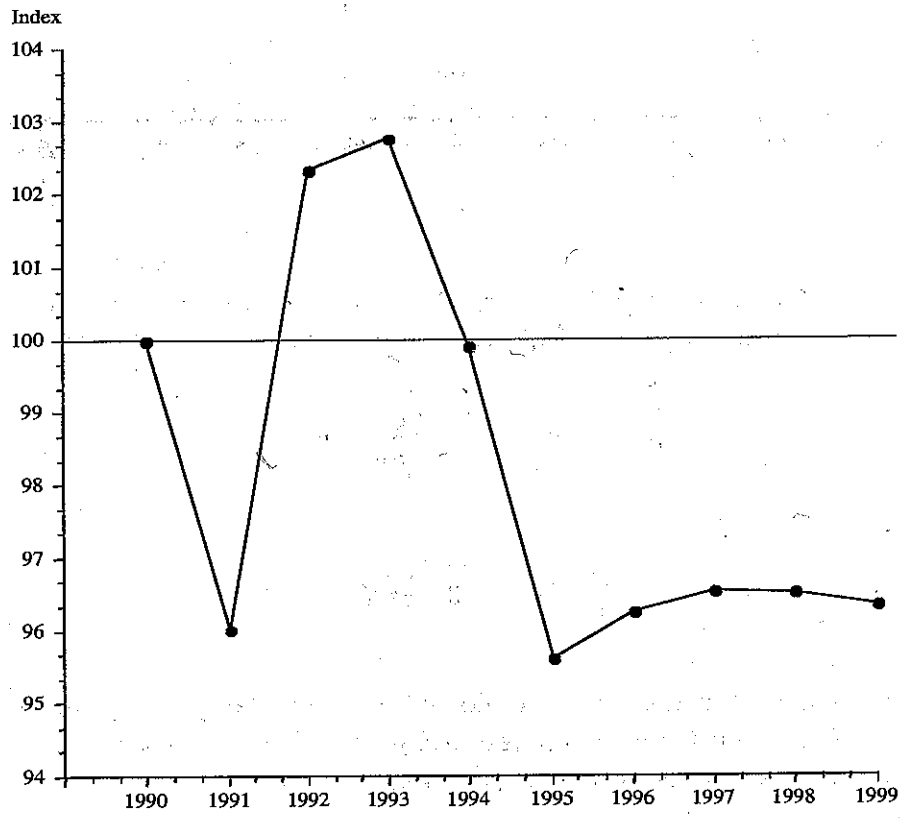
Table 8.6 Index of State and Federal Grants to Masters, Pennsylvania, 1990–1999

Year	*State and federal grants in constant $*	*Index (1990 = 100)*
1990	71,000	100.00
1991	68,182	96.03
1992	72,650	102.32
1993	72,951	102.75
1994	70,930	99.90
1995	67,883	95.61
1996	68,345	96.26
1997	68,534	96.53
1998	68,531	96.52
1999	68,388	96.32

Choosing the Type of Graph to Use

In the previous examples we used only line graphs. Today's managers use a variety of graphs, such as **scatter plots, bar graphs,** and **pie charts**. Each kind of graph presents the information in its own unique way. You should be familiar with each kind of graph and determine which kind of pictorial rendition gets your intended message across in the most effective way.

Figure 8.5 Plot of Yearly Percentage Change in State and Federal Grants to Masters, Pennsylvania, by Indexing



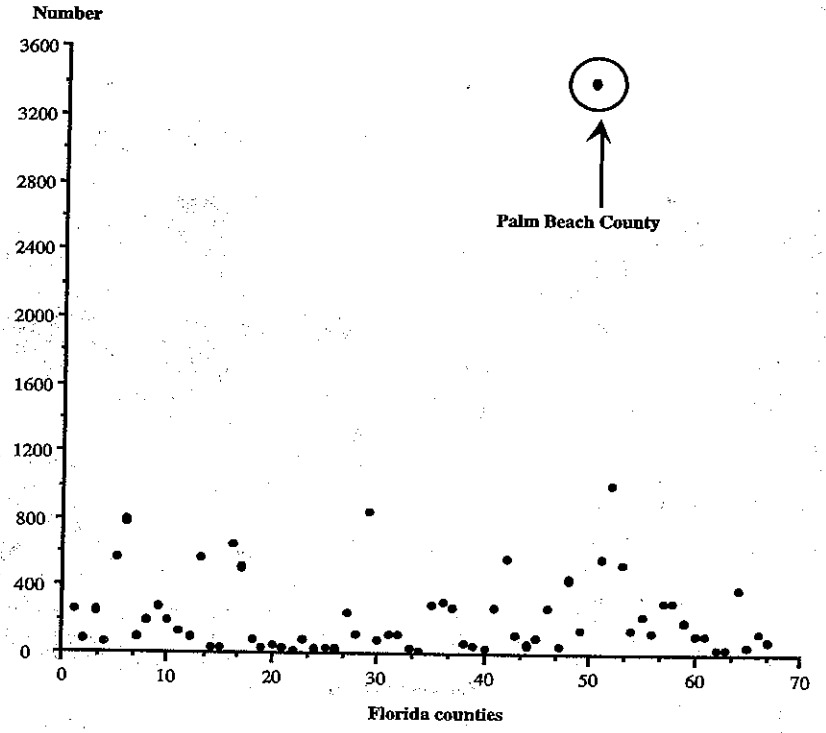
Specifically,

- Line graphs are used for time series data.
- If you have cross-section data, you should use a bar diagram or a scatter plot.
- If the data are in fractions of a total (or in percentages), use a pie chart.

During the now-famous presidential race of 2000, the two candidates, Republican George W. Bush and Democrat Al Gore, were running neck and neck. As the votes were tallied, Florida took on added prominence because the state's twenty-five electoral votes would determine the outcome of the election. Within the state of Florida, the voting of Palm Beach County residents generated an unexpected twist of irony. In the 1992 presidential election, Ross Perot, the candidate of the newly founded Reform Party, drew a respectable number of votes, which many thought gave the presidency to Bill Clinton. In 2000 the Reform Party candidate was Patrick Buchanan. In Palm Beach County, the election commission approved a "butterfly" ballot, a double-faced ballot on which the names of the candidates appeared side by side. The name Al Gore was opposite to that of Pat Buchanan, creating confusion among many Democratic voters. Many Gore supporters made a mistake and voted for Buchanan instead. When the votes were tabulated, Buchanan received 3,400 votes.

In Figure 8.6 I show the number of votes garnered by Buchanan in Florida's sixty-seven counties, before they were hand counted, as a scatter diagram. As you can see, the numbers clearly show an unmistakable spike for Mr. Buchanan.

Figure 8.6 Number of Votes Cast for Patrick Buchanan in the Florida Counties

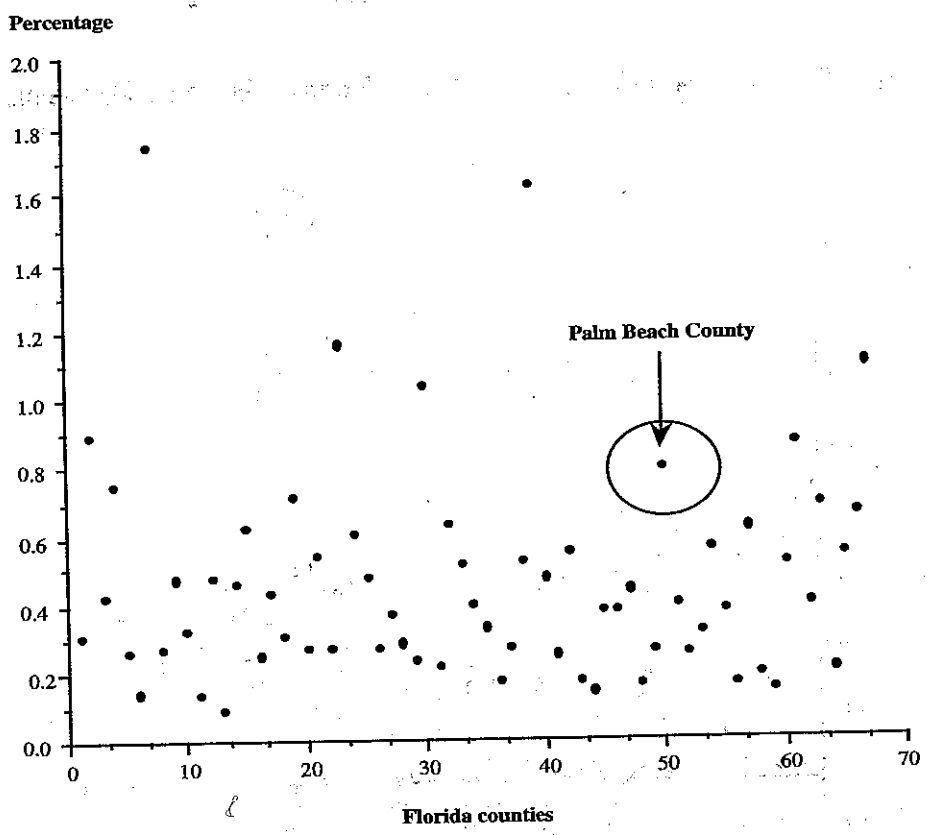


The scatter plot clearly identifies those Palm Beach Buchanan votes as an obvious outlier. However, in the social sciences, for every obvious solution, there is always a different angle. When we calculate the number of votes as a percentage of the total votes in each county, Palm Beach County's record does not appear so strange (see Figure 8.7). In fact, in terms of percentages, at least six small counties (Baker, Charlot, Indian River, Liberty, Suwannee, and Washington) polled higher than did Palm Beach.

Graphical Methods in Decision Making

Graphical presentations describe a situation by visual means. However, they should not be considered as passive tools of description; they can also be used as extremely powerful decision tools. Consider the following situation. The police department in Masters is trying to reach as many youngsters as possible to educate them about the perils of drug use. Last year the department spent considerable effort in arranging school appearances of officers and experts. It also advertised on local radio and television. Suppose last year the city spent $7,500 on school lectures, $13,000 on radio advertisements, and $20,000 on local television ads. A recent survey by the city shows that of the children who are aware of the

Figure 8.7 Percentage of Votes Cast for Patrick Buchanan in Florida Counties



city's drug prevention effort, 35 percent became aware of the issue through face-to-face contact with officers, 15 percent through radio ads, and the remaining 50 percent through watching television. In the two pie charts of Figure 8.8 I show the expenditure for each method of contact.

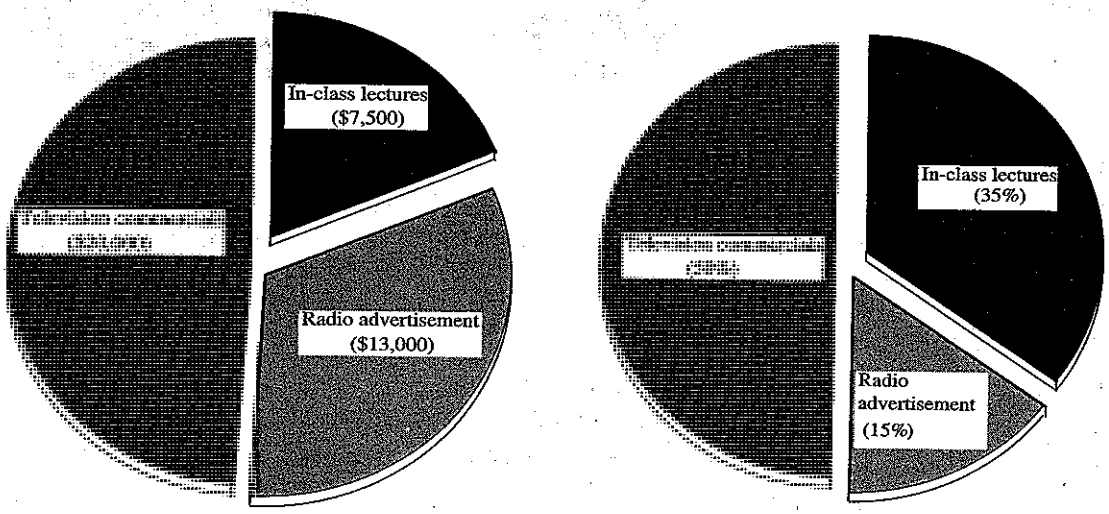
The side-by-side placement of these two diagrams visually demonstrates that face-to-face contacts are the most cost-effective way to disseminate drug prevention information to school-age children. The same information could have been presented as simple percentages in a table. However, as is often the case, a picture is worth a thousand words. For another example of how diagrams may enhance presentation of data, see "A Case in Point: Racial Profiling."

TO TELL THE TRUTH AND NOTHING BUT THE TRUTH

Over the years, statistics have been characterized in less-than-flattering ways. One of British prime minister Benjamin Disraeli's famous quips is, "There are three kinds of lies: lies, damned lies, and statistics." We have come to accept expressions such as "statistical artifacts" or "cooked-up statistics." We must recognize that deception, misunderstood implication, or the existence of a bias in the process of collecting information can cause problems.

The deceptive use of numbers must be defined with respect to the intent of the user. Thus, an individual or organization that puts out information knowing full well that the data have no real-life validity is defrauding or deceiving the user. During times of national emergencies or war, government agencies routinely use data for propaganda purposes. A nation at war may exaggerate or downplay claims about its military or industrial strength, or its war casualties, depending on its strategy. China, in the course of suppressing the prodemocracy movement, underreported the number of student casualties and released figures that were widely

Figure 8.8 Expenditure on Drug Prevention Program

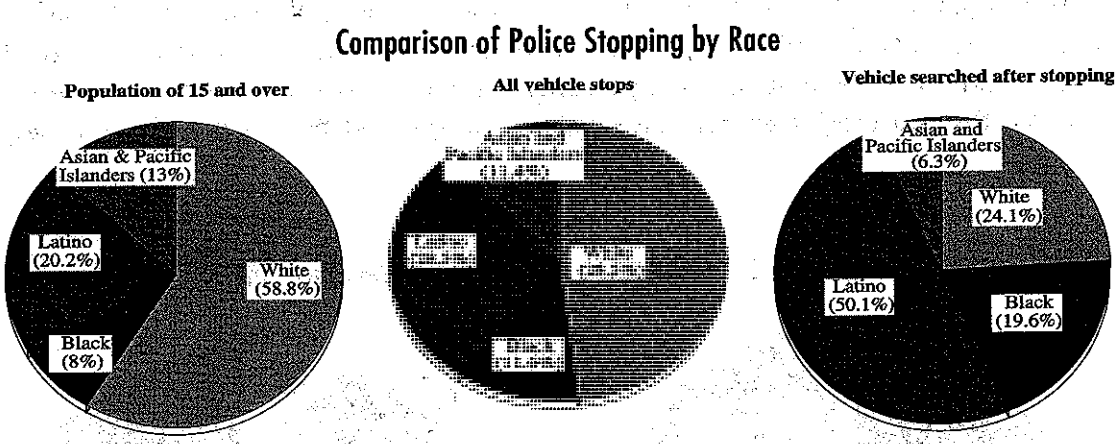


A CASE IN POINT

Racial Profiling

Many minority drivers have been protesting racial profiling. That is, police stop nonwhite motorists and search their vehicles at a disproportionate rate. This practice started the joke that these drivers are guilty of DWB—Driving While Black or Brown. The rising demands from San Diego's minority population prompted the city's police department to initiate a study. The study covered more than 90,000 instances of police stopping drivers between January and June of 2000. The *San Diego Union Tribune* published the data as pie charts (see charts below).[1]

The publication of this data created a big stir, particularly when they were presented visually in pie charts similar to the ones here. "We have known this [racial profiling] for a long time," said an African-American community leader. However, in support of police practices, San Diego's police chief, David Bejarano, pointed out that although Latinos make up only about 20 percent of the city's drivers, because San Diego is a border town, a large number of motorists come from Mexico. Further, he pointed out, the figures might be misleading because police are often deployed in areas close to the border and other high crime areas. Therefore, considering the entire city population does not allow for valid comparison. A more detailed study is required to control for the influence of these two external factors.



Note

1. Mark Arner and Joe Hughes, "Police Stops Blacks, Latinos More Often: Data from Profiling Report Echo Fears of S.D. Minorities," *San Diego Union Tribune,* September 29, 2000.

Discussion Points

1. How effective was the presentation of the data?
2. Can you think of a better way to present the information?
3. What are the points raised by Chief Bejarano? How would you respond to these points?

disputed by such knowledgeable people and agencies as Chinese student groups in the United States and Amnesty International. Similarly, the figures for North Vietnamese war casualties were routinely inflated by Pentagon officials during the Vietnam War.

Another source of contention frequently centers on the "true" implication of a statistic. We often use the per capita gross domestic product (GDP) as a measure of the relative economic development of nations. However, it is obvious that the word "development" used in a national context should imply more than a measure of a country's per capita GDP, because "development" implies a certain degree of progress and maturity in social, political, and economic institutions. Tiny oil-rich nations may have the highest per capita GDP, but one would be hard-pressed to characterize those countries as the most developed in the world.

Biases resulting from other factors may also cause a statistic to be misinterpreted. Valid questions have been raised about whether or not IQ tests measure relative levels of intelligence in children. For years, entrance to the U.S. Civil Service was based on the scores of a multiple-choice examination. However, it was eventually determined that such an examination was biased in favor of white, middle-class males. Therefore, the test score could not be accepted as the best measure of a candidate's suitability for a position.

Then there are data that, by their very nature, call for subjective judgment in the way they are defined and compiled. A good example is the consumer price index, discussed earlier in this chapter (see page 179). As mentioned earlier, a "typical" basket of goods and services that an "average" American consumes yearly is used to measure the rate of inflation. However, we know that each of us has a unique consumption pattern, based not only on our individual tastes but also on various factors such as age, income, race, and geographical location. If the price of skateboards goes up, senior citizens are less likely to be affected than are young people. Similarly, an increase in the cost of health care may not affect single young adults as much as the increase in the index would suggest. Whoever compiles this basket of goods thus faces two problems in making the index relevant to the majority of Americans. First, the compiler must discern what is "typical," in terms of what kinds of goods and services and at what level of consumption. Errors that severely distort the data creep in during the collection of information in many ways; not recognizing this means that the results of analyses may be meaningless, misleading, or even damaging.

You must be extremely wary of accepting data for analysis, and you cannot be too careful in looking at possible sources of biases and errors. At the same time, remember that it is impossible to find a perfect set of data in an otherwise imperfect world. Like the proverbial fastidious eater who dies of starvation, a researcher who is too cautious will know all the flaws of the data and its analyses without being able to draw any useful conclusions from them.

The rule of thumb, then, is to evaluate carefully the sources of bias in the data and be aware of the cost of doing an incorrect analysis. If you are conducting medical research for a new type of vaccine as an antidote for a disease, or calculating

trajectories for the reentry of a space shuttle into the earth's atmosphere, the margin of acceptable error is rather low. However, mercifully, in the areas of social science or public policy research the demand for numerical accuracy may not be that critical. It is most important to be aware of and open about the shortcomings of the data and the possible sources of bias in the analysis and interpretation.

Interpretation and Deception

The last source of skepticism to keep in mind is that to most people, numbers portray a rigid, self-evident truth. In a cocktail party discussion, a friend claimed that homosexuality was purely biological, since every society seems to have homosexuals as 10 percent of its population. It is fairly obvious that this statement is the kind designed to end all discussions, as it purports to present a totally scientific, incontrovertible fact of life. To many people, numbers pose an immediate threat because of their appearance of "scientific" objectivity. However, closer scrutiny will reveal problems resulting from the various biases we have described; and a significant source of disagreement may be that any information (numerical or otherwise) about a complex social situation is bound to be open to interpretation.

In 1954 Darrell Huff wrote an extremely interesting, humorous book, *How to Lie with Statistics*. In it he systematically demonstrated many ways to distort information to suit the purpose of the investigator. In his tongue-in-cheek introduction he states, "This book is a sort of primer in ways to use statistics to deceive. It may seem altogether too much like a manual for swindlers. Perhaps I can justify it in the manner of the retired burglar whose published reminiscences amounted to a graduate course in how to pick a lock and muffle a footfall: The crooks already know these tricks; honest men must learn them in self-defense."[2]

Huff's highly acclaimed book advanced understanding of the various ways one can use descriptive statistics among generations of undergraduate students. However, in all honesty, we may pose the question differently. If the manipulation of data is always suspected of "distorting" the picture, then there must be a truly undistorted version of real life. In other words, are we to assume the universality of truth? Does it always require a statistician to obfuscate an otherwise obvious situation? A famous early-twentieth-century Japanese play, *Rashomon*, by Ryunosuke Akutagawa, brings home the point of relativity of perception. In the play a bandit rapes a young woman traveling with her Samurai husband. A number of different individuals witnessed this terrible act of violence. When they are brought to the trial (including her deceased husband, who speaks through a medium), the incident is found to have variations of interpretation. As the play shows, there may be honest differences of opinion in the way one looks at a situation, even when expressed in "cold, hard, objective numbers." We live in a complex world in which "truth" may have more dimensions than can be effectively captured by any one-dimensional measure. However, if we use multiple indexes to characterize a situation, our cognitive limitations stand in the way of formulating any definitive picture. Like everything else in life, quantification of social phenomena requires a trade-off between the confusion of a total picture and the clarity a limited view offers.

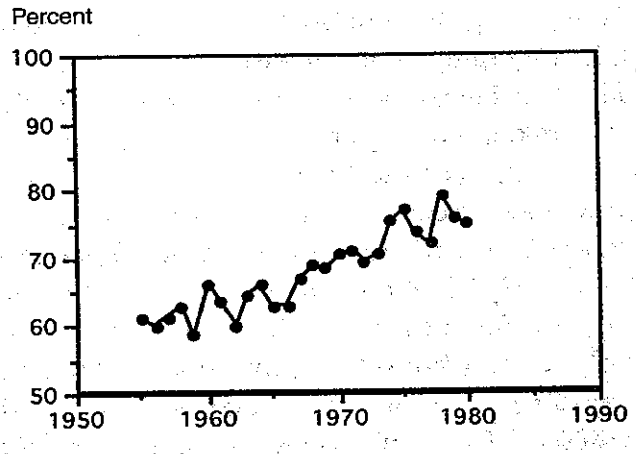
For example, consider our hypothetical city of Masters, Pennsylvania. The demographic composition of Masters is typical of the region, with a large number of working-class people, along with pockets of urban blight, characterized by persistent levels of high unemployment. However, a few areas of the city house extremely wealthy families. Let us focus on three individuals plying their trade in Masters: a real estate broker, a college professor, and a city planner. All three of these individuals want to present the "true" economic picture of this city with a single number—the average income of the city's people. However, these three have different objectives. The real estate broker wants to portray the city as a nice place to live and raise a family. Therefore, in talking to clients he mentions as "average" the mean income of the residents of the town. However, although small in number, the extremely wealthy households influence the mean. The prospective buyer gets a much rosier picture of the average affluence of the city than that espoused by the professor. The professor is conducting research in urban economics, for which he is using the figure of median income. The median, the middle income from the highest to the lowest, presents a less attractive picture of the economic well-being of the city because it is not affected by the presence of the wealthy sector of the community. However, even this number is far superior to the one used by the city planner. The city planner of Masters wants to respond to a request for a grant proposal from the state government to bring in money earmarked for the economically depressed areas. For this proposal, she uses the modal income of the town, which is the most frequently found income of the inhabitants.

Is it possible to pick out which of the three individuals, who use three different measures of average income owing to their different objectives, is engaged in an act of deception? I would argue that none of them can be accused of such an act unless some other kind of deception is present. When it comes to the definition of "average," most people intuitively use the arithmetic mean, median, and mode, in that order. Therefore, by convention, if one uses the term "average" for the mean, one can feel justified. The use of the median may require justification, and the use of the modal income would certainly require its mention in the report, to be ethically fair and aboveboard. However, the use of any of these measures cannot be called a deception. Therefore, we must conclude that without the intent of deception none of the figures can be characterized as a lie; there can be honest difference of opinion, even among those whose business it is to deal with numbers, as to which one of these three represents the most valid picture of the city.

Another source of bias, Huff claimed, comes from the deceptive use of pictorial information—graphs. Because a picture is often worth a thousand words, the desire to convey information by graphical means is rather strong. But in the process one might take advantage of certain trickery. Consider the example in Figure 8.9, where we have depicted nonwhite unemployment as a percentage ratio of white unemployment. In 1955 the unemployment rate within the nonwhite population was 62 percent higher than that within the white population.

Does this presentation of the information suit your needs, or do you want to portray more dramatically the plight of the minority population in the United

Figure 8.9 Ratio of Nonwhite to White Unemployment Rate



Source: U.S. Bureau of Labor Statistics.

States? If you do, you can s-t-r-e-t-c-h the graph for added visual effect showing the same information (Figure 8.10). Obviously, in this case the difference between nonwhite and white unemployment is portrayed in a much more striking fashion.

What if you contend that the situation for the minority population is really not that bad, or that the situation has not changed appreciably over the years? In that case, you can use another trick. You can increase the range of the vertical axis, which allows you to present the same information in a different light. Against a much wider range of possible ratios of unemployment, the ethnic difference in the relative measure of economic deprivation does indeed look small (Figure 8.11).

Figure 8.10 Ratio of Nonwhite to White Unemployment Rate (stretched graph)

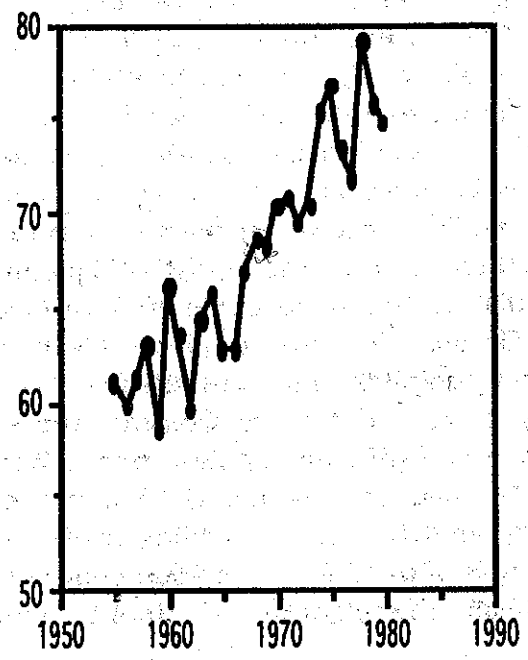
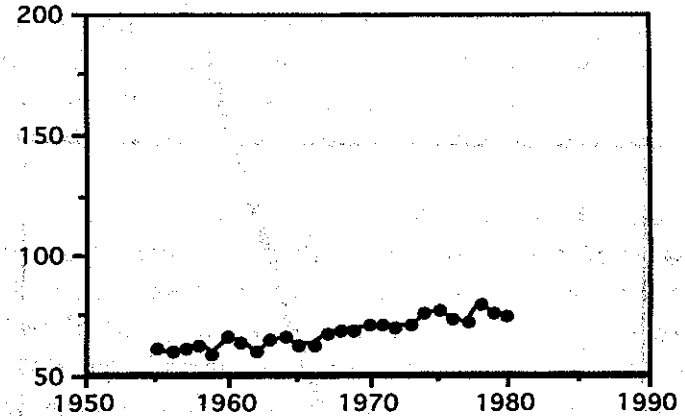


Figure 8.11 Ratio of Nonwhite to White Unemployment Rate (elongated vertical axis)



Yet another effective way of representing a series is to use selective years for comparison. For example, you may want to present the same information contained in the three graphs, but even more emphatically. You may want to show them in either of the following ways without "lying" with your statistics. The two graphs in Figure 8.12 show only three years—1960, 1970, and 1980. This restriction removes the distracting effects of yearly fluctuations and allows us to present long-term trends. Then, by simply manipulating the vertical axis, we have two radically different visual effects.

Now that I have shown you various ways of presenting the same information, which do you think represents the "true" picture? The answer is simple: We do not know which of these diagrams would be classified as a deceptive representation of the reality. However, Figures 8.11 and 8.12 (a) might be interpreted as edging toward questionable practices, since the vertical axes in those figures have wider ranges than is required by the data. But would you call that lying? In real life, truth, like beauty, lies in the eye of the beholder.

In the preceding examples, the difference between interpreting and deceiving might have been subtle, but consider the rendition of the same information in Figure 8.13. In this diagram, you are not only presenting facts but also trying to make

Figure 8.12 Ratio of Nonwhite to White Unemployment Rate (1960, 1970, and 1980)

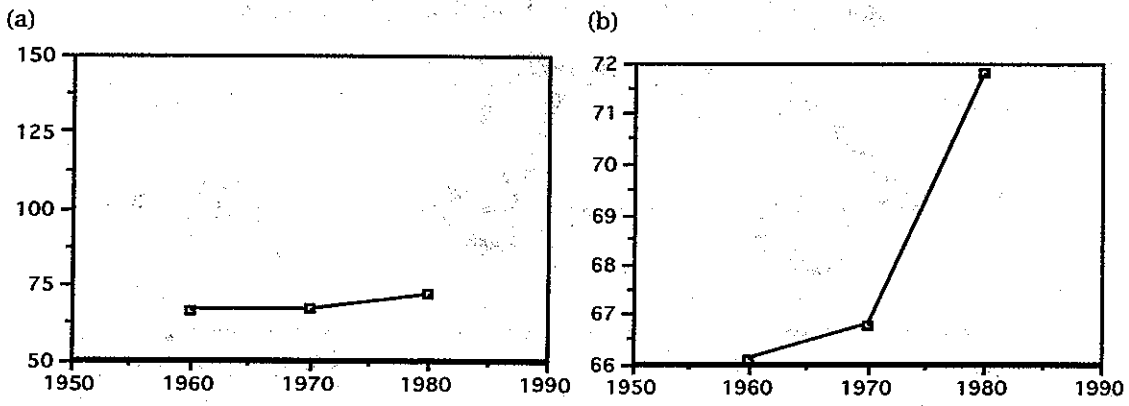
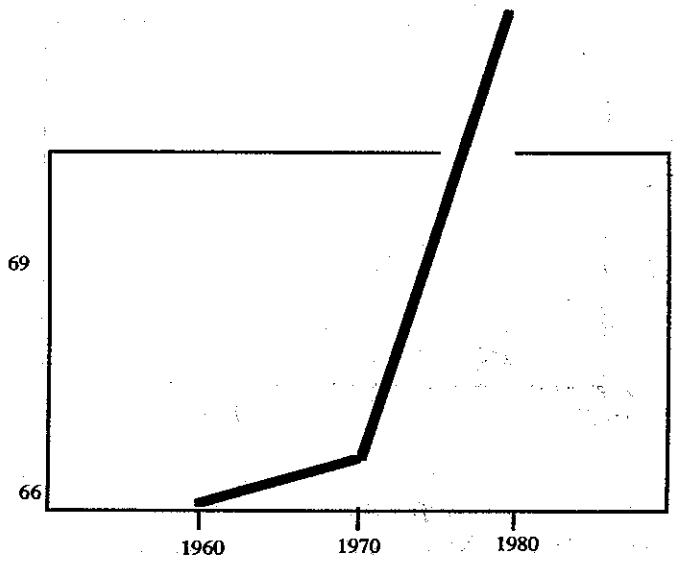


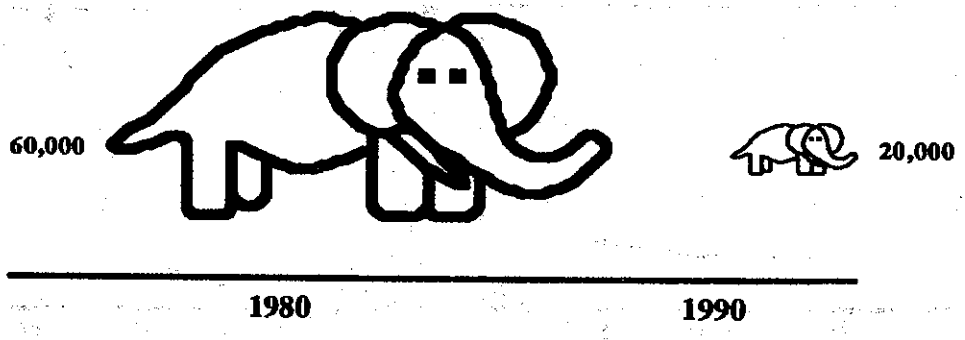
Figure 8.13 Ratio of Nonwhite to White Unemployment Rate (shortened vertical axis)



a rather loud statement with the obvious analogy of the ratio going through the ceiling.

Huff pointed out that, intended or unintended, deception may creep in more readily when one uses graphics instead of lines and bars, which are drawn according to scale. Consider a diagram showing the plight of African elephants (Figure 8.14). It is shocking to realize that in ten years the elephant population has diminished to one-third of its 1980 size. Although the numbers themselves speak volumes about the plight of the hapless pachyderms in Africa, the pictures in the figure are not drawn according to scale. We catch the relative size difference in the picture more readily than the difference in magnitude between the numbers, but we may not be able to catch that the 1980 drawing is much more than three times the size of the 1990 drawing. These kinds of disproportionate pictorial renditions are popular with those who present data at legislative hearings and in administrative decision-making sessions because the message is direct and dramatic. With

Figure 8.14 African Elephant Population, 1980–1990



computer graphics so easy to use, the possibility of some creative minds deceiving the unwary has never been greater.

The power of strong visual presentation is well recognized and well documented. A 1985 *Wall Street Journal* story documents how the use of such a tool helped Caspar Weinberger, then secretary of defense, to avert a cut in the Pentagon budget by David Stockman, director of the Office of Management and Budget.

> In staving off Mr. Stockman's assault on the planned buildup, Mr. Weinberger turned to a tactic for which he has since become famous, the chart and easel. The defense secretary's charts, presented in a meeting with President Ronald Reagan, showed large soldiers bearing large weapons, which were labeled "Reagan budget." They towered above small soldiers with small weapons labeled "OMB budget." President Reagan went along with the "Reagan budget."[3]

Tabular Presentation of Data

A long series of numbers pushes us to the limits of our cognitive capacities. For that reason, the value of an effective table as a means of communication cannot be overstated. Presenting data in useful tabular form is an art, which can be perfected only through practice.

The preparation of an effective table requires considerable thought and time (and therefore, money). First, have a clear idea of exactly what you want to communicate to the reader. Second, choose a title that parsimoniously describes the contents of the table. Third, consider various ways of presenting the raw data so that they make the point you want to make most effectively.

For a table to be useful, its purpose must be absolutely clear from its title. The title should be concise, but not so brief that it does not convey the true intent of your presentation. The numbers, taken together, must tell a coherent story. Consider, for example, Table 8.7. You may notice how much information has been packed into a concise table. You can find out from the table how much money the government spent on selected items in 1995. During that fiscal year, education accounted for 12.10 percent of all public spending, and local governments paid the bulk (68.24 percent) of education costs. The table also tells you that the federal government originated half (50.87 percent) of total government spending.

In my discussion of graphical presentation, I have shown you how to look at a data set from different perspectives. Similarly, before preparing a table, you may draw interesting conclusions by using absolute numbers, ratios, percentages, and so on.

THOSE NOT-SO-INNOCENT NUMBERS

It is relatively easy to define outright deception or lying by the measure of intent and the sheer fabrication of data, but the line between deception and differences in interpretation is murky. Often the intentions of presenters are not obvious; nor

Table 8.7 Spending by Function and Level of Government, 1995

Function	Spending by level of government (in million $)			PS (percent)	Percentage of spending by level of government			
	Federal	State	Local		Federal	State	Local	Total
Defense	327,231	—	—	9.60	100.00	—	—	100.00
Education	27,270	101,510	276,763	12.10	6.72	25.03	68.24	100.00
Highways	731	48,893	30,216	2.38	0.92	61.24	37.85	100.00
Welfare	57,246	160,421	32,669	7.47	22.87	64.08	13.05	100.00
Police	7,563	5,735	52,329	1.96	11.52	8.74	79.74	100.00
Health	26,517	49,487	56,460	3.95	20.02	37.36	42.63	100.00
Administration	19,416	24,781	35,237	2.37	24.44	31.20	44.36	100.00
Insurance	558,291	93,692	13,648	19.86	83.87	14.08	2.05	100.00
Other	680,857	402,563	262,046	39.93	50.60	29.92	19.48	100.00
Total	1,705,122	887,082	759,368	100.00	50.87	26.47	22.66	100.00

Source: U.S. Department of Commerce, Bureau of Census and Tax Foundation, Washington, D.C.

Notes: PS = percentage of total public spending. Some percentages may not add to one hundred because of rounding.

are we capable of detecting purposeful contamination of data. Because we tend to believe in the objectivity of numerical information more readily than in the subjectivity of qualitative statements, deceptive use of statistics can bring incredible misery to people.

In 1896 Frederick L. Hoffman, a nationally famous statistician for the Prudential Insurance Company of America, wrote a book titled, *Race Traits and Tendencies of the American Negro.*[4] Hoffman's thesis was that since their emancipation, African Americans (having left the protective care of their slave owners) had gone back to their "basic racial trait" of "immorality of character." Hoffman based his theory on a number of different statistics that he had collected. He noted that in 1890 there were 567 blacks in prison for rape, which constituted 47 percent of the prison population convicted on rape charges. Because this number was significantly greater than the proportion of the African-American population (about 10 percent at the time), according to Hoffman, rape and other sexual crimes were reflective of the "Negro racial trait." Hoffman thus concluded that "[a]ll the facts brought together in this work prove that the colored population is gradually parting with the virtues and the moderate degree of economic efficiency developed under the regime of slavery. All the facts prove that a low standard of sexual morality is the main and underlying cause of the low and anti-social condition of the race at the present time."[5]

Hoffman then connected the "Negro racial trait of immorality" to the high mortality rate among the black population. On the basis of this causal linkage, disregarding the fact that the census of 1890 showed a steady increase in the size of the black population, Hoffman predicted that African Americans were doomed to face a "gradual extinction of the race." The name of Hoffman's publisher, the American Economic Association, added a dose of respectability to this statistical study, which was widely used as a weapon in promoting white supremacy for decades to come. However, another important consequence of this and other internal statistical studies was that Prudential judged blacks to be bad actuarial risks and promptly started to cancel all black life insurance policies. Within four years, by the end of the century, most insurance companies got out of the business of insuring African Americans.[6] In a similar manner, statistics have been used over the years to perpetrate many kinds of heinous crimes, or their faulty uses have led to extremely inefficient public policies.[7]

STRUCTURE ABOVE A SWAMP

The discussions in this and the previous chapter may be confusing to you. On the one hand, I emphasize the relative nature of truth and, on the other, I advocate objective analysis. A quotation by Karl Popper, the eminent philosopher of science, may resolve this contradiction.

> The empirical basis of objective science has thus nothing "absolute" about it. Science does not rest upon rock-bottom. The bold structure of its theories rise, as it were, above a swamp. It is like a building erected on

> piles. The piles are driven down from above into the swamp, but not down to any natural or "given" base; and when we cease our attempts to drive our piles into a deeper layer, it is not because we have reached firm ground. We simply stop when we are satisfied that they are firm enough to carry the structure, at least for the time being.[8]

Therefore, although we can empirically test "scientific laws" regarding society, we cannot know their truth. The ability to test hypotheses has lent social sciences and policy science a considerable degree of credibility. Hence, on this shifting ground of "truth," we want to achieve objective analyses by being systematic in our definition of goals, consistent about our method of analysis, and forthright about our implicit assumptions.

In psychology it is often held that the strength of one's character can also be the source of one's weakness. Similarly, the appeal of objective methods of policy analysis is the ability to present complex phenomena with simple, easy-to-understand numbers and figures. At the same time, the unquestioned acceptance of these statistical artifacts can lead to serious flaws. Therefore, we should know how to use with skill the extremely useful and powerful tools called statistics. This skill is honed with practice and by knowledge of the methods of manipulation.

Key Words

Bar graphs (p. 184)
Base (p.182)
Consumer price index (p. 179)
Fixed weight (p. 179)
GDP deflator (p. 179)
Nominal income (p. 179)
Pie charts (p. 184)
Real income (p. 179)
Real terms (p. 180)
Scatter plots (p. 184)
Variable weight (p. 179)

Exercises

1. Write an essay on truth and objectivity in quantitative analysis for public policy. Within this context, describe the relative advantages and disadvantages of the various measures of central tendency and dispersion. Provide appropriate examples.
2. Collect data on the growth rate of per capita GDP, rate of inflation, and unemployment from 1950 (consult *Economic Report of the President* and *Statistical Abstract of the United States)*. First make the case that the nation has been better served by the Republican presidents, and then make the case for the Democrats based on the same set of data.
3. Consider the following table, which shows yearly data on percentages of the population living under poverty. Plot the data and derive as many different (and even contradictory) conclusions as you can from them.

Year	All	White	Black
1959	22.4	18.1	58.2
1960	22.2	17.8	56.4
1961	21.9	17.4	56.8

Year	All	White	Black
1962	21.0	16.4	56.1
1963	19.5	15.3	51.1
1964	19.0	14.9	49.8
1965	17.3	13.3	47.1
1966	14.7	11.3	39.7
1967	14.2	11.0	38.2
1968	12.8	10.0	32.8
1969	12.1	9.5	30.9
1970	12.6	9.9	31.6
1971	12.5	9.9	31.3
1972	11.9	9.0	32.4
1973	11.1	8.4	29.3
1974	11.2	8.6	29.7
1975	12.3	9.7	29.8
1976	11.8	9.1	29.5
1977	11.6	8.9	29.0
1978	11.4	8.7	29.4
1979	11.7	9.0	28.1
1980	13.0	10.2	29.9
1981	14.0	11.1	34.2
1982	15.0	12.0	35.6
1983	15.2	12.1	35.7
1984	14.4	11.5	33.8
1985	14.0	11.4	31.3
1986	13.6	11.0	31.1
1987	13.5	10.5	33.1
1988	11.6	8.8	29.1

4. Look at some of the recent news reports presented with quantitative data. See if you can derive a different conclusion from the same information.

Notes

1. You can access data collected by the Bureau of Labor Statistics by logging on to http://stats.bls.gov/blshome.htm.
2. Darrell Huff, *How to Lie with Statistics* (New York: W.W. Norton, 1954), 9.
3. Tim Corrigan, "Weinberger Finds His Well-Worn Strategies Always Succeed in Blunting Defense Budget Axe," *Wall Street Journal*, March 1, 1985.
4. Frederick L. Hoffman, *Race Traits and Tendencies of the American Negro*, quoted in Joel Williamson, *The Crucible of Race: Black-White Relations in the American South Since Emancipation* (New York: Oxford University Press, 1984), 329.
5. Ibid.
6. For a detailed discussion, see Williamson, *The Crucible of Race.*
7. For an excellent discussion of measuring people's abilities with numbers, see Stephen Jay Gould, *The Mismeasure of Man* (New York: Norton, 1981).
8. Karl Popper, *The Logic of Scientific Discovery* (New York: Harper and Row, 1959), 65.