ANNUAL REVIEWS Further
Quick links to online content

Ann. Rev. Public Health. 1987. 8:309–53
Copyright © 1987 by Annual Reviews Inc. All rights reserved

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

GRAPHICAL METHODS IN STATISTICAL ANALYSIS

Lincoln E. Moses

Department of Statistics, Stanford University, Stanford, California 94305

INTRODUCTION

The idea of graphical presentation of statistical data is a relatively recent development as compared with geometry, algebra, or even probability theory (18). The techniques of graphical presentation of data are still rapidly developing, as statistical theory opens new ways of thinking about data, and as using the capabilities of modern computers increasingly reshapes the body of statistical practice—and theory. Before turning to some of the key graphical methods it is well first to look at a few broad issues.

Remarks About the Nature of Graphs

The usefulness of graphical presentation arises partly from the quantity of information that can be displayed compactly. After all, it is easy to allow a pair of numbers to be depicted by a single dot placed suitably on a piece of graph paper; thus, ten dots may depict ten such number pairs in a fashion that makes it easy to compare them and study patterns among them. Observe, ten dots use less ink than even one word; so information can indeed be displayed compactly by graphical means.

Vividness is another source of the appeal of graphical methods. Interesting pictures can replace dry numbers. This feature can help in transmitting information; unfortunately it can also sometimes be harmful. First, distortion can arise simply from perceptual short-cuts that the mind takes, and second, it is not unknown for distortion to be deliberately attempted.

The graphical display carries two kinds of information: the data themselves; and the descriptions of the data, such as labels, scale markers, and the title. With regard to both kinds of information a balance is needed between too little and too much. The temptation to pack a great deal of statistical information

0163-7525/87/0510-0309$02.00

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

into a single chart arises again and again. There can be good reasons to bring many facts together in the same picture: relationships, comparisons, contrasts may be seen easier. But too much information can baffle the eye, boggle the mind, or both. Similarly, thc ink dcvoted to labels, scales, etc can be absolutely necessary, or can be redundant, leading to clutter and even confusion.

Graphs, from Three Perspectives

THE READER Reading a graph demands attention; some systematic patterns of proceeding can be helpful. First, to read the graph, inspect the title, the source of the data, the scales on the margins of the figure, and the labels of any symbols that are used. After these steps it is time to look at the data themselves. If the graph depicts numbers that are listed or tabled in a convenient place it is usually helpful to check thc correspondence between the two representations, tabular and graphical; the aim is not so much to check for error as to make certain that one comprehends correctly just how the numerical data are graphically displayed. Second, to *interpret* the graph, begin with the author's interpretation. Do you understand the basis for it? Does the graph actually support that interpretation? Are there other reasonable interpretations? If the data have large uncertainties, would the writer's interpretation lose credibility or remain reasonablc? If some one particular data point were in error, would the interpretation be strongly affected, or would it still be reasonable? All these questions illustrate the more general notion: *Study the graph*, but after first studying its labels, scales, sources, and title.

THE AUTHOR In preparing a statistical graph one must keep in mind two separate concerns: (*a*) the data, and representing them graphically; (*b*) the intended reader's ease of correctly understanding the resulting graph.

During preparation, the author must choose wisely and explain well the elements of the graph: title, scales, symbols, source(s). Further, he may need to balance the simplicity that is to be had in each of many separate charts, against the gain in fuller understanding that may be available by showing several related things on a single, more complex chart. Further on in this article some ideas bearing on such choices are discussed.

Planning the chart may lead to a much better product; such planning can often gain advantage from some measure of experimenting; thus, alternative ways of graphing a given data set can be executed and tested on one's friends or colleague during the preparation phase.

THE RESEARCHER The researcher may become the author, but before then there can be much to gain by inquisitively graphing the data—typically in

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

several alternative ways. Perhaps only a few of these graphs will see the light of day, after serving their purpose of increasing the researcher's understanding. The tasks here include:

1. Finding suitable levels of aggregation, that is, identifying which subsets of data can be collapsed and combined, and which cannot.
2. Exploring for the relevance of possible interfering variables: Do the data from different interviewers look sufficiently similar? Are subjects with a previous history of disease X different from those without it in our study, which concerns disease Y?
3. Choosing the *scales* on which variables are to be expressed: Should we use travel time? Or its reciprocal, the velocity? Are patterns clearer when log Y is used or Y itself?
4. Assessing the impact of statistical uncertainty on data interpretation, and deciding whether and how to depict the uncertainty.

We have pointed to three roles in which a person may approach a statistical graph: reader, author, and researcher. But in any of these roles, broadly similar issues, ideas, and problems will be met. We turn now to some of these, often addressing the researcher-author, but believing that the reader of graphs can also gain from these considerations.

PRESENTING UNIVARIATE DATA

If each subject of study produces a single measurement, we have univariate data. Additionally, it is not uncommon to acquire information about several variables from each subject under observation: Perhaps several laboratory tests are routinely taken, or perhaps two or three items are reported in the typical pathology report. Any *one* of these variables can be the subject of a *univariate* display in which information concerning the *one* variable is to be graphically depicted, for one or more groups of subjects.

Univariate data can vary continuously, as does weight, or elapsed time; it may have only a few separate (discrete) possible values, like number of children born alive to a woman; it may be scaled in terms of ordered categories, like poor, fair, good, excellent; it may be binary, each observation having one of only two possible values, as with a serological test that can only be positive or else negative. Graphical methods appropriate for these cases vary in some respects, by necessity, as we shall see.

The Dot Diagram

If the number of observations is not very great, say at most a few dozens, then the dot diagram can be very useful to display the data, for one sample, or for several. In Figure 1, the top two panels show one-sample dot diagrams for

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

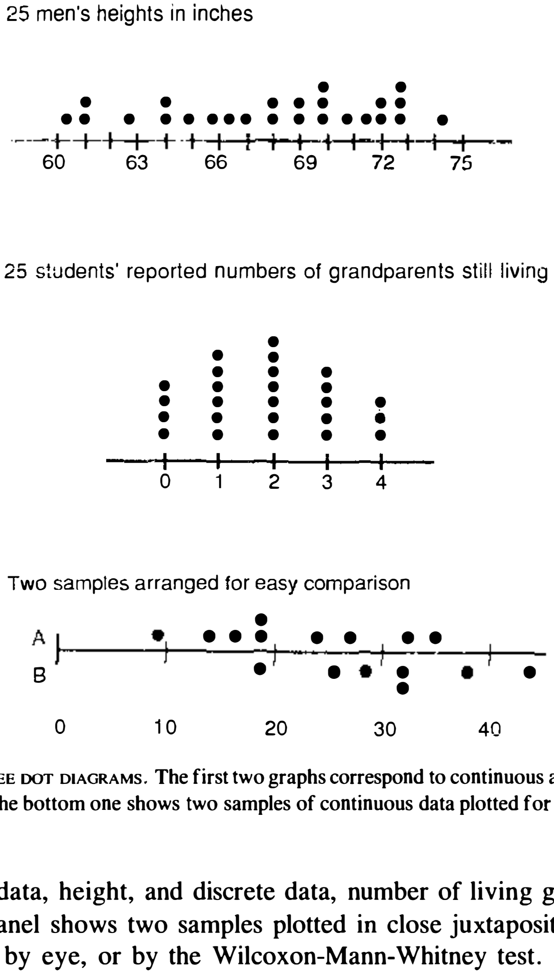


Figure 1 THREE DOT DIAGRAMS. The first two graphs correspond to continuous and discrete data, respectively. The bottom one shows two samples of continuous data plotted for ease of comparison.

continuous data, height, and discrete data, number of living grand parents. The third panel shows two samples plotted in close juxtaposition, for easy comparison by eye, or by the Wilcoxon-Mann-Whitney test. If there were more than two samples to compare, then it would be better to show each on a separate dot diagram, with carefully aligned scales either stacked one above the other or vertically arrayed side by side.

The dot diagram's advantages are ease of construction, ease of interpretation, and precise visualization of the measurements as actually made, because no grouping is imposed on the observations.

The Histogram and a Close Relative

The histogram is useful for data that are quite numerous—more than a few dozens. A system of (usually) equal-length intervals is imposed on the scale, and the number of observations (frequency) belonging to each interval is

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

represented by the height of a bar erected above that interval. In Figure 2, the data from the upper panel of Figure 1 appear twice in histogram form. The left-hand panel shows the histogram that results from using intervals $60^+ - 63$, $63^+ - 66$, . . . and the right hand panel results from the system of intervals $59^+ - 62$, $62^+ - 65$, ... With larger samples, two such interval systems ordinarily produce less dissimilar results. Notice, each histogram has two vertical scales. The one on the left shows frequencies; the heights of the bars, using that scale, add up to 25, the total number of observations. The right-hand scale shows decimal fractions, with .20 at the same elevation as 5, because 5 is .20 of the total sample size; using this scale the sum of the heights of the bars is 1.00. With respect to the right-hand scale, we speak of the picture as a *relative frequency* histogram.

Histograms are natural ways to depict large samples of continuous data. They are also natural for depicting ordered-category data; responses like "improved" or "much improved" embody a range of possible degrees of intensity (like improvement), so letting an interval represent such a category is reasonable (although the interval *widths* may seem problematic).

But if the data are discrete, as with number of living grandparents, then it is more natural to portray the frequencies as spikes at the (only) possible values, 0, 1, 2, 3, 4. Such a figure may be called a "spike diagram." The upper panel of Figure 3 shows the grandparent data in this format.

Histograms are not well adapted to comparing two or more samples. If it be attempted, then let the system of intervals for the two histograms be the same, and plot them as *relative* frequency histograms, so that both pictures will have the same area, facilitating comparison. But two such pictures can be hard to compare by eye, being separated from one another. And if they are superimposed they tend to look tangled unless precautions are taken. Figure 4 offers several ways of displaying two comparable histograms. The reader might try an additional method, such as placing one histogram directly beneath the other.

The superimposed version takes liberties with the data; it represents the data of group I as distributed over the various intervals (as histograms always do), but to avoid tangling, it depicts the data from group II as being concentrated at the interval midpoints. This is the price paid for what may be the easiest visual comparison.

Samples of categorical data (like race or blood-type) for which there is no underlying order to the categories, are sometimes depicted by means of "pie charts," in which a disk is partitioned into segments proportioned (angularly) to the frequencies of the categories. Considerable evidence (4, p. 264) indicated that this representation is an inferior one; the eye is not clever at interpreting angles accurately. It is preferable to use spikes adding to 1.00 for the various categories, as in the lower panel of Figure 3, where the relative

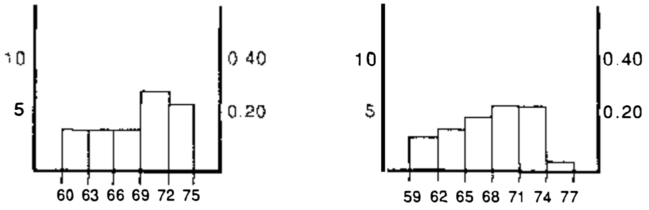


Figure 2 TWO HISTOGRAMS OF THE SAME DATA. The two figures use different nets of intervals to capture the data in the top panel of Figure 1.

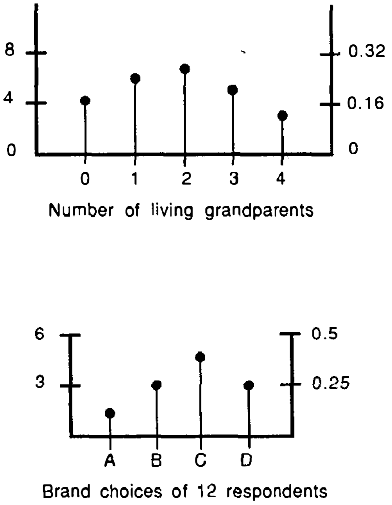


Figure 3 TWO SPIKE DIAGRAMS. The upper panel shows the living grandparent data of Figure 1 and the lower panel shows the preferences for four brands as reported by 12 respondents.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

frequencies of four categories A, B, C, and D are shown. (The horizontal axis has no meaning in this picture; it only serves to start each spike form the same bottom level.)

Cumulative Plots

Histograms plot frequencies, or relative frequencies, belonging to intervals. The same information can be rendered, without loss of information, in cumulative form, as in Table 1.

The table shows the relative frequencies for two samples, as depicted in Figure 4, and then it shows those frequencies *accumulated.* Thus, the 92 found in the table tells us that 92% of the observations in sample I had values of 75 or less. We have offset the cumulative relative frequencies to a level lower than the corresponding relative frequencies to emphasize that those numbers relate to different points on the numerical scale. For example, in sample I the interval $63^{+}-66$ shows 15 in the first column, reporting the

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

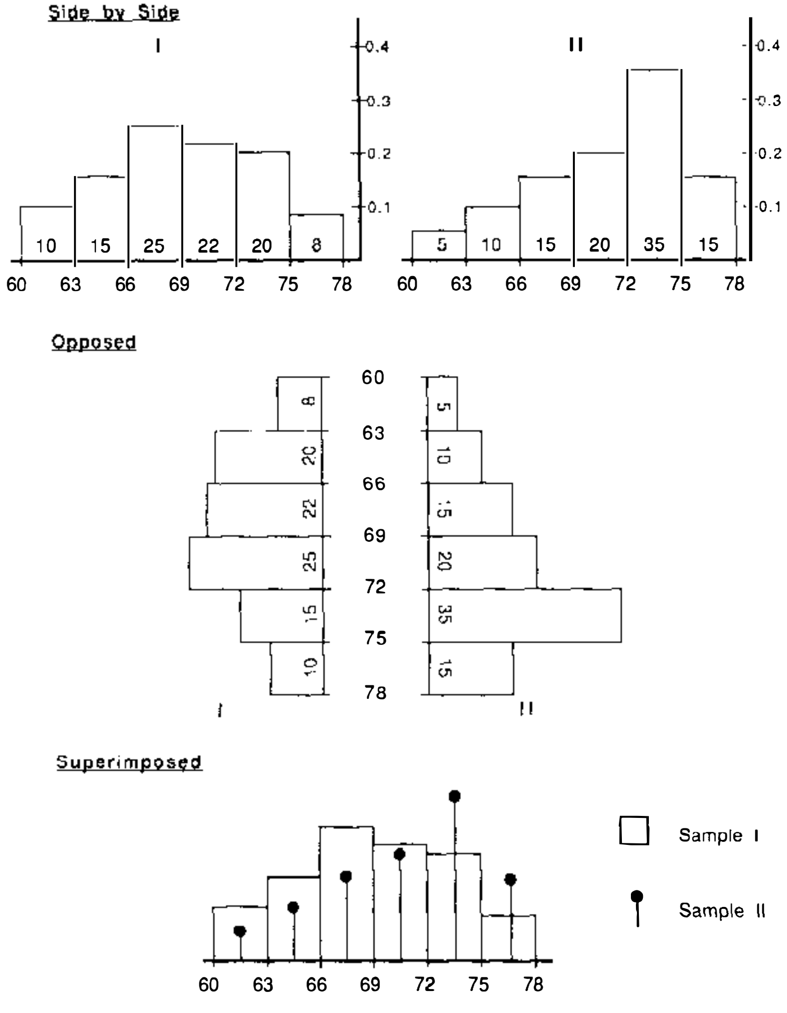


Figure 4 THREE WAYS TO SHOW TWO HISTOGRAMS FOR COMPARING TWO SAMPLES. Not shown is the direct superposition of one upon the other, which produces tangled confusion. The bottom panel approximates such superposition.

percentage of all the sample values that occurred, spread out presumably, through that interval. Also with that interval 25 appears at a lower level, at the boundary between $63^{+}-66$ and the next interval $66^{+}-69$. This location fits well with the meaning of that 25, to wit, that 25% of all of the sample values were *66 or less*. The same idea is captured in the graph, Figure 5, prepared from the two cumulative columns. Note that the cumulative values are plotted at the interval boundaries.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

Table 1 HEIGHTS IN TWO SAMPLES. Note relative frequency is an attribute of an entire interval, but cumulative relative frequency relates to the maximum possible value for the interval

	Sample I		Sample II	
Inches	Rel. freq. (%)	Cumulative (%)	Rel. freq. (%)	Cumulative (%)
60+−63	10		5	
		10		5
63+−66	15		10	
		25		15
66+−69	25		15	
		50		30
69+−72	22		20	
		72		50
72+−75	20		35	
		92		85
75+−78	8		15	
		100		100

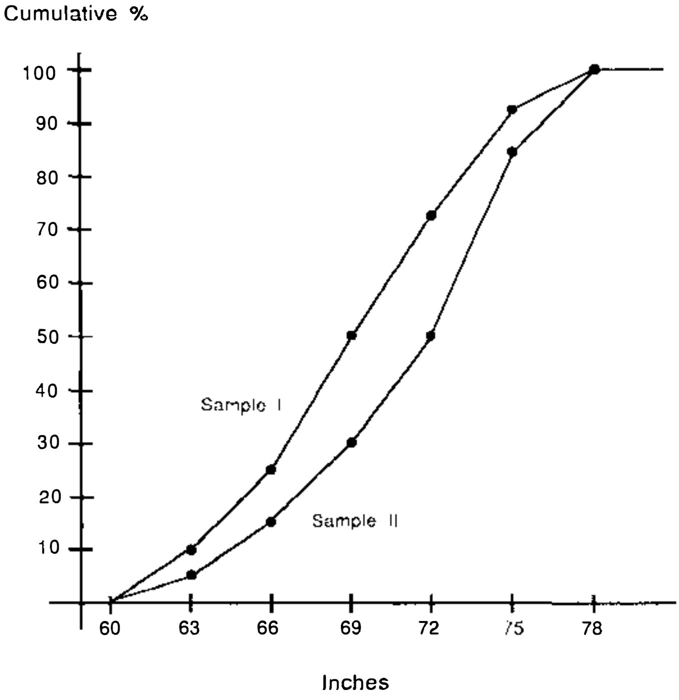


Figure 5 COMPARISON OF TWO SAMPLES USING CUMULATIVE PLOTS. The data of Figure 4 are rendered here in cumulative form.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

The cumulative plot and the histogram contain exactly the same information, for either can be computed from the other. Some advantages attach to the cumulative representation. First, it often "untangles" two or more samples in which the histograms would interweave; this is seen by comparing Figure 5 with Figure 4. Second, the cumulative representation makes it easy to estimate the median (or other percentile) by simply reading off the horizontal value at which the curve attains height .50 (or other desired percentage). Thus the two seventy-fifth percentiles are read off as about 72 ½ inches in sample I and 75 ¼ inches in sample II.

Cumulative representation can help also with data obtained in ordered categories. We illustrate this (and another graphical approach as well) on thc data set (13) found in Table 2, showing the histological grade of 100 tumors of the prostate, falling in five size classes. We regard histological grade as a series of ordered categories, indicating progressively greater abnormality of cells in the tumor. We treat the five size groups as five samples to be compared; this is somewhat artificial, since in fact one sample of routine autopsies furnished the data, and they might most naturally be regarded as 100 bivariate observations, each tumor possessing a grade and a volume; however, we treat the size groups as samples in both the analyses, and both are actually sensiblc ways to display the data; indeed one of them was the form in which the data were originally published.

The data appear in the left-hand panel of Table 2, where the frequencies by grade are shown for the 20 observations in each size class (row) and, at the foot of the table, for the combined samples. The right-hand panel shows cumulative percentages within the row and also, at the foot of the panel, the cumulative percentages for the combined sample.

Graphical representations of these two tables appear in Figure 6. The upper panel depicts the frequencies of the various ordered categories for each size

Table 2 HISTOLOGICAL GRADE OF PROSTATE TUMORS IN FIVE SIZE CLASSES. The columns correspond to increasing abnormality of tissue. Source: Ref. (13)

Size class[a]	Frequencies Grade					Cumulative percentages Grade				
	1–2	3A	3S	4	5	1–2	3A	3S	4	5
A	10	2	7	1	0	50	60	95	100	100
B	3	3	12	2	0	15	30	90	100	100
C	0	5	13	2	0	0	25	90	100	100
D	0	0	12	6	2	0	0	60	90	100
E	0	0	7	10	3	0	0	35	85	100
	13	10	51	21	5	13	23	74	95	100

[a] The size classes A, B, C, D, E, correspond to increasing volumes (in cc) with break points at .054, .171, .464 and 1.42, values chosen to force equal numbers in the five classes.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

class, using shaded bars and "hanging" them at the boundary between classes 3A and 3S. The lower panel plots the five cumulative distributions for the ordered categories, with the category markers located on the horizontal scale in accordance with the combined-data cumulative distribution. Thus the width of each interval is proportional to the number of cases with that histological grade in the combined set of data. On the vertical line at each category division, every cumulative distribution has a plotted point, and in addition the combined cumulative percentage is indicated by an x on each vertical; these lie on the diagonal, which, to avoid clutter has not been added to the figure.

To fix ideas, examine the vertical line with foot at 74, the boundary point on the horizontal axis between 3S and 4. On this vertical line the lowest point, at 35, belongs to group E; this tells us that of group E only 35% had histological grade 3S or less; higher up we find at .55 the point belonging to group D, with the message that 55% of that size group had histological grade 3S or less. The x at .74 attests to the combined sample having 74% of values at 3S or less. B and C agree in having 90% of their members at 3S or less, and in group A, with its point at .95, all but 5% of its members scored 3S or less. This order E, D, C, B, A, seen at this boundary (between 3S and 4) prevails at the three other boundaries as well. The total message is that no matter where you might choose to divide the scale of histology into "less severe" and "more severe," you would find the percentage of "less severe" cases to be greatest in size group A, next in B, then C, D, and E in that order. It is much more difficult to deduce this fact from the upper panel in Figure 6.

Box Plots

The graphical techniques so far discussed portray the entire sample. For some purposes a much briefcr summary will suffice—like simply the sample mean, or the median. In still other cases, such a statistic may tell too little, and yet the histogram or spike diagram is unnecessarily detailed. The *box plot* can help; it gives a useful idea of the sample distribution without portraying it fully. There are several closely related types. To fix ideas we point to this one: Compute the median, the lower quartile, Q_1, (the twenty-fifth percentile) and the upper quartile, Q_3 (the seventy-fifth percentile.) Show them in a format like that in Figure 7.

The Figure shows the following: (*a*) the left-hand sample has a median of 60; thc right-hand one a median of 50; (*b*) the left-hand sample is more compactly distributed around its median than the other; (*c*) the right-hand sample is roughly symmetrically distributed, but the other is quite unsymmetric. In principle we could instead use the sample mean rather than the median, and could put the limits at one standard deviation (or 1.5) above and below the mean. Notice that this would not reveal the information about asymmetry.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

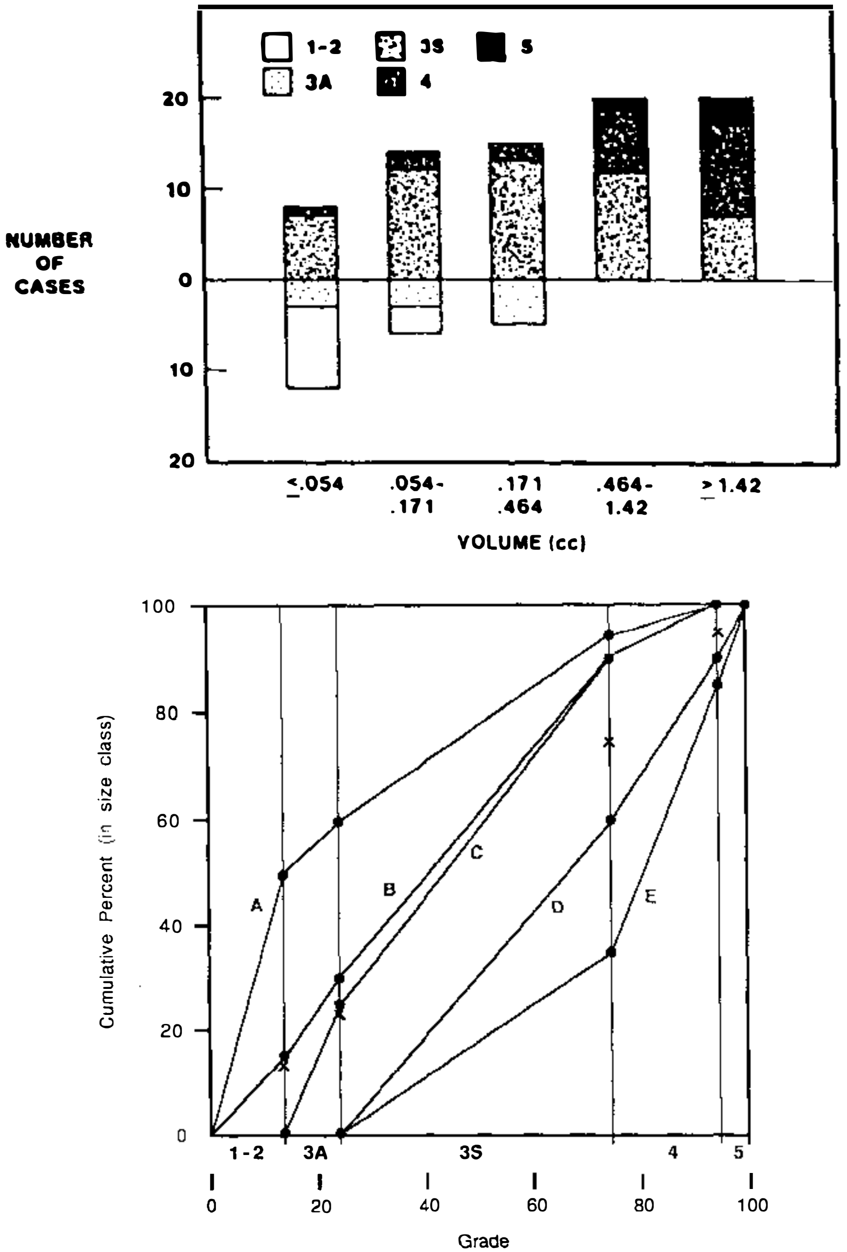


Figure 6 FIVE CUMULATIVE DISTRIBUTIONS OF ORDERED-CATEGORY DATA. The upper panel shows a conventional display, using shading to represent the ordered variable, which is intensity of abnormality. The lower panel shows the same information cast in cumulative form. Source: Ref. (13).

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

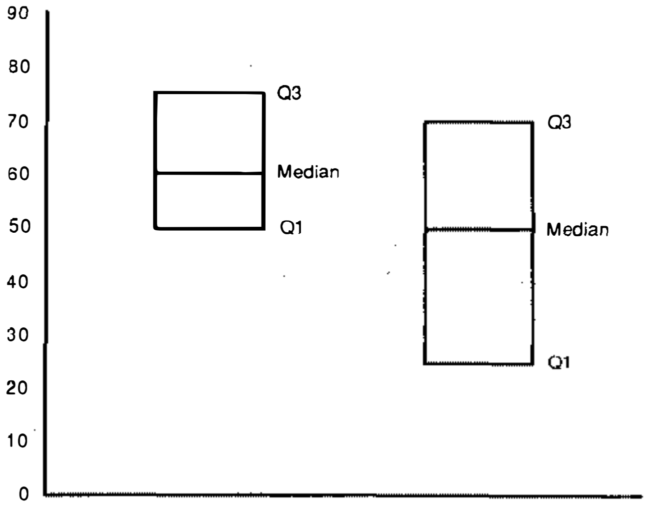


Figure 7 TWO BOX PLOTS. These are a rather minimalist version of the box plot approach, showing only the three quartiles. Other approaches add information concerning the data that lie above and below the quartiles.

Box plots are often usefully elaborated by adding graphical information about data outside the two quartiles. We do not follow this up here, but direct the interested reader elsewhere (5a).

Box plots can quite usefully display the essential features of many samples in one chart, where histograms and cumulative plots would fail.

Figure 8 depicts certain coal mine accident data for 19 large coal mining companies, each with several mines, ranging from 4 to 77, a total of 424 mines in all (5). This picture is based on a table with 424 rates of disabling injuries over the period 1978–1980. The 424 numbers involved are made quite available to the reader by this depiction.

Means and Standard Errors

The graphical devices presented to this point deal with display of all the data, as with dot diagrams, histograms, and cumulative plots, and, with some abridgment, box plots. But sometimes less detail is needed, and then simplicity may commend displaying only sample means (or medians), perhaps supplemented with an indication of statistical uncertainty. In this section we begin with the simplest case, move on to rather more complex ones, and

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

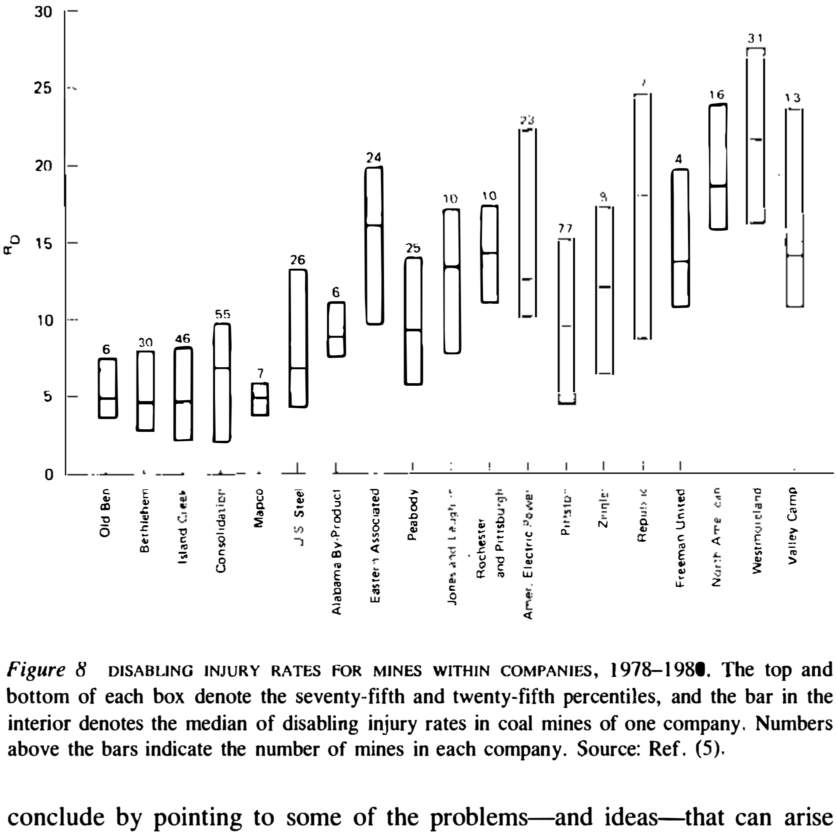


Figure 8 DISABLING INJURY RATES FOR MINES WITHIN COMPANIES, 1978–1980. The top and bottom of each box denote the seventy-fifth and twenty-fifth percentiles, and the bar in the interior denotes the median of disabling injury rates in coal mines of one company. Numbers above the bars indicate the number of mines in each company. Source: Ref. (5).

conclude by pointing to some of the problems—and ideas—that can arise even when each sample is reduced to a single number like the mean.

THE SIMPLEST MEAN: THE BATTING AVERAGE As every baseball fan knows, a player's batting average is a proportion; it tells what proportion of times at bat resulted in the player's getting a hit. This homely example reminds us that a proportion is indeed a (very simple) sample mean. In Figure 9 we see a graphical representation of these data: in Hospital I, of 24 births 11 were female, and in Hospital II, of 50 births, 28 were female. The chart shows the percentages, 46% and 56%. If we wished to indicate the statistical uncertainties, then we would calculate the two standard errors, which are .10 and .07, and show the data as in the lower panel of the Figure.

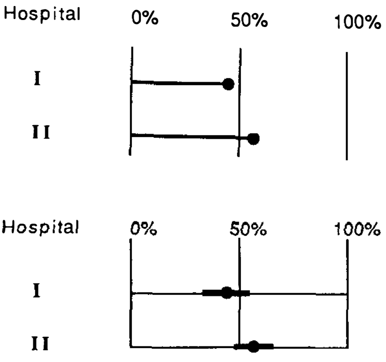


Figure 9 PERCENTAGE OF FEMALE BIRTHS AT TWO HOSPITALS. In the lower panel, bars reach one SE above and below indicated rate.

It is evident that either style is easily adaptable to displaying more than two proportions, say 5 or 10 or 20. It is also evident that either chart could be drawn so that the lines now horizontal would be vertical.

DISPLAYING MANY MEANS In Figure 10, 28 means, with error bars, are shown; they are arranged in descending order of magnitude; the numbers across the top provide identification numbers for the 28 samples. A list of the sample sources, keyed to the identification numbers, would complete the figure.

The error bars have been drawn to length 1.5 standard errors, on each side of the mean. This choice (rather than 1.0 or 2.0) makes it convenient to assess statistical significance, since two means with bars that do not overlap differ at approximate significance level .05, or less.[1] Thus sample 4 can be seen to have a mean significantly exceeding the mean of sample 9, or any one to the right of it.

[1] The standard error of $\bar{x}_i - \bar{x}_j$ is se $= ([\text{se}(\bar{x}_i)]^2 + [\text{se}(\bar{x}_j)]^2)^{1/2}$ and we say the means differ at significance level .05 if

$$|\bar{x}_i - \bar{x}_j| > 2.0 \text{ (se)}. \qquad 1.$$

This standard is appropriate if n_i and n_j are "large." When they are not, then 2.0 should be replaced by the two-sided .05 significance point for t with appropriate degrees of freedom, and 1.5 should be multiplied by one-half of that t value. It can be shown that so long as se($\bar{x}_i$) and se($\bar{x}_j$) are not different by a factor exceeding 2.1, then when the 1.5 bars fail to overlap, Eq. 1 is satisfied and the means differ significantly at .05. Using 1.6 in place of 1.5 would give wider bars ("more conservative") but ensures that non-overlap implies Eg. 1 if se ($\bar{x}_i$) and se ($\bar{x}_j$) differ by larger factors—up to 3.2. With the data in Figure 10 it is apparent that no nearby means have standard errors differing by so large a factor as 2.1, so we may accept the non-overlap significance criterion as applicable in the example.

One might ask whether multiple comparisons issues invalidate this informal significance testing procedure. Not necessarily. If the entire set of means are significantly different by a .05 level F test, then the suggested procedure is a simple approximation to Fisher's Least Significant Difference method, at the .05 level, which is a standard multiple comparisons technique (14).

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

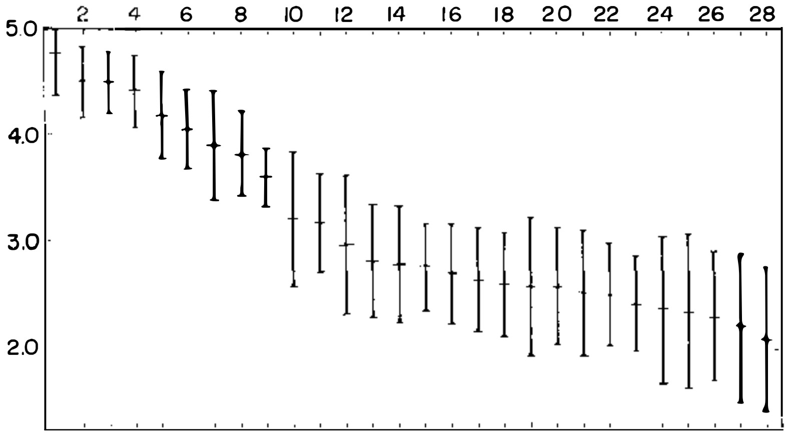


Figure 10 TWENTY-EIGHT LABELED SAMPLE MEANS, ARRANGED IN DESCENDING ORDER OF SIZE, WITH ERROR BARS.

MANY MEANS WITH "STRUCTURE" Our examples have tacitly assumed a kind of symmetry among the means, all being thought of as "on an equal footing." But often this would be an unrealistic stance. We might have means from 12 strains of some yeast, all on an equal footing. But, instead, the twelve groups might correspond to six strains of yeast, each being grown in nutrient media A and B. Or there might be three yeast strains and four nutrient media; or three yeast strains, two media, and two temperatures, in all possible combinations. Perhaps there are two groups, one of five related strains, the other of seven. In every case there are 12 samples, but they bring forth different kinds of questions, and call for different kinds of graphical (as well as numerical) presentation. We see below how such issues of structure bear on the appropriateness of alternative modes of display.

An example with temporal structure Table 3 shows fatal motor accident statistics from Colorado (12). In the years 1964–1968 and again in 1978 and 1979, there was no helmet law. In the years 1970–1976 there was a helmet law. Finally, 1969 and 1977 were years in which a helmet law was in effect for part of the year.

The fatality rates for each year are plotted in Figure 11; in addition, the simple arithmetic averages of the rates in each of the three periods are shown as horizontal lines reaching throughout their periods. (Notice 1969 and 1977 are excluded, because they each comprised two parts, one with and one without a helmet law.) The chart strongly suggests that the fatality rates were lower when the helmet law was in effect. A more delicate and complete

Table 3 FATAL MOTORCYCLE ACCIDENTS IN COLORADO. During 1970 through 1976 a helmet law was in effect, but not in the periods before and after. Source: Ref. (12)

Year	Motorcycle registrations	Fatal accidents	Fatal accident rate[a]
1964	16,645	10	6.02
1965	21,479	10	4.65
1966	24,811	14	5.65
1967	26,034	17	6.53
1968	28,594	23	8.04
1969[b]	34,889	29	8.31
1970	44,851	27	6.01
1971	57,098	21	3.68
1972	68,908	38	5.51
1973	81,871	45	5.59
1974	92,833	39	4.20
1975	95,439	47	4.92
1976	98,051	31	3.16
1977[c]	108,559	57	5.25
1978	110,000	63	5.73
1979	115,000	74	6.43

[a] Accidents per 10,000 registrations.
[b] Helmet law effective July 1, 1969.
[c] Helmet law repealed May 20, 1977.

statistical analysis might employ *weighted* means of the rates within the three periods. The rates of later years have smaller sampling variability than the early ones, and would show vertical error bars (perhaps using the 1.5 convention) for each of the averages.

As it is, a quick assessment of statistical significance is conveniently made. Beside each of the 14 points is shown its rank among the 14, with the smallest rate (in 1976) receiving rank 1, and the largest (1968) receiving rank 14. The sum of the seven ranks for the years with the helmet law is 34, considerably smaller than the null expectation for that sum, which is 52.5 (seven times the average rank, 7.5 which is midway, between 1 and 14). Indeed, that sum, 34, is significantly small, at $p = .02$, two-sided, applying Wilcoxon's two sample test.

This example has shown how graphical display of 16 years' rates elucidates the possible impact of a helmet law in effect during part of the period. The reader is helped greatly by the use of reference lines: the three horizontal lines showing subperiod averages, and the vertical lines defining the subperiods. All information in the graph comes from the table. But the eye and the mind may perceive the information better from the graph.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

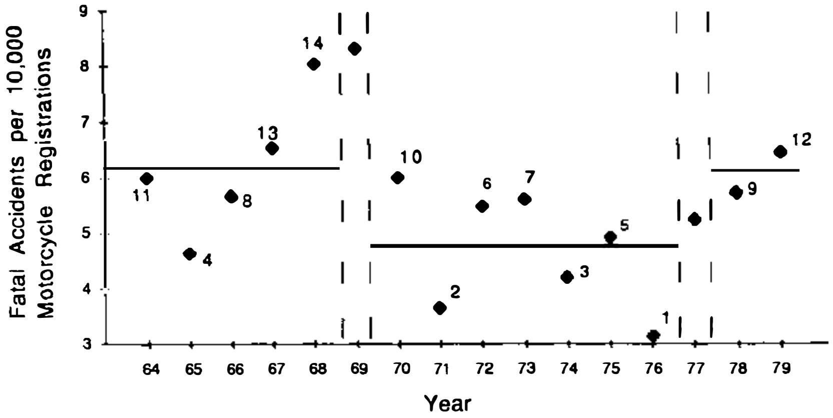


Figure 11 FATAL MOTORCYCLE ACCIDENTS IN COLORADO WITH AND WITHOUT A HELMET LAW. Numbers appearing near the data points are ranks (1 is smallest, 14 is largest) of the observations. Horizontal lines are simple averages of the rates for the period. The ranks justify the conclusion that the lower rate during the helmet law era is not likely to be a chance aberration. The data are the same as in Table 3. Source: Ref. (12).

Factorial structure The previous example was relatively simple, with a three-era structure. Things are much more complex when the data have a factorial structure with, say, one mean for each possible combination of two classifications, one with r classes and the other with k classes. We would have such structure if five kinds of standard specimens were blindly read at each of four laboratories. Graphical display of the 20 (structured) averages could be based on the analysis of variance applied to the data. The 20 means would be summarized in (*a*) the five averages for specimen type, (*b*) the four laboratory averages, (*c*) the 20 differences between each observed average and the value fitted for it by suitably combining the laboratory and specimen-type averages. Each of (*a*), (*b*), and (*c*) could be graphically depicted by methods we have already seen. But we will not take up a detailed treatment of factorially structured means here. Instead we turn, in the next section, to some devices that become available when one or two of the classifications in the factorial structure have only two levels.

Two-Way Displays for Univariate Data

We first look at display of data obtained in pairs. We then turn to an alternative display for means of two samples, and then to several means that arise from a 2 × 2 × k factorial structure.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

MATCHED SAMPLES; PAIRED DATA Data arrive in pairs in many ways. They may be before and after values; they may be ipsilateral and contralateral measurements, or opinions of husband and wife, for example. Whenever data occur in pairs, we should use methods of display and analysis that take account of this feature. It is incorrect to use two-dot diagrams, or two histograms. Correct methods of display fall into two classes. First, one may construct from each pair a single number, like the difference after-minus-before, showing change, or a ratio like A as a percentage of B. Or, one may plot the two values for each pair on a scatter diagram. An interesting data set (11) provides a natural vehicle for exhibiting both approaches.

The data record the number of episodes of apnea, stoppage of breathing (exceeding 20 seconds), per hour of sleep, for each of eight premature infants, under two conditions. Each infant was bedded for two six-hour periods on a bassinet, and for two (sandwiched) six-hour periods on a waterbed. The two apnea rates correspond to the two bedding conditions.

The data are given in Table 4. Note that the differences between x and y are considerably less variable than are the measurements under each condition, as indicated by the much smaller range for d than for x or for y.

The data of Table 4 are displayed in two separate dot diagrams in the top panel of Figure 12. See how similar the two diagrams are. This carries a suggestion of "no difference." But then observe the lines connecting the two observations of each infant. All eight such lines slope downward to the right; all eight attest to reduced apnea on the waterbed. This is a very different—and correct—conclusion. This example illustrates our earlier statement that with paired data it is incorrect simply to display the two dot diagrams without taking precautions; to do so is likely to convey a false impression, as here. The connecting lines work well here because there are few points; they do not offer a generally useful way to patch up the dot diagrams. Two better

Table 4 HOURLY RATE OF APNEIC EPISODES IN PREMATURE INFANTS DURING SLEEP ON TWO KINDS OF BEDS. Each of eight infants was bedded for 12 hours on a waterbed and on a bassinet, in alternating six-hour periods. Source: Ref. (11)

Infant	Waterbed (x)	Bassinet (y)	Difference $d = x - y$
1	0.89	1.36	0.47
2	0.77	1.66	0.89
3	0	0.11	0.11
4	0.65	1.44	0.79
5	0.88	1.63	0.75
6	1.36	1.52	0.16
7	1.22	1.53	0.31
8	0.30	0.48	0.18
Range	1.36	1.55	0.78

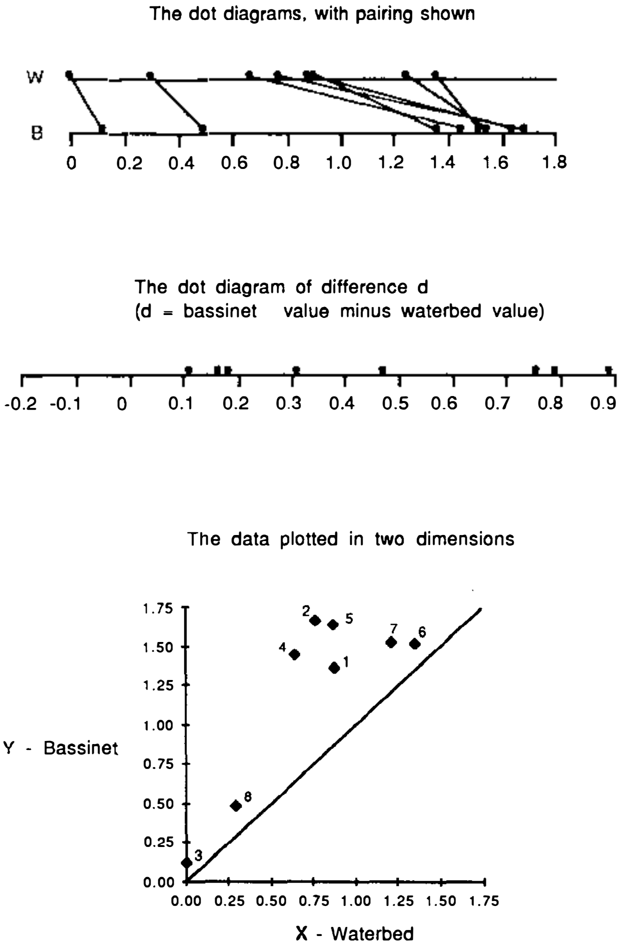


Figure 12 APNEA IN EIGHT PREMATURE INFANTS UNDER TWO SLEEP CONDITIONS. The numerical values represent number of episodes of apnea (exceeding 20 seconds) per hour of sleep on a waterbed (W) or bassinet (B). These are the data of Table 4. Source: Ref. (11).

graphical procedures are offered in the middle and lower panels. In the middle panel the differences (bassinet value minus waterbed value) are plotted on a dot diagram. There are no negative values and the reduced apnea on the waterbed is clearly revealed by the eight positive differences. The bottom panel plots the two values for each infant in a scatter diagram, and shows clearly (*a*) the superiority of the waterbed, since every point above the diagonal denotes an infant with bassinet apnea rate larger than waterbed rate; (*b*) the correlation between the two rates, which arises from a tendency for

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

some infants to be more or less apneic than others on both beds. This correlation accounts for the gain in information available from measuring each infant on both beds, and for the inappropriateness of two separate dot diagrams. Finally, we remark that the display in the bottom panel contains more information than does the middle panel; Observe that from the lower one we could reconstruct the middle one—but not vice versa.

It is conceptually correct to think of this example as really not a univariate one; each infant has provided two measurements—and the data are bivariate. But because both measurements are of the same kind, apneic rate during sleep, they are less obviously bivariate than would be, say, data on apneic rate and heart beat for each infant. So we have employed some expository license in using the term "univariate" at all. Now we take a rather opposite twist, by treating pairs of independent sample means in bivariate coordinates, because of advantages that will become apparent.

PLOTTING TWO MEANS: REVISITED Let us return to the data for the female birth fraction at two hospitals (Figure 9). This information is plotted in an entirely different way in Figure 13.

The two axes show that the plotted point represents the rate .46 for Hospital I and .56 for Hospital II. A point that fell on the diagonal line would denote equal rates for the two hospitals. The uncertainties of the two rates again are shown, and the larger standard error for the rate of Hospital I is visible. The diagonal arrow, constructed as the diagonal of the rectangle defined by the horizontal and vertical standard error segments, represents the standard error of the *difference* between the two rates. When that arrow is rotated to point vertically (or horizontally), it crosses the diagonal line of equality; this means that P_{I}–P_{II} differs from zero by less than one standard error.

EXTENDING THE TWO-MEANS-ONE-POINT PLOT Murray & Bernfield (15) studied incidence of low and very low birth weight as it related to adequacy of

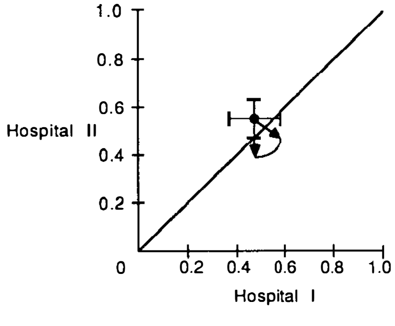


Figure 13 PERCENTAGE OF FEMALE BIRTHS AT TWO HOSPITALS, WITH ERROR BARS. The diagonal arrow shows the standard error of the difference between the rates in these two (independent) samples. Since it is long enough to cross the line of equality, we see that the two rates differ from one another by less than the standard error of that difference.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

Table 5 RACE, ADEQUACY OF PRENATAL CARE, AND FREQUENCY OF SMALL BABIES. Low birthweight (LBW) and very low birthweight (VLBW) infants are more frequent among black mothers than white, and when prenatal care has been less adequate. Source: Ref. (15)

	LBW (≤2500 g)		VLBW (≤1500 g)	
	Black	White	Black	Whitc
Frequencies				
Prenatal Care				
Inadequate	60/400[a]	100/1800	18/400	18/1800
Intermed.	165/1500	450/9000	36/1500	81/9000
Adequate	100/2000	340/17000	22/2000	85/17000
Percentages: *(P)*				
Prenatal Care				
Inadequate	15[b]	6	4.5	1
Intermed.	11	5	2.4	.9
Adcquate	5	2	1.1	.5
1 + log *(P)*				
Prenatal Care				
Inadequate	2.18	1.78	1.65	1.00
Intermediate	2.04	1.70	1.38	.95
Adequate	1.70	1.30	1.04	.70

[a] This ratio reports that of the 400 black mothers with inadequate prenatal care, 60 bore babies weighing 2500 g or less.

[b] This is the percentage of inadequate-care black mothers who bore babies weighing 2500 g or less.

prenatal care and to race of mother. Table 5 is adapted from their data and in its upper panel displays (approximate) numbers of mothers of the two races with infants of "low" birthweight (less than 2500 grams) and of "very low" birthweight (less than 1500 grams), sorted out by adequacy of prenatal care. (Observe that the "very low" birthweight infants are a subset of the "low" birthweight infants. Neither group includes any birth of less than 500 gm, for which survival is very uncommon.) The data in the top panel of the Table consist of 24 numbers, twelve fractions each with a numerator and denominator. Reducing the data to percentage of births cuts the number of entries to 12, displayed in the middle panel.

Now graphical display of these 12 percents is the task at hand. A conventional display might use 12 bars, as in Figure 14. Alternatives to this mode of display are offered in the three panels of Figure 15.

In the top panel of Figure 15 we see six line segments, each showing a black percentage and a white percentage, corresponding to the two races. Plotting the data in the style using two coordinates, in the middle panel, calls for only six points and it is easy to see that (*a*) the fraction of very low birthweight (VLBW) births is much smaller for both races than the fraction

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

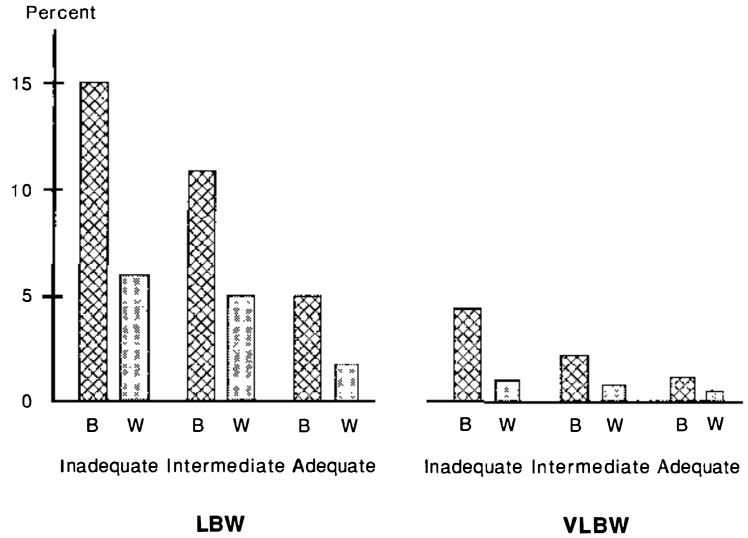


Figure 14 PERCENTAGE OF ALL BIRTHWEIGHTS THAT WERE LESS THAN 2500 GRAMS (LBW) OR LESS THAN 1500 GRAMS (VLBW) BY RACE OF MOTHER AND ADEQUACY OF PRENATAL CARE. This is a bar graph depiction of Table 5. Source: Ref. (15).

low birthweight (LBW) births, since the VLBW points are nearer to the origin, where both rates would be zero; (*b*) for both races the incidence of small babies (LBW and VLBW) is lower among the mothers with better prenatal care; (*c*) the rates for blacks are greater than those for whites in all six weight × care classes, since all six points are below the diagonal of equality. Indeed the picture suggests that in each weight class there may be a constant proportion between the black and white rates, since the three points for each weight class lie near a ray through the origin.

The last observation suggests looking at the rates on a logarithmic scale, where constant proportions are rendered as constant logarithmic differences. The bottom panel of Table 4 displays the values of 1 + log P. (The 1 is added to avoid negative numbers, and *P* denotes percent). When these logarithmic values are plotted, the bottom panel of Figure 15 is the result, and it gives a very simple looking representation. The higher rates for blacks now look roughly constant, since all six points lie close to a line about .4 logarithmic units below the diagonal of equality, denoting a multiplicative factor of 2.5. Indeed, the ratios of black to white rates can be seen from the data in panel *b* to all be close to 2.5, except for the one case of inadequate care and very low birthweight, where the ratio of the black to white rate is 4.5. This discrepant point appears as the white circle in the bottom panel of Figure 15. It is statistically the least precise point of all; reference to the top panel of Table 4 shows that both its coordinates are determined by frequencies of 18, and so both are appreciably affected by sampling error.[2]

PRESENTING BIVARIATE DATA

If we consider two items of infonnation on each observed unit (person, laboratory specimen, experimental animal), we are dealing with *bivariate data*. Either of the two variables may be continuous, discrete, ordered (like histological grade or pain relief on a five-point scale), or categorical.

Where one of the two variables ("components") of each bivariate observation denotes a category, like gender or blood type, and the other component is continuous, then graphical display must consist of separate distributions (or box plots) or means, one for each category. More generally, if one of the two components is categorical, then the result is separate displays, one for each category, of the distribution of the other variable, whether it be discrete-numerical, ordered, or itself categorical.

Preceding parts of this paper presented two devices for displaying bivariate data. The first is to divide one of the continuous variables up into successive size intervals, and then give a univariate display of the other variable separately for each size group (as with the tumor pathology data, where the continuous variable, volume, was broken into five classes A, B, C, D, E, and then the pathology distribution for each volume class was exhibited). The second device is reduction of the two components of each bivariate observation to a single function of them, one value for each subject, with a univariate display of those reduced data. The bassinette/waterbed data exemplified this idea when the apnea rates per hour of sleep were calculated for the two conditions, and then the difference, bassinette-minus-waterbed, was taken as the univariate observation for each of the eight infants. These two earlier glimpses leave much still to be said about displaying bivariate data, and we turn now to a somewhat fuller and more systematic treatment.

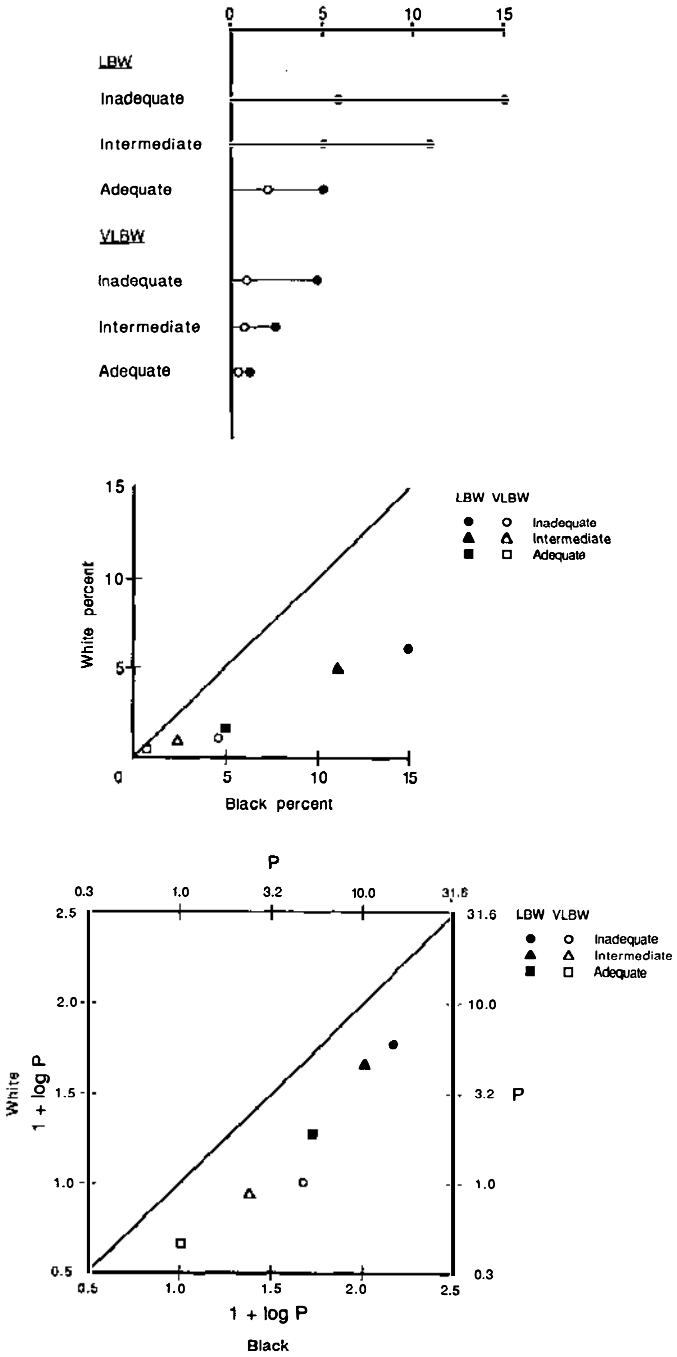
The Fundamental Display for Bivariate Data

With continuous data, the dot diagram is the fundamental data display; from it can be derived histograms, box plots, and so forth. The corresponding fundamental display for bivariate data is the scatter-diagram.

[2] The logarithmic difference between these two percents (4.5 and 1.0) is .65, and its standard error is .33; Thus the white circled point is not significantly removed from .40, the value that summarizes thc othcr fivc points; The apparent discrepancy may reasonably be ascribed to sampling error.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.



Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

With univariate displays, there can be a choice whether to plot *y*, or its reciprocal, or its logarithm, and with bivariate data these choices also present themselvcs, but now for both components of the bivariate observation. One aspect of the art of graphical presentation is choosing the scales on which data can best be presented. In the examples below we see several charts with messages that are clear and crisp, largely because of wise choice of scale transformation on one or both axes of the scatter diagram.

When there are very large quantities of data, the scatter diagram, especially if done by computer, can become difficult to read, because multiple data points are hard to show effectively in the scatter diagram. Helpful ideas for coping with this problem are to be found in Cleveland's fine book (4) on pages 155–62. Another approach, of course, is to construct the analog of the histogram; the intervals on the *x* and *y* axes define a grid, and in each rectangle ("cell") of the grid, some number of bivariate observations from the sample occur (possibly zero in some cells.) To construct the diagram that extends the histogram to bivariate data, let one case be denoted by a "brick" that fits exactly on a grid rectangle; then the number, *f*, of observations occurring in the cell can be indicated by a stack of *f* bricks on that cell. The resulting brick pile, erected on the x,y grid, represents the distribution of the sample values on the x,y plane in a manner analogous to the histogram's representation of the distribution of univariate sample values on the number line. Drawing the brick pile so that it is clearly interpretable requires technique that takes account of perspective in drawing three-dimensional figures. Some computer packages do this well.

Where both bivariate components are discrete, the "natural" representation is a set of spikes, each recording the number of observations that occurred at a point in the plane determined by the possible values for the two components. Such a diagram can be hard to draw well and hard to interpret, depending on how many spikes there are, how much they differ, and so forth. If there are very many such spikes then grouping them into grid cells and making a brick pile might serve well. It may not always prove possible to construct a drawing of doubly discrete bivariate data that will help the reader or the investigator to understand better the numerical data rendered in a table.

GRAPHS AND RELATIONS BETWEEN VARIABLES Bivariate data ordinarily carry with them questions about how the two variables are related. Does lower incidence of dental caries occur in regions with higher levels of fluoride in the drinking water? Does the incidence of heart disease rise with increasing levels

Figure 15 THREE ALTERNATIVE DISPLAYS OF TABLE 5. The first panel replaces 12 bars with 12 points, on six lines. The next two panels show all the data, using six points. The first uses the natural scale of percentages, the second uses a logarithmic scale.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

of sugar consumption? Is the length of the femur related to the long axis of the skull? To the diameter of the wrist? Is socioeconomic status related to health care seeking behavior?

Often one of the two variables is naturally thought of as (possibly) influencing the other, as with fluoride in the drinking water, in which case we would think of caries incidence as possibly *responding* to fluoride levels. In such circumstances it can be informative to study how the *average value* of the response variable relates to different levels of the input variable (often called the "independent variable").

Following common usage, let us name the independent variable x, and the other, whose average we study, y. Then the curve that describes the average of y at varying x is called the *regression of y on x*. We may write this as

$$\text{Ave } (y(x)) = f(x).$$

Only a few values of x can appear in a finite sample; if $x_1, x_2, \ldots, x_N$ are observations on some variable, we can see at most N distinct values of x, perhaps fewer because of repeated observation at some x points. To estimate $f(x)$ for a continuous range of x requires that we somehow combine information from the data in hand, at the observed x's, to describe $f(x)$ for the continuous range of x's that the data relate to. Two approaches to this task are (*a*) model-based regression and (*b*) smoothing. Either approach leads to a line, or other curve, that can be plotted on the scatter diagram, and that represents a sort of trend, depicting how the average value of y changes as x changes. In Figure 16 appears a rather complex example illustrating the idea (9). The data concern cell species of four types: (*a*) RNA viruses and single stranded DNA viruses *(solid squares)*; (*b*) double-stranded DNA viruses *(solid circles)*; (*c*) haploid microorganisms *(shaded circles)*; and (*d*) diploid microorganisms *(shaded squares)*. Each of 31 cell species, accompanied by a numerical identifier, is plotted at a point (x,y) that depicts x, its radiosensitivity, as measured by the logarithm of the dose needed for 63% inhibition of cell reproduction, and y, the logarithm of the mass of its nucleotide material.

With the data plotted on these scales, not only is it easy to see that their sizes and radiosensitivities show family similarities, but also the strong negative relation between size and cell-killing dose is quite evident. To the data of each of the four types has becn fitted a straight line (an example of a model-based method) by least squares. The four lines have been fitted subject to the constraint that they all have a common slope. Inspection of the figure allows consideration of whether the parallel lines offer a reasonable characterization of the relationship between radiosensitivity and nucleotide masses. The original research article offered heuristic interpretations of the numerical value of the common slope, and the acceptability of a common slope was essential to that argument.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

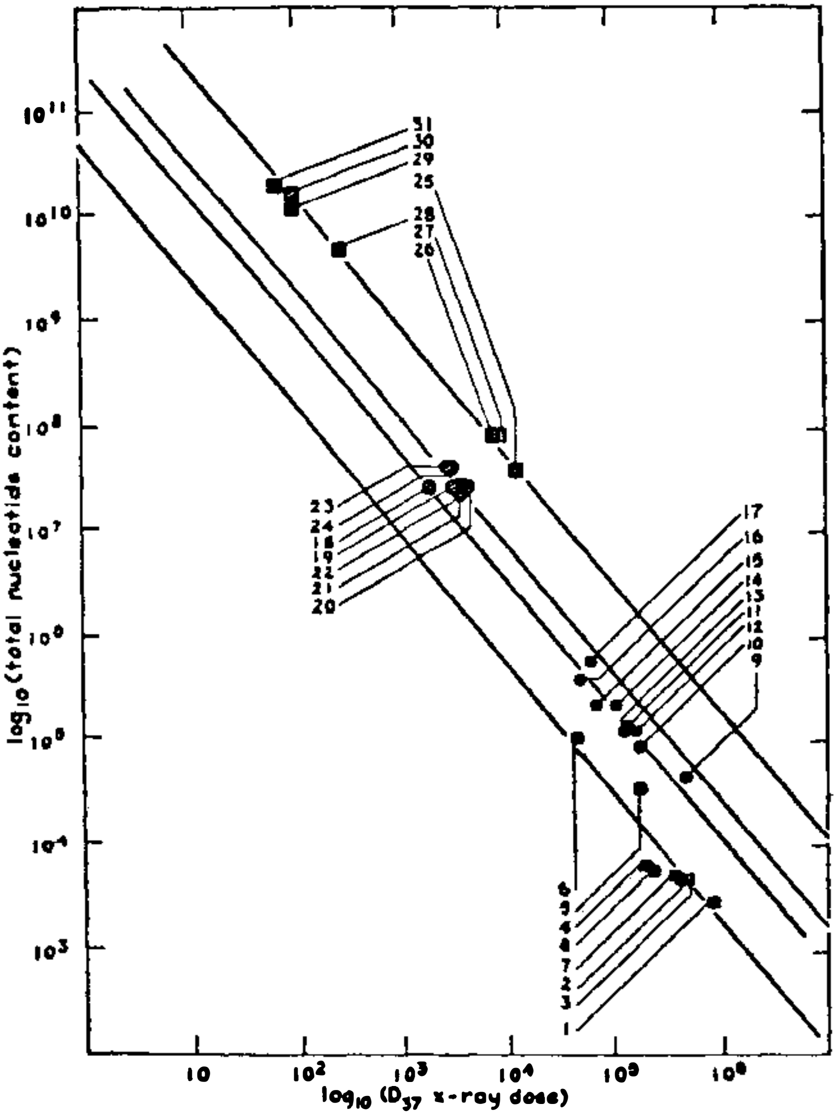


Figure 16 RADIOSENSITIVITY AND NUCLEOTIDE MASS OF 31 SMALL ORGANISMS. The plot is logarithmic on both axes. The lines have been fitted by least squares, under the constraint of a common slope. Source: Ref. (9).

Model-based fitting need not be done by least squares (although it very often is), but it does necessarily produce a prechosen *kind* of regression function; in the example only straight lines, and parallel ones at that, were possible outcomes of the fitting. With smoothing techniques, the *method* is

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

definite enough, but the form of the resulting regression curve is not preordained, except in a very minor respect to be mentioned below.

In Figure 17 both panels depict the same data set (4, pp. 170, 171). Each point represents a hamster; the x coordinate records how many days of the animal's whole lifetime were spent in hibernation, and the vertical coordinate records days of the whole lifetime not spent in hibernation. The eye finds a general pattern of increasing values of y as x increases, and the idea of estimating how, on the average, y changes with x is a natural one. The smoothing algorithm that was used is called LOWESS (3a). At each x is computed a kind of fitted value based on a robust weighted linear regression using only the observations that have x-values "sufficiently near" x_i. At each x_i a separate line is computed to arrive at that observation's fitted value. The concept "sufficiently near" is an adjustable parameter of the smoothing algorithm; the greater it is chosen, the smoother the curve. The left-hand panel used a smaller smoothing parameter and seems to be less satisfactory in this example. The algorithm gives a fitted value at each x_i; these are then joined by straight line segments, and it is in this respect that the result of the smoothing algorithm is slightly "fore-ordained."

In both of the examples above it can be asked how the choice of independent variable was made. In the case of the hamsters it appears that there was known biological reason to regard hibernation time as influencing nonhibernation longevity, rather than vice versa; thus, tracking the average "effect" as it related to the numerical value of the "cause" was chosen as shown in the hamster data.

In the case of the radiosensitivity and nucleotide mass data, the answer is rather complicated. The fitting used radiosensitivity as the dependent variable and nucleotide mass as the independent one. This choice seems intuitively satisfying, but the real reason for that choice was the technical one that while both variables are measured subject to error, nucleotide mass has the smaller uncertainty. Because of conventions in the radiobiology literature the chart then depicted the dependent variable on the *horizontal* axis. Generally, to prevent confusion it is good to take account of the conventions applicable among the chart's readership.

Some Devices that Can Be Useful in Bivariate Graphing

TRANSFORMATION OF VARIABLES Generally, a straight line is a simpler curve to recognize, think about, and characterize numerically, than is a curve. Thus, it can be advantageous to find some simple way to transform one of the variables (or both) to replace a curvilinear relation between the original variables with a straight-line relation between the transformed ones. We have already met this idea in preceding portions of this article. We have seen the

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

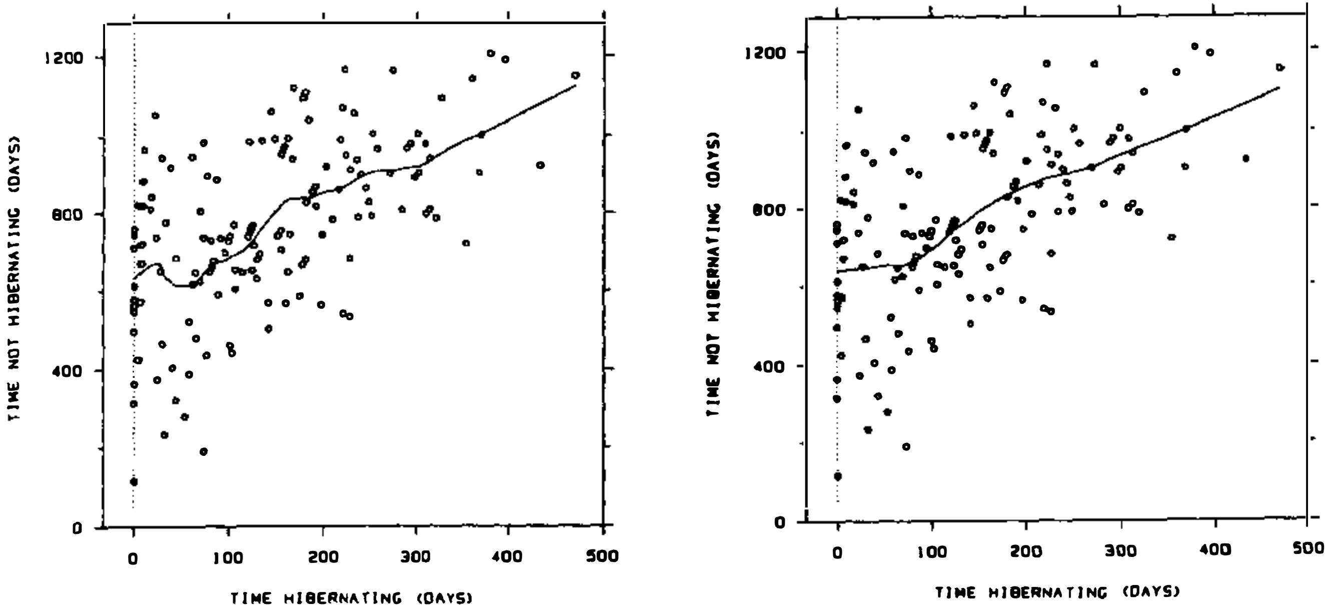


Figure 17 HAMSTER LIFETIMES WITH REGARD TO TIME SPENT IN HIBERNATION. A LOWESS curve is fitted in each panel. The one on the right has used a larger smoothing parameter. Source: Ref. (4). Copyright 1985, Bell Telephone Laboratories, Inc. Reprinted by permission.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

logarithmic transformation applied to the radiosensitivity data, and with ordered category data we have seen the scale of an ordered category variable constructed to make one of the samples (or the average of all of them) yield a straight-line plot. Other transformations, of course, apply in suitable circumstances. The distance y that an object moves down an inclined plane in time t, in the absence of friction, obeys the relation

$$y = ct^2$$

where the constant c depends on the angle of inclination of the plane. *Two* transformations are available to turn this into a linear relation. First, we might replace t^2 by u, then

$$y = cu$$

is a straight line relation, and the slope of the line when plotted would reveal the value of c. Second, we might take logarithms of both sides of our original equation and obtain

$$\log y = \log c + 2 \log t,$$

arriving at a linear plot of log y against log t, with slope 2; the intercept, log c, embodies the information about c. Choice between these two with real, noisy, observed data would be informed by making plots of both kinds and seeing for which one the straight line plot better fitted its plotted points. It could then be used to estimate c.

CHOOSING THE SCALES The ability of the reader to absorb the message of a bivariate graph can depend on choices like which variable should be plotted horizontally and which vertically. In some disciplines the variable thought of as the stimulus or input is given the horizontal coordinate, and the one thought of as response is given the vertical; most statisticians view data this way. In other disciplines the contrary convention is usual. It is well to bear in mind, and use, the convention applicable to the problem under study.

Sometimes it seems more natural to display not y against x but y-x against x. Indeed the hamster data illustrates this device; originally the bivariate data for each animal were longevity in days *(y)* and days spent in hibernation *(x)*; the figures were made by plotting y-x against x.

A feature of the two scales that affects the general gestalt of the chart is the relative compression of the data in the vertical and horizontal directions. This is largely determined by the scale intervals chosen on the two axes. Note in the hamster data that the scales, both in days, are unequal. This has the result of exhibiting the points with approximatcly equal visual spread in both directions; it complicates recognizing that the slope of the regression is, for days of hibernation exceeding 100, about 1.0, a fact that would be reflected in a 45 line if the two scales were equal. Extreme choices of scales can virtually eliminate the appearance of variability. If a sequence of closely related charts is involved, then the same scales for all of them would be a reader's unconscious expectation, and it should be transgressed only with good reason and clear notice to the reader.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

THE USE OF REFERENCE LINES The bivariate display can sometimes gain interpretability from the presence of some lines drawn in for reference. Examples include: (*a*) one or more regression lines, (*b*) the diagonal of equality, (*c*) lines parallel to that diagonal, showing lines of constant difference $y\text{-}x = c$, (*d*) rays through the origin, showing lines of constant ratio $y/x = c$. Other ways to provide reference standards for plotted points might in special circumstances be a family of concentric circles, or of lines $ax+by = c$, or yet others. Considerations of avoiding clutter compete here with aiding the reader to interpret the data.

An ingenious use of reference lines to meet the problem of data overlap is offered by Cleveland, and with permission, we reproduce his example here. Figure 18 depicts brain weight and body weight for many species of animals, belonging to four broad groups: birds, fish, primates, and nonprimate animals. The scales are logarithmic. The three reference lines all have slope 2/3, for theoretical reasons. (And the points fit well, supporting the theory.) This four-fold display is a brilliant substitute for superposing the points on one chart and relying on the usc of different symbols (circles, dots, triangles, shaded, unshaded) to distinguish points from different groups.

USING COLOR TO DISTINGUISH GROUPS The eye apparently can distinguish among small figures more directly in terms of contrasting colors assigned to them than in terms of different shapes (squares, circles, triangles, etc) or shadings. The data-analyst-investigator can easily exploit this principle when graphing a point-swarm by us'ng different colored pens for data points from different groups. Color can be equally effective in presenting the data, but costs and facilities often intervene to make color unavailable. So its primary use, at least in many sett. ıgs, will be the private one of studying one's own data, where it can be invaluable.

GRAPHING MULTIVARIATE DATA

When each observed unit provides more than two items of information, we have multivariate data. Such data arise very frequently, because often more than two numbers are required to describe the important features of the observed unit, be that a patient, a clinic, an experimental animal, or a research

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

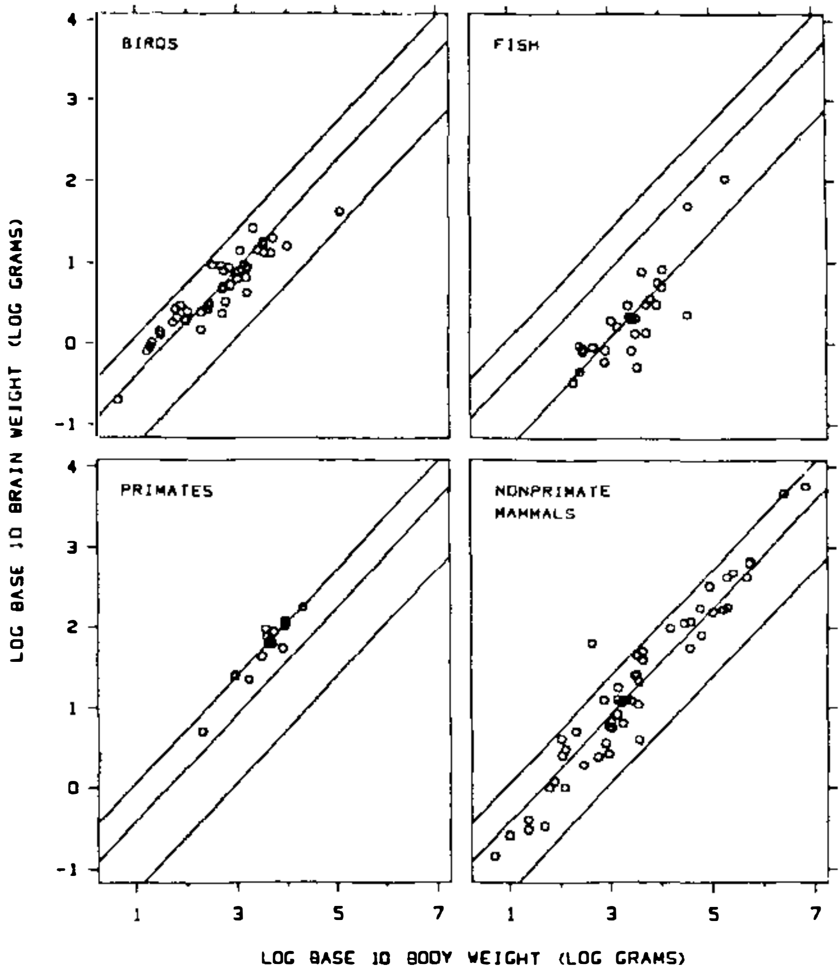


Figure 18 BRAIN WEIGHTS AND BODY WEIGHTS IN FOUR GROUPS OF VERTEBRATES. The coordinates are logarithmic on both axes. The fitted lines have slope 2/3, for theoretical reasons. The lines facilitate comparisons of the data on the different panels. Source: Ref. (4). Copyright 1985, Bell Telephone Laboratories, Inc. Reprinted by permission.

article. When the relations among the variables are of direct interest, bivariate and multivariate exploration and presentation become essential. Thus, a first source of interest in multivariate graphical methods is for finding and representing relations among the variables. A second source of interest grows out of the fact that information about several variables (rather than only one or two) can be helpful in identifying groups of similar observational units (as in distinguishing between rather similar diagnostic groups).

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

Multivariate data can be hard to deal with, in several ways. The "natural" graphical representation of k-variate data is as pooints in a space of k dimensions. For k equal to three, this is already hard to visualize, and for larger values of k the difficulties are multiplied. A table of the data with k columns and one row for each observation is another "natural" representation, but it cries out for effective condensation and summarization. Some help can be brought to the problem by using graphical methods, largely through extension of two methods we have already seen several times: the scatter diagram, and condensing several data items into a single one.

Relationships Among the Variables

Many studies aim at describing (or establishing) relationships between input variables like quality of care, intensity of treatment, or social support available to the patient, and outcome variables like degree of rehabilitation, functional level, life satisfaction, or health. Each of these named concepts defies direct measurement, and instead must be approached through collecting data on many variables that "belong" to the concept; thus, functional level might be captured through tests of agility, endurance, balance, comprehension, strength, etc. Multivariate data sets very commonly originate in efforts to measure some construct, like health, by observing several variables, each of which taps a part of the concept, and typically with some overlap and redundancy among them.

Many multivariate data sets comprise not only some variables that are inputs and some others that are outcomes, but also "interfering variables," which influence outcomes without being part of the inputs under study; examples might include patient's age, gender, and educational level. Thus, the study may involve three kinds of variables: input, outcome, and interfering. Each of these may have several components.

SOME STRATEGIC APPROACHES The tasks of conceptually organizing and then analyzing a data set with several input variables, several output variables, and several interfering variables are in large part a search for legitimate condensation and simplification.

Simplification Sometimes simplification can be achieved by finding that a variable is legitimately ignorable. It may be ignorable because it is redundant, offering information already supplied by other variables; it may be ignorable because it is irrelevant, unrelated to any outcome. The scatter diagram is a tool that may reveal either of these situations.

Simplification can sometimes be achieved by combining several related variables into a single one. If in a study of endurance and body size our data included both right arm length and left arm length, we would almost surely,

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

with no ado whatever, either average the two or use only one of them, say the left arm length. Similarly, with many measured lengths—height, span, foot length, leg length, and arm length—we might "condense" them all to a single measure of "skeletal length" by constructing an additive combination of all of them. We could do this by judgment (as with the two arms) or by recourse to some algorithm, say one producing the additive composite that gives maximum correlation with endurance. (That algorithm is simply multiple regression of endurance on the five length variables.) If a judgment composite is chosen, then directly averaging the variables is likely to be less satisfactory than first dividing each by its standard deviation and then averaging them; the reason is that a variable taking small values, like foot length, would contribute so little as to be almost ignored in a simple average, while division by the standard deviation gives each variable parity with the others.[3]

The algorithmic approach is not necessarily always preferable to the judgment composite. The judgment composite is more easily explained. The task of explanation with multiple regression can be especially uncomfortable if the regression gives a counter-intuitive sign to one of the variables in the composite. But, however the condensation is done, it does reduce a set of several variable to a single one, and each subject has an observed value for this new composite variable, a value that can be plotted in a scatter diagram against other variables, or composites of them.

Identifying variables that are (still) relevant If we are interested in predicting y from u we might well draw a scatter diagram in connection with either the analysis or the presentation. Now, if there is an additional possible predictor, v, how shall we assess whether it can improve the prediction? We begin with the case where v is a binary variable. In Figure 19 the upper left panel shows a rather strong relation between u and y.

The upper right panel is the same diagram, except that 10 of 22 points are darkened to identify the observations where v is at its high level; the 12 light points have v at its low level. This panel clearly indicates that y is related to v, in addition to being related to u, and it points to the possibility of improving prediction of y by taking v as well as u into account. The lower left panel is the same as the previous one, except that it shows two lines of common slope fitted to the dark and light points. The vertical distance between them is indicated near the right edge as d. One can imagine reducing the y value of every black point by d (thus taking account of the influence of high v) and

[3] Sometimes a composite is made from variables among which some point in opposite directions; thus, a clerical skills composite might include a spelling score, typing error rate, and reading comprehension. Then a reasonable judgement composite would not only divide each variable by its standard deviation but would also enter typing error rate with a minus sign.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

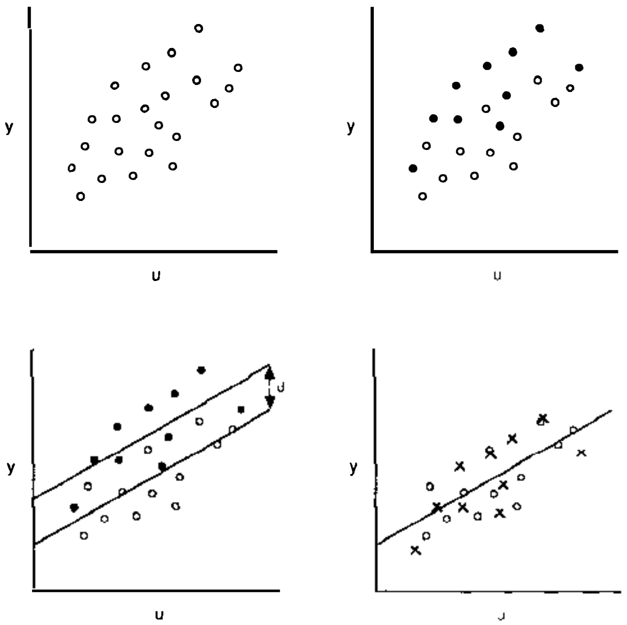


Figure 19 RELATION BETWEEN TWO VARIABLES, y AND u, AT TWO LEVELS OF A THIRD, v. Black dots indicate points at high v, and crosses represent them after downward adjustment by the distance d.

plotting these adjusted points in place of the original black dots. The lower right panel has been made in that way; the adjusted block dots are portrayed as x's among the original light dots. These data now show a much stronger relation between y and u, after adjustment for v, than do the original data shown in the first panel. Thus, using v in addition to u has improved the prediction of y. The value of v as a predictor was directly evident from the second panel, at the upper right. If those black dots had appeared randomly among the 22, then the message would have been that v did not hold promise of improving prediction.

In Figure 20 we show the same swarm of 22 points again. The upper left panel shows y plotted against u. The upper right panel repeats the identification of ten high points in the swarm as being the ones at the high level of v.

The two lower panels are new. The pattern at the left indicates that v has no additional information about y because the relation between y and u looks about the same for the dark points as for the light ones. (We interpret this as, "v is redundant if u is already being used as a predictor of u.") The lower right panel would tell quite a different story. Among the dark points, at high v, there is only a weak relation to be seen between y and u. Similarly, among the

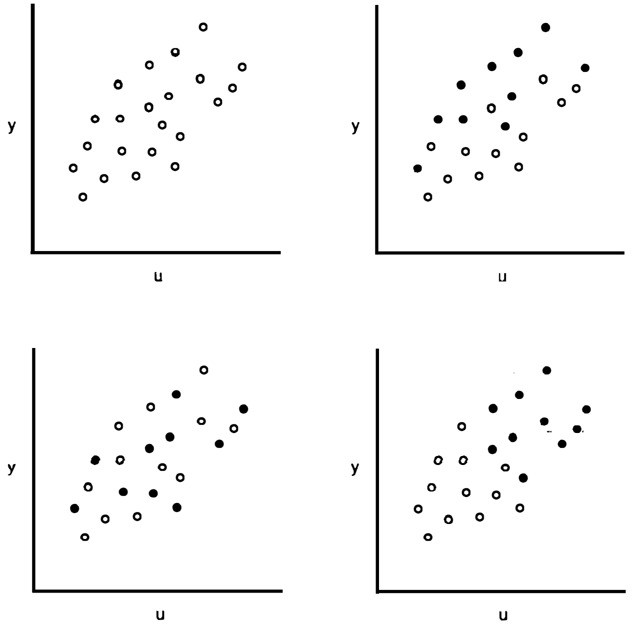


Figure 20 SOME POSSIBLE WAYS THAT THE RELATION BETWEEN TWO VARIABLES MAY BE AFFECTED BY A THIRD. The original relationship is seen in the upper left panel. The upper right panel shows that taking v into account will strengthen the relation (see Figure 19). The lower left panel shows v to be irrelevant, and the lower right panel indicates that the relation may really depend almost entirely upon v.

light points only a weak relation between y and u is to be seen. Identification of the high-and-low-v points has indicated that much, or most, of the original relation can be explained by observing that low-v people (perhaps males) have smaller values of both y and u than do high-v people (females), and in both those subgroups no strong relation between y and u exists. Something very like this would be expected in adults for y = weight lifting ability, u = head circumference, and v = sex, for in neither sex is head circumference much related to strength, but males are both stronger and larger-headed than females.

This device of marking a dichotomous identification on the points of a scatter diagram is immediately applicable to considering whether to adjust the data for sex, or race (black, white), or any other binary classification; further, a continuous variable can be broken into two classes, high and low, and its relevance can thus be assessed. It is possible to break a continuous variable into three or more classes, and that might have advantages in some circum-

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

stances. (For example, if intermediate values of v increased y above the levels associated with either high or low v, that fact would be more easy to find with a three-fold division of v.) If three or more levels are chosen, then a display like that of the brain mass-body mass example might be helpful.

Exploration of the kind we have been describing helps decide when to ignore a classification or a variable, with a gain in simplicity, and of course it can help in strengthening prediction by identifying additional useful variables. Any of the variables we have spoken of, y, u, or v, may be a composite variable. Thus, if y is a dependent variable and u is a regression-based composite, then the method could indicate whether adding a new variable v to the multiple regression would be important.

Groups of Multivariate Observations

CLASSIFICATION In a celebrated paper (6), Sir Ronald Fisher introduced the linear discriminant function. He developed the method to employ the four variables, petal length, petal width, sepal length, and sepal width, for classifying an iris blossom as belonging to the correct one of three species. The data set contained 50 blossoms from each of the three species; the data thus comprised three sets of 50 four-variable observations. He proposed an algebraic-numeric way of using those data to produce a composite of the four variables; it proved to be an effective univariate score for deciding the species of a new specimen. In Figure 21 are shown the petal widths and lengths for a portion of that data set, for ten members of each species (the ones in rows 1, 6, 11, . . . 46 of Fisher's data set in his Table 1.) The sepal data are not shown, because (*a*) two dimensions use up our ability to plot points, and (*b*) inspection of the tabular data indicates that the petal dimensions have less overlap among the species than do the sepal dimensions and so should help more in discrimination among species.

Two lines with the same slope have been drawn in by eye. They can be used for classifying a new specimen. One would measure its petal width and length and plot the point; the species classification would be detennined by the point's position relative to lines one and two. Equivalently, a score (corresponding to the slope of the lines) could be computed by the formula:

$$\text{score} = 2.5 \text{ width} + \text{length}.$$

If that score is less than 4.5, the blossom is classified *Iris setosa;* if greater the 9.5, it is classified as *I. virginica;* and if between those limits as *I. versicolor*. The correspondence between the geometrical recipe (plot the point) and the numerical one originates in the fact that the two lines are defined by the equations.

$$2.5 \text{ width} + \text{length} = 4.5 \quad \text{and}$$

$$2.5 \text{ width} + \text{length} = 9.5$$

for lines one and two, respectively. Points to the right of line two all have scores greater than 9.5, and those to the left of line one all have scores less than 4.5.

The idea in this example is that class membership may be identifiable by observing that different regions of the multivariate space tend to have data points belonging to different classes.

CLUSTERING A similar plotting notion can be applied to a quite different problem: Without knowing what underlying groups may exist, see whether

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

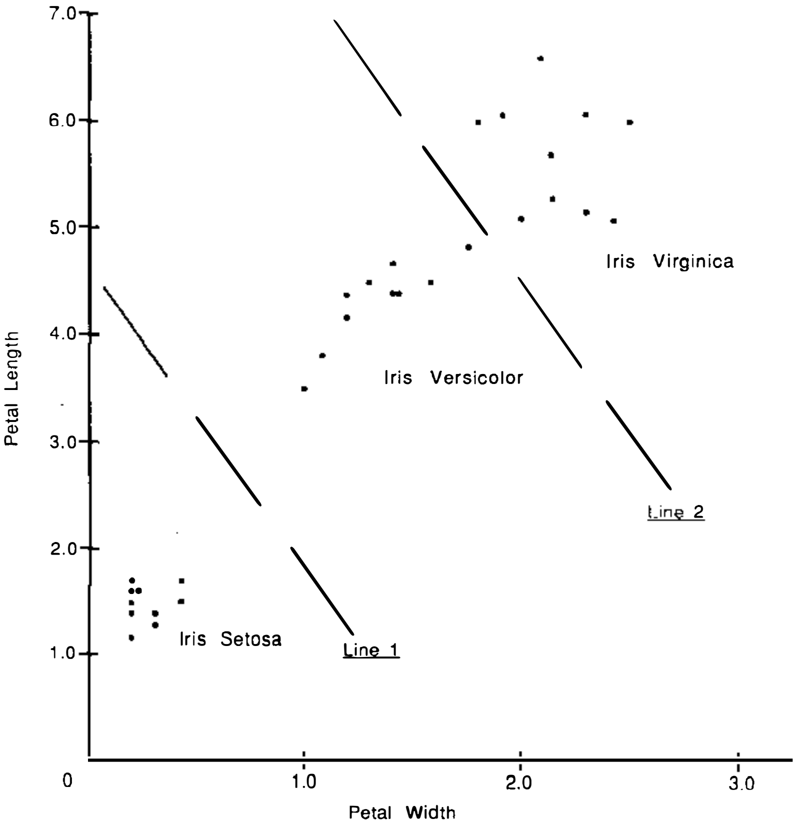


Figure 21 A SUBSET OF FISHER'S IRIS DATA. Fisher used four variables and 50 observations. This subset of the data uses only two of the variables and shows how the bivariate plot allows confident separation of the groups, in terms of a score that corresponds to the slope of the separating lines. Source: Ref. (6).

there appear to be clusters of data points in separated regions; such clusters offer themselves as possibly representing underlying groups. In this task the multivariate character of the data may be tapped by reducing the many variables to only two composite variables, and then displaying each observed unit as a point with those two coordinates.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

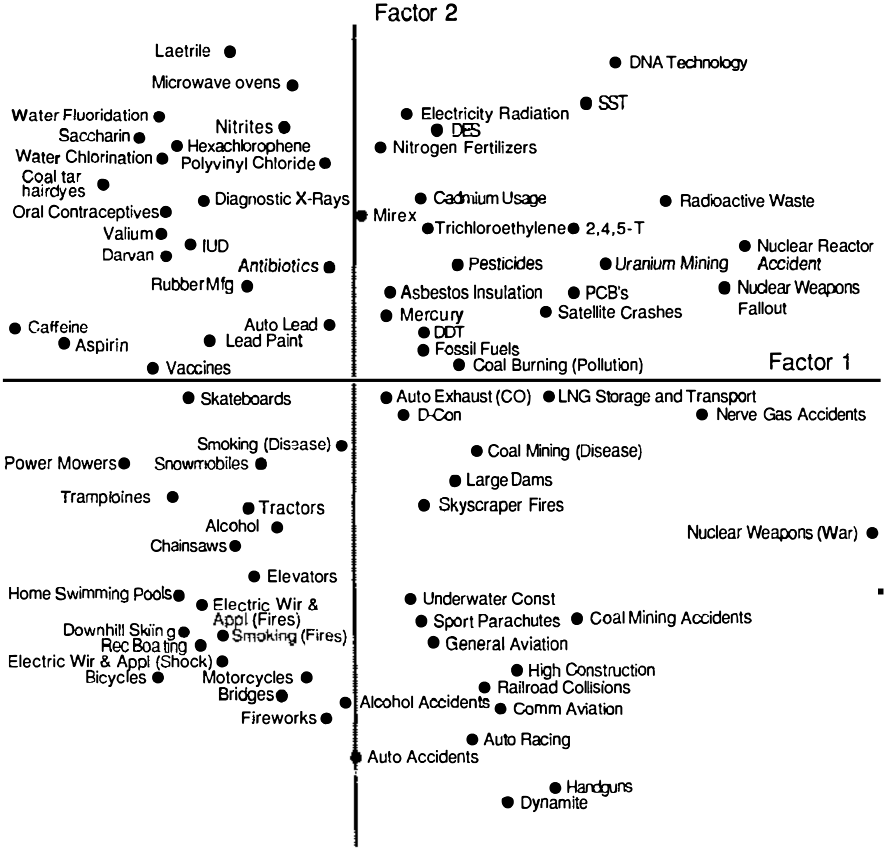
DISPLAY OF MANY MULTIVARIATE ITEMS Slovic et al (17) used composites in explicating people's attitudes toward several dozen hazards of modern living, like motorcycles, nuclear reactors, oral contraceptives, handguns, DDT, etc. The hundreds of respondents scored, on seven-point scales, their feelings about each hazard with regard to a battery of 16 issues, such as: Is the hazard uncontrollable or controllable? Is the effect immediate or delayed? Is exposure to it voluntary or involuntary?

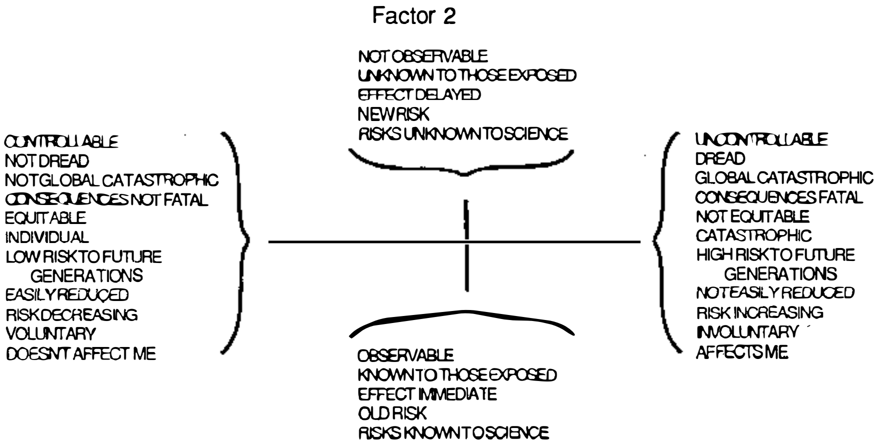
Evidently the data are very highly multivariate. Each of many hazards is scored for each of many attributes. Figure 22 shows the display that the investigators produced for exhibiting the summary information. First they applied factor analysis to the data set. Factor analysis algorithms produce one or more "factors," which are composites of the original variables; here the composites were made up of the scores associated with the 16 issues. The first two factors are the ones that most fully summarize the issue responses. The numerical value of each of the two composites was computed for each hazard. Then a point was plotted for each hazard, using as coordinates these two factor scores. Interpretation of the picture comes largely from the characterization of the factors, which appears at the bottom of the figure. Issues that receive large weight in a composite are listed, showing the direction in which the issue affects the composite's numerical value.

This ingenious display of such a large complex data set allows one to think of attitudes about hazards as depending very largely on judged severity of the hazard and on how unfamiliar it is (here we are over-condensing the factor descriptions to one-word labels) and further to see how the respondents assess each of the many hazards with regard to these two dimensions.

An entirely different approach to displaying many multivariate items is to construct a small diagram for each unit, with one diagrammatic attribute reserved for each multivariate component. A pioneer in this area was Edgar Anderson (1), who proposed "glyphs" for multivariate display. He represented each unit by a small circle, and then he assigned a position on the circumference to each variable; at such a position, a "whisker" would show by its length the magnitude of the variable for that unit. Several later developments are similar in spirit. "Chernoff Faces" (3) assign a feature of a cartoon face to each variable; its numerical value determines the size of the feature. This results in one "face" for each unit, with variations among faces displaying the multivariate numerical information. Other devices include trees and castles (10) and "stars." These last can be thought of a modified Anderson

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.





glyphs, with the whiskers equally spaced around the small disk, each whisker serving as the apex of a triangle (reaching back to the disk.) All these devices show the full numerical information for each unit (which is not true, for example, with the display of Figure 22). It is not clear, however, how effectively the mind copes with the information so presented. The writer's experience is that a large set of stars does not help him much to see patterns, but it is very convenient for quick access to the information in detailed scrutiny of thc data. An informativc, well-illustrated comparative treatment of stars and trees is given by Chambers et al (2).

It would have been possible to display the hazard data as several dozen stars, each with 16 points (or fewer, if some attributes were ignored). This alternative mode of display would provide greater detail of information but would leave the reader to search for patterns; the display offered by Slovic et al supplies a pattern or framework for perception by summarizing much of the information and ignoring the remainder.

PLANNING THE GRAPH

In this closing section we offer some suggestions that may help the maker of a graphical display. The most important single notion is that data often can be graphed in many ways, and a reflective choice among those will usually lead to better results than adopting the first way that comes to mind. We also present a handful of additional ideas:

1. Decide on the primary message of the graph.
2. Be mindful of the tradeoff between quantity of information and the probability of its being read correctly, or read at all.
3. Be mindful of the intended viewers' preparation, skills, and expectations.
4. Since some tasks of visual perception are more accurately done than others, design the graph to call upon the more accurate functions.
5. Be exploratory about the design, sometimes trying more than one approach, sometimes testing early drafts on friends and associates.

Deciding on the Primary Message

Suppose that we had data on the percentage of impurities found in several specimens, and that all these percentages were small, ranging from 1% down to 1/20 of 1%. How should the data for these specimens be depicted? It

Figure 22 TWO FACTOR DISPLAY OF HAZARDS OF MODERN LIVING. Sixteen attributes of each hazard were rated by hundreds of respondents. Those 16 variables were reduced by factor analysis to three factors; the two most important arc the coordinates in this chart. This one picture (partially) summarizes information that in tabular form would involve more than 1000 averages. Source: Ref. (17).

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

depends on the meaning of the data. If the impurities were dangerous or highly toxic, then the difference between 1% and 0.5% might be exceedingly important. Plotting the data as a percentage or log percentage impurity would accent the difference among the specimens. If the impurity were instead the percentage of infertile seed among strains of alfalfa, then all of the samples would be substantially equal, consisting of nearly 100% fertile seeds. That would be the main message, and it would be best conveyed by bars of nearly all the same length, nearly all 100%, rather than by bars showing impurity fractions. Formally, both 100*p*% and 100(1-*p*)% contain equivalent information, but the wrong one is likely to mislead the reader's attention.

Sometimes, to ensure that the main message gets through clearly, it may be wise to hold back supplemental information, or present it in a second related graph. For example, this choice forced itself on the writer in preparing Figure 11, which presented the helmet law data. It was attractive to add to the figure vertical error bars for each year and for each of the three multi-year periods. But both the discussion and the figure would have become considerably more complicated, at the expense of the main message for that figure, which was, "Suitable use of reference lines can help provide a structure for understanding a graphically presented data set." So, error bars and the attendant discussion were foregone in that chart.

THE TRADE-OFF BETWEEN INFORMATION AND EASE OF UNDERSTANDING

One chart can be loaded with enormous amounts of information, but the reader's task of making sense of the chart grows as the amount of information does. A rule that one investigator imposes upon himself in a closely related context, the preparation of a 2 × 2 lantern slide, is informative. Dr. Stephen Pauker holds a slide at arms length; if he cannot figure out its message by reading it in that way he rejects the slide as unsatisfactory and redesigns it (16).

Often a good alternative to packing too much detail onto one chart is to tolerate some redundancy (repetition of information) in replacing one chart by several. Figure 18 does this with the brain mass-body mass data. All four charts display the same three lines, but the reader gains enormously by that; "saving" redundancy by plotting all the data on one figure would effectively hide much that can be seen from the four-chart mode. With the radiosensitivity data (9), a different choice was made; four groups and four lines appear on a single chart. The two data sets differed sharply as to the overlap between groups, and so different graphical strategies were natural.

As remarked above, reference lines can be very helpful. But every element added to a chart can be thought of as clutter, unless it is definitely functional.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org by Monash University on 05/12/10. For personal use only.

The appropriate balance should be sought. Imagine a chart that depicts exponential growth at several different rates. The purpose of that chart really determines the appropriate level of detail for reference lines upon it. If the chart is intended to allow the user to *compute* from it, then both horizontal and vertical rulings will need to be much denser than if the chart's purpose is to show that "small" differences between growth rates, like .03 per year versus .01 per year, can exert profound differences over 70 years, when the one rate has produced an eight-fold growth and the other a two-fold growth. The second purpose is well enough served with very little in the way of reference lines; scales on the axes might be sufficient, without any reference lines at all.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

The Skills and Expectations of Intended Readers

If the intended readers are expert in the subject matter of a graphical display, some things are easier for the author: Technical terms can be used with less explanation; more complexity in the chart is likely to be tolerable. Less flexibility for the author may also result from an expert audience: Convention may almost require that certain transformations be used, or not used, or that certain ordered category boundaries be imposed on data originally acquired on a continuous scale.

Of course it is often true that a readership of both experts and nonexperts is contemplated. In that case technical terms that are routine for many readers may nevertheless require careful definition and explanation for the benefit of other intended readers. Examples that come to mind include "hazard function," "probit," "correlation coefficient."

If most readers are expected not to be expert then even more care should be taken to avoid misperception. Thus, *comment* may be given about the use of broken scales, or the fact that the origin is not part of the figure, or that on one axis the variable is plotted logarithmically, etc.

Take Account of the Relative Accuracy of Various Perceptual Skills

A quantity can be graphically represented in many different ways, for example as a length, area, volume, or intensity of color. An angle can be represented by two lines intersecting at that angle, or in some way that states its numerical value, say a length, area, or volume. Some of these depictions are more accurately perceived than others. Cleveland studied the matter systematically and offered these conclusions (4, p. 254). Accuracy declines as we move downward in this list of perceptual judgments:

1. position along a common scale
2. position along identical, nonaligned scales
3. length

4. angle—slope
5. area
6. volume
7. color hue—color saturation—density

Referring to this list we must conclude, for example, that comparing two years' budgets (of different total size) by means of "pie charts" (with different total areas) is an inferior way to go. Neither angle nor area is as easy to perceive accurately as length; reference to the list encourages us to seek a better way of portraying the two budgets—one using position on aligned scales, for example. In general, the higher up we can go in the list the more accurately the message may be perceived.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

Experiment with Different Representations

The biggest step of all is to consider more than one possibility. There are abundant reasons to do this, with so many dimensions of choice at hand: size, scales, style, transformations of variables, information density, etc. The thought-experiment may sometimes give a satisfying answer. At other times a person may wish to sketch out two or more alternative charts, and decide on the basis of those. If the intended readers may be naive it can be wise to prepare rather finished versions and expose them to naive(!) readers for comment and elicitation of preference.

Pretesting a questionnaire is a well established Good Idea, so is circulating a draft of a manuscript for comment. The message here is that pretesting alternative forms of an important chart can also be a Good Idea.

ADDITIONAL LITERATURE

This chapter begins with the observation that graphical methods are in a state of rapid development. The treatment here is necessarily eclectic and incomplete. The interested reader might pursue (*a*) historical developments (among other topics) in a recent review article (19), (*b*) graphical display in Tufte's book (18), (*c*) data analytic approaches in three recent books (2, 4, 7). All of these cite much additional literature.

Acknowledgment

The writer is grateful to Michael Shing for many useful comments and for producing the graphics in this article, and to Jomarie Ochoa for typing it. Comments on earlier drafts from Byron William Brown, Jr. and Frederick Mosteller helped improve this article.

Annu. Rev. Public. Health. 1987.8:309-353. Downloaded from arjournals.annualreviews.org
by Monash University on 05/12/10. For personal use only.

Literature Cited

1. Anderson, E. 1960. A semi-graphical method for the analysis of complex problems. *Technometrics* 2:387–92
2. Chambers, J. M., Cleveland, W. S., Kleiner, B., Tukey, P. A. 1983. *Graphical Methods for Data Analysis*. Belmont, Calif.: Wadsworth
3. Chernoff, H. 1973. The use of faces to represent points in *k*-dimensional space graphically. *J. Am. Statist. Assoc.* 68:361–68

3a. Cleveland, W. S. 1979. Robust locally weighted regression and smoothing scatter plots. *J. Am. Statist. Assoc.* 74:829–36

4. Cleveland, W. S. 1985. *The Elements of Graphing Data*. Monterey, Calif.: Wadsworth
5. Commission on Engineering and Technical Systems, National Research Council. 1982. *Toward Safer Underground Coal Mines*, p. 112. Washington DC: Natl. Acad. Press

5a. Emerson, J. D., Strenio, J. 1982. Boxplots and batch comparisons of chapt. 3. In *Understanding Robust and Exploratory Data Analysis*, e.d., D. Hoaglin, F. Mosteller, J. W. Tukey, pp. 58–93. New York: Wiley

6. Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugenics* 8:179–88
7. Gnanadesikan, R. 1977. *Methods for Statistical Data Analysis of Multivariate Observations*. New York: Wiley
8. Deleted in proof
9. Kaplan, H. S., Moses, L. E. 1964. Biological complexity and radiosensitivity. *Science* 145:21–25
10. Kleiner, B., Hartigan, J. A. 1981. Representing points in many dimensions by trees and castles. *J. Am. Statist. Assoc.* 76:260–76
11. Korner, A. F., Guilleminault, C., Van dën Hoed, J., Baldwin, R. B. 1978. Reduction of sleep apnea and bradycardia in preterm infants on oscillating water beds: A controlled polygraphic study. *Pediatrics* 61:528–33
12. Krane, S. 1981. Motorcycle crashes, helmet use and injury severity: Before and after helmet law repeal in Colorado. *Symp. on Traffic Safety Effectiveness (Impact) Evaluation Projects, May 29–31, 1981, Chicago,* 1981:330 (Table 1). (Conducted by National Safety Council under contract no. DTNH22-80-C-01564)
13. McNeal, J. E., Bostwick, D. G., Kindrachuk, R. A., Redwine, E. A., Freiha, F. S., et. al. 1986. Patterns of progression in prostate cancer. *Lancet* Jan. 11, pp. 60–64
14. Miller, R. G. Jr. 1980. *Simultaneous Statistical Inference* New York. Springer-Verlag. pp. 26–7, 90–94. 2nd ed.
15. Murray, J., Bernfield, M. 1986. The differential impact of pre-natal care on incidence of low birth weight among blacks and whites in a prepaid health care plan. *N. Engl. J. Med.* Submitted
16. Pauker, S. 1981. Presented in lecture before Health Services Consortium, Stanford Univ., 17 July 1981
17. Slovic, P., Fischoff, B., Lichtenstein, S. 1985. Characterizing perceived risk in *Perilous Progress: Technology as Hazard,* ed. R. W. Kates, C. Hohenemser, J. Kasperson. Boulder, Colo.: Westview
18. Tufte, E. R. 1983. *The Visual Display of Quantitative Information*. Cheshire, Conn.: Graphics Press
19. Wainer, H., Thissen, D. 1981. Graphical data analysis. *Ann. Rev. Psychol.* 32:191–241