# **RESEARCH METHODS IN PSYCHIATRY**

# Speaking Graphically: An Introduction to Some Newer Graphing Techniques

# David L Streiner, PhD<sup>1</sup>

The vast majority of graphs appearing in the psychiatric literature consist of the traditional line graphs, histograms, and bar charts. Over the past decade, new graphing techniques have appeared which make the data easier to read and which present much more information than simply group means and confidence intervals. These methods include horizontal bar charts, dot charts, stem-and-leaf plots, box plots, and notched box plots. This paper describes these new techniques, as well as older ones, such as smoothing, and warns against using some of the options found in graphics programs: 3-dimensional (3-D) graphs, stacked graphs, and pie charts.

(Can J Psychiatry 1997;42:388–394)

Key Words: graphs, statistics, data presentation

In the 1630s, René Descartes developed a radically new way of displaying data, a technique that we now eponymously call Cartesian coordinates. As it is currently used, an independent variable (IV), such as time of day, is plotted along a horizontal axis (the X-axis, or abscissa), and a point is placed corresponding to the value of the dependent variable (DV), such as melatonin level, on the vertical axis (the Y-axis, or ordinate). If we now connect the points, the result is a *line* graph, which shows how changes in one variable relate to changes of the other: in this case, how the melatonin level varies as a function of time. Another version of the graph, used mainly but not exclusively when the IV consists of categories, has a vertical bar, extending upwards from the X-axis, to indicate the value of the DV. This is called a *bar chart* if the IV is a noncontinuous, categorical variable, such as diagnosis, treatment group, or sex, and the bars have spaces between them; it is referred to as a *histogram* if the IV is continuous (for example, age) and the bars abut one another. (For more details about drawing these graphs, see Norman and Streiner [1].)

This new way of looking at the world was so revolutionary that not much happened for the next 300 years or so. If Descartes were still around, he would have no difficulty recognizing the vast majority of graphs that appear in psychiatric journals. Over the past 2 decades, however, a number of new techniques have been developed and are beginning to appear in the literature and in the output of many statistical computer programs. Some of these methods, such as stemand-leaf plots or box plots (which we will describe shortly) are "simply" attempts to convey more aspects of the data than do conventional graphs or to present them in more understandable ways (2); other methods, like horizontal bar charts, are based on empirical studies of human perception and try to minimize errors that people make in reading line and bar charts (3). The purpose of this paper is to introduce some of these newer methods, as well as to rail against some common but erroneous "improvements" in the traditional techniques, alterations attributable mainly to the wide availability of computerized graphing packages that are ill-designed for scientific work.

Manuscript received August 1996, revised January 1997.

This article is the twelfth in the series on Research Methods in Psychiatry. For previous articles please see Can J Psychiatry 1990;35:616–20, 1991; 36:357–62, 1993;38:9–13, 1993;38:140–8, 1994;39:135–40, 1994; 39:191–6, 1995;40:60–6, 1995;40:439–44, 1996;41:137–43, 1996; 41:491–7; and 1996;41:498–502.

<sup>&</sup>lt;sup>1</sup>Professor, Departments of Clinical Epidemiology and Biostatistics and of Psychiatry, McMaster University, Hamilton, Ontario.

Address for correspondence: Dr DL Streiner, Department of Psychiatry, McMaster University, 1200 Main Street West, Hamilton, ON L8N 3Z5 e-mail: streiner@fhs.csu.mcmaster.ca



Figure 1. A horizontal dot chart.

#### Variations on a Bar Chart

Bar charts seem so simple, elegant, and straightforward that there doesn't seem to be too much room left for improvement. What could be easier than a graph where the heights of the bars are proportional to some value? Actually, there are a number of ways to improve the picture. First, as we discussed in a previous article (4), adding error bars showing the width of the 95% confidence interval gives some indication of the precision of our estimate and allows us to do "eye-ball" tests of the data. Another change doesn't quite turn the world on its head, only on its side. Drawing the graph so that the bars project horizontally from the left side accomplishes 2 things. First, a traditional, vertical bar chart can get very messy below the axis if there are many groups; the legends under each bar can run into each other, unless they are turned 45 or 90 degrees, which then makes reading them difficult. Placing the legends on the left side of the chart eliminates this problem, although at the expense of leaving less room on the page for the bars themselves. The second advantage of horizontal bar charts is based on research about perception. People can more accurately gauge the magnitude of the differences among the bars if the numerical axis is on the bottom than if it is on the side (3).

Table 1. A stem-and-leaf plot of the ages of 69 people														
0	1	1	2	3	3	7	8							
1	0	0	1	2	4	4	4	5	8	9				
2	0	2	2	3	3	4	5	7	7	8	9			
3	1	3	4	5	5	6	7	8	9	9	9	9	9	9
4	2	3	3	5	6	6	7	8						
5	0	1	1	2	4	5	5	7	9					
6	0	0	2	3	5	5	6							
8	4													
9	8													
10	4													

Another variation of the horizontal bar chart is the horizontal dot chart, devised by Cleveland (3). Instead of drawing the entire bar, a large dot is placed at the end, often with smaller dots leading up to it, as in Figure 1, which shows the relative use of different psychological tests. If there aren't too many data points (a number determined more by esthetics than by counting), the smaller dots can be omitted. When the graph has a meaningful zero point on the left, then the smaller dots should start at the axis and end at the large dot. If the left axis does not begin at zero, however, the visual impression could be misleading, since the relative lengths of the lines do not reflect true ratios. For example, if all the values in the graph fall between 310 and 350, you would not want to start the graph at zero, since most of the graph will be empty, and all the end points would cluster near each other, making it difficult to detect differences among them. It would make more sense to start the axis at 300, but a line from the base of 300 to a score of 330 would appear on the graph to be 3 times as long as the line going from 300 to 310, even though the difference between 310 and 330 is less than 7%. To minimize this misleading visual effect, the small dots should start at the arbitrary base, go through the large dot, and end at the right-hand side of the graph.

Histograms, which are used primarily to display continuous data, usually have between 5 and 15 bars, mainly for esthetic reasons (1). This means that if the variable we are plotting, like age, has more than this many values, we have to combine categories, such as plotting the number of people who fall within each decade, rather than each year. But this involves a trade-off; what we gain in appearance we lose in information. The graph may tell us that there are, for instance, 47 people in the sample between the ages of 30 and 39, but we cannot tell how many were exactly 30, how many were 31, and so on. (Of course, a very large peak at 39 may alert us to the fact that some people are incapable of counting above this number, especially when it's their years on this planet that they're reporting.)



Figure 2. A 3-D bar chart.

To eliminate this problem, Tukey (2) devised a form of a horizontal histogram called a stem-and-leaf plot. Continuing with our example of age, let's take a look at Table 1. The left-hand column, which is the stem, indicates the decade (that is, the most significant digit, both mathematically and, for age, psychologically), so that the row which has 0 in this column reflects the number of people between birth and 9 years of age, the row with a 1 in the left column counts the number of people between 10 and 19, and so on. Then, the numbers to the right, the leaves, are the least significant digits of all people in that decade. Reading across the first row, we see: 0 1123378, which means that there are 2 children who are 1 year old, one 2-year-old kid, two 3-year-olds, and one child each ages 7 and 8. If each digit takes up the same amount of space (you should not do this with a "smart" word processor, which uses proportional spacing and assigns more room to an 8 than to a 1), then the result is both a horizontal histogram, as well as a table preserving the original data. If the sample is very large, resulting in very long strings of numbers, you can divide each stem in half; that is, you can have 2 rows of 1s, for instance, with the top row reserved for ages 10 to 14, and the bottom for 15 through 19.

Over the years, there have been a number of refinements to the stem-and-leaf plot that have appeared in various computer programs. For example, to the left of the stem, you can indicate the *cumulative percentage* of people: for the first row, it would be the percent of people who fall in that range; for the second row, it would be the percent of people who fall in the first and second rows; and so on. This makes it easy to determine where the 50th percentile is, or where the 5th and 95th percentiles are. To make life even easier, some programs



Figure 3. A stacked bar chart.

put an asterisk next to the stem that contains the 50th percentile; how much more can a person ask for?

Stem-and-leaf plots haven't shown up too much in articles, possibly because they aren't as pretty as histograms. You will, however, encounter them in almost every major statistical program, such as Minitab, SPSS, SAS, and BMDP.

Before we leave the area of histograms, bar charts, and their variants, let me inveigh against 3 "modifications" that should be avoided: 3-D graphs, stacked graphs, and pie charts. Computerized graphics programs have become extremely popular, rivalling computerized poker or solitaire as ways of keeping jaded academics amused. Unfortunately, they're designed for people (who shall remain unidentified, but who usually have MBA following their names) for whom style takes precedence over substance.

It's easy to see the allure of 3-D graphs: you can fool around with the depth and colour of the shading and pretty them up in other ways. Unfortunately, they are as misleading as they are "sexy." Take a look at Figure 2; what is the value of the middle bar? The answer is 60, but your eye is drawn to the uppermost edge, which is closer to 70. The greater the 3-D effect, the more your eye will be deceived. So, use 2-dimensional graphs in articles and talks, and save those 3-D graphs for when you're making a budget presentation to administration (oops, we let slip who the nameless ones are).

Another variant of bar charts regrettably made possible by graphics programs and popularized by newspapers and magazines is the *stacked bar chart*, which is shown in Figure 3. At first glance, this appears to be a useful graph, since we can present, in this case, 5 different values for each of the 3



Figure 4. Pie charts.

groups; quite a lot of information in one figure. The difficulty arises when we try to compare categories across groups. There is no problem with the bottom category (single), since the base is the same for all groups and we can easily tell for which group that segment of the graph is highest. But are the numbers of widowed people the same in all 3 groups, and if not, in which group are they highest? Now the comparison is more difficult. Because the lower 2 categories have different numbers in each group, the bottom of the widowed segments start at different places for each group. We have to mentally move them to a common baseline and then compare the heights. This becomes progressively more difficult to do as the number of groups and the number of categories increase. Again, it looks sexy, but it can be quite misleading.

A similar problem exists with pie charts. A single pie chart, showing the proportion of people in different categories, is fine and can be quite informative. The problem begins as soon as we try to compare 2 pie charts, as in Figure 4. If comparable segments in each pie start with one edge of the wedge exactly at the 12 o'clock position, and extend in the same direction (that is, clockwise or counterclockwise), then, like the bottom segment of a stacked bar chart, we can compare them without too much difficulty. If they begin at different places in their respective pies, however, as with the segment labelled "Widowed," then it becomes almost impossible to see if their areas are equal or not. Some people try to get around this problem by also printing out the actual number or percent of cases inside the wedges. This simply turns the graph into 2 funnylooking, round tables, which is redundant: just present the table and make the journal editor happy, since it costs less money to print a table than a graph. The bottom line is, if you need numbers in order to make the graph tell its story, then

either you don't need a graph or you're using the wrong kind of graph (such as stacked bar charts or pie charts).

#### **Box Plots and Notched Box Plots**

If we use a bar chart to display the mean age of people in a treatment and a control condition, for example, we are showing only one piece of data per group-the mean (or median, if the data are ordinal or highly skewed). We can go a step further and add bars indicating the confidence intervals, but we are still showing only a small amount of the data. Can we convey more information about the groups and still keep the graph readable? Obviously, the answer is yes, or we wouldn't have bothered to ask the question. Let's return to the data in Table 1, which lists the ages of 69 people. To begin with, instead of starting the bar at the bottom of the graph and extending it upward to the mean, let's draw a short horizontal line at the median of the group (for those who have forgotten, the median is the value that divides the group, so that half the people have higher values and half have lower ones). In this case, the median age is 37; 34 people are younger and 34 are older. Next, let's look at those 34 more mature (the euphemism for "older") people and find the median for them. This is at age 51, so we draw another short, horizontal line (which we call the upper quartile, or Q<sub>U</sub>) at this level. Finally, we'll do the same thing for the younger 50% of the people, drawing a line for the lower quartile  $(Q_L)$  at their median, which is at age 19.5. We have thus divided the group into fourths, with roughly an equal number of people in each quarter: 25% above Q<sub>U</sub>; 25% between the median and Q<sub>U</sub>; 25% between the median and  $Q_L$ ; and the remaining 25% below  $Q_L$ . If we now draw vertical lines joining the upper and lower quartiles, we'll end up with a rectangle, which is labelled "Box" in Figure 5. What we've accomplished so far is to display the median of the group (37 years); where the upper quartile falls (age 51); where the lower quartile is (age 19.5); and what the interquartile range (IQR) is, namely, the difference between the upper and lower quartiles (that is, 51 minus 19.5, or 31.5). By definition, the IQR includes the middle 50% of the subjects. (The IQR, by the way, is the measure of variability used with the median, in the same way that the standard deviation is used with the mean.)

This has increased the informational value of the chart quite a bit, but we can go even further by drawing "whiskers" above and below the rectangle. Based on some arcane math which we won't bother to elaborate on here but which is explained in more detail elsewhere (2), 95% of the sample falls within the range of scores of the median  $\pm 1.5 \times IQR$  (in our example, that is  $37 \pm 47.25$ ). By convention (and for the sake of confusion), we don't actually draw the lines at these levels; the upper line is drawn to correspond to the largest actual value in our data that is below this higher limit, and the lower line is at the smallest actual value above the lower limit. The upper limit is 37 + 47.25, or 84.25 years. If we go back



Figure 5. The functional anatomy of a box plot.

to Table 1, the highest age in our data below 84.25 is 84, so that's where the line is drawn. The lower limit is 37 - 47.25, or -10.25 (obviously a very young person), and the first real value above this is 1, so that's where we draw the lower line. In the somewhat convoluted language of box plots, those horizontal lines at the ends of the whiskers are called the inner fences. Next, we define (but don't draw) the outer fences, which are 3.0 times the IQR, and put some symbol, such as a circle, for each subject whose score falls between the inner and outer fences. These cases (there were 2 in our example) are called outliers and have scores that are below the 5th or above the 95th percentile for that group. Extreme outliers, whose scores fall below the 1st or exceed the 99th percentile and lie outside the outer fence, are identified with a different symbol, such as an asterisk; we didn't have any in our data set. (Just to confuse you even more, every computer program uses a different set of symbols or even letters, but they are usually labelled.) Some computer programs even print the case numbers of the outliers and extreme outliers. This makes it extremely easy to identify subjects with very deviant scores, so that the researcher can determine if the data are real or reflect a coding or data entry error.

If more than one group is shown on a graph, we can display even more information: the sample size. This is useful for a number of reasons: estimates of the mean and standard deviation are usually more accurate when based on larger samples,



Figure 6. A notched box plot.

and if groups do not differ, it may be due to a lack of power caused by too few subjects (5). Seeing the differences in sample size across the groups helps us to make these determinations. The sample size is reflected by making the width of the box proportional to the square root of the number of subjects. The square root is used more for perceptual than for mathematical reasons; if one group is twice the size of another, drawing the box twice as wide would actually make it look 3 or 4 times larger, since we are representing a onedimensional number (sample size) with a 2-dimensional figure (the box).

Even with all the information we have displayed so far, 2 important pieces are still missing: the mean and the 95% confidence interval (CI) around the mean. We can easily accommodate these with a variant of the box plot called the *notched box plot*, which is shown in Figure 6. The point of the notch falls at the mean, and the height of the **V** corresponds to the 95% confidence interval, which is defined as:

95% 
$$CI = \overline{X} \pm 1.96 \times \frac{SD}{\sqrt{n}}$$

where X is the mean, SD the standard deviation, and N is the sample size.

As you can see, box plots and notched box plots can display a large amount of data in a small space: 1) the mean; 2) the 95% confidence interval around the mean; 3) the

May 1997



Figure 7. The effect of moving averages.

median; 4) the skewness of the data, by how much the mean is displaced from the middle of the box; 5) the upper quartile; 6) the lower quartile; 7) the interquartile range, within which 50% of the subjects fall; 8) the distance between the fences, which spans 95% of the subjects; and 9) the number and values of outlying data points. These graphs are becoming more common in statistical and psychological journals, and will hopefully soon find their place in psychiatry.

## Getting Rid of Those Ugly Bumps

No, we haven't changed the focus of this article to look at weight-loss procedures. The topic is much more mundane and prosaic (and also much easier to accomplish): how to smooth out graphs in order to uncover trends more easily. This technique is used most often with measures taken repeatedly over time, such as a person's melatonin levels over a course of 24 hours or the diurnal variation in subjective mood. The problem is that any trend in the data may be obscured by errors in measurement, errors attributable to limitations of the measuring tool, normal biological variation, inaccurate recording, and so forth. To illustrate the effectiveness of this technique, we'll use some data for which we know ahead of time what we should see—the 11-year sunspot activity cycle (6).

The top part of Figure 7 shows the number of sunspots per month, averaged over a 3-month span (that is, the first dot is the average of January, February, and March of 1747; the

Table 2. Ave	erage monthly sun	spot activity in	3-month blocks
--------------	-------------------	------------------	----------------

Year	Quarter	Number of sunspots	Moving average 3	Moving average 5
1749	1	44.2		
	2	74.7	66.0	
	3	79.0	86.7	66.8
	4	106.4	71.7	76.5
1750	1	29.7	76.3	80.2
	2	92.8	71.9	78.0
	3	93.2	84.7	68.1
	4	68.1	72.6	73.3
1751	1	56.6	60.2	64.7
	2	55.9	54.1	52.5
	3	49.9	45.9	49.3
	4	31.9	44.6	48.5

second dot is the average of the next 3 months; and so on). The data do tend to follow a sinusoidal path, but it's hard to discern the pattern among all the point-to-point variability. We'll try to reduce some of that clutter with a technique called *moving averages*. The raw data, which are plotted in the top part of Figure 7, are in the "Number of sunspots" column of Table 2. The next column, "Moving average 3," is where the smoothing begins. The first number in this column (66.0) is the average of the first 3 numbers of the previous column (44.2, 74.7, and 79.0); the second value (86.7) is the average of the second (74.7), third (79.0), and fourth (106.4) raw data points; and so on down the table. The results of this averaging are shown in the middle part of Figure 7.

As you can see, the sinusoidal shape is more evident, because the large fluctuations have been reduced considerably. The reason is that if one data point is discrepant, its effect is lessened by averaging it with 2 other (hopefully more typical) values. This is readily apparent with the fifth data point; it is 29.7, while those on either side of it are 106.4 and 92.8. In the top graph of Figure 7, it is responsible for the sharp downward spike at the left, which has been virtually eliminated in the middle graph. If we feel that the data are still too "lumpy," we can smooth them further by averaging over 5 values, which is done in the "Moving average 5" column of Table 2 and shown in the bottom part of Figure 7. Now the pattern appears very clearly.

Should we then average over 6 or 7 numbers? Obviously, there is a trade-off. We can reach a point where we are smoothing out "bumps" that represent useful data rather than error. In fact, if we averaged over 15 numbers in this data set, even the 11-year cycle would disappear. So, we have to rely on our eyeball and clinical judgement to tell us when enough is enough and when any further smoothing would result in the loss of meaningful information. The second negative feature is that we lose some data points. When we average 3 numbers, we lose 2 data points; in general, if we are averaging *k* values, we lose k - 1 points. If the data set is large, this may not matter

too much, but if we start off with only 15 points and average over 4 values, we would lose 20% of the data.

#### Summary

In the past, we were fairly limited in the number of ways in which we could present our findings. Newer techniques allow us to show the data so that they are less prone to misinterpretation (the horizontal and dot charts); to minimize the amount of data which are lost through graphing (stemand-leaf plots); to present a more complete description of the data (box plots and notched box plots); and to reduce the visual clutter due to variation (smoothing). Especially since many statistical programs can output these graphs directly into drawing programs, psychiatric researchers should learn to use them and, in many cases, replace the more traditional but less informative bar and line charts.

#### References

- Norman GR, Streiner DL. Biostatistics: the bare essentials. St Louis: Mosby; 1994.
- 2. Tukey JW. Exploratory data analysis. Reading (MA): Addison-Wesley; 1977.
- 3. Cleveland WS. The elements of graphing. Monterey (CA): Wadsworth; 1985.
- Streiner DL. Maintaining standards: differences between the standard deviation and standard error and when to use each. Can J Psychiatry 1996;41:498–502.
- Streiner DL. Sample size and power in psychiatric research. Can J Psychiatry 1990;35:616–20.
- Andrews DF, Herzberg AM. Data: a collection of problems from many fields for the student and research worker. New York: Springer-Verlag; 1985.

### Résumé

La très grande majorité des graphiques figurant dans la littérature psychiatrique est composée des classiques graphiques linéaires, d'histogrammes et de diagrammes à barres. Au cours de la dernière décennie, de nouvelles techniques graphiques sont apparues, facilitant ainsi la lecture des données et présentant beaucoup plus d'information que de simples moyennes de groupes et des intervalles de confiance. Ces méthodes comprennent les diagrammes à barres horizontales, les diagrammes de dispersion, les schémas arborescents, les tracés en boîte et les tracés en boîte crénée. Dans cet article, on décrit ces nouvelles techniques, ainsi que d'autres, plus anciennes, comme le lissage, et on met le lecteur en garde contre l'utilisation de certaines options des programmes graphiques : graphiques tridimensionnels (3D), graphiques empilés et camemberts.