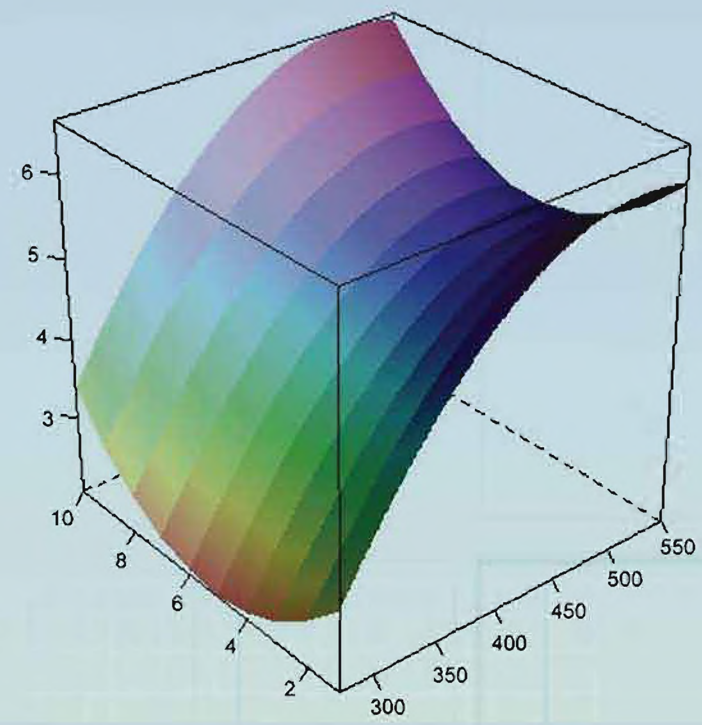
Presenting Scientific Data

Zaid Abdul-Hadi



ICARDA

International Center for Agricultural Research in the Dry Areas

Presenting Scientific Data

Zaid Abdul-Hadi

ICARDA

International Center for Agricultural Research in the Dry Areas

Citation
Abdul-Hadi, Z. (ed.) 2009. Presenting scientific data. International Center for Agricultural Research in the Dry Areas. ICARDA, Aleppo, Syria. iv + 28 pp.

ISBN 92-9127-221-3

International Center for Agricultural Research in the Dry Areas (ICARDA)
P.O. Box 5466, Aleppo, Syria.
Tel: (963-21) 2213433
Fax: (963-21) 2213490
E-mail: ICARDA@cgiar.org
Website: www.icarda.org

About this manual
This manual aims to provide reference material to supplement a training course on Data Presentation and Analysis. It focuses on the use of tables and graphics (created using simple software packages) to present statistical data in a simple, attractive way. The information in this manual was compiled from various published and online sources.

CONTENTS

1 DATA AND GRAPHS

Data analysis comprises basically four steps:

- Looking at the raw data
- Exploring the data
- Analysis
- Presentation

The preliminary examination of most data is facilitated by the use of diagrams. Diagrams prove nothing, but bring outstanding features to the eye; they are therefore no substitute for such critical tests as may be applied to data, but are valuable in suggesting such tests, and in explaining conclusions founded upon them.

RA Fisher
Statistical Methods for Research Workers

Graphical aids may often reveal aspects that might escape notice if the data are presented as numbers in a table. Graphical presentation may help to uncover features of the data that were totally unanticipated prior to the analysis, or to suggest hypotheses which may be further investigated. A mass of figures, even when neatly arranged in columns and tables, only reluctantly reveals its secrets.

A vital part of research is to disseminate the results to other research workers, to students, and to those in a position to implement the findings in a practical way. Unless this dissemination is carried out efficiently, the results will not become widely known or used and much of the time and money expended on the research will have been wasted.

The dissemination of research results can take place in a number of ways, but in most cases some form of visual presentation of all or part of the information is necessary. Printed materials in the form of reports, journal articles (including on-line journals) and books are perhaps the most commonly used medium for publicizing research, and here the presentation is entirely visual. At conferences and in other lecture situations, visual aids such as electronic slides and overhead projector transparencies are used as an essential complement to the spoken word. Posters have also become popular as a means of presenting research results. In this case, a visual summary is presented under circumstances where informal discussions with the author are possible. In teaching situations, CDs, computer-based teaching (e-learning) and PowerPoint presentations are useful visual aids, and in these cases visual information is used in conjunction with the spoken word.

The effective communication of information in visual form, whether it be text, tables, graphs, charts or diagrams, requires an understanding of those factors which determine the legibility, readability and comprehensibility of the information being presented.

- **Legibility:** can the data be clearly seen and easily read?
- **Readability:** is the information set out in a logical way so that its structure is clear and it can be easily scanned?
- **Comprehensibility:** do the data make sense to the audience for whom they are intended? Is the presentation appropriate for their previous knowledge, their present information needs and their information processing capacities? For example, a series of slides on a technical subject for a lay audience is likely to differ somewhat in terms of content and design from a similar series intended for an audience already familiar with the subject. What is comprehensible to the latter may be nonsense to the former.

Graphical techniques have become that much easier to use with sophisticated computer hardware and software including high resolution monitors, electronic data projectors, interactive terminals, intelligent graphic workstations, Windows-based software and efficient display algorithms.

Traditional statistical techniques relied on static and procedural methods. Current methods rely very much on interactive and visual techniques, including dynamic exploration of data in an iterative manner – viewing, analyzing, reviewing and reanalyzing the data. Three dimensional and color displays have developed considerably, opening new avenues for visualizing data.

1.1 Raw Data

Graphical examination of raw data has long been regarded by statisticians as an extremely useful technique and an essential prerequisite to more formal analyses. Let's take a typical experiment from the agricultural research field. The first step is to view the population to get some idea about the distribution of the univariate data. A histogram would therefore be very useful.

The following data are responses of the crop "chickpea" to gamma radiation treatment expressed as weight of dry matter in kg/ha

113 91 84 76 73 64 60 56 50
50 47 45 44 43 42 42 41 39
36 35 35 33 32 32 31 31 31 31
30 30 29 28 28 28 28 27
27 26 26 26 25 22 22 22 22
22 21 21 21 19 19 19 18 18
17 17 17 16 16 16 16 16 15
15 15 14 14 13 13 13 12 12 12
10 10 10 10 10 9 8 8 7 7 7 7 7

Clearly, looking at these figures by themselves will not give us much insight on the effect of the treatment. Plotting the histogram, however, will give a clearer picture of the frequency distribution, as shown in Figure 1a.

We can go on to plot the best curve fit through the frequency distribution to get an idea of the normality or otherwise of the distribution. This is shown in Figure 1b. This process of exploring the raw data should always be undertaken.

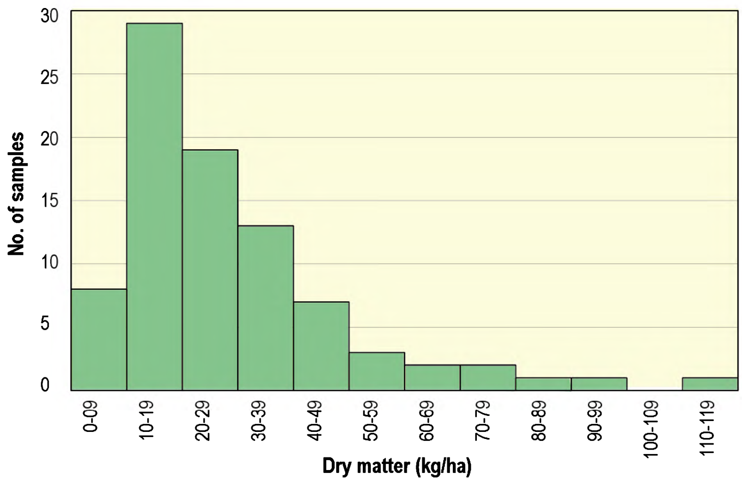


Figure 1a. Chickpea yield response to gamma radiation

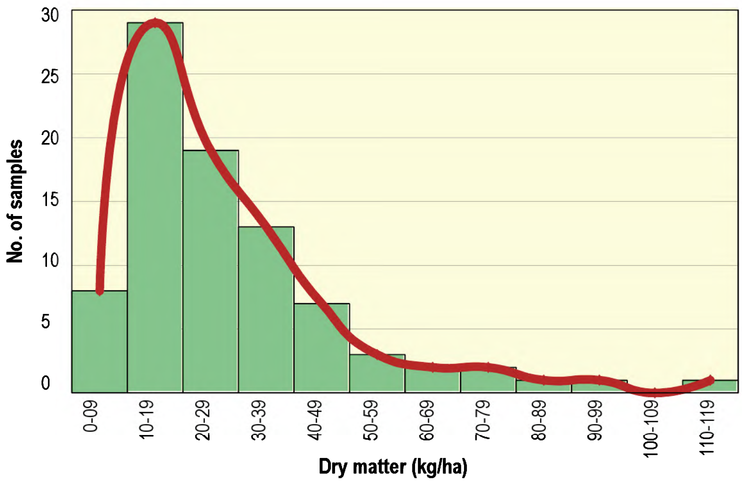


Figure 1b. Chickpea yield response to gamma radiation

1.2 General Principles of Data Presentation

1.2.1 Displaying data

Data are facts that have been collected systematically. In science, data are often numerical and thus are well suited to being displayed in tables, graphs, or maps. These forms are more concise than the written word. They can also be more forceful – if the form suits the purpose and if the display is well designed and executed.

A few general principles hold for all these forms of display. However, because each form organizes data in a different way, some guidelines are outlined for designing the individual types of each form. The guidelines are intended to be applied flexibly and to encompass a range of possibilities. Thus, in the illustrations in this chapter, details vary from one graph or map to another. Not all possible variations are represented.

1.2.2 Selecting the form for a display of data

Before selecting the form for display of data, the author must decide on the purpose of the display. The purpose of the display dictates the form.

- Use a table when exact values are important.
- Use a graph when trends or relationships are more important than exact values; when the meaning of the data needs to be quickly and forcefully expressed; when hidden relationships need revealing.
- Use a distribution map when the location of data is more important than actual numerical values.
- Use an algorithm when the succession of steps used in problem solving needs to be displayed.
- Use a flow chart when processes, sequences, or systems need to be presented in an organized fashion.

1.2.3 General principles

Accuracy. The value of any display of data depends on the integrity and care with which the data were collected and analyzed. No table, graph, or map, however carefully designed and beautifully executed, can redeem poor or inaccurate data.

Selectivity. The number of displays should be limited to the fewest that will cogently express the message of the work. It is more difficult for the reader to pull together the message from 17 tables, graphs, and maps than from five. Similarly, there is a limit to the amount of information that can be conveyed intelligently by any one table, graph, or map. Only the information that is necessary for conveying the message should be included. For example, data from one experiment can be displayed, but data from a parallel experiment yielding similar results can be summarized in the text.

Non-redundancy. Data should be presented only once. The same data should not be presented in both a table and a graph, or in both a table and a map.

Consistency. Similar data should be displayed in similar form. Alternating between graphs, maps and tables merely for variety is distracting.

Focus. A table, graph or map should make a point. The point should be apparent from the design of the display and, whenever possible, should be stated in the title. One point is usually enough. However, two or three points may be made if the display remains simple.

Clarity and simplicity. A table, graph or map should be clear and simple. Undue complexity of data or of explanatory detail, awkward word choice or cryptic abbreviations make the point difficult to grasp. Similarly, in graphs and maps, careless choice and arrangement of design elements such as label and keys cause visual distraction. If the science is complex, the message should be conveyed as simply and clearly as possible. If the science is simple, it should not be made to appear complex.

Visual effectiveness. Graphs and maps are visual media. For graphs and maps to be visually effective, words should be kept to a minimum. Information that cannot be conveyed visually should be omitted.

Impact. A graph or map should be visually convincing. For example, if the point of a graph is that a variable increased, the increase should be readily apparent.

Independence. A table, along with its title and footnotes, and a graph or map, together with its legend, should be understandable without reference to the text.

1.2.4 Preparing tables, charts and graphs
The most important consideration is that charts and graphs be legible – easy to read and easy to interpret. Legibility should never be sacrificed for beauty – achievement of both is a bonus. Some factors involved are: words (number, size, style, spacing); lines (width); shape and arrangement of objects and spaces; value and/or color of background and objects; viewing distance.

Words. Lower-case (small) letters are generally easier to read than upper-case (capital) letters.Short titles can be upper case, but if a title has more than five words lower-case letters are preferable.

Numbers. Individual specifications will be discussed later.

Size. Size depends on viewing distance.

- Use 9-, 10-, 11-, or 12-point type for optimum legibility in printed text.
- Use 11- or 12-point type if the typeface you prefer has a small "x-height" (small lowercase letters).
- Do not use 6- or 7-point type for regular reading material.
- Avoid very small or very large type as they slow reading speed significantly. If you must use 8- and 14-point type, use them sparingly as

Style. Choose a legible letter style. Use a serif typeface for text matter, as style choices and size permit. Readers generally prefer serif styles, such as Times Roman, Garamond, Goudy, etc. Use a sans serif typeface, such as Helvetica, for headlines, visual presentations and signage or for text matter. Several good choices are available.

Try not to mix styles. A type family like Ariel or Times New Roman, with its many different weights and widths, is thus recommended. A bold Ariel can be used with a condensed Ariel or with any other Ariel letter as the style is the same. Avoid overly condensed or very bold or thin letters.

Lines. Lines should not be too thick to confuse or too thin to see.

1.2.5 Shape and arrangement of objects and spaces
Good design can make even the most rigid chart more legible.

Simplicity. Each graph should present one idea only in an easy-to-read and easy-to-understand style. All extraneous information should be deleted.

Emphasis. By use of size or color or shape, one element should be made the center of attention. In a graph, the most important elements are the title and the data. The title has the largest lettering. The curves have the widest line.

Unity. Each element should function as part of a whole rather than as a separate entity. This can be achieved by grouping elements or by creating obvious similarities of line or shape or color.

Balance. Elements should be arranged to create equilibrium. The title or key can often be used to balance an otherwise unstable graphic.

Color. Color is expensive in publications, but generally more effective. Black and white charts and graphs usually are cheaper to reproduce. However, when you do use color, use it for a specific reason and use it well, never just to make the graphic pretty. Use it to add information or to make an object recognizable, to direct attention, to create emphasis, or to differentiate.

Every color has three properties: hue, value, and chroma. Hue is the variation of color as seen in the spectrum, e.g., red, green, violet. From pigments of the three primary hues (blue, red and yellow) many others can be made. The value or tone of a color is its lightness or darkness, e.g., light blue or dark blue, and can be compared with the range of grey from white to black. Chroma, or intensity, refers to the brightness or dullness of a color – yellow has a high chroma, blue has a low chroma. The chroma of a color can be reduced by adding its complement.

In the planning of any colored visual, hue, value and chroma are as important as line, shape, mass and texture. In well-integrated, unified artwork, these visual elements, individually and together as a whole, follow the principles of design: balance, dominance, simplicity or limitation, and movement or rhythm. All colors, no matter whether they are pure spectral hues, darkened with black, lightened with white, or dulled with their complements, can be compared as to lightness or darkness with a value of grey. In designing graphics for any medium, value contrasts are extremely important – a colored graphic may be viewed on a black and white monitor or on a badly adjusted color monitor.

When producing colored graphics, never place white letters on light backgrounds or black letters on dark tones. Dark, low-chroma letters on a light background are more readable than light, high-chroma ones on dark. The most legible color combinations are: shown in Table 1.

Table 1. The most legible color combinations

Lettering or Art	Background
Black	Yellow
Black	White
Dark green	White
Dark blue	White
White	Dark blue

Where possible, let the most important item in a graphic have the most important color and the greatest contrast with its background. ‘Important’ colors are those with high intensities and are at the light end of the value scale. The pure hues – green-yellow, yellow, yellow-orange, orange and orange-red, all warm colors – are considered more attention-getting, especially when used in moderation and when placed against a darker color of low chroma. Warm colors seem to advance while cool colors – red-violet, violet, blue, and blue-green – recede and should be reserved for backgrounds. Avoid placing complementary colors of the same value adjacent to each other. They appear to vibrate and annoy the viewer. In other uses, this can be a most effective device it is not recommended for teaching visuals. Avoid creating large areas of white. On slides, white – especially when contrasted with dark colors or black – flares and is hard to look at.

It is important to bear in mind that color will show differently on various media. What you see on a computer screen will not show the same on paper or on a data projector screen.

1.2.6 Other considerations

There are some other considerations in designing charts and graphs.

- What medium is to be used: print (in a book or in a journal which may have its own rules for style), overhead projection, 35 mm slides, etc? The limitations of each medium will define the size and shape of the graph, the size and number of words, the width of lines, tone patterns, and so on.
- What is the graph intended to stress: trend, magnitude, rate of change? This will dictate the type of graph to be used.
- What audience is expected or intended? How big is it? What is its background? What does it want to learn?
- Is the graphic to stand alone or is it part of a group? Graphics in a series should be prepared in the same size so that all words, lines, and other elements will end up in a uniform size on publication or projection.
- In general, all graphics in a series should be consistent in color, style and appearance. Any one chart or graph in a series can be emphasized by a change in color or format.

1.2.7 Layout

Since European languages are written from left to right and from the top of the page to the bottom, the eye naturally tends to alight first in the top left hand corner of any page, unless there are strong compositional features which draw it elsewhere. Therefore, this is the most important area of any illustration. As far as charts and graphs are concerned, the left and lower margins are also important. The reverse applies to Arabic or languages written from right to left.

1.2.8 Visual emphasis

As a general rule, plotted points and graph lines should be given more 'weight' than the axes. In this way the 'meat' will be easily distinguishable from the 'bones'. Furthermore, an illustration composed of lines of unequal weights is always more attractive than one in which all the lines are of uniform thickness. It may not always be possible to emphasize the data in this way, however. In a scattergram, for example, the more plotted points there are, the smaller they may need to be and this will give them a lighter appearance. Similarly, the more curves there are on a graph, the thinner the lines may need to be. In both cases, the axes may look better if they are drawn with a somewhat bolder line so that they are easily distinguishable from the data.

1.2.9 Title

Use larger lettering for the title. Usually it is centered at the top or bottom of the grid. The title must tell what, where and when. It should be as short as possible: eliminate unnecessary words.

1.2.10 Data points

The plotted points on a graph should always be made to stand out well. They are, after all, the most important feature of a graph, since any lines linking them are nearly always a matter of conjecture. If a multiple-line presentation is used, it will be helpful to code the points by different geometric shapes.

1.2.11 Standard errors

Standard errors of the mean are usually drawn as lines extending vertically above and below a plotted point or the top of a column. Normally these lines are of equal length and therefore only one of the pair needs to be drawn. The choice will depend on the circumstances. Standard error lines should be thinner than lines in the main chart.

1.2.12 Lines linking data points

On many kinds of graphs it is appropriate to link the data points with lines, as this will help the eye to find related points and to follow trends in the data. In multiple-line presentations there is often some degree of overlap between the lines. This can give rise to a great deal of confusion, particularly where the data points for two or more different lines fall very close together. In some cases it may be necessary to code the lines themselves in some way. Color coding of the lines is of some help on multiple-line graphs. This may not be a practical proposition for graphs which are intended for publication in journals, but color can be used on graphs specifically designed for slides or overhead projector transparencies. For black-and-white graphs, different symbols can be used to distinguish data points relating to different lines.

1.2.13 Scale calibrations

The frequency of labeled scale calibrations on the axes of a graph can significantly affect the accuracy with which it is interpreted. As little interpolation as possible should be required of the user, in order to minimize errors. Calibrations should remain clearly visible even after the original has been reduced to fit the requirements of the publication. For this reason we recommend that tick marks should always be drawn on the outside of the axes (pointing towards the number to which they refer). The treatment of scales can be extremely helpful in relating the position of the data to the vertical scale. This effect is achieved by removing the vertical scale altogether and extending the calibrations as thin continuous lines, thus forming a background to the data horizontally. This system works equally well for most graphs and bar charts.

1.2.14 Axes

Every axis must be scaled and identified by a label. The axes of a graph or chart are usually represented by a vertical line on the left and a horizontal line at the bottom of the plot. A line is not always necessary, however. Space can be used to create the impression of a border or line.

The desire to superimpose several different kinds of data which have a common independent variable such as time, needs to be resisted. Some scientists may ask for such data to be illustrated on superimposed axes. The result is more often than not a misleading unscientific compromise, preventing the quick interpretation of the data. The scientific press has examples of several differently measured components all superimposed with several scales both to left and right of the graph. If the research is to be taken seriously then so should the graphic means of communicating it. This in any case may be difficult to do with some graphics packages.

The choice of scales for the axes should be very carefully considered. An ill-advised choice can give a very misleading impression of the data. If a change of scale occurs, once again this must be clearly indicated. A graph that has more than one Y axis is difficult to read. If the labels are brief, they can be centered above the Y axes. It is helpful to include the data point or line pattern in the Y-axis labels to indicate which data relate to each Y axis. One data point in each Y-axis label is sufficient. If a graph having more than one Y axis becomes too complex, each graph should be presented separately.

Labeling of axes. The horizontal axis of a graph is relatively simple to label. The most convenient arrangement is to place the label horizontally beneath the axis. The correct solution is not always so obvious for the vertical axis, however. If the label is written vertically, this will leave more space for the plot itself, but the label will be harder to read. In printed materials this is not a major problem as the reader can easily turn the page through 90 degrees, and study the graph for as long as he or she wishes. With presentation slides, however, the audience will have to turn their heads sideways in order to read the label, and they will have less time to do so. Vertical labels are therefore preferable for printed materials and horizontal ones for slides.

Many illustrations are prepared for both purposes, however, and the question then arises as to which medium should take precedence. The answer should be obvious. Lectures are transient events informing only a few persons at a time, and the audience will usually hear a lecture once only, with no subsequent opportunity of referring to the illustrations. Printed publications, however, provide a much more permanent record which can be referred to time and time again. Published data should therefore be regarded as more important than slides, and labels which waste space on the printed page should be avoided.

Sometimes short vertical axis labels can be accommodated horizontally, but where this is possible, some thought must be given to their positioning. A capital may be used for the first letter of the first word only, never use a capital for the beginning of every word in a label or caption.

Standard abbreviations should be used for all units. The recommended procedure is to describe the variable and then follow this by the units in brackets, for example 'time (sec)', or 'Yield (kg/ha)'. Abbreviations of units should not be pluralized. Grams is g and not gms; the latter would mean grams per meter per second. Similarly, abbreviations should not be followed by full stops unless these are part of the internationally agreed system.

Each axis must be clearly labeled with the name of the variable and the unit of measurement. Conventionally, the name of the variable is given first and the unit of measurement (in SI = Systeme International d'Unites) is given immediately after or below it in parentheses, for example, 'Temperature (°C)'. Exceptions are variables having dimensionless quantities, such as counts (for example, number of seeds), ratios and pH.

Axis labels should be brief. If necessary, a brief label can be explained more fully in the legend. Axis labels should be placed outside the graph. The X-axis label is customarily centered under the X axis. The Y-axis label can be centered perpendicularly above the Y axis, if the label is short, or centered parallel to it. The parallel label is conventional, and is best when horizontal space is limited. However, perpendicular labels are easier to read. Whether perpendicular labels should be placed above or to the left of the Y axis depends on the amount of horizontal space available. In both parallel and perpendicular labels, the unit of measurement can be placed below the name of the variable. Parallel labels should read from bottom to top.

1.2.15 Symbols

Various symbols (● ○ ■ □ ▲+ † *) are available depending on the graphics package used. Symbols should be distinctive, so that they can be easily distinguished. Symbols can be distinguished by color and by shape. The easiest symbols to distinguish are black from white. The easiest shapes to distinguish are circles from triangles. Circles should not fall next to squares because these two symbols are difficult to distinguish, especially after the graph is reduced to publication size. If more than four symbols are needed, squares should be used, and care should be taken to keep squares away from circles. Black symbols are more prominent than white symbols. If only one symbol is needed, black circles should be used.

1.2.16 Legends

The legend of a graph typically defines symbols or abbreviations. Legends should be positioned within the graph if possible. The legend is a descriptive or informative statement, and appears below or next to an illustration to identify and explain the illustration. The legend is needed so that the illustration will be intelligible without reference to the text.

1.2.17 Headings

On illustrations for publication it is not usually necessary to include a heading or title in the artwork because the necessary information will be included in the figure caption. For presentation slides, however, a heading may be necessary as a means of immediately identifying the subject. For dual-purpose illustrations it is often convenient to include a heading, as this can easily be removed subsequently if necessary.

1.2.18 Grid lines

Grid lines may often be omitted. Indicate major divisions by small ticks and minor divisions by smaller ticks inside the scale lines.

1.2.19 Amount scale

On all arithmetic line, bar and column, and frequency distribution charts, the amount scale must start at zero. Not to do so creates visual distortion of the truth. Always write in the zero. On the other hand, a zero base line is not necessary in charts based on an index, e.g. cost of living index.

1.2.20 Curved lines

The curves should be the most prominent feature in any chart. Normally, do not show more than five curves on one chart. When there are many overlapping, crossing curves, enlarge the overcrowded portion of the grid or prepare more than one graph.

1.2.21 Tones

For surface, bar, column, frequency distribution or pie charts, use black and white tone patterns to differentiate areas. Let adjacent values contrast well. The available patterns plus white and black are arranged in ascending order of tone or value or darkness.

1.2.22 Bars and columns

Bars and columns should be wider or narrower than the spaces between them. All bars should be of equal width, preferably half bar width. When there are many bars, the spaces may be eliminated. Give all bars a tone. Do not let one bar or portion be white.

Labeling bars and columns. With divided bars, place the label for each division either above or below the first bar. A key may be necessary here. With divided columns, place labels to the left or right.The use of scale and grid lines makes it unnecessary to give the numerical data in labels, but sometimes the scale can be omitted and actual amounts indicated. Do not place figures or words beyond the ends of the bars or columns, as this creates the illusion of greater length or height and makes visual comparison difficult.

1.2.23 Insets

A portion of a graph can be enlarged and presented as an inset. An inset can also be used to present a related graph or to make a comparison. If an inset is used, it should be within the rectangle implied by the axes, in a spot where there are no data. The data points, letters and numbers in insets should be smaller than those in the main graph, but must still be legible after the figure is reduced to fit the journal's column or page.

1.2.24 Arrows, lines, bars and shading

Arrows can be used to draw attention to important features of a curve or to identify points of special interest. Arrows or vertical dashed lines can be used to indicate an event. Horizontal bars or lightly shaded rectangular areas can be used to indicate the duration of a treatment.

1.2.25 Pie charts

Wherever possible, arrange sectors according to size. Place the largest at the central point of the upper half of the circle and continue with progressively smaller portions in a clockwise manner. No segment should be less than 5%. One segment may be separated to give it extra emphasis, but never separate more than one. Place labels outside the circle.

1.2.26 Composites

Related graphs can be combined into a composite illustration, in which two or more graphs appear side by side or one above the other or both. The composite, along with its legend, should be designed to fit across the journal page or within a column. The graphs within a composite should be of the same proportions and style. Graphs in a composite that are to be compared should also have the same scales. When the same dependent variable appears in more than one graph of a composite, it should be represented by the same symbol, just as in a set.

Each graph within a composite should be identified either by a brief label naming the distinguishing aspect or by a capital letter. The labels or letters should be placed as nearly as possible in the same spot in each graph, preferably in the upper left corner.

1.3 Types of Graphs

A mass of figures, even when neatly arranged in columns and tables, is difficult to interpret. That is why we turn to charts and graphs, and use them to highlight trends and illustrate comparisons implicit in the raw data. Charts and graphs are a means to present statistical information clearly and accurately, and to do this they must be kept as simple and straightforward as possible. The objective is to achieve a balance between simplicity and function. The successful product should be accurate, clear, concise and elegant.

The content of an illustration and the way in which the information is presented will depend on the medium for which it is intended, and the kind of audience. A detailed illustration which is satisfactory for a printed publication may be entirely unsuitable for a PowerPoint slide. Furthermore, a style of presentation which is suitable for one publication may not be suitable for another, and a slide intended for a scientific meeting will not necessarily be suitable for a non-specialist audience. This must be borne in mind when planning illustrations

An illustration will be most effective when its structure and the data it contains are simple rather than complex. Unnecessary information will merely detract from the illustration by increasing its overall complexity. Some scientific findings are unavoidably complex, however. In studies relating to plant pathology, for example, many factors may need to be considered at once. In this case, it may be preferable to build up the data by means of a series of related illustrations, particularly if the information is in slide form. The complex overall picture may then be shown as a final illustration. This method will help the audience or readers to understand the data while making no attempt to pretend that the overall picture is less than complex.

In the case of graphs, the number of lines which can be included on any one illustration will depend largely on how close the lines are and how often they cross one another. Three or four is likely to be the maximum acceptable number. In some instances, there may be an argument for using several graphs with one line each as opposed to one graph with multiple lines. These two arrangements are equally satisfactory if the user wishes to read off the value of specific points. However, if the reader wishes to compare the lines, then the single multi-line graph is superior.

Every item in an illustration should have a direct and simple purpose, and this includes labels and notations of various kinds. Nothing should be added if it does not contribute to the meaning of the illustration.

There are many types of charts and graphs, and each has its own purpose. Arithmetic line or curve charts show trends or movements over time. Column charts compare magnitudes or emphasize differences, also over time. Bar charts are used to compare magnitude or size of different items at the same time. Pie charts, particularly useful to the lay audience, and 100% column charts show relations of component parts of a whole. The more complex semi-logarithmic charts show relative changes and are useful when quantities are vastly different. Frequency distribution charts are specially designed to show patterns of occurrence. Sometimes, instead of abstract lines and bars, symbols can be used to make charts more attractive or to communicate abstract ideas. In every case, the selection of a particular kind of graphic representation is as important as the care with which it is rendered.

1.3.1 Arithmetic line or curve chart

Line or curve charts depict trend or movement over a period of time. Time,or any other independent variable (that which changes regularly) is plotted on the X axis; the dependent variable (that which changes irregularly) is on the Y axis.

1.3.2 Line graphs

A line graph is a graph that has two axes, at right angles to each other, which represent the scales of two ratio-scale variables (e.g. weight, volume, pressure). The axes form a coordinate grid, on which a relationship between the two variables is shown. One variable may be dependent on the other. An independent variable is a variable for which the investigator sets a value or a range of values; the dependent variable changes as the independent variable changes. For example, concentration (dependent variable) may change with time (independent variable).

Conventionally the independent variable is graphed horizontally, along the X axis, and the dependent variable is graphed vertically, along the Y axis. Thus, the graph shows what happens to Y as X changes. The relationship in a line graph is depicted by curves, by data points, or by both.

Three common types of line graph are the linear (arithmetic) scale graph, the semi-logarithmic scale graph (semi-log graph), and the logarithmic scale graph (log-log graph). In linear scale graphs, equal distances represent equal quantities everywhere on each axis. In semi-log graphs, one axis is in linear units and the other is in logarithmic units. In log-log graphs, both axes are in logarithmic units.

To show a relationship between two ratio-scale variables, such as a trend with time or the dependence of a response on the value of an additive, a line graph is used. (Ratio-scale variables have constant intervals between successive units and a zero point that has a physical meaning, e.g. weight, volume, pressure, time, concentration.)

- ***When to use.*** To show trend or movement rather than actual amounts. To illustrate long series of data. To compare several series of data. To interpolate or extrapolate.
- ***When not to use.*** When the emphasis is on change in amounts rather than on movement or trend. Instead, use the semi-logarithmic chart. To compare relative size or to emphasize difference between elements. Instead, use the bar or column chart.

Positive and negative values. To show percentage or other deviations from a norm. The zero base line represents the mean, the curve represents deviations from this mean. Note that the X (the zero base line) and Y axes are the same width.

Special referent line. To compare trend of data with index. The referent line is the same width as the X axis.

Surface or silhouette. To emphasize or dramatize trend. The arithmetic curve is filled in.

Multiple surface or band. To compare trend of a whole and its components (Figure 2). This is not an arithmetic line chart filled in. The width of each band expresses a unique value. Only the topmost line is the total of all values. Place the curve with the least movement next to the horizontal axis, the second smoothest curve next, and so on for no more than five curves. Do not use this chart when movement of curves is erratic. A divided or grouped column chart can illustrate the same information, as can the simple arithmetic chart.

1.3.3 Column charts

Column charts (Figure 3) are simple and adaptable and are readily understood by the lay audience. They are interchangeable with the arithmetic line chart.

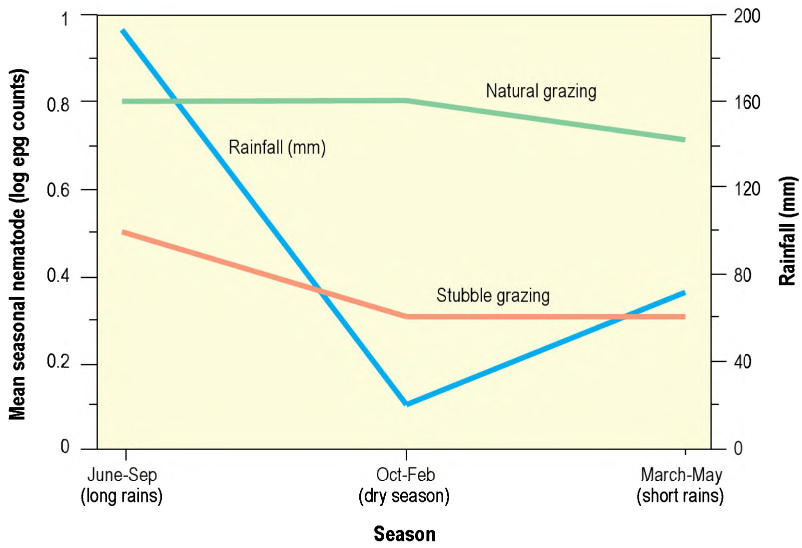


Figure 2. Relationship of mean seasonal rainfall with mean seasonal log epg counts by natural and stubble grazing groups

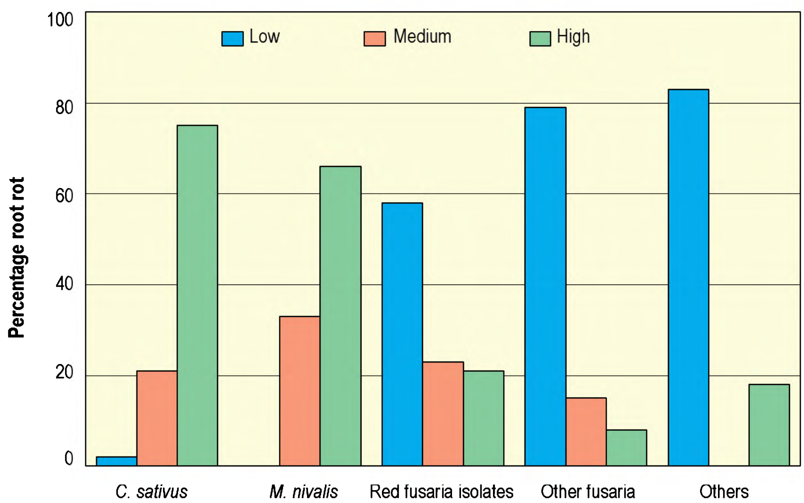


Figure 3. Pathogenecity test, barley root rot survey

- ***When to use***. To compare magnitude or size or to emphasize differences in one item at various periods of time.
- ***When not to use***. Do not use a column chart to illustrate data trends; instead, use an arithmetic line chart.

Grouped column chart. To compare independent series or data over a period of time. Have no more than three bars in one group.

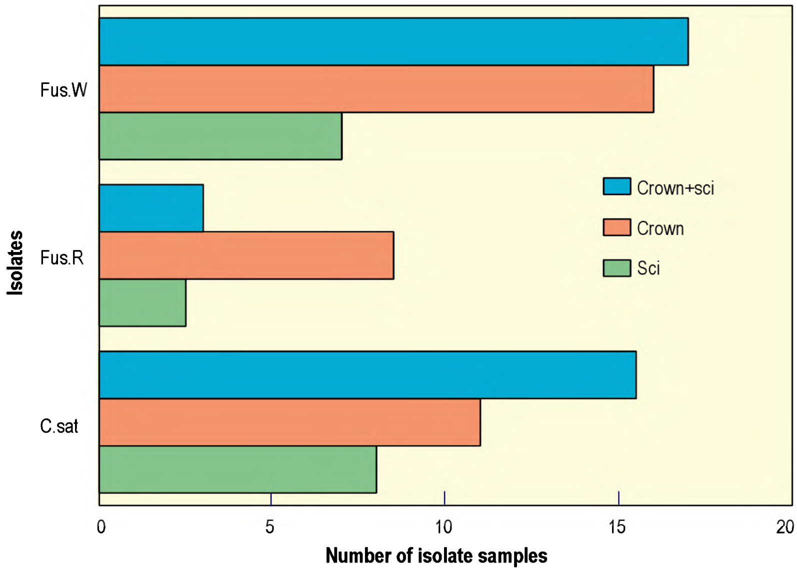


Figure 4. Isolation of C. sativus and Fusaria

Subdivided column chart. To compare totals and sums of totals over a period of time. Because the upper segments do not have a common base line, they are hard to compare.

Deviation column. To portray positive and negative values over a period of time.

100% column chart. To compare components of a whole at one time. This is not a time series. Keep the portion of most interest next to the base line. Lines connecting the sections facilitate comparison.

1.3.4 Bar chart

Bar charts (Figure 4) are simple, adaptable, and readily understood by the lay, audience. To show amounts or frequencies for nominal-scale or ordinal-scale data, a par graph (also called bar chart) is used. Bar graphs are also used for ratio-scale variables, particularly when data were collected at un-equal intervals.

- ***When to use***. To compare magnitude or size or to emphasize difference between many items at one specified time.
- ***When not to use***. To compare difference in size of one item over time.

Types of bar charts. A group bar is used to compare various aspects of several items at the same time. Subdivided bar to compare totals and sums of totals at one time; a 100% bar to compare components of a whole. Paired bar to compare two different types of data. Deviation bar to portray positive and negative values at one time. Arrange bars in descending order where possible.

1.3.5 Semi-log (relative change) charts

This type of chart (Figures 5 and 6), portraying proportional and percentage relationships, is useful because both relative changes and absolute amounts are shown. Relative change for large and small quantities can be compared. In the arithmetic line chart, equal spaces represent equal amounts of change; in the semi-log chart, equal spaces represent equal amounts of relative change.

- ***When to use***. To show relative change when baseline quantities differ greatly. The arithmetic chart portrays comparative fluctuations only when the quantities compared have similar values.
- ***When not to use***. To compare amounts of widely separated values. To illustrate data for a lay audience.

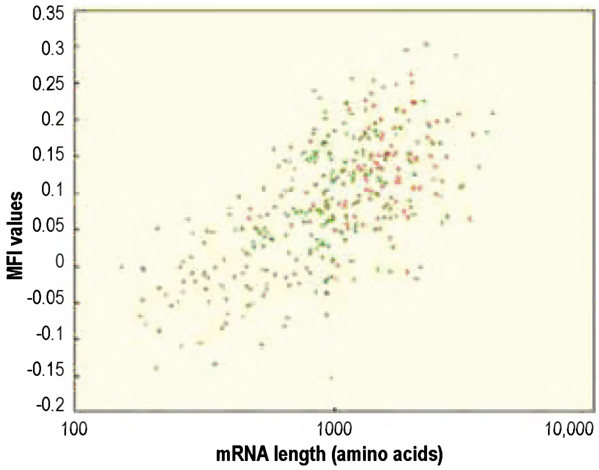


Figure 5. Mitochondrial mRNA localization and mRNA length: MFI values of 358 genes plotted against the corresponding mRNA length (semi-log scale)

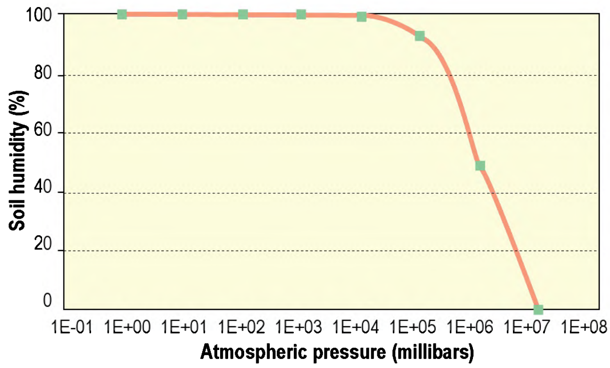


Figure 6. Soil humidity as a function of atmospheric pressure

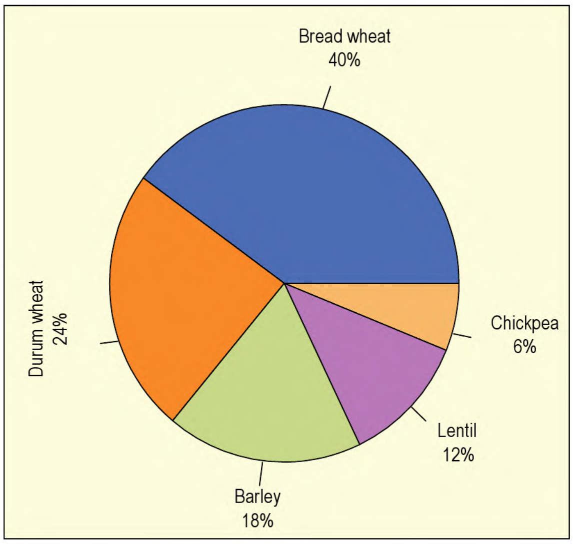


Figure 7. Cereal and legume production

How to read semi-logarithmic charts.
Vertical distance on a semi-logarithmic chart indicates a percentage change. The space between 1 and 2 (or between 10 and 20 or 100 and 200) always represents a change of 100%. This same distance always represents a change of 100%, no matter where in the chart it is measured.

1.3.6 Pie charts

Like the 100%, column chart, the pie chart (Figure 7) portrays relations of component parts. In the pie, however, it is difficult to compare areas accurately. To show the components of a whole, either a component bar graph or a pie graph (also called a pie chart) is used.

- ***When to use.*** To show relations of component parts, especially to the lay audience. Two graphs, the 100% bar or column and the simple bar, will portray the same data.
- ***When not to use.*** To compare component parts of two or more wholes. Instead, use 100% column charts.s

1.3.7 Frequency distribution charts

Histograms and frequency polygons are graphs that display a frequency distribution on a coordinate grid. A frequency distribution is a display that shows the total number of observations for each class of a ratio-scale variable, such as time or weight.

The difference between frequency distribution and arithmetic line charts is in the data portrayed. Discontinuous variables, those with a limited number of distinct values, such as numbers of infested plants, are plotted against continuous variables, such as height, weight, etc., which have an unlimited number of possible values. The latter usually are usually placed on the X axis. To show a frequency distribution of data for a ratio-scale variable, either a histogram or a frequency polygon is used.

In a histogram (Figure 8), the frequency distribution is represented by a series of contiguous rectangles. The width of each rectangle represents a range of values, which constitutes one class of the ratio-scale variable shown on the X axis. The height represents the corresponding frequency, which is shown on the Y axis. The area enclosed by the histogram represents the distribution of the data. Only if the widths of all rectangles are equal do the heights of all rectangles represent the distribution. A histogram may be presented in outline only, with the vertical lines of individual rectangles omitted to emphasize the shape of the distribution or the vertical lines may be retained. The histogram may be shaded dark gray or black or left white.

In a frequency polygon, the frequency of each class of a ratio-scale variable is plotted at its midpoint and the midpoints are joined by lines. So that the entire distribution will be represented, the lines joining the midpoints are extended to reach the baseline (zero frequency) at the midpoints of the classes that lie before the first class and after the last class. The area enclosed by the frequency polygon represents the distribution. Frequency polygons are useful for comparing overlapping distributions; this cannot be done successfully with histograms.

- ***When to use.*** To visualize the distribution of various entities.

Preparation. Three types of frequency distribution charts exist and can be prepared from the same data: histogram, polygon, and smoothed curve.

1.3.8 Area plots

100% surface. Area plots (Figure 9) show the whole area under one graph in one color or shade.

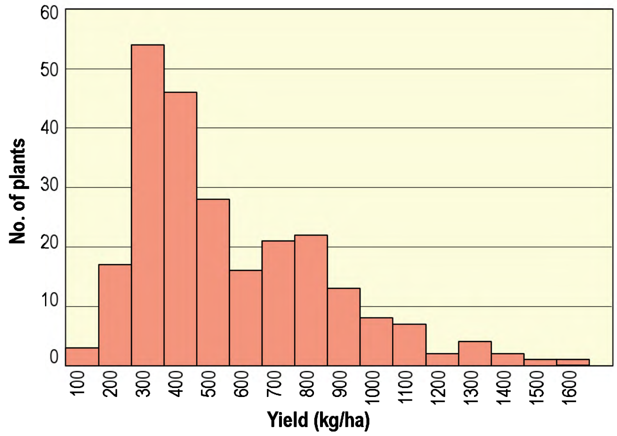


Figure 8. Histogram

- ***When to use.*** Very useful for portraying trend of components of a whole.

1.3.9 Pictorial charts

Pictorial symbols can facilitate the communication of abstract ideas or concepts. Symbols can be used in two ways. As numerical counting units, each symbol represents a specific number. As a support for a conventional chart form, symbols add emphasis and differentiate.

How to use pictorial symbols. Symbols must have strong associations with the idea so that they are readily understood by the audience. Avoid dated or regional symbols. Symbols must be well designed and simple. They must be recognizable on reduction or enlargement and should be easily divisible. Shapes must

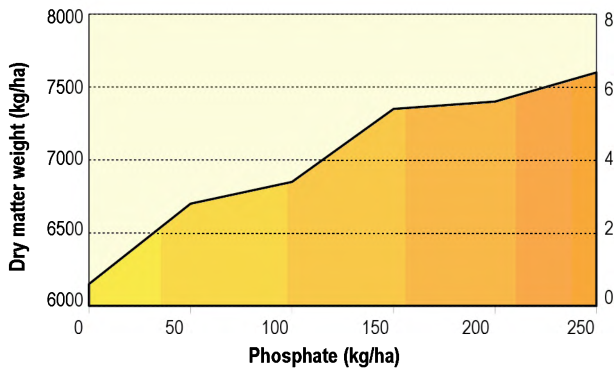


Figure 9. Dry matter response to phosphate doses in lentil

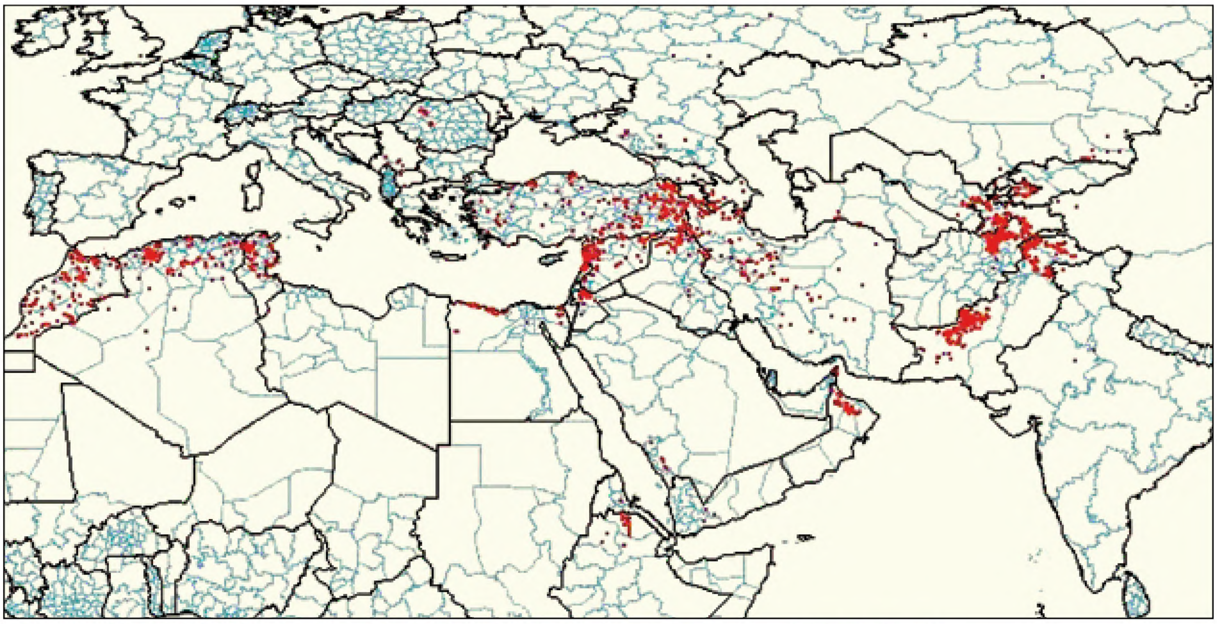


Figure 10. Collection sites for bread wheat landraces held in the ICARDA collection

contrast well. When used as counting units, symbols must all have the same value and should all be approximately the same size. In quantitative data, avoid comparison of size of symbol. It is practically impossible either to draw or visually compare symbols whose volumes vary. An example is shown in Figure 10.

1.3.10 Standard error

In line graphs and bar graphs, variability surrounding the mean value can be indicated by an error bar (a thin vertical line ending in short horizontal lines) centered at each data point or data bar, as shown in Figure 11. The figure legend should state whether the error bars represent standard deviations, standard errors of the mean, confidence limits, or ranges. The number of observations (n) also should be stated, preferably in the legend. Putting n next to each data point or inside each data bar clutters the graph.

Generally, full error bars can be drawn. However, in line graphs, if error bars create a lot of clutter, half error bars can be drawn.

Probability values (P values) should not be put on the graph, either instead of or in addition to symbols indicating statistically significant differences, because P values clutter the graph. The P value, preceded by a statement telling which data are being compared, belongs in the legend.

1.3.11 Curve fitting

Individual points carp either be joined by straight lines as explained earlier, or a curve fitting procedure can be used to generate a

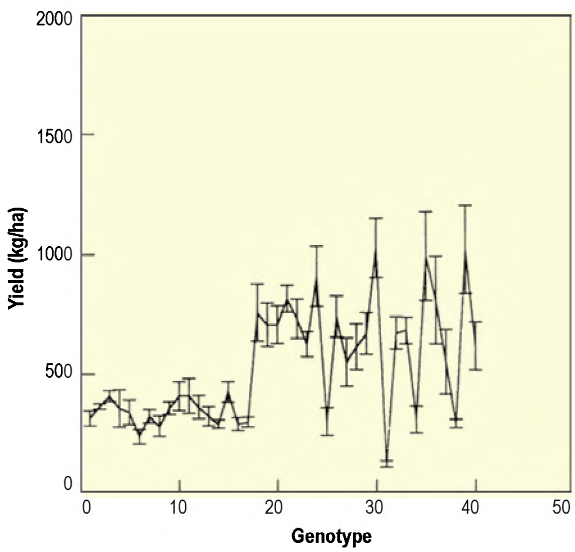


Figure 11. Yield – genotype plot

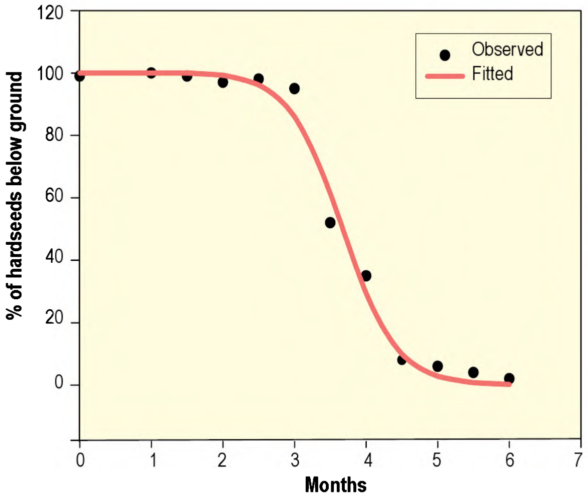


Figure 12. Below-ground hardseeds breakdown in ACC V2650 pasture species

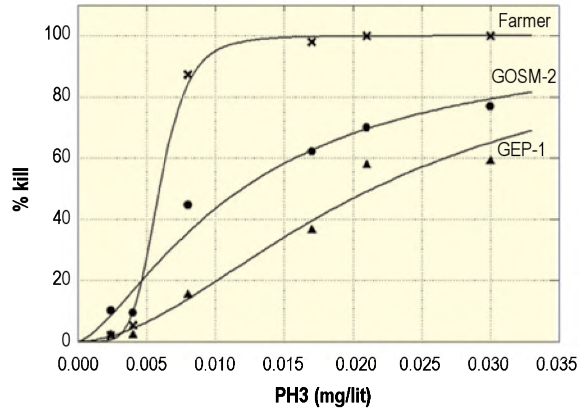


Figure 13. Phosphine resistance in three field strains of Rhizopertha dominca collected from stored products around Aleppo

best-fit curve through the points, as shown in Figures 12 and 13. Linear regression is addressed separately later on, but a polynomial fitting procedure is often used to minimize the errors in the fit through a least-squares procedure.

1.3.12 Three-dimensional plots

Three-dimensional plots (Figures 14 and 15) are highly effective in the display of a function of two variables or the interpolated surface of one variable dependent on two variables. Various interpolation routines are available, some of which can tolerate missing values while others do not.

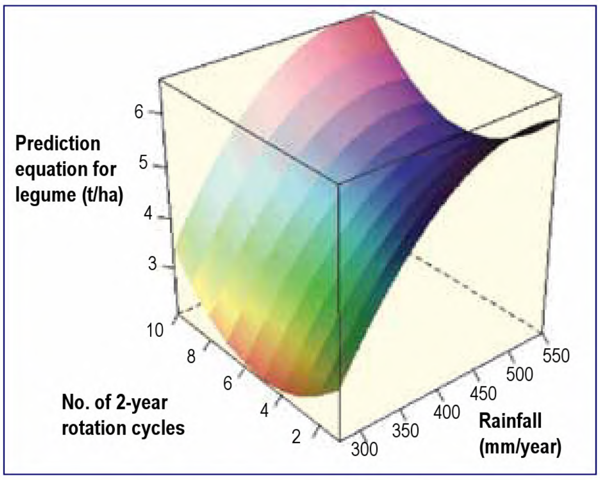


Figure 14. Prediction equation for legume, fallow and continuous cereal systems modeled in terms of rainfall and time as number of 2-year rotation cycles at Hemo Experiment Station in Kamishly, Syria, based on data from 1986/87 to 1995/96

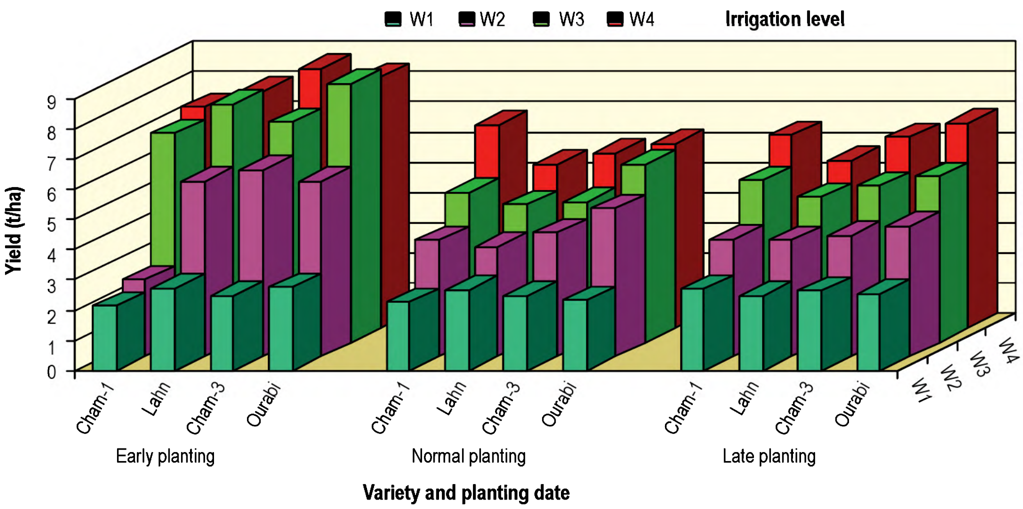


Figure 15. Response of durum wheat to supplemental irrigation, under different planting dates

Usually the surface is depicted as a mesh with the variations clearly visible. The finer the mesh, the greater the effect of a visual continuum. Solid modeling techniques give the possibility of a continuous surface but they require special software and hardware.

1.3.13 Scatter plot

A scattergram (Figure 16) is a graph drawn on a coordinate grid, on which each individual event or value is represented by a data point. Whether the variables shown on the axes are correlated, and if so, to what degree, can be determined by trying to fit a mathematical function to the data points. For example, a linear function can be shown by fitting a straight regression line to the data points. One way to estimate the strength of a linear association is by determining the correlation coefficient (r).

To show whether there is a correlation between two ratio-scale variables, a scatter diagram (scattergram) is used.

1.3.14 Linear regression

Regression lines (Figure 17) can be shown by dashed or solid lines. Many graphical / statistical packages display the regression line and give the values of the parameters of the linear model.

Residuals. The plot of the residuals resulting from any regression process (Figure 18) can give the scientist a lot of insight on the goodness of fit of the regression carried out, by plotting the residuals against the fitted values.

1.3.15 Outliers

Outliers can be highly visible on plotting the initial raw or processed data. Clearly, there are methods for identifying outliers, but graphical presentation can greatly assist the process (Figure 19).

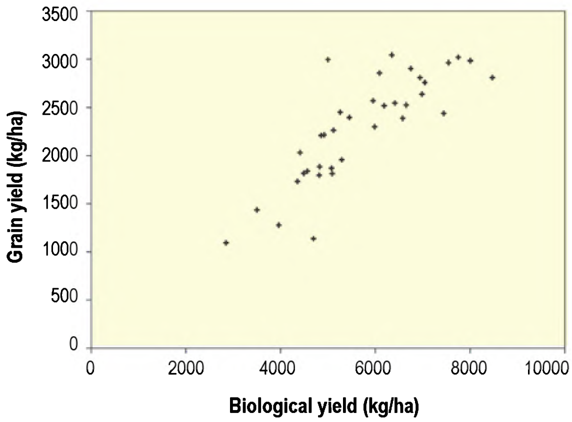


Figure 16. Scatter diagram showing relationship between biological yield and grain yield in chickpea, 1997-98 season, Tel Hadya

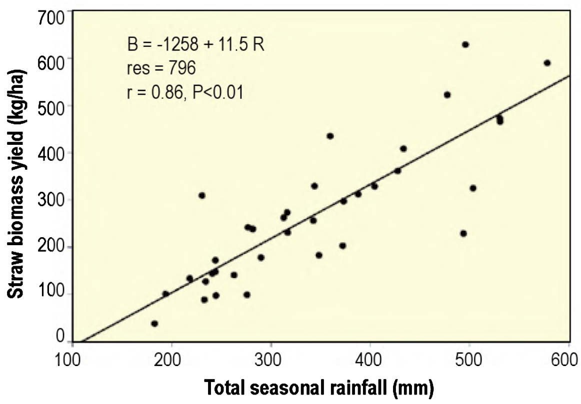


Figure 17. Observed values (circles) and fitted linear regression (lines) of straw biomass yield (B) on total seasonal rainfall (R), for rainfall regimes up to 600 mm. rse = residual standard deviation

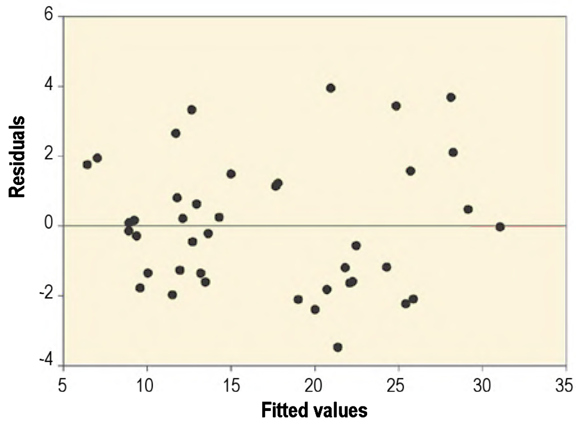


Figure 18. Goodness of fit of near-infrared predictions of straw intake. Residuals versus fitted values

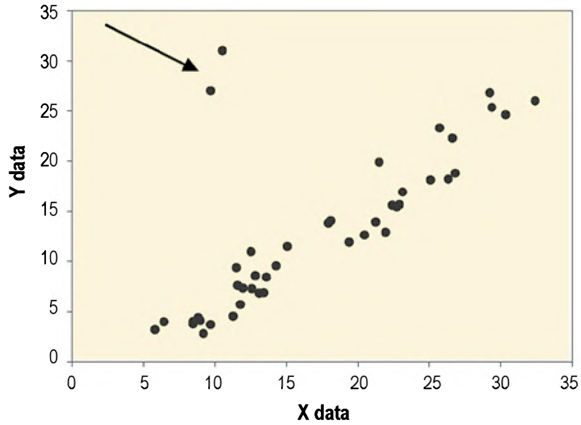


Figure 19. Scatter diagram shows the presence of two outliers

1.3.16 Dendrogram

The dendogram (Figure 20) is a special graphic representation to show the clustering of a certain variable through a schematically structured diagram. The Y axis usually indicates the values used for the clusters while the X axis indicates the groups being clustered.

1.3.17 Box-and-whiskers

Box-and-whisker plots graphically illustrate the statistical distribution (or spread) of a variable. The whiskers can be represented in different ways, depending on how much attention you wish to pay to identifying outliers. The simplest method is to draw the whiskers to the extreme values, the highest and lowest data points. No outliers are displayed. Figure 21 shows the properties of box-and-whisker plots, using Tukey's method:

- A box in the middle of the plot that encloses 50% of the data (the interquartile range)
- A median line in the box; sometimes the mean is also shown
- The upper 'hinge' of the box, i.e. 75th percentile
- The lower 'hinge' of the box, i.e. 25th percentile
- A dotted line, the upper whisker (or tail that extends from the upper 'hinge'

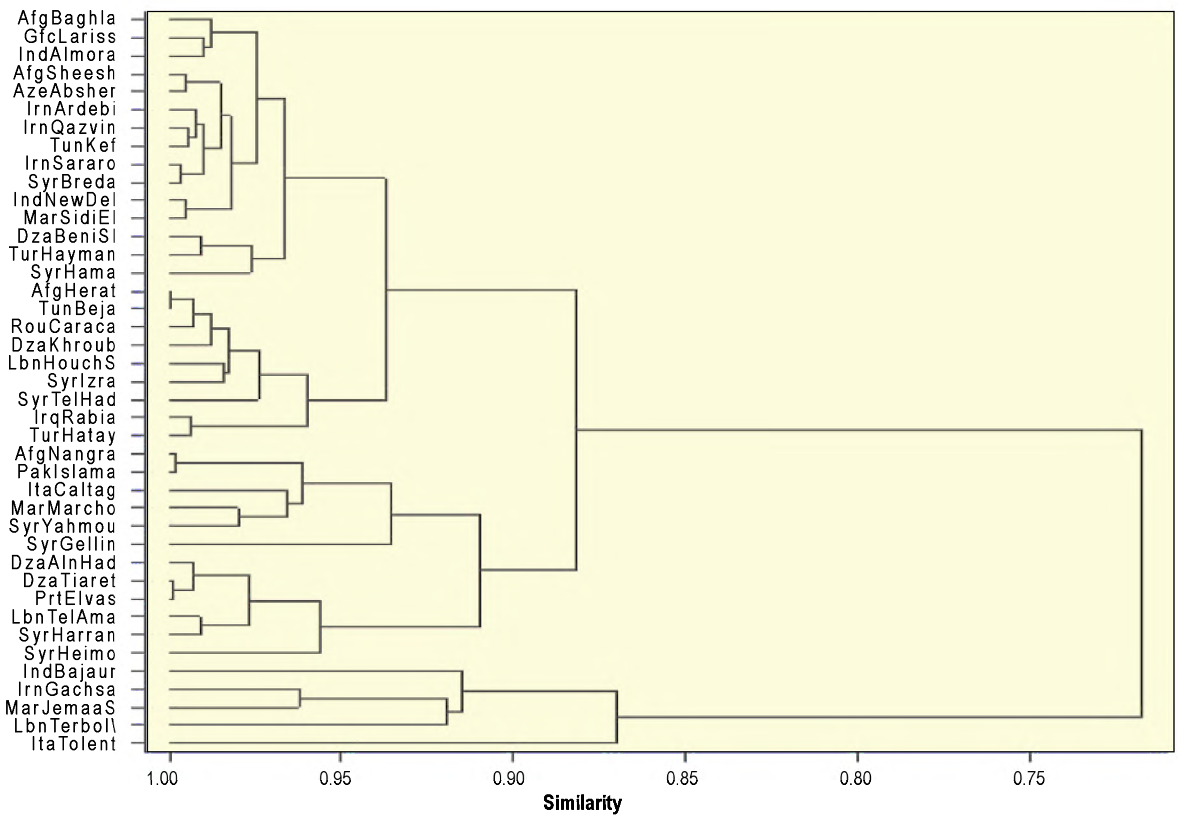


Figure 20. Clustering of locations: average linkage

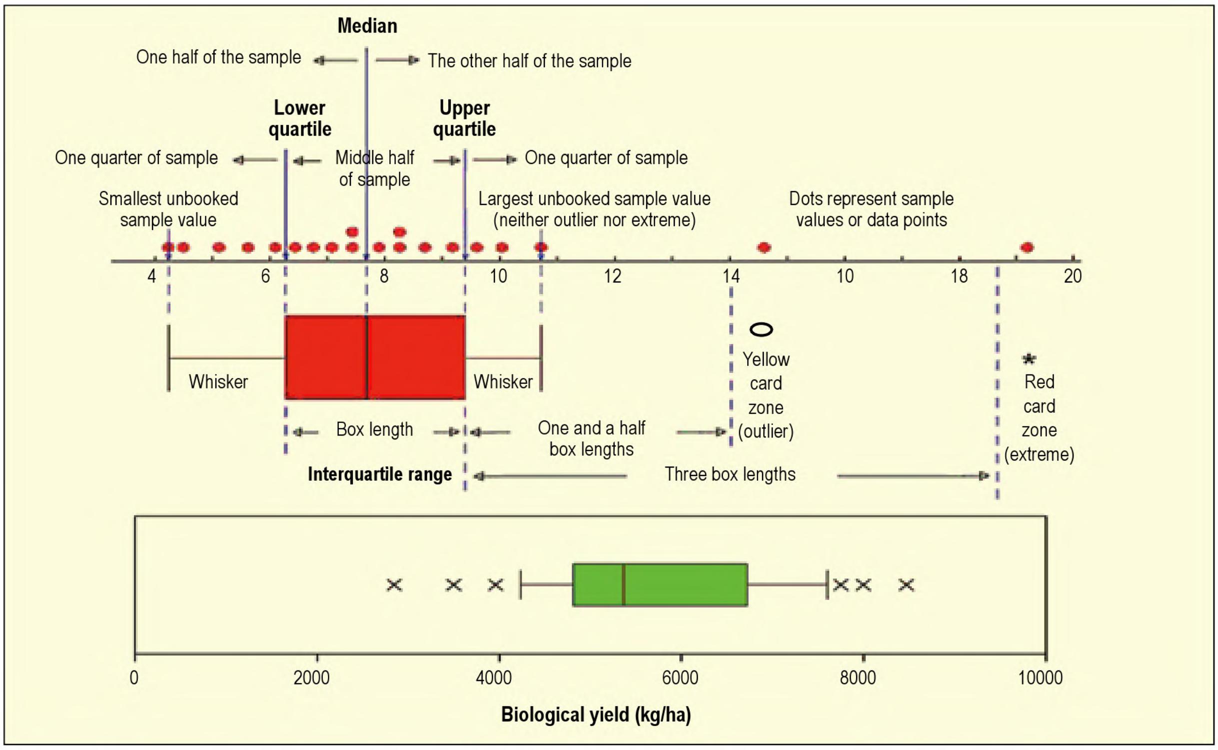


Figure 21. Box-and-whiskers plot: chickpea biological yield histogram, Tel Hadya, 1997-98 season

1.4. Tables

In general, tables are better than graphs for giving structured numeric information, whereas graphs are better for indicating trends, making broad comparisons or showing relationships.

Reference tables contain information that people will look up; they serve an archival function and often need to be laid out for economy of space, while preserving data accurately. It is extremely important that they include good meta-data, the descriptive information which allows the data to be correctly interpreted – usually a comprehensive version of the "what, where and when". Reference tables often appear as appendices.

Demonstration tables are intended to be assimilated quickly by the reader or viewer. It is important that they are clear and well-presented, using reasonable approximations to reduce numbers to relatively few significant digits. They should be included in the text: readers following a general argument tend not to bother flipping backwards and forwards. Over-large demonstration tables are intimidating and users tend to give up on them. If the information is all necessary, it should be split into manageable components. Omit any column which can be readily calculated from data in other columns. Minor, or not very relevant, categories can be combined.

When do you use tables?

- Exact numerical data
- Localized comparisons

Types of tables

- Exploratory
- Resource
- Presentation

Data field

- Keep the number of decimal places to a minimum

- Place notation on the labels not in the data field
- Keep the number of empty cells to a minimum

Table titles
- Briefly describe the data to be presented
- Clearly highlight the difference between this table and others
- Long enough to inform but short enough not to bore
- Do not include a description of every variable

Row labels
- Should be consistent from table to table
- Use indents to show subdivisions
- Create patterns in long lists by skipping a line or half line every fifth entry

Column heads
- Identify what the column is measuring
- Use the same units down the column
- Use simple but informative labels
- Avoid repetitious categories by spacing and grouping
- Make certain that both the numerator and denominator are clear

Create a logical visual pattern
- Order the content according to how the data will be used
- Presentation tables use rank order or other hierarchy
- Resource tables use alphabetical or other logical order
- Look to the numbers to carry out the ordering
- Make patterns and exceptions obvious at a glance
- Place figures that are to be compared close to one another
- Put summary data where you think it will be most useful

1.4.1 Tables of frequencies

The simplest tables arising from surveys, or from coded qualitative information, are of counts or frequencies. If relatively large counts are to be compared in a table with several rows and columns, it is often helpful to present them as percentages: common ways to do this involve making the percentages add up to 100 across rows, or down columns, or across the whole table. These facilitate different types of comparison, and a careful choice should be made. The sizes of sample on which a percentage table is based should be made explicit.

1.4.2 Orientation and order

The orientation of a table can have considerable influence on the readability. It is much easier for a reader to make comparisons within a column of numbers than within a row. Therefore if the purpose of a table is to demonstrate differences between treatments or groups for a number of variables, place the the groups in rows and the variables in columns.

1.4.3 Number of digits and decimal places

The number of digits and decimal places should be the minimum number that is compatible with the purpose of the table. It is often possible to use as few as two significant digits.

Sometimes units can be changed to make numbers more manageable. For example 12,163 kg/ha could be presented as 12.2 t/ha. In other cases numbers can be multiplied or divided by factors such as 1000 or 1 million. For example, most people would find it easier to understand 72 incidents per 1000 population per year than a rate of 0.07189. The number of decimal places should be consistent for each variable. Numbers in a column should be aligned according to the decimal point.

Table 2. Grain yields (t ha) from four rates of nitrogen and four rates of phosphorus on durum wheat cultivar Cham 1, Khan Shekhoun, Hama, 1989

Nitrogen N (kg/ha)	**P_2O_5 (kg/ha)**					
	0	**20**		**40**	**80**	**Mean**
0	1.57	1.51		1.14	1.97	1.55
40	1.79	1.60		1.93	2.03	1.84
80	1.72	1.81		2.09	1.80	1.86
120	1.81	1.76		1.42	2.09	1.72
SE (DF = 15)			±0.14			±0.071
Mean	1.72	1.67		1.65	1.97	1.75
SE (DF = 15)			±0.071			

Evaluation based on the two-factor combinations in an RCBD with 2 replications
SE = standard error, DF = degrees of freedom

Table 3. Mean daily gain and feed intake for 4 breeds and 2 diets (artificial data are shown)

Breed	Weight gain (g/day)			Feed intake (g/day)		
	Protein level			Protein level		
	Low	High	Mean	Low	High	Mean
Menz	38	56	47	639	952	796
Dubasi	33	57	45	603	1008	806
Wello	28	44	36	591	917	754
Watish	29	40	35	628	889	759
Mean	32	49	41	615	942	779

1.4.4 Tables for publications

In publications usually only black and white graphs are possible. The format of the table is often also restricted. Table 2 is a typical example, prepared for a journal publication.

1.4.5 More complex tables

The arrangement of Table 3 facilitates comparisons between the four breeds separately for each protein level as well as averaged over both levels. For example, we see that the first two breeds, Menz and Dubasi, have higher weight gains than Wello and Watish. Note that it would be possible to economize on space by combining the two sub-tables into a single table with two observations per cell. This may be justified in reference tables, but is not recommended in demonstration tables where it can be very confusing. Once again, the table omits sample size or accuracy information.

This kind of table cannot be used when it would need very many columns. Note that enough blank space is needed to separate the weight gain results clearly from the feed intake. Without such space this kind of table can become hard to read.

2 GRAPHICS PACKAGES

Computer graphics offer the scientist new creative opportunities. Graphics in full color (or in black and white) can now be created in a very short time. The possibilities offered by a dedicated, high-speed graphic arts computer, with high resolution and commercial software, are limited only by the buyer's budget and/or imagination. Today's sophisticated computer graphics systems permit the user to draw freely on the computer's monitor, creating images, manipulating their size and position, and developing text, designs and special effects from a selection of 16 million distinct colors.

Computers can receive raw data and then accurately and automatically plot charts and graphs of any design. Graphics software can automatically place statistically accurate curves and standard error bars. Special effects that enhance dimension, create perspective, and sequential spectrum colors can give the effect of a painting or lithograph.

Once artwork is created on the monitor it can be embedded in a PowerPoint presentation or sent to the computer's film recorder, which exposes the images to either color or black-and-white film. The film can then be processed to yield color or black-and-white slides. In black-and-white prints, colors are converted to textures and the images are camera-ready for publication as line art. If a laser printer is hooked up to the computer, color or black-and-white laser prints on paper or transparencies can be produced. These are useful for author's proofs and for filing, if not also as camera-ready copy.

All visual materials created on a graphics computer can be saved and filed indefinitely for quick retrieval and modification. Graphs can also be easily incorporated in a word processed document.

2.1 Overview of Graphics Packages

'Business graphics' systems designed for personal microcomputers now have high-quality graphics suitable for publications. Also, all the statistical and spreadsheet packages now have very good graphing capability. Rapid technological advances have made high quality computer graphics available at low prices.

The use of computer graphics in the agricultural sciences can be roughly classified into three categories: analysis of research data; display of results in color for on-line publications and conference presentation; and production of publication-quality black and white figures for books and journals. Analysis of quantitative data is the oldest and most highly developed of these uses. With the wide availability of powerful software for analyzing data, such analysis is the mainstay of computer graphics, and the quality of output is quite adequate for publication. The use of computer graphics for production of color presentations is more recent, but because of widespread demand and the willingness of business users to pay, development of color graphics for this purpose has evolved rapidly.

The technology for producing publication quality output from computer graphics systems has become the norm. Computer graphics packages have completely replaced manual methods of preparing graphs, charts, diagrams and other presentations whether for printed material, data projectors or animated videos.

2.2 Capabilities and Uses

The main features of these interactive packages are:

- Production of various types of graphs and charts such as XY graphs, pie charts, bar charts and others
- Easy manipulation and customization of the graphs including the titles, axes, grids, colors, symbols, connecting lines, dimensions, fonts, scales and legends
- Linkages to other packages to facilitate the import of data and the export of ready graphs
- Flexibility of output formats and devices
- Free-drawing feature.

2.3 Why Use Computer Graphics for Illustrations?

There are three main advantages of using computer graphics to produce publication-quality illustrations. First, an illustration, once composed, can be easily modified and printed out again. Using computer graphics, corrections and even wholesale changes can be made more quickly and less expensively; this capability encourages researchers to seek perfection in their illustrations rather than settling for the first rendering. Such flexibility is particularly useful for correcting illustrations when last-minute analysis or data editing necessitates changes.

The second major advantage is that the researcher can produce his or her own illustrations without having to interpret the specifications for another person. Although some researchers lack the artistic ability to compose graphics, those who are able can have the process under their own control. This control, together with the ability to make changes rapidly, enables the researcher to produce progressively better drafts, until the final version ideally expresses the scientific point.

The third advantage is the ability to easily display graphics of datasets containing many data points. This is particularly true of data from experiments producing readings from laboratory equipment that may require the resolution of thousands of points for a short line graph, but it may also be true of field trials data involving large numbers of samples.

In the past, such data were often plotted by the laboratory equipment on crude graphs, or by low resolution computer output devices, and were then traced into more finished illustrations by an artist. By converting the analog output signals from laboratory instruments to digital data (so called 'A-to-D conversion'), the researcher can use computer graphics to produce the entire graph complete with appropriate titles, axis labels etc. With the availability of powerful computers and the increasing feasibility of maintaining and analyzing large databases, the need for graphs with large number of data points can be expected to grow in the future.

2.4 The Artist, the Scientist, and Computer Graphics

Although some may view computer graphics as a replacement for the graphics artist, in practice their use has enhanced the artist's position as an efficient producer of publication-quality art and as a collaborator in the scientific process. Graphic systems have reduced response time and increased the profitability of producing text presentations and artistic graphics. Artists now have mastered the computing process to produce publication-quality illustrations, and the combination of their artistic skills with computer efficiency has proved a useful mix for most illustrations, including diagrams, standard charts and graphs, and even free-hand drawings that can be input to the computer. At present, however, most commercial artists cannot easily handle many types of graphs of statistical and laboratory data that must be computed directly from numerical databases.

For the production of publication-quality illustrations, artists have for many years collaborated with scientists by using their computer graphs to produce finished illustrations. As computer output has achieved publication quality, it has become common for artists to accept a computer graphic from an investigator, modify the computer-drawn illustration to overcome limitations in labeling, scale, etc, and produce a final graph for publication. This collaboration between computer-using artists and scientists can be particularly beneficial because it is often quicker for a scientist to have an artist modify a computer graphic than to figure out how to re-program or 'fool' the software to circumvent its limitations.

Graphics systems enable scientists and artists to collaborate more directly by sharing graphics computer files. Scientists are able to compute the most accurate representation of the analysis and artists will be able to access the same computer file for final artistic embellishment.

2.5 General Strategies for Using Computer Graphics

The most common computer-prepared illustrations for scientific journals and books are statistical graphs and free-hand or diagrammatic drawings. To design and set up a computer graphics system, one must understand the basic strategies used by these systems to produce the finished work.

To create a statistical graph, all computer graphics systems perform a set of basic functions. These include defining the type of graph (for example, line graph, bar chart, pie chart); constructing the axes (for example, scaling, numbering, labeling, defining tick marks); writing and positioning title, footnotes, legends and other labels; choosing cross-hatching or coloring of bars or other areas; selecting line types; determining the thickness of axes and lines; choosing and 'fitting' regression lines and confidence limits; and other miscellaneous functions (for example, selecting type fonts or plot symbols). Although these functions are usually taken for granted by computer users, a publication-quality illustration usually requires very careful selection of these parameters.

When using a computer graphics system, the user often wishes to select a few parameters and allow the software's default parameters to determine the rest (for example, the scaling of axes, tick marks, line types, etc). Although graphs can be produced rapidly this way, one cannot expect them to be of publication quality. Acceptable illustrations require the user to become extremely familiar with all of the many parameters and options and to specify virtually all of them in the set-up for the particular graph.

Besides the strategies for producing the usual statistical graphics, many computer graphics systems allow the development of free-hand drawings, complex diagrams, and a wide variety of artistic effects (for example, shading, 'echoing', etc). Generally, these were earlier done by means of a digitizing tablet, a flat drawing surface that enables the scientist to draw the shape with an electronic pen so that the computer can capture it. Free-hand drawings or tracings can be scanned and modified with various drawing functions of computer graphics packages, so it has become possible to draw more effectively. Although a virtually infinite array of artistic effects is possible, these powerful techniques should only be used when they will substantially augment the clarity, precision or informativeness of the illustration; they are not meant simply to be eye-catching or dramatic. Overuse of special effects detracts from the scientific purpose of the publication and may well lead to distortion of scientific objectivity.

2.6 Specific Functions of SigmaPlot, Harvard Graphics, PowerPoint, Excel

These are outlined in the documentation material of these packages. Almost all the good statistical packages (Genstat, SAS, SPSS and others) now have good graphing capabilities in addition to technical or statistical capability.

GLOSSARY

Abscissa	The horizontal or X axis of a statistical grid.
Algorithm	A diagram of a procedure that leads, by a series of choices, to a correct answer, or a diagram of a method of breaking a complicated decision-making sequence into its components
Annotation	An explanatory note forming part of an illustration
Average	A measure of the most 'typical' value in a series of observations. There are three way of expressing averages: arithmetic mean, median or mode
Axis	A fixed line adopted for reference. A line about which a body rotates, or about which the parts of a figure are arranged. Graphs or charts are usually organized on axes whi are at right angles to each other
Balanced layout	A layout which is 'balanced' by eye and not by measurement. A balanced layout is n necessarily symmetrical
Base line	The imaginary line on which the data stand, e.g. the zero line in a bar chart or histogram
Bold face (type)	Thicker than normal type
Cross hatching	The criss-cross patterns made to simulate textures
Dependent variable	A variable which is altered by changes in the independent variable. Dependent variables should always be placed on the vertical axis of a graph
Distribution map	A map that displays spatial relationships of data
Font	A complete family of characters of one size and style of type
Frequency	A set of data is divided into categories, the number of items in each category is know as the frequency distribution
Graph	A diagram that shows amounts, frequencies, trends, or relationships of data. The term 'graph' is sometimes restricted to line graphs, which show relationships between two or more variables. Graphs that show amounts or frequencies of one variable are then referred to as charts. In this book, the broader definition of graph is used
Grid	A framework (giving type areas, margins, illustrations, etc) used as a guide for laying down elements within a given format
Horizontal format, Landscape	The shape of an illustration when the horizontal measurement is greater than the vertical, also known as landscape or comic mode
Independent variable	Any variable whose values are not affected by changes in other variables. Time is an example. Independent variables should always be placed on the X axis of a graph
Italic, ital.	A sloping style of character
Legibility	Can you read it?
Line spacing	The space between each horizontal line of type
Logarithmic scale	Scales which are subdivided according to logarithmic principles, allowing large variation in quantities to be shown on the same illustration. Log scales always start at
Lower case	Letters that are not in capitals
Median	The center value of a series of observations (a form of average)
Standard error of the mean	The square root of the arithmetic mean of the squares of the differences between the observations and their mean, plus one
Variable	Data subject to measured change
Variance	The square of the standard error of the mean
Weighted average	The arithmetic average multiplied by a suitable 'weight' corresponding to importance

REFERENCES AND SOURCES

Reynolds, L. and Simmonds, D. 1983. Presentation of data in science. Martinus Nijhoff Publishers.

MacGregor, A.J. 1981. Graphics simplified. University of Toronto Press.

Council of Biology Editors, 1988. Illustrating science: standards for publication. Council of Biology Editors, USA.

Everitt, B.S. 1978. Graphic techniques for multivariate data. North-Holland Press.

Tukey, J. 1993. Graphic comparison of several linked aspects: alternatives and suggested principles. Journal of Graphical and Computational Statistics 2:1-33.

Energy Information Administration. 2008. EIA guidelines for statistical graphs. http://www.eia.doe.gov/neic/graphs/preface.htm.

Washington University. 2008. Style guide for Washington University publications. http://styleguide.wustl.edu/

Zawitz, M.W. 2008. Data presentation: a guide to good graphics. Washington Statistical Society Methodology Seminars. http://www.scs.gmu.edu/~wss/methods/index.html

Zawitz, M.W. 2008. Data presentation: a guide to good tables. Washington Statistical Society Methodology Seminars. http://www.scs.gmu.edu/~wss/methods/index.html

Presenting Scientific Data

ICARDA-01/500/ January 2010