Questions for review

1 What seven steps need to be taken before the researcher can collect data?
2 What information should researchers record in their journals?
3 What information should be kept on bibliography cards for:
  a a book?
  b an article?
  c a chapter in a book?
4 What is a 'unit of analysis'? Why is it important to know this in preparing data collection sheets and data recording sheets?

Suggestions for further reading

Foddy, William (1993), *Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research*, Cambridge University Press, Cambridge.

Judd, C. M., E. R. Smith and L. H. Kidder (1991), *Research Methods in Social Relations*, Holt, Rinehart and Winston, Fort Worth.

Minichiello, Victor, Rosalie Aroni, Eric Timewell and Loris Alexander (1995), *In Depth Interviewing: Principles, Techniques, Analysis*, 2nd edn, Longman, Melbourne, Chapters 10 and 11.

CHAPTER NINE

Summarizing and presenting data

- **Categories**
- **Tables**
- **Graphs**
- **Means or averages**

You have collected your data. What are you going to do with the stacks of questionnaires, data sheets, or completed interviews? You will have made some tentative decisions about this when you prepared dummy tables earlier. Nonetheless, when confronted with a pile of data, new problems emerge, and further decisions will have to be made. Once data have been collected it is necessary to decide how they are to be summarized and presented.

Since this text presupposes no knowledge of statistics, some methods of data summarization and presentation will, of necessity, not be covered. This text also assumes that the projects undertaken will be very limited in scale so that computer analysis of data is not required. However, the logic of data summarization and presentation is the same regardless of the techniques used. Moreover, there is some merit in doing a few small projects 'by hand' as it were, in order to learn that logic.

To summarize and to organize your data involves three steps. First, *categories* must be selected in which the raw data can be summarized. Second, once the categories are selected the data are *coded*, that is, they are sorted into the categories. Finally, the data are *presented* in a form which facilitates the drawing of conclusions.

Categories

While data are collected in detail, they usually cannot be reported or presented at the same level of detail. In other words, it is unlikely that you will get to report all of the data that you have collected. In order to summarize and present data, tables, graphs, or charts are constructed; averages and percentages are calculated. In order to do this the data must first be categorized. We saw this earlier in the case of research into the effect time spent on studying had on academic performance. Let us presume that the data presented in Figure 9.1 were recorded on the data summary sheet suggested in Chapter 5, p. 67.

Figure 9.1 A completed summary data form for a study of the relation between hours spent in revision and result on a history examination

Student name	Number of hours spent in revision		Examination result	
	Raw score	Code	Raw score	Code
1	30		98	
2	25		99	
3	10		50	
4	12		44	
5	20		65	
6	22		68	
7	25		80	
8	30		75	
9	30		80	
10	20		60	
11	24		65	
12	19		55	
13	18		54	
14	21		58	
15	22		60	
16	24		62	
17	28		70	
18	26		70	
19	27		65	
20	24		60	
21	18		58	
22	19		57	
23	25		68	
24	20		65	
25	21		60	
26	14		45	
27	20		35	
28	22		50	
29	26		55	
30	10		40	

As it stands, no conclusions can be readily or reliably drawn from this data summary form. No pattern emerges from a quick scan of the data. The data are still in too detailed a form. More inclusive categories are required for reporting both amount of time spent in revision and result on the examination. Once the data are in hand it is possible to determine what the extremes were and what the average was. Both of these help to construct categories.

What are the extremes:

- for amount of time spent in revision? ______ most ______ least
- for result on examination? ______ highest ______ lowest

Scan the list and record the results.

What is the average:

- for amount of time spent in revision? ______
- for result on examination? ______

An average or the mean is calculated by totalling the measures (number of hours or result on examination) and dividing by the number of measures (in this instance students).

$$\text{Mean or average history result} = \frac{\text{Total of summation of history results}}{\text{number of students}}$$

Several ways of categorizing these data are now possible. The students could be classified into those who studied more than the average and those who studied less than the average. Similarly, the students could be classified into those whose results were above or below the average. Other ways of classification might include separating those who passed from those who did not. The results could be separated into high pass (65–100), pass (50–64) and fail.

Once the categories are selected the data are coded. That is, the raw data are reclassified into the more inclusive categories. Let us say that you decided to use the categories above average and below average for both number of hours spent in revision and for examination result. Go back to Figure 9.1 and codify the data. That is, after each raw score indicate the category into which it fits. For example:

Figure 9.2 A completed and codified summary data form

	Hours in revision		Result in examination	
	Raw score	Code	Raw score	Code
1 student	30	AH	98	AR
2 student	25	AH	99	AR
3 student	10	BH	50	BR

AH = above average hours
BH = below average hours
AR = above average result
BR = below average result

In this way the raw data are codified and can be more readily analysed.

If your calculations agree with mine, the average number of hours spent in revision was 652 ÷ 30 or 21.7 hours. Hence students who studied more than 21.7 hours (22 or more hours) are coded as 'AH' (above average) and those who studied less (21 and fewer hours) are coded 'BH' (below average). How many students are there in each code (classification)?

Number coded AH = 16
Number coded BH = 14

How about the examination results? What was the average result? My calculations were 1871 ÷ 30 or 62.4. Again, those students who scored over 62.4 (63 and above) were coded 'AR' (above average) and those who scored below 62.4 (62 and less) were coded 'BR' (below average). How many students fell into each category?

Number coded AR = 13
Number coded BR = 17

You have now codified your data and established the frequency of students appearing in each code. You are now ready to present your data in a form which will show the relationship between the two variables.

You can see that if you used different categories the coding would look different. To give yourself practice, copy out Figure 9.1 and codify the data results using high pass (65–100), pass (50–64) and fail (49 or less) as the categories. Whatever categories are used, the aim is to reduce the raw data to a more manageable set of categories. The categories are decided and then the raw data are coded into those categories.

The first two steps have been done. Categories have been selected and the data codified. How are they to be presented? The hypothesis guiding this research asserts that there is a relationship between the amount of time spent in revision and the result on an examination. This means that the way in which you present your data needs to show the relationship between the two variables: time spent in revision and examination result. There are several ways to do this. These are presented in the following figures.

Tables

The first form illustrated is that of tabular presentation.

Figure 9.3 A table for presenting the data from a study of amount of time spent in revision and result on an examination

Result on history examination	**Amount of time spent in revision**	
	Above average	*Below average*
Above average		
Below average		

In order to come up with the numbers to put in the table in Figure 9.3 it is necessary to cross-tabulate your data. That is, you have to locate each case of data collected (in this case each student) in the appropriate box of the table. For this example you would take each student listed on the data summary sheet in Figure 9.1 and place a tick in the appropriate cell (blank square) of a table like that presented in Figure 9.3. The first student listed, Gail, was categorized as 'above average' in both variables, so place a tick in the upper left-hand cell of the table. The second student, John, was also categorized as 'above average' in both variables, so place another tick mark in the upper left-hand cell. Student number three, Tim, was categorized as 'below average' in both variables so place a tick mark in the lower right-hand cell of the table. When all the data have been cross-tabulated in this way your preliminary, table should look like Figure 9.4.

Figure 9.4 The relationship between time spent in revision and result in a history examination (preliminary table)

Result on history examination Number of students	**Amount of time spent in revision Number of students**	
	Above average	*Below average*
Above average		
Below average		

When all the data have been cross-tabulated you add up the tick marks in each cell and put that number in the cell. What do your results look like? They should look like those in the table in Figure 9.5. There were eleven students who were above average both in examination result and in the amount of time they spent in revision. There were five who were below average in result but above average in study time. There were twelve students who were below average on both variables.

Figure 9.5 The presentation of the results tabulated in Figure 9.4

	Amount of time spent in revision Number of students		
Examination result	*Above average*	*Below average*	*Total*
Above average	11	2	13
Below average	5	12	17
Total	16	14	30

The numbers at the side and bottom of the table in figure 9.5 are called marginal totals. They are of course the same as the totals you calculated earlier for the frequencies of each variable. These serve as useful checks to make sure your coding and cross-tabulating were done accurately. It is amazing how many errors can creep in at this stage of the research process. It took me several attempts to get Figure 9.5 correct. The marginals must add to the total used for the construction of the table. They must also add correctly both across the rows and down the columns. It may seem a tedious exercise but it provides a critical check on accuracy.

How would you interpret the table in Figure 9.5? It shows a very clear relation between the two variables. It shows that the two variables are related in such a way that the more there is of one (study time), the more there is of the other (examination result) with few exceptions.

While interpreting Figure 9.5 is relatively straightforward as it stands, sometimes it is better to present the tabular results as percentages. There are two ways of doing this. Since each accurately reflects the data but does so in a slightly different way, the selection depends on which mode of presentation is easiest to interpret. Figure 9.6 presents the findings in Figure 9.5 as percentages of the total. In all tables giving the results as percentages it is very important to indicate the total number upon which the table is based. That is why

'n = 30' (which means the total number is thirty) is placed where it is. It is nearly a universal convention to use the lower case n to refer to the number of cases in a table or graph. Thirty is usually considered the minimum number of cases for the use of percentages in a 2 × 2 table like the one in Figure 9.6. The more cells a table has, the higher the number of cases should be.

Figure 9.6 The relationship between amount of time spent in revision and result on history examination

Result on history examination	Amount of time spent in revision	
	Above average	*Below average*
Above average	36.6	6.7
Below average	16.7	40.0
		n = 30 100%

Figure 9.7 presents the findings in Figure 9.5 in column percentages. That is, each of the columns add to 100 per cent. When you set up a table in the form of one shown in Figure 9.7, you show the impact of the column variable (in this instance, amount of time spent in revision) on the row variable (result in history examination). This is, of course, exactly what you wanted to do, because amount of time spent in revision was your independent variable and examination result the dependent variable. When you construct and interpret tables it is crucial to keep in mind which is the independent and which the dependent variable. Failure to do so can lead to some nonsensical interpretations of data.

Figure 9.7 The percentage of students spending above or below average amounts of time in revision who scored above or below average in their history examination

Result on history examination	Amount of time spent in revision	
	Above average	*Below average*
Above average	68.7	14.3
Below average	31.3	85.7
	100% n = 16	100% n = 14

How you read a table depends in part on which variable is the independent and which is the dependent variable. In this example, time spent in revision was the independent variable and examination result the dependent variable. Figure 9.7 would be read in this way. Among those students who spent an above-average amount of time in revision, 68.7 per cent received above-average examination results, while 31.3 per cent received below-average results. By contrast, among those students who spent a less than average amount of time in revising, 14.3 per cent received an above-average result in the examination and 85.7 per cent received a below-average result. It is therefore concluded that the amount of time spent in revision had a definite and positive effect on the examination results of this group of history students and the hypothesis is confirmed or accepted.

As a general rule, if you are presenting your data in tables using percentages it is best to percentage the independent variable across the dependent variable (as in Figure 9.7). In this way you display the impact of the independent variable on the distribution of the dependent variable, which is, of course, what you are trying to show. Thus while tables may be percentaged in a variety of ways, the most useful is one that percentages the independent variable across the dependent variable.

If you look back over Figures 9.4 to 9.7 it should become clear that the interpretation would be the same in each of these modes of tabular presentation of the data. Tabular presentation of data is very basic and very useful. To give yourself practice at tabular analysis take the data in Figure 9.1 and recode the exam result data into the three categories of high pass (65–100), pass (50–64) and fail (49 and less). Construct tables by cross-tabulating the data again. Present the tables numerically and as percentages of the whole, row percentages and column percentages.

There are other ways of presenting data as well. Remember, data are summarized and presented in such ways as to make clear the relationships that exist between the variables under study. Other ways of summarizing and presenting data include several kinds of graphs, the scattergram, and the use of means.

Graphs

Bar graphs

In order to prepare a graph it is necessary to perform Steps 1 (selecting categories) and 2 (coding the data) of data summarization and presentation. It is also necessary to cross-tabulate the data in some way. Take, for instance, the bar graph or histogram. In both these methods the amount of space given to each variable is proportional to that variable's portion of the sample. Figure 9.8 is a bar graph presenting the same data as Figure 9.4. Essentially, this bar graph presents the information in the top two cells of the table in Figure 9.4. Figure 9.8 shows a bar graph based on the frequency distribution of the data, that is, the numbers falling into each category. Figure 9.9 is a bar graph that gives *all* the data in Figure 9.4.

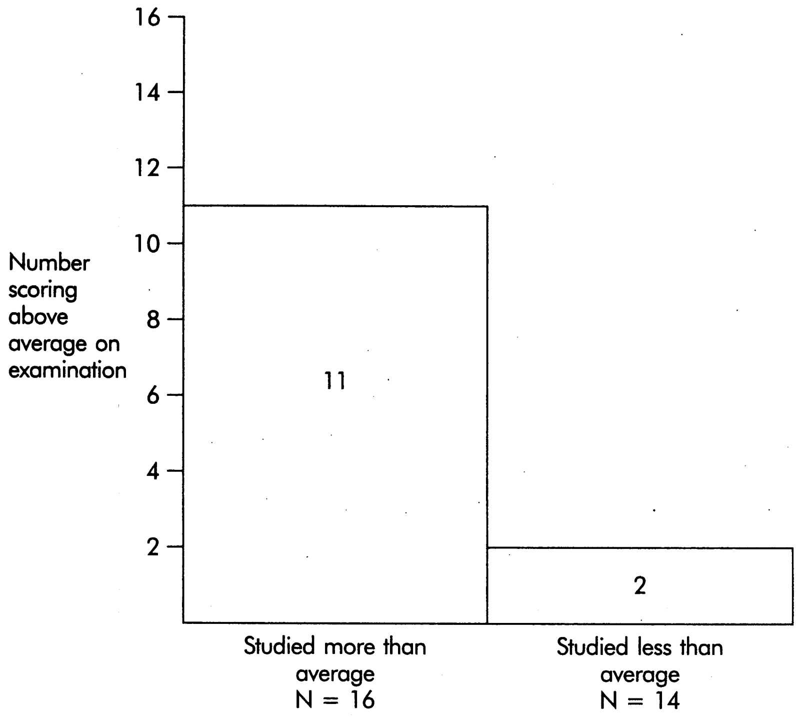

Figure 9.8 A bar graph showing the relationship between hours studied and history examination results

Bar graphs can also be used to present percentage data. Figure 9.10 presents the data in Figure 9.7 in the form of a bar graph. In this instance a table presented as column percentages is converted to a bar graph by making the space in the graph proportional to the percentage of each cell. The essential feature of a bar graph is that the size of the bar is proportional to the size of the variable. Again, it can be seen that different methods of presenting the same data when used correctly do not lead to different conclusions.

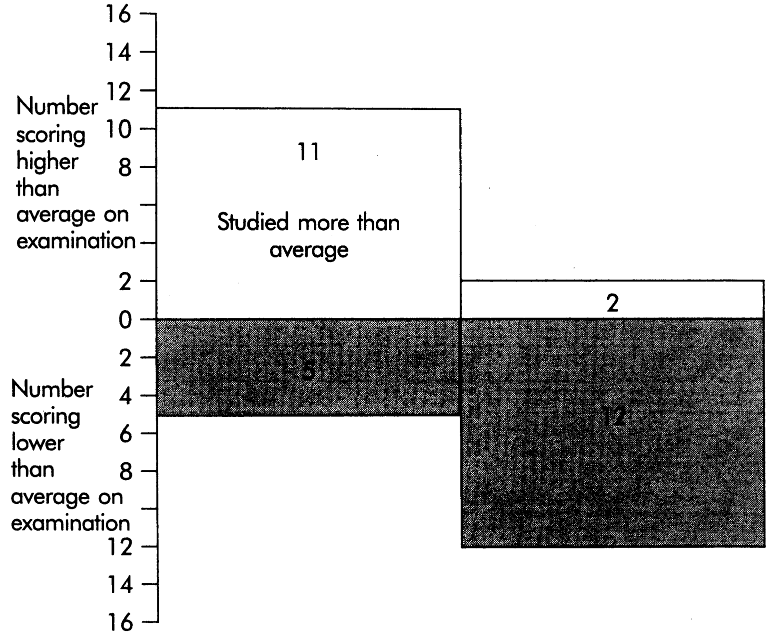

Figure 9.9 A bar graph showing the relationship between amount of time spent in revision and history examination result

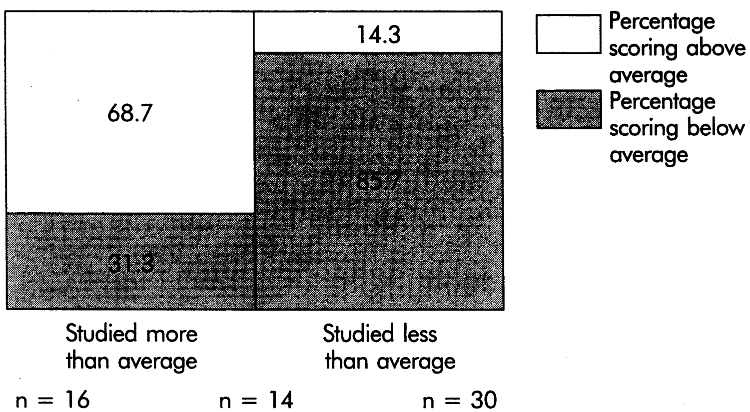

Figure 9.10 A bar graph depicting the relationship between the amount of time spent in revision and history examination result

Pie graphs

The pie graph is not particularly suited to the presentation of the type of data with which we have been dealing. It is much better for

depicting multiple responses, or many categories or sources, when the intent is to show the proportion of each relative to the whole. For example, if the ethnic composition of the students enrolled in your course is:

Asian	10%
Greek	10%
Australian	60%
Italian	8%
Other	12%

a pie graph could be constructed giving each group a wedge equivalent to their proportion of the population. The size of the wedge to represent the Asians must be 10 per cent of the circle's area. This can be measured by calculating what 10 per cent of 360 degrees is. (There are a total of 360 degrees in a circle.) Ten per cent of 360 degrees is 36 degrees. Using a protractor, count 36 degrees, place a dot at 0 and at 36 and draw lines to the centre point of the circle and you have a wedge of the pie equal to 10 per cent of the circle. Repeat this for each group. The next group, the Greeks, would require 10 degrees. Starting where you left off (at 36 degrees) count off 36 and place a dot at 72 degrees and draw a line to the centre point of the circle. Working by guess will not do in the preparation of pie graphs. However, computer programs are now designed to produce accurate pie, bar, and line graphs from data.

A similar pie graph could be constructed for the population of Australia. A comparison of the two graphs would show how the distribution of ethnic groups enrolled in your course compared with the distribution of such groups in Australia. However, so would the following simple table.

Figure 9.11 A table comparing the percentages of several ethnic groups in two populations

	Percentage of students in your course	**Percentage of Australian population* in 1986**
Asian	10	2.7
Greek	10	1.0
Australian	60	77.2
Italian	8	2.0
Other	12	17.1
Total	100	100.0

* These figures are from the Australian Bureau of Statistics and refer to birthplace of respondent.

Pie graphs in general are hard to construct accurately and very difficult to compare. For these reasons they are not usually used in scientific reporting but only in journalistic reporting, like newspapers and magazines. Annual budgets are also often presented in pie graphs. Figure 9.12 presents pie graphs comparing religious affiliation in Australia and New Zealand. The fact that the percentages are given as well as the visual impression of the different sizes of the various wedges helps one to interpret these graphs.

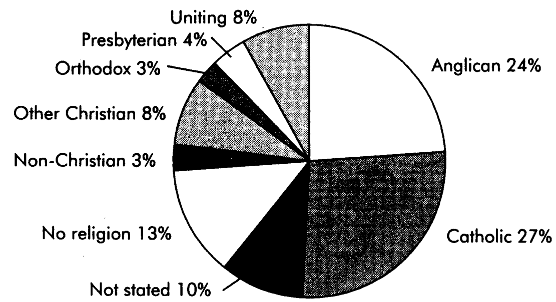

Figure 9.12a Australia

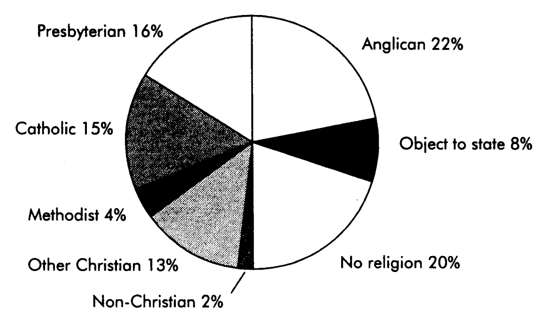

Figure 9.12b New Zealand

Figure 9.12 Examples of pie graphs in the presentation of data comparing religious affiliation in Australia (1991) and New Zealand (1991). *Source*: Australian Bureau of Statistics and New Zealand Bureau of Statistics.

Scattergrams

The scattergram is another way in which data can be summarized and presented. A scattergram is produced by pinpointing each instance of measurement on a grid defined by the two axes of a graph. Figure 9.13 shows such a grid.

The two lines along which the units are marked are called axes and the space between them is defined by the grid formed by the intersecting lines drawn from each unit point along the two axes. The first step in constructing a scattergram is to decide on the scale of units to be used on each axis.

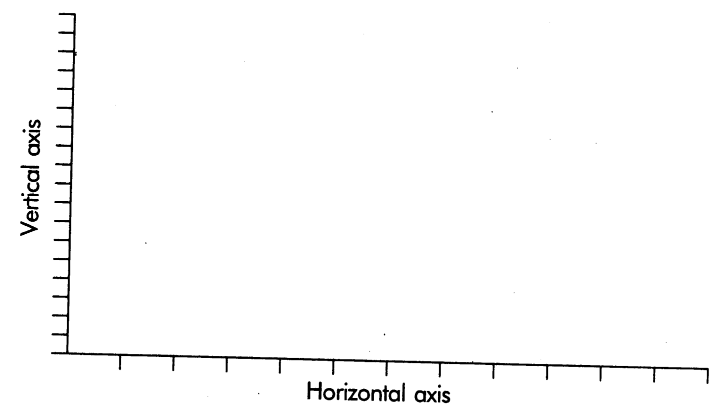

Figure 9.13 A scattergram grid showing horizontal and vertical axes

Data are not usually categorized and coded before constructing a scattergram. Instead the scale of each axis is adjusted so as to accommodate the range of the variable being analysed. Remember I suggested that you do the analysis of the data from the study of the impact of amount of time spent in revision on examination result. I asked you to identify the range of each variable by noting the extremes. This is a very important step if you wish to construct a scattergram. Re-examine the data presented in Figure 9.1:

What is the range of the values recorded for the variable 'Time spent in revision'?
Highest ______ ←————→ ______ Lowest
What is the range of the values recorded for the variable 'Result on an examination'?
Highest ______ ←————→ ______ Lowest

The scale of the units along each axis of the grid upon which you will produce the scattergram must be able to cope with the range of each variable. In this instance the scale of the horizontal axis, the one used to indicate hours spent in revision, must range from 10 (the lowest reported) to 30 (the highest reported). The range for the vertical axis, the axis dealing with examination result, must go from 35 to 99. Figure 9.14 presents a grid upon which a scattergram for the data presented in Figure 9.1 could be constructed. The scattergram is constructed by putting a dot on the grid in the place defined by the two pieces of data for each student. The use of graph paper makes this task much easier.

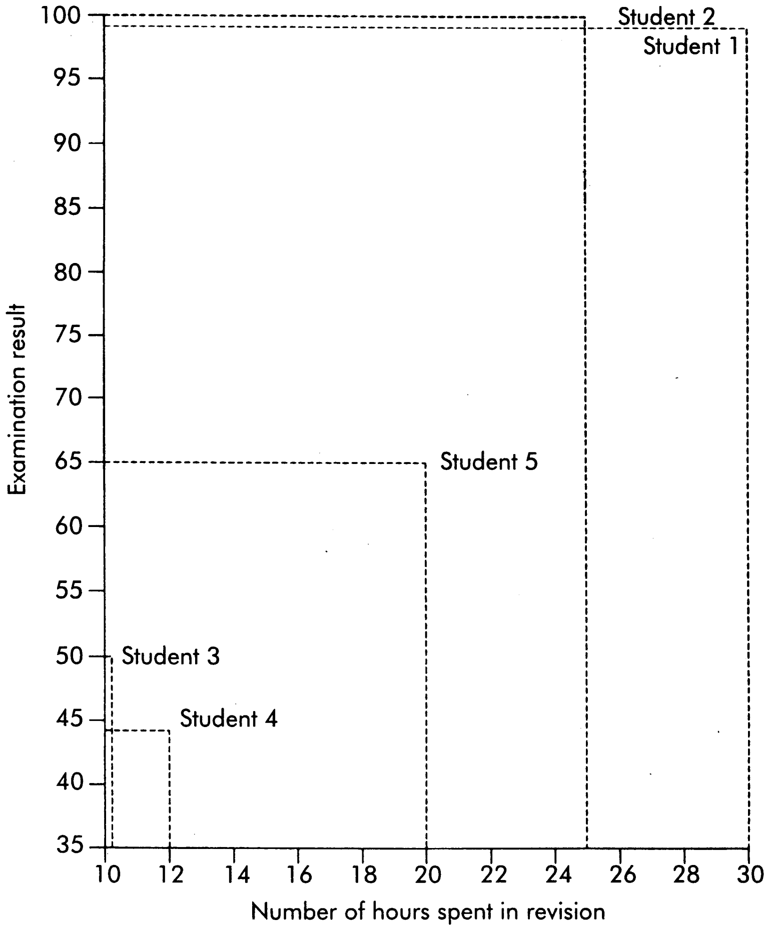

Figure 9.14 A grid for the construction of a scattergram for data on impact of amount of time spent in revision on examination result

The axes are drawn and units marked along them. Now a dot is placed on the grid for each student. Student number 1 studied 30 hours and received a 98, so place a dot at the intersection of a line drawn up from the 30 position on the horizontal axis with a line across from the 98 position on the vertical axis. The positions of students 1 to 5 are drawn in as examples.

Using a sheet of graph paper make a scattergram of all the data in Figure 9.1. Normally the intersecting lines are not drawn on the table. Rather two rulers are used to indicate where the lines intersect and only the dot is placed on the grid. Place two dots close together where two data points are the same. The result is a pattern of dots. What does the pattern of thirty dots tell you?

Line graphs

A line graph shows the relation between two variables by connecting the data points on a grid defined by two axes with a line beginning at the left and moving to the right. Figure 9.16 is an example of a line graph, the data for which were originally presented as in Figure 9.15.

Figure 9.15 Increased number of hours spent on household activities by non-employed persons due to presence of child.

Age of child	**Number of hours increased**
Less than one year	5.2
1st year	4.6
2nd–5th year	4.0
6th–11th year	4.5
12th–17th year	3.6

Source: Data from E. Walker and M. E. Woods (1976), *Time Use: A Measure of Household Production of Family Goods and Services*, Centre for the Family of the American Home Economics Association, Washington, D. C., pp. 50–1.

The data tabled in Figure 9.15 can be presented by a line graph (as in Figure 9.16). A time variable is often placed on the horizontal axis. The number of hours increased is put along the vertical axis. The units are clearly marked along each axis. Then the data points are put in place as for a scattergram. The data points are joined by a line beginning with the first dot on the left and moving to the next dot to the right.

As an exercise, convert your scattergram of the data on the relationship between number of hours spent in revision and examination result. To do so, start with the dot on the far left and

The scale of units selected is very important to graph construction. Figure 9.17 shows that by using large units change is underemphasized. On the other hand, using small units emphasizes the magnitude of change. Let us examine another example of this. There is a lot of talk these days about divorce statistics in Australia. The data on divorce rates in Australia are presented in tabular form in Figure 9.18. The data given in this figure are from the Australian Bureau of Statistics.

Figure 9.18 Divorce rates per 1000 population

Year ended	Australia	NSW	Vic.	SA	WA
1982	2.9	2.7	2.8	3.4	2.9
1983	2.8	2.6	2.6	3.3	2.8
1984	2.8	2.4	2.6	3.0	2.9
1985	2.5	2.2	2.4	3.1	2.8
1986	2.5	2.1	2.3	2.7	2.7
1987	2.4	2.1	2.3	2.9	2.7
1988	2.5	2.1	2.4	2.9	2.6
1989	2.5	2.2	2.4	2.6	2.6
1990	2.5	2.1	2.4	2.8	2.4
1991	2.6	2.2	2.5	2.9	2.7

When the data in Figure 9.18 are presented graphically many things can happen. We will first examine Australian trends in two ways. The same data presented in different ways give a very different impression. First, Figure 9.19 presents all the data.

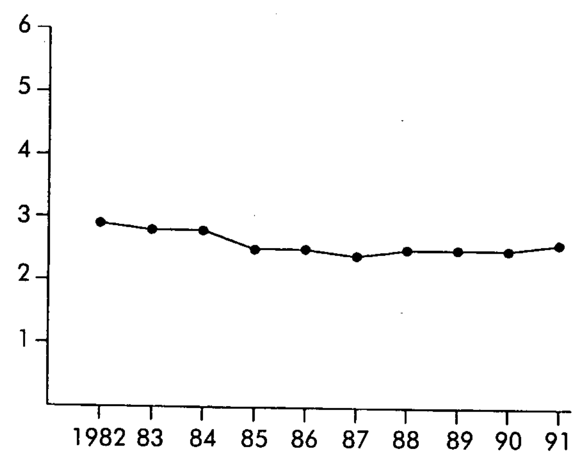

Figure 9.19 A line graph showing the Australian divorce rate from 1982 to 1991

Now let us assume that all the data we could find were the data for the last five years. Notice how two ways of presenting those data give very different impressions. Figures 9.20 and 9.21 show the same data recorded in graphs using different scales of units. Do they look the same? They are both accurate but they give different impressions.

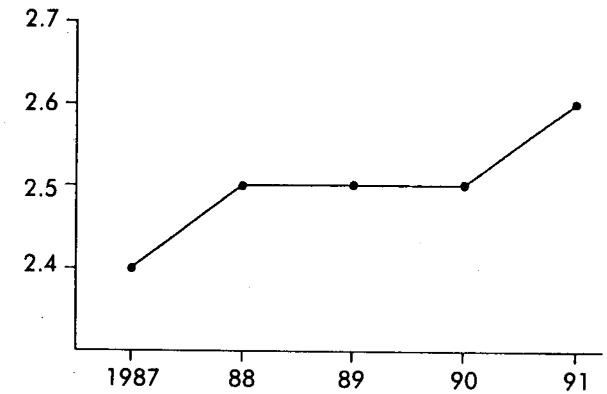

Figure 9.20 Australian divorce rate rising

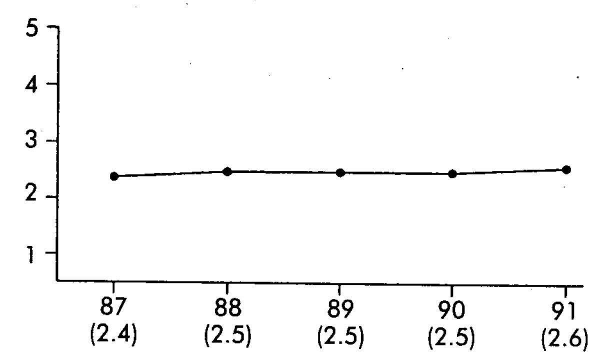

Figure 9.21 Australian divorce rate hardly changing

Both Figures 9.20 and 9.21 are accurate. They give different impressions. They both are risky in that they cover a brief period of time. The data for a decade (Figure 9.19) show that there has been an overall increase even when the hiccup in 1976–78 is accounted for.

As an exercise, make line graphs for several of the states from the data presented in Figure 9.18. Use different scales to see what difference is made. Try using data from every second year, even third year. What impression is given if you take 1976 as your starting point?

Finally, line graphs are very useful for comparing the trends or performance of several groups or persons. Figure 9.23 is a hypothetical line graph comparing the monthly energy consumption of three Australian households over one year. Energy consumption is measured in kilowatt hours (kWh). The raw data for this comparison are presented in Figure 9.22.

Figure 9.22 Energy consumption rates in kilowatt hours for three Australian households

Month	Household A	Household B	Household C
January	205	920	860
February	255	750	620
March	300	605	275
April	350	410	350
May	520	300	360
June	620	275	380
July	880	275	400
August	925	275	450
September	620	290	350
October	540	420	300
November	480	590	580
December	320	830	690

These three households have different patterns of energy consumption. The graph in Figure 9.23 shows these differences more clearly than the columns of numbers in Figure 9.22.

These three households are very different. Household A uses a great deal more electricity in winter than in summer. It may consume a great deal of energy heating the house and drying clothes. Household B may be in Brisbane and spend more on air conditioning in the summer than in winter. Household C might use air conditioning in summer and heating in winter. The usefulness of a line graph to compare energy consumption patterns can be seen clearly in Figure 9.23.

As a further exercise, construct a line graph comparing the divorce rates of New South Wales and South Australia (or any other state), using the data in Figure 9.18. By using a different style of line for each state you can construct a graph that provides a good basis for drawing comparisons among the states on divorce rates. Make sure you get the data points in the right place before you connect the lines.

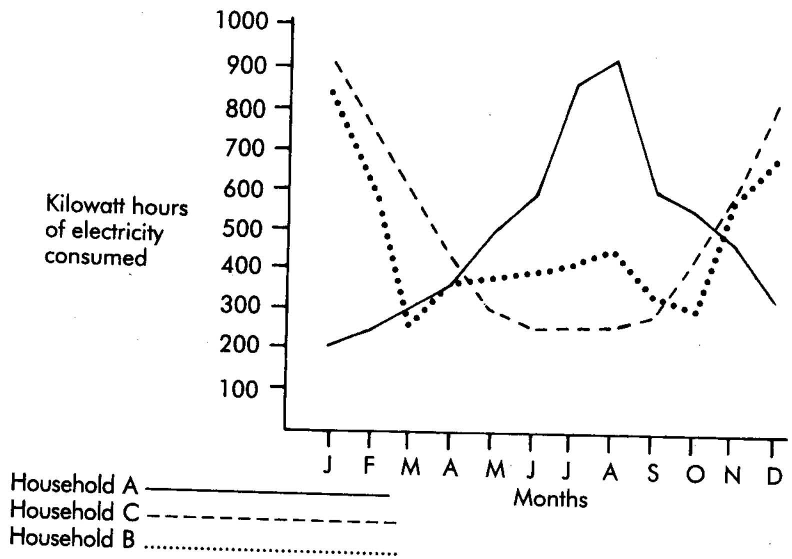

Figure 9.23 A line graph comparing the energy consumption patterns of three fictitious Australian households

Scattergrams and line graphs can be very useful ways of summarizing and interpreting data. They are frequently used in articles and books and research reports.

To construct a line graph you must:

1 Select categories for your data.
2 Code the data into the categories.
3 Select a scale of units for each axis.
4 Plot the data points.
5 Link the data points with lines.

Means or averages

Means, or averages, are often used to compare groups. Means are a useful way to summarize and present data. The average performance of groups, the average rates of consumption, or the average incidence of a particular event may be compared.

To calculate an average, or mean, one adds up all of the individual data and divides by the number of individuals. For example, you already calculated the average number of hours spent in study and the average examination result earlier in this chapter.

Means could be used to summarize and present the data from our study of the impact of the number of hours spent in revision on examination results. The class could be divided into two groups, those who studied more than the average and those who studied less than the average. Once this is done the average examination result for each group could be calculated. First, divide the students into two groups. Remember, the average time spent in revision was 21.7 hours.

Figure 9.24 The calculation of mean test scores for two groups of students

Group A Studied more than average		Group B Studied less than average	
Student	*Mark*	*Student*	*Mark*
1	98	3	50
2	99	4	44
6	68	5	65
7	80	10	60
8	75	12	55
9	80	13	54
11	65	14	58
15	60	21	58
16	62	22	57
17	70	24	65
18	70	25	60
19	65	26	45
20	60	27	35
23	68	30	40
28	50		
29	55		
16	1125	14	746
Group A average = 1125 ÷ 16 = 70.3		Group B average = 746 ÷ 14 = 53.3	

The data in Figure 9.24 would simply be reported in this way. The group of students who studied more than the average received an average result of 70.3 while the group of students who studied less than the average received an average result of 53.3.

As an exercise calculate the average number of hours spent in revision for each of two groups of students. First do it for those who received above-average results and then do it for those who received less-than-average results.

We saw that the move from a scattergram to a line graph involved the loss of a certain level of detail in the presentation of the data. When groups are compared using means, all variation internal to each group is lost.

Scattergram	Presents most information
Line graph, or bar graph, or pie graph	Present less information
Mean or average	Presents least information

Using the data in Figure 9.22 calculate the average yearly and monthly energy consumption of each household. Household A's average monthly energy consumption was 501.25 kWh, while Household B's average was 495.0 kWh and Household C's was 467.92 kWh. This example shows one potential problem with an average. It does not show the variation in the measures. While Households A and B have nearly the same average monthly energy consumption, they consume in very different ways.

While averages are very useful they must be used and interpreted with care. The average tells us nothing for certain about an individual in a group. It is not legitimate to imply that the average for a group applies to an individual. For example, the average household income of residents in a particular suburb might be $100 000 per year. There are many combinations of household income which would lead to that average. For example, one household earning $910 000 and nine each of $10 000 would result in an average income of $100 000. Some averages can be almost meaningless.

On the other hand, we are often interested in group performance and not so interested in the outstanding cases. The average is a useful indication of a characteristic of a group. Trends in averages, like trends in percentages, are particularly useful. The State Electricity Commission, in predicting energy supply requirements for the month of July, will rely on the trends in average energy consumption for the last twenty Julys. It will not be interested in the variations in individual household consumption.

Summary

Once your data are collected they are ready to be summarized and presented. To do this requires you to select categories in which to summarize your data. While you did some preliminary thinking on this when you constructed your dummy tables, the final selection is done when your data are in hand. Once you have selected categories the data are coded into the categories. Then the data are cross-tabulated in some way to show the relationship between the variables in question. We have looked at tables, graphs, and means as the basic techniques for summarizing and presenting your data.

Questions for review

1 What are the three steps involved in summarizing and organizing your data?
2 Why is it necessary to categorize your data?
3 Why is it important to keep straight which variable is the independent and which the dependent variable in constructing tables for the presentation and interpretation of your data?
4 What does it mean to cross-tabulate your data?
5 Describe the difficulties associated with your graphs.
6 What is a scattergram?
7 What are the advantages of using means or averages versus line graphs versus scattergrams in presenting data?

Suggestions for further reading

Babbie, E. R. (1990), *Survey Research Methods*, Wadsworth, Belmont.
Judd, C. M., E. R. Smith and L. H. Kidder (1991), *Research Methods in Social Relations*, Holt, Rinehart and Winston, Fort Worth.
de Vaus, D. A. (1991), *Surveys in Social Research*, Allen and Unwin, Sydney, Parts III and IV.

CHAPTER TEN

Doing qualitative research

- What is qualitative research?
- When to use qualitative research
- Which is better: qualitative or quantitative?
- The qualitative research process
- Phase one
- Phase two
- Phase three

What is qualitative research?

Quantitative research is designed to give numerical results, which can be reported in tables, graphs and charts telling the number of something, the proportion of something, or what the trends are. Qualitative research has a different goal: it answers the question, 'What is going on here?'. Qualitative research is designed to provide an impression; to tell what kinds or types of something there are; to tell what it is like to be, do or think something. Qualitative researchers exercise great discipline in order to find out 'What is going on here?' from the perspective of those who are in the situation being researched. Drawing an absolute line between qualitative and quantitative research is never satisfactory. There are similarities, and some techniques can be used in either type of research.

An example may help here. I recently conducted a study on the role of religion in the way Muslim migrants have settled in Australia (see Further reading). I wanted to gain two different types of information. First, I wanted to know how many Muslims there were in Australia, what their education levels were, what their employment levels were and where they were located. This information could largely be gained by reference to census and other government statistics. However, I also wanted to find out what it was like to be a Muslim living in Australia at this time. This could not be ascertained by reference to census data. To find out what it is like to be a Muslim living in Australia, it was necessary to listen to Muslims describe their lives.