Gerald's Column
by Gerald Fitton

For some time now I have been promising to discuss the use of numerical data to support a statement in a report. This month I shall introduce you to the concept of Illustrative Tables. I shall explain how to create Illustrative Tables which support numerical statements.

Introduction

Numerical data may appear in a publication for two quite distinct reasons. It may exist as a store of quantitative information or it may be used to support a statement.

I shall refer to tables used as a store of quantitative information as 'Reference Tables'. Nowadays the storage medium for this type of data is a computer system.

What I shall refer to as 'Illustrative Tables' are used to illustrate, support or reinforce a comment in a report. Charts (graphs) are also used for this purpose—but I shall discuss 'Illustrative Charts' on some future occasion.

Reference Data

Reference data is obtained directly from a statistical survey conducted by you or someone else. Reference data is stored in a Reference Table.

Reference data must never be stored as a chart because it is difficult and generally less accurate to read numerical values from a chart.

Incidentally, data you collect yourself is called Primary Data. Data collected by somebody else is called Secondary Data. Most people use Secondary Data because it is readily available even if it is not exactly suited to your purpose in the way that your own survey would have been.

Reference data is the starting point for both Illustrative Tables (numbers) and Illustrative Charts (pictures). For my worked example the Reference data used as the starting point has been collected by the Department of Transport (a Government Department) so it is Secondary Data.

Tables or Charts?

Illustrative Tables and Illustrative Charts are used to support written comment in a report.

Illustrative Tables are used to support quantitative (ie numerical) statements whereas Illustrative Charts are used to show qualitative comparisons (such as growing bigger or smaller—but without numbers). I shall return to this question of when to use Table and when to use a Chart sometime in the future! Watch this space!

Reference Tables

Of course Reference Tables can be quite small but it is more usual for them to be the result of large surveys. Many large surveys are conducted by Government Departments including the Department of Transport and the Office of Population Census and Surveys.

The example shown below is a simplified version of a more comprehensive table given in the Central Statistics Office publication Social Trends. I can not emphasise enough that the original table is huge compared with the version I am including here. I have reduced its size (by consolidating data—see below) so that I don't clutter up the magazine with information that you might find totally uninteresting and irrelevant!

Table 1　　Motor vehicles currently licensed in the UK:
by type of vehicle and year.

thousands

Type of Vehicle	1961	1966	1971	1976	1981
Private cars and vehicles	6114	9747	12361	14373	15632
Motorcycles, etc	1842	1430	1033	1235	1386
Public transport vehicles	94	96	108	115	112
Goods vehicles	1490	1611	1660	1796	1771
Agricultural vehicles	481	478	450	414	373
Other vehicles	206	260	247	300	340

Original Source: Department of Transport

Why Use Illustrative Tables

For my worked example I would like you to imagine that you have written a report about the number of vehicles on roads in the UK between 1961 & 1981. Your report includes the following statement which you wish to 'prove' with a Table containing numbers.

"The number of licensed vehicles almost doubled between 1961 and 1981. The increase was mainly due to the growth in the number of private vehicles."

When writing reports it is tempting to photocopy (or nowadays you might scan) the original Reference Table and include it as an Appendix to the report. I have read reports which include lines such as: "As you can see from Table 10 it is obvious that . . . " Then I look at a massive Table 10 and say to myself "It's not obvious to me!". Even in the simplified form shown above it is not easy to see how Table 1 'proves' the two points made in the statement. I can assure you that the original raw data from the Department of Transport (with many sub divisions of the table by engine size and date of first registration etc) is totally incomprehensible without intensive and prolonged study (and many 'sums').

What I am trying to convince you of is this. It is inevitable that a reference table will include more data than you require to illustrate your statement. You have to 'remove' this extraneous data, do some calculations and present the processed data in a more easily understood form if you are to convince your reader that the figures do prove your point.

Table 2 contains the same data after processing. I hope that you find it much more convincing. Please read the statement again whilst looking at Table 2.

"The number of licensed vehicles almost doubled between 1961 and 1981. The increase was mainly due to the growth in the number of private vehicles."

Table 2   Private and other vehicles currently licensed:
by year (1961 to 1981) and type of vehicle.

United Kingdom			thousands
Year	Type of Vehicle		All
	Private	Other	Vehicles
1961	6 100	4 100	10 200
1966	9 700	3 900	13 600
1971	12 400	3 500	15 900
1976	14 400	3 900	18 200
1981	15 600	4 000	19 600
Average	11 600	3 900	15 500

Original Source: Department of Transport

The column 'All Vehicles' has been added. Between 1961 and 1981 the number of vehicles has increased from just over 10 000 to just under 20 000 (thousands). Yes! The total number of vehicles has roughly doubled. 'Type of Vehicle' has been split into 'Private' and 'Other'. The number of 'Other' Vehicles has hardly changed; the increase must have been due to the increase in 'Private' Vehicles—indeed they have increased from 6 100 to 15 600 (thousand)—a considerable increase.

The Five Essentials

Every Table and Chart must include:

1 A Title—this is a description of the data and the way in which it is classified
    For example: Private and other vehicles currently licensed:
             by year (1961 to 1981) and type of vehicle.
    The Title occupies two lines. The first line is a description of the data (Vehicles currently licensed) and is terminated by a colon. The row classification (Year) precedes the column classification (type of vehicle).
2 Where the survey was conducted (usually a geographic area such as the UK)
3 When the survey was conducted (year, sometimes the date and time!)
4 The units used for the data (in our example thousands of vehicles)
5 The Source (who conducted the survey—not the publisher)

Converting Reference Tables to Illustrative Tables

Reference tables are 'converted' to Illustrative tables by the application of six 'Rules'.

These are:

1 Enter either averages or totals of the rows and columns.
2 Put numbers which are to be compared in columns not rows.
3 Round to two effective digits.
4 Arrange the rows and columns in a natural order, eg time or size.
5 Consolidate irrelevant data.
6 Use the layout to guide the eye.

Enter either Averages or Totals of the Rows and Columns

The reason for this rule is that averages or totals allow the reader to see the overall pattern in the data by scanning the margins. This puts the details of the table into the context of the overall trend. In the case of the example of Table 1, it is appropriate to find the total number of vehicles licensed during any year but to average the values for the type of vehicle. In some cases, instead of using the arithmetic mean as the average, a different measure of central tendency such as median or mode might be more appropriate.

By including an "Average" you allow the reader to compare each entry with this average and decide, for example, whether 1971 was a 'good' or 'bad' year for Motorcycles.

Put Numbers which are to be Compared in Columns not Rows

It is much easier to compare numbers which are aligned vertically rather than horizontally. Try to answer this question for the numbers below: "Is the bigger difference between the first or second pair?"

First try with the numbers in the same row: 16 153 & 14 752 & 13 451.

Now try again with the numbers arranged vertically.

    16 153
    14 752
    13 451

In my worked example (the number of vehicles) the statement is about the changes from year to year rather than between different types of vehicle.

Let me remind you: "The number of licensed vehicles almost doubled between 1961 and 1981. The increase was mainly due to the growth in the number of private vehicles."

The comparison from year to year will be easier if the rows and columns are interchanged. Use abbreviations or hyphens in the column headings if it is necessary in order to make the headings fit the available space but do interchange rows with columns if you need to.

For numbers of a thousand or more do not run the numbers together. For example the number written as 15632 in Table 1 is wrong. On the Continent of Europe the use of a comma (as in 15,632) would be interpreted as 15.632 since the comma serves the same function as our (full stop) decimal point.

The recommendation for printed tables is a gap (ie 15 632) rather than a comma.

Round to Two Effective Digits

It is much easier to compare numbers which vary in only two digits.

Try to check if the difference between 4 886 & 6 181 is bigger or smaller than the difference between 6 181 & 7 309; round the numbers (4.9 K & 6.1 K or 6.2 K & 7.3 K) and try again.

I guess that all of you will know about rounding either to significant figures or decimal places but only a few of you will have been told about rounding to 'Effective Digits'. The numbers 2 114 & 2 186 & 2 318 & 2 475 & 2,565 all start with 2 000 so that the '2' is not an effective digit. To two effective digits these numbers round to 2 110 & 2 190 & 2 320 & 2 480 & 2 560.

Original	Rounded
2114	2110
2186	2190
2318	2320
2475	2480
2565	2560

Note that the 'even number rule' for rounding is used to round 2 475 and 2 565. 2 475 is rounded up to 2 480 because 8 is even, whereas the 2 565 is rounded down to 2 560 because 6 is even. This 'even number rounding rule' for a final 5 has been used by mathematicians for centuries but was recently 'rediscovered' by computer programmers only after some serious systematic errors (gradually increasing the total) had some rather drastic results.

A final zero does not need rounding; the digits 1, 2, 3 & 4 are rounded down; the digits 6, 7, 8 & 9 are rounded up. The digit 5 (as a final digit) should be rounded up half the time and rounded down half the time to avoid a systematic increase in a total.

Average or total the numbers before rounding them. Have a look at the row corresponding to 1976 in Table 2. The value of 14 400 + 3 900 is written as 18 200 and not 18 300.

There is a temptation to adjust values after rounding to force the rounded totals to equal the total of the rounded components. You must never do this. Include a footnote such as:

"Components may not add up to the totals because they have been rounded independently".

 I have not included this footnote on Table 2 so that I could surprise you with it now! Normally I would include it immediately below the line containing the "Original Source".

Arrange the Rows and Columns in some Natural Order

Time is an important variable. In government publications it is conventional to put the most recent date as the lowest row or in the column furthest to the right. If the numbers are not time related then put the largest value at the top or left. If there is a change of ranking within the table, then use either the most recent data or the averages (the average used can be the mean, median or mode) to decide the order.

In the example, the rule about ranking by magnitude will require the column for Goods Vehicles to be placed further to the left and Public Transport vehicles further to the right.

Any catch-all category, such as Other Vehicles, is always the last column or lowest row, even if its values are larger. Although the total number of vehicles increases with time, the "latest date last" rule takes priority over the "largest number first" rule—particularly because the statement we are trying to support is about time dependence.

Consolidate Irrelevant Data

In the worked example, this can be effected by reducing the table to only two types of vehicle, namely Private Vehicles and everything else!

Usually the statement requiring illustration is more complex than this example, or the trend could be hidden more deeply in the data.

Have a look at what happened to Motorcycles during this period. In the early 1960s motor cycles were much cheaper than cars to buy and to run. There was no image problem with taking your girl for a ride on a motorcycle or scooter. Cars became cheaper in comparison with two wheelers and our motorcycle industry collapsed. Motor cycles staged a 'come back' as a second vehicle when they were produced in glorious colours, with weather protection and electric starters.

Can I suggest that you might think about how you would illustrate the points I have just made about motorcycles.

It is important not to try to make more than three points with one table or chart. Part of the reason why reducing to two types of vehicle "works" in this instance is because there are more Private Vehicles than all other types added together.

Use the Layout to Guide the Eye

This is a difficult one. Ideally you should talk to a graphic designer!

For tables that you are having printed, consult a graphic designer about fonts, styles, etc. I shall leave this for now because there are many Archive back numbers in which the use of typeface, font size, bold, italic, etc, have been explained and demonstrated in great detail.

In printed tables vertical lines should never be used.

Underlining has a special meaning in both Reference and Illustrative Tables. Look at Table 2. The underline running from "United Kingdom" to "thousands" means that the words immediately above the line apply to everything below the line (up to a 'rule off'—which this table doesn't have). The underline under the words "Type of Vehicle" means that the types of vehicle have been split into the two types "Private" and "Other".

Another Time

All these 'Six Rules' for converting Reference data to an Illustrative Table can be carried out with a suitable spreadsheet. When I was taught about Technical Report Writing the tables had to be constructed from the Reference data using manual techniques. Indeed the Reference data would be printed sheets often of many Foolscap (near A4) pages.

Now that most Reference data is provided as a computer file the use of spreadsheets to apply these 'Six Rules' is a 'must'. I shall explain how to use a spreadsheet for this purpose another time.

Illustrative Charts

Just a quick word about Illustrative Charts. When writing Technical Reports I used to produce these by hand. Well, that isn't quite true because I had a team of 'Computors' (note the 'o') to plot the graphs which would go into my reports! These 'Computors' were invariably young girls (cheaper then men to employ) with a 'Good O Level' in Maths when a 'Good O Level' included the skill of mental and written arithmetic. The young girl who became my wife, Jill, was one of my 'Computors'.

Nowadays these Illustrative Charts are invariably produced with a computer spreadsheet package. Unfortunately graphics packages have encouraged laziness and the Charts are not as Illustrative as perhaps, with a little more effort, they could be. I shall return to this in another article (if I am given some encouragement)!

Finally

Thanks for all your letters and emails. Please write to me at the Abacus Training address which you'll find in Paul's Fact File at the back of Archive.