WP. 15

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Joint UNECE/OECD Work Session on Statistical Dissemination and Communication
(14-15 February 2005, Henley-on-Thames, United Kingdom)

Topic (iii): How to train and educate statisticians to tell the story behind the numbers

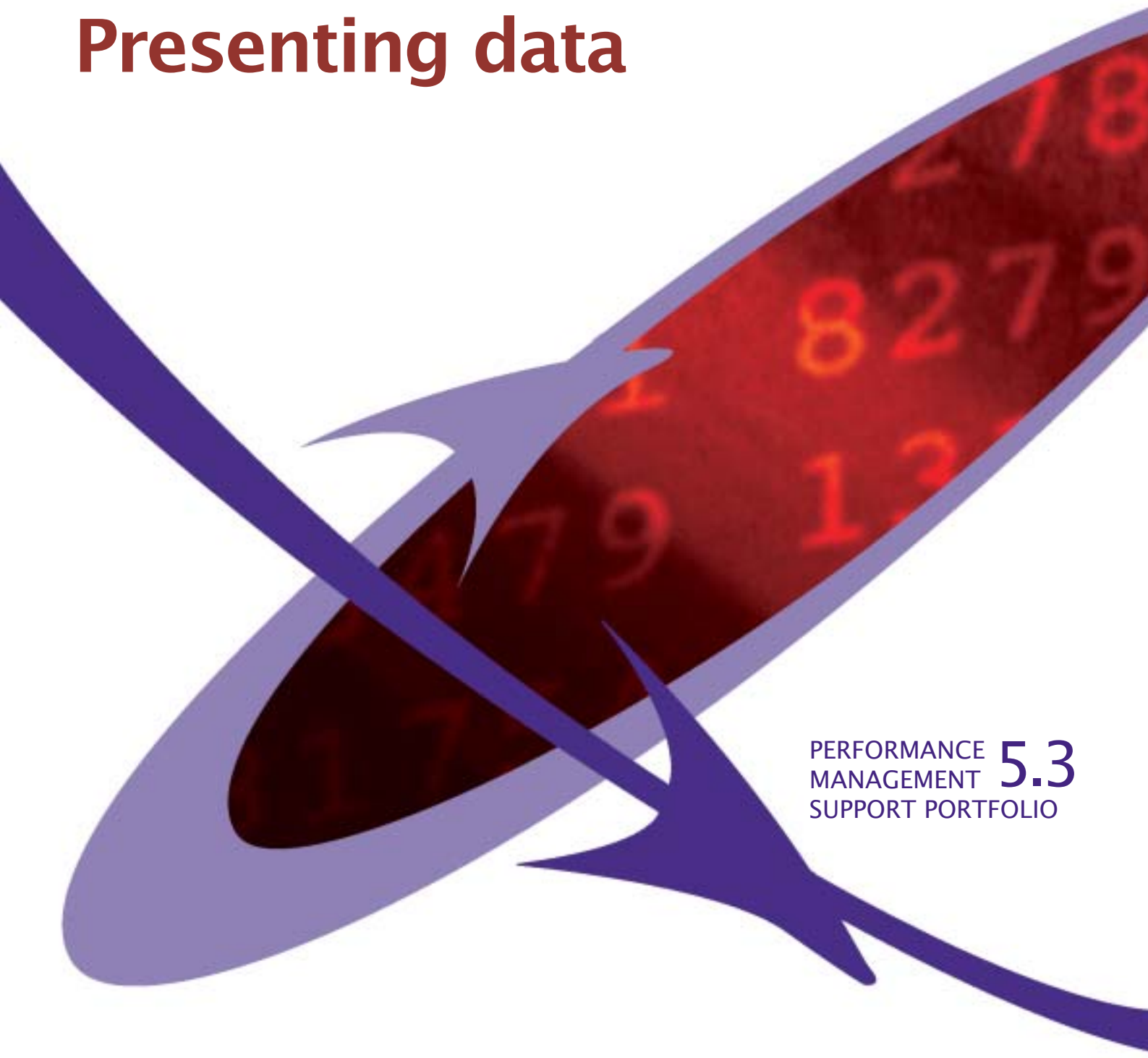
PRESENTING DATA

Supporting Paper

Submitted by Local Government Data Unit, Wales, United Kingdom[1]

[1] Prepared by Ed Swires-Hennessy (ed.swires-hennessy@dataunitwales.gov.uk).

Presenting data



PERFORMANCE MANAGEMENT SUPPORT PORTFOLIO 5.3

UNED DDATA LLYWODRAETH LEOL ~ CYMRU
LOCAL GOVERNMENT DATA UNIT ~ WALES

This document is part of the *Performance Management Support Portfolio* – a series of guides to the key elements of Performance Management, produced by The Local Government Data Unit – Wales and the Welsh Local Government Association.

Each document can be accompanied by the provision of advice, guidance and training support. For further information, or for support in using the Portfolio within your organisation, contact:

Local Government Data Unit – Wales
029 2090 9500

WLGA
029 2046 8600

ISBN - 0-9545990-7-1

Published by:
Local Government Data Unit – Wales
Columbus Walk, Cardiff, CF10 4BY
Telephone 029 2090 9500
Email enquiries@dataunitwales.gov.uk
Web www.dataunitwales.gov.uk

Contents

Presenting data

1. Introduction

The challenge of presenting data faces us all at some time whether for a simple report or for some full analysis of a problem or service. As part of the Performance Management Portfolio of support materials, this paper sets out some principles that will assist with presentation and aid the understanding of the data – and, hence, the overall message.

Numbers should not be treated like our bank balances with every last digit of the pence accounted for. Tables, charts and diagrams should assist with the interpretation of the data – and be representative of the data not, as frequently occurs, obscuring the meaning of the data.

Many of these principles were taught to us in school but we have forgotten some of the rationale with the advent of technology.

Therefore, this paper is aimed at Local Authority Officers who are required to present information in a clear and concise method in order to improve the understanding of the data and who, often, have little or no training in this field.

1.1 Reference data and summarised data

This paper is concerned with summarised data not reference data. Reference data can be defined as the raw counts of data which have a place as the foundation of the summarised data. Summarised data are data taken from the reference data, rounded and extracted, but not showing the full detail of the original data.

As an example, consider the simple table of population data for Wales in the following reference table:

Table 1: Population of Wales

	Number
1901	2,013,012
1911	2,421,265
1921	2,656,621
1931	2,593,982
1951	2,599,654
1961	2,644,211
1971	2,731,456
1981	2,792,935
1991	2,896,474

An example of a summary table from this reference table could be:

Table 2: Population of Wales

	Millions
1901	2.0
1931	2.6
1961	2.6
1991	2.9

Now it is much clearer to the reader what happened to the population from the data given in Table 2 – a steep rise to 1931, stagnation to 1961 and a smaller increase to 1991. The derivation of the same message from Table 1 would have been much harder.

2. Presenting numbers

2.1 Effective digits

The basic principle expounded by many is that the ordinary human being can discriminate only up to two digits in numbers. This principle is key to good presentation. As an example, consider the difference between the following two numbers:

12,573,981 and 11,894,397

If you now think about the process the mind went through:

1. There are too many digits to deal with;

2. Assess the size of the numbers – both two digits of millions (assessed by effectively ignoring six out of the eight digits in each number); and

3. One is '12' and one is '11' millions, thus the second number is smaller than the first.

You may take a calculator to find the actual difference – but that is rarely required to get the understanding of the numbers. Following the basic principle on rounding to **two effective digits**, the numbers would have been written:

12.6 millions and 11.9 millions

Here, over half of the original digits have been thrown away – but the appreciation of the numbers themselves and their differences can be quickly assimilated.

Similarly when numbers are derived, for example percentages, it is tempting to write as many digits as the calculator or spreadsheet provides. Often this is two decimal places: so we could have 77.42 per cent in one category and 22.58 in the other.

Using the rounding principle it would be 77 per cent in one category and 23 per cent in the other – much more quickly perceiving the difference. Another advantage here is that many people can readily discriminate between numbers up to 100.

2.2 Numbers in text

The general rules are as follows:

- Numbers up to and including nine in text should be written in text, ie “one”, “two” etc. Larger numbers should use digits, “11”, “37” etc;
- Use effectively rounded numbers for up to 10,000; so use “6,200” not “6,248”; and
- For numbers over 10,000 use a mixture of digits and words; “13 thousand” not “13,169”; similarly use “1.1 millions” not “1,148,982”.

Often we are tempted to put all the possible digits into our text because we believe this gives the most accurate number. In fact it is the most precise – but this precision implies spurious accuracy and is often misleading. Even with population counts, it is obvious that a number given to the nearest one is precise, but not accurate to that level: some will have been missed. The 2001 Census counters have estimated that there is a two per cent under-recording of the population: that means an extra 60 thousand people in Wales (and over a million for the UK as a whole!).

3. Numbers in tables

3.1 Table layout

Most people believe that they can put data into a table although few do it effectively. Again, some basic principles have to be observed as well as taking account of the fundamental education that the brain has been given.

Many tables of data are today prepared using computer packages. This should provide neat tables but, occasionally, this is at the expense of understanding. The following (abbreviated) tables were extracted from the web site of an international organisation:

Table 3: Citrus prices

	1995/96	1996/97	1997/98	1998/99	1999/2000
	national currency/kg				
ORANGES AND TANGERINES					
Germany (DM) Spanish navels	1.46	1.43	1.38	1.53	1.27
Spanish clementines	2.25	2.16	1.97	2.09	2.04
United States (cents) California navels	66.48	65.58	90.06	132.19	68.11
Japan (yen) Average	181.44	293.58	269.17	264.08	237.1

Table 4: Exports: Total citrus

	1995/96-1997/98	1998/99	1999/00	2000/01
	Average			
World total	**9 438**	**9 005**	**9 478**	**9 518**
Spain	2 783	2 793	3 221	2 859
South Africa	651	722	691	820
Italy	197	112	212	225
Others	2 413	2 416	2 163	2 418

The major difference in the presentation is that the numbers in Table 3 are left justified and those in Table 4 centred within the columns. Back in primary school, in mathematics, we were carefully taught that columns in numbers were important and in sums we always had to observe layout under column heads **"h, t and u"**: for hundreds, tens and units.

For that period, it was the way to handle numbers: now the same principle needs to be employed so that the interpretation of the table can be easily effected. Then numbers that have more digits to the left are 'larger' than the others – and, if we think about how we interpret and compare numbers with many digits, it would be obvious that we look only at the digits on the left of the number to start with, in order to get some basic idea of size.

However, even in this process, we rarely consider more than the first two digits of the numbers! Table 3 destroys this rationalisation in two ways: first the numbers are left justified, taking away the basic comparison; second the last number in the table only has one digit after the decimal point.

These two tables are from the same part of the organisation's web site – and both make the user work in non standard ways to understand what the tables are saying.

All tables in a publication should be in a similar format. A sample table layout is given in Annex 1.

Tables should not be filled with grid-lines. To enable the user to read the data effectively blank lines can be used as a separator. The convention is that, in a long table, rows should be grouped in sets of 5 followed by a space line. This allows the user quickly to identify the number required without getting out a ruler.

When data for a run of years is shown in the table, the summary of the years is not shown in the table title (as in Table 1): where the data are for a single year, the date is shown (as in Table 5).

3.2 Logical ordering

Another principle concerns the illustration of time. If you were presented with three numbers on a page as:

26,300 27,400 20,500

and then told that they represented the number of new jobs in three successive years, you would suppose that the first number was the earliest and the last number was the latest.

If the numbers were presented in a column as:

26,300
27,400
20,500

again you would assume that the 26,300 as the first figure was the first year's number and the 20,500 was the latest year. Why? Because we read from left to right and top to bottom. Our minds presume that time and data run from left to right or from top to bottom. Anyone who presents data the other way round is asking for poor understanding.

A further way to help the reader understand the message concerns the ordering of the rows within a table. Where a table is presented to illustrate some characteristic, rows may be sorted, for example, from highest to lowest: this will assist the reader to see clearly what message is being written about.

Where a table is presented just to show the information, a natural – but user-friendly – order is to be preferred. For example, if the rows relate to electoral divisions, it would not make sense to mark them 1, 2 etc. Instead the names would be given. These should be in alphabetical order – as most readers would expect this. The only real difficulty here, arises when the table is bilingually prepared with the division names in English to the left of the table and the names in Welsh to the right: whichever list is sorted alphabetically, the other will be out of order. Here one has to resort to another logical order – which may be the numeric order of the divisions.

For tables showing local authorities' data in Wales, an agreed order has been defined which is based on geographical position in Wales, starting in the top left (Isle of Anglesey) and concluding with Newport. The ordering of the list can be found in the latest edition of 'Digest of Welsh Local Area Statistics' (see Annex 2).

Another of these taught principles is where the total of rows or columns should appear. If a sum was presented as either:

i. 357+621 or

ii.
```
  357
+ 621
```

then there is no doubt as to where anyone would put the total. In instance (i) it would be to the right of the numbers; in instance (ii) it would be below the numbers.

One Government Department does it counter intuitively and puts the column totals at the top and the row totals to the left: everyone joining that Department has to readjust the ideas which they were taught in Primary School!

3.3 Comparing numbers

When comparing numbers, it is easier to have the numbers in a column than in a row.

Table 5: Population of part of Europe, 1995

Number

Country	Population
France	58,020,376
Denmark	5,215,732
Germany	81,538,628
United Kingdom	58,491,643
Luxembourg	406,589
Total	203,672,968

So if a message is to be derived from a run of numbers, put them, where possible, in a column. So, in Table 5, the understanding of the data is helped by the numbers being in a column, the mind looks to the left digits of the numbers and scans them. In Table 5, three countries have a digit in the most left position Germany – with an '8' is the largest, with the other two having a digit in this position, which are both the same i.e., '5', so we then move to the second digit of the numbers to be able to discriminate, which in this case just happens to be the same, so to the third digit and the United Kingdom has the higher third digit i.e. '4'.

A better presentation of these data is given in Table 6. Here less digits allow the understanding of the data to be grasped more quickly. Note also that the re-ordering by size also assists the understanding of the data; that the descriptor line has been removed and explanations given below the column headings, because the type of data cannot be explained by just one description and that the smaller data entries are rounded to two effective digits to allow Luxembourg an entry into the table, with the other entries rounded to the same number of decimal places.

Table 6: Population of part of Europe, 1995

Country	Population (Millions)	Per cent of total
Germany	81.5	40.0
United Kingdom	58.5	28.7
France	58.0	28.5
Denmark	5.2	2.6
Luxembourg	0.4	0.2
Total	203.7	100.0

Principles for tables:

1. Round data in summary tables;
2. Right justify numbers in columns;
3. Make sure that all tables in a publication are in a similar format;
4. Show time either from left to right or top to bottom on a page;
5. Show row totals to the right and column totals to the bottom of the table;
6. Put data to be compared in columns rather than rows;
7. Use space to separate data not lines; and
8. Keep tables as simple as possible.

4. Using charts and diagrams

4.1 Clarity of message

**'A picture is worth a thousand words'
says the old adage.**

This is no doubt true, but if and only if the picture is a true reflection, in this case, of the data.

So often, particularly with the ability to produce charts and diagrams almost instantly on personal computers, the wrong or an inappropriate diagram is produced which then gives the wrong message.

To be effective in the presentation of numbers in charts, graphs and diagrams, you have to do several things at once:

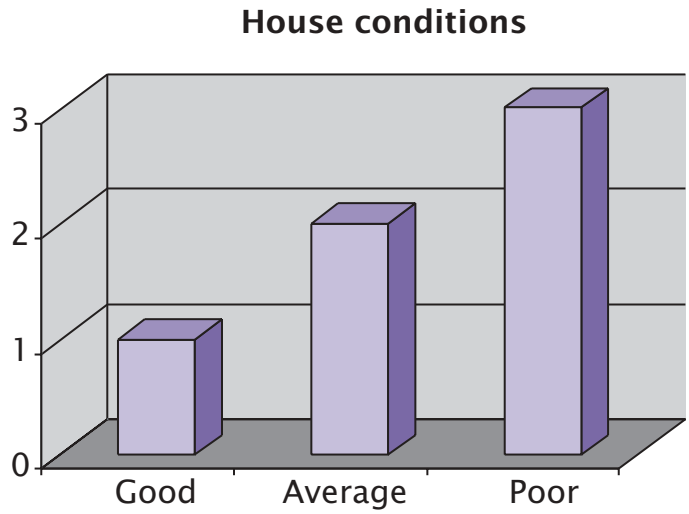
1. ensure that the picture accurately represents the data;

2. make the diagram simple enough so that the reader can quickly assimilate the message; and

3. use colour to assist with the message presentation.

This paper will restrict the scenario to the two dimensional case – that is for those diagrams that are to be displayed on paper or on a screen. The reason a three-dimensional chart does not convey the message well in two dimensions, is that the third dimension cannot be fully appreciated on the two dimensional medium. The following two examples illustrate this principle.

a. House conditions – three dimensional block chart

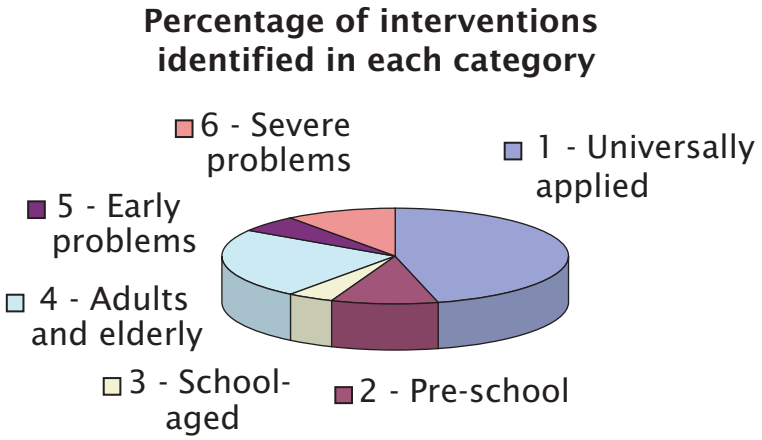


Now the raw data here, is simply
Good = 1 Average = 2 Poor = 3.

The first impression of the reader is that all three bars appear to be below their scale lines – this is the effect of the third dimension. Next notice that, whilst the area of the front and side faces of the blocks seem to be in the proportion 1, 2 and 3, all have the same top surface – so the first block will appear relatively bigger. The data is represented accurately only in the front face of the blocks – hence that is what should be shown.

b. Interventions in various client categories

Presentation of pie-charts like the one below now seem very common. But what message do they give? And is it the correct one?



The purpose of the person drawing this (it was actually published) was to show the relativities of the various categories. So we would expect the relative sizes to be in the right ratios. Even if the third dimension could be ignored, the proportions are not correct: however, the impression given to the eye includes the third dimension – and those parts of the chart with the benefit of the third dimension have a greater impact. To illustrate the impact, let's consider the relative visible areas (what it looks like) and the relative percentages the diagram is trying to represent (what it should be).

Category	Apparent percentage of visible area	Actual percentage from data
1 Universally applied	43	46
2 Pre-school	22	12
3 School- aged	6	4
4 Adults and elderly	17	19
5 Early problems	3	5
6 Severe problems	9	14

Thus the pie-chart does not represent the data well.

4.2 Charts – formatting principles

Most people will be more concerned with the colours used in charts and diagrams than the accuracy of the data representation. The difficulties of interpretation of charts lie as much with the creator of them as with the reader/ interpreter.

A good first principle, before starting to draw a diagram, is to write an aim for the diagram.

For example

'I want to show the relationship between the increasing money spent on a service and the number of assessments undertaken';

or,

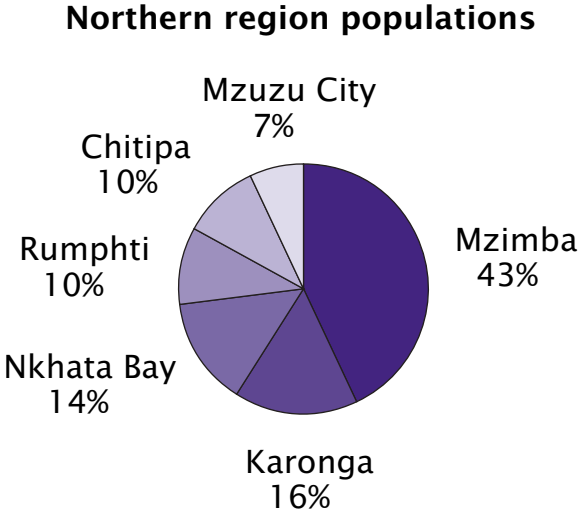
'Show the relative amounts of money spent on various services in a year'.

Given a clear aim, it is then easy to draw a diagram and check whether the outcome meets the aim. It is very easy to give the wrong message from data in diagrams and some principles are discussed on the following pages which, essentially, are standard practice. Each diagram should be judged against them. Remember that most readers try to understand the message of any diagram without reading the scale or other associated information.

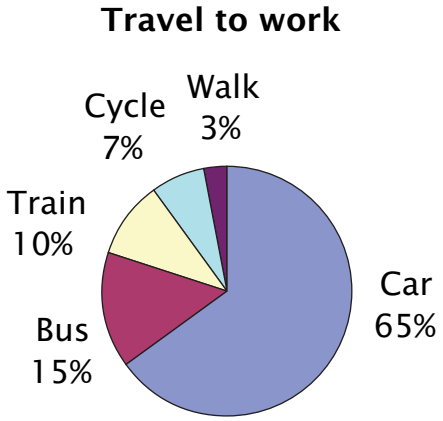
Good diagrams clarify the data and are easy to read. The message is obvious and thus the diagram does not need a lot of study. If it is not the case, the reader will, more than likely, not try to understand.

4.3 Pie charts – formatting principles

Pie charts are used to display the proportion of a total in relative terms. The items can be related, such as people in different areas, or unrelated, such as the modes of travel to work. If the items are related, the pie chart should be in shades of one colour as illustrated below.



If the items are unrelated, the pie chart could be in different colours. The example below shows this type of pie chart. Note that in both cases the start of the chart is at '12 o'clock', to help the reader to have a reference point for the interpretation of the largest segment. Also the segments are ordered by size instantly helping the user to gain information from the diagram.



Always label the chart effectively so that no other reference is needed. Resist the temptation to have a segment for all the possible splits of the data. It is therefore beneficial to arrange the smaller elements into groups. Around six segments are optimum; if more than six appear to be required, the data would probably be better displayed in a bar chart.

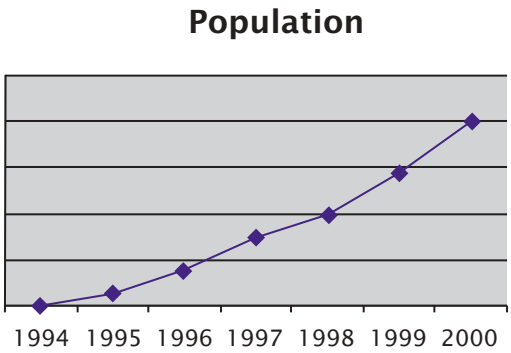
4.4 Graphs – formatting principles

Graphs are normally used for displaying the movement in one or more data items over time, such as the population in the county or number of admissions at the leisure centre.

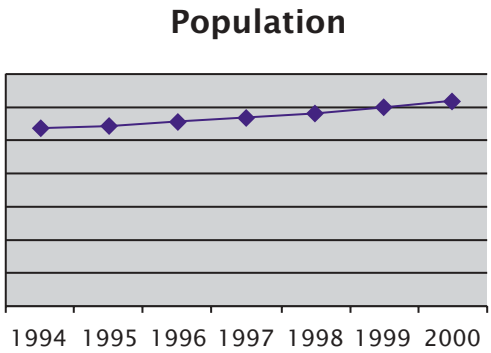
Often one can confuse graphs in two ways. First, the scale can be shown incorrectly so that the wrong message is taken; secondly, two or more data items with different scales can be shown on the same graph – in principle to show the relationship between the two data items over time.

Consider the adjacent two graphs

Graph A

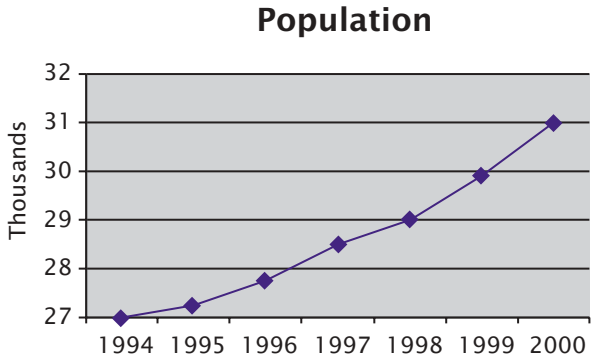


Graph B



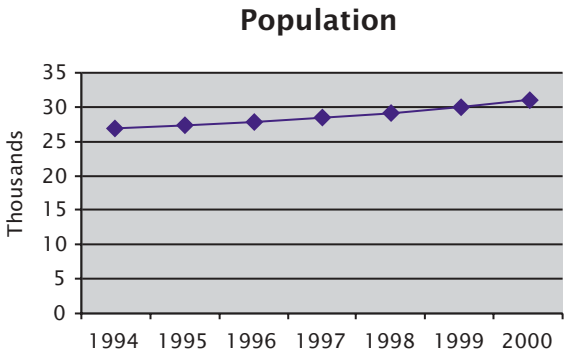
The message to the reader is quite different in Graphs A and B; in A, the population is rising steeply whereas in B the population starts at a high level and increases slowly. Which is right? Yes, the underlying population data is the same in both graphs: so, what is the difference? In Graph A the scale has been truncated – which exaggerates the rise in population. Even if we added the scales, as in Graph C, the visual impression is till of a steeply rising population – which is not true.

Graph C



As Graph D is the proper way to present the data, we may be tempted, having considered the aim of the chart, to rethink its inclusion in the report.

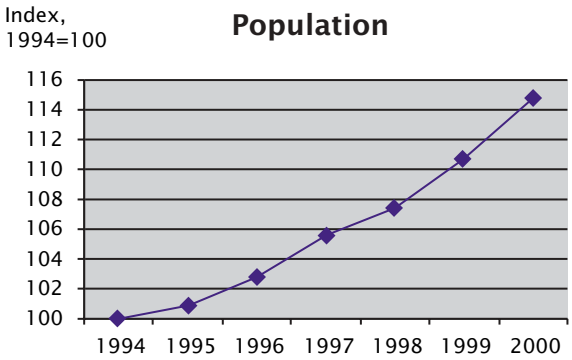
Graph D



One possible solution for this example would be to convert the data into an index, as in Graph E. To do this the first year of the series is taken to represent 100%, following years' values are then calculated as a percentage of this. This would allow the comparison of the change in the population across the years.

This indexed graph is also useful for comparing relative changes in two series where the values are quite different, for example the populations of Wales and England.

Graph E

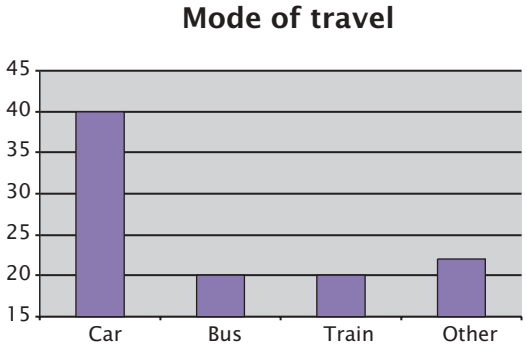


In Graph E, the presentation is of the relative values not the absolute values. The reader thus has to consider the movement of the line not the distance from any starting point.

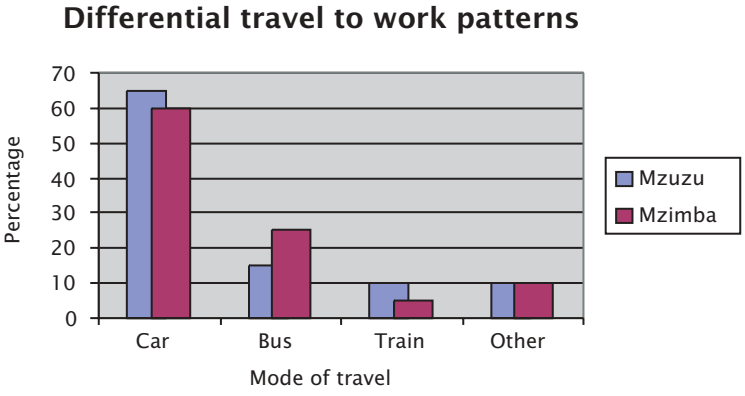
4.5 Bar charts (or block charts) – formatting principles

The purpose of a bar chart is to express the relationship between the data values in terms of the visual area of the bars. As with graphs, it is very important that the bars are correctly drawn with the scale shown – including the zero point. Another problem in presenting bar charts is the use of overlapping bars – hiding some of the area – which gives the wrong relative impression. These two errors are illustrated in Bar charts A and B.

Bar chart A



Bar chart B

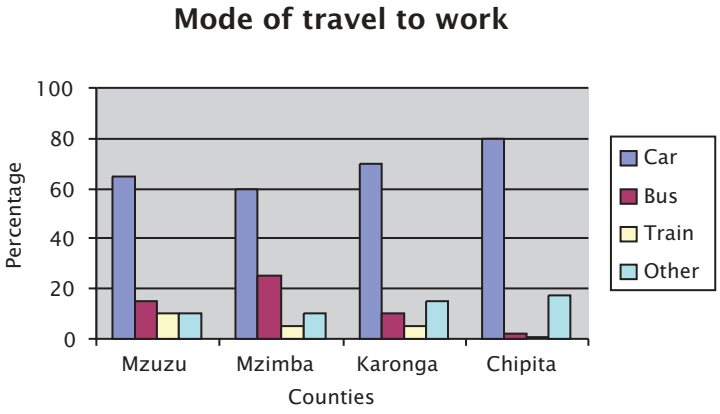


In Bar chart A, the first bar represents '40' and the second bar '20'- that is the first bar is meant to be a factor of two larger. However, the apparent difference is a that the first bar is five times larger than the second. In Bar chart B, the boldness of the colour for Mzimba and the overlapping of the bars gives the impression that the Mzimba bar is much larger than Mzuzu.

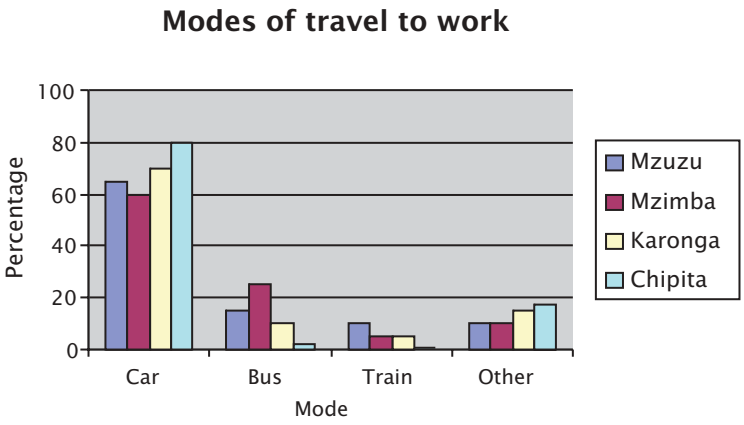
Thus for simple bar charts, **always show full scale** and **don't overlap the bars.**

When more than one data set is shown, it is important to think through the full reasoning of the aim of the chart. In Bar chart C, the question has to be asked whether the prime comparison is within the counties or within the modes of travel to work. If the prime comparison is within the modes of travel to work, then the bar chart should be as in Bar chart D.

Bar chart C



Bar chart D



4.6 Pictograms – formatting principles

In order to present statistics to non-statisticians the pictograms were invented so that data could be easily related to the subject. This could be cars, people or, as in the case below, money.

Charts are usually prepared similar to bar charts with the pictures replacing the bars. One other type of use is in the quick picture. The idea, is to show the growth in a particular item, but instead of replication of the picture as in Pictogram A, the second picture in Pictogram B is double height and double width (which is the equivalent of four times the area). So instead of the ratio of the pictures being 1:2, it is 1:4, and hence incorrect.

Pictogram A

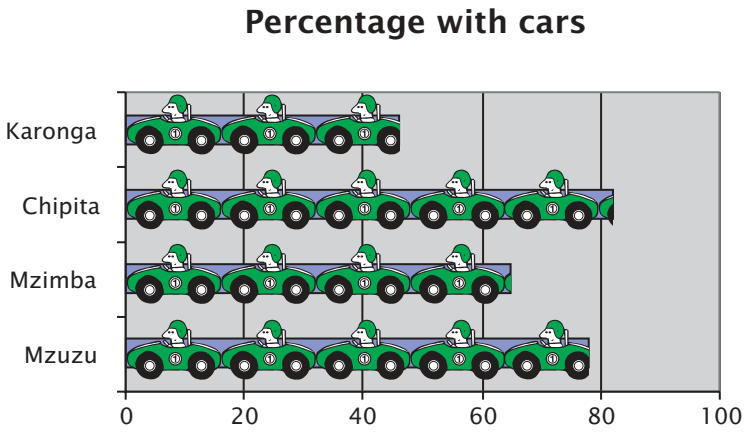
$1 million $2 millions

Pictogram B

$1 million $2 millions

The more common form of pictogram is shown in Pictogram C. Here the bar is replaced with a line of pictures, with the last one being cut appropriately. Such a diagram can help to get the message across to specific interest groups, particularly children. Depending on which picture is chosen, the actual values may be difficult to interpret with such pictures, although this can be assisted by using vertical grid lines in the chart.

Pictogram C



Principles for diagrams (charts and graphs):

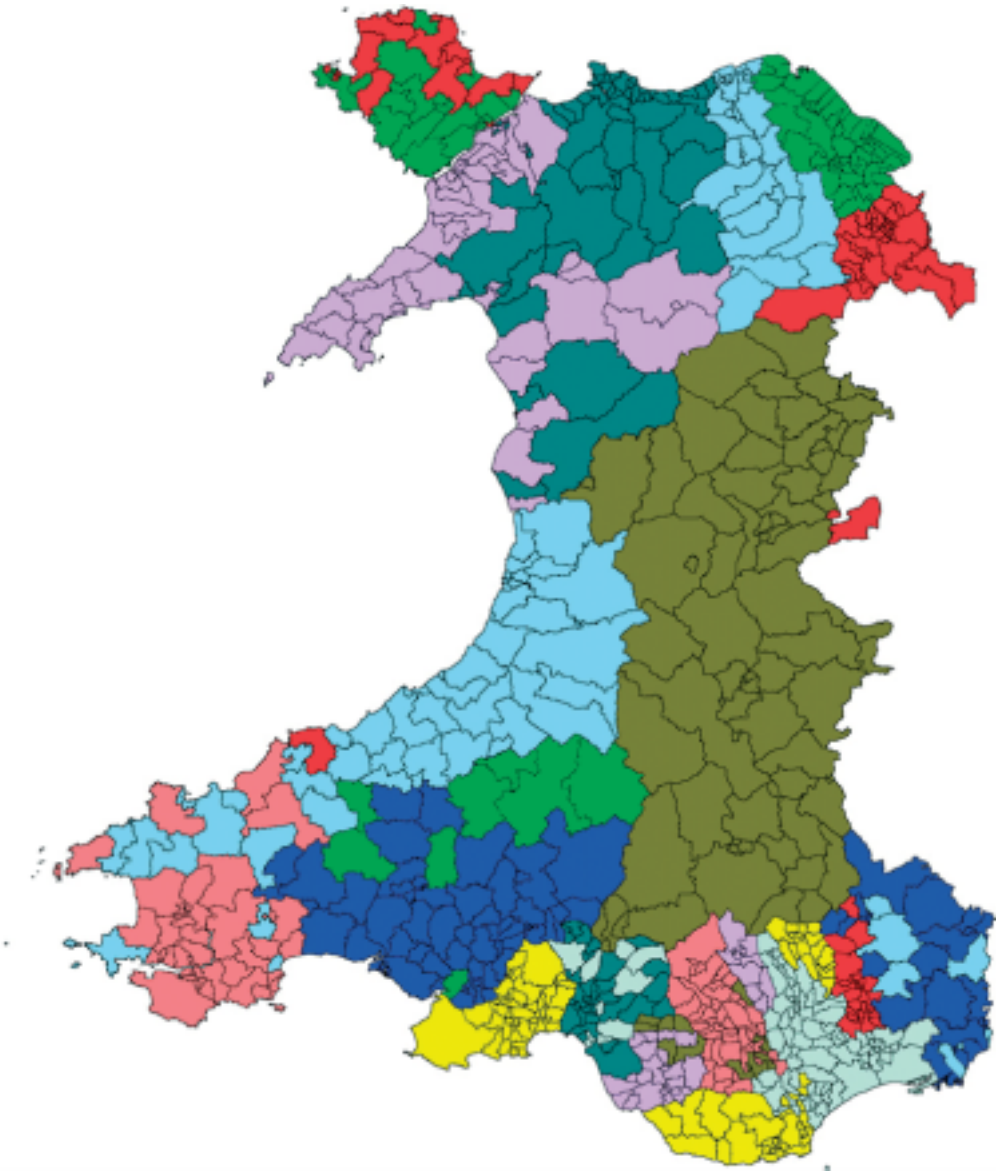
1. Never use three dimensions;
2. Do not overlap bars in bar charts (this distorts the visual comparison);
3. Use colour sparingly: shades of a single colour are often more effective; bold colours may draw attention away from the main purpose;
4. Always put the zero point on the scales when graphing absolute numbers or preparing standard bar charts;
5. Order the segments of pie charts, largest to smallest – and start the first slice at 12 o'clock or 3 o'clock;
6. Always label the diagrams effectively;
7. Always show gridlines to aid interpretation;
8. Always use appropriate rounding of numbers on the axes;
9. When using pictograms, remember these are usually one dimension in measurement, that is the number is represented by one or more pictures of the same size.

5. Using maps

5.1 Basic map principles

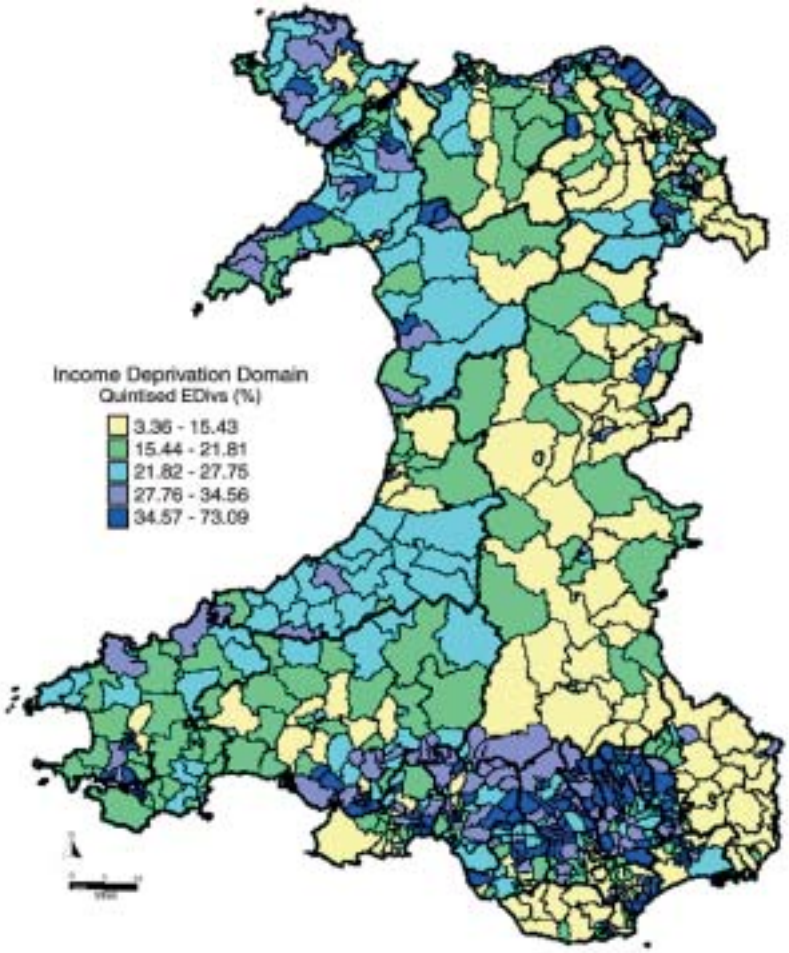
Maps can be a rich means of portraying information that would be very impenetrable and indigestible if put into a table. A temptation exists to put as many different colours into the same map as possible, and that loses one of the main means of communicating the information. Map 1 below, illustrates this.

Map 1



A slightly better example comes from the work on the Welsh Index of Multiple Deprivation. Here information on almost a thousand data items can be portrayed and understood quite quickly. Map 2 below, however, does not follow some of the basic principles.

Map 2 - Welsh Index of Deprivation 2000
Income Deprivation Domain



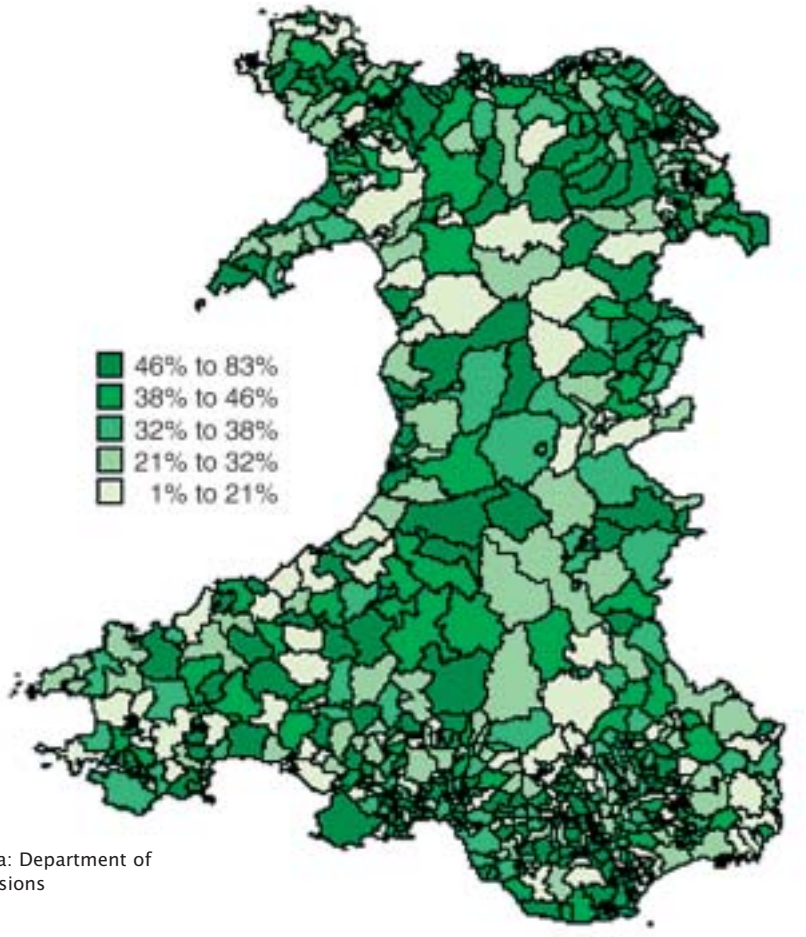
The first point to note is that most people will try to interpret a diagram or map without reading the notes and the scale. Applying this principle to the Map 2, it is quite clear that there is something of a scale with those areas coloured blue and a further assumption that the deeper the colour the larger the 'value'. As the non-blue areas are coloured in two separate colours, the reader will assume that the data is unconnected with the blue area data. Closer examination of the scale reveals that there is a range, with yellow indicating the lowest value, then green, then the palest blue through to the darkest blue. This would have been far more useful to the reader if the whole map had been in shades of the same colour.

Note that if the 'lowest' value is coloured white, the immediate impression given, is that no data exists for that area – even if shown on the scale.

The only really valid use of two colours in statistical presentation, is where the data is distinctly in two parts of a scale, for example, with migration, where inward migration can be coloured in shades of green and outward migration could be in shades of red.

Note, however, that even these two colours could imply meaning: green being positive and red being negative. Some may then choose to use red for inward migration. This makes it easier to understand the scale of the variable across the map – without continually referring to the scale.

Map 3 – Proportion of Children (Under the Age of 16) Living in Families Claiming Means Tested Benefits



Source of Data: Department of Work and Pensions

Map 3 shows how a properly constructed map can assist with the interpretation – and save a thousand words!

Principles for maps:

1. When using maps to display statistics, generally only use shades of a single colour to denote the strength of the values: deeper colours for higher values. The exception would be where the positive and negative values are to be shown on the same map: here use one colour for the positive values and another for negative values; the deeper shades should be at the extremes of the scale;

2. Don't use white for any of the areas – except to denote missing data;

3. Use a maximum of five shades of a colour;

4. Some colours have connotations; e.g. blue - water, green - forestry; and

5. As with diagrams – always write an aim and check whether the resulting map meets the aim.

6. Further reading

Plain Figures, Chapman, M & Wykes, C, The Stationery Office, London.
ISBN 0 11 702039 7

A Primer in Data Reduction, A S C Ehrenberg, John Wiley & Sons.
ISBN 0 471 10135 4,

Annex 1: Standard table layout

Table number: Table title (a)

Table descriptor (e.g. Millions)

Row descriptor	Column heading 1	Column heading 2	Column heading 3
Row heading 1			
Row heading 2 (b)			
Row heading 3			
Row heading 4			
Row heading 5			
Row heading 6			
Total			

Source: Local Government Data Unit

(a) Footnote 1

(b) Footnote 2

Notes:

1. The column headings are centred;

2. Use lower case letters in parentheses to denote a footnote, e.g. (a), (b) etc;

3. Only one footnote to appear against the title – merge if necessary. The footnotes are then sequential across the column headings (left to right) if required and then down the row headings;

4. Footnotes should not appear against individual figures: add to most appropriate row or column heading;

5. The table descriptor is shown only if all of the data items in the table are in the same units: otherwise add a descriptor, in parentheses, to each column or row heading as appropriate; and

6. When giving a breakdown of a total within the rows of a table, use indents to show the relationship to the heading, the sub-heading having a colon following. For example:-

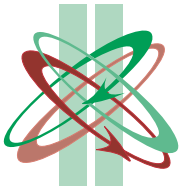
Type of vehicle passing inspection:

Car
Van
Truck
Articulated vehicle
Total

Annex 2:

Standard Ordering List for Welsh Local Authorities

Ynys Môn	Isle of Anglesey
Gwynedd	Gwynedd
Conwy	Conwy
Sir Ddinbych	Denbighshire
Sir y Fflint	Flintshire
Wrecsam	Wrexham
Powys	Powys
Ceredigion	Ceredigion
Sir Benfro	Pembrokeshire
Sir Gaerfyrddin	Carmarthenshire
Abertawe	Swansea
Castell-nedd Port Talbot	Neath Port Talbot
Pen-y-Bont ar Ogwr	Bridgend
Bro Morgannwg	The Vale of Glamorgan
Caerdydd	Cardiff
Rhondda Cynon Taf	Rhondda Cynon Taf
Merthyr Tudful	Merthyr Tydfil
Caerffili	Caerphilly
Blaenau Gwent	Blaenau Gwent
Torfaen	Torfaen
Sir Fynwy	Monmouthshire
Casnewydd	Newport
Cymru	Wales



UNED DDATA LLYWODRAETH LEOL ~ CYMRU
LOCAL GOVERNMENT DATA UNIT ~ WALES

Local Government Data Unit – Wales
8 Columbus Walk, Cardiff, CF10 4BY
Telephone 029 2090 9500
Email enquiries@dataunitwales.gov.uk
Web www.dataunitwales.gov.uk