Making Data User-Friendly

Andrew Ehrenberg

Report 21 for Corporate Members (April 1998)

Most of the data we are asked to look at are hard to understand. The data would communicate far better if people used five simple guidelines.

To illustrate this, Table I gives data for eight brands of gasoline (with theoretical "Dirichlet" norms): market-shares, penetrations and purchase rates.

But what does Table 1 say? What will you remember when you turn the page? What patterns can you see even when you keep on looking at the table? With these questions in mind, Table 1 makes one feel non-numerate. It seems just a jumble of undigested data. Whoever produced the table either did not know what his data were saying, or was not letting on.

In contrast, Table 2 lets us compare each market-share with the single base-line average. We can see that the shares of the top four brands are high (Esso to Jet) and the bottom four are low (Gulf

to Elf). And that it's the same high to low pattern in each column.

Faced with Table 2 we can all cope: the patterns and deviations are "interocular" – they hit us between the eyes – especially once we have been told what the patterns are. Yet Table 1 is still not at all clear, even though we now know what it *should* be telling us.

FIVE SIMPLE RULES
Five simple guidelines worked wonders in thus turning *data* into *information*:
1. Order the rows and/or columns by some measure of size.
2. Round to two effective digits.
3. Give averages as a visual focus.
4. Use table layout to guide the eye.
5. Give a brief verbal summary.

The rules are based on enabling our fragile short-term memory to cope better. In reading a table of numbers

we need to remember them briefly. To see that Esso's penetration at 35% was higher than Elf's at 6%, we need to remember the 35% for an instant. And to note that this drop from 35% to 6% is less steep than that in the market-shares from 18 to 2, we need to keep all *four* numbers in mind, at least briefly.

Mental arithmetic therefore requires short-term memory. Yet it is well established that this is limited to only about seven bits or "chunks" of information. So as we go on reading, we will forget all or almost all the earlier numbers (unless we look back), and especially the onerous three-digit and unstructured ones in Table I. Hence the fewer demands we make on our short-term memory, the better. The five rules apply this notion to the formatting of numerical tables.

RULE 1 : Order by Size
The data in both tables are the same. Yet to find any patterns in the columns in Table 1 requires us to keep its higgledy-piggledy numbers in our mind (which number is the highest in each column, the second highest, the fifth, etc.).

In contrast, the main pattern in Table 2 is easy to see and to remember: simply that the figures in each column go *down* (with an exceptional "high" of 4.8 for Gulf in column 4). And if any column did *not* follow this pattern, that would also be easy to see and remember.

RULE 2: Rounding to Two Effective Digits
Numbers rounded to two effective digits as in Table 2 are much easier to read

Table 1: Gasoline: Quarterly Penetrations and Average Purchase Rates
(8 typical brands: UK 1990: Observed O and Theoretical **T**)

12 weeks	Market Share %	Penetration, %		Average Purchase Rate	
		O	**T**	O	**T**
BP	9.74	27.3	**26.8**	4.14	**4.23**
Elf	2.19	7.4	**6.0**	3.34	**3.79**
Esso	17.58	35.3	**39.8**	5.28	**4.64**
Fina	3.87	12.0	**10.4**	3.48	**3.93**
Gulf	6.33	14.3	**15.8**	4.79	**4.04**
Jet	8.60	23.1	**21.9**	3.93	**4.11**
Shell	12.98	32.2	**31.5**	4.34	**4.28**
Total	3.27	8.7	**9.3**	3.61	**3.90**

Table 2: Penetrations and Purchase Rates
(As in Table 1, but in market-share order; rounded; with averages; layout to guide the eye)

UK 12 weeks 1990	Market Share %	Penetration, %		Average Purchase Rate	
		O	**T**	O	**T**
Esso	18	35	**40**	5.3	**4.6**
Shell	13	32	**31**	4.3	**4.3**
BP	10	27	**26**	4.1	**4.2**
Jet	9	23	**22**	4.0	**4.1**
Gulf	6	14	**16**	(4.8)	**4.0**
Fina	4	12	**10**	3.5	**3.9**
Total	3	7	**8**	3.6	**3.9**
Elf	2	6	**5**	3.3	**3.8**
Average	8	20	**20**	4.1	**4.1**

and to remember than the longer ones in Table I. They are also easier to manipulate mentally. Effective digits are ones that vary in that part of data.

RULE 3: Averages
Row and/or column averages can help to provide a visual/mental focus: In Table 1 we have to compare all the eight 3-digit market-shares with each other (9.74 with 2.19; 2.19 with 17.58; 17.58 back with the 9.74; and so on).

In Table 2 we can compare each 1- or 2- digit market-share with a single base-line (the average 8) to see and remember which figures are high (Esso to Jet) or low (Gulf to Elf). The memory-strain is vastly less.

We can also easily see that the observed and theoretical figures tend to agree (especially on average). And that the purchase rates are all close to their averages (4.1), unlike the market-shares and penetrations. And we can remem-

ber all this at least broadly – the strain on our memory is again vastly less.

RULE 4: Layout to Guide the Eye
In Table 1 the figures were spread apart, as is common practice. But this requires us to move our eyes. Eye-movement interrupts our fragile short-term memory. So when we look at the figures on the right, we will have forgotten the brand names on the left. When we look back at these names, we may forget the figures on the right. Table 2 requires much less eye-movement.

Rows in single spacing with some wider gaps, selective horizontal and vertical lines, and some bold lettering also help us to take in the data. and to see and recall them when we glance back.

RULE 5: A Verbal Storyline
Articulating explicitly in words what the figures in a table are saying helps the reader greatly, e.g. that in Table 2 "All the figures go down'. This gives our memory a "coat-hanger" for the more detailed data, away in: We can them see and remember the patterns much better, and spot the exception (Gulfs high purchase rate of 4.8 again).

More on Rounding
People sometimes worry that in rounding, accuracy is lost because data are actually discarded! But the fuller data are always available on the computer. In any case, even rounded data are more precise than what can usually be read off a graph or predicted from a mathematical model. Precision beyond two

effective digits has rarely been shown to matter.

Drastic rounding is needed because of our short-term memory limitations again. As psychologists have shown? none of us can recall or manipulate long numbers if we are interrupted in any way, even just by our own thoughts.

If we look up a telephone number, such as 723 5473, we can usually dial it. But if we first say to ourselves "What am I going to say to the guy?' we have to look up the number again.

However, we CAN retain numbers in our immediate memory *if they are only two digits long*. This works even when we are interrupted. We cannot easily subtract 17.9% from 35.2%, let alone mentally divide one into the other. But rounded to two digits, 35% is obviously about twice 18%.

With a table of two-digit numbers we can therefore see how *this* ratio is about 2, that one about 3, *that* one 2 again, and so on. We can "read" the table.

There can be no compromise over the rounding rule. Thus saying:

Sales of Bango rose from 29,942 to 94,347, and those of Bingo from 13,492 to 36,051

tells us little as it stands. To be able to see that sales had *trebled* we would have had to round mentally. This could have been done for us:

Sales of Bango rose from 29,000 to 94,000 and those of Bingo from 13,000 to 36,000.

Now we can readily see that Bango increased just *over* three-fold, and Bingo just *under* three-fold.

But we still cannot see this at all easily even with just *three* digits – e.g. 28,900 and 94,300 – without rounding further mentally. With long numbers we are non-numerate. With two-digit ones, we can cope.

Implementation
If as professional people we mostly appear innumerate, the fault is not in ourselves, but in our data. Or with the *producers* of our data. Most tables could easily be made to communicate better. We are to blame only if we opt out of the challenge of bringing this about.

At its simplest, we must learn to return opaque data to their makers rather than do their work for them. Thus if we *insist* on having brief verbal summaries (Rule 5), the analysts themselves may actually learn what (if anything) their data mean before passing them on.

The five rules seem very simple – mere common sense – at least once they have been stated. But they cannot be implemented overnight. This requires effort: there will be set-up costs and upset costs.

Management will need to allocate people and budgets to implement change. It is not enough just to say "Avoid long numbers" or "Use good layout". We need to encourage, train, and *monitor* people (including ourselves) – and then start over again, encouraging, (by setting a good example), training, and monitoring.

We need also to understand *why* the rules work (e.g. the role of short-term memory), so as to use them effectively. One recent writer stressed that the loss in accuracy in rounding Bingo's sales of 13,492 down to 13,000 was only about 3.78%! Such failures to implement the rules occur when the reasons behind them are not understood, e.g. that the rounding error expressed as 4% would be easier to remember, to use, and to relate to other numbers.

Conclusions
Producers of numerical data have been production-oriented, showing little concern for their customers. Most do not realize how unnecessarily incomprehensible their tables usually are.

People are not familiar with the notion that most tables can easily be improved. They think that how data are to be presented depends on one's purpose. But we argue that a table which follows the Five Rules communicates better for *any* purpose.

So we need to remember to:
1. Order by size.
2. Round.
3. Use averages.
4. Use layout to guide the eye.
5. Give a verbal storyline.

Technical Notes
To rehearse some of the main technicalities which can arise in using these rules, we go over the five concepts again in turn:

RULE 1: Ordering by Size
With the rows of a table ordered by some measure of size like the market

shares in Table 2, one can literally see, as on a graph, whether the other columns follow the same pattern. Deviations also stand out.

- To order the rows, one can use any measure of size, e.g. the population (for regional data, say), the row averages, the first column in the table, market shares, etc. – although some may be more relevant than others.
- Use the same order for similar tables.
- Exceptions to Rule 1 are rare (e.g. some time-series and their yearly movements).

RULE 2: Rounding to Two Effective Digits
Using only two effective digits means digits which vary in that set of data. Index numbers like 137, 146, 172, etc., only have two effective digits: 37, 46, 72. The 1's do not vary.

- Exceptions arise in our decimal system for numbers all near 100 say, e.g. 104 and 92. The 104 has only two effective digits (i.e. rounding 104 to 100 would be over-rounding compared with 92).

This is one of the few instances where some judgement is needed. (If the decision seems "difficult", it can hardly matter. Therefore choose what makes it easy fur the reader.)

- Deliberate over-rounding to just one digit can at times be effective (e.g. a correlation coefficient as r = .6, not .63).
- The 0 before a decimal point can give too much visual clutter (not 0.6, but .6).

- Use variable rounding if the readings vary greatly in size. In the market-share column of Table 2 we first rounded to two digits (18, 13, 10), then to one (9,6,4. etc.). With accounting data we should similarly write:

	$		$
Sales	112,000,000	not	112,469,000
Profits	4,800,000	not	4,782,000
Welfare	5,300	not	5,312

- With longer numbers like 17.94% and 35.28% the rounding (to 18 and 35) should not be left for everyone to do mentally. This would not only be inefficient, but there also are deeper reasons. Thus when we look back at the numbers (as one often has to do when working with data), the 18 in our mind would not match the unrounded 17.94 on paper or on the screen. We would have to round that again mentally, to 18. What is more, if it was 18 also on paper, other numbers (like the 35) would come flashing back in our mind without having to look at them (or hardly). That is how our memory functions. We should make the most of it by having the rounding already done on paper or the screen.

RULE 3: Averages

In almost any table, row and column averages should be computed routinely. (They can then be deleted if unhelpful).

- An average need not be "typical". It lets us first see and then summarize the scatter of the readings.
- In reading a table, begin by looking at the averages. In Table 2 we can first see that the Observed 0 and Theoretical **T** agree on average. With that in mind, we can then scan the individual pairs of 0s and **Ts** to see whether they also agree, and spot the exceptions.
- We can also see how the figures in each column vary about their averages – e.g. greatly for the penetrations in Table 2, much less, for purchase rates.

RULE 4: Layout

Making the tables easy to read (or glance at) helps not just those readers who are new to the table, but even the analyst him or herself. We all need possible guidance to see what there is to see.

- Production factors matter little these days (i.e. getting a table actually set up). Yet many publications do nut use vertical rules – which used to be difficult to set – although these can help greatly to guide the eye.

The finer details of table-layout may involve personal judgement. But general factors are:

- A smaller typeface than for the written text reduces the amount of eye-movement and makes it easier to take in the different parts of the table.
- All columns need not be equally spaced (e.g. it helps that in Table 2 the 0 & **T** were closer).
- Selective vertical rules and/or white space also help to group the data visually, as in Table 2.

- Rows of a table should be in single spacing to guide the eye down each column.
- But a half-line gap every fifth row or so, as in Table 2, makes it much easier to read across a biggish table (e.g. that all the figures for BP are two rows above the gap).
- Tables which matter should ALWAYS be shown in the text where they belong (not at the end of the report).
- Numbers are easier to read downwards in a column than across the page in a row, even (or especially) in the "bad" Table I.
  In looking at its first column we can easily compare the first digits or so (9,2, 17.3, 6, etc.) and screen out the rest:

9.74   27.3   26.8   4.14   4.23

2.19

17.58

3.87

6.33

If we also want to read across (as in comparing the pairs of 0 and **T** figures in Table 2), it is much easier if the downward pattern is already explicit {unlike in Table I).

RULE 5: A Verbal Storyline

The table caption, and/or the accompanying text or voice-over, should briefly articulate what the table is saying. Not the whole story necessarily, but to give readers a "way in" – a feature to remember when looking at the detail (e.g. in Table 2: "All the figures go down").

Key References

Ehrenberg, A.S.C. (1982, reprinted 1994), *A Primer in Data Reduction* (Ch. 15 & 16), New York and Chichester: John Wiley.

Making Numbers Work. London: Melrose Learning Resources – fax 01 144 171 622 0421. (A 24-minute training video).

www.MarketingScience.info