

## Case Study: figures, tables, and graphs in Averkiou and Dodgson (2011)

### Source Material

Averkiou, M & Dodgson, NA, "Comparison of relative (mouse-like) and absolute (table-like) interaction with a large stereoscopic work-space", *Stereoscopic Displays & Applications XXII*, Woods, Holliman, Dodgson (eds.), Proc. SPIE 7863, paper 7863-41, 2011

<http://www.cl.cam.ac.uk/~nad10/pubs/SPIE11.pdf> (3MB PDF of the paper)

<http://www.youtube.com/watch?v=ZkI-DW2GpaY> (video of Averkiou's presentation at the conference)

<http://dx.doi.org/10.1117/12.872002> (publisher's website)

### Analysis

**Figure 1** shows three photographs of the equipment. One could argue that these photographs are redundant because the text contains descriptions of the equipment. However, humans are visual creatures and these photographs are included to help readers to visualise the experimental equipment.

**Figure 2** shows two photographs of the experimental tasks. Like the photographs in Figure 1, these help readers to understand and visualise the experimental setup. These are more important than those in Figure 1 because it is essential that readers understand what was meant by "peg-board task" (Figure 2(a)) and get a feeling for the size of the screen and the relative locations of screen space and manipulation space (Figure 2(b)).

**Table 1** and **Table 4** lay out ANOVA results in the standard format. Considerable effort was put into the design of these tables to make them easy to read.

*Rules:* Only two horizontal rules (lines) were used and no vertical rules were used at all. This means that almost all of the ink, and almost all the reader's attention, is focussed on the data.

*Italics:* Italics are used to distinguish the column headings from the body of the table. Italics are perhaps unnecessary, as the rule already divides the headings from the body, but they were chosen for their aesthetic value. That is, to make the table look nicer.

*Alignment:* The columns of real numbers are carefully aligned on the decimal point. Notice how important this is in the *MS* column of Table 1, where the numbers range over two orders of magnitude, between 58 and 2075. With the alignment on the decimal point, it is easy for the reader to compare the magnitude of the numbers in this column. The *df* column is centre-aligned rather than right-aligned. Right alignment would have lined up the digits correctly, but the authors made an aesthetic decision that the table looks more attractive with the numbers centred in the column. The reader's ability to understand the table should over-ride aesthetic considerations, but aesthetics can be used to choose between two options that are equally readable. In this case readability was not affected.

*Precision:* All numbers are given to three decimal digits in the *SS*, *MS*, and *F* columns. This leads to a little over-precision in the *SS* column in Table 1, where there are seven significant digits. The column in which the reader is most interested is, of course, the *p value*. This should be given to three significant digits. Notice that two of the values are shown to only two significant digits; it is not clear why the authors made this minor slip. The *p value* column has an interesting problem of its own: the values differ by 15 orders of magnitude in

Table 1 and four in Table 4. It is not sensible to align these vast differences on the decimal point, so the authors have opted for scientific notation ( $n.nn \times 10^{-n}$ ).

**Table 2** shows results of two-tailed *t*-tests. As above, the *p* values have such a wide range of values that scientific notation is the only sensible alternative. In this case every value is correctly given three significant figures and all of the numbers are aligned on the decimal point. The *Configuration Pair* column took several iterations to get to the format you see. The authors wanted to make it obvious what is being compared with what in each case. They chose to do this in two ways: the numbers of the different conditions followed by a description of the two different conditions. The numbers alone would have required the reader to refer back to a key, so it was thought useful to include an abbreviated key in the Table itself. Despite this effort at making it comprehensible, the complexity of what is being done means that this Table cannot be understood without reading the main text. In situations like this, you do your best to help the reader to understand, but do not replicate vast swathes of text in the caption or table.

**Table 3** is straightforward and neatly laid out. Notice that all of the numbers are carefully aligned on the decimal point and that each is quoted to three decimal places. The text in the *Configuration* column provides a key to understanding Table 2. It is a pity that the two tables are not on the same page, because that would help readers to understand Table 2.

**Figure 3** contains two box-and-whisker plots showing the five quartiles. This was produced in *Excel*, which means that it is limited by what *Excel* is capable of doing. In particular, it required a good understanding of *Excel*'s graphing options to get the medians to display on the same graph as the four other quartiles (which are themselves displayed using one of the stock chart options!). The plots are as attractive as those produced by *R* and, more importantly, do not show outliers, which *R* would show. Outliers are those data points that lie more than 1.5 box heights beyond the first or third quartiles. There are certainly outliers in the right hand graph, and it would help to have shown these.

**Figure 4** contains four scatter graphs produced in *Excel*. The important formatting that has been done here is to ensure that all four graphs have identical axis ranges. This allows easy visual comparison. The *Excel* default is to scale the axis ranges to fit the data in each graph individually, which is not what is desired here. Observation and manual adjustment was needed to get all four graphs to have the same ranges. There is no chart-junk in these graphs. In particular, the authors have removed the horizontal lines that *Excel* puts into all graphs by default. Notice also the trend lines: these are almost horizontal and are included to demonstrate that there is almost no relationship between time and accuracy (see the main text for explanation).

**Figure 5:** these bar charts are OK but would almost certainly work better as tables. Compare the bar charts with the following table. Which uses the least ink to get across the message? From which can you more easily read the important message?

	1	2	3	4	5
How easy was it to <b>learn</b> the <b>absolute</b> positioning technique?	1	1	2	6	
How easy was it to <b>learn</b> the <b>relative</b> positioning technique?	2	2	4	2	
How easy was it to <b>use</b> the <b>absolute</b> positioning technique?			1	4	5
How easy was it to <b>use</b> the <b>relative</b> positioning technique?		2	7	1	
How <b>efficient</b> do you think the <b>absolute</b> positioning technique is?	1	1	4	4	
How <b>efficient</b> do you think the <b>relative</b> positioning technique is?	2	4	3	1	