THE UNIVERSITY OF AUCKLAND

SECOND SEMESTER, 2001
Campus: City

STATISTICS

Information Visualisation

(Time allowed: THREE hours)

NOTE: Attempt all FIVE questions.
All questions are worth equal marks.
You should allot equal time for answering each question.

1. (a) Perspective gives very strong "depth cues" which tells us about the three-dimensional structure of scenes and objects. Describe in detail what perspective is, and how it can be used in statistical graphics.

   (b) There is a problem associated with the (unnecessary) use of perspective in presentation graphics. Using what you know about graphical preception, explain what the problem is.

   (c) There are a number of other effects which give us information on three-dimensional structure. Describe three such effects.

2. (a) It can be argued that colour is inherently a three-dimensional phenomenon. Explain why this is.

   (b) Describe a sensible (i.e. good) use for colour in a statistical graph.

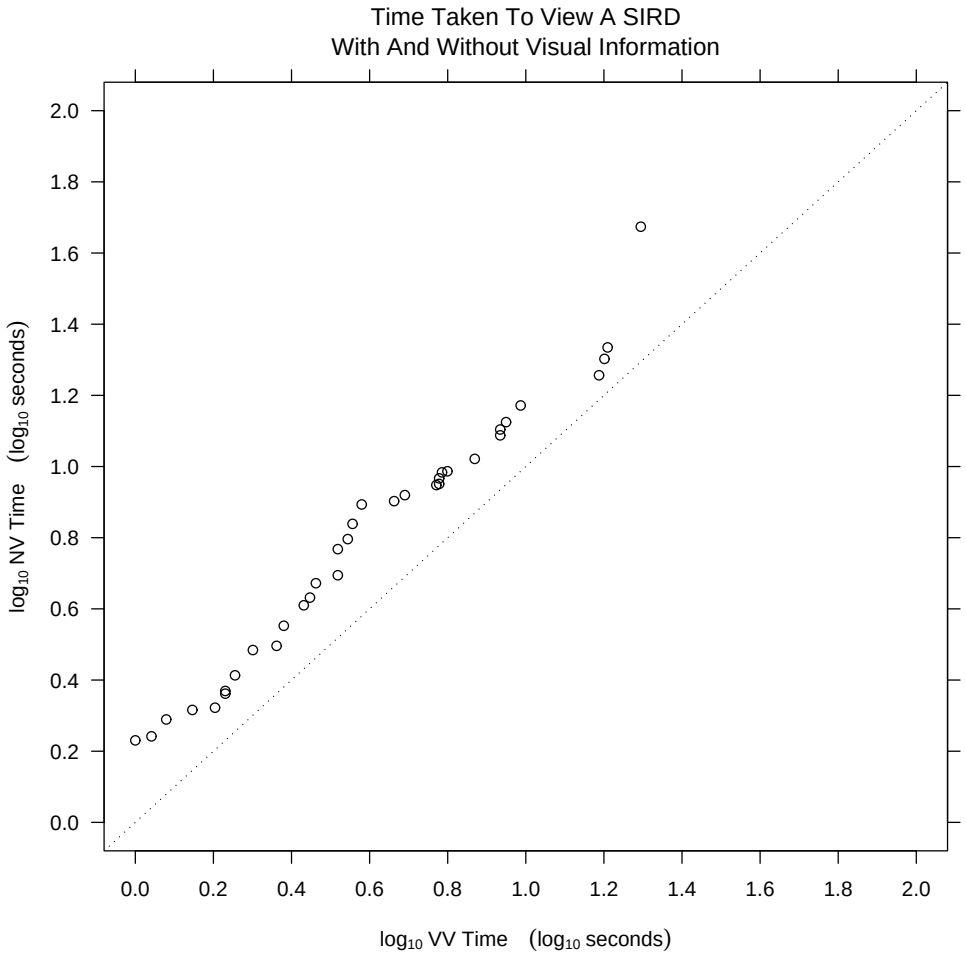
   (c) Give **three** examples of how colour could be used **badly** when drawing a graph.

   (d) Why would it be unwise to use the rainbow of hues ranging from red to blue to encode numerical values?

3. A *single image random dot stereogram* (SIRD) gives a way of hiding a three-dimensional image in an apparently random set of dots (this kind of image was popularised by the "Magic Eye" series of books). A subject views the image by defocussing her eyes and looking "through" the page containing the image.

   An experiment was conducted to see whether giving a subject visual information about the concealed image shortened the time it took for them to see it. The subjects were given varying prior information about the object; 42 receiving either no information or verbal information only (NV), and 35 receiving both verbal and visual information (VV).

   Length of time in seconds taken by subjects to see a three-dimensional object, (a spiral ramp coming out of the page) was measured and the following Q-Q plot comparing the VV and NV groups was obtained.



   (a) Describe how a Q-Q plot for comparing two batches of numerical values is constructed.

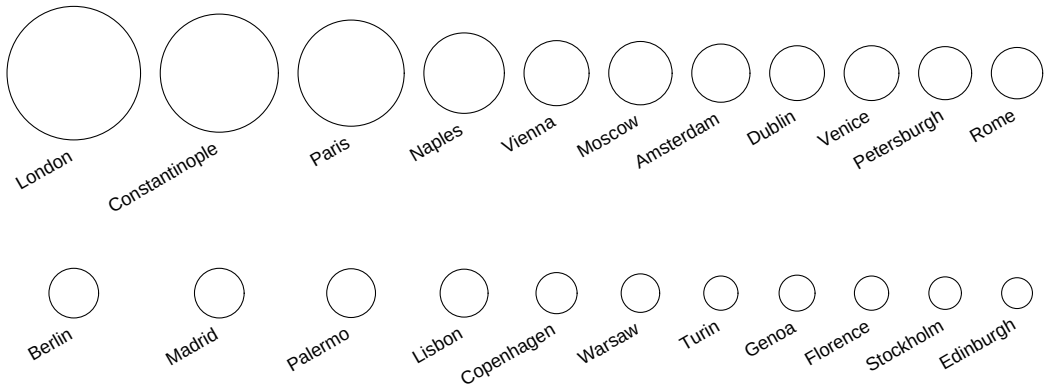
   (b) Use the Q-Q plot given above to interpret how visual information helps or hinders the process of seeing a SIRD. You may (or may not) find the following table of values of $\log_{10} x$ and x to be useful.

Table 3.1 : $\log_{10} x$ with corresponding x values.

$\log_{10} x$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
x	1.26	1.58	2.00	2.51	3.16	3.98	5.01	6.31	7.94	10.00

(c) Explain how you would use R to plot superimposed density traces of for the NV and VV batches.

4. In 1801 William Playfair published his *Statistical Breviary*, which contains the first examples of many graphical displays which we still use today. In one example, Playfair represented the populations of the major European cities of the day as the areas of circles, as shown below.



(a) Is this a good way to represent the city sizes? Justify your answer using perceptual theory.

(b) Suggest a better way of presenting the city populations. Why is your method better?

5. The following graph is redrawn from the Statistics New Zealand publication *Facts New Zealand*. It shows the number of male and female workers in nine different career categories for 1995.

(a) Explain what this graph is shows, and describe the graphical encodings which are being used in the graph.

(b) Give a short explanation of what is good about this graph and what is bad.

(c) Describe an alternative way of presenting the data in this graph which improves on at least one aspect of it. How does your alternative improve on the graph above.

(d) Give a short explanation of how you would produce the graph above using R. You can either give the necessary R statements, or explain in words how to create the graph. You can assume that the data has been stored in the variables: `male`, `female` and `labels`.