Department of Statistics

COURSE STATS 330/762

Assignment 1, 2011

Instructions: Hand in your completed assignment to the Student Resource Centre by **4pm August 4th**

The data sets for this assignment are in the files **oats.txt** (for Question 1) and **animals.txt** (for Question 2) which are available on the course web page.

Question 1.

The data in the file **oats.txt** come from a classic agricultural experiment carried out at the English research station Rothhamsted in the 1930's. The aim of the experiment was to compare the yields of three varieties of oats (Golden. Rain, Marvellous and Victory), when subjected to different amounts of nitrogen fertilizer. The experiment was carried out in six different areas or "blocks", each divided into 12 plots of ground. There were four different amounts of fertilizer used: 0, 0,2, 0.4 and 0.6 hundredweight (this is an obsolete unit of weight), so each of the twelve fertilizer/variety combinations was used in each block. The yield was measured as the total weight harvested from the individual plots. Then variables in the data set are

B: Blocks, levels I, II, III, IV, V and VI.
V: Varieties, 3 levels.
N: Nitrogen (manurial) treatment, levels 0.0cwt, 0.2cwt, 0.4cwt and 0.6cwt, showing the application in cwt/acre.
Y: Yields in 1/4lbs per sub-plot, each of area 1/80 acre.

Instructions

1. Load the data into R, and make a data frame **oats.df** to contain the data. Check for any typographical errors (the data below may be taken to be the correct data, but the data on the web may have been corrupted). Print out the last 10 lines of the data file. [5 marks]

2. What is the relationship between the amount of fertilizer applied and the yield? Does the relationship depend on the block or the variety? If so, how? Draw suitable plots to answer this question [5 marks]

3. Which variety had the highest yield overall?[5 marks]

4. Which block had the highest yield? Did some varieties do better in some blocks than others?[5 marks]

In answering these questions, do not fit any models. Graphs and the calculation of simple averages will suffice. To win the marks there must be some interpretation, not just pictures.

Question 2

In his classic book on statistical graphics (*The Elements of Graphing Data*, Wadsworth, Monterey, California (1985) , William Cleveland argues that several principles should govern the drawing of graphs. One of these principles relates to the placing of labels on scatter diagrams. The two figures overleaf illustrate this.

The data file **Animals.txt** contains a set of data very similar to that graphed in Cleveland's Figure 2.23 reproduced overleaf. There are two variables**, body** and **brain** containing the body weights (in kg) and brain weights (in grams) of 28 different animal species.

Instructions

1. Load the data into R. You may assume that there are no errors in this data set. [5 marks] Note that the data set has row labels: these are not part of the data but are useful for labeling the plot. You can read data like this by omitting the "header=TRUE" argument to read.table.

2. Draw a scatter plot of the data that resembles as much as possible Cleveland's Figure 2.22 shown overleaf. The key idea here is to avoid the bad features in Fig 2.23 and implement good labeling of the species. Pay careful attention to the labeling of the points and the axes. [15 marks]

Total for assignment: 40 marks

R hints

1. Use **read.table** to create the data frames.
2. Use regular R graphics and trellis graphics to help answer Question 1.
3. Don't forget to label the axes. Use **xlab** and **ylab**.
4. Skim the paper "A Tour of Trellis Graphics" on the web site for helpful hints.
5. To draw a scatter plot with logarithmic axes, use plot(x,y, log="xy").
6. The R function **tapply** is good for calculating means for data divided into groups.
7. You can get the row labels from a data frame using the function **row.names.**

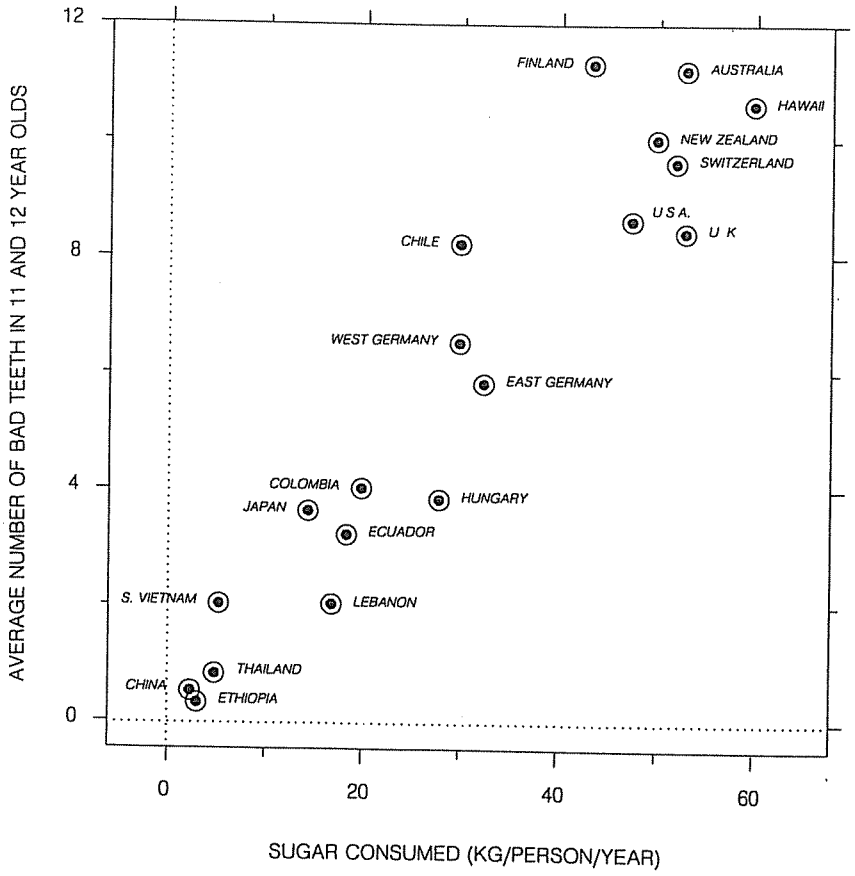

Figure 2.22 DATA LABELS. *Do not allow data labels in the data region to interfere with the quantitative data or to clutter the graph.* The data labels on this graph are needed to convey the names. The visual impact of the labels has been lessened so that they interfere as little as possible with our assessment of the overall pattern of the quantitative data.

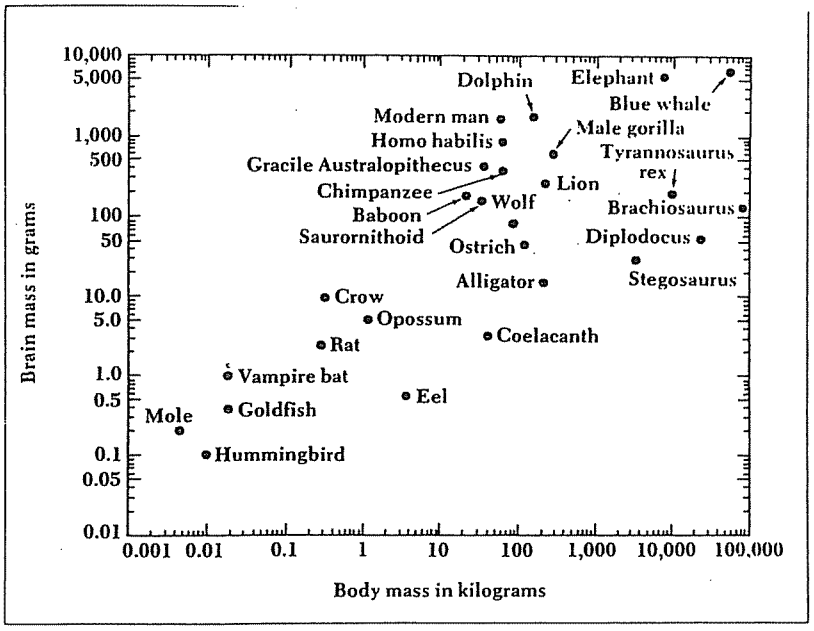

Figure 2.23 DATA LABELS. The data labels interfere with our assessment of the overall pattern of the quantitative data.

Oats Data

	B	V	N	Y
1	I	Victory	0.0cwt	111
2	I	Victory	0.2cwt	130
3	I	Victory	0.4cwt	157
4	I	Victory	0.6cwt	174
5	I	Golden.rain	0.0cwt	117
6	I	Golden.rain	0.2cwt	114
7	I	Golden.rain	0.4cwt	161
8	I	Golden.rain	0.6cwt	141
9	I	Marvellous	0.0cwt	105
10	I	Marvellous	0.2cwt	140
11	I	Marvellous	0.4cwt	118
12	I	Marvellous	0.6cwt	156
13	II	Victory	0.0cwt	61
14	II	Victory	0.2cwt	91
15	II	Victory	0.4cwt	97
16	II	Victory	0.6cwt	100
17	II	Golden.rain	0.0cwt	70
18	II	Golden.rain	0.2cwt	108
19	II	Golden.rain	0.4cwt	126
20	II	Golden.rain	0.6cwt	149
21	II	Marvellous	0.0cwt	96
22	II	Marvellous	0.2cwt	124
23	II	Marvellous	0.4cwt	121
24	II	Marvellous	0.6cwt	144
25	III	Victory	0.0cwt	68
26	III	Victory	0.2cwt	64
27	III	Victory	0.4cwt	112
28	III	Victory	0.6cwt	86
29	III	Golden.rain	0.0cwt	60
30	III	Golden.rain	0.2cwt	102
31	III	Golden.rain	0.4cwt	89
32	III	Golden.rain	0.6cwt	96
33	III	Marvellous	0.0cwt	89
34	III	Marvellous	0.2cwt	129
35	III	Marvellous	0.4cwt	132
36	III	Marvellous	0.6cwt	124
37	IV	Victory	0.0cwt	74
38	IV	Victory	0.2cwt	89
39	IV	Victory	0.4cwt	81
40	IV	Victory	0.6cwt	122
41	IV	Golden.rain	0.0cwt	64
42	IV	Golden.rain	0.2cwt	103
43	IV	Golden.rain	0.4cwt	132
44	IV	Golden.rain	0.6cwt	133
45	IV	Marvellous	0.0cwt	70
46	IV	Marvellous	0.2cwt	89
47	IV	Marvellous	0.4cwt	104

```
48  IV  Marvellous 0.6cwt 117
49   V     Victory 0.0cwt  62
50   V     Victory 0.2cwt  90
51   V     Victory 0.4cwt 100
52   V     Victory 0.6cwt 116
53   V Golden.rain 0.0cwt  80
54   V Golden.rain 0.2cwt  82
55   V Golden.rain 0.4cwt  94
56   V Golden.rain 0.6cwt 126
57   V  Marvellous 0.0cwt  63
58   V  Marvellous 0.2cwt  70
59   V  Marvellous 0.4cwt 109
60   V  Marvellous 0.6cwt  99
61  VI     Victory 0.0cwt  53
62  VI     Victory 0.2cwt  74
63  VI     Victory 0.4cwt 118
64  VI     Victory 0.6cwt 113
65  VI Golden.rain 0.0cwt  89
66  VI Golden.rain 0.2cwt  82
67  VI Golden.rain 0.4cwt  86
68  VI Golden.rain 0.6cwt 104
69  VI  Marvellous 0.0cwt  97
70  VI  Marvellous 0.2cwt  99
71  VI  Marvellous 0.4cwt 119
72  VI  Marvellous 0.6cwt 121
```

Animal data

```
                     body  brain
Mountain beaver      1.350    8.1
Cow                465.000  423.0
Grey wolf           36.330  119.5
Goat                27.660  115.0
Guinea pig           1.040    5.5
Dipliodocus      11700.000   50.0
Asian elephant    2547.000 4603.0
Donkey             187.100  419.0
Horse              521.000  655.0
Potar monkey        10.000  115.0
Cat                  3.300   25.6
Giraffe            529.000  680.0
Gorilla            207.000  406.0
Human               62.000 1320.0
African elephant  6654.000 5712.0
Triceratops       9400.000   70.0
Rhesus monkey        6.800  179.0
Kangaroo            35.000   56.0
Golden hamster       0.120    1.0
Mouse                0.023    0.4
Rabbit               2.500   12.1
Sheep               55.500  175.0
Jaguar             100.000  157.0
```

Chimpanzee	52.160	440.0
Rat	0.280	1.9
Brachiosaurus	87000.000	154.5
Mole	0.122	3.0
Pig	192.000	180.0