1. Introduction

The graphical analysis of data is a necessary complement of inferential statistics and a powerful tool for making statistical results accessible for nonstatisticians. Clear graphical presentation of data is recommended by the Good Statistical Practice guidelines.

Let us suppose that once the term Good Graphical Practice will be introduced. What principles would it lay down? I guess something like:

1. Meaningful graphs have to present as much data as it is possible without damaging the clarity of the image.
2. Unimportant details should not take the place of essential information.
3. Manipulation of data by inadequate weighting, by making breaks of the axes or by linking similar graphs which are differently scaled has to be avoided.
4. To present the precision of data is as important as the data itself.

The next question is how much support can we get in SAS if we want to construct graphs respecting the above principles. If we study the SAS/GRAPH manual more attentively as usual, we find a lot of unexploited possibilities (juggling with the interpol and size design control options can solve many problems). The SAS/STAT manual also contains some new types of graphs related to statistical procedures like LIFETEST, FACTOR, REG, UNIVARIATE etc. However, the third and the most flexible tool is Data Step Graphics Interface (DSGI) which permits us to design our own graphs. I prefer DSGI to annotate because DSGI permits the combination of graphs.

2. Presentation of continuous data

Example: In a clinical trial for comparing the effects of two antiinflammatory drugs, the following white blood cell counts were registered:

<table>
<thead>
<tr>
<th>Visit nr.</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>33</td>
<td>11.70</td>
<td>4.87</td>
<td>5.0</td>
<td>23.0</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>33</td>
<td>10.10</td>
<td>3.22</td>
<td>6.0</td>
<td>23.0</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>32</td>
<td>9.94</td>
<td>4.44</td>
<td>5.0</td>
<td>26.0</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>31</td>
<td>9.87</td>
<td>4.94</td>
<td>4.0</td>
<td>32.0</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>30</td>
<td>7.87</td>
<td>2.39</td>
<td>4.0</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>31</td>
<td>8.52</td>
<td>2.31</td>
<td>4.0</td>
<td>14.0</td>
</tr>
</tbody>
</table>

The analysis of variance for these data resulted the following table:
The most common presentation of these data is like in Figure 1:

![White blood cell counts](image)

**Figure 1.**

Figure 1 (created with *gplot*) illustrates well the decrease of means, the standard deviation of the two samples, but does not show the sample size, thus we don’t know the width of the confidence intervals and we don’t know the proportion of patients outside the normal range as well.

![Individual values](image)

**Figure 2.**
Figure 2. shows the individual values and the normal ranges for the two treatment groups (the two graphs were made with *proc gplot* and they were put together in DSGI). However, Figure 3. seems to be more informative because it reflects not only the proportion of patients outside the normal ranges but the distribution of values inside the normal range as well. The normal region was divided into ten subregions of equal width, thus the figure shows changes in distribution within the normal range (again, the 6 graphs were drawn with *gchart* and they were combined in DSGI). This latter type of graph shows not only the magnitude of the main effects from the analysis of variance but their interaction as well (Cleveland$^1$).

3. **Categorical data**

Graphical methods for this type of data are rather poor. Traditional bar charts are not able to show complex relations in multivariate problems and they are not sufficiently informative in presenting bivariate relationship.

**Example** (Armitage$^2$): The aim of a clinical trial is to compare the effects of PAS and streptomycin in the treatment of pulmonary tuberculosis. The group frequencies obtained are the following
These data can be represented on two linked bargraphs (Figure 4.), the first one showing the number of patients on each treatment while the second one showing their proportion in each outcome category. The two graphs were constructed with proc gchart, they were put together with DSGI, while the links between the two charts were drawn in the Graph Edit window.

![Figure 4.]

The second chart could be more easy to interpret if the bars would not differ in height but in width (spine plots, Hummel').

Such graphs cannot be created with gchart because it does not have an option like the very useful size option in g3d. Therefore we applied proc g3d in order to create a spine plot (Figure 5.). Proc g3d does not accept more z values corresponding to the same (x,y) pair and none of the variables is permitted to have a single level, so some tricks and dummy variables had to be introduced.
Figure 5.

The program which created the two above charts is rather long and complicated (not because of proc g3d but because of the transformations which has to be made before), but their combination in DSGI is simple like this:

data _null_;  
  rc=gset('catalog','rajz','d3');  
  rc=ginit();  
  rc=graph('clear','tot');  
  do j=1 to 2;  
    n=j;  
    ly=0.03;  
    uy=0.97;  
    lx=0.05+(j-1)*0.45;  
    ux=0.05+j*0.45;  
    rc=gset('window',n,0,0,100,100);  
    rc=gset('viewport',n,lx,ly,ux,uy);  
  end;  
  rc=gset('transno',1);  
  rc=graph('insert','g3d');  
  rc=gset('transno',2);  
  rc=graph('insert','g3d1');  
  rc=graph('update');  
  rc=gterm();  
run;
4. Survival times

Various survival plots can be created with the LIFETEST procedure. The problem is that these graphs cannot be customized like those which are the output of a GRAPH’s procedure. For example if more graphs are created, the axes are not equally scaled and their scaling cannot be changed. In order to combine survival plots with the same axes one has to write his own procedure (using gplot and interpol=stepjs).

Figure 6.

Figure 6 presents an example of Box and Cox\(^4\) showing the survival times of 12 animal groups randomly allocated to three poisons and four treatments in order to investigate how to combat the effects of certain toxic agents. As it is expected, the POISON*TREATMENT interaction is significant, that means each poison has a specific treatment. Putting together the 12 lifetables this can be easily read from the diagram.

5. Meta-analysis

Combining the results of several clinical trials requires clear and meaningful presentation. This is a field where the statistician is really alone: no SAS procedure helps his work. But with DSGI even this problem can be solved.

**Example:** Serum cholesterol reduction-trials (Thompson\(^5\))

The overview of 25 clinical trials attempts to quantify the effect of cholesterol reduction on the risk of ischaemic heart disease (IHD). The frequency of IHD in the treated group compared to that of the control group is characterized by the odds ratio and his 95% confidence interval for each trial. The conventional meta-analysis graph is shown on Figure 7.:
The program generating Figure 7. is a very simple one using only the `gdraw` and `gset` functions.

The disadvantage of this presentation is that odds ratios with the largest confidence intervals are dominating it. Galbraith proposes radial plots which reflect better the precision of an estimation - standardised log odds ratios are scattered versus their precision (Figure 8.). The circular scale on the right hand side enables the reading off true numerical values by extrapolating a line from the origin through the respective point. 95% confidence limits can be read off by extrapolating lines from (0,0) through (x,y+2) and (x,y-2), respectively.

Creating the circular scale in DSGI needs only a few program lines:

```plaintext
rc=gdraw('arc',10,56,100,-32,25);
doi=1.1 to 1.3 by 0.1;
    t=left(i);
    alfa=atan(&prop*log(i));
    u1=12+100*cos(alfa);
    u2=56+100*sin(alfa);
    rc=gdraw('text',u1,u2,t);
    rc=gdraw('mark',1,(u1-2),u2);
end;
doi=1.1 to 1.3 by 0.1;
    t=left(put(1/i,4.2));
    alfa=atan(&prop*log(i));
    u1=12+100*cos(alfa);
    u2=56-100*sin(alfa);
    rc=gdraw('text',u1,u2,t);
    rc=gdraw('mark',1,(u1-2),u2);
end;
```

![Figure 7.](image)

---

7
6. Conclusion

SAS/GRAPH®’s procedures applied together with DSGI data steps enables us to create most of the graphs we imagine. However, users would welcome a kind of standardisation of the procedures (for example: similar options and similar syntax for gchart, gplot and g3d) and some new procedures aiding the analysis of categorical data and meta-analysis.

References

1. Cleveland WS: Trellis display: Modelling data from designed experiments. 17th Meeting of Biostatistics, Budapest, 1996.