**2.4 A general approach to graphing data**

There are some general tools and principles for constructing graphs that can be useful. These involve general methods for displaying certain kinds of variables. By thinking of what variables are to be represented, and what types of variables they are, we can use the general tools to construct graphs. Some common ways of representing variables include

* Continuous variables are usually represented by positions along axes (e.g. x and y).
* Categorical variables are usually represented by different symbols or colors.
* Measurements on the same items are usually represented by lines connecting the items. This is especially useful for graphs versus time.
* Less commonly, continuous variables can be represented by varying levels of color or shading or size.
* Categorical variables can also be represented by axis positions.
* Categorical variables can be represented by panels also.

**Example:**

For the data below giving age, blood pressure, drug and status (lived/died) for 12 patients,

a. Graph the data below showing all four variables and comment on any patterns or relations you find. Use R to do this two ways:

i. A single graph, no multiple panels

ii. Using ggplot2 with separate panels for Drug

1. Comment on the effect of the drug on blood pressure and mortality. Note: The wording is important.

Data for the 12 patients: age, blood pressure, treatment (drug) and outcome (lived/died).

Age BP Drug L/D

18 68 1 D

20 64 2 L

22 72 1 D

25 67 2 L

29 80 None D

33 70 None D

34 86 1 D

36 85 None D

37 73 2 L

39 82 None L

41 90 1 D

45 87 2 L

**2.5 Examples**

1. **Cross-sectional data**

A comparison of several distributions is easy using PROC UNIVARIATE with a CLASS statement. Nrows makes 3 graphs vertically. I found this by searching the help files for ‘histogram’ and then noticed the entry ‘comparative’ which sounded promising. There are lots of options such as bar width, see help file.

**proc** **univariate** data=vicki908.countswdemo noprint;

class Group(order=formatted);

histogram dVMSedent dVMVig / vscale=count

vaxislabel='Frequency' turnvlabels nrows=**3** ncols=**1**

cframe=ligr cframeside=gwh cframetop=gwh cfill=gwh;

title 'VM counts by group';

**run**;



My favorite graph for cross-sectional data is the pairwise scatterplot matrix. This shows all bivariate graphs, and the arrangement and common scales allow easy comparisons and following of individual points. Pairwise scatterplot matrices can be made in Splus (or R).

As an example, consider data on five numerical outcomes at baseline in a study of lung function. The data can be imported into Splus and a simple pairs plot can be made using

### The easiest way is to save your Excel file as a type=.csv file. Then in R use

Alan2.fr <- read.csv( "C:/Documents and Settings/Gary/Desktop/Gary/dataforR.csv" )

### R can be picky if the filename path goes to a second line. I horizontally expand the notepad file.

pairs( alan2.fr[, c(1:5)] )



Lots of refinements are possible. For example, the following labels each point with its death status, and makes the default size a bit smaller (cex).

pairs( alan2.fr[, c(1:5)], panel=function(x, y) {

text( x, y, as.character(alan2.fr[, 6]), cex=0.5 )

} )

These graphs can be useful in longitudinal studies also, by defining new variables for changes (e.g. change from baseline to 12 months).

**Scatterplot matrix in SAS (Thanks to MS student Tony Moscoso)**

ods html; /\*turn on html output\*/

ods listing close; /\*turn off list (regular) output window, optional\*/

ods graphics on; /\*turn on ods graphics\*/

**proc** **corr** data = forpairs plots;

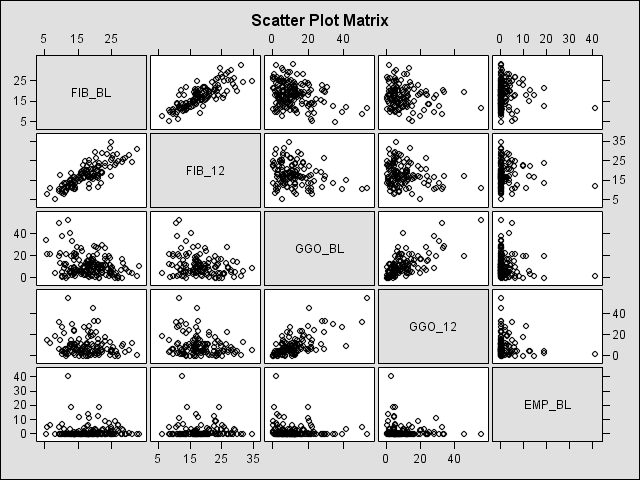
var FIB\_BL FIB\_12 GGO\_BL GGO\_12 EMP\_BL EMP\_12;

**run**;

ods graphics off;

ods html close;

ods listing;



From SAS FAQ (<http://www.ats.ucla.edu/stat/sas/faq/sas9_stat_plots.htm)>. Another MS student, Brandy Ringham, found other code at

[ftp://ftp.sas.com/techsup/download/sample/samp\_lib/imlsampScatterplot\_Matrix.html](ftp://ftp.sas.com/techsup/download/sample/samp_lib/imlsampScatterplot_Matrix.html" \t "_blank)

and said “I found some proc iml code that makes very nice scatterplots.  You only have input your own data set instead of the example and then change the colors so it prints white instead of black.”

There are many different ways of creating scatter matrix plot in SAS, one of the easy way is via proc corr. The option that is required is the plots option in the proc corr statement. Choosing different HTML style will create the same plot with different styles, such as the grayish default style, journal style and statistical style. We want to emphasize though that SAS 9.1.3 still claims that this feature is experimental. This may imply future change in syntax or other aspects. There are two things that one has to do in order for this new feature to work. One is to turn on the html format of output. The other is to turn on the graphics feature of ODS. Some of the procedures require more statement than others. I don’t know if it’s possible to code for groups by color or symbol.

Another useful graph for cross-sectional data with grouping variables is below. The outcome is Y, the predictors are X (takes values 3, 5, 8, 10), and group. Group is coded by color and also shifted a bit.



SAS code to produce graph:

axis2 order=(**0** to **12** by **2**);

axis1 order=(**0** to **20000** by **5000**) label=(a=**90**);

**proc** **gplot** data=datset;

plot Y\*X=group / vaxis=axis1 haxis=axis2;

symbol1 v=dot i=none c=red;

symbol2 v=dot i=none c=blue;

format X F4.1;

format Y F4.1;

label X='# of X';

label Y='# of Y';

title 'X vs Y by Group';

title2 'Red = Group 1, Blue = Group 2';

**run**;

Before creating graph to get offset:

**data** dataset; set dataset;

if group=**2** then X=X+**0.2**;

**run**;

A third predictor variable (e.g. sex) can be added with colors (group is still represented with the shift).



SAS code to produce graph:

axis2 order=(**0** to **12** by **2**);

axis1 order=(**0** to **20000** by **5000**) label=(a=**90**);

**proc** **gplot** data=dataset;

plot Y\*X=gender / vaxis=axis1 haxis=axis2;

symbol1 v=dot i=none c=green;

symbol2 v=dot i=none c=black;

format X F4.1;

format Y F4.1;

label X ='# of X Chosen';

label Y ='# of Y';

title 'X vs Y by Gender';

**run**;

1. **Longitudinal data**



SAS code to produce graph:

symbol1 i=join r=**10** w=**1** color=red;

symbol2 i=join r=**12** w=**1** color=blue;

**proc** **gplot** data=plotdat;

plot x1\*time = idplot / nolegend skipmiss;

\* Data should be sorted by time and idplot;

footnote font=centbi height=**1.2** color=brown 'Note: blue=male and red=female';

label time='Time (mins)';

title 'Variable X1 by Individual';

**run**; **quit**;



symbol1 i=std1mjt l=**1** mode=include w=**2** r=**1** color=red;

symbol2 i=std1mjt l=**1** mode=include w=**2** r=**1** color=blue;

**proc** **gplot** data=plotdat;

plot x1\*time = gender;

label time='Time (mins)';

title 'Variable X1 by Gender';

title2 'Means and 1st Standard Deviation';

**run**; **quit**;