The CHART Procedure

Very often a chart will assist an author in the illustration of points that are being made in the author's text. SAS can produce bar chart, pie charts, and three-dimensional charts that can be included the in the text of an article.

The CHART procedure produces low-resolution vertical and horizontal bar charts. Some summary statistics are also available. These charts would not be considered presentation quality graphics, but they can quickly convey distributional properties in a preliminary data analysis. For presentation quality (high-resolution) graphics, the GCHART procedure in SAS/GRAPH is recommended. The low resolution CHART procedure can be used as an introduction to the programming associated with the GCHART procedure.

The syntax of the CHART procedure is:

```
PROC CHART  DATA=setname   ;
BY variablelist ;                                        (optional statement)
WHERE condition list;                                    (optional statement)
VBAR variablelist  / <option list >  ;
HBAR variablelist  / <option list >   ;
BLOCK  variablelist  / <option list >;
PIE  variablelist   /  < option list >;
RUN;
```

The CHART procedure can be used to generate one or several charts. Each of the statements VBAR through STAR in the syntax above requests that chart(s) of that type be made. One of these statements or several can be used in a single CHART procedure. Available options for each of the statements following PROC CHART are listed in the table below.

BY Statement

A BY statement can be used with PROC CHART to obtain separate analyses on observations in groups defined by the BY variables. PROC SORT using the same BY statement must appear prior to the PROC CHART procedure using a BY statement.

WHERE Statement

The WHERE statement defines a subgroup for which the procedure is to be performed.

VBAR Statement

List the variables for which vertical bar charts are to be produced.

HBAR Statement

List the variables for which horizontal bar charts are to be produced.

BLOCK Statement

List the variables for which 3-D block charts are needed.

PIE Statement

The PIE statement requests a pie chart for each variable listed. The pie's size is determined only by the LINESIZE = and PAGESIZE = system options. See information about the OPTIONS statement to control line size and page size.

Summary of Statement Specific Options

Options Grouped by Function	HBAR	VBAR	BLOCK	PIE	STAR
Separate into groups					
GROUP =	X	X	X		
SUBGROUP =	X	X	X		
G100	X	X	X		
Request statistical analysis					
FREQ	X				
CFREQ	X				
PERCENT	X				
SUM	X				
MEAN	X				
Control Output					
NOLEGEND	X	X	X		
NOSYMBOL	X	X	X		
SYMBOL =	X	X	X		
ASCENDING	X	X			
DESCENDING	X	X			
NOZEROS	X	X			
REF =	X	X			
NOSTATS	X				
NOSPACE		X			
NOHEADER			X	X	X
Standard Options					
AXIS =	X	X	X	X	X
DISCRETE	X	X	X	X	X
FREQ =	X	X	X	X	X
LEVELS =	X	X	X	X	X
MIDPOINTS =	X	X	X	X	X
MISSING	X	X	X	X	X
SUMVAR =	X	X	X	X	X
TYPE =	X	X	X	X	X

Definition of options:

ASCENDING or ASC
prints the bars and any associated statistics in ascending order of size within groups.

AXIS = <*min-value*> *max-value* ;
allows you to specify the minimum and maximum values used in constructing the FREQ, PCT, CFREQ, CPCT, SUM, or MEAN axis. For examples: AXIS= 15 or AXIS = 2 4 6.

CFREQ
prints the cumulative frequency

CPERCENT
prints the cumulative percentages

DESCENDING or DESC
prints the bars and any associated statistics in descending order of size within groups

DISCRETE
is used when the numeric chart variable specified is discrete rather than continuous. If the DISCRETE option is omitted, PROC CHART assumes that all numeric variables are continuous and automatically chooses interval for them unless the MIDPOINTS = or LEVELS = options are used.

FREQ
prints the frequency of each bar to the side of the chart.

FREQ = *variable*
is used when a variable in the data set represent a count for each observation. Normally, each observation contributes a value of one to the frequency counts.

GROUP = *variable*
produces side-by-side charts, with each chart representing the observation have a given value of the GROUP = variable.

G100
is used in conjunction with the GROUP = option to force the bars and statistics to add to 100% for each group.

LEVELS = *number of midpoints*
specifies the number of bars, blocks, or sections representing each chart variable when the variable given in the VBAR statement are continuous.

MEAN
prints the mean of the observations represented by each bar.

MIDPOINTS = *midpoint list*
: defines the range of values each bar or section represents by specifying the range midpoints.

MISSING
: specifies that missing values are to be considered as valid levels for the chart variable

NOHEADER
NOHEADING
: suppresses the default header line normally printed at the top of a chart

NOSTATS
NOSTAT
: suppresses printing statistics on a horizontal bar chart.

NOSYMBOL
NOLEGEND
: is used in conjunction with the SUBGROUP = option to suppress printing of the subgroup legend or symbol table.

NOZEROS
: suppresses any bar with zero frequency.

PERCENT
: prints the percentages of observations having a given value for the chart variable

REF = *value*
: draws a single reference line on the response axis.

SUBGROUP = *variable*
: subdivides each bar into characters that show the SUBGROUP = variables contribution to the bar.

SUM
: prints the total number of observations that each bar represents

SUMVAR
: names the variable to collect summaries for means, sums, or frequencies

SYMBOL = '*character list*'
: is used, when the SUBGROUP = option is not used to define the symbol or symbols in the bars or blocks of the chart.

TYPE = *type*

specifies what the bars of sections in the chart represent. The default TYPE is FREQ. When SUMVAR = option is specified, the default TYPE is SUM. You can specify one of the following keywords to type:

CFREQ
CPERCENT or CPCT
FREQ
MEAN
PERCENT or PCT
SUM

Objective: Run the SAS program on the following page and experiment with the various options for each of the types of charts produced.

You will notice that the images produced by the CHART procedures are not as attractive as some of the images produced by other graphics software. The images produced by the CHART procedure can be saved to a diskette. Since they are not graphics files, but rather, are text files, it requires less disk space to save them.

If you wish to produce a high-resolution image (presentation quality), edit the program and replace CHART with GCHART. GCHART is a SAS/GRAPH procedure that will produce a better image. You should consult the SAS/GRAPH manuals or other SAS/GRAPH reference material for more information. For now, one can view the CHART procedure as an introduction into the more advanced GCHART procedure. GCHART has many more options since the programmer has the ability to select colors and fonts and other enhancements. For content however, the programming of the CHART and the GCHART procedures are quite similar.

Note: GCHART output is not printed in the OUTPUT window but a graphics window. To print from the graphics window image, you must first move to the Graphics window (see the **Window** pull-down menu) and move to the image to be printed by using the **Page Up** and **Page Down** buttons. Then select **FILE - PRINT** from the pull-down menus. Images accumulate in the Graphics window when more than one run of a high-resolution graphics program is done. This window can be "cleared," however, it is not as easily accomplished as clearing the Output and Log windows.

Bar Charts
A SAS Program

Two instructional programs are to be compared. Students independently participate in one of the two programs. Scores at the completion of the instructional period are recorded for each of the students. Information is given in the table below.

Program A	Program B
71 82 88 64 59 78	65 88 92 76 87 89
72 81 83 66 83 91	85 90 81 91 78 81
79 70	86 82 73 79

The SAS code below will generate different bar charts that represent this data. Bring a copy of your program on disk, a hard copy of your program, and a hard copy of your output to the next class.

```
DM 'LOG; CLEAR; OUTPUT; CLEAR; ';
OPTIONS PAGENO=1;
DATA one;
INPUT program $ score @@;
CARDS;
A 71 A 82 A 88 A 64 A 59 A 78 A 72
A 81 A 83 A 66 A 83 A 91 A 79 A 70
B 65 B 88 B 92 B 76 B 87 B 89 B 85
B 90 B 81 B 91 B 78 B 81 B 86 B 82
B 73 B 79
PROC CHART DATA=one;
VBAR score;
TITLE 'Default Vertical Bar Chart for the Scores';

PROC SORT  DATA=one; BY program;
PROC CHART DATA=one; BY program;
VBAR score;
TITLE 'Default Vertical Bar Chart for the Scores for Each Program';

PROC CHART DATA=one;
VBAR score / GROUP=program MIDPOINTS = 65 70 75 80 85 90 95;
TITLE 'Vertical Bar Chart for the Scores for Each Program';
TITLE2 'Midpoints are specified and used for each chart';

PROC CHART DATA=one;
HBAR score;
TITLE 'Default Horizontal Bar for the Scores';
TITLE2 ;

PROC CHART DATA=one;
HBAR score / LEVELS = 4 NOSTATS;
TITLE 'Horizontal Bar Chart ';
TITLE2 'Number of Levels Specified - Frequency Information is Suppressed';
```

```
PROC CHART DATA=one;
HBAR score / FREQ AXIS=15;
TITLE 'Horizontal Bar Chart';
TITLE2 'FREQ and AXIS=15 Options are Illustrated';

PROC CHART DATA=one;
BLOCK score program;
TITLE 'Default Blocked Chart for Both Program and Scores';
TITLE2 'Both are Frequency Charts';

PROC CHART DATA=one;
BLOCK program / DISCRETE TYPE=MEAN SUMVAR=score;
TITLE 'Blocked Horizontal Bar Chart with Number of Blocks Specified';
TITLE2 'Bar Heights are the mean scores as requested by the options';

PROC CHART DATA=one;
PIE program ;
TITLE 'Default Pie Chart';
RUN;
QUIT ;
```