Abstract

By combining the use of procedure options with the annotate facility in SAS/GRAPH®, you have virtually unlimited control over picture production. This paper utilizes this combination to produce customized bar charts with bars constructed by stacking symbols either produced with PROC GFONT or selected from SAS/GRAPH fonts. PROC GFONT is used to construct the custom charting symbol, PROC GPLOT is used to produce axes scaled for the data set being charted, and the annotate facility is used to stack the appropriate number of symbols to produce each bar.

Introduction

Stylized bar charts constructed by stacking symbols which relate to the data being processed have become popular in presentations and the mass media. A bar chart depicting automobile sales across three years might be constructed by stacking the appropriate number of "automobile" symbols for each year's bar, with each symbol representing a fixed number of automobiles sold. This paper presents techniques for constructing vertical bar charts with each bar produced by stacking symbols either selected from available SAS/GRAPH fonts or created by the user with PROC GFONT. For simplicity, only vertical bar charts will be considered and it is assumed that the charting variable is a character variable. However, the techniques presented can be adapted to horizontal bar charts and numeric variables.

Specifically, vertical bar charts will be created to depict the "sun index" in five cities by stacking palm tree symbols from the SAS/GRAPH cartography font. Additionally, "fish" symbols are used to create a bar chart depicting people's preference for three different types of fish—pike, tuna, and salmon. A fish font consisting of three different symbols, one for each of the three species of fish, is created. One chart is produced by stacking the appropriate number of each symbol to form the bars while a second chart is produced by using a single symbol of the appropriate length to form each bar.

Considerations

The actual construction of the charts requires several steps.

1) The selected symbol is digitized and a SAS® data set appropriate for use with PROC GFONT is created. PROC GFONT is then used to build the symbol.

2) The data set being charted is processed with a DATA step to determine the general characteristics of the chart. The number of bars and the height of the smallest and largest bars are determined for use in scaling the axes and in selecting the number of units of the response variable to be represented by a single symbol.

3) An ANNOTATE= data set is created which places the appropriate number of correctly sized symbols in position to form the bars. XSYS, YSYS, and HSYS values of 2' are used to take advantage of the axis placement and scaling done by PROC GPLOT.

4) PROC GPLOT is invoked for the data set being charted. The response and midpoint axes are scaled based on the information obtained in step 2 and the default plotting symbol is suppressed to provide an empty area within the axes for construction of the bars. The ANNOTATE= option on the plot statement is used to create the bars from the annotate data set created in step 3.

5) To make the programs flexible, macro variables are used whenever possible to specify which font and data set are in use, the chart and response variables, and various appearance variables such as color, text, and font.

6) As a final consideration, it should be noted that determining the minimum and maximum numbers of symbols, and hence the size of the symbols, used to construct the bars is a matter of individual taste. Similarly, if the exact height of a bar does not correspond to an even number of symbols, should you approximate the height of the bar with the nearest whole number of symbols or should you make the last symbol on the bar a partial symbol? The latter option was selected in this paper to preserve the accuracy of the chart. However, under certain conditions this might result in the last symbol being so incomplete as to detract from the appearance of the chart.

Example 1 - Single Symbol Bars

This chart is constructed according to the steps outlined above.
Creating the Symbol

It is assumed that you are familiar with PROC GFONT and have built the desired charting symbol. This paper uses a fish font containing three symbols and the SPECIAL font available with SAS/GRAPH. The data set and program for building one symbol in the fish font are shown in Appendix 1.

Creating the Response Data Set

The following SAS data set serves as the response data set for the first example.

DATA CATCH;
  INPUT FISH $ CATCH;
CARDS;
PIKE 4657
SALMON 10235
TUNA 6745
RUN;

Initialization of Macro Variables

The program below sets the values of macro variables used throughout the example.

****** USER INITIALIZATION ****************************;
OPTIONS NOTEST82 CRACKS;
GOPTIONS NOTEXT82 CRACKS;
%LET DATASET = CATCH; %RESPONSE DATA SET;
%LET XC = FISH; %AXIS VARIABLE;
%LET FREQUENCY = FISH; %FREQUENCY VARIABLE;
%LET CONTAINING SYMBOL; %CONTAINING SYMBOL;
%LET FONTNAME = FISH; %FONT NAME
%LET FONTCHAR = FISH; %VAR NAME OF CHARS;
%LET COLOR = BLACK; %COLOR FOR SYMBOL;
%LET LABEL = Index;
*** TITLES *********************************************;
TITLE 'Fish Popularity Index';
TITLE2 H=2;

Determining Characteristics of the Chart

The SAS DATA step shown below reads and processes the data set to be charted and creates macro variables for response axis maximum, major tic mark increment, and midpoint axis length.

***************************************************************************************;
*** DETERMINE MAXIMUM FOR VAXIS (SINCE NEED TO SPECIFY A START AT ZERO, SPECIFYING *** MACRO VARIABLE FOR STOP MAKES PROGRAM *** AUTOMATED RATHER THAN MANUALLY DATA *** DEPENDENT). *** EIGHT TIC MARKS WILL BE ASSUMED *** (TOO MANY CAUSES SHORT AXIS) *** DETERMINE LENGTH OF MIDPOINT AXIS=#BARS*15%. *** SUCH THAT 40% LE LENGTH LE 75% *******************************************************************************;
PROC SORT DATA= &DATASET; BY &FREQVAR ;

Creating the ANNOTATE= Data Set

The SAS DATA step shown below reads the response data set to be charted and creates one observation in an ANNOTATE= data set for each bar. The size (height) of the symbol is based on the response variable in the charted data set.

****** USER INITIALIZATION ****************************;
OPTIONS NOTEST82 CRACKS;
GOPTIONS NOTEXT82 CRACKS;
%LET DATASET = CATCH; %RESPONSE DATA SET;
%LET XC = FISH; %AXIS VARIABLE;
%LET FREQUENCY = FISH; %FREQUENCY VARIABLE;
%LET CONTAINING SYMBOL; %CONTAINING SYMBOL;
%LET FONTNAME = FISH; %FONT NAME
%LET FONTCHAR = FISH; %VAR NAME OF CHARS;
%LET COLOR = BLACK; %COLOR FOR SYMBOL;
%LET LABEL = Index;
*** TITLES *********************************************;
TITLE 'Fish Popularity Index';
TITLE2 H=2;

***************************************************************************************;
*** DETERMINE MAXIMUM FOR VAXIS (SINCE NEED TO SPECIFY A START AT ZERO, SPECIFYING *** MACRO VARIABLE FOR STOP MAKES PROGRAM *** AUTOMATED RATHER THAN MANUALLY DATA *** DEPENDENT). *** EIGHT TIC MARKS WILL BE ASSUMED *** (TOO MANY CAUSES SHORT AXIS) *** DETERMINE LENGTH OF MIDPOINT AXIS=#BARS*15%. *** SUCH THAT 40% LE LENGTH LE 75% *******************************************************************************;
PROC SORT DATA= &DATASET; BY &FREQVAR ;

Creating the ANNOTATE= Data Set

The SAS DATA step shown below reads the response data set to be charted and creates one observation in an ANNOTATE= data set for each bar. The size (height) of the symbol is based on the response variable in the charted data set.
**Producing the Chart**

PROC GPLOT is used to produce axes which are annotated with the charting symbol. The axes are scaled using the macro variables created from processing the response data set and the plotting symbol is suppressed. The annotate data set is used to produce bars consisting of a single symbol of the appropriate size per bar.

---

**Example 2 - Stacked Symbols**

The charts below are produced by modifying the programs in Example 1 to stack an appropriate number of symbols to produce each bar. The first chart uses the custom fish font discussed previously while the second chart uses the palm tree symbol from the SAS/GRAPH cartography font.
The SAS programs to produce the charts are very similar to those discussed in Example 1 and are illustrated in Appendix 2. Three modifications were made to the programs that were used in Example 1.

1) The section of the program that processes the response data set and determines characteristics of the chart is expanded to determine the size of (number of units of the response variable represented by) each symbol.

2) A "bar" symbol is created in the same scale as the font and added to the font. The "bar" symbol is used to white-out a part of a charting symbol by overlaying a white bar symbol on the top symbol of the bar when the height of a bar is not an integer number of charting symbols. This produces a bar of the exact desired height containing a partial symbol as the top symbol. The color white is used for the bar when the picture is sent to the plotter so that the top portion of the last symbol is invisible.

3) The annotate data set produces, within a DO loop, symbols stacked one on top of the other to produce each bar of the correct height.

Summary

This paper presents techniques by which bar charts can be produced by using a single symbol of the appropriate size to construct each bar or by stacking the appropriate number of symbols to create each bar. The annotate facility and PROC GPLOT with selected options are combined to create the charts from customized charting symbols.

SAS and SAS/GRAPH are registered trademarks of SAS Institute, Inc., Cary, NC, USA.

Appendix 1

CREATE THE FAMOUS FISH FONTS

ESOX SP. DATA = PIKE

'P' AND 'P'

DATA PE;
RETAIN CHAR 'P';
INPUT SEGMENT Y X @@;
Y= (Y-.5) *.0413; X=X-.5; X=X.956;
CARDS:
1 0.5 5.0 1 0.5 5.5 1 0.6 5.5 1 1.0 6.0
1 5.0 6.3 1 5.0 7.0 1 7.0 7.5 1 9.0 7.8
1 15.0 8.0 1 22.0 8.1 1 25.0 7.9 1 27.0 7.5
1 29.0 7.3 1 31.0 7.5 1 32.0 6.0 1 31.0 6.8
1 26.0 4.5 1 26.0 4.0 1 23.5 3.0 1 20.0 2.7
1 18.0 3.0 1 16.5 2.6 1 17.0 2.4 1 11.0 2.2
1 10.5 3.0 1 8.7 3.8 1 8.5 2.8 1 6.0 3.3
1 5.0 3.5 1 4.0 3.8 1 2.0 4.4 1 2.5 3.0
2 4.5 6.3 2 4.9 6.5 2 5.3 6.3 2 4.9 6.1
2 4.5 6.3 3 6.6 5.5 3 4.5 4.2 3 4.9 4.5
2 4.0 5.0 3 6.6 5.5 4 22.0 8.1 1 25.0 7.9
6 27.0 7.5 4 28.4 8.0 4 28.9 9.0 4 27.5 11.5
6 25.0 10.6 4 22.0 8.1 5 31.0 7.5 5 32.0 6.0
5 31.0 4.8 5 36.0 2.4 5 36.8 3.0 5 35.0 5.5
5 34.3 6.0 5 35.0 6.5 5 36.8 9.0 5 35.6 10.0
5 31.0 7.5 6 28.0 4.5 6 26.0 4.0 6 23.5 3.0
6 27.0 1.0 6 28.0 0.5 6 29.0 1.0 6 29.0 2.0
6 29.5 3.0 6 28.0 4.5 7 20.0 2.0 7 18.0 3.0
7 16.5 2.6 7 17.0 2.2 7 19.0 1.2 7 21.0 0.6
7 21.5 6.0 7 21.5 1.4 7 20.0 2.7 8 11.0 2.2
8 10.5 3.0 8 8.7 3.8 8 8.5 2.8 8 10.5 0.5
8 11.2 1.0 8 11.0 2.2 9 6.0 3.3 9 8.0 3.5
9 9.8 4.4 9 9.5 5.0 9 9.4 5.8 9 9.3 6.4
9 9.4 5.8 9 9.3 5.0 9 9.0 4.4 9 8.8 3.5
9 6.0 3.3 10 5.0 3.5 10 6.0 4.0 10 7.0 5.0
10 7.0 6.0 10 7.0 5.0 10 6.0 4.0 10 5.0 3.5

DATA EMPTY;
SET EMPTY EMPTY(IN=RIGHT);
IF RIGHT THEN DO;
X=-X;
CHAR=UPCASE(CHAR);
END;
PROC GFONT DATA=FISHBOX NAME=FISHBOX FILLED NODISPLAY;
RUN;
PROC GFONT DATA=FISHBOX NAME=FISHBOX FILLED NODISPLAY;
RUN;
Appendix 2

GOPTIONS NOCHARACTERS NOCELLS;
DATA BOX;
  INPUT x Y;
  RETAIN LP 'P' SEGMENT 1 CHAR 'B';
CARDS;
o 0
o 1
1 1
1 0
0 0
;
PROC GEFONT DATA=BOX NAME=BOX FILLED NODISPLAY;
RUN;
DATA PALMTREE;
  INPUT PLACES $15. SUNINDEX;
CARDS;
Orlando 50
Nassau 70
Negril 85
Rio 95
Cancun 75
OPTIONS DQUOTE;
GOPTIONS NOTEXT CBACK;
*====> USER INITIALIZATION - PROVIDE INFORMATION;
*** IF FONT AND DATASET ARE TEMPORARY. EITHER
RUN THOSE PROGRAMS OR %INCLUDE THEM HERE;
*
DATA TO BE
GRAPHED;

%LET DATASET=PALMTREE; DATASET WITH COUNTS;
%LET XC=PLACES; HORIZONTAL AXIS VARIABLE;
%LET FREQVAR=SUNINDEX; FREQUENCY VARIABLE NAME;
%LET FONTNAME=CARTOG; FONT WITH SYMBOL(S);
%LET FONTCHAR='L'; FONT CHAR OR VAR NAME;
%LET COLOR=.BLACK; COLOR FOR SYMBOL;
%LET BOXFONT=BOX; FONT NAME CONTAINING "BOX";
%LET BOXCHAR='B'; FONT CHAR FOR BOX;
%LET RLABEL=SUNINDEX; RESPONSE AXIS LABELLING;
** TITLES - NOTE: NEED A TITLE LINE THAT
FORCE "SOME EXTRA SPACE SO THAT NO
TITLES ARE OVERWRITTEN;
TITLE1 C=BLACK F=DUPLEX 'HOT VACATION SPOTS'
TITLE2 H=2;
** DETERMINE MAXIMUM FOR VAXIS (SINCE NEED TO
SPECIFY A START AT ZERO, SPECIFYING MACRO
VARIABLE FOR STO CHART MAKES PROGRAM AUTOMATED
RATHER THAN MANUALLY DATA DEPENDENT);
** EIGHT TIC MARKS WILL BE ASSUMED(TOO MANY
CAUSES SHORT AXIS);
** SUCH THAT 40% LE LENGTH LE 75% ;
PROC SORT DATA= &DATASET; BY &FREQVAR
DATA _NULL_;
* USE FIRST AND LAST FREQUENCY VALUES ;
DO I=1,LAST;
  SET &DATASET NOS=LAST POINT=I;
  IF I = 1 THEN
    MIN= &FREQVAR;
  IF I = LAST THEN
    MAX= &FREQVAR;
*** USE CEIL TO ASSURE AXIS GOES HIGH ENOUGH ;
*** USE +8 ON MAX TO ACCOUNT FOR INCREMENT ROUNDED UP;
   CALL SYMPUT('MAX', PUT(CEIL(MAX)+8, 12. ));
   CALL SYMPUT('INC', PUT(CEIL(MAX/8), 12. ));
*** COMPUTE SIZE OF SYMBOL. IF MIN/2
GIVES AT LEAST 5 SYMBOLS BUT NO MORE
THAN 12 FOR MAX VALUE, FINE.
OTHERWISE, USE (MAX-MIN)/7 ;
IF 5 LE MAX/(MIN/2) LE 12
THEN SIZE = MIN/2;
ELSE SIZE = (MAX-MIN)/7 ;
   CALL SYMPUT('SIZE', PUT(SIZE, 12. ));
*** MAX FOR CHART NEEDS TO BE HIGHER
SO THAT "BLANK-OUT" BOX DOESN'T
SHOW ABOVE AXES;
   CALL SYMPUT('MAXCHART',
     PUT(CEIL(MAX+SIZE)+8, 12. ));
   CALL SYMPUT('INCCART',
     PUT(CEIL(MAX/8), 12. ));
*** COMPUTE RESPONSE AXIS LENGTH;
IF 3 LE LAST LE 5 THEN
  LENGTH = LAST+15;
ELSE IF LAST LT 3 THEN LENGTH = 40 ;
ELSE LENGTH = 75;
   CALL SYMPUT('LENGTH', PUT(LENGTH, 2. ));
END;
STOP;

** CREATE THE ANNOTATE DATASET. THIS DATASET
WILL GENERATE THE BARS TO PUT ON THE
PROC G PLOT AXES.
DATA ANNOPLT;
  LENGTH FUNCTION COLOR STYLE $ 8 TEXT $ 1.
  RETAIN XSYS YSYS HSYS '2' POSITION 'B'
  SIZE &SIZE FUNCTION 'LABEL';
  STYLE = "&FONTNAME";
  SET &DATASET ;
  XC= &XC ; TEXT= &FONTCHAR
  GENERATE SYMBOLS FOR BARS;
  DO 1=1 TO &FREQVAR BY &SIZE;
    Y=I; COLOR= "&COLOR";
    OUTPUT;
  END;
  IF A FRACTIONAL SYMBOL IS NEEDED, ABOVE
WILL HAVE PUT ENTIRE SYMBOL OUT. THUS,
BEGINNING AT FREQVAR VALUE, PLACE OUT A
WHITE BOX SO THAT ONLY APPROPRIATE PORTION
OF LAST SYMBOL WILL BE SEEN ;
   CALL SYMPUT('MAX', PUT(MOD(MAX,SIZE)+8, 12. ));
   CALL SYMPUT('INC', PUT(MOD(MAX/8), 12. ));
** USE G PLOT TO PRODUCE THE AXES FOR
THE CUSTOMIZED PLOT ;
PROC G PLOT DATA= &DATASET;
  PLOT &FREQVAR* &XC
  HAXIS=AXIS1 VAXIS=AXIS2
  ANNO=ANNOPLT CAXIS=BLACK;
  SYMBOL V=NONE;
  AXIS1 VALUE=(C=BLACK F=DUPLEX H=1.2)
  LABEL=(C=BLACK F=DUPLEX A=-90 R=90 "&RLABEL";
  AXIS2 OFFSET=(0,2) ORDER=(0 TO &MAX BY &INC)
  VALUE=(C=BLACK F=DUPLEX H=1.2)
  LABEL=(C=BLACK F=DUPLEX H=1.7)
  A=50 B=90 "&RLABEL";
RUN;