Statistical Graphics Procedures

SAS/Graph software has many components. Among the SAS procedures are a collection of procedures referred to as Statistical Graphics Procedures. These procedures are: SGPLOT, SGSCATTER, SGPANEL, SGRENDER, and SGDESIGN. SGPLOT and SGPANEL will be briefly overviewed here.

As with ODS Graphics, as you read through the information below and work through the objectives, you may want to have SAS Help and Documentation 

open.

For these objectives, you will need the AGENTS, CRIME, and HOUSES data tables. For the objectives it is assumed that these data tables are in a library called CLASS.

The SGPLOT Procedure

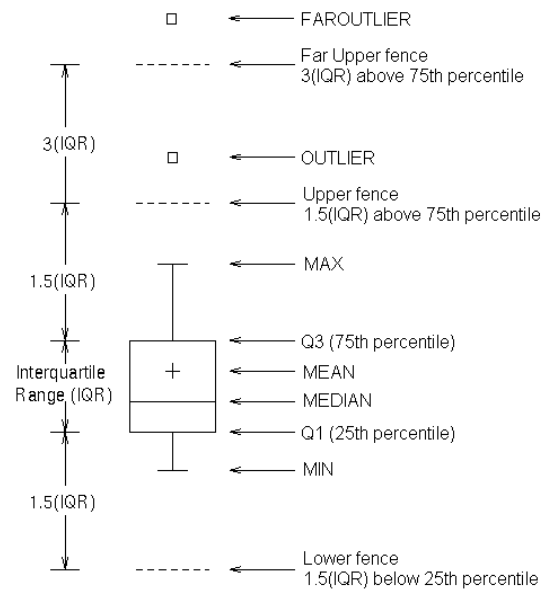
The SGPLOT Procedure can put one or more plots on a single set of axes. Though there are many statements that comprise the SGPLOT procedure, some of the introductory statements or options are presented here.

PROC SGPLOT Syntax

```
PROC SGPLOT DATA=tablename;
HBAR response-variable </options> ;
HBOX response-variable </options> ;
HISTOGRAM response-variable  </options> ;
DENSITY response-variable </options> ;
SCATTER X=xvariable  Y=yvariable </options> ;
SERIES  X=xvariable  Y=yvariable </options> ;
VBAR response-variable </options> ;
VBOX response-variable </options> ;


HBAR response-variable </options> ;
HBAR creates a horizontal bar chart summarizing a category variable.
Options include:
      Bar options:      BARWIDTH== numeric-value
                                   value is 0 to 1; 1 is the default setting
                        FILL | NOFILL
                        FREQ = numeric-value
                        LIMITS = BOTH | UPPER | LOWER
                        ALPHA = numeric-value
                        RESPONSE = variable_summarized by STAT option
                        STAT = FREQ | MEAN | SUM
```

```
      Plot options:     GROUP = variable
                        TRANSPARENCY = numeric-value

HBOX response-variable </options> ;
Box options:
      BOXWIDTH= numeric-value
      CATEGORY= category-variable
      DATALABEL <= variable>
      EXTREME
      FREQ= numeric-variable
      LABELFAR
      MISSING
      PERCENTILE= numeric-value
      SPREAD

Plot options:
      LEGENDLABEL= "text-string"
      NAME= "text-string"
      TRANSPARENCY= numeric-value
      X2AXIS
      Y2AXIS
```

Parts of a Box Plot



`HISTOGRAM` *response-variable* `</options> ;` creates a frequency histogram for the distribution of a numeric variable. When combining images the `HISTOGRAM` statement can only be combined with the `DENSITY` statement.

Histogram options:

```
      FILL | NOFILL     filled or empty bars
      FILLATTRS= style-element | (COLOR= color)  specify bar color
      FREQ= numeric-variable        specify upper limit of frequency axis
      OUTLINE | NOOUTLINE           bars are outlined or not
      SCALE= COUNT | PERCENT | PROPORTION       specify the bar heights
```

Plot options:

```
      LEGENDLABEL= "text-string"
      NAME= "text-string"
      TRANSPARENCY= numeric-value         use when overlaying images.
      X2AXIS                              specifies a second x variable
      Y2AXIS                              specifies a second y variable
```

`DENSITY` *response-variable / TYPE = NORMAL* `;` creates a density curve for the numeric responses variable with the sample mean determining the center of the normal curve.
`TYPE=NORMAL (MU=`*numeric-value* `SIGMA =` *numeric-value*`)` fits a normal density curve "over" the sample data with a hypothesized or imposed mean MU and/or standard deviation SIGMA.

SCATTER X=xvariable Y=yvariable </*options*> ;

SCATTER options:

DATALABEL <= *variable*> displays a label for the data points. If no DATALABEL is specified, the value of the Y variable is the label.

FREQ= *numeric-variable*

Markers used in the plots can be controlled with the following options. In this introduction use the default settings.

MARKERATTRS= *style-element* <(*options*)> | (*options*)

MARKERCHAR= *variable*

MARKERCHARATTRS= *style-element* <(*options*)> | (*options*)

Plot options:

GROUP= *variable*

LEGENDLABEL= "*text-string*"

NAME= "*text-string*"

TRANSPARENCY= *numeric-value*

SERIES X=xvariable Y=yvariable </*options*> ;

SERIES options:

CURVELABEL <= *text-string*>

CURVELABELLOC= INSIDE | OUTSIDE

CURVELABELPOS= MIN | MAX | START | END

DATALABEL <= *variable*>

LINEATTRS= *style-element* <(*options*)> | (*options*)

Plot options:

GROUP= *variable*

LEGENDLABEL= "*text-string*"

NAME= *text-string*"

TRANSPARENCY= *numeric-value*

X2AXIS

Y2AXIS

VBAR *response-variable* </*options*> ;

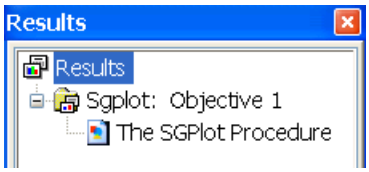
See HBAR statement options.

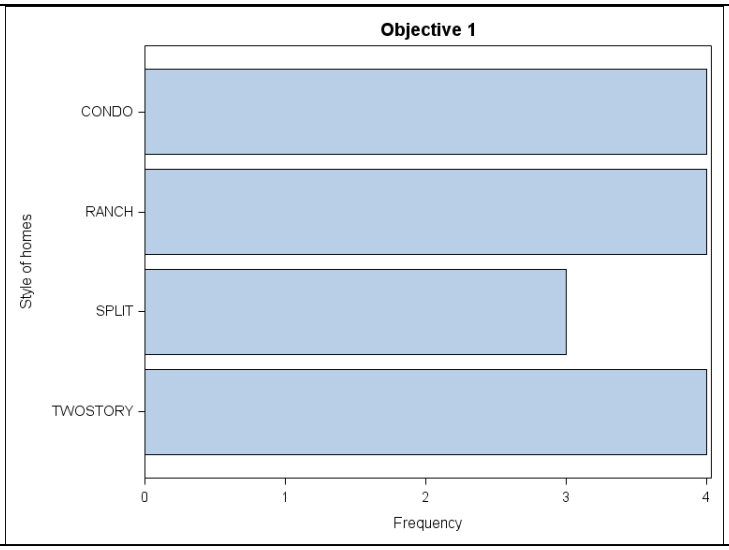
VBOX *response-variable* </*options*> ;

See HBOX statement options.

Objective 1: **Create a horizontal bar chart for the STYLE of homes in the HOUSES data table**.

```
PROC SGPLOT DATA=CLASS.HOUSES ;
HBAR style ;
TITLE 'Objective 1';
RUN;
```





Note that the image produced is a .PNG file. As with ODS Graphics, locate this file in the WORK folder (possibly, C:\Documents and Settings\user) on your computer.

Objective 2*: *Create a horizontal bar chart for STYLE where the bars are only outlined and the bar height is determined by the average selling PRICE of the home.

```
PROC SGPLOT DATA=CLASS.HOUSES ;
HBAR  style / NOFILL    STAT=MEAN    RESPONSE=price ;
TITLE 'Objective 2';
RUN;
```

When the bar heights of two summary variables are similar, the images can be overlayed on a single graph. The category variables in the HBAR (or VBAR) statements must be the same.

Objective 3: Produce overlaid vertical bar chart images for STYLE where the bar heights for one image is the average selling PRICE and for the other image is the average SQFEET. This is an example where the bar heights are <u>not</u> on comparable scales (hence this is an example of a poor graph).

```
PROC SGPLOT DATA=CLASS.HOUSES ;
VBAR style / NOFILL        STAT=MEAN        RESPONSE=price ;
VBAR style / FILL          STAT=MEAN        RESPONSE=sqfeet    BARWIDTH=0.7 ;
TITLE 'Objective 3';
RUN;
```

Objective 4: Modify the code in Objective 3 so that the bar heights are the mean number of baths and the mean number of bedrooms.

```
PROC SGPLOT DATA=CLASS.HOUSES ;
VBAR style / NOFILL        STAT=MEAN        RESPONSE=bedrooms ;
VBAR style / FILL          STAT=MEAN        RESPONSE=baths     BARWIDTH=0.7 ;
```

```
TITLE 'Objective 4';
RUN;
```
Objective 5: Put the BARWIDTH = 0.7 option on the first VBAR statement and rerun.

Objective 6: For the polymers in AGENTS data table, create horizontal box plots of the ph and solids response variables.

```
PROC SGPLOT DATA=CLASS.AGENTS ;
HBOX  ph / CATEGORY = polymer ;
TITLE 'Objective 6';
RUN;
```

Similarly one can produce vertical box plots (VBOX statement) for each polymer.

When the response variable is a numeric variable and the distribution of the variable is to be produced and investigated, then a histogram is appropriate rather than a bar chart.

Objective 7: Produce a histogram for ROBBERY variable in the CRIME data table. Use filled or solid red bars in the histogram.

```
PROC SGPLOT DATA=CLASS.CRIME ;
HISTOGRAM robbery / FILLATTRS = (COLOR=RED);
TITLE 'Objective 7';
RUN;
```

Objective 8: Modify the histogram in Objective 7 by overlaying a normal curve on the histogram. The image in Objective 7 does not suggest normality, and the curve overlay certainly confirms that.

```
PROC SGPLOT DATA=CLASS.CRIME ;
HISTOGRAM robbery / FILLATTRS = (COLOR=RED);
DENSITY robbery / TYPE=NORMAL ;
TITLE 'Objective 8';
RUN;
```

Objective 9: You can overlay multiple plots when the axes for each graph are on the same or similar scales. Examine trends for different types of crimes across the states (using the number assigned to the state).

```
PROC SGPLOT DATA=CLASS.CRIME;
SERIES Y=robbery X=state ;
SERIES Y=assault   X=state ;
SERIES Y=larceny  X=state ;
SERIES Y=burglary X=state ;
RUN;
```

Here burglary and larceny have much greater occurrences. One may argue these should not be on the same graph with the robbery and assault numbers.

Modify this objective by changing the first two SERIES statements to SCATTER statements.

In both images produced in this objective, notice that SAS assigns a legend. The colors, line styles, and symbol choices may not be your preference. With additional information from SAS Help and Documentation, one can find more information about defining legends in the SGPLOT procedure.

The SGPANEL Procedure

The SGPANEL procedure creates multiple graphs in a row, column or lattice (row by column) for one or more classification variables. Within the SGPANEL procedure many of the statements used in SGPLOT are available and these images can be combined in a panel.

PROC SGPANEL Syntax

```
PROC SGPANEL DATA=tablename <options> ;
PANELBY variables </options> ;
KEYLEGEND <"name1" ... "namen" > </options> ;
COLAXIS <options> ;
ROWAXIS <options> ;

HBAR response-variable </options> ;
HBOX response-variable </options> ;
HISTOGRAM response-variable  </options> ;
DENSITY response-variable </options> ;
SCATTER X=xvariable  Y=yvariable </options> ;
SERIES  X=xvariable  Y=yvariable </options> ;
VBAR response-variable </options> ;
VBOX response-variable </options> ;
```

`PANELBY variables </options> ;`

This is the **required** first statement of the SGPANEL procedure. This statement specifies the classification variable(s) for the panel and the type of layout.

Options include:

BORDER | NOBORDER adds or suppresses cell borders

COLHEADERPOS = TOP | BOTTOM | BOTH specifies the location of the column headings in the panel. TOP is the default position. If PANEL=LAYOUT, then this option is not effective.

COLUMNS = n indicates the number of columns in the panel.

ROWHEADERPOS = LEFT | RIGHT | BOTH specifies the location of the row headings. *Default position?* If PANEL=LAYOUT, then this option is not effective.

ROWS = n indicates the number of rows in the panel

LAYOUT = LATTICE | PANEL | COLUMNLATTICE | ROWLATTICE Specify one. LATTICE creates a r x c panel when two classification variables are specified in this statement. PANEL forms an r x c panel with row and column headings available. COLUMNLATTICE arranges the graph cells into a single row, and ROWLATTICE arranges the graph cells into a single column. The default setting is PANEL.

`KEYLEGEND <"name1" ... "namen" > </options> ;`

Options include:

ACROSS = n specifies the number (n) of columns in the legend.

DOWN = n specifies the number (n) of rows in the legend.

You can specify either one or both options. If neither is used, SGPANEL will use a default legend.

BORDER | NOBORDER select one to either place a border around the legend or not. The default setting is BORDER.

POSITION = *value* where *value* is BOTTOM, LEFT, RIGHT, TOP specifies the location or position of the legend. BOTTOM is the default setting.

```
COLAXIS <options> ;
ROWAXIS <options> ;
```

These two statements specify the options for the column axis and the rowaxis.
Here are only a few of the options available:

LABEL = "text string" to supply text for the axis

ALTERNATE supplies reference tick marks for each axis and uses alternate sides for adjacent cells.

GRID creates gridlines at each tick mark of each axis.

VALUES = (values list) specifies the exact values to place tick marks.

The remaining statements are the same as those overviewed here for the SGPLOT procedure. The syntax and format of those statements appears here as well.

Objective 10: Use the AGENTS data table. Create a panel graph of horizontal box plots for pH.

```
PROC SGPANEL DATA=CLASS.AGENTS;
PANELBY polymer ;
HBOX ph ;
TITLE 'Objective 10';
RUN;
```

Objective 11: Modify the PANELBY statement to arrange these boxplots vertically.

```
PANELBY polymer / LAYOUT=ROWLATTICE ;
```

Additionally, investigate the ROWHEADERPOS settings of LEFT and BOTH.

Compare these images with the single image produced by

```
PROC SGPlot DATA=CLASS.AGENTS;
HBOX ph / CATEGORY=POLYMER;
RUN;
```

Objective 12: Change the horizontal box plots to histograms. Overlay a normal curve on each histogram.

```
PROC SGPANEL DATA=CLASS.AGENTS;
PANELBY polymer / LAYOUT=ROWLATTICE;
```

```
HISTOGRAM ph ;
DENSITY ph ;
TITLE 'Objective 12';
RUN;
```

Objective 13: Use the HOUSES data table. For each STYLE of home, create a PRICE by SQFEET scatterplot of the data.

Run this code with different LAYOUT options.

In comparison, the SGPLOT procedure produces a single graph. It may be two images overlapping with each other, but one single set of axes is all that appears in the image.

In SGPANEL multiple sets of axes are in a single panel of graphs. Two graph statements are again allowed, and these will be overlapping images in each cell of the panel. There must be a PANELBY statement when SGPANEL is used. In the SGPANEL overview in SAS Help and Documentation, there are a few very good examples of code and the images produced. See screen capture on next page for how to access the information in SAS Help and Documentation.

For all of the Statistical Graphics procedures, one can also request HTML files using the global commands
ODS HTML;

and

ODS HTML CLOSE;

To locate more information about statements and options for each of the Statistical Graphics procedures covered here, enter the name of the procedure on the Index tab in SAS Help and Documentation, and then go to the ODS Graphics Procedures Guide. You will get many, many syntax options for these and other procedures.