

Creating Bar charts using the proc gplot and annotate function.

Berber Snoeijer, Gerard Hermus Biometric Support, Leiden, The Netherlands

ABSTRACT

The SAS system has a special procedure to produce bar charts. The proc gchart is an easy to use procedure but it has certain limits in freedom. The position of the x-labels, standard deviation-bar setup and the bar width often cannot be specified precisely enough. Difficulties in formatting arise when the bar chart contains both negative and positive values. Rita Tsang (2004) gave a method to make bar plots by using proc gplot and an annotate dataset. We have elaborated her method to be able to make more complicated box plots with finely tuned axes and standard deviation bars. In this paper, we will show how to setup the dataset, produce the annotate dataset and show the possibilities with our method.

INTRODUCTION

The PROC GCHART procedure can be used to create bar charts. However this procedure has limited possibilities. For example, it is not possible to add SD lines to the bars when bars are partly up and partly down. Also it is not possible to add 95% CI lines, other custom lines or stars for significance. Therefore we have created a standard annotate functionality and used this in the PROC GPLOT procedure.

CREATING BAR CHARTS INCLUDING SD LINES USING PROC GPLOT

DATASETS USED

A work dataset is needed with a categorical x-variable (xvar) and a continuous y-variable (yvar). Thereafter summary statistics will be calculated by the categorical x variable including the mean and SD using PROC UNIVARIATE or PROC MEANS. These values are stored in a new dataset. That can be used for further proceedings described below. The variables in this dataset are at least the categorical x-variable, the mean and the SD.

PLOTTING DATASET

The dataset containing the x-variable, the mean and the SD will be used for plotting. However, the beginning of the bar near the x-axis has to be added. This is done by the following code:

```
DATA grwrk2;
  SET grwrk1;
  IF mean>. THEN DO;
    OUTPUT;      /* outputs the original mean */
                /* adding a value slightly above or below the x-axis */
    IF mean < 0 THEN mean=-0.045;
    ELSE mean = 0.045;
    OUTPUT;
  END;
RUN;
```

ANNOTATE DATASET

The annotate dataset is used to create the SD bars. The same dataset as for the plotting dataset is used for this. First, by using the annotate dataset one moves to the starting point of the line which is at the end of the bar. As this end of the bar is equal to the mean, the starting point is also equal to the mean. The end point is equal to the mean plus or minus the SD. If the mean is negative we want the line to go wond and if the mean is positive we want the line to go up. The following code is used for this:

```

DATA sdanno;
  SET grwrk1;
  starty=mean;
  IF mean < 0 THEN stopy=mean-sd; /* define start of the SD line */
  ELSE stopy=mean+sd; /* define end of the SD line */
  function='MOVE'; /* this depends on neg or pos mean */
  color='BLACK'; /* move to start location */
  xsys='2'; /* draw black line (after move) */
  ysys='2'; /* draw line relative to x-data in chart */
  x=xvar; /* draw line relative to y-data in chart */
  y=starty; /* startig point of the line */
  OUTPUT;
  function='DRAW'; /* draw line */
  y=stopy; /* end point of the line */
  OUTPUT;

```

In most cases, one likes to add a small horizontal end line at the end of the SD line. Therefore, 2 extra rows are added to the annotate dataset: The exact amount to be subtracted and added to the x value depends of course on the distribution of the x variable used.

```

  x=x-0.2;
  OUTPUT;
  x=x+0.4;
  OUTPUT;
RUN;

```

Another option would be to change the x settings of xsys to 1 and add and subtract a percentage of the output area. In this case you are not dependent on the x-variable for the width of your horizontal end line.

```

  xsys='1';
  x=x-2;
  OUTPUT;
  x=x+4;
  OUTPUT;
RUN;

```

CREATING THE GRAPHS

PROC GPLOT is intended to be used for creating line graphs. To make boxes instead, the symbols of the lines to be created has to be changed. The most important is that the interpol option (i) will be set to BOX00F. The box option is originally used for creating box plots but can also be used for this purpose. Adding the two zero's results in a box that starts at the first datapoint (near the x-axis) to the last datapoint (the actual mean). The F fills the box with the color specified at CV. The color of the lines around the box are specified with the option CO. The bar width is specified with the option BWIDTH.

```
symbol11 INTERPOL=BOX00F co=DARKYELLOW cv=DARKYELLOW bwidth=8.2;
```

The actual creation of the graphs is simple thereafter. Just use the standard PROC GPLOT procedure and specify your annotate dataset.

```

PROC GPLOT DATA = grwrk2 ANNOTATE=sdanno;
  PLOT mean*xvar ;
RUN;
QUIT;

```

You can create different colors of bars by putting them in different categories (cat) and adding the corresponding variable to the GPLOT statement.

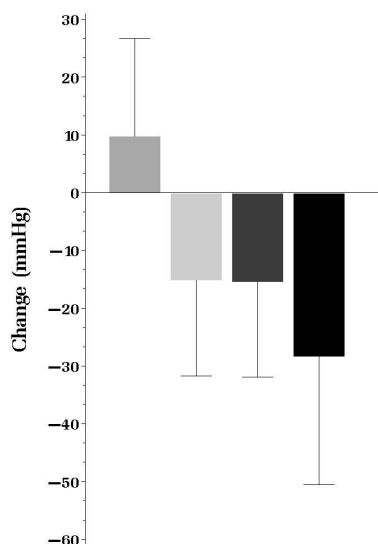
```
PROC GPLOT DATA = grwrk2 ANNOTATE=sdanno;
  PLOT mean*xvar=cat ;
RUN;
QUIT;
```

CREATING BAR CHARTS INCLUDING CI OR MIN-MAX VALUES

With the approach presented above it is very simple to create bar charts including a CI or min-max values. This is done by calculating the CI values and/or MIN/MAX values as well with the PROC UNIVARIATE procedure. The values then have to be used in the annotate dataset instead of the SD. See the following code:

```
DATA sdanno;
  SET grwrk1;
  starty=mean; /* define start of the CI line */
  IF mean < 0 THEN stopy=mean-CImin; /* define end of the CI line */
  ELSE stopy=mean+CImax; /* this depends on neg or pos mean */;
```

An example of a graph created by this method is presented below:



CONCLUSION

In this paper a simple and efficient method is presented to create bar charts including positive or negative SD bars. This method can be useful in case SD bars have to be presented for negative values or instead of SD bars, one wants to show CIs, min/max values etc..

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Berber Snoeijer
 Biometric Support
 Schipholweg 78
 2316 XD Leiden

Work Phone: +31 71 5721828
Fax: + 31 71 5765040
Email: bsnoeijer@biometricsupport.nl
Web: www.biometricsupport.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.