

Bars and Lines: A Quick Introduction To PROC GBARLINE

Garland D. Maddox, Jr., Regions Bank, Birmingham, AL

ABSTRACT

Ever had a need to create a vertical bar chart and then overlay it with one or more plots? PROC GBARLINE provides a convenient and elegant method of doing just that. This paper provides a brief introduction to PROC GBARLINE with an emphasis on using some of the enhancements introduced in SAS® Version 9.2.

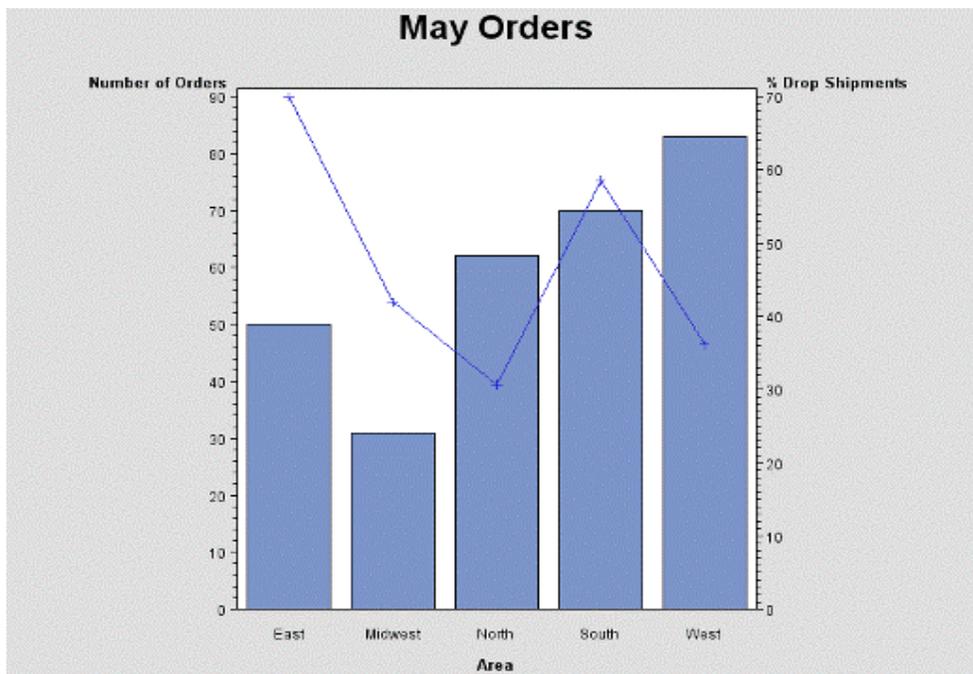
INTRODUCTION

PROC GBARLINE, a SAS/GRAPH® procedure introduced in SAS Version 9.1.3, provides the functionality described in its name. (Berryman and Chapman) It produces graphs which feature both vertical bars and overlaid plots. This paper provides a brief introduction to the procedure and demonstrates some basic features and functions. Enhancements to the procedure were introduced in Version 9.2, and some of these will be examined as well. Lastly, two issues which are somewhat unique to the procedure are discussed along with possible solutions.

Before launching into the basics of the procedure, it would probably be in order to examine why we would even want to produce a graph of this type. There are several uses for a bar-line graph. They make it possible to display relative magnitudes and to provide comparisons for analysis. Bar-line graphs also allow for the examination of the contribution of various parts of a whole, and are capable of displaying a long series of data which can be used to analyze trends and patterns. In particular, managers may have several metrics of interest which may be tied together to confirm or refute some basic assumptions.

THE BASICS

As with PROC GCHART, PROC GBARLINE produces a vertical bar chart based on the values of a chart variable and a response variable. The overlaid line uses the same chart variable and a different response value. Consider this basic bar-line graph of a set of orders which are characterized as to the sales area. In addition, we have an interest in the percentage of drop shipments per area.



Drop shipments are orders which are fulfilled by shipping directly from the manufacturer to the customer as opposed to shipping from inventory on hand. Samples of the data are included in the appendix. The data behind this particular graph was pre-summarized in order to calculate a percentage of drop shipments per area. One feature that readily stands out in the graph is that it has three axes. The horizontal axis shows the categories of values that are based on the chart variable. In this instance, the sales areas are the chart variable, and they are character values which are inherently discrete. Chart variables may also be discrete numeric values, such as years, or continuous numeric variables, such as ages. In the case of continuous numeric variables, the procedure will generate midpoints which represent ranges of values and will calculate the appropriate number of ranges to display. Hence, the chart variable is often referred to as the midpoints, even though the actual values to be charted are discrete.

The left vertical, or response, axis shows the value of the bar statistic, and the right vertical axis shows the value of the line statistic. The procedure handles the scaling of the response axes to accommodate the data.

The following code will generate the basic bar-line graph shown above:

```

/* GBAR is the set of individual orders */
/*Pre-summarize to calculate the percentage of drop shipments */
proc summary data=gbar nway;
class area ;
var sales_price drop_ship;
output out=gbarz2 (rename=(_freq=tot_orders)) sum(sales_price)=revenue
sum(drop_ship)=drop_ship;
run;

data gbarz3;
set gbarz2;
perc_drop_ship=(drop_ship/tot_orders)*100;
run;

title 'May Orders';
goptions device=gif;

ods listing close;
ods html path='c:\sesug 2010\cc' file='gbar3.html';
/* Run the procedure to produce the graph */
proc gbarline data=gbarz3;
bar area / sumvar=tot_orders;
plot / sumvar=perc_drop_ship;
label area='Area' tot_orders='Number of Orders' perc_drop_ship='% Drop Shipments';
run;
quit;

ods html close;
ods listing;

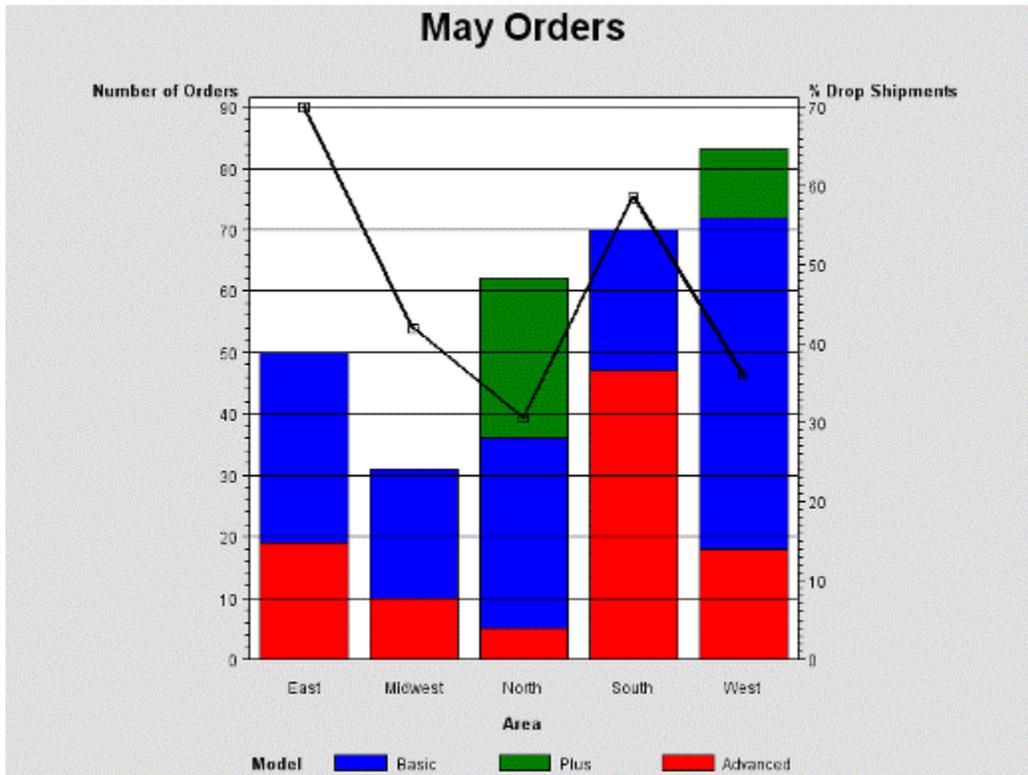
```

There are several items to note. PROC GBARLINE has relatively few procedure level options. Most of the work is done in the BAR and PLOT statements. The SUMVAR option in the code above specifies the response variable for both the bars and the plots. In this case, we want to see the distribution of the orders across areas as depicted by the length of the bars, and the percentage of drop shipments for the corresponding areas. Bars are displayed in alphabetical order for character chart variables in the default scenario. The PLOT statement must always follow the BAR statement. PLOT statements are considered optional and if not included, the procedure will simply produce a bar chart similar to that produced by PROC GCHART. Plot response variables must be numeric. The default statistic for the bars is frequency, and the default statistic for the plot is the sum. The TYPE= option can be used to specify the statistics to be displayed if SUMVAR is not used. For the current SAS versions, PROC GBARLINE can not produce graphics with the JAVA device type.

While relatively few statements are required to produce a basic bar-line chart, it is possible to greatly enhance the chart with the use of standard SAS/GRAPH statements and the many options included for both BAR and PLOT statements. We will examine some of them in the following section.

ENHANCING YOUR BAR-LINE GRAPH

Several enhancements were introduced to PROC GBARLINE in Version 9.2. The SUBGROUP option was added to the BAR statement so that stacked bar graphs could be produced, just as with PROC GCHART. Our data set includes a "Model" variable to indicate whether the ordered product is a Basic, Plus, or Advanced model. We can generate a stacked bar-line graph using that variable by specifying SUBGROUP=model in the BAR statement options. Both BAR and PLOT statements include a wide variety of options that enhance appearance and analytical capabilities. In addition, all of the familiar SAS/GRAPH statements can be used to enhance and modify the graph. For example, the PATTERN statements control bar colors and the SYMBOL statement controls the appearance of the line plot. Legends can be added for both bars and plots. Considerable customization can be done using AXIS statements. Of course, our old friend ANNOTATE can be used to provide supplemental annotations, both at the procedure level and in the BAR and PLOT statements. Here is our basic graph after adding the bar stacking and a few additional options.



In order to display the Model data in the graph, it is necessary to include the Model classification variable in the PROC SUMMARY step. Then, we can calculate the drop shipment percentages, keeping in mind that these represent the Area and Model combinations.

```
proc summary data=gbar nway;
class area model;
var sales_price drop_ship;
output out=gbarz (rename=( _freq_=tot_orders)) sum(sales_price)=revenue
sum(drop_ship)=drop_ship;
run;

data gbarz2;
set gbarz;
perc_drop_ship=(drop_ship/tot_orders)*100;
run;

title 'May Orders';
```

Having re-summarized the data, we can add various statements and options.

```
goptions device=gif;
pattern1 color=red;
pattern2 color=blue;
pattern3 color=green;
symbol1 interpol=join height=10pt value=square line=1 width=2 cv=black ci=black;
legend1 order=('Basic' 'Plus' 'Advanced');

ods listing close;
ods html path='c:\sesug 2010\cc' file='gbar2.html';

proc gbarline data=gbarz2;
bar area / sumvar=tot_orders
           subgroup=model
           legend=legend1
           frame
           cref=black
           autoref;
plot / sumvar=perc_drop_ship type=mean freq=tot_orders;
label area='Area' tot_orders='Number of Orders' perc_drop_ship='% Drop Shipments';
run;
quit;

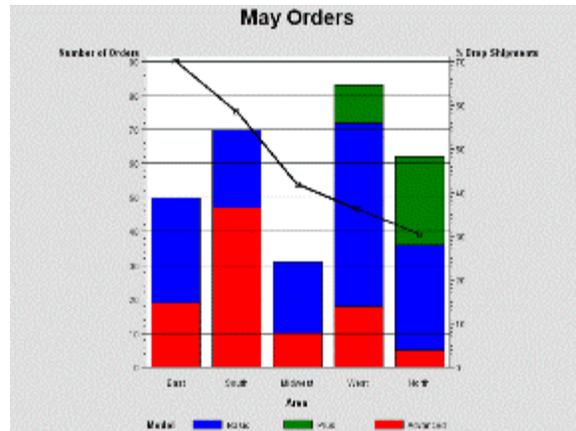
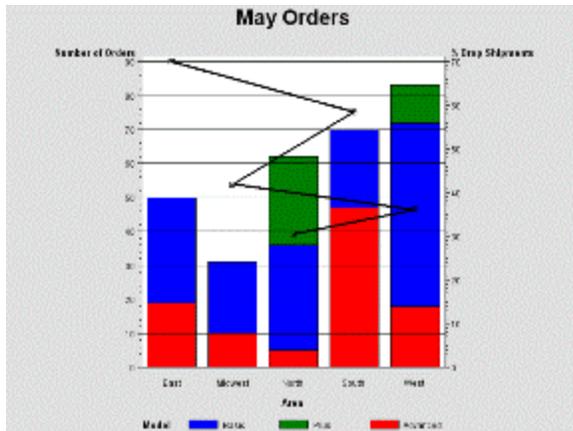
ods html close;
ods listing;
```

These are just a few of the many options that can be used with PROC GBARLINE. Legend, frame, cref, and autoref are added here to produce the legend and add the reference lines and line color. One significant difference is that in the PLOT statement, the mean statistic is requested and a “freq” option. This permits the drop shipment percentages to stay in sync with the bar chart portion of the graph. When we summarized to the area and model area, the calculated percentages are relevant to those combinations. So, we want to calculate a weighted mean to accurately calculate the area percentages.

TWO PROC GBARLINE ISSUES

There are at least two issues to consider when using PROC GBARLINE. The first is the issue of synchronizing the bar and the line data, as discussed in the previous section. If the PLOT statement is not changed to “type=mean freq=tot_orders”, then the percentages would be summed. When summarization is required, extra care needs to be taken that the two displayed statistics are, in fact, comparable.

The other issue is that the bars are normally ordered by the value of the chart variable. Two options, “Ascending” and “Descending”, in the BAR statement can be used to order the bars based on the value of the bar statistic. The bars would then be ordered by shortest to longest if “ascending” is used. If the “descending” option is used, then bars are ordered by longest to shortest. There are similar options in the PLOT statement. However, the bars are not re-ordered by adding the option in the PLOT statement. The plot points are joined in an ascending or descending manner, whichever option is used. This can result in a somewhat “difficult to interpret” graph, as pictured in the figure below on the left. However, the BAR statement features a “midpoints” option which allows the explicit ordering of the midpoints. With just one graph, we can explicitly order the bars. However, it can also be done programmatically by summarizing and sorting the data by the plot response data values. Then, the corresponding chart variable values can be loaded into a macrovariable which could be referenced in the “midpoints” option list. One method of doing this is to use a macro such as the CR8_IN_LIST macro (Sanders, 2009) which must be modified to output a list of chart variable values for a specific graph. The result is a graph with bars ordered by the value of the plot response values., as pictured below on the right. Notice that not only is the plot drawn differently, but the bar order has been changed as well.



CONCLUSION

In summary, bar-line graphs can be a useful analysis tool, and PROC GBARLINE provides a powerful tool for constructing them. The rich option set along with the power of SAS/GRAPH statements can help you go a long way toward creating just the graph that you want.

REFERENCES

Berryman, Pat, and Chapman, Jonathan. 2008. "Modernize Your Business Reports Using ODS and SAS/GRAPH®: A Case Study from SAS® 8.2 to SAS® 9.2." Proceedings of SAS Global Forum 2008. Paper 325.

Sanders, Steve. 2009. "Cr8_In_List – Easy Table Driven Programming." The Proceedings of the SouthEast SAS Users Group, Birmingham, AL, 2009

SAS Online Doc® Version 9.2. Copyright © 2008. SAS Institute Inc, Cary, NC, USA.

SAS System Help for Version 9.2. Copyright SAS Institute Inc., Cary, NC, USA.

RECOMMENDED READING

Tsang, Rita. 2006. "Creating Bar Charts and Scatter Plots on the Same Page, Using PROC GPLOT or PROC GBARLINE in SAS®." The Proceedings of the Thirty-first SAS Users Group International. Paper 175.

CONTACT INFORMATION

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

Garland D. (David) Maddox, Jr.
 Regions Bank
 RCS-3
 250 Riverchase Parkway East
 Birmingham, AL 35244
 205-560-6339
 Garland.MaddoxJr@Regions.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.

APPENDIX

Order Data – Partial Listing – Data Set GBAR

| Order Number | Area | Sales Price | Model | Color | Drop Ship |
|--------------|-------|-------------|----------|-------|-----------|
| 1100 | West | 125 | Basic | Blue | 1 |
| 1101 | West | 155 | Advanced | Blue | 0 |
| 1102 | West | 125 | Basic | Green | 0 |
| 1103 | West | 125 | Basic | Red | 0 |
| 1104 | East | 155 | Advanced | Blue | 1 |
| 1105 | East | 125 | Basic | Red | 0 |
| 1106 | North | 125 | Basic | Blue | 0 |
| 1107 | South | 155 | Advanced | Red | 1 |
| 1108 | South | 125 | Basic | Blue | 0 |
| 1109 | South | 155 | Advanced | Green | 1 |
| 1110 | North | 135 | Plus | Blue | 1 |
| 1111 | West | 125 | Basic | Blue | 1 |
| 1112 | West | 155 | Advanced | Blue | 1 |
| 1113 | West | 125 | Basic | Green | 0 |
| 1114 | West | 125 | Basic | Red | 0 |
| 1115 | East | 155 | Advanced | Blue | 1 |
| 1116 | East | 125 | Basic | Red | 1 |
| 1117 | North | 125 | Basic | Blue | 0 |
| 1118 | South | 155 | Advanced | Red | 1 |
| 1119 | South | 125 | Basic | Blue | 0 |
| 1120 | South | 155 | Advanced | Green | 1 |
| 1121 | North | 135 | Plus | Blue | 0 |
| 1122 | West | 125 | Basic | Blue | 1 |
| 1123 | West | 155 | Advanced | Blue | 1 |
| 1124 | West | 125 | Basic | Green | 0 |

Drop Ship 0= Not drop shipped, fulfilled from inventory

1= Drop shipped, fulfilled directly from the manufacturer