CREATING A CUMULATIVE AREA PLOT

Thomas Miron, Miron InfoTec

A cumulative area plot represents a total as the cumulative result of two or more factors. Cumulative area plots can be created using SAS/GRAPH® software when the input SAS® data set is structured correctly. SAS/GRAPH options give you extensive control over the appearance of your plot.

What is a cumulative area plot?

A cumulative area plot shows a total and the components of that total. The magnitude of each component is shown not by a distance from the horizontal axis, but by the distance from the component below it. The graphic below is a cumulative area plot.

Data for cumulative plots

The SAS data set LIB1.UTILYEAR is shown on the next page. Each observation of LIB1.UTILYEAR holds the gas, electric, and telephone utility expenses for a year. This data set is used as the basis for the example plot. A cumulative area plot for this data is made up of three areas representing the expenditures for gas, electricity, and telephone.
The LIB1.UTILYEAR data set does not contain variables holding cumulative expenditure totals or the total utility costs for the year. In order to create the cumulative plot, an intermediate plotting data set is created that holds cumulative totals. It is these cumulative totals that are plotted instead of the actual values for each utility category from the original data set LIB1.UTILYEAR.

Creating the plot

In this example, two SAS steps are used to create the cumulative area plot. First, a DATA step (program lines 1-16) reads the LIB1.UTILYEAR data set shown above. The DATA step calculates cumulative totals and writes a temporary data set called PLOTIT. PLOTIT is shown at the end of this paper.

A new variable called SOURCE is created. The values assigned to SOURCE are important because they determine the order in which the plot lines will be drawn. In the PROC GPLOT step (program lines 36-43), SOURCE is used as the third-variable so there is one line for each of the three values of SOURCE. The lines are drawn in ascending order of these values. The line for "1_Gas" is drawn first, the line for "2_Elec" second, and the line for "3_Tele" third. The plot areas are also filled in this order. First the area between the horizontal axis and the line for "1_Gas" is filled. Then the area between the first line and the second line is filled. Finally, the area between the second and third lines is filled. The key point is that each line drawn must be higher than the one drawn before it or its area fill will be obscured.

In the example, the relationship between the relative values of the vertical axis variable and the sort sequence of the third-variable is enforced by using "1", "2", and "3" to begin the values of SOURCE as the value of COST increases by accumulating expenses for each utility category.

A LEGEND statement is used to display "Gas", "Electric", and "Communications" instead of "1_Gas", "2_Elec", and "3_Tele."
Restructuring DATA step

The DATA step restructures the utility data to hold cumulative totals.

SYMBOL statements

Areas can only be filled under lines so SYMBOL statements with INTERPOL=JOIN are used. One for each of the three cost categories plotted.

AXIS statement

The AXIS statement defines an appropriate axis label for the vertical plot axis. The AXIS setting is assigned with the VAXIS statement, program line 40.

Third-variable plot request

The third-variable form plot request is used to generate a separate line for each value of SOURCE.

LEGEND= option

The LEGEND=LEGEND1 option applies the LEGEND1 settings, program lines 29-32, to the plot legend.

Keep plotting variables

Only the variables needed to generate the plot are kept.

Read the original data set

LIB1.UTILYEAR is used as input.

Accumulate totals then output

The COST variable is used to accumulate costs from each category. An observation is output after each accumulation. The values of SOURCE are set in an ascending sort sequence to match the ascending value of COST as individual categories are accumulated.

TEMPORARY plotting data set

PLOTIT is a temporary SAS data set holding the restructured plot data.

Keep plotting variables

Only the variables needed to generate the plot are kept.

Read the original data set

LIB1.UTILYEAR is used as input.

Accumulate totals then output

The COST variable is used to accumulate costs from each category. An observation is output after each accumulation. The values of SOURCE are set in an ascending sort sequence to match the ascending value of COST as individual categories are accumulated.

Temporary plotting data set

PLOTIT is a temporary SAS data set holding the restructured plot data.

Keep plotting variables

Only the variables needed to generate the plot are kept.

Read the original data set

LIB1.UTILYEAR is used as input.

Accumulate totals then output

The COST variable is used to accumulate costs from each category. An observation is output after each accumulation. The values of SOURCE are set in an ascending sort sequence to match the ascending value of COST as individual categories are accumulated.

Three PATTERN statements assign a unique color to each of the three plot areas.
Cumulative Plot Results

Axis label
The vertical axis label “Total” is set in the AXIS statement, program line 27, and assigned with the VAXIS=AXIS1 option.

Utility Cost Components

Cumulative area plot
The top line shows the grand total utility cost for each year. The width of each area represents the cost of each component identified in the legend.

Legend
The legend is generated automatically when you use the third-variable form plot request. Legend text and pattern swatches are set in the LEGEND statement, program lines 29-32.

Intermediate Plotting Data Set

PLOTIT data set
PLOTIT is the intermediate data set generated from LIB1.UTILITY. PLOTIT is used to create the cumulative area plot.

<table>
<thead>
<tr>
<th>PLOTIT DATA SET</th>
<th>OBS</th>
<th>YEAR</th>
<th>COST</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1990</td>
<td>5600</td>
<td>1.Gas</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1990</td>
<td>6850</td>
<td>2.Ele</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1990</td>
<td>9300</td>
<td>3.Tel</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1991</td>
<td>6250</td>
<td>1.Gas</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1991</td>
<td>7200</td>
<td>2.Ele</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1991</td>
<td>10200</td>
<td>3.Tel</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1992</td>
<td>6100</td>
<td>1.Gas</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1992</td>
<td>7125</td>
<td>2.Ele</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1992</td>
<td>10525</td>
<td>3.Tel</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1993</td>
<td>6450</td>
<td>1.Gas</td>
<td></td>
</tr>
</tbody>
</table>

SAS and SAS/GRAPH are registered trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration.

947