KEY FEATURE FOR CREATING DECORATIVE INFOGRAPHS

The key feature of the SGPLOT procedure that is used to create the decorative graphs presented in this paper is the ability to use images to define markers using the SYMBOLIMAGE statement. This statement was introduced with the first maintenance release for SAS® 9.4.

```
SymbolImage name=symbolname image="file-name" </ options>;
```

Supported arguments are:

- HOFFSET=*offset*. The horizontal offset to the anchor point to be placed at the (x, y) location.

- ROTATE=*number*. Angle of rotation, in degrees.

- SCALE=*number*. Scale factor for the image. This is used to map to default size of the marker.

- VOFFSET=*offset*. The vertical offset to the anchor point to be placed at the (x, y) location.

We can use this statement to define a symbol with a name specified by the name=symbolname parameter using an image file on the file system specified in the image=' *file-name*' parameter. Then, we can use this symbol to draw markers in the graph using its assigned name just like any other symbol, like Circle or circleFilled. We can also use such symbols for representing groups.

Normally, images found on the web or those taken using a smart phone camera are rectangular in shape, and every pixel in the image is opaque, like the example shown in Figure 3. The image shown in Figure 4 has been processed to make all the background pixels (not part of the object) as transparent. You can do this by using an image processing software such as Photoshop.

The red pixels in Figure 3 are transparent in Figure 4, so when the image is pasted on a background, the background shows through the transparent pixels. I always use transparent images in these examples for reasons shown later in this paper.
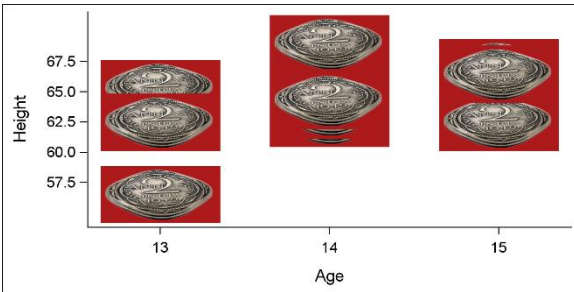

Figure 3. Image with Fully Opaque Background   **Figure 4. Image with Transparent Background**
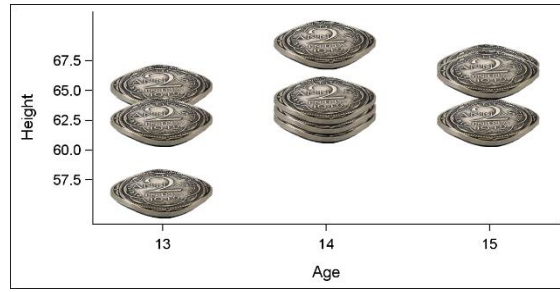
Now, let us define a symbol called coinOpaque using an image from the file system as shown in Figure 3 using the SYMBOLIMGE statement. Next, we create a scatter plot from sashelp.class data and specified the use of this symbol as the marker using the MARKERATTRS option. The graph using this opaque image symbol is shown in Figure 5.

```
proc sgplot data=sashelp.class noborder noautolegend;
  symbolimage name=coinOpaque image="&coinOpaque";
  scatter x=age y=height / markerattrs=(symbol=coinOpaque size=100);
  xaxis offsetmin=0.15 offsetmax=0.15;
  yaxis offsetmin=0.15 offsetmax=0.15;
run;
```

We can create the same graph using the transparent image from Figure 4. The graph using this transparent image symbol is shown in Figure 6. I decided to use this image of the vintge 1942 "Two Annas" coin from British India because of the interesting shape and beautiful engravings on the face.

Figure 5. Graph Using Opaque Symbols

Figure 6. Graph Using Transparent Symbols

Note that in the graphs in Figure 5 and Figure 6, markers are drawn for every observation in the data set. I have set the marker size to 100 pixels. The graph in Figure 5 uses the opaque image for the symbol. So we see the full rectangular image for each marker, and where the markers are close together, the markers are obscured by the background.

In the graph in Figure 6, I have used the transparent image for creating the symbol. Now, only the non-transparent pixels of the symbol are drawn. Whatever was in the graph before a symbol was placed shows through the transparent portion of the symbol. This creates a nice overlap between the coins, including the short stack of three coins in the middle.

Using transparent images as symbols for markers lets you use these as shapes or masks or skins.

USING MARKERS AS SHAPES

When we use transparent images in which the pixels outside the shape of the object in the image are transparent as in Figure 6, we can use markers as shapes for our graphs. Let us review a couple of examples. Some options in the following code are thinned to fit the space. See resources for a link to download the program ZIP file.

COIN STACK BAR CHART

Let us go back to the Coin Stack bar chart to see how we can create this graph using the feature I described above. To start, we have a data set of Revenues by Year as shown in Figure 7. The bar chart shown in Figure 8 can be easily created using the VBAR statement of the SGPLOT procedure as follows:

Obs	Year	Resp
1	2011	8
2	2012	14
3	2013	3
4	2014	12

Figure 7. Data Set by Year

```
title 'Revenues (Millions) by Year';
proc sgplot data=Bar noborder noautolegend;
  vbar year / response=resp dataskin=pressed datalabel);
  xaxis display=(noticks noline nolabel) integer;
  yaxis display=(nolabel noticks noline) integer grid;
run;
```

Now, in order to use the image of the coin as a marker to create the coin stack, we need to use the SCATTER statement instead of the VBAR statement.

```
proc sgplot data=bar noborder noautolegend;
  scatter x=year y=resp / markerattrs=(symbol=circlefilled size=60)
          filledoutlinedmarkers dataskin=gloss;
  text x=year y=resp text=resp / strip position=top backlight=0.75;
  xaxis display=(noticks noline nolabel) offsetmin=0.15 offsetmax=0.15;
  yaxis display=none offsetmin=0.2 offsetmax=0.2;
run;
```

Since we use the same data shown in Figure 7, we will only get one marker for each observation, as shown in Figure 9. We also used the TEXT statement to display the response value on top of each marker per year. XAXIS and YAXIS statements are used to clean up the graph. The results are shown in Figure 9.
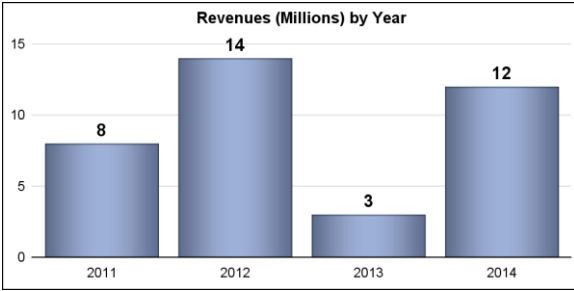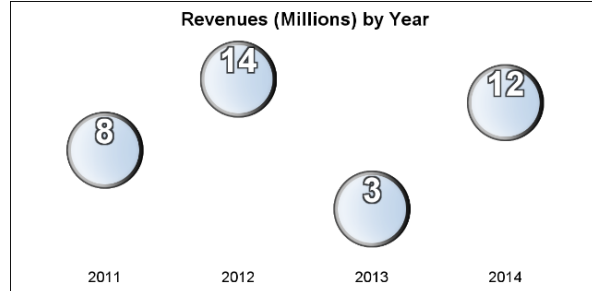
3

Figure 8. Bar Chart of Revenue

Figure 9. Scatter Plot of Revenue

What we really need is a stack of markers for each year, each stack having the number of markers that add up to the revenue. The first stack needs eight markers at values 1 to 8, and so on. Therefore, we need to process the data to create multiple observations per year, as shown in Figure 10.

```
data Coins;
  set bar;
  YVal=resp+1;
  do Val=1 to Resp by 1;
    output;
  YVal=.;
  end;
run;
```

In the code above, we also compute another variable YVAL that has a value just above the revenue for each year. Only one value per year is generated, and all other values are set to missing. The variable YVAL is used in the TEXT statement to display the response value for each year above the stack.

We use this new data set shown in Figure 10 with the scatter plot to create the stack of markers, as shown in Figure 11.

Now that we have a stack of markers, the next step is to replace each marker with a coin. We create a new coin symbol called Indian using the SYMBOLIMAGE statement. We then use this symbol in the MARKERATTRS option in the SCATTER statement with size=110. Now, a coin is displayed in place of each marker in the stack, as shown in Figure 12.

Obs	Year	Val	YVal
1	2011	1	9
2	2011	2	.
3	2011	3	.
4	2011	4	.
5	2011	5	.
6	2011	6	.
7	2011	7	.
8	2011	8	.
9	2012	1	15
10	2012	2	.

Figure 10. New Data Set

```
title 'Revenues (Millions) by Year';
footnote j=l h=7pt italic 'Indian 2 Annas Coin from 1942';
proc sgplot data=Coins noborder noautolegend;
  symbolimage name=Indian image="&Indian";
  scatter x=year y=val / markerattrs=(symbol=Indian size=110);
  text x=year y=YVal text=resp / strip position=top backlight=0.75;
  xaxis display=(noticks noline nolabel) offsetmin=0.15 offsetmax=0.15;
  yaxis display=none offsetmin=0.2 offsetmax=0.2;
run;
```

For this to work, the image of the coin needs to have a bit of a perspective. It is best that the image itself has a perspective view. If not, you could distort the image a bit to meet your needs.
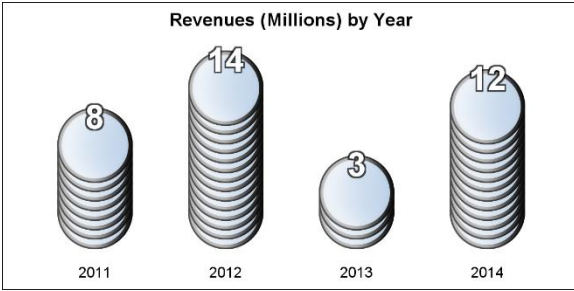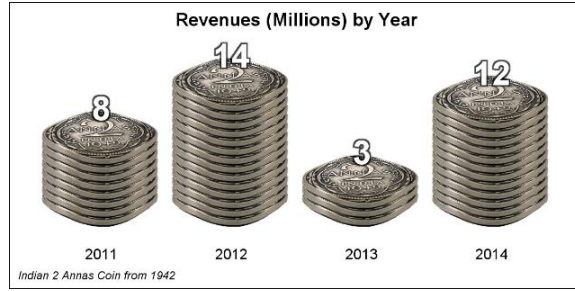
Figure 11. Scatter Plot of Revenue Data

Figure 12. Scatter Plot with Coin Markers

Finally, we use the JITTER option in the SCATTER statement to create stacks that are not perfectly stacked and are a bit more realistic, as shown in Figure 13. We need to create the number of observations in the stack that correspond to the apparent thickness of the coin in the image. The Somalian Silver Shilling is a bigger, thicker coin. Using that coin, we need more spacing between each marker to look right, as shown in Figure 14.
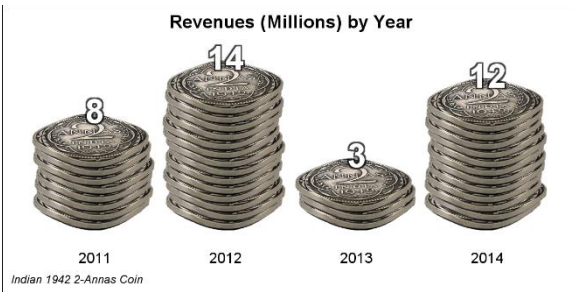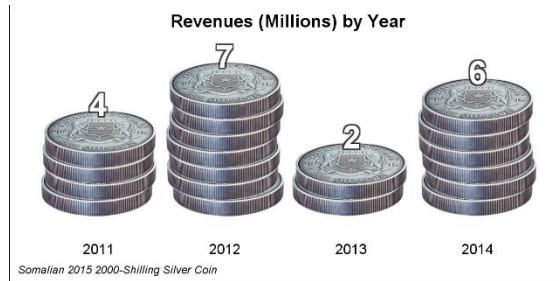

Figure 13. Coin Stack with Jitter

Figure 14. Coin Stack with Thicker Coin

COLLEGE READINESS GRAPH

Let us review another example of using shapes to create decorative infographs.

Recently, there was a request to create a graph comparing the college readiness of students from a local school with the state average. The data for the graph is shown in Figure 15. From a perspective of providing the most effective way to make the comparison, the cluster-grouped graph shown in Figure 16 would be best. It is very easy to compare the performance of students from the local schools with state averages when the results are displayed side by side for each category.

Obs	Value	School	Level
1	100%	This School	Start 9th. Grade
2	72%	This School	Graduate. High School
3	45%	This School	Enter. College
4	24%	This School	Continue in. College
5	100%	State	Start 9th. Grade
6	80%	State	Graduate. High School
7	50%	State	Enter. College
8	31%	State	Continue in. College

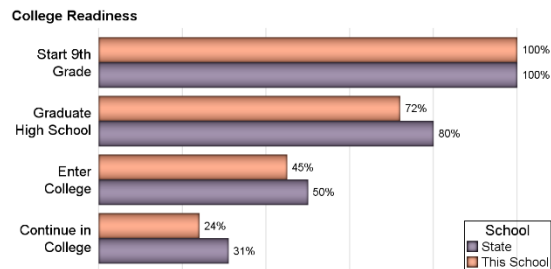**Figure 15. Data for College Readiness**

Figure 16. Graph of College Readiness

Having addressed the issue of effective graphs, let us focus on the needs of this particular user to create a graph that catches the eye of the reader. The desired graph is shown in Figure 18. How would we create a graph like this using the techniques discussed so far?

The first step is to note that the response values are in percent, from 0 to 100. So, we can create an observation for every 10% of the response value. An observation is created for the value This School and another for the value State. We will use the markers shown in Figure 17 to draw the graph. If the value

for the marker is a multiple of 10%, we display it with the full-colored marker (group value A1 or B1). If the marker represents a portion of the 10%, then we use the partial-colored markers with group value of A2 and B2. For the values >response and <100%, we use the gray marker with group value of C1.
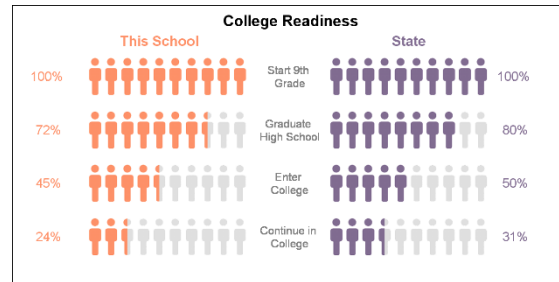

Figure 17. Icons Used in the Graph

Figure 18. Graph of College Readiness

The icons for the symbols for each of the group values A1, A2, B1, B2, and C1 are shown in Figure 17. As in the previous example, we use five SYMBOLIMAGE statements to define the symbols for each group. For this example, we use the STYLEATTRS statement to insert these symbols into the list of group symbols to draw the graph. A separate XAXISTABLE statement is used to display the headers, and the TEXT statement is used to display the values on left, middle, and right of the graph. The SGPLOT code snippet is shown below. See resources for a link to download the program ZIP file. The %RGBHEX macro created by Perry Watts (Watts 2002) is used for setting colors.

```
proc sgplot data=Readiness4 noborder noautolegend;
  symbolimage name=A1 image="&file1";
  symbolimage name=B1 image="&file2";
  symbolimage name=C1 image="&file3";
  symbolimage name=A2 image="&file4";
  symbolimage name=B2 image="&file5";
  styleattrs datasymbols=(A1 A2 C1 B1 B2)
            datacontrastcolors=(%rgbhex(254, 145, 104)
                                %rgbhex(130, 109, 146));
  xaxistable text / x=xlbl colorgroup=text location=inside position=top;
  scatter y=level x=val / group=group markerattrs=(size=40);
  text x=x0 y=level text=valueLblA / textattrs=(color=%rgbhex(254,145,104));
  text x=x270 y=level text=valueLblB /textattrs=(color=%rgbhex(130,109,146));
  text x=x130 y=level text=level / splitchar='.' splitpolicy=splitalways;
  xaxis display=none   max=2.7 offsetmin=0.05 offsetmax=0.0;
  yaxis display=none reverse splitchar='.' offsetmin=0.1
       fitpolicy=splitalways splitjustify=center;
run;
```

USING MARKERS AS MASK AND SHAPE

When we use transparent images in which the pixels outside the shape of the object in the image are transparent (as in 4) we can use markers as shapes for our graphs. We did that in the graphs shown in Figure 13, Figure 14, and Figure 18.

When the outer pixels are opaque (usually of a color that matches the graph background) and the inner pixels are transparent, the marker can be used as a mask. Figure 19 shows an image of a cup in which the pixels inside the cup are all transparent. When this image is pasted on a red background, only the red pixels in the middle show through, while the other pixels are overdrawn by the opaque pixels of the cup image, as shown in Figure 20. We can use this technique to shape the features of a graph.
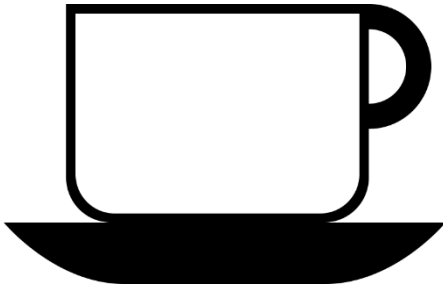
6

Figure 19. Image Used as a Mask

Figure 20. Image Mask on Red Background

COFFEE RECIPES GRAPH

Let us use the mask technique to create a graph showing the ingredients in a cup of coffee. In this example, I want to show you the recipes for four different popular coffee beverages: Espresso, Macchiato, Café Latte, and Café Mocha.

We start with a simple bar chart showing the ingredients and proportions for each beverage. Starting with Espresso only, we mix Espresso plus an equal portion of Milk Foam to make a Macchiato. Then, for Café Latte, we use 1 part Expresso and 2 parts Steamed Milk, topped with 1 part Milk Foam. Lastly for Café Mocha, we use 1 part Expresso, 1 part Chocolate, and 1 part Steamed Milk, topped with 1 part Milk Foam.

We create a data set as shown in Figure 21. Each beverage (Name) uses the required ingredients (Group) with a value. Starting from Y=0, these are added to create the Low and High values that are used by a high-low plot to draw the bars as shown in Figure 22.

Obs	Name	Group	Value	Low	High	Mid	Y	Ys
1	Expresso	Expresso	0.25	0.00	0.25	0.125	0	1.2
2	Macchiato	Expresso	0.25	0.00	0.25	0.125	0	1.2
3	Macchiato	Milk Foam	0.25	0.25	0.50	0.375	0	1.2
4	Cafe Latte	Expresso	0.25	0.00	0.25	0.125	0	1.2
5	Cafe Latte	Steamed Milk	0.50	0.25	0.75	0.500	0	1.2
6	Cafe Latte	Milk Foam	0.25	0.75	1.00	0.875	0	1.2

Figure 21. Data for Coffee Graph

```
title j=l h=1 'Coffee Recipes';
proc sgplot data=Coffee dattrmap=attrmap;
  highlow x=name low=low high=high / group=group type=bar nooutline
          barwidth=0.7 name='a' attrid=Coffee;
  text x=name y=mid text=group / backlight=0.4 textattrs=(color=white);
  xaxis display=( nolabel noticks) offsetmin=0.12 offsetmax=0.12;
  yaxis display=none min=0 max=1 offsetmin=0.15 offsetmax=0.42;
run;
```
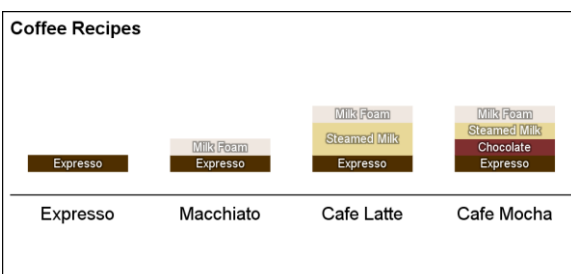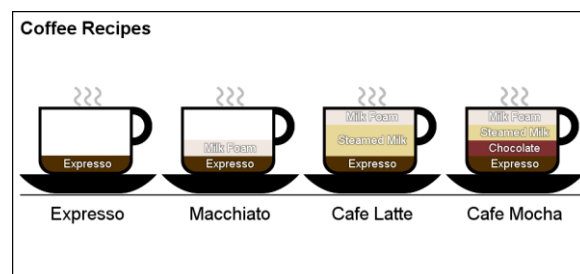

Figure 22. High Low Graph of Ingredients

Figure 23. Graph of Coffee Recipes

Now, comes the key step that turns this ordinary graph into a cool infograph, shown in Figure 23. We use the image of the cup as a mask to sculpt the simple bar chart into one that looks like a coffee cup.

To do this, we use the image of the coffee cup as a symbol using the SYMBOLIMAGE statement. Then, we simply draw a scatter plot with this symbol as its marker right on top of each high-low bar. Recall that in this case the white pixels outside the shape and brown pixels for the part of the cup are opaque. The

pixels in the middle of the cup are transparent. So, the stacked bars previously drawn show through the middle part of the markers.

As a final touch, we display a small marker using an image of steaming vapors on top of the coffee. The code snippet is shown below. Note the use of an attribute map for the color of each ingredient and BACKLIGHT option on the TEXT statement. See resources for a link to download the program ZIP file.

```
title j=l h=1 'Coffee Recipes';
proc sgplot data=Coffee dattrmap=attrmap pad=(bottom=20pct);
  symbolimage name=Cup image="&file1" / voffset=0.15;
  symbolimage name=Steam image="&file2" / voffset=0.0;
  highlow x=name low=low high=high / group=group type=bar nooutline
          barwidth=0.7 name='a' attrid=Coffee;
  scatter x=name y=y / markerattrs=(symbol=Cup size=100);
  scatter x=name y=ys / markerattrs=(symbol=Steam size=20);
  text x=name y=mid text=group / backlight=0.4 textattrs=(color=white);
  xaxis display=( nolabel noticks) offsetmin=0.12 offsetmax=0.12;
  yaxis display=none min=0 max=1 offsetmin=0.15 offsetmax=0.44;
run;
```

USING MARKERS FOR MASK, SHAPE, AND SKIN

For the final example, we use image markers as masks, skins, and shapes, as shown in Figure 24. For this example, we create a graph of soda sales by region. The numbers are made up for illustration purposes only. The idea is to build a decorative infograph of the sales data.

Obs	Region	Value	Low	Mid	Bubble	Y	YLbl
1	USA	0.7	0.01	0.35	0.63	0	0.1
2	Canada	0.6	0.01	0.30	0.54	0	0.1
3	Mexico	0.5	0.01	0.25	0.45	0	0.1
4	Brazil	0.6	0.01	0.30	0.54	0	0.1
5	Chile	0.3	0.01	0.15	0.27	0	0.1

Figure 24. Mask, Skin, and Shape   **Figure 25. Data for Graph**

The first step in creating this graph is to create a bar chart of soda sales by region, as shown in Figure 26. The code for this is shown below. Instead of using a VBAR statement, we have used the HIGHLOW statement. The main reason is to be able to overlay markers later and also use the HIGHLABEL option.

```
title j=l h=1 'Soda Sales by Region (in Millions $)';
proc sgplot data=coke noborder noautolegend nocycleattrs;
  highlow x=Region low=low high=high / type=bar highlabel=high;
  xaxis display=(nolabel noticks) offsetmin=0.12 offsetmax=0.12;
  yaxis display=(nolabel noticks noline) min=0 max=1 offsetmin=0 offsetmax=0;
run;
```

The next step is to sculpt each bar in the shape of the soda bottle. Here I have used an image of a classic coke bottle as a mask. This mask is shown in the left in Figure 24. The image has opaque white pixels everywhere, except in the middle of the bottle defining the shape. These pixels in the middle are fully transparent. When a marker using this image as a symbol is displayed on top of the bar, the color of the bar shows through the transparent pixels, but the rest of the bar is masked. This results in the graph shown in Figure 27. The code snippet is shown below.

```
title j=l h=1 'Soda Sales by Region (in Millions $)';
proc sgplot data=coke noborder noautolegend nocycleattrs;
  symbolimage name=In image="&In" / voffset=0.5 hoffset=0.01;
  highlow x=Region low=low high=high / type=bar nooutline;
  scatter x=Region y=y / markerattrs=(symbol=in size=200);
  xaxis display=( nolabel noticks) offsetmin=0.12 offsetmax=0.12;
  yaxis display=(nolabel noticks noline) min=0 max=1 offsetmin=0 offsetmax=0;
run;
```
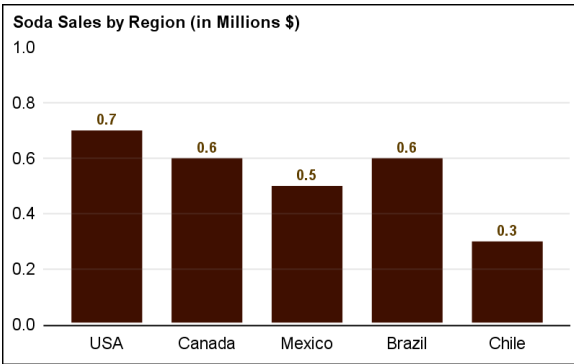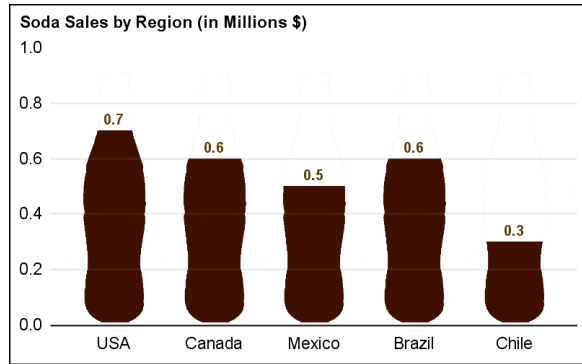

Figure 26. Bar Chart

Figure 27. Bar Chart with Mask

Finally, we use the middle image of the coke bottle in Figure 24 as a skin. This image has the reverse transparency of the mask. The outer pixels are transparent. The pixels making up the shape of the bottle with the highlights and reflections in the glass are opaque. We use another scatter plot to layer this image on top of the previous shape, with a partial transparency. This allows both the dark colors of the shaped bars and the colors of the bottle to show as blended together. The last layer acts like a skin. The result is shown in Figure 28.

Figure 28. Bar Chart with Mask and Skin

If you look closely, you can also see some bubbles in the liquid, placed there using the image of the bubbles in Figure 24. The SGPLOT code snippet is shown below. See resources for a link to download the program ZIP file.

```
title j=l h=1 'Soda Sales by Region (in Millions $)';
proc sgplot data=coke noborder noautolegend nocycleattrs;
  symbolimage name=In image="&In" / voffset=0.5 hoffset=0.01;
  symbolimage name=Out image="&Out" / voffset=0.5;
  symbolimage name=bubbles image="&Bubbles";
  scatter x=Region y=y / markerattrs=(symbol=out size=200) transparency=0;
```

```
   highlow x=Region low=low high=high / type=bar nooutline barwidth=0.7
      fillattrs=(color= cx3f0f00);
   scatter x=Region y=y / markerattrs=(symbol=in size=200);
   scatter x=Region y=y / markerattrs=(symbol=out size=200) transparency=0.4;
   scatter x=Region y=bubble / markerattrs=(symbol=bubbles size=20);
   highlow x=Region low=low high=high / type=bar nooutline barwidth=0.7
      fillattrs=(transparency=1) highlabel=high labelattrs=(color=cx5f3f00);
   refline 0.2 0.4 0.6 0.8  / transparency=0.8;
   xaxis display=( nolabel noticks noline) offsetmin=0.12 offsetmax=0.12;
   yaxis display=(nolabel noticks noline) min=0 max=1
      offsetmin=0 offsetmax=0 values=(0 to 1 by 0.2);
run;
```

CONCLUSION

The SGPLOT procedure has new features with the first maintenance release of SAS 9.4 that support the creation of decorative infographs. In particular, we can use the SYMBOLIMAGE statement that enables us to use any image as a symbol for markers in the graph. When transparent images are used as markers, we can use these as shapes, masks, or skins.

A creative combination of markers as shapes, masks, or skins can be used to make decorative graphs using procedure syntax and data sets. No annotation is required. Now, when you feel the need for a creative and aesthetic visual, the SGPLOT procedure can get you there.

RESOURCES

The PDF file of this paper and the SAS code for all the programs is available at:

Paper: http://support.sas.com/resources/papers/proceedings17/SAS315-2017.pdf

Code: http://support.sas.com/resources/papers/proceedings17/SAS315-2017.zip

REFERENCES

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*, Second Edition. Cheshire, CT: Graphics Press.

Watts, Perry. 2002. "Using ODS and the Macro Facility to Construct Color Charts and Scales for SAS Software Applications." *Proceedings of the Twenty-Seventh Annual SAS Users Group Internation Conference.* Cary, NC: SAS Institute, Inc. Available http://www2.sas.com/proceedings/sugi27/p125-27.pdf.

ACKNOWLEDGMENTS

Thanks to Perry Watts for the %RGBHEX macro for creating a CX formatted string from RGB values.

Thanks to Riley Benson, SAS Institute, for coining the term Decorative Infographs, and for helping with some of the images used in the paper to create the graphs.

RECOMMENDED READING

Matange, Sanjay. "Graphically Speaking." Available http://blogs.sas.com/content/graphicallyspeaking.

Matange, Sanjay. 2016. *Clinical Graphs Using SAS.* Cary, NC: SAS Institute Inc. Available https://www.sas.com/store/prodBK_68179_en.html

Matange, Sanjay and Dan Heath. 2011. *Statistical Graphics Procedures by Example: Effective Graphs Using SAS.* Cary, NC: SAS Institute Inc. Available https://www.sas.com/store/prodBK_63855_en.html.

CONTACT INFORMATION

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

Sanjay Matange
SAS Institute Inc.,
100 SAS Campus Dr.
Cary, NC 27513
Sanjay.Matange@sas.com
http://blogs.sas.com/content/graphicallyspeaking

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.