Paper SAS4425-2020

Pies and Donuts: A New SAS® ODS Graphics Procedure Dessert!

Dan Heath, SAS Institute Inc.

ABSTRACT

Pie charts are a very common graphical display, particularly when showing part-to-whole comparisons. Previously, you needed to use the Graph Template Language (GTL) to create this display in the SAS® ODS Graphics system. But now, with the SGPIE procedure, pie and donut charts are a "piece of cake" to create! This paper focuses not only on the features of this new procedure, but also on the effective use of pie charts in general.

INTRODUCTION

Pie charts are a very common chart seen in reports and publications; It is also one of those charts that can be greatly overused.

The issue with pie charts is a matter of human perception. According to Naomi Robbins, author of *Creating More Effective Graphs* (2005, p. 49**), "We make angle** judgments when we read a pie chart, but we do not judge angles very well. These judgments are biased; we **underestimate acute angles and overestimate obtuse angles.". It turns out that people can** judge lengths much better, as the comparisons you would make in a typical bar chart. However, there are still situations where pie charts can be useful, particularly in part-to-whole comparisons, which we will be discussing later.

Pie chart support has been in the Graph Template Language (GTL) for many years. But until now there has never been an SG procedure that offered pie chart support. Starting with the sixth maintenance release of SAS 9.4, there is the SGPIE procedure, which can produce both pie charts and donut charts. (The procedure was pre-production for that release.) This paper presents the key features of this procedure, while discussing the best practices when using these charts.

PIE CHARTS

To demonstrate the perception problem, observe the charts in Figure 1.
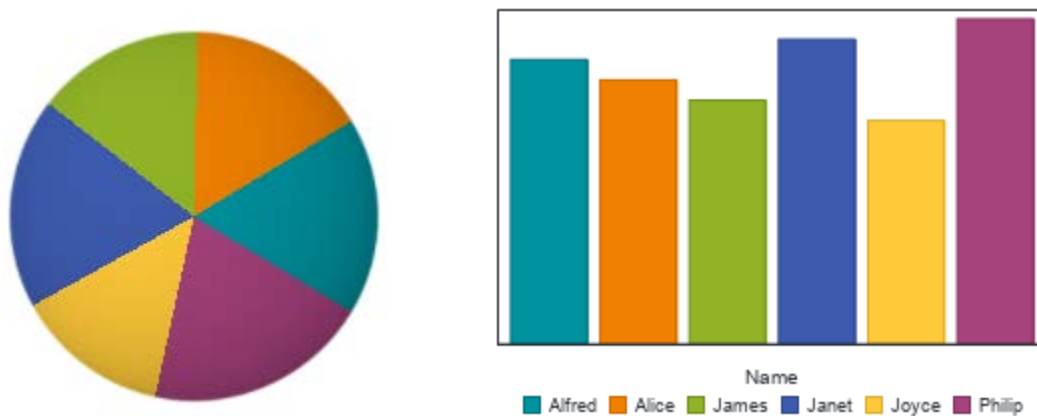

Figure 1. Pie and Bar Perception Comparison

Both the pie and bar are displaying the same data; yet, it is very difficult to compare each slice by size. However, with the bar chart, the differences are much clearer. Pie charts work much better in part-to-whole comparisons, such as one-to-one or one-to-many comparisons.

ONE-TO-ONE COMPARISONS

Figure 2 shows an example of a one-to-one comparison. Notice that the slice labels are in alignment, as well as the center point of both slices. In the SGPIE procedure, the STARTPOS option default is the CENTER of the slice, not the EDGE, which gives you this optimal layout for part-to-whole comparisons.

Heart Study Group Makeup by Gender

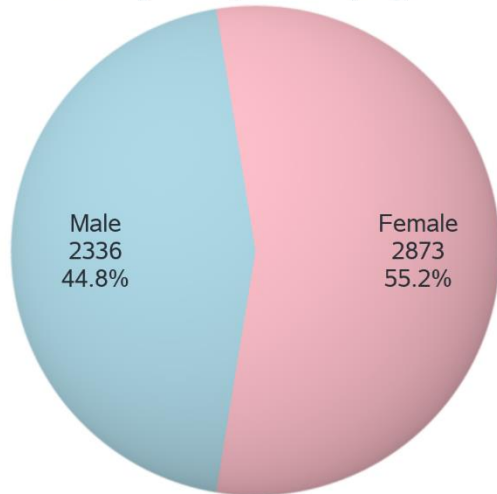

Male
2336
44.8%

Female
2873
55.2%

Figure 2. One-to-One Comparison

This procedure supports discrete attribute maps; but, unlike the other SG procedures, the attribute map is applied to a category variable instead of to a group variable:

```
data attrmap;
length fillcolor $ 9 value $ 6;
retain id "gender";
input value $ fillcolor $;
cards;
Female pink
Male   lightblue
;
run;
title "Heart Study Group Makeup by Gender";
proc sgpie data=sashelp.heart dattrmap=attrmap;
    pie sex / datalabeldisplay=all attrid=gender;
run;
```

ONE-TO-MANY COMPARISONS

One-to-many comparisons are good for showing how one category value compares against the remaining values as a group. It is even possible to compare one group of category values against another to tell a story. One of the best examples I have seen of this case was an article by MarketWatch in 2018. That article used a pie chart to show that the top five market cap companies were worth more than the combined worth of 282 other companies. (See article here). This pie chart tells a story by comparing one group of five slices against another group of 282 slices.

The example in Figure 3 compares the population of Asia against the rest of the world. The STARTPOS=CENTER helps to reinforce this comparison, but you should make sure that Asia is the first slice to get this layout. To do that, set the SLICEORDER to be RESPDESC (response descending) so that the largest slice is drawn first.
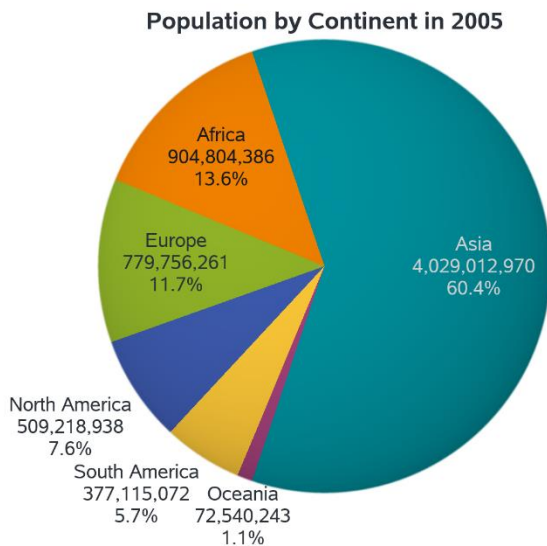
Population by Continent in 2005

Figure 3. One-to-Many Comparison

```
title "Population by Continent in 2005";
proc sgpie data=sashelp.demographics;
   format cont contnames.;
   pie cont / response=pop datalabeldisplay=all
            sliceorder=respdesc;
run;
```

One adjustment that you might want to make to Figure 3 is to reduce the amount of text around the pie. The DATALABELDISPLAY option can be used to control which data values to display around the pie (Figure 4). In addition, anytime the category label is suppressed, an automatic legend is produced to display those values. The legend can be suppressed using the NOAUTOLEGEND option.
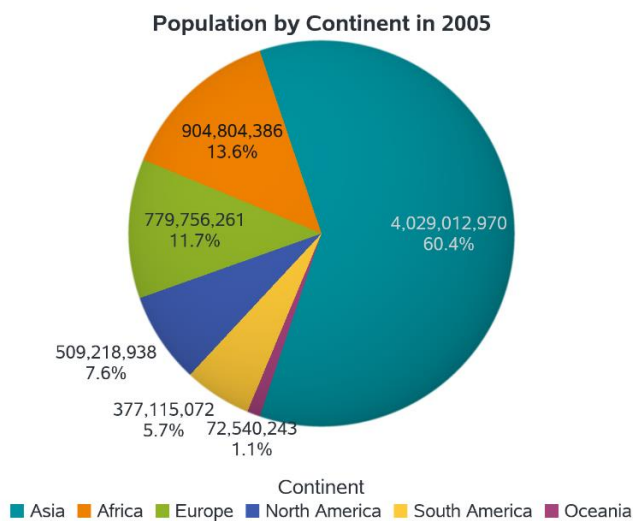
Population by Continent in 2005

Figure 4. Pie with DATALABELDISPLAY and an Automatic Legend

3

```
title "Population by Continent in 2005";
proc sgpie data=sashelp.demographics;
    label cont="Continent";
    format cont contnames.;
    pie cont / response=pop sliceorder=respdesc
               datalabeldisplay=(response percent);
run;
```

OTHER SLICES

You can create your own grouping for comparison by creating an "other" slice. There are two options for creating other slices: MAXSLICES and OTHERPERCENT.

The MAXSLICES option is used to limit the number of pie slices. Setting MAXSLICES=6 produces a pie with five data slices, plus one other slice. In Figure 5, I produced a pie chart that compares the 5 most populous countries against the rest of the world. In addition to setting the MAXSLICES option, you must use the SLICEORDER=RESPDESC option so that the first five slices produced are the largest.
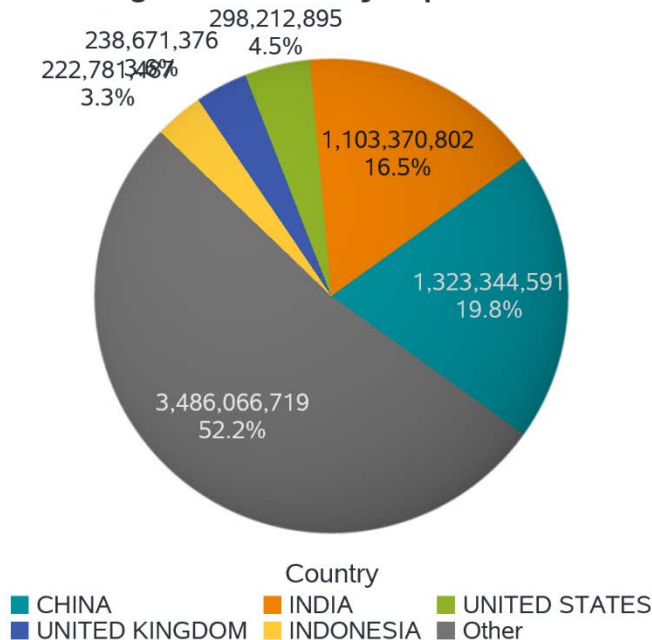

Figure 5. The MAXSLICE Option

```
title "Five Largest Countries by Population in 2005";
proc sgpie data=sashelp.demographics;
    label isoname="Country";
    pie isoname / response=pop sliceorder=respdesc
                  datalabeldisplay=(response percent)
                  maxslices=6;
run;
```

It appears that the biggest slice in Figure 5 was not drawn first; but, from a data perspective, the largest slice was drawn first. China has the largest population of any country, so it was drawn first. The RESPDESC order considers the data before the other slice is created. The smaller countries were lumped into the other slice, which is why it was drawn last.

The layout in Figure 5 is not what you really want for an optimal one-to-many comparison. You would probably prefer the other slice on one side, with the top 5 countries on the other side. Fortunately, there are additional options in the SGPIE procedure that can help you achieve that layout.

In Figure 6, I used the STARTPOS option to set the starting position for the "China" slice to be the EDGE of the slice instead of the CENTER of the slice. Then, I set the STARTANGLE to be 90 degrees so that the pie slices started drawing from the top of the pie. That way, the last slice (the "other" slice, in this case), ends up drawing to the top of the pie.

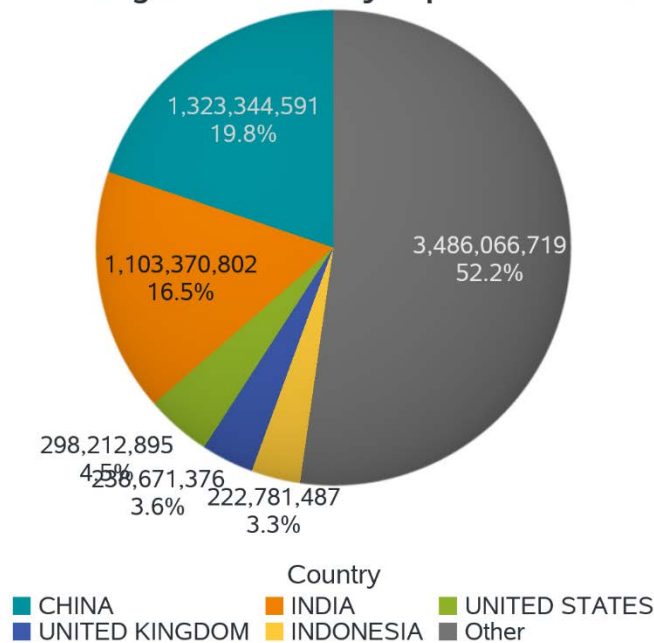**Five Largest Countries by Population in 2005**

Figure 6. The STARTPOS and STARTANGLE Options

```
title "Five Largest Countries by Population in 2005";
proc sgpie data=sashelp.demographics;
    label isoname="Country";
    pie isoname / response=pop sliceorder=respdesc
                  datalabeldisplay=(response percent)
                  maxslices=6 startpos=edge startangle=90;
run;
```

Now that we have the slice layout that we want, there is one other layout item that we might want to address – the label collision. In SGPIE, the DATALABELLOC option gives you the ability to position the labels INSIDE the pie, OUTSIDE the pie, or CALLOUT the labels into an aligned list. With the small slices close together, CALLOUT seems to be the best choice. (See Figure 7).

The other way to create the "other" slice is to do it based on percentage. You can use the OTHERPERCENT option to set a percentage threshold. Any slice with that percentage or less is combined into the "other" slice. Figure 8 demonstrates this option by setting the threshold to 3 percent. In addition, the figure displays the text specified with the OTHERLABEL option for the "other" slice label. The specified label of "Other Countries" overrides the default of "Other".

Five Largest Countries by Population in 2005

1,323,344,591 19.8%

3,486,066,719 52.2%

1,103,370,802 16.5%

298,212,895 4.5%
238,671,376 3.6%
222,781,487 3.3%

Country
- CHINA
- INDIA
- UNITED STATES
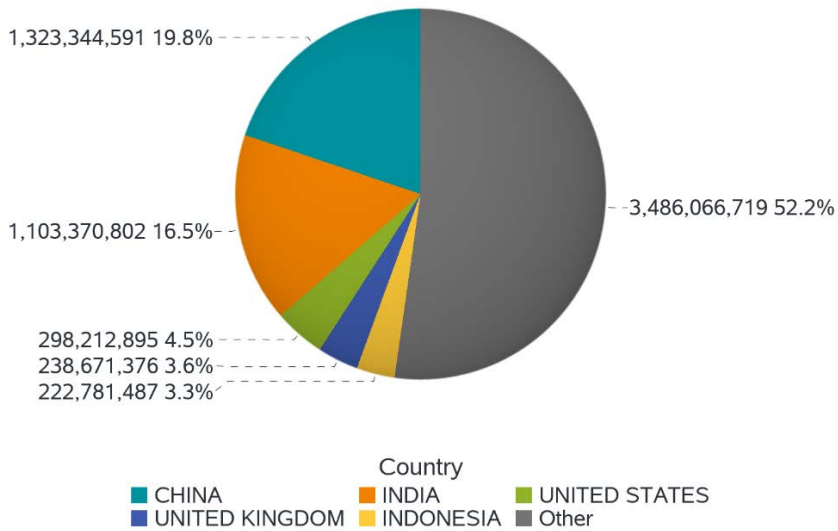- UNITED KINGDOM
- INDONESIA
- Other

Figure 7. The DATALABELLOC Option

```
title "Five Largest Countries by Population in 2005";
proc sgpie data=sashelp.demographics;
    label isoname="Country";
    pie isoname / response=pop sliceorder=respdesc
                  datalabeldisplay=(response percent)
                  maxslices=6 startpos=edge startangle=90
                  datalabelloc=callout;
run;
```

Largest Countries by Population in 2005

1,323,344,591 19.8%

3,486,066,719 52.2%

1,103,370,802 16.5%

298,212,895 4.5%
238,671,376 3.6%
222,781,487 3.3%

Country
- CHINA
- INDIA
- UNITED STATES
- UNITED KINGDOM
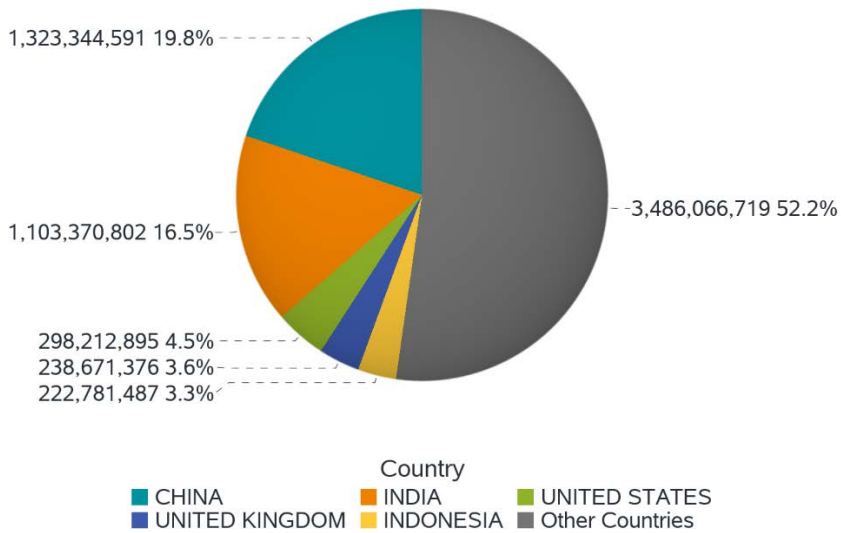- INDONESIA
- Other Countries

Figure 8. The OTHERPERCENT and OTHERLABEL Options

```
title "Five Largest Countries by Population in 2005";
proc sgpie data=sashelp.demographics;
    label isoname="Country";
    pie isoname / response=pop sliceorder=respdesc
                  datalabeldisplay=(response percent)
                  otherpercent=3 startpos=edge startangle=90
                  datalabelloc=callout otherlabel="Other Countries";
run;
```

When using these techniques, you are not limited to the "other" slice being on the right side of the pie. The default direction for adding slices to a pie is in the COUNTERCLOCKWISE direction. However, you can use the DIRECTION option to have the slices added in the CLOCKWISE direction. When combined with our previous options, you get the same pie as **before, but with the "other" slice on the left side of the pie**. (See Figure 9). Also, I added the DATASKIN option to give the pie a 3-D appearance, while retaining the interpretability of a 2-D pie.

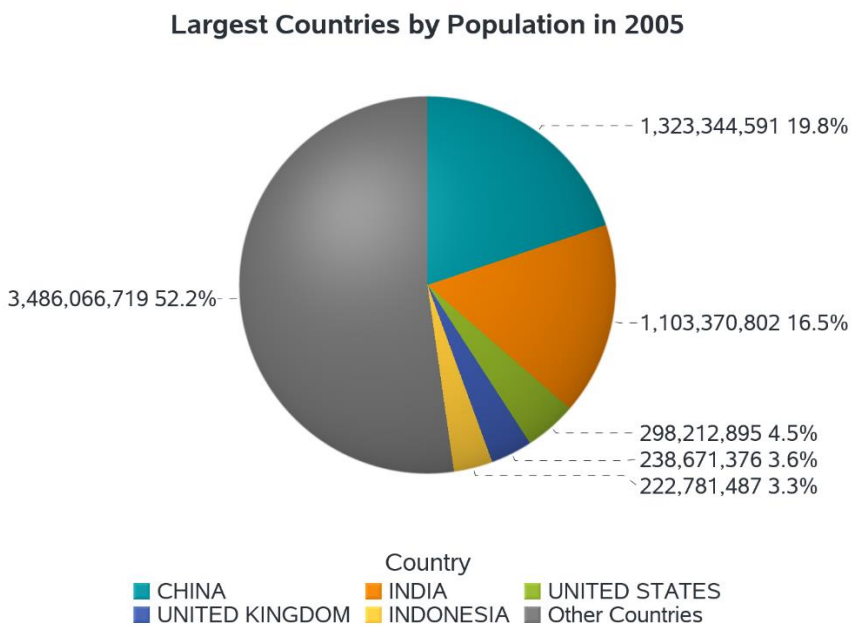### Largest Countries by Population in 2005

Figure 9. The DIRECTION and DATASKIN Options

```
title "Five Largest Countries by Population in 2005";
proc sgpie data=sashelp.demographics;
    label isoname="Country";
    pie isoname / response=pop sliceorder=respdesc
                  datalabeldisplay=(response percent)
                  otherpercent=3 startpos=edge startangle=90
                  datalabelloc=callout otherlabel="Other Countries"
                  direction=clockwise dataskin=pressed;
run;
```

There are a few other pie chart options not mentioned in these examples that are worth noting:

- DATALABELATTRS – controls the text attributes of the pie data labels
- OTHERFILLATTRS – controls the fill attributes of the **"other" slice**

7

- MISSING – shows any "missing" category slices in pie
- URL -- specifies a variable that contains URL addresses for drilldown support. You must specify the IMAGEMAP option in the ODS GRAPHICS statement for this option to work.

The SGPIE procedure also supports the STYLEATTRS statement, as do other SG procedures. However, only two options are supported:

- BACKCOLOR – controls the background color of the pie chart
- DATACOLOR – overrides the fill colors from the ODS style that are used to color the slices

DONUT CHARTS

Basically, donut charts are pie charts with a hole in them. In fact, all the pie features discussed in this paper are also available on the donut chart. However, the donut chart does have some additional abilities regarding the hole. I cover these below.

In Figure 10, I simply took the code from the Figure 8 and changed the statement from PIE to DONUT. The chart appears with the same information as Figure 8, with the addition of the hole.

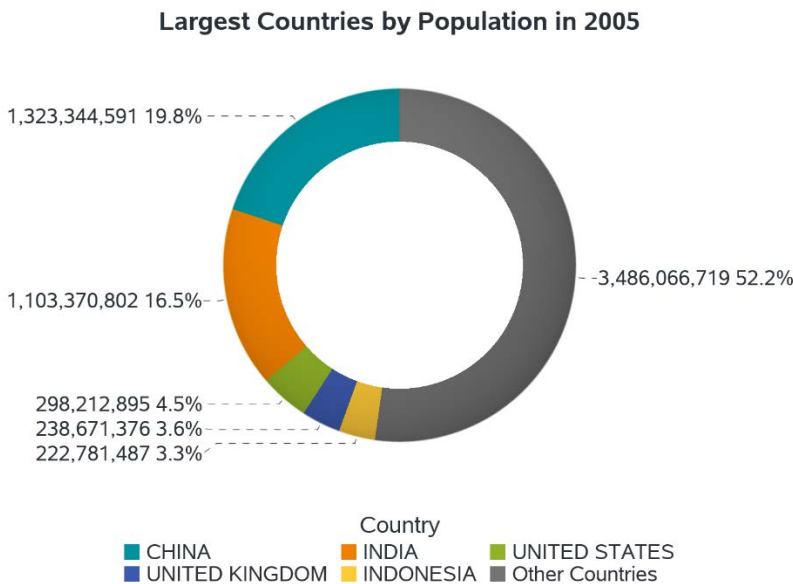**Largest Countries by Population in 2005**

Figure 10. The DONUT Chart

```
title "Five Largest Countries by Population in 2005";
proc sgpie data=sashelp.demographics;
   label isoname="Country";
   donut isoname / response=pop sliceorder=respdesc
                   datalabeldisplay=(response percent)
                   otherpercent=3 startpos=edge startangle=90
                   datalabelloc=callout otherlabel="Other Countries";
run;
```

The data portion of the donut chart is called the "ring". By default, the size of the ring is 30% of the donut radius, meaning that the hole takes up 70%. These percentages are the default, regardless of the size of the donut. However, you can change the percentage size of the ring by using the RINGSIZE option to set a value between 0 and 1.

The center hole gives you the ability to inject additional information into the chart. In Figure 11, the HOLELABEL option adds a custom label **of "Population" to the hole.** When the optional custom text is not provided with this option, the response variable label or name is displayed. I also increased the RINGSIZE to 50% in this example.
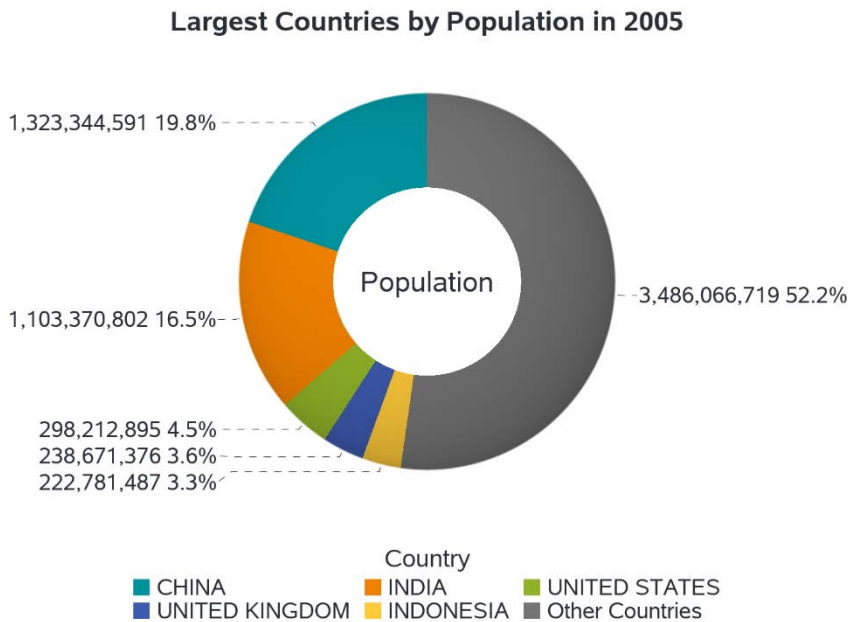

Figure 11. The HOLELABEL option

```
title "Five Largest Countries by Population in 2005";
proc sgpie data=sashelp.demographics;
    label isoname="Country";
    donut isoname / response=pop sliceorder=respdesc
                    datalabeldisplay=(response percent)
                    otherpercent=3 startpos=edge startangle=90
                    datalabelloc=callout otherlabel="Other Countries"
                    holelabel="Population" ringsize=0.5;
run;
```

While the HOLELABEL option can be used to add text, the HOLEVALUE can be used to add numeric values. Large numeric values are displayed using condensed business format, which performs scale extraction and rounding of the number. When you do not specify a number with the HOLEVALUE option, the sum of all slice response values is displayed.

The HOLELABEL and HOLEVALUE options can be used independently or together. In Figure **11, I used the HOLELABEL option to specify a label of "Total"**. But I used HOLEVALUE without a number to display the sum of all slice response values.

There are HOLELABELATTRS and HOLEVALUEATTRS options to control the text attributes as you would normally expect in an SG procedure – except that there is no SIZE option. The reason is that the hole label and hole value text sizes are automatically proportioned to one another, as well as to the size of the hole.
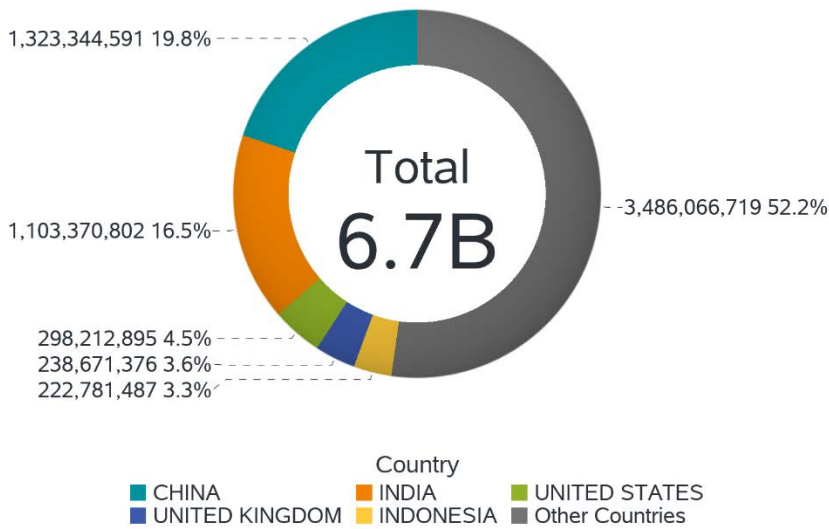
Largest Countries by Population in 2005

Figure 12. The HOLELABEL and HOLEVALUE options together

```
title "Five Largest Countries by Population in 2005";
proc sgpie data=sashelp.demographics;
    label isoname="Country";
    donut isoname / response=pop sliceorder=respdesc
                    datalabeldisplay=(response percent)
                    otherpercent=3 startpos=edge startangle=90
                    datalabelloc=callout otherlabel="Other Countries"
                    holelabel="Total" holevalue;
run;
```

WHAT ABOUT GROUPED PIES?

Before I discuss group support for the SGPIE procedure, it is worth discussing the issues with stacked grouped pies. At the beginning of the pie chart section, I discussed the human perception issues that can happen with a standard pie. Now, imagine stacking those pies on top of one another (Figure 13).
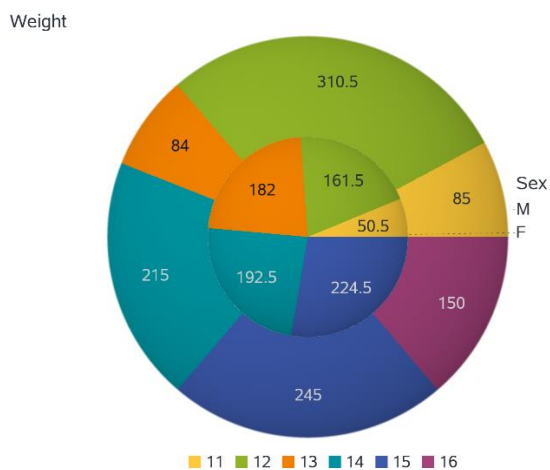

Figure 13. A Stacked, Grouped Pie

As you can see, it is very difficult to interpret the slices from one pie with the other pie superimposed on it. The better option is to break the groups into separate pies and arrange them into a gridded layout.

The SGPIE procedure does not currently support a GROUP option directly on the PIE or DONUT statements, but you can create this type of grouped layout by using BY-grouping and ODS LAYOUT. In the following example, a two-column gridded layout is defined. The advancement to the next cell is triggered by each BY-group value. The BY-lines are disabled, and the TITLE is set to be the BY-group value for each pie. The ODS TEXT statement is used to create an overall title.

```
ods escapechar='^';
ods html path="." (url=none) file="pie_layout.html";
ods text="^S={width=100% just=c fontsize=20pt}Travel Revenue by Hub";
ods layout gridded columns=2 advance=bygroup;
ods graphics / width=3.5in noborder;
options nobyline;
proc sort data=sashelp.revhub2 out=revhub2; by hub; run;
proc sgpie data=revhub2;
   by hub;
   title1 "#byval1";
   pie type / response=revenue;
run;
ods layout end;
ods html close;
```

CONCLUSION

The SGPIE procedure is a useful tool for creating pie and donut charts; but like any other tool, it should be used carefully. Whenever you are creating visualizations to share findings or information, take time to consider the best visualization that will clearly convey the message that you want to share.

REFERENCES

Robbins, Naomi B. 2005. *Creating More Effective Graphs*. Hoboken, NJ: John Wiley & Sons.

Chang, Sue. **"One chart puts mega tech's trillions of marke**t value into eye-popping perspective."MarketWatch July 19, 2018. https://www.marketwatch.com/story/this-1-chart-puts-mega-techs-trillions-of-market-value-into-eye-popping-perspective-2018-07-18.

CONTACT INFORMATION

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

Dan Heath
SAS Institute, Inc.
Dan.Heath@sas.com