

Paper HW-03

Easier than You Think: Creating Maps with SAS® Enterprise Guide®

Stephanie R. Thompson, Datamum, Memphis, TN

ABSTRACT

Have you ever wanted to display data on a map to add more punch to your analysis? Sometimes seeing things geographically can put things into a completely new perspective. Letting your data tell a story on a map is easier than you might think. A two-step approach with SAS® Enterprise Guide® is all you need. From using supplied shape files right through customizing your graph, this paper walks you through the process. Some best practices are also included to make sure your customers are impressed. They will never know how easy it really was.

INTRODUCTION

Creating a map in SAS Enterprise Guide 4.3 is even easier than in earlier versions. The key to creating a Map Chart is making sure your data and the map file have a commonly named variable on which to join. Earlier versions required you to use the Query Builder task to join your data to the shape or map file. In 4.3, this is done automatically. All it takes to create a basic Map Chart is your data with a geographic variable (e.g., state abbreviation or FIPS code) and a variable with something to graph.

Once the basic Map Chart is created, you can customize it to make it look even better. There are numerous options and styles that you can apply. With very little effort, you can create a very professional looking Map Chart.

Data graphed used in this paper is from the 2000 U.S. Census and downloaded from their web site. The shape files are those provided with SAS and located the Maps library. That library is one of the default libraries available during a SAS session. It is installed at C:\Program Files\SAS\SASFoundation\9.2\maps in a default installation for SAS 9.2.

GETTING IN SHAPE

The shape files are, for this paper, the datasets provided by SAS in the Maps library. These datasets contain coordinates used by SAS to generate the map. In total, there are 307 map datasets. There are datasets for the US each state, Europe, the world, and many other countries and areas.

The US.sas7bdat dataset used in the examples in this paper only contains five variables: STATE, SEGMENT, STATECODE, X, and Y. This dataset joins to the census data using STATECODE. Below is the partial output from PROC CONTENTS for the US dataset.

Data Set Name	MAPS.US	Observations	1551
Member Type	DATA	Variables	5
Engine	V9	Indexes	0
Created	Monday, December 22, 2008 02:16:34 PM	Observation Length	32
Last Modified	Monday, December 22, 2008 02:16:34 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	YES
...			

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
2	SEGMENT	Num	5	State Segment Number
1	STATE	Num	5	State FIPS Code
3	STATECODE	Char	2	Two-letter abbrev. for state name
4	X	Num	8	Projected Longitude from Radians
5	Y	Num	8	Projected Latitude from Radians

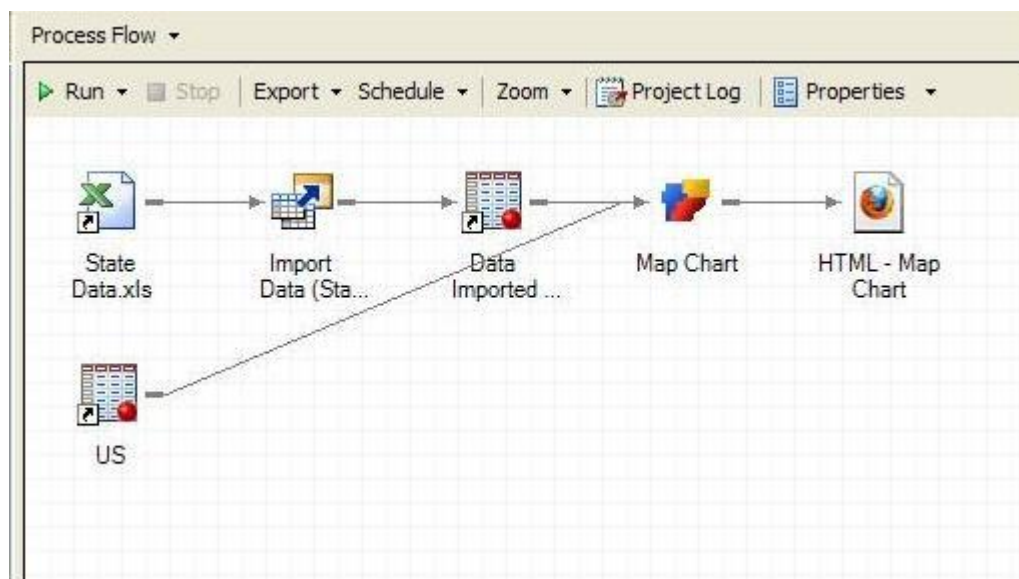
It is worth noting that this dataset has more than the expected 50 or 51 observations. While each state is represented in the dataset, U.S. territories such as Puerto Rico are also included. Every state and territory has multiple observations to delineate the outline of the state. Colorado has eight observations whereas Maryland has 54. It just depends on the complexity of the state's border.

There is another US dataset in the Maps library, US2.sas7bdat. The datasets with a "2" at the end do not work as easily in SAS Enterprise Guide and are not covered in this paper. They contain a different mechanism for defining the geographic borders.

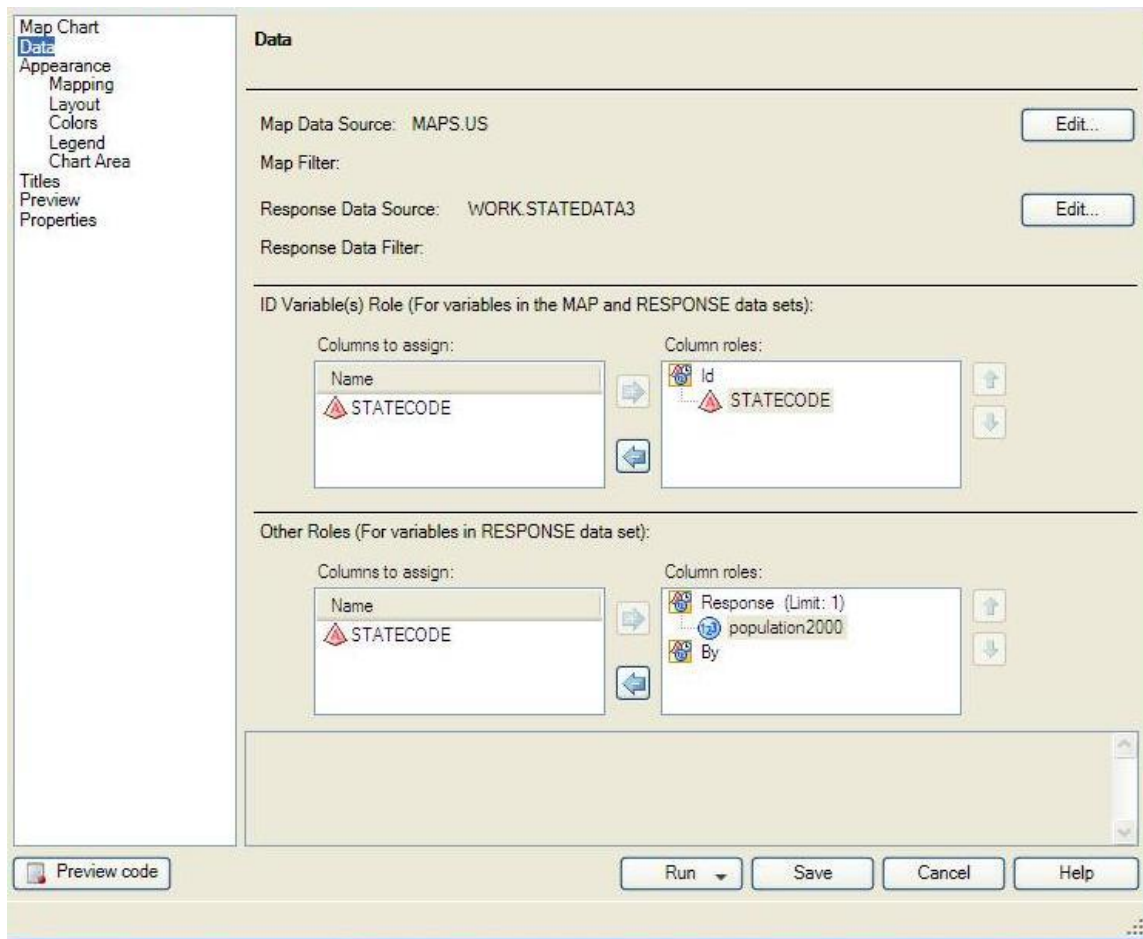
Considering all of the map datasets available, there is a wide assortment of maps that can be easily created. You can use maps at the world level down to the U.S. County level with the maps already part of your SAS installation. If you have a corresponding variable in your dataset, you can map it.

LINKING TO DATA

The entire process flow has four main elements: two datasets, a Map Chart task, and the output. The example shown below has a few more since the data is imported from Excel.



It is interesting to note that there is no Query Builder task where the two datasets come together. The lines just merge before the map chart without any interim task at all. This is because they contain a common variable and automatically join in the first step of the Map Chart task. The screen shot below shows the datasets at the top and the ID variable shows the common variable STATECODE.

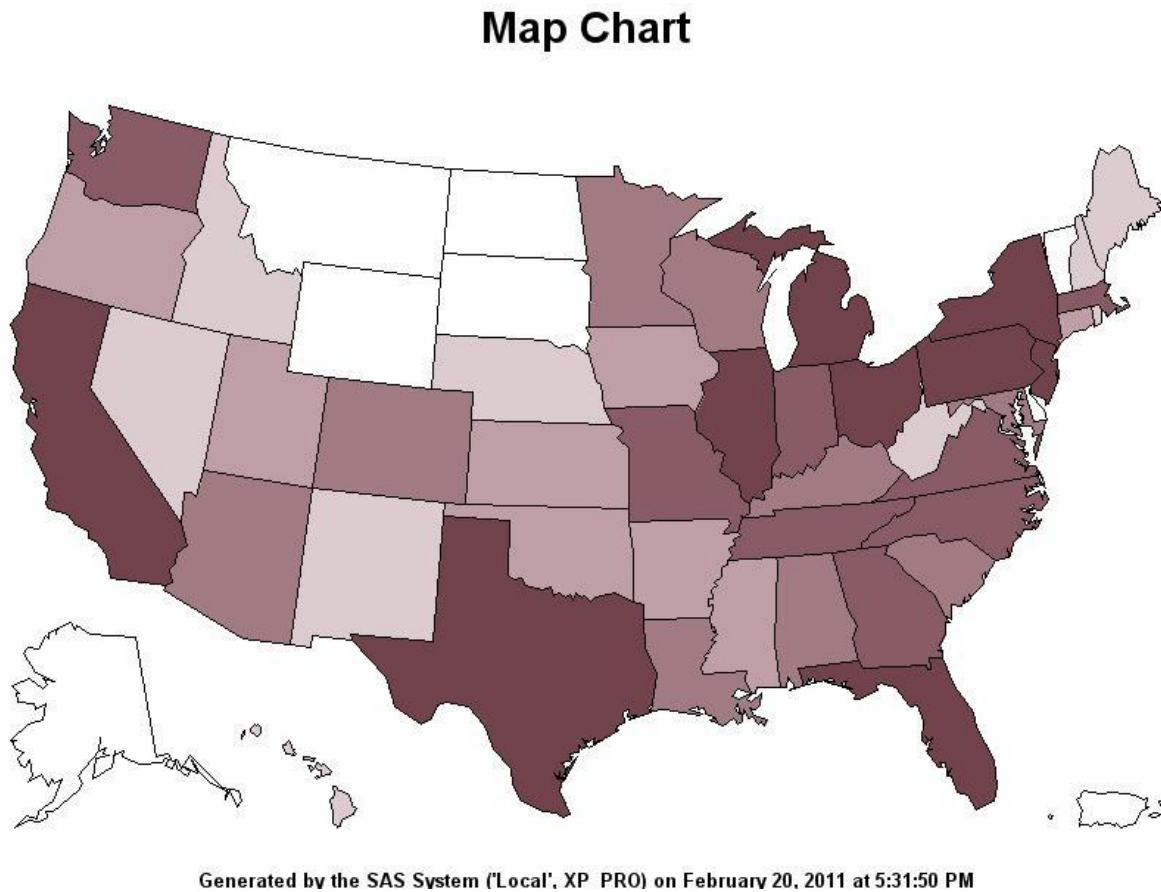


You can have multiple variables to graph in your dataset. You can select only one RESPONSE variable to graph. A by variable can be added and separate Map Charts will be generated for each level of response. If gender was in the statedata3 dataset, graphs of population by state could be created for each gender value in the data.

As previously noted, earlier versions of Enterprise Guide required the additional step of joining the dataset with the Query Builder task. If you cannot seem to get your data to join to the map dataset, you can manually join them prior to graphing. A single dataset can fill the role of Map Data Source and Response Data Source.

CREATING THE PICTURE WORTH A THOUSAND WORDS

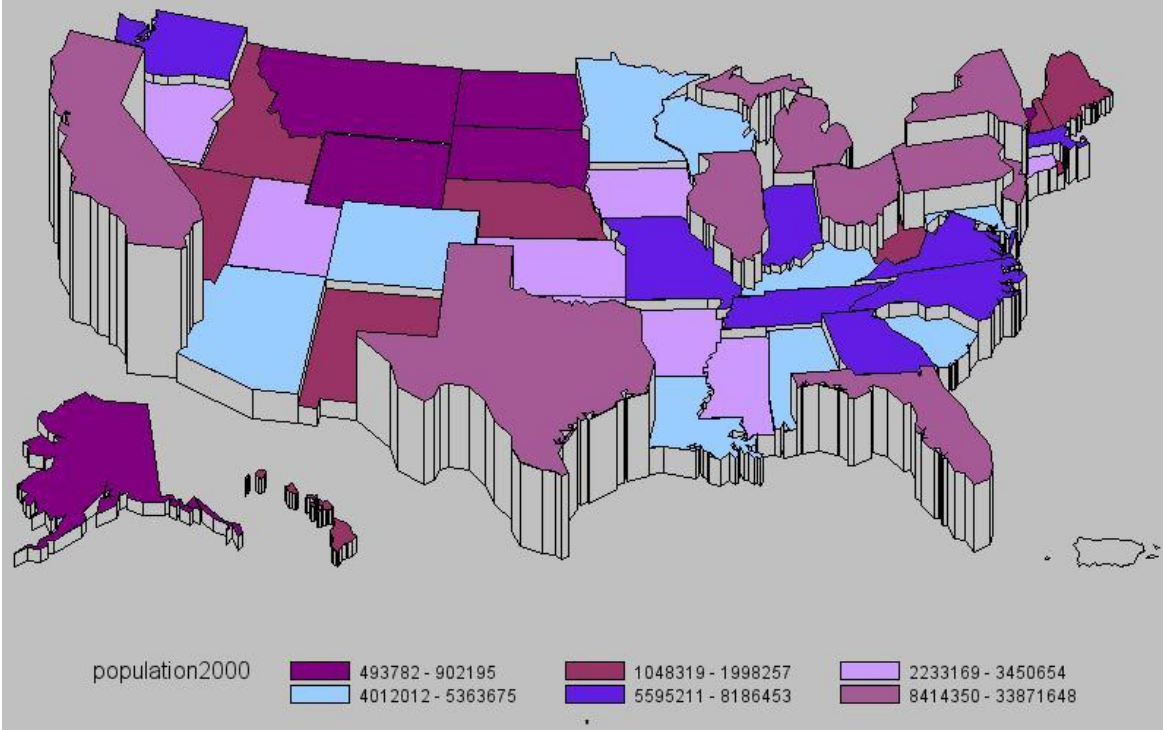
With just a few clicks, the following map was created. It is not fancy and only default settings have been used.



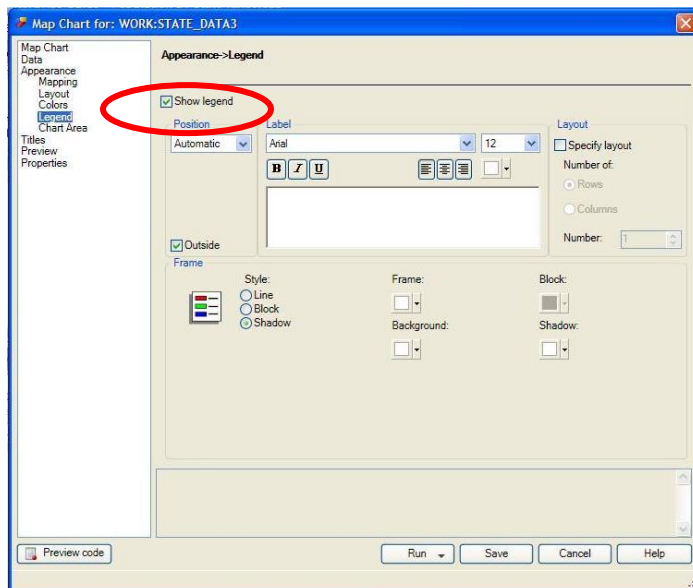
You can see the different levels of population and are able to tell there is a similarity among Montana, North Dakota, South Dakota, Wyoming, Alaska, and Vermont. However, this map is lacking a legend since it is not part of the default Map Chart settings. In order to make your Map Chart more understandable, a little more work is required.

With just a few more clicks, you can have a much nicer looking map. The Map Chart below customizes the title, adds a legend, uses custom colors, and uses the 3-D option. You may not be a fan of the colors chosen, but this is a more informative Map Chart than the first one.

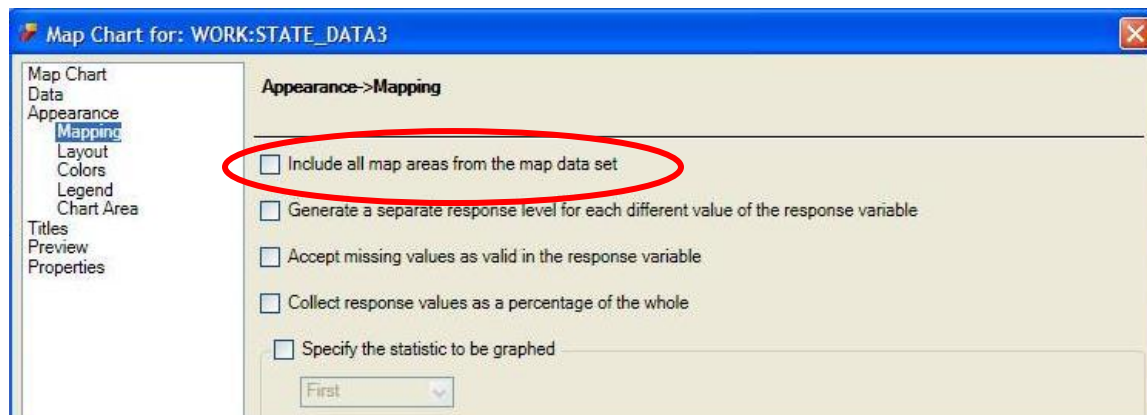
Resident population (April 1) 2000 (complete count)



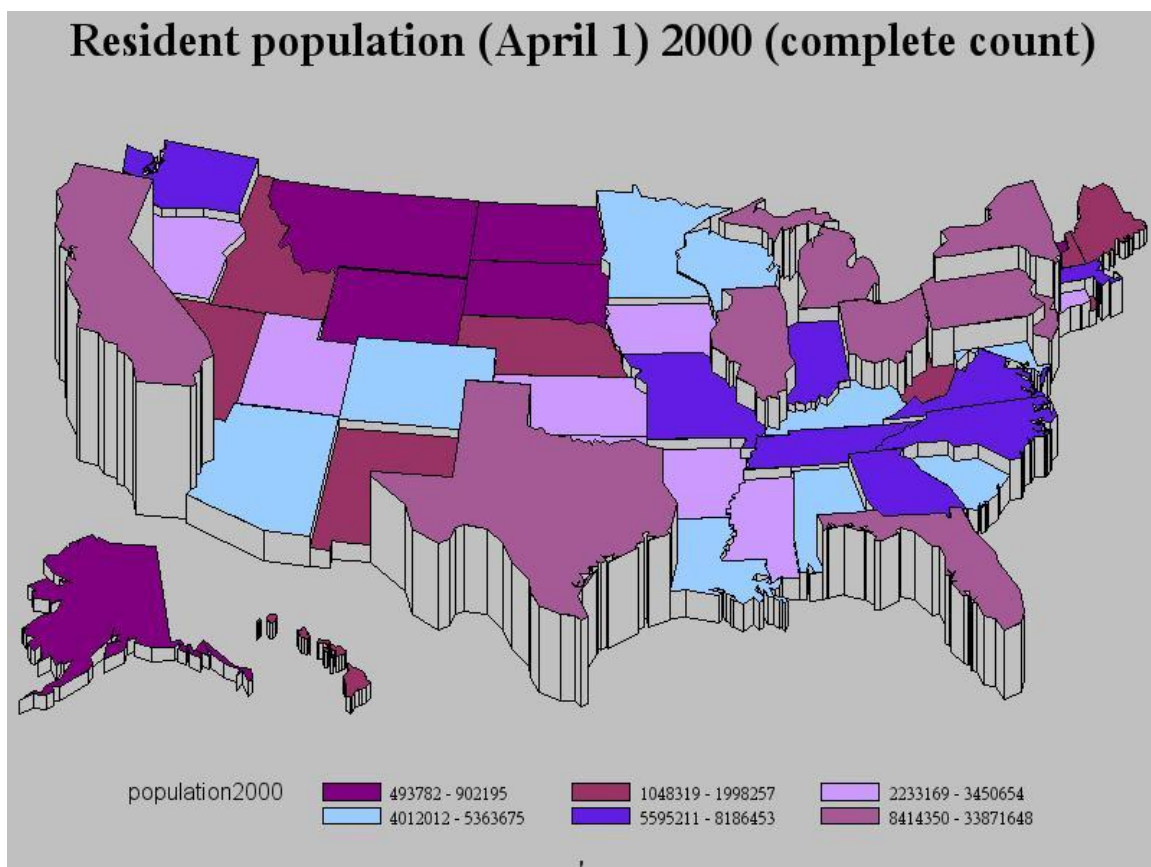
By default, the legend is not part of the map image as previously noted. The setting you need to change is under Appearance then Legend and shown below.



Puerto Rico in both maps contains no population data. In the first case, it is shaded just like the lowest response level. This is coincidental to the null value color matching the low response color on the default style. The second map shows Puerto Rico with no color represented in the legend. You can de-select the option circled below to limit the map to show only states or territories with data. Use caution however, if you plan to chart sparse data where some states have a missing value for the response value you would want to set them to zero beforehand. Otherwise, your map chart will have holes where states should be.

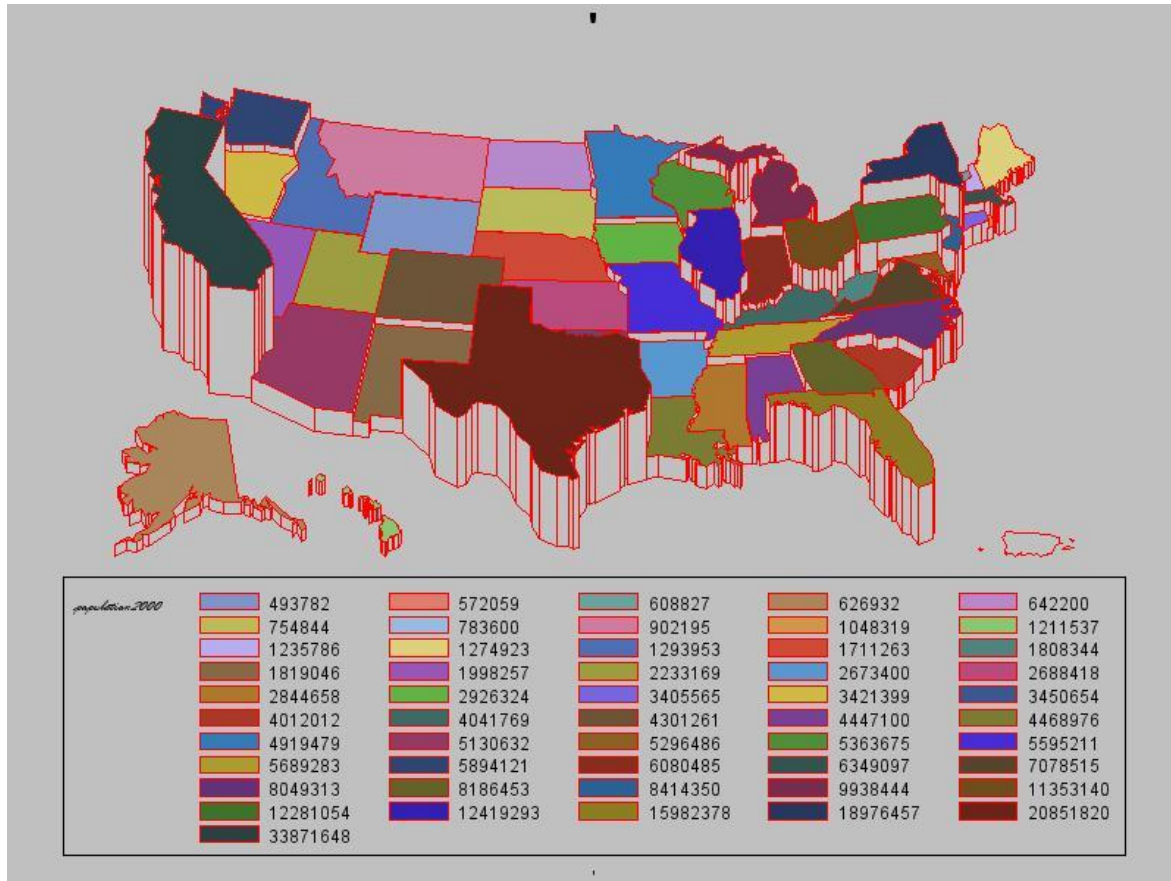


The map below does not show Puerto Rico, but is otherwise the same as the earlier map. Use your judgment on whether or not to show areas with missing data. In the end, it will depend on the story you are trying to tell with the data.



PITFALLS AND PRACTICES

Even though you can make your map quite fancy, you want to make sure not to go overboard. You can easily make the map unreadable and then it will lose its meaning. One challenge is to make sure the legend colors are distinct enough to differentiate. If there are 10 different shades of red, the reader may not know which state is at which level. This section is devoted to making maps readable and easy to understand. Since a picture is worth a thousand words, the map below is an example of what not to do.



Having a different color for each state's population makes this map very hard to interpret. The colors are not helpful either. Trying to match the legend color to the correct state will take the viewer a long time. Ideally, you want the Map Chart to convey the message easily. If you want to show the value for each state, a table might be a better option. The map seems more suited for displaying states with similar values in a range.

The title is missing so you do not know what the map is of. The numbers in the legend could represent any type of population. Is it cows, cars, or people? The text on the legend is hard to read. It is small and in a script style font. It is just another challenge for the reader.

Red is not the best choice of outline color. It seems to glare and clash with some of the 50 different colors used for the states. The goal is to try to make the Map Chart pleasing while it tells the data's story.

Simply stated, the best map charts have

- a meaningful title
- a legend that does not have an excessive amount of differentiation in a color scheme that is easily told apart with a readable font
- outline and background colors that complement and do not compete with the data

It may take some trial and error but with some practice with the different options and styles, you will be making good looking Map Charts in minutes.

THE NEXT FRONTIER

You may find yourself needing a map file that does not come with SAS. Maybe a country's borders changed since you installed SAS. Additional maps can be found on the SAS Maps Online page located through the SAS support site. There is a wide variety of maps available and you can simply download the desired dataset. You will need to log in to your SAS profile to download files.

ArcGIS by ESRI is a powerful spatial analysis tool. There is a lot more you can do to show data spatially with this tool than you can do in SAS. The SAS Bridge for ESRI makes it easy to link your SAS data to ESRI. With ArcGIS, you can show your data on a satellite image or as part of a multi layered view of data from many sources. For example, you could show crime statistics by county layered over population. The possibilities are endless with what you can do if you have the need and software. The good thing is you do not need to extract your data from SAS to use this powerful tool. If you are interested in seeing examples of maps created using this method, the poster display at the conference typically has several or you can look up examples on the ESRI website.

Lastly, as with any project in SAS Enterprise Guide, you can copy the SAS code that is created and put it into your Base SAS program. With a few minor modifications, you can re-create the same high quality graph. This is yet another way to leverage the benefit of a spatial representation of data. The main part of the Map Chart code is shown below.

```
GOPTIONS CBACK= ;
TITLE;FOOTNOTE;

TITLE1 "Map Chart";
FOOTNOTE1 "Generated by the SAS System (&_SASSERVERNAME, &SYSSCPL) on
%TRIM(%QSYSFUNC (DATE()), NLDATE20.) at %TRIM(%SYSFUNC (TIME()), TIMEAMP12.)";

PROC GMAP GOUT=MAPCHART DATA=WORK.MAPCHARTRESPONSEPREP MAP=WORK.MAPCHARTMAPPREP
ALL;
    ID STATECODE;
    CHORO population2000 /
        WOUTLINE=1
        NOLEGEND
    ;
RUN;
QUIT;
```

This is not a lot of code but if you are not familiar with the PROC GMAP syntax and all of the options, it could take longer to directly code your Map Chart than to let SAS Enterprise Guide do it for you. To get some tips on doing this, see the paper cited in the references or any of the other papers on the topic in the SAS Global Forum proceedings or Regional Users Group conference proceedings.

REFERENCES

Data source: http://quickfacts.census.gov/qfd/download_data.html

SAS Maps Online: <http://support.sas.com/rnd/datavisualization/mapsonline/>

SAS / ESRI Partnership: <http://www.sas.com/partners/directory/esri/>

Sample ESRI Maps: <http://www.esri.com/mapping-for-everyone/index.html>

Thompson, Stephanie R. "Graphing the Easy Way with SAS® Enterprise Guide® (or How to Look Good With Less Effort)." Presented at the SouthEast SAS Users Group (SESUG) 2009 Conference. Birmingham, AL. October 26, 2009. <http://analytics.ncsu.edu/sesug/2009/HOW007.Thompson.pdf>

CONCLUSION

As illustrated in this paper, creating a professional looking Map Chart in SAS Enterprise Guide is a simple process that has gotten easier with each release of the software. As long as you have data that can be joined to a map dataset, you are set. It only requires one task and two datasets in the simplest form. With some practice, you will be creating easy to interpret, high quality maps to help your data tell its story.

CONTACT INFORMATION

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

Name: Stephanie R. Thompson
Enterprise: Datamum
Work Phone: 901-326-0030
E-mail: stephanie@datamum.com
Website: www.datamum.com
Linked In: Stephanie Thompson, Analytics Professional
Twitter: @SRT_SESUG
sasCommunity User Name: Stephanie

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.