

Tips and Tricks: Using the new SAS Map data sets

Liz Simon, Darrell Massengill, SAS Institute Inc., Cary, NC

ABSTRACT

SAS Solutions and SAS users have a growing need for good quality SAS Maps with accurate political boundaries; the types of maps provided by SAS mapping technology. In order for SAS to provide good quality maps for our customers we need good, solid and accurate map data. The map data includes worldwide political boundary data for countries, regions, states, counties and provinces. Since SAS is not in the business of generating map data and with the rapidly changing world, the only way to provide this map data is to purchase/license it from a 3rd party source, which specializes in map data. Users needing additional specialized map data could license it directly from the 3rd party and it would all match up with the maps we provide.

This paper will explore the new map data: discussing the problems and limitations with the old map data and features and examples of using the new data.

INTRODUCTION

In the past, we used various sources of map data; some free, some inexpensive and others more costly. We spending huge amounts of time 'munging' this data to fit our needs. This approach will no longer meet our needs. The only way to provide up-to-date, accurate and good quality map data is to have a single standardized source of origin, using the same datum and spatial units. We have partnered with a 3rd party vendor to provide this data with annual updates.

We will first highlight the major differences between the old and new map data, such as new library names, new filenames and new variables. We will include some examples to show coding changes required to create similar output using the new data. Some examples will also highlight features of the new data. We will show how to use some of the new GMAP functions/features with this data.

EXISTING MAP DATA SETS

The old map datasets were based on free or purchased maps from various sources. This process introduced inaccuracies, created problems and cost us a huge amount of time in 'munging' this data. This approach no longer meet our needs:

- Our inexpensive sources of data have basically dried up. For example, we can no longer get the CIA world data for processing our world maps. Without this data, we cannot create our maps.
- Our other sources of data are not in a consistent coordinate plane, occasionally have problems and provide no technical support. We spend a huge amount of time trying to manipulate the data into a usable form. In addition the data was frequently out of date and we had no way to fix political changes around the world except to try to change the political boundary lines by hand. We also do not have the necessary spatial data editing tools for this.

In order for SAS to keep up with the new uses and demands for maps from our customers, we need more accurate map data, so the partnership with 3rd party vendor, *GfK GeoMarketing* (GfK-GeoMarketing.com/SAS) is very exciting.

THE GfK PARTNERSHIP: WHAT WE GET

GfK GeoMarketing was formed in 2006 through the merger of GfK MACON, GfK PRISMA and GfK Regionalforschung and is a part of the international GfK network. According to the website they offer digital postal code and administrative boundary maps for over 240 countries.

GfK GeoMarketing boasts:

- Digital maps from the world's largest archive: GfK GeoMarketing's vector maps offer worldwide
- coverage with detailed administrative and postal maps supplemented by topographical maps.

GfK GeoMarketing's digital maps:

- are part of the most complete collection of administrative and postal maps in the world
- are digitalized and completely overlap-free
- are fully attributed with names and official descriptions
- are updated regularly
- are based on geographic coordinates (WGS84)

The agreement between GfK GeoMarketing and SAS, provides us with Maps of the World at administrative levels 0,1 and 2. Zero (0) level represents the countries of the world. Level one (1) represents the first administrative level, like STATE in the United States. Level two (2) represents the 2nd administrative level, like COUNTY in the United States. The agreement provides levels that cover world country boundaries and boundaries of the first administrative levels worldwide; boundaries of the 2nd level where available and boundaries of the 3rd level, for a few countries, as were provided in the old MAPS data sets. The agreement also provides the NUTS 0, 1, 2 and 3 boundaries of Europe, which our European customers have asked for over the years. These maps are more accurate, more up-to-date and more detailed than our old map data.

The first *production* version of these maps was released with SAS/GRAPH 9.30m2. Due to legal, license agreements with *GfKGeoMarketing* we cannot make these maps available for any pre-9.30M2 Release. The map datasets can only be used with SAS 9.3M2 and beyond and requires a SAS/GRAPH license. The data are currently available at the MapsOnline website

(http://support.sas.com/rnd/datavisualization/maponline/html/gfkmmaps_93m2.html)

Updates (that are ready between releases) will be available for download at the website.

Licensing the data from a 3rd party gives us a single map data source, provides a uniform plane for the whole world, and provides a source for updates and Technical Support. Users needing additional specialized data can license it directly and it would match up with the maps we provide.

Visit the SAS page at GfK-GeoMarketing.com/SAS.

NEW GfK MAP DATA AND SAS/GRAPH

With the fore-knowledge that there would be many differences between the old and new data, we set some goals: (1) make the maps more consistent and easier to use. (2) avoid any unnecessary processing of this data, thus allowing for more frequent updates and (3) avoid breaking compatibility, except where absolutely necessary (for #1 and 2). It was very difficult trying to avoid breaking compatibility and at the same time not over-process the data.

How will the changes and differences impact you, our customer?

There will be a migration period during which both the old and new maps will be shipped.

The old maps are assigned to a library referenced by MAPSSAS and will not be updated; while the new maps are accessible by a library named MAPSGFK. During this transition period, the old maps will continue to be shipped as the default MAPS. MAPS can be redefined to point to either MAPSSAS or MAPSGFK. In the future, MAPSSAS will stop shipping, but the old data is already archived and available for download via the MapsOnline website <http://support.sas.com/rnd/datavisualization/maponline/html/archivedmaps.html>.

Summary of changes.

1. **New License.** The following note will appear in the SAS Log:

NOTE: "The map data sets in library MAPSGFK are based on the digital maps from GfK GeoMarketing and are covered by their Copyright. For additional information, see <http://support.sas.com/maponline/gfkllicense>".

2. **New Libnames.** There are 3 libraries relating to maps: MAPSSAS (default), MAPSGFK and MAPS. MAPSSAS and MAPSGFK are configuration-time-only librefs like SASHELP. MAPS can be assigned in your SAS Session to point to either libref. For example, LIBNAME MAPS (MAPSGFK);

3. **New Files and Filenames.**

- Changes in naming files, including consistency in names, updates, longer-non-truncated, more descriptive names like DR_CONGO vs ZAIRE_CONGO, AFGHANISTAN vs AFGHANIS, US_COUNTIES vs COUNTIES. There are also changes in naming the corresponding attribute tables, like ARGENTINA_ATTR vs ARGENT12.
- Many new map data sets like, KOSOVO, CARIBBEAN, EUROPENUTS0 (NUTS level0 for Europe), (*The NUTS classification is a hierarchical system for dividing up the economic territory of the EU*);
- _ATTR suffix for attribute files like, ALGERIA_ATTR;
- _ALL suffix for countries with dependencies and other territories like, AUSTRALIA_ALL (Australia and dependencies);
- PROJPARM (containing Gproject information for all projected maps).
- Some files were eliminated from this new set like, COUNTY and US2.

4. **New Variables.**

- Many changes to variable names. All the map datasets have ID, SEGMENT, LAT, LONG, X, Y, RESOLUTION and DENSITY. All the _ATTR datasets have ID, IDNAME, ISO, and ISONAME. Many map datasets have additional variables like, ID1, LAKE while the _ATTR datasets have additional variables like ID1, ID1NAME, IDNAMEU, and ID1NAMEU. A few map datasets also have ID2, CONT, STATE, COUNTY, while the _ATTR dataset has ID2, ID2NAME, ID2NAMEU, and CONT. The table below shows examples in various data sets.
- The RESOLUTION variable is a more practical version of GREDUCE. The levels are based on your display size. The various pixel sizes are shown below. You can use the RES= option in GMAP to display a high, medium or low resolution map. In 9.3, the values were "high" "medium" "low". In 9.4, the values are as in the table. For example, "res=4" will display a map using all points with resolution value le 4.

10 =	28800 x	23040
9 =	14400 x	11520
8 =	6000 x	4800
7 =	2400 x	1800
6 =	1600 x	1200
5 =	1280 x	1024
4 =	800 x	600
3 =	640 x	480
2 =	400 x	300
1 =	320 x	240

5. **Data Changes.** The content of some variables have changed.

- ID variables are now character instead of numeric and are unique world wide because they contain the country code.
STATE & ID1 (US_COUNTIES)
STATE = 51; ID1 = US-51 (includes Country code)
- COUNTY & ID (US_COUNTIES)
COUNTY = 15; ID = US-51015 (includes Country and State)
- X and Y are always projected.
- LAT and LONG are always unprojected degrees.

<p>MAPSGFK.US_COUNTIES variables:</p> <ul style="list-style-type: none"> • ID – County code (char) • SEGMENT – ID segment number • X – Projected longitude coordinate vs unprojected radians • Y – Projected longitude coordinate vs unprojected radians • LAT – Unprojected degrees latitude • LONG – Unprojected degrees longitude (East) • DENSITY – Greduce density values • RESOLUTION – Map detail level based on output resolution • ID1 – States code (char) • STATE – Same as MAPS dataset • COUNTY – Same as MAPS data • LAKE – Lake Flag:1-water:2-citytype • STATECODE – Two letter state abbreviation 	<p>MAPSGFK.US_COUNTIES_ATTR variables:</p> <ul style="list-style-type: none"> • ID – County code (char) • IDNAME – County name • ISO – Country ISO number • ISONAME – Country name • STATE – Same as MAPS dataset • COUNTY – Same as MAPS data • STATECODE – Two letter state abbreviation • ID1 – State code (char) • ID1NAME – State Name
<p>AFGHANISTAN variables:</p> <ul style="list-style-type: none"> • ID – Districts code • SEGMENT – Map segment • X – Projected longitude • Y – Projected latitude • LAT – Unprojected latitude (Y) in Degrees • LONG – Unprojected longitude (X) in Degrees • DENSITY – GREDUCE density value • RESOLUTION – Similar to Density, but processed for display size • ID1 – Provinces code 	<p>AFGHANISTAN_ATTR variables:</p> <ul style="list-style-type: none"> • ID – Districts code • IDNAME – Districts name • ISO – ISO Country code • ISONAME – ISO Country name • ID1 – Provinces code • ID1NAME – Provinces name • ID1NAMEU – Unicode version of ID1NAME
<p>BOSNIA variables:</p> <ul style="list-style-type: none"> • ID – Municipalities code • SEGMENT – ID segment number • X – Projected longitude coordinate • Y – Projected latitude coordinate • LAT – Unprojected degrees latitude • LONG – Unprojected degrees longitude • RESOLUTION – Similar to Density, but processed for display size • DENSITY – GREDUCE density value • ID1 – Regions code • ID2 – Districts code 	<p>BOSNIA_ATTR variables:</p> <ul style="list-style-type: none"> • ID – Municipalities code • IDNAME – Municipalities name • ISO – ISO Country code • ISONAME – ISO Country name • ID1 – Regions code • ID2 – Districts code • ID1NAME – Regions name • ID1NAMEU – Unicode version of ID1NAME • ID2NAME – Districts name • COUNTRY – Short form of country name
<p>EUROPE variables:</p> <ul style="list-style-type: none"> • ID – Alpha2 country code • SEGMENT – ID segment number • X – Projected longitude: Albers • Y – Projected latitude: Albers • LAT – Unprojected degrees latitude • LONG – Unprojected degrees longitude • RESOLUTION – Similar to Density, but processed for display size • DENSITY – GREDUCE density value • IDNAME – Country name • LAKE – Lake Flag: 1-water: 2-citytype • ISO – ISO Country code 	<p>EUROPE_ATTR variables:</p> <ul style="list-style-type: none"> • ID – Alpha2 country • IDNAME – ISO country name • ISO – ISO Country code • ISONAME – ISO Country name • IDNAME_alt – Alternate Admin name • ISOALPHA2 – ISO Alpha2-code • ISOALPHA3 – ISO Alpha3-code • CONT – Numeric number

<p>EUROPE1 variables:</p> <ul style="list-style-type: none"> • ID – Admin1 code • SEGMENT – ID segment number • X – Projected longitude coordinate • Y – Projected latitude coordinate • LAT – Unprojected degrees latitude • LONG – Unprojected degrees longitude • RESOLUTION – Similar to Density, but processed for display size • DENSITY – GREduce density value • IDNAME – Admin1 name • ISO – ISO Country code • ISOALPHA2 – ISO Alpha2-code 	<p>EUROPE1_ATTR variables:</p> <ul style="list-style-type: none"> • ID – Alpha2 country • IDNAME – ISO country name • ISO – ISO Country code • ISONAME – ISO Country name • IDNAME_alt – Alternate Admin name • ADMINTYPE – Admin1 type • ISOALPHA2– ISO Alpha2-code • ISOALPHA3– ISO Alpha3-code • CONT– Numeric number
<p>EUROPE2 variables:</p> <ul style="list-style-type: none"> • ID – Admin3 code • SEGMENT – Map segment • X – Projected longitude • Y – Projected latitude • LAT – Unprojected latitude (Y) in Degrees • LONG – Unprojected longitude (X) in Degrees • RESOLUTION – Similar to Density, but processed for display size • DENSITY – GREduce density value • ISO – ISO Country code • ISOALPHA2 – ISO Alpha2-code • LAKE – Lake Flag: 1-water: 2-citytype 	<p>EUROPE2_ATTR variables:</p> <ul style="list-style-type: none"> • ID – “County” code • IDNAME – “County” name • ISO – Country code • ISONAME – Country name • ADMINTYPE – Admin1 type • ISOALPHA2– ISO Alpha2-code • ISOALPHA3– ISO Alpha3-code

MIGRATION ISSUES

Because there are differences between the old MAPS data and the new MAPSGFK data, there will be migration issues. Below are a few items to look for.

- Data set names may be different for example: VENEZUELA vs VENEZUEL
- Maps may contain lower/different levels of political data
- X, Y are always projected
- LAT/LONG contains unprojected coordinates in degrees, not radians. GPROJECT will require the “DEGREES” option.
- LAT/LONG are based on eastern Hemisphere. GPROJECT will require the ‘EASTLONG’ option.
- ID variables (ID, ID1, ID2) are character instead of numeric
- The values in the ID variables are now different and unique across the world
- Corresponding attribute datasets no longer have “2” suffix, but are named with “_ATTR”
- PROC GMAP will use a default RES= option, so all the points will not be displayed unless RES=NONE is specified.

EXAMPLES FOR MIGRATION

The following examples illustrate how to convert existing programs to use the new GfK map data sets.

EXAMPLE 1 – Changing the data set name

MAPS data set:

```
proc GMAP map=maps.afghanis data=maps.afghanis;  
  id id;  
  choro id / nolegend;  
run;
```

MAPSGFK data set:

```
proc GMAP map=mapsgfk.afghanistan data=mapsgfk.afghanistan;  
  id id;  
  choro id / nolegend;  
run;
```

Note that these examples use the map data for the response data. This insures that there is a value for each map area, but isn't useful in the real world. Figure 1 show the results from using the MAPS data set and Figure 2 shows the results from MAPSGFK. As you can see, MAPSGFK has a lower level of polygonal areas.

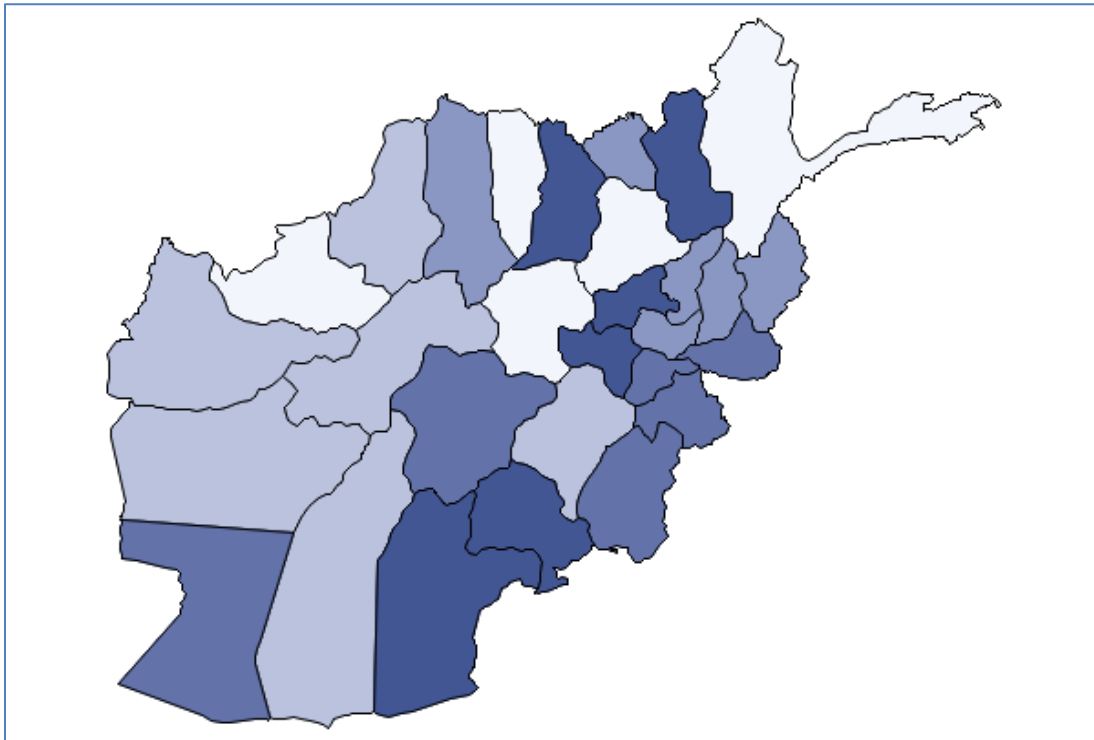


Figure 1 – Output from MAPS data set

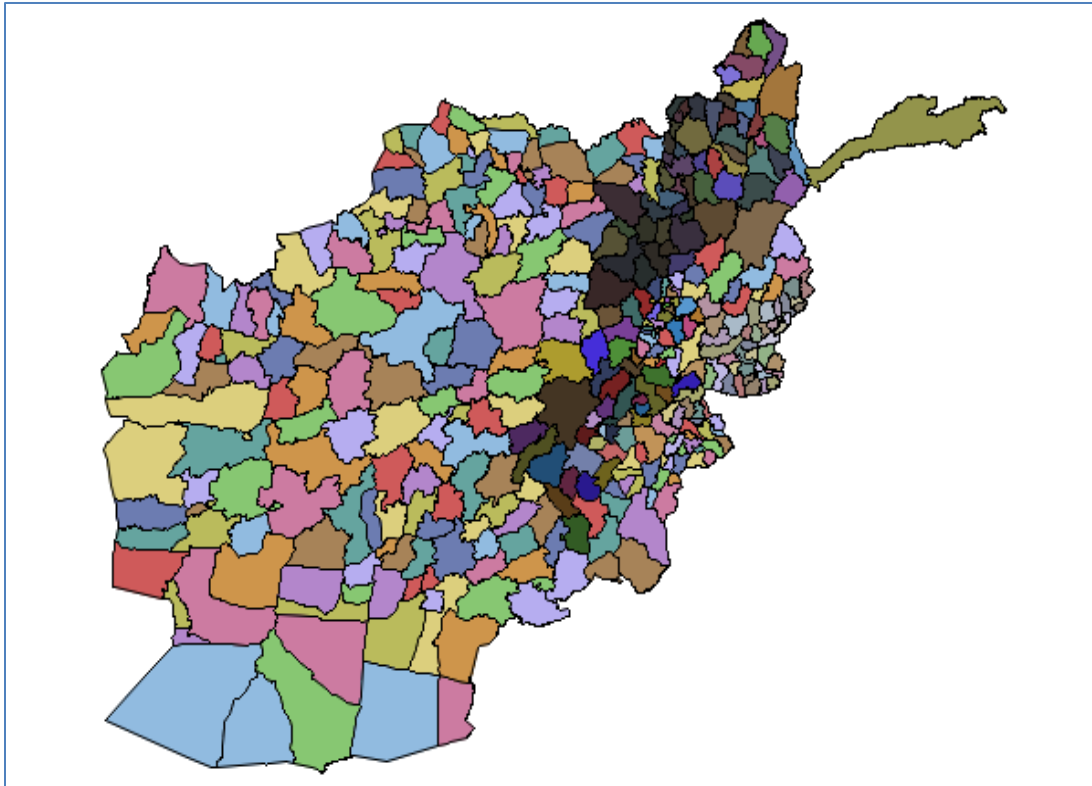


Figure 2 – Output from MAPSGFK data set

EXAMPLE 2 – Reducing the number of levels

In order to create the same map in Example 1, we need to modify the MAPSGFK program to reduce the number of map levels.

MAPSGFK data set:

```

/* Sort the map data by id1 */
proc SORT data=mapsgfk.afghanistan
  out=smymap; by id1 id; run;

/* remove the lowest level of the map data set */
proc GREMOVE data=smymap
  out=mymap; by id1; id id; run;

/* rename id1 to id */
data mymap (drop=id1); set mymap;
  id=id1; run;

proc GMAP data=mymap map=mymap;
  id id;
  choro id / nolegend;
run;

```

Figure 3 shows the output from this program which matches Figure 1.

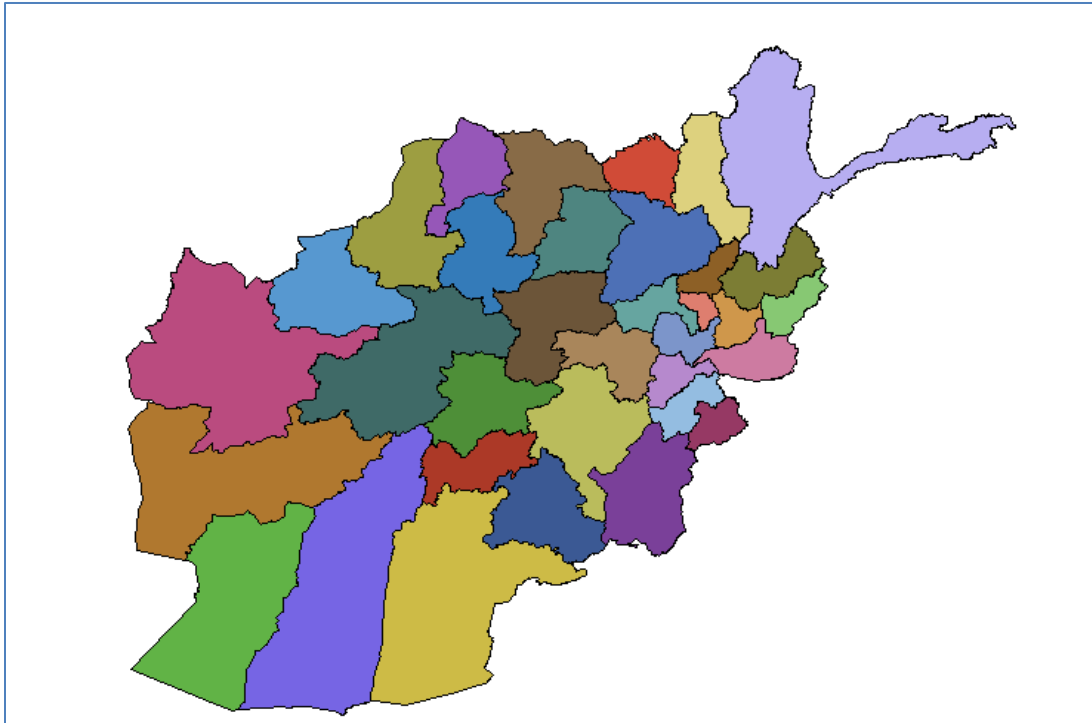


Figure 3 – GREDUCED output from MAPSGFK data set

EXAMPLE 3 – Fixing the Response Data ID values.

Examples 1 and 2 use the map data set for the response data. This causes the ID values to always match. But real programs need to use real response data. In order to do this, you need to make the response data match the ID values used by the map. The following code should be inserted into Example 2 before PROC GMAP.

MAPSGFK data set:

```

/* Subset the Attr file to only one copy of each ID1 value */
/* And rename ID1 to ID and ID1NAME to IDNAME */
/* Finally, uppercase the new IDNAME */
data mymap_attr(keep=ID IDNAME);
  length lastID1 $15; retain lastID1 '';
  set mapsgfk.afghanistan_attr;
  if (id1 ne lastID1) then do;
    lastID1=id1; id=id1; idname=upcase(idname); output;
  end;
run;

/* Read in the response data containing the uppercase NAME of each region*/
data af_vote; infile 'c:\SGF2013\superdemo\2004AFVote.dat';
  input idname $1-15 pervote ;
run;

/* Sort both the response data and the attr data by IDAME */
proc sort data=af_vote; by idname; run;
proc sort data=mymap_attr; by idname; run;

/* Merge the two data sets to add the ID value to the response data */
data mergresp; merge af_vote mymap_attr; by idname; run;

```


EXAMPLE 4 – Character vs. Numeric ID and using LEVELS=

The ID variables are numeric in MAPS and are character in MAPSGFK. If you use LEVELS=1, it will not work the same with MAPSGFK data sets and will be ignored. Another numeric value must be used instead.

MAPS data set:

```
/* want a single-colored map */  
proc gmap data=maps.europe map=maps.europe;  
  id id;  
  choro id / levels=1 nolegend ;  
run;
```

MAPSGFK data set:

```
/* need a numeric variable*/  
proc gmap data= mapsgfk.europe map=mapsgfk.europe;  
  id id segment;  
  choro segment / levels=1;  
run;
```

Figure 4 shows the output from MAPS and Figure 5 shows the output from MAPSGFK. Notice that the map data contains different areas in MAPS and MAPSGFK as indicated by the arrows.

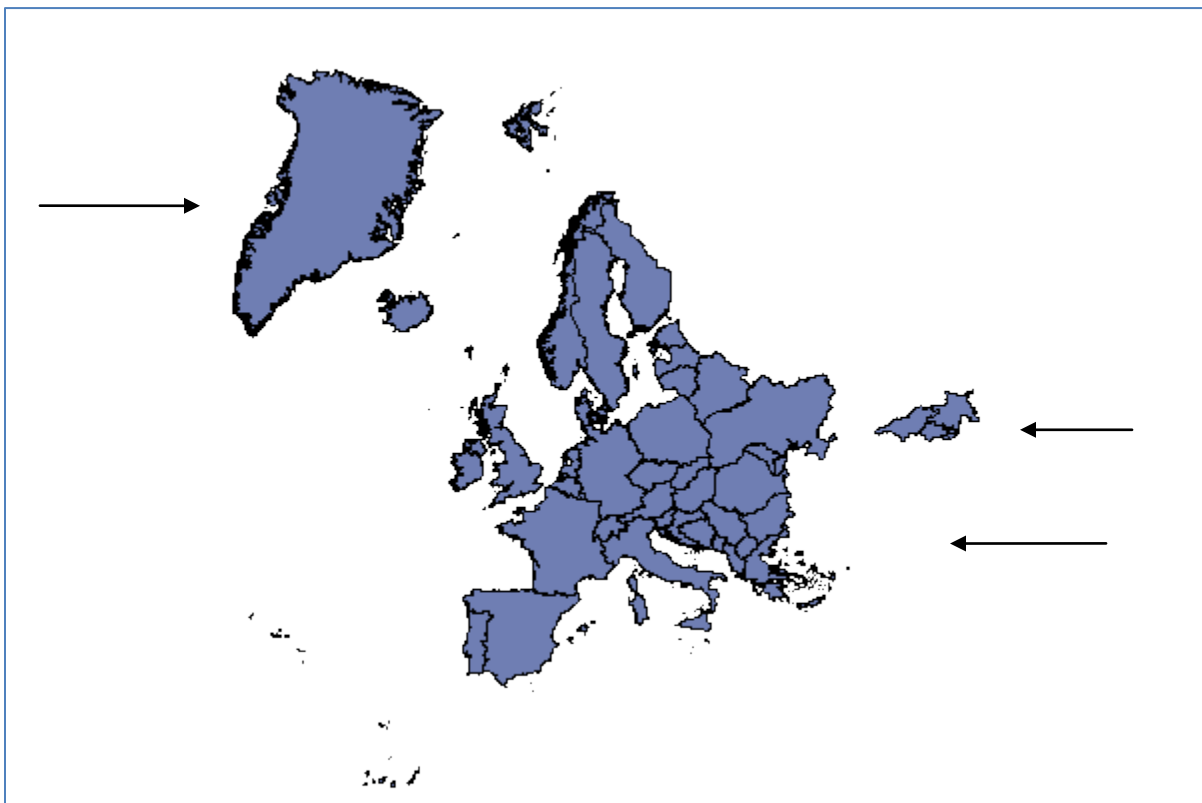


Figure 4 – LEVELS=1 output with MAPS data sets

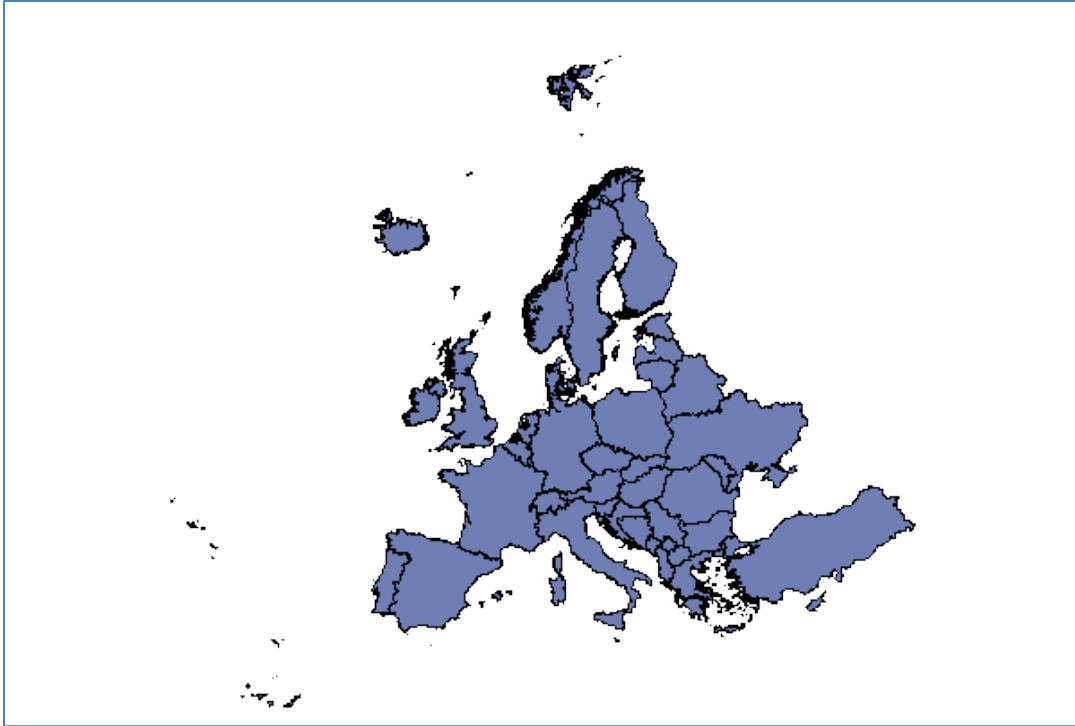


Figure 5 – LEVELS=1 output with MAPSGFK and modified program

EXAMPLE 5 – Projection Differences

In MAPS data sets, there was inconsistency regarding which values were projected. In the US data, X and Y were unprojected in the past. In MAPSGFK, X and Y are always projected. In this example, changing the data set name will leave with a difference in projection.

MAPS data set:

```
proc GMAP map=maps.states data=maps.states;
  id state; /* no id variable */
  choro state / nolegend;
run;
```

MAPSGFK data set:

```
proc GMAP map=mapsgfk.us_states data=mapsgfk.us_states;
  id id;
  choro id / nolegend;
run;
```

Figure 6 shows that MAPS data will appear unprojected and backwards. And Figure 7 shows that the X and Y values will give you a projected map. Notice that there are also some differences in the polygonal areas contained in the maps as indicated by the arrow.

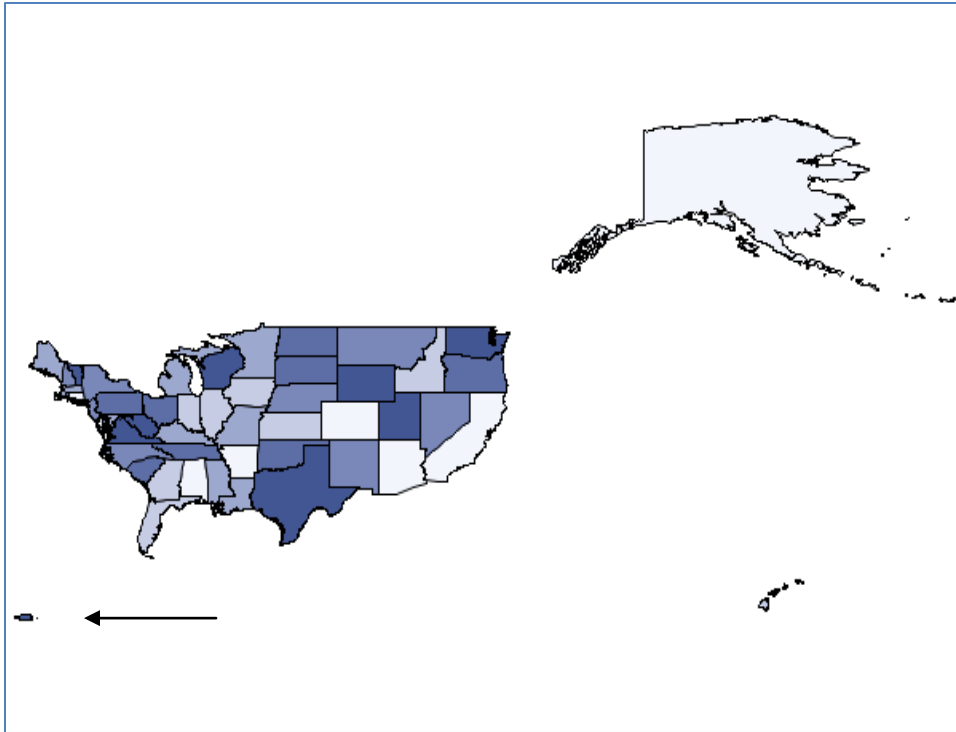


Figure 6 – MAPS data is unprojected

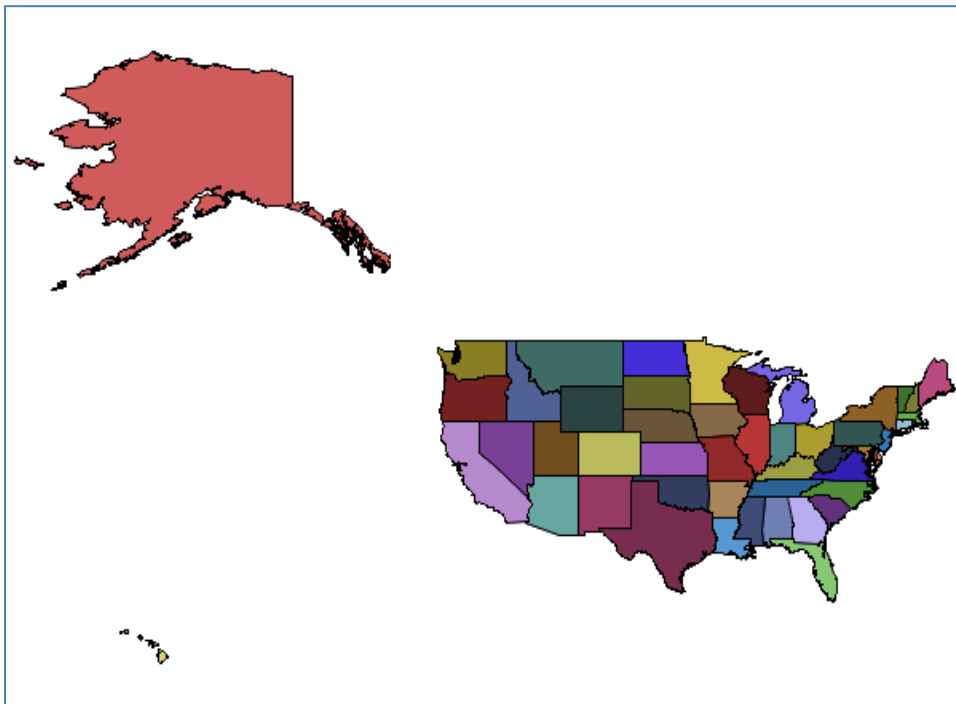


Figure 7 – MAPSGFK is projected

EXAMPLE 6 – GPROJECT usage

Usually PROC GPROJECT was used with MAPS data. This example shows how to convert code using PROC GPROJECT.

MAPS data set:

```
proc GPROJECT data=maps.states
  out=mystates project=miller2;
  id state;
run;

proc GMAP data=mydata
  map=mystates;
  id state;
  choro state / nolegend;
run;
```

MAPSGFK data set:

```
proc GPROJECT data=mapsgfk.us_states
  out=mystates project=miller2
  degrees eastlong latlong;
  id state;
run;

proc GMAP data=mydata
  map=mystates;
  id state;
  choro state / nolegend;
run;
```

Note that because the MAPSGFK data is in degrees and uses the EAST longitude, you must add the DEGREES and EASTLONG options to PROC GPROJECT. The LATLONG option is also used to cause the PROC to get the unprojected coordinates from the LAT and LONG variables instead of X and Y.

EXAMPLE 6B – Using the default projection

Since X and Y are already projected in MAPSGFK data, another way to solve this problem is to use the default projection and remove the PROC GPROJECT code.

MAPSGFK data set:

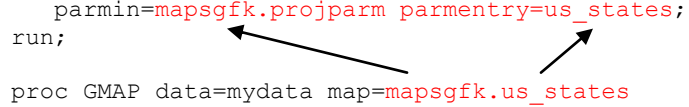
```
/* Notice, GPROJECT was removed */
proc GMAP data=mydata map=mapsgfk.us_states;
  id state;
  choro state / nolegend;
run;
```

EXAMPLE 6C – Projecting Anotate data

In the past, when you had annotate data, you had to combine it with the MAPS data set and project both data sets together. If you are using MAPSGK data, you don't have to re-project the map data if you are using the default projections. This is because a dataset (PROJPARM) is created and saved with each MAPGFK data set as it is projected and this information can be referenced to project the annotate data exactly the same.

MAPSGFK data set:

```
/* ONLY project the annotate date */  
proc GPROJECT data=anno out=annop  
  parmin=mapsgfk.projparm parmentry=us_states;  
run;  
  
proc GMAP data=mydata map=mapsgfk.us_states  
  anno=annop;  
  id state;  
  choro state / nolegend;  
run;
```



CONCLUSION

In this paper, we described the main features and some of the “red-flags” necessary to start using the new Map Data. We hope the examples were sufficient in helping you make the transition from using the old data to the new. We are always open to receiving your feedback and suggestions.

RESOURCES

SAS Maps Online Web Site: <http://support.sas.com/maponline>

For additional maps visit the SAS landing page at the GfK Web Site GfK-GeoMarketing.com/SAS

SAS/Graph Documentation: <http://support.sas.com/rnd/datavisualization/>

SAS Online Documentation: <http://support.sas.com/documentation/>

SAS Customer Support: <http://support.sas.com>

CONTACT INFORMATION

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

Liz Simon
SAS Institute
SAS Campus Drive
Cary, NC 27513
Liz.Simon@sas.com

Darrell Massengill
SAS Institute
SAS Campus Drive
Cary, NC 27513
Darrell.Massengill@sas.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.