

Using SAS/Graph to Map Geographic Data

Presented to Southeast SAS Users' Group, February 15, 1993
by Jay Jacob Wind, Vice-President, American Environmental Institute ¹

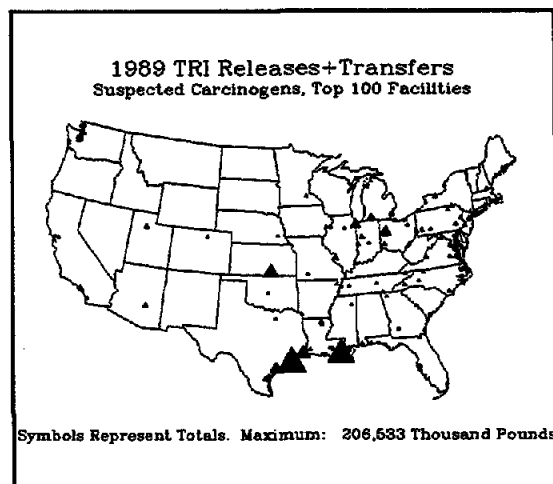
The SAS System provides tools to create excellent maps, including GMAP, GPROJECT, G3D, GPLOT, GCONTOUR, GREMOVE, GREduce, and G3GRID.

SAS tools can map boundaries, e.g., national, provinces, states, counties; points by latitude-longitude; lines, e.g., rivers or streets; gradients of elevation or temperature; and measurements for any of these.

This paper illustrates the following tools: building a thematic map from scratch; converting between coordinate systems; calculating distances between points; orienting a map using a bounding rectangle; adding shapes, sizes, and symbols to show locations; excluding/including Alaska and Hawaii with the lower 48 states; using data in a footnote; enclosing a map in a reference frame with tickmarks for the latitudes and longitudes; and downloading graphics into WordPerfect.

An Example

The map on the right, presented by US EPA at the 1992 international conference of the Organization of Economic Co-operation and Development (OECD), Vienna, Austria, illustrates some of the tools we discuss in this paper. The map excludes Alaska and Hawaii; is reduced to an acceptable level of detail; is pinned to its bounding rectangle to eliminate outliers; is projected to represent equal areas nationwide; is annotated with symbols; and uses data in its footnote.



Building A Thematic Map

Thematic maps use shades or patterns to show values of a variable by nation, province, state, or county. SAS provides map outlines for US, STATES, COUNTIES, COUNTIES, USCENTER, USCITY, USCOUNTY, AFRICA, ASIA, EUROPE, NAMERICA, SAMERICA, WORLDMPJ, CANADA, and more than 80 countries. For a list, use this tool:
PROC CONTENTS DATA=MAPS._ALL_ ; RUN;

How SAS Stores Boundary Data

Boundary files contain an ID, e.g., STATE Federal Information Processing Standard (FIPS) (1=Alabama, 2=Alaska, etc.); a series of X,Y pairs for longitude and latitude; and DENSITY on a 1-to-6 scale (1 = most important).

Creating Data For A Thematic Map

Make a dataset with each state's code (ST), e.g., MD or VA, and a variable, e.g., SALES. translate ST into FIPS codes, use this tool:

```
DATA ONE; SET MY.DATA;  
STATE=STFIPS(ST); RUN;
```

¹ 611 South Ivy Street, Arlington, Virginia 22204 (703) 920-5193
To request copies of the slides accompanying this paper, please
send eMAIL to JWW@NCCIBM.RTPNC.EPA.GOV.

County Names

A county dataset contains FIPS state and county codes and a response variable. For a list of FIPS county codes, use this tool:

```
PROC PRINT DATA=MAPS.CNTYNAME;
  VAR STATE COUNTY COUNTYNM; RUN;
```

How SAS Link Data And Map Files

PROC GMAP automatically links data and map files. Both must contain X, Y, and the ID. Use this tool:

```
PROC GMAP DATA=ONE MAP=MAPS.US ALL;
  ID STATE; CHORO SALES; RUN;
```

How Map Detail Is Controlled

Use DENSITY to filter out high-detail points. MAPS.US (by state) and MAPS.USCOUNTY (by county) don't include DENSITY, but MAPS.STATES and MAPS.COUNTIES do. For state maps with just 30% of the detail, use this tool:

```
DATA MAP; SET MAPS.STATES;
  IF DENSITY LE 3; RUN;
```

For county maps with just 16% of the detail, use this tool:

```
DATA MAP; SET MAPS.COUNTIES;
  IF DENSITY LE 3; RUN;
```

Reducing Detail Is Not Enough

MAPS.US and MAPS.USCOUNTY are pre-projected, but MAPS.STATES and MAPS.COUNTIES aren't. Projected map files don't contain lat-longs. Instead, they contain points "unrolled" onto a flat sheet of paper. Since Earth's surface is round, if you simply plot lat-longs, you see increasing distortion as you move toward the poles. To adjust for Earth's curvature, use this tool:

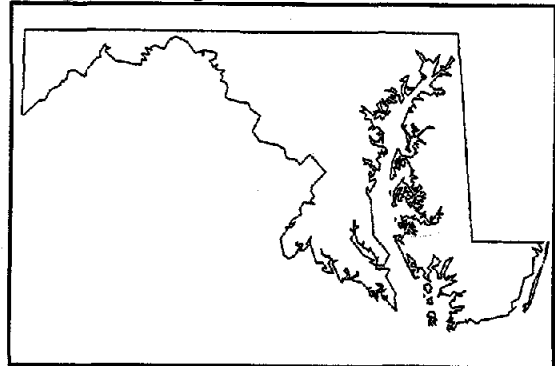
```
PROC GPROJECT DATA=MAPS.STATES
  OUT=MAP; ID STATE; RUN;
```

Then you can use:

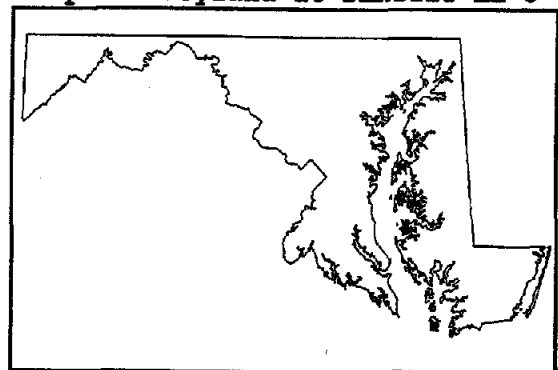
```
PROC GMAP DATA=MAP MAP=MAP ALL;
  ID STATE; CHORO STATE; RUN;
```

The two maps of Maryland on the right illustrate DENSITY levels of 3 and 6.

Map of Maryland at DENSITY LE 3



Map of Maryland at DENSITY LE 6



Patterns

By default, PROC GMAP divides responses into five groups by rounding up the top value and dividing it by five. Also by default, PROC GMAP's legend prints only group midpoints. To overcome these defaults, use PROC FORMAT. Assume you have sales of \$100,000 - \$800,000 per county, and you want four groups. Use this tool:

```
PROC FORMAT; VALUE S
  LOW-200000='UNDER $200K'
  200000-400000='$200K-400K'
  400000-600000='$400K-600K'
  600000-800000='$600K-800K';
PROC GMAP DATA=ONE MAP=MAPS.US;
  ID STATE; CHORO SALES/DISCRETE;
  FORMAT SALES S.; RUN;
```

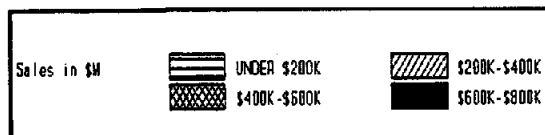
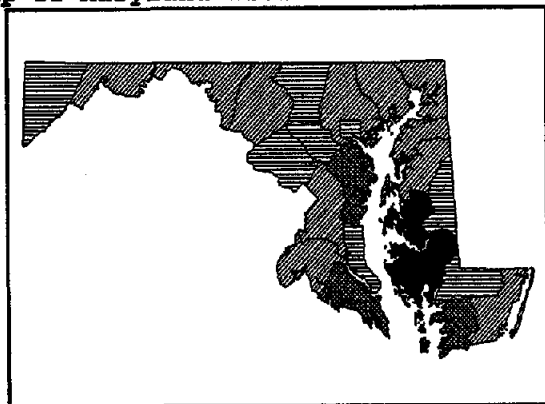
How to Control Patterns

After years of research, we know the four best patterns. Use this tool:

```
PROC GMAP DATA=ONE MAP=MAPS.US;
  ID STATE; CHORO SALES;
  FORMAT SALES S.;
  PATTERN1 V=M3N00; PATTERN2 V=M4N45;
  PATTERN3 V=M4X45; PATTERN4 V=MS;
  RUN;
```

The map of Maryland below shows our favorite patterns.

Map of Maryland with Favorite Patterns



Converting Between Coordinate Systems

Most non-SAS map files store unprojected lat-long data in terms of degrees. While some lat-long files store lat-longs as decimal degrees, e.g., 38.500 = 38½ degrees, many use degrees, minutes, and seconds (DDMMSS). If you encounter DDMMSS format, use this tool:

```
DEGREES=SUBSTR(DDMMSS,1,3)
+SUBSTR(DDMMSS,4,2)/60
+SUBSTR(DDMMSS,6,2)/3600;
```

Radians

SAS map files use RADIANS instead of degrees, where 1 radian = 180/PI degrees = 57.296 degrees. To figure PI to 10 decimals, use this tool:

```
PI=ARCOS(-1);
```

To convert DEGREES to RADIANS, use:

```
X=LONG*PI/180; Y=LAT*PI/180;
```

For RADIANS to DEGREES, use:

```
LONG=X*180/PI; LAT=Y*180/PI;
```

UTMs and Other Systems

Many databases use Universal Transverse Mercator coordinate systems or state grids. We have developed numerous algorithms to convert UTM and state grids into lat-longs.

Calculating Distances Between Points

Since Earth's is about 25,000 miles around, 1 degree at the equator = 69.4 mi. (25,000 mi./360 degrees).

Latitude distance is constant, but longitude varies from 1 degree = 69.4 mi. at the equator to 0 mi. at the poles. Longitude distance decreases with COSINE(Latitude). To figure miles between LONG2, LAT2 and LONG1, LAT1 (in radians), use this tool:

```
DISTANCE=
69.44 * SQRT ((LAT2-LAT1)**2
+ (COS(LAT1)*(LONG2-LONG1))**2);
```

Orienting A Map In A Rectangle

Frequently, point datasets contain outliers beyond their reported area. To exclude points outside the rectangle bounding an area, e.g., for Maryland, use this tool:

```
DATA MAP; SET MAPS.STATES;
IF FIPSTATE(STATE)='MD';
PROC SUMMARY DATA=MAP MISSING NWAY;
VAR X Y;
OUTPUT OUT=ANNO MIN(X)=MIX
MAX(X)=MAX MIN(Y)=MIY MAX(Y)=MAY;
DATA fixed;
RETAIN PI MIX MAX MIY MAY;
SET original;
IF N =1 THEN DO;
PI=ARCOS(-1); SET ANNO;
END;
X=LONG*PI/180; Y=LAT*PI/180;
IF MIX<=X<=MAX AND MIY<=Y<=MAY
THEN OUTPUT;
```

Excluding/Including Alaska, Hawaii, Puerto Rico, And Virgin Islands

MAPS.US and MAPS.STATES contain Alaska and Hawaii. In MAPS.US, their coordinates are shifted and scaled to appear below the California coast. In MAPS.STATES, their lat-longs are real; Alaska is nearly as wide east-to-west as the lower 48 together. For a map excluding Alaska and Hawaii, use this tool:

```
DATA MAP; SET MAPS.STATES;
IF FIPSTATE(STATE)
NOT IN ('AK','HI','RQ','VQ');
```

Shifting Alaska, Hawaii, Puerto Rico, And Virgin Islands To Fit On A Map

```
Use this tool:
IF ST='AK' THEN DO;
LAT=(LAT-51.6)/4+44;
LONG=(LONG-129.9)/8+124.5; END;
IF ST='HI' THEN DO;
LAT=LAT+12.5; LONG=LONG-36; END;
IF STATE>70 THEN DO;
LAT=LAT+22; LONG=LONG+11; END;
```

Adding Shapes, Sizes, And Symbols

To display locations, e.g., toxic waste dumps, you need each site's lat-long. Turn your data into an ANNOTATE file, with the variables: X, Y, XSYS, YSYS, TEXT, SIZE, COLOR, STYLE, FUNCTION, POSITION. After years of research, we have found the best map symbol is TEXT'L' STYLE'SPECIAL' (▲). If we use COLOR, we also use TEXT'J' STYLE'SPECIAL' (=) and TEXT'K' STYLE'SPECIAL' (*). For SIZE, we use minimum of 1 and maximum of 8. We sort to find the maximum, normalize by dividing each value by the maximum, then standardize by multiplying each fraction by 8 and adding the minimum of 1.

To produce the map on the front page, we used this tool:

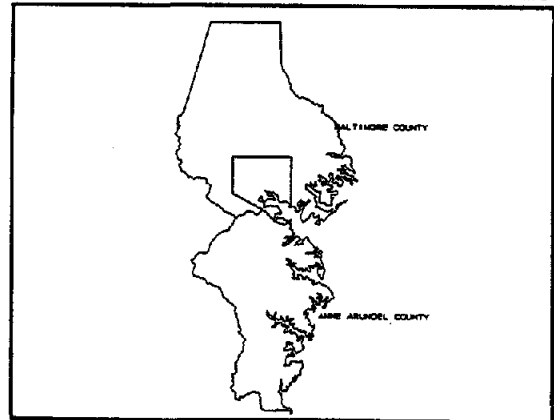
```
* Create base map.;
DATA MAP; RETAIN MAP 1;
  SET MAPS.STATES; IF DENSITY LE 3;
  IF FIPSTATE(STATE) NOT IN
    ('RQ','HI','AK');
* Create bounding rectangle.;
PROC SUMMARY DATA=MAP NWAY; VAR X Y;
  OUTPUT OUT=BOX MIN(X)=MIX
    MIN(Y)=MIY MAX(X)=MAX MAX(Y)=MAY;
* Find top 100 sites.;
PROC SUMMARY DATA=data MISSING NWAY;
  CLASS LAT LONG; VAR KLBY;
  OUTPUT OUT=ONE SUM=;
* Sort top 100 + degrees to radians.;
PROC SORT DATA=ONE;
  BY DESCENDING KLBY;
DATA ONE; RETAIN PI MIX MAX MIY MAY;
  SET ONE; IF LAT AND LONG;
  IF N =1 THEN DO;
    PI=ARCOS(-1); SET BOUNDS;
  END;
  X=LONG*PI/180; Y=LAT*PI/180;
  IF MIX<=X<=MAX AND MIY<=Y<=MAY;
  N+1; IF N LE 100; STATE=N;
* Calculate the max and footnote.;
PROC SUMMARY DATA=ONE MISSING NWAY;
  VAR KLBY; OUTPUT OUT=MAX MAX=;
DATA NULL; SET MAX;
CALL SYMPUT('K',PUT(KLBY,COMMA9.));
CALL SYMPUT('MAX',KLBY); RUN;
FOOTNOTE "Max: &K Thousand Lbs.";
* Create a bounding box for the map.;
DATA BOX; RETAIN MAP 2; SET BOX;
X=MIX; Y=MIY; STATE=99001; OUTPUT;
X=MAX; Y=MIY; STATE=99002; OUTPUT;
X=MAX; Y=MAY; STATE=99003; OUTPUT;
X=MIX; Y=MAY; STATE=99004; OUTPUT;
* Put them together and project.;
DATA ONE; SET ONE TWO MAP;
PROC GPROJECT OUT=ONE; ID STATE;
DATA MAP; SET ONE; IF MAP=1;
```

```
DATA ANNO(DROP=DENSITY SEGMENT STATE);
  LENGTH FUNCTION $8;
  RETAIN WHEN'B' POSITION'5' TEXT'L'
    XSYS YSYS'2' FUNCTION'SYMBOL'
    STYLE'SPECIAL' COLOR'RED';
  SET TWO; IF MAP=0;
* Compute size of each triangle.;
  SIZE=1+8*KLBY/&MAX;
* Generate the map.;
PROC GMAP DATA=MAP MAP=MAP ALL;
  ID STATE; CHORO STATE/LEVELS=1
  DISCRETE NOLEGEND ANNO=TWO; RUN;
```

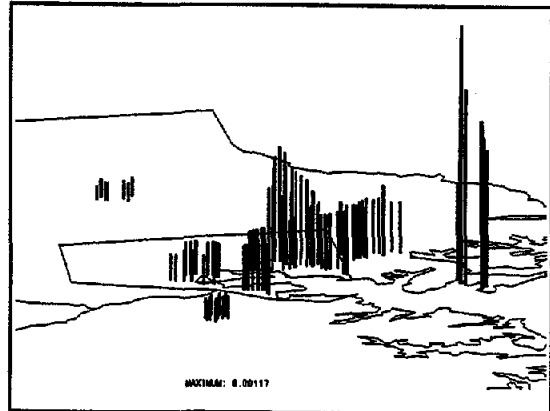
Using Boundary Data In Annotate Files

You can use boundary data in annotate files to take advantage of PROC GPLOT's SCATTER option. In 1986, US EPA studied air pollution health risk near Baltimore and used SAS/GRAPH to illustrate results, as below.

Baltimore and Anne Arundel Counties



Air Pollution Risk Near Baltimore
The Vertical Bars Represent Relative Estimated Excess 70-Year Cancer Risk From All Pollutants From All Sources At Various Urban And Suburban Sites



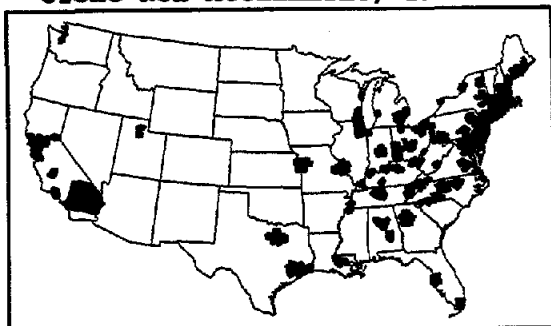
Mapping Counties Only With Data

To map counties only where data actually appears, use this tool:

```
PROC FREQ DATA=MYDATA;
  TABLES STATE*COUNTY/NOPRINT OUT=ONE;
DATA MAP2;
  MERGE ONE (IN=ONE) MAPS.COUNTIES;
  BY STATE COUNTY;
  IF ONE;
DATA MAP; SET MAPS.STATES MAP2;
PROC GPROJECT DATA=MAP OUT=MAP;
  ID STATE COUNTY;
PROC GMAP DATA=ONE MAP=MAP ALL;
  ID STATE COUNTY; CHORO SALES;
RUN;
```

For instance, US EPA uses maps like this one annually to identify counties failing to attain ambient Ozone standards.

Ozone Non-Attainment, 1987-89



Downloading Graphics Into WordPerfect

You can use WordPerfect to view and print graphics. First, program them. Then create a fixed-block 80-column mainframe file. Then use either device HP7475A, CGMWP (portrait mode), or CGMWPL (landscape mode). Then use this tool:

```
FILENAME GSASFILE 'yourid.yourfile';
GOPTIONS DEVICE=CGMWP
GACCESS=GSASFILE
GSFNAME=GSASFILE GSFLen=80
GSFMODE=REPLACE HANDSHAKE=NONE;
PROC GMAP ...; RUN; QUIT;
```

GSASFILE now contains binary code. Download it, e.g., with Kermit, to a PC file named GSASFILE. To import into WordPerfect, use this tool:

```
WP <Enter> <Alt>--<F9> 1 1 --graphics;
6 C --to center;
1 GSASFILE <Enter> --and wait;
9 --to view your map or edit it.
```

Enclosing A Map In A Reference Frame

Biologists evaluate biodiversity using species location maps. Plotting the distribution of two or more species and making visual inferences requires distinct symbols for each.

When you draw a map, you can document its location on the globe by marking latitudes and longitudes as reference points. Scientific journals will not accept maps without latitude and longitude reference marks, and serious documentation must include those data. Reference marks are especially important for local maps that lack recognizable orientation, such as an area within a county.

With the assistance of Dr. Bernard Baum of Agriculture Canada, we have written a macro with the following functions:

- o Ask for four coordinates anywhere on the globe
- o Draw a reference frame connecting the four coordinates
- o Bend longitudes away from equator
- o Project latitudes to represent the Earth's curvature
- o Mark lat-long ticks along the frame
- o To trim borders crossing the frame, use the tool:

```
PROC GPROJECT DATA=ONE OUT=ONE
  MINLAT=y1 MAXLAT=y2
  MINLONG=x1 MAXLONG=x2;
  ID ID;
```

- o Label all countries in the frame
- The 300-line macro is available from the author via SAS-L. An example is shown below.

France and Spain - with Tickmarks

