

Paper RIV-12

Creating ZIP Code-Level Maps with SAS®

Barbara B. Okerson, WellPoint

ABSTRACT

SAS®, SAS/GRAPH®, and ODS graphics provide SAS programmers with the tools to create professional and colorful maps. Provided with SAS/GRAPH are boundary files for U.S. states and territories, as well as internal boundaries at the county level. While much data and results can be displayed at this level, often a higher degree of granularity is needed. The U.S. Census Bureau provides ZIP code boundary files in ESRI shapefile format (.shp) by state for free download and import into SAS using SAS PROC DATAIMPORT. This paper illustrates the use of these ZIP code tabulation area (ZCTA) files with SAS to map data at a ZIP code level. Example maps include choropleth, distance, and heat maps.

Examples included in this paper were created with version 9.2 of SAS on a Windows 64-bit server platform and use Base SAS, SAS/STAT and SAS/GRAPH. The techniques represented in this paper require a SAS/GRAPH license but are not platform-specific and can be adapted by beginning through advanced SAS users.

INTRODUCTION

ZIP codes became mandatory for use on U.S. mail in 1963. ZIP is an acronym for Zone Improvement Plan. The codes were designed to improve the efficiency of mail distribution and delivery. The U.S. Postal Service had implemented postal zones for larger cities beginning in the 1940s and the success of this program led directly to the development of ZIP codes. While postal employee Robert Moon is often given credit for the development of the ZIP code, his plan only led to the first three or regional digits of the ZIP code. In 1983, the U.S. Postal increased granularity of the ZIP code with an expanded system called "ZIP +4" codes.

SAS users are often familiar with the SASHELP.ZIPCODE file that SAS provides containing ZIP code level information for the United States, including ZIP code centroids, area codes, city names and FIPS codes. The file is indexed on ZIP code and is updated regularly by SAS. While this information can be used for marking the center of the ZIP code area, without defined boundaries ZIP code level choropleth maps are not possible. ZIP codes of themselves do not identify a defined geographic location; rather they define mailing addresses associated with a particular mail delivery station. The U.S. Census Bureau has created pseudo boundary files that allow choropleth ZIP code mapping. ZIP Code Tabulation Areas or ZCTAs are generalized spatial representations of the service areas covered by the United States Postal Service ZIP codes.

ZIP CODE TABULATION AREAS

To create ZCTAs from ZIP codes, the Census Bureau collected all of the addresses within each census block to define the list of ZIP codes by block. As a next step, the most frequently occurring ZIP code within each block was assigned to the block as a preliminary ZCTA code. After all of the census blocks with addresses were assigned a preliminary ZCTA code, blocks were aggregated and combined by code to create larger areas. Blocks that contained addresses without a most frequently occurring ZIP code were assigned to the ZCTA of the block with which it shared its longest boundary. Unassigned areas of less than two square miles were assigned to the surrounding ZCTA. For any remaining unassigned blocks, block group boundaries were used by the Census Bureau to identify and group these blocks. They were then also merged into an adjacent ZCTA with the longest shared boundary.

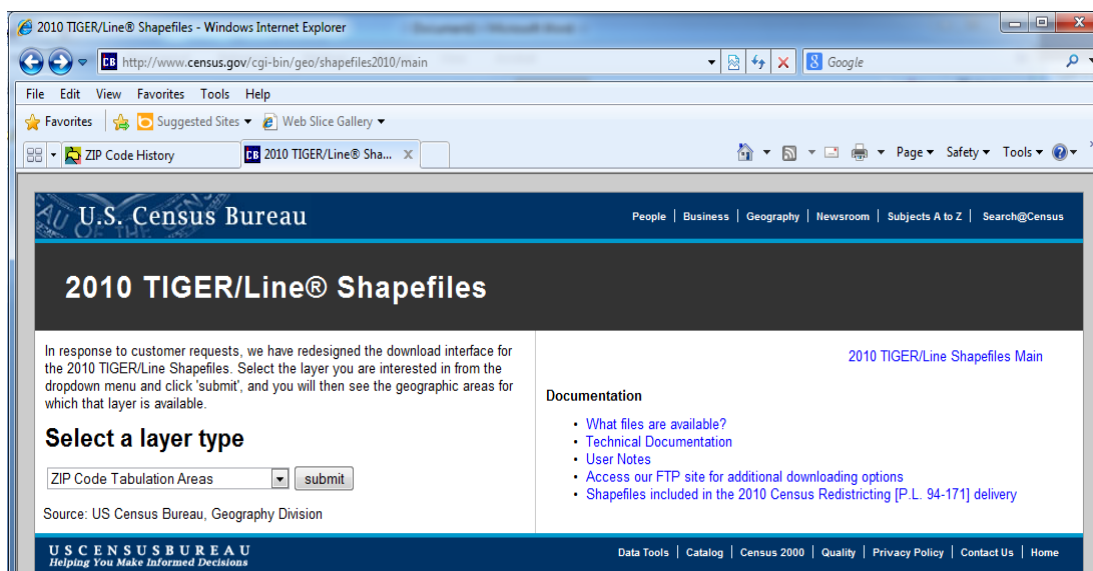
The first year that ZCTAs were available was 2000. For the 2000 ZCTAs the Census Bureau created ZCTAs that ended in "XX" to represent large land areas without addresses or ZIP Codes and "HH" to represent large water areas without ZIP Codes. For the current, or 2010, ZCTAs, these areas were not included. Only residential and non-residential ZIP codes from the Census Bureau's MAF/TIGER (Master Address File/Topologically Integrated Geographic Encoding and Referencing) database were included. In addition, ZIP codes assigned to single delivery point addresses such as a business or school will not necessarily appear as ZCTAs.

Most ZCTA codes are the same as the area's ZIP code. But because the Census Bureau used the most frequently occurring ZIP Code in an area for the ZCTA code some addresses will end up with a ZCTA code that differs from its ZIP Code. For the 2010 Census, there are 33,120 ZCTAs.

ZCTA cartographic boundary files are available from the Census Bureau for free download by state. These are ESRI format shapefiles that can be imported into SAS. To download these files, open the following link:

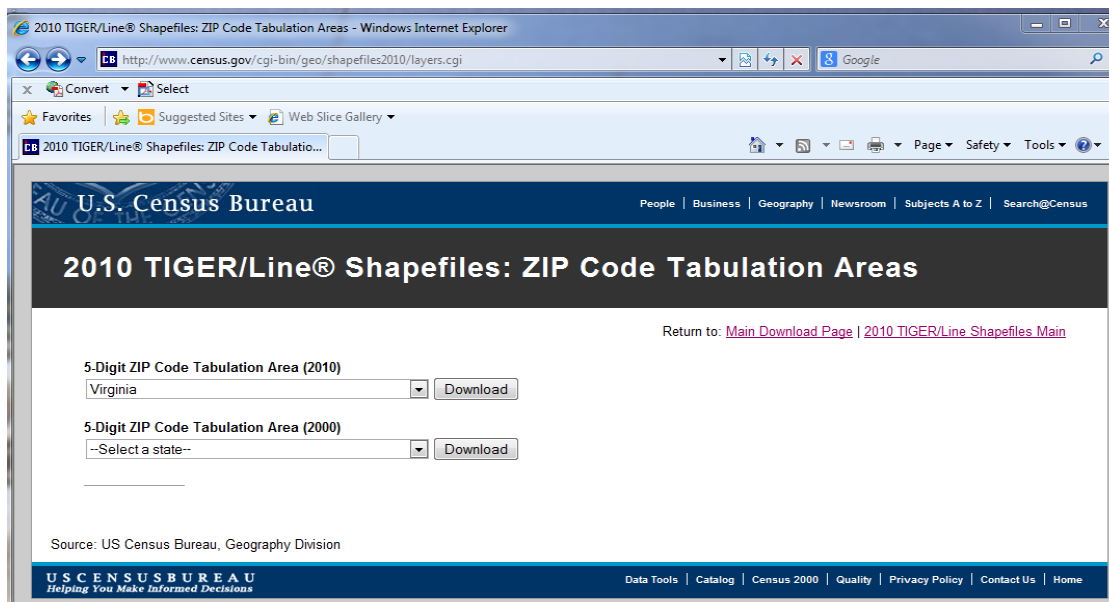
<http://www.census.gov/cgi-bin/geo/shapefiles2010/main>

Then select “Zip Code Tabulation Areas” as illustrated below:



Display 1. Shapefile Layer Selection Screen.

Then select the state that you are planning to map:



Display 2. ZCTA file download screen.

The downloaded file is a compressed file that contains five files which can be extracted into a directory with WinZip or other compressed file extraction software. These files (for Virginia) are listed below. The number ‘51’ in the filenames represents Virginia’s FIPS (Federal Information Processing Standard) number.

- tl_2010_51_zcta510.dbf
- tl_2010_51_zcta510.prj
- tl_2010_51_zcta510.shp
- tl_2010_51_zcta510.shp.xml
- tl_2010_51_zcta510.shx

Only three of these files are required for SAS to be able to import the shapefiles: .DBF, .SHP, and .SHX. These files are designed for small scale thematic mapping. The examples that follow illustrate this use.

SAS PROC MAPIMPORT

The MAPIMPORT procedure does not create any graphics output. The purpose of the procedure is to enable you to import ESRI format shapefiles into SAS/GRAPH map data sets. This procedure only works with ESRI shapefiles.

The code for importing is very straightforward:

```
proc mapimport datafile="r:\tl_2010_51_zcta510.shp" out=va_zshp;
run;
```

While this discussion is for use of MAPIMPORT to import Census Bureau ZCTA files, the MAPIMPORT procedure can be used to import any file in the ESRI shapefile format. While not needed for Census ZCTA files, options that are available for the MAPIMPORT procedure include:

- EXCLUDE – specifies fields to be excluded from SAS map data set
- ID – reorders map polygons by identifier field
- RENAME - renames fields in output data set
- SELECT – selects the fields to be included in the output map data set

EXAMPLE ONE – DISPLAYING CENSUS DATA

Along with the boundary files, gazetteer files containing population and housing regions are available for counties, county subdivisions, named places, census tracts, 111th Congressional districts, American Indian Areas, Alaska Native Areas, Native Hawaiian Areas, school districts, state legislative districts, ZCTAs and urban areas for the 50 states, the District of Columbia, and Puerto Rico. The 2010 files will not provide a one-to-one match with entities from the 2000 or 1990 census due to changed boundaries and creation and dissolution of some geographic entities since the earlier censuses.

The example below uses ZCTAs. The ZCTA gazetteer file, like all other gazetteer files, is a single file containing data for all 50 states, the District of Columbia and Puerto Rico as of the January 1, 2010. The file is ASCII text, one line per record. File layout is as follows:

Column	Label	Description
Column 1	GEOID	Five digit ZIP Code Tabulation Area Census Code
Column 2	POP10	2010 Census population count.
Column 3	HU10	2010 Census housing unit count.
Column 4	ALAND	Land Area (square meters) - Created for statistical purposes only
Column 5	AWATER	Water Area (square meters) - Created for statistical purposes only
Column 6	ALAND_SQMI	Land Area (square miles) - Created for statistical purposes only
Column 7	AWATER_SQMI	Water Area (square miles) - Created for statistical purposes only
Column 8	INTPTLONG	Longitude (decimal degrees) First character is blank or "-" denoting East or West longitude respectively
Column 9	INTPTLAT	Latitude (decimal degrees) First character is blank or "-" denoting North or South latitude respectively

Table 1. ZCTA Gazetteer 2010 file layout

The link to the Census Bureau Gazetteer files is: <http://www.census.gov/geo/maps-data/data/gazetteer2010.html>.

The following example uses the Virginia subset of the ZCTA gazetteer file to map 2010 population location by ZIP code. Twenty-five levels are created using pattern statements. State, county, and ZIP code boundary lines are removed so that natural boundaries by level are formed; adjacent ZIP codes within the same level appear as a single entity. A three-color pattern series is set: green (low), yellow (neutral), and red (high). This color scale is adapted from Darrell Massengill's "SAS Mapping: Technologies, Techniques, Tips, and Tricks" (see references).

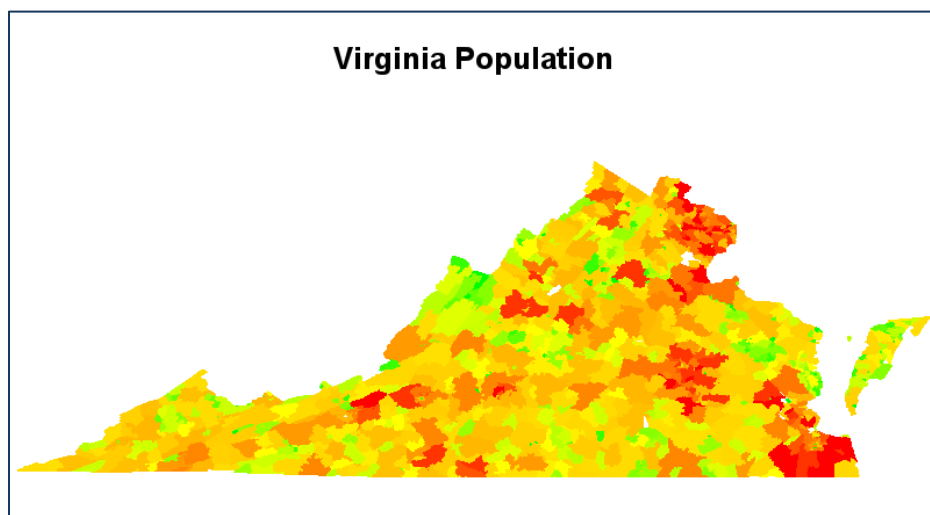


Figure 1. ZIP Code level map of Virginia 2010 population

EXAMPLE ONE SAS CODE:

SAS code for the above map is listed below. The blank footnote statement puts a small amount of white space on the bottom of the map, while the title10 eliminates unwanted white space from the top. The statistical style is used.

```
/*Import ZCTA boundary file*/
Proc MAPIMPORT OUT=vazip Datafile="r:\bokerson\sgf2013\t1_2010_51_ZCTA510.shp";
run;

/*Import gaz_zcta_national.txt as gazetteer*/
data gazetteer; length ZCTA5CE10 $5.;
set gazetteer2;
ZCTA5CE10=left(geoid);
if state='VA';
run;

/* Map the Data*/
ods graphics on;
ODS HTML style=statistical;
goptions ftitle='Arial';
title1 " "; title10 h=2.2 "Virginia Population"; footnote " ";
/* colors from green->red */
pattern1 v=s c=cx00ff00; pattern2 v=s c=cx35ff00; pattern3 v=s c=cx65ff00;
pattern4 v=s c=cx88ff00; pattern5 v=s c=cx9aff00; pattern6 v=s c=cxbaff00;
pattern7 v=s c=cxcfff00; pattern8 v=s c=cxd0ff00; pattern9 v=s c=cxe0ff00;
pattern10 v=s c=cxffff00; pattern11 v=s c=cxffee00; pattern12 v=s c=cxffe000;
pattern13 v=s c=cxffd000; pattern14 v=s c=cxffdc00; pattern15 v=s c=cxffd800;
pattern16 v=s c=cxffd100; pattern17 v=s c=cxffcd00; pattern18 v=s c=cxffc000;
pattern19 v=s c=cxffb700; pattern20 v=s c=cxff9a00; pattern21 v=s c=cxff8700;
pattern22 v=s c=cxff7700; pattern23 v=s c=cxff5400; pattern24 v=s c=cxff3400;
pattern25 v=s c=cxff0000;

proc gmap data=gazetteer map=vazip;
id ZCTA5CE10; /* matches values between response and map dataset */
choro pop10 /* response variable */
/ levels=25 /* pattern/color levels */
nolegend outline=same /* turn off boundary lines*/;
run;quit;

ODS HTML close;
ods graphics off;
```

EXAMPLE ONE VARIATION

A graduated color map can also be created by using Proc TEMPLATE to modify a style using the startcolor and endcolor attributes in place of the pattern statements. While some graphics procedures such as GCONTOUR and SGRENDER support a three color ramp, Proc GMAP does not at this time. Therefore it is not possible to set a neutral color using this method. Here is the above map re-created with a color_ramp style, setting startcolor to a very light green and endcolor to a very dark green.

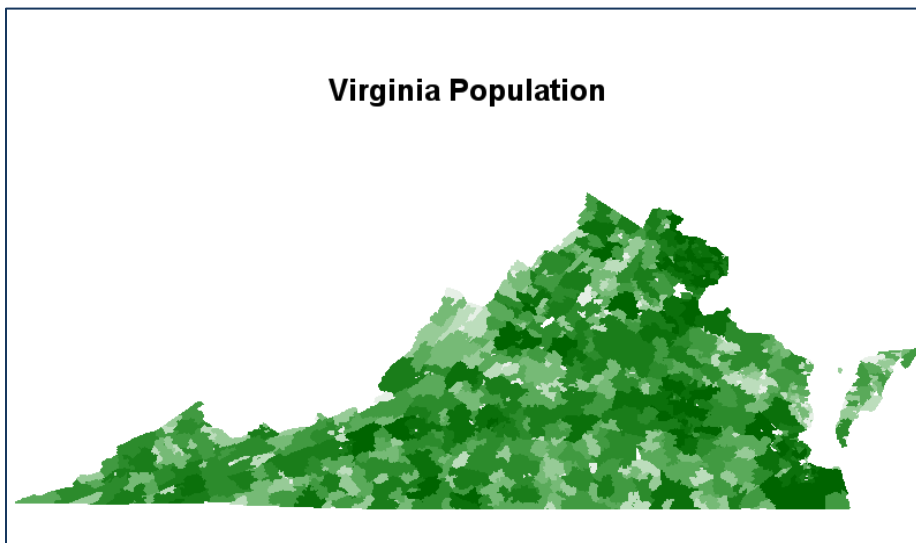


Figure 1a. ZIP Code level map Virginia – variation color map

SAS CODE FOR COLOR RAMP STYLE

With this code, the pattern statements are not needed, nor is the level= in the GMAP procedure. While this is easier to code, it is more difficult to identify population areas than with the three color ramp.

```
ods path sashelp.tmplmst(read) work.templat(update);
ods graphics on;
proc template;
  define style styles.color_ramp;
    parent=styles.statistical;
    style twocolorramp / startcolor=CXE6F0E6 endcolor=CX006400;
  end;
run;
ODS HTML style=color_ramp;
```

EXAMPLE TWO – DISTANCE MAP

Distance maps are often used to visually display the distance between suppliers and consumers in all industries, including health care. Distance maps are especially used in retail location planning to select optimal locations. In this example, the location of public health clinics in an urban area is shown in relation to the location of families with births covered by Medicaid assistance. Map boundaries are ZCTAs.

The map below uses geocoded addresses of the families receiving these services. Geocoding is the process of assigning geographic coordinates to physical addresses. Until recently, geocoding required using online lookup tools for a small number of addresses or commercial products for a large number of addresses. SAS now provides Proc GEOCODE for this purpose. For ZIP +4 applications, the lookup table is not shipped automatically by SAS and needs to be requested or downloaded from www.SAS.com or obtained from another source. Here is SAS code for geocoding ZIP +4 data.

```
proc GEOCODE plus4 lookup=lookup.zip4
data=work.address out=work.geo_address
run;
quit;
```

Here is the ZIP code area plot for the selected urban subset. It is clear that, while one of the public health centers is located in an area of most need, the other is not.

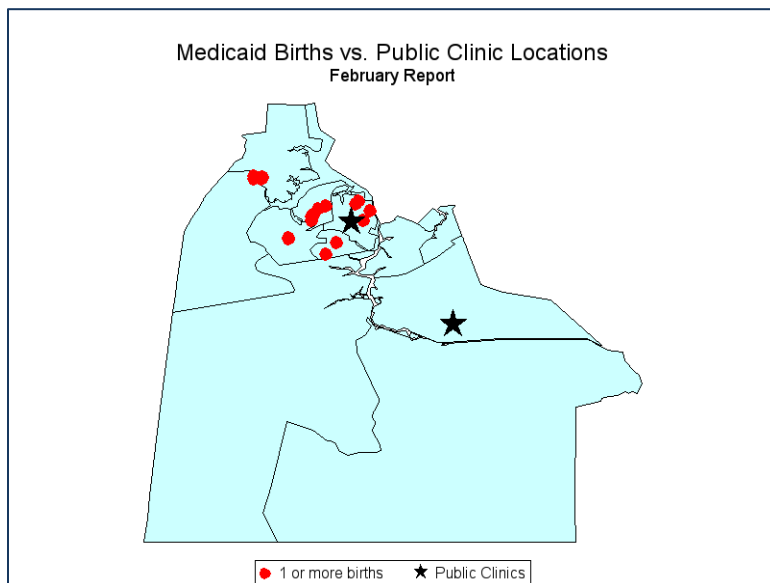


Figure 2. ZIP code boundary map of Medicaid Births

In the map above, the public clinic locations are marked with a large star and areas of residence for births with a red dot. The birth locations and clinic locations are placed on the map with two separate annotate statements. To use two annotate statements with Proc GMAP it is necessary to place one on the choro statement. The legend is generated in the footnote.

EXAMPLE TWO SAS CODE

```
ods graphics on;
ods html style=statistical;
pattern value=msolid repeat=5000 color=CXCCFFFF;
title2 f='Arial' h=2 "Medicaid Births vs. Public Clinic Locations";
title3 f='Arial' h=1.5 "February Report";
/* Map the Data*/
proc gmap data=zip map=zip anno=anno;
id zcta;
choro zcta/discrete coutline=black nolegend anno=dot;
footnote1 f='Arial' box=1 bspace=0 c=red h=1 f=marker 'Z' h=1.3 c=black
        f='Arial' '1 or more births' c=black h=2 f=special 'M' h=1.3
        c=black f='Arial' 'Public Clinics';
run;
quit;
ods graphics off;
ods html close;
```

EXAMPLE THREE – HEAT MAP

A heat map is nothing more than a graphical representation of data where the individual values are represented as a matrix of colors. The difference between a heat map and other data mapping is that the data and the colors represent a hierarchy when displayed on a map. While software designer Cormac Kinney originally coined and trademarked the term "heatmap" in 1991 to describe a 2D grayscale display of financial data, a heat map is any visual display that encodes data values as variations in color. While geographic maps can be used as the underlying structure for heat maps, this is not necessary for this technique and more often this is not the case. In this example, though, we will show a geographic heat map. Like with non-geographic heat maps, the advantage of the heat map is the ability to convey a large amount of information in a relatively small space for speedy visual identification of any existing pattern or patterns. Purists still prefer grayscale to color representations.

In the example that follows, average elevations for ZIP codes on the Virginia Eastern Shore are mapped, first using a grid map, then with a traditional map. To create the grid map, grid squares of latitude/longitude are created and a color value is set based on the combination of elevation for one or more ZIP codes with area within the square. The grid map directly maps the elevation values for each ZCTA.

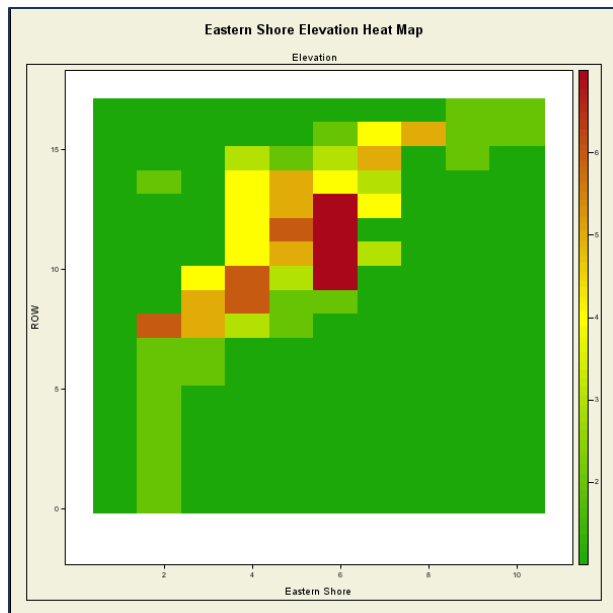


Figure 3a. Eastern Shore ZIP heat map grid

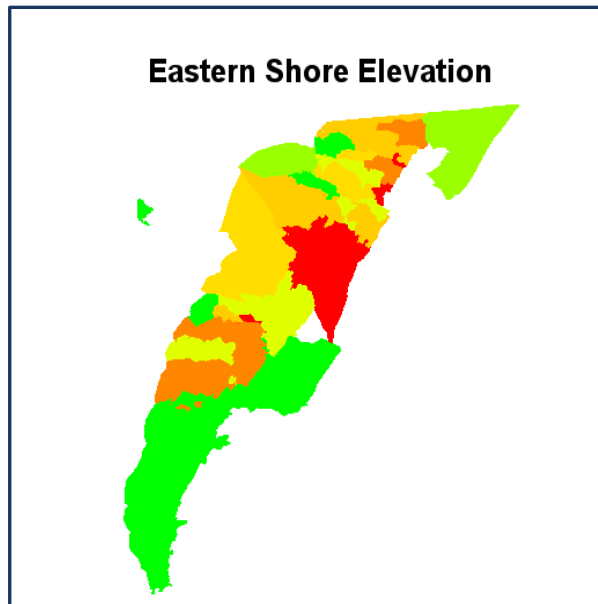


Figure 3b. Eastern Shore ZIP heat map

It should be pointed out that 'high' elevation on the Eastern Shore is really only several feet above sea level. The grid is created using the SGRENDER procedure, a procedure that was created to produce graphical output from templates that are created with the Graph Template Language (GTL). Here is the template and SGRENDER code to create the ZIP code heat map grid:

EXAMPLE THREE SAS CODE:

```
/* Create statistical procedure template*/
proc template;
  define statgraph sgf.Grid;
    mvar wopt hopt; nmvar size1;
    begingraph / designwidth=wopt designheight=hopt;
      entrytitle "Eastern Shore Elevation Heat Map";
      layout gridded / columns=1 columngutter=0px order=rowmajor pad=5px;
        entry "Elevation" / textattrs=GraphLabelText;
      layout overlay / border=true xaxisopts=(label='Eastern Shore' ) pad=(top=5px
        bottom=0px right=0px);
        scatterplot x=name y=row / markercolorgradient=value markerattrs=
        (symbol=squarefilled size=size1) colormodel= threecolorramp name='s2';
        continuouslegend 's2' / orient=vertical location=outside valign=center
        halign=right valuecounthint=10;
      endlayout;
    endlayout;
  endgraph;
end;run;
/* Modify graphics template*/
proc template;
  define Style HeatMap;
    parent = styles.harvest;
    style GraphColors from graphcolors /
      "gramp3cend" = cxaa081b
      "gramp3cneutral" = cxFFFF00
```

```

    "gramp3cstart" = cx1ba808;
end;run;
ods html style=HeatMap image_dpi=100 file='heatmap.html' path='C:\SAS GF';
ods graphics on / reset imagename='ES_Heatmap' imagefmt=gif;
/*Grid the data*/
proc sgrender data=work.grid2 template=sgf.grid;
run;
ods html close;
ods listing;

```

Because the three color ramp was available for SGRENDER, this was used rather than specifying each color. This does not provide a one-to-one match with the geographic map since pattern statements were used for the choropleth colors.

PROC SGRENDER SYNTAX

The SGRENDER procedure allows further customization of ODS Statistical Graphics. SGRENDER only has two statements: SGRENDER and DYNAMIC. While the available options are included in the syntax below, the only required option is TEMPLATE=.

```

Proc SGRENDER TEMPLATE= statgraph-template DATA= input-data-set
  OBJECT= object-name OBJECTLABEL="text-string";
DYNAMIC variable-assignment(s);

```

SGRENDER Options:

- TEMPLATE= StatGraph template that you created
- DATA= Input data set
- OBJECT = ODS object output if desired
- OBJECTLABEL= label the ODS object

The DYNAMIC statement is used to declare dynamic variables. They must also be declared in the StatGraph template.

CONCLUSION

Free data files from the U.S. Census Bureau combined with Proc MAPIMPORT make it very easy to add ZIP code level map data sets to your map libraries. The techniques used in this paper are applicable to ZIP code shapefiles for any U.S. state and territory. There are more shapefiles available, including 3-digit ZIP codes, census tracts and county subdivisions. The SAS code in this paper can be modified to create maps that show distributions of items of interest to you at the ZIP code level as well as at many different levels of geography.

REFERENCES

- Chen Q (2012). "Innovative uses of ODS and GTL." SAS Global Forum 2012 Proceedings.
- Hadden LS, Zdeb M, Allison R (2007). "Wow! You Did That Map with SAS/GRAPH®?" Proceedings of the 31st Annual SAS Users Group International Conference.
- Hadden LS, Zdeb M, Allison R (2006). "ZIP Code 411: A Well-Kept SAS® Secret." Proceedings of the 30th Annual SAS Users Group International Conference.
- Massengill, D. (2004). "SAS Mapping: Technologies, Techniques, Tips, and Tricks." Available at: <http://support.sas.com/rnd/datavisualization/papers/SASMapping.pdf>.
- Massengill, D, Kolovos A (2012). "Together at Last: Spatial Analysis and SAS® Mapping." SAS Global Forum 2012 Proceedings. Available at <http://support.sas.com/rnd/papers>.
- Okerson, B. (2012). "Do SAS® users read books? Using SAS graphics to enhance survey research." SESUG 2012 Proceedings.
- United States Census Bureau. ZIP Code Tabulation Areas (ZCTAs). Available at : <http://www.census.gov/geo/ZCTA/zcta.html>.
- Zdeb, Mike. "From Shapefiles to Maps Via PROC MAPIMPORT and PROC GMAP," NESUG 2011 Proceedings.
- "ZIP Code History". Available at: http://zipboundary.com/zipcode_history.html.

ACKNOWLEDGMENTS

I would like to thank Darrell Massengill, Robert Allison, and the other SAS/GRAPH developers who continue to keep me excited about mapping with SAS.

Healthcare data presented in Example Two has been modified for illustrative purposes and does not represent any actual clinical results.

CONTACT INFORMATION

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

Barbara B. Okerson, Ph.D., CPHQ, FAHM
Senior Health Information Consultant,
West Region Client Advisory & Reporting Services
WellPoint Health & Wellness Solutions
Phone: 804-662-5287
Email: barbara.okerson@wellpoint.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.