SAS® Maps as Tools to Display and Clarify Healthcare Outcomes
Barbara B. Okerson, HMC, Richmond, VA

ABSTRACT

Changes in healthcare and other industries often have a spatial component. Maps can be used to convey this type of information to the user more quickly than tabular reports and other non-graphical formats. SAS®, SAS/GRAPH® and ODS graphics provide SAS programmers with the tools to not only create professional and colorful maps, but also the ability to display spatial data in a meaningful manner that aids in the understanding of changes that have transpired. This paper illustrates the creation of a number of different maps for displaying change over time with examples from the healthcare arena. Examples include choropleth, bubble, and distance maps and introduce the new GEOCODE procedures.

Results included in this paper were created with version 9.1.3 of SAS on a Windows XP platform and use Base SAS, SAS/STAT® and SAS/GRAPH. SAS Version 9.1 or later is required for ODS graphics extensions. SAS Version 9.2 is required for the DATAIMPORT and GEOCODE procedures. The techniques represented in this paper are not platform-specific and can be adapted by both beginning and advanced SAS users.

INTRODUCTION

Health Management Corporation (HMC), a WellPoint company, is one of the nation's largest, most experienced providers of integrated care and total health solutions. Since 1983, HMC has offered members and clients comprehensive programs and services intended to empower members to take control of their health and see positive results. Through a population-based approach, HMC's disease management programs provide solutions for prevention, chronic condition support, lifestyle management and complex condition care. HMC's mission is to help improve the health and financial outcomes of its clients and members through innovative health solutions that consider every single member at his/her level of care.

Within HMC, the informatics team provides a broad range of analytical and data services, including information management, research, advanced analytics, client outcomes, business consulting, and data management that ensure the effectiveness and client satisfaction for the services provided as well as add to general knowledge of the health management arena. SAS is the major analytic software used by HMC and, as such, all of the graphics in this paper were produced with SAS software in conjunction with this work.

Data sets used in this paper are either publicly available or are test data sets developed for the purpose of data modeling and technique demonstration and are not intended to represent the health statistics of an actual population or account.

MAPS AS IMAGES

When generating a map as an image, the user has extensive control over the map's appearance. These maps allow users to do the following:
- For block maps, specify the width of the blocks, the outline colors for the blocks and the map areas, and the angle of view;
- For choropleth maps, select the colors and patterns that fill the map areas, and control the selection of ranges for the response variable;
- For prism maps, control the ranges of the response values and specify the angle of view;
- Hide the legend, or change its features;
- Add titles and footnotes to the map;
- Use annotations to enhance the map; and
- Create non-spatial maps.

JAVA and ACTIVE-X

Both ActiveX and Java allow:
- Enabling a pop-up box that displays data values as the mouse moves over the map areas;
- Panning, rotating, or zooming the map;
- Changing the use of color in the map areas, legend or background;
- Hiding the legend, or changing its features; and
- Displaying or hiding a background image (when used).

Additionally, Java allows:
- Changing the map type so that values for the response data are represented as:
  - blocks that are set on the map areas,
  - patterns or colors in the map areas, or
  - raised polygons in the map areas.

MAPS AS TOOLS FOR OUTCOME DISPLAYS

Maps have an advantage over data tables in that the user can see the outcomes as a complete picture. While maps are used as a tool to display periodic results, they are used far more for ad hoc requests. Maps are used as a major tool in answering questions from clients about their results, especially how their results compare to the population at large. The following examples are used as part of the response to client's requests to understand healthcare outcomes data.

Most of the map examples utilize choropleth maps to display results. A choropleth map is a thematic map in which areas are shaded in proportion to the measurement of the variable being displayed on the map. It provides an easy to see visualization of how the measurement varies across geographic areas. Colors or shading usually form a progression in response to the levels of the mapping variable.

The health of the population is changing, chronic condition numbers are increasing; therefore, it is imperative that we find ways to reverse or slow some of the trends. The maps below show increase in diabetes prevalence from 1998 to 2006.


Diabetes Prevalence 1998

Diabetes Prevalence 2002

Diabetes Prevalence 2006

Diabetes Prevalence by County 2007
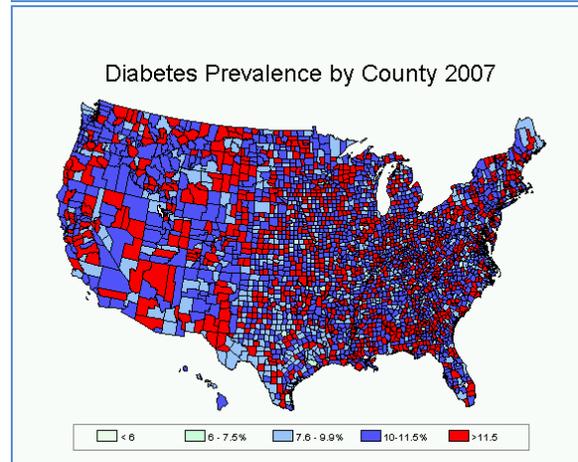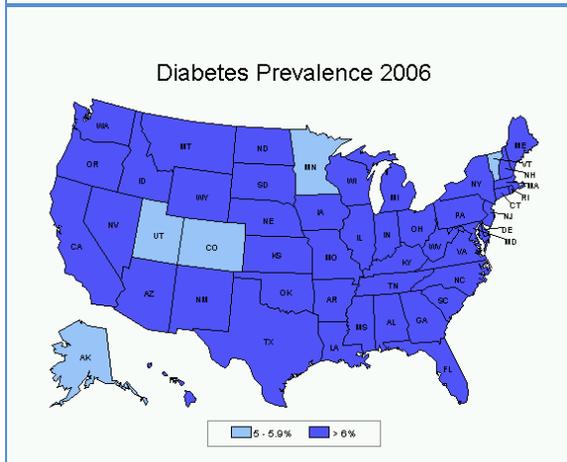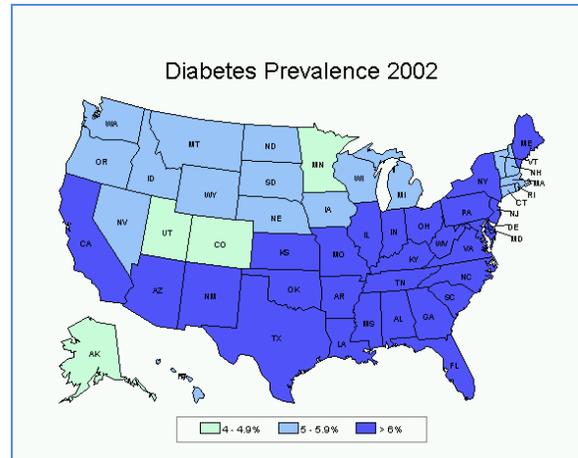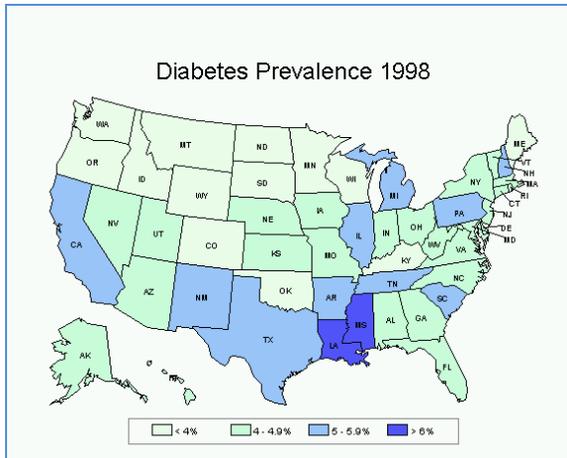
EXAMPLE 1 – Population
Population questions are some of the most common, both static and changes. The questions below are examples.

Question 1a – What is the population universe and how are they served?
This map provides information about the distribution of the reference population as well as providing information about the location of the disease management personnel that work with their clients.

Health Insurance Members and Office Locations
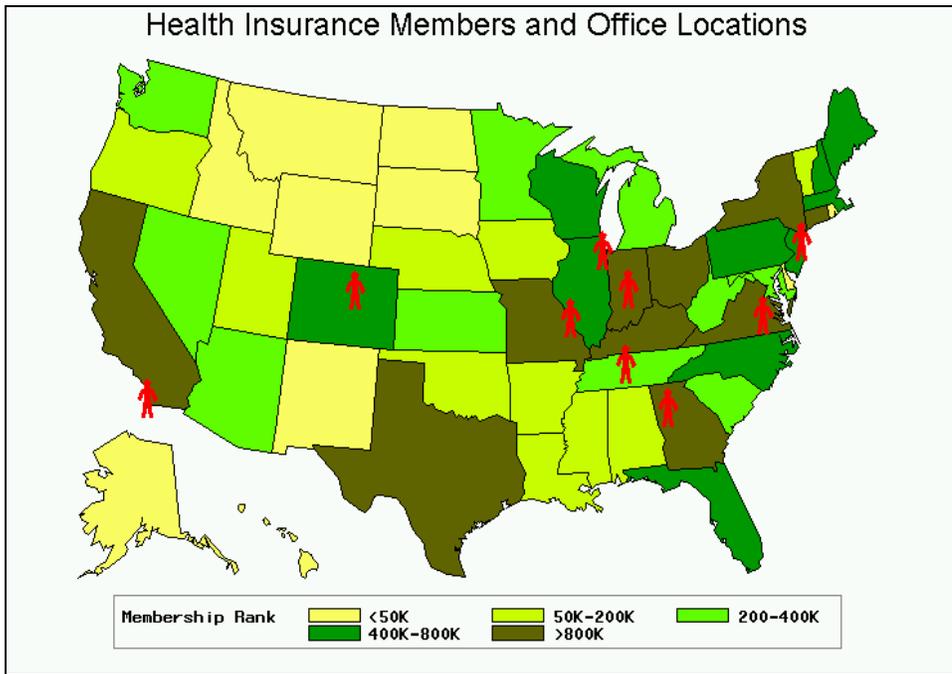
Partial SAS Code for 1a

In the graphic above, the graphic representing the office locations is stored in a format and used in an annotate data set. Here is the code that sets up the format and the annotate data set.

```
 proc format;
    value passfmt 1='Q';
run;
data labels(drop=city pass);
    length function color $ 8 text $ 25;
    retain function 'label' xsys ysys '2' hsys '3' when 'a';
    set office2;
    text=put(pass,passfmt.); position='5'; color='red'; size=7;
       style='Marker'; output;
run;
proc gmap map = maps.us data = maprank;
    id state;
    choro raterank/ levels=5 coutline = black legend=legend1 annotate=labels;
run; quit;
```

Question 1b – How is the client population distributed? Has it changed over time?

When explaining changes in the health of a population, it is often necessary to look at the client's population to see if there were changes in age, gender, and location compared to an earlier time period. The maps below look at a client with major changes in distribution from 2008 to 2010. A choropleth map makes changes easily identifiable.

Population Distribution January 2008

Population Distribution January 2010
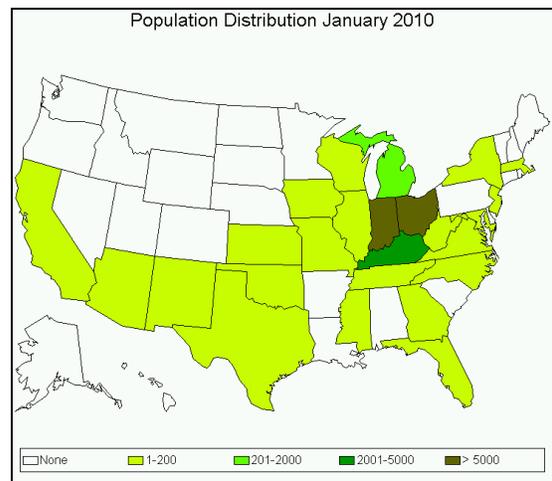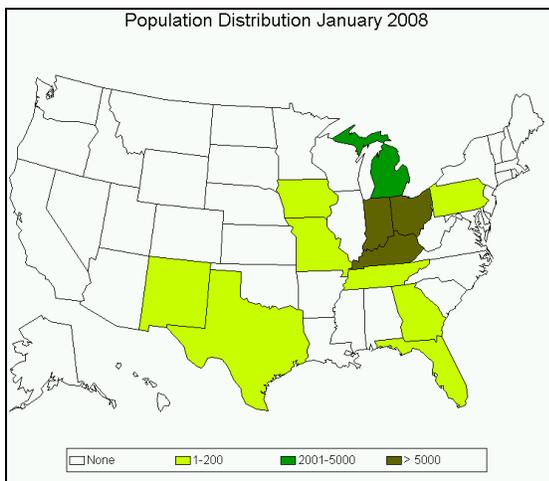
Partial SAS Code for 1b

In this map, the values are captured from the later map to create the ranges used on both maps. The later map has values in all ranges. In this example, as in the earlier example, zip codes from members' addresses are used to identify their states. The SAS zip code data set is used to match zip codes to states, eliminating the territories. Here is code reading in this data set and preparing for merging.

```
proc sort nodupkey data=sashelp.zipcode out=zipstate(where=(statename ^in ('Puerto Rico',
        'Virgin Islands', 'Federated States of Micro','Guam','Marshall Islands',
        'Palau','District of Columbia','Northern Mariana Islands')));
by stfips; run;
```

EXAMPLE 2 – Healthcare Changes

When clients contract for disease/care management, they want to see that the management is impacting their members in a positive way. Questions are often raised as requests to see the program impact.

Question 2a - How has client's healthcare changed over time?

One of the purposes of a disease management program is to improve health outcomes for members. A prenatal health program has a goal of lowering the risk of fetal mortality through better education and use of health services. In these maps, states are divided into quintiles for the baseline period. The remeasurement map uses those same quintiles for mapping the later data. Darker colors are higher rates – lower rates are good. Note that a number of states improved during the time period depicted.
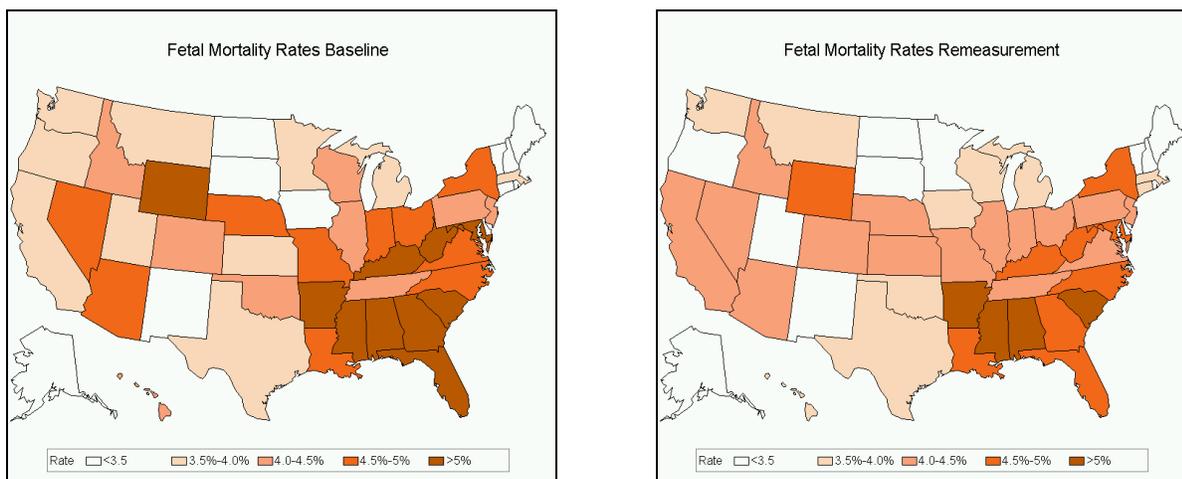

Partial SAS Code for 2a

Proc Rank was used in this example to produce the quintiles for the map. The break points were output and used as the breakpoints for the second map. The code excerpt below shows the goptions (graphic options) statement with the graduated color scheme, the legend statement, and the GMAP procedure code.

```
goptions colors=(white BWH VPAB VLIPB LIB MOPB) ftext='Arial' ctext=black;
legend cborder=gray label=("Rate") value=(j=l "<3.45%" "3.45%-3.78%" "3.79%-4.12%"
"4.13%-4.66%" "4.71%-6.08%");
title3 h=2 'Fetal Mortality Rates Baseline';
proc gmap map=maps.us data=datamap;
id state ;
choro raterank / legend=legend coutline=gray;
run; quit;
```

Question 2b – How has healthcare cost changed over time?

While improving health care outcomes is a goal of a disease management program, so is reducing unnecessary cost. Better condition management usually leads to a reduction in health care costs. The maps below show changes that occurred in cost per member between 2008 and 2010 for a large client. Note that costs increased in some parts of the country and decreased in others. Some increase is inevitable just from inflation.

Medical Costs Per Member by State 2008 / Medical Costs Per Member by State 2010
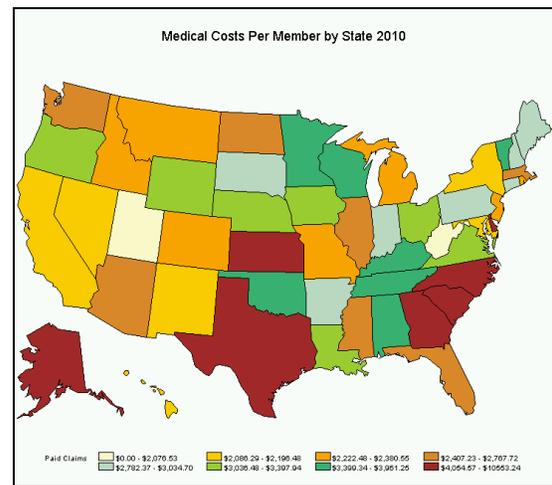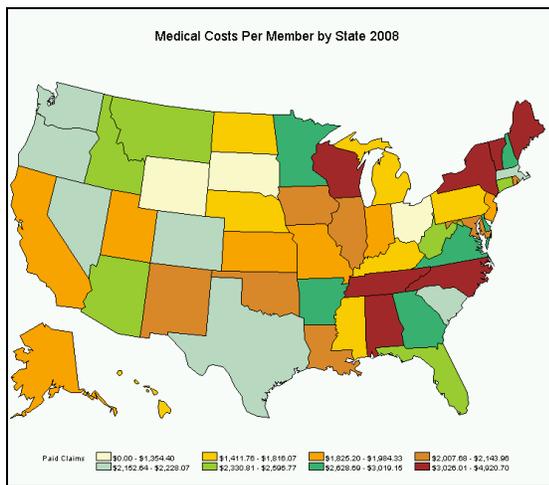
Partial SAS Code for 2b

The maps above calculate the mean per member per state for the beginning of 2008 and 2010 and divides that mean into 8 levels. While it is relatively easy to display graduated change from light to dark using 4 or 5 levels, it is more difficult as the number of levels increase. Yet for some variables, more categories are needed. The code below shows the colors selected for these maps (adapted from Osborne, 2009.)

```
pattern1 v = msolid c = LemonChiffon;
pattern2 v = msolid c = cxffcc00;
pattern3 v = msolid c = orange;
pattern4 v = msolid c = cxd9892b;
pattern5 v = msolid c = vpag;
pattern6 v = msolid c = YellowGreen ;
pattern7 v = msolid c = MediumSeaGreen ;
pattern8 v = msolid c = Brown;
```

EXAMPLE 3 – Geocoding the Population.

The GEOCODE procedure converts address data to geographic coordinates (latitude and longitude values) and requires two SAS data sets:

- The input data set that you want to geocode. This data set will contain variables related to the address such as street address, city, state, and zip code.
- A lookup data set containing the data to transform your address data into geographic locations. By default, sashelp.zipcode is used for zip code or city geocoding.

Additionally a data set will be created that contains the results of the geocoding.

Proc GEOCODE supports 5 geocoding methods:.

- ZIP – ZIP Code centroid geocoding.
- PLUS4 – ZIP+4 geocoding.
- CITY – City and state geocoding.
- RANGE – Matches are made to lookup data containing a range of values. IP addresses are a common use of this type of geocoding
- CUSTOM – This method enables you to use custom lookup data

The procedure syntax is:
```
proc geocode
        method
        input_dataset
        output_dataset
        attribute_variables
```

Distance maps are often thought of in geographic terms, depicting distances between two places. In this example, the area plot option of SAS Proc GPLOT is used to create a distance map without reference to a specific geographic place.

5

Question 3 – Do members have access to client's preferred providers?
In this example, geocoding was used to get exact coordinates of healthcare plan member's residences to calculate distance from providers. The following graphic displays the results. The average distance to 1 provider was 1.9 miles and 4.5 miles to 5 providers
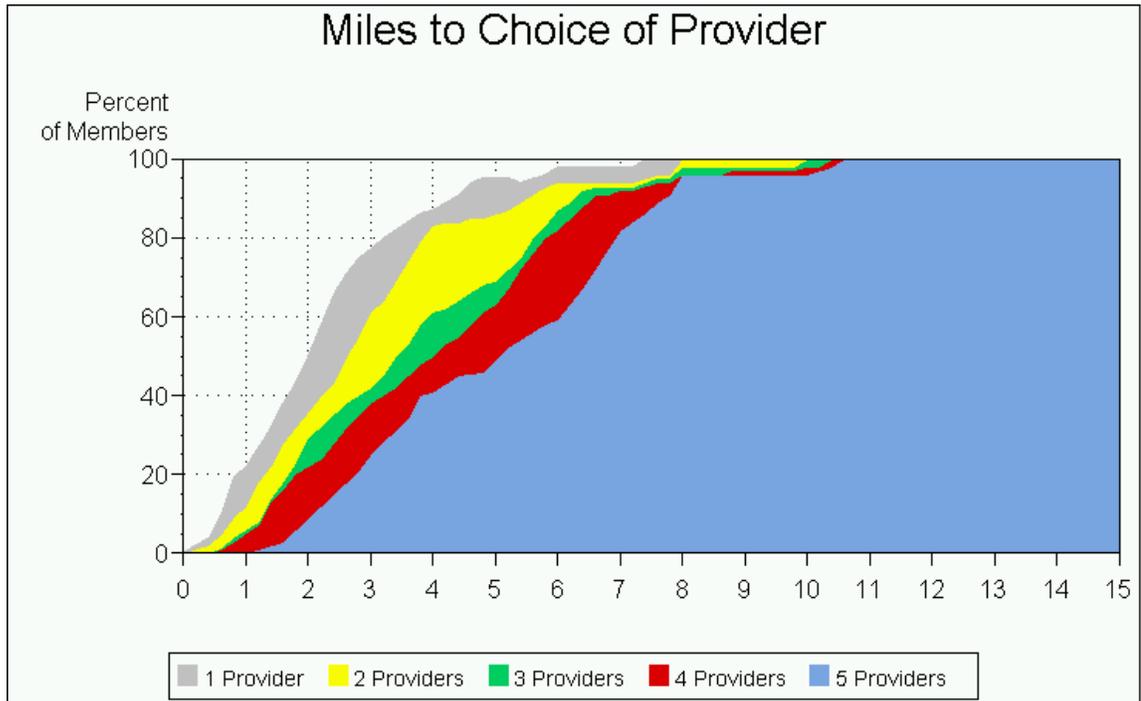
Partial SAS Code for 3
The code below uses the plus4 option. The lookup table for this option is not shipped automatically by SAS and needs to be downloaded or obtained from another source.

```
proc geocode
plus4
lookup=lookup.zip4
data=work.members out=work.geo_members
run; quit;
```

This code creates the plot. The legend was created as a footnote as legends are not supported with multiple plot statements.

```
proc gplot data=provider2;
   plot p5*interval p4*interval p3*interval p2*interval p1*interval  /overlay
   autovref autohref  cvref=black chref=black lautovref=34 lautohref=34
   haxis=axis1 vaxis=axis2  caxis=black  vminor=3
   areas=5;

footnote2 box=1 f=marker c=ltgray 'U' f='Arial' c=black '  1 Provider      '
   f=marker c=cxFFFF00 'U' f='Arial' c=black '   2 Providers     '
   f=marker c=cx00CC66 'U' f='Arial' c=black '   3 Providers     '
   f=marker c=cxD80000 'U' f='Arial' c=black '   4 Providers     '
   f=marker c=cx7Ba7E1 'U' f='Arial' c=black '   5 Providers     ';
run;
quit;
```

EXAMPLE 4 – Bubble Map
Question 4 – Where are preferred provider clinic locations?
While this example does not show locations for preferred providers for a particular client since that information is proprietary, it provides a demonstration of the principle. This map uses the characteristics of a bubble map to identify and name the location of rural health clinics. Additionally, the bubbles are hyperlinked so that Google map information can be obtained by clicking on the clinic locations.
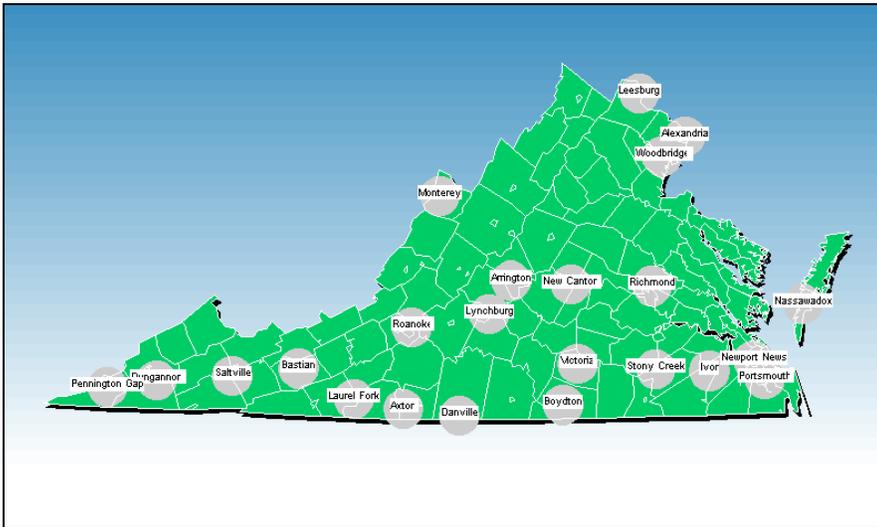
Clicking on Leesburg in the map above, brings up the following information tied to the zip code of the clinic location. The actual location is identified with the red marker on the satellite map.
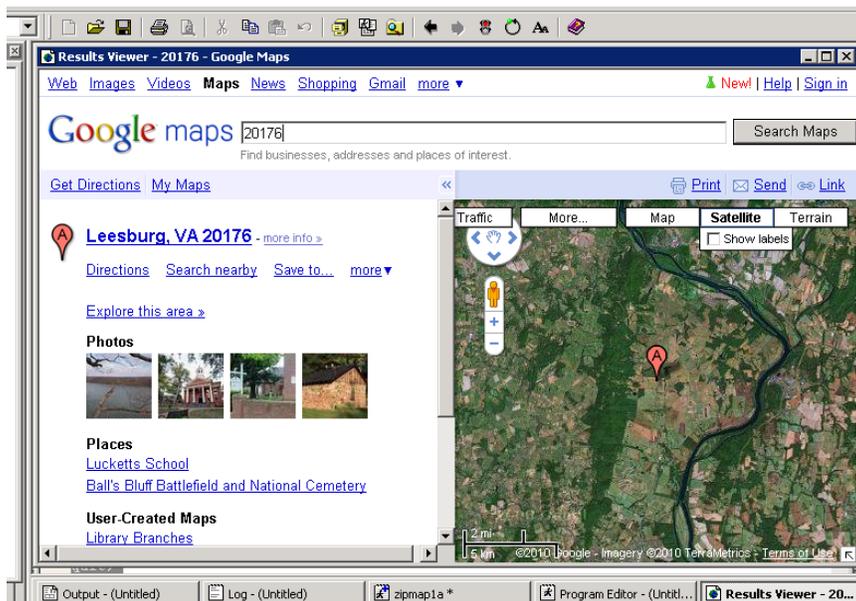
Partial SAS code for 4

The code below is used to create the annotate data set for the bubble circles and they hyperlinks.

```
data circle_anno;
   length function style color $ 8 position $ 1 text $ 20 html $1024;
   retain xsys ysys '2' hsys '3' when 'a';
set circle_anno;
html='title='||quote( trim(left(propcase(city)))||' ('||trim(left(zip))||')' )
 ||'
'||'href='||quote('http://maps.google.com/maps?t=k&hl=en&q='||trim(left(put(zip,z5.))));
function='pie';  color='graycc';  style='psolid';  position='5'; rotate=360; size=3;
   anno_flag=1;
output;
function='label';  position='B'; style='"Arial"';   text=trim(left(city));
   color='black'; cbox='white';  size=2;  anno_flag=3;
output;
run;
```

EXAMPLE 5 – County Map

Sometimes data at the state level does not provide enough information. This example provides national data at a county level.

Question 5 – Are there differences in provider behavior in different regions of the country?

Because corporate clients often have members in diverse geographic areas, they want to ensure that all members are getting the best possible care, preferably the same care. To this end, provider behavior is analyzed in a number of areas. The example shows the results of a referral program analysis.. Identifying referral patterns that exist for members who are referred to and are engaged by specialty programs adds information that is vital for predicting program participation, cost and outcomes. The display (county level) shows that successful behavioral health referrals are more likely in New England and the U.S. Northwest than in other areas.
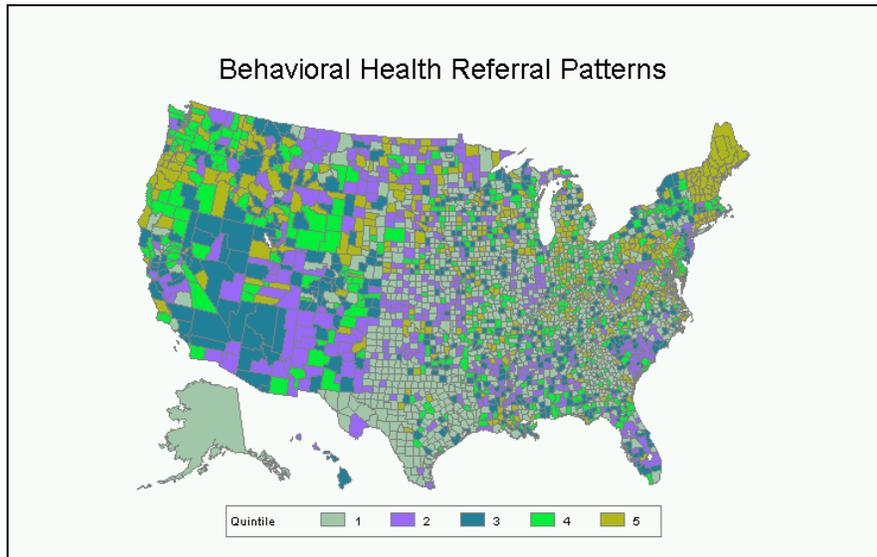
Partial SAS Code for 5

Because the data set contained zip codes with no county information, the sashelp.zipcode data set (as referenced in the geocode example above and regularly updated by SAS) was used to map zip codes to counties. Map colors were set in the GOPTIONS statement with the patterns selected providing a progression from gray to dark gray blue (top quintile). Selected SAS code below shows the RANK procedure used to create the mapping levels.

```
proc rank data=mapbh out=mapbhrank group=5;
   var pctbh;    ranks pctbhgrp;
run;

data mapbhrank(rename=(stfips=state cntyfips=county));
   set mapbhrank; pctbhgrp=pctbhgrp+1; run;

proc sql;
create table bhdatamap as
   select * from maps.uscounty m left join mapbhrank r
   on m.state=r.state and m.county=r.county; quit;
```

CONCLUSION

This paper illustrates several healthcare outcome applications using SAS/GRAPH GMAP as part of disease management and wellness initiatives from a major insurer. Through these and other applications, healthcare and social marketing campaigns can focus on the areas of greatest return on investment, both improving healthcare and saving employer dollars. Geographic data display also aids in the identification of issues such as patient access, provider shortage, and distance to provider, all of which impact quality of care received. The built-in flexibility SAS/GRAPH GMAP, combined with SAS Annotate, and ODS can be used to demonstrate geographical differences and to provide a foundation for further analyses to identify additional relationships and critical factors to enhance the quality improvement process

REFERENCES

Ingenix Geo Networks. http://www.ingenix.com/content/attachments/GeoNetworks_brochure.pdf Accessed February 2010.

Massengill D, Odom E. "PROC GEOCODE: Creating Map Locations from Your Data." SAS Global Forum 2009.

Osborne, Anastasiya. "Let Me Look At It! Graphic Presentation of Any Numeric Variable." SAS Global Forum 2009.

SAS Institute, Inc. 2009. http://support.sas.com/rnd/datavisualization/ Accessed October 2009.

SAS Institute Inc. 2008. SAS OnlineDoc® 9.1.3. Cary, NC: SAS Institute Inc.

ACKNOWLEDGEMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. For more information contact:

Barbara B. Okerson, Ph.D., CPHQ
Senior Health Information Consultant,
National Accounts Outcomes
Health Management Corporation (HMC)
8831 Park Central Drive, Suite 100
Richmond, VA 23227
Office: 804-662-5287
Fax: 804-662-5364
Email: bokerson@choosehmc.com