Distance Mapping in Health and Health Care: SAS® as a Tool for Health Geomatics

Barbara B. Okerson, HMC, Richmond, VA

ABSTRACT

Geomatics is the science concerned with using mathematical methods with spatial data to learn about the earth's surface. Health geomatics is used to improve the understanding of relationships between people, location, time, and health. Geomaticists assist in discovering and eliminating disease, aid in public health disease prevention and health promotion initiatives, and support healthcare service planning and delivery through defining spatial relationships. Distance mapping is one of many tools in the geomatics toolbox that furthers this understanding.

SAS® provides a comprehensive set of graphics tools and procedures as part of the SAS/GRAPH® product that can be used for distance mapping, including the new GEOCODE procedure and the GEODIST function, available with SAS 9.2. This paper combines the graphics available in SAS/GRAPH with other available SAS tools to explore spatial relationships in the health care arena. Examples include maps, grid squares, area charts, and X-Y plots.

The analytic graphics in this paper were developed with version 9.1.3 or 9.2 of SAS executing on a Windows XP platform. SAS 9.2 is required for the GEOCODE procedure. The graphics represented in this paper are not platform-specific and can be adapted by both beginning and advanced SAS users.
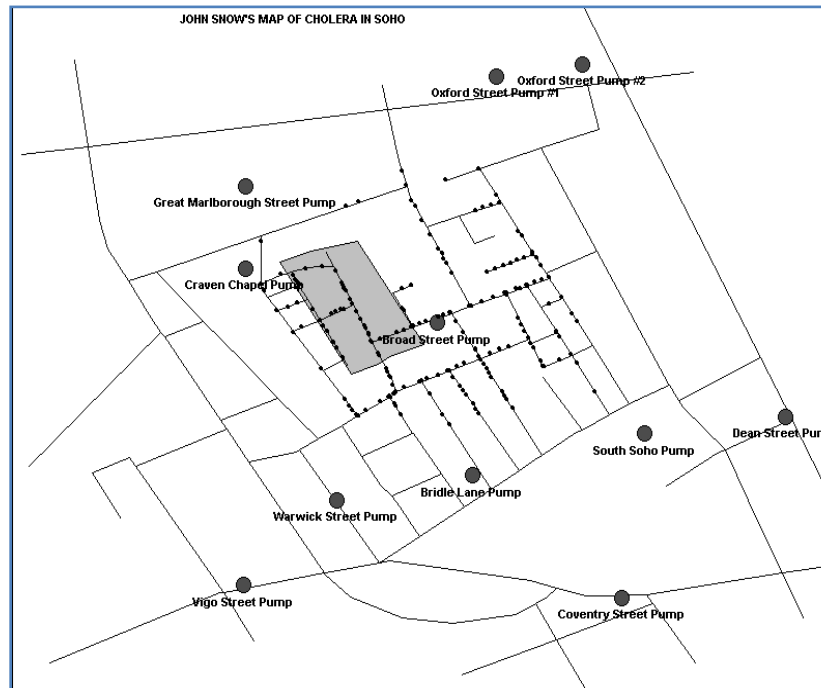
BACKGROUND

Health, disease, and health care differ from place to place and from time to time. While undertaking the analysis of the relationships between health and locations, practitioners in the field of health geomatics can be divided into two distinct areas of study: those seeking information about the geography of disease and those involved with the geography of health care systems; the former concerns itself with the health conditions of a place and the latter is concerned with the response to those health conditions in terms of planning, intervention and infrastructure. The study of the geography of health conditions can increase the general knowledge of health conditions including better understanding of causality. The geography of health care systems is essential in efficient allocation of health care resources in an increasingly stretched economy.

Both areas of study within health geomatics require interdisciplinary approaches, encompassing many academic and operational disciplines. These include data mining, epidemiology, cartography, remote sensing, geographic information systems (GIS), global positioning systems (GPS), database management, computational geometry, statistics and others. By combining all this information, the goal is to better understand health conditions, better target health programs and ultimately lead to better health for the population.

Data used for health geomatics can include, but is not limited to:

- Population data;
- Environmental and ecological data;
- Topography and hydrology data;
- Weather and climate data;
- Land-use and public infrastructure data;
- Transportation networks data;
- Health infrastructure data;
- Epidemiological data, including mortality, morbidity and disease distribution;
- Historical data;
- Comparative and benchmark data;
- Political information; and
- Other data as needed for health-related analyses.

The concept that location can influence health of the inhabitants of that location dates back to Hippocrates and his observations that certain diseases occurred in some locations and not in others. But it was John Snow and his 1854 map of cholera deaths versus pump locations in London, tying that disease to a contaminated water pump, that is considered the beginning of the field of epidemiology. This is illustrated in the map below (Kamel Boulos, Roudsari, Carson, 2001)

SAS 9.2 DISTANCE ANALYSIS TOOLS

Several new tools have been added in SAS 9.2 that aid in location and distance analysis. While these tools will become extremely valuable as 9.2 becomes more widely adopted, except as otherwise noted, all examples in this paper were run with SAS 9.1 and these tools were not available.

- **PROC GEOCODE**—calculates a longitude and latitude for a given address. This is useful for plotting points on a map where only an address or postal code is available.
- **GEODIST Function**—calculates a distance from point to point using geographic longitude and latitude coordinates. The algorithm uses the Vincenty Distance Formula for calculations, making the measurements more accurate because longitudes become closer to each other the farther the latitude is from the equator.
- **ZIPCITYDISTANCE** function is a companion that calculates distances between centers of U.S. Zip codes.
- **%REDUCE macro**— available as "reduce_pixel.sas' from SAS Maps Online, this macro reduces map data optimally for a desired resolution.
- **PROC GINSIDE**—applies regional (polygonal) values to a point. If the point is inside a geographic region, then it inherits the attributes of that region.

FINDING THE DISTANCE BETWEEN TWO POINTS

SAS map data sets have X and Y coordinates represented as longitude and latitude converted to radians. Mileage between two points in an unprojected SAS map data set can be calculated as follows using X1 as the longitude of point 1 and Y1 as the latitude of point1, X2 and Y2 for the next point, etc.

This distance is:

```
Distance = 4000*ARCOS(arc)
        where:
arc = SIN(Y2)*SIN(Y1)+COS(Y2)*COS(Y1)*COS(X2-X1)
```

 Western hemisphere longitudes are negative and eastern hemisphere longitudes are positive. Northern hemisphere latitudes are positive and southern hemisphere latitudes are negative

If your starting coordinates are in decimal degrees, they can be converted to radians as follows:

```
longitude= atan(1)/45*longdeg;
latitude = atan(1)/45*latdeg;
```

If the longitude or latitude is stored in degrees/minutes/seconds, use the following equation to calculate decimal degrees before converting to radians:

```
VALUE = (degrees)+(minutes/60)+(seconds/3600);
```

If you are calculating distance over large areas you can also calculate distance between two pairs of latitude/longitude values in radians using the Great Circle Distance Formula which better compensates for the curvature of the earth. This formula is also known as the haversine formula, since it is often written in terms of the haversine function: (haversin(θ) = sin^2(θ/2) = (1−cos(θ))/2). This formula is:

```
Distance = 3949.99*arcos(sin(Y1)*sin(Y2)+cos(Y1)*cos(Y2)*cos(X1-X2));
```

The following macro uses the Great Circle Distance Formula for calculating the distance between two points.
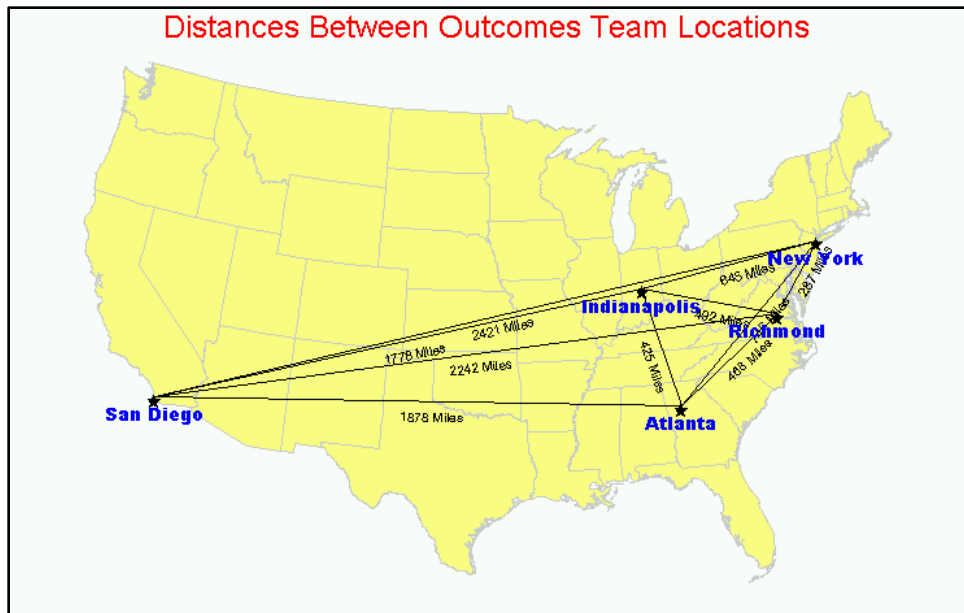
```
%macro distance(lat1,long1,lat2,long2);
%let pi180=0.0174532925199433;
7921.6623*arsin(sqrt((sin((&pi180*&lat2-&pi180*&lat1)/2))**2
      +cos(&pi180*&lat1)*cos(&pi180*&lat2)*(sin((&pi180*&long2-
      &pi180*&long1)/2))**2));
%mend;
```

EXAMPLE ONE – DISTANCE BETWEEN POINTS

In this example, the distances are calculated between the locations of health care team members, which can be important for team members that often have need to support each other at the various locations. For health care, this concept can be very valuable in other areas, including:

- Distance between patient and *nearest* treatment center or provider;
- Distance between patient and *preferred* treatment center or provider;
- Distance from infected individuals for disease condition;
- Distance from emergency services transportation;
- Distance nearest area of contamination.

The city locations use coordinates from the MAPS.USCITY data set, a data set provided by SAS that includes the coordinates for almost 25,000 United States cities. The map is created showing all possible combinations between the city locations.

Distances Between Outcomes Team Locations

PARTIAL CODE FOR EXAMPLE 1:

```
data city_distance;
  keep startcity endcity distance startprojx startprojy endprojx endprojy;
  set locations;
    startx=atan(1)/45*long;
    starty=atan(1)/45*lat;
    startcity=city;

 /* Get the projected values for annotate */
    startprojx=x;
    startprojy=y;

 /* Get the observations for each of the cities */
  do i=1 to &num;
    set locations point=i;
    endx=atan(1)/45*long;
    endy=atan(1)/45*lat;
    endcity=city;
      endprojx=x;
      endprojy=y;

  /* If start and end are the same, delete the observation */
   if startcity = endcity then delete;

  /* Calculate distance between cities with Great Circle Distance Formula*/
  else
    Distance = round(3949.99*arcos(sin(starty)*sin(endy)+
      cos( starty )*cos(endy )*cos( startx - endx ) ));
  output;
 end;
run;
```

Since the distance has been calculated between all points, a standard mileage chart can also be created from the data. To create the mileage chart, delete the step from the above code that removes observations with duplicate start and end values, run a transpose (code below), and then format as desired.
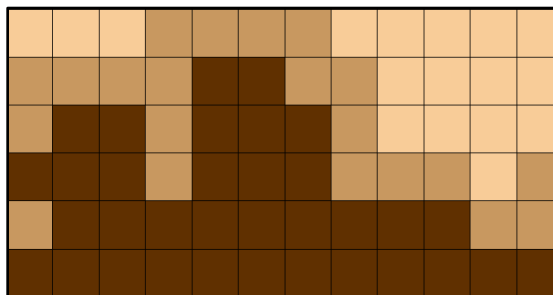
City	Atlanta	Indianapolis	New York	Richmond	San Diego
Atlanta	0	425	746	468	1878
Indianapolis	425	0	645	492	1778
New York	746	645	0	287	2421
Richmond	468	492	287	0	2242
San Diego	1878	1778	2421	2242	0

DISTANCE CHART CODE:

```
proc transpose data=citydistance_forchart out=mchart;
var endcity distance ;
by startcity;
id endcity;
run;
```

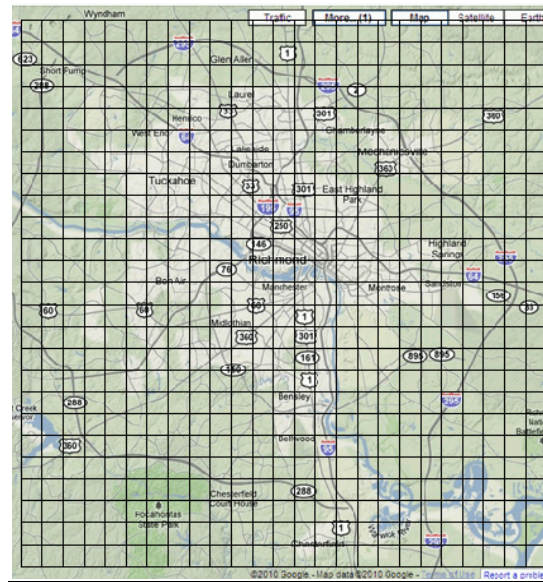## EXAMPLE TWO – GEOGRAPHICALLY BASED DISTANCE GRID

Often when thinking of geographically distance grids, grids like the following come to mind where elevation of the areas within a map square is represented by grid square colors. If covering large distances, this could be useful in health care, since certain conditions, especially cardiac conditions and stroke, are less prevalent in high altitudes.
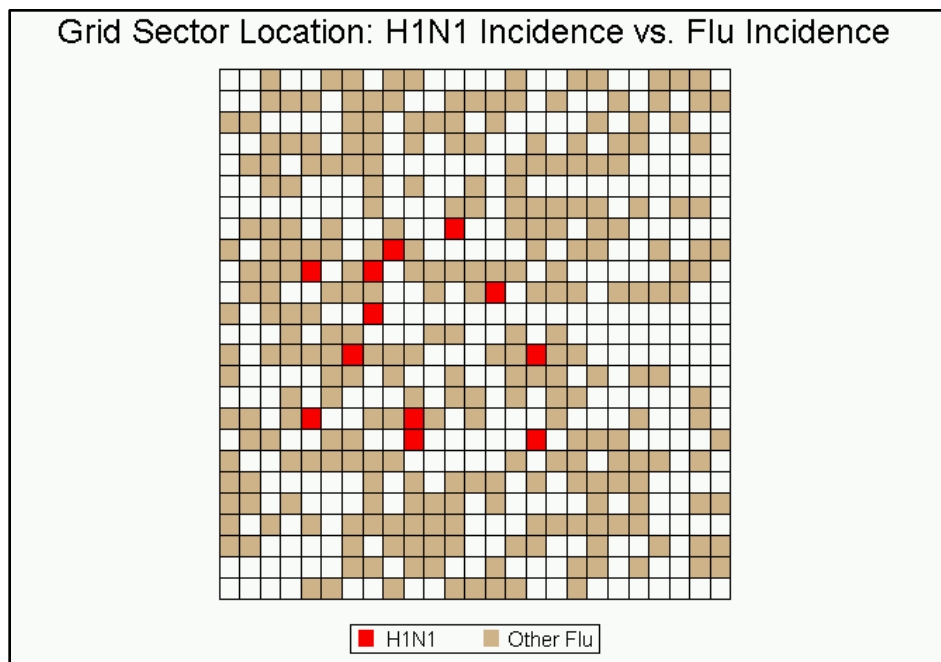

Distance grids become much more useful when applied to smaller geographic areas and health care data for the area can be mapped to the grid. This is especially important for infectious diseases. In the example below, H1N1 and other influenzas are mapped to a geographically based grid to analyze the extent of distribution within a metropolitan area. To create a map like this:

- Overlay grid on map with known and marked geographic coordinates – (coordinates for map below):
  - Lat: 37.3⁰N to 37.9⁰N
  - Long: 77.7⁰W to 77.2⁰W
- Assign an area number (e.g. A1 to A625) to each square of the grid and matching map area.
- Calculate grid line spacing for number of grids that you want and write to data set – in this example, each vertical line is separated by .02 degrees and each horizontal line by .024 degrees.
- Convert to radians for SAS mapping or size to correct size for corner matching.
- Geocode the response data set to be mapped.
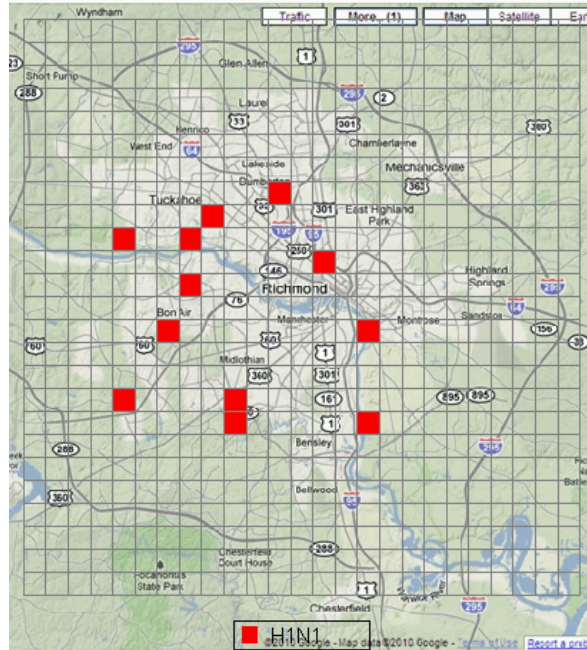- Map the geocoded data from the map area to the grid using SAS/GRAPH GMAP procedure.

This illustration shows the grid matched to the map area. A 25x25 grid was selected to better pinpoint areas with diagnosis, especially H1N1. In this case the area in the map is Richmond, VA and surrounding area, as represented by the coordinates listed above.

After matching to the map and assigning radians to each of the grids, the portion of the population with a flu diagnosis was mapped to the grid squares and displayed. Type of flu is color coded with H1N1 infections colored red and other flu types colored tan. Colors represent at least one diagnosis in that sector. Data has been modified or created for illustrative purposes and does not represent any actual clinical results.

In this map, the grid is matched to the geo-coordinates and the grid/flu data is overlaid on the map to show the geographic sectors where the H1N1 diagnoses occurred.

Again, this data does not represent any actual clinical results and was created through random selection from a previously geocoded file of street addresses of health plan members in the area. The code below shows creation of the grid and assigning geographic coordinates to that grid.

PARTIAL SAS CODE FOR EXAMPLE TWO:
CODE TO CREATE GRID:

```
data space (keep=sq x y);
 do j = 1 to 25;
   do i = 1 to 25;
     sq+1;
     x=i; y=j; output;
     x=i+1 ; output;
     y=j+1 ; output;
     x=i ; output;
   end;
 end;
 do i=1 to 3;
   sq+1;
   output;
 end;
run;
```
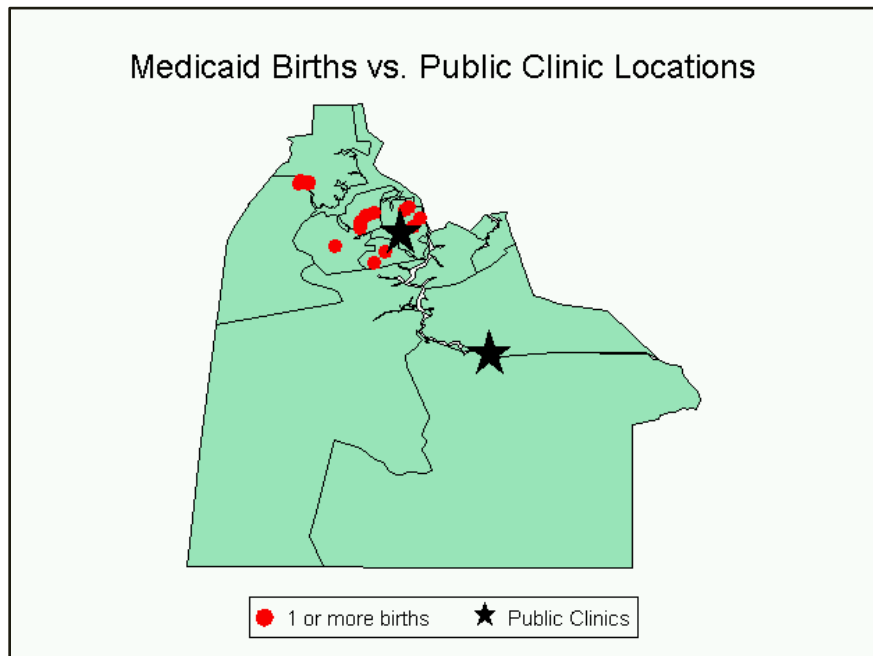
CODE TO MATCH GRID TO GEOGRAPHIC COORDINATES:
This code requires the prior calculation of distance between grid square boundaries as well as the geographic coordinates for the area under the grid. These known coordinates and distance are included in the code. If distance calculation between points is desired, distance can be calculated as 69.17 miles per degree.

```
data grid;
 set space(rename=(x=long y=lat));
   retain xstart 77.7 ystart 37.3 xinc -0.02 yinc 0.024;
   x=xstart + (long-1) * xinc;
   y=ystart + (lat-1)* yinc ;
run;
```

EXAMPLE THREE – LOCATION ANALYSIS MAP

Another example uses a map in a similar manner as the water pump example. Vulnerable populations are often susceptible to inequalities in health care and poorer health outcomes. In this map, using zip code boundary files (ZCTAs), the location of public health clinics in an urban area is shown in relation to the location of those families with children born to families receiving Medicaid assistance. This map shows that, while one of the centers is located in area of most need, the other is not.

PARTIAL CODE FOR EXAMPLE 4

In this example, the birth locations and clinic locations are annotated with two separate annotate statements. The legend is generated in the footnote. Because zip code boundaries are used, the GMAP procedure id is set to zcta. It should be noted that zip codes do not actually have geographic boundaries; the zip code tabulation areas are generated by the U.S. Census Bureau for mapping purposes.

```
pattern value=msolid repeat=5000 color=vlig;
title f='Arial' h=2 "Medicaid Births vs. Public Clinic Locations";

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;
```
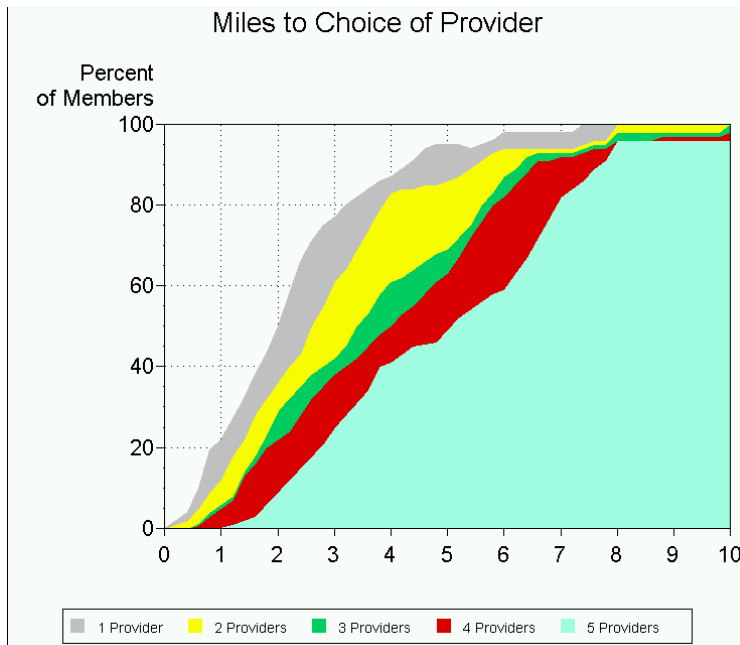
EXAMPLE FOUR – NON-GEOGRAPHIC DISTANCE CHARTS

There is great advantage to tying health care data to geographic coordinates, but all distance mappings in the health care arena do not need radians or latitude/longitude to provide meaningful information. The examples below are illustrations of non-geographic distance mapping.

EXAMPLE 4a: AREA DISTANCE PLOT

In this example, geocoding was used to get exact coordinates of healthcare plan members' 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

Miles to Choice of Provider

PARTIAL SAS CODE FOR EXAMPLE 4a:
The code below uses the Proc GEOCODE plus4 option. The lookup table for this option is not shipped automatically by SAS and needs to be requested or downloaded from www.SAS.com or obtained from another source.

```
proc geocode plus4 lookup=lookup.zip4
  data=work.members out=work.geo_members
run;
quit;
```
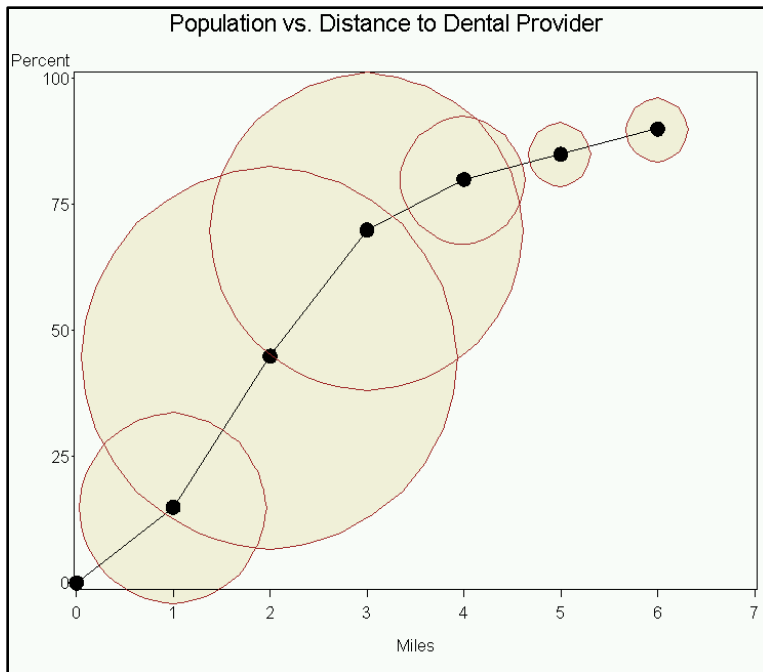
This code creates the plot. The legend was created as a footnote as legends are not yet supported with multiple plot statements.

```
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=cxA5FEE3 'U' f='Arial' c=black '  5 Providers   ';
run;
quit;
```

EXAMPLE 4b – DISTANCE BUBBLE PLOT
It is often important to understand what proportion of a population is reached within a certain distance or mileage. In this additional example of provider to plan member distance, the bubbles represent the proportion of the population at that distance, while the points represent the cumulative percentage at that distance.

9

Population vs. Distance to Dental Provider

PARTIAL SAS CODE FOR EXAMPLE 4b:
The code used to create this graphic uses two annotate data sets, one for the circle and the other for the circle border, since the circle is plotted before the plot data set (dental) and the border after the plot data set.

```
data annoplot; set dental;
  length function $8 text $1;
  retain XSYS YSYS '2' style 'Marker' function 'LABEL'
    when 'b' color 'cxF2F2DF';
  size = size;
  X=distance; y=population;
  text='W';
output;
run;;

data annoplot2; set dental;
  length function $8 text $1;
  retain XSYS YSYS '2' style 'Markere' function 'LABEL'
    when 'a' color 'brown';
  size = size;
  X=distance; y=population;
  text='W';
output;
run;
goptions ftext='Arial' ftitle='Arial' htext=1.4;
title h=2 'Population vs. Distance to Dental Provider';
axis1 order=0 to 7 by 1 minor=none label=("Miles");
axis2 minor=none order= 0 to 100 by 25 label=("Percent");
symbol i=join v=dot h=2;

proc gplot data=dental anno=annoplot;
  plot population*distance/
  vaxis=axis2 haxis=axis1 anno=annoplot2;
run;
```

10

CONCLUSION

Through partnership with geographers, cartographers, epidemiologists and others, health practitioners have the opportunity to use to use health geomatics tools to make strategic contributions to population health in many areas. These include: 1) health service delivery; 2) planning; 3) program implementation and intervention; and 4) policy development that directly impact the health and welfare of all populations. Distance mapping is just one of many techniques that currently enhance this process. With the added functionality of SAS 9.2 and links to GPS coordinates, even more is possible in the future.

AKNOWLEDGEMENTS

I would like to especially thank Marcia Surratt from SAS Institute Technical Support for her assistance in the matching of grid coordinates to geographic coordinates. I also continue to be inspired by the graphics of Robert Allison, Mike Zdeb, Louise Hadden and others. Thanks to them for the willingness to share their work.

REFERENCES

Alexander GL, Kinman EL, Miller LC, Patrick TB (2003). Marginalization and health geomatics. *Journal of Biomedical Informatics* 36: 400–407.

Allison R. Robert Allison's SAS/Graph Examples!. http://robslink.com/SAS/Home.htm. Accessed June 2010.

Fernando, Randima (2004)*. GPU Gems: Programming Techniques, Tips, and Tricks for Real Time Graphics.* Toronto: Addison Wesley Professional.

Gatrell A, Loytonen M, ed (1998). *GIS and health (GISDATA 6).* London: Taylor & Francis, 1998.

Hadden LS (2008). Behind the Scenes with SAS. NESUG 2008 Proceedings.

Hadden LS, Zdeb M, Allison R (2007). Wow! You Did That Map with SAS/GRAPH®? NESUG 2007 Proceedings.

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

Kamel Boulos MN, Roudsari AV, Carson ER (2001). Health Geomatics: An Enabling Suite of Technologies in Health and Healthcare. *Journal of Biomedical Informatics.* 34: 195-219.

Mink K, Zdeb M (2004). Using SAS to Model the Spread of Infectious Disease. SUGI 29 Proceedings.

SAS Institute. Calculate Distances Between Cities. http://support.sas.com/kb/24/897.html. Accessed April 2010.

CONTACT INFORMATION

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

Barbara B. Okerson, Ph.D., CPHQ, FAHM
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