

# Time Series Mapping with SAS®: Visualizing Geographic Change over Time in the Health Insurance Industry

Barbara B. Okerson, WellPoint

## ABSTRACT

Changes in health insurance 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® provides programmers and analysts with the tools to not only create professional and colorful maps, but also the ability to display spatial data on these maps in a meaningful manner that aids in the understanding of the changes that have transpired. This paper illustrates the creation of a number of different maps for displaying change over time with examples from the health insurance arena.

## INTRODUCTION

Maps make it easy to manage and visualize data over different time periods to identify trends and create compelling arguments. They have an advantage over data tables in that the user can see the changes as a complete picture. Maps are used as a tool to display periodic results, answer questions about changes, and show national and global trends. The examples shown here illustrate changes over time in the health insurance industry for:

- Health insurance coverage;
- Hospital cost changes;
- Health insurance type changes; and
- Rates of uninsurance.

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. Choropleths provide easy-to-see visualizations of how the measurements vary across geographic areas. Colors or shading usually form a progression in response to the levels of the mapping variable. Another technique illustrated in this paper is that of dot density mapping. Dots can depict density on their own or can change in both color and size as the values increase or decrease.

Data sets used in this paper are all publicly available and are included in publications by, or available for download from, the U.S. Dept of Health and Human Services, the Centers for Disease Control, United States Census Bureau, and Institute for Health and Socioeconomic Policy. Years selected for comparison were selected solely on the basis of availability and for the purpose of technique demonstration.

Results 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. SAS Version 9.1 or later and a SAS/GRAPH license are required for ODS (Output Delivery System) graphics extensions. The techniques represented in this paper are not platform-specific and can be adapted by beginning through advanced SAS users.

## PROC GMAP

The GMAP procedure is a SAS tool for producing two-dimensional or three-dimensional maps that show variations of values with respect to an area, most often a geographic area. SAS also provides a large assortment of boundary files and other map data sets with SAS/GRAPH software. For a listing of and information about available maps see the METAMAPS data set in the SAS MAPS data library.

Generally maps are thought of as having boundaries with geographic coordinates such as longitude and latitude as in the provided map data sets referenced above. PROC GMAP will draw any shape or shapes that you create. You just need to provide the points to connect the lines and your shape can be a map data set. GMAP is especially useful for the following visual tasks:

- Producing maps of geographic areas;
- Summarizing data that vary by location;
- Showing trends and changes over time by location;
- Displaying variations of data between geographic areas;
- Highlighting regional differences or extremes; and
- Creating non-geographic area maps and layouts.

To place data on the map, you need a response data set. A response data set is a SAS data set that contains data values to be placed on the map; each value of the response variable needs to be associated with a specific map area in the map data set. The response data set also needs to include an identification or ID variable that serves as a link between the map area and the response value. This variable must be the same in both the map and response data set and represent a separate bounded entity in the map data set, e.g. state.

## EXAMPLE ONE – COUNTY LEVEL CHOROPLETH MAP

The first set of maps provides information about changes in health insurance coverage across U. S. counties between 2005 and 2011. These choropleth or area maps show county residents with health insurance as a percentage of total residents in the county. At first glance, the patterns look very similar but closer inspection shows an increase in lighter areas in the second (2011) map. Because the Missouri area is one that appears to have considerable change, the second set of maps shows the Missouri drill-down for further investigation.

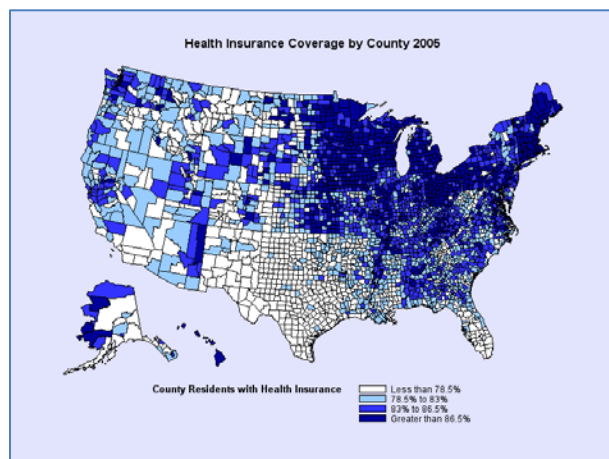


Figure 1. Health Insurance Coverage U.S. County 2005.

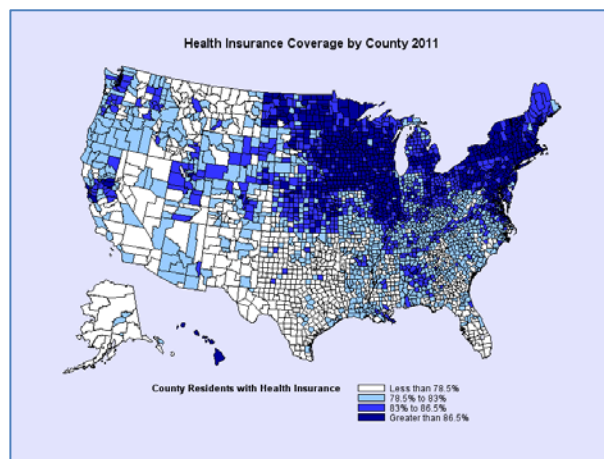


Figure 2. Health Insurance Coverage U.S. County 2011.

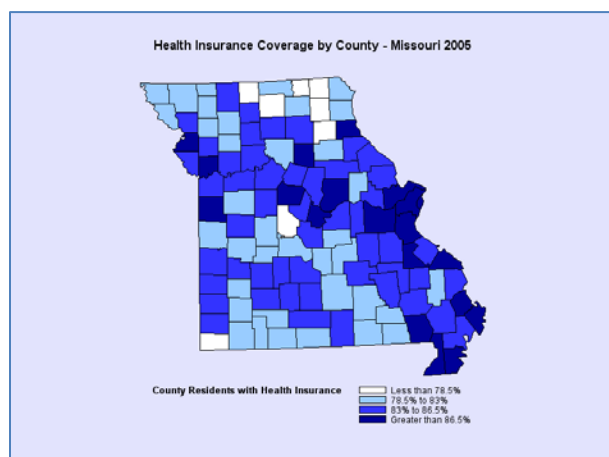


Figure 3. Missouri Coverage 2005.

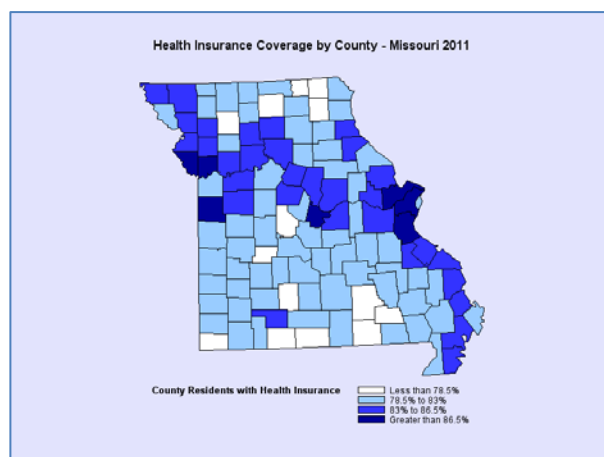


Figure 4. Missouri Coverage 2011.

To create the map and map the data, response data values were first classified into quartiles based on the 2005 values. PROC UNIVARIATE was used to identify the quartiles. These quartile boundaries set the levels for each color as shown in the map legend. Colors here range from white to dark blue as values increase. While SAS allows the use of a large number of colors by name, here hexadecimal codes are used to identify the colors. The CX in front of the color code tells SAS that a hexadecimal color code is being passed to SAS.

The GMAP portion of the code for this example is shown below. The background color for the map was set with a GOPTIONS (graphic options) statement – this statement also draws a border around the map – however this border is not maintained for all output options. GOPTIONS statements stay in effect until the session ends or until changed or reset. The map was created for the ODS HTML destination.

```

goptions cback=CXe2e3fd border;
ODS LISTING CLOSE;
ODS HTML path=odsout body="county.htm"(title="County 2005")style=sasweb;
goptions colors=(white cx99CCFF CX3333FF cx000099);
legend1 label="County Residents with Health Insurance" across=1
      value=(j=1 "Less than 78.5% " j=1 "78.5% to 83%" j=1 "83% to 86.5%"
            j=1 "Greater than 86.5%");
Title2 h=1.4 'Health Insurance Coverage by County 2005';

proc gmap data=hi_2005m map=maps.uscounty;
  id state county;
  choro insured_cat / discrete levels=4 legend=legend1 outline=black;
run; quit;

ODS HTML CLOSE;
ODS LISTING;

```

The map code shown above is for the 2005 data. For 2011, the data set name and titles are changed. To output the Missouri map, the following modification was made to the GMAP statement, using Missouri's FIPS (Federal Information Processing Standard) code of 29.

```

proc gmap data=hi_2005m(where=(state=29)) map=maps.uscounty(where=(state=29));

```

## EXAMPLE TWO – DOT DENSITY MAP

The maps in this example show locations for the one hundred most expensive hospitals in the United States as identified in 2002 and 2012. Most expensive is defined in terms of charge-to-cost ratio. In 2002, the highest charge-to-cost ratio was 1100%, in 2012 1200%. What 1200% means is that the hospital charged \$1,200 for every \$100 of hospital cost. While the top charge-to-cost ratio didn't differ greatly from 2002, the number of hospitals near that top mark did. A standard dot density map is usually a simple scatter of locations; the dot density map below also increases dot size and deepens dot color for each 100% increase in charge-to-cost ratio.

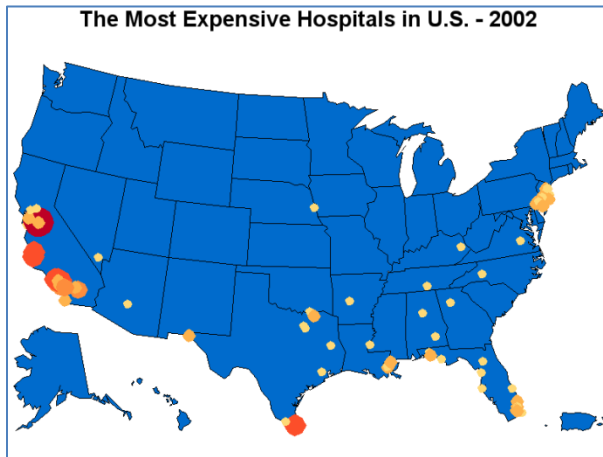


Figure 5. Most Expensive Hospitals U.S. 2002.

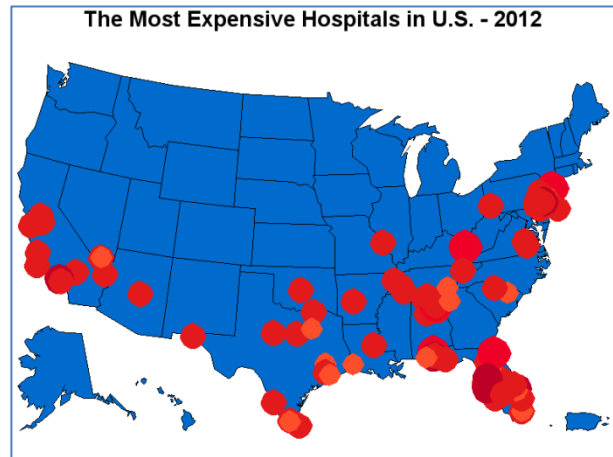


Figure 6. Most Expensive Hospitals U.S. 2012.

Color and size values are stored in the data set based on the cost-to-charge percentage values. These values are then passed to an ANNOTATE data set called DOTS. The ANNOTATE data set is a special SAS data set with specific variables for graphics manipulation. Each observation in the data set is a drawing command defined by values of these variables. The ANNOTATE commands stored in this data set are executed by including the following option in the SAS/GRAPH code. For this paper, ANNOTATE data sets are used with the GMAP procedure.

```

/ANNOTATE=<annotate-data-set>

```

SAS/GRAPH interprets and executes the drawing commands along with the graph and creates output that incorporates the ANNOTATE commands. The following ANNOTATE variables are used in the example below (listed in order of use). The X and Y variables used by ANNOTATE are passed to the ANNOTATE data set from the geographic location in the hospital cost data set.

- XSYS – coordinate system for the X variable.
- YSYS – coordinate system for the Y variable.
- HSYS – The type of units for the size (height) variable.
- STYLE – Font/pattern of a graphics item.
- TEXT – Font value to use in a label, symbol, or comment.
- POSITION – Placement/alignment of text or symbol.
- COLOR – Color of graphics item.
- SIZE – Size of the graphics item specific to the active function. For example, size is the height of the character for a label function.

XSYS and YSYS of '2' ties the coordinates of the ANNOTATE output to the data values. HSYS of '3' sets the coordinate system to percentage of output area. Position of '5' centers the output. For information on the possible values for each of these functions, see SAS documentation at <http://support.sas.com/documentation/>.

For the above dot density maps, the ANNOTATE data set code and the GMAP code is shown below. Note the ANNOTATE= option on the choro statement and that the only response data used by GMAP is through ANNOTATE.

```
data dots;
  length function color $ 8 text $ 8;
  retain function 'label' xsys ysys '2' hsys '3' when 'a' style 'Marker';
  set hosp2002;
  text='W'; position='5'; color=color; size=size;
output;run;

ODS LISTING CLOSE;
ODS HTML path=odsout body="hosp.htm" (title="Hosp 2002") style=sasweb
pattern1 color=cx006BCC;
Title 'The Most Expensive Hospitals in U.S. - 2002';

proc gmap map=maps.us data=maps.us;
  id state;
  choro state/levels=1 annotate=dots nolegend;
run; quit;

ODS HTML CLOSE;
ODS LISTING;
```

### EXAMPLE THREE – MAPPING MULTIPLE FIELDS ON SINGLE MAP

When values do not overlap, it is possible to map values for different variable s or categories on a single map. These maps show the states with the highest percentage of members in three different insurance categories: uninsured, government health insurance and private health insurance. Note the move away from private insurance in the Midwest and the proportional increase in uninsured in the Deep South.

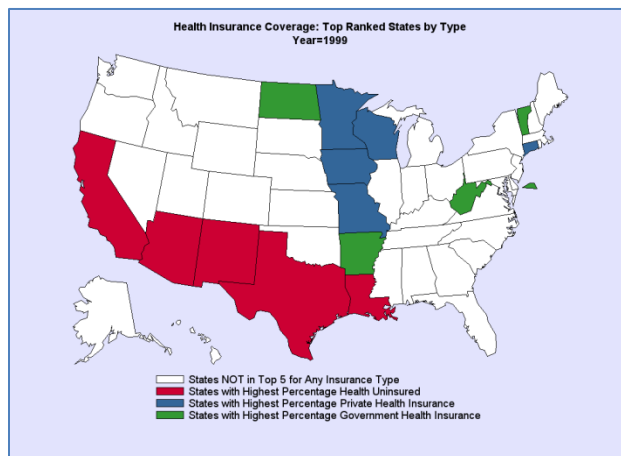


Figure 7. Health Insurance Coverage by Type 1999.

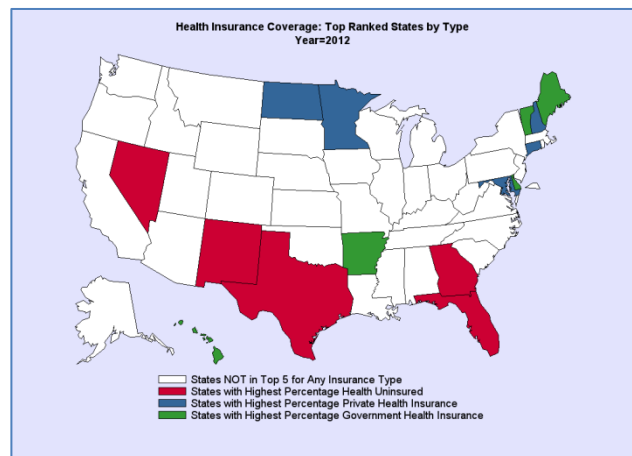


Figure 8. Health Insurance Coverage by Type 2012.

For these maps, the states with top percentages for each insurance category were determined using PROC RANK and then assigned levels in a map data set. Both years were run from a single data set using the SAS BY statement. Because the data set includes three distinct categories, colors from different color families were selected.

```
proc rank data=insurancetype out=ins_rank descending ties=low;
  var uninsured_p totpriv_p totgov_p;
  ranks uninsured_r totpriv_r totgov_r;
by year;run;

data ins_rank;
set ins_rank; mapvar = 0;
  if uninsured_r <= 5 then mapvar=1;
  if totpriv_r <= 5 then mapvar=2;
  if totgov_r <= 5 then mapvar=3;
run;

ODS LISTING CLOSE;
ODS HTML path=odsout body="rank.htm" (title="Ranked Map") style=sasweb;

goptions colors=(white cxC00033 CX336699 CX339933);
legend1 label=none across=1 value=(j=1 "States NOT in Top 5 for Any Insurance
Type" j=1 "States with Highest Percentage Health Uninsured"
j=1 "States with Highest Percentage Private Health Insurance"
j=1 "States with Highest Percentage Government Health Insurance");
Title2 h=1.2 'Health Insurance Coverage: Top Ranked States by Type';

proc gmap data=ins_rank map=all_maps ;
id statecode;
choro plotvar / discrete levels=4 legend=legend1 outline=black;
  by year;
run;quit;

ODS HTML CLOSE;
ODS LISTING;
```

## EXAMPLE FOUR – REGIONAL MAP WITH EXPLODED DISTRICT OF COLUMBIA

The maps below show rate of uninsurance for health insurance for nine U.S. health regional divisions as defined by the National Center for Health Statistics. Data is shown for 2005 and 2013. Rates do not include persons over age 65. Providing separation between the divisions enhances identification of the states included in each division. The explosion of the District of Columbia (D.C.) allows the rate to be visible for that district. Note the increase in rate of uninsurance, especially in the Pacific and East South Central divisions.

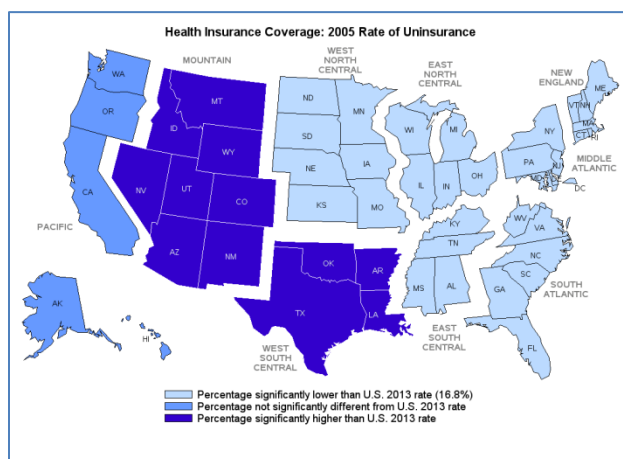


Figure 9. Rate of Uninsurance 2005.

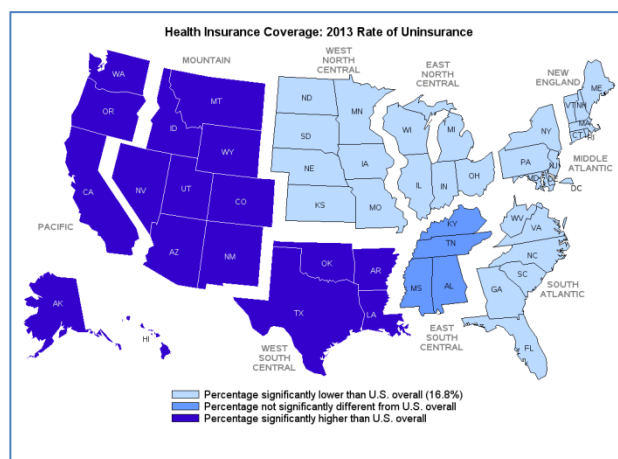


Figure 10. Rate of Uninsurance 2013.

The maps in this example illustrate several techniques, including labeling states with state codes, exploding the D.C. area for data display, separating health service divisions, labeling these divisions, and changing the color of state outlines and labels based on data response values.

To label the states with the two-letter postal abbreviation codes, these codes are first pulled from the MAPS.US data set using the FIPSTATE function. Using this data set allows retaining the latitude and longitude associated with the states. The center of each state was calculated for inclusion in an ANNOTATE data set as illustrated in the code below.

```
proc sql;
create table state_anno as
  select unique fipstate(state), min(x)+(max(x)-min(x))/2 as x,
    min(y)+(max(y)-min(y))/2 as y
from us
group by st;
quit; run;
```

To explode the D.C. area to a size and area where response values are visible, the boundary coordinates were increased by a set factor and then moved to the right of Maryland. The factor can vary depending on how large you would like the D.C. to be. The original coordinates were then dropped from the map data set.

Moving D.C. to the right is similar to separating the regional divisions. To separate the divisions, states are first assigned to the nine displayed divisions. Then the map coordinates of the divisions are adjusted for separation in the data set. Here is partial code for adjusting division locations. How much to move each division is dependent upon the shape of the states; it is NOT a "one size fits all" approach.

```
data us; set us;
if division='West South Central' then do;
  x=x-.005; y=y-.02;
end;
if division='South Atlantic' then do;
  x=x+.05; y=y-.03;
end;
.....
.....
run;
```

The division labels are created in an ANNOTATE data set and placed on the map, shifting the x and y coordinates to move to an area just outside the division from a calculated division center. Calculation follows the form of the state center code above. The data set created is called DIV\_ANN0 and is used in the partial code below to create the ANNOTATE data set with label placement instructions. The label was moved from the center using the graphics unit selected, in this case CELLS or character cells. Other options are CM (centimeters), IN (inches), PCT (percent of output area) or PT (points – there are 72 per inch). The position "B" is centered half a cell above; "E" is centered half a cell below.

```
proc sql;
create table div_anno as
  select unique division, min(x)+(max(x)-min(x))/2 as x,
    min(y)+(max(y)-min(y))/2 as y
from us
group by division;
quit; run;

data div_anno; set div_anno;
length function color $8 style $20 text $100; xsys='2'; ysys='2'; hsys='3';
  when='a';function='label'; size=2; color='gray'; style="albany amt/bold";
  if division='New England' then do;
    y=.225; x=x-.045;
    position='B'; text='NEW'; output;
    position='E'; text='ENGLAND'; output;
  end;
.....
.....
run;
```

Because there are a number of ANNOTATE data sets used to create this map, they are combined into a single data set using the SAS SET statement so that a single ANNOTATE data set can be used by the GMAP procedure. Here is the code to selectively create WHITE state borders and to join that ANNOTATE data set with other ANNOTATE data sets.

```

data annost;
set us ;
  by state segment;
  if insured= 4 then do;
    retain color 'white' xsys '2' ysys '2' when 'a' ;
    if first.segment or (lag(x)=. and lag(y)=.) then function='POLY';
    else function='POLYCONT';
    if x and y then output;
  end;
run;

data all_anno;
set state_anno div_anno annost ;
run; quit;

```

The GMAP procedure code is straightforward and is shown below. Note that the code includes the COUTLINE = BLACK option, which sets state outlines to black. The "WHEN 'a'" setting in the ANNOTATE data set (above) sets the WHITE color only AFTER the map is created, superseding the COUTLINE option. Therefore the correct states have white boundaries.

```

ODS LISTING CLOSE;
ODS HTML path=odsout body="uninsured.htm" style=sasweb;
goptions colors= (cxBBDAff,cx6699FF, cx3300CC) htext=1.2;

legend1 label=none across=1
  value=(j=1 "Percentage significantly lower than U.S. 2013 rate (16.8%)"
    j=1 "Percentage not significantly different from U.S. 2013 rate"
    j=1 "Percentage significantly higher than U.S. 2013 rate");
Title1 " ";
Title2 h=1.4 'Health Insurance Coverage: 2005 Rate of Uninsurance';

proc gmap data=mydata map=all_maps anno=all_anno;
id region division st;
choro ins / discrete legend=legend1 coutline=black
run;quit;

ODS HTML CLOSE;
ODS LISTING;

```

## CONCLUSION

This paper uses the SAS/GRAPH GMAP procedure to illustrate health insurance changes over time using varied display techniques. The built-in flexibility of the SAS/GRAPH GMAP procedure, especially when combined with SAS ANNOTATE and the ODS system, provides all the tools needed for time series analysis with a spatial component. Illustrations produced with the demonstrated techniques can provide a foundation for further analyses for identifying relationships and critical factors impacting health insurance or any industry.

## REFERENCES

- Cohen RA, Martinez ME., "Health insurance coverage: Estimates from the National Health Interview Survey, 2005." National Center for Health Statistics. June 2006. Available from: <http://www.cdc.gov/nchs/nhis.htm>.
- Herman B. "100 Most Expensive Hospitals: National Nurses United Unveils Data." Institute for Health and Socioeconomic Policy. January 2014. Available from: [http://nurses.3cdn.net/e2086b18382cb8e96d\\_0dm6b5ad3.pdf](http://nurses.3cdn.net/e2086b18382cb8e96d_0dm6b5ad3.pdf).
- "The IHSP Hospital 200: The Nation's Most--and Least--Expensive Hospitals." Institute for Health and Socioeconomic Policy. June 2003. Available from: [www.healthwatchusa.org/downloads/IHSP-HospCosts-2003.pdf](http://www.healthwatchusa.org/downloads/IHSP-HospCosts-2003.pdf).
- Martinez ME, Cohen RA. "Health insurance coverage: Early release of estimates from the National Health Interview Survey, January--June 2013." National Center for Health Statistics. December 2013. Available from: <http://www.cdc.gov/nchs/nhis/releases.htm>.
- Mink D, Pasta DJ. "Improving Your Graphics Using SAS/GRAPH® Annotate Facility." *Proceedings of the 31<sup>st</sup> SAS User Group International Conference*. March 2006. Available from: <http://www2.sas.com/proceedings/sugi31/085-31.pdf>



Okerson B. "SAS® Maps as Tools to Display and Clarify Healthcare Outcomes." *Proceedings of the SAS Global 2010 Conference*. Available from: <http://support.sas.com/resources/papers/proceedings10/211-2010.pdf>.

Phillips J. "Outbreak Maps: Visual Discovery in Your Data." *Proceedings of the SAS Global 2011 Conference*. Available from: [http://support.sas.com/md/papers/sgf11/59184\\_Outbreak\\_Maps.pdf](http://support.sas.com/md/papers/sgf11/59184_Outbreak_Maps.pdf).

SAS Institute. SAS/GRAPH 9.2. Reference, Second Edition. Available from: <http://support.sas.com/documentation/cdl/en/graphref/63022/HTML/default/viewer.htm>.

U.S. Census Bureau. Small Area Health Insurance Estimates (SAHIE). Available from: <http://www.census.gov/did/www/sahie/data/index.html>.

U.S. Census Bureau. "Table HIB4. Health Insurance Coverage Status and Type of Coverage by State All People:1999 to 2012." Available from: <http://www.census.gov/hhes/www/hlthins/data/historical/files/hihist4B.xls>.

## **ACKNOWLEDGMENTS**

I would like to acknowledge the members of the WellPoint SAS Users Group, especially Joe Urbi, for encouragement, ideas, and suggestions. Additionally I would like to thank the WellPoint Healthcare Analytics division for their support in the production of this paper.

## **CONTACT INFORMATION**

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

Barbara B. Okerson, Ph.D., CPHQ, FAHM  
Health Information Manager  
West Region Client Reporting & Analytics  
WellPoint Health Care Analytics  
Email: [barbara.okerson@wellpoint.com](mailto: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.