A Map Is Worth A Thousand Words: An Example from Text Pathology Data

Hongyuan Gao • Group Health Research Institute

GroupHealth®
RESEARCH INSTITUTE

Introduction

Most of the healthcare data have a spatial dimension. Maps convey the geographical information more visually than tables or other non-map graphs. However, map creation usually involves specialized software and is time-consuming.

In SAS 9.2, creating highly-customized maps becomes surprisingly easy using GMAP procedure in SAS/GRAPH. This poster aims at showing how easy it is to create maps just within SAS/GRAPH.

Map Creation in SAS/GRAPH

Step 1: Prepare a map data set and a response data set

➢ A map data set contains the information needed to draw map boundaries. The minimum set of variables are X and Y coordinates, and ID. X and Y coordinates are longitude and latitude expressed in radians; ID is the geographic area associated with each pair of X-Y coordinates. SAS/GRAPH offers some map data sets in the MAPS library. Also, lots of shapefiles are available online and can be converted into SAS map datasets by PROC MAPIMPORT.

➢ A response data set contains the information that will be displayed on the map. In order to allow SAS/GRAPH to match the response data set to the map data set, the response data set must contain the same ID variable as the ID variable in the map data set.

Step 2: Match two datasets using PROC GMAP

```
PROC GMAP DATA = response_dateset
    MAP  = map_dataset;
    ID ID_variable;
    CHORO response_variable / options;
RUN;QUIT;
```

Step 3: Let the Annotation do   *Magic!*

The most powerful tool for customizing maps is the Annotation, which is also true for other graphs. It instructs SAS/GRAPH on how to customize graphs as if you were drawing with a pencil on a paper. Tables in the next column are some examples of annotation data sets.

Text Pathology Data

Text pathology reports of GHC from 2005 to the present are stored in the table Daily_HL7 in MS SQL-Server on CTRHS-SQL2K in Pathology database. Text pathology reports between 1976 and 2005 are stored in two other tables in the same database. Data of following graphs are mainly based on text pathology reports in 2010 (excluding pap smear reports and autopsy reports) and enrollment in July 2010. Data from 2005 to 2010 are displayed in: http://ghri-wiki/datawiki/perspective. aspx?action=view&page=data:PathologyReportNumber.

An annotation data set for adding labels:

X	Y	Function	Style	Text	Comment
5	85	Label	Special	M	Add a star in (5,85)
5	85	Label	'Swiss'	Seattle	Add text "Seattle" in (5,85)

An annotation data set for adding a reference line:

X	Y	Function	Color	Size	Comment
5	85	Move			Move to (5,85) to prepare the draw
10	85	Draw	Red	2	Draw a line from (5,85) to (10,85)

An annotation dataset for adding an image:

X	Y	Function	Style	Imgpath	Comment
5	85	Move			The lower left corner of the image is (5,85)
10	95	Image	Fit	Path\image.gif	The upper right corner of the image is (10,95)

An annotation dataset for adding a polygon:

X	Y	Function	Style	Comment
5	85	Poly	Solid	Move to (5,85) to prepare the draw of a polygon
10	85	Polycont	Solid	Draw a line from (5,85) to (10,85)
10	75	Polycont	Solid	Draw a line from (10,85) to (10,75); since this is the end of the polygon, the end point is linked with the start point automatically to form a triangle

An annotation dataset for adding a pie:

X	Y	Function	Rotate	Color	Style	Comment
5	85	Pie	270	Red	Solid	Fill ¾ of the pie with red
5	85	Pie	90	Blue	Solid	Fill ¼ of the pie with blue

Choropleth Maps of Text Pathology Data

Number of Pathology Reports 2005-2010
by county

⭐ **City where Group Health Primary Care Medical Centers locate in 2010**

Number of Pathology Reports in 2010
by PCC and by zipcode

Number of Pathology Reports by PCC:
500 / 1000 / 1500 / 2000 / 2500 / 3000

Number of Pathology Reports by Zipcode:
0-50 / 51-100 / 101-150 / 151-200 / 201-250 / 251-300 / 301-350 / over 351

Number of Pathology Reports and Enrollment
from all enrollees in 2010 by zipcode

Number of Pathology Reports:
100 / 200 / 300 / 400 / 500 / 600

Number of Enrollment:
0-1000 / 1001-2000 / 2001-3000 / 3001-4000 / 4001-5000 / 5001-6000 / 6001-7000 / over 7000

Number of Pathology Reports and Enrollment
from enrollees older than 65 in 2010 by zipcode

Number of Pathology Reports:
50 / 100 / 150 / 200 / 250 / 300

Number of Enrollment:
0-100 / 101-200 / 201-300 / 301-400 / 401-500 / 501-600 / 601-700 / over 700