"Dot Maps"
Made Simple

Robert E. Allison, Jr., North Carolina State University College of Textiles, Raleigh, NC
and SAS Institute Inc., Cary, NC
Moon W. Suh, North Carolina State University College of Textiles, Raleigh, NC

ABSTRACT

This paper describes a simple method to create dot maps with SAS/GRAPH® software. The data values are represented as dots, with the area of the dots scaled proportionally to the data values. The code was developed at North Carolina State University College of Textiles to visualize data in the Textile and Apparel Business Information System (TABIS).

INTRODUCTION

The GMAP procedure provides several methods for creating choropleth, surface, block, and prism maps, but there is no built-in method to create dot maps. This paper describes how to create dot maps and pie-chart maps using the SAS System.

This paper provides very simple example programs that you can use to create dot maps of your data.

The first example shows how to create a US dot map by state. The second example generates a US dot map by county. The last example demonstrates how to turn dot maps into pie-chart maps.

DOT MAPS BY STATE (Example 1)

First, you must generate a SAS data set containing the values to be mapped. The numeric variable state is used to hold the fips state identifier. The character variable st is included for readability. Only a small subset of the actual data is shown here.

data A;
  input st $ state value;
cards;
  CA  6 13.7207
  FL 12  6.6925
  NC 37  2.7424
;

The max_area macro variable defines the area of the dot representing the maximum value in the input data set. You can adjust the value of this variable to accommodate various data values and personal preferences.

%let max_area=20;

The following call to the SQL procedure performs three queries. The first query creates a max_val macro variable containing the maximum data value. The approximate area of each dot is added to data set A by scaling the values to the maximum desired dot area. The radius, or size, is then computed for each dot based on the area and added to data set B along with the coordinates for the center of each state.
proc sql;

select max(value)
format comma5.2 into :max_val from A;

create table A as select unique state, value, (value/max_val)*max_area as area from A;

create table B as select A.*, uscenter.*, sqrt(a.area/3.14) as size from A, maps.uscenter where A.state=uscenter.state;

quit; run;

This code creates an annotate data set containing a dot of the appropriate size at the center coordinate of each state.

data dots; set B;
xsys='2'; ysys='2'; hsys='3'; when='A';
function='PIB'; style='SOLID'; color='BLACK';
lines=0; angles=0.0; rotate=360.0;
if ocean='Y' then output;
run;

These graphics options and titles were used to generate the maps in Figure 1. You can easily adjust the font, size, and text of the titles. Also, depending on your printer, you might need a different target device (ps is used for standard postscript printers).

goptions reset=global gunit=pct
rotate=landscape targetdevice=ps ftitle=zapfb ftext=swissl cback=white ctext=black colors={black} htitle=5 htxt=4;
title1 '1992 Female Clothing Sales';
title2 "(maximum dot represents max_val billion $)";

The GMAP procedure prints the US map, and overlays the annotate data set containing the dots.

pattern v=e r=1000;
proc gmap data=maps.us map=maps.us anno=dots;
id state; choro state / nolegend
discrete coutline=black;
run;

Figure 1

1992 Female Clothing Sales
(maximum dot represents 13.72 billion $)
DOT MAPS BY COUNTY (Example 2)

In this example, the county variable is added to the data structure. Again, only a small subset of the actual data is shown.

```plaintext
data A;
  input state county value name $;
cards;
6  37 100010  CA/LOS_ANGELES
37 183 1499  NC/WAKE
12 95 262   FL/ORANGE
;
```

You will need to scale the data values to smaller dot areas because the counties are much smaller than the states.

```plaintext%
let max_area=7;
```

```plaintext
proc sql;
  select max(value) into :max_val from A;
  select max(value) format=comma6. into :cmax from A;
create table A as select state, county, value,(value/&max_val)/&max_area as area from A;
create table B as select A.*,sqrt(A.area/3.14) as size from A;
```

Although there was already a data set containing the center coordinates of each state, these coordinates must be computed for the county map. You can use this query to create a data set containing the approximate center of each county.

```plaintext
create table center as
  select unique state, county, 
         (max(x)+min(x))/2 as x,
         (max(y)+min(y))/2 as y
  from maps.uscounty group by state, county;
create table B as select B.*, center.x, center.y from B, center
  where B.state=center.state
  and B.county=center.county;
```

The outlines of the counties might obscure the dots and clutter the map, so only the state outlines are printed in this example. Also, since the state outlines for Alaska and Hawaii are not in the same position as the county outlines, and because the SAS maps do not represent counties in Alaska, a query is used to eliminate Alaska and Hawaii from the county dot maps. The maps also print much faster when using the state outlines instead of the county outlines.

```plaintext
create table cont_us as select * from maps.us where state=15 and state=2;
quit; run;
```

The dots/goptions/title code for the county map is very similar to that used in the state map example. The GMAP procedure is the same except only the continental states are used.

```plaintext
data dots; set B;
xsys='2'; ysys='2'; hsys='3'; when='A';
function='PIE'; style='SOLID'; color='BLACK';
lines=0; angle=0.0; rotate=360.0; run;
goptions reset=global quit=pct
  rotate=landscape targetdevice=ps ftitle=zapf
  ftext=swissl cback=white ctext=black
  colors=(black) htitle=5 htext=4;
title1 "1992 Textile Employees per County";
title2 "(maximum dot represents &cmax employees)";
pattern1 v=e r=1000;
proc gmap data=cont_us
  map=cont_us
  anno=dots; id state; choro state
  nolegend discrete coutline=black; run;
```

It is difficult to conduct an unbiased comparison of states with differing sizes and population densities by simply mapping the raw data. So it is often useful to map the data in different ways.

For apparel manufacturing, the employment data can be combined with population data to produce a value representing the percentage of each county's population employed in apparel manufacturing. This shows the importance of apparel manufacturing in the southeast US. Both employment maps are shown in Figures 2 and 3.

The female clothing sales data shown in Figure 1 could be combined with population data to compute per capita sales. The per capita values can be compared among states regardless of the state's size, population, or population density.
Figure 2

1992 Apparel Manufacturing Employees per County
(maximum dot represents 1000 employees)

Figure 3

1992 Apparel Manufacturing Employees per County
as Percent of County Population
(maximum dot represents 17%)

PIE-CHART MAPS (Example 3)

This example shows how to display more information on the same map by changing the dots to pie charts. With the code from Example 1, and these few changes, you can produce pie-chart maps.

First, include variables for the female, male, and total clothing sales in the data structure.

```sas
data A;
input st $ state female male value;
cards;
CA  6   13.72  6.97  20.69
FL  12  6.69   3.04  9.73
NC  37  2.74   1.30  4.04

You may increase the size of the maximum dot to 50 since overlapping pie charts are easier to distinguish than overlapping solid dots.

%let max_area=50;

Also, sort data set A in descending order so that the smaller pie charts are printed in front of the larger ones and not hidden.

```sas
create table A as
select unique
state, female, male, value,
(value/max_val)*max_area as area
from A
order by value descending;
```

Each pie is created in three steps. First, create a solid black slice to represent the female clothing sales. Second, create a solid white slice for the male clothing sales.

```sas
data dots;
length function $ 8; set B;
when='A'; xsys='2'; ysys='2'; hsys='3'; line=0; function='PIE'; style='SOLID';
percent=female/value*100;
rotate=percent*360/100;
if ocean='Y' then output;

percent=male/value*100;
rotate=percent*360/100;
if ocean='Y' then output;
```

Third, create a black circle that surrounds the entire pie and distinguishes the white slice from the white background.

```sas
line=0; style='EMPTY'; angle=0.0;
rotate=360.0; color='BLACK';
if ocean='Y' then output;
```

Add records to the annotate data set to provide a legend. The legend consists of a solid black pie beside the word "female", and an empty pie beside the word "male".

```sas
data legend; length function text $ 8;
xsys='2'; ysys='2'; hsys='3'; when='A';
color='BLACK'; line=0; angle=0.0; rotate=360.0;
function='PIE'; size=1.5;
styles='SOLID';
x=17; y=.23; output;
styles='EMPTY';
x=17; y=.20; output;
function='LABEL';
size=4; styles='SSWISS'; position='6';
text='female';
x=.19; y=.235; output;
text='male ';
x=.19; y=.205; output;
run;
```

Then, merge the legend with the annotate data set containing the dots.

```sas
data dots; set dots legend;
run;
```

As before, create the map using the GMAP procedure, and overlay the dots data set containing both the pie charts and the legend.

```sas
pattern vse=rs1000;
proc gmap data=rnaps.us map=maps.us anno=dots;
id state; choro state / nolegend
discrete coutline=black;
run;
```

This pie-chart map in Figure 4 shows both the total amount of clothing sold in each state (by the size of the pies), and the distribution between male and female clothing sales (by the size of the pie pieces). The pies show that more money is spent on female clothing in all the states.
CONCLUSIONS

With the code presented in this paper, you should be able to create a variety of dot and pie-chart maps to suit your data mapping needs.

REFERENCES


ACKNOWLEDGEMENTS

A special thanks to Carl Priestland, Jack Bulkley, Jay J. Wind, Frank Dilorio, and Paul Marsh for their help and encouragement.

SAS, and SAS/GRAPH are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

The authors welcome questions and comments via e-mail (realliso@tx.ncsu.edu and moon_suh@ncsu.edu) or the following address:

Box 8301
North Carolina State University
College of Textiles
Raleigh, NC 27695-8301 USA