

SAS/Intrnet® and Census Mapping: How Low Would You Like to Get

Lakshmi Pandey, William Joseph (Joey) Smith and David Sjoquist
Fiscal Research Program, Andrew Young School of Policy Studies
Georgia State University, Atlanta, GA 30303

Abstract

This paper details the steps involved with using SAS procedures and program components to create a web-enabled Geographic Information System.

With the increased availability of geographic data via the Internet, data managers are put into a position of trying to make the available data user-friendlier. Access to data now entails more than physically obtaining columns of numbers. It means presenting the data in an understandable way to individuals with little or no technical knowledge. This paper shows how an intermediate SAS user, utilizing features integrated into the base SAS System can easily produce maps of geographic data, and with the addition of ODS (Output Delivery Services) can create live and interactive reports on the web that can be used by individuals without any knowledge of SAS, HTML or any other markup or programming language.

Introduction

An increased emphasis has been placed upon the use of geographic data in the presentation of ideas and the development of strategies. The push has been towards the “democratization” of data through the development of geographic systems that allow the non-programmer access to spatial data. By presenting information in a geographic context people are able to quickly internalize the data and interpret patterns that may be useful in strategic planning and decision-making activities. Furthermore, by using the Internet (or a web browser) as the client interface, the costs associated with acquainting the user with the geographic data system is, at least to some extent, alleviated.

In this paper we detail the steps involved in developing an interactive web-based Geographic Information System to display 1990 and 2000 census data. We demonstrate how components of the SAS system, such as SAS/Intrnet, ODS and PROC GMAP, can be used to develop a web-based geographic information system

intended to analyze geographic data, identify important geographic features, and create high quality reports which are necessary to be competitive in business, government and educational activities.

The goal of this paper is to familiarize the average SAS programmers with the SAS features that allow for the creation of such an application. Furthermore, this paper, through examples, will provide a set of tools that will enable an intermediate-level SAS programmer to create, and administer a web-based data mapping system.

Although some SAS and HTML programming experience are necessary to create the website, no special training or program is required to use the site once it is created. Regardless of whether the resulting website is a simple map that provides store locations based on a submitted address, or a web-based interface for an extensive geographic database, we show SAS users how to leverage SAS programming skills to provide an audience access to information.

Our Project

The amount of information and data coming from the 2000 Census is enormous. Summary File one (SF1) itself has 284 tables for each state. Summary File 3 (SF3), which is in the process of being released, has even more tables than SF1. An individual interested in small subset of the information coming from the census would have to navigate this mountain of tables, along with geographic features like county, tract, block group and blocks. Many people would very quickly find themselves lost in this jungle of information.

There are resources available that make the census more approachable for the average individual. The Bureau of the Census has made an effort at allowing web access to the data; however, this access is limited to “commonly requested” selections. There are also third parties that have made a cottage industry of creating front ends for selections of the census data. Furthermore there are individuals

(academicians and government researchers mostly) who have analyzed these data and have made reports and tables available, but access to information about one's own neighborhood remains relatively difficult to extract. Additionally, tables and numbers do not clearly answer questions like *how has my neighborhood changed in past decade?*

In a joint project the *Fiscal Research Program* of *Andrew Young School of Policy Studies* at *Georgia State University* and the *Atlanta Regional Commission* have developed a website, <http://atlantacensus2000.gsu.edu> that addresses the issue of data democratization using SAS as a foundation. This website provides an easy way to map any geography-based data element from the Census of Population 2000, from state to block-level. It also has provision to make comparison with 1990 Census of Population and Housing so that one can visualize changes or pattern of changes over past decade has brought in.

On the website all the maps, with the exceptions of reports on particular topics of interest, are drawn dynamically and made available through Internet. SAS/Intrnet and Basic SAS procedures are used to create maps and tabular data.

The rest of the paper is organized into the following sections, DATA: descriptions of acquiring necessary data like nodes to draw maps and census demographic information; US MAP: displaying US maps by states; STATES MAP: maps county data for individual states; COUNTY MAP: maps of counties by 2000 census tracts; and TRACTS MAP: maps of ten 2000 tracts around a reference tract.

DATA

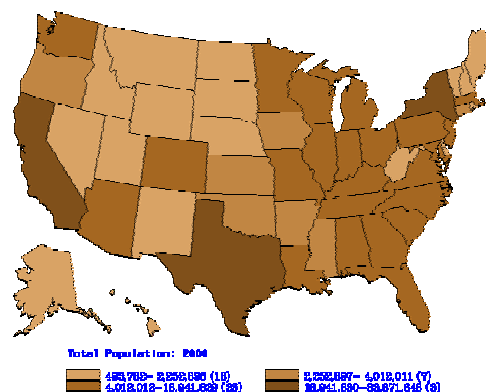
The Bureau of Census provides 1990 and 2000 Tiger/Line files for each county of all 50 states and the District of Columbia and PROC GIS has been used to imports these Tiger files to provide annotation for streets. The U.S. Bureau of Census also provides boundaries for states, counties, tracts and block groups in the form of Arc/Info e00/shape and also in ASCII formats. These ASCII files are used to create map files for PROC GMAP. All the tables of SF1 have been converted at block group levels for each county of the United States.

US Map

With minor modifications to webmap.sas file, which comes with SAS/Intrnet, one can create a US maps as

US Map by State

Total Population: 2000



For District of Columbia [Click Here](#)
[Tabular Data](#)
[Comments and Suggestions](#)

Building a map data set that includes drilldown links for individual states and a micro variable for state, however, ATL option is redundant with JAVA option. This option to Proc GMAP allows us to pass a different SAS program file and also fips of a state as micro variable to next step of mapping as follows

```
data us;set census.staterace;
  html = 'href="' || "&_URL" ||
"?_service=&_service" || "&_debug=" || "&debug" ||
'&_program=' || "&_pgmlib..us2_race.sas"
||
'&PLOTVAR="||"&plotvar"||
'&state=' || left(trim(state)) ||
'" alt="State:' || trim(statename) ||'OD'x ||
left(&showvar) || ""';
run;
```

ODS, pattern and some other statements will go here or could be included by %INCLUDE statements.

State Map

A similar SAS program like the one for US map with the following changes in HTML option will let us pass the state fips as micro variable to facilitate the mapping of counties of a state .The

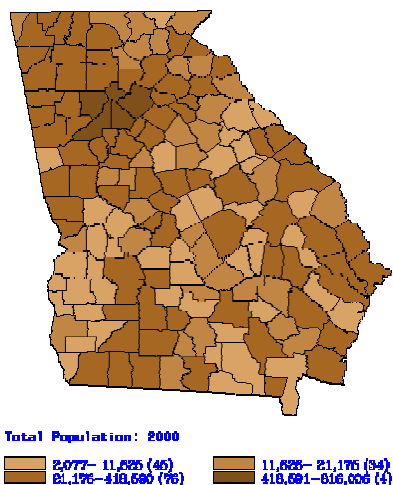
program us2.sas has a statement to display map of a state by county

```
proc gmap map=stmap.state&state data=race;
  id stcnt; format &plotvar ptsfrmt.;
  choro &plotvar/discrete html=html
  legend=legend1 COUTLINE=BLACK;
  run; quit;
```

and it will result into a state map as

State: Georgia

Total Population: 2000



County Map

As in the case of US and state maps, all one has to do is modify the HTML variable in us3.sas to include the census tract so that a county map by census tracts could be drawn as:

```
Data race;set census.race&stcnt;
  html = 'href=' || '&_URL' ||
  '?_service=&_service' || '&_debug=' || '&debug' ||
  '&_program=' ||
  '&_pgmlib..us3_race.sas' ||
  '&PLOTVAR=' || '&plotvar' || '&stctract=' ||
  left(trim(stctract)) || ' alt="State:' ||
  trim(staname) || 'OD'x || trim(countyname) || 'OD'x
  || left(tract) || 'OD'x || left(&showvar) || 'OD'x || '';
```

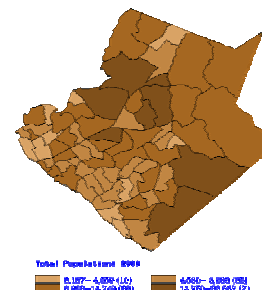
where us3.sas contains following statements to draw county maps by census tract.

```
proc gmap map=tract.tr&stcnt data=race;
  id stctract; format &plotvar ptsfrmt.;
  choro &plotvar/discrete html=html
  legend=legend1 COUTLINE=BLACK;
  run; quit;
```

Following is a map of Gwinnett County of State of Georgia by 2000 census tracts

State: Georgia County: Gwinnett

Total Population: 2000



Census Tracts Map

From the US map, we have clicked our self to a county by tract level map, where each tract shown on the map is tagged with state, county and tract FIPS and given a census tract, one can determine a certain number of census tracts closest to the reference tract or all tracts with centroids within certain distance from the centroid of the reference tract. The following data steps are used to identify the closest eleven census tracts:

```
data center(rename=(xcenter=rx ycenter=ry));set
  tract.trcenter;
  if &stctract=stctract;
  data center(keep=state county tract dis);set
  tract.trcenter;if _n_=1 then set center;
  dis=sqrt((rx-xcenter)**2+(ry-ycenter)**2);
  proc sort;by dis;
  data center(keep=state county tract);
  set center;by dis;
  if _n_<=10;
  run
```

Data set `trcenter` contains the centroids of all the tracts in the country. Depending upon the location of the reference tract, resulting data set `center` of ten tracts may extend to multiple counties or even multiple states. Call to subroutine SYMPUT is used to create macros for each state and county combinations represented in the data set `center` as

```
proc sort data=center;by state county;
  data temp;set center;by state county tract;
  if first.county;
  data _null_;set temp;
```

```
stcnt=int(stctract/10000);
call symput('t'||left(_n_),left(stcnt));
```

These macro variables are used to combine map and spatial data for all counties being represented by ten census tracts and selected for those ten census tracts as

Data

```
apswork.map%sysevalf(&stctract*100,integer);
merge center(in=j) tract.tract&t1 tract.tract&t2
tract.tract&t3 tract.tract&t4 tract.tract&t5
tract.tract&t6 tract.tract&t7 tract.tract&t8
tract.tract&t9 tract.tract&t10;by state county
tract;if j;
```

There should be ten such macros created allowing the possibility of ten counties being represented by the selected ten tracts. Following combines spatial data for maps:

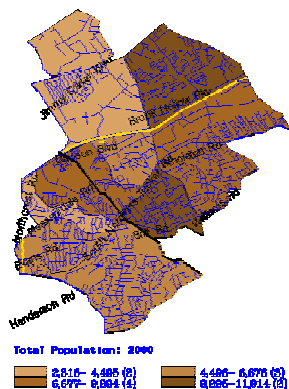
Data

```
apswork.race%sysevalf(&stctract*100,integer);
merge center(in=j) race.race&t1 race.race&t2
race.race&t3 race.race&t4 race.race&t5
race.race&t6 race.race&t7 race.race&t8
race.race&t9 race.race&t10;by state county
tract;if j;
```

Similar statements are used to create annotation of streets. These maps, spatial and annotation data are stored in APSWORK library because they need not to be created again and again when the user just wants to map a different variable. Following is the resulting page a user will be able to display on their browser

State: Georgia County: Gwinnett

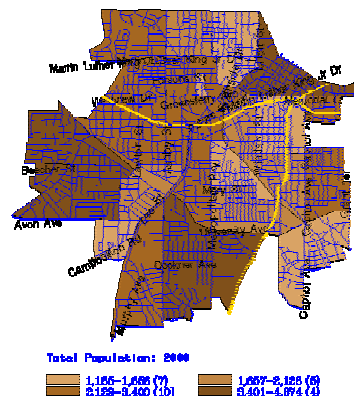
Total Population: 2000



The map above shows the total 2000 population by census tracts for eleven census tracts around reference tract of 504.17 of Gwinnett County in State of Georgia. A similar map could also be drawn for all the tracts within certain distance from the reference tract. However, in rural areas, there might not be any other census tracts within that distance. Following is a map of all the tracts within a 2 miles radius from the centroid of census tract 58 of Fulton County in Georgia:

State: Georgia County: Fulton

Total Population: 2000



Cities and Places Map

City and places are defined at block group levels where only part of the census tract is in a city or place area. Since Census Bureau does not provide a relation between 1990 and 2000 block groups, the best way to visualize the pattern of changes over the past decade is by looking at the 1990 and 2000 maps side by side. PROC GREPLAY has been used to display maps side by side as follows.

```
proc gmap map=city.map00 data=atl00
gout=city.gseg;id county tract bg;
format &plotvar totalv.;
choro &plotvar/discrete coutline=black
html=html name='maps00' legend=legend1
annotate=city.an00;
proc gmap map=city.map90 data=atl90
gout=city.gseg;id county tract bg;
format &plotvar totalv.;
choro &plotvar/discrete coutline=black
html=html name='maps90' legend=legend1
annotate=city.an00;
run;
quit;
```

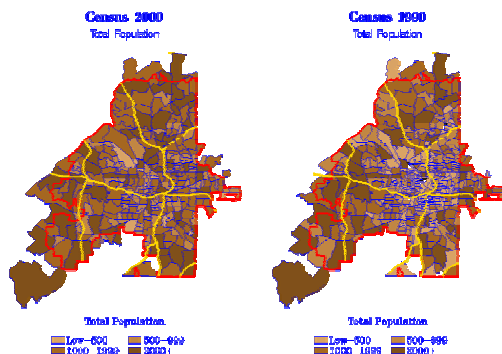
```

options display HSIZE=0;
options display;
proc greplay tc=tempcat nofs;
  igout city.gseg;
tdef newtwo des='two plots of equal size'
1/1lx=0 1ly=0
  ulx=0 uly=100
  urx=50 ury=100
  lrx=50 lry=0
2/1lx=50 1ly=0
  ulx=50 uly=100
  urx=100 ury=100
  lrx=100 lry=0
;
template newtwo;
  treplay 1:maps00
        2:maps90;
run; quit;

```

Following is a map of City of Atlanta displaying 1990 and 2000 Total Population.

City of Atlanta



For the sake of comparison, PATTERN statements are matched to display same color for a population group in both maps. From the above maps one cannot tell which block group has changed by what exact number between 1990 and 2000 but a pattern of changes is quite apparent. It shows that over last decade quite a few of people have moved back to the inner city.

Demography of Neighborhood Where You Live

In the following, a census tract can be selected from the list of 2000 census tracts of a county in a state to define a neighborhood to be mapped. However, if one does not know the tract number, given an address with zip code, American FactFinder, Bureau of Census, can geocode it to

find 2000 census tract. Following three steps let a user select state, county and tract.

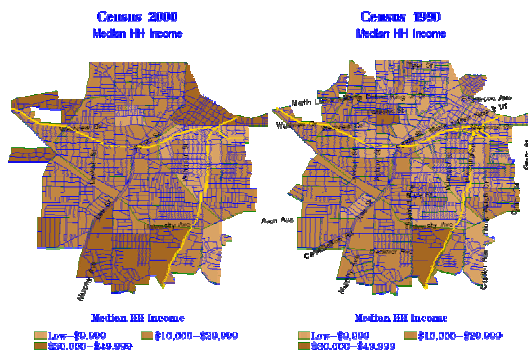
```

data _null_;
  file _webout;
  put '<FORM METHOD="GET" ACTION="'
    '&_url" ""';
  put '  onSubmit="MAPWIDTH.value = '
    '( navigator.appName.indexOf(' "'Microsoft')
    >= 0 ) ?";
  put '    document.body.clientWidth-20 :
window.innerWidth-20;">;
  put '  <INPUT NAME="_PROGRAM"
VALUE=""   'sample.neighbor_us2.sas'   ""
TYPE="HIDDEN">;
  put '  <INPUT NAME="_SERVICE"
VALUE="" "&_service" "" TYPE="HIDDEN">;
  put '  <INPUT NAME="_DEBUG" VALUE=""
"&debug" "" TYPE="HIDDEN">;
  put '<TABLE BORDER="0">;
  put '<TD><SELECT NAME="state">;
  data _null_;set hsql.state;
  file _webout;
  put '  <OPTION VALUE="" state ""
SELECTED>' statename '</OPTION>';
  data _null_;
  file _webout;
  put '  <OPTION VALUE="" SELECTED>
'Select a State </OPTION>';
  put '</SELECT>';
  put '</TD>';
  put '<br>';
  put '  <TD><INPUT TYPE="SUBMIT"
VALUE="Select a State"></TD>';
  put '</TABLE>';
  put '</FORM>';
run;

```

where data set hsql.state has names and fips pair for states. Similar program for drop down menu could let users select a county in a state and a 2000 census tract of county in that state. Users also have options to input miles around the census tract and select different tables of 2000 census. Following the procedure of displaying 11 tracts of counties around a reference tract discussed above and displaying maps of 1990 and 2000 census side-by-side as in the case of City of Atlanta, we can display the maps of a user defined neighborhood as

Demography of Neighborhood Where You Live State and County: Georgia Fulton Tract: 58



Future

This site is very much in development mode. Some of the features being worked out to be included are:

1. Displaying any new data coming out of Census 2000.
2. Selecting census tracts and mapping from centroid of a zip code (alternative to geocoding an address).
3. Creating maps of state and national voting and school districts.
4. Adding other non-census data that are available at the census tract level.
5. Allowing end users to select various features of the map, including color scheme and the number of data intervals.

The principal limitation is that the census data elements available mappings are restricted to those put up on the site. Right now, for example, age is available, but not age by race, by gender, or race and gender. Over time we will expand the complexity of the data elements available.

Summary and Conclusions

In this paper we have tried to demonstrate that starting from US map by states, one could navigate to an area as small as few census tracts. We also showed that it is also possible to display two maps from two different census years side by side for visual comparison. However, there are whole array of features, which are being developed, that will be included in future. With the exception of front page of the web site, each and every page are being generated dynamically by SAS data steps and procedures, and SAS/Intrnet makes it possible to be displayed on an end-users browser.

Contact Information

Any comments and suggestion can be forwarded to

Lakshmi Pandey

Fiscal Research Program

Andrew Young School of Policy Studies

Georgia State University

140 Decatur Street

Atlanta, GA 30303

Work Phone: 404 651 0481

Fax: 404 651 0416

Email: lpandey@gsu.edu

Web: <http://atlantacensus2000.gsu.edu>

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.