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. URL: atlantacensus2000.gsu.edu

Key Words: Base SAS and, SAS programming and Macro Languages, SAS/Graph (Proc Gmap: annotation, HTML and Alternate Text options, Proc GReplay), SAS/GIS (Import Tiger files and SAS/GIS files to SAS/Graph files), ODS (HTML and CSV), SAS/Intrnet

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

Map 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. ASCII file for each boundary has two components, one contains longitude and latitude of centriod and nodes of polygons in degree whereas as the second one provides its geographic identification. Following are few lines from both components for state boundaries:

Nodes:

```
1(ID)-0.152594121361715E+03     0.650971279135803E+02
(centriod)
    -0.147787110884357E+03      0.702453634941157E+02
    -0.147765104000000E+03      0.702198060000000E+02
    -0.147681722000000E+03      0.701999540000000E+02
    -0.147648000000000E+03      0.702032990000000E+02
    -0.147585678000000E+03      0.702033980000000E+02
    -0.147505270000000E+03      0.702003840000000E+02
```

and geographic identifier file

1 (id)

"02" (state FIPS)

"Alaska" (STATE)

"01" (Segment if a state has islands)

" " (filler)

One can read these ASCII files to create SAS/Graph maps files and also could be used to annotate the boundaries of polygons.

Response Data

All the tables of Summary file 3 from Census 2000 and Summary file 3a from Census 1990 have been converted into SAS files up to block group levels for each county of the United States.

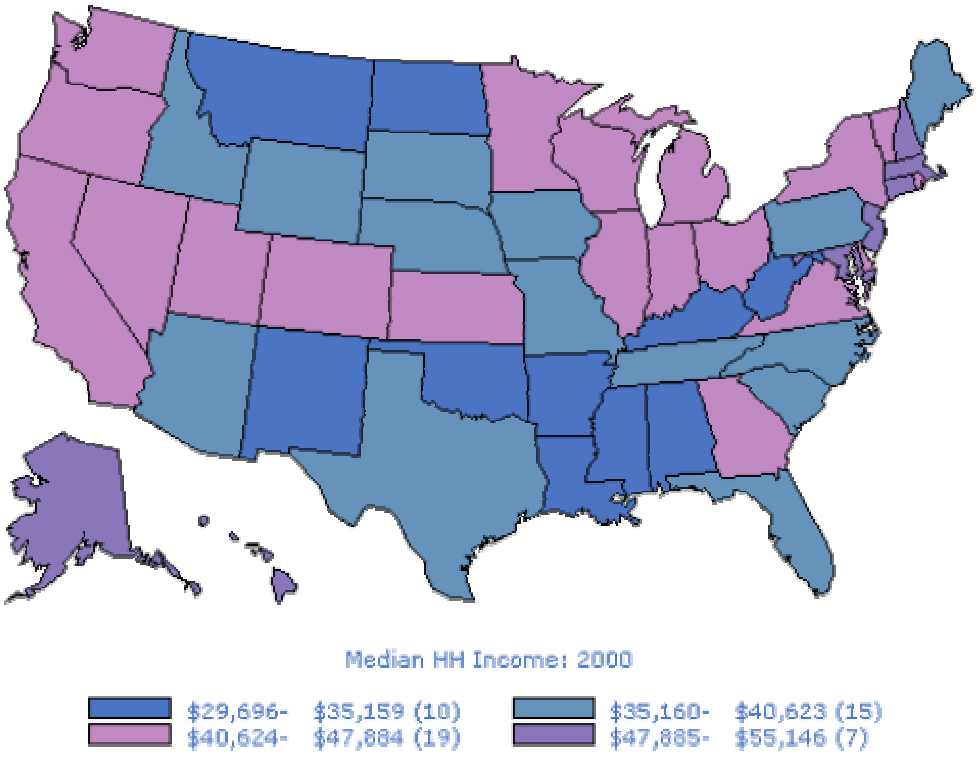
US Map

With minor modifications to webmap.sas file, which comes with SAS/Intrnet, one can create a US maps as shown in Map 1. Building a map data set that includes drilldown links for individual states and a micro variable for state, however, alternate text, 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_income.sas" ||'&PLOTVAR='||"&plotvar"||
      '&state=' || left(trim(state)) ||
      '" alt="State:' || trim(statename) ||'0D'x || left(&showvar) || '"';
 run;
```

Map 1: Household Income Map of 50 States and District of Columbia

Median HH Income: 2000



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

Legend in Map 1 is a result of running PROC FORMAT procedure twice, once to create ranges from minimum to median to maximum values of variable mapped and second to concatenate the frequency counts to those ranges to indicate number of polygons. For example, there are 7 states with median household income in the $47,885-$55,146 range.

State Map

A similar SAS program, us2_income.sas, likes the one for US map will let us map a state by its counties as

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

Where &state macro variable which was passed on from US maps by states (Map 1) is a FIPS for the particular state selected. Above code will result into a map of state by county as shown in Map 2.

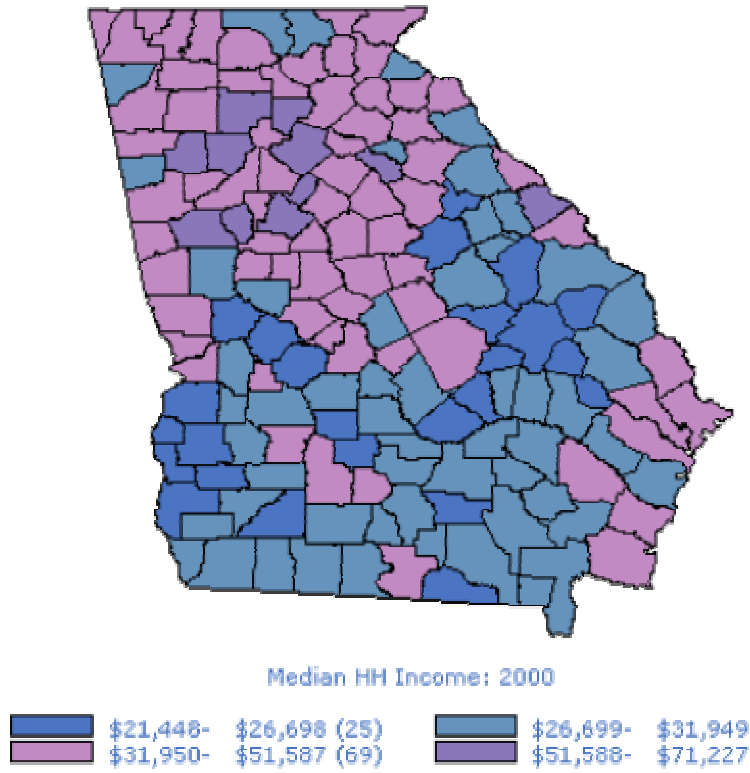
In SAS program us2_income.sas, we can now pass counties FIPS as a micro variable to facilitate the mapping of counties of a state .The program us2_income.sas has a statement to create HTML variable to link each counties of a state with SAS program us3_income.sas as follows:

```
Data income;set census.income&stcnt;
  html = 'href="' || "&_URL" ||
"?_service=&_service" ||'&_debug='||"&debug"||
'&_program=' ||
"&_pgmlib..us3_income.sas" ||
'&PLOTVAR='||"&plotvar"|| '&stcntract=' ||
left(trim(stcntract)) || '" alt="State:' ||
trim(statename) ||'0D'x ||trim(countyname) ||'0D'x
|| left(tract) ||'0D'x ||left(&showvar) ||'0D'x || '";
```

As one can notice, a variable stcnt which is concatenation of state and county FIPS is beings passed on to program us3_income.sas.

Map 2: Household Income Map of State of Georgia by County

State of Georgia
Median HH Income: 2000



County Map

In the program us3_income.sas, a statement containing following steps will produce a county map by its census tracts as shown in Map 3:

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

where &stcnt is state and county concatenated FIPS brought in from state map by its counties.

To further drill-down county map by its tract, one can create a HTML variable in us3_income.sas program to include the census tract so that a subset of tracts could be mapped. Following is the statement to pass census tract as a concatenated FIPS of state, county and tracts, stcntract:

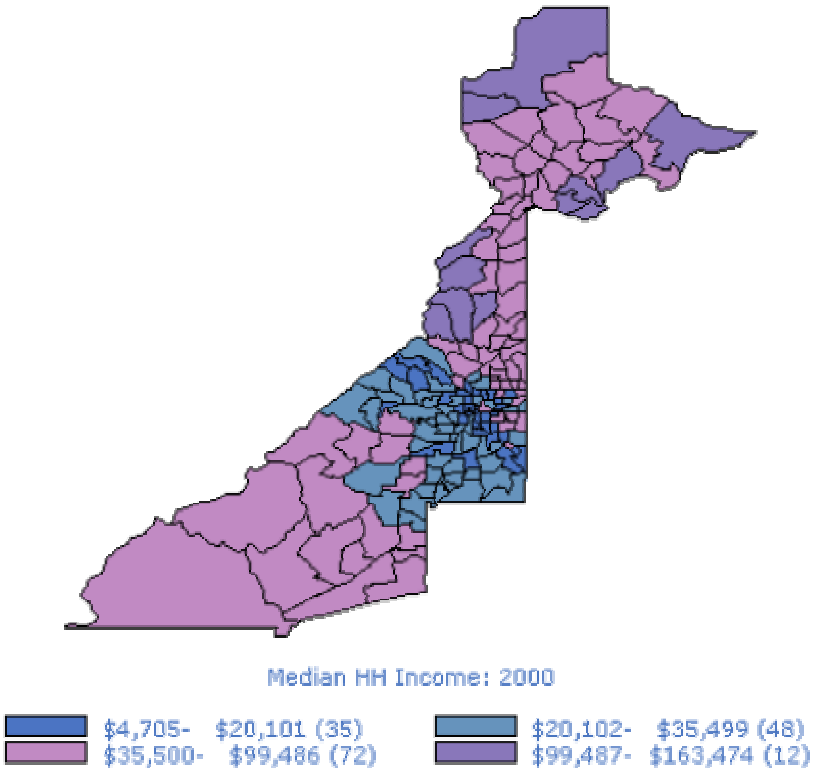
```
Data income;set census.income&stcnt;
 html = 'href="' || "&_URL" || "?_service=&_service" ||'&_debug='||"&debug"|| '&_program=' ||
"&_pgmlib..us4_race.sas" || '&PLOTVAR='||"&plotvar"|| '&stcntract=' || left(trim(stcntract)) || '" alt="State:' || trim(statename) ||'0D'x ||trim(countyname) ||'0D'x || left(tract) ||'0D'x ||left(&showvar) ||'0D'x || '">';
```

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 centriods within certain distance from the centriod of the reference tract. The following data steps are used to identify the closest eleven census tracts:

Map 3: Household Income Map of Fulton County of State Georgia by census tracts

Fulton County of State of Georgia
Median HH Income: 2000




```
data center(rename=(xcenter=rx ycenter=ry));set tract.trcenter;
if &stcntract=stcntract;
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_<=11;
run
```

Data set tract.trcenter contains the centriods of all the tracts in the United States. Depending upon the location of the reference tract, resulting data set center of eleven 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(stcntract/10000);
call symput('t'||left(_n_),left(stcnt));
```

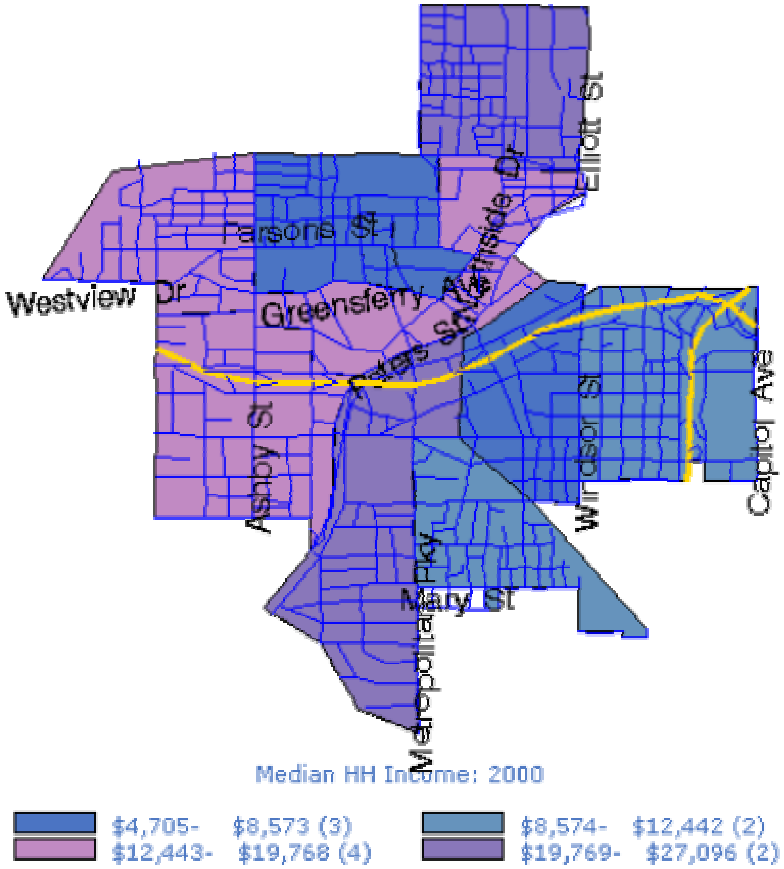
These macro variables are used to combine map and response data for all counties being represented by eleven census tracts and selected for those eleven census tracts as

```
Data apswork.map&stcntract;
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 stcntract;if j;
```

There should be eleven macro variables need to initialized for allowing the possibility of eleven counties being represented by the selected eleven tracts. Following combines the response data for maps:

Map 4: Household Income Map of Eleven tracts of Fulton County of State Georgia by census tracts

Around Tract 43 of Fulton County of State of Georgia
Median HH Income: 2000



Map covers 3.82 square miles area. Click here to remove street layer
Street layer is automatically displayed for map area less than 12 square miles unless removed

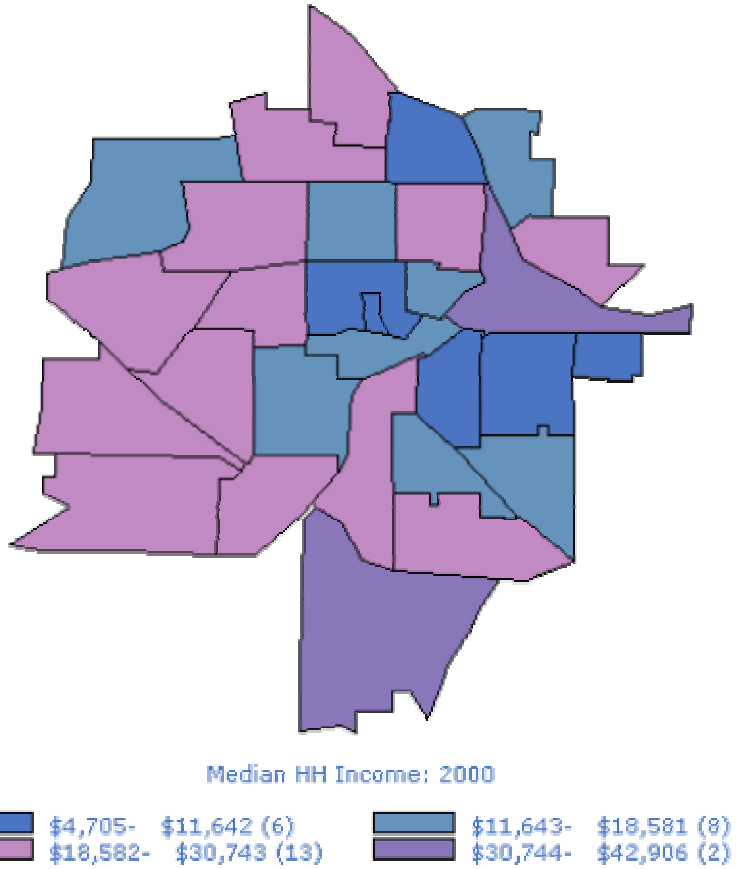
Data
apswork.race%sysevalf(&stcntract*100,integer);
merge center(in=j) race.race&t1 race.race&t2
race.race&t3 race.race&t4 race.race&t5
race.race&t6 race.race&t7 race.race&t8
race.race&t9 race.race&t10;by state county tract;if j;

Similar statements are used to create annotation of streets and also to calculate the area covered by those eleven tracts. These maps, response, annotation and area data sets 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. Map 4 is the resulting page a user will be able to display on their browser

Map 4 above shows the 2000 median household income by census tracts for eleven census tracts around reference tract of 43 of Fulton 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. Map 5 is a map of all the tracts within 2 miles radius from the centriod of census

Map 5: Household Income Map of Eleven tracts of Fulton County of State Georgia by census tracts

Around Tract 43 of Fulton County of State of Georgia
Median HH Income: 2000



Map covers 14.16 square miles area. Click here to display street layer
Street layer is automatically displayed for map area less than 12 square miles unless removed

Tract 43 of Fulton County in Georgia:

Cities and Places Map

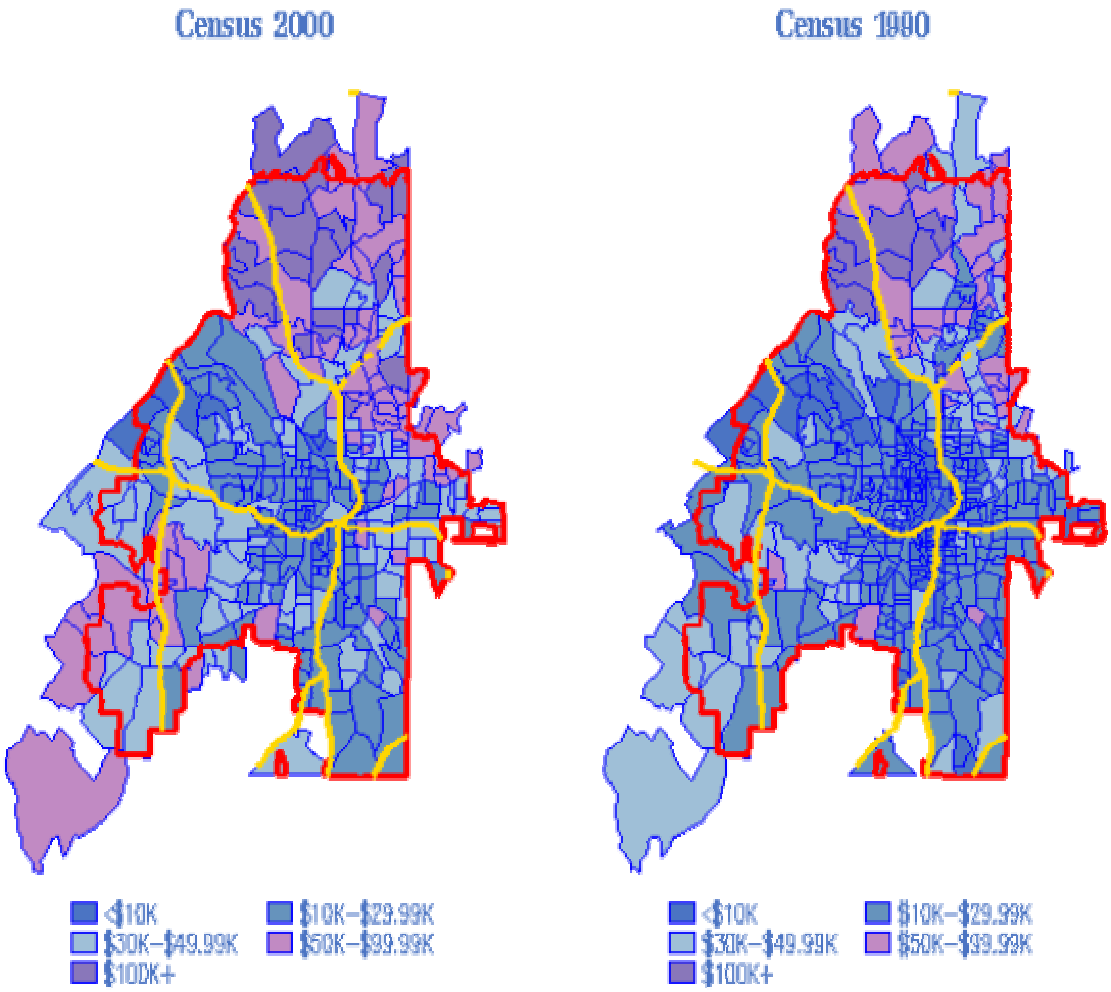
City and places are defined at block group levels where only part of the census tract is in a city or place area. Since Census Bureau does not provide a relation between 1990 and 2000 block groups, the best way to visualize the pattern of changes over the past decade is by looking at the 1990 and 2000 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;
goptions display HSIZE=0;
  goptions display;
  proc greplay  tc=tempcat nofs;
```

```
  igout city.gseg;
tdef newtwo des='two plots of equal size'
1/llx=0   lly=0
  ulx=0   uly=100
  urx=50  ury=100
  lrx=50  lry=0
2/llx=50  lly=0
  ulx=50  uly=100
  urx=100 ury=100
  lrx=100 lry=0
 ;
 template newtwo;
   treplay 1:maps00
           2:maps90;
 run;
quit;
```

Map 6: Household Income Map of City of Atlanta

Median HH Income



Map 6 is a map of City of Atlanta displaying 1990 and 2000 median household income.

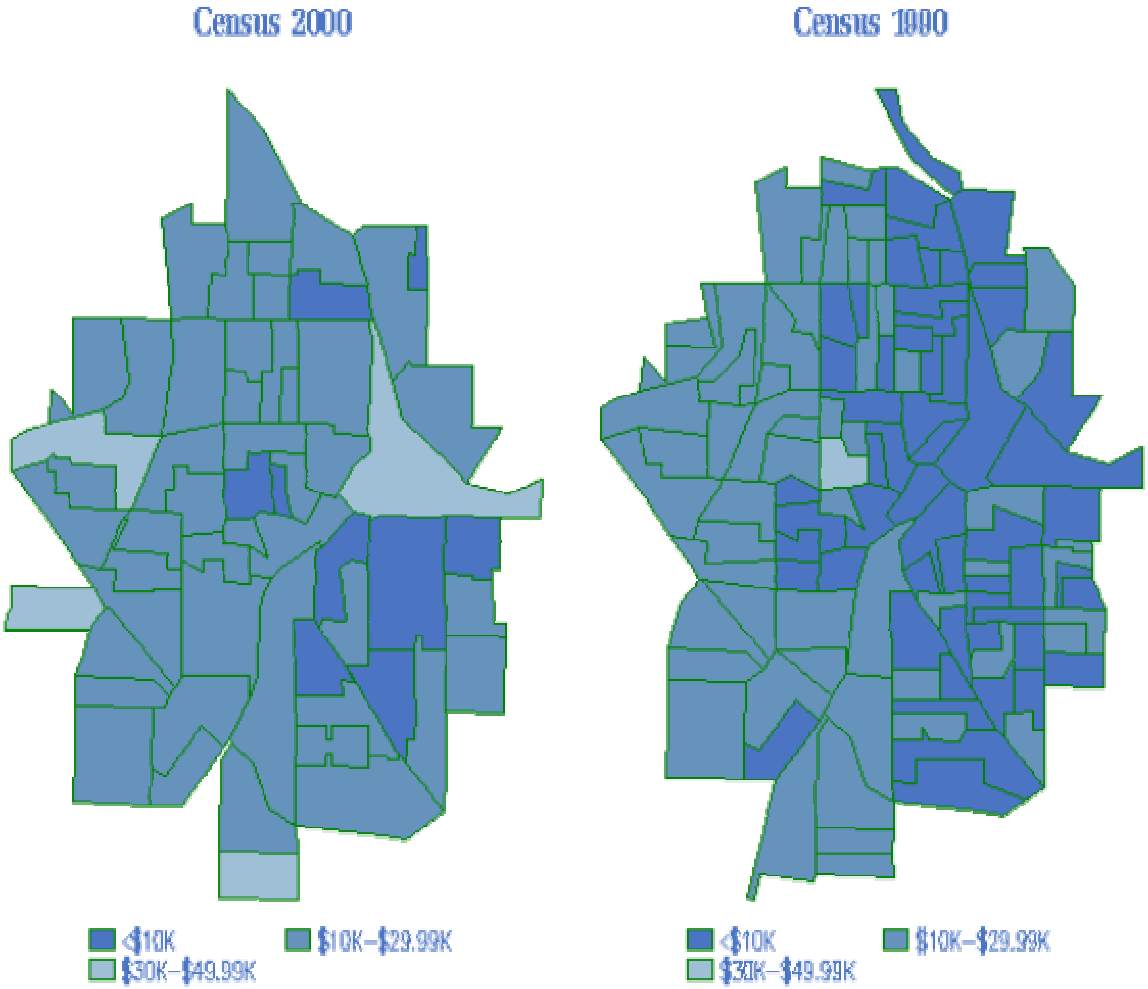
For the sake of comparison, PATTERN statements are matched to display same color for a population group in both maps. From the above maps one cannot tell which block group has changed by what exact number between 1990 and 2000 but a pattern of changes is quite apparent. It shows that over last decade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

Map 7: Household Income Map of user selected Neighborhood

State and County: Georgia Fulton Tract: 43
Median HH Income



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 shown in Map 7.

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 centriod 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.

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