

# CREATING POPULATION TREE CHARTS (USING SAS/GRAPH® SOFTWARE)

Robert E. Allison, Jr. and Dr. Moon W. Suh, College of Textiles, N. C. State University

## ABSTRACT

This paper describes a SAS® program that can be used to generate tree charts of US population data (or projections) to compare any year/sex/race combination on the left side of the tree against any year/state/sex/race combination on the right side.

The program utilized the following SAS procedures and functionality: PROC/GCHART HBAR, PROC/SQL, PROC/TRANPOSE, and Macro Variables.

## INTRODUCTION

Tree charts are often used to analyze Census population projections. This paper describes a SAS program which can be used to compare any two population groups. The program provides the versatility to allow users to easily select any state, year, race, and sex on the left side of the tree, and compare it to any state, year, race and sex on the right side of the tree chart.

The code used in this example was developed as part of the Textile and Apparel Business Information System (TABIS). It is based on the principles presented in Example 17 of the SAS/GRAPH Examples manual.

To fully utilize this sample program, users will need to have access to population data stored in a SAS dataset called "pop" using the variables listed below. (For those who do not have access to such data, a data step with sample data is included in Appendix A.)

|       |         |                                     |
|-------|---------|-------------------------------------|
| STATE | char    | postal abbr. (eg. 'NC', 'SC', etc.) |
| YEAR  | numeric | (eg. 1990)                          |
| RACE  | numeric | (eg. 0=all, 1=white, 2=black)       |
| SEX   | numeric | (eg. 0=both, 1=male, 2=female)      |
| AGE0  | numeric | (number of people age 0)            |
| AGE1  | numeric | (number of people age 1)            |
| ...   |         |                                     |
| AGE98 | numeric | (number of people age 98)           |
| AGE99 | numeric | (number of people age 99 and over)  |

## SAMPLE CODE

To facilitate easily selecting the year, state, sex and race for the two sides of the tree chart, the following macro variables are defined. The variables beginning with an "l\_" are used for the left side of the tree, and variables beginning with an "r\_" denote values for the right side.

```
%let l_year=1990;
%let r_year=2010;
%let l_state='US';
%let r_state='US';
%let l_sex=2;
%let r_sex=2;
%let l_race=0;
%let r_race=0;
```

The population trees in this example split the population into 20 age groups, each covering a 5 year range. The following SAS/SQL code groups the data in this manner, and stores it as G01 ... G20. The variable names are important since the PROC GCHART's HBAR procedure arranges the bars in alphabetical order based on the variable name. G00 is a "dummy" variable used to add extra space at the top of the tree chart.

```
proc sql; create table pop2 as
select unique year, state, sex, race,
sum(AGE0, AGE1, AGE2, AGE3, AGE4 ) as g20,
sum(AGE5, AGE6, AGE7, AGE8, AGE9 ) as g19,
sum(AGE10, AGE11, AGE12, AGE13, AGE14) as g18,
sum(AGE15, AGE16, AGE17, AGE18, AGE19) as g17,
sum(AGE20, AGE21, AGE22, AGE23, AGE24) as g16,
sum(AGE25, AGE26, AGE27, AGE28, AGE29) as g15,
sum(AGE30, AGE31, AGE32, AGE33, AGE34) as g14,
sum(AGE35, AGE36, AGE37, AGE38, AGE39) as g13,
sum(AGE40, AGE41, AGE42, AGE43, AGE44) as g12,
sum(AGE45, AGE46, AGE47, AGE48, AGE49) as g11,
sum(AGE50, AGE51, AGE52, AGE53, AGE54) as g10,
sum(AGE55, AGE56, AGE57, AGE58, AGE59) as g09,
sum(AGE60, AGE61, AGE62, AGE63, AGE64) as g08,
sum(AGE65, AGE66, AGE67, AGE68, AGE69) as g07,
sum(AGE70, AGE71, AGE72, AGE73, AGE74) as g06,
sum(AGE75, AGE76, AGE77, AGE78, AGE79) as g05,
sum(AGE80, AGE81, AGE82, AGE83, AGE84) as g04,
sum(AGE85, AGE86, AGE87, AGE88, AGE89) as g03,
sum(AGE90, AGE91, AGE92, AGE93, AGE94) as g02,
sum(AGE95, AGE96, AGE97, AGE98, AGE99) as g01,
0 as g00
from pop order by year, state, sex, race;
quit; run;
```

The dataset must next be transposed so that the G00 - G20 variable names become values. The following code performs this task, and then converts the population values into millions.

```
proc transpose data=pop2 out=pop2;
  by year state sex race;
run;

proc datasets; modify pop2;
  rename _name_=age_grp;
  rename COL1=pop;
run;

data pop2; set pop2;
  pop=pop/1000000;
run;
```

The macro variables are used in the "WHERE" clauses of the following SAS/SQL queries to select the desired data, and place it into the "left" and "right" datasets. These two datasets are then joined into a dataset called "both". Notice that the values in the left dataset are negative #'s - this forces GCHART to add a zero reference line, and print the values to the left of this line.

```
proc sql;

create table left as
  select unique age_grp, year, state, sex, race,
    (-1*pop) as pop, 'left' as group
  from pop2
  where (year = &l_year) and (state = &l_state)
  and (sex = &l_sex) and (race = &l_race);

create table right as
  select unique age_grp, year, state, sex, race,
    pop, 'right' as group
  from pop2
  where (year = &r_year) and (state = &r_state)
  and (sex = &r_sex) and (race = &r_race);

quit; run;

data both;
  set left right;
run;
```

If you do not have access to Census data, you can use the code in Appendix A instead of the previous code to create a sample "both" dataset.

The following code prepares several of the visual aspects of the chart to make it more readable. In axis1, the "posval" format forces the labels for the negative values to print as positive values. In axis2, a label for each age group is specified. The title2 and title3 statements help document which values were selected for the left and right sides of the tree chart.

```
goptions reset=global gunit=pct rotate=landscape
h1title=4 htext=2.5 ftitle=zapfb ftext=zapf cback=white
c1text=black colors=(black);
```

```
proc format; picture posval low-high='000,009'; run;

axis2 label=('In Millions') major=(number=7);

axis1 label=none value=( 'Age ' '95+' '90-94' '85-89'
'80-84' '75-79' '70-74' '65-69' '60-64' '55-59'
'50-54' '45-49' '40-44' '35-39' '30-34' '25-29'
'20-24' '15-19' '10-14' '5-9' '0-4 ');

title 'Population Distribution';
title2 "Left : state=&l_state race=&l_race sex=&l_sex
year=&l_year";
title3 "Right: state=&r_state race=&r_race sex=&r_sex
year=&r_year";
```

The remaining code creates the tree chart:

```
proc gchart data=both;
  format pop posval.;
  hbar age_grp / discrete space=0 subgroup=group type=sum
  sumvar=pop nostats nolegend frame autoref maxis=axis1
  raxis=axis2
  cframe=white coutline=black caxis=black;
run;
```

## EXAMPLES

In tree charts, it is customary to have all the values on the left and right be the same, except for the one variable that is being compared. The variable most often compared is sex (male versus female). The following examples demonstrate that this program can be used to compare many combinations of variables:

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

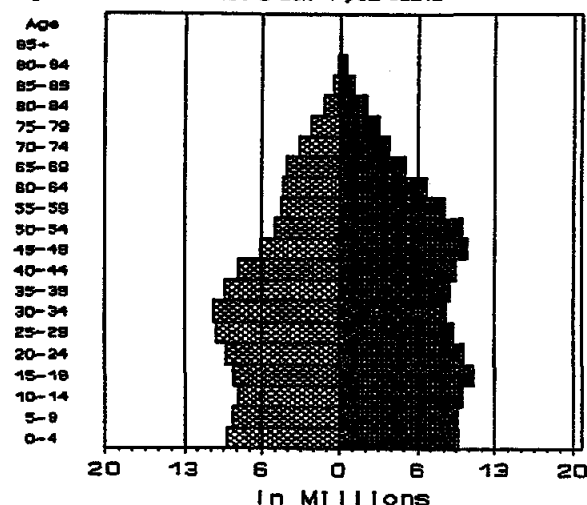
# APPENDIX A. Sample Data

data both;  
input age\_grp \$ year state \$ sex race pop group \$;  
cards;

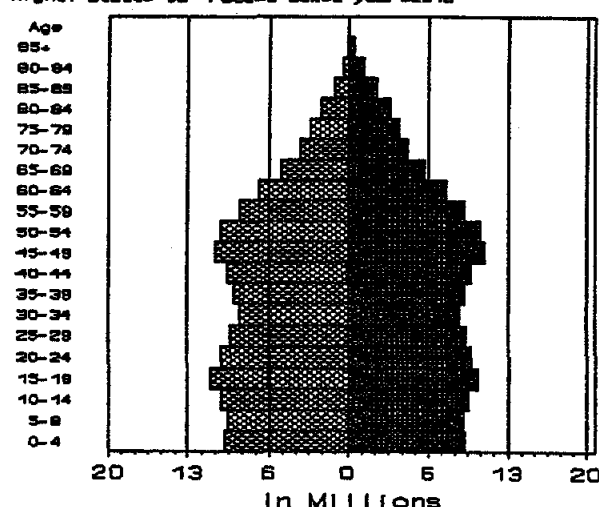
|     |      |    |   |   |          |      |
|-----|------|----|---|---|----------|------|
| G00 | 2010 | US | 1 | 0 | 0.0000   | left |
| G01 | 2010 | US | 1 | 0 | -0.1315  | left |
| G02 | 2010 | US | 1 | 0 | -0.4716  | left |
| G03 | 2010 | US | 1 | 0 | -1.2243  | left |
| G04 | 2010 | US | 1 | 0 | -2.3194  | left |
| G05 | 2010 | US | 1 | 0 | -3.2551  | left |
| G06 | 2010 | US | 1 | 0 | -4.1466  | left |
| G07 | 2010 | US | 1 | 0 | -5.5966  | left |
| G08 | 2010 | US | 1 | 0 | -7.5048  | left |
| G09 | 2010 | US | 1 | 0 | -9.0983  | left |
| G10 | 2010 | US | 1 | 0 | -10.5797 | left |
| G11 | 2010 | US | 1 | 0 | -11.0508 | left |
| G12 | 2010 | US | 1 | 0 | -10.0828 | left |
| G13 | 2010 | US | 1 | 0 | -9.5373  | left |
| G14 | 2010 | US | 1 | 0 | -9.1250  | left |
| G15 | 2010 | US | 1 | 0 | -9.8663  | left |
| G16 | 2010 | US | 1 | 0 | -10.6530 | left |

|     |      |    |   |   |          |       |
|-----|------|----|---|---|----------|-------|
| G17 | 2010 | US | 1 | 0 | -11.5039 | left  |
| G18 | 2010 | US | 1 | 0 | -10.6329 | left  |
| G19 | 2010 | US | 1 | 0 | -10.1170 | left  |
| G20 | 2010 | US | 1 | 0 | -10.2617 | left  |
| G00 | 2010 | US | 2 | 0 | 0.0000   | right |
| G01 | 2010 | US | 2 | 0 | 0.4669   | right |
| G02 | 2010 | US | 2 | 0 | 1.1956   | right |
| G03 | 2010 | US | 2 | 0 | 2.3095   | right |
| G04 | 2010 | US | 2 | 0 | 3.4597   | right |
| G05 | 2010 | US | 2 | 0 | 4.1232   | right |
| G06 | 2010 | US | 2 | 0 | 4.8876   | right |
| G07 | 2010 | US | 2 | 0 | 6.3471   | right |
| G08 | 2010 | US | 2 | 0 | 8.2018   | right |
| G09 | 2010 | US | 2 | 0 | 9.7470   | right |
| G10 | 2010 | US | 2 | 0 | 11.0885  | right |
| G11 | 2010 | US | 2 | 0 | 11.3802  | right |
| G12 | 2010 | US | 2 | 0 | 10.3132  | right |
| G13 | 2010 | US | 2 | 0 | 9.7254   | right |
| G14 | 2010 | US | 2 | 0 | 9.2967   | right |
| G15 | 2010 | US | 2 | 0 | 9.8910   | right |
| G16 | 2010 | US | 2 | 0 | 10.3227  | right |
| G17 | 2010 | US | 2 | 0 | 10.8945  | right |
| G18 | 2010 | US | 2 | 0 | 10.0913  | right |
| G19 | 2010 | US | 2 | 0 | 9.6054   | right |
| G20 | 2010 | US | 2 | 0 | 9.7556   | right |

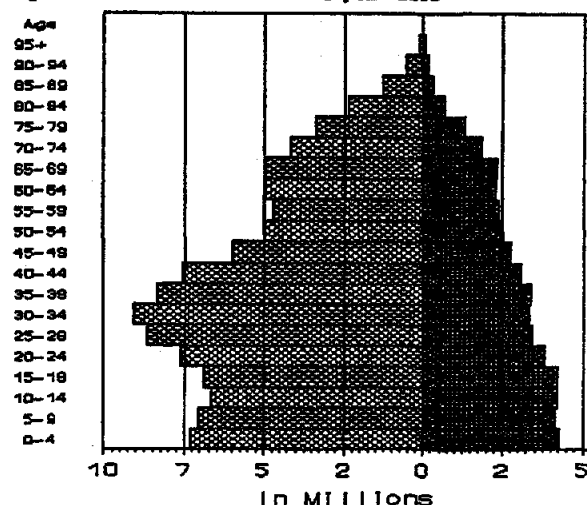
Population Distribution  
left: state=US race=0 sex=1 year=2010



Population Distribution  
left: state=US race=0 sex=2 year=2010



Population Distribution  
left: state=US race=2 sex=0 year=2010



Population Distribution  
left: state=US race=2 sex=1 year=2010

