Making Sense Out of Census Data
E. Taylor Murray, Orkand Corp., Atlanta, GA

This program reads US Bureau of Census data from an ASCII file of 3,000+ state/county records, each with a record length of 14,450 characters. Each record contains identifying information plus 1600 9-digit fields which give the population count data for the 1600 possible age/race/sex/hispanic-origin groupings for one state and county. The task was to output a SAS data set with one record per population figure, adding variables with categorical values for the demographic classifications involved. In addition, the age groupings were to be collapsed from 100 age categories into 17 5-year groups (0-4, 5-9, 10-14, etc.) plus one final group for age 85 and up.

Since the population figures for the different ages for a specific race/sex/ethnic grouping were not adjacent on the input record, the challenge was to compute a running column number in such a way as to scan across the record summing and outputting the population figures as they were encountered rather than, for example, reading the whole file into a (very) large array and outputting from there.

The resulting code is quite compact, less than 100 lines including comments, and contains some interesting usages, including sum statements of the form "X+some arithmetic computation, including a True/False evaluation" rather than just "x+1" or "x+n", negative do-loop increments, use of a variable column specification in an INPUT statement, and shows the use of PUT statements to debug. It produces a SAS® data set that can easily be merged with other data sets containing information ordered by similar demographic variables.

**************
PROGRAM: CENSUS90. Reads ASCII file of 1990 Census data, outputs SAS file with one record per population figure, accompanied by variables for the state/county and demographic variables.
AUTHOR: Taylor Murray.
DO AGE5 = 1 TO 5; DROP AGE5; *Sum 5 ages into 1 grp.;

INPUT @ (COL) POP 9. @; DROP POP;
POPAGE + POP;
if test and age in(1, 2, 13, 17) then put ' at age5 '
age = age5 = col = popage = 0;
COL + 144;
END; *(AGE1-5);

OUTPUT;
END; *(AGE1-17);

AGE = 18; *Our last age grp is ages 85-100;

POPAGE = 0;
DO AGE85UP = 1 TO 16; DROP AGE85UP;
if test then put ' at age85up ' age85up = col = popage = 0;
INPUT @ (COL) POP 9. @;
COL + 144;
POPAGE + POP;
END;
OUTPUT;

END; *(RACE);
END; *(SEX);
END; *(HISP_YN);
RUN;
PROC PRINT DATA = OU.CENSUS90(obs = 200);
RUN;
PROC FREQ DATA = OU.CENSUS90;
WEIGHT POPAGE;
RUN;

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. * indicates USA registration.

E. Taylor Murray
1493 Bates Court
Atlanta, GA 30319
404-638-2720 (h)
770-488-5594 (w)
E-mail: ETMO@CCDOSA1.EM.CDC.GOV