

# Census 2000 Migration PUMS Read Me

## Purpose of files

In- and out-migration flow data and the characteristics of these migrants for states and sub-state areas can only be derived from the individual state 2000 Census Public Use Microdata Samples (PUMS). While in-migrants to any given state can be analyzed using that state's file, analysis of out-migration requires analyzing data from all 50 states, the District of Columbia, and Puerto Rico. The New York State Data Center has created this national migrant file from the individual state and area files in order to facilitate migration analysis of the 2000 Census.

The records included in this file contain the long-form responses for people who lived within a specific geographic area in either 1995 or 2000. Using these responses, data users can develop their own tabulations of the characteristics of these movers and compare the characteristics of people moving into an area with those who left the area. Similar tabulations of the individuals who did not move can be generated directly from the Census Bureau's PUMS files.

## Data Source & File Development

The data records in these files ultimately come from the Census 2000 5% PUMS files for each of the states, the District of Columbia, and Puerto Rico. These files contain samples of the edited individual responses to the Census 2000 long-form designed to protect the respondents' confidentiality.

Using SPSS, the New York State Data Center staff combined these individual state samples into a single national sample. Then the records for individuals living in a different house in 2000 than they did in 1995 were extracted from this file.

Next, a variable, MSTPUMA0, was added that converted the state and 5-percent PUMA of residence in 2000 variables, STATE and PUMA5, to the same coding scheme as the Census Bureau used for the 1995 PUMA of residence. Details on reasons for this variable are included in the "migration / residence coding issues" section below.

Finally, the records of movers residing in a particular state in either 1995 or 2000 were extracted and saved in the desired format.

Because this dataset contains more variables than are allowed in database tables, the dataset is divided into two DBF files. In order to facilitate linking the records in these two DBF files, an additional unique record identification variable, IDNUM, was added to the dataset.

## File Structure and Data Content

### **Files**

This CD contains the following files:

1. Either a single SPSS or SAS data set or two DBF files with the records of people residing in the selected state during either 1995 or 2000. The file names are:
  - MIGPUMSxx.SAV – SPSS datasets
  - MIGPUMSxx.SD7 – SAS versions 7 and later
  - MIGPUMSxx1.DBF and MIGPUMSxx2.DBF – dBase files“xx” is the Postal Abbreviation for the state included in the file.
2. PMLABS.SAS – The file contains value labels associated with each of the variables if the data is provided as a SAS data set.
3. PUMA1\_AREAS.XLS – Total area and land area for each 1-percent Super-PUMA.
4. PUMS.PDF – The Census Bureau’s technical documentation for PUMS. This file contains details on working with these microdata files and on each of the variables included in these files, except those noted below.
5. README.DOC – This file.

### **File contents**

Except as noted below, these files contain all of the variables included in the 5-percent PUMS files distributed by the U.S. Census Bureau. The variable names are those used by the Census Bureau in the technical documentation for the 5-percent PUMS files.

In order to save space, the following PUMS variables have been deleted from these files:

1. RECTYPE – record type (no longer meaningful since this is a flat file)
2. SAMPLE – sample designator (used to identify the records as coming from the 5-percent PUMS files)
3. LNDPUMA1 – 1-percent PUMA land area, in square meters (dropped due to an error when reading the state files, see PUMA1\_AREAS.XLS for the correct values.)

### **SPSS and SAS file structure**

Each record in the SPSS and SAS datasets contain both the household and person data from the Census Bureau’s 5-percent PUMS files in a single record for each person. The data variables are organized in the following general sequence:

1. Household data variables
2. Household data allocation flags
3. Person data variables
4. Person data allocation flags
5. MSTPUMA5 – State and migration PUMA of residence in 1995 (computed)
6. MSTPUMA0 – State and migration PUMA of residence in 2000 (computed)
7. IDNUM – A unique record identification number

### *DBF file structure*

Since there are more variables than can be included in a DBF file, the data is divided into two files as described below:

File 1:

1. Household data variables
2. Person data variables
3. MSTPUMA5 – State and migration PUMA of residence in 1995 (computed)
4. MSTPUMA0 – State and migration PUMA of residence in 2000 (computed)
5. IDNUM – A unique record identification number

File 2:

1. Household data allocation flags
2. Person data allocation flags
3. IDNUM – A unique record identification number

The IDNUM variable can be used to link the records in the two files.

## **Migration / Residence Coding Issues**

The Census Bureau used slightly different coding schemes for the 2000 and 1995 state and PUMA codes.

The Census Bureau coded the 1995 residence PUMA by grouping multiple PUMAs within a county and PUMAs crossing county lines into migration PUMAs containing one or more complete counties. For example, the 10 PUMAs in New York County (Manhattan), NY were combined into a single migration PUMA. Also, the PUMA containing Putnam County, NY and one census tract in Westchester County, NY was combined with the 6 PUMAs entirely within Westchester County to create a single migration PUMA combining Putnam and Westchester Counties.

There is also a difference in the coding of the 2000 and 1995 state of residence variables in the PUMS files. The 2000 state of residence variable, STATE, is a 2-character variable. The 1995 state of residence (which includes foreign countries), MIGST5, is a 3-character variable with a leading zero added to state code. For example, New York State is coded as “36” in STATE, but as “036” in MIGST5.

The MSTPUMA0 (for 2000) and MSTPUMA5 (for 1995) variables were created to identify the residences in the two years with a common coding scheme. Both variables use a 3-character state code followed by the 5-character migration PUMA code for the residence in the respective year.

## **Working with these files**

Except as noted below, these files contain all of the variables included in the original 5% PUMS files produced by the U.S. Census Bureau. Details on these variables and working with these

files can be found in the Census Bureau's technical documentation, PUMS.PDF, included on the CD-ROM. Please note that the documentation includes variable lists for both the 1% and 5% PUMS files – only the 5% PUMS list is appropriate for this file.

When producing tabulations from this file, it is important to weight the observations correctly to get figures that are consistent with those that would be reported in the Census Bureau's Summary File products. There are two weight variables included on each record – HWEIGHT, for tabulations of housing units, households, and families; and PWEIGHT, for tabulations of persons.

Software intended to tabulate large datasets should be used to analyze these files. Examples of this type of software are SAS and SPSS. These files are generally too large to be imported into spreadsheets.

### **Known Data Problems**

There is a significant difference (greater than the 90% confidence interval) in the number of people moving into a given state from other states as calculated from the 5-percent PUMS files and reported on Summary File 3 for 24 states. The Census Bureau indicated that since the 1995 residence was not used in the weighting process, these differences are not unreasonable.

### **Contact Information**

For additional information about these files, contact:

Leonard M. Gaines, Ph.D.  
Empire State Development  
30 South Pearl St.  
Albany, NY 12245  
(518) 292-5300  
[lgaines@empire.state.ny.us](mailto:lgaines@empire.state.ny.us)