

Appendix B. Using Demographic Data from the U.S. Census

This appendix provides additional information on using demographic information from the U.S. Census for determining EJ populations.

Blending and Merging Geographic Spaces

For the 2000 Census, the Bureau of Transportation Statistics and the Bureau of the Census will produce a Census Transportation Planning Package. This package will include occupational data, residential data, travel-to-work data, and related demographic information for each Traffic Analysis Zone (TAZ) defined for a Metropolitan Planning Organization's TAZ base map. For participating organizations, the TAZ has already been defined on the TIGER (Topologically Integrated Geographic Encoding and Referencing system) data base line map file in the 1998, 1999, and 2000-Redistricting editions. The data will be released in 2002.

Until 2002, however, it is necessary to use 1990 *block group* data from the STF3 to define EJ neighborhoods, and to use TAZ geography to report the results of the transportation demand model runs. Merging the data from these two different geographies can be done in several ways, depending on the purpose at hand. Two cases are important here:

Case 1: It must be determined whether a unit from one geographic scheme is within a unit from another geographic scheme. For example, is a TAZ within a particular block group? If so, then it shares the characteristics of the block group, e.g., it may be an EJ

community, or a low-income neighborhood, or the like. The geographic intersect function in most GIS software will make this identification. In ArcView, the *intersect* command of the GeoProcessing option will pick out the TAZ that is within a given block group.

Case 2: It must be determined how much a unit from one geographic scheme contributes to a unit in another geographic scheme. For example, if 10,000 vehicle miles are traveled in a given TAZ, and that TAZ overlaps to a known extent with a block group, how many miles should we allocate to the block group? The geographic *merge* routines in GIS packages do not always do this assignment efficiently or effectively. The standard statistical package, however, allows us to calculate the VMT (Vehicle Miles Traveled) we would assign to each block in the TAZ, then reassemble these blocks—and their associated VMT—into the block groups in which some other value has been tabulated.

To generate workable data for evaluating EJ outcomes, it is necessary to use both of these routines. The allocation techniques are necessary not only for data from different sources, but will also be necessary to blend data from 1990 the 2000 Census information.¹

¹In particular, the PL94-171 Census Redistricting file provides ethnic data on areas that were not mapped or tabulated in the 1990 Census—mostly new housing developments. In order to include data from these areas, the geographies must be merged or allocated using the methods suggested in this subsection.

Inter-Census and Off-Census Data Problems

The most thorough data collection effort in the United States that contributes to effective transportation planning is likely the decennial United States Census. The Census has numerous major parts. Two of the most important are the 100% population count (which most people think of when they think about a census), and the *long form* samples, administered to 1% , 5% or 15% of the population. The *long form* provides very valuable data for understanding the transportation behaviors of U.S. residents, particularly with respect to their journey to work.

The *long form* asks for details about the individual's work life, for example, whether he or she:

- Is in the labor force;
- Works outside the home;
- Travels to work by transit; car or other mode; leaves early or leaves late; or
- Travels a long time or a short distance, etc.

In the 2000 Census, detailed information on the location of the respondent's employer was collected, making possible a determination of starting and ending points of trips to work.

This form, often referred to colloquially as the "5% sample," also provides detail on the educational history, detailed occupation, industry of employment, and work schedule or unemployment pattern. To all of this occupational data, wages, other income, house value, availability of a car, and detailed household composition are added as a broad survey of American demographic characteristics.

Census Confidentiality Limitations on Disclosure of Income Data

The detail on the Census *long form*, however, creates substantial conflict with the Census Bureau's confidentiality policy. In principle, it is against the law for the Census to disclose individual information on any respondent. In historical use, the most detailed data has been tabulated only on large geographic units—states, cities, metropolitan areas, and the like. With the development of consistent Census Tracts, of course publication of data at the tract level became practical. With the development of inexpensive computer tape storage, it became practical to tabulate many-dimensional tables for a wide range of geographic units, ranging from the state down to the smallest, the block—storing the entire tabulation on a large tape file.

These summary files came to be known as the PL94-171 (Congressional redistricting file), the STF1² (Summary Tape File 1) containing counts by ethnic group and age down to the block level; STF2, containing intermediate units, such as Census Designated Places; and STF3, which tabulates only down to the *block group*, a basic Census field administrative unit composed of a few blocks. STF3, however, contains the first small area reports on income and its relationship to ethnicity and other social markers.

The most desired file from the standpoint of Census scholars is the PUMS (Public Use MicroSample) file, which provides a

²For the 2000 Census, the STF series has been renamed the SF series, reflecting the fact that the tables are now released on CD-ROM, rather than magnetic tape.

representation of the 5% sample for each state. The PUMS has been cleaned of personal identifiers. In addition, detailed personal data that might allow the identification of a given person because of his or her unusual occupation, income, or family situation, has been adjusted to reflect somewhat less exotic characteristics. The effect of these adjustments is that it would be impossible to find Bill Gates in the Washington State PUMS, even if he filled out the 5% Census sample instrument—because some of his distinguishing income and asset features would have been “topcoded,” or replaced, by other values that captured the spirit, but not the detail of high income or wealth.

An additional adjustment made in the PUMS is augmentation of missing data by use of a “hot deck,” or a set of sample cases that represent the diversity of the population and have no missing data. The hot deck values are used to replace those of a similar sort of person who did not fill out some specific item on the Census survey form. Additional allocation procedures are used to fill in material that can be reasonably deduced from the other answers the individual provided. Some of the income, housing value, and other household items are augmented by allocation procedures for relatively small fractions of the sample respondents. All allocated fields are identified on an individual basis, so that a user could choose to ignore all allocated values if the study at hand demanded it.

The PUMS is most affected by the Census privacy policies. In general, the Census will not create a PUMA (Public Use Microsample Area) smaller than 100,000 people. This means that in Hawaii, Oahu has about seven PUMAs, the County of Hawaii one PUMA, and Maui and Kauai Counties have been compressed into a single PUMA. In addition, since so much detail is available on sample persons, the place of work designation in the PUMS has been truncated to indicate in which county a Hawaii resident

worked.³

Time Schedule for Typical Census Data Production and Distribution

The Census Bureau publishes a time table for release of Census tabulation and sample data. The first file to be released is always the PL94-171 Redistricting file. This is typically available in the Spring of the year following the Census. (For the 2000 Census, the file began to appear in prototype and complete forms beginning March 7, 2001.) The SF1 (the former STF1) is targeted for release beginning the second week of June, 2001, and the SF3 is planned for Summer, 2002.⁴ The 1% and 5% PUMS samples are expected in 2002 and 2003, respectively.⁵

³Note that the original survey instrument actually collected residence address and location of employer. In principle, the Census Bureau can construct a point-to-point matrix of journey to work movements. The location values are carefully suppressed in the public use files. An alternative source of this work trip data is identified.

⁴This means that for all practical purposes, no income data on small areas will be available before Summer, 2002.

⁵The official Census Release Calendar is found at <http://www.census.gov/population/www/censusdata/c2kproducts.html>

Special Census Programs for Transportation Planning

The Census Transportation Planning Package (CTPP) provides the data for transportation planning that would otherwise be suppressed in the PUMS or other summary tabulations. According to FHWA:⁶

CTPP is a set of special tabulations from the decennial census designed for transportation planners⁷. CTPP contains tabulations by place of residence, place of work, and for trips between home and work.

Key Variables:

Household Size
Household Income
Vehicles per Household

Age and Gender of Workers
Occupation of Workers
Worker Earnings

Usual Mode to Work
Commuting Time
Work Trip Departure Time

Work Location
Time of Arrival at Work

What makes CTPP unique?

Initiated with the 1970 Census, CTPP is the only Census product that summarizes data by place of work and provides information on the trips between home and work.

It is the only source of information provided by the Census Bureau with summary tabulations for TAZs and other small geographic areas.

The current status of the Census 2000 CTPP is given by periodic newsletter updates; according to the FHWA, the most challenging work to date has been interpreting and geocoding places of work so that geographic point-to-point matrices can be constructed.⁸ The package is scheduled for release in 2002.

Conversion from Census Block Groups to Traffic Analysis Zones (TAZ)

The Census block provides a good basis for tallying population counts and demographic data published at the block level. It can pose some complications for constructing traffic analysis zones, however, because the Census enumeration geography may not be

⁶<http://www.fhwa.dot.gov/////////ctpp/about.htm>.

⁷The development process for the 2000 CTPP is documented in the Conference Proceedings, Decennial Census Data for Transportation Planning, Irvine, CA, March 13-16, 1994. The report is available from the Transportation Research Board.

⁸See the current newsletter/status report at:
<http://www.fhwa.dot.gov/////////ctpp/ctppstat.pdf>

designed to recognize automobile or transit system travel impediments, such as streams or narrow bridges. The best TAZ structure for transportation planning will likely involve a few boundary adjustments or other modifications in order to separate areas which might not in fact connect on the road system.

Additional modifications may be caused by distinctly different daytime and nighttime populations. To estimate origin effects, the overnight residential population is important. To estimate job attraction effects, the daytime employment is more important. Split blocks may require equivalency or adjustment tables to take these day/night effects into account. The Department of Planning and Permitting of the City and County of Honolulu anticipates constructing TAZ-to-block equivalency tables to address the block splits or block edits that may be necessary to conform the defined TAZ units to actual transportation and traffic patterns.