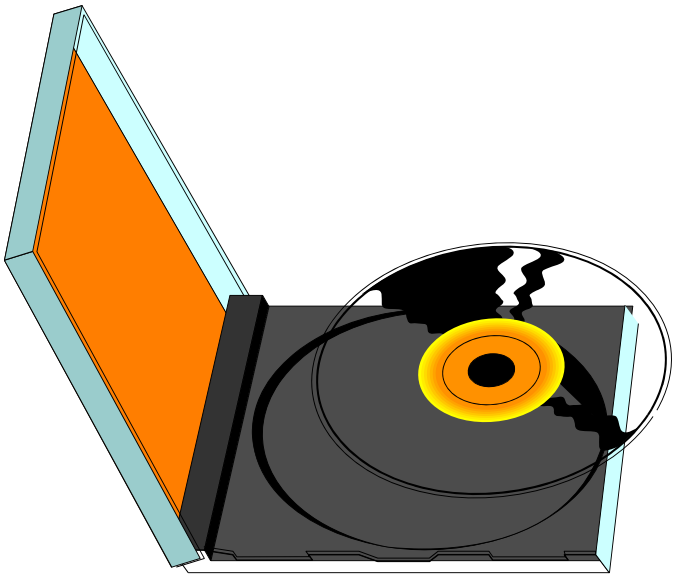
TRACT LEVEL PLANNING DATABASE WITH CENSUS 2000 DATA



Antonio Bruce and J. Gregory Robinson

U.S. Department of Commerce
U.S. Census Bureau
Washington, DC 20233

Table of Contents

TRACT LEVEL PLANNING DATABASE WITH CENSUS 2000 DATA:
STATEMENT OF PURPOSE 1

LIMITATIONS OF THE DATA 1

TRACT LEVEL PLANNING DATABASE WITH CENSUS 2000 DATA:
DESCRIPTION OF VARIABLES IN THE DATABASE 2

Geographic Variables 2
Person/Demographic, Housing, and Operational Variables 3

TRACT LEVEL PLANNING DATABASE WITH CENSUS 2000 DATA:
DERIVED VARIABLES 7

Hard-To-Count Scores 7
Table 1: Percentile Distribution of Selected Hard-to-Count (HTC) Variables for Tracts:
Census 2000 8
Table 2: Comparison of 1990 and Census 2000 Mail Return Rates
by Hard-to-Count Scores 9
Figure 1: Comparison of Trends in 1990 and Census 2000 Mail Return Rates
by Hard-to-Count Scores 9
Table 3: Total Number of Tracts, Total Population, and Total Housing
Units in Census 2000 10

TRACT LEVEL PLANNING DATABASE WITH CENSUS 2000 DATA:
FILE STRUCTURE OF THE DATABASE 11

APPENDIX A. DEFINITION OF CENSUS 2000 DATA UNIVERSE AND ACCURACY
OF THE DATA 12

APPENDIX B. DEFINITION OF CENSUS 2000 MAIL RETURN RATES 13
Census 200 Mail Return Rates Methodology 13
Census 2000 Type of Enumeration Areas and Type of Forms 14

REFERENCES 15

TRACT LEVEL PLANNING DATABASE WITH CENSUS 2000 DATA: STATEMENT OF PURPOSE

The Tract Level Planning Database With Census 2000 Data is a database that assembles a range of housing, demographic, and socioeconomic variables that are correlated with mail nonresponse. Using data from Census 2000, a database containing these variables has been developed for all census tracts in the country.

The variables included in the Tract Level Planning Database With Census 2000 Data (also called the planning database, or PDB) were guided by extensive research conducted by the U.S. Census Bureau and others to measure census coverage and to identify reasons people are missed in the census (de la Puente, 1993). The variables include housing indicators (percent renters, multi-units, crowded housing, lack of telephones, vacancy) and person indicators (poverty, not high school graduate, unemployed, complex households, mobility, language isolation). Other operational and demographic data are also included (such as race/ethnic distributions). Using the 1990 Census as the initial source, a database containing these variables was developed for all tracts in the country for use in the planning, implementation, and evaluation of Census 2000 (U.S. Bureau of the Census, 1999). The PDB contains "hard-to-count" (HTC) scores which summarize the attributes of each tract or block group in terms of enumeration difficulty.

The Tract Level Planning Database With Census 2000 Data is one of many resource tools to aid the planning activities leading up to the 2010 Census. Specific activities which could be supported by the planning database include:

1. Identifying hard-to-count areas (areas with concentrations of attributes that make enumeration difficult);
2. Identifying areas with potentially low mail response rates;
3. Identifying areas where special attention may be needed for:
    a. Questionnaire Assistance Centers
    b. Distribution of Be Counted Forms in languages other than English;
4 Identifying areas where special outreach and promotion efforts could be considered;
5. Planning recruitment activities by Regional Census Centers and Local Census Offices; and
6. Other uses as identified.

LIMITATIONS OF THE DATA

The Tract Level Planning Database With Census 2000 Data is based on Census 2000 data and Census 2000 geographical boundaries. The data collected with the "long form" are subject to sampling error. More up-to-date sources of information may be available through local knowledge of neighborhoods and communities or local administrative records.

TRACT LEVEL PLANNING DATABASE WITH CENSUS 2000 DATA: DESCRIPTION OF VARIABLES IN THE DATABASE

This section contains descriptions and definitions of all variables in the Tract Level Planning Database With Census 2000 Data. More detailed descriptions are provided in Appendix B of the Census 2000 Population and Housing Tract Reports (SF1, SF2, SF3). All basic census data in the database are drawn from the same source files used to produce the Census 2000 Tract Reports.

The total number of tracts delineated for Census 2000 was 65,443. The tract level planning database contains 65,184 total records, one record per tract. Tracts with zero population in the source sample (long form) file of the census are excluded. The data refer to the total housing units and total population of the tract. The variables marked with an "X" are variables included in the calculation of hard-to-count (HTC) scores.

Sources of all variables are indicated: "100%" refers to census data items asked of all persons and all housing units; "Sample" refers to census data items asked of a sample of persons and housing units; "Census geography" refers to the basic geographic data associated with 100% and sample items. See Appendix A for more discussion of the derivation and accuracy of the data. The database layout and description of each variable for tracts are as follows:

Geographic Variables				
Variable Number	**Variable Name**	**Description**	**Source**	**Variable HTC Score**
1	**GIDTRACT**	State/County/Tract - An 11-digit code. The first two digits denote State, the next three digits denote County, and the last six digits denote Tract.	Census geography	
2	**STATE**	FIPS State Code - A 2-digit code.	Census geography	
3	**STUSAB**	State/U.S.-Abbreviation (USPS) - A 2-character postal abbreviation code	Census geography	
4	**COUNTY**	FIPS County Code - A 3-digit code	Census Geography	
5	**ST/CNTY NAME**	State Name and County Name - The primary political division of most states is termed "county". In Alaska, the county equivalents are organized in boroughs and "Census areas". In Louisiana, the county equivalents are organized in "Parishes".	Census Geography	
6	**TRACT**	Census Tract Code - A 6-digit code. Census tracts are delineated for all metropolitan areas and counties. Tracts usually have between 2,500 and 8,000 people, though some have very small populations. When first delineated, tracts are designed to be homogeneous with respect to population characteristics, economic statistics, and living conditions. The spatial size of tracts varies widely depending on the diversity of settlement.	Census geography	
7	**REGION**	Census Geographic Region - A 1-digit code	Census geography	
8	**DIVISION**	Census Geographic Division - A 1-digit code	Census geography	

Person/Demographic, Housing, and Operational Variables				
Variable Number	**Variable Name**	**Description**	**Source**	**Variable HTC Score**
9	**TOT POPULATION**	Total Population for Tract - Total people for 2000 based on a "100-percent" count of the population as of April 1. Includes people living in housing units and in group quarters.	100%	
10	**TOT HOUSING UNIT**	Total Housing Units for Tract - Total housing units for 2000 based on a "100 - percent" count as of April 1. A housing unit is a house, an apartment, a mobile home or trailer, a group of rooms or a single room occupied as separate living quarters or, if vacant, intended for occupancy as separate living quarters.	100%	
11	**HTC**	Hard-To-Count Score - See next section on "Tract Level Planning Database With Census 2000 Data: Derived Variables."	Derived	
12	**MAIL RETURN RATE**	The 2000 Mail Return Rate is calculated by dividing the number of mail returns of occupied units by the number of occupied units in the mailout universe. See Appendix B for more discussion of the definitions of unofficial mail return rates.	Operational	
13	**FLAG**	Flag for "nonrepresentative" Tract -These are tracts with <100 housing units, or with population size <250, or group quarters >50% or high vacancy rates >35%. Given the small size and/or unusual characteristics of these tracts, the HTC scores and other data need to be interpreted with caution. The assigned values are: "1"- represents all tracts that meet the group quarters or vacancy rate criteria; "2" - strictly represents tracts that were List/Enumerate (questionnaires were not mailed) and did not meet the group quarters or vacancy rate criteria; and "3" - strictly represents tracts where the population size < 250 or housing units < 100 and did not meet the group quarters or vacancy rate criteria. State and county names (variable 5) are omitted from tracts failing the size criteria.	Derived	
14	**PCT VACANT HU**	Percent Vacant Units - Number of vacant housing units divided by total housing units (times 100 to convert to percent). A housing unit is vacant if no one is living in it at the time of enumeration, unless its occupants are only temporarily absent.	100%	**X**
15	**PCT SINGLE U STRC**	Percent Single detached or attached Housing Units In Structure - Number of single units in structures containing detached or attached housing units divided by total housing units (times 100). A structure is a separate building that either has open spaces on all sides or is separated by dividing walls that extend from ground to roof.	Sample	
16	**PCT NOT SINGLE U STRC**	Percent of Housing Units that are not Single detached or attached units - Number of units that are not single detached or attached units divided by total housing units (times 100). Housing units defined by this variable include all multi-unit structures (see variables 17 and 18) and trailer/mobile homes (see variable 19).	Sample	**X**
17	**PCT MLT U 10 P STRC**	Percent 10 or More Housing Units In Structure - Number of units in structures containing ten or more housing units divided by total housing units (times 100). A structure is a separate building that either has open spaces on all sides or is separated by dividing walls that extend from ground to roof.	Sample	
18	**PCT MLT U 2 P STRC**	Percent 2 or More Housing Units In Structure - Number of units in structures containing two or more housing units divided by total housing units (times 100).	Sample	
19	**PCT MOBILE HOME**	Percent Trailer/Mobile Home - Number of housing units classified as trailers or mobile homes divided by total housing units (times 100).	Sample	

Person/Demographic, Housing, and Operational Variables				
Variable Number	**Variable Name**	**Description**	**Source**	**Variable HTC Score**
20	**PCT RENTER OCCP HU**	Percent Renter Occupied Unit - Number of renter occupied units divided by total occupied housing units (times 100). A unit is classified as occupied if it is the usual place of residence of the person or group of people living in it at the time of enumeration, or if the occupants are only temporarily absent. A housing unit is "owner-occupied" if the owner or co-owner lives in the unit, even if it is mortgaged or not fully paid for. All other occupied units are classified as "renter occupied", including units rented for cash rent and those occupied without payment of cash rent.	100%	**X**
21	**PCT CROWD OCCP U**	Percent Occupied Units with More Than 1.5 Persons Per Room - Persons per room is obtained by dividing the number of persons in each occupied housing unit by the number of rooms in the unit (times 100). The figures for variable 21 represent the percent of occupied housing units having 1.5 or more occupants than separate rooms; it represents an index of overcrowded housing.	Sample	**X**
22	**PCT NOT_HB WF HH**	Percent Households that are Not Husband/Wife Families - Number of households that are not in husband/wife families divided by total households (times 100). Households are classified by type according to the sex of the householder and the presence of relatives. Two types of householders are distinguished: a family householder and a nonfamily householder. A family household consists of a householder and one or more other persons living in the same household who are related to the householder by birth, marriage, or adoption. A husband/wife family is a family in which the householder and his or her spouse are enumerated as members of the same household. Types of households which are reflected by variable 22 include family households in which the spouse is not present and all nonfamily households (householders who live alone or with nonrelatives only).	100%	**X**
23	**PCT_OCCP U NO PH SRVC**	Percent Occupied Units with No Telephone Service - Number of units without a telephone inside the house, apartment or trailer/mobile home, divided by total occupied units (times 100). Units where the respondent uses a telephone located inside the building but not in the respondent's living quarters are classified as having no telephone.	Sample	**X**
24	**PCT NOT HS GRAD**	Percent Not High School Graduate (Ages 25+) - Number of people 25 years old and over who are not high school graduates (received diploma or its equivalent) divided by total population 25 years old and over (times 100).	Sample	**X**
25	**PCT PRS BLW POV LEV**	Percent People Below Poverty - Number of people classified as below the poverty level divided by the total population for which poverty status was determined (times 100). Families and people were classified as below poverty level if their total family income or unrelated individual income was less than the poverty threshold specified for the applicable family size, age of householder, and number of related children under age 18. Poverty status was determined for all people except institutionalized people, people in military quarters and in college dormitories, and unrelated individuals under 15 years old.	Sample	**X**
26	**PCT PUB ASST INC**	Percent Households with Public Assistance Income - Number of households receiving public assistance or welfare payments divided by the total number of households (times 100). Income data refer to calendar year 1999.	Sample	**X**

Person/Demographic, Housing, and Operational Variables				
Variable Number	**Variable Name**	**Description**	**Source**	**Variable HTC Score**
27	**PCT UNEMPLOY**	Percent of People Unemployed - Number of unemployed people 16 years old and over divided by total population 16 years old and over. Unemployed people are all civilians 16 years old and over who had no employment during the survey week, were available for work and had made specific efforts to find employment.	Sample	**X**
28	**PCT LIHH**	Percent Linguistically Isolated Households - Number of households in which a language other than English is spoken at home and no person (age 14 years or over) speaks English "Very Well" or "English only" divided by total households (times 100).	Sample	**X**
29	**PCT LIHH SPAN**	Percent Linguistically Isolated Spanish Households - Number of households in which Spanish language is spoken at home and no person (age 14 years or over) speaks English "Very Well" or "English only" divided by total households (times 100).	Sample	
30	**PCT LIHH INDO-EURO**	Percent Linguistically Isolated Indo-European Households - Number of households in which an Indo-European language other than Spanish is spoken at home and no person (age 14 years or over) speaks English "Very Well" or "English only" divided by total households (times 100).	Sample	
31	**PCT LIHH API**	Percent Linguistically Isolated Asian and Pacific Islander (API) Households - Number of households in which an API language is spoken at home and no person (age 14 years or over) speaks English "Very Well" or "English only" divided by total households (times 100).	Sample	
32	**PCT LIHH OTHER**	Percent Linguistically Isolated Other Language Households - Number of households in which a language other than Spanish, Indo-European or Asian and Pacific Islander is spoken at home and no person (age 14 years or over) speaks English "Very Well," divided by total households (times 100).	Sample	
33	**PCT_OCCP HU MOVED**	Percent Occupied Units Where Householder Moved Into Unit in 1999-2000 - Householders who reported moving into their house, apartment, or trailer/mobile home in 1999 or 2000 (January - March), divided by total occupied housing units (times 100).	Sample	**X**
34	**PCT WHITE**	Percent White - Number of people who indicated their race "White alone," divided by the total population (times 100).	100%	
35	**PCT BLACK**	Percent Black/African American - Number of people who indicated their race as "Black alone" divided by the total population (times 100).	100%	
36	**PCT AIAN**	Percent American Indian and Alaska Native (AIAN) - Number of people who indicated their race as "American Indian or Alaska Native alone," divided by the total population (times 100).	100%	
37	**PCT ASIAN**	Percent Asian - Number of people who indicated their race as Asian alone, divided by the total population (times 100).	100%	
38	**PCT NHPI**	Percent Native Hawaiian and Pacific Islander (NHPI) - Number of people who indicated their race as NHPI alone , divided by the total population (times 100).	100%	

Person/Demographic, Housing, and Operational Variables				
Variable Number	**Variable Name**	**Description**	**Source**	**Variable HTC Score**
39	**PCT API**	Percent of Asian and Pacific Islander - Composite of number of people indicating their race as Asian alone or NHPI alone, divided by the total population (times 100). This percent does not indicate the population who reported both Asian and NHPI.	100%	
40	**PCT 2P RACE**	Percent Two or more races - People who reported two or more races, divided by the total population (times 100).	100%	
41	**PCT SOR**	Percent Some Other Race - Number of people who indicated their race as Some other race alone, divided by the total population (times 100).	100%	
42	**PCT HISP**	Percent Hispanic Origin (may be of any race) - Number of people who indicated their ethnicity as Hispanic or Latino, divided by the total population (times 100).	100%	
43	**PCT NON HISP WHITE**	Percent Non-Hispanic White - Number of people who indicated their ethnicity as not Hispanic and their race "White alone", divided by the total population (times 100).	100%	
44	**PCT GQ**	Percent Population in Group Quarters - Number of people not living in households and classified by the Census Bureau as living in group quarters, divided by the total population (times 100). The two general categories of people in group quarters include institutionalized people and other people in group quarters (also referred to as "noninstitutional group quarters").	100%	
45	**PCT GQ INST**	Percent Institutionalized Population in Group Quarters - Number of institutionalized people not living in households and classified by the Census Bureau as living in group quarters, divided by the total population (times 100).	100%	
46	**PCT GQ NON INST**	Percent Noninstitutionalized Population in Group Quarters - Number of noninstitutionalized people not living in households and classified by the Census Bureau as living in group quarters, divided by the total population (times 100).	100%	
47	**PCT POP 0-17**	Percent Population Under Age 18 - Number of people classified as less than 18 years of age divided by the total population (times 100). Age is based on the age of the person in complete years as of April 1, 2000.	100%	
48	**PCT POP 65 OVER**	Percent Population Aged 65 and Over - Number of people classified as age 65 or older divided by the total population (times 100.)	100%	

Note: Variables marked with an "X" are used to derive the Hard-to-Count Scores.
For definition of the race groups used for Census 2000, see
http://www.census.gov/prod/2001pubs/c2kbr01-1.pdf

TRACT LEVEL PLANNING DATABASE WITH CENSUS 2000 DATA: DERIVED VARIABLES

Hard-To-Count Scores

The 2000 database file contains "Hard-to-Count" (HTC) scores which summarize the attributes of each tract in terms of enumeration difficulty. A total of 12 variables that were correlated with nonresponse rates in 1990 and 2000 are used to derive the HTC score (see variables marked with an 'X' in previous section).

The set of algorithms used to determine HTC scores is as follows:

(1) the value of each individual variable is sorted across geographic areas from high to low (e.g., sort tracts from highest percent poverty to lowest percent poverty),

(2) scores (0 to 11) are assigned to each variable for each tract (e.g., values of 11 are given to tracts with the highest poverty rates of over 44.3 percent and values of 0 are given to tracts below the national median poverty rate of 9.9 percent in 2000),

(3) the scores assigned to each of the 12 variables for a tract are summed to form a composite HTC score for the tract.

Table 1 illustrates the HTC scores and percent distributions for three specific variables: percent renter, percent not husband/wife household and percent poverty.

With twelve variables used to produce the HTC scores in the tract file, the scores can range from 0 to 132. The comparative standing of areas provides an indicator of the likely degree of difficulty in enumeration. Areas with the highest scores (e.g., over 70) are likely to be the areas with relatively high nonreturn rates and undercount rates while areas with the lowest scores are likely to be areas with low rates.

The predictive effectiveness of the database variables and HTC scores has been proven by testing against empirical measures of mail return rates and net undercoverage rates in the 1990 census, 1995 test census, the Census 2000 Dress Rehearsal, and Census 2000 (see Robinson and Kobilarcik, 1995, Word, 1997, Bruce et al., 2001, and Bruce and Robinson, 2003). In preparing for the 2010 Census, we can capitalize on the targeting power of the database and descriptive statistics for small areas. In particular, the variations in Census 2000 return rates and HTC scores among tracts may be excellent predictors of patterns of public cooperation in returning the 2010 Census questionnaire.

Table 1. Percentile Distribution of Selected Hard-to-Count (HTC) Variables for Tracts: Census 2000

		Values of Percentile Distribution		
Percentile Distribution	**HTC Score**	**% Renter Occupied Unit**	**% Not Husband Wife HH**	**% Persons below Poverty**
97.5 -100	11	91.3 - 100	83.9 - 99.2	44.3 - 100
95 - 97.5	10	82.3 - 91.3	78.8 - 83.9	37.2 - 44.3
90 - 95	9	69.8 - 82.3	72.0 -78.8	29.3 - 37.2
85 - 90	8	60.9 - 69.8	66.9 - 72.0	24.3 - 29.3
80 - 85	7	53.7 - 60.9	62.9 - 66.9	20.6 - 24.3
75 - 80	6	47.8 - 53.7	59.4 - 62.9	18.0 - 20.6
70 - 75	5	42.9 - 47.8	56.4 - 59.4	15.9 - 18.0
65 - 70	4	38.5 - 42.9	53.6 - 56.4	14.0 - 15.9
60 - 65	3	34.5 - 38.5	51.2 - 53.6	12.5 - 14.0
55 - 60	2	31.3 - 34.5	49.0 - 51.2	11.1 - 12.5
50 - 55	1	28.2 - 31.3	46.8 - 49.0	9.9 - 11.1
< 50	0	< 28.2	< 46.8	< 9.9

Note: See text for description of HTC algorithms to assign HTC scores.

Table 2 and Figure 1 illustrate the targeting capability of the Planning Database and associated variables in predicting patterns of public participation in Census 2000. Specifically, they demonstrate the strong association of patterns of Census 2000 mail return rates with 1990 Census rates. The categories in Table 2 span the spectrum of mail return rates ranging from very low mail returns in areas with concentrations of hard-to-count attributes to very high mail return in areas with an absence of hard-to-count characteristics (note the inverse relationship).

We compared patterns of response rates according to HTC scores. The 1990 and 2000 return rates shown in Table 2 and displayed in Figure 1 for tracts classified by HTC score are remarkably similar. The mail return rates vary systematically along the HTC continuum. The Census 2000 return rate was 61.7 percent in 2000 (58.3 in 1990) for the decile of 6,349 tracts with highest concentrations of hard-to-count attributes (HTC scores of 76+); the Census 2000 return rate was a much higher 85.4 percent (84.8 in 1990) in the decile stratum with the lowest concentrations (HTC scores less than 2).

Despite the uniformity of response patterns by HTC decile, differentials are observed in the increase in rates from 1990 to 2000. The mail return rates rose by the largest amount (by 3.4 percentage points) in the most difficult-to-enumerate areas (Strata 1). The second greatest gain (1.8 points) was in the second most difficult strata. The lowest increases in mail return rates were observed in the "easier-to-enumerate" deciles (strata 9 and 10; with slight increases of 0.6 percentage points).

Table 2. Comparison of 1990 and Census 2000 Mail Return Rates by Hard-to-Count Strata

Hard-to-Count Scores	1990		2000		Mail Return Change, 1990 to 2000
	No. of Tracts	Mail Return Rates	No. of Tracts	Mail Return Rates	
76 plus	**5,815**	58.3	**6,349**	61.7	**3.4**
57 to 75	**6,077**	65.4	**6,994**	67.2	1.8
45 to 56	**5,762**	69.5	**6,125**	71.1	1.6
36 to 44	**5,504**	72.5	**5,493**	73.5	1.0
27 to 35	**6,391**	74.9	**6,203**	75.7	0.8
20 to 26	**5,476**	76.8	**5,468**	77.8	1.0
13 to 19	**6,039**	78.4	**6,259**	79.5	1.1
7 to 12	**6,033**	80.3	**6,784**	81.3	1.0
2 to 6	**6,326**	82.8	**7,085**	83.4	0.6
< 2	**4,982**	84.8	**5,839**	85.4	0.6
Total	**58,405**	**74.8**	**62,599**	**76.1**	**1.3**

Source: Bruce and Robinson (2003). Note: Tracts not in mail universe (e.g., List/Enumerate tracts) are excluded. The mail return rates are plotted in Figure 1.

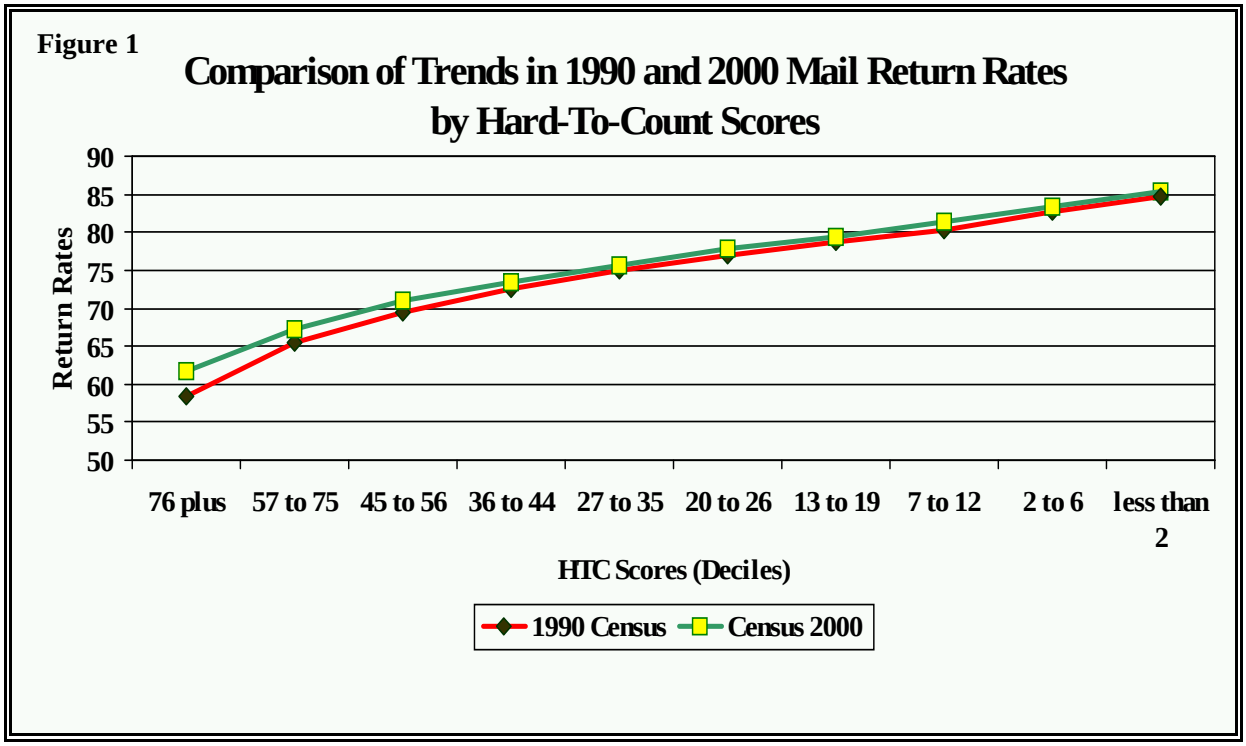


Table 3. Total Number of Tracts, Total Population, and Total Housing Units in 2000, by State

2000 Database for 2010 Planning	State/County Code	Total Number of Tracts	Total Population	Total Housing Units
National	**00**	**65,184**	**281,421,906**	**115,904,641**
AL	01	1,081	4,447,100	1,963,711
AK	02	158	626,932	260,978
AZ	04	1,107	5,130,632	2,189,189
AR	05	623	2,673,400	1,173,043
CA	06	7,040	33,871,648	12,214,549
CO	08	1,057	4,301,261	1,808,037
CT	09	814	3,405,565	1,385,975
DE	10	196	783,600	343,072
DC	11	187	572,059	274,845
FL	12	3,151	15,982,378	7,302,947
GA	13	1,616	8,186,453	3,281,737
HI	15	281	1,211,537	460,542
ID	16	280	1,293,953	527,824
IL	17	2,959	12,419,293	4,885,615
IN	18	1,411	6,080,485	2,532,319
IA	19	791	2,926,324	1,232,511
KS	20	725	2,688,418	1,131,200
KY	21	994	4,041,769	1,750,927
LA	22	1,106	4,468,976	1,847,181
ME	23	344	1,274,923	651,901
MD	24	1,216	5,296,486	2,145,283
MA	25	1,361	6,349,097	2,621,989
MI	26	2,712	9,938,444	4,234,279
MN	27	1,298	4,919,479	2,065,946
MS	28	605	2,844,658	1,161,953
MO	29	1,319	5,595,211	2,442,017
MT	30	270	902,195	412,633
NE	31	502	1,711,263	722,668
NV	32	482	1,998,257	827,457
NH	33	272	1,235,786	547,024
NJ	34	1,938	8,414,350	3,310,275
NM	35	447	1,819,046	780,579
NY	36	4,856	18,976,457	7,679,307
NC	37	1,554	8,049,313	3,523,944
ND	38	227	642,200	289,677
OH	39	2,930	11,353,140	4,783,051
OK	40	989	3,450,654	1,514,400
OR	41	754	3,421,399	1,452,709
PA	42	3,127	12,281,054	5,249,750
RI	44	233	1,048,319	439,837
SC	45	867	4,012,012	1,753,670
SD	46	229	754,844	323,208
TN	47	1,258	5,689,283	2,439,443
TX	48	4,387	20,851,820	8,157,575
UT	49	494	2,233,169	768,594
VT	50	179	608,827	294,382
VA	51	1,529	7,078,515	2,904,192
WA	53	1,318	5,894,121	2,451,075
WV	54	466	1,808,344	844,623
WI	55	1,317	5,363,675	2,321,144
WY	56	127	493,782	223,854

TRACT LEVEL PLANNING DATABASE WITH CENSUS 2000 DATA: FILE STRUCTURE OF THE DATABASE

The 2000 database is available in a Microsoft Excel™ format. The data file contains 65,184 total observations. Tracts with zero population in the source sample (long form) file of the census are excluded. The value of variables in the 2000 database may not be shown for small tracts; on the Excel spreadsheet file, variables with missing data are denoted by a " " (blank).

The database can be:

- Used for linking with spatial map data files, e.g., Tiger/Line
- Used to create thematic maps using commercial mapping software
- Used to export to Microsoft Access™
- Used for generating reports, cross tabulations, and simple analyses

Note: When reading and printing Microsoft Excel™ spreadsheet files, formatting adjustments such as resizing columns may need to be made depending on the configuration of your computer.

APPENDIX A. DEFINITION OF CENSUS DATA UNIVERSE AND ACCURACY OF THE DATA

The data collected for the planning database are based on Census 2000 tabulation geography from the following sources.

1. Summary File 1 (SF1) Hundred Percent Data;
2. Summary File 2 (SF 2) Hundred Percent Data;
3. Sample Edited Detail File (SEDF).
4. Summary File 3 (SF 3) Sample Data;
5. Hundred Percent Edited Detail File (HEDF);

1 *Hundred Percent Data for Census 2000* – The 100-percent data for 2000 contain basic housing and demographic information for all households and people in the United States. Variables in these files include housing unit level variables (e.g., vacancy status, tenure, number of people in the household) and person level variables (e.g., race, Hispanic origin, age.)

2 *Sample Data for Census 2000* – The sample data files for 2000 contain detailed housing and person level information for a sample of housing units that received the long form questionnaire. Variables in these files include housing unit level variables (e.g., type of structure and number of units, and availability of telephone service) and person level variables (e.g., persons in poverty, educational attainment, percent unemployed, and percent linguistically isolated household.)

ACCURACY OF THE DATA

The Tract Level Planning Database With Census 2000 Data is intended for use to help identify areas which may be potentially difficult to enumerate in Census 2010. Variables obtained from two different data sets (100-percent and sample) are included in the database file, and some of these variables are components of the Hard-to-Count Score (variable #11). Percent Households with No Telephone Service (variable #23), Percent of People Not High School Graduates (#24), Percent of People in Poverty (#25), Percent of Households on Public Assistance (#26), Percent Unemployed (#27), Percent of Households that are Linguistically Isolated (#28), and Percent of Households Who Moved in 1999-2000 (#33) are seven of the twelve variables combined to form the Hard-to-Count Score. These variables, along with all other language-based variables, are derived from the Census 2000 sample, commonly called the Long Form. The data are estimates of the actual figures that would have been obtained from a complete count. Estimates derived from a sample are expected to be different from the 100-percent figures because they are subject to sampling and nonsampling errors. Sampling error in data arises from the selection of persons and housing units to be included in the sample. Nonsampling error affects both sample and 100-percent data, and is introduced as a result of errors that may occur during the collection and processing phases of the census. Provided below are the links to the detailed discussion of both types of errors and a description of the sampling and estimation procedures.

http://www.census.gov/prod/cen2000/doc/sf1.pdf
http://www.census.gov/prod/cen2000/doc/sf2.pdf
http://www.census.gov/prod/cen2000/doc/sf3.pdf

APPENDIX B. DEFINITION OF CENSUS 2000 MAIL RETURN RATES

Census 2000 Mail Return Rates Methodology

The Census 2000 mail return rates used in this report are based on tabulation geography -- computed to reflect households in mailback enumeration areas whose final census data were taken from self-response occupied housing unit records, as opposed to households with enumerator-filled occupied housing unit records or with no housing record. The 2000 mail return rate is calculated by dividing the number of mail returns of occupied units by the number of occupied units in the mail back universe (times 100 to convert to a rate). The return rate is the measure included in the database to aid in the identification of potentially hard-to-enumerate areas. Tracts with missing mail return rates are denoted with a "." or "N/A". These were usually List/Enumerate tracts (including American Indian reservations).

Formula for 2000 Mail Return Rate:

Mail Return Rate = $\sum(N_{ip} / D_p)*100$, where N_{ip} represents the number of mailback forms from Record Form Types (RFT) subset (i) in p tract, and D_p represents the total number of mailout forms in occupied housing units in p tract.

The denominator is the number of occupied housing units in the mailout universe of Type of Enumeration Areas (TEAs) 1, 2, 6, 7, and 9.

The numerator is the number of occupied units in the mailback universe of Record Form Type (RFT) '01', '02', '03', '04', and '07'.

NOTE: These mail return rates for tracts are unofficial: they are not consistent with published Census 2000 response rates for tracts[1]. The tract-level rates used for the Planning Database are based on tabulation geography, while the published rates are based on collection geography. Also, the mail return rates are based on occupied housing units, while the denominator of the mail response rates include vacant as well as occupied units. So a return rate will be higher than a response rate, and more closely reflect the participation of households in the census (a vacant unit cannot return a form).

[1] U.S. Census Bureau, April 2, 2004, "Census 2000 Final Response Rates", See http://www.census.gov/dmd/www/response/2000response.html

Census 2000 Type of Enumeration Areas and Type of Forms

2000 Type of Enumeration Areas	Form Types
(1) Mailout/Mailback (MO/MB)	'00'= No return record selected by Decennial Response File (DRF2) processing
(2) Update Leave (UL)	'01'= Short Form MR
(3) List/Enumerate (L/E)	'02'= Long Form MR
(4) Remote Alaska (RA)	'03'= UL Short Form MR
(5) Rural Update/Enumerate (RU/E)	'04'= UL Long Form MR
(6) Military in Update Leave (M-UL)	'05'= Short Form EQ
(7) Urban Update Leave (UUL)	'06'= Long Form EQ
(8) Urban Update/Enumerate (UU/E)	'07'= Be Counted
(9) Update Leave Addition to Address Listing Universe of Blocks	'08'= Not used
	'09'= ICQ Short Form
	'10'= ICQ Long Form
	'11'= ICR Short Form
	'12'= ICR Long Form
	'13'= Not used
	'14'= MCR
	'15'= Not used
	'16'= SCR
	'17'= Enumerator Supplement, Short Form
	'18'= Enumerator Supplement, Short Form
	'19'= Short EQ
	'20'=Long EQ

Key: EQ - Enumerator-return type Questionnaire
ICQ- Individual Census Questionnaire (for service base enumeration (SBE))
ICR- Individual Census Report (for nonmilitary group quarters)
MCR- Military Census Report (for military group quarters)
MR - Mail-Return type questionnaire
SCR - Shipboard Census Report (for shipboard group quarters)

REFERENCES

Akers, Linda and Steve Alnwick. 2001. "Using the Planning Database to Plan Data Collection for Census 2000." Proceedings of the Social Statistics Section, American Statistical Association

Bates, Nancy and Mary H. Mulry. 2008. "Segmenting the Population for the Census 2010 Integrated Communications Program." C2PO 2010 Integrated Communications Research Memoranda Series, No. 1, October 28, 2008.

Bruce, Antonio, J. Gregory Robinson, Erin Love, and Guinevere Mills. 2008. "The Planning Database." Powerpoint presentation to the Steering Committees of the State Data Center and Census Information Center, U.S. Census Bureau, October 8, 2008.

Bruce, Antonio and J. Gregory Robinson. 2003. "The Planning Database: Its Development and Use as an Effective Targeting Tool in Census 2000." Paper presented at the Annual Meetings of the Southern Demographic Association, Arlington, VA, October 24, 2003.

Bruce, Antonio, J. Gregory Robinson, and Monique V. Sanders. 2001. "Hard-to-Count Scores and Broad Demographic Groups Associated with Patterns of Response Rates in Census 2000." Proceedings of the Social Statistics Section, American Statistical Association.

De la Puente, Manuel. 1993. "Why Are People Missed or Erroneously Included by the Census: A Summary of Ethnographic Coverage Reports." Proceedings of the 1993 Research Conference on Undercounted Ethnic Populations. Richmond, VA.

Jacobson, Linda. 2007. Panel Discussion on Perspectives on the American Community Survey at Full Implementation, Population Association of America, New York, NY.

McGovern D. Pamela. 2003. "Documentation of Clustering of Linguistically Isolated Households at the Tract Level in Census 2000." American Community Survey Research and Evaluation Program Discussion Paper. June 20, 2003

Robinson, J. Gregory, Carrie Johanson, and Antonio Bruce. 2006. "The Planning Database: Decennial Census Data for Historical, Real-time, and Prospective Analysis." Paper presented at the 2007 Joint Statistical Meetings, Salt Lake City, Utah.

Robinson, J. Gregory, and Edward L. Kobilarcik. 1995. "Identifying Differential Undercounts at Local Geographic Levels: A Targeting Database Approach." Paper presented at the Annual Meetings of the Population Association of America, San Francisco, April 8, 1995

U.S. Bureau of the Census. 1999. 1990 Data for Census 2000 Planning, CD-ROM with documentation, Washington, D.C. December 1, 1999.

Word, David L. 1997. "Who Responds/Who Doesn't? Analyzing Variation in Mail Response Rates During the 1990 Census." Population Division Working Paper Series, No. 19, U.S. Bureau of the Census.