Chapter 1

Quality of the 1980 Census Data on Ethnicity

Because of the importance of ethnicity, governments and some private organizations have frequently attempted to identify at least some of our people in terms of their ethnicity. The major sources and problems of measurement for the period through the 1970 U.S. Census of Population have been summarized in detail (Price 1980). Clearly, the most comprehensive information on the size and location of ethnic populations has been gathered by the United States government.

The federal government has attempted to record the country of origin of immigrants each year since 1820, and designers of every federal census since the first in 1790 wished to categorize the population ethnically in one or more ways. Yet the difficulties of defining the relevant groups and determining which people are in each are substantial, with definitions and approaches changing over the years. In addition, there is the more basic problem of locating and enumerating all the people in the country. Opinions vary as to the success of all these endeavors over the past decades.

However, the quality of ethnic data from the 1980 census was substantially better than that of previous censuses. A detailed description of the procedures and programs is found in the report of the Bureau of the Census to evaluate the findings are provided in the report of the National Research Council’s Panel on Decennial Census Methodology. The Bi-centennial Census: New Directions for Methodology in 1980 (Citro and Cohen 1985). In general, the specific questions asked on the questionnaire and the great efforts made by the government to count minority groups resulted in much-improved measures of ethnicity. The remainder of this chapter provides a description of the procedures and questions used for the 1980 census, an explanation of the characteristics and quality of the data produced, and cautionary notes regarding interpretation of data mapped in this atlas. Other ethnic data on country of birth and home language are not discussed.

Basic Design and Operations of the 1980 Census

The 1980 census was an attempt to enumerate and determine the basic demographic characteristics of all the people residing in the United States on April 1, 1980. Everyone was to be counted in terms of his or her usual place of residence. Students were counted where they lived while attending school. Citizens of other countries who were either permanent residents, attending school, or working in the U.S. were included together with any immediate family members here too. People residing here illegally were to be counted as well, but temporary visitors from other countries, such as tourists or businessmen, were not.

A questionnaire was provided for each housing unit or essentially separate living space, with seven basic demographic questions to be answered regarding each person in the housing unit. The form could be completed by anyone in the housing unit, with assistance by anyone. A housing unit whose occupants constituted a household, might include people interrelated as a family, but many households contained no such interrelated people. For most households, the seven questions appeared on what was called the short form of the questionnaire. Because these questions were to be asked of everyone, they were the source of what has been called the 100-percent-count data.

An additional set of many questions was combined with the basic seven on a long form of the questionnaire, to be completed only by those included in a random sample of approximately one in every six housing units. The long form covered a wide range of topics such as income, occupation, educational background, ancestry, place of birth, and language use in the home. Questionnaires in the Spanish language were available for households that wished them.

The location of housing units was determined with the assistance of local post office personnel, and it involved additional observations and checks in the field. Ninety-five percent of households received their questionnaires through the mail and were asked to return these by mail after completing them as of April 1. The other 5 percent of households in areas more difficult to reach by mail were canvassed by enumerators who systematically visited each unit in a small assigned area.

Specific Questions as Data Sources

The seven basic questions in the census asked for information on the ages, sex, and marital status of everyone in the household. Also, all persons were supposed to be categorized in terms of their membership in one of a series of 15 groups, most of which were defined in terms of race and various Asian origins. Although in 1980 the word race did not appear on the questionnaire, that question is referred to as the race question. Another question required that each person of Spanish/Hispanic origin be identified in terms of one of four major subgroups. The question regarding ancestry, like other questions on the long form, was asked of all occupants living in the sampled housing units.

The actual questions that were the sources of data used in this atlas have been reproduced below, together with special instructions that accompanied each question.

RACE QUESTION.

Instructions: Fill the circle for the category with which the person most closely identifies. If you fill the Indian (American) or ‘Other’ circle, be sure to print the name of the specific Indian tribe or specific group.
SPANISH-HISPANIC ORIGIN QUESTION.

Instructions: A person is of Spanish/Hispanic origin or descent if the person identifies his or her ancestry with one of the listed groups, that is, Mexican, Puerto Rican, etc. Origin or descent (ancestry) may be viewed as the nationality group, the lineage, or the country in which the person or the person’s parents or ancestors were born.

7. Is this person of Spanish/Hispanic origin or descent?

Fill in circle.

○ No [I left Spanish/Hispanic]
○ Yes, Mexican, Mexican-American, Chicano
○ Yes, Puerto Rican
○ Yes, Cuban
○ Yes, other Spanish/Hispanic

ANCESTRY QUESTION.

Instructions: Print the ancestry group with which the person identifies. Ancestry (or origin or descent) may be viewed as the nationality group, the lineage, or the country in which the person or the person’s parents or ancestors were born before their arrival in the United States. Persons who are of more than one origin and who cannot identify with a single group should print their multiple ancestry (for example, German-Irish). Be specific; for example, if ancestry is “Indian,” specify whether American Indian, Asian Indian, or West Indian. Distinguish Creole Verdiern from Portuguese, and French Canadian from Canadian. A religious group should not be reported as a person’s ancestry.

14. What is this person’s ancestry? If uncertain about how to report ancestry, see instruction guide.

(For example: Afro-Amer., English, French, German, Hutterian, Hungarian, Irish, Italian, Jamaican, Korean, Lebanese, Mexican, Nigerian, Polish, Ukrainian, Venezuelan, etc.)

Processing of Questionnaires

During the few months immediately after the census date of April 1, 1980, the questionnaires that people had mailed to various field offices were checked by the Bureau of the Census for completeness and consistency. Workers at these offices were directed to make several attempts by visits, telephoning, and finally talk with neighbors to obtain, as far as possible, completed and correct questionnaires. Then the forms were microfilmed so that the information could be read electronically, edited, tabulated, and stored in the bureau's computer files.

Where no response could be obtained for a particular question, a computer program at the Bureau of the Census was generally used to allocate a response based on the answer provided by a nearby household of similar age, sex, and race composition. Allocation was used to supply data for those characteristics whose general interrelationships with other characteristics in the same small area were to some degree predictable. It made possible a more complete description of the population than would otherwise have been possible. In cases where people were known to be living in a housing unit but virtually no information on them could be obtained, a category was assigned to an ancestry from a previous household was substituted for those that were missing.

For the United States an average of 1.4 percent of the individuals counted required substitution of their characteristics because of substantial or complete omission of responses. Also, race had to be allocated on an average of 1.2 percent of the individuals, and Spanish origin was allocated for 4 percent of the individuals. Allocation rates were higher for many other items. Because allocation resulted in both increases and decreases in the data values reported for various characteristics, it produced only a slight net shift of these values.

The omission rate for the ancestry question was 10.2 percent (Census 1983a). Ancestry responses were not allocated because of the diversity of ancestry groups and the characteristics of their members. This means that, on an average, the values for ancestry groups are somewhat lower than they would have been if the omitted response could have been supplied by some allocation formula. Lack of ancestry allocation is presumably part of the reason that ancestry data (single plus multiple responses) sometimes show lower values than the Spanish-origin (100-percent count) data for the same groups.

Within the limits of its budget the Bureau of the Census developed systematic procedures that should have reduced the number of errors to a minimum. These are explained in more detail in appendices of the state volumes containing sample data (Census 1983b) and in the technical information chapter of the documentation for the Summary Tape File 4, or STF4, data tape (Census 1983c).

Coping of Ancestry Responses

Special handling was required for responses to the question regarding an individual's ancestry. The Bureau of the Census compiled an almost exhaustive list of the possible written responses anticipated for the ancestry question. Numerical codes were attached to each possible response, with workers assigned to identify the code number for the handwritten ancestry responses. Then, most of the code numbers and anticipated ancestry labels were aggregated by the computer into larger and more commonly recognized ancestry groups. Details of the codes and procedures described in the following paragraphs can be found in appendix C3 of the STF4 documentation (Census 1983f).

Responses were reported in terms of two categories: single-ancestry responses and multiple-ancestry responses. People could indicate either one ancestry or two or more ancestries. In general, those individuals who wrote in at least two responses were reported in the multiple-ancestry category and coded in terms of the first two ancestries listed. However, if the two ancestries listed were alternate names for what was to be coded as one ancestry, then the response was considered to be that of a single ancestry. For example, people who wrote “Bavarian-German” or “Belgian-Fleming” were counted as single-ancestry Germans or Belgians, respectively.

Instructions on the questionnaire indicated clearly that people should not write in the name of a religious group in response to the ancestry question. This is because the Bureau of the Census is barred by law from asking a question regarding religion in any situation wherein an answer is required and not optional. Completion of the census questionnaire is legally required, and for this reason no data on religious affiliation can be collected as part of the census. However, religious affiliation is often an integral part of a person’s ethnic identity. Thus, much important information on ethnicity is not available due to this restriction. If a religious group name was given as a single response or the first part of a multiple response, the response was coded as an unspecified ancestry.

But if the second response was the name of a religious group and the first response was one of the many accepted ancestry responses, then the first response was counted as a single-ancestry response. Thus, “Russian Jew, Syrian Muslim, German Lutheran, Danish Mormon, Swiss Amish, Hungarian Protestant, and Lebanese Maronite” were all treated as single-ancestry responses for the first-named group.

Other responses, like “American,” “White,” “Appalachian,” “Anglo-Saxon,” and “Yankee,” were treated similarly to religious group responses. This means they were in-
cluded in the count of those with unspecified ancestry if they were either a single response or the first part of a multiple response. Yet if they were the second part of a multiple-ancestry response, the first part of the response was considered a single-ancestry response. For this reason "Japanese American," "Filipino white," and "Irish Yankee" were added to the first-listed group's single-ancestry count.

Computer tapes and published reports identified only what the Bureau of the Census considered to be the more important aggregations, and the tapes and reports differ in the particular ancestries identified and aggregated. Decisions regarding exactly which groups to combine and how to present the data were obviously of great importance, because they become the basis for impressions, discussions, and research regarding specific groups in the future.

In some cases, the aggregation procedures have made it difficult and expensive for some researchers to obtain data on smaller ethnic groups. For example, the most basic STF-4 tapes do not identify Hmong, Flemings, Alsatian, and German-Russian ancestries separately. Nor do these tapes identify specific Central and South American nationalities within the broad group of Spanish-speaking countries. People who claimed either French Basque or Spanish Basque ancestry were counted as Basques in the published report (Census 1983a) but were not included in the Basque totals on STF-4, even though they constitute almost half the total Basque-ancestry population. Also, people of Slovak ethnic background who identified the name of the country of origin, Czechoslovakia, were given the same code number as those of Czech ethnic background. Thus, the data make it impossible to know the relative sizes of Czech and Slovak ethnic populations.

Potential Problems and Appropriate Use of the Data

Estimated Undercount

It is difficult to determine the accuracy of a census count because better data against which to compare the census figures are rarely available. The number of people counted in every federal census since the first one in 1790 has presumably been less than the number of actual residents because enumerators did not find everyone. There is no accepted basis for estimating the extent of undercounting in the federal censuses before 1950.

Beginning in 1950, however, the Bureau of the Census used both demographic analysis and postenumeration surveys to provide independent estimates of the numbers of people missed in the census. Results have indicated a lower net undercount with each successive census. This was due largely to the much greater use of late tabulations in 1980 compared to the 1970 count (Baller, Herriot, and Passel 1982; Census 1987b).

Estimates of the people not counted in 1980, based on analysis of birth, death, immigration, and other records, suggest an overall 1 percent undercount of the legally resident population (Census 1987b). However, lack of good measures of illegal immigration and emigration meant that demographic analysis of this type has been flawed (Citron and Cohen 1985). If the total 1980 American population is assumed to include an additional 3 million undocumented aliens living in the country, the estimated undercount rises to 1.4 percent.

Census coverage was also evaluated by a postenumeration program based on the rates of omission in the census of people who were part of the sample chosen by the Bureau of the Census for the Current Population Survey. The results were generally consistent with demographic analysis (Citron and Cohen 1985). Undercounting in 1980 was most serious among blacks, as it had been in previous censuses, but that estimated net undercount was reduced from 7.9 percent in 1970 to 4.5 percent in 1980 (Census 1987b). However, evidence from these and other sources indicated that the undercount of adult black males was much higher than among any other groups. There was probably a net undercount of 10 to 20 percent of this population. Whites and other races were estimated to have been undercounted by 2.1 percent in 1970, but demographic analysis indicated that only .3 percent of that population had been missed in the 1980 census.

A major factor in the low estimated undercount for the nonblack population was probably the much-improved coverage of people of Hispanic origin in 1980 compared to the undercoverage of these groups in 1970. Probably about 2 million undocumented resident aliens (not commuters or short-term visitors) were counted in the 1980 census (Warren and Passel 1983). In 1980 the bureau made much greater efforts to persuade minorities and those illegally resident in this country to be counted, and the net effect was a total count higher by several million than had been expected.

Nevertheless, ethnic group leaders and organizations have often estimated their local group populations to be much larger than the 1980 census count indicated. Such discrepancies seem particularly great with respect to some groups of Middle American and South American origin. In areas where large numbers of people have been living illegally, with the major exception of undocumented Mexicans, who have typically crossed the border illegally, most other peoples who have been residing here illegally entered legally as temporary visitors but overstayed the limits of their visas. Although private estimates of local ethnic populations may be inflated due to weak methodology, self-serving motivations, and a tendency to include marginal individuals who may not consider themselves members, the unknown and possibly much-underestimated size of undocumented resident populations and the understandable motivation of such people to avoid discovery by the government suggest that the much higher private estimates may not be as erroneous as the Bureau of the Census imagines.

However, there is no way that the relative accuracy of the 1980 census results and estimates by knowledgeable local people can be established. In this book the census figures have been taken as the basis for the interpretations.

Subjectivity of Ethnic Identity

In designing questions concerning ethnicity, the Bureau of the Census has not attempted to determine the biological or genealogical heritage of anyone. Even the questions on race and Hispanic origins ask the individual to choose the category with which he or she most closely identifies. This is an appropriate approach, given the United States' relatively small and not well-known ancestries of individuals. Moreover, questions eliciting a subjective sense of identity provide data that is socially more significant than questions probing biological inheritance. The open-ended question on ancestry was a very useful improvement in this regard because it did not force the respondent to choose among a set of categories predefined by the Bureau of the Census.

The subjective interpretation of the meaning of the questions raises questions about the validity and the reliability of the response data (Lowry 1980). There could be significant differences between a person's actual predominant ancestral background and what that person thinks is his or her predominant ancestry. This subjectivity means also that responses are more likely to be influenced by factors such as the wording of questions, social pressures, and the perceived status of ethnic groups (Yancey, Erickson, and Juliani 1976). If membership in a particular ethnic group confers a special status, then even people who have no actual family connection with that group may choose to identify with it.

As a result of government programs and policies during the 1970s, people of mixed white and nonwhite racial backgrounds or mixed Hispanic and non-Hispanic ancestry have often found it advantageous to emphasize the minority part of their heritage. An increased identification with such minorities could have resulted in a higher count of the His-
panic-origin population and races other than white. Variations in the influences from one census to the next may pose difficulties in the interpretation of trends because it will be impossible to know the effect of such changes on the ethnic group numbers.

In the case of American Indians, the fact that some tribes have not been recognized by the U.S. federal government and so do not participate in federal assistance programs suggests that there may be practical reasons for a person of mixed Indian background to select one national or tribal identity over another.

The subjectivity of the race data has been an especially serious problem in Hawaii, where there has been so much mixing of ethnic populations that categorization of people in terms of a single race is usually unrealistic. For instance, between 1970 and 1974 over half of the babies born in Hawaii to Chinese, Korean, or Puerto Rican mothers had fathers of a different race (Lind 1980). For women identified as Japanese, Filipino, Samoan, or Hawaiian mothers during this period, the race of the father was mixed in over 25 percent of the births.

Adopted children probably assumed the ethnic heritage of their adoptive parents, but this was not necessarily the case. A person completing the questionnaire did not necessarily represent the viewpoint of the individual whose ethnic identity was asked. Some adults may not have known how other adults in the household would respond to the questions on the census, but for convenience probably answered for those who were absent at the time. Parents presumably responded for their children on the questionnaires, thus imposing their own opinions regarding ethnic identity on their children.

Single-Response Race Categories for Multiple Identities.

A source of additional and unknown subjectivity is the fact that available response categories in the race question were not mutually exclusive. For example, people of Chinese ethnic background from Vietnam or the Philippines were forced to choose one background. Non-Asian ancestry was the dilemma of those of racially mixed background, particularly the problem of choosing the appropriate response for children whose parents represented two different categories. This difficulty has already occurred in previous censuses, but increasing intermarriage has made the problem more serious. If respondents left that question blank for a child, editing procedures identified the child's race as that of the father in 1970 but that of the mother in 1980. However, a content reexamination of 1980 ancestry responses found among the U.S.-born people with U.S.-born parents a tendency to identity with the ancestry represented by grandparents or more distant immigrants on the father's side (Census 1986).

This procedural shift may have had no significant net effect on the numbers of various groups. Perhaps users of these data assume that children of mixed marriages would on the average be reported evenly between the two parental categories. However, a strong tendency for parents to prefer some categories over others for their children could seriously bias these data.

Identity by Ethnic Group or Country.

The wording of the questions probing ethnicity allowed a person to identify with either a country (political unit) of origin or an ethnic group within that country. As noted in the appropriate sections of the text, this is a major difficulty in interpretation of the ancestry data for people of Eastern European origin. For example, some people chose to identify their ancestry as Yugoslav, whereas others chose Serb or Croat. Some whose ancestors emigrated before World War I from the Austro-Hungarian Empire reported Austrian ancestry, but others preferred Croatian — just one of the ethnic groups within the empire.

Also, it is not known, for example, how people of Italian or German ethnic background who migrated from Chile or Argentina reported their ancestry. The data do not indicate if the people who reported Thai or Lao ancestry were identifying with the country or the ethnic group, and people of Jordanian ancestry might have been Palestinian refugees with Jordanian passports.

Consistency of Race Identities.

As part of the bureau's content reexamination study, a sample of over 24,000 individuals were interviewed about seven months after completion of the 1980 questionnaires in order to determine what extent responses to those questions were consistent over time (McKenny, Fernandez, and Masumura 1984). Whites and blacks were each about 98 percent consistent. All except 4 percent of the Asian and Pacific Islanders identified themselves in the same groups on both occasions; those not consistent had been placed as white in the census.

Consistency of American Indian Identity.

The number of American Indians reported in the 1980 census was 72 percent greater than the 1970 total (Census 1984a). Analysis of Indian births and deaths since 1970 has shown that probably a third of the increase is attributable to the fact that the father in 1970 but that of the mother in 1980. However, a content reexamination of 1980 ancestry responses found among the U.S.-born people with U.S.-born parents a tendency to identity with the ancestry represented by grandparents or more distant immigrants on the father's side (Census 1986).

Every school child is occasionally asked to write a report on his or her family, and thus to identify with a national or tribal identity. However, the tendency to identify with a tribe or nationality rather than the American Indian race probably reduced but slightly the totals of American Indians in the 100-percent-count data for major Indian states. In such states, tribal identities are higher — those people who listed tribal identities (except possibly those who identified as Cherokee) had identified themselves and their family in previous censuses and vital statistics as Indian.

The net effect is that data for states with traditionally important Indian populations are probably good, but data for states with very small Indian populations have been inflated by a shift in the responses of many people from white in 1970 to Indian in 1980. People's tendency in 1980 to identify themselves at least partly as American Indian was even more dramatically evident in the response of several million Americans who reported at least partial Indian ancestry. This total was almost five times greater that the 100-percent count.

Consistency of Spanish/Hispanic-Origin Identity.

The extent of consistency between initial and final's census responses and later interview responses suggested differences in the relative quality of the data items. People reported Hispanic origin with approximately 90-percent consistency, though Cuban identity was only 81 percent and Other Spanish only 55 percent consistent (McKenny, Fernandez, and Masumura 1984). Among people reported as Other Spanish in the interview, in the census 11 percent had indicated a Mexican origin and 31 percent a non-Spanish origin.

Also, comparison of responses to the Hispanic-origin question with ancestry and other 1980 census responses raises additional concern about the meaning of the response and the possibility of an overestimate due to the inclusion of people only marginally, if at all, Hispanic.

Counts based on Hispanic origin were 5 percent greater than could be expected from the total of the various Spanish ancestries (Fernandez and Cresce 1986). And 6 percent of
people listing a Hispanic origin did not indicate by their answer to any of the question (birthplace, ancestry, surname, or language used in the home) that they were likely to be of Hispanic origin (Tienata and Ortiz 1986). In contrast, 98.8 percent of the people who reported their Hispanic origin had no other Hispanic ancestry. Differences in the degree of consistency between the populations of Hispanic and non-Hispanic origin suggest that the Hispanic-origin data include virtually all people who are of Hispanic origin plus many who are very marginally Hispanic.

Another indication of the inclusive nature of the Hispanic-origin data is the fact that the total single- and multiple-ancestry responses for Cubans and Puerto Ricans equaled less than 74 percent of the 100-percent count for those categories. In the case of the Cubans, much of this differential is presumably related to a characteristically inconsistent ethnic identity, probably related to their relatively high status and rapid assimilation in the U.S. (Nelson and Tienda 1985). This suggests that the 100-percent count on Cuban-origin individuals is probably of better quality than the ancestry data because they include people who really were of Cuban origin but were not identified by other indicators.

Nevertheless, the Cuban-origin, Mexican-origin, and Puerto Rican-origin figures may be too low and the other Spanish-origin figures correspondingly too high. If so, this is because so many people who reported an Other Spanish origin later specified a Cuban, Mexican, or Puerto Rican ancestry (Fernandez and Cresce 1986). Seven percent of those reporting Other Spanish listed a Mexican ancestry even though Mexican origin was an available choice in the Hispanic-origin question. In California 12 percent of the Other Spanish-origin group (90,000 people) had Mexican as their only ancestry, and the same error was pronounced with respect to Puerto Rican ancestry in New York and Cuban ancestry in Florida. At the county level over 22,000 people in Cameron County, Texas (representing 11 percent of the county’s population), reported their origin as Other Spanish rather than Mexican. Although some families in this intensely Chi- cano part of south Texas may trace their descent from early Spanish (pre-Mexican) settlers.

The Mexican-origin count was known to be inflated by respondent error in many states outside the Southwest (Census 1982d), as explained in this chapter’s section on consistent and processing errors. It is not known to what extent overcount bias had an undercount in the Southwest, which could be expected in a population with such characteristic poverty, low educational levels, and English-language difficulties. But because the majority of the 2 million estimated undocumented aliens counted in the census (Warren and Pasek 1983) were of Mexican origin, the 1980 census Hispanic-origin count is probably at least equal to the total population reporting an identifiable Hispanic identity. People responding to the Other Spanish-origin question on Hispanic origin represented varied ancestries, but only 27 percent of the Other Spanish reported Spanish, Russian, or other Central or South American ancestry (Fernandez and Cresce 1986). Other common expected ancestries were Filipino, Portuguese, American Indian, Afro-American, and American.

Fully 39 percent of those reporting Other Spanish listed only a general Spanish or Hispanic ancestry response. The high proportion of such responses and the inconsistency between the Other Spanish and the ancestry responses make interpretation of these data difficult. The data can be explained partly by those people from New Mexico and Colorado who have long identified themselves as Spanish Americans; however, the fact that over a third of the Other Spanish in the large states of California, Michigan, Pennsylvania, Ohio, and Texas listed only this general response rather than some Central or South American ancestry suggests weakness in the data. In New York State, where the number of Central and South Americans was particularly large, 23 percent of the Other Spanish listed only the general Spanish-ancestry response. Problems with the Other Spanish count and its interpretation are suggested even more strongly by the 18 percent of Other Spanish respondents who reported a specific ancestry that was not one of the numerous possible Spanish ancestries.

Research in the early 1970s clearly showed the weakening of Hispanic identification with increasing distance from the immigrant generation and a heritage involving intermarriage with people of background other than Hispanic (Census 1974a). Individuals with Hispanic heritage on both sides of the family indicated a Spanish origin in 97 percent of the cases. However, only 21 percent of those whose Hispanic heritage was on but one side of their family indicated that their origin was Hispanic. This finding suggests that in areas where people of Puerto Rican, Cuban, or other Hispanic heritage have had little intermarriage with other groups, the 1980 data on Hispanic origin are not likely to be in error due to a weak or inconsistent identity. However, the 1980 Hispanic-origin data are probably less reliable in areas outside the traditional regions and centers of settlement of the particular group. Also, distinctions between the Other Spanish and the Mexican and Cuban categories sometimes proved difficult for people to make, even in geographic areas of concentration.

### Consistency of Ethnic Ancestries

In 1972, a reinterviewing of over 200,000 people who had first been sampled in a 1971 Current Population Survey established substantial differences in consistency among major groups (Census 1974b). In that study anyone reporting multiple ancestry was categorized as “Other,” a procedure that made consistency difficult to achieve. Responses a year later were more consistent for 88 percent of the people of Italian ancestry and 79 percent of the Polish-ancestry population. Somewhat less consistent were people of German ancestry (66 percent) and Irish ancestry (57 percent). French and Russian ancestries were reported with 62 percent consistency, but only 55 percent of those of English, Scottish, or Welsh ancestries so identified themselves a year later. Smaller ethnic groups were not studied.

Most of the differences among the populations studied can probably be explained by differences in mixed ethnic heritage and generational distance from the immigrant experience. Also, identities that were reported by individuals with a high degree of consistency were presumably those that were more important to the individuals involved. For purposes of comparative population numbers, these variations in consistency probably resulted in slight underestimates of the older, more mixed ancestry groups.

The Bureau of the Census also examined the consistency and bias of responses to the ancestry question in both the 1980 census, with results that showed generally greater consistency than in the early 1970s research (Census 1986). In this content reinterview study, conducted approximately eight months after the census date, interviewers visited the homes of a sample of over 24,000 people who had received the long form of the questionnaire. The ancestry question from the census was asked again, permitting an estimate of the inconsistency of responses. The major inconsistencies concerned persons who identified themselves in the interview as of American or U.S. ancestry but who had identified themselves in the census as of English, French, German, or Irish ancestry. However, disregarding those cases that could not be matched to any 1980 census ancestry response, over 84 percent of people reporting English, German, French, Irish, Scotch, and most other ancestries in the interview responses in the census (1986: 102). Many European ancestries showed a consistency of over 90 percent.

The design of the content reinterview study permitted an assessment of the bias or net differences in various ancestry responses between the census and the interview. The responses given during the interview were assumed to be closer to the truth than those written on the census form. In general, comparison of responses from the two sources showed only minor net differences, positive for some ancestry groups while negative for others. However, English and many smaller ancestries not specified in the report were
Comparability with Results of Previous Censuses

This atlas does not examine in detail the changes in sizes of ethnic populations. However, the historical growth of ethnic populations is frequently the focus of research and the text of this atlas does identify some trends.

In the case of the censuses, the choice of questions and data categories used have been dependent on changing perceptions of data needs so that the numbers representing various ethnic groups are often not directly comparable from one census year to others. Thus, there are special difficulties in describing changes in ethnic populations over time (Price 1980). For recent decades the Bureau of the Census has analyzed how ethnic population numbers vary as a function of the different ethnic indicators used (Levin and Farley 1982; McKenney, Farley, and Levin 1983).

The most important problem in comparability is the fact that until the 1970 census, ethnic groups within the white population were measured only in terms of foreign-born or foreign-stock populations. From 1850 through 1980 there are data on the numbers of people born in various foreign countries, but no other indicators have been used consistently. Data on the foreign-born and their children (often aggregated and labeled “the foreign stock”) have been provided with some consistency from 1870 through 1970, identified usually by country of birth and occasionally by mother tongue. These data were more objective than data based on self-identification. However, the widespread assumption that immigrant ethnicity would be lost by the grandchildren of immigrants and later generations meant that, for most of this century, no questions regarding white ethnicity beyond the foreign stock were asked. The many millions of people that were missed by such questions are indicated by the fact that 80 percent of American whites in 1940 had not one immigrant grandparent or parent and were not immigrants themselves (Lieberson and Santi 1985).

A pattern of greater strength of identity in traditional areas of settlement was found in some 1980 ancestry data. Analysis of state variations in proportions of single- and multiple-ancestry responses did not show marked regional patterns for many groups. However, the proportion of single-ancestry responses for some of the larger groups did vary regionally. For instance, the proportion of single-ancestry responses for American Indians was highest in the Dakotas, Montana, New Mexico, Arizona, and for the French, the highest proportions of single-ancestry responses were in Louisiana and the New England states. It is reasonable to expect that identity is stronger and more consistent in places where ethnic group numbers are greater.

In the 1980 census a completely new and open-ended question on ancestry was introduced into the attempt to measure ethnicity more appropriately. The attempt made possible much-improved data on the ethnicity of small groups and those long in this country, but it is impossible to compare the results with any previous census counts. However, a supplement to the bureau’s November 1979 Current Population Survey included several questions on ancestry, mother tongue, and current language use so as to provide a link between the 1970 and 1980 ethnic data (Census 1982a).

Data on race groups can generally be compared from one census to another although problems remain with subjectivity and the dilemma of identity of those of mixed race. There have been changes in labels and categories used for nonwhite groups, but the relatively greater consistency of measurement of race and its obvious importance have meant that these data have been assumed to be adequate for the study of historical trends. The several new categories added in 1970 and 1980 to the traditional race groups have resulted in better 100-per-cent data for more groups. However, historical ethnic indicators of the white population is difficult because 40 percent of the Hispanic-origin population who on the race question reported “Other,” were shifted arbitrarily into the white category in 1970 but not in 1980.

Although in 1970 the Bureau of the Census changed its method for identifying a person’s race, it is not known if this had any effect on the historical comparability of data. Prior to the 1970 census (and in many rural areas in 1970) the enumerator assigned race identities to members of households on the basis of visual inspection and interviews. More recently, respondents have identified their own race and that of members of their household on the questionnaires, thus eliminating most errors due to misperception.

Changes in the several indicators used to measure the size of the populations of Hispanic heritage mean that historical comparisons must be made with extra care. In addition to questions regarding place of birth and mother tongue for the foreign-born and foreign stock, some special ethnic indicators have been devised beginning in 1960. There have been modifications in their use from one census to another, and some indicators were applied only to populations in the five states of the Southwest. Even what might appear to be similar indicators can pose difficulties. Comparisons between 1970 and 1980 numbers would appear to be appropriate at the national level because a 5-percent sample of the total U.S. population in 1970 was asked an open-ended question or descent that was quite similar to the Hispanic-origin question asked of all people in 1980. The difficulty lies in the fact that coverage of the Hispanic populations (including people illegally resident) was significantly better in 1980. Increases
in numbers were a function of both population growth and improved coverage.

Respondent and Processing Errors
People in the households who completed the census forms may have misread or not understood the questions or may have accidentally or purposefully given wrong answers.

One particular error of this type was especially noticeable. In eastern states a surprisingly large number of people of Mexican origin were reported from areas where such people had been traditionally few in number, particularly in the Southeast (Census 1982b). Telephone checks with a large sample of such respondents from states in the eastern half of the country indicated that many people had wrongly reported themselves as of Mexican origin. The error was great enough to have misled people who may have misinterpreted the meaning of the letters "Amer." in the phrase "Mex-Amer." Thus the reported totals of Mexican-origin people in many states were probably twice as large as they should have been.

The introduction of error due to the gathering and processing of the data has been presumably reduced by the mailing of questionnaires. In earlier censuses the data were more dependent on the honesty and thoroughness of thousands of enumerators who personally visited homes to record the information needed, but it is not known how successful these efforts were. In the 1980 census, omitted and inconsistent responses were noted at check-off fields, and the work of those who were responsible for getting the correct information was checked as much as possible. Some errors, such as the counting on several Indian reservations, were found and corrected in later publications, but totals for Blackfoot, Micmac, Mohawk, Falu, Seminole, and Shoshone may be low in some states due to processing errors (Census 1984a, 1985).

One county’s unusually poor rate of response to the long-form questionnaire affected some maps. Only 22 percent of the sampled households in Nevada’s Esmeralda County (population 772) answered two or more of those questions (Census 1983b, table C-3). Because so few people responded to the ancestry question and because it had been this bureau’s decision not to allocate missing ancestry responses, the percentages of large ancestry groups (particularly English and Irish) reported in the county’s total population were exceptionally low. Thus, Esmeralda County’s high visibility on certain maps probably reflects these data weaknesses rather than any significant differences in real ancestry proportions compared to surrounding counties.

Responses to the question regarding ancestry had to be read and coded by workers hired by the Bureau of the Census. This operation introduced the possibility of coding errors, especially since ancestry and other coded items were all given the same code. Examination of ancestry totals by state and county in relation to expected patterns of local ethnic settlement suggests that coding errors were not widespread. However, some anomalies may be explained by confusion on the part of coders between numerical state codes and ancestry codes of the same number. For example, the unexpectedly large number of French Basques in Nebraska (2,707 total responses) suggests this type of error since both the ancestry code and one of the two state codes were number 31. Also, it is likely that the Lummi Indians reported in North Carolina represent a miscoding for Lumbee (Census 1985).

All coding took place at only three different locations, and questionnaires were not assigned to coders on the basis of small geographical units like counties. Several coders worked on questionnaires from any county, and any one coder handled questionnaires from many counties. If some one or two individuals made somewhat systematic errors that were not spotted during checking by the supervisor, such errors could have had no significant effect on the results for any one county compared to other counties in a state.

The fact that a microfilm copy of responses for most items was read electronically into the computer system reduced the potential for human error in later processing.

Estimated Error Due to Sampling
Any sample is only an estimate of what the data would have shown if the information had been supplied by all people rather than only a sample. Characteristics reported for the particular sample chosen will differ from those reported for a different sample taken from the same population. Because the ancestry data in this atlas are based on a sample, it is important to understand this source of weakness in the data, the significance of which is greater for smaller data values in relatively small total populations.

The statistic that describes the relative importance of the potential error due to sampling is called the standard error of the estimate. It is possible to calculate this for a particular data item if an estimate of the total population of the area involved, the percentage of the population sampled, and the number reported for that data item in the census tables. The standard error of the estimate is an indicator of the range of numbers that could be expected from a series of different samples taken from the same total population. It shows the range of data values that could be expected in 68 percent (1 s.e.) or 95 percent (2 s.e.) of all possible samples of that population.

The method of calculation is explained in appendix D of the published volumes of sample data for each state (Census 1983b) and in the documentation for ST44 (Census 1983b). In the given calculations there is a standard error adjustment factor that varies slightly from one ancestry to another. These factors can be found in table 1 of Ancestry of the Population by State: 1880 (Census 1883a).

To aid in the interpretation of the many maps based on ancestry data it is most useful to present a few examples that illustrate the sorts of standard errors that are applicable to these data. The following situations and calculations assume a sampling rate of one-sixth of the housing units.

Sampling Errors of Ancestry Data in Selected Counties.
In Barton County, Kansas, the 31,343 estimated people of German ancestry (single ancestry only) represent 34.4 percent of the county’s total population. With such a high percentage involved, the standard error is .5 percent. This means that if the chances are about 95 percent that other samples of the same total population would have produced a result within the range of the 1.6 percent above or below the 34.4 percent figure.

In a similar example involving large data values for the sample, there were 43,473 Haitians (single ancestry only) estimated to be within the total New York City population of 7,071,639. Calculation shows the standard error of this estimate to be 838. There is a 95-percent chance that the number of Haitians counted would be within the range ±1,676 from the 43,473 figure. In other words, there is a 95-percent chance that the number of single-ancestry Haitians in New York City reported by a complete census would have been between 41,797 and 45,149.

Reliability of estimated data values is decreased substantially if the data value or the total population is small. The potential error due to sampling is so significant that on most maps the data categories of less than 500 (shown by two sizes of dots) should be interpreted with some caution. For example, in Fayette County, Kentucky, within the total population of 20,165 there were an estimated 1 people people who had listed Danish as their ancestry if a 100-percent count had been taken. But the standard error of the estimate is 9: there is a 95 percent chance that the complete count would have listed between 13 and 89 Danish single-ancestry responses. Similarly, the standard error of the estimate of 102 Ukrainians in Wyoming County, Pennsylvania (total population of 26,433) is 26. Thus there is a 95 percent chance that other samples taken would have shown between 50 and 154 Ukrainians.

The ancestry data in appendix 2 include estimates of
county ancestry populations that are far too small to be considered at all reliable. In most cases, reported numbers in less than 7 reflected only a single individual’s response. Depending on both the sample size of key demographic subgroups in the county and other factors, that one response was given varying weights as indicated in the data.

**Sampling Error in Rural Counties.** The sampling error is much less in counties that are predominantly rural because in such areas the sampling rate was usually between 43 and 50 percent. Standard errors can be calculated for various rates of sampling and ancestry numbers estimated in different counties. Sampling rates and population totals for each county in a state can be found in table D of the General Social and Economic Characteristics (Census 1983b) volume for that state.

The calculation is made by means of a standard formula, such as that found in Arkin and Colton (1963: 22) and the appropriate ancestry group multiplier as determined by the Bureau of the Census. For example, Nemaha County, Kansas, has a total population of 11,211; and Edwards County, South Dakota, a total population of 5,159. Both had sampling rates of 50 percent. In Nemaha County, where 214 Swis were reported, calculation of the standard error shows that there is a 95-percent chance that a complete count would have found between 180 and 248 Swis. Similarly, the census estimated 62 Czechs in Edwards County. There is a 95-percent chance that the number counted if there had been no sampling would have been between 43 and 81.

Users of this atlas should generally focus on those counties reporting data values over 500 for various ancestry groups unless the county of interest is predominantly rural. In this way they will be less likely to overestimate the precision of the ancestry data.

**Sampling Error of PUMS Migration Data.** The special Public-Use Microdata Sample A file (Census 1983c) has been used only for special ancestry tabulations not available in STF4B and for maps of net interstate migration flows of certain race and Hispanic-origin groups. This file represents a sample of 5 percent of American households, with records of over 11 million individuals, with name and local addresses deleted for confidentiality. Over a quarter of the households that received the long form of the census questionnaire were included in this file, but with small race groups in small states there is substantial standard error of the estimate.

Formulas from which standard errors may be calculated can be found in the documentation for the PUMS file (Census 1983d). Nearly all of the net migrations mapped were so large that sampling errors would not significantly change the relative sizes of flows. However, some samples provide an indication of the sampling error associated with some of the smaller net migrations. For people of Mexican origin only, the lifetime migration flows between Texas and Indiana showed an estimated 8,400 people born in Texas but living in Indiana in 1980, and 4,320 people born in Indiana but living in Texas in 1980. Thus, the estimated net migration of the Mexican-origin population from Texas to Indiana was 4,080. The standard error of the net estimate is a function of the variance of estimated migrations in both directions (Census 1983d: 24). Including the 1.2 adjustment factor for race and Hispanic-origin data, there is a standard error of ±479. Then, 95 percent of all possible samples (2.5 s.e.) would produce estimates of the net migration between 3,122 and 5,038.

Of all specific net interstate migrations shown in this map series, the largest standard error relative to the net flow was that of Chinese from Arizona to California. Although most arrow widths are based on estimates with much less sampling error, the estimation error in the data is large.

There were an estimated 380 Chinese who were born in California but living in Arizona in 1980, with those born in Arizona but living in California estimated at 1,120. This produces a standard error of ±203 from the estimated 740 net migrants, with 95 percent of all possible samples resulting in estimates between 334 and 1,146 net migrants.

**Conclusions on Data Quality for Ethnic Populations.**

The best ethnic distributional data in U.S. history was produced by the 1980 census. The few race categories of previous censuses were expanded to include many more Asian and Pacific Island peoples, and a question on specific Hispanic origin was included for the first time on the short form. Thus, complete-count data that are especially useful in characterizing small areas and populations are available for many more groups.

The race and Hispanic-origin questions provided what is generally considered the best census data ever on minority numbers. However, the data on Indians are probably good for traditional areas of Indian settlement but inflated in states with substantial small or highly urbanized populations (e.g., West Virginia, Alabama, California). People of Other Spanish origin include so many who gave unspecific Spanish ancestry responses or were inconsistent in expressing this identity over time that there are probably weaknesses in data based on this question. The children of people who were born in Mexico was probably an overcount in eastern, midwestern, and especially southern states, with the exception of Texas. However, the fact that all these data represent information gathered from everyone means that they contain no errors due to sampling.

The sample data on ancestry measured the diversity of ethnic groups better than did any other recent census. Ancestry was reported for the entire sample population, not just for the immigrant generation and their children, as had generally been the case during the last 120 years. The result was a more appropriate measure of those ethnic groups characterized by relatively little recent immigration. Also, responses to the open-ended question were coded into a great many ancestry categories, so that many less-prominent or smaller ethnic groups could be identified for the first time. Interviewing approximately eight months after the 1980 census indicated that most of the larger ancestry groups had been reliably counted in the census. However, the population counts of the smaller groups may be overestimated.

Legal restrictions on obtaining religious group membership data prevented the reporting of an important aspect of ethnicity. Distinctions between Christian and Jew in Eastern European ancestries and among Christian, Muslim, and Jew in Middle Eastern ancestries are socially important, as is the distinction between Catholic and Protestant among several European ancestries. The census data cannot illuminate these patterns, however.

It is not known how people of mixed race or partial Hispanic origin responded to the categories available on the questionnaire. Individual whites, blacks, and Asians have tended to identify their race consistently over a short period of time, but many people seem to vary their identities between white and American Indian and between Other Spanish and non-Spanish identities. People of different European ancestries have varied in the consistency with which an ethnic identity is reported, but the option of single or multiple categories probably increased the reliability of responses over time.

Biases and errors that were not due to sampling may well have occurred within any group. The error due to undercounting of people who avoided the census because they were living in the U.S. illegally cannot be known for the many ethnic groups which probably included such people, but it
seems likely that undocumented aliens were clustered in the larger ethnic settlements, which then might well have been substantially undercounted. Other errors probably occurred in a geographically random pattern. The subjectivity and ill-defined nature of the identity sought in the questions are matters of concern, but the potential effects on distributional patterns are probably small. Although some weaknesses or errors in the data may remain, they are more likely to affect the total numbers of an ethnic population in the United States than to bias the data on the spatial pattern of the group’s population. This atlas is not essentially concerned with the total number of Americans measured in the various groups. The inherent focus of the maps on the relative numbers in different places effectively means that, except for sampling errors in small populations, the patterns portrayed in this atlas are probably substantially accurate for most ethnic groups.