Chapter 2

Data Selection and Map Design

This chapter presents the reasons for selecting particular sets of data and for using certain cartographic designs and procedures.

In this atlas most maps can be grouped into several basic types, each of which normally portrays different bodies of data. The most important maps were prepared from U.S. census computer tapes and show 1980 county patterns of race, Hispanic origin, and ancestry populations. These are the full-page and quarter-page maps that constitute the great majority of the maps in the atlas. There is also a similar map of the estimated Jewish population.

Four special cartograms portray certain 1980 census data for states or large metropolitan areas. These cartograms show the ethnic components of certain populations by means of graduated squares of different colors. The two that are in the final chapter show many different ethnic populations and are considered summary graphics for this atlas.

Major changes in distributions over recent decades are indicated by maps of net lifetime migration of the larger race and Hispanic-origin groups and a 1981-1984 net migration map of Southeast Asian refugees. Finally, an indication of the historical antecedents of the 1980 patterns is provided by some 30 maps of ethnic populations as of 1920.

The atlas has been designed to minimize the effect of problems in historic comparability of data. The geographical perspective represented by the many maps based on the 1980 data takes advantage of the strength of the census data — the fact that these were collected and processed in ways that were highly uniform from place to place.

**County Maps: 1980 Data**

**Sources for Census Data**

Results of the 1980 census have appeared in numerous published volumes and computer tapes. National and state totals can be found in the supplementary reports on race (Census 1981a), Spanish origin (Census 1982d), and ancestry (Census 1983a). County and urban place data for race and Hispanic-origin groups and for the largest ancestry groups are available in the published volumes General Population Characteristics and General Social and Economic Characteristics (Census 1982c, 1983b), each of which is composed of a volume for each state and a U.S. summary. However, ancestry data on smaller groups at the county level cannot be found in any published report.

In order to reduce the burden of processing tremendous amounts of data, tapes were used for the preparation of this atlas as much as possible. Summary Tape File 1C (Census 1981b) provided the short-form, complete-count (or 100-percent-count) data. This was the source used for mapping populations of the different races and Hispanic origins. The figures are the same as in the General Population Characteristics volumes (Census 1982c) and reflect minor corrections from tape provisional race data that appeared in the 1981 publication on race groups (Census 1981a). Summary Tape File 4B (Census 1983e) supplied the long-form data for the sample population and made possible the mapping of various ancestry populations. Data for states and the U.S. were taken from the U.S. Summary of the General Population Characteristics and the supplementary reports (1982c and 1983a).

Complete-count data have been used in this atlas where possible, so as to avoid sampling errors in the small ethnic populations of many counties. For several race and Hispanic-origin groups there are both 100-percent data and sample data. These figures differ slightly, with final counts based on the sample typically higher. This is because the Bureaus of the Census had more time to edit the responses of the sampled population. For example, in the race data, some sampled individuals who had been reported originally as Other were shifted to a specific group on the basis of their responses to other questions. Although the sample data may for this reason be slightly more accurate for large populations, the sampling error for small county populations makes the 100-percent count superior for this atlas.

Appendix 2 presents the raw data on which the maps are based. It includes county-level data on 39 ancestries for which the Bureau of the Census has published no figures for areal units smaller than states.

Data from the census files were transferred and typeset for the appendix by computer. This reduced the potential for error and none was found in numerous final checks. The only inconsistency with the data in Bureau of the Census publications concerns Portuguese ancestry: because Madeirans were not included with the Portuguese in this atlas, their presence in certain counties is indicated by the slightly larger numbers of Portuguese in census publications. Users of the appendix are cautioned that sampling error in the ancestry data will be very large for populations under 50 or 100, especially in counties that are not completely rural. (See the discussion of sampling error in chapter 1.)

Although the maps show only totals for New York City, appendix 2 has the data for each of the five counties (boroughs) that make up that city. It also contains figures for three subgroup within the aggregated Arab nationality ancestries, and Guyanese are enumerated separately from Trinidadian and Tobagonians.

**Coverage of Ethnic Groups**

Most ethnic groups for which data on the tape files are available have been mapped. However, the smallest ancestry groups are not covered in this atlas. Generally, ethnic groups for which there were fewer than about 20,000 single-ancestry responses were not included.

In some cases small ancestry groups have been included in this atlas as part of a larger ethnic group. This has been done for ancestry groups in East and West African countries and also those from smaller, formerly British territories in the Caribbean Sea. There was no way to separate Christian from
Muslim ancestries in the Middle East, except for the (Christian) Greeks. Thus, people who listed their ancestry as one of the predominantly Arab countries, both in the Middle East and North Africa, were added to those identifying an Arab ancestry. The map of French ancestry includes people who identify themselves as French Canadian, just as French Canadian ancestry numbers have been added to those reporting Portuguese ancestry. Similarly, Guyanese ancestry numbers were combined with those from Trinidad and Tobago. Ancestries for most individual Central and South American countries could be mapped only at the state level because data were not included on STAF.

Ancestries that represented an ethnically varied region or country, such as Scandinavian, Yugoslavian, and Eastern European, were not mapped, especially if more ethnically specific data could distinguish the distributions of the ethnic groups involved. In the Yugoslavian case this resulted in separate maps for Croatian, Slovene, and Serbian ancestries. However, because about 200,000 Americans chose to identify their country as Yugoslavian compared to a total of about 220,000 reporting the three ethnic ancestries, the symbols on the Slovene, Croatian, and Yugoslav maps suggest unrealistically small populations.

This atlas does not cover two additional ancestry categories for which data are available. The population of Austrian ancestry (23,324) was barely sufficient to warrant treatment, but their status as an ethnic group was so doubtful (Parkin 1980) that they were not included. Also, Canadian ancestry was not mapped because the numbers represented an unusually small proportion of immigrants and their descendants. This was probably due to a tendency for people from Canada to form their ethnic identity in terms of their European ancestry prior to immigration to Canada. This has certainly been the case with French Canadians. Moreover, a substantial number of Scottish Canadian or Italian Canadian ancestry for example, probably identified more with that Scottish or Italian heritage than with the Canadian. An indication of this fact is the reminder of Canadian-born people in the U.S. in 1980 was three times the size of the Canadian ancestry population (Lieberson and Santi 1985). In addition, people who reported Canadian ancestry represented 12 percent of the British Canadian white foreign-stock population in 1910 (1.8 million people), indicating that any map could portray a small and perhaps unrepresentative portion of Americans of Canadian background.

Choice of the County as the Basic Mapping Unit

Most of the 50 states of this country are so large and varied that maps of ethnic populations at the state level are usually too generalized to be very useful. Ethnic settlement has typically been highly localized in specific places or types of places, and these features cannot be suggested on small-scale maps. Alternatively, ethnic populations could be identified in terms of individual towns and cities, but for national coverage this method represents more detail than would normally be desired. The over-3,100 counties and equivalent areas (e.g., independent cities in five states, census divisions in Alaska, and parishes in Louisiana) are appropriate units for relatively detailed treatment of the country as a whole.

In addition, decentralization of cities since 1920 has meant that many members of ancestry groups in metropolitan areas no longer live within the central cities identified on the 1920 maps. This means that such city data are inappropriate for many ethnic groups in 1980. The automobile and suburbanization of homes, jobs, and most city functions have made both the older central cities and nearby suburbs highly accessible. For those who wish to participate, the ethnic community could be effectively knit together as in the period of higher density sixty or more years ago even though people's homes are much more scattered today. As a result, the larger area units of the county or the metropolitan area are more appropriate for describing local community sizes in 1980. The county unit permits uniform treatment of all areas within the 50 states and was chosen as the basic areal unit for most 1980 maps.

Use of Single-Ancestry Data

The data for ancestries have been presented by the Bureau of the Census in two forms: single-ancestry and mixture-ancestry sets. Single-ancestry data are used in this atlas because many of the ancestry categories presented separately in the single-ancestry data were aggregated into larger groups in the multiple-ancestry tables of the STAF files. Thus, the multiple-ancestry data could not be interpreted into any consistent treatment of ancestry groups.

The sizes of ancestry groups can vary a great deal depending on whether single- or multiple-ancestry data or an aggregation of the two is used as the basis for measurement. This is because groups differ in the relative proportions of single- as opposed to multiple-ancestry responses. Those groups with histories of predominantly recent immigration had high proportions of single-ancestry responses. Conversely, those with long histories in this country, much geographical mixing, and high rates of intermarriage with other groups had understandably much higher proportions of multiple-ancestry responses.

The single-ancestry data are excellent indicators of most non-European ethnic populations, in which the introversion of the people have usually immigrated within the last few decades. In these cases the single-ancestry numbers far exceed the multiple-ancestry totals because intermarriage with people of other ancestries that have been low and ethnic identity is less often blurred or diluted.

However, for the European ancestry groups, which have longer histories in this country, the single-ancestry numbers are satisfactory but less than ideal. For any one county they underestimate the size and percentage values of the ancestry groups because, of course, they exclude all people who listed more than one ancestry. However, National Opinion Research Center (NORC) surveys during the 1972-1977 period showed that, in addition to the 53 percent of Americans who expressed one clear ethnic identity, another 25 percent, who were of mixed or ambiguous background, could choose one dominant ethnicity if they were urged to do so (Smith 1980). Moreover, similar surveys in the 1977-1980 period indicate that ethnic identity, especially for young people who have attended college, tended to identify with a single ancestry more readily than older people of mixed ancestry (Alba and Chamlin 1983). This is so despite the increasing social and cultural assimilation experienced by these younger people, and may be partly due to the reawakening of interest in roots and ethnicity during those years. Altogether, these results suggest that, despite the fact that most Americans have backgrounds including more than one ethnic group, the single-ancestry totals were as large as they were, even for those of European ancestry.

Single-ancestry data are perfectly adequate for comparisons between counties in the relative sizes and percentages of ancestry populations, which is the major concern of this atlas. Moreover, the somewhat low values for European-ancestry populations are generally small enough to allow for small problems which are emphasized in this atlas—counties where an ancestry group represents a relatively high percentage of the total county population. For example, in Bay, especially in County, Massachusetts, the single-ancestry Irish population in 1980 was almost 50 percent larger than the multiple-ancestry Irish population. Also, in most strongly German counties of the Midwest the single-ancestry German population was from 2 to 6 times the size of the multiple-ancestry German population.

The single-ancestry numbers appear to be adequate for most comparisons between counties as to the relative strength or intensity of an ethnic population.

In the text of this atlas, references to a population of a certain ancestry indicate the single-ancestry population only. The total ancestry population of any ethnic group is the sum of both the single- and the multiple-ancestry numbers. Any
reference to multiple-ancestry numbers is specifically indicated.

Non-Census Data on Jews

Because Congress decided that questions pertaining to religion cannot be asked in the decennial censuses, the 1980 census can provide no data on Jewish populations. However, Jews are an ethnic group, as well as a religious group, and clearly have a sense of shared identity as a special people. A similar argument has sometimes been made for other populations, such as Mormons and white Southerners, who also have a strong sense of collective identity. However, for this atlas, only Jews were judged so obviously an ethnic group that their distribution could not be omitted.

The number of Jews affiliated with many temples and synagogues is available from a large multidimensional national survey of religious group membership in 1980 (Quinn et al. 1982), but only about half of American Jews are formally affiliated with such organizations (Cohen 1982). Moreover, that count of Jews (totalled only about 800,000 for the entire United States—only 14 percent of the 5.7 million Jews estimated in the source that was chosen for this atlas.

Better data can be found in the estimates of sizes of Jewish communities that are published in the American Jewish Year Book (Chenkin and Miran 1980). Unlike the U.S. census, however, the Year Book figures were not collected through either a census or a sample survey of individuals or households. Rather, the numbers have been developed from estimates prepared annually by Jewish federations across the country and, in areas not served by a federation, less frequent estimates handled through local synagogues (Chenkin 1982). A federation's estimate of its community size is typically based on a list of households known to be Jewish, adjusted to include others thought to be in the area but not listed, and then multiplied by a figure representing the average number of people in a household. Surveys of the Jewish population in 12 major metropolitan areas have shown an average household size of about 2.4 persons—a decrease from the 2.8 figure established by the 1971 National Jewish Population Study and used as the multiplier for the 1980 data (Tobin and Lipsman 1984). This factor alone suggests that the 1980 estimates for most counties may be too high although other weaknesses in the data are probably of greater significance.

Also, some Florida federations may have counted seasonal visitors included in the estimates of northern communities (Chenkin and Miran 1980).

The estimates vary in their quality, and some are based on formal local population studies. Federations that anticipate a more thorough community study have sometimes chosen not to revise outdated estimates until the study has been completed. The estimates for Toledo and Oklahoma City, for example, were substantially reduced in subsequent editions of the Year Book, suggesting that the 1980 figures were probably already too high at that time.

However, only in the case of the New York City area were the post-1980 estimates considered so significantly different that changes in the published 1980 figures were made for this atlas. The 1970 estimate for New York was repeated in each Year Book until 1982, when a formal study produced new counts for each of the five boroughs (counties) of the city and three nearby metropolitan counties (Chenkin 1982). For each of these areas, a corrected 1980 figure was calculated, assuming a linear trend between the 1970 and 1982 figures and interpolating to 1980.

Before being used, the data were also adjusted to exclude non-Jewish members of Jewish households. A national survey has estimated that the average 100 households defined as Jewish there are 7 or 8 people who do not consider themselves Jews, almost always non-Jewish spouses (Massarrk 1974). For this reason, the estimates of Jews presented in the Year Book were reduced by 7.5 percent.

The locations of the Jewish population were converted from the particular city and towns listed to the appropriate counties. In some cases the Year Book presented the distribution in terms of individual counties, but for many other places the distribution was expressed in terms of totals for groups of counties. Where total population data showed that one county clearly predominated in size, the estimated Jewish population was allocated to that largest county. Otherwise, the Jewish population was assigned to a county that was geographically in the center of the group of counties. For a dozen generally large metropolitan areas the Year Book provides estimates of the entire area's total Jewish population. Because there was no good basis for allocating the numbers to particular counties of residence, the estimated squares for central counties represent the much larger numbers of entire metropolitan areas. Suburban counties with large Jewish populations are frequently (but wrongly) identified as having few or no Jews. This problem meant that it would have been improper to calculate percentages of Jews in total county populations.

Because estimates of Jews in rural areas and small towns are frequently far out of date and generally less reliable, estimates of fewer than 100 Jews in a place have not been published in the Year Book. However, the appendix of the atlas includes unpublished estimates of some county Jewish populations of less than 100. This addition produced no significant change in the mapped pattern, but the unpublished data do indicate a greater dispersal of Jews than would otherwise be evident. Because of the less reliable nature of the small community estimates, in the appendix those counties with estimated Jewish populations of 50 to 100 are indicated only by an asterisk. However, state totals of Jews are the sum of all counties with 25 or more estimated Jews.

Processes of Map Construction

The distributions of ethnic populations in this atlas are presented on a base map, parts of which were originally created for the National Atlas of the United States of America (U.S. Geological Survey 1970). Existing plates from that atlas were revised and used for the basic layout, as well as for the political boundaries and water areas of the Great Lakes. The coastline and lettering on the full-page maps and all features of smaller maps were drawn specifically for We the People. The four reference maps showing county boundaries and names were adapted from the similar map in the National Atlas, with new names and boundaries in Alaska and Virginia appropriately updated as of 1980.

The computer and plotter generated the graduated symbols for maps of that type. Bureau of the Census data tapes ST4FIC and ST4F4B were linked with the PICADAD tape (Census 1978), which identified the longitude and latitude coordinates of the population centroids of all counties. These coordinates had to be modified, especially in Alaska, to take account of changes in county boundaries between 1970 and 1980. Next a program was written to convert the coordinates to the specific vertical and horizontal dimensions of centroid location on the Albers Equal-Area projection used in this atlas. The sizes and locations of graduated symbols could then be calculated by the computer and drawn by the Calcomp 1051 plotter. For choropleth maps, the computer was programmed to mark on the plot the location of counties in each percentage category to be mapped. These plotter-generated patterns were then placed beneath photomechanical masking film as a guide for identifying counties.

Design of Graduated-Square Maps

The dilemma of whether to show absolute numbers or the proportion of each ethnic group within a county's population was generally solved by incorporating aspects of both into each map. But because most ethnic groups represent very small percentages of each county's population, most maps were designed primarily to portray absolute numbers of ethnic populations by means of graduated square and diamond-shaped point symbols.

The area of the squares was made proportional to the ethnic population reported in each county with the scaling
shown in each map's legend. It is possible to measure the area of a square to determine the numbers of an ethnic population in a county. The scale of graduated symbols was designed to keep the squares small so as to avoid a great deal of overlap. This meant that counties with fewer than 50 persons reporting any ethnic group had to be shown by the two diamond types of point symbols. In contrast to the proportionately scaled graduated squares, each of the two diamond symbols represents a range of magnitudes.

Ethnic populations of fewer than 50 were not mapped. This low threshold for inclusion on maps was chosen in order to portray small ethnic populations effectively and provide a consistent treatment for groups that ranged greatly in size. Also, the widespread dispersal of most ethnic populations outside their areas of concentration can be more easily grasped when a low threshold is used. On the other hand, the large-sampling error for small populations argued for a higher threshold, perhaps 100 or more. However, comparison between different maps requires a consistent treatment and the choice of a percentage appropriate compromise.

Full-page maps were constructed for ethnic groups in which the race, Hispanic-origin population, or single-ancestry population totaled over 200,000 nationally. This large size permitted inclusion of county boundaries to assist in orientation. Quarter-page maps were used for ethnic groups with populations totaling between 20,000 and 200,000. Excepting the map of Jews, scaling of symbols within both of these size groups is consistent, making possible comparisons among maps of each type.

The major disadvantage of mapping the numbers of ethnic populations is that variations in their proportions in the total population are not normally shown. Yet it was important to identify those counties in which the ethnic population was especially large relative to the total county population; this would provide an indication of unusual ethnic intensity or strength. The percentages which each ethnic population represented in each county were calculated and ranked in order to determine those counties that fell in the highest 10 percent of these percentages. Counties with ethnic populations of fewer than 50 were excluded from these calculations in order to avoid sampling error and overemphasis of small populations. On the maps of each ethnic population those 10 percent of the counties with the highest percentages in the county population have been shown in a different color, providing a visual indication of an ethnic population numbered at least 500 each in a minimum of five counties.

Design of Choropleth Maps. The map of estimated numbers of Jews in counties was sized and designed distinctively to call attention to the non-census source of these data. Ethnic populations of this size (over 5 million) were normally mapped by the choropleth design, in which colors show percentages in total population. The sizes of the squares have been scaled differently from the other full-page distribution maps. The reader is reminded that the estimated number of Jews in many metropolitan areas is shown by a single graduated square that represents the Jews in several counties. Due to the same problem of unknown distributions within many metropolitan areas, the map of estimated Jewish population does not use a more intense tone or different color to distinguish areas with especially high percentages of Jews. For this reason, only a very few unaggregated counties were labeled on the map.

Design of Choropleth Maps. The largest ethnic groups (i.e., those of over 3,000,000 population) seemed an appropriate compromise. This was done because these populations were frequently so numerous and over extensive areas that graduated symbols overlapped to a confusing degree on important sections of the maps. On choropleth maps such as these the range of percentage values for all counties is divided into categories, each of which is indicated on the map by a different tone or color. For easier map interpretation, increasingly higher percentage values are shown by a logical sequence of color changes that visually suggests the progression. All counties must be assigned to a category, regardless of the size of the population and sampling error. However, among all the choropleth maps there is only one case in which a high percentage county with fewer than 50 members of the ethnic group had to be shown: Loving County, Texas, on the map of Irish ancestry.

Although the usual choropleth maps show variations in absolute intensity or proportion of ethnic populations, they indicate poorly the areas with large populations of an ethnic group where that group represents a small proportion of the total population. This is frequently the case in the largest cities and metropolitan areas. This problem was partly overcome by adding a set of 20 hollow graduated square symbols to indicate the counties (or aggregated county groupings) with the largest and smallest scale of symbolization had to be reduced from that used on the normal graduated symbol maps to avoid overlapping squares. Nevertheless, the scaling is consistent among these choropleth maps of large ethnic populations, meaning that the sizes of the squares on different maps can be compared. The choropleth mapping technique was particularly useful in handling what were likely weaknesses in the data on the Other-Spanish- and Mexican-origin populations. The visual impression conveyed by these maps accentuates areas of greater data quality. In the case of the map showing Mexican origin there was evidence of a substantial overcount in the Eastern Southeast (Census 1982c), and particularly the Southeast (Census 1982d), potentially misleading visual impression in that region was mostly corrected by eliminating one choropleth category near the bottom end of the scale.

Predominant Ethnic Population Map. This summary map (chapter 13) integrates the census counts of ethnic groups to identify the largest one in each county. Because neither complete-count nor sample data alone provided county figures for all groups, the two data sets were made approximately equivalent. However, there were national single-ancestry totals for all the major race and Hispanic-origin groups, and all county figures for each such complete-count group were reduced by the same proportion that the national single-ancestry population total represented of the national complete-total. For example, in the U.S. the Afro-American single-ancestry total was 77 percent of the total black population, and the Mexican single-ancestry total was 80 percent of the complete-count total. Thus, in each county the figure used for blacks was 77 percent of that county's complete count of blacks. Similarly, the county Mexican ethnic population was considered to be 80 percent of the listed Mexican-origin figure. Because the single-ancestry American Indian total was larger than the complete-count race total, the complete-count figures were used without modification. Also indicated on this map are the counties that are the most ethnically diverse. An entropy index was calculated for all 3,137 American counties or county equivalents. The index incorporates the same data used for maps of ethnic populations. However, only those 37 groups totaling over 200,000 and shown on full-page maps, plus Aleuts and Eskimos in Alaska, were included in the calculations. The highest value of the index occurs when all ethnic groups are equally represented; lowest values are found when many groups are not represented and one group constitutes most of the entire population (White 1986b). Here, the index ranged from 22 to 2.89. The counties in the highest 2 percent in this range (63 counties with entropy index values over 2.541) were identified on the map by a simple pattern.

Place Identification. Readers should be able to identify the places that are indicated by the map as significant. Yet, the inclusion of too many place-names clutters the map and obscures details of the patterns. To provide a general orientation to most maps in
Aggregation of Selected Counties

Areas of unusually high population density have often been aggregated for a clearer presentation on the maps. The five boroughs (or counties) of New York City have been combined so that graduated symbols and percentage values reflect the city as a whole. On the choropleth maps, the values for St. Louis City and the independent cities of Virginia were each assigned to an adjacent county, and the Washington, D.C., figures were incorporated into the Prince George's County, Maryland, totals, making possible larger areal units for better visibility on the maps. In Hawaii, Kauai was also aggregated with Maui County for this reason.

On a few maps the graduated symbols for counties in the area between Rhode Island and Washington, D.C., were often so crowded and overlapping that it was impossible to distinguish specific counties. For those populations with unusually concentrated settlement in parts of that area, the numbers for counties were combined to produce a clearer map. In each area indicated, the total ethnic population in all listed counties was allocated to the county shown in italics.

The following aggregations were made for graduated symbols on the maps: (1) Puerto Rican origin, Russian, Italian, Polish, Irish, and German ancestries; and estimated Jewish population:

1. Rhode Island: Providence County includes entire state.
2. Connecticut: Hartford County includes entire state.
3. New York area: New York County (Manhattan Borough) includes Kings, Queens, Nassau, Putnam, Richmond, Rockland, Suffolk, and Westchester counties.
5. Southern New Jersey: Camden County includes all New Jersey counties not listed in item 4.
6. Philadelphia area: Delaware County includes Bucks, Chester, Montgomery, and Philadelphia counties.
7. Baltimore area: Baltimore City includes Anne Arundel, Baltimore, Carroll, Harford, and Howard counties.

For Haitian and Romanian ancestries and Asian Indians, only the aggregations for New York and Northern New Jersey were made.

State and Metropolitan Area Cartograms: 1980 Data

These graphs represent populations as nested rectangles and squares, the areas of which are proportional to the sizes of the populations. Because the figures comprising the cartogram are not in their correct relative positions based on real-world distances and directions, the graphs are not maps. However, the cartogram does preserve some general aspects of relative location so that the reader can be aware of regional patterns. The various squares and rectangles were drawn by the computer plotter and then repositioned for photography.

Ancestries of State Populations

The cartogram of major ancestries is intended as a graphic summary of ethnic identities in America and represents an appropriate adaptation to the state data (Census 1983a).

Because the older European ancestry groups had low percentages of single-ancestry responses, multiple ancestries were incorporated in a special way. State totals for each ancestry group were produced by adding together the single-ancestry responses and one-third the multiple-ancestry responses. Simply adding the two response categories for each ancestry group, as has been done in some U.S. census publications, overweights the multiple-ancestry responses because those individuals have been in effect, counted twice. The division of the multiple-ancestry totals by one-third rather than one-half was done in recognition of the fact that an unknown number of multiple-ancestry responses listed three or more ancestries. The Bureau of the Census has had only the first two responses in its counts of multiple-ancestry numbers, except for a group of 17 unique triple-ancestry combinations. Individuals who reported any of these combinations were counted in each of the three ancestry groups they listed. The relative frequency of questionnaires with three of more ancestries listed can only be suggested by the large numbers reported for these few special combinations. Under these conditions of unknown dilution represented in the multiple-ancestry counts, division of those counts by three served to give an appropriately greater weight to the single-ancestry responses.

The sizes of squares representing the total ancestries of each state needed an additional adjustment because of the different ways of defining multiple ancestry in the data source—total of persons listing any multiple ancestry (Census 1983a). State totals of persons listing any multiple ancestry (table 1 of the above source) were many fewer than the total count of ancestries that were part of multiple-ancestry response (table 3b). For this reason, square sizes that included only one-third of the people giving any multiple-ancestry response could not accommodate the total of specific ancestries listed by those people. Because each person had to have made at least two ancestry responses, the person data in table 1 was doubled before multiplication by one-third. Thus, the sizes of squares were calculated as the sum of the people reporting a single ancestry and those reporting American or U.S. ancestry, plus two-thirds of the people reporting any multiple ancestry. In this way the 5.8 percent of Americans who gave their ancestry as American or U.S. have also been included as an ancestry group and appear on the cartogram where appropriate.

In each state only those ancestries that were proportionately significant were identified specifically or included in the aggregations. For the cartogram an ancestry had to represent at least 4 percent of the state's total to be shown specifically. In the appendix, if a specific ancestry was not one of the 20 largest (single plus one-third multiple), it was assigned to the category Other. Colorado provides an example. Because neither Finnish nor Puerto Rican ancestry was among the 20 leading groups, their numbers were included as Other rather than in the Northern European and Mexican-Puerto Rican totals.

Some ancestries have been combined: British includes English, Scottish, and Welsh ancestries. Northern European includes Danish, Norwegian, Swedish, and Finnish ancestries; and Asian includes Japanese, Chinese, Filipino, and Korean ancestries. The Mexican and Puerto Rican ancestry...
totals were combined, but Cubans were included only in Florida due to their small numbers outside Florida and the fact that they contrasted strongly with the Mexican and Puerto Rican populations in demographic, occupational, and income characteristics (Navarro and Tienda 1985). All Hispanic/Spanish ancestry has been of such numerical and cultural importance in the populations of New Mexico and Colorado that those responses have been added to the Mexican/Puerto Rican totals for those two states. For Alaska only, Eskimo and Aleut ancestries were included with American Indian ancestry.

The numerical results of these calculations for each state, as well as the percentage of each ancestry group in the total of all ancestries for each state, are provided in appendix 1. These figures represent a good approximation of the ethnic proportions in different states’ populations.

Minorities in Metropolitan Areas

The cartogram of metropolitan areas of over one million people focuses on the relative proportions in those places where the nonwhite and Hispanic-origin populations have been concentrated. The black minority population is not included separately. However, the category Other Asians included Vietnamese, Koreans, and Asian Indians. Also, Cubans have been combined with Other Spanish. The Hispanic-origin and minority-race groups are not mutually exclusive, in that some people who reported a Hispanic origin also reported black, Filipino, or possibly other races. Such individuals were included in both groups because the effect on the graphics was insignificant in nearly all cases. Where a group makes up less than 4 percent of the minority population within the group of counties designated the Bureau of the Census as a metropolitan statistical area (MSA), it is included in the category Other.

The sizes of squares representing the total minority population of the MSAs are proportional to the sum of the Hispanic-origin population and the total population after people of white race were subtracted. This results in a double counting of Hispanic-origin people who reported their race as Other (a large minority of that population) or any nonwhite race. For example, about 10 percent of the Puerto Ricans and Cubans in New York and Pennsylvania identified themselves as black (Census 1982d). However, because the cartogram is designed for comparisons of proportions of various minorities in different places, doubling the size of the total minority population is not a serious problem.

The STFC data were aggregated for the counties included in the metropolitan areas, because in the case of Boston the data for entire counties were used, the figure for Boston represents the New England county metropolitan area rather than the usual metropolitan statistical area.

Central and South American and Dominican Ancestries

The national origins of many people in the category of Other Spanish can be specified by this cartogram. The graph is based on the published single-ancestry data (Census 1983a) for 15 Latin American countries of predominantly Spanish language. People whose origin lay in Spain were not included because the small numbers reporting a Spanish ancestry suggested that many such people were not correctly identified (Fernandez and Cresce 1986).

The size of the rectangles and squares for each state is proportional to the state populations of the groups. However, the extreme contrast between states in population numbers and problems of sampling error made it inappropriate to show specific ancestry groups in states totaling fewer than 2,000 altogether. For larger states, ancestry groups that constitute between 4 and 7 percent of the state total are shown as small squares on top of the rectangles, but smaller groups have been combined into a general category.

American Indian Ethnic Groups: Nations/Tribes

Because most American Indians have a strong sense of identity with their nation or tribe, totals based on the fill-in portion of the race question were used for this cartogram (Census 1987a). These represent a sample of the complete-count data, but one that received adequate correct reporting and completeness. The size of each state’s square is proportional to the number of people who reported a specific group identity. Since many of those who identified themselves as Indian did not show a tribal or national affiliation, the totals in each state are smaller than the counts of Indians in STFC. However, the relative proportions of groups are probably well represented.

Colored identify the major linguistic families of the groups (Voegelin and Voegelin 1977; Spicer 1980). Specific groups have been identified in the figure for each state only if they constituted at least 200 individuals and 2 percent of the reported state total. Groups representing between 2 and 4 percent of the reported state total are shown by squares above the rectangles except for four cases in which groups of just over 3 percent could not be indicated because of insufficient space on the line. In North Carolina the 1,677 estimated Lummi Indians have been included with the Lumbees to compensate for apparent coding errors (Census 1985).

Maps of Net Migration

Census Data from the PUMS File

The maps of migration flows of selected race and Hispanic groups are based on data from the Public-Use Microdata Sample (PUMS) tape file (Census 1983a). This file is a record of the responses on over one-quarter of the long-form questionnaires and represents a 5-percent sample of the U.S. population. Data for all individuals in each sampled household are included, but there are no identifying names or addresses. Sampling errors associated with this use of the PUMS file are discussed in chapter 1.

In order to focus on the long-term redistribution of each population by internal migration, the state-of-birth data for U.S.-born individuals in selected race and Hispanic-origin groups were used. The District of Columbia was treated as a state. The PUMS file indicates the state of birth for all sampled individuals residing in each state. The gross migration between any two states is the total number of people born in one state but living in the other in 1980. The net migration is the difference between the numbers migrating in one direction from those migrating in the opposite direction. For each ethnic group mapped, a 51 × 51 matrix was developed, showing gross and net migrations between each pair of states.

Many of the people represented in these data have migrated between states numerous times. Some of these were not even counted as migrants at all because they returned to their state of birth by 1980 after decades elsewhere. Thus, the data indicate not rates of migration over time but net shifts of population.

Refugee Migration Data

Because there were so few U.S.-born Vietnamese or other Southeast Asians in the United States in 1980, the census contained no data regarding them comparable to the data used for the net migration maps of other populations. Moreover, because the large immigration from Southeast Asia began in May of 1975, the census question concerning place of residence on April 1, 1975, could not be used to identify their interstate migrations. Nevertheless, it is known that Southeast Asians have been redistributing themselves rapidly after their initial settlement, and the recent influx into America has continued and most Southeast Asians who arrived in the early 1980s were at first given refugee rather than immigrant status, data on interstate refugee migrations could approximate the more recent population shifts.

The data on refugee migration is based on reports by state agencies providing social services to refugees, usually only during their first three years in this country. Because in 1984 about 60 percent of the Vietnamese and Laotians and 49 percent of the Cambodians who were eligible were receiving at least some assistance, the data represent almost 60 percent of recently arriving refugees from Southeast Asia (U.S. Office of Refugee Resettlement 1985b). Refugees who were receiving assistance or services as of June 30, 1984, were identified
by their social security numbers, which show by their first three digits the state in which the person was living when he or she originally applied for the card. Tabulation of these states of origin compared to the states in which people were living when they received assistance produces a gross migration data matrix, from which net flows can be calculated.

These data include some refugees from regions other than Southeast Asia, but in 1984 Southeast Asians comprised 83 percent of all refugees requesting assistance (U.S. Office of Refugee Resettlement 1985b). The approximately 40 percent of eligible refugees not receiving government assistance may have had somewhat different net migration patterns and perhaps lower rates of migration. In particular, it is likely that California, with its very high rate of cash assistance (85 percent) among eligible refugees, is overestimated as a destination in these particular data. However, there is no reason to question its role as by far the most popular destination for refugees making secondary migrations.

Design of Net Migration Maps

The maps of net interstate lifetime migration identify major population flows that partially explain changes in distribution over the last several decades. The various ancestry groups within the white population were thought to be generally similar in their choice of California and Florida as most typical destinations, although their states of origin would differ just as the 1920 distributions differed. For this reason, and for economy of treatment, there are no maps of net migrations of ancestry groups. The larger race and Hispanic-origin groups were mapped because greater contrasts in patterns were expected among them. For a few groups the maps provide a visual link between the 1920 and the 1980 distributions.

Arrows on the map vary in width according to the importance of each net interstate migration relative to the U.S.-born population of the group in the United States. (People born in Puerto Rico are considered part of the U.S.-born population.) The maps can be compared to each other despite great differences in absolute numbers of migrants because all arrow widths are based on the percentage of that base population represented in the specific net flow between any two states. Arrow bases and points may be positioned at any place inside the boundaries of a state and should not be interpreted as indicating anything about intrastate regions of origin or destination.

Since people of all ages were part of the sample, the maps are best used as indicating generalized patterns over the last half century. To keep maps uncluttered and avoid problems of sampling error with smaller populations, only the largest net interstate flows of each group were included. Omission of a particular flow means only that it was below the threshold chosen for that map; it does not mean that the migration was small or insignificant.

State Maps: 1920 Data

Small maps show the distribution of certain ethnic groups as of 1920. These have been designed to complement the text interpretation of the 1980 map patterns. The data and design for the 1920 and 1980 maps make them not directly comparable. However, general similarities in the 1920 and 1980 distributions indicate visually the importance of earlier settlements in any explanation of 1980 patterns. For some ethnic groups the geographical stability in the older areas of settlement is pronounced, while for others there have been major changes in the last 60 years.

In a few cases the 1920 maps identify ethnic populations that had to be aggregated with others in the 1980 data. For example, the contrasting distributions of the French Canadian foreign stock and the foreign stock whose country of origin was France are evident. To the extent that these have persisted from 1920 to 1980 they aid in the interpretation of the map of French ancestry. Also, the 1920 map of people whose mother tongue was Flemish suggests the areas where that portion of the Belgian-ancestry population was probably strong in 1980.

U.S. Census Data

All the 1920 census data mapped in this atlas represent an attempted complete count of the population at their usual place of residence. This was different from the procedures for the 1980 census, in which only a sample of the population received the ancestry question. The earlier census was conducted over several weeks beginning in January, 1920, and enumerators were supposed to visit each place of residence to obtain answers to the questions. Communication with people who had little knowledge of English was a probable source of error in addition to undercounting, but these data are some of the best for the early 20th century (Census 1922a: 31–76, 891–1037; Census 1922b: 1158–1193).

Maps of blacks, Chinese, Japanese, and American Indians make use of data on race groups. In the 1920 census the enumerators identified a person's race primarily by observation. People who appeared to be a mixture of white and some nonwhite racial heritage were counted in that nonwhite group. People who seemed to represent a mixture of two nonwhite races were assigned according to the larger group in the neighborhood. The data thus include only those people who were visibly recognizable as Indian or living in an Indian society. People of Mexican origin were counted as white unless they were clearly Indian or in some other nonwhite group.

Because so many European peoples had arrived in the United States in the 1850–1920 period, it is appropriate to use data population. This data on the foreign-stock population included all people who were born outside the United States and American-born children of these immigrants. These numbers were tabulated only for whites. People who had been born in the U.S. but had foreign-born parents from two different countries or speaking two different languages were not included in these data.

Some ethnic data pertain only to the foreign-born population, but these are generally of less value than the data on foreign stock. Including the second generation more than doubles the size of ethnic populations and covers more effectively those older communities where more people arrived several decades earlier.

Because ancestral language is usually a better indicator of ethnic identity than country of birth, the most useful data concern the mother tongue of people of white foreign stock. A mother tongue was defined as the language commonly used in a person's home before he or she immigrated. The names of certain languages were frequently confused and misreported that census officials decided to avoid errors by aggregating data on Dutch and Frisian, Lithuanian and Latvian (Lettish), Yiddish and Hebrew, and Syriac and Arabic mother tongues. The mother tongue data are, however, not as definitive as might be thought because the Eastern European mother tongue populations include substantial numbers of Jews. Although most Jews whose origin was in that area used Yiddish as their mother tongue, many also knew and used the language of the local Christian population. Because this had a known important effect on the Russian and Romanian mother tongue data, no 1920 maps for these groups were included.

In cases wherein the mother tongue does not identify an ethnic group effectively, the country-of-origin data are much more useful. This is particularly important in differentiating the several English-speaking ethnic groups. American-born whites with both parents born in the same foreign country or with only one parent born in a foreign country were included as white foreign stock with origin in that country. In addition, for a foreign-born person, the country of birth of that person's father was reported as the country of origin. Although this definition was inconsistent with previous censuses, it should have the advantage of better identifying some origins. It acknowledged, for example, the fact that Welsh, Irish, or Scottish identity was retained by English-born immigrants to America long after their fathers had migrated from Wales, Ireland, or Scotland to England.
Data on Jews
The one map regarding 1920 data not based on the 1920 census is the map of Jewish numbers, which is based on estimates made by local Jewish community leaders and published in the American Jewish Year Book. U.S. census data on the people whose mother tongue was Yiddish or Hebrew are available. However, well those data might locate the Eastern European Jewish population, they fail to indicate those German Jews and Sephardic Jews who arrived much earlier. Also, the majority of Jews who were not officially members of synagogues or temples were not counted in the U.S. census counts of religious organizations and membership made around that time. Thus, the private Year Book estimates are more complete than any other data source for that time period (Linfield 1924).

Design of Maps
These maps were designed to aid comparisons between ethnic groups regarding relative distributions. Ethnic groups that were more highly concentrated are indicated if the map shows from one to a few states with over 10 percent of the group's U.S. population and not many others in the 2- to 10-percent categories. Maps were kept simple by showing only those 10 cities with the largest populations of each group, but the three sizes of dot symbols provide some indication of relative ethnic population numbers.

Most maps portray the 48 states of the United States in 1920. However, some ethnic populations are shown on larger-scale maps of just the northeastern quarter of the country because no state outside that area had more than 2 percent of the group's national population or any of its ten largest urban concentrations. The map of American Indians includes Alaska, and the maps of the Chinese, Japanese, and Portuguese include Hawaii because these groups had especially large numbers in those territories. Because nearly 80 percent of the 27,454 Filipinos counted were in Hawaii, no map was made for that population.

The 1920 maps indicate the specific cities of concentration at a time when different immigrant groups were often distinctively located in certain cities or types of cities. The city and its neighborhoods were the bases for identifying locations at that time. Also, because the pattern of the 1920 city concentrations is often similar to the 1980 county concentrations of the same ethnic group, the more easily recognized city names are a useful identifying label.