UC Berkeley
Earlier Faculty Research

Title

Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area

Permalink

https://escholarship.org/uc/item/1839w13f

Authors

Miller, Vincent P.
Quigley, John M.

Publication Date

1990-03-01

eScholarship.org Powered by the California Digital Library
University of California

Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area

Vincent P. Miller
John M. Quigley

March 1990
Reprint, No. 14

The University of California Transportation Center

University of California
Berkeley, CA 94720

The University of California Transportation Center

The University of California Transportation Center (UCTC) is one of ten regional units mandated by Congress and established in Fall 1988 to support research, education, and training in surface transportation. The UC Center serves federal Region IX and is supported by matching grants from the U.S. Department of Transportation, the California State Department of Transportation (Caltrans), and the University.

Based on the Berkeley Campus, UCTC draws upon existing capabilities and resources of the Institutes of Transportation Studies at Berkeley, Davis, and Irvine; the Institute of Urban and Regional Development at Berkeley; the Graduate School of Architecture and Urban Planning at Los Angeles; and several academic departments at the Berkeley, Davis, Irvine, and Los Angeles campuses. Faculty and students on other University of California campuses may participate in Center activities. Researchers at other universities within the region also have opportunities to collaborate on selected studies. Currently faculty at California State University, Long Beach, and at Arizona State University, Tempe, are active participants.

UCTC's educational and research programs are focused on strategic planning for improving metropolitan accessibility, with emphasis on the special conditions in Region IX. Particular attention is directed to strategies for using transportation as an instrument of economic development, while also accommodating to the region's persistent expansion and while maintaining and enhancing the quality of life there.

The Center distributes reports on its research in working papers, monographs, and in reprints of published articles. For a list of publications in print, write to the address below.

University of California Transportation Center

108 Naval Architecture Building
Berkeley, California 94720
Tel: 415/643-7378
FAX: 415/643-5456

Authors of papers reporting on UCTC-sponsored research are solely responsible for their content. This research was supported by the U.S. Department of Transportation and the California State Department of Transportation, neither of which assumes liability for its content or use.

Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area

Vincent P. Miller
John M. Quigley

Department of Economics
Graduate School of Public Policy
University of California at Berkeley

Reprint, No. 14

Urban Studies
Vol. 27, No. 1, pp. 3-21, 1990

The University of California Transportation Center
University of California at Berkeley

Urban Studies, Vol. 27, No. 1, 1990 3–21

Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area

Vincent P. Miller and John M. Quigley

[Paper first received, November 1988; in final form, May 1989]

Summary. This paper considers residential segregation by race and by type of household in 1970 and 1980. The paper presents entropy indices of segregation for the San Francisco Bay Area and its five metropolitan areas. The methodology permits an investigation of the effects of group definition upon segregation measures, and an analysis of the degree of independence in the segregation of households by race and demographic group. The results indicate that the levels of segregation by race and by household type have declined modestly during the 1970s, at least in this region. More importantly, however, the results indicate a remarkable independence in the spatial distribution of households by race and demographic group. Only a very small fraction of the observed levels of segregation by race could be 'explained' by the prior partitioning of households by demographic group. The principal results of the analysis are invariant to changes in the definition of racial or household groups.

I. Introduction

Even the most casual observer notices that residential patterns in American urban areas are highly segregated by race. It is only slightly less obvious that urban areas are segregated by income, by household size and composition, and by other demographic characteristics. Presumably, residential segregation by socio-demographic group reflects a similarity of tastes for local public goods and locational amenities and a similarity in disposable income. Residential segregation by race may reflect the same phenomenon. It may also reflect the outcomes of a discriminatory market in which minority households are denied access to the entire housing stock or in which minority households feel less threatened by choosing to reside in close proximity.

Disentangling 'natural' segregation by socio-demographic group from that which arises from racial discrimination is no easy task. Yet the distinction is important in interpreting trends in segregation and in promoting equal opportunity. If, for example, levels of housing market discrimination have *declined* while demographic differences among races have increased, the observable result may be *increased* tendencies towards spatial segregation by race. This situation would imply that the determinants of housing segregation are increasingly rooted in economic differences among households and not in the resistance of actors in the real estate market to the granting of equal access to rental and sales markets.[1]

This issue of interpretation arises in

Vincent P. Miller and John M. Quigley are at the Department of Economics and Graduate School of Public Policy, University of California, Berkeley, California 04720, USA. *This research was supported by the Center for Real Estate and Urban Economics and by the Transportation Center, University of California, Berkeley.*

other contexts. For example, in the decade between the 1970 and 1980 census, the poverty rate for US black households declined very modestly from 29.5 per cent to 28.9 per cent. Underlying this persistently high average poverty rate were two offsetting trends. The first was the increase in the income of two-adult black households and intact families. The second was the increase in the numbers of single-adult households, especially those headed by single mothers with limited income opportunities. In fact, if the composition of black households in 1980 had been the same as in 1970, then the decline in poverty rates by household type would have led to a 9.6 percentage point decline in the black poverty rate (to 19.9 per cent), instead of the 0.6 percentage point decline actually observed (Green and Welniak, 1982).

These developments in household composition have spatial consequences. It is important to examine the extent to which racial segregation reflects reductions in the residential segregation of household types counteracted by shifts in household composition towards single-earner (or no-earner) households who are more likely to be in poverty and thus to have fewer options for housing and location.

Unfortunately most empirical studies of residential segregation have ignored these distinctions among discrete demographic groups in focusing attention on the occupancy patterns of two racial groups. Even at this aggregate level, however, similar problems in interpretation arise. For example, reductions in the level of racial segregation between blacks and non-blacks reported in the 1970s may reflect, to an unknown extent, the increasing spatial integration of blacks with other minorities (especially hispanics) combined with the increasing segregation of white and minority households (Massey, 1979).

These problems of interpretation arise because analyses of segregation are often based empirically upon a binary representation of residential location—black–non-black or white–non-white—by census tract or urban neighbourhood.[2]

In this paper we consider the decomposition of residential segregation by several distinct household types and races, as well as by location. The paper begins with a careful definition of residential segregation and a cursory comparison of some common measures of the phenomenon, and then presents an analysis of segregation patterns in the San Francisco Bay Area as indicated by the 1970 and 1980 census reports.

We focus on the San Francisco area for several reasons. First, it is a large and economically diverse metropolitan area. Segregation patterns observed are thus more likely to be comparable with those of other large and diverse metropolitan areas. Secondly, the region is among the fastest growing in the United States, and demographic trends are starkly visible. Thirdly, the region has several large and well-defined ethnic populations, including black, hispanic and Asian Americans. Finally, the spatial character of the region is well defined. Like most metropolitan areas in the western United States, its development is relatively recent. However, unlike most others, the density and housing price gradients reveal a monotonic decline from a central business district. Section II below introduces the methodological issues in the comparisons. Sections III and IV introduce the data and the comparative measures employed. Section V presents an extensive analysis of segregation by race, demographic group and location in the region.

II. Segregation

Residential segregation can refer to both a process and an outcome. This paper is concerned with measures of the outcome.

A. Definitions

Consider households' choices to reside at various points in an urban area, where conflicting choices are resolved by some impersonal mechanism. Now, if the conditional distribution of households by race in space differs from the unconditional distribution, the population may be said to be segregated by race. Of course, if conflicting choices are resolved by a price mechanism, then differences in income or wealth among races may lead to this segregation. Also, in these circumstances systematic differences by race in other factors which affect preferences for location, for example, family size and household composition, can cause some degree of residential segregation by race.

This definition of segregation is scale invariant (Allison, 1978); it is independent of the number of racial groups involved in the analysis, the size of the total population and the overall distribution of area households among races. Scale invariance permits direct comparisons of segregation to be made for differing geographical areas or time periods. Of course, scale invariance may not be an appropriate property for all purposes. For example, suppose segregation was of concern principally because it inhibited social interaction among people of different races, and that the likelihood of interaction was a function of the probability that one saw members of another racial group. Then a doubling of population density would lead to a reduction in the impact of segregation (since it would raise the expected number of individuals of a different race within a fixed distance of any household).[3]

The spatial implication of these definitions is that, in the absence of segregation, the distribution of households by race will be similar in each of the subdivisions of a larger area. This suggests that the *presence* of segregation can be tested by partitioning the area to be analysed into subareas and by examining the racial composition of each. For example, in the multiracial context, the homogeneity of several multinomial distributions can be tested by a simple Chi-squared comparison.[4] Alternatively, measures of the *degree* of segregation may be constructed from variations in racial composition across subareas. Two problems arise with this approach to the analysis of either the presence or the degree of segregation. One is that, as the size of the subareas increases, the same physical area appears less segregated. At the limit, when the subarea subsumes everything, the metropolitan area must be 'integrated'. The second is that the way in which the area is partitioned can affect conclusions about the presence of segregation or the degree of segregation. For example, a checker-board pattern of residential occupancy by race can give rise to extreme differences in residential occupancy by subarea or to identical measures of the racial composition of subareas, depending only upon how the checker-board is partitioned.

Despite the potential importance of this partitioning problem, any empirical analysis of patterns of US residential segregation must ultimately begin with counts of individuals or households by predetermined geographical areas: census tracts or perhaps block faces.[5] Census tracts were established to have stable boundaries, and were "designed to be relatively homogeneous areas with respect to population characteristics, economic status and living conditions" (US Bureau of the Census, 1982, p. 8). Any measure of segregation is conditional upon the prior partitioning of the urban area into these geographical subareas.

B. Measurement

The empirical analysis in this paper relies principally upon the entropy measure of the degree of residential segregation. This measure is suited to the problem of analysing segregation jointly by race and demographic group. The discussion begins with a cursory review of other, more common measures of segregation.[6]

As noted above, any quantitative measure of segregation must begin with counts of households by racial or other group residing in subareas (census tracts).

Let n_{it} be the number of individuals of group i residing in census tract t. Thus $n_{*t}=\Sigma_i n_{it}$ is the total number of individuals residing in t, and $n_{i*}=\Sigma_t n_{it}$ is the total number of individuals of type i in the entire area. Finally, $n_{**}=\Sigma_i n_{i*}=\Sigma_t n_{*t}$ is the area population.

The most common quantitative measure of the level of residential segregation computed from these counts of individuals by census tract is the so-called 'dissimilarity index' popularised by Taeuber and Taeuber (1965).

The dissimilarity of index measures the level of segregation between two groups. Suppose each neighbourhood t is composed of n_{1t} and n_{2t} individuals of group 1 and 2 respectively. The dissimilarity index D,

$$D=\frac{\Sigma_t n_{*t}|(n_{1t}/n_{*t})-(n_{1*}/n_{**})|}{2n_{**}(n_{1*}/n_{**})(n_{2*}/n_{**})}, \qquad (1)$$

is the normalised sum of the absolute deviations of the racial proportions of census tracts from the overall racial proportion.

The index does have an appealing intuitive interpretation; its value represents the minimum proportion of the population that would have to relocate to eliminate segregation. The traditional measure is seriously deficient, however, on other grounds. First, its properties violate the common-sense principle of transfers (Allison, 1978), namely, that an index of segregation should decrease in value when members of a group move from an area of higher group concentration to one of lower concentration.[7] Secondly, the index is not well defined when there are more than two groups. Increasingly, US metropolitan areas are characterised by several identifiable minority groups, and the dissimilarity index is deficient in representing that heterogeneity.[8]

An alternative index measures the 'exposure' of one group of residents to others. The exposure index is the weighted average proportion of agents in each area who are not members of the same group. The exposure of any population group i, to all other groups, $\bar{i}$, is defined as

$$E_{i\bar{i}}=(1/n_{i*})\Sigma_t n_{*t}(n_{it}/n_{*t})(1-n_{it}/n_{*t}) \qquad (2)$$

In contrast to the dissimilarity index, the value of the exposure index does depend upon the distribution of population groups within each subarea. Moreover, the exposure index can be decomposed into a weighted average of the exposure of members of group j to each of the other subgroups:

$$E_{i\bar{i}}=\Sigma_{k\neq i}(n_{k*}/n_{**})E_{ik} \qquad (3)$$

The exposure index cannot, however, be decomposed spatially or geographically.

In contrast, the entropy index (H) is the only measure of segregation which satisfies the properties of symmetry, continuity and full additivity. The entropy index is defined as

$$H=\Sigma_t\Sigma_i(n_{it}/n_{*t})\log(n_{*t}/n_{it}) \qquad (4)$$

The principal advantage of the entropy index in representing the segregation of households by household type, race and location is illustrated in the analysis that follows.[9]

III. The Data

As noted above, this analysis of spatial segregation is based upon data from the San Francisco Bay Area (The 'San Francisco–San Jose–Oakland Consolidated Metropolitan Statistical Area'), which includes nine counties and five Standard Metropolitan Statistical Areas (SMSAs). The analysis is based upon census tract data for 1970 and 1980 which refer to 1079 census tracts (according to 1970 boundaries).[10] Figure 1 presents, in schematic terms, the five SMSAs which make up the Bay Area, the central cities of each SMSA,[11] and the census tracts which form the ultimate building blocks for the analysis.

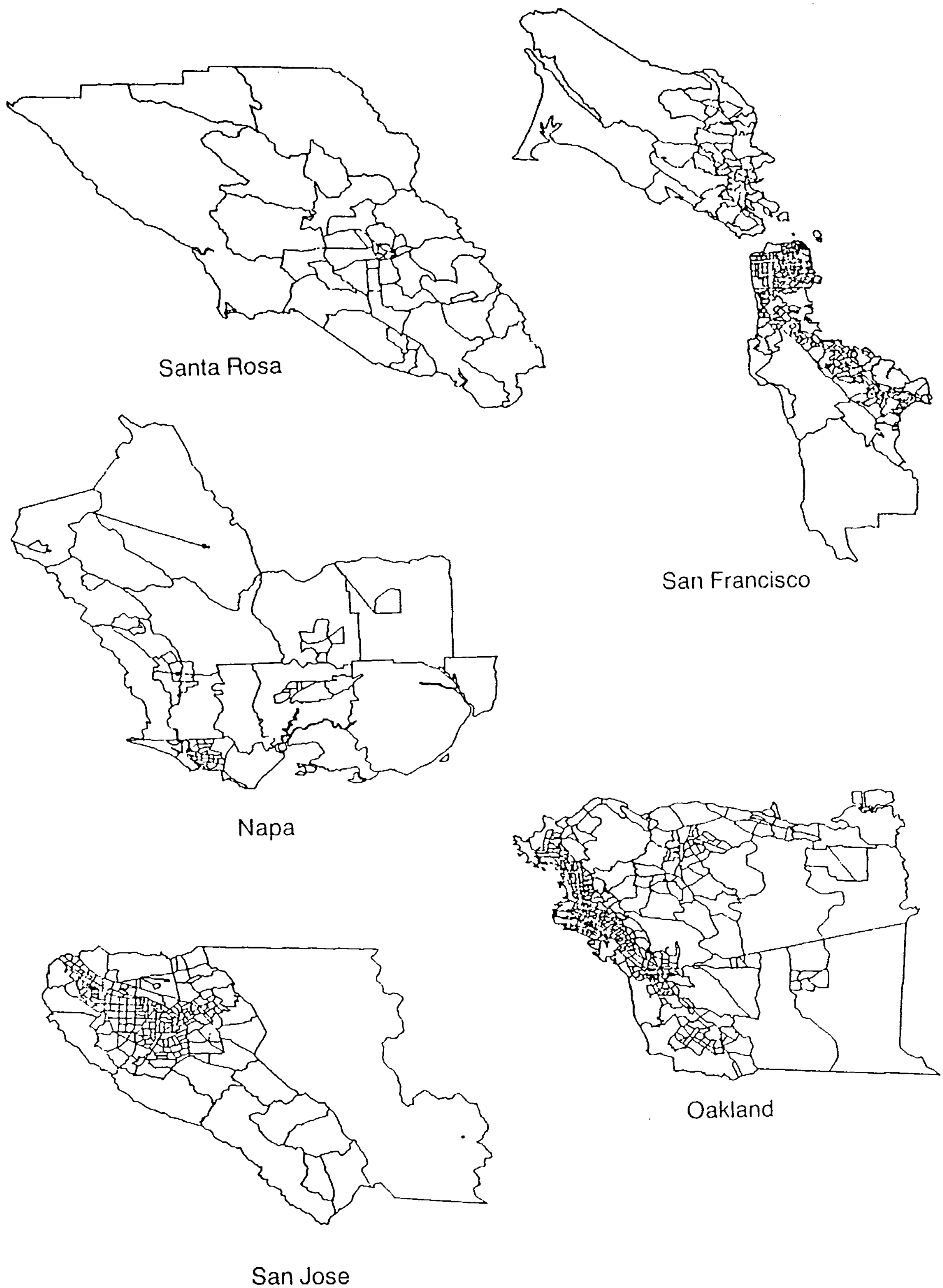

Figure 1. Census tracts and metropolitan areas comprising the San Francisco Bay Area (1970 definitions of census tracts).

Table 1. Race and hispanic origin, San Francisco Bay Area, 1970–80 (individuals)

Race	1980					1970	
	Reported origin			Revised			
	Hispanic	Non-hispanic	Total	Total	%	Total	%
Hispanic	—	—	—	632 650	12.2	587 503[a]	12.8
Black	7 474	458 800	466 274	458 800	8.9	365 893	7.9
Asian	0	462 890	462 890	462 890	8.9		
White	334 255	3 605 829	3 940 084	3 605 829	69.6	3 641 962[b]	79.3
Native American	32 025	5 162	37 187	5 162	0.1		
Other	258 896	14 453	277 349	14 453	0.3		
Total	632 650	4 547 134	5 179 784	5 179 784	100.0	4 595 358	100.0

[a] Spanish-Americans.
[b] Non-black, non-Spanish-Americans.
Source: Calculations by authors from 1980 Census STF3, and from 1970 Census of Housing, Fourth Count Summary.

The empirical analysis considers the segregation of households by race and household type as well as location. Census data for race and household type are available in more detail for 1980 than for 1970. For both years, however, race and hispanic origin were assessed in separate questions. For our analysis we have created separate hispanic racial groups in both census years by reallocating into a separate category all persons of any race who reported hispanic origin. The result for 1980 is a mutually exclusive six category race–ethnicity classification comparable to, but more detailed than, the three category classification generated for 1970. The classification for 1970 includes blacks, Spanish-Americans and all others (Massey and Denton, 1987).

Table 1 reports the racial composition of the Bay Area for 1970 and 1980 in these mutually exclusive categories. The table also reports the raw data for 1980 from which these totals were calculated, as well as the less detailed racial composition reported in the 1970 census. For the nine county regions as a whole, about 70 per cent of the population is classified as white and about 12 per cent as of hispanic origin. About 9 per cent of the population is classified as black, and a similar proportion is classified as Asian. During the decade of the 1970s, the total population of the region increased by almost 13 per cent. The hispanic population increased by 45 000, and the black population, by about 100 000.

In contrast, the non-black non-hispanic population declined by 36 000 people. One suspects that the Asian population increased substantially during this period, but census data provide no evidence on this.

The classification of population into household types is less problematic. According to US census conventions, the population is counted by family and by household. Families are defined on the basis of relationships; households, on the basis of living quarters.[12] Households are of two basic types. Family households include two or more related persons living together. Non-family households are persons living alone or sharing living quarters with persons to whom they are not related.[13]

Average household size in the Bay Area is 2.6 persons, and in 1980 97.7 per cent of the population resided in households. Table 2 presents data for 1980 on the distribution of Bay Area households by the six races defined above and by seven major types of household. These types include traditional husband–wife families with and without children; single adults living alone, by sex; single-parent households, by sex; and non-family households containing two or more adults.[14] Note that Asian, hispanic, and 'other' households are far more likely to involve married couples with children than is true for white, black, or native American households. Also, black households are three times more likely to be made up of an unmarried female head with children than is the case for other groups. Among households with children, 45 per cent of black households are headed by single women, compared to 16 per cent for all other groups. Twenty-seven per cent of all the households in the Bay Area are white, non-family households. Only 22 per cent of all households are white married couples with children. Married couples of all races with children account for only 27 per cent of Bay Area households.

Less detail about household types is available from the 1970 census. In particular, this census did not distinguish between families with children and those without children. Households were recorded in only four categories: families with married couples, those headed by unmarried males or females, and non-family groups. Table 3 presents the comparable race and household type information available for the Bay Area for 1970 and 1980. For each of these four household types, counts are available separately for black, hispanic and all other (i.e. non-black, non-hispanic)

Table 2. Household type by race, Bay Area, 1980: 42 race–household categories (households)

	White		Black		Native		Asian		Hispanic		Other		Total	
Household type	Number	Prop.	Number	Prop.	Number	Prop.	Number	Prop.	Number	Prop.	Number	Prop.	Number	Prop.
Family														
Married couple														
With children	331 493	0.16	28 834	0.01	2 314	0.00	44 208	0.02	70 110	0.04	31 657	0.02	508 616	0.26
No children	427 324	0.22	25 863	0.01	1 582	0.00	29 979	0.02	36 530	0.02	20 243	0.01	541 521	0.27
Male householder (unmarried)														
With children	16 900	0.01	5 612	0.00	2 623	0.00	3 592	0.00	3 574	0.00	3 757	0.00	36 058	0.01
No children	19 821	0.01	3 410	0.00	161	0.00	3 505	0.00	4 209	0.00	1 715	0.00	32 821	0.02
Female householder (unmarried)														
With children	66 317	0.03	27 706	0.01	1 116	0.00	4 628	0.00	17 346	0.01	6 861	0.00	123 974	0.06
No children	45 004	0.02	9 679	0.00	397	0.00	4 914	0.00	7 937	0.00	2 973	0.00	70 904	0.04
Non-family	524 036	0.27	53 845	0.03	2 786	0.00	29 160	0.01	41 741	0.02	21 184	0.01	672 752	0.34
Total	1 430 895		154 949		10 979		119 986		181 447		88 390		1 986 646	
Proportion		0.72		0.08		0.01		0.06		0.09		0.05		1.00

Source: Calculations by authors from 1980 census, STF3.
Note: Male and female 'householder' classes may include other adults.

households. The table presents a comparison of these 12 categories as reported in the 1970 and 1980 census.

Between 1970 and 1980 the number of black households increased by 39 per cent, and the number of hispanic households increased by 16 per cent. Despite these increases, however, the number of black households consisting of married couples increased by only 4 per cent, and the number of hispanic households consisting of married couples actually declined by 7 per cent. The largest comparable race household category in 1970 and in 1980 is other (non-black, non-hispanic) married couples, but the fraction of Bay Area households comprising this category fell from 53 per cent to 45 per cent during the decade.

IV. Residential Segregation by Demographic and Racial Group[15]

Households in the Bay Area are partitioned into race categories and household types, and into a variety of spatial groupings (for example, central city or suburban location) in each of the five SMSAs.

Again, let the subscript i denote the category of household, race or household type, let t denote an index of the geographical areas (census tracts) included in the area, and let M_s be the set of census tracts in some spatial aggregation s, such as a central city or a metropolitan area. Thus,

$$\rho_{it} = n_{it}/n_{*t} \tag{5}$$

is the number of households of category i as a fraction of all households in census tract t, and

$$\omega_t = n_{*t}/n_{**} \tag{6}$$

is the number of households in census tract t as a fraction of all households. Similarly,

$$W_s = \Sigma_{t \in M_s} \omega_t \tag{7}$$

is the number of households in aggregation s (a central city or a metropolitan area, for example) as a fraction of area total. Following (4), the entropy at the level of census tract is

$$H_t = \Sigma_i \rho_{it} \log(1/\rho_{it}), \tag{8}$$

and the entropy for some aggregate level s is

$$H_s = \Sigma \rho_{is} \log(1/\rho_{is}), \tag{9}$$

where ρ_{is} is the proportion of population of type i in region s

$$\rho_{is} = \Sigma_{t \in M_s} (\omega_t/W_s)\rho_{it}. \tag{10}$$

The average entropy of the census tracts in region s is

$$\bar{H}_s = \Sigma_{t \in M_s} (\omega_t/W_s) H_t. \tag{11}$$

Note that the entropy measure in (11) is a simple linear combination of the entropy at the level of the census tract. Clearly, for any number of household categories, I, entropy is maximised when each of the underlying probabilities in (8) is equal to $(1/I)$. It follows that the maximum entropy of any region depends upon the aggregate distribution of population among each of the categories.[16] Less obvious is the fact that the region's entropy, H_s, is the maximum for the average entropy of the census tracts in that region. Thus,

$$Z_s = (H_s - \bar{H}_s)/H_s \tag{12}$$

measures the relative reduction in entropy arising from the spatial segregation of household types in the entire region or in any aggregation s.

The previous discussion deals with classifications in one dimension, say racial categories. The extension to the bivariate case—the joint distribution of race, r, and household type, h—is straightforward. As before,

$$\rho_{rht} = n_{rht}/n_{**t} \tag{13}$$

is the number of households of race r and housing type h as a fraction of all households in census tract t. The probabilities of the two marginal distributions are

$$\begin{aligned} \rho_{r*t} &= \Sigma_h \rho_{rht} \\ \rho_{*ht} &= \Sigma_r \rho_{rht}, \end{aligned} \tag{14}$$

and the entropies of these distributions are

Table 3. Household type by race, Bay Area, 1970 and 1980: 12 race–household categories (households)

House type	All other		Black		Hispanic		Total	
	Number	Prop.	Number	Prop.	Number	Prop.	Number	Prop.
1980								
Family								
Married couple	888 738	0.45	54 673	0.03	106 640	0.05	1 050 051	0.53
Male householder (unmarried)	42 015	0.02	6 497	0.00	7 783	0.00	56 295	0.03
Female householder (unmarried)	131 964	0.07	37 333	0.02	25 283	0.01	194 580	0.10
Non-family	577 116	0.29	53 822	0.03	41 741	0.02	672 679	0.34
Total	1 639 832	0.83	152 326	0.08	181 447	0.09	1 973 605	1.00
1970								
Family								
Married couple	817 030	0.53	50 088	0.03	117 437	0.08	984 546	0.64
Male householder (unmarried)	25 917	0.02	2 838	0.00	3 932	0.00	32 668	0.02
Female householder (unmarried)	93 195	0.06	21 940	0.01	15 598	0.01	130 710	0.08
Non-family	344 742	0.22	29 168	0.02	23 297	0.02	397 175	0.26
Total	1 280 883	0.83	104 034	0.07	160 264	0.10	1 545 099	1.00

Source: Calculated by the authors from 1980 census, STF3, and from 1970 Census of Housing, Fourth Count Summary.

Table 4. Indices of residential segregation; percentages by race and household type, San Francisco Bay Area, 1980

	Race		Household type	
	Six groups	Three groups	Seven groups	Four groups
Entire Region	22.43	26.58	8.19	9.32
SMSAs				
Oakland	25.16	31.43	8.49	9.87
San Francisco	23.22	24.52	8.53	9.90
San Jose	12.06	14.38	6.36	6.97
Santa Rosa	8.73	4.21	2.94	2.80
Napa	13.25	13.67	5.16	4.86
Central City/Suburbs				
Oakland				
Central city	21.28	23.69	7.54	8.37
Suburbs	18.79	23.26	7.40	8.45
San Francisco				
Central city	20.32	25.34	8.33	9.22
Suburbs	22.66	20.13	5.50	6.27
San Jose				
Central city	11.80	13.87	5.93	6.39
Suburbs	9.42	11.06	6.12	7.00
Santa Rosa				
Central city	12.68	4.92	4.68	4.69
Suburbs	4.64	3.39	1.53	1.29
Napa				
Central city	12.45	13.13	5.56	5.51
Suburbs	8.29	6.67	3.48	2.20

Note: Table entries are $(H_s - \bar{H}_s)/H_s$ where H_s, is the maximum entropy possible in each geographical region and $\bar{H}_s$ is the average entropy computed from the census tracts in that region. Values of H appear in Table 5.

$$
\begin{aligned}
H(r)_t &= \Sigma_r\, p_{r*t} \log(1/p_{r*t}) \\
H(h)_t &= \Sigma_h\, p_{*ht} \log(1/p_{*ht}) \qquad (15) \\
H(r, h)_t &= \Sigma_r\, \Sigma_h\, p_{rht} \log(1/p_{rht}).
\end{aligned}
$$

These entropies can clearly be aggregated to some spatial level by analogy to equations (9) and (10). Further it can be shown that

$$
\begin{aligned}
H(r, h) &= H(r) + H_r(h) \\
&= H(h) + H_h(r) \qquad (16)
\end{aligned}
$$

where $H_h(r)$ and $H_r(h)$ are the average entropies of r conditional upon h and vice versa:

$$
\begin{aligned}
H_h(r) &= \Sigma_r\, \Sigma_h\, p_{rh} \log(p_{r*}/p_{rh}) \\
H_r(h) &= \Sigma_r\, \Sigma_h\, p_{rh} \log(p_{*h}/p_{rh}). \qquad (17)
\end{aligned}
$$

Here we have omitted the subscript t for notational simplicity.

These conditional entropies have a convenient interpretation in terms of segregation. $H_h(r)$, the average conditional racial entropy, measures the extent of racial integration of a geographical area conditional upon the extent of segregation by household type. Similarly, $H_r(h)$, the average conditional household type entropy, measures the extent of integration of household types conditional upon the extent of segregation by race. These conditional entropies must always be smaller than the unconditional entropies unless the distributions of race and household type are completely independent.[17]

V. Results

Table 4 compares, for 1980, the household type and racial entropy of the geographical components of the San Francisco Bay Area with the maximum entropy possible. The comparison is based upon both the four and six racial classifications and the four and seven household classifications noted in Tables 1–3. Considering all six races, the maximum racial entropy in the region is 0.978, which would be obtained if each and every census tract had the racial composition of the region as a whole—that is, if each tract had the racial proportion reported on the last line of Table 2. The actual racial entropy of the region is lower, 0.759, due to the segregation of races (see Table 5). The reduction in entropy due to racial segregation is 0.219, or 22.43 per cent of the maximum.

Taking the five SMSAs individually, the maximum racial entropy is largest in San Francisco and Oakland, the two SMSAs with the smallest fractions of white households. The measures of segregation are also largest in these two SMSAs, 25.16 per cent and 23.22 per cent respectively. The least segregated SMSA is clearly Santa Rosa, but it is also the one with the smallest non-white population.

Table 4 presents similar information for the central city and suburban rings of each SMSA. These entries must be interpreted somewhat judiciously since the maximum possible entropy is conditional upon the racial composition of only a part of each SMSA. The table indicates that the level of segregation within the suburbs of each SMSA is substantially lower than the level of segregation in the central cities. This indicates that minorities fortunate enough to reside in the suburbs are less segregated than minorities in central cities. It should be noted, however, that the maximum possible entropy is as much as 50 per cent higher in the central cities, reflecting the intense segregation of minorities into the central cities of these SMSAs.

The levels of entropy by racial and demographic grouping, and the interpretation of the segregation indices themselves, are dependent upon the prior classification of the underlying population into meaningful groups. If the groups are too finely divided, their spatial integration will be less remarkable (as when Danish-American and Norwegian-American households are observed to live in adjacent houses). If the groups are too aggregated, their spatial integration may be misleading (as when the increasing integration of blacks and Puerto Ricans in Spanish Harlem is reported as representing a decrease in the level of black–non-black segregation). Accordingly, the second column of Table 4 presents segregation indices computed at a higher level of aggregation, using the three racial groupings noted in Table 3.

A comparison of the first and second columns of Table 4 reveals that the index of segregation is increased significantly from 22.4 per cent to 26.6 per cent when the population is divided into three racial groups (blacks, hispanics and others) rather than six. This difference reflects the relatively greater integration of the Asian and white communities. Spatial integration of Asians with whites (and also with the small population of native Americans and others) 'counts' in the disaggregated analysis in the sense that it leads to a reduction in the measure of racial segregation. In the second column, all non-black, non-hispanic households are considered together. The difference in the index of segregation reported for the Oakland SMSA is particularly striking.

The third and fourth columns present analogous information for 1980 on the segregation of households by demographic type within the region. In the third column the comparison is based on the seven classifications of household type noted in Table 2. For the region as a whole, the maximum entropy is 1.485, which would be obtained if each census tract had a distribution of household types identical to that reported in the last column of Table 2. The maximum entropy by household type is considerably larger than the racial entropy, reflecting in part the more equal classification of households into groups. For the region as a whole, segregation by

Table 5. Actual entropy levels by race and household type, San Francisco Bay Area, 1980

	Race, $H(r)$		Household, $H(h)$		Joint, $H(r, h)$	
Race groups:	Six	Three			Six	Three
Household groups:			Seven	Four	Seven	Four
Entire Region	0.759	0.419	1.363	0.936	2.059	1.338
SMSAs						
Oakland	0.767	0.455	1.376	0.937	2.079	1.373
San Francisco	0.764	0.402	1.312	0.940	2.001	1.322
San Jose	0.814	0.431	1.394	0.932	2.153	1.347
Santa Rosa	0.463	0.231	1.425	0.964	1.853	1.187
Napa	0.762	0.446	1.381	0.903	2.088	1.333
Central City/Suburbs						
Oakland						
Central city	0.960	0.684	1.403	1.047	2.262	1.695
Suburbs	0.715	0.393	1.369	0.907	2.030	1.286
San Francisco						
Central city	0.908	0.469	1.240	0.942	2.052	1.386
Suburbs	0.627	0.337	1.380	0.937	1.952	1.386
San Jose						
Central city	0.915	0.505	1.416	0.939	2.274	1.427
Suburbs	0.697	0.345	1.369	0.924	2.014	1.256
Santa Rosa						
Central city	0.474	0.219	1.391	0.955	1.827	1.166
Suburbs	0.456	0.240	1.449	0.970	1.872	1.202
Napa						
Central city	0.790	0.468	1.388	0.916	2.117	1.364
Suburbs	0.684	0.388	1.362	0.866	2.009	1.246

household type reduces actual entropy to 1.363 (see Table 5), or by 8.19 per cent. Thus, for the region as a whole racial segregation is about two and a half times more intense than segregation by demographic group. When the entropy measures are disaggregated by SMSA, the results are similar. The index of segregation varies from 2.9 per cent in the Santa Rosa SMSA to 8.5 per cent in the Oakland and San Francisco SMSAs. In contrast, the index of racial segregation varies from 8.7 per cent in Santa Rosa to 23.2 per cent in Oakland and 25.2 per cent in San Francisco.

Table 4 also indicates the level of segregation by household type within the central city and suburbs of each SMSA. In contrast to the results by race, there is no systematic difference in the maximum possible entropy between central cities and their surrounding rings. However, there seems to be a slightly greater level of segregation within central cities than within suburbs (in at least four of the five SMSAs), but the differences are rather small. Spatial segregation by household type is far less intense than segregation by race, and differences between central cities and suburbs are far less pronounced.

The fourth column in the table presents the indices of residential segregation by household type in 1980 when households are classified into four groups only: families headed by married couples, single females, single males and non-family households. Computed this way, the indices of spatial segregation are slightly larger. The overall segregation measure is 1.1 percentage points higher; for three of the five SMSAs the index is also higher, by 0.6 to 1.4 percentage points. The differences are rather small, however, and the base is also rather small.

For either of these groupings, the level of

Table 6. Proportionate differences (%) in conditional and unconditional entropies, San Francisco Bay Area, 1980

	Race		Household type	
Race groups: / Household groups:	Six / Seven	Three / Four	Six / Seven	Three / Four
Entire Region	8.30	4.06	4.62	1.82
SMSAs				
Oakland	8.34	4.18	4.65	2.03
San Francisco	9.82	4.98	5.72	2.13
San Jose	6.76	3.71	3.95	1.71
Santa Rosa	7.56	3.46	2.46	0.83
Napa	7.22	3.59	3.98	1.72
Central City/Suburbs				
Oakland				
Central city	10.52	5.26	7.20	3.44
Suburbs	7.55	3.56	3.94	1.54
San Francisco				
Central city	10.57	5.33	7.74	2.65
Suburbs	8.77	4.15	3.99	1.49
San Jose				
Central city	6.23	3.37	4.03	1.81
Suburbs	7.46	3.77	3.80	1.41
Santa Rosa				
Central city	8.02	3.65	2.73	0.84
Suburbs	7.24	3.33	2.28	0.82
Napa				
Central city	7.72	4.27	4.39	2.18
Suburbs	5.41	2.06	2.72	0.92

Note: For columns 1 and 2, table entries are $[H(r)-H_h(r)]/H(r)=J(r, h)/H(r)$; for columns 3 and 4, table entries are $[H(h)-H_r(h)]/H(h)=J(r, h)/H(h)$.

racial segregation is estimated to be about two and a half times as intense as the level of segregation by household type.

Table 6 uses the actual entropy levels reported in Table 5 to compare the conditional and unconditional entropies by race and household type for the various geographical components of the San Francisco Bay Area. Knowledge of the three unconditional entropies, $H(r)$, $H(h)$, $H(r, h)$, permits the average conditional entropies to be calculated from equation (16), as well as the expected mutual information, $J(r, h)$:

$$\begin{aligned} J(r, h) &= H(r)-H_h(r) \\ &= H(h)-H_r(h) \\ &= H(r)+H(h)-H(r, h). \end{aligned} \tag{18}$$

It is clear from (18) that $J(r, h)=0$ when the distributions of r and h are independent; otherwise $J(r, h)>0$. As indicated in Table 5, the values of $J(r, h)$, and hence the differences between the conditional and unconditional entropies, are quite small indeed for the region as a whole and for each of its subareas. It can be shown that the upper bound of J is the smaller of the two marginal entropies, but in fact the values of J for this region are only about one-tenth as large as the smaller marginal entropy. This indicates a substantial degree of independence in the spatial distribution of households by race and household type—for the region as a whole and for its various components. Stated another way, incorporating prior knowledge of the spatial distribution of household types does not affect the expected level of racial segregation very much.

Table 7. Indices of residential segregation; percentages by race and household type and proportionate differences in conditional and unconditional entropies, San Francisco Bay Area, 1970

	Segregation index		Proportionate difference in conditional and unconditional entropies	
	Race (three groups)	Household type (four groups)	Race	Household type
Entire Region	26.91	11.78	4.02	2.08
SMSAs				
Oakland	31.63	11.87	3.59	1.98
San Francisco	25.97	11.82	4.30	2.07
San Jose	11.76	9.46	3.77	2.12
San Rosa	4.52	3.64	3.77	2.12
Napa	15.42	6.49	3.14	1.92
Central City/Suburbs				
Oakland				
Central city	29.02	8.38	4.44	2.68
Suburbs	26.93	10.99	3.27	1.72
San Francisco				
Central city	25.47	9.67	4.62	3.12
Suburbs	21.51	8.06	3.36	1.39
San Jose				
Central city	11.87	10.61	4.31	2.85
Suburbs	9.15	8.11	3.87	1.82
Santa Rosa				
Central city	3.97	4.97	3.40	0.92
Suburbs	4.58	2.42	2.59	0.83
Napa				
Central city	16.68	7.29	3.70	2.05
Suburbs	7.54	3.12	2.32	1.15

Note: See Tables 4 and 6 for definitions.

Table 6 indicates the proportionate change in the conditional and unconditional entropies for 1980. The entries in the table have a convenient interpretation. Suppose the spatial distribution of household types in the metropolitan region is governed by 'economic forces'. Under these circumstances, recognising the known and prior distribution of household types explains only a small fraction of the observed segregation of households by race. Using the most disaggregated definitions of race and household type, and for the region as a whole, only 8.3 per cent of the racial segregation observed could be attributed to segregation by household type. For the central cities of San Francisco and Oakland, only about 10.5 per cent of the racial segregation observed could be attributed to the segregation by household type arising from economic forces.

Alternatively, only about 4.6 per cent of the spatial segregation of household types could be explained by the prior segregation of households by race. For the largest central cities of San Francisco and Oakland, the upper limit was less than 8 per cent.

Using more aggregated groupings, the mutual information is even smaller, and the proportionate changes in conditional and unconditional entropies are even less. These results are also reported in Table 6. Despite the differences arising from group definition, the more aggregated analysis confirms the principal results. *Only a small*

Table 8. Actual entropy levels by race and household type, San Francisco Bay Area, 1970

	Race, $H(r)$	Household, $H(h)$	Joint, $H(r, h)$
Race groups:	Three		Three
Household groups:		Four	Four
Entire Region	0.418	0.564	1.05
SMSAs			
Oakland	0.447	0.562	1.067
San Francisco	0.413	0.583	1.110
San Jose	0.415	0.531	0.937
Santa Rosa	0.253	0.599	0.970
Napa	0.417	0.554	1.017
Central City/Suburbs			
Oakland			
Central city	0.590	0.651	1.395
Suburbs	0.397	0.531	0.953
San Francisco			
Central city	0.488	0.609	1.251
Suburbs	0.325	0.553	0.947
San Jose			
Central city	0.472	0.524	0.926
Suburbs	0.357	0.538	0.978
Santa Rosa			
Central city	0.233	0.605	0.964
Suburbs	0.268	0.594	1.108
Napa			
Central city	0.426	0.553	1.034
Suburbs	0.383	0.561	0.955

fraction of segregation by household type can be explained by a prior segregation of households by race. An even smaller fraction of segregation by race can be explained by economic forces leading to a clustering by demographic group.

The analysis of 1980 census data in more aggregated race–household categories provides a confirmation that the findings do not depend upon the definitions of race or type of household. Consideration of these groupings of households does, however, permit a direct comparison of levels of segregation during the 1970s. Exactly the same information is available for the same census tracts from the 1970 census: counts of households by each of the three racial and four household types (Tables 7 and 8).[18]

Table 7 summarises the identical analysis of segregation by race and household type using 1970 census information for the same 1079 census tracts and for the same racial and household definitions. The index of racial segregation for the region as a whole is 26.91 per cent compared with 26.58 per cent in 1980. During the 1970s, the level of racial segregation declined in four of the five metropolitan areas—by 2 to 3 percentage points in the Oakland, San Francisco and Napa SMSAs. Increases in the level of racial integration were most pronounced within the central city of Oakland and the suburbs of the San Francisco SMSA. Residential segregation by race increased in the San Jose metroplitan area, both in the central city and its suburbs.

A comparison of Tables 4 and 7 also indicates that the residential segregation of households of different types declined during the 1970s. For the area as a whole, the index of segregation declined from 11.78 per cent in 1970 to 9.32 per cent in 1980. This reduction was observed in each of the five SMSAs, and in all 10 central city/suburban subareas.

Finally, a comparison of the proportionate differences in conditional and unconditional entropies in 1970 and 1980 is obtained by comparing the results presented in Tables 6 and 7. For the entire region, the proportionate difference in racial entropy went from 4.02 per cent to 4.62 per cent and the difference by household type went from 2.08 per cent to 1.82 per cent.

The evidence that the spatial distributions of households by race and household type were independent in 1970 as well as in 1980 is equally compelling. The level of mutual information is quite small relative to the joint entropy, and the differences between the conditional and unconditional entropies by race and by household type are very small indeed. For all the entries for 1970 and 1980 reported in Tables 6 and 7, the proportionate difference is in the order of 4 per cent by race and 2 per cent by household type.[19]

The socio-economic forces which lead to spatial clustering of different types of households explain practically none of the spatial segregation of the races in 1980. They explained practically none in 1970, either.

VI. Conclusions

This paper considers residential segregation by race and by type of household using census tract data for 1970 and for 1980. The paper presents entropy indices of segregation for the San Francisco Bay Area and its five constituent metropolitan areas. The methodology permits an investigation of the effects of group definition upon segregation measures, and an analysis of the degree of independence in the segregation of households by race and demographic group. The methodology also permits a comparison by metropolitan area as well as a separate analysis for the central city and the suburban rings of each metropolitan area.

The results indicate that during the 1970s the level of segregation by race declined only slightly, while the level of segregation by household type declined more substantially, at least in this region. Specifically, for the region as a whole the entropy index of segregation by household type declined from 11.78 per cent to 9.32 per cent during the decade. Over this period the entropy index of racial segregation declined from 26.91 per cent to 26.58 per cent. More importantly, however, the results indicate a remarkable independence in the spatial distribution of households by race and demographic group. Only a very small fraction of the observed levels of segregation by race can be 'explained' by the prior partitioning of households by demographic group. Similarly, only a small fraction of the observed segregation by household type can be 'explained' by the prior partitioning of households by race. For example, in 1970 only about 2 per cent of the segregation by household type could be explained by the pattern of racial segregation, and only about 4 per cent of the segregation by race could be explained by the pattern of segregation by household type. These figures are essentially identical in 1980. The principal results of the analysis are invariant to changes in the definition of racial or household groups.

Notes

1. See Yinger (1986) for an extensive discussion.
2. There are important exceptions to the analysis of segregation as a binary phenomenon (see especially Massey and Denton, 1987, which relies upon the same racial definitions as those utilised below), but few studies which systematically measure segregation by (two or more) racial groups and segregation by demographic characteristics.
3. See Lichter (1985), for a discussion of racial concentration, density and racial segregation.
4. See Mood *et al.* (1974, pp. 448–452) for details. Note, however, that this approach

does not yield a measure of the degree of segregation.

5. Any analysis of data by block faces is severely compromised by Census Bureau confidentiality rules which lead to the suppression of population counts by categories, including simple counts of households by race or housing type.
6. See White (1986) for a more extensive comparison of many common measures of segregation.
7. Schnare (1980) has presented some examples of the curious properties of the dissimilarity index in the context of residential segregation by race.
8. A number of papers have used the dissimilarity index to analyse the segregation of hispanic households. Massey (1979) and Massey and Bitterman (1985) considered the segregation of hispanic from 'other' households in two metropolitan areas, while Hwang *et al.* (1985) considered the segregation of hispanics in Texas. The latter study computed dissimilarity indices for three pairwise combinations of households: white–black, white–hispanic and hispanic–black. In addition, at least two papers have attempted to generalise the dissimilarity index to three or more groups (Morgan and Norbury, 1981; Sakoda, 1981). There is, however, no convenient way to extend the dissimilarity index to several groups (see Theil, 1972, for an extensive discussion).
9. One deficiency with the index remains. The entropy index does not overcome the ambiguity arising from the arbitrary way that a region is partitioned into subareas. Other measures of segregation (e.g. the exposure and dissimilarity indices) also suffer from this deficiency. It could be addressed crudely by accounting for the distance between each pair of subareas and assuming that population is concentrated at the centroid of each subarea. See White (1983, 1984) for a discussion. We do not pursue this extension here, because, in contrast to the entropy index, these distance-related measures do not preserve the property of additivity. They also require extensive geocoding of subareas.
10. Census tract data for 1970 were obtained from the Fourth Count Housing Summary Tape, File A. Data for 1980 were obtained from Summary Tape File 3b.
11. Note that one SMSA, the 'Vallejo–Napa–Fairfield' SMSA (hereafter 'Napa'), contains three central cities.
12. Persons not living in households live in group quarters without separate cooking facilities, such as college dormitories.
13. This latter category includes the famous 'Persons of Opposite Sex Sharing Living Quarters' category. Of particular importance in San Francisco, this category also includes homosexual couples.
14. Race is defined by the race of the 'householder', generally the adult cited first by the census respondent.
15. A more extensive discussion and proof of the results summarised in this section is contained in Theil (1972).
16. This merely restates the commonplace observation that all schools cannot be integrated 'fifty–fifty' when 80 per cent of the aggregate student body is composed of members of one race.
17. Specifically

$$H_h(r) \leqslant H(r)$$
$$H_r(h) \leqslant H(h)$$

and the equality holds if and only if

$$\rho_{rht} = \rho_{r*t}\rho_{*ht} \quad \text{for all } r, h \text{ and } t.$$

18. The only difference is in the suppression of data for small samples. In the 1970 census, household counts by race were suppressed if the number of households of that race was less than five. The analysis underlying Tables 7 and 8 was undertaken by distributing the number of suppressed households in any census tract into household types according to the distribution of housing types by that race in the central city or suburban ring containing that census tract.
19. As one further check on the importance of group definition in reaching these conclusions, we conducted the entire analysis using two groups (black–non-black) and two household types (female headed–all others) for 1970 and 1980. The qualitative results were the same as those reported in the text for the more disaggregated groupings.

References

ALLISON, P.D. (1978) Measures of inequality, *American Sociological Review*, 43, pp. 865–880.

CORTES, C.F. *et al.* (1976) Further considerations of the methodological analysis of segregation indices, *American Sociological Review*, 41, pp. 630–637.

DUNCAN, O.D. and DUNCAN, B. (1955) A methodological analysis of segregation indexes, *American Sociological Review*, 20, pp. 210–217.

GREEN, G. and WELNIAK, E. (1982) *Changing Family Composition and Income Differentials*. Special Demographic Analysis CDS-80-7. Washington, DC: US Department of Commerce, Bureau of the Census.

HWANG, S.S. *et al.* (1985) The effects of race and socioeconomic status on residential segregation in Texas, 1970–1980, *Social Forces*, 63, pp. 732–747.

JAMES, D.R. and TAEUBER, K.E. (1985) Measures of segregation, in: N. TUMA (Ed.) *Sociological Methodology*, pp. 1–32. San Francisco, CA: Jossey-Bass.

LICHTER, D.T. (1985) Racial concentration and segregation across US counties, 1950–1980, *Demography*, 22, pp. 603–609.

LIEBERSON, S. and CARTER, D.K. (1982) A model for inferring the voluntary and involuntary causes of residential segregation, *Demography*, 19, pp. 511–526.

MASSEY, D.S. (1979) Residential segregation of Spanish Americans in United States urban areas, *Demography*, 16, pp. 553–564.

MASSEY, D.S. and BITTERMAN, B. (1985) Explaining the paradox of Puerto Rican segregation, *Social Forces*, 64, pp. 306–331.

MASSEY, D.S. and DENTON, N.A. (1987) Trends in the residential segregation of blacks, hispanics and Asians: 1970–1980, *American Sociological Review*, 52, pp. 802–825.

MOOD, A.M., GRAYBILL, F.A. and BOES, D.C. (1974) *Introduction to The Theory of Statistics*, 3rd edn. New York: McGraw-Hill.

MORGAN, B.S. and NORBURY, J. (1981) Some further observations on the index of residential differentiation, *Demography*, 18, pp. 251–256.

SAKODA, J. (1981) A generalized index of dissimilarity, *Demography*, 18, pp. 245–250.

SCHNARE, A.B. (1980) Trends in residential segregation by race: 1960–1970, *Journal of Urban Economics*, 7, pp. 293–301.

TAEUBER, K.E. and TAEUBER, A.F. (1965) *Negroes in Cities*. Chicago, IL: Aldine.

TAEUBER, K.E. and TAEUBER, A.F. (1976) A practitioner's perspective on the index of dissimilarity, *American Sociological Review*, 41, pp. 884–889.

THEIL, H. (1972) *Statistical Decomposition Analysis*. Amsterdam: North-Holland.

WHITE, M.J. (1983) The measurement of spatial segregation, *American Journal of Sociology*, 88, pp. 1008–1018.

WHITE, M.J. (1984) Reply to Mitra, *American Journal of Sociology*, 90, pp. 189–191.

WHITE, M.J. (1986) Segregation and diversity measures in population distribution, *Population Index*, 52, pp. 198–221.

YINGER, J. (1986) Measuring racial discrimination with fair housing audits: caught in the act, *American Economic Review*, 76, pp. 881–893.

ZOLOTH, B.S. (1976) Alternative measures of school segregation, *Land Economics*, 52, pp. 278–298.