

About the Measure

Domain:	Social Environments
Measure:	Race/Ethnic Residential Segregation
Definition:	Race/Ethnic Residential Segregation is a measure of neighborhood race/ethnic residential segregation, based on data from the U.S. Census Bureau.
Purpose:	This measure examines various population characteristics to determine the degree of race/ethnic residential segregation, the degree to which various groups reside in different neighborhoods (Iceland & Douzet, 2006). Race/ethnic residential segregation, particularly when resulting from discrimination, can have negative consequences for minority group members. Race/ethnic residential segregation can limit residential choice, constrain economic and educational opportunities by limiting people's access to good schools and jobs, serve to concentrate poverty in disadvantaged neighborhoods, and contribute to social exclusion and alienation (Massey & Denton, 1988). Residential segregation also affects the nature and quality of intergroup relations in society: segregation reduces contact between groups and is usually thought to cause and reflect polarization across communities (Reardon, 2006). Following Reardon (2006), a region is segregated to the extent to which individuals of a different group live in different neighborhoods in a region. That is, the term <i>segregation</i> does not apply to individual neighborhoods but to larger regions (e.g., school districts, counties, metropolitan statistical areas).
Essential PhenX Measures:	Current Address Residential History
Related PhenX Measures:	None
Measure Release Date:	

About the Protocol

Protocol Release Date:	
PhenX Protocol Name:	Separation (S) Index, Unbiased
Keywords:	ACS, American Community Survey, Neighborhood, Neighborhood Disadvantage, Residential Segregation, Social Determinants of Health, U.S. Census

Protocol Name from Source:	American Community Survey (ACS), 5-year estimates
Description:	<p>The protocol is based on extracting data from the U.S. Census Bureau on a set of variables related to the concept of residential segregation. Residential segregation describes the distribution of different race/ethnic groups across smaller areal units (e.g., census tracts) in larger areas (e.g., counties or metropolitan statistical areas[MSAs]). The Separation Index (also known as the eta squared) is one of the most commonly used race/ethnic residential segregation measures. All the relevant variables are available from the decennial censuses or the American Community Survey (ACS) 5-year estimates. Once the data are extracted, the Separation Index can be calculated.</p> <p><i>Which data set should be used?</i></p> <p>Users interested in using measures of residential segregation in conjunction with the Neighborhood Concentrated Disadvantage protocol should use data from the ACS 5-year estimates for consistency of data sources. However, users should be aware that segregation index values calculated using sample data will be inflated in comparison with scores calculated using 100% count data. The reason for this is that measures of uneven distribution register deviations from “parity,” and these will be more common when using sample data because of the effect of sampling error.</p> <p>Users who are interested in using the 100% count data rather than estimates or making comparisons of residential segregation in metropolitan areas across time (e.g., 1990 vs. 2010) should use data from the decennial censuses. The protocol here describes the process using 5-year estimates from the ACS. Users interested in using the decennial census data should refer to the alternate protocol.</p>
Specific Instructions:	<p>Assuming that information on current address (see PhenX Demographics domain, Current Address measure) has been collected for a study respondent, then it is possible to use geocoding to link the address of a study participant to his or her local neighborhood (or other large geographical unit).</p> <p>It is necessary to extract data for smaller units (e.g., census tracts) to calculate the Dissimilarity Index for each larger unit. To aid comparability between studies, the Social Environment Working Group recommends that researchers set the smaller area to the census tract and the larger area to the metropolitan statistical area.</p> <p>Additionally, researchers can use the census variables to calculate more basic diversity scores at the census-tract level, such as the entropy index.</p> <p>The most common conceptualization of residential segregation is based on the dimension of evenness (Massey & Denton, 1988; Reardon & O’Sullivan, 2004; Taeuber & Taeuber, 1965; White, 1986). The most widely used measure of residential segregation is the Dissimilarity Index, sometimes referred to as <i>D</i>. This measure is computationally straightforward to calculate from census data, and although <i>D</i> was originally applied in a comparison of two population groups (most often whites and blacks), recent papers have extended this measure to the multiple race/ethnic group case (Reardon & Firebaugh, 2002). Others have extended the two populations and multigroup measure by incorporating the spatial dimension using data from adjacent or proximate census units and weighting accordingly (Reardon et al., 2008; Reardon & O’Sullivan, 2004; White, 1983; Wong, 1993).</p>

The WG recommends that investigators calculate values of the Separation Index (*S*) to supplement and compare with values of *D*. *S* has been used extensively in previous research but under a variety of names (e.g., the variance ratio, eta squared, Zoloth's *S*, Coleman's *r*, and more). *S* consistently fares better than *D* in reviews on technical criteria for segregation measures (Fossett, 2017; Reardon & Firebaugh, 2002; White, 1986; Zoloth, 1976) and is far less susceptible than *D* to the problem of index bias (Fossett, 2017; Winship, 1977).

Massey and Denton (1988) state that "residential segregation is the degree to which two or more groups live separately from one another, in different parts of the urban environment." Based on this definition, it is useful to calculate and compare values of *S* and *D* because values of *S* provide a more certain indication that uneven distribution involves separation of groups into ethnically homogeneous areas. In contrast, values of *D* can be high even when the groups in the comparison live together in areas that differ only modestly on ethnic composition (Fossett, 2017). This can occur because *D* responds strongly to small departures from parity that do not involve separation of groups into ethnically homogeneous areas. The only way to identify this pattern is to calculate values of *D* and *S* and compare them (Fossett, 2017).

When comparing the standard *S*-index to the standard *D*-index, if the calculated numbers are similar, either calculation can be used. However, if the numbers differ, it is recommended to use the standard *S*-index formula. Further comparison can be made between the standard *S*-index and unbiased *S*-index. If the calculated numbers are similar, either calculation can be used. However, if the numbers differ, it is recommended to use the unbiased *S*-index formula.

Protocol: The ACS data used in this protocol can be accessed by using Excel to read the Summary Files or using the "Download Center" at the U.S. Census Bureau's American FactFinder portal at <http://factfinder.census.gov>. Users can find additional information on these tools at the following locations:

Using Excel to Access Summary Files: http://www2.census.gov/programs-surveys/acs/summary_file/2014/documentation/tech_docs/ACS_SF_Excel_Import_Tool.pdf

Using the Download Center: http://www2.census.gov/programs-surveys/acs/summary_file/2014/documentation/tech_docs/How_to_Access_ACS_Estimates_AFF.pdf

The technical documentation for the ACS summary files is available online at <http://www.census.gov/programs-surveys/acs/technical-documentation.html>. Select the "Summary File Documentation" link, and then select the data set of interest. Users not familiar with Census data should consult the technical materials.

The key race/ethnicity data in the ACS are found in "Table B03002: Hispanic or Latino by Race." This table is preferred over other possible race and race/ethnic tables available, as it provides data on the main race/ethnic groups in the United States and explicitly incorporates data on Hispanic or Latino populations, otherwise not available in the race-only tables.

Variable Code	Variable Name
B03002001	Total:

B03002002	Not Hispanic or Latino:
B03002003	White alone
B03002004	Black or African American alone
B03002005	American Indian and Alaska Native alone
B03002006	Asian alone
B03002007	Native Hawaiian and Other Pacific Islander alone
B03002008	Some other race alone
B03002009	Two or more races:
B03002010	Two races including Some other race
B03002011	Two races excluding Some other race, and three or more races
B03002012	Hispanic or Latino:
B03002013	White alone
B03002014	Black or African American alone
B03002015	American Indian and Alaska Native alone
B03002016	Asian alone
B03002017	Native Hawaiian and Other Pacific Islander alone
B03002018	Some other race alone
B03002019	Two or more races:
B03002020	Two races including Some other race
B03002021	Two races excluding Some other race, and three or more races

The race/ethnic data are available for all small census geographies—such as census block, census block group, and census tract—and can be easily extracted for almost any geographic level. Note: Although block group data have long been available from the Census File Transfer Protocol site, the Census Bureau did not make block groups available for download at American FactFinder until the release of the 2009-2013 ACS. Information about accessing block group data for earlier years is available at http://www.census.gov/library/video/acs_block_group.html.

Researchers can use the data in this table to easily calculate basic variables (e.g., the percentage of any race and/or ethnicity group) or to combine groups (e.g., all minorities).

Unbiased Versions of S via Difference of Means Calculations

$$\text{Index Score} = \bar{Y}_1 - \bar{Y}_2 = (1/N_1)\sum n_{1i}y_i - (1/N_2)\sum n_{2i}y_i$$

where

n_1 and n_2 are the counts for the reference and comparison groups, respectively, in spatial unit i ,

N_1 and N_2 are the counts for the reference and comparison groups, respectively, for the larger area as a whole,

y_i is a score for “scaled contact with the reference group” assigned on the basis of an index-specific function of the reference group proportion in the population of spatial unit i given by $p_i = n_{1i}/(n_{1i}+n_{2i})$, and

\bar{Y}_1 and \bar{Y}_2 are group means for “scaled contact with reference group”.

In the case of S , the functions for assigning scores on scaled contact with the reference group (y_i) based on the reference group proportion in the population of spatial unit i (p_i) is simple and easy to implement.

For S , $y_i = p_i$. Accordingly, S registers the simple group difference in average contact with the reference group.

S takes value of 0 when the two groups have identical levels of contact with the reference group. This occurs when the two groups live together in smaller areas in the same proportions seen for the larger area as a whole. S takes value of 1 when the comparison group has no contact with the reference group and the reference group has only contact with itself. This occurs when the two groups live apart in areas that are homogeneous.

These formulations of S are mathematically equivalent to the “standard” formulas for S given earlier (derivations are provided in Fossett 2017). They thus yield scores that are identical to the scores obtained using the standard formulas and thus will have the same bias components.

Obtaining Unbiased Index Scores for S

Bias is eliminated from S by calculating the value of p_i as follows:

for members of the reference group, $p_i = (n_{1i}-1)/(n_{1i}+n_{2i}-1)$, and

for members of the comparison group, $p_i = (n_{1i}-0)/(n_{1i}+n_{2i}-1)$.

The resulting adjusted values of p_i are applied as before. The values of S obtained using the adjusted values of p_i in the difference of means formula will be free of bias (Fossett 2017).¹

The adjustment to p_i shown above removes the impact of self-contact on the value p_i . In so doing, it completely eliminates index bias at the point of initial measurement. The basis for this welcome result is simple. The expected value of contact with the reference group among neighbors (excluding the individual under consideration) is unbiased; it is the same for both groups. But the expected value of contact with the reference group based from self-contact is biased; it is always positive for members of the reference group (larger in value when counts involved are small) and always zero for members of the comparison group. Extending the Dissimilarity Index and the Separation Index: The Multigroup Analog

While much early research on segregation looked at two groups (e.g., black and white, or majority and minority), today’s society is multiethnic. Two-group measures are useful but limited for describing complex patterns of segregation. The choice to use a two-group or multigroup D or S depends on the specific question of interest. In a region where the population is composed of three groups (e.g., white non-Hispanic, black non-Hispanic, and Hispanic), we may be interested in

a) segregation between two specific groups (e.g., How segregated are white from black residents?); or

b) segregation among all three groups (e.g., How segregated are white non-Hispanic, black non-Hispanic, and Hispanic residents from each other?).

The two-group measure can still be used by comparing all possible pairs of population groups (Morrill, 1995), but these are not comprehensive, and multiple groups are not treated simultaneously. To address segregation among multiple groups requires a multigroup analog to D (Morgan et al., 1975; Sakoda, 1981). The multigroup analog describes the extent to which two or more population groups are similarly distributed among subareas. The formulas for multigroup dissimilarity (D) and multigroup separation (S) (from Reardon & Firebaugh, 2002) are:

$$S = \sum_{m=1}^M \sum_{j=1}^J \frac{t_j}{T} (\pi_{jm} - \pi_m)^2$$

where

T is total population,

M is the number of groups m ,

J is the number of subareas or units j ,

t_j is number of individuals in subarea j ,

π_m is the proportion in group m ,

π_{jm} is the proportion in group m , of those in unit j , and

I is the Simpson's Interaction Index, given by

$$I = \sum_{m=1}^M \pi_m (1 - \pi_m)$$

In the Stata statistical software package, the command `seg` (installed by typing "ssc install seg" from within Stata) will compute both two-group and multigroup versions of S (Reardon & Firebaugh, 2002).²

Researchers have extended segregation measures by incorporating the spatial dimension (White, 1983; Wong, 1993; Reardon & O'Sullivan, 2004). Fossett (2017) introduces spatial formulations of S and other popular measures of uneven distribution.

Unbiased versions of multigroup indices have not been developed.

¹ There is one further adjustment. Singleton individuals—individuals who happen to be the only member of either group residing in the spatial unit, are excluded from the calculations as the adjusted calculation of p_i will be undefined for them. In practice, this is a rare occurrence.

² The `seg` program calculates S under multiple mathematically equivalent formulations including the "normalized exposure index" and the "squared coefficient of variation index".

Selection Rationale:

The Separation Index (S) provides an objective measure of racial/ethnic residential segregation using U.S. Census Bureau data. A questionnaire that relies on subjective

	<p>judgment based on retrospective ascertainment is likely to be unreliable.</p> <p>Winship (1977) established that the Dissimilarity Index (D), and to a lesser extent S, are potentially subject to non-negligible upward bias under certain circumstances. The bias component of D can be large and create misleadingly high values when areal units have small population counts for one or both groups under even distribution. The problem is well known to researchers and has prevented them from assessing segregation involving small groups or from assessing segregation in smaller communities where segregation would need to be assessed using block data with small population counts.</p> <p>“After-the-fact” adjustments to remove the unwanted impact of bias on index scores have been proposed (e.g., Carrington & Troske, 1997; Winship, 1977), but they do not perform well in practical applications (Fossett, 2017) and have not gained wide use.</p> <p>The unbiased version of S is obtained by measuring segregation with the “difference of means” framework introduced in Fossett (2017). This framework casts all widely used measures of uneven distribution in the following formulation.</p> <p>A recent methodological study (Fossett, 2017) has introduced formulas for calculating a refined version of D that is “unbiased”; that is, the formulas for the unbiased version yield scores for D that are free of the potentially serious problem of upward index bias discussed in Winship (1977). When index bias is not a problem, they yield scores identical to scores obtained using “standard” formulas. When index bias is a problem, they yield scores that are appropriately lower because upward bias has been eliminated.</p> <p>Researchers should calculate values of S to supplement and compare with values of D. S has been used extensively in previous research but under a variety of names (e.g., the variance ratio, eta squared, Zoloth’s S, Coleman’s r, and more). S consistently fares better than D in reviews on technical criteria for segregation measures (Fossett, 2017; Reardon & Firebaugh, 2002; White, 1986; Zoloth, 1976) and is far less susceptible than D to the problem of index bias (Fossett, 2017; Winship, 1977).</p>
Source:	<p>Brown University. (2019). <i>Spatial structures in social sciences</i>. Retrieved May 28, 2019, from https://www.brown.edu/academics/spatial-structures-in-social-sciences/</p> <p>Fossett, M. (2017). <i>New methods for measuring and analyzing segregation</i>. Springer.</p> <p>U.S. Census Bureau. (2019). <i>American Community Survey (ACS) data products (5-year estimates)</i>. Retrieved May 28, 2019, from http://www.census.gov/programs-surveys/acs</p> <p>U.S. Census Bureau. (2019). American FactFinder. Retrieved May 28, 2019, from http://factfinder.census.gov</p>
Availability:	Publicly available
Life Stage:	Infant, Toddler, Child, Adolescent, Adult, Senior, Pregnancy
Language:	English

Participant:	Not applicable; derived from publicly available secondary data
Personnel and Training Required:	<p>Knowledge of census data products and websites, such as American FactFinder and/or publicly available data portals (e.g., https://nhgis.org/) and/or commercial geospatial data products, such as those provided by vendors like GeoLytics or Social Explorer.</p> <p>The extracted data need to be manipulated, and the Index of Dissimilarity needs to be calculated.</p>
Equipment Needs:	Access to a desktop or laptop computer with Internet access to download raw data from the U.S. Census Bureau's American FactFinder website. Statistical packages (e.g., SPSS, SAS) for data manipulation.
General References:	<p>Carrington, W. J., & Troske, K. R. (1997). On measuring segregation in samples with small units. <i>Journal of Business & Economic Statistics</i>, 15(4), 402–409.</p> <p>Fossett, M. (2017). <i>New methods for measuring and analyzing segregation</i>. Springer.</p> <p>Iceland, J., & Douzet, F. (2006). Measuring racial and ethnic segregation. <i>Herodote</i>, 122(3), 25–43.</p> <p>Iceland, J., Weinberg, D. H., & Steinmetz, E. (2002). <i>Racial and ethnic residential segregation in the United States: 1980-2000</i> (U.S. Census Bureau, Series CENSR-3). Washington, DC: U.S. Government Printing Office.</p> <p>James, D. R., & Taeuber, K. E. (1985). Measures of segregation. <i>Sociological Methodology</i>, 15, 1–32.</p> <p>Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. <i>Social Forces</i>, 67, 281–315.</p> <p>Morgan, P. M., Murphy, R. F., Willis, R. A., Hubbard, D. W., & Norton, J. M. (1975). Dental health of Louisiana residents based on the ten-state nutrition survey. <i>Public Health Reports</i>, 90(2), 173–178.</p> <p>Morrill, R. L. (1995). Aging in place, age specific migration and natural decrease. <i>Annals of Regional Science</i>, 29(1), 41–66.</p> <p>Reardon, S. F. (2006). A conceptual framework for measuring segregation and its associations with population outcomes. In J. M. Oakes & J. S. Kaufman (Eds.), <i>Methods in social epidemiology</i> (pp. 169–192). San Francisco, CA: Wiley and Sons/Jossey-Bass.</p> <p>Reardon, S. F., & Firebaugh, G. (2002). Measures of multi-group segregation. <i>Sociological Methodology</i>, 32, 33–67.</p> <p>Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). The geographic scale of metropolitan racial segregation. <i>Demography</i>, 45(3), 489–514.</p>

	<p>Reardon, S. F., & O'Sullivan, D. (2004). Measures of spatial segregation. <i>Sociological Methodology</i>, 34, 121–162.</p> <p>Sakoda, J. M. (1981). A generalized index of dissimilarity. <i>Demography</i>, 18(2), 245–250.</p> <p>Taeuber, K. E., & Taeuber, A. F. (1965). <i>Negroes in cities: Residential segregation and neighborhood change</i>. Chicago, IL: Aldine.</p> <p>Theil, H. (1972). <i>Statistical decomposition analysis</i> (vol. 14). Amsterdam, Netherlands: North-Holland.</p> <p>White, M. J. (1983). The measurement of spatial segregation. <i>American Journal of Sociology</i>, 88, 1008–1018.</p> <p>White, M. J. (1986). Segregation and diversity measures in population distribution. <i>Population Index</i>, 52, 198–221.</p> <p>Winship, C. (1977). A revaluation of indexes of residential segregation. <i>Social Forces</i>, 55(4), 1058–1066.</p> <p>Wong, D. S. (1993). Spatial indices of segregation. <i>Urban Studies</i>, 30, 559–572.</p> <p>Zoloth, B. S. (1976). Alternative measures of school segregation. <i>Land Economics</i>, 52(3), 278–298.</p>										
Mode of Administration:	Secondary data analysis										
Derived Variables:	None										
Requirements:	<table border="1" data-bbox="391 1293 1312 1629"> <thead> <tr> <th data-bbox="391 1293 1013 1373">Requirements Category</th> <th data-bbox="1013 1293 1312 1373">Required (Yes/No):</th> </tr> </thead> <tbody> <tr> <td data-bbox="391 1373 1013 1430">Major equipment</td> <td data-bbox="1013 1373 1312 1430">No</td> </tr> <tr> <td data-bbox="391 1430 1013 1482">Specialized training</td> <td data-bbox="1013 1430 1312 1482">No</td> </tr> <tr> <td data-bbox="391 1482 1013 1556">Specialized requirements for biospecimen collection</td> <td data-bbox="1013 1482 1312 1556">No</td> </tr> <tr> <td data-bbox="391 1556 1013 1629">Average time of greater than 15 minutes in an unaffected individual</td> <td data-bbox="1013 1556 1312 1629">No</td> </tr> </tbody> </table>	Requirements Category	Required (Yes/No):	Major equipment	No	Specialized training	No	Specialized requirements for biospecimen collection	No	Average time of greater than 15 minutes in an unaffected individual	No
Requirements Category	Required (Yes/No):										
Major equipment	No										
Specialized training	No										
Specialized requirements for biospecimen collection	No										
Average time of greater than 15 minutes in an unaffected individual	No										
Annotations for Specific Conditions:	No annotations at this time										

Race/Ethnic Residential Segregation

Date of SC final approval

Process and Review:	Not applicable
----------------------------	----------------