|  |  |
| --- | --- |
| **Domain:** | Social Environments |
| **Measure:** | Race/Ethnic Residential Segregation |
| **Definition:** | 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 both 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 within a region. That is, the term segregation does not apply to individual neighborhoods but to larger regions (e.g., school districts, counties, metropolitan statistical areas. |
| **Essential PhenX Measures:** |  |
| **Related PhenX Measures:** |  |
| **Collections:** |  |
| **Keywords:** | Neighborhood, Residential Segregation, Neighborhood Disadvantage, U.S. Census, Social Environments |

|  |  |
| --- | --- |
| **Protocol Release Date:** | October 8, 2010 |
| **PhenX Protocol Name:** | Race/Ethnic Residential Segregation |
| **Protocol Name from Source:** | This section will be completed when reviewed by an Expert Review Panel. |
| **Description:** | 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) within larger areas (e.g., counties or metropolitan statistical areas [MSAs]). All the relevant variables are available from the 1990 and 2000 decennial Censuses. Once the data are extracted the proposed measure, the Dissimilarity Index can be calculated. The protocol here describes the process using Summary File 1 (SF1) files (i.e., 100% sample data) from the 2000 Census. The Dissimilarity Index is one of the most commonly used race/ethnic residential segregation measures. All the necessary variables used in calculating residential segregation will be available from the 2010 Census. At time of writing (summer 2010) the 2010 Census has not been released, but once available it should be possible to calculate the Dissimilarity Index using input data from census tract and census block-group levels. It is mandated by law that the first population count data should be released before March 31, 2011. |
| **Specific Instructions:** | Assuming that information on current address (see PhenX Demographics domain, Current Address measure) and any previous address(es) (see PhenX Environmental Exposures domain, Residential History measure) has been collected for a study respondent, then via geocoding it is possible to link the address of a participant to his or her county or metropolitan area (or other large geographical unit). 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. Additionally, researchers can use the census variables to calculate more basic diversity scores at the census tract level such as the entropy index. The most common conceptualization of residential segregation is based on the dimension of evenness (Taeuber & Taeuber, 1965; White, 1986; Massey & Denton, 1988; Reardon & O'Sullivan, 2004), and the most widely used measure of residential segregation is the Dissimilarity Index, sometimes referred to as D. This measure is computationally straightforward to calculate from Census data, and while the index of dissimilarity was originally applied in a comparison of two different population groups (most often Whites and Blacks), recent papers have extended this measure to the multiple race/ethnic group case (Reardon & Firebaugh, 2002), and others have extended the 2 and multigroup measure by incorporating the spatial dimension using data from adjacent or proximate census units and weighting accordingly (see White, 1983; Wong, 1993; Reardon & O'Sullivan, 2004; Reardon et al., 2008). |
| **Protocol:** | The Dissimilarity Index is based on U.S. Census data.  The technical documentation for the 1990 and 2000 decennial Censuses are available online at American Factfinder (http://factfinder.census.gov). Select Data Sets link, and then select Decennial Census. Different tabs allow the user to select 2000 or 1990 data and/or materials. Specific data sets are then selected by clicking in data-specific radio buttons (e.g., for Census 2000 options, include SF1 or SF3). SF1 refers to Summary File 1 (the short form of the U.S. Census collected on everyone, and referred to as 100% data), and SF3 refers to Summary File 3 (the long form of the U.S. Census collected on detailed population and housing variables from a one-in-six [16%–17%] sample and weighted to represent the total population). For each data set, the user can create tables and maps, extract data, or consult technical documentation. Users not familiar with Census data should consult the technical materials. The technical documentation for SF1 and SF3 data sets are available at the U.S. Census Bureau website:  SF1: http://www.census.gov/prod/cen2000/doc/sf1.pdf  SF3: http://www.census.gov/prod/cen2000/doc/sf3.pdf  Race and Ethnicity data can be found in SF1 and SF3 (as well as other Census data products). Here, we focus on SF1 and data based on the 100% sample in 2000.  The key race/ethnicity data in SF1 (100%) are found in "Table P8. 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.  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.  Table: SF1 Table P8 on Hispanic or Latino by Race  Universe: Total population  SF1 Table P8 is reproduced below:   |  |  | | --- | --- | | **Variable** | **Reference Cell #** | | Total Population | P008001 | | Not Hispanic or Latino | P008002 | | White alone | P008003 | | Black or African American alone | P008004 | | American Indian and Alaska Native alone | P008005 | | Asian alone | P008006 | | Native Hawaiian and Other Pacific Islander alone | P008007 | | Some other race alone | P008008 | | Two or more races alone | P008009 | | Hispanic or Latino | P008010 | | White alone | P008011 | | Black or African American alone | P008012 | | American Indian and Alaska Native alone | P008013 | | Asian alone | P008014 | | Native Hawaiian and Other Pacific Islander alone | P008015 | | Some other race alone | P008016 | | Two or more races alone | P008017 |   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).  The Dissimilarity Index provides data on larger areas (e.g., metropolitan statistical areas) using smaller level data.  The most common conceptualization of residential segregation is based on the dimension of evenness. Evenness refers to the differential distribution of the subject population across neighborhoods in a large area (e.g., metropolitan area). It ranges from 0 (complete integration) to 1 (complete segregation) and indicates the percentage of a group's population that would have to change residence for each neighborhood to have the same percentage of that group as the metropolitan area overall. It is computed as:  where  *n* is the number of tracts in the larger area (e.g., a metropolitan area),  *xi*is the population size of the minority group of interest in tract *i*,  *X* is the population of the minority group in the larger area (e.g., metropolitan area) as a whole,  *yi*is the population of the reference group (usually non-Hispanic Whites) in tract *i*, and,  *Y* is the population of the reference group in the larger area (e.g., metropolitan) area as a whole.  The calculation requires the computation of the totals for each group across all subareas within a larger region (e.g., all census tracts within a county), the proportion of each group within each subarea, the absolute difference between the proportions, and the sum of the absolute differences. The latter number is multiplied by 0.5 to generate a result between 0.0 and 1.0. A value of 0.0 would indicate there were the same proportions of majority and minority group populations in each subarea, as in the larger regions' population. If all subareas within the region contain members of just one group (i.e., there is no co-residence) then *D* equals 1.0, indicating complete segregation.  Extending the *Dissimilarity* 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* 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 formula for multigroup dissimilarity (from Reardon & Firebaugh, 2002) is:  where  *T* is total population,  *M* is the number of groups *m,*  *J* is the number of subareas or units *j,*  *tj*is number of individuals in subarea *j,*  *πm*is the proportion in group *m,*  *πjm* is the proportion in group m, of those in unit *j,* and  *I* is the Simpson's Interaction Index, given by  The interpretation of multigroup *D* (sometimes labeled as *D*(*m*)) is the same as *D* (see Wong, 1993).  In the Stata statistical software package, the command *seg* (installed by typing "ssc install seg" from within Stata) will compute *D* (Reardon, 2002).  Researchers have extended segregation measures by incorporating the spatial dimension (see White, 1983; Wong, 1993; Reardon & O'Sullivan, 2004). There are spatially modified versions of the *D* index (see Wong, 1993). |
| **Selection Rationale:** | The PhenX Social Environments Working Group preferred an objective measure of racial/ethnic residential segregation using Census data. A questionnaire that relies on subjective judgment based on retrospective ascertainment is likely to be unreliable. |
| **Source:** | Recommended data sources include:  U.S. Census Bureau decennial Census (1990 and 2000).  U.S. Census Bureau. (2001). Census 2000, Summary File 1, Technical Documentation, Available from http://www.census.gov/prod/cen2000/doc/sf1.pdf.  American Factfinder website: http://factfinder.census.gov.  Note that several online sources provide Dissimilarity Index scores for selected metropolitan statistical areas, counties, and school districts (and across Census years). See, for example, the American Communities Project and the School Segregation Project at the Brown and Lewis Mumford Center at Albany (http://www.s4.brown.edu/cen2000/index.html; http://www.s4.brown.edu/schoolsegregation/index.htm) as well as the Spatial Impact Factor Web Data at RTI International (http://rtispatialdata.rti.org/). |
| **Life Stage:** | Any age |
| **Language of source:** | English |
| **Participant:** | Not applicable: Derived from publicly available secondary data |
| **Personnel and Training Required:** | Knowledge of Census data products and websites, such as American Factfinder, or commercial geospatial data products, such as that provided by vendors such as GeoLytics (http://www.geolytics.com). The extracted data need to be manipulated, and the Index of Dissimilarity needs to be calculated. |
| **Equipment Needs:** | Access to a desktop/laptop computer with internet access to download raw data from the U.S. Census Bureau's American Factfinder website (http://factfinder.census.gov). Statistical Packages (e.g., SPSS, SAS) for data manipulation. |
| **Standards:** | |  |  |  |  | | --- | --- | --- | --- | | **Standard** | **Name** | **ID** | **Source** | | Common Data Element (CDE) | Social Environment Race/Ethnic Residential Segregation Assessment Score | 3151013 | [CDE Browser](https://cdebrowser.nci.nih.gov/CDEBrowser/search?elementDetails=9&FirstTimer=0&PageId=ElementDetailsGroup&publicId=3151013&version=1.0) | | Logical Observation Identifiers Names and Codes (LOINC) | Race - ethnic resid segregation proto | 63038-4 | [LOINC](http://s.details.loinc.org/LOINC/63038-4.html?sections=Web) | |
| **General references:** | Iceland, J., & Douzet, F. (2006). Measuring racial and ethnic segregation. *Hrodote, 122*(3): 25–43.  Iceland, J., Weinberg, D. H., & Steinmetz, E. (2002). *Racial and ethnic residential segregation in the United States: 1980–2000* (U.S. Census Bureau, Series CENSR‑3). Washington DC: U.S. Government Printing Office.  Available from http://www.census.gov/prod/2002pubs/censr-3.pdf  Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces,* *67,* 281–315.  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. *Public Health Reports, 90*(2), 173-178.  Morrill, R.L. (1995). Aging in place, age specific migration and natural decrease. *The Annals of Regional Science, 29*(1), 41-66.  Reardon, S. F. (2006). A conceptual framework for measuring segregation and its associations with population outcomes. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in social epidemiology* (pp. 169–192). San Francisco, CA: Wiley and Sons/Jossey-Bass.  Reardon, S. F., & Firebaugh, G. (2002). Measures of multi-group segregation.  *Sociological Methodology, 32*, 33–67.  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. *Demography,* *45*(3), 489–514.  Reardon, S. F., & O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology, 34*, 121–162.  Sakoda, J.M. (1981). A generalized index of dissimilarity. *Demography, 18*(2), 245-50.  Taeuber, K. E., & Taeuber, A. F. (1965). *Negroes in cities: Residential segregation and neighborhood change*. Chicago, IL: Aldine.  Theil, H. (1972). *Statistical decomposition analysis* (Vol. 14). Amsterdam, The Netherlands: North-Holland.  White, M. J. (1983). The measurement of spatial segregation. *American Journal of Sociology,* *88,* 1008–1018.  White, M. J. (1986). Segregation and diversity measures in population distribution. *Population Index,* *52,* 198–221.  Wong, D. S. (1993). Spatial indices of segregation. *Urban Studies*, *30,* 559–572. |
| **Mode of Administration:** | Secondary Data Analysis |
| **Derived Variables:** | None |
| **Requirements:** | |  |  | | --- | --- | | **Requirement Category** | **Required** | | Major equipment | No | | Specialized training | No | | Specialized requirements for biospecimen collection | No | | Average time of greater than 15 minutes in an unaffected individual | No | |
| **Process and Review:** | This section will be completed when reviewed by an Expert Review Panel. |