Graphic Detail

Geographic Information Systems organize and clarify the patterns of human activities on Earth's surface and their interaction with each other. GIS data, in the form of maps, can quickly and powerfully convey relationships to policymakers and the public. This department of Cityscape *includes maps that convey important housing or community development policy issues or solutions. If you have made such a map and are willing to share it in a future issue of* Cityscape*, please contact david.e.chase@hud.gov.*

Visualizing Racial Segregation Differently—Exploring Changing Patterns From the Effect of Underlying Geographic Distributions

Ronald E. Wilson
U.S. Department of Housing and Urban Development

It has long been observed that minorities are often concentrated in impoverished areas and lack access to resources such as jobs, educational opportunities, good-quality food, life services, and face other disadvantages. As such, racial and ethnic segregation is an important factor to consider when crafting public policy. A first step in many analyses of segregation is the production of single racial or ethnic group percentage maps that show geographic patterns where one group is predominant and the other is not. At lower levels of geography, such as census block groups or tracts, these percentage maps are less problematic because the size of the unit is small enough and can reveal a concentration of single racial or ethnic groups within a small area. When examining trends across the United States at the city and county level, however, the percentage mapping approach becomes disadvantageous for two primary reasons.

First, important patterns are lost when racial or ethnic percentages are the result of summed-up population counts from lower geographies and displayed on a map. This pattern loss is the result of the Modifiable Areal Unit Problem (MAUP), which has adverse consequences for data analysis

because the unit of geography changes but the observation data do not.[1] This straight summation of data has a dilution effect that can erase variation in local racial and ethnic settlement patterns in the underlying geography, leaving true levels of segregation more difficult to detect. This is a crucial point, because segregation can be more acute at the local level than across many cities or counties. When these local patterns are not apparent, incorrect assumptions can be made about larger regional segregation patterns associated with social and economic ills.

A second reason the approach is flawed is because maps drawn this way convey the assumption that segregation is the same throughout the region. Just because a racial or ethnic group is concentrated within a region does not indicate homogeneous distribution at the local level. There is a far greater direct spatial disconnect at the local level than at regional or national levels, because the mismatch between segregated populations and needed resources to maintain a certain quality of life is more succinctly separated. At the city and county levels, there is a reduction in the spatial disconnect, but the dynamics of segregation and associated problems become the cumulative result of those local effects that combine with adjacent jurisdictions and can depress the regional economy, making it less vibrant. Thus, simple thematic maps of the presence or absence of one racial or ethnic group at the national level do not depict very well the extent and magnitude of segregation and misguide the formation of questions toward uncovering the true problems associated with segregation at the regional level. Factoring in levels of local segregation within cities and counties can instead give an indication of how integrated racial or ethnic groups are across a region and the nation as a whole.

This article demonstrates an alternative approach that accounts for local-level segregation below the city and county level that minimizes the effects of MAUP. This approach shifts the context of the analysis from a solely segregation standpoint to one that also exhibits integration among racial or ethnic groups. This approach recasts regional patterns of segregation in a way that highlights other regional problems associated with the pervasive problems resulting from the concentration of minorities. This article is the first of two that examine segregation at the regional level.

The data used for this analysis are from the American Community Survey (ACS) 5-year estimates from 2005 to 2009 by the U.S. Census Bureau. Percentages and measures are derived to reveal changing patterns of segregation and integration from the factoring of local variation into the analysis. This analysis will identify not just succinct jurisdictions (cities and counties) of segregation but large areas (clusters of contiguous jurisdictions) that form subregions. The focus of this analysis is between the White and Black populations in the contiguous United States.

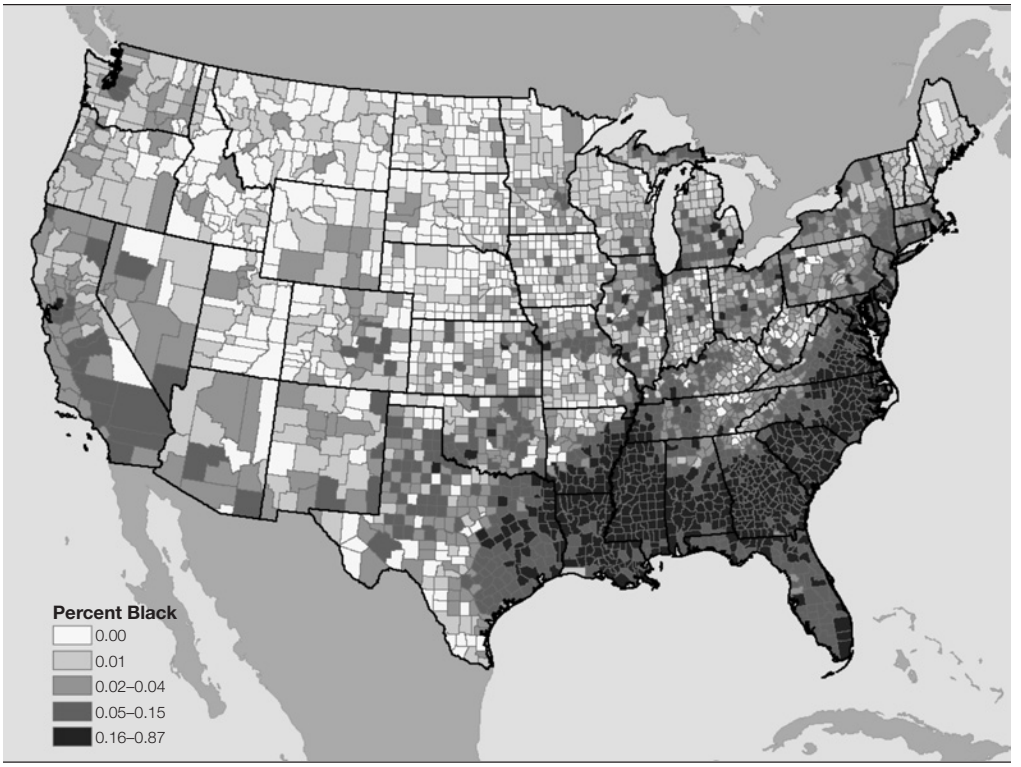
The United States has a long history of regional segregation between the Black and White populations. The Black population constitutes 13 percent of the total population with most living in 19 percent of the cities and counties across the United States. Translated into geography, the Black population primarily resides in only 9 percent of the entire country. A look at the geographic distribution of those percentages shows a very clear pattern of regional segregation (see exhibit 1).[2]

[1] For a thorough discussion of MAUP, see Openshaw (1994).

[2] See the appendix for corresponding frequency distributions for all maps.

Exhibit 1

Distribution of the Black Population for the Contiguous 48 States—Quantile Classification



The concentration of the Black population appears to form a solid belt of cities and counties from east Texas straight across the southern states and continuing up the east side of the Carolinas, ending mainly around Baltimore, Maryland. Also, several small and loosely coupled regions of Black concentrations are in the northeastern, midwestern, and western states. The north-central states are relatively devoid of the Black population.

Several problems emerge when examining segregation this way. First, this map displays only the results of one value and has no larger comparative context. The percentage is a comparative measure, only in a relative sense, as a proportion and the resulting map leaves an assumption about the distribution of the White population. It can be assumed that the White population lives in areas where there are low percentages of Blacks or the two populations are more balanced where the percentages are moderate, but this cannot be ascertained without actual values incorporated into the analysis. Second, when one group of the population is unevenly distributed across an area, a thematic mapping classification scheme must be selected that allows patterns in the data to be distinguished. In this instance, the quantile classification scheme is used because of a significant amount of variation in the frequency distribution. This scheme[3] nicely highlights clusters of like

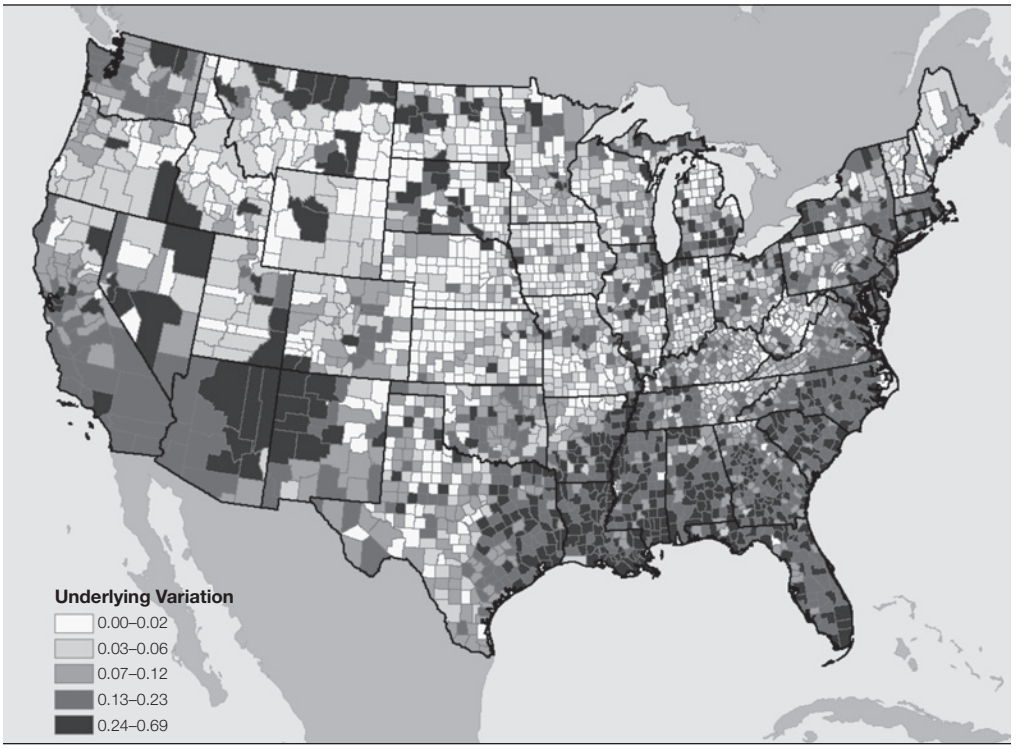
[3] The quantile classification groups equal numbers of jurisdictions into each data partition.

values but includes too wide of a range of values in the highest partition; the range of data values in this partition is exaggerated when the distribution is highly skewed. A look at the legend shows that 16 to 87 percent of the Black population is contained within that belt. A geographic pattern forms that includes many jurisdictions that do not have similar percentages. This geographic pattern is exaggerated with respect to the size of the regional concentration of the Black population, and using any of the other classification schemes does not offer a solution. The other schemes either hide patterns or create inappropriate data partitions on skewed frequency distributions. As a result of these problems, percentage maps do not communicate the message about segregation as precisely as they should; they should depict the true geographic extents and patterns of regional segregation. The final problem, though, is that city and county population counts are straight summations of lower level tallies in which variation from the lower geographies is erased by the MAUP. This summation can significantly affect the display of geographic patterns, regardless of the thematic mapping classification scheme used.

The way to get a first indication of hidden levels of segregation within a jurisdiction is through mapping local variation of percentage point differences between the White and Black populations. Variation is measured as the standard deviation of the absolute differences between the percent of the White and Black population across the underlying census tracts for each jurisdiction (see exhibit 2).

Exhibit 2

Variation of White and Black Population Percentage Differences Within Cities and Counties—Quantile Classification



The resulting data frequency distribution again requires the quantile thematic mapping classification scheme to map the data. A low standard deviation will indicate a more even distribution of each racial group across the jurisdiction because of little variation in the percentage point differences from the tracts within the jurisdiction. Conversely, a high standard deviation signifies a large variation in percentage point differences between the underlying tracts. It cannot be discerned whether or not the percentage point differences are small or large, because this method reveals only whether a variation exists across tracts within a jurisdiction.

There is a significant change in the geographic patterns compared with the single percentage map of the Black population in exhibit 1. The belt across the southern states and into the northeastern states is still present, but the pattern has become diluted, revealing a more dispersed level of regional segregation. What is revealed, however, is that many of the jurisdictions within the belt have similar percentages of the White population, indicating a more integrated population. A large regional cluster of high variation between the two groups is now present in the four-corner states of Arizona, Colorado, New Mexico, and Utah. Several small clusters of high variation between groups also now appear in the north-central states. Finally, several jurisdictions in the northeastern states have high variations in the differences between the two groups, indicating they are more segregated than shown in the single percentage map. These pattern changes hint that segregation is a localized phenomenon that varies significantly.

Although the map in exhibit 2 compares one population with another, it still does not situate either group in the context of the combined populations to determine how segregated the two groups are within a jurisdiction. Using this map makes it difficult to discern which racial group is dominant across a region, but historical knowledge of segregation in the United States, in general, is a good indicator. Nevertheless, the advantage of using this map is that it gives an indication of the differences between the White and Black populations that may prompt an examination of jurisdictions that are more segregated than others at the local level. Further, the skewed shape of the distribution continues to prevent the use of a thematic map classification scheme that can equalize the class partitions. The use of a diversity index can help alleviate these continuing problems by transforming the data into a form that addresses the previous technical and substantive issues.

Several dissimilarity and diversity index measures can be applied to data to depict levels of segregation.[4] In this analysis, Theil's entropy index[5] is used because it has mathematical properties that are sensitive to disproportionality changes between two or more groups that matches the theoretical aspects of changes in segregation levels in place. The resulting index allows for an examination of segregation between the two groups in comparison to the total population between them (see exhibit 3).[6] More importantly, with the re-expression of data, the equal interval classification scheme can be used, because the variation across the full range of values has been reduced and standardized. This scheme partitions data values into equal ranges and is not affected by the distribution of the data. This grouping makes the ranges comparable with each other and, subsequently, affects the map by limiting the geographies in the highest partition to those that are truly different with respect to what is being analyzed, which, in this case, is the levels of segregation of the Black population.

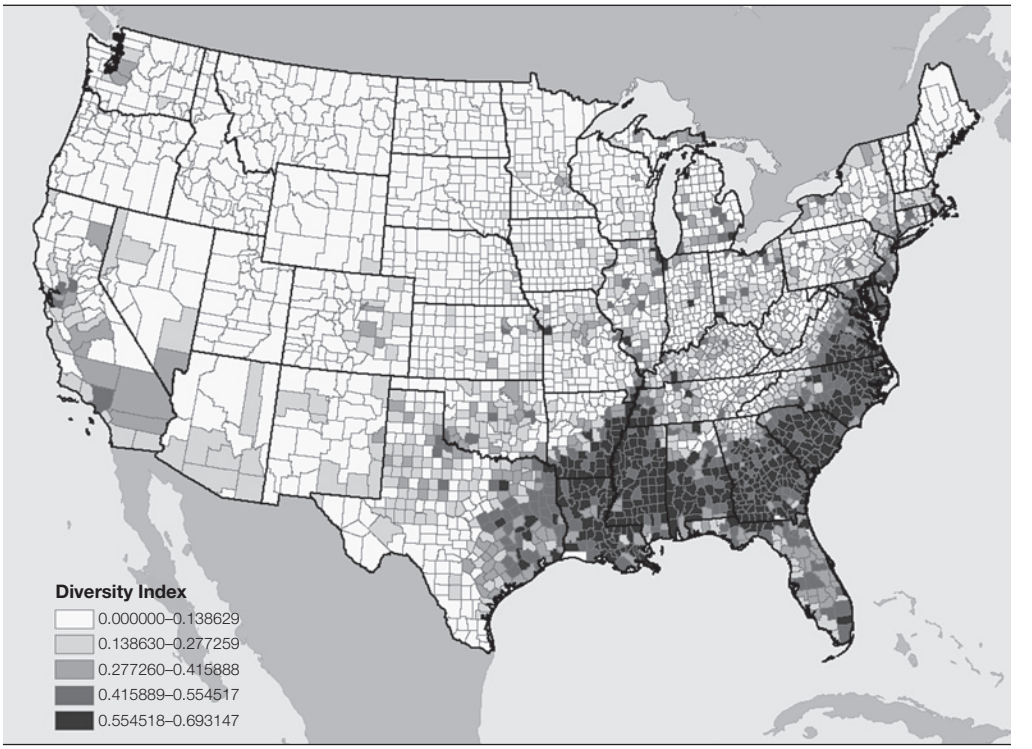
[4] For a full analysis of several common indexes, see Reardon and Firebaugh (2002).

[5] For the mathematical details of Theil's entropy index, see the appendix.

[6] Values closer to 0 indicate more segregation and values closer to 1 indicate more integration.

Exhibit 3

Diversity Levels of Cities and Counties—Equal Interval Classification



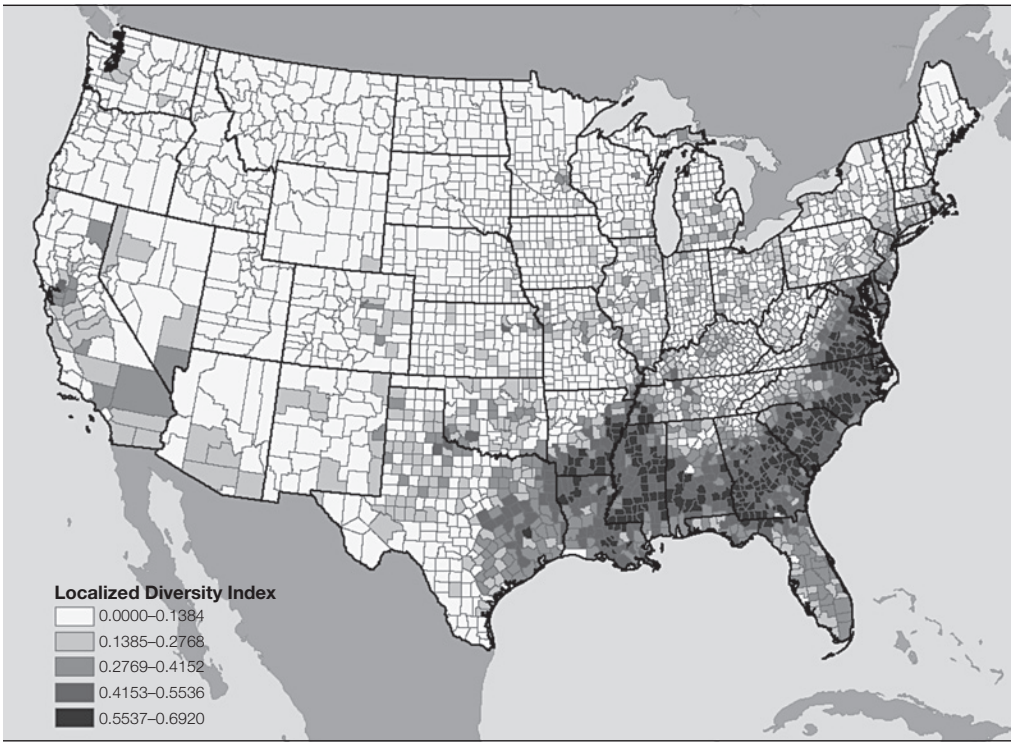
The map in exhibit 3 has three distinct advantages over the original percentage map in exhibit 1. The first advantage is a more succinct regional segregation pattern. The index has produced values that express the magnitude of difference between the two groups based on how integrated each group is in comparison with the total population of each other. The regional belt of Black population concentration from east Texas to northern Maryland is still prevalent but has been thinned out and slightly broken up. The concentration of the Black population is now in 14 percent of the jurisdictions across the United States, a reduction of 5 percentage points from the percentage map in exhibit 1. The index, however, detects levels of segregation only between two groups and does not indicate which group dominates on either end of the range. Again, a historical knowledge of the distribution of racial groups in the United States will indicate which group is more dominant in a particular region. The second advantage is that the map more accurately shows where the two groups are segregated and integrated. Many jurisdictions have indexes that are in the upper partitions of the distribution with several states in the belt showing a significant level of integration between Whites and Blacks. With this map, Florida is much more distinguishable as being segregated within the state because the index values show greater variation. Third, the remainder of the jurisdictions across the United States continues to be widely distributed, but the indexes are now in the lowest partition compared with the percentage map in exhibit 1. The exception is California, in which several jurisdictions have indexes in the middle partitions indicating a greater level of statewide segregation.

An interesting result is that the cities of Chicago, Detroit, St. Louis, and Indianapolis show up in the highest partition, indicating a high level of integration between the two populations, although it is well known that they are very segregated cities. The previous results are still subjugated to MAUP, because the data used at the city and county levels demonstrate a loss of information because the population counts used are summed from the underlying geography. This same effect is also likely occurring in many other jurisdictions; this likelihood is of concern. To compensate for this effect, the diversity index can be adjusted for the underlying local segregation between the two racial groups across the census tracts contained within a jurisdiction.[6] (See exhibit 4.)

The main observation in exhibit 4 is that the belt that stretches across the southern states and up the east coast has thinned out more and broken apart, forming small regions of integration with a coherent pattern of regional segregation. This map is more accurate because the two populations have been placed in context with each other and local levels of segregation have been factored in. The concentration of the Black population in exhibit 4 is now in 8 percent of the jurisdictions across the United States. Two specific examples exemplify the corrective adjustment the localized index makes. In exhibit 3, Chicago, Detroit, Indianapolis, and St. Louis each has an index in the highest class, which indicates that it is very integrated between the two populations. The adjustment

Exhibit 4

Localized Diversity Levels of Cities and Counties—Equal Interval Classification



[6] For mathematical details of the localized diversity index, see the appendix.

for local variation minimizes the effect of the MAUP and shifts the index to a level that is more reflective of the segregation in the underlying geography. A similar instance is in Bibb County, Alabama, which is located south of the city of Birmingham in the center of the state. This jurisdiction is now an island of high segregation within a region of high integration in the surrounding jurisdictions. The city of Birmingham, which is just northeast of Bibb County, was also reduced from the highest partition to the middle and now matches the surrounding counties, all of which remained in the same partition after the adjustment. The two jurisdictions now stand in stark contrast to each other. The variation across the rest of the United States is further attenuated, indicating a more realistic portrait of much of the county having little in the way of a Black population. In addition, the original percentage map in exhibit 1 showed that nearly all jurisdictions in Mississippi were in the highest partition of the Black population. Mississippi now has clear interior geographic patterns of segregation and integration. The geographic patterns in the underlying variation in exhibit 2 also revealed this trend, but the effects of the quantile thematic mapping scheme made it unclear whether the changes resulted from a wide range of high and low values being partitioned together. Also, the racial groups in exhibit 2 were not within the context of their combined populations. California still maintains several jurisdictions with indexes in the middle partitions with little adjustment for levels of segregation previously.

Exhibit 4 contains several pieces of information that were not available in the original percentage map of the Black population in exhibit 1. The original percentage map in exhibit 1 showed a limited aspect of segregation, which was just a continuum of percentages of where the Black population was and was not located. That map also did not situate the Black population within a context of another population for comparison and account for local segregation. The final map now meets several elements of excellence in statistical graphics that are not in the single percentage map.[7] The final map more accurately identifies core areas of segregation and reveals systemic and anomalous patterns that facilitate the exploration of factors associated with segregation such as regional poverty, education, or economic deprivation. Further, because the data have been standardized, a more direct comparison can be made with those factors in and between specific cities or counties. More informed questions can be developed about spatial connections with the local levels and the broader region in which they sit. The use of a diversity index, localized or not, is much more analytically robust and brings out several aspects often hidden in single percentage maps.

The main point of this article is to push thinking away from simple percentage maps of segregation at the national level. As the United States becomes a more diverse nation, the percentage maps become more misleading about segregation and its association with related social and economic ills. The Graphic Detail article in the next issue of *Cityscape* (volume 13, number 3) will take this analysis further and examine the variation in segregation in the context of economic activity areas.

[7] Tufte (1983).

Appendix

Diversity Index (Theil's Entropy Index)

Theil's entropy index is formally defined as:

$$Diversity_j = \sum_{j=1}^{J} \pi_j \log\left(\frac{1}{\pi_j}\right) \quad (1)$$

where j is a jurisdiction and is π proportion of each racial group in jurisdiction j.[8] The result is D_j, which is a continuous value that ranges between 0 and $\log\left(\frac{1}{\pi}\right)$. Values closer to 0 indicate higher levels of segregation by one group or the other. Higher values indicate greater levels of diversity between the two groups.

In the context of place, the result is a nonlinear curve in which each unit increment or decrement along the scale has a different rate and magnitude of change. Values on either end of the curve have different qualitative meanings. Geographies are not simply the additive sum of their assets.[9] The addition of each new asset transfers benefit in such a way as to compound that benefit to a greater degree as assets are accumulated through the increased combinations among other assets. The converse is also true. As assets are removed from a place, the decline compounds the negative effects. Social, economic, and political behavior reacts to place changes and further induces other positive or negative effects, depending on whether assets are added or removed. A linear function does not capture this dynamic, because a constant is applied that produces no change in the rate or magnitude of the curve and only shifts the curve up or down uniformly. Lack of any change in the shape of the curve would represent a static impact on a place from assets being added or subtracted with no compounding effect, which is theoretically and empirically not supported.

This reexpression is due to the log function standardizing values by confining them to consistent range with each other. Logarithms applied to skewed data compress values in the lower tails while systematically enlarging values in the middle and in the upper tail. In particular, the *natural log* (as used in the entropy index) preserves the dispersion of the distribution in the transformed data, making it comparable to the original distribution given they are approximate each other. This preservation permits the multiplication of π_j with the log of π_j to produce a meaningful result, because values close together in the original distribution remain close in the transformed distribution. The spreads between the two distributions have been stabilized so they can be added together to create the index from appropriately scaled components to produce a meaningful result. Using a transformation that does not preserve the dispersion would result in the multiplication of mismatched original and transformed values of the same unit and distort the resulting curve.

In addition, outliers are made more prominent through the multiplication operation and are scaled relative to the other values in the lowest and highest classes. Large outliers are amplified so that the

[8] For a review of the mechanics of this index, see Wong (2003).

[9] An asset is defined as (1) developmental, (2) commercial, (3) recreational, (4) physical, and (5) social, where a greater combination of each asset leads to an improved quality of life.

group at the upper end and small outliers at the lower end do not significantly increase beyond their original value.

Localized Entropy Index

The localized index adjusts for lower level geographic variation and is formally defined as:

$$\overline{Diversity}_j = \left(\sum_{i=\Omega_j}^{J} N_i * D_i\right) / \left(\sum_{i=\Omega_j}^{J} N_i\right) \quad (2)$$

where $\overline{D}$ is now the average level of diversity of all tracts in jurisdiction j accounting for the variation, Ω is the number of tracts in jurisdiction j, N_i is the total population of the two racial groups in tract i, and D_i is the same index in equation 1 but applied to each tract i. The two population groups for each tract i are multiplied by the local diversity index individually and summed up to jurisdiction j, forming a local interaction effect that weights the diversity commensurate with the size of the population of the two groups. The interactions capture the variation at the local level before the result is summed to the jurisdiction, thus adjusting for local levels of segregation. After the interactions are summed up for each jurisdiction, the result is rescaled back to the relative size of the total population of the two groups to create an average level of diversity. Now $D_j \geq \overline{D}_j$ and preserves all the mathematical properties and interpretations from the original diversity index but is now adjusted for underlying geographical variation of segregation.

Frequency Distributions

Each appendix exhibit number corresponds with the respective map exhibit number in the main body of the article (for example, exhibit 1 corresponds with exhibit A-1).

Exhibit A-1

Distribution of the Black Population for the Contiguous 48 States—Quantile Classification

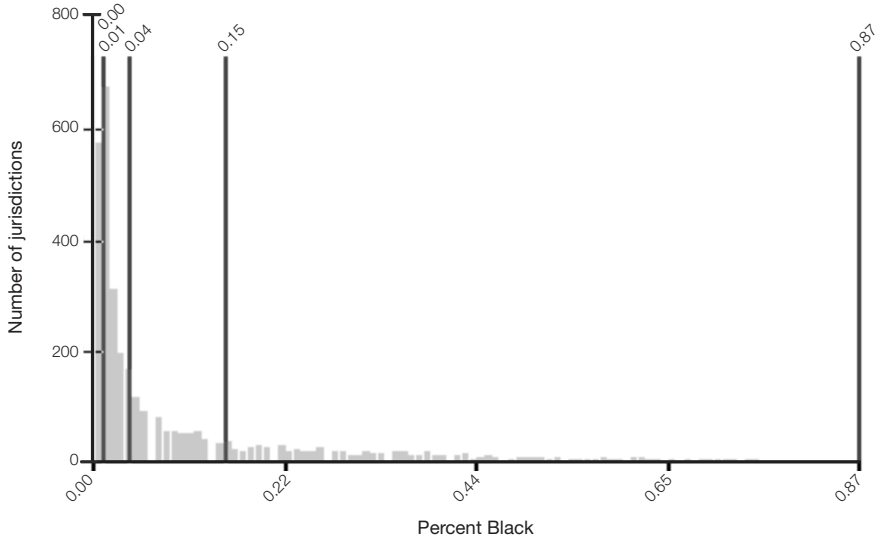


Exhibit A-2

Variation of White and Black Population Percentage Differences Within Cities and Counties—Quantile Classification

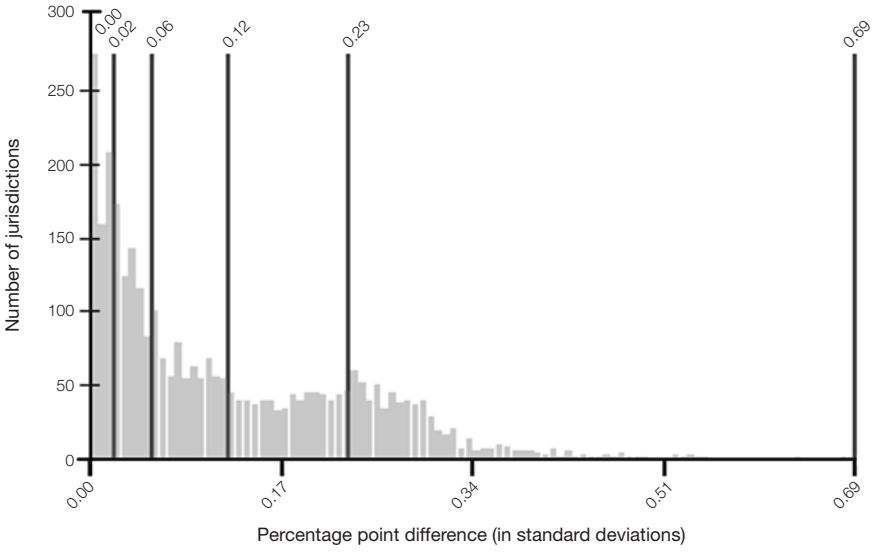


Exhibit A-3

Diversity Levels of Cities and Counties—Equal Interval Classification

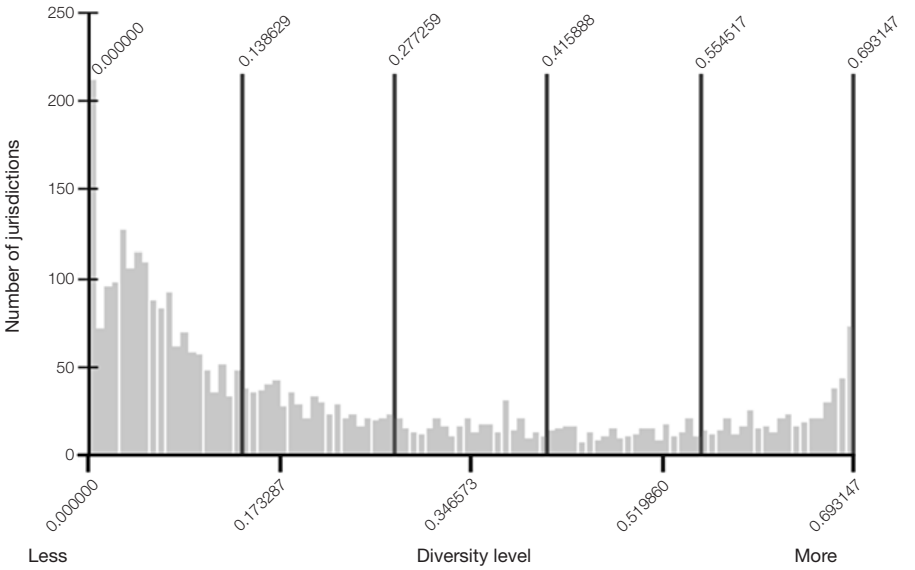
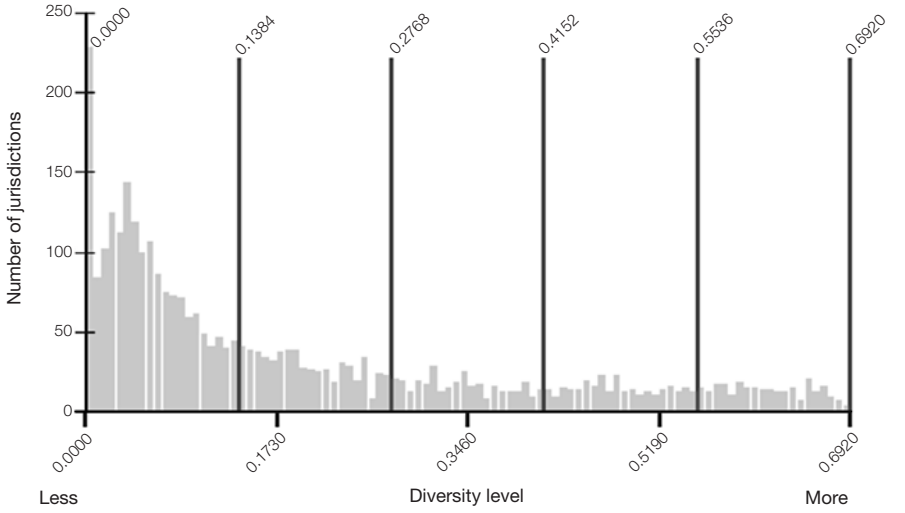


Exhibit A-4

Localized Diversity Levels of Cities and Counties—Equal Interval Classification



Acknowledgments

The author thanks Michael O'Leary from Towson University for providing valuable comments toward clearly describing the mathematical operations behind the statistical equations used in this analysis. The author also thanks Jay Lee from Kent State University for providing comments on the geographical analysis.

The views expressed in this article are those of the author and do not represent the official positions or policies of the U.S. Department of Housing and Urban Development.

Author

Ronald E. Wilson is a social science analyst in the Office of Policy Development and Research at the U.S. Department of Housing and Urban Development.

References

Openshaw, Stan. 1994. *The Modifiable Areal Unit Problem*. London, United Kingdom: Geo Books. http://qmrg.org.uk/files/2008/11/38-maup-openshaw.pdf (accessed March 23, 2011).

Reardon, Sean F., and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation," *Sociological Methodology* 32: 33–67.

Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Wong, David W.S. 2003. "Spatial Decomposition of Segregation Indices: A Framework Toward Measuring Segregation at Multiple Levels," *Geographical Analysis* 35 (3): 179–194.