



Spatio-temporal changes in racial segregation and diversity in large US cities from 1990 to 2020: a visual data analysis

Anna Dmowska^{1*}  and Tomasz F. Stepinski²

*Correspondence:

dmowska@amu.edu.pl

¹Institute of Geoecology and Geoinformation, Adam Mickiewicz University, Bogumiła Krygowskiego 10, 61-680 Poznań, Poland
Full list of author information is available at the end of the article

Abstract

Urban populations in large US cities exhibit racial and ethnic diversity, yet they remain residentially segregated. The examination of temporal trends in segregation and diversity is crucial for sociologists and urban planners. In this study, we investigate the spatio-temporal changes in segregation and diversity across 61 major US cities, utilizing data from four US Censuses conducted between 1990 and 2020. Unlike previous studies, our approach relies on visual data analysis, enabling us to capture the overarching changes in racial coresidence during this period. We employ four distinct perspectives – geographical, temporal, groups evolution, and desegregation scale limit – to visualize and analyze the data. Geographical analysis uncovers a decrease in regional disparities in urban diversity and segregation since 1990, as urban racial integration extends beyond West Coast and Southwestern cities to encompass the entire US. Through temporal analysis, we observe a general trend of rapidly increasing diversity and gradual reduction in segregation, albeit with varying rates across different cities. Groups evolution analysis reveals that cities grouped based on their diversity and segregation metrics in 1990 follow the overall trend toward larger diversity and smaller segregation while preserving group's coherence but not their distinctiveness. Finally, the desegregation scale limit perspective suggests that, on average, over the 1990 to 2020 period, the desegregation scale has started to subceed the lower limit of the census block. By employing these diverse analytical perspectives, our study provides a comprehensive understanding of the changes in racial segregation and diversity within US cities over the past three decades.

Keywords: Racial segregation; US Census data; Racial dynamics; Visual data analysis

1 Introduction

Patterns of spatial distribution in large US cities provide evidence of segregation, whereby each racial group exhibits its distinct spatial distribution. Segregation occurs when these distributions show minimal overlap, and the degree of overlap serves as a measure of segregation, with smaller overlaps indicating higher levels of segregation. To quantify segregation, the field of racial demography has developed various metrics (for a comprehensive review of segregation metrics, refer to [1]). Among these metrics, the information the-

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

ory index H is commonly used to quantify racial segregation in multiracial populations [2, 3]. In addition to measuring segregation, another important metric in studying multiracial populations is racial diversity. A population is considered diverse when multiple racial groups significantly contribute to the overall population composition. The population diversity is the number of distinct groups that make significant contributions to the total population [4].

To analyze spatio-temporal changes in urban diversity and segregation across the United States, a sample of cities representing various regions was selected, and metrics of diversity and segregation were calculated using data from multiple past censuses. The findings of such data analysis have been published since the early 21st century and continue to be published to this day [5–16]. Typically, these results were presented in tabular form, providing values of diversity and segregation metrics for different cities at various census years. However, this approach only allows for a limited comparison of average indices across different census years, resulting in an extreme compression of the extensive information available in past censuses.

In this study, we deviate from the conventional approach in racial demography by employing visual data analysis methods to examine the spatio-temporal changes in urban racial segregation. Visualization is a powerful tool for intuitively analyzing complex phenomena [17] and has been widely utilized in the field of sociology [18], which is the primary source of racial segregation studies. However, until recently [19, 20] visualization has mainly been used for illustrating rather than analyzing data in the context of racial segregation studies. The Racial Landscape (RL) method, introduced by Dmowska et al. [19], represents a significant advancement in this area. It provides a geospatial dataset that visualizes the high-resolution distribution of all racial groups within a single map. The RL visualization resembles a detailed “image” of the land, indicating the racial composition of its inhabitants. Moreover, the RL includes a tool that enables segregation calculations for any given area without relying on census boundaries. Another relevant contribution is the Segplot proposed by Elbers [20], which is a graphical tool designed to visualize patterns of segregation. Unlike RL, Segplot is an effective aspatial data visualization method. However, both of these approaches primarily focus on visualizing and analyzing racial data within a single city rather than across multiple cities.

In this paper, we employed visual data analysis techniques to examine the temporal changes in segregation among a diverse set of cities representing various regions across the United States. To achieve this, we compiled a comprehensive dataset of diversity and segregation metrics using census data from the years 1990, 2000, 2010, and 2020. The dataset consisted of 61 the largest US cities, strategically distributed throughout the conterminous 48 states. Subsequently, we conducted a visual analysis of this dataset from four distinct perspectives: geographical, temporal, groups evolution, and desegregation scale limit.

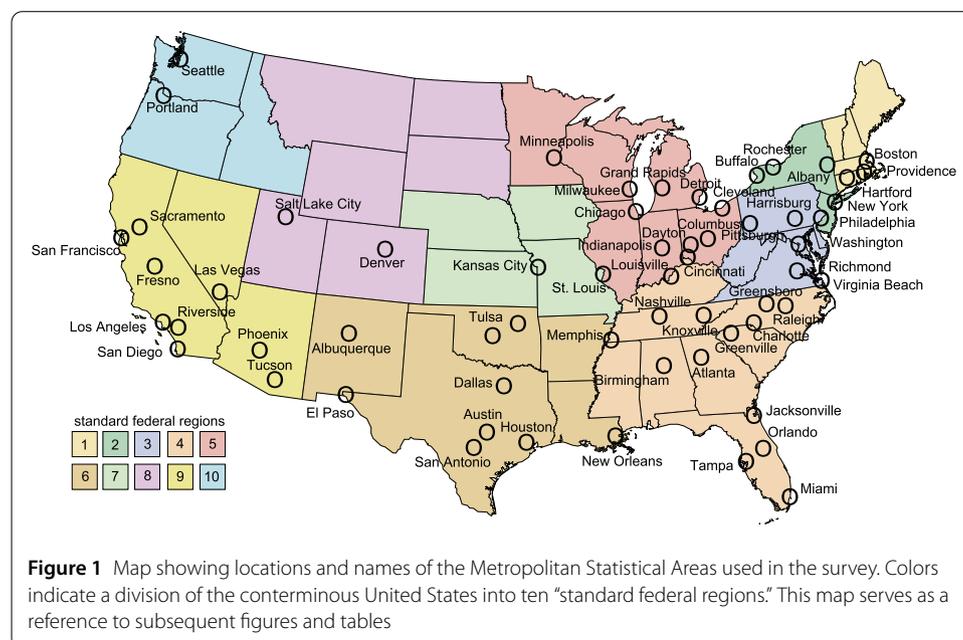
This paper introduces several novel contributions to the field. Firstly, our approach utilizes visual data analysis, allowing for a comprehensive visualization of the entire dataset while effectively highlighting the overall trend. This approach offers a novel perspective and facilitates a more intuitive understanding of the data. Secondly, we propose a transformation of the standard metric of racial diversity (entropy) into a more direct estimation of the number of distinct groups that significantly contribute to a city’s total population. Whereas entropy is a measure of diversity, the new index (the Hill’s number) is the di-

versity [4]. Thirdly, we conduct groups evolution analysis to investigate the persistence of similarities among cities that shared similar values of diversity and segregation metrics in 1990. This analysis sheds light on the long-term patterns of urban racial dynamics and highlights the evolving nature of these cities over time. Lastly, we employ desegregation scale limit analysis to explore whether there are changes in the spatial scale of desegregation over the study period. This analysis uncovers important insights into the spatial dynamics of segregation and offers valuable information about the shifting patterns of urban residential segregation.

2 Data and method

We obtained the racial composition data for our analysis from the U.S. Census Bureau, specifically the 1990, 2000, 2010, and 2020 datasets at both the tract and block levels of spatial aggregation. These datasets were accessed from the National Historical Geographic Information System (NHGIS) [21]. Our analysis focuses on a sample of 61 the largest metropolitan statistical areas (MSAs) based on their 2020 boundaries. For brevity, we refer to these areas as “cities” throughout the paper, although we are analyzing the entire MSAs.

In Fig. 1, we present a visual representation of the geographical locations and names of the cities included in our survey. This figure also shows a division of the United States into ten standard Federal Regions; these regions have no official names, and they are identified by numbers from 1 to 10. Our analysis involves the calculation of metrics related to racial diversity and composition using the following subpopulations: White (W), Black (B), Asian (including Hawaiian/Pacific Islanders) (A), Hispanics (H), and others (including American Indians) (O). It is important to note that all racial subpopulations, except for Hispanics, are categorized as non-Hispanic. For the sake of conciseness, we refer to each subpopulation as a “race,” but we acknowledge that the Hispanic category represents ethnicity.



2.1 Diversity and segregation metrics

Entropy E is frequently used as a diversity metric of the multiracial population consisting of K different subpopulations,

$$E = - \sum_{k=1}^K f_k \ln(f_k), \quad (1)$$

where \ln represents the natural logarithm and $F = f_1, \dots, f_K$ denotes a set of fractions (or shares) representing the total population count in K subpopulations, with $\sum_k f_k = 1$. Each value f_k can be interpreted as the probability that a randomly selected person from the population belongs to race k . It is worth noting that entropy is a functional, which means it takes another function as its argument and yields a numerical value. In this case, the argument is the normalized population histogram F , and the resulting number E quantifies the shape of the histogram. For a narrow histogram that represents a population dominated by a single race, the entropy value is minimum, denoted as $E_{\min} = 0$. On the other hand, a maximally broad histogram where all subpopulations have an equal share results in the maximum entropy value, represented as $E_{\max} = \ln K$. The entropy metric captures the level of uncertainty regarding the race of a randomly selected individual from the population, with larger entropy values indicating greater diversity.

It is important to emphasize that the entropy value, denoted as E , remains unaffected by the specific assignment of races to the histogram bins. Therefore, if the races were assigned to different bins, the entropy would remain the same. For instance, a city comprising 50% Whites, 30% Blacks, and 20% Hispanics would have the same entropy value as a city with 50% Hispanics, 30% Whites, and 20% Blacks.

As mentioned in the [Introduction](#), entropy is not an intuitive measure of diversity [4]. Intuitively, we would expect that a city with $2K$ equally common subpopulations is twice as diverse as a city with K equally common subpopulations. However, entropy does not align with this expectation. For instance, when $K = 4$, Eq. (1) indicates that $E(2K \text{ city})/E(K \text{ city}) \approx 1.5$. Moreover, entropy values can be ambiguous because they depend on the choice of logarithm base used in Eq. (1) and whether the entropy is standardized or not. The demographic literature typically employs standardized entropy calculated with the natural logarithm, but this choice is based on tradition. Some more recent studies, such as Stepinski and Dmowska [22], have used a logarithm with a base of 2 and have not applied standardization.

These issues are addressed by applying a simple transformation $E \rightarrow a^E = N_H$, where a represents the base of the logarithm (in this paper, we use the Euler number e to remain consistent with the demographic literature). The resulting quantity, N_H , is known as Hill's number [23]. Hill's number is referred to as the effective diversity or the effective number of subpopulations because it represents the number of equally abundant subpopulations that would yield the same entropy value as the actual subpopulation composition. In practice, Hill's number is often not an integer, so we estimate the number of substantial groups by rounding the value of N_H to the nearest integer (see [24] for further details). The major advantage of Hill's number over entropy is that it does not require interpretation; it simply express a diversity [4]. Moreover, unlike entropy, its interpretation is unambiguous.

A segregation metric of the multiracial population is most frequently calculated using the information theory index H [2, 25],

$$H = \frac{E^a - \sum_{i=1}^n \frac{t_i}{T} E_i^s}{E^a} = \frac{E^a - \langle E^s \rangle}{E^a} = \frac{MI}{E^a} \quad (2)$$

where E^a represents the entropy of the entire area, and E_i^s represents the entropy of the i th subdivision within this area. The numerator in Eq. (2) corresponds to the difference between the diversity of the entire area and the population-weighted average of diversities in individual subdivisions (denoted as $\langle E^s \rangle$). In information theory [26], this quantity is referred to as mutual information (MI). The value of MI indicates the extent to which we have reduced uncertainty (on average) regarding the race of a randomly chosen person by considering the population of a specific subdivision rather than the entire population. Thus, MI serves as a measure of segregation by quantifying the reduction in diversity.

The denominator in Eq. (2) serves as the normalizing constant, thus H can be interpreted as the reduction of uncertainty at the subdivision's population level *relative* to the uncertainty at the entire population level. The relative nature of H ensures that a value of $H = 1$ corresponds to complete separation (at the scale of subdivisions used or lower) of subpopulations, regardless of the number of subpopulations present. However, this relative nature of H prevents its transformation into a more interpretable metric expressed in terms of the ratio of the number of subgroups in the entire area to an average number of subgroups in a subdivision, as would be the case if segregation were measured using MI . The use of MI as a measure of segregation has been discussed by various authors [22, 27–29]; however, in this paper, we employ H to align with the prevailing trend in the demographic literature.

3 Results

The complete collection of diversity and segregation metrics for the 61 analyzed cities in the years 1990, 2000, 2010, and 2020 can be found in Table S1 of the Additional file 1. In the main body of the paper, our focus is on visually analyzing these metrics from four distinct perspectives: geographical, temporal, group evolution, and desegregation scale limit.

3.1 Geographical perspective

The geographical perspective serves the purpose of visually depicting spatial variations in the racial characteristics of cities across the contiguous United States. Figure 2 comprises four maps corresponding to the years 1990, 2000, 2010, and 2020, respectively. Cities are represented by disks of varying sizes and colors. In order to facilitate visual data analysis, the size of each disk is proportional to the diversity metric (N_H), while the darkness of its color corresponds to the segregation metric (H).

The primary utility of Fig. 2 lies in visually examining a map for a particular year to discern regional disparities in the diversity and segregation of US cities, and to observe how these disparities evolve over time. For instance, the 1990 map clearly indicates that cities in regions 9 (California) and 6 (Texas) exhibited, on average, higher levels of diversity and lower levels of segregation compared to cities in the other regions. The 2020 map reveals that, on average, cities in regions 9 (California) and 6 (Texas) experienced a slight increase in diversity and a slight decrease in segregation over the span of 30 years. However, during the same period, cities in other regions witnessed more substantial increases in diversity and decreases in segregation, on average.

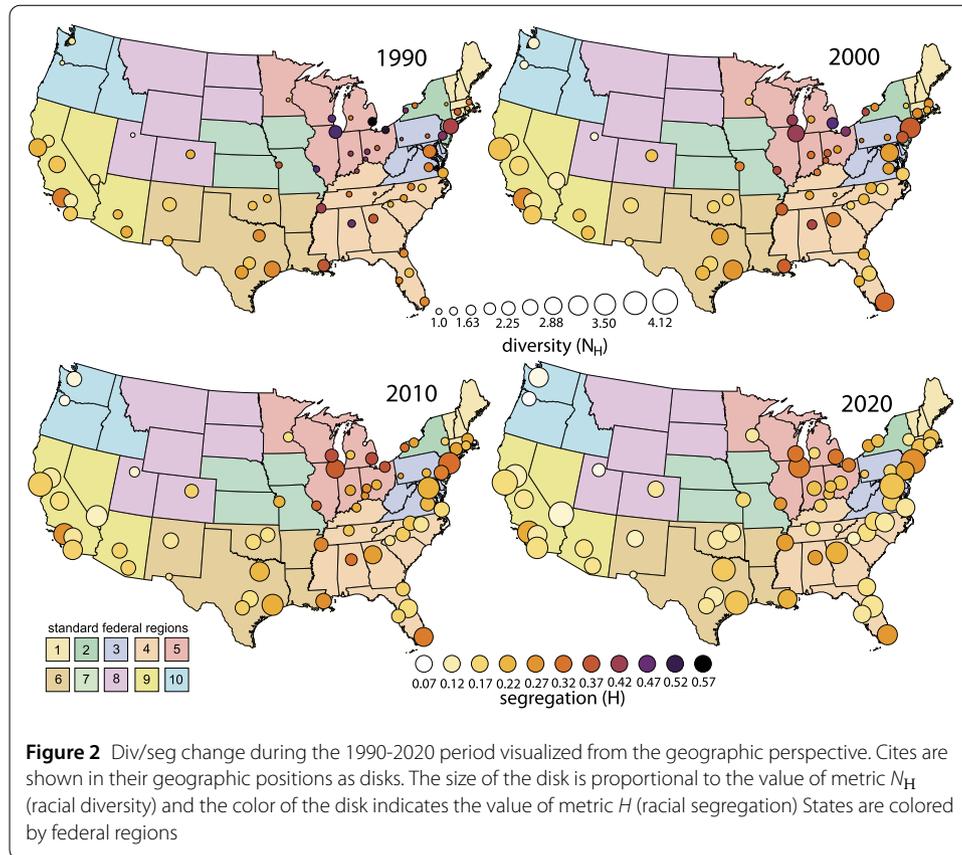


Figure 2 Div/seg change during the 1990-2020 period visualized from the geographic perspective. Cities are shown in their geographic positions as disks. The size of the disk is proportional to the value of metric N_H (racial diversity) and the color of the disk indicates the value of metric H (racial segregation) States are colored by federal regions

Table 1 Diversity and segregation metrics for standard federal regions

Region	City	1990			2000			2010			2020			1990-2020	
		Div	Seg	Change	Change										
1	3	1.70	0.30	22%	2.07	0.26	-12%	2.38	0.24	-10%	2.86	0.20	-16%	68%	-34%
		0.35	1.92	17%	0.31	2.25	10%	0.27	2.47	10%	0.22	2.89	17%	-37%	51%
3	6	2.00	0.34	17%	2.33	0.28	-12%	2.65	0.23	-16%	3.08	0.19	-19%	54%	-44%
		0.31	1.88	24%	0.25	2.33	14%	0.21	2.66	14%	0.18	3.06	15%	-43%	62%
5	10	1.80	0.43	18%	2.11	0.35	-18%	2.36	0.29	-17%	2.75	0.15	-19%	53%	-44%
		0.25	2.43	7%	0.22	2.72	12%	0.19	2.90	7%	0.24	3.22	11%	-37%	32%
7	2	1.78	0.43	15%	2.05	0.35	-19%	2.29	0.29	-17%	2.67	0.22	-23%	50%	-49%
		0.15	1.81	22%	0.15	2.22	11%	0.14	2.47	11%	0.11	2.81	14%	-26%	54%
9	9	2.69	0.43	18%	3.17	0.35	-7%	3.36	0.29	-7%	3.60	0.24	-13%	34%	-25%
		0.14	1.72	29%	0.10	2.22	17%	0.09	2.59	17%	3.11	20%	-13%	81%	-41%

The regional disparities depicted in Fig. 2 are quantified in Table 1. This table provides the average diversity (Div) and segregation (Seg) metrics for cities located in ten different regions for the period spanning 1990 to 2020, as well as the changes observed in these

metrics between consecutive censuses, expressed as a percentage change in the values of N_H (or H for segregation) relative to the preceding census. The first column is the ID number of the region, the second column is the number of cities in the region, and the third column lists values of N_H (top) and H (bottom) in 1990. The next six columns correspond to the years 2000, 2010, and 2020. There are two columns for each year, the first shows the values of N_H (top) and H (bottom), and the second shows values of percentage change of N_H (top) and H (bottom) from the previous census year. The last column represents the percentage change from 1990 to 2020.

The findings obtained from the geographical perspective analysis can be summarized as follows:

- Throughout each decade, diversity increased and segregation decreased in all regions. In 1990, cities, on average, consisted of two sizable racial groups, whereas in 2020, the average city had three sizable racial groups.
- In 1990, the levels of diversity and segregation exhibited strong regional disparities among US cities. Over the course of the next 30 years, these regional differences persisted, albeit to a lesser extent. Cities in regions 9 (California) and 6 (Texas) remained the most diverse and least segregated, on average. Cities in regions 5 and 7 (Midwest) remained the least diverse and most segregated, on average. Cities in region 10 (Portland, OR, Seattle, WA) exhibited relatively low segregation in 1990 and managed to maintain this low level while increasing their diversity over the subsequent 30 years.
- The level of diversity seems to reach a threshold at the presence of the four major racial groups. This observation may be attributed to the classification system used by the U.S. Census, which only distinguishes four significantly populous racial groups.

3.2 Temporal perspective

The objective of the temporal perspective is to visually compare the rates of change in diversity and segregation indices across different cities. The visualization method employed is the same for both diversity, quantified by values of N_H , and segregation, quantified by values of H .

To analyze diversity, we begin by arranging the cities in ascending order based on their 1990 values of N_H . Figure 3 represents this ranked list as a blue chain of points (rank, N_H). By construction, the values of N_H in this chain increase monotonically with the rank, ranging from the least diverse city (Knoxville, TN) to the most diverse city (Los Angeles, CA). Subsequently, using the 1990 city ranking, we plot their corresponding values of N_H for the years 2000 (a chain of yellow points), 2010 (a chain of green points), and 2020 (a chain of red points).

The first notable observation is that, in general, the yellow chain lies above the blue chain, the green chain lies above the yellow chain, and the red chain lies above the green chain. Therefore, racial diversity tends to increase monotonically over time in the majority of cities. The average (\pm standard deviation) increases in diversity between consecutive censuses are $\Delta N_H = 0.38 \pm 0.19$, 0.26 ± 0.13 , and 0.37 ± 0.12 for the periods 1990-2000, 2000-2010, and 2010-2020, respectively. Notably, El Paso, TX deviates from this trend. Racial diversity in El Paso, TX decreased from 1990 to 2000 and again from 2000 to 2010, before experiencing a slight increase from 2010 to 2020. This is attributed to the fact that the population of El Paso, TX has predominantly become Hispanic since 1990.

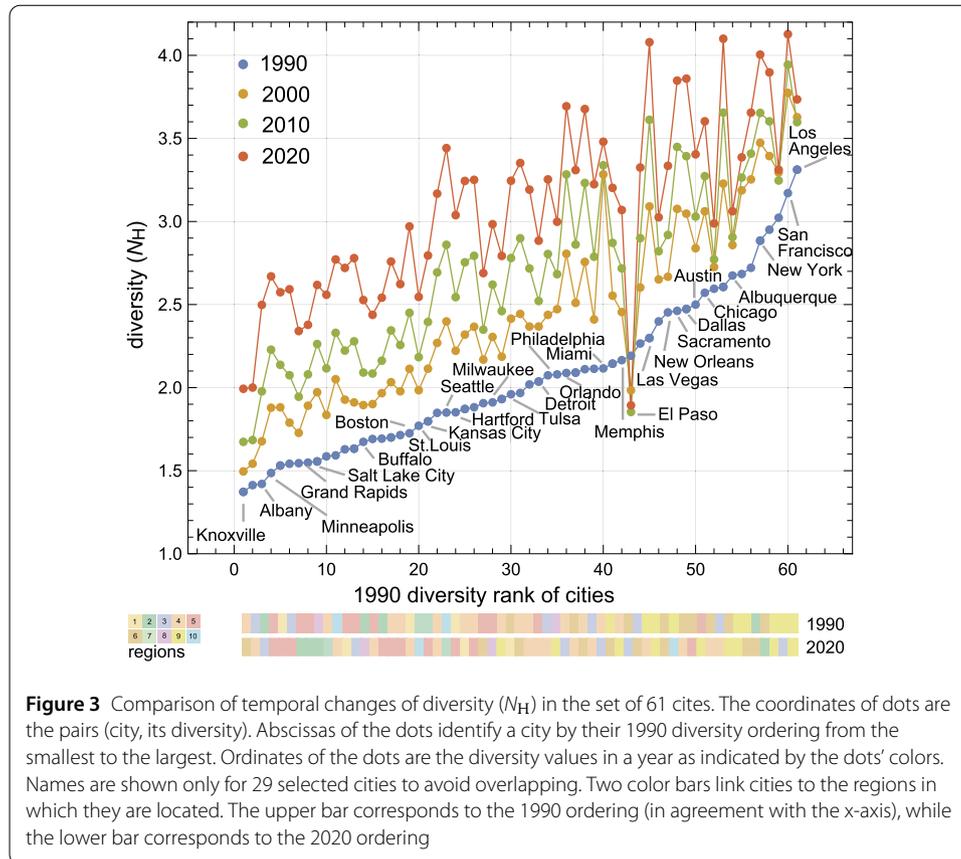


Figure 3 Comparison of temporal changes of diversity (N_H) in the set of 61 cities. The coordinates of dots are the pairs (city, its diversity). Abscissas of the dots identify a city by their 1990 diversity ordering from the smallest to the largest. Ordinates of the dots are the diversity values in a year as indicated by the dots' colors. Names are shown only for 29 selected cities to avoid overlapping. Two color bars link cities to the regions in which they are located. The upper bar corresponds to the 1990 ordering (in agreement with the x-axis), while the lower bar corresponds to the 2020 ordering

Another observation is that, unlike the blue chain, the yellow, green, and red chains do not exhibit a monotonic increase in relation to the 1990 rank, resulting in a zig-zagged visual pattern. This indicates that the diversity-based rankings of cities are reshuffled after each census due to varying degrees of diversity growth in different cities (as evident from the relatively large standard deviations of diversity growth mentioned earlier).

On average, there was an increase in diversity during the entire 1990-2020 period, with $\Delta N_H = 1.01 \pm 0.35$ across all 61 cities. It is important to note that N_H represents the effective number of distinct population groups in a city. Therefore, an average increase of N_H by approximately 1 indicates that, on average, the racial composition of a city saw an increase of one significant group during the 1990-2020 period. Some of the smallest changes in diversity occurred in cities that were already highly diverse in 1990 (located in the upper right corner of Fig. 3). Los Angeles, CA serves as a good example, as it already had $N_H = 3.3$ in 1990. Given that Hill's number represents the number of distinct subpopulations significantly contributing to the population, and considering that the census lists only four significant populations (Whites, Blacks, Hispanics, and Asians), there is limited room for the growth of N_H in Los Angeles from its 1990 value.

In Fig. 3, two color bars positioned below the x-axis establish a connection between the locations of individual cities (by region) and their diversity rankings. The upper bar represents the 1990 ordering, while the lower bar represents the 2020 ordering. Analyzing these two bars qualitatively provides insights into regional diversity trends. For instance, we observe that region 9 (California) consistently dominates high diversity rankings in both 1990 (five of the top ten) and 2020 (four of the top ten).

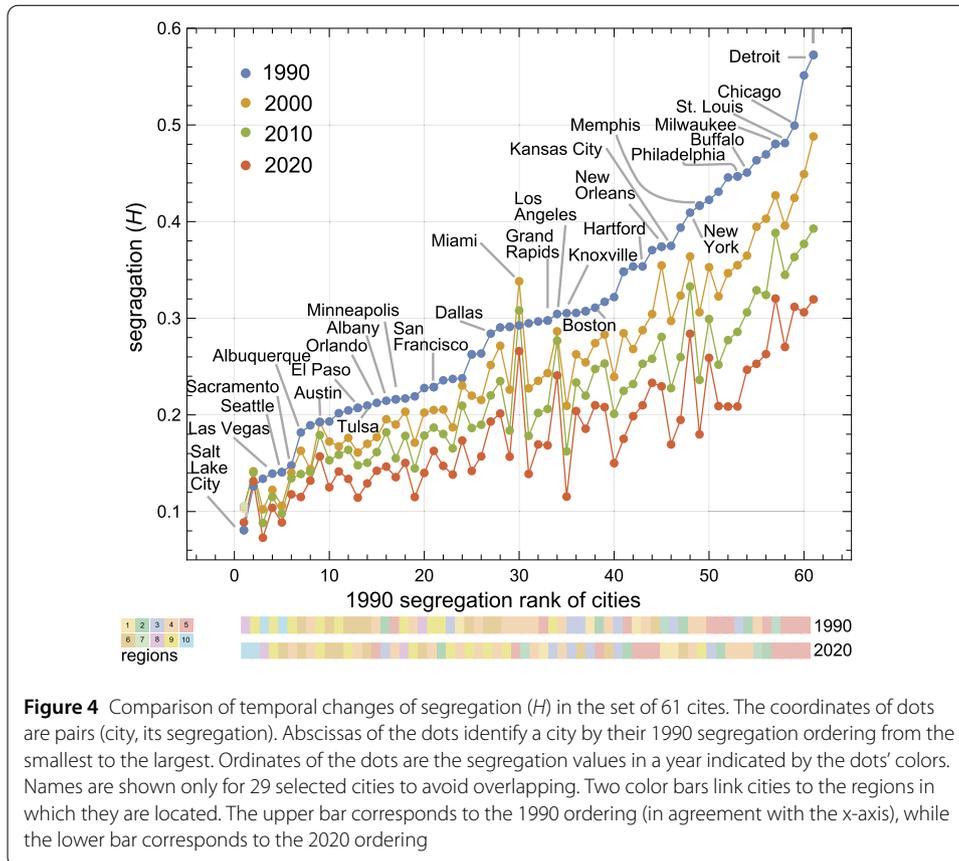
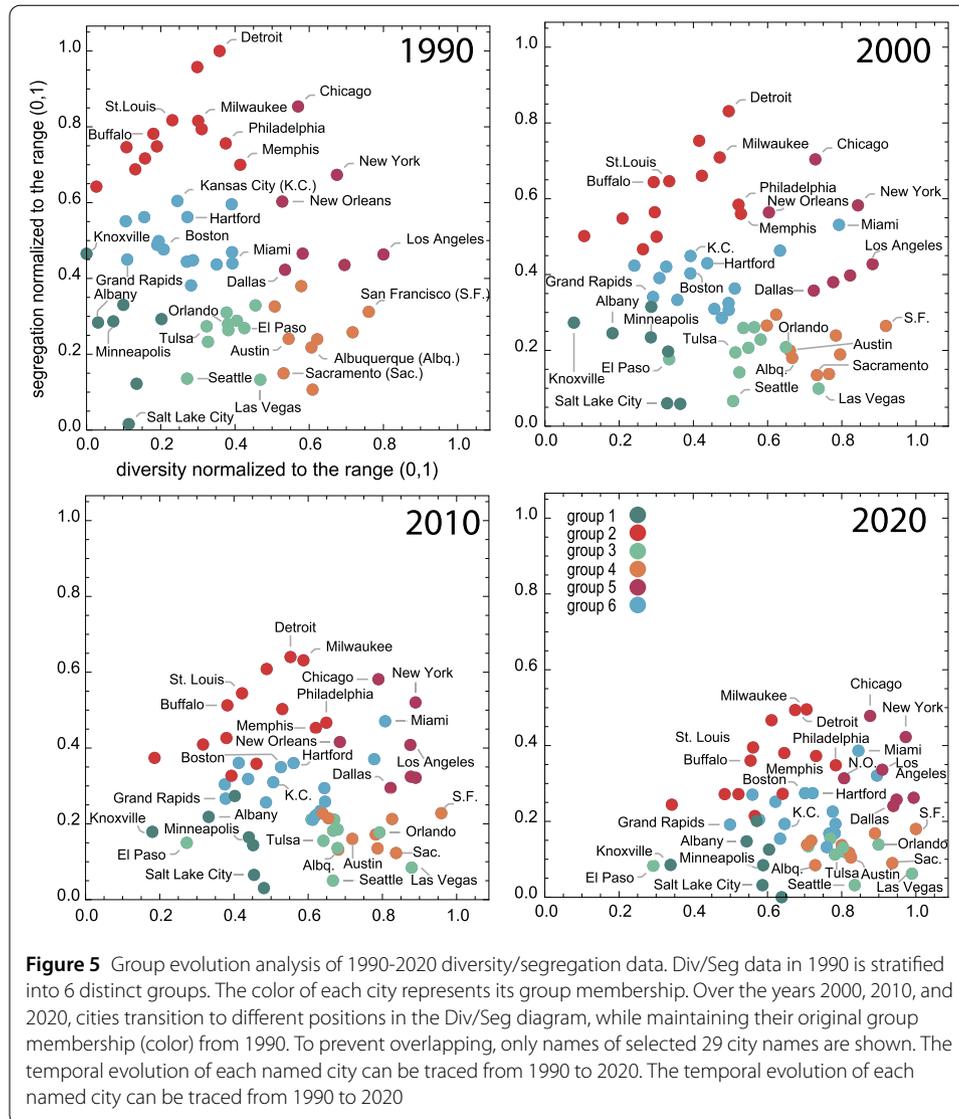


Figure 4 Comparison of temporal changes of segregation (H) in the set of 61 cites. The coordinates of dots are pairs (city, its segregation). Abscissas of the dots identify a city by their 1990 segregation ordering from the smallest to the largest. Ordinates of the dots are the segregation values in a year indicated by the dots' colors. Names are shown only for 29 selected cities to avoid overlapping. Two color bars link cities to the regions in which they are located. The upper bar corresponds to the 1990 ordering (in agreement with the x-axis), while the lower bar corresponds to the 2020 ordering

Figure 4 presents a comparison of segregation rankings. The construction of this figure follows the same approach as that of Fig. 3: cities are ranked based on their 1990 segregation values (measured by H). The figure illustrates that, in general, segregation has been decreasing nationwide over time. However, there is a notable exception in Miami, FL, where segregation increased from 1990 to 2000 before starting to decrease. On average (\pm standard deviation), there is a decrease in segregation between consecutive censuses: $\Delta H = -0.05 \pm 0.03, -0.04 \pm 0.02, -0.04 \pm 0.02$ for the periods 1990-2000, 2000-2010, and 2010-2020, respectively. The average decrease in segregation during the entire 1990-2020 period is $\Delta H = -0.12 \pm 0.07$. It is worth noting the relatively large standard deviations in all years, indicating varying rates of desegregation across different cities. This variability contributes to the zig-zagged pattern observed in the chains of dots representing 2000, 2010, and 2020 in Fig. 4.

The analysis of Figs. 3 and 4 yields the following findings:

- Racial diversity in US cities has exhibited a consistent upward trend from 1990 to 2020. On average, there has been an increase of approximately one significant racial group ($N_H \sim 1$) in the population of an average US city over the course of 30 years.
- Racial segregation in US cities has shown a consistent downward trend from 1990 to 2020. On average, there has been a decrease of $H = -0.12$ in segregation over the 30-year period. The magnitude of this decrease in the index H does not have a straightforward intuitive interpretation.



3.3 Groups evolution perspective

In this analysis, our focus is on the temporal evolution of city groups rather than individual cities. Each group consists of cities that exhibited similar characteristics in the (N_H, H) space in 1990. The objective is to investigate whether these groups maintain their coherence and distinctiveness over time.

The initial distribution of the data in the (N_H, H) space in 1990 is depicted in the upper-left panel of Fig. 5. Without considering the color labels, it can be observed that the 1990 data points are evenly dispersed across the (N_H, H) space, lacking any discernible inherent structures. Nevertheless, we proceed with the stratification of the 1990 data. Stratification involves dividing the dataset into approximately homogeneous subsets or groups based on specific criteria. In this case, the criteria are the similarity of the diversity and segregation metrics, N_H and H .

To stratify the 1990 data, we utilize the k -means algorithm [30] with the number of groups set to 5, 6, or 7. Given the absence of inherent structure in the (N_H, H) space, the choice of k is arbitrary. However, we explore different values of k to ensure the robustness

Table 2 Temporal change of the values of inhomogeneity and silhouette metrics in 6 1990-defined groups

Group	Group description	Cites	1990	2000	2010	2020
			<i>Inh.</i> <i>Sil.</i>	<i>Inh.</i> <i>Sil.</i>	<i>Inh.</i> <i>Sil.</i>	<i>Inh.</i> <i>Sil.</i>
1 (dark green)	low diversity and low segregation	Providence, Albany, Knoxville, Greenville, Minneapolis, Salt Lake City, Portland	0.2 0.26	0.18 0.29	0.17 0.33	0.15 0.34
2 (red)	low diversity and high segregation	Buffalo, Pittsburgh, Philadelphia, Memphis, Louisville, Birmingham, Milwaukee, Indianapolis, Detroit, Dayton, Cleveland, Cincinnati, St. Louis	0.19 0.39	0.21 0.28	0.21 0.17	0.19 0.10
3 (light green)	medium diversity and low segregation	Virginia Beach, Raleigh, Orlando, Tulsa, Oklahoma City, El Paso, Denver, Phoenix, Las Vegas, Seattle	0.11 0.51	0.15 0.16	0.18 -0.16	0.19 -0.30
4 (orange)	high diversity and low segregation	San Antonio, Austin, Albuquerque, Tucson, San Francisco, San Diego, Sacramento, Riverside, Fresno	0.16 0.31	0.15 0.22	0.14 0.12	0.13 0.08
5 (dark red)	high diversity and moderate to high segregation	New York, Washington, Chicago, New Orleans, Houston, Dallas, Los Angeles	0.24 0.14	0.21 0.22	0.17 0.29	0.14 0.37
6 (blue)	low diversity and medium segregation	Hartford, Boston, Rochester, Richmond, Harrisburg, Tampa, Nashville, Miami, Jacksonville, Greensboro, Charlotte, Atlanta, Grand Rapids, Columbus, Kansas City	0.15 0.35	0.19 0.10	0.18 -0.05	0.16 -0.14

of our findings. Although the three stratifications result in some variations in the grouping of cities, all three approaches yield the same overarching conclusions, as outlined at the end of this subsection. Figure 5 and Table 2 present the results for stratification with $k = 6$.

Table 2 summarizes the cities belonging to each group, and the short description for each groups assigned based on the segregation/diversity level. It is important to note that the stratification of cities is solely based on the values of N_H and H , and no information regarding population sizes or racial compositions is considered.

Over the years 2000, 2010, and 2020, cities transition to different positions in the Div/Seg diagram, while maintaining their original group membership (color) from 1990. To prevent overlapping, only selected city names are shown.

In the remaining three panels of Fig. 5, we depict the data from the years 2000, 2010, and 2020 in the (N_H, H) space while maintaining their 1990 group color labeling. The purpose of this representation is to track the temporal evolution of groups of cities that were initially similar in terms of N_H and H in 1990 within the (N_H, H) space. Figure 5 offers three key observations.

- 1 The entire dataset shows a shift towards the lower-right corner of the (N_H, H) diagram, reflecting the overall trend of increasing diversity and decreasing segregation.
- 2 As a result of this trend, all groups, except for group #4 (orange, highly diverse and low segregated cities), do not maintain their initially assigned characteristics (i.e., their positions in the (N_H, H) diagram) from 1990. Group #4 retains its characteristics because it was already identified as high-diversity/low-segregation in 1990, and there is no alternative location on the (N_H, H) diagram where it could be shifted by the overall trend.

3 The other groups experience shifts but mostly preserve their coherence. In 2020, cities within these groups occupy different positions on the (N_H, H) diagram compared to 1990, yet they remain similar to one another in terms of their diversity and segregation metrics, N_H and H . One exception is group #3, which “lose” one city (El Paso, TX) that deviates from the overall trend of increasing diversity.

Table 2 presents two metrics, namely inhomogeneity (*inh.*) and silhouette (*sil.*), which quantify the observations discussed in the preceding paragraph. The inhomogeneity measures the similarity of the cities in a group, and it can range from 0 to 1. The smaller the value of inhomogeneity, the more similar the cities are. On the other hand, the silhouette metric [31] assesses the distinctiveness of a given group compared to other groups. It ranges from -1 to 1 , with larger values indicating higher distinguishability. In our context, the silhouette metric measures the degree of similarity between cities within a group relative to their similarity with cities in other groups.

Each entry in Table 2 provides a quantitative assessment of the temporal evolution within a specific group. It includes the group’s identification number, member cities, and the values of the inhomogeneity metric (upper row) and silhouette metric (lower row) from 1990 to 2020. The inhomogeneity metric values for a given group do not exhibit systematic changes over time, supporting our observation of coherence preservation. The only group that shows a systematic increase in the inhomogeneity value is group #3, which includes El Paso, TX. The values of the silhouette metric demonstrate systematic changes over time. Specifically, they increase over time in group #1 (cities with low diversity and low segregation) and, particularly, in group #5 (highly diverse cities with moderate to high segregation). Consequently, in terms of the similarity of their member cities based on diversity and segregation metrics, group #5 (which includes many of the largest US cities) and group #1 become more distinguishable from other groups in 2020 compared to 1990. Conversely, values of the silhouette metric systematically decrease over time in the remaining groups.

Evaluation of our survey data from the perspective of group evolution reveals the following findings:

- Cities that were grouped together in 1990 based on their similarity in terms of diversity and segregation metrics continue to exhibit similarity in 2020. This indicates that the trend towards increasing diversity and decreasing segregation has impacted all cities within the 1990 groups in a similar manner. This finding is intriguing because the groups consist of cities from different regions of the US, with their only commonality in 1990 being the values of N_H and H . However, over the course of 30 years, the evolution of racial geography has influenced their N_H and H values in a similar fashion. One exception is group #3, where El Paso, TX, located on the US-Mexico border, has maintained its relatively low diversity due to its predominantly Hispanic population.
- The majority of groups identified in 1990 have lost their distinctiveness by 2020. The evolution of racial geography, characterized by increased diversity and decreased segregation, has compressed the (N_H, H) space into a smaller domain compared to 1990. As a result, the groups defined in 1990 now overlap on the (N_H, H) diagram. However, groups #1 and #5 are exceptions to this trend, as they not only maintained their distinctiveness but actually increased it in relation to the other groups. However, with different groupings ($k = 5$ or $k = 7$) such exceptions are absent.

3.4 Spatial scale limit of desegregation

The ultimate goal of desegregation would be to achieve city-wide subpopulation shares across all measurement scales. Imagine a city with equal shares of different subpopulation at the scale of the city. In an ideal scenario of perfect desegregation, a one-person-per-dot map [32] of a this city would look like a random noise of dots colored by the race of inhabitants. However, the reality is quite different, as the dot maps of actual US cities deviate significantly from this random noise pattern. Instead, they exhibit notable spatial auto-correlation, which is indicative of segregation (for example, refer to Fig. 1 in Dmowska et al. [19]).

Spatial auto-correlation of racial maps is most pronounced at the smallest available measurement scale, namely the scale of the census block. This phenomenon becomes evident when examining the diversity values of these blocks. On a racial level, blocks tend to exhibit a high degree of homogeneity. For instance, in 2010, the average diversity value (N_H) for urban blocks was only 1.28, whereas the average diversity value for urban tracts (larger units than blocks) stood at 2.90 [33]. In terms of population, census blocks typically range from a few hundred to a few thousand people, while census tracts are generally an order of magnitude more populous. It's important to note, however, that the size of the population does not directly impact the values of N_H and H . It should be noted that segregation metrics cannot be directly calculated for blocks. Calculating H metric required to divide area into subdivisions, and blocks are the smallest available subdivisions provided by the US Census. However, given their propensity to resemble monoracial enclaves, blocks serve as a lower limit for the scale of desegregation.

The aim of this analysis is to examine whether the lower limit on the scale of desegregation has weakened over the period from 1990 to 2020. A direct approach to this analysis would involve calculating the diversities of blocks at each of the four census years and comparing their values. However, in order to maintain consistency with the methodology used in Sect. 3.2, we employ an indirect approach that compares two different ways of calculating the segregation metric, H , for the entire city. One approach utilizes tracts as subdivisions of the city, while the other employs blocks as subdivisions. For each city in each census year, we calculate $\Delta H = H_b - H_t$, where the subscripts b and t refer to the division into blocks and tracts, respectively. The value of ΔH is guaranteed to be positive because tracts are more diverse than blocks (see Eq. (2)). However, if ΔH decreases over time, it suggests that the diversity of blocks is increasing relative to the diversity of tracts. In other words, the lower limit on the scale of desegregation is weakening.

Figure 6 displays the values of H_t (represented by blue dots) and H_b (represented by red dots) as a function of the rank of H_t for the 61 cities in each census year. It is worth noting that the abscissas of the red dots are identical to those of the blue dots; only the ordinates differ. In this graphical representation, the horizontal distance between a red dot and a blue dot represents the value of ΔH for a particular city. Consistently, the red dots are positioned above the blue dots as expected. However, a slight trend toward smaller values of ΔH over time is observed. This observation is quantified in Table 3.

Based on the analysis of Fig. 6 and Table 3, the following findings emerge.

- During the period from 1990 to 2020, there has been a slight decrease in the gap between segregation values calculated from blocks and tracts for cities. This suggests that blocks have experienced an increase in diversity relative to tracts, thereby weakening the lower limit of desegregation scale.

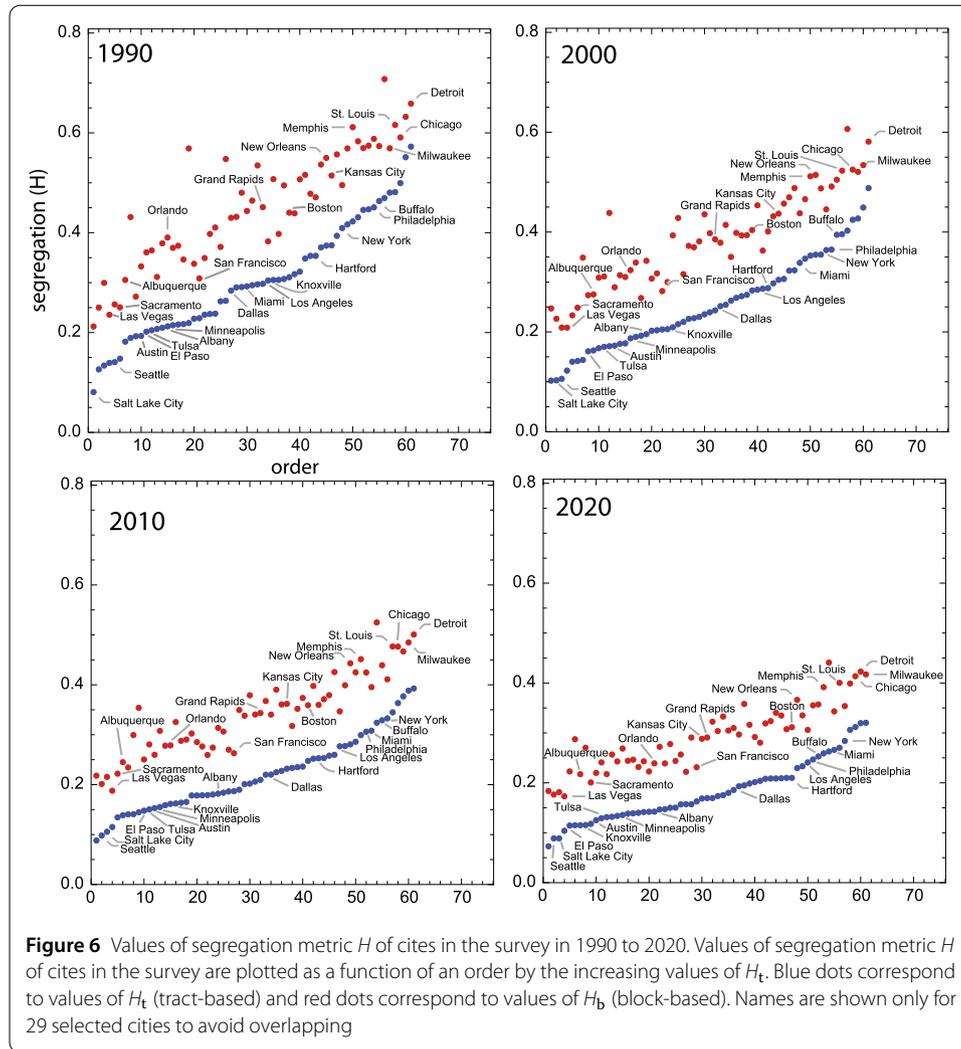


Figure 6 Values of segregation metric H of cities in the survey in 1990 to 2020. Values of segregation metric H of cities in the survey are plotted as a function of an order by the increasing values of H_t . Blue dots correspond to values of H_t (tract-based) and red dots correspond to values of H_b (block-based). Names are shown only for 29 selected cities to avoid overlapping

Table 3 Segregation: block-based versus tract-based

Year	$\langle \Delta H \rangle$	$SD(\Delta H)$	$Max(\Delta H)$	$Min(\Delta H)$
1990	0.146	0.050	0.35 (Greenville 0.57-0.22)	0.08 (Los Angeles 0.38-0.30)
2000	0.130	0.037	0.27 (Greenville 0.44-0.17)	0.08 (San Diego 0.27-0.19)
2010	0.120	0.030	0.21 (Greenville 0.35-0.14)	0.07 (Los Angeles 0.35-0.28)
2020	0.110	0.025	0.18 (Birmingham 0.35-0.14)	0.06 (San Diego 0.22-0.16)

- The standard deviation of segregation gaps among the surveyed cities has significantly decreased throughout the 1990-2020 period. This indicates that the differences in segregation gaps have become more uniform across cities, suggesting a convergence in the patterns of racial segregation.
- In 1990, cities in region 4 (Southeast) exhibited the highest values of ΔH , indicating greater disparities between block-based and tract-based segregation measures. Conversely, cities in regions 9 (California) and 5 (Midwest) had the smallest values of ΔH , indicating lower differences between the two measures. By 2020, the regional disparities in ΔH had diminished to some extent. However, cities in region 4 still

displayed relatively high values of ΔH , while cities in region 9 (excluding region 5) continued to show relatively low values of ΔH .

4 Conclusions and discussion

Residential racial segregation is a significant topic in American urban studies [34]. Sociologists generally associate racial segregation with racial inequality [35–37], and a decrease in segregation is seen as an indicator of social progress. Consequently, following each US decennial census, comparisons between the latest and previous segregation data are conducted to assess the state of this aspect of social progress (see references in the [Introduction](#)).

In our study, we have conducted such analyses using the most recent 2020 US Census data, comparing it with data from the 1990, 2000, and 2010 US Censuses. To the best of our knowledge, this is the first comprehensive comparison of multigroup metrics of urban segregation and diversity that encompasses all four recent censuses. Our paper is organized in a way that presents specific conclusions from each investigative perspective in bullet lists at the end of Sects. 3.1, 3.2, 3.3, and 3.4. Therefore, we will not provide a detailed repetition of those conclusions here. Instead, we will compare our findings with those obtained in previous studies when there is overlapping context.

Elbers [38] published a concise 3-page paper presenting a graph depicting the temporal variation of the population-weighted average value of H over the 1990–2020 period. The analysis was conducted on a sample of 228 US cities. The findings from Elbers' study align with the results obtained from our temporal perspective analysis (Sect. 3.2), particularly when we consider the population-weighted averages of our segregation and diversity metrics. However, as emphasized in the [Introduction](#), relying solely on sample-averaged values of segregation and diversity metrics significantly compresses the data, limiting our ability to extract comprehensive information about temporal changes in residential racial configuration. To gain a more nuanced understanding, our visual analysis approach allows for the examination of trends for each city in the sample individually, as well as the trend for the entire sample as a whole. This enables us to capture a broader range of insights regarding residential racial dynamics over time.

Logan et al. [16] conducted a comprehensive evaluation of segregation change spanning the 1980–2020 period. However, their analysis primarily concentrated on binary segregation, specifically examining the segregation of a particular group from the rest of the population and the segregation between two individual groups. In a similar vein, Frey [15] examined changes in diversity and binary segregation over the 2000–2020 period. While these studies offer valuable insights, their focus differs from our temporal perspective analysis, which specifically emphasizes multigroup segregation. As a result, these studies serve as complementary investigations rather than direct comparisons to our findings.

A subset of the findings from our geographical perspective analysis (Sect. 3.1) can be compared to the study conducted by Bellman et al. [14]. In their work, they presented normalized values of E and H for the years 2000 and 2010, not only for their entire sample, but also for four sub-samples categorized by the geographical locations of the cities: Northeast, Midwest, South, and West. To facilitate the comparison, we aggregated our 2000–2010 results based on the standard federal regions into four groups: regions 1, 2, and 3 were combined to represent the Northeast, regions 5, 7, and 8 were combined for the Midwest, regions 4 and 6 were grouped for the South, and regions 9 and 10 were considered as the West. Subsequently, we recalculated the values of N_H in Table 1 to obtain

Table 4 Regional values of E and H : This study versus [14]

	This study			[14]		
	Year	E	H	Year	E	H
Northeast	2000	0.58	0.28	2000	0.64	0.47
	2010	0.66 (+14%)	0.25 (-11%)	2010	0.72 (+13%)	0.42 (-11%)
Midwest	2000	0.55	0.28	2000	0.55	0.52
	2010	0.63 (+15%)	0.24 (-14%)	2010	0.61 (+11%)	0.45 (-13%)
South	2000	0.68	0.235	2000	0.69	0.43
	2010	0.74 (+9%)	0.20 (-15%)	2010	0.76 (+10%)	0.37 (-14%)
West	2000	0.75	0.225	2000	0.75	0.34
	2010	0.79 (+5%)	0.19 (-15%)	2010	0.79 (+5%)	0.30 (-12%)

the values of normalized E . This allows for a rough comparison to the results reported by Bellman et al. The outcomes of this comparison are presented in Table 4.

The analysis of the data presented in Table 4 reveals a notable agreement between our study and the work conducted by Bellman et al [14]. The values and growth rates of diversity metrics during the 2000-2010 period are remarkably similar between the two studies. However, there is a difference in the values of segregation indices, with our study reporting smaller values compared to those listed by Bellman et al. This disparity can be attributed to the fact that the two studies employ different measurement scales for assessing segregation. Our use of a larger measurement scale naturally results in smaller segregation index values. It is important to note that despite the variation in segregation index values, the percentage decline rates of segregation during the 2000-2010 period are very similar between the two studies. It should be acknowledged that Bellman et al. did not investigate changes during the 1990-2000 or 2010-2020 periods.

Our study encompasses two additional investigative perspectives, namely groups evolution (Sect. 3.3) and desegregation scale limit (Sect. 3.4), which, to the best of our knowledge, have not been explored previously. The group evolution analysis revealed that in 2020, the groups of cities established in 1990 maintained their group coherence but experienced a loss of distinctiveness. This loss of distinctiveness can be attributed to the overall trend observed during the 1990-2020 period, which was characterized by increasing diversity and decreasing segregation. Thus, the cities belonging to the 1990 groups with low diversity and high segregation shifted towards the higher diversity and lower segregation sector of the (N_H, H) diagram. Meanwhile, the cities from the 1990 groups with higher diversity and lower segregation remained relatively unchanged as the value of the diversity is constrained by the number of groups considered in the census. This results in an overlap of the 1990 groups in 2020. The desegregation scale limit analysis aims to evaluate changes in the “texture” of segregation. The results indicate that the texture of segregation is becoming “finer” as census blocks are losing their monoracial character. This desegregation scale limit perspective analysis is another unique contribution of our study.

Overall, our study reveals that over the course of three decades following 1990, the residential composition and spatial distribution of racial subgroups in US cities have exhibited a consistent trend: an increase in diversity accompanied by a decrease in segregation. However, it is important to note that the rates of change for these two metrics varied across different regions of the US (refer to Table 1). Generally, cities that were already diverse and relatively desegregated in 1990 exhibited slower rates of change, while those with lower initial diversity and higher levels of segregation experienced more rapid changes.

This finding suggests the existence of thresholds for the maximum value of N_H , which can be attributed to the fact that the US census identifies only four sub-populations with significant shares. On the other hand, the observed lower limit of segregation is likely influenced by the individual choices of inhabitants. This hypothesis is supported by the results of our desegregation scale limit analysis, which indicate that desegregation has been slow to penetrate the finer spatial scales of sub-tracks (refer to Table 3). Spatially-explicit forecasting studies [39, 40] also support this trend, indicating that the trajectory of increasing diversity (in cities where diversity can still increase based on the available census data) and decreasing segregation is expected to continue until the year 2030.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-023-00408-3>.

Additional file 1. The file Supplementary material is a MS Excel spreadsheet that contains the values of segregation and diversity metrics calculated for 61 cities based on US census block and tract level data for 1990, 2000, 2010, 2020. This dataset was used to perform all analysis presented in the paper (XLS 23 kB)

Acknowledgements

Not applicable.

Availability of data and materials

Data are available as a supporting material to this paper (<https://osf.io/fqpx4/>).

Declarations

Competing interests

The authors declare that they have no competing interests. The authors have no relevant financial or non-financial interests to disclose.

Author contributions

Both authors contributed to the study conception and design. AD collected data and performed calculation of diversity and segregation indices. TS prepare Figures included in the paper and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Author details

¹Institute of Geoecology and Geoinformation, Adam Mickiewicz University, Bogumila Krygowskiego 10, 61-680 Poznan, Poland. ²Department of Geography and GIS, Space Informatics Lab, University of Cincinnati, 215 Braunstein Hall, 45221 Cincinnati, OH, US.

Received: 13 April 2023 Accepted: 27 July 2023 Published online: 16 August 2023

References

1. Yao J, Wong DW, Bailey N, Minton J (2018) Spatial segregation measures: a methodological review. *Tijdschr Econ Soc Geogr* 110(3):235–250
2. Theil H, Finizza AJ (1971) A note on the measurement of racial integration of schools by means of informational concepts. *J Math Sociol* 1(2):187–193
3. Reardon SF, Firebaugh G (2002) Measures of multigroup segregation. *Sociol Method* 32(1):33–67
4. Jost L (2006) Entropy and diversity. *Oikos* 113(2):363–375
5. Iceland J, Weinberg D, Steinmetz E (2002) Racial and ethnic residential segregation in the United States: 1980–2000. Technical report
6. Iceland J (2004) Beyond black and white metropolitan residential segregation in multi-ethnic America. *Soc Sci Res* 33:248–271. [https://doi.org/10.1016/S0049-089X\(03\)00056-5](https://doi.org/10.1016/S0049-089X(03)00056-5)
7. Fischer MJ (2003) The relative importance of income and race in determining residential outcomes in US urban areas, 1970–2000. *Urban Aff Rev* 38(5):669–696
8. Fischer CS, Stockmayer G, Stiles J, Hout M (2004) Distinguishing the geographic levels and social dimensions of US metropolitan segregation, 1960–2000. *Demography* 41(1):37–59
9. Farrell CR (2008) Bifurcation, fragmentation or integration? The racial and geographical structure of US metropolitan segregation, 1990–2000. *Urban Stud* 45:467–499. <https://doi.org/10.1177/0042098007087332>
10. Farrell CR, Lee BA (2011) Racial diversity and change in metropolitan neighborhoods. *Soc Sci Res* 40(4):1108–1123
11. Parisi D, Lichter DT, Taquino MC (2011) Multi-scale residential segregation: black exceptionalism and America's changing color line. *Soc Forces* 89(3):829–852

12. Lee BA, Iceland J, Farrell CR (2014) Is ethnoracial residential integration on the rise? Evidence from metropolitan and micropolitan America since 1980. In: Diversity and disparities: America enters a new century. Russell Sage Foundation, pp 415–456
13. Fowler CS, Lee BA, Matthews SA (2016) The contributions of places to metropolitan ethnoracial diversity and segregation: decomposing change across space and time. *Demography* 53(6):1955–1977
14. Bellman B, Spielman SE, Franklin RS (2018) Local population change and variations in racial integration in the United States, 2000–2010. *Int Reg Sci Rev* 41(2):233–255
15. Frey W (2022) A 2020 census portrait of America's largest metro areas: population growth, diversity, segregation, and youth. Brookings Mountain West
16. Logan JR, Stults BJ et al (2022) Metropolitan segregation: no breakthrough in sight. Technical report, Working Papers 22-14, Center for Economic Studies, U.S. Census Bureau
17. Tufte ER (1983) The visual display of quantitative information. Graphics Press, Cheshire, 197 pp.
18. Healy K, Moody J (2014) Data visualization in sociology. *Annu Rev Sociol* 40:105–128
19. Dmowska A, Stepinski TF, Nowosad J (2020) Racial landscapes—a pattern-based, zoneless method for analysis and visualization of racial topography. *Appl Geogr* 122:102239
20. Elbers B, Gruijters R (2022) Segplot: A new tool for visualizing patterns of segregation. *SocArXiv*. <https://doi.org/10.31235/osf.io/rwv4g>
21. Manson S, Schroeder J, Van Riper D, Ruggles S (2021) IPUMS National Historical Geographic Information System: version 16.0 [dataset]. Technical report, Minneapolis, MN. IPUMS. <https://doi.org/10.18128/D050.V16.0>
22. Stepinski TF, Dmowska A (2019) Imperfect melting pot—analysis of changes in diversity and segregation of US urban census tracts in the period of 1990–2010. *Comput Environ Urban Syst* 76:101–109
23. Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* 54:427–432
24. Saeedimoghaddam M, Stepinski T, Dmowska A (2020) Rényi's spectra of urban form for different modalities of input data. *Chaos Solitons Fractals* 139:109995
25. Theil H (1972) Statistical decomposition analysis. North-Holland Publishing Co., Amsterdam,
26. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
27. Frankel DM, Volij O (2007) Measuring segregation. Technical report, Economics Working Papers (2002-2016) 180, Iowa State University
28. Mora R, Ruiz-Castillo J (2008) A defence of an entropy based index of multigroup segregation. Technical report, Working Paper 07-76 Economics Series 45, Departamento de Economía, Universidad Carlos III de Madrid
29. Roberto E (2016) The divergence index: a decomposable measure of segregation and inequality. *arXiv preprint*. [arXiv:1508.01167](https://arxiv.org/abs/1508.01167)
30. Lloyd S (1982) Least squares quantization in PCM. *IEEE Trans Inf Theory* 28(2):129–137
31. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65
32. Dmowska A, Stepinski TF (2019) Racial dot maps based on dasymmetrically modeled gridded population data. *Soc Sci* 8(5):157
33. Dmowska A, Stepinski TF (2022) Improving assessment of urban racial segregation by partitioning a region into racial enclaves. *Environ Plan B: Urban Anal City Sci* 49:290–303
34. Angotti T, Morse S (2023) Zoned out!: race, displacement, and city planning in New York city. New Village Press
35. Logan JR (2013) The persistence of segregation in the 21st century Metropolis. SAGE Publications Sage CA, Los Angeles
36. Jargowsky PA (2018) The persistence of segregation in the 21st century. *Law Inequal* 36:207
37. Richardson R (2021) Racial segregation and the data-driven society: how our failure to reckon with root causes perpetuates separate and unequal realities. *Berkeley Technol Law J* 36:1051
38. Elbers B (2021) Trends in US residential racial segregation, 1990 to 2020. *Socius* 7:1–3
39. Stepinski TF, Dmowska A (2022) Machine-learning models for spatially-explicit forecasting of future racial segregation in US cities. *Mach Learn Appl* 9:100359
40. Kinkhabwala YA, Barron B, Hall M, Arias TA, Cohen I (2021) Forecasting racial dynamics at the neighborhood scale using Density-Functional Fluctuation Theory. *arXiv preprint*. [arXiv:2108.04084](https://arxiv.org/abs/2108.04084)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
