Measuring Race and Ethnicity With the 2020 Census Redistricting Data

Leslie Reynolds

Program on Applied Demographics

Cornell University

Program on Applied Demographics

Web: https://pad.human.cornell.edu   Email: PADinfo@cornell.edu

Introduction

Despite evolutions in the way we conceptualize, measure, and quantify race and ethnicity in the past two and a half decades, the standards for measurement of these concepts have not kept up. The 1997 minimum requirements as provided by the Office of Management and Budget (OMB) Standards includes two groups for ethnicity and a minimum of five options for race, which must be recorded using two separate questions (Table 1). Two or more races is an optional racial category that is encouraged for use, especially when write ins and multiple selections are included as options in the questionnaire. Other race is not mentioned as a required or optional category by the federal register notice (OMB, 1997). The 2020 Redistricting Data (P.L. 94-171) produced a snapshot of race and ethnicity in the United States unlike others before it. Since the 2010 Decennial Census, the count of people identifying as two or more races or Some Other race in the U.S. increased by 276% and 129%, respectively (Jones et al., 2021a). The uniqueness of these results is generally credited to the improvements made in the design of the race and ethnicity questions, processing, and coding by the Census Bureau. Although the content likely differs greatly from previous Census years due to various methodological and social changes, the 2020 data is still being presented as a mirror of past Census years.

Table 1: 1997 Office of Management and Budget Standards for Federal Data on Race and Ethnicity[1]

Ethnicity		
Hispanic or Latino		Not Hispanic or Latino
A person from, or having heritage in, a Latin American or other Spanish speaking country.		A person with no identified Hispanic or Latino origins or heritage.
Race		
American Indian or Alaskan Native		Asian
A person with heritage or origins in any original peoples or tribes of North and South America (including Central America) and who maintains tribal affiliations.		A person originating or descending from any original peoples of the Far East, Southeast Asia, or the Indian subcontinent.
Black or African American	Native Hawaiian or Other Pacific Islander	White
A person having origins or heritage in any black racial groups of Africa. ''Haitian'' or ''Negro'' are also acceptable terminologies.	A person having origins or heritage in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.	A person originating in or descending from any original peoples of Europe, the Middle East, or North Africa.

Observed increases in reported identities other than non-Hispanic White alone have been accompanied by discussions of increasing racial and ethnic diversity in the United States. Colloquially, diversity refers to observed differences in a characteristic such as race, ethnicity, or gender. As a measurement, diversity is generally expressed using the diversity index, which describes the likelihood that two people randomly selected from a given area will be of a different race and ethnicity from one another. Both informally and officially though, talks of racial and ethnic diversity are reaching the same conclusions- America is becoming more diverse. However, with the noted changes in how race and ethnicity is measured by the Census, it is difficult to determine whether

---

[1] Measures were adapted from the 1997 Office of Management and Budget Standards, outlined in Federal Register Notice 62: 210 (OMB, 1997).

growth in diversity is truly driven by population change, or by the increased flexibility in individual race reporting and response coding and processing than in previous years. Moreover, methods of measuring diversity vary greatly across organizations and the literature. Thus, consequently, diversity estimates are often not equivalent to one another though they theoretically express the same concept. Therefore, the purpose of this paper is three-fold. First, we detail the changes in Census race reporting from 2010 to 2020 and discuss the lack of comparability offered by the recent counts. Second, we describe and compare multiple methods of calculating diversity indices and their substantive meaning. Thirdly, we illustrate estimates of race and ethnicity in New York State and discuss nuances by ethnicity and geographic area. All calculations for this paper were conducted using the 2020 Census Redistricting data (P.L. 94-171) released in August of 2021.

Changes in Census Race Reporting from 2010 to 2020

From 2010 to 2020, there were multiple integral changes made by the Census Bureau regarding how responses on race and ethnicity were collected, coded, and processed (Jones et al., 2021a). On the surface, the two Decennial Census years appear largely the same. In accordance with the previously outlined OMB Standards for Race and Ethnicity (displayed in Table 1), the Census Bureau asks separate questions for ethnicity and race, and classifies written responses based on this guidance. Despite research and recommendations from the bureau regarding a combined race and ethnicity question and the inclusion of an ethnic category for people of Middle Eastern/North African (MENA) origins or descent, lack of official action on the part of the OMB before the decennial Census count resulted in the continued use of the previous classifications for 2020 (U.S. Census Bureau, 2021). However, even without these changes, significant differences are still observed between the two questionnaires (Figures 1-3), the most notable of which are the provided write-in spaces for racial origins. Write-in responses were requested of all racial groups in 2020, compared with just a select number of groups

Figure 1: 2010 Race and Ethnicity Questionnaire

in 2010 (see Figure 1). Although selecting more than one race has been allowed since 2000, the impact of this request combined with the corresponding coding and processing changes (discussion to follow) is exemplified by the tremendous growth in the multiracial population between 2010 and 2020. In New York state for example, the number of people identifying as more than one race seemingly increased about 200% (from 590,182 to 1,767,463). Although most of this change is likely driven by the coding and processing of these responses, requesting write-ins of racial origins may be itself problematic due to the potential colloquial conflation of nationality with racial origins.

Figure 2: 2020 Ethnic Origin Question

Questionnaire differences only scratch the surface of the discrepancies in race and ethnicity reporting between the two most recent decennial censuses. Though the mere request of one's origins likely caused some problems regarding homogeneity of responses, lack of comparability between count years appears to stem from the coding and processing of responses. In efforts to consider the fluidity of personal racial and ethnic identities, the Census modified the way these responses were coded, including an expansion of how many characters could be captured. In 2010, the coding of the race and ethnicity questions was limited to two write-in responses and a maximum of 30 characters per line (Jones et al., 2021b). Because of this limit, Hispanic responses were prioritized in the coding of Hispanic origin, while racial and tribal responses were prioritized in

the coding of race. For example, if a respondent were to list Dominican, Spaniard, and Samoan in their write in box for ethnicity, the response of Samoan would be disregarded even though that may well be an important part of their ethnic identity. In 2020, coding limitations were relaxed such that the Census could accommodate up to six write-in responses and 200 characters per write-in line, with no prioritization of responses. However, the lack of response prioritization for the race and ethnicity questions may have led to the observed increase in those coded as White in Combination, Other in Combination, and Two or More races, especially among Hispanic/Latino people. In the estimates displayed in Tables 2 and 3 below, we use New York State as an example to verify the expected growth observed for each of these categories (though concluding the exact source of change is outside the scope of this paper). Non-Hispanic and Hispanic White alone were the only groups to see a decline in count between these years. Conversely, the groups of Hispanic White in Combination, Hispanic Other in Combination, and Hispanic Two or More Races each experienced relative increases over 300%, with surges in their raw counts all over 640,000. While the Hispanic and Some Other race alone group also experienced an equally sizeable increase in count, this only amounted to a 48% increase from its 2010 numbers.

Figure 3: 2020 Racial Origin Question

Table 2: Change in Racial/Ethnic Responses, 2010-2020

	Hispanic		Non-Hispanic	
	White Alone	White in Combination	White Alone	White in Combination
2010	1,436,727	183,559	11,304,247	230,741
2020	544,442	827,036	10,598,907	563,652
Absolute Change	-892,285	+ 643,477	-705,340	+ 332,911
Percent Change	-62.1%	350.6%	-6.2%	144.3%

Table 3: Change in Racial/Ethnic Responses, 2010-2020 cont.

	Hispanic			Non-Hispanic		
	Other Alone	Other in Combination	Two or More Races	Other Alone	Other in Combination	Two or More Races
2010	1,359,943	189,063	259,815	81,620	53,762	326,034
2020	2,013,526	974,292	1,046,616	197,107	240,630	720,847
Absolute Change	+ 653,583	+ 785,229	+ 786,801	+ 115,487	+ 186,868	+ 394,813
% Change	48.1%	415.3%	302.8%	141.5%	347.6%	121.1%

Arguably, the most impactful aspect of the 2020 Census racial and ethnic reporting to consider, especially when trying to compare it to previous estimates, is the coding and processing of responses for those who are now considered multiracial. For the current Census, write-in responses that matched with an existing racial category as defined by OMB standards (e.g. Black/African American, White, Asian etc.) were coded as belonging to that category. For example, one who indicated being of French and Middle Eastern origin would be coded as White alone since both origins are encompassed by the OMB category of White. However, when these responses do not fit within the same OMB category, the processing gets more complicated. For instance, since Hispanic is not a recognized racial category, any response associated with Hispanic or Latino origins in the question for race would be considered as Some Other race. Therefore, if one were to identify as White but write Argentinian in the box for racial origin, this person would be considered multiracial. Similarly, any write-in response for race that differed from those delimited by the OMB standards for that category led to a code of multiracial. These processing changes create two primary issues. First, the expected count of those considered two or more races would be expected to differ drastically from past years. Second, though often treated as such, the category of two or more races is not a homogeneous group but one created from numerous smaller groups. Therefore, the population being described as multiracial in these two Censuses are likely not similar and cannot be compared. Again viewing New York state as an example, we display the observed changes between 2010 and 2020 among the five largest multiracial groups in the state in Table 4. While the overall change for those of two or more races was already over 200%, those counted as White and some other race, and White, Black, and some other race both seemingly grew by over 500% (a six-fold increase). In general, those who expressed racial identities in combination with some other race experienced the largest growth. Like being multiracial, "some other race" is not a homogenous grouping, and the addition of those with Hispanic write-in responses for race likely only made this group even more diverse. Thus, researchers should proceed with caution when making generalizations across all groups in the Some Other Race and Two or More Race categories without further considering within group differences, and when comparing race and ethnicity estimates from the 2020 Census with estimates from previous years. With this in mind, the remainder of this paper will consider the 2020 data alone, rather than attempting to make comparisons.

Table 4: Change in New York's Largest Multiracial Groups, 2010-2020

Multiracial Category	2010	2020	% Change
Total: Two Or More Races	585,849	1,767,463	201.7%
White; Some Other Race	125,731	840,481	568.5%
Black or African American; Some Other Race	51,132	226,733	343.4%
White; American Indian and Alaska Native	41,342	113,950	175.6%
American Indian and Alaska Native; Some Other Race	12,869	33,617	161.2%
White; Black or African American; Some Other Race	4,813	31,594	556.4%

Measuring Diversity Using the 2020 Census

   Population counts and estimates can be sizably influenced by the measurement of concepts. Even a seemingly basic, yet integral topic such as diversity is difficult to quantify due to the subjectivity of measurement building and the importance of validity. Typically based on the diversity index, estimates of diversity usually express the probability or likelihood (ranging from 0 to 1) that two people randomly selected from a given area will be of a different race and ethnicity. Initially, measures of diversity were intended to summarize general variation in a characteristic within a specific population and to compare this variation over time and across areas. As diversity is becoming an increasingly relevant topic of discussion and reporting, however, it is important to consider the limitations and pitfalls of building and interpreting this measure. Though it is common to allocate respondents to selected categories based on one overarching characteristic, this practice places people within boxes that they may not necessarily belong to or would not put themselves in. Furthermore, in the mentioned index responses falling within a specific box are treated as being homogeneous to all other responses in that box- which is not the case. For example, within the Asian population, those who are Indian would likely have a different cultural and racial background than someone who is Chinese. Additionally, even within the racial group of White someone who is Arabic or Middle Eastern likely has a different cultural background than someone who is Brazilian or Italian. Similarly, all responses in a different box are considered completely disparate from other boxes, which is also likely not always true. For example, the lines between those who respond as White alone, Other alone, and the multiracial category White and Other are quite blurry. The diversity index, though flawed, is still an important descriptive measure and can be quite effective if designed and interpreted with care.

   There are multiple configurations that can be used to generate a diversity index for race and ethnicity. The substantive meaning of the estimates, however, may change depending on the formula and characteristic groupings used. Most calculations of racial and ethnic diversity in the field of social sciences utilize a form of Simpson's diversity index. This index, originally designed to measure biodiversity and ecological diversity, was adapted by Peter Blau in 1977 for use in the social sciences (Appendix B) (Blau, 1977; Rushton, 2008). The original Simpson's Index and the Blau adaptation produce the same results, given the categories are exhaustive of the population. However, if groups are selected such that the sum of their proportions is not one, then the results produced using these two formulas will be different. Even when using the same general formula, the configuration of racial and ethnic categories can influence the estimates produced. To illustrate the impact of these measurement decisions, we will display and discuss three different diversity estimates (all based upon the Blau formula). The categories used in the diversity index can be allocated in an infinite number of ways, but for the purposes of this paper we will examine the two most prevalent options, and a version that we developed. All methodology for our Diversity Index calculations can be found in Appendix B. The first index will be generated using the Census Bureau method. Their index expresses the probability that two randomly selected people in a given area will be of a different race and ethnicity and includes eight racial/ethnic categories. The next index we will display is a recent version used by the USA Today which calculates the probability that two randomly selected people in a given area will be of a different race *or* ethnicity. The final index is a fourteen-category index fully separated by ethnicity (denoted as PAD-14), designed as an expansion of the Census Bureau methodology.

Some key differences likely produce noticeable variation in the estimates. For example, the USA Today is the only index used here that treats ethnicity and race as separate concepts, and consequently does not utilize some other race as a group in their index. Conceptually, this is due to its "high error" from write-ins of Hispanic origins in the racial question and its overall collinearity with ethnicity (Meyer & McIntosh, 1992; Overberg, 2014). Therefore, the original index included just five racial categories and two categories for ethnicity, while members of the "Other" group were assigned to the other racial groups in proportion to their incidence in the population (Meyer & McIntosh, 1992). For our purposes, those who responded as Other race alone were allocated to the White category, while those reporting as one of the five races outlined by OMB standards in combination with Some Other race were placed in one of those five racial groups and removed from the category of two or more races. Though by design the original USA Today version also did not include Two or More races as a category (Meyer & McIntosh, 1992; Overberg, 2014), we include it in our calculations to maximize the comparability of this index with the others used here. The Census Bureau, unlike the USA Today version, does utilize the categories of Some Other race and Two or More races. However, this index assumes that those who are Hispanic feel more strongly about their ethnicity than their race, while those who are non-Hispanic prioritize their race. To alleviate this issue, our version of the diversity index includes all seven racial categories (including Some Other race and Two or More Races) each broken down by ethnicity. We posit that this method avoids the prioritization of race over ethnicity (or vice versa) among respondents, and instead allows for these groups to become as homogeneous as possible without making any undue assumptions.

In Table 5, we illustrate the three variations in calculated diversity for New York State and its economic regions. When using any of these indices, interpretation can be tricky. The index itself assumes that the relative size of any particular subgroup does not matter, and instead prioritizes general between-group heterogeneity (Rushton, 2008). For example, one could select two regions with similar values of the PAD index from the table (for example, Long Island and Mid-Hudson regions) and claim that they are approximately equal in their diversity. However, no conclusions can be made regarding which groups (or the sizes of those groups) create this diversity. Using the Census Bureau index for example, we can only conclude that if you select two people at random from either the Long Island or Mid-Hudson region, there will be around a 60% likelihood that they will be of different races and ethnicities. We could not, for instance, interpret that the share of White, Asian, or Black individuals is any larger or smaller among either region.
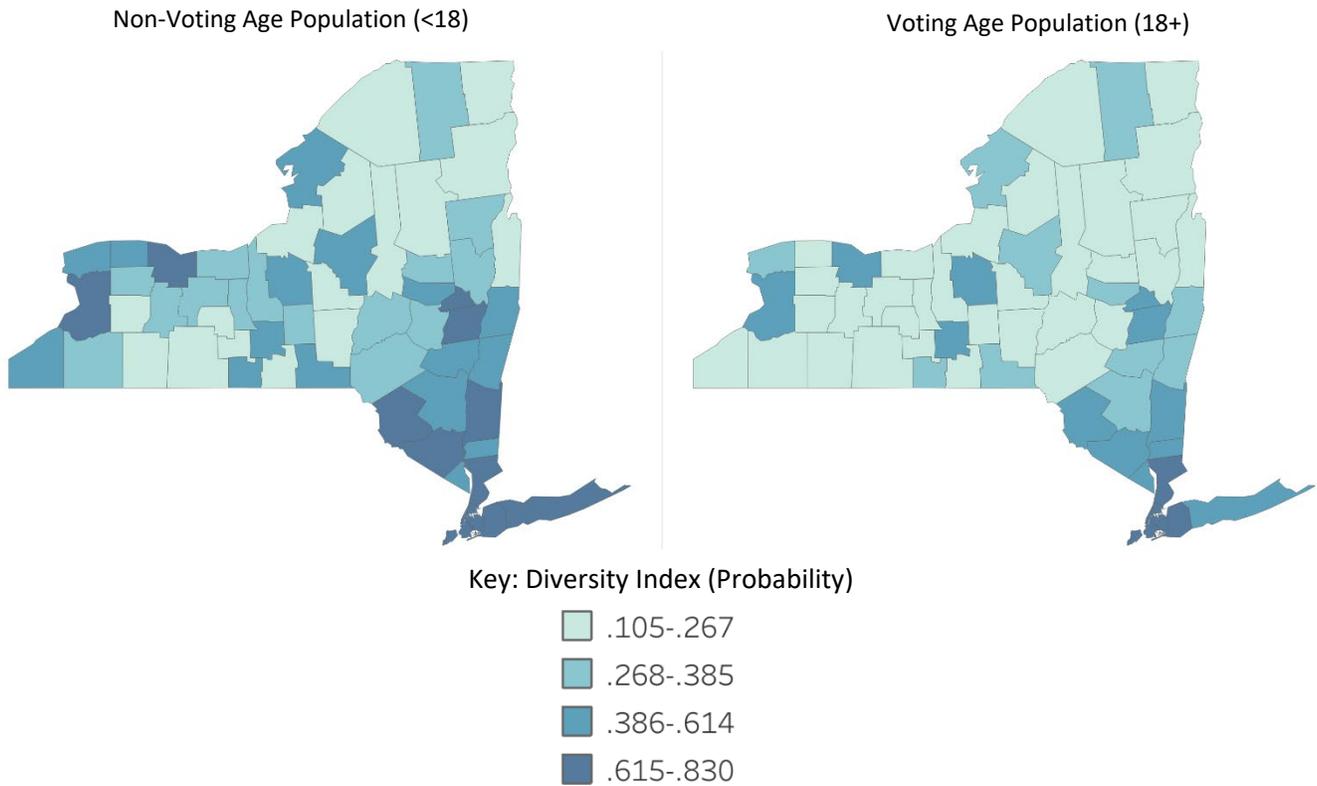
Table 5: Measuring Diversity Across New York, 2020

Region	Census Bureau	USA Today	PAD-14
New York State	0.658	0.638	0.682
Capital District	0.399	0.375	0.401
Central New York	0.363	0.345	0.365
Finger Lakes	0.427	0.411	0.431
Long Island	0.587	0.564	0.614
Mid-Hudson	0.599	0.577	0.630
Mohawk Valley	0.320	0.301	0.323
North Country	0.266	0.249	0.267
New York City	0.758	0.770	0.808
Southern Tier	0.301	0.281	0.303
Western NY	0.416	0.401	0.418

Keeping this in mind, there are clear differences between the values produced by these indices. While the Census Bureau calculation for New York State overall results in a .658 probability of selecting two people of a different race and ethnicity in New York State, this likelihood is found to be slightly larger using PAD-14 (.682). The USA Today index produces the lowest value of diversity for the state and has a slightly different interpretation; the probability of two randomly selected people in New York state being of a different race *or* ethnicity is .638 (about a 64% "chance"). The modest estimates of this index are likely due to the construction of its calculation, and the subtraction of the multiplicative probabilities of race and ethnicity from 1 (see Appendix B). Because of the greater number of categories used for PAD-14, larger estimates were expected. This method consistently produced the highest diversity estimates, although they were often quite similar to those produced using the Census Bureau index (e.g., .266 vs .267 in the North Country). In general, PAD-14 and the Census produced the largest estimates while the USA Today generated the smallest. One exception to these generalizations though is New York City, for which PAD-14 produced the largest estimate (.808), followed by the USA Today (.770), and the Census Bureau (.758). It may be that because New York City has such a wide range of racial and ethnic groups, PAD-14- the most exhaustive of the indices- picks up additional nuances of diversity that other methods do not. However, it is also possible that the PAD-14 index is overestimating the amount of diversity in NYC, either due to the methodology or some characteristic the city possesses compared to the rest of the state (i.e., a larger spread across both Non-Hispanic and Hispanic racial categories). Interestingly, despite these observed differences, index methodology does not seem to have an impact on the rank order of diversity by geographic area. For each index, New York City is the most diverse while the North Country is the least diverse. If your intent is to use these indices as designed (to compare diversity among geographic areas or over time) it matters less which index is used and more that you consistently use the same one.

Because of the significant changes in race reporting between 2010 and 2020, in this paper we will not compare changes in diversity over the past decade. However, splitting the diversity calculation by age still allows for a snapshot in differences in race reporting between generations. In Figure 4, we display county-level diversity within New York state for the voting age (18+) and non-voting age (< 18) populations. For these calculations we utilized the PAD-14 index so that we may capture the meaning and distribution of reported racial and ethnic identities as accurately as possible. Both maps were displayed using the same color scale, shown below.

Non-Voting Age Population (<18)

Voting Age Population (18+)

Key: Diversity Index (Probability)

- [ ] .105-.267
- [ ] .268-.385
- [ ] .386-.614
- [ ] .615-.830

From Figure 4, it appears that youth in 2020 were generally more diverse than the adult population in most counties. There were some areas of exception, however. Multiple counties in the New York City area (Bronx, Kings, New York, Queens, and Rockland) as well as St. Lawrence and Wyoming counties exhibited larger diversity estimates for those 18 and older than for their counterparts under 18 (Appendix A). This is a key finding because the general assumption is that youth in America are more diverse than adults (Jensen et al. 2021). While this tended to be the case, it is important to acknowledge the geographic variation in this estimate and consider the potential uniqueness of each population. Indeed, while the smallest diversity estimates for both groups were observed in Lewis County (.133 for youth, .079 for adults), the largest probabilities of diversity were found in Kings County for the non-voting age population (.744) and in Bronx County among adults (.784). Therefore, the most diverse population in New York state was found among adults in 2020, not youth. However, by excluding the New York City metropolitan area (Bronx, Kings, Nassau, New York, Putnam, Queens, Richmond, Rockland, Suffolk, and Westchester counties), the picture becomes more "typical". When the ten counties comprising the NYC metropolitan statistical area are removed, the diversity index ranges from .133-.683 for youth, and .079-.592 for adults, with the new maxima lying in Schenectady and Orange counties, respectively. Therefore, the most diverse areas for the voting age population lie in the New York city metropolitan area and thus the range of diversity among the voting age population becomes lower at both ends than that of the non-voting age population. The literal numerical significance of these estimates, though, are slim because of their limitations regarding relative group size and distribution. In the next section, we will discuss the racial and ethnic composition of New York, and how this impacts the interpretation of "diversity" throughout the state.

Racial and Ethnic Composition and the Meaning of "Diversity"

   Like many other statistics, the diversity index generates a single number without necessarily giving us a sense of what that estimate means. In this section, we will depict the racial and ethnic composition in various parts of the state to better visualize the distribution that makes New York state and its regions uniquely diverse. Like the rest of the United States, the majority of New York state's racial composition is non-Hispanic White (52.47%). However, this share makes up just over half the state's population and thus leaves room for considerable diversity. New York's next largest racial/ethnic group is non-Hispanic Black (13.66%), followed by Hispanic Other race (9.97%), and non-Hispanic Asian (9.49%) (shown in Figure 5).

Figure 5: Racial and Ethnic Composition of New York State
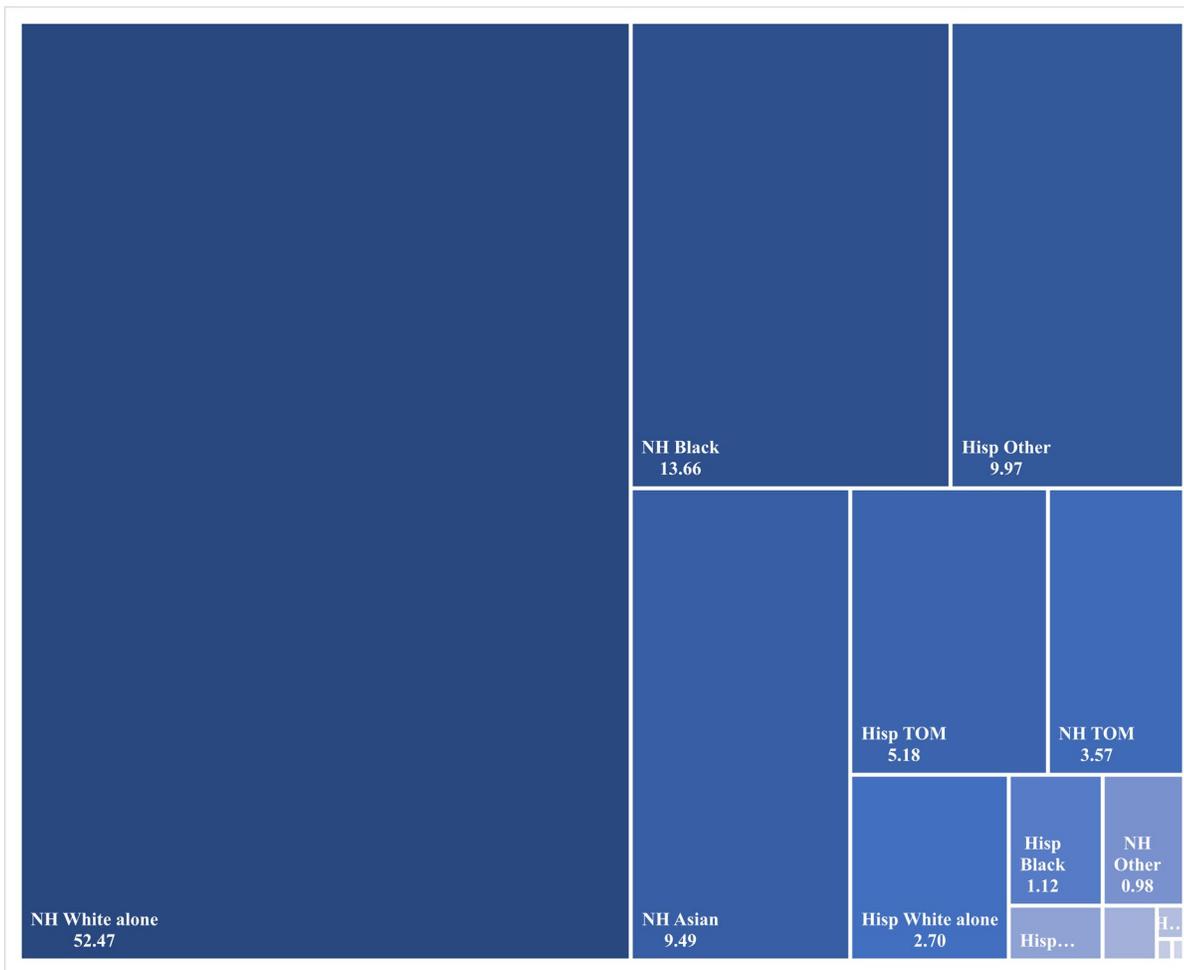
Not Labeled:	
Hispanic AIAN	0.47%
NH AIAN	0.27%
Hispanic Asian	0.08%
NH Pacific Islander	0.03%
Hispanic Pacific Islander	0.02%

The Hispanic and non-Hispanic Two or More race groups were the fifth and sixth largest racial/ethnic groups, at 5.18% and 3.57% respectively. The smallest racial and ethnic combinations observed for New York state were Hispanic and non-Hispanic Pacific Islander (.02% and .03%), followed by Hispanic Asian (.08%). When compared with the diversity estimates for the state, we once again observe the limitations of index calculations, especially regarding the degree of diversity being expressed. Table 6 illustrates the share of the three largest Hispanic and non-Hispanic racial and ethnic groups observed in each region of New York state along with the diversity estimate for that region (using the PAD-14 index). Most notably, Hispanic Other alone composes 15.6% of the NYC population and 10% of the state's population. Although White alone was a sizeable group among both non-Hispanic and Hispanic respondents, the other most common groups among each ethnicity do not necessarily coincide. For example, while Asian is a substantial group within the non-Hispanic population, it is more rarely combined with Hispanic identity. This table exemplifies the variation within each ethnic group that may be masked by conceptualizing all individuals of the same ethnicity into one homogeneous group, without considering potential racial, national, or cultural nuances. Since "Hispanic/Latino" is already a very heterogeneous group, it is even more important to split this group by race to better capture key subgroups. Similarly, having comparable diversity estimates does not mean that the racial composition of two regions look the same. For example, though the Long Island and Mid-Hudson regions both have probabilities of diversity around .62, the second most prevalent racial groups for these regions are Hispanic Other alone and non-Hispanic Black alone, respectively. Though the regions mirror each other in some ways, they are not directly comparable despite similar aggregate measures of diversity. Especially in areas with a higher diversity score (i.e., Long Island, New York City and Mid-Hudson regions), we observe higher shares of the Hispanic Other alone and Hispanic Two or More race groups, which are diverse groups in themselves. However, because New York City is a major cultural hub, the composition of the Two or More Races group is quite diverse from the rest of the state.

Table 6: Regional Diversity and Racial Composition Across New York State, 2020

Region	Diversity Index	NH White Alone	NH Black Alone	NH Asian Alone	Hispanic White Alone	Hispanic Other Alone	Hispanic Two or More Races
New York State	0.682	52.5%	13.7%	9.5%	2.7%	10.0%	5.2%
Capital District	0.401	76.7%	7.1%	4.4%	1.3%	1.6%	2.0%
Central NY	0.365	79.1%	7.7%	2.9%	1.2%	1.3%	1.6%
Finger Lakes	0.431	74.4%	10.2%	2.9%	1.7%	2.9%	2.4%
Long Island	0.614	59.8%	8.7%	7.8%	3.2%	9.7%	6.4%
Mid-Hudson	0.630	58.2%	10.6%	4.8%	3.4%	10.4%	6.5%
Mohawk Valley	0.323	82.0%	4.2%	2.8%	1.8%	2.1%	2.0%
North Country	0.267	85.4%	3.4%	1.1%	1.5%	1.0%	1.1%
New York City	0.808	30.9%	20.2%	15.6%	3.2%	15.6%	6.7%
Southern Tier	0.303	83.2%	3.6%	3.6%	1.2%	1.1%	1.4%
Western New York	0.418	75.3%	10.6%	3.5%	1.5%	1.8%	1.9%

In Table 7, we separate the Two or More races category into the most commonly reported racial combinations and separate the count for New York City (i.e., Bronx, Kings, New York City, Queens, and Richmond counties) from the state estimates to emphasize its uniqueness. Unsurprisingly, New York City alone accounts for over half of the state's multiracial population. Only three of the displayed multiracial groups are smaller in NYC than the state overall- White and Black, White and Some Other race, and White and American Indian/Alaskan Native. Higher observed counts of White and American Indian individuals for upstate New York versus New York City are

not unexpected, due to the relative share of American Indian reservations outside versus within the city. This same observation among the other two groups is more surprising, especially since White and some other race is a commonly processed identity among Hispanics and New York City houses the majority of Hispanic individuals in the state (about 63.1%). However, because New York City is so culturally and ethnically diverse it may be possible that those in the city more often report being of nationalities or origins that contrast with the standardized OMB categories and are thus processed as three or more races. Consistent with this argument, the count of individuals in NYC belonging to three or more races (including those in the White, Black, and Other category) includes over 14,000 more individuals than the respective count for the state. However, to confirm this notion and learn more about the within-group distribution of racial origins we would need access to the detailed write-in origin data from the 2020 Census. Even without the reported stratification by write-in origin, one can still observe the heterogeneity within the group of Two or More Races which is made up of over fifty sub-groups (only seven which are shown here). Therefore, regardless of group size it is still important to include and count as many groups as possible to reduce bias and amplify representation. By lumping "smaller" groups, still made up of tens or hundreds of thousands of people, into one homogeneous category we begin to blur or even erase diversity. Thus, methodological decisions to reduce groups and dimensions within the diversity index should be made with caution while trying to keep the number of and integrity of the groups reported as exact as possible. Moreover, interpretations of these measures should keep in mind the groups not being represented by these estimates.

Table 7: Count of Multiracial Groups in New York State and New York City

Combination of Races	New York City	New York State (excluding NYC)
Total: Two or more races	887,097	880,366
White; Some Other Race	386,774	453,707
Black or African American; Some Other Race	180,766	45,967
White; Asian	79,879	69,048
White; Black or African American	62,704	112,982
White; American Indian and Alaska Native	14,965	98,985
American Indian and Alaska Native; Some Other Race	22,721	10,896
White; Black or African American; Some Other Race	20,042	11,552
Other Three or More Races**	46,193	40,447

**Other three or more races includes all individuals reporting as three to six races, minus those reporting as White, Black and Some Other race which is displayed in this table.

Conclusions and Looking Forward

   Because of the efforts made by the Census Bureau to allow for more detailed self-reporting of race, the coding and processing of race and ethnicity reports were greatly altered. We wanted to write this paper to explain the extensive changes made between the Censuses and demonstrate the consequences for comparability across years as well as shed light upon assumptions made by quantifying such a fluid and complex concept as diversity. Although these changes were made with good intentions, they may have more substantial costs for interpretability and validity than expected, especially the alterations made to response processing. Additionally, the request of racial origins may offer the potential for more detailed information on racial identity, but the conflation of the terms origins and nationality may lead to problems especially in the self-identification of respondents. Therefore, the potential for response validity and misclassification renders these changes slightly problematic. We are looking forward to the release of the Demographic and Housing Characteristic (DHC) and the Detailed DHC files to learn more about the demographic and housing estimates for 2020 (particularly among detailed racial and ethnic groups) to see what will be done with this large intake of new information. We are also hoping to see more separation within these files by age group to better disentangle demographic trends. We hope that this paper serves as a guide to quantifying race and ethnicity using the 2020 Census Redistricting Data (P.L. 94-171) that will assist researchers and data users in producing the best estimates possible.

References

Blau, P. M. (1977). Inequality and heterogeneity: A primitive theory of social structure. The Free Press.

Jensen, E., Jones, N., Rabe, M., Pratt, B., Medina, L., Orozco, K., & Spell, L. (2021, October 15). *The chance that two people chosen at random are of different race or ethnicity groups has increased since 2010*. Census.gov. Retrieved December 7, 2021, from https://www.census.gov/library/stories/2021/08/2020-united-states-population-more-racially-ethnically-diverse-than-2010.html.

Jones, N., Marks, R., Ramirez, R., & Ríos-Vargas, M. (2021a, August 31). *2020 Census Illuminates Racial and Ethnic Composition of the Country*. U.S. Census Bureau. Retrieved December 7, 2021, from https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html.

Jones, N., Ramirez, R., & Marks, R. (2021b, September 28). *Improvements to the 2020 census race and Hispanic origin question designs, data processing, and coding procedures*. The United States Census Bureau. Retrieved December 7, 2021, from https://www.census.gov/newsroom/blogs/random-samplings/2021/08/improvements-to-2020-census-race-hispanic-origin-question-designs.html.

Krogstad, J. M., & Cohn, D. V. (2014, March 14). *U.S. Census looking at big changes in how it asks about race and ethnicity*. Pew Research Center. Retrieved December 7, 2021, from https://www.pewresearch.org/fact-tank/2014/03/14/u-s-census-looking-at-big-changes-in-how-it-asks-about-race-and-ethnicity/.

Meyer, P., & McIntosh, S. (1992). The USA today index of ethnic diversity. *International Journal of Public Opinion Research*, *4*(1), 51–58. https://doi.org/10.1093/ijpor/4.1.51

Office of Management and Budget (1997) Federal Register Notice 62:210. https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf.

Overberg, P. (2014, October 21). *Changing face of America: About This report*. USA Today. Retrieved December 7, 2021, from https://www.usatoday.com/story/news/nation/2014/10/21/diversity-index-data-how-we-did-report/17432103/.

Rushton, M. (2008). A note on the use and misuse of the racial diversity index. *Policy Studies Journal*, *36*(3), 445–459. https://doi.org/10.1111/j.1541-0072.2008.00276.x

Schilling, M. (2002). Measuring diversity in the United States. *Math Horizons*, *9*(4), 29–30. https://doi.org/10.1080/10724117.2002.11975161

U.S. Census Bureau. (2021a, October 18). *Research to improve data on race and ethnicity*. Census.gov. Retrieved December 7, 2021, from https://www.census.gov/about/our-research/race-ethnicity.html.

U.S. Census Bureau. (2021b, October 19). *Race and ethnicity in the United States: 2010 census and 2020 census*. Census.gov. Retrieved December 7, 2021, from https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html.

County	< 18	18+	Total	County cont.	<18	18+	Total cont.
New York State	0.748	0.663	0.682	Niagara	0.488	0.278	0.325
Albany	0.673	0.484	0.525	Oneida	0.547	0.332	0.383
Allegany	0.183	0.164	0.168	Onondaga	0.609	0.402	0.451
Bronx	0.770	0.787	0.786	Ontario	0.383	0.210	0.247
Broome	0.516	0.336	0.373	Orange	0.722	0.603	0.637
Cattaraugus	0.339	0.194	0.228	Orleans	0.386	0.254	0.283
Cayuga	0.334	0.215	0.240	Oswego	0.257	0.169	0.189
Chautauqua	0.454	0.262	0.305	Otsego	0.296	0.207	0.223
Chemung	0.426	0.275	0.309	Putnam	0.593	0.402	0.445
Chenango	0.240	0.141	0.162	Queens	0.829	0.810	0.815
Clinton	0.265	0.239	0.244	Rensselaer	0.546	0.350	0.392
Columbia	0.485	0.274	0.314	Richmond	0.743	0.620	0.650
Cortland	0.290	0.223	0.237	Rockland	0.575	0.609	0.600
Delaware	0.315	0.210	0.230	St. Lawrence	0.196	0.208	0.206
Dutchess	0.667	0.493	0.530	Saratoga	0.361	0.217	0.248
Erie	0.621	0.428	0.472	Schenectady	0.702	0.485	0.539
Essex	0.228	0.174	0.184	Schoharie	0.269	0.175	0.193
Franklin	0.349	0.327	0.333	Schuyler	0.202	0.121	0.136
Fulton	0.318	0.192	0.218	Seneca	0.295	0.233	0.247
Genesee	0.332	0.210	0.235	Steuben	0.262	0.163	0.185
Greene	0.421	0.291	0.315	Suffolk	0.702	0.535	0.574
Hamilton	0.252	0.106	0.127	Sullivan	0.644	0.498	0.533
Herkimer	0.266	0.146	0.172	Tioga	0.229	0.148	0.166
Jefferson	0.443	0.335	0.361	Tompkins	0.531	0.442	0.457
Kings	0.781	0.767	0.771	Ulster	0.583	0.386	0.425
Lewis	0.160	0.105	0.118	Warren	0.283	0.161	0.184
Livingston	0.292	0.199	0.217	Washington	0.227	0.198	0.204
Madison	0.256	0.178	0.194	Wayne	0.370	0.205	0.243
Monroe	0.674	0.482	0.528	Westchester	0.771	0.684	0.706
Montgomery	0.548	0.344	0.395	Wyoming	0.194	0.233	0.226
Nassau	0.733	0.625	0.650	Yates	0.173	0.121	0.133

Note: Diversity estimates were calculated using the fourteen category PAD index. Methodology for this index can be found in Appendix B.

Appendix B: Diversity Index Methodology

Simpson Diversity Index: $\qquad D_S = \sum_{i=1}^{k} p_i(1 - p_i)$

Blau (1977) Diversity Index: $\qquad D_B = 1 - \sum_j p_j^2$

Index Version	Calculation
Census Bureau	$1-(H^2 + NHW^2 + NHB^2 + NHAI^2 + NHAS^2 + NHO^2 + NHTOM^2)$
USA Today	$1-((W^2 + B^2 + AI^2 + AS^2 + PI^2 + TOM^2) * (NH^2 + H^2))$
PAD-14	$1-(NHW^2 + HW^2 + NHB^2 + HB^2 + NHAI^2 + HAI^2 + NHAS^2 + HAS^2 + NHPI^2 + HPI^2 + NHO^2 + HO^2 + NHTOM^2 + HTOM^2)$

Additional Methodological Notes:

- The racial categories for the Census Bureau and PAD-14 indices did not include those who were part of an OMB race in combination with Other.

  - For the USA Today Index, those counted as Other in combination with a specified OMB racial category were removed from the count of Two or More Races. Respondents coded as Other alone were included in the White alone category.

- In the above methodological chart, NH represents non-Hispanic while H represents Hispanic; the absence of either notation indicates that category was not separated by ethnicity.