

Improvements to Race and Ethnicity Coding in the 2020 Census

The National Academies of Sciences, Engineering, and Medicine

Nicholas Jones, Roberto Ramirez, and Rachel Marks

U.S. Census Bureau

September 28, 2021

U.S. Office of Management and Budget (OMB) Standards for Race and Ethnicity (1997)

OMB ethnicity categories:

- Hispanic or Latino
- Not Hispanic or Latino

OMB race categories:

- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- White

The Census Bureau is required by Congress to use the category “Some Other Race” as a sixth race category.

2020 Census

Separate Ethnicity Question

→ NOTE: Please answer BOTH Question 6 about Hispanic origin and Question 7 about race. For this census, Hispanic origins are not races.

6. Is this person of Hispanic, Latino, or Spanish origin?

- ☐ No, not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin – *Print, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.*

- Question designs for 2020 Census adhere to 1997 OMB standards for race and ethnicity.
- The Census Bureau did not use combined question format for collecting race and ethnicity.
- 1997 OMB standards require two separate questions for self-response.
- “Middle Eastern or North African” (MENA) category not used, but detailed MENA responses collected.
- Significant changes from 2010 Census questions for race and ethnicity.

Separate Race Question

7. What is this person's race?

Mark ☒ one or more boxes AND print origins.

- ☐ White – *Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.*

- ☐ Black or African Am. – *Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.*

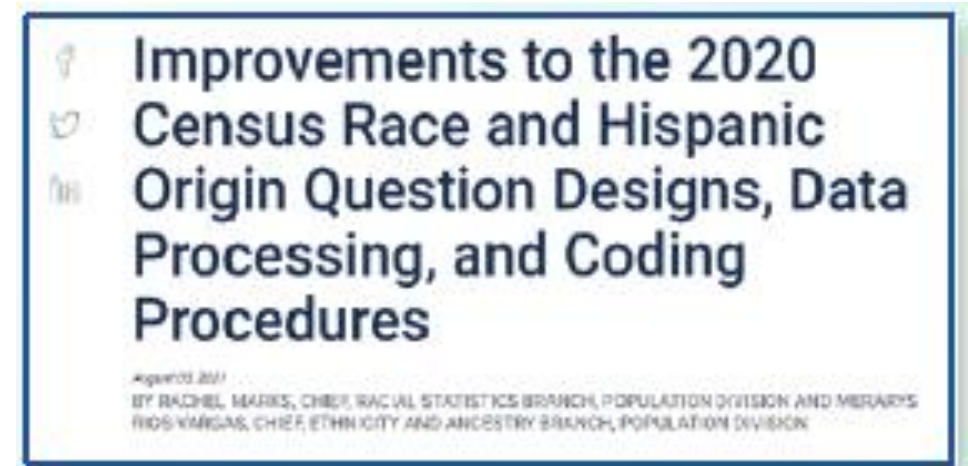
- ☐ American Indian or Alaska Native – *Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.*

- | | | |
|--|-------------------------------------|---|
| ☐ Chinese | ☐ Vietnamese | ☐ Native Hawaiian |
| ☐ Filipino | ☐ Korean | ☐ Samoan |
| ☐ Asian Indian | ☐ Japanese | ☐ Chamorro |
| ☐ Other Asian –
<i>Print, for example, Pakistani, Cambodian, Hmong, etc.</i> | | ☐ Other Pacific Islander –
<i>Print, for example, Tongan, Fijian, Marshallese, etc.</i> |

- ☐ Some other race – *Print race or origin.*

Improvements to 2020 Census Question Designs, Data Processing, and Coding Procedures

- Yield a more accurate portrait of how people self-identify
- Reveal the U.S. population is much more multiracial and more diverse



Improvements to the 2020 Census Hispanic Origin Question

- The instruction to “Print origin, for example” was revised to “Print, for example.”
- The example groups were revised from “Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.” to “Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.”

→ **NOTE:** Please answer BOTH Question 6 about Hispanic origin and Question 7 about race. For this census, Hispanic origins are not races.

6. Is this person of Hispanic, Latino, or Spanish origin?

- ☐ No, not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin – *Print, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.* ↗

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Improvements to the 2020 Census Race Question

- Write-in response areas were added for the White and Black or African American racial categories.
- Six examples were provided for each of the write-in fields allocated to the "White," "Black or African American," and "American Indian or Alaska Native" groups.
- The category "Black, African Am., or Negro" was changed to "Black or African Am." on paper.
- The detailed Asian and Native Hawaiian or Other Pacific Islander checkboxes were re-ordered by population size.
- The checkbox category "Guamanian or Chamorro" was changed to "Chamorro."
- The write-in instructions for the "Some Other Race" category were updated to "Print race or origin."

7. What is this person's race?

Mark ☒ one or more boxes **AND** print origins.

☐ White – *Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.* ↴

☐ Black or African Am. – *Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.* ↴

☐ American Indian or Alaska Native – *Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.* ↴

☐ Chinese	☐ Vietnamese	☐ Native Hawaiian
☐ Filipino	☐ Korean	☐ Samoan
☐ Asian Indian	☐ Japanese	☐ Chamorro
☐ Other Asian – <i>Print, for example, Pakistani, Cambodian, Hmong, etc.</i> ↴	☐ Other Pacific Islander – <i>Print, for example, Tongan, Fijian, Marshallese, etc.</i> ↴	

☐ Some other race – *Print race or origin.* ↴

Code List Improvements for the 2020 Census

- In the 2010 Census code list, many detailed White groups and detailed Black groups did not have unique codes
- We also added many new White and Black detailed groups that we collect in the ACS question on ancestry to the 2020 Census race code list
- For the 2020 Census, we conducted extensive research and consulted with experts to improve our code list
- The addition of these detailed codes gives us the ability to tabulate data for more detailed groups than we have in the past
- Plans for data products on detailed race and ethnicity groups are still under development

<u>EUROPEAN (EXCEPT SPANISH)</u>		<u>1010-1749</u>
Albanian		1010-1014
Albanian		
Not Used		
Alsatian	Motswana	3220-3224
Alsatian	Motswana (Botswana)	3220
Not Used	Not Used	3221-3224
Andorran	Alaska Native	5020-5499
Andorran	Alaskan Athabascan	5020
Not Used	Ahtna, Inc. Corporation	5021
	Alanvik	5022
	Alatna Village	5023
	Alexander	5024
	Allakaket Village	5025
	Anvik Village	5026
	Arctic Village	5027
	Beaver Village	5028
	Birch Creek Tribe	5029
	Chalkyitsik Village	5030
	Cheesh-Na Tribe (Chistochina)	5031
	Chickaloon Native Village	5032
	Circle Native Community	5033
	Cook Inlet	5034
	Copper River	5035
	Doyon	5036
	Eklutna Native Village	5037
	Evansville Village (Bettles Field)	5038
	Galena Village (Louden Village)	5039
	Gulkana Village Council	5040
	Healy Lake Village	5041
Armenian	Namibian	
Armenian	Namibian	
Not Used	Not Used	
Austrian	Nigerian	
Austrian	Not Used	
Tyrolean	Nigerian (Nigeria)	
Not Used	Yoruba	
	Igbo	
	Not Used	
	Nigerien	
	Nigerien (Niger)	
	Not Used	

The Autocoding and Residual Coding Processes

Automated coding

- Responses that had a match in the master file were automatically coded
- About 99% of write-ins were automatically coded

Residual coding

- Responses that could not be matched and automatically coded were “residual” responses
- Residual responses were sent to the National Processing Center (NPC) to be coded using the Residual Coding System (RCS)

Autocoding Rates and Total Number of Write-in Responses: 2010 and 2020

(M = millions)

Subject	2010 Census Number of Write-ins	2020 Census Number of Write-ins	2010 to 2020 Difference, Number of Write-ins
Race	37.7 M	335.5 M	297.8 M
Hispanic origin	17.1 M	15.0 M	-2.1 M
Total	54.7 M	350.5 M	295.8 M

Source: 2010 Census Internal Residual Coding System Reports and 2020 Census Internal Residual Coding System Reports

How Data on Hispanic Origin and Race are Processed and Coded in the 2020 Census Compared to the 2010 Census

2010 Census

- Limited to coding two write-in responses and 30 characters per write in line.
- Due to the limit, in the Hispanic origin question, Hispanic origin responses were prioritized over race responses.
- Due to the limit, in the race question, race and tribal responses were prioritized over Hispanic origin responses.

2020 Census

- Up to six write-in responses and 200 characters were coded per write in line.
- There was no prioritization of responses, all responses were treated equally.

We fully tested these coding and question changes in the 2015 National Content Test and finalized them in the 2018 Census Test.

2010 Census Data Collection Operation Captured Up to 30 Characters and Coded Up to Two Groups

MEXICAN AMERICAN INDIAN AND PORTUGUESE AND AFRICAN
AMERICAN

2020 Census Data Collection Operation Captured Up to 200 Characters and Coded Up to Six Groups

MEXICAN AMERICAN INDIAN AND PORTUGUESE AND AFRICAN
AMERICAN

Improvements for 2020 Census Hispanic Origin and Race Code List

2010 Census

- Used two separate code lists for race and Hispanic origin.
- Race code list included limited detailed codes for White, Black, and Hispanic groups.
- Hispanic origin code list included limited detailed codes for race groups.

2020 Census

- Used one combined code list for race and Hispanic origin.
- The combined race and Hispanic origin code list included thousands of detailed codes for all race and Hispanic origin groups.

We fully tested these coding and question changes in the 2015 National Content Test and finalized them in the 2018 Census Test.

Coding Rules for Hispanic Origin: 2010 Census and 2020 Census

- In the 2010 Census, if someone provided more than two write-in responses in the Hispanic origin question write-in area, we prioritized coding Hispanic groups over race groups or other types of responses.
- In 2020, there was no prioritization of responses, enabling all responses to be treated equally.

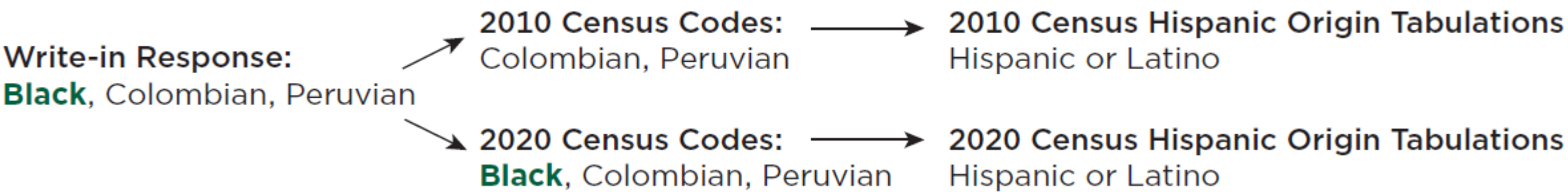
Coding Rules for Hispanic Origin Question: 2010 Census and 2020 Census

Write-in response	How it was coded in the 2010 Census	How it was coded in the 2020 Census
Black, Peruvian	Black <i>and</i> Peruvian	Black <i>and</i> Peruvian
White, Puerto Rican	White <i>and</i> Puerto Rican	White <i>and</i> Puerto Rican
Spaniard, White, Honduran	Spaniard <i>and</i> Honduran	Spaniard <i>and</i> White <i>and</i> Honduran
Mexican, Black, Colombian	Mexican <i>and</i> Colombian	Mexican <i>and</i> Black <i>and</i> Colombian

Note: 2010 Census coding rules prioritized Hispanic origin responses over race responses; 2020 Census coding rules did not.

Hispanic Origin Coding and Tabulation: 2010 Census and 2020 Census

Coding Impacts on Tabulation in the Hispanic Origin Question



Coding Rules for Race: 2010 Census and 2020 Census

- In 2010, if more than two groups were part of a write-in text string on the same line in the race question, we prioritized coding race groups over Hispanic origin groups because we were limited to only coding two responses.
- In 2020, there was no prioritization of responses, enabling all responses to be treated equally.

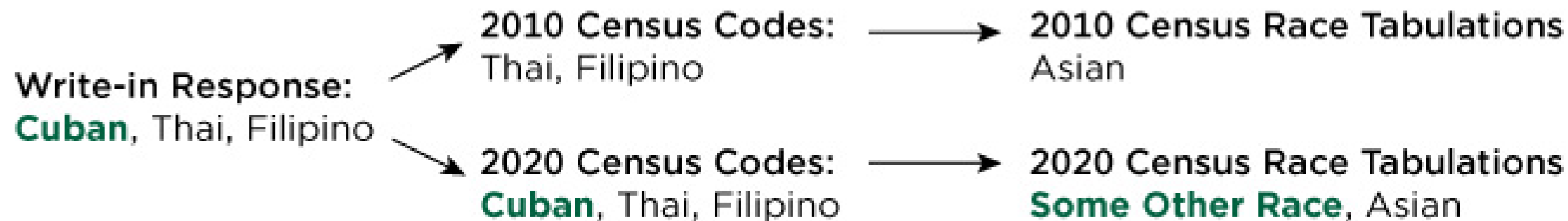
Coding Rules for Race Question: 2010 Census and 2020 Census

Write-in response	How it was coded in the 2010 Census	How it was coded in the 2020 Census
Hispanic, White	Hispanic <i>and</i> White	Hispanic <i>and</i> White
Black, Latino	Black <i>and</i> Latino	Black <i>and</i> Latino
Hispanic, White, Chinese	White <i>and</i> Chinese	Hispanic <i>and</i> White <i>and</i> Chinese
Spanish, Mexican, Samoan, Chamorro	Samoan <i>and</i> Chamorro	Spanish <i>and</i> Mexican <i>and</i> Samoan <i>and</i> Chamorro

Note: 2010 Census coding rules prioritized race responses over Hispanic origin responses; 2020 Census coding rules did not.

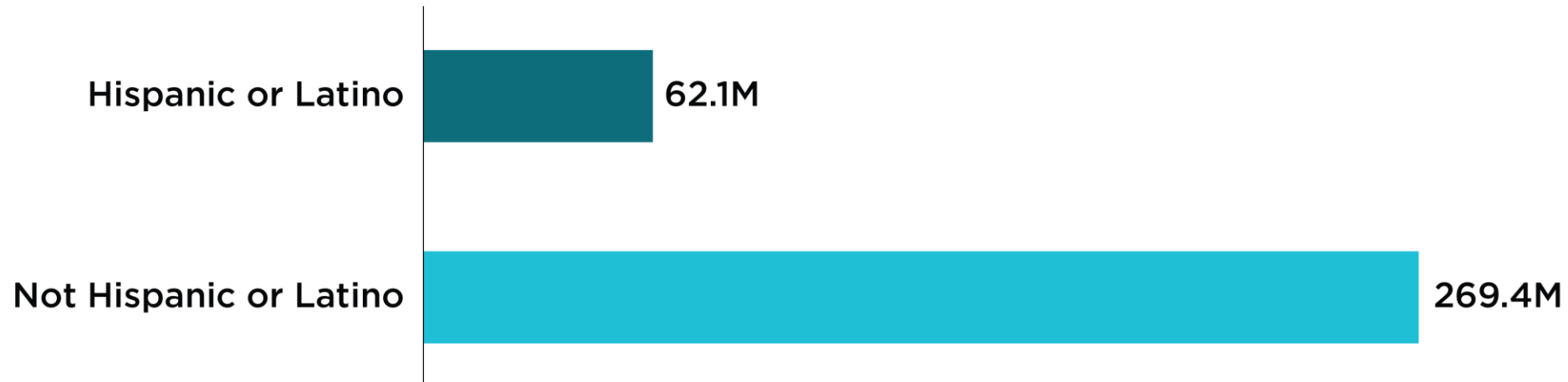
Race Coding and Tabulation: 2010 Census and 2020 Census

Coding Impacts on Tabulation in the Race Question



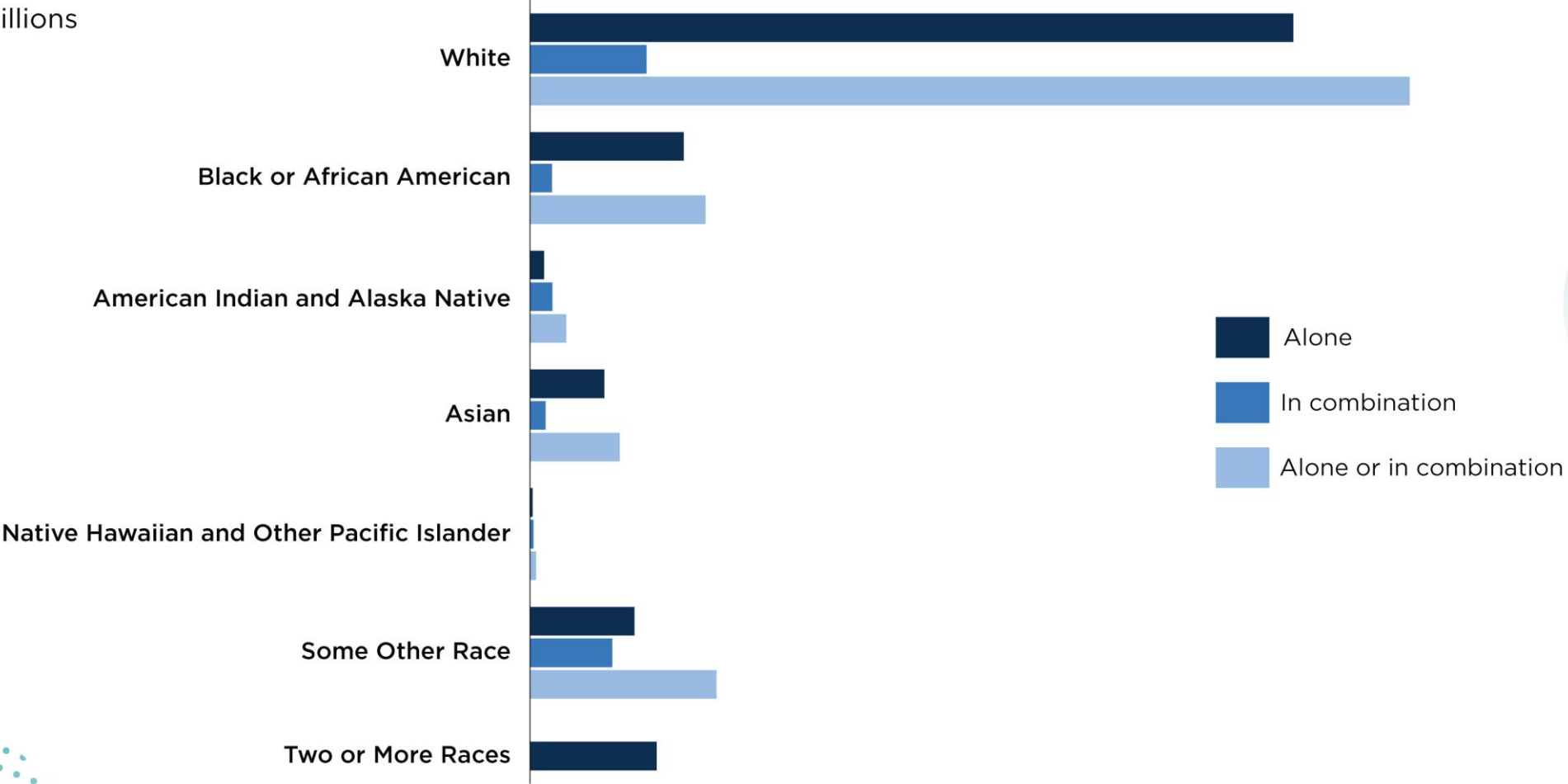
2020 Census Illuminates U.S. Racial and Ethnic Composition

M = Millions



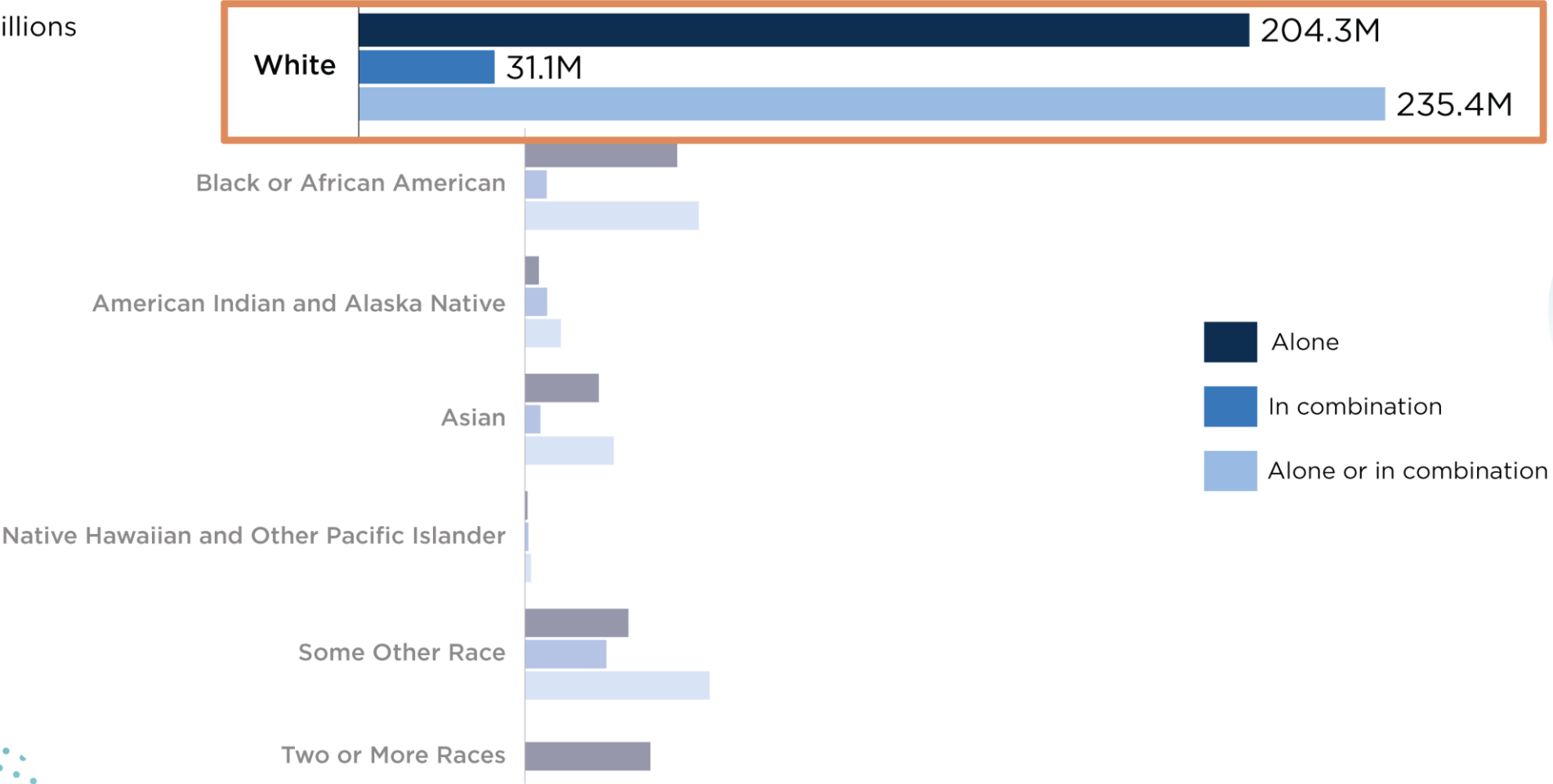
2020 Census Illuminates U.S. Racial and Ethnic Composition

M = Millions



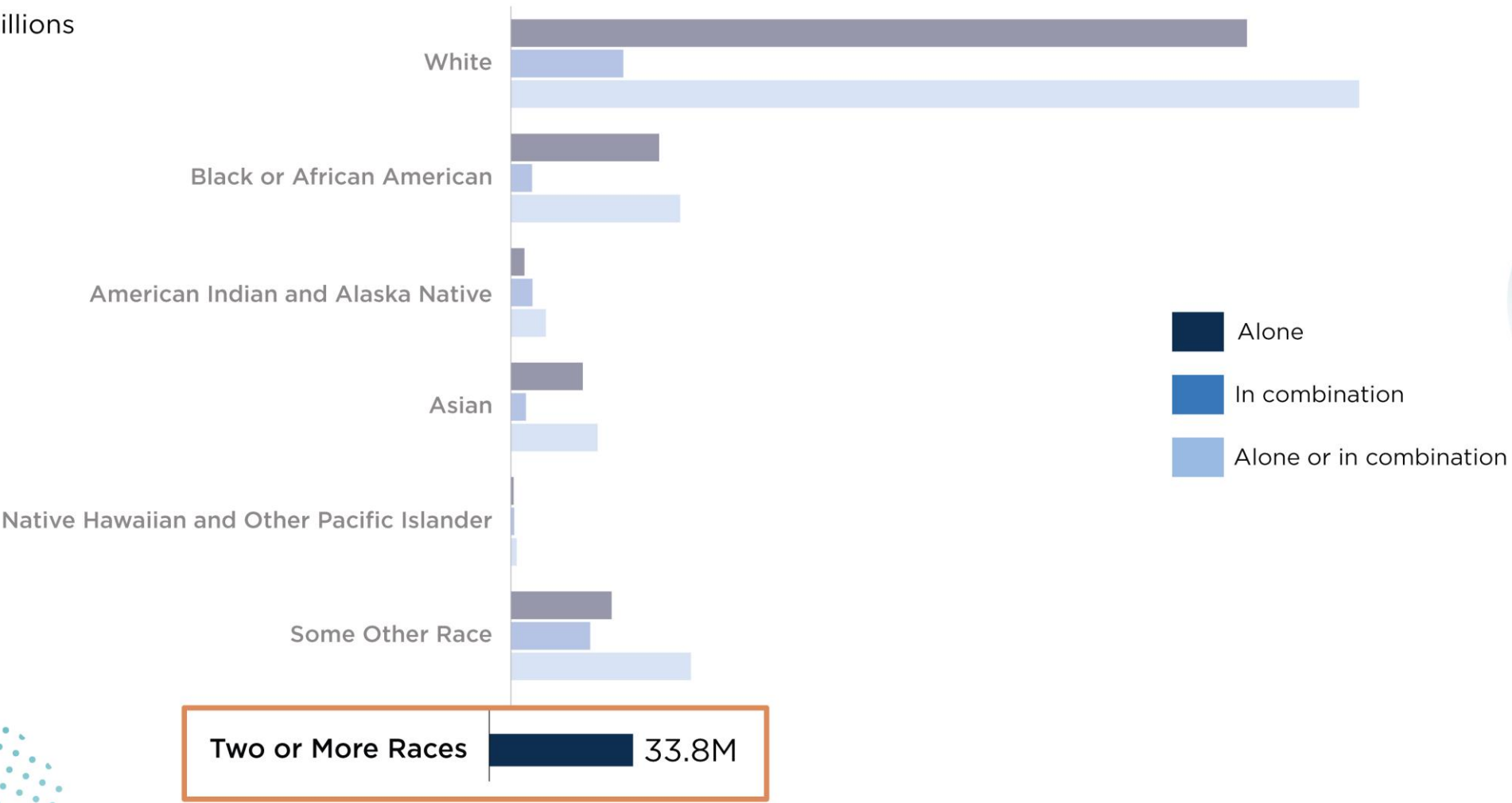
2020 Census Illuminates U.S. Racial and Ethnic Composition

M = Millions



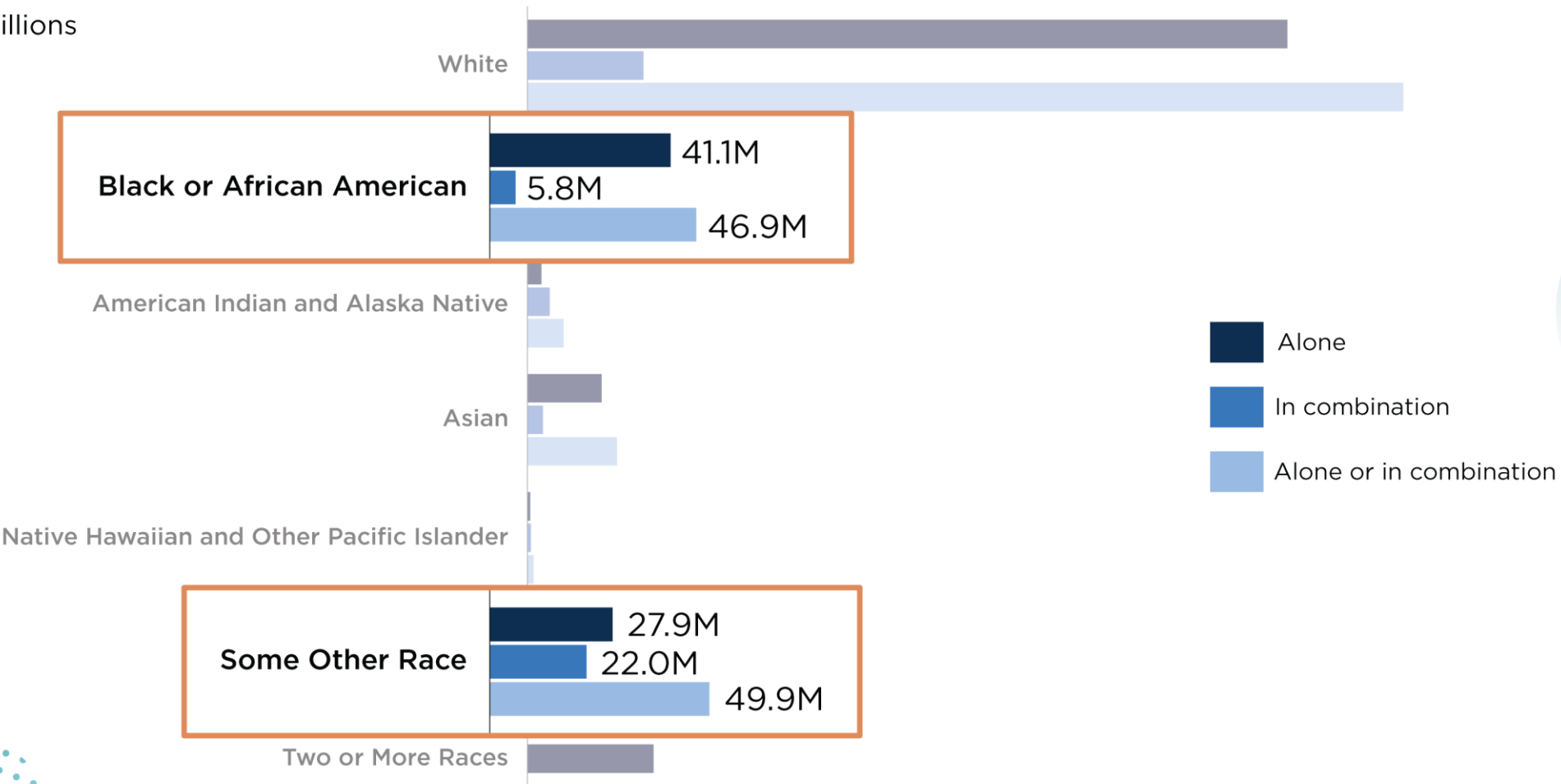
2020 Census Illuminates U.S. Racial and Ethnic Composition

M = Millions



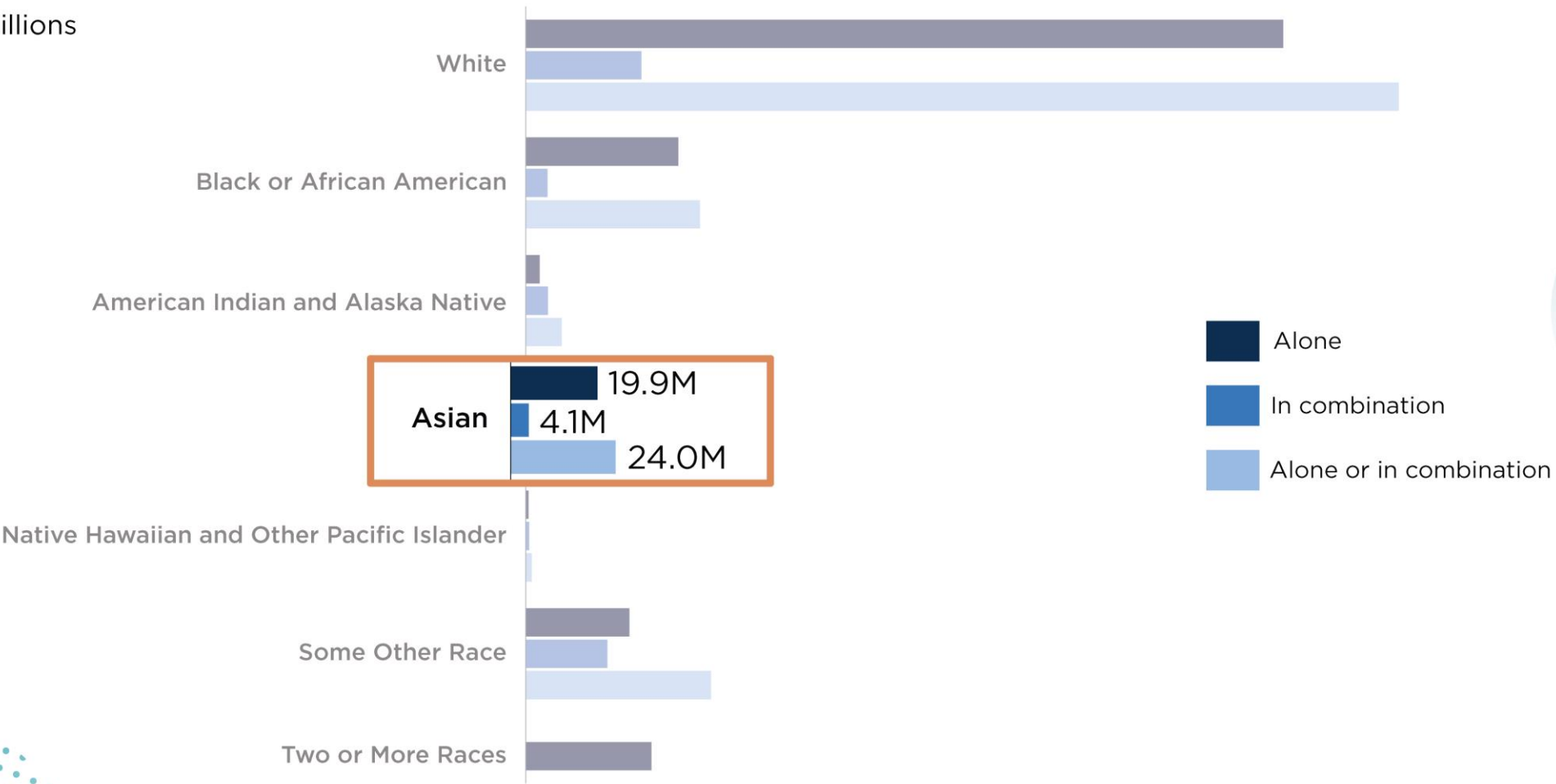
2020 Census Illuminates U.S. Racial and Ethnic Composition

M = Millions



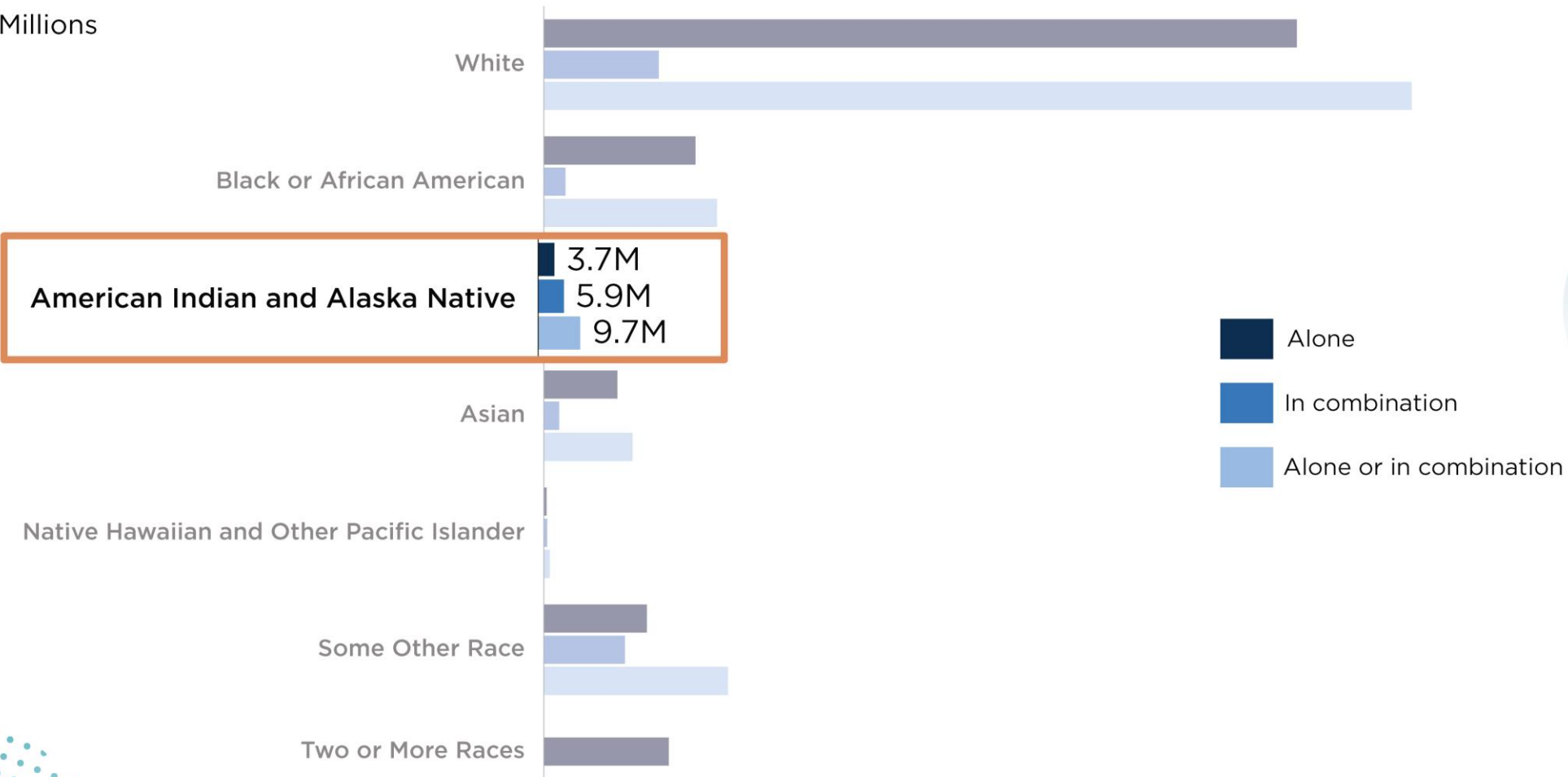
2020 Census Illuminates U.S. Racial and Ethnic Composition

M = Millions



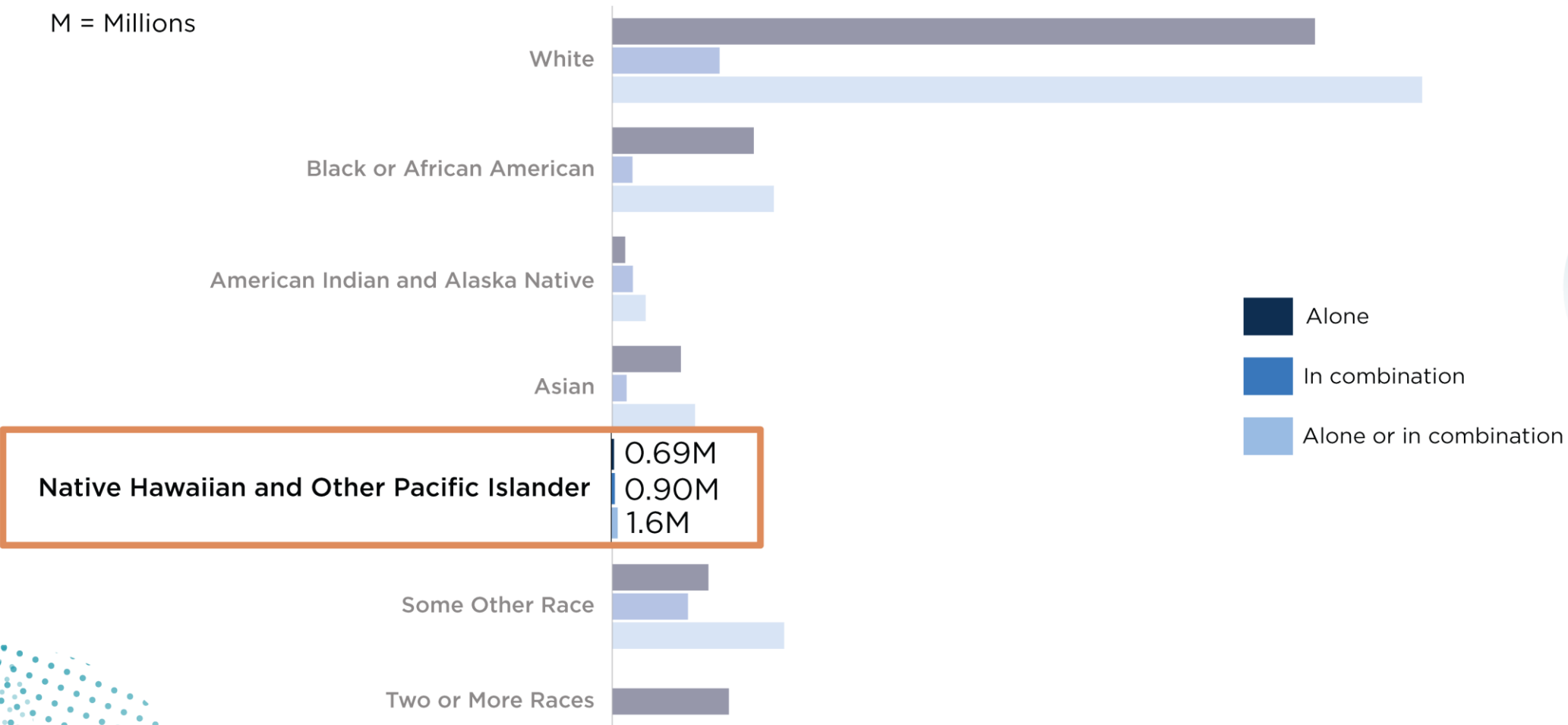
2020 Census Illuminates U.S. Racial and Ethnic Composition

M = Millions

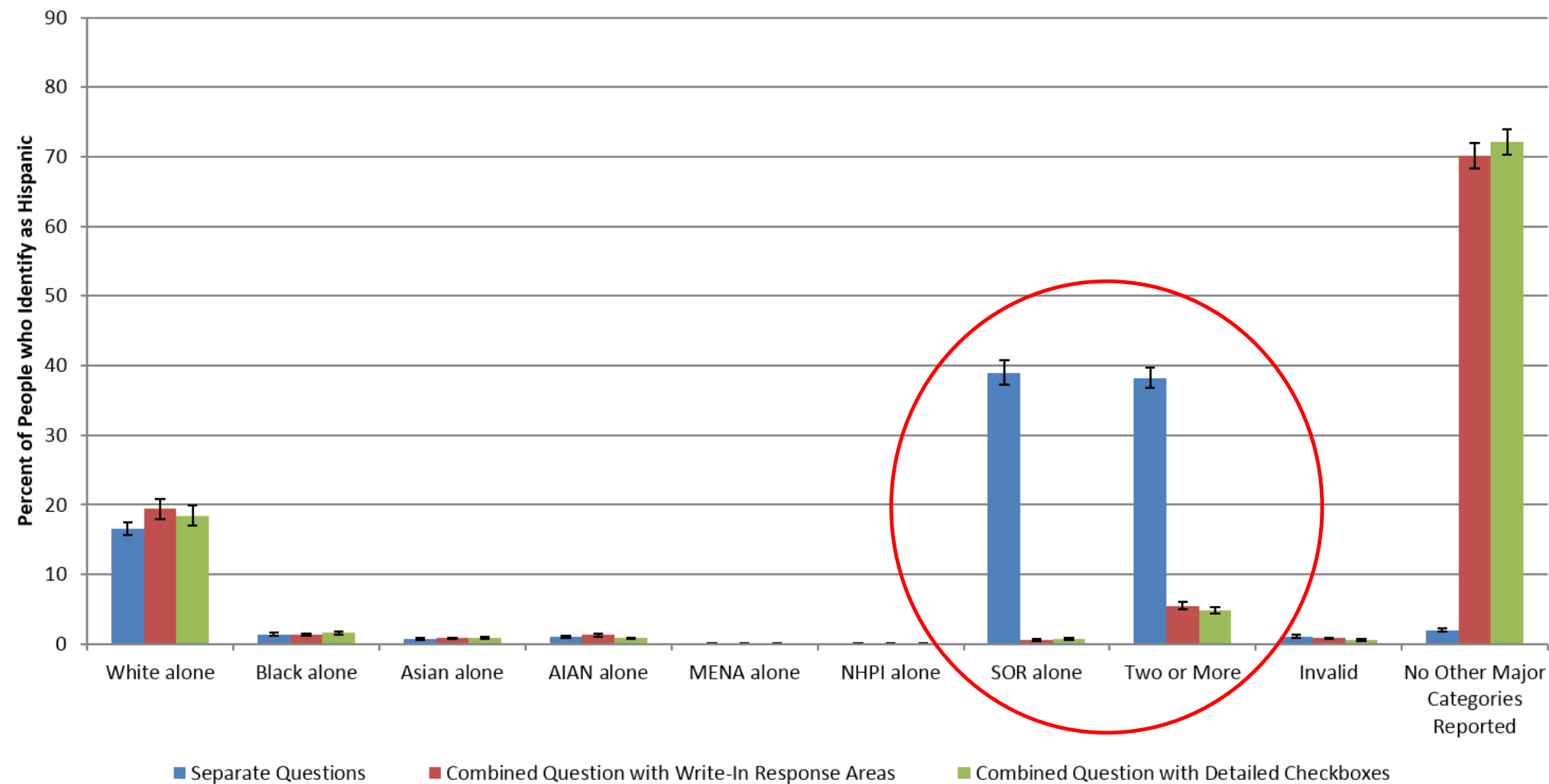


2020 Census Illuminates U.S. Racial and Ethnic Composition

M = Millions



Percent Distribution of Race for Hispanic Origin Respondents by Question Format in the 2015 National Content Test



Source: 2015 National Content Test data.
Note: 95% confidence intervals are depicted by the vertical black lines at the top of each bar.

2020 Census Shows Different But Reasonable and Expected Distributions

- Improving the 2020 Census questions on Hispanic origin and race, along with our coding procedures, enable us to have a more complete picture of the detailed identities reported by the U.S. population in 2020.
- The 2020 Census and 2010 Census race data comparisons should be made with caution, taking into account improvements
- Results are not surprising and align with expert research and findings
- These improvements more accurately illustrate the richness and complexity of how people identify their race and ethnicity in the 21st century within the OMB Standards

Next Steps

- Continue to investigate 2020 Census race and ethnicity data at lower levels of geography to understand changes in the data and demographic patterns
- Develop and solicit stakeholder feedback on detailed race, ethnicity, and tribal data products
- Finalize mid-decade research plans based on the results of the 2020 Census and in consultation with outside experts, stakeholders, and federal partners

Questions?