



Complexity of classifying and presenting race and ethnicity data

Reynaldo Quiroz & Dan Hubbard

University of North Texas

February 2023

About UNT

- ▶ Located in the DFW Metroplex
- ▶ 44,405 students (F22)
- ▶ 237 academic programs (89 in Nation's Top 100)
- ▶ 36% students are Pell eligible
- ▶ 37% students utilize federal unsubsidized loans
- ▶ Carnegie R1 Hispanic Serving Research Institution
- ▶ 54.3% non-white
- ▶ 6.6% International
- ▶ Awarded 10,500+ degrees last year
- ▶ Record growth in enrollment
- ▶ Peoplesoft & Salesforce campus

About | DATA, ANALYTICS, & INSTITUTIONAL RESEARCH

- ▶ 12 amazing professionals across IR, Data Modeling, & Data Management
- ▶ Launched Insights program in 2017, Insights 2.0 launched in 2020
- ▶ 2019 CIO 100 Awardee
- ▶ 1200 users, 1100 terms, and 22 analytic products
- ▶ Leverages SAS Viya, BDN, DIS, EG

About Dan

- Responsible for leadership and implementation of data governance, data management, analytic development, and training involving internal and external data resources at UNT.
- Helps manage Insights, UNT's comprehensive data warehousing, analytics, and governance Program.
- Holds a Bachelor's degree in Psychology, Master's Degree in Cognition and Neuroscience, and a Ph.D. in Cognition and Neuroscience from UT Dallas.



About Reynaldo

- Responsible for providing data analysis of faculty information, preparing reports for the state and federal education agencies, and developing analytics for Insights and UNT leadership.
- Holds an MBA from UNT, a Master's degree Finance, and a Bachelor's degree in Commercial Engineering from EMI.
- Currently working on his dissertation to complete his Ph.D. in Computer Science.



Overview



1

Collecting and classifying ethnicity/race data – more challenging than aggregating individual observations.

2

Complexity of dealing with ethnicity and race data reporting.

3

Zero-Sum game theory concept and Race reclassification.

4

Implications of a Zero-Sum and Non-Zero-Sum game in ethnicity/race data.

Importance of race/ethnicity data to us



- Race/ethnicity data affects funding programs that provide services for specific groups.
- At the same time, it provides the foundations for more equitable education, opportunities, and services for our students, faculty, and staff.



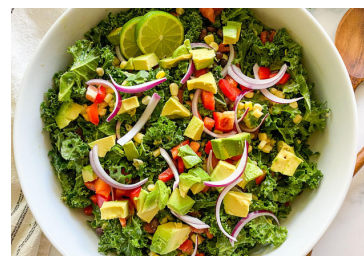
Ethnicity/race concepts



Race and ethnicity are frequently used interchangeably, although there are differences in their meanings.



- Race is more related to physical or biological traits such as skin color, facial features, hair texture, and so on. Although, it could also be used to refer to cultural patterns.
- Ethnicity, on the other hand, represents the shared cultural traditions, beliefs, history, celebrations, and language.



We can consider ourselves a “melting pot” or a “salad bowl.”

The U.S. Office of Management and Budget (OMB) defined the minimum standards for maintaining, collecting, and presenting data on race and ethnicity



Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity

- **American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.
- **Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
- **Black or African American.** A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American."
- **Hispanic or Latino.** A person of Cuban, Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race. The term, "Spanish origin," can be used in addition to "Hispanic or Latino."
- **Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
- **White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

As result: two question format is required



Minimum standards:

- **Ethnicity:**
 - Hispanic or Latino
 - Not Hispanic or Latino
- **Race:**
 - American Indian or Alaska Native
 - Asian
 - Black or African American
 - Native Hawaiian or Other Pacific Islander
 - White

Sample of the ethnicity/race questionnaire

Ethnic Categories*	Select One
Hispanic or Latino	
Not-Hispanic or Latino	
Racial Categories*	Select All that Apply
American Indian or Alaska Native	
Asian	
Black or African American	
Native Hawaiian or Other Pacific Islander	
White	
Other	

The ethnicity/race data grouping paradox



"Birds of a feather flock together"



"My relationship to my Asian identity, it's always been good and healthy. And I love it."

(Keanu Reeves)

However, he has mixed feelings about being referred to as a person of color.

Source: David Lilly, The Canadian Nature Photographer.

An example of the race/ethnicity classification which originates data bias



Are you Hispanic or Latino?

☐ Yes

☒ No

Regardless of your answer to the prior question, please indicate how you identify yourself. (Select one or more)

☒ American Indian or Alaska Native

☐ Asian

☒ Black or African American

☐ Native Hawaiian or Other Pacific Islander

☒ White



IPEDS to the rescue



IPEDS provides guidelines to make this work easier for higher education entities.

If the individual self identifies as...	Report to IPEDS as...	Are you Hispanic or Latino?
Hispanic only, or Hispanic and any race category	Hispanic	<input type="radio"/> Yes <input checked="" type="radio"/> No
Not Hispanic; American Indian or Alaska Native only	American Indian or Alaska Native	Regardless of your answer to the prior question, please indicate how you identify yourself. (Select one or more) <input checked="" type="checkbox"/> American Indian or Alaska Native <input type="checkbox"/> Asian <input checked="" type="checkbox"/> Black or African American <input type="checkbox"/> Native Hawaiian or Other Pacific Islander <input checked="" type="checkbox"/> White
Not Hispanic; Asian only	Asian	
Not Hispanic; Black or African American only	Black or African American	
Not Hispanic; Native Hawaiian or Other Pacific Islander only	Native Hawaiian or Other Pacific Islander	
Not Hispanic; White only	White	
Not Hispanic; more than one race category	Two or more races	

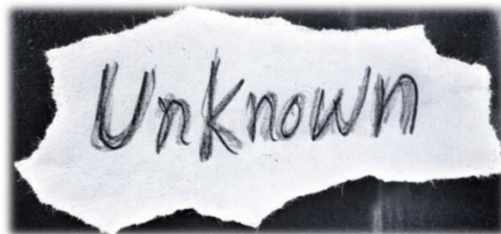
Source: National Center for Education Statistics (NCES)

... challenges with ethnicity/race data classification may persist



If the individual...	Report to IPEDS as...
Refuses to respond to both questions	Unknown race and ethnicity
Responds No to the Hispanic question, but does not respond to the race question	Unknown race and ethnicity
Responds to the race question, but does not respond to the Hispanic question	Report race as outlined above, as if individual self-identified as non-Hispanic
Is a nonresident alien according to the visa and citizenship information on record at the institution	Nonresident alien

Source: National Center for Education Statistics (NCES)



These assumptions are not free of data bias.

THECB ethnicity/race data requirements



Race and ethnicity of regularly enrolled students (CBM002)

White
Black
Hispanic
Asian
American Indian/Alaskan Native
Native Hawaiian/Other Pacific Islander
Multi-Racial
International
Ethnic Origin/Race Unknown

Univ. faculty race and ethnicity report (CBM008)

White
Black
Hispanic
Asian
American Indian/Alaskan Native
Native Hawaiian/Other Pacific Islander
Multi-Racial
International
Ethnic Origin/Race Unknown

Source: THECB, Reporting and Procedures Manual Texas for Public Universities, Current Version, Summer 2022.

Zero-sum game, a win-lose relationship



- There is one winner and one loser – as in most games such as tennis, chess, arm-wrestling, or dominoes.
- Reclassifying some race categories within the whole data set, could drive us to a zero-net gain, one's race category gain could be another's race loss or vice versa.
- It implies that there is an equilibrium.

Prisoner's dilemma



Assumptions



From Home Alone 1 and 2

- The Wet Bandits were caught.
- They were separated in different cells.
- There is no honor among thieves.
- If both cooperate and remain silent, they will go to jail for 1 year.
- If Harry blames Marv and Marv confess, Marv will get 20 years in jail, and Harry will go free.
- If Marv blames Harry and Harry confess, Harry will get 20 years in jail, and Marv will go free.
- If both confess, they will get 5 years in jail each.

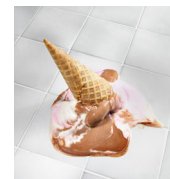
		Harry	
		Confess	No confess
Marv	Confess	5 , 5	0 , 20
	No confess	20 , 0	1 , 1

Game theory advocates that both confess.

Non-Zero-Sum game



- In a non-zero-sum game, the fortune of both or more race categories can rise or fall together.
- In a win-win situation, two or more race categories can benefit from the exchange, net positive.
- In a lose-lose situation, two or more race categories can lose, net loss (**we have to be careful!**).
- For our case, the adoption of a particular reclassification method can impact on the minorities.



Our methodology to reclassify race data



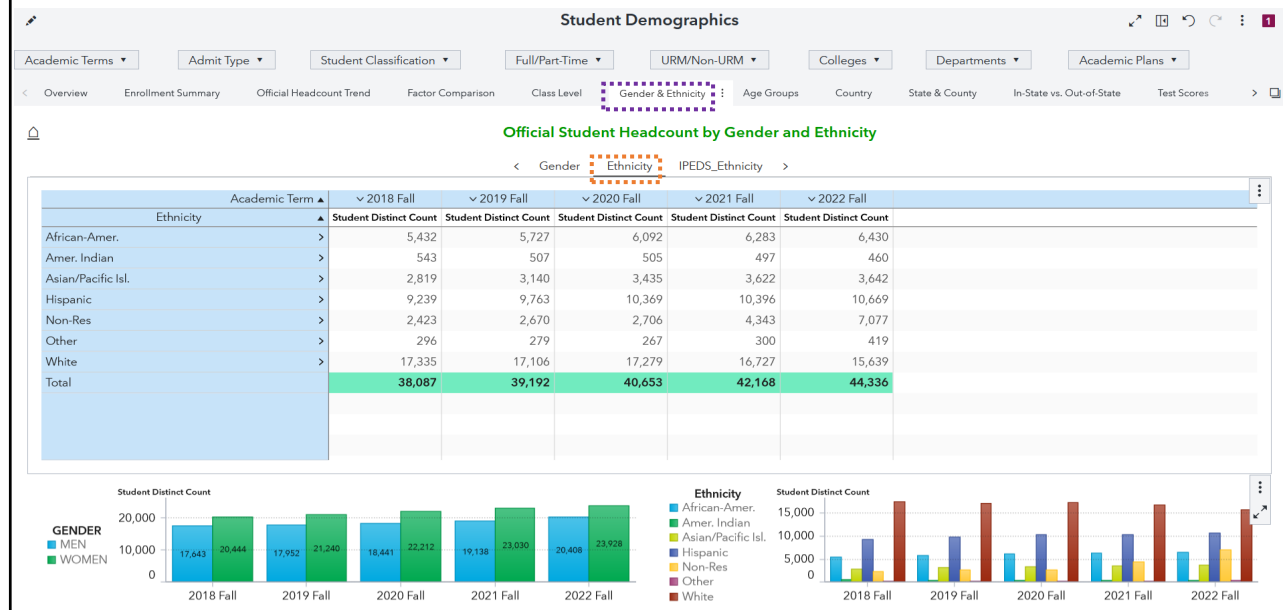
- Revisit our original data.
- Check for the IPEDS and THECB guidelines.
- Check the current weight of races under “Unknown”, “Two or more races”, and “Other.”
- Create a priority rule.
- Reprocess and check the changes.

“Priority rule” when more than one race is marked

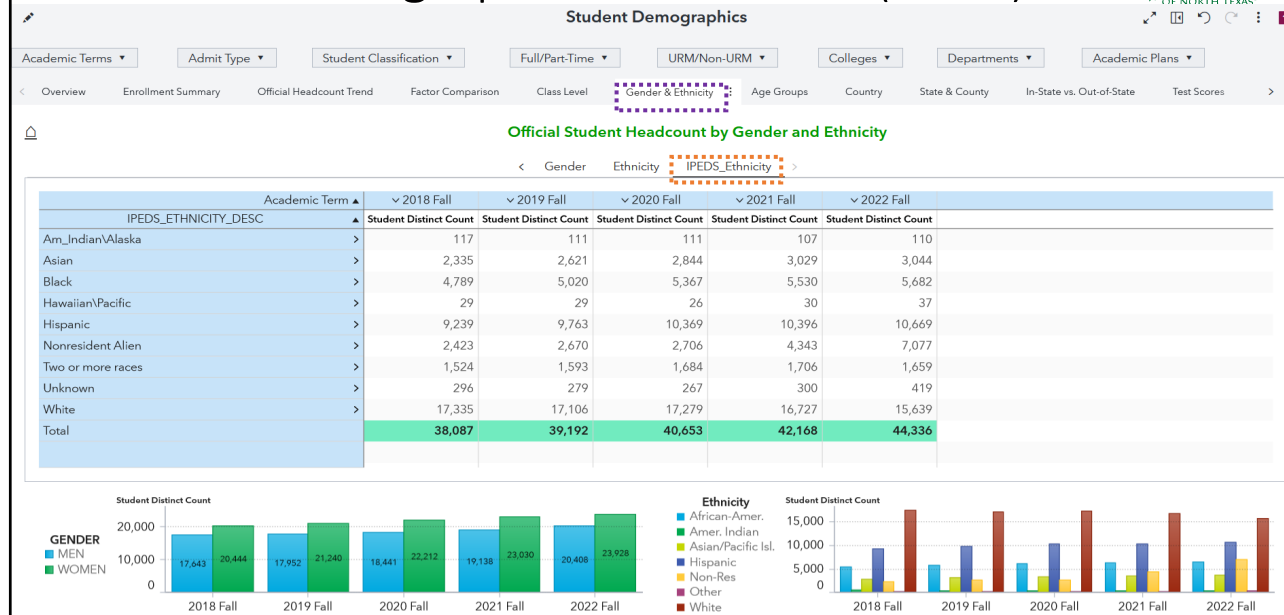


- If ethnicity is Hispanic then race = Hispanic
- If ethnicity is not Hispanic then race = which go first?
 - Black
 - American Indian
 - Hawaiian
 - Asian
 - White
 - Unknown

Student Demographics dashboard



Student Demographics dashboard (cont.)



UNT student ethnicity/race enrollment data



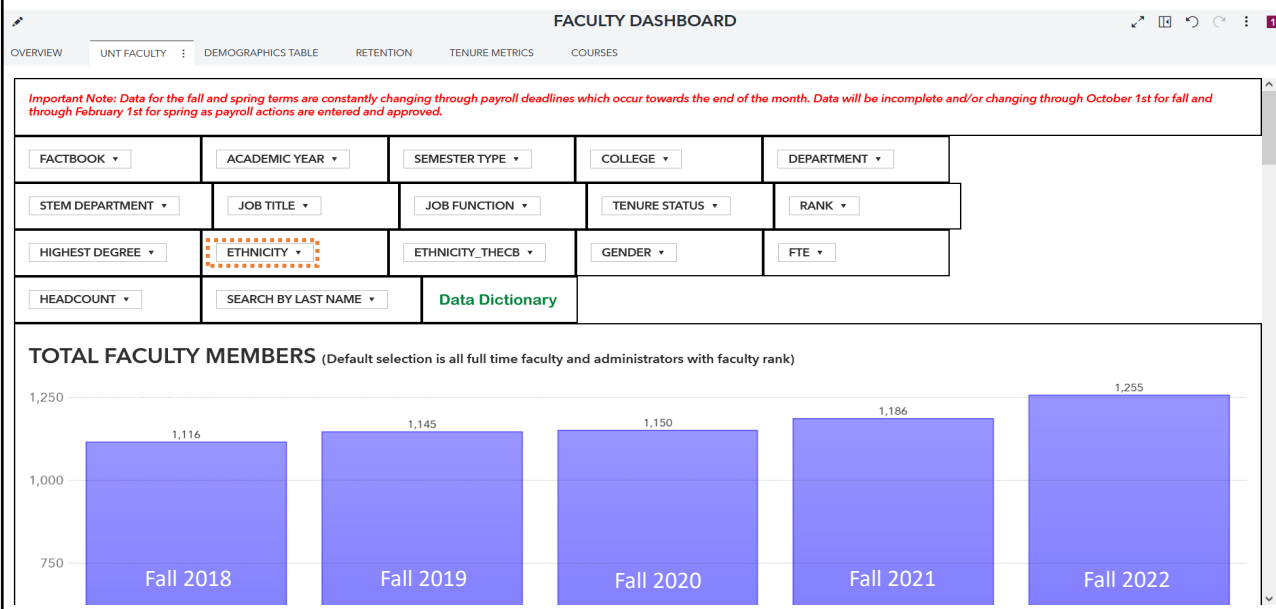
Ethnicity	2018 Fall	2019 Fall	2020 Fall	2021 Fall	2022 Fall
African American	5,432	5,727	6,092	6,283	6,430
American Indian	543	507	505	497	460
Asian/Pacific Islander	2,819	3,140	3,435	3,622	3,642
Hispanic	9,239	9,763	10,369	10,396	10,669
Nonresident Alien	2,423	2,670	2,706	4,343	7,077
Other	296	279	267	300	419
White	17,335	17,106	17,279	16,727	15,639
Total	38,087	39,192	40,653	42,168	44,336

IPEDS Ethnicity description	2018 Fall	2019 Fall	2020 Fall	2021 Fall	2022 Fall
Black	4,789	5,020	5,367	5,530	5,682
American Indian/Alaska	117	111	111	107	110
Asian	2,335	2,621	2,844	3,029	3,044
Hawaiian/Pacific	29	29	26	30	37
Hispanic	9,239	9,763	10,369	10,396	10,669
Nonresident Alien	2,423	2,670	2,706	4,343	7,077
Two or more races	1,524	1,593	1,684	1,706	1,659
Unknown	296	279	267	300	419
White	17,335	17,106	17,279	16,727	15,639
Total	38,087	39,192	40,653	42,168	44,336

Source: UNT Insights – Student Demographics Dashboard

Is there a Zero-Sum Game?

Faculty Dashboard (Ethnicity)

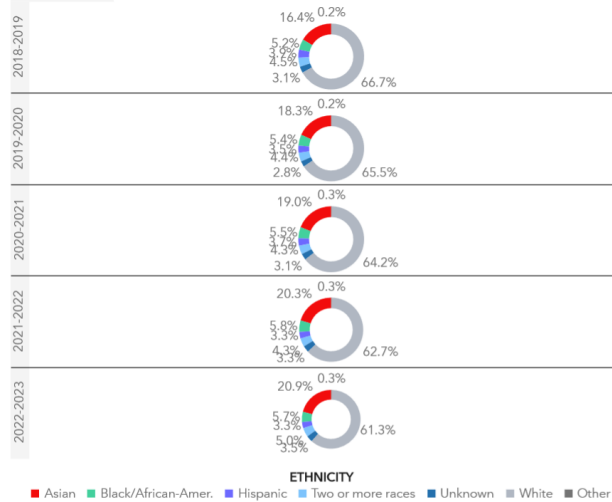


Comparing both criteria of faculty ethnicity



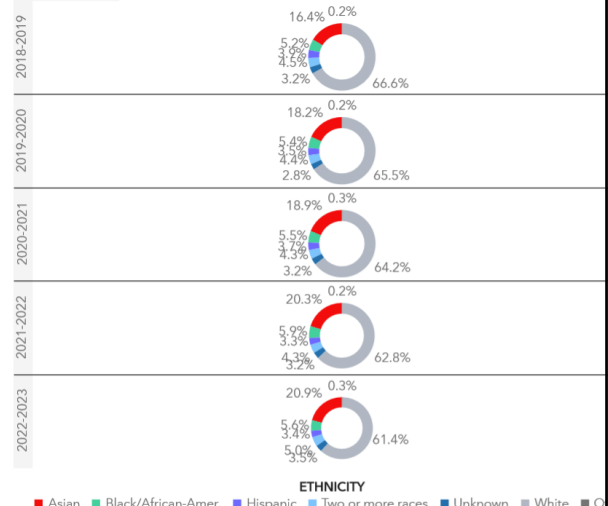
FACULTY MEMBERS BY ETHNICITY

ACADEMIC YEAR



FACULTY MEMBERS BY ETHNICITY

ACADEMIC YEAR



The UNT faculty ethnicity/race data adjusting the THECB criteria



Faculty ethnicity	2018 Fall	2019 Fall	2020 Fall	2021 Fall	2022 Fall
American Indian/Alaskan	1	1	2	1	2
Asian	184	209	217	239	259
Black/African-American	58	62	63	69	69
Hawaiian/Pacific Islander	1	1	1	1	2
Hispanic	43	40	42	39	42
Two or more races	51	51	50	50	62
Unknown	35	32	36	38	43
White	739	745	733	736	763
Total	1,112	1,141	1,144	1,173	1,242

Source: UNT Insights – Faculty Dashboard

Our methodology	2018 Fall	2019 Fall	2020 Fall	2021 Fall	2022 Fall
American Indian	8	6	6	5	6
Asian	140	165	170	194	217
Black	59	64	65	69	70
Native Hawaiian	2	2	2	2	4
Hispanic	72	69	71	70	83
International/Nonresident Alien	85	87	91	86	85
Unknown	32	29	33	35	40
White	714	719	706	712	737
Total	1,112	1,141	1,144	1,173	1,242

Is there a Non Zero-Sum Game among minorities?

Ethnicity criteria adopted by other universities

IPED Ethnicity/Race Categories	MIT	Stanford	Princeton
—American Indian or Alaska Native	American Indian or Alaskan Native	American Indian or Alaskan Native	
—Asian	Asian	Asian	Asian
—Black or African American	Black or African American	Black or African American	Black
—Hispanic or Latino	Hispanic or Latino	Hispanic or Latino	Hispanic
—Native Hawaiian or Other Pacific Islander	Native Hawaiian or Other Pi	Native Hawaiian or Other Pacific Islander	
—White	White	White	White
	International	International	
			Multiracial
—Two or more Races	Two or more Races	Two or more Races	
—Unknown		Unknown	Unknown

Source: Universities' websites where they published ethnicity and race data.

Changes to the 2022-23 IPEDS Data Collection

- Hispanic or Latino, regardless of race

For Non-Hispanic/Latino individuals:

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

In addition, the following categories may be used:

- U.S. Nonresident **alien**
- Race and ethnicity unknown

Racial/ethnic descriptions - Racial/ethnic designations as used in this survey do not denote scientific definitions of anthropological origins. The categories are:

- Hispanic or Latino- A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.
- American Indian or Alaska Native- A person having origins in any of the original peoples of North and South America (including Central America) who maintains cultural identification through tribal affiliation or community attachment.
- Asian- A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian Subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
- Black or African American- A person having origins in any of the black racial groups of Africa.
- Native Hawaiian or Other Pacific Islander- A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
- White - A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

Other descriptive categories

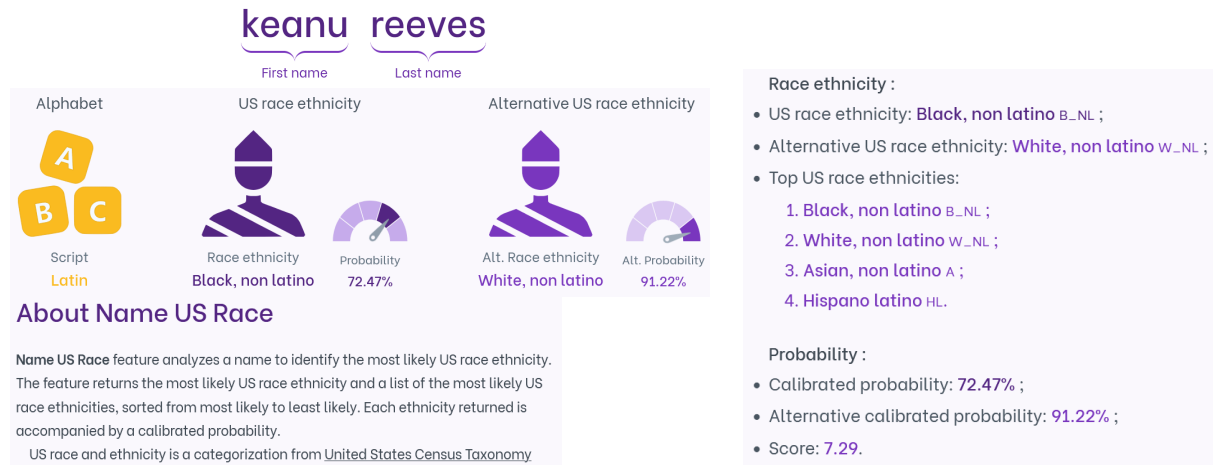
- U.S. Nonresident **alien** - A person who is not a citizen or national of the United States and who is in this country on a visa or temporary basis and does not have the right to remain indefinitely. **Do not include DACA, undocumented, or other eligible noncitizens in this category.** NOTE - U.S. Nonresidents **aliens** are to be reported separately, in the boxes provided, rather than included in any of the seven racial/ethnic categories. **Other eligible (for financial aid purposes) non-citizens who are not citizens or nationals of the United States and who have been admitted as legal immigrants for the purpose of obtaining permanent resident status (and who hold either an alien registration card (Form I-551 or I-151), a Temporary Resident Card (Form I-688), or an Arrival-Departure Record (Form I-94) with a notation that conveys legal immigrant status such as Section 207 Refugee, Section 208 Asylee, Conditional Entrant, Parolee or Cuban-Haitian) are to be reported in the appropriate racial/ethnic categories along with United States citizens.**

Source: IPEDS Data Collection System. Changes to the 2022-23 IPEDS Data Collection.

No fundamental changes have been made to help improve our current situation.



Can AI help identify individual's ethnicity?



Source: NAMSOR, name checker for gender, origin and ethnicity determination.
<https://namsor.app/>.



Conclusions

- Challenges collecting, classifying, and presenting ethnicity/race data.
- Federal and state education agencies' ethnicity and race data presentation standards help organize and tabulate the data, but not considering reviewing it could be costly for some races and ethnicities.
- Game theory concepts can be applied to formulate scenarios where races can be considered agents, and their strategies can be reclassified under other race categories looking for optimal solutions.
- There are ways to improve our ethnicity/race data presentation by addressing the education agencies' standards and increasing the presence of minority groups.

