

As Arnold Ventures' [Advancing Pretrial Policy and Research](#) (APPR) [National Research Partner](#), RTI International independently studies six [Research-Action Sites](#) that receive intensive technical assistance through APPR. In keeping with APPR's principles of pretrial justice, our research aims to understand how changes are made to pretrial systems; to understand the impacts of those changes; and to identify opportunities for more equitable, fair, and transparent practices for people involved with the pretrial system. To support our commitment to transparency, equity, and accountability, we have developed this series of APPR Research Briefs to share information about our research processes and findings.

Research Insights: Using Data to Explore Racial and Ethnic Disparities in the Criminal Legal System

APPR Research Brief, August 2022

by Elizabeth Tibaduiza, MA, Megan Comfort, PhD, and Emily Burtch, MS

Jurisdictions across the country have committed to decreasing racial and ethnic disparities in their criminal legal systems. A key step to achieving this goal is to identify where disparities exist and then assess what the potential sources of those disparities are, what reforms might mitigate the disparities, and whether implemented reforms have the intended effects. Having accurate and precise data is necessary to this process. However, the collection and interpretation of race and ethnicity data are not as straightforward as they may seem. This Research Brief provides context to our understanding of the complexities related to data collection on race and ethnicity. We then outline high-level considerations to help jurisdictions think through their current data collection processes and how those processes might be improved. Finally, we present an example of how RTI is using data to help the Research-Action Sites assess disproportionality in their jail bookings.

Research-Action Sites (Cohort 1):

- Fulton County, GA
- Montgomery County, AL
- Pierce County, WA
- Thurston County, WA

Research-Action Sites (Cohort 2):

- Pulaski County, AR
- Ramsey County, MN

The Complexities of Race and Ethnicity Data: Key Issues to Consider

Race and ethnicity data can help facilitate an understanding of disparities and the identification of inequity in the criminal legal system. However, it is important when interpreting these data and drawing conclusions to consider the challenges and complexities surrounding their meaning and collection.

What Are Race and Ethnicity? It's Complicated.

Although race and ethnicity have become heavily associated with physical appearance and biological traits, people invented the idea of dividing humans into different groups. This is what is meant by the statement that "race is a social construct."¹ The definitions of racial and ethnic categories have changed throughout history and continue to be regularly examined, challenged, and modified.^{2,3} The U.S. Census Bureau race and ethnicity categories—currently guided by the standards set by the U.S. Office of Management and Budget in 1997⁴—have been revised numerous times to reflect "changes in society."^{5,6} For example, Census Bureau researchers recommended adding a race category of "Middle Eastern or North African" and combining "Hispanic, Latino, or Spanish origin," currently classified as an ethnicity, into the race category.^{5,7} While these recommendations ultimately were not incorporated into the 2020 Census, they reflect ongoing conversations about how race and ethnicity are defined.

U.S. Census Race Categories

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

U.S. Census Ethnicity Category

- Hispanic, Latino, or Spanish origin

Continually revising definitions and expanding understandings of our racial and ethnic identities can affect the reliability and validity of race and ethnicity data.⁸⁻¹¹ How people identify does not necessarily align with classification systems used by federal, state, and local government agencies and can lead to the selection of "other" categories^{12,13} or of multiple categories, which adds complexity to analysis and interpretation. For example, in the 2010 Census, the third most selected race category was "Some Other Race," which was largely driven by people who identified as "Hispanic, Latino, or Spanish."¹⁴ In the 2020 Census, selection of "Some Other Race" (alone or in combination with another group) increased 129%, becoming the second most selected category. The "Two or More Races" (or multiracial) population also increased by 276% from 2010.¹⁵

Variations in Data Collection Approaches

Taking stock of how data are collected and categorized across the criminal legal system is an important step in better understanding where and how disparities arise and what policy changes hold the most promise for eliminating them. The lack of standardization in how race and ethnicity data are classified and collected across the criminal legal system can complicate their use.¹⁶⁻¹⁸ Using the Research-Action Sites' administrative jail data as an example, each site uses categories that vary from the census categories, such as combining race categories, excluding a race category, or combining race and ethnicity (these variations are further detailed below). Another potential issue is that the categories used by agencies within the same jurisdiction may also vary. These inconsistencies can make it difficult to compare demographics at decision points (e.g., arrest, booking) to local population demographics to examine disparities within a jurisdiction and across jurisdictions. As one Research-Action Site stakeholder observed during an [RTI site visit interview](#), strengthening system actors' understanding of how the data they enter are used could help improve data quality:

Some [race and gender] data is compromised, given how it's inputted. Some of it's reliant upon observations, as opposed to a confirmation from the individual. It comes from a lot of different data points [e.g., law enforcement contact, booking, charging documents], and I think what we're finding is that not all of them are congruent. I also think too that it gets impacted by whether or not the person who's inputting some information believes that it's valuable information. Is someone going to rely on this so that I need to be careful about what it is that I'm writing or am I just filling in a box? I think if we knew the importance of that, if people understood that data point and how it gets used, then they might be more apt to confirm it with the individual.

The National Center for State Courts developed broad [guidelines](#) and considerations to assist courts in collecting race and ethnicity data.¹⁹ These materials support the use of categories that are consistent with the U.S. Census, and they encourage jurisdictions to add more detailed options within an overarching category as needed to reflect the demographic composition of a community (e.g., specifying tribal affiliation for people who identify as American Indian or Alaska Native). When jurisdictions consider which categories to include, the center recommends that they avoid combining groups into an "other" category, a practice that results in the loss of nuanced data and representation of groups and limits the ability for comparison.²⁰⁻²²

Data collection methods also vary, with some agencies relying on self-classification, others relying on another person's observation (i.e., assumption), and some using previously reported data (e.g., from criminal history records). Moreover, different data collection methods may be used in agencies in the same jurisdiction, which can produce inconsistent information. The different ways in which race and ethnicity are measured (e.g., self-classified, observed) represent different dimensions of race and ethnicity.^{23,24} How people classify themselves captures something different from how someone else classifies them. Self-classification is the most accurate objective measure; however, which measure is most appropriate to use may depend on the outcomes of interest. For example, Hanna and colleagues²⁴ suggest self-classification to examine demographic change or observation to examine discrimination. Comparison to local population demographics is often the objective for criminal legal agencies, in which case self-classification is the most helpful method. Some researchers contend that observed race or ethnicity is more likely to influence disparities in systems than self-identified race or ethnicity, and it may be useful to collect both measures.^{2:24} The National Center for State Courts guidelines say that data collectors should record whether race and ethnicity are self-identified or observed.¹⁹

Where to Start? Considerations for Collaborative and Practical Change

Racial and ethnic groups often experience disparate treatment while involved in the criminal legal system, and it is important to assess and monitor these disparities.^{2:25:26} Knowing where disparities exist allows for exploration of why they exist, what consequences they have, and how they can be addressed. Below, we offer suggestions and considerations for jurisdictions that want to improve their collection and quality of race and ethnicity data. It is important to note that there is not a one-size-fits-all approach, and jurisdictions are encouraged to design their approaches to meet their local needs (e.g., assessing racial disparity at specific decision points such as arrest, booking, charging, pretrial release).

Read the National Center for State Courts Court Statistics Project's brief on [Collecting Race & Ethnicity Data](#) for additional questions and guidance.¹⁹

- Consider carefully who should be engaged in this process. Ideally, a variety of system stakeholders and community members, including people who have been directly affected by the criminal legal system, will be given meaningful opportunity to participate and provide input.
 - Engaging representatives from agencies that are or will be responsible for data collection increases the likelihood of identifying and resolving barriers before implementation. It also provides opportunity to consistently apply improvements to measures and practices across local agencies. This consistency is important and should be prioritized.
 - Including a wide variety of stakeholders can help to foster collaboration in your jurisdiction and increase buy-in to any proposed improvements.
- If your jurisdiction is just starting to have conversations about race and ethnicity data collection, these questions and others may be helpful to explore:
 - What are your jurisdiction's data priorities? How do you want to use these data?
 - What are current challenges to collecting the data?

- What needs improvement? What resources are needed?
- What questions and concerns (e.g., time, funding, data security) do people have about collecting these data?
- Catalog the data currently being collected by your local criminal legal system. Know whether race and ethnicity data are collected; by whom, when, and how data are collected; how race and ethnicity data are coded; and how to access data that have been collected.
 - Align your data priorities with the data currently available to assess whether any changes need to be made.
- Develop your race and ethnicity categories in accordance with local priorities, accounting for populations that are important to capture that are not represented in the census categories. If more detailed categories are needed, structure them such that they can be grouped to align with the broad census categories to facilitate comparisons to local demographics. It may be helpful to develop a crosswalk, as described later in this brief.
- Consider which measures are important to capture and can realistically be documented consistently. The [National Open Court Data Standards](#) state that self-identification is preferred.²⁷ Researchers are increasingly advocating for multi-method approaches to measure race and ethnicity. Consider these examples when determining what is best for your jurisdiction:^{24; 28}
 - Self-identified race and ethnicity (how a person classifies their own race/ethnicity)
 - If several categories can be selected, the race or ethnicity the person identifies with the most
 - Perception of socially assigned race or ethnicity (how a person thinks others generally classify their race or ethnicity)²⁹
 - Observed race and ethnicity (the data collector’s perception of race/ethnicity)
- Develop clear data definitions and procedures and train staff who collect and ensure data to ensure standardization and quality.
- Transparency can help to build public trust in systems. Consider regularly disseminating analyses that assess disparities or providing the community with easily accessible, de-identified data.
- Incorporate racial equity awareness into data visualization and presentation by being mindful of the words, structure, images, and colors that are used. This [article](#) from the Urban Institute³⁰ provides guidance and links to other resources to help make inclusive and equitable choices.

Measures for Justice’s [The Power and Problem of Criminal Justice Data](#)³¹ is another helpful resource that connects the need for broad data improvements to concrete examples of what we miss without reliable data and what can be examined when such data are available. In the next section, we provide our own example, describing how RTI is using data from the Research-Action Sites to help sites assess disproportionality in jail bookings.

Data in Action: Examining Disproportionality in Jail Bookings in Research-Action Sites

As part of RTI’s APPR research activities, we developed site-specific dashboards to provide analytical tools for Research-Action Site policy teams, stakeholders, and technical assistance providers. This work led to an exploration of how the sites categorize race and ethnicity data and how their categories compare to census data. We then developed a standardized process to match a site’s race categories to census categories, allowing for comparison.

Research-Action Site Dashboards: Historical Jail Data, 2017–2021

RTI obtained administrative jail data from the first cohort of Research-Action Sites for all individuals booked and admitted into jail from January 1, 2017, through various end dates in 2021, depending on the site.^a These data are used to populate dashboards that allow users to explore booking and charging trends in the local criminal legal system through variables such as race, ethnicity, gender, bond, and type of crime. The dashboard features can also be used to inspect jail trends—the number of bookings and releases, the average length of stay, and the jail population. The Research-Action Sites continue to provide administrative jail data to RTI to keep the dashboards current. As they become available, court and criminal history data are incorporated as well.

Matching Research-Action Site and Census Bureau Race and Ethnicity Categories

To facilitate comparison between Research-Action Site data and local demographic data, we have incorporated data from the U.S. Census Bureau’s 2019 American Community Survey (ACS)³² into the dashboards. This feature allows the sites to explore whether there is racial or ethnic disproportionality. As a first step in this process, it was necessary to examine the racial and ethnic category definitions used by local jails and align them with the racial and ethnic category definitions used in the 2019 ACS.

For full definitions of the 2019 ACS race and ethnicity questions, including how nationalities and ethnic groups are grouped into each category, see the American Community Survey and Puerto Rico Community Survey: 2019 Subject Definitions.³¹

^aFor the first cohort of Research-Action Sites, RTI has obtained administrative jail data for all individuals booked and admitted into jail within the following date ranges: Fulton County, GA, 1/1/2017–7/31/2021; Montgomery County, AL, 1/1/2017–6/30/2021; Pierce County, WA, 1/1/2017–9/30/2021; and Thurston County, WA, 1/1/2017–9/30/2021.

RTI’s [APPR Racial and Community Justice Committee](#) and the dashboard data analysts developed a crosswalk of the ACS and Research-Action Site categories. The ACS uses the following race categories: White; Black or African American; American Indian or Alaska Native; Asian; Native Hawaiian or Other Pacific Islander; and Some Other Race.³³ ACS respondents are also asked separately to indicate if they identify as Hispanic, Latino, or Spanish.³⁴ **Exhibit 1** shows how the Cohort 1 Research-Action Sites’ definitions align. Although there is overlap, there are also notable differences. “American Indian or Alaska Native” is not included in one site’s data; three sites use the combined category of “Asian/Pacific Islander;” and most of the sites combine race and ethnicity (i.e., Hispanic/Latinx) into one variable. All the sites also include an “unknown” or “other” option (or both).

Exhibit 1. Comparing Census Bureau and Cohort 1 Research-Action Site Race and Ethnicity Categories

| Category | Census | Site 1 | Site 2 | Site 3 | Site 4 |
|-----------|---|------------------------------------|--|------------------------------------|------------------------------------|
| Race | Black or African American | Black | Black | Black | Black |
| | American Indian or Alaska Native | American Indian/ Native Alaskan | No match to census | American Indian/ Native Alaskan | American Indian/ Native Alaskan |
| | Asian | Asian/Pacific Islander | Chinese, Filipino, Japanese, Korean, Laotian, Vietnamese, Other Asian | Asian/Pacific Islander | Asian/Pacific Islander |
| | Native Hawaiian or Other Pacific Islander | | Guamanian, Hawaiian, Pacific Islander, Samoan | | |
| | White | White | White | White | White |
| | Some other race, Two or more races | Other, Multiple races | Other | Other | Other |
| | Not an option | Unknown | Unknown | Unknown | Not an option |
| Ethnicity | Hispanic, Latino, or Spanish origin [S] | Hispanic [C] | Hispanic, Latin, Mexican [C] | Hispanic [S & C] | Hispanic [S] |

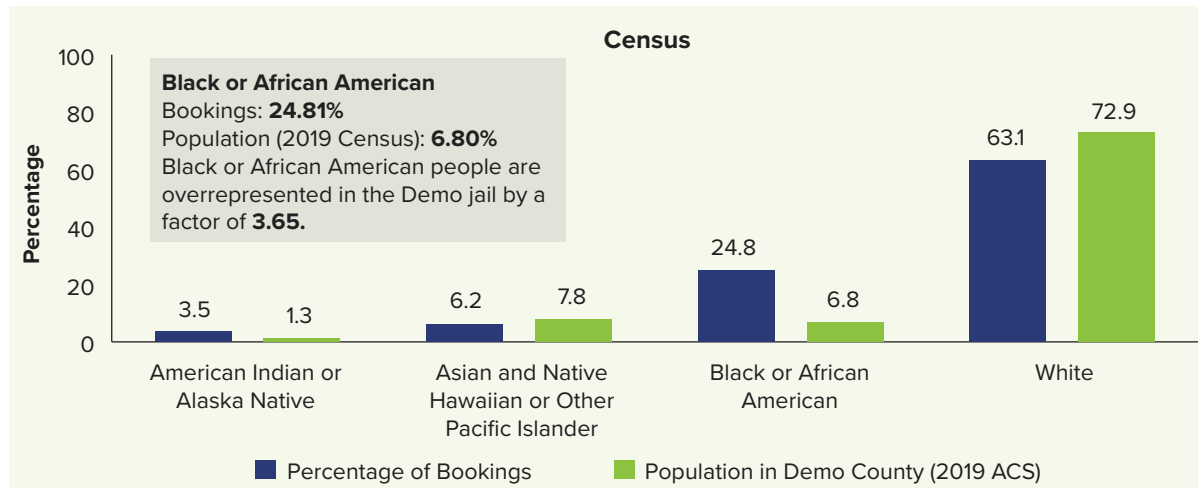
Note: [S] indicates Hispanic/Latinx is collected separately from race (i.e., as its own variable). [C] indicates it is combined with the race categories. One site collects this information both ways (separately and combined).

Assessing and Visualizing Disproportionality

Once the categories were matched, the data allowed for examination of potential racial disproportionalities. We did this by creating a grouped bar chart for each Research-Action Site dashboard to compare the percentage of each race represented in the jail booking data to the percentage represented in the local population (i.e., the ACS data). An example of this type of chart is shown in **Exhibit 2**. Similar comparisons for Hispanic/Latinx ethnicity cannot be made for most of the sites, as the ACS categories and jail booking data collection procedures do not align.

We created a ratio to describe whether each racial population was over- or underrepresented in the Research-Action Site jail bookings by dividing each jail booking race category percentage by its corresponding ACS percentage. **Exhibit 2** shows that Black or African American individuals are strongly overrepresented in these example data (24.81% of bookings, and 6.8% of the population). This statistic helps contextualize the jail population and prompts viewers to consider whether over-policing or discrimination occurs in their local criminal legal system. Furthermore, jurisdictions with a focus on remediating racial disparities can use this dashboard as a guidepost to situate where their bookings lie in proportion to the general population—and how this might change over time.

Exhibit 2. Example Graph Comparing Jail Bookings to Local Population by Race



You Can't Change What You Can't See

Race and ethnicity and their measurement are broad and complicated topics that this brief cannot fully delve into; we have included many references and encourage reading them to further explore this topic! We want to amplify that collecting data on race and ethnicity is essential to rooting out disparities in the criminal legal system. Without being able to identify where there are disparities, jurisdictions cannot efficiently and effectively advance policy goals related to racial equity. We also acknowledge that this is just one step in the process to achieve those goals. As RTI continues its APPR research with the Research-Action Sites, we will also continue to share what we are learning to spark thoughtful discussion and help build knowledge in jurisdictions that are seeking to advance justice in their systems.

The [Racial and Community Justice Committee \(RCJC\)](#) is an internal committee of RTI researchers. The RCJC was established to ensure that RTI maintains a focus on racial equity and community engagement across all of our work on APPR and to help ensure that APPR's study of pretrial change is grounded in a commitment to equity and informed by people directly affected by pretrial systems. For questions about this brief, please contact Megan Comfort at mcomfort@rti.org or (510) 665-8275.

References

- Milner IV, H. R. (2015). Getting real about race. (Excerpted from *Rac(e)ing to Class: Confronting Poverty and Race in Schools and Classrooms*, by H. R. Milner IV, 2015, Harvard Education Press). *Teaching Tolerance Magazine*, 51, 29–31.
- Strmic-Pawl, H. V., Jackson, B. A., & Garner, S. (2018). Race counts: Racial and ethnic data on the U.S. census and the implications for tracking inequality. *Sociology of Race and Ethnicity*, 4(1), 1–13. <https://doi.org/10.1177/2332649217742869>
- Hans, V. P., & Martinez, R. (1994). Intersections of race, ethnicity, and the law. *Law and Human Behavior*, 18(3), 211–221.
- U.S. Office of Management and Budget. (1997, October). *Revisions to the standards for the classification of federal data on race and ethnicity*. https://obamawhitehouse.archives.gov/omb/fedreg_1997standards
- U.S. Census Bureau. (2021, April). *Research to improve data on race and ethnicity*. <https://www.census.gov/about/our-research/race-ethnicity.html>
- Brown, A. (2020, February 25). *The changing categories the U.S. census has used to measure race*. Pew Research Center. <https://www.pewresearch.org/fact-tank/2020/02/25/the-changing-categories-the-u-s-has-used-to-measure-race/>
- Cohn, D. (2017, April 20). *Seeking better data on Hispanics, Census Bureau may change how it asks about race*. Pew Research Center. <https://www.pewresearch.org/fact-tank/2017/04/20/seeking-better-data-on-hispanics-census-bureau-may-change-how-it-asks-about-race/>
- Perez, A. D., & Hirschman, C. (2009). The changing racial and ethnic composition of the US population: emerging American identities. *Population and Development Review*, 35(1), 1–51. <https://doi.org/10.1111/j.1728-4457.2009.00260.x>
- Gupta, S. (2020, March). To fight discrimination, the U.S. census needs a different race question. *Science News*, 197(5), 16–21.
- Cohn, D. (2014, May 5). *Millions of Americans changed their racial or ethnic identity from one census to the next*. Pew Research Center. <https://www.pewresearch.org/fact-tank/2014/05/05/millions-of-americans-changed-their-racial-or-ethnic-identity-from-one-census-to-the-next/>
- Deerinwater, J. (2019, December 10). *Paper genocide: The barriers to an accurate count of Native people in 2020*. Rewire News Group. <https://rewirenewsgroup.com/article/2019/12/10/paper-genocide-the-barriers-to-an-accurate-count-of-native-people-in-2020/>

12. Lopez, M., & Krogstad, J. (2014, April 4). 'Mexican,' 'Hispanic,' 'Latin American' top list of race write-ins on the 2010 census. Pew Research Center. <https://www.pewresearch.org/fact-tank/2014/04/04/mexican-hispanic-and-latin-american-top-list-of-race-write-ins-on-the-2010-census/>
13. Wang, H. (2021, September 30). 1 in 7 people are 'some other race' on the U.S. Census. That's a big data problem. NPR. <https://www.npr.org/2021/09/30/1037352177/2020-census-results-by-race-some-other-latino-ethnicity-hispanic>
14. Mathews, K., Phelan, J., Jones, N. A., Konya, S., Marks, R., Pratt, B. M., Coombs, J., & Bentley, M. (2017). 2015 national content test race and ethnicity analysis report. U.S. Census Bureau.
15. U.S. Census Bureau. (2021, August). 2020 Census statistics highlight local population changes and nation's racial and ethnic diversity. Local Population Changes and Nation's Racial and Ethnic Diversity. census.gov
16. Center for Open Data Enterprise. (2021). Open data for racial equity: A briefing paper on policing data. <http://reports.opendataenterprise.org/Policing-Data-Briefing-Paper.pdf>
17. Diaz, S., Bernal, A., Aguilar, J., Morales, M., Applegarth, M., Casas, R., Foxen, P., Rincon-Gallardo, M., & Villarruel, F. (2020). The Latinx data gap in the youth justice system. <https://latino.ucla.edu/wp-content/uploads/2020/08/LPPI-CJ-JJ-Report-copy.pdf>
18. Donnelly, M., & Bergeron, P. (2021, January 24). Judges on race: Lack of data deters criminal justice reform. <https://www.law360.com/articles/1347287/judges-on-race-lack-of-data-deters-criminal-justice-reform>
19. National Center for State Courts. (2020, June). Collecting race & ethnicity data: A Court Statistics Project data governance special topic. https://www.ncsc.org/_data/assets/pdf_file/0019/42256/Race_special_topic_final.pdf
20. Daniel, R. (2020, April 22). Since you asked: What data exists about Native American people in the criminal justice system? Prison Policy Initiative. <https://www.prisonpolicy.org/blog/2020/04/22/native/>
21. Hu, C., & Esthappan, S. (2017, May 23). Asian Americans and Pacific Islanders, a missing minority in criminal justice data. Urban Institute. <https://www.urban.org/urban-wire/asian-americans-and-pacific-islanders-missing-minority-criminal-justice-data>
22. Goodman, P. (2008). "It's just Black, White, or Hispanic": An observational study of racializing moves in California's segregated prison reception centers. *Law & Society Review*, 42(4), 735–770. <https://doi.org/10.1111/j.1540-5893.2008.00357.x>
23. Roth, W. D. (2016). The multiple dimensions of race. *Ethnic and Racial Studies*, 39(8), 1310–1338. <https://doi.org/10.1080/01419870.2016.1140793>
24. Hanna, A., Denton, E., Smart, A., & Smith-Loud, J. (2020, January). Towards a critical race methodology in algorithmic fairness. 2020 Conference on Fairness, Accountability, and Transparency, pp. 501–512.
25. Stewart, E. A., Warren, P. Y., Hughes, C., & Brunson, R. K. (2017). Race, ethnicity, and criminal justice contact: Reflections for future research. *Race and Justice*, 10(2), 119–149. <https://doi.org/10.1177/2153368717738909>
26. Walker, M. L. (2016). Race making in a penal institution. *American Journal of Sociology*, 121(4), 1051–1078. <https://doi.org/10.1086/684033>
27. National Open Court Data Standards. (2020, April). User Guide. https://www.ncsc.org/_data/assets/pdf_file/0014/34025/NODS-User-Guide.pdf
28. Campbell, M. E., Bratter, J. L., & Roth, W. D. (2016). Measuring the diverging components of race: An introduction. *American Behavioral Scientist*, 64(4), 381–389. <https://doi.org/10.1177/0002764215613381>
29. White, K., Lawrence, J. A., Tchangalova, N., Huang, S. J., & Cummings, J. L. (2020). Socially-assigned race and health: A scoping review with global implications for population health equity. *International Journal for Equity in Health*, 19(25). <https://doi.org/10.1186/s12939-020-1137-5>
30. Schwabish, J., & Feng, A. (2020, September 3). Applying racial equity awareness in data visualization. <https://osf.io/x8tbw/>
31. Measures for Justice. (2021, June). The power and problem of criminal justice data: A 20-state review. https://www.measuresforjustice.org/about/docs/The_Power_And_Problem_Of_Criminal_Justice_Data.pdf
32. U.S. Census Bureau. (n.d.). American Community Survey and Puerto Rico Community Survey: 2019 Subject Definitions. U.S. Census Bureau. https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2019_ACSSubjectDefinitions.pdf
33. U.S. Census Bureau. (2020, October). About the topic of race. U.S. Census Bureau. <https://www.census.gov/topics/population/race/about.html>
34. U.S. Census Bureau. (2020, October). About the Hispanic population and its origin. U.S. Census Bureau. <https://www.census.gov/topics/population/hispanic-origin/about.html>

RTI International (www.rti.org) is the independent National Research Partner for APPR.

RTI International is a trade name of Research Triangle Institute. RTI and the RTI logo are U.S. registered trademarks of Research Triangle Institute.

