

IV. THE RACIAL DATA TRACKER EXPERIENCE



In this Section, we describe the experience of the RDT team collecting data in the summer of 2021. We describe the RDT methodology, detail the challenges and data deficiencies the RDT team encountered for each issue area it studied, and summarize the implications for antiracist policymaking.

A. RDT Methodology

In the summer of 2021, the RDT team, which consisted of Center faculty, staff, postdoctoral fellows, and student interns, examined racial and ethnic data in the areas of houselessness, criminal arrests, and police violence, and identified many deficiencies in the available information. The team began by collecting data at the national level to see what information was available and disaggregated for each issue area. It then tried to fill data gaps by obtaining information directly from states and the fifty largest cities, but often could not do so because data were unavailable.⁹² The RDT data collection efforts are ongoing, but for purposes of this report, the data collection period began in May 2021 and ended in August 2021.

The RDT team gathered data from federal, state, and city websites and data dashboards, as well as reports and data published by research institutions. In some instances, the team informally contacted cities' statistics departments, health departments, and social services departments to supplement missing data. Data ranged widely in its accessibility, with some data easily downloadable from publicly available websites and others only obtainable upon request to public officials. The team focused its efforts on collecting data in the form of cumulative counts.

B. The Challenges of Obtaining Houselessness Data by Race and Ethnicity

The RDT team encountered several deficiencies in houselessness data concerning race and ethnicity. First, the data sources employ different methodologies. The datasets vary in terms of whether they reflect houselessness data collected on a given night, during a three-day period, or during a one-year period. Second, houselessness data are aggregated in a manner that does not aid in the creation of evidence-based policies that address particular geographic or jurisdictional needs: generally such data are publicly reported as aggregated national statistics. Additionally, when the data are disaggregated beyond the national level, they are often disaggregated at units of geography that do not correspond to the jurisdictions of governments that make policy decisions (such as cities). Compounding this issue is the fact that the jurisdictions that do report disaggregated data are not fixed, and their boundaries often change. Third, longitudinal data are severely lacking, which prevents both an analysis of trends over time and an evaluation of the efficacy of policy interventions. These challenges demonstrate the inadequate and piecemeal nature of houselessness race and ethnicity data.

1. Datasets with Race and Ethnicity Information Are Deficient and Vary in Methodology

The RDT team examined four major national sources of houselessness data by race and ethnicity: Point-in-Time (PIT) counts, the Homeless Management Information System (HMIS), the federal Census, and the American Community Survey (ACS).⁹³ These sources are generally compliant with the OMB racial and ethnic categories except for the ACS, as explained below. These datasets each contain limitations and posed challenges for the RDT team.

a) Deficiencies of Datasets That Rely on Continuums of Care for Reporting

A major obstacle to robust race and ethnicity data analysis in the houselessness context is the wide reliance on Continuums of Care (CoCs) for reporting. CoCs are federally designated planning bodies responsible for coordinating the funding and delivery of services for people experiencing houselessness.⁹⁴ Two of the major data repositories studied by the RDT—PIT counts and the HMIS—are collected by the U.S. Department of Housing and Urban Development (HUD), which in turn obtains data from CoCs.⁹⁵ HUD depends on CoCs to report point-in-time and yearly counts of people experiencing houselessness.⁹⁶ CoCs are incentivized to provide data, including racial and ethnic data, to HUD as a condition of receiving federal funding.⁹⁷

Reliance on CoCs for data reporting is problematic for many reasons. CoCs operate at various jurisdictional levels that rarely match the geographic boundaries of local government or other Census geographies that engage in policy making, as demonstrated in Figure 11. Specifically, CoCs often encompass multiple city or

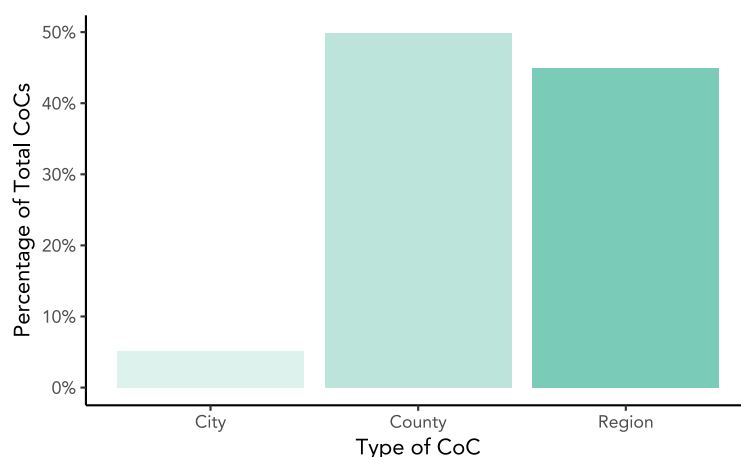


Figure 11: Level of Jurisdiction of Continuums of Care Reporting to HUD

county jurisdictions, or portions of such jurisdictions. They also do not necessarily operate as part of a local government. “In 2020, nearly 60% of Continuums of Care surveyed were organized as a non-governmental entity.”⁹⁸ Consequently, CoC data are usually not disaggregated at the city or town level where houselessness policies are often made, as discussed in more detail below. The arbitrary CoC boundaries also make it challenging to link Census data—including racial and ethnic data—to CoCs in any kind of analysis. Additionally, CoCs use varying collection and data storage methodologies which can bias final counts because they are not comparable (i.e. one CoC may collect data for unique users and another may collect data for every user per day).⁹⁹ And when CoCs merge, as they often do, they change the boundaries under which they operate, precluding longitudinal analysis.

The amount of race and ethnicity data HUD has managed to obtain from CoCs has also decreased over time, likely because CoCs have decreased in number. The RDT’s case study of CoC participation in data reporting for three separate years (2007, 2013, and 2016) vis-a-vis 2020 demonstrates that even the same CoCs do not always participate in data reporting, and that overall participation has decreased since 2007.¹⁰⁰ The RDT team investigated the reasons why some CoCs did not participate in data reporting in 2020, but this information was not widely available. Those that did provide a reason indicated that the CoC had merged with another CoC, had ceased to exist, or had not applied for HUD funding and thus was not required to report.

Separate from their reliance on CoCs, PIT counts and HMIS have additional limitations regarding the quality of their race and ethnicity data. The most complete publicly available, disaggregated data (at the CoC level) on race and ethnicity comes from the PIT counts, which capture only a point-in-time snapshot of houselessness on a given night each year.¹⁰¹ PIT counts add up the number of shelter users and unsheltered individuals during a given night in January.¹⁰² PIT counts have limited value because they cannot be easily compared to other sources of houselessness data, which reflect data collected over the course of a year.

By contrast, data from HMIS estimate the number of unique shelter users in a fiscal year as recorded by shelters' administrative records, but HMIS only publicly reports information on race and ethnicity as aggregated, national counts. Instead of disaggregating data by geographic jurisdiction, these publicly available datasets provide national counts that are disaggregated by type of shelter option, such as family or emergency shelters. Additionally, HMIS estimates rely on shelter reports, which omit information about unhoused people who are not in shelter settings and may introduce other inconsistencies in the data.¹⁰³ HMIS collects race and ethnicity data at the CoC level as well, but these data are not publicly available.¹⁰⁴

b) Deficiencies of Datasets That Do Not Rely on Continuums of Care for Reporting

Houselessness datasets that do not rely on CoCs for data, namely the Census and the ACS, have their own limitations and variations regarding race and ethnicity information. The 2010 Decennial Census included racial and ethnic data on people experiencing houselessness during a three-day period.¹⁰⁵ Advantages of this database include that it employs collection efforts that follow the same methodology throughout the country, unlike those done by individual CoCs, and that it includes some information on unsheltered individuals.¹⁰⁶ However, like the one-day focus of the PIT counts, the three-day unit of measurement is likely not representative of the overall yearly population of people experiencing houselessness, and is difficult to analyze in relation to other standard measures of houselessness that are collected over the course of a year. Additionally, the Census aggregates race and ethnicity data at the state and national level, precluding a better understanding of where and how racial inequities arise at the local level.

Finally, the ACS has collected micro-level data about people in emergency and transitional shelters with sleeping facilities, but it omits data on unsheltered individuals, and the information it provides is not publicly available. One advantage of the ACS is that it collects data in some major cities, which is missing from other sources of collection. While the database includes categories for race and ethnicity, it is unclear what those categories are, as the RDT team could not access the repository.¹⁰⁷

c) Differences in Methodologies Cannot Be Reconciled

The RDT team examined the differences between the houselessness data repositories to investigate whether they could be combined to overcome some of the aforementioned data gaps. The RDT team compiled Table 9, which reveals key differences and deficiencies in the datasets that prevent them from being used in a complementary way.

Table 9. Comparison of Data Sources

	HUD Point-in-Time	HUD Homeless Management Information System	ACS	2010 Decennial
Race data	yes	yes	yes	yes
Ethnicity data	yes	yes	yes	yes
Years Available	2015-2020 (with variables on race, otherwise since 2007)	2007-2017	2006-2020	2010
Jurisdiction available	CoC level, state level, US level	US level	State level, US level, some major cities	State level, US level
Public Availability	yes	yes	no	yes
Period of Collection	One night in January	Annual number of shelter users	Yearly	Conducted march 29-31, 2010
Data on Sheltered Individuals	yes	yes	yes	yes
Data on Unsheltered Individuals	yes	no	no	some
Method of Collection	Individual Reports from CoCs	Individual Reports from CoCs	Survey of People residing in shelter facilities + imputation techniques	Survey of People residing in shelter facilities

2. Race and Ethnicity Data Are Not Sufficiently Disaggregated at Local Levels

Racial and ethnic data on houselessness are generally not disaggregated by municipality or locality, which makes it difficult to design and evaluate houselessness policies that originate at the local level. While many important policy decisions are made by city officials responding to city problems, city-level racial and ethnic demographic data are lacking in the datasets described above. At best, houselessness data are disaggregated at the CoC level, but as explained above, CoC boundaries often do not correspond with local governments. PIT data are available at the CoC, state, and national levels, HMIS data are only available at the national level, Census data are collected for states and the national level, and ACS data are collected for some cities, all states, and at the national level.

When the RDT team tried to supplement these datasets by contacting the relevant departments of the fifty largest cities to request yearly counts on houselessness,

only about 54% responded and, of those that did, none were able to provide the information requested—indeed, most cities referred the RDT team back to the PIT counts collected by their corresponding CoCs. Given the highly localized nature of houselessness policies, the absence of publicly available, geographically disaggregated racial and ethnic data makes it difficult to craft and evaluate the rules, laws, and ordinances that may have the greatest impact on racial and ethnic inequities in the houselessness context.

3. Race and Ethnicity Data Are Not Consistently Available across Time

None of the repositories discussed above consistently or comprehensively collected racial and ethnic demographic data over time. While PIT data have generally been available since 2007, data on race and ethnicity were not included in PIT counts until 2015. HMIS data, by contrast, included race and ethnicity information only between 2007 and 2017. Census data are only collected once every ten years, and only included comprehensive racial and ethnic data in 2010, following partial availability of racial and ethnic data in 2000 and 1990. Finally, racial and ethnic data from the ACS have been available on a yearly basis since 2006, but are not publicly accessible. Figure 12 demonstrates the scarcity of longitudinal data on houselessness and the inconsistency across data sources of the time periods for which race and ethnicity data are available. The absence of such information precludes a comprehensive picture of racial and ethnic disparities over time, which in turn prevents researchers from evaluating the effectiveness of policies intended to promote racial equity.

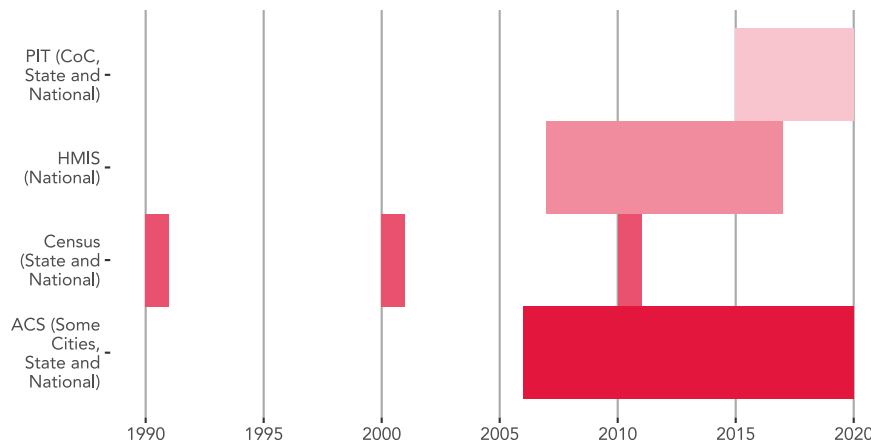


Figure 12: Timeline of Data Availability by Race/Ethnicity of Four main Repositories of Data on Houselessness

C. The Challenges of Obtaining Arrest and Police Violence Data by Race and Ethnicity

The RDT team also identified several critical problems with major national repositories of racial and ethnic data regarding criminal arrests and police violence. For this study, the RDT tried to collect data on (1) overall arrests, (2) arrests for murder and non-negligent manslaughter, and (3) police use of violence. As with the houselessness databases, these sources varied significantly in their methodologies, had many gaps, lacked longitudinal data, and were not sufficiently disaggregated.

1. Datasets Vary in Collection Methodologies

a) Arrest Data by Race and Ethnicity

The RDT team examined three national datasets with racial and ethnic data on arrests: the Uniform Crime Reporting (UCR) Program from the FBI, the Bureau of Justice Statistics, and the National Crime Victimization Survey (NCVS), each of which has its own methodologies and particular limitations regarding race and ethnicity data.¹⁰⁸

The most comprehensive source of racial and ethnic data for arrests in the United States is the UCR. Agencies voluntarily report and submit their data to the federal UCR through a state Uniform Crime Reporting System or directly to the federal UCR via the National Incident Based Reporting System (NIBRS). UCR data are reported at the national, state, and agency level. Information for both instances of arrest (murder and non-negligent manslaughter) was available between 1985 and 2019 through the Crime Data Explorer (CDE).¹⁰⁹ The UCR does not employ the OMB categories; data are disaggregated for race but not ethnicity,¹¹⁰ and the category of “Native Hawaiian” does not include “other Pacific Islanders.”¹¹¹

A second source of race and ethnicity data on arrests is the Bureau of Justice Statistics, which has data on murder, non-negligent manslaughter, and thirty-one other offenses by race between 1980 and 2014. Like UCR data, information from the Bureau of Justice Statistics is available at the national, state, and agency levels. These data are also non-compliant with the OMB standards, as the only racial categories available are white, Black, AIAN and Asian Pacific Islander (“API”), and no data on ethnicity is included.

A third source of data on arrests is the NCVS, a study that has been administered yearly in the United States since 1973¹¹² by the Census Bureau on behalf of the Bureau of Justice Statistics. This survey is presented to a nationally representative sample of approximately 169,000 people ages 12 or older in the United States. The survey asks about reported and unreported incidents of crime, why some incidents were not reported, the contexts of these incidents, experiences with the criminal legal system, self-protective measures used, and substance use in the past six months. For this study, the race and ethnicity of victims per type of crime

was collected from 2005 through 2019, but only at the national level. The only categories used for the data were white, Black, Hispanic, and other.¹¹³

The RDT’s examination of these datasets revealed particularly significant gaps in the availability of data regarding the ethnicity of people who are arrested, so the RDT team attempted to manually collect ethnicity data by searching law enforcement agencies’ websites and states’ websites. Only 32% of states reported data on ethnicity locally for all arrests and for non-negligent manslaughter arrests. The absence of ethnicity data was worse in some places than others: such data were available in 29% of states in the South, 33% of states in the Midwest, 38% of states in the West, and 56% of states in the Northeast. The lack of available ethnicity data regarding arrests precludes tailored policy responses to ethnic inequities in the criminal legal system.

b) Police Violence Data by Race and Ethnicity

Racial and ethnic data on police violence are scarce, and there is no nationwide repository of such information. While the Deaths in Custody Reporting Act of 2014 imposes financial penalties on states that do not comply with certain police violence data reporting requirements, there has been no thorough enforcement of the Act to date.¹¹⁴ The FBI began collecting data on reporting compliance in 2019. As Figure 13 illustrates, the percentage of law enforcement agencies that reported police violence data per state in 2021 was very low: 6,543 out of 18,514 federal, state, local, and tribal law enforcement agencies submitted data (approximately 35%). These data included information on the race of the person against whom the police used violence, officer information (including race), and incident information. However, the only data publicly available to date are the number of agencies reporting and number of incidents reported for each; data on race and ethnicity remain unavailable.¹¹⁵

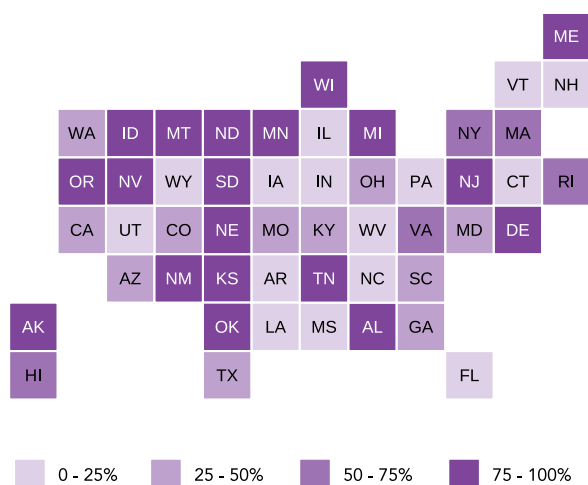


Figure 13: Percentage of Law Enforcement Agencies That Reported Police Violence Per State in 2021

Some states have their own requirements for reporting racial and ethnic data on police violence, but such piecemeal efforts are insufficient to provide comprehensive information about racism in policing. Data from the National Conference of State Legislatures shows that, at the state level, at least twenty-one states require some sort of data collection on law enforcement use-of-force incidents¹¹⁶ but only eight of these states require that the data be publicly reported, only fifteen states specify the need to collect data on race, and only eleven require data on the ethnicity of victims.¹¹⁷ Moreover, definitions of police use of violence and type of data collected vary greatly by state, making the data difficult to compare or aggregate. For example, only sixteen states collect race and ethnicity data on officer-involved deaths.¹¹⁸

In the absence of a reliable nationwide data repository, the Mapping Police Violence Organization is the most comprehensive dataset that collects data on police violence, including racial and ethnic data, but it, too, comes with important limitations. This database includes over 9,000 killings by police nationwide between 2013 and 2020, based on data compiled from a variety of sources, including (1) police use-of-force data collection programs in the small set of states that report publicly; (2) nationwide data from the Fatal Encounters database, a crowdsourced database on police killings; and (3) searches in social media, obituaries, criminal record databases, police records, and other sources. This approach allows the Mapping Police Violence Organization to identify and report the race and ethnicity of 90% of the victims of police violence.¹¹⁹ While deeply impressive in its scope, this database relies on incomplete and often unofficial sources of information. It also generally follows the OMB categories, except that the category of “Pacific Islanders” omits “Hawaiian,” and the “Hispanic” category does not include “Latino” in its name.

Other non-governmental agency groups assemble and analyze race and ethnicity data on policing, but their repositories are also incomplete. The Police Scorecard, for example, collects data on police violence, accountability, racial bias, and policing outcomes for over 16,000 municipal and county law enforcement agencies in the United States. The data are collected from police arrests, personnel, funding, incarceration rates, and homicide clearance rates from official federal and state databases such as the UCR, the Bureau of Justice Statistics’ Annual Survey of Jails, the U.S. Census Bureau’s Survey of State and Local Government Finances, and the California Department of Justice’s Open Justice database. This information is complemented with agency publications and media reports, including the data from the Mapping Police Violence dataset. Each agency is assigned a score by focusing on a number of criteria. Some of their measures take into account race, specifically the variables: “racial disparities in deadly force,” “racial disparities in drug arrests,” “police violence by race,” and “percent of homicides unsolved by race.” While these scorecards are an important indication of agency performance, the data on race are not disaggregated beyond the state level and are not provided in cumulative counts. Moreover, as with other private organizations, the Police

Scorecard relies on sources of information that are incomplete, and it can only create a database that is as complete as its sources.

Finally, the *Washington Post* has a database of fatal shootings by a police officer in the line of duty since January 1, 2015, which contains some race and ethnicity information. The data are collected by looking at local news reports, law enforcement websites and social media, and by monitoring independent databases such as “Killed by Police” and “Fatal Encounters.”¹²⁰ The categories of race and ethnicity are mostly consistent with the OMB regulations, but the category of “Hawaiian and other Pacific Islander” is missing. While the *Washington Post’s* database is useful, it is also incomplete, as it does not include any police shootings or incidents of police violence that do not attract media attention.

In sum, racial and ethnic data concerning police violence are woefully incomplete and missing, hindering the creation of policies that effectively address racialized policing.

2. Longitudinal Data Are Scarce

Longitudinal data by race are generally available for the past several decades for criminal arrests, but are essentially nonexistent for police violence.

The longitudinal arrest data come with several caveats. The UCR has longitudinal data on arrests with racial data (but no information on ethnicity) that spans from 1985 to 2019. The Bureau of Justice Statistics provides racial (but, again, not ethnic) data from 1980 to 2014, but with many gaps in reporting from law enforcement agencies. The NCVS has longitudinal data on race and ethnicity from 1973 to 2019, but this is also of limited value because the racial and ethnic categories it utilizes have changed over time. Moreover, because the NCVS focuses on self-reported victimization data, it uses a very different collection methodology than the prevalence counts¹²¹ from UCR and the Bureau of Justice Statistics. Finally, all these sources of information use different race and ethnicity categories. As a result, it is harder to compare, complement, or impute data between data sources because they do not represent the same groups of people.

Efforts to collect race and ethnicity data on police violence are too recent to provide meaningful longitudinal data. Mapping Police Violence has tracked police violence data since 2013, and the *Washington Post* has tracked such data since 2015. Nationwide data before those dates is unavailable.¹²²

3. The Level of Race and Ethnicity Data Reporting Varies Greatly from State to State

Crime-related data submission to the federal government is voluntary, so many agencies simply do not submit data, including race and ethnicity data, to the UCR. As a consequence, race and ethnicity data submission levels vary greatly from region to region and from year to year. Figure 14 shows the percentage of agencies reporting

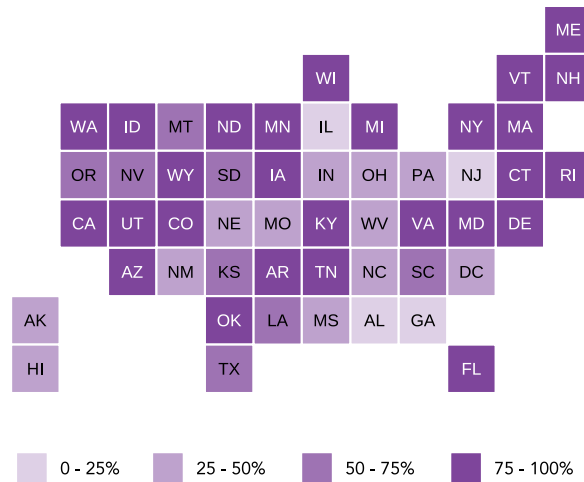


Figure 14: Percentage of Law Enforcement Agencies That Reported Per State in 2019

per state in 2019. The differences are stark: while 100% of agencies reported race and ethnicity data in Connecticut, less than 0.1% of agencies provided that data in Illinois. Year-to-year collection also varies greatly; for instance, the city of Boston did not submit race and ethnicity data in 2019 (for the most recent dataset) but did submit such data for 2018.

4. Geographic Boundaries for Data Collection Often Overlap

There are 18,000 different law enforcement agencies nationwide, which greatly complicates efforts to systematically collect racial and ethnic data on arrests and police violence. These agencies operate at different units of geography, including city, county, and regional levels. Figure 15 demonstrates how law enforcement jurisdictions correspond with various units of geography, including city, county, and regional jurisdictions among the fifty largest cities in the United States.

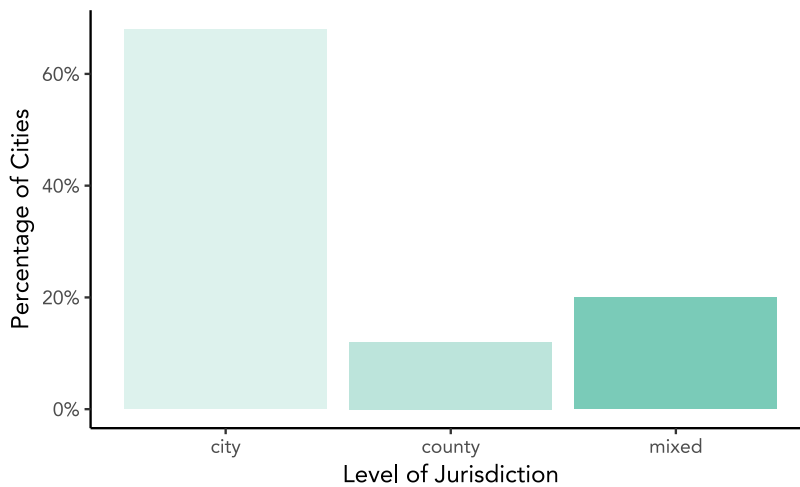


Figure 15: Type of Jurisdiction(s) Operating in the 50 Largest Cities in the U.S.

While a majority of cities have one law enforcement agency operating at the city level, at least 20% of them have more than two agencies that have the authority to arrest individuals in the city. This creates agency overlap, where multiple agencies engage in policing within one jurisdiction. For example, in New York City, the New York Police Department (NYPD) and New York City Transit can both make arrests, which they separately report to the federal government. Similar overlap exists with housing authorities and university-based police departments. Some cities also rely on county sheriff's offices for some or all of their policing. In Los Angeles, the Los Angeles Police Department (LAPD) and Los Angeles County Sheriff Department (LASD) are separate entities; the former is for the city of Los Angeles, but the LASD policed public hospitals, nine community colleges, and public transit (until policing of public transit transitioned to LAPD in 2017). Figure 16 shows the percentage of overlapping law enforcement agencies that operate in the fifty largest cities of the United States.

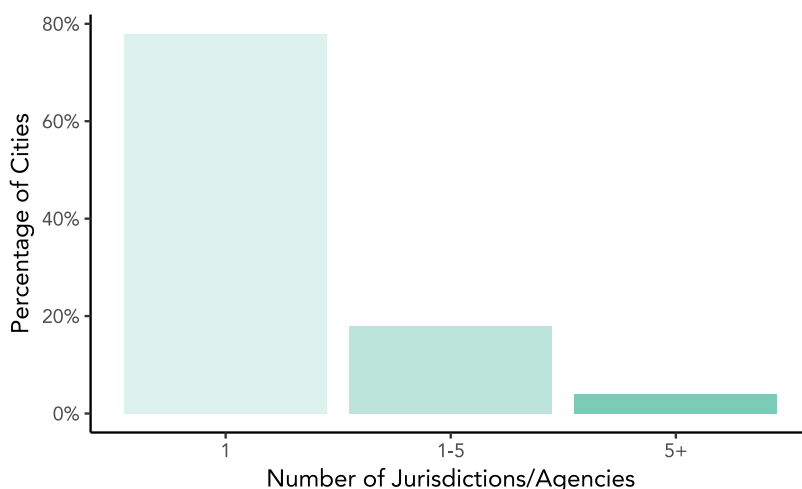


Figure 16: Number of Law Enforcement Agencies Overlapping in the 50 Largest Cities in the U.S.

Agency overlap complicates racial and ethnic data collection. In attempting to collect such data from police jurisdictions for the fifty largest cities in the United States, the RDT team identified a series of problems that have been corroborated in other studies.¹²³ When two or more overlapping law enforcement entities have separate collecting repositories, separate jurisdictions, and potentially different collection methodologies, data on race and ethnicity are likely to be incomplete. Just obtaining race and ethnicity data from the city police force, for example, might miss data on arrests made by transit, county, and university police forces. This is compounded by the fact that policing jurisdictions change over time. Moreover, no geocoded national data exist that would allow researchers to identify the number of police forces operating within a jurisdiction—and to assemble their race and ethnicity data in a way that aligns with geographic boundaries. Nor is it always clear which law enforcement agencies have the power to arrest within city limits, even if they do patrol a city. Additionally, while city level authorities will enact policy on

crime for their jurisdiction, they will not be able to evaluate the exact prevalence of the issue by race and ethnicity if the data are collected at another level. This array of challenges precludes the accurate measurement of racialized law enforcement at the local level, as well as targeted public policies that might redress it.

D. Summary

The RDT's first wave of data collection efforts confirms that existing datasets on houselessness, arrests, and police violence that report counts by race and ethnicity do not provide sufficient information to policymakers and advocates interested in combating racial inequities, particularly at the local level. Like the state-reported data that informed the CRDT, the local, regional, and state data on houselessness, arrests, and police violence are incomplete, uncoordinated, and unreliable. Existing datasets cannot be used side by side to try to fill these gaps because their methodologies are too varied. Due to the organization and structure of existing data collection entities, data disaggregated at the local level (where many policy choices are made) are unavailable.

The work of the CRDT and RDT, analyzed together, affirms that state and local data collection infrastructure must be financed and strengthened, and that such systems should report to a single, standardized, nationwide system of data collection and reporting by race and ethnicity.