



Why policing data matters to safety and equity

Scarlet Neath, Tracy Kawabata-Perrett, and Damon McCullough

Community members and local government officials in St. Louis, Missouri have been using data analysis as part of their ongoing work to reimagine public safety. Last year, CPE found that Black people in St. Louis were [2.3 times more likely to be stopped while walking](#) than White pedestrians. This evidence of disparities motivated continued partnership, and CPE analyzed the city's policing—or “departmental”—data again more recently to answer questions of how to reduce this biased behavior and improve how the city delivers public safety.

The resulting [report](#) shed light on why many residents felt that their safety needs were not being met. Data analysis showed that patrol officers were being deployed to all neighborhoods at roughly the same number, despite some neighborhoods having

significantly more residents, calls for service, or both. The data also pointed to potential solutions. It showed that 18% of routine calls for service in the city were for minor issues, such as parking violations, that could be redirected to non-police responders.

This is just one example of how using police data to reduce harm not only lays the groundwork for measuring any problems that exist in police behavior. It can also identify promising paths forward to create structural change.

CPE believes that publishing complete and transparent data to measure the harm inflicted by policing upon Black and Brown communities is an important step on the road toward ending that harm through large-scale changes. Measuring harm is important because it can motivate immediate action toward structural change. It is also of value because it provides a record of truth. By helping communities to accurately document racial disparities in police contact, we aim to support longstanding, critical efforts to acknowledge, accurately measure, and address harms inflicted on Black and Brown people. This includes Ida B. Wells' 19th century [work](#) to capture and publish statistics about lynchings. It also includes recent initiatives to [document](#) where homicides are frequently unsolved, and to [accurately count](#) where incarcerated people live, to name a few.

Departmental data can act as a powerful fuel to ignite change. It can be used alongside many other types of data to reveal the source of a problem. It can also drive more informed decisions about where communities should invest their resources to maximize public safety. Once these decisions are made, data can also identify what's working over the long run to reduce harm and deliver safety. This helps build important evidence for policies that can be widely adopted.

Too often, however, departmental data is unusable—it's buried in spreadsheets in police computers. The communities who are most affected by disparities and other harms in

policing deserve access to comprehensive and accessible data about public safety issues. This brief explains how CPE and others are working to make data work for communities.

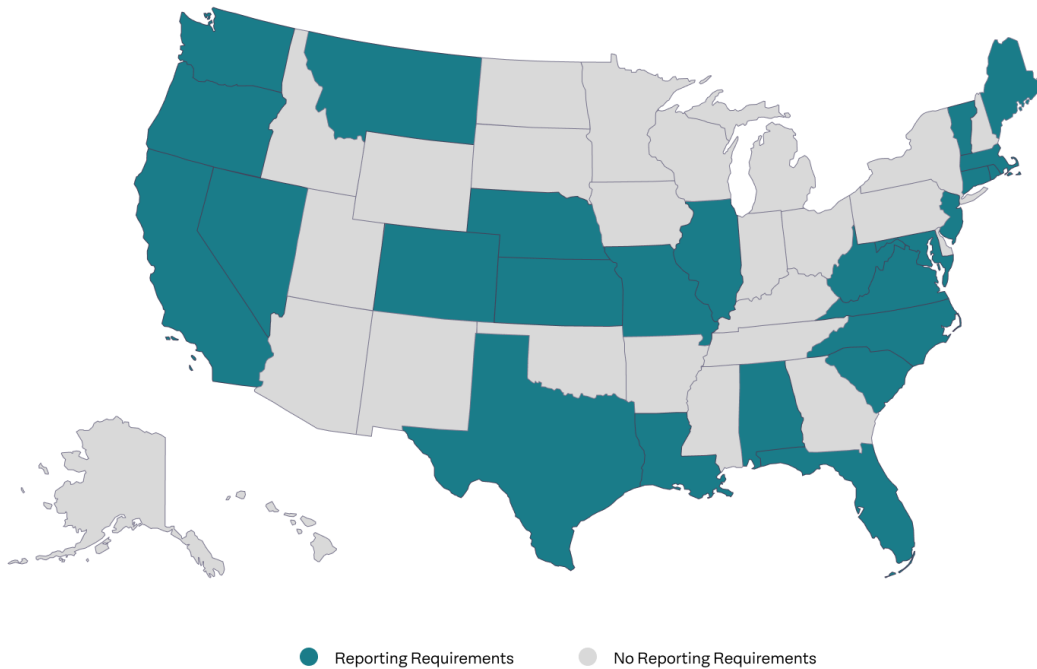
The current landscape of policing data collection

Most police departments collect some kind of data about their basic activities, such as stops. But they do so in very different ways, and may not publish their data. That means it's difficult to consistently use departmental data to answer even basic questions about how police interact with the communities they are expected to serve. For example, some departments don't include the racial group of the person stopped in the data. This means that data can't be used to uncover whether there are racial disparities in their officers' behavior. When hundreds or even thousands of cities don't collect and publish data with the racial group, it makes it difficult to answer bigger questions about the scale of racial disparities in policing and what is working to limit them.

Because of a lack of nationwide data collection standards, departments are largely responsible for deciding what data to collect, analyze, and share. The FBI has published common data standards for tracking some types of use of force incidents nationally. This program aims to use the data collected according to its standards to publish a public database. But collecting use of force data is still voluntary in many states, and low participation by local police departments [threatens](#) the future of the FBI database project.

States are taking initiative to pass their own legislation requiring every local police department to collect data on police stops and use of force incidents. However, these requirements vary widely, and largely do not require that departments collect and publish complete data on common police interactions with community members. CPE is working with communities and policymakers to improve policing data standards in legislation. We

have also published guidance on [best practices and lessons learned](#) for data collection so far.



Use of force reporting requirements by state

Even when departmental data is collected well, it can be difficult to analyze alongside other data sources to answer questions about disparities and public safety. CPE’s Justice Navigator platform helps communities maximize departmental data to drive decisions about public safety. To do this, CPE collects, cleans, transforms, merges, and analyzes departments’ data. The results of this process are displayed in a [Justice Navigator assessment](#) that is accessible to the public. The Justice Navigator currently analyzes departmental data on stops and use of force incidents, as well as public crime and demographic data. And soon, it will also analyze data on communities’ calls for service. Results from the Justice Navigator assessments—as well as other CPE analyses of

departmental data, such as in St. Louis—are collected and analyzed as part of the [National Justice Database](#).

CPE's [methods](#) are not the only way to responsibly prepare and analyze policing and other public safety data. But too often, what goes on behind the scenes of data analysis is not shared with the users of that data. We hope that the examples in this brief of how we make decisions with data can help inform the development of responsible standards for data analysis. We also hope this information will shed light on what makes the Justice Navigator a unique and valuable tool. With this information, community members can more effectively advocate for their department to receive an assessment. Ultimately, this helps more people use policing data as a tool for change.

Why standardize data?

“Standardized” data shows information in a clean, consistent format so that it can be easily analyzed. Our approach to standardizing data means what you’re seeing in the Justice Navigator has been rigorously checked and meets a high level of consistency and completeness. Standardizing data also makes it possible to run automated analyses without having to reinvent the wheel each time, for each of the thousands of departments in the country with different approaches to data collection. It also makes policing data more useful: It creates the ability to compare data across different departments to uncover common trends. These trends are summarized in other JN Data Briefs in this series.

How the Justice Navigator put policing data to work

CPE puts policing data through a standardization process to make sure that the data in the Justice Navigator assessments are measuring disparities in actual police behavior as accurately as possible.

Below is an overview of the steps we take to get policing data from an officer's computer into the Justice Navigator.

Request and review the data

We first give departments a detailed list of the exact data they need to pull from their computer system in a spreadsheet format. Departments give us spreadsheets of data along with a codebook—or guide to what data labels mean—so that we can decode them. For example, the codebook would tell us that the shorthand label “AN” corresponds to “Asian” and not “Alaska Native.”


We then review a sample of the pulled data to make sure that it's capturing what we want. For example, we might find that data on use of force doesn't include incidents with officer-involved shootings, which were stored in another file. We then ask the department to provide that file, so that we are analyzing data on use of force that are as complete as possible. We also review the sample to make sure that we minimize any sensitive data. We confirm that all names of officers and individuals are replaced with “unique identifying numbers” that allow us to track specific incidents responsibility, without risking their privacy.

Transform the data

Once we've assembled the most complete set of data files the department can provide, we start analyzing the data. To make the data into something we can confidently analyze for racial disparities, we need to answer two main questions:

1. Are incidents counted in a way that measures the actual amount of police contact?

We want to make sure the data is measuring the full story of the recorded experiences community members have with police. CPE counts one incident of police contact—captured in one row of data in a spreadsheet—as a single person experiencing contact with police. The number of officers involved in an incident does not change the fact that a single community member was targeted for police contact in that incident. We often see that departments' data count "incidents" differently than we do. For example, the same incident might appear in 5 different rows of data, because there were 5 officers there. CPE "deduplicates" the data, or deletes these extra incidents, as part of transforming the data into our standardized data language.



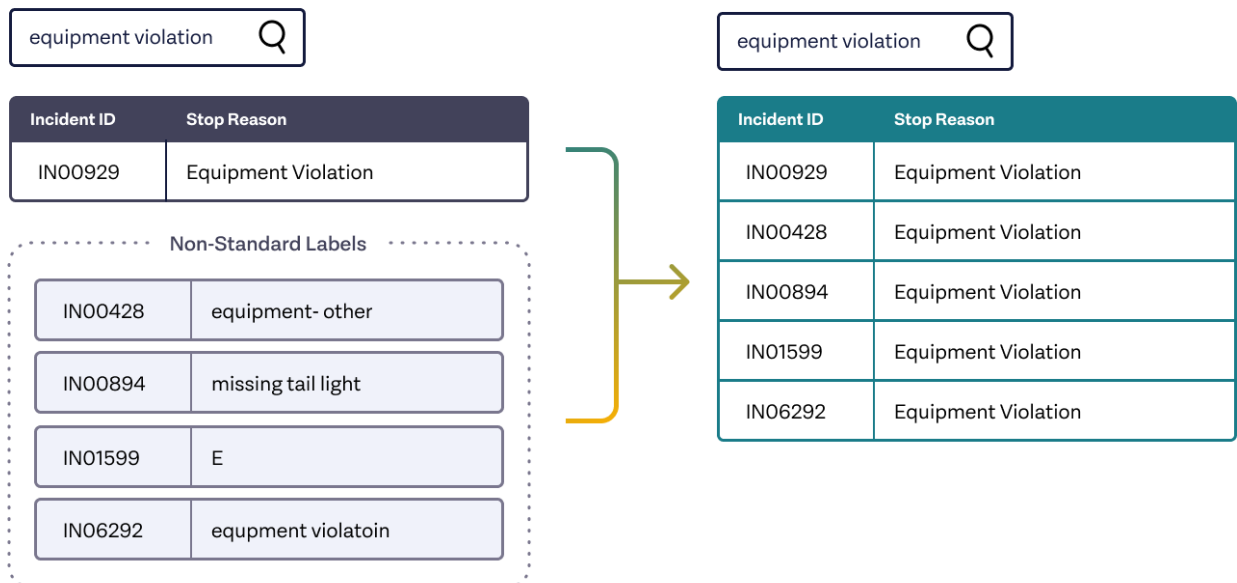
Incident ID	Incident Date	Person Stopped	Officer ID
IN00117	07/07/2018	CIT25	OF001C
IN00019	03/20/2018	CIT16	OF000A
IN00019	03/20/2018	CIT16	OF000B
IN00019	03/20/2018	CIT16	OF000C
IN00019	03/20/2018	CIT16	OF000D
IN00019	03/20/2018	CIT16	OF000E
IN00722	04/29/2019	CIT33	OF00BB

Incident ID	Incident Date	Person Stopped	Officer ID
IN00117	07/07/2018	CIT25	OF001C
IN00019	03/20/2018	CIT16	OF000A OF000B OF000C OF000D OF000E
IN00722	04/29/2019	CIT33	OF00BB

Incident data before and after deduplication

2. Do the department's categories match our common language?

We then need to standardize the different data labels (or “values”) that capture what happened at each incident. These values might include searches, type of force used, the reason for a stop, or an officer’s work assignment. Transforming a department's value of “missing tail light” into our standard label of “equipment violation” helps us more easily reveal important stories about disparities in reasons for vehicle stops. Equipment-related reasons for stopping cars may be less related to public safety than, for example, speeding violations. Grouping these values together allows us to see at a glance whether they are together contributing to any racial disparities in vehicle stops. The details of this translation and matching work are available in each Justice Navigator assessment’s [Data Notes page](#).

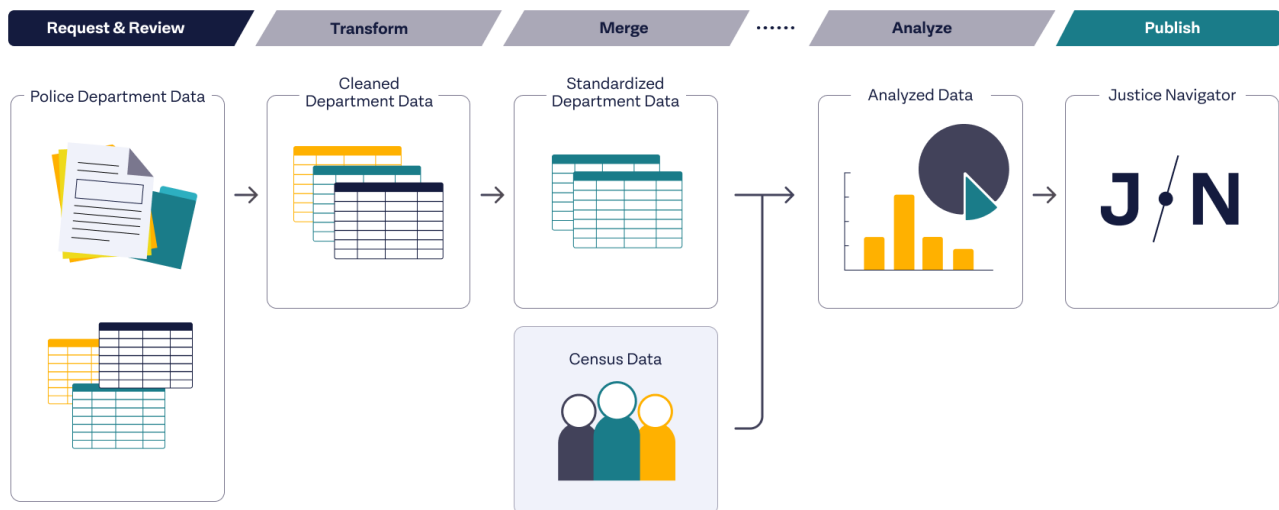


Incident data before and after standardization

Merge the data

By itself, departmental data can tell us big-picture information about police behavior. For example, it can show that there were 3,127 pedestrian stops of Black people in a year, but it doesn't tell us whether that is a high or low number, compared to the number of Black residents. To answer questions about racial disparities in police behavior, we combine the standardized departmental data with public Census data about the city's resident population. Because public safety problems and police responses are hyper-local, we compare police behavior and the resident population at the neighborhood level.

To add in Census data, we first map the address where a given incident took place to geographic coordinates. Then we group those coordinates into Census tracts, or neighborhood-sized areas defined by the Census Bureau. We can then match the department's dataset to Census population data about specific tracts to identify disparities at the neighborhood level. We can also match the department's data to population data about the whole jurisdiction to identify disparities at the city level.



Check the data

We use a combination of automation and human-led processes to check our results for possible errors. For example, an automated check might flag that there are no stops in the month of May. This looks like a potential error because there were more than 100 stops in April and June of that year. We would confirm with the department whether no stops occurred in May, or whether the missing data can be explained by a human or computer error.

Each of our Justice Navigator analyses—the charts and visualizations displayed as results—has different standards. These standards help us tell a more complete and accurate story about the officers' recorded interactions with community members. Many of these standards are for data "missingness." Data missingness means the number of incidents that are missing values, such as the racial group of the person stopped, or the reason for the stop. One important missingness standard is: Are more than 15% of values that are relevant for an analysis missing? For CPE, the racial group of the person who experienced police contact is often a relevant value. This is because we're always analyzing the data to answer questions about racial disparities. We also have standards about the number of incidents required to run an analysis. For example, even if data is complete, we do not perform [complex statistical analyses](#) for datasets with only a few dozen incidents for a racial group. These advanced statistics require a certain number of incidents to produce results we can be confident in.

Ask questions and analyze the data

Data that is cleaned and standardized can then be put to work to reveal useful information. We use a variety of statistical methods to answer our standard [research questions](#) about racial disparities in policing. These research questions, and our approaches to answering them, evolve based on the latest scientific knowledge and community priorities. We are

continuously improving how we use data to answer urgent questions about racial disparities.

The research questions we ask aim to reveal the size of racial disparities in police behavior. They also uncover potential sources of those disparities. For example, to analyze pedestrian stop data, we look at [who was stopped while walking, who was searched after being stopped](#), and whether any racial disparities [can be explained](#) by the influence of different neighborhood-level factors. Standardized data can also be analyzed to answer specific questions that come up in the community. For example, in St. Louis, we examined the geographic location of use of force incidents, as well as the different types of calls that were associated with incidents in different neighborhoods.

The Justice Navigator is a custom website that transforms our research questions and departments' standardized data in spreadsheets into easy-to-read charts and graphs. CPE staff then conduct a final series of human-led checks on the visuals, and publish them only once we are confident that the results reflect the original data correctly.

Make the data results public

Publishing the results of any data analysis and information about the process used is an essential part of making policing data useful. Access to data analyses allows communities to engage with the department and other stakeholders to design public safety systems that meet their needs. Sometimes a barrier to change is lack of agreement about what needs to change and why. Publicly accessible data analyses—such as those found in the Justice Navigator—can create opportunities for data-driven solutions by establishing a common understanding of the size and scope of disparities in policing.

To ensure that data are available to communities, we require all departments we partner with to make their Justice Navigator assessments public. We also support departments in

improving their data collection practices, using Justice Navigator results to drive change, and planning events to share their results with their communities.

How to improve policing data that is collected

This brief gives an overview of how the Justice Navigator transforms policing data to reveal useful information. However, to be able to participate in the Justice Navigator (or similar initiatives) departments must consistently collect specific kinds of data. To learn about what data departments should collect, and how to make sure data are accurately collected, visit www.justicenavigator.org/for-communities/improve-your-data.

Acknowledgments

Thank you to Payton Dougherty, Kasey Eickmeyer, April Garner, Matthew Graham, Robert Kenter, Tracey Lloyd, Christopher Mebius, Ural Mishra, Kerry Mulligan, Hilary Rau, Joelle Lee-Silcox, Lucy Odigie-Turley, and Sarah Beth Weintraub for their valuable feedback.