Data & Civil Rights: Criminal Justice Primer

by Alex Rosenblat, Kate Wikelius, danah boyd, Seeta Peña Gangadharan, and Corrine Yu
Produced for Data & Civil Rights Conference / October 30, 2014

Discrimination and racial disparities persist at every stage of the U.S. criminal justice system, from policing to trials to sentencing. The United States incarcerates a higher percentage of its population than any of its peer countries, with 2.2 million people behind bars. The criminal justice system disproportionately harms communities of color: while they make up 30 percent of the U.S. population, they represent 60 percent of the incarcerated population. There has been some discussion of how "big data" can be used to remedy inequalities in the criminal justice system; civil rights advocates recognize potential benefits but remained fundamentally concerned that data-oriented approaches are being designed and applied in ways that also disproportionately harms those who are already marginalized by criminal justice processes.

Like any other powerful tool of governance, data mining can empower or disempower groups. The values that go into an algorithm, and the metrics it optimizes for, are baked into its design. Data could be used to identify discrimination in current practices, or to predict where certain combinations of data points are likely to lead to an erroneous conviction. When algorithms are designed to improve how law enforcement regimes are deployed, the question that data analytics raises is, which efficiencies are we optimizing for? Who are the stakeholders, and where do they stand to gain or lose? How do these applications intersect with core civil rights concerns? Where can we use big data techniques to improve the structural conditions criminal justice system that lead to disparate impacts on marginalized communities? How do we measure that impact, and the factors that lead to it?

Background: Discrimination in Criminal Justice

Major themes and existing challenges in the U.S. criminal justice system:

- *War on Drugs*: Even though race/ethnicity are not a significant factor in the use or distribution of drugs, Blacks and Hispanics comprise 62 percent of those in state prisons for drug offenses.[1] According to a 2012 federal report, more than seventy percent of all persons sentenced for federal drug trafficking offenses were either Black (25.9 percent) or Hispanic (46.2 percent), many of whom often face harsh mandatory sentences.[2]
- *Racial Profiling*: Law enforcement actions that single out individuals based not on individual behavior, but instead on the basis of race, ethnicity, national origin, or religion, disproportionately target minorities as criminal suspects, skewing at the outset the racial and ethnic composition of the population ultimately charged, convicted, and incarcerated.
- *Police Misconduct:* While strides have been made in the areas of police misconduct and brutality, incidents such as the shooting of unarmed African-American teenager Michael Brown in Ferguson, Missouri show us that police continue to use force disproportionately (both in terms of frequency and intensity) against people of color.
- *Mandatory Minimums:* The proliferation of mandatory minimum penalties, particularly at

the federal level as a result of the War on Drugs, has harmed minority communities and fueled the country's incarceration rates. In an analysis of nearly 80,000 cases in 2010, the U.S. Sentencing Commission found that nearly 25 percent of offenders were sentenced to a mandatory minimum penalty.[3]

- *Barriers to Re-Entry*: Incarcerated individuals, especially racial minorities, face a number of challenges during their imprisonment and upon re-entry, including restrictions on interaction with their families, limited access to medical care, voting rights restoration, and employment discrimination.

"Big Data" and Criminal Justice

Over the last decade, many states have adopted big data technologies and practices, compiling large databases on their populations and deploying risk-assessment tools that analyze this data to set individuals' conditions of confinement, probation, or parole.[4] Other arms of the criminal justice system, like the police, are adopting data-driven techniques for targeting potential offenders, as well as predicting crime "hot spots" or areas of town likely to contain high rates of criminal activity.

The application of big data tools and practices in a criminal justice context raises questions about the kinds of data used for analysis and consequences of error, bias, or inaccuracies, including problems of cumulative disadvantage.[5] Data mining works most effectively with data containing binary characteristics: an email is spam or it's not.[6] The rules of categorization for these two types of email are clear, and the potential consequences of a misclassification are fairly minor: routine scanning of email in the spam box can easily rectify the problem. When an algorithm calculates the profile of a likely or potential criminal, or of someone who deserves a short sentence or a long sentence, classificatory schemes entail complex (non-binary) determinations. Data-driven outcomes represent the potential for bias and error to be systematically propagated on a much larger, non-local scale, and criminal justice professionals may not have the technical expertise to detect or address these risks.

Data mining techniques use past data to "train" algorithms and generate predictions about new situations. As a result, biases in the training data can lead to biases in algorithmically p outcomes. For instance, as law professor Frank Pasquale observes, "Drug or gun possession is as likely among whites as it is among racial minorities, but in New York City, racial minorities comprise the vast majority of persons who are stopped and frisked. Disproportionately more nonwhites than whites, therefore, will end up with criminal records for gun or drug possession."[7] If an algorithm uses this data on drug or gun possession to predict who is likely to be in possession of these in future, than these disproportions could be reflected in how an algorithm learns to predict which characteristics, like race, are indicators of potential criminal activity. However, an algorithm that is constantly updating probabilities based on new data inputs could potentially weaken the prejudicial element, if it was not present in the evolving data set.

More broadly, incarceration rates tend to affect disadvantaged communities, and particularly communities of color. The rate of incarceration per 100,000 people in 2005 was 412 for Whites, 742 for Hispanics, and 2,290 for Blacks.[8] Approximately half of all imprisoned offenders are incarcerated again within three years of their release.[9] When algorithms rely on the characteristics of convicted or arrested populations to predict persons who are likely to commit crime, they solidify a history of bias against those already disproportionately targeted by the criminal justice system. Similarly, a data-driven sentencing algorithm may reflect that past presumptions of criminal justice professionals about which traits best correlate with crime. Algorithms designed to find correlations from these prejudiced data would produce discriminatory outcomes.

Human and Machine Bias

Though inaccurate classifications (false positives and false negatives)[10] can result from both human-driven assessment systems and data-driven ones, society-wide faith in machine based-judgment can often overshadow problems of error, bias, and inaccuracies in automated decisions. Proponents often tout data analytics as a way of removing human bias from a range of law enforcement activities. Predictive analytics suggest that if a prosecutor or a probation officer can punch in the characteristics or actions of the subject in question, the algorithm can provide probabilities about that subject's future actions based on how similar previous known individuals have responded to interventions. In theory, such an approach could serve to standardize results across the board and equalize the treatment of different populations. This is particularly important because there is tremendous evidence that shows unequal treatment and remedies. For example, a recent Justice Department investigation into the Shelby County juvenile court system in Memphis that found that black children were consistently punished more harshly than white children.[11]

When algorithmic determinations are flawed, and individuals can manually override computer decisions (and these overrides can be biased, too), they may be hesitant to do so. "The algorithm told me to" can become a guiding rationale for people as they become more reliant on technology, to the point that acting contrary to algorithmic suggestions produces anxieties about being held liable for doing the wrong thing. For example, police held an African-American woman at gunpoint when an automated license-plate reader misidentified her vehicle as stolen. Though an officer noticed a discrepancy, police arrested the woman for possessing a stolen vehicle on the basis of the red flag generated by the license plate reader.[12] While manual overrides of computerized results and individualized decisions are not necessarily more fair, it is important to consider how automated decisions often come with an implicit, technophilic promise of accuracy and fairness that they do not necessarily deliver (even if the users are cautioned about their limitations by the designers). Given that any machine learning system will produce results that have error rates, how do we ensure that the people applying these technologies understand their limitations? How do we balance between the biases introduced by people and those introduced by technology?

Algorithmic analysis can also outpace a human's ability to accurately categorize patterns of behavior, raising questions about whether, when, and how algorithmic determinations should complement or replace human judgment. Since 1994, the New York City Police Department has been using a data-driven management system called CompStat, which organizes all of the data the police receive from official sources on crime development efficiently; it has a geographical component that produces maps of crime hot spots. The program has been adopted widely by other U.S. cities.[13] There is some evidence that computerized geographic mapping of crime hotspots have made policing more effective, partially because there was a significant drop in violent crime after CompStat, and other similar systems, were deployed, though it is not conclusive, and other factors might better explain the reductions in crime.[14] Generally, causal connections are hard to draw in this area because there is a small body of research into the kinds of policing strategies that are the most effective in reducing crime in the long-term, particularly with a focus on hotspots, and because confounding variables make it difficult to compare the effectiveness of different policing strategies.[15] However, its proponents do assert that data-driven policing improves public safety.[16]

Currently, individuals have few means to confront or challenge flawed algorithmic determinations. Chicago's police department recently took the mapping process a step further, adopting an algorithm that generates a 'heat map' not of places, but of *people* deemed at risk for perpetrating violent offenses.[17] Research by sociologist Andrew Papachristos suggests that the people who are more at risk for violent crime are visible through their social networks (e.g. within a given neighborhood, or even a hotspot, some people are more at risk than others). Violent crime is 'thicker' around certain nodes of a network, and thus, predictions can be made around who is likely to be at risk for involvement in a violent crime.[18] When a member of the police department showed up at the house of Robert McDaniel to announce that police had identified him as at risk and placed him under informal police supervision, McDaniel was incredulous: he had never committed a violent offense, nor interacted with the police recently, and yet the algorithm pointed to him as a likely culprit.[19] How does an individual like McDaniel challenge that algorithmic calculation? Is an algorithmic profiling mechanism preferable to other forms of profiling? How does social-graph mapping interact with our notions of justice, fairness, and safety? In defining the power that algorithms have, we can locate culpability and responsibility in ways that are meaningful and provide channels of recourse to the people affected by algorithmic outputs.

Selecting Attributes for Analysis

Some of the factors now used in criminal justice algorithms put pressure on basic notions of justice, fairness and due process. In examining sentencing algorithms, law professor Sonja B. Starr describes, "The basic problem is that the risk scores are not based on the defendant's crime. They are primarily or wholly based on prior characteristics: criminal history (a legitimate criterion), but also factors unrelated to conduct. Specifics vary across states, but common factors include unemployment, marital status, age, education, finances, neighborhood, and family background, including family members' criminal history."[20] When a sentencing algorithm translates these other factors into a risk score, it can impose disproportionate punishment on those who carry the socio-economic markers of poverty, relative to others convicted of the same crime. Even when such an algorithm excludes protected class characteristics from its calculations, other factors or characteristics can act as accurate proxies for these, which can pick out the same populations of color for special disadvantage.[21]

However, not all data-driven risk assessments involve suspect variables. For instance, researchers at the Laura and John Arnold Foundation found that low-risk defendants are frequently imprisoned to await trials, and that higher-risk defendants accused of violent crimes are often released.[22] After developing a pretrial risk-assessment tool called the Public Safety Assessment-Court (PSA-Court)—one which excluded education, socio-economic status, and neighborhood from its calculations—researchers found that a defendant's criminal history and the charges pending against them most reliably predict future criminal behavior.[23] (Nevertheless, disproportionately racialized arrest and incarceration rates mean that communities of color will still be systematically penalized by any risk assessment tool that uses criminal history as a legitimate criterion.)

A risk assessment tool that avoids using obvious markers of socio-economic status may reduce disparities in patterns of imprisonment. For instance, a low-risk offender who is sitting in jail awaiting his trial likely does not have the money to pay bail or obtain a good lawyer who can articulate to a judge that her client is not a flight risk. If an algorithm makes up for a defendant not having resources to contest their pre-trial standing, does this make our justice system more efficient? Alternately, does this use of data perpetuate a broken system that otherwise might be reformed to avoid such errors?

Potential Uses of Big Data for Civil Rights

Currently, the focus of data analytics and its application to the criminal justice system is on predictive policing algorithms and data-driven sentencing. However, these applications do not solely define the application of big data techniques to the field of criminal justice. Some see data analytics solutions as a method for removing the human bias factor from a range of law enforcement activities. As indicated above, the possibility of standardizing results and equalizing the treatment of different populations is a significant and important driver in developing big data techniques.

There are also tremendous opportunities to use large-scale data to better understand dynamics of racial profiling and police misconduct. In the state of North Carolina, the Southern Coalition for Social Justice has worked with police data on traffic law enforcement stops, a dataset that reaches back to 2000 (for state highway patrol) and 2002 (for all other police agencies), to discern patterns of racial profiling.[24] By standardizing data collection practices and increasing certain types of data collection, there are increased opportunities to perform comparative analysis. For example, Measures for Justice (MFJ) designs tools to assess the comparative performance of criminal justice system across jurisdictions; the goal is to aggregate data from local criminal justice systems to get the big picture on systematic inefficiencies and inequalities.[25] For MJF, the absence of empirical data with which to compare the performances of each part of the system—including prosecutors, administrators, defense attorneys, etc.—is a major barrier to identifying nodes in the network that could benefit most from intervention. MFJ is using big data to create transparency, such that anyone, rather than an expert, can look at the data and identify widespread problems; and policymakers can see more easily envision roadmaps for change.[26]

Beyond their enforcement capacities, law enforcement can also use inferential models to protect vulnerable groups. For instance, in the finance sector, banks have started inferring patterns of human trafficking from the financial transactions they process, and sharing that information with law enforcement.[27] JP Morgan Chase reasoned that since both human trafficking and money laundering involve hidden transactions, they can apply analytics technology to detect both kinds of criminal behavior. Palantir Technologies and the National Center for Missing and Exploited Children work together to analyze persons, businesses, and websites that potentially involve human trafficking and automate the identification of red flags.[28] The same mobile technologies that generate 'big' data and facilitate trafficking can also inform law enforcement strategies to combat it and to support the civil rights of vulnerable people.[29]

Questions for Data, Civil Rights, and Criminal Justice

1. Where is anti-discrimination law unable to meet the challenges presented by data mining and networked data outputs?
2. Where do automated systems create efficiencies, and what are the potential costs and benefits of these efficiencies for marginalized communities?
3. How can data analytics be used to correct historical biases in the criminal justice system, minimize inequities, or to reduce high rates of incarcerations overall?
4. How can data analytics be used to measure which variables lead to the most or least discriminatory impact on marginalized communities?
5. What policies or tools can we have in place to remedy errors, or to hold data-driven decision-making processes accountable? How can individuals confront flawed algorithmic determinations?
6. How do we identify which part of an algorithmic calculation leads to a discriminatory result? What are the technical and policy issues at play?
7. What incentives are driving the development of different technologies? How do we evaluate and debate these incentives before they are built into the tools that are deployed?
8. If we are going to make changes to our criminal justice system using big data techniques, what should we optimize for?
    a. Should we aim to reduce incarceration rates overall?
    b. Can data analytics help us identify new variables that have a maximum impact on racial and minority disparities in criminal justice?
9. Should we eliminate the use of certain factors altogether (such as education, socio-economic status or outside information like social media information) or conversely only use certain factors (like type of offense) in making determinations?

---

[1] Bureau of Justice Statistics, *Prisoners In 2012: Trends in Admissions and Releases, 1991-2012*, by E. Ann Carson and Daniela Golinelli, NCJ 243920 (Washington, D.C., 2013), 12-13. See table 10.
[2] U.S. Sentencing Commission, *Annual Report*, USSCFY12, (Washington, D.C., 2013), Table 34. http://www.ussc.gov/sites/default/files/pdf/research-and-publications/annual-reports-and-sourcebooks/2012/Table34.pdf.
[3] U.S. Sentencing Commission, *Report to Congress: Mandatory Minimum Penalties in the Federal Criminal Justice System*, (Washington, D.C., 2011). http://www.ussc.gov/sites/default/files/pdf/news/congressional-testimony-and-reports/mandatory-minimum-penalties/20111031-rtc-pdf/Executive_Summary.pdf.
[4] Massimo Calabresi, "Statistics Can Predict Criminal Risk. Can They Deliver Equal Justice?", *TIME*, July 31, 2014. http://time.com/3061893/holder-to-oppose-data-driven-sentencing/.
[5] Oscar Gandy, *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage* (Burlington, VT: Ashgate, 2009).
[6] Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *SSRN*, October 19, 2014, 8. http://ssrn.com/abstract=2477899.
[7] Frank Pasquale, *The Black Box Society* (Cambridge, MA: Harvard University Press, 2015), 48.
[8] The Sentencing Project, "Uneven Justice: State Rates of Incarceration By Race and Ethnicity," by Marc Mauer and Ryan S. King (Washington, D.C., 2007), 6. http://www.sentencingproject.org/doc/publications/rd_stateratesofincbyraceandethnicity.pdf.

[9] "Center on Sentencing and Corrections," Vera Institute of Justice, accessed September 9, 2014, http://www.vera.org/centers/center-sentencing-and-corrections?qt-centers_sentencing_and=8&qt-centers_center_on_sentencing_and=8#qt-centers_center_on_sentencing_and.

[10] Barocas and Selbst, "Big Data", 10.

[11] Carrie Johnson, "After Discrimination Finding, Jury's Out On Memphis Juvenile Courts," *NPR*, August 6, 2014. http://www.npr.org/2014/08/06/338086226/after-discrimination-finding-jurys-out-on-memphis-juvenile-courts.

[12] Ali Winston, "Privacy, Accuracy Concerns as License-Plate Readers Expand," *SFGate*, June 17, 2014. http://www.sfgate.com/crime/article/Privacy-accuracy-concerns-as-license-plate-5557429.php.

[13] "What is CompStat?", University of Maryland, accessed October 19, 2014. http://www.compstat.umd.edu/what_is_cs.php.

[14] Steven D. Levitt, "Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not," *Journal of Economic Perspectives* 18, 1 (Winter 2004): 163-190. http://pricetheory.uchicago.edu/levitt/Papers/LevittUnderstandingWhyCrime2004.pdf

[15] Leighton Walter Kille, "Hot spots policing and crime prevention strategies: Ongoing research," *Journalist's Resource,* July 11, 2013. http://journalistsresource.org/studies/government/criminal-justice/hot-spot-policing-strategies#.

[16] Zach Friend, "Predictive Policing: Using Technology to Reduce Crime," *FBI Law Enforcement Bulletin*, April 9, 2013. http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/2013/April/predictive-policing-using-technology-to-reduce-crime.

[17] Monte Reel, "Chronicle of a Death Foretold," *Harper's Magazine*, March, 2014. http://harpers.org/archive/2014/03/chronicle-of-a-death-foretold-2/.

[18] Whet Moser, "The Small Social Networks at the Heart of Chicago Violence," *Chicago Magazine*, December 9, 2013. http://www.chicagomag.com/city-life/December-2013/The-Small-Social-Networks-at-the-Heart-of-Chicago-Violence/.

[19] Matt Stroud, "The Minority Report: Chicago's New Police Computer Predicts Crimes, but Is It Racist?", *The Verge*, February 19, 2014. http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist.

[20] Sonja B. Starr, "Sentencing, by the Numbers," *The New York Times*, August 10, 2014. http://www.nytimes.com/2014/08/11/opinion/sentencing-by-the-numbers.html.

[21] Barocas and Selbst, "Big Data," 23.

[22] Anne Milgram, "To Minimize Injustice, Use Big Data," *Laura and John Arnold Foundation*, August 20, 2014. http://www.arnoldfoundation.org/news/minimize-injustice-use-big-data.

[23] Milgram, "Minimize Injustice."

[24] Kimberly Johnson, "Driving While Black; Carolina City Crafts Racial Awareness Policy Policy," *Al Jazeera America*, September 9, 2014. http://america.aljazeera.com/articles/2014/9/9/racial-profilingnorthcarolina.html; see also: North Carolina General Statute 114-10.01. http://www.ncga.state.nc.us/EnactedLegislation/Statutes/HTML/BySection/Chapter_114/GS_114-10.01.html.

[25] "Our Mission and Approach," Measures of Justice, accessed October 13, 2014. http://measuresforjustice.org/who-we-are/mission/.

[26] "The Measures," Measures of Justice, accessed October 13, 2014. http://measuresforjustice.org/know-more/the-measures.

[27] Organization for Security and Co-operation in Europe, "Expert Seminar on Leveraging Anti-Money Laundering Regimes to Combat Human Trafficking," (conference, Vienna, Austria, October 3-4, 2011). http://www.osce.org/secretariat/107665.

[28] Judge Herbert B. Nixon, Jr., "Human Trafficking and the Internet* (*and Other Technologies, too)," *The Judge's Journal* 52, 1. (2013). http://www.americanbar.org/publications/judges_journal/2013/winter/human_trafficking_and_internet_and_other_technologies_too.html.

[29] Mark Latonero, et al. "The Rise of Mobile and the Diffusion of Technology-Facilitated Tracking," *Technology and Human Trafficking.* (Los Angeles, CA: USC Annenberg Center on Communication, Leadership & Policy, 2012). https://technologyandtrafficking.usc.edu/files/2012/11/HumanTrafficking2012_Nov12.pdf.