Capstone Project Milestone Report

JUSTFAIR State – North Carolina

Authors: Preet Khowaja​, Dorothy Hou, Chuhan Zhou​, and Clarissa Aché​.

**Introduction**

*“The United States is home to 5 percent of the world’s population but 25 percent of its prisoners. We have more total prisoners than any other country in the world”.*

The U.S. has a longstanding history of racial bias in the criminal justice system, depicted by the disproportionate makeup of African American prisoners as opposed to the general populace[[1]](#footnote-2). Black Americans comprise only 13% of the U.S. population but constitute 40% of those incarcerated[[2]](#footnote-3). As stated in the Vera Institute’s 2018 report, *An Unjust Burden: The Disparate Treatment of Black Americans in the Criminal Justice System*[[3]](#footnote-4), the deep rootedness of this systemic bias makes it a complex and challenging problem to solve, but one without which true justice cannot be served. Investigating this bias requires a robust data-oriented approach based on legal sentencing guidelines and historical sentencing records. According to the U.S. Constitution, citizens have a right to access criminal trial proceedings. However, justice proceedings’ records are difficult to obtain in a way that court decisions can be analyzed, therefore in practice, the right is rarely granted. In other words, “the public does not have access to high quality, large-scale information about the federal criminal justice system, and therefore, much of what happens in criminal courts remains opaque”[[4]](#footnote-5). While sentencing guidelines exist to inform decision-makers of the corresponding penalties for every processed trial, among other rules that govern the court system processes in NC, prosecutors and, ultimately, appointed judges hold discretion on each case’s outcome, allowing personal and systemic biases to obstruct justice.

There have been many recent efforts investigating bias in criminal sentencing. One source that has amalgamated the results of multiple studies over the years is that of The Sentencing Project[[5]](#footnote-6). They published a report that chronicles the racial disparity in each stage of the U.S. criminal justice system and proposed some underlying causes throughout the system[[6]](#footnote-7). In addition, the research conducted by Thompson[[7]](#footnote-8) examines the racialized history of the American criminal justice system, and the regularly discriminatory application of the law as well as the consistent lack of equal justice under the law over time. The study thus offers a historical and analytical context for understanding racial disparities in the U.S. legal system. However, previous work is either outdated or does not take a scientific approach to address this problem, which leads to the initiative of this project.

JUSTFAIR is an initiative presented by the Institute for the Quantitative Study of Inclusion, Diversity, and Equity (QSIDE) that aims to bring transparency to the criminal justice system in the United States by making archived court records publicly available in an aggregated format to facilitate analysis of the decisions made in courts of all levels. This initiative is a leader in investigating biases in the justice system using data-driven approaches. So far, JUSTFAIR has built a database of criminal sentencing decisions and analyzed sentencing disparity at a federal level. This project we describe in this report is a subsequent effort of the JUSTFAIR initiative to expand JUSTFAIR to the state-level scope. Therefore, this project will have the same objective as JUSTFAIR Federal but will focus on the level of state courts, specifically criminal sentences in North Carolina, using more recent data.

The first output of this project will be, to the extent allows by the privacy of the information obtained, the raw or summarized data that links information about defendants and their demographic characteristics with information about their federal crimes, sentences, and the identity of the sentencing judge for the state of North Carolina. We intend to construct this database to the standards of the larger JUSTFAIR initiative. The publication of the database or the summarized data will allow for independent research of the criminal justice system of North Carolina. The possibility of scrutiny can promote accountability to elected and appointed public officials participating in the system and bring legitimacy to the claims we make about potential systemic bias through reproducibility.

The second deliverable of this project is to analyze whether the decisions made in the criminal court system of North Carolina are biased against groups of people of different demographics, including race. This analysis entails separating specific court decision outcomes such as adjudications of sentences, bail bonds, court fees, or any type of punishment assigned to defendants to the recorded factors that lead to these decisions, to then, compares these decisions on aggregated scales using data science tools and finally answer whether people of different demographics are prosecuted and adjudicated the same way, that is, whether justice is being served for all people equally.

Finally, our deliverables would serve as a template project for others to replicate this analysis in a different state. QSIDE, the organization behind the JUSTFAIR project, intends to expand the initiative to all 50 states and all levels of courts. Our research can not only help bring more transparency to the prosecution system in North Carolina but help other researchers complete similar work and multiply the impact.

This project intends to expose biases and other faults in the law enforcement system that put groups of people at a disadvantage compared to others. The outcome of this study is useful for both the residents and the appointed and elected officials of North Carolina. Residents can hold public officials accountable for the responsibility given. Additionally, public officials can help remediate potentially biased processes in the court system and become aware of their own potential unconscious biases to make better decisions and use of their power fairly.

**Methods**

**Data**

We obtained the primary data source from Automated Criminal/Infractions System (ACIS) Criminal and Infraction Statistical Extract[[8]](#footnote-9), which contains only criminal and infraction case information as recorded by the clerks of the superior court in the 100 clerks’ offices in North Carolina.

The dataset consists of a 5-year direct extract from the ACIS system of 14,081,546 cases that are matched to 23,039,404 individual offences, in total. The creation date of cases can go back as far as the 1970s, however, the extract contains every case that has been created or updated in the years between January 1st of 2017 and December 1st of 2021. Given cases do not get usually updated after a year of their creation, most of the cases in this extract have been created in this five-year range.

Chart, pie chart

Description automatically generated

***Figure 1.*** *In the Courts 5-year system extract, 91.5% of cases are managed at the district court level. Only 8.5% of cases were handled by superior courts.*

Most of the cases in the ACIS Court System extract are handled by district courts compared to superior courts (**Figure 1**), that means most cases committed in the five-year range of the dataset are of relative low severity. In NC, both district and superior courts are trial courts, meaning they hold trials and gather juries to determine outcomes of cases. District courts, however, handle serious, but lower-level crimes, whereas Superior courts hear criminal cases involving felony crimes[[9]](#footnote-10).

Each case in the dataset contains information about the person who committed the crime, the court processes the case goes through until it's closed, the offenses charged, and outcomes of each individual offense prosecution. The completeness of different features of a case and offense in the dataset also explains the processes some types of cases go through and which they do not. **Figure 2** shows demographic information about the defendant is complete in most cases, however, case specific features such as the Bond Amount assigned, is missing in more than half of the cases. We attribute the missing information to the fact that not all cases are severe enough for the defendant to have to pay a bail bond to remain free before conviction. For offenses, the intuition is similar. Depending on the disposition of the initially charged offense, the information available varies. A more detailed discussion on the missing offense information is shown in the Results section.

A picture containing background pattern

Description automatically generated

Background pattern

Description automatically generated

***Figure 2.*** *Patterns in missing case and offense information may explain different prosecution routes in the criminal justice system.*

Returning to the Bail Bonds example from the last paragraph, the amount set as bail for offenses committed offers an opportunity to visualize and quantify potential racial and gender bias that may exist in the system. Bail bonds are not always a dollar amount, but in most records found in the dataset where the bond is the amount you must pay to be released from jail until your criminal case is resolved, an arbitrary amount is set by a court magistrate that is proportional to the likelihood of the defendant of not returning to court for their trial appointments. In North Carolina, between 2017-2021, the offenses for which you would most frequently get a bond amount assigned are shown in **Figure 3**. “Driving while Impaired” is the most common offence for which a defendant may get a dollar amount bond.

Chart, histogram

Description automatically generated

***Figure 3****. Frequency of most common offense types that are assigned a bail bond amount (that is greater or equal to $0). ‘Driving while impaired’ is the most common offense that gets bail assigned.*

Understanding the frequency of offenses that are assigned bail amount is important to be able to draw meaningful comparisons in bail amounts that were set by magistrates for different people. **Figure 4** shows the distribution of bail bond amounts assigned for a specific type of offense, “Driving while impaired”, for defendants of different races and genders. Given ~45% of the bail amounts for the offences of type "driving while impaired" do not have a bail amount determined (are missing in the data, not $0) and we do not have confirmation of why these values are missing, we cannot draw many conclusions from this visualization. However, if the missing values were intentionally not determined for conscious process-specific reasons, the distributions of the bail amounts could be hinting that the differences in these distributions are very subtle, meaning the bail amount variation may not be explained to a large degree by the race and gender of the defendant. A regression model that includes the effect of all other measurable conditions in the data and is able to explain better the variations in the dollar amounts set for bail in different instances of the same crime for different individuals could explain with certainty whether the differences could ever be attributed to racial-gender bias. However, this analysis is scheduled as a future step of this project.

Chart, box and whisker chart

Description automatically generated

***Figure 4****. Comparison of distributions of bail bond amounts set for offenses categorized as ‘Driving while impaired’ for different combinations of sex and gender of defendants. The median values are all at $1000, however, the distributions have subtle differences for each defendant group based on gender-race combinations”*

There is a large variety of court decision outcomes that we can analyze to find evidence of racial and gender bias. However, criminal sentences are the formal legal consequences to a conviction, that means, sentences are the means through which the court imposes justice for all types of confessed or proved wrongdoings. A sentence is also the court decision with largest impact for any individual that flows through the criminal justice system, and where if in any way the decisions made in court are being more disadvantageous for some groups of people, it would be important to make it visible before it continues to be deleterious for the community the court serves.

**The Punishment Charts**

To understand the study methods, it is important to first describe the outputs we are interested in. The length of criminal sentences, which is the main variable that is adjusted to the severity of the crime convicted, and thus, a potential source of disparity among convictions of different groups of people. However, the court decision process is long and fraught with many points of decision-making that reflect disparity. In addition to sentence length, we are also interested in bail amounts, fine amounts and other types of restitution paid by defendants. First, we look at the length of sentences.

In theory, in North Carolina, the criminal sentence length of any crime is informed by the Structured Sentencing Law, which aims to bring consistency to punishments under the legal system. This law contains two punishment charts (updated last in 2013), in which, a sentence minimum length is determined by 3 factors:

1. The type of crime (i.e., whether it is a felony or a misdemeanor)
2. The class of the crime, which is a classification of standard misconducts
3. The prior-record points, which apprise the frequency and severity of previously committed crimes of a defendant (Fig 5, Fig 6).

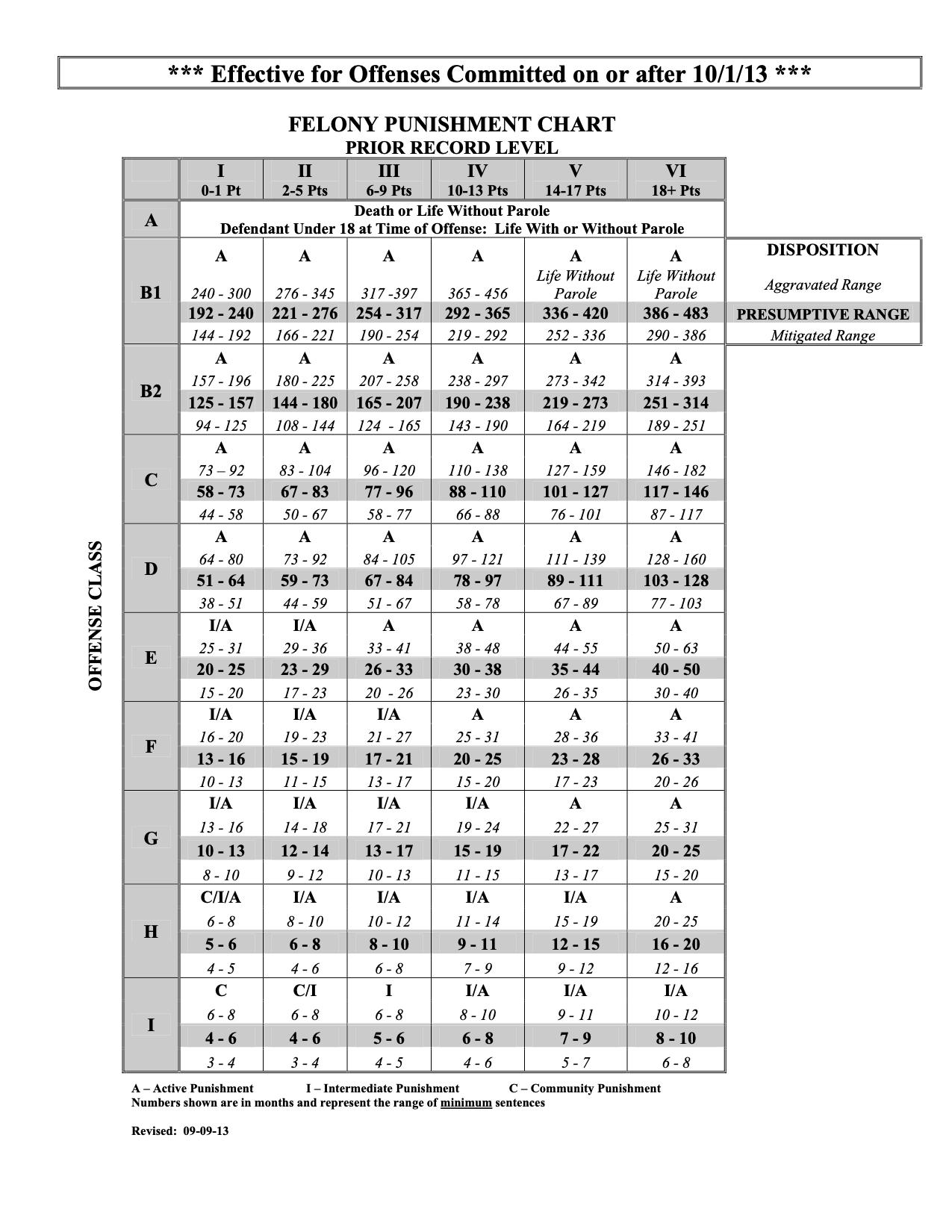
The grids also provide a recommendation for the maximum sentence for each minimum sentence length selected.

The three ranges in each sentencing bracket reflect the variability in an individual’s circumstances, which would be to the discretion of the decision maker to select for each conviction. The ranges, in order of severity, from smallest to largest, are referred to as mitigated, presumptive and aggravated. For example, a mitigated sentence may be selected for a low-income individual stealing to provide for themself.

Table

Description automatically generated

***Figure 5****. Misdemeanor Punishment Chart and Minimum/Maximum Table for Offenses Committed. Last update to this document was made in September 20th, 2013.*



***Figure 6****. Felony Punishment Chart and Minimum/Maximum Table for Offenses Committed on or after October 1, 2013*

**Comparing sentence severity by race and gender groups**

A reasonable expectation in an unbiased system is that criminal sentence length variation should not be impacted by race, gender or a combination of the two. We intend to run statistical tests to determine whether the distributions of criminal sentences lengths for race and gender groups of defendants are significantly (statistically) different for comparable crimes, within the same court level. The outcomes of this test can be informative in answering whether black defendants, for example, are given more severe sentences disproportionately compared to defendants in other racial and gender groups. The gender-racial groups we will compare include: white men, white women, black men, black women, Hispanic men, Hispanic women. Although these groups do not represent the complete spectrum of diversity that exists in the community the North Carolina Courts serve, the demographic information reported in the ACIS data and race diversity found in most NC districts constrain our analysis to these 6 groups. The court level of study will be the district court given a larger number of cases found at this level.

To accomplish this, we will build regression models and include racial and gender information as factors. In addition, we will filter the data to only include cases where crimes are represented with enough observations, so that the conclusions we draw will be more reliable since they are based on a sufficient sample size.

There are two expected outcomes of this part of the analysis: First is the answer to whether the courts of North Carolina are following (and to which extent) the sentencing guidelines proposed in 2013. Second, using the sentencing guidelines are a measure of severity the courts themselves have proposed for their own decision-making, we will be able to answer whether the sentences adjudicated for different racial-gender groups of convicted individuals who committed comparable crimes are or not, in a measurable way, different.

The next step of this analysis will be to attribute the differences to individual district courts and find potential explanations to the disparities based on the particular instance of the judicial system where they are happening.

**Comparing initial vs. convicted charges**

In addition to the variation within each Sentencing Chart “cell”, we also look at the difference between the initial and final sentences for members of different racial-gender groups. Public prosecutors often negotiate with defendants and offer them a conviction for some or fewer crimes than they have initially been charged with the condition of having them pled guilty for the lesser charges, this process is called plea bargaining. Plea bargains, currently, ensure that approximately >90% of criminal cases in North Carolina are not deliberated in courts[[10]](#footnote-11), where bias can play out in a subtle and nuanced manner. We plan to investigate whether the difference in the initial and final charge is constant by race and gender of the defendant.

**Results**

**Punishment Grid Analysis**

Prior to analyzing any results, it is essential to know that there exists a steep funnel of the data that is available to analyze as **Figure 7** shows. In spite of the fact that all the cases have initial charged offenses available, only 24.4% have convicted offense codes. Moreover, there are only about 5% of cases which have sentencing and prior records information available for us to compare them with the punishment guidelines. Therefore, the following analysis on punishment guidelines are limited to the scope of data with sentencing and prior records information available.



***Figure 7.*** *Funnel of case data availability from initial charges to final sentencing*

As the sentencing length suggested in the punishment guidelines are determined by offense class and prior record points, we first examined if the cases with all the information available are generally following the guidelines. After matching the offense class and prior record points with the punishment grids and comparing the given sentencing length and suggested length in the grids, only 56.3% fell within the suggested sentencing range. The remaining cases are either under-sentenced or over-sentenced compared to the punishment grids. To obtain a bigger picture of racial disparities presented in the data, we conducted a preliminary analysis on cases which fell outside the punishment guidelines. Among those cases which are over-sentenced, i.e., have sentencing length greater than the maximum of aggravated range, 50.6% are black defendants while 44.9% are white defendants. This percentage of black defendants is disproportionately higher compared to the distribution of racial groups in the entire offense data, which consists of 49.9% white defendants and 43.4% black defendants. This result indicates a sign of racial disparity from the perspective of over-sentencing. For the cases that are under-sentenced, i.e., have sentencing length less than the minimum of mitigated range, 47.8% are white defendants and 46.0% are black defendants, which shows that no clear evidence of racial disparity is observed in under-sentenced cases.

**Next Steps**

Based on our analysis so far, we have the following plans for the Spring semester:

Since there is currently only a very small portion of data (~5%) that have complete information for us to analyze the NC punishment guidelines, we plan to further dig into reasons for the missing values in the data and validate our assumptions to determine if we can reasonably impute some missing values and therefore utilize more data for our analysis. In addition, we can deepen our analysis on punishment grids by examining which range do people fall in the grid by race and identify if there exists a discrepancy for different races.

Furthermore, as stated in the Methods section above, we will build a regression model which includes race, gender and all potential control variables as predictors and sentencing length as the response variable. After controlling for the confounding variables, we can see if race and gender are statistically significant in predicting sentencing length. Moreover, we would like to further try to attribute an increase or decrease of sentencing length to a specific court or judge in order to identify where in the judicial system do the disparities occur.

**References:**

1. Hinton E, Henderson L & Reed C (2018), An Unjust Burden: The Disparate Treatment of Black Americans in the Criminal Justice System. https://www.vera.org/publications/for-the-record-unjust-burden

2. Gilbert C (2022), Beneath the Statistics: The Structural and Systemic Causes of Our Wrongful Conviction Problem. https://www.georgiainnocenceproject.org/2022/02/01/beneath-the-statistics-the-structural-and-systemic-causes-of-our-wrongful-conviction-problem/

3. Ciocanel M-V, Topaz CM, Santorella R, Sen S, Smith CM, Hufstetler A (2020) JUSTFAIR: Judicial System Transparency through Federal Archive Inferred Records. PLoS ONE 15(10): e0241381. <https://doi.org/10.1371/journal.pone.0241381>

4. The Sentencing Project, “Racial Disparity in Sentencing”, January 2005. <https://www.opensocietyfoundations.org/publications/racial-disparity-sentencing>

5. Thompson, H (2019). THE RACIAL HISTORY OF CRIMINAL JUSTICE IN AMERICA. Du Bois Review: Social Science Research on Race, 16(1), 221-241. doi:10.1017/S1742058X19000183

6. Sentencing and Policy Advisory Commission, Punishment Grids. <https://www.nccourts.gov/documents/publications/punishment-grids>

7. North Carolina Judicial Branch, RPA Criminal Extracts. https://www.nccourts.gov/services/remote-public-access-program/rpa-criminal-extracts

8. North Carolina Judicial Branch, Type of Courts. <https://www.nccourts.gov/learn/types-of-courts>

9. Devers, L (2011). Plea and charge bargaining - Bureau of Justice Assistance. https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/PleaBargainingResearchSummary.pdf

10. The Sentencing Project (2018). Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System. https://www.sentencingproject.org/publications/un-report-on-racial-disparities/

1. Hinton E, Henderson L & Reed C (2018), An Unjust Burden: The Disparate Treatment of Black Americans in the Criminal Justice System. https://www.vera.org/publications/for-the-record-unjust-burden [↑](#footnote-ref-2)
2. Gilbert C (2022), Beneath the Statistics: The Structural and Systemic Causes of Our Wrongful Conviction Problem. https://www.georgiainnocenceproject.org/2022/02/01/beneath-the-statistics-the-structural-and-systemic-causes-of-our-wrongful-conviction-problem/ [↑](#footnote-ref-3)
3. Hinton E, Henderson L & Reed C (2018), An Unjust Burden: The Disparate Treatment of Black Americans in the Criminal Justice System. https://www.vera.org/publications/for-the-record-unjust-burden [↑](#footnote-ref-4)
4. Ciocanel M-V, Topaz CM, Santorella R, Sen S, Smith CM, Hufstetler A (2020) JUSTFAIR: Judicial System Transparency through Federal Archive Inferred Records. PLoS ONE 15(10): e0241381. https://doi.org/10.1371/journal.pone.0241381 [↑](#footnote-ref-5)
5. The Sentencing Project, “Racial Disparity in Sentencing”, January 2005. https://www.opensocietyfoundations.org/publications/racial-disparity-sentencing [↑](#footnote-ref-6)
6. The Sentencing Project (2018). Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System. https://www.sentencingproject.org/publications/un-report-on-racial-disparities/ [↑](#footnote-ref-7)
7. Thompson, H. (2019). THE RACIAL HISTORY OF CRIMINAL JUSTICE IN AMERICA. *Du Bois Review: Social Science Research on Race,* *16*(1), 221-241. doi:10.1017/S1742058X19000183 [↑](#footnote-ref-8)
8. North Carolina Judicial Branch, RPA Criminal Extracts. https://www.nccourts.gov/services/remote-public-access-program/rpa-criminal-extracts [↑](#footnote-ref-9)
9. North Carolina Judicial Branch, Type of Courts. https://www.nccourts.gov/learn/types-of-courts [↑](#footnote-ref-10)
10. Devers, L (2011). Plea and charge bargaining - Bureau of Justice Assistance. Retrieved from https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/PleaBargainingResearchSummary.pdf [↑](#footnote-ref-11)