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, this information has been difficult to obtain, and 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 inform decision-makers of the corresponding penalties for every processed trial, appointed judges still hold discretion on each case’s outcome, allowing personal 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-8). Similar research offers a historical and analytical context for understanding racial disparities in the American criminal justice system[[7]](#footnote-10). However, previous work is either outdated or does not take a data science approach to address this problem.

JUSTFAIR is an initiative presented by QSIDE 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. 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 is a subsequent effort of the JUSTFAIR initiative to expand the project to the state-level scope. Therefore, the project described in this document will have the same objective 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 the database 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 construct this database to the standards of the larger JUSTFAIR initiative. The publication of the database 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.

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. We will answer whether people of different demographics are prosecuted 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 behaviors and make better decisions and use of their power.

**Methods**

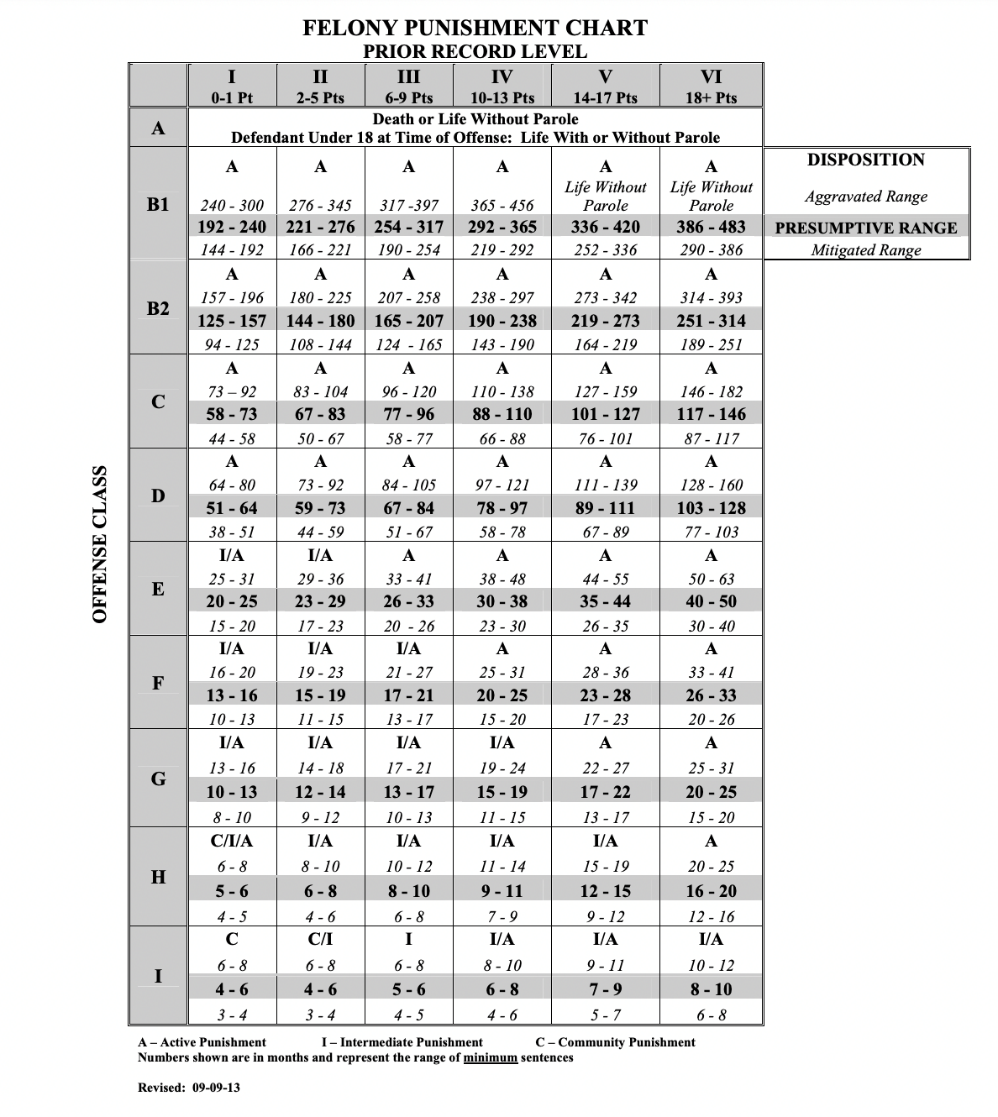
**Data**

We obtained the primary data source from Automated Criminal/Infractions System (ACIS) Criminal and Infraction Statistical Extract[[8]](#footnote-11), which contains criminal and infraction case information as recorded by the clerks of the superior court in the 100 clerks’ offices in North Carolina. The ACIS statistical extract represents case activity for the latest five-year period. Thus, our project focuses on criminal records data from the year 2017 to 2021. For data preprocessing, we will merge data tables representing different dimensions using ‘Case Root Record Key’ as a unique case identifier to obtain all the variables we need for investigation. Given the complexity of the criminal justice system, it is critical to understand each available field presented in the ACIS data and identify variables necessary to incorporate in the analysis, using JUSTFAIR Federal as a reference and general guideline.

In order to investigate bias in criminal sentencing, we use a two-step framework as follows:

1. Comparing sentence length by demographic
2. Regression analysis to investigate the factors that influence the sentence an individual receives

To understand the study methods, it is important to first describe the output we are interested in: the length of criminal sentences. In theory, the criminal sentence length is decided by the type of felony an individual commits and their historical criminal record (Fig 1). Ranges in sentencing reflect the variability in an individual’s circumstances—for example, a mitigated sentence for a low-income individual stealing to provide for themself.



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

**Comparing sentence length by demographic**

A reasonable expectation in a just world is that criminal sentence length variation should not be impacted by race, gender or a combination of the two. Looking at the distributions of race and gender within offense class categories can be informative in answering whether black defendants are given the aggravated ranges more often than other defendants. This analysis is also broken up by gender to compare black women to white women, and black men to white men.

In addition to within cell variation, we also look at the difference between the initial and final sentences for members of different racial groups. Public prosecutors often decide to negotiate with defendants and prosecute them for some or fewer crimes than they have initially been charged with in a system called plea bargaining. Plea bargains make up over 90% of criminal cases in North Carolina, where bias can play out in a subtle and nuanced manner. We investigate whether the difference in the initial and final charge is constant by race.

**Regression Analysis**

In the second step of this study, we are interested in modelling the sentence that an individual is likely to receive based on factors such as race, gender, history of crime, and type of felony. This is a complicated model since the target variable does not lie on a linear axis, but takes the shape of a matrix. We have yet to finalize the details of this model, but we will use either a linear or logistic regression since these models have high interpretability. We are interested in the R-square and statistical significance of the coefficients in our regression. We hope to see a significant impact of race on the length of an individual’s sentence.

(We acknowledge the complexity of the legal system and know this is a difficult effect to measure. We have recently obtained a sample of our data and are still making the effort to obtain the whole dataset. We are working with our client to follow a modelling framework like what has been done at the federal level, and the details will take a little longer to hash out. We are also working on submitting an IRB to obtain the full dataset.)

**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

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. https://www.sentencingproject.org/publications/un-report-on-racial-disparities/ [↑](#footnote-ref-8)
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-10)
8. https://www.nccourts.gov/services/remote-public-access-program/rpa-criminal-extracts [↑](#footnote-ref-11)