Olivia Farrell & Ashu Tayal

PPHA 30536

Final Project Write-Up

**Project Title**: Exploring Racial Biases in CCSAO Sentencing

**Data source**: Cook County State's Attorney Office's (CCASO) Open Data Portal

Initiation and Dispositions Data for 2015-2019

**Link to Folder**: [Google Drive](https://drive.google.com/drive/folders/1HNWlXUyzuT6yzukMdnMnKdV6jekArBZ2?usp=sharing)

**Source URL**: [Cook County Data Catalog](https://datacatalog.cookcountyil.gov/browse?tags=state%27s+attorney+case-level&sortBy=most_accessed)

**Questions**:

1. How does race impact charges over time?

2. How likely is the initial charge to be changed depending on race?

**Overview**

For our final project, we set out to explore racial biases in Cook County State’s Attorney’s Office (CCSAO) practices in charging individuals at disposition. Specifically, we wanted to answer if race impact charges over time and how likely is the initial charge to be changed depending on race. In order to do so, we have pulled initiation and disposition data for the CCSAO for the past five years. We then visually explored the relationship between race and the verdicts over time, the relationship between race and the fraction of charges changed from initiation to disposition, and the geographical relationship between the level of charge. Additionally, we conducted multiple regressions exploring these relationships.

**Data**

The first step of this project was to obtain the necessary data from the CCSAO. This was done by utilizing their data portal[[1]](#footnote-1). The datasets span multiple decades and cover every felony case processed by the CCSAO during those years. Needless to say, the files were quite large. The merged file had over 9 million rows. When we first attempted to add these files to our repository, the size of these files proved to be the first roadblock in our analysis. GitHub has an 100 Mb file size limit. To continue with our analysis, we had to download and utilize GitHub Large File Storage[[2]](#footnote-2). This process forced us to learn more about the working of GitHub and how to utilize Git Bash. After several hours, we successfully uploaded our files, but the size of the data would continue to be a burden in our analysis through increasing the time for code to run and limiting our ability to visually inspect the data.

**Cleaning and Merging**

Once the data was uploaded, we merged our files. As the CCSAO data is case-level data, we were able to merge based on case IDs and other consistent variables between the datasets. Doing so gave us one file that captured the status of the cases and charges against an individual at initiation and disposition. This data captures the experience of individual immediately following arrest and just prior to sentencing. Following the merge, we began to clean it to be ready for analysis and visualization.

**Choice of variables**

The size of the data meant that we had to be selective in what data we chose to clean. A full analysis, beyond the scope of this project, would likely require coding and grouping thousands of different unique data points, such as the unique charges. For our purposes, we ensured consistency in naming and grouped variables of interest. Specifically, this meant fixing terms for gender, race, and charge offense title. Grouping the charge offense title required significant research on the meaning of the numerous legal terms and abbreviations to be able to group them into simple categories for analysis of guilty, not guilty, dropped, and other/unknown.

Once the data had been cleaned, we began by conducting a simple analysis of the data by determining basic summary statistics to have a better understanding of the data. These are shown in the appendix. The given counts of races and their relative percentages led to our decision to focus on only white, Black, and Latinx individuals for our analysis as these three groups account for more than 95 percent of cases.

**Analysis and Findings**

We then graphed the verdicts over time by race (Graphs 1-3 in the appendix). Black individuals had much higher counts of all three verdicts. For both guilty and not guilty, the count of Black individuals has gone down over the years, while the number has been generally trending upward for dropped cases. For Latinx and white individuals the counts have been more stagnant, with a slight downward trend. There are several possible reasons for these trends. We might be seeing a reflection of the over policing of Black individuals resulting in higher arrests of invalid charges and the increasing amount of dropped cases.

To further explore this relationship between verdict and race, we used bar plots to visually display the counts of guilty, not guilty, and dropped cases by race for each year. These bar graphs were similar for each of the years. There seemed to be a higher proportion of dropped cases to guilty and not guilty for both Black and Latinx populations compared to whites. This could possibly reflect the same issue as the line graphs, that these populations are over policed and often arrested on trumped up charges which are later dropped.

Next, we wanted to determine if race played a role in the decision for a charge to be changed from initiation to disposition. We chose to only focus on those with a guilty verdict for this analysis, as if the case was dropped or if the individual was declared not guilty, it is of less relevance if the charges were altered. This analysis does not capture whether or not the charges were upgraded or downgraded. In order to make this determination, each individual charge would have to be coded with the multitude of potential upgrades and downgrades for that charge. Given there are over 2,300 charges, each with numerous charges they could be changed to, this would mean accounting for easily hundreds of thousands of possible combinations. We assume, however, for the purpose of this project, that we are most likely seeing charges downgraded, as this happens at a much higher frequency. What we see for all three races, is a general trend of fewer charges being changed over time, with an increase from 2018 to 2019, especially for whites. With a view towards ending mass incarceration, the general thought would be a desire for more downgraded charges. Kim Foxx, the current state’s attorney has pushed for this viewpoint. She was elected in 2016 but may have pushed additional initiatives in 2019 that led to this increase of charges being downgraded. Research into such initiatives, would provide more context for this data.

Another approach to analyzing the role of race is by looking at where the cases initiate. In order to do this, we utilized geopandas to plot the incident city. To allow for an interactive plot, this was done in Jupyter. We created a heat map of the counts of each felony per city, mapped onto the municipality borders for Cook County[[3]](#footnote-3). This was done for each felony level as defined in the data dictionary[[4]](#footnote-4). In decreasing order of severity these are First degree murder (as a separate class of felony Class M), Class X felonies, Class 1 felonies, Class 2 felonies, Class 3 felonies, Class 4 felonies. Additionally, some misdemeanors are included in the data if a charge was pled down or coincided with a felony charge. In decreasing order of severity these are captured by Class A, Class B, and Class C. The results maps could be compared to a visual displaying of demographic data to begin to understand patterns in policing and how the racial demographics of the cities are reflected in the resulting cases.

After developing our visualizations, we were interested in formalizing our analyses through regressions. We started by filtering the necessary data and creating the necessary dummy variables. We then utilized Statsmodels to run the following regressions:

Equation 1:

Equation 2:

Verdict = whether or not the individual was found guilty of any charges

Change in charge = whether or not the charge was changed from initiation to disposition

Black = dummy variable for whether or not an individual is Black

Latinx = dummy variable for whether or not an individual is Latinx

Gender = variable capturing all gender options

Age = the age of the induvial when the alleged crime occurred

Court = the court district where the case took place

Year = year in which the incident took place

The first regression gave a coefficient of -0.0831 for Black and 0.0078 for Latinx, with significant t-values. This indicates that Black individuals are 8.31% less likely to be found guilty than their white counterparts, while Latinx individuals are just slightly (<1%) more likely to be found guilty. This analysis, however, did not take into consideration the actual crime the individual was charged with. In order to extend further meaning to these results, additional analysis would be needed.

The second regression gave a coefficient of -0.0404 for Black and -0.0021 for Latinx, however the t-statistic for the Latinx value was not significant. These values indicate that Black individuals are 4.04% less likely to have their charge altered from initiation to disposition, while Latinx individuals do not have a statistically significant difference in changing of charges when compared to their white counterparts. As mentioned above, this does not take into consideration as to whether these charges or upgraded or downgraded.

We ended our project by reviewing our code and revising it to incorporate additional functions and best practices. There are many limitations of our analysis, that have been highlighted throughout this write-up. To truly answer our research questions, individual charges would need to be coded. Additionally, demographics data for the population and arrest data would be needed to provide additional context to the trends seen in the CCSAO data.

**Appendix: Graphs and Tables**

Summary Statistics

|  |  |  |  |
| --- | --- | --- | --- |
| 2015 | Race | Count | Percentage |
| 0 | American Indian | 784 | 0.001292575 |
| 1 | Asian | 8154 | 0.013443444 |
| 2 | Biracial | 20 | 3.30E-05 |
| 3 | Black | 387052 | 0.638129986 |
| 4 | Latinx | 143466 | 0.236531413 |
| 5 | Unknown | 6626 | 0.010924241 |
| 6 | White | 60439 | 0.099645366 |

|  |  |  |  |
| --- | --- | --- | --- |
| 2016 | Race | Count | Percentage |
| 0 | American Indian | 118 | 0.00014218 |
| 1 | Asian | 7147 | 0.008611539 |
| 2 | Biracial | 27 | 3.25E-05 |
| 3 | Black | 562742 | 0.678057144 |
| 4 | Latinx | 158676 | 0.191191337 |
| 5 | Unknown | 6107 | 0.007358425 |
| 6 | White | 95116 | 0.114606842 |

|  |  |  |  |
| --- | --- | --- | --- |
| 2017 | Race | Count | Percentage |
| 0 | American Indian | 251 | 0.0002525 |
| 1 | Asian | 20989 | 0.0211125 |
| 2 | Biracial | 15 | 1.51E-05 |
| 3 | Black | 681437 | 0.6854462 |
| 4 | Latinx | 163865 | 0.1648291 |
| 5 | Unknown | 3066 | 0.003084 |
| 6 | White | 124528 | 0.1252606 |

|  |  |  |  |
| --- | --- | --- | --- |
| 2018 | Race | Count | Percentage |
| 0 | American Indian | 1015 | 0.001089 |
| 1 | Asian | 6608 | 0.00709 |
| 2 | Biracial | 567 | 0.0006084 |
| 3 | Black | 643173 | 0.6900872 |
| 4 | Latinx | 160276 | 0.1719668 |
| 5 | Unknown | 2172 | 0.0023304 |
| 6 | White | 118206 | 0.1268282 |

|  |  |  |  |
| --- | --- | --- | --- |
| 2019 | Race | Count | Percentage |
| 0 | American Indian | 4 | 5.63E-06 |
| 1 | Asian | 3013 | 0.004237306 |
| 2 | Biracial | 37 | 5.20E-05 |
| 3 | Black | 520990 | 0.732689698 |
| 4 | Latinx | 101209 | 0.142334386 |
| 5 | Unknown | 1211 | 0.001703079 |
| 6 | White | 84601 | 0.118977871 |

|  |  |  |  |
| --- | --- | --- | --- |
| 2015 | Verdict | Count | Percentage |
| 0 | Dropped | 446537 | 0.736202499 |
| 1 | Guilty | 104511 | 0.172306571 |
| 2 | Not Guilty | 55476 | 0.091462902 |
| 3 | Undetermined | 17 | 2.80E-05 |

|  |  |  |  |
| --- | --- | --- | --- |
| 2016 | Verdict | Count | Percentage |
| 0 | Dropped | 631272 | 0.76063007 |
| 1 | Guilty | 114720 | 0.13822803 |
| 2 | Not Guilty | 83937 | 0.10113708 |
| 3 | Undetermined | 4 | 4.82E-06 |

|  |  |  |  |
| --- | --- | --- | --- |
| 2017 | Verdict | Count | Percentage |
| 0 | Dropped | 812239 | 0.81701774 |
| 1 | Guilty | 101903 | 0.10250254 |
| 2 | Not Guilty | 79987 | 0.0804576 |
| 3 | Undetermined | 22 | 2.21E-05 |

|  |  |  |  |
| --- | --- | --- | --- |
| 2018 | Verdict | Count | Percentage |
| 0 | Dropped | 801496 | 0.859963176 |
| 1 | Guilty | 95112 | 0.102050188 |
| 2 | Not Guilty | 35402 | 0.037984489 |
| 3 | Undetermined | 2 | 2.15E-06 |

|  |  |  |  |
| --- | --- | --- | --- |
| 2019 | Verdict | Count | Percentage |
| 0 | Dropped | 614434 | 0.86410384 |
| 1 | Guilty | 75206 | 0.1057653 |
| 2 | Not Guilty | 21425 | 0.03013086 |

Graph 1: Count of Guilty by Race Over Time

Chart, line chart

Description automatically generated

Graph 2: Count of Not Guilty by Race Over Time

Chart, line chart

Description automatically generated

Graph 3: Count of Dropped Cases by Race Over Time

Chart, line chart

Description automatically generated

Graph 4: 2015 Verdicts by Race

Chart, bar chart

Description automatically generated

Graph 5: 2016 Verdicts by Race

Chart, bar chart

Description automatically generated

Graph 6: 2017 Verdicts by Race

Chart

Description automatically generated

Graph 7: 2018 Verdicts by Race

Chart

Description automatically generated

Graph 8: 2019 Verdicts by Race

Chart, waterfall chart

Description automatically generated

Graph 9: Changes in Charge for Guilty Verdicts

Chart, line chart

Description automatically generated

Regression 1 Output:

A picture containing text, screenshot, receipt

Description automatically generated

Regression 2 Output:

A picture containing text, receipt, screenshot

Description automatically generated

1. CCSAO data portal: <https://datacatalog.cookcountyil.gov/browse?tags=state%27s+attorney+case-level&sortBy=most_accessed> [↑](#footnote-ref-1)
2. More information on this system and the process for installing and configuring the system can be found here: <https://docs.github.com/en/free-pro-team@latest/github/managing-large-files/configuring-git-large-file-storage> [↑](#footnote-ref-2)
3. Source: <https://hub-cookcountyil.opendata.arcgis.com/datasets/534226c6b1034985aca1e14a2eb234af_2?selectedAttribute=MUNICIPALITY> [↑](#footnote-ref-3)
4. Data dictionary: <https://www.cookcountystatesattorney.org/sites/default/files/files/documents/column_by_dataset_glossary_final_1.pdf> [↑](#footnote-ref-4)