Crime and Equity in Sentencing within the Cook County Justice System

Team 0 Final Project Report

Annie Cui, Deanna Emery, Ryan Wong
MIDS W200, Fall 2021

Team GitHub Repository:

https://github.com/UC-Berkeley-I-School/Project2_Cui_Emery_Wong

Overview and Data Set

A nation's justice system is intended to be blind, impartial, and fully equitable so that even criminals may seek justice in their punishment. However, recent social movements such as the Black Lives Matter movement have revealed that the US justice system may not be as equitable as it strives to be. Cook County, IL was a hotpot of racial justice protests, its borders including Chicago, IL and also purporting to be the second most-populated county nation-wide. In an effort towards transparency, Cook County offers its criminal sentencing data publicly available. Using the publicly available data, our team sought to investigate what the criminal environment looks like within the county, and also whether or not there may be evidence of unjust activity within a major justice system.

Our team analyzed the criminal sentencing data from Cook County, IL, which is available here: https://datacatalog.cookcountyil.gov/Courts/Sentencing/tg8v-tm6u/data

We also analyzed census data for Cook County as well, which is also available here: https://data.census.gov/cedsci/table?q=cook%20county%20IL&tid=ACSDP1Y2018.DP05

We also analyzed unemployment data from the U.S. Bureau of Labor Statistics for the Chicago, IL area, which is available here:

https://data.bls.gov/timeseries/LAUDV171697400000003?amp%253bdata_tool=XGtable&output_view=data&include_graphs=true

In order to understand the criminality landscape and to investigate the possibility of unfair sentencing, we chose to analyze the data to answer the following five research questions:

1. Are people of different age groups sentenced differently per crime?
2. Are people sentenced differently based on demographics?
3. Does unemployment lead to higher rates of crime?
4. Do people of different demographics spend differing amounts of time within the system?
5. What do the demographics of the criminal population look like compared to the demographics of the civilian population?

Data Cleaning

In order to maintain maximum data hygiene for our sentencing data set, the data had to be cleaned and modified in a number of ways. Below are the steps taken to clean the sentencing data in order to make it maximally usable. This cleaned data was then used for every research question.

- Converted all date columns to datetime data types.
- Only kept sentences where both the ARREST_DATE and RECEIVED_DATE are from 2011 and onward.
- Only kept sentences where the CURRENT_SENTENCE_FLAG is True. This is to remove any sentences that were reversed.
- Removed entries where sentences were in non-time values, such as Life, Dollar Value, or “Term”.
- Striped non-numeric characters from the sentence terms (COMMITMENT_TERM) and created the “COMMITMENT_TERM_DAYS” field to standardize sentence lengths to days.
- Created a severity ranking for the DISPOSITION_CHARGED_CLASS.
- Dropped sentences for criminals who were already issued a sentence with a PRIMARY_CHARGE_FLAG value of False, thus removing additional secondary charges.
- Deduplicated CASE_PARTICIPANT_ID, keeping only the most severe sentence for the individual.
- Removed any offense categories that have fewer than 100 counts so that the means and medians would be more reliable.
- Cleaned and standardized the race and gender columns

Data Assumptions

In order to draw meaningful conclusions from our analysis, we have to make a number of assumptions about the data. While we may not know to what extent these assumptions are correct, we consider these assumptions to be reasonable enough to be plausible. Regarding all of our data sets, the following assumptions have been made:

1. All data is accurate to the real world facts.
2. All data represents an accurate representation of the overall population being investigated, or encompasses the entire population in question (e.g. all crimes have been captured in the sentencing data).
3. There were no outstanding factors during the time period being investigated that would have influenced our data and findings. For example, the COVID-19 pandemic during 2020 and 2021 are assumed to have not had an outstanding impact on the data and outcomes.
4. All of the data cleaning steps done have not compromised or misconstrued the data.

Research Question and Analyses

Question 1: What do the demographics of the criminal population look like compared to the demographics of the civilian population?

Crime is a symptom of existing issues within a society, but does crime affect all of society equally? By extension, does the justice system apply to all demographics equally? Our analysis compared demographic 2019 data from the US Census for Cook County against the demographics of individuals being sentenced in Cook County's legal system during the same period of time (2011-2021) in an effort to see which groups are overrepresented or underrepresented within the Cook County legal system.

Additional Data Cleaning and Sanity Checks

- Removed duplicate entries for, "CASE_PARTICIPANT_ID" in order to remove duplicate individuals and avoid double counts.
- Dropped users without a stated Male or Female gender
- Converted "AGE_AT_INCIDENT" to Float in order to make it numerizable from strings.

Research Analysis

The 2019 census data was imported into a CSV file and then analyzed with Python along with the sentencing data. Both data sets were stored in appropriate Pandas data frames.

The first calculation was to find the size of both the sentencing data and the census data. For the sentencing data, this was done by finding the size of the Pandas data frame that the sentencing data was stored in, since each row of that cleaned data represents a single person. For the census data, the data already provided a value for the population size for their data. Using both population size values, we found the percentage of criminals in the civilian population by dividing the size of the sentencing data against the size of the civilian population. This data was compiled into a table which can be found in the Appendix as *Figure A.1.1*.

With the size for both population bases determined, we computed and compared the census and sentencing data across three main demographic focuses: gender, age group, and race.

For computing the gender comparisons, the census data already provided a percentage number for the % female population. The sentencing data required that this percentage be found manually, which was done by counting the number of sentencing entries that were marked as female and then finding what percentage those entries had compared to the total number of entries. This data was compiled into a table which can be found in the Appendix as *Figure A.1.2*.

For computing the age groups, the census data already provided a percentage number for each age group's population. The sentencing data similarly required that these percentages be found, which was done by counting the number of sentencing entries that were marked by a particular age, categorizing them into an appropriate age group, and then computing the percentage those entries had compared to the total number of entries. In order to match the census data, the sentencing data was organized into the following age ranges: 20-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75-84, and 85-100. This data was compiled into two visualizations, a table as *Figure A.1.3* and a graph as *Figure A.1.4.*

For computing race groups, the census data already provided a percentage number for each race's population. The sentencing data similarly required that these percentages be found, which was done by counting the number of sentencing entries that were marked by a particular race, categorizing them into an appropriate racial group, and then computing the percentage those entries had compared to the total number of entries. In order to match the census data, the sentencing data was organized into the following age ranges: American Indian, Asian, Black, Hispanic, White, and Biracial. This data was compiled into a table which can be found in the Appendix as This data was compiled into two visualizations, a table as *Figure A.1.5* and a graph as *Figure A.1.6.*

Research Findings

Our research discovered a number of substantial differences between the census data and the sentencing data. This suggests that the general demographic makeup of the criminal population is drastically different than that of the general civilian population.

The most dramatic difference is in the overall population number, with *Figure A.1.1* showing that the number of criminals is approximately only 2.43% the number of Cook County's general population. This suggests that criminality is still a relatively rare occurrence in the general population, as only 1 in 50 people have been convicted of a crime.

In comparing the gender differences, the sentencing data had a much lower % female ratio than the census data. As seen in *Figure 1.1*, sentencing data was only 12.36% female, compared to 51.4% female from the census data, a difference of 39.04%. This suggests that women are 3.15x less represented in the criminal population than the general population.

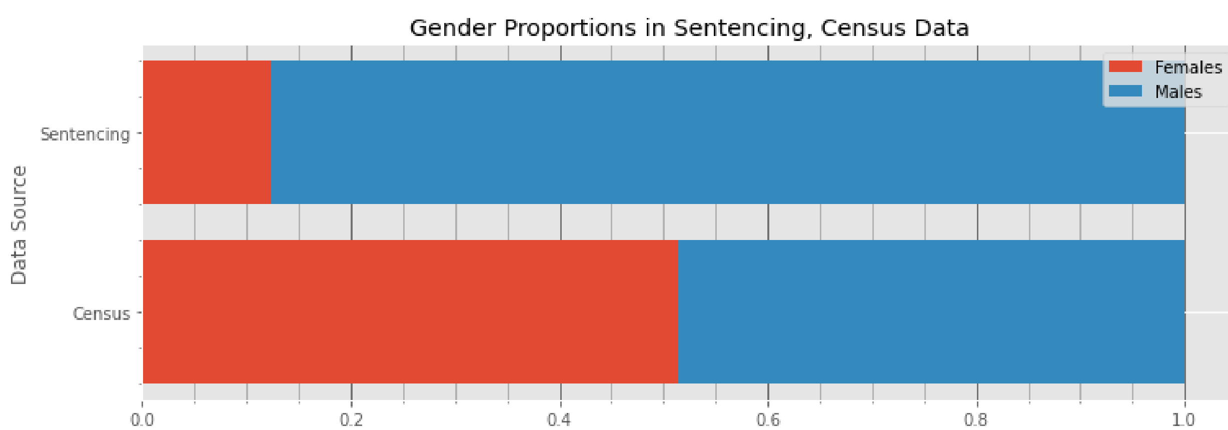

Figure 1.1: Gender Proportions in Sentencing, Census Data

In comparing the age group , the sentencing data showed a significant overrepresentation in younger people and a significant underrepresentation of older people. As seen in *Figure 1.2*, Individuals aged 20-24 were 57.82% over-represented compared to the general population, and individuals aged 25-34 were 12.95% over-represented. Meanwhile, individuals aged 55-85% were significantly under-represented, with 55-59 year olds being

52.86% under-represented and individuals over 75-84 and over being over 98.98% under-represented. Among the 85-100 community, less than 0.27% were sentenced criminals.

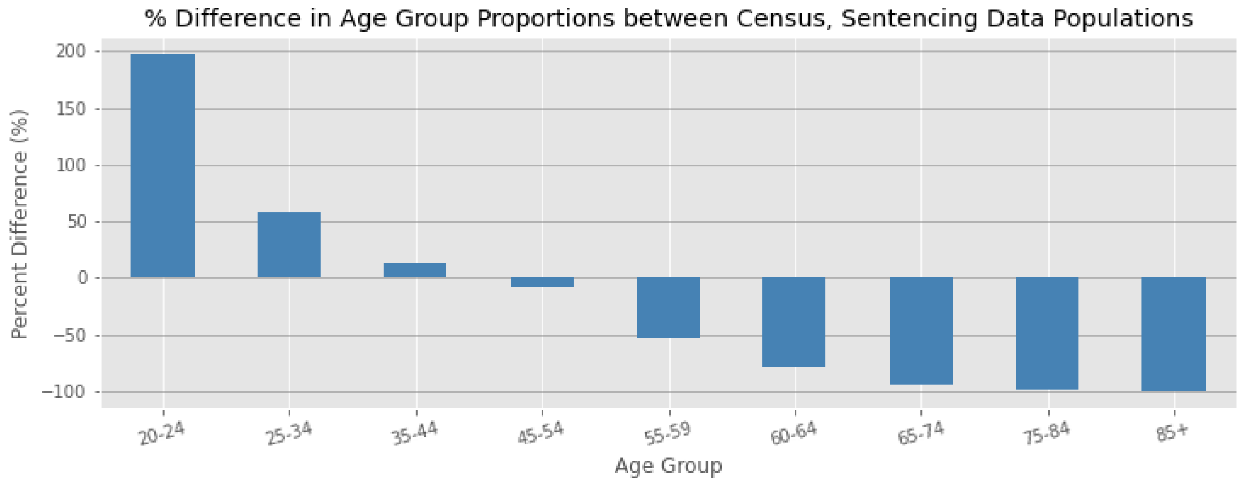

Figure 1.2: % Difference in Age Group Proportions between Census, Sentencing Data Populations

In comparing the racial groups, the sentencing data also showed significant underrepresentation in most groups but a substantial overrepresentation in one racial group. As seen in *Figure 1.3*, Black individuals represented around 66.60% of the entries in the sentencing data, but only represented 21.49% of the general population, presenting a staggering overrepresentation of 188.31%. The other racial groups appear to be under-represented in the sentencing data, with the least under-represented group being Hispanic individuals with 30.81% under-representation. The most under-represented racial group is Asian individuals with a 92.42% under-representation value.

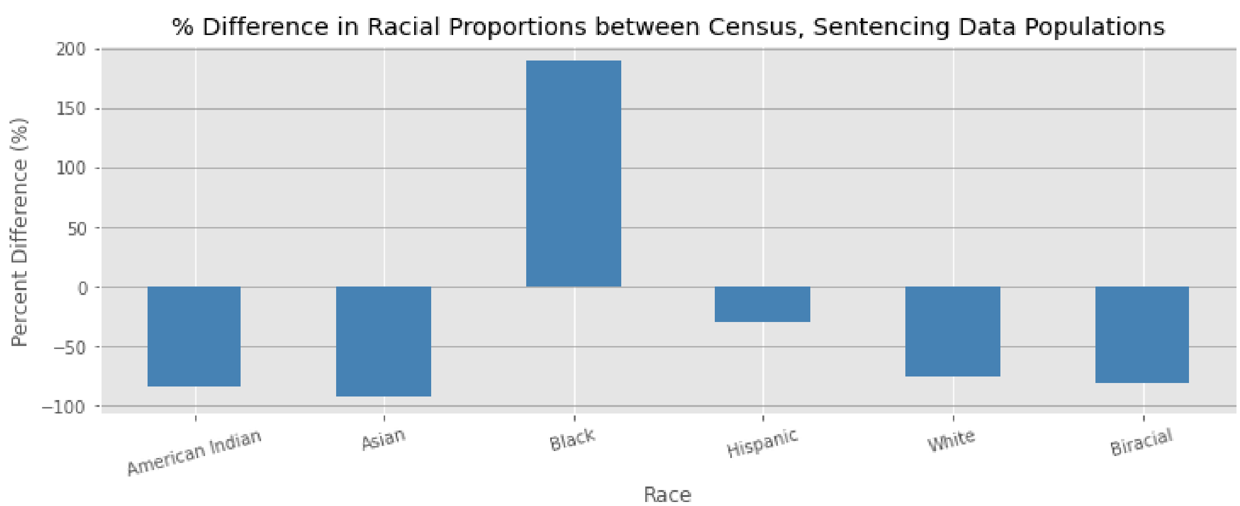

Figure 1.3: % Difference in Racial Group Proportions between Census, Sentencing Data Populations

While our findings show clear trends in the data, indicating groups that are under-represented and over-represented in the criminal sentencing data compared to the general population, we did want to reiterate that these findings are not causal statements. We do not believe that belonging to a particular demographic group has a direct impact on criminal activity. However, we do recognize that there are clear trends in the sentencing data that may warrant further investigation, which itself may lead to uncovering inequities based on demographics.

Question 2: Does unemployment lead to higher rates of crime?

We want to investigate external factors in demographics that may be correlated with crime rates in Cook County. For this question, we will be looking at unemployment data from the Bureau of Labor Statistics and compare the rate of unemployment to the rate of crime in Cook County.

Additional Data Cleaning and Sanity Checks

- Load in BLS unemployment data
- Remove non-numeric characters from the unemployment rates and convert them to floats
- Separate annual unemployment rates from the monthly unemployment rates
- The unemployment and sentencing data must be merged by date. For the sentencing data, use the minimum of the ARREST_DATE and RECEIVED_DATE columns. Break the date into year and month to enable merging with the unemployment data. We use the minimum of these two dates because there are instances when the crime occurs and is documented before the suspect is found or arrested. Since the goal is to compare crime with unemployment at the time, we want to approximate the time of the crime as accurately as possible.

Research Analysis

Experiment 1: Look for correlations between unemployment and overall crime rate.

- From the sentencing data, calculate the annual count of arrests and the monthly count of arrests separately. Merge this data with the corresponding annual and monthly unemployment data.
- Calculate the correlation (using the pandas built in method) between the count of arrests and the unemployment rate for the corresponding time period to determine if there is a relationship between unemployment and crime rate. Because the population of the Chicago area has remained relatively stagnant, I have not attempted to normalize the arrest counts by the population size. Refer to *Figure A.2.1* in the appendix to view the labor force size over time.

Experiment 2: Look for correlation between unemployment and crime rate by category and severity of crime.

- Follow the same procedure as Experiment 1, but calculate the annual and monthly counts of arrests for each offense category separately. Similarly, calculate the annual and monthly counts of arrests for each crime severity class.
- Calculate the correlations between the count of arrests and the unemployment rate for each offense category and for each crime severity class separately to determine if particular types of crime are more influenced by unemployment rates than others.

Experiment 3: Look for correlations between crime rates and past unemployment rates.
- This time, we followed a similar procedure as in Experiment 1 and 2, but instead of comparing unemployment rates corresponding to the time of arrest, we compared unemployment rates from months/years prior to the time of arrest.

Research Findings

Experiment 1

When comparing unemployment rates to overall crime, the correlations are relatively low. Using the yearly data, we find a correlation between crime and unemployment of 0.22. Using the monthly data, we find a correlation of only 0.08. This seems to indicate that there is no significant relationship between crime and unemployment. *Figure 2.1* is a plot of the unemployment rate as a function of crime counts.

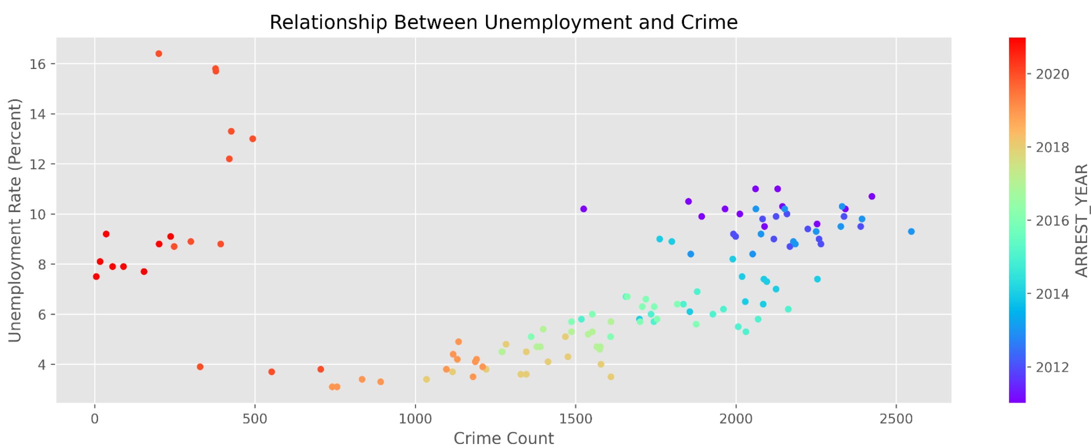

Figure 2.1: Relationship Between Unemployment and Crime

The data points are colored by time, so you can see that crime counts have been consistently decreasing with time, but unemployment has recently increased significantly. It is likely this large jump in unemployment that is impacting the correlation values.

Experiment 2: Crime Category

When we segment the monthly and yearly data by crime category, we find that some crimes are more correlated than others. *Figure 2.2* displays the correlations of each crime category with yearly unemployment rates. Only crime categories with at least 500 counts between 2011-2021 were plotted.

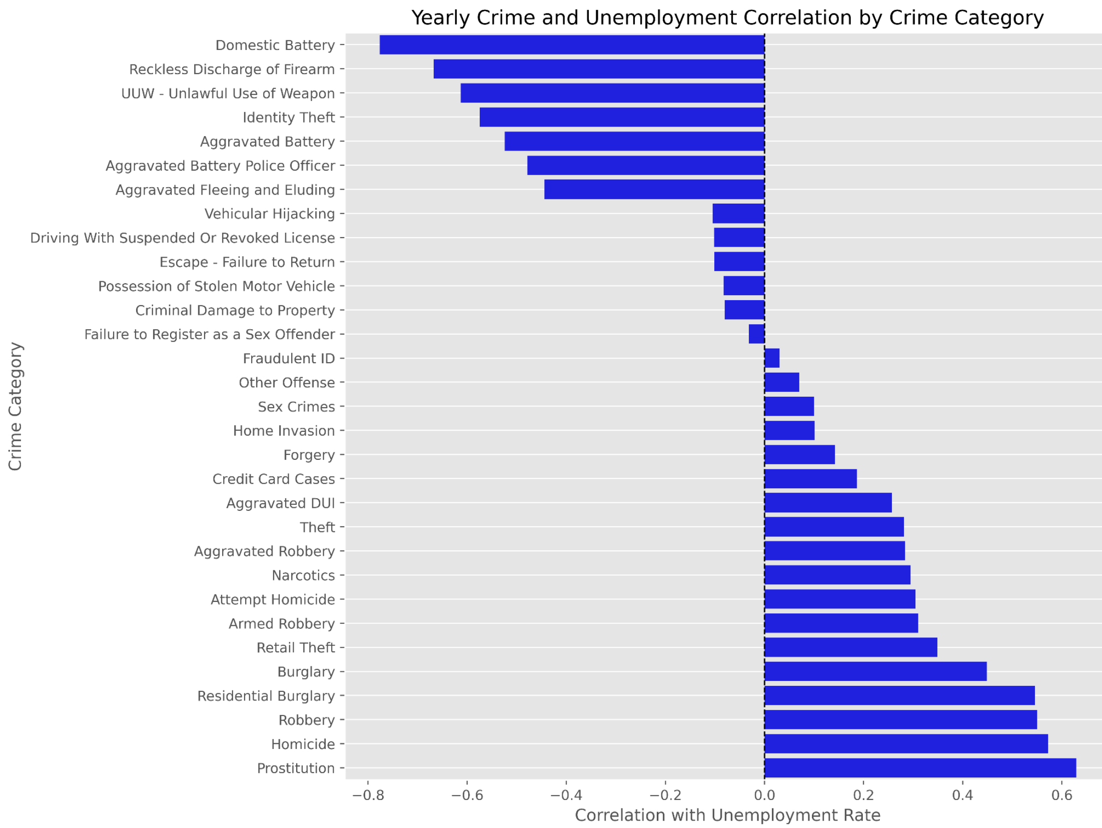

Figure 2.2: Yearly Crime and Unemployment Correlation by Crime Category

Consistent with intuition, you can see that crimes such as prostitution, homicide and varying forms of theft are most positively correlated with unemployment. As individuals become more financially desperate, it is reasonable that they are more likely to commit such crimes. Crimes such as domestic battery, and unlawful use of weapons are most negatively correlated with unemployment. Interestingly, studies show that the reverse is true for domestic battery — that unemployment leads to more domestic abuse. When we plot unemployment over time, we see that it is steadily decreasing until the COVID era, as shown in *Figure 2.3*.

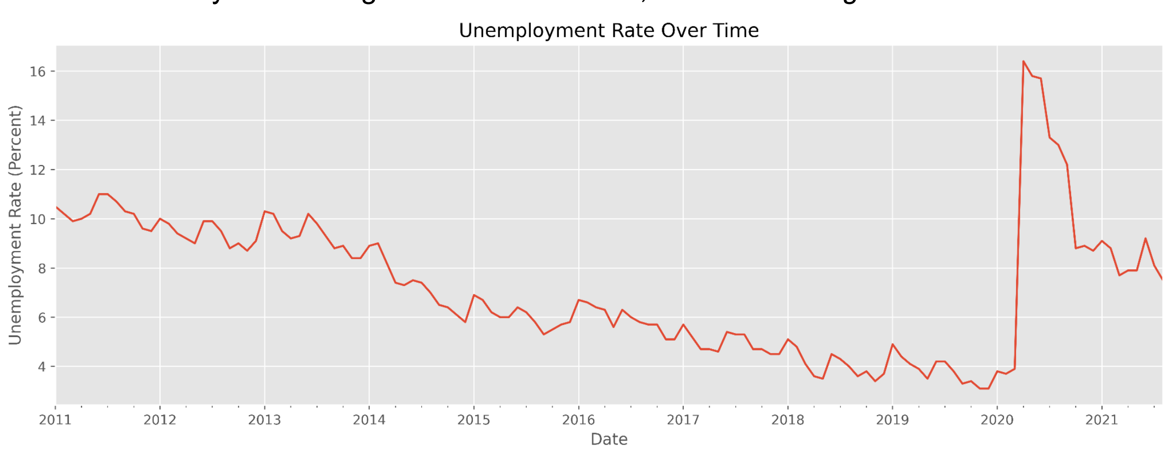

Figure 2.3: Unemployment Rate Over Time

A similar plot of crime counts for the most correlated categories is shown below in *Figure 2.4*.

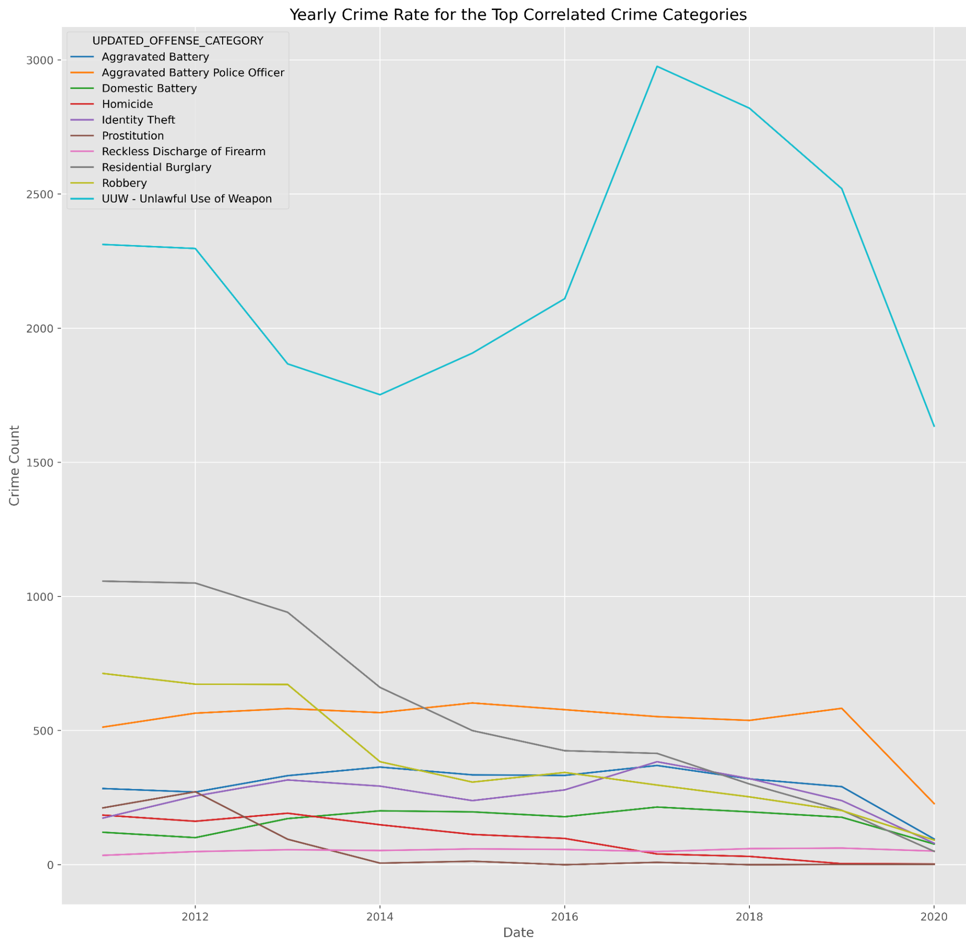

Figure 2.4: Yearly Crime Rate for the Top Correlated Crime Categories

Strangely, we do not see a corresponding jump in domestic battery accompanying the increase in unemployment. It is possible, however, that cases of domestic battery have not been documented or reported to law enforcement, or that delays in court hearings caused by COVID have caused events not to show up in our dataset.

<u>Experiment 2: Crime Class</u>

The relationship between crime class and unemployment is more obvious. The scatter plot in *Figure 2.5* shows the correlations between crime count and unemployment for each class.

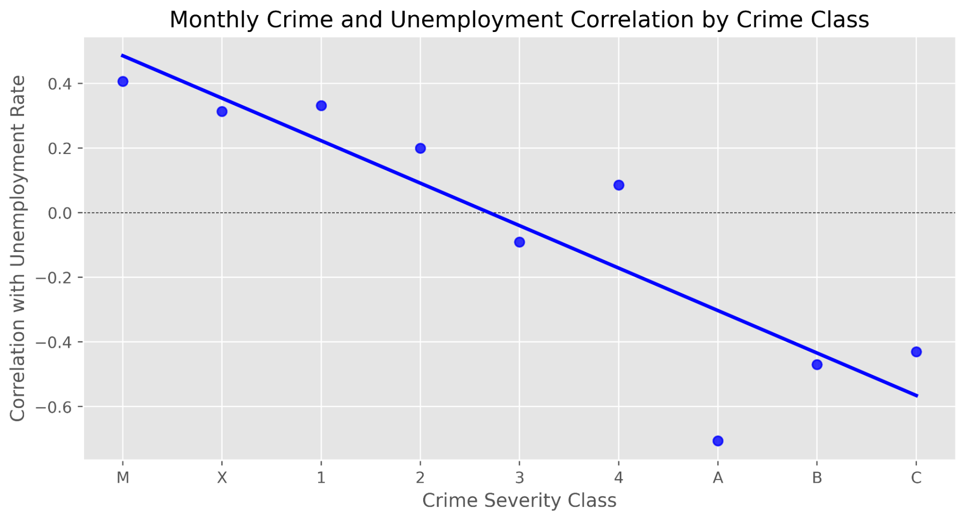

Figure 2.5: Monthly Crime and Unemployment Correlation by Crime Class

As highlighted by the trend line, there is a clear linear relationship between correlation with unemployment and crime severity. High severity crimes are positively correlated with unemployment, while low severity crimes are negatively correlated with unemployment. While it seems intuitive that high severity crimes would increase with unemployment, it is not immediately apparent why misdemeanors would increase when unemployment decreases. One possible explanation is that arrests for misdemeanors decrease when arrests for felonies increase as a result of limited policing resources. When unemployment increases, police prioritize making arrests for higher class crimes.

Interestingly, this linear relationship differs by race, as demonstrated in *Figure 2.6*.

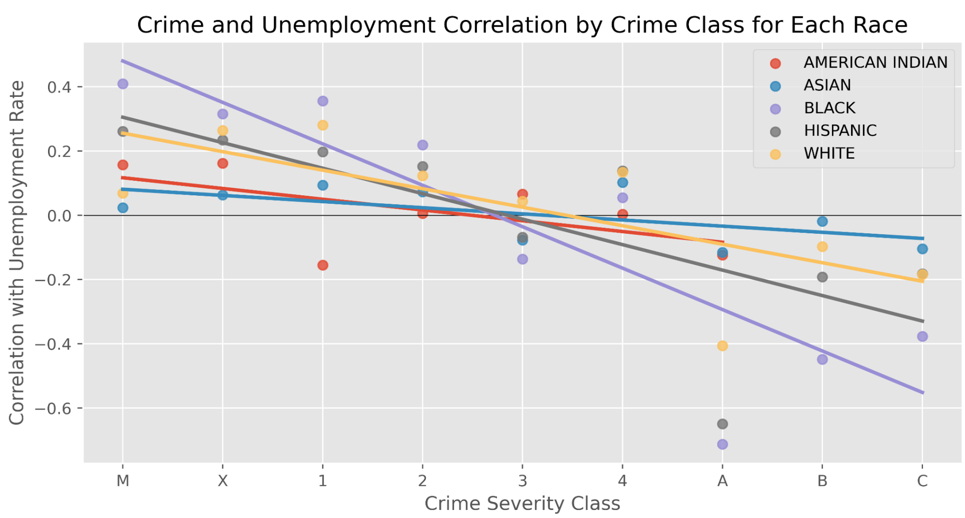

Figure 2.6: Crime and Unemployment correlation by Crime Class for Each Race

Black and Hispanic individuals have higher absolute correlations with unemployment than White or Asian individuals. This relationship makes sense because black and hispanic demographics have the highest unemployment rates in the Chicago area.

Experiment 3

Finally, we applied the correlation analysis between crime rate and past unemployment rates with the goal of detecting whether there is a lag between unemployment and crime. It appears that whether we compare correlations with the monthly or the yearly data, in both instances, the correlation is highest (with a value of 0.91) when the unemployment is shifted by 2 years or 25 months (which is also approximately 2 years). While this appeared to be an interesting conclusion at first, we realized that by shifting unemployment by 2 years, we cutoff the COVID era of unemployment from our analysis. Over the last two years, unemployment has soared to staggering highs. While the previous two years could be considered anomalous, we do not believe that the results from shifting unemployment can be considered as reliable. We would need additional data following COVID to accurately represent these correlations.

In conclusion, while unemployment and crime may not be inherently related, certain crimes (especially high class crimes) are highly correlated with unemployment.

Question 3: Do people of different demographics spend differing amounts of time within the system?

The motivation of this question stems from analyzing the process of the criminal trial process in Chicago with the data provided from Cook County jail.
The independent variable measured for this question will be time in the system. More specifically, the engineered variable, (SENTENCE_DATE - RECEIVED_DATE), to capture this feature.

Data Cleaning and Sanity Checks

- How to measure duration of time in the criminal system / how long the process of trial and sentencing takes?
    - Use sentence date minus initial intake date
- Are there variables that are skewing the correlation between variables in question?
- Are there iffy values such age ages > 100 and years beyond 2021?

Research Analysis

To further understand the details of what happened during the criminal trial process, we will look at the variable of time (from arrest to sentencing) in comparison to the following variables: Age_at_incident, Race, Gender, and Offense_category.

I will be creating tables for each of these comparisons with time as the dependent variables and grouping on the independent variable. All times are from the years 2011 - 2021 and the Race and Gender categories are cleaned as the standard to match the rest of the research questions.

First, take a look at the distribution of length of time from arrest date to sentencing date to see what the mean and median time lengths are. At first glance, there is a heavy skew right towards larger time lengths (>7000). The mean of the distribution is around 400 days and the median is in the range 50-200 days. I then decided to narrow down the distribution to time lengths less than 2000 days and that left us with 90% of the original data. Removing the larger time length values will prevent larger outliers from skewing further correlation results.

Research Findings

By looking at the sentencing time for individuals in Cook County jail across different demographics and criminal offenses, it can be seen that the offense category is one variable that contributes most to differing sentence times. Age, race, and gender provide weaker signs of contribution.

As seen in *Figure A.3.1*, most cases spend around 50-200 days from their arrest date to their sentence date. The histogram skewed right which indicates that there are a trickled number of cases that take longer than 200 days from arrest to sentencing. This skew increases the overall average number of days taken from arrest to sentencing.

The top three offense categories that correlate to the highest number of days on average from arrest to sentencing are homicides (1295 days), pl ossession of explosives (890 days), and sex crimes (870 days). It is shocking to see that these top offense categories take more than around 3+ years to go through the criminal trial process. However, it makes sense that the crimes having this degree of severity take longer on average. The rest of the offense categories and their respective time lengths can be seen in *Figure A.3.2*.

When looking at time from arrest to sentence by race category, there is not a significantly large difference between the different racial categories. Asian (372 days), Hispanic (356 days), Black (342 days), White (322 days), American Indian (317 days). However when looking at the difference in time spent in the criminal system across race and offense category, the differences in time spent for each racial group differs largely by offense category. *Figure A.3.3* shows the complete set of differences in time from arrest to sentence by offense category and race. *Figure 3.1* below zooms in on a few offense categories to show the significance of unequal times spent in the criminal trial process incurred by different racial groups.

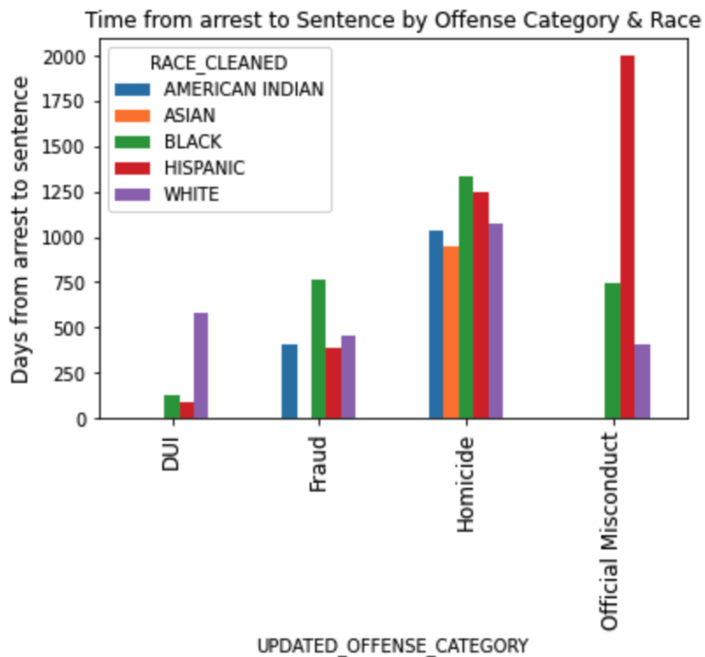

Figure 3.1 Time from arrest to Sentence by Offense Category & Race

Question 4: Are people of different age groups sentenced differently per crime?

Race, gender, wealth, and their impacts on criminal sentencing have been a popular topic for social and criminal justice activists, especially in light of recent movements such as Black Lives Matter. However, one key demographic that seems to have evaded activist interest is age. Our research investigated whether age may be a factor in determining sentence severity.

Research Analysis

The sentencing data was imported from a CSV file into Python using the Pandas library. After the data was cleaned using the steps above, the cleaned data was separated into groups based upon the five most prevalent criminal offense types; the offense types were: Narcotics, Unlawful Use of Weapon, Retail Theft, Burglary, and Aggravated DUI. Within each offense category, the sentencing data for a particular offense was then further categorized into age ranges for the offending criminal at the time of their arrest; the data was organized into the following age ranges: 20-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75-84, and 85-100. With each criminal offense being segmented into age ranges, we computed the mean and median sentence lengths for each age range by analyzing the normalized, “sentence_in_days” field that was generated.

With each criminal offense type now segmented into age groups, and each age group having an associated mean and median value for their sentence length, we begin analyzing the sentence severity across all age groups within an offense type. This was accomplished by computing the following:

- The mean and median values for the average length of sentence across all age groups within a particular offense type (e.g. all age ranges for Narcotics offenses).
- The difference in the mean length of sentence for a particular offense and age range and the mean length of sentence for the offense including all age ranges.

Firstly, the mean and median values for the average length of sentence across all age groups within a particular offense type was calculated, providing baseline values for the average sentence length. Secondly, the difference between the mean and the median across all age groups was computed to determine how a baseline difference between the total mean and median value was, indicating how influential extreme sentencing values were in computing the mean. The table computing these values were compiled into the table *Figure A.4.1*, which can be found in the Appendix.

With the mean and median sentence length values computed for an entire offense category, we then proceeded to analyze the mean, median, and difference values within each offense type's age ranges. For each age group, we computed the difference between the age range's mean sentence and the offense group's mean sentence across all age ranges. The same was done to compute the difference between the age range's median sentence and the offense group's median sentence across all age ranges. These computations were compiled into tables, which can be found in the Appendix as *Figure A.4.2* and *Figure A.4.3.* Lastly, we computed the percentage difference between the mean differences and the median differences respectively.

Lastly, we used all of the computed sentencing difference data to create an aggregated set of values to represent how fairly each age group is sentenced. For each age group, we summed the mean difference values for each offense category, thus creating an aggregate value "total_mean_difference", also known as the Total Mean Sentence Difference, indicating how fairly an age range is sentenced on average. If the value is highly positive, this indicates that that age range is sentenced more severely on average compared to all other age ranges; if the value is highly negative, this indicates that the age range is sentenced less severely on average compared to other age ranges; if the value is close to 0, this indicates that the age range is sentenced equally severely on average compared to other age ranges. This method for computing the total mean difference value was similarly applied for computing the Total Median Sentence Difference value as well.

Research Findings

Our research discovered a number of substantial differences in sentence severity between age ranges. There appears to be a definitive association between age range and sentencing severity in many cases and circumstances. While there is not enough evidence to suggest that age range is a direct factor in sentencing, the severity values may be high enough to warrant further investigation.

The findings are best seen in the figures below.

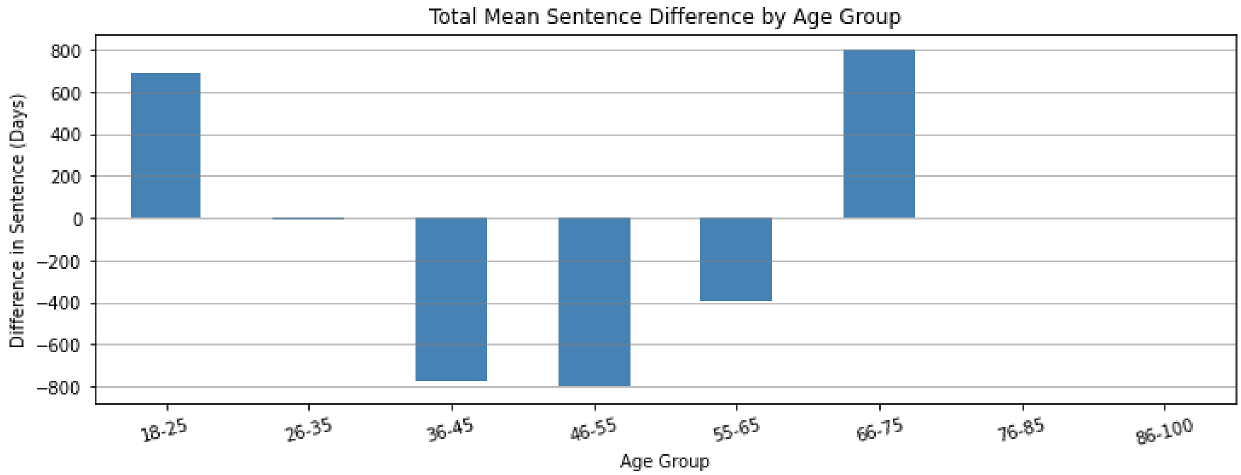

Figure 4.1, Offense Sentence Mean Difference by Age Group

Figure 4.1 shows the overall difference value in the mean sentence for each age range and offense category compared to the offense category's mean sentence across all age ranges.

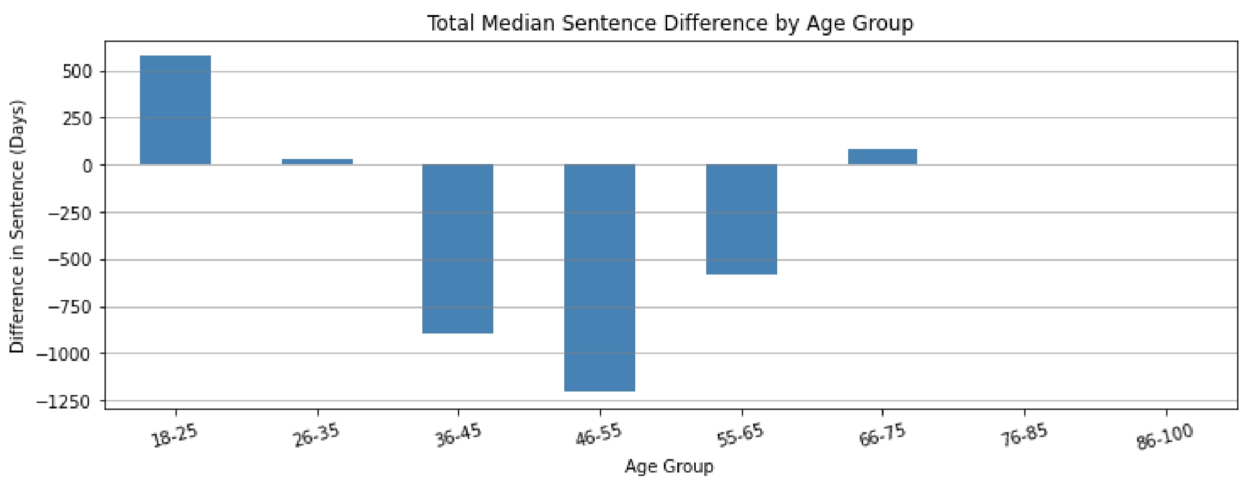

Figure 4.2, Offense Sentence Median Difference by Age Group

Figure 4.2 shows the difference value in the mean sentence for each age range and offense category compared to the offense category's mean sentence across all age ranges.

These figures provide a number of interesting insights into how one's age group and offense_type is associated with the severity of their sentence. Below are some of the most striking insights provided:

- The age range 18-25 had a higher average sentence length of an additional 690 days for the mean sentence, and an additional 575 days for the median sentence across all offense types. This supports the belief that younger offenders are sentenced more severely than other age groups.
- The age range 36-65 all saw moderate to high decreases in sentencing severity for both total mean difference and total median difference. The 46-55 age range was subjected to

the lowest sentencing severities, being sentenced to on average 312 fewer days for the mean sentence across all offenses, and 1204.5 fewer days for the median sentence across all offenses.

- The 66-75 age range has a substantial difference in the mean difference sentence and the median difference sentence, suggesting that there are individuals in that age range who received extremely high sentences that introduced a bias into the mean average.
- It seems like only the 26-35 age range is sentenced fairly and in-line with the average sentence severity across all age ranges.
- Lastly, it appears that not enough data exists for individuals aged between 76 and 100 for the crime offense types that were being investigated.

While our findings show clear trends in the data, indicating certain groups are given sentences of differing severities, we did want to reiterate that these findings are not causal statements. While we could not account for data missing from the sentencing dataset, such as criminal history or the severity of the crime, we do recognize that there are clear trends in the sentencing data that may warrant further investigation, which itself may lead to uncovering inequities based on demographics.

Question 5: Are people sentenced differently based on demographics?

Recent social justice movements such as Black Lives Matters have cast a light on how race and gender play influential roles in how people are treated, especially when interacting with the criminal justice system. Cook County is a highly populated, racially diverse region that received a lot of attention during the Black Lives Matters protests in 2020 and 2021. With that in mind, our research investigated whether or not gender and race were an influence in one's sentencing, and if so to what extent?

Research Analysis

Gender

Performing a simple average on the sentence terms, we found that women were typically given sentences that were 336 days shorter than the average sentence length for men (refer to *Figure A.5.1* in the appendix). However, comparing the averages for each demographic is insufficient. Demographics vary across crime classes and crime categories, as shown in *Figures 5.1* and *5.2* (you can also see the gender distributions across each crime class and category in *Figures A.5.2* and *A.5.3* in the appendix).

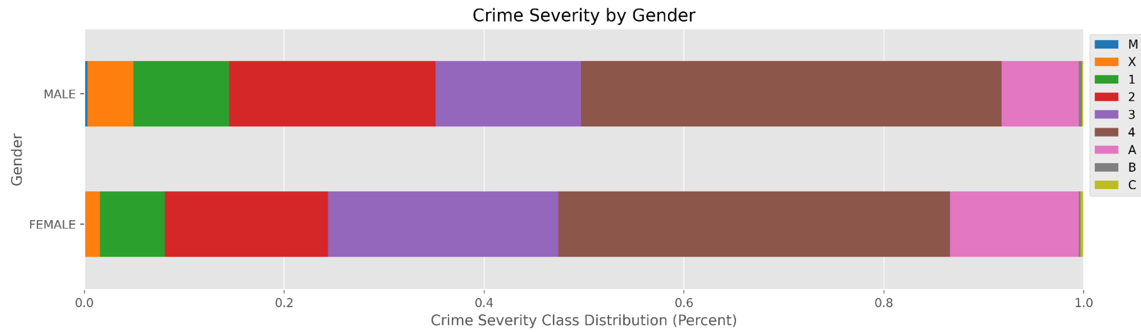

Figure 5.1: Crime Severity by Gender

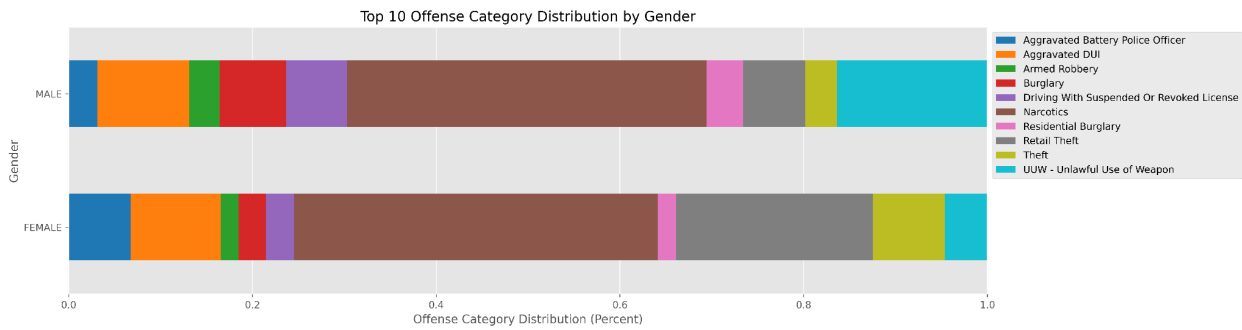

Figure 5.2: Top 100 Offense Category Distribution by Gender

Women make up a smaller proportion of high severity crimes. As a result, the average sentence for women would naturally be lower (refer to *Figures A.5.4* and *A.5.5* in the appendix to see the average sentence for each crime class and category). To account for this, we normalized sentences by calculating the percent deviation of a given sentence from the mean for its corresponding crime class and category. We then averaged the percent deviations across each demographic group. The results are shown in *Figure 5.3* below.

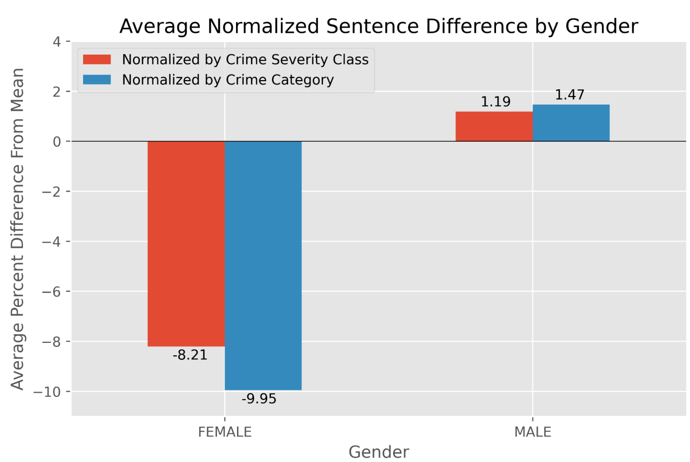

Figure 5.3: *Average Normalized Sentence Difference by Gender*

Based on these results, we can conclude that:

- Women are given less severe sentences in comparison to men for the same crimes. Note that the data only accounts for a binary representation of gender.
- On average, women have 9% shorter sentences compared to men for the same crime class, and they have 11% shorter sentences compared to men for the same crime category.

Race

Following the process detailed above for gender, we performed a similar analysis to determine whether individuals are sentenced differently by racial demographic. *Figures 5.4* and *5.5* below demonstrate the varying demographic distribution across crime classes and categories. Refer to *Figures A.5.6* and *A.5.7* in the appendix to see how races are distributed within crime classes and categories.

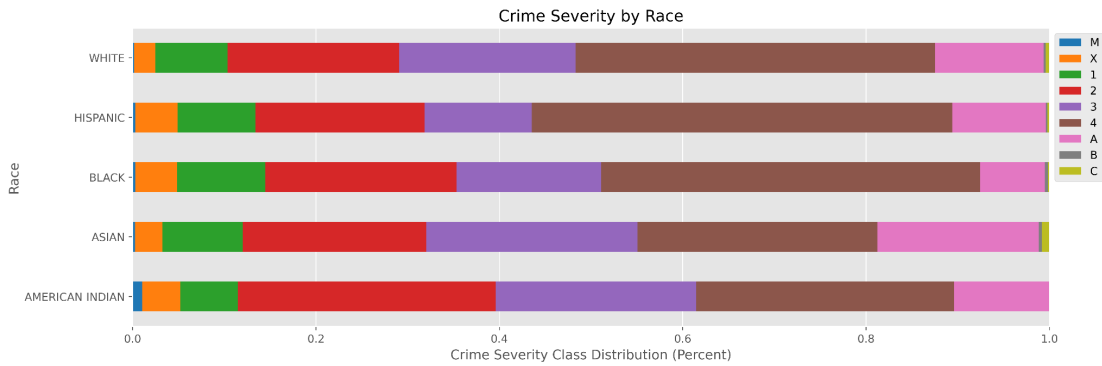

Figure 5.6: Crime Severity by Race

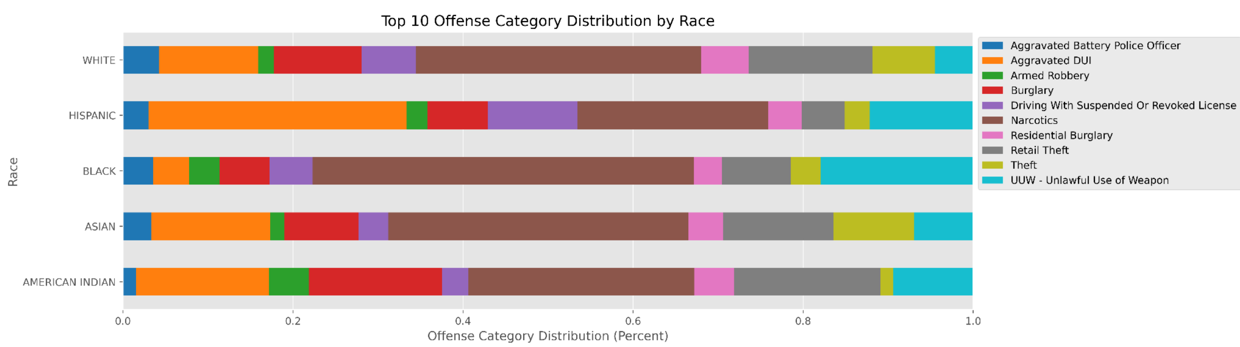

Figure 5.7: *The 10 Offense Category Distribution by Race*

You can see from *Figure 5.6* that American Indian individuals have a higher proportion of class M (first degree murder) crimes compared to the other race categories. It is, therefore, unsurprising that American Indian individuals had the highest overall average for sentences. For this reason, we again normalized the sentences by the mean for the corresponding crime class and category using the methods detailed in the *Gender* section. The results of this analysis are shown in *Figure 5.8* below.

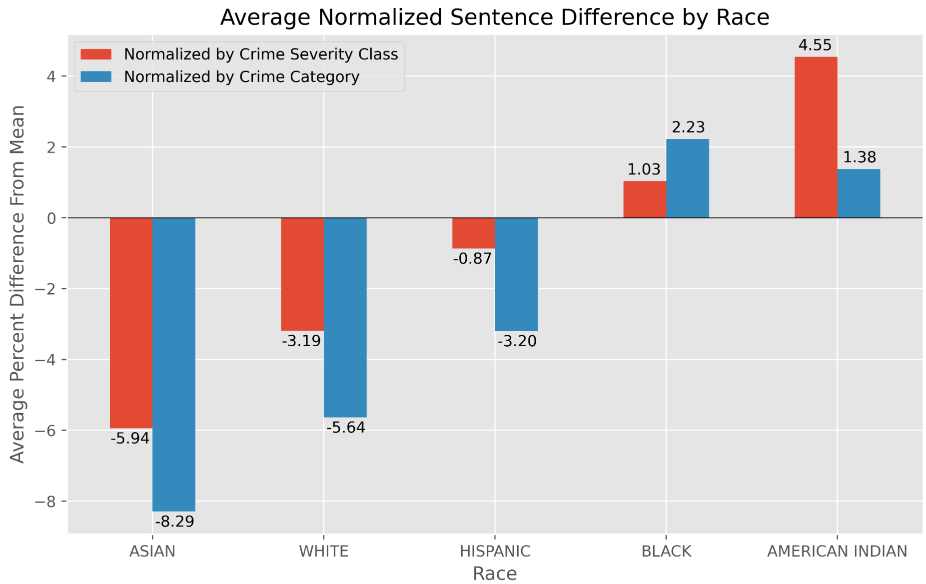

Figure 5.8: Average Normalized Sentence Difference by Race

- We found that sentencing is not equal across racial groups. Asian individuals on average have the shortest sentences, while American Indian individuals have the longest sentences on average. Note, however, that there are only 96 entries for sentences against Native American individuals, and only 1193 such entries for Asian individuals. Consequently, the results for these two categories may have low significance.
- While American Indian individuals have the longest sentences when normalizing by crime class, Black individuals have the longest when normalizing by crime category. This is likely due to the fact that of the homicides committed, for black individuals, 54% of homicides were classified as Class M homicides (the most severe) while for American Indian individuals, only 33% were labeled as Class M.

This last finding led us to investigate how crimes within the same offense category are classified per race group. With the classes ranked in order of most severe (1) to least severe (9), we normalized the crime severities by calculating the difference between an individual sentence severity and the average severity for the corresponding crime category. We then averaged this normalized severity across each race group. The results are shown in *Figure 5.9*.

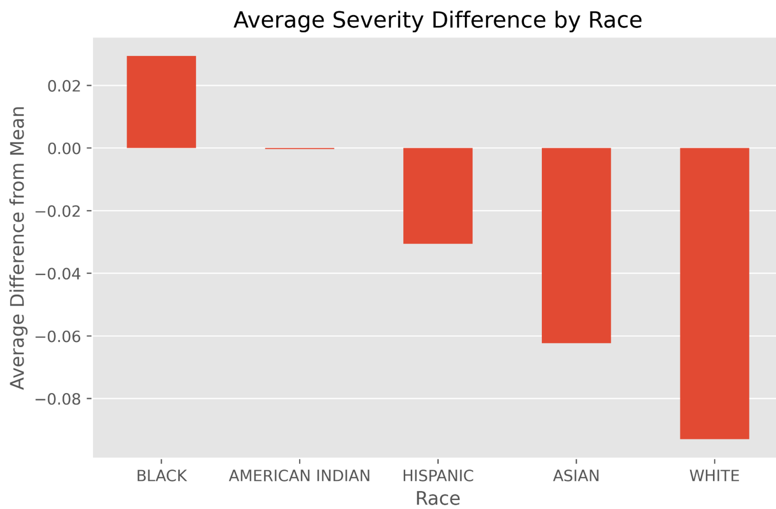

Figure 5.9: Average Severity Difference by Race

On average, crimes committed by black individuals are classified most severely, while crimes committed by white individuals are classified least severely. These findings, in conjunction with the discrepancies in average normalized sentence length demonstrate that the severity of punishment is not equal across race groups.

Overall Findings and Conclusions

After cleaning and analyzing this data, it is clear that there are a number of significant trends connecting demographics, crime, and justice in Cook County. These injustices and unfair treatments are summarized within the following trends:

- Criminal population is not representative of the general population, with men, young adults, African Americans being highly over-represented in the criminal population.
- Unemployment is associated with spikes in certain types of crimes.
- Younger individuals are sentenced more harshly than middle-aged or older individuals.
- Certain races spend more time in the justice system, notably African Americans.
- Race and gender are not sentenced the same, notably African Americans, American Indians, and men were sentenced more harshly.

These findings illustrate that there is injustice within the county's justice system. More research is needed to get to the bottom of this. If we were to continue with this project, we can further investigate Cook County's arrest data to study individual case participants' profiles. Even though we discovered a number of disturbing trends, we do applaud Cook County for making this data open source; by having this data available to the public to scrutinize, we can identify that such issues exist and therefore take the necessary steps to solve them.

Appendix

Question 1:

Figure A.1.1: Sentencing and Census Population Size Comparison
Total Users in Sentencing data: 125347
Total Users in Census data: 5150233
% of criminals in civilian population: 2.4338122178161647 %

Figure A.1.2: Sentencing and Census Gender Comparison
% Female in Sentencing data: 12.313816844439835 %
% Female in Census data: 51.4 %
% difference in sentencing data, census data: 23.956842109805127 %

age_group	sentencing_prop	census_prop	difference	%_representation_in_sentencing
20-24	24.76	8.32	16.43	197.49
25-34	33.89	21.49	12.4	57.69
35-44	20.14	17.87	2.27	12.69
45-54	14.94	16.29	-1.34	-8.25
55-59	3.86	8.18	-4.32	-52.8
60-64	1.67	8.03	-6.36	-79.23
65-74	0.67	11.38	-10.7	-94.09
75-84	0.06	5.86	-5.8	-98.96
85+	0.01	2.58	-2.58	-99.72

Figure A.1.3: Sentencing and Census Age Comparison

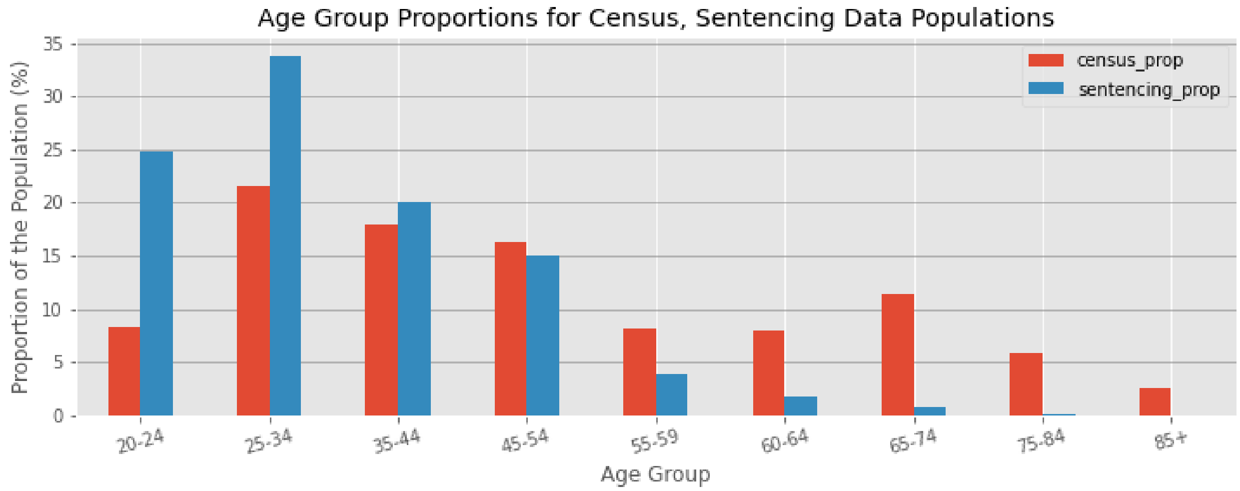

Figure A.1.4: Age Group Proportions for Census, Sentencing Data Populations

race	**sentencing_prop**	**census_prop**	**difference**	**%_representation_in_sentencing**
American Indian	0.05	0.3	-0.25	-84.58
Asian	0.57	7.7	-7.13	-92.55
Black	66.72	23.1	43.62	188.82
Hispanic	17.81	25.6	-7.79	-30.45
White	13.96	56.6	-42.64	-75.34
Biracial	0.55	2.8	-2.25	-80.51

Figure A.1.5: Sentencing and Census Race Comparison

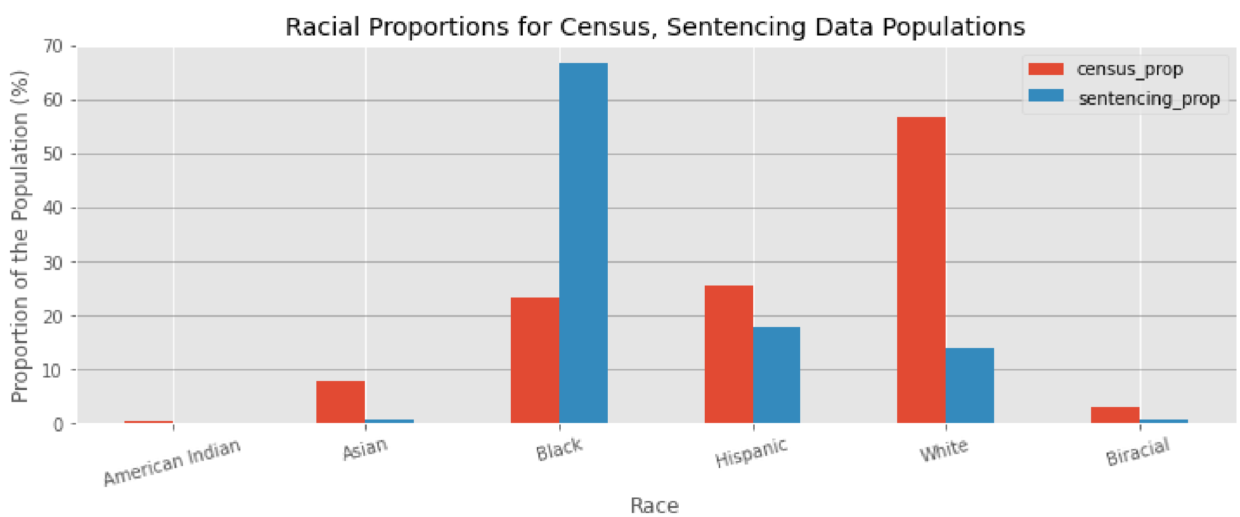

Figure A.1.6: Racial Proportions for Census, Sentencing Data Populations

Question 2:

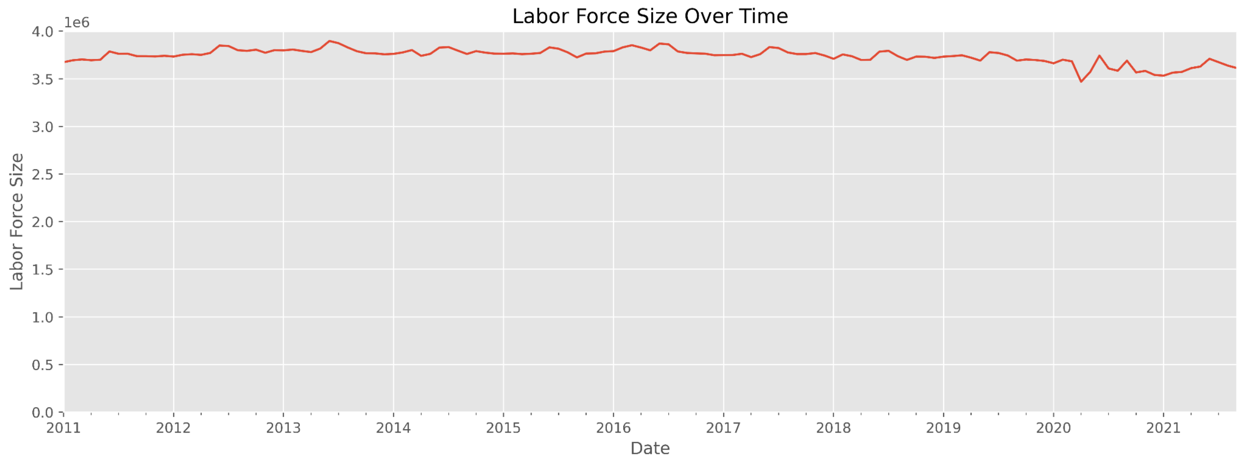

Figure A.2.1: Labor Force Size Over Time

Question 3:

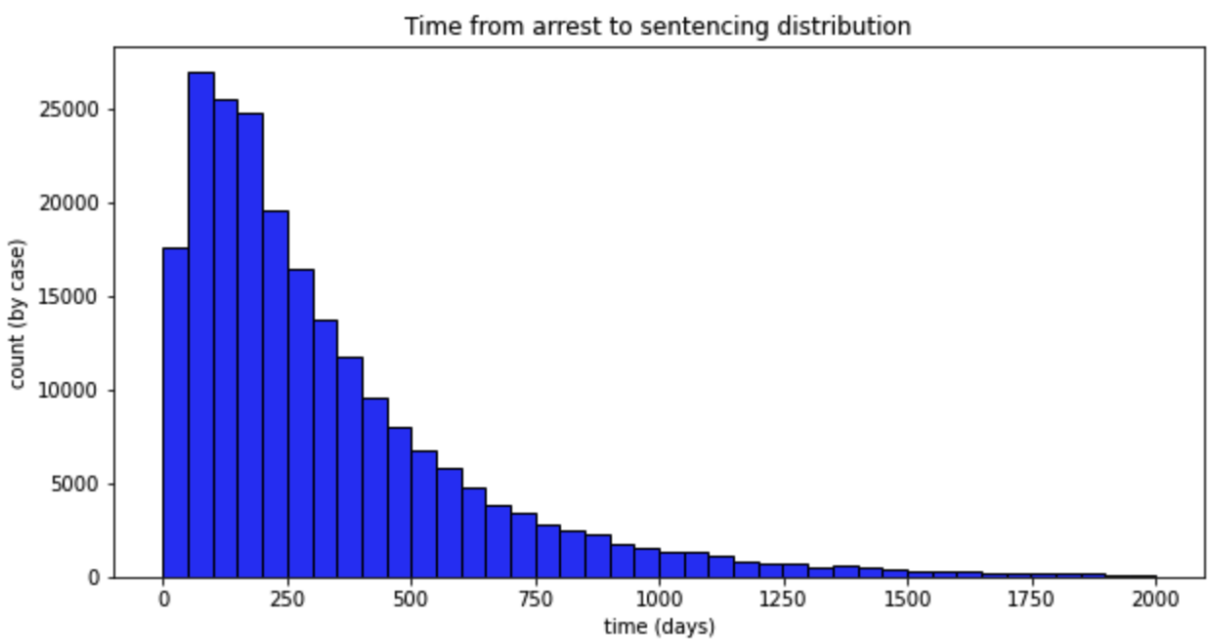

Figure A.3.1: Histogram that shows the count of cases by number of days taken from arrest to sentencing

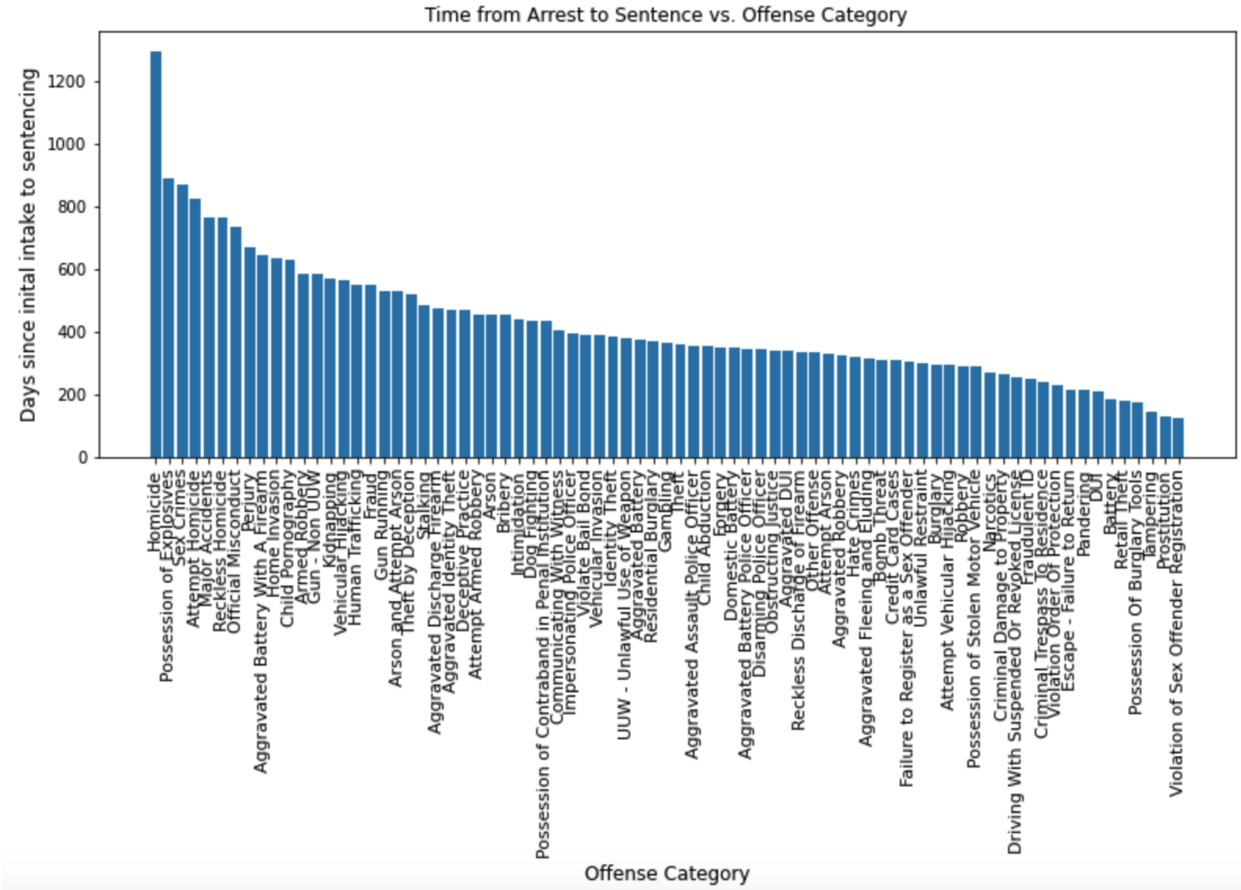

Figure A.3.2: Time from Arrest to Sentence by Offense Category

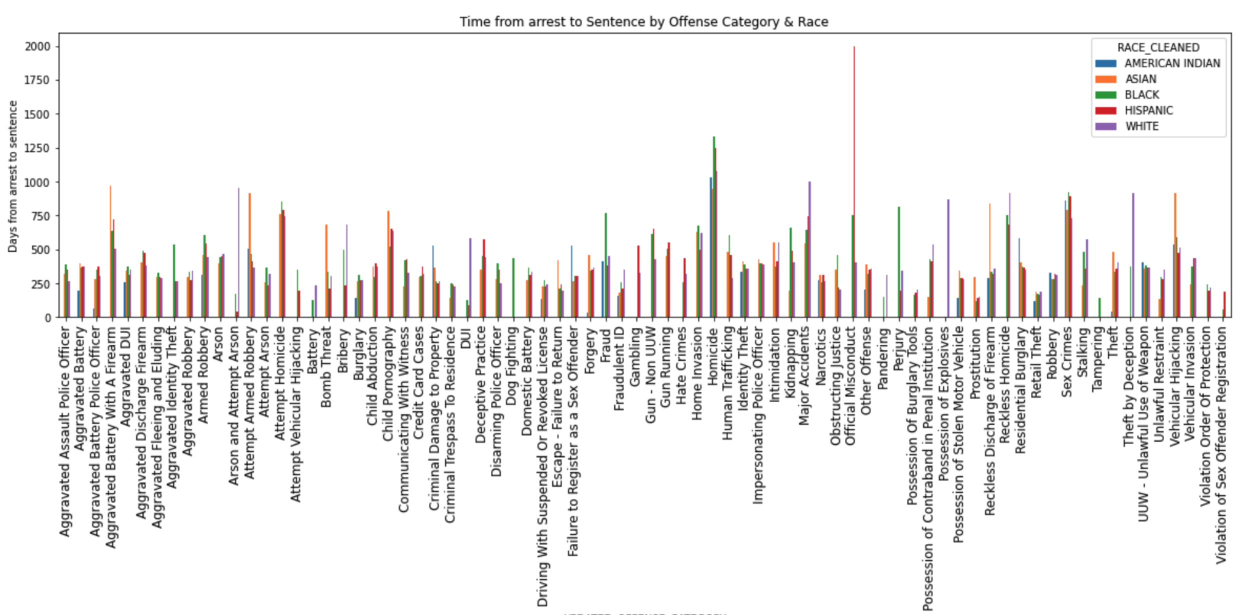

Figure A.3.3: Time from Arrest to Sentence by Offense Category and Race

Question 4:

Figure A.4.1: Total Mean and Median Sentence Length Across All Offenses
Mean: 1062.148348122105
Median: 730.0

age_group	narcotics_mean	uwu_mean	retail_theft_mean	burglary_mean	aggrivated_dui	total_mean_difference
18-25	95.25	243.34	-34.52	302.86	83.40	690.33
26-35	17.43	-278.26	-3.31	130.73	128.78	-4.63
36-45	-71.13	-339.37	-13.40	-222.64	-127.41	-773.95
46-55	-91.90	-339.28	7.63	-312.34	-64.25	-800.13
55-65	1.72	-161.46	52.98	-135.84	-151.75	-394.35
66-75	242.55	315.37	86.54	9.95	143.09	797.50
76-85	nan	nan	246.45	nan	-452.85	nan
86-100	nan	nan	nan	nan	nan	nan

Figure A.4.2: Heatmap of Offense Sentence Means and Total Mean Difference by Age Range

age_group	narcotics_median	uwu_median	retail_theft_median	burglary_median	aggrivated_dui	total_median_difference
18-25	0.00	285.00	-190.00	200.00	280.00	575.00
26-35	0.00	-165.00	0.00	0.00	190.00	25.00
36-45	-170.00	-365.00	0.00	-365.00	0.00	-900.00
46-55	-182.50	-292.00	0.00	-730.00	0.00	-1204.50
55-65	0.00	0.00	147.50	-365.00	-365.00	-582.50
66-75	0.00	0.00	175.00	0.00	-87.50	87.50
76-85	nan	nan	175.00	nan	-635.00	nan
86-100	nan	nan	nan	nan	nan	nan

Figure A.4.3: Heatmap of Offense Sentence Medians and Total Median Difference by Age Range

Question 5:

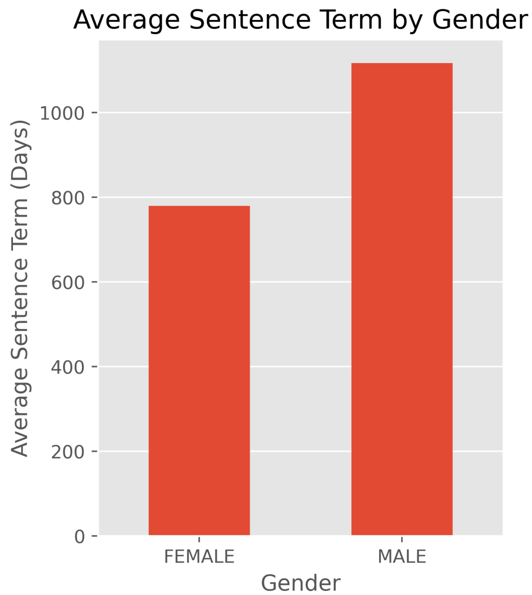

Figure A.5.1: Average Sentence Term by Gender

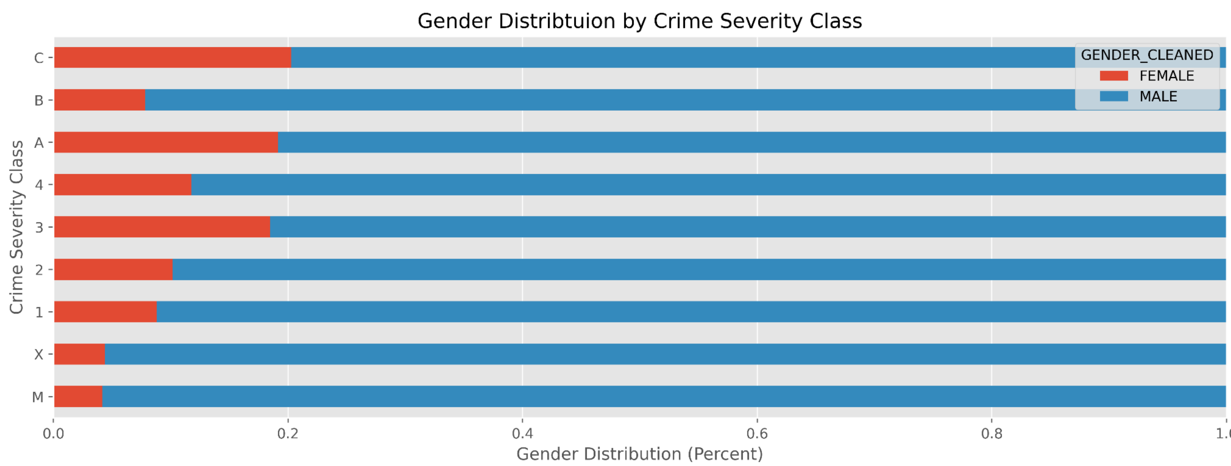

Figure A.5.2: Gender Distribution by Crime Severity Class

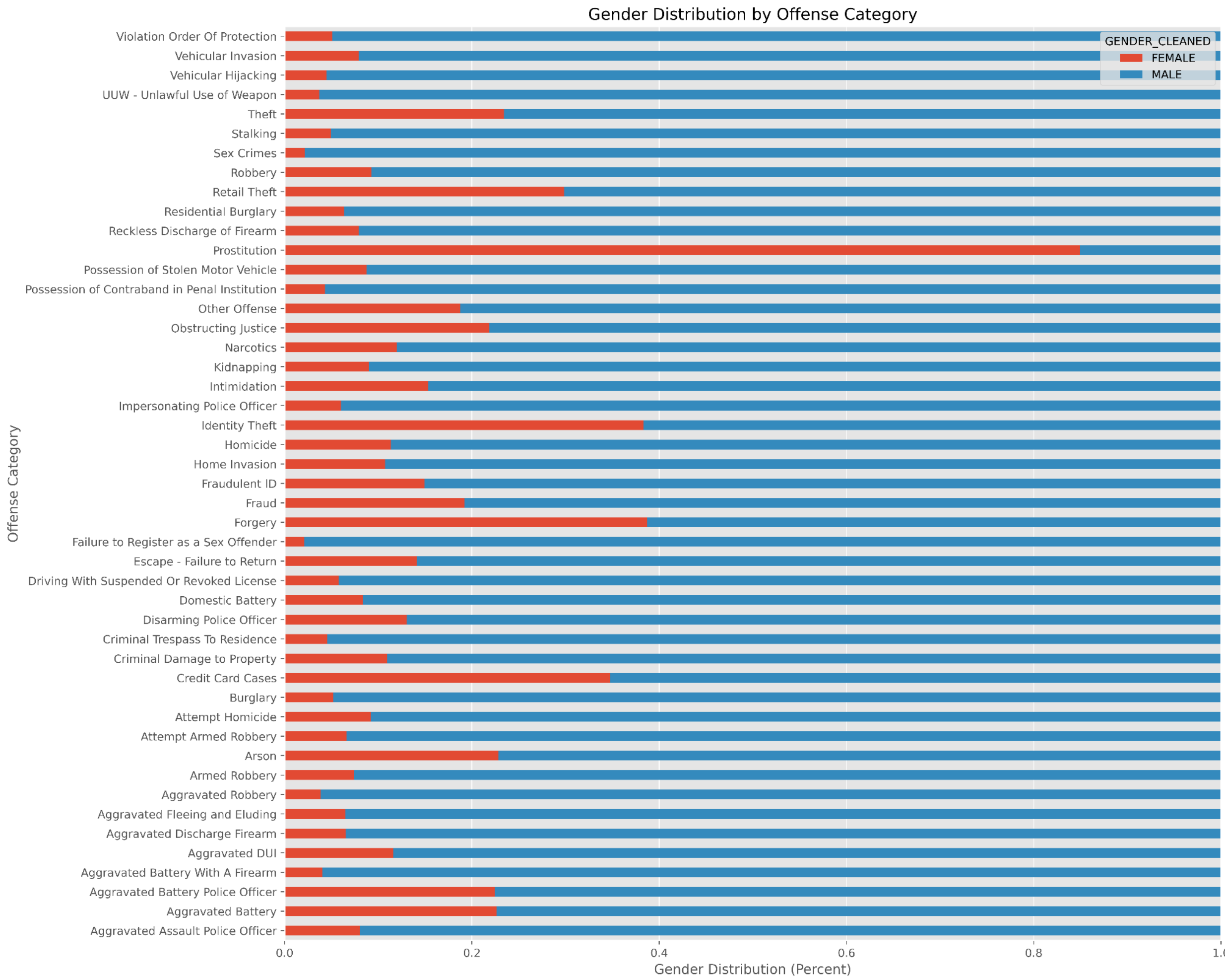

Figure A.5.3: Gender Distribution by Offense Category

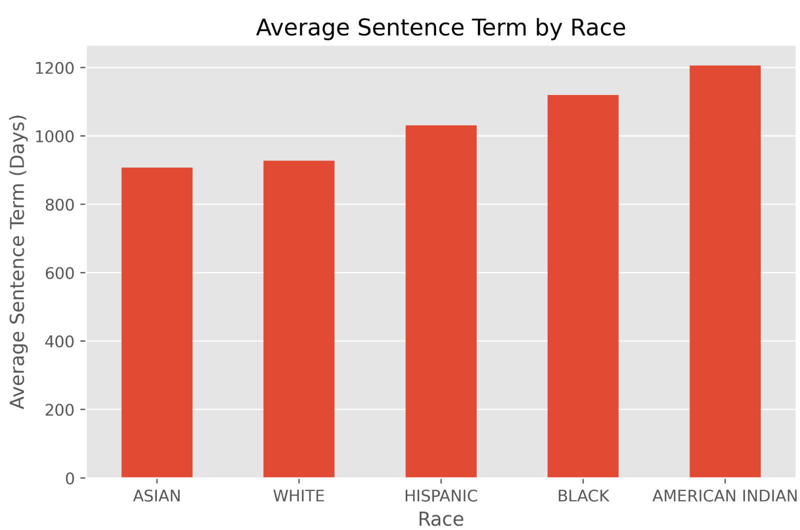

Figure A.5.4: Average Sentence Term by Race

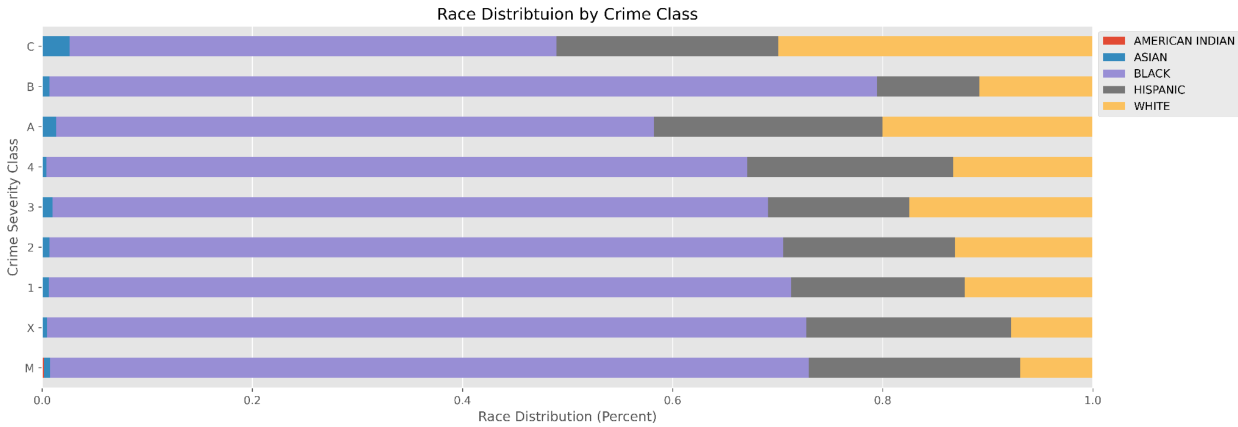

Figure A.5.5: Race Distribution by Crime Class

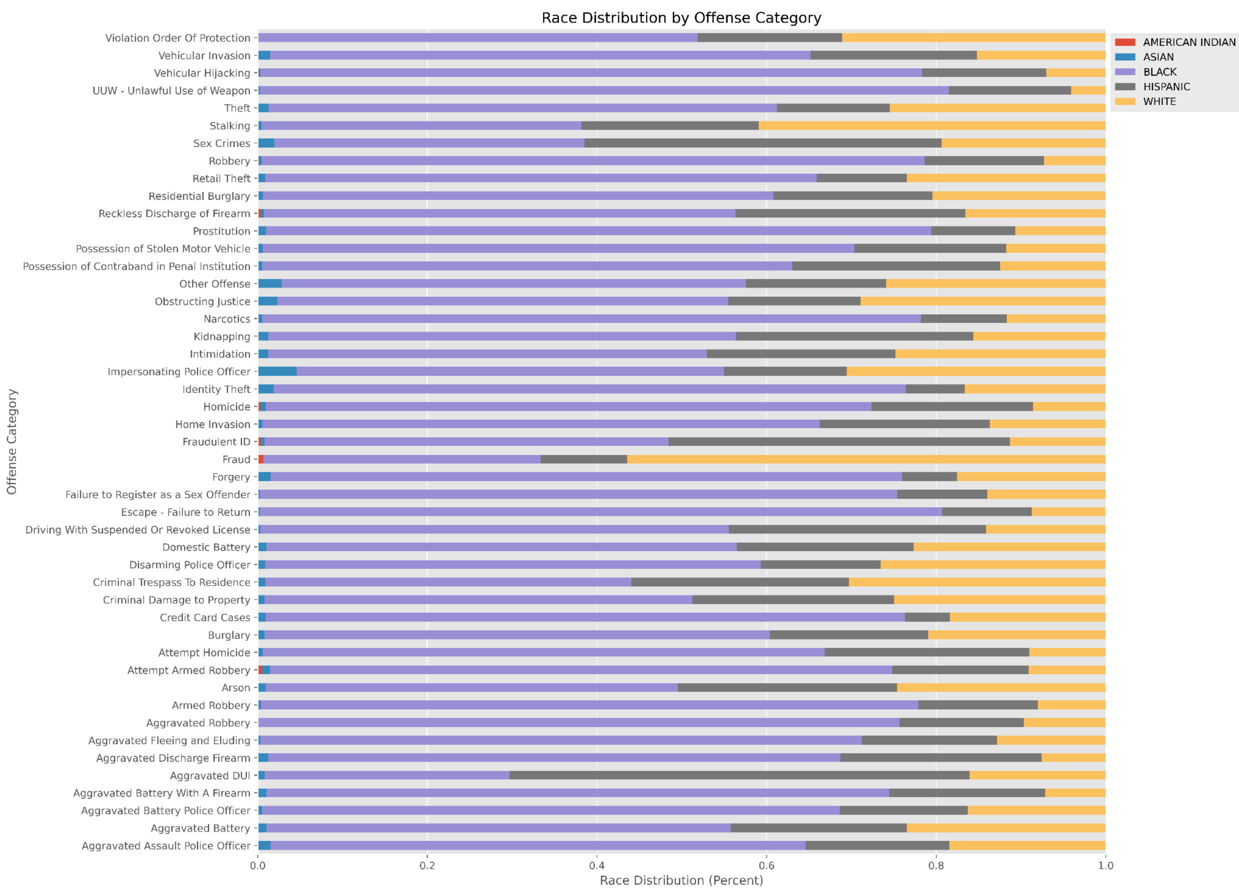

Figure A.5.6: Race Distribution by Offense Category