Race and Incarceration in America

December 12, 2019

1 Introduction

The United States' criminal justice system is a large complicated machine that seeks to deliver justice when an offense has been committed. This system has been slowly evolving as our society and culture have been changing. Many things that Americans take as natural in our criminal justice systems are quite abnormal among justice systems worldwide. Since the 1990s, America has seen a drastic increase in the incarcerated population, despite a sharp decrease in crime[6]. Many Americans may believe that this drastic increase in incarceration is a result of increased rates of crime, and that this heightened rate is natural and just. To many, it is unclear who is most affected by this drastic change in the application of justice in America, and even more unclear is how they are so affected.

There is a lot of existing research exploring incarceration and the criminal justice system. There is a consensus that America incarcerates a larger proportion of its population than any other nation and that people of color are disproportionately affected by this high incarceration rate[5].

In this project I am interested in exploring the ways in which the law is being applied differently to people in America. I will explore different ways to quantify claims about mass incarceration and racial bias. I will also be examining things that factor into sentence length including the offense committed and race. There are many things that contribute to sentence length, however the scope of this project is limited to these factors.

2 Data

2.1 Source and Credibility

The data that I will be using in this analysis is gathered from primarily two sources. The first is the Bureau of Justice Statistics and the second is a link to a database hosted on Data.gov and maintained by the State of Connecticut Department of Corrections. These are highly credible sources because they are primary sources for the data. These organizations are official government agencies which collect, maintain, and report on this data.

2.2 Gathering and Cleaning

All the data which I am using in this report are freely available to the public. Collection and cleaning was relatively simple as the source data was well maintained. The data that I collected from the Bureau of Justice Statistics (BJS) needed to be formatted in a way that is easily read by the Python packages I will be using. This data was prepared in .xlsx files as to be easily human readable, however this is not generally easily ingested by programs. I extracted data that I found to be relevant into separate .csv files and kept the original files for reference. The files are

incarceration_counts.csv
incarceration_by_race.csv
crime_data.csv.

The file that I obtained from the Connecticut Department of Corrections is a very well maintained database. The largest issue I had with this file was mild inconsistency with the way in which certain data was encoded (ex. race was encoded as both `WHITE` and `WHITE\t`). This was the data that I spent the most time working to engineer as it is a data set that I intend to use for different regression-related analyses. The files are

individuals.csv
regression_df.csv.

2.3 Contents

2.3.1 Bureau of Justice Statistics Data

Here I will describe generally each data file and its contents, as well as give a small sample from each file [3,4,5].

First we will examine `incarceration_trends.csv`. This data set records total jail and prison populations across the United States over time. This is useful in understanding general trends in the U.S. over time. The prison population columns are raw populations of incarcerated individuals while the "population" column is the U.S. population in millions for that year.

	Year	State prisons	Population
24	1949	146881	149.19
61	1986	485553	240.13
75	2000	1209130	282.16

`incarceration_by_race.csv` contains race demographic data for incarcerated populations by institution. This will allow us to understand state incarcerated populations.

	Geography	Total	White	Black	White_rate	Black_rate
42	South Dakota	6327	3708	476	530	4663
3	Arizona	67767	36160	8246	775	3184

`crime_data.csv` records crime rates over time in this U.S. This data set will help us understand how crime relates to incarceration. The rates are given in offense per 100,000.

	Year	Violent crime	Murder	Rape	Robbery	Assault
46	2006	473.6	5.8	31.6	150.0	292.0
37	1997	611.0	6.8	35.9	186.2	382.1
9	1969	328.7	7.3	18.5	148.4	154.5

2.3.2 Connecticut Department of Corrections Data

This data set contains individual information for 7.77 million people that have been processed by the justice system and recorded by the Connecticut Department of Corrections. Each individual is recorded along with their age, gender, race, offense, and sentence length, among other things [6].

I also created one-hot encoded versions of this data set in order to run regressions on the data. Because of the size of the data, the regression data sets are only random subsets of the larger data set.

Because there is so much to consider in what is found in the data set, I chose not to engineer more features as to avoid unneeded complexity.

	LATEST ADMISSION DATE	AGE	RACE	SENTENCE DAYS
4149965	06/22/2018	29	WHITE	731
7570861	01/31/2019	34	WHITE	548
3340323	02/02/2018	43	WHITE	548

We ought to disclose the sample sizes among different races that are found in the Connecticut Department of Justice data. The sample size for American Indians and Asians is much smaller than that of Whites, Hispanics, and Blacks, hence we may see some irregular outcomes in the analysis related to these racial groups.

```
Sample size for Blacks: 3287596
Sample size for Whites: 2393949
Sample size for Hispanic: 2039297
Sample size for American Indian: 21133
Sample size for Asian: 35660
```

3 Analysis and Visualization

3.1 Increased Rates

Here we can see the drastic increase of the amount of incarceration in the U.S. Something that is interesting to note is that jails are defined as places for people who have a sentence less than 1 year, or who are awaiting trial. So we see that at its peak in 2008, there were more people awaiting trial than there were being held in federal prison in 1991.

Here I have incarceration rates plotted against the violent crime rate in the United States. Near the beginning of the crime rate data we might assume some amount of inaccuracy, since the crime rate seems to be less than the incarceration rate, however there is an indisputable spike in crime rates in the 1980s and 1990s. Something to note is that the incarceration rate seems to lag behind about 20 years. Another thing to note is that there has been a strict decrease in violent crime (and in all crime) since the 90s, however we do not see the same decrease in the incarceration rate.

Some would argue that we see this decrease in crime because of the increase of incarceration. I would disagree. Consider the scales of the different curves we see on the chart. Both are in terms of the rate per 100,000, but the crime rate is more than 200 times higher than the incarceration rate in state prisons. Unless 0.5% of people who committed crime in the 90s were committing more than half of all crime, I would argue that there is some other cause for the decrease in crime rates.

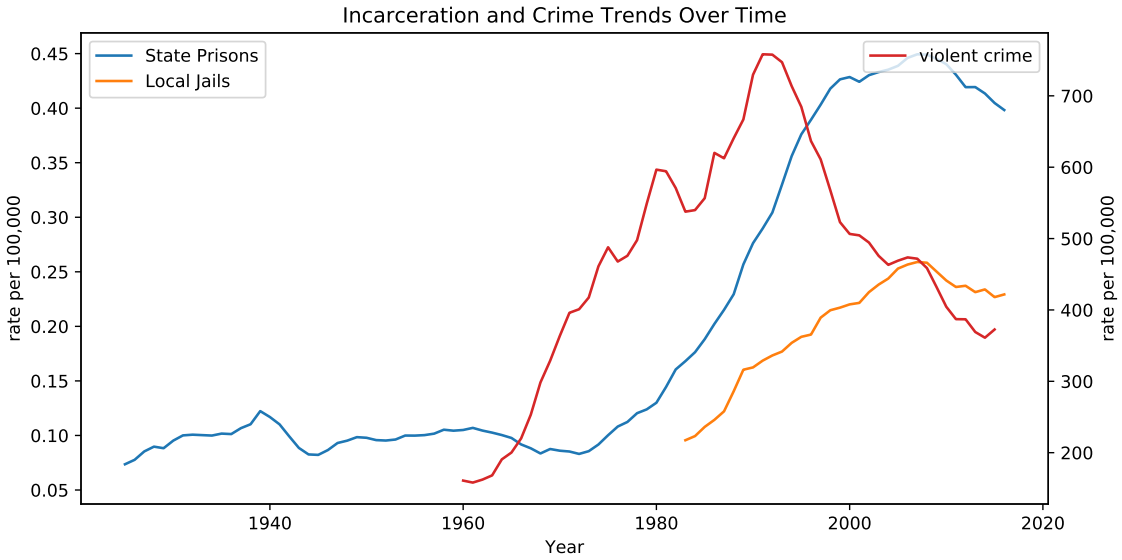

Figure 1: Incarceration rates over time

3.2 Artifacts of Prison Policy

One might expect sentence lengths to be distributed somewhat smoothly. However there are standard sentence lengths and mandatory minimum sentences that influence the distribution of sentence lenghts, making the distribution not smooth.

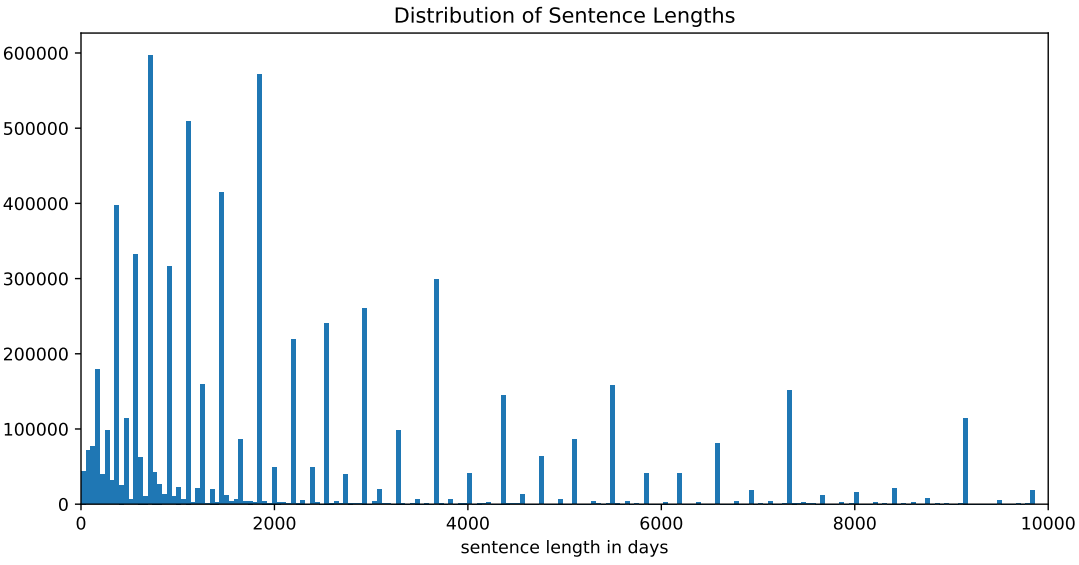

Figure 2: Distribution of sentences across all inmates in inmates.csv

3.3 Racial Disparities

In this section we will explore how the criminal justice system affects people of different races.

3.3.1 Incarceration Disparity

These simple histograms, which plot the incarceration rates of Whites, Blacks, Hispanics, and Asians among the 50 states, show the huge racial disparity in incarceration in America. By incarceration rate we mean: of 100,000 people of a given race, how many are incarcerated? A simple examination of the x-axis, which is rate of incarceration shows that the histogram for Whites barely overlaps with the histograms for Blacks and Hispanics. This means that in states with a low incarceration rate for Blacks, the Black incarceration rate is higher than the highest rates seen among Whites.

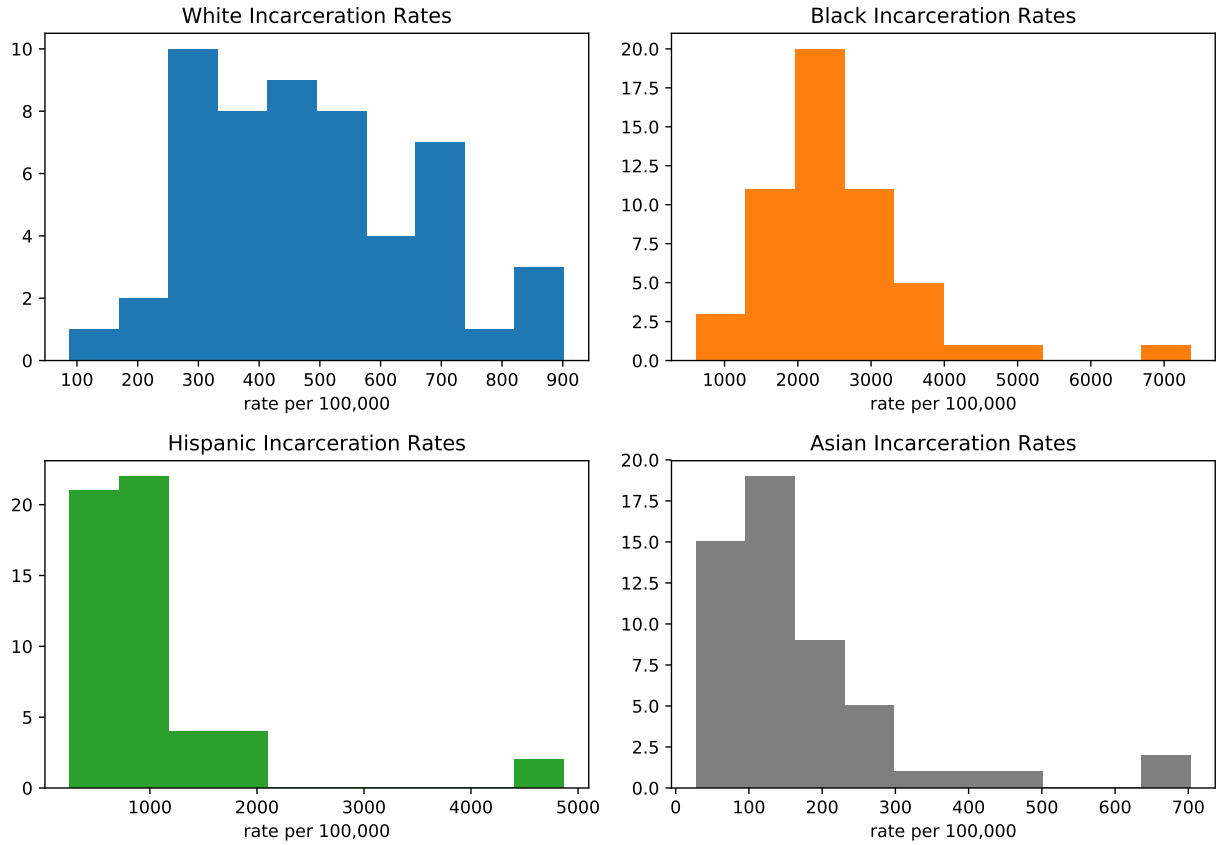

Figure 3: Distribution of states' mean incarceration rate, blocked by race

3.3.2 Distribution of Sentence Lengths

The next aspect we will explore in this section is the distribution of sentence lengths among people of different races. Here we are using the data of more than 7.7 million individuals processed by the criminal justice system. The medians are comparable on the scale at which sentence length is given. Asians and American Indians have the longest median sentence length, however the

standard deviation in their sentence lengths is much less than what is seen in the other racial groups. I would attribute this to the low sample size of Asians and American Indians.

```
Median sentence lengths:
 RACE
AMER IND    2192
ASIAN       2557
BLACK       1826
HISPANIC    1826
WHITE       1461
Name: SENTENCE DAYS, dtype: int64


Standard deviation of sentence lengths:
 RACE
AMER IND     3208.584729
ASIAN        5122.011265
BLACK       37306.146883
HISPANIC    35763.282330
WHITE       46882.705552
Name: SENTENCE DAYS, dtype: float64
```

This box plot is where we begin to see the disparity in the way people of different races are sentenced. While the first two quartiles of each racial group's sentence lengths are similar, we see that the top two quartiles vary widely.

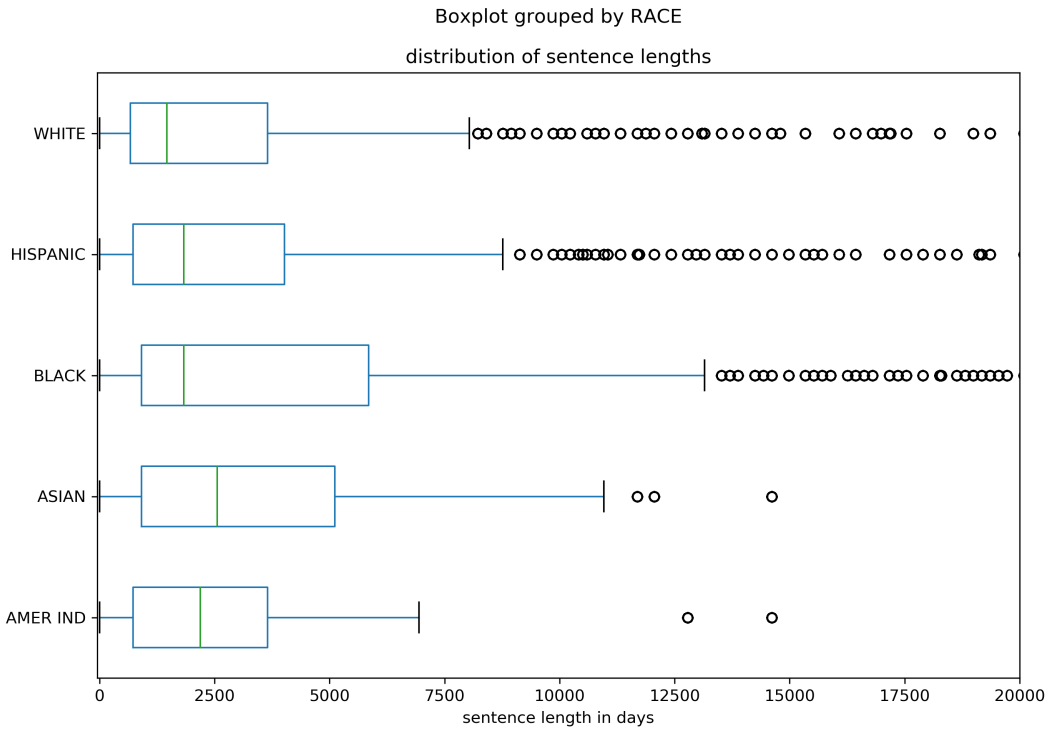

Figure 4: Boxplot distribution of sentence length for inmates in individuals.csv, blocked by race

The following histograms will assist us in understanding these distributions.

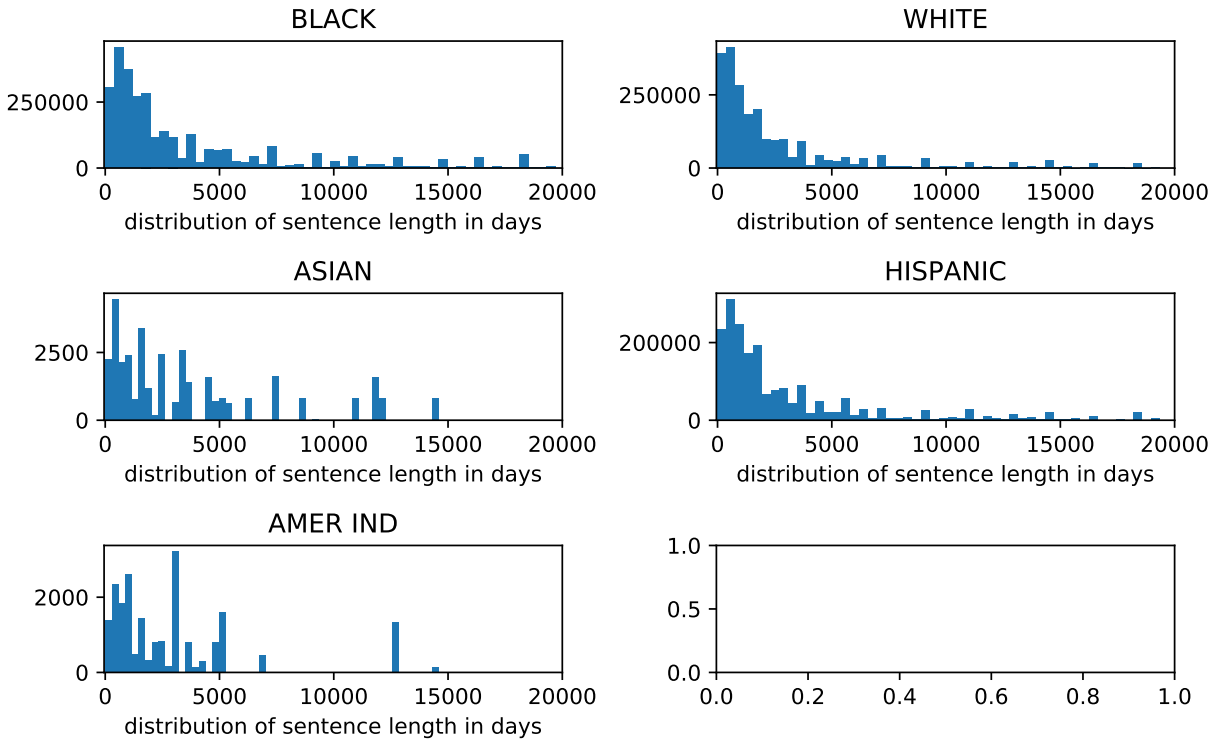

Figure 5: Histogram distribution of sentence length for inmates in individuals.csv, blocked by race

While it may appear that these sentences are distributed fairly equivalently, we can examine the kurtosis of the distribution to understand how much of the weight of the distribution is found in the extremities. Groups with high kurtosis have a higher probability of receiving a sentence that is far from the mean[8].

As we can see, Blacks and Hispanics have the greatest kurtosis, meaning they are much more likely to receive an extreme sentence length. It is difficult to explain the small kurtosis in American Indians and Asians because of the small sample size.

```
White Kurtosis: 54.01344161453346
Black Kurtosis: 86.11479232153843
Hispanic Kurtosis: 96.03075354970996
American Indian Kurtosis: 4.0085627062942955
Asian Kurtosis: 7.45297417968173
```

3.3.3 Predicting Sentence Length

As indicated by the regression below, attempting to predict sentence length based only on the race of the offender is not possible. A larger regression, contained in the appendix, shows that a simple linear regression does a decent job with an R^2 value of 0.52, when offenses are included as variables in the regression. It is inaccurate to claim that race alone is the biggest predictor of sentence length and unreasonable to claim race is more important than the offense when in comes to sentence length.

So how can we explore more deeply the relationship between race and sentence length?

```
Regression Results:

        Method: Least Squares
        R-squared value: 0.03601534317319244
        AIC: 19410096.12270456
        BIC: 19410165.748789284
```

3.3.4 Racial Disparity

In the following section we will examine only individuals that have committed the same crime. Then, among those individuals we will examine distribution of sentence length. This will show how people of different races are treated differently, when in comparable situations.

We will begin by examining the distribution of sentence lengths among offenders convicted of "criminal liability for another person".

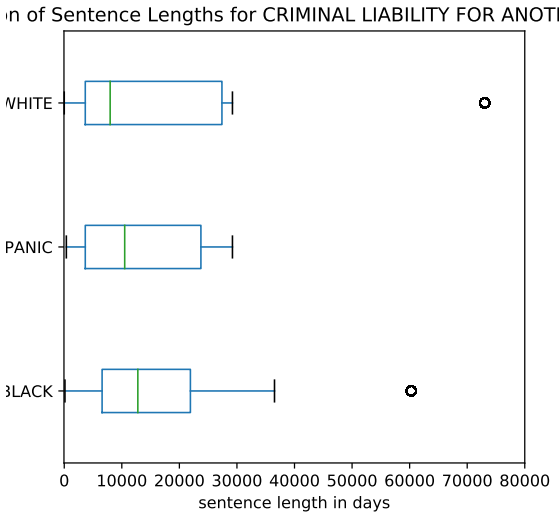

Figure 6: Distribution of sentence length of inmates convicted of criminal liability, blocked by race

Again we can notice that the distributions don't seem to be all that different. However, we can examine kurtosis to see if there is a significant difference between these distributions that is lost in a box plot.

```
White Kurtosis for
        CRIMINAL LIABILITY FOR ANOTHER PERSON: 12.421485976410926
Black Kurtosis for
        CRIMINAL LIABILITY FOR ANOTHER PERSON: 23.752981544095174
Hispanic Kurtosis for
        CRIMINAL LIABILITY FOR ANOTHER PERSON: 0.5345890525784038
```

Here it is made clear that Blacks and Hispanics are much more likely to receive an extreme sentence than other racial groups for the crime of criminal liability.

Following is a sample of a few offenses, the distributions of sentence lengths, and the kurtosis of those distributions.

Figure 7: Distribution of sentence length of inmates convicted of murder, blocked by race

```
White Kurtosis: 3.815929855470677
Black Kurtosis: 35.42506786574221
Hispanic Kurtosis: 23.661202825848072
```

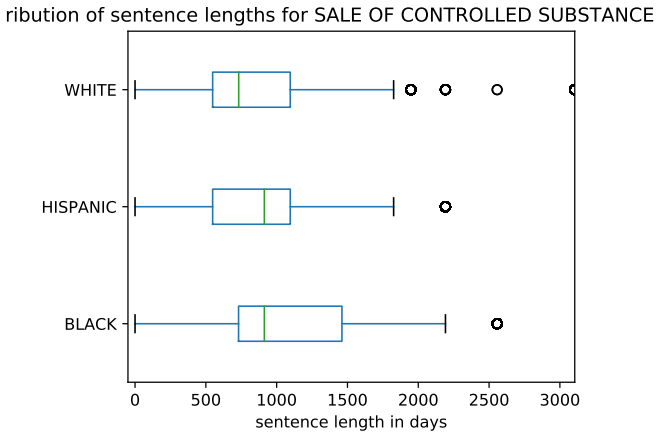

Figure 8: Distribution of sentence length of inmates convicted of sale of controlled substances, blocked by race

```
White Kurtosis: 3.2523247666826185
Black Kurtosis: 4.946745832035741
Hispanic Kurtosis: 40.024525483281515
```

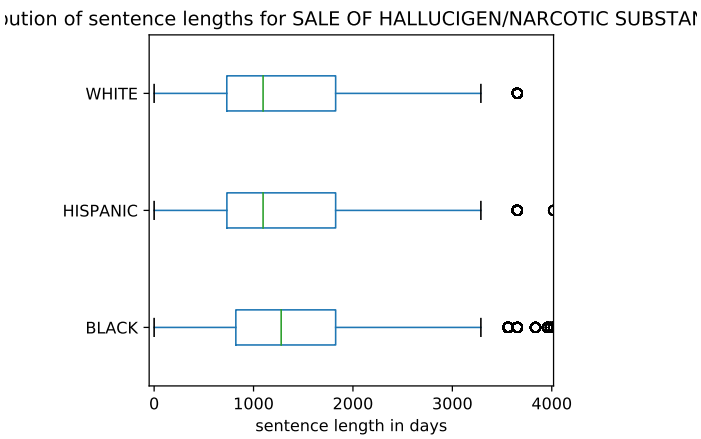

Figure 9: Distribution of sentence length of inmates convicted of sale of hallucinogen/narcotic substance, blocked by race

```
White Kurtosis: 26.642120595723625
Black Kurtosis: 123.56848013313629
Hispanic Kurtosis: 0.751155900894612
```

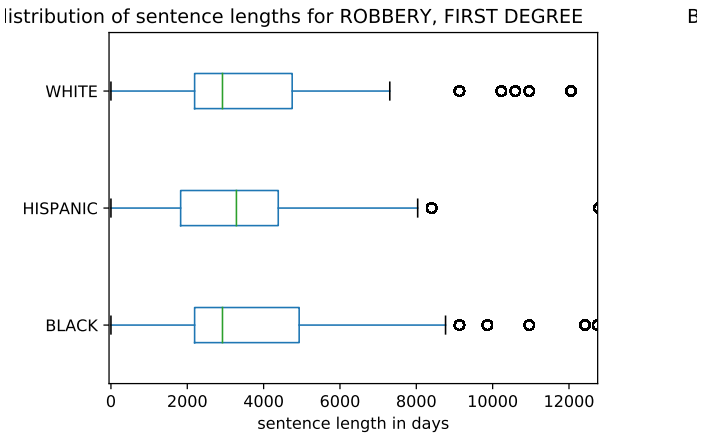

Figure 10: Distribution of sentence length of inmates convicted of robbery, blocked by race

```
White Kurtosis: 41.205491339214646
Black Kurtosis: 75.82664857484097
Hispanic Kurtosis: 99.72485672864623
```

Here we can see that in many cases Blacks and Hispanics are treated differently when it comes to sentencing and incarceration.

4 Conclusion

The American criminal justice system had been criticized by many. Some claim that it is an extension of slavery, others claim that is biased and broken, while still others claim that it is effective[7]. Dozens of organizations study the American prison system with the goal of reforming it, while private, for profit, prisons lobby to maintain the status quo. We have explored a few of the ways in which the American criminal justice system is biased, though understanding the biases is only the first step to formulating sound policy that can affect positive change. I hope the facts and opinions presented in this report become impetus to research further and understand the ways in which American society continues to disenfranchise Americans of color today.

References

[1] *Data Collection: Annual Survey of Jails (ASJ).* URL: https://www.bjs.gov/index.cfm?ty=dcdetail&iid=261. Accessed 18 November, 2019.

[2] *Data Collection: Census of Jail Inmates.* URL: https://www.bjs.gov/index.cfm?ty=dcdetail&iid=404. Accessed 18 November, 2019.

[3] *Data Collection: Census of State and Federal Adult Correctional Facilities (CSFACF).* URL: https://www.bjs.gov/index.cfm?ty=dcdetail&iid=255. Accessed 18 November, 2019.

[4] M. Lansing. *Sentenced Inmates in Correctional Facilities*. URL: https://catalog.data.gov/dataset/sentenced-inmates-in-correctional-facilities. Accessed 2 December, 2019.

[5] A. Liptak. "U.S. prison population dwarfs that of other nations." In: (Apr. 2008). URL: https://www.nytimes.com/2008/04/23/world/americas/23iht-23prison.12253738.html. Accessed 12 December, 2019.

[6] C. Robertson. "Crime Is Down, Yet U.S. Incarceration Rates Are Still Among the Highest in the World." In: (Apr. 2019). URL: https://www.nytimes.com/2008/04/23/world/americas/23iht-23prison.12253738.html. Accessed 12 December, 2019.

[7] *The Legacy Museum: From Enslavement to Mass Incarceration*. URL: eji.org.https://museumandmemorial.eji.org/museum. Accessed 12 December, 2019.

[8] Peter H. Westfall. "Kurtosis as Peakedness, 1905–2014. R.I.P." In: (). DOI: 10.1080/00031305.2014.917055. URL: https://www.tandfonline.com/doi/full/10.1080/00031305.2014.917055?scroll=top&needAccess=true.