ANALYSIS OF RACIAL BIAS IN NORTHPOINTE'S COMPAS ALGORITHM

A THESIS

SUBMITTED ON THE SIXTH DAY OF MAY 2019

TO THE DEPARTMENT OF MATHEMATICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE SCHOOL OF SCIENCE AND ENGINEERING

OF TULANE UNIVERSITY

FOR THE DEGREE

OF

MASTER OF SCIENCE

BY

Adrienne Brackey

Adrienne Brackey

APPROVED: Ricardo Cortez

Ricardo Cortez, Ph.D.
Advisor

Michelle Lacey, Ph.D.

Scott McKinley

Scott McKinley, Ph.D.

1 Abstract

This study will evaluate the effectiveness and validity of an algorithm that is widely used in the criminal justice system. In the year 2000, Northpointe introduced a risk assessment algorithm, called Correctional Offender Management Profiling for Alternative Sanctions, often referred to as COMPAS, in attempt to optimize the pretrial and sentencing stages of the process. COMPAS computes scores for defendants and these scores assist judges in deciding jail terms, sentencing, and probation. Some argue that such algorithms are more accurate and less biased than humans, but COMPAS has been highly criticized for perpetuating the systemic racial bias that is currently present in the criminal justice system.

This analysis will test the logical bases of the algorithm and its predictive ability. The results of this analysis cast doubt on the usefulness of the COMPAS algorithm if fairness is the goal of the criminal justice system.

A logistic regression analysis will show that the two most important variables for predicting recidivism are a defendant's age and the number of prior offenses committed. A linear model using only these two predictors will be shown to yield better predictive accuracy than COMPAS. The analysis will also show that COMPAS scores were unevenly distributed between different racial groups. In general, African Americans were more likely than Caucasians to be given higher scores regardless of their recidivism rates. Further, more frequently than Caucasians, African Americans who were assigned lower scores did not recidivate.

Keywords: recidivism, decile score, systemic racism, logistic regression analysis

2 Acknowledgements

Throughout the process of completing this thesis, I have received valuable support and guidance from many. I would first like to express my gratitude to Dr. Cortez, my advisor, for meeting with me regularly to keep me on schedule and for continually challenging me with new ideas for this project. Your guidance has helped me not only with this thesis, but also with completing my degree as a whole.

I would also like to thank Dr. Lacey for helping me to understand the mathematical methods that are central to this thesis and for answering any statistics related questions I had throughout the past two semesters.

I would like to thank these two professors, along with Dr. McKinley, for acting as my thesis committee and taking the time to read my thesis.

This thesis could not have been completed without the research done by Jeff Larson, et al. at *ProPublica*. Their study was the basis of my project and inspired me to explore this topic further.

Contents

List of Tables

List of Figures

3 Introduction

Algorithms are heavily relied upon in order to make certain decisions and tasks more convenient. In many cases, algorithms do exactly this. Algorithms have been introduced into the United States' criminal justice system by acting as risk assessment tools that aim to optimize the pretrial and sentencing stages of a defendant's time within the criminal justice system. They do so by predicting a defendant's future behavior to help judges make appropriate sentencing decisions.

Although there are a few different algorithms of this type, Northpointe's COMPAS algorithm is one of the most widely used today. One area where COMPAS is used is in Broward County, Florida where public records are readily accessible. Upon a defendant's arrest, he or she is prompted to complete a 137 question survey [1]. The first 30 questions are answered by the police officer and the remaining by the defendant. Some of the questions answered by the defendant are:

- How many of your friends have ever been arrested?
- If you lived with both parents and they separated, how old were you?
- How many of your friends do illegal drugs?
- How often do you barely have enough money to get by?
- What were your usual grades in school?

Whereas the majority of questions answered by the police officer are objective and concern the defendant's criminal history, one of the subjective questions answered by the police officer is "Is this person a suspected gang member?"

Based on the responses, the defendant is assigned three different scores each ranging from one to ten. Scores of one to four are considered low scores, five to seven considered medium, and eight to ten considered high. The three different categories that each defendant is scored for are "Risk of Recidivism," "Risk of Violent Recidivism," and "Risk of Failure to Appear." According to Northpointe's practitioners guide, *recidivism* is defined as the act of recommitting a crime within the two years that follow the assignment of a COMPAS score. This definition will be used throughout the entirety of the paper. In concordance with Northpointe's definition, the United States Sentencing Commission found that the majority of defendants who recommit crimes do so within the two years after their served jail time.

The Risk of Recidivism score, which is also referred to as the *decile score*, indicates to a judge the likelihood of a defendant committing a nonviolent crime in the future. Theft, fraud, prostitution, and drug related crimes are all examples of nonviolent crimes. Similarly, the Risk of Violent Recidivism indicates the likelihood of committing a violent crime in the future. Violent crimes are those such as assault, rape, and manslaughter. The Risk of Failure to Appear indicates the likelihood of a defendant not showing up to trial in the event that he or she is not required to stay in jail between the time of arrest and his or her trial. A high score corresponds to having a high likelihood, whereas a low score corresponds to having a low likelihood. For example, having a decile score of one means that the defendant, based on the survey responses, is not predicted to commit another crime. The nonviolent decile score is the only score utilized in the regression analyses throughout this study.

While Northpointe asserts that their software is fair in the treatment of all racial groups, an analysis by Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin published in *ProPublica* proves otherwise. Prompted by Northpointe's refusal to release the technical workings of the algorithm, the researchers at ProPublica created a linear model in order to understand what variables are most correlated with COMPAS scores. Essentially, they hoped to gain information about how Northpointe uses a defendant's demographics to assign a decile score.

In the following section, the dataset curated by the researchers at *ProPublica* will be introduced. Section 5 will derive the methods used to create a logistic regression model. In Section 6.1, the accuracy of COMPAS decile scores will be calculated. In Section 6.2, the logistic regression model created by *ProPublica* will be reproduced. In Section 6.3, this study will attempt to produce a more fair model that predicts recidivism with comparable or better predictive accuracy than COMPAS and without the introduction of racial bias. To do so, several logistic regression models will be created. The ultimate goal in creating these linear models will be to question the impact of certain factors in predicting recidivism and to attempt to hinder the introduction of racial bias throughout the process.

Finally, in Section 7, the supposed inequalities will be investigated and the idea of fairness in the context of algorithms will be debated. Also, the social, economic, and political injustices that exist and perpetuate the systemic racism seen in the criminal justice system today will be discussed. The research that follows will show that COMPAS is not en entirely reliable algorithm and thus its importance in the courtroom should be questioned.

4 The Data

4.1 Obtaining the Data

The researchers at *ProPublica* gathered all arrest records from Broward County between the years of 2013 and 2016. Of these defendants, the researchers were also able to gather the COMPAS scores assigned to 18,000 people who were arrested during 2013 and 2014. From the 80,000 total arrest records, they matched COMPAS scores to defendants using first and last name and date of birth. Since all arrest records are entered manually, there were a small number of discrepancies with COMPAS scores being assigned to the incorrect people. Of a random sample of 400 people, ProPublica found an error rate of 3.75% [2].

This resulted in a comprehensive dataset that contained

- A defendant's demographics (age, race, and gender),
- the number of crimes he or she committed prior to the current one,
- the current crime for which he or she was assigned the given COMPAS score,
- and any other crimes that occurred within the two years following when the given score was assigned.

Using all of the above, the researchers could determine whether the crime for which each defendant was scored was either the first and only crime they committed or one of the many they may have committed. This was crucial in understanding offenders' patterns in relation to their COMPAS scores.

4.2 Overview of the Dataset

A sample subset of the raw dataset can be found in the table below.

Age	Crime Cat.	Race	Score Cat.	Sex	Priors	Decile Sc.	Two Yr. Recid
24	M	African-Amer.	Medium	Female	2	6	0
32	F	Caucasian	Low	Male	1	1	1
22	F	Hispanic	Medium	Female	0	6	0
27	F	Other	Low	Male	1	4	0
27	M	Caucasian	Low	Male	3	3	0
63	F	Hispanic	Low	Male	1	1	0
27	F	African-Amer.	Medium	Male	2	7	1
29	F	Caucasian	High	Male	9	8	1
24	F	African-Amer.	Medium	Male	1	6	1
49	F	Caucasian	Low	Male	0	1	0

4.2.1 The Variables

Each row in the dataset corresponds to a defendant and each column in the dataset is a variable that contains the defendant's demographic information.

The column "Crime Cat." describes the type of crime for which the defendant is being scored. This column may contain the letter "M," "F," or "O." An "O" refers to an ordinary or petty offense, such as a traffic offense, which are not typically punished with jail time. An "M" refers to a misdemeanor charge. Misdemeanors may result in jail time, although sentences are typically shorter than felonies, that is served in local jails. For most misdemeanor charges, defendants who cannot afford a private lawyer are not granted the right to a public defender. An "F" refers to a felony charge. Felonies result in longer jail sentences served in state or federal prisons. Defendants facing felony charges are granted the right to a public defender in the case that they cannot afford a private lawyer.

The "Race" column indicates the race of the defendant. The six possible race categories in the Broward County database are African American, Asian, Caucasian, Hispanic, Native American, and Other.

The "Score Cat." column classifies the category that the defendant's decile score, which is recorded in the "Decile Sc." column, belongs to. The decile score may be classified as either "Low," "Medium," or "High." Scores ranging from one to four are considered low, five to seven considered medium, and eight to ten considered high.

The column "Priors" corresponds to the number of crimes the defendant has committed prior to the current one for which the decile score is assigned. This variable will often be referred to as "prior counts." For this dataset, the minimum number of prior counts across all defendants is zero and the maximum is 38.

The column "Two Yr. Recid" indicates whether the defendant is a recidivist. The column contains either a "1" or a "0." A "0" in this column indicates that the defendant is not a recidivist, meaning that the defendant did not recidivate within the two years following his or her COMPAS score assignment. A "1" in this column indicates that the defendant did recidivate within the two years following the COMPAS score assignment.

4.3 Filtering the Data

Before running a regression analysis on the dataset, *ProPublica* found that not all rows were applicable to the analysis and some rows were missing values. In order to create a relevant and accurate model, they filtered the dataset in the following ways.

In Broward, COMPAS scores are most commonly used to decide whether a defendant should be released between the time of arrest and the time of trial and to decide what type of sentence a defendant should receive. Hence, any COMPAS score that was utilized after the pretrial and sentencing stages of a defendant's process was removed from the overall dataset. For example, if the score was used to make parole or probation decisions, it was removed from the dataset. This narrowed the sample from 18,000 to 11,757 defendants.

Furthermore, if any row had an "O" in the "Crime Cat." column, it was removed from the dataset. This type of charge refers to a petty crime and does not result in a jail sentence.

Due to the definition of recidivism given by Northpointe, *ProPublica* only kept those rows that contained defendants who either

1. recidivated within the two year period following their score assignment or

2. surpassed the two year period without recidivating.

Those who failed to appear at their court date were not considered recidivists, so these rows were removed from the dataset as well.

Of the remaining 7,214 data points, 1,042 defendants had an elapsed time of more than 30 days in between the crime for which he or she was assigned the given COMPAS score and the date for which they were arrested for the supposed crime. As a result, the researchers at *ProPublica* assumed that the COMPAS score in the dataset corresponded to a different crime and deleted these points from the dataset.

This filtering process resulted in the final dataset which is used in all of the following analyses. It contains 6,172 defendants' information.

4.4 Demographics of the Filtered Data

After filtering the data, the data has the following demographic summary:

Table 1: Race Category Breakdown

Caucasian	African-American	Asian	Hispanic	Native American	Other
2103	3175	31	509	11	343

Table 2: Gender Category Breakdown

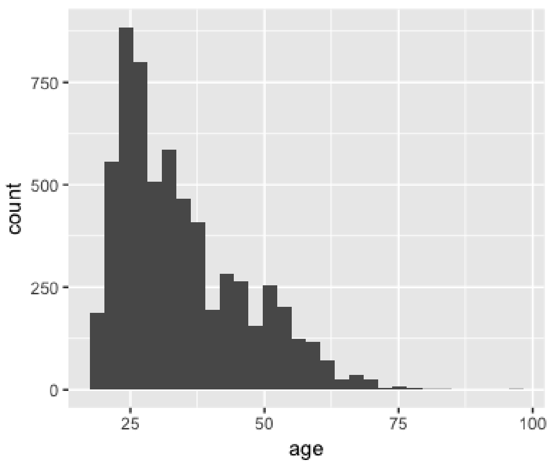
Male	Female
4997	1175

Table 3: Score Category Breakdown

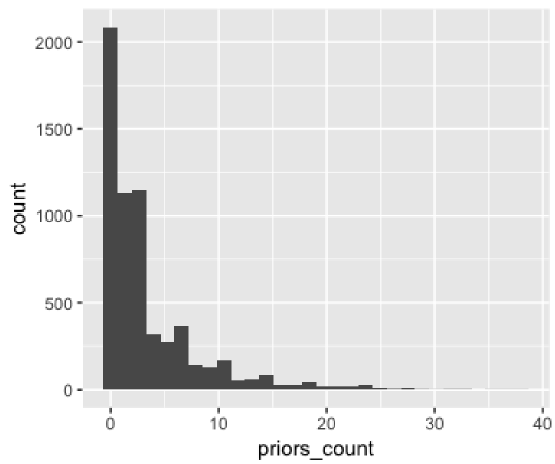
Low Score	Medium Score	High Score
3421	1607	1144

Table 4: Age Category Breakdown

Less than 25	25 - 45	Greater than 45
1347	3532	1293



(a) Distribution of Ages



(b) Distribution of Prior Counts

Figure 1: Relevant Demographic Distributions

5 Methodology

As stated previously, the goal of this paper is to compute the predictive accuracy of COMPAS, reproduce *ProPublica's* linear model, and create new linear models to determine what parameters are good predictors of recidivism. Combined, these efforts help the reader understand the shortcomings of COMPAS and the possible improvements to be made.

In order to create these models, one must first be familiar with regression analysis techniques.

For all of the models produced in this paper, the response variable, which is the variable to be predicted, is a binary value. With the exception of *ProPublica's* model, the response variable for the models created is the two year recidivism factor. The variable is either a "0" or a "1;" a defendant is either a recidivist or a non-recidivist. When the response variable is binary, logistic regression models are often the appropriate regression models to be used for the analysis. A logistic regression model exhibits the relationship between the response variable, two year recidivism, and several predictor variables.

Logistic regression is particularly useful because the output is more informative than other regression models. The output not only indicates the significance of each predictor variable, but it also indicates the direction of the significance of each predictor. Another benefit of using logistic regression is that the response and predictor variables are not required to have a normal distribution.

An overview of logistic regression is given below.

5.1 Logistic Regression Using Newton's Method

Given a set of predictive values, $\vec{x_k}$'s, and a response variable, y_k, for each defendant k, the goal of logistic regression is to solve the linear combination in Equation 1 for the regression coefficients, $\vec{\beta}$, in order to determine how statistically significant each component of the vector $\vec{\beta}$ is in predicting our response variable. For each defendant $k = 1, \ldots, n$, l_k can be understood as the *logarithm of the odds*, commonly referred to as the *log odds*.

$$l_k = \vec{\beta} \cdot \vec{x_k} \tag{1}$$

The vector $\vec{x_k}$ contains the values of the m parameters used in the model that correspond to defendant k. For all defendants, $x_{k,0} = 1$ and is the coefficient of β_0.

$$\vec{x_k} = \begin{bmatrix} x_{k,0} & x_{k,1} & \cdots & x_{k,m-1} \end{bmatrix}$$

The following representation will be useful for understanding our dataset.

$$\begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ \vdots \\ l_n \end{bmatrix} = \beta_0 + \beta_1 \begin{bmatrix} x_{1,1} \\ x_{2,1} \\ x_{3,1} \\ \vdots \\ x_{n,1} \end{bmatrix} + \beta_2 \begin{bmatrix} x_{1,2} \\ x_{2,2} \\ x_{3,2} \\ \vdots \\ x_{n,2} \end{bmatrix} + \ldots + \beta_m \begin{bmatrix} x_{1,m} \\ x_{2,m} \\ x_{3,m} \\ \vdots \\ x_{n,m} \end{bmatrix}$$

In the models created in Section 6.3, $\vec{l}$ gives the logarithm of the odds of recidivating based on the given predictor variables. If, for example, β_1 is the regression coefficient representing prior counts, the corresponding vector $\vec{x}$ contains the number of prior counts for each defendant in the dataset. As opposed to the response variable $\vec{y}$, the values of $\vec{x_k}$ can be either binary or continuous real values.

The distinction between the odds of recidivating and the probability of recidivating must be made. Odds and probabilities differ in the following ways.

$$odds(recidivating) = \frac{\#ofrecidivists}{\#ofnon - recidivists} \tag{2}$$

$$probability(recidivating) = \frac{\#ofrecidivists}{\#ofrecidivists + \#ofnon - recidivists} \tag{3}$$

In order to convert the log odds to probabilities, the *logistic function* is used. The logistic function maintains the smoothness of the log odds but ensures that the output is within $(0, 1)$. The logistic function $p(\vec{x}_j, \beta_j)$ is defined in Equation 4.

$$p_k = \frac{1}{1 + e^{-l_k}} \tag{4}$$

The next step in logistic regression is to maximize the likelihood that each point in the dataset predicts the response variable correctly. This approach is called *maximum likelihood estimation* and involves estimating the values of β that maximize the *likelihood function*. The likelihood function expresses the probability of outputting the observed value of y_k given the values of the parameters used in the model for each defendant. The likelihood function for each defendant is calculated using Equation 5.

$$L_k = p_k^{y_k}(1-p_k)^{1-y_k} \tag{5}$$

Once the likelihood function is obtained for each defendant, the product of the likelihoods is taken over all defendants.

$$L = \prod_{k=1}^{n} p_k^{y_k}(1-p_k)^{1-y_k} \tag{6}$$

Multiplying n likelihoods, which are all between 0 and 1, results in a likelihood function of degree 10^{-n}. For large samples (in our case, $n = 6172$), this is computationally expensive. Thus, the log of the likelihood function is taken, converting the product to a summation. The *log likelihood function* is

$$LL = \log\big[\prod_{k=1}^{n} p_k^{y_k}(1-p_k)^{1-y_k}\big] = \sum_{k=1}^{n} y_k \log\big(p_k\big) + (1-y_k)\log\big(1-p_k\big) \tag{7}$$

In order to maximize Equation 7, the gradient is taken with respect to $\vec{\beta}$ and set equal to zero. The log likelihood function is strictly concave, ensuring that there will only be one solution.

$$\begin{aligned}
&\sum_{k=1}^{n} \nabla\big[y_k \log\big(p_k\big)\big] + \nabla\big[(1-y_k)\log\big(1-p_k\big)\big] = \\
&\sum_{k=1}^{n} y_k \frac{1}{p_k}\nabla p_k + (1-y_k)\frac{1}{1-p_k}\nabla(1-p_k) = \\
&\sum_{k=1}^{n} y_k \frac{1}{p_k}\nabla p_k - (1-y_k)\frac{1}{1-p_k}\nabla p_k = \\
&\sum_{k=1}^{n} \frac{y_k \nabla p_k}{p_k(1-p_k)} - \frac{\nabla p_k}{1-p_k}
\end{aligned} \tag{8}$$

The gradient of p_k is

$$\begin{aligned}\nabla p_k =&\nabla\left[\frac{1}{1+e^{-l_k}}\right] = \frac{\vec{x}_k e^{-l_k}}{(1+e^{-l_k})^2} = \frac{\vec{x}_k e^{-l_k}}{(1+e^{-l_k})}p_k \\ =& \frac{\vec{x}_k}{1+e^{l_k}}p_k = \vec{x}_k\big(1-p_k\big)p_k\end{aligned} \tag{9}$$

This implies that

$$\frac{\nabla p_k}{p_k(1-p_k)} = \vec{x}_k \tag{10}$$

So Equation 8 becomes

$$\sum_{k=1}^{n}(y_k - p_k)\vec{x}_k \tag{11}$$

Maximizing Equation 11 equates to

$$\sum_{k=1}^{n}(y_k - p_k)\vec{x}_k = \vec{0} \tag{12}$$

There is clearly not a closed form solution of Equation 12 and thus, a root finding method is required. Newton's Method is used in order to solve for each component of the vector $\vec{\beta}$'s. The function of interest is

$$\vec{V}(\vec{\beta}) = \sum_k (y_k - p_k)\vec{x}_k$$

The n, l entry of the Jacobian is

$$J_{n,l} = \frac{\partial V_n}{\partial \beta_l} = -\sum_k p_k(1-p_k)x_{k,n}x_{k,l}$$

and Newton's method for this equation is

$$\vec{\beta}^{j+1} = \vec{\beta}^{j} - J^{-1}\vec{V}(\vec{\beta})$$

until converging to the values of β that maximize the log likelihood.

Newton's method requires the solution of a linear system with matrix J at each step j which is seemingly more expensive than other root finding algorithms, but is made up for by having a higher order of convergence (order 2).

Due to the fact that Newton's method is an iterative method, an initial guess must be made for the $\beta's$. Once the values of β are found that maximize the log likelihood, the output can be predicted by using the logistic function to convert the log odds to probabilities, setting a threshold for these probabilities, and converting these probabilities to equal either 0 or 1 depending on the threshold. A prediction is correct if the prediction is equal to the observed outcome y_k.

This process can be manually implemented in MATLAB or we can use the glm() command in R to run a logistic regression model. This type of regression model is used continually throughout Section 6.

6 Results

The accuracy and fairness of the COMPAS algorithm has been called into question by many. Although some claim that algorithms like COMPAS make courtroom decisions more efficient and less biased, it is necessary that we seriously investigate the validity of the software. After all, the pretrial and sentencing decisions that are justified by COMPAS scores determine the fate of a defendant's future.

In this section, the accuracy of COMPAS is calculated, the logistic regression model created by *ProPublica* is recreated, and some original linear models are introduced.

6.1 Accuracy of COMPAS

In order to calculate the accuracy of COMPAS scores in predicting recidivism, consider the truth table in Table 5 that shows the breakdown of score categories as assigned by COMPAS and the corresponding recidivism rates. Scores are partitioned into two groups by grouping both medium and high scores and relabeling them as high scores. The rows represent these newly labeled score categories and the columns represent two year recidivists. Again, the column denoted "0" corresponds to those defendants who do not recidivate and "1" corresponds to those who do.

Throughout this study, the use of the term accuracy refers to the number of correctly identified predictions and is calculated by summing up the number of correct predictions and dividing by the sum of both correct and incorrect predictions. Predictions that are correct are those where

1. defendants are assigned low scores and do not recidivate in the two years following and
2. defendants are assigned high scores and do recidivate in the two years following.

Table 5: COMPAS Truth Table

	0	1
LowScore	2345	1076
HighScore	1018	1733

By the above definition, COMPAS has an accuracy of 66.07%. COMPAS scores correctly predict future recidivism only two-thirds of the time.

6.2 ProPublica's Model

Since Northpointe will not disclose the details of the algorithm used to compute decile scores, *ProPublica*'s linear model attempts to mirror the suspected inner workings of the COMPAS algorithm to understand what variables potentially influence how scores are assigned. The model measures how well gender, age category (less than 25, 25-45, and greater than 45), race category (African American, Asian, Caucasian, Hispanic, Native American, Other), priors count, crime category (M or F), and two year recidivism (0 or 1) predict score category (low or high). This model yields the summary found in Table 6.

```
model <- glm(score_factor ~ gender_factor + age_factor + race_factor +
    priors_count + crime_factor + two_year_recid, family="binomial",
    data=df)
```

Table 6: ProPublica Logistic Regression Model Summary

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.5255	0.0785	-19.43	< 2e-16***
gender_factorFemale	0.2213	0.0795	2.78	0.0054**
age_factorGreater than 45	-1.3556	0.0991	-13.68	< 2e-16***
age_factorLess than 25	1.3084	0.0759	17.23	< 2e-16***
race_factorAfrican-American	0.4772	0.0693	6.88	5.93e-12***
race_factorAsian	-0.2544	0.4782	-0.53	0.5947
race_factorHispanic	-0.4284	0.1281	-3.34	0.0008***
race_factorNative American	1.3942	0.7661	1.82	0.0688
race_factorOther	-0.8263	0.1621	-5.10	3.43e-07***
priors_count	0.2689	0.0111	24.22	< 2e-16***
crime_factorM	-0.3112	0.0665	-4.68	2.91e-06***
two_year_recid	0.6859	0.0640	10.71	< 2e-16***

The (Intercept) term in Table 6 is the value of the β_0 in Equation 1. It describes the log odds of recidivating when all of the x_i's are zero, i.e. when a defendant is a Male, is between the ages of 25 and 45, is Caucasian, has zero prior counts, has a felony charge, and has not recidivated.

The "Estimate" column gives the value of each regression coefficient. These coefficients may be thought of as weights, in a sense. If the predictor variable is continuous, the coefficient explains how much the log odds change with every one unit increase in the parameter. For example, for every additional prior count, the log odds of a defendant recidivating increase by 0.2689. If the predictor variable is discrete, the coefficient explains how the log odds change when going from the given discrete value to the another. For example, going from male to female results in an increase of the log odds by .2213 for this particular model. Similarly, the log odds of recidivating increase by .4772 if going from Caucasian to African American.

The Pr(>|z|) column gives the p value of the coefficient. Typically, a p value less than .05 indicates that the null hypothesis may be rejected and that the coefficient is statistically significant. A small p value corresponds to having a large z value. Z values are calculated by dividing the coefficient estimate by the standard error. Notice that regression coefficient estimates that are larger in magnitude also have z values that are larger in magnitude.

According to this model, gender factor, age factor, race factor, prior counts, crime factor, and two year recidivism are significant predictors of score factor. Among the races, African American, Hispanic, and Other are the most predictive of score factor.

The truth table for *ProPublica's* is found below.

Table 7: Truth Table for ProPublica's Model

	LowScore	HighScore
FALSE	2744	854
TRUE	677	1897

Score factor is correctly predicted 73.19% of the time using only these six factors according to Table 7. Although this model does not correctly predict score factor for every defendant, this model ultimately implies that gender factor, age factor, race factor, prior counts, crime factor, and two year recidivism may be some of the most significant variables used in the COMPAS algorithm to compute decile scores.

6.3 Our Logistic Regression Models

Whereas *ProPublica's* model attempts to understand what variables predict score factor best, the logistic regression models in this section aim to understand what variables best predict two year recidivism. Clearly, COMPAS scores are not always indicative of a defendants' future behaviors. In fact, as it was shown above in Table 5, one-third of the time COMPAS misidentifies the score factor as it relates to recidivism. Thus, these models attempt to extract what variables are truly indicative of future recidivism.

The following models do not use any information other than a defendant's demographics to predict recidivism rates. Since the details of COMPAS are not public, we assume no knowledge of score factors. Additionally, because ProPublica reassigns score factors (by combining medium and high Scores and relabeling them both as high Scores) in the creation of their model, leaving this factor out can help improve the true accuracy of the results.

The first model is similar in structure to the model *ProPublica* created, but replaces score factor with the two year recidivism factor, and uses a subset of the same predictor variables.

```
model5 <- glm(two_year_recid ~ gender_factor + age_factor + race_factor
    + priors_count + crime_factor , family="binomial", data=df)
```

Table 8: Model5 Summary

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.6083	0.0645	-9.43	$<$ 2e-16***
gender_factorFemale	-0.3486	0.0719	-4.85	1.22e-06***
age_factorGreater than 45	-0.6682	0.0761	-8.78	$<$ 2e-16***
age_factorLess than 25	0.7335	0.0689	10.64	$<$ 2e-16***
race_factorAfrican-American	0.0960	0.0627	1.53	0.1258
race_factorAsian	-0.5508	0.4278	-1.29	0.1979
race_factorHispanic	-0.1704	0.1088	-1.57	0.1174
race_factorNative American	-0.2707	0.6507	-0.42	0.6773
race_factorOther	-0.1549	0.1277	-1.21	0.2251
priors_count	0.1655	0.0081	20.52	$<$ 2e-16***
crime_factorM	-0.2187	0.0588	-3.72	0.0002***

Table 9: Truth Table for Model5

	0	1
FALSE	2564	1248
TRUE	799	1561

This model yields 66.83% accuracy, meaning that these five factors correctly predict if a defendant recidivates about two-thirds of the time. This is slightly better than the accuracy of COMPAS score factors (66.07%). The summary of the model also shows that none of the race categories are significant predictors of recidivism. This is the central aspect of this study: race is not indicative of one's future criminal behavior but COMPAS scores certain races differently as if to imply that race is a factor in future criminal behavior. This will be explored more in depth in the next section.

Since race is not a significant predictor for recidivism, the next model excludes the race category but maintains the rest of the factors. At minimum, this model should have comparable accuracy as the one above.

```
model4 <- glm(two_year_recid ~ gender_factor + age_factor + priors_
    count + crime_factor , family="binomial", data=df)
```

Table 10: Model4 Summary

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.5963	0.0521	-11.43	$<$ 2e-16***
gender_factorFemale	-0.3409	0.0717	-4.76	1.98e-06***
age_factorGreater than 45	-0.6854	0.0754	-9.10	$<$ 2e-16***
age_factorLess than 25	0.7486	0.0685	10.92	$<$ 2e-16***
priors_count	0.1703	0.0079	21.47	$<$ 2e-16***
crime_factorM	-0.2265	0.0586	-3.87	0.000111***

Table 11: Truth Table for Model4

	0	1
FALSE	2568	1239
TRUE	795	1570

This results in accuracy of 67.04%. From the summary above, the predictor with the lowest absolute value of z value is crime factor. Since crime factor is now the least significant predictor of recidivism, the next model does not include crime factor as a predictor.

```
model3 <- glm(two_year_recid ~ gender_factor + age_factor + priors_count, family="binomial", data=df)
```

Table 12: Model3 Summary

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.6906	0.0463	-14.91	<2e-16***
gender_factorFemale	-0.3499	0.0715	-4.89	1e-06***
age_factorGreater than 45	-0.6970	0.0752	-9.26	<2e-16***
age_factorLess than 25	0.7747	0.0681	11.37	<2e-16***
priors_count	0.1747	0.0079	22.16	<2e-16***

Table 13: Truth Table for Model3

	0	1
FALSE	2461	1127
TRUE	902	1682

The accuracy of the model increases to 67.13%. It is now clear that age factor and priors count are the most significant predictors of recidivism. Consider the following model that uses only these two predictors.

```
model2 <- glm(two_year_recid ~ age_factor + priors_count, family="
    binomial", data=df)
```

Table 14: Model2 Summary

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.7711	0.0435	-17.71	<2e-16***
age_factorGreater than 45	-0.6903	0.0751	-9.19	<2e-16***
age_factorLess than 25	0.7851	0.0680	11.55	<2e-16***
priors_count	0.1788	0.0079	22.70	<2e-16***

Table 15: Truth Table for Model2

	0	1
FALSE	2393	1176
TRUE	970	1633

The accuracy of the model drops to 65.23%. This is explained by the fact that the model was reduced by leaving out the gender factor, which although is not as significant as age and priors count, is clearly still quite predictive. Rather than looking only at the age categories, we can see how using age as a continuous variable affects our results.

```
model2continuous <- glm(two_year_recid ~ age + priors_count, family="
    binomial", data=df)
```

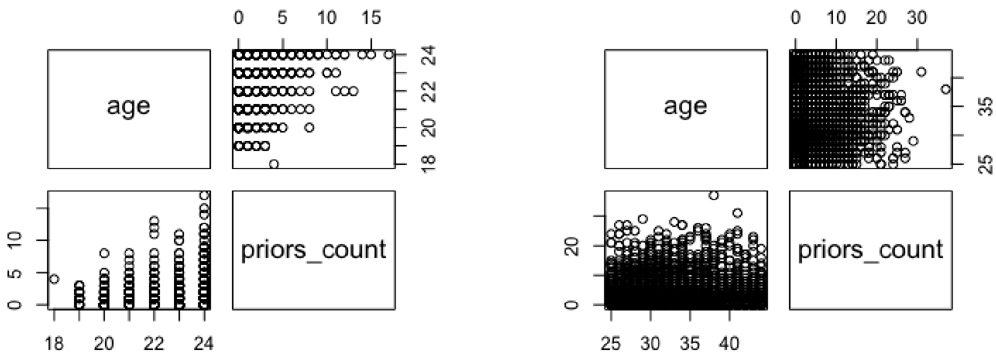
Table 16: Model2Continuous Summary

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.8788	0.0890	9.87	<2e-16***
age	-0.0470	0.0026	-17.86	<2e-16***
priors_count	0.1748	0.0077	22.57	<2e-16***

In fact, this model produces 67.87% accuracy of predicting recidivism, trumping COMPAS's accuracy of 66.07% and improving the accuracy from the previous model by 2.5%. Using continuous data such as one's age rather than discrete data is almost always preferred because it allows a better visual representation of the data and helps to account for any variability in the data. To see why continuous age yields more accurate results, consider the pairs plots of age versus prior counts for each age category found in Figures 2 and 3.

Table 17: Truth Table for Model2Continuous

	0	1
FALSE	2718	1338
TRUE	645	1471



(a) Age Category = Less than 25 (b) Age Category = 25 - 45

Figure 2: Age vs. Prior Counts Pairs Plot, Age Category: Less than 25, 25 - 45

For young defendants, there is less of a likelihood of having a large number of prior counts simply due to the fact that they have had less time to commit crimes than someone who is older. The majority of defendants under the age of 25 have less than five prior counts and the average number of prior counts for this age group is only one (1.35). As seen in the pairs plot, there is one eighteen year old defendant who has had four prior counts. Data points such as these are outliers and can affect the accuracy when they are grouped according to age category rather than by each individual age.

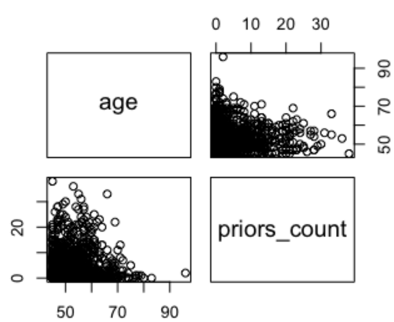


Figure 3: Age vs. Prior Counts Pairs Plot, Age Category: Greater than 25

For those defendants between the ages of 25 and 45, who make up 57% of the total population, the amount of prior counts has great variability. The average number of prior counts for this age group is nearly four (3.77). However, it is clear that several defendants in this category have more than four prior counts and that these defendants vary in age.

For defendants above the age of 45, the number of prior counts starkly decreases as age increases. Although there is a higher likelihood of having a large number of prior counts as you age, 34% of defendants in this age group have had no prior counts. Again, the 96 year old with 2 prior counts is yet another outlier in the set.

All of these outliers within each age category may be accounted for by using continuous age as a predictor. Rather than creating noise in the data for the entire age category, these outliers now only affect the individual age groups.

6.3.1 Likelihood Ratio Tests for our Logistic Regression Models

Although predictive accuracy is a good measure of how well each model performs, a likelihood ratio test may be used to compare any two models. The only requirement is that one of the models be more complex and the other must be more simple, such that the more complex model only differs from the simple model by at least one added parameter. This test computes the log likelihood of each model. As mentioned in previous sections, the purpose of a logistic regression model is to maximize the log likelihood. Hence, the model that has the largest log likelihood is the optimal model to use. In summary, the likelihood ratio test reveals which of the two models fits the dataset significantly better. Likelihood ratio tests comparing each successive model produce the output in Tables 18-20.

Table 18: Likelihood Ratio Test Summary for Model5 (1) vs. Model4 (2)

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	11	-3788.09			
2	6	-3794.03	-5	11.86	0.0367

For each successive model, the model is reduced by one predictor variable. As the number of

Table 19: Likelihood Ratio Test Summary for Model4 (1) vs. Model3 (2)

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	6	-3794.03			
2	5	-3801.51	-1	14.96	0.0001

Table 20: Likelihood Ratio Test Summary for Model3 (1) vs. Model2 (2)

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	5	-3801.51			
2	4	-3813.66	-1	24.31	0.0000

predictors decreases, the log likelihood also decreases. The log likelihood of "model5" is -3788.09, whereas the log likelihood of "model2" is -3813.66. By definition of the test, "model5" is a better fit for our dataset than "model2."

Moreover, it was shown earlier that "model2continuous," which uses continuous age rather than age category as a variable, is more accurate in predicting recidivism than "model2." This can also be seen in the likelihood ratio test comparing these two models.

Table 21: Likelihood Ratio Test Summary for Model2 (1) vs. Model2Continuous (2)

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	4	-3813.66			
2	3	-3785.26	-1	56.79	0.0000

In summary, "model2continuous" has the highest percentage of accuracy out of all of the models created. The likelihood ratio test comparing "model5" and "model2continuous" produces the output below.

Table 22: Likelihood Ratio Test Summary for Model5 (1) vs. Model2Continuous (2)

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	11	-3788.09			
2	3	-3785.26	-8	5.66	0.6854

The results of the likelihood ratio test indicate that "model2continuous" is a slightly better fit for our dataset because the log likelihood of "model2continuous" is greater than that of "model5". Hence "model2continuous" has the largest log likelihood of all of the above models and is therefore the best fit model for this dataset.

7 Analysis

7.1 Fairness of the COMPAS Algorithm

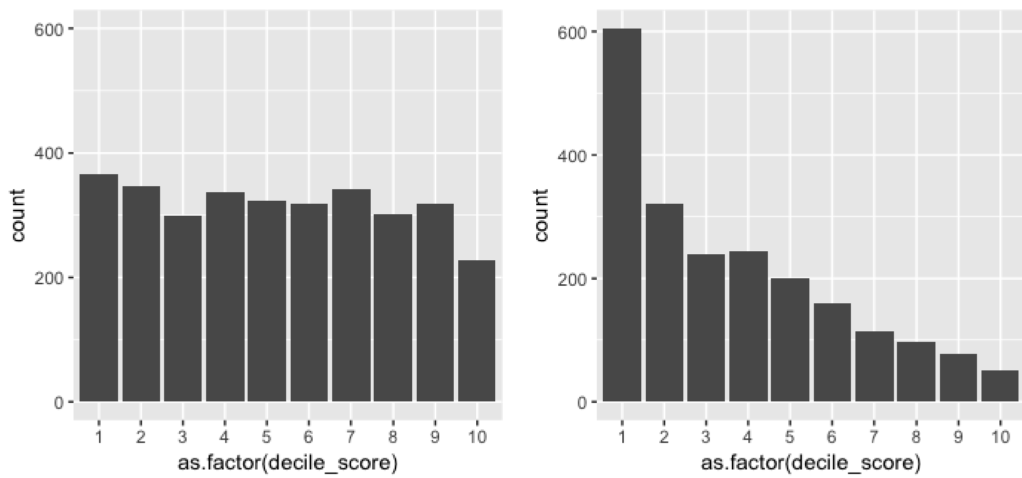
As shown in Section 6.1, COMPAS only has about 66% accuracy and therefore should be questioned before being heavily relied upon to make such impactful decisions in the courtroom. Throughout Section 7, the drawbacks of COMPAS and the effects that these drawbacks have on different groups are shown.

Before analyzing what it means for COMPAS to be "fair" in its scoring process, the concept of fairness in the context of predictive algorithms needs to be defined. By Northpointe's argument, fairness is quantified by the percent of predictions that are correct in identifying recidivists. They are not concerned with the correctness of predictions made for non-recidivists. This is exactly where COMPAS fails to treat different racial groups equally.

First, take into account the distribution of decile scores for both African Americans and Caucasians assigned by COMPAS in Figure 4.

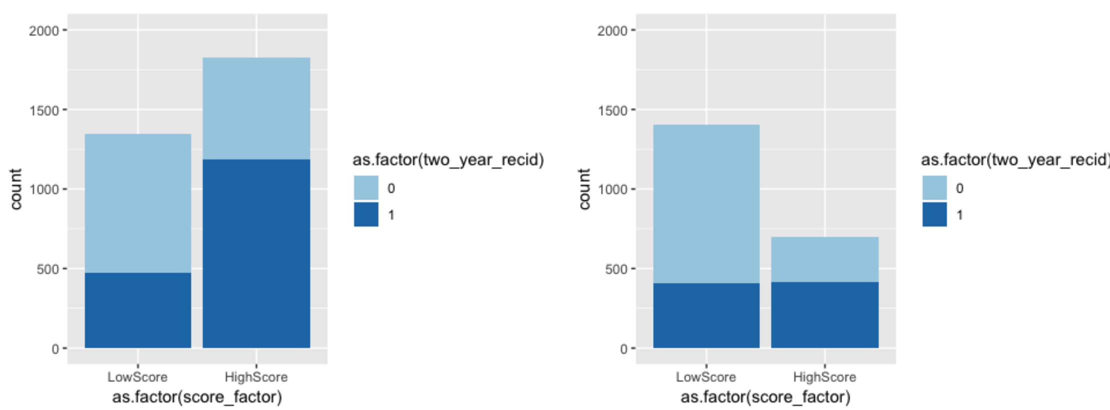
Clearly, African-Americans are more likely to receive Medium or High scores than Caucasians. 57.6% of African Americans are given high scores compared to 33.1% of Caucasians. The distribution of Caucasians significantly decreases as scores increase from low to high, whereas the distribution of African Americans is uniformly distributed across all decile scores.

If we consider fairness to mean equality in accuracy for only those defendants who did re-offend, then Northpointe would be correct in their assumption that COMPAS is fair. Of those who do recidivate, African Americans are almost equally as likely to receive Medium or High Scores as Caucasians (64.95% vs. 59.48%). This can be seen in the plots in Figure 5. The light blue portion of each graph represents those who do not recidivate and the dark blue represents those who do.



(a) Distribution of Scores for African Americans (b) Distribution of Scores for Caucasians

Figure 4: Distribution of Decile Scores for Different Race Categories



(a) Score Factor vs. Recidivism for African Amer. (b) Score Factor vs. Recidivism for Caucasians

Figure 5: Score Factor vs. Recidivism for Different Race Categories

Again, this is why Northpointe asserts that COMPAS is technically fair: because it correctly predicts recidivism at nearly the same rate for those who do recidivate. COMPAS may achieve equal accuracy for recidivists with high scores, but it fails to do so for defendants who do not recidivate. African Americans who do not re-offend are roughly twice as likely as Caucasians to receive high COMPAS scores (42.34% vs. 22.01%). African Americans who are given low scores recidivate at lower rates than Caucasians. These findings are summarized in Table 23.

Table 23: Inaccurate Predictions Made by COMPAS

	African Americans	Caucasians
Given High Score and Don't Recidivate	42.34%	22.01%
Given Low Score and Do Recidivate	28.48%	49.64%

An algorithm that results in the wrongful sentencing of one racial group more than it does for another group cannot be considered a fair algorithm. Rather than comparing COMPAS scores to future recidivism, using the probabilities found while running "model2" and comparing them to future recidivism might yield more fair results. These probabilities indicate the probability of recidivating according to the factors used "model2." Any row with probability greater than the threshold is predicted to recidivate and any row with probability less than the threshold is predicted not to recidivate. Typically, a threshold of value 0.5 is used. If we consider only those defendants who do not recidivate, African Americans are still more likely to be predicted to recidivate than Caucasians, although less harsh in difference compared to COMPAS (37.91% vs. 20.61%). One idea to improve the fairness is to adjust the threshold value for different races. If the threshold is lowered for Caucasians, and the threshold for African Americans remains constant, more accuracy equity is achieved. A threshold of 0.5 for African Americans and 0.36 for Caucasians yields nearly equal rates. Of those who do not recidivate, African Americans are only slightly more likely than Caucasians to be predicted to recidivate (37.91% vs. 34.89%).

Another way in which COMPAS is not fair in its assignment of decile scores can be seen in the different gender categories. In *ProPublica's* study, they found that women are 19.4% more likely to receive a high score even though women recidivate at lower rates than men (35.15% vs. 47.95%) [2].

Clearly, COMPAS fails to treat different groups of both recidivists and non-recidivists equally. Without knowing every detail of the actual algorithm, it is hard to conclude what the single source causing these disparities may be. While Northpointe claims that race is not a factor in the assignment of scores, *ProPublica's* model indicates otherwise. In *ProPublica's* model, the most significant factors in predicting a defendant's COMPAS score are prior counts, age (with being younger more significant than being older), two year recidivism, and race (with African American being the most significant out of all of the race categories). Being an African American increases the log odds of having a high score by .4772 in comparison to being a Caucasian. This implies that race may be a factor used in the COMPAS algorithm.

Possible insight may come from training *ProPublica's* model on the subset of the dataset that corresponds to the Caucasian population and testing the model on the African American population afterwards. Table 24 shows that 1253 of 2103 Caucasians are predicted to have low scores and do in fact have low COMPAS scores. That figure decreases by half in Table 25 and, more notably, most of the African American population is now predicted to not receive a high score even if COMPAS assigned them a high score. If COMPAS trained their model on those racial groups who don't face discrimination, more fair results might be achieved for the other racial groups.

Table 24: Truth Table for Model Trained on Caucasians

	LowScore	HighScore
FALSE	1253	326
TRUE	154	370

Table 25: Truth Table for Model Tested on African Americans

	LowScore	HighScore
FALSE	671	908
TRUE	226	298

If "model2continuous" is held to the same standard of fairness that COMPAS is in the above paragraphs, it can be seen that the ways in which "model2continuous" predict recidivism incorrectly differ from the ways in which COMPAS does. Without controlling for race, recall the truth table for "model2continuous."

	0	1
FALSE	2718	1338
TRUE	645	1471

This model results in 1338 cases where the defendant is not predicted to recidivate but later does and 645 cases where the defendant is predicted to recidivate but later does not. On the other hand, COMPAS makes equally incorrect predictions for both cases. COMPAS results in 1076 cases where the defendant is not predicted to recidivate but later does and 1018 cases where the defendant is predicted to recidivate but later does not. These results are summarized in Table 26.

Table 26: Inaccurate Predictions Made by COMPAS vs. Model2Continuous

	COMPAS	Model2Continuous
Not Predicted to Recidivate but Does	1076	1338
Predicted to Recidivate but Does Not	1018	645

It was calculated that "model2continuous" has 67.87% accuracy which is 1.8% more accurate than COMPAS. While this may seem like an insignificant improvement, this results in 111 more correct predictions using just age and prior counts. Age has been proven by the United States Sentencing Commission to have a significant correlation to future criminal behavior [3]. The United States Sentencing Commission has also done an extensive study confirming that past criminal history (i.e. prior counts) is a strong predictor of recidivism [4].

The use of only these two factors to predict recidivism reduces the opportunity for racial bias to affect the results of the predictions. The COMPAS survey contains several subjective questions that influence the assignment of scores. Some of the questions asked are:

- How many of your friends have ever been arrested?
- If you lived with both parents and they separated, how old were you?
- How many of your friends do illegal drugs?
- How often do you barely have enough money to get by?
- What were your usual grades in school?

Through these questions, systemic racial bias is introduced into the scoring process. The way that these questions are answered by each defendant is highly dependant upon the social and economic background of the defendant. Therefore, marginalized groups are more at risk of receiving higher COMPAS scores simply because of how they answer these questions. These scores indicate to a judge the risk that the defendant has of recidivating, falsely implicating that the systemic inequalities faced by African Americans inherently make this group more likely to recidivate.

Although the accuracy of "model2continuous" is not perfect, this model is a better predictor of recidivism than COMPAS scores because

1. it has greater accuracy,
2. it prevents the introduction of bias by using only two factors,
3. and it is more fair in that it decreases the number of cases where defendants are predicted to recidivate, given harsher jail/parole/probation sentences because of this, and don't go on to recidivate in the future.

The ways in which systemic racism permeate every stage of the criminal justice system will be outlined in Section 7.2.

7.2 How Systemic Racism Affects the Criminal Justice System

While it is not entirely known in what capacity race is used in the COMPAS algorithm, there are several ways in which racial bias has had historical implications that, even today, impact the assignment of decile scores. The concept of systemic racism is often hard to grasp due to its elusive nature. Systemic racism is not an overt act of individual racism, but is instead the presence of unequal access to opportunities or goods that are available to the majority. Eventually, this unequal access is normalized through institutions like the government, private companies, and schools.

African Americans have been treated as inferior to Caucasians since the founding of the United States. Although slavery was abolished in 1865 with the ratification of the Thirteenth Amendment, a clause in the Thirteenth Amendment states that slavery is only permissible as a punishment for crime. Because of this, many criminologists and psychologists consider the United States prison system as a continuation of slavery. With the passing of the Thirteenth Amendment also came the passing of the Jim Crow laws by white government officials. These laws allowed for the continuation of legal segregation of African Americans, forcing African Americans to live in the most socioeconomically deprived and disorganized residential areas and reserving nicer residential areas for whites. These urban neighborhoods where African Americans were forced to live were densely populated and lacked resources necessary to survive. Access to affordable and/or healthy food was limited. The effects of this segregation are still felt today.

This segregation was also evident in the public school system. Educational opportunities that were available to African Americans were inferior to those available to Caucasians. To this day, federal funding for public schools is directly correlated the income level of the neighborhood of the school [5]. This leads to fewer job opportunities for those who do graduate from these schools and results in significantly lower wages. In addition to having less opportunity, inherent bias, false stereotypes, and blatant discrimination all perpetuate this cycle of racism. For example, as determined by the U.S. Supreme Court, it is legal to not hire someone who has dreadlocks since dreadlocks are not "immutable characteristic of black persons" [6]. Systemic racism has severely decreased the opportunity for African Americans to have residential stability, positive social environments, and high paying jobs. The COMPAS survey asks 30 questions that pertain to these three areas of a defendant's life. In doing so, the COMPAS survey allows historical racism to influence what score an African American defendant is assigned.

Systemic racism is not an idea of the past, but rather one that is continually revived. In addition to the historical inequalities that impact decile scores, systemic racism is introduced throughout each stage of the criminal justice system. During the policing stage of the criminal justice system, one of the ways in which racism reveals itself is through the allocation of police resources. Police associate high levels of crime with ethnically diverse neighborhoods, and therefore send most of their units to these areas. In some cities, new police officers are the ones assigned to these areas. Due to a higher presence of police in these areas, some of those who have zero years of experience, more arrests are made in these neighborhoods in comparison to wealthy, predominantly white neighborhoods.

Even though African Americans are being arrested at higher rates than Caucasians, African Americans are not necessarily committing crimes at higher rates than Caucasians. In fact, African Americans and Caucasians commit drug related crimes at essentially the same rate. Additionally, African Americans are substantially more likely than Caucasians to be the victims of violent crimes [7]. However, due to the unequal presence of the police force in different neighborhoods, whites are less likely to be arrested. Thus, simply because they are more likely to be caught, African Americans have a higher chance of being arrested, thus being labeled as recidivists and having a higher number of prior counts than whites. As shown in *ProPublica's* model, prior counts are the most significant predictors of receiving high decile scores. This is one explanation as to why African Americans are typically assigned higher decile scores.

Racial bias inherent directly in the police force can also explain the disparity at hand. In a New York Times study of 15 high profile shootings, all of which included an African American being killed by a white police officer, only three cases resulted in the conviction of the officer [8]. Also, African Americans are consistently pulled over more than whites, three times as likely to be searched once pulled over, and twice as likely to be arrested in these scenarios [9]. This is another reason why African Americans have a higher number of prior counts and higher recidivism rates than whites. Furthermore, police are responsible for completing the first thirty questions of the COMPAS survey when arresting a defendant. One of the questions asks, "Is this person a suspected gang member?" A subjective question such as this one allows an officer personal bias to influence what decile score the defendant is assigned.

During the sentencing period, there are a few factors that contribute to the racial disparity present in the criminal justice system. When it comes to sentences that have mandatory minimums, it has been shown that prosecutors are two times more likely to charge African Americans than they are Caucasians with similar criminal backgrounds [10]. Again, this leads to African Americans having more serious criminal charges (felonies as compared to misdemeanors) and having more prior counts than Caucasians.

Unfortunately, the described bias is just as present in written policy as it is in human discretion.

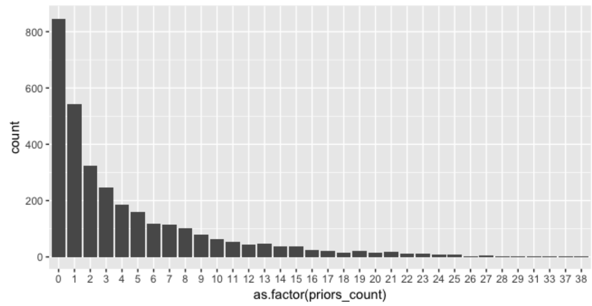
One policy that implicitly targets specific racial groups and low income populations is the drug free school zone law, which requires harsher sentencing for drug crimes committed in school zones. These drug free school zones are large and, due to the fact that most schools are in urban areas where concentrations of people of color live, disproportionately contribute to the number of arrests of African Americans. The *Los Angeles Times* studied the effects that the three strike rule has had on the African American population in California. This rule requires that any defendant who has committed three crimes serve a life sentence at the discretion of the court. In California, three misdemeanor charges qualify a defendant for a life sentence in prison. Of those who commit a third crime in their lifetimes, African Americans are given life sentences 13.3% more than whites are [11]. Both of these laws contribute to the increased number of prior counts of African Americans.

Once arrested, defendants facing misdemeanor charges who cannot afford a private lawyer are not granted the right to a public defender. Those defendants who come from impoverished backgrounds are at a disadvantage in the courtroom without the knowledge and expertise of a lawyer. Although misdemeanors most often do not result in jail time, not having a lawyer may lead to a defendant being sentenced to jail time in some cases. Defendants facing felony charges who cannot afford a private lawyer are granted the opportunity to have a public defender aid in their defense, but the quality of representation available is decreasing as the majority of states have cut public defender budgets. Public defenders are taking on impossible workloads and cannot provide the best quality case for every defendant given the circumstances [12]. Again, this most likely results in a defendant receiving a harsher sentence than someone who could afford a private lawyer. The lack of quality representation in both cases undoubtedly supports the argument that African Americans have higher numbers of prior counts due to unequal opportunity.

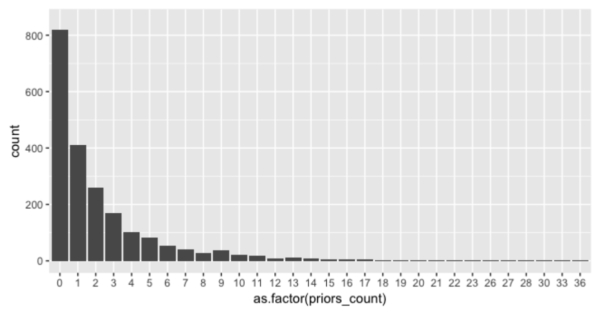
Just as racial bias in police forces adds to the growing African American prison population, racial bias in correctional officers has the same effect during the parole stage of the process. Often times, parole terms are recommended by the correctional officer. In a New York Times study of the New York state prison system, the largest disparities were found when correctional officers were given the authority to determine the punishment for inmate violations that could not be proven with hard evidence, such as a defendant refusing to listen to an order. In one New York prison, African Americans were four times as likely than whites with similar offenses to be sent to solitary confinement and stayed nearly one and a half times the amount whites did in isolation [13]. This racism exhibited by some correctional officers affects the ways in which African Americans can, or in this case cannot, utilize the resources typically offered in prisons. Inmates in bad standing are unable to work while in jail, access education and therapeutic treatment [9]. These all worsen the chance of parole. Of those granted parole, transitioning back into society can prove to be very challenging.

Once out of the prison system, those with felony charges lose access to several basic resources necessary to survive. Not only do these citizens no longer have access to food stamps, financial aid programs, and housing but they also face employment discrimination, travel restrictions, and the risk of losing parental rights. The loss of these basic rights often makes it near impossible for one to hold and keep a job, to be able to afford food, and to spend time with family. The lack of financial stability may put parolees at risk of becoming homeless. The absence of financial and familial stability are highly correlated with crime, meaning that these disadvantages, established and regulated by the government, put parolees at a greater risk of being rearrested.

Systemic racism is rooted in social, political, and economic discrimination that threatens the progression of the African American community. The perpetuity of systemic racism is in part due to the United States' criminal justice system, one of the largest institutions in the world. All of the above examples of inherent racism are what contribute to the continually increasing prison population, the majority of which is comprised of African Americans. Bias during the policing, sentencing, and parole stages of a defendant's process all explain why African Americans are more likely than whites to have a larger number of prior counts and to be labeled as recidivists. Our dataset supports the claim that African Americans have more prior counts than Caucasians, which can be seen in Figure 6 below. As shown in *ProPublica's* model, prior counts is the single most important factor in predicting whether a defendant is assigned a high score by COMPAS. Thus, even though African Americans do not necessarily commit crimes at higher rates than Caucasians, it makes sense according to *ProPublica's* model why African Americans generally have higher COMPAS decile scores.



(a) Distribution of Prior Counts for African Americans. Mean = 4.24



(b) Distribution of Prior Counts for Caucasians. Mean = 2.29

Figure 6: Prior Counts Distributions for Different Race Categories

8 Conclusion

This study concludes by arguing that Northpointe's COMPAS algorithm is not a reliable or efficient tool for use in the courtroom. The findings conclude that COMPAS has an accuracy of 66.07% of predicting recidivism based on a defendant's assigned score. In addition, these findings provide insight as to how incorrect COMPAS predictions affect different racial groups. African Americans are substantially more likely than Caucasians to be given high decile scores and to not recidivate thereafter. For those who do recidivate, COMPAS predictions were equally accurate for both groups. In recreating *ProPublica's* model, it was shown that prior counts, age, two year recidivism, and race were the most significant factors in predicting a defendant's COMPAS score. These results demonstrate a strong correlation between race and the decile score assigned.

The model "model2continuous" which uses only two factors, continuous age and prior counts, performs slightly better than COMPAS in predicting recidivism with an accuracy of 67.87%. While race is shown to be a significant predictor of COMPAS scores in the model created by the researchers at *ProPublica*, the results of this model show that race is not indicative of future criminal behavior.

In order to achieve a criminal justice system that is more fair in the treatment of different racial groups, the decile scores produced by COMPAS must not be so blindly utilized. The algorithm has the potential to let systemic racial bias influence the sentencing of defendants belonging to certain racial groups. These decile scores are calculated from the answers given on the COMPAS survey. Many of these questions are subjective, asking about the defendant's familial and financial stability, educational background, social environment, and employment history. In doing so, the survey allows systemic racism to influence the computation of COMPAS scores.

Rather than using COMPAS, perhaps judges should consider the use of a model similar in structure to "model2continuous." In the future, one could train this model on the data belonging to Caucasian defendants and test its performance when the model is applied to African Americans. This is one way in which the discrimination evident in the COMPAS algorithm could potentially be minimized.

Racial bias throughout history, police forces, correctional officers, laws, etc. all lead to African Americans having a higher number of prior counts than Caucasians. And although inherent racial bias may never be eradicated, the use of "model2continuous" together with a more equal dispersion of police resources might also offer a better alternative to COMPAS in that these efforts will result in more accurate predictions of recidivism. Surely, "model2continuous" decreases the ways in which racial bias can interfere with the accuracy of the recidivism predictions by eradicating the use of the COMPAS survey and ensuring that race is not an explicit factor in score calculations.

References

[1] Angwin, Julia. "COMPAS 'CORE' Risk and Needs Assessment." ProPublica, 2016.

[2] Larson, Jeff, et al. *How We Analyzed the COMPAS Recidivism Algorithm.* ProPublica, 2016, *How We Analyzed the COMPAS Recidivism Algorithm,* www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

[3] Pryor, William H. *The Effects of Aging on Recidivism Among Federal Offenders.* United States Sentencing Commission, 2017, *The Effects of Aging on Recidivism Among Federal Offenders*, www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2017/20171207_Recidivism-Age.pdf.

[4] Pryor, William H. *The Past Predicts the Future: Criminal History and Recidivism of Federal Offenders.* United States Sentencing Commission, 2017, *The Past Predicts the Future: Criminal History and Recidivism of Federal Offenders*, https://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2017/20170309_Recidivism-CH.pdf.

[5] "Jim Crow: Now & Then." *Jim Crow: Now Then,*

[6] United States Court of Appeals, Eleventh Circuit. *Equal Employment Opportunity Commission, Plaintiff-Appellant, v. Catastrophe Management Solutions, Defendant-Appellee.* https://www.naacpldf.org/files/about-us/CMS%20-%20Cert%20Petitio%20FINAL.PDF

[7] Ghandnoosh, N. (2014). *Race and Punishment: Racial Perceptions of Crime and Support for Punitive Policies.* Washington, D.C.: The Sentencing Project.

[8] Lee, Jasmine C., and Haeyoun Park. "15 Black Lives Ended in Confrontations With Police. 3 Officers Convicted." *New York Times*, 5 Oct. 2018.

[9] *Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System.* The Sentencing Project, 2018, *Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System,* www.sentencingproject.org/publications/un-report-on-racial-disparities/.

[10] Starr, S. B. and Rehavi, M. M. (2013). *Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of.* The Yale Law Journal, 123(2), 2-80.

[11] Krikorian, Gref. “More Blacks Imprisoned Under ’3 Strikes,’ Study Says.” *Los Angeles Times*, 5 Mar. 1996, www.latimes.com/archives/la-xpm-1996-03-05-mn-43270-story.html.

[12] Oppel, Richard A, and Jugal K Patel. “One Lawyer, 194 Felony Cases, and No Time.” *New York Times*, 31 Jan. 2019, www.nytimes.com/interactive/2019/01/31/us/public-defender-case-loads.html.

[13] Schwirtz, Michael, et al. “The Scourge of Racial Bias in New York State’s Prisons.” *New York Times*, 3 Dec. 2016, www.nytimes.com/2016/12/03/nyregion/new-york-state-prisons-inmates-racial-bias.html.

[14] Corbett-Davies, Sam, and Emma Pierson. “A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It’s Actually Not That Clear.” *Washington Post*, 17 Oct. 2016.

[15] Dieterich, William, et al. *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity.* NorthPointe Inc., 2016, *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity,* www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html.

[16] Dressel, Julia, and Hany Farid. “ The Accuracy, Fairness, and Limits of Predicting Recidivism.” *Science Advances*, 30 Mar. 2018, advances.sciencemag.org/content/advances/4/1/eaao5580.full.pdf.

[17] Has, Karen. “AI Is Sending People to Jail—and Getting It Wrong.” *MIT Technology Review*, 21 Jan. 2019, www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/cibonayrae.github.io/jimcrow/index.html.

[18] Jones, J. M. (1997). *Prejudice and Racism (2nd ed.)*. New York: McGraw-Hill.

[19] Larson, Jeff. “COMPAS Analysis.” github.com/propublica/compas-analysis. Github, 2016.