Reproduction of Bias Behind Bars

Ankhee Paul

16/12/2020

Abstract

On October 24th, 2020, The Globe and Mail Canada launched 'Bias behind Bars' by their crime justice reporter Tom Cardoso who investigates the systematic bias in the justice system against Black and indigenous people while establishing their risk assessment scores. In this paper we aim to reproduce Tom Cardoso's article. We use multinomial logistic regression model to determine the impact of racial background on the two most important risk assessment scores of male and female inmates. Our results show that black inmates have a higher likelihood of receiving the worst offender security level compared to white prisoners whereas black and indigenous male prisoners have a higher probability of receiving the worst reintegration potential score compared to other male prisoners of other racial backgrounds.

Introduction

On October 24, 2020 the Globe and Mail Canada launched "Bias behind bars" ; a two-year inquiry by The Globe's crime and justice reporter, Tom Cardoso, that uncovered systemic bias in the Canadian judicial prison system. By 26th October, 2020, the investigation gathered unanimous support of Members of Parliament, calling for an independent study on systemic discrimination in federal prisons, including inmate risk assessments. In the article published, Cardoso explores the risk assessment received by federal inmates and uses logistic regression to determine how race affects the scores received by the prisoners.

When an individual enters the Correctional System, they immediately a series of psychological evaluations, cognitive tests and questionnaires, based on which a risk assessment report is prepared. This initial report, handed to correctional officers, create an impression on the officer and further determines their future in the correction system. Length and type of their sentence, their access to rehabilitative and treatment programs, security level of the facility they would be placed in as well as potential parole- all depend on a person's risk assessment. It is basically the score that determines how much threat an individual poses to the society. Cardoso, through his analysis, discovered that black and indigenous people have are likely to receive worse risk assessment score on average implying a bias in the correctional system.

In this paper, we attempt to reproduce Cardoso's work to the best of our ability. In order to do so we use the data set provided to us by the Globe and Mail that was originally used by Cardoso. We use a multinomial logistic regression model to determine the impact of racial background on the two most important risk assessment score, namely, Offender Security level and the reintegration potential score. Our results yield significant in determining that black inmates have a higher probability of receiving the worst offender security level compared to white prisoners whereas black and indigenous male prisoners have a higher probability of receiving the worst reintegration potential score compared to other male prisoners of other racial backgrounds.

In Section 2, we represent the dataset and plot various variables in it. We also describe modifications made to the data set to allow us to run our regression models. Section 3 highlights the logistic regression models and establishes the validity of the models. Section 4 displays the results yielded from our models. Finally, Section 5 concludes with a discussion of our models and the yielded results as well as their implications and weaknesses.

2.Data

2.1. Data Information

This paper works with the data used by Cardoso in the original paper. In late 2018, when the Globe and Mail decided to undertake this investigation, they reached out to the Correctional Services of Canada and filed a freedom of information request with them. The agency released data with around 744,958 rows and 25 columns, with details of around 50,116 people in custody or supervised release.

When an individual receives a sentence of two years or longer, the CSC enters their details as a new inmate to their records database. The database logs all information about the inmates: their age, gender, religion, psychological evaluations, risk assessment scores, correctional plans, progress reports, parole recommendations and more. This dataset contains data from 2012 to 2018, with a snapshot of the database captured on March 31st, the last day of the CSC's fiscal year.

When an individual enters the correctional system, they are immediately evaluated with the predominant motivation of ascertaining how much of a risk would the individual pose to the society. A preliminary assessment report, handed to the parole officer, has great importance in feeding into future assessments and decisions – access to rehabilitative programs, frequency of meetings with parole officers, security level of the facility assigned, determination of solitary confinement and most crucially, chances of getting parole. In order to determine that, a series of cognitive tests, questionnaires, as well as the professional judgement of correctional officers are taken into consideration. This ends up with a tally of their risk assessment scores. The data contains five different types of risk assessments along with age, gender, religion, race, details of length of a prisoner's sentence, whether they are in a minimum-medium or maximum facility or whether they are on parole. It also displays "Offender number" which is a unique id provided to each individual. "Offence ID" represents a unique offence, "Warrant ID" represents a unique warrant of committal issued by a judge and "Sentence ID" is a unique identifier for each inmate's sentence. A prisoner may have multiple offence ids, warrant ids and sentence ids.

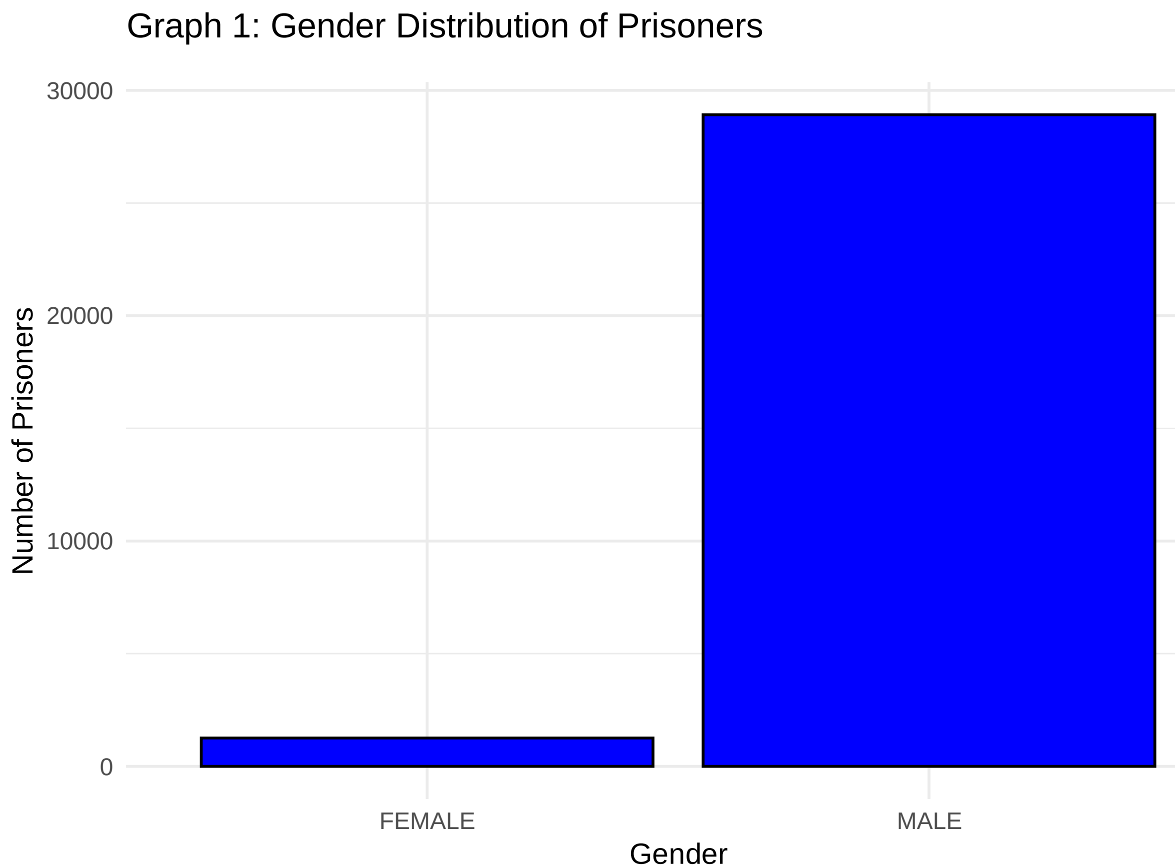
The most important scores that is pertinent in our study is the "Offender Security Level" and the "REINTEGRATION POTENTIAL". The "OFFENDER SECURITY LEVEL" is an ordinal categorical variable, that is, it is qualitative following a certain order of "Maximum", "Medium" and "Minimum". It describes the security level designated to each inmate upon their arrival as a measure of the security risk they pose and impacts their access to treatment programs as well as determines the security level of the facility they would be assigned to. The "REINTEGRATION SCORE" is a score with levels "LOW', "MEDIUM" and "HIGH" which basically evaluates the potential of reformation of an inmate. This score is used to estimate whether an inmate is eligible to re-enter society and how much of a security risk they pose if the former were to occur. The variable "STATIC RISK" is a proxy for criminal record. It is a CSC tool determined by Statics Factor Assessment that measures a person's past involvement with the justice system. If a person has "HIGH" static risk, it implies that the person had significant involvement in the past with the justice system.

In order to clean the data, we first filter out people under provincial jurisdiction that leaves us with 741,738 rows and 49,165 unique individuals. We then remove missing values and categorize the race of an individual under "Black", "White" and "Other", with "Other" comprising of all other racial backgrounds such as South-East Asian, North American, etc. As Race is self-reported by an individual upon entering the system and can change over time, Cardoso considered the race that an individual provided upon their first entry. Moreover, we select variables such as age, gender, religion, length and type of sentence received by an inmate as well as the two most important risk assessment scores "REINTEGRATION POTENTIAL" and "OFFENDER SECURITY LEVEL".

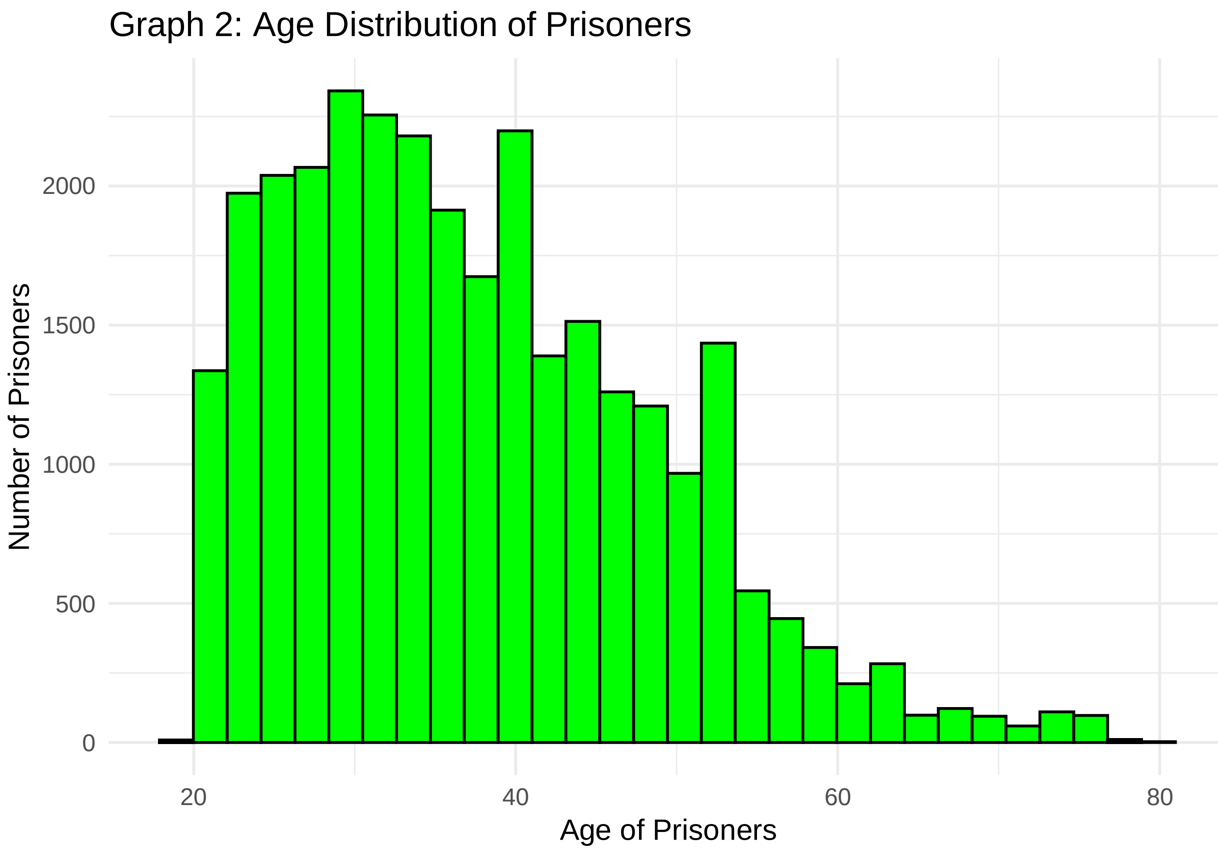
In his methodology, Cardoso simplified people's charges that are described in text form. In order to determine the severity of offence, Cardoso hand-matched more than seven hundred charges to Uniform Crime Reporting Survey offence categories which he then cross-referred with Statistics Canada's crime severity index weights – a system where seriousness and severity of crime is measure by assigning numerical weights to the crimes. For example, a weight of 7656.16 is assigned for a charge of first-degree murder. The weights increase with severity of the crime and to find an inmate's most serious offence, Cardoso picked the largest weight for an inmate's sentence in a given year. Since we were unable to find Statistics Canada's crime severity index weights, we were unable to determine the most serious offence committed by an individual. This left an individual with multiple charges occupying multiple entries in the data set. Hence, we grouped together by an individual's "Offence Number" which is unique to each inmate.

After cleaning up our data, we plotted it using bar plots and histograms in ggplot.

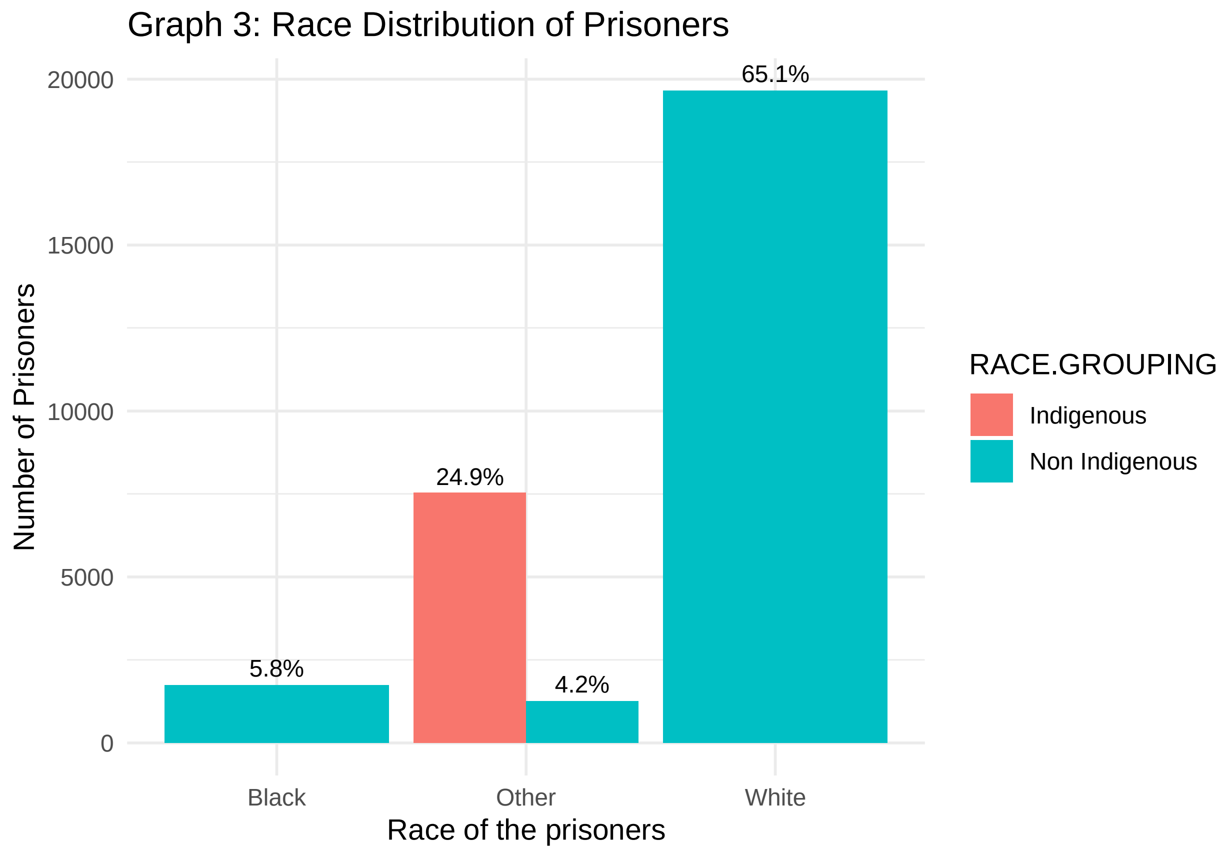
2.2 Display of Data



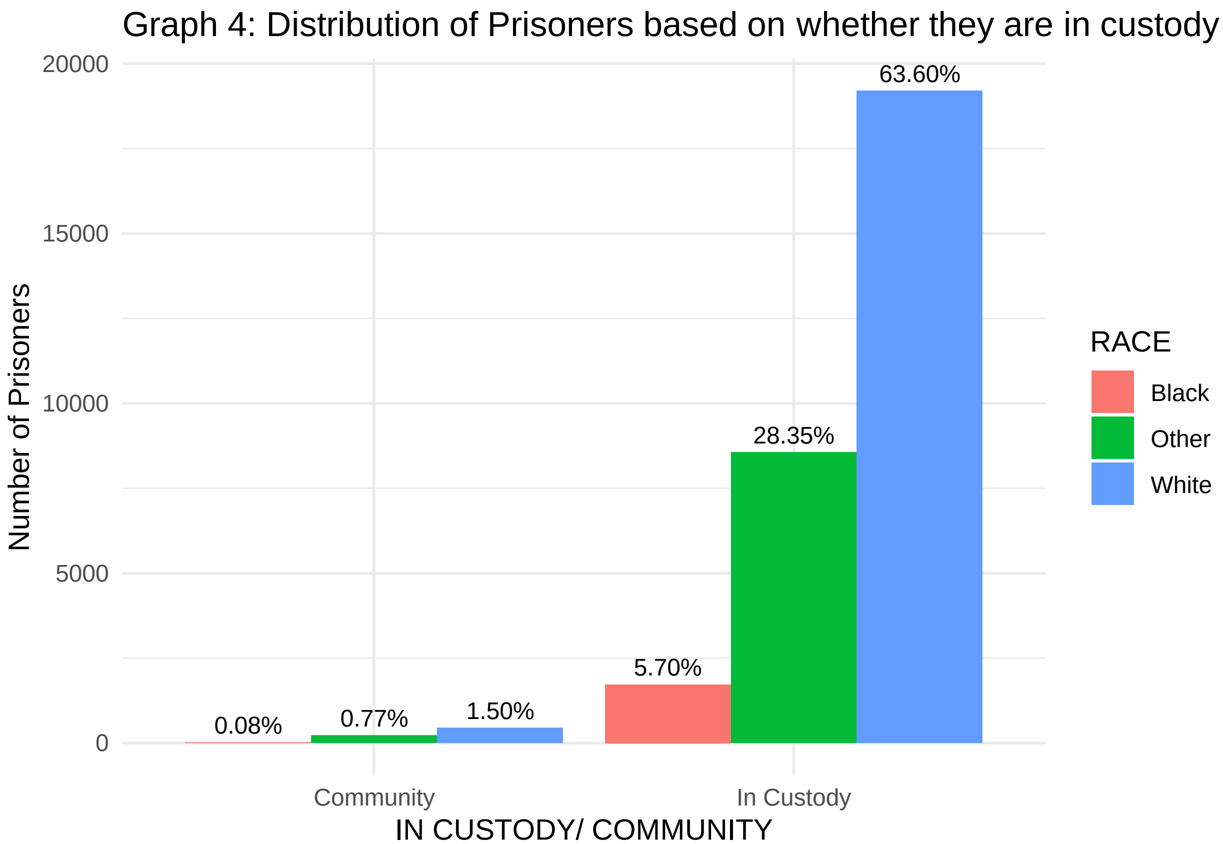
Graph 1 shows us the gender distribution of the inmates. The distribution is overwhelmingly male implying a higher number of male prisoners than female prisoners.

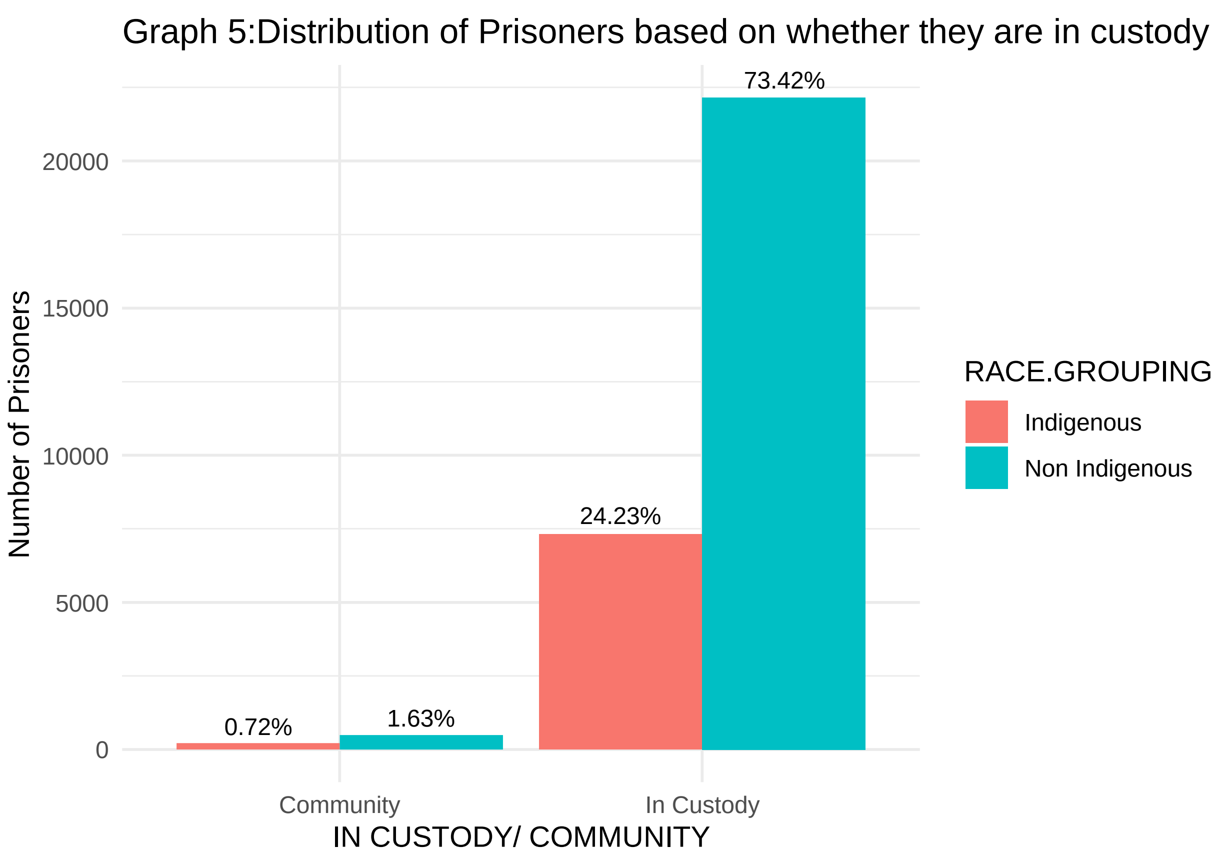


Graph 2 plots the age distribution of the prisoners in a histogram. The plot shows that a higher number of prisoners are between the ages of 25year to 50 years with the mean age being around 40 years. This shows that the prisoner demographic comprises of the younger to middle-aged generations. The histogram is right-skewed or positively skewed implying that the mean and median ages are greater than the mode of the distribution.

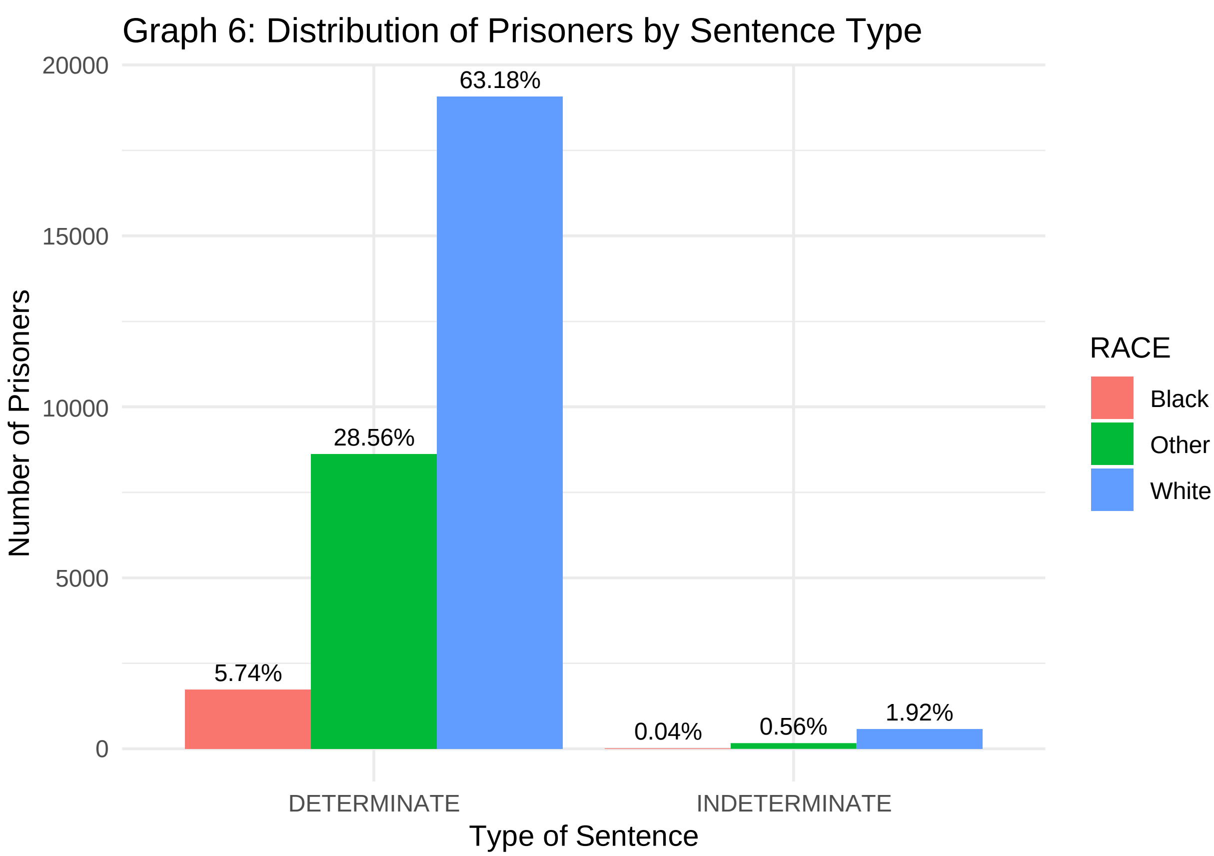


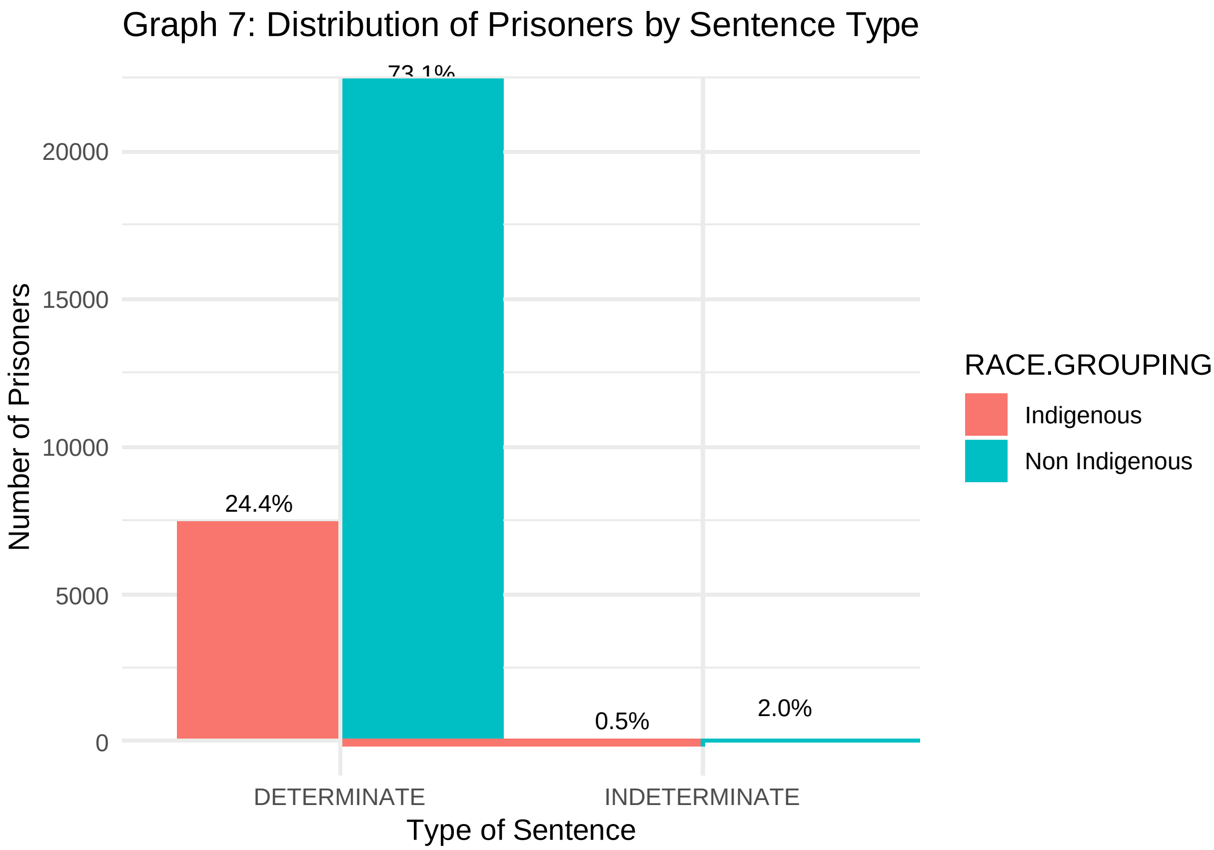
Graph 3 displays the racial distribution of the inmates as per their racial grouping, that is, whether they identify as "Indigenous" or "Non-Indigenous". Although we see a higher number of white individuals than other races, this is due to the population distribution in the country. In his article, Cardoso mentions that given the representation of different races in the Canadian population in general, Black people and Indigenous people are over- represented. Using Cardoso's census figures mentioned in the article, Black people and Indigenous people represented 3.5% and 4.8% of the population respectively in 2016. Graph 3 shows that Black people and Indigenous people represent 6.89% and 20.61% of the prison population respectively implying a over representation as mentioned by Cardoso.


Graph 4: Distribution of Prisoners based on whether they are in custody
20000
15000
10000
5000
0
Number of Prisoners
63.60%
28.35%
5.70%
0.08%
0.77%
1.50%
Community
In Custody
IN CUSTODY/ COMMUNITY
RACE
Black
Other
White

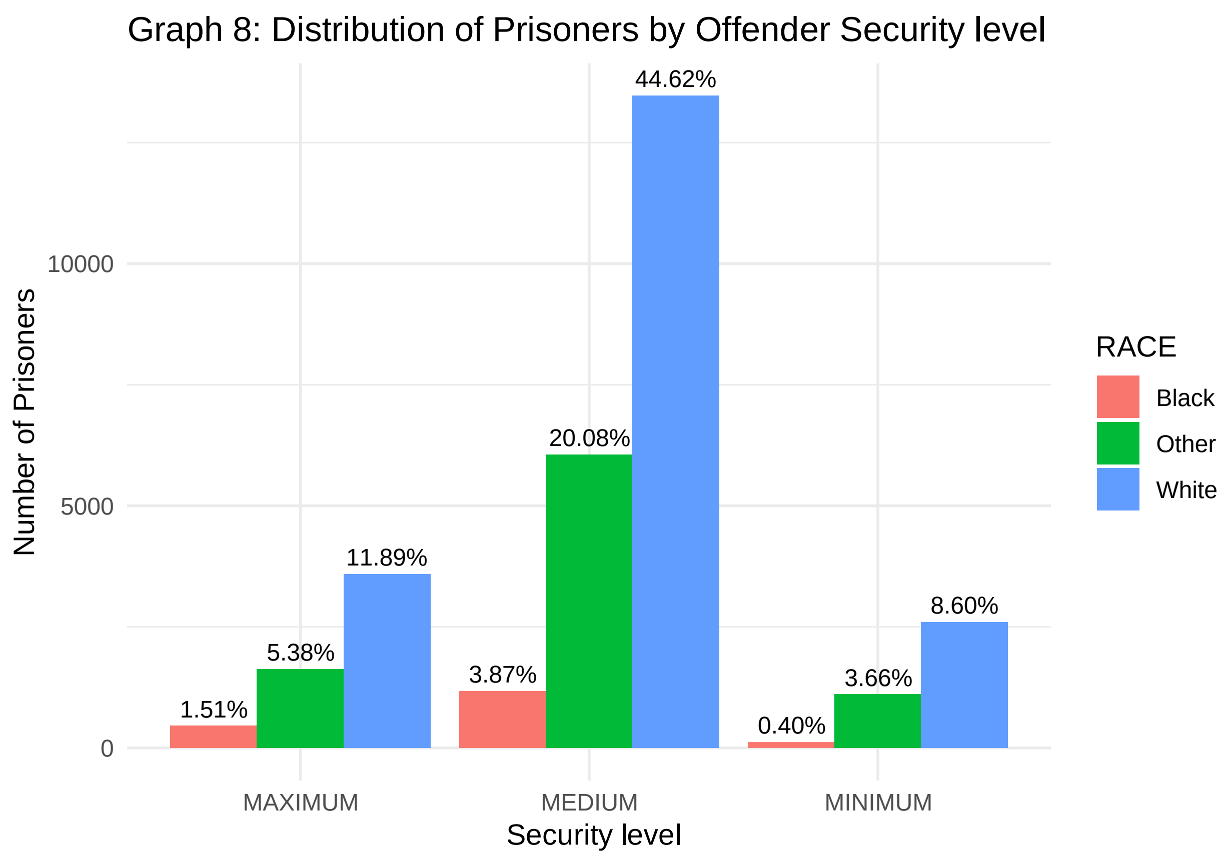

Graph 5:Distribution of Prisoners based on whether they are in custody
20000
15000
10000
5000
0
Number of Prisoners
73.42%
24.23%
0.72%
1.63%
Community
In Custody
IN CUSTODY/ COMMUNITY
RACE.GROUPING
Indigenous
Non Indigenous

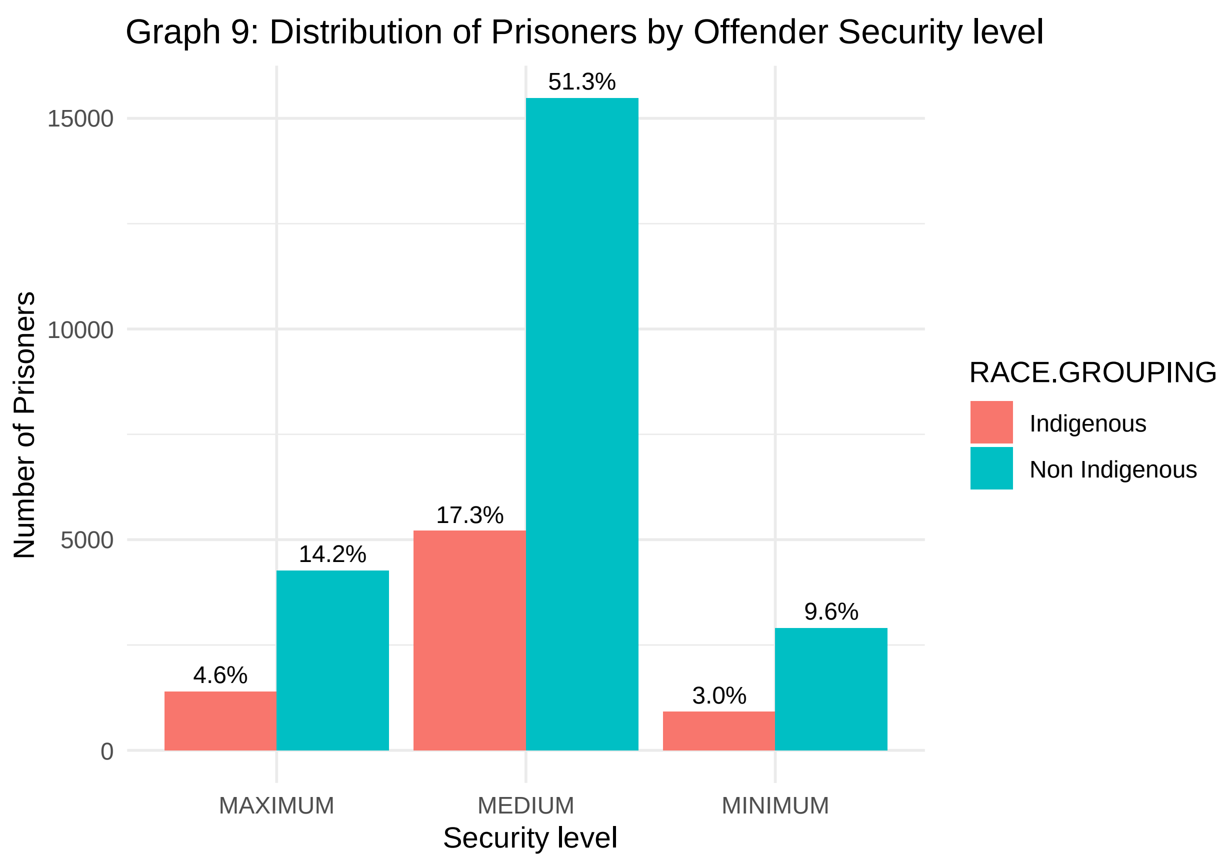
Graphs 4 and 5 shows the distribution of prisoners based on whether they are in custody or on parole. As can be seen the ratio between inmates in custody versus on parole for white people is almost 1:2 whereas the ratio for Black as well as Indigenous people is almost 1:3, suggesting that a lower proportion of black and indigenous prisoners are considered for parole.



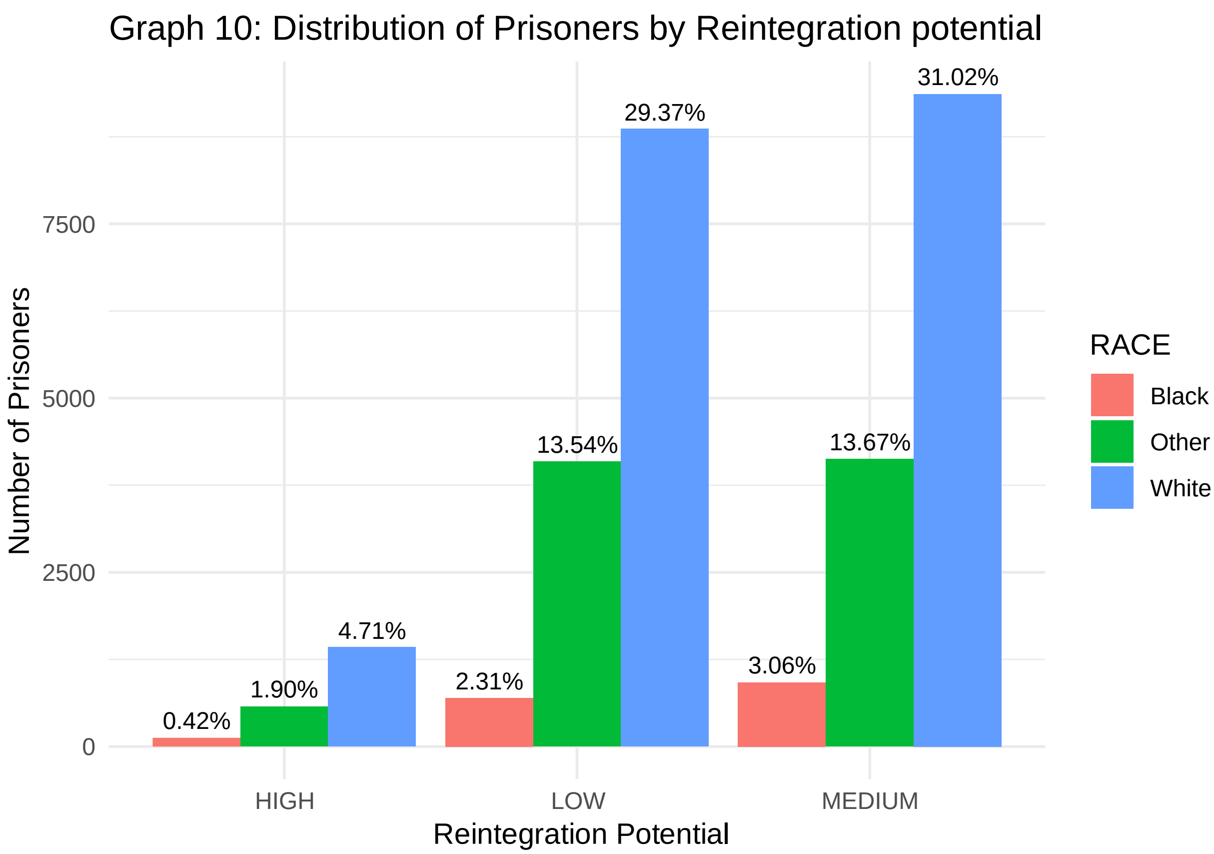


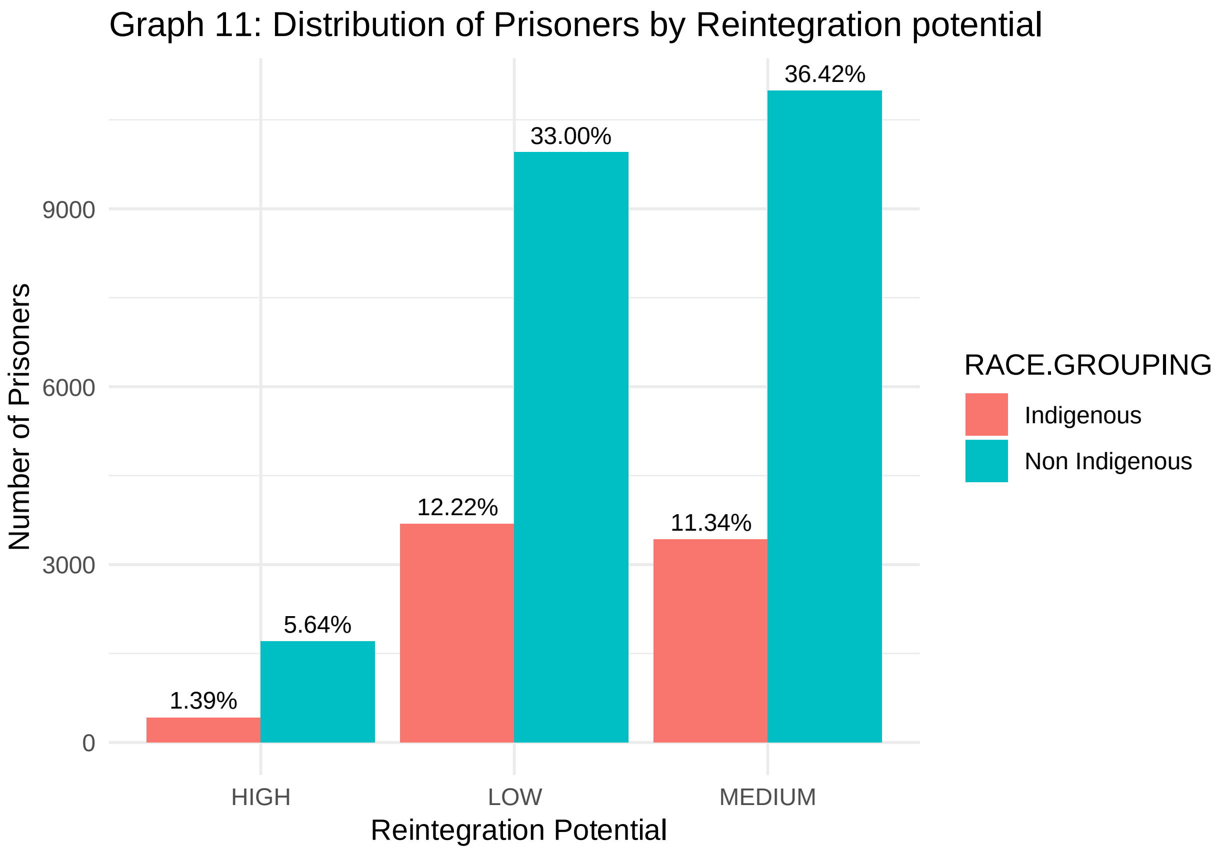
Graphs 6 and 7 show the distribution of prisoners based on their sentence type, whether it's a determinate sentence with a fixed number of years or life imprisonment, in which case, the sentence type would be "INDETERMINATE". In Graph 6, the ratio between "DETERMINATE" versus "INDETERMINATE" is 6:1 for black people compared to roughly a ratio of 5:1 for white people. However, for indigenous people the ratio is almost 5:1 which is roughly the same as one for white people. Considering that the indigenous community is over-represented among the prisoners and under-represented in the population compared to white people, there is a significantly high rate of indigenous people receiving life sentences.


Graph 8: Distribution of Prisoners by Offender Security level
Number of Prisoners
10000
5000
0
44.62%
20.08%
11.89%
5.38%
1.51%
3.87%
8.60%
3.66%
0.40%
MAXIMUM
MEDIUM
MINIMUM
Security level
RACE
Black
Other
White


Graph 9: Distribution of Prisoners by Offender Security level
Number of Prisoners
15000
10000
5000
0
51.3%
17.3%
14.2%
4.6%
9.6%
3.0%
MAXIMUM
MEDIUM
MINIMUM
Security level
RACE.GROUPING
Indigenous
Non Indigenous

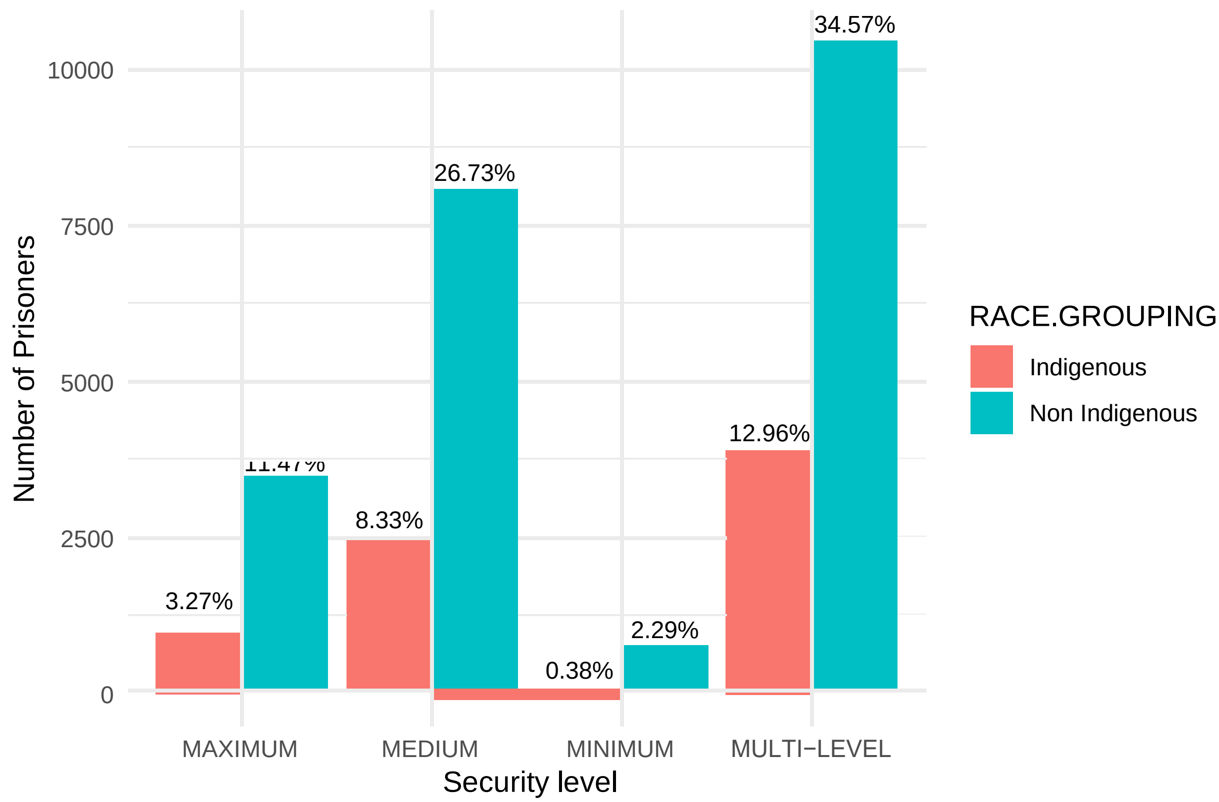
Graphs 8 and 9 show the distribution of prisoners based on their Offender security level score. We condition on the "MEDIUM" score, that is, the score received by a prisoner given the number of prisoners who have received a "MEDIUM" score (since that score has the highest recipients). It is immediately seen in graph 8 that Black people have a ratio of about 1:2 of receiving a "MAXIMUM" score compared to white people with a ratio of almost 1:4. Similarly conditioning on the "MEDIUM" score for Graph 9 that indigenous people have a ratio of approximately 1:3 of getting a "MAXIMUM" score. This implies that Black people get a worse score than indigenous people and white people.



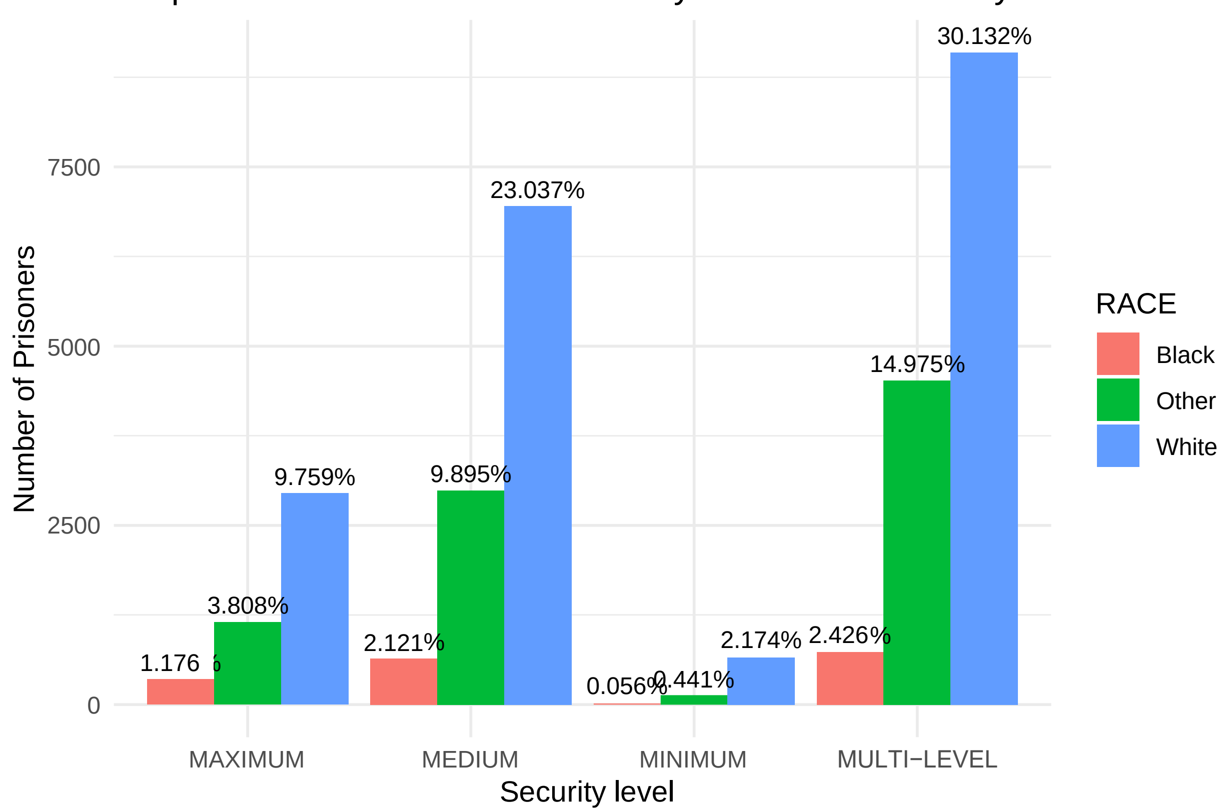


Graphs 10 and 11 plot the distribution of prisoners based on their Reintegration potential score. Since "MEDIUM" score has the most recipients, we condition on it. It is observed in Graph 10 that Black people have a ratio of almost 1:0.75 of getting a "LOW" reintegration potential score compared to white people with a ratio of almost 1:0.95 of getting "LOW". Similarly, in Graph 11 indigenous people have a ratio almost of 1.07:1 of getting a "LOW" score implying that indigenous people get a worse score than black people and white people.

Graph 12: Distribution of Prisoners by Institutional Security level



Graph 13: Distribution of Prisoners by Institutional Security level



Graph 12 and 13 plot the distribution of prisoners based on the security level of the facility they are placed at. Despite "MULTI-LEVEL" security having the highest recipients, we condition "MEDIUM" security level. It is noticed that the ratio for Black people getting assigned to a "MAXIMUM" security facility is almost 1:1 compared to white people who have a ratio slightly greater than 1:2. For indigenous people this ratio is almost 1:3 implying that black people get assigned to a "MAXIMUM" facility more often than indigenous people.

3. Model

3.1 Regression Models

In this study we use a multinomial logistic regression model to determine whether the Race and Race grouping of individual inmates affect the Offender security level score and the Reintegration potential score. To accomplish these tasks, we will be using R to carry out our analysis. Multinomial Logistic Regression is used to classify subjects based on values of a set of predictor variables. This type of regression is similar to logistic regression, but it is more general because the dependent variable is not restricted to two categories unlike generalized logistic regression.

Under generalized logistic regression, which is a type of generalized linear regression, the dependent variable is binary, meaning the outcome can take only two values. Unlike that, multinomial logistic regression generalizes logistic regression to multiclass problems, that is, with more than two possible discrete outcomes. Logistic Regression models are primarily used to predict the odds of an event occurring, given the inputs as predictor variables. Due to the logistic model's link function logit, these odds are interpreted as the log-odds of seeing our event occur, expressed as the log ratio of the probability of success to the probability of failure. For example, we can determine which candidate wins an election using logistic regression given predictor variables like voter demographics (age, sex, religion, etc). We choose to use this model since in both the models that we plan to analyze, the dependent/response variable has more than two outcomes. Both, "REINTEGRATION SCORE" and "OFFENDER SECURITY LEVEL" , are ordinal categorical variables with three different levels.

Like Cardoso, we create two models analyzing the impact of racial background on inmate's two most important risk assessment score, namely "OFFENDER SECURITY LEVEL" and "REINTEGRATION SCORE". Each of these two models are run twice, once for male inmates and once for female inmates. To analyze the impact of racial background on "OFFENDER SECURITY LEVEL", we subset the data into male and female inmates and condition on the "MAXIMUM" score which is the worst score an inmate can receive with "RACE" and "RACE GROUPING" as our predictor variables. Similarly, to check for the impact on reintegration score we subset the data into male and female inmates who are currently in custody since reintegration score is calculated every year. We check the likelihood an inmate ends up with a "LOW" which is the worst score against the odds of receiving a "MEDIUM" or "HIGH" score.

3.2 Model Validation

We divide our data subsets into training and testing sets with 60% of the data in the training set, and use a 10-fold cross validation analysis to find a cross validated estimated accuracy for the different models. This means when we split our survey data into 10 groups and cross validate 10 times using each of these subsamples as a testing set, we compute an estimated accuracy. Taking into account that we are on a log-odds scale, this value represents the total number of correct predictions made by our model over the total predictions.

For the models determining the impact of racial background on offender security level score we get an accuracy of 0.68 for the model with male inmates and 0.69 for the model with female inmates. This suggests that the proportion of correct predictions made by our models over the total number of predictions is 0.68% and 0.69%.

For the models determining the impact of racial backgrounds on reintegration potential score, we get an accuracy of 0.49 for the model with male inmates and 0.68 for the model with female inmates. The low accuracy could be due to the fact that the classes are not separable enough given the features we have.

3.3 Final Model

4. Results

The results and summary statistics for each model is displayed in tables using the kable() function from the "KableExtra" package .The tables include the predictor's; estimate, standard error, z-value, and p-value. Utilizing these values, we can determine the strength and significance of the predictor variables of racial backgrounds on the offender security level score and reintegration potential score. The estimate value relays the change in the log-odds of inmates receiving a maximum offender security level and low reintegration potential score, given the scores received by the inmates and the standard error will tell us the expected error we will see in this estimate value. The z-value and p-value, together, tell us whether or not we can reject the null hypothesis that our estimate value is truly zero, which helps determine the significance of our predictions. To conform with a 95% confidence interval, we are looking for z-values with a magnitude greater than 1.96 or smaller than -1.96, and p-values with values smaller than 0.05. Our estimates are set up as categorical variables and not like usual continuous or count data implying that for each predictor, the estimates are conditioned on a specific response and their values indicate the difference we expect to see from this conditional response.

Table 1: Table 1: Summary statistics for Offender Security Level with Male inmates

y.level	term	estimate	std.error	statistic	p.value
1	(Intercept)	-1.18	0.13	-9.25	0.00
1	RACEOther	-0.29	0.12	-2.40	0.02
1	RACEWhite	-0.31	0.08	-3.98	0.00
1	RACE.GROUPINGNon Indigenous	0.02	0.10	0.21	0.83

Table 2: Table 2: Summary statistics for Offender Security Level with Female inmates

y.level	term	estimate	std.error	statistic	p.value
1	(Intercept)	8.14	20.34	0.40	0.69
1	RACEOther	-10.15	20.34	-0.50	0.62
1	RACEWhite	-2.78	0.35	-7.88	0.00
1	RACE.GROUPINGNon Indigenous	-7.51	20.33	-0.37	0.71

Table 1 and 2:

The tables shows the results for our model predicting the likelihood of receiving an offender security level of "MAXIMUM" based on racial backgrounds. Table 1 displays the result for the model comprising of male inmates and Table 2 displays the result for the model comprising of female inmates.

Race estimates: For this variable we are conditioning on black people. In Table 1 we see that estimates are negative values which means that black males are more likely to get a maximum score compared to other races and white males. Since white males have the lowest estimate of -0.33, they have the least probability of getting a maximum score. The z-values for other races is bigger than -1.96 while its p- value is 0.05 implying that we cannot reject the null hypothesis. The z-value and p-value for white people conform to our requirements and therefore we can reject the null hypothesis.

In Table 2, once again, conditioning on black females, we see that the coefficients have negative values implying that black female inmates are most likely to receive a maximum score than other races and white female inmates. Other races have the least probability of receiving the worst score with the lowest value of -10.15. P-values and z scores for the white people relay significance at 0 and -7.88 respectively but in case of other races, the values imply that we cannot reject our null hypothesis.

Racial Grouping: For this variable we condition on Indigenous people. The coefficient is positive at 0.05 in Table 1 implying that non-indigenous male inmates are more likely to get a worse score. The corresponding p-value and the z-value do not conform to our confidence interval and therefore we fail to reject the null hypothesis.

For female inmates, as displayed in Table 2, the coefficient is negative implying that indigenous female inmates are more likely to get a worse score. With a z value of -0.37 and high p-value of 0.71, this predictor fails to be significant.

Table 3: Table 3: Summary statistics for Reintegration potential with Male inmates

y.level	term	estimate	std.error	statistic	p.value
1	(Intercept)	0.36	0.11	3.39	0
1	RACEOther	-0.34	0.10	-3.36	0
1	RACEWhite	0.23	0.07	3.47	0
1	RACE.GROUPINGNon Indigenous	-0.77	0.09	-8.97	0

Table 4: Table 4: Summary statistics for Reintegration potential with Female inmates

y.level	term	estimate	std.error	statistic	p.value
1	(Intercept)	0.17	0.70	0.24	0.81
1	RACEOther	-1.33	0.69	-1.92	0.05
1	RACEWhite	-0.33	0.36	-0.93	0.35
1	RACE.GROUPINGNon Indigenous	-1.24	0.62	-2.00	0.05

Table 3 and 4:

The tables show the results for our model predicting the likelihood of receiving a reintegration potential score of "LOW" based on racial backgrounds. Table 1 displays the result for the model comprising of male inmates in custody and Table 2 displays the result for the model comprising of female inmates in custody.

Race estimates: Conditioning on black male inmates Table 3 shows that the coefficients are negative for other races and positive for white males. This means that white people have a higher probability of receiving the worse score compared to black male inmates. The p-values and z-values conform to our requirements rendering our predictors statistically significant.

Conditioning on black female inmates, Table 4 shows that coefficients are negative implying that black female inmates stand a higher chance of receiving the worst score while other races stand the least chance of it. The corresponding p-values and z-values, however, do not conform to our requirements and leads to us failing to reject the null hypothesis.

Racial Grouping estimates: This variable conditions on indigenous people. In table 3, the coefficient is negative suggesting that indigenous male inmates have a higher probability of receiving the worst score while the z-values and p-values render this predictor statistically significant. Table 4 shows that female indigenous inmates stand a higher chance of a worse score whereby its z-values and p-values meet our requirements and we can reject the null hypothesis.

Figure 1: Offender Security level for Male

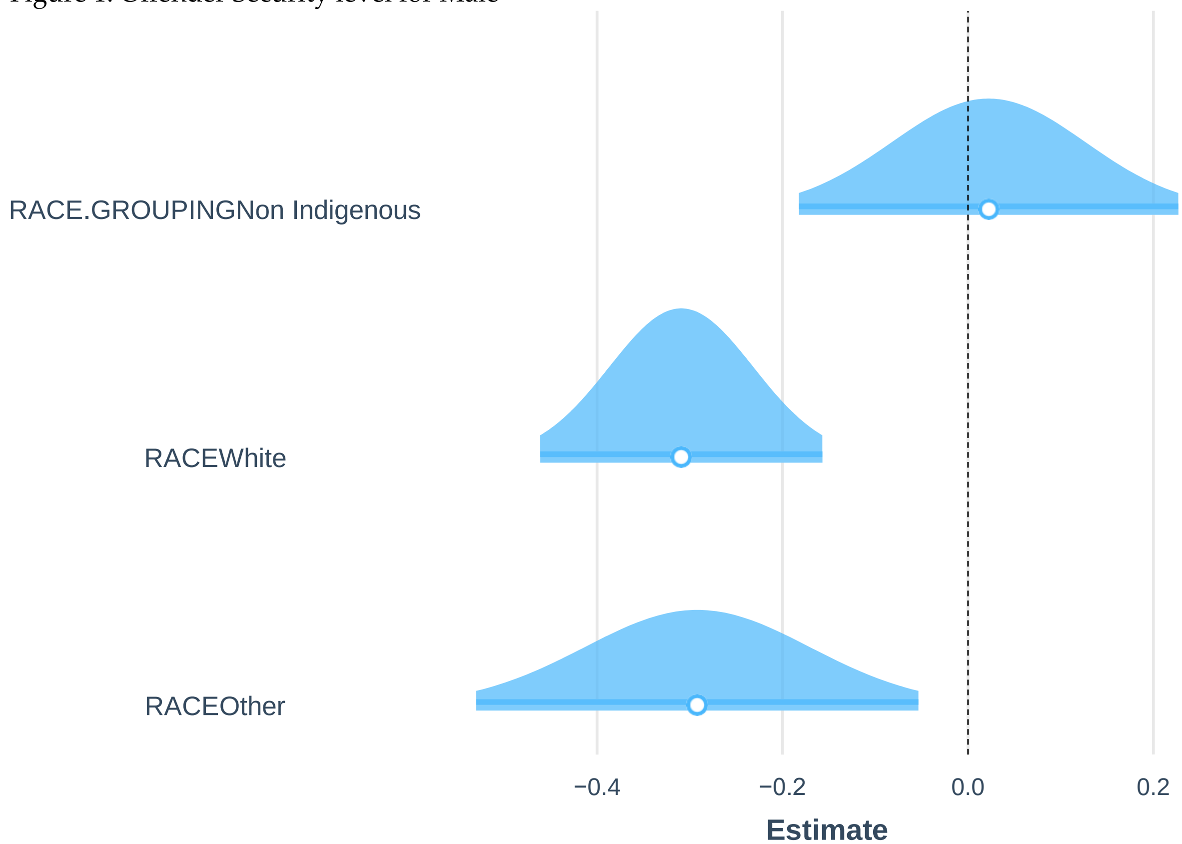


Figure 2: Offender Security level for Female

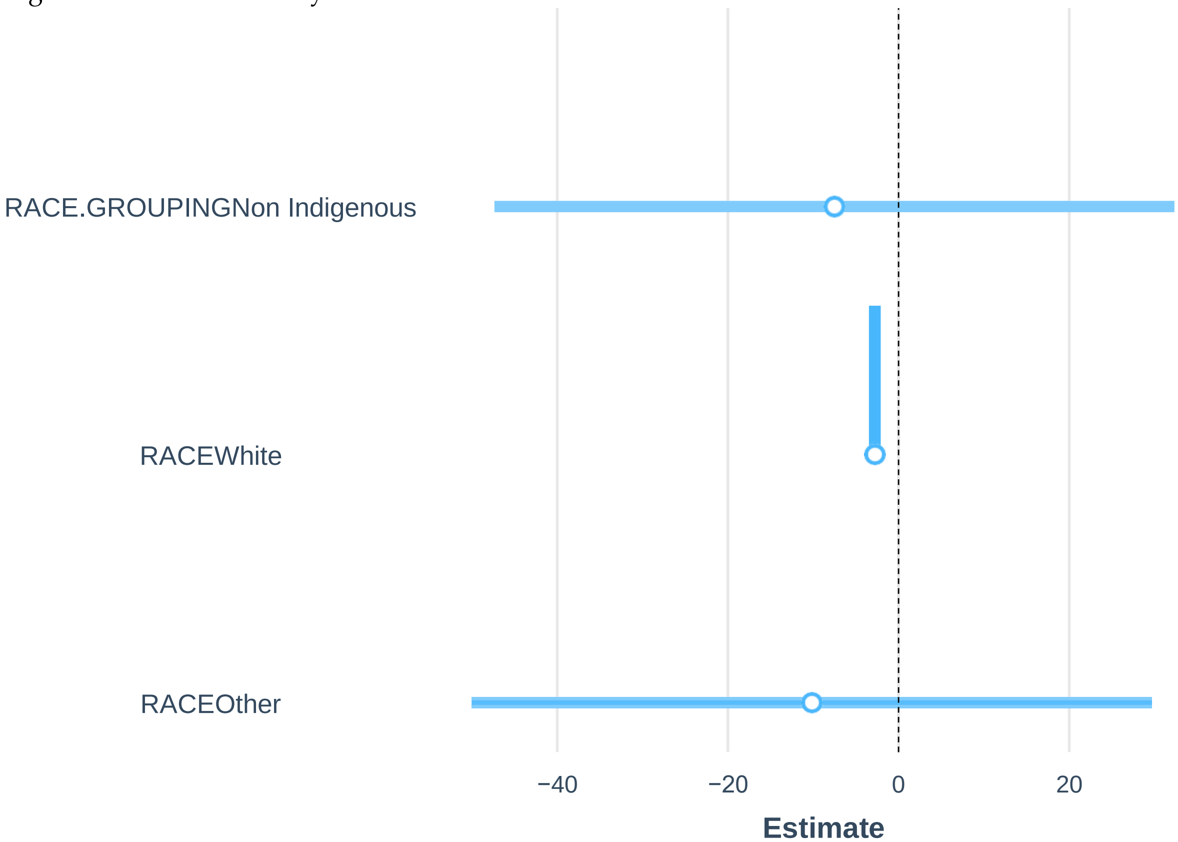


Figure 3: Reintegration Potential for Male

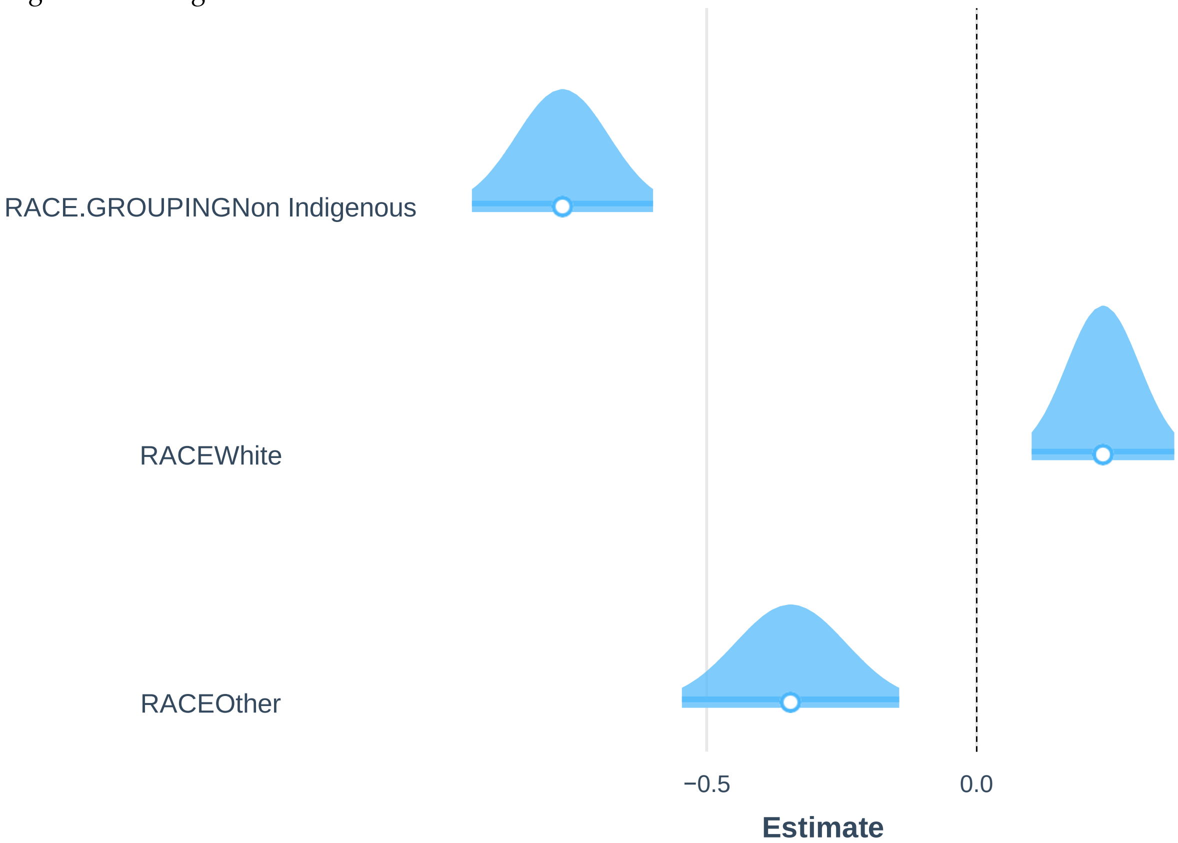
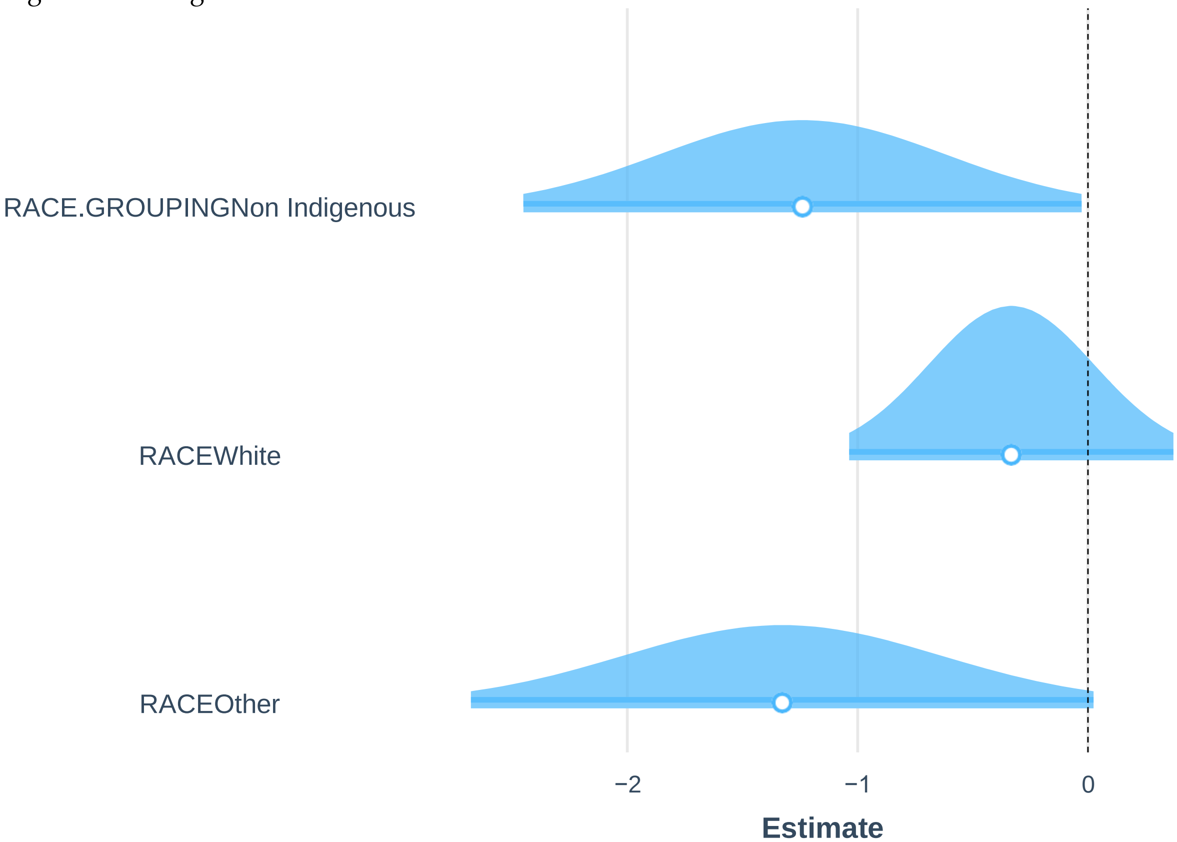


Figure 4: Reintegration Potential for Female



Using the plot_summs() function from the jtools package, we plot our models. The figures provide us with a visualization of estimates, the disribution around these estimates and their 95% confidence intervals. It relays the log-odds estimates of each racial background's likelihood of being deterministic of their risk assessment scores.

##5. Discussion

##5.1 Weaknesses

##citations ##Appendix