Racial Discrimination Behind Bars: Evidence from CSC Risk Scores

Jiang Anan

November 19, 2020

Abstract

Recent anti-racism movements around the world have brought the existing systemic discrimination against visible minorities back to public notice. Inspired by the investigation of the potential biased federal inmates' risk assessments documented in The Globe and Mail (Cardoso 2020), which reveals the discriminating practices on black and indigenous people, I followed the statistical methodology that the author published and attempt to reproduce the analysis. I study how race is associated with two particular risk indicators - the security level and reintegration score using Ordinal Logistic Regression. I present two main results. First, I find evidence that black men are more likely than white men to receive maximum security level scores and low reintegration potential score. Second, black and indigenous women are more likely than white women to end up with maximum security level. There is also higher possibility that black women would get a low reintegration potential score. These findings suggest that the racial discrimination creates inequality in prison and would have an impact on inmates' time behind the bars.

Keywords

Race, Discrimination, Prison, Risk Assessment, Ordinal Logistic Regression

1. Introduction

In 2020, worldwide anti-racism movements have raised awareness about discrimination against people of color. Racism has always been a social crisis. Persistent racial injustice represents the inequality in all aspects of people's life, from education to healthcare, employment and justice system. With the analysis done by The Globe and Mail (Cardoso 2020), I looked into the systemic discrimination in prison. The author used inmates' federal risk assessment as a measurable set of indicators for evaluating inequality behind bars. Security level and reintegration potential score are two specific risk indexes due to the impact they have on prisoners' length of sentence in prison.

In this paper I utilize the dataset obtained from The Globe and Mail (Cardoso 2020) to identify the relationship between race and risk indicators. I used ordinal logistic regression to build models on categorical risk indicators and three main races that are mentioned in the article.

In section 2, I present the data collection and data cleaning process. Because of the access to limited amount of data and information, the data cleaning result is slightly different from the cleaned data which the article used. In section 4, I interpret the model and generate the results. It is shown that there are slight differences in analysis outcome between my results and the ones presented in the article. In discussion section, I clarify the potential causes of the difference and the validation of the model.

2. Data Collection & Data Cleaning

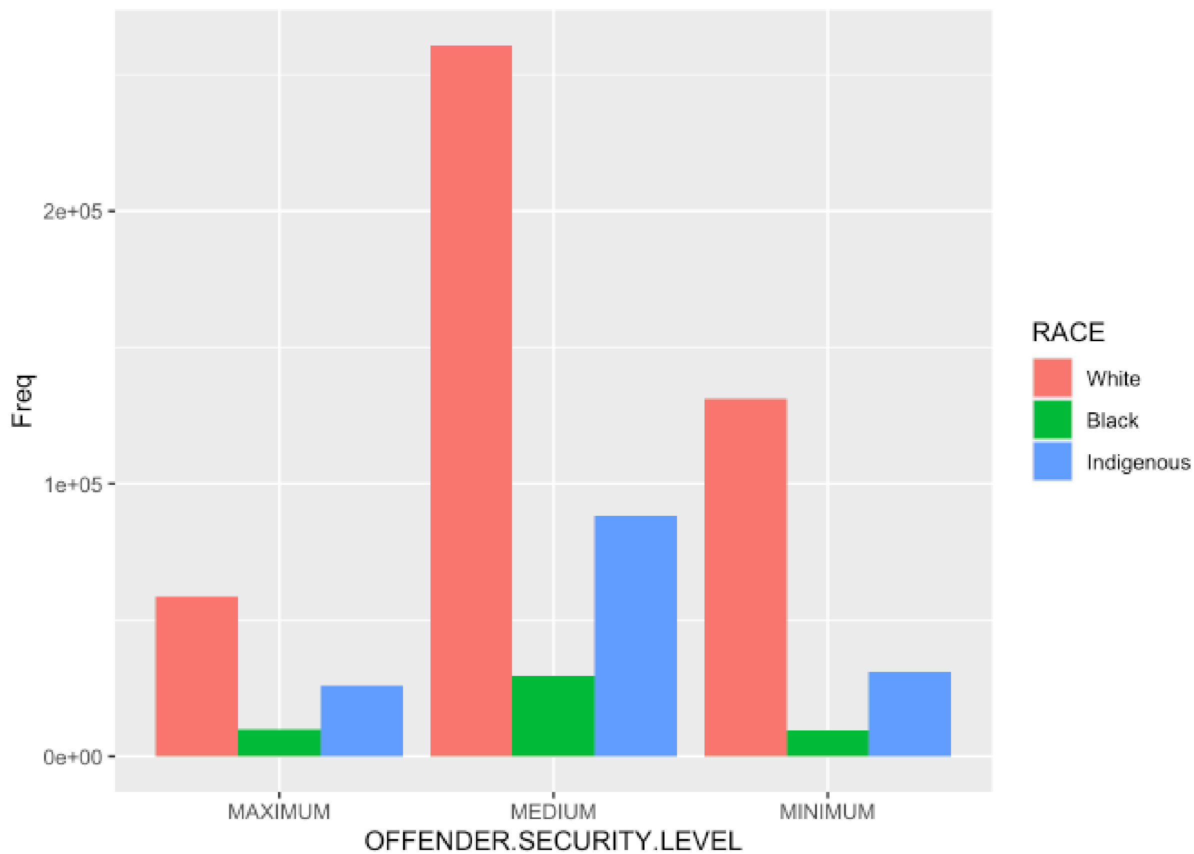
The data used in this paper is collected from The Globe and Mail (Cardoso 2020). The dataset contains detailed information on inmates' demography and risk assessment indicators. The data cover the period from 2012 to 2018. The unit of observation in the data is the jail-enter record. The same inmate may enter the data for multiple times when he or she re-enter the prison. So there could be multiple entries for same person. For each entry, there is race, gender, age, sentence length, security level, reintegration potential, etc. I focus on the demographic information, security level and reintegration potential, which it is reasonable because security level and reintegration score are two major indicators, and reintegration score is calculated by taking dynamic need and static risk into account. Table I, II, III and IV provide the descriptive statistics of the gender ratio, indigenous people ratio and race ratio in the dataset, respectively .

The approach is to focus on the effect of being black, indigenous or white on their risk assessment scores, summarized as two particular risk scores – security level and reintegration level. Therefore, I performed data cleaning on raw data and created a new dataset containing race, age, security level and reintegration potential for a more straight forward representation.

Data cleaning process consists of categorizing races except white, indigenous and black into "other" category, removing entries with missing value in specific columns and split the dataset by gender into two subset – male dataset and female dataset, since the regression is run separately on male and female data.

Graph I and graph II show the distribution of races under three ordered category in offender security level and reintegration potential, respectively.

Graph I. Race by Offender Security Level



Graph II. Race by Reintegration Potential Level

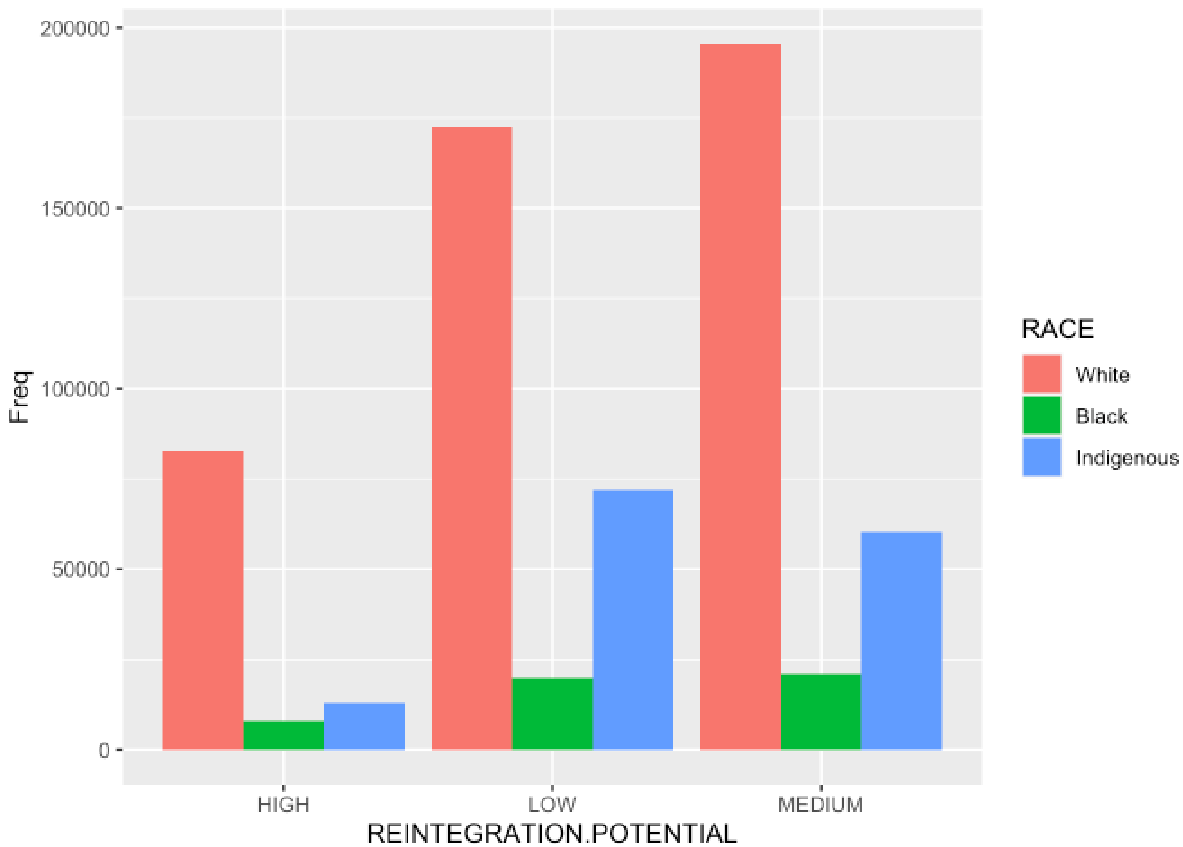
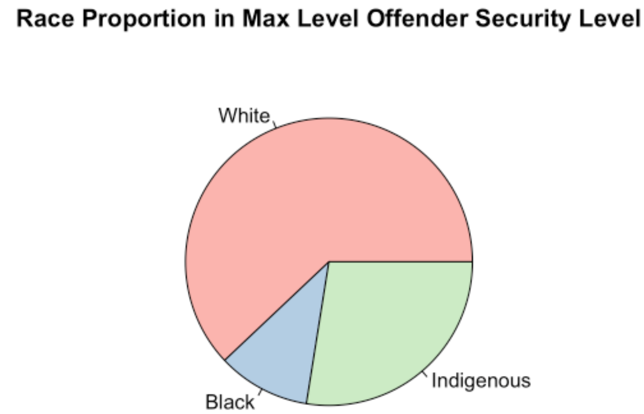


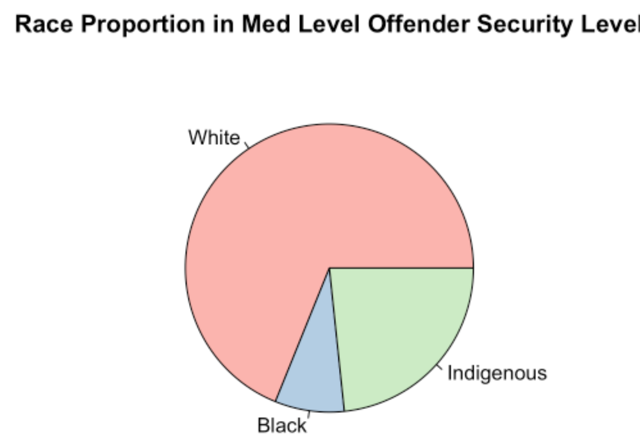
Table I uses pie chart to present the ratio of race in offender security level by levels. Table II present the ratio of three security levels by race. Table III shows the race ratio in different reintegration levels. Table IV shows the ratio of reintegration level in different races.

Table I. Race Ratio by Security Level

Race	Proportion
White	0.6203663
Black	0.1044727
Indigenous	0.2751610



Race	Proportion
White	0.68880034
Black	0.07781114
Indigenous	0.23338852



Race	Proportion
White	0.7647093
Black	0.0553495
Indigenous	0.1799412

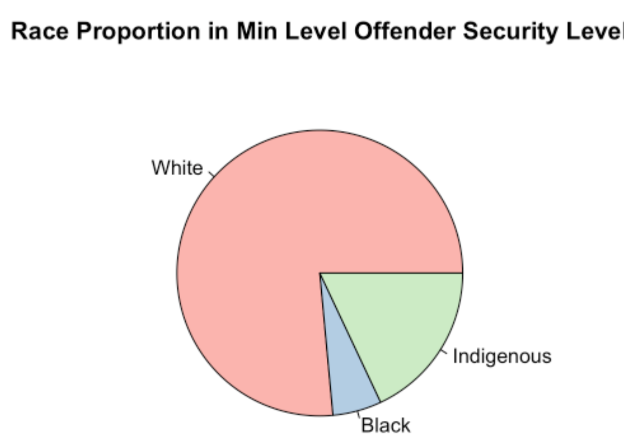
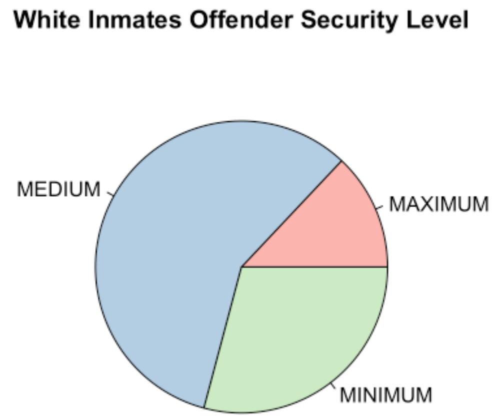
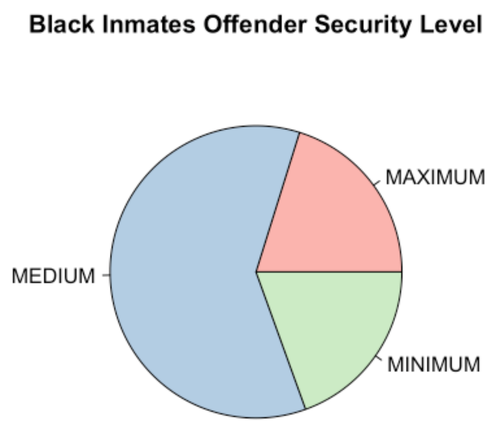


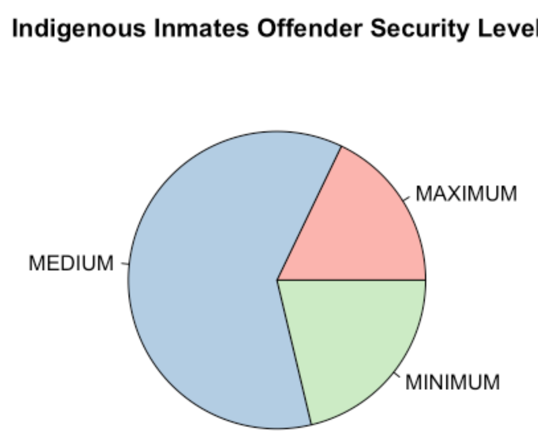
Table II. Security Level by Race



Offender.Security.Level	Proportion
MAXIMUM	0.1297837
MEDIUM	0.5787966
MINIMUM	0.2914197

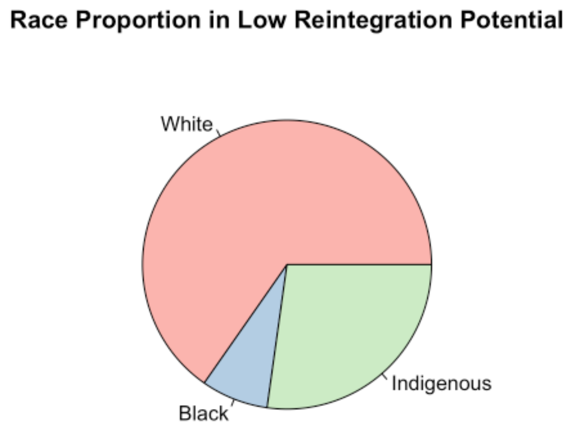


Offender.Security.Level	Proportion
MAXIMUM	0.2017492
MEDIUM	0.6035475
MINIMUM	0.1947033

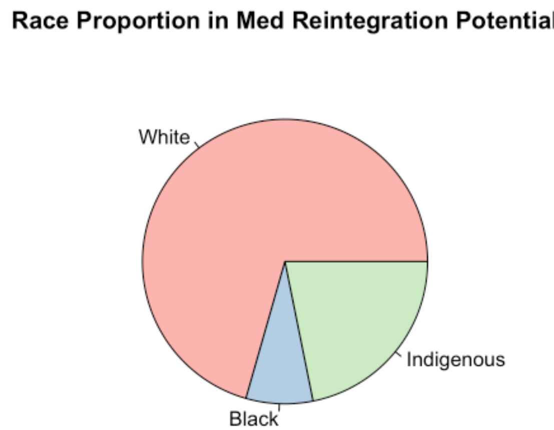


Offender.Security.Level	Proportion
MAXIMUM	0.1786327
MEDIUM	0.6085753
MINIMUM	0.2127920

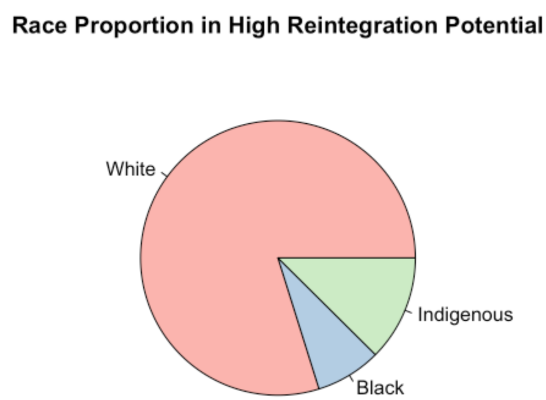
Table III. Race Ratio by Reintegration Potential



Race	Proportion
White	0.65292902
Black	0.07503143
Indigenous	0.27203956



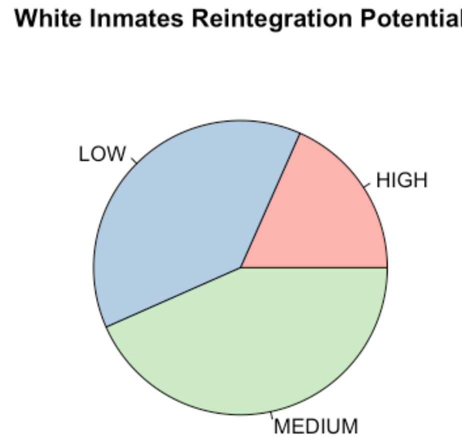
Race	Proportion
White	0.70579699
Black	0.07584312
Indigenous	0.21835989



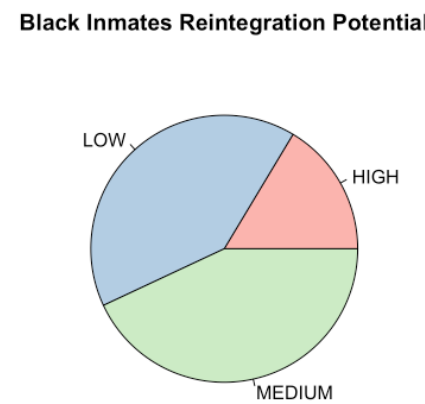
Race	Proportion
White	0.79835915
Black	0.07718739
Indigenous	0.12445345

Table IV. Reintegration Potential by Race

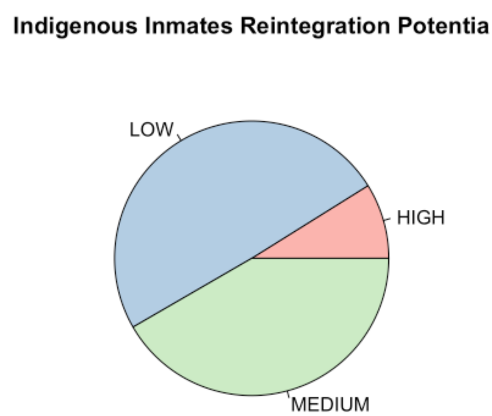
Reintegration.Potential	Proportion
HIGH	0.1835344
LOW	0.3826477
MEDIUM	0.4338179



Offender.Security.Level	Proportion
HIGH	0.1637958
LOW	0.4058948
MEDIUM	0.4303095



Offender.Security.Level	Proportion
HIGH	0.0887827
LOW	0.4947291
MEDIUM	0.4164882



3. Model

The work of Tom Cardoso (2020) show that race is a important element that affects inmates' risk assessment score. The approach is to examine whether being black or indigenous would bias the risk assessment process and to measure the bias prisoners face. I examine the effect of race on inmate's security level and the effect of race on reintegration score.

The offender's security level is based on three factors: how the offender will adjust to the institution, the risk of escape and public safety (CSC 2019). It is set to three classification from minimum to maximum. Maximum security level is described as the worst score, leading to locating at a maximum security institution or a longer time behind bars. The reintegration score is the likelihood of successfully re-entering society without committing a new offence. It is organized into three level- high, medium and low, with low being the worst score (Cardoso 2020).

The outcome variables- security level and reintegration potential are categorical but also they follow an order from a low level to a high level on Likert scale. Thus, I use ordinal logistic regression model to analyze the effect of race. The first model is examining the effect of race on offender security level. The offender security levels are ordered from minimum to maximum.

$$logit\left(P\left(Y\leq 1\right)\right)=\beta_0-\beta_1\,Race-\beta_2\,Age-\beta_3\,Reintegration\ Potential \quad (1)$$

$$logit\left(P\left(Y\leq 2\right)\right)=\beta_4-\beta_1\,Race-\beta_2\,Age-\beta_3\,Reintegration\ Potential \quad (2)$$

The Y variable is the ordinal outcome offender security level, which is set to three level- minimum, medium and maximum. In equation (1), $P\left(Y\leq 1\right)$ is the cumulative probability of being placed in minimum security level. The right hand side of equation (1) can be interpreted as the log odds of receiving a minimum security classification. The right hand side of equation (2) can be interpreted as the log odds of receiving a minimum or medium security classification versus receiving a maximum security classification.

The second model is to examine the effect of race on reintegration potential, with reintegration score being the dependent variable. The reintegration potential categories are ordered from low to high.

$$logit\left(P\left(Z \leq 1\right)\right) = \alpha_0 - \alpha_1 Race - \alpha_2 Age - \alpha_3 Offender\ Security\ Level \quad (3)$$

$$logit\left(P\left(Z \leq 2\right)\right) = \alpha_4 - \alpha_1 Race - \alpha_2 Age - \alpha_3 Offender\ Security\ Level \quad (4)$$

The Y variable is the ordinal outcome reintegration potential score, which is set to three level- low, medium and high. In equation (3), $P(Z \leq 1)$ is the probability of being placed in low reintegration potential level. The right hand side of equation (3) can be interpreted as the log odds of receiving a low reintegration potential classification. The right hand side of equation (4) can be interpreted as the log odds of receiving a low or medium reintegration potential level versus receiving a high reintegration potential level.

In this set-up, I ran the both two models on male and female dataset and we can infer the effect of race on inmates' risk assessment score.

4. Results

In this section, I will present the findings from the models I ran in section 3. First I analyze the effect of race on security level score. Then I analyze how race effects the reintegration potential score.

For the first model, I set the reference level as white people with medium reintegration potential score. As shown in Table V, the coefficients of black and indigenous are 0.22 and -0.19, respectively. It means that black men are more likely to receive a maximum security classification and indigenous men are more likely to receive a minimum security classification compared to white men. In Table VI, for black men, the odds of being placed in maximum security category is 1.25 times (i.e. 24.52% more likely) that of white men, holding all other variables constant. For indigenous men, the odds of being placed in maximum security category is 17% lower than white men, holding all other variables constant.

For female inmates, both black women and indigenous women are more likely to receive maximum security classification compared to white women. The odds of being more likely to receive a maximum security classification for black women and indigenous women are 1.07 (i.e. 7% more likely) and 1.48 (i.e. 48% more likely) times that of white women, respectively. And all the results are statistically significant based on the valid p-values.

Table V. Summary Statistics of Security Level and Race in Male

	Value	Std. Error	t value	p value
RACEBlack	0.2192642	0.0105730051	20.73812	1.569489e-95
RACEIndigenous	-0.1882170	0.0068899959	-27.31743	2.633371e-164
AGE	-0.0504092	0.0002317483	-217.51701	0.000000e+00
REINTEGRATION.POTENTIALHIGH	-1.8619043	0.0087992360	-211.59841	0.000000e+00
REINTEGRATION.POTENTIALLOW	2.3613420	0.0078414022	301.13772	0.000000e+00
MINIMUM|MEDIUM	-2.9843641	0.0113122132	-263.81788	0.000000e+00
MEDIUM|MAXIMUM	1.2134459	0.0113815377	106.61528	0.000000e+00

Table VI. Odds Ratio of Security Level and Race in Male

RACEBlack	RACEIndigenous
1.2451602	0.8284349
AGE	REINTEGRATION.POTENTIALHIGH
0.9508403	0.1553765
REINTEGRATION.POTENTIALLOW	
10.6051741	

Table VII. Summary Statistics of Security Level and Race in Female

	Value	Std. Error	t value	p value
RACEBlack	0.07050851	0.058153119	1.212463	2.253352e-01
RACEIndigenous	0.39350387	0.026909312	14.623334	1.993909e-48
AGE	-0.04134098	0.001239706	-33.347391	7.947514e-244
REINTEGRATION.POTENTIALHIGH	-1.95156042	0.038694597	-50.434959	0.000000e+00
REINTEGRATION.POTENTIALLOW	2.59635978	0.039649282	65.483147	0.000000e+00
MINIMUM|MEDIUM	-1.67047136	0.049700504	-33.610753	1.168411e-247
MEDIUM|MAXIMUM	1.75803896	0.053982710	32.566705	1.214818e-232

Table VIII. Odds Ratio of Security Level and Race in Female

RACEBlack	RACEIndigenous
1.0730537	1.4821650
AGE	REINTEGRATION.POTENTIALHIGH
0.9595019	0.1420522
REINTEGRATION.POTENTIALLOW	
13.4148162	

The second model shows the association of race and reintegration potential. The reference level is set as white people with medium offender security level. In Table IX, the coefficient of black is 0.09 and the coefficient of indigenous is -0.58, which indicates that black men are more likely to receive a high reintegration score, and indigenous men are more likely to receive a low reintegration score, holding all other variables constant. The odds of ending up with a high reintegration score for black men is 9.22% higher than that of white men. The odds of ending up with a high reintegration score for indigenous men is 43.78% lower than that of white men. The patterns are the same for female inmates, with black women being 33.19% more likely to receive a high reintegration score and indigenous women being 31.34% less likely to receive a high reintegration score.

Table IX. Summary Statistics of Reintegration Potential and Race in Male

	Value	Std. Error	t value	p value
RACEBlack	0.08821708	0.0100376458	8.788622	1.514054e-18
RACEIndigenous	-0.57574067	0.0065849283	-87.433096	0.000000e+00
AGE	-0.01935080	0.0002107599	-91.814443	0.000000e+00
OFFENDER.SECURITY.LEVEL.L	-3.19057989	0.0082392848	-387.239909	0.000000e+00
OFFENDER.SECURITY.LEVEL.Q	0.34237491	0.0051479555	66.506968	0.000000e+00
LOW|MEDIUM	-1.28379909	0.0099379634	-129.181306	0.000000e+00
MEDIUM|HIGH	1.58754915	0.0100569245	157.856326	0.000000e+00

Table X. Odds Ratio of Reintegration Potential and Race in Male

RACEBlack	RACEIndigenous	AGE
1.0922252	0.5622882	0.9808352
OFFENDER.SECURITY.LEVEL.L	OFFENDER.SECURITY.LEVEL.Q	
0.0411480	1.4082882	

Table XI. Summary Statistics of Reintegration Potential and Race in Female

	Value	Std. Error	t value	p value
RACEBlack	0.286596795	0.052651402	5.443289	5.230568e-08
RACEIndigenous	-0.375985201	0.026750446	-14.055287	7.149139e-45
AGE	-0.003671369	0.001157621	-3.171476	1.516662e-03
OFFENDER.SECURITY.LEVEL.L	-3.103875481	0.037865345	-81.971403	0.000000e+00
OFFENDER.SECURITY.LEVEL.Q	0.158122501	0.024827253	6.368908	1.903784e-10
LOW|MEDIUM	-1.591095924	0.048958173	-32.499087	1.098422e-231
MEDIUM|HIGH	2.252431213	0.050094976	44.963215	0.000000e+00

Table XII. Odds Ratio of Reintegration Potential and Race in Female

```
               RACEBlack           RACEIndigenous                      AGE
              1.33188708               0.68661249               0.99633536
OFFENDER.SECURITY.LEVEL.L OFFENDER.SECURITY.LEVEL.Q
              0.04487495               1.17130967
```

5. Discussion

While anti-racism social activities are gaining more and more attention in recent years, there is still discrimination against people of color in all kinds of fields of life. In this paper, I address the effect of race on inmates' risk scores by analyzing data collected from Correctional Service of Canada covering seven-year entries from 2012 to 2018.

This inmates' risk assessment dataset is consist of various of categorical variables and some numeric variables. The characteristic of variables inspired me to use ordinal logistic regression to fit the model. The results of the regression are all statistically significant at a 5% level. And I use AUC to measure the test performance. AUC results are included in the Appendix.

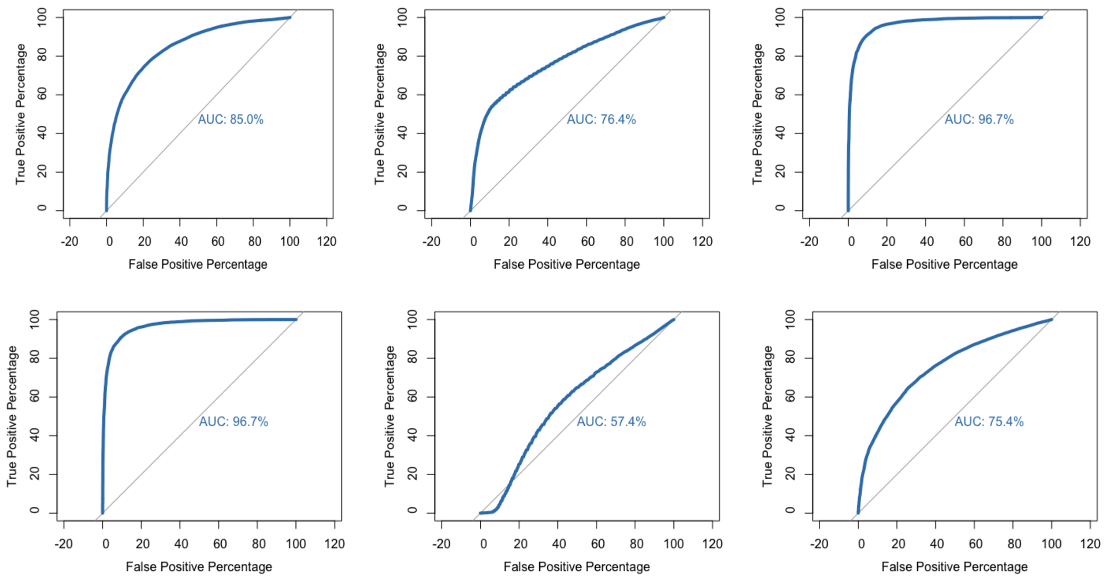
As shown in section 4, in ordinal logistic regression, we can see the bias that prisoners face are solely based on their races. For offender security level, both black men and women are more likely to receive a maximum security classification compared to white people, while indigenous men are less likely to be placed in maximum classification but indigenous women are facing the same bias as black women. When we look into the race bias in reintegration level, it is found surprisingly that black men and women are more likely to receive high reintegration score compared to white men. Although things are opposite for indigenous people, indigenous people regardless of their gender, are more likely to have low reintegration score.

The direction of the effect from my analysis is similar to the results from the article (Cardoso 2020). But there are slight differences in the magnitudes of the effect, which I believe is caused by the data cleaning process. With limited information about the risk score system, I can't look into the score calculation process, which might lead to collinearity. And there is no information about inmates' re-entering record, which would make me not be able to filter out certain entries like what the author did in the article (Cardoso 2020).

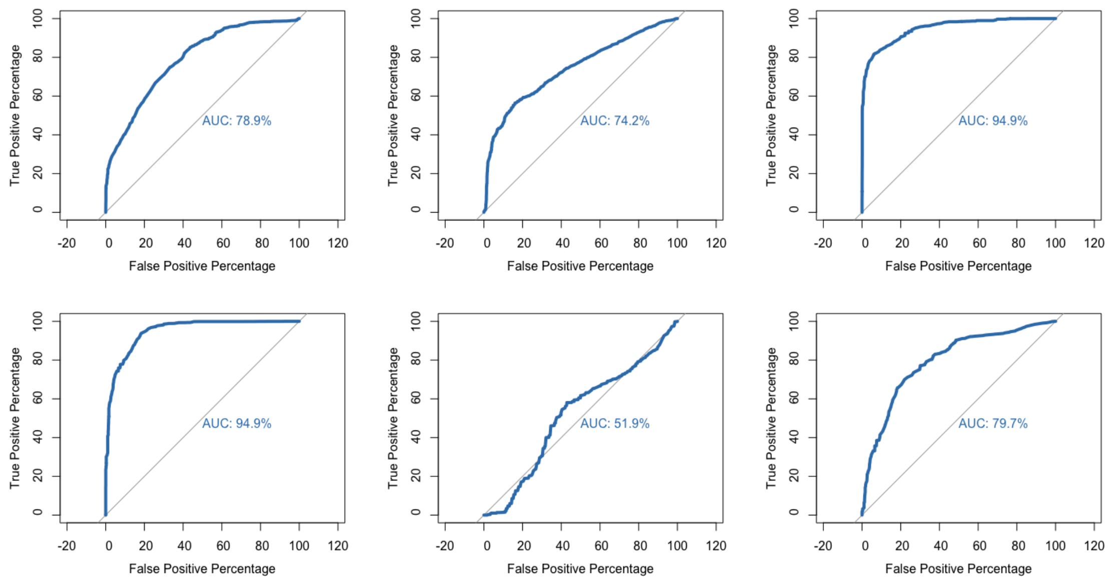
This study is an evidence of the existing of racial discrimination in governmental justice system. I believe this study will be an insightful information for discovering and understanding the racial discrimination and improvement of greater racial and ethnic diversity.

Appendix

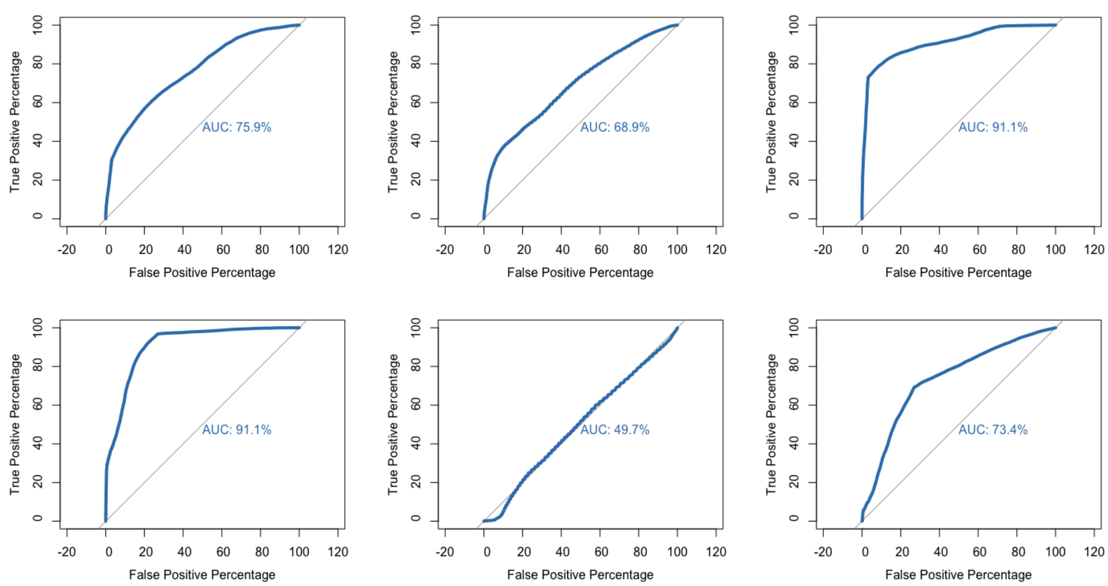
Appendix Graph I. AUC of Offender Security Level and Race for Male



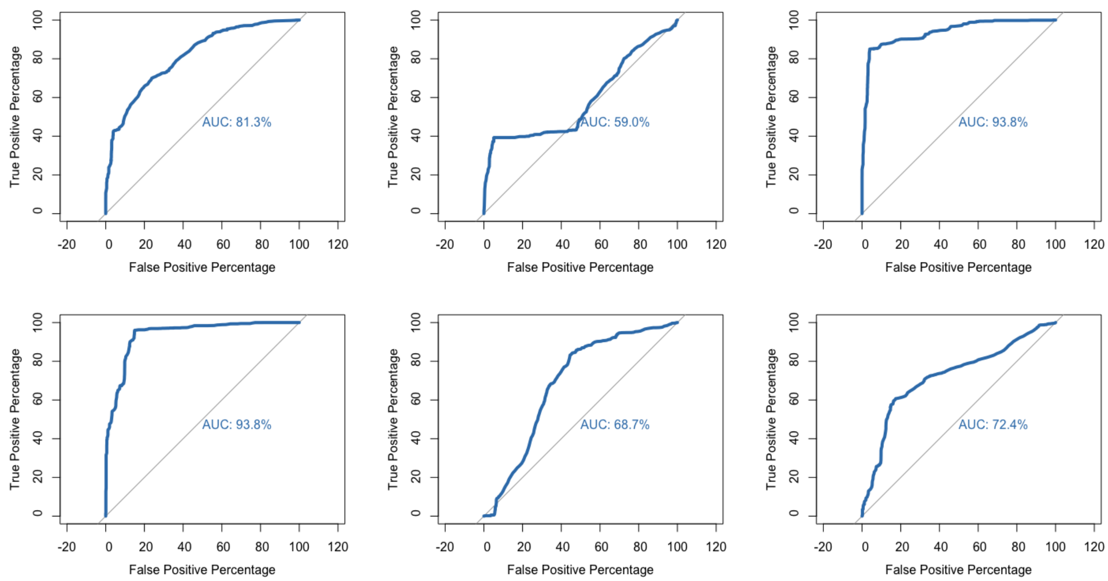
Appendix Graph II. AUC of Offender Security Level and Race for Female



Appendix Graph III. AUC of Reintegration Potential and Race for Male



Appendix Graph IV. AUC of Reintegration Potential and Race for Female



Code for this project can be found at:

https://github.com/Anan-Jiang/sta304_Final_project.git

References

Government of Canada, C. (2019, August 02). Security classifications. Retrieved from https://www.csc-scc.gc.ca/security/001003-1000-eng.shtml

Hadley Wickham (2020). tidyr: Tidy Messy Data. R package version 1.1.2. https://CRAN.R-project.org/package=tidyr

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. https://ggplot2.tidyverse.org

Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686

Lionel Henry and Hadley Wickham (2020). rlang: Functions for Base Types and Core R and 'Tidyverse' Features. R package version 0.4.9. https://CRAN.R-project.org/package=rlang

Matt Dowle and Arun Srinivasan (2019). data.table: Extension of `data.frame`. R package version 1.12.8. https://CRAN.R-project.org/package=data.table

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 0.8.5. https://CRAN.R-project.org/package=dplyr

Hadley Wickham (2019). forcats: Tools for Working with Categorical Variables (Factors). R package version 0.4.0. https://CRAN.R-project.org/package=forcats

Gregory Demin (2020). expss: Tables, Labels and Some Useful Functions from Spreadsheets and 'SPSS' Statistics. R package version 0.10.2. https://CRAN.R-project.org/package=expss

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0. http://www.stats.ox.ac.uk/pub/MASS4

Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller (2011). pROC: an open- source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, p. 77. DOI: 10.1186/1471-2105-12-77 http://www.biomedcentral.com/1471-2105/12/77/

Yihui Xie (2020). tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents. R package version 0.28.

Yihui Xie (2019) TinyTeX: A lightweight, cross-platform, and easy-to-maintain LaTeX distribution based on TeX Live. TUGboat 40 (1): 30--32. http://tug.org/TUGboat/Contents/contents40-1.html