Examining the Effectiveness of a Written Methodology Discussion based on a Reproducibility Paper

Matthew Caringi

9 December 2020

Abstract

In this paper, the examination and reproducibility of a written methodology discussion is preformed to view its effectiveness and ability to maintain the author's originality. Assumptions on data filtering, variable grouping, and type of linear regression model will be observed to define the differences between the results of this article to the original article. Preliminary results reveal similar trends to the original article thus supporting this methodology mode. To strengthen this argument, more reproducibility papers must be preformed due to the fact that the ability to reproduce articles are based on the reader's interpretation which is highly variable.

Keywords Reproducibility; Tom Cardoso; Risk assessments; Bayesian linear regression

The following article was written using R (R Core Team (2020)) and the final report was complied with R markdown (Allaire et al. (2020)). The following packages were used in this paper: Tidyverse (Wickham et al. (2019)), ggplot2 (Wickham (2016)).

Introduction

In the Globe and Mail, an article, "Bias behind bars: A Globe investigation finds a prison system stacked against Black and Indigenous inmates" (Cardoso (2020a)), was written describing racial bias through the analysis of the inmate's risk assessments using data from the Correctional Service of Canada. The author/crime and justice reporter, Tom Cardoso, investigated this aspect by observing the inmate's initial assessment scores set by correctional officers to determine the inmate's security level. They noticed how Black and Indigenous inmates were subjected to harder scoring thus resulted in a more stricter security level such as maximum security. To support these results, Cardoso utilized multiple linear regression models of the CSC dataset in order to examine the influence of an inmate's personal characteristics on their security level. Additionally, these initial risk assessment scores judge the inmates on certain aspects which heavily influences the decisions after conviction, for example parole request. Cardoso also makes use of more linear regression models to investigate inmate reintegration back into society. This revealed that Black and Indigenous inmates would reintegrate better and are less likely to reoffend than White inmates. Overall, Cardoso present relevant conclusions on how Black and Indigenous inmates are subjected to more stricter assessments while showing better reintegration potential.

Within research articles, reproducibility is extremely important for the validation of results. When a paper is not reproducible, this decreases the credibility of the paper and results in the article being less impactful. If a paper cannot be reproduced by another person, this is normally due to the original's author lack of instructions in the methodology or in the worst case, the original author's manipulation of the results to favor their hypothesis. To avoid this reduction in credibility, the original work or R code can be linked. Unfortunately, in the professional field this is not a common practice due to the fact that authors do not want others to steal and plagiarism their work. Therefore, in this paper we attempt to reproduce Tom Cardoso's article based on the posted methodology (Cardoso (2020b)). Our research aim examines the

effectiveness of a written methodology to provide adequate instructions for reproducibility with maintaining the author's recognition for their original work.

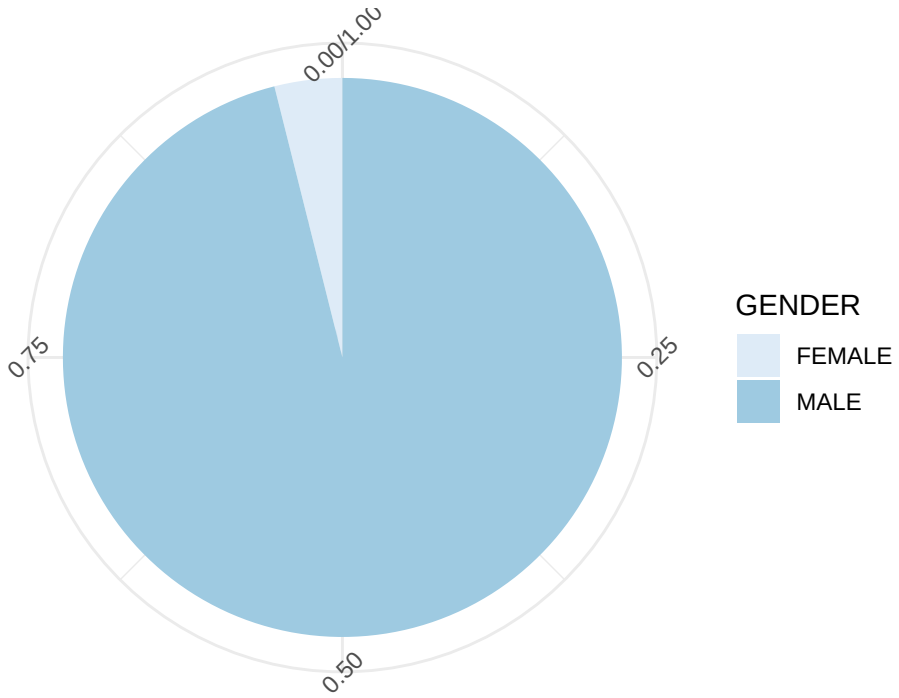
In this paper, we reproduced the data results and analysis found in Tom Cardoso's article based on a written story-like methodology. To note, the R code was not posted or linked thus decisions are based on our assumptions from reading the methodology discussion. Following the methodology, the Correctional Service of Canada dataset provided by the Globe and Mail was used. Filtering and grouping of the dataset were followed as best as possible based on our interpretation and hints from the author such as 'filtering this allowed these many observations/rows'. One major interpretation was the specific type of linear regression model used. In this paper, we used multiple Bayesian linear regression models while it is unclear which type of linear regression model was used in Cardoso's original article.

Data

NOT COMPLETE

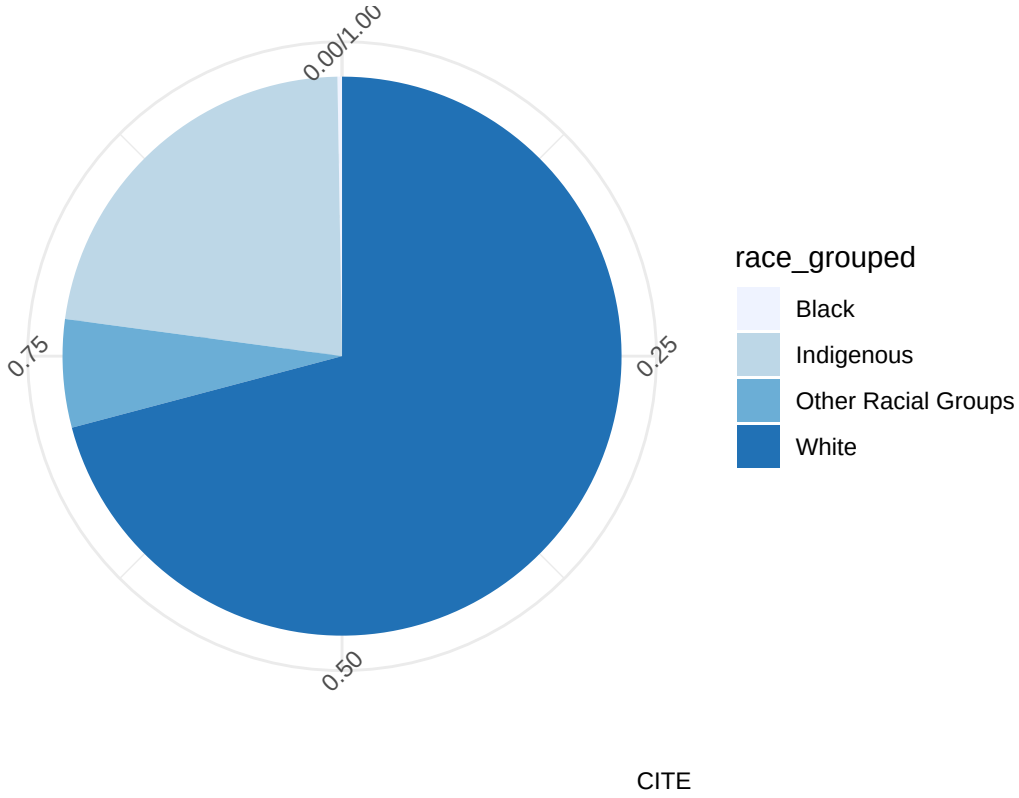
PLAN - touch up graphs ex. add citations, text

Figure 1: Distribution of Gender



CITE

Figure 2: Distribution of Race Grouping



CITE

Figure :

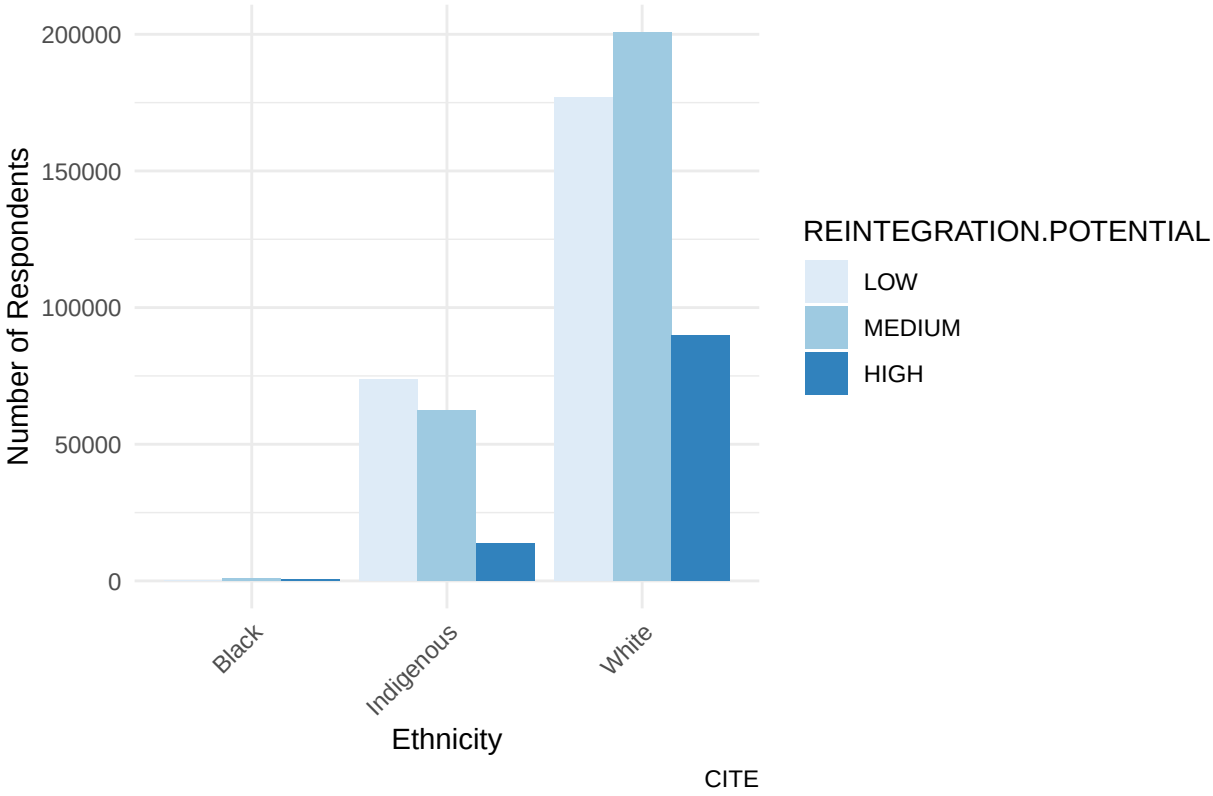
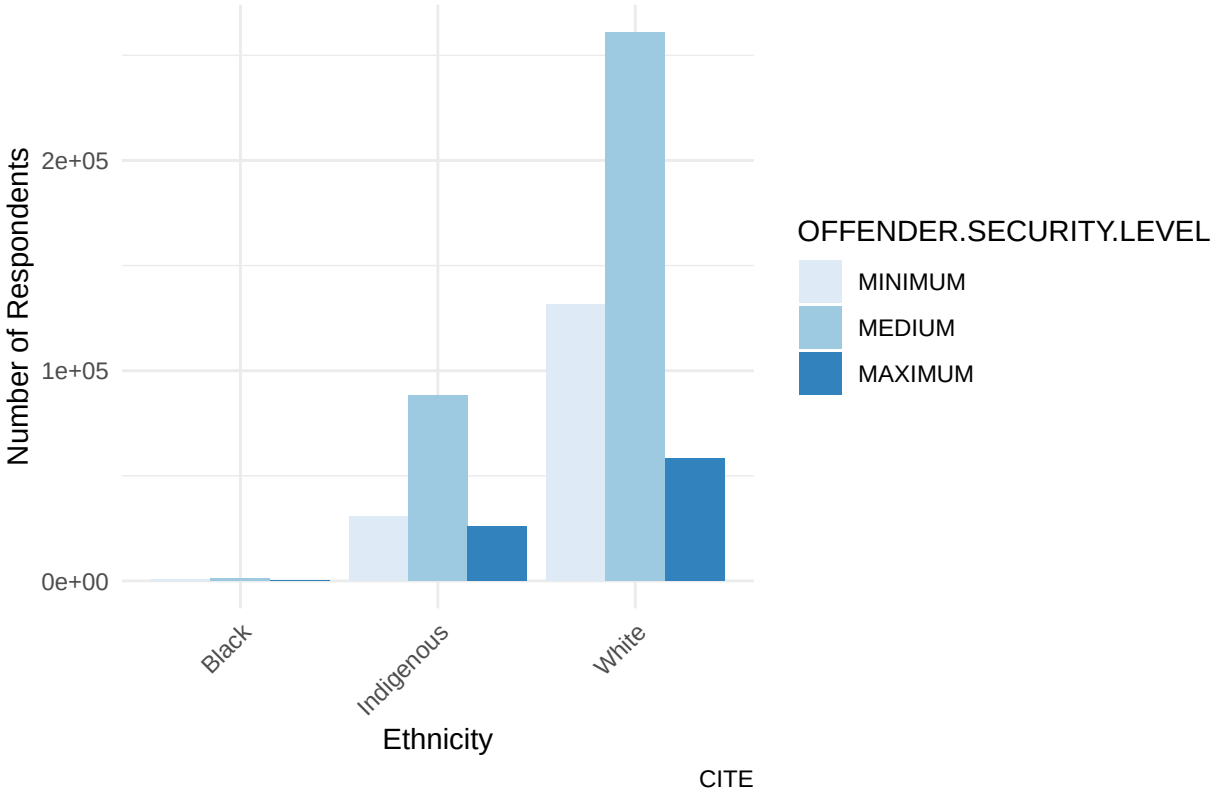

200000
150000
100000
50000
0
Number of Respondents
Black
Indigenous
White
Ethnicity
REINTEGRATION.POTENTIAL
LOW
MEDIUM
HIGH
CITE

Figure :



Model

NOT COMPLETE

PLAN - Bayesian linear regression model - have to weigh the offenses by points

Results

NOT COMPLETE

PLAN - state the data from the graphs (anything off etc) - look at the result of my models (anything off, significant etc)

Discussion

NOT COMPLETE

PLAN - compare data graphs to original article - does my assumptions make a huge difference - do my models have a different outcome - future work - more reproducibility articles from other people (increase sample size to strengten argument) -

References

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2020. *Rmarkdown: Dynamic Documents for R*. https:

//github.com/rstudio/rmarkdown.

Cardoso, Tom. 2020a. "Bias Behind Bars: A Globe Investigation Finds a Prison System Stacked Against Black and Indigenous Inmates." *The Globe and Mail*. https://www.theglobeandmail.com/canada/article-investigation-racial-bias-in-canadian-prison-risk-assessments/.

———. 2020b. "How We Did It: How the Globe Uncovered Systemic Bias in Prisoners' Risk Assessments." *The Globe and Mail*. https://www.theglobeandmail.com/canada/article-investigation-racial-bias-in-canadian-prisons-methodology/.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.