Replicate 'Bias Behind Bars' under Bayesian setting*

Logistic regression on risk scores of prison inmates under Bayesian setting

Su

16 December 2020

Abstract

Racial discrimination against the First Nations has been a problem in Canada since the first colonist step on the sides of St. Lawrence river. While this problem has been a controversial among the general public, is it still the possible case in the scope of prison systems? The Globe and Mail investigation indicates that Black and Indigenous prisoners experience worse assessment scores solely base on their race. In this report, we focus on replicating the results of the Globe's investigation under Bayesian settings ,instead of the original Frequentist settings, using the same dataset from Correctional Service of Canada.
keywords: Logistic Regression, Bias, Racial Discrimination, Correctional Service of Canada, The Globe and Mail

1 Introduction

For majority of the general public, the part of the world behind bars is likely to be one of the least favorable place to be. Not only because that prisons have a less favorable physical environment, but also because it is filled by heinous criminals. This is, to some extent, a known fact. However, are the prisoners really all that heinous? The answer would be no. Although they all did commit severe crime to be put in the prison, but it does not necessary mean they are all heinous since the severity of their crime are all different. In the prison system, there are various scores, designed by the Correctional Service of Canada (CSC), assessing a criminal's potential threat to inmates of the prison, and also the potential ability to reintegrate to the society after their sentence. These scores are crucial to the inmates since these determines the overall correctional plan and will affect the inmates' life after sentence.

In general, these scores are risk assessments, assessing an inmate's future risk to public using all past and present information for design of a correctional plan. The problem arises from these scores, while these measures are crucial to a prisoner's in-prison life and out-prison life, these are not all determined under a standardized and objectective measure, thus bias naturally exist when subjectivity enter. Some of these are measured purely by standardized tests and objective data, but some are purely based on officer's judgment. In case of bias due to subjectivity, it is natural to consider some of the long-lasted bias among the general public like race and gender bias. Racial discrimination against the First Nations have been a problem in North America even before Canada exist as a country, and it would not be surprise if some degree of discrimination exist among prison systems. The Globe and Mail investigation (Cardoso 2020a) identified that there is bias against Black and First Nations (indigenous) inmates among the assessments. The focus of this report is to replicate the Globe's findings using a Bayesian setting instead of the original Frequentist setting, and examine if the findings still hold statistically. We replicate exactly the original procedure using the same dataset originally obtained from CSC (Canada 2018). However, instead of replicating all three logistic models, we only replicate the first two model analyzing the impact of race on security and reintegration scores.

The article introducing the original methodology (Cardoso 2020b) is not a detailed instruction paper and there are many hidden details to the exact methodology. When encountering such an issue, we proceed using our interpretation to the context of the procedure. The final result is slightly different with the article due to

*Code and data are available at: LINK.

these issues, as well as the use of the Bayesian settings. However, the result should agree with the article that there are some bias against Black and Indigenous inmates on CSC assessment scores. We will discuss more detail on the hidden detail interpretation issues in the sections that encounters it. This reports validates the reproducibility on the first two statistical models discussed in "Bias Behind Bars"

In order to replicate the result of the Globe (Cardoso 2020b), we focus on two of the risk scores, the offender security level and reintegration level, for having the largest impact on an inmate's life behind the bars. Both of these scores are assessed by the CSC officers using a series of specialized tests and interviews, which are all assigned when the inmates first arrive at the federal prison. The offender security level is a set of measure from minimum, medium, to maximum. Each inmate with following measure will be assigned to facility or area suiting their security level, for example, an inmate with maximum score will be assigned to area with maximum security level. The reintegration score estimates an inmate's potential to re-enter the society without committing a new offend, and this score is also crucial for the parole hearings. The reintegration is also designed to have low, medium, and high levels, assigned using a set of actuarial and non-actuarial assessments. More detailed discussion of the scores is included in *Section 2*.

For the remaining part of the report, *Section 2* includes discussion on the CSC dataset, and some exploratory data analysis for deeper understanding of the dataset. The specific procedure of the investigation by the Globe and the logistic models used are included in *Section 3*. The model results and comparisons to the original results are included in *Section 4*. Lastly, we make discuss about some potential future improvements of the procedure, and draw conclusion base on the comparison results in *Section 5*.

2 The CSC 2012-2018 Data

The dataset we use is a Correctional Service of Canada (CSC) dataset from their inmate's database, and this dataset was requested through several levels of bureaucracy (inlusding consent from head of CSC) by Tom Cordoso, the author of "Bias Behind Bars" (Cardoso 2020a). The dataset records inmate entries from 2012 to 2018. It is free of all personal information, containing 744958 entries from 50116 inmates and 25 variables in total (technically it contains 26, but the "judge" variable was set all blank by CSC). Specifically, each year's data given by the CSC is a snapshot of their full database on March 31 of the corresponding year. March 31 is the end of CSC's fiscal year, and for that reason, the unit of year in this dataset is fiscal year like YE1112 means the fiscal year 2011-2012, instead of the "common" year unit like 2015.

The variables recorded include the offender security score and reintegration potential score, which are the two response variables we are interested in. As well as the characteristic that should be controlled to achieve less biased results, including age, gender, year recorded, sentence type, and static score. The sentence types are separated into determinate and indeterminate (life sentence), indicating severity of crime to some degree. The static score corresponds to the risk level of the offender, it is assigned using Static Factors Assessment, a CSC tool measuring the inmates's past involvement with the criminal justice system (some are crime records). Thus, a higher static score means the inmate has had a decent record with the criminal justice system in the past, and thus indicates an inmate's crime history to some extent.

The article also generated numerical offence severity, by hand-matching over 700 offence types recorded in this dataset with the Uniform Crime Reporting Survey's offence category codes (product of Statistics Canada). In that way, the new offence codes can be matched with Statistics Canada's weighting system for Crime Severity Index (CSI), which represents a numerical measure of severity of the offence. However, this procedure of hand-matching over 700 offence types is not very realistic given the time constraints of this report especially when it was available later in the time line. Nevertheless, we still thank Rohan Alexander for gathering the complete 2018 CSI weight document from Tom Cordoso (it is a public document, but is not open for download on Statistics Canada). This action will cause our replicate model to have one less control variable to include, which is the most sever offence of an inmate. The resultant issues will be discussed in *Section 5*.

For the remaining of the report, we perform data manipulation, result visualization, and statistical model training in `R` (R Core Team 2020). In the data manipulation steps, the packages `tidyr` (Wickham 2020), `tidyverse` (Wickham et al. 2019), `skimr` (Waring et al. 2020), and `naniar` (Tierney et al. 2020) were used. The replication of data cleaning begins with removing all entries for inmates on provincial jurisdiction. This

step was not an explicitly step written in the methodology instruction, it was included in a bracket, and we assume it is a necessary step to take. However, we already come to a difference with the article results in the first step, the author mentioned he had 741,738 entries after removing the provincial jurisdictions, but our result showed 741829 entries without missing value in jurisdictions. Fortunately this is not a major concern, but it rings the bell that differences should be expected in most further steps.

Next, we implement our usual steps when dealing with raw data. We removed the missing values in the variables we are interested in (three sores, age, gender, year, sentence type). This reduce the dataset size to 686540 entries. For the 34 race categories in the dataset, we replicate the article methodology of classifying Indigenous races to "Indigenous" race category, and classify all races other than White, Black, and Indigenous, into "other" race category. This classification allows the model to focus on differences between White, Black, and Indigenous groups. The Indigenous classification reference we used in directly from the race grouping categories in the CSC dataset, while the article used Ontario's Data Standards for the Identification and Monitoring of Systemic Racism, which is a guide for identifying racial disparities in data. Assuming the CSC follows a similar, and ethic classification system, we should not encounter major differences with the article, if any, in the classification of races.

Although not explicitly mentioned in the article, we recognize a significant amount of entries that have the same fiscal year ID, sentence ID, and offender ID, and the only difference between these entries is the offence ID. This due to multiple offences in one crime action, and the system had to put multiple entries in to accommodate that, we make the assumption that an offender only gets a true record of scores for each sentence ID given same fiscal year. After removing all replicated observations, the datset size is reduced to 142700 entries. In *Section 2.1* and *Section 2.2*, the high degree of similarity in size of the model specific sub-dataset verifies that this action is appropriate, and should be similar to the actual procedure of the article.

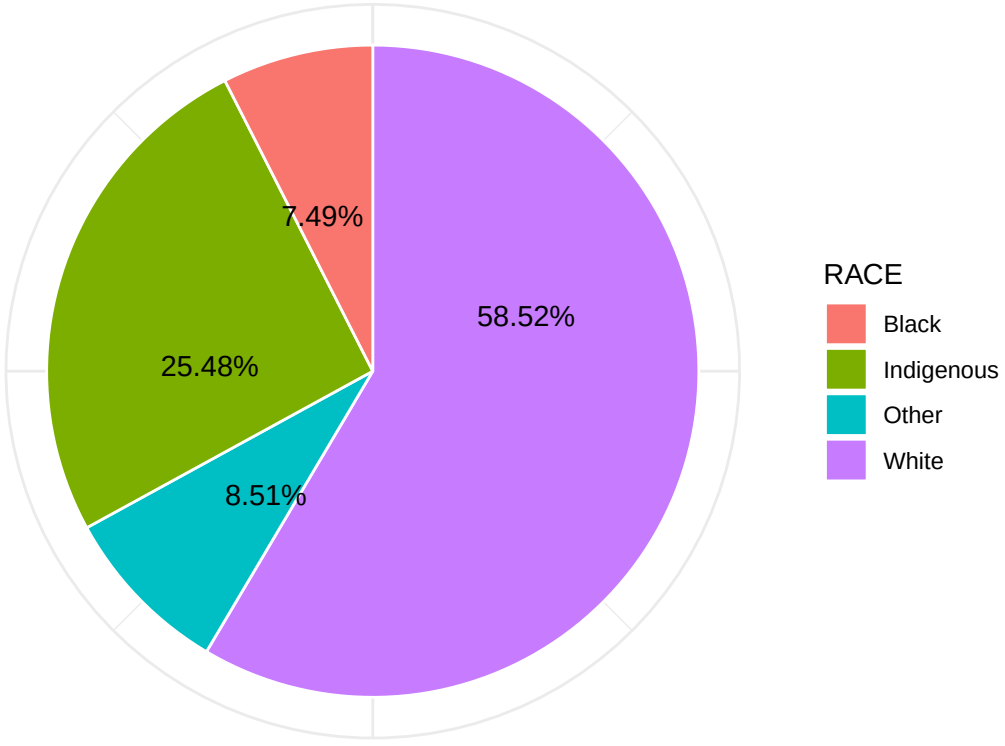


Figure 1: Distribution of race in 2018

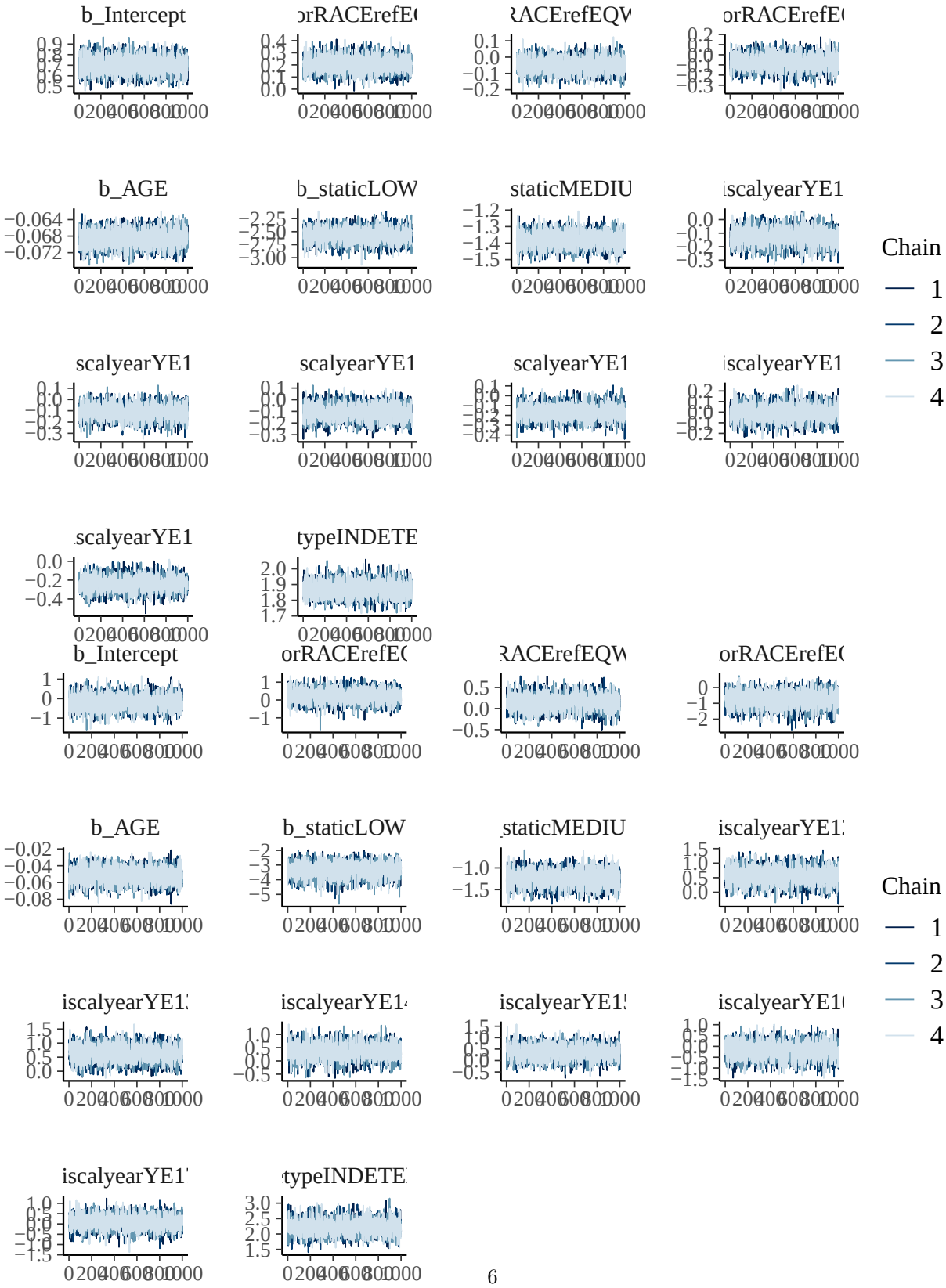
The article tested three models in total, two model assessing impact of race on probability of getting the worst offender security score and reintegration potential score, and one model assessing impact of reintegration score on probability of re-offending. Due to the time constraint of this report and ambiguity in the methodological

instructions of the third model, we only replicated the first two models assessing the impact of race on the two assessment scores. For each model, we subset the CSC dataset into sub-datasets since the offender security score assessments are different from reintegration potential scores.

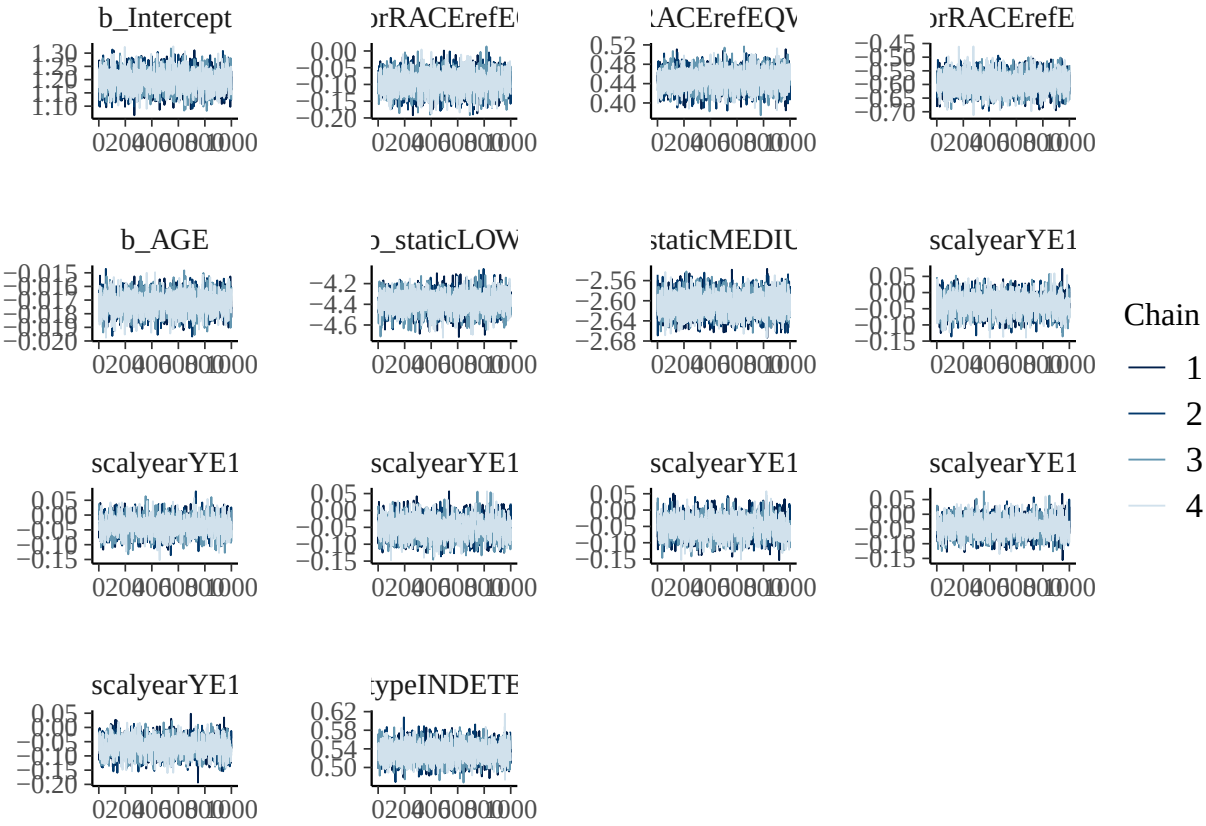
2.1 Security Socre Model Subset

2.2 Reintegration Score Model Subset

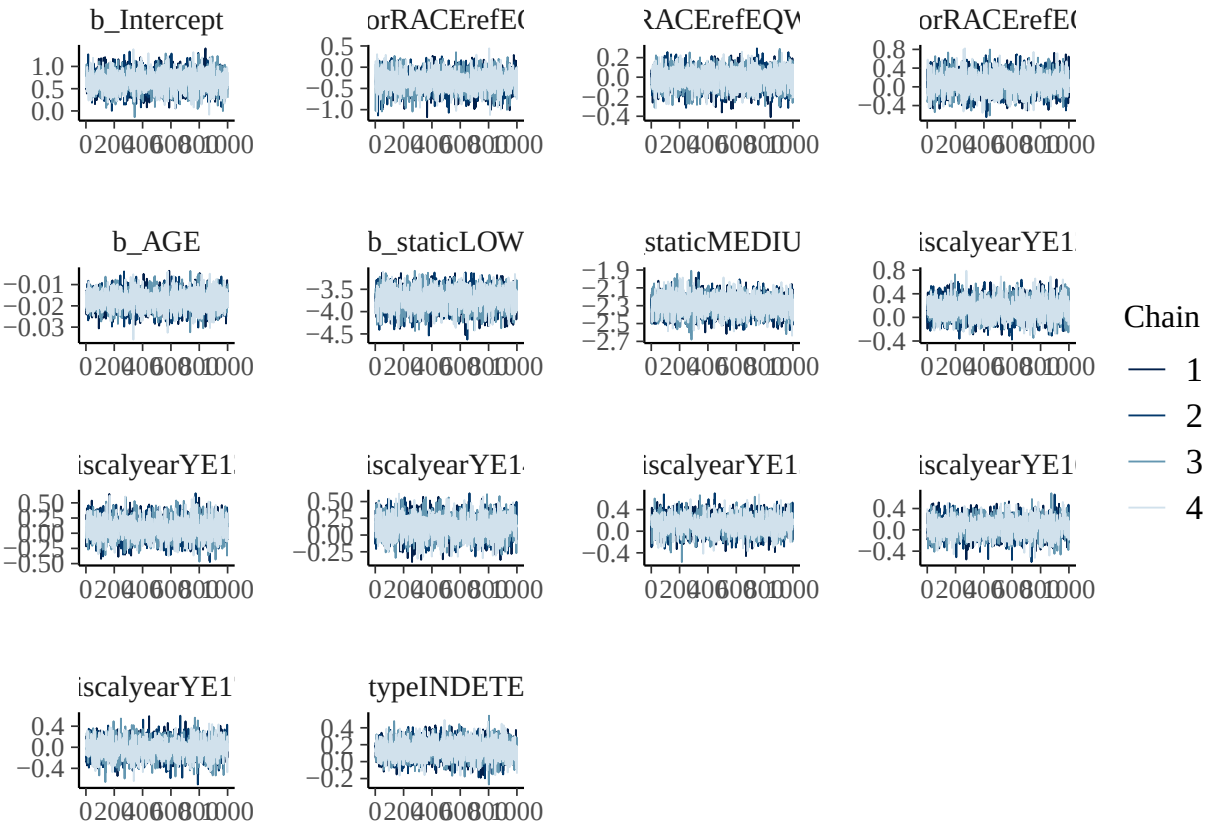
3 Model



```
## No divergences to plot.
```




```
## No divergences to plot.
```



4 Results

4.1 Model 1

4.2 Model 2

5 Discussion and Conclusion

References

Canada, Correctional Service of. 2018. *Commissioner's Directives*. www.csc-scc.gc.ca/acts-and-regulations/005006-0001-en.shtml.

Cardoso, Tom. 2020a. "Bias Behind Bars: A Globe Investigation Finds a Prison System Stacked Against Black and Indigenous Inmates." *The Globe and Mail*. 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/.

Tierney, Nicholas, Di Cook, Miles McBain, and Colin Fay. 2020. *Naniar: Data Structures, Summaries, and Visualisations for Missing Data*. https://CRAN.R-project.org/package=naniar.

Waring, Elin, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu, and Shannon Ellis. 2020. *Skimr: Compact and Flexible Summaries of Data*. https://CRAN.R-project.org/package=skimr.

Wickham, Hadley. 2020. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

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.