

# VCOSS Cleaning Notes 2020

Ben Cole

23 Sep 2020

## Contents

<b>Previous work</b>	<b>1</b>
<b>1 Individual Charities Data</b>	<b>2</b>
1.1 Data Source . . . . .	2
1.2 Data_Cleaning.R script . . . . .	2
1.2.1 Recoding booleans . . . . .	2
1.2.2 Checking for Invalid ABNs . . . . .	2
1.2.3 Main Activity . . . . .	2

## Previous work

The Victorian Council of Social Service (VCOSS) has produced reports in past years using ACNC data cleaned with proprietary software requiring subscriptions. The R language was chosen to clean the data sets for this new VCOSS report so that the cleaning procedures could be transparent and accessible without need for purchasing software.

Previous reports from VCOSS had not cleaned the 2018 ACNC data as it had not been released at the time. Considering the 2018 data needed to be cleaned and filtered to VCOSS guidelines, it was decided to clean the ACNC data from years 2014 - 2018 inclusive for consistency. Some new decisions were made while cleaning the data that produced different data sets than previously used by VCOSS.

# 1 Individual Charities Data

## 1.1 Data Source

The ACNC data was sourced from **the data.gov.au website** for **2014, 2015, 2016, 2017, and 2018** at the beginning of the project.

## 1.2 Data\_Cleaning.R script

As with any script of code, the *Data\_Cleaning.R* script serves as a record of the cleaning procedures as well as a reusable+reproducible means for repetitive cleaning of the data.

### 1.2.1 Recoding booleans

Boolean columns in the data were stored with the values “y” and “n” throughout the years, so were recoded to TRUE and FALSE which are native to R.

### 1.2.2 Checking for Invalid ABNs

This is legacy code from previous work by the Future Social Service Institute and was included in initial stages of cleaning. Subsequent consultation with VCOSS led to the decision being made not to remove charities with invalid ABNs. The code is left in place, but the filter is not applied.

### 1.2.3 Main Activity

Selecting charities based on their reported *Main Activity* is the first filter applied by the script. If a charity reported that they performed any of the activities from the list below they were retained, and any charity that did not was removed.

The following activities varied in their naming structure throughout the years, so variants were deliberately used to catch them. For example, Main Activity in the 2015 data contained

“Civic and advocacy activities”, while the 2016 data contained **both** “Civic and advocacy activities” and “Advocacy and civic activities”.

Aged care activities  
Civic and advocacy activities  
Economic, social, and community development  
Emergency relief  
Employment and training  
Housing activities  
Income support and maintenance  
International activities  
Law and legal services  
Mental health and crisis intervention  
Other education  
Other health service delivery  
Social services