COVID CNS: Cleaning ethnicity

Jessica Mundy

14/01/2022

Set up

```
rm(list=ls())
```

```
source(file = "../functions/add_numeric_1.R")
source(file = "../functions/remove_duplicates.R")
source(file = "../functions/sumscores.R")
source(file = "../functions/package_check.R")
source(file = "../functions/imp_check.R")
```

Note: always load tidyverse last

```
packages = c(
  "summarytools",
  "sjlabelled",
  "Amelia",
  "knitr",
  "gtsummary",
  "tidyverse"
  )
package_check(packages)
```

```
Loading required package: summarytools

Warning: package 'summarytools' was built under R version 4.0.5

Registered S3 method overwritten by 'pryr':
  method      from
  print.bytes Rcpp

Loading required package: sjlabelled

Attaching package: 'sjlabelled'

The following object is masked from 'package:summarytools':

    unlabel
```

```
Loading required package: Amelia

Warning: package 'Amelia' was built under R version 4.0.5

Loading required package: Rcpp

Warning: package 'Rcpp' was built under R version 4.0.5

##
## Amelia II: Multiple Imputation
## (Version 1.8.0, built: 2021-05-26)
## Copyright (C) 2005-2022 James Honaker, Gary King and Matthew Blackwell
## Refer to http://gking.harvard.edu/amelia/ for more information
##

Loading required package: knitr

Warning: package 'knitr' was built under R version 4.0.5

Loading required package: gtsummary

Warning: package 'gtsummary' was built under R version 4.0.5

Loading required package: tidyverse

Warning: package 'tidyverse' was built under R version 4.0.5

-- Attaching packages --------------------------------------- tidyverse 1.3.1 --

v ggplot2 3.3.5     v purrr   0.3.4
v tibble  3.1.5     v dplyr   1.0.7
v tidyr   1.1.4     v stringr 1.4.0
v readr   2.0.2     v forcats 0.5.1

Warning: package 'ggplot2' was built under R version 4.0.5

Warning: package 'tibble' was built under R version 4.0.5

Warning: package 'tidyr' was built under R version 4.0.5

Warning: package 'readr' was built under R version 4.0.5

Warning: package 'purrr' was built under R version 4.0.5

Warning: package 'dplyr' was built under R version 4.0.5

Warning: package 'stringr' was built under R version 4.0.5

Warning: package 'forcats' was built under R version 4.0.5
```

```
-- Conflicts ---------------------------------------- tidyverse_conflicts() --
x forcats::as_factor() masks sjlabelled::as_factor()
x dplyr::as_label()    masks ggplot2::as_label(), sjlabelled::as_label()
x dplyr::filter()      masks stats::filter()
x dplyr::lag()         masks stats::lag()
x tibble::view()       masks summarytools::view()
```

```
date <- Sys.Date()
date
```

```
[1] "2022-02-22"
```

```
source("../credentials/paths.R")
```

Read in the data

```
ethn <- read_rds(file =
                   paste0(ilovedata, "/data_raw/latest_freeze/covid_cns/baseline/dem_covid_cns.rds")
                 )

# check
ethn %>%
  dim()
```

```
[1] 235 133
```

```
ethn %>%
  colnames()
```

```
  [1] "externalDataReference"
  [2] "startDate"
  [3] "endDate"
  [4] "dem.day"
  [5] "dem.month"
  [6] "dem.year"
  [7] "dem.required_question_eligibility_criteria.txt"
  [8] "dem.what_gender_do_you_identify_with"
  [9] "dem.what_gender_do_you_identify_with.txt"
 [10] "dem.do_you_consider_yourself_to_be_transgender"
 [11] "dem.have_you_ever_been_pregnant"
 [12] "dem.what_is_your_sexual_orientation"
 [13] "dem.what_is_your_sexual_orientation.txt"
 [14] "dem.what_is_your_current_maritalrelationship_status"
 [15] "dem.what_is_your_current_maritalrelationship_status.txt"
 [16] "dem.how_would_you_describe_your_vision"
 [17] "dem.how_would_you_describe_your_hearing"
 [18] "dem.which_hand_do_you_usually_write_with"
 [19] "dem.college_or_university_degree"
 [20] "dem.a_levelsas_levels_or_equivalent"
```

```
[21] "dem.o_levelsgcses_or_equivalent"
[22] "dem.cses_or_equivalent"
[23] "dem.nvq_or_hnd_or_hnc_or_equivalent"
[24] "dem.other_professional_qualifications_"
[25] "dem.other_professional_qualifications_text.txt"
[26] "dem.none_of_the_above"
[27] "dem.prefer_not_to_say"
[28] "dem.british_mixed_british"
[29] "dem.irish"
[30] "dem.northern_irish"
[31] "dem.any_other_white_background"
[32] "dem.white_and_black_caribbean"
[33] "dem.white_and_black_africa"
[34] "dem.white_and_asian"
[35] "dem.any_other_mixed_background"
[36] "dem.indian_or_british_indian"
[37] "dem.pakistani_or_british_pakistani"
[38] "dem.bangladeshi_or_british_bangladeshi"
[39] "dem.any_other_asian_background"
[40] "dem.caribbean"
[41] "dem.african"
[42] "dem.any_other_black_background"
[43] "dem.chinese"
[44] "dem.any_other_ethnic_group"
[45] "dem.other"
[46] "dem.othertext.txt"
[47] "dem.english"
[48] "dem.scottish"
[49] "dem.welsh"
[50] "dem.cornish"
[51] "dem.cypriot_"
[52] "dem.greek"
[53] "dem.greek_cypriot"
[54] "dem.italian"
[55] "dem.irish_traveller"
[56] "dem.traveller"
[57] "dem.gypsyromany"
[58] "dem.polish"
[59] "dem.republics_made_ussr"
[60] "dem.kosovan"
[61] "dem.albanian"
[62] "dem.bosnian"
[63] "dem.croatian"
[64] "dem.serbian"
[65] "dem.republics_made_yugoslavia"
[66] "dem.mixed_white"
[67] "dem.other_white_european_european_unspecified_european_mix"
[68] "dem.black_and_asian"
[69] "dem.black_and_chinese"
[70] "dem.black_and_white"
[71] "dem.chinese_and_white"
[72] "dem.asian_and_chinese"
[73] "dem.other_mixed_mixed_unspecified"
[74] "dem.other_mixed_mixed_unspecifiedtext.txt"
```

```
 [75] "dem.mixed_asian"
 [76] "dem.punjabi"
 [77] "dem.kashmiri"
 [78] "dem.east_african_asian"
 [79] "dem.tamil"
 [80] "dem.sinhalese"
 [81] "dem.british_asian"
 [82] "dem.caribbean_asian"
 [83] "dem.other_asian_asian_unspecified"
 [84] "dem.other_asian_asian_unspecifiedtext.txt"
 [85] "dem.somali"
 [86] "dem.mixed_black"
 [87] "dem.nigerian"
 [88] "dem.black_british"
 [89] "dem.other_black_black_unspecified"
 [90] "dem.other_black_black_unspecifiedtext.txt"
 [91] "dem.is_english_your_first_language"
 [92] "dem.what_is_your_first_language"
 [93] "dem.what_is_your_first_language.txt"
 [94] "dem.please_select_your_preferred_units_of_measurement"
 [95] "dem.what_is_your_current_height"
 [96] "dem.what_is_your_current_height.1"
 [97] "dem.what_is_your_current_height.2"
 [98] "dem.pregnant_weigh_weight_provide"
 [99] "dem.pregnant_weigh_weight_provide.1"
[100] "dem.pregnant_weigh_weight_provide.2"
[101] "dem.pregnant_weighed_weight_provide"
[102] "dem.pregnant_weighed_weight_provide.1"
[103] "dem.pregnant_weighed_weight_provide.2"
[104] "dem.highest_weight"
[105] "dem.stopped_growing_adult_height"
[106] "dem.stopped_growing_adult_height.1"
[107] "dem.stopped_growing_adult_height.2"
[108] "dem.body_suffered_injury_involving"
[109] "dem.middle_wake_night_covid19"
[110] "dem.middle_wake_night_covid19.1"
[111] "dem.medical_history_birth_relevant"
[112] "dem.affects_concerned_live_memory"
[113] "dem.memory_problem_worse_year"
[114] "dem.based_confirm_living_question"
[115] "dem.diagnosed_required_question_covid19"
[116] "dem.long_ago_diagnosed_required"
[117] "dem.long_ago_diagnosed_required.1"
[118] "dem.diagnosed_covid19_experienced_similar"
[119] "dem.quality_rate_life"
[120] "dem.energy_everyday_life"
[121] "dem.opportunity_leisure_activities"
[122] "dem.money_day"
[123] "dem.middle_wake_night_trouble"
[124] "dem.affects_concerned_live_memory.1"
[125] "dem.affects_concerned_live_memory.2"
[126] "dem.has_your_memory_got_progressively_worse"
[127] "dem.vietnamese"
[128] "dem.filipino"
```

```
[129] "dem.malaysian"
[130] "dem.any_other_group"
[131] "dem.any_other_grouptext.txt"
[132] "dem.lowest_weight_adult_height"
[133] "dem.happy_general_health"
```

Specify columns to be excluded from add_numeric function Continuous variables should be excluded, as they are already numeric

```
exclude_cols_numeric <- c(
  "ID",
  "sample",
  "startDate",
  "endDate",
  "dem.othertext.txt"
  )
```

Select & rename relevant columns

```
ethn_id <- ethn %>% #new dataset with ID
  drop_na(externalDataReference) %>% # Drop NAs
  distinct(externalDataReference, .keep_all = TRUE) %>% # Changed to distinct due to NA coercion
  add_column(sample = "COVIDCNS",
             .after = "externalDataReference") %>% # Create new sample column
  select(
         ID = externalDataReference, # ID
         sample,
         startDate,
         endDate,
         dem.british_mixed_british,
         dem.irish,
         dem.northern_irish,
         dem.any_other_white_background,
         dem.white_and_black_caribbean,
         dem.white_and_black_africa,
         dem.white_and_asian,
         dem.any_other_mixed_background,
         dem.indian_or_british_indian,
         dem.pakistani_or_british_pakistani,
         dem.bangladeshi_or_british_bangladeshi,
         dem.any_other_asian_background,
         dem.caribbean,
         dem.african,
         dem.any_other_black_background,
         dem.chinese,
         dem.any_other_ethnic_group,
         dem.other,
         dem.othertext.txt
        ) %>%
   add_numeric_1(exclude = exclude_cols_numeric)

# Inspect colnames
```

```
ethn_id %>%
  colnames()
```

```
 [1] "ID"
 [2] "sample"
 [3] "startDate"
 [4] "endDate"
 [5] "dem.british_mixed_british"
 [6] "dem.irish"
 [7] "dem.northern_irish"
 [8] "dem.any_other_white_background"
 [9] "dem.white_and_black_caribbean"
[10] "dem.white_and_black_africa"
[11] "dem.white_and_asian"
[12] "dem.any_other_mixed_background"
[13] "dem.indian_or_british_indian"
[14] "dem.pakistani_or_british_pakistani"
[15] "dem.bangladeshi_or_british_bangladeshi"
[16] "dem.any_other_asian_background"
[17] "dem.caribbean"
[18] "dem.african"
[19] "dem.any_other_black_background"
[20] "dem.chinese"
[21] "dem.any_other_ethnic_group"
[22] "dem.other"
[23] "dem.othertext.txt"
[24] "dem.british_mixed_british_numeric"
[25] "dem.irish_numeric"
[26] "dem.northern_irish_numeric"
[27] "dem.any_other_white_background_numeric"
[28] "dem.white_and_black_caribbean_numeric"
[29] "dem.white_and_black_africa_numeric"
[30] "dem.white_and_asian_numeric"
[31] "dem.any_other_mixed_background_numeric"
[32] "dem.indian_or_british_indian_numeric"
[33] "dem.pakistani_or_british_pakistani_numeric"
[34] "dem.bangladeshi_or_british_bangladeshi_numeric"
[35] "dem.any_other_asian_background_numeric"
[36] "dem.caribbean_numeric"
[37] "dem.african_numeric"
[38] "dem.any_other_black_background_numeric"
[39] "dem.chinese_numeric"
[40] "dem.any_other_ethnic_group_numeric"
[41] "dem.other_numeric"
```

```
# Inspect dimensions
dim(ethn_id)
```

```
[1] 228  41
```

```
# Differences
ethn_excluded <- dim(ethn_id)[1]-dim(ethn)[1]
ethn_excluded
```

```
[1] -7
```

`ethn_excluded` COVID CNS participants excluded due to missing data

Inspect numeric variables

```
ethn_id %>%
  select(all_of(ends_with("numeric"))) %>%
  tbl_summary(missing_text = "Missing")
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	144 (63%)
Missing	1
What is your ethnic origin?	4 (1.8%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	13 (5.7%)
Missing	1
What is your ethnic origin?	5 (2.2%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	2 (0.9%)
Missing	1
What is your ethnic origin?	2 (0.9%)
Missing	1
What is your ethnic origin?	3 (1.3%)
Missing	1
What is your ethnic origin?	3 (1.3%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	3 (1.3%)
Missing	1
What is your ethnic origin?	14 (6.2%)
Missing	1
What is your ethnic origin?	11 (4.8%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	0 (0%)
Missing	1
What is your ethnic origin?	25 (11%)

Characteristic	N = 228
Missing	1

```
ethn_id %>%
  select(ends_with("numeric")) %>%
  freq()
```

```
Frequencies
ethn_id$dem.british_mixed_british_numeric
Label: What is your ethnic origin?
Type: Numeric

              Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
----------- ------ --------- -------------- --------- --------------
          0     83     36.56          36.56     36.40          36.40
          1    144     63.44         100.00     63.16          99.56
       <NA>      1                               0.44         100.00
      Total    228    100.00         100.00    100.00         100.00

ethn_id$dem.irish_numeric
Label: What is your ethnic origin?
Type: Numeric

              Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
----------- ------ --------- -------------- --------- --------------
          0    223     98.24          98.24     97.81          97.81
          1      4      1.76         100.00      1.75          99.56
       <NA>      1                               0.44         100.00
      Total    228    100.00         100.00    100.00         100.00

ethn_id$dem.northern_irish_numeric
Label: What is your ethnic origin?
Type: Numeric

              Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
----------- ------ --------- -------------- --------- --------------
          0    226     99.56          99.56     99.12          99.12
          1      1      0.44         100.00      0.44          99.56
       <NA>      1                               0.44         100.00
      Total    228    100.00         100.00    100.00         100.00

ethn_id$dem.any_other_white_background_numeric
Label: What is your ethnic origin?
Type: Numeric

              Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
----------- ------ --------- -------------- --------- --------------
          0    214     94.27          94.27     93.86          93.86
          1     13      5.73         100.00      5.70          99.56
       <NA>      1                               0.44         100.00
      Total    228    100.00         100.00    100.00         100.00
```

ethn_id$dem.white_and_black_caribbean_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	222	97.80	97.80	97.37	97.37
1	5	2.20	100.00	2.19	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.white_and_black_africa_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	226	99.56	99.56	99.12	99.12
1	1	0.44	100.00	0.44	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.white_and_asian_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	225	99.12	99.12	98.68	98.68
1	2	0.88	100.00	0.88	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.any_other_mixed_background_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	225	99.12	99.12	98.68	98.68
1	2	0.88	100.00	0.88	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.indian_or_british_indian_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	224	98.68	98.68	98.25	98.25
1	3	1.32	100.00	1.32	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.pakistani_or_british_pakistani_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	224	98.68	98.68	98.25	98.25
1	3	1.32	100.00	1.32	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.bangladeshi_or_british_bangladeshi_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	226	99.56	99.56	99.12	99.12
1	1	0.44	100.00	0.44	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.any_other_asian_background_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	224	98.68	98.68	98.25	98.25
1	3	1.32	100.00	1.32	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.caribbean_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	213	93.83	93.83	93.42	93.42
1	14	6.17	100.00	6.14	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.african_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	216	95.15	95.15	94.74	94.74
1	11	4.85	100.00	4.82	99.56
<NA>	1			0.44	100.00

Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.any_other_black_background_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	226	99.56	99.56	99.12	99.12
1	1	0.44	100.00	0.44	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.chinese_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	226	99.56	99.56	99.12	99.12
1	1	0.44	100.00	0.44	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.any_other_ethnic_group_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	227	100.00	100.00	99.56	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

ethn_id$dem.other_numeric
Label: What is your ethnic origin?
Type: Numeric

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	202	88.99	88.99	88.60	88.60
1	25	11.01	100.00	10.96	99.56
<NA>	1			0.44	100.00
Total	228	100.00	100.00	100.00	100.00

Data cleaning

Numeric variables

Select numeric ethnicity variables for cleaning

```
ethn_vars_numeric <- c(
   "dem.british_mixed_british_numeric",
   "dem.irish_numeric",
   "dem.northern_irish_numeric",
   "dem.any_other_white_background_numeric",
   "dem.white_and_black_caribbean_numeric",
   "dem.white_and_black_africa_numeric",
   "dem.white_and_asian_numeric",
   "dem.any_other_mixed_background_numeric",
   "dem.indian_or_british_indian_numeric",
   "dem.pakistani_or_british_pakistani_numeric",
   "dem.bangladeshi_or_british_bangladeshi_numeric",
   "dem.any_other_asian_background_numeric",
   "dem.caribbean_numeric",
   "dem.african_numeric",
   "dem.any_other_black_background_numeric",
   "dem.chinese_numeric",
   "dem.any_other_ethnic_group_numeric",
   "dem.other_numeric"
)
```

Vector of plausible values for the numeric ethnicity variables

```
ethn_values_numeric <- c(
  0,
  1,
  -777,
  NA
)
```

```
imp_check(data = ethn_id,
          variables = ethn_vars_numeric,
          values = ethn_values_numeric)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	144 (63%)
Missing	1
What is your ethnic origin?	4 (1.8%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	13 (5.7%)
Missing	1
What is your ethnic origin?	5 (2.2%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	2 (0.9%)
Missing	1
What is your ethnic origin?	2 (0.9%)
Missing	1
What is your ethnic origin?	3 (1.3%)
Missing	1
What is your ethnic origin?	3 (1.3%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	3 (1.3%)
Missing	1
What is your ethnic origin?	14 (6.2%)
Missing	1
What is your ethnic origin?	11 (4.8%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	1 (0.4%)
Missing	1
What is your ethnic origin?	0 (0%)
Missing	1
What is your ethnic origin?	25 (11%)
Missing	1

Non-numeric variables

British, Mixed British

Select "British, Mixed British" variable for cleaning

```
british_variable <- c(
  "dem.british_mixed_british"
)
```

Vector of plausible values for "British, Mixed British" variable

```
british_values <- c(
  "British, Mixed British",
  "Not British, Mixed British",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = british_variable,
          values = british_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not British, Mixed British	83 (37%)
British, Mixed British	144 (63%)
Missing	1

Irish

Select "Irish" variable for cleaning

```
irish_variable <- c(
  "dem.irish"
)
```

Vector of plausible values for "Irish" variable

```
irish_values <- c(
  "Irish",
  "Not Irish",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = irish_variable,
          values = irish_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not Irish	223 (98%)
Irish	4 (1.8%)
Missing	1

Northern Irish

Select "Northern Irish" variable for cleaning

```
n_irish_variable <- c(
  "dem.northern_irish"
)
```

Vector of plausible values for "Northern Irish" variable

```
n_irish_values <- c(
  "Northern Irish",
  "Not Northern Irish",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = n_irish_variable,
          values = n_irish_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not Northern Irish	226 (100%)
Northern Irish	1 (0.4%)
Missing	1

Any other White background

Select "Any other White background" variable for cleaning

```
aowb_variable <- c(
  "dem.any_other_white_background"
)
```

Vector of plausible values for "Any other White background" variable

```
aowb_values <- c(
  "Any other White background",
  "Not Any other White background",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = aowb_variable,
          values = aowb_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not Any other White background	214 (94%)
Any other White background	13 (5.7%)
Missing	1

White and Black Caribbean

Select "White and Black Caribbean" variable for cleaning

```
wbcarab_variable <- c(
  "dem.white_and_black_caribbean"
)
```

Vector of plausible values for "White and Black Caribbean" variable

```
wbcarab_values <- c(
  "White and Black Caribbean",
  "Not White and Black Caribbean",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = wbcarab_variable,
          values = wbcarab_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 228**
What is your ethnic origin?	
Not White and Black Caribbean	222 (98%)
White and Black Caribbean	5 (2.2%)
Missing	1

White and Black Africa

Select "White and Black Africa" variable for cleaning

```
wbafrica_variable <- c(
  "dem.white_and_black_africa"
)
```

Vector of plausible values for "White and Black Africa" variable

```
wbafrica_values <- c(
  "White and Black Africa",
  "Not White and Black Africa",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = wbafrica_variable,
          values = wbafrica_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not White and Black Africa	226 (100%)
White and Black Africa	1 (0.4%)
Missing	1

White and Asian

Select "White and Asian" variable for cleaning

```
w_asian_variable <- c(
  "dem.white_and_asian"
)
```

Vector of plausible values for "White and Asian" variable

```
w_asian_values <- c(
  "White and Asian",
  "Not White and Asian",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = w_asian_variable,
          values = w_asian_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not White and Asian	225 (99%)
White and Asian	2 (0.9%)
Missing	1

Any other mixed background

Select "Any other mixed background" variable for cleaning

```
aomb_variable <- c(
  "dem.any_other_mixed_background"
)
```

Vector of plausible values for "Any other mixed background" variable

```
aomb_values <- c(
  "Any other mixed background",
  "Not Any other mixed background",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = aomb_variable,
          values = aomb_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not Any other mixed background	225 (99%)
Any other mixed background	2 (0.9%)
Missing	1

Indian or British Indian

Select "Indian or British Indian" variable for cleaning

```
ind_br_ind_variable <- c(
  "dem.indian_or_british_indian"
)
```

Vector of plausible values for "Indian or British Indian" variable

```
ind_br_ind_values <- c(
  "Indian or British Indian",
  "Not Indian or British Indian",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = ind_br_ind_variable,
          values = ind_br_ind_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 228**
What is your ethnic origin?	
Not Indian or British Indian	224 (99%)
Indian or British Indian	3 (1.3%)
Missing	1

Pakistani or British Pakistani

Select "Pakistani or British Pakistani" variable for cleaning

```
p_br_p_variable <- c(
  "dem.pakistani_or_british_pakistani"
)
```

Vector of plausible values for "Pakistani or British Pakistani" variable

```
p_br_p_values <- c(
  "Pakistani or British Pakistani",
  "Not Pakistani or British Pakistani",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = p_br_p_variable,
          values = p_br_p_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not Pakistani or British Pakistani	224 (99%)
Pakistani or British Pakistani	3 (1.3%)
Missing	1

Bangladeshi or British Bangladeshi

Select "Bangladeshi or British Bangladeshi" variable for cleaning

```
b_br_b_variable <- c(
  "dem.bangladeshi_or_british_bangladeshi"
)
```

Vector of plausible values for "Bangladeshi or British Bangladeshi" variable

```
b_br_b_values <- c(
  "Bangladeshi or British Bangladeshi",
  "Not Bangladeshi or British Bangladeshi",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = b_br_b_variable,
          values = b_br_b_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not Bangladeshi or British Bangladeshi	226 (100%)
Bangladeshi or British Bangladeshi	1 (0.4%)
Missing	1

Any other Asian background

Select "Any other Asian background" variable for cleaning

```
aoab_variable <- c(
  "dem.any_other_asian_background"
)
```

Vector of plausible values for "Any other Asian background" variable

```
aoab_values <- c(
  "Any other Asian Background",
  "Not Any other Asian Background",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = aoab_variable,
          values = aoab_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not Any other Asian Background	224 (99%)
Any other Asian Background	3 (1.3%)
Missing	1

Caribbean

Select "Caribbean" variable for cleaning

```
carab_variable <- c(
  "dem.caribbean"
)
```

Vector of plausible values for "Caribbean" variable

```
carab_values <- c(
  "Caribbean",
  "Not Caribbean",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = carab_variable,
          values = carab_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 228**
What is your ethnic origin?	
Not Caribbean	213 (94%)
Caribbean	14 (6.2%)
Missing	1

African

Select "African" variable for cleaning

```
african_variable <- c(
  "dem.african"
)
```

Vector of plausible values for "African" variable

```
african_values <- c(
  "African",
  "Not African",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = african_variable,
          values = african_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not African	216 (95%)
African	11 (4.8%)
Missing	1

Any other Black Background

Select "Any other Black Background" variable for cleaning

```
aobb_variable <- c(
  "dem.any_other_black_background"
)
```

Vector of plausible values for "Any other Black Background" variable

```
aobb_values <- c(
  "Any other Black Background",
  "Not Any other Black Background",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = aobb_variable,
          values = aobb_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N = 228
What is your ethnic origin?	
Not Any other Black Background	226 (100%)
Any other Black Background	1 (0.4%)
Missing	1

Chinese

Select "Chinese" variable for cleaning

```
chinese_variable <- c(
  "dem.chinese"
)
```

Vector of plausible values for "Chinese" variable

```
chinese_values <- c(
  "Chinese",
  "Not Chinese",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = chinese_variable,
          values = chinese_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 228**
What is your ethnic origin?	
Not Chinese	226 (100%)
Chinese	1 (0.4%)
Missing	1

Any other ethnic group

Select "Any other ethnic group" variable for cleaning

```
aoeg_variable <- c(
  "dem.any_other_ethnic_group"
)
```

Vector of plausible values for "Any other ethnic group" variable

```
aoeg_values <- c(
  "Any other ethnic group",
  "Not Any other ethnic group",
  "Seen but not answered",
  NA
)
```

```
imp_check(data = ethn_id,
          variables = aoeg_variable,
          values = aoeg_values)
```

```
[1] "There are no implausible values in the dataset. Can leave these variables as they are."
```

```
Table printed with `knitr::kable()`, not {gt}. Learn why at
http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 228**
What is your ethnic origin?	
Not Any other ethnic group	227 (100%)
Any other ethnic group	0 (0%)
Missing	1

Save clean data

Export variables

```
export_variables <- c(
  "ID",
  "sample"  ,
  "startDate",
  "endDate",
  "dem.british_mixed_british" ,
  "dem.irish",
  "dem.northern_irish",
  "dem.any_other_white_background",
  "dem.white_and_black_caribbean",
  "dem.white_and_black_africa",
  "dem.white_and_asian",
  "dem.any_other_mixed_background",
  "dem.indian_or_british_indian",
  "dem.pakistani_or_british_pakistani",
  "dem.bangladeshi_or_british_bangladeshi",
  "dem.any_other_asian_background",
  "dem.caribbean",
  "dem.african",
  "dem.any_other_black_background",
```

```
  "dem.chinese",
  "dem.any_other_ethnic_group",
  "dem.other",
  "dem.othertext.txt",
  "dem.british_mixed_british_numeric",
  "dem.irish_numeric",
  "dem.northern_irish_numeric",
  "dem.any_other_white_background_numeric",
  "dem.white_and_black_caribbean_numeric",
  "dem.white_and_black_africa_numeric",
  "dem.white_and_asian_numeric",
  "dem.any_other_mixed_background_numeric",
  "dem.indian_or_british_indian_numeric",
  "dem.pakistani_or_british_pakistani_numeric",
  "dem.bangladeshi_or_british_bangladeshi_numeric",
  "dem.any_other_asian_background_numeric",
  "dem.caribbean_numeric",
  "dem.african_numeric",
  "dem.any_other_black_background_numeric",
  "dem.chinese_numeric",
  "dem.any_other_ethnic_group_numeric",
  "dem.other_numeric"
)
```

```
ethn_id %>%
  select(all_of(export_variables)) %>%
  saveRDS(file =
            paste0(ilovedata, "/data/latest_freeze/covidcns/ethnicity_covidcns_clean.rds"))
```