Highest qualifications

Chelsea Mika Malouf

31/01/2022

This script is to clean highest qualification. This consists of the following question: Which of the following qualifications do you have? A variable was created in the end of the script representing the highest qualification/level of education.

#Set up

```
remove(list = ls())
```

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

```
packages <- c("summarytools",
              "sjlabelled",
              "Amelia",
              "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: 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(file = "../credentials/paths.R")
```

Read in covid cns data

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

# Check variable names in dataframe
cns_dat %>%
  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"
```

```
# Inspect dimensions of dataframe
cns_dat %>%
  dim()
```

```
[1] 235 133
```

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

Select & rename relevant columns

```r
cns_dat_id <- cns_dat %>% #new dataset with ID
  drop_na(externalDataReference) %>% # Drop participants with no ID
  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.college_or_university_degree,
         dem.a_levelsas_levels_or_equivalent,
         dem.o_levelsgcses_or_equivalent,
         dem.cses_or_equivalent,
         dem.nvq_or_hnd_or_hnc_or_equivalent,
         dem.other_professional_qualifications_,
         dem.other_professional_qualifications_text.txt,
         dem.none_of_the_above,
         dem.prefer_not_to_say
         ) %>%
  add_numeric_1(exclude = exclude_cols_numeric)

# Inspect colnames
cns_dat_id %>%
  colnames()
```

```
 [1] "ID"
 [2] "sample"
 [3] "startDate"
 [4] "endDate"
 [5] "dem.college_or_university_degree"
 [6] "dem.a_levelsas_levels_or_equivalent"
 [7] "dem.o_levelsgcses_or_equivalent"
 [8] "dem.cses_or_equivalent"
 [9] "dem.nvq_or_hnd_or_hnc_or_equivalent"
[10] "dem.other_professional_qualifications_"
[11] "dem.none_of_the_above"
[12] "dem.prefer_not_to_say"
[13] "dem.other_professional_qualifications_text.txt"
[14] "dem.college_or_university_degree_numeric"
[15] "dem.a_levelsas_levels_or_equivalent_numeric"
[16] "dem.o_levelsgcses_or_equivalent_numeric"
[17] "dem.cses_or_equivalent_numeric"
```

```
[18] "dem.nvq_or_hnd_or_hnc_or_equivalent_numeric"
[19] "dem.other_professional_qualifications__numeric"
[20] "dem.none_of_the_above_numeric"
[21] "dem.prefer_not_to_say_numeric"
```

```
# Inspect dimensions of new data set
cns_dat_id %>%
  dim()
```

```
[1] 228  21
```

```
# Inspect number of rows dropped
cns_excluded <- dim(cns_dat_id)[1] - dim(cns_dat)[1]
cns_excluded
```

```
[1] -7
```

```
cns_dat_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.
```

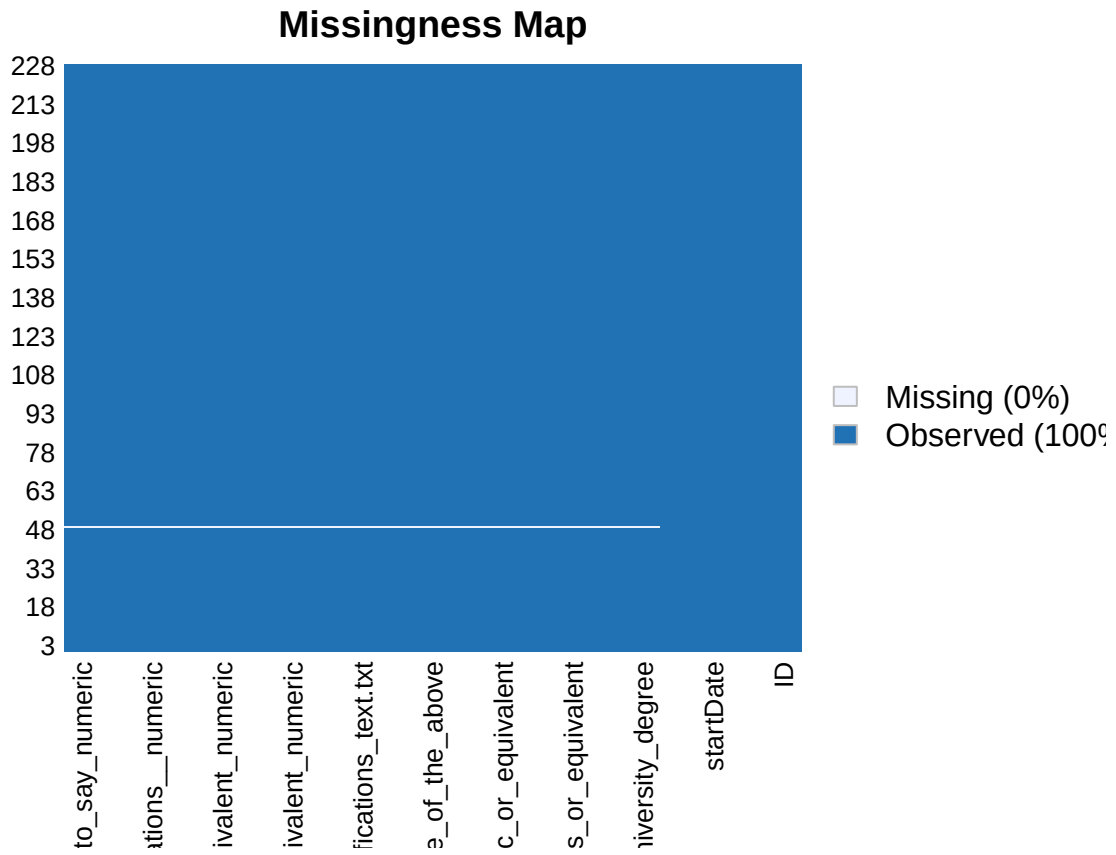
Characteristic	**N = 228**
Which of the following qualifications doyou have?	81 (36%)
Missing	1
Which of the following qualifications doyou have?	49 (22%)
Missing	1
Which of the following qualifications doyou have?	92 (41%)
Missing	1
Which of the following qualifications doyou have?	31 (14%)
Missing	1
Which of the following qualifications doyou have?	40 (18%)
Missing	1
Which of the following qualifications doyou have?	43 (19%)
Missing	1
Which of the following qualifications doyou have?	30 (13%)
Missing	1
Which of the following qualifications doyou have?	2 (0.9%)
Missing	1

```
miss_map <- cns_dat_id %>%
  missmap()
```

```
Warning: Unknown or uninitialised column: `arguments`.

Warning: Unknown or uninitialised column: `arguments`.
```

```
Warning: Unknown or uninitialised column: ‘imputations‘.
```




```
miss_map
```

```
NULL
```

```
 dat <- cns_dat_id %>%
   mutate(across(ends_with("numeric"),
                 ~case_when(
                   . == -55 ~ -555,
                   . == -77 ~ -777,
                   . == -88 ~ -888,
                   . == -99 ~ -999,
                   TRUE ~ .)))
```

Cleaning Categorical variables

```
variables_categorical_university <-
  c(
    "dem.college_or_university_degree"
    )
variables_categorical_university
```

```
[1] "dem.college_or_university_degree"
```

```
values_categorical_university <-
  c(
    "Not College or university degree",
    "College or university degree",
    NA
  )
values_categorical_university
```

```
[1] "Not College or university degree" "College or university degree"
[3] NA
```

```
imp_check(data = dat,
          variables = variables_categorical_university,
          values = values_categorical_university)
```

```
[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**
Which of the following qualifications doyou have?	
Not College or university degree	146 (64%)
College or university degree	81 (36%)
Missing	1

```
#There are no implausible values in the dataset. Can leave these variables as they are
```

```
variables_categorical_a_levels <-
  c(
    "dem.a_levelsas_levels_or_equivalent"
    )
variables_categorical_a_levels
```

```
[1] "dem.a_levelsas_levels_or_equivalent"
```

```
values_categorical_a_levels <-
  c(
    "Not A levels/AS levels or equivalent",
    "A levels/AS levels or equivalent",
    NA
  )
values_categorical_a_levels
```

```
[1] "Not A levels/AS levels or equivalent"
[2] "A levels/AS levels or equivalent"
[3] NA
```

```
imp_check(data = dat,
          variables = variables_categorical_a_levels,
          values = values_categorical_a_levels)
```

```
[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**
Which of the following qualifications doyou have?	
Not A levels/AS levels or equivalent	178 (78%)
A levels/AS levels or equivalent	49 (22%)
Missing	1

```
#There are no implausible values in the dataset. Can leave these variables as they are
```

```
variables_categorical_o_levels <-
  c(
    "dem.o_levelsgcses_or_equivalent"
    )
variables_categorical_o_levels
```

```
[1] "dem.o_levelsgcses_or_equivalent"
```

```
values_categorical_o_levels <-
  c(
    "Not O levels/GCSEs or equivalent",
    "O levels/GCSEs or equivalent",
    NA
  )
values_categorical_o_levels
```

```
[1] "Not O levels/GCSEs or equivalent" "O levels/GCSEs or equivalent"
[3] NA
```

```
imp_check(data = dat,
          variables = variables_categorical_o_levels,
          values = values_categorical_o_levels)
```

```
[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
Which of the following qualifications doyou have?	
Not O levels/GCSEs or equivalent	135 (59%)
O levels/GCSEs or equivalent	92 (41%)
Missing	1

```
#There are no implausible values in the dataset. Can leave these variables as they are
```

```
variables_categorical_cses <-
  c(
    "dem.cses_or_equivalent"
    )
variables_categorical_cses
```

```
[1] "dem.cses_or_equivalent"
```

```
values_categorical_cses <-
  c(
    "Not CSEs or equivalent",
    "CSEs or equivalent",
    NA
  )
values_categorical_cses
```

```
[1] "Not CSEs or equivalent" "CSEs or equivalent"     NA
```

```
imp_check(data = dat,
          variables = variables_categorical_cses,
          values = values_categorical_cses)
```

```
[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
Which of the following qualifications doyou have?	
Not CSEs or equivalent	196 (86%)
CSEs or equivalent	31 (14%)
Missing	1

```
#There are no implausible values in the dataset. Can leave these variables as they are
```

```
variables_categorical_nvq <-
  c(
    "dem.nvq_or_hnd_or_hnc_or_equivalent"
    )
variables_categorical_nvq
```

```
[1] "dem.nvq_or_hnd_or_hnc_or_equivalent"
```

```
values_categorical_nvq <-
  c(
    "Not NVQ or HND or HNC or equivalent",
    "NVQ or HND or HNC or equivalent",
    NA
  )
values_categorical_nvq
```

```
[1] "Not NVQ or HND or HNC or equivalent" "NVQ or HND or HNC or equivalent"
[3] NA
```

```
imp_check(data = dat,
          variables = variables_categorical_nvq,
          values = values_categorical_nvq)
```

```
[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**
Which of the following qualifications doyou have?	
Not NVQ or HND or HNC or equivalent	187 (82%)
NVQ or HND or HNC or equivalent	40 (18%)
Missing	1

```
#There are no implausible values in the dataset. Can leave these variables as they are
```

```
variables_categorical_other <-
  c(
    "dem.other_professional_qualifications_"
    )
variables_categorical_other
```

```
[1] "dem.other_professional_qualifications_"
```

```
values_categorical_other <-
  c(
    "Not Other professional qualifications",
    "Other professional qualifications",
    NA
  )
values_categorical_other
```

```
[1] "Not Other professional qualifications"
[2] "Other professional qualifications"
[3] NA
```

```
imp_check(data = dat,
          variables = variables_categorical_other,
          values = values_categorical_other)
```

```
[1] "The number of implausible values in the dataset is 227. Please investigate."
```

```
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**
Which of the following qualifications doyou have?	
Not Other professional qualifications (e.g. nursing, teaching)	184 (81%)
Other professional qualifications (e.g. nursing, teaching)	43 (19%)
Missing	1

```
#The number of implausible values in the dataset is 227. Please investigate.
```

Recode variable dem.other_professional_qualifications_

```
dat <- dat %>%
  mutate(dem.other_professional_qualifications_ =
           fct_recode(dem.other_professional_qualifications_,
    "Not Other professional qualifications" = "Not Other professional qualifications (e.g. nursing, tea
    "Other professional qualifications" = "Other professional qualifications (e.g. nursing, teaching)")
  )
```

```
imp_check(data = dat,
          variables = variables_categorical_other,
          values = values_categorical_other)
```

```
[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**
Which of the following qualifications doyou have?	
Not Other professional qualifications	184 (81%)
Other professional qualifications	43 (19%)
Missing	1

```
#There are no implausible values in the dataset. Can leave these variables as they are
```

```
variables_categorical_none <-
  c(
    "dem.none_of_the_above"
    )
variables_categorical_none
```

```
[1] "dem.none_of_the_above"
```

```
values_categorical_none <-
  c(
    "Not None of the above",
    "None of the above",
    NA
  )
values_categorical_none
```

```
[1] "Not None of the above" "None of the above"     NA
```

```
imp_check(data = dat,
          variables = variables_categorical_none,
          values = values_categorical_none)
```

```
[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**
Which of the following qualifications doyou have?	
Not None of the above	197 (87%)
None of the above	30 (13%)
Missing	1

```
#There are no implausible values in the dataset. Can leave these variables as they are
```

```
variables_categorical_prefer_not_to_say <-
  c(
    "dem.prefer_not_to_say"
    )
variables_categorical_prefer_not_to_say
```

```
[1] "dem.prefer_not_to_say"
```

```
values_categorical_prefer_not_to_say <-
  c(
    "Not Prefer not to say",
    "Prefer not to say",
    NA
  )
values_categorical_prefer_not_to_say
```

```
[1] "Not Prefer not to say" "Prefer not to say"     NA
```

```
imp_check(data = dat,
          variables = variables_categorical_prefer_not_to_say,
          values = values_categorical_prefer_not_to_say)
```

```
[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**
Which of the following qualifications doyou have?	
Not Prefer not to say	225 (99%)
Prefer not to say	2 (0.9%)
Missing	1

```
#There are no implausible values in the dataset. Can leave these variables as they are
```

Numeric variables

```
variables_numeric_0_1 <-
  c(
    "dem.college_or_university_degree_numeric",
    "dem.a_levelsas_levels_or_equivalent_numeric",
    "dem.o_levelsgcses_or_equivalent_numeric",
    "dem.cses_or_equivalent_numeric",
    "dem.nvq_or_hnd_or_hnc_or_equivalent_numeric",
    "dem.other_professional_qualifications__numeric",
    "dem.none_of_the_above_numeric",
    "dem.prefer_not_to_say_numeric"
    )
variables_numeric_0_1
```

```
[1] "dem.college_or_university_degree_numeric"
[2] "dem.a_levelsas_levels_or_equivalent_numeric"
[3] "dem.o_levelsgcses_or_equivalent_numeric"
```

```
[4] "dem.cses_or_equivalent_numeric"
[5] "dem.nvq_or_hnd_or_hnc_or_equivalent_numeric"
[6] "dem.other_professional_qualifications__numeric"
[7] "dem.none_of_the_above_numeric"
[8] "dem.prefer_not_to_say_numeric"
```

```
values_numeric_0_1 <-
  c(
    -777,
    0,
    1,
    NA
  )
values_numeric_0_1
```

```
[1] -777    0    1   NA
```

```
imp_check(data = dat,
          variables = variables_numeric_0_1,
          values = values_numeric_0_1)
```

```
[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**
dem.college_or_university_degree_numeric	81 (36%)
Missing	1
dem.a_levelsas_levels_or_equivalent_numeric	49 (22%)
Missing	1
dem.o_levelsgcses_or_equivalent_numeric	92 (41%)
Missing	1
dem.cses_or_equivalent_numeric	31 (14%)
Missing	1
dem.nvq_or_hnd_or_hnc_or_equivalent_numeric	40 (18%)
Missing	1
dem.other_professional_qualifications___numeric	43 (19%)
Missing	1
dem.none_of_the_above_numeric	30 (13%)
Missing	1
dem.prefer_not_to_say_numeric	2 (0.9%)
Missing	1

```
#There are no implausible values in the dataset. Can leave these variables as they are.
```

Making a categorical variable where highest_education is the variable and qualifications are categories

```
dat <- dat %>%
  mutate(
    dem.highest_education_numeric =
      case_when(
        dem.none_of_the_above_numeric == "1" ~ 0,
        dem.college_or_university_degree_numeric == "1" ~ 1,
        dem.a_levelsas_levels_or_equivalent_numeric == "1" ~ 2,
        dem.o_levelsgcses_or_equivalent_numeric == "1" ~ 3,
        dem.cses_or_equivalent_numeric == "1" ~ 4,
        dem.nvq_or_hnd_or_hnc_or_equivalent_numeric == "1" ~ 5,
        dem.other_professional_qualifications__numeric == "1" ~ 6,
        dem.prefer_not_to_say_numeric == "1" ~ -999
        )
    )

#recode the numeric version into a factor
dat <- dat %>%
  mutate(
    dem.highest_education =
      recode_factor(
        dem.highest_education_numeric,
        `0` = "None of the above" ,
        `1` = "College or university degree",
        `2` = "A levels/AS levels or equivalent",
        `3` = "O levels/GCSEs or equivalent",
        `4` = "CSEs or equivalent",
        `5` = "NVQ or HND or HNC or equivalent",
        `6` = "Other professional qualifications",
        `-999` = "Prefer not to say"
        )
    )

dat %>%
  freq(dem.highest_education)
```

```
Frequencies
dat$dem.highest_education
Type: Factor

                                      Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
-------------------------------------- ------ --------- -------------- --------- --------------
                        None of the above     30     13.22          13.22     13.16          13.16
             College or university degree     81     35.68          48.90     35.53          48.68
       A levels/AS levels or equivalent       22      9.69          58.59      9.65          58.33
           O levels/GCSEs or equivalent       51     22.47          81.06     22.37          80.70
                      CSEs or equivalent      17      7.49          88.55      7.46          88.16
         NVQ or HND or HNC or equivalent      13      5.73          94.27      5.70          93.86
       Other professional qualifications      11      4.85          99.12      4.82          98.68
                      Prefer not to say        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	

Save cleaned data

Export variables

```
export_variables <-
  c(
    "ID",
    "sample",
    "startDate",
    "endDate",
    "dem.college_or_university_degree",
    "dem.college_or_university_degree_numeric",
    "dem.a_levelsas_levels_or_equivalent",
    "dem.a_levelsas_levels_or_equivalent_numeric",
    "dem.o_levelsgcses_or_equivalent",
    "dem.o_levelsgcses_or_equivalent_numeric",
    "dem.cses_or_equivalent",
    "dem.cses_or_equivalent_numeric",
    "dem.nvq_or_hnd_or_hnc_or_equivalent",
    "dem.nvq_or_hnd_or_hnc_or_equivalent_numeric",
    "dem.other_professional_qualifications_",
    "dem.other_professional_qualifications__numeric",
    "dem.none_of_the_above",
    "dem.none_of_the_above_numeric",
    "dem.prefer_not_to_say",
    "dem.prefer_not_to_say",
    "dem.highest_education",
    "dem.highest_education_numeric"
  )
```

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