Codebook for baseline

Autogenerated data summary from dataReporter

2022-03-24 12:16:43

# Data report overview

The dataset examined has the following dimensions:

|  |  |
| --- | --- |
| Feature | Result |
| Number of observations | 5059 |
| Number of variables | 16 |

# Codebook summary table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Label | Variable | Class | # unique values | Missing | Description |
|  | [**prim\_key**](#prim_key) | AsIs | 5059 | 0.00 % |  |
|  | [**alc\_consumption**](#alc_consumption) | factor | 3 | 55.43 % |  |
|  | [**alc\_freq**](#alc_freq) | factor | 6 | 76.87 % |  |
|  | [**alc\_amount**](#alc_amount) | factor | 4 | 96.92 % |  |
|  | [**sex\_risk**](#sex_risk) | factor | 3 | 4.74 % |  |
|  | [**sex\_work**](#sex_work) | factor | 3 | 3.70 % |  |
|  | [**sex\_work\_condom**](#sex_work_condom) | factor | 3 | 99.58 % |  |
|  | [**sex\_hiv**](#sex_hiv) | factor | 3 | 5.04 % |  |
|  | [**sex\_partners**](#sex_partners) | integer | 48 | 11.17 % |  |
|  | [**sex\_partners\_24**](#sex_partners_24) | integer | 12 | 3.48 % |  |
|  | [**hiv\_positive**](#hiv_positive) | factor | 3 | 0.67 % |  |
|  | [**age**](#age) | integer | 69 | 0.83 % |  |
|  | [**marital\_status**](#marital_status) | factor | 3 | 0.08 % |  |
|  | [**c\_bd\_country**](#c_bd_country) | factor | 3 | 0.10 % |  |
|  | [**c\_bd\_educ4**](#c_bd_educ4) | ordered | 5 | 0.34 % |  |
|  | [**sex**](#sex) | factor | 2 | 0.00 % |  |

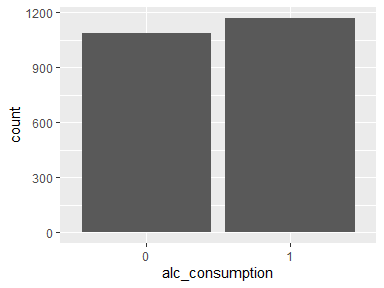
# Variable list

## prim\_key

* The variable is a key (distinct values for each observation).
* The variable has class AsIs which is not supported by dataReporter.

## alc\_consumption

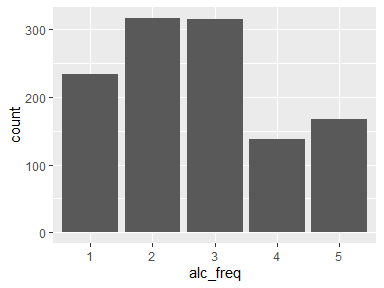
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 2804 (55.43 %) |
| Number of unique values | 2 |
| Mode | “1” |
| Reference category | 0 |



* Observed factor levels: "0", "1".

## alc\_freq

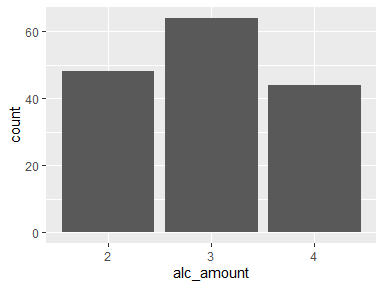
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 3889 (76.87 %) |
| Number of unique values | 5 |
| Mode | “2” |
| Reference category | 1 |



* Observed factor levels: "1", "2", "3", "4", "5".

## alc\_amount

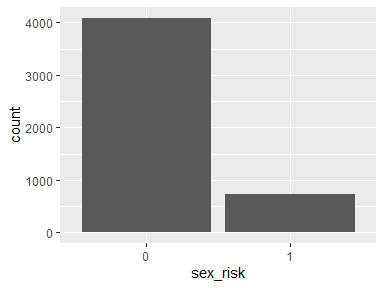
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 4903 (96.92 %) |
| Number of unique values | 3 |
| Mode | “3” |
| Reference category | 2 |



* Observed factor levels: "2", "3", "4".

## sex\_risk

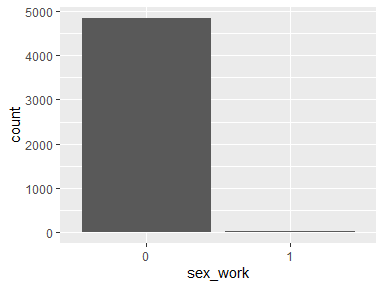
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 240 (4.74 %) |
| Number of unique values | 2 |
| Mode | “0” |
| Reference category | 0 |



* Observed factor levels: "0", "1".

## sex\_work

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 187 (3.7 %) |
| Number of unique values | 2 |
| Mode | “0” |
| Reference category | 0 |



* Observed factor levels: "0", "1".

## sex\_work\_condom

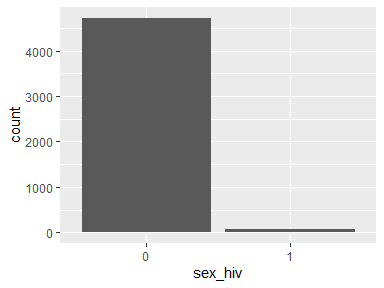
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 5038 (99.58 %) |
| Number of unique values | 2 |
| Mode | “0” |
| Reference category | 0 |



* Observed factor levels: "0", "1".

## sex\_hiv

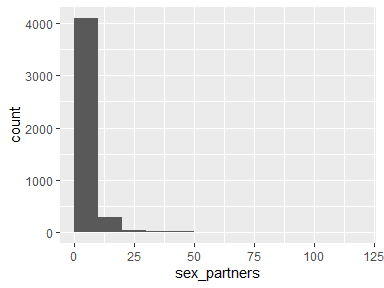
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 255 (5.04 %) |
| Number of unique values | 2 |
| Mode | “0” |
| Reference category | 0 |



* Observed factor levels: "0", "1".

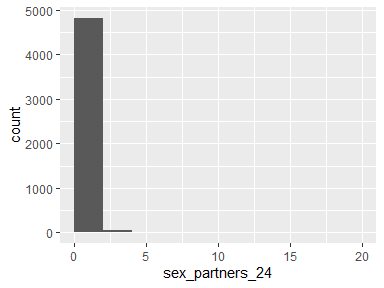
## sex\_partners

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | integer |
| Number of missing obs. | 565 (11.17 %) |
| Number of unique values | 47 |
| Median | 2 |
| 1st and 3rd quartiles | 1; 5 |
| Min. and max. | 0; 120 |



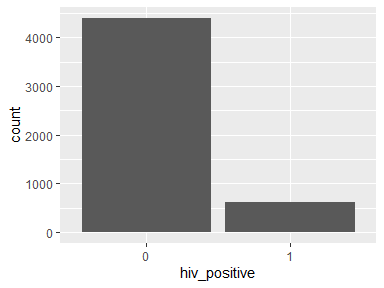
## sex\_partners\_24

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | integer |
| Number of missing obs. | 176 (3.48 %) |
| Number of unique values | 11 |
| Median | 1 |
| 1st and 3rd quartiles | 0; 1 |
| Min. and max. | 0; 20 |



## hiv\_positive

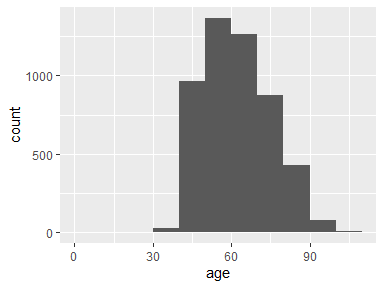
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 34 (0.67 %) |
| Number of unique values | 2 |
| Mode | “0” |
| Reference category | 0 |



* Observed factor levels: "0", "1".

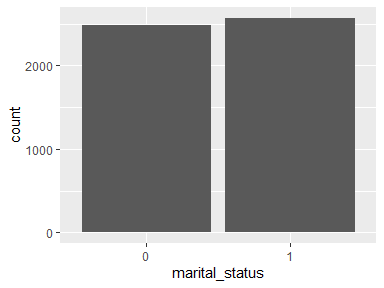
## age

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | integer |
| Number of missing obs. | 42 (0.83 %) |
| Number of unique values | 68 |
| Median | 62 |
| 1st and 3rd quartiles | 52; 72 |
| Min. and max. | 2; 105 |



## marital\_status

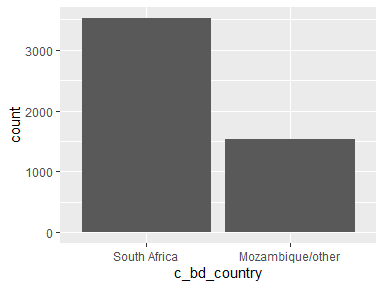
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 4 (0.08 %) |
| Number of unique values | 2 |
| Mode | “1” |
| Reference category | 0 |



* Observed factor levels: "0", "1".

## c\_bd\_country

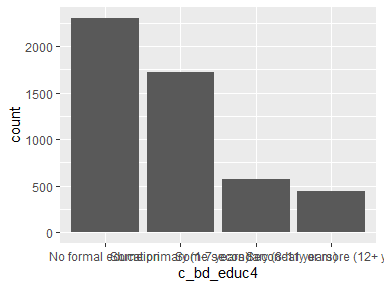
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 5 (0.1 %) |
| Number of unique values | 2 |
| Mode | “South Africa” |
| Reference category | South Africa |



* Observed factor levels: "Mozambique/other", "South Africa".

## c\_bd\_educ4

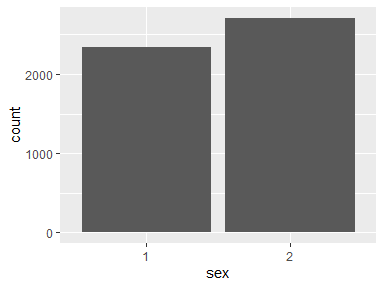
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | ordered |
| Number of missing obs. | 17 (0.34 %) |
| Number of unique values | 4 |
| Mode | “No formal education” |
| Reference category | No formal education |



* Observed factor levels: "No formal education", "Secondary or more (12+ years)", "Some primary (1-7 years)", "Some secondary (8-11 years)".

## sex

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 2 |
| Mode | “2” |
| Reference category | 1 |



* Observed factor levels: "1", "2".

Report generation information:

* Created by: Could not determine from system (username: Unknown)
* Report creation time: Thu Mar 24 2022 12:16:43
* Report was run from directory: C:/Users/XinYQ/Documents/Duke/glhlth562/final/proposal
* dataReporter v1.0.2 [Pkg: 2021-11-11 from CRAN (R 4.1.2)]
* R version 4.1.1 (2021-08-10).
* Platform: x86\_64-w64-mingw32/x64 (64-bit)(Windows 10 x64 (build 19044)).
* Function call: dataReporter::makeDataReport(data = baseline, mode = c("summarize", "visualize", "check"), smartNum = FALSE, file = "codebook\_baseline.Rmd", replace = TRUE, checks = list(character = "showAllFactorLevels", factor = "showAllFactorLevels", labelled = "showAllFactorLevels", haven\_labelled = "showAllFactorLevels", numeric = NULL, integer = NULL, logical = NULL, Date = NULL), listChecks = FALSE, maxProbVals = Inf, codebook = TRUE, reportTitle = "Codebook for baseline")