

Package ‘ihsMW’

June 4, 2026

Title Clean and Harmonise 'Malawi Integrated Household Survey' Data

Version 0.2.1

Description An offline suite of tools to clean, aggregate, and harmonise data from the 'Malawi Integrated Household Survey' ('IHS'). Provides crop-specific unit conversions, stratified winsorization, and automatic cross-round harmonisation for complex survey designs.

License MIT + file LICENSE

Depends R (>= 4.1.0)

URL <https://github.com/vituk123/ihsMW>

BugReports <https://github.com/vituk123/ihsMW/issues>

Imports dplyr (>= 1.1.0), readr (>= 2.1.0), rlang (>= 1.1.0), cli (>= 3.6.0)

Suggests srvyr (>= 1.2.0), survey (>= 4.2.0), testthat (>= 3.0.0), usethis (>= 2.2.0), pkgdown (>= 2.0.0), knitr (>= 1.40), rmarkdown (>= 2.20), withr (>= 2.5.0), jsonlite (>= 1.8.0)

Encoding UTF-8

RoxygenNote 7.3.3

VignetteBuilder knitr

Config/testthat/edition 3

Language en-US

NeedsCompilation no

Author Vitumbiko Kayuni [aut, cre] (ORCID:
<<https://orcid.org/0009-0008-4331-3839>>)

Maintainer Vitumbiko Kayuni <vitumbikokayuni@gmail.com>

Repository CRAN

Date/Publication 2026-06-04 20:30:02 UTC

Contents

ihs_aggregate	2
ihs_clean	2
ihs_convert_units	3
ihs_crosswalk_check	4
ihs_harmonise	4
ihs_search	5
ihs_standardize_missing	5
ihs_winsorize	6

Index	7
--------------	----------

ihs_aggregate	<i>Smart Aggregation to Household Level</i>
---------------	---

Description

Automatically detects variable types and applies sensible aggregations (e.g., ‘sum’ for continuous quantities, ‘max’ or logical OR for dummies). Throws warnings for ambiguous columns rather than failing silently.

Usage

```
ihs_aggregate(data, group_col = "case_id")
```

Arguments

data	A data.frame at the individual or plot level
group_col	The column name identifying the household (e.g., "case_id" or "y4_hhid")

Value

A data.frame aggregated to the household level

ihs_clean	<i>Clean and Harmonise IHS Data</i>
-----------	-------------------------------------

Description

This wrapper function applies standard cleaning procedures to Malawi IHS data. It handles missing value conversions, winsorization of continuous variables, and returns an audit log of all transformations applied.

Usage

```
ihs_clean(
  data,
  winsorize_vars = NULL,
  winsorize_by = NULL,
  probs = c(0.01, 0.99)
)
```

Arguments

<code>data</code>	A data.frame (typically loaded from a '.dta' file)
<code>winsorize_vars</code>	Character vector of continuous variables to winsorize (e.g., consumption, harvest)
<code>winsorize_by</code>	Optional character string of a grouping variable (e.g., region) for stratified winsorization
<code>probs</code>	Numeric vector of length 2 specifying the lower and upper quantiles for winsorization. Default is 'c(0.01, 0.99)'.

Value

A data.frame with cleaning applied. The returned object has an 'ihs_audit' attribute containing a log of modifications.

`ihs_convert_units` *Convert Agricultural Units to Kilograms*

Description

Converts reported harvest units (e.g., Pails, Ox carts, Heaps) into standard kilograms using official NSO crop-specific conversion factors.

Usage

```
ihs_convert_units(data, qty_col, unit_col, crop_col, unmapped = "warn")
```

Arguments

<code>data</code>	A data.frame
<code>qty_col</code>	The name of the column containing the quantity
<code>unit_col</code>	The name of the column containing the unit code or name
<code>crop_col</code>	The name of the column containing the crop code
<code>unmapped</code>	Action to take when a unit cannot be mapped: "warn" (default), "error", or "ignore".

Value

A data.frame with a new `qty_col_kg` column.

ihs_crosswalk_check *Check the comparability of variables across IHS rounds*

Description

Evaluates the completeness and comparability of variables across the available IHS rounds (IHS2, IHS3, IHS4, IHS5) using the bundled crosswalk.

Usage

```
ihs_crosswalk_check(verbose = TRUE)
```

Arguments

verbose Logical. If TRUE (default), prints a summary report to the console using cli.

Value

A tibble containing the full crosswalk. If verbose is TRUE, also prints a summary.

Examples

```
## Not run:
# Check the crosswalk and print a report
cw <- ihs_crosswalk_check()

## End(Not run)
```

ihs_harmonise *Harmonise Raw IHS Data*

Description

Takes a raw data.frame loaded from a Malawi IHS survey round (e.g. from a '.dta' file) and renames its columns to the standard harmonised variable names defined in the crosswalk.

Usage

```
ihs_harmonise(data, round = "IHS5", extra = FALSE)
```

Arguments

data A data.frame, typically read from a '.dta' file using haven::read_dta.
round A character string specifying the IHS round (e.g., "IHS5", "IHS4").
extra Logical. If FALSE (default), drops columns that are not in the harmonisation crosswalk or standard ID columns. If TRUE, keeps all original columns.

Value

A data.frame with columns renamed to standard 'harmonised_name's where applicable.

ihs_search	<i>Search across all IHS rounds for variables manually mapped</i>
------------	---

Description

Searches the manual harmonisation crosswalk bundled within ihsMW for specific variables.

Usage

```
ihs_search(keyword, round = NULL, fields = c("name", "label", "module"))
```

Arguments

keyword	A single search string to find (case-insensitive).
round	Limits search to a specific round. Valid inputs are "IHS2", "IHS3", "IHS4", "IHS5". Defaults to NULL (all rounds).
fields	A character vector of fields to include in the search. Valid fields are "name", "label", and "module".

Value

A tibble with cross-round harmonised search results.

Examples

```
ihs_search("consumption")
ihs_search("expenditure", round = "IHS5")
ihs_search("age", fields = c("name", "label"))
```

ihs_standardize_missing	<i>Standardize Survey Missing Codes</i>
-------------------------	---

Description

Converts common negative missing codes (like -99 for "Refused" or -98 for "Don't Know") into standard R 'NA' values to prevent them from skewing numeric calculations.

Usage

```
ihs_standardize_missing(data)
```

Arguments

data A data.frame

Value

A data.frame with missing values standardized

ihs_winsorize *Winsorize Continuous Variables*

Description

Caps extreme outliers at specified percentiles. Crucially, this function allows for stratified winsorization (e.g., by region) to avoid over-trimming poor/rich areas, and it creates new ‘_w’ suffixed columns to preserve raw data provenance.

Usage

```
ihs_winsorize(data, vars, by = NULL, probs = c(0.01, 0.99))
```

Arguments

data A data.frame
vars Character vector of column names to winsorize
by Optional grouping variable name (e.g., "region") for stratified thresholds
probs Numeric vector of lower and upper quantiles. Default ‘c(0.01, 0.99)’

Value

A data.frame with new ‘*_w’ columns added.

Index

ihs_aggregate, 2
ihs_clean, 2
ihs_convert_units, 3
ihs_crosswalk_check, 4
ihs_harmonise, 4
ihs_search, 5
ihs_standardize_missing, 5
ihs_winsorize, 6