

Package ‘BEMPdata’

March 31, 2026

Title Access the Bangladesh Environmental Mobility Panel Dataset

Version 0.2.3

Description Provides functions to download and work with the Bangladesh Environmental Mobility Panel (BEMP), a household panel survey tracing the impacts of riverbank erosion and flooding on (im)mobility, socio-economic outcomes, and political attitudes along the Jamuna River in Bangladesh (2021-2024). Wave datasets (20 files across 14 survey rounds) are hosted on Zenodo ([doi:10.5281/zenodo.18229497](https://doi.org/10.5281/zenodo.18229497)) and downloaded on demand with local caching. Bundled data include a merged cross-wave codebook and wave metadata.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

Imports utils, tools, readr, haven, shiny

Suggests bslib, DT, ggplot2, dplyr, testthat (>= 3.0.0), knitr, rmarkdown

Config/testthat/edition 3

VignetteBuilder knitr

URL <https://github.com/janfreihardt/BEMPdata>,
<https://CRAN.R-project.org/package=BEMPdata>

BugReports <https://github.com/janfreihardt/BEMPdata/issues>

NeedsCompilation no

Author Jan Freihardt [aut, cre]

Maintainer Jan Freihardt <fjan@ethz.ch>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2026-03-31 08:10:09 UTC

Contents

bemp_cache_clear	2
bemp_cache_info	3
codebook	3
get_codebook	4
get_wave	5
lookup_variable	6
run_app	7
wave_overview	7

Index	9
--------------	----------

bemp_cache_clear	<i>Clear the local BEMPdata cache</i>
------------------	---------------------------------------

Description

Deletes all locally cached BEMP files. The next call to `get_wave()` or `get_codebook()` will re-download from Zenodo.

Usage

```
bemp_cache_clear(confirm = FALSE)
```

Arguments

confirm	Logical. Set to TRUE to skip the interactive confirmation prompt (useful in scripts). Default FALSE.
---------	--

Value

Invisibly returns TRUE if the cache was cleared, FALSE if the user declined.

Examples

```
if (interactive()) {
  bemp_cache_clear(confirm = TRUE)
}
```

bemp_cache_info	<i>Show information about the local BEMP data cache</i>
-----------------	---

Description

Reports the location and size of the local cache directory where downloaded BEMP files are stored.

Usage

```
bemp_cache_info()
```

Value

Invisibly returns the cache directory path. Called for its side effect of printing cache information.

Examples

```
bemp_cache_info()
```

codebook	<i>Merged BEMP codebook across all waves</i>
----------	--

Description

A data frame containing the combined codebook for all 20 BEMP wave datasets. Each row describes one variable in one wave. Built from the per-wave codebook CSV files hosted on Zenodo; regenerate with `source("data-raw/build_data.R")`.

Usage

```
codebook
```

Format

A data frame with one row per variable per wave. Key columns:

- `wave`: Wave identifier, e.g. "w1", "w6_m", "w14_v".
- `variable_name`: Variable name as it appears in the dataset.
- `variable_label`: Short English label for the variable.
- `block`: Thematic block within the questionnaire.
- `question`: Short question description.
- `question_type`: Question type (e.g. single-choice, numeric).
- `question_text`: Full question text in English.
- `question_text_bn`: Full question text in Bangla.

- `appears_in_waves`: String listing all waves in which this variable appears, with observation counts.

Additional columns `val_0`, `val_1`, ... contain English value labels for coded response options; `0_bn`, `1_bn`, ... contain the corresponding Bangla value labels. Other columns include `item_source`, `item_text`, `item_text_bn`, `skip_logic`, `block_display_logic`, `block_randomization`, `question_display_logic`, `comment`, and `dataset`.

Source

Zenodo (concept DOI 10.5281/zenodo.18229497), `bemp_codebooks_as_csv.zip`

get_codebook

Get the codebook for one or all BEMP waves

Description

Returns the codebook for a specific wave, or the merged codebook across all waves. Codebooks are downloaded from Zenodo on first use (~4 MB) and cached locally. Alternatively, use the pre-built [codebook](#) object that ships with the package for offline use.

Usage

```
get_codebook(wave = "all", refresh = FALSE)
```

Arguments

<code>wave</code>	Character. Wave identifier (e.g. "w1", "w6_M") or "all" (default) to return the merged codebook for every wave. See wave_overview .
<code>refresh</code>	Logical. Re-download from Zenodo even if already cached. Default FALSE.

Value

A tibble with one row per variable. Key columns:

- `wave`: wave identifier
- `variable_name`, `variable_label`
- `block`, `question`, `question_type`
- `question_text`, `question_text_bn` (Bangla)
- `appears_in_waves`: cross-wave appearance string
- Value label columns (`val_0`, `val_1`, ...)

Examples

```
# Use a temporary cache so downloaded files are cleaned up after the session
old_cache <- Sys.getenv("BEMPDATADIR")
Sys.setenv(BEMPDATADIR = file.path(tempdir(), "BEMPdata"))

# Codebook for the baseline wave
cb_w1 <- get_codebook("w1")

# Merged codebook (all waves)
cb_all <- get_codebook()

Sys.setenv(BEMPDATADIR = old_cache)
```

get_wave

*Download and return a BEMP wave dataset***Description**

Downloads the requested wave dataset from Zenodo on first use and returns it as a tibble. All wave files share a single zip archive per format, so the first call downloads every wave at once (~6 MB for CSV, ~14 MB for Stata); subsequent calls are instant because files are cached locally.

Usage

```
get_wave(wave, format = "csv", refresh = FALSE)
```

Arguments

wave	Character. Wave identifier. Use lowercase with an underscore suffix for migrant (_M), non-migrant (_N), or village-profile (_V) sub-questionnaires. Examples: "w1", "w6_M", "w12_N", "w14_V". See wave_overview for the full list.
format	Character. "csv" (default) or "dta" (Stata). CSV returns a tibble ; DTA returns a labelled tibble via haven::read_dta with value labels attached.
refresh	Logical. Re-download from Zenodo even if already cached. Default FALSE.

Value

A tibble with one row per survey respondent.

Examples

```
# Use a temporary cache so downloaded files are cleaned up after the session
old_cache <- Sys.getenv("BEMPDATADIR")
Sys.setenv(BEMPDATADIR = file.path(tempdir(), "BEMPdata"))

# Baseline in-person wave
w1 <- get_wave("w1")
```

```
# Wave 6, migrant questionnaire (accepts upper or lower case)
w6m <- get_wave("w6_M")

# Village profile, Wave 14, in Stata format with value labels
w14v <- get_wave("w14_V", format = "dta")

Sys.setenv(BEMPDATADIR = old_cache)
```

lookup_variable	<i>Search for variables across all BEMP waves</i>
-----------------	---

Description

Searches the bundled [codebook](#) for variables whose name, label, or question text matches a pattern. Useful for finding a variable when you know a keyword but not the exact variable name.

Usage

```
lookup_variable(pattern, fields = c("name", "label", "question"))
```

Arguments

pattern	Character. A regular expression (case-insensitive). For simple keyword searches just supply a word, e.g. "income".
fields	Character vector. Which fields to search. One or more of "name", "label", "question". Default: all three.

Value

A tibble with columns wave, variable_name, variable_label, block, and question for all matching variables, sorted by wave.

Examples

```
# Find all income-related variables
lookup_variable("income")

# Search only variable labels for migration-related items
lookup_variable("migrat", fields = "label")
```

`run_app`*Launch the BEMP Data Explorer Shiny app*

Description

Opens an interactive data explorer with three tabs:

Codebook Browser Search and filter variables across all waves by keyword, thematic block, or wave. Click any row to see the full question text and value labels.

Variable Inspector Select a wave and variable to view its distribution (bar chart for categorical, histogram for numeric) and summary statistics.

Download Assistant Select a wave and a subset of variables, preview the data, and download as CSV.

Usage

```
run_app(...)
```

Arguments

... Additional arguments passed to `shiny::runApp()`, e.g. `port = 4321` or `launch.browser = FALSE`.

Value

Called for its side effect of launching a Shiny app.

Examples

```
if (interactive()) {  
  run_app()  
}
```

`wave_overview`*Overview of BEMP survey waves*

Description

A small reference table describing each of the 20 BEMP wave datasets: survey round, mode (in-person / phone), and questionnaire type.

Usage

```
wave_overview
```

Format

A data frame with 20 rows and 4 columns:

wave Wave identifier (character), e.g. "w1", "w6_M".

survey_round Integer survey round number (1–14).

type Survey mode: "in-person" or "phone".

questionnaire Questionnaire type: "main", "migrant", "non-migrant", or "village profile".

Examples

```
wave_overview  
wave_overview[wave_overview$type == "in-person", ]
```

Index

* datasets

- codebook, [3](#)
- wave_overview, [7](#)

bemp_cache_clear, [2](#)
bemp_cache_info, [3](#)

codebook, [3](#), [4](#), [6](#)

get_codebook, [4](#)
get_codebook(), [2](#)
get_wave, [5](#)
get_wave(), [2](#)

haven::read_dta, [5](#)

lookup_variable, [6](#)

run_app, [7](#)

shiny::runApp(), [7](#)

tibble, [5](#)

wave_overview, [4](#), [5](#), [7](#)