Package 'vismi'

February 3, 2026

Type Package

Title Visual Diagnostics for Multiple Imputation

Version 0.9.5

Description A comprehensive suite of static and interactive visual diagnostics for assessing the quality of multiply-imputed data obtained from packages such as 'mixgb' and 'mice'. The package supports inspection of distributional characteristics, diagnostics based on masking observed values and comparing them with re-imputed values, and convergence diagnostics.

URL <https://agnesdeng.github.io/vismi/>

BugReports <https://github.com/agnesdeng/vismi/issues>

License GPL (>= 3)

Encoding UTF-8

Language en-GB

LazyData true

Depends R (>= 4.3.0)

Imports cli, data.table, dplyr, GGally, ggplot2 (>= 4.0.1), ggtext, gridExtra, ggridges, mixgb (>= 2.2.3), patchwork, plotly, purrr, rlang, stats, scales, tidyr, trelliscopejs

RoxygenNote 7.3.3

NeedsCompilation no

Author Yongshi Deng [aut, cre] (ORCID: <https://orcid.org/0000-0001-5845-859X>), Thomas Lumley [ths]

Maintainer Yongshi Deng <agnes.yongshideng@gmail.com>

Repository CRAN

Date/Publication 2026-02-03 14:10:02 UTC

Contents

---

imp_newborn   *Precomputed mixgb imputed datasets for 'newborn'*

Description

A small precomputed list object containing 5 imputed datasets generated by 'mixgb::mixgb()' on the 'newborn' example data. This dataset is included so that users can run plotting examples without installing 'mixgb'.

Usage

```
data(imp_newborn)
```

Format

A list of 5 data.frames (each a completed dataset) created by 'mixgb::mixgb()' in development.

Source

Generated during package development with 'mixgb::mixgb()'.

| imp_nhanes3 | *Precomputed mixgb imputed datasets for 'nhanes3'* |

Description

A small precomputed list object containing 5 imputed datasets generated by 'mixgb::mixgb()' on the 'nhanes3' example data. This dataset is included so that users can run plotting examples without installing 'mixgb'.

Usage

```
data(imp_nhanes3)
```

Format

A list of 5 data.frames (each a completed dataset) created by 'mixgb::mixgb()' in development.

Source

Generated during package development with 'mixgb::mixgb()'.

| newborn | *NHANES III (1988-1994) newborn data* |

Description

This dataset is extracted from the NHANES III (1988-1994) for the age class Newborn (under 1 year). Please note that this example dataset only contains selected variables and is for demonstration purposes only.

Usage

```
data(newborn)
```

Format

A data frame of 2107 rows and 16 variables, adapted from the NHANES III dataset. Nine variables contain missing values. Variable names and factor levels have been renamed for clarity and easier interpretation.

household_size Household size. An integer variable ranging from 1 to 10. The original variable name in the NHANES III dataset is HSHSIZER.

age_months Age at interview (screener), in months. An integer variable ranging from 2 to 11. The original variable name in the NHANES III dataset is HSAGEIR.

sex Sex of the subject. A factor variable with levels Male and Female. The original variable name in the NHANES III dataset is HSSEX.

race Race of the subject. A factor variable with levels `White`, `Black`, and `Other`. The original variable name in the NHANES III dataset is `DMARACER`.

ethnicity Ethnicity of the subject. A factor variable with levels `Mexican-American`, `Other Hispanic`, and `Not Hispanic`. The original variable name in the NHANES III dataset is `DMAETHNR`.

race_ethinicity Combined race–ethnicity classification. A factor variable with levels `Non-Hispanic White`, `Non-Hispanic Black`, `Mexican-American`, and `Other`. The original variable name in the NHANES III dataset is `DMARETHN`.

head_circumference_cm Head circumference, in centimetres. Numeric. The original variable name in the NHANES III dataset is `BMPHEAD`.

recumbent_length_cm Recumbent length, in centimetres. Numeric. The original variable name in the NHANES III dataset is `BMPRECUM`.

first_subscapular_skinfold_mm First subscapular skinfold thickness, in millimetres. Numeric. The original variable name in the NHANES III dataset is `BMPSB1`.

second_subscapular_skinfold_mm Second subscapular skinfold thickness, in millimetres. Numeric. The original variable name in the NHANES III dataset is `BMPSB2`.

first_triceps_skinfold_mm First triceps skinfold thickness, in millimetres. Numeric. The original variable name in the NHANES III dataset is `BMPTR1`.

second_triceps_skinfold_mm Second triceps skinfold thickness, in millimetres. Numeric. The original variable name in the NHANES III dataset is `BMPTR2`.

weight_kg Body weight, in kilograms. Numeric. The original variable name in the NHANES III dataset is `BMPWT`.

poverty_income_ratio Poverty income ratio. Numeric. The original variable name in the NHANES III dataset is `DMPPIR`.

smoke Whether anyone living in the household smokes cigarettes inside the home. A factor variable with levels `Yes` and `No`. The original variable name in the NHANES III dataset is `HFF1`.

health General health status of the subject. An ordered factor with levels `Excellent`, `Very Good`, `Good`, `Fair`, and `Poor`. The original variable name in the NHANES III dataset is `HYD1`.

Source

https://wwwn.cdc.gov/nchs/nhanes/nhanes3/datafiles.aspx

References

U.S. Department of Health and Human Services (DHHS). National Center for Health Statistics. Third National Health and Nutrition Examination Survey (NHANES III, 1988-1994): Multiply Imputed Data Set. CD-ROM, Series 11, No. 7A. Hyattsville, MD: Centers for Disease Control and Prevention, 2001. Includes access software: Adobe Systems, Inc. Acrobat Reader version 4.

| nhanes3 | *A small subset of the NHANES III (1988-1994) newborn data* |

Description

This dataset is a small subset of newborn. It is for demonstration purposes only. More information on NHANES III data can be found on https://wwwn.cdc.gov/Nchs/Data/Nhanes3/7a/doc/mimodels.pdf

Usage

```
data(nhanes3)
```

Format

A data frame of 500 rows and 6 variables. Three variables have missing values.

age_months Age at interview (screener), in months. An integer variable ranging from 2 to 11. The original variable name in the NHANES III dataset is HSAGEIR.

sex Sex of the subject. A factor variable with levels Male and Female. The original variable name in the NHANES III dataset is HSSEX.

ethnicity Ethnicity of the subject. A factor variable with levels Mexican-American, Other Hispanic, and Not Hispanic. The original variable name in the NHANES III dataset is DMAETHNR.

head_circumference_cm Head circumference, in centimetres. Numeric. The original variable name in the NHANES III dataset is BMPHEAD.

recumbent_length_cm Recumbent length, in centimetres. Numeric. The original variable name in the NHANES III dataset is BMPRECUM.

weight_kg Body weight, in kilograms. Numeric. The original variable name in the NHANES III dataset is BMPWT.

Source

https://wwwn.cdc.gov/nchs/nhanes/nhanes3/datafiles.aspx

References

U.S. Department of Health and Human Services (DHHS). National Center for Health Statistics. Third National Health and Nutrition Examination Survey (NHANES III, 1988-1994): Multiply Imputed Data Set. CD-ROM, Series 11, No. 7A. Hyattsville, MD: Centers for Disease Control and Prevention, 2001. Includes access software: Adobe Systems, Inc. Acrobat Reader version 4.

---

overimp | *Overimpute main function*

Description

Overimp main function to call different imputation methods.

Usage

```
overimp(
  data,
  m = 5,
  p = 0.2,
  test_ratio = 0,
  method = "mixgb",
  seed = NULL,
  ...
)
```

Arguments

data	A data frame with missing values.
m	The number of imputation.
p	The extra proportion of missing values.
test_ratio	The proportion of test set. Default is 0, meaning no test set.
method	Can be one of the following: "mixgb","mice", and more in the future.
seed	Random seed.
...	Other arguments to be passed into the overimp function.

Value

An `overimp` object containing imputed training, test data (if applicable) and essential parameters required for plotting.

Examples

```
obj <- overimp(data = nhanes3, m = 3, p = 0.2, test_ratio = 0.2, method = "mixgb")
```

---

print.vismi   *print method for vismi objects*

Description

vismi Print method for vismi objects

Usage

```
## S3 method for class 'vismi'
print(x, ...)
```

Arguments

x	An object of class 'vismi' created by the `vismi.data.frame()` function.
...	Additional arguments (not used).

Value

A `vismi` object, returned invisibly.

---

trellis_vismi   *Trelliscope Visualisation of Distributional Characteristics*

Description

Generates a Trelliscope display for distributional characteristics across all variables.

Usage

```
trellis_vismi(
  data,
  imp_list,
  m = NULL,
  imp_idx = NULL,
  integerAsFactor = FALSE,
  title = "auto",
  subtitle = "auto",
  color_pal = NULL,
  marginal_x = "box+rug",
  nrow = 2,
  ncol = 4,
  path = NULL,
  verbose = FALSE,
  ...
)
```

Arguments

`data`	A data frame containing the original data with missing values.
`imp_list`	A list of imputed data frames.
`m`	An integer specifying the number of imputed datasets to plot. It should be smaller than `length(imp_list)`. Default is NULL (plot all).
`imp_idx`	A vector of integers specifying the indices of imputed datasets to plot. Default is NULL (plot all).
`integerAsFactor`	
	A logical value indicating whether to treat integer variables as factors (TRUE) or numeric (FALSE). Default is FALSE.
`title`	A string specifying the title of the plot. Default is "auto" (automatic title based on `x,y,z` input). If NULL, no title is shown.
`subtitle`	A string specifying the subtitle of the plot. Default is "auto" (automatic subtitle based on `x,y,z` input). If NULL, no subtitle is shown.
`color_pal`	A named vector of colors for different imputation sets. If NULL (default), a default color palette is used.
`marginal_x`	A character string specifying the type of marginal plot to add for the x variable in 2D plots. Options are "hist", "box", "rug", "box+rug", or NULL (default, no marginal plot) when interactive = TRUE. Options are "box", "rug", "box+rug", or NULL (default, no marginal plot) when interactive = FALSE.
`nrow`	Number of rows in the Trelliscope display. Default is 2.
`ncol`	Number of columns in the Trelliscope display. Default is 4.
`path`	Optional path to save the Trelliscope display. If NULL, the display will not be saved to disk.
`verbose`	A logical value indicating whether to print extra information. Default is FALSE.
`...`	Additional arguments passed to the underlying plotting functions, such as point_size, alpha, nbins, width, and boxpoints.

Value

A Trelliscope display object visualising distributional characteristics for all variables.

Examples

```
trellis_vismi(data = nhanes3, imp_list = imp_nhanes3, marginal_x = "box")
```

```
trellis_vismi_converge
```
Trelliscope Visualisation of Convergence Diagnostics

Description

Generates a Trelliscope display for convergence diagnostics across all variables.

Usage

```
trellis_vismi_converge(
  obj,
  tick_vals = NULL,
  color_pal = NULL,
  title = "auto",
  subtitle = "auto",
  nrow = 2,
  ncol = 4,
  path = NULL,
  verbose = FALSE,
  ...
)
```

Arguments

obj	An object of class 'mixgb' or 'mids' containing intermediate imputed result for each iteration.
tick_vals	A numeric vector specifying the tick values for the x-axis (iterations). If NULL, default tick values will be used.
color_pal	A vector of colors to use for the imputation lines. If NULL, default colors will be used.
title	A string specifying the title of the plot. If NULL, no title is shown. If "auto", a title will be generated based on the input. Default is "auto".
subtitle	A string specifying the subtitle of the plot. If NULL, no subtitle is shown. If "auto", a title will be generated based on the input. Default is "auto".
nrow	Number of rows in the Trelliscope display. Default is 2.
ncol	Number of columns in the Trelliscope display. Default is 4.
path	Optional path to save the Trelliscope display. If NULL, the display will not be saved to disk.
verbose	A logical value indicating whether to print extra information. Default is FALSE.
...	Additional arguments to customize the Trelliscope display.

Value

A Trelliscope display object visualising convergence diagnostics for all variables.

Examples

```
library(mixgb)
set.seed(2026)
mixgb_obj <- mixgb(data = nhanes3, m = 3, maxit = 4, pmm.type = "auto", save.models = TRUE)
trellis_vismi_converge(obj = mixgb_obj)
```

---

trellis_vismi_overimp   *Trelliscope Visualisation of Overimputation Diagnostics*

Description

Generates a Trelliscope display for overimputation diagnostics across all variables.

Usage

```
trellis_vismi_overimp(
  obj,
  m = NULL,
  imp_idx = NULL,
  integerAsFactor = FALSE,
  title = "auto",
  subtitle = "auto",
  num_plot = "cv",
  fac_plot = "cv",
  train_color_pal = NULL,
  test_color_pal = NULL,
  stack_y = FALSE,
  diag_color = "white",
  seed = 2025,
  nrow = 2,
  ncol = 4,
  path = NULL,
  verbose = FALSE,
  ...
)
```

Arguments

obj	An object of class 'overimp' containing imputed datasets and parameters.
m	A single positive integer specifying the number of imputed datasets to plot. It should be smaller than the total number of imputed datasets in the object. Default is NULL (plot all).
imp_idx	A vector of integers specifying the indices of imputed datasets to plot. Default is NULL (plot all).
integerAsFactor	
	A logical indicating whether integer variables should be treated as factors. Default is FALSE (treated as numeric).

title	A string specifying the title of the plot. Default is "auto" (automatic title). If NULL, no title is shown.
subtitle	A string specifying the subtitle of the plot. Default is "auto" (automatic subtitle). If NULL, no subtitle is shown.
num_plot	A character string specifying the type of plot for numeric variables. Options are "cv" (cross-validation), "ridge", or "density". Default is "cv".
fac_plot	A character string specifying the type of plot for categorical variables. Options are "cv" (cross-validation), "bar", or "dodge". Default is "cv".
train_color_pal	
	A vector of colors for the training data. If NULL, default colors will be used.
test_color_pal	A vector of colors for the test data. If NULL, default colors will be used.
stack_y	A logical indicating whether to stack y-values in the plots. Default is FALSE.
diag_color	A color specification for the diagonal line in the plots. Default is NULL.
seed	An integer seed for reproducibility. Default is 2025.
nrow	Number of rows in the Trelliscope display. Default is 2.
ncol	Number of columns in the Trelliscope display. Default is 4.
path	Optional path to save the Trelliscope display. If NULL, the display will not be saved to disk.
verbose	A logical value indicating whether to print extra information. Default is FALSE.
...	Additional arguments to customize the plots, such as point_size, xlim, ylim.

Value

A Trelliscope display object visualising overimputation diagnostics for all variables.

Examples

```
obj <- overimp(data = nhanes3, m = 3, p = 0.2, test_ratio = 0, method = "mixgb")
trellis_vismi_overimp(obj = obj, stack_y = TRUE)
```

---

vismi   *Visualise Multiple Imputations Through Distributional Characteristics*

Description

This function provides visual diagnostic tools for assessing multiply imputed datasets created with 'mixgb' or other imputers through inspecting the distributional characteristics of imputed variables. It supports 1D, 2D, and 3D visualisations for numeric and categorical variables using either interactive or static plots.

Usage

```
vismi(
  data,
  imp_list,
  x = NULL,
  y = NULL,
  z = NULL,
  m = NULL,
  imp_idx = NULL,
  interactive = FALSE,
  integerAsFactor = FALSE,
  title = "auto",
  subtitle = "auto",
  color_pal = NULL,
  marginal_x = "box+rug",
  marginal_y = NULL,
  verbose = FALSE,
  ...
)
```

Arguments

data	A data frame containing the original data with missing values.
imp_list	A list of imputed data frames.
x	A character string specifying the name of the variable to plot on the x axis. Default is NULL.
y	A character string specifying the name of the variable to plot on the y axis. Default is NULL.
z	A character string specifying the name of the variable to plot on the z axis. Default is NULL.
m	An integer specifying the number of imputed datasets used for visualisation. It should be smaller than `length(imp_list)`. Default is NULL (plot all).
imp_idx	A vector of integers specifying the indices of imputed datasets to plot. Default is NULL (plot all).
interactive	A logical value indicating whether to create an interactive plotly plot (TRUE by default) or a static ggplot2 plot (FALSE).
integerAsFactor	
	A logical value indicating whether to treat integer variables as factors (TRUE) or numeric (FALSE). Default is FALSE.
title	A string specifying the title of the plot. Default is "auto" (automatic title based on `x,y,z` input). If NULL, no title is shown.
subtitle	A string specifying the subtitle of the plot. Default is "auto" (automatic subtitle based on `x,y,z` input). If NULL, no subtitle is shown.
color_pal	A named vector of colors for different imputation sets. If NULL (default), a default color palette is used.

marginal_x   A character string specifying the type of marginal plot to add for the x variable in 2D plots. Options are "hist", "box", "rug", "box+rug"(default), or NULL when interactive = TRUE. Options are "box", "rug", "box+rug"(default), or NULL when interactive = FALSE.

marginal_y   A character string specifying the type of marginal plot to add for the y variable in 2D plots. Options are "hist", "box", "rug", "box+rug", or NULL (default, no marginal plot) when interactive = TRUE. Options are "box", "rug", "box+rug", or NULL (default, no marginal plot) when interactive = FALSE.

verbose   A logical value indicating whether to print extra information. Default is FALSE.

...   Additional arguments passed to the underlying plotting functions, such as point_size, alpha, nbins, width, and boxpoints.

Value

A plotly or ggplot2 object visualising the multiply-imputed data.

Examples

```
vismi(data = nhanes3, imp_list = imp_nhanes3, x = "weight_kg", y = "head_circumference_cm", z="sex")
```

---

vismi_converge   *Visualise convergence diagnostics*

Description

This function generates convergence diagnostic plots showing the mean and standard deviation (SD) of imputed values for a specified variable across iterations.

Usage

```
vismi_converge(
  obj,
  x,
  xlim = NULL,
  mean_lim = NULL,
  sd_lim = NULL,
  title = "auto",
  subtitle = "auto",
  tick_vals = NULL,
  color_pal = NULL,
  linewidth = 0.8,
  ...
)
```

Arguments

`obj`	A 'mixgb' object returned by `mixgb()` function or a 'mids' object returned by the `mice()` function.
`x`	The name of the variable to plot convergence for.
`xlim`	Optional numeric vector of length 2 specifying the x-axis limits for iterations.
`mean_lim`	Optional numeric vector of length 2 specifying the y-axis limits for mean values of the variable.
`sd_lim`	Optional numeric vector of length 2 specifying the y-axis limits for standard deviation values of the variable.
`title`	A string specifying the title of the plot. If NULL, no title is shown. If "auto", a title will be generated based on the input. Default is "auto".
`subtitle`	A string specifying the subtitle of the plot. If NULL, no subtitle is shown. If "auto", a title will be generated based on the input. Default is "auto".
`tick_vals`	Optional numeric vector specifying x-axis tick values for iterations.
`color_pal`	A vector of m color codes (e.g., hex codes). If NULL, default colors will be used.
`linewidth`	The line width for the plot lines. Default is 0.8.
`...`	Additional arguments.

Value

Two side-by-side ggplot2 object showing the mean and standard deviation (SD) of imputed values for a specified variable across iterations.

Examples

```
library(mixgb)
set.seed(2026)
mixgb_obj <- mixgb(data = nhanes3, m = 3, maxit = 4, pmm.type = "auto", save.models = TRUE)
vismi_converge(obj = mixgb_obj, x = "recumbent_length_cm")
```

---

vismi_overimp	*Visualise Multiple Imputation Through Overimputation*

Description

This function provides overimputation diagnostics for assessing imputations generated by 'mice', 'mixgb' or other imputers. It supports evaluation on both training and test data.

Usage

```
vismi_overimp(
  obj,
  x = NULL,
  y = NULL,
  z = NULL,
  m = NULL,
  imp_idx = NULL,
  integerAsFactor = FALSE,
  title = "auto",
  subtitle = "auto",
  num_plot = "cv",
  fac_plot = "cv",
  train_color_pal = NULL,
  test_color_pal = NULL,
  stack_y = FALSE,
  diag_color = NULL,
  seed = 2025,
  ...
)
```

Arguments

obj	Overimputation object of class 'overimp' created by the overimp() function.
x	A character string specifying the name of the variable to plot on the x axis. Default is NULL.
y	A character string specifying the name of the variable to plot on the y axis. Default is NULL.
z	A character string specifying the name of the variable to plot on the z axis. Default is NULL.
m	A single positive integer specifying the number of imputed datasets to plot. It should be smaller than the total number of imputed datasets in the object.
imp_idx	A vector of integers specifying the indices of imputed datasets to plot.
integerAsFactor	
	A logical indicating whether integer variables should be treated as factors. Default is FALSE (treated as numeric).
title	A string specifying the title of the plot. Default is "auto" (automatic title based on x,y,z input). If NULL, no title is shown.
subtitle	A string specifying the subtitle of the plot. Default is "auto" (automatic subtitle based on x,y,z input). If NULL, no subtitle is shown.
num_plot	A character string specifying the type of plot for numeric variables.
fac_plot	A character string specifying the type of plot for categorical variables.
train_color_pal	
	A vector of colors for the training data. If NULL, default colors will be used.
test_color_pal	A vector of colors for the test data. If NULL, default colors will be used.

stack_y	A logical indicating whether to stack y values in certain plots. Default is FALSE.
diag_color	A character string specifying the color of the diagonal line in scatter plots. Default is NULL.
seed	An integer specifying the random seed for reproducibility. Default is 2025.
...	Additional arguments to customize the plots, such as position, point_size, linewidth, alpha, xlim, ylim, boxpoints, width.

Value

An `overimp_plot` object displaying the overimputation plots for training and test data (if users set `test_ratio > 0` in the `overimp()` function.)

Examples

```
obj <- overimp(data = nhanes3, m = 3, p = 0.2, test_ratio = 0.2, method = "mixgb")
vismi_overimp(obj = obj, x = "head_circumference_cm", num_plot = "cv")
```

Index