Package 'vlda'

June 26, 2020

Type Package

Title Visualization of Multidimensional Longitudinal Data

Version 1.1.5

Author Seongwon Ryu [aut, cre],
Bohui Lee [aut],
Yongseok Choi [ctb]

Maintainer Seongwon Ryu <seongwon6154@gmail.com>

Depends R (>= 3.6.0)

Imports dplyr, ggrepel, ggplot2, ggiraph, ggsci

Description Assists in producing a plot that more effectively expresses changes over time for two different types (long format and wide format) using a consistent calling scheme for longitudinal data. It provides the ability to projection supplementary information (supplementary objects and variables) that can often occur in longitudinal data to graphs, as well as provides a new interactive implementation to perform the additional interpretation, so it is also useful for longitudinal data visuals analysis (see <http://lib.pusan.ac.kr/resource/e-article/?app=eds&mod=detail&record_id=edsker.000004649097&db_id=edsker> for more information).

License MIT + file LICENSE

URL https://github.com/pnuwon/vlda

BugReports https://github.com/pnuwon/vlda/issues

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-06-26 08:50:02 UTC

R topics documented:

`Depression` *Depression data*

Description

Data comparing two drugs to treat patients suffering from depression.

Usage

```
data(Depression)
```

Format

Case Number of patients

Diagnosis Classification based on the initial symptoms of depression (1 : Mild, 2 : Severe)

Drug Drugs given to patients (1 : New, 2 : Standard)

1week Depression symptoms 1week after taking the drug (1 : Abnormal, 2 : Normal)

2weeks Depression symptoms 2weeks after taking the drug (1 : Abnormal, 2 : Normal)

4weeks Depression symptoms 4weeks after taking the drug (1 : Abnormal, 2 : Normal)

Details

Patients in each group were randomly assigned to standard or new drugs, and the degree of each patient suffering from depression was classified as normal or abnormal after 1 week, 2 weeks, and 4 weeks of treatment.
A data frame with 800 rows and 6 variables

Case	Diagnosis	Drug	1week	2weeks	4weeks
1	1	2	2	2	2
2	1	2	2	2	2

3	1	2	2	2	2
4	1	2	2	2	2
5	1	2	2	2	2
6	1	2	2	2	2
7	1	2	2	2	2
8	1	2	2	2	2
9	1	2	2	2	2
10	1	2	2	2	2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
796	2	1	1	1	1
797	2	1	1	1	1
798	2	1	1	1	1
799	2	1	1	1	1
800	2	1	1	1	1

References

Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H. and Lehnen, R. C.(1977). A General Methodology for the Analysis of Experiments with Repeated Measurement of Categorical Data. *Biometrics*, **33**, 133-158.

Depression_column *Supplementary data to be added to Depression data*

Description

Artificially created data to add the degree and sex of depression after 6 weeks.

Usage

```
data(Depression_column)
```

Format

6week.1 A value of 1 indicates that depression is "Abnormal" after 6 weeks

6week.2 A value of 1 indicates that depression is "Normal" after 6 weeks

sex.1 A value of 1 indicates that the gender is "Male"

sex.2 A value of 1 indicates that the gender is "Female"

Details

As supplementary variables, for 800 patients, response at fourth time point (after 6 weeks) and gender that could affect depression were added to the columns.
Indicator matrix of 800 rows and 4 dummy variables.

	6weeks.1	6weeks.2	sex.1	sex.2
1	0	1	1	0
2	0	1	1	0
3	0	1	1	0
4	0	1	1	0
5	0	1	1	0
6	0	1	1	0
7	0	1	1	0
8	0	1	1	0
9	0	1	1	0
10	0	1	1	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
796	0	1	0	1
797	0	1	0	1
798	0	1	0	1
799	0	1	0	1
800	0	1	0	1

`Depression_row` *Supplementary data to be added to Depression data*

Description

Artificially generated data to add a placebo effect that affects the degree of depression.

Usage

```
data(Depression_row)
```

Format

Diagnosis.1 A value of 1 indicates the "Severe" of depression

Diagnosis.2 A value of 1 indicates the "Mild" of depression

Drug.1 A value of 1 indicates that the drug being taken is a “New drug”

Drug.2 A value of 1 indicates that the drug being taken is a “Standard drug”

1week.1 A value of 1 indicates that depression is "Abnormal" after 1 week

1week.2 A value of 1 indicates that depression is "Normal" after 1 week

2week.1 A value of 1 indicates that depression is "Abnormal" after 2 weeks

2week.2 A value of 1 indicates that depression is "Normal" after 2 weeks

4week.1 A value of 1 indicates that depression is "Abnormal" after 4 weeks

4week.2 A value of 1 indicates that depression is "Normal" after 4 weeks

Details

Supplementary 100 objects are patients who take placebo.
Indicator matrix of 100 rows and 10 dummy variables

	Diag.1	Diag.2	Drug.1	Drug.2	.	.	.	2week.2	4week.1	4week.2
1	1	0	0	0	.	.	.	0	1	0
2	1	0	0	0				0	1	0
3	1	0	0	0	.	.	.	0	1	0
4	1	0	0	0				0	0	1
5	1	0	0	0	.	.	.	0	0	1
6	1	0	0	0				0	0	1
7	1	0	0	0	.	.	.	0	0	1
8	1	0	0	0				0	0	1
9	1	0	0	0	.	.	.	1	1	0
10	1	0	0	0				1	1	0
.	.	.	.	.	.			.	.	.
.	.	.	.	.		.		.	.	.
.	.	.	.	.			.	.	.	.
96	0	1	0	0	.	.	.	1	0	1
97	0	1	0	0				1	0	1
98	0	1	0	0	.	.	.	1	0	1
99	0	1	0	0				1	0	1
100	0	1	0	0	.	.	.	1	0	1

`indicator` *Indicator matrix*

Description

Convert values of categorical variables into indicator matrix

Usage

```
indicator(x)
```

Arguments

`x`	A data frame of categorical data coded in numbers.

Value

Dummy_variables

Examples

```
## Long form
data(PTSD)
PTSD <- as.data.frame(PTSD)
# Transform a string or continuous class variable into factor
PTSD[,2:4] <- apply(PTSD[,2:4], 2, function(x) ifelse(x >= 3, 1, 0))
PTSD[,5] <-  ifelse(PTSD[,5] >= 6 , 1, 0)
PTSD <- data.frame(lapply(PTSD[,-1], function(x) as.factor(x)))
indicator(PTSD)


## Wide form
data(Depression)
str(Depression)
indicator(Depression[,-1])
```

`print.vlda` *Print a vlda object*

Description

Print method for vlda

Usage

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

Arguments

`x`	A vlda object to print
`...`	Other arguments not used by this method

Value

Invisibly returns the result of vlda, which is a list of components that contain the data itself, information etc.

PTSD *Post Traumatic Stress Disorder data*

Description

This data of 316 patients who survived the fire, each patient was measured at 3, 6 and 12 months after the fire.

Usage

```
data(PTSD)
```

Format

subject Patient number

control Self-control (A numeric vector)

problems The number of life problems (A numeric vector)

stress The number of stress events (A numeric vector)

cohesion Family cohesion (A numeric vector)

time Measured at 3, 6 and 12 months after the fire (1 : 3 months, 2 : 6 months, 3 : 12 months)

ptsd Post traumatic stress disorder, Outcome variable (Categorical vector) (0 : No, 1 : Yes)

Details

Control, problems, and stress were divided into upper and lower levels based on 3, and cohesion was divided into upper and lower levels based on 6. (0 : Low, 1 : high)
A data frame with 948 rows and 7 variables

	subject	control	problems	stress	cohesion	time	ptsd
1	15	3.22	5.62	1	8	1	0
2	15	3.17	5.38	0	8	2	0
3	15	3.28	3.75	1	8	3	0
4	18	2.56	9.25	0	8	1	1
5	18	3.44	4.38	0	8	2	0
6	18	3.33	2.38	0	8	3	0
7	19	2.72	7.75	1	8	1	1
8	19	2.78	7.75	1	7	2	1
9	19	2.78	7.50	1	7	3	0
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
943	570	3.72	2.75	0	7	1	0
944	570	3.89	2.25	0	7	2	0
945	570	3.67	1.25	0	7	3	0
946	571	3.56	3.00	0	7	1	0
947	571	2.94	1.88	0	7	2	0
948	571	3.50	2.75	0	7	3	0

Source

Allison (1991, chapter 8).

References

Allison, P. D. (2001). *Logistic Regression Using the SAS System, Theory and Application.* SAS Institute Inc.

`PTSD_column` *Supplementary data to be added to PTSD data*

Description

Artificially created data to add drinking level to PTSD data.

Usage

```
data(PTSD_column)
```

Format

Drinking.0 A value of 1 indicates that the degree of drinking is low

Drinking.1 A value of 1 indicates that the degree of drinking is high

Details

The degree of drinking (low, high) that can affect PTSD is added to the columns corresponding to the first to third time points for 316 patients.
Indicator matrix of 948 rows and 2 dummy variables.

	Drinking.0	Drinking.0
1	1	0
2	1	0
3	1	0
4	0	1
5	1	0
6	1	0
7	0	1
8	0	1
9	0	1
10	1	0
.	.	.
.	.	.
.	.	.
944	1	0
945	1	0
946	1	0

947	1	0
948	1	0

PTSD_row *Supplementary data to be added to PTSD data*

Description

Artificially created data to add variables after 18 months to PTSD data.

Usage

```
data(PTSD_row)
```

Format

control.0 A value of 1 indicates low control

control.1 A value of 1 indicates high control

problems.0 A value of 1 indicates that the degree of problems is low

problems.1 A value of 1 indicates that the degree of problems is high

stress.0 A value of 1 indicates that the degree of stress is low

stress.1 A value of 1 indicates that the degree of stress is high

cohesion.0 A value of 1 indicates that the degree of stress is low

cohesion.1 A value of 1 indicates that the degree of stress is high

time.1 Zero vector (All elements is zero)

time.2 Zero vector (All elements is zero)

time.3 Zero vector (All elements is zero)

ptsd.0 A value of 1 indicates a low post-traumatic stress disorder

ptsd.1 A value of 1 indicates a high post-traumatic stress disorder

Details

This data is a long form of control, problem, stress, stress, stress and PTSD added to the row, and is intended for 316 patients after 18 months.
Indicator matrix of 316 rows and 13 dummy variables.

	control.0	control.1	.	.	.	time.1	time.2	time.3	ptsd.0	ptsd.1
1	0	1	.	.	.	0	0	0	1	0
2	0	1				0	0	0	1	0
3	0	1	.	.	.	0	0	0	1	0
4	0	1				0	0	0	1	0
5	0	1	.	.	.	0	0	0	1	0

6	0	1			0	0	0	1	0
7	0	1	.	.	0	0	0	1	0
8	0	1			0	0	0	1	0
9	0	1	.	.	0	0	0	1	0
10	0	1			0	0	0	1	0
.	.	.	.		.	.	.	.	.
.	.	.		.	.	.	.	.	.
.	.	.		.	.	.	.	.	.
312	1	1	.	.	0	0	0	1	0
313	1	1			0	0	0	1	0
314	1	1	.	.	0	0	0	1	0
315	1	1			0	0	0	1	0
316	1	1	.	.	0	0	0	1	0

`Respiratory` *Patient's respiratory staus data*

Description

This is part of the data on the patient's respiratory status.

Usage

```
data(Respiratory)
```

Format

subject Number of patients

gender Patient gender (0 : Female, 1 : Male)

age Patient age (0 : Under 30, 1 : Over 30)

month Measurement time (0 : before, 1 : 1 month, 2 : 2months)

status Measurement status after taking placebo, response variable (0 : poor, 1 : good)

Details

57 patients were measured for good and bad by taking 3 measurements before, 1 and 2 months after taking placebo A data frame with 171 rows and 5 variables

	subject	gender	age	month	status
1	1	0	1	0	0
2	1	0	1	1	0
3	1	0	1	2	0
4	2	0	0	0	0
5	2	0	0	1	0

6	2	0	0	2	0
7	4	0	1	0	1
8	4	0	1	1	1
9	4	0	1	2	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
166	104	1	1	0	1
167	104	1	1	1	0
168	104	1	1	2	1
169	106	1	1	0	0
170	106	1	1	1	0
171	106	1	1	2	0

Source

Davis. (1977).

References

Davis, C. S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine*, **10**, 1959—1980.

`vlda` *Visualization of Longitudinal Data Analysis*

Description

Visualization of multidimensional longitudinal data based on the projection method using the indicator matrix.

Usage

```
vlda(x, object, time, type = c("long", "wide"))
```

Arguments

`x`	A data frame consisting of categorical data coded in numbers. Its n samples(`object`) should have been repeatedly measured through multiple time points; its p variables will be represented as variable coordinate. To keep track of which observation occurred in which time point, you must have included a variable, `Time`.
`object`	A vector of length n samples. The `object` who would have made repeatedly measure through multiple time points; the object is indicated by the name of the observation coordinate.
`time`	A time point of longitudinal data. Accepts a character string that denotes the name of the time variable.
`type`	A type of longitudinal data.

Details

The value returned by vlda is using as the main argument of vlda_plot and vlda_add function, the corresponding model. long-format is that each row is one time point per `object` So each `object` has `T` rows. All `T` values for each `object` are stacked–they're all in the one column; wide-format is that a `object` repeated responses will be in a single row, and each response is in a separate column. so $(Y_1, \ldots, Y_T)$ are the response variables obtained at time $t(= 1, \ldots, T)$.`type = c(long,wide)`

Value

`obs.coordinate`	A tibble data class of row coordinates. Each row represents row coordinates and the observations corresponding to each row are included in the `obs_list`
`var.coordinate`	The column coordinate.
`Eigen`	Summarize the `principal inertias(Eigenvalues)` that as a result of applying the above algorithm using the indicator matrix
`GOF`	`Goodness-of-fit` of the Approximation for 2-dimensional VLDA plot.

See Also

`vlda_add`
`vlda_plot`

Examples

```
## longform of the PTSD data
data(PTSD)
PTSD <- as.data.frame(PTSD)
PTSD[,2:4] <- apply(PTSD[,2:4], 2, function(x) ifelse(x >= 3, 1, 0))
PTSD[,5] <-  ifelse(PTSD[,5] >= 6 , 1, 0)
PTSD <- data.frame(lapply(PTSD, function(x) as.factor(x)))
vlda(x = PTSD, object = "subject", time = "time", type = "long")


## Wideform od the Depression data
data(Depression)
head(Depression)
vlda(Depression, object = "Case", time = c("1week", "2weeks", "4weeks"), type = "wide")
vlda(Depression, "Case", c("1week", "2weeks", "4weeks"), "wide")
```

`vlda_add` *Supplementary Objects and Variables*

Description

Add objects or variables with new information to the two-dimensional VLDA plot proposed for multidimensional longitudinal data.

Usage

```
vlda_add(fit, add.col = NULL, add.row = NULL, time.name = NULL)
```

Arguments

fit	An object returned by vlda()
add.col	A data matirx, The type of indicator matrix. Additional data sets in column format. $p \geq 2$
add.row	A data matirx, The type of indicator matrix. Additional data sets in row format. Supplemental data should have the same variable name as `fit$ind.mat` returned by `vlda`, and if it is not an indicator matrix, you can use it after generate an indicator matrix using `indicator` function built into vlda.
time.name	If supplemental data to add contains a time variable, it requires argument a character string that specifies the name of the time variable.

Details

The longitudinal data inevitably has the characteristic that supplementary data is added such as:

* Outcome variables measured at additional time points, such as $T+1, T+2, ...$ after the last time point T.
* New objects that are not previously measured.
* Other covariates that indicate the characteristics of objects.

Find coordinates representing objects and variables that are added in the VLDA plot already provided, through a method obtain that find coordinates on low-dimensional space for supplementary elements.

Value

...	Same as the result of vlda
sup.coordiante	A tibble data class. The coordinates of the new object created when adding supplemental data to the already provided vlda plot.

See Also

vlda

Examples

```
#### Supplementary row and column indicator matrix added ####
### long form ###
data(PTSD)
PTSD <- as.data.frame(PTSD)
PTSD[,2:4] <- apply(PTSD[,2:4], 2, function(x) ifelse(x >= 3, 1, 0))
PTSD[,5] <-  ifelse(PTSD[,5] >= 6 , 1, 0)
PTSD <- data.frame(lapply(PTSD, function(x) as.factor(x)))
fit <- vlda(x = PTSD, object = "subject", time = "time", type = "long")
```

```
data(PTSD_column) # The degree of drinking that may affect PTSD
PTSD_column <- as.matrix(PTSD_column)

data(PTSD_row) # Added to the row, and is intended for 316 patients after 18 months.
PTSD_row <- as.matrix(PTSD_row)


vlda_add(
fit,
add.row = PTSD_row,
add.col = PTSD_column
)


### Wide form ###
data(Depression)
fit2 <- vlda(x = Depression, object = "Case", time = c("1week", "2weeks", "4weeks"), type = "wide")

# Response after 6 weeks and gender were added the columns for 800 existing patients.
data(Depression_column)
Depression_column <- as.matrix(Depression_column)

# 100 patients who took placebo in each group of mild and severe were added to the rows.
data(Depression_row)
Depression_row <- as.matrix(Depression_row)


vlda_add(
fit2,
time.name = "6weeks",
add.row = Depression_row,
add.col = Depression_column
)
```

---

vlda_plot *VLDA Plot*

Description

Assists in producing a plot that more effectively expresses changes over time for two different types (long format and wide format) using a consistent calling scheme for longitudinal data. It provides the ability to projection supplementary information (supplementary objects and variables) that can often occur in longitudinal data to graphs, as well as provides a new interactive implementation to perform the additional interpretation, so it is also useful for longitudinal data visuals analysis.

Usage

```
vlda_plot(fit, rename = NULL, interactive = TRUE,
          title = NULL, title.col = NULL, title.size = 15, title.hjust = 0,
          subtitle = NULL, sub.col = NULL, sub.size = 15, sub.hjust = 0,
          labels = NULL, lab.col = NULL, lab.size = NULL, lab.face = NULL,
          legend.position = "bottom", legend.justification = NULL,
          linetype = 2, line.col = "red", font.size = 1.0, var.size = 2.5,
          obs.col = "darkgray", obs.size = 2.5, add.obs.col = "#666666",
          arrow.col = "orange", arrow.size = 0.5, arrow.type = "closed")
```

Arguments

fit	An object returned by vlda() or supplement()
rename	Rename a variable to another name
interactive	Use the interactive graphical elements (default TRUE)
title	Plot title. If NULL, the title is not shown (default NULL)
title.col	Title color (default color is black)
title.size	Title font size (default size = 15)
title.hjust	Alignment of title (Number from 0 (left) to 1 (right): left-aligned by default)
subtitle	Subtitle for the plot which will be displayed below the title
sub.col	Sub-title color (default color is black)
sub.size	Sub-title font size (default size = 15)
sub.hjust	Alignment of sub-title (Number from 0 (left) to 1 (right): left-aligned by default)
labels	Legend labels
lab.col	Legend labels color
lab.size	Legend labels size
lab.face	Legend labels font c("plain","bold","italic","bold.italic") default = "plain"
legend.position	The position of legends ("none","left","right","bottom","top", or two-element numeric vector) default is "bottom"
legend.justification	Anchor point for positioning legend inside plot ("center" or two-element numeric vector) or the justification according to the plot area when positioned outside the plot
linetype	Line types can be specified with: An integer or name: 0 = blank, 1 = solid, 2 = dashed, 3 = dotted, 4 = dotdash, 5 = longdash, 6 = twodash, as shown below:
line.col	Axis line color
font.size	Font size (left-aligned by default size = 1.0)
var.size	Variable coordinate point size of plot
obs.col	Observation coordinate point color of plot

obs.size	Observation coordinate point size on plot
add.obs.col	Color of added observation coordinate points
arrow.col	Arrow color (default color = "orange")
arrow.size	Arrow size (default size = 0.5)
arrow.type	One of "open" or "closed" indicating whether the arrow head should be a closed triangle

Details

Coordinates in opposite directions on each axis can be considered to be different groups. And if the distance between the coordinates is close, it indicates that the group has a similar tendency. Even if the explanatory variable is not significant, a small tendency can confirm because the coordinate is placed in consideration of the relative influence.

Value

...	Same as the result of vlda
graphics	As a result of vlda, it creates a two-dimensional graph. provides interactive graphics, so when the mouse cursor points to the observation coordinates, it provides a tooltip that displays observations of having the same coordinates and displays the row and column coordinate. In the case of long-form, the tooltip displays a time point, besides, coordinate having the same time point are filled with the yellow color on the graph, to make it easier to distinguish the same time points of observations with colors. In the case of a wide form, the combinations that the explanatory variables can have are grouped and the coordinates points of the corresponding observations are shown in yellow on the graph. changes in time points are indicated by orange arrows on the graph.

See Also

vlda

Examples

```
### Long form ###
data(PTSD)
PTSD[,2:4] <- apply(PTSD[,2:4], 2, function(x) ifelse(x >= 3, 1, 0))
PTSD[,5] <-  ifelse(PTSD[,5] >= 6 , 1, 0)
PTSD <- data.frame(lapply(PTSD, function(x) as.factor(x)))
PTSD
str(PTSD)
head(PTSD, 10)
fit <- vlda(x = PTSD, object = "subject", time = "time", type = "long")
vlda_plot(fit)

 ## row and column ##
data(PTSD_row)
data(PTSD_column)
PTSD_row <- as.matrix(PTSD_row)
```

```
PTSD_column <- as.matrix(PTSD_column)

fit2 <- vlda_add(fit, add.row = PTSD_row, add.col = PTSD_column)
vlda_plot(fit2)


### Wide form ###
data(Depression)
wide.fit <-
vlda(
  x = Depression,
  object = "Case",
  time = c("1week", "2weeks", "4weeks"),
  type = "wide"
)
vlda_plot(wide.fit)
```

Index