

Package ‘panelView’

September 24, 2023

Type Package

Title Visualizing Panel Data

Version 1.1.17

Date 2023-09-23

Maintainer Yiqing Xu <yiqingxu@stanford.edu>

Description Visualizes panel data. It has three main functionalities: (1) it plots the treatment status and missing values in a panel dataset; (2) it visualizes the temporal dynamics of a main variable of interest; (3) it depicts the bivariate relationships between a treatment variable and an outcome variable either by unit or in aggregate. For details, see <[doi:10.18637/jss.v107.i07](https://doi.org/10.18637/jss.v107.i07)>.

URL <https://yiqingxu.org/packages/panelview/index.html>

License MIT + file LICENSE

Imports ggplot2 (>= 2.1.0), gridExtra, grid, dplyr (>= 1.0.0)

Depends R (>= 2.10)

RoxygenNote 6.0.1

NeedsCompilation no

Author Hongyu Mou [aut],
Licheng Liu [aut],
Yiqing Xu [aut, cre] (<<https://orcid.org/0000-0003-2041-6671>>)

Repository CRAN

Date/Publication 2023-09-24 00:00:02 UTC

R topics documented:

capacity	2
panelview	2
simdata	5
turnout	6
Index	7

capacity	<i>State capacity dataset</i>
----------	-------------------------------

Description

Cross-national panel data on democracy and state capacity.

Format

dataframe

References

The democracy measure (demo) comes from Boix, Miller, and Rosato (2013). The state capacity measure (Capacity) comes from Hanson and Sigman (2013.)

Reference: Erik H. Wang and Yiqing Xu (2018). "Awakening Leviathan: the Effect of Democracy on State Capacity, 1960-2009." *Research and Politics*, Vol. 5, Iss. 2, April-June 2018, pp. 1-7.

panelview	<i>Panel data visualization</i>
-----------	---------------------------------

Description

Visualizes missing values, treatment and outcome variables, and their relationships in panel data

Usage

```
panelview(data, formula = NULL, Y = NULL, D = NULL,
          X = NULL, index,
          ignore.treat = FALSE, type = "treat",
          outcome.type = "continuous",
          treat.type = NULL, by.group = FALSE, by.group.side = FALSE,
          by.timing = FALSE, theme.bw = TRUE,
          xlim = NULL, ylim = NULL,
          xlab = NULL, ylab = NULL,
          gridOff = FALSE, legendOff = FALSE,
          legend.labs = NULL, main = NULL,
          pre.post = NULL, id = NULL, show.id = NULL,
          color = NULL, axis.adjust = FALSE, axis.lab = "both",
          axis.lab.gap = c(0, 0), axis.lab.angle = NULL, shade.post = TRUE,
          cex.main = 15, cex.main.sub = 12, cex.axis = 8,
          cex.axis.x = NULL, cex.axis.y = NULL,
          cex.lab = 12, cex.legend = 12, background = NULL,
          style = NULL, by.unit = FALSE, lwd = 0.2, leave.gap = FALSE,
          display.all = NULL, by.cohort = FALSE,
          collapse.history = NULL, report.missing = FALSE)
```

Arguments

<code>data</code>	a data frame. The panel does not have to be balanced.
<code>formula</code>	an object of class "formula": a symbolic description of the model to be fitted. The first variable on the right-hand-side is designated as the treatment indicator if <code>ignore.treat = FALSE</code> . If there is not any covariates, the formula should be like $Y \sim 1$, where Y is the outcome variable.
<code>Y</code>	variable name of the outcome. Ignored if <code>formula</code> is provided.
<code>D</code>	variable name of the treatment. Ignored if <code>formula</code> is provided.
<code>X</code>	variable name of the time-varying covariates. Ignored if <code>formula</code> is provided.
<code>index</code>	a two-element string vector specifying the unit (group) and time indicators. Must be of length 2.
<code>ignore.treat</code>	a logical flag indicating whether there is a treatment variable. Default value is <code>ignore.treat = FALSE</code> .
<code>type</code>	a string that specifies the type of the plot. Must be either "treat" (default), which plots the treatment status of each unit at each time point, "missing", which plots the missing-data, "outcome", which plots the raw outcome data, or "bivariate", which plots time series of outcome and treatment in one graph.
<code>outcome.type</code>	a string that specifies the type of outcome variable. Must be either "continuous" (default) or "discrete". For a continuous variable, time series lines for specified units will be plotted, and for discrete response, jitter-ed points at each time period will be plotted.
<code>treat.type</code>	a string that specifies the type of treatment variable. Must be either "continuous" or "discrete". The default is NULL, which means the option will be decided based on the number of unique treatment values: if the number is bigger than 10, it will be set as "continuous"; otherwise, it will be set as "discrete".
<code>by.group</code>	a logical flag indicating whether the data should be plotted in a column in separate groups based on treatment status changes for the outcome plot.
<code>by.group.side</code>	a logical flag indicating whether to arrange subfigures of <code>by.group = TRUE</code> in a row rather than in a column.
<code>by.timing</code>	a logical flag indicating whether the units should be sorted based on the timing of receiving the treatment for the treat plot.
<code>theme.bw</code>	a logical flag specifying whether to use a black-and-white theme.
<code>xlim</code>	a two-element numeric vector specifying the range of x-axis. When the class of time variable is string, must specify the range of strings to be shown, e.g. <code>xlim=c(1,30)</code> .
<code>ylim</code>	a two-element numeric vector specifying the range of y-axis.
<code>xlab</code>	a string indicating the label of the x-axis.
<code>ylab</code>	a string indicating the label of the y-axis.
<code>gridOff</code>	a logical flag controlling whether to show the grid lines on the treat plot..
<code>legendOff</code>	a logical flag controlling whether to show the legend.
<code>legend.labs</code>	a vector specifying the legend labels. Ignored when <code>legendOff=TRUE</code> .
<code>main</code>	a string that controls the title of the plot.

pre.post	a logical flag indicating whether to distinguish control status of treated units from that of control units. Only used for staggered data in the treat and outcome plots.
id	a vector specifying units to be shown in the plot. Useful when the number of units is very large.
show.id	a numeric vector or sequence specifying the sorted order of units to be shown in the "treat" plot. Useful when the number of units is very large. Ignored if !is.null("id").
color	a string vector specifying color setting for the plot.
axis.adjust	a logic flag indicating whether to adjust labels on the x-axis. Useful when the class of time variable is string and there are many time periods.
axis.lab	a string indicating whether labels on the x- and y-axis will be shown. There are four options: "both" (default): labels on both axes will be shown; "unit": only labels on y-axis will be shown; "time": only labels on the x-axis will be shown; "none": no labels will be shown.
axis.lab.gap	a numeric vector setting the gaps between labels on the x- or y-axis for the plot. Default is axis.lab.gap = c(0, 0), which means that all labels will be shown. Useful for datasets with large N or T.
axis.lab.angle	a numeric value setting the angle (degrees) of the labels shown on the x-axis. Must be between 0 and 90.
shade.post	a logical flag controlling whether to shade the post-treatment periods. Ignored if type = "treat" or no treatment variable is supplied.
cex.main	a numeric value (pt) specifying the fontsize of the main title.
cex.main.sub	a numeric value (pt) specifying the fontsize of the subtitles. Ignored if type = "treat" or by.group = FALSE.
cex.axis	a numeric value (pt) specifying the fontsize of the texts on the axes; overwritten by cex.axis.x or cex.axis.y.
cex.axis.x	a numeric value (pt) specifying the fontsize of the texts on the x-axis.
cex.axis.y	a numeric value (pt) specifying the fontsize of the texts on the y-axis.
cex.lab	a numeric value (pt) specifying the fontsize of the axis titles.
cex.legend	a numeric value (pt) specifying the fontsize of the legend.
background	a character specifying the background color.
by.unit	a logic flag indicating whether to plot by each specified units or to plot mean D and Y against time in the same graph.
style	a string vector to set line/connected line/bar styles for the outcome and treatment variables.
lwd	a numeric value (pt) specifying the line width when plotting time series of treatment and outcome variables.
leave.gap	a logical flag indicating whether to keep time gaps as white bars if time is not evenly distributed (possibly due to missing data). Default value is leave.gap = FALSE.
display.all	a logical flag indicating whether to show all units if the number of units is more than 500, otherwise we randomly select 500 units to present.

- `by.cohort` a logical flag indicating whether to plot the average outcome lines based on unique treatment histories in an "outcome" plot.
- `collapse.history` a logical flag indicating whether to collapse units by treat history in a "treat" plot.
- `report.missing` a logical flag indicating whether to report missingness in the included variables.

Details

`panelview` visualizes the treatment status, missing values, and raw outcome data of a time-series cross-sectional dataset.

Author(s)

Hongyu Mou <hongyumou@g.ucla.edu>

Licheng Liu <liulch@mit.edu>

Yiqing Xu <yiqingxu@stanford.edu>

References

Hongyu Mou, Licheng Liu and Yiqing Xu (2023). "Panel Data Visualization in R (`panelView`) and Stata (`panelview`)." *Journal of Statistical Software*, 107(7), pp. 1–20. <doi:10.18637/jss.v107.i07>

Examples

```
library(panelView)
data(panelView)
panelview(turnout ~ policy_edr + policy_mail_in + policy_motor,
           data = turnout, index = c("abb", "year"))
```

simdata

A simulated dataset

Description

A simulated panel dataset with a discrete outcome variable.

Format

dataframe

References

Hongyu Mou, Licheng Liu and Yiqing Xu (2023). "Panel Data Visualization in R (`panelView`) and Stata (`panelview`)." *Journal of Statistical Software*, 107(7), pp. 1–20. <doi:10.18637/jss.v107.i07>

turnout

Voter turnout data.

Description

State-level voter turnout data.

Format

dataframe

References

Melanie Jean Springer (2014). *How the States Shaped the Nation: American Electoral Institutions and Voter Turnout, 1920-2000*. University of Chicago Press.

Yiqing Xu (2017). "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis*, 25(1), pp. 57–76.

Index

* datasets

capacity, [2](#)

simdata, [5](#)

turnout, [6](#)

capacity, [2](#)

panelview, [2](#)

simdata, [5](#)

turnout, [6](#)