

Package ‘APCtools’

October 12, 2022

Title Routines for Descriptive and Model-Based APC Analysis

Version 1.0.1

Maintainer Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>

Description Age-Period-Cohort (APC) analyses are used to differentiate relevant drivers for long-term developments.
The ‘APCtools’ package offers visualization techniques and general routines to simplify the workflow of an APC analysis.
Sophisticated functions are available both for descriptive and regression model-based analyses. For the former, we use density (or ridgeline) matrices and (hexagonally binned) heatmaps as innovative visualization techniques building on the concept of Lexis diagrams.
Model-based analyses build on the separation of the temporal dimensions based on generalized additive models, where a tensor product interaction surface (usually between age and period) is utilized to represent the third dimension (usually cohort) on its diagonal.
Such tensor product surfaces can also be estimated while accounting for further covariates in the regression model.
See Weigert et al. (2021) <[doi:10.1177/1354816620987198](https://doi.org/10.1177/1354816620987198)> for methodological details.

License MIT + file LICENSE

URL <https://bauer-alex.github.io/APCtools/>

BugReports <https://github.com/bauer-alex/APCtools/issues>

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

Imports ggpubr, checkmate, knitr, ggplot2, colorspace, dplyr, mgcv, scales, tidyr

Suggests testthat (>= 3.0.0), rmarkdown, covr

VignetteBuilder knitr

Depends R (>= 3.5)

Config/testthat/edition 3

NeedsCompilation no

Author Alexander Bauer [aut, cre] (<<https://orcid.org/0000-0003-3495-5131>>),
 Maximilian Weigert [aut] (<<https://orcid.org/0000-0003-4400-134X>>),
 Hawre Jalal [aut] (<<https://orcid.org/0000-0002-8224-6834>>),
 Pauline Hohenemser [ctb]

Repository CRAN

Date/Publication 2022-01-11 10:22:51 UTC

R topics documented:

calc_density	3
capitalize_firstLetter	3
compute_xCoordinate	4
compute_yCoordinate	4
create_APCsummary	4
create_groupVariable	6
create_highlightDiagonalData	6
create_modelSummary	7
create_oneAPCsummaryTable	8
drug_deaths	8
ensure_segmentsInPlotRange	9
extract_summary_linearEffects	10
get_plotGAMobject	10
gg_addReferenceLines	11
gg_highlightDiagonals	11
plot_1Dsmooth	12
plot_APCheatmap	13
plot_APChexamap	16
plot_density	18
plot_densityMatrix	20
plot_density_categorical	23
plot_density_metric	24
plot_jointMarginalAPCEffects	25
plot_linearEffects	26
plot_marginalAPCEffects	27
plot_partialAPCEffects	28
plot_variable	30
travel	32

Index

34

calc_density	<i>Internal helper to calculate the (group-specific) density of a variable</i>
--------------	--

Description

Internal helper function that is called in `plot_density` to calculate the density of a metric variable. If `plot_density` is called from within `plot_densityMatrix` (i.e., when some of the columns `c("age_group", "period_group", "cohort_group")` are part of the dataset, the density is computed individually for all respective APC groups.

Usage

```
calc_density(dat, y_var, weights_var = NULL, ...)
```

Arguments

<code>dat</code>	Dataset with columns <code>period</code> and <code>age</code> and the main variable specified through argument <code>y_var</code> .
<code>y_var</code>	Character name of the main variable to be plotted.
<code>weights_var</code>	Optional character name of a weights variable used to project the results in the sample to some population.
<code>...</code>	Additional arguments passed to <code>density</code> .

Value

Dataset with the calculated densities.

capitalize_firstLetter	<i>Internal function to capitalize the first letter of a character</i>
------------------------	--

Description

Internal helper function to capitalize the first letter of a character value. The use case is to create a plot label like 'Age' from a variable name like 'age'.

Usage

```
capitalize_firstLetter(char)
```

Arguments

<code>char</code>	Character value whose first letter should be capitalized
-------------------	--

compute_xCoordinate *Internal helper to tilt the x-axis for the hexamap plot*

Description

Internal helper function to be called in [plot_APChexamap](#), to tilt the x-axis for the hexamap plot.

Usage

```
compute_xCoordinate(period_vec)
```

Arguments

period_vec Numeric vector of period values.

compute_yCoordinate *Internal helper to tilt the x-axis for the hexamap plot*

Description

Internal helper function to be called in [plot_APChexamap](#), to tilt the x-axis for the hexamap plot.

Usage

```
compute_yCoordinate(period_vec, age_vec)
```

Arguments

period_vec Numeric vector of period values.
age_vec Numeric vector of age values.

create_APCsummary *Create a summary table for multiple estimated GAM models*

Description

Create a table to summarize the overall effect strengths of the age, period and cohort effects for models fitted with [gam](#) or [bam](#). The output format can be adjusted by passing arguments to [kable](#) via the ... argument.

Usage

```
create_APCsummary(model_list, dat, digits = 2, apc_range = NULL, ...)
```

Arguments

<code>model_list</code>	A list of regression models estimated with gam or bam . If the list is named, the names are used as labels.
<code>dat</code>	Dataset with columns <code>period</code> and <code>age</code> . If <code>y_var</code> is specified, the dataset must contain the respective column. If <code>model</code> is specified, the dataset must have been used for model estimation with gam or bam .
<code>digits</code>	Number of digits for numeric columns. Defaults to 2.
<code>apc_range</code>	Optional list with one or multiple elements with names <code>"age"</code> , <code>"period"</code> , <code>"cohort"</code> to filter the data. Each element should contain a numeric vector of values for the respective variable that should be kept in the data. All other values are deleted before producing the table.
<code>...</code>	Optional additional arguments passed to kable .

Details

If the model was estimated with a log or logit link, the function automatically performs an exponential transformation of the effect.

Value

Table created with [kable](#).

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>

Examples

```
library(APCtools)
library(mgcv)

data(travel)
model_pure <- gam(mainTrip_distance ~ te(age, period), data = travel)
model_cov  <- gam(mainTrip_distance ~ te(age, period) + s(household_income),
                  data = travel)

model_list <- list("pure model"      = model_pure,
                  "covariate model" = model_cov)

create_APCsummary(model_list, dat = travel)
```

create_groupVariable *Internal helper to create a group variable as base for a density matrix*

Description

Internal helper function to create a group variable based on the categorization of either age, period or cohort. To be called from within [plot_densityMatrix](#).

Usage

```
create_groupVariable(dat, APC_var, groups_list)
```

Arguments

dat	Dataset with a column "age", "period" or "cohort", dependent on the specified APC_var.
APC_var	One of c("age", "period", "cohort").
groups_list	A list with each element specifying the borders of one row or column in the density matrix. E.g., if the period should be visualized in decade columns from 1980 to 2009, specify groups_list = list(c(1980,1989), c(1990,1999), c(2000,2009)). The list can be named to specify labels for the categories.

Value

Vector for the grouping that can be added as additional column to the data.

create_highlightDiagonalData
Internal helper to create a dataset for ggplot2 to highlight diagonals

Description

Internal helper function to create a dataset for ggplot2 that can be used to highlight specific diagonals in a density matrix.

Usage

```
create_highlightDiagonalData(dat, highlight_diagonals)
```

Arguments

dat	Dataset with columns period and age and the main variable specified through argument y_var.
highlight_diagonals	Optional internal parameter which is only specified when plot_density is called from within plot_densityMatrix. See plot_densityMatrix for details.

create_modelSummary *Create model summary tables for multiple estimated GAM models*

Description

Create publication-ready summary tables of all linear and nonlinear effects for models fitted with [gam](#) or [bam](#). The output format of the tables can be adjusted by passing arguments to [kable](#) via the ... argument.

Usage

```
create_modelSummary(model_list, digits = 2, ...)
```

Arguments

model_list	list of APC models
digits	number of displayed digits
...	additional arguments to kable

Details

If the model was estimated with a log or logit link, the function automatically performs an exponential transformation of the effects.

The table for linear coefficients includes the estimated coefficient (coef), the corresponding standard error (se), lower and upper limits of 95% confidence intervals (CI_lower, CI_upper) and the p-values for all coefficients apart from the intercept.

The table for nonlinear coefficients include the estimated degrees of freedom (edf) and the p-value for each estimate.

Value

List of tables created with [kable](#).

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>

Examples

```
library(APCtools)
library(mgcv)

data(travel)
model <- gam(mainTrip_distance ~ te(age, period) + residence_region +
             household_size + s(household_income), data = travel)

create_modelSummary(list(model), dat = travel)
```

```
create_oneAPCsummaryTable
```

Internal helper to create a summary table for one estimated GAM model

Description

Internal helper function to be called in `create_APCsummary`. This function creates the summary table for one model estimated with `gam` or `bam`.

Usage

```
create_oneAPCsummaryTable(model, dat, apc_range = NULL)
```

Arguments

<code>model</code>	Optional regression model estimated with <code>gam</code> or <code>bam</code> to estimate a smoothed APC surface. Only used if <code>y_var</code> is not specified.
<code>dat</code>	Dataset with columns <code>period</code> and <code>age</code> . If <code>y_var</code> is specified, the dataset must contain the respective column. If <code>model</code> is specified, the dataset must have been used for model estimation with <code>gam</code> or <code>bam</code> .
<code>apc_range</code>	Optional list with one or multiple elements with names <code>"age"</code> , <code>"period"</code> , <code>"cohort"</code> to filter the data. Each element should contain a numeric vector of values for the respective variable that should be kept in the data. All other values are deleted.

Value

`data.frame` containing aggregated information on the individual effects.

```
drug_deaths
```

Drug deaths of white men in the United States

Description

Dataset on the number of unintentional drug overdose deaths in the United States for each age group between 1999 and 2019, retrieved from the CDC WONDER Online Database. The data only cover white men.

Usage

```
data(drug_deaths)
```


Format

A dataframe containing

period Calendar year

age Age group.

deaths Number of observed unintentional drug overdose deaths in the respective age group and calendar year.

population Number of white men in the respective age group and calendar year in the U.S. population.

mortality_rate Drug overdose mortality rate for the respective age group and calendar year, reported as the number of deaths per 100,000 people. Calculated as $100000 * \text{deaths} / \text{population}$.

Details

The data were exported from the CDC WONDER Online Database (see link in references down below), based on the following settings:

- Group by *Year* and by *Single-Year Ages*
- Demographics: Gender *Male*; Ethnicity *White*
- Cause of death: *Drug / Alcohol Induced Causes*. Then select the more specific category *Drug poisonings (overdose) Unintentional (X40-X44)*.

References

Jalal, H., & Burke, D. S. (2020). Hexamaps for Age-Period-Cohort Data Visualization and Implementation in R. *Epidemiology (Cambridge, Mass.)*, 31(6), e47. doi: [10.1097/EDE.0000000000001236](https://doi.org/10.1097/EDE.0000000000001236).

Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999-2019 on CDC WONDER Online Database, released in 2020. Data are from the Multiple Cause of Death Files, 1999-2019, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at wonder.cdc.gov/ucd-icd10.html on 6 December 2021.

ensure_segmentsInPlotRange

Internal helper for gg_addReferenceLines to keep diagonal lines in the plot range

Description

Internal helper function to be called from within `gg_addReferenceLines`. This function takes the dataset prepared for adding diagonal reference lines in the plot, checks if some diagonals exceed the plot limits, cuts them accordingly, if necessary, and again returns the corrected dataset.

Usage

```
ensure_segmentsInPlotRange(dat_segments, plot_dat)
```

Arguments

dat_segments Dataset containing information on the diagonal reference lines.
 plot_dat Dataset used for creating the heatmap.

extract_summary_linearEffects

Internal helper to extract summary of linear effects in a gam model

Description

Internal helper function to create a data.frame containing the linear effects summary of a model fitted with [gam](#) or [bam](#).

Usage

```
extract_summary_linearEffects(model)
```

Arguments

model Model fitted with [gam](#) or [bam](#).

Details

If the model was estimated with a log or logit link, the function automatically performs an exponential transformation of the effect.

get_plotGAMobject

Extract returned values of plot.gam() while suppressing creation of the plot

Description

Internal helper function to extract the values returned of [plot.gam](#) while suppressing creation of the plot.

Usage

```
get_plotGAMobject(model)
```

Arguments

model GAM model fitted with [gam](#) or [bam](#).

gg_addReferenceLines *Internal helper to add reference lines in an APC heatmap*

Description

Internal helper function to add reference lines in an APC heatmap (vertically, horizontally or diagonally). The function takes an existing list of ggplot objects, adds the specified reference lines in each plot and returns the edited ggplot list again. To be called from within [plot_APChemap](#).

Usage

```
gg_addReferenceLines(  
  gg_list,  
  dimensions,  
  plot_dat,  
  markLines_list,  
  markLines_displayLabels  
)
```

Arguments

gg_list	Existing list of ggplot objects where the reference lines should be marked in each individual ggplot.
dimensions	Character vector specifying the two APC dimensions that should be visualized along the x-axis and y-axis. Defaults to c("period", "age").
plot_dat	Dataset used for creating the heatmap.
markLines_list	Optional list that can be used to highlight the borders of specific age groups, time intervals or cohorts. Each element must be a numeric vector of values where horizontal, vertical or diagonal lines should be drawn (depends on which APC dimension is displayed on which axis). The list can maximally have three elements and must have names out of c("age", "period", "cohort").
markLines_displayLabels	Optional character vector defining for which dimensions the lines defined through markLines_list should be marked by a respective label. The vector should be a subset of c("age", "period", "cohort"), or NULL to suppress all labels. Defaults to c("age", "period", "cohort").

gg_highlightDiagonals *Internal helper to add the diagonal highlighting to a ggplot*

Description

Internal helper function to highlight diagonals in a density matrix. The function takes an existing ggplot object, adds the diagonal highlighting and returns the edited ggplot object again.

Usage

```
gg_highlightDiagonals(gg, dat, dat_highlightDiagonals)
```

Arguments

`gg` Existing ggplot object to which the diagonal highlighting should be added.

`dat` Dataset with columns `period` and `age` and the main variable specified through argument `y_var`.

`dat_highlightDiagonals` Dataset created by [create_highlightDiagonalData](#) to highlight specific diagonals in a density matrix.

plot_1Dsmooth	<i>Plot 1D smooth effects for gam models</i>
---------------	--

Description

Plots 1D smooth effects for a GAM model fitted with [gam](#) or [bam](#).

Usage

```
plot_1Dsmooth(
  model,
  plot_ci = TRUE,
  select,
  alpha = 0.05,
  ylim = NULL,
  return_plotData = FALSE
)
```

Arguments

`model` GAM model fitted with [gam](#) or [bam](#).

`plot_ci` If TRUE CIs are plotted. Only used if `plot_type = 1`.

`select` Index of smooth term to be plotted.

`alpha` (1-alpha) CIs are calculated. The default 0.05 leads to 95% CIs.

`ylim` Optional limits of the y-axis.

`return_plotData` If TRUE, the dataset prepared for plotting is returned. Defaults to FALSE.

Details

If the model was estimated with a log or logit link, the function automatically performs an exponential transformation of the effect.

Value

ggplot object

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>

Examples

```
library(APCtools)
library(mgcv)

data(travel)
model <- gam(mainTrip_distance ~ te(age, period) + residence_region +
             household_size + s(household_income), data = travel)

plot_1Dsmooth(model, select = 2)
```

plot_APHeatmap

Heatmap of an APC surface

Description

Plot the heatmap of an APC structure. The function can be used in two ways: Either to plot the observed mean structure of a metric variable, by specifying `dat` and the variable `y_var`, or by specifying `dat` and the `model` object, to plot some mean structure represented by an estimated two-dimensional tensor product surface. The model must be estimated with [gam](#) or [bam](#).

Usage

```
plot_APHeatmap(
  dat,
  y_var = NULL,
  model = NULL,
  dimensions = c("period", "age"),
  apc_range = NULL,
  bin_heatmap = TRUE,
  bin_heatmapGrid_list = NULL,
  markLines_list = NULL,
  markLines_displayLabels = c("age", "period", "cohort"),
  y_var_logScale = FALSE,
  plot_CI = TRUE,
  legend_limits = NULL
)
```

Arguments

<code>dat</code>	Dataset with columns <code>period</code> and <code>age</code> . If <code>y_var</code> is specified, the dataset must contain the respective column. If <code>model</code> is specified, the dataset must have been used for model estimation with <code>gam</code> or <code>bam</code> .
<code>y_var</code>	Optional character name of a metric variable to be plotted.
<code>model</code>	Optional regression model estimated with <code>gam</code> or <code>bam</code> to estimate a smoothed APC surface. Only used if <code>y_var</code> is not specified.
<code>dimensions</code>	Character vector specifying the two APC dimensions that should be visualized along the x-axis and y-axis. Defaults to <code>c("period", "age")</code> .
<code>apc_range</code>	Optional list with one or multiple elements with names <code>"age"</code> , <code>"period"</code> , <code>"cohort"</code> to filter the data. Each element should contain a numeric vector of values for the respective variable that should be kept in the data. All other values are deleted.
<code>bin_heatmap</code> , <code>bin_heatmapGrid_list</code>	<code>bin_heatmap</code> indicates if the heatmap surface should be binned. Defaults to <code>TRUE</code> . If <code>TRUE</code> , the binning grid borders are defined by <code>bin_heatmapGrid_list</code> . This is a list with each element a numeric vector and a name out of <code>c("age", "period", "cohort")</code> . Can maximally have three elements. Defaults to <code>NULL</code> , where the heatmap is binned in 5 year steps along the x-axis and the y-axis.
<code>markLines_list</code>	Optional list that can be used to highlight the borders of specific age groups, time intervals or cohorts. Each element must be a numeric vector of values where horizontal, vertical or diagonal lines should be drawn (depends on which APC dimension is displayed on which axis). The list can maximally have three elements and must have names out of <code>c("age", "period", "cohort")</code> .
<code>markLines_displayLabels</code>	Optional character vector defining for which dimensions the lines defined through <code>markLines_list</code> should be marked by a respective label. The vector should be a subset of <code>c("age", "period", "cohort")</code> , or <code>NULL</code> to suppress all labels. Defaults to <code>c("age", "period", "cohort")</code> .
<code>y_var_logScale</code>	Indicator if <code>y_var</code> should be <code>log10</code> transformed. Only used if <code>y_var</code> is specified. Defaults to <code>FALSE</code> .
<code>plot_CI</code>	Indicator if the confidence intervals should be plotted. Only used if <code>y_var</code> is not specified. Defaults to <code>TRUE</code> .
<code>legend_limits</code>	Optional numeric vector passed as argument <code>limits</code> to <code>scale_fill_gradient2</code> .

Details

See also `plot_APChexamap` to plot a hexagonal heatmap with adapted axes.

If the plot is created based on the `model` object and the model was estimated with a log or logit link, the function automatically performs an exponential transformation of the effect.

Value

Plot grid created with `ggarrange` (if `plot_CI` is `TRUE`) or a `ggplot2` object (if `plot_CI` is `FALSE`).

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>, Maximilian Weigert <maximilian.weigert@stat.uni-m

References

Weigert, M., Bauer, A., Gernert, J., Karl, M., Nalmpatian, A., Küchenhoff, H., and Schmude, J. (2021). Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances. *Tourism Economics*. doi:10.1177/1354816620987198.

See Also

plot_APHeatmap

Examples

```
library(APCtools)
library(mgcv)

data(travel)

# variant A: plot observed mean structures
# observed heatmap
plot_APHeatmap(dat = travel, y_var = "mainTrip_distance",
               bin_heatmap = FALSE, y_var_logScale = TRUE)

# with binning
plot_APHeatmap(dat = travel, y_var = "mainTrip_distance",
               bin_heatmap = TRUE, y_var_logScale = TRUE)

# variant B: plot some smoothed, estimated mean structure
model <- gam(mainTrip_distance ~ te(age, period) + residence_region +
             household_size + s(household_income), data = travel)

# plot the smooth tensor product surface
plot_APHeatmap(dat = travel, model = model, bin_heatmap = FALSE, plot_CI = FALSE)

# ... same plot including the confidence intervals
plot_APHeatmap(dat = travel, model = model, bin_heatmap = FALSE)

# the APC dimensions can be flexibly assigned to the x-axis and y-axis
plot_APHeatmap(dat = travel, model = model, dimensions = c("age", "cohort"),
               bin_heatmap = FALSE, plot_CI = FALSE)

# add some reference lines
plot_APHeatmap(dat = travel, model = model, bin_heatmap = FALSE, plot_CI = FALSE,
               markLines_list = list(cohort = c(1910, 1939, 1955, 1980)))

# default binning of the tensor product surface in 5-year-blocks
plot_APHeatmap(dat = travel, model = model, plot_CI = FALSE)

# manual binning
```

```

manual_binning <- list(period = seq(min(travel$period, na.rm = TRUE) - 1,
                                   max(travel$period, na.rm = TRUE), by = 5),
                      cohort = seq(min(travel$period - travel$age, na.rm = TRUE) - 1,
                                   max(travel$period - travel$age, na.rm = TRUE), by = 10))
plot_APChexamap(dat = travel, model = model, plot_CI = FALSE,
               bin_heatmapGrid_list = manual_binning)

```

plot_APChexamap *Hexamap of an APC surface*

Description

Plot the heatmap of an APC structure using a hexagon-based plot with adapted axes. In this way, the one temporal dimension that is represented by the diagonal structure is visually not underrepresented compared to the other two dimensions on the x-axis and y-axis.

The function can be used in two ways: Either to plot the observed mean structure of a metric variable, by specifying `dat` and the variable `y_var`, or by specifying `dat` and the model object, to plot some mean structure represented by an estimated two-dimensional tensor product surface. The model must be estimated with [gam](#) or [bam](#).

Usage

```

plot_APChexamap(
  dat,
  y_var = NULL,
  model = NULL,
  apc_range = NULL,
  y_var_logScale = FALSE,
  obs_interval = 1,
  iso_interval = 5,
  color_vec = NULL,
  color_range = NULL,
  line_width = 0.5,
  line_color = gray(0.5),
  label_size = 0.5,
  label_color = "black",
  legend_title = NULL
)

```

Arguments

<code>dat</code>	Dataset with columns <code>period</code> and <code>age</code> . If <code>y_var</code> is specified, the dataset must contain the respective column. If <code>model</code> is specified, the dataset must have been used for model estimation with gam or bam .
<code>y_var</code>	Optional character name of a metric variable to be plotted.

model	Optional regression model estimated with gam or bam to estimate a smoothed APC surface. Only used if <code>y_var</code> is not specified.
apc_range	Optional list with one or multiple elements with names "age", "period", "cohort" to filter the data. Each element should contain a numeric vector of values for the respective variable that should be kept in the data. All other values are deleted.
y_var_logScale	Indicator if <code>y_var</code> should be log10 transformed. Only used if <code>y_var</code> is specified. Defaults to FALSE.
obs_interval	Numeric specifying the interval width based on which the data is spaced. Only used if <code>y_var</code> is specified. Defaults to 1, i.e. observations each year.
iso_interval	Numeric specifying the interval width between the isolines along each axis. Defaults to 5.
color_vec	Optional character vector of color names, specifying the color continuum.
color_range	Optional numeric vector with two elements, specifying the ends of the color scale in the legend.
line_width	Line width of the isolines. Defaults to 0.5.
line_color	Character color name for the isolines. Defaults to gray.
label_size	Size of the labels along the axes. Defaults to 0.5.
label_color	Character color name for the labels along the axes.
legend_title	Optional character title for the legend.

Details

See also [plot_APHeatmap](#) to plot a regular heatmap.

If the plot is created based on the `model` object and the model was estimated with a log or logit link, the function automatically performs an exponential transformation of the effect.

Value

Creates a plot with base R functions (not `ggplot2`).

Author(s)

Hawre Jalal <hjalal@pitt.edu>, Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>

References

Jalal, H., Burke, D. (2020). Hexamaps for Age–Period–Cohort Data Visualization and Implementation in R. *Epidemiology*, 31 (6), e47-e49. doi: 10.1097/EDE.0000000000001236.

See Also

[plot_APHeatmap](#)

Examples

```
library(APCtools)
library(mgcv)
library(dplyr)

data(drug_deaths)

# restrict to data where the mortality rate is available
drug_deaths <- drug_deaths %>% filter(!is.na(mortality_rate))

# hexamap of an observed structure
plot_APChexamap(dat      = drug_deaths,
                y_var    = "mortality_rate",
                color_range = c(0,40))

# hexamap of a smoothed structure
model <- gam(mortality_rate ~ te(age, period, bs = "ps", k = c(8,8)),
            data = drug_deaths)

plot_APChexamap(dat = drug_deaths, model = model)
```

plot_density

Plot the density of one metric or categorical variable

Description

Create a density plot or a boxplot of one metric variable or a barplot of one categorical variable, based on a specific subset of the data.

Usage

```
plot_density(
  dat,
  y_var,
  plot_type = "density",
  apc_range = NULL,
  highlight_diagonals = NULL,
  y_var_cat_breaks = NULL,
  y_var_cat_labels = NULL,
  weights_var = NULL,
  log_scale = FALSE,
  xlab = NULL,
  ylab = NULL,
  legend_title = NULL,
  ...
)
```

Arguments

dat	Dataset with columns period and age and the main variable specified through argument y_var.
y_var	Character name of the main variable to be plotted.
plot_type	One of c("density", "boxplot"). Only used if the y_var column is metric.
apc_range	Optional list with one or multiple elements with names "age", "period", "cohort" to filter the data. Each element should contain a numeric vector of values for the respective variable that should be kept in the data. All other values are deleted.
highlight_diagonals	Optional internal parameter which is only specified when plot_density is called from within plot_densityMatrix. See plot_densityMatrix for details.
y_var_cat_breaks	Optional numeric vector of breaks to categorize y_var based on calling function cut . Only used to highlight the categories based on different colors. And only used if the y_var column is numeric.
y_var_cat_labels	Optional character vector for the names of the categories that were defined based on y_var_cat_breaks. The length of this vector must be one shorter than length(y_var_cat_breaks). Only used if the y_var column is numeric.
weights_var	Optional character name of a weights variable used to project the results in the sample to some population.
log_scale	Indicator if the main variable should be log10 transformed. Only used if the y_var column is numeric. Defaults to FALSE.
xlab, ylab, legend_title	Optional plot annotations.
...	Additional arguments passed to density .

Details

If plot_density is called internally from within [plot_densityMatrix](#) (i.e., if the dataset contains some of the columns c("age_group", "period_group", "cohort_group")), this function will calculate the metric densities individually for these groups.

Value

ggplot object

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>, Maximilian Weigert <maximilian.weigert@stat.uni-m

Examples

```
library(APCtools)
data(travel)

plot_density(dat = travel, y_var = "mainTrip_distance")
```

```
plot_density(dat = travel, y_var = "mainTrip_distance")
```

plot_densityMatrix *Create a matrix of density plots*

Description

This function creates a matrix of individual density plots (i.e., a *ridgeline matrix*) or boxplots (for metric variables) or of individual barplots (for categorical variables). The age, period or cohort information can each either be plotted on the x-axis or the y-axis.

Usage

```
plot_densityMatrix(
  dat,
  y_var,
  dimensions = c("period", "age"),
  age_groups = NULL,
  period_groups = NULL,
  cohort_groups = NULL,
  plot_type = "density",
  highlight_diagonals = NULL,
  y_var_cat_breaks = NULL,
  y_var_cat_labels = NULL,
  weights_var = NULL,
  log_scale = FALSE,
  legend_title = NULL,
  ...
)
```

Arguments

dat	Dataset with columns period and age and the main variable specified through argument y_var.
y_var	Character name of the main variable to be plotted.
dimensions	Character vector specifying the two APC dimensions that should be visualized along the x-axis and y-axis. Defaults to c("period", "age").
age_groups, period_groups, cohort_groups	Each a list with one element specifying the borders of one row or column in the density matrix. E.g., if the period should be visualized in decade columns from 1980 to 2009, specify period_groups = list(c(1980, 1989), c(1990, 1999), c(2000, 2009)). The list can be named to specify labels for the categories. Only the two arguments should be passed that were specified by the dimensions argument.

plot_type	One of c("density", "boxplot"). Only used if the y_var column is metric.
highlight_diagonals	Optional list to define diagonals in the density that should be highlighted with different colors. Each list element should be a numeric vector stating the index of the diagonals (counted from the top left) that should be highlighted in the same color. If the list is named, the names are used as legend labels.
y_var_cat_breaks	Optional numeric vector of breaks to categorize y_var based on calling function <code>cut</code> . Only used to highlight the categories based on different colors. And only used if the y_var column is numeric.
y_var_cat_labels	Optional character vector for the names of the categories that were defined based on y_var_cat_breaks. The length of this vector must be one shorter than length(y_var_cat_breaks). Only used if the y_var column is numeric.
weights_var	Optional character name of a weights variable used to project the results in the sample to some population.
log_scale	Indicator if the main variable should be log10 transformed. Only used if the y_var column is numeric. Defaults to FALSE.
legend_title	Optional plot annotations.
...	Additional arguments passed to <code>plot_density</code> .

Value

ggplot object

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>, Maximilian Weigert <maximilian.weigert@stat.uni-m

References

Weigert, M., Bauer, A., Gernert, J., Karl, M., Nalmpatian, A., Küchenhoff, H., and Schmude, J. (2021). Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances. *Tourism Economics*. doi:10.1177/1354816620987198.

Examples

```
library(APCtools)

# define categorizations for the main trip distance
dist_cat_breaks <- c(1,500,1000,2000,6000,100000)
dist_cat_labels <- c("< 500 km", "500 - 1,000 km", "1,000 - 2,000 km",
                    "2,000 - 6,000 km", "> 6,000 km")

age_groups <- list(c(80,89),c(70,79),c(60,69),c(50,59),c(40,49),c(30,39),c(20,29))
period_groups <- list(c(1971,1979),c(1980,1989),c(1990,1999),c(2000,2009),c(2010,2018))
cohort_groups <- list(c(1980,1989),c(1970,1979),c(1960,1969),c(1950,1959),c(1940,1949),
```

```

                                c(1930,1939),c(1920,1929))

plot_densityMatrix(dat          = travel,
                   y_var       = "mainTrip_distance",
                   age_groups   = age_groups,
                   period_groups = period_groups,
                   log_scale    = TRUE)

# highlight two cohorts
plot_densityMatrix(dat          = travel,
                   y_var       = "mainTrip_distance",
                   age_groups   = age_groups,
                   period_groups = period_groups,
                   highlight_diagonals = list("born 1950 - 1959" = 8,
                                             "born 1970 - 1979" = 10),
                   log_scale    = TRUE)

# also mark different distance categories
plot_densityMatrix(dat          = travel,
                   y_var       = "mainTrip_distance",
                   age_groups   = age_groups,
                   period_groups = period_groups,
                   log_scale    = TRUE,
                   y_var_cat_breaks = dist_cat_breaks,
                   y_var_cat_labels = dist_cat_labels,
                   highlight_diagonals = list("born 1950 - 1959" = 8,
                                             "born 1970 - 1979" = 10),
                   legend_title   = "Distance category")

# flexibly assign the APC dimensions to the x-axis and y-axis
plot_densityMatrix(dat          = travel,
                   y_var       = "mainTrip_distance",
                   dimensions   = c("period","cohort"),
                   period_groups = period_groups,
                   cohort_groups = cohort_groups,
                   log_scale    = TRUE,
                   y_var_cat_breaks = dist_cat_breaks,
                   y_var_cat_labels = dist_cat_labels,
                   legend_title   = "Distance category")

# use boxplots instead of densities
plot_densityMatrix(dat          = travel,
                   y_var       = "mainTrip_distance",
                   plot_type    = "boxplot",
                   age_groups   = age_groups,
                   period_groups = period_groups,
                   log_scale    = TRUE,
                   highlight_diagonals = list("born 1950 - 1959" = 8,
                                             "born 1970 - 1979" = 10))

# plot categorical variables instead of metric ones
plot_densityMatrix(dat          = travel,

```

```
y_var           = "household_size",
age_groups      = age_groups,
period_groups   = period_groups,
highlight_diagonals = list("born 1950 - 1959" = 8,
                           "born 1970 - 1979" = 10))
```

plot_density_categorical

Internal helper to plot a categorical density

Description

Internal helper function to plot one categorical density, to be called from within [plot_density](#).

Usage

```
plot_density_categorical(
  dat,
  y_var,
  dat_highlightDiagonals = NULL,
  weights_var = NULL,
  xlab = NULL,
  ylab = NULL
)
```

Arguments

dat	Dataset with columns period and age and the main variable specified through argument y_var.
y_var	Character name of the main variable to be plotted.
dat_highlightDiagonals	Optional dataset created by create_highlightDiagonalData to highlight specific diagonals in a density matrix.
weights_var	Optional character name of a weights variable used to project the results in the sample to some population.
xlab	Optional plot annotations.
ylab	Optional plot annotations.

plot_density_metric *Internal helper to plot a metric density*

Description

Internal helper function to plot one metric density, to be called from within [plot_density](#).

Usage

```
plot_density_metric(
  dat,
  y_var,
  plot_type = "density",
  dat_highlightDiagonals = NULL,
  y_var_cat_breaks = NULL,
  y_var_cat_labels = NULL,
  weights_var = NULL,
  log_scale = FALSE,
  xlab = NULL,
  ylab = NULL,
  legend_title = NULL,
  ...
)
```

Arguments

dat	Dataset with columns period and age and the main variable specified through argument y_var.
y_var	Character name of the main variable to be plotted.
plot_type	One of c("density", "boxplot"). Only used if the y_var column is metric.
dat_highlightDiagonals	Optional dataset created by create_highlightDiagonalData to highlight specific diagonals in a density matrix.
y_var_cat_breaks	Optional numeric vector of breaks to categorize y_var based on calling function cut . Only used to highlight the categories based on different colors. And only used if the y_var column is numeric.
y_var_cat_labels	Optional character vector for the names of the categories that were defined based on y_var_cat_breaks. The length of this vector must be one shorter than length(y_var_cat_breaks). Only used if the y_var column is numeric.
weights_var	Optional character name of a weights variable used to project the results in the sample to some population.
log_scale	Indicator if the main variable should be log10 transformed. Only used if the y_var column is numeric. Defaults to FALSE.

xlab	Optional plot annotations.
ylab	Optional plot annotations.
legend_title	Optional plot annotations.
...	Additional arguments passed to <code>density</code> .

plot_jointMarginalAPCEffects

Joint plot to compare the marginal APC effects of multiple models

Description

This function creates a joint plot of the marginal APC effects of multiple estimated models. It creates a plot with one pane per age, period and cohort effect, each containing one lines for each estimated model.

Usage

```
plot_jointMarginalAPCEffects(
  model_list,
  dat,
  vlines_list = NULL,
  ylab = NULL,
  ylim = NULL
)
```

Arguments

model_list	A list of regression models estimated with <code>gam</code> or <code>bam</code> . If the list is named, the names are used as labels.
dat	Dataset with columns <code>period</code> and <code>age</code> . If <code>y_var</code> is specified, the dataset must contain the respective column. If <code>model</code> is specified, the dataset must have been used for model estimation with <code>gam</code> or <code>bam</code> .
vlines_list	Optional list that can be used to highlight the borders of specific age groups, time intervals or cohorts. Each element must be a numeric vector of values on the x-axis where vertical lines should be drawn. The list can maximally have three elements and must have names out of <code>c("age", "period", "cohort")</code> .
ylab, ylim	Optional ggplot2 styling arguments.

Details

If the model was estimated with a log or logit link, the function automatically performs an exponential transformation of the effect.

Value

Plot grid created with `ggarrange`.

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>, Maximilian Weigert <maximilian.weigert@stat.uni-m

Examples

```
library(APCtools)
library(mgcv)

data(travel)
model_pure <- gam(mainTrip_distance ~ te(age, period), data = travel)
model_cov <- gam(mainTrip_distance ~ te(age, period) + s(household_income),
                 data = travel)

model_list <- list("pure model" = model_pure,
                  "covariate model" = model_cov)

plot_jointMarginalAPCEffects(model_list, dat = travel)

# mark specific cohorts
plot_jointMarginalAPCEffects(model_list, dat = travel,
                             vlines_list = list("cohort" = c(1966.5,1982.5,1994.5)))
```

plot_linearEffects *Plot linear effects of a gam in an effect plot*

Description

Create an effect plot of linear effects of a model fitted with [gam](#) or [bam](#).

Usage

```
plot_linearEffects(model, variables = NULL, return_plotData = FALSE)
```

Arguments

model	Model fitted with gam or bam .
variables	Optional character vector of variable names specifying which effects should be plotted. The order of the vector corresponds to the order in the effect plot. If the argument is not specified, all linear effects are plotted according to the order of their appearance in the model output.
return_plotData	If TRUE, the dataset prepared for plotting is returned. Defaults to FALSE.

Details

If the model was estimated with a log or logit link, the function automatically performs an exponential transformation of the effect.

Value

ggplot object

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>

Examples

```
library(APCtools)
library(mgcv)

data(travel)
model <- gam(mainTrip_distance ~ te(age, period) + residence_region +
             household_size + s(household_income), data = travel)

plot_linearEffects(model)
```

plot_marginalAPCEffects

Plot of marginal APC effects based on an estimated GAM model

Description

Plot the marginal effect of age, period or cohort, based on an APC model estimated as a semiparametric additive regression model with [gam](#) or [bam](#). This function is a simple wrapper to [plot_partialAPCEffects](#), called with argument `hide_partialEffects = TRUE`.

Usage

```
plot_marginalAPCEffects(
  model,
  dat,
  variable = "age",
  vlines_vec = NULL,
  return_plotData = FALSE
)
```

Arguments

model	Optional regression model estimated with gam or bam to estimate a smoothed APC surface. Only used if <code>y_var</code> is not specified.
dat	Dataset with columns <code>period</code> and <code>age</code> . If <code>y_var</code> is specified, the dataset must contain the respective column. If <code>model</code> is specified, the dataset must have been used for model estimation with gam or bam .

variable	One of c("age", "period", "cohort"), specifying the temporal dimension for which the partial effect plots should be created.
vlines_vec	Optional numeric vector of values on the x-axis where vertical lines should be drawn. Can be used to highlight the borders of specific age groups, time intervals or cohorts.
return_plotData	If TRUE, a list of the datasets prepared for plotting is returned instead of the ggplot object. The list contains one dataset each for the overall effect (= evaluations of the APC surface to plot the partial effects) and for each marginal APC effect (no matter the specified value of the argument variable). Defaults to FALSE.

Value

ggplot object

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>, Maximilian Weigert <maximilian.weigert@stat.uni-m

References

Weigert, M., Bauer, A., Gernert, J., Karl, M., Nalmpatian, A., Küchenhoff, H., and Schmude, J. (2021). Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances. *Tourism Economics*. doi:10.1177/1354816620987198.

Examples

```
library(APCtools)
library(mgcv)

data(travel)
model <- gam(mainTrip_distance ~ te(age, period), data = travel)

plot_marginalAPCEffects(model, dat = travel, variable = "age")

# mark specific cohorts
plot_marginalAPCEffects(model, dat = travel, variable = "cohort",
                        vlines_vec = c(1966.5, 1982.5, 1994.5))
```

plot_partialAPCEffects

Partial APC plots based on an estimated GAM model

Description

Create the partial APC plots based on an APC model estimated as a semiparametric additive regression model with `gam` or `bam`.

Usage

```
plot_partialAPCeffects(
  model,
  dat,
  variable = "age",
  hide_partialEffects = FALSE,
  vlines_vec = NULL,
  return_plotData = FALSE
)
```

Arguments

<code>model</code>	Optional regression model estimated with <code>gam</code> or <code>bam</code> to estimate a smoothed APC surface. Only used if <code>y_var</code> is not specified.
<code>dat</code>	Dataset with columns <code>period</code> and <code>age</code> . If <code>y_var</code> is specified, the dataset must contain the respective column. If <code>model</code> is specified, the dataset must have been used for model estimation with <code>gam</code> or <code>bam</code> .
<code>variable</code>	One of <code>c("age", "period", "cohort")</code> , specifying the temporal dimension for which the partial effect plots should be created.
<code>hide_partialEffects</code>	If TRUE, only the marginal effect will be plotted. Defaults to FALSE.
<code>vlines_vec</code>	Optional numeric vector of values on the x-axis where vertical lines should be drawn. Can be used to highlight the borders of specific age groups, time intervals or cohorts.
<code>return_plotData</code>	If TRUE, a list of the datasets prepared for plotting is returned instead of the <code>ggplot</code> object. The list contains one dataset each for the overall effect (= evaluations of the APC surface to plot the partial effects) and for each marginal APC effect (no matter the specified value of the argument <code>variable</code>). Defaults to FALSE.

Details

If the model was estimated with a log or logit link, the function automatically performs an exponential transformation of the effect.

Value

`ggplot` object (if `hide_partialEffects` is TRUE) or a plot grid created with `ggarrange` (if FALSE).

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>, Maximilian Weigert <maximilian.weigert@stat.uni-m

References

Weigert, M., Bauer, A., Gernert, J., Karl, M., Nalmpatian, A., Küchenhoff, H., and Schmude, J. (2021). Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances. *Tourism Economics*. doi:10.1177/1354816620987198.

Examples

```
library(APCtools)
library(mgcv)

data(travel)
model <- gam(mainTrip_distance ~ te(age, period), data = travel)

plot_partialAPCEffects(model, dat = travel, variable = "age")

# mark specific cohorts
plot_partialAPCEffects(model, dat = travel, variable = "cohort",
                        vlines_vec = c(1966.5, 1982.5, 1994.5))
```

plot_variable

Distribution plot of one variable against one APC dimension

Description

Plot the distribution of one variable in the data against age, period or cohort. Creates a bar plot for categorical variables (see argument `geomBar_position`) and boxplots or a line plot of median values for metric variables (see `plot_type`).

Usage

```
plot_variable(
  dat,
  y_var,
  apc_dimension = "period",
  log_scale = FALSE,
  plot_type = "boxplot",
  geomBar_position = "fill",
  legend_title = NULL,
  ylab = NULL,
  ylim = NULL
)
```

Arguments

dat	Dataset containing columns age and period.
y_var	Character name of the variable to plot.
apc_dimension	One of c("age", "period", "cohort"). Defaults to "period".
log_scale	Indicator if the visualized variable should be log10 transformed. Only used if the variable is numeric. Defaults to FALSE.
plot_type	One of c("boxplot", "line", "line-points"), specifying if boxplots or a line plot of median values should be drawn for metric variables. "line-points" adds points to the line plot where observations are available.
geomBar_position	Value passed to <code>geom_bar</code> as position argument. Only used if the visualized variable is categorical. Defaults to "fill".
legend_title	Optional character title for the legend which is drawn for categorical variables.
ylab, ylim	Optional arguments for styling the ggplot.

Value

ggplot object

Author(s)

Alexander Bauer <alexander.bauer@stat.uni-muenchen.de>

Examples

```
library(APCtools)
data(travel)

# plot a metric variable
plot_variable(dat = travel, y_var = "mainTrip_distance",
              apc_dimension = "period", log_scale = TRUE)
plot_variable(dat = travel, y_var = "mainTrip_distance",
              apc_dimension = "period", log_scale = TRUE, plot_type = "line")

# plot a categorical variable
plot_variable(dat = travel, y_var = "household_size", apc_dimension = "period")
plot_variable(dat = travel, y_var = "household_size", apc_dimension = "period",
              geomBar_position = "stack")
```

travel

Data from the German Reiseanalyse survey

Description

This dataset from the [Reiseanalyse survey](#) comprises travel information on German travelers between 1971 and 2018. Data were collected in a yearly repeated cross-sectional survey of German pleasure travels, based on a sample representative for the (West) German citizens (until 2009) or for all German-speaking residents (starting 2010). Travelers from former East Germany are only included since 1990. Note that the sample only contains trips with at least five days of trip length. For details see [Weigert et al. \(2021\)](#).

Usage

```
data(travel)
```

Format

A dataframe containing

period Year in which the respondent traveled.

age Age of the respondent.

sampling_weight Individual weight of each respondent to account for a not perfectly representative sample and project the sample results to the population of German citizens (until 2009) or of German-speaking residents (starting 2010). Only available since 1974.

german_citizenship Indicator if the respondent is German citizen or not. Only available since 2010. Until 2009, all respondents were German citizens.

residence_region Indicator if the respondent's main residence is in a federal state in the former area of West Germany or in the former area of East Germany.

household_size Categorized size of the respondent's household.

household_income Joint income (in €) of the respondent's household.

mainTrip_duration Categorized trip length of the respondent's *main trip*. The main trip is the trip which the respondent stated was his/her most important trip in the respective year.

mainTrip_distance Distance (in km) between the center of the respondent's federal state and the center of the country of destination, for the *main trip*. The main trip is the trip which the respondent stated was his/her most important trip in the respective year.

Details

The data are a 10% random sample of all respondents who undertook at least one trip in the respective year, between 1971 and 2018. We thank the [Forschungsgemeinschaft Urlaub und Reisen e.V.](#) for allowing us to publish this sample.

References

Weigert, M., Bauer, A., Gernert, J., Karl, M., Nalmpatian, A., Küchenhoff, H., and Schmude, J. (2021). Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances. *Tourism Economics*. doi: [10.1177/1354816620987198](https://doi.org/10.1177/1354816620987198).

Forschungsgemeinschaft Urlaub und Reisen e.V. (FUR) (2020b) *Survey of tourist demand in Germany for holiday travel and short breaks*. Available at: https://reiseanalyse.de/wp-content/uploads/2019/08/RA2020_Infoflyer.pdf (accessed 22 November 2021).

Index

- * **datasets**
 - drug_deaths, 8
 - travel, 32
- bam, 4, 5, 7, 8, 10, 12–14, 16, 17, 25–27, 29
- calc_density, 3
- capitalize_firstLetter, 3
- compute_xCoordinate, 4
- compute_yCoordinate, 4
- create_APCsummary, 4, 8
- create_groupVariable, 6
- create_highlightDiagonalData, 6, 12, 23, 24
- create_modelSummary, 7
- create_oneAPCsummaryTable, 8
- cut, 19, 21, 24
- density, 3, 19, 25
- drug_deaths, 8
- ensure_segmentsInPlotRange, 9
- extract_summary_linearEffects, 10
- gam, 4, 5, 7, 8, 10, 12–14, 16, 17, 25–27, 29
- geom_bar, 31
- get_plotGAMobject, 10
- gg_addReferenceLines, 9, 11
- gg_highlightDiagonals, 11
- ggarrange, 14, 25, 29
- kable, 4, 5, 7
- plot.gam, 10
- plot_1Dsmooth, 12
- plot_APCheatmap, 11, 13, 17
- plot_APChexamap, 4, 14, 16
- plot_density, 3, 18, 21, 23, 24
- plot_density_categorical, 23
- plot_density_metric, 24
- plot_densityMatrix, 3, 6, 19, 20
- plot_jointMarginalAPCeffects, 25
- plot_linearEffects, 26
- plot_marginalAPCeffects, 27
- plot_partialAPCeffects, 27, 28
- plot_variable, 30
- scale_fill_gradient2, 14
- travel, 32