

Package ‘eikosograms’

January 11, 2026

Type Package

Title Visualizing Probabilities, Frequencies, and Conditional Independence for Categorical Variates

Version 1.0.0

URL <https://github.com/rwoldford/eikosograms>,
<https://rwoldford.github.io/eikosograms/>

BugReports <https://github.com/rwoldford/eikosograms/issues>

Description An eikosogram (ancient Greek for probability picture) divides the unit square into rectangular regions whose areas, sides, and widths represent various probabilities associated with the values of one or more categorical variates. Rectangle areas are joint probabilities, widths are always marginal (though possibly joint margins, i.e. marginal joint distributions of two or more variates), and heights of rectangles are always conditional probabilities. Eikosograms embed the rules of probability and are useful for introducing elementary probability theory, including axioms, marginal, conditional, and joint probabilities, and their relationships (including Bayes' theorem as a completely trivial consequence). They provide advantages over Venn diagrams for this purpose, particularly in distinguishing probabilistic independence, mutually exclusive events, coincident events, and associations. They also are useful for identifying and understanding conditional independence structure. Eikosograms can be thought of as mosaic plots when only two categorical variates are involved; the layout is quite different when there are more than two variates. Only one categorical variate, designated the “response”, presents on the vertical axis and all others, designated the “conditioning” variates, appear on the horizontal. In this way, conditional probability appears only as height and marginal probabilities as widths. The eikosogram is ideal for response models (e.g. logistic models) but equally useful when no variate is distinguished as the response. In such cases, each variate can appear in turn as the response, which is handy for assessing conditional independence in discrete graphical models (i.e. “Bayesian networks” or “BayesNets”). The eikosogram and its value over Venn diagrams in teaching probability is described in W.H. Cherry and R.W. Oldford (2003) <<https://math.uwaterloo.ca/~rwoldfor/papers/eikosograms/paper.pdf>>, its value in exploring conditional independence structure and relation to graphical and log-linear models is described in R.W. Oldford (2003) <<https://math.uwaterloo.ca/~rwoldfor/papers/eikosograms/independence/paper.pdf>>, and a number of problems, puzzles, and paradoxes that are easily explained with eikosograms are given in R.W. Oldford (2003) <<https://math.uwaterloo.ca/~rwoldfor/papers/eikosograms/examples/paper.pdf>>.

License GPL-3

Depends R (>= 3.5.0)

Imports grid, stats

Suggests knitr, rmarkdown, gridExtra

VignetteBuilder knitr, rmarkdown

Encoding UTF-8

LazyLoad yes

LazyData true

RoxygenNote 7.3.3

NeedsCompilation no

Author Wayne Oldford [aut, cre],
 Erle Holgersen [aut],
 Ben Lafreniere [aut],
 Tianlu Zhu [aut]

Maintainer Wayne Oldford <rwoldford@uwaterloo.ca>

Repository CRAN

Date/Publication 2026-01-11 21:00:02 UTC

Contents

eikos	2
eikos.default	5
eikos.formula	7
eikos_data	10
eikos_legend	10
eikos_x_labels	11
eikos_x_probs	11
eikos_y_labels	12
eikos_y_probs	13
medicalRecords	13
tuberculosis_1910	15

Index	17
--------------	-----------

eikos

Generic method for creating an eikosogram

Description

Generic method for creating an eikosogram

Usage

```
eikos(
  y,
  x = NULL,
  data = NULL,
  marginalize = NULL,
  main = "",
  main_size = 16,
  ylabs = TRUE,
  ylab_rot = 0,
  yname_size = 12,
  yvals_size = 12,
  yaxs = TRUE,
  yprobs = NULL,
  yprobs_size = 8,
  xlabs = TRUE,
  xlab_rot = 0,
  xname_size = 12,
  xvals_size = 12,
  xaxs = TRUE,
  xprobs = NULL,
  xprobs_size = 8,
  vertical_xprobs = TRUE,
  ispace = list(bottom = 8, left = 2, top = 2, right = 5),
  legend = FALSE,
  col = NULL,
  bottomcol = "steelblue",
  topcol = "snow2",
  lcol = "black",
  draw = TRUE,
  newpage = TRUE,
  lock_aspect = TRUE
)
```

Arguments

<code>y</code>	Either the name of a variable in the data set (<code>eikos.default</code>), or a formula of such variables (<code>eikos.formula</code>).
<code>x</code>	name(s) of any conditional variable(s) (horizontal axis). Should be null if formula given.
<code>data</code>	data frame or table
<code>marginalize</code>	variable(s) to marginalize on, or NULL if none. Marginalized variables still appear in plot.
<code>main</code>	title of plot
<code>main_size</code>	font size of title (in points)
<code>ylabs</code>	logical, whether y labels should appear or not.

<code>ylab_rot</code>	rotation of y labels
<code>yname_size</code>	font size of vertical axis names (in points)
<code>yvals_size</code>	font size of labels for values of y variable (in points)
<code>yaxs</code>	logical, whether y axis should appear or not.
<code>yprobs</code>	probabilities to be shown on y-axis. NULL if they should be calculated from the data.
<code>yprobs_size</code>	font size of labels for horizontal probabilities (in points)
<code>xlabs</code>	logical, whether x labels should appear or not.
<code>xlab_rot</code>	rotation of x labels
<code>xname_size</code>	font size of horizontal axis names (in points)
<code>xvals_size</code>	font size of labels for values of x variables (in points)
<code>xaxs</code>	logical, whether x axis should appear or not.
<code>xprobs</code>	probabilities to be shown on x-axis. NULL if they should be calculated from the data.
<code>xprobs_size</code>	font size of labels for horizontal probabilities (in points)
<code>vertical_xprobs</code>	logical, whether probabilities on x axis should be rotated vertically.
<code>ispace</code>	list of four items (bottom, left, top, right) indicating the margins separating the text around the diagram. Each value is a positive integer giving a measure in "points".
<code>legend</code>	logical, whether to include legend
<code>col</code>	a vector of colours to match the response values. If NULL (the default), the colours are constructed as a smooth transition from 'bottomcol' to 'topcol' via 'grDevices::colorRampPalette
<code>bottomcol</code>	bottom colour
<code>topcol</code>	top colour
<code>lcol</code>	colour of lines
<code>draw</code>	logical, whether to draw eikosogram.
<code>newpage</code>	logical, whether to draw on a newpage.
<code>lock_aspect</code>	logical, whether to force entire plot to 1:1 aspect ratio.

See Also[eikos.default](#)[eikos.formula](#)**Examples**

```
eikos("Hair", "Eye", data=HairEyeColor, legend = TRUE)
eikos(gear ~ cyl, data = mtcars)
eikos(Admit ~ Gender + Dept, data = UCBA admissions,
      yaxs = FALSE, xaxs = FALSE,
      lock_aspect = FALSE,
      xlab_rot = 90, xvals_size = 8,
      ispace = list(bottom = 15))
```

eikos.default	<i>Create a new eikosogram</i>
---------------	--------------------------------

Description

Return a grid graphic object (grob) and draw an eikosogram if draw = TRUE.

Usage

```
## Default S3 method:
eikos(
  y,
  x = NULL,
  data = NULL,
  marginalize = NULL,
  main = "",
  main_size = 16,
  ylabs = TRUE,
  ylab_rot = 0,
  yname_size = 12,
  yvals_size = 12,
  yaxs = TRUE,
  yprobs = NULL,
  yprobs_size = 8,
  xlabs = TRUE,
  xlab_rot = 0,
  xname_size = 12,
  xvals_size = 12,
  xaxs = TRUE,
  xprobs = NULL,
  xprobs_size = 8,
  vertical_xprobs = TRUE,
  ispace = list(bottom = 8, left = 2, top = 2, right = 5),
  legend = FALSE,
  col = NULL,
  bottomcol = "steelblue",
  topcol = "snow2",
  lcol = "black",
  draw = TRUE,
  newpage = TRUE,
  lock_aspect = TRUE
)
```

Arguments

y Either the name of a variable in the data set (eikos.default), or a formula of such variables (eikos.formula).

x	name(s) of any conditional variable(s) (horizontal axis). Should be null if formula given.
data	data frame or table
marginalize	variable(s) to marginalize on, or NULL if none. Marginalized variables still appear in plot.
main	title of plot
main_size	font size of title (in points)
ylabs	logical, whether y labels should appear or not.
ylab_rot	rotation of y labels
yname_size	font size of vertical axis names (in points)
yvals_size	font size of labels for values of y variable (in points)
yaxs	logical, whether y axis should appear or not.
yprobs	probabilities to be shown on y-axis. NULL if they should be calculated from the data.
yprobs_size	font size of labels for horizontal probabilities (in points)
xlabs	logical, whether x labels should appear or not.
xlab_rot	rotation of x labels
xname_size	font size of horizontal axis names (in points)
xvals_size	font size of labels for values of x variables (in points)
xaxs	logical, whether x axis should appear or not.
xprobs	probabilities to be shown on x-axis. NULL if they should be calculated from the data.
xprobs_size	font size of labels for horizontal probabilities (in points)
vertical_xprobs	logical, whether probabilities on x axis should be rotated vertically.
ispace	list of four items (bottom, left, top, right) indicating the margins separating the text around the diagram. Each value is a positive integer giving a measure in "points".
legend	logical, whether to include legend
col	a vector of colours to match the response values. If NULL (the default), the colours are constructed as a smooth transition from 'bottomcol' to 'topcol' via 'grDevices::colorRampPalette'
bottomcol	bottom colour
topcol	top colour
lcol	colour of lines
draw	logical, whether to draw eikosogram.
newpage	logical, whether to draw on a newpage.
lock_aspect	logical, whether to force entire plot to 1:1 aspect ratio.

Examples

```
eikos("Hair", "Eye", data=HairEyeColor, legend = TRUE)
eikos("Hair", "Eye", data=HairEyeColor,
      legend = TRUE, ylabs = FALSE,
      yname_size = 16, yvals_size = 8)
eikos("Hair", "Eye", data=HairEyeColor,
      legend = TRUE, ylabs = FALSE,
      yprobs = seq(0.2, 1, .2))
eikos("Eye", "Hair", data=HairEyeColor, yprobs = seq(0,1,0.25),
      yname_size = 20, xname_size = 20,
      col = c("sienna4", "steelblue", "darkkhaki", "springgreen3"),
      lcol = "grey10",
      lock_aspect = FALSE)
```

eikos.formula	<i>Draw eikosogram using a formula to identify response and conditioning variates</i>
---------------	---

Description

Draw eikosogram using a formula to identify response and conditioning variates

Usage

```
## S3 method for class 'formula'
eikos(
  y,
  x = NULL,
  data = NULL,
  marginalize = NULL,
  main = "",
  main_size = 16,
  ylabs = TRUE,
  ylab_rot = 0,
  yname_size = 12,
  yvals_size = 12,
  yaxis = TRUE,
  yprobs = NULL,
  yprobs_size = 8,
  xaxis = TRUE,
  xlab_rot = 0,
  xname_size = 12,
  xvals_size = 12,
  xaxis = TRUE,
  xprobs = NULL,
  xprobs_size = 8,
```

```

vertical_xprobs = TRUE,
ispace = list(bottom = 8, left = 2, top = 2, right = 5),
legend = FALSE,
col = NULL,
bottomcol = "steelblue",
topcol = "snow2",
lcol = "black",
draw = TRUE,
newpage = TRUE,
lock_aspect = TRUE
)

```

Arguments

y	Either the name of a variable in the data set (eikos.default), or a formula of such variables (eikos.formula).
x	name(s) of any conditional variable(s) (horizontal axis). Should be null if formula given.
data	data frame or table
marginalize	variable(s) to marginalize on, or NULL if none. Marginalized variables still appear in plot.
main	title of plot
main_size	font size of title (in points)
ylabs	logical, whether y labels should appear or not.
ylab_rot	rotation of y labels
yname_size	font size of vertical axis names (in points)
yvals_size	font size of labels for values of y variable (in points)
yaxs	logical, whether y axis should appear or not.
yprobs	probabilities to be shown on y-axis. NULL if they should be calculated from the data.
yprobs_size	font size of labels for horizontal probabilities (in points)
xlabs	logical, whether x labels should appear or not.
xlab_rot	rotation of x labels
xname_size	font size of horizontal axis names (in points)
xvals_size	font size of labels for values of x variables (in points)
xaxs	logical, whether x axis should appear or not.
xprobs	probabilities to be shown on x-axis. NULL if they should be calculated from the data.
xprobs_size	font size of labels for horizontal probabilities (in points)
vertical_xprobs	logical, whether probabilities on x axis should be rotated vertically.

ispace	list of four items (bottom, left, top, right) indicating the margins separating the text around the diagram. Each value is a positive integer giving a measure in "points".
legend	logical, whether to include legend
col	a vector of colours to match the response values. If NULL (the default), the colours are constructed as a smooth transition from 'bottomcol' to 'topcol' via 'grDevices::colorRampPalette'
bottomcol	bottom colour
topcol	top colour
lcol	colour of lines
draw	logical, whether to draw eikosogram.
newpage	logical, whether to draw on a newpage.
lock_aspect	logical, whether to force entire plot to 1:1 aspect ratio.

Examples

```
eikos(Eye ~ Hair + Sex, data=HairEyeColor)
eikos(Hair ~ ., data=HairEyeColor,
      yaxs = FALSE, ylabs = FALSE,
      legend = TRUE,
      col = c("black", "sienna4",
              "orangered", "lightgoldenrod" ))
eikos(Hair ~ ., data=HairEyeColor, xlab_rot = 30,
      yprobs = seq(0.1, 1, 0.1),
      yvals_size = 10,
      xvals_size = 8,
      ispace = list(bottom = 10),
      bottomcol = "grey30", topcol = "grey70",
      lcol = "white")
eikos(Hair ~ ., data=HairEyeColor, xlab_rot = 30,
      marginalize = "Eye",
      yvals_size = 10,
      xvals_size = 8,
      ispace = list(bottom = 10),
      bottomcol = "grey30", topcol = "grey70",
      lcol = "white")
eikos(Hair ~ ., data=HairEyeColor, xlab_rot = 30,
      marginalize = c("Eye", "Sex"),
      yvals_size = 10,
      xvals_size = 8,
      ispace = list(bottom = 10),
      bottomcol = "grey30", topcol = "grey70",
      lcol = "white")
```

eikos_data	<i>Create eikosogram data frame</i>
------------	-------------------------------------

Description

Eikos helper function used to convert data.

Usage

```
eikos_data(y, x, data, marginalize = NULL)
```

Arguments

y	response variable.
x	conditional variables.
data	data frame or table to be converted.
marginalize	name of variable to marginalize on, NULL if none.

eikos_legend	<i>Create eikosogram legend</i>
--------------	---------------------------------

Description

Eikos helper function used to create legend.

Usage

```
eikos_legend(
  labels,
  title = NULL,
  yname_size = 12,
  yvals_size = 12,
  col,
  margin = unit(2, "points"),
  lcol = "black"
)
```

Arguments

labels	labels to be included in legend
title	if non-NULL a string to give as the legend title
yname_size	font size of vertical axis names (in points)
yvals_size	font size of labels for values of y variable (in points)
col	colours od
margin	unit specifying margin between legend entries
lcol	line colour

eikos_x_labels	<i>eikos helper function. Returns grob with x axis labels.</i>
----------------	--

Description

eikos helper function. Returns grob with x axis labels.

Usage

```
eikos_x_labels(
  x,
  data,
  margin = unit(10, "points"),
  xname_size = 12,
  xvals_size = 10,
  lab_rot = 0
)
```

Arguments

x	vector of conditional variables
data	data frame from eikos_data.
margin	unit specifying margin
xname_size	font size for x axis variable names (in points)
xvals_size	font size of labels for values of x variables (in points)
lab_rot	integer indicating the rotation of the label, default is horizontal

Value

gList with x labels and x-axis names as grob frames.

eikos_x_probs	<i>Create grob with eikosogram x-axis probabilities</i>
---------------	---

Description

Creates x axis grob to be placed on eikosogram. Called by eikos functions.

Usage

```
eikos_x_probs(
  data,
  xprobs = NULL,
  xprobs_size = 8,
  margin = unit(2, "points"),
  rotate = TRUE
)
```

Arguments

data	data frame from eikos_data object
xprobs	vector of probabilities to be shown. NULL if they should be calculated from the data.
xprobs_size	font size of labels for horizontal probabilities (in points)
margin	unit specifying margin between y axis and eikosogram
rotate	logical, whether probabilities should be rotated vertically.

Value

textGrob with x-axis probabilities.

eikos_y_labels	<i>eikos helper function. Returns grob with y axis labels.</i>
----------------	--

Description

eikos helper function. Returns grob with y axis labels.

Usage

```
eikos_y_labels(
  y,
  data,
  margin = unit(2, "points"),
  yname_size = 12,
  yvals_size = 10,
  lab_rot = 0
)
```

Arguments

y	response variable
data	data frame from eikos_data.
margin	unit specifying margin
yname_size	font size for y axis variable names (in points)
yvals_size	font size of labels for values of y variable (in points)
lab_rot	integer indicating the rotation of the label, default is horizontal

Value

grobFrame with response variable labels and axis text

eikos_y_probs	<i>Create grob with eikosogram y-axis probabilities</i>
---------------	---

Description

Creates y axis grob to be placed on eikosogram. Called by eikos functions.

Usage

```
eikos_y_probs(data, yprobs, yprobs_size = 8, margin = unit(2, "points"))
```

Arguments

data	data frame from eikos_data object
yprobs	vector of probabilities to be shown. NULL if they should be calculated from the data.
yprobs_size	font size of labels for horizontal probabilities (in points)
margin	unit specifying margin between y axis and eikosogram

Value

textGrob with y-axis probabilities.

medicalRecords	<i>Mining medical records (fictional)</i>
----------------	---

Description

An entirely artificially constructed data set and context designed for classroom discussion and analysis.

A medical data mining context is given in detail below. In light of the context, interesting scientific questions will arise as to the data collection, and how the results should, or should not, be interpreted. It should also raise questions on what might be done in any follow up studies.

Instructors might choose to invent their own context.

Format

A data frame with 16 rows and 5 variables (providing the counts for a 2x2x2x2 contingency table).

Age A two level factor recording one of two age groups: "20-39" or "40-59".

Sex A two level factor recording sex: "Male" or "Female".

Treatment A two level factor recording the treatment received: "A" or "B".

Outcome A two level factor recording patient outcome after treatment: "Recovered" or "Died".

Freq The frequency count of patients having that combination of factors.

Details

One fictional context (constructed in March 2020) for this data set is given below (in the PPDAC style of Mackay and Oldford (2000)).

Problem:

A disease epidemic has broken out in the population of some country. It is thought that adults under the age of 60 appear to be particularly vulnerable. Both men and women contract the disease and need to be treated. Those who go untreated die within 5 days of contracting the disease.

The medical community has tried two quite different approaches to treat patients having the disease – call these ‘Treatment A’ and ‘Treatment B’. For the health of the country, it is important to determine which of these two treatments is more effective.

Plan:

To investigate which is the better treatment, it was decided to mine the medical records from another country of those who had contracted the disease and had been treated with one of the two treatments. Patients treated with either A or B survive the disease and recover fully; some however still die.

Electronic medical records available from several of the more populous districts are accessible. These can be searched to provide records from patients that have received treatment. It was decided that there should be the same number of records drawn for each treatment.

Moreover, concern was raised that the investigation have gender balance (i.e. equal numbers of males and females). So, to make sure that both sexes were equally represented, it was also decided that the number of female patients would be the same as the number of male patients.

Finally, it was desirable to detect even small differences in success rates of the two treatments since small differences could mean many more lives being saved. A sample size of about $n = 3,000$ was decided on.

Records would be collected until 3,000 were found, 1500 of which were treated with ‘A’, 1500 with ‘B’, and there were equal numbers of males and females in the study.

Data:

In this stage, the plan is executed. Instead of 1500 records of treatment ‘A’ and ‘B’, 1600 of each were found. The number of males and females was kept equal (now 1600 of each sex).

The process was to search the records in order, selecting those first encountered to get 1600 for each treatment and 1600 of each sex. Many records might be discarded whenever one quota was met and the search continued to meet the other quotas. It was also noticed that the patient’s age was available for each record, so that the effect of treatment on younger and older adults might also be considered.

The counts which fell into the various categories were assembled into the data presented here.

Author(s)

R.W. Oldford

References

R.J. MacKay and R.W. Oldford 2000, ‘Scientific Method, Statistical Method, and the Speed of Light’, *Statistical Science*, Volume 15, No. 3, pp. 254-278. <doi:10.1214/ss/1009212817>

tuberculosis_1910

Tuberculosis 1910 death rates in New York and in Richmond

Description

Tuberculosis 1910 death rates in New York and in Richmond

Format

A 2 x 2 x 2 table of counts involving three binary variables

Group One of two racial groups: "White" or "Colored"

City One of two U.S. cities: "New York" or "Richmond"

Total Either total deaths from tuberculosis or total population in 1910: "Deaths" or "Population"

Details

These are historical data taken from page 449 of Cohen and Nagel's 1934 "Introduction to Logic and Scientific Method". For this reason, the original names of the racial groups have been retained.

The data are of special historical interest in Statistics because they are one of the earliest recorded instances of a real Simpson's paradox (Simpson 1951) occurring in practice (see Blyth 1971). Preserving this historical context, the questions posed by Cohen and Nagel (1934) are also recorded here using their own words. The data and questions appear at the back of their book as exercises on "Chapter XVI: Statistical Methods".

In their table, Cohen and Nagel (1934, p. 449) include the "death rates from tuberculosis in Richmond, Virginia, and in New York City in 1910". These rates (in number per 100,000) are easily calculated and so have been excluded from the table given here.

In their words, Cohen and Nagel (1934, p. 449) pose the following two questions as exercise (*emphasis* is theirs):

"a. Does it follow that tuberculosis caused a greater mortality in Richmond than in New York?

b. Notice that the death rate for whites and that for Negroes were *lower* in Richmond than in New York, although the *total* death rate was *higher*. Are the two populations compared really *comparable*, that is, homogeneous?"

Author(s)

R.W. Oldford.

Source

"An Introduction to Logic and Scientific Method" by Morris R. Cohen and Ernest Nagel, (1934), Harcourt, Brace and Company, New York.

References

Blyth, Colin R. 1972. On Simpson's Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association*, 67, pp.364-366.

Cohen, Morris R.; Nagel, Ernest. 1934. *An Introduction to Logic and Scientific Method*. Harcourt, Brace and Company. New York.

Simpson, E.H. 1951. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society, Series B*, 13, pp. 238-241.

Index

- * **Simpson's paradox**
 - tuberculosis_1910, 15
 - * **categorical**
 - medicalRecords, 13
 - * **contingency tables**
 - tuberculosis_1910, 15
 - * **datasets**
 - tuberculosis_1910, 15
 - * **fictional**
 - medicalRecords, 13
 - * **history of statistics**
 - tuberculosis_1910, 15
 - * **medicine**
 - medicalRecords, 13
 - * **paradox**
 - medicalRecords, 13
 - * **teaching**
 - medicalRecords, 13
- eikos, 2
- eikos.default, 4, 5
- eikos.formula, 4, 7
- eikos_data, 10
- eikos_legend, 10
- eikos_x_labels, 11
- eikos_x_probs, 11
- eikos_y_labels, 12
- eikos_y_probs, 13
- medicalRecords, 13
- tuberculosis_1910, 15