Package 'ezEDA'

October 13, 2022

Type Package

Title Task Oriented Interface for Exploratory Data Analysis

Version 0.1.1

URL https://github.com/kviswana/ezEDA

BugReports https://github.com/kviswana/ezEDA/issues

Maintainer Viswa Viswanathan <kv.viswana@gmail.com>

Description Enables users to create visualizations using functions
based on the data analysis task rather than on plotting mechanics. It hides
the details of the individual 'ggplot2' function calls and
allows the user to focus on the end goal. Useful for quick preliminary explorations.
Provides functions for common exploration patterns. Some of the ideas in this
package are motivated by Fox (2015, ISBN:1938377052).

Depends R (>= 3.1)

Imports ggplot2 (>= 3.1.0), dplyr (>= 0.8.0.1), rlang (>= 0.2.1),
tidyr (>= 0.8.3), GGally (>= 1.4.0), scales (>= 1.0.0),
magrittr (>= 1.5), purrr (>= 0.3.3)

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.1

Suggests testthat, knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Viswa Viswanathan [aut, cre] (<https://orcid.org/0000-0003-4395-5061>)

Repository CRAN

Date/Publication 2021-06-29 04:40:10 UTC

R topics documented:

`category_contribution` *Plot the contribution of different categories to a measure*

Description

Plot the contribution of different categories to a measure

Usage

```
category_contribution(data, category, measure)
```

Arguments

`data`	A data frame or tibble
`category`	Unquoted name of category (can be factor, character or numeric)
`measure`	Unquoted name of measure

Value

A ggplot plot object

Examples

```
category_contribution(ggplot2::diamonds, cut, price)
category_contribution(ggplot2::diamonds, clarity, price)
```

category_tally *Plot counts of a category*

Description

Plot counts of a category

Usage

```
category_tally(data, category_column)
```

Arguments

data	A data frame or tibble
category_column	Unquoted column name of category (can be factor, character or numeric)

Value

A ggplot plot object

Examples

```
category_tally(ggplot2::mpg, class)
category_tally(ggplot2::diamonds, cut)
```

col_to_factor *Private utility function: given a possibly non-factor column passed as a quosure, convert into a factor*

Description

Private utility function: given a possibly non-factor column passed as a quosure, convert into a factor

Usage

```
col_to_factor(data, col_enquo)
```

Arguments

data	A data frame or tibble
col_enquo	A quosure

Value

A data frame or tibble with the corresponding column converted to factor if nevessary

ezeda *ezeda: A package for task oriented exploratory data analysis*

Description

The ezeda package provides functions for visualizations for exploratory data analysis. Whereas graphic packages generally provide many functions that users assemble to create suitable plots, each ezeda function warps ggplot and other code to generate a complete plot for common exploratory data analysis task corresponding to a recurring pattern.

Details

ezeda provides five categories of functions: tally, contribution, measure distribution, measure relationship, and measure trend

tally functions

- category_tally
- two_category_tally

contribution functions

- category_contribution
- two_category_contribution

measure distribution functions

- measure_distribution
- measure_distribution_by_category
- measure_distribution_by_two_categories
- measure_distribution_by_time

measure relationship functions

- two_measures_relationship
- multi_measure_relationship

measure trend functions

- measure_change_over_time
- measure_change_over_time_long

measure_change_over_time_long *Plot the change of a measure (or set of measures) over time where the data is in "long" format That is, all measures are in one column with another column labeling each measure value*

Description

Plot the change of a measure (or set of measures) over time where the data is in "long" format That is, all measures are in one column with another column labeling each measure value

Usage

```
measure_change_over_time_long(
  data,
  time_col,
  measure_labels,
  measure_values,
  ...
)
```

Arguments

data	A data frame or tibble
time_col	Unquoted column name with time values to plot on the x axis
measure_labels	Unquoted column name containing the name of the measure in the corresponding measure_values (see below) row (up to 6 measures)
measure_values	Unquted column name of the column with the measure values to be plotted
...	Unquoted names of measures to plot (up to 6 measures)

Value

A ggplot plot object

Examples

```
measure_change_over_time_long(ggplot2::economics_long, date, variable, value, pop, unemploy)
```

`measure_change_over_time_wide` *Plot the change of a measure (or set of measures) over time where each measure is in a different column*

Description

Plot the change of a measure (or set of measures) over time where each measure is in a different column

Usage

```
measure_change_over_time_wide(data, time_col, ...)
```

Arguments

`data`	A data frame or tibble
`time_col`	Unquoted column name with time values to plot on the x axis
`...`	Unquoted column names of one or more measures to plot (up to 6 measures)

Value

A ggplot plot object

Examples

```
measure_change_over_time_wide(ggplot2::economics, date, pop, unemploy)
```

`measure_distribution` *Plot the distribution of a numeric (measure) column*

Description

Plot the distribution of a numeric (measure) column

Usage

```
measure_distribution(data, measure, type = "hist", bwidth = NULL)
```

Arguments

`data`	A data frame or tibble
`measure`	Unquoted column name of containing numbers (measure)
`type`	Histogram ("hist") or Boxplot ("box")
`bwidth`	width of bin for histogram (by default uses binwidth for 30 bins)

Value

A ggplot plot object

Examples

```
measure_distribution(ggplot2::diamonds, price)
measure_distribution(ggplot2::mpg, hwy)
measure_distribution(ggplot2::mpg, hwy, bwidth = 2)
measure_distribution(ggplot2::mpg, hwy, "hist")
measure_distribution(ggplot2::mpg, hwy, "box")
```

---

measure_distribution_by_category

Plot the distribution of a numeric (measure) column differentiated by a category

---

Description

Plot the distribution of a numeric (measure) column differentiated by a category

Usage

```
measure_distribution_by_category(
  data,
  measure,
  category,
  type = "hist",
  separate = FALSE,
  bwidth = NULL
)
```

Arguments

data	A data frame or tibble
measure	Unquoted column name of measure (containing numbers)
category	Unquoted column name of category (can be factor, character or numeric)
type	Histogram ("hist") or Boxplot ("box")
separate	Boolean specifying whether to plot each category in a separate facet
bwidth	width of bin for histogram (by default uses binwidth for 30 bins)

Value

A ggplot plot object

Examples

```
measure_distribution_by_category(ggplot2::diamonds, price, cut)
measure_distribution_by_category(ggplot2::mpg, hwy, class)
measure_distribution_by_category(ggplot2::diamonds, price, cut, separate = TRUE)
measure_distribution_by_category(ggplot2::mpg, hwy, class, separate = TRUE)
measure_distribution_by_category(ggplot2::mpg, hwy, class, "box")
```

---

measure_distribution_by_two_categories

Plot the distribution of a numeric (measure) column differentiated by two categories

---

Description

Plot the distribution of a numeric (measure) column differentiated by two categories

Usage

```
measure_distribution_by_two_categories(
  data,
  measure,
  category1,
  category2,
  bwidth = NULL
)
```

Arguments

`data`	A data frame or tibble
`measure`	Unquoted column name of containing numbers (measure)
`category1, category2`	Unquoted column names of categories (can be factor, character or numeric)
`bwidth`	width of bin for histogram (by default uses binwidth for 30 bins)

Value

A ggplot plot object

Examples

```
measure_distribution_by_two_categories(ggplot2::mpg, hwy, class, fl)
measure_distribution_by_two_categories(ggplot2::diamonds, carat, cut, clarity)
```

measure_distribution_over_time

Plot the change of distribution of a numeric (measure) column over time

Description

Plot the change of distribution of a numeric (measure) column over time

Usage

```
measure_distribution_over_time(data, measure, time, bwidth = NULL)
```

Arguments

data	A data frame or tibble
measure	Unquoted column name of containing numbers (measure)
time	Unquoted name of column containing the time object
bwidth	width of bin for histogram (by default uses binwidth for 30 bins)

Value

A ggplot plot object

Examples

```
h1 <- round(rnorm(50, 60, 8), 0)
h2 <- round(rnorm(50, 65, 8), 0)
h3 <- round(rnorm(50, 70, 8), 0)
h <- c(h1, h2, h3)
y <- c(rep(1999, 50), rep(2000, 50), rep(2001, 50))
df <- data.frame(height = h, year = y)
measure_distribution_over_time(df, h, year)
```

multi_measures_relationship

Plot the relationship between many measures

Description

Plot the relationship between many measures

Usage

```
multi_measures_relationship(data, ...)
```

Arguments

data	A data frame or tibble
...	Unquoted column names of numeric columns (measures)

Value

A ggplot plot object

Examples

```
multi_measures_relationship(ggplot2::mpg, hwy, displ)
multi_measures_relationship(ggplot2::mpg, cty, hwy, displ)
```

---

two_category_contribution

Plot the contribution to a measure by combinations of two categories

---

Description

Plot the contribution to a measure by combinations of two categories

Usage

```
two_category_contribution(
  data,
  category1,
  category2,
  measure,
  separate = FALSE
)
```

Arguments

data	A data frame or tibble
category1, category2	Unquoted names of category columns (can be factor, character or numeric)
measure	Unquoted name of measure
separate	Boolean to indicate whether the plots for different combinations should be in different facets

Value

A ggplot plot object

Examples

```
two_category_contribution(ggplot2::diamonds, cut, clarity, price)
two_category_contribution(ggplot2::diamonds,  clarity, cut, price, separate = TRUE)
```

two_category_tally *Plot counts of combinations of two category columns*

Description

Plot counts of combinations of two category columns

Usage

```
two_category_tally(
  data,
  main_category,
  sub_category,
  separate = FALSE,
  position = "stack"
)
```

Arguments

data	A data frame or tibble
main_category, sub_category	Unquoted column names of two categories (can be factor, character or numeric)
separate	Boolean indicating whether the plot should be faceted or not
position	"stack" or "dodge"

Value

A ggplot plot object

Examples

```
two_category_tally(ggplot2::mpg, class, drv)
two_category_tally(ggplot2::mpg, class, drv, position = "dodge")
two_category_tally(ggplot2::mpg, class, drv, separate = TRUE)
two_category_tally(ggplot2::diamonds, cut, clarity)
two_category_tally(ggplot2::diamonds, cut, clarity, separate = TRUE)
```

two_measures_relationship *Plot the relationship between two measures and optionally highlight a category*

Description

Plot the relationship between two measures and optionally highlight a category

Usage

```
two_measures_relationship(data, measure1, measure2, category = NULL)
```

Arguments

`data`	A data frame or tibble
`measure1, measure2`	Unquoted column names of measures
`category`	Unquoted name of a category (can be factor, character or numeric)

Value

A ggplot plot object

Examples

```
two_measures_relationship(ggplot2::diamonds, carat, price)
two_measures_relationship(ggplot2::diamonds, carat, depth)

two_measures_relationship(ggplot2::mpg, displ, hwy)
two_measures_relationship(ggplot2::mpg, cty, hwy)
two_measures_relationship(ggplot2::mpg, displ, hwy, class)
```

Index