

# Package ‘psre’

October 14, 2022

**Type** Package

**Title** Presenting Statistical Results Effectively

**Version** 0.1.2

**Description** Includes functions and data used in the book “Presenting Statistical Results Effectively”, Andersen and Armstrong (2022, 978-1446269800).

**biocViews**

**Imports** stats, MASS, metap, car, nortest, lawstat, DAMisc, ggplot2, boot, grid, cowplot, fANCOVA, dplyr, tidyr, utils, magrittr, tibble, sm, multcomp, rlang, ggrepel, mgcv

**Suggests** ggeffects, nnet

**Depends** R (>= 3.5.0)

**License** GPL (>= 2)

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.2.0

**NeedsCompilation** no

**Author** Dave Armstrong [aut, cre],  
Robert Andersen [aut],  
Justin Esarey [cph],  
John Fox [cph],  
Michael Friendly [cph],  
Adrian Bowman [cph],  
Adelchi Azzalini [cph]

**Maintainer** Dave Armstrong <dave@quantoid.net>

**Repository** CRAN

**Date/Publication** 2022-08-08 15:40:01 UTC

**R topics documented:**

assocfun . . . . .	2
boot_imp . . . . .	3
caption . . . . .	3
ces . . . . .	4
cld.ss . . . . .	5
dfbhist . . . . .	6
gg_hmf . . . . .	7
glmImp . . . . .	8
gss . . . . .	9
inc_ineq . . . . .	10
india . . . . .	11
letter_plot . . . . .	12
lsa . . . . .	13
normBand . . . . .	14
optCL . . . . .	15
print.ss . . . . .	17
qqPoints . . . . .	18
relig_imp . . . . .	19
repress . . . . .	20
rrPlot . . . . .	21
shuffle . . . . .	22
simple_slopes . . . . .	23
srr_imp . . . . .	23
tpb . . . . .	24
transNorm . . . . .	25
wvs . . . . .	26
<b>Index</b>	<b>29</b>

---

assocfun	<i>Association Function</i>
----------	-----------------------------

---

**Description**

Calculates the R-squared from a LOESS regression of y on x. Can be used with outer to produce the a non-parametric correlation matrix.

**Usage**

```
assocfun(xind, yind, data)
```

**Arguments**

xind	column index of the x-variable
yind	column index of the y-variable
data	data frame from which to pull the variables.

**Value**

a squared correlation.

---

boot_imp	<i>Bootstrap Importance Function</i>
----------	--------------------------------------

---

**Description**

Function to calculate bootstrap measures of importance. This function must be passed to the boot function.

**Usage**

```
boot_imp(data, inds, obj)
```

**Arguments**

data	A data frame
inds	Indices to be passed into the function.
obj	An object of class lm.

**Value**

A vector of standard deviation of predictions for each term in the model.

---

caption	<i>Caption Grob</i>
---------	---------------------

---

**Description**

Create a caption grob

**Usage**

```
caption(lab, x = 0.5, y = 1, hj = 0.5, vj = 1, cx = 1, fs = 12, ft = "Arial")
```

**Arguments**

lab	Text giving the caption text.
x	Scalar giving the horizontal position of the label in $[0, 1]$ .
y	Scalar giving the vertical position of the label in $[0, 1]$ .
hj	Scalar giving horizontal justification parameter.
vj	Scalar giving vertical justification parameter.
cx	Character expansion factor
fs	Font size
ft	Font type

**Value**

A text grob.

---

ces

*Canadian Election Study*

---

**Description**

These data are a subset of the Canadian Election Study telephone sample (Stephenson et. al. 2020).

**Format**

A data frame with 2799 rows and 29 variables

**vote** Vote for Parliament - This variable is used to make all of the “Vote for . . .” variables. These are actual self-reported votes from the post-election study, not campaign-period vote intention. We coded those who indicated did not vote, none, don’t know and refused as missing.

**gender** Binary variable indicating respondent sex.

**agegrp** Age Group - Age is calculated by subtracting the year of birth from the survey year. Then observations are put into age-groups (18-34, 35-54, 55+)

**relig** Religious Affiliation - Respondents are coded into four groups - no religious affiliation/Agnostic, Catholic, Non-Catholic Christians (incl. Anglican, Baptist, Eastern Orthodox, Johova’s Witness, Lutheran, Pentecostal, Presbyterian, Protestant, United Church of Canada, Christian, Salvatian Army, Mennonite) and Other (incl. Buddhist, Hindu, Hewish, Muslim, Sikh). We also include an indicator variable for Catholic vs non-Catholic.

**educ** Educational Attainment coded into three categories HS or Less (incl. No schooling, some elementary, completed elementary, some secondary, completed secondary), Some Post-secondary (incl. some echnical/community college, completed technical/community college, some university) and Univ Grad (incl. bachelor’s degree, master’s degree, professional degree)

**region** Provinces are coded into four regions: Atlantic (Newfoundland and Labrador, PEI, Nova Scotia, New Brunswick), Quebec, Ontario and the West (Manitoba, Saskatchewan, Alberta and British Columbia)

**province** Province of respondent

**pid** Party with which respondent identifies. These are coded into Liberal, Conservative, NDP, Green, Bloc Quebecois and Other.

**retroper** Retrospective Personal Economic Perceptions - Whether respondent thinks his or her personal economic situation has gotten better, stayed the same or gotten worse in the past year.

**retrocan** Retrospective National Economic Perceptions - Whether respondent thinks Canada’s economic situation has gotten better, stayed the same or gotten worse in the past year.

**sp\_defence** Respondent’s opinion of how much defence spending should change in three categories - Less (much less, less), Stay the same, More (more or much more).

**sp\_envir** Respondent’s opinion of how much spending on the environment should change in three categories - Less (much less, less), Stay the same, More (more or much more).

- immig** Respondent's opinion about how immigration levels should change - Increase, Stay the same/Don't Know, Decrease
- usties** Respondent's opinion about how ties between Canada and the US should change - Much more distant, Somewhat more distant, Stay the Same/Don't Know, Somewhat closer, Much closer.
- jobspriv** Level of agreement with the following statement - The government should leave it ENTIRELY to the private sector to create jobs: Strongly disagree, Disagree, Don't know, Agree, Strongly agree.
- blame** Level of agreement with the following statement - People who don't get ahead should blame themselves, not the system: Strongly disagree, Disagree, Don't know, Agree, Strongly agree.
- poorgap** How much should be done to reduce the gap between rich and poor in Canada - Much less, Somewhat less, About the same/Don't know, Somewhat more, Much more.
- stayhome** Level of agreement with the following statement - Society would be better off if fewer women worked outside the home: Strongly disagree, Disagree, Don't know, Agree, Strongly agree.
- feelgays** Feeling thermometer for homosexuals.
- dowomen** How much do you think should be done for women: Much less, Somewhat less, About the same/Don't know, Somewhat more, Much more.
- leader\_lib** Feeling thermometer for Justin Trudeau, leader of the Liberal Party.
- leader\_con** Feeling thermometer for Andrew Scheer, leader of the Conservative Party.
- leader\_ndp** Feeling thermometer for Jagmeet Singh, leader of the NDP.
- leader\_bloc** Feeling thermometer for Yves-Francois Blanchet, the leader of the Bloc Quebecois.
- market** Market liberalism – additive scale of jobspriv, poorgap and blame variables.
- moral** Moral traditionalism – additive scale of dowomen, stayhome and feelgays.
- union** Whether respondent is a union member - yes or no.
- weight\_CES** Weighting variable for the CES.

## References

Stephenson, Laura B, Allison Harell, Daniel Rubenson, Peter John Loewen. (2020). "2019 Canadian Election Study - Phone Survey", doi:[10.7910/DVN/8RHLG1](https://doi.org/10.7910/DVN/8RHLG1), Harvard Dataverse, V1.

---

cld.ss

*Compact Letter Display for Simple Slopes*

---

## Description

Calculates a letter matrix for a simple-slopes output.

## Usage

```
## S3 method for class 'ss'
cld(x, level = 0.05, ...)
```

**Arguments**

x	An object of class 'ss'
level	Confidence level used for the letters.
...	Other arguments to be passed to generic function.

**Value**

A compact letter matrix

---

dfbhist	<i>Hybrid Plot for DFBETAS</i>
---------	--------------------------------

---

**Description**

Plots a hybrid histogram, dot plot for DFBETAS. A histogram is plotted for the observations below `cutval`. Observations above `cutval` are plotted and labelled with individual points.

**Usage**

```
dfbhist(
  data,
  varname,
  label,
  cutval = 0.25,
  binwidth = 0.025,
  xlab = "DFBETAS",
  ylab = "Frequency",
  xrange = NULL,
  yrange = NULL,
  nudge_x = NULL,
  nudge_y = NULL
)
```

**Arguments**

data	A data frame of DFBETAS values
varname	The name of the variable to plot
label	Name of variable that holds the labels that will go with the points
cutval	The value that separates the histogram from the individual points.
binwidth	The bin width for the histogram part of the display.
xlab	Label to put on the x-axis.
ylab	Label to put on the y-axis.
xrange	Alternative range to plot on the x-axis.
yrange	Alternative range to plot on y-axis
nudge_x	Vector of values to nudge labels horizontally.
nudge_y	Vector of values to nudge labels vertically.

**Value**

A ggplot.

**Examples**

```
data(wvs)
wvs <- na.omit(wvs[,c("country", "secpay", "gini_disp", "democrat")])
lmod <- lm(secpay ~ gini_disp + democrat, data=wvs)
dba <- dfbetas(lmod)
dbd <- wvs
dbd$dfb_ginil <- dba[,2]^2
dbd$dfb_democ1 <- dba[,3]^2
dfbhist(dbd, "dfb_ginil", "country")
```

---

gg\_hmf

*Heatmap Fit Plot using GGplot*


---

**Description**

Makes a Heatmap Fit plot (Esary and Pierce, 2012) using GGPlot rather than lattice that the heatmapFit package uses.

**Usage**

```
gg_hmf(
  observed,
  prob,
  method = c("loess", "gam"),
  span = NULL,
  nbin = 20,
  R = 1000,
  ...
)
```

**Arguments**

observed	Vector of observe (0/1) values used in a binary regression model.
prob	Vector of predicted probabilities from the model with observed as the dependent variable.
method	Method for making the line - LOESS or GAM (from the mgcv package.)
span	Optional span parameter to be passed in. If NULL, AICc will be used to find the appropriate span for the loess smooth.
nbin	Number of bins for the histogram.
R	Number of bootstrap resamples
...	Currently unimplemented.

**Value**

Two ggplots - the main heatmap Fit plot and a histogram that can be included as a marginal density.

**Examples**

```
data(india)
india$bjp <- ifelse(india$in_prty == 2, 1, 0)
mod1 <- glm(bjp ~ educyrs + anti_immigration,
            data=india, family=binomial)

gh1 <- gg_hmf(model.response(model.frame(mod1)),
              fitted(mod1),
              method="loess")
```

---

 glmImp

*Importance Measure for Generalized Linear Models*


---

**Description**

Calculates importance along the lines of Greenwell et al (2018) using partial dependence plots.

**Usage**

```
glmImp(obj, varname, data, level = 0.95, ci_method = c("perc", "norm"), ...)
```

**Arguments**

obj	Model object, must be able to use <code>predict(obj, type="terms")</code> .
varname	Character string giving the name of the variable whose importance will be calculated.
data	A data frame used to estimate the model.
level	Confidence level used for the confidence interval.
ci_method	Character string giving the method for calculating the confidence interval - normal or percentile.
...	Other arguments being passed down to <code>aveEffPlot</code> from the <a href="#">DAMisc</a> package.

**Value**

A data frame of importance measures with optimal bootstrapped confidence intervals.

**References**

Greenwell, Brandon M., Bradley C. Boehmke and Andrew J. McCarthy. (2018). "A Simple and Effective Model-Based Variable Importance Measure." arXiv1805.04755 [stat.ML]



## Examples

```
data(gss)
mod <- glm(childs ~ sei10 + sex + educ + age,
           data=gss, family=poisson)
g_imp1 <- glmImp(mod, "age", gss)
```

gss

*General Social Survey*

## Description

This is a subset of the 2016 US General Social Survey (Smith et. al. 2016).

## Format

A data frame with 2867 rows and 14 variables

**aidhouse** On the whole, do you think it should or should not be the government's responsibility to provide decent housing for those who can't afford it? (Definitely Should Be, Probably Should Be, Probably Should Not Be, Definitely Should Not Be)

**partyid** Combination of questions regarding partisan affiliation and strength of affiliation. Results in 7-point scale from Strong Democrat to Strong Republican along with Other Party affiliation coded separately as 8.

**realinc** Total family income in constant US dollars.

**aid\_scale** Additive scale of items with same general form as aidhouse, but including items about: decent standard of living for the old, industry with the help it needs to grow, decent standard of living for the unemployed, give financial help to university students and low-income families. Items were standardized and reversed so higher vales indicated greater generosity.

**age** Respondent age.

**sei10** Socio-economic Status indicator - theoretically ranges from 0 to 1.

**sex** Binary indicator of respondent sex.

**tax** Are your federal income taxes too high, about right or too low?

**newsfrom** Where do you get most of your information about current news events? (newspapers, magazines, the Internet, books or other printed materials, TV, radio, government agencies, family, friends, colleagues, some other source)

**educ** Total number of years of formal education completed.

**sparts** Please tell me whether you would like to see more or less government spending on culture and the arts. Remember, that if you say "much more" it might require a tax increase to pay for it. Five-point Scale from Spend Much More to Spend Much Less.

**wtssnr** Survey Weighting variable

**party3** Party ID variable that puts leaners, independents and others together in Other; Strong and moderate Democrats are coded as Democrat while strong and moderate Republicans are coded Republican.

**childs** Number of children in respondent's household.

## References

Smith, Tom W, Peter Marsden, Michael Hout, and Jibum Kim. (2016). General Social Surveys, 1972-2016 [machine-readable data file] Principal Investigator, Tom W. Smith; Co-Principal Investigator, Peter V. Marsden; Co-Principal Investigator, Michael Hout; Sponsored by National Science Foundation. -NORC ed.- Chicago: NORC at the University of Chicago [producer and distributor]. Data accessed from the GSS Data Explorer website at <https://gss.norc.org/get-the-data>.

---

inc\_ineq

*Income and Inequality Data*

---

## Description

This merges the Gini coefficient measured in disposable income from the Standardized World Income Inequality Data (Solt, 2020) with GDP and population data from the Penn World Tables version 10 (Feenstra et. al., 2015).

## Format

A data frame with 12810 rows and 6 variables

**country** Country Name

**year** Year

**rgdpe** Expenditure-side real GDP at chained PPPs (in mil. 2017 US Dollars). Useful for making cross-country/cross-time comparisons of relative living standards. [PWT]

**pop** Population in millions. [PWT]

**gdp\_cap** GDP divided by population from PWT.

**gini** Gini Coefficient (Disposable Income) [SWIID]

## References

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" American Economic Review, 105(10), 3150-3182, available for download at <https://www.rug.nl/ggdc/productivity/pwt/>.

Solt, Frederick. 2020. "Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database." Social Science Quarterly 101(3):1183-1199. SWIID Version 9.0, October 2020.

india

*Indian Election Data***Description**

These data are from the International Social Survey Programme: National Identity III survey (ISSP Research Group 2015). This subset contains only the data from India.

**Format**

A data frame with 1530 rows and 22 variables

**patriotism** Additive scale of level of agreement regarding statements about patriotism (5-point scale Agree Strongly to Disagree Strongly): Strengthens India's place in the world (-), leads to intolerance in India (+), is needed for India to remain united (-), leads to negative attitudes toward immigrants in India (+). All items are standardized before being summed.

**imp\_roots** Additive scale of level of agreement regarding statements about importance of the following things for being truly Indian (5-point scale Agree Strongly to Disagree Strongly, all indicators positively associated with the scale): being born in India, having Indian citizenship, having lived in India most of your life, ability to speak Hindi, to be Hindu, to respect India's political institutions and laws, to feel Indian and to have Indian ancestry. All items are standardized and reversed before being summed.

**pride\_country** Additive scale of level of agreement regarding statements about pride in the following things about India (5-point scale Agree Strongly to Disagree Strongly, all indicators positively associated with the scale): the way democracy works, India's political influence in the world, India's economic achievements, its social security system, its scientific and technological achievements, its achievements in sports, its achievements in the arts and literature, India's armed forces, its history, its fair and equal treatment of all groups in society. All items are standardized and reversed before being summed.

**country\_first** Additive scale of level of agreement regarding statements about the following things regarding India relationships with other countries (5-point scale Agree Strongly to Disagree Strongly, all indicators positively associated with the scale): India should limit the import of foreign products to protect national economy, India should follow its own interests even if that leads to conflict, foreigners should not be allowed to buy land in India, India's television should give preference to Indian films and programs. All items are standardized and reversed before being summed.

**anti\_immigration** Additive scale of level of agreement regarding statements about pride in the following things about immigrants (5-point scale Agree Strongly to Disagree Strongly): immigrants increase crime rates (+), immigrants are generally good for India's economy (-), immigrants take jobs away from people born in India (+), immigrations improve India's society by bringing new ideas and cultures (-), India's culture is generally undermined by immigrants (-), legal immigrants to India who are not citizens should not have the same rights as Indian citizens (+), India should take stronger measures to exclude illegal immigrants (+), legal immigrants should have equal access to public education as Indian citizens (-). All items standardized before being summed.

**educyrs** Years of formal education, capped at 20.

**age** Respondent age.

**sb** Dummy indicator for Scheduled or Backward Caste.

**sex** Binary indicator of respondent gender.

**partliv** Living in a steady relationship with a partner.

**religgrp** Religious group to which respondent belongs.

**attend** Frequency of attendance at religious services.

**topbot** Self-placement in socio-economic status decile.

**in\_ethn1** Respondent ethnicity.

**hhchildr** Number of children under 18 in the household.

**in\_inc** Income group in local currency.

**urbrural** Urban-rural category of residence.

**work** Ever had paying work (currently, previously, never).

**mainstat** Main current employment status.

**union** Union membership (current, previous, never).

**vote\_bjp** Vote for the BJP in most recent election.

**vote\_le** Vote turnout in last election.

**in\_prty** Party voted for in most recent parliamentary election.

**party\_lr** Party voted for in most recent parliamentary election in terms of ideological position.

## References

ISSP Research Group (2015): International Social Survey Programme: National Identity III - ISSP 2013. GESIS Data Archive, Cologne. ZA5950 Data file Version 2.0.0, [doi:10.4232/1.12312](https://doi.org/10.4232/1.12312)

---

letter\_plot

*Dot Plot with Letter Display*

---

## Description

Produces an dot plot with error bars along with a compact letter display

## Usage

```
letter_plot(fits, letters)
```

## Arguments

fits	Output from ggpredict from the <b>ggeffects</b>
letters	A matrix of character strings giving the letters from a compact letter display. This is most often from a call to <code>cld</code> from the <b>multcomp</b> package.

**Value**

A ggplot.

**Examples**

```
library(psre)
library(ggeffects)
library(multcomp)
library(dplyr)
library(ggplot2)
data(wvs)
wvs$civ <- with(wvs, case_when(
  civ == 4 ~ "Islamic",
  civ == 6 ~ "Latin American",
  civ == 7 ~ "Orthodox",
  civ == 8 ~ "Sinic",
  civ == 9 ~ "Western",
  TRUE ~ "Other"))
wvs$civ = factor(wvs$civ, levels=c("Western",
                                  "Sinic",
                                  "Islamic",
                                  "Latin American",
                                  "Orthodox",
                                  "Other"))

mod <- lm(resemaival ~ civ + gdp_cap +
          pct_secondary + pct_univ_degree +
          pct_high_rel_imp, data=wvs)

eff <- ggpredict(mod,
                 "civ",
                 ci.lvl = .95)

pwc <- summary(glht(mod, linfct=mcp(civ = "Tukey")),
               test=adjusted(type="none"))
cld1 <- cld(pwc)
lmat <- cld1$mcletters$LetterMatrix
eff$x <- reorder(eff$x, eff$predicted, mean)
letter_plot(eff, lmat) +
  labs(x="Predicted Emancipative Values\n(95% Confidence Interval)")
```

**Description**

Produces a linear scatterplot array with marginal histograms

**Usage**

```
lsa(
  formula,
  xlabel = NULL,
  ylab = NULL,
  data,
  psize = 1,
  pshape = 1,
  pcolor = "gray65"
)
```

**Arguments**

formula	Formula giving the variables to be plotted.
xlabel	Vector of character strings giving the labels of variables to be used in place of the variable names.
ylab	Character string giving y-variable label to be used instead of variable name.
data	A data frame that holds the variables to be plotted.
psize	Size of points.
pshape	Shape of points.
pcolor	Color of points.

**Value**

A cowplot object.

**Examples**

```
data(wvs)
lsa(formula = as.formula(sacsecval ~ resemaval + moral +
  pct_univ_degree + pct_female +
  pct_low_income),
  xlabel = c("Emancipative Vals", "Moral Perm",
    "% Univ Degree", "% Female", "% Low Income"),
  ylab = "Secular Values",
  data=wvs)
```

---

normBand

*Kernel Density with Normal Density Overlay*


---

**Description**

Calculates a kernel density estimate of the data along with confidence bounds. It also computes a normal density and confidence bounds for the normal density with the same mean and variance as the observed data.

**Usage**

```
normBand(x, ...)
```

**Arguments**

x                    A vector of values whose density is to be calculated  
...                   Other arguments to be passed down to `sm.density`.

**Details**

The function is largely cribbed from the **sm** package by Bowman and Azzalini

**Value**

A named vector of scalar measures of fit

**Author(s)**

Dave Armstrong, A.W. Bowman and A. Azzalini

**References**

A.W. Bowman and A. Azzalini, R package `sm`: nonparametric smoothing methods (version 5.6).

---

optCL

*Calculate the Optimal Visual Testing Confidence Level*

---

**Description**

Calculates the Optimal Visual Testing (OVT) confidence level. The OVT level is a level you can use to make confidence intervals such that the overlapping (or non-overlapping) of confidence intervals preserves the pairwise testing results. That is, statistically different estimates have confidence intervals that do not overlap and statistically indistinguishable intervals have confidence intervals that do overlap. It does not always work perfectly, but it generally results in fewer inferential errors than the nominal level.

**Usage**

```
optCL(  
  obj = NULL,  
  varname = NULL,  
  b = NULL,  
  v = NULL,  
  resdf = Inf,  
  level = 0.95,  
  quasi_vars = NULL,  
  add_ref = TRUE,
```

```

grid_range = c(0.75, 0.99),
grid_length = 100,
adjust = p.adjust.methods[c(8, 1:7)]
)

```

### Arguments

obj	A model object, on which coef and vcov can be called. Either obj and varname or b and v must be specified.
varname	The name of a variable whose coefficients will be used.
b	Optional vector of coefficients to be passed into the function. it overrides the coefficients in obj. Either obj and varname or b and v must be specified.
v	Optional variance-covariance matrix. This can be specified even if obj and varname are specified. It replaces the variance-covariance matrix from the model.
resdf	If only b and v are passed in, this gives the residual degrees of freedom for the t-statistics.
level	The confidence level to use for testing.
quasi_vars	An optional vector of quasi-variances that will be used to make the confidence intervals.
add_ref	If obj and varname are passed in, an optional 0 is added to the front of the vector of coefficients, along with a leading row and column of zeros on the variance-covariance matrix to represent the reference category.
grid_range	The range of values over which to do the grid search.
grid_length	The number of values in the grid.
adjust	String giving the method used to adjust the p-values for multiplicity. All methods allowed in p.adjust.methods are permitted. None is the default.

### Value

A list with the following elements:

- opt\_levels** The optimal confidence levels that all have identical minimal error rates.
- opt\_diffs** The sum of differences between upper and lower bounds that characterize the appropriate visual tests. Larger numbers are better.
- opt\_errors** The proportion of errors across all simple contrasts.
- lev\_errors** The proportion of errors made at the nominal significance level.
- tot\_comps** The total number of comparisons
- lev\_dat** If there are inferential errors at the nominal level, this is a data frame that has all of the information about which comparisons are not appropriately represented by the overlaps in confidence intervals.
- err\_dat** If there are inferential errors at the optimal level, this is a data frame that has all of the information about which comparisons remain not appropriately represented by the overlaps in optimized confidence intervals.



**Examples**

```
data(wvs)
wvs$civ2 <- "Other"
wvs$civ2 <- ifelse(wvs$civ == 9,
                  "Western",
                  wvs$civ2)
wvs$civ2 <- ifelse(wvs$civ == 6,
                  "Latin American",
                  wvs$civ2)
wvs$civ2 <- as.factor(wvs$civ2)

intmod <- lm(resemaaval ~ civ2 * pct_secondary,
            data=wvs)

ss2 <- simple_slopes(intmod,
                    "pct_secondary",
                    "civ2")
o2 <- optCL(b=ss2$est$slope, v=ss2$v)
```

---

print.ss

*Print Method for Simple Slopes*

---

**Description**

Prints the results of the Simple Slopes function

**Usage**

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

**Arguments**

x                    An object of class ss.  
...                   Other arguments passed down to print

**Value**

Printed output

qqPoints

*Quantile Comparison Data***Description**

Makes data that can be used in quantile comparison plots.

**Usage**

```
qqPoints(
  x,
  distribution = "norm",
  line = c("quartiles", "robust", "none"),
  conf = 0.95,
  ...
)
```

**Arguments**

<code>x</code>	vector of values whose quantiles will be calculated.
<code>distribution</code>	String giving the theoretical distribution against which the quantiles of the observed data will be compared. These need to be functions that have <code>q</code> and <code>d</code> functions in R. Defaults to "norm".
<code>line</code>	String giving the nature of the line that should be drawn through the points. If "quartiles", the line is drawn connecting the 25th and 75th percentiles. If "robust" a robust linear model is used to fit the line.
<code>conf</code>	Confidence level to be used.
<code>...</code>	Other parameters to be passed down to the quantile function.

**Value**

A data frame with variables `x` observed quantiles, `theo` the theoretical quantiles and `lwr` and `upr` the confidence bounds. The slope and intercept of the line running through the points are returned as `a` and `b` as an attribute of the data.

**Examples**

```
x <- rchisq(100, 3)
qqdf <- qqPoints(x)
a <- attr(qqdf, "ab")[1]
b <- attr(qqdf, "ab")[2]
l <- min(qqdf$theo) * b + a
u <- max(qqdf$theo) * b + a
library(ggplot2)
ggplot(qqdf, aes(x=theo, y=x)) +
  geom_ribbon(aes(ymin=lwr, ymax=upr), alpha=.15) +
  geom_segment(aes(x=min(qqdf$theo), xend=max(qqdf$theo), y = 1, yend=u)) +
```

```
geom_point(shape=1) +
theme_classic() +
labs(x="Theoretical Quantiles",
      y="Observed Quantiles")
```

---

relig\_imp

*World Values Survey Religious Importance*


---

### Description

A subset of data from the second through fifth waves of the World Values Survey measuring religious importance.

### Format

A data frame with 224 rows and 4 variables

**country** Country of respondent residence.

**relig\_imp** Response Category for the religious importance variable: Very Important, Rather Important, Not Very Important and Not At All Important.

**n** Proportion of observation in each country-response category.

**I** The average value of religious importance on the 1-4 scale.

### Details

These data come from the same source as the wvs data. These are aggregated responses to the question about religious importance by country and religious importance response.

The dataset has 224 rows and 4 variables. The variables are as follows:

### References

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014a. World Values Survey: Round Two - Country-Pooled Datafile Version: <https://www.worldvaluessurvey.org/WVSDocumentationWV2.jsp>. Madrid: JD Systems Institute.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014b. World Values Survey: Round Three - Country-Pooled Datafile Version: <https://www.worldvaluessurvey.org/WVSDocumentationWV3.jsp>. Madrid: JD Systems Institute.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014c. World Values Survey: Round Four - Country-Pooled Datafile Version: <https://www.worldvaluessurvey.org/WVSDocumentationWV4.jsp>. Madrid: JD Systems Institute.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014d. World Values Survey: Round Five - Country-Pooled Datafile Version: <https://www.worldvaluessurvey.org/WVSDocumentationWV5.jsp>. Madrid: JD Systems Institute.

---

repress

*State Repression Dataset*

---

## Description

These data consider the democracy-repression nexus. While they are different data than used in previous studies, they are similar in spirit to the data used in Poe and Tate (1994) and in Davenport and Armstrong (2004).

## Format

A data frame with 1530 rows and 22 variables

**gwno** Gleditsch and Ward numeric country code

**year** Year of observation

**pts\_s** Political Terror Scale coding of State Department country reports.

**rgdpe** Penn World Tables measure of GDP in millions \$USD.

**pop** Population in millions from the Penn World Tables.

**pr** Freedom House's Political Rights measure (0-40)

**cwar** Civil War indicator from the UCDP Armed Conflict Database.

**iwar** Interstate War indicator from the UCDP Armed Conflict Database.

## References

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer 2015. 'The Next Generation of the Penn World Table' *American Economic Review*, 105(10), 3150-3182, available for download at <https://www.rug.nl/ggdc/productivity/pwt/>.

Freedom House. (2020). *Freedom in the World 2020*. New York: Freedom House.

Gleditsch, Nils Petter; Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg & Havard Strand, 2002. 'Armed Conflict 1946–2001: A New Dataset', *Journal of Peace Research* 39(5): 615–637.

Gibney, Mark, Linda Cornett, Reed Wood, Peter Haschke, Daniel Arnon, Attilio Pisano, Gray Barrett, and Baekwan Park. 2020. 'The Political Terror Scale 1976-2019.' Date Retrieved, from the Political Terror Scale website: <https://www.politicalterrorsscale.org/>.

---

rrPlot	<i>Residual-Residual Plot</i>
--------	-------------------------------

---

**Description**

Produces a linear scatterplot array with marginal histograms. The plots have OLS regression lines and a 45-degree line.

**Usage**

```
rrPlot(
  formula,
  xlabel = NULL,
  ylab = NULL,
  data,
  return = c("grid", "grobs"),
  psize = 1,
  ptshape = 1,
  ptcol = "gray65"
)
```

**Arguments**

formula	Formula giving the variables to be plotted.
xlabels	Vector of character strings giving the labels of variables to be used in place of the variable names.
ylab	Character string giving y-variable label to be used instead of variable name.
data	A data frame that holds the variables to be plotted.
return	A string identify what to return. If 'grid', then a cowplot object is returned with all plots printed. If 'grobs' then a list with all of the individual ggplots/grobs is returned.
psize	Size of points.
ptshape	Shape of points.
ptcol	Color of points.

**Value**

A cowplot object.

**Examples**

```
data(wvs)
library(MASS)
lmod <- lm(secpay ~ gini_disp + democrat + log(pop), data=wvs)
e1_m <- rlm(secpay ~ gini_disp + democrat + log(pop),
            data=wvs, method="M")$residuals
```

```
e1_mm <- rlm(secpay ~ gini_disp + democrat + log(pop),
             data=wvs, method="MM")$residuals
e1dat <- data.frame(OLS = lmod$residuals,
                   M = e1_m,
                   MM = e1_mm)
rrPlot(OLS ~ M + MM, data=e1dat)
```

---

shuffle

*Shuffle coefficients and standard errors together*


---

### Description

Function shuffles together coefficients and standard errors with a significance flag.

### Usage

```
shuffle(b, pv, se, alpha = 0.05, digits = 3, names = NULL)
```

### Arguments

b	Vector of coefficients
pv	Vector of p-values corresponding to b
se	Vector of standard errors corresponding to b
alpha	Alpha level for the significance flag
digits	Number of digits to print
names	A character vector of coefficient names as long as b

### Value

A character vector of printed output

### Examples

```
library(nnet)
data(repress)
mrm <- multinom(pts_s ~ pr + cwar + iwar + log(rgdpe) + log(pop), data=repress)
b <- coef(mrm)
v <- vcov(mrm)
b <- c(t(b))
se <- sqrt(diag(v))
pv <- 2*pnorm(abs(b/se), lower.tail=FALSE)
tab11_7 <- matrix(shuffle(b, pv, se), ncol=4)
rownames(tab11_7) <- rep("", 12)
rownames(tab11_7)[seq(1, 12, by=2)] <- colnames(coef(mrm))
colnames(tab11_7) <- paste0("PTS = ", 2:5)
noquote(tab11_7)
```

---

simple_slopes	<i>Calculate Simple Slopes</i>
---------------	--------------------------------

---

**Description**

Calculates Simple Slopes from an interaction between a categorical and quantitative variable.

**Usage**

```
simple_slopes(mod, quant_var, cat_var)
```

**Arguments**

mod	A model object that contains an interaction between a quantitative variable and a factor.
quant_var	A character string giving the name of the quantitative variable in the interaction.
cat_var	A character string giving the name of the factor variable in the interaction.

**Value**

A data frame giving the conditional partial effect along with standard errors, t-statistics and p-values.

---

srr_imp	<i>Absolute Importance Measure</i>
---------	------------------------------------

---

**Description**

Calculates absolute importance along the lines consistent with relative importance as defined by Silber, Rosenbaum and Ross (1995)

**Usage**

```
srr_imp(  
  obj,  
  data,  
  boot = TRUE,  
  R = 250,  
  level = 0.95,  
  pct = FALSE,  
  combine_terms = NULL,  
  ...  
)
```

**Arguments**

obj	Model object, must be able to use <code>predict(obj, type="terms")</code> .
data	A data frame used to estimate the model.
boot	Logical indicating whether bootstrap confidence intervals should be produced and included.
R	If <code>boot=TRUE</code> , the number of bootstrap samples to be used.
level	Confidence level used for the confidence interval.
pct	Logical indicating whether importance figures should be turned into percentages. Default is <code>TRUE</code> .
combine_terms	A named list of the names of terms to be combined into one.
...	Other arguments being passed down to <code>boot</code> .

**Value**

A data frame of importance measures with optimal bootstrapped confidence intervals.

**References**

Silber, J. H., Rosenbaum, P. R. and Ross, R N (1995) Comparing the Contributions of Groups of Predictors: Which Outcomes Vary with Hospital Rather than Patient Characteristics? *JASA* 90, 7–18.

**Examples**

```
data(gss)
mod <- glm(childs ~ sei10 + sex + educ + age,
           data=gss, family=poisson)
srr_imp(mod, data=gss)
```

---

 tpb

---

*Truncated Power Basis Functions*


---

**Description**

Makes truncated power basis spline functions.

**Usage**

```
tpb(x, degree = 3, nknots = 3, knot_loc = NULL)
```

**Arguments**

x	Vector of values that will be transformed by the basis functions.
degree	Degree of the polynomial used by the basis function.
nknots	Number of knots to use in the spline.
knot_loc	Location of the knots. If <code>NULL</code> they will be placed evenly along the appropriate quantiles of the variable.



**Value**

A  $n \times \text{degree} + n\text{knots}$  matrix of basis function values.

**Examples**

```
library(psre)
data(wvs)
smo3 <- lm(secpay ~ tpb(gini_disp, degree=3, knot_loc=.35) + democrat, data=wvs)
summary(smo3)
```

---

transNorm

*Transform Variables to Normality*


---

**Description**

Note, that we do not use the Doornik-Hansen test because the implementation in 'normwh.test' has been archived. We continue to use the other methods prescribed in Velez et al.

**Usage**

```
transNorm(
  x,
  start = 0.01,
  family = c("bc", "yj"),
  lams,
  combine.method = c("Stouffer", "Fisher", "Average"),
  ...
)
```

**Arguments**

x	Vector of values to be transformed to normality
start	Positive value to be added to variable to ensure all values are positive. This follows the transformation of the variable to have its minimum value be zero.
family	Family of test - Box-Cox or Yeo-Johnson.
lams	A vector of length 2 giving the range of values for the transformation parameter.
combine.method	String giving the method used to to combine p-values from normality tests.
...	Other arguments, currently unimplemented.

**Details**

Uses the method proposed by Velez, Correa and Marmolejo-Ramos to normalize variables using Box-Cox or Yeo-Johnson transformations.

**Value**

A scalar giving the optimal transformation parameter.

**References**

Velez Jorge I., Correa Juan C., Marmolejo-Ramos Fernando. (2015) "A new approach to the Box-Cox Transformation" *Frontiers in Applied Mathematics and Statistics*.

**Examples**

```
data(wvs)
library(car)
lam <- transNorm(wvs$gdp_cap,
                 family="yj",
                 lams =c(-2,2))
wvs$trans_gdp <- yjPower(wvs$gdp_cap,
                        lambda=lam)
```

---

wvs

---

*World Values Survey Aggregate Data*


---

**Description**

A subset of data from the second through fifth waves of the World Values Survey.

**Format**

A data frame with 162 rows and 26 variables

**country** Country of respondent residence.

**wave** Wave of the survey.

**year** Year of the survey.

**pct\_high\_rel\_imp** Religious importance is coded as Very, Rather, Not very or Not at all important in the individual data. This variable is the proportion of respondents who indicated Very or Rather.

**pct\_female** Proportion of observations identifying as female.

**mean\_lr** Left-right self-placement is coded on a 1 (Left) to 10 (Right) scale in the individual data. The mean\_lr variable is the country-wave average of left-right self-placement.

**pct\_less\_secondary, pct\_secondary, pct\_some\_univ, pct\_univ\_degree** In the individual data, education is coded as Less than secondary, Secondary complete, Some university and University degree or more. In the aggregate data, we calculate the proportion of observations in each category.

**pct\_low\_income, pct\_mid\_income, pct\_high\_income** In the individual data, income is coded in decies (i.e., a 1-10 scale). In the aggregate data, we calculate the proportion of observations in categories 1-3 (Low), 4-7 (Middle) and 8-10 (High) categories.

- moral** In the individual data, we created an additive scale of variables about how justifiable the following actions are: Illegally claiming government benefits, Avoiding a fare on public transport, Cheating on taxes, Accepting a bribe, Homosexuality, Divorce, Abortion, Prostitution, Euthanasia, Suicide on a scale from 1 (Never justifiable) to 10 (Always Justifiable). In the aggregate data, we calculate the country-wave average of this scale.
- sacsecval** Secular Values - opposite of traditional values wherein religion, parent-child ties, deference to authority and traditional family values are paramount. In the aggregate data, we take the country-wave average of this scale.
- secpay** Imagine two secretaries, of the same age, doing practically the same job. One finds out that the other earns considerably more than she does. The better paid secretary, however, is quicker, more efficient and more reliable at her job. In your opinion, is it fair or not fair that one secretary is paid more than the other? The secpay variable is the proportion of people in each country indicating that the pay discrepancy is unfair.
- resemaval** Emancipative Values - preference for gender and racial equality, liberty and personal autonomy. In the aggregate data, we take the country-wave average of this scale.
- rgdpe** Expenditure-side real GDP at chained PPPs (in mil. 2017US\$). Useful for making cross-country/cross-time comparisons of relative living standards. Obtained from Penn World Tables.
- rel\_soc** Dummy variable indicating places where at least 75 respondents identified religion as being important.
- pop** Population in Millions, obtained from Penn World Tables.
- gdp\_cap** GDP/capita:  $rgdpe/pop$ .
- gini\_disp** Gini coefficient in terms of disposable income from the SWIID.
- gini\_mkt** Gini coefficient in terms of market prices from the SWIID.
- polrt** Measure of the violation of political rights from the Freedom in the World Project. Coded on a scale from 1 (fewest violations) to 7 (most violations).
- civlib** Measure of the violation of civil liberties from the Freedom in the World Project. Coded on a scale from 1 (fewest violations) to 7 (most violations).
- democrat** Using the freedom status variable, we code a country as a democracy if in the past 15 years it was always at least partly free and was free for at least 50 percent of the time. This follows the work of Weakliem et. al. (2005).
- civ** Categories defining the civilization in which each country belongs. Other=0, African=1, Buddhist=2, Hindu=3, Islamic=4, Japanese=5, Latin American=6, Orthodox=7, Siinic=8, Western=9.

## Details

We started with waves 2 (Inglehart et. al., 2014a), 3 (Inglehart et. al., 2014b), 4 (Inglehart et. al., 2014c) and 5 (Inglehart et. al., 2014d) of the World Values Survey (WVS). The WVS is a cross-national survey effort aimed at describing the character of value systems around the globe. From each survey, we capture country and survey year, several demographic variables (Religious Importance, fairness, left-right self-placement, education, income, sex and age) along with some values scales (emancipative values and secular values). We also capture several questions about the extent to which several controversial actions are morally justifiable. We add data from several other projects to these data. To measure inequality, we use the Standardized World Income Inequality

Data (Solt, 2020). From this dataset, we capture the Gini Coefficient (both in disposable income and market income, though we tend to use the former in models). We obtain GDP and population data from the Penn World Tables version 10 (Feenstra et al., 2015). We gather data on political rights, civil liberties and freedom status from the Freedom in the World Project (Freedom House, 2020). We use the civilizations codes from Henderson and Tucker (2001), which were used to test Huntington's (1996) argument about the "clash of civilizations". Finally, we get the human development index (HDI) from the United Nations Development Programme (2020). The combined dataset has 237,787 individual observations nested within 84 countries. Most countries appear in only one or two waves (65), but nine appear in three waves and 10 in four waves.

We aggregate the variables in the individual dataset by country-wave to produce a more manageable data set. The aggregate dataset has 162 rows and 38 variables. The variables are as follows:

## References

- Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" *American Economic Review*, 105(10), 3150-3182, available for download at <https://www.rug.nl/ggdc/productivity/pwt/>.
- Freedom House. (2020). *Freedom in the World 2020*. New York: Freedom House.
- Henderson, Errol A. and Richard Tucker. 2001. "Clear and Present Strangers: The Clash of Civilizations and International Conflict." *International Studies Quarterly*, 45(2):317–338.
- Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014a. *World Values Survey: Round Two - Country-Pooled Datafile Version*: <https://www.worldvaluessurvey.org/WVSDocumentationWV2.jsp>. Madrid: JD Systems Institute.
- Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014b. *World Values Survey: Round Three - Country-Pooled Datafile Version*: <https://www.worldvaluessurvey.org/WVSDocumentationWV3.jsp>. Madrid: JD Systems Institute.
- Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014c. *World Values Survey: Round Four - Country-Pooled Datafile Version*: <https://www.worldvaluessurvey.org/WVSDocumentationWV4.jsp>. Madrid: JD Systems Institute.
- Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014d. *World Values Survey: Round Five - Country-Pooled Datafile Version*: <https://www.worldvaluessurvey.org/WVSDocumentationWV5.jsp>. Madrid: JD Systems Institute.
- Solt, Frederick. 2020. "Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database." *Social Science Quarterly* 101(3):1183-1199. SWIID Version 9.0, October 2020.

# Index

- \* **datasets**
  - ces, 4
  - gss, 9
  - inc\_ineq, 10
  - india, 11
  - relig\_imp, 19
  - repress, 20
  - wvs, 26
- assocfun, 2
- boot\_imp, 3
- caption, 3
- ces, 4
- cld.ss, 5
- DAMisc, 8
- dfbhist, 6
- gg\_hmf, 7
- glmImp, 8
- gss, 9
- inc\_ineq, 10
- india, 11
- letter\_plot, 12
- lsa, 13
- normBand, 14
- optCL, 15
- print.ss, 17
- qqPoints, 18
- relig\_imp, 19
- repress, 20
- rrPlot, 21
- shuffle, 22
- simple\_slopes, 23
- srr\_imp, 23
- tpb, 24
- transNorm, 25
- wvs, 26