

# Package ‘ggRandomForests’

May 9, 2022

**Type** Package

**Title** Visually Exploring Random Forests

**Version** 2.2.0

**Date** 2022-05-09

**Author** John Ehrlinger <john.ehrlinger@gmail.com>

**Maintainer** John Ehrlinger <john.ehrlinger@gmail.com>

**License** GPL (>= 3)

**Encoding** UTF-8

**Language** en-US

**URL** <https://github.com/ehrlinger/ggRandomForests>

**BugReports** <https://github.com/ehrlinger/ggRandomForests/issues>

**Description** Graphic elements for exploring Random Forests using the 'randomForest' or 'randomForestSRC' package for survival, regression and classification forests and 'ggplot2' package plotting.

**Depends** R (>= 3.5.0), randomForestSRC (>= 1.5.5), randomForest

**Imports** survival, parallel, tidyr, ggplot2

**Suggests** testthat, rmdformats, bookdown, RColorBrewer, MASS, dplyr, knitr, rmarkdown, plot3D, lintr, datasets

**RoxygenNote** 7.1.2

**LazyData** true

**LazyDataCompression** bzip2

**VignetteBuilder** knitr

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2022-05-09 14:30:02 UTC

**R topics documented:**

ggRandomForests-package	3
cache_rfsrc_datasets	4
calc_auc	5
calc_roc.rfsrc	6
combine.gg_partial	8
gg_error	9
gg_interaction	11
gg_minimal_depth	13
gg_minimal_vimp	16
gg_partial	18
gg_partial_coplot.rfsrc	21
gg_rfsrc.rfsrc	23
gg_roc.rfsrc	25
gg_survival	26
gg_variable	28
gg_vimp	30
interaction_data	32
kaplan	35
nelson	36
partial_coplot_data	37
partial_data	39
partial_surface_data	42
plot.gg_error	45
plot.gg_interaction	47
plot.gg_minimal_depth	49
plot.gg_minimal_vimp	51
plot.gg_partial	53
plot.gg_partial_list	56
plot.gg_rfsrc	59
plot.gg_roc	61
plot.gg_survival	62
plot.gg_variable	63
plot.gg_vimp	66
print.gg_minimal_depth	68
quantile_pts	69
rfsrc_data	70
shift	73
surface_matrix	74
vargsel_data	75

---

ggRandomForests-package

*ggRandomForests: Visually Exploring Random Forests*

---

## Description

ggRandomForests is a utility package for randomForestSRC (Ishwaran et.al. 2014, 2008, 2007) for survival, regression and classification forests and uses the ggplot2 (Wickham 2009) package for plotting results. ggRandomForests is structured to extract data objects from the random forest and provides S3 functions for printing and plotting these objects.

The randomForestSRC package provides a unified treatment of Breiman's (2001) random forests for a variety of data settings. Regression and classification forests are grown when the response is numeric or categorical (factor) while survival and competing risk forests (Ishwaran et al. 2008, 2012) are grown for right-censored survival data.

Many of the figures created by the ggRandomForests package are also available directly from within the randomForestSRC package. However, ggRandomForests offers the following advantages:

- Separation of data and figures: ggRandomForest contains functions that operate on either the `rfsrc` forest object directly, or on the output from randomForestSRC post processing functions (i.e. `plot.variable`, `var.select`, `find.interaction`) to generate intermediate ggRandomForests data objects. S3 functions are provide to further process these objects and plot results using the ggplot2 graphics package. Alternatively, users can use these data objects for additional custom plotting or analysis operations.
- Each data object/figure is a single, self contained object. This allows simple modification and manipulation of the data or ggplot2 objects to meet users specific needs and requirements.
- The use of ggplot2 for plotting. We chose to use the ggplot2 package for our figures to allow users flexibility in modifying the figures to their liking. Each S3 plot function returns either a single ggplot2 object, or a list of ggplot2 objects, allowing users to use additional ggplot2 functions or themes to modify and customize the figures to their liking.

The ggRandomForests package contains the following data functions:

- `gg_rfsrc`: randomForest[SRC] predictions.
- `gg_error`: randomForest[SRC] convergence rate based on the OOB error rate.
- `gg_roc`: ROC curves for randomForest classification models.
- `gg_vimp`: Variable Importance ranking for variable selection.
- `gg_minimal_depth`: Minimal Depth ranking for variable selection (Ishwaran et.al. 2010).
- `gg_minimal_vimp`: Comparing Minimal Depth and VIMP rankings for variable selection.
- `gg_interaction`: Minimal Depth interaction detection (Ishwaran et.al. 2010)
- `gg_variable`: Marginal variable dependence.
- `gg_partial`: Partial (risk adjusted) variable dependence.
- `gg_partial_coplot`: Partial variable conditional dependence (computationally expensive).

- `gg_survival`: Kaplan-Meier/Nelson-Aalen hazard analysis.

Each of these data functions has an associated S3 plot function that returns `ggplot2` objects, either individually or as a list, which can be further customized using standard `ggplot2` commands.

## References

- Breiman, L. (2001). Random forests, *Machine Learning*, 45:5-32.
- Ishwaran H. and Kogalur U.B. (2014). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.5.5.12.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R. *R News* 7(2), 25–31.
- Ishwaran H., Kogalur U.B., Blackstone E.H. and Lauer M.S. (2008). Random survival forests. *Ann. Appl. Statist.* 2(3), 841–860.
- Ishwaran, H., U. B. Kogalur, E. Z. Gorodeski, A. J. Minn, and M. S. Lauer (2010). High-dimensional variable selection for survival data. *J. Amer. Statist. Assoc.* 105, 205-217.
- Ishwaran, H. (2007). Variable importance in binary regression trees and forests. *Electronic J. Statist.*, 1, 519-537.
- Wickham, H. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.

---

cache\_rfsrc\_datasets *Recreate the cached data sets for the ggRandomForests package*

---

## Description

Recreate the cached data sets for the `ggRandomForests` package

## Usage

```
cache_rfsrc_datasets(set = NA, save = TRUE, pth, ...)
```

## Arguments

<code>set</code>	Defaults to all sets (NA), however for individual sets specify one or more of <code>c("boston", "iris")</code>
<code>save</code>	Defaults to write files to the current data directory.
<code>pth</code>	the directory to store files.
<code>...</code>	extra arguments passed to <code>randomForestSRC</code> functions.

## Details

Constructing random forests are computationally expensive, and the `ggRandomForests` operates directly on `randomForestSRC` objects. We cache computationally intensive `randomForestSRC` objects to improve the `ggRandomForests` examples, diagnostics and vignettes run times. The set of precompiled `randomForestSRC` objects are stored in the package data subfolder, however version changes in the dependent packages may break some functionality. This function was created to help the package developer deal with those changes. We make the function available to end users to create objects for further experimentation.

There are five cached data set types: ’

- `rfsrc_data` - `rfsrc` objects.
- `varsel_data` - `var.select` minimal depth variable selection objects.
- `interaction_data` - `find.interaction` minimal depth, pairwise variable interaction matrices.
- `partial_data` - `plot.variable` objects (`partial=TRUE`) for partial variable dependence.
- `partial_coplot_data` - `plot.variable` objects (`partial=TRUE`) for partial variable dependence.

For the following data sets: #’

- `_iris` - The `iris` data set.
- `_boston` - The `boston` housing data set (MASS package).

## See Also

[iris](#) [airqmtcars](#) [Boston](#) [pbc](#) [veteran](#) [rfsrc\\_data](#) [varsel\\_data](#) [interaction\\_data](#) [partial\\_data](#) [partial\\_coplot\\_data](#)

---

calc\_auc

*Area Under the ROC Curve calculator*

---

## Description

Area Under the ROC Curve calculator

## Usage

```
calc_auc(x)
```

## Arguments

x [gg\\_roc](#) object

## Details

`calc_auc` uses the trapezoidal rule to calculate the area under the ROC curve.

This is a helper function for the [gg\\_roc](#) functions.

**Value**

AUC. 50% is random guessing, higher is better.

**See Also**

[calc\\_roc](#) [gg\\_roc](#)

[plot.gg\\_roc](#)

**Examples**

```
##
## Taken from the gg_roc example
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
data(rfsrc_iris)

## Not run:
gg_dta <- gg_roc(rfsrc_iris, which.outcome=1)

calc_auc(gg_dta)

## End(Not run)

gg_dta <- gg_roc(rfsrc_iris, which.outcome=2)

calc_auc(gg_dta)

## randomForest tests
rf_iris <- randomForest::randomForest(Species ~ ., data = iris)
gg_dta <- gg_roc(rfsrc_iris, which.outcome=2)

calc_auc(gg_dta)
```

---

calc\_roc.rfsrc

*Receiver Operator Characteristic calculator*

---

**Description**

Receiver Operator Characteristic calculator

**Usage**

```
## S3 method for class 'rfsrc'
calc_roc(object, dta, which.outcome = "all", oob = TRUE, ...)
```

## Arguments

object	rfsrc or predict.rfsrc object containing predicted response
dta	True response variable
which.outcome	If defined, only show ROC for this response.
oob	Use OOB estimates, the normal validation method (TRUE)
...	extra arguments passed to helper functions

## Details

For a randomForestSRC prediction and the actual response value, calculate the specificity (1-False Positive Rate) and sensitivity (True Positive Rate) of a predictor.

This is a helper function for the [gg\\_roc](#) functions, and not intended for use by the end user.

## Value

A gg\_roc object

## See Also

[calc\\_auc](#) [gg\\_roc](#)

[plot.gg\\_roc](#)

## Examples

```
## Taken from the gg_roc example
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
data(rfsrc_iris)

gg_dta <- calc_roc(rfsrc_iris, rfsrc_iris$yvar,
  which.outcome=1, oob=TRUE)
gg_dta <- calc_roc(rfsrc_iris, rfsrc_iris$yvar,
  which.outcome=1, oob=FALSE)

rf_iris <- randomForest(Species ~ ., data = iris)
gg_dta <- calc_roc(rf_iris, rf_iris$yvar,
  which.outcome=1)
gg_dta <- calc_roc(rf_iris, rf_iris$yvar,
  which.outcome=2)
```

---

combine.gg\_partial      *combine two gg\_partial objects*

---

### Description

The `combine.gg_partial` function assumes the two `gg_partial` objects were generated from the same `rfsrc` object. So, the function joins along the `gg_partial` list item names (one per partial plot variable). Further, we combine the two `gg_partial` objects along the group variable.

Hence, to join three `gg_partial` objects together (i.e. for three different time points from a survival random forest) would require two `combine.gg_partial` calls: One to join the first two `gg_partial` object, and one to append the third `gg_partial` object to the output from the first call. The second call will append a single `lbls` label to the `gg_partial` object.

### Usage

```
combine.gg_partial(x, y, lbls, ...)
```

### Arguments

<code>x</code>	<code>gg_partial</code> object
<code>y</code>	<code>gg_partial</code> object
<code>lbls</code>	vector of 2 strings to label the combined data.
<code>...</code>	not used

### Value

`gg_partial` or `gg_partial_list` based on class of `x` and `y`.

### Examples

```
## Not run:  
# We need to create this dataset  
cache_rfsrc_datasets(set="pbc")  
  
# Load a set of plot.variable partial plot data  
data(partial_pbc)  
  
# A list of 2 plot.variable objects  
length(partial_pbc)  
class(partial_pbc)  
  
class(partial_pbc[[1]])  
class(partial_pbc[[2]])  
  
# Create gg_partial objects  
ggPrtl.1 <- gg_partial(partial_pbc[[1]])  
ggPrtl.2 <- gg_partial(partial_pbc[[2]])
```



```

# Combine the objects to get multiple time curves
# along variables on a single figure.
ggpart <- combine.gg_partial(ggPrt1.1, ggPrt1.2,
                           lbls = c("1 year", "3 years"))

# Plot each figure separately
plot(ggpart)

# Get the continuous data for a panel of continuous plots.
ggcont <- ggpart
ggcont$edema <- ggcont$ascites <- ggcont$stage <- NULL
plot(ggcont, panel=TRUE)

# And the categorical for a panel of categorical plots.
nms <- colnames(sapply(ggcont, function(st) {st}))
for(ind in nms) {
  ggpart[[ind]] <- NULL
}
plot(ggpart, panel=TRUE)

## End(Not run)

```

---

gg\_error

*randomForest error rate data object*


---

## Description

Extract the cumulative (OOB) randomForestSRC error rate as a function of number of trees.

## Usage

```
gg_error(object, ...)
```

## Arguments

object	<a href="#">rfsrc</a> object.
...	optional arguments (not used).

## Details

The `gg_error` function simply returns the `rfsrc$err.rate` object as a data.frame, and assigns the class for connecting to the S3 `plot.gg_error` function.

## Value

`gg_error` data.frame with one column indicating the tree number, and the remaining columns from the `rfsrc$err.rate` return value.

## References

- Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.
- Ishwaran H. and Kogalur U.B. (2013). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.4.

## See Also

[plot.gg\\_error](#) rfsrc plot.rfsrc

## Examples

```
## Examples from RFSRC package...
## -----
## classification example
## -----
## ----- iris data
## You can build a randomForest
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
# ... or load a cached randomForestSRC object
data(rfsrc_iris, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_error(rfsrc_iris)

# Plot the gg_error object
plot(gg_dta)

## RandomForest example
rf_iris <- randomForest::randomForest(Species ~ ., data = iris)
gg_dta <- gg_error(rf_iris)
plot(gg_dta)

gg_dta <- gg_error(rf_iris, training=TRUE)
plot(gg_dta)
## -----
## Regression example
## -----
## Not run:
## ----- airq data
rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality, na.action = "na.impute")

# Get a data.frame containing error rates
gg_dta<- gg_error(rfsrc_airq)

# Plot the gg_error object
plot(gg_dta)

## End(Not run)

## ----- Boston data
```

```

data(rfsrc_boston, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_error(rfsrc_boston)

# Plot the gg_error object
plot(gg_dta)

## Not run:
## ----- mtcars data

# Get a data.frame containing error rates
gg_dta<- gg_error(rfsrc_mtcars)

# Plot the gg_error object
plot(gg_dta)

## End(Not run)

## -----
## Survival example
## -----
## Not run:
## ----- veteran data
## randomized trial of two treatment regimens for lung cancer
data(veteran, package = "randomForestSRC")
rfsrc_veteran <- rfsrc(Surv(time, status) ~ ., data = dta$veteran, ...)

gg_dta <- gg_error(rfsrc_veteran)
plot(gg_dta)

## ----- pbc data
# Load a cached randomForestSRC object
data(rfsrc_pbc, package="ggRandomForests")

gg_dta <- gg_error(rfsrc_pbc)
plot(gg_dta)

## End(Not run)

```

---

gg_interaction	<i>Minimal</i>	<i>Depth</i>	<i>Variable</i>	<i>Interaction</i>	<i>data</i>	<i>object</i>
	( <a href="#">find.interaction</a> ).					

---

## Description

Converts the matrix returned from [find.interaction](#) to a `data.frame` and add attributes for S3 identification. If passed a `rfsrc` object, `gg_interaction` first runs the [find.interaction](#) function with all optional arguments.

**Usage**

```
gg_interaction(object, ...)
```

**Arguments**

`object` a `rfsrc` object or the output from the `find.interaction` function call.  
`...` optional extra arguments passed to `find.interaction`.

**Value**

`gg_interaction` object

**References**

- Ishwaran H. (2007). Variable importance in binary regression trees and forests, *Electronic J. Statist.*, 1:519-537.
- Ishwaran H., Kogalur U.B., Gorodeski E.Z, Minn A.J. and Lauer M.S. (2010). High-dimensional variable selection for survival data. *J. Amer. Statist. Assoc.*, 105:205-217.
- Ishwaran H., Kogalur U.B., Chen X. and Minn A.J. (2011). Random survival forests for high-dimensional data. *Statist. Anal. Data Mining*, 4:115-132.

**See Also**

`rfsrc` `find.interaction` `max.subtree` `var.select` `vimp` `plot.gg_interaction`

**Examples**

```
## Examples from randomForestSRC package...
## -----
## find interactions, classification setting
## -----
## ----- iris data
## iris.obj <- rfsrc(Species ~., data = iris)
## TODO: VIMP interactions not handled yet...
## randomForestSRC::find.interaction(iris.obj, method = "vimp", nrep = 3)
## interaction_iris <- randomForestSRC::find.interaction(iris.obj)
data(interaction_iris, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_iris)

plot(gg_dta, xvar="Petal.Width")
plot(gg_dta, panel=TRUE)

## -----
## find interactions, regression setting
## -----
## Not run:
## ----- air quality data
## airq.obj <- rfsrc(Ozone ~ ., data = airquality)
##
## TODO: VIMP interactions not handled yet....
```

```

## randomForestSRC::find.interaction(airq.obj, method = "vimp", nrep = 3)
## interaction_airq <- randomForestSRC::find.interaction(airq.obj)
data(interaction_airq, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_airq)

plot(gg_dta, xvar="Temp")
plot(gg_dta, xvar="Solar.R")

plot(gg_dta, panel=TRUE)

## End(Not run)

## ----- Boston data
data(interaction_boston, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_boston)

plot(gg_dta, panel=TRUE)

## Not run:
## ----- mtcars data
data(interaction_mtcars, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_mtcars)

plot(gg_dta, panel=TRUE)

## End(Not run)
## Not run:
## -----
## find interactions, survival setting
## -----
## ----- pbc data
## data(pbc, package = "randomForestSRC")
## pbc.obj <- rfsrc(Surv(days,status) ~ ., pbc, nsplit = 10)
## interaction_pbc <- randomForestSRC::find.interaction(pbc.obj, nvar = 8)
data(interaction_pbc, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_pbc)

plot(gg_dta, xvar="bili")
plot(gg_dta, panel=TRUE)

## ----- veteran data
data(interaction_veteran, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_veteran)

plot(gg_dta, panel=TRUE)

## End(Not run)

```

**Description**

the `[randomForestSRC]{var.select}` function implements random forest variable selection using tree minimal depth methodology. The `gg_minimal_depth` function takes the output from `[randomForestSRC]{var.select}` and creates a `data.frame` formatted for the `plot.gg_minimal_depth` function.

**Usage**

```
gg_minimal_depth(object, ...)
```

**Arguments**

<code>object</code>	A <code>[randomForestSRC]{rfsrc}</code> object, <code>[randomForestSRC]{predict}</code> object or the list from the <code>[randomForestSRC]{var.select.rfsrc}</code> function.
<code>...</code>	optional arguments passed to the <code>[randomForestSRC]{var.select}</code> function if operating on an <code>[randomForestSRC]{rfsrc}</code> object.

**Value**

`gg_minimal_depth` object, A modified list of variables from the `[randomForestSRC]{var.select}` function, ordered by minimal depth rank.

**See Also**

`[randomForestSRC]{var.select}`  
[plot.gg\\_minimal\\_depth](#)

**Examples**

```
## Examples from RFSRC package...
## -----
## classification example
## -----
## ----- iris data
## You can build a randomForest
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
# varsel_iris <- randomForestSRC::var.select(rfsrc_iris)
# ... or load a cached randomForestSRC object
data(varsel_iris, package="ggRandomForests")

# Get a data.frame containing minimaldepth measures
gg_dta<- gg_minimal_depth(varsel_iris)

# Plot the gg_minimal_depth object
plot(gg_dta)

## -----
## Regression example
## -----
## Not run:
```

```

## ----- air quality data
# rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality, na.action = "na.impute")
# varsel_airq <- randomForestSRC::var.select(rfsrc_airq)
# ... or load a cached randomForestSRC object
data(varsel_airq, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_minimal_depth(varsel_airq)

# Plot the gg_minimal_depth object
plot(gg_dta)

## End(Not run)

## ----- Boston data
data(varsel_boston, package="ggRandomForests")

# Get a data.frame containing error rates
plot(gg_minimal_depth(varsel_boston))

## Not run:
## ----- mtcars data
data(varsel_mtcars, package="ggRandomForests")

# Get a data.frame containing error rates
plot.gg_minimal_depth(varsel_mtcars)

## End(Not run)

## -----
## Survival example
## -----
## Not run:
## ----- veteran data
## veteran data
## randomized trial of two treatment regimens for lung cancer
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time, status) ~ ., data = veteran, ntree = 100)
# varsel_veteran <- randomForestSRC::var.select(rfsrc_veteran)
# Load a cached randomForestSRC object
data(varsel_veteran, package="ggRandomForests")

gg_dta <- gg_minimal_depth(varsel_veteran)
plot(gg_dta)

## ----- pbc data
data(varsel_pbc, package="ggRandomForests")

gg_dta <- gg_minimal_depth(varsel_pbc)
plot(gg_dta)

## End(Not run)

```

---

gg_minimal_vimp	<i>Minimal depth vs VIMP comparison by variable rankings.</i>
-----------------	---

---

**Description**

Minimal depth vs VIMP comparison by variable rankings.

**Usage**

```
gg_minimal_vimp(object, ...)
```

**Arguments**

object	A <code>rfsrc</code> object, <code>predict.rfsrc</code> object or the list from the <code>var.select.rfsrc</code> function.
...	optional arguments passed to the <code>var.select</code> function if operating on an <code>rfsrc</code> object.

**Value**

gg\_minimal\_vimp comparison object.

**See Also**

[plot.gg\\_minimal\\_vimp](#) [var.select](#)

**Examples**

```
## Examples from RFSRC package...
## -----
## classification example
## -----
## ----- iris data
## You can build a randomForest
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
# varsel_iris <- randomForestSRC::var.select(rfsrc_iris)
# ... or load a cached randomForestSRC object
data(varsel_iris, package="ggRandomForests")

# Get a data.frame containing minimaldepth measures
gg_dta<- gg_minimal_vimp(varsel_iris)

# Plot the gg_minimal_depth object
plot(gg_dta)

## -----
## Regression example
## -----
## Not run:
```



```

## ----- air quality data
# rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality,
#                   na.action = "na.impute")
# varsel_airq <- randomForestSRC::var.select(rfsrc_airq)
# ... or load a cached randomForestSRC object
data(varsel_airq, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_minimal_vimp(varsel_airq)

# Plot the gg_minimal_vimp object
plot(gg_dta)

## End(Not run)

## ----- Boston data
data(varsel_boston, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_minimal_vimp(varsel_boston)

# Plot the gg_minimal_vimp object
plot(gg_dta)

## Not run:
## ----- mtcars data
data(varsel_mtcars, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_minimal_vimp(varsel_mtcars)

# Plot the gg_minimal_vimp object
plot(gg_dta)

## End(Not run)
## -----
## Survival example
## -----
## Not run:
## ----- veteran data
## randomized trial of two treatment regimens for lung cancer
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time, status) ~ ., data = veteran,
#                       ntree = 100)
# varsel_veteran <- randomForestSRC::var.select(rfsrc_veteran)
# Load a cached randomForestSRC object
data(varsel_veteran, package="ggRandomForests")

gg_dta <- gg_minimal_vimp(varsel_veteran)
plot(gg_dta)

## ----- pbc data
data(varsel_pbc, package="ggRandomForests")

```

```
gg_dta <- gg_minimal_vimp(varsel_pbc)
plot(gg_dta)

## End(Not run)
```

---

gg\_partial

*Partial variable dependence object*

---

## Description

The `plot.variable` function returns a list of either marginal variable dependence or partial variable dependence data from a `rfsrc` object. The `gg_partial` function formulates the `plot.variable` output for partial plots (where `partial=TRUE`) into a data object for creation of partial dependence plots using the `plot.gg_partial` function.

Partial variable dependence plots are the risk adjusted estimates of the specified response as a function of a single covariate, possibly subsetted on other covariates.

An option named `argument` can name a column for merging multiple plots together

## Usage

```
gg_partial(object, ...)
```

## Arguments

<code>object</code>	the partial variable dependence data object from <code>plot.variable</code> function
<code>...</code>	optional arguments

## Value

`gg_partial` object. A `data.frame` or list of `data.frames` corresponding the variables contained within the `plot.variable` output.

## References

Friedman, Jerome H. 2000. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics* 29: 1189-1232."

## See Also

[plot.gg\\_partial](#)  
[plot.variable](#)

**Examples**

```

## -----
## classification
## -----
## ----- iris data

## iris "Petal.Width" partial dependence plot
##
# rfsrc_iris <- rfsrc(Species ~., data = iris)
# partial_iris <- plot.variable(rfsrc_iris, xvar.names = "Petal.Width",
#                             partial=TRUE)
data(partial_iris, package="ggRandomForests")

gg_dta <- gg_partial(partial_iris)
plot(gg_dta)

## -----
## regression
## -----
## Not run:
## ----- air quality data
## airquality "Wind" partial dependence plot
##
# rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality)
# partial_airq <- plot.variable(rfsrc_airq, xvar.names = "Wind",
#                             partial=TRUE, show.plot=FALSE)
data(partial_airq, package="ggRandomForests")

gg_dta <- gg_partial(partial_airq)
plot(gg_dta)

gg_dta.m <- gg_dta[["Month"]]
plot(gg_dta.m, notch=TRUE)

gg_dta[["Month"]] <- NULL
plot(gg_dta, panel=TRUE)

## End(Not run)

## ----- Boston data
data(partial_boston, package="ggRandomForests")

gg_dta <- gg_partial(partial_boston)
plot(gg_dta, panel=TRUE)

## Not run:
## ----- mtcars data
data(partial_mtcars, package="ggRandomForests")
gg_dta <- gg_partial(partial_mtcars)

gg_dta.cat <- gg_dta
gg_dta.cat[["disp"]] <- gg_dta.cat[["wt"]] <- gg_dta.cat[["hp"]] <- NULL

```

```

gg_dta.cat[["drat"]] <- gg_dta.cat[["carb"]] <- gg_dta.cat[["qsec"]] <- NULL

plot(gg_dta.cat, panel=TRUE, notch=TRUE)

gg_dta[["cyl"]] <- gg_dta[["vs"]] <- gg_dta[["am"]] <- NULL
gg_dta[["gear"]] <- NULL
plot(gg_dta, panel=TRUE)

## End(Not run)

## -----
## survival examples
## -----
## Not run:
## ----- veteran data
## survival "age" partial variable dependence plot
##
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time,status)~., veteran, nsplit = 10,
#   ntree = 100)
#
## 30 day partial plot for age
# partial_veteran <- plot.variable(rfsrc_veteran, surv.type = "surv",
#   partial = TRUE, time=30,
#   xvar.names = "age",
#   show.plots=FALSE)
data(partial_veteran, package="ggRandomForests")

gg_dta <- gg_partial(partial_veteran[[1]])
plot(gg_dta)

gg_dta.cat <- gg_dta
gg_dta.cat[["celltype"]] <- gg_dta.cat[["trt"]] <- gg_dta.cat[["prior"]] <- NULL
plot(gg_dta, panel=TRUE)

gg_dta.cat[["karno"]] <- gg_dta.cat[["diagtime"]] <-
  gg_dta.cat[["age"]] <- NULL
plot(gg_dta.cat, panel=TRUE, notch=TRUE)

gg_dta <- lapply(partial_veteran, gg_partial)
gg_dta <- combine.gg_partial(gg_dta[[1]], gg_dta[[2]] )

plot(gg_dta[["karno"]])
plot(gg_dta[["celltype"]])

gg_dta.cat <- gg_dta
gg_dta.cat[["celltype"]] <- gg_dta.cat[["trt"]] <- gg_dta.cat[["prior"]] <- NULL
plot(gg_dta, panel=TRUE)

gg_dta.cat[["karno"]] <- gg_dta.cat[["diagtime"]] <-
  gg_dta.cat[["age"]] <- NULL
plot(gg_dta.cat, panel=TRUE, notch=TRUE)

```

```

## ----- pbc data
data("partial_pbc", package = "ggRandomForests")
data("varsel_pbc", package = "ggRandomForests")
xvar <- varsel_pbc$topvars

# Convert all partial plots to gg_partial objects
gg_dta <- lapply(partial_pbc, gg_partial)

# Combine the objects to get multiple time curves
# along variables on a single figure.
pbc_ggpart <- combine.gg_partial(gg_dta[[1]], gg_dta[[2]],
                               lbls = c("1 Year", "3 Years"))

summary(pbc_ggpart)
class(pbc_ggpart[["bili"]])

# Plot the highest ranked variable, by name.
#plot(pbc_ggpart[["bili"]])

# Create a temporary holder and remove the stage and edema data
ggpart <- pbc_ggpart
ggpart$edema <- NULL

# Panel plot the remainder.
plot(ggpart, panel = TRUE)

#plot(pbc_ggpart[["edema"]], panel=TRUE) #,
# notch = TRUE, alpha = .3, outlier.shape = NA)

## End(Not run)

```

---

```
gg_partial_coplot.rfsrc
```

*Data structures for stratified partial coplots*

---

## Description

Data structures for stratified partial coplots

## Usage

```

## S3 method for class 'rfsrc'
gg_partial_coplot(
  object,
  xvar,
  groups,
  surv_type = c("mort", "rel.freq", "surv", "years.lost", "cif", "chf"),
  time,
  ...
)

```

## Arguments

object	rfsrc object
xvar	list of partial plot variables
groups	vector of stratification variable.
surv_type	for survival random forests, c("mort", "rel.freq", "surv", "years.lost", "cif", "chf")
time	vector of time points for survival random forests partial plots.
...	extra arguments passed to <code>plot.variable</code> function

## Value

gg\_partial\_coplot object. An subclass of a `gg_partial_list` object

## Examples

```
## Not run:
# Load the forest
data(rfsrc_pbc, package="ggRandomForests")

# Create the variable plot.
ggvar <- gg_variable(rfsrc_pbc, time = 1)

# Find intervals with similar number of observations.
copper_cts <- quantile_pts(ggvar$copper, groups = 6, intervals = TRUE)

# Create the conditional groups and add to the gg_variable object
copper_grp <- cut(ggvar$copper, breaks = copper_cts)

## We would run this, but it's expensive
partial_coplot_pbc <- gg_partial_coplot(rfsrc_pbc, xvar = "bili",
                                       groups = copper_grp,
                                       surv_type = "surv",
                                       time = 1,
                                       show.plots = FALSE)

## so load the cached set
data(partial_coplot_pbc, package="ggRandomForests")

# Partial coplot
plot(partial_coplot_pbc) #, se = FALSE)

## End(Not run)
```

---

gg_rfsrc.rfsrc	<i>Predicted response data object</i>
----------------	---------------------------------------

---

**Description**

Extracts the predicted response values from the `rfsrc` object, and formats data for plotting the response using `plot.gg_rfsrc`.

**Usage**

```
## S3 method for class 'rfsrc'
gg_rfsrc(object, oob = TRUE, by, ...)
```

**Arguments**

object	<code>rfsrc</code> object
oob	boolean, should we return the oob prediction , or the full forest prediction.
by	stratifying variable in the training dataset, defaults to NULL
...	extra arguments

**Details**

`surv_type` ("surv", "chf", "mortality", "hazard") for survival forests  
`oob` boolean, should we return the oob prediction , or the full forest prediction.

**Value**

gg\_rfsrc object

**See Also**

[plot.gg\\_rfsrc](#) [rfsrc](#) [plot.rfsrc](#) [gg\\_survival](#)

**Examples**

```
## -----
## classification example
## -----
## ----- iris data
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
data(rfsrc_iris, package="ggRandomForests")
gg_dta<- gg_rfsrc(rfsrc_iris)

plot(gg_dta)

## -----
## Regression example
## -----
```

```

## Not run:
## ----- air quality data
# rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality, na.action = "na.impute")
data(rfsrc_airq, package="ggRandomForests")
gg_dta<- gg_rfsrc(rfsrc_airq)

plot(gg_dta)

## End(Not run)

## ----- Boston data
data(rfsrc_boston, package="ggRandomForests")
plot(gg_rfsrc(rfsrc_boston))

### randomForest example
data(Boston, package="MASS")
rf_boston <- randomForest::randomForest(medv ~ ., data = Boston)
plot(gg_rfsrc(rf_boston))

## Not run:
## ----- mtcars data
data(rfsrc_mtcars, package="ggRandomForests")
gg_dta<- gg_rfsrc(rfsrc_mtcars)

plot(gg_dta)

## End(Not run)
## -----
## Survival example
## -----
## Not run:
## ----- veteran data
## randomized trial of two treatment regimens for lung cancer
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time, status) ~ ., data = veteran, ntree = 100)
data(rfsrc_veteran, package = "ggRandomForests")
gg_dta <- gg_rfsrc(rfsrc_veteran)
plot(gg_dta)

gg_dta <- gg_rfsrc(rfsrc_veteran, conf.int=.95)
plot(gg_dta)

gg_dta <- gg_rfsrc(rfsrc_veteran, by="trt")
plot(gg_dta)

## ----- pbc data
## We don't run this because of bootstrap confidence limits
data(rfsrc_pbc, package = "ggRandomForests")

gg_dta <- gg_rfsrc(rfsrc_pbc)
plot(gg_dta)

```



```

gg_dta <- gg_rfsrc(rfsrc_pbc, conf.int=.95)
plot(gg_dta)

gg_dta <- gg_rfsrc(rfsrc_pbc, by="treatment")
plot(gg_dta)

## End(Not run)

```

---

gg_roc.rfsrc	<i>ROC (Receiver operator curve) data from a classification random forest.</i>
--------------	--

---

## Description

The sensitivity and specificity of a randomForest classification object.

## Usage

```

## S3 method for class 'rfsrc'
gg_roc(object, which.outcome, oob, ...)

```

## Arguments

object	an <code>rfsrc</code> classification object
which.outcome	select the classification outcome of interest.
oob	use oob estimates (default TRUE)
...	extra arguments (not used)

## Value

gg\_roc data.frame for plotting ROC curves.

## See Also

[plot.gg\\_roc rfsrc randomForest](#)

## Examples

```

## -----
## classification example
## -----
## ----- iris data
rfsrc_iris <- rfsrc(Species ~ ., data = iris)
data(rfsrc_iris, package="ggRandomForests")

# ROC for setosa

```

```
gg_dta <- gg_roc(rfsrc_iris, which.outcome=1)
plot(gg_dta)

# ROC for versicolor
gg_dta <- gg_roc(rfsrc_iris, which.outcome=2)
plot(gg_dta)

# ROC for virginica
gg_dta <- gg_roc(rfsrc_iris, which.outcome=3)
plot(gg_dta)

rf_iris <- randomForest::randomForest(Species ~ ., data = iris)

# ROC for setosa
gg_dta <- gg_roc(rf_iris, which.outcome=1)
plot(gg_dta)

# ROC for versicolor
gg_dta <- gg_roc(rf_iris, which.outcome=2)
plot(gg_dta)

# ROC for virginica
gg_dta <- gg_roc(rf_iris, which.outcome=3)
plot(gg_dta)
```

---

gg\_survival

*Nonparametric survival estimates.*

---

## Description

Nonparametric survival estimates.

## Usage

```
gg_survival(
  interval = NULL,
  censor = NULL,
  by = NULL,
  data,
  type = c("kaplan", "nelson"),
  ...
)
```

## Arguments

`interval` name of the interval variable in the training dataset.

censor	name of the censoring variable in the training dataset.
by	stratifying variable in the training dataset, defaults to NULL
data	name of the training data.frame
type	one of ("kaplan","nelson"), defaults to Kaplan-Meier
...	extra arguments passed to Kaplan or Nelson functions.

### Details

gg\_survival is a wrapper function for generating nonparametric survival estimates using either [nelson](#)-Aalen or [kaplan](#)-Meier estimates.

### Value

A gg\_survival object created using the non-parametric Kaplan-Meier or Nelson-Aalen estimators.

### See Also

[kaplan nelson](#)  
[plot.gg\\_survival](#)

### Examples

```
## Not run:
## ----- pbc data
data(pbc, package="randomForestSRC")
pbc$time <- pbc$days/364.25

# This is the same as kaplan
gg_dta <- gg_survival(interval="time", censor="status",
                     data=pbc)

plot(gg_dta, error="none")
plot(gg_dta)

# Stratified on treatment variable.
gg_dta <- gg_survival(interval="time", censor="status",
                     data=pbc, by="treatment")

plot(gg_dta, error="none")
plot(gg_dta)

# ...with smaller confidence limits.
gg_dta <- gg_survival(interval="time", censor="status",
                     data=pbc, by="treatment", conf.int=.68)

plot(gg_dta, error="lines")

## End(Not run)
```

---

gg_variable	<i>Marginal variable dependence data object.</i>
-------------	--

---

### Description

`plot.variable` generates a `data.frame` containing the marginal variable dependence or the partial variable dependence. The `gg_variable` function creates a `data.frame` of containing the full set of covariate data (predictor variables) and the predicted response for each observation. Marginal dependence figures are created using the `plot.gg_variable` function.

Optional arguments `time` point (or vector of points) of interest (for survival forests only) `time_labels` If more than one time is specified, a vector of time labels for differentiating the time points (for survival forests only) `oob` indicate if predicted results should include oob or full data set.

### Usage

```
gg_variable(object, ...)
```

### Arguments

<code>object</code>	a <code>rfsrc</code> object
<code>...</code>	optional arguments

### Details

The marginal variable dependence is determined by comparing relation between the predicted response from the `randomForest` and a covariate of interest.

The `gg_variable` function operates on a `rfsrc` object, or the output from the `plot.variable` function.

### Value

`gg_variable` object

### See Also

[plot.gg\\_variable](#)  
[plot.variable](#)

### Examples

```
## -----
## classification
## -----
## ----- iris data
## iris
#rfsrc_iris <- rfsrc(Species ~., data = iris)
data(rfsrc_iris, package="ggRandomForests")
```

```

gg_dta <- gg_variable(rfsrc_iris)
plot(gg_dta, xvar="Sepal.Width")
plot(gg_dta, xvar="Sepal.Length")

plot(gg_dta, xvar=rfsrc_iris$xvar.names,
      panel=TRUE) # , se=FALSE)

## -----
## regression
## -----
## Not run:
## ----- air quality data
rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality)
gg_dta <- gg_variable(rfsrc_airq)

# an ordinal variable
gg_dta[, "Month"] <- factor(gg_dta[, "Month"])

plot(gg_dta, xvar="Wind")
plot(gg_dta, xvar="Temp")
plot(gg_dta, xvar="Solar.R")

plot(gg_dta, xvar=c("Solar.R", "Wind", "Temp", "Day"), panel=TRUE)

plot(gg_dta, xvar="Month", notch=TRUE)

## End(Not run)
## Not run:
## ----- motor trend cars data
rfsrc_mtcars <- rfsrc(mpg ~ ., data = mtcars)

gg_dta <- gg_variable(rfsrc_mtcars)

# mtcars$cyl is an ordinal variable
gg_dta$cyl <- factor(gg_dta$cyl)
gg_dta$am <- factor(gg_dta$am)
gg_dta$vs <- factor(gg_dta$vs)
gg_dta$gear <- factor(gg_dta$gear)
gg_dta$carb <- factor(gg_dta$carb)

plot(gg_dta, xvar="cyl")

# Others are continuous
plot(gg_dta, xvar="disp")
plot(gg_dta, xvar="hp")
plot(gg_dta, xvar="wt")

# panels
plot(gg_dta, xvar=c("disp", "hp", "drat", "wt", "qsec"), panel=TRUE)
plot(gg_dta, xvar=c("cyl", "vs", "am", "gear", "carb"), panel=TRUE,
      notch=TRUE)

```

```

## End(Not run)
## ----- Boston data
data(Boston, package="MASS")

rf_boston <- randomForest::randomForest(medv~., data=Boston)
gg_dta <- gg_variable(rf_boston)
plot(gg_dta)

## -----
## survival examples
## -----
## Not run:
## ----- veteran data
## survival
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time,status)~., veteran, nsplit = 10,
#                       ntree = 100)
data(rfsrc_veteran, package="ggRandomForests")

# get the 1 year survival time.
gg_dta <- gg_variable(rfsrc_veteran, time=90)

# Generate variable dependence plots for age and diagtime
plot(gg_dta, xvar = "age")
plot(gg_dta, xvar = "diagtime", )

# Generate coplots
plot(gg_dta, xvar = c("age", "diagtime"), panel=TRUE, se=FALSE)

# If we want to compare survival at different time points, say 30, 90 day
# and 1 year
gg_dta <- gg_variable(rfsrc_veteran, time=c(30, 90, 365))

# Generate variable dependence plots for age and diagtime
plot(gg_dta, xvar = "age")

## End(Not run)
## ----- pbc data

```

---

gg\_vimp

*Variable Importance (VIMP) data object*


---

### Description

gg\_vimp Extracts the variable importance (VIMP) information from a a `rfsrc` object.

### Usage

```
gg_vimp(object, nvar, ...)
```

**Arguments**

object	A <a href="#">rfsrc</a> object or output from <a href="#">vimp</a>
nvar	argument to control the number of variables included in the output.
...	arguments passed to the <a href="#">vimp.rfsrc</a> function if the <a href="#">rfsrc</a> object does not contain importance information.

**Value**

gg\_vimp object. A data.frame of VIMP measures, in rank order.

**References**

Ishwaran H. (2007). Variable importance in binary regression trees and forests, *Electronic J. Statist.*, 1:519-537.

**See Also**

[plot.gg\\_vimp.rfsrc](#)

[vimp](#)

**Examples**

```
## -----
## classification example
## -----
## ----- iris data
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
data(rfsrc_iris, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_iris)
plot(gg_dta)

## -----
## regression example
## -----
## Not run:
## ----- air quality data
# rfsrc_airq <- rfsrc(Ozone ~ ., airquality)
data(rfsrc_airq, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_airq)
plot(gg_dta)

## End(Not run)

## ----- Boston data
data(rfsrc_boston, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_boston)
plot(gg_dta)

## ----- Boston data
data(Boston, package="MASS")
```

```

rf_boston <- randomForest::randomForest(medv~., Boston)
gg_dta <- gg_vimp(rf_boston)
plot(gg_dta)

## Not run:
## ----- mtcars data
data(rfsrc_mtcars, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_mtcars)
plot(gg_dta)

## End(Not run)
## -----
## survival example
## -----
## Not run:
## ----- veteran data
data(rfsrc_veteran, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_veteran)
plot(gg_dta)

## ----- pbc data
data(rfsrc_pbc, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_pbc)
plot(gg_dta)

# Restrict to only the top 10.
gg_dta <- gg_vimp(rfsrc_pbc, nvar=10)
plot(gg_dta)

## End(Not run)

```

---

interaction\_data

*Cached [find.interaction](#) matrix objects for examples, diagnostics and vignettes. Data sets storing [find.interaction](#) matrix objects corresponding to training data according to the following naming convention:*

- *interaction\_iris - from a `randomForestSR[C]` for the iris data set.*
  - *interaction\_boston - from a `randomForestS[R]C` for the Boston housing data set (MASS package).*
  - *interaction\_pbc - from a `randomForest[S]RC` for the pbc data set (randomForestSRC package)*
- 

## Description

Cached [find.interaction](#) matrix objects for examples, diagnostics and vignettes.

Data sets storing [find.interaction](#) matrix objects corresponding to training data according to the following naming convention:



- `interaction_iris` - from a `randomForestSRC` for the `iris` data set.
- `interaction_boston` - from a `randomForestSRC` for the Boston housing data set (MASS package).
- `interaction_pbc` - from a `randomForestSRC` for the `pbc` data set (`randomForestSRC` package)

## Format

`find.interaction` matrix

## Details

Constructing the minimal depth interaction matrices on `randomForestSRC` objects are computationally expensive. We cache `find.interaction` matrix objects to improve the `ggRandomForests` examples, diagnostics and vignettes run times. (see `cache_rfsrc_datasets` to rebuild a complete set of these data sets.)

For each data set listed, we build a `rfsrc` (see `rfsrc_data`), then calculate the minimal depth variable interaction table with `find.interaction`. Each data set is built with the `cache_rfsrc_datasets` with the `randomForestSRC` version listed in the `ggRandomForests` DESCRIPTION file.

- `interaction_iris` - The famous (Fisher's or Anderson's) `iris` data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. Build a classification random forest for predicting the species (`setosa`, `versicolor`, and `virginica`) on 5 variables (columns) and 150 observations (rows).
- `interaction_airq` - The `airquality` data set is from the New York State Department of Conservation (ozone data) and the National Weather Service (meteorological data) collected in New York, from May to September 1973. Build regression random forest for predicting Ozone on 5 covariates and 153 observations.
- `interaction_mtcars` - The `mtcars` data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). Build a regression random forest for predicting `mpg` on 10 covariates and 32 observations.
- `interaction_boston` - The Boston housing values in suburbs of Boston from the MASS package. Build a regression random forest for predicting `medv` (median home values) on 13 covariates and 506 observations.
- `interaction_pbc` - The `pbc` data from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo controlled trial of the drug D-penicillamine. 312 cases participated in the randomized trial and contain largely complete data. Data from the `randomForestSRC` package. Build a survival random forest for time-to-event death data with 17 covariates and 312 observations (remaining 106 observations are held out).
- `interaction_veteran` - Veteran's Administration randomized trial of two treatment regimens for lung cancer. Build a survival random forest for time-to-event death data with 6 covariates and 137 observations.

## References

- #————— randomForestSRC —————
- Ishwaran H. and Kogalur U.B. (2014). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.5.5.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R. R News 7(2), 25-31.
- Ishwaran H., Kogalur U.B., Blackstone E.H. and Lauer M.S. (2008). Random survival forests. Ann. Appl. Statist. 2(3), 841-860.
- #————— Boston data set —————
- Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. Regression Diagnostics. Identifying Influential Data and Sources of Collinearity. New York: Wiley.
- Harrison, D., and D.L. Rubinfeld. 1978. "Hedonic Prices and the Demand for Clean Air." J. Environ. Economics and Management 5: 81-102.
- #————— Iris data set —————
- Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) The New S Language. Wadsworth \& Brooks/Cole. (has iris3 as iris.)
- Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7, Part II, 179-188.
- Anderson, Edgar (1935). The irises of the Gaspé Peninsula, Bulletin of the American Iris Society, 59, 2-5.
- #————— pbc data set —————
- Fleming T.R and Harrington D.P., (1991) Counting Processes and Survival Analysis. New York: Wiley.
- T Therneau and P Grambsch (2000), Modeling Survival Data: Extending the Cox Model, Springer-Verlag, New York. ISBN: 0-387-98784-3.

## See Also

[iris](#) [Boston](#) [pbc](#) [find.interaction](#) [rfsrc\\_data](#) [cache\\_rfsrc\\_datasets](#) [gg\\_interaction](#) [plot.gg\\_interaction](#)

## Examples

```
## Not run:
#-----
# iris data - classification random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_iris, package="ggRandomForests")

# The interaction table
interaction_iris <- find.interaction(rfsrc_iris)

# plot the forest interaction table
gg_dta <- gg_interaction(interaction_iris)
plot(gg_dta, panel=TRUE)
```

```

#-----
# MASS::Boston data - regression random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_boston, package="ggRandomForests")

# The interaction table
interaction_boston <- find.interaction(rfsrc_boston)

# plot the forest interaction table
gg_dta <- gg_interaction(interaction_boston)
plot(gg_dta, panel=TRUE)

#-----
# randomForestSRC::pbc data - survival random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_pbc, package="ggRandomForests")

# The interaction table
interaction_pbc <- find.interaction(rfsrc_pbc)

# plot the forest interaction table
gg_dta <- gg_interaction(interaction_pbc)
plot(gg_dta, panel=TRUE)

## End(Not run)

```

---

kaplan

*nonparametric Kaplan-Meier estimates*


---

### Description

nonparametric Kaplan-Meier estimates

### Usage

```
kaplan(interval, censor, data, by = NULL, ...)
```

### Arguments

interval	name of the interval variable in the training dataset.
censor	name of the censoring variable in the training dataset.
data	name of the training set data. frame
by	stratifying variable in the training dataset, defaults to NULL
...	arguments passed to the survfit function

**Value**

`gg_survival` object

**See Also**

`gg_survival` `nelson` `plot.gg_survival`

**Examples**

```
## Not run:
# These get run through the gg_survival examples.
data(pbc, package="randomForestSRC")
pbc$time <- pbc$days/364.25

# This is the same as gg_survival
gg_dta <- kaplan(interval="time", censor="status",
                 data=pbc)

plot(gg_dta, error="none")
plot(gg_dta)

# Stratified on treatment variable.
gg_dta <- gg_survival(interval="time", censor="status",
                     data=pbc, by="treatment")

plot(gg_dta, error="none")
plot(gg_dta)

## End(Not run)
```

---

nelson

*nonparametric Nelson-Aalen estimates*

---

**Description**

nonparametric Nelson-Aalen estimates

**Usage**

```
nelson(interval, censor, data, by = NULL, weight = NULL, ...)
```

**Arguments**

<code>interval</code>	name of the interval variable in the training dataset.
<code>censor</code>	name of the censoring variable in the training dataset.
<code>data</code>	name of the survival training data.frame
<code>by</code>	stratifying variable in the training dataset, defaults to NULL
<code>weight</code>	for each observation (default=NULL)
<code>...</code>	arguments passed to the <code>survfit</code> function

**Value**

`gg_survival` object

**See Also**

`gg_survival` `nelson` `plot.gg_survival`

**Examples**

```
## Not run:
# These get run through the gg_survival examples.
data(pbc, package="randomForestSRC")
pbc$time <- pbc$days/364.25

# This is the same as gg_survival
gg_dta <- nelson(interval="time", censor="status",
                 data=pbc)

plot(gg_dta, error="none")
plot(gg_dta)

# Stratified on treatment variable.
gg_dta <- gg_survival(interval="time", censor="status",
                     data=pbc, by="treatment")

plot(gg_dta, error="none")
plot(gg_dta, error="lines")
plot(gg_dta)

gg_dta <- gg_survival(interval="time", censor="status",
                     data=pbc, by="treatment",
                     type="nelson")

plot(gg_dta, error="bars")
plot(gg_dta)

## End(Not run)
```

---

`partial_coplot_data` *Cached `plot.variable` objects for examples, diagnostics and vignettes. Data sets storing `rfsrc` objects corresponding to training data according to the following naming convention:*

- `partial_coplot_boston` - *randomForestS[R]C for the Boston housing data set (MASS package).*
-

## Description

Cached `plot.variable` objects for examples, diagnostics and vignettes.

Data sets storing `rfsrc` objects corresponding to training data according to the following naming convention:

- `partial_coplot_boston` - `randomForestS[R]C` for the Boston housing data set (MASS package).

## Format

List of `plot.variable` objects

## Details

Constructing random forests are computationally expensive. We cache `rfsrc` objects to improve the `ggRandomForests` examples, diagnostics and vignettes run times. (see `cache_rfsrc_datasets` to rebuild a complete set of these data sets.)

For each data set listed, we build a `rfsrc`. Tuning parameters used in each case are documented in the examples. Each data set is built with the `cache_rfsrc_datasets` with the `randomForestSRC` version listed in the `ggRandomForests` DESCRIPTION file.

- `partial_coplot_boston` - The Boston housing values in suburbs of Boston from the MASS package. Build a regression random forest for predicting `medv` (median home values) on 13 covariates and 506 observations.

## References

#————— `randomForestSRC` —————

Ishwaran H. and Kogalur U.B. (2014). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.5.5.

Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R. R News 7(2), 25-31.

Ishwaran H., Kogalur U.B., Blackstone E.H. and Lauer M.S. (2008). Random survival forests. Ann. Appl. Statist. 2(3), 841-860.

#————— Boston data set —————

Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. Regression Diagnostics. Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Harrison, D., and D.L. Rubinfeld. 1978. "Hedonic Prices and the Demand for Clean Air." J. Environ. Economics and Management 5: 81-102.

#————— `pbc` data set —————

Fleming T.R and Harrington D.P., (1991) Counting Processes and Survival Analysis. New York: Wiley.

T Therneau and P Grambsch (2000), Modeling Survival Data: Extending the Cox Model, Springer-Verlag, New York. ISBN: 0-387-98784-3.

**See Also**

[Boston.plot.variable cache\\_rfsrc\\_datasets](#)

**Examples**

```
## Not run:
#-----
# MASS::Boston data - regression random forest
#-----
data(Boston_rfsrc, package="ggRandomForests")

# Cut the codependent variable
rm_pts <- cut_distribution(rfsrc_boston$xvar$rm, groups=6)
rm_grp <- cut(rfsrc_boston$xvar$rm, breaks=rm_pts)

# plot.variable for lstat on subsets of rm (this will take some time.)
partial_coplot_boston <- gg_partial_coplot(rfsrc_boston, xvar="lstat",
                                           groups=rm_grp,
                                           show.plots=FALSE)

## End(Not run)
```

---

partial\_data

*Cached [plot.variable](#) objects for examples, diagnostics and vignettes. Data sets storing [plot.variable](#) objects corresponding to training data according to the following naming convention:*

- partial\_iris - from a randomForestSR[C] for the iris data set.
  - partial\_boston - from a randomForestS[R]C for the Boston housing data set (MASS package).
  - partial\_pbc - from a randomForest[S]RC for the pbc data set (randomForestSRC package)
- 

**Description**

Cached [plot.variable](#) objects for examples, diagnostics and vignettes.

Data sets storing [plot.variable](#) objects corresponding to training data according to the following naming convention:

- partial\_iris - from a randomForestSR[C] for the iris data set.
- partial\_boston - from a randomForestS[R]C for the Boston housing data set (MASS package).
- partial\_pbc - from a randomForest[S]RC for the pbc data set (randomForestSRC package)

**Format**`plot.variable`**Details**

Constructing partial plot data with the `randomForestSRC::plot.variable` function are computationally expensive. We cache `plot.variable` objects to improve the `ggRandomForests` examples, diagnostics and vignettes run times. (see `cache_rfsrc_datasets` to rebuild a complete set of these data sets.)

For each data set listed, we build a `rfsrc` (see `rfsrc_data`), then calculate the partial plot data with `plot.variable` function, setting `partial=TRUE`. Each data set is built with the `cache_rfsrc_datasets` with the `randomForestSRC` version listed in the `ggRandomForests` DESCRIPTION file.

- `partial_iris` - The famous (Fisher's or Anderson's) `iris` data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. Build a classification random forest for predicting the species (`setosa`, `versicolor`, and `virginica`) on 5 variables (columns) and 150 observations (rows).
- `partial_boston` - The Boston housing values in suburbs of Boston from the `MASS` package. Build a regression random forest for predicting `medv` (median home values) on 13 covariates and 506 observations.
- `partial_pbc` - The `pbc` data from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo controlled trial of the drug D-penicillamine. 312 cases participated in the randomized trial and contain largely complete data. Data from the `randomForestSRC` package. Build a survival random forest for time-to-event death data with 17 covariates and 312 observations (remaining 106 observations are held out).

**References**

#————— `randomForestSRC` —————

Ishwaran H. and Kogalur U.B. (2014). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.5.5.

Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R. *R News* 7(2), 25-31.

Ishwaran H., Kogalur U.B., Blackstone E.H. and Lauer M.S. (2008). Random survival forests. *Ann. Appl. Statist.* 2(3), 841-860.

#————— Boston data set —————

Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Harrison, D., and D.L. Rubinfeld. 1978. "Hedonic Prices and the Demand for Clean Air." *J. Environ. Economics and Management* 5: 81-102.

#————— Iris data set —————

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole. (has `iris3` as iris.)



Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, Part II, 179-188.

Anderson, Edgar (1935). The irises of the Gaspe Peninsula, *Bulletin of the American Iris Society*, 59, 2-5.

#————— pbc data set —————

Fleming T.R and Harrington D.P., (1991) *Counting Processes and Survival Analysis*. New York: Wiley.

T Therneau and P Grambsch (2000), *Modeling Survival Data: Extending the Cox Model*, Springer-Verlag, New York. ISBN: 0-387-98784-3.

### See Also

[iris MASS::Boston](#) [pbc plot.variable](#) [rfsrc\\_data](#) [cache\\_rfsrc\\_datasets](#) [gg\\_partial](#) [plot.gg\\_partial](#)

### Examples

```
## Not run:
#-----
# iris data - classification random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_iris, package="ggRandomForests")

# The plot.variable call
partial_iris <- plot.variable(rfsrc_iris,
                             partial=TRUE, show.plots=FALSE)

# plot the forest partial plots
gg_dta <- gg_partial(partial_iris)
plot(gg_dta, panel=TRUE)

#-----
# MASS::Boston data - regression random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_boston, package="ggRandomForests")

# The plot.variable call
partial_boston <- plot.variable(rfsrc_boston,
                               partial=TRUE, show.plots = FALSE )

# plot the forest partial plots
gg_dta <- gg_partial(partial_boston)
plot(gg_dta, panel=TRUE)

#-----
# randomForestSRC::pbc data - survival random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_pbc, package="ggRandomForests")
```

```

# The plot.variable call -
# survival requires a time point specification.
# for the pbc data, we want 1, 3 and 5 year survival.
partial_pbc <- lapply(c(1,3,5), function(tm) {
  plot.variable(rfsrc_pbc, surv.type = "surv",
               time = tm,
               xvar.names = xvar,
               partial = TRUE,
               show.plots = FALSE)
})

# plot the forest partial plots
gg_dta <- gg_partial(partial_pbc)
plot(gg_dta)

## End(Not run)

```

---

partial\_surface\_data *Cached `plot.variable` objects for examples, diagnostics and vignettes. Data sets storing `plot.variable` objects corresponding to training data according to the following naming convention:*

- partial\_boston\_surf - *from a randomForestS[R]C for the Boston housing data set (MASS package).*
  - partial\_pbc\_surf - *from a randomForest[S]RC for the pbc data set (randomForestSRC package)*
  - partial\_pbc\_time - *from a randomForest[S]RC for the pbc data set (randomForestSRC package)*
- 

## Description

Cached `plot.variable` objects for examples, diagnostics and vignettes.

Data sets storing `plot.variable` objects corresponding to training data according to the following naming convention:

- partial\_boston\_surf - *from a randomForestS[R]C for the Boston housing data set (MASS package).*
- partial\_pbc\_surf - *from a randomForest[S]RC for the pbc data set (randomForestSRC package)*
- partial\_pbc\_time - *from a randomForest[S]RC for the pbc data set (randomForestSRC package)*

## Format

list of `plot.variable` objects

## Details

Constructing partial plot data with the `randomForestSRC::plot.variable` function are computationally expensive. We cache `plot.variable` objects to improve the `ggRandomForests` examples, diagnostics and vignettes run times. (see `cache_rfsrc_datasets` to rebuild a complete set of these data sets.)

For each data set listed, we build a `rfsrc` (see `rfsrc_data`), then calculate the partial plot data with `plot.variable` function, setting `partial=TRUE`. Each data set is built with the `cache_rfsrc_datasets` with the `randomForestSRC` version listed in the `ggRandomForests` DESCRIPTION file.

- `partial_boston` - The Boston housing values in suburbs of Boston from the MASS package. Build a regression random forest for predicting `medv` (median home values) on 13 covariates and 506 observations.
- `partial_pbc` - The `pbc` data from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo controlled trial of the drug D-penicillamine. 312 cases participated in the randomized trial and contain largely complete data. Data from the `randomForestSRC` package. Build a survival random forest for time-to-event death data with 17 covariates and 312 observations (remaining 106 observations are held out).

## References

#————— `randomForestSRC` —————

Ishwaran H. and Kogalur U.B. (2014). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.5.5.

Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R. R News 7(2), 25-31.

Ishwaran H., Kogalur U.B., Blackstone E.H. and Lauer M.S. (2008). Random survival forests. Ann. Appl. Statist. 2(3), 841-860.

#————— Boston data set —————

Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. Regression Diagnostics. Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Harrison, D., and D.L. Rubinfeld. 1978. "Hedonic Prices and the Demand for Clean Air." J. Environ. Economics and Management 5: 81-102.

#————— `pbc` data set —————

Fleming T.R and Harrington D.P., (1991) Counting Processes and Survival Analysis. New York: Wiley.

T Therneau and P Grambsch (2000), Modeling Survival Data: Extending the Cox Model, Springer-Verlag, New York. ISBN: 0-387-98784-3.

## See Also

[Boston](#) [pbc](#) [plot.variable](#) [rfsrc\\_data](#) [cache\\_rfsrc\\_datasets](#) [gg\\_partial](#) [plot.gg\\_partial](#)

**Examples**

```

## Not run:
#-----
# MASS::Boston data - regression random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_boston, package="ggRandomForests")

# The plot.variable call
partial_boston <- plot.variable(rfsrc_boston,
                               partial=TRUE, show.plots = FALSE )

# plot the forest partial plots
gg_dta <- gg_partial(partial_boston)
plot(gg_dta, panel=TRUE)

#-----
# randomForestSRC::pbc data - survival random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_pbc, package="ggRandomForests")

# Restrict the time of interest to less than 5 years.
time_pts <- rfsrc_pbc$time.interest[which(rfsrc_pbc$time.interest<=5)]

# Find the 50 points in time, evenly space along the distribution of
# event times for a series of partial dependence curves
time_cts <-quantile_pts(time_pts, groups = 50)

# Generate the gg_partial_coplot data object
system.time(partial_pbc_time <- lapply(time_cts, function(ct) {
  plot.variable(rfsrc_pbc, xvar = "bili", time = ct,
               npts = 50, show.plots = FALSE,
               partial = TRUE, surv.type="surv")
}))
# user system elapsed
# 2561.313 81.446 2641.707

# Find the quantile points to create 50 cut points
alb_partial_pts <-quantile_pts(rfsrc_pbc$xvar$albumin, groups = 50)

system.time(partial_pbc_surf <- lapply(alb_partial_pts, function(ct) {
  rfsrc_pbc$xvar$albumin <- ct
  plot.variable(rfsrc_pbc, xvar = "bili", time = 1,
               npts = 50, show.plots = FALSE,
               partial = TRUE, surv.type="surv")
}))
# user system elapsed
# 2547.482 91.978 2671.870

## End(Not run)

```

---

plot.gg\_error                      *Plot a gg\_error object*

---

### Description

A plot of the cumulative OOB error rates of the random forest as a function of number of trees.

### Usage

```
## S3 method for class 'gg_error'  
plot(x, ...)
```

### Arguments

x                      gg\_error object created from a [rfsrc](#) object  
...                    extra arguments passed to ggplot functions

### Details

The gg\_error plot is used to track the convergence of the randomForest. This figure is a reproduction of the error plot from the [plot.rfsrc](#) function.

### Value

ggplot object

### References

- Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.
- Ishwaran H. and Kogalur U.B. (2013). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.4.

### See Also

[gg\\_error](#) [rfsrc](#) [plot.rfsrc](#)

**Examples**

```

## Not run:
## Examples from RFSRC package...
## -----
## classification example
## -----
## ----- iris data
## You can build a randomForest
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
# ... or load a cached randomForestSRC object
data(rfsrc_iris, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_error(rfsrc_iris)

# Plot the gg_error object
plot(gg_dta)

## -----
## Regression example
## -----
## ----- airq data
# rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality, na.action = "na.impute")
# ... or load a cached randomForestSRC object
data(rfsrc_airq, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_error(rfsrc_airq)

# Plot the gg_error object
plot(gg_dta)

## ----- Boston data
data(rfsrc_boston, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_error(rfsrc_boston)

# Plot the gg_error object
plot(gg_dta)

## ----- mtcars data
data(rfsrc_mtcars, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_error(rfsrc_mtcars)

# Plot the gg_error object
plot(gg_dta)

## -----
## Survival example

```

```
## -----
## ----- veteran data
## randomized trial of two treatment regimens for lung cancer
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time, status) ~ ., data = veteran, ntree = 100)

# Load a cached randomForestSRC object
data(rfsrc_veteran, package="ggRandomForests")

gg_dta <- gg_error(rfsrc_veteran)
plot(gg_dta)

## ----- pbc data
# Load a cached randomForestSRC object
data(rfsrc_pbc, package="ggRandomForests")

gg_dta <- gg_error(rfsrc_pbc)
plot(gg_dta)

## End(Not run)
```

---

plot.gg\_interaction *plot.gg\_interaction* Plot a [gg\\_interaction](#) object,

---

## Description

plot.gg\_interaction Plot a [gg\\_interaction](#) object,

## Usage

```
## S3 method for class 'gg_interaction'
plot(x, xvar, lbls, ...)
```

## Arguments

x	gg_interaction object created from a <a href="#">rfsrc</a> object
xvar	variable (or list of variables) of interest.
lbls	A vector of alternative variable names.
...	arguments passed to the <a href="#">gg_interaction</a> function.

## Value

ggplot object

## References

Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.  
 Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.  
 Ishwaran H. and Kogalur U.B. (2013). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.4.

**See Also**

[rfsrc](#) [find.interaction](#) [max.subtree](#) [var.select](#) [vimp](#) [plot.gg\\_interaction](#)

**Examples**

```
## Not run:
## Examples from randomForestSRC package...
## -----
## find interactions, classification setting
## -----
## ----- iris data
## iris.obj <- rfsrc(Species ~., data = iris)
## TODO: VIMP interactions not handled yet...
## find.interaction(iris.obj, method = "vimp", nrep = 3)
## interaction_iris <- find.interaction(iris.obj)
data(interaction_iris, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_iris)

plot(gg_dta, xvar="Petal.Width")
plot(gg_dta, xvar="Petal.Length")
plot(gg_dta, panel=TRUE)

## -----
## find interactions, regression setting
## -----
## ----- air quality data
## airq.obj <- rfsrc(Ozone ~ ., data = airquality)
##
## TODO: VIMP interactions not handled yet...
## find.interaction(airq.obj, method = "vimp", nrep = 3)
## interaction_airq <- find.interaction(airq.obj)
data(interaction_airq, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_airq)

plot(gg_dta, xvar="Temp")
plot(gg_dta, xvar="Solar.R")
plot(gg_dta, panel=TRUE)

## ----- Boston data
data(interaction_boston, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_boston)

plot(gg_dta, panel=TRUE)

## ----- mtcars data
data(interaction_mtcars, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_mtcars)

plot(gg_dta, panel=TRUE)

## -----
## find interactions, survival setting
```



```
## -----
## ----- pbc data
## data(pbc, package = "randomForestSRC")
## pbc.obj <- rfsrc(Surv(days,status) ~ ., pbc, nsplit = 10)
## interaction_pbc <- find.interaction(pbc.obj, nvar = 8)
data(interaction_pbc, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_pbc)

plot(gg_dta, xvar="bili")
plot(gg_dta, xvar="copper")
plot(gg_dta, panel=TRUE)

## ----- veteran data
data(interaction_veteran, package="ggRandomForests")
gg_dta <- gg_interaction(interaction_veteran)

plot(gg_dta, panel=TRUE)

## End(Not run)
```

---

plot.gg\_minimal\_depth *Plot a [gg\\_minimal\\_depth](#) object for random forest variable ranking.*

---

## Description

Plot a [gg\\_minimal\\_depth](#) object for random forest variable ranking.

## Usage

```
## S3 method for class 'gg_minimal_depth'
plot(x, selection = FALSE, type = c("named", "rank"), lbls, ...)
```

## Arguments

x	<a href="#">gg_minimal_depth</a> object created from a <a href="#">rfsrc</a> object
selection	should we restrict the plot to only include variables selected by the minimal depth criteria (boolean).
type	select type of y axis labels c("named","rank")
lbls	a vector of alternative variable names.
...	optional arguments passed to <a href="#">gg_minimal_depth</a>

## Value

ggplot object

## References

- Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.
- Ishwaran H. and Kogalur U.B. (2014). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.5.

## See Also

[var.select](#) [gg\\_minimal\\_depth](#)

## Examples

```
## Not run:
## Examples from RFSRC package...
## -----
## classification example
## -----
## ----- iris data
## You can build a randomForest
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
# varsel_iris <- var.select(rfsrc_iris)
# ... or load a cached randomForestSRC object
data(varsel_iris, package="ggRandomForests")

# Get a data.frame containing minimaldepth measures
gg_dta<- gg_minimal_depth(varsel_iris)

# Plot the gg_minimal_depth object
plot(gg_dta)

## -----
## Regression example
## -----
## ----- air quality data
# rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality, na.action = "na.impute")
# varsel_airq <- var.select(rfsrc_airq)
# ... or load a cached randomForestSRC object
data(varsel_airq, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_minimal_depth(varsel_airq)

# Plot the gg_minimal_depth object
plot(gg_dta)

## ----- Boston data
data(varsel_boston, package="ggRandomForests")

# Get a data.frame containing error rates
plot(gg_minimal_depth(varsel_boston))
```

```
## ----- mtcars data
data(varsel_mtcars, package="ggRandomForests")

# Get a data.frame containing error rates
plot.gg_minimal_depth(varsel_mtcars)

## -----
## Survival example
## -----
## ----- veteran data
## veteran data
## randomized trial of two treatment regimens for lung cancer
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time, status) ~ ., data = veteran, ntree = 100)
# varsel_veteran <- var.select(rfsrc_veteran)
# Load a cached randomForestSRC object
data(varsel_veteran, package="ggRandomForests")

gg_dta <- gg_minimal_depth(varsel_veteran)
plot(gg_dta)

## ----- pbc data
data(varsel_pbc, package="ggRandomForests")

gg_dta <- gg_minimal_depth(varsel_pbc)
plot(gg_dta)

## End(Not run)
```

---

plot.gg\_minimal\_vimp *Plot a [gg\\_minimal\\_vimp](#) object for comparing the Minimal Depth and VIMP variable rankings.*

---

### Description

Plot a [gg\\_minimal\\_vimp](#) object for comparing the Minimal Depth and VIMP variable rankings.

### Usage

```
## S3 method for class 'gg_minimal_vimp'
plot(x, nvar, lbls, ...)
```

### Arguments

x	<a href="#">gg_minimal_depth</a> object created from a <a href="#">var.select</a> object
nvar	should the figure be restricted to a subset of the points.
lbls	a vector of alternative variable names.
...	optional arguments (not used)

**Value**

ggplot object

**See Also**

[gg\\_minimal\\_vimp var.select](#)

**Examples**

```
## Not run:
## Examples from RFSRC package...
## -----
## classification example
## -----
## ----- iris data
## You can build a randomForest
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
# varsel_iris <- var.select(rfsrc_iris)
# ... or load a cached randomForestSRC object
data(varsel_iris, package="ggRandomForests")

# Get a data.frame containing minimaldepth measures
gg_dta<- gg_minimal_vimp(varsel_iris)

# Plot the gg_minimal_depth object
plot(gg_dta)

## -----
## Regression example
## -----
## ----- air quality data
rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality, na.action = "na.impute")
varsel_airq <- var.select(rfsrc_airq)

# Get a data.frame containing error rates
gg_dta<- gg_minimal_vimp(varsel_airq)

# Plot the gg_minimal_vimp object
plot(gg_dta)

## ----- Boston data
data(varsel_boston, package="ggRandomForests")

# Get a data.frame containing error rates
gg_dta<- gg_minimal_vimp(varsel_boston)

# Plot the gg_minimal_vimp object
plot(gg_dta)

## ----- mtcars data
data(varsel_mtcars, package="ggRandomForests")
```

```

# Get a data.frame containing error rates
gg_dta<- gg_minimal_vimp(varsel_mtcars)

# Plot the gg_minimal_vimp object
plot(gg_dta)

## -----
## Survival example
## -----
## ----- veteran data
## randomized trial of two treatment regimens for lung cancer
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time, status) ~ ., data = veteran, ntree = 100)
# varsel_veteran <- var.select(rfsrc_veteran)
# Load a cached randomForestSRC object
data(varsel_veteran, package="ggRandomForests")

gg_dta <- gg_minimal_vimp(varsel_veteran)
plot(gg_dta)

## ----- pbc data
data(varsel_pbc, package="ggRandomForests")

gg_dta <- gg_minimal_vimp(varsel_pbc)
plot(gg_dta)

## End(Not run)

```

---

plot.gg\_partial

*Partial variable dependence plot, operates on a [gg\\_partial](#) object.*


---

## Description

Generate a risk adjusted (partial) variable dependence plot. The function plots the `rfsrc` response variable (y-axis) against the covariate of interest (specified when creating the `gg_partial` object).

## Usage

```

## S3 method for class 'gg_partial'
plot(x, points = TRUE, error = c("none", "shade", "bars", "lines"), ...)

```

## Arguments

<code>x</code>	<code>gg_partial</code> object created from a <code>rfsrc</code> forest object
<code>points</code>	plot points (boolean) or a smooth line.
<code>error</code>	"shade", "bars", "lines" or "none"
<code>...</code>	extra arguments passed to <code>ggplot2</code> functions.

**Value**

ggplot object

**References**

Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.

Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.

Ishwaran H. and Kogalur U.B. (2013). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.4.

**See Also**

[plot.variable](#) [gg\\_partial](#) [plot.gg\\_partial\\_list](#) [gg\\_variable](#) [plot.gg\\_variable](#)

**Examples**

```
## Not run:
## -----
## classification
## -----
## ----- iris data

## iris "Petal.Width" partial dependence plot
##
# rfsrc_iris <- rfsrc(Species ~., data = iris)
# partial_iris <- plot.variable(rfsrc_iris, xvar.names = "Petal.Width",
#                             partial=TRUE)
data(partial_iris, package="ggRandomForests")

gg_dta <- gg_partial(partial_iris)
plot(gg_dta)

## -----
## regression
## -----
## ----- air quality data
## airquality "Wind" partial dependence plot
##
# rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality)
# partial_airq <- plot.variable(rfsrc_airq, xvar.names = "Wind",
#                             partial=TRUE, show.plot=FALSE)
data(partial_airq, package="ggRandomForests")

gg_dta <- gg_partial(partial_airq)
plot(gg_dta)

gg_dta.m <- gg_dta[["Month"]]
plot(gg_dta.m, notch=TRUE)

gg_dta[["Month"]] <- NULL
plot(gg_dta, panel=TRUE)
```

```

## ----- Boston data
data(partial_boston, package="ggRandomForests")

gg_dta <- gg_partial(partial_boston)
plot(gg_dta)
plot(gg_dta, panel=TRUE)

## ----- mtcars data
data(partial_mtcars, package="ggRandomForests")

gg_dta <- gg_partial(partial_mtcars)

plot(gg_dta)

gg_dta.cat <- gg_dta
gg_dta.cat[["disp"]] <- gg_dta.cat[["wt"]] <- gg_dta.cat[["hp"]] <- NULL
gg_dta.cat[["drat"]] <- gg_dta.cat[["carb"]] <-
  gg_dta.cat[["qsec"]] <- NULL

plot(gg_dta.cat, panel=TRUE)

gg_dta[["cyl"]] <- gg_dta[["vs"]] <- gg_dta[["am"]] <- NULL
gg_dta[["gear"]] <- NULL
plot(gg_dta, panel=TRUE)

## -----
## survival examples
## -----
## ----- veteran data
## survival "age" partial variable dependence plot
##
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time,status)~., veteran, nsplit = 10,
#                       ntree = 100)
#
## 30 day partial plot for age
# partial_veteran <- plot.variable(rfsrc_veteran, surv.type = "surv",
#                                 partial = TRUE, time=30,
#                                 xvar.names = "age",
#                                 show.plots=FALSE)
data(partial_veteran, package="ggRandomForests")

gg_dta <- gg_partial(partial_veteran[[1]])
plot(gg_dta)

gg_dta.cat <- gg_dta
gg_dta.cat[["celltype"]] <- gg_dta.cat[["trt"]] <- gg_dta.cat[["prior"]] <- NULL
plot(gg_dta, panel=TRUE)

gg_dta.cat[["karno"]] <- gg_dta.cat[["diagtime"]] <-
  gg_dta.cat[["age"]] <- NULL
plot(gg_dta.cat, panel=TRUE, notch=TRUE)

```

```

gg_dta <- lapply(partial_veteran, gg_partial)
length(gg_dta)
gg_dta <- combine.gg_partial(gg_dta[[1]], gg_dta[[2]] )

plot(gg_dta[["karno"]])
plot(gg_dta[["celltype"]])

gg_dta.cat <- gg_dta
gg_dta[["celltype"]] <- gg_dta[["trt"]] <- gg_dta[["prior"]] <- NULL
plot(gg_dta, panel=TRUE)

gg_dta.cat[["karno"]] <- gg_dta.cat[["diagtime"]] <-
  gg_dta.cat[["age"]] <- NULL
plot(gg_dta.cat, panel=TRUE, notch=TRUE)

## ----- pbc data

## End(Not run)

```

---

plot.gg\_partial\_list *Partial variable dependence plot, operates on a gg\_partial\_list object.*

---

## Description

Generate a risk adjusted (partial) variable dependence plot. The function plots the `rfsrc` response variable (y-axis) against the covariate of interest (specified when creating the `gg_partial_list` object).

## Usage

```

## S3 method for class 'gg_partial_list'
plot(x, points = TRUE, panel = FALSE, ...)

```

## Arguments

<code>x</code>	gg_partial_list object created from a <a href="#">gg_partial</a> forest object
<code>points</code>	plot points (boolean) or a smooth line.
<code>panel</code>	should the entire list be plotted together?
<code>...</code>	extra arguments

## Value

list of ggplot objects, or a single faceted ggplot object



## References

- Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.
- Ishwaran H. and Kogalur U.B. (2013). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.4.

## See Also

[plot.variable](#) [gg\\_partial](#) [plot.gg\\_partial](#) [gg\\_variable](#) [plot.gg\\_variable](#)

## Examples

```
## Not run:
## -----
## classification
## -----
## ----- iris data

## iris "Petal.Width" partial dependence plot
##
# rfsrc_iris <- rfsrc(Species ~., data = iris)
# partial_iris <- plot.variable(rfsrc_iris, xvar.names = "Petal.Width",
#                             partial=TRUE)
data(partial_iris, package="ggRandomForests")

gg_dta <- gg_partial(partial_iris)
plot(gg_dta)

## -----
## regression
## -----
## ----- air quality data
## airquality "Wind" partial dependence plot
##
# rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality)
# partial_airq <- plot.variable(rfsrc_airq, xvar.names = "Wind",
#                              partial=TRUE, show.plot=FALSE)
data(partial_airq, package="ggRandomForests")

gg_dta <- gg_partial(partial_airq)
plot(gg_dta)

gg_dta.m <- gg_dta[["Month"]]
plot(gg_dta.m, notch=TRUE)

gg_dta[["Month"]] <- NULL
plot(gg_dta, panel=TRUE)

## ----- Boston data
data(partial_boston, package="ggRandomForests")
```

```

gg_dta <- gg_partial(partial_boston)
plot(gg_dta)
plot(gg_dta, panel=TRUE)

## ----- mtcars data
data(partial_mtcars, package="ggRandomForests")

gg_dta <- gg_partial(partial_mtcars)

plot(gg_dta)

gg_dta.cat <- gg_dta
gg_dta.cat[["disp"]] <- gg_dta.cat[["wt"]] <- gg_dta.cat[["hp"]] <- NULL
gg_dta.cat[["drat"]] <- gg_dta.cat[["carb"]] <- gg_dta.cat[["qsec"]] <- NULL

plot(gg_dta.cat, panel=TRUE)

gg_dta[["cyl"]] <- gg_dta[["vs"]] <- gg_dta[["am"]] <- NULL
gg_dta[["gear"]] <- NULL
plot(gg_dta, panel=TRUE)

## -----
## survival examples
## -----
## ----- veteran data
## survival "age" partial variable dependence plot
##
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time,status)~., veteran, nsplit = 10,
#                       ntree = 100)
#
## 30 day partial plot for age
# partial_veteran <- plot.variable(rfsrc_veteran, surv.type = "surv",
#                                 partial = TRUE, time=30,
#                                 xvar.names = "age",
#                                 show.plots=FALSE)
data(partial_veteran, package="ggRandomForests")

gg_dta <- gg_partial(partial_veteran[[1]])
plot(gg_dta)

gg_dta.cat <- gg_dta
gg_dta.cat[["celltype"]] <- gg_dta[["trt"]] <- gg_dta[["prior"]] <- NULL
plot(gg_dta, panel=TRUE)

gg_dta.cat[["karno"]] <- gg_dta.cat[["diagtime"]] <-
  gg_dta.cat[["age"]] <- NULL
plot(gg_dta.cat, panel=TRUE, notch=TRUE)

gg_dta <- lapply(partial_veteran, gg_partial)
length(gg_dta)
gg_dta <- combine.gg_partial(gg_dta[[1]], gg_dta[[2]] )

```

```

plot(gg_dta[["karno"]])
plot(gg_dta[["celltype"]])

gg_dta.cat <- gg_dta
gg_dta[["celltype"]] <- gg_dta[["trt"]] <- gg_dta[["prior"]] <- NULL
plot(gg_dta, panel=TRUE)

gg_dta.cat[["karno"]] <- gg_dta.cat[["diagtime"]] <-
  gg_dta.cat[["age"]] <- NULL
plot(gg_dta.cat, panel=TRUE, notch=TRUE)

## ----- pbc data

## End(Not run)

```

---

plot.gg\_rfsrc                      *Predicted response plot from a gg\_rfsrc object.*

---

## Description

Plot the predicted response from a [gg\\_rfsrc](#) object, the [rfsrc](#) prediction, using the OOB prediction from the forest.

## Usage

```
## S3 method for class 'gg_rfsrc'
plot(x, ...)
```

## Arguments

x                      [gg\\_rfsrc](#) object created from a [rfsrc](#) object  
 ...                    arguments passed to [gg\\_rfsrc](#).

## Value

ggplot object

## References

Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.  
 Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.  
 Ishwaran H. and Kogalur U.B. (2013). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.4.

## See Also

[gg\\_rfsrc](#) [rfsrc](#)

**Examples**

```

## Not run:
## -----
## classification example
## -----
## ----- iris data
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
data(rfsrc_iris, package="ggRandomForests")
gg_dta<- gg_rfsrc(rfsrc_iris)

plot(gg_dta)

## -----
## Regression example
## -----
## ----- air quality data
# rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality, na.action = "na.impute")
data(rfsrc_airq, package="ggRandomForests")
gg_dta<- gg_rfsrc(rfsrc_airq)

plot(gg_dta)

## ----- Boston data
data(rfsrc_boston, package="ggRandomForests")
plot(rfsrc_boston)

## ----- mtcars data
data(rfsrc_mtcars, package="ggRandomForests")
gg_dta<- gg_rfsrc(rfsrc_mtcars)

plot(gg_dta)

## -----
## Survival example
## -----
## ----- veteran data
## randomized trial of two treatment regimens for lung cancer
# data(veteran, package = "randomForestSRC")
# rfsrc_veteran <- rfsrc(Surv(time, status) ~ ., data = veteran, ntree = 100)
data(rfsrc_veteran, package = "ggRandomForests")
gg_dta <- gg_rfsrc(rfsrc_veteran)
plot(gg_dta)

gg_dta <- gg_rfsrc(rfsrc_veteran, conf.int=.95)
plot(gg_dta)

gg_dta <- gg_rfsrc(rfsrc_veteran, by="trt")
plot(gg_dta)

## ----- pbc data
data(rfsrc_pbc, package = "ggRandomForests")
gg_dta <- gg_rfsrc(rfsrc_pbc)

```

```
plot(gg_dta)

gg_dta <- gg_rfsrc(rfsrc_pbc, conf.int=.95)
plot(gg_dta)

gg_dta <- gg_rfsrc(rfsrc_pbc, by="treatment")
plot(gg_dta)

## End(Not run)
```

---

plot.gg\_roc                      *ROC plot generic function for a [gg\\_roc](#) object.*

---

## Description

ROC plot generic function for a [gg\\_roc](#) object.

## Usage

```
## S3 method for class 'gg_roc'
plot(x, which.outcome = NULL, ...)
```

## Arguments

x                      [gg\\_roc](#) object created from a classification forest  
which.outcome        for multiclass problems, choose the class for plotting  
...                    arguments passed to the [gg\\_roc](#) function

## Value

ggplot object of the ROC curve

## References

- Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.
- Ishwaran H. and Kogalur U.B. (2013). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.4.

## See Also

[gg\\_roc](#) [rfsrc](#)

**Examples**

```
## Not run:
## -----
## classification example
## -----
## ----- iris data
#rfsrc_iris <- rfsrc(Species ~ ., data = iris)
data(rfsrc_iris, package="ggRandomForests")

# ROC for setosa
gg_dta <- gg_roc(rfsrc_iris, which.outcome=1)
plot.gg_roc(gg_dta)

# ROC for versicolor
gg_dta <- gg_roc(rfsrc_iris, which.outcome=2)
plot.gg_roc(gg_dta)

# ROC for virginica
gg_dta <- gg_roc(rfsrc_iris, which.outcome=3)
plot.gg_roc(gg_dta)

# Alternatively, you can plot all three outcomes in one go
# by calling the plot function on the forest object.
plot.gg_roc(rfsrc_iris)

## End(Not run)
```

---

plot.gg\_survival      *Plot a [gg\\_survival](#) object.*

---

**Description**

Plot a [gg\\_survival](#) object.

**Usage**

```
## S3 method for class 'gg_survival'
plot(
  x,
  type = c("surv", "cum_haz", "hazard", "density", "mid_int", "life", "proplife"),
  error = c("shade", "bars", "lines", "none"),
  ...
)
```

**Arguments**

x	gg_survival or a survival gg_rfsrc object created from a rfsrc object
type	"surv", "cum_haz", "hazard", "density", "mid_int", "life", "proplife"
error	"shade", "bars", "lines" or "none"
...	not used

**Value**

ggplot object

**Examples**

```
## Not run:
## ----- pbc data
data(pbc, package="randomForestSRC")
pbc$time <- pbc$days/364.25

# This is the same as kaplan
gg_dta <- gg_survival(interval="time", censor="status",
                      data=pbc)

plot(gg_dta, error="none")
plot(gg_dta)

# Stratified on treatment variable.
gg_dta <- gg_survival(interval="time", censor="status",
                      data=pbc, by="treatment")

plot(gg_dta, error="none")
plot(gg_dta)

# ...with smaller confidence limits.
gg_dta <- gg_survival(interval="time", censor="status",
                      data=pbc, by="treatment", conf.int=.68)

plot(gg_dta, error="lines")

## End(Not run)
```

---

plot.gg\_variable      *Plot a gg\_variable object,*

---

**Description**

Plot a gg\_variable object,

**Usage**

```
## S3 method for class 'gg_variable'
plot(
  x,
  xvar,
  time,
  time_labels,
  panel = FALSE,
  oob = TRUE,
  points = TRUE,
  smooth = TRUE,
  ...
)
```

**Arguments**

x	gg_variable object created from a rfsrc object
xvar	variable (or list of variables) of interest.
time	For survival, one or more times of interest
time_labels	string labels for times
panel	Should plots be faceted along multiple xvar?
oob	oob estimates (boolean)
points	plot the raw data points (boolean)
smooth	include a smooth curve (boolean)
...	arguments passed to the ggplot2 functions.

**Value**

A single ggplot object, or list of ggplot objects

**References**

Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.

Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.

Ishwaran H. and Kogalur U.B. (2013). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.4.

**Examples**

```
## Not run:
## -----
## classification
## -----
## ----- iris data
## iris
#rfsrc_iris <- rfsrc(Species ~., data = iris)
```



```

data(rfsrc_iris, package="ggRandomForests")

gg_dta <- gg_variable(rfsrc_iris)
plot(gg_dta, xvar="Sepal.Width")
plot(gg_dta, xvar="Sepal.Length")

## !! TODO !! this needs to be corrected
plot(gg_dta, xvar=rfsrc_iris$xvar.names,
      panel=TRUE, se=FALSE)

## -----
## regression
## -----
## ----- air quality data
#rfsrc_airq <- rfsrc(Ozone ~ ., data = airquality)
data(rfsrc_airq, package="ggRandomForests")
gg_dta <- gg_variable(rfsrc_airq)

# an ordinal variable
gg_dta[, "Month"] <- factor(gg_dta[, "Month"])

plot(gg_dta, xvar="Wind")
plot(gg_dta, xvar="Temp")
plot(gg_dta, xvar="Solar.R")

plot(gg_dta, xvar=c("Solar.R", "Wind", "Temp", "Day"), panel=TRUE)

plot(gg_dta, xvar="Month", notch=TRUE)

## ----- motor trend cars data
#rfsrc_mtcars <- rfsrc(mpg ~ ., data = mtcars)
data(rfsrc_mtcars, package="ggRandomForests")
gg_dta <- gg_variable(rfsrc_mtcars)

# mtcars$cyl is an ordinal variable
gg_dta$cyl <- factor(gg_dta$cyl)
gg_dta$am <- factor(gg_dta$am)
gg_dta$vs <- factor(gg_dta$vs)
gg_dta$gear <- factor(gg_dta$gear)
gg_dta$carb <- factor(gg_dta$carb)

plot(gg_dta, xvar="cyl")

# Others are continuous
plot(gg_dta, xvar="disp")
plot(gg_dta, xvar="hp")
plot(gg_dta, xvar="wt")

# panel
plot(gg_dta, xvar=c("disp", "hp", "drat", "wt", "qsec"), panel=TRUE)
plot(gg_dta, xvar=c("cyl", "vs", "am", "gear", "carb"), panel=TRUE)

## ----- Boston data

```

```

## -----
## survival examples
## -----
## ----- veteran data
## survival
data(veteran, package = "randomForestSRC")
rfsrc_veteran <- rfsrc(Surv(time,status)~., veteran, nsplit = 10,
                      ntree = 100)

# get the 1 year survival time.
gg_dta <- gg_variable(rfsrc_veteran, time=90)

# Generate variable dependance plots for age and diagtime
plot(gg_dta, xvar = "age")
plot(gg_dta, xvar = "diagtime")

# Generate coplots
plot(gg_dta, xvar = c("age", "diagtime"), panel=TRUE)

# If we want to compare survival at different time points, say 30, 90 day
# and 1 year
gg_dta <- gg_variable(rfsrc_veteran, time=c(30, 90, 365))

# Generate variable dependance plots for age and diagtime
plot(gg_dta, xvar = "age")
plot(gg_dta, xvar = "diagtime")

# Generate coplots
plot(gg_dta, xvar = c("age", "diagtime"), panel=TRUE)

## ----- pbc data

## End(Not run)

```

---

plot.gg_vimp	<i>Plot a <a href="#">gg_vimp</a> object, extracted variable importance of a <a href="#">rfsrc</a> object</i>
--------------	---

---

## Description

Plot a [gg\\_vimp](#) object, extracted variable importance of a [rfsrc](#) object

## Usage

```

## S3 method for class 'gg_vimp'
plot(x, relative, lbls, ...)

```

**Arguments**

x	<a href="#">gg_vimp</a> object created from a <a href="#">rfsrc</a> object
relative	should we plot vimp or relative vimp. Defaults to vimp.
lbls	A vector of alternative variable labels. Item names should be the same as the variable names.
...	optional arguments passed to <a href="#">gg_vimp</a> if necessary

**Value**

ggplot object

**References**

- Breiman L. (2001). Random forests, *Machine Learning*, 45:5-32.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R, *Rnews*, 7(2):25-31.
- Ishwaran H. and Kogalur U.B. (2013). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.4.

**See Also**

[gg\\_vimp](#)

**Examples**

```
## Not run:
## -----
## classification example
## -----
## ----- iris data
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
data(rfsrc_iris, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_iris)
plot(gg_dta)

## -----
## regression example
## -----
## ----- air quality data
# rfsrc_airq <- rfsrc(Ozone ~ ., airquality)
data(rfsrc_airq, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_airq)
plot(gg_dta)

## ----- Boston data
data(rfsrc_boston, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_boston)
plot(gg_dta)

## ----- mtcars data
```

```

data(rfsrc_mtcars, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_mtcars)
plot(gg_dta)

## -----
## survival example
## -----
## ----- veteran data
data(rfsrc_veteran, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_veteran)
plot(gg_dta)

## ----- pbc data
data(rfsrc_pbc, package="ggRandomForests")
gg_dta <- gg_vimp(rfsrc_pbc)
plot(gg_dta)

## End(Not run)

```

---

```
print.gg_minimal_depth
```

*Print a `gg_minimal_depth` object.*

---

## Description

Print a `gg_minimal_depth` object.

## Usage

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

## Arguments

`x` a `gg_minimal_depth` object.  
`...` optional arguments

## Examples

```

## -----
## classification example
## -----
## You can build a randomForest
# rfsrc_iris <- rfsrc(Species ~ ., data = iris)
# varsel_iris <- var.select(rfsrc_iris)
# ... or load a cached randomForestSRC object
data(varsel_iris, package="ggRandomForests")

```

```

# Get a data.frame containing minimaldepth measures
gg_dta <- gg_minimal_depth(varsel_iris)
print(gg_dta)

## -----
## regression example
## -----
## Not run:
# ... or load a cached randomForestSRC object
data(varsel_airq, package="ggRandomForests")

# Get a data.frame containing minimaldepth measures
gg_dta<- gg_minimal_depth(varsel_airq)
print(gg_dta)

# To nicely print a rfsrc::var.select output...
print(varsel_airq)

## End(Not run)

# ... or load a cached randomForestSRC object
data(varsel_boston, package="ggRandomForests")

# Get a data.frame containing minimaldepth measures
gg_dta<- gg_minimal_depth(varsel_boston)
print(gg_dta)

# To nicely print a rfsrc::var.select output...
print(varsel_boston)

```

---

quantile\_pts

*Find points evenly distributed along the vectors values.*


---

### Description

This function finds point values from a vector argument to produce groups intervals. Setting groups=2 will return three values, the two end points, and one mid point (at the median value of the vector).

The output can be passed directly into the breaks argument of the cut function for creating groups for coplots.

### Usage

```
quantile_pts(object, groups, intervals = FALSE)
```

**Arguments**

object	vector object of values.
groups	how many points do we want
intervals	should we return the raw points or intervals to be passed to the cut function

**Value**

vector of groups+1 cut point values.

**See Also**

cut [gg\\_partial\\_coplot](#)

**Examples**

```
data(rfsrc_boston)

# To create 6 intervals, we want 7 points.
# quantile_pts will find balanced intervals
rm_pts <- quantile_pts(rfsrc_boston$xvar$rm, groups=6, intervals=TRUE)

# Use cut to create the intervals
rm_grp <- cut(rfsrc_boston$xvar$rm, breaks=rm_pts)

summary(rm_grp)
```

---

rfsrc\_data

*Cached [rfsrc](#) objects for examples, diagnostics and vignettes.*

---

**Description**

Data sets storing [rfsrc](#) objects corresponding to training data according to the following naming convention:

- rfsrc\_iris - randomForestSR[C] for the iris data set.
- rfsrc\_boston - randomForestS[R]C for the Boston housing data set (MASS package).
- rfsrc\_pbc - randomForest[S]RC for the pbc data set (randomForestSRC package)

**Format**

[rfsrc](#) object

## Details

Constructing random forests are computationally expensive. We cache `rfsrc` objects to improve the `ggRandomForests` examples, diagnostics and vignettes run times. (see `cache_rfsrc_datasets` to rebuild a complete set of these data sets.)

For each data set listed, we build a `rfsrc`. Tuning parameters used in each case are documented in the examples. Each data set is built with the `cache_rfsrc_datasets` with the `randomForestSRC` version listed in the `ggRandomForests DESCRIPTION` file.

- `rfsrc_iris` - The famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. Build a classification random forest for predicting the species (setosa, versicolor, and virginica) on 5 variables (columns) and 150 observations (rows).
- `rfsrc_boston` - The Boston housing values in suburbs of Boston from the MASS package. Build a regression random forest for predicting `medv` (median home values) on 13 covariates and 506 observations.
- `rfsrc_pbc` - The pbc data from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo controlled trial of the drug D-penicillamine. 312 cases participated in the randomized trial and contain largely complete data. Data from the `randomForestSRC` package. Build a survival random forest for time-to-event death data with 17 covariates and 312 observations (remaining 106 observations are held out).

## References

- # randomForestSRC
- Ishwaran H. and Kogalur U.B. (2014). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.5.5.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R. *R News* 7(2), 25-31.
- Ishwaran H., Kogalur U.B., Blackstone E.H. and Lauer M.S. (2008). Random survival forests. *Ann. Appl. Statist.* 2(3), 841-860.
- # Boston data set
- Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity.* New York: Wiley.
- Harrison, D., and D.L. Rubinfeld. 1978. "Hedonic Prices and the Demand for Clean Air." *J. Environ. Economics and Management* 5: 81-102. # Iris data set
- Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language.* Wadsworth & Brooks/Cole. (has `iris3` as iris.)
- Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, Part II, 179-188.
- Anderson, Edgar (1935). The irises of the Gaspe Peninsula, *Bulletin of the American Iris Society*, 59, 2-5.
- # pbc data set

Fleming T.R and Harrington D.P., (1991) Counting Processes and Survival Analysis. New York: Wiley.

T Therneau and P Grambsch (2000), Modeling Survival Data: Extending the Cox Model, Springer-Verlag, New York. ISBN: 0-387-98784-3.

### See Also

[iris](#) [Boston](#) [pbc](#) [rfsrc](#) [cache\\_rfsrc\\_datasets](#) [gg\\_rfsrc](#) [plot.gg\\_rfsrc](#) [gg\\_error](#) [plot.gg\\_error](#)

### Examples

```
## Not run:
#-----
# iris data - classification random forest
#-----
# rfsrc grow call
rfsrc_iris <- rfsrc(Species ~., data = iris)

# plot the forest generalization error convergence
gg_dta <- gg_error(rfsrc_iris)
plot(gg_dta)

# Plot the forest predictions
gg_dta <- gg_rfsrc(rfsrc_iris)
plot(gg_dta)

#-----
# MASS::Boston data - regression random forest
#-----
# Load the data...
data(Boston, package="MASS")
Boston$chas <- as.logical(Boston$chas)

# rfsrc grow call
rfsrc_boston <- rfsrc(medv~., data=Boston)

# plot the forest generalization error convergence
gg_dta <- gg_error(rfsrc_boston)
plot(gg_dta)

# Plot the forest predictions
gg_dta <- gg_rfsrc(rfsrc_boston)
plot(gg_dta)

#-----
# randomForestSRC::pbc data - survival random forest
#-----
# Load the data...
# For simplicity here. We do a bit of data tidying
# before running the stored random forest.
data(pbc, package="randomForestSRC")
```



```

# Remove non-randomized cases
dta.train <- pbc[-which(is.na(pbc$treatment)),]

# rfsrc grow call
rfsrc_pbc <- rfsrc(Surv(years, status) ~ ., dta.train, nsplit = 10,
                  na.action="na.impute")

# plot the forest generalization error convergence
gg_dta <- gg_error(rfsrc_pbc)
plot(gg_dta)

# Plot the forest predictions
gg_dta <- gg_rfsrc(rfsrc_pbc)
plot(gg_dta)

## End(Not run)

```

---

shift                                    *lead function to shift by one (or more).*

---

### Description

lead function to shift by one (or more).

### Usage

```
shift(x, shift_by = 1)
```

### Arguments

x	a vector of values
shift_by	an integer of length 1, giving the number of positions to lead (positive) or lag (negative) by

### Details

Lead and lag are useful for comparing values offset by a constant (e.g. the previous or next value)

Taken from: <http://ctszkin.com/2012/03/11/generating-a-laglead-variables/>

This function allows me to remove the dplyr::lead depends. Still suggest for vignettes though.

### Examples

```

d<-data.frame(x=1:15)
#generate lead variable
d$df_lead2<-ggRandomForests:::shift(d$x,2)
#generate lag variable
d$df_lag2<-ggRandomForests:::shift(d$x,-2)

```

---

surface_matrix	<i>Construct a set of (x, y, z) matrices for surface plotting a gg_partial_coplot object</i>
----------------	--

---

### Description

Construct a set of (x, y, z) matrices for surface plotting a gg\_partial\_coplot object

### Usage

```
surface_matrix(dta, xvar)
```

### Arguments

dta	a gg_partial_coplot object containing at least 3 numeric columns of data
xvar	a vector of 3 column names from the data object, in (x, y, z) order

### Details

To create a surface plot, the plot3D::surf3D function expects 3 matrices of n.x by n.y. Take the p+1 by n gg\_partial\_coplot object, and extract and construct the x, y and z matrices from the provided xvar column names.

### Examples

```
## Not run:
## From vignette(randomForestRegression, package="ggRandomForests")
##
data(rfsrc_boston)
rm_pts <- quantile_pts(rfsrc_boston$xvar$rm, groups=49, intervals=TRUE)

# Load the stored partial coplot data.
data(partial_boston_surf)

# Instead of groups, we want the raw rm point values,
# To make the dimensions match, we need to repeat the values
# for each of the 50 points in the lstat direction
rm.tmp <- do.call(c,lapply(rm_pts,
                          function(grp) {rep(grp, 1
                                              length(partial_boston_surf))}))

# Convert the list of plot.variable output to
partial_surf <- do.call(rbind,lapply(partial_boston_surf, gg_partial))

# attach the data to the gg_partial_coplot
partial_surf$rm <- rm.tmp

# Transform the gg_partial_coplot object into a list of three named matrices
# for surface plotting with plot3D::surf3D
```

```

srf <- surface_matrix(partial_surf, c("lstat", "rm", "yhat"))

## End(Not run)

## Not run:
# surf3D is in the plot3D package.
library(plot3D)
# Generate the figure.
surf3D(x=srf$x, y=srf$y, z=srf$z, col=topo.colors(10),
       colkey=FALSE, border = "black", bty="b2",
       shade = 0.5, expand = 0.5,
       lighting = TRUE, lphi = -50,
       xlab="Lower Status", ylab="Average Rooms", zlab="Median Value"
)

## End(Not run)

```

---

varsel\_data

*Cached `var.select` objects for examples, diagnostics and vignettes. Data sets storing `var.select` objects corresponding to training data according to the following naming convention:*

- `varsel_iris` - from a `randomForestSR[C]` for the `iris` data set.
  - `varsel_boston` - from a `randomForestS[R]C` for the Boston housing data set (MASS package).
  - `varsel_pbc` - from a `randomForest[S]RC` for the `pbc` data set (`randomForestSRC` package)
- 

## Description

Cached `var.select` objects for examples, diagnostics and vignettes.

Data sets storing `var.select` objects corresponding to training data according to the following naming convention:

- `varsel_iris` - from a `randomForestSR[C]` for the `iris` data set.
- `varsel_boston` - from a `randomForestS[R]C` for the Boston housing data set (MASS package).
- `varsel_pbc` - from a `randomForest[S]RC` for the `pbc` data set (`randomForestSRC` package)

## Format

`var.select` object

## Details

Constructing minimal depth variable selection with the `randomForestSRC::var.select` function is computationally expensive. We cache `var.select` objects to improve the `ggRandomForests` examples, diagnostics and vignettes run times. (see `cache_rfsrc_datasets` to rebuild a complete set of these data sets.)

For each data set listed, we build a `rfsrc` (see `rfsrc_data`), then calculate the minimal depth variable selection with `var.select` function, setting `method="md"`. Each data set is built with the `cache_rfsrc_datasets` with the `randomForestSRC` version listed in the `ggRandomForests DESCRIPTION` file.

- `varsel_iris` - The famous (Fisher's or Anderson's) `iris` data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. Build a classification random forest for predicting the species (`setosa`, `versicolor`, and `virginica`) on 5 variables (columns) and 150 observations (rows).
- `varsel_boston` - The Boston housing values in suburbs of Boston from the `MASS` package. Build a regression random forest for predicting `medv` (median home values) on 13 covariates and 506 observations.
- `varsel_pbc` - The `pbc` data from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo controlled trial of the drug D-penicillamine. 312 cases participated in the randomized trial and contain largely complete data. Data from the `randomForestSRC` package. Build a survival random forest for time-to-event death data with 17 covariates and 312 observations (remaining 106 observations are held out).

## References

- #————— `randomForestSRC` —————
- Ishwaran H. and Kogalur U.B. (2014). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.5.5.
- Ishwaran H. and Kogalur U.B. (2007). Random survival forests for R. *R News* 7(2), 25-31.
- Ishwaran H., Kogalur U.B., Blackstone E.H. and Lauer M.S. (2008). Random survival forests. *Ann. Appl. Statist.* 2(3), 841-860.
- #————— Boston data set —————
- Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*. New York: Wiley.
- Harrison, D., and D.L. Rubinfeld. 1978. "Hedonic Prices and the Demand for Clean Air." *J. Environ. Economics and Management* 5: 81-102.
- #————— Iris data set —————
- Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole. (has `iris3` as `iris`.)
- Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, Part II, 179-188.
- Anderson, Edgar (1935). The irises of the Gaspé Peninsula, *Bulletin of the American Iris Society*, 59, 2-5.
- #————— `pbc` data set —————
- Fleming T.R and Harrington D.P., (1991) *Counting Processes and Survival Analysis*. New York: Wiley.

T Therneau and P Grambsch (2000), *Modeling Survival Data: Extending the Cox Model*, Springer-Verlag, New York. ISBN: 0-387-98784-3.

### See Also

`iris` `Boston` `pc` `var.select` `rfsrc_data` `cache_rfsrc_datasets` `gg_minimal_depth` `plot.gg_minimal_depth` `gg_minimal_vimp` `plot.gg_minimal_vimp`

### Examples

```
## Not run:
#-----
# iris data - classification random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_iris, package="ggRandomForests")

# The var.select call
varsel_iris <- var.select(rfsrc_iris)

# plot the forestminimal depth ranking
gg_dta <- gg_minimal_depth(varsel_iris)
plot(gg_dta)

#-----
# MASS::Boston data - regression random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_boston, package="ggRandomForests")

# The var.select call
varsel_boston <- var.select(rfsrc_boston)

# plot the forestminimal depth ranking
gg_dta <- gg_minimal_depth(varsel_boston)
plot(gg_dta)

#-----
# randomForestSRC::pbc data - survival random forest
#-----
# load the rfsrc object from the cached data
data(rfsrc_pbc, package="ggRandomForests")

# The var.select call
varsel_pbc <- var.select(rfsrc_pbc)

# plot the forestminimal depth ranking
gg_dta <- gg_minimal_depth(varsel_pbc)
plot(gg_dta)
```

```
## End(Not run)
```

# Index

## \* datasets

- interaction\_data, 32
  - partial\_coplot\_data, 37
  - partial\_data, 39
  - partial\_surface\_data, 42
  - rfsrc\_data, 70
  - varsel\_data, 75
- Boston, 5, 34, 39, 43, 72, 77
- cache\_rfsrc\_datasets, 4, 33, 34, 38–41, 43, 71, 72, 75–77
- calc\_auc, 5, 7
- calc\_roc, 6
- calc\_roc (calc\_roc.rfsrc), 6
- calc\_roc.rfsrc, 6
- combine.gg\_partial, 8
- combine.gg\_partial\_list (combine.gg\_partial), 8
- find.interaction, 5, 11, 12, 32–34, 48
- gg\_error, 3, 9, 45, 72
- gg\_interaction, 3, 11, 34, 47
- gg\_minimal\_depth, 3, 13, 49–51, 68, 77
- gg\_minimal\_vimp, 3, 16, 51, 52, 77
- gg\_partial, 3, 8, 18, 41, 43, 53, 54, 56, 57
- gg\_partial\_coplot, 3, 70
- gg\_partial\_coplot (gg\_partial\_coplot.rfsrc), 21
- gg\_partial\_coplot.rfsrc, 21
- gg\_partial\_list, 22
- gg\_partial\_list (gg\_partial), 18
- gg\_rfsrc, 3, 59, 63, 72
- gg\_rfsrc (gg\_rfsrc.rfsrc), 23
- gg\_rfsrc.rfsrc, 23
- gg\_roc, 3, 5–7, 61
- gg\_roc (gg\_roc.rfsrc), 25
- gg\_roc.rfsrc, 25
- gg\_survival, 4, 23, 26, 36, 37, 62, 63
- gg\_variable, 3, 28, 54, 57, 63, 64
- gg\_vimp, 3, 30, 66, 67
- ggRandomForests-package, 3
- interaction\_boston (interaction\_data), 32
- interaction\_boston, (interaction\_data), 32
- interaction\_data, 5, 32
- interaction\_iris (interaction\_data), 32
- interaction\_iris, (interaction\_data), 32
- interaction\_pbc (interaction\_data), 32
- kaplan, 27, 35
- max.subtree, 12, 48
- nelson, 27, 36, 36, 37
- partial\_boston (partial\_data), 39
- partial\_boston, (partial\_data), 39
- partial\_boston\_surf (partial\_surface\_data), 42
- partial\_boston\_surf, (partial\_surface\_data), 42
- partial\_coplot\_boston (partial\_coplot\_data), 37
- partial\_coplot\_boston, (partial\_coplot\_data), 37
- partial\_coplot\_boston2 (partial\_coplot\_data), 37
- partial\_coplot\_boston2, (partial\_coplot\_data), 37
- partial\_coplot\_data, 5, 37
- partial\_coplot\_pbc (partial\_coplot\_data), 37
- partial\_coplot\_pbc, (partial\_coplot\_data), 37
- partial\_coplot\_pbc2 (partial\_coplot\_data), 37
- partial\_data, 5, 39

partial\_iris (partial\_data), 39  
partial\_iris, (partial\_data), 39  
partial\_pbc (partial\_data), 39  
partial\_pbc\_surf  
    (partial\_surface\_data), 42  
partial\_pbc\_surf,  
    (partial\_surface\_data), 42  
partial\_pbc\_time  
    (partial\_surface\_data), 42  
partial\_surface\_data, 42  
pbc, 5, 34, 41, 43, 72, 77  
plot.gg\_error, 9, 10, 45, 72  
plot.gg\_interaction, 12, 34, 47, 48  
plot.gg\_minimal\_depth, 14, 49, 77  
plot.gg\_minimal\_vimp, 16, 51, 77  
plot.gg\_partial, 18, 41, 43, 53, 57  
plot.gg\_partial\_list, 54, 56  
plot.gg\_rfsrc, 23, 59, 72  
plot.gg\_roc, 6, 7, 25, 61  
plot.gg\_survival, 27, 36, 37, 62  
plot.gg\_variable, 28, 54, 57, 63  
plot.gg\_vimp, 31, 66  
plot.rfsrc, 45  
plot.variable, 5, 18, 22, 28, 37–43, 54, 57  
predict.rfsrc, 7, 16  
print.gg\_minimal\_depth, 68  
  
quantile\_pts, 69  
  
randomForest, 25  
rfsrc, 3, 5, 7–9, 11, 12, 16, 18, 22, 23, 25, 28,  
    30, 31, 33, 37, 38, 40, 43, 45, 47–49,  
    53, 56, 59, 63, 64, 66, 67, 70–72, 76  
rfsrc\_boston (rfsrc\_data), 70  
rfsrc\_boston, (rfsrc\_data), 70  
rfsrc\_data, 5, 33, 34, 40, 41, 43, 70, 76, 77  
rfsrc\_iris (rfsrc\_data), 70  
rfsrc\_iris, (rfsrc\_data), 70  
rfsrc\_pbc (rfsrc\_data), 70  
rfsrc\_pbc, (rfsrc\_data), 70  
rfsrc\_pbc\_test (rfsrc\_data), 70  
  
shift, 73  
surface\_matrix, 74  
  
var.select, 5, 12, 16, 48, 50–52, 75–77  
var.select.rfsrc, 16  
varsel\_boston (varsel\_data), 75  
varsel\_boston, (varsel\_data), 75  
varsel\_data, 5, 75  
varsel\_iris (varsel\_data), 75  
varsel\_iris, (varsel\_data), 75  
varsel\_pbc (varsel\_data), 75  
veteran, 5  
vimp, 12, 31, 48  
vimp.rfsrc, 31