

# Package ‘VDCutil’

April 17, 2009

**Title** utilities supporting the VDC system

**Version** 1.15

**Author** Micah Altman, Akio Sone

**Date** 2008-05-01

**Description** The VDC system is an open source digital library system for quantitative data. This is a package to support on-line analysis using VDC and Zelig, accuracy, and R2HTML

**Maintainer** VDC Development Team <vdc-dev@latte.harvard.edu>

**Depends** R (>= 2.1.0), Zelig, R2HTML, accuracy

**Suggests** UNF, foreign

**License** GPL-2

**URL** <http://thedata.org/>

**Repository** CRAN

**Date/Publication** 2008-05-02 10:10:15

## R topics documented:

hplot . . . . .	2
HTML.summary.zelig . . . . .	3
HTML.VDCxtabs . . . . .	4
print.VDCxtabs . . . . .	5
printZeligSchemaInstance . . . . .	6
recodeVDCdf . . . . .	7
VDCcrossTabulation . . . . .	8
VDCgenAnalysis . . . . .	9
VDCWebSearch . . . . .	11
VDCxtabs . . . . .	12
zeligModelDependencies . . . . .	13
<b>Index</b>	<b>14</b>

---

hplot *plots an object, with html output*

---

### Description

This calls `plot()` to plot the object, `bitmap()` to output it as bitmaps, and `HTMLInsertGraph` to generate HTML referring to it

### Usage

```
hplot(x,y, ... , htmlFile=NULL, graphfileBase=NULL, bitmapArgs=NULL,htmlArgs=NULL)
```

### Arguments

<code>x</code>	object to plotted
<code>y</code>	optional object to plotted
<code>...</code>	additional parameters passed to <code>plot.default</code>
<code>htmlFile</code>	location of html file to write to, will default to contents of <code>.HTML.file</code> or <code>tempdir/index.html</code>
<code>graphfileBase</code>	base name for graph files , defaults to "graph"
<code>bitmapArgs</code>	arguments to pass to <code>bitmap</code>
<code>htmlArgs</code>	arguments to pass to <code>HTMLInsertFile</code>

### Value

None

### Author(s)

Micah Altman (vdc-devlatte.harvard.edu) <http://thedata.org>

### See Also

[HTMLInsertGraph](#), [bitmap](#), [plot](#)

### Examples

```
# Note: requires R2HTML
# start HTML output
HTMLInitFile()
hplot(1:10) # plot(1:10) embedded as html

# lm produces multiple plots, which are all captured sequentially
hplot(lm(Employed~.,longley))
#change default options
hplot(2:10,bitmapArgs=list(type="jpeg",res=300),htmlArgs=list(GraphBorder=0,Caption="Thi
```

```
HTMLEndFile()
```

---

```
HTML.summary.zelig pretty print zelig sim summaries
```

---

## Description

This is used with R2HTML to pretty-print zelig simulation summaries

## Usage

```
HTML.summary.zelig(x,digits = getOption("digits"), print.x = FALSE, ...)  
HTML.summary.MCMCZelig(x,digits = max(3, getOption("digits") - 3), ...)  
HTML.summary.zelig.strata(x,subset = NULL, ...)
```

## Arguments

x	object to plotted
digits	number of digits to print
subset	strata subset
print.x	include x values
...	additional parameters passed to plot.default

## Value

None

## Author(s)

Micah Altman (vdc-devlatte.harvard.edu) <http://thedata.org>

## See Also

[HTML](#), [sim](#)

## Examples

```
# should be Employed~. , but bug in Zelig 2.52  
z.out = zelig(Employed~Year,"ls",longley)  
sim.out =sim(z.out, setx(z.out))  
summary(sim.out)  
# HTML methods called implicitly  
HTML(summary(sim.out),file="")
```

HTML.VDCxtabs      *pretty print VDCxtabs*

---

### Description

This is used with R2HTML to pretty-print VDCxtabs

### Usage

```
HTML.VDCxtabs(x, ...,
              wantPercentages=T, wantTotals=T, wantStats=T, nameLength=15)
```

### Arguments

x	object to converted to HTML
...	additional parameters passed to HTML
wantPercentages	include percentages in the output
wantTotals	include marginals in the output
wantStats	include chi-squared statistics in the output
nameLength	maximum length of variable labels

### Value

None

### Author(s)

Micah Altman <vdc-devlatte.harvard.edu> <http://thedata.org>

### See Also

[HTML](#), [VDCxtabs](#)

### Examples

```
vxt=VDCxtabs(Freq~., as.data.frame(Titanic))
# HTML methods called implicitly
HTML(vxt, file="")
```

---

print.VDCxtabs      *print VDCxtabs*

---

## Description

This is used to pretty-print VDCxtabs

## Usage

```
## S3 method for class 'VDCxtabs':  
print(x, ...,  
      wantPercentages=T, wantTotals=T, wantStats=T, nameLength=15 )
```

## Arguments

x	object to printed
...	additional parameters passed to print.default
wantPercentages	include percentages in the output
wantTotals	include marginals in the output
wantStats	include chi-squared statistics in the output
nameLength	maximum length of variable labels

## Value

None

## Author(s)

Micah Altman (vdc-devlatte.harvard.edu) <http://thedata.org>

## See Also

[VDCxtabs](#)

## Examples

```
# print methods called implicitly  
VDCxtabs(Freq~, as.data.frame(Titanic))
```

---

```
printZeligSchemaInstance  
    generate list of Zelig models as single XML document
```

---

**Description**

This is used to generate a single XML file describing the supported Zelig models

**Usage**

```
printZeligSchemaInstance(filename=NULL, serverName=NULL, vdcAbsDirPrefix=NU
```

**Arguments**

<code>filename</code>	output file
<code>serverName</code>	server name, for schema location
<code>vdcAbsDirPrefix</code>	prefix for schema location, directory prefix

**Value**

None. Outputs to file.

**Author(s)**

Micah Altman (thedata-userslists.sourceforge.net) <http://thedata.org>

**See Also**

[zelig](#)

**Examples**

```
printZeligSchemaInstance(filename="", serverName="thedata.org", vdcAbsDirPrefix="/zel
```

---

`recodeVDCdf`*Recode's VDC data frame using embedded attributes.*

---

**Description**

VDC embeds variable types, labels, value labels, and additional missing value codes in attributes. This method recodes a VDC data frame for analysis in R.

**Usage**

```
recodeVDCdf(x, recodemiss=TRUE, recodenames=FALSE, recodefactors=TRUE, dropfactorlevels=FALSE)
```

**Arguments**

<code>x</code>	VDC data frame to be recoded
<code>recodemiss</code>	recode VDC missing data codes to NA's in the data frame
<code>recodenames</code>	recode VDC variable labels to var names in the data frame
<code>recodefactors</code>	recode discrete numeric and character variables to factors, and apply VDC value labels as factor labels
<code>dropfactorlevels</code>	drop unused factor levels in recoding
<code>orderfactors</code>	use heuristic to determine when to recode to ordered factors. If false, will return unordered factors only

**Value**

Returns recoded data frame if VDC metadata is available, unaltered data frame otherwise.

**Note**

Warning: Recoding integers as factors alters coding sequences. This does not affect most analysis (in fact, most zelig models expect factors). However recoded data frames should not be saved, as information is lost. Use the original VDC data frame.

Also, we use a heuristic to determine whether a factor is ordered. This works in most cases but can mistakenly treat integer variables with more than 3 real categories as ordered.

**Author(s)**

Micah Altman (vdc-devlatte.harvard.edu) <http://thedata.org>

**See Also**

[VDCgenAnalysis](#),

---

VDCcrossTabulation *VDC cross tabs, with pretty printing*

---

### Description

This calculate cross tabs with marginals and proportions.

### Usage

```
VDCcrossTabulation(data=parent.env(), classificationVars=NULL, freqVars=NULL,
  wantPercentages=T, wantTotals=T, wantStats=T, wantExtraTables=FALSE,
  HTMLfile="", ...)
```

### Arguments

<code>data</code>	data to be analyzed
<code>classificationVars</code>	list of variables to form right hand side of formula for <code>xtabs</code>
<code>freqVars</code>	optional list of variables to form left hand side of formula for <code>xtabs</code>
<code>wantPercentages</code>	include percentages in the output
<code>wantTotals</code>	include marginals in the output
<code>wantStats</code>	include chi-squared statistics in the output
<code>wantExtraTables</code>	include results for aggregated tables
<code>HTMLfile</code>	file argument to pass to HTML
<code>...</code>	arguments to pass to <code>xtabs</code>

### Value

None. Writes HTML output to designated file.

### Author(s)

Micah Altman (vdc-devlatte.harvard.edu) <http://thedata.org>

### See Also

[xtabs](#), [VDCxtabs](#)

### Examples

```
VDCcrossTabulation(as.data.frame(Titanic), classificationVars=c("Class", "Sex", "Age", "Survived"))
```

---

VDCgenAnalysis      *A generalized analysis framework used by the VDC system*

---

## Description

This runs a statistical analysis through Zelig and summarizes the results as html output. Options include simulation results, sensitivity analysis, and plots.

## Usage

```
VDCgenAnalysis (
  formula, model, data, by=NULL,
  setxArgs=NULL, setx2Args=NULL, simArgs=NULL, zeligArgs=NULL,
  sensitivityArgs=list("ptb.s"=sensitivityQ),
  HTMLInitArgs=list(Title="VDC Analysis"),
  HTMLbodyHeading = HTMLInitArgs$Title,
  HTMLnote= "<em>The following are the results of your requested analysis.</em>",
  naMethod="exclude",
  sensitivityQ=NULL,
  outDir=tempdir(),
  wantSummary=T, wantPlots=F, wantSensitivity=F, wantSim=F,
  wantBinOutput=F,
  debug=F)
```

## Arguments

formula	formula to be passed to Zelig
model	model to be passed to Zelig
data	data for analysis, passed to Zelig
by	strata, passed to Zelig
setxArgs	argument list to pass to setx
setx2Args	argument list, if present, setx is called a second time
simArgs	argument list, passed to sim
zeligArgs	argument list, passed to zelig, which in turn passes any non-zelig arguments to the model function itself
sensitivityArgs	argument list, passed to pzelig
HTMLInitArgs	argument list, passed to HTMLInitFile
HTMLbodyHeading	main heading for html page
HTMLnote	note text below main heading
naMethod	if omit na.omit is used, if exclude na.exclude is used
sensitivityQ	q-value passed to sensitivity analysis

outDir	directory in which to create output files
wantSummary	Flag, summary output wanted
wantPlots	Flag, plot output wanted
wantSensitivity	Flag, sensitivity analysis wanted
wantSim	Flag, simulation output wanted
wantBinOutput	Flag, binary output wanted, in addition to html
debug	debugging flag

**Value**

The paths to html index file and the binary save file.

**Note**

Only formula, model and data are required. Everything else can be defaulted. And in the formula if you use `list()` to enclose multiple outcome variables the appropriate zelig special will automatically be supplied.

VDCxtabs will call `recodeVDCdf` to apply VDC metadata attributes for labels, and missing values.

**Author(s)**

Micah Altman ([vdc-devlatte.harvard.edu](mailto:vdc-devlatte.harvard.edu)) <http://thedata.org>

**See Also**

[hplot](#), [HTML](#), [pzelig](#), [zelig](#),

**Examples**

```
# using required fields and defaults, output will be summary html
VDCgenAnalysis(Employed~Year,"ls",longley)

# shows additional output options, passing arguments to sim()
# and through zeligArgs to lm()

# should be Employed~. but Zelig 2.52 currently has a bug in setx wrt to dots
VDCgenAnalysis(Employed~GNP.deflator+GNP+Unemployed+Armed.Forces+Population+Year,"ls",longley,
  wantPlots=TRUE, wantSim=TRUE, wantBinOutput=TRUE, wantSensitivity=TRUE,
  naMethod="omit",
  sensitivityQ=0.95,
  simArgs = list(num=c(2000,200)),
  setxArgs = list(Year=1970),
  setx2Args = list(Year=1990),
  zeligArgs = list(singular.ok=TRUE)
)
```

---

`VDCWebSearch`*Launch a search for data through the VDC network*

---

## Description

This launches a search for data files contained in the VDC network. The VDC network offers both public and private data, most available in R binary format through the VDC extraction interface.

## Usage

```
VDCWebSearch(search="", host="vdc.hmhc.harvard.edu")
```

## Arguments

<code>search</code>	terms to search for , see below for examples
<code>host</code>	VDC host to search

## Value

Opens a web browser with the search results.

## Author(s)

Micah Altman (thedata-userslists.sourceforge.net) <http://thedata.org>

## See Also

[zelig](#), [printZeligSchemaInstance](#)

## Examples

```
## Not run:
# search for election data
VDCWebSearch('elections')
# can use booleans, phrases
VDCWebSearch('elections AND "united states"')
# can also use fielded searches
VDCWebSearch('author=Altman OR author=King')
# search your local host
VDCWebSearch('author=Colby', host="murraydata.hmhc.harvard.edu")
## End (Not run)
```

---

`VDCxtabs`*VDC cross tabs, with pretty printing*

---

**Description**

This calculate cross tabs with marginals and proportions.

**Usage**

```
VDCxtabs(formula, data=parent.env(), ...)
```

**Arguments**

<code>data</code>	data to be analyzed
<code>formula</code>	crosstab formula suitable for <code>xtabs</code>
<code>...</code>	arguments to pass to <code>xtabs</code>

**Value**

Returns a crosstabulation, essentially a combination of `xtabs` with proportions and marginals.

**Note**

VDCxtabs will call `recodeVDCdf` to apply VDC metadata attributes for labels, and missing values.

**Author(s)**

Micah Altman ([vdc-devlatte.harvard.edu](mailto:vdc-devlatte.harvard.edu)) <http://thedata.org>

**See Also**

[xtabs](#), [recodeVDCdf](#), [VDCcrossTabulation](#)

**Examples**

```
VDCxtabs(Freq~., as.data.frame(Titanic))
```

---

`zeligModelDependencies`*Derivative Zelig interface functions*

---

**Description**

Zelig interface functions. Used by VDC DSB to communicate with Zelig.

**Usage**

```
zeligModelDependencies (inZeligOnly=T, schemaVersion="1.1", uninstalledOnly=T,
```

**Arguments**

<code>inZeligOnly</code>	Flag, include only models in official Zelig distribution
<code>uninstalledOnly</code>	Flag, include only uninstalled models
<code>repos</code>	URL of default repository to use
<code>schemaVersion</code>	version of Zelig schema

**Value**

Use `zeligModelDependencies` to generate a list of package dependencies for models.

**Author(s)**

Micah Altman (thedata-userslists.sourceforge.net) <http://thedata.org>

**See Also**

[zelig](#), [printZeligSchemaInstance](#)

**Examples**

```
zeligModelDependencies (uninstalledOnly=FALSE)
```

# Index

## \*Topic **IO**

- hplot, 1
- HTML.summary.zelig, 2
- HTML.VDCxtabs, 3
- print.VDCxtabs, 4
- printZeligSchemaInstance, 5
- recodeVDCdf, 6
- VDCcrossTabulation, 7
- VDCgenAnalysis, 8
- VDCWebSearch, 10
- VDCxtabs, 11
- zeligModelDependencies, 12

## \*Topic **iplot**

- hplot, 1
- HTML.summary.zelig, 2
- HTML.VDCxtabs, 3
- print.VDCxtabs, 4
- recodeVDCdf, 6
- VDCgenAnalysis, 8

## \*Topic **print**

- hplot, 1
- HTML.summary.zelig, 2
- HTML.VDCxtabs, 3
- print.VDCxtabs, 4
- printZeligSchemaInstance, 5
- recodeVDCdf, 6
- VDCcrossTabulation, 7
- VDCgenAnalysis, 8
- VDCWebSearch, 10
- VDCxtabs, 11
- zeligModelDependencies, 12

bitmap, 2

hplot, 1, 9  
HTML, 3, 4, 9  
HTML.summary.MCMCZelig  
    (HTML.summary.zelig), 2  
HTML.summary.zelig, 2  
HTML.VDCxtabs, 3

HTMLInsertGraph, 2

plot, 2  
print.VDCxtabs, 4  
printZeligSchemaInstance, 5, 10, 12  
pzelig, 9

recodeVDCdf, 6, 11

sim, 3

VDCcrossTabulation, 7, 11  
VDCgenAnalysis, 7, 8  
VDCWebSearch, 10  
VDCxtabs, 4, 5, 8, 11

xtabs, 8, 11

zelig, 6, 9, 10, 12  
zeligModelDependencies, 12