

Easy Data.Frame

Smart Labels for R

Huashan Chen

2015年12月19日

- 1 Easy Data.Frame
- 2 标签的存储
 - 2.1 变量标签
 - 2.2 数值标签
- 3 制表函数
 - 3.1 tbl()
 - 3.2 ctbl()
 - 3.3 ftable()
- 4 选项

1 Easy Data.Frame

ezdf 的目的是使 R 支持类似 SPSS 或 Stata 那样的标签输出。Easy data.frame 包的目的并不是定义一套新的制表函数，而是控制相关制表函数（如 pander）在输出时，能够自动带上对应的标签。除此之外，Easy data.frame 也封装了几个常用的制表方法。

众所周知，在 R 体系中，并无变量标签或者数值标签的定义。在 R 中导入 SPSS 或 Stata 等传统统计软件的数据格式可有多个包来实现，例如 foreign、readStata13、haven、sas7bdat 等等。这些包在导入数据时，都能原数据中所定义的标签。然而所有这些包目前来说各有优缺点，即使对同一个格式也做不到支持各个版本的导入，因此难以提供一揽子解决方案。并且，各个包导入数据之后所定义的标签属性的名称也是各不相同，更不必谈在制作表格或者统计结果输出时，能够让 R 做到标签友好。

本包目前支持采用 foreign 或 haven 导入 Stata 以及 SPSS 格式数据。在 haven 包中，导入数据后，变量标签的定义位于数据库每列字段的 label 属性当中，数值标签的定义在每列字段的 labels 属性当中。在 foreign 包中，导入数据后，变量标签的定义在数据框的 variable.labels 属性当中，数值标签则由参数 use.value.labels 控制是否转化为 factor。

2 标签的存储

2.1 变量标签

Easy data.frame 定义了一个继承自 data.frame 的类 ez.data.frame，并将变量标签存储在 meta 属性当中。meta 可为 data.frame 或 matrix 类，且至少包括两列字段，其中第一列为变量名，第二列为变量标签。

若要创建一个新的 ez.data.frame 对象，可用 as.ez(dt, meta)，例如：

```
data(iris)
library(ezdf)
dl = as.ez(iris)
class(dl)
## [1] "ez.data.frame" "ezdf"          "data.table"    "data.frame"
```

示例的 iris 数据集并不包含任何标签信息，可以用 setmeta() 方法设置自定义的标签。meta 参数必须是至少包括两列的 data.frame 对象，其第一列被当作变量名，第二列被当作变量标签。setmeta() 将清除数据框中原有的变量标签设置，并返回一个 ezdf 对象。

```
d1$test = sample(5, size = nrow(iris), replace = T)
setmeta(d1, data.frame(var= 'test', lbl = 'Test VAR'))
attr(,"meta")
##      var      lbl
## 1: test Test VAR
```

若要获取数据框中某些变量的标签，可用 `varLabels()` 方法。

```
varLabels(ezdt, c('Species','test'))
## Error in varLabels(ezdt, c("Species", "test")): object 'ezdt' not found

# 用 default = "var" 控制空标签的输出为变量名。该参数默认值为 "", 即输出空标签。
varLabels(ezdt, c('Species','test'), default = "var")
## Error in varLabels(ezdt, c("Species", "test"), default = "var"): object
'ezdt' not found

# 返回所有已定义的变量标签
varLabels(ezdt)
## Error in varLabels(ezdt): object 'ezdt' not found

# 设置变量标签
varLabels(ezdt, "test") <- "New Label"
## Error in varLabels(ezdt, "test") <- "New Label": object 'ezdt' not found
varLabels(ezdt, "test")
## Error in varLabels(ezdt, "test"): object 'ezdt' not found
```

2.2 数值标签

数值标签的存储采用 `haven` 的格式，即数据框中的数值型字段可有一个属性 `labels`，其数据类型为命名的数值型，例如

```
c(V1 = 1, V2 = 2, V3 = 3, MI = 9)
..
```

注意，对于在非数值型字段（例如：字符型、factor、逻辑型等）使用的 ``labels`` 属性，``ezdf`` 并不能在表格输出时生成调用数值标签。

数值标签的获取和设置 ``valueLabels()`` 方法，例如：

```
v11 = valueLabels(ezdt, 'test')
## Error in valueLabels(ezdt, "test"): object 'ezdt' not found
v12 = v11 + c(MI=9, MM=8)
## Error in eval(expr, envir, enclos): object 'v11' not found
valueLabels(ezdt, 'test') = v12
## Error in eval(expr, envir, enclos): object 'v12' not found
```

3 制表函数

3.1 `tbl()`

生长表统计量，命令为 `tbl(ez, expr, func = 'mean', N = FALSE, sort = TRUE)`，其中 `expr` 为 formula 对象，当公式右端只有一个自变量时，忽略 `func` 参数，输出频次统计量；`func` 为需要采用的统计量，例如 `mean`、`sum`、`var`。

```
tbl(d1, ~test)
##      Test VAR  N
## 1:      1 37
## 2:      2 29
## 3:      3 24
## 4:      4 31
## 5:      5 29

# 分组求均值, 添加样本数
tbl(d1, Sepal.Length ~ Species + test, 'mean', N = T)
##      Species Test VAR Sepal.Length  N
## 1:      setosa      1   5.120000  10
## 2:      setosa      2   4.966667   9
## 3:      setosa      3   5.027273  11
## 4:      setosa      4   4.927273  11
## 5:      setosa      5   4.988889   9
## 6: versicolor      1   5.875000  16
## 7: versicolor      2   5.872727  11
## 8: versicolor      3   5.866667   6
## 9: versicolor      4   5.970000  10
## 10: versicolor      5   6.185714   7
## 11: virginica      1   6.963636  11
## 12: virginica      2   6.466667   9
## 13: virginica      3   6.300000   7
## 14: virginica      4   6.680000  10
## 15: virginica      5   6.438462  13

# tbl() 默认按照公式右端 x 的值排序, 如果取消排序
tbl(d1, Sepal.Length ~ Species + test, 'mean', N = T, sort = F)
##      Species Test VAR Sepal.Length  N
## 1:      setosa      2   4.966667   9
## 2:      setosa      5   4.988889   9
## 3:      setosa      3   5.027273  11
## 4:      setosa      1   5.120000  10
## 5:      setosa      4   4.927273  11
## 6: versicolor      1   5.875000  16
## 7: versicolor      2   5.872727  11
## 8: versicolor      5   6.185714   7
## 9: versicolor      4   5.970000  10
## 10: versicolor      3   5.866667   6
## 11: virginica      1   6.963636  11
## 12: virginica      5   6.438462  13
## 13: virginica      4   6.680000  10
## 14: virginica      3   6.300000   7
## 15: virginica      2   6.466667   9
```

3.2 ctbl()

ctbl() 是对 table() 的封装, 采用 ctbl(ez, expr) 的调用方式。

```
ctbl(d1, Sepal.Length ~ Species + test)
## , , test = 1
##
##      Species
## Sepal.Length setosa versicolor virginica
```

```

##      4.3      0      0      0
##      4.4      1      0      0
##      4.5      1      0      0
##      4.6      0      0      0
##      4.7      0      0      0
##      4.8      0      0      0
##      4.9      1      0      0
##      5      1      1      0
##      5.1      1      1      0
##      5.2      0      0      0
##      5.3      1      0      0
##      5.4      2      1      0
##      5.5      1      1      0
##      5.6      0      2      0
##      5.7      1      2      0
##      5.8      0      2      0
##      5.9      0      0      0
##      6      0      1      0
##      6.1      0      0      0
##      6.2      0      2      1
##      6.3      0      0      1
##      6.4      0      0      2
##      6.5      0      0      0
##      6.6      0      0      0
##      6.7      0      2      1
##      6.8      0      0      0
##      6.9      0      0      1
##      7      0      1      0
##      7.1      0      0      0
##      7.2      0      0      2
##      7.3      0      0      0
##      7.4      0      0      0
##      7.6      0      0      0
##      7.7      0      0      2
##      7.9      0      0      1
##
## , , test = 2
##
##      Species
## Sepal.Length setosa versicolor virginica
##      4.3      0      0      0
##      4.4      2      0      0
##      4.5      0      0      0
##      4.6      0      0      0
##      4.7      1      0      0
##      4.8      0      0      0
##      4.9      0      1      0
##      5      2      0      0
##      5.1      1      0      0
##      5.2      1      0      0
##      5.3      0      0      0
##      5.4      1      0      0
##      5.5      1      1      0
##      5.6      0      3      0
##      5.7      0      1      0

```

```

##      5.8      0      0      1
##      5.9      0      1      0
##      6       0      0      1
##      6.1      0      1      1
##      6.2      0      0      0
##      6.3      0      0      1
##      6.4      0      1      0
##      6.5      0      0      1
##      6.6      0      1      0
##      6.7      0      1      1
##      6.8      0      0      2
##      6.9      0      0      0
##      7       0      0      0
##      7.1      0      0      0
##      7.2      0      0      1
##      7.3      0      0      0
##      7.4      0      0      0
##      7.6      0      0      0
##      7.7      0      0      0
##      7.9      0      0      0
##
## , , test = 3
##
##      Species
## Sepal.Length setosa versicolor virginica
##      4.3      0      0      0
##      4.4      0      0      0
##      4.5      0      0      0
##      4.6      2      0      0
##      4.7      0      0      0
##      4.8      1      0      0
##      4.9      2      0      1
##      5       1      1      0
##      5.1      2      0      0
##      5.2      1      0      0
##      5.3      0      0      0
##      5.4      1      0      0
##      5.5      0      1      0
##      5.6      0      0      0
##      5.7      1      0      0
##      5.8      0      1      0
##      5.9      0      0      1
##      6       0      1      0
##      6.1      0      1      0
##      6.2      0      0      0
##      6.3      0      0      1
##      6.4      0      0      1
##      6.5      0      0      1
##      6.6      0      0      0
##      6.7      0      0      1
##      6.8      0      1      0
##      6.9      0      0      0
##      7       0      0      0
##      7.1      0      0      0
##      7.2      0      0      0

```

```

##      7.3      0      0      0
##      7.4      0      0      1
##      7.6      0      0      0
##      7.7      0      0      0
##      7.9      0      0      0
##
## , , test = 4
##
##      Species
## Sepal.Length setosa versicolor virginica
##      4.3      1      0      0
##      4.4      0      0      0
##      4.5      0      0      0
##      4.6      2      0      0
##      4.7      0      0      0
##      4.8      2      0      0
##      4.9      0      0      0
##      5       2      0      0
##      5.1      3      0      0
##      5.2      0      1      0
##      5.3      0      0      0
##      5.4      0      0      0
##      5.5      0      2      0
##      5.6      0      0      0
##      5.7      0      0      0
##      5.8      1      0      1
##      5.9      0      1      0
##      6       0      1      0
##      6.1      0      2      0
##      6.2      0      0      1
##      6.3      0      1      2
##      6.4      0      0      1
##      6.5      0      1      1
##      6.6      0      1      0
##      6.7      0      0      1
##      6.8      0      0      0
##      6.9      0      0      0
##      7       0      0      0
##      7.1      0      0      0
##      7.2      0      0      0
##      7.3      0      0      1
##      7.4      0      0      0
##      7.6      0      0      1
##      7.7      0      0      1
##      7.9      0      0      0
##
## , , test = 5
##
##      Species
## Sepal.Length setosa versicolor virginica
##      4.3      0      0      0
##      4.4      0      0      0
##      4.5      0      0      0
##      4.6      0      0      0
##      4.7      1      0      0

```

```
##      4.8      2      0      0
##      4.9      1      0      0
##      5       2      0      0
##      5.1      1      0      0
##      5.2      1      0      0
##      5.3      0      0      0
##      5.4      1      0      0
##      5.5      0      0      0
##      5.6      0      0      1
##      5.7      0      2      1
##      5.8      0      0      1
##      5.9      0      0      0
##      6       0      1      1
##      6.1      0      0      1
##      6.2      0      0      0
##      6.3      0      2      1
##      6.4      0      1      1
##      6.5      0      0      1
##      6.6      0      0      0
##      6.7      0      0      1
##      6.8      0      0      0
##      6.9      0      1      2
##      7       0      0      0
##      7.1      0      0      1
##      7.2      0      0      0
##      7.3      0      0      0
##      7.4      0      0      0
##      7.6      0      0      0
##      7.7      0      0      1
##      7.9      0      0      0
# 等价于
table(d1$Sepal.Length, d1$Species, d1$test)
## , , d1$test = 1
##
##      d1$Species
## d1$Sepal.Length setosa versicolor virginica
##      4.3      0      0      0
##      4.4      1      0      0
##      4.5      1      0      0
##      4.6      0      0      0
##      4.7      0      0      0
##      4.8      0      0      0
##      4.9      1      0      0
##      5       1      1      0
##      5.1      1      1      0
##      5.2      0      0      0
##      5.3      1      0      0
##      5.4      2      1      0
##      5.5      1      1      0
##      5.6      0      2      0
##      5.7      1      2      0
##      5.8      0      2      0
##      5.9      0      0      0
##      6       0      1      0
##      6.1      0      0      0
```

```

##      6.2      0      2      1
##      6.3      0      0      1
##      6.4      0      0      2
##      6.5      0      0      0
##      6.6      0      0      0
##      6.7      0      2      1
##      6.8      0      0      0
##      6.9      0      0      1
##      7       0      1      0
##      7.1      0      0      0
##      7.2      0      0      2
##      7.3      0      0      0
##      7.4      0      0      0
##      7.6      0      0      0
##      7.7      0      0      2
##      7.9      0      0      1
##
## , , dl$test = 2
##
##      dl$Species
## dl$Sepal.Length setosa versicolor virginica
##      4.3      0      0      0
##      4.4      2      0      0
##      4.5      0      0      0
##      4.6      0      0      0
##      4.7      1      0      0
##      4.8      0      0      0
##      4.9      0      1      0
##      5       2      0      0
##      5.1      1      0      0
##      5.2      1      0      0
##      5.3      0      0      0
##      5.4      1      0      0
##      5.5      1      1      0
##      5.6      0      3      0
##      5.7      0      1      0
##      5.8      0      0      1
##      5.9      0      1      0
##      6       0      0      1
##      6.1      0      1      1
##      6.2      0      0      0
##      6.3      0      0      1
##      6.4      0      1      0
##      6.5      0      0      1
##      6.6      0      1      0
##      6.7      0      1      1
##      6.8      0      0      2
##      6.9      0      0      0
##      7       0      0      0
##      7.1      0      0      0
##      7.2      0      0      1
##      7.3      0      0      0
##      7.4      0      0      0
##      7.6      0      0      0
##      7.7      0      0      0

```



```
##          7.9      0      0      0
##
## , , d1$test = 3
##
##          d1$Species
## d1$Sepal.Length setosa versicolor virginica
##          4.3      0      0      0
##          4.4      0      0      0
##          4.5      0      0      0
##          4.6      2      0      0
##          4.7      0      0      0
##          4.8      1      0      0
##          4.9      2      0      1
##          5      1      1      0
##          5.1      2      0      0
##          5.2      1      0      0
##          5.3      0      0      0
##          5.4      1      0      0
##          5.5      0      1      0
##          5.6      0      0      0
##          5.7      1      0      0
##          5.8      0      1      0
##          5.9      0      0      1
##          6      0      1      0
##          6.1      0      1      0
##          6.2      0      0      0
##          6.3      0      0      1
##          6.4      0      0      1
##          6.5      0      0      1
##          6.6      0      0      0
##          6.7      0      0      1
##          6.8      0      1      0
##          6.9      0      0      0
##          7      0      0      0
##          7.1      0      0      0
##          7.2      0      0      0
##          7.3      0      0      0
##          7.4      0      0      1
##          7.6      0      0      0
##          7.7      0      0      0
##          7.9      0      0      0
##
## , , d1$test = 4
##
##          d1$Species
## d1$Sepal.Length setosa versicolor virginica
##          4.3      1      0      0
##          4.4      0      0      0
##          4.5      0      0      0
##          4.6      2      0      0
##          4.7      0      0      0
##          4.8      2      0      0
##          4.9      0      0      0
##          5      2      0      0
##          5.1      3      0      0
```

```
##      5.2      0      1      0
##      5.3      0      0      0
##      5.4      0      0      0
##      5.5      0      2      0
##      5.6      0      0      0
##      5.7      0      0      0
##      5.8      1      0      1
##      5.9      0      1      0
##      6      0      1      0
##      6.1      0      2      0
##      6.2      0      0      1
##      6.3      0      1      2
##      6.4      0      0      1
##      6.5      0      1      1
##      6.6      0      1      0
##      6.7      0      0      1
##      6.8      0      0      0
##      6.9      0      0      0
##      7      0      0      0
##      7.1      0      0      0
##      7.2      0      0      0
##      7.3      0      0      1
##      7.4      0      0      0
##      7.6      0      0      1
##      7.7      0      0      1
##      7.9      0      0      0
##
## , , dl$test = 5
##
##              dl$Species
## dl$Sepal.Length setosa versicolor virginica
##      4.3      0      0      0
##      4.4      0      0      0
##      4.5      0      0      0
##      4.6      0      0      0
##      4.7      1      0      0
##      4.8      2      0      0
##      4.9      1      0      0
##      5      2      0      0
##      5.1      1      0      0
##      5.2      1      0      0
##      5.3      0      0      0
##      5.4      1      0      0
##      5.5      0      0      0
##      5.6      0      0      1
##      5.7      0      2      1
##      5.8      0      0      1
##      5.9      0      0      0
##      6      0      1      1
##      6.1      0      0      1
##      6.2      0      0      0
##      6.3      0      2      1
##      6.4      0      1      1
##      6.5      0      0      1
##      6.6      0      0      0
```

```
##          6.7      0      0      1
##          6.8      0      0      0
##          6.9      0      1      2
##          7      0      0      0
##          7.1      0      0      1
##          7.2      0      0      0
##          7.3      0      0      0
##          7.4      0      0      0
##          7.6      0      0      0
##          7.7      0      0      1
##          7.9      0      0      0
```

3.3 ftable()

`ftable.ez.data.frame` 方法是对 `ftable()` 的封装，其调用方式为 `ftable(ez, formula, style = 1, prop_margin = 1, ...)`。其中 `style = 1` 输出频次；`style = 2` 输出百分比，由 `prop_margin` 参数指定行百分比还是列百分比；`style = 3` 输出百分比和行加总频次。

```
ftable(d1, Species~test)
##
##      setosa versicolor virginica
## 1      10         16         11
## 2       9          11          9
## 3      11          6          7
## 4      11         10         10
## 5       9          7          13
ftable(d1, Species~test, style = 2)
##
##      setosa versicolor virginica
## 1 0.2702703 0.4324324 0.2972973
## 2 0.3103448 0.3793103 0.3103448
## 3 0.4583333 0.2500000 0.2916667
## 4 0.3548387 0.3225806 0.3225806
## 5 0.3103448 0.2413793 0.4482759
(t1 = ftable(d1, Species~test, style = 3))
##      setosa versicolor virginica N
## 1 0.2702703 0.4324324 0.2972973 37
## 2 0.3103448 0.3793103 0.3103448 29
## 3 0.4583333 0.2500000 0.2916667 24
## 4 0.3548387 0.3225806 0.3225806 31
## 5 0.3103448 0.2413793 0.4482759 29

# pander 输出
pander::pander(t1, ez = d1)
```

setosa	versicolor	virginica	N
0.2703	0.4324	0.2973	37
0.3103	0.3793	0.3103	29
0.4583	0.25	0.2917	24
0.3548	0.3226	0.3226	31
0.3103	0.2414	0.4483	29

4 选项

目前支持下述三个选项:

- `options('ezdfKeepVal' = T)`
- `options('ezdfValueLabelSep' = '=')`
- `options('ezdfKeepVarName' = T)`

```
options('ezdfKeepVal' = T)
options('ezdfValueLabelSep' = '=')
#options('ezdfKeepVarName' = T)
```

```
tbl(d1, ~test)
##      Test VAR  N
## 1:      1 37
## 2:      2 29
## 3:      3 24
## 4:      4 31
## 5:      5 29
```