

Easy Data.Frame

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

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包来实现，例如 `foreign`、`readStata13`、`haven`、`sas7bdat` 等等。这些包在导入数据时，或多或少保留了部分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，各个包所定义的标签其用法也各有不同，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，变量标签的定义在每列字段的 `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。

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

```
data(iris)
library(ezdf)
d1 = as.ez(iris)
class(d1)
## [1] "ez.data.frame" "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

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

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，在非数值型字段（例如：字符型、`factor`、逻辑型等）应用 ``labels`` 属性，并不会在输出时自动调用数值标签。

设置数值标签可用 ``set_value_labels()`` 方法，例如：

```
set_value_labels(d1, 'test', c(V1 = 1, V2 = 2, V3 = 3, V3 = 3))
```

3 制表函数

3.1 `tbl()`

生成长表统计量，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:      V1 26
## 2:      V2 24
## 3:      V3 32
## 4:           32
## 5:           36

# 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   V1   5.027273 11
## 2:      setosa   V2   5.240000  5
## 3:      setosa   V3   5.120000 10
## 4:      setosa           4.963636 11
## 5:      setosa           4.846154 13
## 6: versicolor   V1   5.800000  8
## 7: versicolor   V2   5.960000 10
## 8: versicolor   V3   5.950000 10
## 9: versicolor           6.033333 12
## 10: versicolor           5.890000 10
## 11: virginica   V1   6.828571  7
## 12: virginica   V2   6.388889  9
## 13: virginica   V3   6.383333 12
## 14: virginica           6.711111  9
## 15: virginica           6.700000 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   V3   5.120000 10
## 2:      setosa           4.963636 11
## 3:      setosa           4.846154 13
## 4:      setosa   V2   5.240000  5
## 5:      setosa   V1   5.027273 11
## 6: versicolor   V1   5.800000  8
## 7: versicolor           6.033333 12
## 8: versicolor   V2   5.960000 10
## 9: versicolor   V3   5.950000 10
## 10: versicolor           5.890000 10
## 11: virginica   V1   6.828571  7
## 12: virginica   V3   6.383333 12
## 13: virginica           6.711111  9
## 14: virginica   V2   6.388889  9
## 15: virginica           6.700000 13
```

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)
## , , Sepal.Length = 4.3
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 1 0
## versicolor 0 0 0 0 0
## virginica  0 0 0 0 0
##
## , , Sepal.Length = 4.4
##
##          test
## Species    1 2 3 4 5
## setosa     1 0 1 0 1
## versicolor 0 0 0 0 0
## virginica  0 0 0 0 0
##
## , , Sepal.Length = 4.5
##
##          test
## Species    1 2 3 4 5
## setosa     1 0 0 0 0
## versicolor 0 0 0 0 0
## virginica  0 0 0 0 0
##
## , , Sepal.Length = 4.6
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 1 3
## versicolor 0 0 0 0 0
## virginica  0 0 0 0 0
##
## , , Sepal.Length = 4.7
##
##          test
## Species    1 2 3 4 5
## setosa     1 0 0 1 0
## versicolor 0 0 0 0 0
## virginica  0 0 0 0 0
##
## , , Sepal.Length = 4.8
##
##          test
## Species    1 2 3 4 5
## setosa     0 1 1 1 2
## versicolor 0 0 0 0 0
## virginica  0 0 0 0 0
##
## , , Sepal.Length = 4.9
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 1 3
## versicolor 1 0 0 0 0
## virginica  0 1 0 0 0
##
## , , Sepal.Length = 5
##
##          test
## Species    1 2 3 4 5
## setosa     4 0 1 1 2
## versicolor 1 0 1 0 0
## virginica  0 0 0 0 0
##
## , , Sepal.Length = 5.1

```

```

##
##          test
## Species    1 2 3 4 5
##  setosa    0 1 4 2 1
##  versicolor 0 0 1 0 0
##  virginica  0 0 0 0 0
##
## , , Sepal.Length = 5.2
##
##          test
## Species    1 2 3 4 5
##  setosa    1 0 0 2 0
##  versicolor 0 1 0 0 0
##  virginica  0 0 0 0 0
##
## , , Sepal.Length = 5.3
##
##          test
## Species    1 2 3 4 5
##  setosa    1 0 0 0 0
##  versicolor 0 0 0 0 0
##  virginica  0 0 0 0 0
##
## , , Sepal.Length = 5.4
##
##          test
## Species    1 2 3 4 5
##  setosa    0 2 2 0 1
##  versicolor 0 0 0 1 0
##  virginica  0 0 0 0 0
##
## , , Sepal.Length = 5.5
##
##          test
## Species    1 2 3 4 5
##  setosa    1 1 0 0 0
##  versicolor 1 1 0 0 3
##  virginica  0 0 0 0 0
##
## , , Sepal.Length = 5.6
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 0 0 0
##  versicolor 2 1 1 0 1
##  virginica  0 0 1 0 0
##
## , , Sepal.Length = 5.7
##
##          test
## Species    1 2 3 4 5
##  setosa    1 0 0 1 0
##  versicolor 0 0 2 1 2
##  virginica  0 0 0 1 0
##
## , , Sepal.Length = 5.8
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 1 0 0
##  versicolor 0 1 0 2 0
##  virginica  0 1 2 0 0
##
## , , Sepal.Length = 5.9
##
##          test
## Species    1 2 3 4 5

```

```

## setosa      0 0 0 0 0
## versicolor 0 0 0 2 0
## virginica  0 1 0 0 0
##
## , , Sepal.Length = 6
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 0 0
## versicolor 0 3 0 0 1
## virginica  0 0 0 0 2
##
## , , Sepal.Length = 6.1
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 0 0
## versicolor 0 0 1 2 1
## virginica  0 0 1 0 1
##
## , , Sepal.Length = 6.2
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 0 0
## versicolor 1 1 0 0 0
## virginica  1 0 1 0 0
##
## , , Sepal.Length = 6.3
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 0 0
## versicolor 0 0 1 2 0
## virginica  2 0 2 2 0
##
## , , Sepal.Length = 6.4
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 0 0
## versicolor 0 1 0 1 0
## virginica  1 1 2 0 1
##
## , , Sepal.Length = 6.5
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 0 0
## versicolor 0 0 1 0 0
## virginica  0 1 0 1 2
##
## , , Sepal.Length = 6.6
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 0 0
## versicolor 1 0 0 0 1
## virginica  0 0 0 0 0
##
## , , Sepal.Length = 6.7
##
##          test
## Species    1 2 3 4 5
## setosa     0 0 0 0 0
## versicolor 0 0 1 1 1
## virginica  0 2 0 2 1

```

```

##
## , , Sepal.Length = 6.8
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 0 0 0
##  versicolor 0 0 1 0 0
##  virginica  0 0 1 0 1
##
## , , Sepal.Length = 6.9
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 0 0 0
##  versicolor 0 1 0 0 0
##  virginica  0 1 0 0 2
##
## , , Sepal.Length = 7
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 0 0 0
##  versicolor 1 0 0 0 0
##  virginica  0 0 0 0 0
##
## , , Sepal.Length = 7.1
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 0 0 0
##  versicolor 0 0 0 0 0
##  virginica  0 0 0 1 0
##
## , , Sepal.Length = 7.2
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 0 0 0
##  versicolor 0 0 0 0 0
##  virginica  0 0 1 0 2
##
## , , Sepal.Length = 7.3
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 0 0 0
##  versicolor 0 0 0 0 0
##  virginica  1 0 0 0 0
##
## , , Sepal.Length = 7.4
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 0 0 0
##  versicolor 0 0 0 0 0
##  virginica  0 0 0 1 0
##
## , , Sepal.Length = 7.6
##
##          test
## Species    1 2 3 4 5
##  setosa    0 0 0 0 0
##  versicolor 0 0 0 0 0
##  virginica  1 0 0 0 0
##
## , , Sepal.Length = 7.7
##

```

```

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

```



```

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

```

```

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

```

```
##          6.5      0      0      2
##          6.6      0      1      0
##          6.7      0      1      1
##          6.8      0      0      1
##          6.9      0      0      2
##          7       0      0      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      0
##          7.9      0      0      1
```

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
## V1      11         8         7
## V2       5        10         9
## V3      10        10        12
##          11        12         9
##          13        10        13
ftable(d1, Species~test, style = 2)
##
##      setosa versicolor virginica
## V1 0.4230769 0.3076923 0.2692308
## V2 0.2083333 0.4166667 0.3750000
## V3 0.3125000 0.3125000 0.3750000
##      0.3437500 0.3750000 0.2812500
##      0.3611111 0.2777778 0.3611111
(t1 = ftable(d1, Species~test, style = 3))
##      setosa versicolor virginica N
## V1 0.4230769 0.3076923 0.2692308 26
## V2 0.2083333 0.4166667 0.3750000 24
## V3 0.3125000 0.3125000 0.3750000 32
##      0.3437500 0.3750000 0.2812500 32
##      0.3611111 0.2777778 0.3611111 36

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

nchar width > width 1979 3, [32] nchar width > width 1358954515 3, [36]

	setosa	versicolor	virginica	N
V1	0.4231	0.3077	0.2692	26
V2	0.2083	0.4167	0.375	24
V3	0.3125	0.3125	0.375	32

	setosa	versicolor	virginica	N
0.343	8 0.	375 0.	2812 32	
0.361	1 0.2	778 0.	3611 36	

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=V1 26
## 2:      2=V2 24
## 3:      3=V3 32
## 4:      4=  32
## 5:      5=  36
```