

Easy Data.Frame

Huashan Chen

2015年12月19日

一	Easy Data.Frame	2
二	标签的存储	3
2.1	变量标签	3
2.2	数值标签	4
三	制表函数	5
3.1	tbl()	5
3.2	ctbl()	6
3.3	ftable()	17
四	选项	19

第一章 Easy DataFrame

`ezdf` 的目的是使 R 支持类似 SPSS 或 Stata 那样的标签输出。`Easy data.frame` 包的目的是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`。

第二章 标签的存储

第2.1节 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:

```
1 data(iris)
2 library(ezdf)
3 d1 = as.ez(iris)
4 class(d1)
5 ## [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。`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。

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

若要获取数据框中某些变量的标签, 可用 `varLabels()` 方法。

```
1 varLabels(d1, c('Species','test'))
2 ## [1] ""          "Test VAR"
3
4 # 用 default = "var" 控制空标签的输出为变量名。该参数默认值为 "",
5 即输出空标签。
6 varLabels(d1, c('Species','test'), default = "var")
7 ## [1] "Species"  "Test VAR"
8
9 # 返回所有已定义的变量标签
10 varLabels(d1)
11 ##      var      lbl
```

```
12     ## 1: test Test VAR
13
14     # 设置变量标签
15     varLabels(d1, "test") <- "New Label"
16     varLabels(d1, "test")
17     ## [1] "New Label"
```

第2.2节 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

```
1     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：

```
1     v11 = valueLabels(d1, 'test')
2     v12 = v11 + c("Class1"=1, "Class2"=2, "Class3"=3, MI=9, MM=8)
3     valueLabels(d1, 'test') = v12
```

第三章 制表函数

第3.1节 `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`。

```
1   tbl(d1, ~test)
2   ##      test\nNew Label  N
3   ## 1:      1 = Class1 29
4   ## 2:      2 = Class2 36
5   ## 3:      3 = Class3 26
6   ## 4:           4 =   26
7   ## 5:           5 =   33
8
9   # 分组求均值，添加样本数
10  tbl(d1, Sepal.Length ~ Species + test, 'mean', N = T)
11  ##      Species test\nNew Label Sepal.Length  N
12  ## 1:   setosa      1 = Class1      5.160000 10
13  ## 2:   setosa      2 = Class2      4.958333 12
14  ## 3:   setosa      3 = Class3      5.075000  8
15  ## 4:   setosa           4 =      4.966667  9
16  ## 5:   setosa           5 =      4.900000 11
17  ## 6: versicolor  1 = Class1      5.811111  9
18  ## 7: versicolor  2 = Class2      6.000000 10
19  ## 8: versicolor  3 = Class3      5.783333 12
20  ## 9: versicolor           4 =      5.983333  6
21  ## 10: versicolor           5 =      6.092308 13
22  ## 11: virginica  1 = Class1      6.710000 10
23  ## 12: virginica  2 = Class2      6.528571 14
24  ## 13: virginica  3 = Class3      6.550000  6
25  ## 14: virginica           4 =      6.727273 11
26  ## 15: virginica           5 =      6.400000  9
27
```

28

```
# 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\nNew Label Sepal.Length  N
##  1:      setosa      5 =      4.900000 11
##  2:      setosa      3 = Class3   5.075000  8
##  3:      setosa      2 = Class2   4.958333 12
##  4:      setosa      1 = Class1   5.160000 10
##  5:      setosa      4 =      4.966667  9
##  6: versicolor      1 = Class1   5.811111  9
##  7: versicolor      3 = Class3   5.783333 12
##  8: versicolor      5 =      6.092308 13
##  9: versicolor      4 =      5.983333  6
## 10: versicolor      2 = Class2   6.000000 10
## 11: virginica      2 = Class2   6.528571 14
## 12: virginica      3 = Class3   6.550000  6
## 13: virginica      4 =      6.727273 11
## 14: virginica      5 =      6.400000  9
## 15: virginica      1 = Class1   6.710000 10
```

第3.2节 `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      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      1          0          0
##           5        3          1          0
##           5.1      1          0          0
##           5.2      0          0          0
```

```

1    ##      5.3      1      0      0
    ##      5.4      2      0      0
    ##      5.5      0      1      0
    ##      5.6      0      2      0
    ##      5.7      1      2      0
    ##      5.8      0      1      0
    ##      5.9      0      0      1
    ##      6       0      0      0
    ##      6.1      0      0      0
    ##      6.2      0      0      0
    ##      6.3      0      0      2
    ##      6.4      0      1      0
    ##      6.5      0      0      2
    ##      6.6      0      0      0
    ##      6.7      0      0      1
    ##      6.8      0      0      1
    ##      6.9      0      0      0
    ##      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      1
    ##      7.9      0      0      0
    ##
    ## , , 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      1      0      0
    ##      4.7      0      0      0
    ##      4.8      2      0      0
    ##      4.9      1      0      1
    ##      5       3      0      0

```

```

2    ##      5.1      0      1      0
    ##      5.2      0      0      0
    ##      5.3      0      0      0
    ##      5.4      1      0      0
    ##      5.5      1      0      0
    ##      5.6      0      2      0
    ##      5.7      1      1      0
    ##      5.8      0      1      0
    ##      5.9      0      0      0
    ##      6       0      1      0
    ##      6.1      0      1      1
    ##      6.2      0      0      1
    ##      6.3      0      0      2
    ##      6.4      0      0      3
    ##      6.5      0      0      1
    ##      6.6      0      1      0
    ##      6.7      0      1      1
    ##      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      1
    ##      7.4      0      0      0
    ##      7.6      0      0      0
    ##      7.7      0      0      0
    ##      7.9      0      0      1
    ##
    ## , , 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      1      0      0
    ##      4.7      1      0      0
    ##      4.8      0      0      0

```



```

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

```

```

4    ##      4.7      0      0      0
    ##      4.8      1      0      0
    ##      4.9      1      1      0
    ##      5      0      0      0
    ##      5.1      3      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      1      1
    ##      5.7      0      0      0
    ##      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      1      0
    ##      6.4      0      0      0
    ##      6.5      0      1      1
    ##      6.6      0      1      0
    ##      6.7      0      0      1
    ##      6.8      0      0      0
    ##      6.9      0      0      1
    ##      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      1
    ##      7.7      0      0      2
    ##      7.9      0      0      0
    ##
    ## , , test = 5
    ##
    ##           Species
    ## Sepal.Length setosa versicolor virginica
    ##           4.3      1      0      0
    ##           4.4      1      0      0

```

```

5    ##      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      1      0      0
    ##      5.2      1      0      0
    ##      5.3      0      0      0
    ##      5.4      1      0      0
    ##      5.5      1      3      0
    ##      5.6      0      0      0
    ##      5.7      0      1      0
    ##      5.8      0      0      1
    ##      5.9      0      2      0
    ##      6       0      0      1
    ##      6.1      0      1      0
    ##      6.2      0      1      1
    ##      6.3      0      2      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      1      2
    ##      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      0
    ##      7.9      0      0      0

```

等价于

```
table(d1$Sepal.Length, d1$Species, d1$test)
```

```
## , , d1$test = 1
```

```
##
```

```
##           d1$Species
```

```

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

```

7

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

8

```
##
## , , 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      1      0      0
##           4.7      1      0      0
##           4.8      0      0      0
##           4.9      0      0      0
##           5        1      1      0
##           5.1      3      0      0
##           5.2      1      1      0
##           5.3      0      0      0
##           5.4      0      1      0
##           5.5      0      1      0
##           5.6      0      0      0
##           5.7      0      1      1
##           5.8      1      1      1
##           5.9      0      0      0
##           6        0      2      0
##           6.1      0      2      0
##           6.2      0      1      0
##           6.3      0      0      1
##           6.4      0      1      0
##           6.5      0      0      0
##           6.6      0      0      0
##           6.7      0      0      1
##           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      0
##           7.6      0      0      0
```

```

9    ##          7.7      0      0      1
    ##          7.9      0      0      0
    ##
    ## , , d1$test = 4
    ##
    ##          d1$Species
    ## d1$Sepal.Length setosa versicolor virginica
    ##          4.3      0      0      0
    ##          4.4      0      0      0
    ##          4.5      1      0      0
    ##          4.6      1      0      0
    ##          4.7      0      0      0
    ##          4.8      1      0      0
    ##          4.9      1      1      0
    ##          5      0      0      0
    ##          5.1      3      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      1      1
    ##          5.7      0      0      0
    ##          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      1      0
    ##          6.4      0      0      0
    ##          6.5      0      1      1
    ##          6.6      0      1      0
    ##          6.7      0      0      1
    ##          6.8      0      0      0
    ##          6.9      0      0      1
    ##          7      0      0      0
    ##          7.1      0      0      0
    ##          7.2      0      0      0
    ##          7.3      0      0      0

```

```

10  ##          7.4      0      0      1
    ##          7.6      0      0      1
    ##          7.7      0      0      2
    ##          7.9      0      0      0
    ##
    ## , , d1$test = 5
    ##
    ##          d1$Species
    ## d1$Sepal.Length setosa versicolor virginica
    ##          4.3      1      0      0
    ##          4.4      1      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      1      0      0
    ##          5.2      1      0      0
    ##          5.3      0      0      0
    ##          5.4      1      0      0
    ##          5.5      1      3      0
    ##          5.6      0      0      0
    ##          5.7      0      1      0
    ##          5.8      0      0      1
    ##          5.9      0      2      0
    ##          6      0      0      1
    ##          6.1      0      1      0
    ##          6.2      0      1      1
    ##          6.3      0      2      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      1      2
    ##          7      0      0      0
    ##          7.1      0      0      0

```



```

11  ##           7.2      0      0      0
    ##           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

```

第3.3节 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 = Class1      10           9         10
## 2 = Class2      12          10         14
## 3 = Class3       8          12          6
## 4 =              9           6         11
## 5 =             11          13          9
ftable(d1, Species~test, style = 2)
##
##           setosa versicolor virginica
## 1 = Class1 0.3448276 0.3103448 0.3448276
## 2 = Class2 0.3333333 0.2777778 0.3888889
## 3 = Class3 0.3076923 0.4615385 0.2307692
## 4 =         0.3461538 0.2307692 0.4230769
## 5 =         0.3333333 0.3939394 0.2727273
(t1 = ftable(d1, Species~test, style = 3))
##           setosa versicolor virginica N
## 1 = Class1 0.3448276 0.3103448 0.3448276 29
## 2 = Class2 0.3333333 0.2777778 0.3888889 36
## 3 = Class3 0.3076923 0.4615385 0.2307692 26
## 4 =         0.3461538 0.2307692 0.4230769 26
## 5 =         0.3333333 0.3939394 0.2727273 33

```

```
1 # pander 输出
  library(pander)
  pander(t1, ez = d1)
```

	setosa	versicolor	virginica	N
1 = Class1	0.3448	0.3103	0.3448	29
2 = Class2	0.3333	0.2778	0.3889	36
3 = Class3	0.3077	0.4615	0.2308	26
4 =	0.3462	0.2308	0.4231	26
5 =	0.3333	0.3939	0.2727	33

第四章 选项

目前支持下述三个选项:

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

```
1 options('ezdfKeepVal' = T)
2 options('ezdfValueLabelSep' = '=')
3 options('ezdfKeepVarName' = F)
4
5 tbl(d1, ~test)
6 ##      New Label  N
7 ## 1:  1=Class1 29
8 ## 2:  2=Class2 36
9 ## 3:  3=Class3 26
10 ## 4:         4= 26
11 ## 5:         5= 33
12 pander(tbl(d1, ~test))
```

New Label	N
1=Class1	29
2=Class2	36
3=Class3	26
4=	26
5=	33

```
1 options('ezdfKeepVarName' = T)
2 options('ezdfValueLabelSep' = '++')
3 pander(tbl(d1, ~test))
```

test New Label	N
1++Class1	29
2++Class2	36
3++Class3	26
4++	26
5++	33
