

Foundations of R Software

Lecture 34

Factors – Class and Unclass

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Factors

The **factor** function encodes the vector of discrete values into a factor.

Usage

```
factor(x = character(), levels, labels =  
levels, exclude = NA, ...)
```

levels : Determines the categories of the factor variable.

Default is the sorted list of all the distinct values of **x**.

labels : (Optional) Vector of values that will be the labels of the categories in the **levels** argument.

exclude : (Optional) It defines which levels will be classified as **NA** in any output using the factor variable.

Factors

A vector can be turned into a factor with the command `as.factor`:

```
> x = c(3, 4, 5, 6, 1, 2, 3, 3, 4, 4, 5, 6)
```

```
> x
```

```
[1] 3 4 5 6 1 2 3 3 4 4 5 6
```

```
> y = as.factor(x)
```

```
> y
```

```
[1] 3 4 5 6 1 2 3 3 4 4 5 6
```

```
Levels: 1 2 3 4 5 6
```

The single levels are ordered numerically:

```
1 -> 2 -> 3 -> 4 -> 5 -> 6
```

Factors

Example

```
> x = factor( c("lemonade", "lemonade",  
"juice", "lemonade", "water") )
```

```
> x
```

```
[1] lemonade  lemonade  juice  lemonade  water
```

```
Levels: juice lemonade water
```

The single levels are ordered alphabetically:

```
juice --- lemonade --- water
```

Factors

Example

```
R Console
> x = c(3, 4, 5, 6, 1, 2, 3, 3, 4, 4, 5, 6)
> x
[1] 3 4 5 6 1 2 3 3 4 4 5 6
> y = as.factor(x)
> y
[1] 3 4 5 6 1 2 3 3 4 4 5 6
Levels: 1 2 3 4 5 6
> |
```

```
R Console
> x = factor( c("lemonade", "lemonade", "juice", "lemonade", "water") )
> x
[1] lemonade lemonade juice    lemonade water
Levels: juice lemonade water
> |
```

Factors

class function :

All objects in R have a class and function **class** reports it.

For simple vectors, this is just the mode, e.g. **"numeric"**, **"logical"**, **"character"**, **"list"**, **"matrix"**, **"array"**, **"factor"** and **"data.frame"**.

A special attribute class of the object is used to allow for an object-oriented style of programming in R.

Factors

class function :

```
> class(9)
[1] "numeric"
```

```
> class("9")
[1] "character"
```

```
> class(print)
[1] "function"
```

```
> x = matrix(nrow=2, ncol=2, data=1:4)
> class(x)
[1] "matrix" "array"
```

Factors

class function :

```
R Console

> class(9)
[1] "numeric"
>
> class("9")
[1] "character"
>
> class(print)
[1] "function"
>
> x=matrix(nrow=2, ncol=2, data=1:4)
> class(x)
[1] "matrix" "array"
> x
      [,1] [,2]
[1,]     1     3
[2,]     2     4
> |
```

Factors

`unclass` function

For example if an object has class `"data.frame"`, it will be printed in a certain way, the `plot()` function will display it graphically in a certain way etc.

`unclass()` is used to temporarily remove the effects of class.

Use `help("unclass")` to get more information.

Factors

Examples: `unclass` function

Consider a group of strings and store it as factors

```
> brands = c("A", "A", "B", "B", "B", "B", "C")
```

```
> brands
```

```
[1] "A" "A" "B" "B" "B" "B" "C"
```

```
> brands_fac = factor(brands)
```

```
> brands_fac
```

```
[1] A A B B B B C
```

```
Levels: A B C
```

Factors

Examples: `unclass` function

Apply `unclass` to this group of factors.

`unclass` will convert the factors to their numbers, like

```
> unclass(brands_fac)
```

```
[1] 1 1 2 2 2 2 3
```

```
attr(,"levels")
```

```
[1] "A" "B" "C"
```

Suppose we have a vector of colors as

```
> colours = c("blue", "green", "red")
```

```
> colours
```

```
[1] "blue" "green" "red"
```

Factors

Examples: `unclass` function

With the following

```
> unclass(brands_fac)
```

```
[1] 1 1 2 2 2 2 3
```

```
attr(,"levels")
```

```
[1] "A" "B" "C"
```

numbers you can convert the factors to colors, by doing:

```
> colours [unclass(brands_fac)]
```

```
[1] "blue"  "blue"  "green" "green" "green"
"green" "red"
```

```
Recall brands = c("A", "A", "B", "B", "B", "B", "C")
```

Factors

Examples: `unclass` function

R Console

```
> colours = c("blue", "green", "red" )
> colours
[1] "blue" "green" "red"
>
> brands = c("A", "A", "B", "B", "B", "B", "C")
> brands
[1] "A" "A" "B" "B" "B" "B" "C"
>
> brands_fac = factor(brands)
> brands_fac
[1] A A B B B B C
Levels: A B C
>
> unclass(brands_fac)
[1] 1 1 2 2 2 2 3
attr(,"levels")
[1] "A" "B" "C"
>
> colours [unclass(brands_fac)]
[1] "blue" "blue" "green" "green" "green" "green" "red"
> |
```

Factors

Examples: `unclass` function

The command `unclass` shows, an integer is assigned to every factor level:

```
> x = factor( c( "lemonade", "lemonade",  
"juice", "lemonade", "water") )
```

```
> unclass(x)
```

```
[1] 2 2 1 2 3
```

```
attr(,"levels")
```

```
[1] "juice" "lemonade" "water"
```

Factors

```
R Console
> x = factor( c( "lemonade", "lemonade", "juice", "lemonade", "water") )
> x
[1] lemonade lemonade juice    lemonade water
Levels: juice lemonade water
>
> unclass(x)
[1] 2 2 1 2 3
attr(,"levels")
[1] "juice"    "lemonade" "water"
>
```

Factors

If a different assignment is desired, the parameter **levels** can be used:

```
> x = factor( c("lemonade", "lemonade",  
"juice", "lemonade", "water"),  
levels=c("water", "juice", "lemonade") )
```

```
> x
```

```
[1] lemonade lemonade juice      lemonade water
```

```
Levels: water juice lemonade
```

Factors

```
> unclass(x)
```

```
[1] 3 3 2 3 1
```

```
attr(,"levels")
```

```
[1] "water"      "juice"      "lemonade"
```

```
> levels(x)
```

```
[1] "water"      "juice"      "lemonade"
```

Factors

R Console

```
> x = factor( c("lemonade", "lemonade", "juice", "lemonade", "water"),  
+ levels=c("water", "juice", "lemonade") )  
> x  
[1] lemonade lemonade juice    lemonade water  
Levels: water juice lemonade  
>  
> unclass(x)  
[1] 3 3 2 3 1  
attr(,"levels")  
[1] "water"    "juice"    "lemonade"  
>  
> levels(x)  
[1] "water"    "juice"    "lemonade"  
> |
```

Factors

Example for an ordered factor:

```
> income = ordered(c("high", "high", "low",  
"medium", "medium"), levels=c("low", "medium",  
"high"))
```

```
> income  
[1] high    high    low     medium medium  
Levels: low < medium < high
```

```
> unclass(income)  
[1] 3 3 1 2 2  
attr(,"levels")  
[1] "low"      "medium"   "high"
```

Factors

R Console

```
> income = ordered(c("high", "high", "low", "medium", "medium"),  
+ levels=c("low", "medium", "high"))  
> income  
[1] high  high  low   medium medium  
Levels: low < medium < high  
>  
> unclass(income)  
[1] 3 3 1 2 2  
attr(,"levels")  
[1] "low"      "medium"   "high"  
> |
```