

Factores, variables y valores etiquetados

Factores

Se usan para representar variables categóricas. También permiten mostrar un vector de strings en un orden que no sea alfabético.

En R los factores están compuestos por los valores en los datos y por los `levels`, lo que debe considerarse cuando se crea un factor.

El vector `mes` que se detalla a continuación es el siguiente:

```
mes <- c("Marzo", "Abril", "Febrero", "Marzo", "Marzo", "Enero", "Abril")
```

El tipo de este vector es `character` y cuando se ordenan sus valores se muestran en orden alfabético, lo que no refleja el orden temporal de los meses del año.

```
sort(mes)
```

```
## [1] "Abril" "Abril" "Enero" "Febrero" "Marzo" "Marzo" "Marzo"
```

En este caso será útil convertir el vector `mes` a factor, lo que requiere especificar los `levels`.

```
mes_level <- c("Enero", "Febrero", "Marzo", "Abril", "Mayo", "Junio",  
              "Julio", "Agosto", "Septiembre", "Octubre", "Noviembre", "Diciembre")
```

Para crear un factor, escriba:

```
mes_factor <- factor(mes, levels = mes_level)
```

Ahora los meses del año aparecen ordenados de acuerdo a cómo se especificó `levels`.

```
sort(mes_factor)
```

```
## [1] Enero  Febrero Marzo  Marzo  Marzo  Abril  Abril  
## 12 Levels: Enero Febrero Marzo Abril Mayo Junio Julio Agosto ... Diciembre
```

Uno de los paquetes que será de gran ayuda al trabajar con factores es `forcats`. Para mayor información se recomienda revisar:

<https://cran.r-project.org/web/packages/forcats/forcats.pdf>

Todas las funciones incluidas en este paquete comienzan con `fct_`. Entre las funciones más utilizadas de este paquete se encuentran:

Función	Descripción
<code>fct_reorder()</code>	reordenar un factor por otra variable
<code>fct_recode()</code>	cambiar los valores de los niveles
<code>fct_infreq()</code>	reordenar un factor por la frecuencia de valores
<code>fct_relevel()</code>	Cambiar manualmente el orden de un factor
<code>fct_lump()</code>	colapsar los valores menos/más frecuentes de un factor en la categoría "other"
<code>fct_rev()</code>	<code>dfgdh</code>

Variables y valores etiquetados

```
library(dplyr)
library(haven)
library(janitor)
```

Muchos datos importados a R incluyen variables etiquetadas y también sus valores con una etiqueta. A continuación se presenta un ejemplo en que la variable `sexo` está etiquetada como “Sexo de la persona” y los valores que toma están etiquetados como: 1 = Mujer, 2 = Hombre

```
sexo <- labelled(
  c(2, 1, 1, 2, 2),
  labels = c(Mujer = 1, Hombre = 2),
  label = "Sexo de la persona"
)
edad <- c(21, 32, 28, 45, 24)

tb <- tibble::tibble(sexo, edad)
tb
```

```
## # A tibble: 5 x 2
##       sexo  edad
##   <dbl+lbl> <dbl>
## 1 2 [Hombre]    21
## 2 1 [Mujer]    32
## 3 1 [Mujer]    28
## 4 2 [Hombre]    45
## 5 2 [Hombre]    24
```

```
class(tb$sexo)
```

```
## [1] "haven_labelled" "vctrs_vctr"      "double"
```

```
typeof(tb$sexo)
```

```
## [1] "double"
```

Note que la variable `sexo` tiene un formato `dbl+lbl`, esto es, es numérica pero está etiquetada, lo que también se puede ver a continuación:

```
str(tb)
```

```
## tibble [5 x 2] (S3: tbl_df/tbl/data.frame)
## $ sexo: dbl+lbl [1:5] 2, 1, 1, 2, 2
## ..@ labels: Named num [1:2] 1 2
## .. ..- attr(*, "names")= chr [1:2] "Mujer" "Hombre"
## ..@ label : chr "Sexo de la persona"
## $ edad: num [1:5] 21 32 28 45 24
```

Al construir una tabla de frecuencias de la variable `sexo` no se muestran las etiquetas de los valores, sino sólo los valores numéricos:

```
tb %>% tabyl(sexo)
```

```
##  sexo n percent
##    1 2    0.4
##    2 3    0.6
```

A continuación la variable `sexo` se convertirá en factor usando `as_factor(sexo)` que usará como `levels` a los `labels` especificados anteriormente.

```
tb <- tb %>% mutate(sexo = as_factor(sexo))
tb
```

```
## # A tibble: 5 x 2
##   sexo    edad
##   <fct>  <dbl>
## 1 Hombre    21
## 2 Mujer     32
## 3 Mujer     28
## 4 Hombre    45
## 5 Hombre    24
```

```
str(tb)
```

```
## tibble [5 x 2] (S3: tbl_df/tbl/data.frame)
##  $ sexo: Factor w/ 2 levels "Mujer","Hombre": 2 1 1 2 2
##    ..- attr(*, "label")= chr "Sexo de la persona"
##  $ edad: num [1:5] 21 32 28 45 24
```

Ahora la tabla de frecuencias de la variable `sexo` muestra como valores las etiquetas de los valores que ahora son los `levels` del factor:

```
tb %>% tabyl(sexo)
```

```
##   sexo n percent
##   Mujer 2    0.4
##   Hombre 3    0.6
```