

forcats

Entornos de Análisis de Datos: R

Alberto Torres Barrán

2020-02-06

forcats

Factores

- Representan variables categóricas, con un número posible de valores:
 - fijo
 - *pequeño*
 - conocido
- Ejemplos: meses del año, provincias, sexo, etc.
- Muchas funciones de R convierten los vectores de cadenas de caracteres a factores automáticamente
- La librería `forcats` implementa funciones para manipular factores

Creando factores

- Función `factor()`
- Necesita dos argumentos:
 - Datos a convertir (típicamente vector de cadenas)
 - Valores posibles (**niveles**)

```
x <- c("Primavera", "Verano", "Verano", "Verano", "Otoño")
f <- factor(x, levels = c("Primavera", "Verano", "Otoño", "Invierno"))
f
## [1] Primavera Verano    Verano    Verano    Otoño
## Levels: Primavera Verano Otoño Invierno
```

- Si un valor no está en los niveles se convierte a NA

```
x <- c("Primavera", "Verano", "Verano", "Verano", "Otono")
f1 <- factor(x, levels = c("Primavera", "Verano", "Otoño", "Invierno"))
f1
## [1] Primavera Verano Verano Verano <NA>
## Levels: Primavera Verano Otoño Invierno
```

- Si no se indican los niveles, se toman como los valores únicos del vector

```
unique(x)
## [1] "Primavera" "Verano" "Otono"
```

```
factor(x)
## [1] Primavera Verano Verano Verano Otono
## Levels: Otono Primavera Verano
```

También se pueden crear a partir de una variable numérica

```
edad <- c(20, 25, 15, 38)
cut(edad, breaks = c(10, 20, 30, 40))
## [1] (10,20] (20,30] (10,20] (30,40]
## Levels: (10,20] (20,30] (30,40]
```

Algunas funciones útiles

- `fct_reorder()` para reordenar los niveles (se mueven al principio)

```
fct_relevel(f, "Invierno", "Primavera")  
## [1] Primavera Verano Verano Verano Otoño  
## Levels: Invierno Primavera Verano Otoño
```

- `fct_recode()` para cambiar el nombre de los niveles o eliminarlos (asignando el valor `NULL`)

```
fct_recode(f, Inv = "Invierno", Prim = "Primavera", Ver = "Verano", Oto = "Otoño")  
## [1] Prim Ver Ver Ver Oto  
## Levels: Prim Ver Oto Inv
```

- `fct_recode()` también puede colapsar niveles asignando el mismo nombre

```
fct_recode(f, Frio = "Invierno", Calor = "Primavera", Calor = "Verano", Frio = "Ot  
## [1] Calor Calor Calor Calor Frio  
## Levels: Calor Frio
```

- Otra alternativa es `fct_collapse()`

```
fct_collapse(f, Frio = c("Invierno", "Otoño"), Calor = c("Primavera", "Verano"))  
## [1] Calor Calor Calor Calor Frio  
## Levels: Calor Frio
```