

Clase 1.1

Estructuras de datos

Marcos Rosetti y Luis Pacheco-Cobos

Estadística y Manejo de Datos con R (EMDR) — Virtual

Factores

Factores

- Son vectores cuyos valores están organizados en **categorías**.
- Las categorías se llaman *levels* y son valores de texto.
- Esta nueva capa de información es útil para calcular estadísticos descriptivos.

Factores

- Construcción de un factor

```
mons <- c("March", "April", "January", "November", "January", "September",  
         "October", "September", "November", "August", "January",  
         "November", "November", "February", "May", "August", "July",  
         "December", "August")
```

```
mons
```

```
## [1] "March"      "April"      "January"    "November"   "January"    "September"  
## [7] "October"    "September"  "November"   "August"     "January"    "November"  
## [13] "November"   "February"   "May"        "August"     "July"       "December"  
## [19] "August"
```

Factores

- Construcción de un factor

```
fmons <- factor(mons)
fmons
```

```
## [1] March      April      January   November  January   September October
## [8] September November  August    January   November  November  February
## [15] May         August     July      December  August
## 11 Levels: April August December February January July March May ... September
```

Factores

- Ahora removamos `April`

```
fmons <- fmons[which(fmons != "April")]  
fmons
```

```
## [1] March      January  November January  September October  September  
## [8] November  August   January  November November February May  
## [15] August    July     December August  
## 11 Levels: April August December February January July March May ... September
```

Factores

- ¿Qué pasa con April?

```
table(fmons)
```

```
## fmons
##      April      August  December  February  January      July      March      May
##         0         3         1         1         3         1         1         1
## November  October  September
##         4         1         2
```

```
levels(fmons)
```

```
## [1] "April"      "August"     "December"   "February"   "January"    "July"
## [7] "March"      "May"        "November"   "October"    "September"
```

Factores

- ¿Y ahora?

```
fmons <- droplevels(fmons)
fmons
```

```
## [1] March      January  November January  September October  September
## [8] November  August   January  November November February May
## [15] August     July     December August
## 10 Levels: August December February January July March May November ... September
```


Factores

- ¿Y ahora?

```
levels(fmons)
```

```
## [1] "August" "December" "February" "January" "July" "March"  
## [7] "May" "November" "October" "September"
```

```
table(fmons)
```

```
## fmons  
## August December February January July March May November  
## 3 1 1 3 1 1 1 4  
## October September  
## 1 2
```

Factores

```
weights <- rnorm(n = 100, mean = 50, sd = 10)
fw <- cut(weights, breaks = 3)
table(fw)
```

```
## fw
## (28.5,47.2] (47.2,65.9] (65.9,84.7]
##          43          50          7
```

```
fw <- cut(round(weights), breaks = quantile(weights, probs=seq(0, 1, 0.25)),
          labels = c('1stQ', '2ndQ', '3rdQ', '4thQ'))
```

```
table(fw) # advertencia : produce valores NA para los outliers
```

```
## fw
## 1stQ 2ndQ 3rdQ 4thQ
##   26   24   25   24
```

Licencia CC BY



Estadística y Manejo de Datos con R (EMDR) por Marcos F. Rosetti S. y Luis Pacheco-Cobos se distribuye bajo una [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).