

# Factors withforcats :: CHEATSHEET



The **forcats** package provides tools for working with factors, which are R's data structure for categorical data.

## Factors

R represents categorical data with factors. A **factor** is an integer vector with a **levels** attribute that stores a set of mappings between integers and categorical values. When you view a factor, R displays not the integers, but the levels associated with them.

stored	displayed
1 1=a 3 2=b 2 3=c 1	a 1=a b 2=b c 3=c a

*Create a factor with factor()*  
**factor**(x = character(), levels, labels = levels, exclude = NA, ordered = is.ordered(x), nmax = NA) Convert a vector to a factor. Also **as\_factor()**.  
**f <- factor(c("a", "c", "b", "a"), levels = c("a", "b", "c"))**

*Return its levels with levels()*  
**levels(x)** Return/set the levels of a factor. **levels(f); levels(f) <- c("x", "y", "z")**

*Use unclass() to see its structure*

## Inspect Factors

1=a 2=b 3=c a	f n a 2 b 1 c 1
------------------------	--------------------------

**fct\_count(f, sort = FALSE, prop = FALSE)** Count the number of values with each level. **fct\_count(f)**

**fct\_match(f, lvls)** Check for lvls in f. **fct\_match(f, "a")**

**fct\_unique(f)** Return the unique values, removing duplicates. **fct\_unique(f)**

## Combine Factors

a 1=a c 2=c a	b 1=a a 2=b a	a 1=a c 2=c b a
---------------------	---------------------	--------------------------

**fct\_c(...)** Combine factors with different levels. Also **fct\_cross()**.  
**f1 <- factor(c("a", "c"))**  
**f2 <- factor(c("b", "a"))**  
**fct\_c(f1, f2)**

**fct\_unify(fs, levels = lvls\_union(fs))** Standardize levels across a list of factors. **fct\_unify(list(f2, f1))**

## Change the order of levels

**a c b a** → **a c b a** **fct\_relevel(f, ..., after = 0L)** Manually reorder factor levels. **fct\_relevel(f, c("b", "c", "a"))**

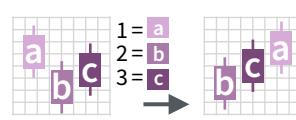
**c c a** → **c c a** **fct\_infreq(f, ordered = NA)** Reorder levels by the frequency in which they appear in the data (highest frequency first). Also **fct\_inseq()**.  
**f3 <- factor(c("c", "c", "a"))**  
**fct\_infreq(f3)**

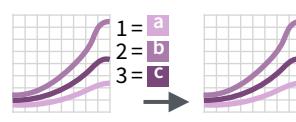
**b a** → **b a** **fct\_inorder(f, ordered = NA)** Reorder levels by order in which they appear in the data. **fct\_inorder(f2)**

**a b c** → **a b c** **fct\_rev(f)** Reverse level order.  
**f4 <- factor(c("a", "b", "c"))**  
**fct\_rev(f4)**

**a b c** → **a b c** **fct\_shift(f)** Shift levels to left or right, wrapping around end. **fct\_shift(f4)**

**a b c** → **a b c** **fct\_shuffle(f, n = 1L)** Randomly permute order of factor levels. **fct\_shuffle(f4)**

 **fct\_reorder(f, .x, .fun = median, ..., .desc = FALSE)** Reorder levels by their relationship with another variable. **boxplot(PlantGrowth, weight ~ fct\_reorder(group, weight))**

 **fct\_reorder2(f, .x, .y, .fun = last2, ..., .desc = TRUE)** Reorder levels by their final values when plotted with two other variables. **ggplot(diamonds, aes(carat, price, color = fct\_reorder2(color, carat, price))) + geom\_smooth()**

## Change the value of levels

**a c b a** → **v z x v** **fct\_recode(f, ...)** Manually change levels. Also **fct\_relabel()** which obeys purrr::map syntax to apply a function or expression to each level.  
**fct\_recode(f, v = "a", x = "b", z = "c")**  
**fct\_relabel(f, ~ paste0("x", .x))**

**a c b a** → **2 1=2  
1 2=1  
3 3=3  
2** **fct\_anon(f, prefix = "")** Anonymize levels with random integers. **fct\_anon(f)**

**a c b a** → **x c x x** **fctCollapse(f, ..., other\_level = NULL)** Collapse levels into manually defined groups. **fct\_collapse(f, x = c("a", "b"))**

**a c b a** → **a 1=a  
2=Other  
3=Other** **fct\_lump\_min(f, min, w = NULL, other\_level = "Other")** Lumps together factors that appear fewer than min times. Also **fct\_lump\_n()**, **fct\_lump\_prop()**, and **fct\_lump\_lowfreq()**. **fct\_lump\_min(f, min = 2)**

**a c b a** → **a 1=a  
2=b  
3=Other** **fct\_other(f, keep, drop, other\_level = "Other")** Replace levels with "other." **fct\_other(f, keep = c("a", "b"))**

## Add or drop levels

**a b** → **a b** **fct\_drop(f, only)** Drop unused levels.  
**f5 <- factor(c("a", "b"), c("a", "b", "x"))**  
**f6 <- fct\_drop(f5)**

**a b** → **a b** **fct\_expand(f, ...)** Add levels to a factor. **fct\_expand(f6, "x")**

**a b** → **a b x** **fct\_na\_value\_to\_level(f, level = "(Missing)")** Assigns a level to NAs to ensure they appear in plots, etc. **f7 <- factor(c("a", "b", NA))**  
**fct\_na\_value\_to\_level(f7, level = "(Missing)")**