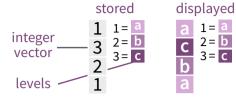
# Factors with forcats:: cheatsheet

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

#### **Factors**

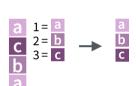
R represents categorical data with factors. A factor is an integer vector with a **levels** attribute that stores a set of mappings between



integers and categorical values. When you view a factor, R displays not the integers, but the levels associated with them.

#### 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"))



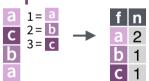
Return its levels with levels()

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**

c 2= b 3= c



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")



**Combine Factors** 

a 1= a + b 1= a = a 1= a C 2= C

2= C 3= b

2=b 3=c

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

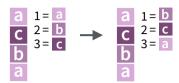
fct\_c(...) Combine factors

with different levels.

Also fct\_cross().

fct\_unify(list(f2, f1))

# Change the order of levels



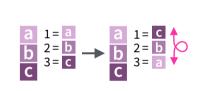
fct\_relevel(.f, ..., after = 0L) Manually reorder factor levels. fct relevel(f, c("b", "c", "a"))



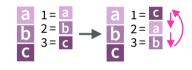
fct\_infreq(f, ordered = NA) Reorder levels by the frequency in which they appear in the data (highest frequency first). Also fct\_inseq(). f3 <- factor(c("c", "c", "a")) fct infrea(f3)



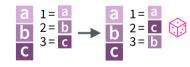
fct\_inorder(f, ordered = NA) Reorder levels by order in which they appear in the data. fct\_inorder(f2)



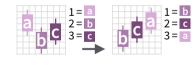
fct\_rev(f) Reverse level order. f4 <- factor(c("a","b","c"))



fct\_shift(f) Shift levels to left or right, wrapping around end.

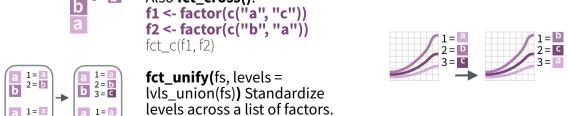


fct\_shuffle(f, n = 1L) Randomly permute order of factor levels. fct shuffle(f4)



fct\_reorder(.f, .x, .fun = median, ..., .desc = FALSE) Reorder levels by their relationship with another variable.

PlantGrowth. weight ~ fct\_reorder(group, weight)

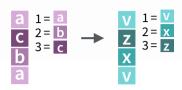


fct\_reorder2(.f, .x, .y, .fun = last2, ..., .desc = TRUE) Reorder levels by their final values when plotted with two other variables.

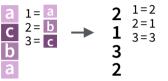
color = fct\_reorder2(color, carat, price)

geom smooth()

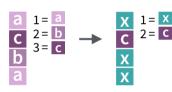
## Change the value of levels



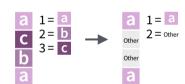
fct\_recode(.f, ...) Manually change levels. Also fct relabel() which obeys purrr::map syntax to apply a function or expression to each level. fct\_recode(f, v = "a", x = "b", z = "c") fct\_relabel(f, ~ paste0("x", .x))



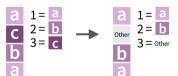
fct anon(f, prefix = "") Anonymize levels with random integers. fct anon(f)



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

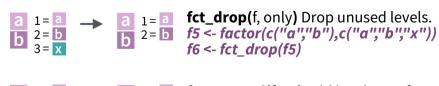


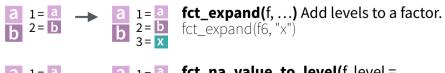
fct\_lump\_min(f, min, w = NULL, other\_level = "Other") Lumps together factors that appear fewer than min times. Also fct\_lump\_n(), fct\_lump\_prop(), and fct\_lump\_lowfreq(). fct lump min(f, min = 2)



fct\_other(f, keep, drop, other level = "Other") Replace levels with "other."  $fct_other(f, keep = c("a", "b"))$ 

### Add or drop levels





a 1= fct\_na\_value\_to\_level(f, level = b 2 = b "(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)")

