

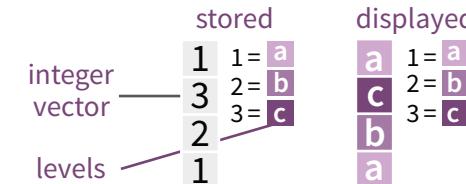
Factors withforcats :: CHEAT SHEET

The **forcats** パッケージは、Rのカテゴリカルデータ構造「ファクター」の操作を行うツールを提供します

ファクター

Rでは、カテゴリカルデータを「ファクター」として表現します。ファクターはINT整数ベクトルと、「レベル」文字列がマッピングされたデータ形式です。

文字列でなくファクターとして保存されたデータは、読みやすさを保ったまま、処理の際にはINT整数として演算されるために高速な処理が可能になります。



factor(x = character(), levels, labels = levels, exclude = NA, ordered = is.ordered(x), nmax = NA)
f <- factor(c("a", "c", "b", "a"), levels = c("a", "b", "c"))

levels(x) Return/set the levels of a factor. **levels(f); levels(f) <- c("x", "y", "z")**
unclass() 関数で構造を見てみましょう

※**as_factor()** 関数もよく使われます

ファクターを調べる

fct_count(f, sort = FALSE)
 レベルの数をカウントして返します **fct_count(f)**

fct_unique(f)
 ユニークなレベル文字列を返します **fct_unique(f)**

ファクターを合成する

fct_c(...) 異なるレベルを持つ複数のファクターを合成します
f1 <- factor(c("a", "c"))
f2 <- factor(c("b", "a"))
fct_c(f1, f2)

fct_unify(fs, levels = lvl_union(fs)) 複数のファクター間のレベルを統一します
fct_unify(list(f2, f1))

ファクターを並び替える

fct_relevel(.f, ..., after = 0L)
 手動でレベルを並び替えます
fct_relevel(f, c("b", "c", "a"))

fct_infreq(f, ordered = NA)
 「データセット内の出現頻度」順で、レベルを並び替えます(降順)
f3 <- factor(c("c", "c", "a"))
fct_infreq(f3)

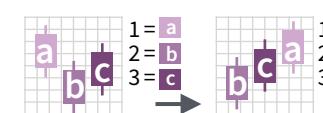
fct_inorder(f, ordered = NA)
 「データセット内で最初に出現した」順で、レベルを並び替えます
fct_inorder(f2)

fct_rev(f) 並び替え：逆順
f4 <- factor(c("a", "b", "c"))
fct_rev(f4)

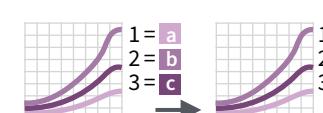
fct_shift(f) 並び替え：位置シフト
 前後にずらします(始点・終点はループします)
fct_shift(f4)

fct_shuffle(f, n = 1L) シャッフル
 ランダムに位置を入れ替えます。
fct_shuffle(f4)

fct_reorder(.f, .x, .fun = median, ..., .desc = FALSE) 他のひとつの変数との関連性に基づき、レベルを並び替えます。



boxplot(data = iris, Sepal.Width ~ fct_reorder(Species, Sepal.Width))



fct_reorder2(.f, .x, .y, .fun = last2, ..., .desc = TRUE) 他のふたつの変数との関連性に基づき、レベルを並び替えます。
ggplot(data = iris, aes(Sepal.Width, Sepal.Length, color = fct_reorder2(Species, Sepal.Width, Sepal.Length))) + geom_smooth()

レベルの値を書き換える

fct_recode(.f, ...) 手動での書き換え
fct_relabel 関数/構文での書き換え
purrr::map 文法が使えます
fct_recode(f, v = "a", x = "b", z = "c")
fct_relabel(f, ~ paste0("x", .x))

fct_anon(f, prefix = "")
 ランダムな整数に置換え、元のレベルを匿名化します
fct_anon(f)

fct_collapse(f, ...)
 指定したレベルグループのみを残して不要なレベルをXに書き換えます
fct_collapse(f, x = c("a", "b"))

fct_lump(f, n, prop, w = NULL, other_level = "Other", ties.method = c("min", "average", "first", "last", "random", "max"))
 出現頻度の最も多い・少ないレベルをひとつにまとめます
fct_lump(f, n = 1)
fct_lump_min 関数も使用できます

fct_other(f, keep, drop, other_level = "Other") Replace levels with "other."
 指定したレベル以外を「other」としてひとつにまとめます
fct_other(f, keep = c("a", "b"))

レベルをドロップする

fct_drop(f, only) 未使用的レベルを削除
f5 <- factor(c("a", "b"), c("a", "b", "x"))
f6 <- fct_drop(f5)

fct_expand(f, ...) レベルを追加
fct_expand(f6, "x")

fct_explicit_na(f, na_level = "(Missing)")
 NA値にレベルを割り当て、プロットなどでそれらの値が消滅してしまう事を防ぎます。
fct_explicit_na(factor(c("a", "b", NA)))