

Défaire un groupe répété aplati

`pivot_longer()` avec `.value`, l'appel unique qui retransforme un export CommCare ou Kobo en une ligne par personne — plus l'avertissement de `pivot_wider()` qui signifie que votre clé n'en est pas une.

Pierre Robentz Cassion

Leçon 7 sur 8

R et le tidyverse pour les données de programme — Restructurer et répéter

Ce que couvre cette leçon

- Ce qu'une plateforme de collecte fait d'un groupe répété
- `pivot_longer()` avec `.value`
- Vérifiez le nombre de lignes obtenu
- `pivot_wider()` et l'avertissement qui veut dire quelque chose
- Séparer et réunir des colonnes
- L'imbrication, brièvement
- Où cela rejoint le cours Python
- Ce qui vient ensuite

Ce qu'une plateforme de collecte fait d'un groupe répété — Exemple

hh	member_1_age	member_1_sex	member_2_age	member_2_sex
H1	34	f	8	m

`pivot_longer()` avec `.value` — En R

```
library(tidyr)

long <- wide |>
  pivot_longer(
    starts_with("member_"),
    names_to = c("member", ".value"),
    names_pattern = "member_(\\d+)_(.*)"
  )
```

`pivot_longer()` avec `.value` — Exemple

```
# A tibble: 2 × 4
  hh    member    age sex
<chr> <chr> <dbl> <chr>
1 H1      1      34 f
2 H1      2       8 m
```

```
pivot_longer() avec .value
```

- Le motif

`pivot_longer()` avec `.value` — En R

Separateur fixe, sans expression reguliere.

```
pivot_longer(wide, starts_with("member_"),  
             names_to = c("member", ".value"), names_sep = "_(?=[a-z]+$)")
```

Une seule colonne de valeurs, nom conserve entier.

```
pivot_longer(wide, starts_with("crop_"),  
             names_to = "slot", names_prefix = "crop_", values_to = "crop",  
             values_drop_na = TRUE)
```

Vérifiez le nombre de lignes obtenu — En R

```
stopifnot(nrow(long) == sum(!is.na(wide$member_1_age)) + sum(!is.na(wide$member_2_age)))
```

Vérifiez le nombre de lignes obtenu — En R

```
long |>  
  count(hh, name = "members_found") |>  
  left_join(households |> select(hh, hh_size), by = "hh") |>  
  filter(members_found != hh_size)
```

`pivot_wider()` et l'avertissement qui veut dire quelque chose — En R

```
by_commune_month |>  
  pivot_wider(names_from = month, values_from = cases, values_fill = 0)
```

`pivot_wider()` et l'avertissement qui veut dire quelque chose — Exemple

Warning: Values from `v` are not uniquely identified; output will contain list-cols.

`pivot_wider()` et l'avertissement qui veut dire quelque chose

- Cet avertissement signifie que votre clé n'en est pas une — La cellule contient désormais une liste des valeurs entrées. . .

`pivot_wider()` et l'avertissement qui veut dire quelque chose — En R

```
by_commune_month |> count(commune, month) |> filter(n > 1)
```

`pivot_wider()` et l'avertissement qui veut dire quelque chose

- `values_fill`

`pivot_wider()` et l'avertissement qui veut dire quelque chose — En R

```
pivot_wider(..., values_fill = 0)
```

Séparer et réunir des colonnes — En R

```
tibble(x = c("2024-01", "2024-02")) |>  
  separate_wider_delim(x, delim = "-", names = c("year", "month"))
```

Séparer et réunir des colonnes — Exemple

```
# A tibble: 2 × 2
  year month
<chr> <chr>
1 2024 01
2 2024 02
```

Séparer et réunir des colonnes — En R

```
separate_wider_delim(x, "-", names = c("year", "month"), too_few = "align_start")
```

Séparer et réunir des colonnes — En R

```
unite(df, "key", district, community, sep = "-", remove = FALSE)
```

L'imbrication, brièvement — En R

```
by_commune <- muac |>  
  tidyr::nest(.by = commune)
```

```
by_commune$data[[1]]      # le tibble complet de la premiere commune
```

Où cela rejoint le cours Python

- Le cours Python couvre la même restructuration sous `melt` et `pivot`.

Ce qui vient ensuite

- Les données ont la bonne forme.

Lire la leçon complète, avec le code exécutable

<https://data-analysis.cassion.dev/fr/cours/r-for-programme-data/r-07-pivots>