

La zone dont les dossiers ont doublé est celle qui s'est améliorée

Le flux mensuel de Saint-Louis-du-Nord passe de 5,8 dossiers à 10,2 après juillet. Rien de la violence dans cette zone n'a changé. Un service a ouvert, et une courbe de dossiers mesure si les gens peuvent atteindre un bureau — jamais combien en avaient besoin.

Pierre Robentz Cassion

Leçon 7 sur 8

Données de protection et VBG — Ce que les nombres signifient

Ce que couvre cette leçon

- La courbe, et la lecture évidente
- Ce qu'un compte de dossiers compte
- Le test qui les distingue
- Dites-le dans le rapport, dans la phrase à côté du nombre
- La conséquence que l'on manque
- D'où vient réellement la prévalence
- La suite

La courbe, et la lecture évidente — En Python

```
import pandas as pd

cases = pd.read_csv("protection-case-management-2024.v1.csv")
monthly = cases.groupby(["admin2", "opened_month"]).size().unstack(fill_value=0)
print(monthly)
```

La courbe, et la lecture évidente — En R

```
library(dplyr)

cases |> count(admin2, opened_month) |>
  tidyr::pivot_wider(names_from = opened_month, values_from = n, values_fill = 0)
```

La courbe, et la lecture évidente — En Python

```
area = cases[cases["admin2"] == "Saint-Louis-du-Nord"]
month = area["opened_month"].str[5:].astype(int)
before = (month <= 6).sum() / 6
after = (month >= 7).sum() / 6
print(f"Jan-Jun: {before:.1f} cases a month")
print(f"Jul-Dec: {after:.1f} cases a month  ({after / before - 1:+.0%})")
```

```
cases |>
  filter(admin2 == "Saint-Louis-du-Nord") |>
  mutate(half = if_else(as.integer(substr(opened_month, 6, 7)) <= 6, "H1", "H2")) |>
  count(half) |> mutate(per_month = n / 6)
```

- 5,8 dossiers par mois avant juillet, 10,2 après — une hausse de 74 % — tandis qu'aucune autre zone ne bouge de plus de. . .

Ce qu'un compte de dossiers compte

Le compte dépend de	Ce qui change quand
L'existence d'un service à portée	L'un ouvre, l'un ferme, une route est coupée
Le fait que les gens le sachent	Une campagne de sensibilisation a lieu
La disposition à se confier	La confiance monte ou baisse ; une mauvaise expérience circule
Le fait que quelqu'un ouvre un dossier	Effectifs, charge de travail, pratique de référencement
L'incidence dans la population	—

Ce qu'un compte de dossiers compte

- Les quatre premières ont bougé à Saint-Louis-du-Nord en juillet, parce qu'un service a ouvert — La cinquième n'est pas. . .

Ce qu'un compte de dossiers compte — En Python

```
print("Case count = incidence x reporting rate x service availability x recording")  
print("A change in the product does not tell you which factor moved.")
```

Ce qu'un compte de dossiers compte — En R

One equation, five unknowns, one observation.

- Quelque chose a-t-il changé dans le service ? — L'indice le plus fort et le moins statistique
- La composition a-t-elle changé ? — La part de protection générale à Saint-Louis-du-Nord passe de 31,4 % à 41,0 % tandis. . .

Le test qui les distingue — En Python

```
def mix(frame):  
    return frame["case_category"].value_counts(normalize=True).round(3)  
  
print("before July:", mix(area[month <= 6]).to_dict())  
print("after July: ", mix(area[month >= 7]).to_dict())
```

Le test qui les distingue — En R

```
cases |> filter(admin2 == "Saint-Louis-du-Nord") |>  
  mutate(half = if_else(as.integer(substr(opened_month, 6, 7)) <= 6, "H1", "H2")) |>  
  count(half, case_category) |> mutate(share = n / sum(n), .by = half)
```

- Les autres zones ont-elles bougé? — Une hausse confinée à une zone avec une explication administrative est un effet de...

Le test qui les distingue — En Python

```
halves = cases.assign(half=(cases["opened_month"].str[5:].astype(int) > 6))
shift = halves.groupby(["admin2", "half"]).size().unstack()
shift["change"] = shift[True] / shift[False] - 1
print((shift["change"] * 100).round(1).sort_values())
```

Le test qui les distingue — En R

```
cases |> mutate(half = as.integer(substr(opened_month, 6, 7)) > 6) |>  
  count(admin2, half) |> tidyr::pivot_wider(names_from = half, values_from = n)
```

- Une zone à $+74,3\%$; les cinq autres entre $-9,4\%$ et $+11,3\%$ — Une seule zone bougeant sept fois plus que la plus large. . .

Dites-le dans le rapport, dans la phrase à côté du nombre — Exemple

Cases opened, Saint-Louis-du-Nord

Jan-Jun	35 cases	5.8 per month	
Jul-Dec	61 cases	10.2 per month	+74%

A case management service opened in the area in July. This figure measures the number of people who reached a service and disclosed; it is not a measure of the number of incidents. The rise is consistent with improved access and cannot be read as a rise in violence.

The other five areas moved between -9% and +11% over the same period.

Dites-le dans le rapport, dans la phrase à côté du nombre

- L'avertissement va à côté du nombre, non en note de bas de page — Une courbe de dossiers de protection sans cette. . .

La conséquence que l'on manque — En Python

```
outcomes = pd.DataFrame({
    "observation": ["cases rise", "cases fall"],
    "good reading": ["outreach and access improved",
                    "incidents fell"],
    "bad reading": ["violence increased",
                   "service became unreachable, or trust was lost"],
})
print(outcomes)
```

La conséquence que l'on manque — En R

The same number supports opposite conclusions. The context decides.

La conséquence que l'on manque

- Un programme de protection ne peut donc pas employer le volume de dossiers comme indicateur de performance dans un sens. . .

D'où vient réellement la prévalence

- Des enquêtes de prévalence en population — conçues pour cela, avec une formation spécialisée des enquêteurs et des. . .
- Jamais de données de service — à quelque niveau de sophistication que ce soit

D'où vient réellement la prévalence — En Python

```
print("Question: how many women in this district experienced violence this year?")  
print("Answerable from this dataset: no")  
print("Instrument that answers it: a population-based prevalence survey")
```

D'où vient réellement la prévalence — En R

Saying "we cannot answer that with this" is the answer.

- La dernière leçon assemble le tout en un rapport de protection — y compris la section qui énonce quelles analyses ont été demandées et refusées, et pourquoi cette section a sa place dans le document plutôt que dans un courriel que personne ne conserve.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/protection-and-gbv-data/
prot-07-reporting-not-prevalence](https://data-analysis.cassion.dev/fr/cours/protection-and-gbv-data/prot-07-reporting-not-prevalence)