The area whose cases doubled is the one that improved

Saint-Louis-du-Nord's monthly intake goes from 5.8 cases to 10.2 after July. Nothing about violence in that area changed. A service opened, and a case curve measures whether people can reach a desk — never how many of them needed to.

Pierre Robentz Cassion

Lesson 7 of 8

Protection and GBV Data — What the numbers mean

What this lesson covers

- The curve, and the obvious reading
- What a case count is a count of
- The test that distinguishes them
- Say it in the report, in the sentence next to the number
- The consequence people miss
- Where prevalence actually comes from
- What comes next

The curve, and the obvious reading — In 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)
```

The curve, and the obvious reading — In R

```
library(dplyr)

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

The curve, and the obvious reading — In 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%})")
```

The curve, and the obvious reading — In R

```
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)
```

The curve, and the obvious reading

- 5.8 cases a month before July, 10.2 after — a 74% rise — while no other area moves by more than eleven points either way

What a case count is a count of

The count depends on	Which changes when
A service existing within reach	One opens, one closes, a road is cut
People knowing it exists	An outreach campaign runs
People being willing to disclose	Trust rises or falls; a bad experience circulates
Someone opening a file	Staffing, workload, referral practice
Incidence in the population	—

What a case count is a count of

- Every one of the first four moved in Saint-Louis-du-Nord in July, because a service opened — The fifth is not measured...

What a case count is a count of — In 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.")
```

What a case count is a count of — In R

```
# One equation, five unknowns, one observation.
```

The test that distinguishes them

- Did anything change in the service? — The strongest evidence and the least statistical
- Did the composition change? — Saint-Louis-du-Nord's general protection share goes from 31.4% to 41.0% while GBV falls...

The test that distinguishes them — In 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())
```

The test that distinguishes them — In 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)
```

The test that distinguishes them

- Did other areas move? — A rise confined to one area with an administrative explanation is a reporting effect

The test that distinguishes them — In 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())
```

The test that distinguishes them — In R

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

The test that distinguishes them

- One area up 74.3%; the other five range from −9.4% to +11.3% — A single area moving seven times as far as the widest of...

Say it in the report, in the sentence next to the number — Example

```
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.
```

Say it in the report, in the sentence next to the number

- The disclaimer belongs beside the number, not in a footnote — A protection case curve without that sentence will be...

The consequence people miss — In 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)
```

The consequence people miss — In R

```
# The same number supports opposite conclusions. The context decides.
```

The consequence people miss

- So a protection programme cannot use case volume as a performance indicator in either direction — and the indicators...

Where prevalence actually comes from

- Population-based prevalence surveys — designed for the purpose, with specialised interviewer training and ethical...
- Never from service data — at any level of sophistication

Where prevalence actually comes from — In 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")
```

Where prevalence actually comes from — In R

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

What comes next

- The last lesson assembles everything into a protection report — including the section that states which analyses were requested and declined, and why that section belongs in the document rather than in an email nobody keeps.

Read the full lesson, with runnable code

https://data-analysis.cassion.dev/courses/protection-and-gbv-data/prot-07-reporting-not-prevalence