

La courbe n'est pas l'épidémie

975 cas, 47 sans date de début, et un district dont le registre est rempli après coup depuis le cahier d'admission — si bien que son délai médian d'admission affiche zéro alors que sa létalité est la plus élevée des trois.

Pierre Robentz Cassion

Leçon 3 sur 8

Santé publique et épidémiologie appliquée — Une épidémie, un cas à la fois

Ce que couvre cette leçon

- À quoi sert la courbe
- Tracez-la
- Ce que font les dates manquantes
- Le district dont la courbe n'est pas une courbe
- Lire la forme
- Tracez des cas, non des taux
- La suite

À quoi sert la courbe

- Une courbe épidémique, ce sont les cas par date de début des symptômes.

Tracez-la — En Python

```
import pandas as pd

cases = pd.read_csv("cholera-line-list-2024.v1.csv", parse_dates=["onset_date"])

with_onset = cases[cases["onset_date"].notna()]
print(f"{len(with_onset)} of {len(cases)} cases have an onset date")

curve = (
    with_onset.assign(week=with_onset["onset_date"].dt.to_period("W").dt.start_time)
    .groupby("week").size()
)
print(curve)
print(f"peak week {curve.idxmax().date()} with {curve.max()} cases")
```

```
library(dplyr)

cases <- readr::read_csv("cholera-line-list-2024.v1.csv")

curve <- cases |>
  filter(!is.na(onset_date)) |>
  mutate(week = lubridate::floor_date(onset_date, "week")) |>
  count(week)
```

Ce que font les dates manquantes — En Python

```
missing = cases[cases["onset_date"].isna()]
print(missing.groupby("district").size())
print(f"{len(missing) / len(cases):.1%} of all cases")
```

Ce que font les dates manquantes — En R

```
cases |> filter(is.na(onset_date)) |> count(district)
```

Ce que font les dates manquantes

- Les exclure — Le défaut, et généralement le bon
- Les imputer d'après la date d'admission — Tentant, et cela décale toute la courbe vers la droite du délai médian, ce. . .
- Les tracer en bande séparée — Honnête et rarement fait — une barre empilée avec « début inconnu » dans une teinte. . .

Ce que font les dates manquantes — En Python

```
imputed = cases.copy()
fallback = pd.to_datetime(imputed["admission_date"], errors="coerce")
imputed["onset_or_admission"] = imputed["onset_date"].fillna(fallback)

both = pd.DataFrame({
    "excluded": curve,
    "imputed": (imputed.assign(
        week=imputed["onset_or_admission"].dt.to_period("W").dt.start_time)
        .groupby("week").size()),
}).fillna(0).astype(int)
print(both)
```

Ce que font les dates manquantes — En R

```
cases |>
  mutate(onset_or_admission = coalesce(onset_date, admission_date),
         week = lubridate::floor_date(onset_or_admission, "week")) |>
  count(week)
```

Ce que font les dates manquantes

- Quel que soit votre choix, énoncez le dénominateur sur le graphique — « $n = 928$ des 975 cas avec une date de début. . .

Le district dont la courbe n'est pas une courbe — En Python

```
delay = (  
    pd.to_datetime(cases["admission_date"], errors="coerce") - cases["onset_date"]  
).dt.days  
  
by_district = cases.assign(delay=delay).dropna(subset=["delay"]).groupby("district")  
print(by_district["delay"].median())  
print((by_district["delay"].apply(lambda s: (s == 0).mean()) * 100).round(0))
```

Le district dont la courbe n'est pas une courbe — En R

```
cases |>
  filter(!is.na(onset_date), !is.na(admission_date)) |>
  mutate(delay = as.integer(admission_date - onset_date)) |>
  summarise(median_delay = median(delay), zero_day = mean(delay == 0), .by = district)
```

Le district dont la courbe n'est pas une courbe

District	Délai médian	Même jour
Nord	0 jour	80 %
Centre	2 jours	13 %
Sud	2 jours	20 %

Le district dont la courbe n'est pas une courbe

- Les cas de Nord semblent atteindre le traitement le jour même où ils tombent malades, quatre fois plus souvent que. . .
- La courbe épidémique de Nord est une courbe d'admissions — Elle est décalée vers la droite du temps que ses cas ont. . .
- La statistique de délai de Nord est inutilisable — et c'est celle que la leçon suivante emploie pour expliquer la. . .

Le district dont la courbe n'est pas une courbe — En Python

```
print("Nord: median delay 0 days, case fatality 6.30%")  
print("Centre: median delay 2 days, case fatality 2.00%")
```

Le district dont la courbe n'est pas une courbe — En R

The contradiction is the finding: the fastest-looking district dies most.

Le district dont la courbe n'est pas une courbe

- Une contradiction entre deux indicateurs d'un même registre est une question de données avant d'être une question. . .

- **Un pic unique et net** suggère une source ponctuelle — un point d'eau contaminé, un événement — avec des cas. . .
- **Une série de pics espacés d'environ une période d'incubation** suggère une propagation interhumaine.
- **Un long plateau** suggère une exposition à source commune persistante, ce qui pour le choléra signifie en général que. . .

Tracez des cas, non des taux

- Un dernier point pratique.

- La courbe dit quand.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/public-health-epidemiology/
epi-03-epidemic-curves](https://data-analysis.cassion.dev/fr/cours/public-health-epidemiology/epi-03-epidemic-curves)