

Quinze litres, trente minutes, cinq cents mètres

13,4 % de ces ménages passent sous le minimum Sphère de quantité. Deux des colonnes qui produisent ce nombre portent des erreurs d'unité, et l'une d'elles fait monter un ménage d'un échelon plutôt que sortir d'un graphique.

Pierre Robentz Cassion

Leçon 3 sur 8

Analyse EAH — Ce que Sphère demande et qu'une échelle ne dit pas

Ce que couvre cette leçon

- Ce que Sphère ajoute à l'échelle
- Deux erreurs d'unité, et une seule est visible
- L'attente est incluse dans l'aller-retour, et c'est un problème
- La distance n'est pas du tout dans le fichier
- Rapportez la distribution face au seuil
- La suite

Ce que Sphère ajoute à l'échelle

Standard	Seuil	Dans ce fichier
Quantité	Au moins 15 litres par personne et par jour	litres_per_person_day
Temps d'attente	30 minutes au plus	Inclus dans round_trip_minutes
Distance	500 m au plus jusqu'au point le plus proche	Non enregistrée ; le temps sert de proxy

Ce que Sphère ajoute à l'échelle — En Python

```
import pandas as pd

households = pd.read_csv("wash-household-survey-2024.v1.csv")
litres = households["litres_per_person_day"]

print(f"median {litres.median():.1f} L/p/d")
print(f"below Sphere minimum: {(litres < 15).mean():.1%}")
print(litres.describe().round(1))
```

Ce que Sphère ajoute à l'échelle — En R

```
library(dplyr)

households |> summarise(
  median = median(litres_per_person_day),
  below_15 = mean(litres_per_person_day < 15),
  n = n()
)
```

Ce que Sphère ajoute à l'échelle

- 13,4 % sous 15 litres — avec une médiane de 23,7

Deux erreurs d'unité, et une seule est visible — En Python

```
print(litres.nlargest(8).round(1).tolist())
```

Deux erreurs d'unité, et une seule est visible — En R

```
households |> slice_max(litres_per_person_day, n = 8) |>  
  select(household_id, household_size, litres_per_person_day)
```

Deux erreurs d'unité, et une seule est visible — En Python

```
suspect = households[litres > 80]
recovered = suspect["litres_per_person_day"] / suspect["household_size"]
print(recovered.round(1).describe())
```

Deux erreurs d'unité, et une seule est visible — En R

```
households |>  
  filter(litres_per_person_day > 80) |>  
  mutate(recovered = litres_per_person_day / household_size) |>  
  summarise(min = min(recovered), max = max(recovered))
```

Deux erreurs d'unité, et une seule est visible

- La seconde erreur est la dangereuse — parce qu'elle rend une valeur plus petite et non plus grande

Deux erreurs d'unité, et une seule est visible — En Python

```
short = households[households["round_trip_minutes"].between(1, 8)]  
print(f"{len(short)} households report a round trip of 1 to 8 minutes")  
print(short["water_source"].value_counts())
```

Deux erreurs d'unité, et une seule est visible — En R

```
households |> filter(between(round_trip_minutes, 1, 8)) |> count(water_source)
```

Deux erreurs d'unité, et une seule est visible

- Une erreur qui ramène une valeur dans la plage normale ne se trouve pas en cherchant des valeurs aberrantes — Elle se. . .

L'attente est incluse dans l'aller-retour, et c'est un problème — En Python

```
print("round_trip_minutes includes walking, queueing and return.")  
print(f"over 30 minutes: {(households['round_trip_minutes'] > 30).mean():.1%}")  
print("Queue time alone: not collected. Recommend adding it to the next round.")
```

L'attente est incluse dans l'aller-retour, et c'est un problème — En R

One line in the limitations section beats a queue statistic that is a walk.

La distance n'est pas du tout dans le fichier

- Là où un standard nomme une unité dont vous ne disposez pas, rappez le proxy sous son propre nom — Rapporter le. . .

Rapportez la distribution face au seuil — Exemple

Water quantity, 2,403 households

Median	23.7 L/person/day	
Below Sphere minimum (15 L)	13.4%	323 households
Below 7.5 L	2.2%	54 households

Collection time

Over 30 minutes 39.3% 944 households

Queue time not separated from walking time in this round.

Distance not recorded; time is the proxy and is not equivalent.

11 records held household total litres in the per-person column and were corrected; 11 held collection time in hours and were corrected.

- La quantité et le temps disent si l'eau est arrivée et à quel prix d'effort.

Lire la leçon complète, avec le code exécutable

```
https://data-analysis.cassion.dev/fr/cours/wash-analysis/  
wash-03-quantity-queue-and-distance
```