

Fifteen litres, thirty minutes, five hundred metres

13.4% of these households fall below the Sphere minimum for quantity. Two of the columns that produce that number carry unit errors, and one of them moves a household up a ladder rung rather than off the end of a chart.

Pierre Robentz Cassion

Lesson 3 of 8

WASH Analysis — What Sphere asks that a ladder does not

What this lesson covers

- What Sphere adds to the ladder
- Two unit errors, and only one of them is visible
- Queueing is inside the round trip, and that is a problem
- Distance is not in the file at all
- Report the distribution against the threshold
- What comes next

What Sphere adds to the ladder

Standard	Threshold	In this file
Quantity	At least 15 litres per person per day	litres_per_person_day
Queueing time	No more than 30 minutes	Inside round_trip_minutes
Distance	No more than 500 m to the nearest point	Not recorded; time is the proxy

What Sphere adds to the ladder — In 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))
```

What Sphere adds to the ladder — In R

```
library(dplyr)

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

What Sphere adds to the ladder

- 13.4% below 15 litres — with a median of 23.7

Two unit errors, and only one of them is visible — In Python

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

Two unit errors, and only one of them is visible — In R

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

Two unit errors, and only one of them is visible — In Python

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

Two unit errors, and only one of them is visible — In R

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

Two unit errors, and only one of them is visible

- The second error is the dangerous one — because it makes a value smaller rather than larger

Two unit errors, and only one of them is visible — In 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())
```

Two unit errors, and only one of them is visible — In R

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

Two unit errors, and only one of them is visible

- An error that moves a value into the normal range cannot be found by looking for outliers — It has to be found by a...

Queueing is inside the round trip, and that is a problem — In 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.")
```

Queueing is inside the round trip, and that is a problem — In R

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

Distance is not in the file at all

- Where a standard names a unit you do not have, report the proxy under the proxy's own name — Reporting round-trip time. . .

Report the distribution against the threshold — Example

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.

What comes next

- Quantity and time say whether water arrived and how hard it was to get.

Read the full lesson, with runnable code

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