

Four fifths improved, half with service

82.3% of these households are on an improved source and 56.3% have at least basic drinking water service. The twenty-six points between the two are collection time, and they are the whole reason the ladder has rungs.

Pierre Robentz Cassion

Lesson 1 of 8

WASH Analysis — The ladder is not the source

What this lesson covers

- The number most WASH reports give
- The ladder the SDG indicator actually uses
- Why the thirty minutes is in the definition
- The rung this survey cannot compute
- Report the distribution, not the top rung
- What comes next

The number most WASH reports give — In Python

```
import pandas as pd

households = pd.read_csv("wash-household-survey-2024.v1.csv")

IMPROVED = {
    # the JMP list, delivered water included
    "piped-into-dwelling", "piped-into-yard", "public-tap",
    "borehole", "protected-well", "protected-spring", "tanker-truck",
}

improved = households["water_source"].isin(IMPROVED)
print(f"improved source: {improved.mean():.1%} of {len(households)} households")
```

The number most WASH reports give — In R

```
library(dplyr)

improved <- c("piped-into-dwelling", "piped-into-yard", "public-tap",
             "borehole", "protected-well", "protected-spring",
             "tanker-truck")

households |> summarise(improved = mean(water_source %in% improved), n = n())
```

The number most WASH reports give

- 82.3% — It is a correct calculation of a thing nobody asked about

The ladder the SDG indicator actually uses

Rung	Definition
Safely managed	An improved source on premises, available when needed, and free from contamination
Basic	An improved source, round trip 30 minutes or less including queueing
Limited	An improved source, round trip more than 30 minutes
Unimproved	An unprotected well or spring
Surface water	River, dam, lake, pond, canal, irrigation channel

The ladder the SDG indicator actually uses — In Python

```
def ladder(row):  
    if row["water_source"] == "surface-water":  
        return "surface water"  
    if row["water_source"] not in IMPROVED:  
        return "unimproved"  
    minutes = row["round_trip_minutes"]  
    return "basic" if minutes <= 30 else "limited"  
  
households["service"] = households.apply(ladder, axis=1)  
print(households["service"].value_counts(normalize=True).round(3))
```

The ladder the SDG indicator actually uses — In R

```
households |>
  mutate(service = case_when(
    water_source == "surface-water" ~ "surface water",
    !water_source %in% improved ~ "unimproved",
    round_trip_minutes <= 30 ~ "basic",
    TRUE ~ "limited"
  )) |>
  count(service) |>
  mutate(share = n / sum(n))
```

The ladder the SDG indicator actually uses

Rung	Households	Share
Basic (including on premises)	1,352	56.3%
Limited	626	26.1%
Unimproved	326	13.6%
Surface water	99	4.1%

The ladder the SDG indicator actually uses

- At least basic drinking water service: 56.3% — The source-type figure was 82.3%, and every one of the twenty-six points. . .

Why the thirty minutes is in the definition

- households collect **less** water, so quantity falls below the Sphere minimum more often;
- collection is done by women and girls, so the time comes out of school attendance and paid work;
- households supplement from a nearer, worse source, so the recorded source is not the only source.
- A definition that looks bureaucratic usually encodes a finding — Before arguing that a threshold is arbitrary, look for. . .

The rung this survey cannot compute — In Python

```
on_premises = (households["round_trip_minutes"] == 0)
tested = households["ecoli_cfu_100ml"].notna()

print(f"on premises: {on_premises.mean():.1%}")
print(f"water quality tested: {tested.mean():.1%}")
print(f"both: {(on_premises & tested).mean():.1%}")
```

The rung this survey cannot compute — In R

```
households |> summarise(  
  on_premises = mean(round_trip_minutes == 0),  
  tested = mean(!is.na(ecoli_cfu_100ml))  
)
```

Report the distribution, not the top rung — Example

Drinking water service level, 2,403 households

At least basic	56.3%	1,352
of which on premises	27.7%	665
Limited	26.1%	626
Unimproved	13.6%	326
Surface water	4.1%	99

JMP ladder. Safely managed not computable: availability was not asked and water quality was tested on 33.4% of households.

What comes next

- Water has one ladder and two more are waiting.

Read the full lesson, with runnable code

```
https://data-analysis.cassion.dev/courses/wash-analysis/  
wash-01-the-drinking-water-ladder
```