

Reef data lab: lesson 1 of 4

Clean before you count

Real reef data is messy.
Scientists check it before
they trust it.



MCP divers surveying. Photo: MCP



Meet the dataset

Every row is one survey: two divers, one 30 m tape, one depth zone. Every column is one thing they recorded.

2,720

fish surveys, 2022 to 2026

4,340

coral (seabed) surveys,
2018 to 2026

11

sites: 8 marine reserves and
3 open, unprotected reefs

Teaching dataset based on real MCP survey records from southern Negros Oriental. Site and surveyor names are replaced with codes, enforcement levels are invented, and some values have been changed to make patterns clearer for learning. Not for research use.



How the data was collected

Stratified random sampling

Surveys spread across shallow, medium and deep zones at every site.

30 m transects

Divers count fish along the tape, and record the seabed at 120 points.

Two divers

Each diver counts separately, so their counts can be compared.

Repeated

Sites are surveyed again and again through the year.



Why clean the data?

Every real dataset has mistakes: a typo, a missing value, a number typed in the wrong place. If you analyse them without checking, your answer can be completely wrong.

Scientists call this 'garbage in, garbage out'. Checking data is not boring admin. It is part of the science.



Your mission: find the problems

1. Open the fish surveys (raw) sheet.
2. Sort or filter each column. Look for values that are impossible, misspelled, missing or repeated.
3. Check that date, year and month agree with each other.
4. Record every problem in your change log, and what you did about it.

Look for

Impossible temperatures or visibility

Site names spelled two ways

Blank cells

The same survey twice

A date that can't exist



Spreadsheet tricks

Sort

Biggest and smallest values jump to the top.

Filter

The filter list shows every different spelling in a column.

COUNTIF

Count how often a survey_id appears. More than 1 means a duplicate.

Conditional formatting

Colour cells outside a sensible range.



The surveyor puzzle

Now look at the surveyors. Work out the average total_fish for each surveyor_1. Then work out what share of the fish diver 1 counted, compared with their buddy.

One surveyor stands out. How? And what could explain it?

What should a scientist do about it?



Rules for honest cleaning

1. Always keep an untouched copy of the raw data.
2. Write down every change you make, and why.
3. Only fix a value when you are sure what it should be. Otherwise, flag it or leave it out, and say so.
4. Never delete data just because you don't like the result.



Talk about it

Which problem was hardest to spot?

What might explain the surveyor who stands out: skill, honesty, or something else?

Why do survey teams use two divers?

Next time: do protected areas have more fish?



Photos and sources

Photos: survey divers by MCP. All illustrations drawn for this lesson by Marine Conservation Philippines.

Sources:

- Marine Conservation Philippines. Marine Monitoring.
- Teaching dataset based on MCP survey records, 2018 to 2026, adjusted for learning.

Full links are in the teacher reading. Free to use and share for education. Made by Marine Conservation Philippines. More free lessons: marineconservationphilippines.org/resources/teachers-resources/

