



Clean before you count

Lesson plan: explore a real reef survey dataset, find and log planted data errors, and detect a surveyor whose counts don't fit



Ages	14 to 18 (grades 9 to 12)
Time	60 to 75 minutes
Group	Pairs, one computer per pair
Where	Computer room, or laptops with Excel, Google Sheets or LibreOffice
You need	The slides, the Reef data lab workbook, one cleaning log and data dictionary per pair

What students learn

- How a real reef survey dataset is structured, and how the data was collected.
- How to find common data errors with sorting, filtering and simple formulas.
- How to spot a surveyor whose records don't fit, and how scientists handle it.
- Rules for honest data cleaning.

Before the lesson

1. Put the **Reef data lab workbook** (or the CSV files) where students can open it. Google Sheets opens the .xlsx directly.
2. Print one **cleaning log** and **data dictionary** per pair.
3. Keep the clean fish file for yourself until lesson 2.

The lesson

- 1** **Meet the dataset 10 minutes, slides 1 to 3.** Explain what a row and a column are, and how MCP collects the data. Point out the note that the data has been adjusted for teaching.
- 2** **Why clean? 5 minutes, slide 4.** Garbage in, garbage out.
- 3** **Find the problems 25 minutes, slides 5 and 6.** Pairs hunt for errors and fill in the log.
- 4** **The surveyor puzzle 15 minutes, slide 7.** Pairs compare surveyors. A pivot table or AVERAGEIFS by surveyor_1 works well.
- 5** **Honest cleaning 5 minutes, slides 8 and 9.** Discuss the rules and one or two questions.

Answers and what to listen for

The planted errors (17):

- F2683: site is 'reserve a' (should be Reserve A).
- F0421: site is 'Reserve A' (should be Reserve A).
- F1989: site is 'Open reef1' (should be Open reef 1).
- F0261: site is 'Open reef 3 ' (should be Open reef 3).
- F0843: site is 'Reserve C.' (should be Reserve C).
- F0217: water_temp_c is '82' (should be 30).
- F0606: water_temp_c is '2.9' (should be 29).
- F2521: water_temp_c is '290' (should be 29).
- F0280: visibility_m is '140' (should be 8).
- F0052: visibility_m is '-5' (should be 8).
- F2041: depth is '' (should be Deep).
- F0912: depth is '' (should be Medium).
- F0203: water_temp_c is '' (should be 29).
- F0687: water_temp_c is '' (should be 29).
- F0178: date is '2025-02-30' (should be 2022-10-03).
- F1318: appears twice. Delete one copy.
- F1775: appears twice. Delete one copy.

The date error also disagrees with its year and month columns, which students can use to spot it. Accept 29 or 30 for the 2.9°C slip.

The surveyor puzzle: surveyor S217 is surveyor 1 on 46 surveys. Their surveys average 2,276 fish, against 753 for everyone else, and they counted about 86% of the fish on each survey while their buddies counted the rest, against a usual share of about 56%. They also record several times as many large fish. The buddy counts are normal, which points to S217 rather than the reef. The honest response is to flag these surveys, check with the team and leave them out of the analysis, saying so. Also worth noticing: a handful of surveys list the same code as both divers.

Making it easier or harder

- **Less confident groups:** give pairs the list of error types to look for, one column at a time.
- **Older groups:** write a short data-quality report, as a scientist would, before any analysis.
- **No computers:** print the first 60 rows of the raw sheet. Several planted errors are near the top once sorted by survey_id.

Curriculum links

- **Maths and statistics:** data handling, averages, outliers.
- **Science:** scientific method, reliability of data.
- **Philippine senior high school:** research skills and General Mathematics.
- **IB and GCSE:** evaluating data quality in investigations.

The teacher reading has the science behind every slide, with links to the sources.

Free to use and share for education. Made by Marine Conservation Philippines. More free lessons: marineconservationphilippines.org/resources/teachers-resources/ Photo by MCP. Updated September 2026.