LESSON 6

Collect & Analyse Data

Gather evidence consistently and find honest patterns.

About 105 min | Reading 40 min | Slides 60 min | Quiz 5 min

> **Learning objective.** By the end you will be able to collect data carefully, organise it, and find patterns that answer your research question.

Key points

- Collect data the same way every time, and keep a simple log as you go.
- Summarise numbers with averages and charts; group words into themes.
- A pattern isn't proof of cause. Say clearly what your data can't show.

Learn

Collecting data well means being consistent. Ask every participant the same questions in the same way, and keep your wording neutral so it doesn't push people toward an answer. Record everything as you go, including surprises and mistakes, and keep a simple log of when, where and how each piece of data was gathered. Store it securely and remove names wherever you can.

How you analyse it depends on the kind of data. Numbers (quantitative data) can be summarised with averages, medians, ranges and charts like bar graphs or scatterplots. Words (qualitative data), such as interview answers, open responses or historical documents, are analysed by reading closely, labelling recurring ideas and grouping them into themes. Many projects use existing datasets, from sources like national statistics offices or Our World in Data, instead of collecting new ones.

Be honest about limits. A small or unrepresentative sample can't speak for everyone, and a pattern between two things doesn't prove that one causes the other. Analysis often sends you back a step to collect more data, refine your question or rethink your design. That loop is a normal part of research.

Example

> Ravi surveys 40 classmates about weekly hours spent gaming and their stress level from 1 to 5. In a spreadsheet he calculates the median gaming hours and makes a scatterplot. It shows only a weak pattern, so he writes that the two seem loosely related and that his sample is a single school. His open-ended answers mention 'relaxing with friends' again and again, so he tags that as a theme, which gives him an idea for follow-up interviews.

Across disciplines

Scientists and social scientists often use statistics and charts, and qualitative researchers code interviews for themes. Humanities scholars analyse texts and sources closely for meaning and context, and engineers analyse test results against design requirements.

Watch

Correlation Doesn’t Equal Causation: Crash Course Statistics #8, CrashCourse (12 min): youtube.com/watch?v=GtV-VYdNt_g

> **Takeaway.** Collect data consistently and ethically, use analysis that fits your data type, and report honestly what your results can and can’t show.

Try it: Mini data detective

Name: _____ Date: _____

1. Ask 8–10 friends or family members one neutral question with a number answer and one short open-ended question. Keep it anonymous and low-risk, and get their okay.

2. Enter the number answers in a spreadsheet and find the mean, median and range.

3. Read the open-ended answers, highlight repeated ideas, and group them into two or three themes.

4. Write two sentences: one pattern you found, and one reason to be cautious about it.

Lesson 6 quiz

Circle one answer for each question. Pass mark: 4 of 5.

1. Which statement about collecting more data is accurate?

(A) A bigger sample removes both random error and bias from the results
(B) A bigger sample matters less than using a precise instrument
(C) A bigger sample shrinks random error but does not fix bias
(D) A bigger sample turns a correlation into a causal finding

2. A student studying household spending finds that a few families have very high incomes. Which measure of the centre is safest to report?

(A) The mean, because it uses every value in the data set equally
(B) The range, because it shows the highest and lowest values
(C) The correlation, because it shows how strong the pattern is
(D) The median, because it is less affected by extreme values

3. A biology student randomly assigns seedlings to two light levels, keeping water and soil the same. Taller seedlings grow in more light. What may the student claim?

(A) Only an association, because all biology data are treated as observational
(B) A causal claim about light and growth may be reasonable in that setup
(C) Light causes growth in all plants everywhere, with no limitations
(D) Nothing, because correlations can never be reported in science

4. A student surveying canteen satisfaction on a 1–5 scale finds one rating of 7. What is the best way to handle it?

(A) Change it to 5, the closest valid value on the scale
(B) Delete it quietly so the results look tidy and complete
(C) Keep it as recorded, since removing any data counts as cherry-picking
(D) Apply a cleaning rule set in advance and report the removal

5. One of Noor's interview themes (invented example) is that students choose louder spaces for social reasons. Why does this matter for her simulated results?

(A) It proves that noise is the only cause of lower ratings
(B) It shows her phone-app sound readings were inaccurate
(C) It points to self-selection as a possible confounder
(D) It means her survey sample was random after all

Further reading

- **Analyzing Your Primary Data**, Purdue OWL, Purdue University. How to make sense of interview, survey and observation data. https://owl.purdue.edu/owl/research_and_citation/conducting_research/conducting_primary_research/analyzing_primary_data.html
- **Creating Good Interview and Survey Questions**, Purdue OWL, Purdue University. How to avoid biased, leading and double-barrelled questions. https://owl.purdue.edu/owl/research_and_citation/conducting_research/conducting_primary_research/interview_and_survey_questions.html

- **Summarizing quantitative data**, Khan Academy. Free lessons on mean, median, spread and describing distributions.
  https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data
- **Statistics in Schools**, U.S. Census Bureau. Free real-data activities for secondary students across many subjects.
  https://www.census.gov/programs-surveys/sis.html