

Data Evidence Packet

Optional practice for Topic 5: not graded, and nothing to submit (changed 10/9; eCampus shows current grading). Protect raw data, document cleaning, verify formulas and chart claims, and keep the final analysis defensible.

Name	Date	Status
		Optional practice: not graded, nothing to submit.

DEFAULT DATA

Use the supplied synthetic CSV
Do not replace it with a real gradebook, survey, customer file, health record, workplace export, or personally identifiable dataset.

OPTIONAL

Nothing to submit
Topic 5 has no graded assignment. Work through any part at home as practice, and keep it for yourself.

Access: Use the instructor-supplied CSV with hand checks (the default), the pre-written AI proposals below, or an AI tool of your choice coached by the two prompts below; the hand checks stay yours. No AI account, paid product, live upload, spreadsheet MCP connection (MCP, the Model Context Protocol, lets an AI tool connect to outside apps such as a spreadsheet), or Copilot license is required. This editable HTML page is the primary digital route; it saves locally in your browser. You may instead use a spreadsheet editor and calculator, the tagged printable PDF, or paper/manual work.

Optional prompts for any AI (no skill or account is needed). For a formula: "Explain this formula: what it divides by what and the unit, work one row, and say what happens with a blank, a zero or text. Don't change my sheet." For a whole workbook: "Check this workbook for formula errors, inconsistent ranges, typed-in numbers and unit mismatches. Report what you find; don't change anything." Then do the hand check yourself; you still decide which rows count.

Data boundary: Never upload private, confidential, proprietary, indirectly identifying, education, health, employment, customer, or financial records to any tool. Never share credentials, account screens, device evidence, or full private chat history. Product approval and dataset authorization are separate requirements.

How to complete this packet: This page saves your draft only in this browser's local storage; it does not send responses anywhere. Clearing browser/site data may erase the draft, so keep a backup if needed. Use only the synthetic dataset. "Print or save a copy" expands your responses and omits the later Completed Example. You may instead print this page and write by hand.

Work plan: Reserve one focused 90-minute block. You may cross-reference completed issue-log rows instead of rewriting the same decision in Parts 3 and 6.

Instructor-Supplied Simulation: Pre-written AI Proposals

These AI proposals were pre-written for class; no prompt is sent to a tool. For each proposal, accept, reject, or revise it and explain the source evidence.

Closed source: week-05-synthetic-admin-dataset.csv. It contains invented event records and deliberately includes a lowercase category, missing attendance, possible duplicate, text in a cost field, and possible attendance outlier.

Field guide and units: Event is a label; Date uses YYYY-MM-DD; Category is a grouping label; Expected and Attended are counts of people; Cost should be cash expense in USD; Sponsor is an explicitly fictional exercise label; Notes preserve source comments. The file has no unique event ID, so it cannot resolve the duplicate candidate by itself. The text `free` is deliberately ambiguous and is not a verified numeric zero.

1. Standardize arts to Arts.
2. Fill Career Panel attendance with 40 because Expected is 40.
3. Delete the second Robotics Night row as a duplicate.
4. Change Food Drive cost from `free` to 0.
5. Change Debate Prep attendance from 85 to 18 because 85 looks too high.

Formula proposals: (A) `=F2/E2` for cost per attendee; (B) `=AVERAGE(E2:E9)` for “average attendance across all events”; (C) copy Expected into every blank Attended cell.

Output A — first chart and memo proposal: Title “Best Event Categories by Cost”; all eight rows included; axis labeled “Count”; no formula, exclusions, limitation, or disclosure shown; memo says, “STEM events are most efficient because they have the lowest cost.”

Output B — revised chart plan and memo: Title “Recorded cost per attendee for verified event rows”; include Art Jam, AI Poster Studio, and Parent Info Night; axis labeled USD per attendee; list excluded unresolved rows; disclose the simulation; memo says Parent Info Night is lowest only among these three rows and this metric does not measure event quality.

Part 1 — Permission, source, and Raw

Source and authorization note

Raw preservation evidence

Part 2 — Five-issue data log

Complete all five rows. A finding is not automatically a repair. Name the exact row/field, issue type, decision, evidence needed, and downstream effect.

Observed field or row	Issue type	Keep, fix, flag, exclude, or ask?	Reason, source check, and effect
Art Jam category: arts			
Career Panel Attended: blank			
Second Robotics Night row			
Food Drive Cost: free			
Debate Prep Attended: 85			

Part 3 — Analysis brief and cleaning decisions

Source-bound analysis brief

Decision summary without repetition

One before/after cleaning example

Part 4 — Formulas and independent row checks

Explain two formulas in plain language and check at least two row results independently. State units, eligible rows, blank/zero handling, and unresolved exclusions.

Formula or calculation	Plain-English meaning and unit	Manual row check	Edge rule / limitation

Part 5 — Chart, critique, and three-bullet memo

Chart or chart plan

Critique of the simulated “STEM is most efficient” claim

Three-bullet insight memo

Part 6 — Reflection eval, trace, rejection, and disclosure

Use exactly five checks for both outputs: (1) named eligible source rows, (2) formula shown and spot-checked, (3) metric labels and units, (4) claim fits the evidence, and (5) limitations plus assistance disclosure. Mark missing evidence **Not demonstrated**; do not assume it passed.

Supplied Output A vs revised Output B

First weak workflow step and component check

One rejected proposal and evidence

AI tool, pre-written proposals, or hand-check disclosure

Self-check

- ☐ I used the supplied synthetic CSV or another dataset explicitly authorized for this practice.
- ☐ I preserved Raw and completed all five issue-log rows.
- ☐ I did not invent missing values or silently convert estimates into observations.
- ☐ I explained two formulas and independently checked at least two rows with units and edge rules.
- ☐ My chart and three memo bullets identify eligible evidence and limitations.
- ☐ I compared A and B, located a weak step, rejected one proposal with evidence, and disclosed meaningful assistance.
- ☐ My copy contains no private data, credentials, account/device evidence, or full private chat history.

Quick vocabulary

Raw

The unchanged authorized source. Cleaning happens in a copy or separate representation.

Eligible rows

The rows a documented rule allows into a specific calculation or chart.

Edge rule

What the workflow does with blanks, zeros, text values, duplicates, or other special cases.

Claim strength

How certain, broad, comparative, or causal a statement is relative to its evidence.

Component check

A focused test of one weak step before rerunning the complete workflow.

What strong evidence looks like

Use this to check your own practice; nothing is graded. Strong evidence is complete, specific and traceable; fabricated values, unsupported claims and private data are never acceptable.

Criterion	Evidence expected
Permission, source, and Raw	Authorized synthetic source, protected-data boundary, and unchanged Raw evidence are explicit.
Five-issue log and cleaning decisions	All five issues are classified; decisions distinguish observation from repair and name evidence/ downstream effects.
Analysis brief and change evidence	Brief fixes source, goal, fields, outputs, constraints, and checks; before/after decisions are traceable.
Formulas and row verification	Two formulas are explained; at least two results are checked independently with units, eligible rows, and edge rules.
Chart and memo	Metric, rows, labels, units, grouping, claim, three bullets, and limitations are aligned.
Reflection, trace, rejection, disclosure, completeness	A/B checks, first weak step, component check, rejected proposal, disclosure, and a readable packet are complete.