

THURSDAY, AUGUST 20 4:00–5:00 P.M.

Lab 2: From a draft file to a policy briefing

Commission the work, test the evidence, and decide what can be said

The assignment

You are preparing a short briefing for a policy director. Using the available project files, construct a defensible dataset and report how income and under-five mortality differed across World Bank income groups in 2022. You have one hour. Decide what evidence is trustworthy, document any material data problem, and state what the results do—and do not—show.

The analysis dataset must contain one observation per economy–year from 2000 through 2022, GDP per capita (PPP), under-five mortality, a natural-log measure of income, and the economy information needed for interpretation. Observations without positive income or observed mortality do not enter this analysis.

Four working products

1. `code/02-health-agent-pipeline.R`
2. `outputs/02-health-analysis.csv`
3. `outputs/02-health-briefing-by-income.csv`
4. `BRIEFING.md`

The first two products come from a bounded Codex commission and your review of it. You add the grouped briefing calculation to the same R script. The final briefing is yours: Codex may check it, but it may not decide what you release.

Route through the hour

Minutes	Work
00–08	Read the request; specify the unit, key, and one failure that could change the briefing.
08–18	Ask Codex to review the draft, then independently verify one important finding in R.
18–28	Decide which source is authoritative and commission the construction.
28–40	Observe the run; inspect the generated script and saved analysis file.
40–48	Audit the candidate and make an accept, revise, or reject decision.
48–56	Build and verify the grouped 2022 briefing table.
56–60	Complete <code>BRIEFING.md</code> and make the release decision.

Can we trust the draft?

Decision gate 1. What would make the briefing wrong?

Before opening the draft, write down:

- what one row in the intended analysis file should represent;
- which variables should identify that row uniquely; and
- one record-level problem that could change a count, median, plot, or conclusion in the director's briefing.

Open `code/lab-2-starter.R` in RStudio and run the section titled **Gate 1**. Read the results before asking Codex to review the project.

Act as a skeptical data reviewer for a policy director's briefing on income and under-five mortality.

Read `LAB-2-HANDOUT.pdf`, `data/lab-2/health-briefing-draft.csv`, `data/derived/math-camp-wdi-2000-2022.csv`, and `data/documentation/indicator-dictionary.csv`. Work read-only.

Assess whether the draft can support the requested analysis. Examine the unit of observation, economy-year key, period, required variables, missingness, ranges, and any disagreement among project files. For every consequential claim, identify the source file, variable, economy-year when relevant, and observed value. Separate what the files establish from your explanation of how a problem might have arisen.

End with (1) a release recommendation, (2) the two findings that matter most for the policy briefing, and (3) R commands that independently verify those findings. Do not edit, deduplicate, or rebuild any file.

Decision gate 2. Can you verify the agent's most important finding?

Choose the Codex finding most likely to change the briefing. Write and run a small R check that tests it directly against the project files. Record whether the evidence confirms, contradicts, or only partly supports the finding. Then separate what the files establish from Codex's explanation of why it happened.

Codex finding

R check and result

Unsupported explanation

Briefing verdict

Can we trust the constructed dataset?

Before commissioning the build, answer this question:

If two project files disagree, what makes one authoritative: its filename, its modification date, an agent's confidence, or the project's documentation and provenance? What evidence would persuade you?

Then give Codex the following commission.

Prepare the analysis dataset for this policy briefing: How did income and under-five mortality differ across World Bank income groups in 2022?

Use project documentation and the available data to determine which input is authoritative. State the evidence for that choice. Create and run `code/02-health-agent-pipeline.R`, and save `outputs/02-health-analysis.csv`.

The analysis file must have one row per iso3c + year for 2000 through 2022 and contain iso3c, country, region, current income classification, year, GDP per capita (PPP), under-five mortality, and log income. Exclude missing or nonpositive income and missing mortality. Define log income as the natural log of GDP per capita (PPP).

You may create or edit only `code/02-health-agent-pipeline.R` and files under `outputs/`. Do not modify source data, setup files, the starter audit, or the independent verifier. Run the pipeline. Report the input chosen, exclusions, files changed, checks performed, and any decision that remains contestable.

While Codex works, watch which files it reads, which actions it requests, what input it selects, and whether it actually runs the code it writes. Then inspect the source choice, selection rules, transformation, tests, and saved output.

Open `code/lab-2-starter.R` and run the section titled **Gate 3**.

Decision gate 3. Would you release the constructed dataset?

Make three separate judgments:

1. **Execution:** Did the script run and create the requested file?
2. **Evidence:** Does the output have the required shape and agree with the documented project data?
3. **Method:** Would the result survive a change in row order or an unsupported source file?

Decide *accept*, *revise*, or *reject*. A successful run is not itself a release decision. If every check passes, name the strongest failure you attempted to falsify and the evidence that ruled it out.

Can we trust the briefing?

Use only a candidate you accepted or revised. Add a clearly marked STUDENT BRIEFING TABLE section to code/02-health-agent-pipeline.R. Begin by comparing rows with distinct economies:

```
health_analysis |>
  filter(year == 2022) |>
  group_by(income_level_current) |>
  summarise(
    rows = n(),
    economies = n_distinct(iso3c),
    .groups = "drop"
  )
```

Why can rows and economies differ? Which quantity belongs in this briefing? Resolve that question before creating one row per current income group with the number of economies, median GDP per capita (PPP), and median under-five mortality. Arrange the groups from lowest to highest median mortality and save outputs/02-health-briefing-by-income.csv.

Ask Codex to independently recompute the grouped values from the accepted analysis file and compare them with your saved table. Codex may propose a correction, but it may not rewrite your calculation or briefing.

Decision gate 4. What can you tell the policy director?

Complete BRIEFING.md. Your release decision must answer:

Would you send this briefing to the policy director? Identify the strongest evidence supporting your decision, one failure you ruled out, one numerical conclusion the table supports, and one statement you would refuse to make.

One row in the grouped table represents a current income group in 2022. This is an unweighted descriptive comparison: each observed economy contributes once. Current income classification in the snapshot is not a historical country-year classification, and the grouped pattern does not identify a causal effect of income on mortality.

Keep the prompt, generated script, audit evidence, grouped table, and final briefing together. The point of the lab is not to prove that Codex is right or wrong. It is to make enough of the analytical process visible that you can tell.