

Claude Code for Accounting Research

Module 2 — The Empirical Loop on SEC FSDS

John Barrios

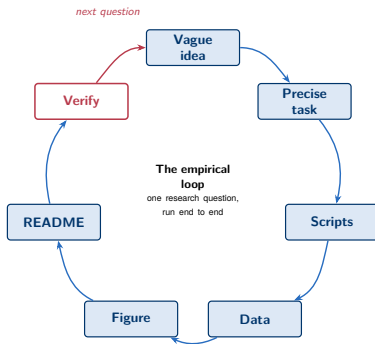
Yale School of Management

Today's Arc

- **0:00–0:40** Lecture: dataset anatomy, the loop, reproducibility habits
- **0:40–1:15** Live demo: R&D intensity by industry, empty folder to figure
- **1:15–1:25** Break
- **1:25–2:45** Lab: your own measure, empty folder to verified figure
- **2:45–3:05** Debrief, exit ticket, **WRDS checkpoint (hard gate)**

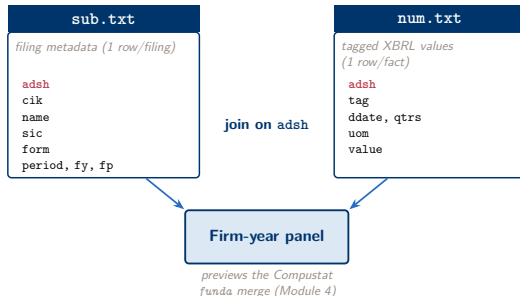
Recap: Yesterday's Model, Today's Test

- Module 1: files persist, trust but verify, be specific — one script, one number
- Module 2: one accounting question, the *full* pipeline, end to end
- Every stage is a file on disk, not a chat reply



SEC FSDS: Two Files, One Key

- `sub.txt` — one row per filing:
adsh, cik, sic, form, period
- `num.txt` — one row per tagged value:
adsh, tag, ddate, qtrs, value
- Join key: adsh — state row counts before and after



Free, quarterly, credential-free — and it previews the shape of Compustat.

Vague Idea → Precise Task, Again

- “How has R&D intensity evolved across industries?” — a good question, an unusable prompt
- Module 1’s four components: file, operation, aggregation, output
- Today: a multi-file pipeline adds two more decisions — the join, the paper trail

Decomposing the R&D Question

- **Which files?** which FSDS quarters, which two files inside each
- **Which join?** on what key, checked how
- **Which measure?** R&D over assets — which XBRL tags, filtered how
- **Which aggregation?** what statistic, by what grouping, over what time unit
- **Which output?** what figure, saved under what name, alongside what table

Reproducibility as Code, Not Chat Lore

- **Empty-folder discipline** — `mkdir ~/lab2 && cd ~/lab2 && claude`
- **Named-script discipline** — delete the chat: does the pipeline still run?
- **Paper-trail prompting** — request `DECISIONS.md` / `LOG.md` / attrition rows *before* any code runs

The Opening-Prompt Template: Six Parts

DATA: *FSDS quarters, sub.txt + num.txt, columns named*

JOIN: *merge on adsh; state row counts before/after; flag unmatched*

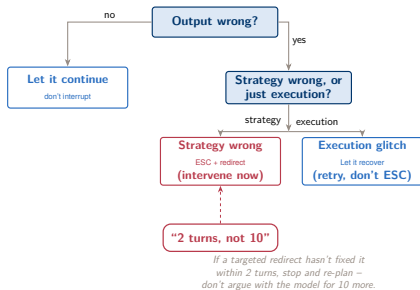
MEASURE: *R&D expense / total assets; tags named; qtrs handling; form filter*

AGGREGATION: *median by SIC2 $\times$ year, across all quarters*

OUTPUT: *rd_intensity.png + rd_intensity_by_sic2_year.csv, named in advance*

PAPER TRAIL: *named scripts only; DECISIONS.md; LOG.md; attrition row every sample-size change*

Iteration and Recovery



- First figure is (almost) never right — normal, not broken
- Small targeted rounds: axis/labels, then bucket granularity — count them
- Watch autonomous recovery (the 403); intervene only on strategy

Sanity Checks: Two, Minimum

- **Benchmark comparison** — pharma/tech should beat utilities/retail; does it?
- **Plausibility check** — does the series show a signature around 2020?
- Name both checks *before* the figure renders

Common Failure

Treating a plausible-looking figure as a verified one

Clean axis labels and smooth lines say nothing about whether the tag selection, filter, or join was correct. Cosmetic polish and verification are different things.

Verify This

Name your two checks before you look at the figure

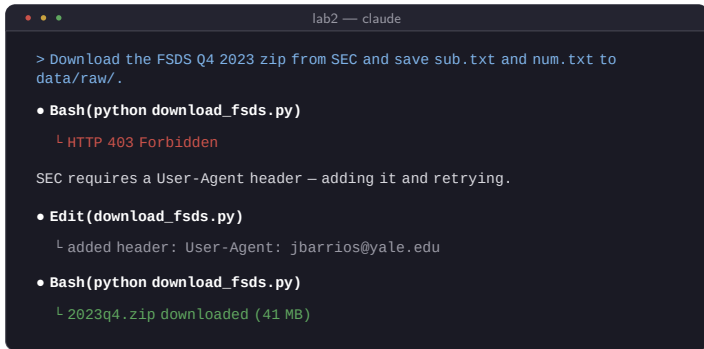
Decide the benchmark comparison and the plausibility check *before* the figure renders — and write both, with sources, into the README.

The Three Mantras, Carried Forward

Same three sentences, more work at once

1. Files persist; context doesn't — DECISIONS.md, LOG.md, attrition rows, at every stage
2. Trust, but verify — two named sanity checks, not a vague sense it looks fine
3. Be specific — six parts now: data, join, measure, aggregation, output, paper trail

What You'll See: Error Recovery



```
lab2 — claude

> Download the FSDS Q4 2023 zip from SEC and save sub.txt and num.txt to
data/raw/.

• Bash(python download_fsds.py)
  ↳ HTTP 403 Forbidden

SEC requires a User-Agent header — adding it and retrying.

• Edit(download_fsds.py)
  ↳ added header: User-Agent: jbarrios@yale.edu

• Bash(python download_fsds.py)
  ↳ 2023q4.zip downloaded (41 MB)
```

Stylized session transcript (not a screenshot).

Autonomous error recovery: read the error, fix the cause, retry — no prompt from you.

A bare 403 Forbidden, diagnosed and fixed by the agent without you typing a word —
watch, don't drive.

Next: Live Demo (0:40--1:15)

- Opening prompt read aloud, slowly — the teaching object (8 min)
- Download + 403 recovery — watch, don't drive (5 min)
- Clean + join on `adsh`; attrition rows print live (7 min)
- First-draft figure → two refinement rounds (8 min)
- Sanity checks, sign off, README written last (7 min)

What You'll See: Prompt and Join

- Six-part prompt projected and read aloud — not retyped, not rushed
- Same 403-and-recover pattern from Module 1, on a bigger file
- Row counts before/after the join; unmatched `adsh` flagged; one `DECISIONS.md` entry read aloud live

What You'll See: Iteration and Sign-Off

- First draft renders wrong on purpose — count the rounds together
- Round 1: axes and labels. Round 2: industry-bucket granularity
- Both sanity checks run and narrated; README written only after they pass

Next: The Lab (1:25--2:45)

- Empty folder → verified figure + README, one continuous 80-minute block
- Choose one measure: **base** leverage / **standard** R&D variant / **stretch** revenue concentration
- Disjoint from today's demo variable — your verification has to be real
- Requirements: named scripts, DECISIONS.md + LOG.md requested upfront, attrition table, one figure

Lab Verify Checklist

- ☐ Pipeline reruns start-to-finish from the README, in a fresh clone
- ☐ `sub/num` joined on `adsh` — row counts stated before/after
- ☐ Attrition table present; every sample-size change explained in `DECISIONS.md`
- ☐ Two sanity checks written down, with sources
- ☐ Figure produced by a script — no interactive-only outputs

WRDS Checkpoint --- Hard Gate

Due today: submit one of the two

1. A `tunnel_status` screenshot: `ssh wrds` → Duo push → success
 2. A Duo problem flag on your exit ticket
- Triaged tonight; fixes announced Module 3 morning
 - Two days of runway is the minimum lead time to fix an account or Duo problem
 - No working WRDS by Module 4 → pre-assigned to the public-data fallback track, decided now, not scrambled later

Debrief and Exit Ticket

- Whose first-draft figure was actually right? (Expect close to zero — that's the point)
- What did Claude recover from alone; where did you have to step in?
- Exit ticket: standard 3 questions + sub-agent teaser (why might one have appeared today?)
- **WRDS checkpoint submitted — see previous slide**