Your Data Folder, File by File — Step Zero

Everything on this site assumes you have a folder named `dashboard_data` containing your college's numbers. **You provide every file in it** — this page tells you exactly what the folder looks like, what goes in each file, and the rules the data must follow.

You don't have to start from a blank page: the **sample data download** on this site is a complete, filled-in example (synthetic data for a fictional college). Copy its shapes and replace the numbers with yours.

The file tree

```
dashboard_data/
├── README.md                          ← describes each file (start from the sample's)
├── data_notes.json                    ← your college's name, data source, disclaimers
├── metric_definitions.json            ← plain-language definition of every metric
├── narrative.json                     ← your college's VAR narrative comments
├── kpis_by_year.csv                   ← the headline numbers, one row per year
├── awards_by_type_year.csv            ← degrees & certificates by type and year
├── equity_outcomes.csv                ← outcome rates by student group
├── equity_participation_share.csv     ← who programs reach vs. the college overall
├── students_served_by_program.csv     ← unduplicated students served, by program
├── services_by_program_term.csv       ← service contacts by program and term
├── services_by_category_term.csv      ← service contacts/dollars/hours by category
└── direct_aid_by_type.csv             ← direct aid disbursed, by type and year
```

Your college seal or logo is **not** part of this folder — you'll provide it separately when you build (the step-by-step guides show where).

What goes in each file

File	One row per	Columns
`kpis_by_year.csv`	academic year	`academic_year`, `headcount`, `awards_conferred`, `transfers_to_four_year`, `median_units_at_completion`, `timely_completion_rate_3yr`
`awards_by_type_year.csv`	year × award type	`academic_year`, `award_type` (AA / AS / ADT / Certificate), `awards`
`equity_outcomes.csv`	metric × student group	`metric`, `group_type` (all / race_ethnicity / gender / pell / special_population), `group`, `students`, `rate`

File	One row per	Columns
`equity_participation_share.csv`	program × student group	`program_code`, `program`, `students_served`, `group`, `share_of_served`, `share_of_collegewide`, `proportionality_index`, `di_flag` (Y/N), `verdict` (plain-language, e.g. "No disproportionate impact found")
`students_served_by_program.csv`	year × program, plus an ALL row	`academic_year`, `program_code`, `program`, `students_served` (unduplicated)
`services_by_program_term.csv`	term × program	`term`, `academic_year`, `program_code`, `program`, `services_delivered`, `students_served`
`services_by_category_term.csv`	term × service category	`term`, `academic_year`, `service_category`, `unit` (count / dollars / hours), `total_amount`, `students_served`
`direct_aid_by_type.csv`	year × aid type	`academic_year`, `service_code`, `aid_type`, `unit`, `total_amount`, `students_receiving`
`narrative.json`	—	your college's short VAR narrative comments, organized by Vision 2030 goal (a few sentences each — they appear on the dashboard next to the charts they describe)
`metric_definitions.json`	—	for every metric: a plain-language definition, how it's computed, and any caveat
`data_notes.json`	—	your college's name, data source, years covered, suppression rule, disclaimers — this feeds the dashboard's methodology footer, and **it's where the dashboard learns your college's name**
`README.md`	—	one paragraph on what the folder is, plus the file table and your suppression rules — copy the sample's and edit

The rules your data must follow

These aren't style preferences — the dashboard is public, and the prompts are written to expect these conventions.

1. **Aggregate data only.** Counts, rates, and totals. Never one row per student; no names, IDs, or birthdates anywhere.
2. **Suppress small counts — and write them as `<10`.** Any count based on fewer than 10 students appears in the cell as the literal text `<10`. Don't delete the value or leave it blank — the dashboard renders `<10` as a visibly masked cell.
3. **Watch the subtraction trap.** If a published total would let someone recover a masked number by subtracting the visible ones, mask the next-smallest category too.
4. **No rate with a small denominator.** Don't publish any rate computed from fewer than 10 students.
5. **Blank means "not yet observable" — never zero, never suppressed.** Leave a cell blank only when the data genuinely can't exist yet (e.g. a rate for the newest cohort). If it's suppressed, write `<10`; if it's zero, write 0.
6. **Never sum different units.** Service contacts, dollars, and hours each stay in their own lane — the `unit` column keeps them apart.
7. **Protect sensitive groups.** Don't cross sensitive program membership (disability services, foster youth, veterans, CalWORKs/CARE) with demographics.

When in doubt about any of this, ask your institutional research office before the data leaves your college.

How to actually build this

Your institutional research office almost certainly has these numbers already — in MIS referential reports, program review files, or Student Centered Funding Formula reporting. The work is reshaping them into the files above, and Claude is good at that:

1. Download the sample data from this site and this spec.
2. Give Claude both, along with your college's **already-aggregated** reports, and ask: *"Help me reshape my reports into the dashboard_data files this spec describes."*
3. Only aggregated numbers, please — the same safety rules apply to data you give Claude for reshaping as to data you publish.

Missing a file's worth of data? That's fine — include what you have, note what's missing in `README.md` and `data_notes.json`, and ask Claude to adapt the dashboard prompt to skip what you don't have (the step-by-step guides suggest doing exactly that as your first step).