

Dashboards and Reporting

Build a dashboard someone opens twice, and write the donor report around it — including the refresh, the caveats and the definition panel that stops the monthly argument.

Instructor	Pierre Robentz Cassion
Level	Intermediate
Learner hours	18 h
Taught in	Python · R
Updated	July 29, 2026
Sectors	WASH · Nutrition · Protection · Public Health · Food Security
Standards and methodologies	OECD DAC evaluation criteria · Core Humanitarian Standard (CHS) · UNICEF indicator definitions · Sphere Standards · Integrated Food Security Phase Classification (IPC)

Read and run the course online at data-analysis.cassion.dev/courses/dashboards-and-reporting

Course content is licensed CC BY 4.0. You may adapt and redistribute it, including commercially, with attribution to Cassion.

What you will be able to do

- Decide what a dashboard is for, and refuse the twenty-tile request with something better
- Lay out a page so the three numbers that drive a decision are read first
- Carry provenance on every figure — the as-of date, the source file, the refresh cadence
- Write an indicator definition panel so a number and its definition travel together
- Write a narrative report whose findings, limitations and recommendation are separable
- Present a finding to a steering committee when it is not the finding they wanted

Prerequisites

- Data Visualisation for Programmes

Contents

I	What it is for	1
1	The twenty-tile request is a symptom	2
1.1	The request as it arrives	2
1.2	The question to ask instead	2
1.3	What each of the seven dashboards here is for	2
1.4	Three tiles, not twenty	3
1.5	What to do with the other seventeen	3
1.6	The tile that should not be there	3
1.7	Report it whole	4
1.8	What comes next	4
2	The three numbers first, and everything else lower	5
2.1	The order a page is read in	5
2.2	The tile	5
2.3	What goes behind a link	6
2.4	The five-second test, applied to a page	6
2.5	Print, and the phone	6
2.6	What the layout cannot fix	6
2.7	Report it whole	7
2.8	What comes next	7
II	What travels with the number	8
3	As of when, from what, and how often	9
3.1	The three questions	9
3.2	What the answers look like	9
3.3	Refresh, stated as a cadence and an owner	10
3.4	Completeness belongs next to the number	10
3.5	What the platform's dashboards state	10
3.6	Where provenance is thin, say so	10
3.7	Report it whole	11
3.8	What comes next	11
4	The panel that caught a wrong number in this course	12
4.1	What happened	12
4.2	Why the documentation was the one that was right	12
4.3	What a definition entry contains	13
4.4	Where the panel lives	13
4.5	The check the panel makes possible	14
4.6	The argument this prevents every month	14
4.7	Report it whole	14
4.8	What comes next	15

III	The report around it	16
5	Findings, limitations, recommendation — in that order and separable	17
5.1	Who reads what	17
5.2	The order the platform's reports use	17
5.3	The recommendation paragraph	17
5.4	Limitations that are not apologies	18
5.5	The three sentences every report needs	18
5.6	Length, and what to cut	19
5.7	Report it whole	19
5.8	What comes next	19
6	An annex of only successful analyses is the one nobody believes	20
6.1	What the annex is for	20
6.2	The test table	20
6.3	Exclusions, with counts	20
6.4	Deviations from the plan	21
6.5	The exploratory section, labelled	21
6.6	What an annex should not contain	21
6.7	The reproduction note	22
6.8	Report it whole	22
6.9	What comes next	22
IV	The room	23
7	Presenting the finding nobody commissioned	24
7.1	The three unwelcome findings	24
7.2	Lead with the decision, not the finding	24
7.3	Bring the thing that would have been true	24
7.4	What not to do	25
7.5	The questions that will come, and the answers	25
7.6	The number in the last three proposals	25
7.7	Report it whole	26
7.8	What comes next	26
8	Seven dashboards, and what each one refuses to show	27
8.1	What is actually shipped	27
8.2	The four constraints that decided the design	27
8.3	What each one is for, and what it refuses	27
8.4	The one that found an error	28
8.5	What these dashboards do not do	28
8.6	The pattern, portable	28
8.7	Report it whole	29
8.8	What comes next	29

About this course

A dashboard that nobody opens twice is a dashboard that failed, and the reason is almost never the charting. It is that the page answers no question anyone was about to decide — so the first lesson is about refusing a request rather than fulfilling it.

This course teaches against seven dashboards the platform actually ships. They are static HTML, built from committed datasets by `pnpm project:dashboards`, and every claim the course makes about them is checkable.

Three of those claims shape it.

A dashboard with no JavaScript is not a limitation here, it is the design. The Content-Security-Policy forbids inline script bodies, and the audience works offline and on metered connections. What is left — tiles, a table, an embedded figure, about eight kilobytes — opens on a phone in a field office and prints from a laptop with no network at all.

Every one of the seven ends with a section saying what it does not show. The protection dashboard names the nineteen cells of one that a four-way disaggregation would produce. The vaccination one says coverage above 100% is a denominator rather than performance. That section is the part a reviewer reads first and the part that is written last.

A definition panel is not documentation, it is a check. Building the WASH dashboard from the same file as the project's indicator documentation surfaced a disagreement — 82.3% against 79.7% on improved water coverage — because the course had omitted delivered water from the JMP improved list. The documentation was right, the lesson was wrong, and nothing would have found it except two artefacts computing the same indicator and being compared.

Both Python and R throughout. You need *Data Visualisation for Programmes* first: this course assumes you can already choose a mark and put an interval on it, and spends its time on what surrounds the chart rather than on the chart.

Part I

What it is for

CHAPTER 1

The twenty-tile request is a symptom

Lesson 1 of 8 · 135 min

Somebody asks for a dashboard with every indicator on it. That request is real and the specification in it is wrong, and the useful response is a question rather than a page — because a dashboard that answers no pending decision is one nobody opens twice.

1.1 The request as it arrives

"Can you build us a dashboard? We want coverage, admissions, defaulting, stock-outs, staffing, funding burn, beneficiary numbers by district and by month, and the indicators from the logframe."

That is twenty tiles and it is a real request. Somebody wants to stop asking for numbers by email, and the dashboard is what they have a word for.

It is also a specification that guarantees a page nobody opens twice. Twenty numbers with no ordering is a table with more colours, and the reader has to supply the question the page should have answered.

1.2 The question to ask instead

"What decision are you about to make, and what would change it?"

Ask it once and one of three things happens, and all three are useful.

They name a decision. "Which six communities get rehabilitation." Now you know the page has one job, the ordering is by that job, and most of the twenty tiles are not on it.

They name a recurring meeting. "The monthly cluster call." Then the dashboard's job is to make that call shorter, and the design question is what gets argued about every month — which is usually a definition, not a number.

They cannot answer. That is the most common outcome and it is a finding rather than a failure. It usually means the numbers are wanted for assurance rather than for a decision, and the honest response is a monthly one-page summary, not a dashboard.

Write the answer down at the top of the page, in the words they used. Every later argument about what to include is settled by reading it.

1.3 What each of the seven dashboards here is for

Every project on this platform states its decision in the entry, and the dashboard is built against it.

Project	The decision it informs
WASH coverage	Which six of eighteen communities get water point rehabilitation
Nutrition screening	Where the next screening round and CMAM capacity go
Protection pathway	Which gate in the referral pathway to fix first
Food security	Which instrument the caseload figure is based on
School attendance	Which schools need an attendance response, and which look bad for a reporting reason
SMART survey	Whether the survey's quality permits publishing a prevalence at all
Vaccination coverage	Whether a coverage figure reflects service or a denominator

None of them is "monitor the programme". Each is a sentence with a verb in it, and the tiles are chosen by whether they move that verb.

1.4 Three tiles, not twenty

The rule that survives contact with a real request.

Three headline numbers, at the top, in the size of a headline. They are the ones the decision turns on.

One table underneath, which is the disaggregation the decision needs — by district, by commune, by stage.

One figure, if a shape matters. Most pages need none; a pathway or a trend needs one.

Everything else goes behind a link, or into the report, or nowhere.

```
TILES = ["households surveyed", "improved source", "basic service", "no E. coli"]
assert len(TILES) <= 4, "if this fails, something has to leave the page"
```

```
# An assertion in the build is a more durable specification than an agreement.
```

Four is the practical ceiling and three is better. A row of tiles is read left to right and then abandoned.

1.5 What to do with the other seventeen

They are not worthless — they are the wrong artefact.

The ask	Where it belongs
Every logframe indicator	The quarterly report's annex table
Historic monthly series	A downloadable CSV, linked from the page
Beneficiary counts by everything	A pivotable export, not a page
Anything updated yearly	The report, not the dashboard
Anything nobody could name a use for	Nowhere, and say so

Offer the alternative in the same sentence as the refusal. "Not on the dashboard, but I will put every logframe indicator in the quarterly annex and send you the CSV" is a different conversation from "no".

1.6 The tile that should not be there

A tile earns its place only if a plausible value would change what somebody does.

```
def earns_its_place(tile) -> bool:
    """Would any value of this number change a decision this quarter?"""
    return tile.decision is not None and tile.threshold is not None
```

```
# If you cannot state the threshold, the tile is decoration.
```

Ask for the threshold. "At what value of this would you do something different?" A tile whose owner cannot answer is a tile that is being watched rather than used, and watching is what a monthly report does more cheaply.

Total beneficiaries is the usual example. It goes up, everybody likes it, and no value of it changes anything. It belongs in the report.

1.7 Report it whole

Dashboard scope, WASH coverage project

Question the page answers: which six of eighteen communities receive water point rehabilitation in the next funding cycle.

On the page: households surveyed, improved source, basic service, share with no detectable E. coli; the district table; the source-against-service figure.

Requested and not included: monthly trend (the survey is a single round), every logframe indicator (quarterly annex), beneficiary totals (no threshold could be stated), staffing and funding burn (a different audience and a different page).

Agreed with the WASH coordinator on 2025-02-11. Revisit when the second survey round exists, which makes a trend meaningful for the first time.

The "requested and not included" paragraph is the deliverable of this conversation, and writing it down is what stops the same twenty tiles being requested again in June by someone who was not in the room.

1.8 What comes next

A page with three numbers on it can still bury the one that matters. The next lesson is about the order they are read in, and about what belongs behind a link rather than on the page.

CHAPTER 2

The three numbers first, and everything else lower

Lesson 2 of 8 · 135 min

A reader gives a dashboard about eight seconds before deciding whether it is for them. What they see in those eight seconds is decided by the top of the page, and most dashboards spend it on a logo, a date range picker and a total.

2.1 The order a page is read in

Top-left, then across, then down. That is not a preference, it is what a reader does before deciding whether to keep reading — so the page's argument has to be in that order.

Position	Carries
Title and one line	What this is and what it decides
The tiles	The three or four numbers the decision turns on
The table	The disaggregation the decision needs
The figure	A shape a table cannot show
The context	Who it is for and what they are deciding
The method and the limits	What it does not show

Every dashboard on this platform is in that order, and the order is the same in both languages because it is about reading rather than about wording.

The two things that usually occupy the top and should not: a logo, and a filter control. A logo tells a reader nothing they did not know, and a filter asks them to do work before the page has told them why.

2.2 The tile

```
56.3%  
Basic service  
improved and within 30 minutes
```

Three parts, and each is doing a job.

The value, large. One number, at the precision the interval supports. 56.3%, never 56.26%.

The label, plain. What it is, in the words the audience uses. "Basic service", not `basic_water_pct`.

The note, small. The definition compressed to a clause — "improved and within 30 minutes" — because the tile is where the misunderstanding starts and one clause prevents most of it.

Colour only where it is a state. The dashboards here use the signal band on a tile only when the number is a severity or a data-quality flag, for the reason the visualisation course gave: spending green, amber and red on ordinary categories teaches the reader to stop reading them as measurements.

2.3 What goes behind a link

The phrase in the spine is "the detail behind a click", and on a static page the click is a link to another artefact rather than an interaction.

On the page	Behind a link
Three headline numbers	Every indicator in the logframe
One disaggregation	Every disaggregation
The current round	The historic series, as a CSV
The finding	The notebook that produced it
What it does not show	The technical documentation that says why

A static page can do all of this with an anchor tag, and it survives being printed, emailed and opened offline — which an interactive filter does not.

This is a real trade and it is worth naming. An interactive dashboard lets a reader ask their own question; a static one answers the question it was built for and points at the artefact for everything else. **For an audience on a metered connection, the second is usually the better bargain**, and it is the one this platform took.

2.4 The five-second test, applied to a page

Show it to somebody who was not in the analysis and ask what it says.

If they read the tiles and state the finding, it works.

If they ask "what am I looking at", the title is a file name. "WASH coverage and water quality by district" tells a reader what it is. "Dashboard" does not.

If they start scrolling immediately, the tiles are not the numbers they need.

If they ask what the number means, the tile has no note.

2.5 Print, and the phone

The two displays a programme dashboard is actually read on.

```
@media print {
  body { background: #fff; padding: 0; font-size: 11pt; }
}
```

A dashboard that cannot be printed is one that cannot go in a meeting pack.

Somebody will print it. It goes into a meeting pack, it gets photocopied, and the version everyone in the room holds is monochrome. The tiles have to survive that, which is the greyscale rule from the visualisation course arriving as a layout constraint.

And somebody will open it on a phone. Tiles in a grid that reflows to one column, a table that scrolls horizontally inside its own container rather than pushing the page sideways, and no fixed width anywhere.

2.6 What the layout cannot fix

A page whose question is wrong. Ordering the wrong three numbers well produces a well-ordered wrong page, and the last lesson is the one that catches that.

A number nobody trusts. Layout does not create confidence; provenance does, and that is the next lesson.

A finding nobody wants. A dashboard cannot make an unwelcome result welcome, and the attempt — burying it below the fold, choosing a flattering disaggregation — is the failure this course ends on.

2.7 Report it whole

Dashboard layout

Reading order: title and purpose, four tiles, district table, figure, commissioning context, method and limits.

Tiles carry a value at the precision the interval supports, a plain label, and a one-clause definition. Signal colours appear only where the value is a severity or a quality flag.

The historic series, the full indicator set and the analysis notebook are linked rather than embedded. The page is 8 kB, prints to one side of A4 in greyscale, and reflows to a single column below 640 px.

The size is worth stating. Eight kilobytes is a number a reader on a metered connection can weigh, and it is a claim the file itself either supports or does not.

2.8 What comes next

A well-laid-out number that nobody can source is a number that gets challenged every month. The next lesson attaches the as-of date, the source file and the refresh cadence to every figure on the page.

Part II

What travels with the number

CHAPTER 3

As of when, from what, and how often

Lesson 3 of 8 · 135 min

The three questions asked of every number on a dashboard, every month, by someone who was not in the analysis. Answering them on the page costs one line each and converts a recurring argument into a recurring glance.

3.1 The three questions

They arrive in this order, from someone who trusts you and needs to defend the number to someone who does not.

"As of when?" A coverage figure with no date is a figure that will be quoted next March.

"From what?" Which export, which version, which extract of the database — and whether it is the same one the last report used.

"How often does it change?" Whether the number in front of them is this morning's or last quarter's, and whether checking again next week is worth doing.

Answer all three on the page. Each is one line, and the alternative is answering them by email eleven times.

3.2 What the answers look like

Source: wash-household-survey-2024.v1.csv
Data as of: 2024-11-30 (survey round 1)
Refreshed: when a new survey round is published; this is a single round

Name the file, not the system. "From DHIS2" is where it came from; the export's filename and version is what somebody can go and get.

Version by filename. .v1.csv is what makes "the same file the March report used" a checkable statement. A correction ships as .v2 and never overwrites, which is the rule the datasets on this platform follow and the reason a notebook pinned to v1 keeps reproducing.

Distinguish the data date from the build date. They are different and only one of them matters to a reader.

	Means	Put it on the page?
Data as of	The last observation in the file	Yes, prominently
Extract date	When the export was pulled	If they differ materially
Build date	When the HTML was written	No

The build date is the one everybody puts on and nobody needs. It changes on every rebuild, tells a reader nothing about the data, and — as the workflows course established — makes the output non-deterministic so a diff stops meaning anything.

None of the seven dashboards here carries a generation date, and that is why rebuilding them produces byte-identical files.

3.3 Refresh, stated as a cadence and an owner

Refreshed monthly, on the 5th, from the screening database export.
Owner: the M&E officer. If the export has not arrived by the 10th, the page still shows the previous month and says so.

A cadence without an owner is a wish. The useful line names who does it and what happens when it does not happen.

Say what a stale page looks like. The failure mode is a dashboard silently showing last month while everyone assumes it is current. A page that states its data date makes staleness visible without any machinery.

3.4 Completeness belongs next to the number

This is the provenance question this sector has that others do not.

```
submitted = reports[reports["report_submitted"]].shape[0]
print(f"reporting completeness: {submitted / len(reports):.1%}")

mean(reports$report_submitted)
```

A coverage figure computed from 76% of facilities is a different number from one computed from 99%, and the epidemiology course spent a lesson on a month where completeness fell to 28.9% and the case count fell with it.

So completeness goes on the page, beside any indicator whose denominator depends on who reported. The vaccination dashboard here carries it as a tile for exactly that reason.

And a month with no reports is drawn as a gap, never as a zero. Zero is a claim that nothing happened.

3.5 What the platform's dashboards state

Source: protection-referrals-2024.v1.csv
How these numbers are produced: computed from the committed dataset file named above by scripts/export/build-project-dashboards.mjs. Nothing is typed by hand and the page contains no script, no external request and no generation date, so rebuilding it on unchanged data produces an identical file.

The last clause is the strongest provenance claim available, and it is checkable by anyone with the repository: rebuild, and git status is empty.

"Nothing is typed by hand" is the part that matters to a reviewer. A number in a caption that was typed can drift from the chart above it, and nothing detects it — which is the failure the visualisation course's last lesson exists to prevent.

3.6 Where provenance is thin, say so

Three cases, and naming them is better than the alternative.

A number from another agency. "Cluster 4W, February submission" is provenance; "cluster data" is not. If you cannot version it, say when you received it and from whom.

A denominator from a projection. Population figures are usually projected from a census, and the projection's base year is the provenance a reader needs. The education course's gross enrolment ratio of 109.2% is a denominator story rather than an enrolment story.

A number somebody adjusted. If a facility's obviously-wrong entry was corrected before analysis, that is a decision and it belongs in the technical documentation, with the page pointing at it.

3.7 Report it whole

Provenance panel

Source	wash-household-survey-2024.v1.csv 2,403 households, 18 communities, 3 districts
Data as of	2024-11-30, survey round 1
Refresh	single round; no monthly cadence. A second round is planned for 2025 and will make a trend meaningful.
Completeness	water quality tested on 33.4% of households; every quality figure states that denominator separately
Produced by	scripts/export/build-project-dashboards.mjs, from the committed file. No hand-entered numbers, no generation date, byte-identical on rebuild.
Definitions	docs/indicator-definitions.md

The last line is the handover to the next lesson. A provenance panel says where the number came from; a definition panel says what it is, and the two together are what stop the same argument recurring monthly.

3.8 What comes next

Two artefacts on this platform computed improved water coverage from the same file and got different answers. The next lesson is about the panel that would have caught it — and about the fact that it did.

CHAPTER 4

The panel that caught a wrong number in this course

Lesson 4 of 8 · 135 min

Two artefacts on this platform computed improved water coverage from the same file and got 82.3% and 79.7%. The technical documentation was right and the lesson was wrong, and nothing except a definition panel and a second computation would have found it.

4.1 What happened

The WASH project has carried an indicator definitions document since it was written. It states improved water coverage as **82.3%** and basic service as **56.3%**, with the numerator, the denominator and the standard for each.

The WASH course lesson, the committed figure and the first version of the project dashboard all said **79.7%** and **54.7%**.

```
# What the lesson had
IMPROVED = {
  "piped-into-dwelling", "piped-into-yard", "public-tap",
  "borehole", "protected-well", "protected-spring",
}

# What the JMP definition actually lists
improved <- c("piped-into-dwelling", "piped-into-yard", "public-tap",
             "borehole", "protected-well", "protected-spring",
             "tanker-truck") # packaged or delivered water
```

The JMP improved list includes packaged or delivered water, which is tanker-truck in this dataset. The lesson omitted it — in a course that names the JMP ladder and SDG 6.1.1 explicitly — and understated improved coverage by 2.6 points and basic service by 1.6.

Nothing detected it for months. The lesson was internally consistent, the figure agreed with the lesson because it used the same set, the tests passed, and the number was plausible. **It surfaced the day a second artefact computed the same indicator from the same file and disagreed.**

4.2 Why the documentation was the one that was right

Because it stated the definition rather than implementing it.

```
### 2.1 Improved source

| Numerator   | Households whose main drinking water source is piped, borehole,
|             | protected well, protected spring, rainwater or packaged water
| Denominator | All surveyed households
| Value       | 0.823
| Valid at    | Household, community, district
| Standard    | WHO/UNICEF JMP service ladders
```

Writing "or packaged water" in prose made it checkable against the standard. A Python set of six strings does not read like anything; a numerator written as a sentence can be compared to the definition it claims to implement.

That is the whole argument for the panel. It is not documentation of the code — it is the specification the code is supposed to satisfy, written in a form a sector specialist can audit without reading Python.

4.3 What a definition entry contains

Six fields, and the last two are the ones usually missing.

Field	Example
Numerator	Households with an improved source and a round trip of 30 minutes or less
Denominator	All surveyed households
Value	0.563
Valid at	Household, community, district
Standard	JMP service ladders; SDG 6.1.1
Decision informed	Which communities receive rehabilitation

"Valid at" prevents the commonest misuse. A number computed at district level and quoted at community level is the same arithmetic answering a different question, and stating the level is what stops it.

"Decision informed" is what justifies the indicator existing. The WASH documentation writes "None. It is reported so the report can show what changed" against improved coverage — which is an unusually honest entry and exactly right: the figure is retained to quantify a gap, not to be used.

4.4 Where the panel lives

Three places, and they are not alternatives.

Where	What it holds
On the dashboard	One clause per tile — "improved and within 30 minutes"
In the technical documentation	The full entry, six fields, per indicator
In the code	A comment naming the standard and the reason

```
# The JMP improved list, which includes packaged or delivered water --
# `tanker-truck` here. Omitting it understates improved coverage by 2.6
# points and disagrees with the project's own indicator documentation.
IMPROVED = {...}
```

```
# The comment is what makes the next reader check rather than assume.
```

The tile clause is what a reader sees; the document is what a challenger reads; the comment is what the next analyst edits. All three, or the three drift.

4.5 The check the panel makes possible

Compute the indicator twice, from different code, and compare. That is what found the error here, and it is a check you can run deliberately rather than by accident.

```
from_pipeline = summary["basic_service"]
from_documentation = 0.563          # the value the definition panel states
assert abs(from_pipeline - from_documentation) < 0.001, (
    f"pipeline {from_pipeline:.3f} disagrees with the definition panel"
)
```

```
stopifnot(abs(basic_service - 0.563) < 0.001)
```

A stated value in the definition panel is a test fixture. Once the panel carries the number, any pipeline that computes it differently fails loudly — which is the workflows course's rule arriving in a place nobody expects it.

And when they disagree, decide which is authoritative before changing either. Here the standard settled it: JMP lists delivered water, so the documentation was right and four other artefacts were corrected to match.

4.6 The argument this prevents every month

The indicator definition panel exists because of a specific recurring meeting.

"Coverage is 82%." — "We reported 76% last quarter." — "That was a different definition." — "Which one is right?"

Both are right and they are different indicators. Without a panel that conversation takes twenty minutes every quarter and is never resolved; with one it takes as long as opening the document.

The commonest three disagreements in this sector, all of which a panel settles:

Source against service. Improved source, 82.3%. Basic service, 56.3%. Twenty-six points, and both are correct.

Numerator population against denominator population. Gross enrolment of 109.2% is a projected denominator, not more children than exist.

Screening against survey. A screening rate is not a prevalence estimate, and the nutrition dashboard here says so in its own "what this does not show" section.

4.7 Report it whole

Indicator definitions

Full definitions in docs/indicator-definitions.md: numerator, denominator, the level it is valid at, the standard applied, and the decision it informs, for each of the nine indicators in this report.

Each dashboard tile carries the definition compressed to one clause.

The pipeline asserts its computed values against the values stated in the definitions document. A disagreement fails the build rather than reaching a report.

```
Change log: 2026-07-29, improved source corrected from 79.7% to 82.3%.  
Delivered water (tanker truck) was omitted from the JMP improved list.  
The definitions document had it right; the analysis code did not.
```

The change log entry is the part worth copying. An indicator that changed definition needs a dated line saying what changed and which direction, or every comparison with an earlier report is silently wrong.

4.8 What comes next

The dashboard answers a question. The report is where the answer gets argued, and the next lesson is about writing one whose findings, limitations and recommendation can be read separately — because they will be.

Part III

The report around it

CHAPTER 5

Findings, limitations, recommendation — in that order and separable

Lesson 5 of 8 · 135 min

Somebody will read only the recommendation, somebody else only the limitations, and a third person will forward one paragraph with no context. Write the report so each of those readings is survivable, because all three happen to every report.

5.1 Who reads what

A programme report has four readers and none of them reads all of it.

Reader	Reads	Needs
The decision-maker	The recommendation	What to do, and what it costs
The technical reviewer	The limitations and the annex	Whether to believe it
The donor	The summary	Whether the money did something
Whoever forwards it	One paragraph	That paragraph to be true alone

Write for the fourth. A report whose paragraphs survive being separated is a report that survives its own circulation, and every other reader is served by the same discipline.

5.2 The order the platform's reports use

Every project report here has the same shape, and it is not the academic one.

```
## Recommendation
## What changed in the measurement
## The ranking
## Why the list can be defended
## Water quality, reported separately
## What this does not establish
## Recommended next steps
```

The recommendation is first. Not the methodology, not the background — the thing you are asking someone to do. A reader who stops after one section should have the answer.

"What this does not establish" is a numbered section, not a footnote. It is where the technical reviewer goes first, and burying it signals that you would rather they did not.

The methodology is not in the report at all. It is in the indicator definitions document, linked, because a reader who wants it wants all of it and a reader who does not wants none.

5.3 The recommendation paragraph

Four things, and it fails without any of them.

What to do. A verb and an object. "Rehabilitate the water points in these six communities."

Why these and not others. One sentence of criterion, not a summary of the analysis.

What it costs, or what it displaces. A recommendation with no cost is a wish.

What would change it. The condition under which this advice stops being the advice — which is what makes it a recommendation rather than an instruction.

Rehabilitate water points in the six communities listed in section 3.

They are the six with the highest combined need score: households above the thirty-minute collection threshold, households below the Sphere quantity minimum, and households on an unimproved source. The score is defined in docs/indicator-definitions.md section 4.

This is the full rehabilitation budget for the cycle. Three of the six were not in the previous list, which was ranked on improved-source counts.

If the second survey round shows the three new communities improving without intervention, the list should be recomputed before the following cycle.

The third paragraph is the one that gets the recommendation accepted or rejected, because it says what changes rather than what is true.

5.4 Limitations that are not apologies

There is a register that undermines a report and a register that strengthens it, and the difference is whether the limitation is quantified.

Not this: "The data have some limitations and results should be interpreted with caution." This says nothing and reads as a disclaimer.

This: "Water quality was tested in 33.4% of households. The quality figures are reported on that subsample and are not comparable with the access figures, which use all 2,403."

Weak	Strong
"Small sample in some areas"	"Three communes have fewer than 200 screenings; their intervals are twice as wide"
"Data quality issues"	"One team's mean sits beyond the SMART concern threshold; its clusters are excluded"
"Results are indicative"	"This survey cannot distinguish one commune from its neighbour; it can distinguish t

A quantified limitation is a finding. It tells a reader what the analysis can and cannot support, which is more useful than a general warning and much harder to ignore.

5.5 The three sentences every report needs

They come from the statistics and impact evaluation courses and they belong in every narrative report this audience writes.

The estimate with its interval, in programme units. "19.4 percentage points, or 38 cases that did not reach a service."

The smallest effect the design could detect. "This evaluation could not have detected anything below 4.9 points."

What the analysis cannot rule out. "Effects between zero and three points — the range the programme is arguing about — are not excluded."

Most reports have the first and not the second or third, and the absence of the second is what lets a null result be read as a refutation.

5.6 Length, and what to cut

Four to six pages for a programme report, and the discipline is in what is not there.

Cut	Because
The methodology section	It is the definitions document, linked
The literature review	Nobody reading this is deciding whether the field is real
Every chart that does not carry a finding	A chart is a claim; a decorative one is a claim about nothing
The background on the crisis	The audience lives there
Anything the annex holds	That is what an annex is

The one thing never to cut is the "does not establish" section. It is the first thing an editor reaches for and the last thing that should go.

5.7 Report it whole

Report structure

Recommendation (1 page), what changed in the measurement, the ranked list, why the list can be defended, water quality reported separately, what this does not establish, next steps.

Methodology is not in the report. Indicator definitions, with numerator, denominator and standard for each of nine indicators, are in docs/indicator-definitions.md and linked from the recommendation.

Each section is written to be readable alone, because the recommendation will be forwarded without the rest of it.

The report states the interval on every estimate and names one thing the analysis cannot establish for each of its three findings.

"Written to be readable alone" is a testable claim. Send one section to somebody and ask what it says.

5.8 What comes next

A report that states its limitations invites a reviewer to check them. The next lesson is about the annex that lets them — and about why an annex full of only successful analyses is the one nobody believes.

CHAPTER 6

An annex of only successful analyses is the one nobody believes

Lesson 6 of 8 · 135 min

The annex exists so a reviewer can check the report without rerunning it. That means it holds the tests that found nothing, the analyses that were run and dropped, and the count of comparisons — because a reviewer who cannot see the denominator cannot weigh the findings.

6.1 What the annex is for

Not completeness — a reviewer will not read forty pages. **The annex holds exactly what somebody needs in order to disagree with you competently.**

The annex holds	So a reviewer can
Every test run, including the null ones	Weigh the findings against the count
The definitions, or a link to them	Check the indicator is what they think
Deviations from the analysis plan	See that changes were decided, not discovered
The exclusions, with counts and reasons	Reconstruct the denominator
The code, or where it is	Rerun it

Everything else belongs in the report or nowhere.

6.2 The test table

Comparison	Test	Statistic	p	Effect
Completion by disability	two-sample z	z = -5.21	<0.001	-19.4 pts
Over-age by sex	two-sample z	z = 1.84	0.066	+5.5 pts
Attendance by sex	two-sample z	z = 2.28	0.023	+0.56 pts
Attendance by feeding	Welch t	t = 3.11	0.008	+5.2 pts
Over-age by sex, 24 schools	two-sample z x24	2 of 24 at p<0.05		0 survive Bonferroni

Comparisons run in total: 29. Findings reported in the body: 2.

Three of those five rows are in the annex precisely because they found nothing. A non-significant gap, a significant irrelevance and a family of null tests are all things a reviewer needs to be able to find.

The last line is the one that changes how the report is read. Two findings from twenty-nine comparisons is a defensible ratio stated openly; two findings with no denominator is a claim a reviewer has no way to weigh.

An annex containing only successful tests tells a reviewer the count was managed, and there is no way to un-signal that.

6.3 Exclusions, with counts

Every exclusion is a decision, and a reviewer reconstructing your denominator needs all of them.

```

audit = []
audit.append(("all rows", len(survey)))
survey = survey[survey["age_months"] <= 60]
audit.append(("age above 60 months", len(survey)))
survey = survey.dropna(subset=["muac_mm"])
audit.append(("no MUAC recorded", len(survey)))

for label, remaining in audit:
    print(f"{label:28} {remaining:>6,}")

```

Build the audit trail as you filter, not from memory afterwards.

Print the count after each step, in order. A reviewer who sees 4,218 in the report and 4,206 in the annex knows exactly which twelve rows went and why.

The commonest silent exclusion is a join. An inner join that drops a third of the data looks exactly like one that drops nothing, which is why the workflows course's most useful assertion is a row count across every join.

6.4 Deviations from the plan

Deviations from the analysis plan

```

Planned: primary analysis on all enrolled students.
Done:    restricted to the 585 with both assessment rounds.
Why:     156 students have no endline. Attrition analysis added; those who
         left scored 39.3% at baseline against 56.8% for those who stayed,
         so the panel is not representative and this is a stated limitation.

```

A listed deviation is a strength and an unlisted one found by a reviewer is the end of the report's credibility, and the difference is four lines written at the time.

6.5 The exploratory section, labelled

Two headings, and the label is the whole mechanism.

B.1 Pre-specified analyses	1 model, reported in the body
B.2 Exploratory analyses	3 models, generating hypotheses only

A subgroup finding is exploratory unless the subgroup was named in advance. The statistics course found two significant schools out of twenty-four on an effect that was exactly zero; a subgroup chosen after seeing the data is the same machine with the label removed.

6.6 What an annex should not contain

Raw data. Especially not here. A protection annex with a case-level table is a disclosure, and the aggregation that prevents it is the same suppression rule the protection course teaches.

Screenshots of code. The code, or the repository, or nothing.

Every chart that did not make the report. An annex is not a graveyard; a chart that carried no finding carries none in the annex either.

Output nobody can read. A pasted regression summary with forty rows of software formatting is not evidence, it is intimidation. Format it as the table the body would have used.

6.7 The reproduction note

The last thing in the annex, and the one that makes everything above it checkable.

Reproduction

```
Code: github.com/<org>/nutrition-surveillance, commit a3f9c21
Data: muac-screening-artibonite-2024.v1.csv, SHA-256 4f2a9c1b8e...
Environment: Python 3.12, dependencies pinned in uv.lock
```

```
uv sync && uv run python run.py
```

```
Running this on a clean checkout reproduces every number in this report
and this annex. Verified 2026-03-14.
```

"Verified" with a date is the claim. The workflows course's whole argument lands here: an analysis nobody can rerun is one whose other properties have to be taken on trust.

6.8 Report it whole

Annex contents

- A. Indicator definitions (linked; numerator, denominator, level, standard, decision informed, for nine indicators)
- B.1 Pre-specified analyses (1)
- B.2 Exploratory analyses (3, labelled, not presented as findings)
- C. All comparisons run: 29, with statistic, p and effect size. Two are reported as findings in the body.
- D. Exclusions: 4,218 screenings, 12 excluded for an implausible age, with the count after each step.
- E. Deviations from the analysis plan: 1, listed with its reason.
- F. Reproduction: repository, commit, data checksum, environment, and the two commands. Verified 2026-03-14.

Six sections and none of them is long. An annex is judged on whether a reviewer can find the thing they doubt, not on its weight.

6.9 What comes next

The report is written and the annex holds the working. The next lesson is the room it gets presented in — specifically the case where the finding is not the one the committee commissioned, which is the hardest thing this course asks anyone to do.

Part IV

The room

CHAPTER 7

Presenting the finding nobody commissioned

Lesson 7 of 8 · 135 min

The committee asked whether the programme works. The answer is that the evaluation could not have told them, the seven-point gain was the school year, and the number in the last three proposals is not an effect. Here is how to say that in a room.

7.1 The three unwelcome findings

They are not equally hard and the difference matters.

"The programme did not do what you thought." Hard, and it is the one people prepare for.

"We cannot tell." Harder, because it sounds like a failure of the analyst rather than of the design. The impact evaluation course's whole last unit is this case.

"The number you have been using is wrong." Hardest, because somebody put it in a proposal, and the room contains the person who did.

All three arrive in this sector regularly, and the difference between a career-limiting meeting and a useful one is almost entirely preparation.

7.2 Lead with the decision, not the finding

The instinct is to build to the result. It is the wrong order in a room.

Not this: methodology, then data, then results, then the awkward part at minute eighteen when half the room has stopped listening and the other half has worked out where it is going.

This: the recommendation, then the one sentence that supports it, then the detail for whoever wants it.

"My recommendation is that you do not expand the programme on the basis of this evaluation, and do not discontinue it either.

Literacy rose seven points in programme schools. It also rose in the schools without the programme, by slightly more. The seven points are the school year.

Separately, this evaluation was not large enough to detect the three-point difference you would act on, which was knowable before it ran. I can show you what a study that could detect it would need."

Sixty seconds, and the room now knows what you are asking for. Everything after that is answering questions rather than delaying the point.

7.3 Bring the thing that would have been true

An unwelcome finding lands very differently when it arrives with what a positive finding would have looked like.

"Here is what we would have seen if the programme were working at the size you care about." Then show that it is not that, on the same chart, with the same axis.

And bring the cost of knowing. "Sixty schools instead of twenty-four" is a concrete, priceable thing, and a committee that cannot act on your finding can act on that.

A recommendation makes an unwelcome finding survivable. Without one the meeting is about whether you are right; with one it is about what to do.

7.4 What not to do

Four moves that feel helpful in the moment and are not.

Do not soften the number. "Roughly flat" for a -0.96 with an interval from -3.5 to $+1.6$ is a kindness that gets quoted as "flat" and then as "slightly positive".

Do not offer the flattering disaggregation. There is always a subgroup where the programme looks like it worked. Offering it is how the two significant schools out of twenty-four end up in a proposal, and the statistics course showed the generator set that effect to exactly zero.

Do not blame the data. "The data quality was poor" is usually false and always sounds like an excuse. If the design was underpowered, say the design was underpowered — that is a decision somebody made, and naming it is how the next one is made better.

Do not apologise for the finding. You did not cause it. Apologising for it invites the room to treat it as your position rather than as the result.

7.5 The questions that will come, and the answers

Question	The answer
"Can we look at it by district?"	Yes, and here is the multiple-comparison count first
"The other agency reports 76%"	Different indicator; here is the definitions panel
"Can you re-run it excluding those schools?"	Yes, as an exploratory analysis, labelled
"So it doesn't work?"	No — the study could not detect an effect of the size you care about
"What do we tell the donor?"	This, in three sentences, and I will draft it

The last row is the one to volunteer. The committee's real problem is often not your finding but what they now have to say about it, and offering the three sentences turns you from the bearer of the problem into part of the answer.

7.6 The number in the last three proposals

The hardest case, and it needs its own handling.

Establish it privately first. Take it to the person who owns the number before the meeting. Not to protect them — to make sure you are right, because you might not be, and a room is the worst place to discover that.

Separate the error from the person. "The improved-source figure omitted delivered water, which the JMP definition includes" is about a set of strings. "The previous analysis was wrong" is about whoever did it.

Say which direction and how much. 79.7% to 82.3% is a correction that makes coverage look *better*, and saying so immediately stops the room bracing for the worst.

Bring the corrected version, the change log entry and the check that would have caught it earlier. A correction with a mechanism attached is a process improvement; one without is an accusation.

Change log: 2026-07-29, improved source corrected from 79.7% to 82.3%.
Delivered water was omitted from the JMP improved list. The definitions

document had it right; the analysis code did not. The pipeline now asserts its computed values against the definitions document, so the two cannot disagree silently again.

7.7 Report it whole

Steering committee, 14 March

Presented: no detectable effect on literacy (-0.96 points, 95% CI -3.5 to $+1.6$), and the finding that the design's minimum detectable effect was 4.9 points against a threshold of interest of 3.

Recommendation: neither expand nor discontinue on this evidence. A design able to answer the question needs about 60 schools; randomising the order of the next roll-out costs nothing and would give a comparison group that does not have to be argued for.

Agreed: the 7-point gain will not be used in the next proposal. The programme team will consider randomising the 2026 expansion order.
Deferred: whether to fund a larger evaluation.

The "agreed" line is the meeting's output, and writing it in the room is what stops the seven-point gain reappearing in June.

7.8 What comes next

The last lesson turns the course on the seven dashboards this platform ships — what each one is for, what each refuses to show, and the one that exists because a number in this very course was wrong.

CHAPTER 8

Seven dashboards, and what each one refuses to show

Lesson 8 of 8 · 135 min

Eight kilobytes, no JavaScript, no external request, byte-identical on rebuild. Each of the seven states the decision it informs and ends by naming what it does not show — and one of them found an error in this platform’s own teaching.

8.1 What is actually shipped

```
pnpm project:dashboards
ls apps/data-analysis/public/projects/*/dashboard/
```

Seven projects, two languages, fourteen files, about 8 kB each. Static HTML built from committed CSVs, with the project’s committed SVG figure inlined.

No JavaScript, no external request, no generation date. Rebuilding on unchanged data produces byte-identical files, which is the property that makes a diff mean something.

8.2 The four constraints that decided the design

The audience works offline and on metered connections. That removed anything needing a live server or a CDN before any other consideration.

The Content-Security-Policy forbids inline script bodies. `script-src` has no `‘unsafe-inline’`, so a script in a static file under `public/` is refused by the browser — silently, which is the failure the slide reader shipped once. Inline styles are allowed, and tiles, tables and a figure need nothing else.

The deploy has no Python, no TeX and no pandoc. Rendering the project notebook would need Quarto, Python, pandas and a network fetch. The builder reads the CSV directly in bare Node instead, so `pnpm project:dashboards` runs anywhere `pnpm` does.

The output has to be reviewable. No clock, no build stamp, no generated identifier — so the page is a function of the data and the project entry.

Three of those four are constraints rather than preferences, and the design is what is left once they are respected. That is usually how a good constraint feels from the inside.

8.3 What each one is for, and what it refuses

Project	Decision	Refuses
WASH coverage	Which six communities get rehabilitation	Unweighted proportions from a clustered survey
Nutrition screening	Where the next round goes	The commune ranking beyond the three highlights
Protection pathway	Which gate to fix first	Four-way disaggregation — nineteen cells of one
Food security	Which instrument the caseload rests on	Averaging the four prevalences; none is an IPC
School attendance	Which schools need a response	Counting the strike period as absence; any cause
SMART survey	Whether a prevalence is publishable	The prevalence itself, until quality control passes
Vaccination	Whether coverage is service or denominator	Months with no report drawn as zero

The refusal column is the one to read first. It is the section each page ends with, and it took longer to write than the numbers.

The SMART row is the strongest of the seven. The dashboard for a nutrition survey declines to show the prevalence, because one team's mean sits beyond the SMART concern threshold and quality control has to pass before a prevalence is publishable. **A dashboard whose main number is withheld, with the reason, is doing its job.**

8.4 The one that found an error

Building the WASH dashboard meant computing improved water coverage from the same file the project's indicator documentation used. They disagreed: **82.3% in the documentation, 79.7% in the dashboard, the lesson and the figure.**

The documentation was right. The JMP improved list includes packaged or delivered water — tanker-truck here — and the course had omitted it, in a course that names the JMP ladder and SDG 6.1.1 explicitly.

Nothing else would have found it. The lesson was internally consistent, the figure used the same set so it agreed, the tests passed, and 79.7% is a plausible number. It surfaced because a second artefact computed the same indicator and was compared.

That is the argument for a definition panel, made by a failure rather than by assertion, and it is why the correction ships with a change-log entry and an assertion in the pipeline rather than only a fix.

8.5 What these dashboards do not do

Being honest about the gaps, as the workflows course insisted.

No interactivity. A reader cannot filter, drill or re-slice. Everything beyond the one question is a link to the notebook or the CSV. For an audience on a metered connection that is the right bargain and it is still a bargain.

No live refresh. The page is rebuilt when the data is, by a script somebody runs. A monthly programme dashboard would need that script on a schedule, and this platform's data does not change.

No access control. They are public static files, which is fine because every dataset here is synthetic — and would not be fine for a protection caseload. **A real deployment of this pattern needs to answer the access question before the first file is written.**

No accessibility audit. They are semantic HTML with a documented reading order and they have not been tested with a screen reader, which is a gap rather than a claim.

8.6 The pattern, portable

Five properties, and any tool can have them.

One file per audience-question, not per data source.

Every number computed by a script from a versioned file, with nothing typed.

Provenance and definitions on the page, not in a separate document nobody opens.

A section naming what the page does not show, written before the page is considered finished.

Deterministic output, so "is this current?" is a question the version control system answers.

8.7 Report it whole

Project dashboards

Seven projects, two languages, 14 files, ~8 kB each. Static HTML built by `scripts/export/build-project-dashboards.mjs` from the committed dataset files, with each project's committed SVG figure inlined.

No JavaScript (the CSP forbids inline script bodies), no external request, no generation date. Rebuilding on unchanged data is byte-identical.

Each page states the decision it informs and ends with what it does not show. `dashboard.ready` is true on all seven projects and `project-files.test.ts` fails if a declared file is missing.

Not done: no interactivity, no scheduled refresh, no access control, no screen-reader audit.

The last line is the one to copy. A dashboard section without a "not done" is a section nobody has audited.

8.8 What comes next

That is the course, and it is the last one in the programme's spine.

Twenty-one courses and one argument. A number in a programme report is a claim, and everything here is about what has to be true for the claim to hold: a denominator you can defend, an interval that says how sure you are, a comparison group that is one, a design that could have detected what matters, a chart that does not say more than the data, a pipeline that produces the same answer when somebody else runs it — and, finally, a page and a report that put all of that in front of the person who has to decide.

The last step is the one the others are for. An analysis that never reaches a decision is an analysis that did not happen, and this audience is judged on defending a number in a review rather than on producing it.

About this edition

This PDF is generated from the same source as the web version of the course, so the two cannot drift. The web edition carries the runnable code, the downloadable datasets and any corrections issued since this file was produced.

Every dataset referenced in this course is synthetic or fully de-identified. No record traces to a real person.

This course is published in English and in French. The French edition is a professional adaptation using the terminology UNICEF, WHO, WFP, OCHA and national ministries publish in — not a literal translation.

Read and run the course online at data-analysis.cassion.dev/courses/dashboards-and-reporting

Course content is licensed CC BY 4.0. You may adapt and redistribute it, including commercially, with attribution to Cassion.

Generated July 29, 2026.