AN EXPERIMENTAL FIELD GUIDE / 2026

THE MISSING.

An atlas of everything the evidence failed to capture.

THE WORLD DOES NOT BECOME COMPLETE JUST BECAUSE THE DATASET DOES.

A long-form inquiry into missing data, imperfect ecological detection, archival silences, and the ethics of making conclusions from partial records. This print guide complements the working browser experiment at sahiltayal.com/missing/.

Original text and design: Sahil Tayal. Published 1 November 2026. Original material: CC BY 4.0 with attribution; third-party works retain their own rights.

HOW TO USE THIS GUIDE

Start with the boundaries. Visit the browser experiment when you reach Chapter 3: it shows visually how a constant population can produce different observed averages under different selection processes. The paper exercises are for reflection, not diagnosis or scoring.

FACT / INTERPRETATION / SIMULATION

Scholarship is attributed in the reading ladder. The hypothetical circuit, institutional examples, and invented archive are illustrative scenarios. The numerical experiment contains 120 made-up observations and no real-world estimates. The conceptual bridge between statistical loss, nondetection and historical silences is editorial synthesis rather than a single established theory.

CONTENTS

01 / The proposition

02 / The boundary

03 / The Absence Engine

04 / The lineage

05 / Competing explanations

06 / The archive is not a camera

07 / Two case files

08 / A serious counterexample

09 / Five mistakes

10 / Who pays for the blind spot?

11 / What This Does NOT Show

12 / Three-column exercise

13 / Annotated reading ladder

14 / Concluding insight

FIELD NOTE 01 / THE PROPOSITION

The proposition

What we know is not a transparent window onto everything that happened. It is the result of a chain: something occurs; some part is noticed; some observation is recorded; a record survives; an analyst retrieves it; someone decides it matters.

A failure at any point may leave an empty place where confident interpretation wants a fact. An empty place is not automatically proof of concealment. Equally, it is not automatically proof that nothing occurred.

This inquiry distinguishes three questions often muddled together: *Was it present? Was it detectable? Was it recorded?*

FIELD NOTE 02 / THE BOUNDARY

The boundary

In statistical analysis, missing data are values that belong to a defined analysis but are unobserved. If a study plans to measure participants' outcomes at six months and one person's outcome is not collected, that is a missing value. But a number that never had a meaningful definition is not necessarily missing: the National Academies, for example, distinguish an undefined post-death quality-of-life outcome from a collected outcome that was simply absent. [2]

Nondetection is different. A species can occupy a site yet go unseen on a survey. The distinction is between the organism's actual presence and the observation process. [4]

Archival silence is different again: a historical event, perspective or testimony may be inadequately documented, selectively preserved, omitted from an archive, or marginalised in later narratives. This is not just a problem of empty spreadsheet cells. [5]

These fields share a family resemblance: the gap between the world and the surviving account. They do *not* share one universal mathematical fix. A statistical assumption does not settle a dispute about history, and a historical interpretation cannot replace a survey design.

FIELD NOTE 03 / THE ABSENCE ENGINE

The Absence Engine

THE WORKING EXHIBIT

The interactive experiment is designed for the browser. Open https://sahiltayal.com/missing/#machine to change the mechanism, missingness pressure, and trial, then reveal the synthetic population. The printed record below documents the model and its limitations.

One population, three observation processes. A fixed set of 120 invented cases contains equal numbers in recorded group A and B. Group B has a higher invented outcome distribution. The true values never change. The interface uses pseudo-random numbers to determine which observations are hidden.

RANDOM LOSS / MCAR

The loss probability is the same for all cases and independent of their group and value. A particular observed average may differ by chance.

GROUP DEPENDENCE / MAR

Group B has a higher chance of nonresponse than group A; conditional on recorded group, loss does not depend on the hidden outcome. The unadjusted sample composition can shift the observed mean.

VALUE DEPENDENCE / MNAR

Higher outcome values have higher loss probability even conditional on recorded group. The incomplete sample alone cannot reveal its own missing high outcomes.

AN IMPORTANT NUANCE

The slider controls the mechanism's loss probabilities, not an exact guarantee about the percent missing. The model is a teaching device, not a validated empirical estimator.

WORKED EXAMPLE / RECORDED-GROUP SELECTION

Imagine 100 invented records: 50 in group A with an outcome of 40 each and 50 in group B with an outcome of 80 each. The full-population average is 60. Suppose only 45 group-A records and 10 group-B records are observed. The raw complete-case average is (45 × 40 + 10 × 80) / 55 = 47.3, far below 60. If group membership and the 50/50 population split are known, reweighting the observed group means equally recovers 60 in this simplified illustration. The demonstration assumes all observed and unobserved values within a group are identical; real missing-data problems rarely offer such certainty.

THE ABSENCE ENGINE / MACHINE 001

One synthetic population of 120 cases. The underlying values do not change. You change how observations become unavailable. Watch the reported average change — sometimes through chance, sometimes through systematic selection.

How to read the machine. MCAR can give an unrepresentative sample by chance, but its complete cases are not systematically selected on the modelled variables. MAR permits missingness to depend on observed information; careless complete-case averages can still be biased. MNAR lets missingness depend on the unobserved value itself, making recovery especially assumption-sensitive. This demonstration illustrates mechanisms, *not* real-world bias magnitudes or a validated estimator. [1] [2] [3]

Try this: keep the pressure fixed and switch between the three mechanisms. Then reveal the hidden values. Ask whether a bigger visible sample would necessarily repair a patterned absence.

FIELD NOTE 04 / THE LINEAGE

The lineage

Before data science had a name

Historians have long read sources against their omissions, and naturalists have long worried about creatures that were present but not counted. The contemporary challenge is not the discovery that observation is incomplete. It is learning *which kind* of incompleteness we face, and what follows.

1976 / a statistical turning point

Donald B. Rubin's article *Inference and Missing Data* helped formalise conditions under which a missingness process could be ignored in specified forms of inference. Popular summaries reduce this to a three-part vocabulary — missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). The technical conditions are more exacting than the everyday phrase “random missingness” suggests. [1] [2]

1995 / the making of an archive

Michel-Rolph Trouillot's *Silencing the Past* examined how unequal access to making sources, assembling archives, constructing narratives and assigning later significance shapes historical memory. The issue is not just that a historian has too few documents. Institutions and social positions help determine which documents have a chance of existing. [5]

2002 / ecology learns to count the unseen

MacKenzie and colleagues developed occupancy models that explicitly account for imperfect detection. In their formulation, failing to see a species is not identical to proving that it does not occupy a site. Repeated, properly designed surveys help estimate detection probability rather than pretending it is perfect. [4]

Interpretation: the disciplines differ in object and method, but all challenge the same shortcut: treating the surviving observation as a complete description of the world.

FIELD NOTE 05 / COMPETING EXPLANATIONS

Competing explanations

Chance removal

Some records are lost independently of the variables of interest. A small remaining sample can still deviate from the full population simply through sampling variability. This is not a licence to call an actual dataset MCAR without checking the study design and plausible alternative mechanisms.

Observable selection

A follow-up response can depend on a known attribute — such as the recorded treatment group or location. Conditioned on suitable observed information, the missingness mechanism may meet an MAR assumption. That does not guarantee a raw average of observed outcomes is representative. [2]

Hidden selection

People with especially high or low unobserved values may be more or less likely to report them. Because the selection depends on what is not observed, several different underlying worlds can fit the same visible records. Conclusions need assumptions, sensitivity analysis, and sometimes new data. [3]

Institutional production

A person may be absent from an archive because they lacked access to institutions that could record their experience; a surviving record may be excluded from canonical histories. This is a historically situated interpretation, not a formal missing-data category. [5]

FIELD NOTE 06 / THE ARCHIVE IS NOT A CAMERA

The archive is not a camera

An interpretation inspired by Michel-Rolph Trouillot, not a formal probabilistic model. Four stages at which a history can become partial:

1. Making sources

Whose words, acts and testimony were documented at the time?

2. Assembling archives

Who decided which traces were worth preserving?

3. Retrieving narratives

Which surviving sources do researchers actually consult?

4. Assigning significance

Which stories enter textbooks, memorials and public memory?

Explore four different places where a historical account can lose part of its world. Inspired by Trouillot's analysis of the production of history; these fictional examples are interpretive, not empirical reconstructions.

FIELD NOTE 07 / TWO CASE FILES

Two case files

The animal you did not see

A conservation team visits a wetland and does not encounter a particular amphibian. A hasty report might label the site unoccupied. But an animal may be hidden, silent, inactive, or missed because conditions were unfavourable. The observed outcome of a visit is therefore a blend of ecological presence and detection probability.

In their 2002 *Ecology* paper, MacKenzie and colleagues developed a likelihood-based approach to estimating site occupancy when detection is imperfect. Using monitoring data for amphibians at Maryland wetlands, the paper illustrated how accounting for detection can change estimated occupancy relative to the naïve proportion of places with sightings. [4]

Notice the scientific design principle: repeated surveys under assumptions about closure and detection provide information that a single zero cannot. A model does not conjure the animal out of absence. It uses the pattern of detections and nondetections across visits to estimate a latent state with uncertainty.

What this demonstrates: an absent sighting is not a clean absence of the target. **What it does not demonstrate:** that any claimed unseen species must be present, or that occupancy estimates are valid without their sampling assumptions.

The revolution a dominant narrative struggled to see

The Haitian Revolution poses a different challenge. Between the late eighteenth and early nineteenth centuries, enslaved and formerly enslaved people played decisive roles in a revolution that transformed Haiti's political order. Yet later accounts of the Atlantic world have often marginalised the event or treated it as difficult to fit within inherited categories of who could enact revolutionary history.

In *Silencing the Past*, Michel-Rolph Trouillot uses the Haitian Revolution to examine how silences and unequal power shape both historical sources and the narratives built from them. His argument is not simply that an event was forgotten, nor that every historian intentionally erased it. It concerns how the categories and institutions used to produce history can make some events difficult to render intelligible. [5]

A historian must therefore ask several separate questions. What contemporary traces survive? Who could create them? Which repositories preserved them? What assumptions shaped their interpretation? Which stories later acquired institutional authority? The silences are a research problem, not a shortcut to claims unsupported by archives.

What this demonstrates: a historical narrative can be systematically partial even when built from genuine records. **What it does not demonstrate:** that any preferred alternative story becomes true just because existing archives are incomplete.

FIELD NOTE 08 / A SERIOUS COUNTEREXAMPLE

A serious counterexample

Imagine an automated instrument that records whether a controlled electrical circuit was closed during a specified interval. Its sensors are calibrated, continuously logging, independently checked, and every expected timestamp is present. If the instrument registers no closure during that interval, the evidence can support a strong negative finding — subject to specified detection limits, calibration and failure checks.

Likewise, in an ecological setting, repeated high-quality surveys may make a species' presence increasingly unlikely. The correct lesson from imperfect detection is not “absence can never be shown.” It is that claims about absence should be calibrated to how likely evidence would have been observed *if the thing were present*.

This is an illustrative counterexample, not a claim of infallible instrumentation. It shows why the project's thesis should not be confused with blanket skepticism. We need not refuse conclusions; we need conclusions proportionate to observation quality.

> *Not every blank hides a suppressed truth. Sometimes it marks a tested negative. Sometimes it marks an unanswered question. The job is to distinguish them.*

FIELD NOTE 09 / FIVE MISTAKES

Five mistakes

Not necessarily. In the technical MAR framework, missingness may depend on observed variables; what matters is the relevant conditional dependence on missing values. [2]

If the mechanism persistently excludes the same sort of case, collecting more of the same observed cases can produce very precise estimates of the wrong quantity.

Reporting may depend on access, incentives, detection, resources or archival survival. Importance and observability are different properties.

Omission can be deliberate, routine, accidental, structural, or simply the consequence of limited measurement. Motive needs independent evidence.

Some missingness is not identifiable from surviving data alone. Sensitivity analyses reveal dependence on assumptions; they do not automatically certify the lost truth.

FIELD NOTE 10 / WHO PAYS FOR THE BLIND SPOT?

Who pays for the blind spot?

When an institution uses records to allocate attention, money or protection, its omissions can have material consequences. Consider a health system that learns primarily from people able to

attend follow-up visits, a city that receives infrastructure complaints mostly from neighbourhoods with better connectivity, or a workplace survey in which the least secure workers hesitate to reply. These are plausible scenarios, not claims that any named institution behaves this way.

The central ethical problem is not simply “collect more data.” More collection may invade privacy or expose vulnerable people. A responsible investigation asks whether participation is safe, whose preferences matter, whether exclusions can be measured, and whether people can challenge conclusions drawn from partial records.

Different disciplines offer different remedies. Statistical analysis can model assumptions and use sensitivity checks. Field ecology can improve sampling design. Archives can preserve provenance, diversify collections and foreground uncertainties. Governance can make appeals possible. Each remedy has trade-offs; there is no single tool that repairs all absences.

Connections across the border

Machine learning: datasets represent collection processes, not society in miniature. **Medicine:** dropout can distort outcomes and interpretation. **Climate and ecology:** a missing observation may reflect sensor coverage or imperfect detection. **Organisational behaviour:** silence in a survey may reflect multiple mechanisms, not one psychological motive. **History:** who had the capacity to create and preserve a trace shapes inherited narratives.

These connections are conceptual bridges. They invite domain-specific research; they should not be treated as evidence that the same mechanism operates identically in every setting.

FIELD NOTE 11 / WHAT THIS DOES NOT SHOW

What This Does NOT Show

- This production does not establish that everything absent was deliberately hidden or historically suppressed.
- The Absence Engine uses synthetic records, invented scores and programmed selection mechanisms. Its numbers are illustrations, not empirical estimates, simulations calibrated to a real study, or forecasts.
- The three statistical mechanisms do not exhaust the social, ethical and historical meanings of silence. Trouillot's framework is not a statistical model.
- From observed data alone, an analyst usually cannot prove that the MAR assumption holds rather than an MNAR alternative. Good judgment requires design knowledge and sensitivity analysis. [3]

- Imperfect detection is not evidence of certain presence; reliable negative evidence remains possible when a detection process is strong and explicitly evaluated.
- Stories about excluded voices demand careful primary-source research. The moral appeal of a hypothesis is not proof of its factual details.

This work is an educational synthesis, not an original peer-reviewed study, clinical instrument, forensic method or formal training in missing-data estimation.

FIELD NOTE 12 / THREE-COLUMN EXERCISE

Three-column exercise

Choose a decision, memory, headline, dataset or workplace report that seems complete. Do not try to “catch” anybody. Instead, distinguish what the record shows from what it cannot yet tell you.

No score. No prescription. Keep it private or print a blank page in the Field Guide.

THE THREE COLUMNS

OBSERVED

What is actually in the record? Which primary source supports it?

UNKNOWN

What might matter but was not observed, detected, or preserved?

TESTABLE

Which observation or comparison would reduce the uncertainty?

DISCONFIRMING QUESTION: What would persuade me that the missingness does not change the conclusion?

FIELD NOTE 13 / ANNOTATED READING LADDER

Annotated reading ladder

[2] National Research Council, 2010.

The Prevention and Treatment of Missing Data in Clinical Trials, relevant introductory chapter. Defines missing outcomes and offers a careful overview of MCAR, MAR and MNAR. Best first read for practical distinctions.

[3] University of Melbourne, Statistical Consulting Centre.

Missing data. Accessible practical discussion of missingness and why "ignorable" does not mean harmless. Useful for thinking about assumptions and choices of analysis.

[1] Rubin, Donald B., 1976.

"Inference and Missing Data." *Biometrika*, 63(3), 581–592. Foundational work on conditions for ignoring missingness processes under specified inferential approaches. Technical reading.

[4] MacKenzie and colleagues, 2002.

"Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One." *Ecology*, 83(8), 2248–2255. Shows the logic of accounting for imperfect detection in occupancy research.

[5] Trouillot, Michel-Rolph, 1995.

Silencing the Past: Power and the Production of History. Beacon Press. A major historical and anthropological account of silences at several stages of producing historical knowledge.

DIRECT SOURCE LINKS / VERIFIED AT AUTHORING

https://www.nationalacademies.org/read/12955/chapter/3

https://scc.ms.unimelb.edu.au/resources/preparing-your-data/missing-data

https://doi.org/10.1093/biomet/63.3.581

https://www.usgs.gov/publications/estimating-site-occupancy-rates-when-detection-probabilities-are-less-one

https://nationalhumanitiescenter.org/fellows-books/silencing-the-past-power-and-the-production-of-history/

FIELD NOTE 14 / CONCLUDING INSIGHT

Concluding insight

Good inquiry does not fill every silence with a story. It learns to tell a missing fact from a negative finding — and to keep the question open when the distinction is not yet known.

METHODS & ATTRIBUTION

The original browser model uses a fully fabricated population and deterministic pseudo-random selection. It is intended to make conditional missingness easier to reason about. It does not implement multiple imputation, inverse-probability weighting, occupancy estimation, or historical source verification. In particular, identifying MNAR from incomplete outcomes alone is generally not possible without further assumptions or information. This is an educational artwork, not statistical advice.

Original textual material and visual design licensed Creative Commons Attribution 4.0 International (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/. Attribution: “The Missing — A Sahiltayal.com Production, by Sahil Tayal,” with link to https://sahiltayal.com/missing/. Third-party research, titles and quotations are not relicensed by this guide.

PUBLICATION NOTE

Prepared 1 November 2026. Primary editorial research references are linked in the Reading Ladder. No claims of peer review, original ecological research, or empirical data collection are made. Refer to original publications for technical methodological detail.